

ANALISIS MOLEKULER DAN BIOINFORMATIKA



Dewi Rahmawati
Hasyati Shabrina



Published by
SEAMEO BIOTROP
Southeast Asian Regional Centre for Tropical Biology
Bogor, Indonesia
www.biotrop.org

ANALISIS MOLEKULER DAN BIOINFORMATIKA

**Disusun oleh:
Dewi Rahmawati
Hasyiyati Shabrina**

**SEAMEO BIOTROP
SOUTHEAST ASIAN REGIONAL CENTRE FOR TROPICAL BIOLOGY
2024**

Analisis Molekuler dan Bioinformatika
ISBN: 978-979-8275-99-9

Copyright ©2024 SEAMEO BIOTROP

Penulis : Dewi Rahmawati, MSi dan Dr. Hasyati Shabrina

Editor : Sri Ismawati Soerianegara, MSc

Desain Sampul : Dewi Rahmawati, MSi

Cetakan Pertama: September 2024

Buku ini diterbitkan dengan Dana DIPA SEAMEO BIOTROP 2024

Buku ini tidak diperjualbelikan.

Hak cipta dilindungi oleh undang-undang.

Dilarang memperbanyak buku ini tanpa izin tertulis dari penerbit.

Diterbitkan oleh:



KATA PENGANTAR

Puji dan syukur kami panjatkan ke hadirat Allah SWT karena hanya dengan kasih sayang dan rahmat-Nya, maka Buku Analisis Molekuler dan Bioinformatika ini dapat diselesaikan dengan baik. Buku ini disusun bagi kalangan pengguna laboratorium atau praktisi, serta mahasiswa dengan tujuan untuk meningkatkan pemahaman dan *skill*/keterampilan dalam mengolah data molekuler dari hasil data primer laboratorium maupun data sekunder. Kombinasi kemajuan bioteknologi dan teknologi informasi yaitu Bioinformatika merupakan penggunaan IT untuk menyelesaikan permasalahan yang berhubungan dengan Biologi Molekuler, yang dilatarbelakangi oleh ledakan data observasi biologi sebagai hasil yang dicapai dari kemajuan bioteknologi. Eksperimen biologi molekuler telah mengumpulkan banyak sekali data ekspresi genetika. Data dalam jumlah banyak tersebut, perlu diolah menjadi informasi yang berguna. Perkembangan Biologi Molekuler yang sangat pesat tersebut, maka data hasil eksperimen laboratorium/basah semakin lama menjadi berlimpah. Buku ini diharapkan dapat memberikan tambahan pengetahuan cara mengakses dan pemanfaatan sistem bioinformatika melalui metode analisis data molekuler dengan *software* terkini, serta pengorganisasian data agar hasil analisis yang diperoleh semakin akurat.

SEAMEO BIOTROP memiliki laboratorium bioteknologi yang bergerak di bidang molekuler, salah satu *output* nya adalah data-data molekuler. Bioinformatika dapat digunakan dalam manajemen informasi di bidang penyimpanan data *in silico* dari eksperimen biologi molekuler. Pemahaman dalam bidang bioinformatika sangatlah penting untuk meningkatkan kemampuan komputasi dan analisis data melalui pengadaan sumber daya komputasi dan penggunaan kluster informatika publik.

Masukan yang membangun sangatlah penulis harapkan untuk perbaikan Buku Analisis Molekuler dan Bioinformatika ini di masa mendatang. Akhir kata, penulis berharap semoga isi buku ini dapat bermanfaat bagi para pembacanya.

Bogor, September 2024

Penulis

SAMBUTAN DIREKTUR SEAMEO BIOTROP

SEAMEO BIOTROP (*Southeast Asian Regional Centre for Tropical Biology*) adalah lembaga regional yang bergerak dalam bidang biologi tropika. Mandat SEAMEO BIOTROP adalah penelitian, pelatihan dan diseminasi hasil penelitian biologi tropika. SEAMEO BIOTROP aktif dalam penyebaran informasi untuk peningkatan pengetahuan dan kompetensi dan berperan penting dalam program-program strategis yang memiliki pengaruh terhadap kemajuan pendidikan nasional dan regional ke depan. Sebagai lembaga yang telah berkiprah dalam hal menjembatani transfer ilmu dan teknologi bagi masyarakat umum, sekolah-sekolah, maupun universitas. SEAMEO BIOTROP yang telah membuktikan komitmennya dalam meningkatkan kompetensi bagi komunitas di tingkat nasional dan Asia Tenggara. Salah satunya penyebaran hasil-hasil dan penerapan teknik biologi molekuler di Laboratorium Bioteknologi melalui penerbitan buku *Analisis Molekuler dan Bioinformatika*. Buku ini terbit berdasarkan dengan kebutuhan bioteknologi modern yang dirasakan semakin penting akhir-akhir ini seiring dengan semakin berkembangnya program Bioinformatika.

Buku *Analisis Molekuler dan Bioinformatika* ini merupakan penerapan teknologi informasi pada bidang biologi dengan mengorganisasikan dan menganalisis data-data genetik dari *database* DNA, RNA dan protein menjadi sebuah informasi biologi yang bermakna.

Dalam rangka mendukung kebijakan Kementerian Pendidikan, Kebudayaan, Riset dan Teknologi Republik Indonesia terkait kebijakan Merdeka Belajar dan Kampus Merdeka, SEAMEO BIOTROP menerbitkan Buku *Analisis Molekuler dan Bioinformatika* ini. Saya mengucapkan terima kasih kepada penulis yang telah meluangkan waktu, tenaga, dan pemikirannya untuk menulis Buku *Analisis Molekuler dan Bioinformatika* ini. Buku ini diharapkan dapat memberi manfaat, membuka wawasan dan meningkatkan kompetensi bagi para pembacanya.

Terimakasih.

Bogor, September 2024
Direktur SEAMEO BIOTROP

Dr. Zulhamsyah Imran

TESTIMONI PEMBACA



Nama : Dr Deden D Matra, SP, MAgr

Instansi : Departemen Agronomi dan Holtikultura, Fakultas Pertanian. IPB University

Buku Analisis Molekuler dan Bioinformatika ini sudah sangat lengkap menyampaikan dasar-dasar molekuler dan aplikasinya khususnya dalam penggunaan komputasi dalam biologi molekuler terkini yaitu pemanfaatan *Next-Generation-Sequencing*. Semoga buku ini dapat mengantarkan calon-calon ahli biologi molekuler dan bioinformatika menjadi ilmuwan hebat.



Nama : Dr Fitri Indriani, SHut.

Instansi : Badan Riset dan Inovasi Nasional.

Buku ini dapat dijadikan sebagai panduan untuk analisis molekuler DNA dan RNA *sequencing*. Langkah-langkah pengolahan data-data molekuler disajikan dengan menggunakan *software* terkini yang dapat dengan mudah diikuti oleh pembaca. Buku ini akan sangat berguna bagi mahasiswa, peneliti maupun praktisi yang bekerja dibidang biologi molekuler.



Nama : Dr. (cand.) Aditya Nugroho.

Instansi : Mahasiswa Doktoral Prodi Silvikultur Tropika - IPB University.

Buku ini begitu lengkap dan mendetail, menjabarkan tahap-tahap penggunaan *software* dengan runut. Buku ini sangat membantu untuk pemula yang sedang belajar molekuler dan bioinformatik.

DAFTAR ISTILAH

A

Accession adalah nomor pengidentifikasi unik di *Database GenBank*.

Adenin (A) adalah salah satu dari dua basa N purina yang digunakan dalam membentuk nukleotida dari asam nukleat DNA dan RNA. Pada DNA, adenin berikatan dengan timin melalui dua ikatan hidrogen untuk membantu menstabilkan struktur asam nukleat.

Alignment adalah suatu metode atau cara untuk mensejajarkan.

Amino acid sequence adalah urutan asam amino yang muncul sebagai heliks (huruf besar) atau lembaran (huruf kecil) yang diidentifikasi, di mana lebar setiap segmen adalah representasi kasar dari jumlah asam amino di setiap segmen.

Amplicon length adalah panjang produk PCR. Panjang yang ideal berkisar antara 100-500 pasang basa.

Annotation adalah proses untuk mendapatkan informasi terkait identitas biologis suatu gen.

Asam nukleat adalah biomolekul berukuran besar sebagai informasi genetik yang dapat diwariskan dari satu generasi ke generasi berikutnya.

Assembly adalah proses untuk menyatukan sekuens pendek yang dihasilkan menjadi sekuens transkriptom-transkriptom yang utuh.

Avoid Cross Homology dilakukan untuk menghindari *cross homology* dengan cara menganalisis homologi primer dengan DNA genome melalui BLAST-NCBI. *Cross homology* yaitu kejadian primer menempel pada daerah di luar gen target.

B

Base pair (bp) adalah satuan panjang utas DNA yang berupa satu nukleotida.

Biologi molekuler adalah cabang ilmu pengetahuan yang mempelajari hubungan antara struktur dan fungsi molekul-molekul hayati, serta kontribusi hubungan tersebut terhadap pelaksanaan dan pengendalian berbagai proses biokimia.

Bioteknologi adalah cabang ilmu biologi yang mempelajari pemanfaatan makhluk hidup (bakteri, fungi, virus, dan lain-lain) maupun produk dari makhluk hidup dalam proses produksi untuk menghasilkan barang dan jasa yang dapat digunakan oleh manusia.

Bioinformatika adalah ilmu yang mempelajari penerapan teknik komputasional untuk mengelola dan menganalisis informasi biologis.

Bit score adalah skor yang mengukur kesamaan sekuens terlepas dari panjang sekuens dan ukuran *query* pada *database*.

BLAST (*Basic Local Alignment Search Tool*) adalah program yang digunakan untuk analisis pensejajaran atau suatu program untuk pencarian kemiripan sekuens (*sequence similarity*) dan merupakan alat dalam identifikasi gen dan karakter genetik. BLAST juga dapat digunakan untuk melakukan pencarian sekuens melalui perbandingan dengan *database* DNA dalam waktu singkat.

Bootstrap adalah metode untuk menguji seberapa baik model dengan cara *resample* kolom pada sekuens-sekuens yang disejajarkan untuk membentuk pensejajaran baru.

C

Coverage adalah jumlah rata-rata sekuens yang meng-cover area/tumpang tindih tertentu pada genom target.

Conserved domain (CD) adalah unit yang berulang-ulang sebagai akibat evolusi molekuler yang dapat dilihat dari analisis sekuens (protein atau nukleotida) maupun struktur (protein); mengandung pola atau motif yang dapat terdeteksi ketika dilakukan sekuensing (nukleotida atau protein).

CAP3 adalah perangkat lunak yang digunakan untuk meng-assemble *contig*.

CDHIT adalah perangkat lunak yang mengklusterisasi dan membandingkan sekuens protein atau nukleotida.

Coding sequence (daerah pengkodean gen) adalah bagian dari DNA atau RNA gen yang mengkode protein yang mempelajari panjang, komposisi, regulasi, penyambungan, struktur, dan fungsi daerah pengkode dibandingkan dengan daerah non-pengkode pada spesies dan periode waktu yang berbeda dapat memberikan sejumlah besar informasi penting mengenai organisasi gen dan evolusi prokariota dan eukariota.

Contig adalah DNA/RNA yang saling *overlap* (tumpang tindih) yang akan disusun menjadi satu *full length genome/transkriptom*.

Corset adalah perangkat lunak untuk mengklusterisasi transkrip yang sudah di-assembly secara *de novo* serta menghitung *reads* yang *overlap*.

Cross Dimer adalah primer yang berikatan dengan primer pasangannya (*reverse* dan *forward*).

D

Database adalah data sekuens DNA atau protein yang didapatkan melalui percobaan laboratorium yang biasanya disimpan di dalam file komputer. Setiap file dari suatu sekuens berisi informasi mengenai asal organisme, nama sekuens dan juga nomor akses yang digunakan untuk mengidentifikasi sekuens tersebut.

Database nukleotida adalah suatu koleksi sekuens dari beberapa sumber, termasuk di antaranya *GenBank*, *Reference Sequence* (RefSeq), *Third Party Annotation* (TPA) dan *Protein Data Bank* (PDB).

Data biner adalah sistem bilangan biner atau sistem bilangan basis dua yang merupakan suatu sistem penulisan angka dengan menggunakan dua simbol, yaitu 0 dan 1.

Data *genotipe* terdapat dalam bentuk *data* genetik seperti DNA atau RNA.

De novo sequencing adalah *sequencing* transkrip dari spesies yang belum memiliki *reference*.

Differential expression analysis, yaitu analisis perbedaan ekspresi yang dihasilkan dari individu yang transkriptomnya dibandingkan.

Differential Expression Gene (DEG) yaitu analisis untuk mengidentifikasi kandidat gen-gen yang diekspresikan secara signifikan pada jaringan organisme.

DNA (*Deoxyribonucleic acid*) adalah sejenis biomolekul yang menyimpan dan menyandi instruksi-instruksi genetika dari setiap organisme yang berperan penting dalam pertumbuhan, perkembangan, dan fungsi organisme. Molekul DNA terdiri dari dua unit biopolimer yang berpilin satu sama lainnya membentuk heliks ganda.

DNA *barcoding* adalah suatu metode pengidentifikasian spesies dengan menggunakan urutan gen pendek dari genom organisme.

DNA *ladder* adalah *ladder* yang digunakan untuk menentukan ukuran fragmen DNA dengan elektroforesis di laboratorium biologi molekuler.

Domain adalah unit fungsional dan/atau struktural yang berbeda dari protein; unit lipat secara independen dari polipeptida yang membawa fungsi spesifik.

Down-regulated adalah ekspresi gen yang lebih rendah pada sampel yang diberi perlakuan ketika dibandingkan dengan ekspresi gen pada sampel kontrol atau dengan perlakuan sebaliknya.

E

EdgeR adalah *package* pada *software* R yang digunakan untuk menganalisis perbedaan ekspresi gen.

Entrez adalah sistem pencarian informasi dalam NCBI yang menyediakan akses terintegrasi untuk melakukan sekuensing, pemetaan (*mapping*), taksonomi dan data struktural. *Entrez* juga menyediakan gambaran grafis untuk *mapping* sekuens dan kromosom. Ciri khas dan keunggulan *Entrez* adalah kemampuan untuk pencarian informasi terkait sekuens, struktur dan referensi.

E-value adalah Nilai E (nilai yang diharapkan) yang merupakan angka untuk menjelaskan berapa kali Anda mengharapkan kecocokan secara kebetulan dalam *database* sebesar itu. Semakin rendah nilai E, semakin signifikan kecocokannya.

Exon adalah urutan nukleotida yang diekspresikan. *Exon* berada di sisi intron. Sedangkan intron merupakan nukleotida yang tidak dapat diekspresikan, sehingga harus dihilangkan selama proses ekspresi gen.

F

Filogeni adalah bidang ilmu yang menjelaskan perkembangan makhluk hidup dalam evolusi melalui pertimbangan morfologi dan molekuler. Filogeni merupakan gambaran klasifikasi yang menunjukkan hubungan kekerabatan suatu spesies dengan nenek moyang dan hubungan evolusioner antara organisme.

Frekuensi alel adalah proporsi ataupun perbandingan keseluruhan *copy* gen yang terdiri dari suatu varian gen tertentu.

Format FASTA adalah format berbasis teks untuk mewakili sekuens nukleotida atau sekuens protein, di mana pasangan basa atau asam amino yang diwakili menggunakan kode huruf tunggal. Penamaan dan penjelasan dibedakan dari sekuens yang diletakkan pada baris pertama yang didahului oleh simbol “>”.

F-statistik atau uji F merupakan metode pengujian statistika yang dilakukan secara serentak dengan dua atau lebih objek sebagai pembanding. Uji statistika F ini digunakan untuk menguji hipotesis.

G

GenBank adalah *database* sekuens genetik dalam biologi molekuler berupa koleksi sekuens DNA yang dapat diketahui oleh publik.

Gen adalah bagian dari kromosom yang berfungsi sebagai unit pewarisan sifat bagi organisme hidup.

Gen terbentuk dari sejumlah asam nukleat yang tersusun dalam makromolekul. Bentuk fisik gen berupa urutan DNA yang menyandi suatu protein, polipeptida, atau seuntai RNA.

Genetika adalah ilmu tentang pewarisan sifat.

Gene Ontology adalah metode pengorganisasian gen untuk mendeskripsikan fungsi gen dan produk dari gen tersebut. Gen diklasifikasikan ke dalam 3 domain yaitu: fungsi molekuler, proses biologis, dan lokasi seluler.

Genom adalah keseluruhan informasi genetik yang dimiliki oleh suatu sel atau organisme, atau khususnya keseluruhan asam nukleat yang memuat informasi tersebut. Secara fisik, genom dapat terbagi menjadi molekul-molekul asam nukleat yang berbeda (sebagai kromosom atau plasmid), sementara secara fungsi, genom dapat terbagi menjadi gen-gen.

Genomics adalah setiap usaha untuk menganalisis atau membandingkan seluruh komponen genetik dari satu spesies atau lebih.

Guanin adalah satu dari dua basa N yang menyusun DNA dan RNA.

H

Hairpins adalah interaksi intramolekuler dalam primer.

Heterozigositas adalah jumlah rata-rata individu dengan lokus polimorfik.

Heterozigot adalah alel yang tidak sama pada satu atau lebih lokus.

Homology alignment adalah pensejajaran sekuens yang hanya terjadi ketika dua atau lebih sekuens homolog atau memiliki kesamaan dengan derajat tertentu.

Homozigot adalah salah satu keadaan genotipe di mana alel-alel identik dalam satu lokus.

Homozigositas adalah jumlah rata-rata individu dengan lokus monomorfik.

I

Identity adalah jumlah nukleotida dalam urutan *query* yang identik dengan urutan subjek.

Intron adalah setiap urutan nukleotida di dalam gen yang dibuang di dalam penjalinan RNA selama pematangan produk RNA akhir.

Intra-alelik adalah interaksi alel-alel pada lokus yang sama.

Isoform adalah dua atau lebih protein fungsional dengan fungsi yang sama, tetapi dengan susunan asam amino yang berbeda.

Jarak genetik adalah perselisihan genetik antara spesies atau antara populasi dalam satu spesies tertentu. Jarak genetik diukur dengan berbagai parameter. Jarak genetik yang kecil menunjukkan hubungan genetik yang dekat. Jarak genetik yang besar menunjukkan hubungan genetik yang jauh.

J

Jarak genetik adalah tingkat perbedaan gen (perbedaan genomik) pada suatu populasi atau spesies yang diukur melalui kuantitas numerik. Hasil jarak genetik digunakan untuk mengetahui pohon filogenetik. Jarak genetik yang kecil menunjukkan hubungan genetik yang dekat, sedangkan jarak genetik yang besar menunjukkan hubungan genetik yang jauh.

K

KEGG (*Kyoto Encyclopedia of Genes and Genomes*) adalah *database* yang berisi data-data terkait *high-level function* dan kegunaan suatu sistem biologi seperti sel, organisme, dan ekosistem berdasarkan informasi molekuler, khususnya *dataset* besar hasil sekuensing.

K-mer adalah motif dengan panjang *k* yang diamati lebih dari satu kali dalam hasil sekuens genom. K-mers digunakan untuk mencari dan meng-cover pengulangan dan elemen seluler, pemrosesan awal FASTAQ, *de novo assembly* acacacacac ("AC" dinukleotida) gacgacgacgac ("GAC" trinukleotida).

Kodominan adalah dua alel di dalam suatu gen yang menghasilkan produk berbeda dengan alel yang satu tidak dipengaruhi oleh alel yang lain.

Kodon (kode genetik) adalah deret nukleotida pada mRNA yang terdiri atas kombinasi tiga nukleotida berurutan yang menyandi suatu asam amino tertentu (kodon triplet). Asam amino yang disandikan contohnya metionin oleh urutan nukleotida ATG (AUG pada RNA). Banyak asam amino yang disandikan oleh lebih dari satu jenis kodon. Kodon berada pada molekul mRNA. Penerjemahan mRNA menjadi protein dilakukan pada ruas penyandi yang diapit oleh kodon awal (AUG) dan kodon akhir (UAA, UAG atau UGA), ruas ini disebut gen. Kodon pada molekul mRNA dapat menyandi asam-asam amino dengan bantuan interpretasi kodon oleh tRNA. Setiap tRNA membawa satu jenis asam amino sesuai dengan tiga urutan nukleotida atau triplet yang disebut dengan antikodon yang berada pada simpul antikodon tRNA. Antikodon mengikatkan diri secara komplementer pada kodon di mRNA, sehingga asam amino yang dibawa oleh tRNA sesuai dengan kodon yang ada pada mRNA. Pesan genetik ditranskripsi kodon demi kodon dengan cara tRNA membawa asam-asam amino sesuai antikodon yang komplementer dengan kodon dan ribosom yang menyambungkan asam-asam amino tersebut menjadi suatu rantai polipeptida. Ribosom menambahkan tiap asam amino yang dibawa oleh tRNA ke ujung rantai polipeptida yang sedang tumbuh.

L

Library adalah material genetik (DNA maupun RNA) yang diberi perlakuan pra-proses dalam proses sekuensing menggunakan NGS.

Linkage adalah keterikatan suatu sifat pada sifat yang menurun akibat letak gen yang saling berdekatan pada kromosom homolog yang sama. *Linkage* disebut juga pautan atau kaitan.

Linux adalah sebuah sistem operasi yang berbasis Unix yang disebarluaskan ke masyarakat secara gratis dan berada di bawah lisensi GNU *General Public License* (GPL), yang berarti bahwa Linux didistribusikan berikut dengan *source code*-nya.

Lokus adalah posisi atau letak kedudukan suatu gen pada suatu kromosom.

M

Materi genetik atau faktor hereditas adalah informasi yang dimiliki oleh setiap sel makhluk hidup yang dapat diwariskan kepada keturunannya. Materi genetik terdiri dari kromosom, gen, DNA, dan RNA.

Mapping adalah proses mensejajarkan hasil *reads* ke transkriptom *reference*.

Marker dominan adalah marka yang hanya mengungkapkan alel dominan tunggal saja tetapi pada lokus yang sama.

Marker ko-dominan adalah salah satu marka yang dapat mengidentifikasi semua alel yang ada pada suatu lokus tertentu.

- Max score* adalah ukuran *statistical significance* dari *alignment* (*Score* tinggi menunjukkan derajat kesamaan yang lebih tinggi ketika membandingkan dua sekuens).
- Metagenomik adalah suatu ilmu yang mempelajari metagenom yaitu seluruh DNA dari suatu ekosistem secara lengkap, misalnya segenggam tanah, sepuluh mililiter air laut, atau isi perut manusia.
- Metabarcoding* adalah *barcode* DNA/RNA (atau eDNA/eRNA) dengan cara yang memungkinkan terjadinya identifikasi simultan dari banyak taksa dalam sampel yang sama.
- Mikrosatelit* atau *Simple Sequence Repeat* (SSR) adalah motif nukleotida berulang yang terdapat pada DNA dengan ukuran 6-15 pasang basa.
- Missing data* adalah data yang tidak tersedia pada *variable* yang diobservasi.
- Mispriming* adalah tingginya persentase penempelan primer di tempat lain yang tidak diinginkan yang akan menyebabkan berkurangnya spesifisitas dari primer tersebut.
- Motif adalah pola sekuens tertentu yang tersebar pada genom atau transkriptom.
- Multiple alignment sequences* adalah proses *alignment* yang melibatkan dua atau lebih sekuens untuk memprediksi kesamaan dan *homologinya*.

N

- Next Generation Sequencing* (NGS) adalah teknologi *sequencing* modern/terbaru yang mampu melakukan pembacaan jutaan sekuens basa fragmen DNA, di antaranya *whole genome sequencing*, *targeted sequencing*, dan *de novo sequencing*. RNA *sequencing* untuk menganalisis ekspresi gen dan transkriptomik serta Metagenomik.
- Nukleotida adalah senyawa organik yang terdiri dari sebuah nukleosida dan sebuah gugus fosfat yang berperan sebagai monomer yang menyusun polimer berupa asam nukleat, yaitu asam deoksiribonukleat dan asam ribonukleat, yang merupakan biomolekul penting yang menyusun makhluk hidup di Bumi.

P

- Percent identity* (identitas persen) adalah angka yang menggambarkan seberapa mirip urutan *query* dengan urutan target (berapa banyak karakter di setiap urutan yang identik). Semakin tinggi persen identitasnya, semakin signifikan kecocokannya.
- Phylogenetic tree* (pohon filogenetik) adalah visualisasi dalam bentuk pohon filogenetik untuk menganalisis kekerabatan organisme.
- Primer DNA adalah sekuens DNA yang komplementer terhadap sekuens yang akan diamplifikasi, terutama dalam reaksi berantai *polymerase/Polymerase Chain Reaction* (PCR).
- Primer *forward* adalah primer maju.
- Primer *reverse* adalah primer mundur yang bekerja berlawanan arah dengan primer *forward*.
- Primer Melting Temperature* (*T_m*) merupakan suhu yang diperlukan oleh separuh primer dupleks agar mengalami disosiasi/lepas ikatan.
- Polimorfik adalah keadaan di mana terdapat variasi alel di dalam suatu populasi.
- Polymerase* adalah jenis enzim yang ditemukan di dalam sel-sel yang terlibat dalam proses polimerisasi polinukleotida, atau penciptaan molekul DNA atau RNA.
- Populasi adalah sekumpulan individu dengan ciri-ciri yang sama (spesies) yang hidup di tempat yang sama dan memiliki kemampuan bereproduksi di antara sesamanya.

Q

Query sequence identity adalah jumlah karakter yang memiliki kesamaan antara dua sekuens.

Query cover adalah angka yang menjelaskan seberapa banyak urutan *query* yang dicakup oleh urutan target. Jika urutan target dalam *database* mencakup seluruh urutan *query*, maka *cover query* adalah 100% (persentase menampilkan bagian dari urutan *query* yang serupa/identik dengan urutan subjek).

R

Read(s) adalah sekuens yang berasal dari fragmen pendek yang terbaca oleh mesin NGS yang belum di-assembly.

Repeats adalah nukleotida yang berulang dalam primer.

Runs adalah primer yang sebaiknya tidak memiliki urutan basa yang diulang secara terus menerus.

RNA (*Ribonucleic Acid*) adalah makromolekul yang berfungsi sebagai penyimpan dan penyalur informasi genetik.

S

Scaffold adalah bagian dari sekuens genom yang tersusun dari *contig-contig* yang telah berhasil disekuens dan di-gap dari daerah yang belum diketahui sekuensnya.

Score adalah angka yang digunakan untuk menilai relevansi biologis dari suatu temuan. Angka yang lebih tinggi sesuai dengan kemiripan yang lebih tinggi.

Self Dimer adalah primer yang menempel dengan primer lainnya yang dalam satu reaksi PCR sehingga jumlah hasil amplifikasi DNA sedikit bahkan tidak ada.

Sekuens DNA adalah sebuah seri huruf-huruf (A, C, G, T) yang mewakili struktur primer dari molekul DNA atau *strand* nyata (informasi genetik yang terdiri dari deretan nukleotida).

Sequence alignment adalah metode penataan dua sekuens (DNA/RNA dan asam amino) atau lebih, untuk menentukan kesamaan dan homologinya atau disejajarkan posisinya untuk melihat jarak kedekatan antar masing-masing sekuens.

Sequence assembly adalah proses komputasi untuk menyatukan fragmen-fragmen sekuens pendek menjadi sekuens yang utuh.

Sekuensing adalah teknik untuk menentukan urutan basa nukleotida dari urutan suatu DNA seperti adenin, timin, guanisin, dan sitosin.

Sekuens consensus adalah sekuens hasil gabungan dari pembacaan primer *forward* dan *reverse*.

Sitosin adalah basa nitrogen pirimidin yang terdapat pada DNA dan RNA. Sitosin berikatan dengan guanin.

Start codon adalah kodon inisiasi berupa AUG pada mRNA yang berfungsi sebagai sinyal “start” pada tahap inisiasi translasi. AUG memiliki antikodon UAC yang membawa asam amino metionin.

Stop codon adalah triplet basa nitrogen yang berupa UAA, UAG, dan UGA. Kodon stop berfungsi sebagai sinyal untuk menghentikan translasi agar tidak lagi mengkode asam amino.

Suhu melting (*T_m*) atau *Primer Melting Temperature* (*T_m*) atau suhu leleh merupakan suhu yang diperlukan oleh primer untuk mengalami disosiasi/lepas ikatan. Suhu leleh primer yang digunakan harus sama untuk memastikan kinerja yang konsisten pada pasangan primer.

Suhu *annealing* (*T_a*) adalah suhu optimal yang diperlukan dalam proses penempelan primer, sehingga diharapkan bisa diperoleh DNA yang digunakan dalam pengujian *in-vitro* berbasis molekuler. *Annealing* merupakan tahapan penempelan primer yang sangat mempengaruhi keberhasilan proses amplifikasi menggunakan PCR.

T

Timin (T) adalah salah satu dari empat nukleobasa nitrogen yang membentuk komponen dasar asam deoksiribonukleat (DNA).

Transkriptom adalah satu set RNA yang diekspresikan oleh individu makhluk hidup atau sekumpulan sel/jaringan tertentu.

Trinity adalah perangkat lunak untuk melakukan *de novo assemble* pada RNA-seq yang mengkonstruksi *contig*, mengelompokkan *contig* yang terbentuk, serta mengekstrak isoform yang terbentuk dari *contig-contig* tersebut.

Transkrip adalah *reads* yang sudah digabungkan menjadi bacaan utas RNA.

U

Unrooted tree adalah menggambarkan keterkaitan ‘*leaf nodes*’ tanpa membuat asumsi tentang leluhur.

Up-regulated adalah ekspresi gen yang lebih tinggi pada sampel dengan perlakuan ketika dibandingkan dengan ekspresi gen pada sampel control atau perlakuan sebaliknya.

Urasil adalah satu dari dua basa N pirimidina yang dijumpai pada RNA dan dapat berikatan dengan adenin melalui dua ikatan hidrogen.

Catatan:

DAFTAR ISI

KATA PENGANTAR.....	iii
SAMBUTAN DIREKTUR SEAMEO BIOTROP	iv
TESTIMONI PEMBACA	v
DAFTAR ISTILAH	vi
DAFTAR ISI.....	1
BAB I ANALISIS MOLEKULER BERBASIS FRAGMEN DNA	3
1.1 Skoring pita hasil elektroforesis menggunakan <i>software</i> PyElph	4
1.2 Klasterisasi menggunakan perangkat lunak NTSYSPC.....	12
1.3 Pengolahan data pita genetik berbasis biner (1 dan 0) dengan menggunakan program <i>software</i> DARwin.	20
1.4 Pengolahan data fragmen DNA menggunakan GenAlEx 6.5	36
1.5 Analisis struktur genetik menggunakan <i>software structure</i> v. 2.3.4	46
BAB II PENGANTAR BIOINFORMATIKA.....	59
BAB III PENGENALAN BIOINFORMATIKA.....	65
BAB IV ANALISIS DATA SEKUENS PENDEK	93
4.1 <i>Editing</i> sekuens menggunakan aplikasi Bioedit.....	95
4.2 Pembuatan <i>alignment</i> multi sekuens, analisis kekerabatan, dan pembuatan pohon filogenetik dari sekuens menggunakan <i>software</i> MEGA11.....	109
BAB V DESAIN PRIMER	123
5.1 Contoh desain primer menggunakan fitur pada NCBI.....	130
5.2 Analisis struktur sekunder menggunakan NetPrimer	133
5.3 Mendesain primer menggunakan primer3plus	136
BAB VI NEXT GENERATION SEQUENCING.....	143
6.1 Pengolahan data RNA-Seq dengan komputer	145
6.2 Pengolahan data RNA-Seq menggunakan <i>Cloud Computing</i> pada MASER (Kinjo <i>et al.</i> 2017)	162
6.3 <i>Cloud Computing</i> pada Platform Galaxy	171
6.4 Teknologi <i>Nanopore Sequencing</i>	175
6.5 <i>Microsatellite finding</i> dan desain primer <i>microsatellite</i>	176
DAFTAR PUSTAKA.....	186

Catatan:

BAB I

ANALISIS MOLEKULER BERBASIS FRAGMEN DNA

Terdapat 4 jenis marka yang digunakan untuk analisis yaitu marka morfologi, kromosom, biokimia dan molekuler. Keempat marka tersebut digunakan untuk saling melengkapi informasi tentang spesies yang diteliti. Marka morfologi adalah marka yang dapat diamati secara langsung seperti pengamatan pohon di lapangan, struktur jaringan batang kayu dan variasi karakter morfologi. Kelebihan marka morfologi adalah kemudahan pengamatannya, sedangkan kelemahannya adalah sangat dipengaruhi oleh tahap perkembangan tanaman dan lingkungan serta sulit dibedakan secara genotipe. Marka kromosom merupakan berupa struktural dan variasi numerik kromosom. Marka biokimia jumlahnya sangat terbatas, memiliki tingkat polimorfisme yang relatif rendah, tergantung umur dan bagian tanaman serta ekspresinya dipengaruhi oleh lingkungan. Marka molekuler mengacu pada polimorfisme pita berupa fragmen DNA atau variasi di tingkat DNA sekuen. Keunggulan utama marka molekuler antara lain:

- a) Tingginya keakuratan.
- b) Tingginya tingkat polimorfismenya.
- c) Tidak dipengaruhi oleh lingkungan yang berpengaruh terhadap ekspresi gen.
- d) Dapat dilakukan pada semua bagian dan tingkat perkembangan tanaman.
- e) Tidak tergantung pada organisme organisme yang diuji.
- f) Mempercepat proses seleksi yang selanjutnya menghemat dan mengefisienkan pengujian di lapangan.
- g) Mendeteksi variasi pada *coding* dan *non-coding* sekuen.
- h) Biasanya diekspresikan dalam kodominan atau jenis dominan kurang disukai.
- i) Bersifat multi-alel.
- j) Lokusnya spesifik.
- k) Berjumlah banyak dan didistribusikan ke seluruh genom.

Teknik biologi molekuler saat ini telah digunakan secara luas, karena dapat memberikan informasi keragaman genetik dari suatu organisme secara cepat dan akurat, sehingga dapat digunakan untuk mendukung hasil pengamatan secara morfologis. Penggunaan marka molekuler seperti RAPD, RFLP, AFLP, SNP, dan SSR serta sekuensing region DNA tertentu (rDNA, region ITS rDNA dan mtDNA) secara langsung dapat menggambarkan keanekaragaman spesies.

Kedepannya, penggunaan teknik biologi molekuler dapat memberikan informasi yang lebih banyak, cepat, dan akurat. Akhir-akhir ini perkembangan Teknik biologi molekuler semakin pesat, seiring dengan semakin banyaknya fenomena kehidupan yang dapat dijelaskan melalui biologi

molekuler. Aplikasi biologi molekuler terdapat di berbagai bidang kehidupan, antara lain: pertanian, kesehatan, kelautan, perternakan, industri dan berbagai bidang lainnya. Analisis biologi molekuler yang lebih lanjut dibutuhkan agar data yang diperoleh di dalam penelitian dapat menjadi bermakna dan memberi informasi yang lebih jelas serta kongkret.

Untuk mempelajari struktur populasi dari suatu kelompok individu termasuk analisis kekerabatan (*similarity*) atau jarak genetik (*genetic distance*) dan penyebaran kelompok tertentu di dalam populasi, perlu dilakukan analisis klaster. Langkah awal dari klasterisasi individu adalah penentuan "parameter" atau dasar pengelompokan yang digunakan, kemudian diikuti oleh analisis statistika dengan menggunakan metode *multivariate* atau *principle component analysis* (PCA). Parameter pengelompokan dapat berupa sifat morfologi, fisiologi, ketahanan terhadap cekaman faktor abiotik atau biotik tertentu dan lain-lain.

Kemajuan di bidang biologi molekuler membuka kesempatan untuk klasterisasi kelompok individu berdasarkan pola pita atau profil DNA pada gel elektroforesis. Profil DNA dapat dihasilkan melalui penggunaan teknologi PCR atau marka molekuler. Berdasarkan profil DNA, individu tertentu dapat dikelompokkan secara objektif dengan membandingkan "laju migrasi pita-pita DNA" dari dua individu yang berbeda pada gel elektroforesis. Berdasarkan skoring pita-pita DNA (*band*) pada gel elektroforesis dapat disusun matriks kekerabatan atau jarak geneti. Selanjutnya dilakukan klasterisasi berdasarkan data matriks yang dihasilkan.

Pada populasi yang relatif kecil dengan menggunakan parameter sederhana, klasterisasi dapat dilakukan secara manual. Akan tetapi pada populasi yang relatif besar atau dengan menggunakan berbagai parameter (beberapa primer PCR atau beberapa metode analisis DNA), maka klasterisasi secara manual akan menjadi sangat kompleks. Untuk itu, klasterisasi dapat dilakukan melalui perangkat lunak/*software* yang tersedia.

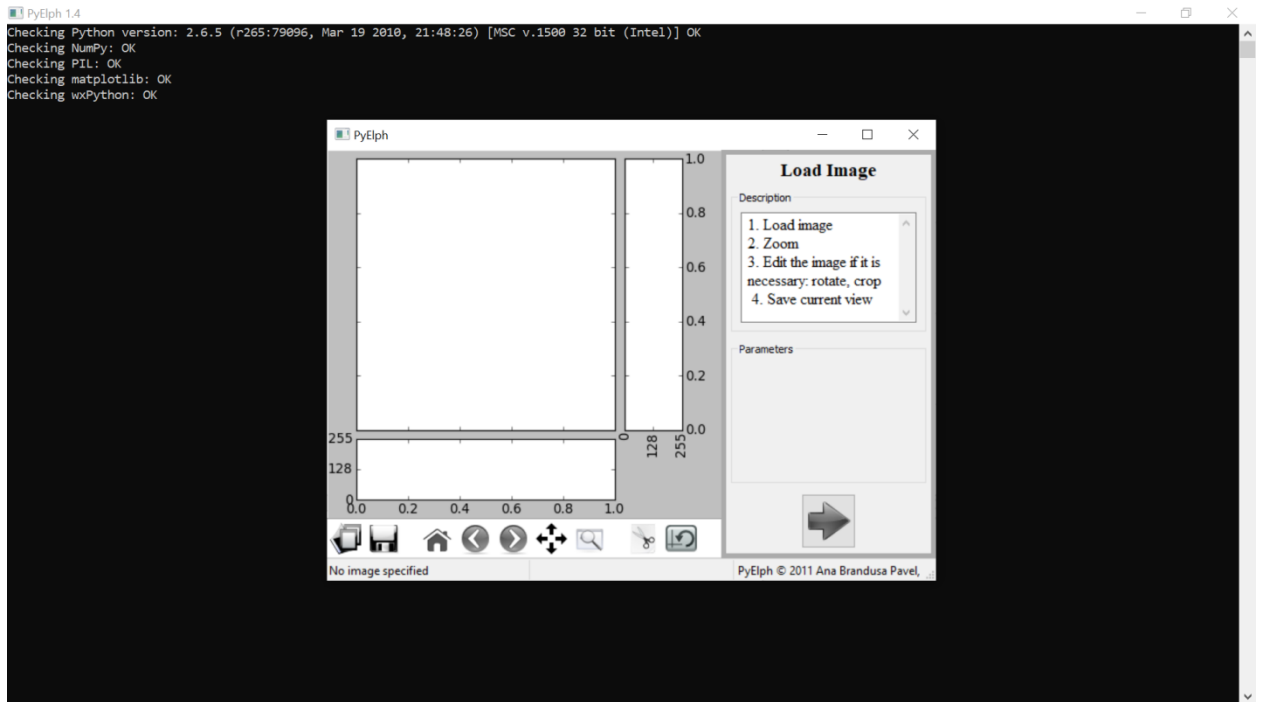
1.1 Skoring pita hasil elektroforesis menggunakan *software* PyElph

Skoring adalah kegiatan untuk menilai keberadaan pita DNA hasil PCR yang divisualisasi dengan elektroforesis pada gel. Pada marker kodominan, proses skoring mungkin tidak akan terlalu sulit karena jumlah pita yang tidak banyak. Contohnya adalah pada *mikrosatelit* maksimal terdapat dua pita yang dihasilkan untuk satu sampel. Sementara itu, marker yang bersifat dominan, seperti AFLP, RAPD, ISSR memiliki jumlah pita yang cukup banyak sehingga terkadang skoring secara manual cukup menyulitkan.

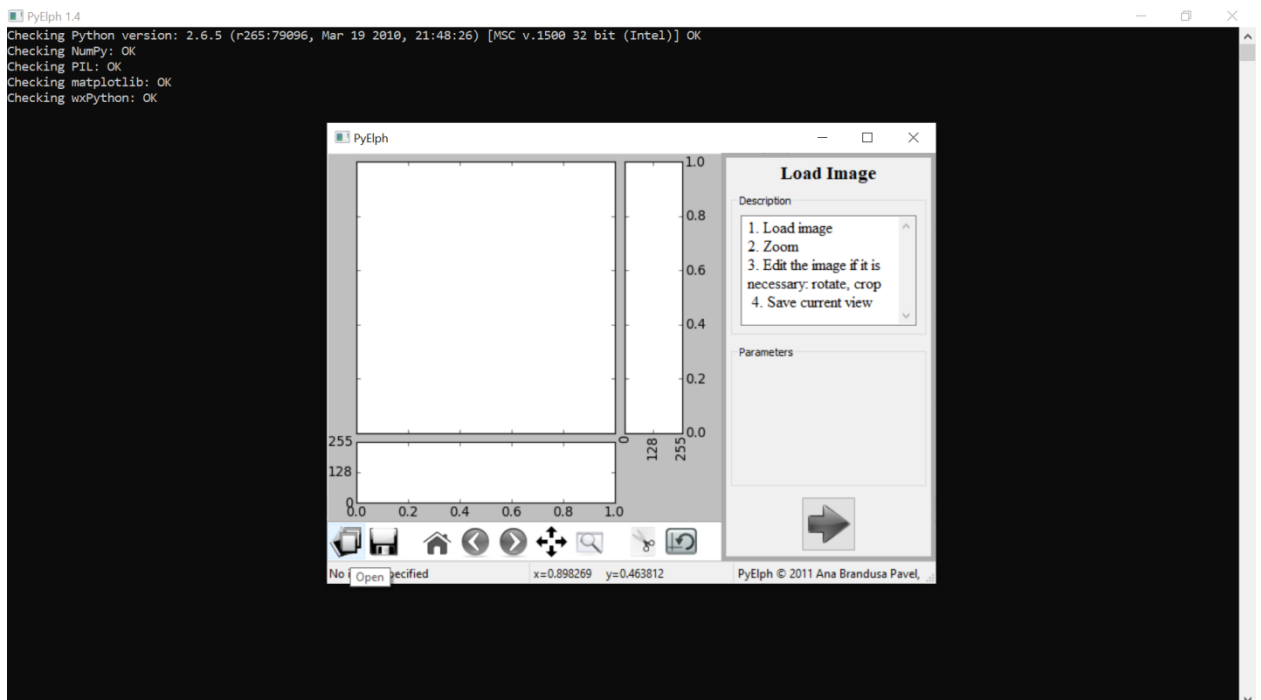
Saat ini, telah tersedia beberapa perangkat lunak yang dapat membantu dalam melakukan skoring pita DNA hasil elektroforesis. Salah satu perangkat lunak yang mudah digunakan untuk skoring marker dominan yaitu PyElph. PyElph (Pavel & Vasile 2012) adalah aplikasi skoring yang

mengeksktrak data dari foto hasil elektroforesis. PyElph berbasis bahasa pemrograman *python* dengan *graphic user interface* yang cukup mudah dipahami. Aplikasi PyElph dapat diunduh di <https://pyelph.software.informer.com/1.4/>. Skoring pita DNA dilakukan mengikuti tahapan berikut:

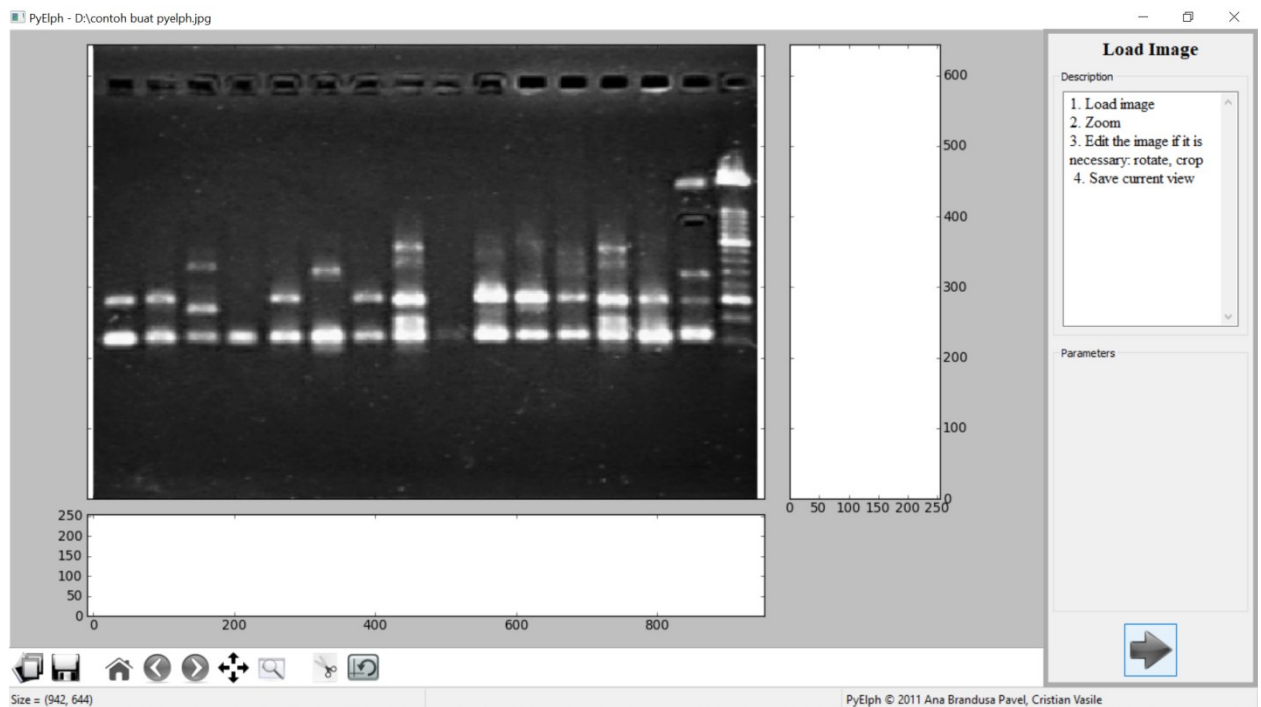
1. Buka aplikasi PyElph



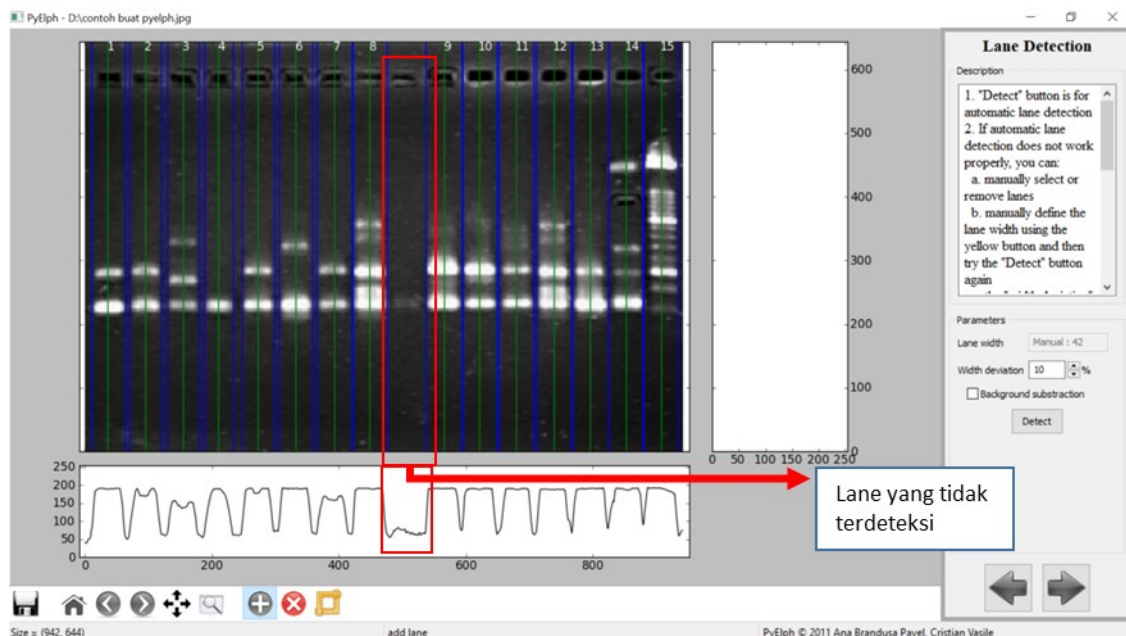
2. Arahkan kursor ke ikon pojok kiri bawah untuk membuka foto hasil elektroforesis.



- Setelah gambar berhasil diunggah, jika tidak ada editing, lanjutkan dengan menekan panah di pojok kanan bawah.

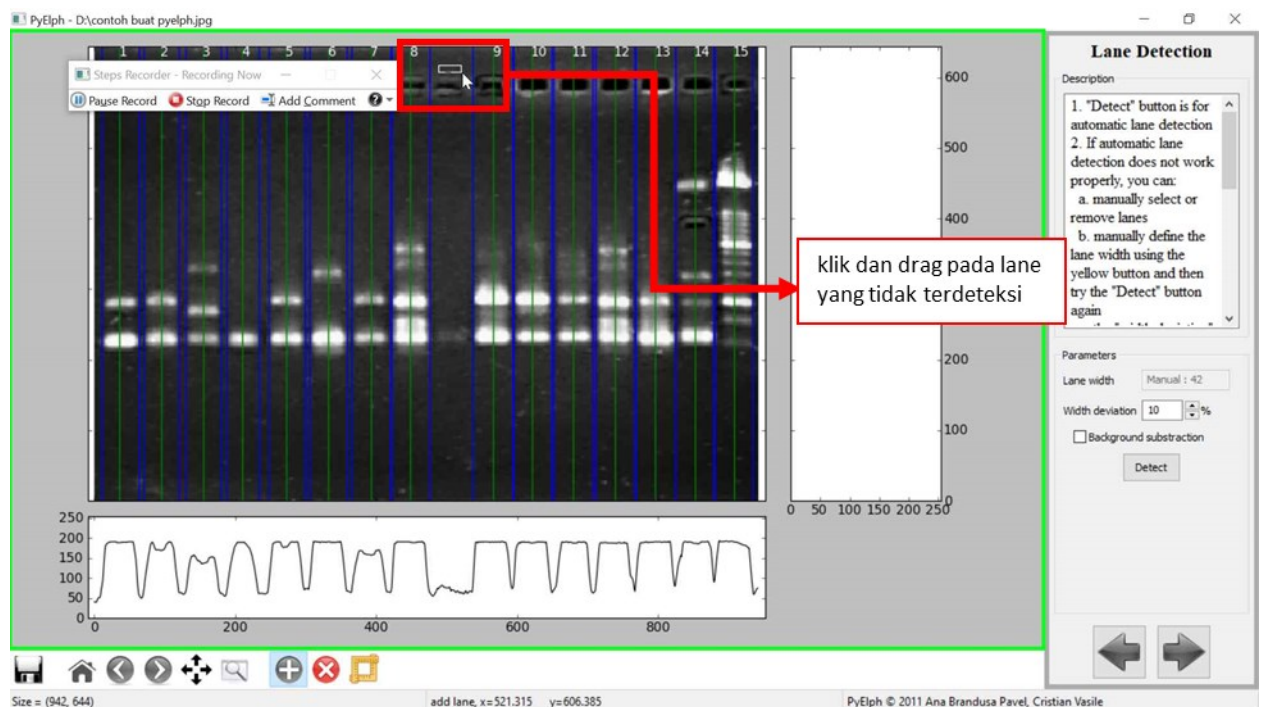
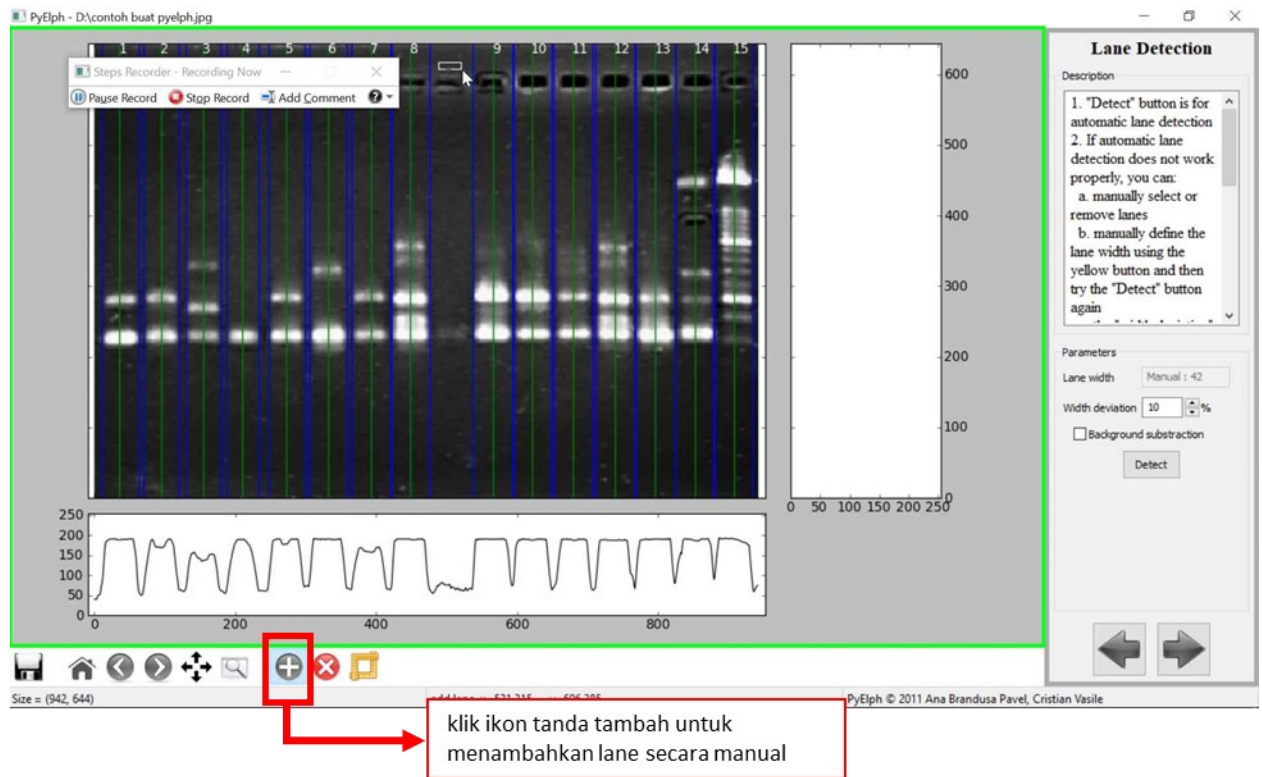


- Lane detection* digunakan untuk mendeteksi lebar pita. *Lane detection* dapat dilakukan secara otomatis dengan tombol *detect*.

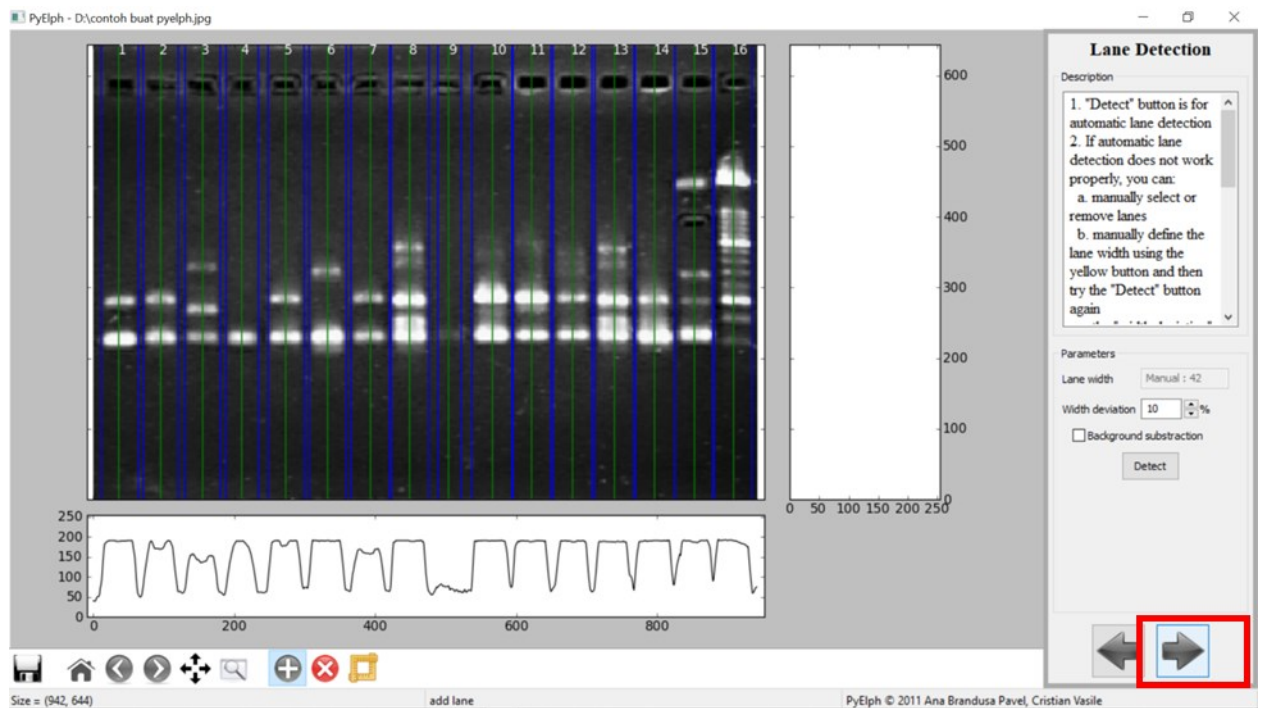


Contoh: setelah *detect* otomatis, ada lane yang tidak terdeteksi sehingga ditambah secara manual.

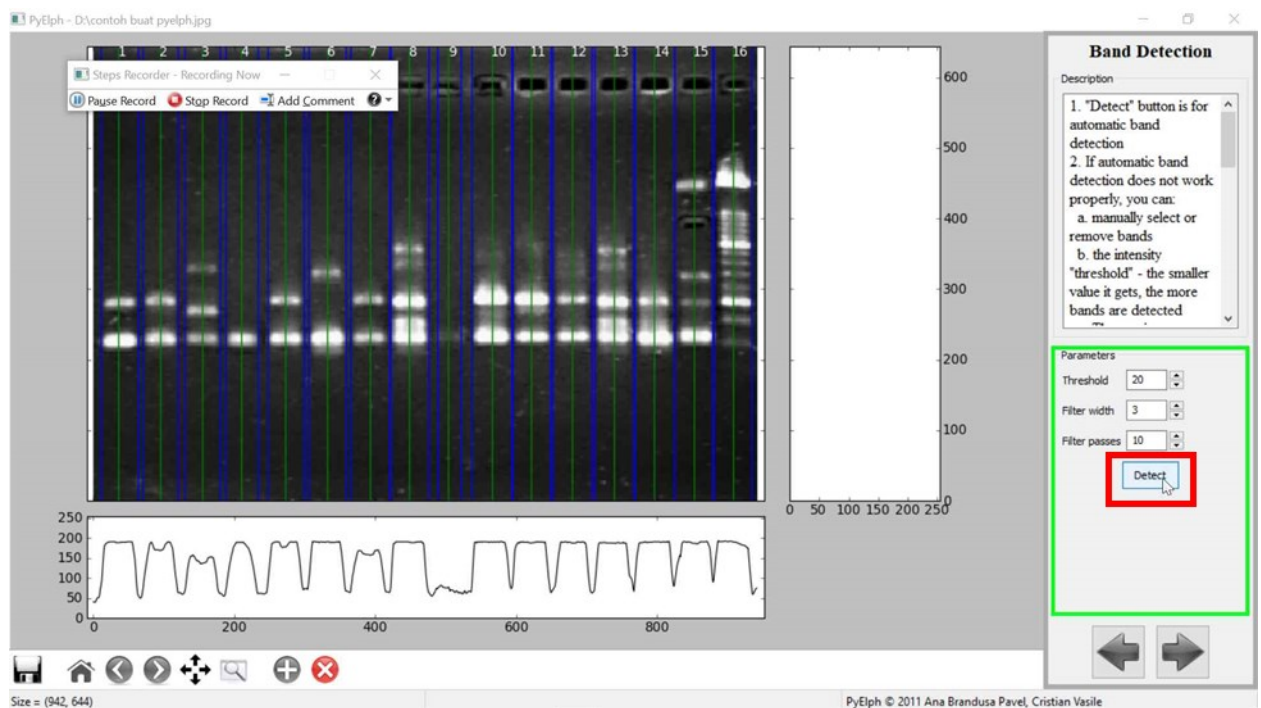
5. *Lane detection* dapat pula dilakukan secara manual dengan mengklik ikon tambah di bawah, kemudian memilih *lane* yang diinginkan dengan klik tanda tambah, lalu geser dan tahan pada jalur sumur yang diinginkan.



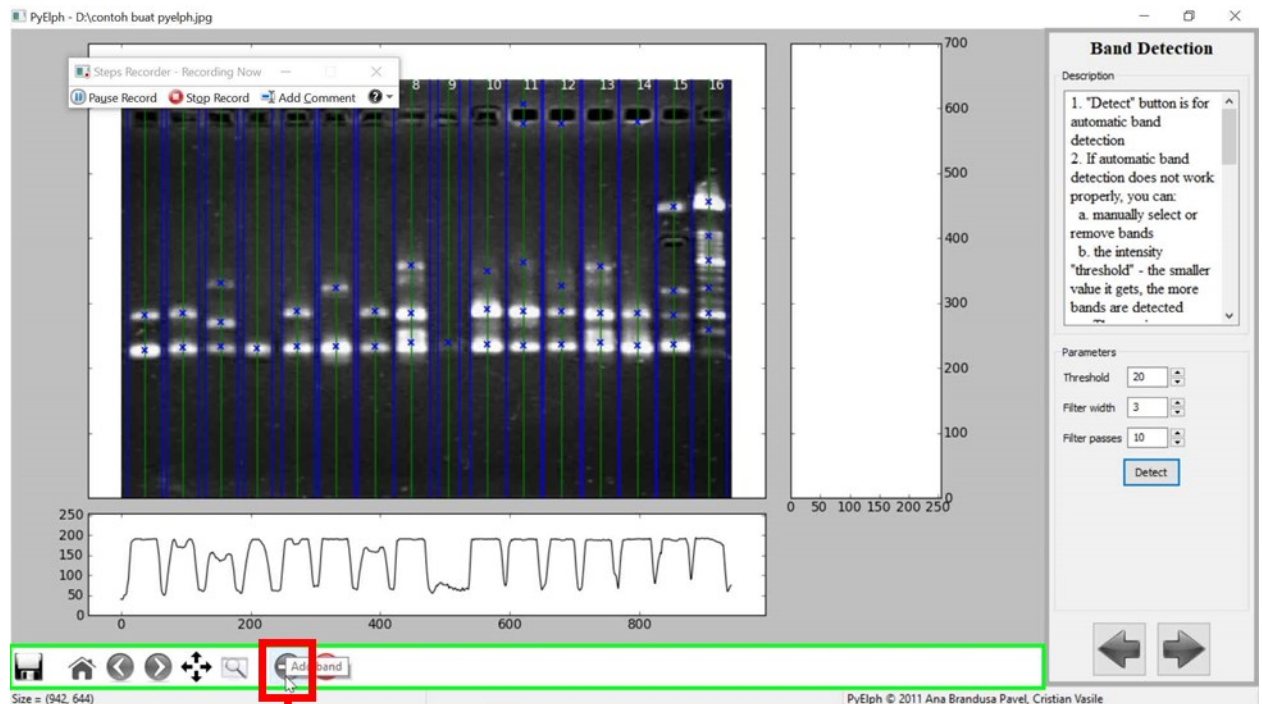
Lanjutkan dengan klik panah ke kanan di pojok kanan bawah.



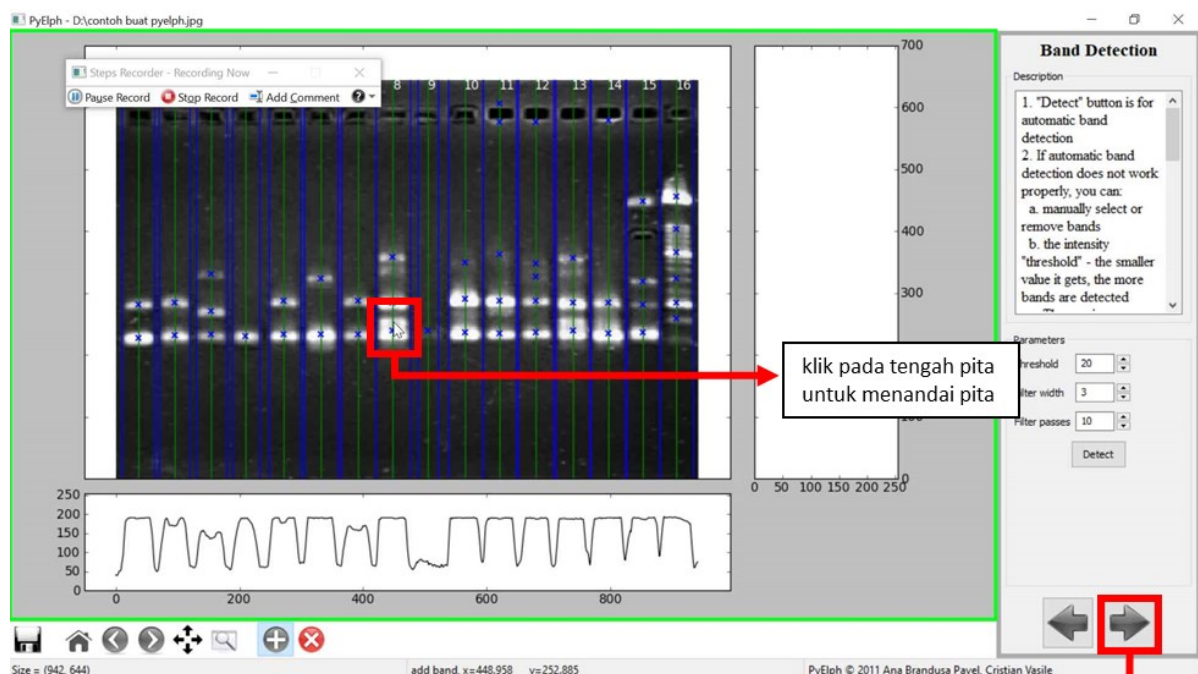
6. Deteksi pita dapat dilakukan secara otomatis dengan klik *detect*.



7. Jika masih ada pita yang tidak terdeteksi dapat ditambahkan dengan mengklik ikon tambah, kemudian klik pada pita yang belum ditambahkan. Jika sudah, lanjutkan dengan mengklik panah ke kanan.



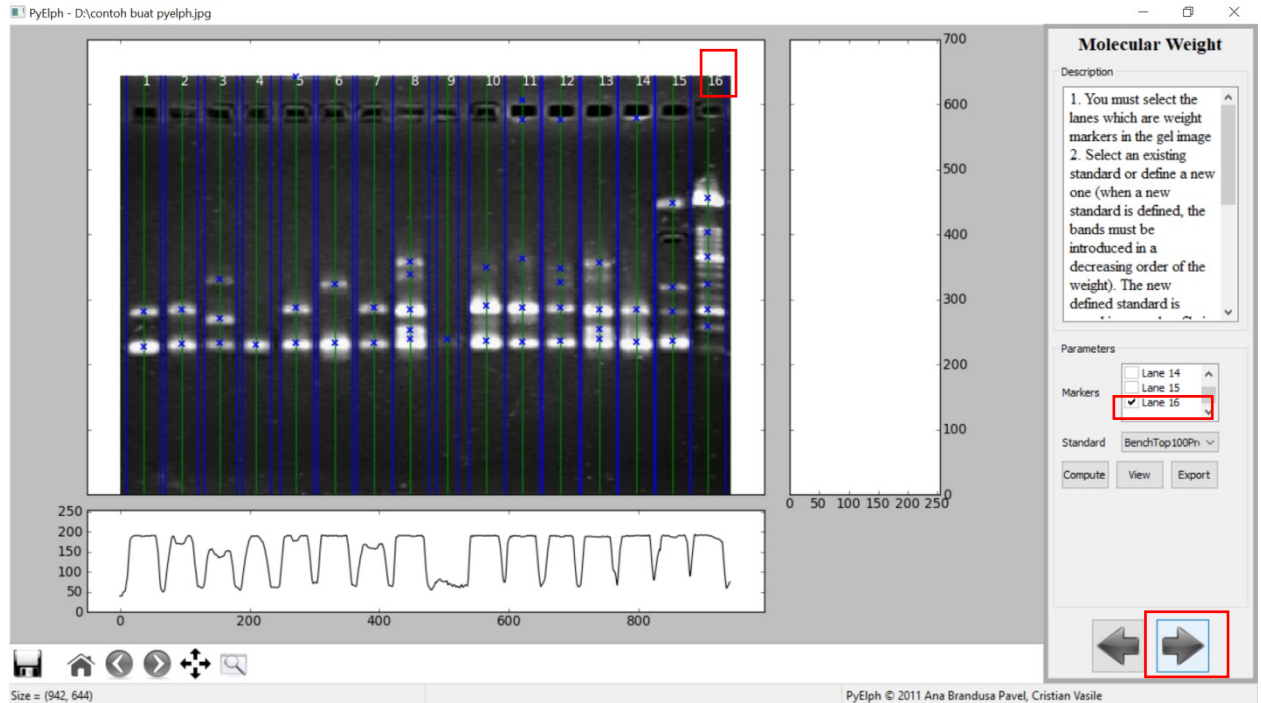
Klik tanda tambah untuk menandai pita secara manual



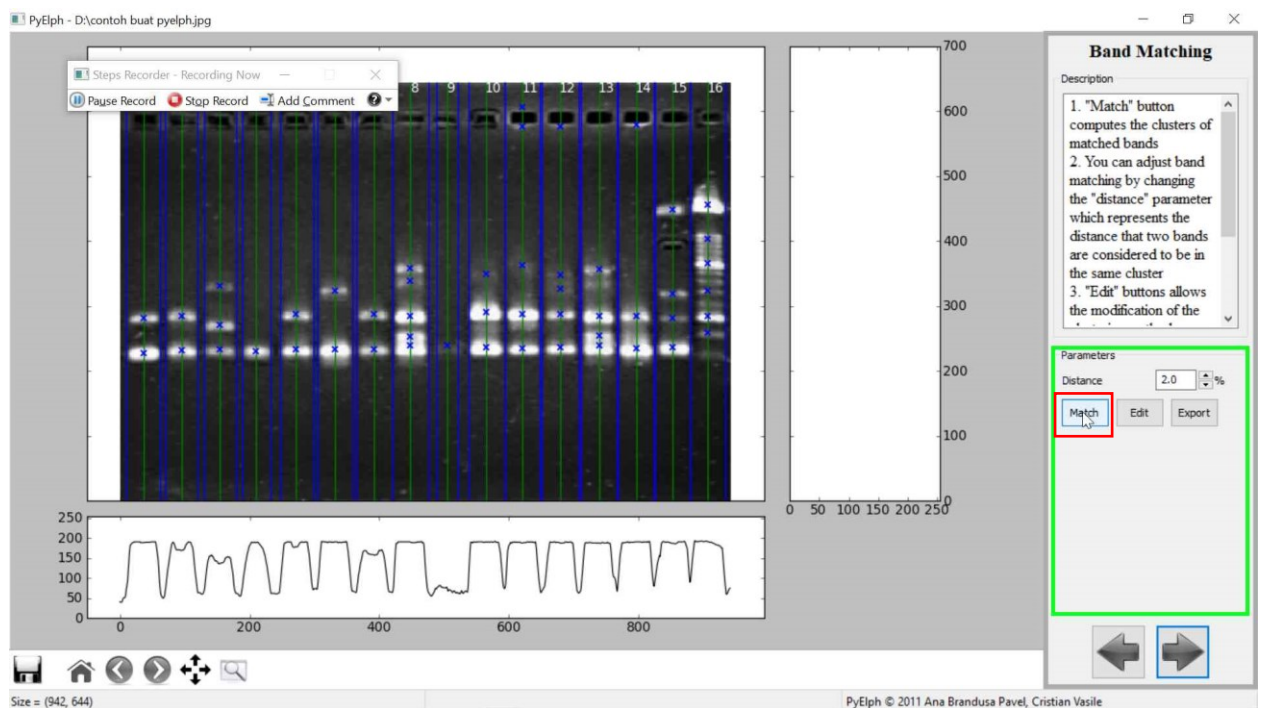
klik pada tengah pita untuk menandai pita

setelah selesai, klik tanda panah ke kanan untuk masuk ke proses selanjutnya

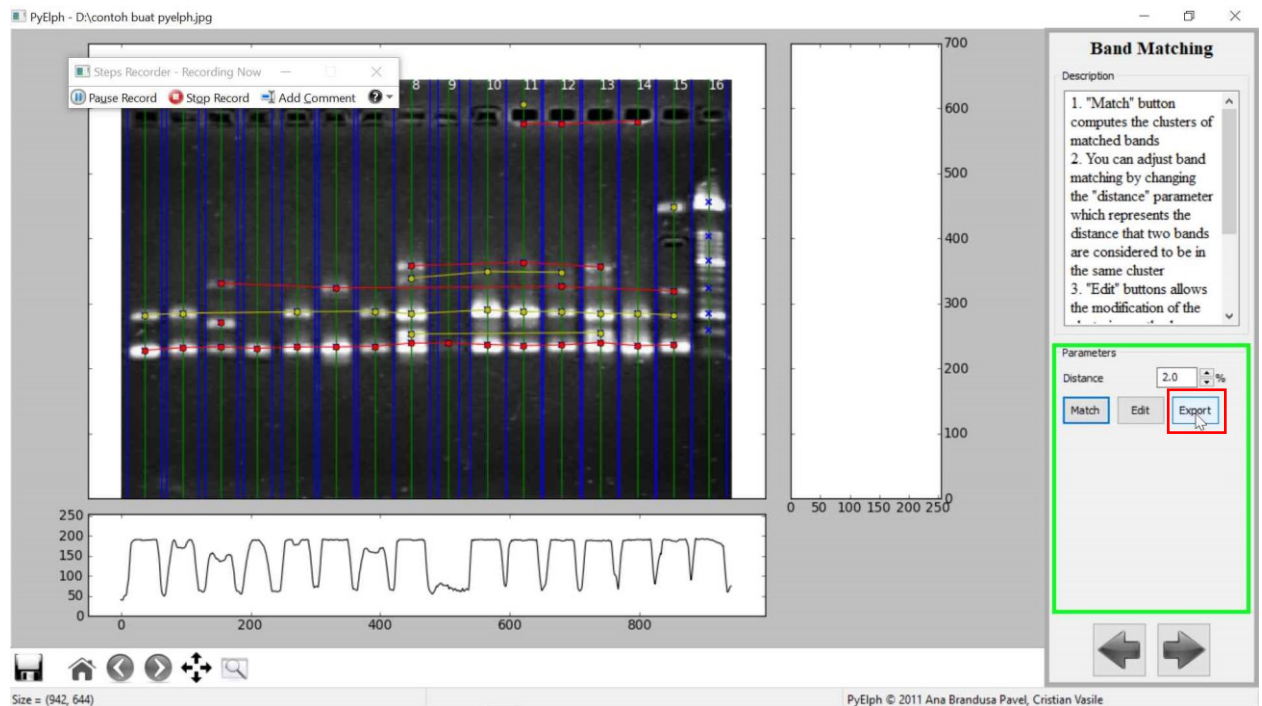
8. Pilih *lane 16* sebagai *lane molecular weight*. Pada foto yang digunakan sebagai contoh, *lane 16* bukan sampel, tetapi *DNA ladder*. Jika sudah lanjutkan dengan klik panah ke kanan di pojok kanan bawah.



9. Pensejajaran pita dan skoring dilakukan dengan mengklik *match* untuk pensejajaran dan skoring secara otomatis.



10. Jika garis pensejajaran sudah sesuai maka hasil skoring dapat dilihat dengan menekan tombol *export*. Jika belum sesuai dapat menekan tombol *Edit* untuk melakukan penyesuaian.



11. Hasil skoring berdasarkan *lane* yang berupa sampel. Contoh untuk skoring sampel pada *lane* 1 yaitu 00000001001, dibaca dari atas ke bawah.

The screenshot shows a Notepad window with the title "skoring - Notepad". The menu bar includes File, Edit, Format, View, and Help. The main text area contains a 16x16 grid of binary digits (0s and 1s). The first column of this grid is enclosed in a red rectangular box.

0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	1	0	0	1	0	1	0	1	0	0
0	0	0	0	0	0	1	0	1	0	1	0	1	0	0	0
0	0	1	0	0	1	0	0	0	0	1	0	0	1	0	0
1	1	0	0	1	0	1	1	0	1	1	1	1	1	1	1
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

The status bar at the bottom indicates the cursor position as "Ln 1, Col 1", the zoom level as "100%", the encoding as "Windows (CP1252)", and the font as "UTF-8".

1.2 Klasterisasi menggunakan perangkat lunak NTSYSPC

NTSYS-PC (*Numerical Taxonomy System for Personal Computer*) merupakan perangkat lunak berbasis statistika yang bertujuan untuk mencari dan menampilkan struktur dari data *multivariate*. Analisis yang dilakukan dengan menggunakan NTSYSPC biasanya adalah analisis kluster (*cluster analysis*). Beberapa metode yang dapat digunakan dalam klasterisasi yaitu (Prokic 2010):

1. UPGMA (*Unweighted Pair Group with Arithmetic Mean*);
2. WPGMA (*Weighted Pair Group with Arithmetic Mean*);
3. UPGMC (*Unweighted Pair Group Method with Centroid*);
4. WPGMC (*Weighted pair Group Method with Centroid*);
5. *Ward's method* atau metode variasi minimum.

Contoh analisis kluster yang disampaikan pada Bab 1 ini menggunakan metode UPGMA dengan koefisien DICE. Koefisien DICE adalah koefisien kemiripan atau similarity yang dikenal juga dengan nama *Sørensen–Dice index*, yang mengkalkulasikan kemiripan antara dua data. Tahapan klasterisasi menggunakan NTSYS-PC adalah sebagai berikut:

1.1 Memasukkan Data

1. Masukkan data biner atau data skoring ke dalam tabel *Microsoft Excel* dengan contoh sebagai berikut:

	A	B	C	D	E	F
1	1	6	5			
2		Sample1	Sample2	Sample3	Sample4	Sample5
3	Band1	1	1	0	1	1
4	Band2	1	0	1	1	1
5	Band3	0	0	1	1	1
6	Band4	1	1	1	0	1
7	Band5	1	1	1	1	1
8	Band6	0	1	1	1	0

Keterangan:

- ◆ Ketik angka 1 di kolom A1 sebagai tanda ada pita (present). Contoh: 1
- ◆ Ketik jumlah pita di kolom B1. Contoh: 6

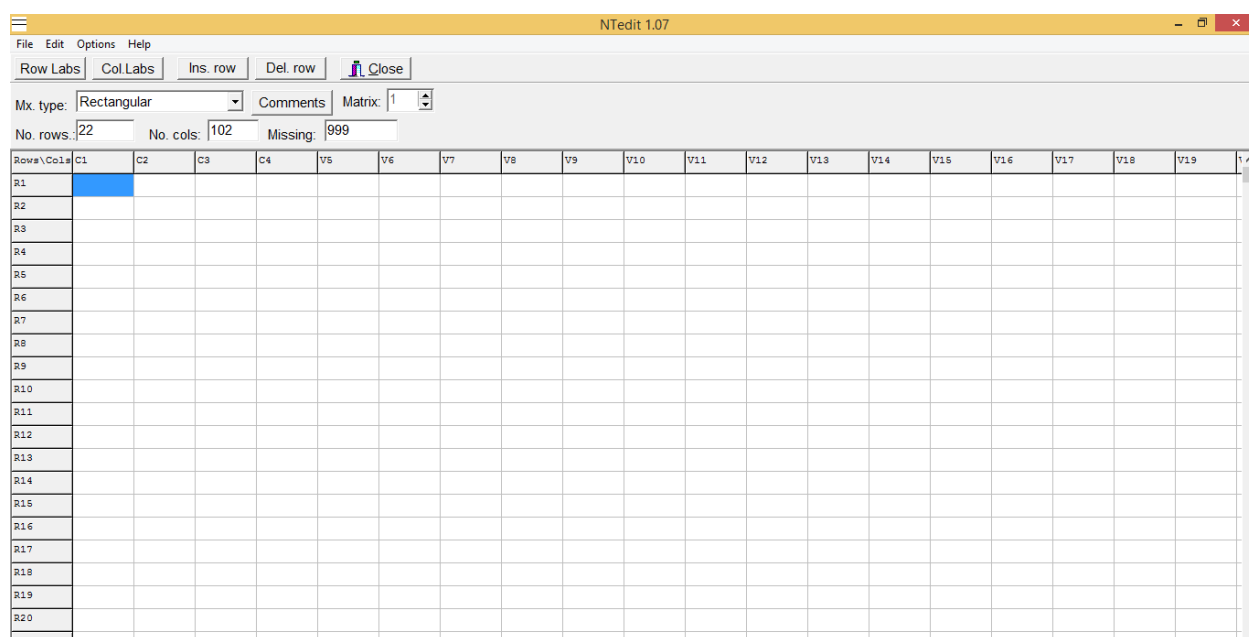
- ♦ Ketik jumlah sample di kolom C1. Contoh: 5
 - ♦ Kode sampel dan band ditulis dengan huruf (bukan angka) dan tidak menggunakan spasi. Bila menggunakan angka, diawali dengan huruf. Contoh: Sample1; Band1
2. Beri nama untuk tabel tersebut dan simpan di dalam folder tersendiri dengan memberi nama file tersebut. Contoh: contoh1.xlsx.

1.2 Langkah-langkah untuk menjalankan Program NTsys

Menjalankan Ntedit

1. Buka program NTsys. Terdapat dua menu yaitu NTsys dan NTedit.

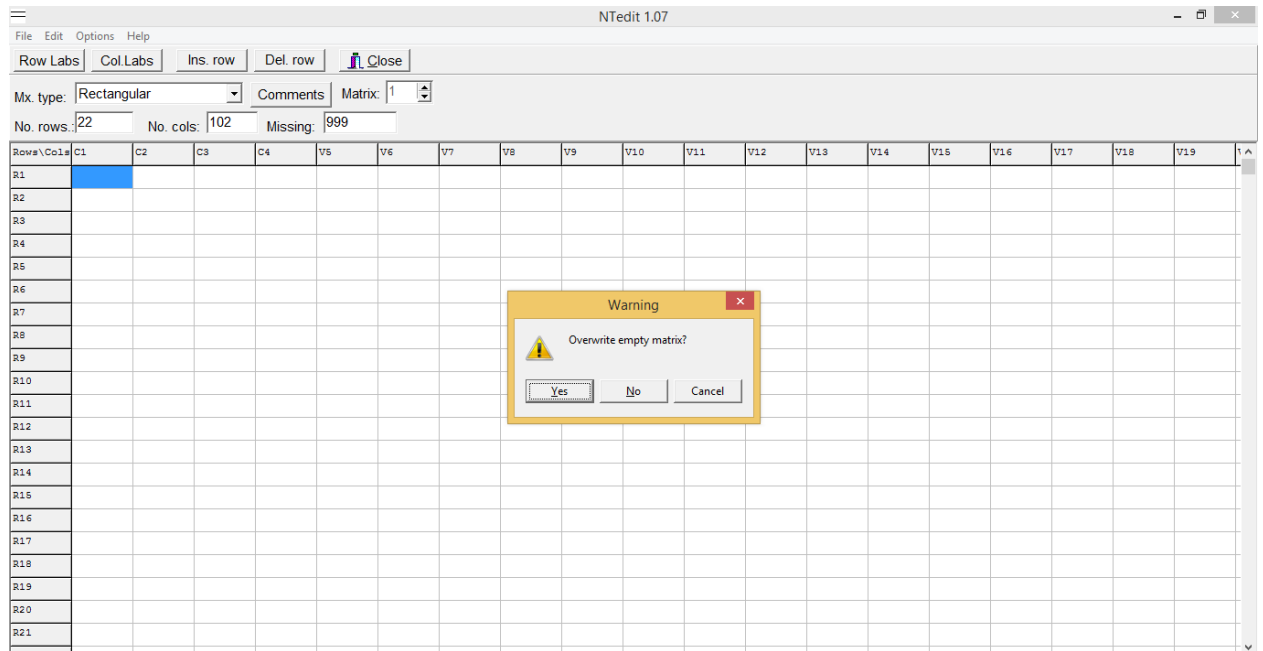
Buka NTedit, isi *row* dan *column* sesuai jumlah data yang kita punya, termasuk *missing data*.



2. Selanjutnya klik *File*, lalu klik *Import Excel* dan klik *using OLE*.



3. Tampilan selanjutnya dan klik *YES*.



4. Tampilan setelah di klik *Yes*.

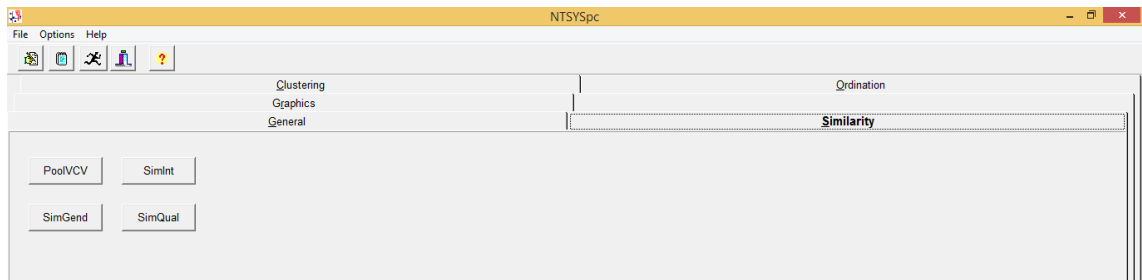
The screenshot shows the NTedit 1.07 interface after clicking 'Yes'. The main grid now displays a matrix of 0s and 1s. The rows are labeled with sample IDs: SCa12, SCa6, GC7, ICS60, DR2, DR38, DRc16, ICS13, IOCRI3, IOCRI4, RCC70, RCC71, RCC72, IOCRI1, IOCRI2, PA300, PA303, TSH858, DR1, WIC4, and UF667. The columns are labeled C1 through C102. The matrix data is as follows:

Rows/Col	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22
SCa12	1	0	1	0	1	0	1	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0
SCa6	1	0	1	0	1	0	1	0	0	0	1	0	0	0	1	0	1	0	1	0	1	0
GC7	0	1	1	0	1	0	1	0	0	0	1	0	1	0	0	0	1	0	1	0	1	0
ICS60	1	0	1	0	1	0	0	1	0	0	0	1	0	1	0	0	1	0	0	0	0	0
DR2	1	0	1	0	1	0	0	1	0	0	0	1	0	0	1	0	1	0	0	0	0	0
DR38	1	0	1	0	1	0	0	1	0	0	0	1	1	0	0	0	1	0	0	0	0	0
DRc16	1	0	1	0	1	0	1	0	0	0	1	0	0	0	1	0	1	0	0	0	0	0
ICS13	1	0	1	0	0	1	0	1	0	0	0	1	0	1	0	0	0	1	0	0	0	0
IOCRI3	1	0	1	1	0	0	1	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0
IOCRI4	1	0	1	1	0	0	1	0	0	0	1	0	0	0	1	0	1	0	1	0	1	0
RCC70	1	0	1	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	0	0	0
RCC71	1	0	1	0	1	0	0	1	0	0	1	0	0	0	1	0	1	0	1	0	1	0
RCC72	1	0	1	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	0	0	0
IOCRI1	1	0	1	0	1	0	1	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0
IOCRI2	1	0	1	0	0	1	1	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0
PA300	0	1	1	0	1	0	0	1	0	1	0	0	1	0	0	0	1	0	1	0	1	0
PA303	0	1	1	0	1	0	0	1	0	1	0	0	0	0	1	0	1	0	1	0	1	0
TSH858	1	0	1	0	1	0	0	1	0	1	0	0	1	0	0	0	1	0	0	0	0	0
DR1	1	0	1	0	1	0	1	0	0	0	1	0	0	1	0	0	1	0	1	1	0	0
WIC4	1	0	1	0	0	1	1	0	0	0	0	1	0	0	1	0	0	1	1	1	0	0
UF667	1	0	1	1	0	1	1	0	0	0	1	0	0	0	1	0	0	1	1	1	0	0

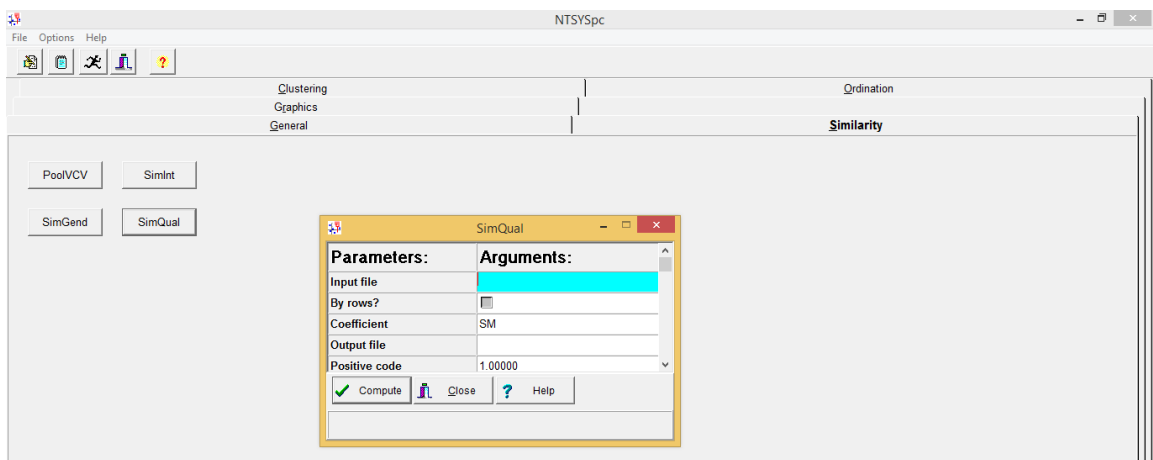
5. Setelah data terinput, *SAVE As* dalam bentuk .NTS (misalnya data1.NTS). Tutup file.

- Menjalankan NTsys

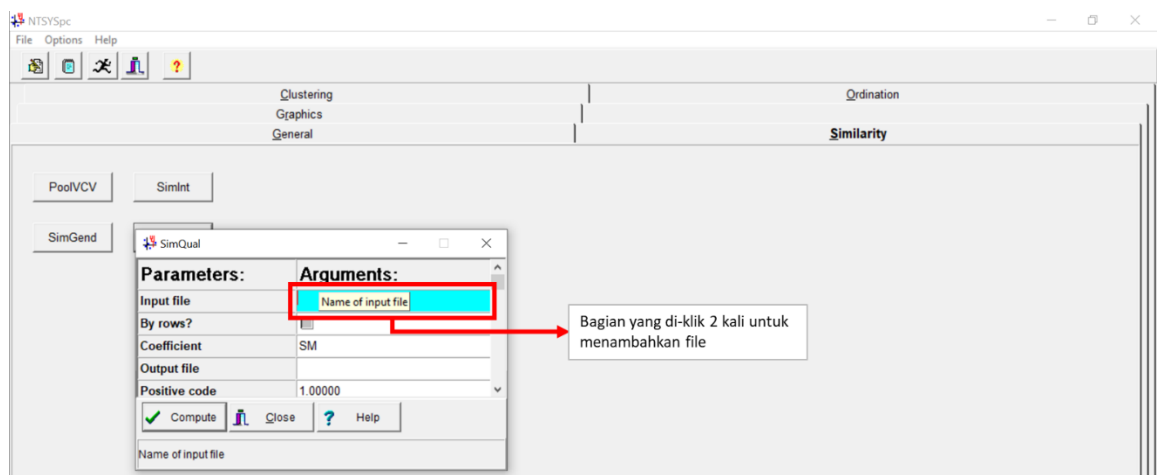
1. Buka file NTsys, tampilannya sebagai berikut:



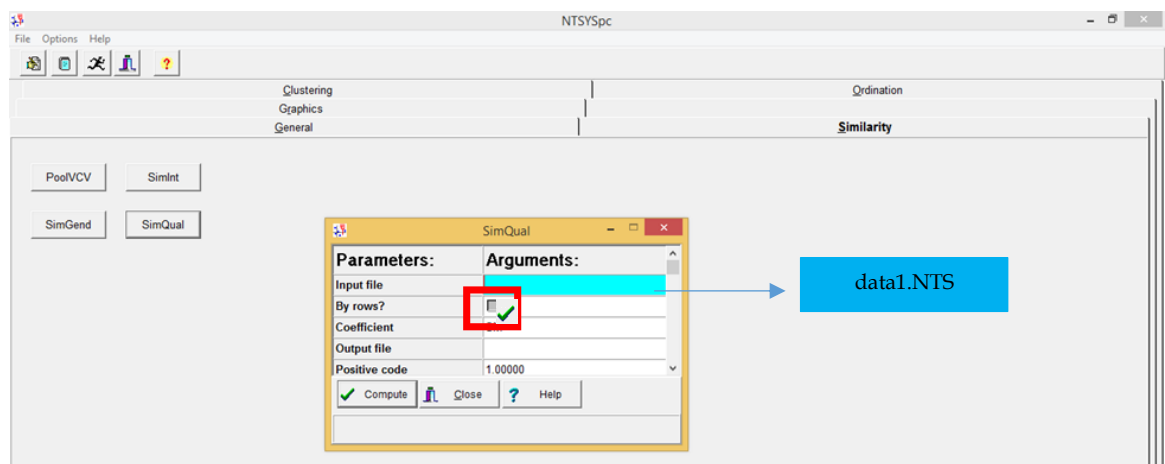
2. Pilih *similarity* dan klik *SimQual*, tampilannya sebagai berikut:



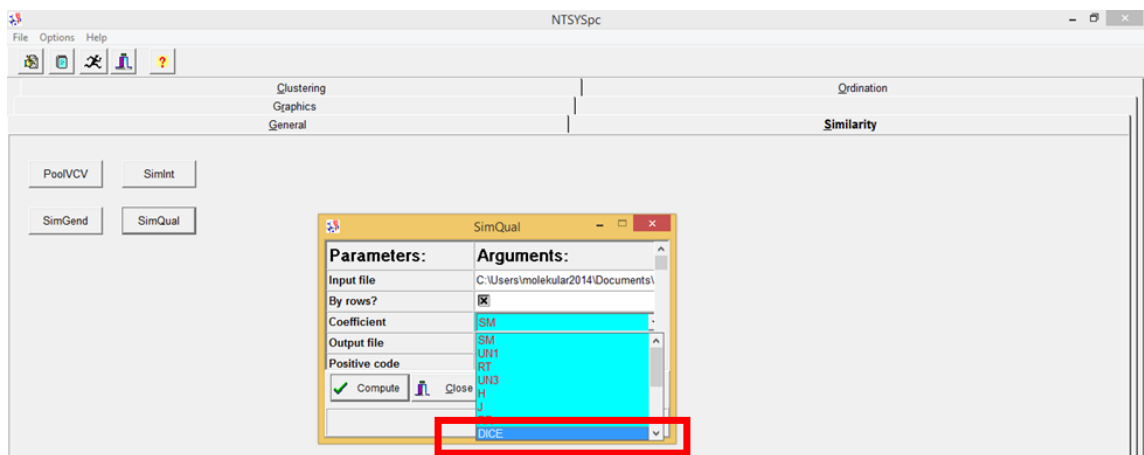
3. Pada baris *Input file* pilih file data1.NTS, yaitu berdasarkan data file yang sudah kita *save* dan berisi data yang kita buat ketika menjalankan program NTedit. Cara memilih file pada baris *Input file* adalah dengan mengklik 2 kali di baris *input file* dalam bentuk <nama file>.NTS



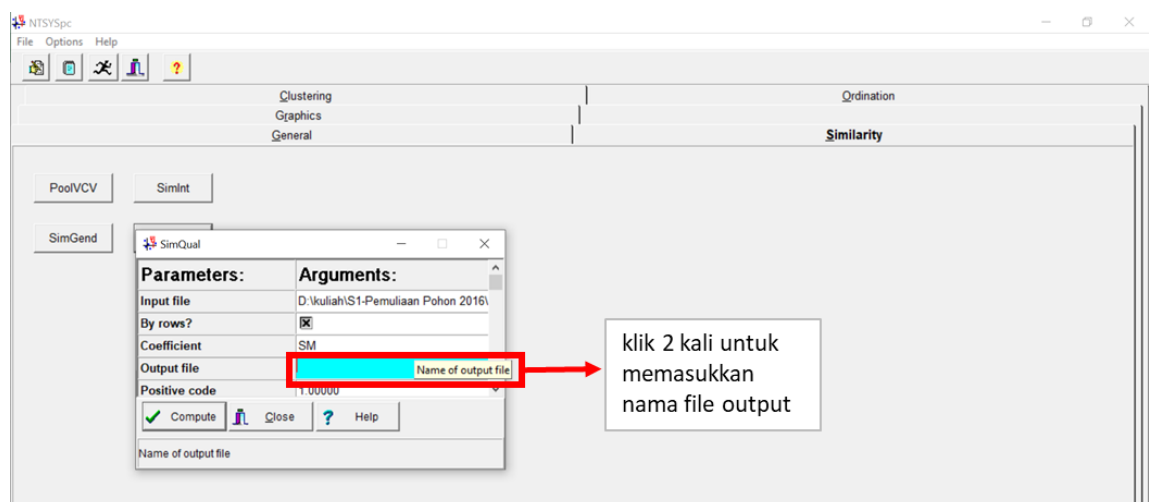
4. Letakkan tanda *checkbox* di baris *By row* jika data yang digunakan tersusun dalam baris.



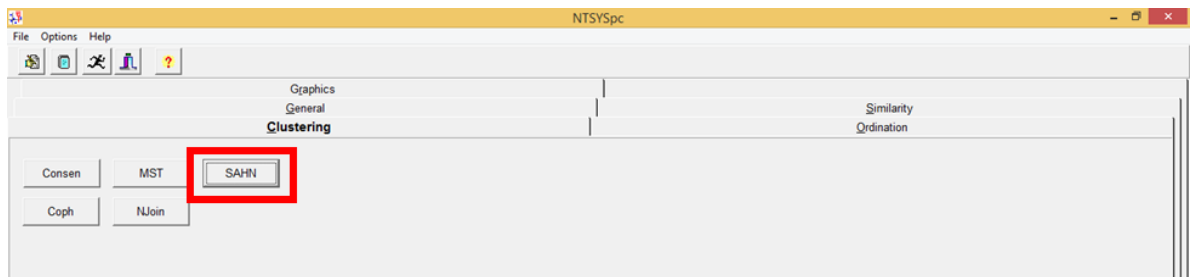
5. Pada baris *Coefficient* pilih DICE.



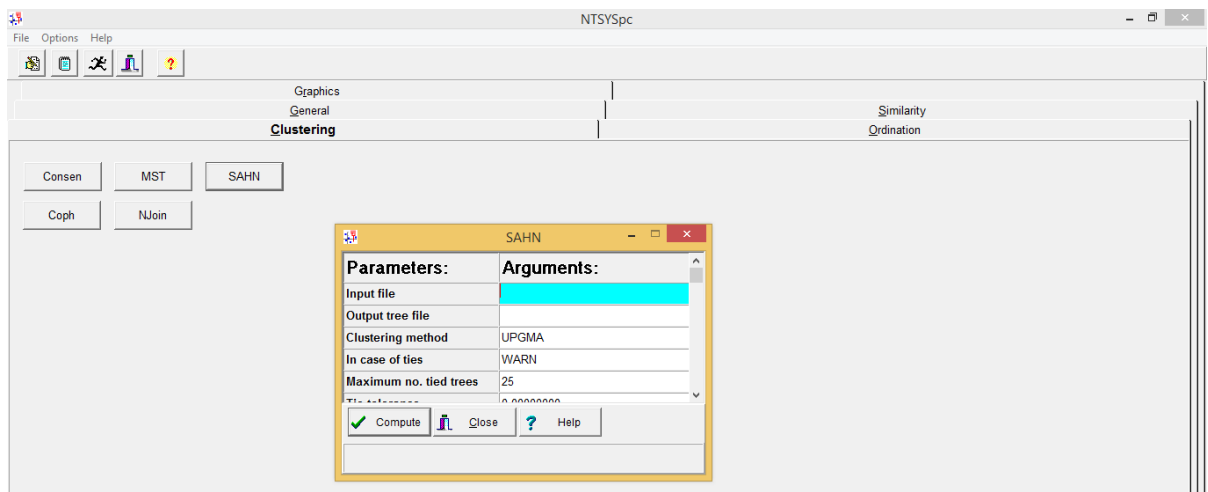
6. Pada baris *Output file* isi simqualdata1, kemudian klik *save* di dalam folder yang sudah kita buat dan klik *Compute*. Tutup.



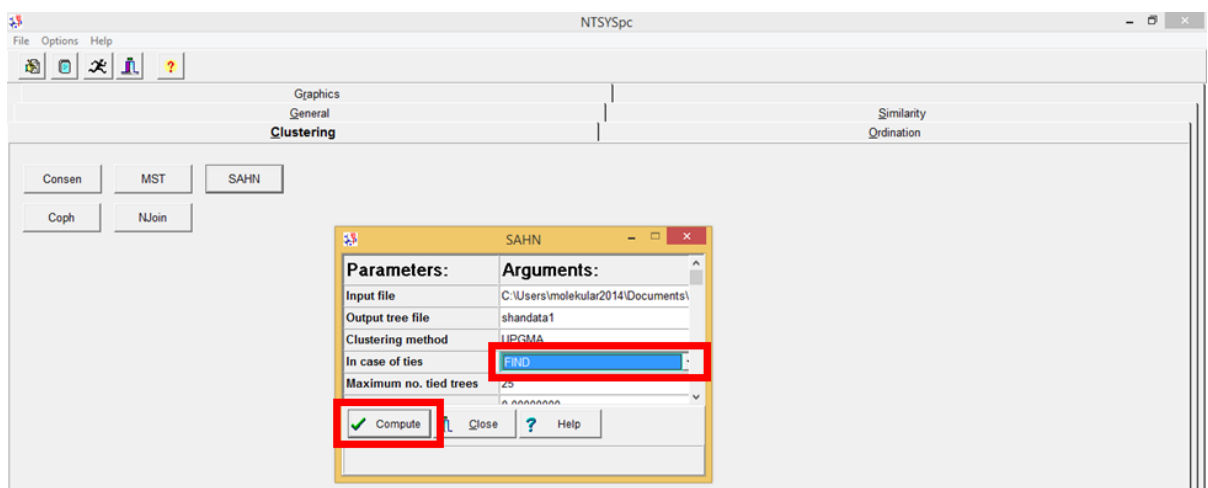
Kembali ke menu awal NTSys dan klik *Clustering* dan pilih *SAHN*



7. Akan muncul tampilan sebagai berikut:

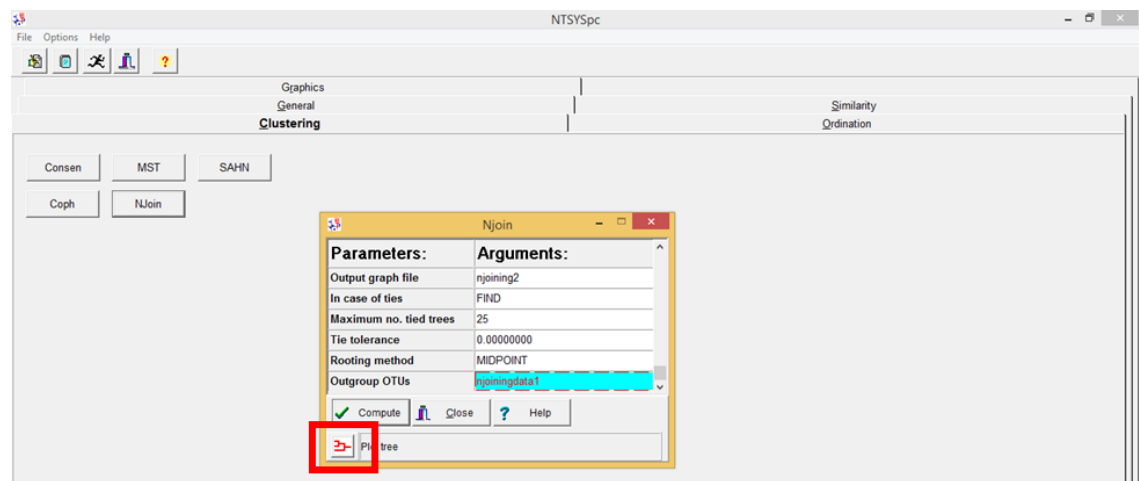


8. Pada baris *Input file* pilih Simqual data1 yang sebelumnya diisikan sebagai *Output file* di point 7 dengan cara mengklik 2 kali folder yang sudah kita buat di dalam poin 6. Selanjutnya pada baris *output tree file* isi Shandata1 kemudian *save*. Pada baris *In case of ties* pilih FIND.

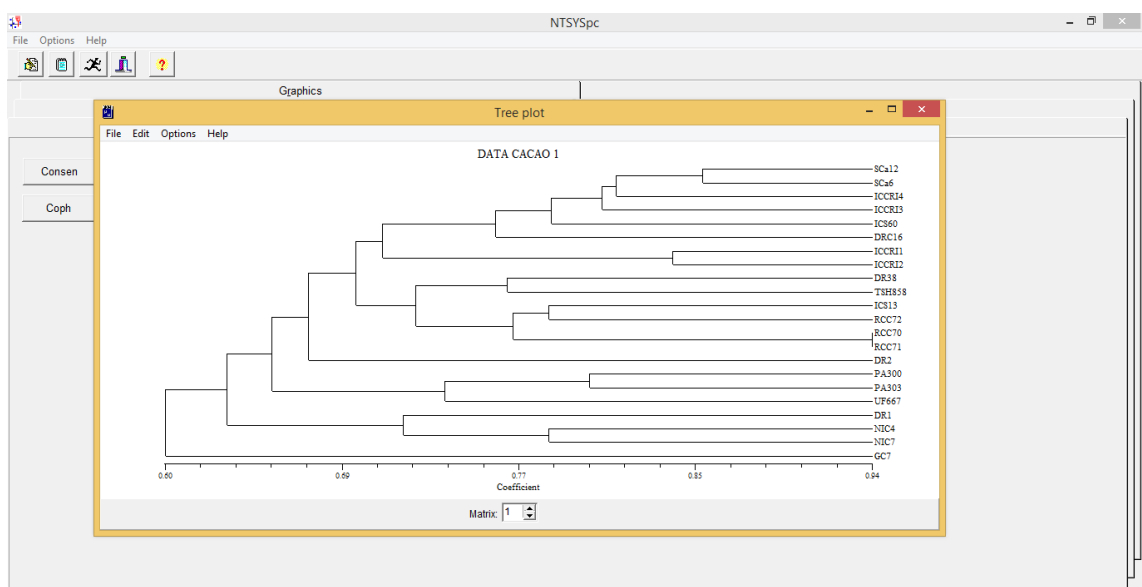


9. Klik *Compute*.

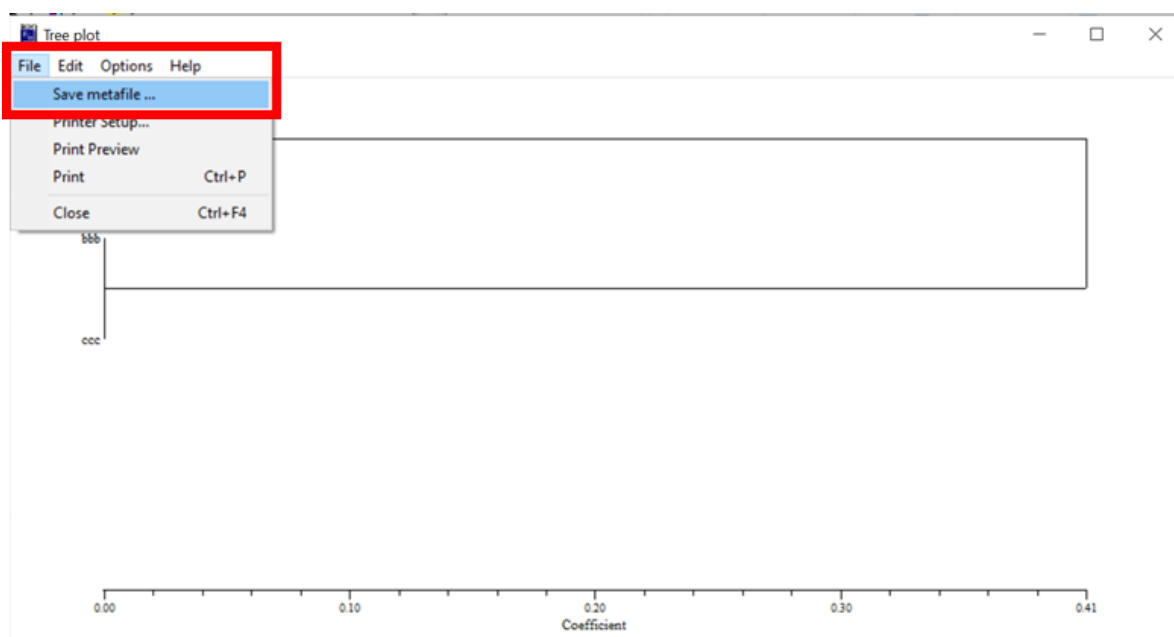
Tampilan hasil *compute* bisa dilihat langsung pada *tree* filogenetiknya (*Plot tree*) dengan mengklik warna merah yang ada di pojok bawah.



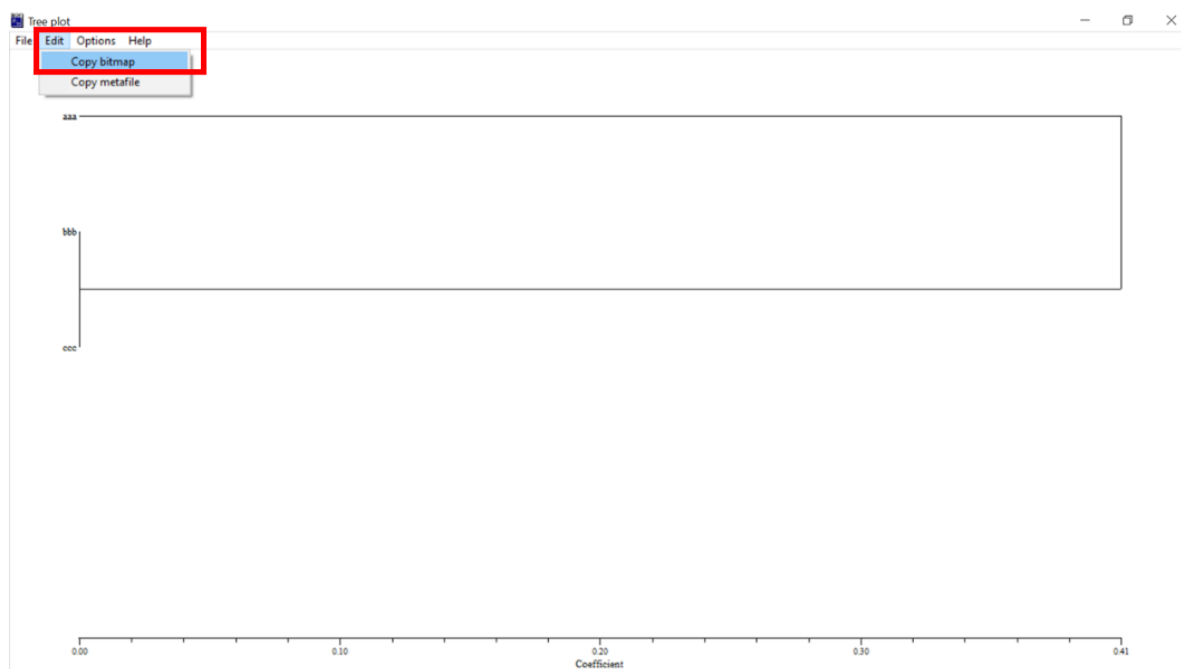
10. Tampilan yang akan muncul adalah sebagai berikut:



11. *Save metafile* gambar pohon filogenetik yang sudah diperoleh.



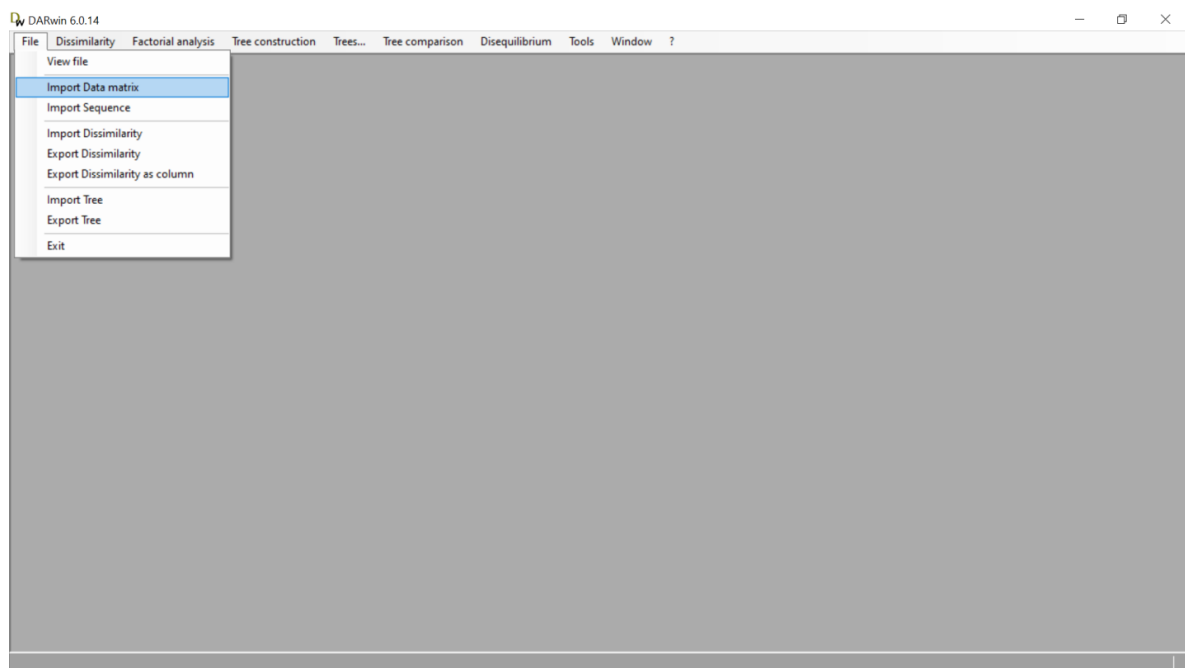
12. Jika ingin menyimpan gambar ke dalam format gambar (JPG, PNG) bisa dilakukan dengan menyalin gambar sebagai bitmap kemudian di-paste ke aplikasi editor gambar, kemudian disimpan ke dalam format gambar



1.3 Pengolahan data pita genetik berbasis biner (1 dan 0) dengan menggunakan program *software* DARwin.

DARwin (*Dissimilarity Analysis and Representation on Windows*) adalah aplikasi atau perangkat lunak yang dikembangkan untuk analisis keragaman dan analisis filogenetik berdasarkan perbedaan evolusioner. DARwin menggunakan metode standar untuk representasi pohon dan faktorial yang ditingkatkan dengan pendekatan awal dan spesifik, terutama menjawab pertanyaan sensitivitas terhadap akurasi data. Data jarak genetik disajikan di dalam *file extension* yang bernama *dissimilarity* (.dis) (Perrier & Jacquemoud-Collet 2006). DARwin dapat digunakan untuk mengolah data dari marker dominan seperti AFLP dan RAPD, kodominan seperti *mikrosatelit* dan sekuens. DARWin dapat diakses secara gratis untuk pendidikan pada laman <https://darwin.cirad.fr/product.php>. Langkah-langkah untuk menjalankan program DARwin adalah sebagai berikut:

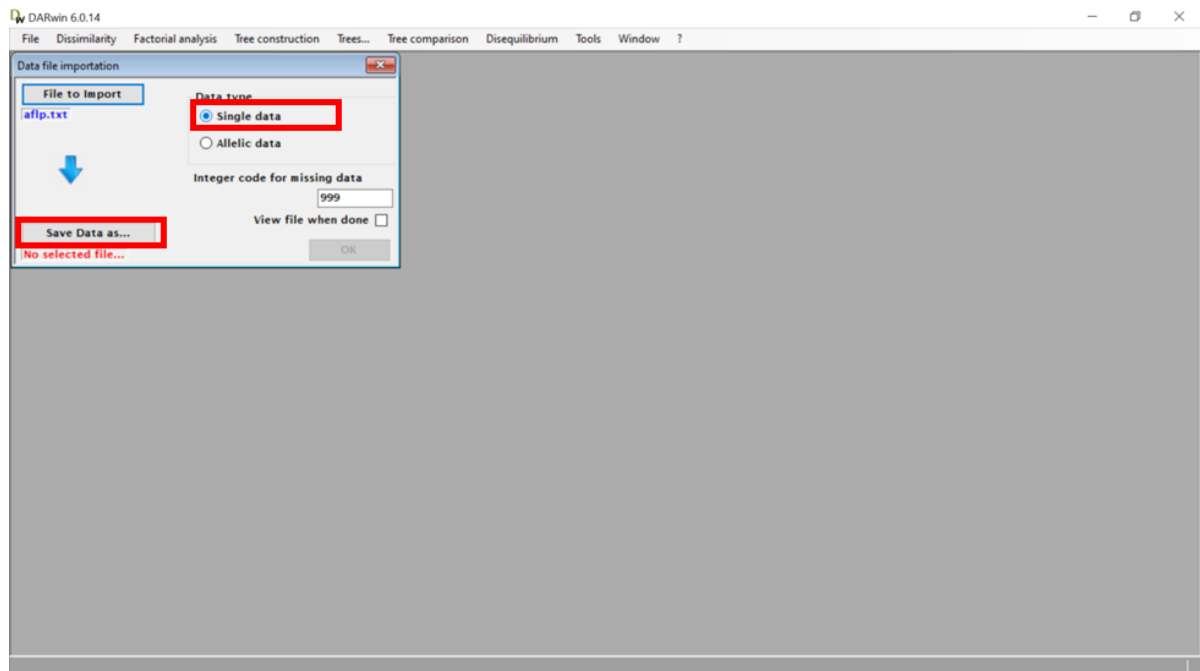
1. Memasukkan matriks data dengan cara mengklik *File* dan memilih *Import Data Matrix*.



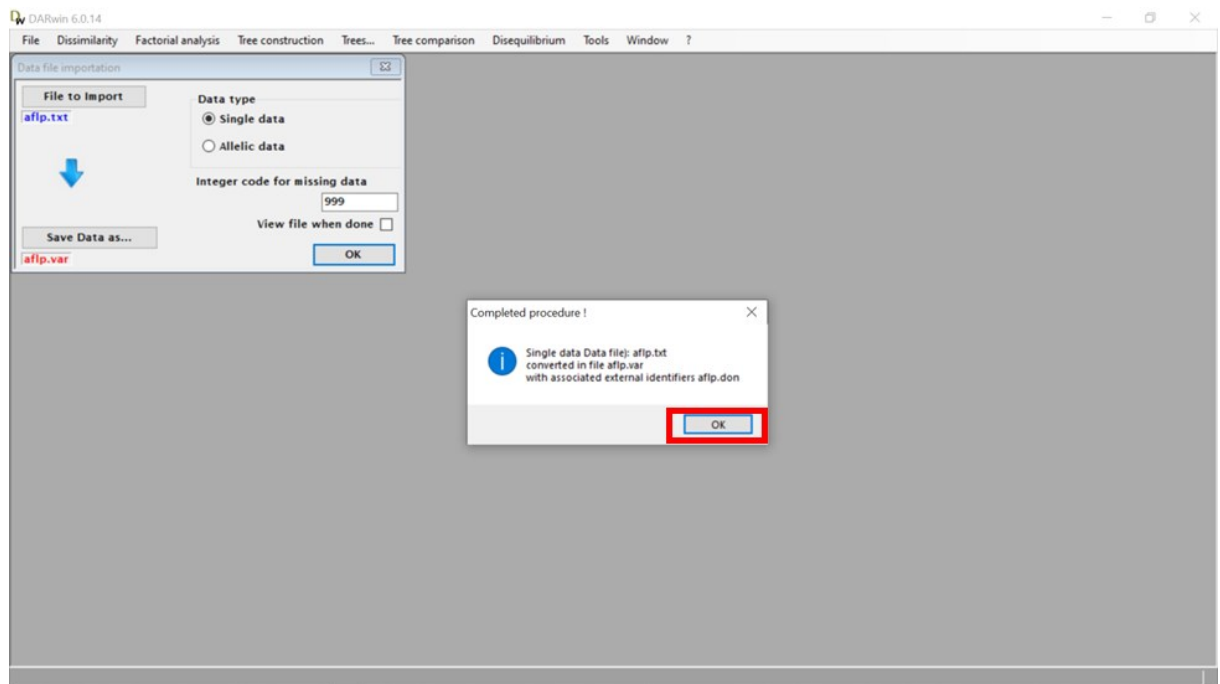
2. Matriks data yang digunakan untuk DARwin berupa file .txt dengan format seperti gambar di bawah. Pada baris paling atas, masukkan kode individu, bisa berupa sampel, aksesori dan sebagainya. Kode individu tersebut diikuti dengan nama marker yang digunakan. Tampilan data yang sudah dimasukkan adalah sebagai berikut:

Individual	Marker 1	Marker 2	Marker 3	Marker 4	Marker 5	Marker 6	Marker 7	Marker 8	Marker 9	Marker 10	Marker 11	Marker 12	Marker 13	Marker 14	Marker 15	Marker 16	Marker 17	Marker 18	Marker 19	Marker 20
B01_R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B02_R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B03_R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B04_R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B05_R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B06_R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B07_R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B08_R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B09_R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B10_R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B11_R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B12_R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B13_R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B14_R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B15_R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B16_R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B01_S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B02_S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B03_S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B04_S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B05_S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B06_S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B07_S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B08_S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B09_S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B10_S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B11_S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B12_S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B13_S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B14_S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B15_S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B16_S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C01_R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C02_R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C03_R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C04_R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

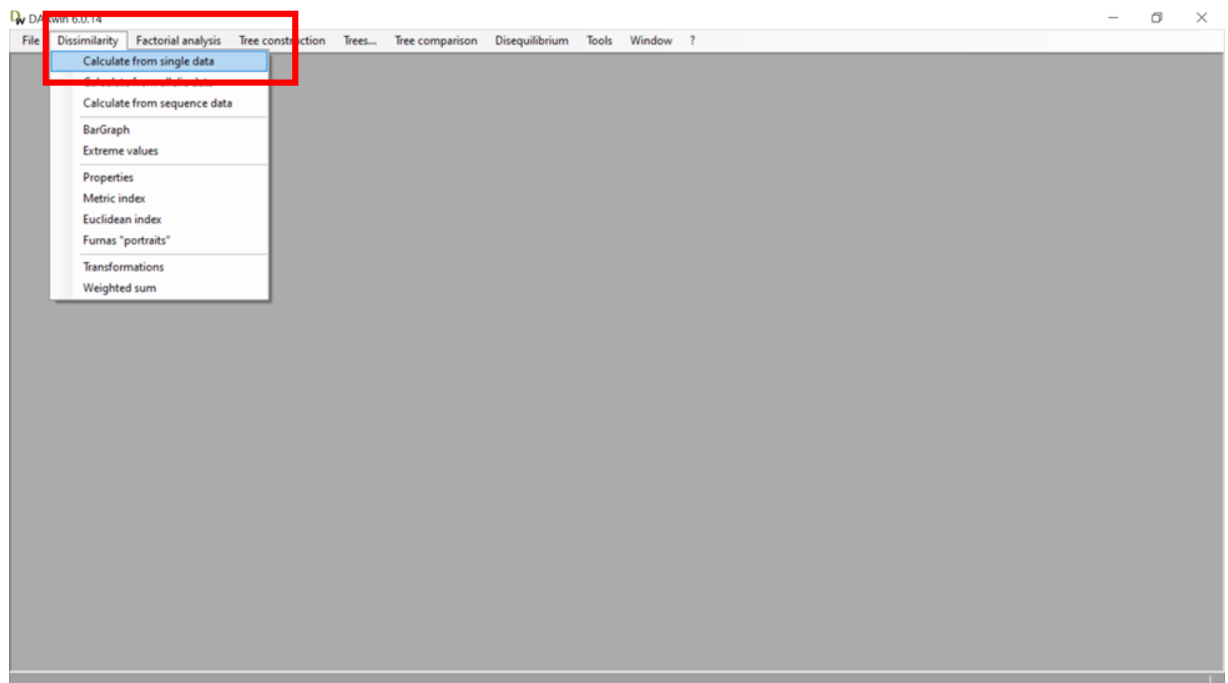
3. Karena contoh data yang digunakan merupakan data biner, pilih *Single* data kemudian klik *Save Data as* dan simpan data.



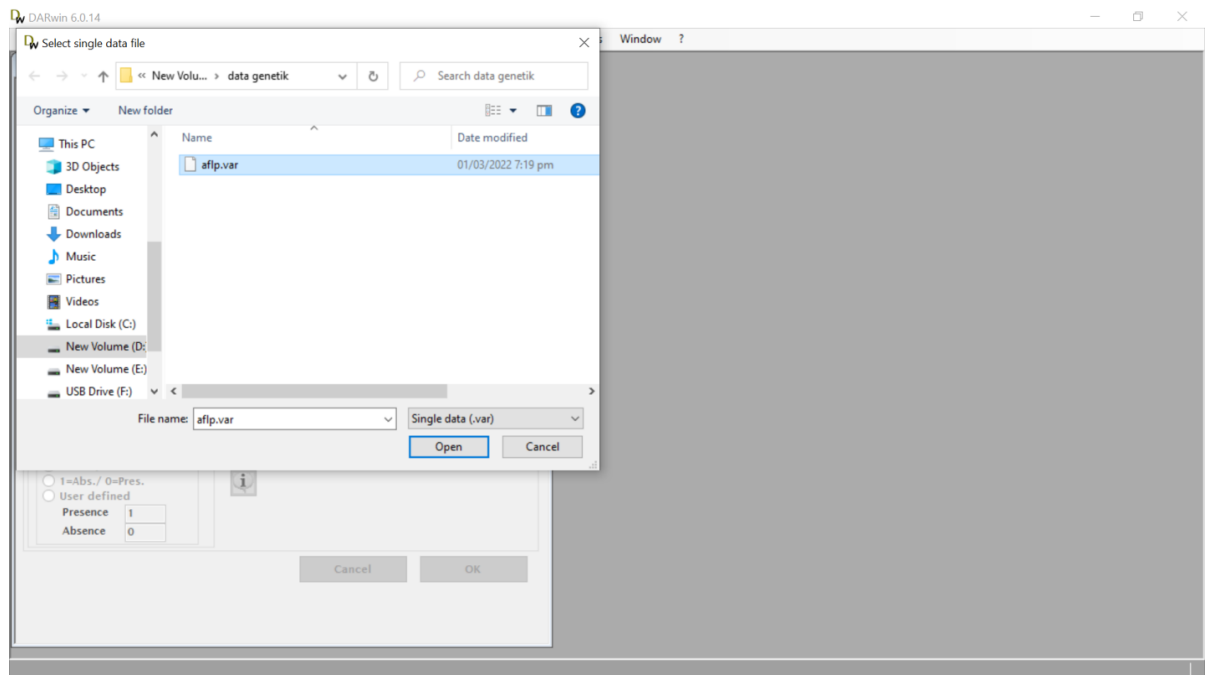
- Setelah data tersimpan dan mengklik OK akan muncul tampilan seperti gambar di bawah ini. File akan tersimpan dengan ekstensi .don sebagai *object identifier*. Klik OK. Lalu layar akan kembali ke menu utama.



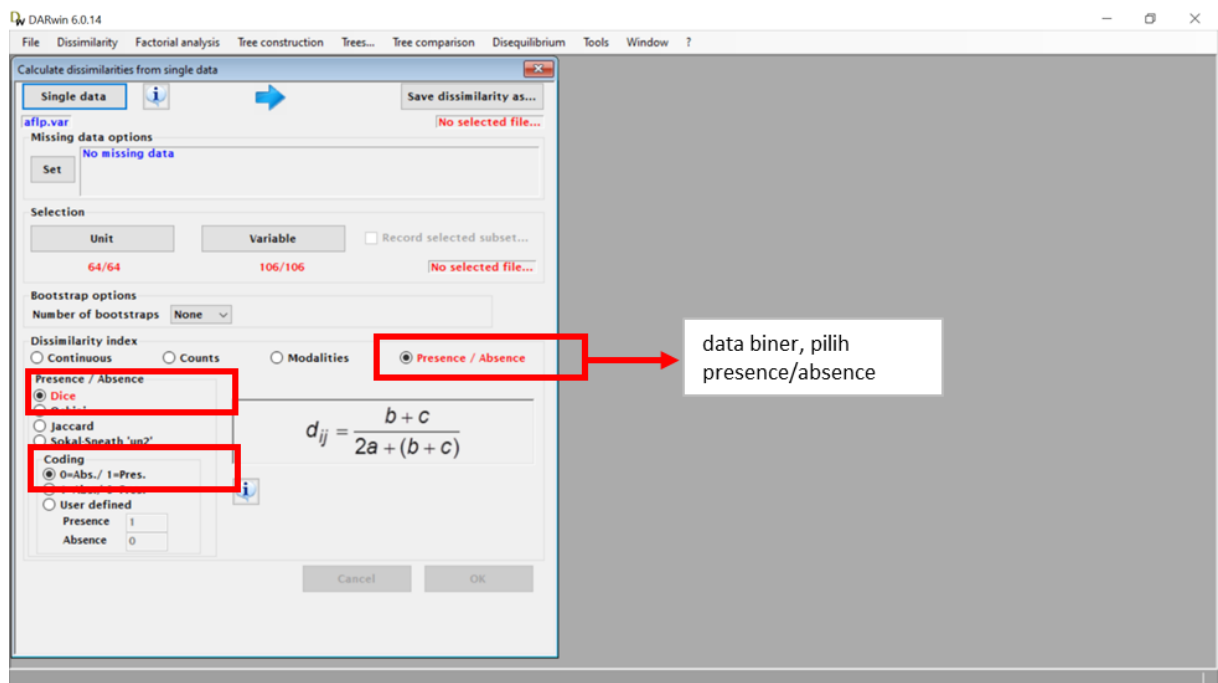
- Pada menu utama, pilih *Dissimilarity*, kemudian pilih *Calculate from single data*.



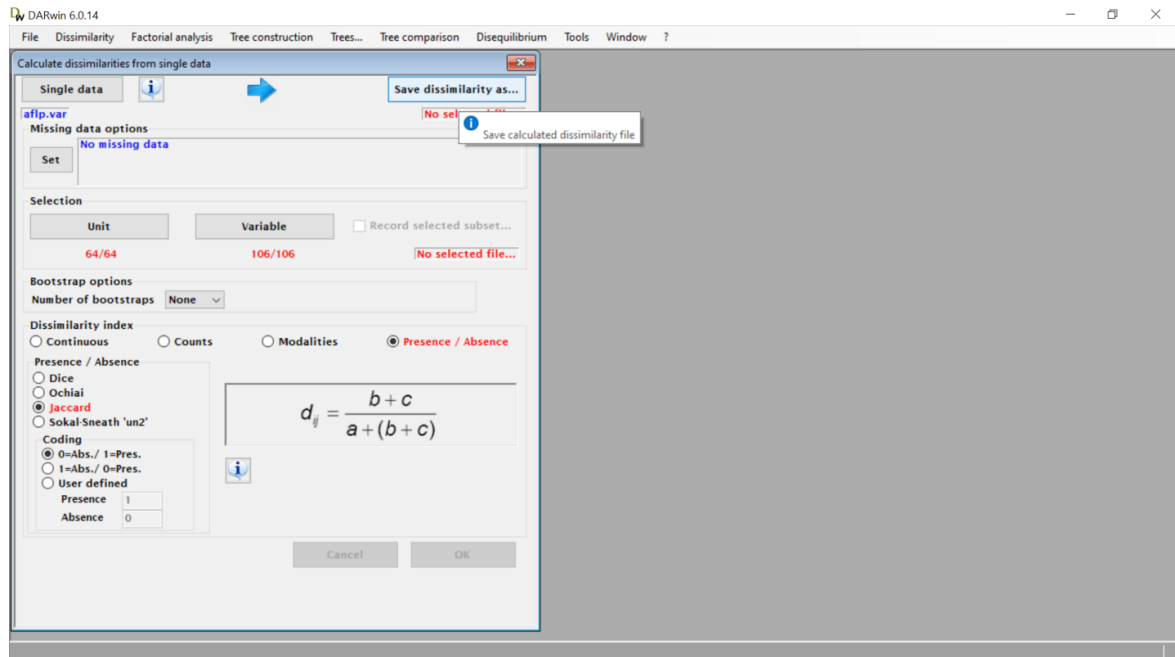
6. Kemudian pilih file yang tersimpan dengan ekstensi.var



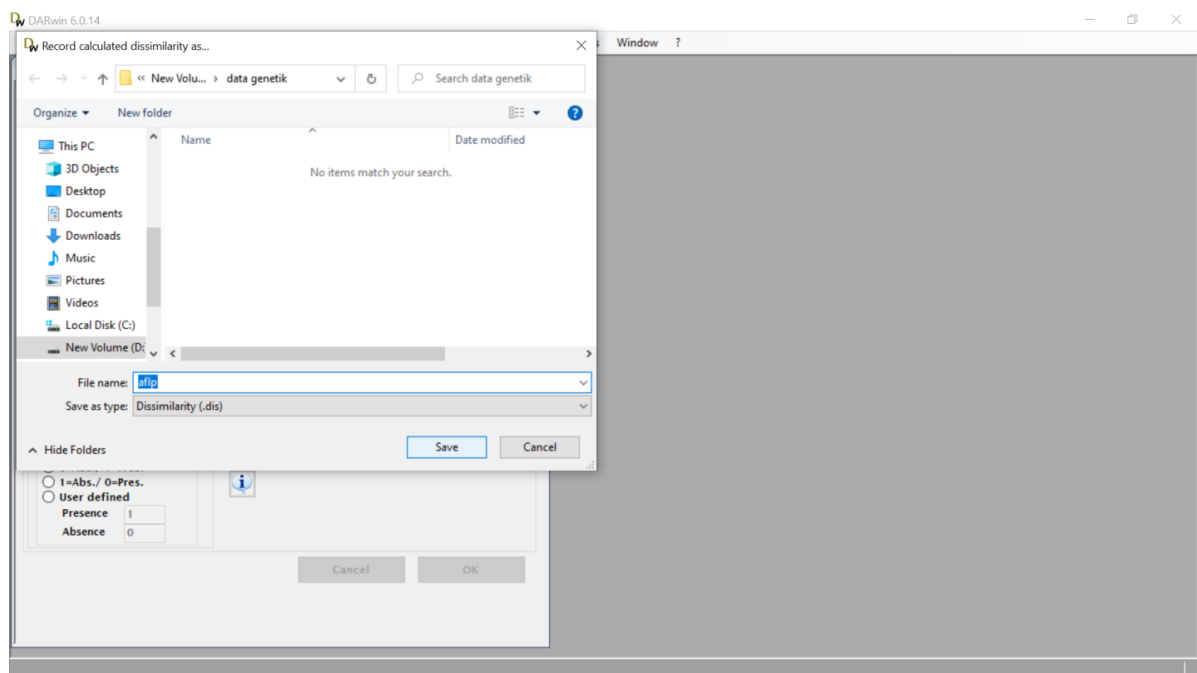
7. Setelah terbuka, akan muncul tampilan seperti di gambar di bawah ini. Untuk koefisien yang digunakan, jika individu atau sampel berkerabat dekat, sebaiknya menggunakan koefisien Jaccard atau Dice (Dalirsefat *et al.* 2009).



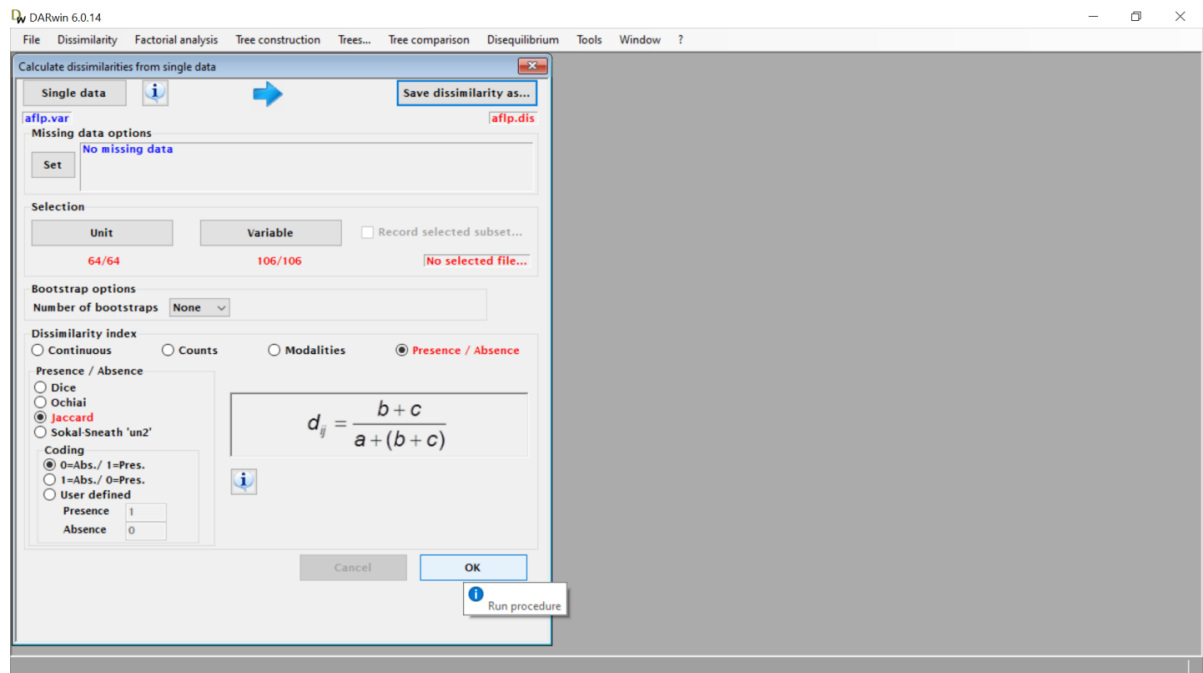
8. Kemudian klik *Save dissimilarity as* untuk menyimpan file *output dissimilarity*.



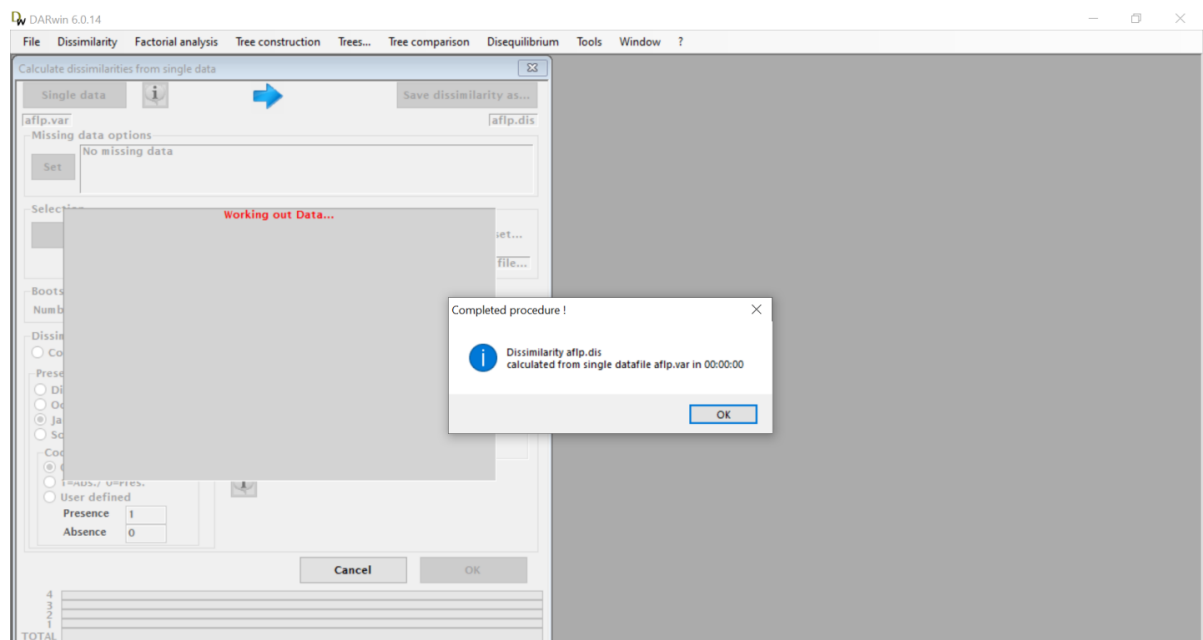
9. Data akan tersimpan dalam ekstensi .dis



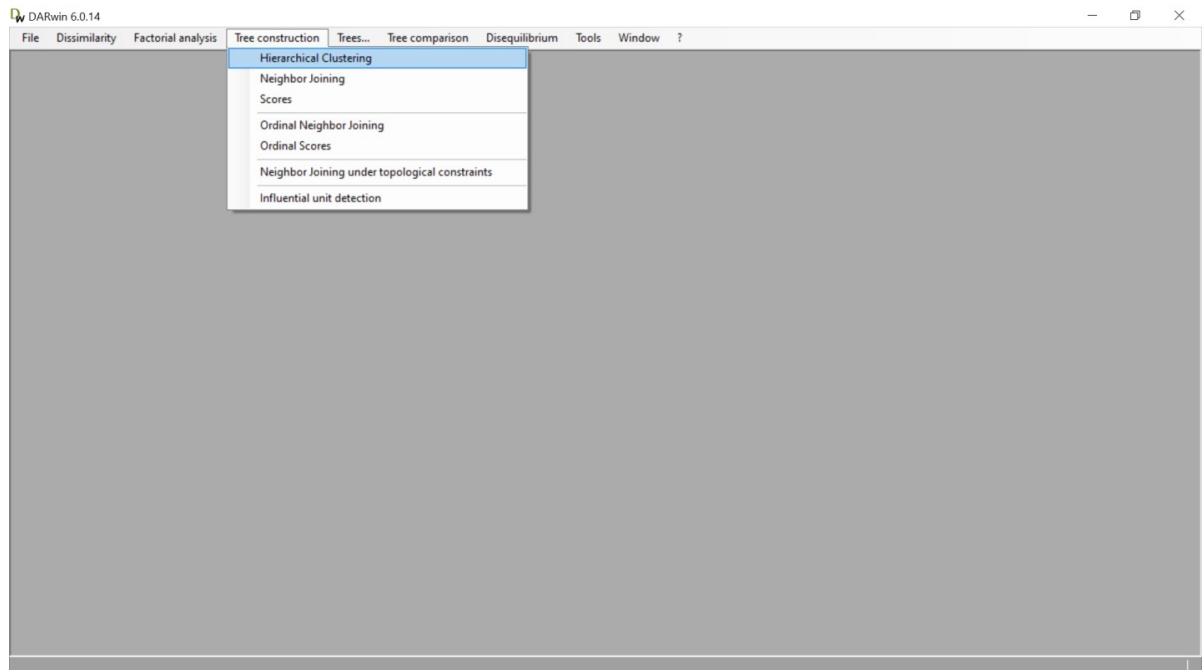
10. Setelah semua parameter dipilih dan nama output sudah disiapkan, klik ok.



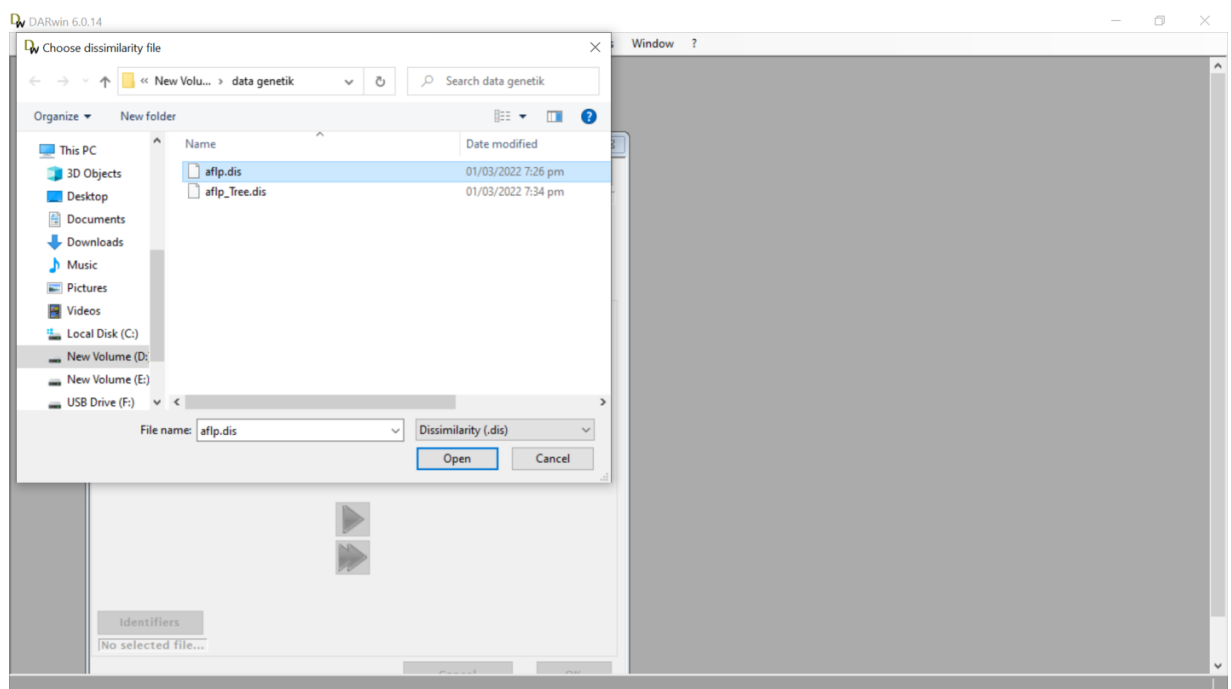
11. Jika analisis jarak genetik sudah selesai dilakukan, akan muncul tampilan seperti di bawah ini.



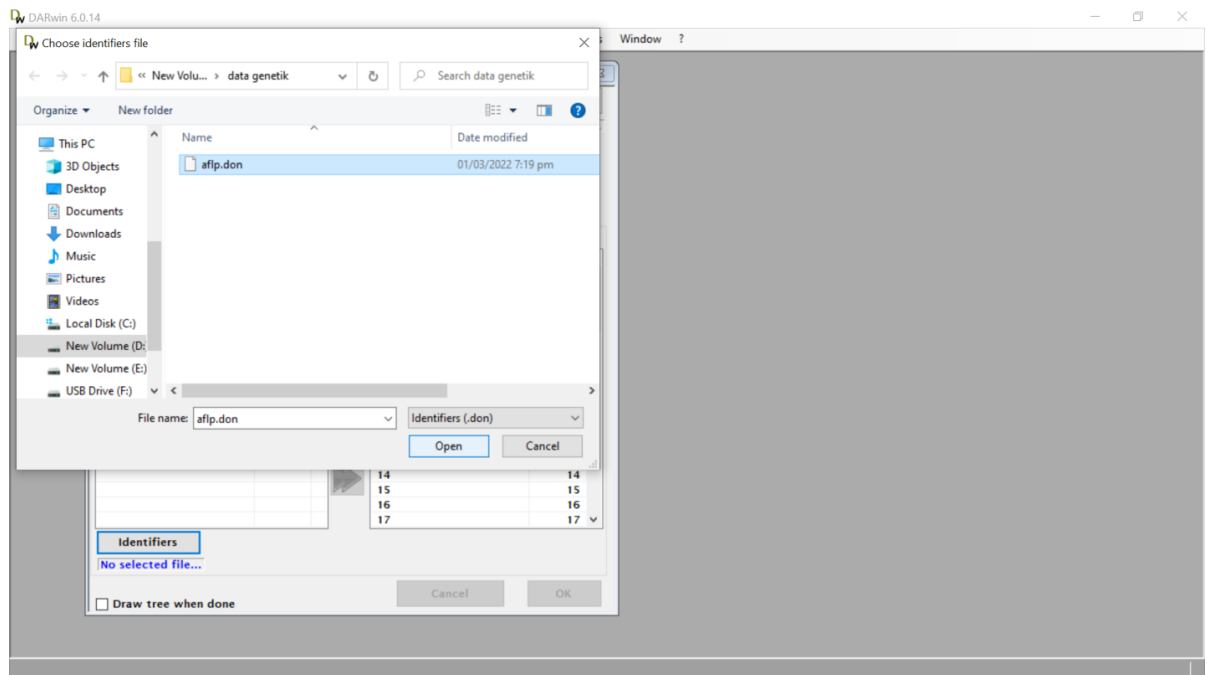
12. Pada *toolbar*, pilih *tree construction* kemudian pilih metode yang ingin digunakan. Untuk contoh digunakan *hierarchial clustering*.



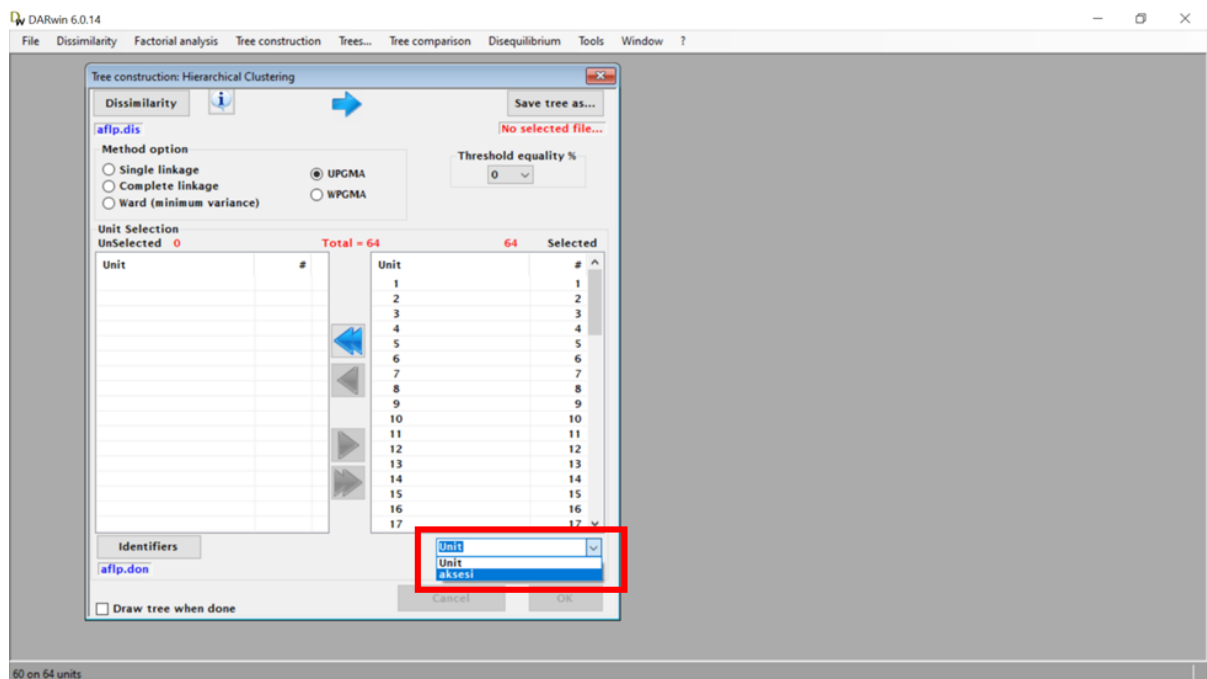
13. Kemudian pilih file *output dissimilarity* (.dis) yang akan digunakan untuk membuat pohon filogeni



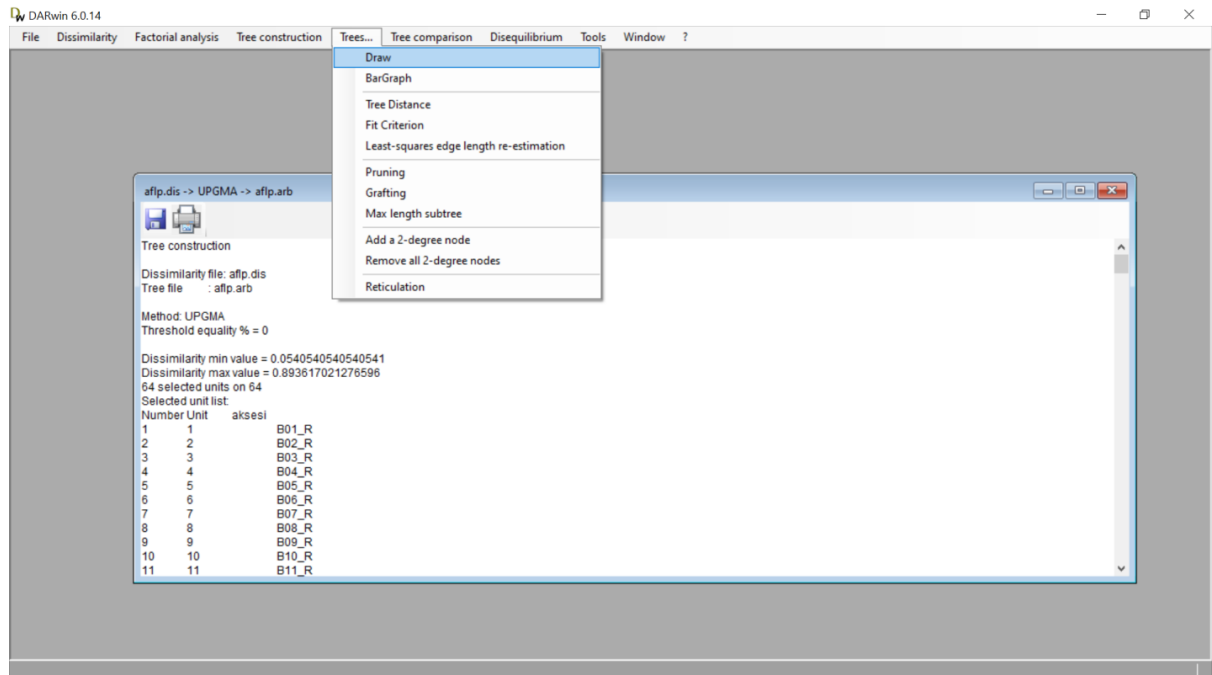
14. Pada tab *tree construction*, pilih *identifiers*, kemudian pilih file dengan ekstensi *.don*.



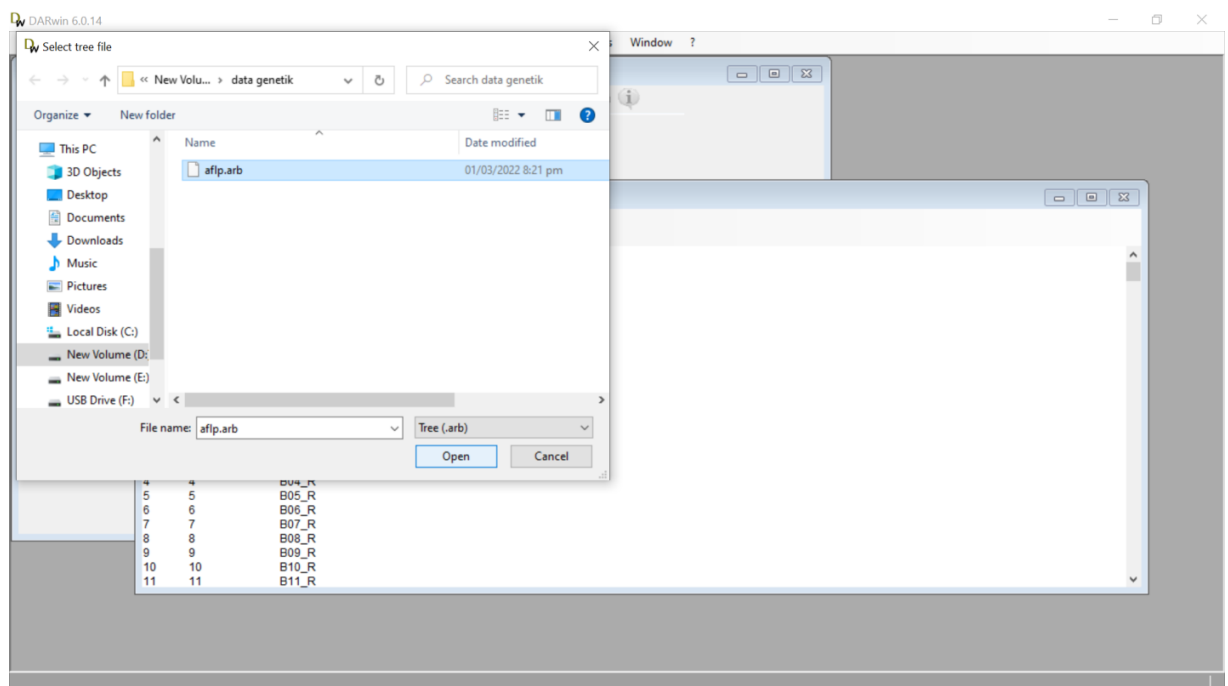
15. Kemudian pilih aksesi pada pilihan unit. Selanjutnya, pilih metode yang akan digunakan (pada tampilan contoh, metode yang dipilih yaitu UPGMA). Pilih *save tree as* untuk menyimpan file pohon filogeni. Jika ingin *tree* (pohon filogenetik) ditampilkan secara otomatis, pilih *draw tree when done*.



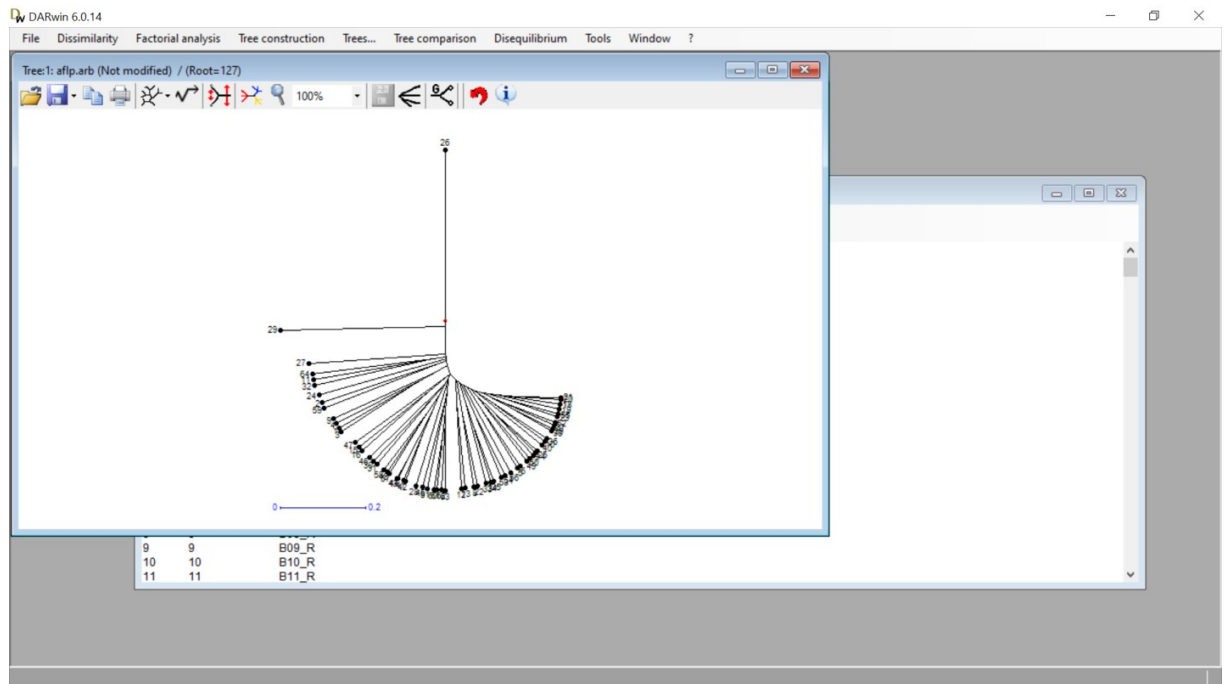
16. Jika tidak memilih *Draw tree when done*, munculkan *tree* otomatis dengan cara klik *tab Trees* pada menu utama, kemudian langsung klik *draw*.



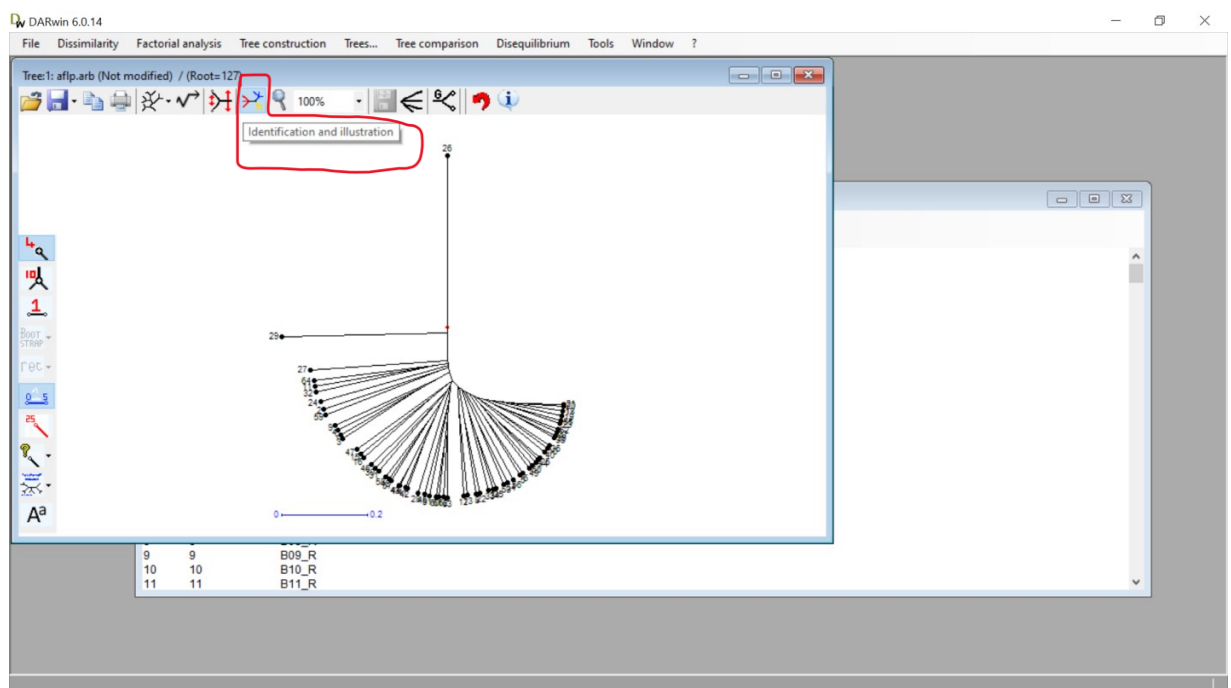
17. Pilih file dengan ekstensi .arb.



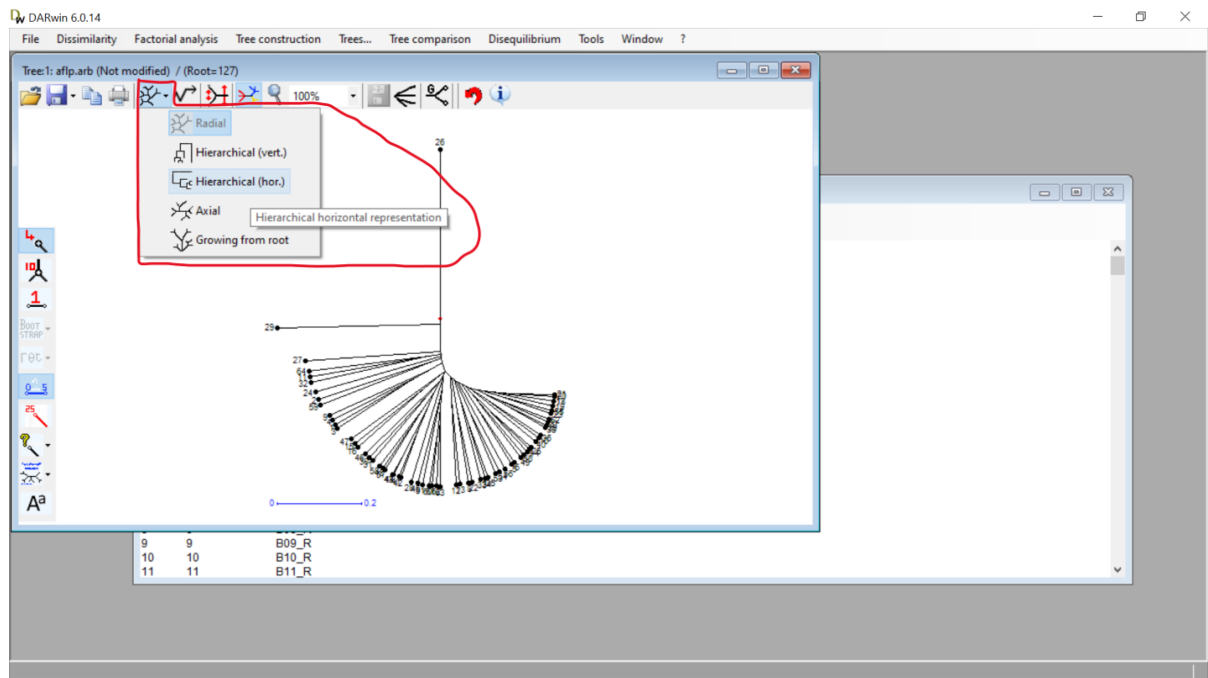
18. Akan muncul tampilan *tree* seperti gambar di bawah ini



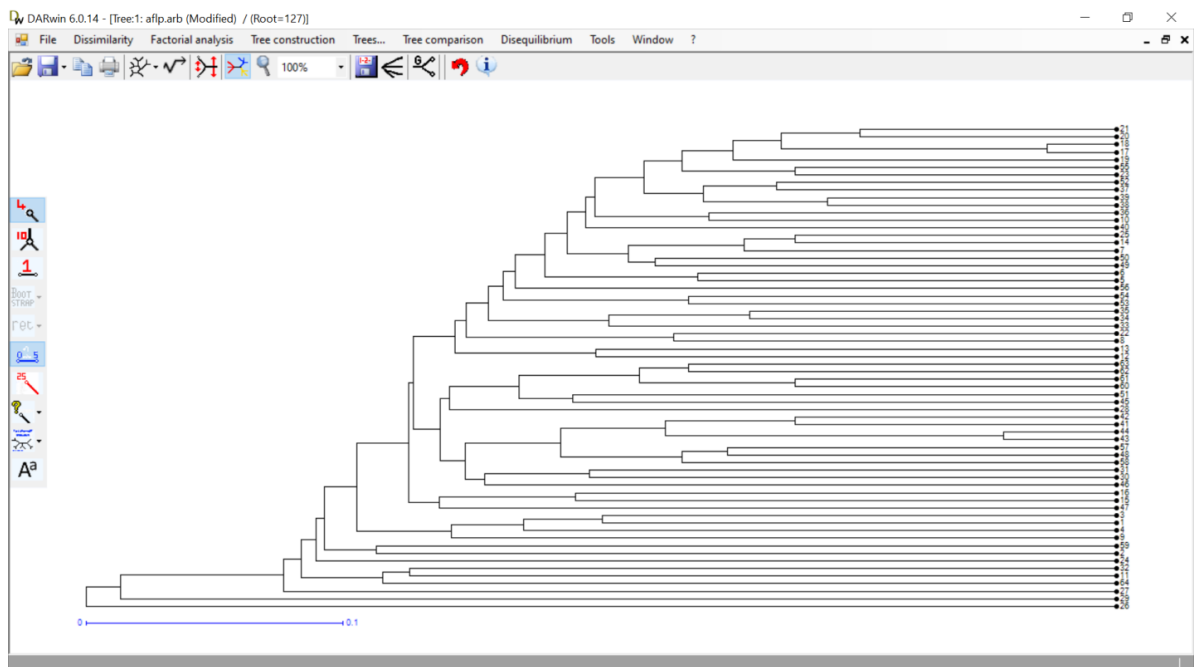
19. Untuk mengubah tampilan *tree*, klik ikon di sebelah ikon zoom. Ikon tersebut berwarna merah, biru, dan kuning. Untuk memperjelas, ikon yang dilingkari dengan warna merah.



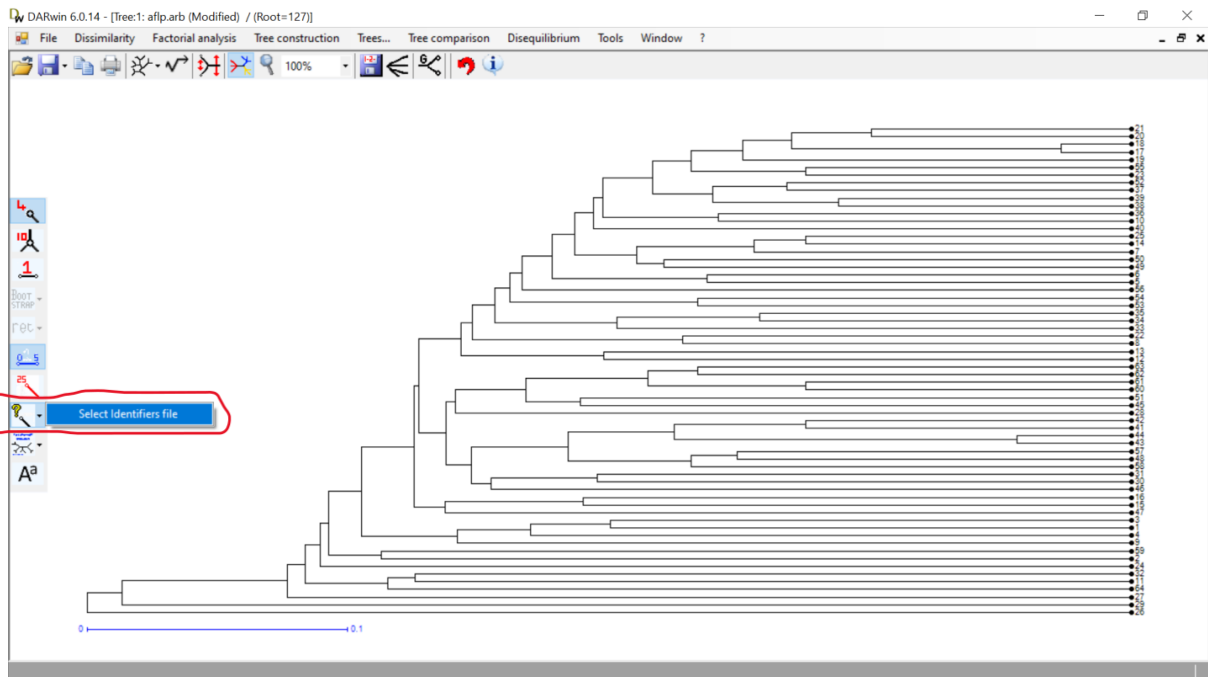
20. Untuk mengubah bentuk *tree*, klik ikon cabang di samping ikon *print*. Pada contoh, *tree* diubah menjadi bentuk *hierarchial horizontal*. Untuk memperjelas, ikon cabang dilingkari dengan warna merah.



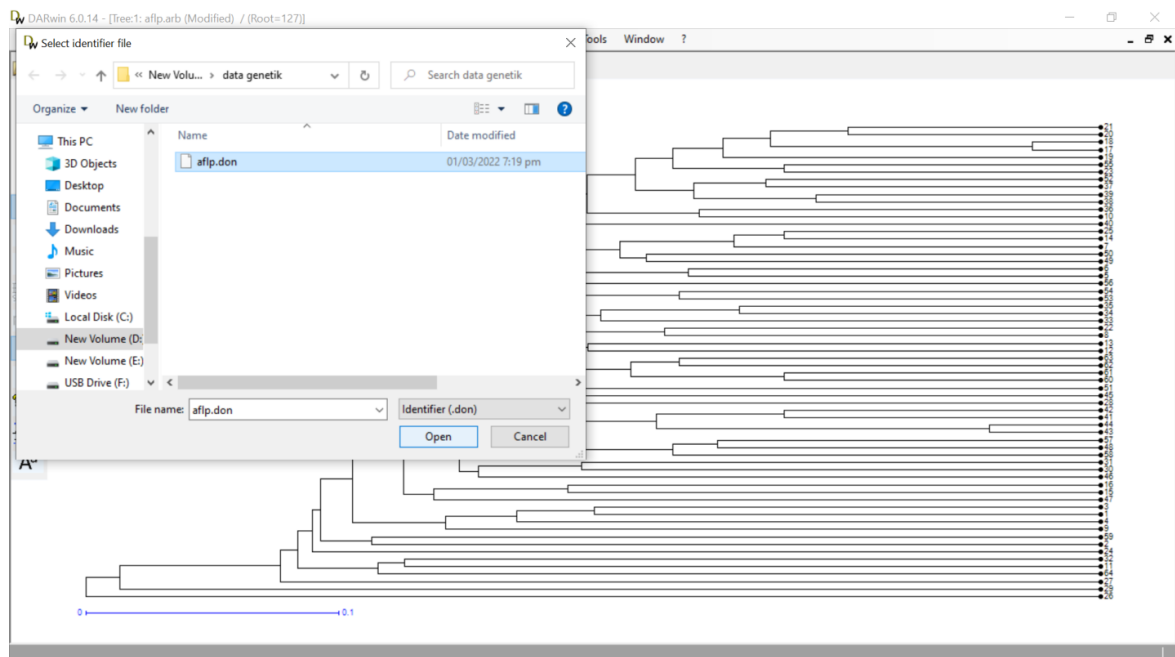
21. Hasilnya adalah seperti gambar di bawah ini



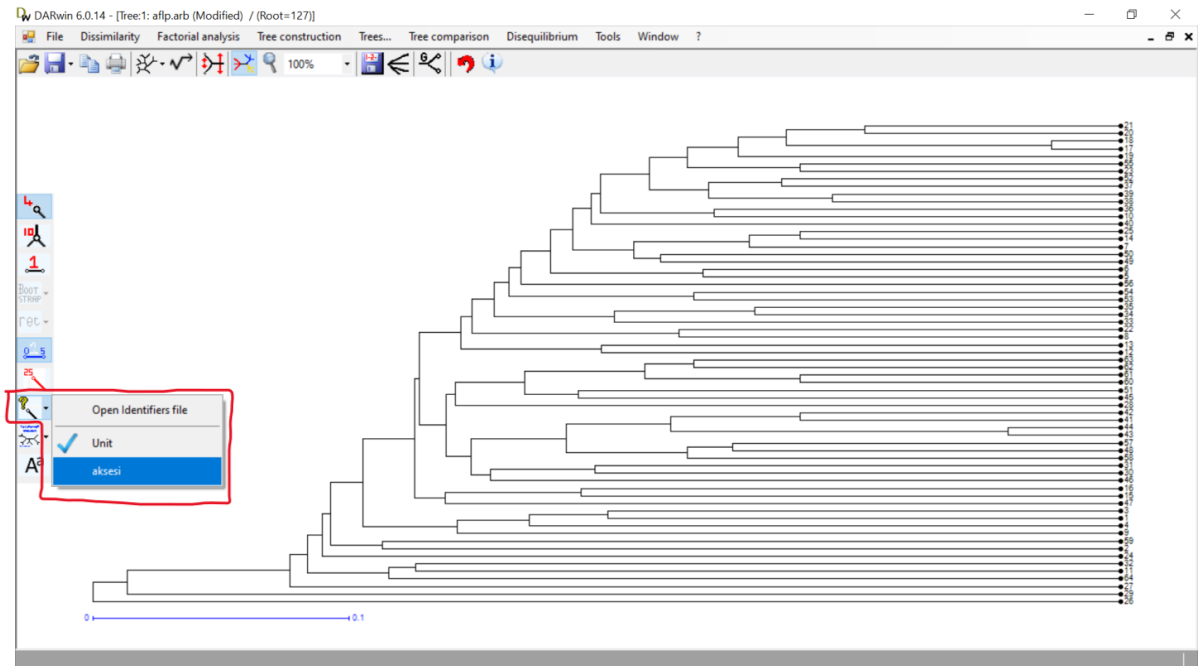
22. Untuk mengubah tampilan nama sampel di ujung *tree*, pilih ikon wand dengan tanda tanya di kiri, kemudian klik *select identifier file*. Untuk lebih jelasnya, ikon dilingkari dengan warna merah.



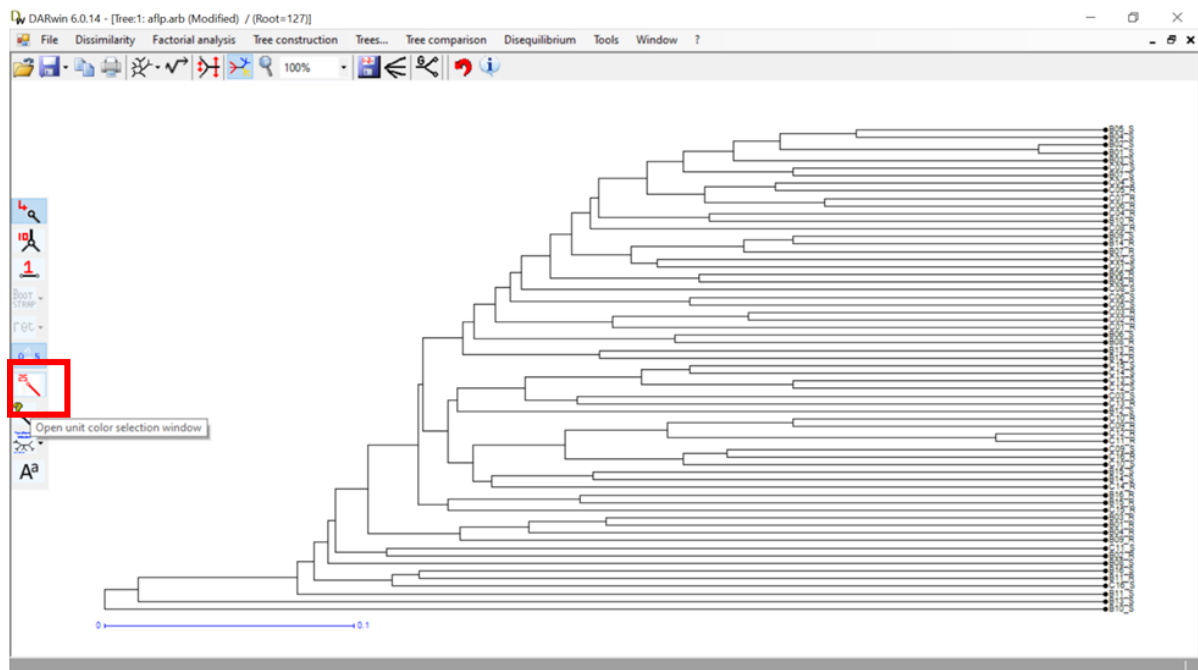
23. Pilih file dengan ekstensi .don.



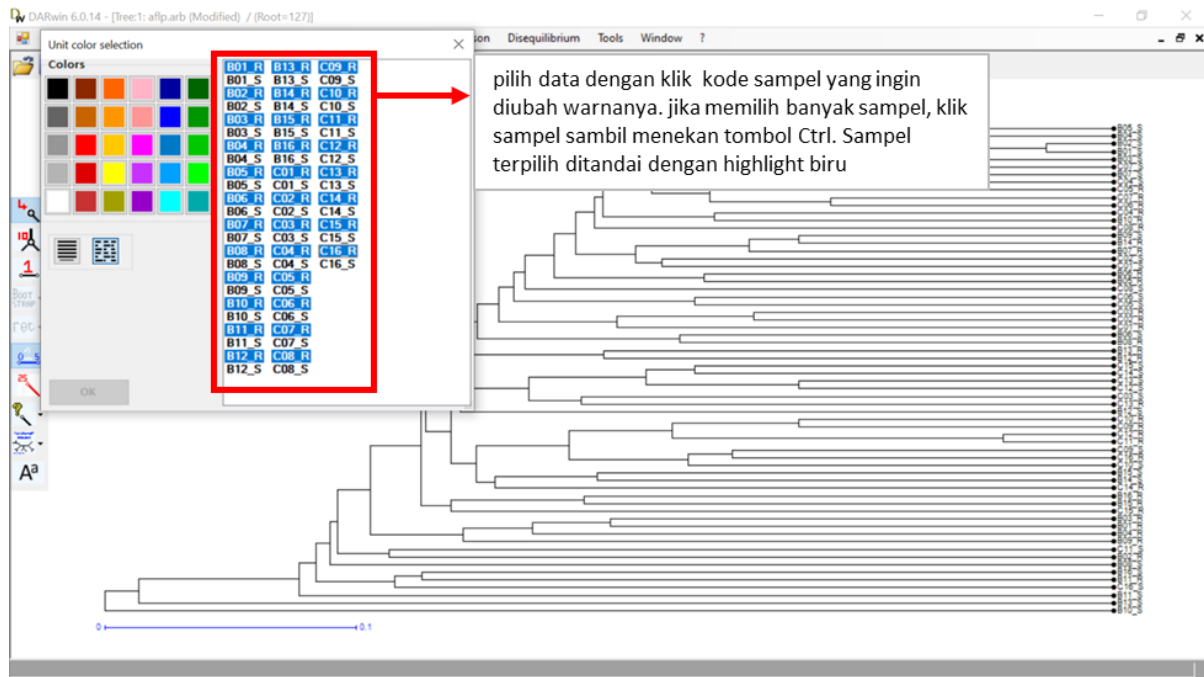
24. Setelah itu klik lagi ikon *identifier* dan pilih akses. Untuk lebih jelasnya, ikon dilingkari dengan warna merah.



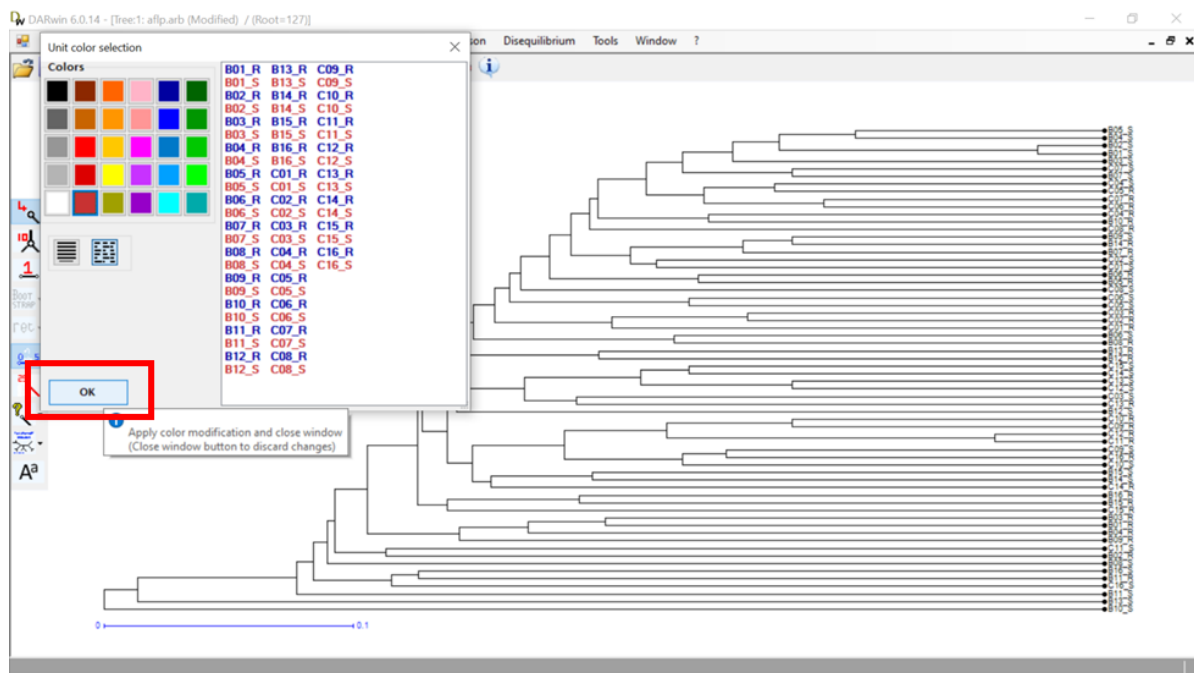
25. Jika ingin membedakan sampel menggunakan warna, pilih ikon di atas ikon *identifier*. Untuk lebih jelasnya, ikon dilingkari dengan warna merah.



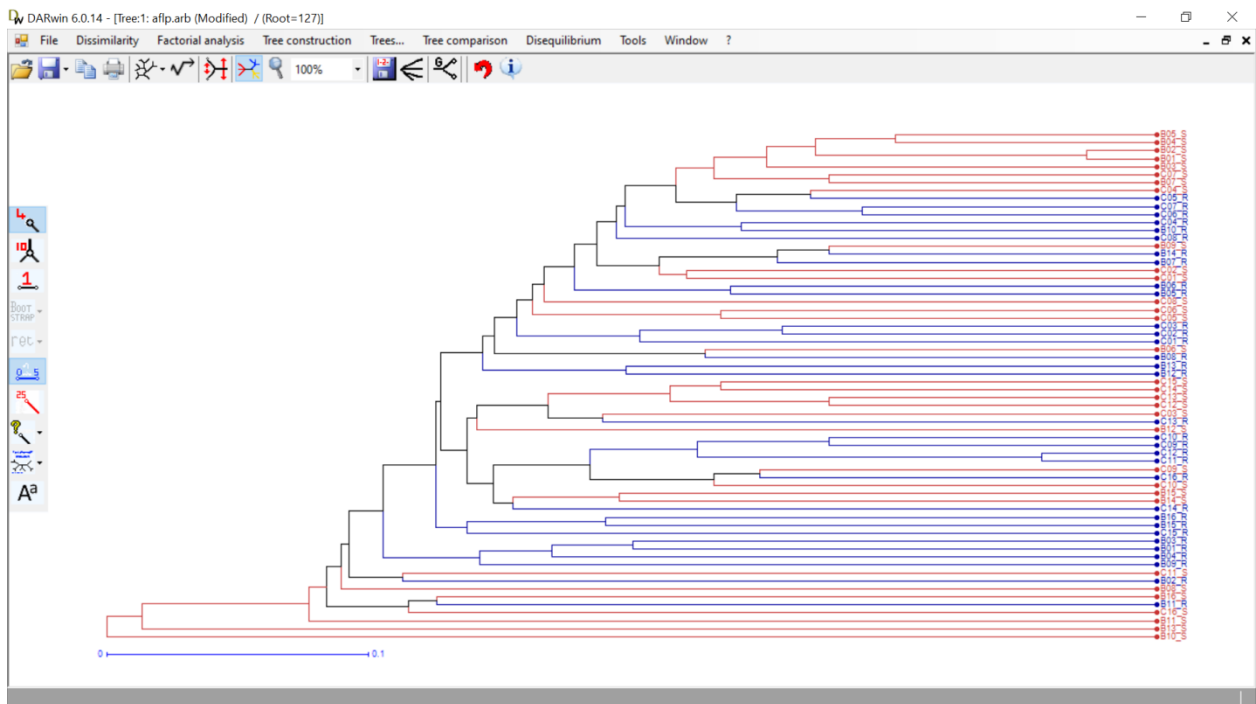
26. Kemudian pilih sampel yang akan diubah warnanya.



27. Bisa dilakukan pemilihan *multisample* dengan menekan tombol *control* (ctrl) pada keyboard saat memilih sampel. Kemudian pilih warna yang diinginkan. Setelah selesai, klik Ok.



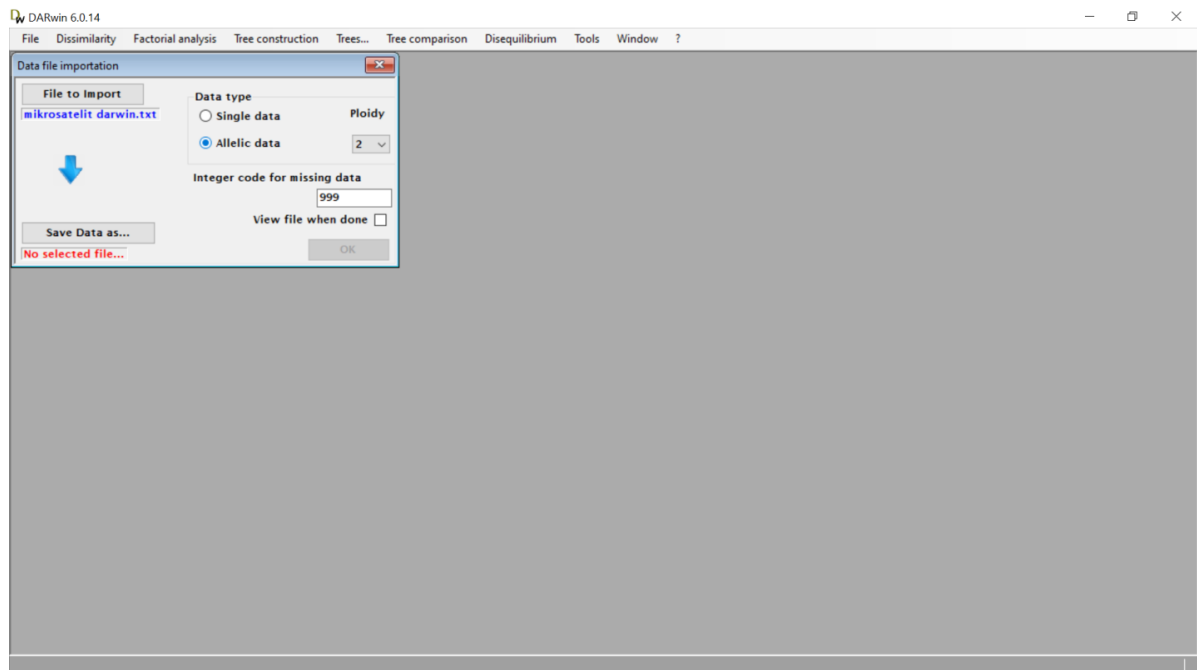
28. Gambar dapat dicopy ke aplikasi lain untuk disajikan atau disimpan.



29. Selain menggunakan data mentah/*raw data* berupa data biner, data yang digunakan untuk analisis menggunakan DARwin dapat berupa data *distance/dissimilarity* dari NTsys atau PHYLIP. Data tersebut di-input dengan memilih *import dissimilarity* pada menu file.



30. Jika ingin menggunakan data *mikrosatelit* maka pada saat Import file pilih *Allelic data* pada tab *import file*.



31. Format data berupa text file (.txt) dengan format seperti gambar di bawah ini.

mikrosatelit darwin - Notepad

Pop	AI_5	AI_11	AI_13	AI_34	AI_4
Nagrak1	1	2	1	2	2
Nagrak2	1	2	1	2	2
Nagrak3	1	2	1	3	2
Nagrak4	1	2	2	2	2
Nagrak5	1	2	2	2	1
Nagrak6	1	2	2	2	3
Nagrak7	1	2	2	2	2
Nagrak8	1	2	2	2	2
Nagrak9	1	2	2	2	2
Nagrak10	1	2	1	2	2
Nagrak11	1	2	1	2	2
Nagrak12	1	2	1	2	2
Nagrak13	1	2	1	2	2
Nagrak14	1	2	1	2	2
Nagrak15	1	2	1	2	2
Nagrak16	1	2	1	2	2
Nagrak17	1	2	1	2	2
Nagrak18	1	2	1	2	2
Nagrak19	1	2	1	2	1
Nagrak20	1	2	1	2	1
Babakah Rema1	1	1	1	2	1
Babakah Rema2	1	1	1	2	1
Babakah Rema3	1	1	1	2	1
Babakah Rema4	1	1	1	2	1
Babakah Rema5	1	1	1	2	1
Babakah Rema6	1	1	1	2	1
Babakah Rema7	1	1	1	2	1
Babakah Rema8	1	1	1	2	1
Babakah Rema9	1	1	1	2	1
Babakah Rema10	1	1	1	2	1
Babakah Rema11	1	1	1	2	1
Babakah Rema12	1	1	1	2	1
Babakah Rema13	1	1	1	2	1
Babakah Rema14	1	1	1	2	1
Babakah Rema15	1	1	1	2	1
Babakah Rema16	1	1	1	2	1

Ln 121, Col 30 100% Windows (CRLF) UTF-8

32. Untuk menghitung *Disiimilarity*, pada tab *Dissimilarity* pilih *calculate from allelic data*.



33. Jika *dissimilarity* telah didapatkan, langkah selanjutnya sama seperti pada data biner.

1.4 Pengolahan data fragmen DNA menggunakan GenAlEx 6.5

GenAlEx (*Genetic Analysis in Excel*) merupakan *software extension* atau *add-ins* pada *Microsoft Excel* yang berfungsi untuk mengolah data genetik berupa haploid, biner, alelik, maupun sekuens (Peakall & Smouse 2012). Analisis yang dapat dilakukan yaitu analisis berbasis frekuensi (F-statistik, heterozigositas, HWE, kekerabatan), dan berdasarkan jarak genetik (AMOVA, PCoA, tes mantel, korelasi spasial), dan lain-lain. Pada contoh yang akan disampaikan, data yang akan digunakan berupa data alelik *mikrosatelit* hasil skoring dari gel elektroforesis. Untuk data *mikrosatelit*, jika sampel tidak divisualisasi dengan gel elektroforesis tetapi dianalisis ukuran fragmennya dengan mesin sekuenser maka angka skoring diganti dengan ukuran fragmen (bp). Langkah-langkah analisis menggunakan GenAlEx adalah sebagai berikut:

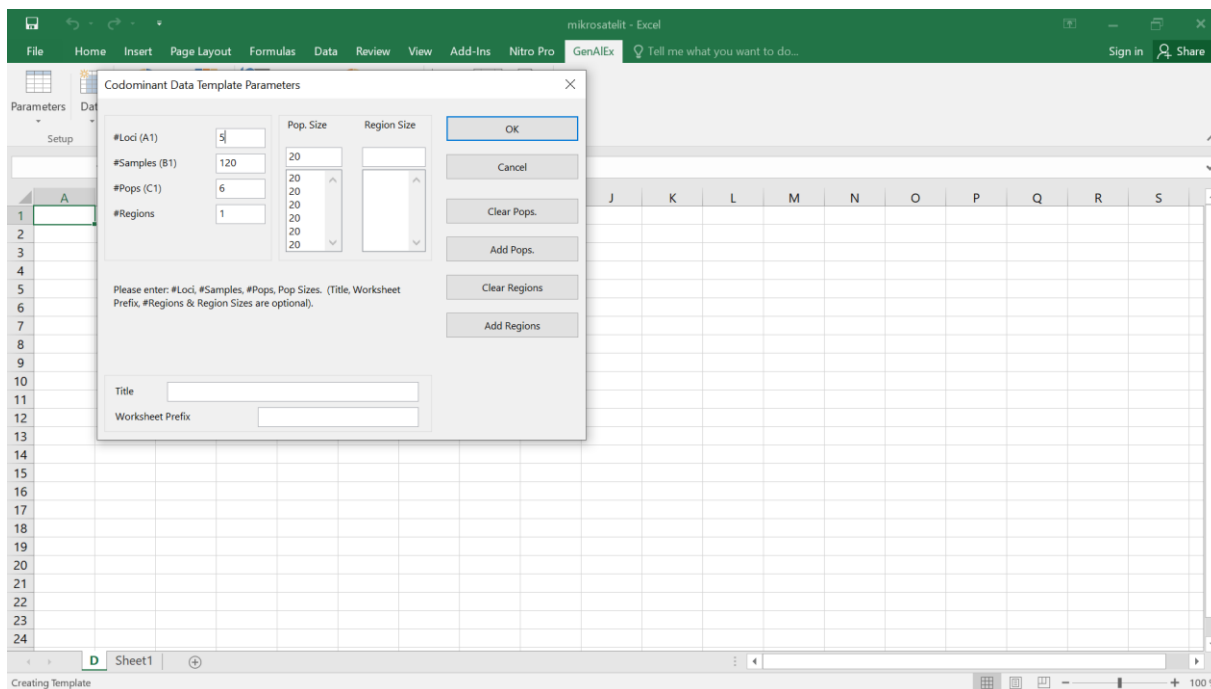
1. Siapkan data berupa tabel dalam *Spreadsheet Excel*. Format baris pertama adalah sample, pop, kemudian nama primer yang digunakan. Nama primer dituliskan di atas kolom pertama lokus yang dihasilkan oleh primer tersebut.

Sample	Pop	AI_5	AI_11	AI_13	AI_34	AI_4
1 Nagrak1		1	2	1	2	1
2 Nagrak2		1	2	1	2	1
3 Nagrak3		1	2	1	3	1
4 Nagrak4		1	2	1	2	1
5 Nagrak5		1	2	1	2	1
6 Nagrak6		1	2	1	2	1
7 Nagrak7		1	2	1	2	1
8 Nagrak8		1	2	1	2	1
9 Nagrak9		1	2	1	2	1
10 Nagrak10		1	2	1	2	1
11 Nagrak11		1	2	1	2	1
12 Nagrak12		1	2	1	2	1
13 Nagrak13		1	2	1	2	1
14 Nagrak14		1	2	1	2	1
15 Nagrak15		1	2	1	2	1
16 Nagrak16		1	2	1	2	1
17 Nagrak17		1	2	1	2	1
18 Nagrak18		1	2	1	2	1
19 Nagrak19		1	2	1	1	1
20 Nagrak20		1	2	1	1	1
21 Babakah Rema1		1	1	1	1	1
22 Babakah Rema2		1	1	1	1	1
23 Babakah Rema3		1	1	1	1	1

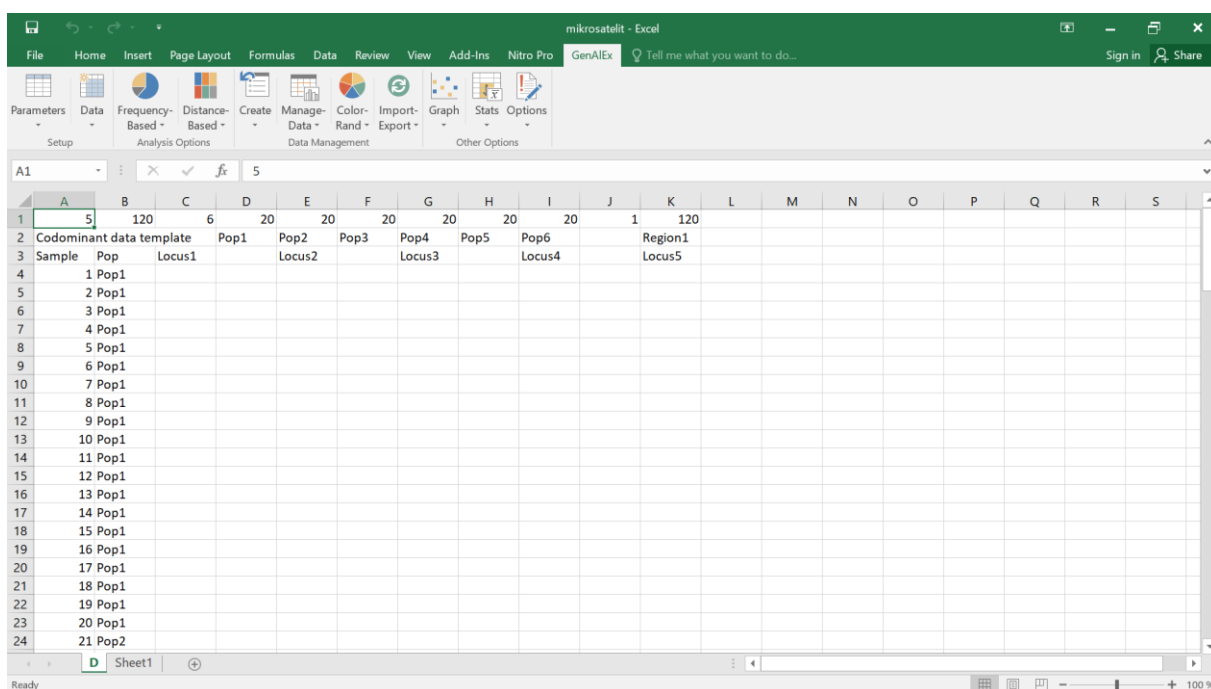
2. Untuk data *mikrosatelit*, klik *Create* lalu langsung klik *Codominant Template*, sedangkan untuk data biner klik *Create*, lalu langsung klik *Binary Template*.

Sample	Pop	AI_5	AI_11	AI_13	AI_34	AI_4
1 Nagrak1		1	2	1	2	1
2 Nagrak2		1	2	1	2	1
3 Nagrak3		1	2	1	3	1
4 Nagrak4		1	2	1	2	1
5 Nagrak5		1	2	1	2	1
6 Nagrak6		1	2	1	2	1
7 Nagrak7		1	2	1	2	1
8 Nagrak8		1	2	1	2	1
9 Nagrak9		1	2	1	2	1
10 Nagrak10		1	2	1	2	1
11 Nagrak11		1	2	1	2	1
12 Nagrak12		1	2	1	2	1
13 Nagrak13		1	2	1	2	1
14 Nagrak14		1	2	1	2	1
15 Nagrak15		1	2	1	2	1
16 Nagrak16		1	2	1	2	1
17 Nagrak17		1	2	1	2	1
18 Nagrak18		1	2	1	2	1
19 Nagrak19		1	2	1	1	1
20 Nagrak20		1	2	1	1	1
21 Babakah Rema1		1	1	1	1	1
22 Babakah Rema2		1	1	1	1	1
23 Babakah Rema3		1	1	1	1	1

- Data yang digunakan di dalam contoh berikut ini terdiri dari 120 sampel dari 6 lokasi (masing-masing lokasi memiliki 20 sampel) dan menggunakan 5 primer. Populasi ditambahkan dengan menekan tombol *Add Pops*. Jika ingin membagi populasi berdasarkan wilayah, misalnya populasi lokasi di Jawa Barat dan Jawa Tengah, maka bisa ditambahkan dengan menekan tombol *Add Regions*.



- Setelah meng-klik OK, akan muncul tampilan *template* seperti gambar di bawah ini. Pada kolom yang kosong, *copy-paste* data dari *sheet data* (dalam gambar tertera sebagai '*Sheet1*') ke kolom-kolom tersebut. Nama sampel di kolom Pop dapat diganti dengan nama sampel asli.



5. Hasil tabel data yang sudah terisi sebagai berikut:

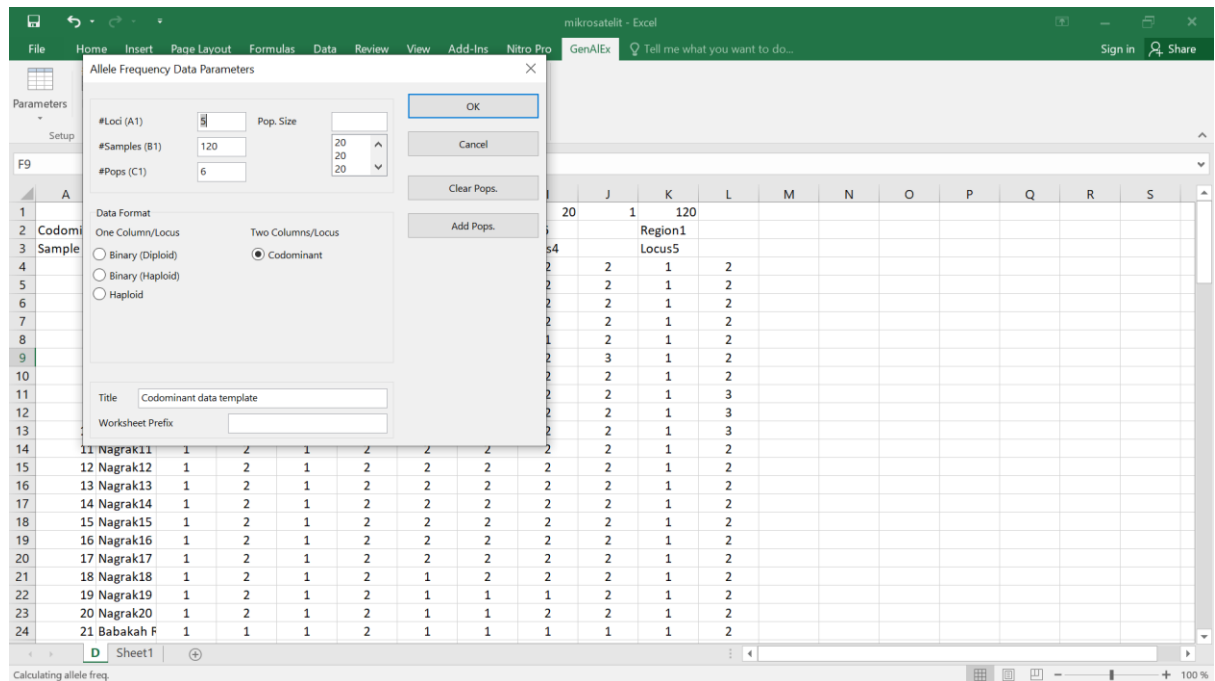
The screenshot shows the GenAIEx Excel spreadsheet with the following data:

Sample	Pop	Locus1	Pop1	Pop2	Pop3	Pop4	Pop5	Pop6	Region1
1 Nagrak1	1	2	1	2	1	2	2	2	1
2 Nagrak2	1	2	1	2	1	2	2	2	1
3 Nagrak3	1	2	1	2	1	3	2	2	1
4 Nagrak4	1	2	1	2	2	2	2	2	1
5 Nagrak5	1	2	1	2	2	2	1	2	1
6 Nagrak6	1	2	1	2	2	2	2	3	1
7 Nagrak7	1	2	1	2	2	2	2	2	1
8 Nagrak8	1	2	1	2	2	2	2	2	1
9 Nagrak9	1	2	1	2	2	2	2	2	1
10 Nagrak10	1	2	1	2	2	2	2	2	1
11 Nagrak11	1	2	1	2	2	2	2	2	1
12 Nagrak12	1	2	1	2	2	2	2	2	1
13 Nagrak13	1	2	1	2	2	2	2	2	1
14 Nagrak14	1	2	1	2	2	2	2	2	1
15 Nagrak15	1	2	1	2	2	2	2	2	1
16 Nagrak16	1	2	1	2	2	2	2	2	1
17 Nagrak17	1	2	1	2	2	2	2	2	1
18 Nagrak18	1	2	1	2	1	2	2	2	1
19 Nagrak19	1	2	1	2	1	1	2	2	1
20 Nagrak20	1	2	1	2	1	1	2	2	1
21 Babakah F	1	1	1	2	1	1	1	1	2

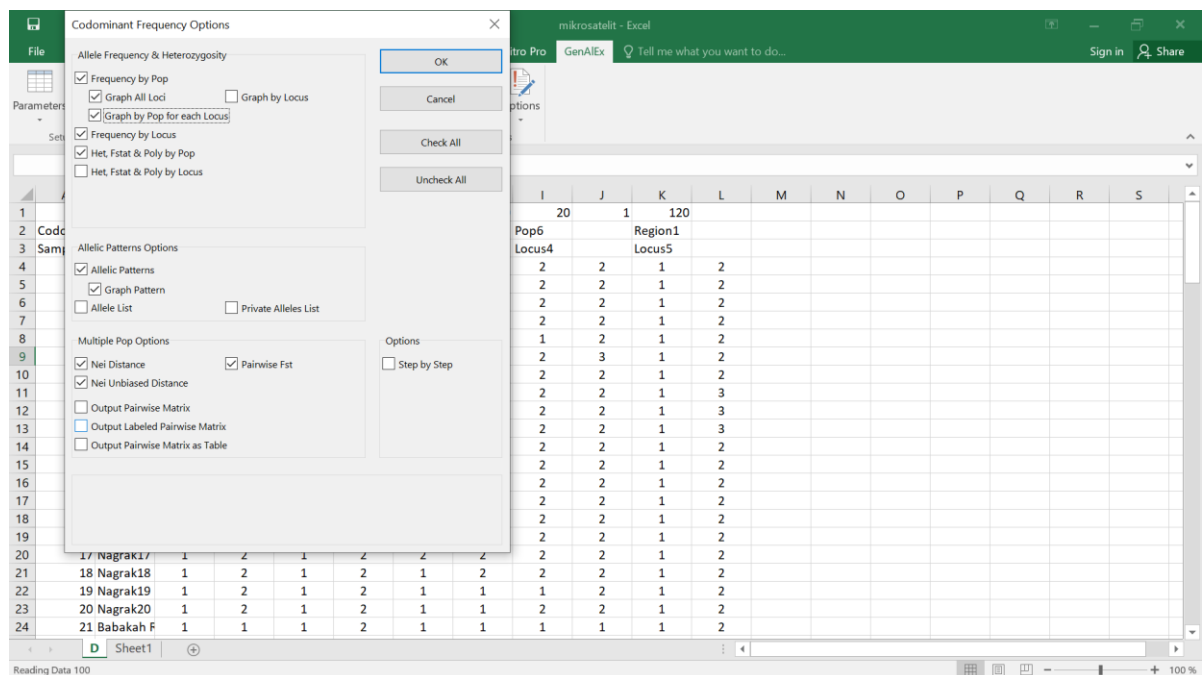
6. Untuk analisis berbasis frekuensi, klik *Frequency Based*, lalu langsung klik *Frequency*..

The screenshot shows the GenAIEx Excel spreadsheet with the 'Frequency' menu option highlighted. The data is the same as in the previous screenshot.

- Kemudian secara otomatis akan muncul tab berisi detail data. Secara otomatis akan terisi kodominan.

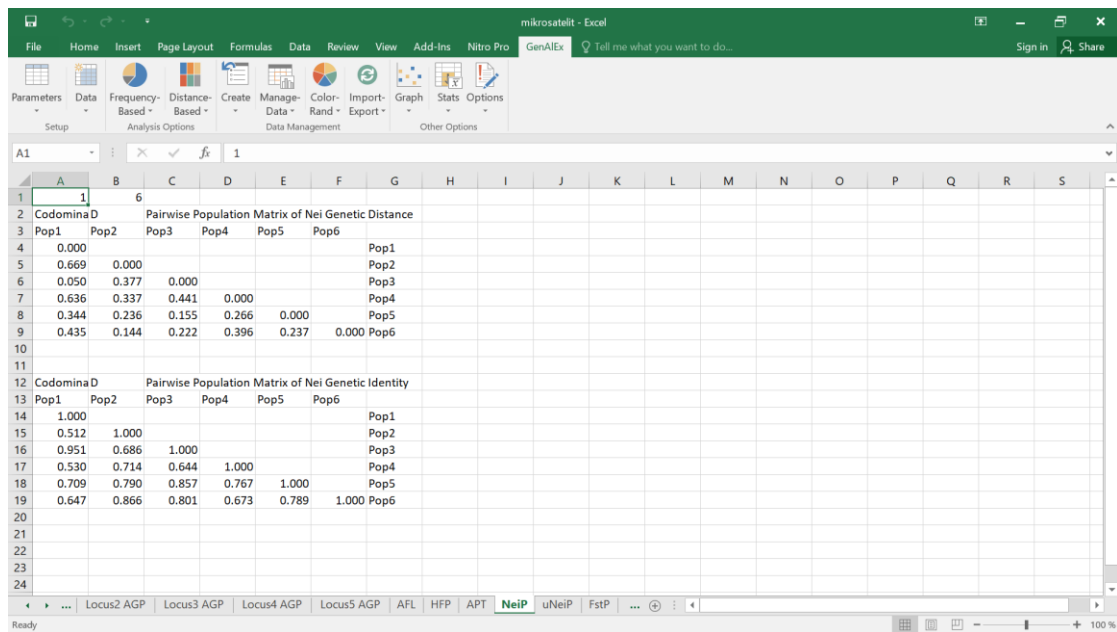


- Setelah itu, pilih analisis apa saja yang kita inginkan, kemudian klik OK.



9. Terlampir beberapa contoh *sheet* hasil (jarak genetik Nei, heterozigositas, frekuensi alel):

a. Jarak genetik Nei (Nei Genetic distance).

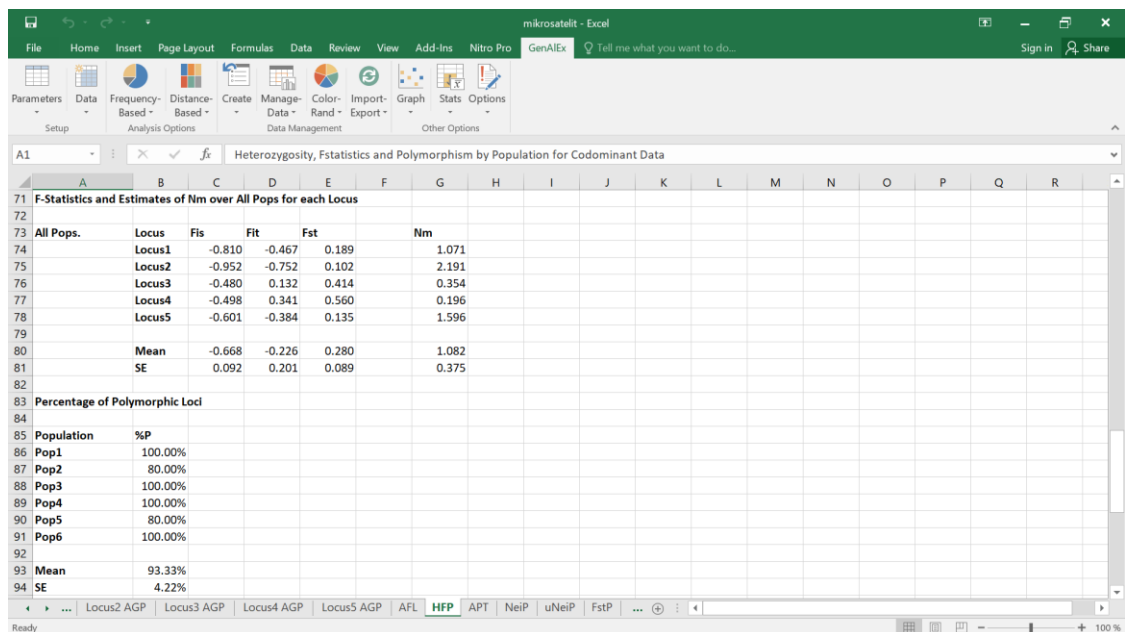


The screenshot shows two worksheets in the mikrosatellit - Excel application. The first worksheet, 'NeiP', displays the Pairwise Population Matrix of Nei Genetic Distance. The second worksheet, 'uNeiP', displays the Pairwise Population Matrix of Nei Genetic Identity. Both matrices are for six populations (Pop1 to Pop6).

	Pop1	Pop2	Pop3	Pop4	Pop5	Pop6
Pop1	0.000					
Pop2	0.669	0.000				
Pop3	0.050	0.377	0.000			
Pop4	0.636	0.337	0.441	0.000		
Pop5	0.344	0.236	0.155	0.266	0.000	
Pop6	0.435	0.144	0.222	0.396	0.237	0.000

	Pop1	Pop2	Pop3	Pop4	Pop5	Pop6
Pop1	1.000					
Pop2	0.512	1.000				
Pop3	0.951	0.686	1.000			
Pop4	0.530	0.714	0.644	1.000		
Pop5	0.709	0.790	0.857	0.767	1.000	
Pop6	0.647	0.866	0.801	0.673	0.789	1.000

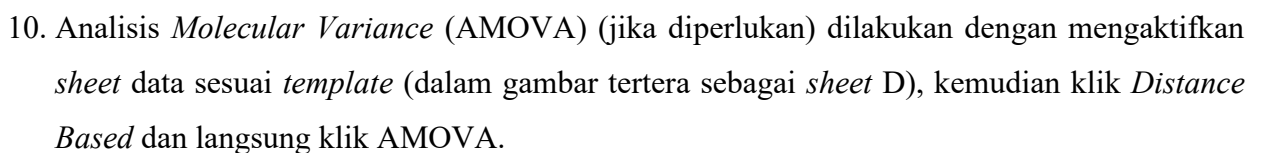
b. Heterozigositas (*Heterozygosity, Fstatistics, Polymorphism*)



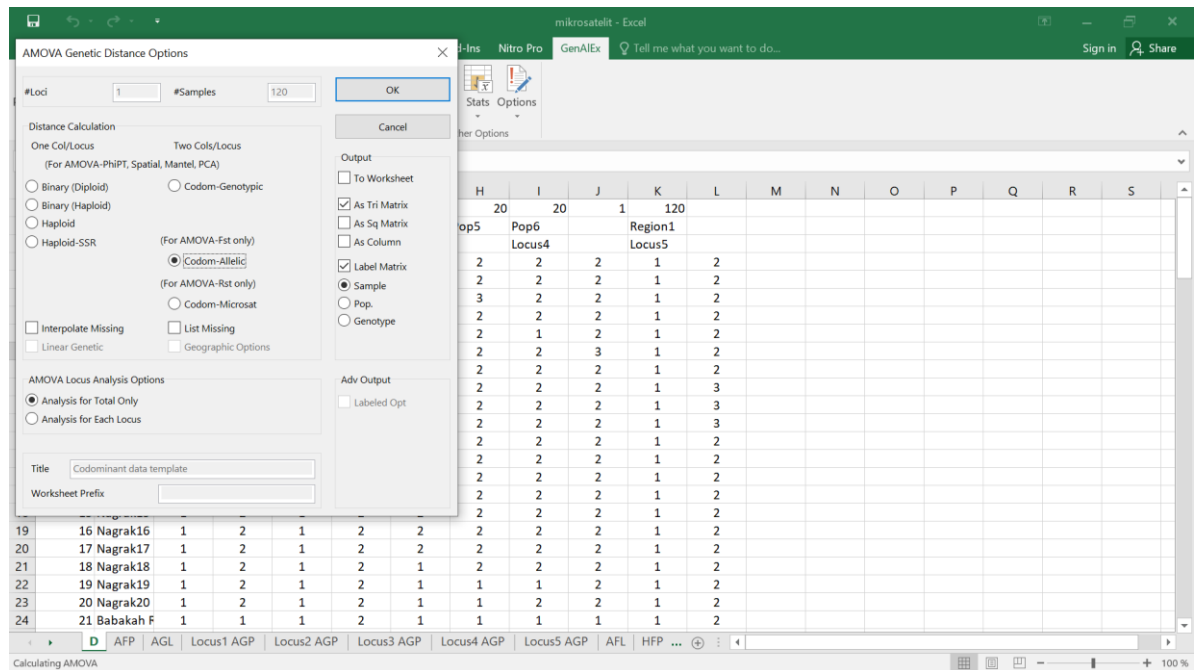
The screenshot shows the 'Heterozygosity, Fstatistics and Polymorphism by Population for Codominant Data' worksheet. It contains two main sections: 'F-Statistics and Estimates of Nm over All Pops for each Locus' and 'Percentage of Polymorphic Loci'.

All Pops.	Locus	Fis	Fit	Fst	Nm
	Locus1	-0.810	-0.467	0.189	1.071
	Locus2	-0.952	-0.752	0.102	2.191
	Locus3	-0.480	0.132	0.414	0.354
	Locus4	-0.498	0.341	0.560	0.196
	Locus5	-0.601	-0.384	0.135	1.596
	Mean	-0.668	-0.226	0.280	1.082
	SE	0.092	0.201	0.089	0.375

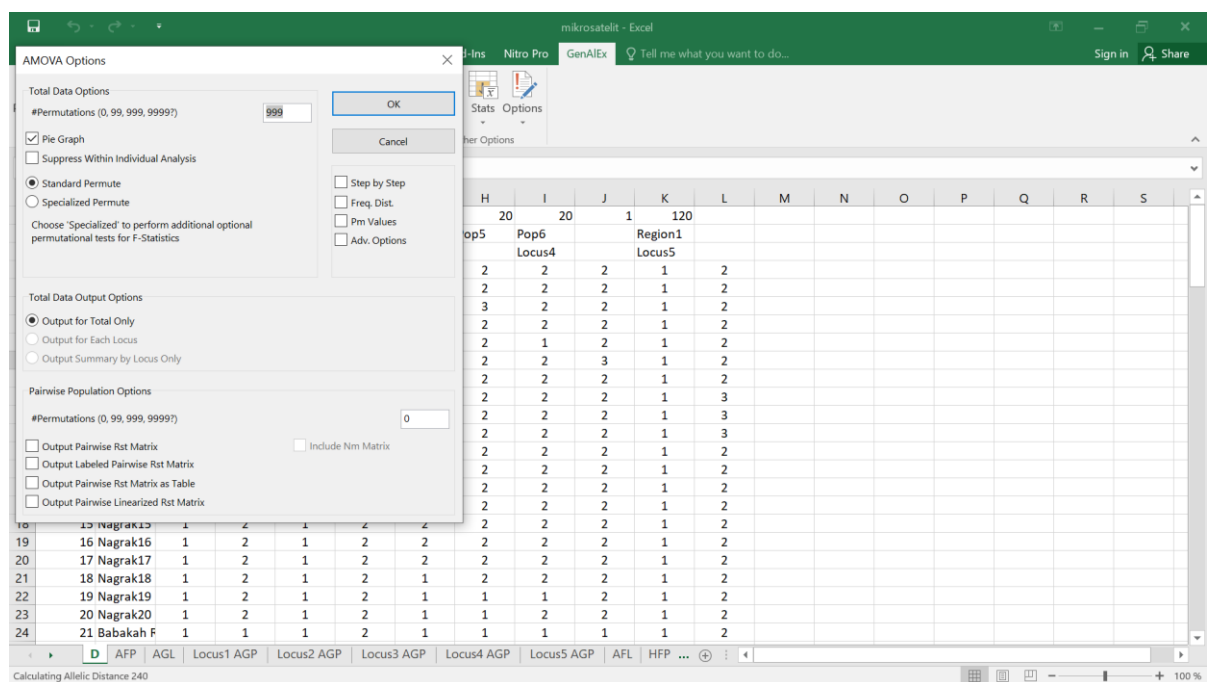
Population	%P
Pop1	100.00%
Pop2	80.00%
Pop3	100.00%
Pop4	100.00%
Pop5	80.00%
Pop6	100.00%
Mean	93.33%
SE	4.22%



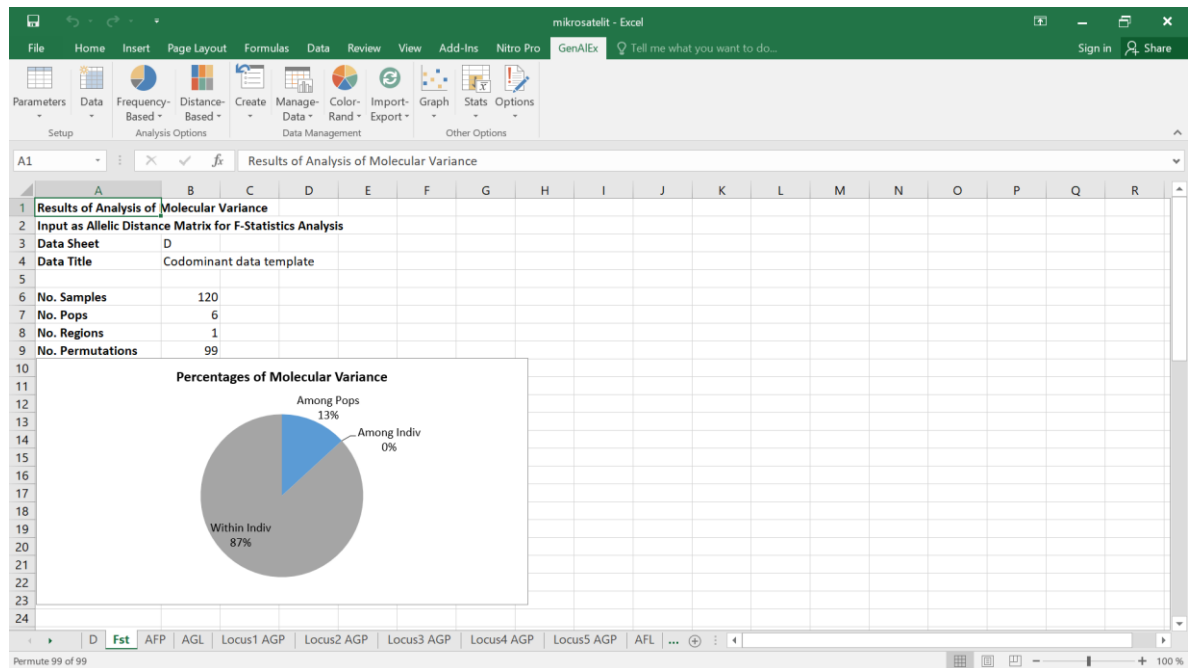
11. Selanjutnya akan muncul Tab menu lanjutan yang secara otomatis terisi *Codom-Allelic*, karena data yang digunakan merupakan data kodominan. Klik Ok.



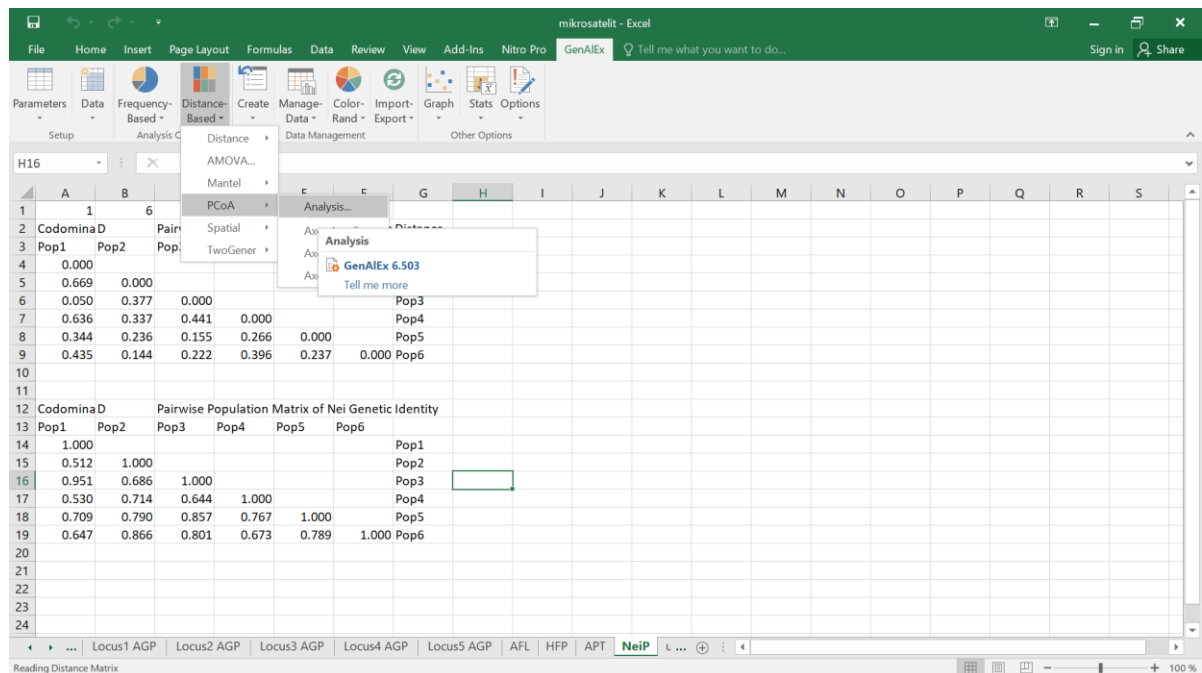
12. Kemudian akan muncul *Tab menu AMOVA Option*. Pilih option yang diinginkan.



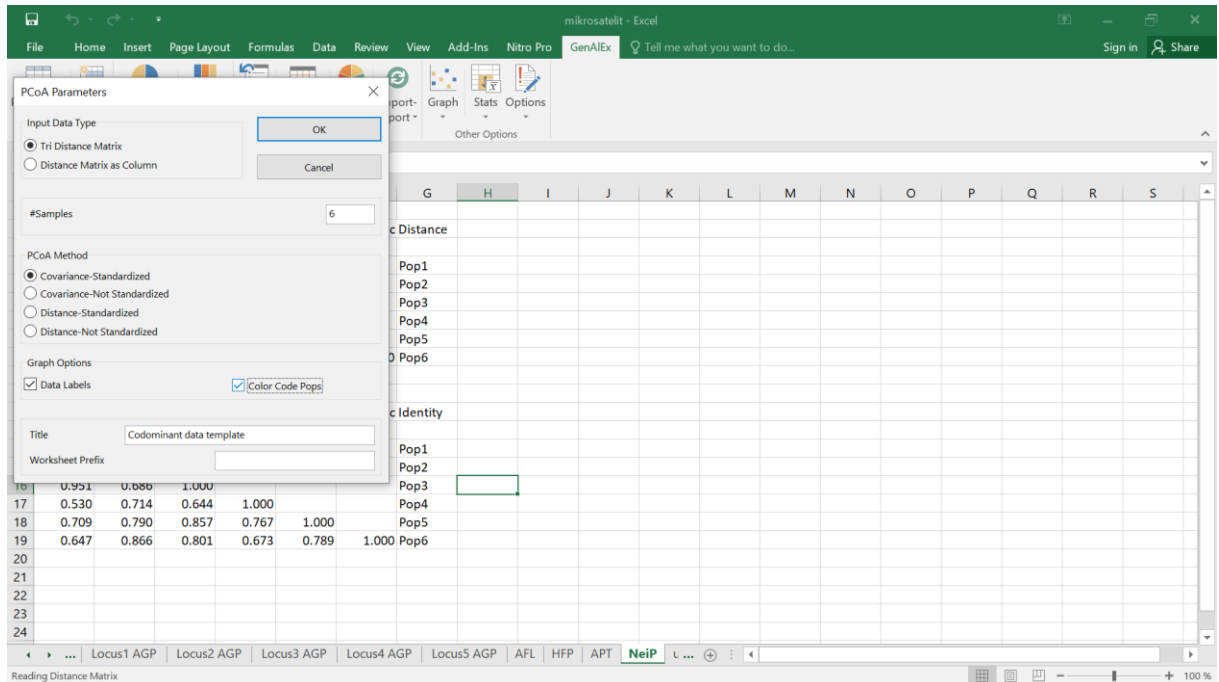
13. Terlampir hasil AMOVA. Terlihat dari *pie chart* bahwa keragaman antar individu adalah 87% sementara keragaman antar populasi adalah 13%.



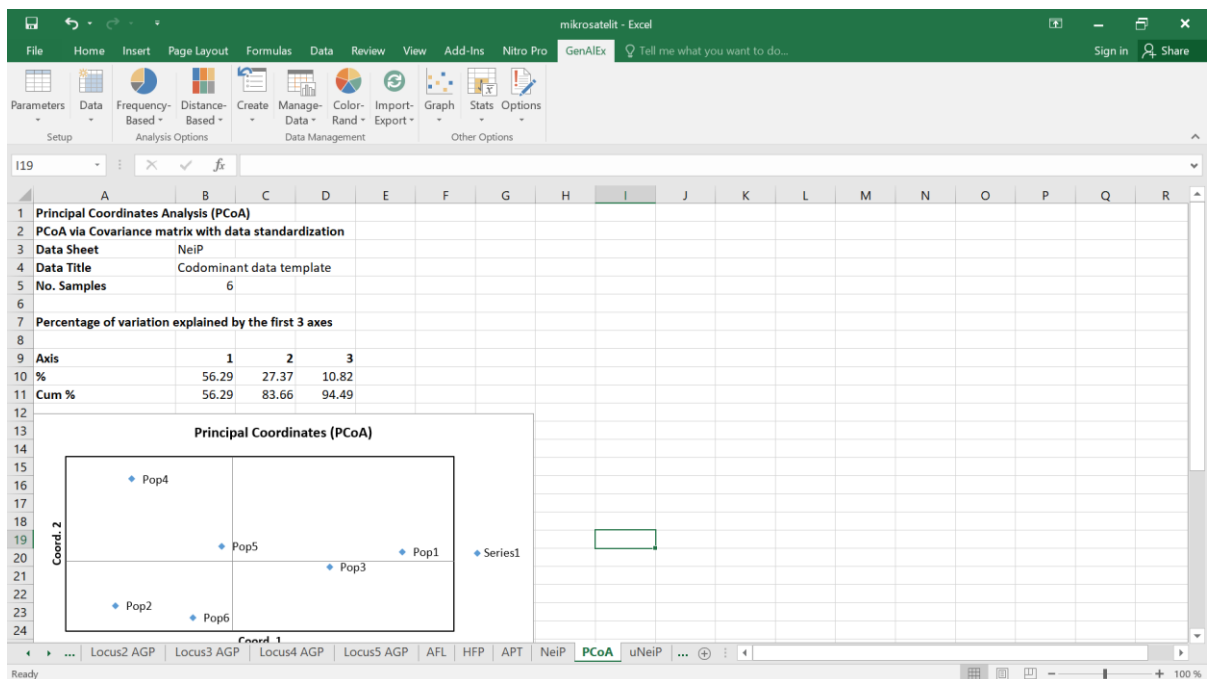
14. *Principal Coordinate Analysis (PCoA)* digunakan untuk memvisualisasikan *similarity/dissimilarity* data (jika diperlukan). Untuk menggunakan PCoA, aktifkan *sheet* jarak genetik Nei (dalam gambar tertera sebagai *sheet* NeiP, kemudian pilih *Distance Based*, lalu langsung klik PCoA, kemudian langsung klik *Analysis*.



15. Selanjutnya akan muncul tab *PCoA parameters*. Pilih *Tri Distance matrix* sebagai *Input Data Type* dan metode PCoA yang diinginkan. Lalu klik OK.



16. Terlampir hasil PcoA.



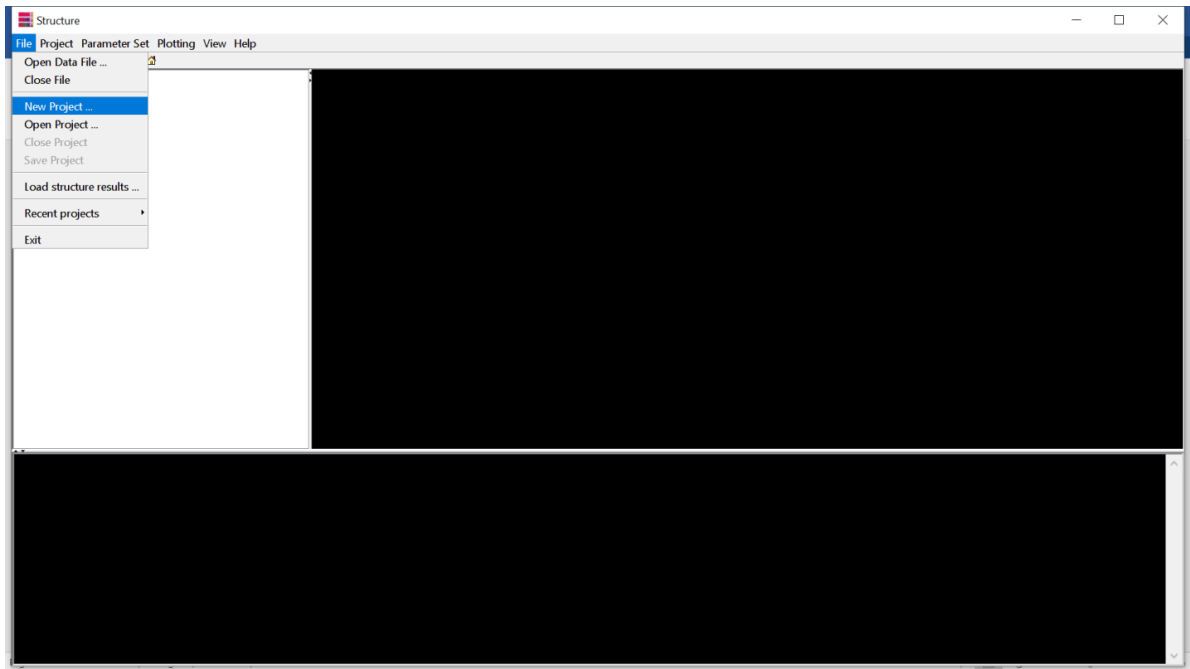
1.5 Analisis struktur genetik menggunakan *software structure* v. 2.3.4

Structure adalah *software* yang digunakan untuk melakukan klasterisasi individu ke dalam populasi berdasarkan informasi genetik dan menganalisis proporsi genom dari setiap individu dengan menggunakan metode klastering kuantitatif (Pritchard *et al.* 2000). *Structure* dapat diakses pada laman https://web.stanford.edu/group/pritchardlab/structure_software/release_versions/v2.3.4/html/structure.html. Data awal yang dapat digunakan dalam *structure* yaitu hasil skoring baik biner maupun alelik, dalam format text (.txt). Langkah-langkah analisis struktur populasi menggunakan *software structure* adalah sebagai berikut:

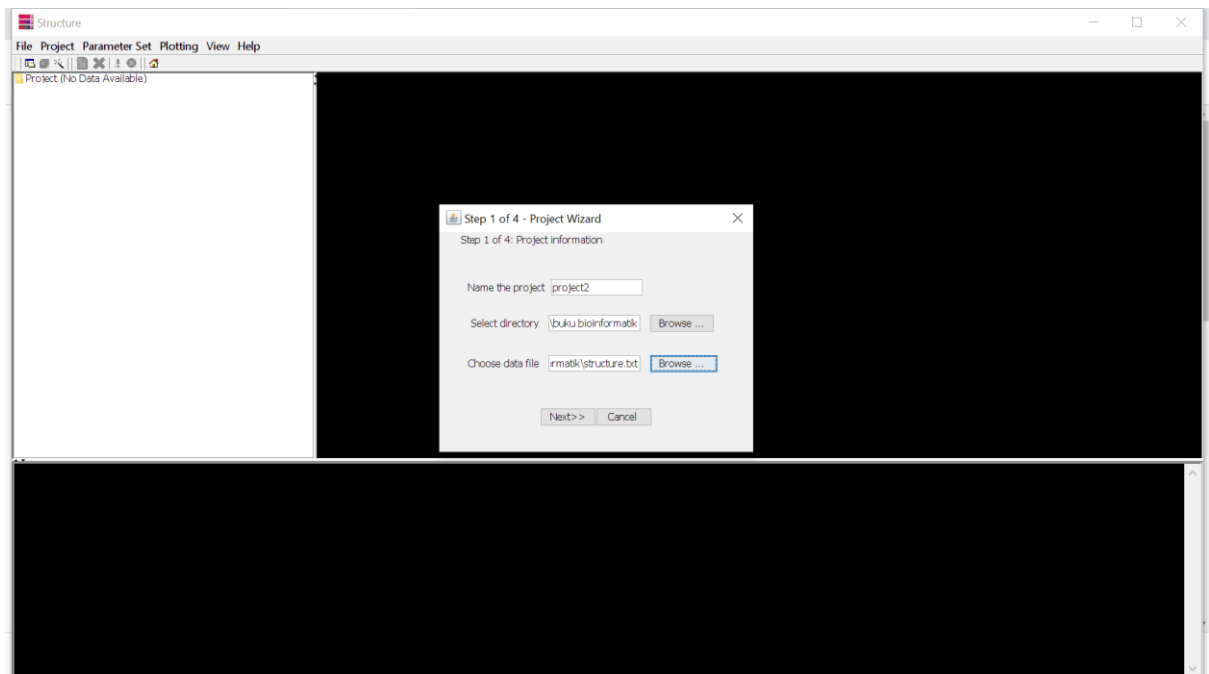
1. Format data yang digunakan adalah sebagai berikut: nama sampel tidak memiliki *header* kolom, setiap lokus diberikan nama lokus<nomor>. Data yang digunakan dalam contoh kali ini adalah sebanyak 64 individu dari 4 populasi dari 2 region yang dianalisis dengan marka dominan dan menghasilkan 155 lokus. Karena data berasal dari sampel diploid, maka setiap sampel dibuat 2 baris yang sama. Contohnya pada gambar di bawah ini terdapat sampel B01 R ada sebanyak 2 baris.

[illegible]

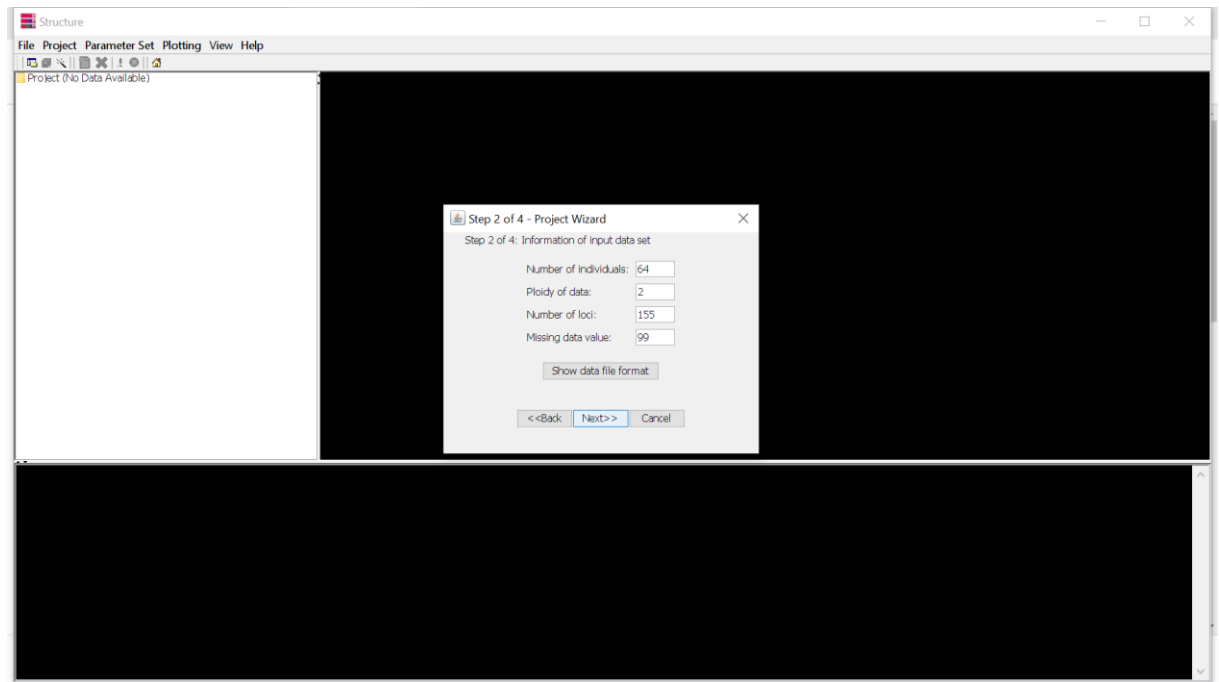
2. Untuk mulai menggunakan *Structure*, klik file, lalu langsung klik *New Project*.



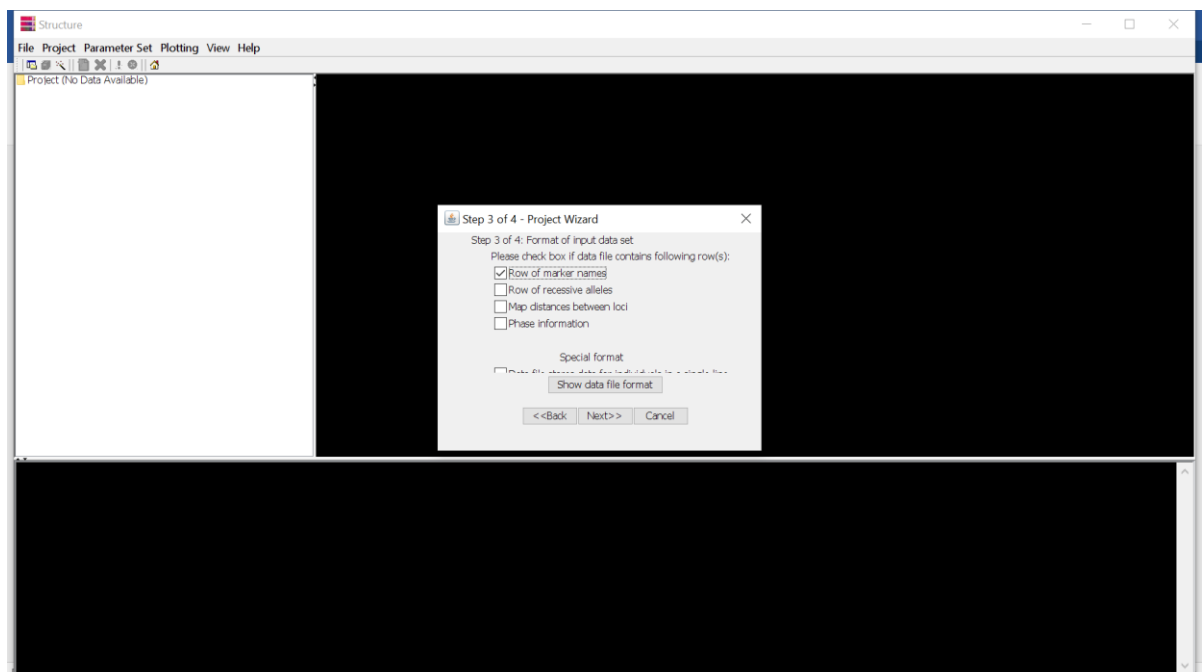
3. Ada 4 *steps* yang diperlu dilengkapi. Pada *Step* Isi nama *project* (dalam gambar tertera “*project2*” sebagai nama *project*). Setelah itu, pilih folder tempat menyimpan data dan pilih data yang akan digunakan, lalu klik *Next* untuk selanjutnya melengkapi informasi di *Step* 2.



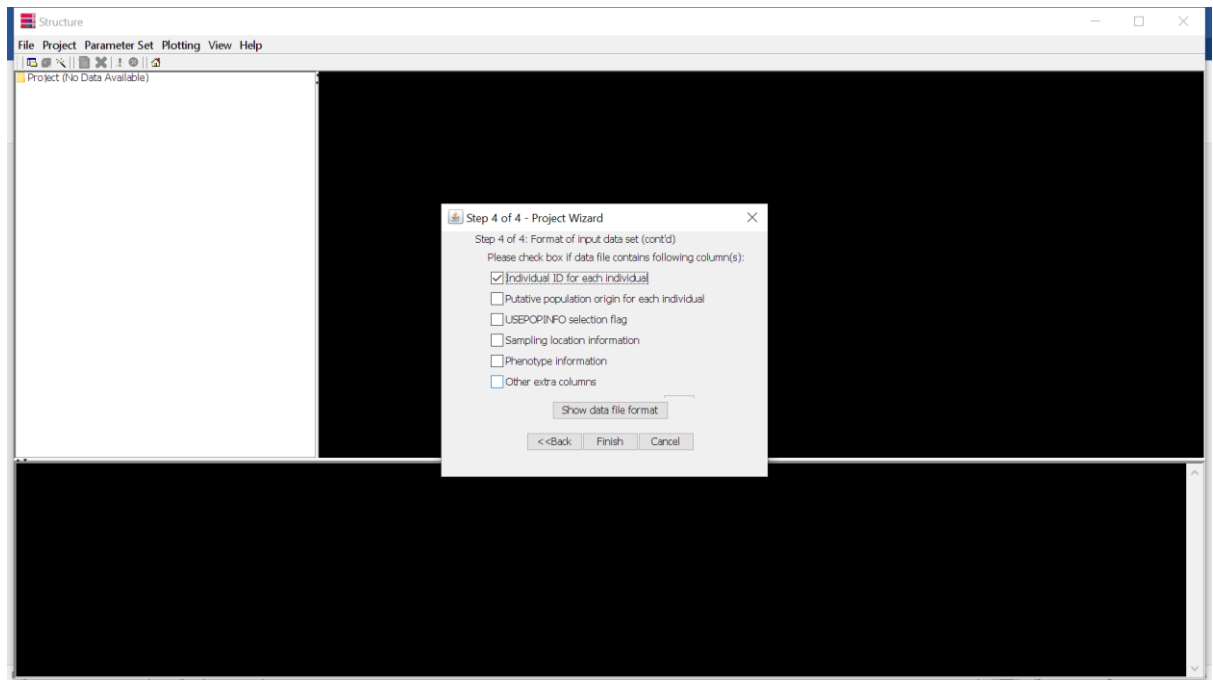
4. Pada *Step 2*, isi jumlah individu, *ploidy*, dan lokus. *Missing data* diisi sesuai dengan format data. Setelah itu klik *Next*, untuk selanjutnya melengkapi informasi di *Step 3*.



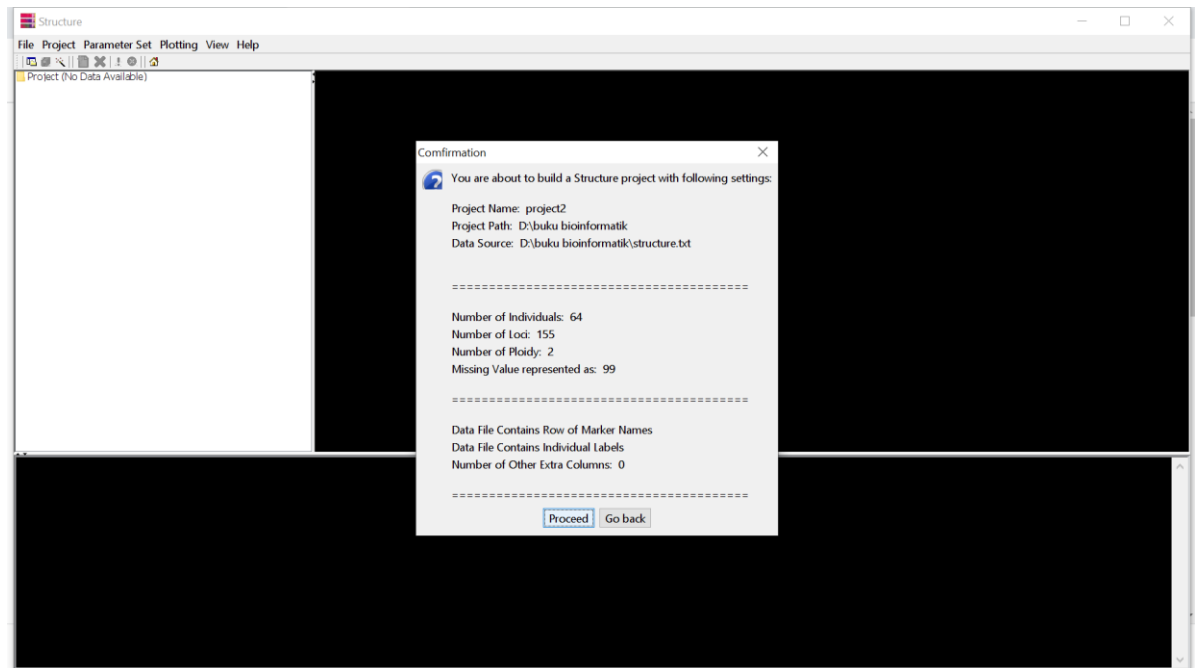
5. Pada *Step 3*, beri tanda ceklis pada pilihan *row of marker names*, lalu klik *Next* untuk selanjutnya melengkapi informasi di *Step 4*.



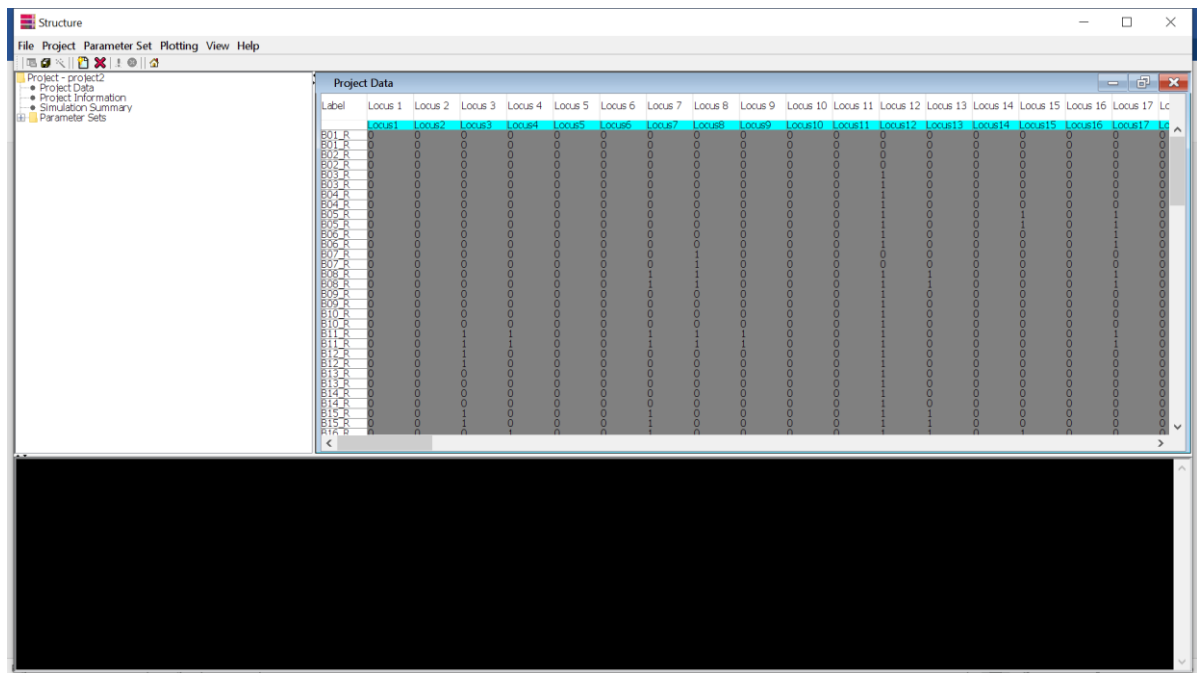
6. Pada *Step 4*, pilih *individual ID for each individual*, lalu klik Finish. Jika terdapat informasi *sampling* dan lain-lain dapat ditambahkan pada data.



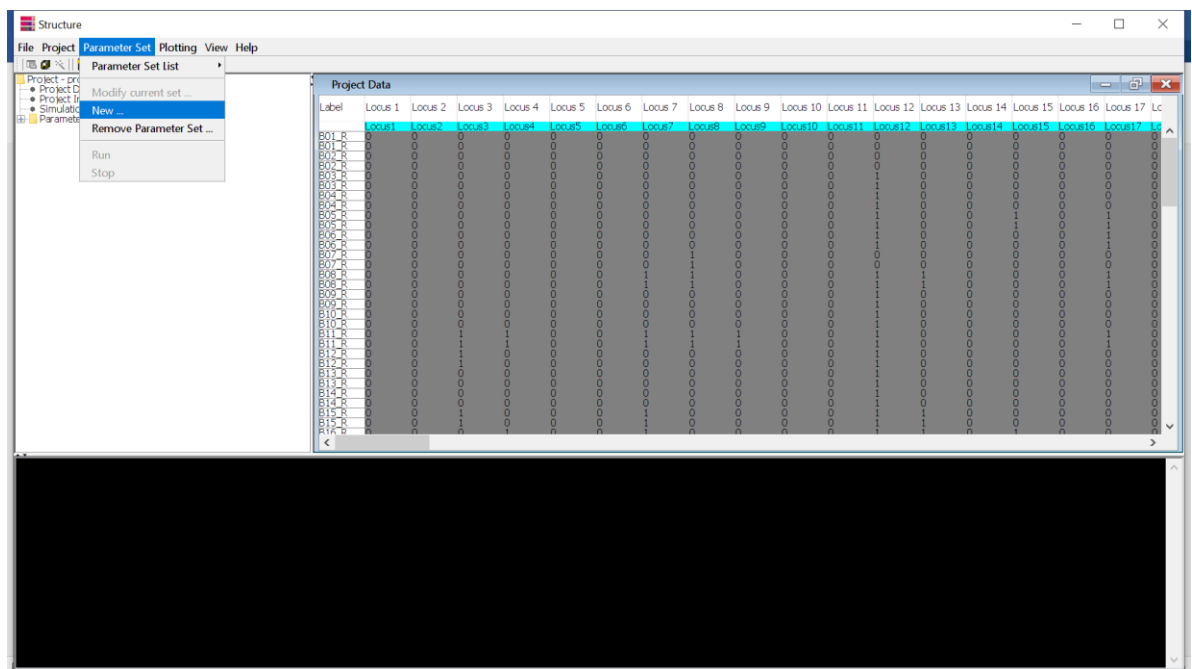
7. Setelah muncul *tab confirmation*, klik *proceed*.



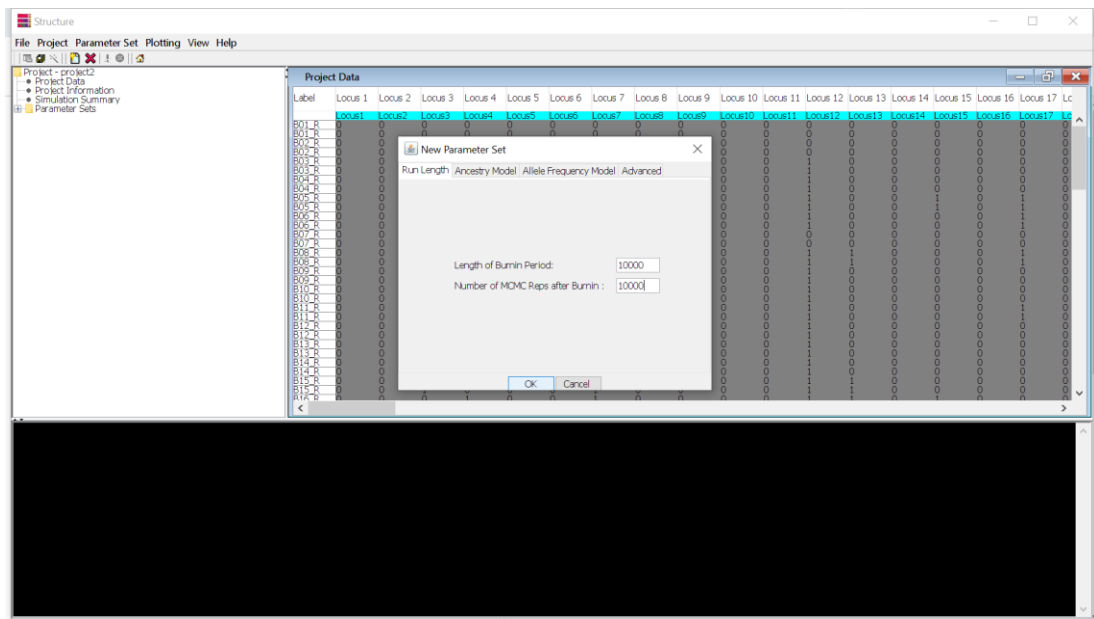
8. Terlampir tampilan setelah data terinput.



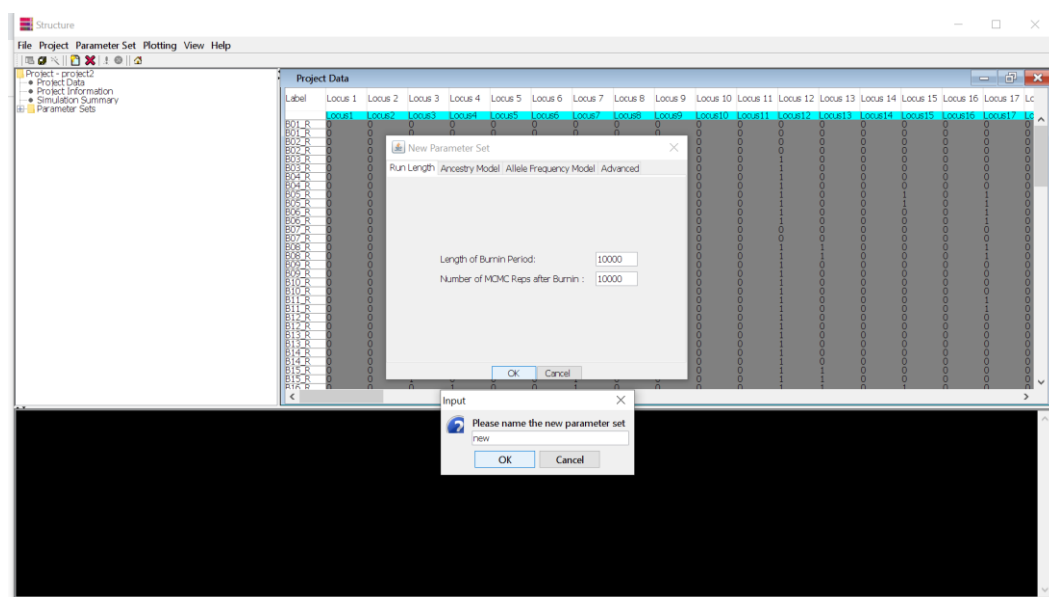
9. Pilih tab *Parameter set*, lalu klik *New*.



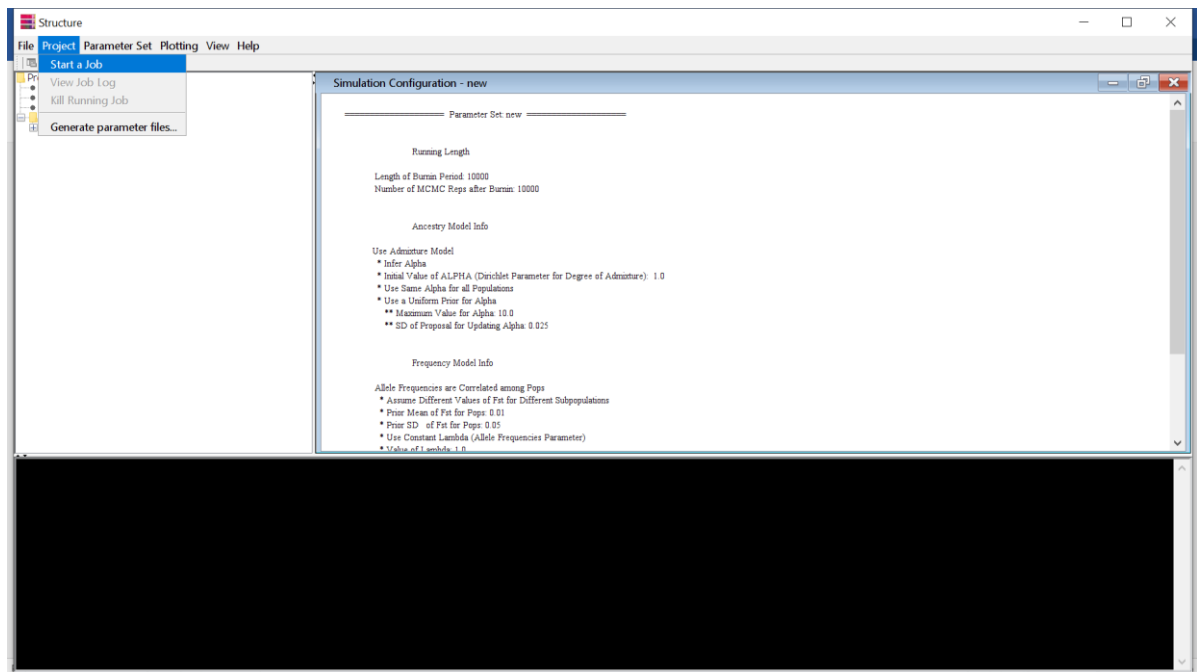
10. Setelah itu akan muncul tampilan *burnin period*. *Burnin period* menentukan berapa lama simulasi akan dijalankan sebelum mendapatkan data untuk mencegah efek *starting configuration*. MCMC adalah *Markov Chain Monte Carlo* yang digunakan untuk mengassign individu secara random pada grup tertentu. *Burnin period* dan MCMC yang biasa digunakan berkisar antara 10.000-100.000. Setelah mengisi *Length of Burnin Period* dan *Number of MCMC* klik OK.



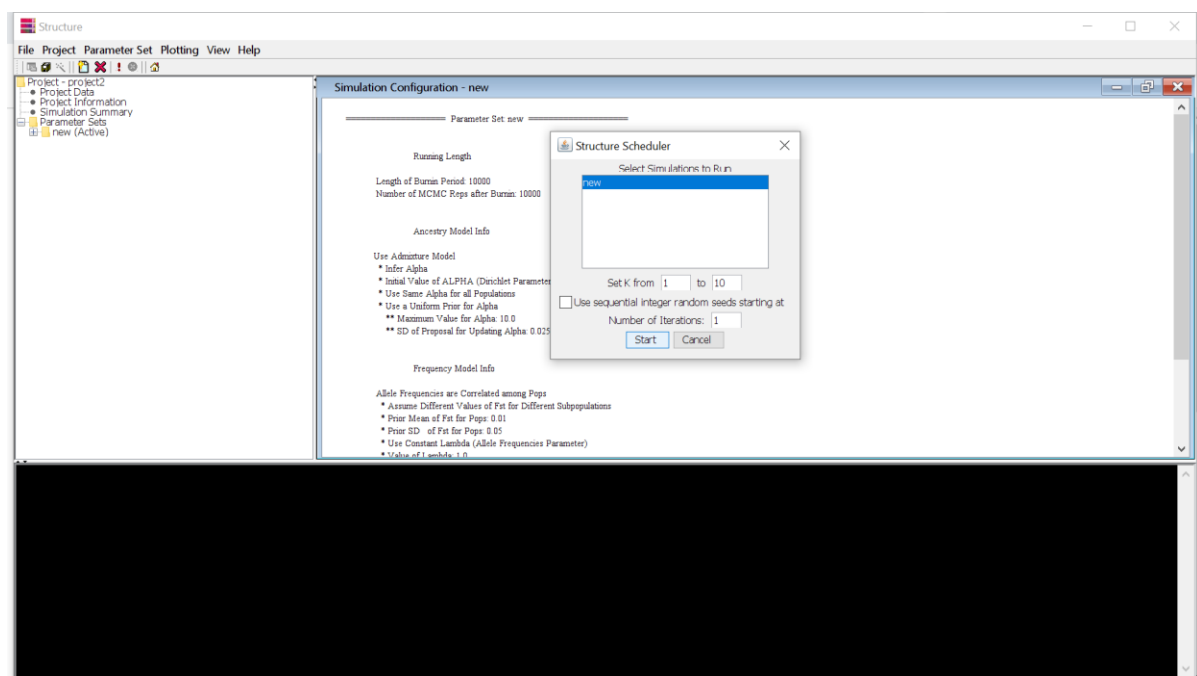
11. Setelah klik OK, akan muncul *tab input* untuk memberi nama *Parameter Set*. Dalam gambar terlihat *tab Input* dan permintaan “Please name the new parameter set adalah “new”. Setelah memberi nama *Parameter Set*, lalu klik OK. (Dalam gambar tertera nama *parameter set* adalah “new”).



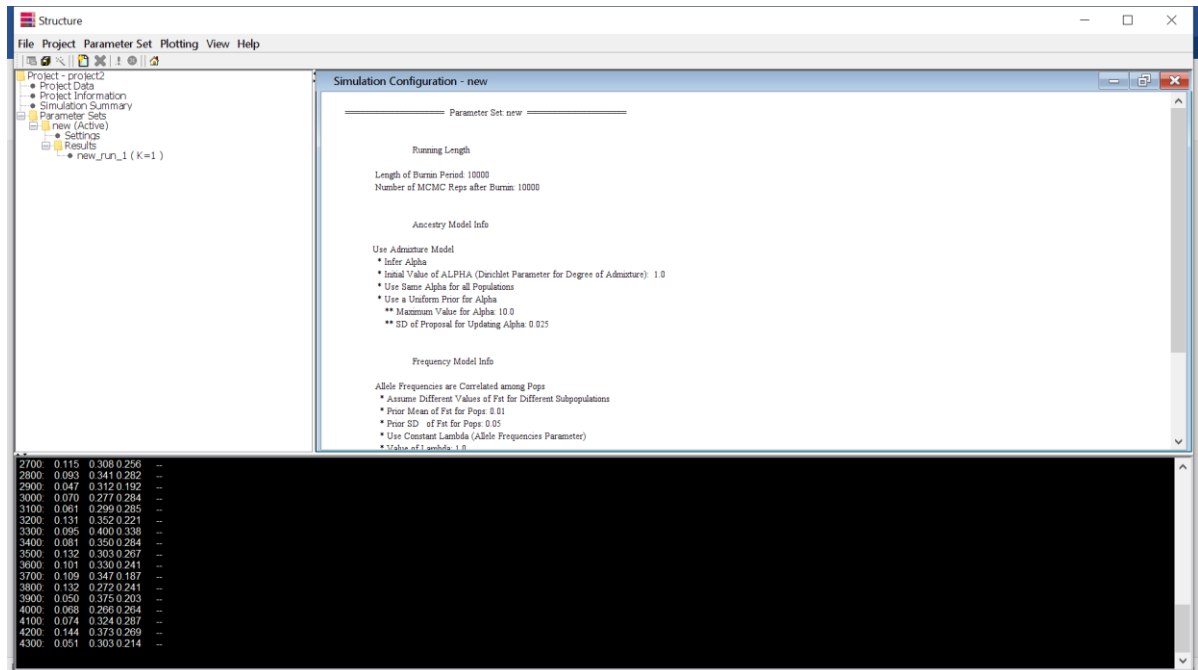
12. Setelah itu, kembali ke menu utama, dan pilih *tab Project*, lalu langsung klik *Start a job*.



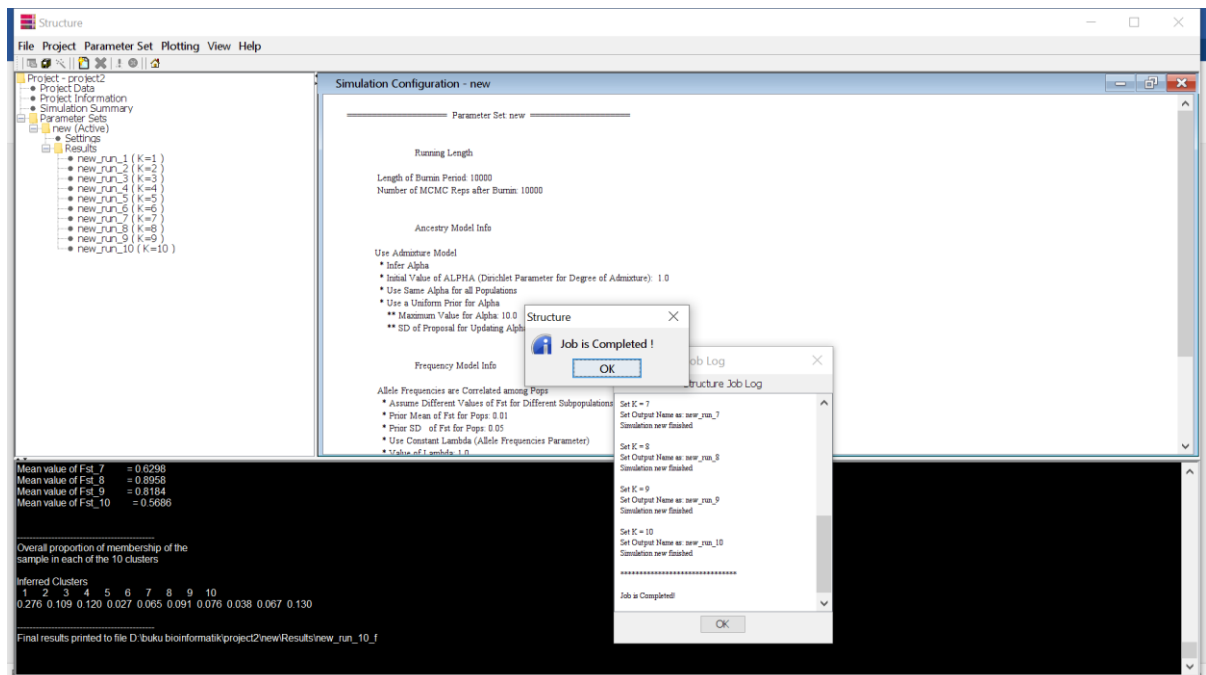
13. Akan muncul *tab Structure Schedule*. Pada bab ini, pilih *parameter set* (dalam gambar tertera “new” sebagai *parameter set*, yaitu yang telah dibuat pada poin 11. Setelah itu isilah jumlah K (Dalam gambar tertera “Set K from 1 to 10”). K adalah asumsi jumlah kelompok yang akan terbentuk dari data.



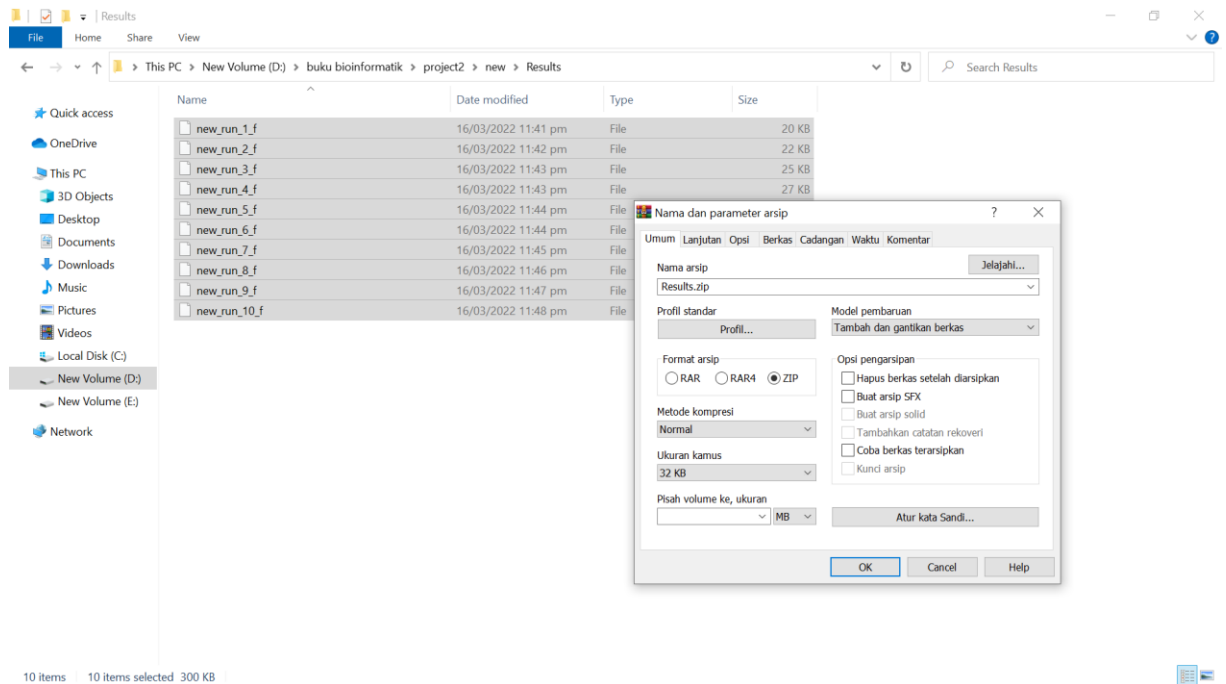
14. Dimulainya analisis ditandai dengan munculnya angka-angka di *window* bagian bawah.



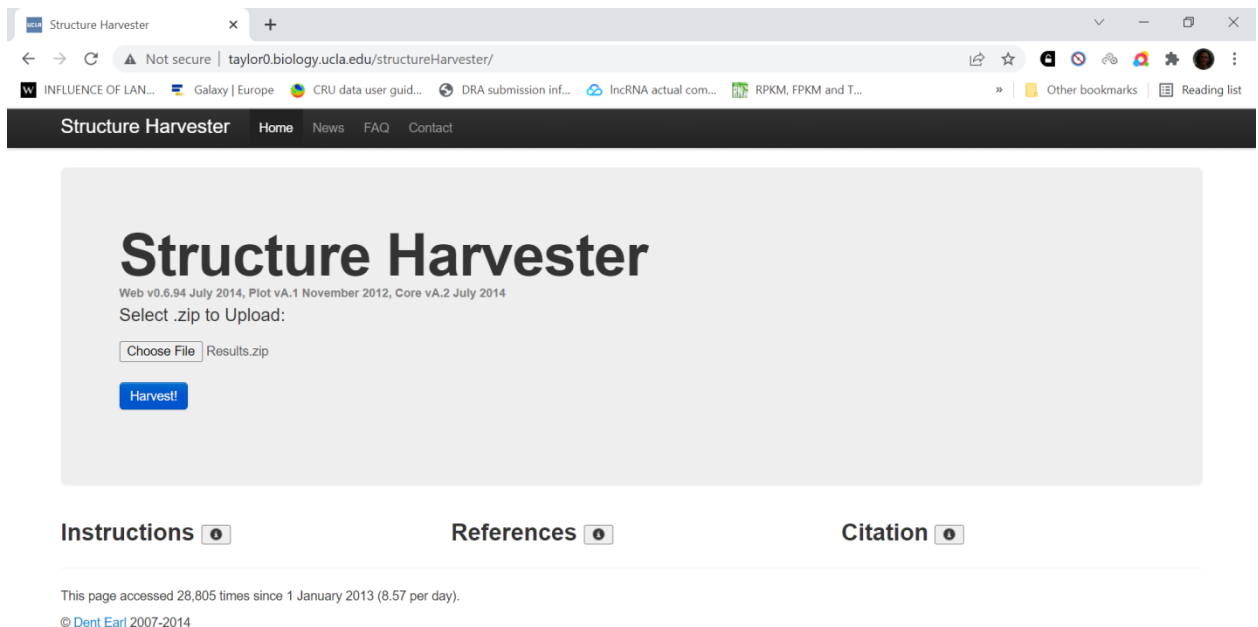
15. Jika sudah selesai akan muncul notifikasi bahwa “*Job is Completed*”. Hasil analisis terlihat ada di *File Explorer* di dalam *subfolder Results*.



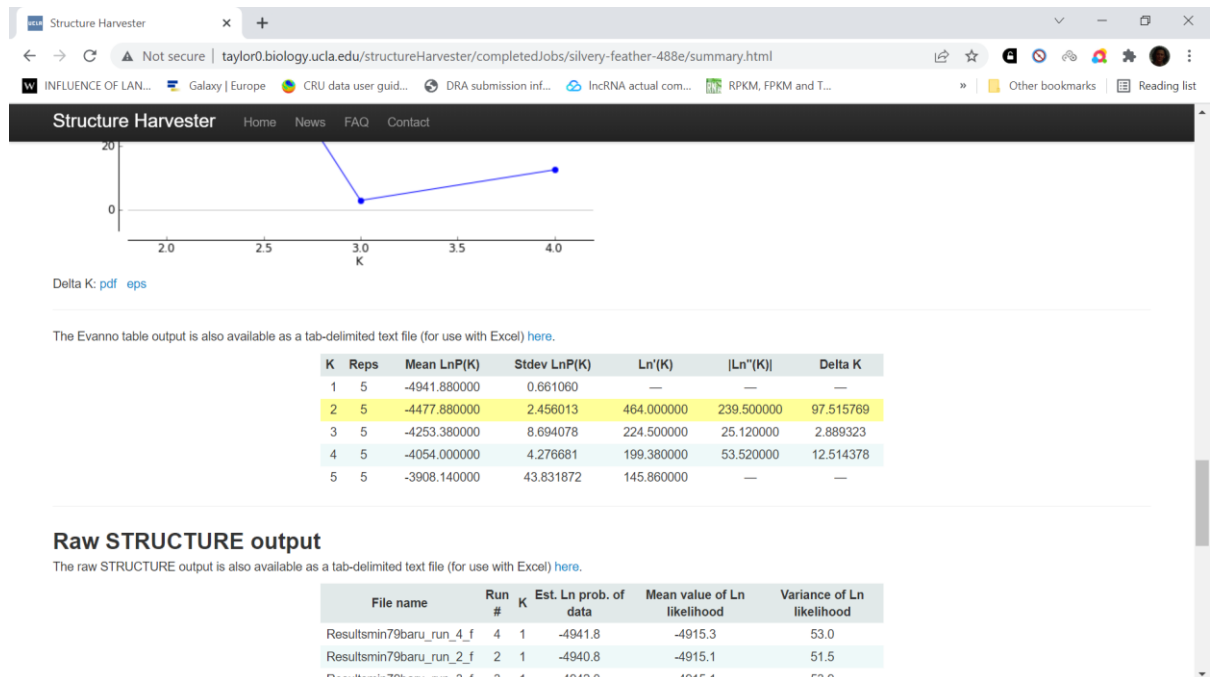
16. Selanjutnya buka *File Explorer* kemudian masuk ke *Subfolder Results*. Pilih semua hasil analisis kemudian kompres ke dalam format .zip, lalu klik OK.



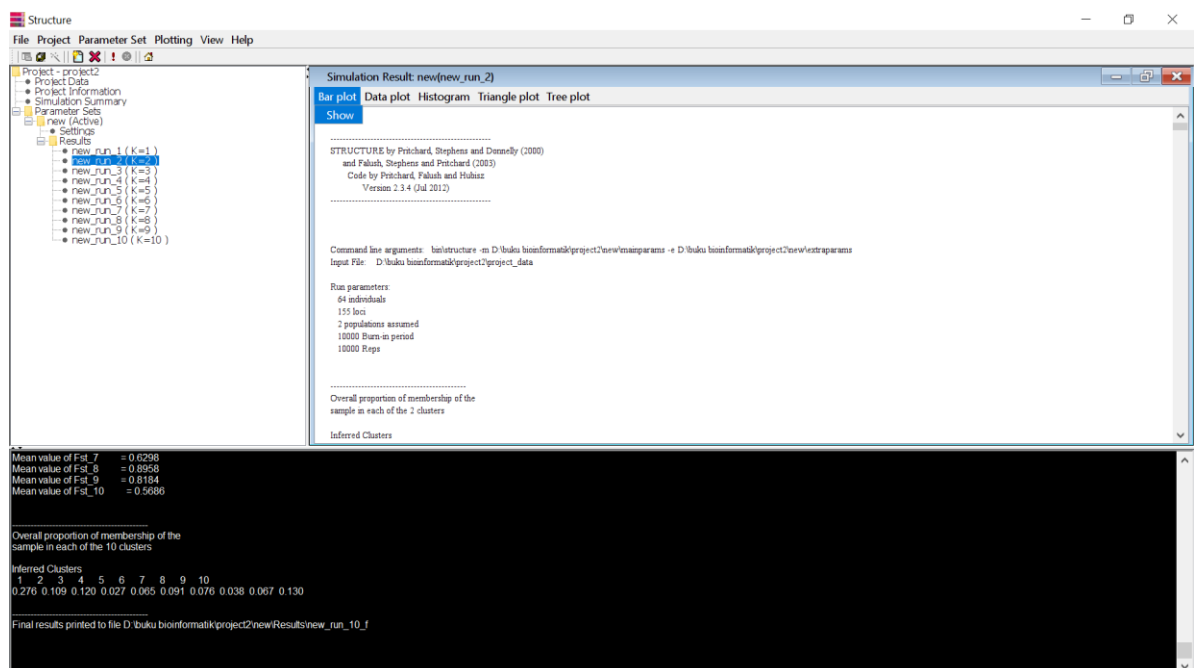
17. Kemudian masuk ke laman *Structure Harvester* untuk mendapatkan nilai K yang sebenarnya (<http://taylor0.biology.ucla.edu/structureHarvester/>) (Earld *et al.* 2012). Pilih file hasil analisis yang telah dikompres ke dalam bentuk .zip, kemudian klik *Harvest*.



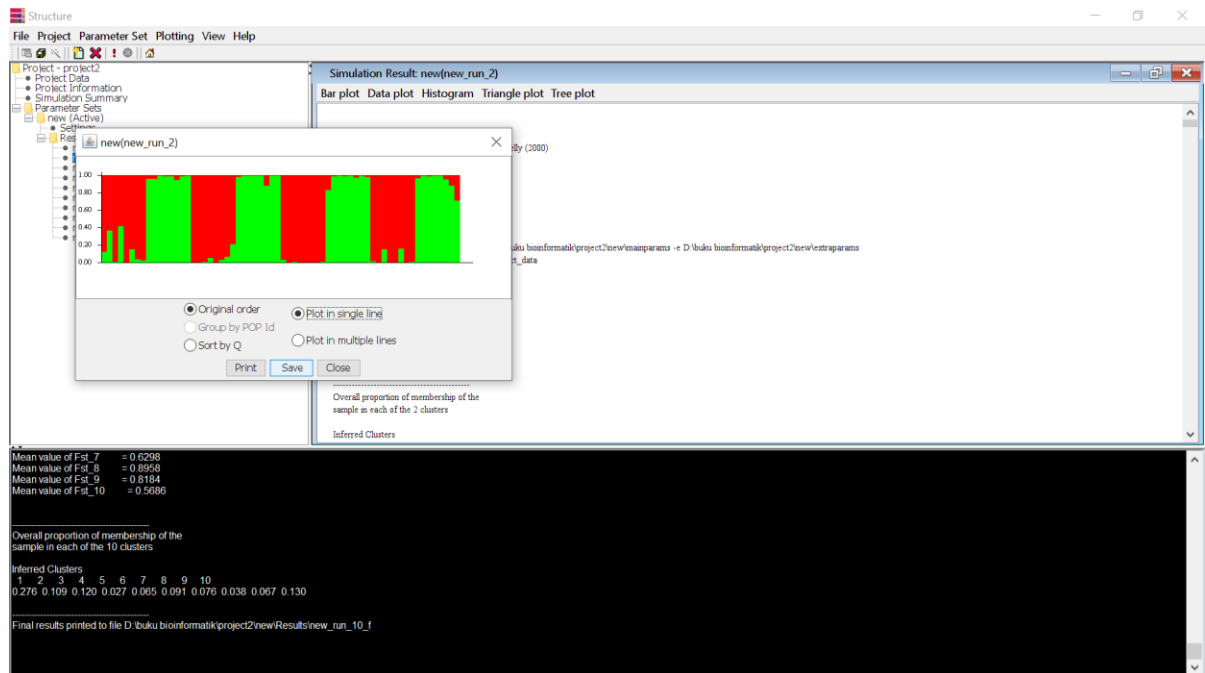
18. Nilai K yang sebenarnya ditandai dengan *highlight* berwarna kuning pada tabel yang muncul. Dalam contoh kali ini, terlihat bahwa nilai K yang diberi *highlight* berwarna kuning oleh *Harvest* adalah 2, berarti nilai K yang sebenarnya adalah 2.



19. Setelah mencatat nilai K yang sebenarnya (yaitu 2), maka kembali ke *software* Structure. Pada menu utama Structure, klik File dan pilih *subfolder* Results. Kemudian klik file hasil analisis dengan nilai K=2. Setelah itu klik *Bar plot*, lalu langsung klik *Show*.



20. Muncul hasil sebaran genetik pada tiap individu dari populasi. Urutan sampel dimulai dari sebelah kiri. Terlihat bahwa 4 populasi memiliki 2 tipe genetik yang berbeda yang ditunjukkan oleh warna merah dan hijau muda.



Catatan:

Catatan:

BAB II

PENGANTAR BIOINFORMATIKA

Bioinformatika meliputi kajian yang memadukan disiplin ilmu biologi molekuler dan teknik informasi. Bioinformatika merupakan aplikasi dari alat komputasi untuk melakukan manajemen data termasuk: 1) pengumpulan, visualisasi, integrasi, dan analisis data; 2) *modeling* dan prediksi berdasarkan data yang tersedia; 3) penyimpanan, penyebaran dan aplikasi informasi biologis serta 4) interpretasi data biologi molekuler hasil kegiatan di laboratorium basah atau kering.

Bioinformatika pertama kali muncul atas ide para ahli biologi molekuler dan ahli statistika, berdasarkan pemikiran bahwa semua gejala yang ada di alam ini dapat dibuat secara *artificial* melalui simulasi dari berbagai data yang ada. Bioinformatika membuka jalan untuk mempelajari informasi biologi dengan menggunakan konsep ilmu komputer, statistika, dan *engineering*. Bioinformatika dikategorikan sebagai manajemen informasi dan biologi. Di dalam Bioinformatika, data-data genetika merupakan inti pembentukan simulasi yang berperan sangat penting, dalam hal manajemen data biologi molekuler, terutama sekuens DNA dan informasi genetika. Bioinformatika mengakomodasi keperluan akan pusat data (*database*) dan bank data yang diperoleh dari penelitian biologi molekuler. Peran bioinformatika antara lain:

- Secara teknologi, bioinformatika merupakan teknologi yang ‘*powerful*’ untuk mengelola, menyimpan, menganalisis dan melakukan interpretasi dari data biologi molekuler.
- Sebagai metodologi, bioinformatika merupakan pendekatan yang menyeluruh, luas, mulai dari aspek hulu hingga ke hilir yang menghasilkan hipotesis baru, pola baru dan penemuan elemen fungsional baru.

Bioteknologi modern ditandai dengan kemampuan manusia untuk memanipulasi kode genetik DNA. Bioinformatika menjadi tren yang lahir dan berkembangnya dipacu oleh perkembangan biologi molekuler. Pada tahun 1953 Watson-Crick berhasil merumuskan struktur DNA yang secara teoritis menggunakan data penelitian Rosalind Franklin. Perkembangan penting selanjutnya adalah teknologi rekombinan DNA yang berhasil ditemukan pada tahun 1970, diikuti oleh teknologi PCR (*Polymerase Chain Reaction*) pada 1980an.

Kemajuan perkembangan bioinformatika dalam bidang biologi molekuler dan bioteknologi dimulai dari *genome project* yang dimulai sejak Oktober 1990 yang dilaksanakan di seluruh dunia dan menghasilkan banyak informasi gen dari berbagai makhluk hidup, mulai dari makhluk hidup tingkat rendah sampai makhluk hidup tingkat tinggi. Hasil pembacaan dari *genome project* ini adalah *database* genom. Keberadaan *database* adalah syarat utama dalam analisis bioinformatika. *Database* yang dihasilkan dari *genome project* merupakan susunan data yang rapi, sehingga dapat digunakan

untuk berbagai keperluan, termasuk keperluan penelitian maupun keperluan pengambilan keputusan. Bioinformatika juga mempercepat penyelesaian *genome project* ini karena bioinformatika mensuplai berbagai program yang diperlukan untuk proses pembacaan genom. Dengan bioinformatika, berbagai data tersebut dapat disimpan teratur dengan jangka waktu proses yang singkat dan tingkat akurasi yang tinggi serta sekaligus dianalisis dengan berbagai program yang dibuat untuk tujuan tertentu.

Pada tahun 2001, pembacaan sekuens genom manusia yang dilakukan oleh perusahaan bioteknologi *Celera Genomics* (Amerika Serikat) menjadi lebih cepat dan lebih akurat dibandingkan dengan usaha konsorsium lembaga riset publik di Amerika Serikat, Eropa, dan negara maju lainnya, akibat kontribusi teknologi bioinformatika. Lahirnya bioinformatika digunakan untuk mengorganisasi dan menganalisis data biologi molekuler menjadi sebuah informasi biologis yang bermakna dan mudah dipahami. Ketersediaan *database* dasar (DNA, protein) yang bersifat gratis merupakan peluang besar untuk menggali informasi kandidat-kandidat gen yang potensial. Teknologi *chip* DNA semakin mempercepat ledakan informasi dari kemajuan bioteknologi seperti data sekuens DNA dari pembacaan genom, data sekuens dan struktur protein, hingga data transkripsi RNA.

Cakupan bioinformatika dalam bidang biologi molekuler dan bioteknologi terkait dengan DNA/Genom (kode kehidupan) dan Omics (turunannya) yaitu:

- a. Analisis sekuens (*sequence analysis*)
 - *Genome sequencing*
 - *Gene finding* dan *genome annotation*
 - *Sequence comparison*
- b. Analisis transkriptomik (*transcriptome analysis*)
 - *Microarray analysis*
 - *Tiling arrays*
 - *Regulatory sequence analysis*
- c. *Computational proteomics*
- d. *Metabolomics*
- e. *Ontologies*
- f. *Databases*

Perkembangan teknologi di dunia yang begitu pesat mendorong kinerja penelitian menjadi lebih cepat. Kecepatan dan ketepatan dalam penelitian biologi molekuler dapat didukung dengan keberadaan sistem basis data bioinformatika secara daring. Keberadaan beberapa basis data sekuens yang ada (*existing*) belum sepenuhnya dimanfaatkan oleh para peneliti untuk mendukung penelitian mereka di tingkat pendahuluan atau pemula di dalam berbagai bidang kehidupan. Namun demikian,

tren penelitian berbasis bioinformatika diperkirakan akan semakin meningkat di Indonesia, karena penelitian berbasis bioinformatika tergolong “murah” dan mampu menghasilkan data prediktif yang mendekati kondisi nyata di laboratorium. Meskipun demikian, peneliti perlu memiliki keahlian khusus untuk melakukan validasi sekuens untuk memberikan makna biologis dari sekuens tersebut. Penggunaan bioinformatika sangat intensif di berbagai bidang seperti bidang kesehatan, kedokteran molekuler, terapi gen, farmasi dan pengembangan obat-obatan, aplikasi genom mikrobiologi, pemuliaan tanaman, analisis forensik, bioteknologi, studi evolusi, pertanian, lingkungan serta studi perubahan iklim. Bioinformatika juga dapat digunakan untuk membantu penemuan gen/protein hasil modifikasi yang dapat dipatenkan. Berikut ini beberapa contoh aplikasi bioinformatika di beberapa bidang:

1. Bidang forensik:

Perkembangan aplikasi sidik jari DNA memudahkan identifikasi korban kecelakaan pesawat terbang atau pengeboman dengan sisa jaringan jenazah, meskipun hanya tersisa sedikit.

2. Bidang kesehatan:

Banyak dilakukan penelitian bioinformatika untuk mengatasi penyakit pada manusia, diantaranya pemanfaatan informasi genom untuk mempelajari interaksi berbagai protein dalam rangka memahami basis molekuler dari sebuah penyakit.

3. Bidang klinis:

Bioinformatika digunakan untuk mencari informasi klinis dan melakukan identifikasi mutasi gen-gen penyebab penyakit dalam rangka menentukan terapi gen dan pengobatan individual sesuai profil genetik setiap pasien. Bioinformatika juga berperan dalam identifikasi agen penyakit baru, diagnosis penyakit baru dan dalam penemuan obat. Penemuan yang terbaru adalah pemanfaatan tes PCR untuk mendiagnosis COVID-19.

4. Bidang pertanian:

Bioinformatika digunakan untuk mendukung eksperimen laboratorium pada bidang pertanian dengan hadirnya teknologi *Next Generation Sequencing* (NGS). Penelitian berbasis bioinformatika telah digunakan dalam pengembangan penanda molekuler, desain primer untuk analisis ekspresi gen diferensial, pengembangan peta genetik dan analisis ekspresi gen. Pemanfaatan Bioinformatika dalam ilmu terapan di bidang pertanian juga mengarah ke desain produk aplikatif untuk pengendalian hama dan penyakit serta perlindungan varietas tanaman. Di dalam bidang pertanian, aplikasi bioinformatika sangat umum digunakan dalam anotasi sekuens genomik dan proteomik dalam jumlah besar.

Saat ini, aplikasi bioinformatika semakin maju dan luas, antara lain dalam hal:

- a. Analisis molekuler, biodiversitas dan filogenetik.
- b. Pengembangan varietas tanaman baru.
- c. Analisis pensejajaran sekuens DNA dan RNA (*sequence alignment*).
- d. Pendugaan ekspresi gen.
- e. Pendugaan *conserved region*.
- f. Pemahaman dan pembelajaran evolusi gen hingga organisme.
- g. Pendugaan interaksi antar protein dan fungsi protein.
- h. Pengelolaan informasi biologis.
- i. Prediksi dan desain struktur sekunder RNA.
- j. Prediksi dan desain struktur protein (2D dan 3D).
- k. Identifikasi dan karakterisasi protein.
- l. Analisis *Fingerprint*.
- m. Analisis proteomik dan genomik.

Perangkat utama Bioinformatika adalah *software* yang didukung oleh ketersediaan internet dan *server word wide web* (www) untuk menganalisis data biologi dan menyimpan sejumlah data biologi yang dihasilkan oleh penelitian genom. Kebutuhan akan bioinformatika meningkat seiring dengan meningkatnya *database* yang tersedia. Bioinformatika menyediakan metode yang efisien untuk: 1) Penyimpanan data; 2) Anotasi; 3) Pencarian dan pengambilan data; 4) Integrasi data; 5) Analisis data.

Pengolahan data untuk mempelajari, memahami dan memanipulasi sandi/kode genetik DNA yang didukung oleh perkembangan teknologi bioinformatika telah mempercepat perkembangan dan kemajuan ilmu biologi. Beberapa *software* yang dibuat untuk mendukung perkembangan bioinformatika antara lain:

1. PRIMER 3.
2. MEGA.
3. GENEIOUS.
4. WEBSNAPER.
5. MAPMAKER.
6. UGENE.
7. GENALEX.
8. DARWIN.
9. MAPCHART.

10. CERVUS.
11. NTSYS.
12. Krait.
13. Omicsbox.
14. R statistics
15. TASSEL.
16. Trinity
17. TopHat
18. GeneScan
19. Geneious
20. Genetyx

Aplikasi bioinformatika menjadi semakin penting di masa depan, tidak hanya dalam hal mengakselerasi kemajuan bioteknologi, namun juga menjembatani dua gelombang ekonomi baru yaitu, teknik informasi dan bioteknologi. Bioinformatika menjelma menjadi bisnis besar yang bernilai ekonomi tinggi. Teknologi informasi saat ini telah menjadi mesin penggerak ekonomi sekaligus tren gaya hidup manusia modern. Bioinformatika juga diyakini telah menjadi lokomotif penggerak ekonomi masa depan, sehingga siapa saja yang menguasai bioteknologi dan teknik informasi (bioinformatika), maka dia akan menguasai ekonomi dunia.

Catatan:

BAB III

PENGENALAN BIOINFORMATIKA

Ilmu bioinformatika pertama kali hadir pada tahun 1960 yang diawali dengan pembuatan pangkalan data dan pengembangan algoritma untuk analisis sekuens biologi. Amerika Serikat merupakan pihak yang sukses merintis pangkalan data protein. Pada tahun 1970, pangkalan data protein mulai dikerjakan secara bersama-sama oleh konsorsium Amerika Serikat dan Jerman di laboratorium Biologi Molekuler Eropa (*European Molecular Biology Laboratory*). Akhirnya pada tahun 1970 dan 1990-an terjadi ledakan jumlah sekuens DNA organisme. Pada tahun 1980-an, bioinformatika mengacu pada penerapan ilmu komputer dengan perintisnya adalah Paulien Hogeweg.

Penemuan teknologi lainnya, yaitu jaringan internet, mendukung perkembangan ilmu bioinformatika. Dengan penemuan jaringan internet, para peneliti mampu menghubungkan pangkalan data bioinformatika, sehingga memudahkan peneliti lain dalam mengumpulkan hasil sekuensing ke dalam pangkalan data tersebut serta memperoleh sekuens biologi sebagai bahan analisis. Selain itu, jaringan internet mempermudah para ilmuwan dalam mengakses program, aplikasi, dan data bioinformatika. Kemajuan ilmu bioinformatika ini lebih didorong lagi oleh *genome project* yang dilaksanakan di seluruh dunia dan menghasilkan banyak informasi gen dari makhluk hidup tingkat rendah hingga makhluk hidup tingkat tinggi. Pengorganisasian data memastikan informasi biologis tersedia melalui pengawasan kualitas, pemeriksaan silang, dan standarisasi untuk selanjutnya dapat diakses atau ditelusuri dengan cepat dan mudah.

Seiring dengan berjalannya waktu, perkembangan bioteknologi semakin pesat. Hal tersebut berdampak pada melimpahnya jumlah data terkait spesimen biologis yang ada. Selain itu, saat ini kehidupan manusia memasuki era *data-openness*, atau era keterbukaan informasi dalam rangka mempercepat penelitian untuk meningkatkan kesejahteraan hidup manusia dan konservasi biodiversitas. Berbagai hal tersebut menyebabkan lembaga-lembaga di dunia membentuk beberapa *database* atau basis data yang berfungsi sebagai tempat penyimpanan berbagai data biologis, di antaranya sekuens DNA, RNA, genom, protein, taksonomi, dan sebagainya. Pertumbuhan pesat *database* observasi biologi, dibuktikan dengan adanya *database GenBank*. *GenBank* merupakan pusat *database* yang menyimpan semua data biologi molekuler dari seluruh penjuru dunia.

Database bioinformatika terbagi menjadi 3 jenis, yaitu:

1. *Database* primer, merupakan arsip data yang memuat seluruh data yang diunggah dan dapat diakses secara luas. Data yang terunggah memiliki kode unik untuk identifikasi. Contoh *database* primer: GenBank dan DDBJ.

2. *Database* sekunder, merupakan arsip data yang memuat berbagai data yang terkurasi dan menampilkan hasil analisis dari *database* primer. Contoh: InterPro, UniProt, dan Protein Data Bank (PDB).

3. *Database* komposit, merupakan *database* yang berisi data dari *database* primer yang telah dikomparasi dan difilter berdasarkan kriteria tertentu. Contoh: SwissProt+TrEMBL.

Database bioinformatika juga terbagi berdasarkan data yang termuat di dalamnya, misalnya *database* khusus DNA, *database* khusus penyedia sekuens DNA *barcoding*, *database* protein, *database* taksonomi, dan sebagainya. Selain itu, terdapat juga *database* yang memuat berbagai tipe data sekaligus seperti NCBI.

Beberapa lembaga membangun bank data mengenai sekuens-sekuens yang telah dipublikasi dan mengembangkan banyak *software* untuk menganalisis berbagai sekuens tersebut. Tabel berikut merupakan beberapa situs yang umum digunakan di dalam bidang bioinformatika

No.	Deskripsi Online Data Base.
-----	-----------------------------

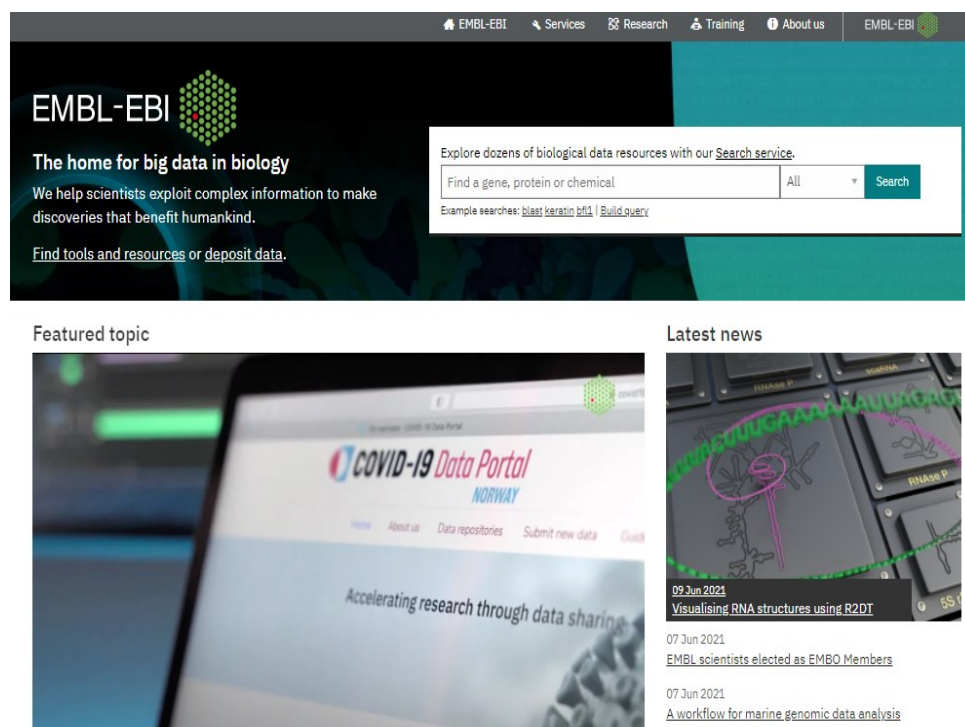
1. *European Bioinformatics Institute* (EBI).

<https://www.ebi.ac.uk/>

European Bioinformatics Institute (EBI) didirikan pada tahun 1997 sebagai bagian dari *the European Molecular Biology Laboratory* (EMBL).

Misi EBI:

- Menyediakan *open access data* dan pelayanan bioinformatika.
- Menyediakan pelatihan bioinformatika secara online.



2 National Center for Biotechnology Information (NCBI).

<https://www.ncbi.nlm.nih.gov/>

3. DNA Data Bank of Japan (DDBJ).

<https://www.ddbj.nig.ac.jp/index-e.html>

4. UniProt.

<https://www.uniprot.org/>

UniProt mengkurasi data dari SwissProt dan TrEMBL. Data yang terdapat di dalam UniProt sudah terkurasi secara lengkap dan berisi fungsi, struktur, sekuens, penggunaan dalam bioteknologi dan patologi, serta rumpun protein yang masih berada di dalam satu famili protein tersebut.

The screenshot displays the UniProt website interface. The top navigation bar includes links for BLAST, Align, Retrieve/ID mapping, Peptide search, SPARQL, and Help/Contact. A prominent banner announces 'The new UniProt website is here!' with a link to the UniProt BETA site. Below the banner, the mission statement is provided: 'The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.'

The main content area is divided into several sections:

- UniProtKB** (UniProt Knowledgebase): Includes Swiss-Prot (566,996) with manually annotated and reviewed records, and TrEMBL (230,328,648) with automatically annotated records.
- UniRef**: Sequence clusters.
- UniParc**: Sequence archive.
- Proteomes**: Proteome sets.
- Supporting data**: Literature citations, Taxonomy, and Subcellular locations.
- News**: A section for forthcoming changes and planned updates, including 'UniProt release 2022_01' and 'UniProt release 2021_04'.

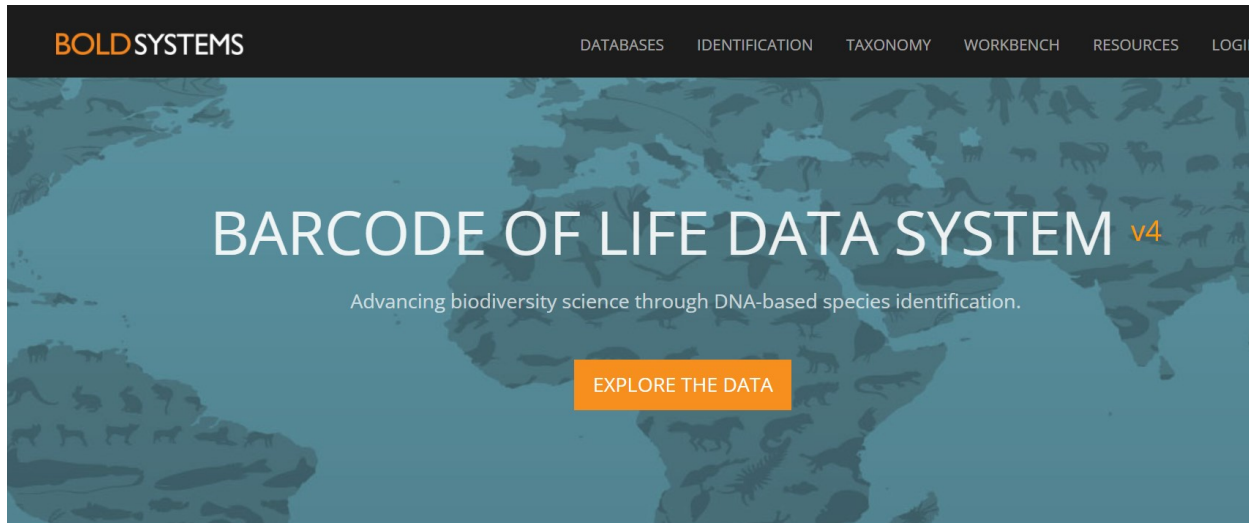
The bottom section shows a detailed view of the UniProtKB entry for **P0DTD1 (R1AB_SARS2)**. The entry is for the **Replicase polyprotein 1ab** (rep) from **Severe acute respiratory syndrome coronavirus 2 (2019-nCoV) (SARS-CoV-2)**. The status is 'Reviewed' with an annotation score of 5 stars, indicating experimental evidence at the protein level.

The 'Function' section provides detailed information:

- Replicase polyprotein 1ab**: Multifunctional protein involved in the transcription and replication of viral RNAs. Contains the proteinases responsible for the cleavages of the polyprotein.
- Host translation inhibitor nsp1**: Inhibits host translation by associating with the open head conformation of the 40S subunit (PubMed:33479166, PubMed:33080218, PubMed:32680882, PubMed:32908316). The C-terminus binds to and obstructs ribosomal mRNA entry tunnel (PubMed:33479166, PubMed:33080218, PubMed:32680882, PubMed:32908316). Thereby inhibits antiviral response triggered by innate immunity or interferons (PubMed:33080218, PubMed:32680882, PubMed:32979938). The nsp1-40S ribosome complex further induces an endonucleolytic cleavage near the 5'UTR of host mRNAs, targeting them for degradation (By similarity). Viral mRNAs less susceptible to nsp1-mediated inhibition of translation, because of their 5'-end leader sequence (PubMed:32908316, PubMed:33080218).
- Non-structural protein 2**: May play a role in the modulation of host cell survival signaling pathway by interacting with host PHB and PHB2. Indeed, these two proteins play a role in maintaining the functional integrity of the mitochondria and protecting cells from various stresses.
- Papain-like protease nsp3**: Responsible for the cleavages located at the N-terminus of the replicase polyprotein. Participates together with nsp4 in the assembly of virally-induced cytoplasmic double-membrane vesicles necessary for viral replication (By similarity).
- Antagonizes innate immune induction of type I interferon** by blocking the phosphorylation, dimerization and subsequent nuclear translocation of host IRF3 (PubMed:32733001).

5. Barcode of Life (BOLD System).

<https://www.boldsystems.org/>



DESIGNED TO SUPPORT THE GENERATION & APPLICATION OF DNA BARCODE DATA

BOLD is a cloud-based data storage and analysis platform developed at the Centre for Biodiversity Genomics in Canada. It consists of four main modules, a data portal, an educational portal, a registry of BINs (putative species), and a data collection and analysis workbench.

Beberapa *database* tersebut saling berintegrasi dalam menggabungkan, memaparkan, maupun memperbaiki informasi tentang sekuens DNA atau protein dari suatu organisme. Contohnya, data sekuens transkriptom sengan yang diunggah ke DDBJ juga tersedia dan dapat diakses melalui laman NCBI.

Berbagai Fitur pada NCBI (National Center for Biotechnology Information)

NCBI merupakan bagian dari *United States National Library of Medicine* (NLM), bagian dari *the National Institutes of Health*. NCBI yang berada di Bethesda, Maryland ini didirikan pada tahun 1988 dan merupakan suatu institusi yang menyediakan sumber informasi terkait perkembangan biologi molekuler. *Website* NCBI bersifat komprehensif untuk para ahli biologi di seluruh dunia karena *database* yang berbasis biologi tersebut merupakan salah satu bank data gen, protein dan literatur terlengkap serta menyediakan *software* untuk analisis data molekuler, termasuk *software* untuk menganalisis DNA, data genom, protein 3D, pencarian/*design primer*, pencarian *conserve domain* dan lain sebagainya. NCBI mempunyai sistem yang mengakomodasi penyimpanan dan pengambilan data. *Database* NCBI yang dapat diakses oleh publik tersebut diperbaharui (*update*) sesuai dengan penemuan-penemuan terkini yang menyangkut DNA, protein, senyawa aktif dan taksonomi.

Website NCBI menyediakan program BLAST (*Basic Local Alignment Search Tool*) yaitu program untuk menganalisis kesamaan dan didesain untuk mengeksplorasi semua *database* sekuens, baik DNA mau protein. Program BLAST digunakan untuk menyediakan identitas dan fungsi urutan *query*, membantu mengarahkan desain eksperimen untuk membuktikan fungsi dari urutan sekuens, dan untuk menemukan urutan sekuens yang sama dengan organisme lain. Baik sekuens nukleotida maupun protein dapat digunakan sebagai materi (*query*) dalam program BLAST. Selain BLAST, terdapat aplikasi lain dalam bioinformatika yang dapat digunakan untuk memperdalam analisis biologi molekuler berdasarkan data yang tersedia. *Training* dan tutorial NCBI dapat dilakukan oleh pemula dengan memanfaatkan *GenBank*.

The image displays two screenshots of the NCBI (National Center for Biotechnology Information) website. The top screenshot shows the NCBI Home page with a sidebar menu on the left. The 'Training & Tutorials' link in the sidebar is highlighted with a red circle. The main content area features a 'Welcome to NCBI' message and several interactive buttons: 'Submit', 'Download', 'Learn', 'Develop', 'Analyze', and 'Research'. The right sidebar lists 'Popular Resources' including PubMed, Bookshelf, and BLAST. The bottom screenshot shows the 'Training & Tutorials' page, which includes a search bar and a list of links to various resources such as the 'NCBI C++ Toolkit Manual', 'NCBI Education Page', 'NCBI Glossary', 'NCBI Handbook', and 'NCBI Help Manual'.

Database dan Software NCBI

1. *Entrez*

- Sistem pencarian informasi dalam NCBI menyediakan akses terintegrasi untuk melakukan sekuensing, pemetaan (*mapping*), taksonomi, dan data protein dalam bentuk struktur 3D
- Menyediakan gambaran grafis untuk pemetaan sekuens dan kromosom.
- Kemampuan untuk pencarian informasi terkait sekuens, struktur dan referensi.
- *Entrez Gene* dapat diakses pada <http://www.ncbi.nlm.nih.gov/Entrez/>

2. *GenBank*

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

3. *Database of Expressed Sequence Tags (dbEST)*

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

4. *Database of Genome Survey Sequences (dbGSS)*

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

5. *BLAST (Basic Local Alignment Search Tool)*

BLAST menentukan homologi suatu urutan DNA atau asam amino dengan data yang ada di NCBI.

Pemanfaatan program NCBI terkait nukleotida ada 3 yaitu:

- Search Nucleotida* merupakan pencarian nukleotida berdasarkan kata kunci yang diinginkan. Kata kunci diketikkan pada laman <https://www.ncbi.nlm.nih.gov/nucleotide/>. Kata kunci dapat berupa spesies, gen, atau daerah tertentu pada genom. Semakin spesifik kata kunci maka hasil pencarian nukleotida akan semakin mengerucut pada hasil pencarian yang diinginkan. Hasil pencarian pada halaman awal (no 1-20 atau 1-50) biasanya merupakan hasil yang paling sesuai dengan kata kunci. Contoh: mencari sekuens protein *spike* pada virus SARS-COV-2. Ketika kata kunci yang dimasukkan hanya SARS-COV-2 maka ditemukan hasil pencarian sebanyak 2.595.896 hasil pencarian. Untuk mendapatkan hasil pencarian yang diinginkan, maka perlu ditambahkan kata kunci yang lebih spesifik.

NCBI Resources How To Sign in to NCBI

Nucleotide SARS-CoV-2 Search

Create alert Advanced Help

COVID-19 Information

Public health information (CDC) | Research information (NIH) | SARS-CoV-2 data (NCBI) | Prevention and treatment information (HHS) | Español

Species Summary 20 per page Sort by Default order Send to: Filters: Manage Filters

Animals (5,491)
Fungi (8)
Bacteria (81)
Viruses (2,589,266)
Customize ...

Molecule types
genomic DNA/RNA (2,589,755)
mRNA (5,095)
Customize ...

Source databases
INSDC (GenBank) (2,594,799)
RefSeq (906)
Customize ...

Sequence type
Nucleotide (7,605,805)

REFERENCE GENOME

Was this helpful?

[Severe acute respiratory syndrome coronavirus 2 \(SARS-CoV-2\) reference genome](#)

Severe acute respiratory syndrome coronavirus 2 (Host: human, vertebrates)

ssRNA(+)

RefSeq: NC_045512.2

[NCBI Virus resources](#) [RefSeq genomic segments \(1\)](#) [RefSeq Proteins \(38\)](#) [NCBI SARS-CoV-2](#)

Results by taxon

Top Organisms [Tree](#)

Severe acute respiratory syndrome-related coronavirus (258912)
Homo sapiens (5021)
synthetic construct (796)
Sarbecovirus sp. (147)
Rhinacovirus sp. (85)
All other taxa (935)
More...

Find related data

Database: Select

- ☐ [Homo sapiens clone CnC2t1p1_D6 anti-SARS-CoV-2 immunoglobulin heavy chain variable region mRNA, partial cds](#)
2595894. 378 bp linear mRNA
- Accession: MT658808.1 GI: 1862605895
- [Protein](#) [Taxonomy](#)
- [GenBank](#) [FASTA](#) [Graphics](#)
- ☐ [Homo sapiens clone CnC2t1p1_B4 anti-SARS-CoV-2 immunoglobulin heavy chain variable region mRNA, partial cds](#)
2595895. 358 bp linear mRNA
- Accession: MT658807.1 GI: 1862605893
- [Protein](#) [Taxonomy](#)
- [GenBank](#) [FASTA](#) [Graphics](#)
- ☐ [Homo sapiens clone CnC2t1p1_B10 anti-SARS-CoV-2 immunoglobulin heavy chain variable region mRNA, partial cds](#)
2595896. 367 bp linear mRNA
- Accession: MT658806.1 GI: 1862605891
- [Protein](#) [Taxonomy](#)
- [GenBank](#) [FASTA](#) [Graphics](#)

<< First < Prev Page 129 of 129795 Next > Last >>

Summary 20 per page Sort by Default order

Send to:

Dalam kasus ini, dicoba untuk menambahkan kata kunci *surface glycoprotein*. *Surface glycoprotein* adalah protein pada permukaan virus yang berfungsi menempelkan virus pada sel target. Hasil pencarian berkurang menjadi 1.345.876 hasil pencarian.

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

☐ [Reverse genetics vector pCCI-4K-SARS-CoV-2-NanoLuc, complete sequence](#)

1345873. 35,853 bp circular other-genetic

Accession: MT926412.1 GI: 1906746442

[Protein](#) [PubMed](#) [Taxonomy](#)

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

☐ [Reverse genetics vector pCCI-4K-SARS-CoV-2-mCherry, complete sequence](#)

1345874. 36,048 bp circular other-genetic

Accession: MT926411.1 GI: 1906746389

[Protein](#) [PubMed](#) [Taxonomy](#)

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

☐ [Reverse genetics vector pCCI-4K-SARS-CoV-2-Wuhan-Hu-1, complete sequence](#)

1345875. 35,283 bp circular other-genetic

Accession: MT926410.1 GI: 1906746340

[Protein](#) [PubMed](#) [Taxonomy](#)

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

☐ [Reverse genetics vector pCCI-4K-SARS-CoV-2-ZsGreen, complete sequence](#)

1345876. 36,033 bp circular other-genetic

Accession: MW289908.1 GI: 1986871386

[Protein](#) [PubMed](#) [Taxonomy](#)

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

<< First < Prev Page 672 of 67294 Next > Last >>

Masih di dalam kasus tersebut, dicoba lagi untuk menambahkan kata kunci negara asal sampel virus SARS-COV-2 yaitu dari Indonesia. Hasil pencarian menunjukkan hanya 9 data.

NCBI Resources How To Sign in to NCBI

Nucleotide Nucleotide SARS-COV-2 surface glycoprotein Indonesia Search

Create alert Advanced Help

[gene, complete sequence; and nucleocapsid phosphoprotein \(N\) and ORF10 protein \(ORF10\)](#)
[genes, complete cds](#)
 29,709 bp linear RNA
 Accession: MZ570426.1 GI: 2067818915
[Protein](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

☐ [Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/IDN/RSDS-](#)

8. [RCVTD-UNAIR-33-B/2020 ORF1ab polyprotein \(ORF1ab\), ORF1a polyprotein \(ORF1ab\), surface glycoprotein \(S\), ORF3a protein \(ORF3a\), envelope protein \(E\), membrane glycoprotein \(M\), ORF6 protein \(ORF6\), ORF7a protein \(ORF7a\), and ORF7b \(ORF7b\), genes, complete cds; ORF8 gene, complete sequence; and nucleocapsid phosphoprotein \(N\) and ORF10 protein \(ORF10\)](#)
[genes, complete cds](#)
 29,777 bp linear RNA
 Accession: MZ570402.1 GI: 2067817437
[Protein](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

☐ [Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/IDN/RSDS-](#)

9. [RCVTD-UNAIR-35-B/2020 ORF1ab polyprotein \(ORF1ab\), ORF1a polyprotein \(ORF1ab\), surface glycoprotein \(S\), ORF3a protein \(ORF3a\), envelope protein \(E\), membrane glycoprotein \(M\), ORF6 protein \(ORF6\), ORF7a protein \(ORF7a\), and ORF7b \(ORF7b\), genes, complete cds; ORF8 gene, complete sequence; and nucleocapsid phosphoprotein \(N\) and ORF10 protein \(ORF10\)](#)
[genes, complete cds](#)
 29,777 bp linear RNA
 Accession: MZ570273.1 GI: 2067814842
[Protein](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

Summary 20 per page Sort by Default order Send to:

- B. BLAST *Nucleotide to Database* yaitu melakukan *alignment* atau pensejajaran nukleotida dengan nukleotida yang tersedia di dalam *database*. Contohnya adalah identifikasi fungi menggunakan DNA *Barcoding* yang hasilnya di BLAST pada *database*.
- C. BLAST *Nucleotide to SRA DB*, yaitu pensejajaran nukleotida pada *database Sequence Read Archive (SRA)*. SRA adalah repositori yang menyimpan data-data sekuens hasil *High-Throughput Sequencing* yang biasanya berukuran cukup besar. Sekuens yang berada pada repositori SRA biasanya berupa sekuens genom keseluruhan, transkriptom, dan metagenomik. Contoh: BLAST salah satu sekuens RNA sengan pada *database SRA* adalah sebagai berikut:

Selain memasukkan sekuens, kolom *SRA Experiment Set* harus terisi dengan nama spesies yang kita ingin sejajarkan dengan sekuens kita. Hasil BLAST dengan SRA menunjukkan sekuens yang kita masukkan memiliki kesamaan dengan beberapa sekuens dari SRA. Berbagai sekuens dari *database* yang memiliki kesamaan dengan sekuens yang kita masukkan berasal dari satu unggahan SRA.

BLAST® » blastn suite-SRA » results for RID-VBKCTAZZ01R

Home Recent Results Saved Strategies Help

< Edit Search Save Search Search Summary v How to read this report? BLAST Help Videos Back to Traditional Results Page

Job Title Nucleotide Sequence

RID VBKCTAZZ01R Search expires on 12-13 17:23 pm Download All v

Program BLASTN Citation v

Database SRA See details v

Query ID IdlQuery_13267

Description None

Molecule type dna

Query Length 1163

Other reports Distance tree of results MSA viewer v

Descriptions Graphic Summary Alignments

Sequences producing significant alignments Download v Manage columns Show 100 v

select all 100 sequences selected Graphics Distance tree of results MSA Viewer

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
DRX174243	267	267	12%	4e-68	98.67%	150	SRA.DRR183762.31364289.1
DRX174243	267	267	12%	4e-68	98.67%	150	SRA.DRR183762.13169362.2
DRX174243	261	261	12%	2e-66	98.00%	150	SRA.DRR183762.26899784.1
DRX174243	261	261	12%	2e-66	98.00%	150	SRA.DRR183762.24324055.1
DRX174243	261	261	12%	2e-66	98.00%	150	SRA.DRR183762.5745404.1
DRX174243	261	261	12%	2e-66	98.00%	150	SRA.DRR183762.82052.1
DRX174243	259	259	12%	6e-66	97.99%	150	SRA.DRR183762.13169362.1
DRX174243	252	252	12%	1e-63	97.32%	150	SRA.DRR183762.31678879.2

Database dan *software (analysis tools)* NCBI yang sering digunakan untuk analisis adalah sebagai berikut:

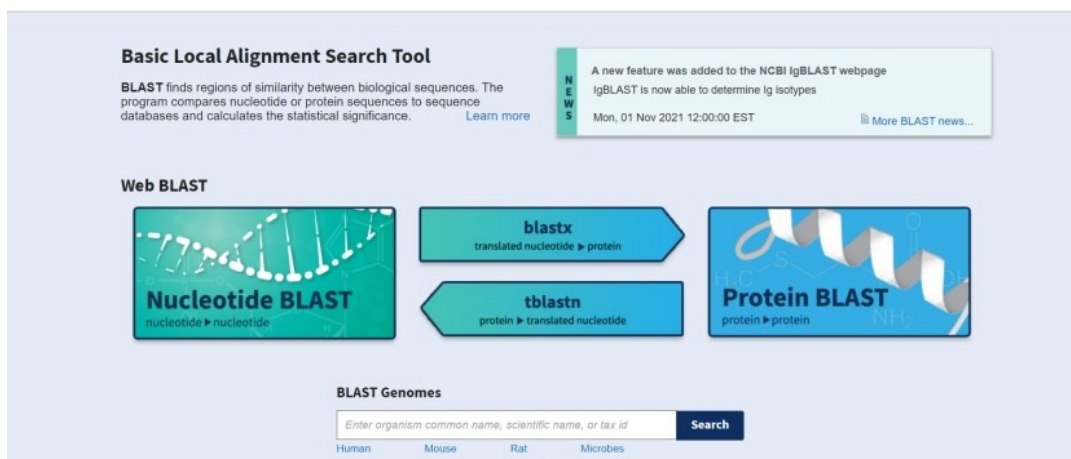
A. DNA-RNA TOOLS:

GenBank

GenBank merupakan bagian dari *database* nukleotida internasional yang bekerja sama dengan *Data Bank of Japan (DDBJ)*, *the European Molecular Biology Laboratory (EMBL)*, and *Genbank* di NCBI. *BioSystems Database* berisi tentang korelasi biologi dari gen, protein dan *small molecule* berdasarkan literatur. *BioSystems Database* dapat digunakan untuk menganalisis fungsi protein dan interaksi protein yang diteliti pada level sel. Sistem ini terhubung dengan beberapa *database biosystem* seperti *reactome* dan KEGG.

Contoh: Untuk mengetahui keberadaan/fungsi protein atau gen di dalam sel, maka masukkan *keyword* ke dalam *box*, maka akan diperoleh hasil yang menunjukkan fungsi gen.

- B. *Basic Local Alignment Search Tool (BLAST)* adalah algoritma dan program berbasis *web* untuk membandingkan sekuens biologis, seperti asam amino (protein), DNA, maupun RNA. Prinsip penggunaan BLAST adalah membandingkan sampel sekuens nukleotida, yang selanjutnya disebut *query*, dengan sekuens-sekuens yang memiliki kemiripan dengan *query* berdasarkan kriteria tertentu pada *database*. BLAST dapat diakses pada <https://blast.ncbi.nlm.nih.gov/Blast.cgi>.

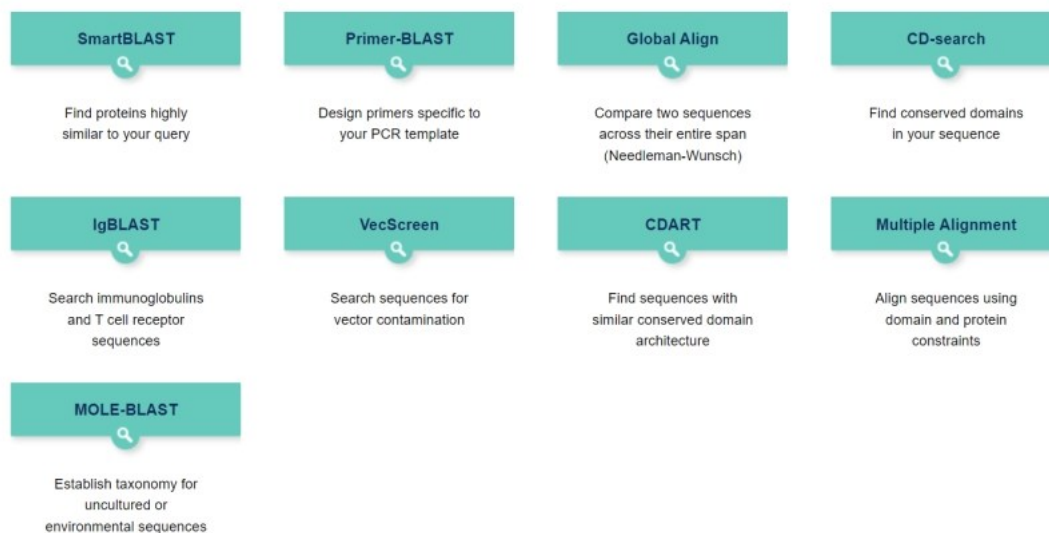


Pada laman BLAST di NCBI, terdapat lima program utama yang dapat digunakan:

1. BLASTn: *nucleotide BLAST*, yaitu program untuk membandingkan *query* (sekuens nukleotida) dengan sekuens nukleotida yang terdapat di dalam *database*. Contoh: membandingkan sekuens DNA dari sampel yang belum diketahui spesiesnya dengan data sekuens DNA yang terdapat di dalam *database*.

2. BLASTx: program untuk membandingkan *query* (sekuens nukleotida) yang ditranslasikan menjadi asam amino, bisa berbentuk sekuens RNA, dengan data sekuens protein yang terdapat di dalam *database*. Contoh: membandingkan RNA hewan dengan protein yang dihasilkan.
 3. tBLASTn: kebalikan dari BLASTx. Program ini membandingkan *query* berupa sekuens protein dengan data sekuens nukleotida yang ditranslasikan pada *database*. Contoh: sekuens protein *spike* pada virus corona di-BLAST untuk mengetahui urutan nukleotida gen yang menghasilkan protein tersebut.
 4. BLASTp: protein BLAST, program untuk membandingkan *query* berupa sekuens protein dengan data sekuens protein yang terdapat di dalam *database*. Contoh: membandingkan protein sampel dengan *treatment* tertentu tetapi belum diketahui jenis proteinnya, dengan data protein yang terdapat di dalam *database* untuk mengetahui jenis protein yang dihasilkan.
 5. tBLASTx: program untuk membandingkan *query* berupa nukleotida dengan data sekuens nukleotida yang telah ditranslasikan dan terdapat di dalam *database*. Contoh: ingin mengetahui nama *query* sekuens RNA yang diekstraksi dari sampel dengan *treatment* tertentu dengan melakukan tBLASTx sehingga diketahui kemungkinan RNA tersebut mengandung kode gen atau asam amino yang mana.
- Selain kelima program utama tersebut, BLAST juga memiliki beberapa program tambahan yaitu:

Specialized searches



1. SmartBLAST: membandingkan *query* berupa protein dengan *database* dan memberikan hasil berupa tiga *best-matched* protein atau yang paling mendekati *query* dari *non-redundant* protein yang terdapat di dalam *database* serta tambahan dua *best-matched* protein dari *reference* spesies yang paling baik.
2. Primer-BLAST: program untuk mendesain primer berdasarkan sekuens nukleotida yang diunggah (di-upload).

3. *Global Align*: program untuk melakukan *alignment* atau pensejajaran antara *query* dengan *database* yang diketahui memiliki kemiripan yang sangat signifikan.
4. *CD-search*: program untuk mencari *Conserved Domain* (CD) yang berupa daerah struktural dan fungsional pada protein. *Conserved domain* adalah struktur 3 dimensi dari protein yang terdiri dari satu atau lebih sekuens motif protein. Sekuens motif protein dalam hal ini merupakan sekuens asam amino yang membentuk kombinasi struktur sekunder yang penting dalam fungsi protein.
5. *IgBLAST*: program untuk mencari sekuens immunoglobulin dan sekuens *reseptor T-cell*.
6. *Vec Screen*: program untuk mendeteksi terdapatnya kontaminasi sekuens dari vektor pada *query*. Umumnya kontaminasi vektor terjadi pada *query* sampel hasil kloning.
7. *CDART: Conserved Domain Architecture Retrieval Tool*, berfungsi untuk menemukan kesamaan antar protein berdasarkan *conserved domain*.
8. *Multiple alignment*: program untuk melakukan pensejajaran lebih dari 3 *query* sekaligus yang panjangnya hampir sama.
9. *Mole-BLAST*: program yang membantu untuk menemukan spesies terdekat dari *query* yang kita unggah.

Hasil *alignment* atau pensejajaran menggunakan BLAST akan menampilkan beberapa nilai, yaitu:

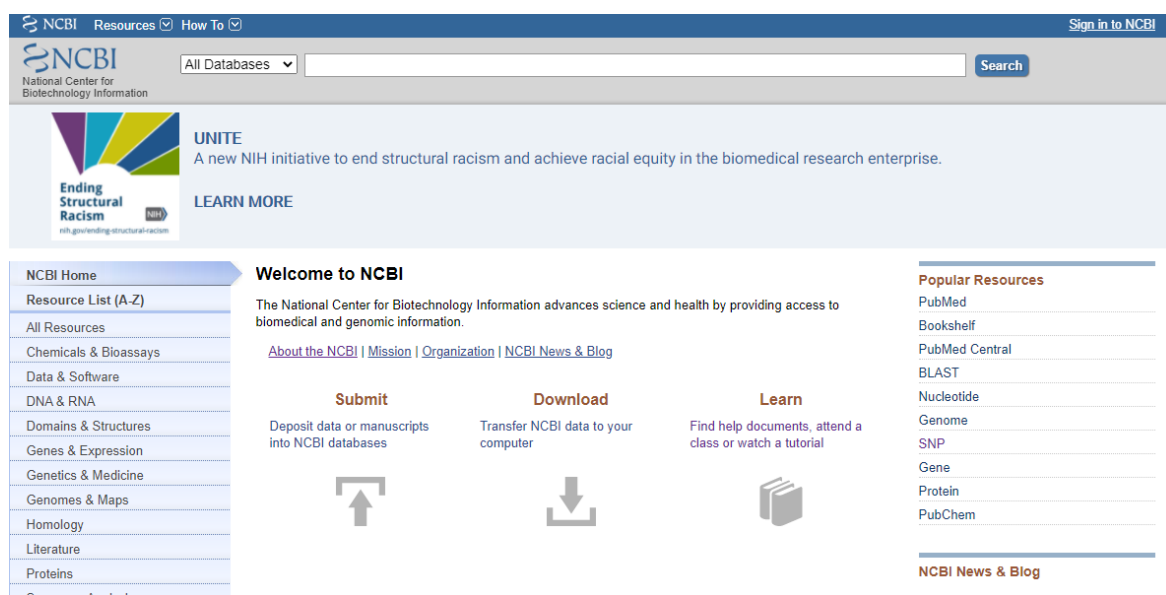
- Score* : Angka yang digunakan untuk menilai relevansi biologis dari suatu temuan. angka yang lebih tinggi sesuai dengan kemiripan yang lebih tinggi.
- Query cover* : Menampilkan bagian dari urutan *query* yang serupa/identik dengan urutan subjek.
- E-value* : Nilai ekspektasi atau nilai harapan yang mewakili jumlah *alignments* yang berbeda dengan skor yang setara dengan atau lebih baik daripada yang diharapkan terjadi dalam pencarian basis data secara kebetulan. Semakin rendah nilai E, semakin signifikan skor dan *Alignment*.
- Identity* : jumlah nukleotida dalam urutan *query* yang identik dengan urutan subjek.
- Accession* : Nomor pengidentifikasi unik di *Database GenBank*.
- Domain* : unit fungsional dan/atau struktural yang berbeda dari protein; Unit lipat secara independen dari polipeptida yang membawa fungsi spesifik.
- Phylogenetic tree* : Diagram atau pohon percabangan yang menunjukkan hubungan evolusi yang disimpulkan dari antara berbagai spesies biologis atau entitas lain berdasarkan filogeninya pada persamaan dan perbedaan dalam karakteristik fisik atau genetiknya.

Unrooted Tree : Menggambarkan keterkaitan '*leaf nodes*' tanpa membuat asumsi tentang leluhur.

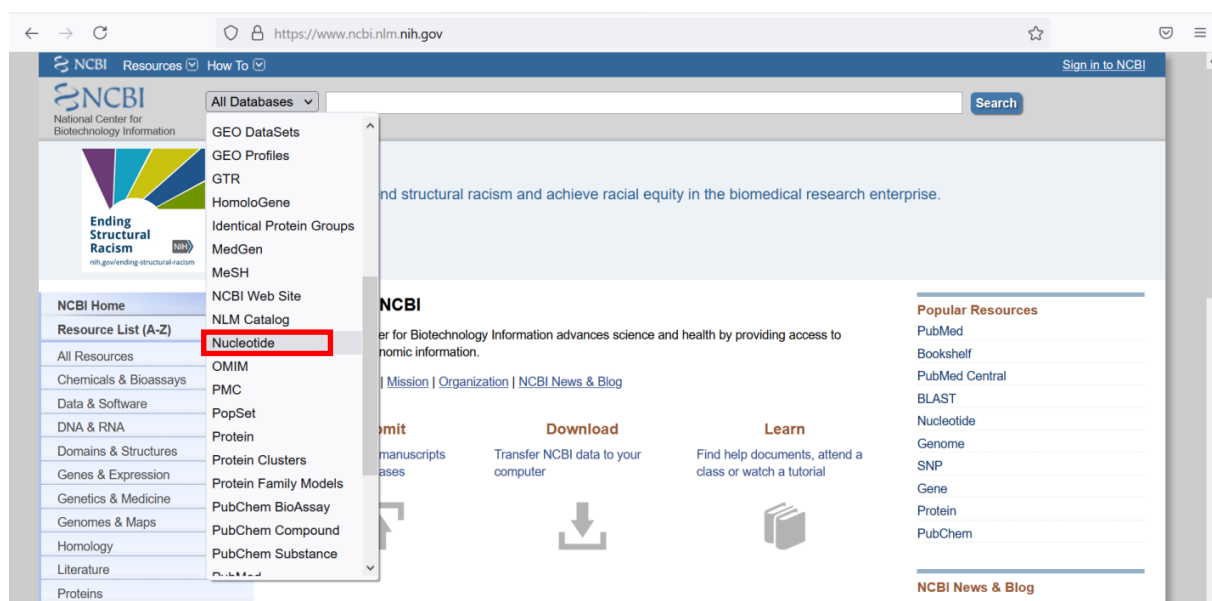
Berikut adalah langkah-langkah untuk mencari dan mendapatkan data dari *GenBank*.

Contoh 1: untuk mencari sekuens *Musa paradisiaca* (pisang), maka dilakukan langkah-langkah sebagai berikut:

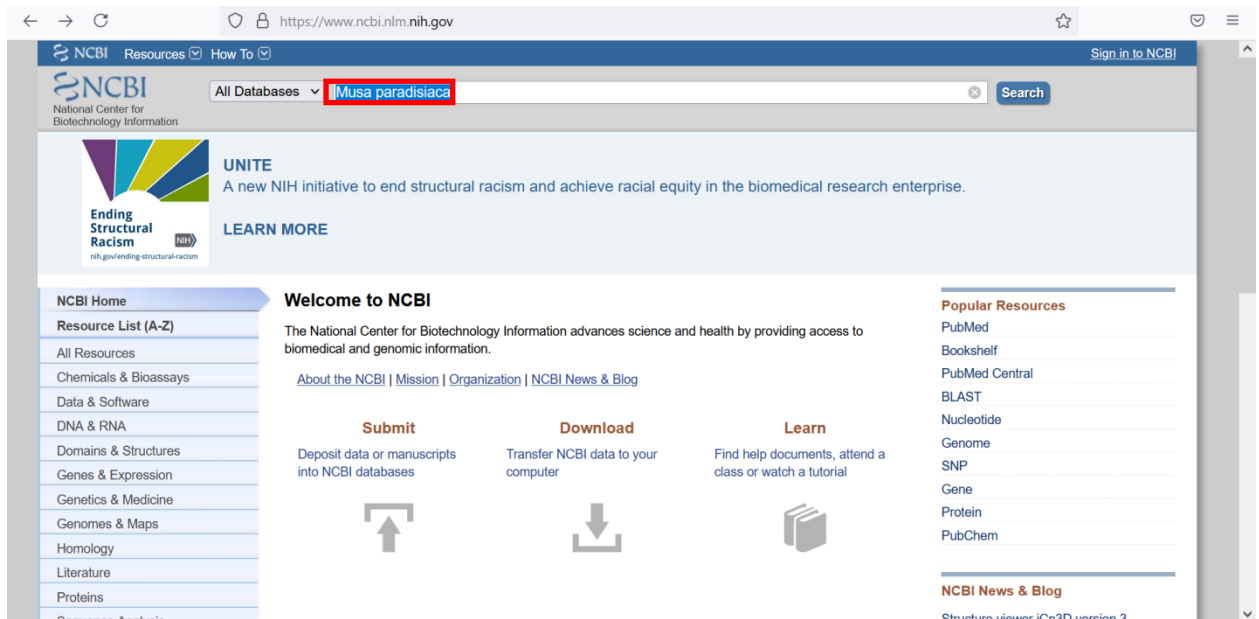
1. Membuka situs lembaga yang berfungsi sebagai *GenBank* dengan situs <https://www.ncbi.nlm.nih.gov/>, dan akan tampil seperti di bawah ini:



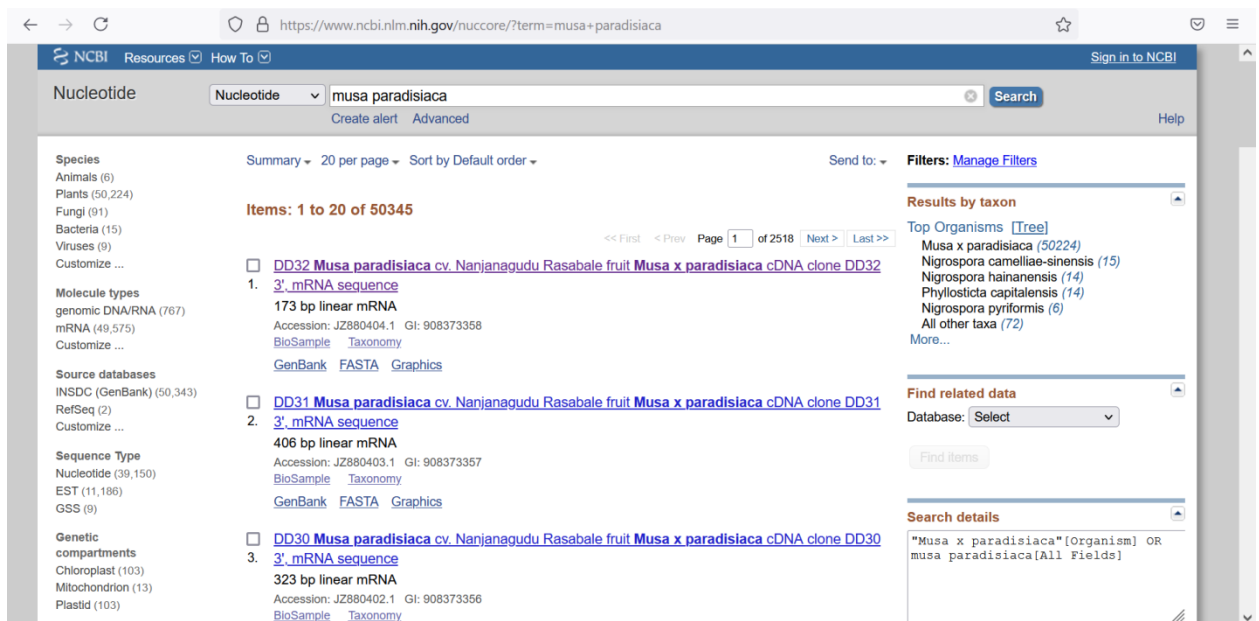
2. Klik kolom *All database* dan pilih *nucleotide*.



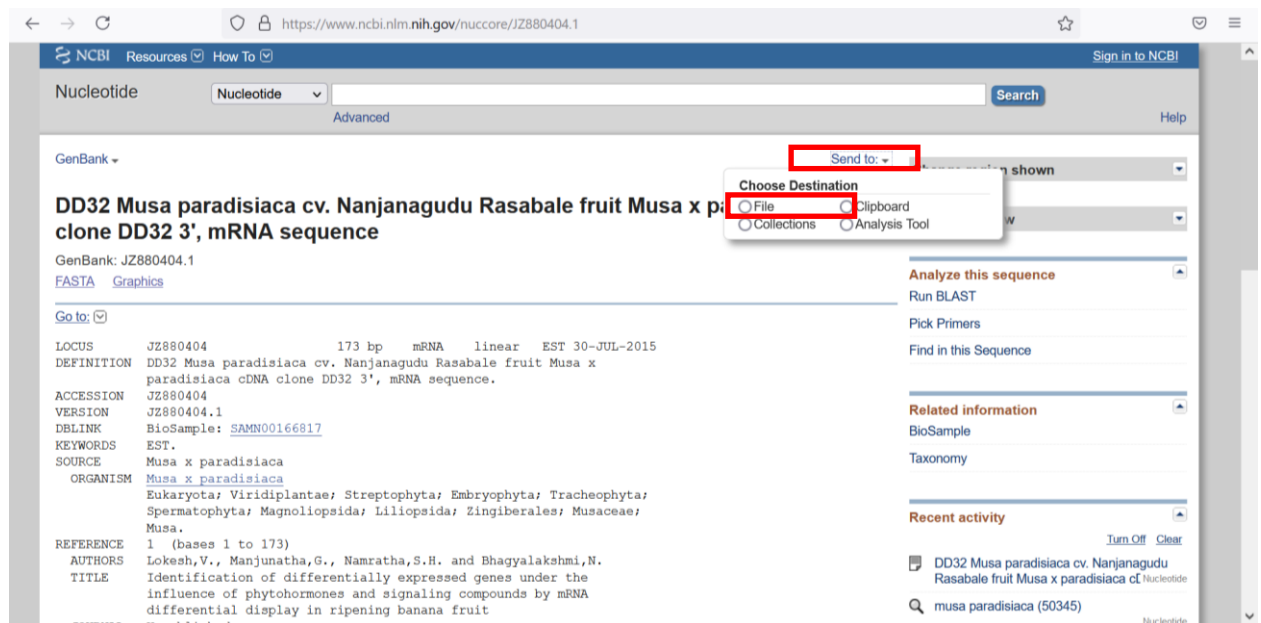
3. Diperlukan *keyword* yang spesifik, misalnya ketik *Musa paradisiaca*.



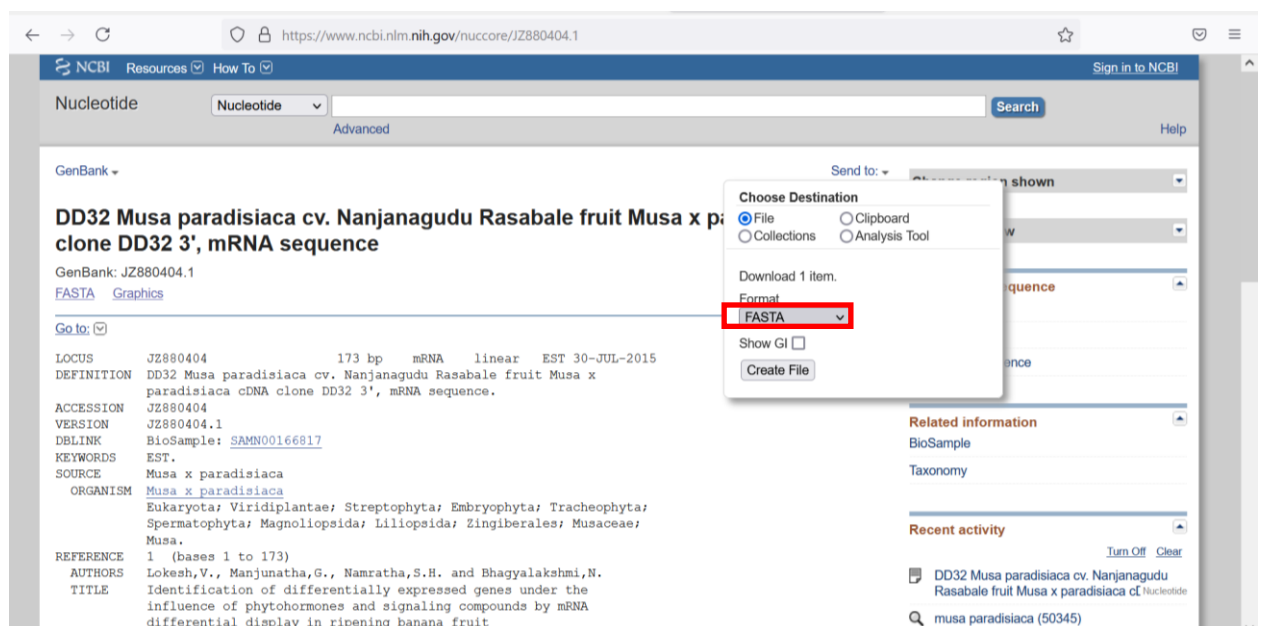
4. Klik *search* tampilan, lalu pilih salah satu dari pilihan yang telah tersedia di layar/monitor.

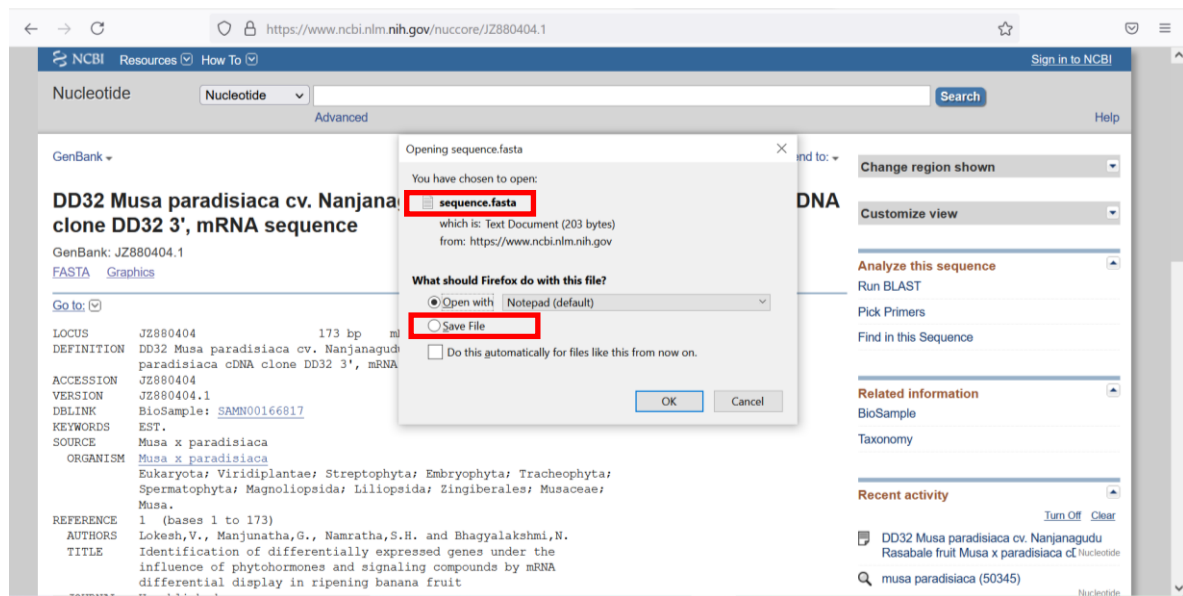


5. Klik *send to*, dan pilih *file*.

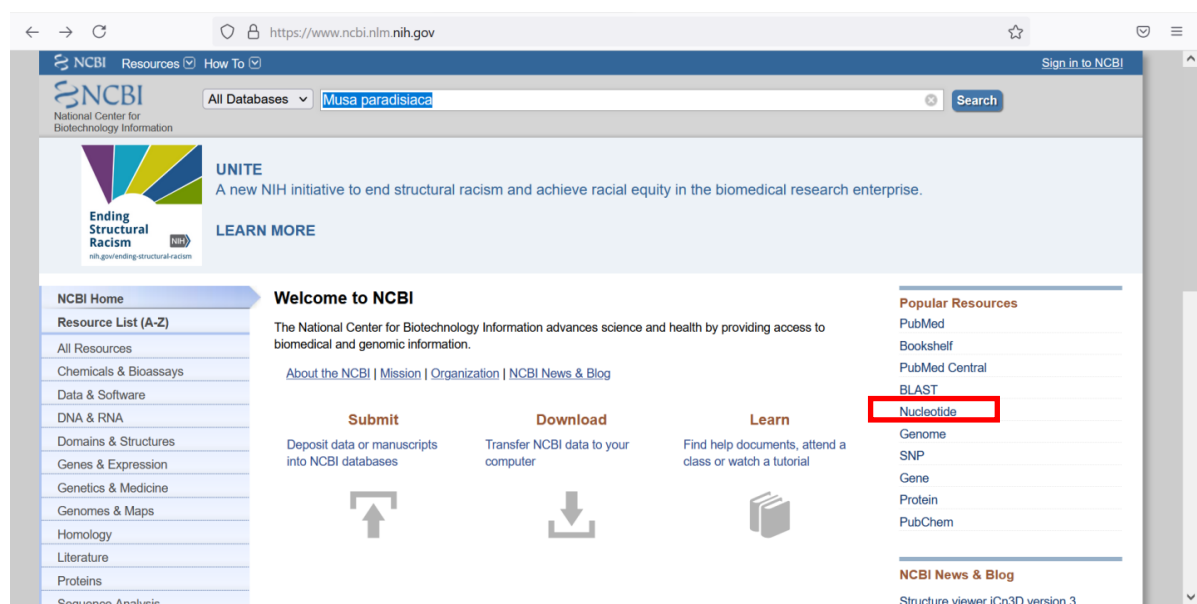


6. Klik format FASTA.

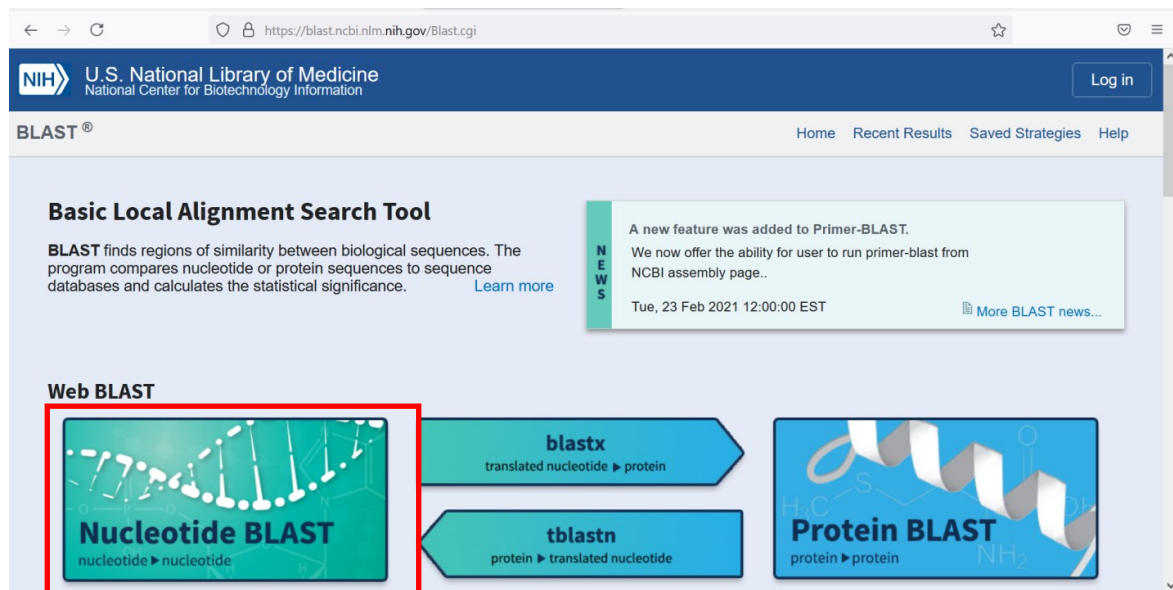


7. Klik *Save file* dan *sequence.fasta*.

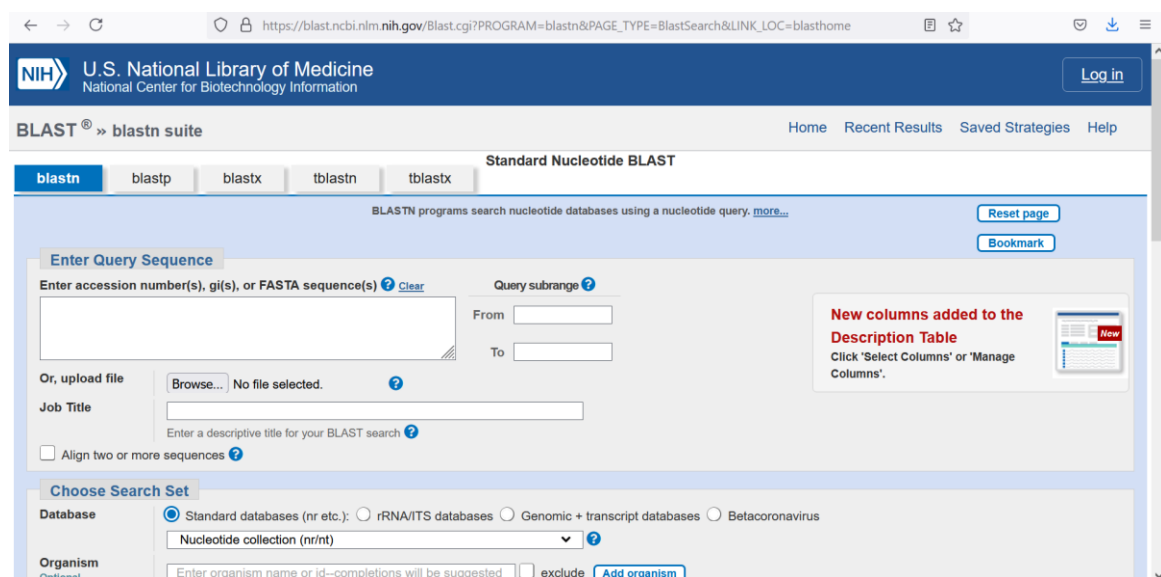
8. Kembali ke menu awal dan klik BLAST.



9. Pada tampilan di layar/monitor klik *Nucleotida BLAST*.



10. Tampilan yang muncul adalah *Sequence Alignment: BLAST (Basic Local Alignment Search)*.



11. Input *copy/paste* urutan nukleotida-format FASTA atau *upload file*.Database : *nucleotide collection*

U.S. National Library of Medicine
National Center for Biotechnology Information

BLAST® » blastn suite

Standard Nucleotide BLAST

BLASTn programs search nucleotide databases using a nucleotide query. [more...](#)

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#)

Query subrange [?](#)

From

To

Or, upload file **Browse...** [sequence.fasta](#) [?](#)

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 [Optional](#) ☐ exclude [Add organism](#)

[New columns added to the Description Table](#)
Click 'Select Columns' or 'Manage Columns'.

12. Pilih *program selection* -----Optimize for -----Highly similar sequence (megablast) kemudian klik BLAST.

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 [Optional](#) ☐ exclude [Add organism](#)

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown [?](#)

Exclude [Optional](#) ☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

Limit to [Optional](#) ☐ Sequences from type material

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

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

+ Algorithm parameters

13. Hasil/output BLAST.

Sequences producing significant alignments

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
PREDICTED: Musa acuminata subsp. malaccensis putative aminoacylate hydrolase RutD (LOC103993790), mRNA	Musa acuminata...	311	311	98%	1e-80	99.42%	1604	XM_009413980.2
Musa acuminata subsp. malaccensis strain Doubled-haploid Pahang (DH-Pahang) genome assembly, chromosome: A08	Musa acuminata...	167	319	100%	3e-37	100.00%	50124231	HG996472.1

Max score: ukuran *statistical significance* dari *alignment* (Score tinggi menunjukkan kelebihan sama satu sekuens daripada sekuens yang lainnya, dari antara dua sekuens).

The E-value (expectation value): nilai E-value rendah menunjukkan kelebihan sama satu sekuens daripada sekuens yang lainnya, dari antara dua sekuens.

PREDICTED: Musa acuminata subsp. malaccensis putative aminoacylate hydrolase RutD (LOC103993790), mRNA

Sequence ID: [XM_009413980.2](#) Length: 1604 Number of Matches: 1

Range 1: 1063 to 1233

Score	Expect	Identities	Gaps	Strand
311 bits(168)	1e-80	170/171(99%)	0/171(0%)	Plus/Minus

Query 1 AGTTATAAACGGCTAGAAGATTGCTTGCCACACCATTTGTTCTTGGTCAATGTCAATG 60
 Sbjct 1233 AGTTATAAACGGCTAGAAGATTGCTTGCCACACCATTTGTTCTTGGTCAATGTCAATG 1174

Query 61 GAGTTCCTATCATGAGCCATCCGGCTTGCTTCTCTTGCATGTTGGACCACTCTTCAGGCT 120
 Sbjct 1173 GAGTTCCTATCATGAGCCATCCGGCTTGCTTCTCTTGCATGTTGGACCACTCTTCAGGCT 1114

Query 121 GAAGCTTGGAATTTGAAGCATTATCAGTCCAAAAGAGAAAGAATAACCT 171
 Sbjct 1113 GAAGCTTGGAATTTGAAGCATTATCAGTCCAAAAGAGAAAGAATAACCT 1063

Range 1: 3516568 to 3516657

Musa acuminata subsp. malaccensis strain Doubled-haploid Pahang (DH-Pahang) genome assembly, chromosome: A08

Sequence ID: [HG996472.1](#) Length: 50124231 Number of Matches: 2

Length : Panjang *alignment*, menunjukkan perbandingan panjang antara dua segmen dari beberapa sekuens yang disejajarkan

Identities : Persentase kesamaan dari nukleotida yang sama dari antara dua sekuens yang disejajarkan

Gaps : Nukleotida yang tidak disejajarkan.

14. Klik *Distance tree of result*.

U.S. National Library of Medicine
National Center for Biotechnology Information

BLAST® » blastn suite » results for RID-CC6JM5X9016

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](#)

Job Title JZ880404.1 DD32 Musa paradisiaca cv. Nanjanagadu
RID CC6JM5X9016 Search expires on 06-14 22:46 pm
[Download All](#)

Program BLASTN [Citation](#)
Database nt [See details](#)
Query ID lcl|Query_27343
Description JZ880404.1 DD32 Musa paradisiaca cv. Nanjanagadu R ...
Molecule type dna
Query Length 173
Other reports [Distance tree of results](#) [MSA viewer](#)

Filter Results

Organism only top 20 will appear ☐ exclude
 Type common name, binomial, taxid or group name
[+ Add organism](#)

Percent Identity to **E value** to **Query Coverage** to
[Filter](#) [Reset](#)

Alignments | Descriptions | Graphic Summary | Taxonomy

Alignment view: Pairwise ☐ CDS feature [Restore defaults](#) [Feedback](#)

15. Tampilan di layar/monitor:

BLAST tree view

This tree was produced using BLAST pairwise alignments. [more...](#)

[Reset Tree](#)

BLAST RID CC6JM5X9016 **Query ID** lcl|Query_27343 **Database** nt

Tree method Fast Minimum Evolution **Max Seq Difference** 0.75 **Sequence Label** Sequence Title (if avail)

Mouse over an internal node for a subtree or alignment. Click on tree label to select sequence to download

Find: all

JZ880404.1 DD32 Musa paradisiaca cv. Nanjanagadu Rasabale fruit Musa x paradisiaca cDNA clone DD32 3' mRNA sequence

PREDICTED: Musa acuminata subsp. malaccensis putative aminoacylate hydrolase RutD (LOC103993790), mRNA

Musa acuminata subsp. malaccensis strain Doubled-haploid Pahang (DH-Pahang) genome assembly, chromosome: A08

Success Nodes 5(0 selected) View port at (0,0) of 1092x269 0.0006

BLAST is a registered trademark of the National Library of Medicine

NCBI National Center for Biotechnology Information, U.S. National Library of Medicine

[Support center](#) [Mailing list](#)

Contoh 2.

Hasil sekuens fungi yang belum diketahui jenisnya, dapat diidentifikasi dengan bantuan *DNA barcoding*, menggunakan primer ITS

Langkah pengerjaan sama dengan contoh *Musa* sp. (Contoh 1, di halaman 77 - 84)

```
>A|
TTTCCTCCGCTTATTGATATGCTTAAGTTCAGCGGGTATTCTACCTAATCCGAGGTCAACCACTAGAAGTATAGGGCTT
TACGGCTGGCAGCTGAGGCCGCTCTGCAAGCGAGGCAATGGAAAATACTACGCTTAGAGTGTGCTACAACCTCCGCCACT
GACTTTGAGGAAGTGCAGGCTGGGCCGCAAGTCCCAACGCTAAGCAACTGGGGCTTAAGGGTTCGAAATGACGCTCGAAT
AGGCATGCCCCACTAGAATACTAATGGGCGCAATGTGCGTTCAAAGATTCGATGATTCACTGAATTCTGCAATTCACATTA
CTTATCGCATTTTCGCTGCGTTCCTCATCGATGCCAGAACCAAGAGATCCGTTGTTGAAAGTTTTAACTTATTTTATAATA
AATTTTATATGAGTTTTACTGGGGAGAAATAAAAGAGTTTTGGTACGGTCCTTCCGGCGCTGCAGCGGCCTAGAGGCCGCG
GCAGTACGCCGAGGAAACAGTGGTAGGTTGCAGAAAGGGTTGGAGTTGTTAGAACTCAGTAATGATCCCTCCGCAGGTT
CACCTACGGAAGCA
```

Hasil BLAST menunjukkan spesies dengan kemiripan tertinggi berdasarkan *E-value* adalah *Annulohypoxyton nitens*

BLAST® » blastn suite » results for RID-VBF0DDV6013

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](#)

Job Title: Nucleotide Sequence

RID: VBF0DDV6013 Search expires on 12-13 16:08 pm [Download All](#)

Program: BLASTN [Citation](#)

Database: nt [See details](#)

Query ID: lcl|Query_15977

Description: None

Molecule type: dna

Query Length: 574

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

Filter Results

Organism: only top 20 will appear ☐ exclude

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 [Select columns](#) Show 100 [?](#)

☒ select all 100 sequences selected

[GenBank](#) [Graphics](#) [Distance tree of results](#) [MSA Viewer](#)

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☒ Annulohypoxyton nitens isolate 91022108 18S ribosomal RNA gene, partial sequence, internal transcribed spacer...	Annulohypoxyton ...	1033	1033	98%	0.0	99.47%	569	EF026138.1
☒ Annulohypoxyton nitens isolate A2S6-D13 18S ribosomal RNA gene, partial sequence, internal transcribed spacer...	Annulohypoxyton ...	1016	1016	99%	0.0	98.78%	597	KJ774045.1
☒ Annulohypoxyton nitens isolate BPEF25 18S ribosomal RNA gene, partial sequence, internal transcribed spacer...	Annulohypoxyton ...	1014	1014	98%	0.0	98.94%	582	KF151846.1
☒ Annulohypoxyton bahnhadengense strain STMA13115 18S ribosomal RNA gene, partial sequence, internal tran...	Annulohypoxyton ...	1000	1000	96%	0.0	99.28%	575	KX376338.1
☒ Annulohypoxyton nitens isolate HSAB01 internal transcribed spacer 1, partial sequence, 5.8S ribosomal RNA gen...	Annulohypoxyton ...	965	965	93%	0.0	99.25%	550	MW715603.1
☒ Annulohypoxyton nitens isolate HY11 internal transcribed spacer 1, partial sequence, 5.8S ribosomal RNA gene a...	Annulohypoxyton ...	953	953	93%	0.0	98.70%	541	KX722230.1
☒ Fungal sp. isolate N2 III (1-3 b) internal transcribed spacer 1, partial sequence, 5.8S ribosomal RNA gene and int...	fungal sp.	929	929	92%	0.0	98.31%	537	MW603257.1

Urutan hasil yang ditampilkan dapat diubah berdasarkan kriteria yang digunakan (*Percent Identity*, *Query Cover*, *E-value*, *Total Score*, maupun *Max Score*), yaitu dengan mengklik kolom kriteria pada tabel hasil di laman hasil BLAST. Kriteria yang digunakan untuk melihat urutan tertinggi kemiripan spesies tergantung pada tujuan penelitian.

Contoh: hasil BLAST dengan kemiripan terbesar berdasarkan *Percent Identity*.

BLAST® » blastn suite » results for RID-VBF0DDV6013 [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](#)

Job Title **Nucleotide Sequence**

RID [VBF0DDV6013](#) Search expires on 12-13 16:08 pm [Download All](#)

Program BLASTN [Citation](#)

Database nt [See details](#)

Query ID lcl|Query_15977

Description None

Molecule type dna

Query Length 574

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

Filter Results

Organism only top 20 will appear ☐ exclude

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](#) [Select columns](#) Show

☒ select all 100 sequences selected [GenBank](#) [Graphics](#) [Distance tree of results](#) [MSA Viewer](#)

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☒ Annulohypoxyton nitens ITS1_5.8S rRNA gene and ITS2 isolate H154	Annulohypoxyton...	900	900	84%	0.0	100.00%	487	FM209453.1
☒ Annulohypoxyton nitens 18S ribosomal RNA gene partial sequence internal transcribed spacer 1 5.8S ribosomal...	Annulohypoxyton...	761	1099	97%	0.0	100.00%	612	EU714403.1
☒ Annulohypoxyton nitens isolate A2S6-D6 internal transcribed spacer 1 partial sequence 5.8S ribosomal RNA ge...	Annulohypoxyton...	850	850	80%	0.0	99.78%	463	KJ774044.1
☒ Annulohypoxyton nitens ITS1_5.8S rRNA gene and ITS2 isolate H189	Annulohypoxyton...	889	889	84%	0.0	99.59%	488	FM209459.1
☒ Annulohypoxyton nitens isolate 91022108 18S ribosomal RNA gene partial sequence internal transcribed spacer...	Annulohypoxyton...	1033	1033	98%	0.0	99.47%	569	EF026138.1
☒ Annulohypoxyton nitens ITS1_5.8S rRNA gene and ITS2 isolate H157	Annulohypoxyton...	883	883	84%	0.0	99.39%	488	FM209455.1

Contoh: hasil BLAST dengan kemiripan terbesar berdasarkan *Total Score*.

BLAST® » blastn suite » results for RID-VBF0DDV6013 [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](#)

Job Title **Nucleotide Sequence**

RID [VBF0DDV6013](#) Search expires on 12-13 16:08 pm [Download All](#)

Program BLASTN [Citation](#)

Database nt [See details](#)

Query ID lcl|Query_15977

Description None

Molecule type dna

Query Length 574

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

Filter Results

Organism only top 20 will appear ☐ exclude

Type common name, binomial, taxid or group name

[+ Add organism](#)

Percent Identity to E value to Query Coverage to

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

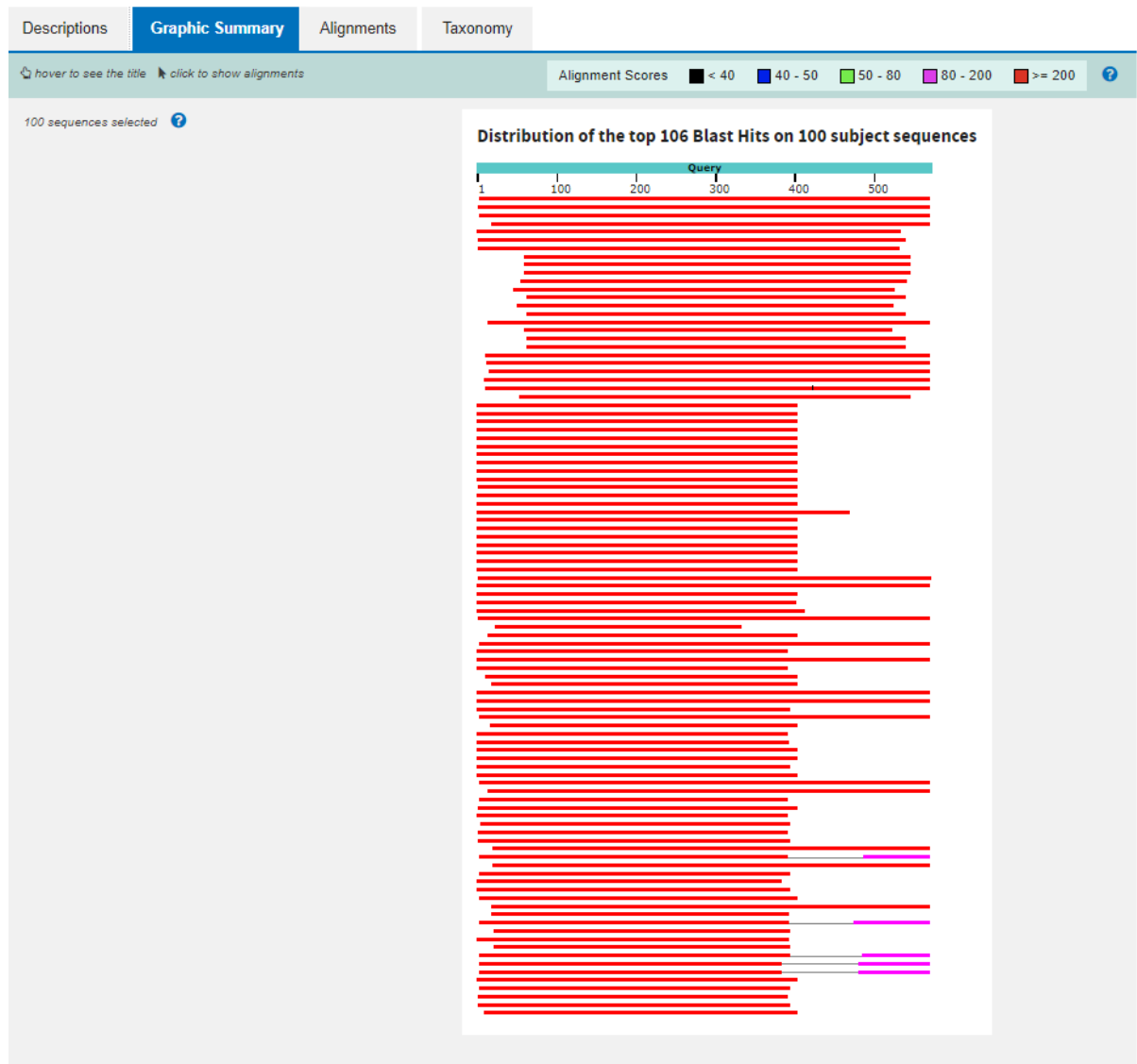
Descriptions [Graphic Summary](#) [Alignments](#) [Taxonomy](#)

Sequences producing significant alignments [Download](#) [Select columns](#) Show

☒ select all 100 sequences selected [GenBank](#) [Graphics](#) [Distance tree of results](#) [MSA Viewer](#)

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☒ Annulohypoxyton nitens 18S ribosomal RNA gene partial sequence internal transcribed spacer 1 5.8S ribosomal...	Annulohypoxyton...	761	1099	97%	0.0	100.00%	612	EU714403.1
☒ Annulohypoxyton nitens isolate 91022108 18S ribosomal RNA gene partial sequence internal transcribed spac...	Annulohypoxyton...	1033	1033	98%	0.0	99.47%	569	EF026138.1
☒ Annulohypoxyton nitens isolate A2S6-D13 18S ribosomal RNA gene partial sequence internal transcribed spac...	Annulohypoxyton...	1016	1016	99%	0.0	98.78%	597	KJ774045.1
☒ Annulohypoxyton nitens isolate BPEF25 18S ribosomal RNA gene partial sequence internal transcribed spacer...	Annulohypoxyton...	1014	1014	98%	0.0	98.94%	582	KF151846.1
☒ Annulohypoxyton bahngphadengense strain STMA13115 18S ribosomal RNA gene partial sequence internal tra...	Annulohypoxyton...	1000	1000	96%	0.0	99.28%	575	KX376338.1
☒ Annulohypoxyton nitens isolate H5AB01 internal transcribed spacer 1 partial sequence 5.8S ribosomal RNA ge...	Annulohypoxyton...	965	965	93%	0.0	99.25%	550	MW715603.1
☒ Annulohypoxyton nitens isolate HY11 internal transcribed spacer 1 partial sequence 5.8S ribosomal RNA gene...	Annulohypoxyton...	953	953	93%	0.0	98.70%	541	KX722230.1

Pada menu *Graphic summary*, dapat dilihat kemiripan sekuens contoh yang akan diidentifikasi dengan sekuens yang terdapat di dalam *database*, dengan syarat sekuens contoh memiliki kesamaan dengan sekuens yang terdapat di dalam *database* atau ter-cover oleh sekuens dari *database*. Garis-garis merah yang terdapat pada *graphic summary* menunjukkan panjang dan letak sekuens pada database yang homolog dengan sekuens yang di-BLAST.



Pada menu *Alignment*, dapat dilihat urutan basa yang dimulai dan berakhir pada sekuens contoh yang memiliki kesamaan dengan sekuens yang terdapat di dalam *database*.

Descriptions Graphic Summary **Alignments** Taxonomy

Alignment view Pairwise ☐ CDS feature [Restore defaults](#) Download

100 sequences selected

[Download](#) [GenBank](#) [Graphics](#) [Next](#) [Previous](#) [Descriptions](#)

Annulohypoxylon nitens isolate 91022108 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete sequence; and 28S ribosomal RNA gene, partial sequence
Sequence ID: [EF026138.1](#) Length: 569 Number of Matches: 1

Range 1: 1 to 569 [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Identities	Gaps	Strand
1033 bits(559)	0.0	566/569(99%)	1/569(0%)	Plus/Minus

Query 3 TCCTCCGCTTATTGATATGCTTAAGTTTCAGCGGGTATTCTACCTAATCCGAGGTCAACC 62
Sbjct 569 TCCTCCGCTTATTGATATGCTTAAGTTTCAGCGGGTATTCTACCTAATCCGAGGTCAACC 510

Query 63 ACTA-GAAGTATAGGGCTTTACGGCTGGCAGCTGAGGCCGCTCTGCAAGCGAGGCAATGG 121
Sbjct 509 ACTAAGAAATATAGGGCTTTACGGCTGGCAGCTGAGGCCGCTCTGCAAGCGAGGCAATGG 450

Query 122 AAAAATACTACGCTTAGAGTGTGCTACAACCTCCGCACTGACTTTGAGGAACTGCGGCT 181
Sbjct 449 AAAAATACTACGCTTAGAGTGTGCTACAACCTCCGCACTGACTTTGAGGAACTGCGGCT 390

Query 182 GGGCCGAGACTCCCAACGCTAAGCAACTGGGGCTTAAGGGTCGAAATGACGCTCGAATA 241
Sbjct 389 GGGCCGAGACTCCCAACGCTAAGCAACTGGGGCTTAAGGGTCGAAATGACGCTCGAATA 330

Query 242 GGCATGCCCACTAGAATACTAATGGGCGCAATGTGCGTTCAAGATTGATGATCACTG 301
Sbjct 329 GGCATGCCCACTAGAATACTAATGGGCGCAATGTGCGTTCAAGATTGATGATCACTG 270

Query 302 AATTCTGCAATTACATTACTTATCGCATTTGCTGCGTTCTTCATCGATGCCAGAACCA 361
Sbjct 269 AATTCTGCAATTACATTACTTATCGCATTTGCTGCGTTCTTCATCGATGCCAGAACCA 210

Query 362 AGAGATCCGTTGTTGAAAGTTTTAACTTATTTTATAATAAATTTATATGAGTTTACTG 421
Sbjct 209 AGAGATCCGTTGTTGAAAGTTTTAACTTATTTTATAATAAATTTATATGAGTTTACTG 150

Query 422 GGGAGAAATAAAGAGTTTGGTACGGTCTTCCGGCGCTGACGCGGCTAGAGGCCCGCG 481
Sbjct 149 GGGAGAAATAAAGAGTTTGGTACGGTCTTCCGGCGCTGACGCGGCTAGAGGCCCGCG 90

Query 482 CAGTACGCCGAGGAAACAGTGGTAGGTTGCAAAAGGGTTGGAGTTGTTAGAACTCAGT 541
Sbjct 89 CAGTACGCCGAGGAAACAGTGGTAGGTTGCAAAAGGGTTGGAGTTGTTAGAACTCAGT 30

Query 542 AATGATCCCTCCGAGGTTACCTACGGA 570
Sbjct 29 AATGATCCCTCCGAGGTTACCTACGGA 1

[Download](#) [GenBank](#) [Graphics](#) [Next](#) [Previous](#) [Descriptions](#)

Annulohypoxylon nitens isolate A2S6-D13 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete sequence; and 28S ribosomal RNA gene, partial sequence
Sequence ID: [KJ774045.1](#) Length: 597 Number of Matches: 1

Range 1: 23 to 594 [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Identities	Gaps	Strand
1016 bits(550)	0.0	565/572(99%)	2/572(0%)	Plus/Minus

Selain BLAST, terdapat beberapa aplikasi pada laman NCBI yang sering digunakan, antara lain:

1. NCBI *tree viewer*, yaitu laman untuk melihat tampilan grafis pohon filogenetik. Aplikasi ini dapat diakses di <https://www.ncbi.nlm.nih.gov/projects/treeview/>
2. *Gnomon*, yaitu program untuk memprediksi dan menganotasi gen dari genom apapun.

Umumnya digunakan untuk menganalisis genom yang pernah dianotasi.

Contoh: memprediksi gen-gen pada lebah *Apis mellifera*, yang mana informasi tentang transkriptom lebah tersebut dari penelitian-penelitian sebelumnya adalah terbatas. Dengan demikian digunakan *Gnomon* untuk memprediksi gen lebah tersebut.

3. *Structure*, yaitu program yang digunakan untuk memprediksi struktur protein. Aplikasi tersebut terdapat di laman <https://www.ncbi.nlm.nih.gov/Structure> dan digunakan dengan cara memasukkan kode Protein Data Bank (PDB) dari protein yang perlu dilihat strukturnya.

Catatan:

Catatan:

BAB IV

ANALISIS DATA SEKUENS PENDEK

Riset sekuensing genom di Indonesia digunakan dalam rangka mengembangkan komoditas unggulan pertanian untuk menemukan gen-gen unggul untuk mempercepat program pemuliaan tanaman. Berkembangnya teknologi sekuensing modern dan biayanya yang semakin terjangkau telah mengubah paradigma penanganan, karakterisasi, proteksi, dan pemanfaatan Sumber Daya Genetika (SDG) tanaman dari yang semula berdasarkan data morfologi tanaman menjadi berdasarkan data genomika. Sekuensing genom tersebut dapat menganalisis genotipe SDG tanaman secara cepat, tepat, akurat, dan komprehensif sehingga pemanfaatannya untuk program pemuliaan tanaman juga lebih komprehensif dan berpresisi tinggi.

Teknologi sekuensing genom juga telah merevolusi teknologi pemuliaan tanaman dari seleksi berdasarkan data morfologi (fenotipe) menjadi seleksi berdasarkan data genom yang jauh lebih efisien, efektif, komprehensif, dan berpresisi tinggi. Informasi mengenai genotipe SDG bisa membantu mempercepat penemuan berbagai variasi genetik untuk mendukung program pemuliaan tanaman secara terarah, lebih cepat, dan lebih akurat dengan menggunakan metode baru, yaitu pemuliaan tanaman berbasis data genom. Di masa depan, program pemuliaan tanaman akan lebih fokus pada perbandingan komposisi genom individu plasma nutfah, sehingga akan membuka peluang penggunaan kombinasi strategi baru dalam pemetaan genetik dan analisis evolusi untuk penemuan dan pemanfaatan keragaman genetik SDG tanaman secara optimal untuk menunjang program pemuliaan yang bersinambungan.

Peta genom acuan suatu spesies tanaman akan menjadi fondasi untuk penemuan gen-gen penting yang berkontribusi pada berbagai karakter yang bernilai komersial. Salah satu aplikasi teknologi sekuensing ialah karakterisasi genom spesies tanaman untuk membuat sekuens genom sebagai acuan bagi suatu spesies tanaman. Sekuens genom acuan ini menjadi pedoman dalam membuat peta genom untuk studi lanjutan berbagai individu aksesori anggota spesies tanaman tersebut. Pada umumnya riset resequencing (*resequencing*) dilakukan untuk mengetahui perbedaan/variasi DNA individu SDG tanaman dibandingkan dengan sekuens genom acuan yang ada. Data variasi di tingkat DNA sangat bermanfaat untuk berbagai macam keperluan, mulai dari perlindungan varietas hingga pemuliaan, karena DNA dapat diambil dari berbagai organ individu yang dipelajari serta tidak dipengaruhi oleh faktor lingkungan. Dengan demikian identifikasi suatu individu maupun sifat-sifat unggul yang dibawanya dapat dilakukan kapan saja sepanjang sampel DNA dapat diperoleh dari individu tersebut.

Untaian sekuens pendek (100–150 bp) tidak perlu disambung (*assembly*) satu dengan yang lainnya, tetapi dipetakan kembali ke sekuens genom acuan untuk mengidentifikasi perbedaan genetik antara sekuens acuan dan sekuens individu kedua, ketiga, dan seterusnya. Dengan demikian sejumlah perbedaan dapat didokumentasikan dan perbedaan genetik tersebut digunakan untuk menjelaskan berbagai perbedaan antar genotipe (varietas) dalam suatu spesies. Contohnya daya hasil, ketahanan hama dan penyakit, waktu pembungaan, serta toleransi kekeringan dan banjir (Long & Ort 2010). Riset resequencing bisa digunakan untuk survei keragaman dalam spesies, *genotyping* dengan sekuensing genom total, konstruksi peta genetik resolusi tinggi, dan pembuatan *Genome Wide Association Studies* (GWAS). Kegiatan resequencing berbagai galur unggul (*superior lines*) dan individu aksesori plasma nutfah dapat mengidentifikasi keunikan setiap genotipe SDG. Hal ini akan menjadi identitas dari varietas, galur, maupun individu aksesori sebagai sidik jari dari individu tersebut.

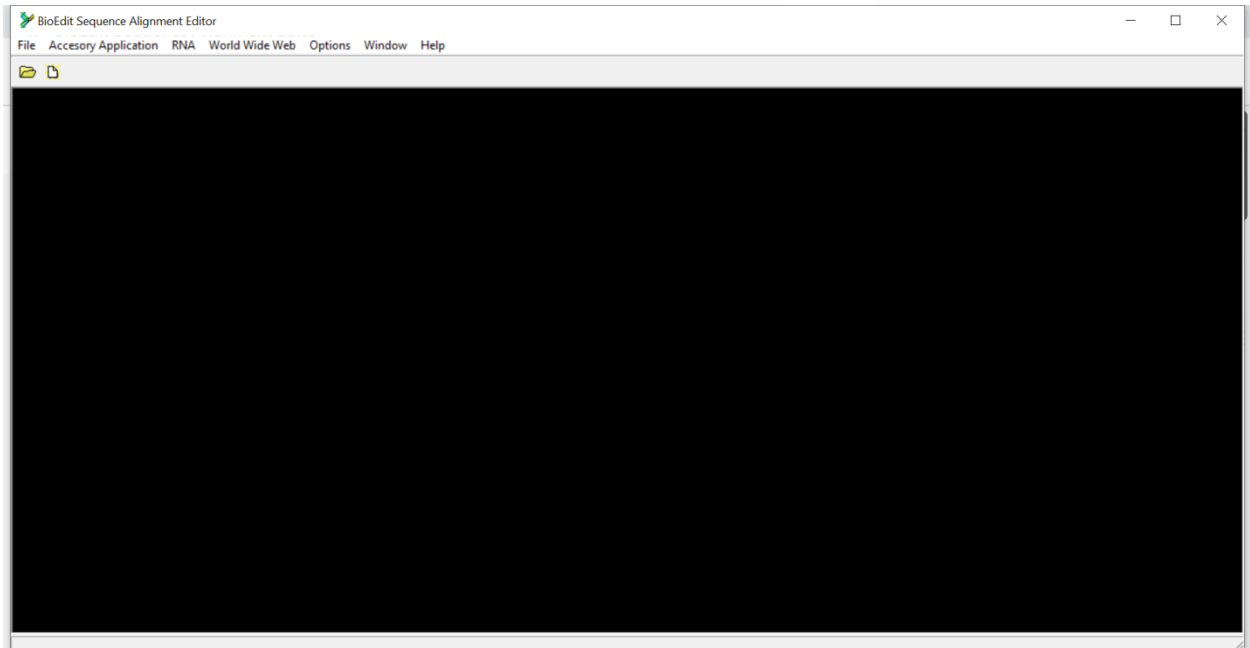
Sekuensing DNA (*DNA sequencing*) adalah proses untuk mendapatkan informasi urutan basa nukleotida dari DNA. Metode yang paling umum digunakan untuk melakukan sekuensing adalah metode Sanger *sequencing*. Di masa kini, sekuensing dengan metode Sanger umumnya tidak lagi menggunakan gel elektroforesis, tetapi menggunakan mesin sekuenser dengan sensor fluoresens untuk mendeteksi basa dan langsung terhubung ke komputer untuk ekstraksi data. Mesin sekuensing dengan metode Sanger umumnya mampu mensekuens 200 – 1000 bp dari satu sampel dalam satu waktu. Umumnya sekuensing metode sanger digunakan dalam DNA *barcoding* konvensional dan sekuensing gen atau fragmen tunggal dari DNA bagian tertentu.

Tahapan awal setelah diperolehnya data sekuens dari mesin yaitu melakukan *quality control* terhadap hasil sekuens tersebut. *Quality control* dalam analisis sekuens dilakukan dengan cara mengedit sekuens. *Editing* yang dilakukan yaitu menghilangkan ujung-ujung sekuens yang memiliki hasil bacaan yang kurang baik, mengkonfirmasi basa-basa yang ambigu, dan menggabungkan sekuens *forward* dan *reverse* jika sekuensing dilakukan secara dua arah. Beberapa *software* yang dapat digunakan untuk *editing* sekuens antara lain *CodonCode Aligner*, ATGC (Genetyx, Jepang), UGENE, *SeqTools*. Pada contoh kali ini, *software* yang digunakan untuk *editing* adalah BioEdit (<https://bioedit.software.informer.com/7.2/>). Sementara itu, untuk *alignment* digunakan *software* Mega X.

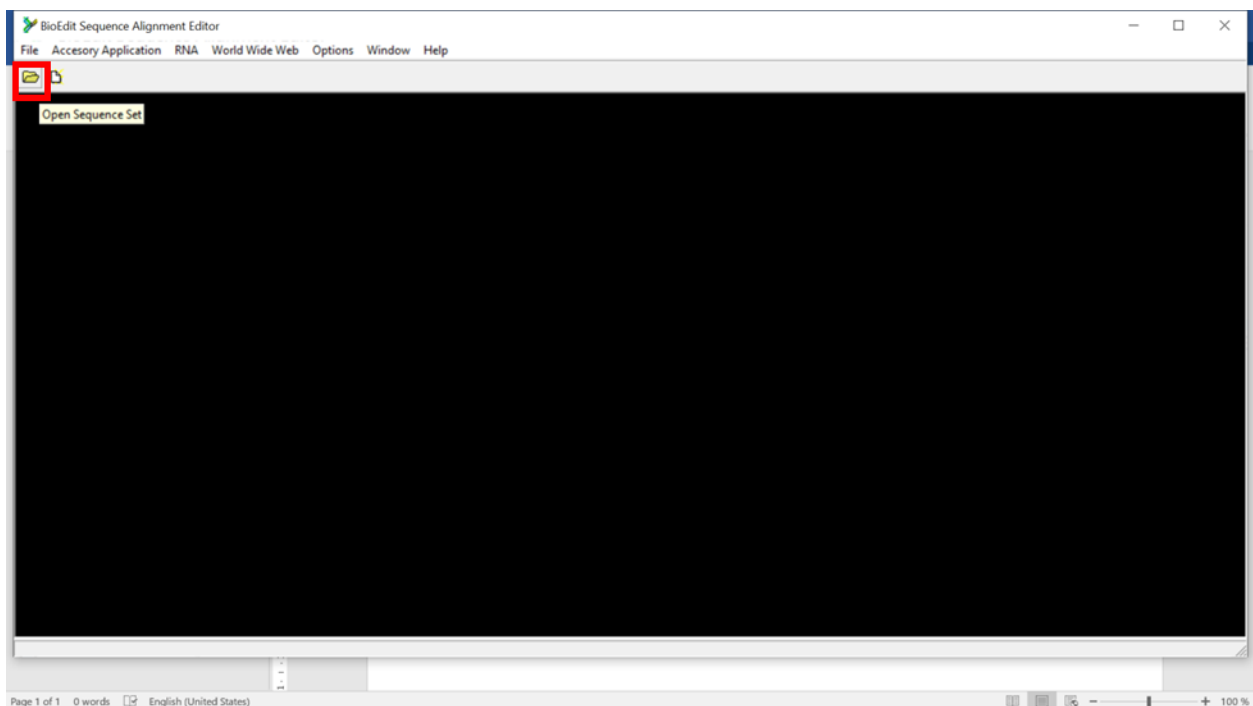
4.1 Editing sekuens menggunakan aplikasi Bioedit

Bioedit (Hall 1999) adalah salah satu *free software* yang dapat digunakan untuk melihat kualitas basa hasil sekuensing, menyatukan sekuens *forward* dan *reverse*, serta beberapa fungsi tambahan seperti *alignment* dan translasi ke protein. Tahapan *editing* dan pembuatan sekuens *consensus* (gabungan *forward reverse*) disajikan sebagai berikut:

1. Buka aplikasi Bioedit.

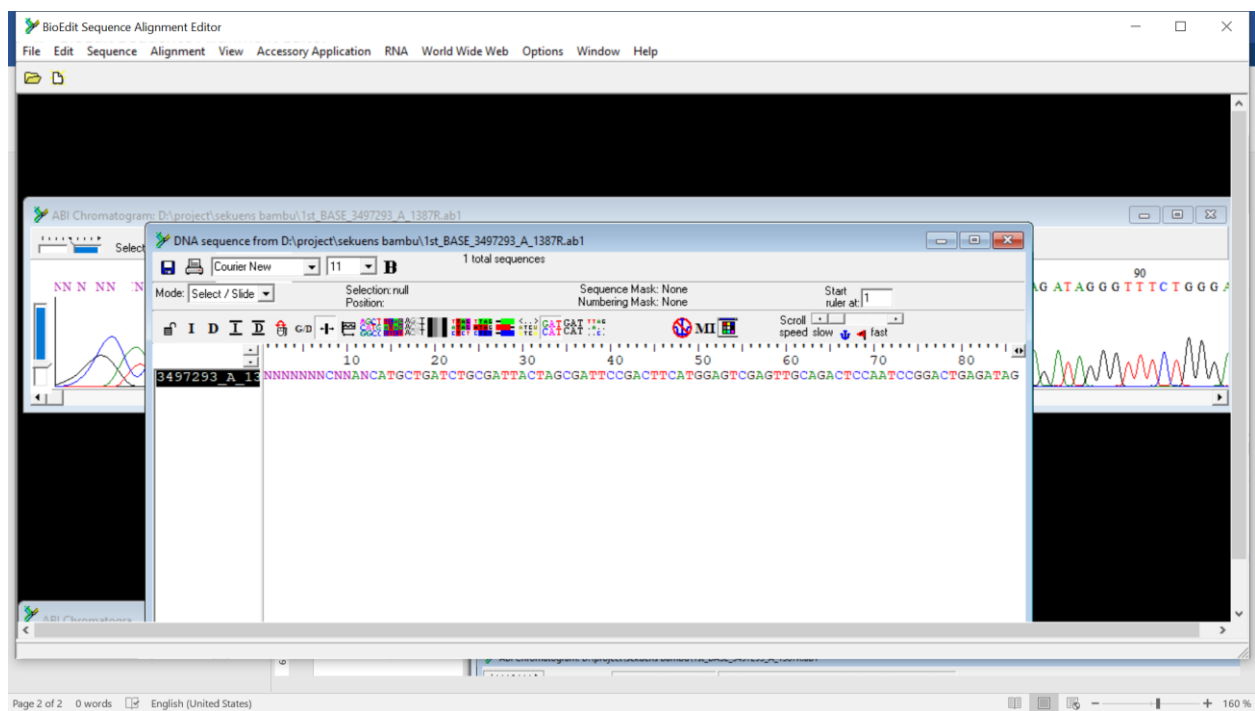


2. Buka sekuens yang akan di-*edit*. Jika sekuens berupa *forward* dan *reverse*, buka kedua sekuens.

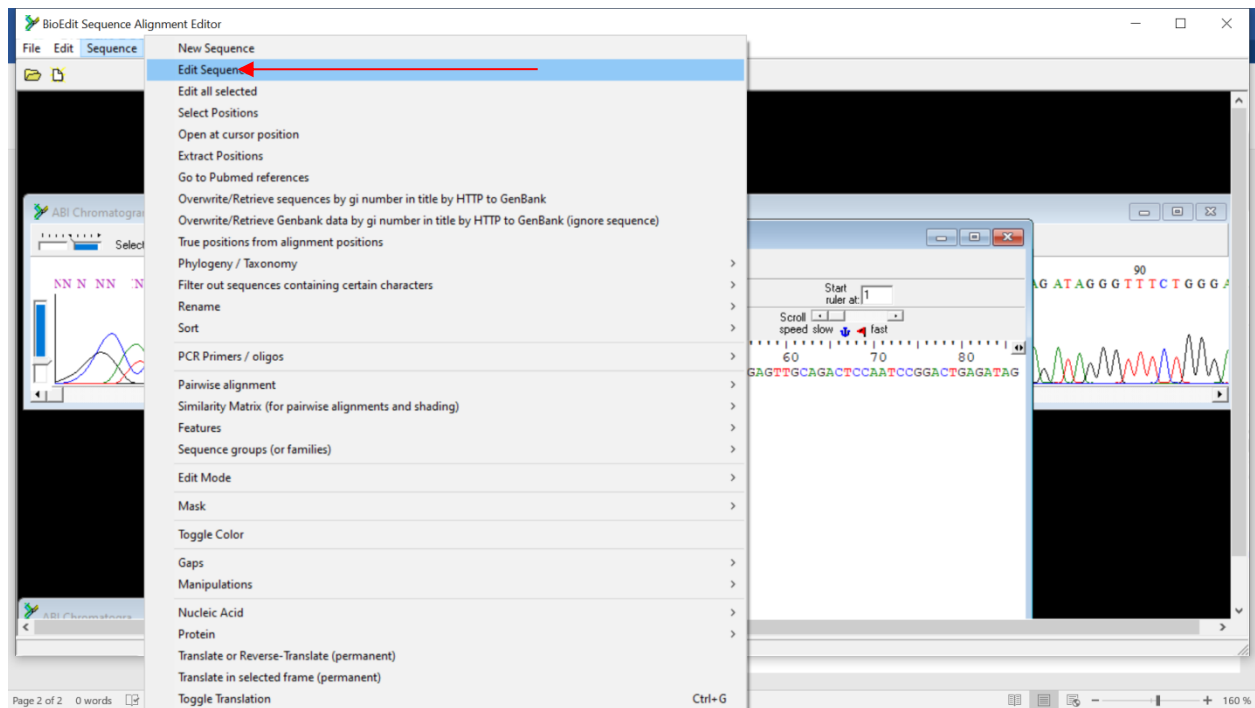




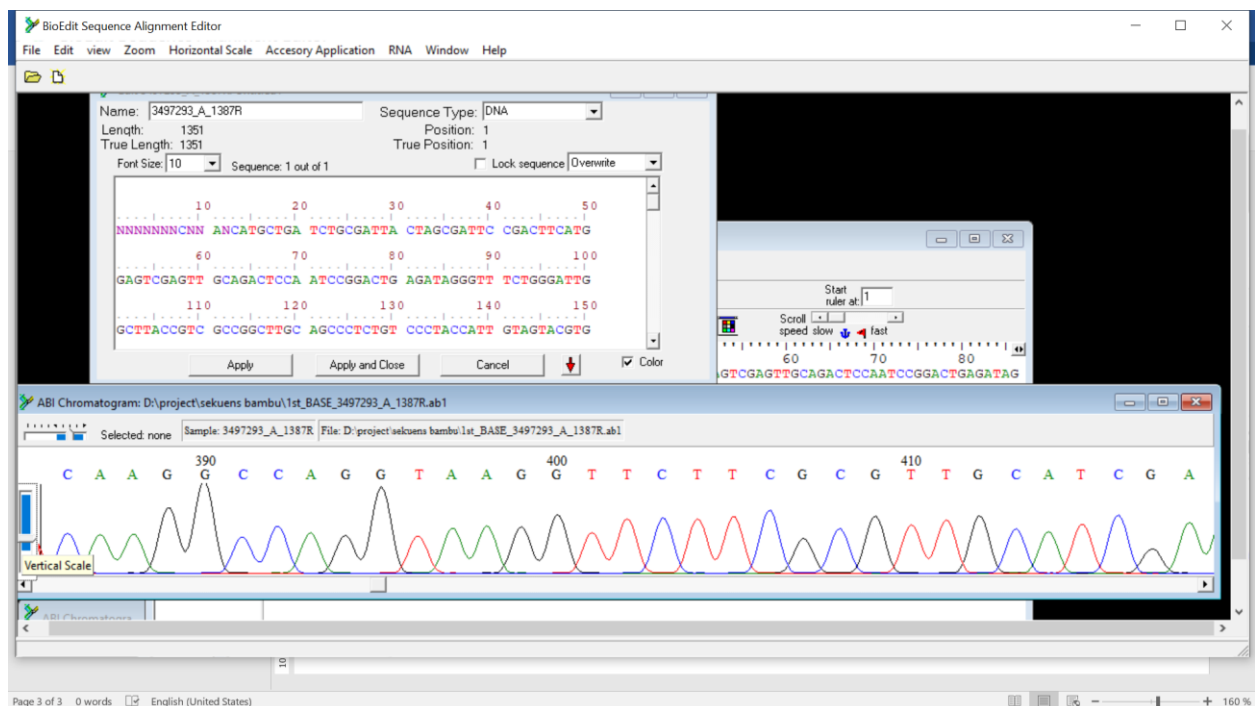
3. Pilih sekuens yang akan di-edit.



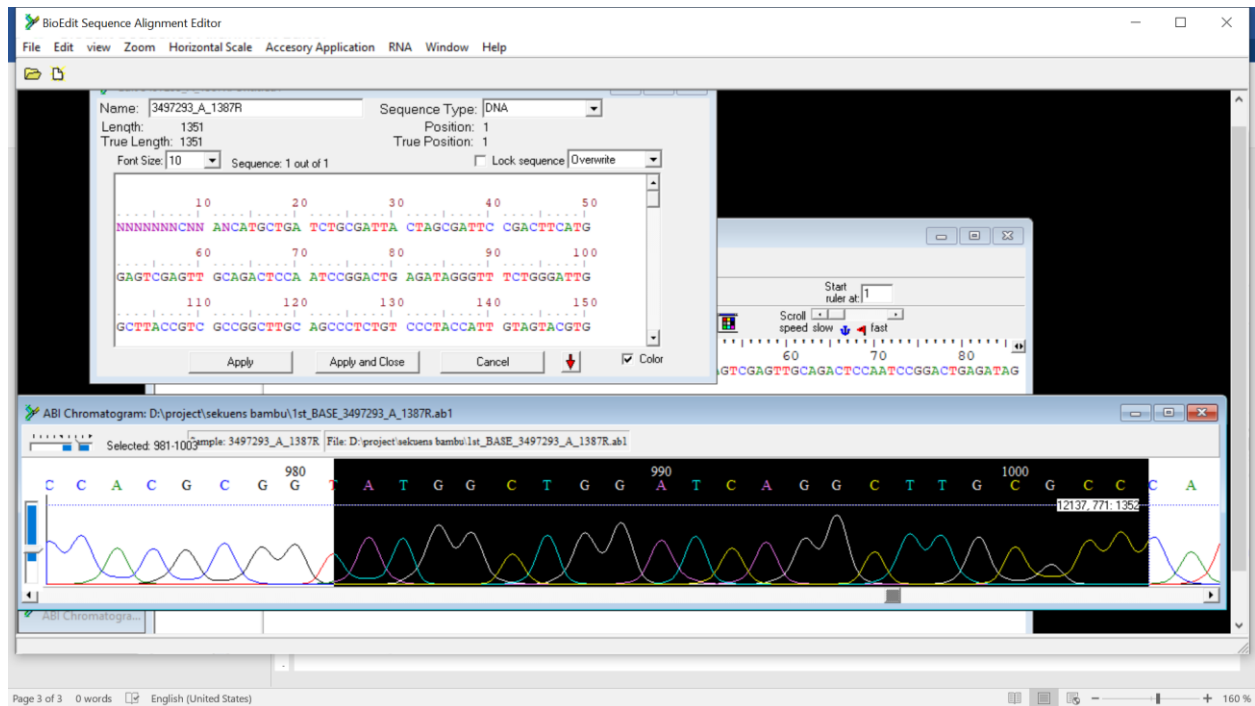
4. Pilih *edit sequence* pada menu *sequence* untuk memudahkan *editing*.



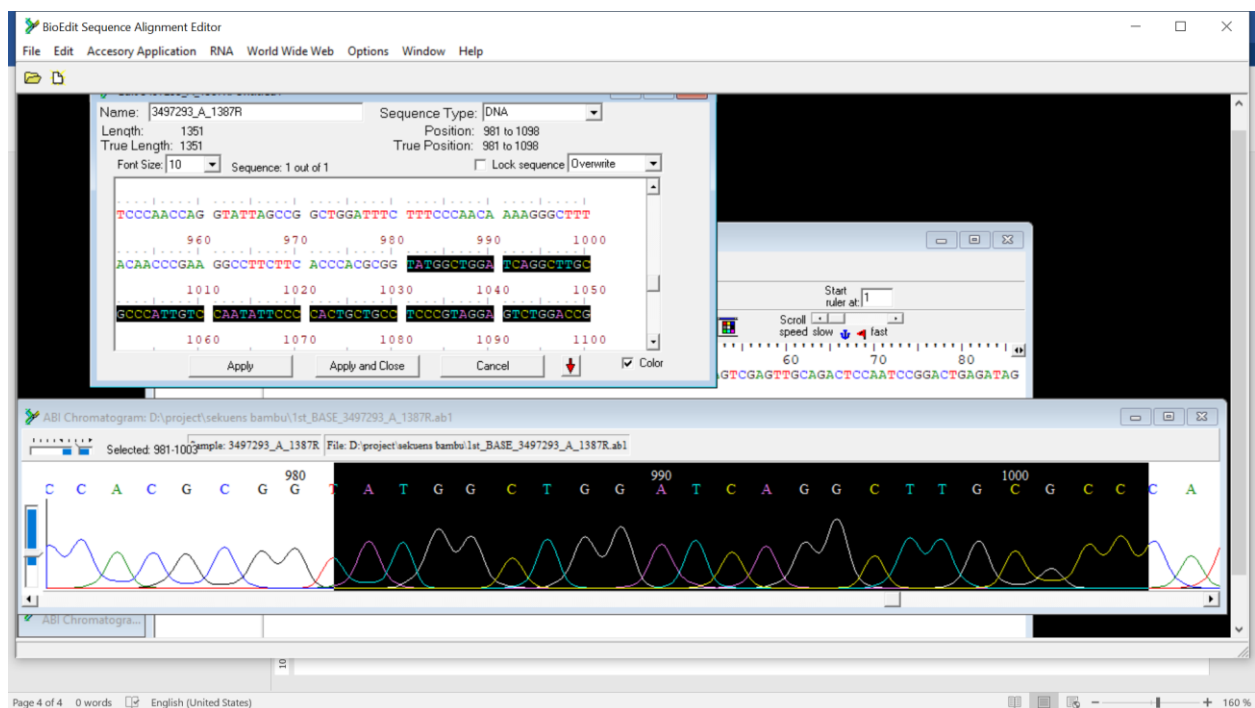
5. Bandingkan urutan sekuens dengan *chromatogram*.

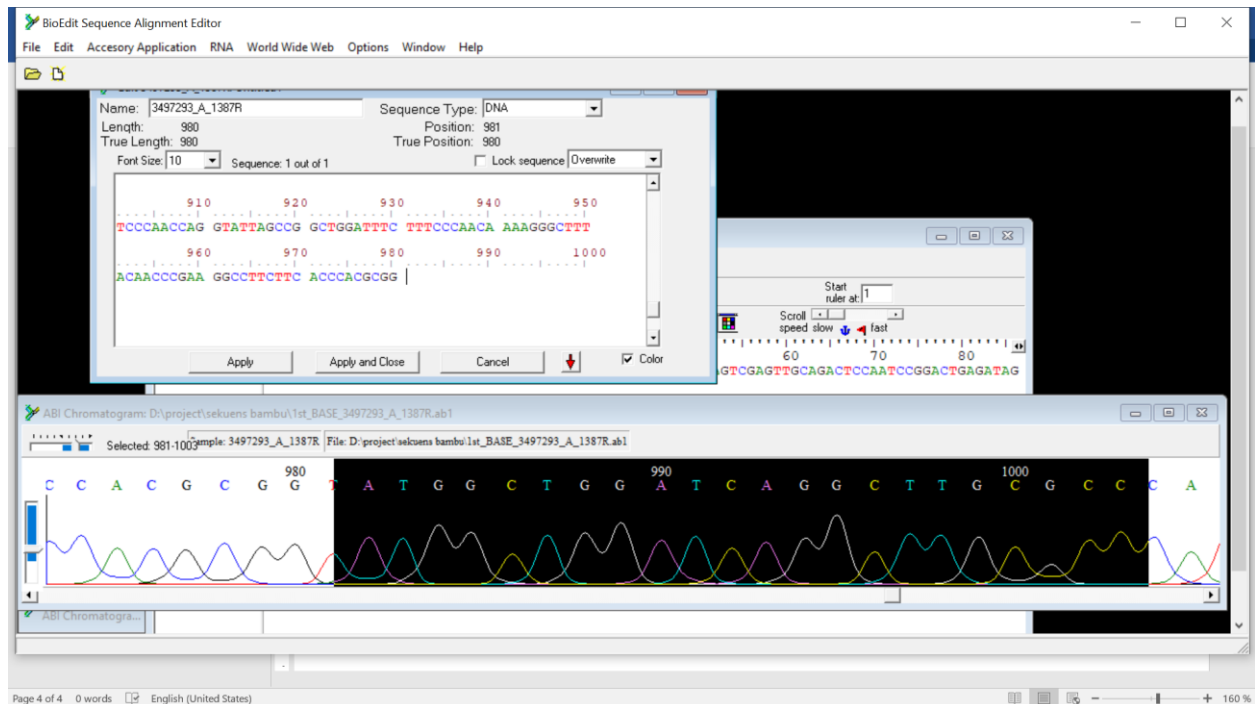


6. Pada umumnya, bagian ujung awal dan akhir sekuens memiliki *chromatogram* yang kurang baik sehingga perlu dihilangkan. Pada contoh, basa nomor 981 dan seterusnya perlu dihilangkan.

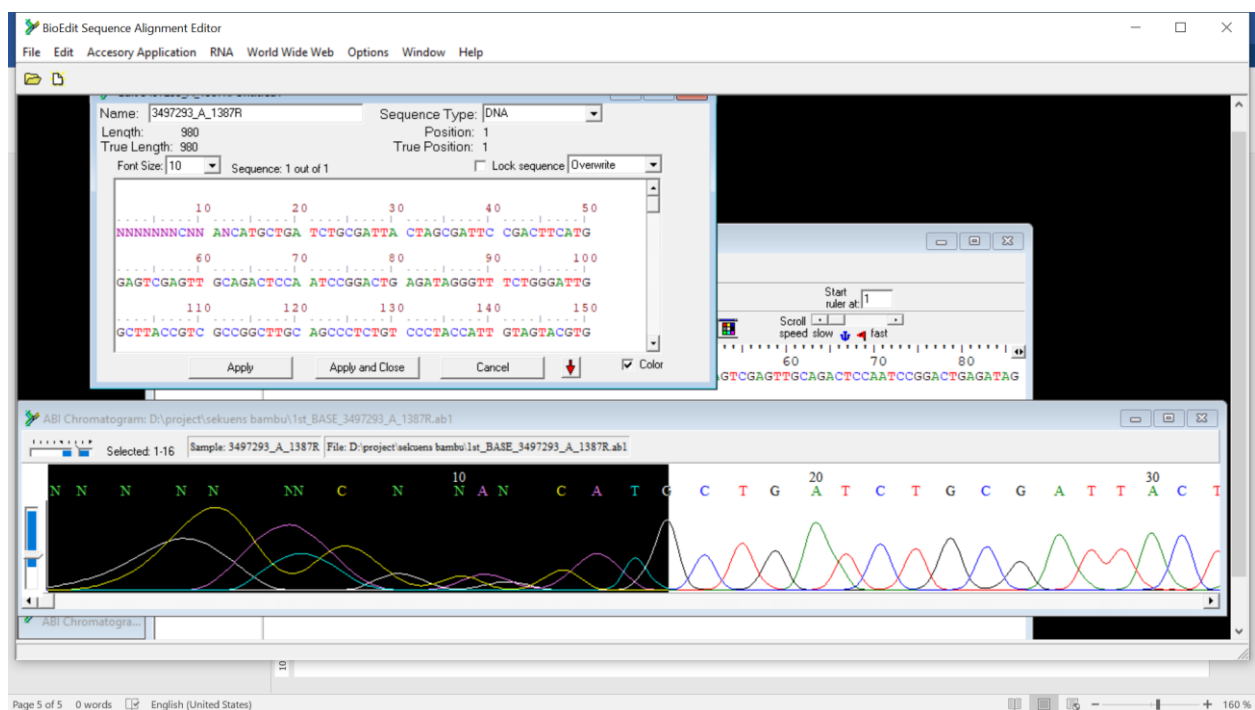


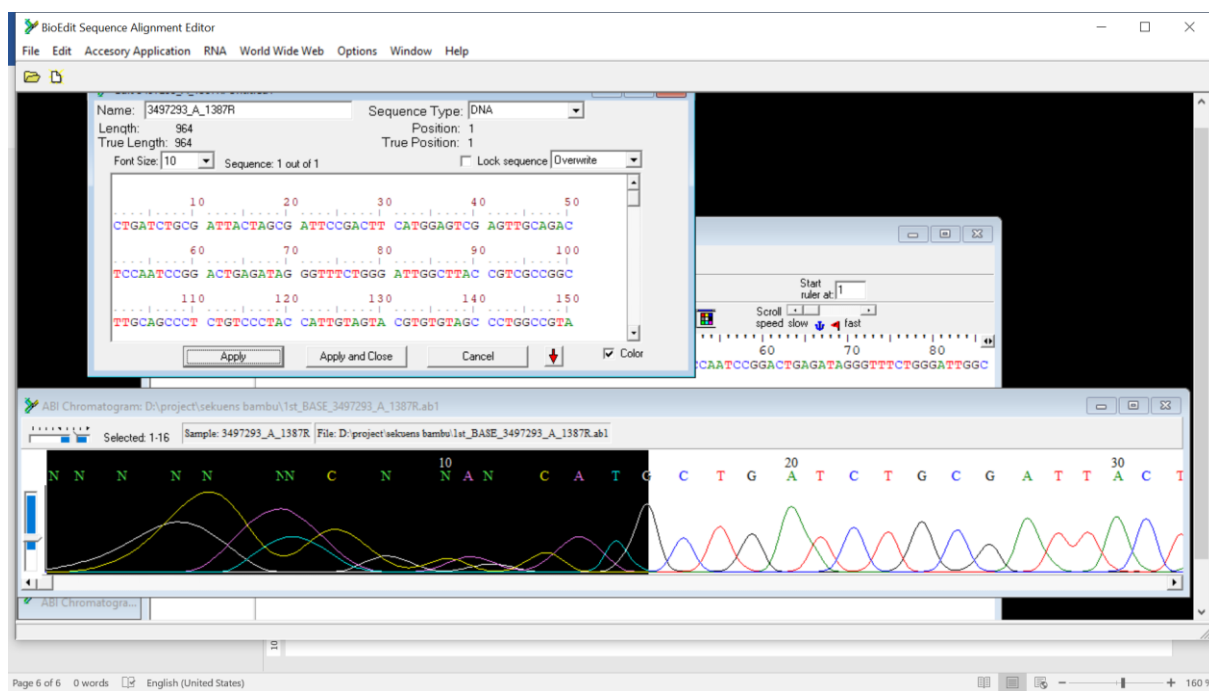
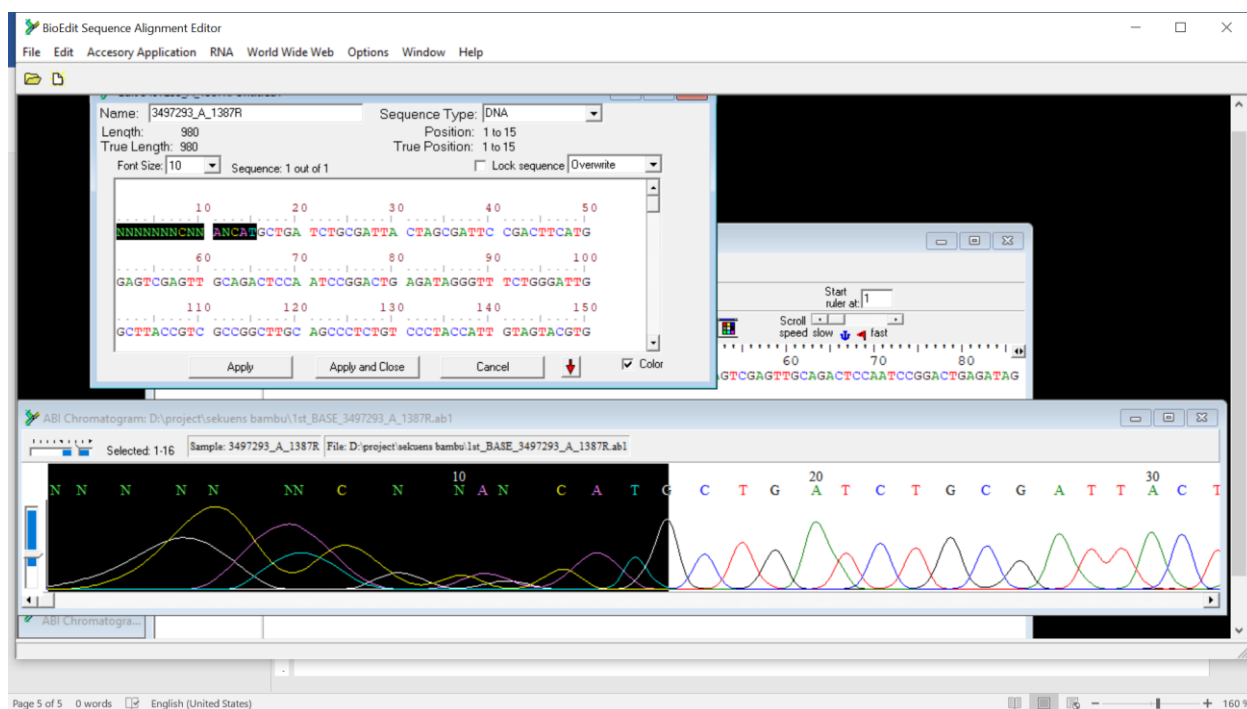
7. Pada *window edit sequence*, pilih basa nomor 981 dan seterusnya kemudian hapus (*delete*).



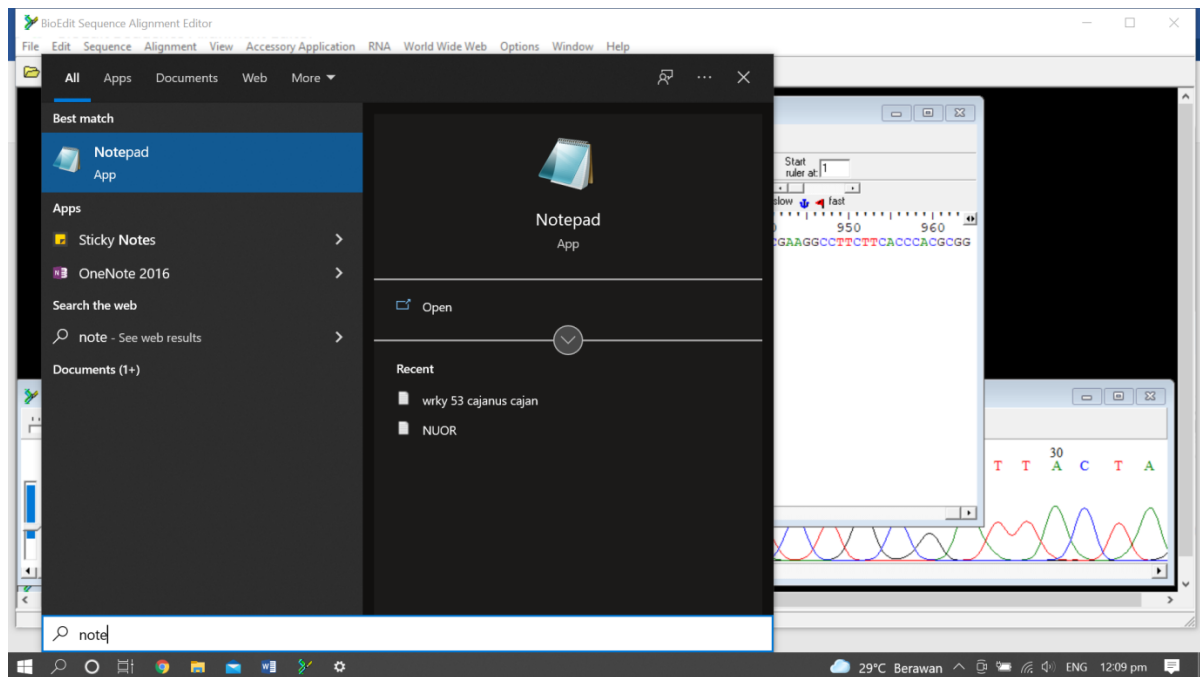
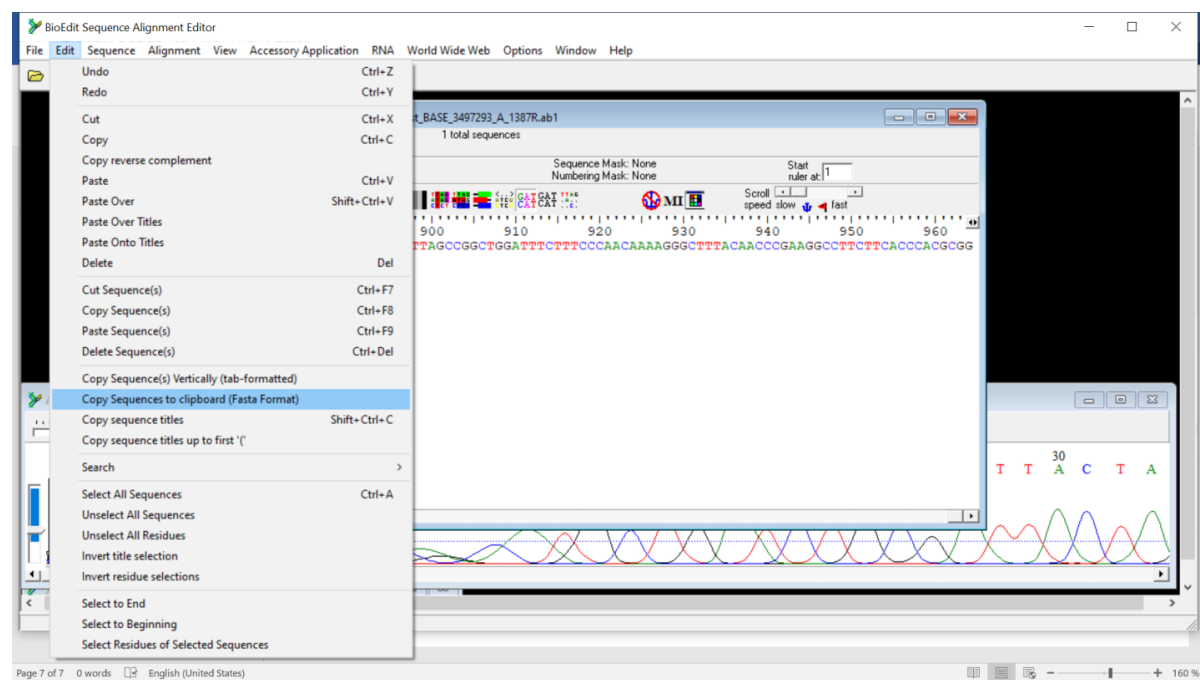


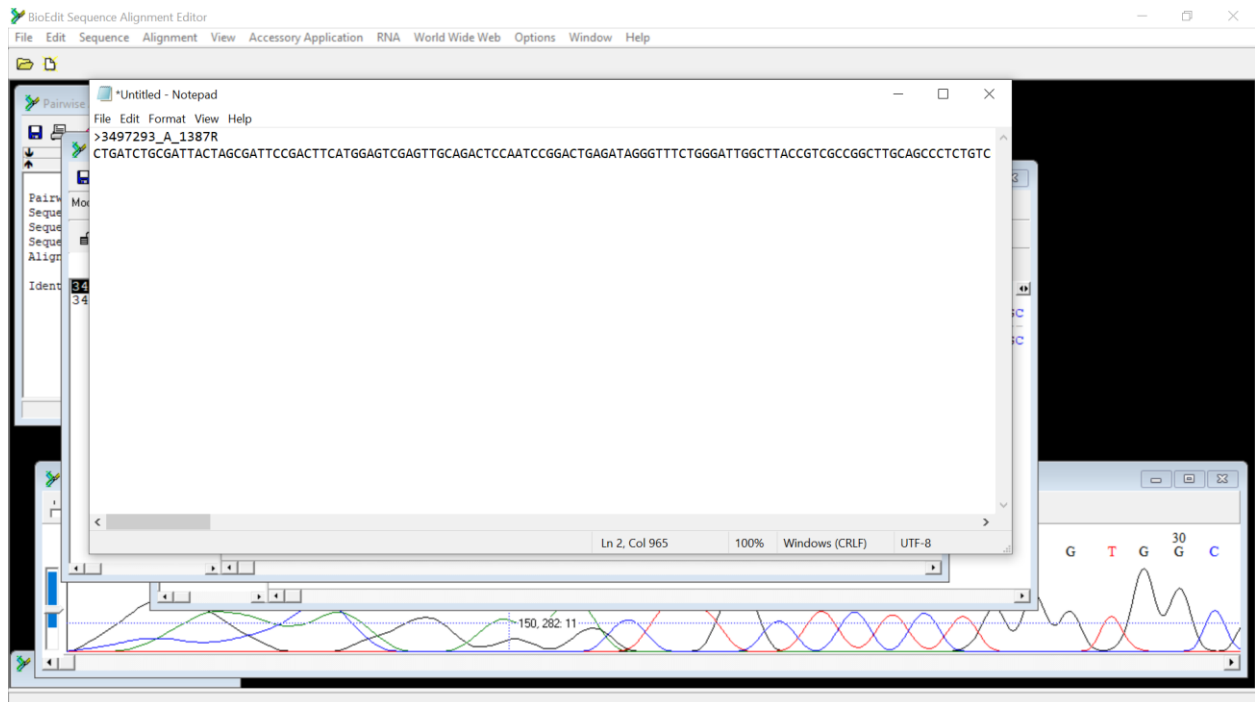
8. Pada bagian awal sekuens, bacaan *chromatogram* yang cukup baik dimulai dari basa nomor 16 sehingga basa 1-15 perlu dihilangkan/dihapus/delete.



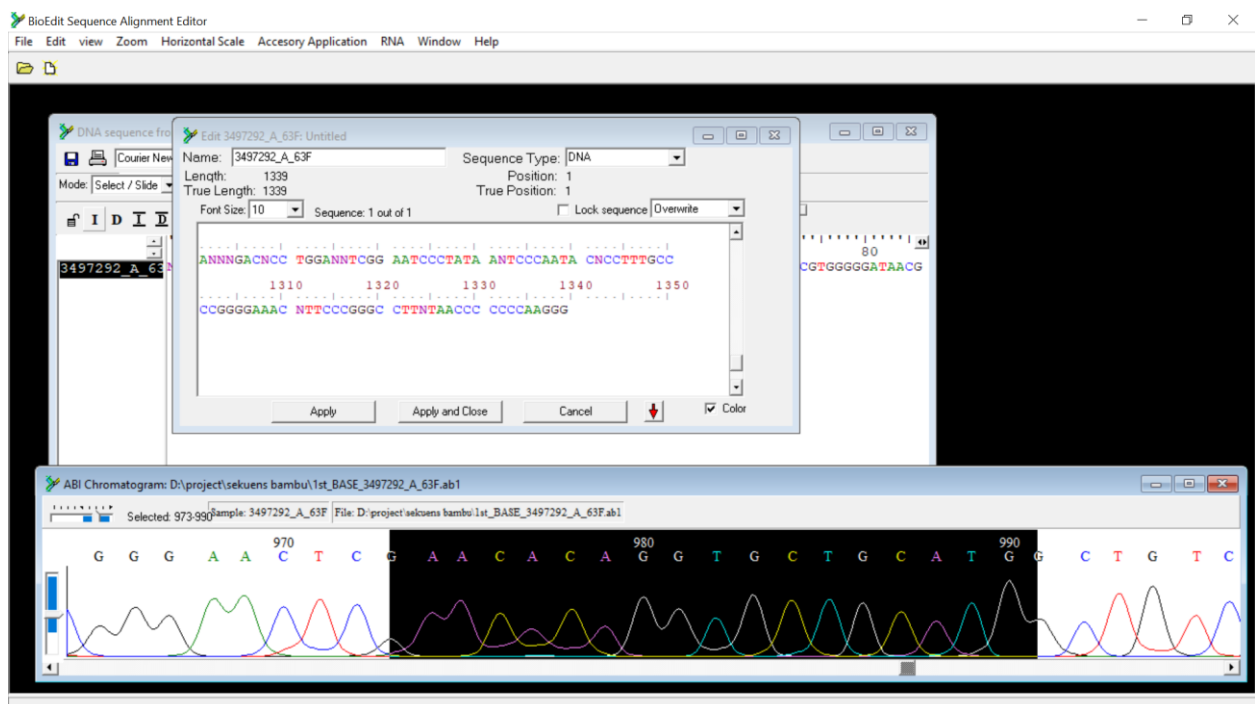


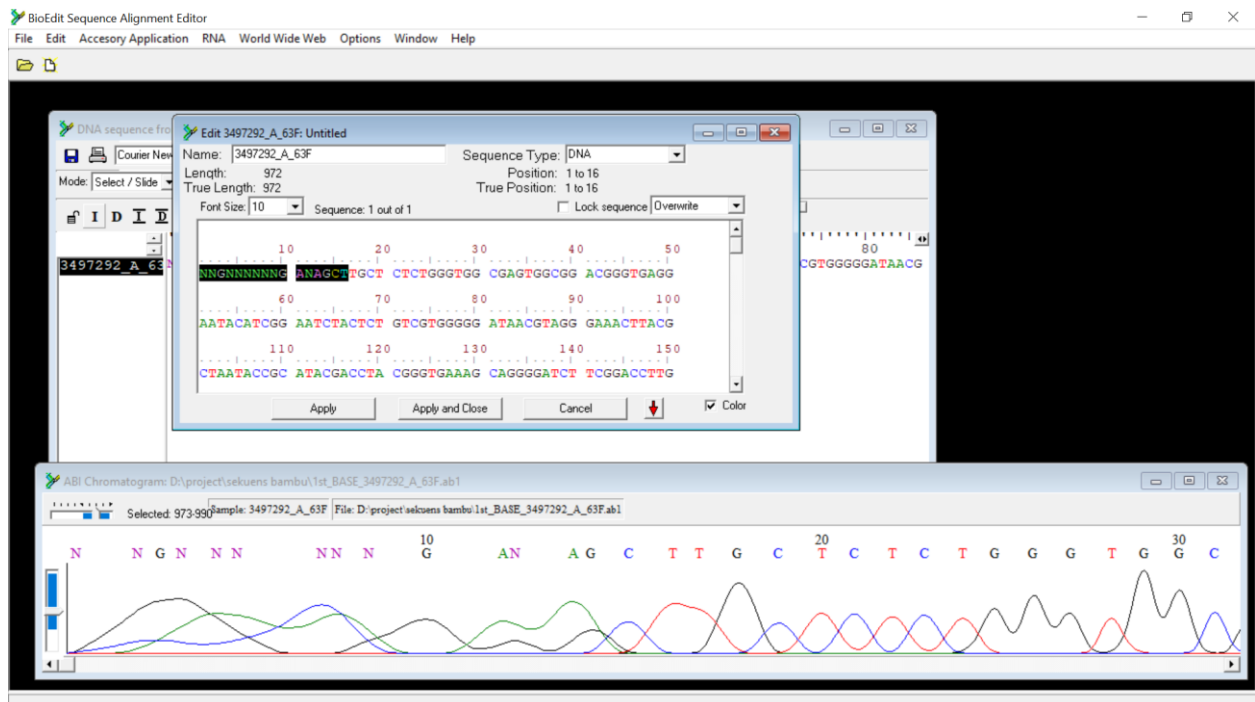
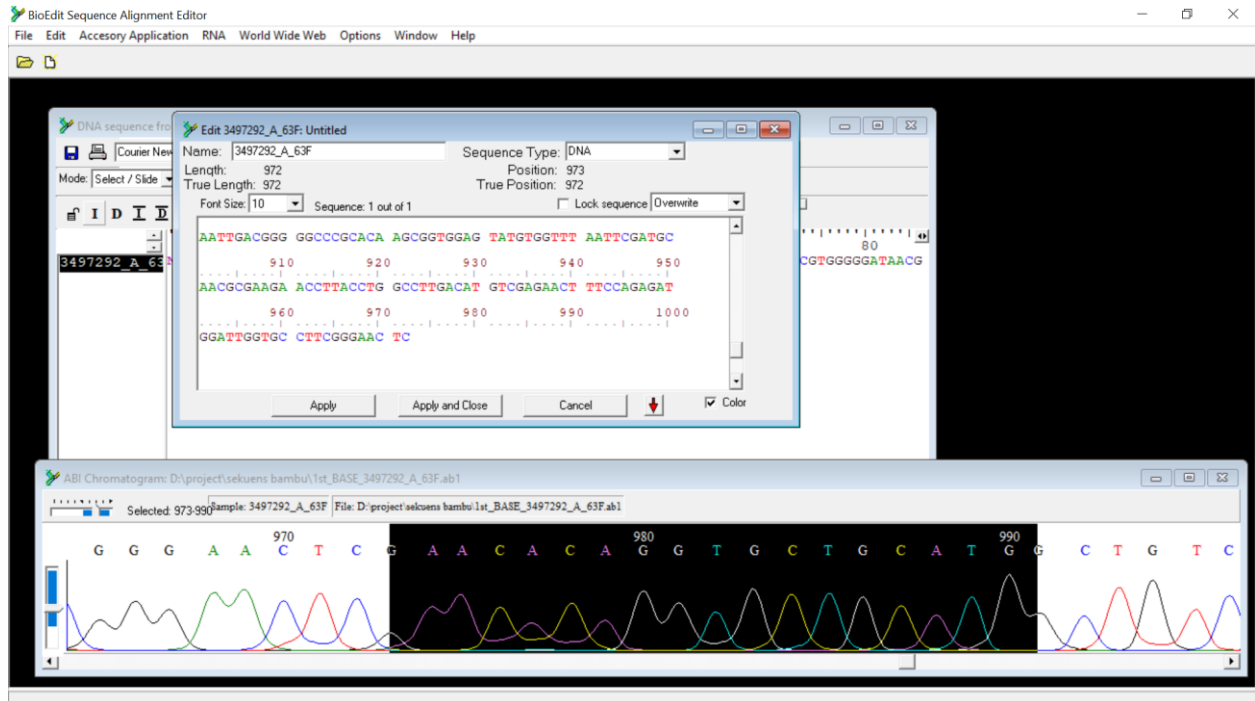
Pada saat *editing*, perlu dilihat apakah di tengah-tengah sekuens yang sudah dilakukan pemotongan pada ujung-ujungnya masih terdapat basa N. Jika masih terdapat basa N, maka perlu dilakukan verifikasi pada *chromatogram* untuk memastikan basa yang benar.

9. Buka aplikasi *Notepad*.10. Pilih sekuens, kemudian pada menu *edit* pilih *copy to clipboard fasta format*, kemudian *paste* pada *Notepad*.

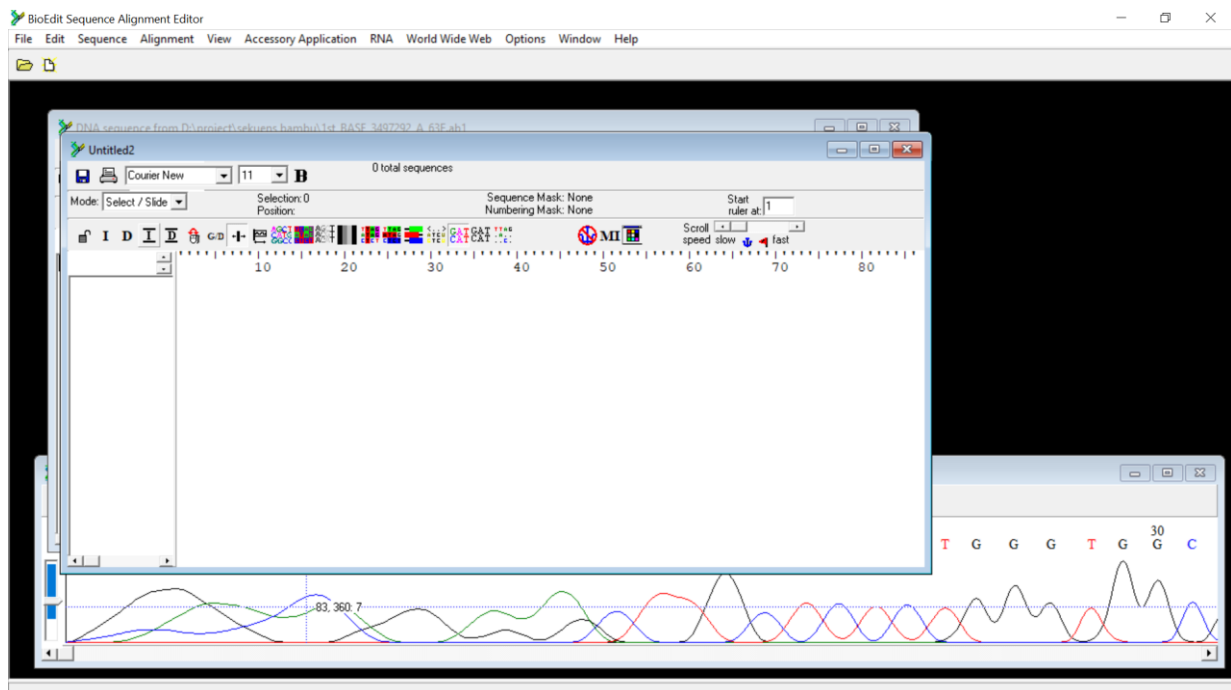


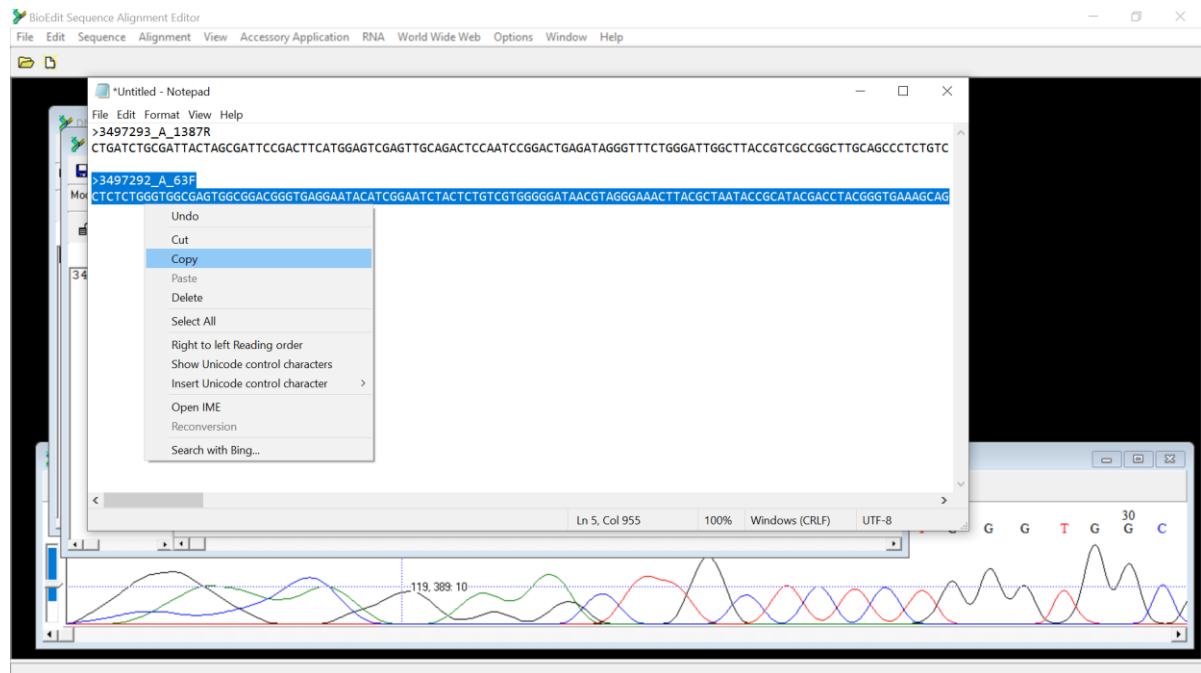
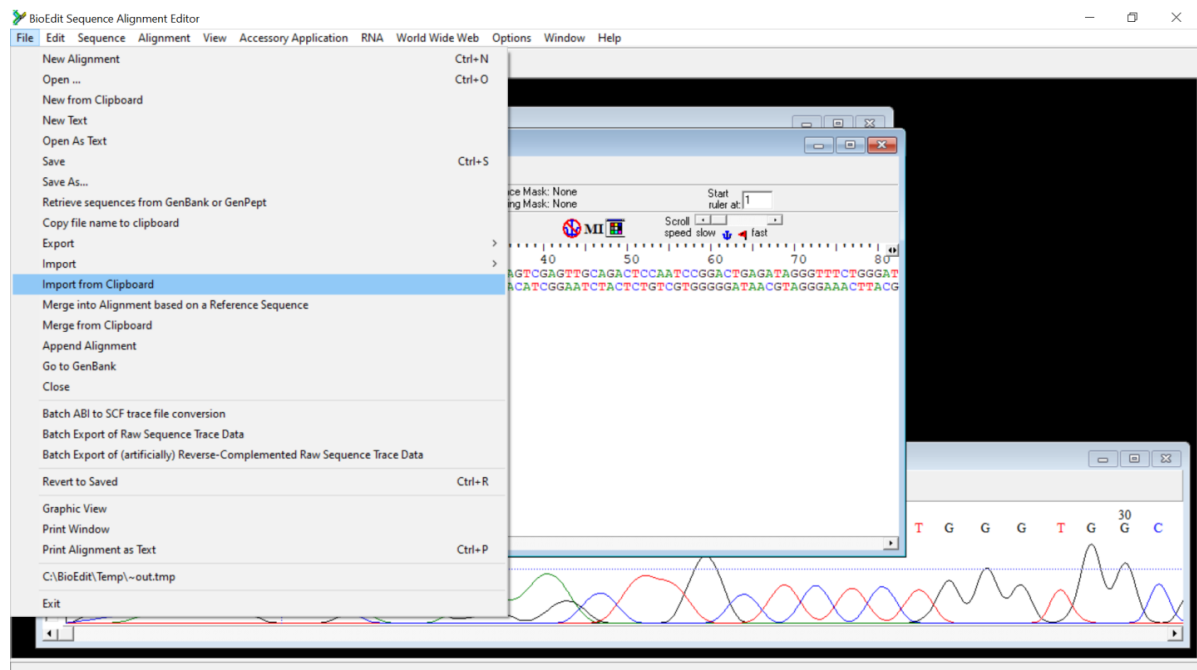
11. Lakukan proses *editing* yang sama pada sekuens pasangannya.



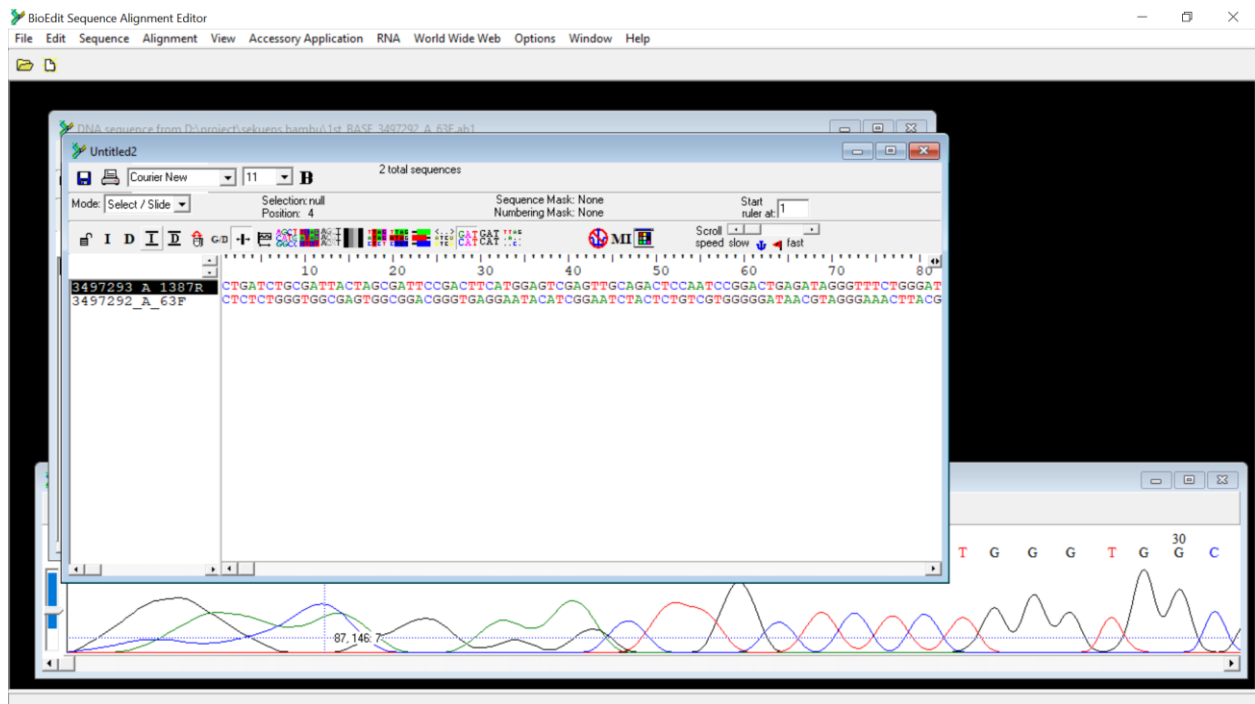


12. Klik ikon kertas putih (*new alignment*) untuk membuat *alignment* baru.

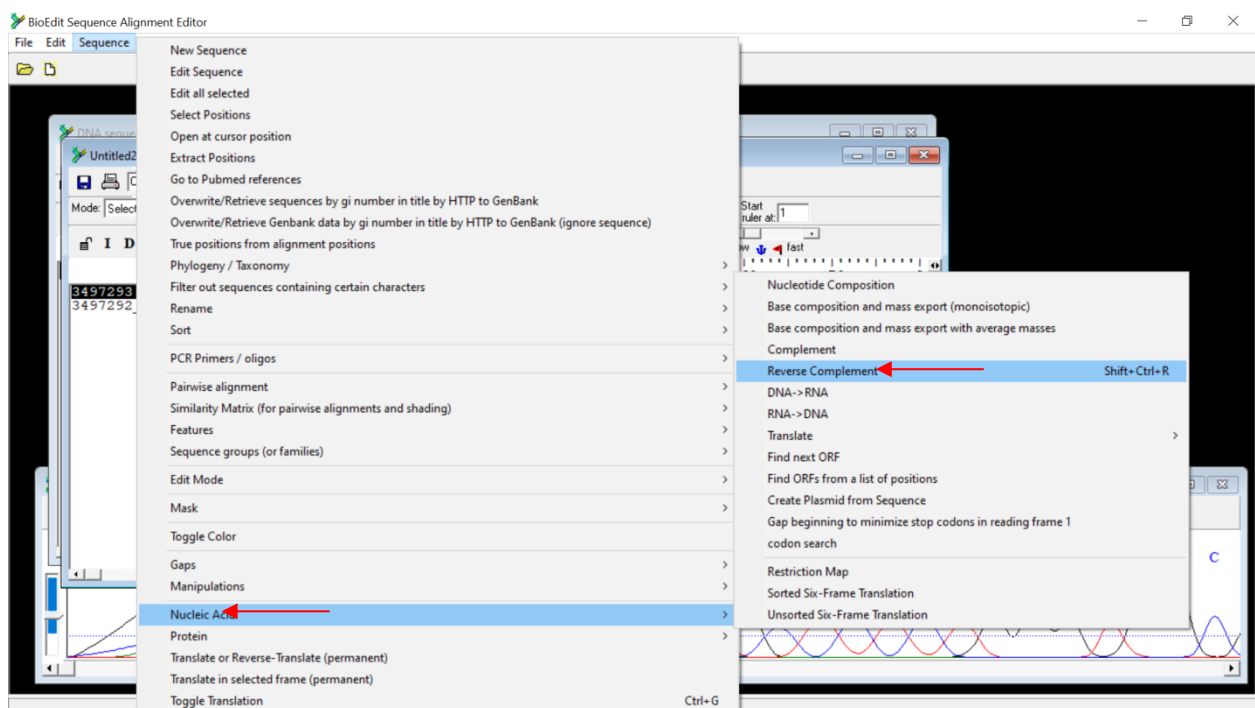


13. Pilih sekuens dengan cara meng-copy dari *Notepad*.14. Masukkan sekuens yang telah di-copy dengan cara *import from clipboard* pada menu *file*.

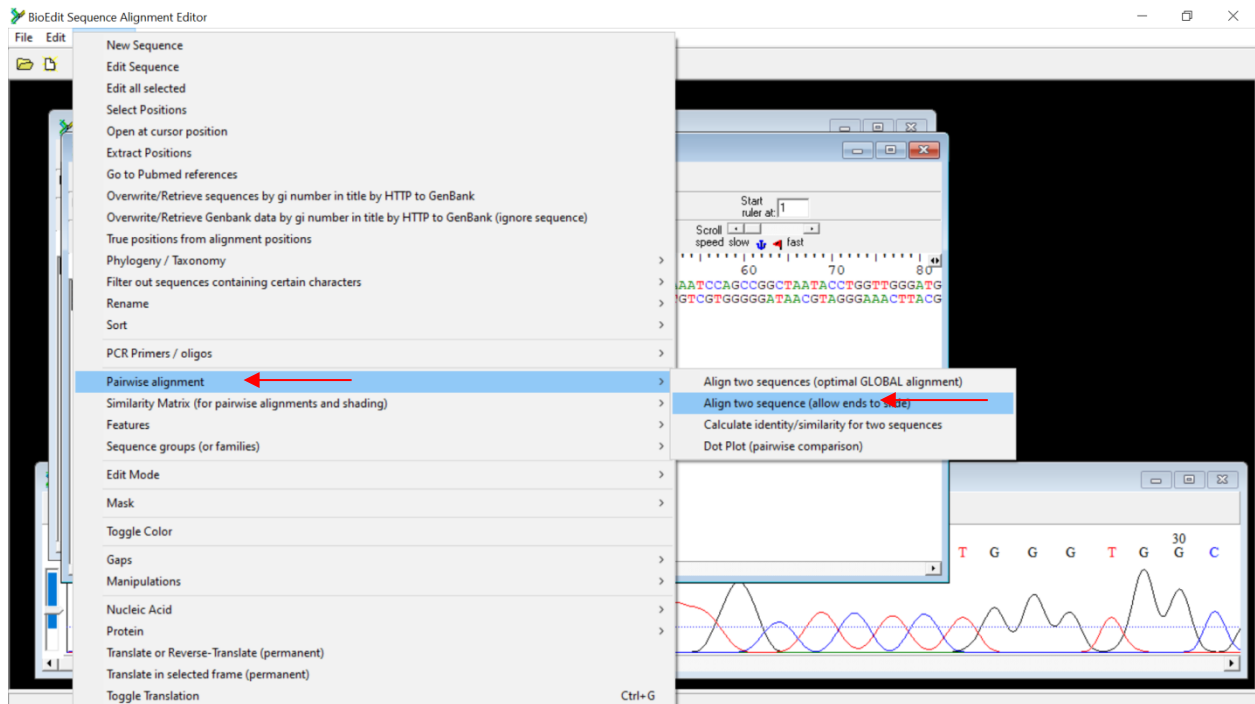
15. Pilih sekuens dengan pembacaan *reverse*.



16. Pada menu *sequence*, terdapat *tools nucleic acid*, kemudian pilih *reverse complement*.



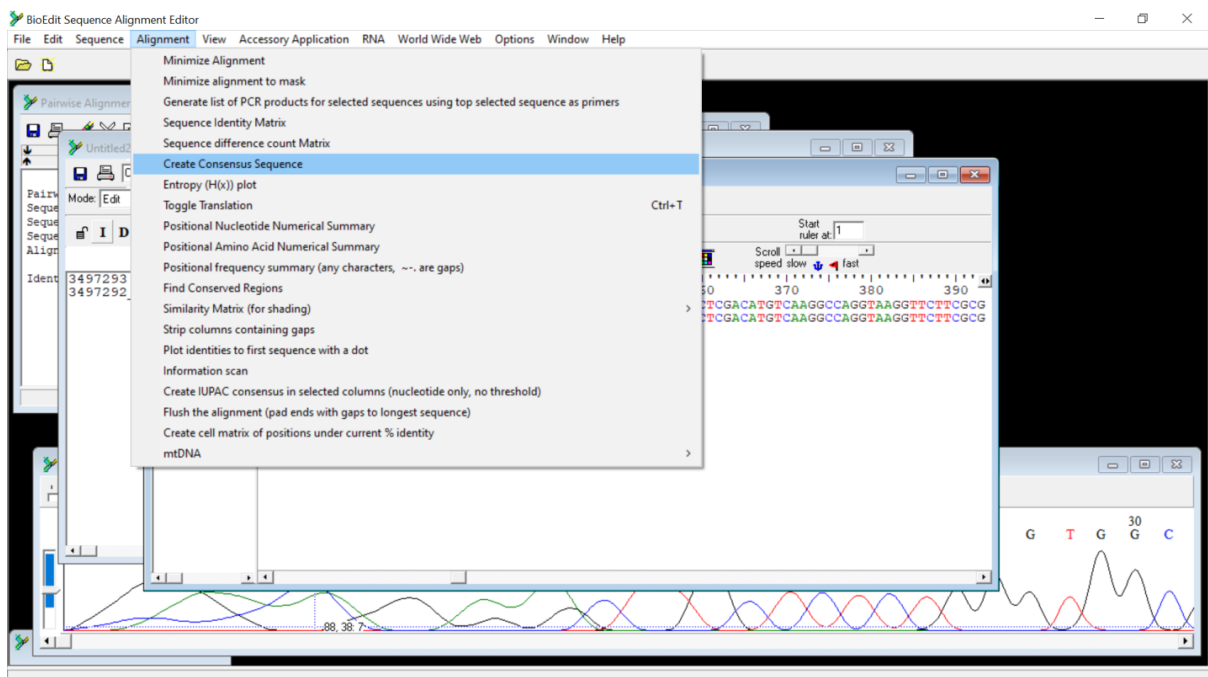
17. Untuk mensejajarkan sekuens *forward* dan *reverse*, pada menu sekuens, pilih *pairwise alignment*, kemudian pilih *align two sequence (allow end to slide)*.



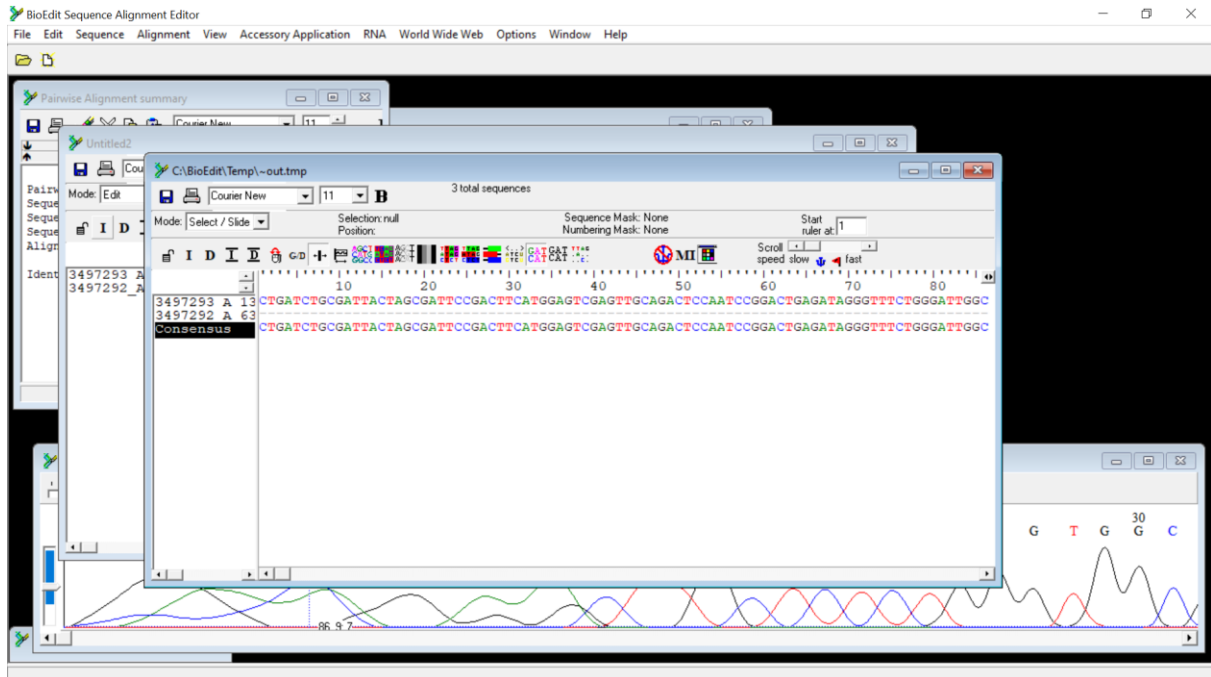
18. Untuk membuat sekuens *consensus* atau gabungan sekuens *forward* dan *reverse*, pilih kedua sekuens.



19. Pada menu *alignment* pilih *create consensus sequence*.



20. Hasil *consensus* sekuens akan muncul. Sekuens dapat disimpan di dalam format pasta dengan cara di-copy ke *Notepad* dengan fasta format. Ketika akan disimpan, pilih *All files* kemudian simpan file dengan format namafile.fasta. Sekuens siap untuk digunakan lebih lanjut.

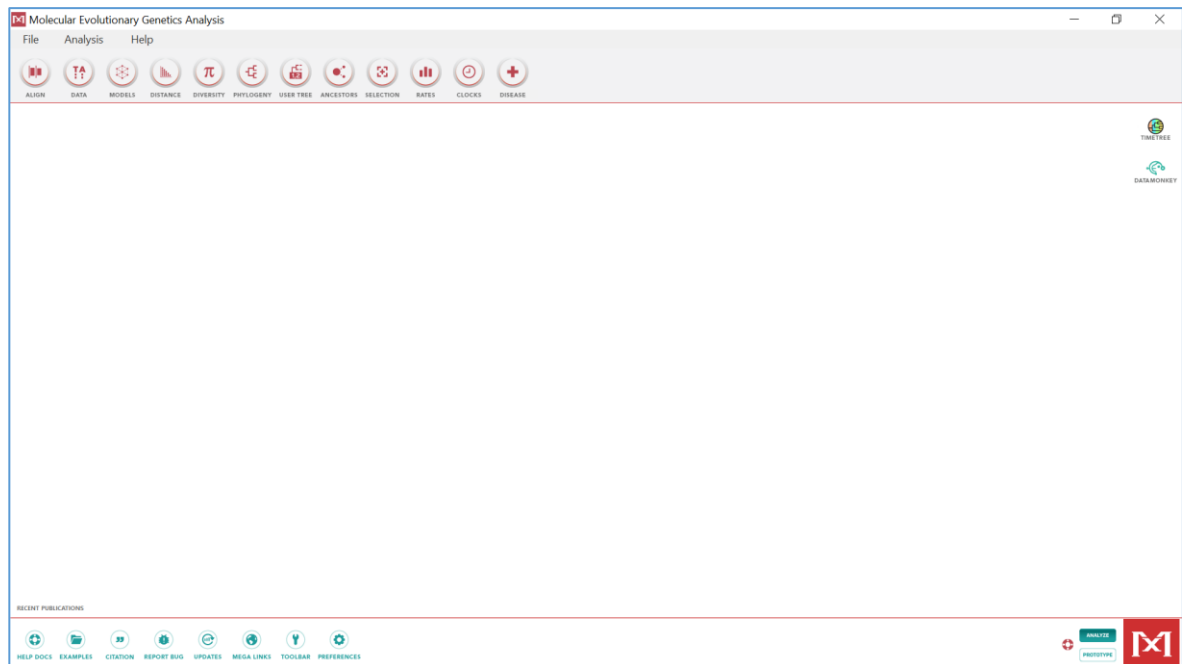


4.2 Pembuatan *alignment* multi sekuens, analisis kekerabatan, dan pembuatan pohon filogenetik dari sekuens menggunakan *software* MEGA11.

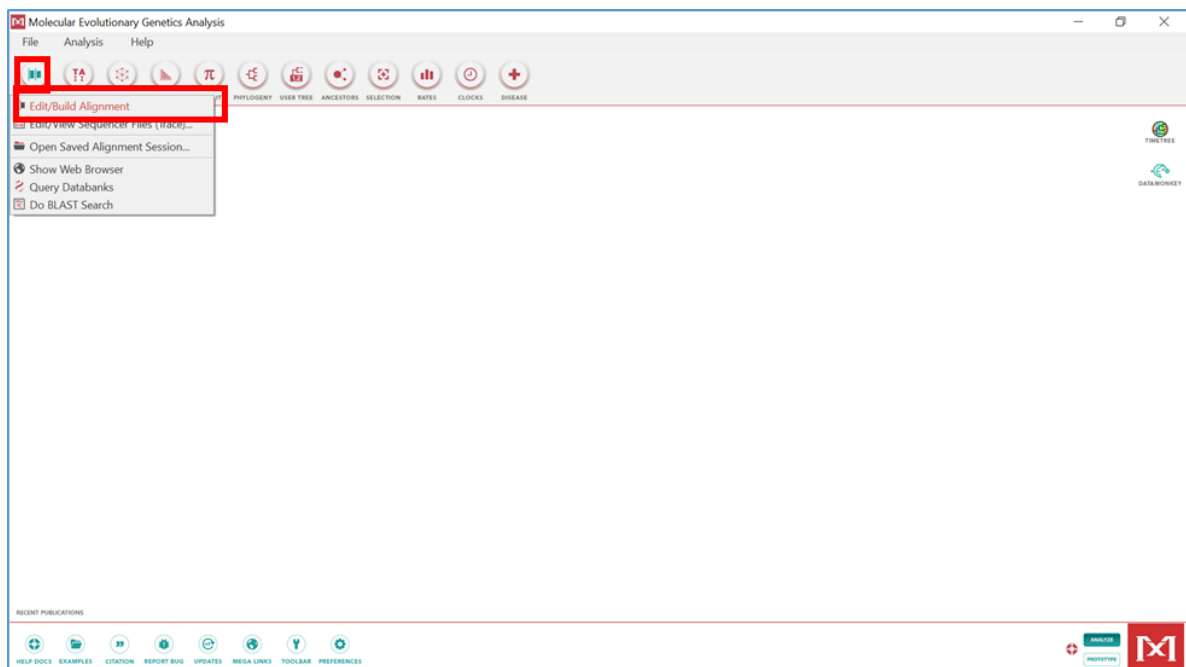
MEGA (*Molecular and Evolutionary Genetic Analysis*) adalah *software* yang dirancang untuk melakukan analisis komparatif antar sekuens dengan penekanan khusus untuk menganalisis hubungan evolusi (Kumar *et al.* 2008).

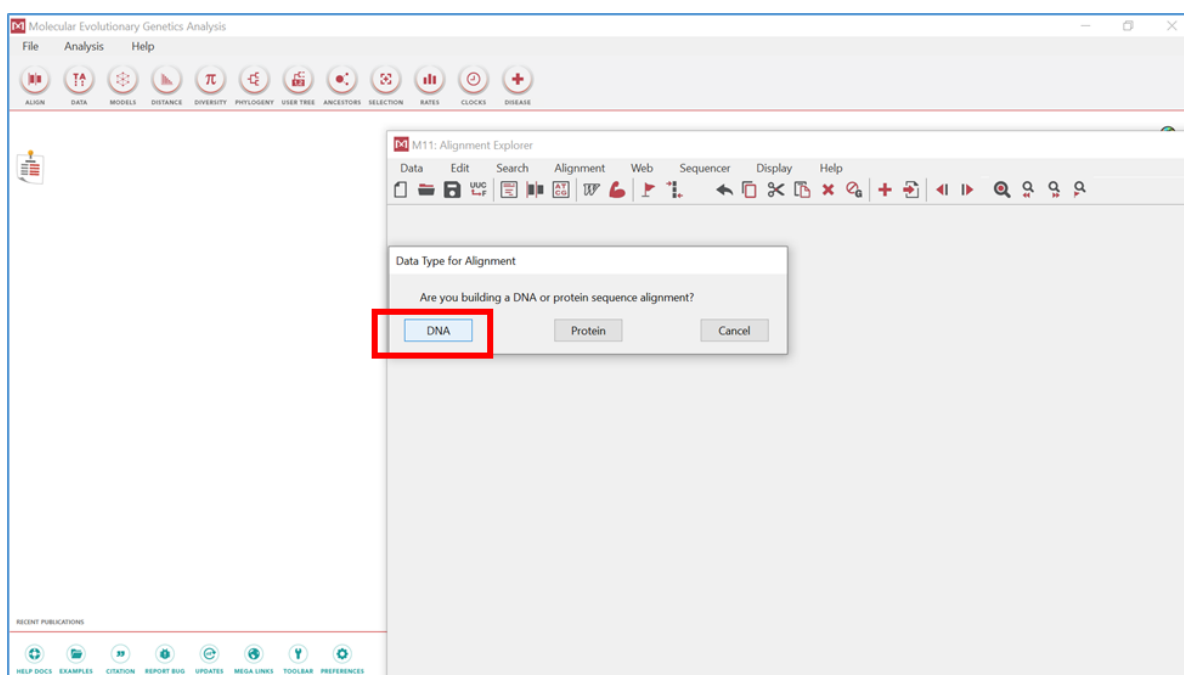
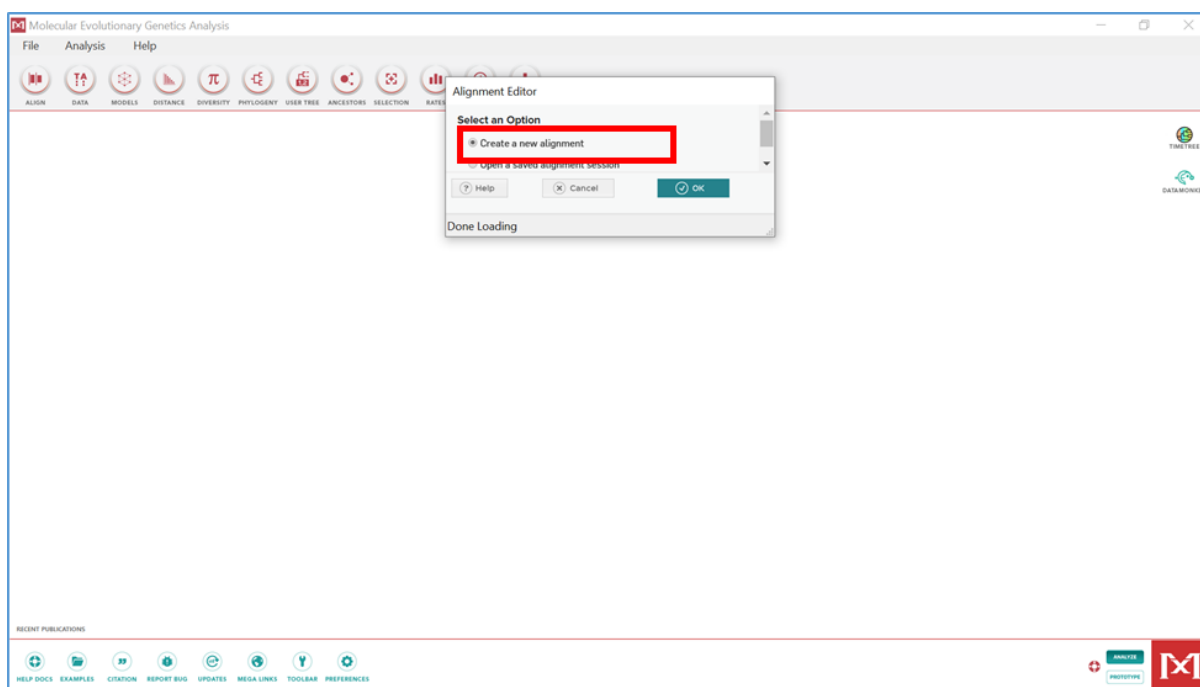
Langkah-langkah untuk melaksanakan analisis dengan *software* MEGA11 adalah sebagai berikut:

1. Buka *software* MEGA 11.

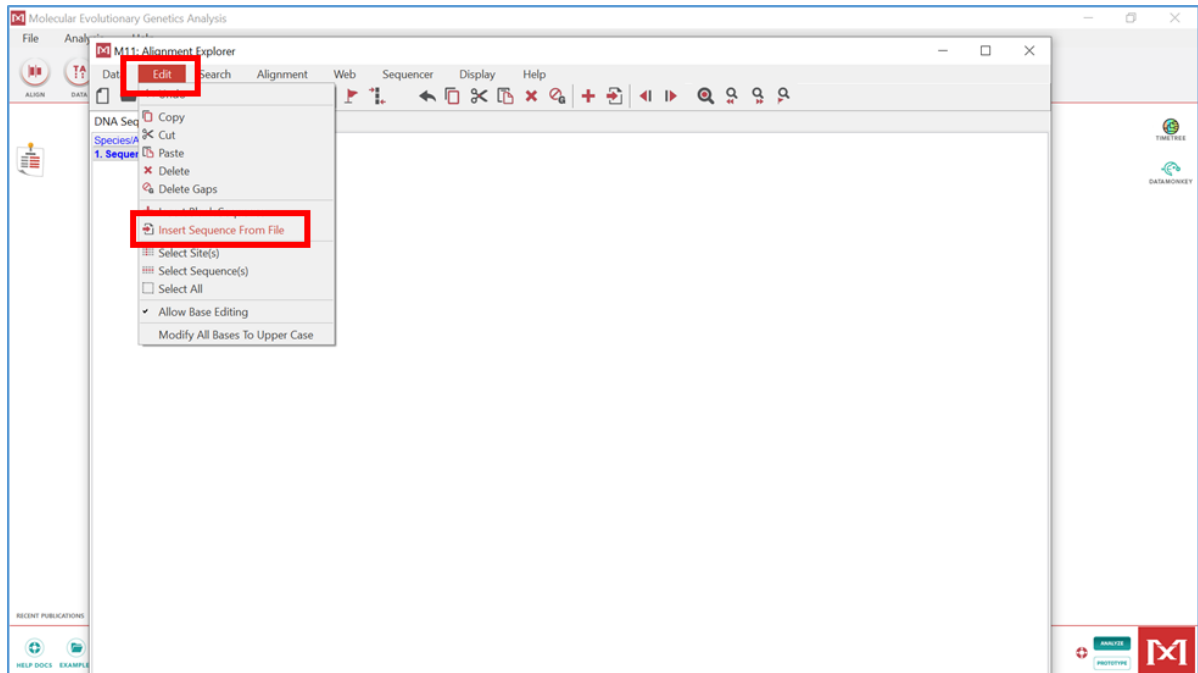


2. Pilih *edit/build alignment*, kemudian pilih *create new alignment*. Di dalam contoh kali ini, sekuens yang digunakan adalah DNA. Oleh sebab itu, pilih DNA.

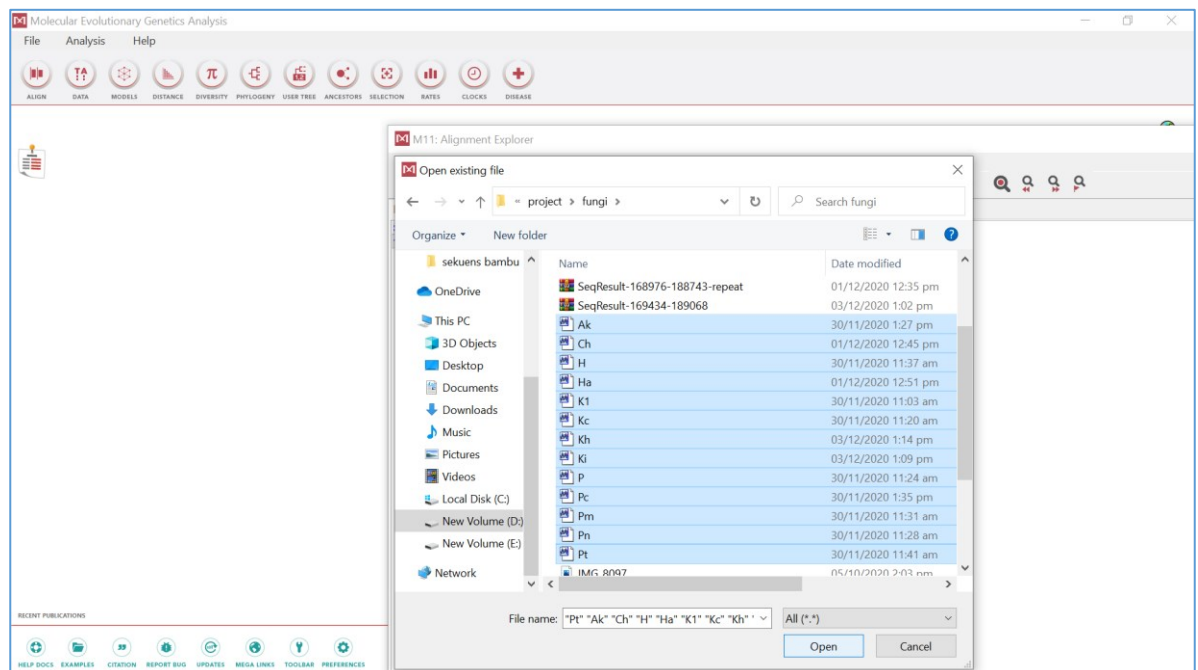


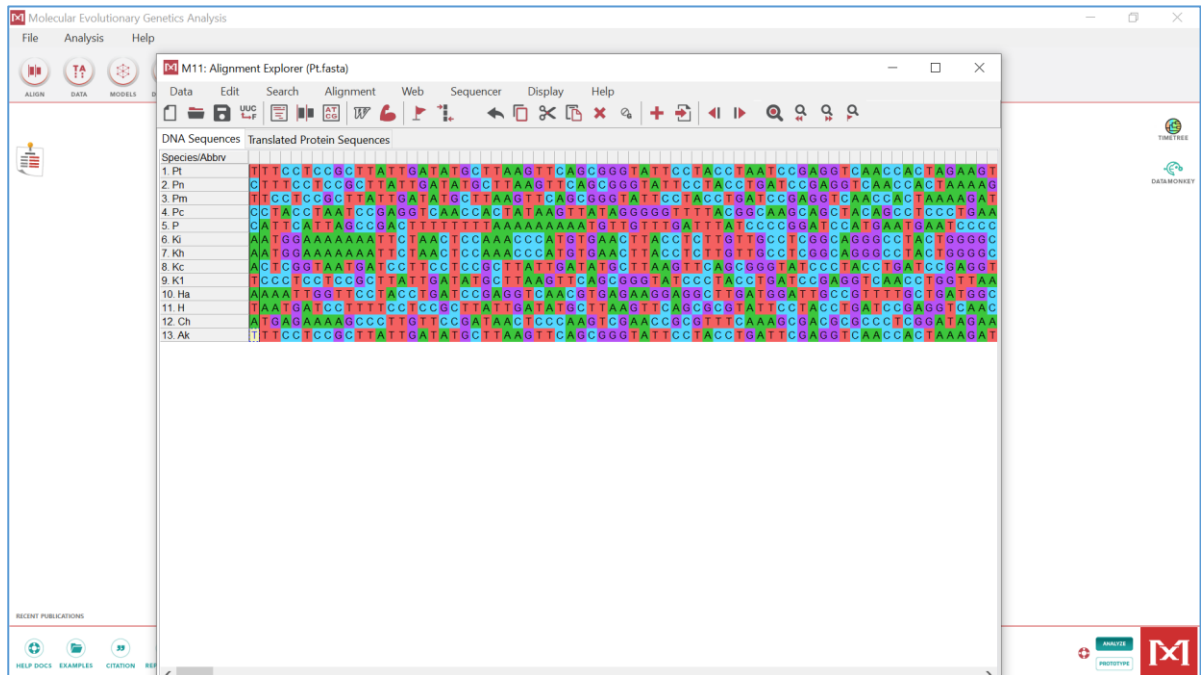


- Untuk memasukkan sekuens yang akan digunakan, pada menu edit di *window alignment*, pilih *insert sequence from file*.

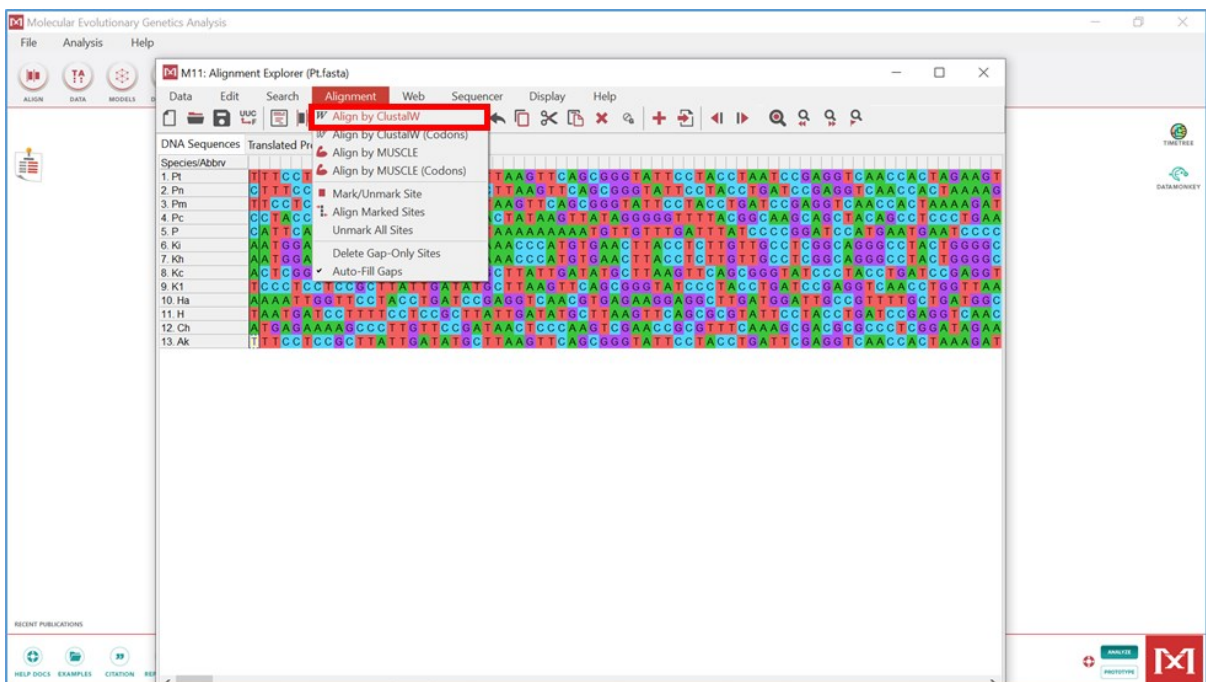


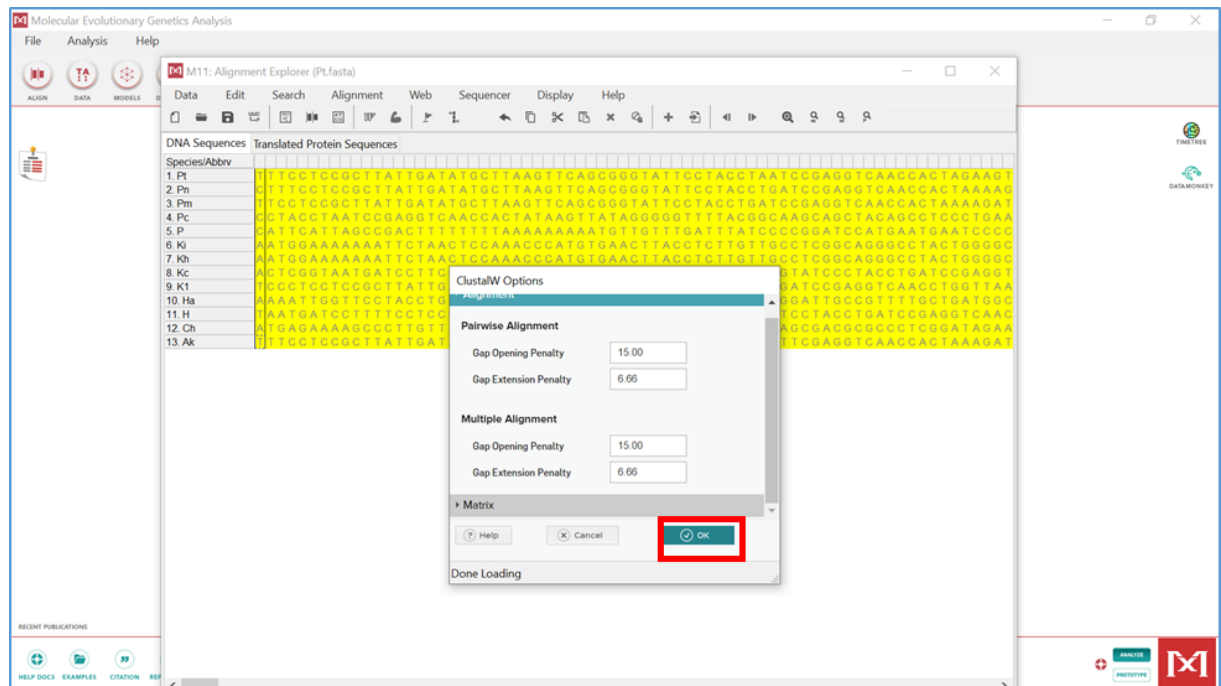
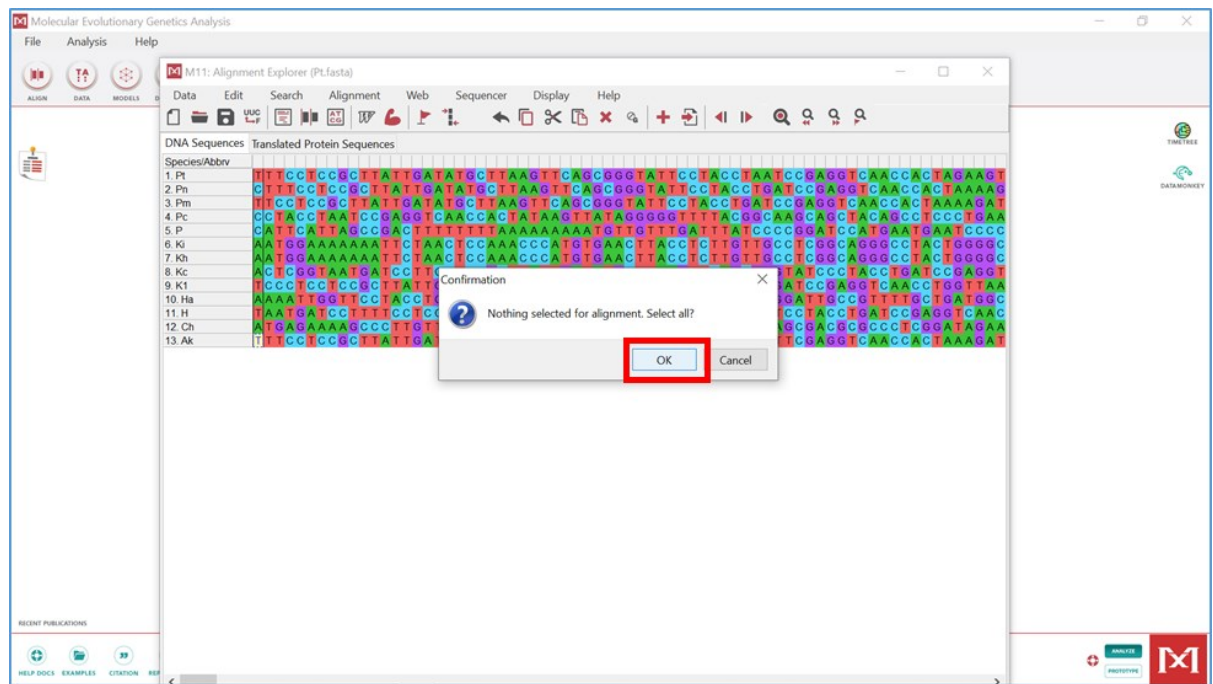
- Pilih sekuens yang akan disejajarkan. Kemudian, setelah seluruh sekuens terunggah (*upload*), maka *delete* sekuens kosong (*sequence1*) sehingga tersisa sekuens yang akan kita masukkan.



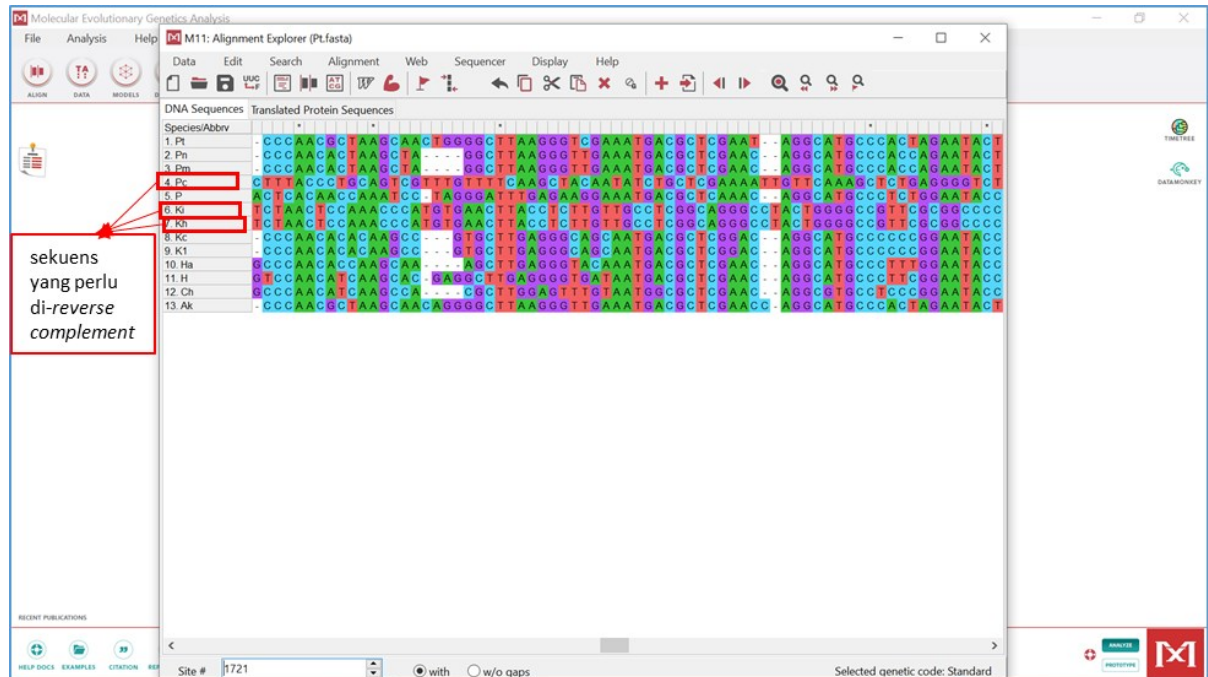


5. Pada menu *alignment*, pilih *align by clustal W*.

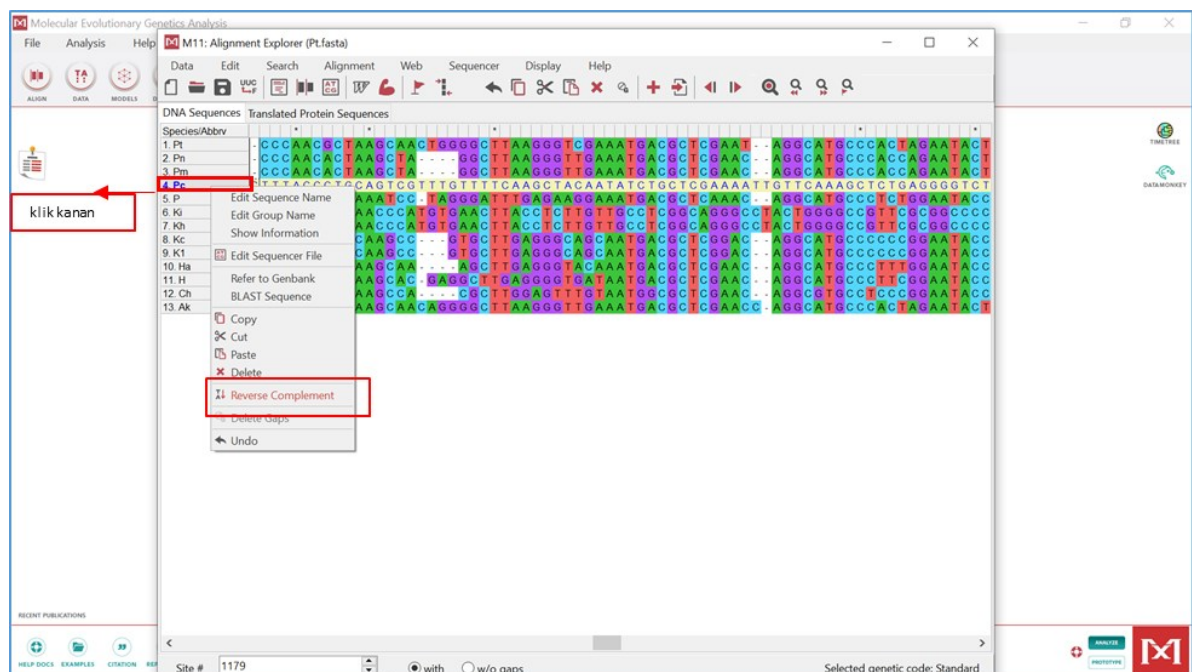




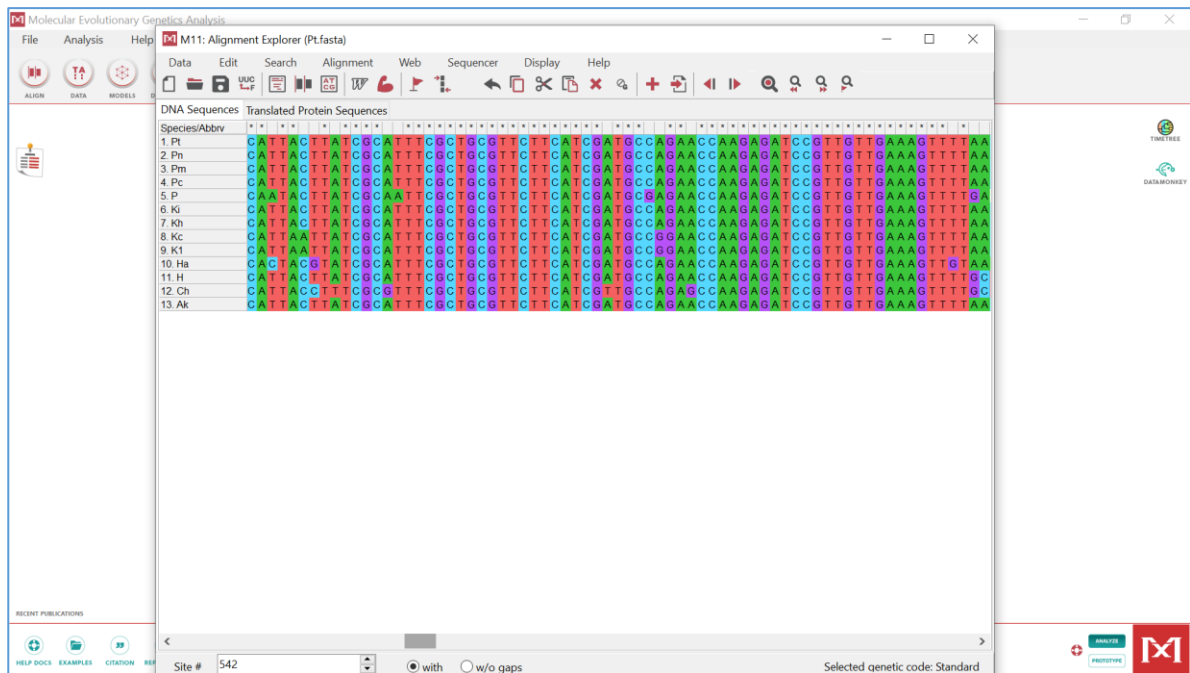
6. Hasil *alignment* menunjukkan terdapatnya beberapa sekuens yang dibaca dengan arah yang berbeda dari sekuens lainnya (Pc, Ki, Kh). Sekuens tersebut perlu di-*reverse complement* terlebih dahulu, untuk kemudian dilakukan pensejajaran ulang.



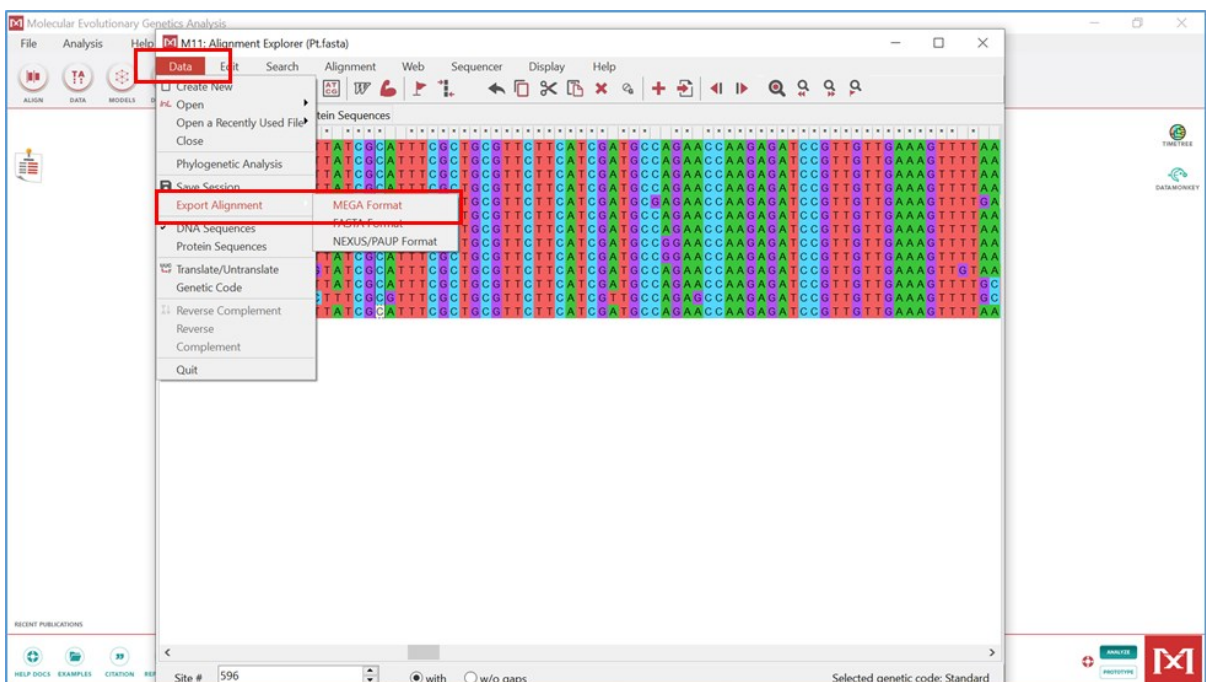
7. Klik kanan pada nama sekuens, kemudian pilih *reverse complement*. Jika sudah, lakukan *alignment* kembali.



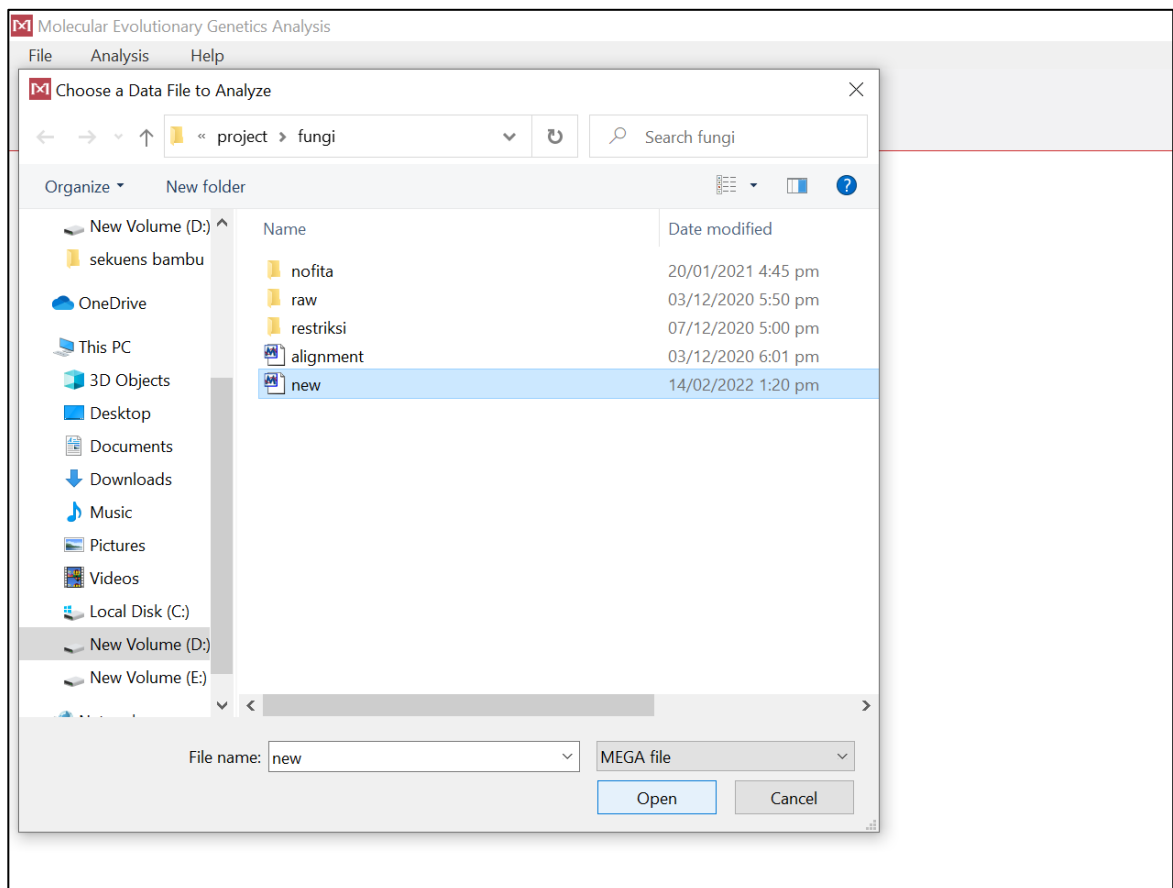
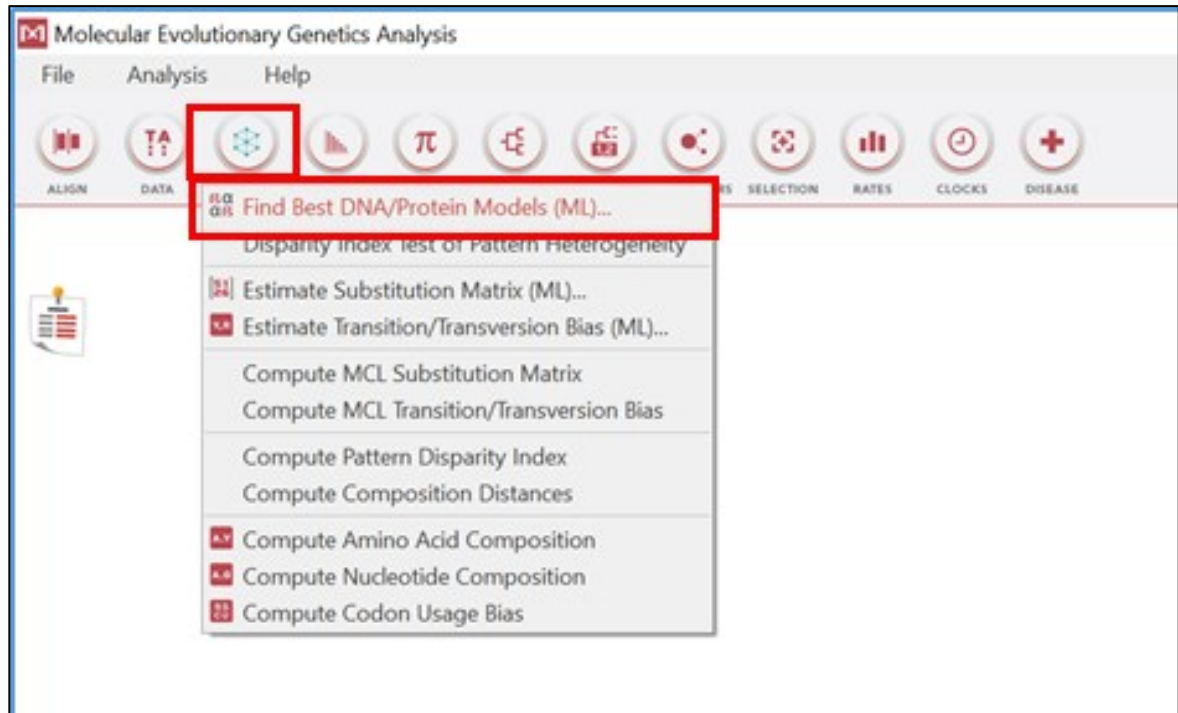
8. Hasil *alignment* sudah benar jika terdapat *region* atau bagian dari seluruh sekuens yang telah disejajarkan dan hasilnya benar-benar sejajar. Sekuens yang disejajarkan pada contoh tidak akan sama 100% karena berasal dari spesies yang berbeda.

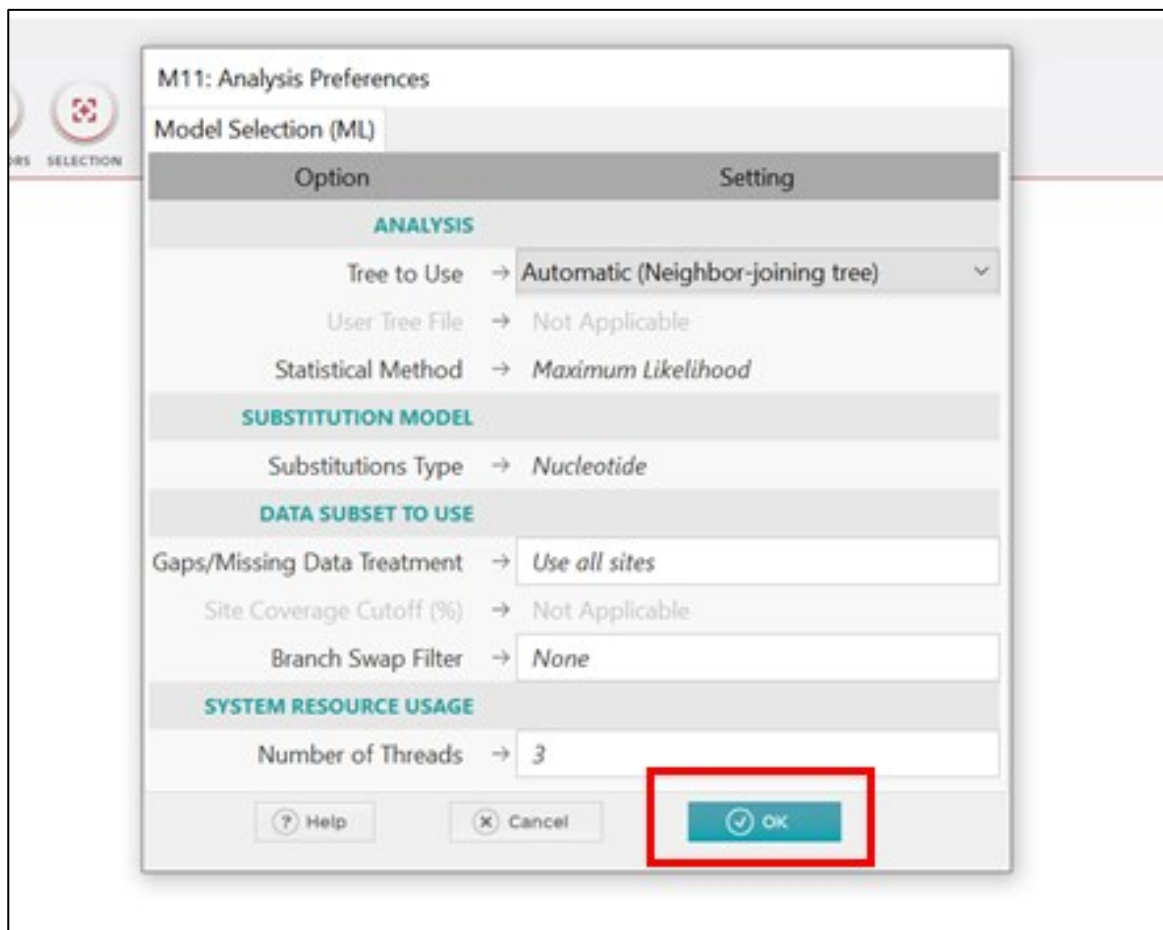


9. *Export file alignment* ke mega format melalui menu data.



10. Untuk mengetahui model terbaik yang akan digunakan dalam pembuatan pohon filogenetik, maka pada *window* utama, pilih menu *models*, kemudian pilih *find best model*. Pilih file *alignment* yang sudah disimpan sebelumnya. Pada pilihan metode *default*, pilih ok.





11. Pada window hasil model terdapat beberapa model yang ditampilkan. Dari sekian tampilan model, maka model yang terbaik adalah model yang teratas. Pada contoh yaitu K2+G+I. model terbaik dipilih dengan mempertimbangkan jumlah parameter, *Bayesian Information Criterion* (BIC), dan *Akaike Information Criterion corrected* (AICc).

MEGA Caption Expert: Find Best-Fit Substitution Model (ML)

FileEditViewHelp

Results

Table: Maximum Likelihood fits of 24 different nucleotide substitution models

Model	Parameters	BIC	AICc	lnL	(+G)	R	(+I)	(+G)	(+I)	(+G)	(+I)	(+G)	(+I)	(+G)	(+I)	(+G)	(+I)	(+G)	(+I)	(+G)	(+I)	(+G)	(+I)	(+G)	(+I)
K2+G+I	26	20653.679	20463.521	-10205.897	0.14	3.00	1.33	0.250	0.250	0.250	0.250	0.054	0.054	0.143	0.054	0.143	0.054	0.054	0.143	0.054	0.143	0.054	0.054	0.143	0.054
TN2+G+I	27	20655.573	20458.105	-10201.984	0.14	2.98	1.33	0.258	0.258	0.242	0.242	0.055	0.052	0.139	0.055	0.139	0.052	0.055	0.147	0.052	0.147	0.055	0.052	0.147	0.055
HKY+G+I	29	20673.833	20461.749	-10201.796	0.14	3.01	1.33	0.262	0.253	0.240	0.244	0.054	0.052	0.139	0.056	0.137	0.052	0.056	0.145	0.052	0.149	0.054	0.052	0.149	0.054
K2+I	25	20682.522	20499.672	-10224.778	0.15	n/a	1.18	0.250	0.250	0.250	0.250	0.057	0.057	0.135	0.057	0.135	0.057	0.135	0.057	0.135	0.057	0.135	0.057	0.135	0.057
TN2+G	26	20684.369	20494.210	-10221.042	0.15	n/a	1.18	0.258	0.258	0.242	0.242	0.059	0.056	0.131	0.059	0.131	0.056	0.059	0.140	0.056	0.140	0.059	0.056	0.140	0.059
K2+G	25	20692.674	20509.824	-10229.854	n/a	0.82	1.44	0.250	0.250	0.250	0.250	0.051	0.051	0.147	0.051	0.147	0.051	0.147	0.051	0.147	0.051	0.147	0.051	0.147	0.051
TN2+G	26	20694.530	20504.371	-10226.122	n/a	0.82	1.44	0.258	0.258	0.242	0.242	0.053	0.050	0.143	0.053	0.143	0.050	0.053	0.152	0.050	0.152	0.053	0.152	0.053	0.152
TN93+G+I	30	20695.122	20475.730	-10207.781	0.14	3.12	1.12	0.262	0.253	0.240	0.244	0.060	0.057	0.151	0.062	0.104	0.058	0.062	0.110	0.058	0.162	0.060	0.057	0.162	0.060
HKY+I	28	20702.744	20497.968	-10220.911	0.15	n/a	1.18	0.262	0.253	0.240	0.244	0.058	0.055	0.132	0.060	0.130	0.056	0.060	0.137	0.056	0.142	0.058	0.055	0.142	0.058
GTR+G+I	33	20706.354	20485.040	-10199.419	0.14	3.05	1.33	0.262	0.253	0.240	0.244	0.059	0.042	0.146	0.061	0.131	0.045	0.046	0.138	0.065	0.156	0.047	0.064	0.156	0.047
HKY+G	28	20712.648	20507.871	-10225.863	n/a	0.82	1.44	0.262	0.253	0.240	0.244	0.052	0.049	0.144	0.054	0.142	0.050	0.054	0.149	0.050	0.154	0.052	0.049	0.154	0.052
TN93+I	29	20728.876	20516.792	-10229.318	0.15	n/a	0.97	0.262	0.253	0.240	0.244	0.064	0.061	0.136	0.066	0.102	0.062	0.066	0.108	0.062	0.146	0.064	0.061	0.146	0.064
GTR+I	32	20734.563	20500.556	-10218.183	0.15	n/a	1.19	0.262	0.253	0.240	0.244	0.063	0.045	0.134	0.065	0.129	0.048	0.050	0.136	0.069	0.144	0.050	0.067	0.144	0.050
GTR+G	32	20745.871	20511.864	-10223.837	n/a	0.82	1.43	0.262	0.253	0.240	0.244	0.056	0.042	0.156	0.058	0.129	0.045	0.046	0.136	0.059	0.167	0.047	0.058	0.167	0.047
TN93+G	29	20823.595	20611.511	-10226.677	n/a	2.22	1.15	0.262	0.253	0.240	0.244	0.059	0.056	0.139	0.061	0.120	0.057	0.061	0.127	0.057	0.148	0.059	0.056	0.148	0.059
JC+G+I	25	20831.063	20648.214	-10299.049	0.14	3.52	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
JC+I	24	20850.762	20675.222	-10313.557	0.15	n/a	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
JC+G	24	20867.451	20691.911	-10321.902	n/a	0.88	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
K2	24	21150.633	20975.094	-10463.493	n/a	n/a	0.99	0.250	0.250	0.250	0.250	0.063	0.063	0.125	0.063	0.125	0.063	0.063	0.125	0.063	0.125	0.063	0.125	0.063	0.125
TN2	25	21152.192	20969.342	-10459.613	n/a	n/a	0.99	0.258	0.258	0.242	0.242	0.065	0.061	0.121	0.065	0.121	0.061	0.065	0.128	0.061	0.128	0.065	0.061	0.128	0.065
HKY	27	21170.862	20973.394	-10459.629	n/a	n/a	0.99	0.262	0.253	0.240	0.244	0.064	0.060	0.122	0.066	0.120	0.061	0.066	0.126	0.061	0.131	0.064	0.060	0.131	0.064
TN93	28	21178.113	20973.336	-10458.595	n/a	n/a	0.99	0.262	0.253	0.240	0.244	0.064	0.060	0.128	0.066	0.113	0.061	0.066	0.119	0.061	0.138	0.064	0.060	0.138	0.064
GTR	31	21202.066	20975.367	-10456.594	n/a	n/a	1.00	0.262	0.253	0.240	0.244	0.067	0.054	0.128	0.069	0.113	0.056	0.059	0.119	0.070	0.138	0.058	0.069	0.138	0.058
JC	23	21292.208	21123.978	-10538.939	n/a	n/a	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083

NOTE: Models with the lowest BIC scores (Bayesian Information Criterion) are considered to describe the substitution pattern the best. For each model, AICc value (Akaike information Criterion, corrected), Maximum Likelihood value (lnL), and the number of parameters (including branch lengths) are also presented [1]. Non-uniformity of evolutionary rates among sites may be modeled by using a discrete Gamma distribution (+G) with 5 rate categories and by assuming that a certain fraction of sites are evolutionarily invariable (+I). Whenever applicable, estimates of gamma shape parameter and/or the estimated fraction of invariant sites are shown. Assumed or estimated values of transition/transversion bias (R) are shown for each model, as well. They are followed by nucleotide frequencies (f) and rates of base substitutions (r) for each nucleotide pair. Relative values of instantaneous r should be considered when evaluating them. For simplicity, sum of r values is made equal to 1 for each model. For estimating ML values, a tree topology was automatically computed. This analysis involved 13 nucleotide sequences. There were a total of 2023 positions in the final dataset. Evolutionary analyses were conducted in MEGA11 [2].

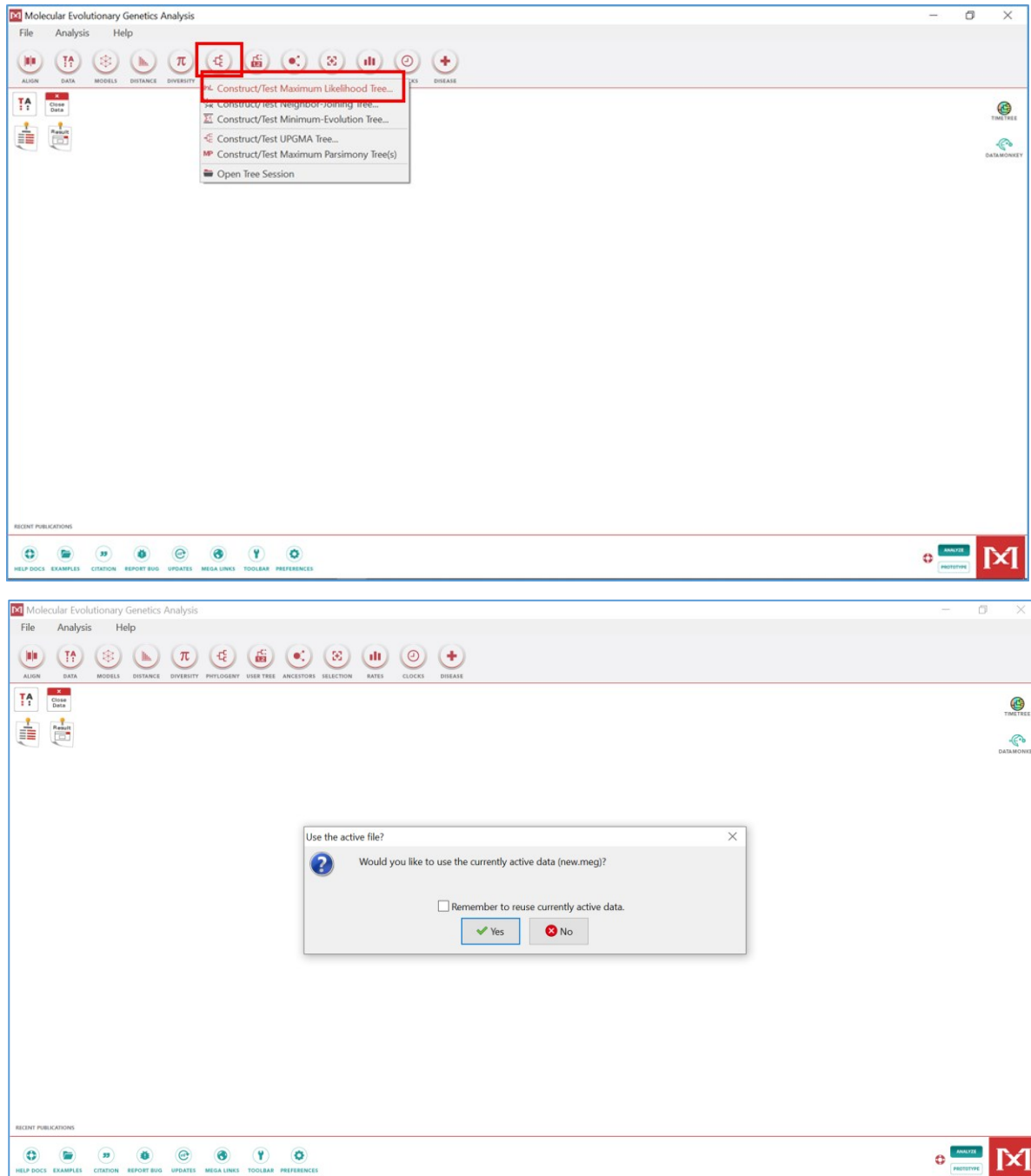
Abbreviations: TR: General Time Reversible; HKY: Hasegawa-Kishino-Yano; TN93: Tamura-Nei; TN92: Tamura 2-parameter; K2: Kimura 2-parameter; JC: Jukes-Cantor (dN).

1. Nei M. and Kumar S. (2000). Molecular Evolution and Phylogenetics. Oxford University Press, New York.

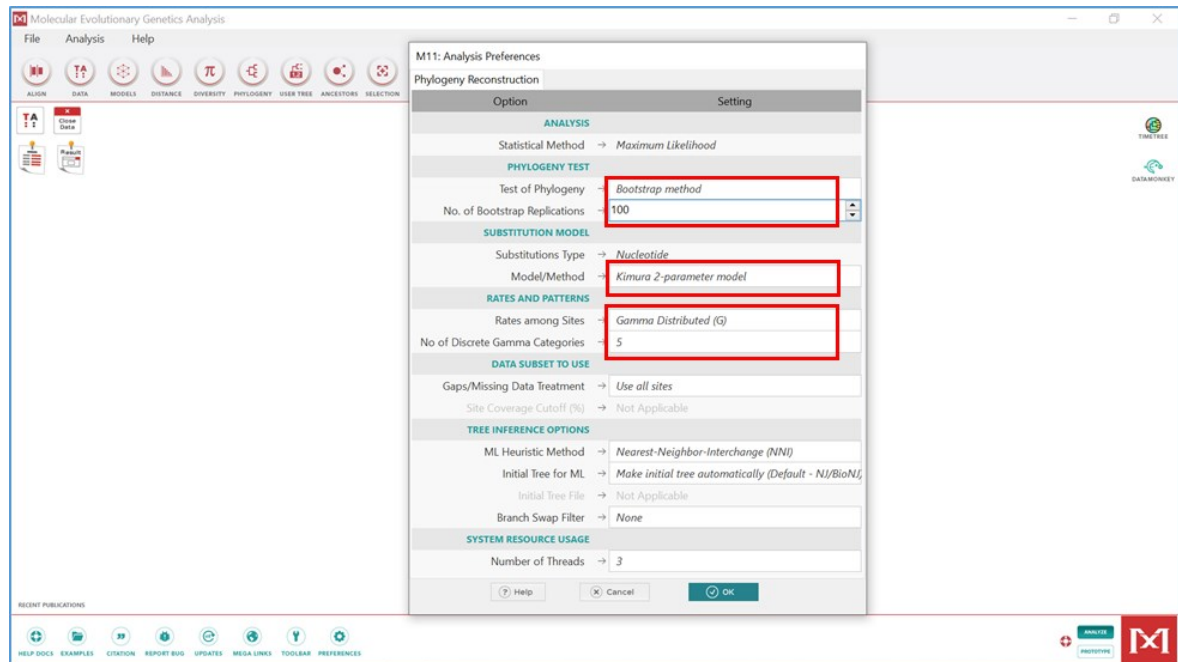
2. Tamura K., Stecher G. and Kumar S. (2021). MEGA 11: Molecular Evolutionary Genetics Analysis Version 11. Molecular Biology and Evolution <https://doi.org/10.1093/molbev/mbaa120>.

Disclaimer: Although utmost care has been taken to ensure the correctness of the caption, the caption text is provided "as is" without any warranty of any kind. Authors advise the user to carefully check the caption text for its use for any purpose and report any errors or problems to the authors immediately (seameo@seameo.net). In no event shall the authors and their employers be liable for any damages, including but not limited to special, consequential, or other damages. Author specifically disclaims all other warranties expressed or implied, including but not limited to the representation of accuracy of the caption text for a specific purpose, use, or application.

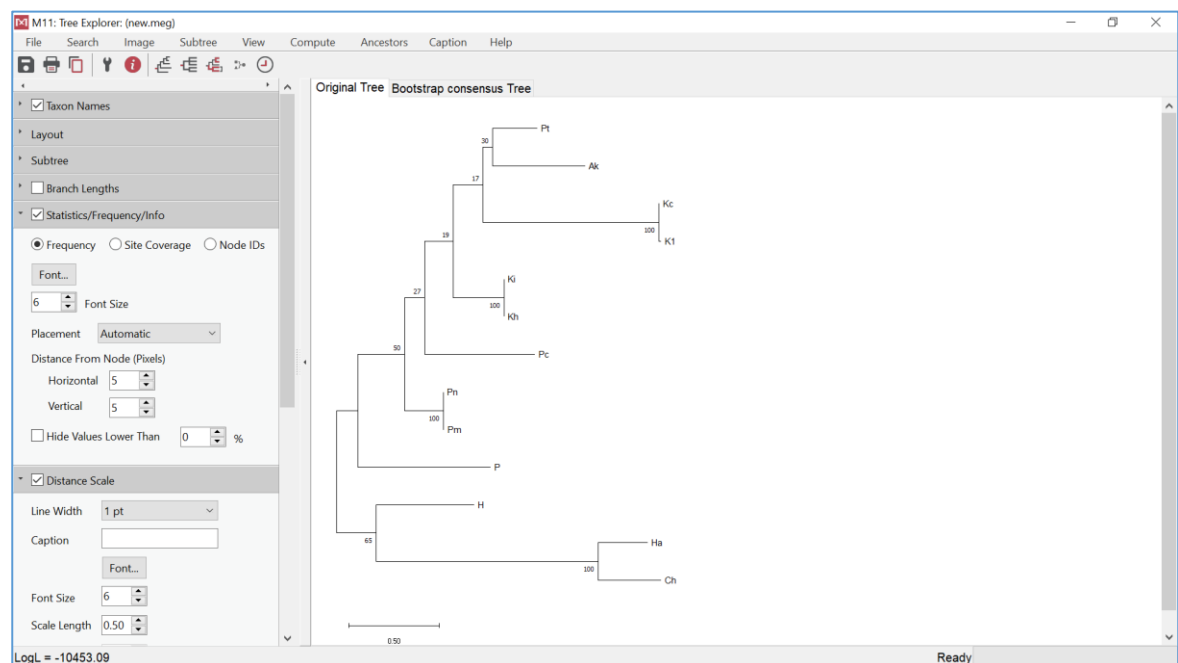
12. Selanjutnya, untuk pembuatan pohon filogenetik, maka pada *window* utama pilih menu *phylogeny*, kemudian pilih metode yang akan digunakan dalam pembuatan pohon filogenetik. Pada contoh kali ini dipilih metode *maximum likelihood*.



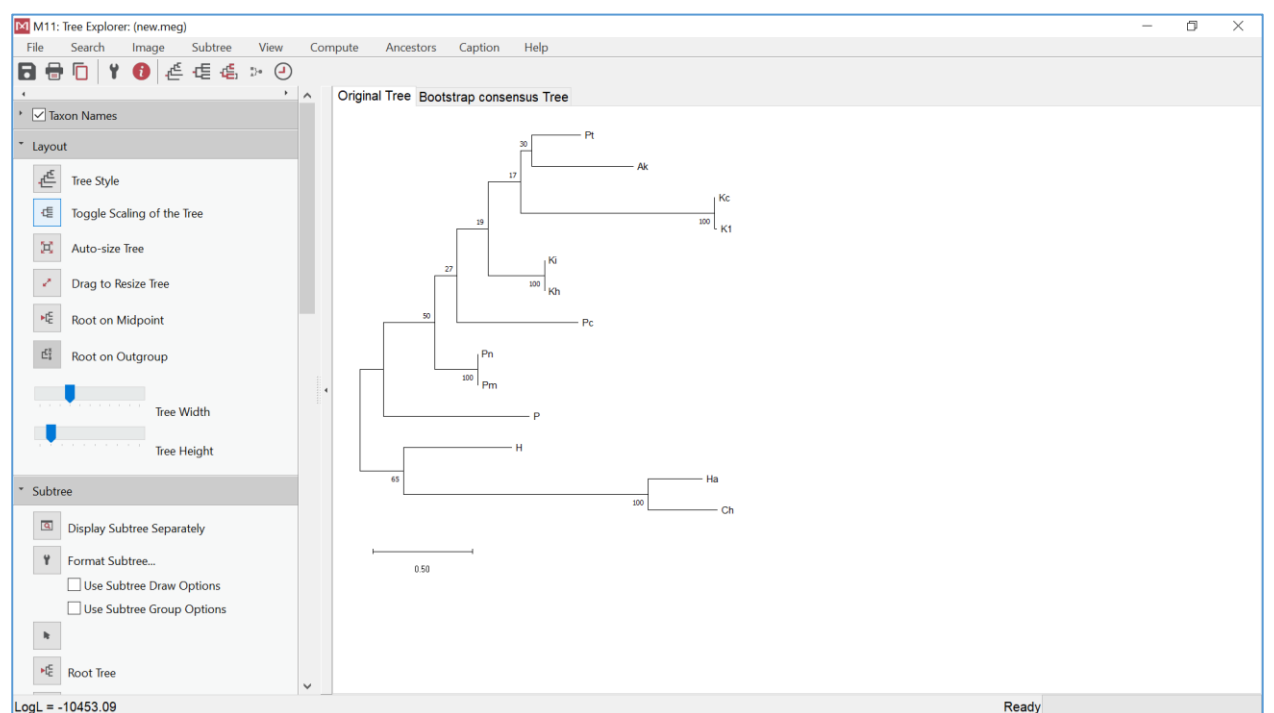
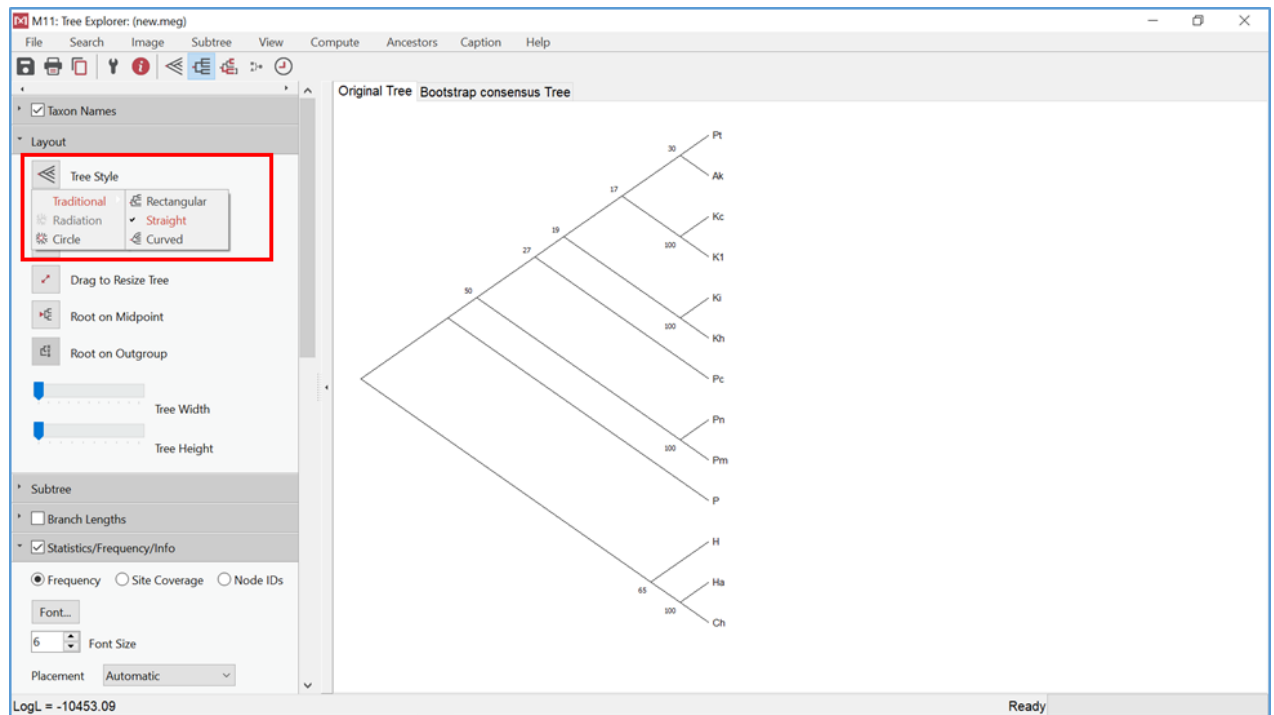
13. Pada pilihan berikutnya, ubah model agar sesuai dengan hasil *best model*. Ubah model menjadi Kimura 2-parameter, *rates* menjadi *gamma*, dan *number of discrete gamma categories* menjadi 5. Selain itu, dapat juga dilakukan penambahan uji filogeni menggunakan *bootstrap*. Nilai *bootstrap* dapat diubah sesuai jumlah pengulangan data secara random yang kita inginkan; umumnya antara 1-10.000. Semakin tinggi nilai *bootstrap* maka analisis membutuhkan waktu yang semakin lama. Pohon filogeni hasil *bootstrap* menunjukkan apakah *clade* atau cabang yang dihasilkan tanpa *bootstrap* masih dapat dihasilkan.



14. Hasil pohon filogenetik.



15. Tampilan pohon filogenetik dapat diubah menggunakan *tools* yang terletak di samping diagram.



Catatan:

BAB V

DESAIN PRIMER

Primer adalah sekuens DNA sintetis yang komplementer dengan bagian yang akan dianalisis dari total DNA/RNA dalam proses *Polymerase Chain Reaction* (PCR). Primer dapat diperoleh dengan cara melakukan desain primer. Desain primer dilakukan secara *in silico*, yaitu merancang primer dengan bantuan suatu *software* Bioinformatika, sehingga mendapatkan suatu primer sesuai kriteria primer yang baik untuk amplifikasi. *Software* tersebut akan secara otomatis menentukan primer yang layak pada suatu sekuens untuk kemudian digunakan di dalam amplifikasi.

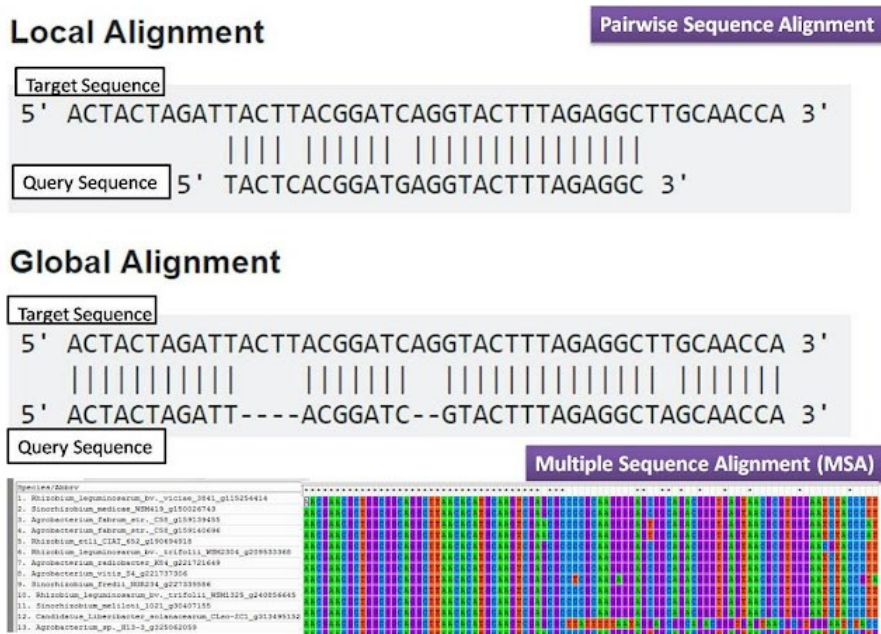
Aktivitas perancangan/desain primer dilakukan untuk memperoleh primer yang dapat digunakan dalam amplifikasi DNA dengan metode PCR. Desain primer yang baik sangat penting, karena keberhasilan amplifikasi DNA sangat bergantung pada ketepatan primer yang digunakan. Primer yang digunakan dalam proses PCR harus dapat membatasi daerah yang akan diamplifikasi/diperbanyak. Primer yang didesain mempunyai fungsi sebagai pembatas fragmen DNA target yang akan diamplifikasi. Desain primer yang bagus sangat penting untuk keberhasilan reaksi PCR. Oleh karena itu, desain primer ini penting sebagai kunci untuk amplifikasi spesifik dengan hasil tinggi. Beberapa *software* yang dapat membantu melakukan desain primer di antaranya adalah:

1. *Oligo Analyzer* 1.0.2.
2. *Oligo Explorer* 1.1.0.
3. Primer3.
4. NetPrimer.
5. Eurofins Genomics' Primer Design Tools.
6. PrimerQuest.
7. Primer BLAST NCBI.
8. SnapGene.
9. RExPrimer, dan lain-lain.
10. Mega X.

Beberapa *software* harus di-*install* seperti *Oligo Explorer*, sementara kebanyakan *software* berbasis web. Sekuens primer sebaiknya tidak mempunyai daerah yang dapat berikatan baik dengan internal primer tersebut maupun dengan primer pasangannya.

Selanjutnya primer menggunakan pasangan basa yang komplementer dan tidak memiliki banyak *binding sites* dalam genom target. Primer yang dirancang ini juga tidak menempel pada banyak molekul dan hanya dapat menempel dengan baik pada daerah spesifik yang ditargetkan.

Konsep pensejajaran sekuens (*sequence alignment*) merupakan pengaturan timbal balik dari dua sekuens atau lebih. Pensejajaran yang optimal adalah yang memperhatikan paling banyak kesamaan dan kemiripan sekuens serta kemungkinan sekuens-sekuens tersebut memiliki kesamaan fungsi. Dari jumlah sekuensnya, *sequence alignment* dapat dibagi menjadi dua kategori, yaitu: *pairwise alignment* dan *multiple sequence alignment*. Sementara itu, *sequence alignment* ditinjau dari lingkup pensejajarannya dibagi menjadi dua yaitu: *global alignment* dan *local alignment*. *Global alignment* adalah mensejajarkan dua sekuens yang panjangnya relatif sama dan bertujuan untuk mensejajarkan setiap basa yang dapat disejajarkan dari ujung 5' hingga ujung 3'. *Local alignment* bertujuan untuk mencari apakah sekuens yang disejajarkan memiliki kesamaan motif atau *region* tertentu, yang mana umumnya panjang sekuens memiliki perbedaan yang cukup besar. *Pairwise alignment* yaitu mensejajarkan dua utas sekuens, sedangkan *multiple sequence alignment* bertujuan mensejajarkan beberapa sekuens sekaligus. Berbagai perbedaan *alignment* yang digunakan dapat dilihat pada gambar di bawah ini:



Sumber: <https://www.majordifferences.com/2016/05/difference-between-pairwise-and-multiple-sequence-alignment.html>

Tujuan dari *alignment* adalah untuk melihat homologi atau kesamaan relatif dari suatu gen atau produk gen, baik pada data yang kita miliki maupun pada *database*. Informasi kesejajaran sekuens ini berguna untuk menjawab pertanyaan, misalnya:

1. Mengidentifikasi sekuens dari struktur yang belum diketahui.
2. Mengidentifikasi kesamaan *region* yang terkait dengan evolusi molekuler.
3. Mencari motif tertentu untuk pembuatan primer.

Pada umumnya tahapan desain primer dari suatu sekuens DNA adalah:

1. Menyiapkan data hasil sekuens gen atau *region* yang akan dibuat primernya dalam format FASTA, atau
2. Mencari sekuens DNA organisme target dengan membuka situs *web GenBank*
<https://www.ncbi.nlm.nih.gov/>
<http://www.ebi.ac.uk>
 dengan memasukkan nama gen atau target *region* yang diinginkan beserta nama spesiesnya.
3. Menentukan bagian sekuens DNA spesifik, jika ada bagian yang harus ikut dalam produk PCR.
4. Menetapkan dan meng-*input* kriteria primer yang diinginkan.

Kriteria pemilihan hasil BLAST untuk desain primer spesifik gen dari isolat yang telah diketahui nama spesiesnya untuk digunakan pada saat *alignment* dan analisis pemilihan *region* konservatif yaitu:

- Pilih hasil BLAST yang memiliki *identity* paling tinggi dan spesies yang sama dengan isolat yang kita miliki.
- Gunakan sebanyak mungkin agar dapat melakukan *multiple alignment*.
- Gunakan juga spesies isolat lain dengan *identity* yang rendah sebagai pembanding.
- Simpan file ke dalam format fasta/multifasta.
- Cek *region* konservatif untuk digunakan sebagai *region* desain primer (*software BioEdit*).

Aplikasi primer antara lain:

1. Kloning gen (*gene specific primer (GSP)*, *degenerate primer*, *adapter primer*).
2. Analisis ekspresi gen berbasis PCR dan *Real Time PCR*.
3. Konstruksi *probe* untuk analisis *Northern-blot*, *Southern-blot*, dan *Micro-array*.
4. Peruntukan DNA (*DNA Sequencing*).

Stabilitas suatu primer mempengaruhi penempelan primer pada *template*. Primer yang baik adalah primer yang memenuhi kriteria parameter primer. Akan tetapi, kriteria primer yang baik berbeda-beda tergantung aplikasinya. Contohnya primer yang digunakan untuk analisis qPCR akan berbeda kriterianya dengan primer untuk analisis *mikrosatelit*. Primer yang digunakan untuk mengambil sekuens gen yang akan dikloning akan berbeda kriterianya dengan primer yang akan digunakan untuk PCR koloni bakteri hasil kloning gen. Perancangan primer harus memenuhi beberapa parameter kriteria antara lain:

1. Panjang primer

Pada umumnya panjang primer berkisar antara 18–30 basa. Panjang primer lebih dari 30 basa tidak akan meningkatkan spesifisitas primer secara bermakna dan hal ini akan menyebabkan harga yang lebih mahal. Sedangkan primer dengan panjang kurang dari 18 basa akan menjadikan spesifisitas primer menjadi rendah. Pada primer yang berukuran pendek kemungkinan terjadinya *mispriming* (penempelan primer di tempat lain yang tidak diinginkan) akan tinggi; yang akan menyebabkan berkurangnya spesifisitas dari primer tersebut. Hal ini nantinya akan berpengaruh pada efektivitas dan efisiensi proses PCR.

2. Komposisi primer.

Urutan nukleotida yang sama perlu dihindari, karena dapat menurunkan spesifisitas primer sehingga memungkinkan terjadinya *mispriming* di tempat lain. Jumlah Basa G dan C (*GC content*) di dalam primer yang ideal sekitar 40-60%. Kandungan komposisi (G+C)) (persentase jumlah G dan C) sebaiknya sama atau lebih besar daripada kandungan (G+C) DNA target, karena primer dengan persentase jumlah (G+C) rendah tidak akan mampu berkompetisi untuk menempel secara efektif pada *template*, sehingga menurunkan efisiensi proses PCR. Urutan nukleotida pada ujung 3' sebaiknya G atau C. Nukleotida A atau T lebih toleran terhadap *Mismatch* daripada G atau C, sehingga akan dapat menurunkan spesifisitas primer. Jumlah basa G dan C yang terdapat pada 5 basa terakhir (3') disebut dengan *GC clamp*. *GC clamp* yang baik adalah sekitar 3 basa G/C dan tidak melebihi 5 basa G/C. Keberadaan G/C di ujung 3' primer sangat membantu terjadinya stabilitas ikatan antara primer dengan DNA *template* yang diperlukan untuk inisiasi *polymerase* DNA pada proses PCR. Persentase basa G/C pada primer berbeda-beda tergantung tujuan PCR, tetapi pada umumnya berkisar antara 40-60%.

3. Panjang Produk PCR.

Panjang produk yang menjadi target PCR juga menjadi pertimbangan penting dalam mendesain primer. Contohnya untuk aplikasi RT-PCR panjang produk sebaiknya berkisar antara 80-150 bp karena terkait dengan kemampuan alat *sensor fluoresens* pada mesin RT-PCR. Sedangkan untuk PCR *mikrosatelit* panjang produk yang diinginkan berkisar antara 100-300 bp.

4. *Melting Temperature* (T_m)

Primer Melting Temperature (T_m) merupakan suhu yang diperlukan oleh separuh primer dupleks agar mengalami disosiasi/lepas ikatan, yaitu suhu di mana 50% untai ganda DNA terpisah. Pemilihan T_m sangat berpengaruh di dalam pemilihan suhu *annealing*. T_m ini berkaitan dengan komposisi primer dan panjang primer. Primer dengan T_m berkisar antara 50–65 °C sangat ideal, sedangkan T_m di atas 65 °C akan mengurangi efektivitas *annealing* sehingga proses amplifikasi DNA kurang berjalan baik. T_m ini sangat ditentukan oleh jumlah basa GC. Menghitung T_m dapat dilakukan dengan cara:

1. Manual dengan formula rumus.
2. Menggunakan T_m Calculator yang tersedia pada beberapa program.

Secara teoritis T_m primer dapat dihitung dengan menggunakan rumus sebagai berikut:

A. $T_m (^{\circ}\text{C}) = ((G+C) \times 4) + ((A+T) \times 2)$ formula secara kasar (kurang akurat).

B. $T_m (^{\circ}\text{C}) = \{\Delta H / \Delta S + R \ln(C)\} - 273,15$ formula secara akurat.

5. *Annealing temperature* (T_a)

Primer *annealing temperature* (T_a) merupakan suhu yang diperkirakan untuk suatu primer dapat berikatan dengan DNA *template* secara stabil (DNA-DNA *hybrid stability*). T_a yang tinggi akan sangat menyulitkan terjadinya ikatan primer dengan DNA *template*, sehingga menghasilkan produk PCR yang rendah dan kurang efisien. Namun, jika T_a terlalu rendah akan menyebabkan terjadinya penempelan primer pada DNA *template* yang tidak spesifik. T_a dapat dihitung dengan menggunakan rumus di bawah ini:

$$T_a = 0,3 \times T_m(\text{primer}) + 0,7 T_m (\text{produk PCR}) - 14,9$$

$$T_m (\text{primer}) = T_m \text{ primer}$$

$$T_m (\text{product}) = T_m \text{ produk PCR}$$

6. Primer *Secondary Structures* :

- a) *Hairpins* merupakan interaksi intramolekuler dalam primer, yang membentuk struktur *loop/hairpins* pada primer. *Hairpins* dalam primer dapat mengganggu proses penempelan primer pada *template* dalam proses PCR. Jika *hairpins* tidak terdapat di dalam primer yang didesain, maka primer ini cukup baik untuk dapat digunakan di dalam proses PCR. *Hairpins* sebaiknya dihindari. Akan tetapi, sangat sulit untuk memperoleh primer tanpa memiliki struktur *hairpins*. *Hairpins* yang terdapat pada ujung 3' dengan $\Delta G = -2$ kkal/mol dan *hairpins* internal dengan $\Delta G = -3$ kkal/mol masih dapat ditoleransi (Catatan: ΔG adalah energi yang diperlukan untuk memecah struktur *hairpins*).
 - b) *Self Dimer* yaitu primer yang berikatan dengan primer lainnya yang sejenis. *Dimer* pada ujung 3' primer sebaiknya tidak lebih dari 3 basa karena dapat menurunkan spesifisitas primer. *Self Dimer* pada ujung 3' dengan $\Delta G = -5$ kkal/mol dan *self Dimer* pada bagian internal dengan $\Delta G = -6$ kkal/mol masih dapat ditoleransi. Primer sebaiknya tidak mempunyai 3 atau lebih basa G atau C pada 3' *dimer*, karena dapat menstabilkan *annealing* primer non-spesifik. *Dimer* pada ujung 3' masing-masing primer yang dirancang adalah 1 *dimer* pada primer *forward* dan 2 *dimer* pada primer *reverse* dan tidak melebihi batas dari parameter *dimer* suatu primer.
 - c) *Cross Dimer* yaitu primer yang berikatan dengan primer pasangannya (*reverse* dan *forward*). *Cross dimer* terjadi karena adanya bagian yang homolog antara dua utas primer tersebut. *Dimer* pada ujung 3' dengan ΔG sebesar - 5 kkal/mol dan cross-dimer internal pada ujung 3' dengan $\Delta G = -5$ kkal/mol dan *self- dimer* pada bagian internal dengan $\Delta G = -6$ kkal/mol masih dapat ditoleransi.
7. *Repeats* merupakan nukleotida yang berulang dalam primer. Adanya *repeats* dapat menyebabkan terjadinya penempelan primer di tempat yang tidak diinginkan (*mispriming*). *Repeats* primer sebaiknya tidak memiliki urutan pengulangan dari 2 basa. Maksimum pengulangan 2 basa sebanyak 4 kali masih dapat ditoleransi, misalnya ATATATAT.
 8. *Runs* yaitu primer yang memiliki pengulangan basa secara berurutan. *Runs* sebaiknya tidak terjadi secara terus menerus. Pengulangan basa berurutan sampai 4 kali masih dapat ditoleransi. Misalnya AGCGGGGGATGGGG memiliki urutan basa G diulang 5 kali berturut-turut.

9. *Avoid Cross Homology*. Untuk menghindari *Cross Homology* dapat dilakukan analisis homologi primer dengan DNA genom melalui BLAST-NCBI.
10. *Amplicon Length* yaitu panjang PCR produk yang ideal berkisar antara 100-500 pasang basa.
11. *Optimum annealing temperature* (T_a Opt) yaitu suhu *annealing* optimum. Suhu ini sangat mempengaruhi hasil PCR. T_a Opt ini dapat dihitung dengan rumus:

$$T_a \text{ Opt} = 0,3 \times (T_m \text{ primer}) + 0,7 \times (T_m \text{ produk PCR}) - 25.$$
12. *Primer Pair T_m Mismatch* yaitu perbedaan T_m sepasang primer, yang sebaiknya tidak lebih dari 5 °C.

Beberapa kriteria spesifik pada perancangan primer untuk RT-PCR antara lain:

1. Panjang produk PCR target.
 Panjang produk RT-PCR yang diharapkan yaitu 80-150 bp. Proses RT-PCR tidak mengambil seluruh gen karena sensor pada mesin RT-PCR mampu mendeteksi fluoresensi yang dihasilkan dari produk RT-PCR dengan kisaran panjang yang terbatas. Panjang minimal produk PCR perlu ditetapkan agar tidak dapat dibedakan dari primer-dimer yang mungkin terbentuk.
2. Panjang primer.
 Panjang primer untuk RT-PCR yaitu 17-25 bp.
3. Kandungan GC (%GC) berkisar antara 40-60%, lebih ideal lagi jika antara 45-55%. Hal tersebut untuk menghindari ketidakcocokan atau *mismatch* serta kestabilan primer dan produk PCR.
4. *Melting temperature* (T_m)
 T_m yang baik yaitu berkisar antara 60-65 °C. Selain itu, pada pasangan primer, perbedaan T_m yang baik tidak lebih dari 5 °C.
5. Ujung 3' primer.
 Ujung 3' pada masing-masing primer berupa basa G atau C. Hal tersebut dilakukan karena basa G dan C memiliki ikatan hidrogen yang lebih kuat, sehingga kestabilan primer lebih baik. Selain itu, basa T pada ujung primer meningkatkan potensi *mismatch*.
6. Hindari adanya struktur sekunder dan pengulangan basa berurutan (*runs*) lebih dari 3 basa. Hindari juga terbentuknya ikatan primer-dimer (primer berikatan dengan primer pasangannya, bukan dengan nukleat target) yang dapat membuat ambigu dan terbaca oleh mesin. Struktur sekunder, *runs*, dan primer dimer dapat diuji menggunakan aplikasi seperti NetPrimer.

7. Suhu *annealing* untuk RT-PCR rata-rata adalah 60 °C. Beberapa *reagen* RT-PCR dapat mengamplifikasi primer yang berbeda pada suhu 60 °C. Sementara itu, beberapa *reagen* lain butuh penyesuaian suhu *annealing*, tetapi tidak jauh dari 60 °C.

LANGKAH-LANGKAH MENDESAIN PRIMER

5.1 Contoh desain primer menggunakan fitur pada NCBI

1. Buka *website* <http://www.ncbi.nlm.nih.gov/tools/primer-blast/>

The screenshot shows the NCBI Primer-BLAST web interface. The top navigation bar includes the NIH logo and the text "U.S. National Library of Medicine National Center for Biotechnology Information". The main heading is "Primer-BLAST" with the subtitle "A tool for finding specific primers". Below this, there are tabs for "Primers for target on one template" and "Primers common for a group of sequences". The "PCR Template" section has a text input field for "Enter accession, GI, or FASTA sequence (A refseq record is preferred)" which is highlighted with a red box and labeled "kotak tempat memasukkan sekuens yang di-copy". There are also fields for "Forward primer" and "Reverse primer". The "Primer Parameters" section includes fields for "Min" and "Max" values for PCR product size, number of primers to return, and primer melting temperatures. The "Exon/intron selection" section has fields for "Exon junction span" and "Exon junction match". The "Primer Pair Specificity Checking Parameters" section includes a "Specify check" checkbox, "Search mode" dropdown, "Database" dropdown, "Exclusion" checkbox, "Organism" dropdown, and "Entrez query" field. The "Get Primers" button is at the bottom left of the form.

- Masukkan urutan DNA dalam bentuk FASTA ke dalam kotak yang telah disediakan serta mengganti kriteria primer sesuai kriteria primer yang akan dibuat. Untuk RT-PCR, kriteria panjang primer 80-150 bp dengan T_m 60-65 °C.

Primer-BLAST A tool for finding specific primers

Finding primers specific to your PCR template (using Primer3 and BLAST).

Primers for target on one template Primers common for a group of sequences

[Retrieve recent results](#) [Publication](#) [Tips for finding specific primers](#)

PCR Template

Enter accession, gi, or FASTA sequence (A refseq record is preferred) Clear

GGAAGTCCCAGTAAGATGTCAGCAGAAGTGTATCAGTCACAGGTATCATCGATAAATATTTT
 CCTACCAGTAACAGAAT
 CGACAATTTTCAGCAGCCCTTCGTTCTCAGATGTTACATCTAGAGCTGTTTATGAGGAGTTGG
 TAATCAACCAAGGGCAA
 AGACCCTTGAAACACCTCCTCTGAATGTCTCGCGTGCGCCATTATTGCATGTTAACTCAGTTT
 AGCAATGTAAGGAT

or, upload FASTA file Choose File No file chosen

Range Clear

Forward primer From To

Reverse primer

Primer Parameters

Use my own forward primer (5'→3' on plus strand) Clear

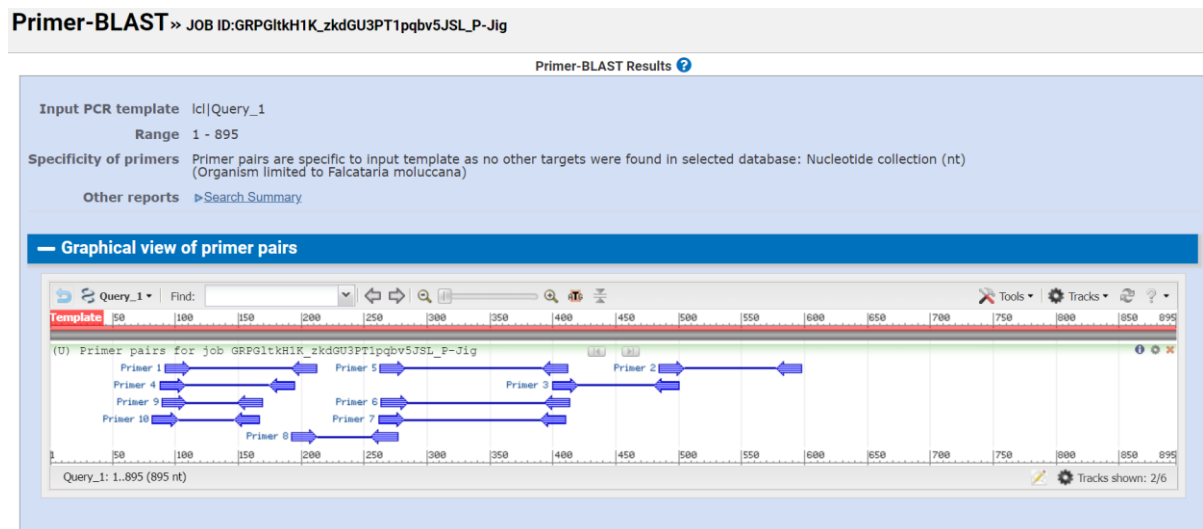
Use my own reverse primer (5'→3' on minus strand) Clear

PCR product size Min Max

of primers to return

Primer melting temperatures (T_m) Min Opt Max Max T_m difference ?

- Diperoleh hasil seperti di bawah ini:



— Detailed primer reports

Primer pair 1

	Sequence (5'→3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	GCAGCCCTTCGTTCTCAGAT	Plus	20	92	111	60.11	55.00	5.00	3.00
Reverse primer	ACATGCAATAATGGCGCACG	Minus	20	212	193	60.25	50.00	5.00	2.00
Product length	121								

Primer pair 2

	Sequence (5'→3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	TGGGCGTAGGCTCAGAGTTA	Plus	20	485	504	60.32	55.00	6.00	3.00
Reverse primer	AACAGCTCCGAGAAGCTAC	Minus	20	597	578	60.11	55.00	5.00	0.00
Product length	113								

Primer pair 3

	Sequence (5'→3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	AGCGACAGCGATGATTGACA	Plus	20	400	419	60.11	50.00	4.00	3.00
Reverse primer	TCTGAGCCTACGCCACATA	Minus	20	500	481	60.40	55.00	6.00	2.00
Product length	101								

Primer pair 4

	Sequence (5'→3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	TTCAGCAGCCCTTCGTTCTC	Plus	20	88	107	60.32	55.00	3.00	0.00
Reverse primer	CGCGAGACATTCAGAGGAGG	Minus	20	194	175	60.25	60.00	4.00	0.00
Product length	107								

Primer pair 5

	Sequence (5'→3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	GGAAGCGCCTTCTTTCTGCT	Plus	20	263	282	60.96	55.00	6.00	0.00
Reverse primer	ATCGCTGTCGCTGATGTTGA	Minus	20	411	392	60.11	50.00	3.00	1.00
Product length	149								

Primer pair 6

	Sequence (5'→3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	GAAGCGCCTTCTTTCTGCTTC	Plus	21	264	284	60.14	52.38	6.00	0.00
Reverse primer	TCATCGCTGTCGCTGATGTT	Minus	20	413	394	60.11	50.00	5.00	1.00
Product length	150								

Primer pair 7

	Sequence (5'→3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	AGGAAGCGCCTTCTTTCTGCTC	Plus	20	262	281	60.96	55.00	6.00	2.00
Reverse primer	TCGCTGTCGCTGATGTTGAA	Minus	20	410	391	60.32	50.00	3.00	3.00
Product length	149								

Primer pair 8

	Sequence (5'→3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	CGTGCGCCATTATTGCATGT	Plus	20	193	212	60.25	50.00	5.00	2.00
Reverse primer	AAGAAGCGCGCTTCTCAAGTTT	Minus	21	276	256	60.20	47.62	6.00	0.00
Product length	84								

Primer pair 9

	Sequence (5'→3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	CAGCAGCCCTTCGTTCTCA	Plus	19	90	108	60.00	57.89	3.00	1.00
Reverse primer	CAAGGGTCTTTGCCCTTGGT	Minus	20	169	150	60.47	55.00	7.00	0.00
Product length	80								

Primer pair 10

	Sequence (5'→3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	GACAATTTGAGCAGCCCTTCG	Plus	21	82	102	60.14	52.38	4.00	2.00
Reverse primer	GGGTCTTTGCCCTTGTTGA	Minus	20	166	147	60.47	55.00	3.00	1.00
Product length	85								

5.2 Analisis struktur sekunder menggunakan NetPrimer

Untuk mengakses NetPrimer, perlu membuat akun dengan registrasi e-mail pada laman <https://www.premierbiosoft.com/netprimer/>.

PREMIER Biosoft
Accelerating Research in Life Sciences

HOME COMPANY PRODUCTS ORDERING SUPPORT BLOG

Net Primer

Free! Primer Analysis Software

"I wish I had analyzed my primers before ordering for the third time. Now I check all my primers with NetPrimer before ordering" - Stacey McCann, Cancer Biology Program, Stanford University

[Access NetPrimer](#) [Request More](#) [View More Tools](#)

About NetPrimer

NetPrimer combines the latest primer analysis algorithms with a web-based interface allowing the user to analyze primers over the Internet. All primers are analyzed for **primer melting temperature** using the nearest neighbor thermodynamic theory to ensure accurate T_m prediction. Primers are analyzed for all **primer secondary structures** including hairpins, self-dimers, and cross-dimers in primer pairs. This ensures the availability of the primer for the reaction as well as minimizing the formation of primer dimer. The program eases quantitation of primers by calculating primer molecular weight and optical activity. To facilitate the selection of an optimal primer, each primer is given a rating based on the stability of its secondary structures. A comprehensive analysis report can be printed for individual primers or primer pairs.

Major Features

- ✓ **Primer Melting Temperature (T_m):** Predicts primer melting temperature using the accurate nearest neighbor thermodynamic algorithm.
- ✓ **Primer Properties:** T_m, molecular weight, GC%, optical activity (both in nmol/A260 & µg/A260), DG, 3' end stability, DH, DS, and 5' end DG.
- ✓ **Primer Secondary Structures:** Hairpins, dimers, cross dimers, palindromes, repeats and runs.
- ✓ **Primer Rating:** Quantitative prediction of the efficiency of a primer.
- ✓ **Comprehensive Report:** Prints complete primer analysis for an individual primer or primer pair.

This website uses cookies to ensure you get the best experience on our website. [Cookie Policy](#) [I Understand](#)

PREMIER Biosoft
Accelerating Research in Life Sciences

HOME COMPANY PRODUCTS DOWNLOADS ORDERING SUPPORT BLOG

Welcome Dr. [REDACTED] [Log Out](#)

[Launch NetPrimer](#)

For SYBR Green and TaqMan primer analysis and properties calculation, use the [free qpcr design software](#)

[Download Manual](#)

- Masukkan sekuens primer *forward* dan *reverse*. Sebagai contoh, pasangan primer yang dimasukkan adalah pasangan primer nomor 4 dari hasil pencarian primer di NCBI.

NetPrimer Free web based tool to analyze primers

Oligo Analysis

Name :

Description :

Sequence #1 (5' to 3')

Length : 20

Sequence #2 (5' to 3')

Length : 20

Reaction Conditions

Oligo Concentration : 250.0 pM

Monovalent Ion Concentration : 50.0 mM

Free Mg++ Ion Concentration : 1.5 mM

Total Na[+] Equivalent : 204.92 mM

Temperature for Free Energy Calculation : 25.0 °C

5. Setelah meng-klik *analyze* kemudian akan muncul hasil berupa *cross dimer* dengan energi yang dibutuhkan untuk memutus *cross dimer* sebesar 5,75 kkal/mol. Akan muncul pula *self dimer* pada primer *reverse* dengan energi untuk memutus ikatan pada *self dimer* sebesar 10,36 kkal/mol. Selain itu, ditemukan adanya *runs* pada primer *forward* sebanyak 3 basa yang sama berada secara berurutan.

Analysis Results #1: TTCAGCAGCCCTTCGTTCTC				
Rating	:	100.0		
Molecular Wt	:	5994.99		
Tm	:	60.29	°C	
GC%	:	55.0		
GC Clamp	:	1		
nmol/A ₂₆₀	:	5.84		
ug/A ₂₆₀	:	35.04		
ΔG	:	-35.3	kkal/mol	
3' end stability	:	-6.69	kkal/mol	
ΔH	:	-155.1	kkal/mol	
ΔS	:	-0.4	kkal/°K/mol	
5' end ΔG	:	-7.07	kkal/mol	
Self Dimer (ΔG)	:		kkal/mol	
Hairpin (ΔG)	:		kkal/mol	
Repeats (# of pairs)	:		kkal/mol	
Run (# of bases)	:	3	kkal/mol	

Analysis Results #2: CGCGAGACATTCAGAGGAGG				
Rating	:	81.0		
Molecular Wt	:	6216.12		
Tm	:	60.48	°C	
GC%	:	60.0		
GC Clamp	:	2		
nmol/A ₂₆₀	:	4.89		
ug/A ₂₆₀	:	30.4		
ΔG	:	-34.94	kkal/mol	
3' end stability	:	-9.31	kkal/mol	
ΔH	:	-151.9	kkal/mol	
ΔS	:	-0.39	kkal/°K/mol	
5' end ΔG	:	-11.94	kkal/mol	
Self Dimer (ΔG)	:	-10.36	kkal/mol	
Hairpin (ΔG)	:		kkal/mol	
Repeats (# of pairs)	:		kkal/mol	
Run (# of bases)	:		kkal/mol	

Cross Dimer (ΔG)	:	-5.75	kkal/mol	
------------------	---	-------	----------	--

Print

*Keterangan:

- Kualitas hasil analisis dapat dilihat dari nilai rating primer *forward/reverse* (F/R).
- Nilai Tm dan parameter lain dapat dilihat langsung pada *display software*.
- Untuk melihat sensitivitas primer, gunakan BLAST ke NCBI. Selanjutnya hasil dapat di-*print* dan primer dapat di-*order*.

Contoh lain yaitu primer nomor 1.

Oligo Analysis			
Name	:		
Description	:		
Sequence #1 (5' to 3')			
Length	:	20	
GCAGCCCTTCGTTCTCAGAT			
Sequence #2 (5' to 3')			
Length	:	20	
ACATGCAATAATGGCGCACG			
Reaction Conditions			
Oligo Concentration	:	250.0	pM
Monovalent Ion Concentration	:	50.0	mM
Free Mg++ Ion Concentration	:	1.5	mM
Total Na[+] Equivalent	:	204.92	mM
Temperature for Free Energy Calculation	:	25.0	°C
Default Analyze			

Hasil analisis menunjukkan bahwa pada primer *reverse* terdapat struktur *hairpins* dengan energi yang dibutuhkan untuk memutus ikatan yang membentuk struktur *hairpins* sebesar 1,89 kkal/mol

Analysis Results #1: GCAGCCCTTCGTTCTCAGAT			
Rating	:	92.0	
Molecular Wt	:	6044.02	
Tm	:	59.25	°C
GC%	:	55.0	
GC Clamp	:	1	
nmol/A ₂₆₀	:	5.57	
ug/A ₂₆₀	:	33.69	
ΔG	:	-34.83	kcal/mol
3' end stability	:	-6.6	kcal/mol
ΔH	:	-154.6	kcal/mol
ΔS	:	-0.4	kcal/°K/mol
5' end ΔG	:	-9.83	kcal/mol
Self Dimer (ΔG)	:	-4.17	kcal/mol
Hairpin (ΔG)	:		kcal/mol
Repeats (# of pairs)	:		kcal/mol
Run (# of bases)	:	3	kcal/mol
Analysis Results #2: ACATGCAATAATGGCGCACG			
Rating	:	77.0	
Molecular Wt	:	6135.09	
Tm	:	62.66	°C
GC%	:	50.0	
GC Clamp	:	2	
nmol/A ₂₆₀	:	5.01	
ug/A ₂₆₀	:	30.75	
ΔG	:	-36.93	kcal/mol
3' end stability	:	-10.05	kcal/mol
ΔH	:	-160.1	kcal/mol
ΔS	:	-0.41	kcal/°K/mol
5' end ΔG	:	-6.72	kcal/mol
Self Dimer (ΔG)	:	-10.89	kcal/mol
Hairpin (ΔG)	:	-1.89	kcal/mol
Repeats (# of pairs)	:		kcal/mol
Run (# of bases)	:		kcal/mol
Cross Dimer (ΔG)	:	-5.95	kcal/mol
Print			

5.3 Mendesain primer menggunakan primer3plus

Situs primer3plus yaitu <http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi> memunculkan beberapa informasi *tab* yang perlu diperhatikan antara lain:

1. *Main*
2. *General Setting*
3. *Advance Setting*
4. *Penalty Weight*
5. *Sequence Quality*

Langkah-langkah primer3plus :

1. Isi *Sequence ID*.



Primer3Plus
pick primers from a DNA sequence

[Primer3Manager](#) [Help](#)
[About](#) [Source Code](#)

Task: Detection Select primer pairs to detect the given template sequence. Optionally targets and included/excluded regions can be specified. [Pick Primers](#) [Reset Form](#)

Main **General Settings** **Advanced Settings** **Internal Oligo** **Penalty Weights** **Sequence Quality**

Sequence Id:

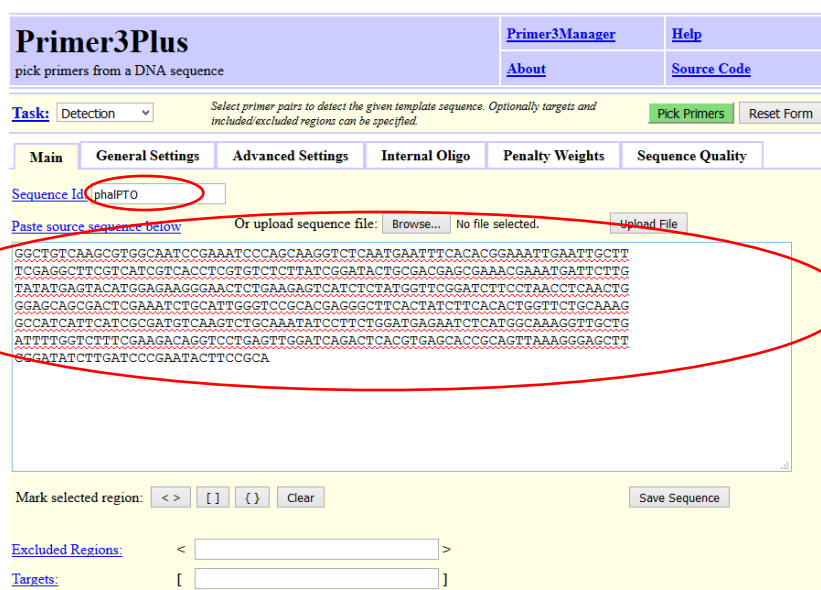
Paste source sequence below Or upload sequence file: [Browse...](#) No file selected. [Upload File](#)

Mark selected region: [< >](#) [\[\]](#) [{ }](#) [Clear](#) [Save Sequence](#)

Excluded Regions: [<](#) [>](#)

Targets: [\[](#) [\]](#)

2. Lanjut isi FASTA dengan cara *copy-paste* sekuens primer.



Primer3Plus
pick primers from a DNA sequence

[Primer3Manager](#) [Help](#)
[About](#) [Source Code](#)

Task: Detection Select primer pairs to detect the given template sequence. Optionally targets and included/excluded regions can be specified. [Pick Primers](#) [Reset Form](#)

Main **General Settings** **Advanced Settings** **Internal Oligo** **Penalty Weights** **Sequence Quality**

Sequence Id:

Paste source sequence below Or upload sequence file: [Browse...](#) No file selected. [Upload File](#)

GGCTGTCAAGCGTGGCAATCCGAAATCCAGCAAGGCTCAATGAATTCACACGGAAATGAATTGCTT
TCGAGGCTTCGTCATCGTCACCTCGTCTCTTATCGGATACTCGACGAGCGAAACGAATGATCTTG
TATATGAGTACATGGAGAAGGGAACCTCTGAAGAGTCATCTCTATGTTTGGATCTTCACTTCACTG
GGAGCAGGACTCGAAATCTGCAATGGGTCCGACGAGGGCTTCACTATCTTCACTGTTTCTGCAAG
GCCATCATTCGCGATGTCAAGTCTGCAAAATATCTTCTGGATGAGAATCTCATGGCAAAGGTTGCTG
ATTTGGTCTTTGGAAGACAGGCTCTGAGTTGGATCAGACTACCTGAGCACCAGCTTAAAGGGAGCTT
CCGATATCTTGATCCCGAATACTTCGCA

Mark selected region: [< >](#) [\[\]](#) [{ }](#) [Clear](#) [Save Sequence](#)

Excluded Regions: [<](#) [>](#)

Targets: [\[](#) [\]](#)

3. Di bagian *task* pilih primer untuk kloning.

Primer3Plus
pick primers from a DNA sequence

[Primer3Manager](#) [Help](#)
[About](#) [Source Code](#)

Task: Detection

Select primer pairs to detect the given template sequence. Optionally targets and included/excluded regions can be specified.

Main **Settings** **Advanced Settings** **Internal Oligo** **Penalty Weights** **Sequence Quality**

Sequences:

Paste source sequence below Or upload sequence file: No file selected.

GGCTGTCAAGCGTGGCAATCCGAAATCCAGCAAGGCTCAATGAATTCACACGGAATGAATTGCTT
TCGAGGCTTCGTATCGTCACCTCGTGTCTCTTATCGGATACGCGACGAGCGAAACGAAATGATTCTTG
TATATGAGTACATGGAGAAGGGAACCTGAAGAGTCATCTCTATGGTTCGGATCTTCTAACCTCAACTG
GGAGCAGCGACTCGAAATCTGCATTGGTCCGACGAGGGCTTCACTATCTTCACACTGGTTCTGCAAG
GCCATCATTCATCGCGATGCTCAAGTCTGCAATATCCTTCTGGATGAGAATCTCATGGCAAAGGTTGCTG
ATTTTGGTCTTTCGAAGACAGGCTCTGAGTTGGATCAGACTCACGTGAGCACCAGCTTAAAGGGAGCTT
CGGATATCTTGATCCCGAATACCTCCGCA

Mark selected region:

Excluded Regions: < >

Targets: []

Beri kurung kurawal pada pasta ke-2 di awal dan baris ke-3 [], artinya PCR produk berisi di sekuens ke-2, *left/forward primer*.

4. Klik *general setting* maka tampilan akan muncul seperti:

Primer3Plus
pick primers from a DNA sequence

[Primer3Manager](#) [Help](#)
[About](#) [Source Code](#)

Task: Detection

Mark an included region to pick primers fixed at its the boundaries. The quality of the primers might be low.

Main **General Settings** **Advanced Settings** **Internal Oligo** **Penalty Weights** **Sequence Quality**

150-250 100-300 301-400 401-500 501-600 601-700 701-850 851-1000

Primer Size Min: 18 Opt: 20 Max: 27
Primer Tm Min: 57.0 Opt: 60.0 Max: 63.0 **Max Tm Difference:** 100.0
Primer GC% Min: 20.0 Opt: Max: 80.0 **Fix the** 5 **prime end of the primer**

Concentration of monovalent cations: 50.0 **Annealing Oligo Concentration:** 50.0
Concentration of divalent cations: 0.0 **Concentration of dNTPs:** 0.0

Mispriming/Repeat Library: NONE

Load and Save
Please select special settings here: Default (use Activate Settings button to load the selected settings)
To upload or save a settings file from your local computer, choose here:
 No file selected.

Keterangan:

- ☼ *Product size range*: ukuran amplikon yang ditulis bervariasi dalam beberapa ukuran.
- ☼ Ukuran primer optimal adalah 18-28 bp.
- ☼ Tm: minimal 57; maksimal 63; opt 60, maksimum perbedaan Tm *forward* dan *reverse* adalah 3.
- ☼ Primer GC: minimal 20; maksimal 80. Nilai optimal untuk %GC adalah: 50-60%.

Selanjutnya di-klik *advance setting* dan akan muncul tampilan sebagai berikut:

Keterangan:

- ☀ *Max poly-X* → : 3 *promote mispriming* (menyebabkan *mispriming*)
- ☀ *Number to return*: Jumlah pilihan primer F/R hasil pencarian (*set to 10/diatur pada 10*)
- ☀ *Max 3' self complementarity*: *Set to 0* (diatur pada 0).

5. Setelah di-klik *pick primer* di atas (warna hijau), maka akan muncul tampilan primernya :

Keterangan:

- ❖ Jika primer gagal didesain maka ubah daerah amplifikasi, dengan cara tidak memilih *high complementary*. Jika memilih *high complementary*, maka akan menyebabkan primer berikatan dengan diri sendiri.
- ❖ Pilih TM yg paling kecil di antara *forward* dan *reverse* primer ex, yaitu pilih: 55.

- ❖ Semakin pendek primer maka primer semakin tidak spesifik. Sebaliknya jika primer semakin panjang, maka primer semakin spesifik.
 - ❖ Jika ada beberapa pilihan yang ditampilkan oleh *software*, maka primer urutan *pair1* adalah yang terbaik. Jika ada lebih dari satu, maka cek keseluruhan pasangan primer yang kita inginkan dengan menggunakan Net primer (<https://www.premierbiosoft.com/netprimer/>)
6. Setelah diperoleh primer, maka klik *Send to primer3 manager*.

Primer3Plus
pick primers from a DNA sequence

[Primer3Manager](#) [Help](#)
[About](#) [Source Code](#)

< Back

Pair 1:

☒ Left Primer 1: phalPTO_F
Sequence: CGTCACCTCGTGTCTTTAT
Start: 86 Length: 20 bp Tm: 55.3 °C GC: 50.0 % ANY: 3.0 SELF: 2.0

☒ Right Primer 1: phalPTO_R
Sequence: TTAGGAAGATCGGAACCAT
Start: 201 Length: 20 bp Tm: 54.8 °C GC: 40.0 % ANY: 5.0 SELF: 2.0

Product Size: 116 bp Pair Any: 4.0 Pair End: 3.0

[Send to Primer3Manager](#) [Reset Form](#)

1	GGCTGTCAG	CGTGGCAATC	CGAAATCCCA	GCAAGGTCTC	AATGAATTC
51	ACACGGAAAT	TGAATTGCTT	TCGAGGCTTC	GTCAATCGTCA	CCTCGTGTCT
101	CTTATCGGAT	ACTGCGACGA	GCGAAACGAA	ATGATTCTTG	TATATGAGTA
151	CATGGAGAG	GGAACCTCTGA	AGAGTCATCT	CTATGGTTCG	GATCTTCCTA
201	ACCTCAACTG	GGAGCAGCGA	CTCGAAATCT	GCATTGGGTC	CGCACGAGGG
251	CTTCACTATC	TTCACACTGG	TTCTGCAAAG	GCCATCATTC	ATCGCGATGT
301	CAAGTCTGCA	AATATCCTTC	TGGATGAGAA	TCTCATGGCA	AAGGTGCTG
351	ATTTTGGTCT	TTCAAGACCA	GGTCTGAGT	TGGATCAGAC	TCACGTGAGC
401	ACGCGAGTTA	AAGGGAGCTT	CGGATATCTT	GATCCCGAAT	ACTTCCGCA

☐ Select all Primers

7. Tampilan setelah *Sent to Prime3Manager*

Primer3Manager
manage your primer library

[Primer3Plus](#) [Help](#)
[About](#) [Source Code](#)

Order selected Primers Refresh Reset Form Delete selected Primers

To upload or save a primer file from your local computer, choose here:
Browse... No file selected. Upload File Save File

Select	Name	Sequence	Designed on	Check! BLAST!
☒	phalPTO_F	CGTCACCTCGTGTCTTTAT	09.11.2015	Check! BLAST!
☒	phalPTO_R	TTAGGAAGATCGGAACCAT	09.11.2015	Check! BLAST!

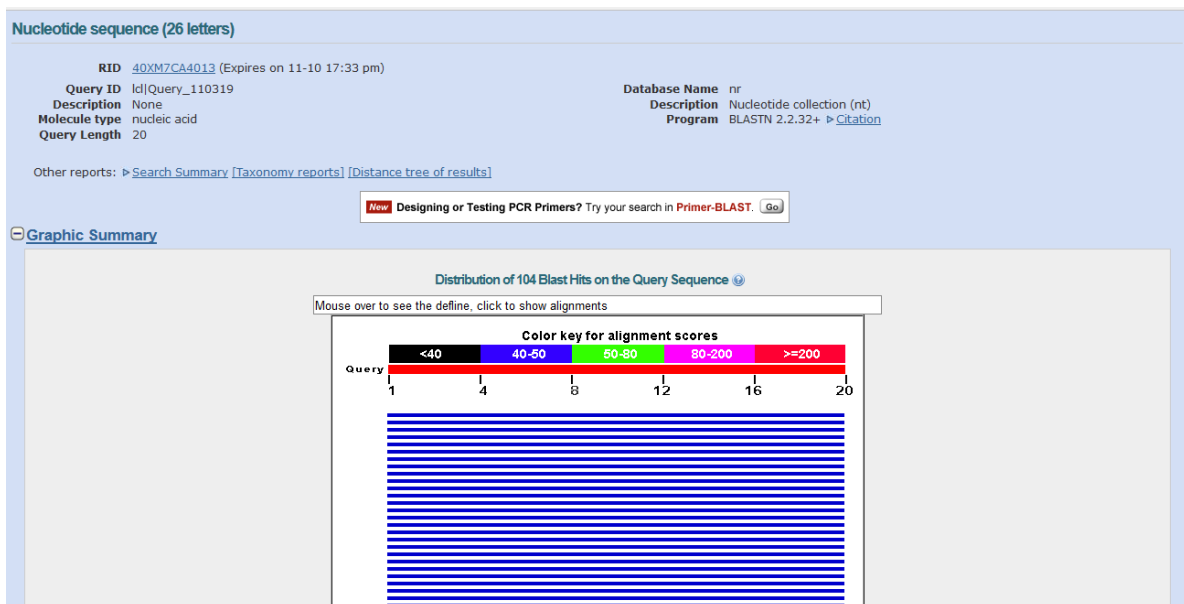
Order selected Primers Refresh Reset Form Delete selected Primers

8. Klik *Blast* di bagian paling bawah untuk masing-masing sekuens.

The screenshot shows the NCBI BLAST web interface. Annotations include:

- A red arrow pointing to the "Enter Query Sequence" text box with the label "Copy-paste sekuens gen".
- A red arrow pointing to the "Job Title" field with the label "Pilih nukleotida".
- A red arrow pointing to the "Organism" dropdown menu with the label "Pilih organisme sesuai asal contoh sekuens dari fungi atau bakteri dll.".
- A red circle around the "BLAST" button at the bottom left.
- A red circle around the "Show results in a new window" checkbox.

9. Tampilan hasil BLAST.



Blast di NCBI digunakan untuk mengetahui apakah sekuens primer yang kita desain memiliki kemiripan dengan organisme tertentu, misalnya OPT (Organisme Pengganggu Tanaman). Jika ada kemiripan, maka ganti primer, karena jika primer tersebut diamplikasikan pada tanaman yang terserang OPT tersebut, maka hasil positif dapat saja berasal dari OPT.

Catatan:

Catatan:

BAB VI

NEXT GENERATION SEQUENCING

Next-generation sequencing (NGS) adalah *platform* sekuensing yang masif dan paralel. Teknologi NGS memungkinkan proses sekuensing menjadi lebih cepat. Selain itu, jika dikalkulasi biaya per-basanya lebih murah dibandingkan dengan metode sekuensing Sanger (Wheeler *et al.* 2008). Saat ini, *platform* untuk NGS antara lain: Illumina, IonTorrent, Roche, dan BGI-Seq.

Penggunaan metode NGS dalam sekuensing telah mempercepat proses sekuensing genom manusia yang awalnya membutuhkan waktu 15 tahun, tapi sekarang dapat dilakukan dalam waktu dua bulan (Kulski 2016). Penggunaan NGS dapat diaplikasikan pada banyak hal, antara lain:

1. *Whole-Genome Sequencing* (WGS)

Teknik komprehensif yang digunakan dalam proses pengurutan sekuens DNA menjadi suatu gambaran genom utuh (*whole genome sequence*) dengan menggunakan teknologi *Next Generation Sequencing* (NGS).

2. *Whole-Exome Sequencing* (WES).

Teknik genomik untuk mengurutkan semua daerah pengkode protein gen dalam genom (dikenal sebagai *exome*). Tujuannya adalah untuk mengidentifikasi varian genetik yang mengubah sekuens protein.

3. *RNA-Sequencing*

Pengurutan (*sequencing*) asam nukleat yang memungkinkan untuk mengetahui kode genetik dari molekul RNA.

4. *Metabarcoding*

DNA *barcoding* yang dapat mengidentifikasi banyak spesies dari satu sampel yang sama. Sampel yang digunakan biasanya berupa eDNA/eRNA.

5. Epigenetik

Studi yang mempelajari bagaimana perilaku dan lingkungan individu bisa mengubah cara kerja gennya.

6. Deteksi variasi pada mutasi (*genomic surveillance* untuk mendeteksi keberadaan mutasi pada SARS-CoV-2).

Prinsip kerja NGS mirip dengan sekuensing Sanger, yaitu dengan memanfaatkan elektroforesis pada kapiler. Perbedaannya pada NGS, kapilernya disusun dalam bentuk *array* atau *plate* dengan jumlah kapiler yang banyak, sehingga proses sekuensing dilakukan secara paralel. Proses sekuensing secara paralel ini mengakibatkan waktu yang dibutuhkan untuk melakukan sekuensing pada material genetik yang berukuran besar menjadi lebih singkat.

Sebelum proses sekuensing, material genetik yang akan di-sekuens dipersiapkan melalui beberapa tahap pra-proses. Setiap *platform* memerlukan tahap pra-proses yang berbeda-beda. Pada umumnya, ada beberapa tahap yaitu:

1. Fragmentasi, yaitu material genetik dipotong-potong menjadi ukuran yang lebih kecil.
2. Ligasi adaptor dan *end-repair* yaitu persiapan agar fragmen-fragmen tersebut dapat dikenali oleh sensor kapiler pada mesin sekuenser.
3. Amplifikasi, yaitu untuk menambahkan komponen tertentu dan memperbanyak material sekuensing agar cukup untuk dibaca oleh mesin sekuenser.

Adanya fragmentasi menyebabkan hasil sekuensing dengan NGS yang dihasilkan masih berupa *raw* sekuens dengan kondisi terpotong-potong. Oleh karena itu, perlu dilakukan pengolahan data lebih lanjut untuk mendapatkan sekuens yang diinginkan. Pengolahan data yang dilakukan tergantung pada aplikasi data yang digunakan atau material genetik asal.

Data yang dihasilkan dari proses NGS umumnya cukup besar sehingga diperlukan komputer dengan kapasitas yang mumpuni untuk mengolah data tersebut. Di beberapa laboratorium, spesifikasi komputer yang digunakan yaitu minimal 8-core CPU 3GHz, 16 GB RAM, dengan kemampuan penyimpanan (*storage*) minimum 200 GB (*harddisk*). Pengolahan data hasil NGS umumnya menggunakan *software* berbasis Linux untuk melakukan *assembly*. Saat ini, selain dengan komputer *hardware* yang kita miliki, pengolahan data hasil NGS juga dapat dilakukan melalui layanan *cloud computing* yang diprakarsai oleh beberapa lembaga dunia seperti Galaxy (<http://usegalaxy.org>) dan MASER (*Management and Analysis System for Enormous Reads*) (https://cell-innovation.nig.ac.jp/maser/index_en.html).

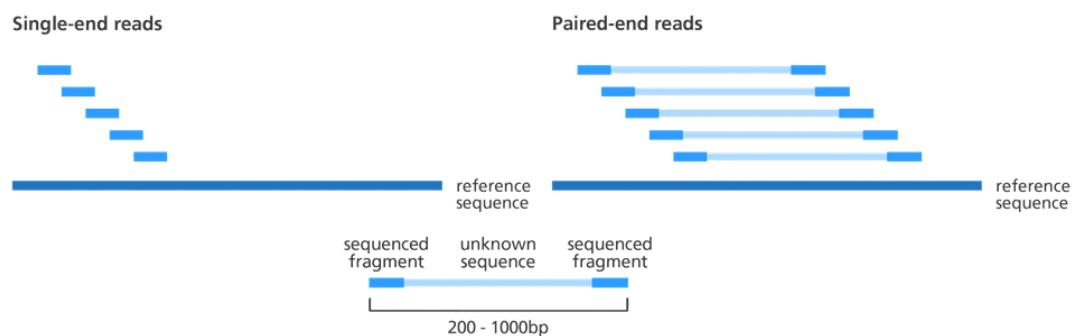
Sebelum melakukan pengolahan data, terdapat beberapa hal yang perlu diperhatikan antara lain:

1. Mesin sekuenser yang digunakan dan data yang dihasilkan dari mesin sekuenser tertentu terkadang memiliki kompatibilitas dengan *software* tertentu saja, seperti contoh di bawah ini:

<i>Sequencing platform</i>	<i>Software untuk assembly pada whole genome sequencing</i>
Illumina	Velvet (https://www.ebi.ac.uk/~zerbino/velvet/) SPAdes (http://bioinf.spbau.ru/spades)
Ion Torrent	MIRA (http://www.Chevreur.Org/projects_mira.html)
Roche 454	Newbler (http://454.com/contact-us/software-request.asp)
PacBio SMRT	SPAdes, HGAP, and the Celera-MHAP assembler

Sumber: <https://www.cd-genomics.com/bioinformatics-workflow-for-whole-genome-sequencing.html>

2. Tipe data yang dihasilkan apakah berupa *single-end* atau *paired-end*. Pada data *single-end*, mesin sekuenser membaca sekuens hanya dari satu sisi saja. Sedangkan pada *paired end*, mesin sekuenser juga membaca sekuens dari arah yang berlawanan. Gambaran perbedaan *single-end* dan *paired end* disajikan pada gambar di bawah ini



Sumber: Devon Ryan (2017); https://www.yourgenome.org/sites/default/files/images/illustrations/bioinformatics_single-end_pair-end_reads_yourgenome.png

6.1 Pengolahan data RNA-Seq dengan komputer

RNA-Sequencing atau RNA-Seq adalah salah satu aplikasi NGS untuk menganalisis sekuens RNA yang dihasilkan oleh makhluk hidup. RNA-seq dapat menghasilkan gambaran secara utuh dari transkriptom pada suatu waktu dan kondisi tertentu pada jaringan yang digunakan dan dapat digunakan untuk mengkarakterisasi dan mengkuantifikasi ekspresi gen (Wong *et al.* 2011; Song *et al.* 2016; Shih *et al.* 2018). RNA-seq dapat digunakan pada jenis-jenis non-model yang belum

memiliki *reference genome* dengan metode *De Novo* (Unamba *et al.* 2016). Adapun tahapan pengolahan data sekuensing RNA-Seq (Korpelainen *et al.* 2014) yaitu:

1. *Quality Control* dan pra-pengolahan data, yaitu proses untuk melihat kualitas *raw data* dan menghilangkan data yang berkualitas rendah.
2. *Assembly*, yaitu proses untuk menyatukan sekuens pendek yang dihasilkan menjadi sekuens transkriptom-transkriptom yang utuh.
3. *Mapping*, yaitu proses mensejajarkan hasil *reads* ke transkriptom *reference*.
4. *Annotation*, yaitu proses untuk emndapatkan informasi terkait identitas biologis suatu gen.
5. *Differential expression analysis*, yaitu analisis perbedaan ekspresi yang dihasilkan dari individu yang transkriptomnya dibandingkan.

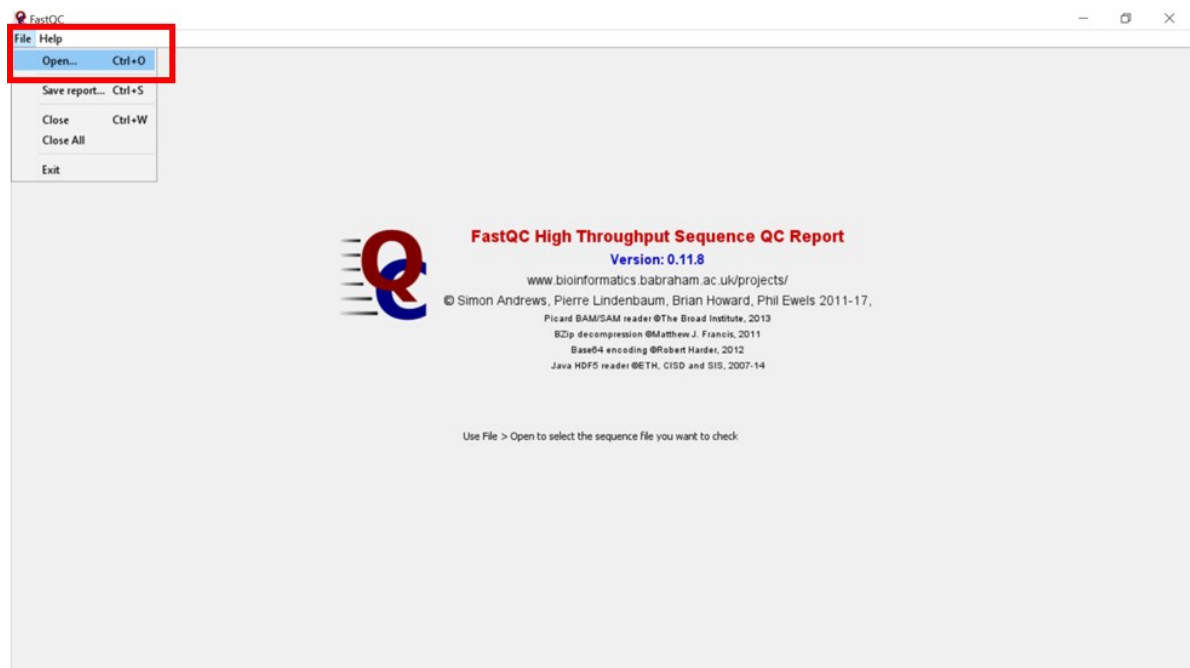
A.1. Melihat kualitas data dengan FastQC

FastQC adalah program berbasis Java yang dapat digunakan untuk melihat kualitas *raw data* yang dihasilkan (Andrew 2010). FastQC dapat diunduh pada laman <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> . Ketika akan menggunakan FastQC, pastikan komputer sudah ter-*install* Java (<https://www.java.com/en/download/manual.jsp>). FastQC dapat digunakan pada perangkat dengan sistem operasi Windows, Linux, dan Mac. Langkah-langkah penggunaan FastQC adalah sebagai berikut:

1. Pada folder FastQC, dan pilih file `run_fastqc`.

Configuration	04/10/2018 1:13 pm	File folder	
Help	04/10/2018 1:13 pm	File folder	
net	04/10/2018 1:13 pm	File folder	
org	04/10/2018 1:13 pm	File folder	
Templates	04/10/2018 1:13 pm	File folder	
uk	04/10/2018 1:13 pm	File folder	
cisd-jhdf5	21/12/2017 1:50 pm	Executable Jar File	9,051 KB
fastqc	04/10/2018 12:35 pm	File	15 KB
fastqc_icon	21/12/2017 1:50 pm	Icon	3 KB
INSTALL	21/12/2017 1:50 pm	Text Document	7 KB
jbzip2-0.9	21/12/2017 1:50 pm	Executable Jar File	49 KB
LICENSE	21/12/2017 1:50 pm	File	18 KB
LICENSE	21/12/2017 1:50 pm	Text Document	35 KB
LICENSE_JHDF5	21/12/2017 1:50 pm	Text Document	12 KB
README.md	21/12/2017 3:21 pm	MD File	2 KB
README	21/12/2017 3:21 pm	Text Document	3 KB
RELEASE_NOTES	04/10/2018 1:10 pm	Text Document	37 KB
run_fastqc	26/04/2018 11:48 am	Windows Batch File	1 KB
sam-1.103	21/12/2017 1:51 pm	Executable Jar File	630 KB

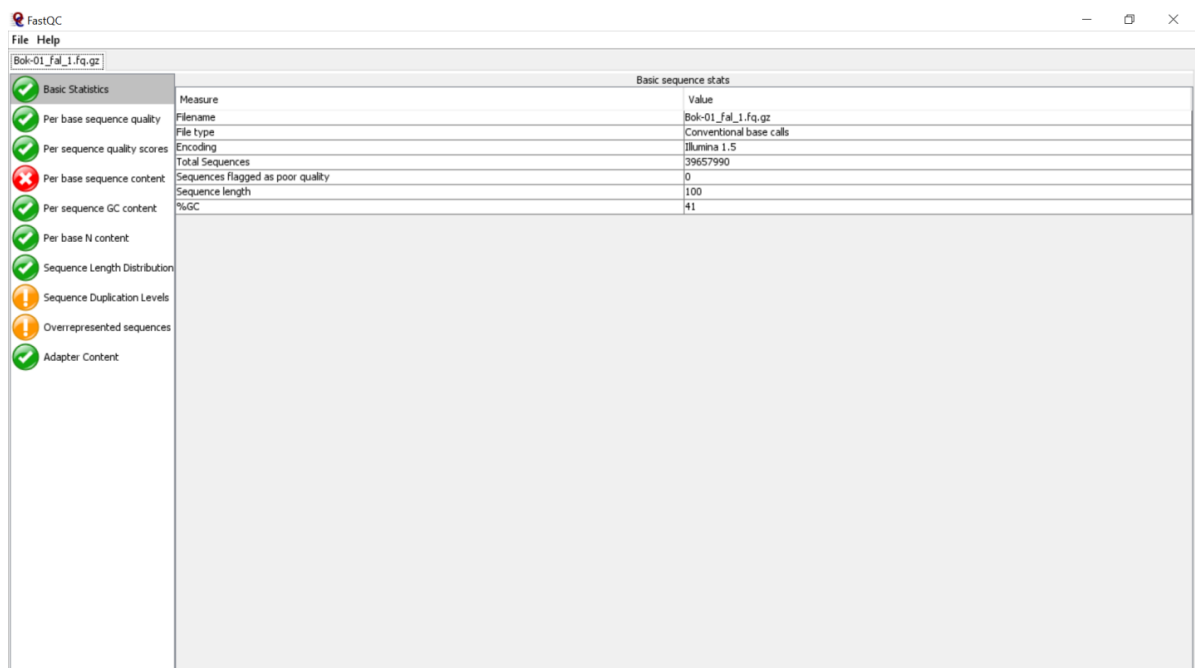
2. Unggah file *raw sequence* (dalam bentuk fastq.zip) dengan memilih File → Open



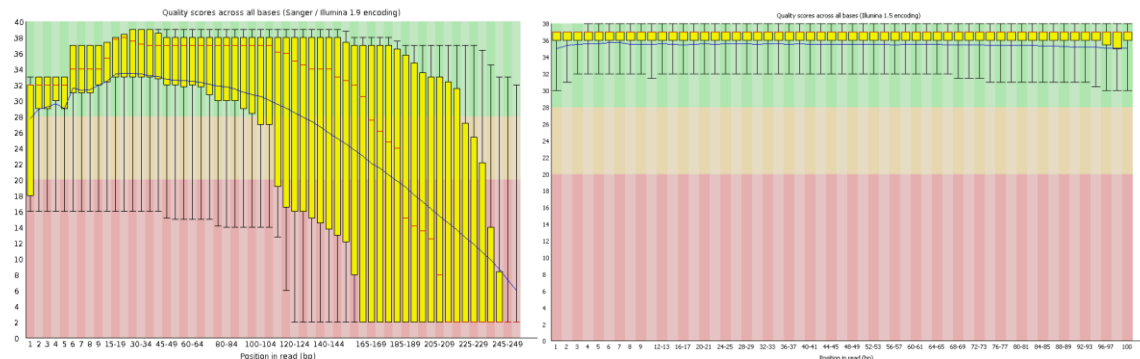
3. Untuk satu sampel terdapat 2 file dan keduanya perlu dianalisis satu persatu.

	Bok-01_fal_1.fq	19/10/2018 1:51 pm	Arsip WinRAR	3,070,355 KB
	Bok-01_fal_2.fq	19/10/2018 1:48 pm	Arsip WinRAR	3,089,474 KB

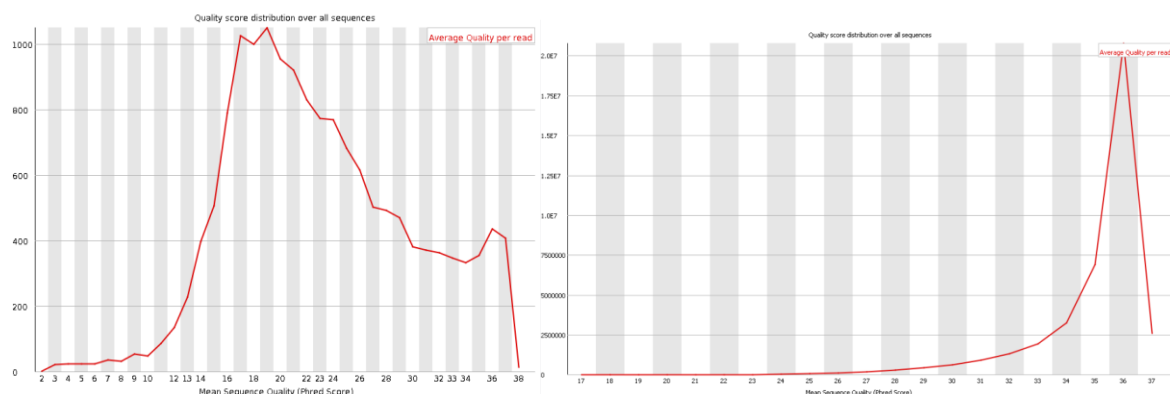
4. Pilih salah satu file dan akan muncul halaman utama hasil.



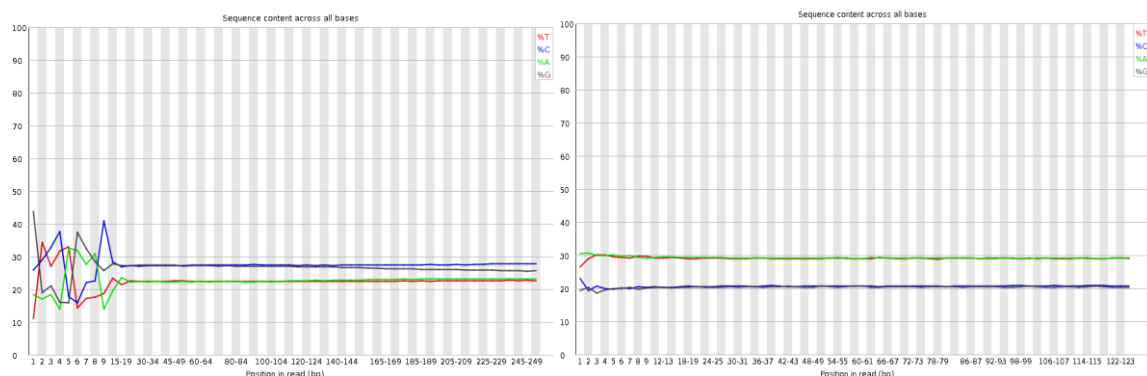
- Contoh kualitas basa rendah (kiri) dan contoh kualitas basa baik (kanan) pada menu *per base sequence quality*. Raw sekuens yg cukup baik yaitu ketika sebagian besar berada pada skor 28-40.



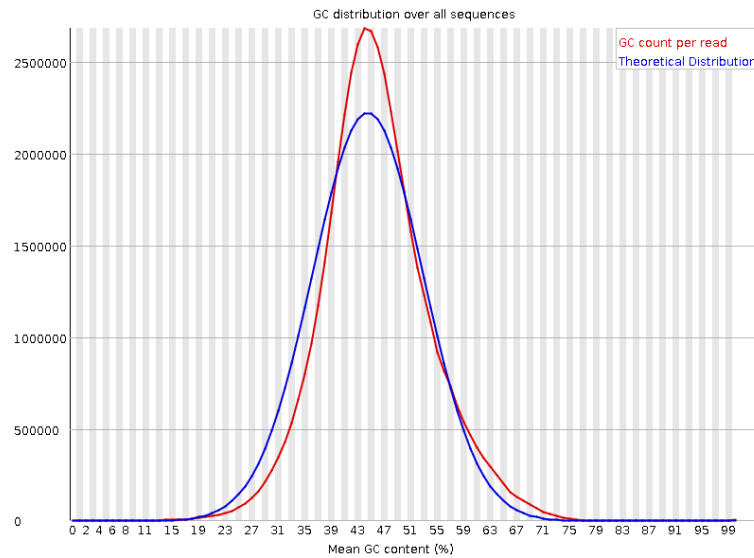
- Contoh kualitas sekuens rendah (kiri) dan contoh kualitas sekuens baik (kanan) pada menu *per sequence quality scores*. Sekuens yang baik rata-rata skornya cenderung berada di ujung kanan pada grafik.



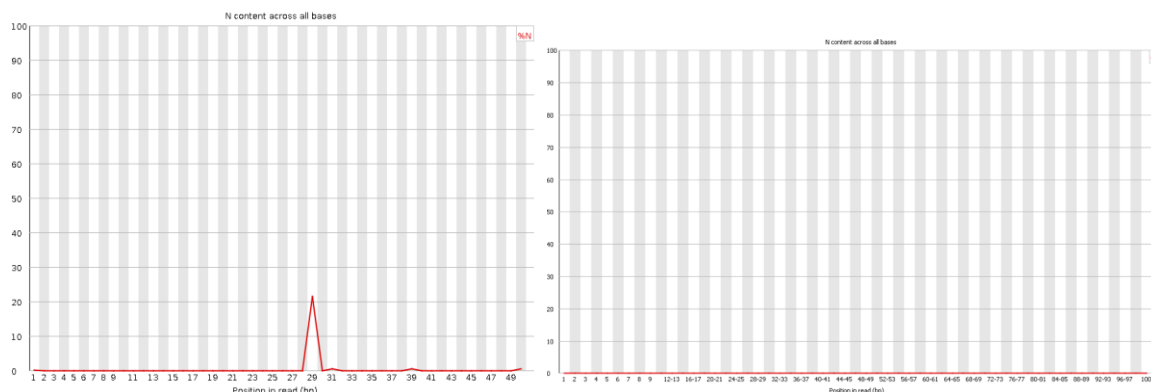
- Contoh kualitas sekuens rendah (kiri) dan contoh kualitas sekuens baik (kanan) pada menu *per base sequence content*. Terkadang meskipun hasil analisis dengan FastQC menunjukkan grafik seperti gambar yang di kiri, tetapi sekuens dianggap normal karena pada metode preparasi pra-sekuensing dilakukan penambahan primer untuk *barcode* pada mesin, yang menyebabkan 10-15 basa pertama memiliki distribusi yang tidak normal.



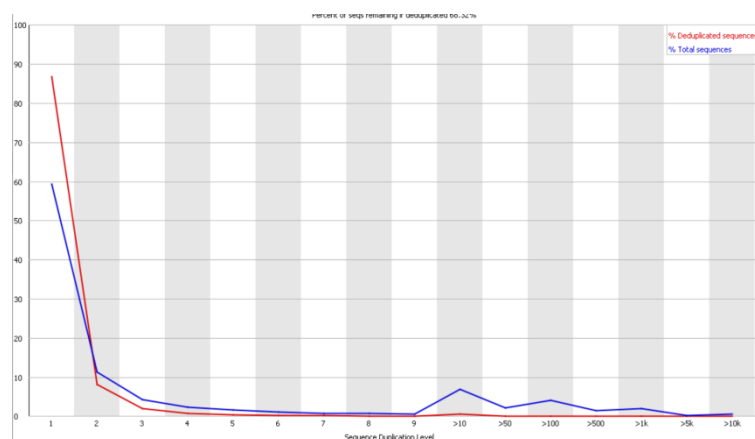
8. GC content yang baik tidak akan menyimpang terlalu jauh dari distribusi normal.



9. Jumlah basa yang tidak bisa diidentifikasi dapat dilihat pada menu *per base N content*. Adanya *spike* menunjukkan adanya *error* saat proses sekuensing. Gambar di sebelah kiri menunjukkan terjadinya kegagalan instrumen dalam membaca 20% basa pada posisi 29.



10. Pada RNA-seq, adanya duplikasi pada hasil pembacaan merupakan hal yang normal, karena adanya ekspresi dari transkrip-transkrip tertentu yang lebih tinggi jika dibandingkan dengan *whole genome sequencing* (diharapkan sekuens yang terbaca adalah sekuens yang unik dan tidak terduplikasi). Oleh karena itu, jika pada *sequence duplication levels* atau *overrepresented sequence* terdapat *warning*, pengolahan data dapat dilanjutkan.



11. Pada *adapter content*, idealnya tidak ada sekuens *adapter* yang ikut terbaca. Akan tetapi, karena *library* yang terbentuk sebelum di-sekuens panjangnya bermacam-macam, maka terkadang *adapter* ikut terbaca.

A.2. *Trimming* atau pemotongan dengan *trimmomatic*

Trimmomatic adalah salah satu tools yang digunakan untuk memfilter sekuens berdasarkan kualitas yang diinginkan. *Trimmomatic* digunakan pada proses sekuensing menggunakan platform Illumina (Bolger *et al* 2014). *Trimmomatic* dapat dijalankan pada *OS Linux* melalui Terminal maupun *Windows* melalui *Command Prompt*. *Command* yang digunakan untuk melakukan filter pada basa-basa dengan kualitas rendah menggunakan *Trimmomatic* yaitu:

```
java -jar trimmomatic-0.32.jar PE -phred64 reads1.fastq.gz reads2.fastq.gz paired1.fq.gz
unpaired1.fq.gz paired2.fq.gz unpaired2.fq.gz avgqual:20
```

dengan input berupa:

file reads1/2.fastq.gz, output yaitu paired1/2.fq.gz, dan data yang terbuang unpaired1/2.fq.gz

A.3. *Assembly* dengan metode *de novo*

Assembly adalah proses penggabungan *reads* menjadi *full-length transkriptom* atau RNA yang utuh. Karena jumlah spesies yang sudah memiliki *reference* genom masih terbatas, terkadang proses *assembly* dilakukan secara *de novo*. Untuk mendapatkan *full length transkriptom* yang ideal, proses *assembly* dilakukan dengan beberapa tahap:

1. *Assembly* menggunakan *trinity*. *Trinity* adalah *software* yang mengkonstruksi *contig*, mengelompokkan *contig* yang terbentuk, serta mengekstrak isoform yang terbentuk dari *contig-contig* tersebut (Grabherr *et al.* 2013). *Trinity* dapat dijalankan pada terminal *Linux* dengan *command*:

```
Trinity.pl --seqType fq --left chr18_1.fq --right chr18_2.fq --CPU 4
```

Input berupa file dengan ekstensi .fq. *Left sequence* adalah sekuens dengan nomor 1, *right sequence* adalah sekuens dengan kode 2. *Command* tersebut dapat dimodifikasi dengan menambahkan memori komputer maksimal yang digunakan (*--max_memory* XXG *--*), minimal panjang *contig* (*--min_contig_length* XXX), mencegah terdapat file besar yang tertinggal (*--full_cleanup*), dan nama file yang menjadi output (*--output namafile*).

2. Untuk mengurangi kelompok transkrip yang hampir sama dan memperoleh *contig* yang lebih lengkap, *contig-contig* yang terbentuk kemudian di-*reassembly* menggunakan *software* CAP3 (Huang & Madan 1999) dan CD-HIT-EST (Li & Godzik 2006) dengan *identity threshold* 90% untuk menghilangkan *contig-contig* yang ambigu. *Command* untuk CAP3 dan CDHIT:

```
cap3 namafileoutputTrinity.fasta -p 90 > namafileoutput.cap3
```

```
cd-hit-est -I namafileoutput.cap3 -o namafileoutput.cdhit -c 0.90 -M 0 -T 0 > cdhit report
```

3. Memetakan *reads* ke *reference* menggunakan bowtie2 (Langmead & Salzberg 2012), membuat referensi:

```
bowtie2-build namafileoutput.cdhit namafileoutput.cdhit.index
```

proses *mapping*:

```
bowtie2 -p 22 -x namafileoutput.cdhit.index -1 leftsequence.fq -2 rightsequence.fq -S namafileoutput.cdhit.sam
```

dilanjutkan menggunakan *samtools* untuk mengubah ke format *gene expression*:

```
samtools view -S -b namafileoutput.cdhit.sam > namafileoutput.cdhit.bam
```

dilanjutkan dengan *Corset* dengan nilai ekspresi dalam *Counts per Million* (CPM) >1 (Davidson & Oshlack 2014)

```
corset -r true-stop namafileoutput.cdhit.bam
```

```
corset -m 0 -I corset namafileoutput.cdhit.bam.corset-reads
```

4. *Contig* yang diperoleh tersebut kemudian dibandingkan atau dianotasikan dengan sekuens dari beberapa basis data, antara lain dengan *non-redundant* protein (nr) NCBI dan SwissProt menggunakan algoritma *Basic Local alignment Search Tools* (BLAST) dengan *cut-off E-value* 1E-5 (Altschul *et al.* 1990). Identifikasi kemiripan sekuens dari beberapa *contig* dalam grup yang homolog dan deteksi *contig* dengan *Open Reading Frames* (ORFs) dilakukan dengan perangkat BLASTX pada *NCBI ORF Finder*. *Database* yang digunakan untuk BLAST dapat diunduh dari NCBI menggunakan FTP.

Command pada program BLAST yang telah dijalankan di Terminal Linux:

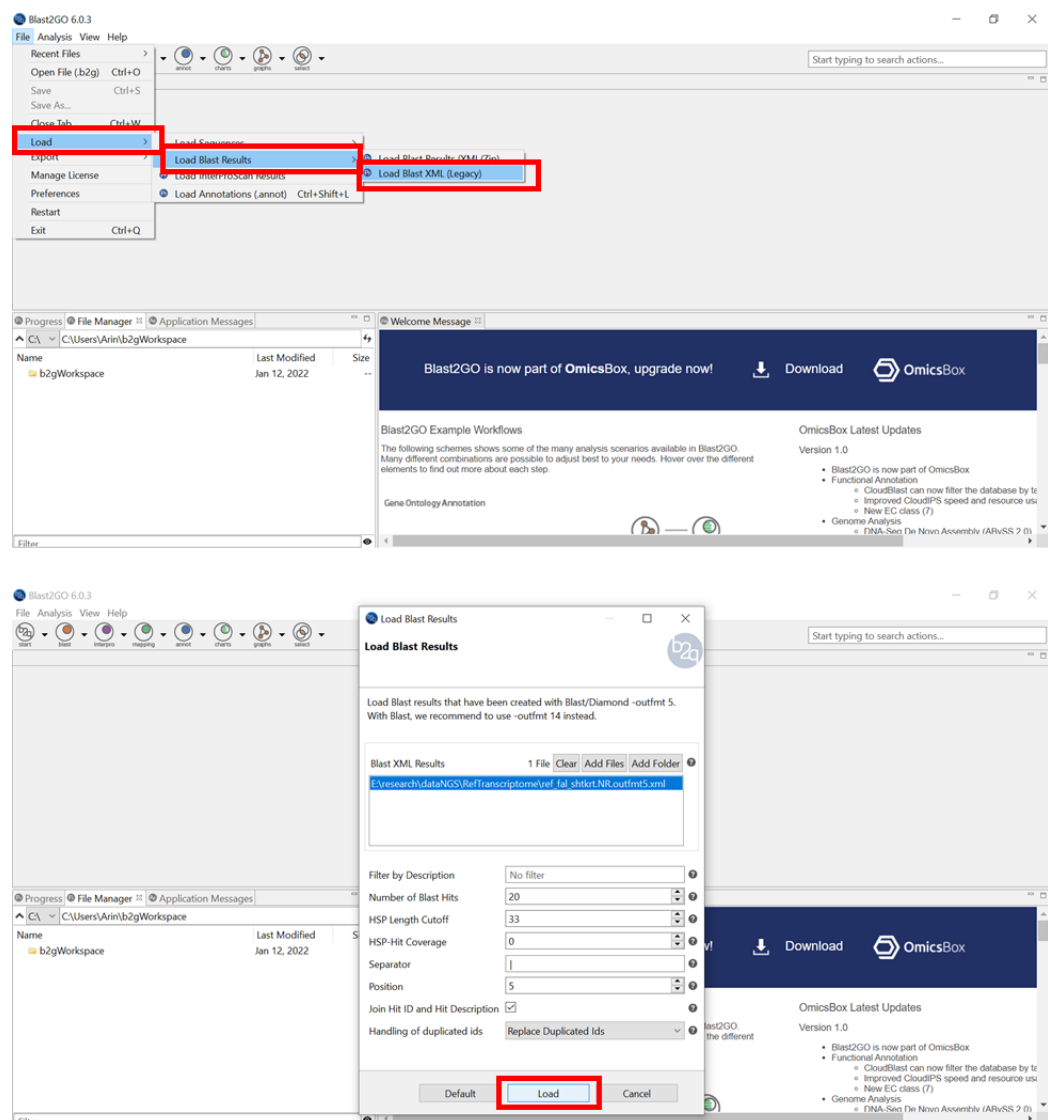
```
Blastx -query namafiletranskriptomhasil.fasta -db /pathway menuju file database (contoh SwissProtDB)
```


Blastx -query namafiletranskriptom.fasta -db /D/database/protein/swissprotDB -num_threads 23 -max_target_seqs 1 -outfmt5 e -evalue 1e-5 -max_hsps 1 -out namafiletranskriptom.swissprot.outfmt5 (untuk file blast2go)

blastx -query namafiletranskriptom.fasta -db /D/database/protein/swissprotDB -num_threads 23 -max_target_seqs 1 -outfmt 6 "6 qseqid sseqid qstart qend pident length evalue bitscore score stitle" -evalue 1e-5 -max_hsps 1 -out ref_fal_shtkrt.SP.outfmt6 (untuk digabungkan dengan file DEG)

5. Analisis *Gene Ontology* dan *KEGG pathway* dilakukan menggunakan perangkat lunak Blast2GO (Conesa *et al.* 2005, Götz *et al.* 2008).

a. Load file hasil blast (file dengan ekstensi xml) di Blast2go:



b. Lakukan *mapping*. Sekuens yang ter-*mapping* akan berwarna hijau.

Blast2GO 6.0.3

File Analysis View Help

Start typing to search actions...

Table: ref_fal_shkrt.NR.outfmt5

Run GO Mapping Ctrl+Shift+M

Id	Test	Description	Length	Bits	E-value	Sim mean	MSD	GO IDs	GO Names	Enzyme Codes	Enzyme Names	InterPro IDs	InterPro GO IDs	InterPro GO Names
1	NO-BLAST	Fa0K1080.0.7.81... -Nk...	749											
2	BLASTED	Fa0K1080.0.7.81... Acute-induced prote...	1198	1	7.26E-137	78.72%								
3	BLASTED	Fa0K1080.0.7.791... hypothetical prote...	1016	1	6.4E-73	78.82%								
4	NO-BLAST	Fa0K1080.0.7.81... -Nk...	1117											
5	BLASTED	Fa0K1080.0.7.783... hypothetical prote...	911	1	8.25E-49	78.92%								
6	BLASTED	Fa0K11950.0.7.81... transmembrane 9 su...	464	1	4.67E-59	91.59%								
7	BLASTED	Fa0K18951.0.7.78... hypothetical prote...	924	1	6.69E-33	90.14%								
8	NO-BLAST	Fa0K1080.0.7.79... -Nk...	848											
9	BLASTED	Fa0K1080.0.7.73... UBR2_3 domain-con...	821	1	1.42E-12	83.25%								
10	BLASTED	Fa0K1080.0.7.73... uncharacterized prot...	1039	1	4.38E-12	47.43%								
11	NO-BLAST	Fa0K1080.0.7.81... -Nk...	808											
12	BLASTED	Fa0K1080.0.7.73... uncharacterized prot...	752	1	1.31E-150	92.83%								
13	BLASTED	Fa0K1080.0.7.73... hypothetical prote...	1830	1	1.3E-151	79.57%								
14	NO-BLAST	Fa0K1080.0.7.73... -Nk...	1264											

Progress: File Manager Application Messages

100% Mapping (ref_fal_shkrt.NR.outfmt5): done [1m41s]

100% Load Blast Results: done [34s]

Welcome Message

Blast2GO is now part of OmicsBox, upgrade now! Download OmicsBox

Blast2GO Example Workflows

The following schemes shows some of the many analysis scenarios available in Blast2GO. Many different combinations are possible to adjust best to your needs. Hover over the different elements to find out more about each step.

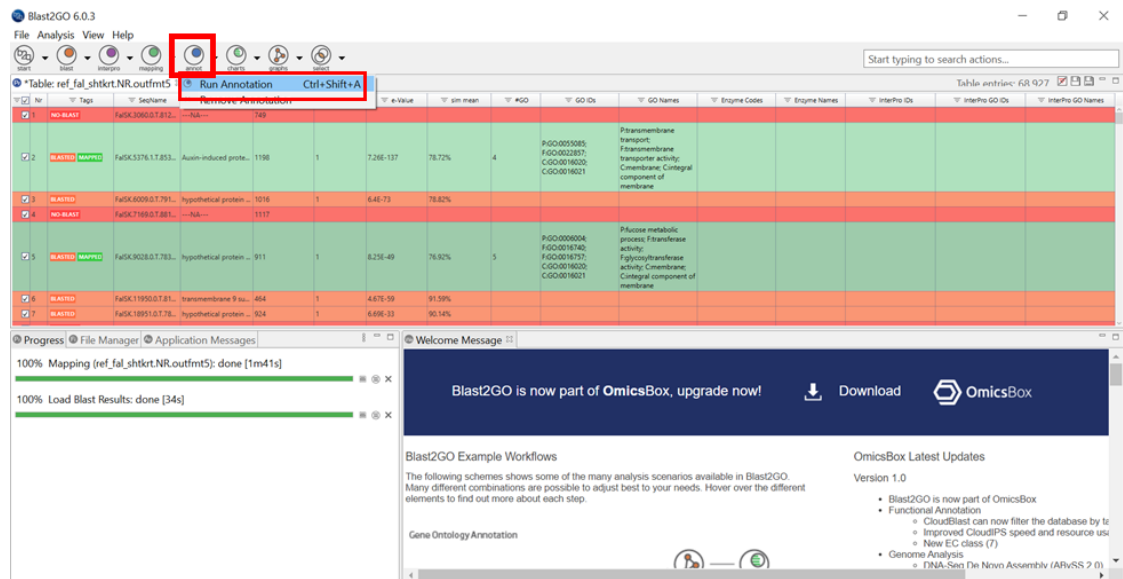
Gene Ontology Annotation

OmicsBox Latest Updates

Version 1.0

- Blast2GO is now part of OmicsBox
- Functional Annotation
 - CloudBlast can now filter the database by tax
 - Improved CloudIPS speed and resource use
 - New EC class (7)
- Genome Analysis
 - DNA-Seq The Nvsn Assembly (ARvSS 2.0)

c. Lakukan *Annotation*. Sekuens yang terannotasi di Blast2go akan berwarna biru.



Blast2GO 6.0.3

File Analysis View Help

Start typing to search actions...

Table: ref_fal_shtkrt.NR.outfmt5

Remove Annotation Ctrl+Shift+A

Id	Tags	SeqName	Description	Length	bits	e-value	sim mean	AGO	GO IDs	GO Names	Enzyme Codes	Enzyme Names	InterPro IDs	InterPro GO IDs	InterPro GO Names
1	NO-BLAST	FaSK3060.0.7.812...	...	749											
2	BLASTED	FaSK3376.1.7.853...	Axin-induced prote...	1198	1	7.26E-137	78.72%	4	P:GO:0055085 P:GO:0022857 C:GO:0016030 C:GO:0016021	Transmembrane transport; Transmembrane transporter activity; Membrane; Integral component of membrane					
3	BLASTED	FaSK6009.0.7.791...	hypothetical protein...	1016	1	6.4E-73	78.82%								
4	NO-BLAST	FaSK7169.0.7.881...	...	1117											
5	BLASTED	FaSK9020.0.7.783...	hypothetical protein...	911	1	8.25E-49	76.92%	3	P:GO:0006004 P:GO:0016740 P:GO:0016757 C:GO:0016030 C:GO:0016021	Flucose metabolic process; Transferase activity; Fatty acid transferase activity; Membrane; Integral component of membrane					
6	BLASTED	FaSK11950.0.7.81...	transmembrane 9 su...	464	1	4.67E-59	91.59%								
7	BLASTED	FaSK18951.0.7.78...	hypothetical protein...	924	1	6.69E-33	90.14%								

Progress: File Manager Application Messages

100% Mapping (ref_fal_shtkrt.NR.outfmt5): done [1m41s]

100% Load Blast Results: done [34s]

Welcome Message

Blast2GO is now part of OmicsBox, upgrade now! Download OmicsBox

Blast2GO Example Workflows

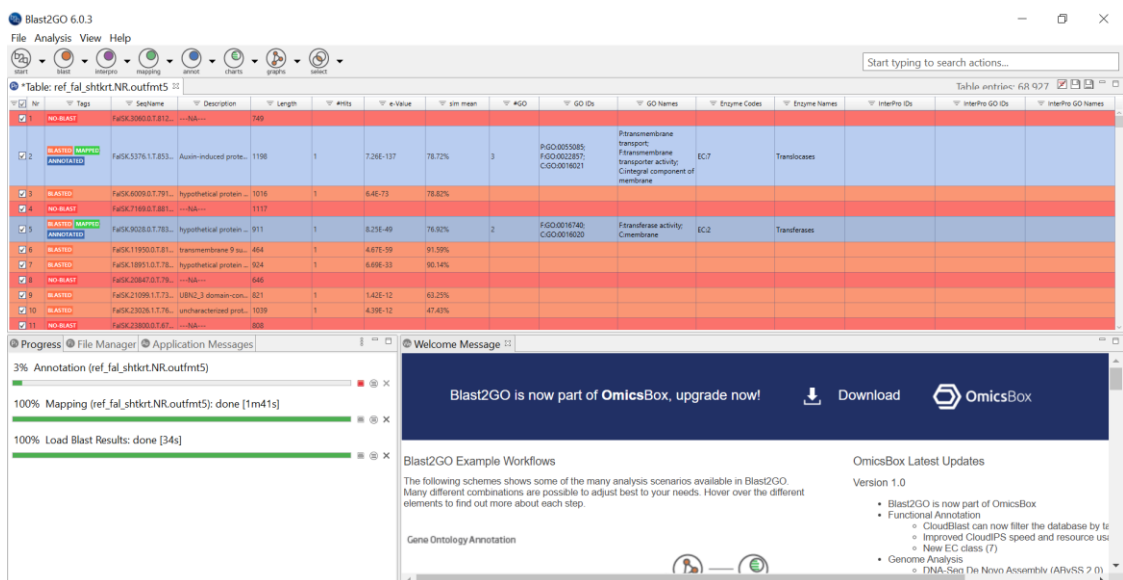
The following schemes shows some of the many analysis scenarios available in Blast2GO. Many different combinations are possible to adjust best to your needs. Hover over the different elements to find out more about each step.

Gene Ontology Annotation

OmicsBox Latest Updates

Version 1.0

- Blast2GO is now part of OmicsBox
- Functional Annotation
 - CloudBlast can now filter the database by tax
 - Improved CloudIPS speed and resource use
 - New EC class (7)
- Genome Analysis
 - RNA-Seq De Novo Assembly (ARVSS 2.0)



Blast2GO 6.0.3

File Analysis View Help

Start typing to search actions...

Table: ref_fal_shtkrt.NR.outfmt5

Remove Annotation Ctrl+Shift+A

Id	Tags	SeqName	Description	Length	bits	e-value	sim mean	AGO	GO IDs	GO Names	Enzyme Codes	Enzyme Names	InterPro IDs	InterPro GO IDs	InterPro GO Names
1	NO-BLAST	FaSK3060.0.7.812...	...	749											
2	BLASTED	FaSK3376.1.7.853...	Axin-induced prote...	1198	1	7.26E-137	78.72%	3	P:GO:0055085 P:GO:0022857 C:GO:0016030 C:GO:0016021	Transmembrane transport; Transmembrane transporter activity; Membrane; Integral component of membrane	EC:7	Translocases			
3	BLASTED	FaSK6009.0.7.791...	hypothetical protein...	1016	1	6.4E-73	78.82%								
4	NO-BLAST	FaSK7169.0.7.881...	...	1117											
5	BLASTED	FaSK9020.0.7.783...	hypothetical protein...	911	1	8.25E-49	76.92%	2	P:GO:0016740 C:GO:0016030	Transferase activity; Membrane	EC:2	Transferases			
6	BLASTED	FaSK11950.0.7.81...	transmembrane 9 su...	464	1	4.67E-59	91.59%								
7	BLASTED	FaSK18951.0.7.78...	hypothetical protein...	924	1	6.69E-33	90.14%								
8	NO-BLAST	FaSK20847.0.7.78...	...	646											
9	BLASTED	FaSK21099.1.7.73...	UBN3 domain-con...	821	1	1.42E-12	63.25%								
10	BLASTED	FaSK23026.1.7.76...	uncharacterized prot...	1039	1	4.39E-12	47.43%								
11	NO-BLAST	FaSK23800.0.7.87...	...	808											

Progress: File Manager Application Messages

3% Annotation (ref_fal_shtkrt.NR.outfmt5)

100% Mapping (ref_fal_shtkrt.NR.outfmt5): done [1m41s]

100% Load Blast Results: done [34s]

Welcome Message

Blast2GO is now part of OmicsBox, upgrade now! Download OmicsBox

Blast2GO Example Workflows

The following schemes shows some of the many analysis scenarios available in Blast2GO. Many different combinations are possible to adjust best to your needs. Hover over the different elements to find out more about each step.

Gene Ontology Annotation

OmicsBox Latest Updates

Version 1.0

- Blast2GO is now part of OmicsBox
- Functional Annotation
 - CloudBlast can now filter the database by tax
 - Improved CloudIPS speed and resource use
 - New EC class (7)
- Genome Analysis
 - RNA-Seq De Novo Assembly (ARVSS 2.0)

- d. Analisis gen *Ontology* dilakukan untuk mengetahui letak, proses biologis, dan fungsi molekuler dari gen yang terannotasi pada transkriptom yang sudah ter-assembly. Sekuens yang berhasil dianalisis gen ontology-nya akan berwarna krem.

Blast2GO 6.0.3

File Analysis View Help

GO-Slim Run GO-Slim Ctrl+Shift+G

Enzyme Code and KEGG Enrichment Analysis (Fisher's Exact Test)

Remove GO-Slim

	Enzyme Code and KEGG	Enrichment Analysis (Fisher's Exact Test)	Length	#hits	e-value	sim mean	#GO	GO IDs	GO Names	Enzyme Codes	Enzyme Names	InterPro IDs	InterPro GO IDs	InterPro GO Names
1	NO BLAST	FaSK30602.0.7.812...-Nik...	749											
2	BLASTED	FaSK3376.1.7.853...-Nik...	1198	1	7.26E-137	78.72%	3	PSO005085; PSO002807; CGO0018021	Transmembrane transport; Transmembrane transporter activity; Integral component of membrane	EC7	Translocases			
3	BLASTED	FaSK4009.0.7.791...-Nik...	1016	1	6.4E-73	78.82%								
4	BLASTED	FaSK7189.0.7.881...-Nik...	1117											
5	BLASTED	FaSK3020.0.7.783...-Nik...	911	1	8.23E-49	76.92%	2	PSO0016740; CGO0018023	Transferase activity; Crenolase	EC2	Transferases			
6	BLASTED	FaSK11950.0.7.81...-Nik...	464	1	4.67E-59	91.59%								
7	BLASTED	FaSK18951.0.7.78...-Nik...	924	1	6.69E-33	90.14%								
8	NO BLAST	FaSK20847.0.7.79...-Nik...	646											
9	BLASTED	FaSK21099.1.7.73...-Nik...	821	1	1.43E-12	83.23%								
10	BLASTED	FaSK23026.1.7.76...-Nik...	1039	1	4.39E-12	47.43%								
11	NO BLAST	FaSK23800.0.7.67...-Nik...	808											

Progress: File Manager Application Messages

100% Annotation (ref_fal_shkrt.NR.outfmt5): done [20s]

100% Mapping (ref_fal_shkrt.NR.outfmt5): done [1m41s]

100% Load Blast Results: done [34s]

Welcome Message

Blast2GO is now part of OmicsBox, upgrade now! Download OmicsBox

Blast2GO Example Workflows

The following schemes shows some of the many analysis scenarios available in Blast2GO. Many different combinations are possible to adjust best to your needs. Hover over the different elements to find out more about each step.

Gene Ontology Annotation

OmicsBox Latest Updates

Version 1.0

- Blast2GO is now part of OmicsBox
- Functional Annotation
 - CloudBlast can now filter the database by taxonomic group
 - Improved CloudBlast speed and resource usage
 - New EC class (7)
- Genome Analysis
 - DNA-Seq De Novo Assembly (ARVSS 2.0)

Blast2GO 6.0.3

File Analysis View Help

GO-Slim Run GO-Slim Ctrl+Shift+G

Enzyme Code and KEGG Enrichment Analysis (Fisher's Exact Test)

Remove GO-Slim

	Enzyme Code and KEGG	Enrichment Analysis (Fisher's Exact Test)	Length	#hits	e-value	sim mean	#GO	GO IDs	GO Names	Enzyme Codes	Enzyme Names	InterPro IDs	InterPro GO IDs	InterPro GO Names
1	NO BLAST	FaSK30602.0.7.812...-Nik...	749											
2	BLASTED	FaSK3376.1.7.853...-Nik...	1198	1	7.26E-137	78.72%	2	PSO005085; PSO002807; CGO0018021	Transmembrane transport; Transmembrane transporter activity	EC7	Translocases			
3	BLASTED	FaSK4009.0.7.791...-Nik...	1016	1	6.4E-73	78.82%								
4	BLASTED	FaSK7189.0.7.881...-Nik...	1117											
5	BLASTED	FaSK3020.0.7.783...-Nik...	911	1	8.23E-49	76.92%	1	PSO0016740	Transferase activity	EC2	Transferases			
6	BLASTED	FaSK11950.0.7.81...-Nik...	464	1	4.67E-59	91.59%								
7	BLASTED	FaSK18951.0.7.78...-Nik...	924	1	6.69E-33	90.14%								
8	NO BLAST	FaSK20847.0.7.79...-Nik...	646											
9	BLASTED	FaSK21099.1.7.73...-Nik...	821	1	1.43E-12	83.23%								
10	BLASTED	FaSK23026.1.7.76...-Nik...	1039	1	4.39E-12	47.43%								
11	NO BLAST	FaSK23800.0.7.67...-Nik...	808											

Progress: File Manager Application Messages

100% GO-Slim (ref_fal_shkrt.NR.outfmt5): done [11s]

100% Annotation (ref_fal_shkrt.NR.outfmt5): done [20s]

100% Mapping (ref_fal_shkrt.NR.outfmt5): done [1m41s]

100% Load Blast Results: done [34s]

Welcome Message

Blast2GO is now part of OmicsBox, upgrade now! Download OmicsBox

Blast2GO Example Workflows

The following schemes shows some of the many analysis scenarios available in Blast2GO. Many different combinations are possible to adjust best to your needs. Hover over the different elements to find out more about each step.

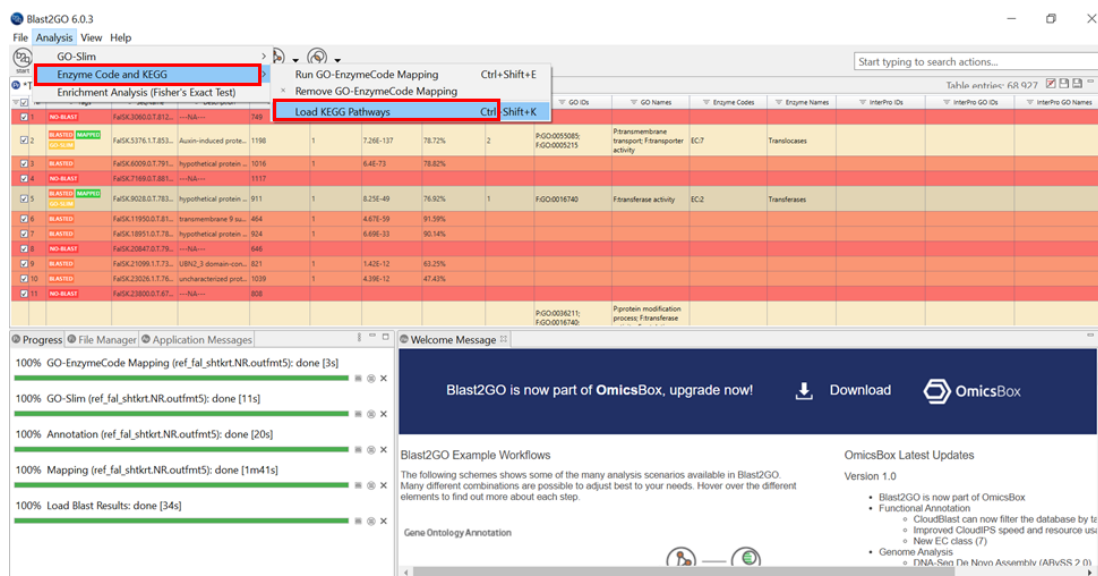
Gene Ontology Annotation

OmicsBox Latest Updates

Version 1.0

- Blast2GO is now part of OmicsBox
- Functional Annotation
 - CloudBlast can now filter the database by taxonomic group
 - Improved CloudBlast speed and resource usage
 - New EC class (7)
- Genome Analysis
 - DNA-Seq De Novo Assembly (ARVSS 2.0)

- e. *Database KEGG pathway* adalah salah satu bagian dari KEGG database yang berisi kumpulan peta *pathway* yang merepresentasikan interaksi molekuler dalam proses metabolisme, proses informasi genetik, proses informasi lingkungan, proses seluler, sistem organisme, penyakit, dan pembuatan obat-obatan. KEGG (*Kyoto Encyclopedia of Genes and Genomes*) adalah database yang berisi data-data terkait *high-level function* dan kegunaan suatu sistem biologi seperti sel, organisme, dan ekosistem berdasarkan informasi molekuler, khususnya *dataset* besar hasil sekuensing



Blast2GO 6.0.3

File Analysis View Help

GO-Slim

Enzyme Code and KEGG

Run GO-EnzymeCode Mapping Ctrl+Shift+E

Remove GO-EnzymeCode Mapping

Load KEGG Pathways Ctrl+Shift+K

Table entries: 68 927

GO IDs GO Names Enzyme Codes Enzyme Names InterPro IDs InterPro GO Names

1. BLASTED MAPRED GO-Slim FalsK3376.1.T.851... Asin-induced prote... 1198 1 7.26E-137 78.72% 2 PGO0055085 FGO0055215 Transmembrane transport Transporter activity EC:7 Translocases

2. BLASTED MAPRED GO-Slim FalsK4009.0.T.791... hypothetical protein ... 1016 1 6.4E-73 78.82% PGO0016740 Transferrase activity EC:2 Transferrases

3. BLASTED MAPRED GO-Slim FalsK7169.0.T.881... ---Nk--- 1117 1 8.25E-49 76.92% 1 FGO0016740 Transferrase activity EC:2 Transferrases

4. BLASTED MAPRED GO-Slim FalsK3002.0.T.783... hypothetical protein ... 911 1 4.67E-59 91.59% PGO0016740 Transferrase activity EC:2 Transferrases

5. BLASTED MAPRED GO-Slim FalsK11952.0.T.81... transmembrane 5 su... 464 1 6.69E-31 90.14% PGO0016740 Transferrase activity EC:2 Transferrases

6. BLASTED MAPRED GO-Slim FalsK1891.0.T.73... hypothetical protein ... 924 1 1.43E-12 83.25% PGO0016740 Transferrase activity EC:2 Transferrases

7. BLASTED MAPRED GO-Slim FalsK20847.0.T.79... ---Nk--- 648 1 4.39E-12 47.43% PGO0016740 Transferrase activity EC:2 Transferrases

8. BLASTED MAPRED GO-Slim FalsK21099.1.T.73... UBR2_3 domain-con... 821 1 4.39E-12 47.43% PGO0016740 Transferrase activity EC:2 Transferrases

9. BLASTED MAPRED GO-Slim FalsK23026.1.T.73... uncharacterized prot... 1039 1 4.39E-12 47.43% PGO0016740 Transferrase activity EC:2 Transferrases

10. BLASTED MAPRED GO-Slim FalsK23805.0.T.67... ---Nk--- 808 1 4.39E-12 47.43% PGO0016740 Transferrase activity EC:2 Transferrases

11. BLASTED MAPRED GO-Slim FalsK23805.0.T.67... ---Nk--- 808 1 4.39E-12 47.43% PGO0016740 Transferrase activity EC:2 Transferrases

Progress: File Manager Application Messages

100% GO-EnzymeCode Mapping (ref_fal_shtkrt.NR.outfmt5): done [3s]

100% GO-Slim (ref_fal_shtkrt.NR.outfmt5): done [11s]

100% Annotation (ref_fal_shtkrt.NR.outfmt5): done [20s]

100% Mapping (ref_fal_shtkrt.NR.outfmt5): done [1m41s]

100% Load Blast Results: done [34s]

Welcome Message: Blast2GO is now part of OmicsBox, upgrade now! Download OmicsBox

Blast2GO Example Workflows

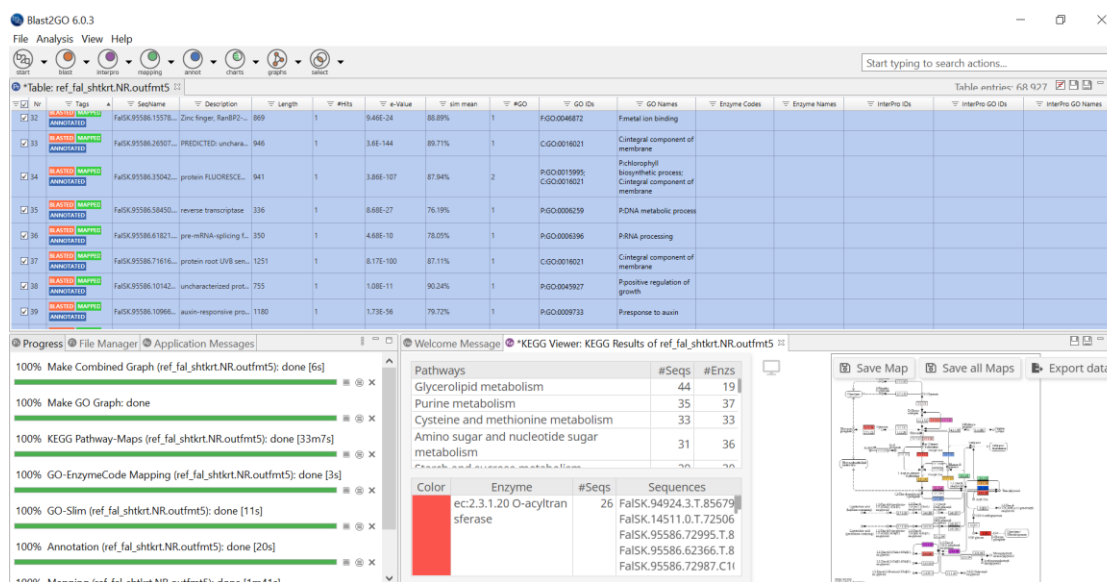
The following schemes shows some of the many analysis scenarios available in Blast2GO. Many different combinations are possible to adjust best to your needs. Hover over the different elements to find out more about each step.

Gene Ontology Annotation

OmicsBox Latest Updates

Version 1.0

- Blast2GO is now part of OmicsBox
- Functional Annotation
 - CloudBlast can now filter the database by tax
 - Improved CloudBlast speed and resource use
 - New EC class (F)
- Genome Analysis
 - DNA-Seq De Novo Assembly (ARVSS 2.0)



Blast2GO 6.0.3

File Analysis View Help

Table: ref_fal_shtkrt.NR.outfmt5

Table entries: 68 927

Pathways #Seqs #Enzs

Glycerolipid metabolism 44 19

Purine metabolism 35 37

Cysteine and methionine metabolism 33 33

Amino sugar and nucleotide sugar metabolism 31 36

Color Enzyme #Seqs Sequences

ec:2.3.1.20 O-acyltransferase 26 FalsK.94924.3.T.85679

FalsK.14511.0.T.72506

FalsK.95586.72995.T.8

FalsK.95586.62366.T.8

FalsK.95586.72987.C11

Progress: File Manager Application Messages

100% Make Combined Graph (ref_fal_shtkrt.NR.outfmt5): done [6s]

100% Make GO Graph: done

100% KEGG Pathway-Maps (ref_fal_shtkrt.NR.outfmt5): done [33m7s]

100% GO-EnzymeCode Mapping (ref_fal_shtkrt.NR.outfmt5): done [3s]

100% GO-Slim (ref_fal_shtkrt.NR.outfmt5): done [11s]

100% Annotation (ref_fal_shtkrt.NR.outfmt5): done [20s]

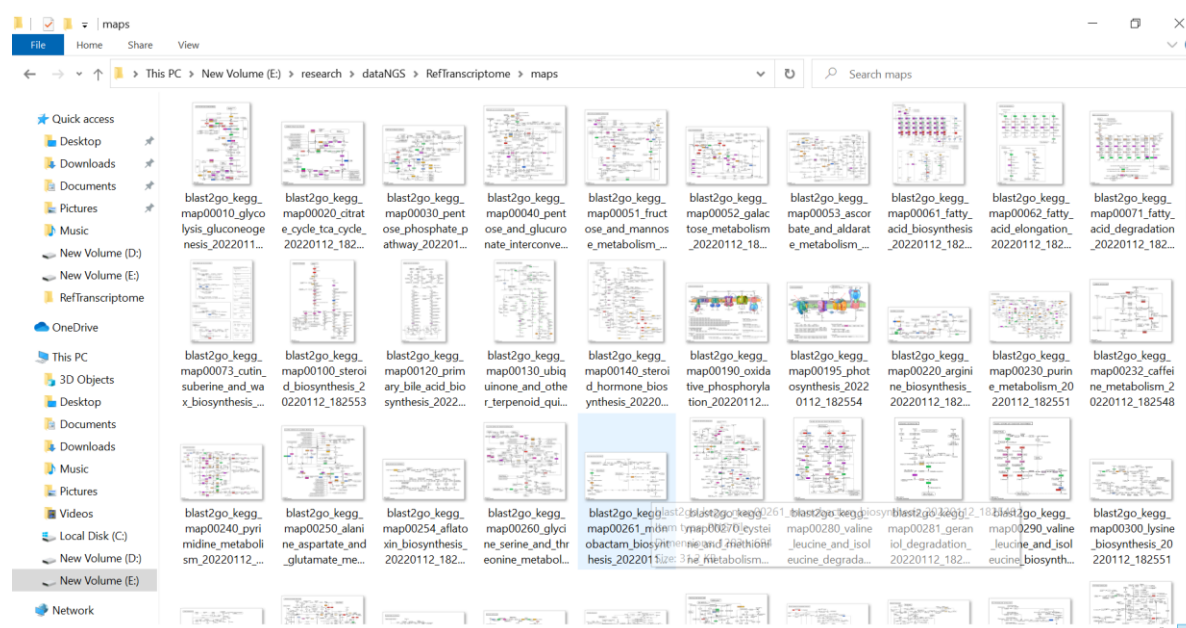
100% Mapping (ref_fal_shtkrt.NR.outfmt5): done [1m41s]

Welcome Message: *KEGG Viewer: KEGG Results of ref_fal_shtkrt.NR.outfmt5

Save Map Save all Maps Export data

Jangan lupa di akhir nama file ketika akan *export* ditambahkan extension: .txt

Contoh hasil *pathway* yang dihasilkan:



6. Ekstraksi jumlah ekspresi gen

Untuk menganalisis perbedaan ekspresi gen antar sampel, perlu dilakukan kuantifikasi sekuens-sekuens yang dihasilkan. Salah satunya menggunakan *software* eXpress (Roberts & Pachter 2012). *Software* ini dapat dijalankan pada Linux, Mac, OS X, maupun Windows.

Command untuk melakukan kuantifikasi pada program eXpress yang sudah berjalan:

```
Express -no-bias-correct -o namasampel.eXpress namafileoutput.fasta
namaalignedreads.(sam/bam)
```

7. Melakukan analisis *Differentially Expressed Gene* (DEG)

Analisis *differential expression gene* (DEG) yaitu analisis untuk mengidentifikasi kandidat gen-gen yang diekspresikan secara signifikan pada jaringan sengon yang terserang dan tidak terserang karat tumor. Nilai ekspresi gen dinyatakan dalam *Counts per million reads* (CPM). Semakin tinggi nilai CPM, maka semakin tinggi nilai ekspresi gen tersebut. Kriteria penetapan DEG adalah dengan nilai log *Fold Change* (FC) sebesar ≥ 2 dan nilai peluang relatif (p) < 0.05 . Nilai $\log FC \geq 2$ artinya adalah nilai $\log CPM$ gen yang akan dibandingkan adalah ≥ 2 x lebih tinggi atau lebih rendah dibandingkan dengan gen pembanding. Analisis DEG dilakukan dengan *package* edgeR pada R *statistical software* (Robinson *et al.* 2010).

Hal pertama yang perlu dilakukan adalah menggabungkan data eXpress.

```
read.table("file1.eXpress.txt",sep="\t",header=FALSE,
stringsAsFactors=FALSE)->namafile1
read.table("file2.eXpress.txt",sep="\t",header=FALSE,
stringsAsFactors=FALSE)->namafile2
```

```
merge(namafile1,namafile2, by="V1")->m
write.table(m,"namafilegabungan.txt",sep="\t",row.names=FALSE, col.names=FALSE,quote =
FALSE)
```

Selanjutnya dilakukan analisis *Differentially Expressed Gene* (DEG) menggunakan *package edgeR*.

```
library(edgeR)
read.table("namafilegabungan.txt", header=F, row.names=1)->namasingkatfile
c("sampel1", "sampel2")->names(namasingkatfile)
factor(c("sampel1", "sampel2"))->group
DGEList(counts=namasingkatfile, genes=row.names(namasingkatfile), group=group)->y
calcNormFactors(y)->y1
y1$samples
0.2->bcv
exactTest(y1, dispersion=bcv^2)->exacttest
exacttest$table->namafileoutput
summary(namasingkatfile<-decideTestsDGE(exacttest))
topTags(exacttest, 94255)$table->namaoutput2
write.table(namaoutput2, file="namafileoutput2.txt", sep="\t", na="NA", quote=FALSE)
```

keterangan:

bcv= koefisien variasi biologis, umumnya berkisar dari 0.1-0.4.

Kemudian untuk mengetahui gen-gen apa saja yang memiliki perbedaan ekspresi, satukan data hasil analisis *Differentially Expressed Gene* (DEG) dengan hasil annotation (.outfmt6).

```
read.delim("namafileoutput2.txt", header=F)->namasingkatdeg
names(namasingkatdeg)<-c("seq", "logFC", "logCPM", "pValue", "significant")
read.delim("namafileannotation.outfmt6", header=F)->namasingkatannotation
names(namasingkatannotation)<-c("seq", "ann", "a", "b", "c", "d", "e", "f", "g", "swissprot")
merge(namasingkatdeg, namasingkatannotation, by="seq")->namafilegabungan
head(namafilegabungan)
write.table(namafilegabungan, file="namafilegabungan.txt", sep="\t", na="NA",
quote=FALSE)
```


Contoh hasil data *Differentially Expressed Gene* (DEG) yang digabungkan dengan hasil *annotation*

ID sekuen	logFC	FDR	Nama Gen
<i>Down-regulated</i>			
			<i>Mitogen-activated protein</i>
Fm170150.37989.T.115515.c0gli6	-14.08	4.2E-40	<i>kinase kinase kinase 5</i>
			<i>Anthocyanin regulatory C1</i>
Fm170150.62813.T.115909.c1g2i2	-13.37	7.4E-35	<i>protein</i>
			<i>Anthocyanin regulatory C1</i>
Fm170150.88424.T.115909.c1g2i1	-13.36	7.9E-35	<i>protein</i>
<i>Up-regulated</i>			
Fm170150.71692.Fm195586.92343.			
T.58420.c0g2i1	12.29	9.8E-26	<i>Transcription factor UPBEAT1</i>
Fm44288.2.Fm1113171.0.T.80273.c2			
g2i3	12.13	1.4E-24	<i>Ribonuclease 3</i>
			<i>Retrovirus-related Pol</i>
			<i>polyprotein from transposon</i>
Fm170150.74932.T.128668.c0g3i1	11.97	1.6E-23	<i>RE2</i>

Untuk memvisualisasikan jumlah gen yang terekspresi, baik secara *up-regulated* ($\log FC \geq 2$) maupun *down-regulated* ($\log FC \leq -2$) dapat menggunakan *volcano plot*.

```

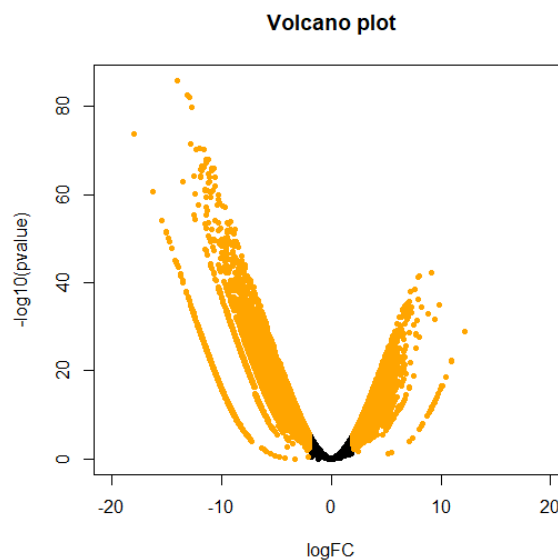
namasingkatfile <- read.table("namafiledeg.txt", header=TRUE)
with(namasingkatfile, plot(logFC, -log10(pvalue), pch=20, main="Volcano plot", xlim=c(-20,20)))
with(subset(namasingkatfile, abs(logFC)>2), points(logFC, -log10(pvalue), pch=20, col="orange"))
with(subset(namasingkatfile, padj<.05 & abs(logFC)>2), points(logFC, -log10(pvalue), pch=20, col="green"))
with(subset(namasingkatfile, padj<.05 & abs(logFC)>10), textxy(logFC, -log10(pvalue), labs=Gene, cex=.3))

```

Adjusted p-value atau FDR adalah rasio jumlah hasil yang *false positive* dibandingkan dengan total hasil yang positif. Beberapa metode yang dapat digunakan untuk mencari nilai *adjusted p-value* yaitu Hochberg, Holm, Hommel, Bonferroni, Benjamini-Holm (BH), Benjamini-Yekutieli (BY). Kalkulasi nilai *adjusted p-value* dilakukan dengan memilih salah satu metode tersebut. Jika nilai *adjusted p-value* atau *false discovery rate* (FDR) belum diketahui, nilai tersebut dapat dicari terlebih dahulu sebelumnya dari file *deg* dengan kode. Nilai *p* tersebut kemudian ditambahkan pada tabel hasil gabungan, dapat dilakukan secara manual (*copy paste*) maupun menggunakan fungsi *merge*.

```
p.adjust(pValue, method = c("holm", "hochberg", "hommel", "bonferroni", "BH",
"BY", "fdr", "none"), n = length(pValue))
```

Contoh hasil *volcano plot*:

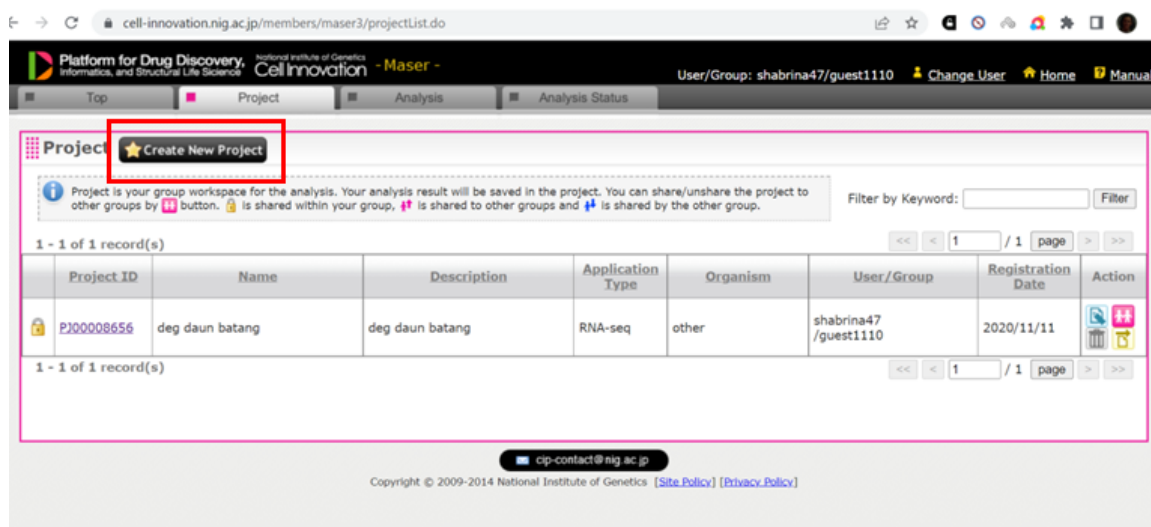


Volcano plot hasil analisis *Differentially Expressed Gene* (DEG).

Keterangan : Bagian berwarna kuning dengan nilai *logFC* negatif merupakan gen terekspresi secara *down-regulated*, sedangkan yang bernilai positif terekspresi secara *up regulated*.

6.2 Pengolahan data RNA-Seq menggunakan *Cloud Computing* pada MASER (Kinjo *et al.* 2017)

1. Buat akun pada laman MASER (https://cell-innovation.nig.ac.jp/ums_wus/UserRegistrationInputForm) email. Akun hanya bisa diaktifkan satu kali dan tidak ada opsi *reset password*.
2. Jika sudah berhasil login, akan muncul laman *project*. Pada laman tersebut tambahkan *project* baru dengan menekan *create new project* (proyek baru).



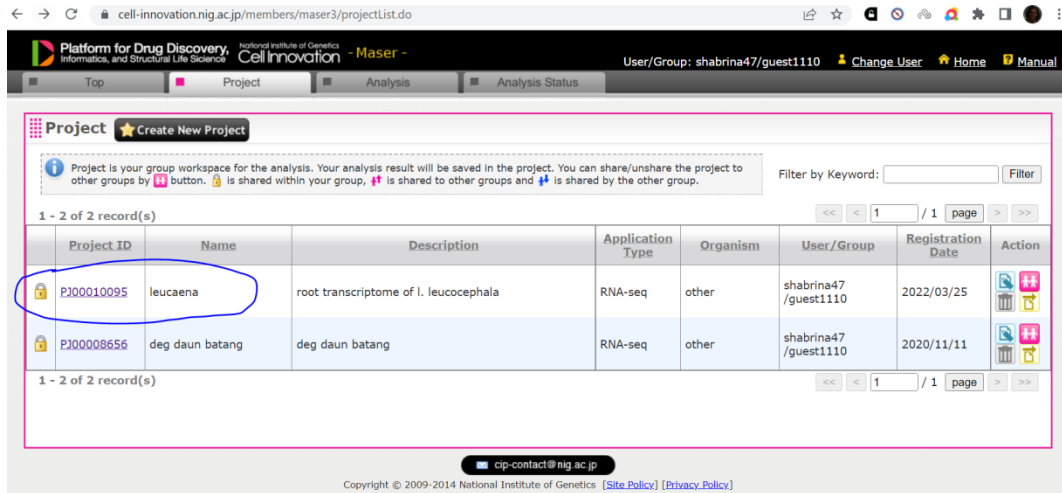
Contoh *new project*:

The screenshot shows the MASER Project Create form. It includes a 'Project Information' section with the following fields:

- Name:** Describe freely the words meaning of this project. (Input: leucaena)
- Description:** Describe freely the words detail of this project. (Input: root transcriptome of 1. leucocephala)
- Application Type:** Select the classification of sequence application. (Dropdown: RNA-seq)
- Organism:** Select the organism species. (Dropdown: Other (other))
- User/Group:** Owner of this project. (Input: shabrina47/guest1110)

At the bottom, there are 'OK' and 'Cancel' buttons.

3. *New Project* muncul pada laman. Klik *project ID* untuk mulai menambahkan data.



Platform for Drug Discovery, Informatics, and Structural Life Sciences
National Institute of Genetics
Cell Innovation - Maser

User/Group: shabrina47/guest1110 Change User Home Manual

Project Create New Project

Project is your group workspace for the analysis. Your analysis result will be saved in the project. You can share/unshare the project to other groups by button. is shared within your group, is shared to other groups and is shared by the other group.

Filter by Keyword: Filter

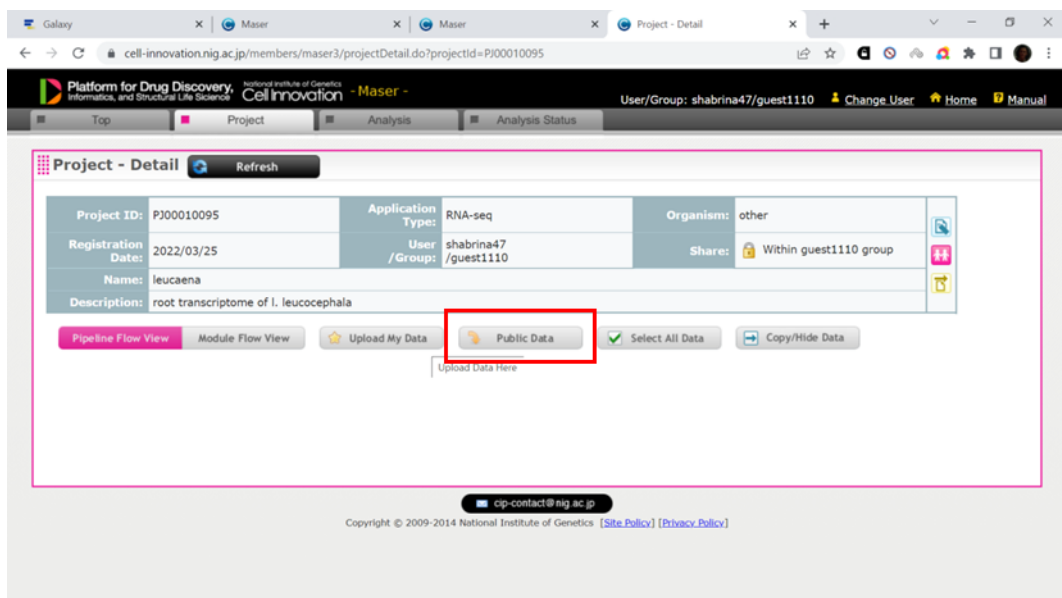
1 - 2 of 2 record(s)

Project ID	Name	Description	Application Type	Organism	User/Group	Registration Date	Action
P300010095	leucaena	root transcriptome of l. leucocephala	RNA-seq	other	shabrina47/guest1110	2022/03/25	
P300008656	deg daun batang	deg daun batang	RNA-seq	other	shabrina47/guest1110	2020/11/11	

1 - 2 of 2 record(s)

Copyright © 2009-2014 National Institute of Genetics [Site Policy] [Privacy Policy]

4. Unggah data dengan memilih *upload* data. Jika ingin menggunakan data yang tersedia di *domnain public* seperti *database*, klik *public data*.



Platform for Drug Discovery, Informatics, and Structural Life Sciences
National Institute of Genetics
Cell Innovation - Maser

User/Group: shabrina47/guest1110 Change User Home Manual

Project - Detail Refresh

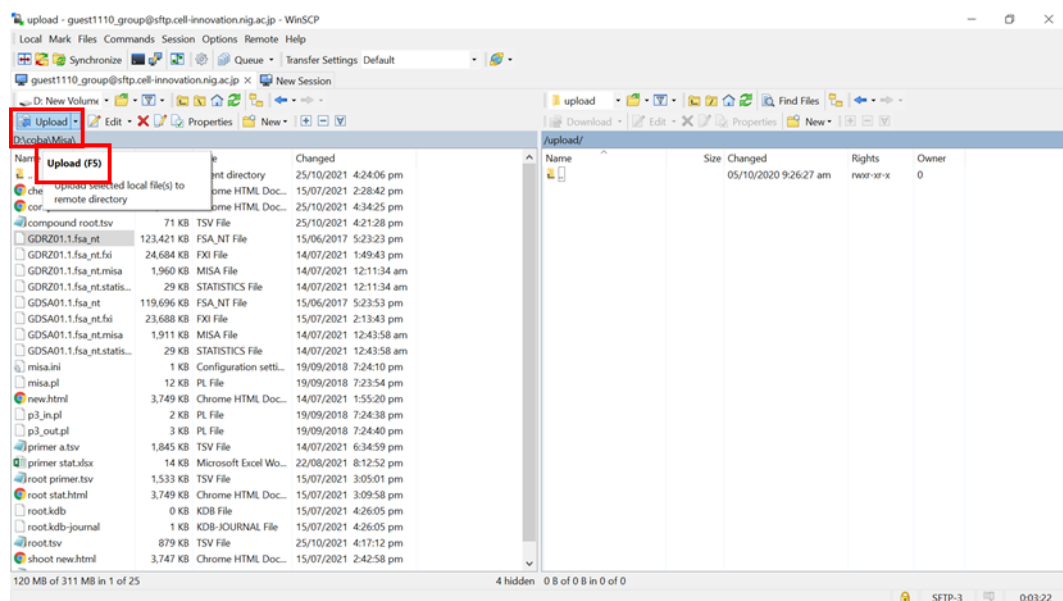
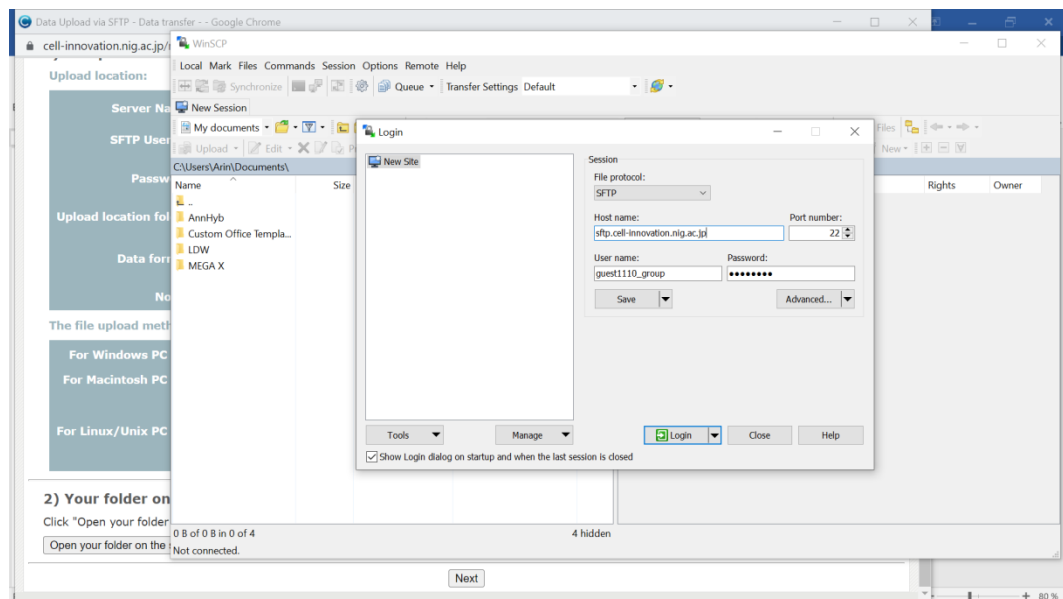
Project ID:	P300010095	Application Type:	RNA-seq	Organism:	other
Registration Date:	2022/03/25	User /Group:	shabrina47/guest1110	Share:	Within guest1110 group
Name:	leucaena				
Description:	root transcriptome of l. leucocephala				

Pipeline Flow View Module Flow View Upload My Data **Public Data** Select All Data Copy/Hide Data

Upload Data Here

Copyright © 2009-2014 National Institute of Genetics [Site Policy] [Privacy Policy]

5. Jika ukuran file besar (>100 MB) unggah *file* menggunakan WinSCP (*Windows*) atau *Cyberduck* pada Mac.



Untuk melihat apakah data sudah terunggah, maka masuk ke dalam *folder* di dalam *server*.

Upload location:

Server Name: **sftp.cell-innovation.nig.ac.jp**

SFTP User ID: **guest1110_group**

Password: **acae28a6**
Caution) The password is appear only on first time. Please keep it at hand.

Upload location folder: **/upload**

Data format: The supported data formats are listed [here](#).
 If you transfer compressed files with extension of "gz" or "bz2", your files will be decompressed automatically in the next step.

Notes: This folder is shared within the guest1110_group user, but it isn't referred to from other users.

The file upload method via SFTP:

For Windows PC user: Please use [WinSCP](#) application.

For Macintosh PC user: Please use [Cyberduck](#) application.

For Linux/Unix PC user: Please use sftp command.
 ex.)
`$ sftp guest1110_group@sftp.cell-innovation.nig.ac.jp:/upload`
`sftp> put mydata.fastq`
`sftp> exit`

2) Your folder on the server

Click "[Open your folder on the server](#)" to confirm the files has been transferred properly.

[Open your folder on the server](#)

Next

6. Contoh jika menggunakan *public data*, pilih data yang diinginkan.

Project ID	Description	Type	Date
SRX6088942	s; RNA-Seq	SRA	2019/10/09
SRX6088944	GSM3897044: ADSCD21P3; Homo sapiens; RNA-Seq	SRA	2019/10/09
SRX6088943	GSM3897043: ADSCD14P3L; Homo sapiens; RNA-Seq	SRA	2019/10/09
SRX6088940	GSM3897040: ADSCD10P3; Homo sapiens; RNA-Seq	SRA	2019/10/07
SRX3630292	GSM2971255: Bacillus subtilis_pnpA_Rend_seq_LB_25s_frag; Bacillus subtilis subs p. subtilis str. 168; RNA-Seq	SRA	2019/06/29
srauser01-0004	Project PJ00007015 related data	User Data	2019/05/24
SRX4505322	GSM3318398: Mouse-stomach-development-W8; Mus musculus; RNA-Seq	SRA	2019/05/06
SRX997767	GSM1659693: RORgamma_liver_ZT10; Mus musculus; ChIP-Seq	SRA	2019/04/05
SRX997766	GSM1659692: RORalpha_liver_ZT10; Mus musculus; ChIP-Seq	SRA	2019/01/09
☒ SRX203189	High-throughput RNA sequencing of a FFP E autopsy lung tissue samples from the 2009 influenza pandemic	SRA	2018/07/09
SRX203200	High-throughput RNA sequencing of a FFP E autopsy lung tissue sample from the 1918 influenza pandemic	SRA	2018/07/09
ERX2160003	Illumina HiSeq 2500 paired end sequencing	SRA	2018/06/17
ERX2157054	Illumina HiSeq 2500 paired end sequencing; RNA-Seq analysis of mouse dopamine neurons identifies transcriptional repression of PlexinC1 by Lmx1a and Lmx1b	SRA	2018/06/17

Tampilan jika data sudah terunggah adalah sebagai berikut:

Project - Detail

Project ID: PJ00010095 Application Type: RNA-seq Organism: other

Registration Date: 2022/03/25 User / Group: shabrina47 / guest1110 Share: Within guest1110 group

Name: leucaena

Description: root transcriptome of l. leucocephala

Pipeline Flow View Module Flow View Upload My Data Public Data Select All Data Copy/Hide Data

SRR611716 (DS00734567) Select

Copyright © 2009-2014 National Institute of Genetics

7. Klik *select*, kemudian klik *analysis*.

Project - Detail

Project ID: PJ00010095 Application Type: RNA-seq Organism: other

Registration Date: 2022/03/25 User / Group: shabrina47 / guest1110 Share: Within guest1110 group

Name: leucaena

Description: root transcriptome of l. leucocephala

Pipeline Flow View Module Flow View Upload My Data Public Data Select All Data Copy/Hide Data

SRR611716 (DS00734567) Select

Clipboard

Please select a bulk operation for these data:

Genome Viewer Download Analysis

Filter by Data Type: Analysis

Dataset ID	Data Label	Data Type	Action
DS00734567	SRR611716	fastq (single-end)	Unselect

Copyright © 2009-2014 National Institute of Genetics

8. Pada laman *analysis*, pilih RNA-Seq

Project selected: leucaena (PJ00010095)

Dataset selected: SRR611716 (DS00734567)

The number of running processes: 9 total, 0 yours

Select a category on the left and click **Analysis** button of the pipeline you want to run.

Filter by Keyword: Filter

Category

- NGS applications (40)
 - Data processing (3)
 - RNA-seq (10)**
 - Standard (2)
 - Other (8)
 - SAGE-seq (4)
 - CAGE-seq (5)
 - ChIP-seq (5)
 - Bisulphate-seq (3)
 - Resequencing (Exome-seq) (2)
 - De novo Genome sequencing (4)
 - Metagenome (4)
 - Visualization (0)
 - Genetic statistics (0)
 - Tools (43)
 - Misc. (29)

Pipeline

Action	Name	Summary
Analysis	TopHat2, CuffLinks2 and CummeRbund + GE (SE)	TopHat2, CuffLinks2, CummeRbund, and the mapping result is visualized by Genome Explorer (GE) on the web. [IN: up to 10 fastq (single-end) files]
Analysis	Trinity, Bowtie, eXpress and DEGseq (SE)	Trinity, Bowtie, eXpress and DEGseq [IN: fastq (single-end)]
Analysis	Assemble: Trinity (SE)	Assemble: Trinity (SE)
Analysis	mapped read count with BWA-mem (SE)	mapped read count with BWA-mem (SE)
Analysis	rSeq calcGeneExp from fastq	Calculate gene expression with rSeq from fastq file.
Analysis	TMAP-eXpress (SE)	TMAP-eXpress (SE)
Analysis	TopHat2, CuffLinks2 and CummeRbund + GE (SE) v1.1	TopHat2, CuffLinks2, CummeRbund, and the mapping result is visualized by Genome Explorer (GE) on the web. [IN: up to 10 fastq (single-end) files]
Analysis	TopHat2, CuffLinks2 and CummeRbund + GE (SE) v1.2	TopHat2, CuffLinks2, CummeRbund, and the mapping result is visualized by Genome Explorer (GE) on the web. [IN: up to 10 fastq (single-end) files]
Analysis	Trinity, Bowtie2, clustering and blast annotation (SE)	Trinity, Bowtie2, clustering and blast annotation (SE)
Analysis	Trinity, TMAP, eXpress and DEGseq (SE)	Trinity, TMAP, eXpress and DEGseq [IN: fastq (single-end)]

9. Pilih *pipeline analysis* yang diinginkan. Sebagai contoh kali ini adalah analisis yang dilingkari biru.

The screenshot shows the Galaxy web interface. On the left, under 'Category', 'RNA-seq (10)' is selected. On the right, under 'Pipeline', a table lists various analysis actions. The pipeline 'Trinity, Bowtie, eXpress and DEGseq (SE)' is highlighted with a blue circle.

Action	Name	Summary
Analysis	TopHat2, CuffLinks2 and CummeRbund + GE (SE) ?	TopHat2, CuffLinks2, CummeRbund, and the mapping result is visualized by Genome Explorer(GE) on the web. [IN: up to 10 fastq (single-end) files]
Analysis	Trinity, Bowtie, eXpress and DEGseq (SE) ?	Trinity, Bowtie, eXpress and DEGseq [IN: fastq (single-end)]
Analysis	Assemble: Trinity (SE) ?	Assemble: Trinity (SE)
Analysis	mapped read count with BWA-mem (SE) ?	mapped read count with BWA-mem (SE)
Analysis	rSeq calcGeneExp from fastq ?	Calculate gene expression with rSeq from fastq file.
Analysis	TMAP-eXpress (SE) ?	TMAP-eXpress (SE)
Analysis	TopHat2, CuffLinks2 and CummeRbund + GE (SE) v1.1 ?	TopHat2, CuffLinks2, CummeRbund, and the mapping result is visualized by Genome Explorer(GE) on the web. [IN: up to 10 fastq (single-end) files]
Analysis	TopHat2, CuffLinks2 and CummeRbund + GE (SE) v1.2 ?	TopHat2, CuffLinks2, CummeRbund, and the mapping result is visualized by Genome Explorer(GE) on the web. [IN: up to 10 fastq (single-end) files]
Analysis	Trinity, Bowtie2, clustering and blast annotation (SE) ?	Trinity, Bowtie2, clustering and blast annotation (SE)
Analysis	Trinity, TMAP, eXpress and DEGseq (SE) ?	Trinity, TMAP, eXpress and DEGseq [IN: fastq (single-end)]

10. Pada laman yang muncul setelah di-klik *analysis*, maka akan muncul data dan *project* yang akan dianalisis untuk konfirmasi. Kemudian pilih *set option and run*.

The screenshot shows the Galaxy web interface for the 'Analysis' page. It displays input datasets and a project selection dropdown. The 'Set option and run' button is highlighted with a red box.

4. add BLASTN (NR) annotation to expression table
add BLASTN (NR) annotation to expression table (require sequence column)
1. add BLASTN (NT) annotation to expression table

5. GO analysis for Trinity-eXpress
GO analysis for Trinity-eXpress (require GO id and cluster number)
1. GO analysis for Trinity-eXpress

Please select input datasets of analysis pipeline.

* Required

1. fastq (single-end) * SRR611716(DS00734567) Select

2. fastq (single-end) * Select

3. fastq (single-end) Select

4. fastq (single-end) Select

5. fastq (single-end) Select

6. fastq (single-end) Select

7. fastq (single-end) Select

8. fastq (single-end) Select

9. fastq (single-end) Select

10. fastq (single-end) Select

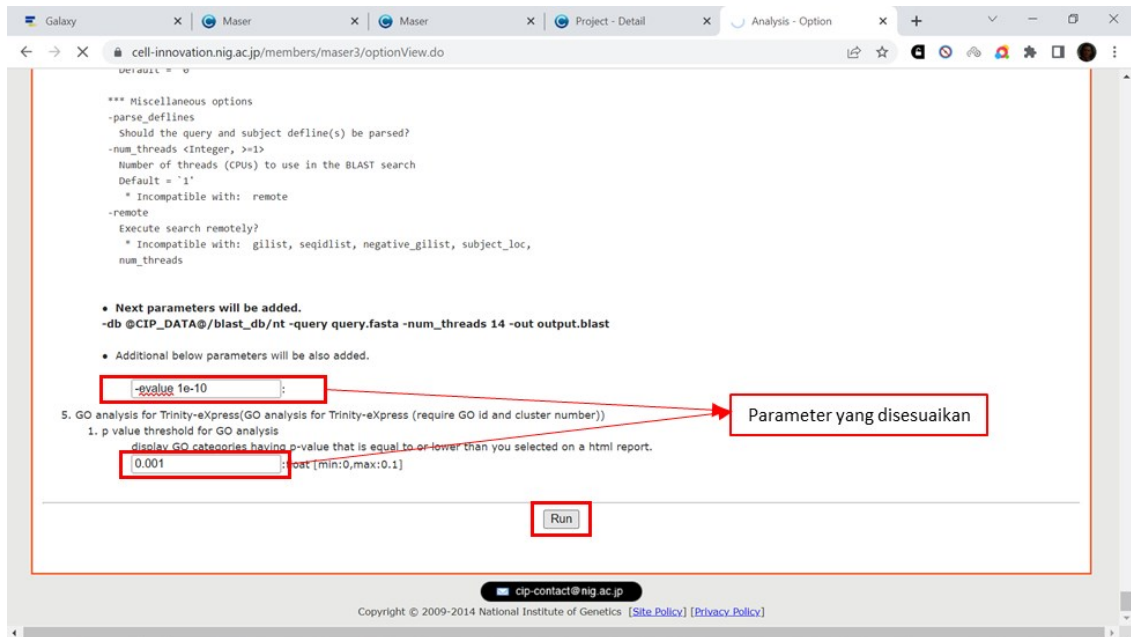
Please select a destination project for the analysis result.

* Required

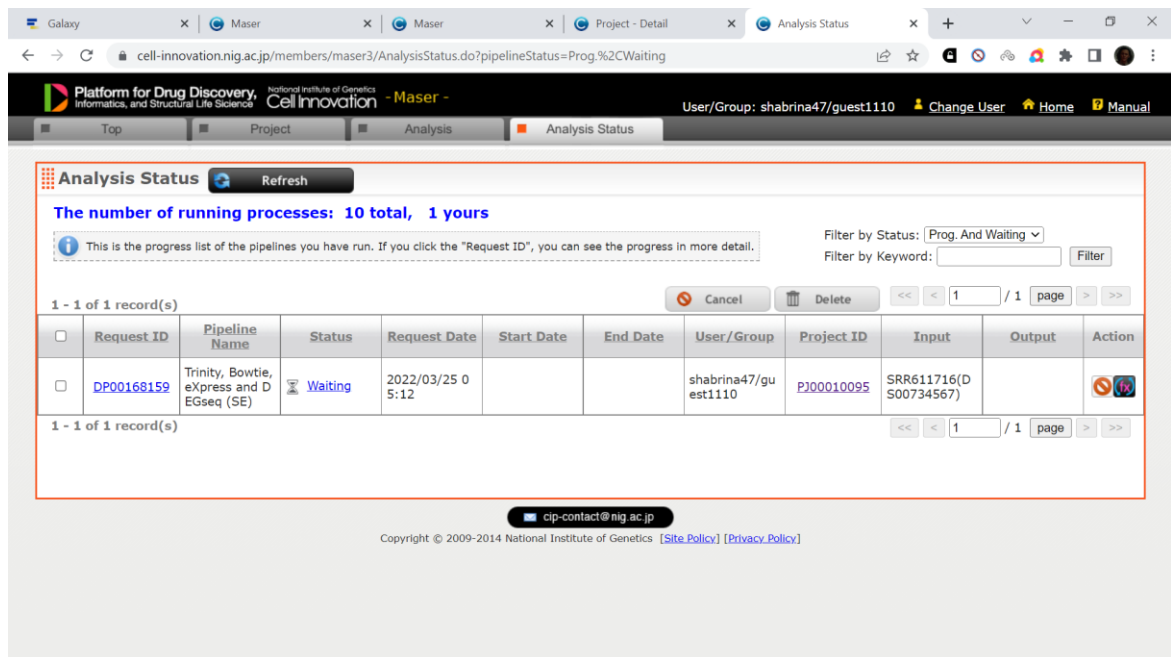
1. Project * PJ00010095 Select

Set option and run Run

11. Selanjutnya *set parameter* untuk *gen ontology*, dan lainnya. Kemudian klik *run*.

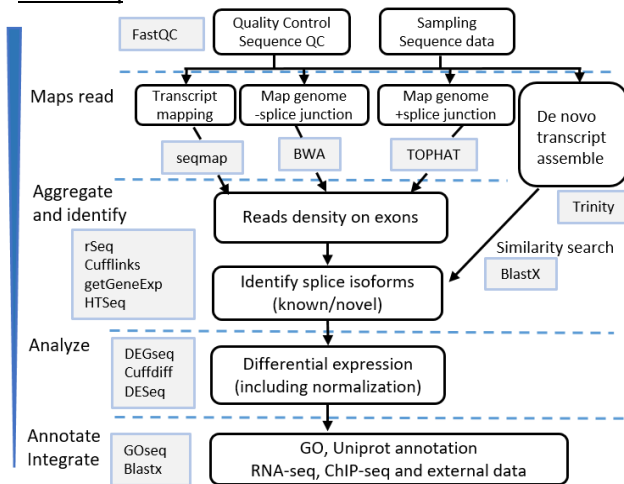


12. Analisis sedang berlangsung.

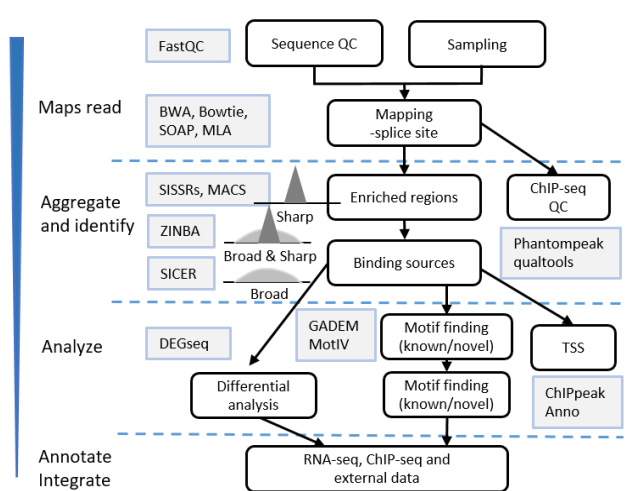


Keunggulan MASER adalah kita tidak perlu memikirkan program apa saja yang akan kita gunakan karena program-program tersebut sudah tersusun di dalam *pipeline-pipeline* yang dapat kita pilih. Selain untuk RNA-Seq, MASER juga menyediakan layanan untuk ChIP-Seq (*Chromatin Immunoprecipitation sequencing*), BS-Seq (*Bisulfate sequencing* untuk mendeteksi sitosin termetilasi pada DNA genom), CAGE-Seq (*Cap Analysis Gene Expression sequencing* untuk menganalisis ujung 5' pada RNA), *De Novo Genome Sequencing*, Metagenome analisis, dan *genom resequencing*. Pipeline atau alur kerja pengerjaan beberapa pengolahan data NGS dapat dilihat pada gambar di bawah ini:

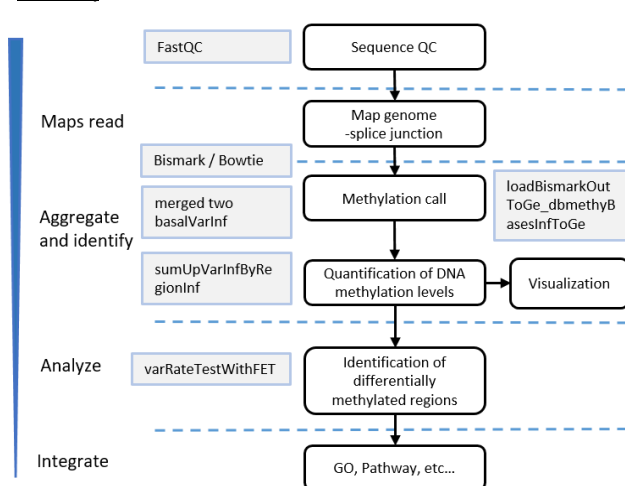
RNA-Seq



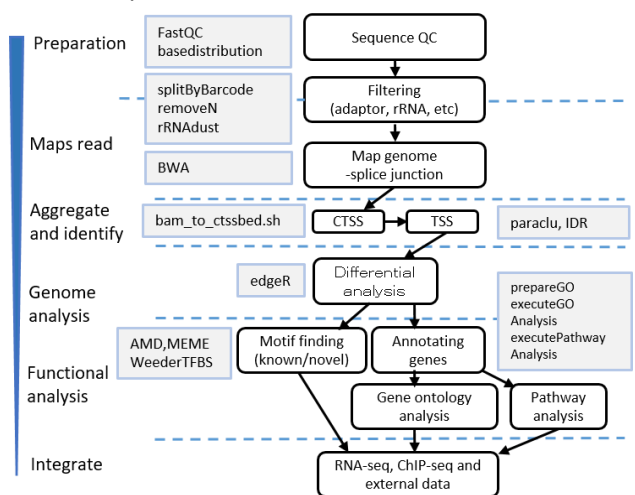
ChIP-Seq



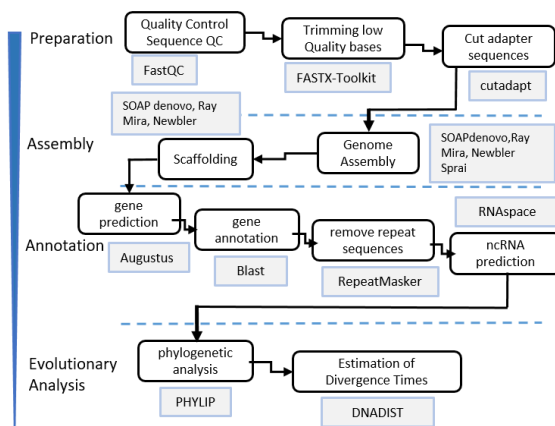
BS-Seq



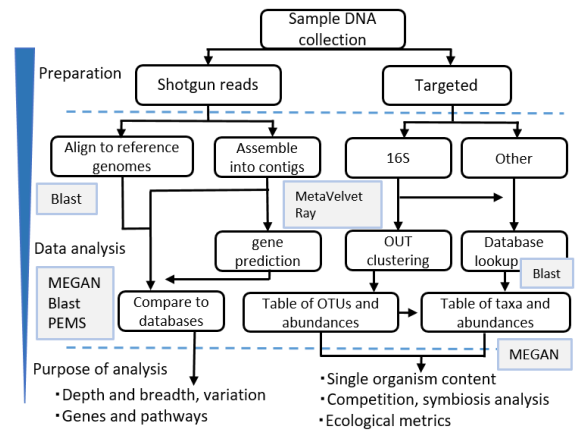
CAGE-Seq



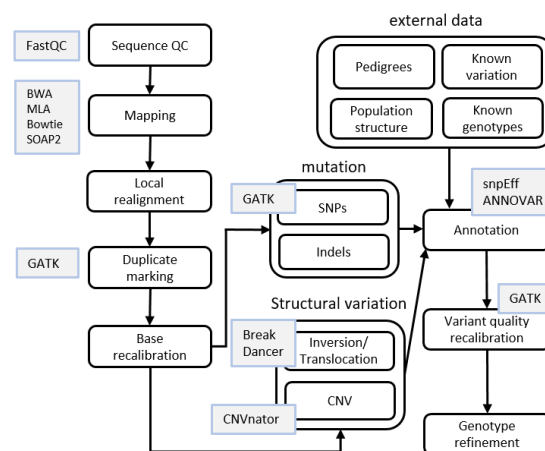
De Novo Genome sequencing



Metagenome



Genome resequencing



Sumber: https://cell-innovation.nig.ac.jp/maser/index_en.html

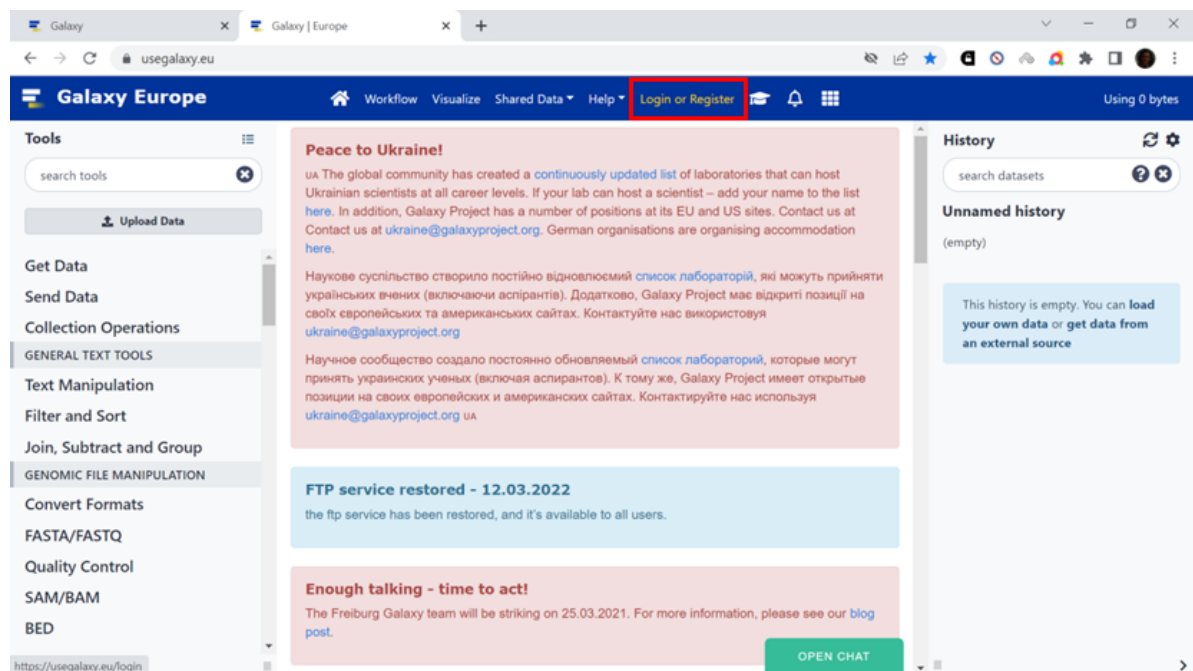
6.3 Cloud Computing pada Platform Galaxy

Galaxy adalah *platform* berbasis *web* yang dikembangkan sejak tahun 2005 dengan tujuan memungkinkan peneliti yang tidak memiliki akses sistem untuk melakukan berbagai analisis komputasi pada data biologis (Afgan *et al.* 2016). Terdapat tiga server Galaxy yang dapat digunakan yaitu:

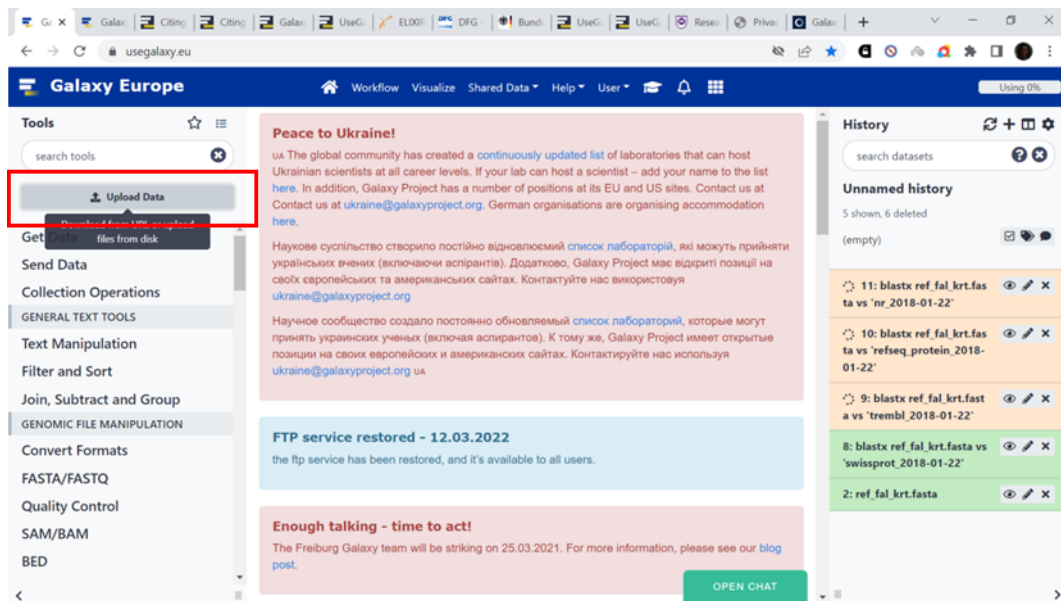
1. Usegalaxy.org yang merupakan server utama yang disponsori oleh Galaxy project.
2. Usegalaxy.eu, server Galaxy yang bekerja sama dengan ELIXIR (<https://elixir-europe.org/>), *Deutsche Forschungsgemeinschaft* (DFG, <https://www.dfg.de/>), serta Kementerian Pendidikan dan Riset Jerman.
3. Usegalaxy.org.au, server Galaxy yang bekerja sama dengan *Research Computing Center*, *Bioplatforms Australia*, dan *Australian Research Data Commons*.

Berikut adalah langkah untuk melakukan analisis data pada *platform* Galaxy:

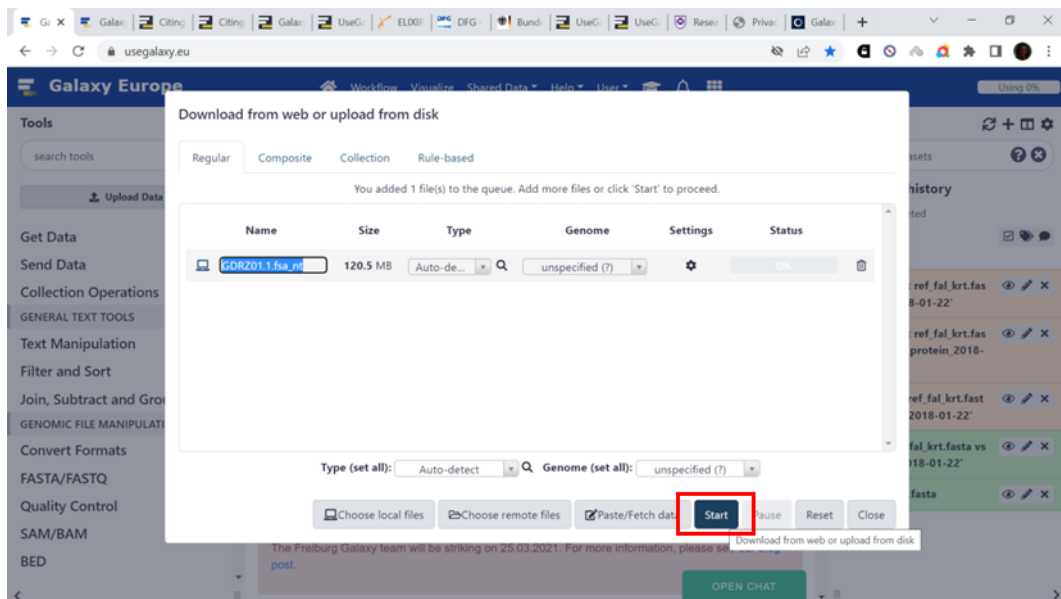
1. Buat akun pada laman Galaxy yang akan digunakan.



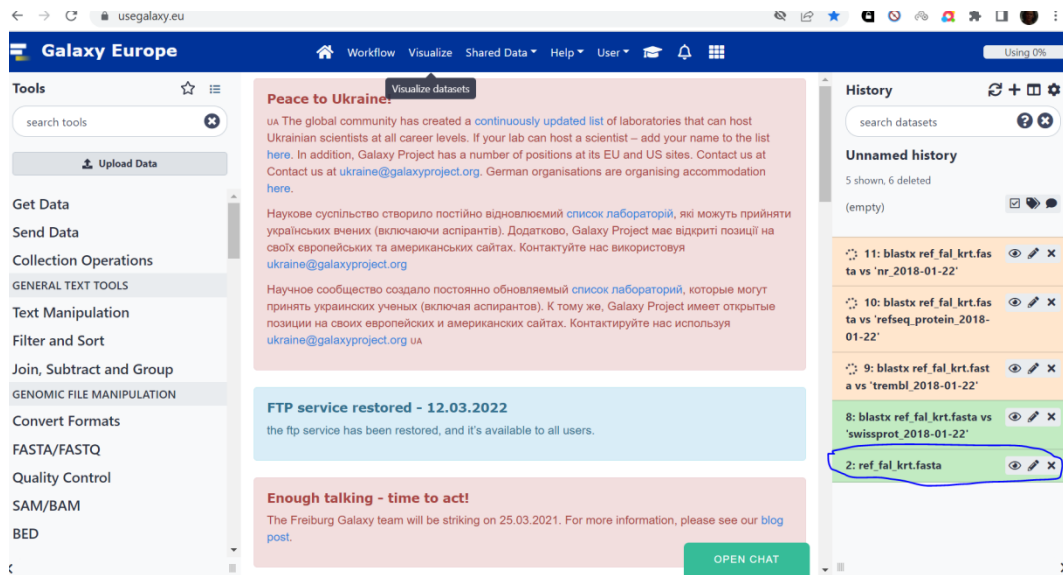
2. Setelah berhasil *login*, unggah data kita ke dalam *server* Galaxy.



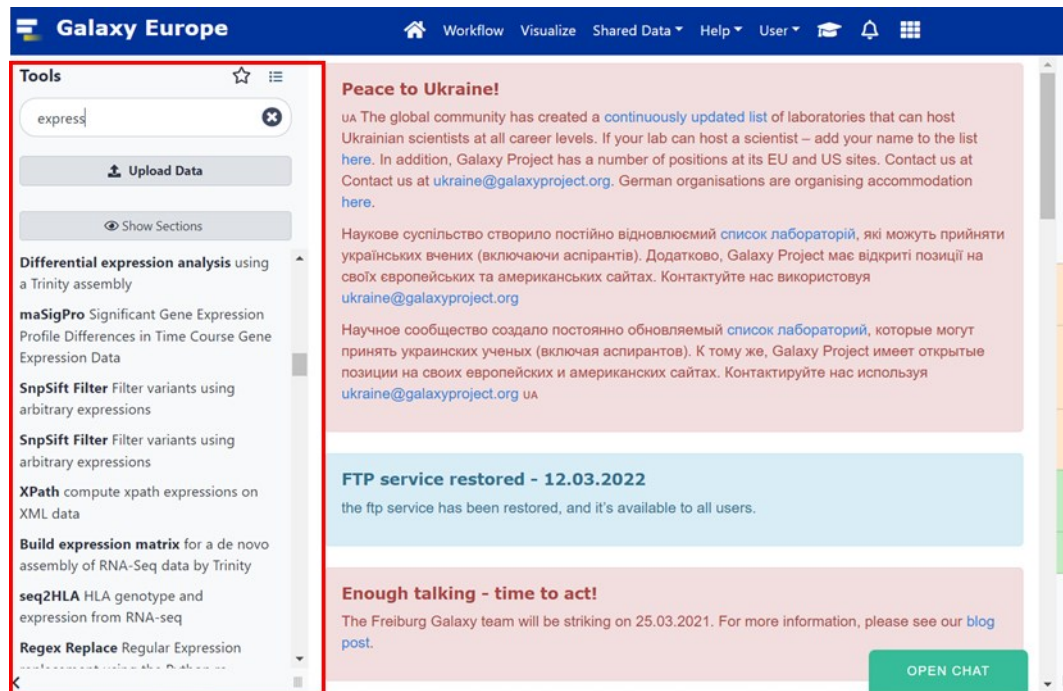
Pilih *file* yang akan diunggah ke dalam *server* Galaxy, kemudian klik *Start*



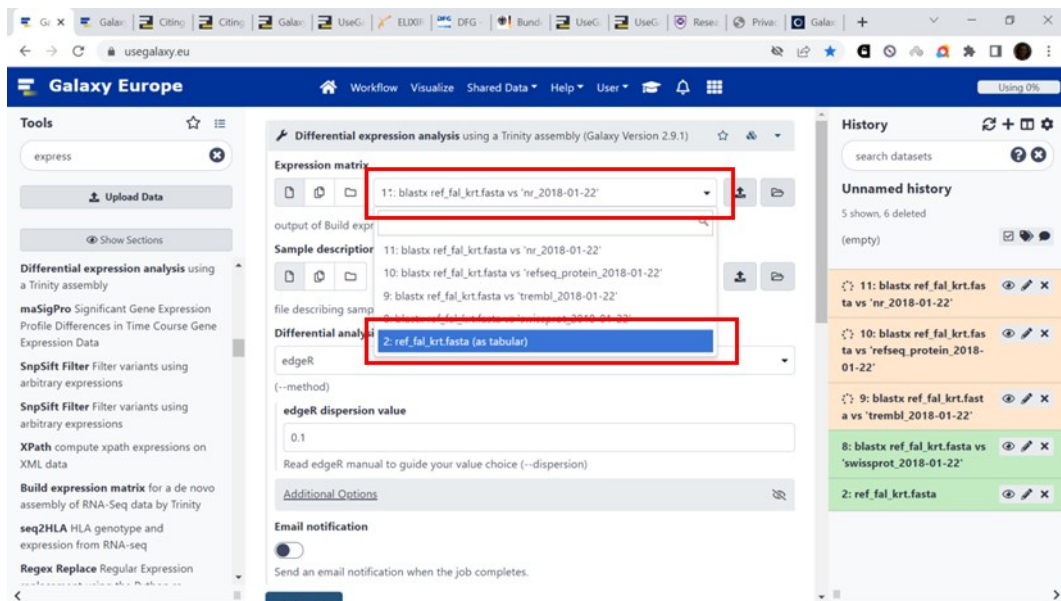
3. Data yang sudah terunggah akan muncul pada *history* Galaxy di bagian kanan.



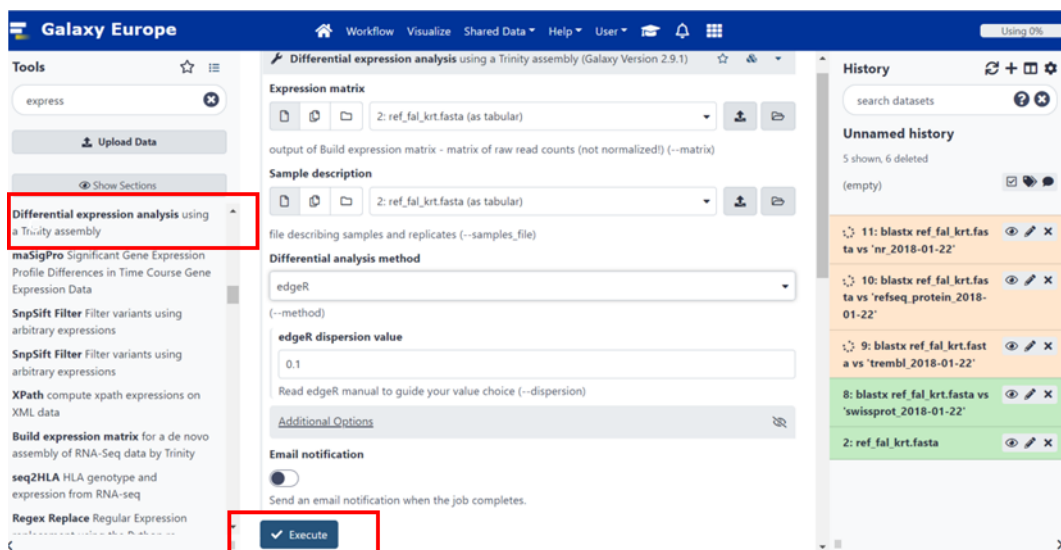
4. Selanjutnya cari dan pilih analisis yang akan dilakukan pada *search tools*. Sebagai contoh adalah analisis ekspresi gen.



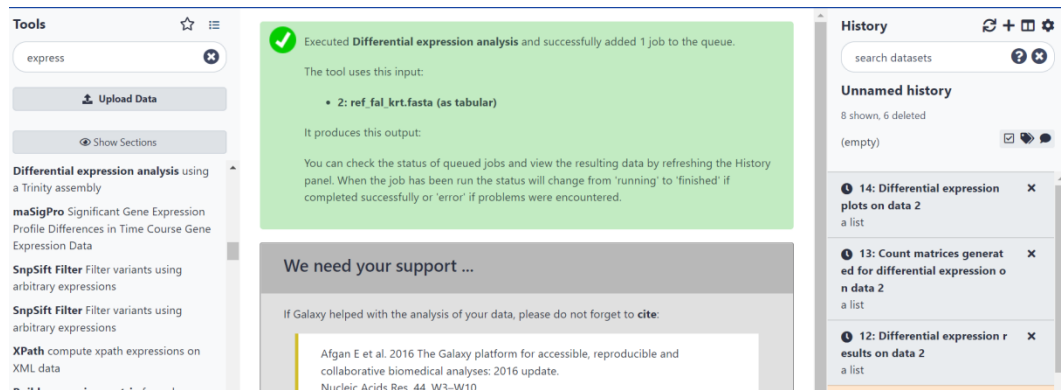
5. Pilih *file* yang akan dianalisis.



6. Pilih metode yang akan dilakukan, kemudian klik *execute*.



7. Tampilan di bawah ini menunjukkan bahwa analisis berhasil dimasukkan dan sedang menunggu antrian komputasi pada *server*. Analisis yang sudah selesai dilakukan akan muncul pada bagian *history* dengan warna hijau. Warna kuning menunjukkan analisis sedang berlangsung.



6.4 Teknologi Nanopore Sequencing

Teknologi sekuensing terkini adalah *nanopore*. Teknologi ini dikembangkan oleh *Oxford Nanopore* sejak tahun 2005 untuk mendapatkan teknologi yang disruptif, elektronik, dengan sistem deteksi *single-molecule* berbasis *nanopore*. Sekuensing yang dilakukan dengan teknologi *nanopore* adalah DNA dan RNA (Jain *et al.* 2016). *Nanopore* merupakan pori yang berukuran nano yang terbuat dari protein. Alat sekuensing *nanopore* tersusun dari alat sensor yang berupa *nanopore-nanopore* pada *chip* yang disebut *flowcell*, yang digunakan bersamaan dengan *Application-Specific Integrated Circuit* (ASIC) yang mengontrol dan mengukur proses sekuensing. Alat yang menggabungkan *sensor nanopore* pada *flowcell* dengan komputer yang digunakan untuk merekam data dinamakan MinIon. MinIon merupakan perangkat sekuensing terkecil yang tersedia saat ini. Tampilan di bawah ini adalah MinIon.



Sumber: <https://nanoporetech.com/products/minion>

Prinsip kerja *nanopore sequencer* yaitu *nanopore* yang terbuat dari protein mendeteksi perubahan arus listrik yang diakibatkan oleh material DNA atau RNA yang melewati *nanopore*. Perubahan arus listrik tersebut kemudian direkam dengan *MinKnow software* yang memproses rangkaian gelombang arus listrik dalam bentuk *FAST5 file*. Data sinyal arus listrik yang tersimpan di dalam format *FAST5* kemudian diproses dengan *basecalling algorithm* untuk mengubah sinyal tersebut menjadi sekuens dengan format *FASTQ*. Setelah itu, data tersebut dianalisis kualitasnya, kemudian di-*filter* dengan kriteria Q-score >7 dengan panjang sekuens >500 bp. Tahapan selanjutnya yaitu *trimming* dan *assembly*. Analisis data hasil sekuens dengan *nanopore* juga dapat dilakukan pada *platform* Galaxy dan MASER.

Keunggulan dari teknologi *Oxford Nanopore* yaitu:

1. *Portable* karena ukurannya yang kecil sehingga proses sekuensing dapat dilakukan di mana saja. Teknologi *portable* ini menyebabkan proses sekuensing virus zika dan ebola di Afrika dapat dilaksanakan dengan lebih cepat.
2. Harga instrumen yang jauh lebih murah jika dibandingkan dengan *platform* sekuensing NGS lainnya.
3. Sekuens yang dihasilkan berupa *long-reads* atau pembacaan sekuens berukuran panjang. Kemampuan pembacaan sekuens oleh *Oxford Nanopore* berkisar antara 500 kbp hingga 2.3 Mbp, yang jauh lebih panjang dibandingkan dengan *platform* NGS lainnya.

Keberadaan teknologi sekuensing *nanopore* menimbulkan percepatan yang amat pesat dalam berbagai hal, seperti sekuensing genom virus SARS-Cov-2 yang menyebabkan pandemi sejak 2020 hingga saat ini. Selain itu, banyak organisme yang awalnya belum tercatat susunan genomnya pada *database*, saat ini mulai tersedia datanya. Teknologi *nanopore* yang *portable* juga menyebabkan semakin banyaknya proses sekuensing *in-situ* sehingga kegiatan konservasi biodiversitas berbasis genetika bisa lebih optimal dan efisien dari segi waktu.

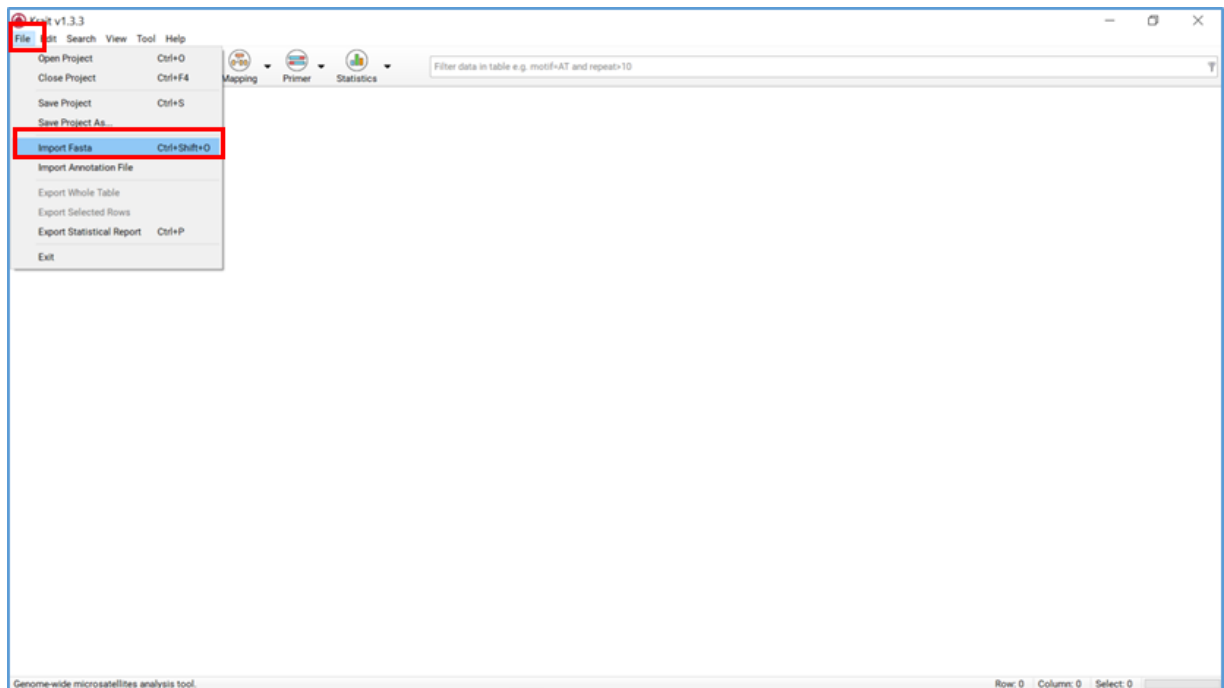
6.5 *Microsatellite finding* dan desain primer *microsatellite*

Microsatellite atau *Short Sequence Repeat (SSR)* adalah motif nukleotida berulang yang terdapat pada DNA maupun RNA. Motif SSR dapat berupa pengulangan dari di-, tri-, tetra-, penta-, atau heksanukleotida dengan panjang tertentu, biasanya minimal 5-6 pengulangan motif. *Microsatellite* terdapat pada ribuan titik pada genom. Situs *microsatellite* memiliki laju mutasi yang tinggi sehingga memiliki keragaman yang cukup tinggi. Keragaman yang tinggi menyebabkan situs *microsatellite* banyak dijadikan sebagai penanda molekuler. *Microsatellite* dapat ditemukan pada sekuens hasil sekuensing pada DNA *whole genome sequencing* maupun RNA.

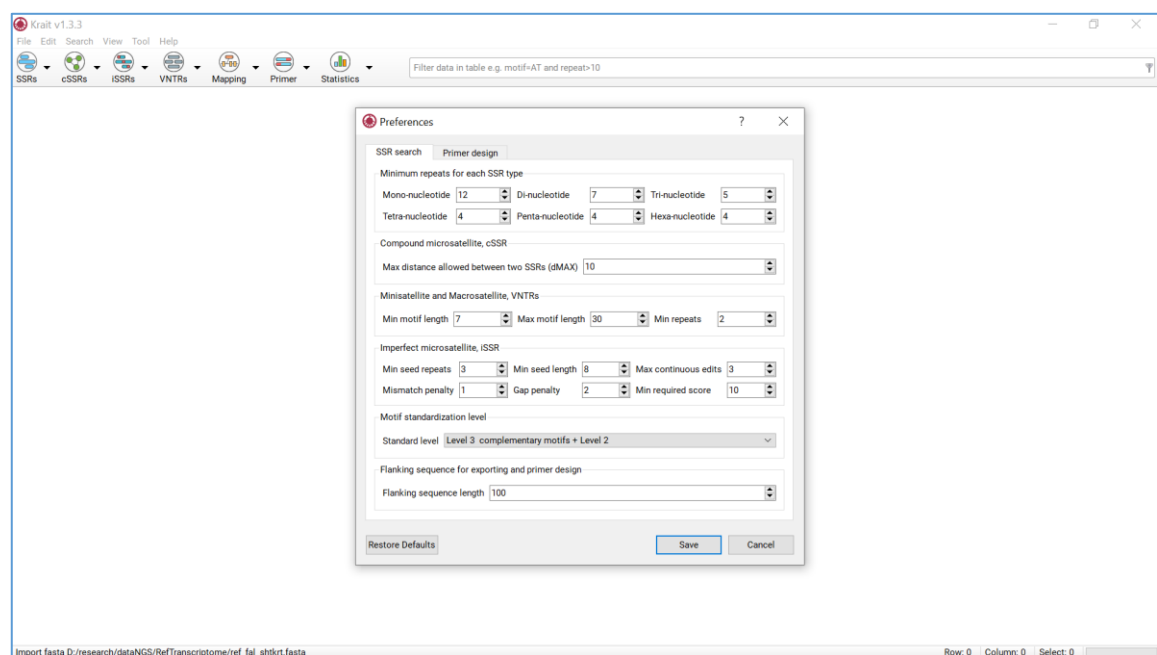
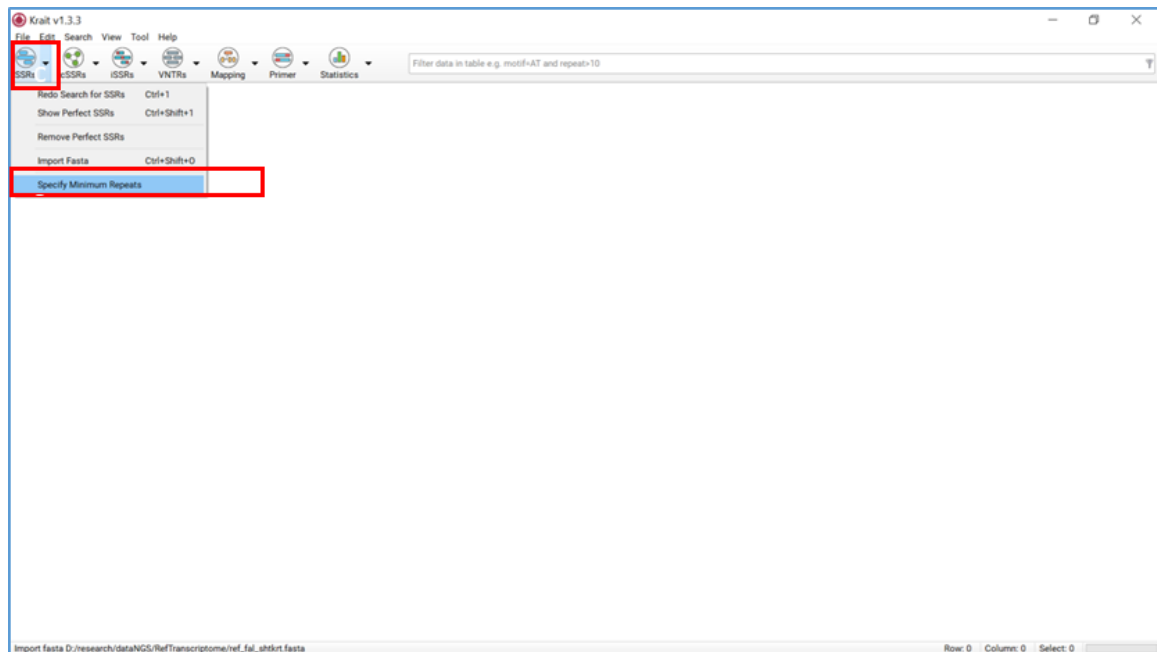
Salah satu program yang dapat digunakan untuk mencari situs *microsatellite* sekaligus mendesain primer yaitu Krait (Du *et al.* 2018). Kelebihan Krait yaitu pengoperasiannya berdasarkan *Graphic User Interface* (GUI), bukan menggunakan *command prompt* atau terminal seperti pada program MISA. Selain itu, Krait juga dapat melakukan analisis terhadap *compound microsatellite* atau *microsatellite* tandem, ISSR, *Variable Number Tandem Repeat* (VNTR), dan menggabungkan data *mikrosatelit* dengan hasil anotasi. Kelebihan lainnya yaitu program Krait dapat secara cepat mendesain seluruh primer *microsatellite* yang terdeteksi dan memenuhi kriteria primer tanpa memasukkan kembali sekuens satu per satu seperti pada tahap pembuatan primer secara konvensional.

Tahapan analisis *microsatellite* menggunakan Krait adalah sebagai berikut:

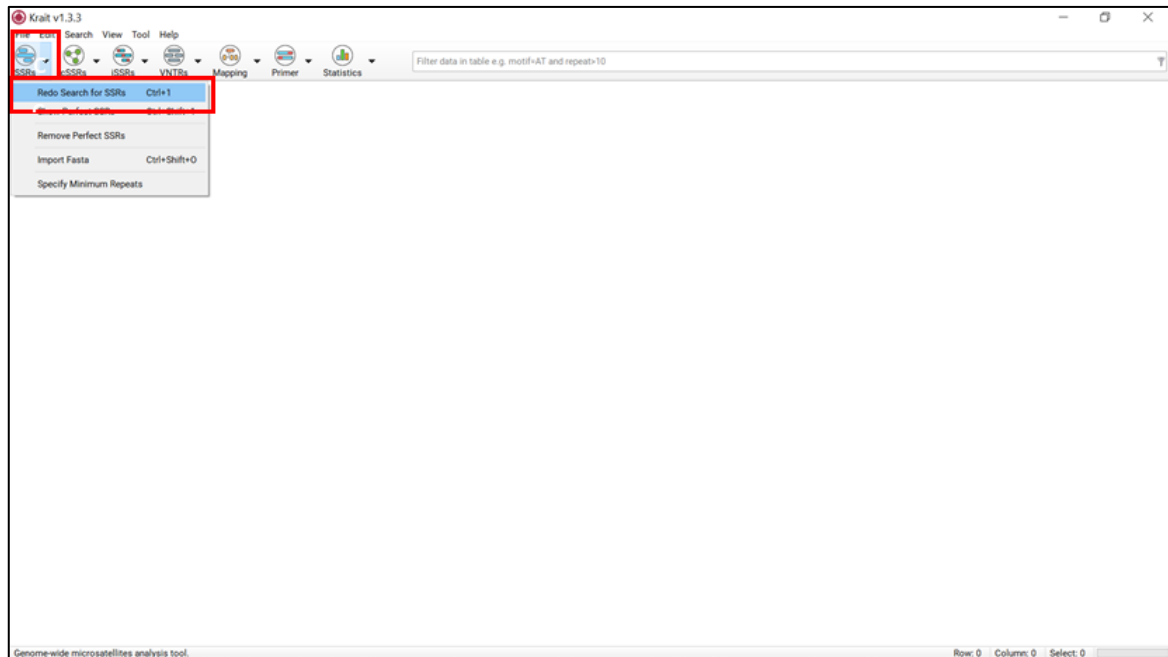
1. *Import fasta*



- Atur jumlah minimum pengulangan motif. Umumnya pengulangan motif minimum yang digunakan yaitu 10 ulangan untuk mononukleotida, 6 ulangan untuk dinukleotida, dan 5 ulangan untuk tri-, tetra-, penta-, dan hexanukleotida.



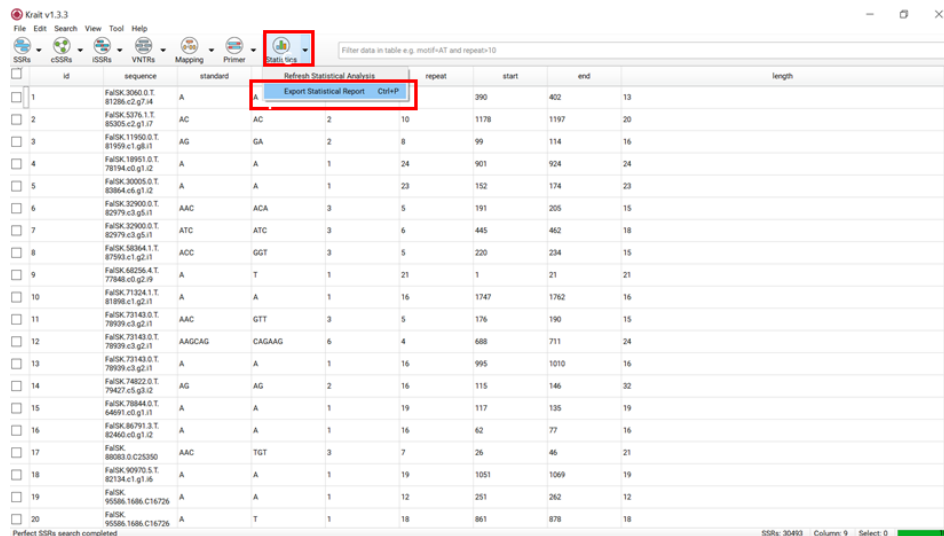
3. Klik SSR, lalu pilih *search* SSR untuk mencari *perfect SSR* dan hasilnya adalah sebagai berikut:



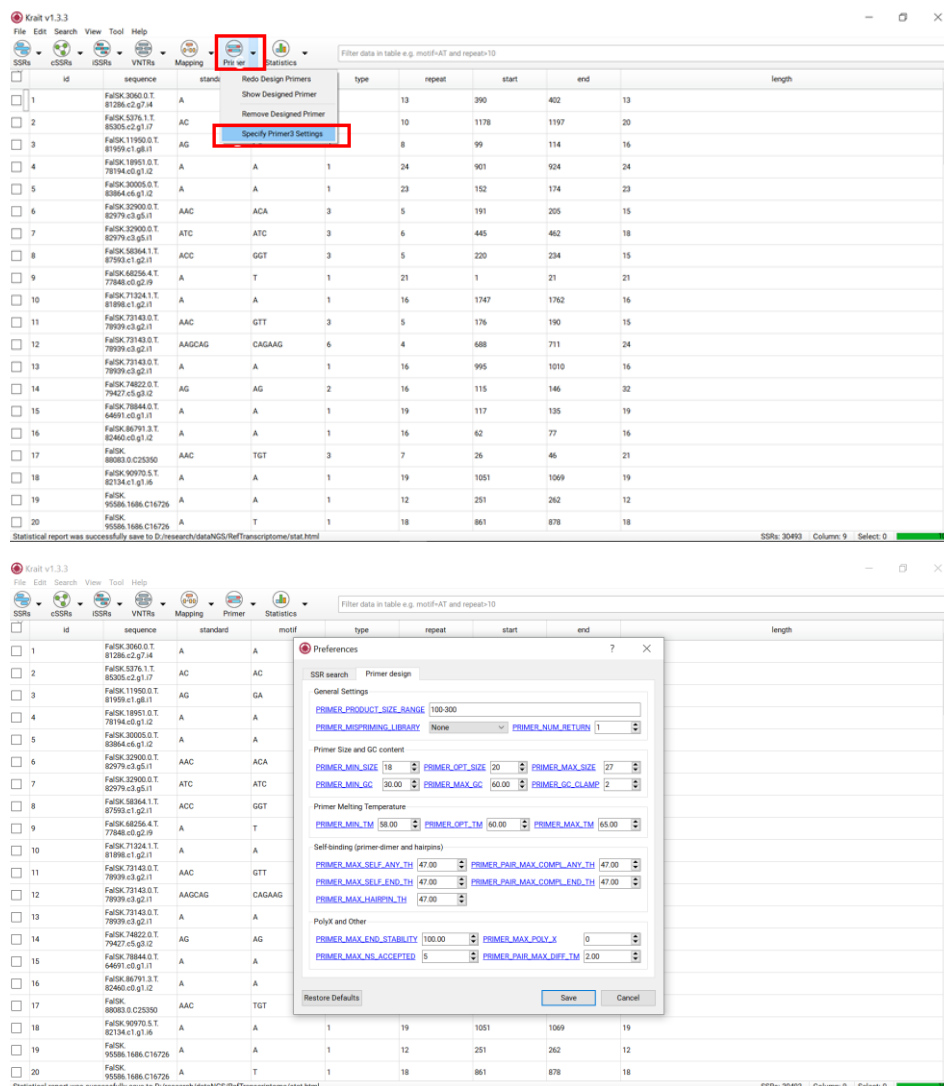
The screenshot shows the Krait v1.3.3 software window with the results of the SSR search displayed in a table. The table has columns for id, sequence, standard, motif, type, repeat, start, end, and length. The status bar at the bottom indicates 'Perfect SSRs search completed' and 'SSRs: 30493 Column: 9 Select: 0'.

id	sequence	standard	motif	type	repeat	start	end	length
1	FaSK 3060.0.T 81286.c2.g7.i4	A	A	1	13	390	402	13
2	FaSK 5376.1.T 85305.c2.g1.i7	AC	AC	2	10	1178	1197	20
3	FaSK 11950.0.T 81559.c1.g8.i1	AG	GA	2	8	99	114	16
4	FaSK 18951.0.T 78194.c0.g1.i2	A	A	1	24	901	924	24
5	FaSK 30005.0.T 83864.c6.g1.i2	A	A	1	23	152	174	23
6	FaSK 32900.0.T 82979.c3.g5.i1	AAC	ACA	3	5	191	205	15
7	FaSK 32900.0.T 82979.c3.g5.i1	ATC	ATC	3	6	445	462	18
8	FaSK 58364.1.T 87593.c1.g2.i1	ACC	GGT	3	5	220	234	15
9	FaSK 68256.4.T 77848.c0.g2.i9	A	T	1	21	1	21	21
10	FaSK 71324.1.T 81898.c1.g2.i1	A	A	1	16	1747	1762	16
11	FaSK 73143.0.T 78939.c3.g2.i1	AAC	GTT	3	5	176	190	15
12	FaSK 73143.0.T 78939.c3.g2.i1	AAGCAG	CAGAA	6	4	688	711	24
13	FaSK 73143.0.T 78939.c3.g2.i1	A	A	1	16	995	1010	16
14	FaSK 74822.0.T 79427.c5.g3.i2	AG	AG	2	16	115	146	32
15	FaSK 78844.0.T 64691.c0.g1.i1	A	A	1	19	117	135	19
16	FaSK 86791.3.T 82460.c0.g1.i2	A	A	1	16	62	77	16
17	FaSK 88083.0.C25350	AAC	TGT	3	7	26	46	21
18	FaSK 90970.5.T 82134.c1.g1.i6	A	A	1	19	1051	1069	19
19	FaSK 95586.1686.C16726	A	A	1	12	251	262	12
20	FaSK 95586.1686.C16726	A	T	1	18	861	878	18

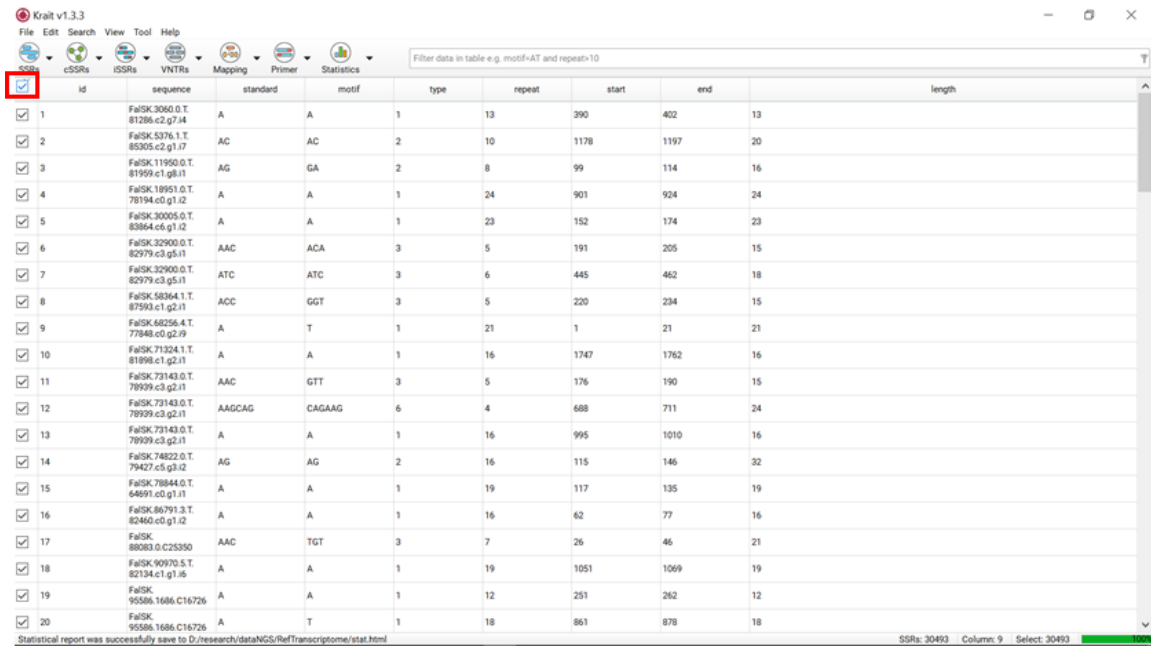
4. Klik *statistic* → *export statistical report* untuk melihat ringkasan data.



5. Klik primer → *specify primer settings* untuk mengatur kriteria primer yang sesuai untuk *microsatellite*.



6. Klik kotak di pojok kiri data untuk memilih semua data, kemudian pilih primer, *redo design primers*.



Krait v1.3.3

File Edit Search View Tool Help

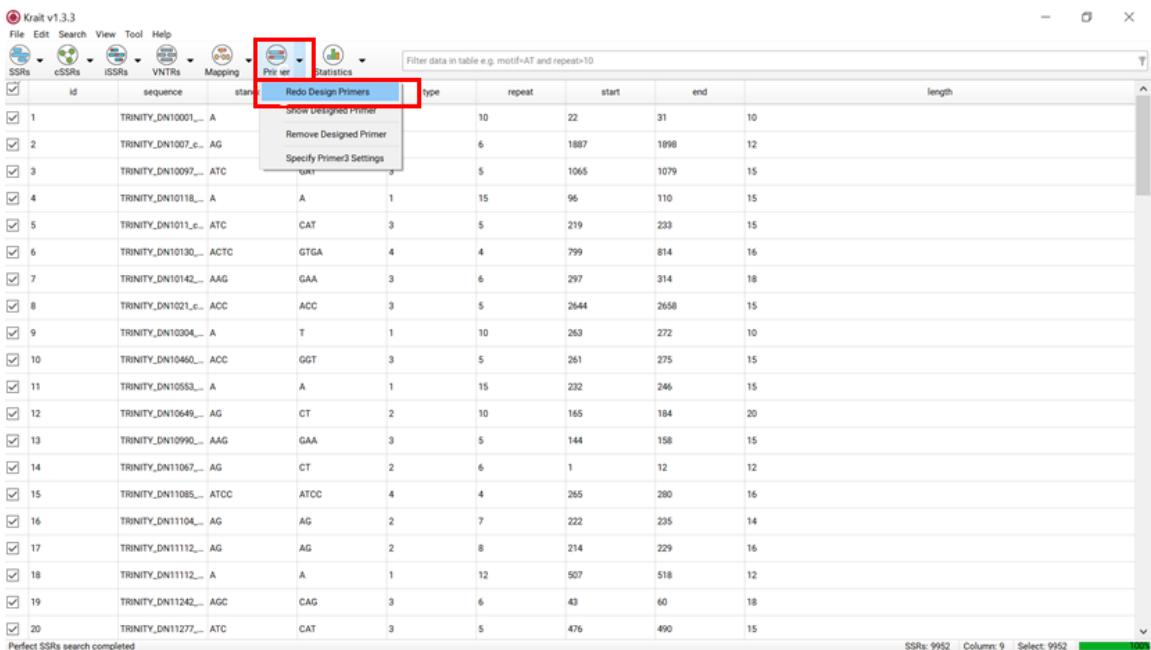
SSRs cSSRs VNTRs Mapping Primer Statistics

Filter data in table e.g. motif=AT and repeat=10

	id	sequence	standard	motif	type	repeat	start	end	length
☒	1	FaISK_3060.0.T 81286.c2.g1.14	A	A	1	13	390	402	13
☒	2	FaISK_5376.1.T 85305.c2.g1.17	AC	AC	2	10	1178	1197	20
☒	3	FaISK_11950.0.T 81959.c1.g8.11	AG	GA	2	8	99	114	16
☒	4	FaISK_16951.0.T 78194.c0.g1.12	A	A	1	24	901	924	24
☒	5	FaISK_30005.0.T 83864.c6.g1.12	A	A	1	23	152	174	23
☒	6	FaISK_32900.0.T 82979.c3.g5.11	AAC	ACA	3	5	191	205	15
☒	7	FaISK_32900.0.T 82979.c3.g5.11	ATC	ATC	3	6	445	462	18
☒	8	FaISK_58364.1.T 87593.c1.g2.11	ACC	GGT	3	5	220	234	15
☒	9	FaISK_68256.4.T 77848.c0.g2.19	A	T	1	21	1	21	21
☒	10	FaISK_71524.1.T 81988.c1.g2.11	A	A	1	16	1747	1762	16
☒	11	FaISK_73143.0.T 78939.c3.g2.11	AAG	GTT	3	5	176	190	15
☒	12	FaISK_73143.0.T 78939.c3.g2.11	AAGCAG	CAGAAG	6	4	688	711	24
☒	13	FaISK_73143.0.T 78939.c3.g2.11	A	A	1	16	995	1010	16
☒	14	FaISK_74822.0.T 79427.c5.g3.12	AG	AG	2	16	115	146	32
☒	15	FaISK_78844.0.T 64691.c0.g1.11	A	A	1	19	117	135	19
☒	16	FaISK_86791.3.T 82460.c0.g1.12	A	A	1	16	62	77	16
☒	17	FaISK_88083.0.C25350	AAG	TGT	3	7	26	46	21
☒	18	FaISK_90970.5.T 82134.c1.g1.16	A	A	1	19	1051	1069	19
☒	19	FaISK_95586.1686.C16726	A	A	1	12	251	262	12
☒	20	FaISK_95586.1686.C16726	A	T	1	18	861	878	18

Statistical report was successfully save to D:\research\data\NGS\ref\Transcriptome\stat.html

SSRs: 30493 Column: 9 Select: 30493



Krait v1.3.3

File Edit Search View Tool Help

SSRs cSSRs VNTRs Mapping Primer Statistics

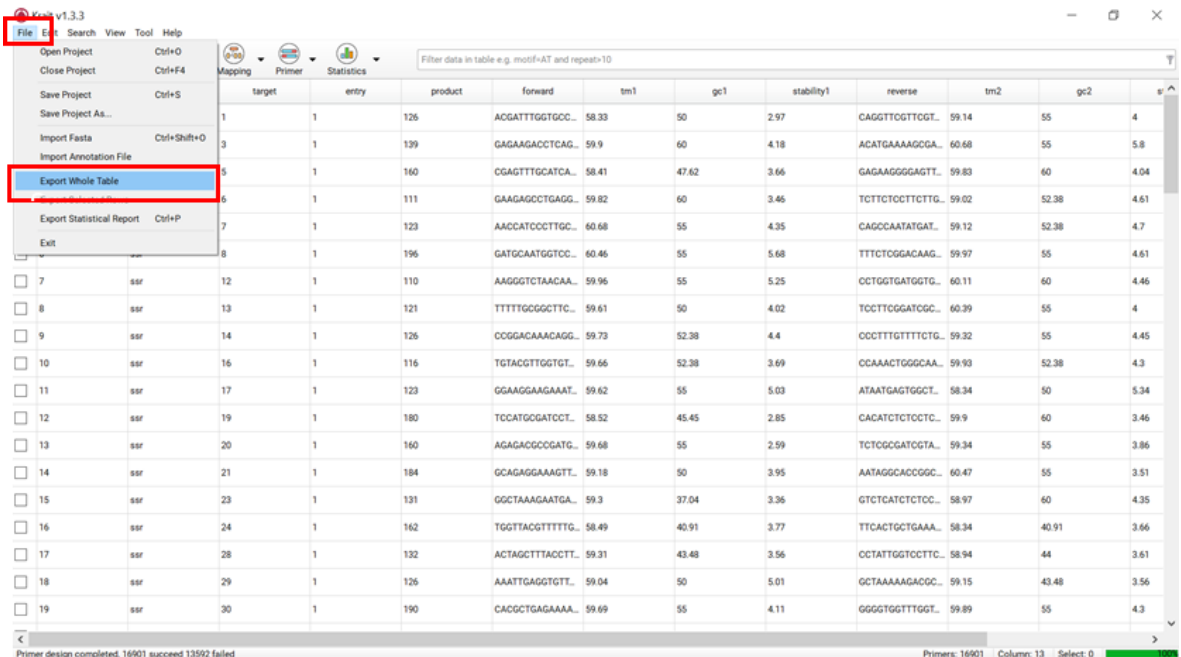
Filter data in table e.g. motif=AT and repeat=10

	id	sequence	standard	motif	type	repeat	start	end	length
☒	1	TRINITY_DN10001_..._	A			10	22	31	10
☒	2	TRINITY_DN1007_..._	AG			6	1887	1898	12
☒	3	TRINITY_DN10097_..._	ATC			5	1065	1079	15
☒	4	TRINITY_DN10118_..._	A			15	96	110	15
☒	5	TRINITY_DN10111_..._	ATC	CAT	3	5	219	233	15
☒	6	TRINITY_DN10130_..._	ACTC	GTGA	4	4	799	814	16
☒	7	TRINITY_DN10142_..._	AAG	GAA	3	6	297	314	18
☒	8	TRINITY_DN1021_..._	ACC	ACC	3	5	2644	2658	15
☒	9	TRINITY_DN10304_..._	A	T	1	10	263	272	10
☒	10	TRINITY_DN10460_..._	ACC	GGT	3	5	261	275	15
☒	11	TRINITY_DN10553_..._	A	A	1	15	232	246	15
☒	12	TRINITY_DN10649_..._	AG	CT	2	10	165	184	20
☒	13	TRINITY_DN10990_..._	AAG	GAA	3	5	144	158	15
☒	14	TRINITY_DN11067_..._	AG	CT	2	6	1	12	12
☒	15	TRINITY_DN11085_..._	ATCC	ATCC	4	4	265	280	16
☒	16	TRINITY_DN11104_..._	AG	AG	2	7	222	235	14
☒	17	TRINITY_DN11112_..._	AG	AG	2	8	214	229	16
☒	18	TRINITY_DN11112_..._	A	A	1	12	507	518	12
☒	19	TRINITY_DN11242_..._	AGC	CAG	3	6	43	60	18
☒	20	TRINITY_DN11277_..._	ATC	CAT	3	5	476	490	15

Perfect SSRs search completed

SSRs: 9952 Column: 9 Select: 9952

- Pilih file → *export whole table* untuk mengekspor tabel hasil, baik tabel SSR yang dihasilkan maupun primer. Jika ingin mengekspor tabel hasil pencarian situs SSR, maka jangan lakukan pencarian primer terlebih dahulu.



Catatan:

Catatan:

Catatan:

DAFTAR PUSTAKA

- Afgan E, Baker D, van den Beek M, Blankenber D, Bouvier D. 2016. The galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nuc Acid Res.* 44:W3-W10.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol.* 215(3): 403–10.
- Andrews S. 2010. FastQC: A quality control tool for high throughput sequence data [internet]. Tersedia pada <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> . Diakses tanggal 18 Oktober 2018.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15): 2114-20.
- Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. 2005. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21: 3674–6.
- Dalirsefat SB, da Silva Meyer A, Mirhoseini SZ. Comparison of similarity coefficients used for cluster analysis with amplified fragment length polymorphism markers in the silkworm, *Bombyx mori*. *J Insect Sci* 2009(9):1-8. DOI: 10.1673/031.009.7101. PMID: 20050782; PMCID: PMC3011968.
- Davidson NM, Oshlack A. 2014. Corset: Enabling differential gene expression analysis for *de novo* assembled transcriptomes. *Genome Biol* 15: 410.
- Du L, Zhang C, Liu Q, Zhang X, Yue B. 2018. Krait: An ultrafast tool for genome-wide survey of microsatellites and primer design. *Bioinformatics* 34(4): 681-3.
- Earld DA, vonHoldt BM. 2012. Structure Harvester: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Res* 4(2): 359–61.
- Hall TA. 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41:95-8.
- Huang X, Madan A. 1999. CAP3: A DNA sequence assembly program. *Genome Res* 9(9):868–77.
- Jain M, Olsen HE, paten B, Akeson M. 2016. The Oxford Nanopore MinION: Delivery of nanopore sequencing to the genomics community. *Genome Biology* 17: 239.
- Korpelainen E, Tuimala J, Somervuo P, Huss M, Wong G. 2014. RNA-Seq Data Analysis: A practical approach. Florida (US): CRC Press.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie2. *Nat Methods* 9:357–9.

- Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22(13): 1658–9.
- Long, SP, Ort DR. 2010. More than taking the heat: Crops and global change. *Curr Opin Pl Biol* 13: 240–7.
- Peakall R, Smouse PE. 2012. GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* 28:2537–9.
- Perrier X, Jacquemoud-Collet JP. 2006. DARwin software <http://darwin.cirad.fr/>
- Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155(2):945-59.
- Prokic J. 2010. Families and Resemblance. [Dissertation]. University of Gröningen.
- Roberts A, Pachter L. 2013. Streaming fragment assignment for real-time analysis of sequencing experiments. *Nat Methods* 10: 71-3.
- Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: A bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26(1): 139–40.
- Shih TH, Lin SH, Huang MY, Sun CW, Yang CM. 2018. Transcriptome profile of cup-shaped galls in *Litsea acuminata* leaves. *PloS ONE*. 13(10): e0205265.
- Song H, Wang P, Lin JY, Zhao C, Bi Y, Wang X. 2016. Genome-Wide identification and characterization of WRKY gene family in peanut. *Front Plant Sci* 7:534.
- Kinjo S, Monma N, Misu S, Kitamura N, Imoto J, Yoshitake K, Ikeo K. 2017. Maser: A cloud-based platform to manage, analyze, and visualize NGS data for basic researches in life science. (in prep) <http://cell-innovation.nig.ac.jp/maser/>.
- Tasma MI. 2015. Pemanfaatan teknologi sekuensing genom untuk mempercepat program pemuliaan tanaman. *J. Litbang Pert.* 34(4):159-68.
- Unamba CIN, Nag A, Sharma RK. 2015. Next generation sequencing technologies: The doorway to the unexplored genomics of non-model plants. *Front Plant Sci* 6: 1074.
- Wong MML, Cannon CH, Wickneswari R. 2011. Identification of lignin genes and regulatory sequences involved in secondary cell wall formation in *Acacia auriculiformis* and *Acacia mangium* via *de novo* transcriptome sequencing. *BMC Genomics* 12: 342.

TENTANG PENULIS



Dewi Rahmawati adalah staf SEAMEO BIOTROP (Southeast Asian Regional Centre for Tropical Biology) dan staf PNS IPB (Institut Pertanian Bogor) di bawah Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi. Beliau menyelesaikan strata satu dari Departemen Biologi, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Pertanian Bogor (IPB University) dan strata dua dari Program Studi Pemuliaan dan Bioteknologi Tanaman, IPB University. Beliau banyak membimbing mahasiswa magang dan mahasiswa yang melakukan penelitian dari berbagai universitas di Indonesia. Pengajar di program Merdeka Belajar Kampus Merdeka Kegiatan MSIB (Magang dan Studi Independen Bersertifikat) mata kuliah Bioteknologi Modern (2021-sekarang). Beliau juga menjadi narasumber pelatihan molekuler dan berpengalaman di penelitian molekuler dan kultur jaringan serta sudah melahirkan beberapa publikasi ilmiah di bidang Bioteknologi. Buku yang pernah disusun antara lain: Teknik Pengoperasian Laboratorium Bioteknologi (2020), Teknologi Kultur Jaringan di Laboratorium Bioteknologi (2021), Biodiversitas dan Pemodelan DNA, Teknik Kultur Jaringan Pisang Cavendish dan Teknik Molekuler untuk Mengidentifikasi Spesies Tanaman (unpublished). Artikel beliau berjudul “The Morphological, Wood Anatomical and Phytochemical Characterization of Susceptible and Resistant Sengon (*Falcateria moluccana*) Tree To Gall Rust Disease” terbit di Jurnal Biodiversitas Volume 20, Number 3, March 2019, halaman 907-913. Artikel lainnya berjudul “Fingerprinting Sengon (*Falcateria moluccana*) Accessions Resistant to Boktor Pest and Gall Rust Disease using Microsatellite Markers juga terbit di Jurnal Biodiversitas Volume 20, Nomor 9, September 2019, halaman 1698-1706. Pada tahun 2022, artikel beliau berjudul “Pengaruh Radiasi Sinar Gamma terhadap Daya Kecambah *Acasia mangium* Generasi M2” terbit di Jurnal Perbenihan Tanaman Hutan Volume 10 Nomor 1, Agustus 2022, halaman 23-36. Beliau juga aktif menulis di majalah populer, di antaranya: *Modern Biology for the Future*, terbit di BIOTROP Courier Newsletter Volume XXIII, Quarter 1 March 2021, ISSN 0303-8807 (media cetak); *NanoPhotometer Technology* terbit di BIOTROP Courier Newsletter, Vol XXIII, Quarter 3 September 2021, ISSN 0303-8807 (media cetak); dan *Biotechnology for Tropical Plant Breeding in SEAMEO BIOTROP* terbit di majalah BIODIVERS Volume 1, Nomor 2 tahun 2022. *Biodiversity Genetic Conservation as a Key to Support Agro-Eco-Edu-Tourism Program in Southeast Asia* terbit di majalah BIODIVERS Vol.2 No.1 tahun 2023. *Micropropagation of Banana Plant Cavendish (*Musa acuminata* L.) Using Shoot-tip Culture to Support Food Security* terbit di majalah BIODIVERS Vol. 2 No. 2 tahun 2023.

TENTANG PENULIS



Hasyati Shabrina adalah dosen di Program Studi Kehutanan, Fakultas Pertanian, Universitas Mataram. Beliau menyelesaikan program S1 di Departemen Silvikultur, Fakultas Kehutanan Institut Pertanian Bogor (IPB University), dan S3 di Program Studi Silvikultur Tropika, Sekolah Pascasarjana IPB University dengan Beasiswa Program Magister menuju Doktor untuk Sarjana Unggul (PMDSU). Beberapa tulisan beliau yang telah terbit antara lain: The Dataset of de novo Transcriptome Assembly of *Falcataria moluccana* Cambium from Gall Rust (*Uromycladium falcatarium*) Infected and Non-infected tree (Data in Brief Volume 26 tahun 2019, halaman 104489), Expressed

Sequence Tag Simple Sequence Repeats (EST-SSRs) Mining and Marker Development from *Leucaena leucocephala* Root Transcriptome (IOP Conference Series: Earth and Environmental Science Volume 959, Nomor 1, 2022), dan Konfirmasi Jenis dan Keragaman Genetik Sengon Resisten dan Rentan Infeksi Karat Tumor menggunakan Penanda DNA Kloroplas (Jurnal Penelitian Hutan Tanaman, Volume 17, Nomor 2, 2020).

Buku ini terbit seiring dengan kebutuhan bioteknologi modern yang sejalan dengan semakin berkembangnya program-program bioinformatika. Kombinasi kemajuan bioteknologi dan teknologi informasi yaitu Bioinformatika merupakan penggunaan IT untuk menyelesaikan permasalahan yang berhubungan dengan Biologi Molekuler, yang dilatarbelakangi oleh ledakan data observasi biologi sebagai hasil pencapaian kemajuan bioteknologi.

Eksperimen Biologi Molekuler telah mengumpulkan banyak sekali data ekspresi genetika. Data dalam jumlah banyak tersebut perlu diolah menjadi informasi yang berguna. Kemajuan ilmu bioinformatika ini lebih didesak lagi oleh *genome project* yang dilaksanakan di seluruh dunia dan menghasilkan tumpukan informasi gen dari berbagai makhluk hidup, mulai makhluk hidup tingkat rendah sampai makhluk hidup tingkat tinggi, sehingga diperlukan sistem otomatis dan efisien untuk menyimpan dan menganalisis data. Kemajuan di dunia biologi molekuler, khususnya bioinformatika, sangatlah cepat. Pengorganisasian data untuk mengelola informasi biologis pensejajaran sekuens (*sequence alignment*) yang ada sangatlah berguna untuk hasil analisis yang lebih akurat. Pembahasan di bidang bioinformatika ini tidak terlepas dari perkembangan Biologi Molekuler modern, salah satunya peningkatan pemahaman manusia dalam bidang genomik yang terdapat di dalam molekul DNA.

Di bidang biologi molekuler, salah satu *output*-nya adalah data-data molekuler, yang memerlukan analisis molekuler lebih lanjut. Bioinformatika dapat digunakan sebagai manajemen informasi di bidang penyimpanan data *in silico* dari eksperimen biologi molekuler. Dalam eksperimen biologi molekuler, data yang dihasilkan perlu diolah dengan pendekatan khusus untuk menjadi informasi yang bermanfaat. Pemahaman bioinformatika sangat penting untuk meningkatkan kemampuan komputasi dan analisis data melalui pengadaan sumber daya komputasi dan penggunaan kluster informatika publik.

Buku Analisis Molekuler dan Bioinformatika ini merupakan penerapan teknologi informasi pada bidang biologi dengan mengorganisasikan dan menganalisis data-data genetika dari *database* DNA, RNA dan protein menjadi sebuah informasi biologi yang bermakna. Buku ini diharapkan dapat memberikan tambahan pengetahuan tentang cara mengakses dan memanfaatkan sistem bioinformatika melalui metode analisis data molekuler dengan *software* terkini. Materi buku ini disusun secara berurutan yang terdiri dari beberapa topik utama materi dan tahapan, yaitu: 1) Analisis molekuler berbasis fragmen DNA (A. skoring pita hasil elektroforesis menggunakan *software* PyElph; B. Klasterisasi menggunakan perangkat lunak NTSYSPC; C. Pengolahan data pita genetik berbasis biner (1 dan 0) menggunakan program *software* DARWIN; D. Pengolahan data fragmen DNA menggunakan GenAlEx 6.5; E. Analisis struktur genetika menggunakan *software* Structure v. 2.3.4), 2) Pengantar Bioinformatika, 3) Pengenalan *database* Bioinformatika; 4) Analisis Data Sekuens Pendek (A. *Editing* sekuens menggunakan aplikasi Bioedit; B. Pembuatan *alignment* multi sekuens, analisis kekerabatan, dan pembuatan pohon filogenetik dari sekuens menggunakan *software* MEGA11); 5) Desain Primer (A. Contoh desain primer menggunakan fitur pada NCBI, B. Analisis struktur sekunder menggunakan NetPrimer, C. Mendesain primer menggunakan primer3plus); 6) *Next Generation Sequencing* (A. P, pengolahan data RNA-Seq dengan komputer, B. Pengolahan data RNA-Seq menggunakan *cloud computing* pada MASER, C. *Cloud Computing* pada platform Galaxy, D. Teknologi *Nanopore Sequencing*, E. *Mikrosatelit finding* dan desain primer *mikrosatelit*). Buku ini diharapkan dapat memberi manfaat, membuka wawasan dan meningkatkan kompetensi bagi para pembacanya, sesuai dengan perkembangan zaman sehingga dapat mengembangkan metode pembelajaran dan mengimplementasikan ilmu dengan optimal.