

I. PENGENALAN NATIONAL CENTRE FOR BIOTECHNOLOGY INFORMATION (NCBI)

A. PENDAHULUAN

NCBI (*National Centre for Biotechnology Information*) merupakan suatu institusi yang menyediakan sumber informasi terkait perkembangan biologi molekuler. NCBI membuat database yang dapat diakses oleh publik dan mengembangkan software penganalisis data genom. Situs NCBI dapat di akses pada : www.ncbi.nlm.nih.gov.

Databases and Software

1. Entrez


Entrez merupakan 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 sekuen dan kromosom. Ciri khas dan keunggulan Entrez adalah kemampuan untuk pencarian informasi terkait sekuen, struktur dan referensi. Literatur jurnal yang tersedia dapat diakses melalui PubMed. PubMed merupakan alat penghubung pencarian di web yang menyediakan akses ke lebih dari 11 juta sitasi jurnal di MEDLINE. Entrez Gene adapat diakses pada www.ncbi.nlm.nih.gov/gene.

2. Nucleotide Database

Database nukleotida merupakan suatu koleksi sekuen dari beberapa sumber, termasuk diantaranya GenBank, Reference Sequence (RefSeq), Third Party Annotation (TPA) dan Protein Data Bank (PDB).

a. GenBank

GenBank merupakan database sekuen genetik dari NIH (*National Institutes of Health*), berupa koleksi sekuen DNA yang dapat diketahui oleh publik. Database GenBank dibiayai dan didistribusikan oleh NCBI. Data sekuen dikirim ke GenBank oleh peneliti dari seluruh dunia.


NCBI

GenBank Overview

[PubMed](#)
[Entrez](#)
[BLAST](#)
[OMIM](#)
[Books](#)
[Taxonomy](#)
[Structure](#)

Search for

[NCBI Home](#)
[NCBI Site Map](#)
[Submit to GenBank](#)
[Submit an update](#)
[Search GenBank](#)
[GenBank and RefSeq:
a comparison](#)
[BLAST](#)

► **What is GenBank?**

GenBank® is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences ([Nucleic Acids Research, 2008 Jan;36\(Database issue\):D25-30](#)). There are approximately 106,533,156,756 bases in 108,431,692 sequence records in the traditional GenBank divisions and 148,165,117,763 bases in 48,443,067 sequence records in the WGS division as of August 2009.

The complete [release notes](#) for the current version of GenBank are available on the NCBI ftp site. A new release is made every two months. GenBank is part of the [International Nucleotide Sequence Database Collaboration](#), which comprises the DNA DataBank of Japan (DDBJ), the European Molecular Biology Laboratory (EMBL), and GenBank at NCBI. These three organizations exchange data on a daily basis.

Format GenBank disajikan pada gambar dibawah ini :

[Display Settings:](#) ☒ GenBank

Pyrococcus woesei beta-galactosidase gene, complete cds

GenBank: AF043283.1

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

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

```

LOCUS       AF043283                1533 bp    DNA        linear    BCT 25-MAY-2001
DEFINITION  Pyrococcus woesei beta-galactosidase gene, complete cds.
ACCESSION   AF043283
VERSION     AF043283.1   GI:2811285
KEYWORDS    .
SOURCE      Pyrococcus woesei
  ORGANISM  Pyrococcus woesei
            Archaea; Euryarchaeota; Thermococci; Thermococcales;
            Thermococcaceae; Pyrococcus.
REFERENCE   1  (bases 1 to 1533)
  AUTHORS   Daabrowski,S., Sobiewska,G., Maciunska,J., Synowiecki,J. and Kur,J.
  TITLE     Cloning, expression, and purification of the His(6)-tagged
            thermostable beta-galactosidase from Pyrococcus woesei in
            Escherichia coli and some properties of the isolated enzyme
  JOURNAL   Protein Expr. Purif. 19 (1), 107-112 (2000)
  PUBMED   10833397
REFERENCE   2  (bases 1 to 1533)
  AUTHORS   Dabrowski,S., Maciunska,J. and Synowiecki,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1998) Food Preservation, Technical University of
            Gdansk, Narutowicza 11/12, Gdansk 80-952, Poland

```

```

FEATURES                     Location/Qualifiers
     source                   1..1533
                              /organism="Pyrococcus woesei"
                              /mol_type="genomic DNA"
                              /strain="DSM3773"
                              /db_xref="taxon:2262"
     CDS                      1..1533
                              /codon_start=1
                              /transl_table=11
                              /product="beta-galactosidase"
                              /protein_id="AAB97862.1"
                              /db_xref="GI:2811286"
                              /translation="MFPEKFLWGVAQSGFQFEMGDKLRNIDTNTDWWHVVDRDKNIE
KGLVSGDLPEEGINNYELYEKDHEIARKLGLNAYRIGIEWSRIFPWPTTFIDVDYSYN
ESYNLIEDVKITKDTLEELDEIANKREVAYYRSVINSLSKGFKVIVNLSNHTLPYWL
HDPLEARERALTNKRNGWVNPRTVIEFAKYAAYIAYKFGDIVDMWSTFNEPMVVVELG
YLAPYSGFPPGVLNPEAAKLAILHMINAHALAYRQIKKFDTEKADKDSKEPAEVGIIY
NNIGVAYPKDPNDSKDVKAAENDNFFHSGLFFEAHKGKLNIEFDGETFIDAPYLKGN
DWIGVNYTREVVTYQEPMPFSIPLITFKGVQGYGYACRPGTLSKDDRPVSDIGWELY
PEGMYDSIVEAHKYGVPPVYVTENGIADSKDILRPYYIASHIKMTAKAFEDGYEVKGYF
HWALTDNFEWALGFRMRFLYEVNLTIRIPREKSVSIFREIVANNGVTKKIEEELL
RG"

```

Coding Sequence

Amino acid sequence

Start codon

```

ORIGIN
1  atgctccctg aaaagtccct ttgggggtgtg gcacaatcgg gttttcagtt tgaaatgggg
61  gataaactca ggaggaatat tgacactaac actgattggt ggcactgggt aaggggataag
121  acaaatatag agaaaggcct cgttagtggg gatcttcccg aggaggggat taacaattac
181  gagctttatg agaaggacca tgagattgca agaaagctgg gtcttaatgc ttacagaata
241  ggcataagat ggagcagaat attcccatgg ccaacgacat ttattgatgt tgattatagc
301  tataatgaat catataacct tatagaagat gtaaatgaca ccaaggacac ttggaggag
361  ttgatgaga tcgccaacaa gagggaggtg gcctactata ggtcagtcac aaacagcctg
421  aggagcaagg ggtttaaggt tatagttaat ctaaatcact tcacccttcc atattggttg
481  catgatccca ttgaggctag ggagaggcgg ttaactaata agagggaacgg ctgggttaac
541  ccaagaacag ttatagagtt tgcaaatgat gccgcttaca tagcctataa gtttgagat
601  atagtggata tgtggagcac gtttaatgag cctatgggtg ttgttgagct tggctaccta
661  gccccctact ctggcttccc tccagggtgt ctaaatccag aggccgcaaa gctggcgata
721  cttcacatga taaatgcaca tgctttagct tataggcaga taaagaagtt tgacactgag
781  aaagctgata aggattctaa agagcctgca gaagttggta taatttacia caacattgga
841  gttgcttata ccaaggatcc gaacgattcc aaggatgtta aggcagcaga aaacgacaa
901  ttcttccact cagggtctgt cttcgaggcc atacacaaag gaaaacttaa tatagagttt
961  gacggtgaaa cgtttataga tgccccctat ctaaaaggga atgactggat aggggttaat
1021  tactacacaa gggaagtagt tacgtatcag gaaccaatgt ttccttcaat cccgctgac
1081  acctttaagg gagttcaagg atatggctat gcctgcagac ctgggactct gtcaaggat
1141  gacagaccgg tcagcgacat aggatgggaa ctctatccag aggggatgta cgattcaata
1201  gttgaagctc acaagtacgg cgttccagtt tacgtgacgg agaacggaat agcggattca
1261  aaggacatcc taagacctta ctacatagcg agccacataa agatgacaga gaaggccttt
1321  gaggatgggt atgaagttaa gggctacttc cactgggcat taactgacaa cttcgagtgg
1381  gctctcgggt ttgaatgcg ctttggcctc tacgaagtca acctaattac aaaggagaga
1441  attcccgagg agaagagcgt gtcgatattc agagagatag tagccaataa tgggtgttacg
1501  aaaaagattg aagaggaatt gctgagggga tga

```

//

Stop codon

3. Blast

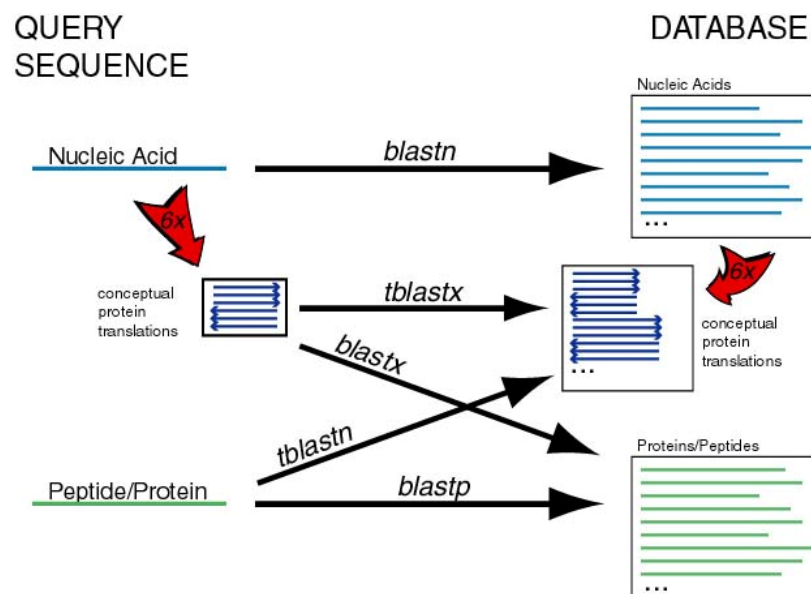
BLAST (*Basic Local Alignment Search Tool*) merupakan suatu program untuk pencarian kemiripan sekuen (*sequence similarity*) dan merupakan alat dalam identifikasi gen dan karakter genetik. Blast dapat melakukan pencarian sekuen melalui perbandingan

dengan database DNA dalam waktu singkat (kurang dari 15 detik). Keterangan lengkap mengenai program BLAST dapat dilihat pada :
<http://www.ncbi.nlm.nih.gov/blast/producttable.shtml#>

Ada 5 program utama dalam BLAST, yaitu :

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

Gambaran mengenai program Blast dapat dilihat pada gambar berikut ini :



HOMOLOGI PROTEIN

Protein merupakan polipeptida yang tersusun atas asam amino. Sekuens asam amino dari masing-masing protein ditentukan oleh gen yang meng-kodenya. Gen ditranskripsi menjadi mRNA dan selanjutnya mRNA ditranslasikan menjadi protein oleh ribosom. Urutan basa yang berbeda pada mRNA akan menghasilkan asam amino yang berbeda. Beberapa organisme dapat memiliki jenis protein yang sama. Sebagai contoh, hormon insulin terdapat pada manusia, babi, sapi dan beberapa hewan mamalia lainnya. Fungsi insulin pada masing-masing organisme tersebut adalah sama, yaitu mengubah glukosa menjadi glikogen sehingga kadar gula darah dapat terkontrol. Namun demikian, ternyata insulin pada masing-masing organisme memiliki struktur primer yang berbeda karena adanya perbedaan beberapa sekuens asam amino.

Oleh sebab itu diperlukan penelitian mengenai homologi protein antar-organisme sehingga dapat dibandingkan kemiripan asam amino yang menyusun protein. Melalui program *blastp* pada situs NCBI dapat dilakukan pencarian protein homolog dari berbagai macam organisme. Homologi bukan berarti sama persis, namun terdapat kemiripan antara satu dengan yang lainnya dengan persentase kemiripan tertentu. Untuk tujuan terapi, misalnya diinginkan suatu protein pengganti dari hewan tertentu, tentu akan dipilih protein yang memiliki persen kemiripan yang paling tinggi dengan protein yang kita miliki, sehingga diharapkan reaksi alergi tidak terjadi. Dalam bidang enzimologi, homologi protein diperlukan untuk mengetahui kemiripan sekuens asam amino suatu enzim yang dihasilkan oleh suatu organisme dibandingkan dengan organisme lain.

ACARA 1. ANALISIS HOMOLOGI SEKUEN NUKLEOTIDA DAN PROTEIN

TUJUAN

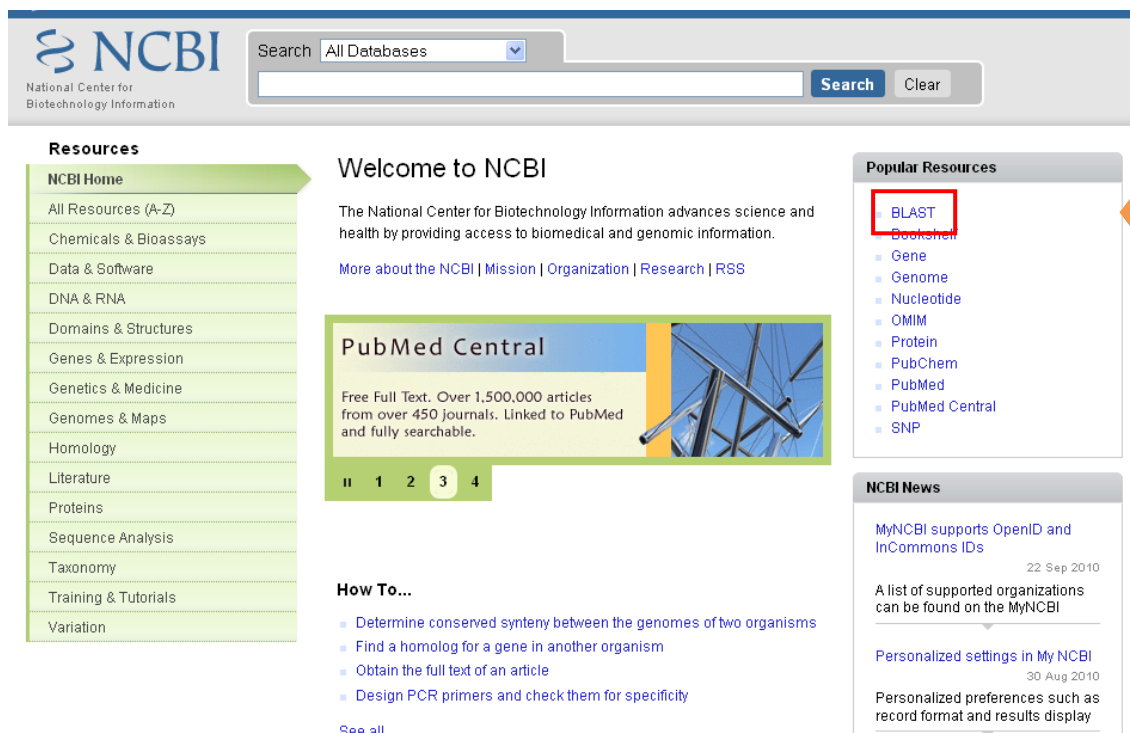
1. Mahasiswa mampu memahami dan menggunakan program Blastn untuk identifikasi sekuen nukleotida melalui database Genbank
2. Mahasiswa dapat melakukan analisis homologi suatu protein dari organisme tertentu dengan organisme lain menggunakan database Genbank menggunakan Blastp.

LANGKAH KERJA

a. Blastn

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

1. Buka situs www.ncbi.nlm.nih.gov



The screenshot shows the NCBI (National Center for Biotechnology Information) homepage. At the top, there is a search bar with a dropdown menu set to 'All Databases' and buttons for 'Search' and 'Clear'. Below the search bar, the page is divided into several sections. On the left, there is a 'Resources' sidebar with a list of links including 'NCBI Home', 'All Resources (A-Z)', 'Chemicals & Bioassays', 'Data & Software', 'DNA & RNA', 'Domains & Structures', 'Genes & Expression', 'Genetics & Medicine', 'Genomes & Maps', 'Homology', 'Literature', 'Proteins', 'Sequence Analysis', 'Taxonomy', 'Training & Tutorials', and 'Variation'. The main content area features a 'Welcome to NCBI' message, a 'PubMed Central' banner, and a 'How To...' section with links for determining conserved synteny, finding homologs, obtaining full text, and designing PCR primers. On the right, there is a 'Popular Resources' section where the 'BLAST' link is highlighted with a red box and an orange arrow. Below this, there is an 'NCBI News' section with updates on MyNCBI support for OpenID and InCommons IDs, and personalized settings.

2. Pilih tool "BLAST", akan muncul tampilan pilihan program BLAST.

Untuk mencari gen suatu sekuen nukleotida dari database nukleotida pilih "nucleotide blast" (blastn).

BLAST	Basic Local Alignment Search Tool		
Home	Recent Results	Saved Strategies	Help

NCBI/ BLAST Home

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

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

BLAST Assembled RefSeq Genomes

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

<input type="checkbox"/> Human	<input type="checkbox"/> Oryza sativa	<input type="checkbox"/> Gallus gallus
<input type="checkbox"/> Mouse	<input type="checkbox"/> Bos taurus	<input type="checkbox"/> Pan troglodytes
<input type="checkbox"/> Rat	<input type="checkbox"/> Danio rerio	<input type="checkbox"/> Microbes
<input type="checkbox"/> Arabidopsis thaliana	<input type="checkbox"/> Drosophila melanogaster	<input type="checkbox"/> Apis mellifera

Basic BLAST

Choose a BLAST program to run.

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

3. Setelah tampilan muncul, entri sekuen nukleotida (query) yang akan dicari; pilih setting pencarian dari database "others" (jika belum diketahui spesiesnya); pilih program "megablast"; klik "BLAST" untuk memulai proses searching.

Pada latihan/contoh digunakan sekuen nukleotida DNA berikut ini :

```
ATGTTCCCTGAAAAGTTCCTTTGGGGTGTGGCACAATCGGGTTTTTCAGTTTGAAATGGGGGATAAACTCA
GGAGGAATATTGACACTAACACTGATTGGTGGCACTGGGTAAAGGGATAAGACAAATATAGAGAAAGGCC
CGTTAGTGGAGATCTTCCCGAGGAGGGGATTAACAATTACGAGCTTTATGAGAAGGACCATGAGATTGCA
AGAAAGCTGGGTCTTAATGCTTACAGAATAGGCATAGAGTGGAGCAGAATATTTCCCATGGCCAACGACAT
TTATTGATGTTGATTATAGCTATAATGAATCATATAACCTTATAGAAGATGTAAAGATCACCAAGGACAC
TTTGGAGGAGTTAGATGAGATCGCCAACAAGAGGGAGGTGGCCTACTATAGGTCAGTCATAAACAGCCTG
AGGAGCAAGGGGTTTAAGGTTATAGTTAATCTAAATCACTTCACCCTTCCATATTGGTTGCATGATCCCA
TTGAGGCTAGGGAGAGGGCGTTAACTAATAAGAGGAACGGCTGGGTAAAC
```

blastn blastp blastx tblastn tblastx

Enter Query Sequence

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

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

Enter query help

From

To

Or, upload file [Browse...](#)

Job Title

AF043283:Pyrococcus woesei beta-galactosidase...

Enter a descriptive title for your BLAST search

☐ Align two or more sequences

Choose Search Set

Database

☐ Human genomic + transcript ☐ Mouse genomic + transcript ☒ Others (nr etc.):

Nucleotide collection (nr/nt)

Organism

Optional

Exclude

Optional

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

Entrez Query

Optional

Enter an Entrez query to limit search

Program Selection

Optimize for

☒ Highly similar sequences (megablast)

☐ More dissimilar sequences (discontiguous megablast)

☐ Somewhat similar sequences (blastn)

Choose a BLAST algorithm

BLAST

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

☒ Show results in a new window

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

Nucleotide Sequence (440 letters)

Query ID Id|64235

Description None

Molecule type nucleic acid

Query Length 440

Other reports: [Search Summary](#) [Taxonomy reports](#) [Distance tree of results](#)

Database Name nr

Description All GenBank+EMBL+DDBJ+PDB sequences (but no EST, STS, GSS, environmental samples or phase 0, 1 or 2 HTGS sequences)

Program BLASTN 2.2.24+ [Citation](#)

▼ Graphic Summary

Distribution of 4 Blast Hits on the Query Sequence

Mouse-over to show define and scores, click to show alignments

Color key for alignment scores

Query

1 80 160 240 320 400

▼ Descriptions

Legend for links to other resources: [U](#) UniGene [G](#) GEO [G](#) Gene [S](#) Structure [M](#) Map Viewer [P](#) PubChem BioAssay

Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
AE009950.1	Pyrococcus furiosus DSM 3638, complete genome	813	813	100%	0.0	100%	
AF043283.1	Pyrococcus woesei beta-galactosidase gene, complete cds	813	813	100%	0.0	100%	
U60214.1	Pyrococcus furiosus beta-mannosidase (bmnA) gene, complete cds	813	813	100%	0.0	100%	G
FU294509.1	Uncultured bacterium clone pWTS4 beta-galactosidase gene, complete cds	272	272	95%	1e-69	78%	


```
>gb|AF043283.1|AF043283 Pyrococcus woesei beta-galactosidase gene, complete cds
Length=1533
```

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

```
Query 1 ATGTTCCCTGAAAAGTTCCTTTGGGGTGTGGCACAATCGGGTTTTAGTTTGAAATGGGG 60
      |||
Sbjct 1 ATGTTCCCTGAAAAGTTCCTTTGGGGTGTGGCACAATCGGGTTTTAGTTTGAAATGGGG 60

Query 61 GATAAACTCAGGAGGAATATTGACACTAACACTGATTGGTGGCACTGGGTAAAGGATTAAG 120
      |||
Sbjct 61 GATAAACTCAGGAGGAATATTGACACTAACACTGATTGGTGGCACTGGGTAAAGGATTAAG 120

Query 121 ACAAATATAGAGAAAAGGCTCTGTTAGTGGAGATCTTCCCGAGGAGGGGATTAAACAATTAC 180
      |||
Sbjct 121 ACAAATATAGAGAAAAGGCTCTGTTAGTGGAGATCTTCCCGAGGAGGGGATTAAACAATTAC 180

Query 181 GAGCTTTATGAGAAAGGACCATGAGATTGCAAGAAAAGCTGGGTCTTAATGCTTACAGAAATA 240
      |||
Sbjct 181 GAGCTTTATGAGAAAGGACCATGAGATTGCAAGAAAAGCTGGGTCTTAATGCTTACAGAAATA 240

Query 241 GGCATAGAGTGGAGCAGAAATATTCCTCATGGCCAAACGACATTTATTGATGTTGATTATAGC 300
      |||
Sbjct 241 GGCATAGAGTGGAGCAGAAATATTCCTCATGGCCAAACGACATTTATTGATGTTGATTATAGC 300

Query 301 TATAATGAATCATATTAACCTTATAGAAAGATGTAAAGATCACCAAGGACACTTTGGAGGAG 360
      |||
Sbjct 301 TATAATGAATCATATTAACCTTATAGAAAGATGTAAAGATCACCAAGGACACTTTGGAGGAG 360

Query 361 TTAGATGAGATCGCCAAACAGAGGGAGGTGGCCTACTATAGGTCAGTCATAAACAGCCTG 420
      |||
Sbjct 361 TTAGATGAGATCGCCAAACAGAGGGAGGTGGCCTACTATAGGTCAGTCATAAACAGCCTG 420

Query 421 AGGAGCAAGGGGTTTAAAGGTTATAGTTAATCTAAATCACTTCACCCCTTCCATATTGTTG 480
      |||
Sbjct 421 AGGAGCAAGGGGTTTAAAGGTTATAGTTAATCTAAATCACTTCACCCCTTCCATATTGTTG 480

Query 481 CATGATCCCATTTGAGGCTAGGGAGAGGGCGTTAACTAATAGAGGAACGGCTGGGTTAAC 540
      |||
Sbjct 481 CATGATCCCATTTGAGGCTAGGGAGAGGGCGTTAACTAATAGAGGAACGGCTGGGTTAAC 540
```

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

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

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

Pyrococcus woesei beta-galactosidase gene, complete cds

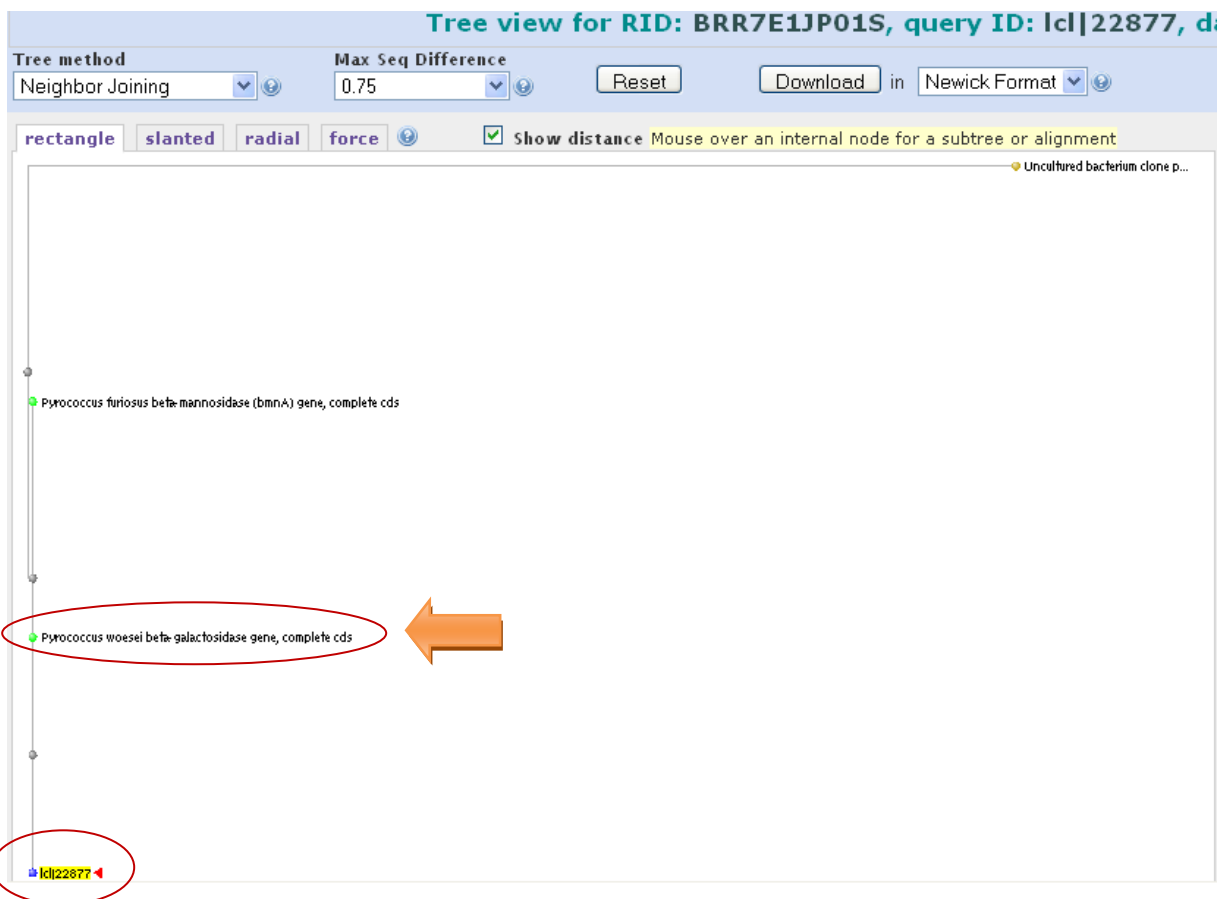
GenBank: AF043283.1

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

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

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

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



b. Blastp

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

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

Pada latihan / contoh digunakan query sekuen protein berikut ini :

```
mfpekflwgv aqsgfqfemg dklrrnidtn tdwwhwvrdk tniekglvsg dlpeeginny  
elyekdheia rklglnayri giowsrifpw pttdidvdys ynesynlied vkitkdtlee  
ldeiankrev aayrsvinsl rskgfkviwn lnhftlpywl hdpiearera ltnkrngwn  
prtviefaky aayiafkfgd ivdmwstfne pmvvvelgyl apysgfppgv lnpeaaklai
```

► NCBI/ BLAST/ blastp suite

The screenshot shows the NCBI BLASTp web interface. An orange arrow points to the 'Enter Query Sequence' section, specifically to the text input field containing the protein sequence. Another orange arrow points to the 'Choose Search Set' section, specifically to the 'Database' dropdown menu which is set to 'Non-redundant protein sequences (nr)'. A third orange arrow points to the 'Program Selection' section, specifically to the 'blastp (protein-protein BLAST)' radio button, which is selected. A fourth orange arrow points to the 'BLAST' button at the bottom left. The 'Job Title' field is empty. The 'Exclude' section has 'Models (XM/XP)' and 'Uncultured/environmental sample sequences' checked. The 'Entrez Query' field is empty. The 'Search database' is 'Non-redundant protein sequences (nr)' using 'Blastp (protein-protein BLAST)'. The 'Show results in a new window' checkbox is checked.

Enter Query Sequence

Enter accession number, gi, or FASTA sequence

Clear

Query subrange

From

To

Or, upload file

Browse...

Job Title

Enter a descriptive title for your BLAST search

☐ Align two or more sequences

Choose Search Set

Database

Non-redundant protein sequences (nr)

Organism

Optional

Exclude

Optional

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

Entrez Query

Optional

Enter an Entrez query to limit search

Program Selection

Algorithm

☒ blastp (protein-protein BLAST)

☐ PSI-BLAST (Position-Specific Iterated BLAST)

☐ PHI-BLAST (Pattern Hit Initiated BLAST)

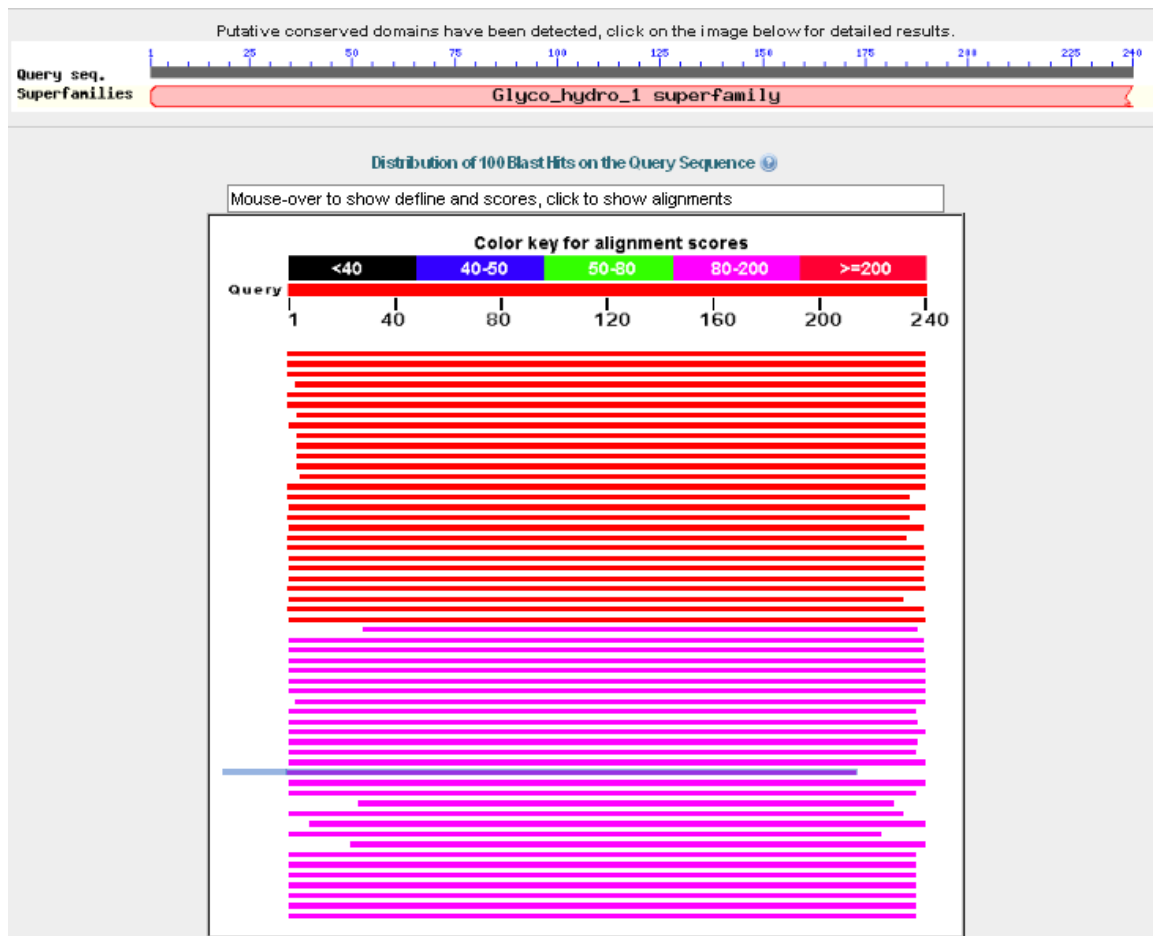
Choose a BLAST algorithm

BLAST

Search database Non-redundant protein sequences (nr) using Blastp (protein-protein BLAST)

☒ Show results in a new window

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



Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value
Q52629.1	RecName: Full=Beta-galactosidase; Short=Lactase >gb AAB97862.1 beta-galactosidase	494	494	100%	2e-138
NP_578937.1	beta-mannosidase [Pyrococcus furiosus DSM 3638] >gb AAC44387.1 beta-mannosidase	494	494	100%	4e-138
NP_126617.1	Beta-mannosidase [Pyrococcus abyssi GE5] >emb CAB49848.1 bgaL-1 beta-galactosidase	426	426	100%	1e-117
YP_002993739.1	Beta-galactosidase [Thermococcus sibiricus MM 739] >gb ACS89390.1 Beta-galactosidase	337	337	98%	1e-90
ABV49715.1	beta-galactosidase [uncultured bacterium]	328	328	100%	3e-88
YP_921073.1	glycoside hydrolase family protein [Thermophilum pendens Hrk 5] >gb ABL79070.1 glycosidase	315	315	100%	3e-84
YP_184174.1	beta-galactosidase [Thermococcus kodakarensis KOD1] >dbj BAD85950.1 beta-glycosidase	312	312	98%	3e-83
YP_003668816.1	glycoside hydrolase family 1 [Staphylothermus hellenicus DSM 12710] >gb ADI31917.1	308	308	99%	4e-82
BAA78713.1	beta-glycosidase [Thermococcus kodakaraensis]	306	306	98%	1e-81
NP_142473.1	beta-mannosidase [Pyrococcus horikoshii OT3] >dbj BAA29589.1 483aa long hypothetical protein	303	303	98%	1e-80
NP_127217.1	beta-galactosidase [Pyrococcus abyssi GE5] >emb CAB50447.1 bgaL-2 beta-galactosidase	299	299	98%	2e-79
NP_578085.1	beta-galactosidase [Pyrococcus furiosus DSM 3638] >gb AAL80480.1 beta-galactosidase	299	299	98%	2e-79
ZP_04878911.1	beta-galactosidase [Thermococcus sp. AM4] >gb EEB74957.1 beta-galactosidase [Thermococcus]	297	297	97%	7e-79

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

beta-galactosidase [Pyrococcus woesei]

GenBank: AAB97862.1

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

[Go to:](#) ☐

LOCUS AAB97862 510 aa linear BCT 25-MAY-2001
DEFINITION beta-galactosidase [Pyrococcus woesei].
ACCESSION AAB97862
VERSION AAB97862.1 GI:2811286
DBSOURCE accession [AF043283.1](#)
KEYWORDS .
SOURCE Pyrococcus woesei
ORGANISM [Pyrococcus woesei](#)
Archaea; Euryarchaeota; Thermococci; Thermococcales;
Thermococcaceae; Pyrococcus.
REFERENCE 1 (residues 1 to 510)
AUTHORS Daabrowski,S., Sobiewska,G., Maciunska,J., Synowiecki,J. and Kur,J.
TITLE Cloning, expression, and purification of the His(6)-tagged
thermostable beta-galactosidase from Pyrococcus woesei in
Escherichia coli and some properties of the isolated enzyme
JOURNAL Protein Expr. Purif. 19 (1), 107-112 (2000)
PUBMED [10833397](#)
REFERENCE 2 (residues 1 to 510)
AUTHORS Dabrowski,S., Maciunska,J. and Synowiecki,J.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-1998) Food Preservation, Technical University of
Gdansk, Narutowicza 11/12, Gdansk 80-952, Poland
COMMENT Method: conceptual translation.

FEATURES Location/Qualifiers
source 1..510
/organism="Pyrococcus woesei"
/strain="DSM3773"
/db_xref="taxon:[2262](#)"
[Protein](#) 1..510
/product="beta-galactosidase"
[Region](#) 1..504
/region_name="Glyco_hydro_1"
/note="Glycosyl hydrolase family 1; cl01046"
/db_xref="CDD:[174504](#)"
[CDS](#) 1..510
/coded_by="AF043283.1:1..1533"
/transl_table=[11](#)

ORIGIN
1 mfpekflwgv aqsgfqfemg dklrrnidtn tdwwhwvrdk tniekglvsg dlpeeginny
61 elyekdheia rklglnayri giewsrifpw pttfidvdys ynesynlied vkitkdtlee
121 ldeiankrev aayrsvinsl rskgfkviwn lnhftlpywl hdpiearera ltnkrngwvn
181 prtviiefaky aayiaykfgd ivdmwstfne pmvvelgyl apysgfppgv lnpeaaklai
241 lhminahala yrqikkfdte kadkdskepa evgiynnig vaypkdpnds kdvkaaendn
301 ffhsglffea ihkgklinief dgetfidapy lkgndwigvn yytrevvtyq epmfpsiqli
361 tfkgvqgygy acrpgtlskd drpvsdigwe lypegmydsi veahkygvpv yvtengiads
421 kdilrpyyia shikntekaf edgyevkgyf hwaltdnfew algfrmrfgl yevnlitker
481 ipreksvsif reivannngvt kkieeellrg

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

