

NCBI resources, search and retrieval functions

The screenshot shows a GenBank sample record for *Saccharomyces cerevisiae*. The record includes fields for LOCUS, DEFINITION, ACCESSION, VERSION, SOURCE, ORGANISM, REFERENCE, and FEATURES. The DEFINITION field states: "Saccharomyces cerevisiae TCT1-beta gene, partial cds, and Aali2 (AALI2) and Rev7p (REV7) genes, complete cds." The SOURCE field indicates the organism is *Saccharomyces cerevisiae* (Baker's yeast). The ORGANISM field lists the taxonomic lineage: Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces. The REFERENCE field lists three references, including a 2009 paper by Toppey et al. and a 1994 paper by Romer et al. The FEATURES field includes a CDS (Coding Sequence) and a GFF (Gene Feature File) section.

- Narrow your search using particular fields: [definition], [organism], [author], [subtree], [lineage], etc. (otherwise, it will search all fields)

and

- logical Boolean operators: AND, OR, XOR, and NOT

<https://www.ncbi.nlm.nih.gov/Sitemap/samplerecord.html>

3

TEHOMHKA: BLAST & alignment tools, 5 august 2017, Cpeca, #3

NCBI resources, search and retrieval functions

The screenshot shows the NCBI homepage. The top navigation bar includes links to NCBI Home, Resources, and How To. The left sidebar contains a list of resources, including Chemicals & Bioassays, Data & Software, CNA & RNA, Genomes & Structures, Genes & Expression, Genomics & Medicine, Genomes & Maps, Homology, Literature, Proteins, Sequence Analysis, Taxonomy, Training & Tutorials, and Variation. The main content area is titled "All Resources" and lists various tools and services, such as "Find nucleic acids in which a given drug is active", "Find nucleic acids that test a particular disease or protein target", "Submit data to NCBI", "Save text searches and set up automated searches with E-mailed results", "Download NCBI Software", "Retrieve all sequences for an organism or taxon", "Find the function of a gene or gene product", "Find expression patterns", "Find genes associated with a phenotype or disease", "Compare protein homologs between two microbial genomes", "Visualize features around an object or between two objects on a chromosome", "Find sequenced genomes, including those in progress, for a taxonomic group", "Download the complete genome for an organism", "Display genomic annotation graphically", "Submit sequence data to NCBI", "Convert feature coordinates between genomic assemblies", "Intersect consensus synteny between the genomes of two organisms", "Find a homolog for a gene in another organism", "Obtain the full text of an article", "Find articles about a topic similar to that in a given article", "View the 3D structure of a protein", "Find a curated version of a sequence record (NCBI Reference Sequence)", "Align two or more 3D structures to a given structure", "Find published information on a gene or sequence", "Find transcript sequences for a gene", "Link from an object on a map to another resource", "Design PCR primers and check them for specificity", "Automate BLAST searches performed on NCBI servers", "Obtain genomic sequence for a gene, marker, transcript or protein", "Compare your sequence to the RefSeqGene/RefSeq standard", "Run BLAST software on a local computer", "Submit multiple query sequences in a single BLAST search", "Find the complete taxonomic lineage for an organism", "Generate a Common Tree for a set of taxa", "Complete an NCBI tutorial", "Find out what's new at NCBI", "Learn about an NCBI resource", "Learn about the basics of molecular biology and bioinformatics", "Download a large, custom set of records from NCBI", "Find human variations associated with a phenotype or disease (clinical associations)", "View a mutation site in a 3D structure", "View all SNPs associated with a gene", and "View genotype frequency data for a gene, disease or short genetic variation".

4

TEHOMHKA: BLAST & alignment tools, 5 august 2017, Cpeca, #3

Introduction to BLAST: Basic Local Alignment Search Tool

- widely used search tool to find similar nucleotide or amino acid sequences developed in 1990 and 1997 (Stephen Altschul)
- heuristic approach for performing local alignments through searches of high scoring segment pairs (HSP's) and based on Smith-Waterman algorithm
- inferring function of a query sequence from its similarity with well-studied and annotated sequences
- uses statistics to predict significance of initial matches and to find best local alignments
- provides statistical significance for alignments
- accurate and fast
- a suit of tools (www, standalone, network clients, etc.)



ЛЕХОМНКА: BLAST & alignment tools, 5 апреля 2017, Москва, #3



BLAST Concepts for Sequence Similarity Searching

- One sequence by itself is not informative; it must be analyzed by comparative methods against existing sequence databases to develop hypothesis concerning structure and function.
- looks for clusters of nearby or locally dense “similar or homologous” k -tuples
- uses “look-up” tables to shorten search time
- uses larger “word size” than FASTA to accelerate the search process
- performs both Global and Local alignment
- fastest and most frequently used sequence alignment tool

10

ЛЕХОМНКА: BLAST & alignment tools, 5 апреля 2017, Москва, #3



How Does BLAST Really Work?

- The BLAST programs improved the overall speed of searches while retaining good sensitivity (important as databases continue to grow) by breaking the query and database sequences into fragments ("words"), and initially seeking matches between fragments.
- "Word" hits are then extended in either direction in an attempt to generate an alignment with a maximum score value "S".

11

ЛЕОМОНКА: BLAST & alignment tools, 5 апреля 2017, Москва, #3



Alignment

Query sequence: AAC**CG**TTC---T**ACA**ATT**AC**CTAGGC
 ===-==== =-====-=====

Best match sequence: AAC**GT**TT**C**AGT**CA**AA**T**AG**C**TAGGC

Hits (+1 per nucleotide): 1×18 (matching nucleotides) = 18

Penalty for mismatching (-2 per nucleotide): -2×5
 (mismatching nucleotides) = -10

Penalty for gaps opening (-2 per gap) \times # of gaps + penalty
 for extension (-1 per nucleotide): -2×1 (# of gaps) + -1×3
 (# of nucleotides in the gap)

Score = $18 \times 1 + 5 \times (-2) + 1 \times (-2) + 3 \times (-1) = 3$

12

ЛЕОМОНКА: BLAST & alignment tools, 5 апреля 2017, Москва, #3



Alignment

Global Alignment:

- compares total length of two sequences
 - [Needleman, S.B. and Wunsch, C.D.](#) A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol. 48(3):443-53(1970)

Local Alignment:

- compares segments of sequences
- finds cases when one sequence is a part of another sequence, or they only match in parts.
 - [Smith, T.F. and Waterman, M.S.](#) Identification of common molecular subsequences. J Mol Biol. 147(1):195-7 (1981)

13

ЛЕОМОНКА: BLAST & alignment tools, 5 апреля 2017, Москва, #3



The BLAST algorithm

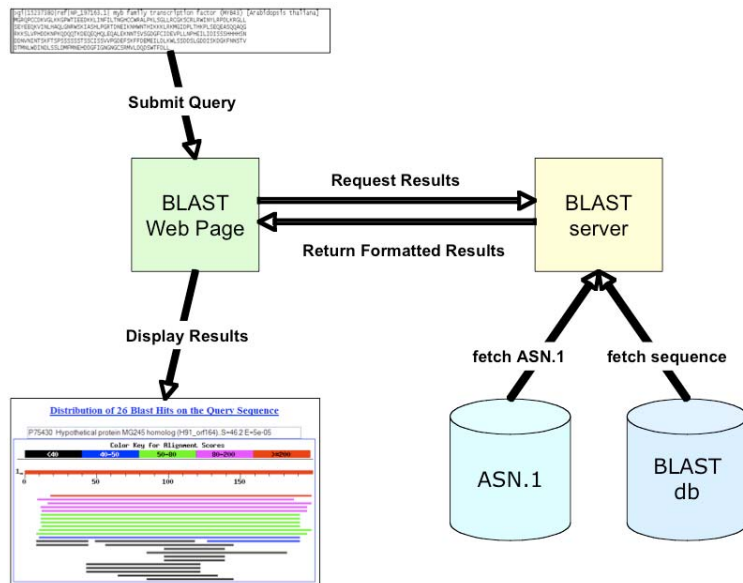
- The BLAST programs (Basic Local Alignment Search Tools) are a set of sequence comparison algorithms introduced in 1990 that are used to search sequence databases for optimal local alignments to a query.
 - Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) “Basic local alignment search tool.” J. Mol. Biol. 215:403-410.
 - Altschul SF, Madden TL, Schaeffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) “Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.” NAR 25:3389-3402.

14

ЛЕОМОНКА: BLAST & alignment tools, 5 апреля 2017, Москва, #3



The BLAST algorithm



15

LEHOMIKA: BLAST & alignment tools, 5 august 2017, Creta #3

BLAST Algorithm

- Scoring of matches done using scoring matrices
- Sequences are split into words (default n=3)
 - speed, computational efficiency
- BLAST algorithm extends the initial “seed” hit into an HSP
 - HSP = high scoring segment pair = Local optimal alignment

16

LEHOMIKA: BLAST & alignment tools, 5 august 2017, Creta #3

Sequence Similarity Searching – The statistics are important

- Discriminating between real and artifactual matches is done using an estimate of probability that the match might occur by chance (expectation values or E-values).
- We'll talk more about the meaning of the scores (S) and e-values (E) that are associated with BLAST hits

17

ЛЕОМОНКА: BLAST & alignment tools, 5 апреля 2017, Москва, #3



What do the Score and the E-value really mean?

- The quality of the alignment is represented by the **Score (S)**.
 - The **score** of an alignment is calculated as the **sum of substitution and gap scores**. Substitution scores for amino acid sequence alignment are given by a look-up substitution matrix (PAM, BLOSUM) whereas gap scores are assigned empirically .
- The significance of each alignment is computed as an **E-value (E)**.
 - **Expectation value**. The number of different alignments with scores equivalent to or better than S that are expected to occur in a database search by chance. The lower the E value, the more significant the score.

18

ЛЕОМОНКА: BLAST & alignment tools, 5 апреля 2017, Москва, #3



Notes on E-values

- Low E-values suggest that sequences are homologous
 - ◉ can't show non-homology
- Statistical significance depends on both the size of the alignments and the size of the sequence database
 - important consideration for comparing results across different searches
 - E-value increases as database gets bigger
 - E-value decreases as alignments get longer

19

ЛЕОМОНКА: BLAST & alignment tools, 5 апреля 2017, Москва, #3



Where does the score (S) come from?

- The quality of each pair-wise alignment is represented as a score (S) and the scores are ranked.
- **Scoring matrices** are used to calculate the score of the alignment base by base (DNA) or amino acid by amino acid (protein).
- The alignment score is the sum of the scores for all positions together.

20

ЛЕОМОНКА: BLAST & alignment tools, 5 апреля 2017, Москва, #3



What is a scoring matrix?

- Substitution matrices are used for amino acid alignments.

– each possible residue substitution is given a score

alanine Ala A
cysteine Cys C
aspartic acid Asp D
glutamic acid Glu E
phenylalanine Phe F
glycine Gly G
histidine His H

	A	C	D	E	F	G	H
A	4	0	-2	-1	-2	0	-2
C	0	9	-3	-4	-2	-3	-3
D	-2	-3	6	2	-3	-1	-1
E	-1	-4	2	5	-3	-2	0
F	-2	-2	-3	-3	6	-3	-1
G	0	-3	-1	-2	-3	4	-2
H	-2	-3	-1	0	-3	-2	4

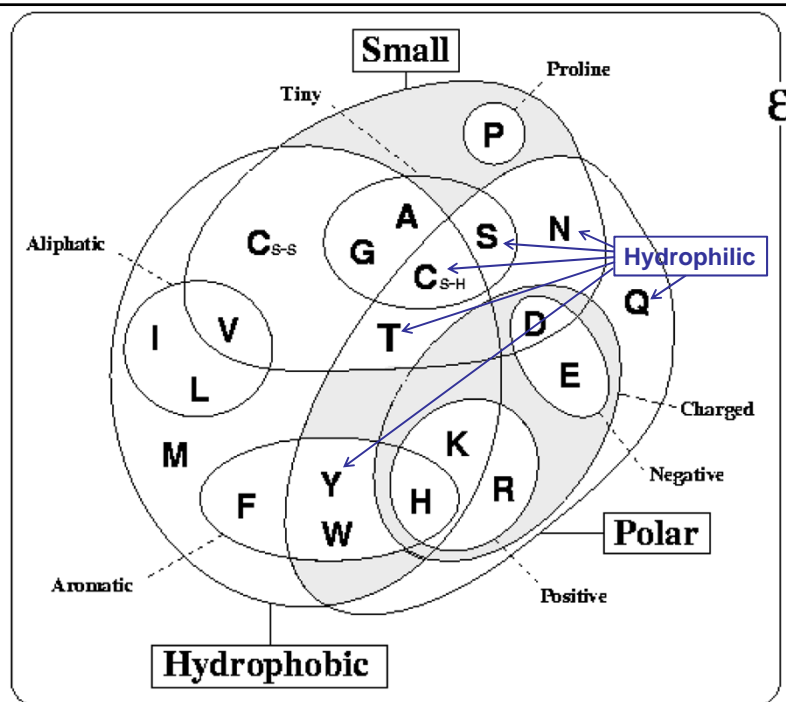
BLOSUM 62

BLOSUM - BLOcks SUBstitution Matrix
(Henikoff & Henikoff 1992)

- A simpler unitary matrix is used for nucleotide pairs (+1 for match, -2 mismatch)

21

CEHOMHKA: BLAST & alignment tools, 5 апреля 2017, Cisca, #3

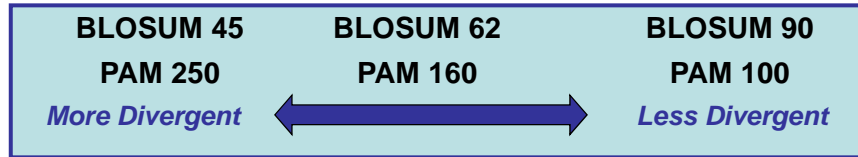


22

CEHOMHKA: BLAST & alignment tools, 5 апреля 2017, Cisca, #3



BLOSUM vs. PAM



- **BLOSUM 62 (BLOcks SUBstitution Matrix)** is the default matrix in BLAST 2.0. Though it is tailored for comparisons of moderately distant proteins, it performs well in detecting closer relationships. A search for distant relatives may be more sensitive with a different matrix, such as **PAM (Point Accepted Mutation)**.

23

LEHOMHKA: BLAST & alignment tools, 5 апреля 2017, Москва, #3



Suggested BLAST Cutoffs

- For nucleotide based searches, one should look for hits with E-values of 10^{-6} or less and sequence identity of 70% or more
- For protein based searches, one should look for hits with E-values of 10^{-3} or less and sequence identity of 25% or more

Take Home Message:
Always look at your alignments

Chapter 11 in "Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins" by Andreas D. Baxevanis and B. F. Francis Ouellette (Editors)

24

LEHOMHKA: BLAST & alignment tools, 5 апреля 2017, Москва, #3



Homology: Some Guidelines

- Similarity can be indicative of homology
- Generally, if two sequences are significantly similar over entire length they are likely homologous
- Low complexity regions can be highly similar without being homologous
- Homologous sequences not always highly similar

25

ЛЕОМОНКА: BLAST & alignment tools, 5 апреля 2017, Москва, #3



What BLAST tells you ...

- reports alignments (sometimes surprising)
 - similarities imply evolutionary homology, descent from a common ancestor, but does not always imply similar function
- provides statistical support that alignments are not by chance (E-values)
- provide level of identity and similarity

26

ЛЕОМОНКА: BLAST & alignment tools, 5 апреля 2017, Москва, #3



Phosphatase and tensin homolog (PTEN)

```

>|gb|AA108419.1| PTEN [Takifugu rubripes]
Length=412

Score = 197 bits (501), Expect = 2e-49, Method: Composition-based stats.
Identities = 95/100 (95%), Positives = 98/100 (98%), Gaps = 0/100 (0%)

Query 2 IVSRNKRRYQEDGFDDLTYIYPNIIAMGFPAERLEGVYRNNIDVVRFLDSKHKHNYKI 61
      +VSRNKRRYQEDGFDDLTYIYPNIIAMGFPAERLEGVYRNNIDVVRFLDSKHKHNYKI
Sbjct 8 MVS RNKRRYQEDGFDDLTYIYPNIIAMGFPAERLEGVYRNNIDVVRFLDSKHKHNYKI 67

Query 62 YNLCAERHYDTAKFNCRVAQYPPFEDHNPQLELIKPFQK 101
      YNLCAERHYD AKFNCRVAQYPPFEDHNPQLELIKPF ++
Sbjct 68 YNLCAERHYDAKFNCRVAQYPPFEDHNPQLELIKPFCE 107

```

Resulting alignment is called the HSP (high scoring segment pair = local optimal alignment) – more than one HSP per hit possible

```

Score = 83.6 bits (205), Expect = 4e-15, Method: Composition-based stats.
Identities = 60/103 (58%), Positives = 68/103 (66%), Gaps = 32/103 (31%)

Query 99 KQNKMLKKDKMFHFWNTFFIPGPPEV-----D 126
      KQNKMKKKDKMFHFWNTFFIPGPPEE +
Sbjct 260 KQNKMMKKDKMFHFWNTFFIPGPPEESRDKLENGAVNNADSQQGVPAPGGQGPQSAECRE 319

Query 127 NDKEYLVLTLTkndldkankdkanRYFSPNFKVLYFTKTVEE 169
      +D++YL+LTL+KND DKANKDKANRYFSPNFKVKL F+KTVEE
Sbjct 320 SDRDYLLTLTKNDRDKANKDKANRYFSPNFKVKLCFSKTVEE 362

```



```


>|gb|AAH93110.1| UC Ptenb protein [Danio rerio]
Length=289

Score = 197 bits (500), Expect = 2e-49, Method: Composition-based stats.
Identities = 95/99 (95%), Positives = 98/99 (98%), Gaps = 0/99 (0%)

Query 3 VSRNKRRYQEDGFDDLTYIYPNIIAMGFPAERLEGVYRNNIDVVRFLDSKHKHNYKIY 62
      VSRNKRRYQEDGFDDLTYIYPNIIAMGFPAERLEGVYRNNIDVVRFLDSKHK+HYKIY
Sbjct 9 VSRNKRRYQEDGFDDLTYIYPNIIAMGFPAERLEGVYRNNIDVVRFLDSKHKDHYKIY 68


Query 63 NLCAERHYDTAKFNCRVAQYPPFEDHNPQLELIKPFQK 101
      NLCAERHYDTAKFNCRVAQYPPFEDHNPQLELIKPF ++
Sbjct 69 NLCAERHYDTAKFNCRVAQYPPFEDHNPQLELIKPFCE 107

```



27 TEHOMIKA: BLAST & alignment tools, 5 august 2017, Cncaia #3

BLAST programs	
Program	Description
blastp	Compares an amino acid query sequence against a protein sequence database.
blastn	Compares a nucleotide query sequence against a nucleotide sequence database.
blastx	Compares a nucleotide query sequence translated in all reading frames against a protein sequence database. You could use this option to find potential translation products of an unknown nucleotide sequence.
tblastn	Compares a protein query sequence against a nucleotide sequence database dynamically translated in all reading frames.
tblastx	Compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database.



28 TEHOMIKA: BLAST & alignment tools, 5 august 2017, Cncaia #3

more BLAST programs

Program		Notes
Megablast	Contiguous	Nearly identical sequences
	Discontiguous	Cross-species comparison
Position Specific	PSI-BLAST	Automatically generates a position specific score matrix (PSSM)
	RPS-BLAST	Searches a database of PSI-BLAST PSSMs



nucleotide only



protein only

29

ЛЕОМНКА: BLAST & alignment tools, 5 апреля 2017, Москва, #3



Other BLAST tools and services

- **MEGABLAST** - for comparison of large sets of long DNA sequences
- **RPS-BLAST** - Conserved Domain Detection
- **BLAST 2 Sequences** - for performing pairwise alignments for 2 chosen sequences
- **Genomic BLAST** - for alignments against select human, microbial or malarial genomes
- **VecScreen** - for detecting cloning vector contamination in sequenced data

30

ЛЕОМНКА: BLAST & alignment tools, 5 апреля 2017, Москва, #3



BLAST Access

- **NCBI BLAST**

- <https://blast.ncbi.nlm.nih.gov/Blast.cgi>

- **Canadian Bioinformatics Resource: Similarity Searching and Classification**

- https://bioinformatics.ca/links_directory/category/sequence-comparison/similarity-searching-and-classification

- **European Bioinformatics Institute BLAST**

- <http://www.ebi.ac.uk/Tools/sss/ncbiblast/nucleotide.html>

31

LEHOMIKA: BLAST & alignment tools, 5 august 2017, Creta #3



BLAST at the NCBI website

32

LEHOMIKA: BLAST & alignment tools, 5 august 2017, Creta #3



To learn about a genome of interest, visit NCBI → Taxonomy → TaxBrowser → Genome

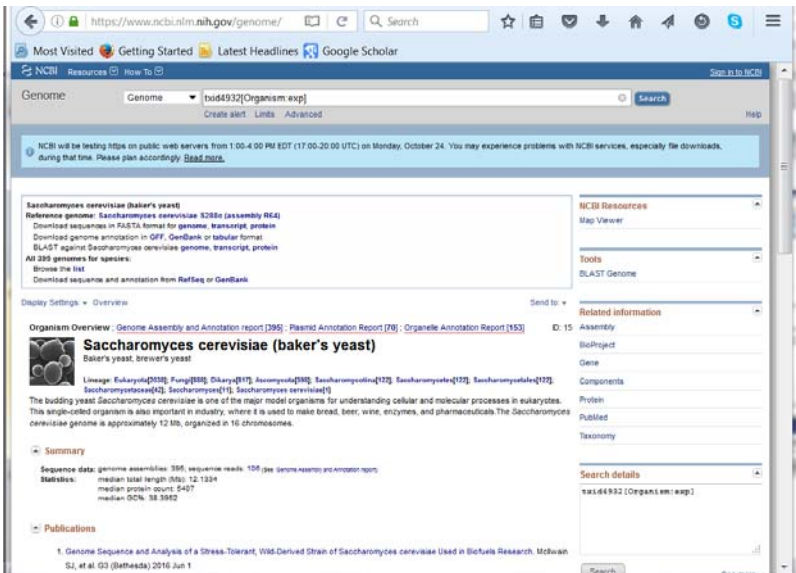
The screenshot shows the NCBI Taxonomy browser interface. The search bar contains 'Saccharomyces cerevisiae'. The 'Taxonomy' tab is selected. The page displays the following information:

- Search for:** Saccharomyces cerevisiae
- Display:** 0 levels using filter: none
- Saccharomyces cerevisiae**
 - Taxonomy ID: 4932
 - Genbank common name: baker's yeast
 - Inherited blast name: ascomycetes
 - Rank: species
 - Genetic code: Translation table 1 (Standard)
 - Mitochondrial genetic code: Translation table 2 (Chloroplast)
 - Other names:
 - Saccharomyces uvarum var. mellissae
 - Saccharomyces eublastus
 - Saccharomyces Italianus
 - Saccharomyces capensis
 - Common name: yeast
 - Common name: lager beer yeast
 - Common name: brewer's yeast
 - Assembly: Candida rebausa
 - Type material: NCRL V-12632
 - Type material: CBS 1171
 - Type material: ATCC 18224
- Entrez records**

Database name	Database	Direct link
Nucleotide	128,223	111,292
Nucleotide EST	3,451	38,912
Nucleotide GSS	2,838	2,818
Protein	104,451	40,184
Structure	3,390	2,481
Gene	1	1
Protein	320	320
Protein	6	3
GOI Diagrams	45,804	45,804
PubMed Central	77,119	77,119
Gene	7,000	6,000
Protein/Genome	3,390	3,100
RNA Experiments	47,805	41,843
Protein	104,451	104,451
Assembly	431	254
Pro Protein	2,471	2,110
Pro Sample	63,418	60,471
Bio Systems	1,484	1,725
GO Profiles	1,047,411	1,047,411
PubChem Bioassay	1,016	1,017
Taxonomy	318	1

A red arrow points from the 'Taxonomy' tab to the 'Entrez records' table.

To learn about a genome of interest, follow the TaxBrowser → Genome links



The screenshot shows the NCBI Taxonomy browser interface. The main content area displays information for *Saccharomyces cerevisiae* (baker's yeast). Key details include:

- Genome Overview:** Genome Assembly and Annotation report [396], Plasmid Annotation Report [70], Organellar Annotation Report [153].
- Sequence data:** genome assemblies: 396, sequence reads: 158, genome reports and annotation reports.
- Statistics:** median total length (Mb): 12.134, median protein count: 5497, median GCN: 58.3952.
- Publications:** 1. Genome Sequence and Analysis of a Stress-Tolerant, Wild-Derived Strain of *Saccharomyces cerevisiae* Used in Biotech Research. Mchwan Si, et al. (3) (Bethesda) 2016 Jun 1.

The right sidebar contains links to NCBI Resources, Map Viewer, Tools, BLAST Genome, Related information, Assembly, BioProject, Gene, Components, Protein, Published, Taxonomy, and Search details.

Saccharomyces cerevisiae (...

https://www.ncbi.nlm.nih.gov/genome/

Most Visited Getting Started Latest Headlines Google Scholar

Representative (genome information for reference and representative genomes)

Reference genome: [see all organisms]

Submitter: Saccharomyces Genome Database

Line	Type	Name	RefSeq	Size (kb)	GC%	Protein	rRNA	tRNA	Other tRNA	Gene	Pseudogene
Nuc	Chr	I	NC_001133.8	840,845.2	62.32	84	-	4	2	101	1
Nuc	Chr	II	NC_001134.8	840,845.2	62.32	408	-	13	4	425	-
Nuc	Chr	III	NC_001135.5	840,845.2	62.32	163	-	10	4	179	2
Nuc	Chr	IV	NC_001136.10	840,845.2	62.32	755	-	28	4	788	1
Nuc	Chr	V	NC_001137.3	840,845.2	62.32	279	-	20	9	309	1
Nuc	Chr	VI	NC_001138.8	840,845.2	62.32	128	-	10	4	140	1
Nuc	Chr	VII	NC_001139.9	840,845.2	62.32	550	-	30	10	570	-
Nuc	Chr	VIII	NC_001140.6	840,845.2	62.32	202	-	11	4	297	-
Nuc	Chr	IX	NC_001141.2	840,845.2	62.32	211	-	10	3	220	8
Nuc	Chr	X	NC_001142.9	840,845.2	62.32	359	-	24	0	389	-
Nuc	Chr	XI	NC_001143.9	840,845.2	62.32	313	-	16	5	304	-
Nuc	Chr	XII	NC_001144.6	840,845.2	62.32	509	12	21	18	862	2
Nuc	Chr	XIII	NC_001145.3	840,845.2	62.32	451	-	21	15	497	-
Nuc	Chr	XIV	NC_001146.0	840,845.2	62.32	396	-	14	6	410	-
Nuc	Chr	XV	NC_001147.0	840,845.2	62.32	550	-	20	11	509	2
Nuc	Chr	XVI	NC_001148.4	840,845.2	62.32	455	-	17	6	490	2
MT	Chr	MT	NC_001224.1	16,569.1	62.32	19	2	24	1	48	-

Chromosomes

Click on chromosome case to open MapViewer

External Resources

35

TEHOMIKA: BLAST & alignment tools, 5 august 2017, Cocos #3

To explore human chromosome 21 at NCBI

→ Find MapViewer

→ Choose Primates

→ Click [Annotation Release 107](#) and, then, chromosome 21

NCBI National Center for Biotechnology Information

Search [All Databases] Search Clear

Resources

NCBI Home

All Resources (A-Z)

Chemicals & Bioassays

Resources How To

3 A B C D E F G H I J L M N O P R S T U V

Taxonomy browser (Homo... Map Viewer

https://www.ncbi.nlm.nih.gov/mapview

Most Visited Getting Started Latest Headlines Google Scholar

NCBI Home GenBank BLAST

Map Viewer Home

The Map Viewer provides a wide variety of genome mapping and sequencing data. More

Search Select Group or Organism for Go

Tools Legend

Search or Browse the Genome BLAST

Y: Vertebrates (160)

M: Mammals (160)

Primates (15)

Scientific name	Common name	Build	Tools
<i>Callithrix jacchus</i>	white-tufted-ear marmoset	Annotation Release 102	Map Viewer
<i>Chlorocebus sabaeus</i>	green monkey	Annotation Release 102	Map Viewer
<i>Quilla quilla</i>	western quilla	Annotation Release 102	Map Viewer
<i>Homo sapiens</i>	human	Annotation Release 102	Map Viewer
<i>Macaca fascicularis</i>	crab-eating macaque	Annotation Release 102	Map Viewer
<i>Macaca mulatta</i>	Japanese macaque	Annotation Release 102	Map Viewer

36

TEHOMIKA: BLAST & alignment tools, 5 august 2017, Cocos #3

37

TFHOMHKA: BLAST & alignment tools, 5 august 2017, Cncaia, #3

References and Credits

- Materials for this presentation have been adapted from the following sources:
 - NCBI HelpDesk - Field Guide Course Materials
 - Bioinformatics: A practical guide to the analysis of genes and proteins
- Strongly recommend BLAST tutorial on NCBI site
 - <http://www.ncbi.nlm.nih.gov/BLAST/tutorial/Altschul-1.html>
- Further “Bioinformatics for quantitative geneticists course notes” J. McEwan
 - http://www-personal.une.edu.au/~jvanderw/aabc_materials2004.htm#ModuleC