



# Геномные базы данных, часть I

*к.б.н.  
Левинский В.Г.*

Кафедра эволюционной биологии ФЕННГУ



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

The screenshot shows the NCBI homepage in a Microsoft Internet Explorer browser window. The address bar displays <http://www.ncbi.nlm.nih.gov/>. The page features the NCBI logo and the text "National Center for Biotechnology Information" and "National Library of Medicine National Institutes of Health". A navigation menu includes links for PubMed, Entrez, BLAST, OMM, Books, TaxBrowser, and Structure. A search bar is present with "Entrez" selected in the dropdown and a "Go" button. The main content area is divided into several sections: "What does NCBI do?" (Established in 1988 as a national resource for molecular biology information, NCB1 creates public databases, conducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information - all for the better understanding of molecular processes affecting human health and disease. [More...](#)), "Hot Spots" (Clusters of orthologous groups, Electronic PCR, Gene expression omnibus, Genes and disease, Human genome resources, Human/mouse homology maps, LocusLink, Malaria genetics & genomics, Map Viewer, MHC, Mouse genome resources, NCB1 Handbook), "PubMed Central" (An archive of life sciences journals, Free fulltext, Over 100,000 articles from over 130 journals, Linked to PubMed and fully searchable, Use of PubMed Central requires no registration or fee. Access it from any computer with an Internet connection.), and "dbMHC" (A new NCB1 resource that provides a platform for genetic and clinical data related to the human Major Histocompatibility Complex (MHC) where users can submit, edit, view, exchange and analyse MHC data.). A left sidebar contains a "SITE MAP" with links to "Guide to NCB1 resources", "About NCB1" (An introduction for researchers, educators, and the public), "GenBank" (Sequence submission support and software), "Literature databases" (PubMed, OMM, Books and PubMed Central), "Molecular databases" (Sequences, structures, and taxonomy), and "Genomic biology" (The human genome whole).



# Multiple Database entry

<http://www.ncbi.nlm.nih.gov/gquery/gquery.fcgi>

The screenshot shows the Entrez cross-database search page in Microsoft Internet Explorer. The browser's address bar displays <http://www.ncbi.nlm.nih.gov/gquery/gquery.fcgi>. The page features the NCBI logo and the Entrez logo with the tagline "Entrez, The Life Sciences Search Engine". A navigation bar includes links for "PubMed", "Entrez", "Human Genome", "GenBank", "Map Viewer", and "BLAST". Below this is a search bar with the text "Search across databases" and buttons for "GO" and "GO" (with a question mark icon), along with a "Help" link. A welcome message reads: "Welcome to the new Entrez cross-database search page". The page lists various databases in two columns, each with an icon and a question mark icon for more information:

- PubMed: biomedical literature citations and abstracts
- PubMed Central: free, full-text journal articles
- Journals: detailed information about journals in Entrez
- MeSH: detailed information about NLM's controlled vocabulary
- Books: online books
- OMIM: online Mendelian Inheritance in Man
- Site Search: NCBI web and FTP sites
- nucleotide sequence database (GenBank)
- Protein: sequence database
- Genomes: whole genome sequences
- Structure: three-dimensional macromolecular structures
- Taxonomy: organisms in GenBank
- SNP: single nucleotide polymorphism
- UniGene: gene-oriented clusters of transcript sequences
- COG: conserved protein domain database
- 3D Domains: domains from Entrez Structure
- UniSTS: markers and mapping data
- PopSet: population study data sets
- GEO: expression and molecular abundance profiles
- GEO Datasets: experimental sets of GEO data

At the bottom, instructions are provided: "Enter terms and click 'GO' to run the search against ALL the databases, OR Click Database name or icon to go directly to the Search Page for that database, OR Click Question Mark for a short explanation of that database."



# PubMedCentral FullText paper, NCBI

<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pmc>

The screenshot shows the PubMed Central homepage. At the top, there is a navigation bar with links for PubMed, Medline, Pubmed, Genes, Structure, PMC, Taxonomy, OMIM, and Books. Below this is a search bar with the text 'Search PMC for' and buttons for 'Go' and 'Clear'. To the right of the search bar are two checked options: 'Free full-text only' and 'SmartSearch'. Below the search bar are links for 'Limits', 'PreviewIndex', 'History', 'Clipboard', and 'Details'. On the left side, there is a blue sidebar with links for 'About Entrez', 'PubMed Central', 'About PMC', 'Help / FAQ', 'Journal List', 'Citation Search', 'Related Resources', 'PubMed', 'Cubby', 'Consumer Health', 'Clinical Alerts', 'ClinicalTrials.gov', and 'Privacy Policy'. The main content area contains a list of search tips:

- Enter one or more search terms, or click [PreviewIndex](#) for advanced searching.
- Enter [author names](#) as *smith jr.* Initials are optional.
- Enter [journal titles](#) in full or as MEDLINE abbreviations. Use the [Journal Database](#) to find journal titles.
- **SmartSearch** analyzes the complete text of an article and, when searching, ignores 'noise' words that are of little significance in the content of the article. To search against every word in the text, uncheck the SmartSearch box above.

Below the list is a yellow highlighted box containing the text: "PubMed Central (PMC) is the U.S. National Library of Medicine's digital archive of life sciences journal literature. Access to the full text of articles in PMC is free, except where a journal requires a subscription for access to recent articles."

At the bottom of the page, there are links for "Write to the Help Desk", "NCBI | NLM | NIH", "Department of Health & Human Services", and "Freedom of Information Act | Disclaimer".





# Genome – whole genomes, NCBI



Entrez - Genome - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Genome

NCBI

Genome

Search Genome for Go Clear

Links Preview/Index History Clipboard Details

About Entrez

Entrez Genomes Help

Submitting Genome Project Genome sequences

Microbial Genomes Complete Genomes List of projects PDB neighbors

Genomic BLAST Microbial Eukaryotic

Archaea Genome Plasmids Unfinished

Bacteria Genome Plasmids Unfinished

The whole genomes of over 1000 viruses and over 100 microbes can be found in Entrez Genome. The genomes represent both completely sequenced organisms and those for which sequencing is in progress. All three main domains of life - **bacteria**, **archaea**, and **eukaryota** - are represented, as well as many **viruses** and **organelles**.

**New releases**  
[Caenorhabditis elegans](#)  
release WGS7 of the assembled and annotated genome sequence

**Related resources**  
**Microbial**  
reference sequences and resources  
**Organisms**  
reference sequences and tools  
**Viruses**  
reference sequences and tools  
**SARS Coronavirus Resource**  
sequence data and analyses  
**Plant Genomes Center**  
major plant genome projects  
**Eukaryotic Genome Projects**  
**Tools and Analysis**  
[Map Viewer](#)  
genome browser for eukaryotic genomes  
[TaxPlot](#)

**Recently completed genome: *Photorhabdus luminescens* subsp. *lummondii* T101**  
Publication Date October 2003  
Reference Duchaud E, et al. Complete genome sequence of the entomopathogenic bacterium *Photorhabdus luminescens*. Nat Biotechnol 11 (1) (2003).  
Size 5,698,987 bp  
The complete genome sequence of *Photorhabdus luminescens* subsp. *lummondii* T101 is deposited under GenBank Accession Number [BX470251](#).

*Photorhabdus luminescens* is a bioluminescent, gram-negative bacterium, and an insect pathogen with an interesting lifestyle. Its natural vector is a nematode in whose gut it establishes a symbiotic relationship. The nematode causes small lesions to insect larvae which are then rapidly killed by the bacteria.

**Recently completed genome: *Glossobacter***

eukaryota

eukaryotic genome project

MapView (геномная карта)

major sequencing centers (ниже на странице)



# Eukaryota, links at NCBI

[http://www.ncbi.nlm.nih.gov/genomes/static/euk\\_g.html](http://www.ncbi.nlm.nih.gov/genomes/static/euk_g.html)

The screenshot shows the NCBI Genomes website in a Microsoft Internet Explorer browser. The page title is "Eukaryote Genomes / List". The main content area displays a list of genomes under the heading "Complete genome". Each entry includes a species name and its corresponding chromosomes. An arrow points from the Russian text "ВИДЫ И НОМЕРА ХРОМОСОМ" to the list of genomes.

**ВИДЫ И НОМЕРА ХРОМОСОМ**

- [7] *Drosophila dentissima*  
chromosomes X, 2, 3
- [7] *Drosophila melanogaster*  
chromosomes I, II, III, IV, V
- [4] *Gemmatimonas sanguis*  
chromosomes I, II, III, IV, V, X
- [7] *Drosophila melanogaster*  
chromosomes 1, 2, 3, 4, X
- [11] *Brachyditoma cerasi genome*  
chromosomes I, II, III, IV, V, VI, VII, VIII, IX, X, XI
- [7] *Chlamydomonas reinhardtii genome*  
chromosomes I, 2, 3
- [14] *Saccharomyces cerevisiae*  
chromosomes I, II, III, IV, V, VI, VII, VIII, IX, X, XI, XII, XIII, XIV, XV, XVI
- [14] *Paramecium tetraureum*  
chromosomes 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14
- [1] *Schistosoma mansoni genome*  
chromosomes I, II, III

**Maps -> See genomes in Map Viewer**

- [14] *Homo sapiens*  
chromosomes 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, X, Y
- [11] *Mus musculus*  
chromosomes 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, X, Y



# Eukaryotic genome project, links at NCBI

[http://www.ncbi.nlm.nih.gov/genomes/static/EG\\_T.html](http://www.ncbi.nlm.nih.gov/genomes/static/EG_T.html)

The screenshot shows a Microsoft Internet Explorer browser window displaying the NCBI website. The address bar shows the URL: [http://www.ncbi.nlm.nih.gov/genomes/static/EG\\_T.html](http://www.ncbi.nlm.nih.gov/genomes/static/EG_T.html). The page title is "Eukaryotic Genome Projects in Progress".

The page content includes a search bar at the top with a "Go" button and a "Clear" button. Below the search bar, there are navigation tabs: "Limits", "Preview/Index", "History", and "Clipboard".

The main heading is "Eukaryotic Genome Projects in Progress". Below this heading, there is a paragraph of text:

NCBI is working in close collaboration with the various sequencing centers to provide new and up-to-date relevant information about the ongoing genome sequencing programs for organisms of all three domains of life. Entrez Genome is in the process of updating its pages and will display in a short future a complete list of all eukaryotic sequencing efforts for which genomic data is publicly available. Please note the color-coded organism title, which reflects the sequencing effort in progress: **Large-scale genome sequencing**; **Small-scale genome sequencing**; **EST, GSS, or cDNA sequencing**.

The page is organized into sections based on taxonomic groups:

- Metazoa [26]:**
  - Arthropoda [9]:**
    - Drosophila melanogaster*
    - Anopheles gambiae*
    - Culex pipiens*
  - Chordata [12]:**
    - Caenorhabditis briggsae*
    - Caenorhabditis elegans*
    - Danio rerio*
    - Gallus gallus*
    - Homo sapiens*
    - Macaca mulatta*
    - Monodelphis domestica*
    - Rattus norvegicus*
    - Sus scrofa*
    - Xenopus laevis*
    - Xenopus tropicalis*
  - Nematoda [3]:**
    - Brugia malayi*
  - Trematoda [2]:**
    - Schistosoma japonicum*
    - Schistosoma mansoni*
- Fungi [14]:**
  - Ascomycota [11]:**
    - Aspergillus fumigatus*
    - Aspergillus nidulans*
    - Aspergillus parasiticus*





# Major sequencing centers, at NCBI

<http://www.ncbi.nlm.nih.gov/genomes/static/links.html>

**GENOME CENTERS AND DATABASES**

BLAST    PubMed    Entrez    Human Genome    Microbial Genomes    Taxonomy

**Entrez Genomes**

List by Organism

- ACGT
- Baylor
- BCUJ
- CBC
- CEPH
- CHC
- DCBJ
- DOE-JGI
- EMBL
- FlyBase
- FruitFly
- Genethon
- Genoscope
- GGL
- GSC
- Jackson
- JMSTC
- Kazusa
- KEGG
- L. Berkeley
- Los Alamos NL
- L. Urmere NL
- MazdaDB
- MosquitDB
- NHGRI
- NISC
- NITE
- Nimble

**GENOME CENTERS**    **DATABASES**

**ACGT (Univ. Oklahoma)**

<i>Anabaena</i>	<i>Arabidopsis thaliana</i>	<i>Aspergillus nidulans</i>	<i>Brassica napus</i>	<i>Clostridium botulinum</i>	<i>Drosophila melanogaster</i>	<i>Escherichia coli</i>	<i>Haemophilus influenzae</i>	<i>Homo sapiens</i>	<i>Mycobacterium tuberculosis</i>	<i>Neisseria meningitidis</i>	<i>Rhizobium loti</i>	<i>Saccharomyces cerevisiae</i>	<i>Staphylococcus aureus</i>	<i>Streptococcus pneumoniae</i>	<i>Yersinia enterocolitica</i>
-----------------	-----------------------------	-----------------------------	-----------------------	------------------------------	--------------------------------	-------------------------	-------------------------------	---------------------	-----------------------------------	-------------------------------	-----------------------	---------------------------------	------------------------------	---------------------------------	--------------------------------

**Baylor College of Medicine**

<i>Brucella abortus</i>	<i>Brucella melitensis</i>	<i>Brucella abortus</i>	<i>Brucella abortus</i>	<i>Brucella abortus</i>
-------------------------	----------------------------	-------------------------	-------------------------	-------------------------

**Biotechnology Center, Univ. Illinois (BCUI)**

<i>Brucella abortus</i>	<i>Brucella abortus</i>	<i>Brucella abortus</i>	<i>Brucella abortus</i>	<i>Brucella abortus</i>
-------------------------	-------------------------	-------------------------	-------------------------	-------------------------

**CBC**

<i>Brucella abortus</i>	<i>Brucella abortus</i>	<i>Brucella abortus</i>	<i>Brucella abortus</i>	<i>Brucella abortus</i>
-------------------------	-------------------------	-------------------------	-------------------------	-------------------------

**CEPH-Genethon, France**    [Top](#)

Human Genome





База данных LocusLink обеспечивает интегрированный доступ к различным локусам (районам) геномов эукариот

The screenshot shows the LocusLink website interface. At the top, there's a navigation bar with tabs for PubMed, Entrez, BLAST, OMM, Map Viewer, Taxonomy, and Structure. Below this is a search bar with a dropdown menu set to 'LocusLink', a 'Display' dropdown set to 'Brief', and an 'Organism' dropdown set to 'All'. A 'Query:' field is present with 'Go' and 'Clear' buttons. Below the search bar is a horizontal navigation bar with letters A-Z. The main content area has sections for 'Introduction', 'Sequence accessions', 'Data can be accessed by clicking one of the letters above to browse loci sorted by symbols, or by entering a query into the search form. Use of wild cards (\*) is supported. Additional information and query tips are provided in the [Help](#) documentation.', and 'New features' dated October 22, 2003.

поле быстрого поиска



<http://www.ncbi.nlm.nih.gov/LocusLink/list.cgi>

## Результаты поиска по названию гена *Hsp27*

NCBI LocusLink

Search: LocusLink | Display: Brief | Organism: All

Query: hsp27

18 loci found

LocusID	Org	Symbol	Description	Position	Links
25389	Rs	Atf3	activating transcription factor 3	6q27	[Icons]
1410	Hs	CRYAB	crystallin, alpha B	11q22.3-q23.1	[Icons]
13614	Mm	Edn1	endothelin 1	13 26.0 cM	[Icons]
39078	Dm	Hsp27	Heat shock protein 27	2-81	[Icons]
24471	Rs	Hspb1	heat shock 27kDa protein 1	12q12	[Icons]
3315	Hs	HSPB1	heat shock 27kDa protein 1	7q11.23	[Icons]
3316	Hs	HSPB2	heat shock 27kDa protein 2	11q22-q23	[Icons]
69253	Mm	Hspb2	heat shock protein 2	9 29.0 cM	[Icons]
11009	Hs	IL24	interleukin 24	1q32	[Icons]
114294	Hs	LACTB	lactamase, beta	15q22.1	[Icons]
81649	Rs	Mapk14	mitogen activated protein kinase 14	20p12	[Icons]

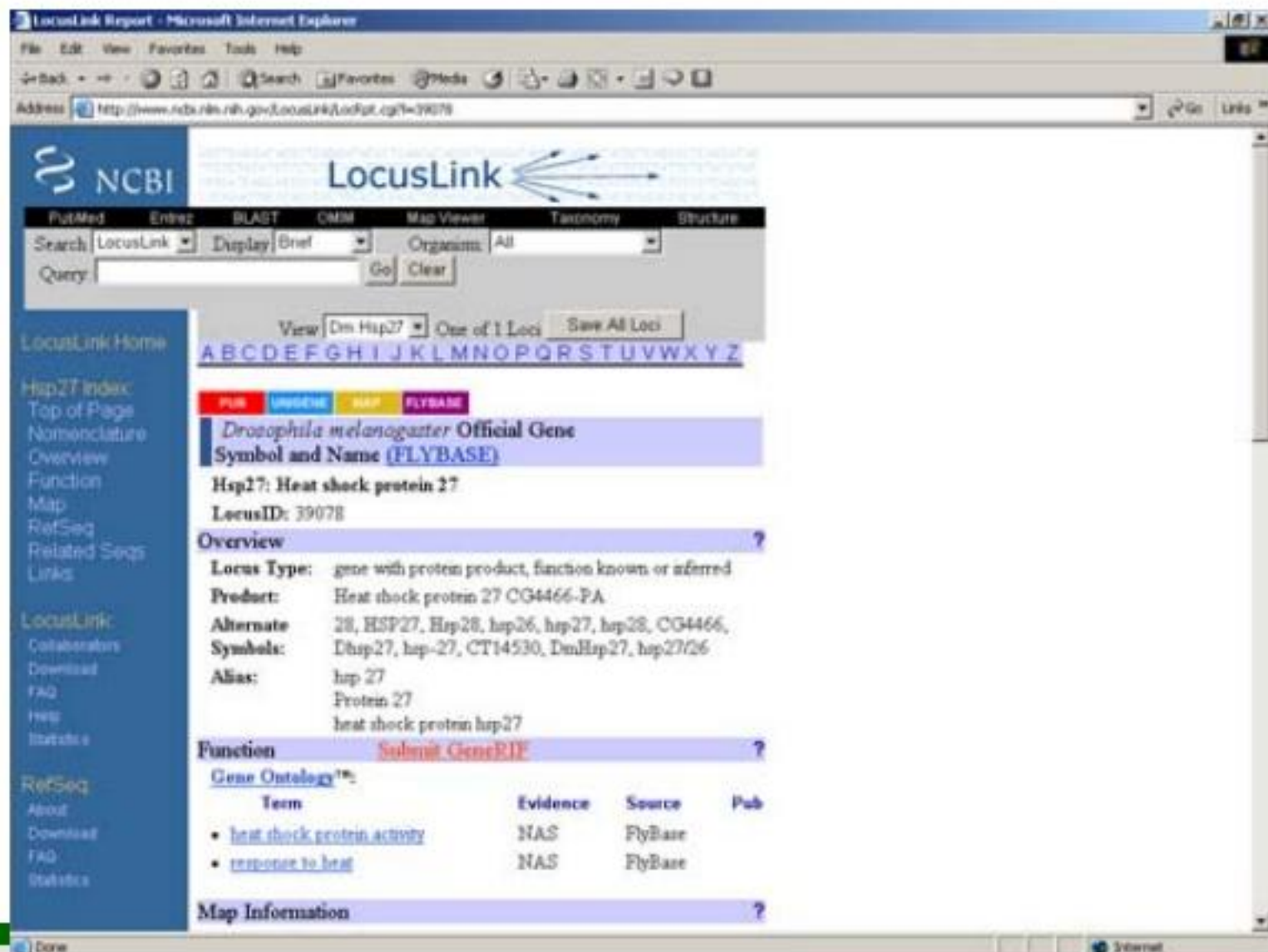
ВИД

название  
локуса

описание

хромосомная  
локализация

## Описание в базе гена *Hsp 27 Drosophila melanogaster*



The screenshot shows the NCBI LocusLink interface for the gene *Hsp27* in *Drosophila melanogaster*. The page includes a search bar, navigation tabs (PubMed, Entrez, BLAST, OMIM, Map Viewer, Taxonomy, Structure), and a sidebar with links to various sections like Overview, Function, and Map Information. The main content area displays the gene's symbol and name, locus ID, and a table of Gene Ontology terms.

NCBI LocusLink

Search: LocusLink | Display: Brief | Organism: All

View: Dm Hsp27 | One of 1 Loci | Save All Loci

ABCDEFGHIJKLMNOPQRSTUVWXYZ

PubMed | UniProt | RefSeq | FLYBASE

***Drosophila melanogaster* Official Gene**  
**Symbol and Name (FLYBASE)**

**Hsp27: Heat shock protein 27**  
LocusID: 39078

**Overview** ?

**Locus Type:** gene with protein product, function known or inferred  
**Product:** Heat shock protein 27 CG4466-PA  
**Alternate Symbols:** 28, HSP27, Hsp28, hsp26, hsp27, hsp28, CG4466, Dhsp27, hsp-27, CT14530, DmHsp27, hsp27/26  
**Alias:** hsp 27, Protein 27, heat shock protein hsp27

**Function** [Submit GeneRIF](#) ?

**Gene Ontology™**

Term	Evidence	Source	Pub
<a href="#">heat shock protein activity</a>	NAS	FlyBase	
<a href="#">response to heat</a>	NAS	FlyBase	

**Map Information** ?





# Геномные карты (MapView), NCBI <http://www.ncbi.nlm.nih.gov/mapview>

The screenshot shows the NCBI Map Viewer interface. At the top, there is a search bar with the text "Search [Select Organism] for [ ] [Go]". Below the search bar, there is a green box with the following text: "New! - searching for map objects in any user-determined subset of all plant genomes presented by NCBI. A map object includes, but is not limited to, a locus, probe name, GenBank accession, gene or name of BAC clone. Select "All plants" from the search menu at the top of the page or click (S) to the right of the Plants node." Below this, there is a taxonomic tree with several nodes. The nodes are: "Other Vertebrates" (with a sub-node "Danio rerio (zebrafish)"), "Invertebrates" (with sub-nodes "Amphioxus japonicus (amphioxus)", "Caenorhabditis elegans (nematode)", and "Drosophila melanogaster (fruit fly)"), "Fungi" (with sub-nodes "Saccharomyces cerevisiae (baker's yeast)", "Schizosaccharomyces pombe (fission yeast)", and "Neurospora crassa"), "Plants" (with sub-nodes "Arabidopsis thaliana (thale cress)", "Rosa canina (rose)", "Medicago sativa (sage)", "Oryza sativa (rice)", "Festuca arvensis (wheat)", "Suaeda (suaeda)", "Lycopersicon esculentum (tomato)", and "Glycine max (soybean)"), and "Protocera" (with a sub-node "Plasmodium falciparum"). There are also links for "Bacteria", "Organisms", and "Viruses".

беспозвоночные  
млекопитающие  
растения  
грибы



# Интерфейс программы BLAT, <http://genome.ucsc.edu/> (BLAT - аналог программы парного выравнивания BLAST)

Fruitfly BLAT Search - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address <http://genome.ucsc.edu/cgi-bin/blast>

Home Genome Browser Blast Table Browser FAQ Help

### Fruitfly BLAT Search

## BLAT Search Genome

Genome:  Assembly:  Query type:  Sort output:  Output type:

Please paste in a query sequence to see where it is located in the genome. Multiple sequences can be searched at once if separated by a line starting with > and the sequence name.

Rather than pasting a sequence, you can choose to upload a text file containing the sequence.

Upload sequence:

Only DNA sequences of 25,000 or less bases and protein or translated sequence of 5000 or less letters will be processed. If multiple sequences are submitted at the same time, the total limit is 50,000 bases or 12,500 letters.

BLAT on DNA is designed to quickly find sequences of 95% and greater similarity of length 40 bases or more. It may miss more divergent or shorter

поле ввода  
анализируемой  
последователь-  
ности ДНК



# Результаты запроса в программу BLAT (запрос – ген *Drosophila melanogaster* *Hsp27* с флангами)

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CSDO	STRAND	START	END
<a href="#">Browser</a> <a href="#">Details</a>	YourSeq	3619	1	3619	3619	100.0%	3L	+	9341575	9345193
<a href="#">Browser</a> <a href="#">Details</a>	YourSeq	30	369	399	3619	100.0%	X	+	2055469	2055504
<a href="#">Browser</a> <a href="#">Details</a>	YourSeq	29	295	497	3619	48.4%	X	-	5656298	5656350
<a href="#">Browser</a> <a href="#">Details</a>	YourSeq	23	773	795	3619	100.0%	3L	+	7609558	7609580
<a href="#">Browser</a> <a href="#">Details</a>	YourSeq	22	372	393	3619	100.0%	3R	-	18344217	18344238
<a href="#">Browser</a> <a href="#">Details</a>	YourSeq	20	776	795	3619	100.0%	3L	-	9342350	9342369
<a href="#">Browser</a> <a href="#">Details</a>	YourSeq	19	777	795	3619	100.0%	3L	-	14708755	14708753
<a href="#">Browser</a> <a href="#">Details</a>	YourSeq	19	777	795	3619	100.0%	2R	-	9506395	9506413
<a href="#">Browser</a> <a href="#">Details</a>	YourSeq	19	777	795	3619	100.0%	X	+	5000000	5000000

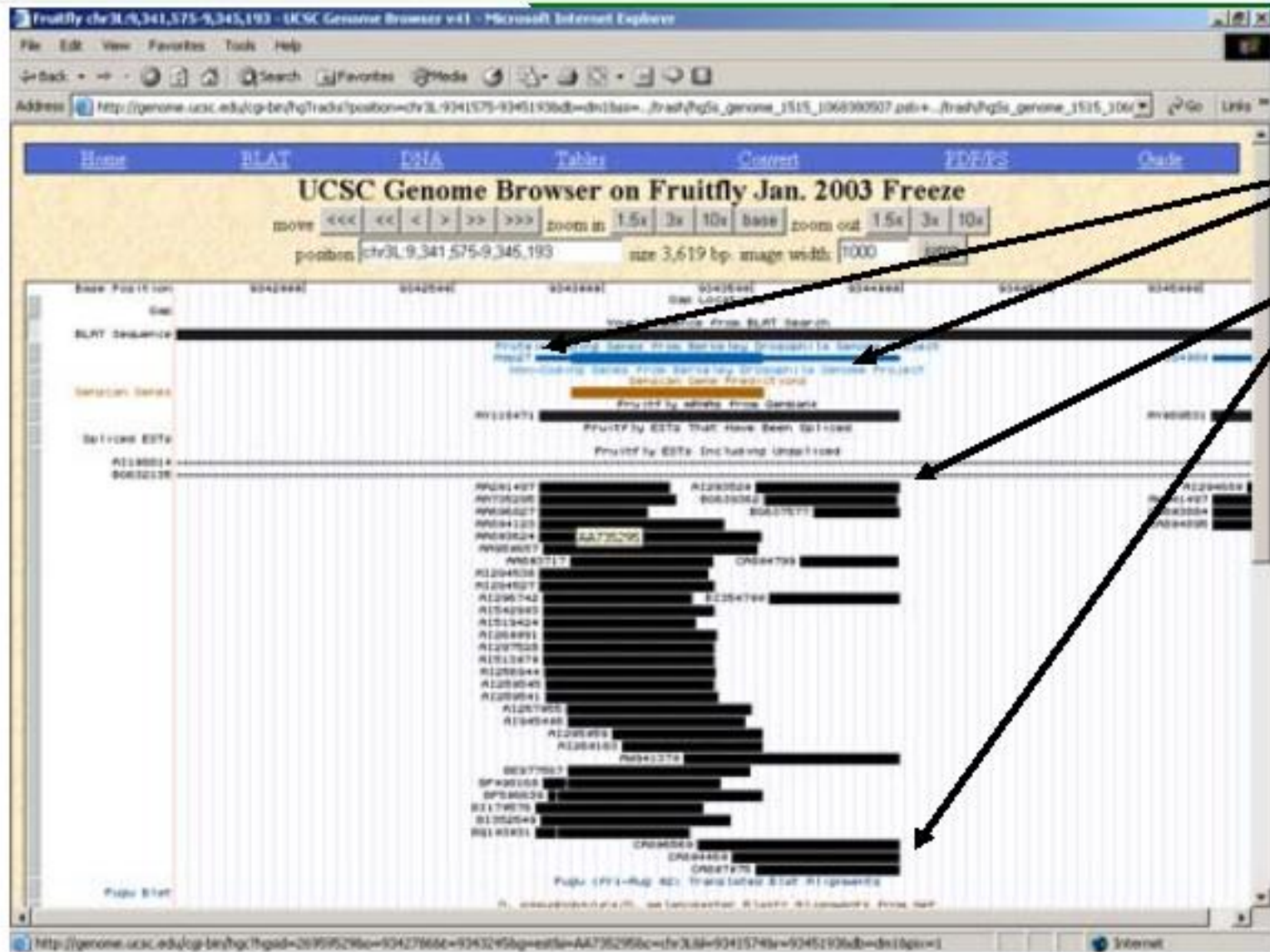
высокое и  
обширное  
сходство  
с запросом

низкое и  
ограниченное  
сходство  
с запросом





# Результаты запроса в программу BLAT – карта положения гена *HSP27* на хромосоме

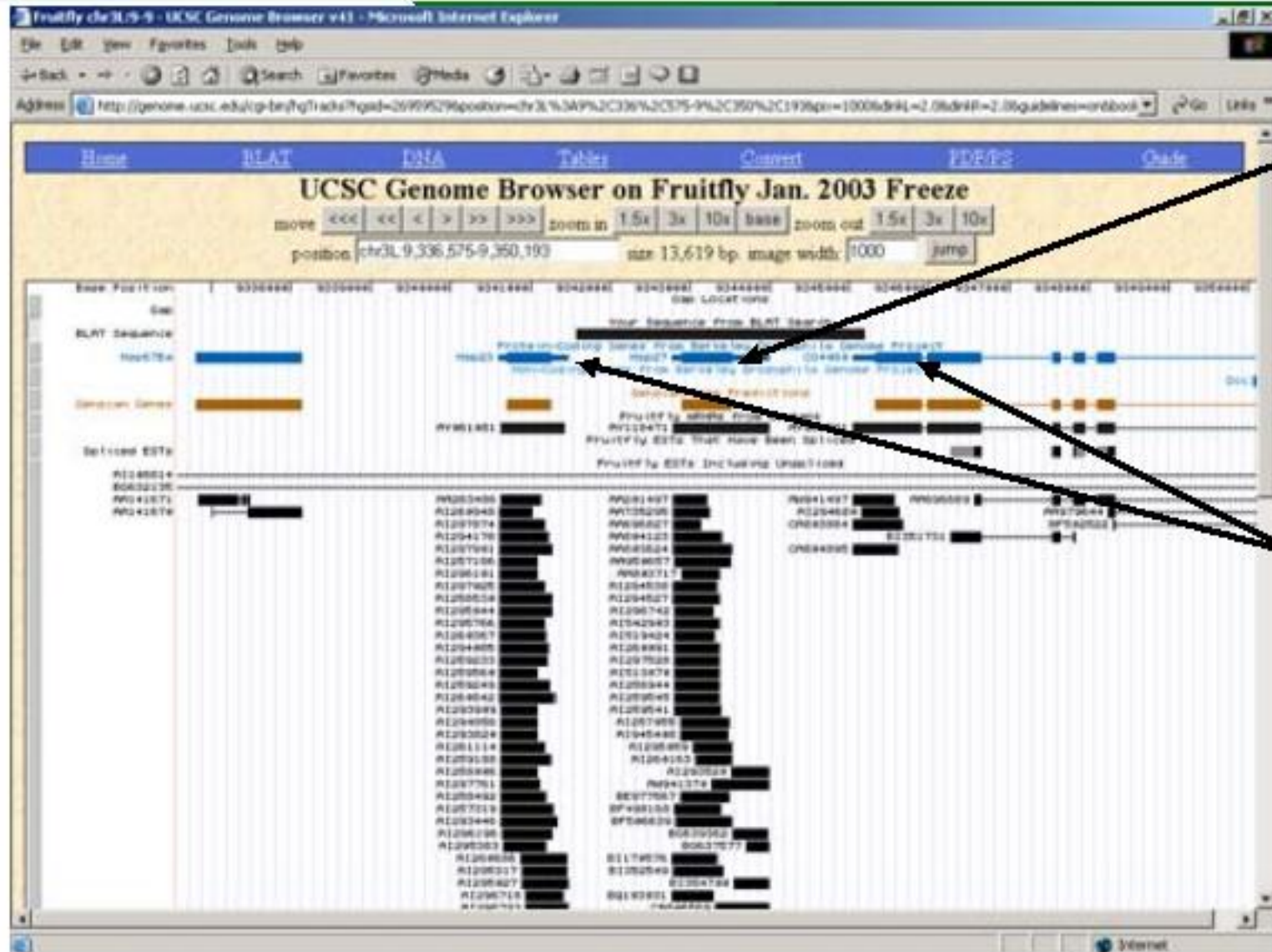


Ген HSP27

EST,  
фрагменты  
экспрессиру-  
ющихся  
последова-  
тельностью



# Опция изменения масштаба – в карту положения гена *HSP27* на хромосоме включены соседние гены



Ген *HSP27*

соседние  
с *HSP27*  
ГЕНЫ



# Saccharomyces Genome Database, <http://www.yeastgenome.org/>

Quick Search:  Submit

[Site Map](#) | [Help](#) | [Full Search](#) | [Home](#)

[Community Info](#) | [Submit Data](#) | [BLAST](#) | [Protein](#) | [Pathways](#) | [Genetic Resources](#) | [Virtual Library](#) | [Contact SGD](#)

SGD™ is a scientific database of the molecular biology and genetics of the yeast *Saccharomyces cerevisiae*, which is commonly known as baker's or budding yeast.

## New and Noteworthy

- **Improved gene product descriptions in SGD**  
SGD curators are in the process of revising the information currently associated with the Description field, for all protein-coding and non-coding RFA genes. This project will ensure that the Description, intended to be a concise summary of the function and biological context of each gene product, contains the most up-to-date information and is written in a consistent style. New Descriptions will be added to SGD gradually over the coming weeks. (Posted October 31, 2003)
- **Large-scale localization data added to SGD**  
Based on the genome-wide protein localization analysis published by [Huh et al. \(2003\) Nature 425:686-691](#) and protein abundance data published by [Chaemmaghani et al. \(2003\) Nature 425:737-741](#), Cellular Component GO terms have been added for 1326 ORFs that were previously annotated to "cellular\_component unknown". In addition, links to the [YeastJEP](#) database are available on the pull-down

поле  
быстрого  
поиска

Введём в поле название гена *MFA2*





# Результаты быстрого поиска по запросу “MFA2”

SGD Quick Search:  Submit Site Map | Help | Full Search | Home

Community Info Submit Data BLAST Images PathMap GeneSeq Resources Virtual Library Contact SGD

## MFA2/YNL145W [Help](#)

[Alternative single page format](#)

### MFA2 BASIC INFORMATION

Standard Name	MFA2
Systematic Name	YNL145W
Feature Type	ORF, Verified
GO Annotations	<a href="#">MFA2 GO evidence and references</a>
Molecular Function	<ul style="list-style-type: none"><li><a href="#">pheromone activity</a></li></ul>
Biological Process	<ul style="list-style-type: none"><li><a href="#">signal transduction during conjugation with cellular fusion</a></li></ul>
Cellular Component	<ul style="list-style-type: none"><li><a href="#">extracellular</a></li></ul>
Description	lipopeptide mating pheromone
Gene Product	a-factor mating pheromone precursor
Phenotype	<a href="#">MFA2 Phenotype details and references</a>

### MFA2 RESOURCES

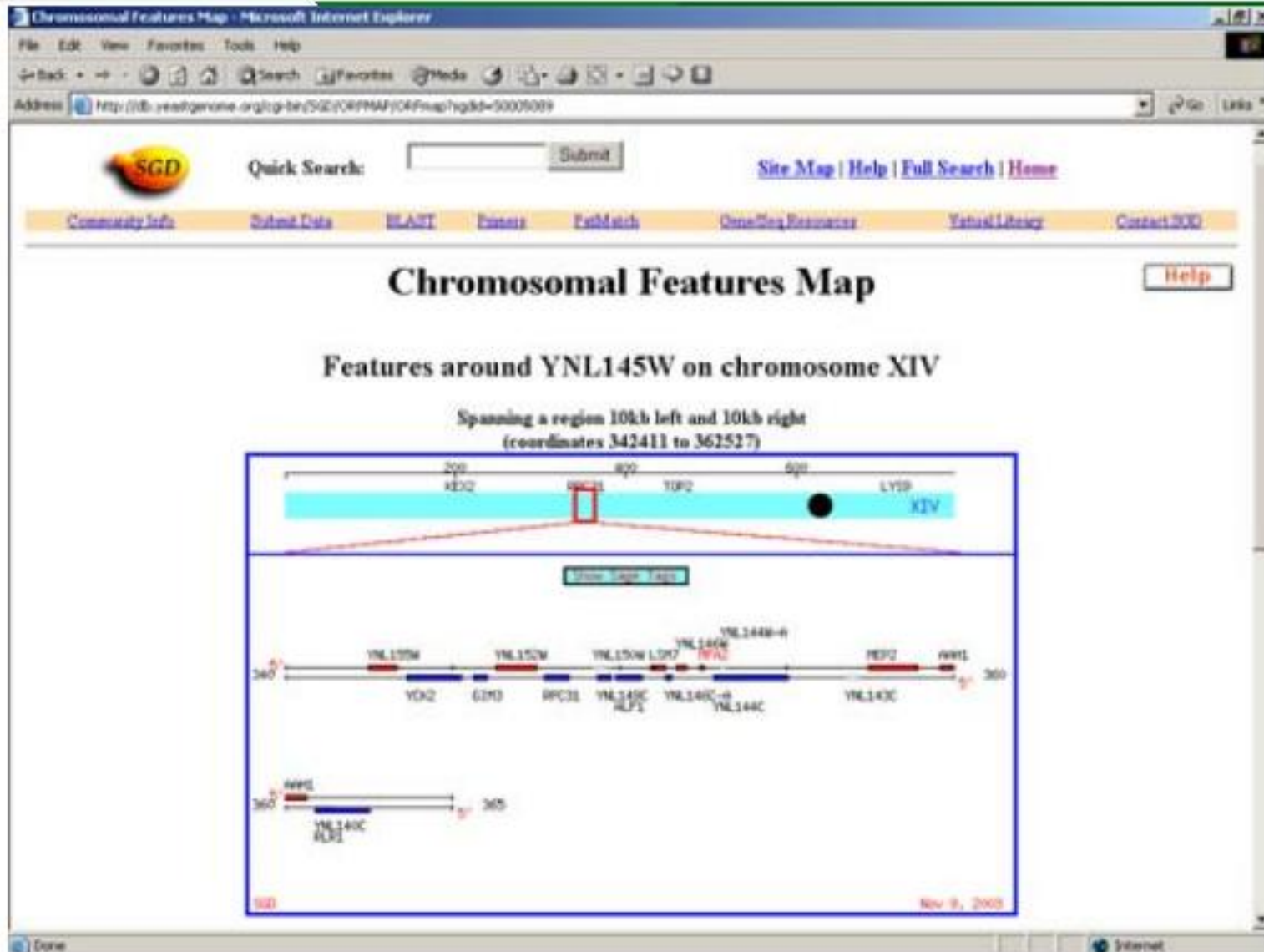
Click on map for expanded view

- Literature  
Literature Guide
- Retrieve Sequences  
Genomic DNA
- Sequence Analysis Tools  
BLASTP
- Protein Info & Structure

ЛОКАЛИЗАЦИЯ  
ПОСЛЕДОВАТЕЛЬНОСТИ  
ДНК на  
хромосоме

ДОСТУП К  
ПОСЛЕДОВАТЕЛЬНОСТИ  
ДНК

# Локализация последовательности гена *MFA2* на хромосоме 14





# Последовательность ДНК гена *MFA2* (транскрипт +5' и 3' районы гена запрошенной ранее пользователем длины)

The screenshot shows a Microsoft Internet Explorer browser window displaying the SGD (Saccharomyces Genome Database) website. The address bar shows the URL: <http://db.yeastgenome.org/cgi-bin/Seq/seq?seq=YNL145W/mfa2+1000&dir=+1000&map=s3map>. The page title is "Sequence for a region of YNL145W/MFA2". The SGD logo is visible in the top left. A navigation bar includes links for "Community Info", "Dataset Data", "BLAST", "Clones", "FullMatch", "GeneSeq Resources", "Yeast Library", and "Contact SGD". Below the navigation bar, there is a "Quick Search" field and a "Submit" button. Links for "Site Map", "Help", "Full Search", and "Home" are also present. The main heading is "Sequence for a region of YNL145W/MFA2". Below this, there are links for "Send questions or suggestions to SGD" and "BLAST search | FASTA search | Genome Restriction Map | Design Primers for this sequence". The text states: "The currently selected gene/sequence is: **YNL145W/MFA2** on chromosome XIV from coordinates 352411 to 352527 plus 1000 basepairs of upstream sequence and 1000 basepairs of downstream sequence." Below this, there is a link for "Other Formats Available" and a "GO" button. The DNA sequence is displayed in a monospaced font, starting with ">YNL145W Chr 14" and followed by 20 lines of nucleotide bases.

Sequence for a region of YNL145W/MFA2

Send questions or suggestions to SGD

[BLAST search](#) | [FASTA search](#) | [Genome Restriction Map](#) | [Design Primers](#) for this sequence

The currently selected gene/sequence is:

**YNL145W/MFA2** on chromosome XIV from coordinates 352411 to 352527 plus 1000 basepairs of upstream sequence and 1000 basepairs of downstream sequence.

Other Formats Available [GO](#)

```
>YNL145W Chr 14
TTCTTACAGTTGAAAAACAAAATCAAAATCAAGGATATGCGATTTCCTTTACATATTCCTCAAT
ACCTGACTGACATAGTATGCGCTACATAGTTTTGTCTATTCTTTTACACCGGGATTTCG
AAGTCTAAAGCCAGAGAATGAGGAAAAACGAAACAAACTGAGATCATTGTATAAAGGGA
GTCACTGTACAGCTGCCAATACCAATTTCTCTGCGAAATCAAAAAAGTATCGGCATA
CACATACAGCCCTGCTCTACAACTTAAATTAATTAATGATTCCTTTGATTTGATTCGCA
TATGAGCAATATAAGCATACCACTTCTCATCATGGAACCAAAAAGGATCATTATGCT
TACATTTTATTCATACATATAATGATATTTGGAACCAAGTGTCTCAGAAATGAC
ATTCAGAACTACTGCCAGATTTTACAAAAAGAAATTCAGACTCAGTTAATTAATAA
CAGGCCATCTATTTCTGGAAGAACTTGGTGGATTTGATGATGTTAAGCATGGTCA
TAAAGCCAGACTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT
ATAAATTAAGAAAATTAAGCAAGCATGTTTTTATTTGAAACAAATACTAATGCAATAATA
```





<http://www.xenbase.org/>

**Xenbase: a *Xenopus* web resource**

**Jump to XMMR**

A database of information pertaining to the cell and developmental biology of the frog, *Xenopus* ([introduction to \*Xenopus\*](#))

**Request for Community Input**

- [community](#)
- [Xenopus genome!](#)
- [genomics tools](#)
- [gene expression](#)
- [NIH X resources](#)

**New REA (2003)**

- [white pages](#) (people)
- [green pages](#) (labs)
- [jobs](#) (vacancies)
- [events & courses](#)
- [blast trace files](#) (5.8 Gbp)
- [download draft](#)
- [visit XGI project](#)
- [ESTs](#) (>492,000)
- [microarrays](#), bioinformatics
- [by gene name](#)
- [by site of expression](#)
- [CGAP XGC \(\*Xenopus\* Gene Collection\)](#)
- [NIH initiative dbEST](#)
- [PubMed BLAST](#)

**search**

- search Xenbase

**cell biology**

- [cell cycle](#)
- [cytoskeleton](#)
- signal transduction.

**literature**

- [e-journals](#), frog [books](#)
- [automated searches of the \*Xenopus\* literature](#) ... (e.g. [Xenopus heart](#))

**genetics**

- [transgenics](#)

**methods/materials**

- [obtain frogs](#)
- [care for frogs](#)
- [cDNA libraries](#), RT-PCR, [antisense](#), [morpholinos](#)






<http://www.sanger.ac.uk/>

The screenshot shows a Microsoft Internet Explorer browser window displaying the homepage of the Wellcome Trust Sanger Institute. The browser's address bar shows the URL <http://www.sanger.ac.uk/>. The website header features the Sanger Institute logo, which includes a stylized DNA double helix and the text "The Wellcome Trust Sanger Institute". Below the logo is a brief description of the institute's mission: "The Sanger Institute is a genome research institute primarily funded by the Wellcome Trust. Our purpose is to further the knowledge of genomes, particularly through large scale sequencing and analysis. Wellcome Trust Genome Campus, Hinxton, Cambs, CB11 1TA, UK".

The main content area is divided into two columns. The left column contains a navigation menu with the following items:

- **Information, Teams, Faculty & People**  
News, Introduction, Your Genome, Press, Seminars, Travel, Library
- **Ensembl Genome Data Resources**  
Human, Mouse, Rat, Pig, Zebrafish, Musashi, Fly, Trachea Resources
- **Software, Blast & Databases**  
Acdb, Artemis, GeneDB, Hman, Hman 2.0, Human BLAST, CeleraBLAST
- **Genome Sequencing & Annotation**  
Yeast, Human, Mouse, Zebrafish, Microbes, Worm, Drosophila, Genome, *S. cerevisiae*
- **Genomics & Genetics**  
Cancer Genome Project, Human Genetics, Mouse, *C. elegans*, Escherichia
- **Genome Infrastructure**  
Gene Resources, Microarrays, Genomics, Bioinformatics, Robotics
- **Careers At The Sanger Institute**  
Graduate, Postdoctoral & Faculty, Advertised Jobs

The right column features a news section with several articles:

- 5th Nov 2003** [More](#)  
**Ensembl version 1.8 released**  
The Ensembl Developers are pleased to announce the release of Ensembl 1.8. 
- 3rd Nov 2003** [More](#)  
**Sanger Institute completes first billion genetic letters**  
The Wellcome Trust Sanger Institute has broken a genetic decoding record on its tenth anniversary which could literally take 
- 29th Oct 2003** [More](#)  
**PhD opportunities at the Sanger Institute**  
The Sanger Institute PhD programme provides research opportunities and training for graduate students who wish to study for a PhD. 
- 28th Oct 2003** [More](#)  
**WormBase WS112 released**

At the bottom of the page, a status bar indicates "Finished Sequence: 2,010,545,837 bases". The browser's status bar at the very bottom shows "Internet".




<http://genome.wustl.edu/>

The screenshot shows a Microsoft Internet Explorer browser window displaying the homepage of the Genome Sequencing Center at Washington University in St. Louis. The browser's address bar shows the URL <http://genome.wustl.edu/>. The website header features the center's name and the university's logo. The main content area includes a descriptive paragraph about the center's role in the Human Genome Project, an image of an open -80 freezer with the caption "Archival Ceres in -80 Freezer", and a paragraph about the open sharing of genomics methodology. A sidebar on the right lists recent news items with dates and "More" links. At the bottom, there are navigation menus for "Center Information", "Human Genome", "Mouse Genome", and "C. elegans Genome".

Genome Sequencing Center  
Washington University in St. Louis  
School of Medicine

The Genome Sequencing Center at [Washington University Medical School](#) focuses on the large scale generation and analysis of DNA sequence. We play a leadership role in [The Human Genome Project](#), constructing the clone map and contributing 25% of the finished sequence. To better understand the human genome sequence and to advance the study of biology, we also have sequenced the genomes of other species as can be found in these pages.



Archival Ceres in -80 Freezer

The open sharing of genomics methodology and the rapid release of sequence data via our website serves to empower other researchers and facilitates the elucidation of gene function, forming the ultimate basis for an understanding of how DNA sequence dictates the form and function of an organism, both in health and in disease.

- **Center Information**  
[General Information](#), [Employment Opportunities](#), [Map and Directions](#), [Staff Home Pages](#), [Presentations](#), [Outreach Program](#), [Meeting Abstracts](#), [Site Feedback](#)
- **Human Genome**  
[Overview](#), [BAC database](#), [Clone Library Sources](#), [Sequence Data](#), [G15 Human Genome Finishing Standards](#)
- **Mouse Genome**  
[Genome Database](#), [Chromosome](#)
- **C. elegans Genome**  
[Genome Map](#), [GDB](#)

09 Jul 03 [More](#)  
Sequence of Human Chromosome 7 is Fine-Tuned and Finished ([publication at Nature](#))

29 Jun 03 [More](#)  
Researchers Discover Use of Novel Mechanism Preserves Y Chromosome Genes

30 May 03 [More](#)  
Yeast Genomes Reveal New Sites of Gene Control ([publication at Science](#))

05 Dec 02 [More](#)  
Initial Sequencing and Analysis of the Mouse Genome

15 Jul 02 [More](#)  
New C. briggsae WGS assembly BLAST database available

06 Jun 02 [More](#)  
Maize Mitochondrial Sequence Available See the [data release policy](#). Data available





<http://www.fruitfly.org/>

The screenshot shows the FlyBase website interface. The browser window title is "FlyBase @ Flybase.bio.indiana.edu - Microsoft Internet Explorer". The address bar shows "http://flybase.org". The main content area features the "FlyBase" logo and the subtitle "A Database of the Drosophila Genome". Below this, there are two columns of links: "Data Classes" and "Selected Searches & Tools". The "Data Classes" column includes links for Maps, Genes, Sequences, Stocks, Transgenes & Transposons, Aberrations, Anatomy & Images, and References. The "Selected Searches & Tools" column includes links for Cytologic maps, Annotated Genome, Gene Products, Gene Expression, EST project cDNAs, and Stock Collections. At the bottom of the main content area, there is a search box with a "Search" button and a "Clear" button. The search box contains the text "BLAST sequence search at BDGP, EDGP" and "for words: [input field]". Below the search box, there is a dropdown menu labeled "in these sections" with "All sections" selected. The left sidebar contains various navigation links such as "Site Map", "Genus Started", "Documents", "News, meetings & announcements", "Drosophila resources", "FlyBase mirrors", and "Important News". The "Important News" section includes "FlyBase Next Generation Preview" (November 2003), "Call for Inbred Strains" (October 2003), and "Evoking EP Lines" (October 2003).



спасибо за внимание!