

INTRODUCTION TO BIOINFORMATICS BIOLOGICAL DATA COLLECTIONS

Katarzyna Kubiak, PhD, 30.09.2014

Lecture 1

Introduction to bioinformatics

- Basic definitions
- □ Scope of the course

Internet bio-data resources

- EBI
- ExPASy
- Sanger Institute
- **KEGG**
- BRENDA



I UNDERSTAND, therefore I am!

Bioinformatics is entirely necessary for biotechnology in era of growing amounts of biological data.

What does "data" mean in Bio-sciences?

We need to have an easy access to all existing information in our research field.

Amount of information extends traditional means of sharing it.

Hopefully bioinformatics analysis will enable us to discover the most reliable "truths" (the ones which are in accordance with all/the most of existing data).



Biological papers flood



Growth of GenBank

From 1982 to the present, the number of bases in GenBank has doubled approximately every 18 months



GenBank Data (release 203 - Aug 2014)

Uncompressed, the Release 203.0 flat files require roughly 652 GB

ftp://ftp.ncbi.nih.gov/genbank/gbrel.txt

Growing understanding of complexity of life

NCBI Taxonomy new Nodes: 2014

NCBI Taxonomy new Nodes: all dates

Ranks:	higher taxa	genus	species	lower taxa	total	Ranks:	higher taxa	genus	species	lower taxa	total
Archaea	7	5	8	0	20	Archaea	143	140	524	0	807
Bacteria	64	144	220	35	463	Bacteria	1370	2611	13318	819	18118
Eukaryota	812	2628	16941	1374	21755	Eukaryota	20436	67549	296479	22545	407009
Fungi	121	262	1379	57	1819	Fungi	1549	4618	29222	1104	36493
Metazoa	459	2066	9601	695	12821	Metazoa	14665	45477	144665	11390	216197
Viridiplant ae	182	213	5556	602	6553	Viridiplan tae	2622	14667	113417	9788	140494
Viruses	33	34	102	0	169	Viruses	617	442	2349	0	3408
All taxa	918	2811	17268	1409	22406	All taxa	22595	70749	312701	23364	429409

Definitions

According to National Institutes of Health (NIH):

□ "**BIOINFORMATICS** is:

research, development, or **application of computational tools** and approaches **for expanding the use of biological**, medical, behavioral or health **data**, **including those to** acquire, **store**, **organize**, **analyze**, **or visualize such data**"

□ "COMPUTATIONAL BIOLOGY is:

the development and application of data-analytical and theoretical methods, mathematical modeling and computational simulation techniques to the study of biological, behavioral, and social systems".



Bioinformatics course



Scope of the Bioinformatics course

LECTURES (15 hours) Dr Katarzyna Kubiak Tue 12:15-14:00

- □ ECTS 2 credits
- □ Last lecture (25.Nov): **TEST**

Final mark up to 4.5 = sum of test & practical results

□ 5 – after oral questioning

PRACTICALS (30 hours) MSc Natalia Kwiatos Mon 8:30 – 11:00 Wed 11:15-14:00

START: 6th of October

- □ 6 individual exercises
- Small project in pairs finished with presentation

The course website

http://bioinformatics.p.lodz.pl

- Timetable with lectures/laboratoty topics
- Weekly updated downloads:
 - Whole lectures & printouts
- Useful links
- Now available: lectures from the last academic year Passwords:
- L1 INTRO; L2 DATA; L3 BLAST; L4 MULTI; L5 PROT; L6 GENE
- All software used during this course is free available online



Textbooks in English (available at my office: room9, BiNoZ ground floor)

The course has no required textbook.

The lectures in this course correspond to a few chapters of Jonathan Pevsner Bioinformatics and Functional genomics, Wiley-Blackwell, 2009. Author's lectures and additional documents are there available on the publisher website (http://www.bioinfbook.org/)



Information sheet for the course

Code 05 36 0015 01

ECTS credits: 2

Course name	Bioinformatics
Goal	 The aim of the course is to enable students to acquire knowledge in the field of the contents of biological databases (catalogs of sequences, structural and literature data, catalogs of enzymes and model organisms) available online. The aim of the course is to develop/ shape students' skills in usage of basic bioinformatic freeware to solve typical problems in molecular biology and in biotechnological experiments design.
Learning outcomes	 Having accomplished the course a student will have been able to: Find in biological databases the scientific and patent literature and the biological information on: genes and proteins sequences, genomes sequences, enzymes properties, proteins structures, structural and functional patterns in protein domains, taxonomic names of organisms. Explain application of basic bioinformatic freeware, such as: BLAST with all subtypes, PyMOL, Clustal, MUSCLE, Tcoffee, PrimerBLAST, NEBcutter, MFOLD, Glimmer, GeneMark. Explain the principles of applications (and relevant algorythms) computing pairwise and multiple sequence alignment (FASTA, BLAST, megaBLAST, Clustal, MUSCLE, Tcoffee, ProbCons) and choose appropriate scoring matrices, gap penalties and word sizes in above mentionned applications according incoming data used as well as interpret the results (clarify the minning of alignment score and evaluate parameters, finding conservative regions/residues, conclude about relationship of aligned sequences). Use online versions of basic bioinformatic freeware (BLAST with all subtypes, PyMOL, Clustal, MUSCLE, Tcoffee, PrimerBLAST, NEBcutter, MEGA, MFOLD, Glimmer, GeneMark, TaxPlot, mummer) choosing parameters in these programs properly to solve typical issues in biological experiments design, their results interpretation and anlysis of data from biological databases. Explain application of algorithms using machine learning (especially using hidden Markov model) in solving such bioinformatic problems as: ORF finding in DNA sequences or domains in protein sequences. Formulate hypotheses about expected protein function based on bioinformatic analysis of the sequence of unknown DNA fragment encoding this protein. Explain how informatics methods are used in results analysis of DNA microarray experiment (statistics, R project) and in proteomics (2D gels analysis, analysis of mass spectrometry results).

Lecture 1

Introduction to bioinformatics

- Basic definitions
- □ Scope of the course

Internet bioinformatic resources

- ExPASy
- Sanger Institute
- □ KEGG
- BRENDA

Databases used in bioinformatics



Common websites with access to data and tools

NCBI - National Center for Biotechnology Information (USA)

- GenBank; RefSeq, PubMed; OMIM
- BLAST

EBI – European Bioinformatics Institute

- ENA; UniProt; ArrayExpress;
- Euro PubMed;



NCBI

EMBL-EB

ExPASy - Expert Protein Analysis System

- UniProt; prosite; 2D-page results
- Great source of protein & proteomic tools (e.g. SWISS Model)

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

SNCBI Resources 🕑 How To 🕑

National Center for Biotechnology Information

Search All Databases • for

katarzynakubiak | My NCBI | Sig

Searc

Resources

NCBI Home
All Resources (A-Z)
Literature
DNA & RNA
Proteins
Sequence Analysis
Genes & Expression
Genomes
Maps & Markers
Domains & Structures
Genetics & Medicine
Taxonomy
Data & Software
T 0 T

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

More about the NCBI | Mission | Organization | Research | RSS

PubMed Central

Free Full Text. Over 1,500,000 articles from over 450 journals. Linked to PubMed and fully searchable.

II 1 2 3 4

How To...



November and October 02 Dec 2009 News

http://www.ncbi.nlm.nih.gov NAVIGATION

You are here: NCBI > National Center for Biotechnology Information

GETTING STARTED NCBI Education NCBI Help Manual NCBI Handbook Training & Tutorials RESOURCES Chemicals & Bioassays Data & Software DNA & RNA Domains & Structures Genes & Expression Genetics & Medicine Genomes & Maps Homology Literature Proteins Sequence Analysis Taxonomy Training & Tutorials Variation

POPULAR PubMed Nucleotide BLAST PubMed Central Gene Bookshelf Protein OMIM Genome SNP

FEATURED

Genetic Testing Registry PubMed Health GenBank Reference Sequences Map Viewer Human Genome Mouse Genome Influenza Virus Primer-BLAST Sequence Read Archive

Write

- NCBI INFORMATION About NCBI Research at NCBI
- NCBI Newsletter
- NCBI on Facebook
- NCBI on Twitter
- NCBI on YouTube

Copyright | Disclaimer | Privacy | Accessibility | Contact

National Center for Biotechnology Information, U.S. National Library of Medicine 8600 Rockville Pike, Bethesda MD, 20894 USA



http://www.ncbi.nlm.nih.gov SEARCH

🗧 NCBI 🛛 Resources 🕑 How	To 🗹	katarzynakubiak My NCBI S
National Center for Biote	echnology Information Search All Databases	or Sea
Resources	Malagma to NCDI	
NCBI Home		Popular Resources
All Resources (A-Z)	The National Center for Biotechnology Information advances science and	PubMed
Literature	health by providing access to biomedical and genomic information.	 PubMed Central
DNA & RNA	More about the NCBI Mission Organization Research RSS	Bookshelf
Proteins		Gene
Sequence Analysis	Rub Mad Cantral	Nucleotide
Genes & Expression	Publied Central	GEO
Genomes	Free Full Text, Over 1,500,000 articles	Conserved Domains
Maps & Markers	from over 450 journals. Linked to PubMed and fully searchable.	Structure PubChem
Domains & Structures		- r abonem
Genetics & Medicine	II 1 2 3 4	NCBI News
Taxonomy		
Data & Software	How To	November and October 02 Dec 2009

Entrez – an easy access to all NCBI databases



۲	3	9	3	9	۲	3	۲	3	9	3	3	3	3	
dbGaP: genotype and phenotype	UniGene: gene-oriented clusters of transcript sequences	CDD: conserved protein domain database	3D Domains: domains from Entrez Structure	UniSTS: markers and mapping data	PopSet: population study data sets	GEO Profiles: expression and molecular abundance profiles	GEO DataSets: experimental sets of GEO data	Cancer Chromosomes: cytogenetic databases	PubChem BioAssay: bioactivity screens of chemical substances	PubChem Compound: unique small molecule chemical structures	PubChem Substance: deposited chemical substance records	Protein Clusters: a collection of related protein sequences	Peptidome: MS/MS proteomic experiments	
	Ø		2		8					S		٩		
275	68	none	509	20	29	737	10	none	ŝ	none	8	none	none	
3	۲	۲	3	۲	۲	۲	3	۲	3	Э	۲	9	۲	3
2985 Nucleotide: Core subset of nucleotide sequence records	8348 5 EST: Expressed Sequence Tag	6989 (1) records	7594 Protein: sequence database	12 () Genome: whole genome sequences	128 Structure: three-dimensional macromolecular structures	non GenBank	none M Polymorphism	99 Gene: gene-centered information	none 🕕 ska: Sequence Read Archive	none Systems of interacting molecules	5 HomoloGene: eukaryotic homology groups	2866 Second Seco	153 Probe: sequence-specific reagents	224 Genome Project: genome project information

\mathfrak{X}^{NCBI} Res	sources 🗸	How	To 🔇					
GenBank		Nucle	eotide					
				J				
GenBank	Submit	•	Genomes 🔻	WGS 🔹	HTGs 🔻	EST/GSS <	Metagenomes 👻 TPA 👻	TSA
GenBank (Overviev	3					GenBank Res	ource
							<u>GenBank Home</u>	

What is GenBank?

sequences (Nucleic Acids Research , 2011 Jan;39(Database issue):D32-7). There are approximately GenBank $^{\odot}$ is the NIH genetic sequence database, an annotated collection of all publicly available DNA 126,551,501,141 bases in 135,440,924 sequence records in the traditional GenBank divisions and 191,401,393,188 bases in 62,715,288 sequence records in the WGS division as of April 2011.

Update GenBank Recol

Submission Types

Submission Tools

Search GenBank

<u>Database Collaboration</u> , 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 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 oasis.

An example of a GenBank record may be viewed for a Saccharomyces cerevisiae gene.

Access to GenBank

There are several ways to search and retrieve data from GenBank.

- divided into three divisions: CoreNucleotide (the main collection), dbEST (Expressed Sequence Search GenBank for sequence identifiers and annotations with Entrez Nucleotide, which is Tags), and <u>dbGSS</u> (Genome Survey Sequences).
- Alignment Search Tool). BLAST searches CoreNucleotide, dbEST, and dbGSS independently; Search and align GenBank sequences to a query sequence using <u>BLAST</u> (Basic Local see BLAST info for more information about the numerous BLAST databases. •

S NCBI Resources	🗹 How To 🖾	Sign in to N
Pub Med.gov	PubMed •	Searc
JS National Library of Medicine National Institutes of Health	Advanced	



PubMed

PubMed comprises more than 24 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full-text content from PubMed Central and publisher web sites.

PubMed Commons

Featured comment - Sep 29 Subtyping #schizophrenia? Group of genetics researchers critique GWAS analysis, via G Breen (@psychgenomics) <u>1.usa.gov/1ru8Q6u</u>

ch

Using PubMed

PubMed Tools

More Resources

PubMed Quick Start Guide

PubMed Mobile

MeSH Database

MEDLINE = Medical Literature, Analysis, and Retrieval System Online

the world biggest life sciences bibliographic database contains citations form over 5,600 worldwide biomedical journals

MEDLINE also covers life sciences vital to biomedical practitioners, researchers, and educators, including aspects of biology, environmental science, marine biology, plant and animal science as well as biophysics and chemistry. Increased coverage of life sciences began in 2000.

Boolean operators – e.g. literature search



Search your query only in described fields

PubMed syntax: ≻[AUTHOR] ≻ [ABSTRACT] ≻ [TITLE]

Affiliation [AD]	ISBN [ISBN]	Place of Publication
Article Identifier	Issue [IP]	[PL]
[AID]	Journal Title [TA]	PMCID & MID
All Fields [ALL]	Language [LA]	Publication Date
Author [AU]	Last Author [LASTAU]	[DP]
Book [book]	Location ID [LID]	Publication Type
Comment Corrections	MeSH Date [MHDA]	<u>[PT]</u>
Corporate Author	MeSH Major Topic [MAJR]	Secondary Source
[CN]	MeSH Subheadings [SH]	ID [SI]
Create Date [CRDT]	MeSH Terms [MH]	Subset [SB]
EC/RN Number [RN]	NLM Unique ID [JID]	Substance Name
Editor [ED]	Other Term [OT]	[NM]
Entrez Date [EDAT]	Owner	Text Words [TW]
Filter [FILTER]	Pagination [PG]	<u>Title [TI]</u>
First Author Name	Personal Name as Subject	<u>Title/Abstract</u>
[1AU]	[PS]	[TIAB]
Full Author Name	Pharmacological Action MeSH	Transliterated Title
[FAU]	Terms [PA]	<u>[TT]</u>
Full Investigator		UID [PMID]
Name [FIR]		Volume [VI]
Grant Number [GR]		
Investigator [IR]		

Query may be build by typing or with "Advanced search" tool

Fubmeu Auvanceu Search	« Back to PubMed
Search Box	Limits Details Help Search Preview Clear
Search Builder	
Affiliation All Fields Author Book Corporate Author	Show Index
Create Date EC/RN Number Editor Entrez Date Filter First Author	
Full Author Name Full Investigator Name Grant Number ISBN Investigator Issue Journal	

PubMed list of results



MeSH Medical Subject Headings

🕞 🎯 www.r	icbi.nlm. nih.gov /mes	h/						∀ Cł	🚷 👻 obraz	Q	☆	ê	÷	俞	8	
S NCBI	Resources 🗹	How To 🕑										3	Sign	in to	NCE	<u>31</u>
MeSH		MeSH												Sea	rch	Í
			Ĩ	Limits	A	dvanc	ed.								Hel	р



MeSH

MeSH (Medical Subject Headings) is the NLM controlled vocabulary thesaurus used for indexing articles for PubMed.

Using MeSH

Help

Tutorials

More Resources

E-Utilities

NLM MeSH Homepage

MeSH a database of scientific terms

Me	SH	MeSH	✓ beta globin	1				🙁 Sea	rch
			Save search	Limits	Advanced				Help
Dis	play Setting	gs: 🗹 Summary				Send to: 🕑		0 80 202	_
Do	culto: A						PubMed Sear	rch Builder	• 📩
Re	Sult5. 4								
	beta-Glo	bins							
1.	Members	of the beta-globin famil	ly. In humans, they a	re encode	ed in a gene	cluster on			
	CHROMC	SOME 11. They include	epsilon-globin, gam	ima-globin	n, delta-glob	in and			d
	beta-glob	on. There is also a pseu	dogene of beta (thet	a-beta) in	the gene cl	uster. Adult	Add to search	n builder	
	HEMOGL	OBIN IS comprised of tw	0 ALPHA-GLUBIN C	chains and	i two beta-g	IODIN			
	Chains.	ucod: 2000					Coorch Dubly	lod	
	Tear Introdu	uced. 2009					Search Publy	lea	
0	Erythremi	ia, Beta-Globin Type [S	Supplementary Conc	cept]				Vou The T	Interial
2.	Date introdu	uced: November 5, 2012					Find related data	TOUTION	
0	Methemo	globinemia, Beta-Globi	in Type [Supplemen	itary Conc	cept]		Database:		
3.	Date introdu	uced: November 5, 2012					Select	•	
1	beta-glot	oin readthrough protein.	, Oryctolagus cunicu	ilus [Supp	ementary	Concept]	Find items).	

MeSH as a tool for PubMed search building

MeSH MeSH	✓ Limits Advanced		Search Help
<u>Display Settings:</u>		Send	to: PubMed Search Builder
Members of the beta-globin family. In hum epsilon-globin, gamma-globin, delta-globi cluster. Adult HEMOGLOBIN is comprised Year introduced: 2009 PubMed search builder options Subheadings:	nans, they are encoded in a ger n and beta-globin. There is also d of two ALPHA-GLOBIN chain	ne cluster on CHROMOSOME 11. They include a pseudogene of beta (theta-beta) in the genesis and two beta-globin chains.	Add to search builder AND Search PubMed
administration and dosage	chemistry	immunology	
agonists	classification	isolation and purification	Related information
analysis	cytology	metabolism	PubMed
anatomy and histology	deficiency	pharmacology	PubMed - Major Topic
antagonists and inhibitors	drug effects	physiology	Clinical Queries
biosynthesis	etiology	therapeutic use	
chemical synthesis	genetics	ultrastructure	NLM MeSH Browser
			PubChem Substance
Restrict to MeSH Major Topic.			
Do not include MeSH terms found bel	ow this term in the MeSH hierai	rchy.	

Common websites with access to data and tools



EMBL-EBI

- NCBI National Center for Biotechnology Information (USA)
- GenBank; RefSeq, PubMed; OMIM
- BLAST

EBI – European Bioinformatics Institute

- ENA; UniProt; ArrayExpress
- EuroPubMed



ExPASy

- UniProt; PROSITE; 2D-page results
- Great source of protein & proteomic **tools** (e.g. SWISSModel)

http://www.ebi.ac.uk/

EMBL-EBI

Services Research Training About us

The European Bioinformatics Institute

Part of the European Molecular Biology Laboratory

EMBL-EBI provides freely available data from life science experiments, performs basic research in computational biology and offers an extensive user training programme, supporting researchers in academia and industry.

Find a gene, protein or chemical:

Examples: blast, keratin, bfl1...

News from EMBL-EBI



BLUEPRINT: new perspectives on white blood cells

Taking the guesswork out of genetic analysis New method by EMBL-EBI and Microsoft Research improves



Search

Advance in stem-cell technology

Scientists revert human stem cells to pristine state European Molecular Biology Laboratory

Visit EMBL.org



Popular

& Services	H Jobs				
88 Research	🕆 Visit us				
III Training	🔷 EMBL				
News	8 Contacts				

Events

Animal Genome Informatics Sep 29 2014 -Oct 3 2014 Registration deadline: Aug 15 2014

diXa Open Meeting - 29-30 September 2014 Sep 29 2014 -Sep 30 2014 Registration deadline: Sep 12 2014

Argentinian Bioinformatics Workshop, 2014 Sep 30 2014 -Oct 1 2014

Major international initiative to

Services Overview A to Z Support			
Bioinformatics So We maintain the world's most comp databases. Developed in collaborati perform complex queries and analys data and software, or use our web so	ONICOS rehensive range of freely available a ion with our colleagues worldwide, our se the results in different ways. You o services to access our resources progr	nd up-to-date <u>molecular</u> services let you share data, an work locally by downloading our ammatically.	Popular ef Ensembl UniProt BLAST A Literature A Train online A Support
DNA & RNA genes, genomes & variation	C Gene expression RNA, protein & metabolite expression	Proteins sequences, families & motifs	Bioinformatics training
Structures Molecular & cellular structures	Systems reactions, interactions & pathways	Chemical biology chemogenomics & metabolomics	Guide to resources
A Ontologies taxonomies & controlled vocabularies	Literature scientific publications & patents	Conservation of the software cross-domain tools & resources	
Programmatic access EMBL-EBI web services allow you to d	query our large biological databases	orogrammatically, so that you can	

Services Research Training Industry About us

EMBL-EBI

0

EMBL-EBI web services allow you to query our large biological databases programma develop data analysis pipelines or integrate public data with your own applications.

European sequence database

EMBL-EBI		Services	Research	Training	About us	
European Nucleotide Archive	an Nucleotide Archive		1000065, histone			
Home Search & Browse Submit & Update Software About ENA Support						
European Nucleotide Archive (ENA) provides a comprehensive record of the world's nucleotide sequencing information, covering raw sequencing data, sequence assembly information and functional annotation. More about ENA Access to ENA data is provided though the browser, through search tools, large scale file download and through the API.		Popular • Submit and update • Sequence submissions • Genome assembly submissions • Submitting environmental sequences • Citing ENA data • Rest URLs for data retrieval • Rest URLs to search ENA				
Text Search ENA = European Nucleotide Ar Examples: BN000065, histone EMBL-Bank (European / Search SPA (Sequence Poord A)	r chive is con Molecular B	nposed of: iology Laboratory	' Bank)		de	
Advanced search SKA (Sequence Redd A Sequence Search Trace Archive – trace =	SKA (Sequence Read Archive)- NGS (Next Gen Sequencing) reads Trace Archive – trace = raw data from automatic sequencers					

Enter or paste a nucleotide sequence or accession number


A citation browser with text meanning functions

 Europe PubMed Central (Europe PMC) offers fraccess to biomedical literature resources includin PubMed abstracts (about 22 million) Europe PMC full text articles (about 2.2 million, of which over 400,000 are Open Access) Patent abstracts (over 4 million European, US, and International) National Health Service (NHS) clinical guidelines Agricola records (500,000) Supplemented with Chinese Biological Abstracts and the Citeseer database. 	ee Recent Activity In About Overview PubMed + patent & agriculture literature Contact us etwork of PMC	Clipboard						
 PubMed abstracts (about 22 million) Europe PMC full text articles (about 2.2 million, of which over 400,000 are Open Access) Patent abstracts (over 4 million European, US, and International) National Health Service (NHS) clinical guidelines Agricola records (500,000) Supplemented with Chinese Biological Abstracts and the Citeseer database. Europe PMC is based on PubMed Central (PMC), developed at the NCBI in the USA and is part of a ne International (PMCI) repositories that also includes PMC Canada. It is supported by 19 funders of biol research, including charities and government organisations in the UK, Austria, and Italy, led by the We Europe PMC is developed by the European Bioinformatics Institute. The University of Maschaster (Mine)	Overview PubMed + patent & agriculture literature Contact us							
Europe PMC is based on PubMed Central (PMC), developed at the NCBI in the USA and is part of a n International (PMCI) repositories that also includes PMC Canada. It is supported by 19 funders of bio research, including charities and government organisations in the UK, Austria, and Italy, led by the We	etwork of PMC							
Europe PMC is based on PubMed Central (PMC), developed at the NCBI in the USA and is part of a network of PMC International (PMCI) repositories that also includes PMC Canada. It is supported by 19 funders of biomedical research, including charities and government organisations in the UK, Austria, and Italy, led by the Wellcome Trust. Europe PMC is developed by the European Bioinformatics Institute, The University of Manchester (Mimas and NecTable) and the British Library								
Reasons to use Europe PMC as your primary biomedical information resource:								
 It is free, comprehensive, fast and reliable. All of the abstracts and articles can be accessed through one search box. You can sort the search results by publication date, relevance, or the number of times an article cited. Links to public databases such as UniProt, Protein Data Bank (PDBe), and the European Nucl (ENA) are provided. Through textmining technologies, you can highlight and browse keywords such as gene name and diseases. 	e has been eotide Archive s, organisms							

- · Roadtest new tools based on Europe PMC content in Europe PMC labs.
- In Europe PMC plus, PIs supported by the Europe PMC funders can link grants to publication information,



PanG, a New Ketopantoate Reductase Involved in Pantothenate Synthesis. (PMID:23243306)



Clipboard

Recent Activity

(-)

Read Article at asm.org (Subscription required)

Patents in Europe PubMed

Europe PubMed	vitamin biosynthesis		् 🌣
Joonna			
COMPOSITION CONTAINING COLLAGEN SKIN CARE (PAT - US2011160137)	PEPTIDE FOR IMPROVING	 Recent Activity Back to Results 	🛱 Clipboard
KIM JEONG KEE, LEE JI HAE, YANG MI SUK, LEE JI EUN AMOREPACIFIC CORP Patent: Patent Appl. Publ. within the TVPP - United States	, KIM WAN GI	Formats Abstract Patent at EPO 5	
Application: US200913060825 on 2009/08/27 Publication: 2011/06/30		PDF Export citation (RIS)	
Abstract Gene Ontology(3) Genes/P	Highlight Terms 2 roteins(1) Species(1) Chemicals(3)	Email citation	
The present invention relates to an oral composition for i exhibits the effects of reducing skin wrinkles and inhibitir contains a collagen peptide and at least one selected fro	Abstract is always linked to the espac	available ar enet – EPO	

browser (European Patent Office)

hyaluronic acid and vitamin C. Particularly, the composition contains the collagen peptide, the elastin protein, hyaluronic acid and vitamin C at the optimum ratio, and when it is taken into the human body, it has no side effect, maximizes the biosynthesis of collagen in the skin dermal layer, shows excellent in vivo retention rate, and exhibits the effects of inhibiting skin wrinkle formation, maintaining or improving skin elasticity and moisturizing the skin. Thus, the composition will be useful as a health functional food for improving the beauty of the skin and preventing skin aging. **Priority number:**

- KR20080083759
- WO2009KR04793

 About 	Espacenet O	ther EPO online services	•		Patent w Espacenet	
Search	Result list	🜟 My patents list (0)	Query history	Settings	Help	

US2011160137 (A1)	Bibliographic	data: US2011160137 (A1) — 2011-06-30	
Bibliographic data	🖈 In my patents list	↗ EP Register → Report data error	P Print
Description			
Claims	COMPOSITION		
Mosaics	COMPOSITION C	ON TAINING COLLAGEN PEPTIDE FOR IMPROVING SKIN CARE	
Original document	Page bookmark	US2011160137 (A1) - COMPOSITION CONTAINING COLLAGEN PEPTIDE FOR IMPROVING SKIN CARE	
Cited documents	Inventor(a)		
Citing documents	inventor(s):	NIM JEONG REE [RR], LEE JI HAE [RR], YANG MI SUR [RR], LEE JI EUN [RR], RIM WAN GI [RR] \pm	
INPADOC legal status	Applicant(s):	AMOREPACIFIC CORP [KR] ±	
INPADOC patent family	Classification:	- international: A61K8/65; A61Q19/00; A61Q19/08	
		- cooperative: <u>A61K8/365; A61K8/64; A61K8/65; A61K8/676; A61K8/735; A61Q19/00; A61Q19/08</u>	
Quick help –	Application number	US200913060825 20090827	
→ What does A1, A2, A3 and B stand for after a European	Priority number(s):	KR20080083759 20080827 ; WO2009KR04793 20090827	
publication number? → What happens if I click on "In my patents list"?	Also published as:	KR20100025500 (A) WO2010024608 (A2) WO2010024608 (A3) JP2012501320 (A) CN102131492 (A)	
What happens if I click on the "Register" button?			
→ Why are some sidebar options deactivated for certain	Abstract of US2	011160137 (A1)	
documents?	Translate this text into	i	
How can I bookmark this page? Why does a list of documents	Chinese	patenttranslate powered by EPO and Google	
 with the heading "Also published as" sometimes appear, and what are these documents? Why do I sometimes find the abstract of a corresponding document? What happens if I click on the red "sotest translate" button? 	The present invention of and inhibiting wrinkle for protein, hyaluronic acid vitamin C at the optimur skin dermal layer, show skin elasticity and mois and preventing skin ag	elates to an oral composition for improving the beauty of the skin, which exhibits the effects of reducing sommation. The composition contains a collagen peptide and at least one selected from the group consisting of and vitamin C. Particularly, the composition contains the collagen peptide, the elastin protein, hyaluronic action natio, and when it is taken into the human body, it has no side effect, maximizes the biosynthesis of colla vs excellent in vivo retention rate, and exhibits the effects of inhibiting skin wrinkle formation, maintaining of turizing the skin. Thus, the composition will be useful as a health functional food for improving the beauty of the skin.	kin wrinkle of elastin sid and gen in the r improving of the skin

"patent translate" button?

European Patent Office

Europäisches Patentamt European Patent Office Office européen des brevets	Site searc Enter search term → Advanced search	h Patent search Search	Deutsch English Français About us I Contact Service & support
Hom Searching for patents	Applying for a patent	Law & practice News & issues	Learning & events
 Start searching European patent register European publication server Espacenet - worldwide patent search Patent translate More online searching 	 Before you search Search essentials Patent information tour 	Image: constrained of the second of the s	We are recruiting 20 scientists and engineers to work as patent examiners • Apply now! • Apply now! • Read more • Read more • State • State <td< th=""></td<>
Website updates	News	About us	53 901 60 776 60 937 Online services and

6.3.2013 OPS possible outages and disruptions between 8 and 11 March 2013

4.3.2013 New online open consultation 100 10 T 1 1 1 1 1 1 1

6.3.2013 A record year for the EPO

6.3.2013 EU Science conference: President Benoît Battistelli highlights EPO's role



software

Please select

Your central access point for online tools, software downloads and availability information

Europäisches Patentamt	Site search	Patent search	Deutsch English Françai
European Patent Office	Enter search term	Search	About us Conta
Office européen des brevets	Search current area only →.	Advanced search	Service & suppo
Home Searching for patent	s Applying for a patent Law & p	practice News & issues Learnin	g & events
Free online services Subscriptio	n products Asian patent information	Patent information centres Essentials	Data
Espacenet - patent search			
European patent register			
Third-party observations			Website updates
European publication server			
European Patent Bulletin	Before you search		6.3.2013
Open Patent Services			OPS possible outages and disruptio
EBD	Search essentials		petween 8 and 11 March 2013
IPscore	→ Patent information tour		25.2.2013
European patent applications It se and specifications	sarch		Nouce concerning the survey about online publication of the Official Jour
Common Citation Document			20.2.2013
Patent translate			guides for Korean patent information updated
			19.2.2013 Temporary unavailability of OPS
			14.2.2013
Data from the EPO	Asian patent information	Upcoming events	Asian patent information: Search quides for Chinese sources updated
Database collections	Virtual helpdesk	12.3.2013 Virtual classroom, online	7.2.2013
	Help with patent information from:	An overview of EPO patent information products	Registration for East meets West 20 - Forum on Asian patent information
	+ Chinese Taipei	15.3.2013 Virtual classroom, online	now open
A Dow data download area	↓ India	Patent information essentials	4.2.2013
About raw data products and sample data	↓ <u>Korea</u>	18.3.2013 Vienna, Austria Patent searching for beginners - EPO Potent Information Boginners Comingr	An e-learning module on Cooperativ Patent Classification (CPC) is now available
	Other services		5 FUL C V

esp@cenet jest dostępny w wersji polskojęzycznej

Europäisches Patentamt European Patent Office Office européen des brevets	Site search Pater Enter search term	t search Search	Deutsch English Français About us I Contact Service & support	
Home Searching for particular Free online services Subsc Home → Searching for patents → F	tents Applying for a patent Law & practice Ne ription products Asian patent information Patent information reconline services → Espacenet - patent search	ws & issues Learning & e on centres Essentials Data	events	
Espacenet - patent search	Espacenet		Print 🖨	
Release notes & known Issues	Patent search		See also	
About searching in Espacenet	Economic offers free economic to more than 70 million potent decum	anto worldwide, containing	→ Patent Information News	
Fair use charter	information about inventions and technical developments from 1836	ô to today.		
European patent register				
Third-party observations	Direct access			
European publication server	Access Espacenet at the EPO			
European Patent Bulletin	Search the EPO's databases in German, English and French			
Open Patent Services	Onen Fenarcanet at the EPO			
EBD	open Lapacenet at the LF o			
IPscore	Access Espacenet at the European Commission			
European patent applications and specifications	Search the EPO's databases in German, English and French	Here are avaic	able all countries w	hich may
Common Citation Document	Open Espacenet at the EC			/
Patent translate	Access Espacement at national and regional offices	access espacen	iet.	
	Search using an interface in your language	Exotic localizat	ions offen are atro	tive due to
	Please select Go	transfer speed		
	→ Available interface languages	-		
	Getting started			



Espacenet Wyszukiwanie patentów

yszukuj Wynik 🚖 Mój	wykaz patentów (0) Historia zapytań	Ustawienia Pomoc	
teligentne wyszukiwanie	Wyszukiwanie zaawansowa	ne It is possible to shoose los	
yszukiwanie Proste			cui
yszukiwanie zaawansowane	1.Wybierz bazę danych	the one "from which you o	cor
yszukiwanie według numeru	Wybierz baze danych w której chcesz wyszukiw	ać i	
vszukiwanie wo ECLA	Worldwide - Pełen zbiór dokumentacji zołoszer	niowej z przeszło 80 krajów	\mathbf{v}
bka Pomoc –	2.Wprowadź wyrażenia wyszukiwawc	ze	
e terminów można wprowadzić	Wprowadź słowa kluczowe (w bazach: Worldwid - w językach danego kraju). Kombinacja klawisz	Le de, EP i WO - w jęz. angielskim, a w bazach narodowy y Ctrl+Enter powiększa pole, w którym wprowadzasz	ych
zy mogę wyszukiwać za	wyłażenia wyszukiwawcze.		
omoca kombinacji prowadzapych słów?	Słowo(a) kluczowe w tytule: i	plastic and bicy	cle
ak wprowadzać słowa z opisów			1
b zastrzeżeń?			
<u>zy mozna stosowac symbole</u> jeloznaczne?	Słowo(a) kluczowe w tytule lub w skrócie: 1	h	nair
ak wprowadzać numery:			-//
ublikacji, zgłoszenia,	Numer publikacji: i	WO20080145	520
dnoszace sie do literatury			//
iepatentowej?			
ow do I enter the names of	Nr zgłoszenia lub nr publikacji zgłoszenia: i	DE199710316	696
ersons and organisations? aka jest różnica pomiedzy MKP			-//
ECLA?	Numer pierwozeństwo:	WO100ELIS1E0	125
<u>akie formaty dat obowiązują w</u>		WO 199505 159	.25
<u>ystemie?</u>			11
an jest dopuszczainy format			

ECLA Classification ...

biotechnology belongs to class C (chemistry; metallurgy)

Wyszukaj za pomocą klasyfikacji europejskiej

	Znajdź symbole klasyfikacji dla	słów kluczowych		Wyszukuj	
		Idź Index A B C D E F G H Y		Następny: /	4
HU	MAN NECESSITIES			A	
PE	REORMING OPERATIONS; T	RANSPORTING		B	
СН	IEMISTRY; METALLURGY			C 🛛	
TE	XTILES; PAPER			D	
FD	KED CONSTRUCTIONS			E	
ME PU	CHANICAL ENGINEERING; L	IGHTING; HEATING; WEAPONS; BLASTING ENGINES OF	٤	F	
PH	YSICS			G	
EL	ECTRICITY			н	
GE	NERAL TAGGING OF NEW T	ECHNOLOGICAL DEVELOPMENTS [N0403]		Υ	
	🗖 pokaż uwagi	Kopiuj do formularza poszukiwań:	Kopiuj	Wyczyść	

CLASSES – letters A-H + Y

Biotechnology = subclass C12

ORGANIC CHEMISTRY (such compounds as the oxides, sulfides, or oxysulfides of carb cyanogen, phosgene, hydrocyanic acid or salts thereof C01; products obtained from laye base-exchange silicates by ion-exchange with organic compounds such as ammonium, phosphonium or sulfonium compounds or by intercalation of organic compounds C01B33, macromolecular compounds C08; dyes C09; fermentation products C12; fermentation or enzyme-using processes to synthesise a desired chemical compound or composition or optical isomers from a racemic mixture C12P; production of organic compounds by elect	on, ered /44; or to separate trolysis or	C07 🗖	
electrophoresis C25B3/00, C25B7/00) ORGANIC MACROMOLECULAR COMPOUNDS; THEIR PREPARATION OR CHEMICA WORKING-UP; COMPOSITIONS BASED THEREON (manufacture or treatment of ar threads, fibres, bristles or ribbons D01) [C9410]	L tificial	C08 🗖	
DYES; PAINTS; POLISHES; NATURAL RESINS; ADHESIVES; MISCELLANEOUS		C09 🔲	
PETROLEUM, GAS OR COKE INDUSTRIES; TECHNICAL GASES CONTAINING CARE	BON	C10 🗖	
ANIMAL AND VEGETABLE OILS, FATS, FATTY SUBSTANCES AND WAXES; FATTY THEREEROM: DETERGENTS: CANDLES (edible oil or fat compositions A23)	ACIDS	C11 🔳	
BIOCHEMISTRY; BEER; SPIRITS; WINE; VINEGAR; MICROBIOLOGY; ENZYMOLOGY MUTATION OR GENETIC ENGINEERING	/;	C12 🗖	
SUGAR OR STARCH INDUSTRY (polysaccharides, e.g. starch, derivatives thereof C08B	; malt	C13 🔳	
SKINS; HIDES; PELTS; LEATHER		C14 🔲	
METALLURGY OF IRON		C21 🔳	
METALLURGY (of iron C21); FERROUS OR NON-FERROUS ALLOYS; TREATMENT OF OR NON-FERROUS METALS (production of metals by electrolysis or electrophoresis C2	F ALLOYS 25)	C22 🗖	
COATING METALLIC MATERIAL; COATING MATERIAL WITH METALLIC MATERIAL metallising textiles D06M11/83; decorating textiles by locally metallising D06Q1/04); CH SURFACE TREATMENT; DIFFUSION TREATMENT OF METALLIC MATERIAL; COATIN VACUUM EVAPORATION, BY SPUTTERING, BY ION IMPLANTATION OR BY CHEMIC VAPOUR DEPOSITION, IN GENERAL (for specific applications, see the relevant places,	(by IEMICAL G BY AL e.g. for	C23 🗖	
manufacturing resistors H01C17/06); INHIBITING CORROSION OF METALLIC MATER INCRUSTATION IN GENERAL (treating metal surfaces or coating of metals by electroly electrophoresis C25D, C25F)	SUBCLASSES	– n	umbers

More precise calssification - subsubclasses

INE; OTHER ALCOHOLIC BE	/ERAGES; PREPARATION THEREOF (beer C12C)	C12G
STEURISATION; STERILISA	TION; PRESERVATION; PURIFICATION; CLARIFICATION;	C12H
SEING		
NEGAR; ITS PREPARATION		C12J 📃
TCHING OR DEPITCHING MA	CHINES; CELLAR TOOLS (cleaning of casks B08B9/00)	C12L
PARATUS FOR ENZYMOLO	GY OR MICROBIOLOGY (installation for fermenting manure	C12M
1C3/02 ; preservation of living	parts of humans or animals A01N1/02 ; physical or chemical	
paratus in general B01; maltin	g or mashing apparatus C12C1/00 ; brewing apparatus C12C13/00 ;	
mentation apparatus for wine (12G; apparatus for preparing vinegar C12J1/10)	\sim
CRO-ORGANISMS OR ENZY	MES; COMPOSITIONS THEREOF (biocides, pest repellants or	C12N 🗷
ractants, or plant growth regul	ators, containing micro-organisms, viruses, microbial fungi,	
zymes, fermentates or substa	nces produced by or extracted from micro-organisms or animal	
aterial A01N63/00 ; food comp	sitions A21, A23; medicinal preparations A61K; chemical aspects	
or use of materials for, banda	jes, dressings, absorbent pads or surgical articles A61L; fertilisers	
5); PROPAGATING, PRESER	VING OR MAINTAINING MICRO-ORGANISMS (preservation of	
ng parts of humans or animals	A01N1/02); MUTATION OR GENETIC ENGINEERING; CULTURE	
DIA (micro-biological testing i		
		C12P
XTUDE (formantation process	v OR TO SEPARATE OPTICAL ISOMERS FROM A RACEMIC	
	es to form a food composition A21, A25, compounds in general,	
e relevant compound class le c	C01 C07 : brewing of beer C12C : producing vinegar C12U	
cesses for producing enzyme	s C12N9/00 · DNA or RNA concerning genetic engineering vectors	
a plasmids or their isolation	preparation or purification C12N15/00)	
ASURING OR TESTING PRO	CESSES INVOLVING ENZYMES OR MICRO-ORGANISMS	C120 🔲
munoassay G01N33/53); COI	IPOSITIONS OR TEST PAPERS THEREFOR: PROCESSES OF	
REPARING SUCH COMPOSIT	ONS: CONDITION RESPONSIVE CONTROL IN	
CROBIOLOGICAL OR ENZYM	OLOGICAL PROCESSES	
OCESSES USING MICRO-O	RGANISMS	C12R 🔲

SUBSUBCLASSES –letters A-Z

Kontakt z EPO Zmiana kraju 🔸				🗗 Drukuj	÷			Data pierwszeństwa 2007-11-21		Data pierwszeństwa 2005-02-11	VITAMIN C	Data pierwszeństwa 2006-06-07
	I	l		wsze strony (0)		Sort	NTENT IN CREASED	Informacja o publikacji: US2009133157 (A1) 2009-05-21		Informacja o publikacji: US2010248315 (A1) 2010-09-30 US8318462 (B2) 2012-11-27	MPOSITION WITH HIGH	Informacja o publikacji: W02007141420 (A2)
				e Pobierz pierv		> 6	AMIN C.CO	MKP: A01H5/00 A01H5/08 C12N15/82		MKP: C12N1/21 C12P17/04	ICE, AND C	MKP:
	·	a Pomoc		N XLS)	la: ssification	wania Descendir	ANT WITH VIT	CPC: C12 N15/8242		CPC: <u> C12 N9/0008</u> C12 P17/04	RUS FRUIT JUI	CPC: A23L1/302 A23L2/06
		zapytan Ustawieni	0	me tytuły 🕒 Eksport (CS	w bazie danych Worldwide d as the Cooperative Patent Clai	alizacji 🔌 Porządek sorto alizacji	VSZENSTWB APSICUM PL	CY :: INCORP ADM AGENCY NAT L AGR & [JP]	C production	Zgłaszający: DSM IP ASSETS B.V	ING AND PRO <mark>c</mark> essing <mark>c</mark> h	Zgłaszający: TERRAS JEAN-LOUIS [FR]
spacenet szukiwanie patentów	: EPO i UPRP 🔸	az patentow (U) Historia	YKAZ WYNIKÓW 🖥	Wybierz wszystko 📃 Sa	Znaleziono około 43 wyników vitarnin C w tytule AND C12 i	Sortuj według Data aktu	1. METHOD OF F Tworca	Twórca: CPC TSUDA SHINYA [JP]	2. Fermentative vitamin.	Twórca: BEUZELIN-OLLIVIER MARIE-GABRIELLE [CH] CHEVREUX BASTIAN [DE] (+13)	3. METHOD FOR EXTRA CI	Twórca: TERRAS JEAN-LOUIS [FR]
	 4 0 Espacenecie Inne usługi online 	Wyszukuj Wynik 📉 Moj wyka <u>Uściślij wyszukiwanie</u> → Strona z wynikam	Inteligentne wyszukiwanie W	Wyszukiwanie zaawansowane Wyszukiwanie wg ECLA	Zybka Pomoc -		 Wykazem (lista) wynikow? Czy mogę wykonać eksport 	wyników z listy?	→ Dlaczego lista jest ograniczona do 500 wyników?	 Can I deactivate the highlighting? Diaczego niektóre dokumenty nie sa wyświetlane na liście wywików? Czy można sortować liste wyników? Co spowoduje klikniecie ikony 	q <u>wiazdki?</u> + <u>Co to jest dokument XP?</u> + Czv można zapamietać zadane	<u>pytanie?</u> Xobacz również +

"Bioinformatic" websites



EMBL-EBI

- NCBI National Center for Biotechnology Information (USA)
- GenBank; RefSeq, PubMed; OMIM
- BLAST...
- EBI European Bioinformatics Institute
 - ENA; UniProt; ArrayExpress
 - EuroPubMed



ExPASy

- UniProt; PROSITE; 2D-page results
- Great source of proteins & proteomics tools (e.g. SWISS Model)

$\mathsf{ExPASy:}\ \mathsf{proteomic}\ \mathsf{server} \to \mathsf{bioinformatics}\ \mathsf{resource}$

				Fiorne. 7	toour con
	Query all databases	×	search he		
isual Guidance	ExPASV is the SIB Bioinformatics Resource Po	rtal which provides access to s	cientific databases	Popular resources	
ategories	and software tools (i.e., resources) in differen	nt areas of life sciences incl	luding proteomics,	O. Halbanko	
roteomics	genomics, phylogeny, systems biology, population the left menu). On this portal you find resources t	1 genetics, transcriptomics etc.	(see Categories in		
enomics	institutions.	rom many unicrem oib groups	as well as external	STRING	
tructural bioinformatics				PROSITE	
systems biology	Featuring today				
hylogeny/evolution	· · · · · · · · · · · · · · · · · · ·				
opulation genetics	TCS			Latest News	
ranscriptomics	Predict interaction specificity in bact signalling (two-component systems	erial s)		MetaNetX is available -	2013-02-22
lophysics	[details]	111		A new Systems Biology w	ebsite ca
maging					1
Linfractructura	•••••	\sim		EasyProt is available - 2	013-02-06
minastructure				A new mass spectrometry	tool calle
frug design				EasyProt has been added	

ExPASy offers an easy access to numerous bioinformatic tools

www.expasy.org/	proteomics	
	Query all databases	× search help
'isual Guidance	SIB resources	
ategories	Databases	Tools
proteomics		
protein sequences and identification	UniProtKB • functional information on proteins • [more]	SWISS-MODEL Workspace • structure homology-modeling • Imprel
nass spectrometry and 2-DE data	UniProtKB/Swiss-Prot • protein sequence database • [more]	SwiceDock - protein ligand decking conver - [more]
rotein characterisation and function	STRING • protein-protein interactions • [more]	Swissbock - broten udang gocknig server - [unde]
amilies, patterns and profiles	SWISS-MODEL Repository • protein structure homology	771P • Prediction of leucine zinner domains • [more]
post-translational modification	DDOOTE antheir damains and families. Inseed	30f5 • find user-defined natterns in protein sequences • [more]
protein structure	PROSITE • protein domains and families • [more]	AACompleant + protein identification by as
rotein-protein interaction	ViralZone • portal to viral UniProtKB entries • [more]	composition • [more]
imilarity search/alignment	a nextProt • human proteins • [more]	AACompSim • amino acid composition comparison • [more]
enomics		Agadir • Prediction of the helical content of peptides • [more]
tructural bioinformatics	Courses • [more]	ALF • simulation of genome evolution • [more]
ystems biology	ENZYME • enzyme nomenclature • [more]	Alignment tools • Four tools for multiple alignments • [more]
hylogeny/evolution	GlycoSuiteDB • glycan database • [more]	AllAll • protein sequences comparisons • [more]
	GPSDB • gene and protein synonyms • [more]	APSSP • Advanced Protein Secondary Structure
opulation genetics	HAMAP • UniProtKB family classification and annotation	Prediction • [more]
anscriptomics	• [more]	Ascalaph • Molecular modeling software • [more]
ophysics	📥 MetaNetX • Metabolic Network Repository & Analysis • [more]	
naging	MIAPEGeIDB • MIAPE document edition • [more]	🐣 Biochemical Pathways • Biochemical Pathways • [more]
infractructure	MyHits • protein domains database and tools • [more]	📥 BLAST • sequence similarity search • [more]
	ANDITplus • protein families and domains resources • [more]	BLAST (UniProt) • BLAST search on the UniProt web
rug design	🔼 PaxDb • protein abundance database • [more]	site = [more]
esources AZ	Prolune • Popular science articles (in French) • [more]	BLAST - NCBI • Biological sequence similarity search • [more
nks/Documentation	Protein Model Portal • structural information for a protein • [more]	BLAST - PBIL • BLAST search on protein sequence databases • [more]
	Protoin Conflict - Informally written reviews on	Blast2Fasta • Blast to Fasta conversion • [more]



UniProt is the most comprehensive, centralized protein sequence catalog, it consists of 3 databases:

- Swiss-Prot best-annotated protein database at Swiss Institute of Bioinformatics
- The Translated EMBL (TrEMBL) Nucleotide Sequence Database Library provides automated annotations of proteins not in Swiss-Prot
- PIR (Protein Information Resource) Protein Sequence Database curated by experts from Georgetown University







UniProt – protein sequence database

🗲 🛞 www.uniprot.org			≂ C 🛐 × o	braz 👂 🏠 💼 🖊 🏫	6
UniProt	UniProtKB -			Advanced - Q Help Cont	tact
Welcome to the new UniF	Prot website! We hope you e	njoy the new design. If yo	u're not quite ready yet, you	can still go back to the old site.	
S The mission of UniProt i	is to provide the scientifi al information.	c community with a c	omprehensive, high-qualit	y and freely accessible resource of proteir	n
UniProtKB Swiss-Prot (546,238)	UniRef Sequence clusters	UniParc Sequence archive	Proteomes	News I its own game New human variant types available on the FTP site UniProt release 2014_08	N
annotated and reviewed.	S	Supporting dat	а	Lark or owl? PER3 is the answer Cross-references to CCDS and	
TrEMBL (82,126,897) Automatically	Literature citations	Taxonomy	Subcellular locations	GeneReviews UniParc cross- references with protein and gene names FTP UniProt release 2014_07	-
annotated and not reviewed.	Cross-ref. databases	XXX XXX	Keywords	Rews archive	×.



Lecture 1

Introduction to bioinformatics

- Basic definitions
- □ Scope of the course

Internet bio-data resources

- **B**EBI
- □ ExPASy
- Sanger Institute
- □ KEGG
- BRENDA

http:// www.sanger.ac.uk

Home Research Scientific resources Work	About us
Mouse Zebrafish Data Software Da	atabases Technologies Talks & training
Scientific resources	
As part of its own research or as part of colla DNA sequences, develops databases such as	aborative projects, the Wellcome Trust Sanger Institute produces raw data such a Ensembl, and generates biological resources such as embryonic stem cells.
onn sequences, acreiops databases such as	ensense, and generates biological resources such as emptyonic stelli tells.
These are available from these pages with the min collaboration. You can also access a catalogue of o	nimum of conditions. You can find out how to access our technologies: most often this is thro our research papers and other science publications.
These are available from these pages with the min collaboration. You can also access a catalogue of o	nimum of conditions. You can find out how to access our technologies: most often this is thro our research papers and other science publications.
These are available from these pages with the min collaboration. You can also access a catalogue of o Quick access to resources in the fol	nimum of conditions. You can find out how to access our technologies: most often this is thro our research papers and other science publications. Ilowing areas:
These are available from these pages with the min collaboration. You can also access a catalogue of o Quick access to resources in the fol > <u>Cancer genome project</u>	nimum of conditions. You can find out how to access our technologies: most often this is thro our research papers and other science publications. Ilowing areas:
These are available from these pages with the min collaboration. You can also access a catalogue of o Quick access to resources in the fol > Cancer genome project > Mouse resources	nimum of conditions. You can find out how to access our technologies: most often this is thro our research papers and other science publications. llowing areas:
These are available from these pages with the min collaboration. You can also access a catalogue of or Quick access to resources in the fol > <u>Cancer genome project</u> > <u>Mouse resources</u> > <u>Zebrafish resources</u> > <u>Xenopus tropicalis resources</u>	nimum of conditions. You can find out how to access our technologies: most often this is thro our research papers and other science publications. Ilowing areas:
These are available from these pages with the min collaboration. You can also access a catalogue of or Quick access to resources in the fol > <u>Cancer genome project</u> > <u>Mouse resources</u> > <u>Zebrafish resources</u> > <u>Xenopus tropicalis resources</u>	nimum of conditions. You can find out how to access our technologies: most often this is thro our research papers and other science publications.
These are available from these pages with the min collaboration. You can also access a catalogue of or Quick access to resources in the fol > <u>Cancer genome project</u> > <u>Mouse resources</u> > <u>Zebrafish resources</u> > <u>Xenopus tropicalis resources</u> Software	himum of conditions. You can find out how to access our technologies: most often this is thro our research papers and other science publications. Ilowing areas: Databases
These are available from these pages with the min collaboration. You can also access a catalogue of or Quick access to resources in the fol > <u>Cancer genome project</u> > <u>Mouse resources</u> > <u>Zebrafish resources</u> > <u>Zebrafish resources</u> > <u>Xenopus tropicalis resources</u> Software Manual Institute software programs developed	himum of conditions. You can find out how to access our technologies: most often this is thro our research papers and other science publications. Ilowing areas: Patabases Collections of analysed data, such as Projects
These are available from these pages with the min collaboration. You can also access a catalogue of or Quick access to resources in the fol . <u>Cancer genome project</u> . <u>Mouse resources</u> . <u>Zebrafish resources</u> . <u>Zebrafish resources</u> . <u>Xenopus tropicalis resources</u> . <u>Software</u> . Institute software programs developed for genome analysis, such as Artemis for bacterial expenses as well as DMA	himum of conditions. You can find out how to access our technologies: most often this is thro our research papers and other science publications. Ilowing areas: Collections of analysed data, such as the Ensembl genome browser and the DECURPTION databases of characteristics
These are available from these pages with the min collaboration. You can also access a catalogue of or Quick access to resources in the fol > <u>Cancer genome project</u> > <u>Mouse resources</u> > <u>Zebrafish resources</u> > <u>Zebrafish resources</u> > <u>Xenopus tropicalis resources</u> > <u>Software</u> Institute software programs developed for genome analysis, such as Artemis for bacterial genomes, as well as DNA sequence production and file formats	himum of conditions. You can find out how to access our technologies: most often this is thro our research papers and other science publications. Ilowing areas:

ARNIE	BLAST
Expression	The sequencing projects' Blast Search BLAST Services
OSMIC	DECIPHER
Database of somatic mutation information in human cancers	Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources
NCODE	GUono
Aims to identify all functional elements in the human genome sequence	Annotation from a growing number of organisms is available from the GeneDB web database
nsembl Genome Browser	
Ensembl produces and maintains automatic	MEROPS
annotation on selected eukaryotic genomes	Mr.Kors Provides the internationally recognised classification of peptidases and their inhibitors
LIDERS	avout
Genome-wide linkage disequilibrium repository and search engine	Rfam
	Provides a classification of RNA families - Rfom using covariance models
Pfam	more

GLIDE	RS	1 BINI
U	Genome-wide linkage disequilibrium repository and search engine	Rfam Provides a classification of RNA families Rfam using covariance models
iPfam		Tone
iPfam	Describes Pfam domain interactions that are observed in PDB entries	TreeFam Provides curated phylogenetic trees of animal names that nive reliable ortholog
Pfam		and paralog assignments
Pfam	Provides a classification of proteins into families and domains using hidden Markov models	Wormbase The primary database on the biology and genome of the model organism
Tiffin		Caenorhabditis elegans
	A database of predicted regulatory motifs, a subset with predicted functional annotation	ZF-Models Use of zebrafish to make models for human disease and development
VEGA	Genome Browser	Thore
Vega	Central database repository for high quality manual annotation of vertebrate finished genome sequence	

http:// www.sanger.ac.uk

sanger	Search	www.sanger.ac.uk 🛐 🖲	for Enter search here
Home Research Scientific resources	Work & study About us		A
Mouse Zebrafish Data Software	Databases Technologies	Talks & training	
Colontific recourses			
Scientific resources			
As part of its own research or as part o	f collaborative projects, the W	ellcome Trust Sanger Instit	tute produces raw data such as
UNA sequences, develops databases su	ch as Ensembl, and generates	biological resources such a	s embryonic stem cells.
These are available from these pages with the collaboration. You can also access a catalogue	ne minimum of conditions. You can ue of our research papers and othe	find out how to access our tec er science publications.	chnologies: most often this is throu
These are available from these pages with the collaboration. You can also access a catalogue	ne minimum of conditions. You can ue of our research papers and othe	find out how to access our teo ar science publications.	chnologies: most often this is throu
These are available from these pages with the collaboration. You can also access a catalogu Quick access to resources in the second sec	ne minimum of conditions. You can ue of our research papers and othe e following areas:	find out how to access our teo ar science publications.	chnologies: most often this is throu
These are available from these pages with the collaboration. You can also access a catalogue Quick access to resources in the > Cancer genome project	ne minimum of conditions. You can ue of our research papers and othe e following areas:	find out how to access our teo er science publications.	hnologies: most often this is throu
These are available from these pages with the collaboration. You can also access a catalogue Quick access to resources in the Cancer genome project	ne minimum of conditions. You can ue of our research papers and othe e following areas:	find out how to access our tec er science publications.	chnologies: most often this is throu
These are available from these pages with the collaboration. You can also access a catalogue Quick access to resources in the <u>Cancer genome project</u> <u>Mouse resources</u> <u>Zebrafish resources</u> <u>Xenopus tropicalis resources</u> 	ne minimum of conditions. You can ue of our research papers and othe e following areas:	find out how to access our teo er science publications.	hnologies: most often this is throu
These are available from these pages with the collaboration. You can also access a catalogue Quick access to resources in the <u>Cancer genome project</u> <u>Mouse resources</u> <u>Zebrafish resources</u> <u>Xenopus tropicalis resources</u> 	ne minimum of conditions. You can ue of our research papers and othe e following areas:	find out how to access our tec er science publications.	hnologies: most often this is throu
These are available from these pages with the collaboration. You can also access a catalogue Quick access to resources in the <u>Cancer genome project</u> <u>Mouse resources</u> <u>Zebrafish resources</u> <u>Xenopus tropicalis resources</u> Software	ne minimum of conditions. You can ue of our research papers and othe e following areas: Databases	find out how to access our tec er science publications.	chnologies: most often this is throu Related Links
These are available from these pages with the collaboration. You can also access a catalogue Quick access to resources in the <u>Cancer genome project</u> <u>Mouse resources</u> <u>Zebrafish resources</u> <u>Xenopus tropicalis resources</u> Software Qetext(1) Institute software programs developed 	e minimum of conditions. You can ue of our research papers and othe e following areas: Databases oped Collections	find out how to access our teo er science publications. s of analysed data, such as	chnologies: most often this is throu Related Links <u>Projects</u>
These are available from these pages with the collaboration. You can also access a catalogue Quick access to resources in the <u>Cancer genome project</u> <u>Mouse resources</u> <u>Zebrafish resources</u> <u>Xenopus tropicalis resources</u> Software Institute software programs developed for genome analysis, such as Arter 	oped mis	find out how to access our tec er science publications. s of analysed data, such as abl genome browser and the	chnologies: most often this is throu Related Links <u>Projects</u> <u>Academic Faculty</u>
These are available from these pages with the collaboration. You can also access a catalogue Quick access to resources in the <u>Cancer genome project</u> <u>Mouse resources</u> <u>Zebrafish resources</u> <u>Zebrafish resources</u> <u>Xenopus tropicalis resources</u> Software Institute software programs develop for genome analysis, such as Arter for bacterial genomes, as well as I sequence production and file form 	oped mis DNA ats	find out how to access our tec er science publications. s of analysed data, such as abl genome browser and the database of chromosome ments	chnologies: most often this is throu Related Links Projects Academic Faculty Work and study
These are available from these pages with the collaboration. You can also access a catalogue Quick access to resources in the <u>Cancer genome project</u> <u>Mouse resources</u> <u>Zebrafish resources</u> <u>Zebrafish resources</u> <u>Xenopus tropicalis resources</u> Software Magger Institute software programs develor for genome analysis, such as Arter for bacterial genomes, as well as the sequence production and file form	oped mis DNA ats ore b	find out how to access our tec er science publications. s of analysed data, such as abl genome browser and the database of chromosome ments more	Related Links Projects Academic Faculty Work and study The Genome Campus

Model organisms

- BACTERIA: Escherichia coli K-12
 4.6 Mb genome sequenced in 1997
- YEAST: Saccharomyces cerevisiae (12.5 Mb)
 First Eukaryotic genome sequenced (in1996)
- PLANT: Arabidopsis thaliana (125 Mb)
 First plant (second Eukaryotic) genome sequenced in 2000
- FRUITFLY: Drosophila melanogaster (137 Mb) model for genetics; sequencing finished in 2000
- NEMATODE: Caenorhabditis elegans (100.2 Mb) first multicellular animal sequenced (in 1998), exacly 595 somatic cells, transparent body
- FISH: Danio rerio (1.8Gb !) short generation time, large amount of progeny, well annotated genes
- MAMMAL: Mus musculus sequenced in 2002 (3 Gb) Rodents diverged from primetes ~80MYA and share most genes with humans

Example of model organism webpage

Home Genome Blast / Blat WormMart Batc	h Sequences Markers Genetic Maps Submt Searches Site M
Wormbase Release WS 198	NormBase 3 Cal
Sanger WormBase Mirror . Master is at www.wormbase.org	
Find: Any Gene	Search
Exact match Results	as XML 🔲 Literature Search 🗹 Wormbase Suggest
Web Site Directory	News and Notes
Release Notes New/Changed Genes, release notes	October 25, 2008: New release of WormBase: WS195
General Searches WormBase Class Browser, Wormbase Query Language Search, AQL Search	 October 08, 2008: Martin Chalfie wins Nobel Prize for GFP in C. elegans Martin Chalfie (with Osamu Shimomura and Roger Y. Tsien) has won the 2008 Nobel Prize in Chemistry for pioneering the use of GFP in C. elegans. This is t third Nabel Prize and the second sec
Gene, Blast / Blat, e-PCR, Gene Ontology, Synteny Viewer, Cis-Elements (CisOrtho),	 September 26, 2008: Award-winning C. elegans in the last six years. Victor Ambros and Gary Ruvkun (with David Baulcombe) have won the Albert Lasker Basic Medical Research Award for work on small regulatory RNAs, wh
Cells and Gene Expression Cell and Pedigree, Neurons, Expression Pattern, Expression profile	Susan Mango has won a MacArthur Fellowship for work on the genomics of organogenesis.
Genetics, Strains, and Phenotypes Genetic interval, Rearrangements, Clone, Allele, SNPs, Markers, and Strains,	 Old News. Old News.
Strain Veport Ubenetypes UNA	

Mouse resources

This resource page gives one-stop access to the Sanger Institute's DNA, genetic and biological resources for mouse researchers.

generating and characterising mouse mutants and is dedicated to making these resources freely The Sanger Institute plays a lead role in sequencing mouse genomes, discovering variants, available to researchers.

- Access by resource programme Our mouse resources collected by their programme, such as gene trapping.
- Access by resource type Our mouse resources collected by type of material, such as genomes, vectors, mutant ES cells, and mice.
- Mouse research laboratories Six of the Institute's research teams focus on the mouse as a model for human disease.



[Wellcome Library, London]

Related Links

Mouse resource portal



Zebrafish resources Academic Faculty Projects gene targeting vectors, targeted embryonic stem (ES) cells, mutant mouse lines, and phenotypic Unified access to major mouse resources from the Institute or its collaborators, including BACs, data.

A anom

Cancer genome project The Genome Campus

What we do

Access by resource programme

- High throughput gene targeting A high-throughout gene targeting pipeline to generate conditional targeted mutations in C57BL/6 ES cells.
- Sanger Institute Gene Trap Resource (SIGTR) More than 10,000 characterised lacZ-tagged insertional mutations in 129 ES cells.
- Mutagenic Insertion and Chromosome Engineering Resource (MICER) A genome-wide collection of targeting vectors for chromosome engineering and gene-targeting.
- > Email contacts for more information about mouse resources.

Access by resource type

Mouse lines and phenotype data

- Mouse cell lines and phenotype data are available from our Mouse resource portal.
- > Request a mouse strains from the Sanger Institute Mouse Genetics Programme

Mutant FS call lines

KEGG

Kyoto **E**ncyclopedia of **G**enes and Genomes

KEGG Home Introduction

> Overview Release notes Current statistics

KEGG XML

KEGG API

KEGG Identifiers

Pathway maps

Brite hierarchies

Extremely useful for methabolic pathways – focused searches



Japanese

KEGG: Kyoto Encyclopedia of Genes and Genomes

A grand challenge in the post-genomic era is a complete computer representation of the cell, the organism, the ecosystem, and the biosphere, which will enable computational prediction of higher-level complexity of cellular processes and organism behaviors from genomic and molecular information. Towards this end we have been developing a bioinformatics resource named KEGG as part of the research projects of the Kanehisa Laboratories in the Bioinformatics Center of Kyoto University and the Human Genome Center of the University of Tokyo.

KEGG FTP	2 11 11 11 11 11 11 11 11 11 11 11 11 11	
VonToole	 Main entry point to 	the KEGG web service
Regioois	KEGG2	KEGG Table of Contents Update notes Help
	🚽 🖉 Data-oriented entry	points
GenomeNet	KEGG PATHWAY	Pathway maps and pathway modules Pathway maps
	KEGG BRITE	Functional hierarchies and ontologies Brite hierarchies
DBGET/LinkDB	KEGG DISEASE	Human diseases Disease classification
Feedback	KEGG DRUG	Drugs ATC drug classification
	KEGG ORTHOLOGY	KO system and ortholog annotation KO system
	KEGG GENES	Genes and proteins
	KEGG GENOME	Genomes KEGG organisms
	KEGG COMPOUND	Chemical compounds Compound classification
	KEGG GLYCAN	Glycans
-	KEGG REACTION	Reactions
	🥥 Organism-specific e	ntry points
	KEGG Organisms	Select Organism Go (example) has
	🥥 Analysis tools	
	KEGG Mapper New!	KEGG PATHWAY and BRITE mapping tools
	KEGG Atlas	Navigation tool to explore KEGG global maps
	KAAS	KEGG automatic annotation server
	BLAST/FASTA	Sequence similarity search
	SIMCOMP	Chemical structure similarity search
	PathPred	Biodegradation/biosynthesis pathway prediction

Besides searches with key words you can choose a pathway from the list...

1. Metabolism

1.1 Carbohydrate Metabolism

Glycolysis / Gluconeogenesis Citrate cycle (TCA cycle) Pentose phosphate pathway Pentose and glucuronate interconversions Fructose and mannose metabolism Galactose metabolism Ascorbate and aldarate metabolism Starch and sucrose metabolism Amino sugar and nucleotide sugar metabolism Pvruvate metabolism Glyoxylate and dicarboxylate metabolism Propanoate metabolism Butanoate metabolism C5-Branched dibasic acid metabolism Inositol phosphate metabolism

1.2 Energy Metabolism

Oxidative phosphorylation Photosynthesis Photosynthesis - antenna proteins Carbon fixation in photosynthetic organisms Reductive carboxylate cycle in photosynthetic bacteria Methane metabolism Nitrogen metabolism Sulfur metabolism

1.3 Lipid Metabolism

Fatty acid biosynthesis Fatty acid elongation in mitochondria Fatty acid metabolism Synthesis and degradation of ketone bodies Steroid biosynthesis Primary bile acid biosynthesis Secondary bile acid biosynthesis

Enzymes Compounds with biological roles

Photosynthesis proteins

Lipids Lipid biosynthesis proteins

Green boxes = enzymes active in selected organism



BRENDA at the Institute of Biochemistry and Bioinformatics at the Technical University of Braunschweig, Germany

iinche reität achweig
achivelg
ed by
,
arameter
arameter
Parameter

Enzyme name search



Record in BRENDA may be limitted to organism of interest, there are some text minning tools available



Information on EC 1.1.1.1 - alcohol dehydrogenase:

			PRINT
Mark a special word or phrase in this record:	Mark!		
Select one or more organisms in this record:	All organisms		
	Acetobacter pasteurianus		
	Acinetobacter calcoaceticus		
	Aeropyrum pernix		
	Alligator mississippiensis	Ŧ	Submit



EC NUMBER COMMENTARY

1.1.1.1

RECOMMENDED NAME GeneOntology No.







EC 1.1.1.1 - alcohol dehydrogenase

PubMed	PubMed
Thermoplasma acidophilum	Homo sapiens
2009	2008
61-67	65-75
42	245
Mol. Biotechnol.	Toxicology
Heterologous expression and characterization of an alcohol dehydrogenase from the archeon Thermoplasma acidophilum	Oxidation of alcohols and reduction of aldehydes derived from methyl- and dimethylpyrenes by cDNA-expressed human alcohol dehydrogenases
Marino-Marmolejo, E.N.; De Leon-Rodriguez, A.; de la Rosa, A.P.; Santos, L.	Kollock, R.; Frank, H.; Seidel, A.; Meinl, W.; Glatt, H.
700124	701322

LINKS TO OTHER DATABASES (specific for EC-Number 1.1.1) EXPASY KEGG MetaCVC MetaCVC NCBI: PubMed, Protein, Nucleotide, Structure, Genome, OMIM, Domains UBMB Enzyme Nomenclature Protein Nutant Database of protein families and domains Protein Mutant Database InterPro (database of protein families, domains and functional sites)

Take-home messages

- Bioinformatics help scientists to understand their laboratory results and to keep their research up to date
- Biological data is collected in databases, most of them are freely available online
- Data in bioinformatics may be: scientific publications, sequences & structures of macromolecules, organism – focused information, metabolic pathways etc...
- There are bioinformatics websites enabling the access to all these kinds of information (such as NCBI, EBI and ExPASy)
- Metabolic information is gathered by such databases as KEGG and MetaCyc; specialized enzymatic database is called BRENDA


Thank you for your attention

NEXT WEEK:

PRACTICAL 1: LITERATURE SEARCH Mon 06.Oct 8:30; WED 08.Oct 11:15**LECTURE 2:**PAIRWISE SEQUENCE ALIGMENT Tue, 07.Oct, 12:15