KEGG

(Kyoto Encyclopedia of Genes and Genomes)

京都基因与基因组百科全书

郭静

KEGG

GenomeNet

Bioinformatics Center
Institute for Chemical Research
Kyoto University

English 日本語

http://www.genome.jp/

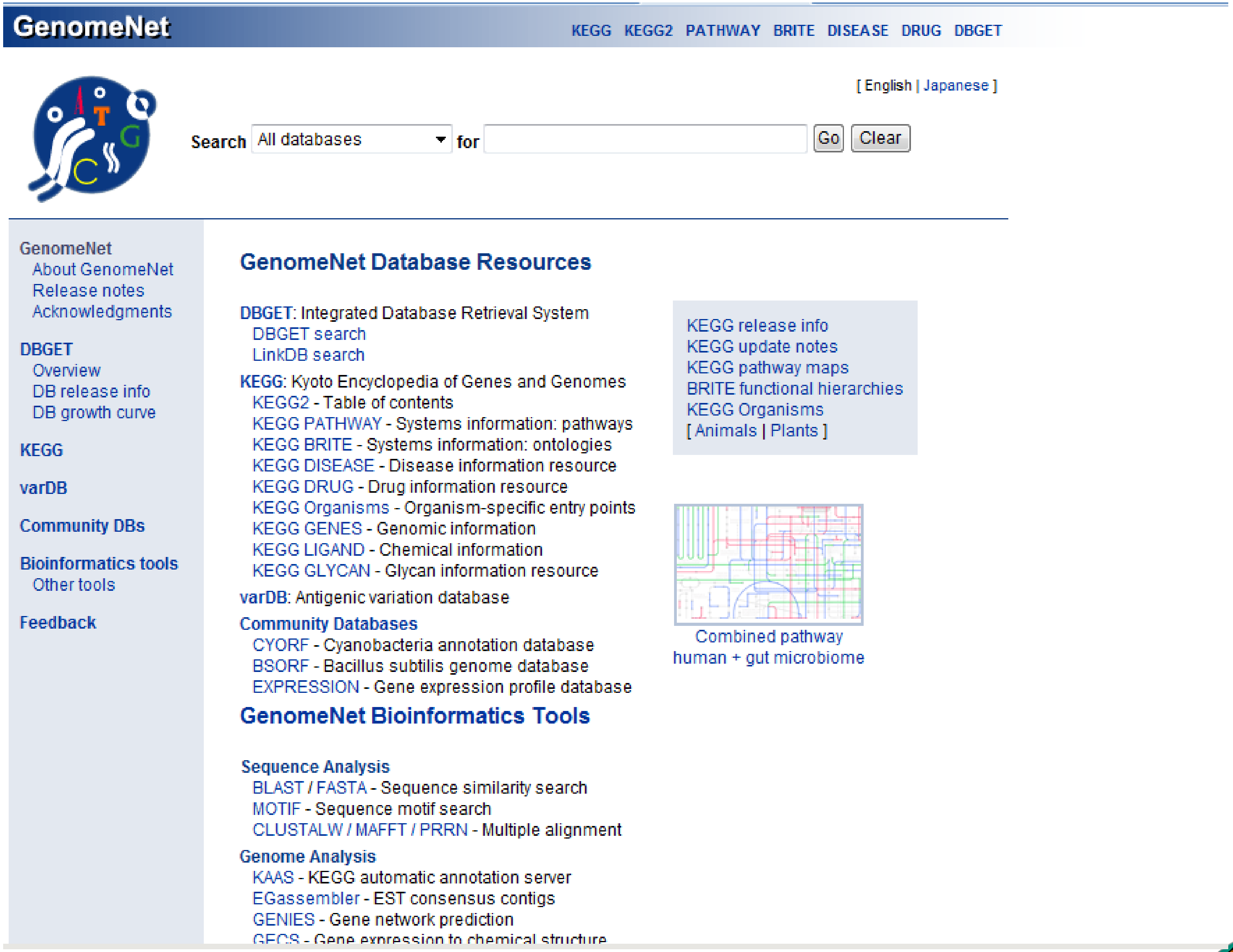
Kyoto Encyclopedia of Genes and Genomes

访问地址：

http://www.genome.jp/kegg/

KEGG FTP：

ftp://ftp.genome.jp/pub/kegg/



GenomeNet
  About GenomeNet
  Release notes
  Acknowledgments

DBGET
  Overview
  DB release info
  DB growth curve

KEGG

varDB

Community DBs

Bioinformatics tools
  Other tools

Feedback

GenomeNet Database Resources

DBGET: Integrated Database Retrieval System
  DBGET search
  LinkDB search

KEGG: Kyoto Encyclopedia of Genes and Genomes
  KEGG2 - Table of contents
  KEGG PATHWAY - Systems information: pathways
  KEGG BRITE - Systems information: ontologies
  KEGG DISEASE - Disease information resource
  KEGG DRUG - Drug information resource
  KEGG Organisms - Organism-specific entry points
  KEGG GENES - Genomic information
  KEGG LIGAND - Chemical information
  KEGG GLYCAN - Glycan information resource

varDB: Antigenic variation database

Community Databases
  CYORF - Cyanobacteria annotation database
  BSORF - Bacillus subtilis genome database
  EXPRESSION - Gene expression profile database

KEGG release info
KEGG update notes
KEGG pathway maps
BRITE functional hierarchies
KEGG Organisms
[Animals | Plants]

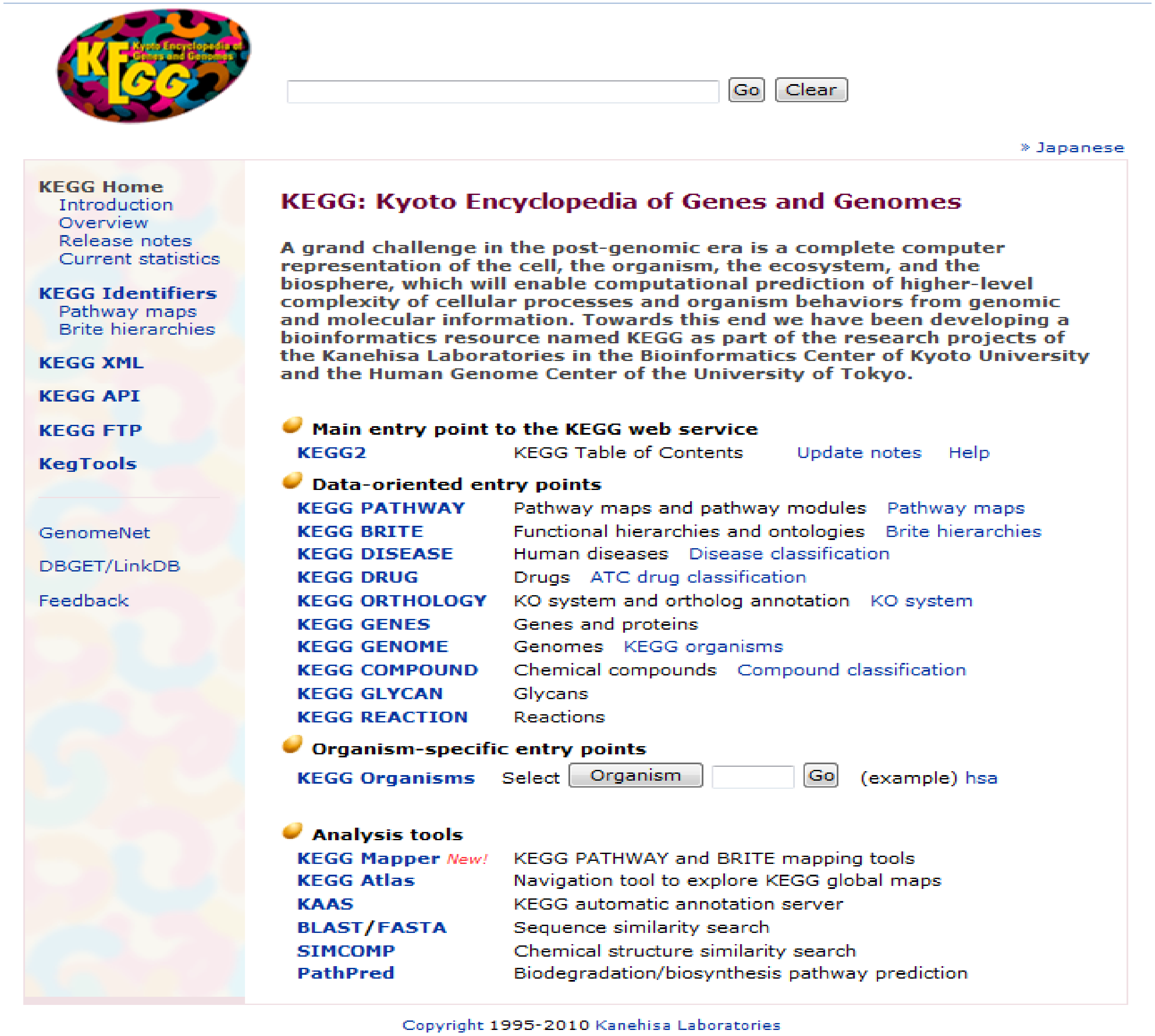
Combined pathway
human + gut microbiome

GenomeNet Bioinformatics Tools

Sequence Analysis
  BLAST / FASTA - Sequence similarity search
  MOTIF - Sequence motif search
  CLUSTALW / MAFFT / PRRN - Multiple alignment

Genome Analysis
  KAAS - KEGG automatic annotation server
  EGassembler - EST consensus contigs
  GENIES - Gene network prediction
  GECS - Gene expression to chemical structure

http://www.genome.jp/



KEGG Home
- Introduction
- Overview
- Release notes
- Current statistics

KEGG Identifiers
- Pathway maps
- Brite hierarchies

KEGG XML

KEGG API

KEGG FTP

KegTools

GenomeNet

DBGET/LinkDB

Feedback

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.

Main entry point to the KEGG web service

KEGG2	KEGG Table of Contents	Update notes	Help

Data-oriented entry points

KEGG PATHWAY	Pathway maps and pathway modules	Pathway maps
KEGG BRITE	Functional hierarchies and ontologies	Brite hierarchies
KEGG DISEASE	Human diseases	Disease classification
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 entry points

KEGG Organisms Select (example) hsa

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

Copyright 1995-2010 Kanehisa Laboratories

http://www.genome.jp/kegg/

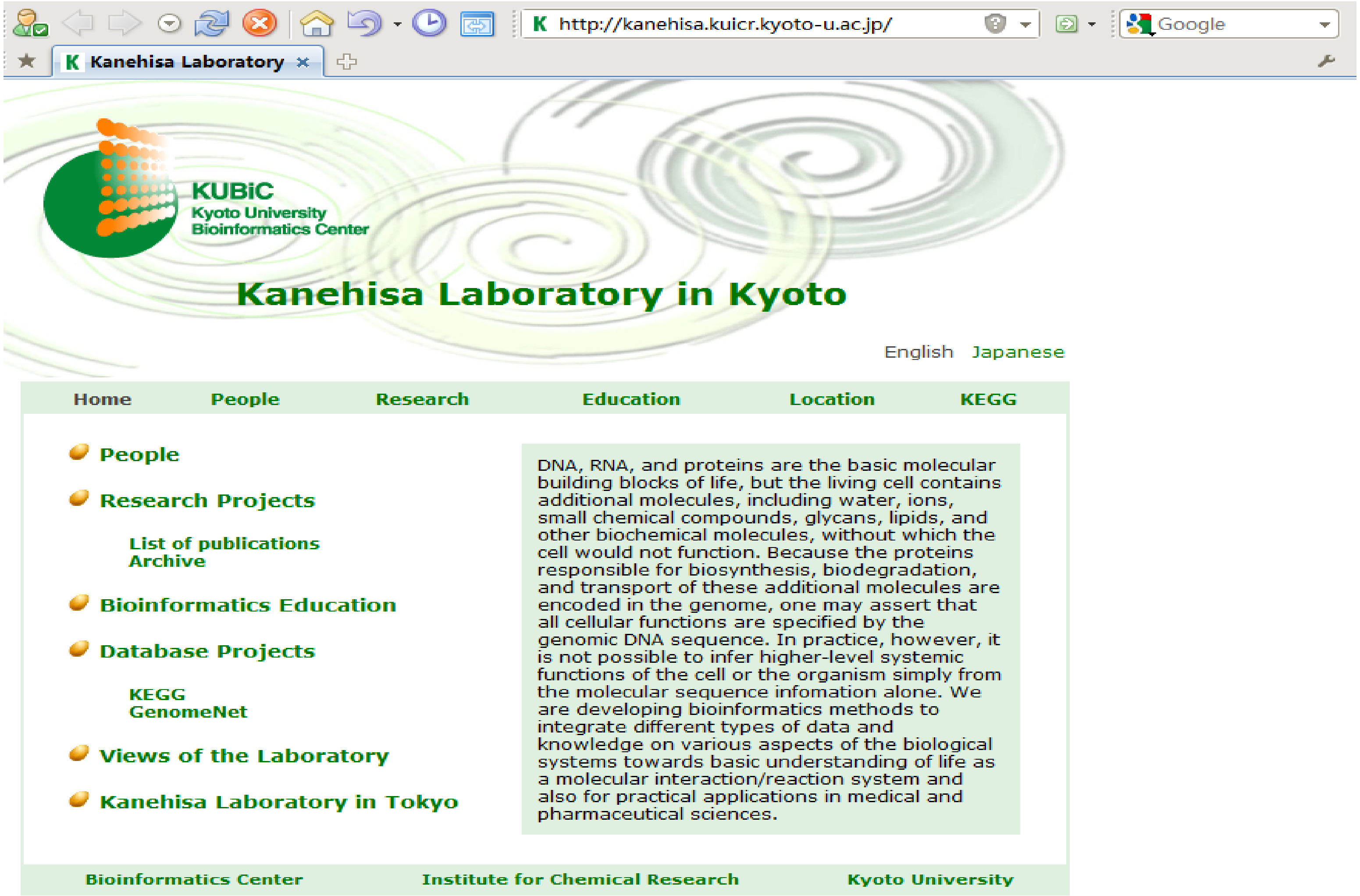
KEGG的产生背景

KEGG是系统分析基因功能的数据库，它将基因组的信息与基因功能联系起来，旨在揭示生命现象的遗传与化学蓝图。如何借助计算机全面地展示细胞和生物所包含的生物学信息是后基因组时代的重大挑战之一。科学家期望能够根据基因组中的信息，用计算机计算或者预测出比较复杂的细胞中的通路或者生物的复杂行为。出于这个目的，日本京都大学生物信息学中心的Kanehisa实验室建立了一个生物信息学数据库KEGG。

1995年5月，日本的教育、科学、体育和文化部携手建立了KEGG数据库。经过了15年的发展，数据库的规模不断的扩大，其应用价值也得到了世界范围内的认可。

KEGG是系统地分析基因功能、链接基因组信息和功能信息的数据库。数据库的维护人员不定期的根据最近出版的一些学术论文和生物学实验得到的数据对该数据库进行更新，来保证数据库的信息与最近的科研成果同步。

Kanehisa Laboratory



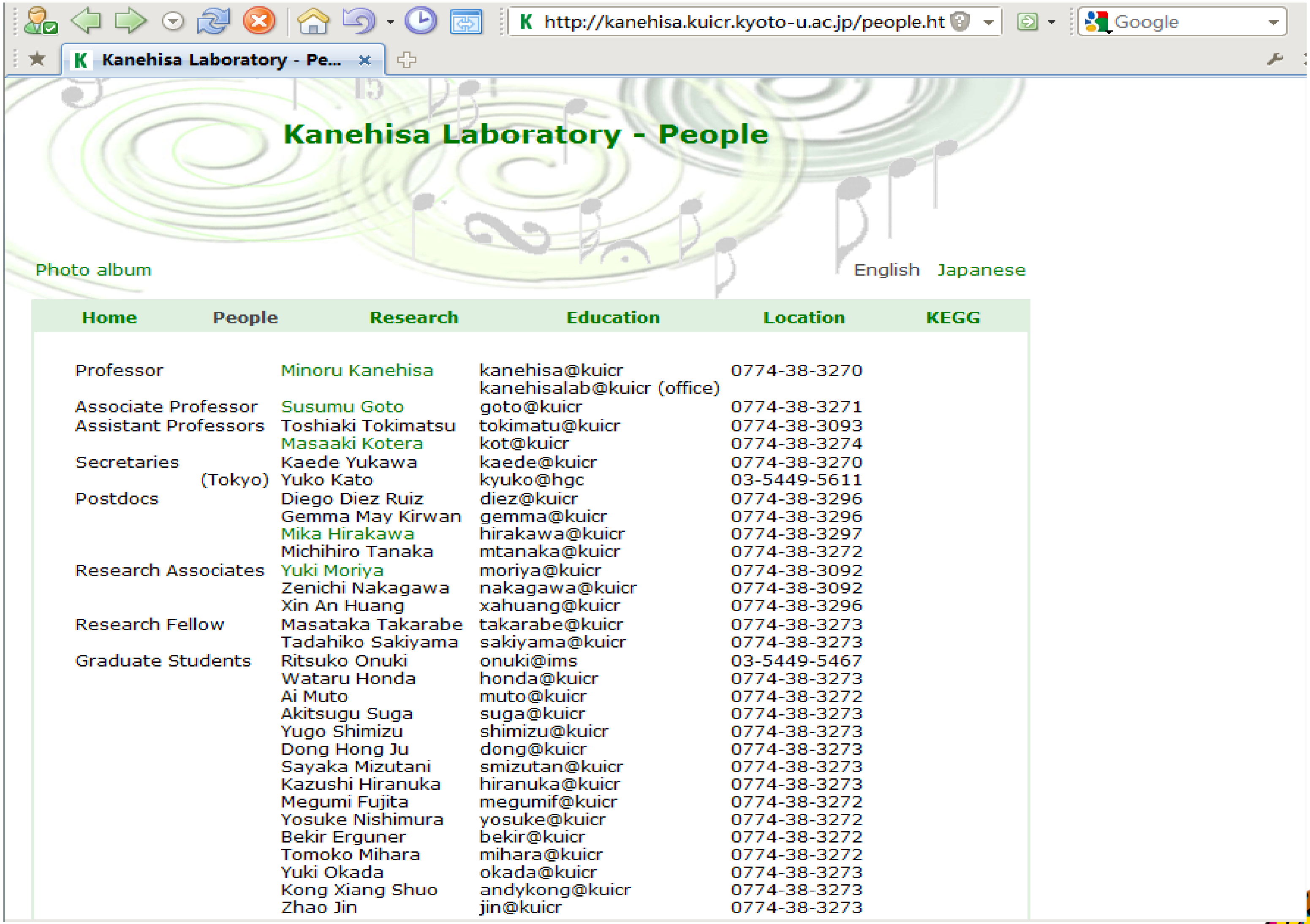
Kanehisa Laboratory in Kyoto

- People
- Research Projects
  - List of publications
  - Archive
- Bioinformatics Education
- Database Projects
  - KEGG
  - GenomeNet
- Views of the Laboratory
- Kanehisa Laboratory in Tokyo

DNA, RNA, and proteins are the basic molecular building blocks of life, but the living cell contains additional molecules, including water, ions, small chemical compounds, glycans, lipids, and other biochemical molecules, without which the cell would not function. Because the proteins responsible for biosynthesis, biodegradation, and transport of these additional molecules are encoded in the genome, one may assert that all cellular functions are specified by the genomic DNA sequence. In practice, however, it is not possible to infer higher-level systemic functions of the cell or the organism simply from the molecular sequence infomation alone. We are developing bioinformatics methods to integrate different types of data and knowledge on various aspects of the biological systems towards basic understanding of life as a molecular interaction/reaction system and also for practical applications in medical and pharmaceutical sciences.

http://kanehisa.kuicr.kyoto-u.ac.jp/

Kanehisa Laboratory



Kanehisa Laboratory - People

Photo album

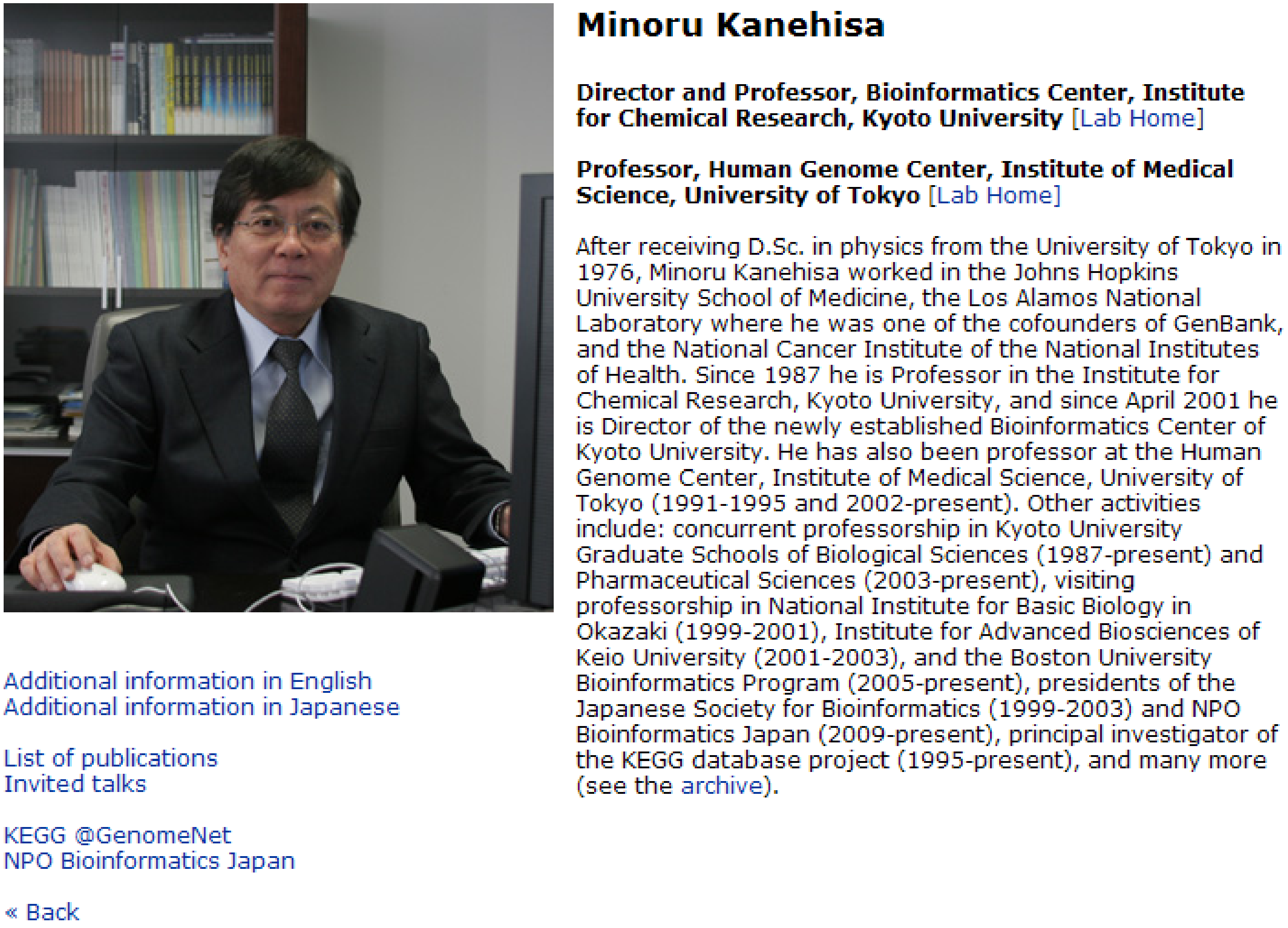
English Japanese

Home People Research Education Location KEGG

Professor	Minoru Kanehisa	kanehisa@kuicr kanehisalab@kuicr (office)	0774-38-3270
Associate Professor	Susumu Goto	goto@kuicr	0774-38-3271
Assistant Professors	Toshiaki Tokimatsu	tokimatu@kuicr	0774-38-3093
	Masaaki Kotera	kot@kuicr	0774-38-3274
Secretaries	Kaede Yukawa	kaede@kuicr	0774-38-3270
(Tokyo)	Yuko Kato	kyuko@hgc	03-5449-5611
Postdocs	Diego Diez Ruiz	diez@kuicr	0774-38-3296
	Gemma May Kirwan	gemma@kuicr	0774-38-3296
	Mika Hirakawa	hirakawa@kuicr	0774-38-3297
	Michihiro Tanaka	mtanaka@kuicr	0774-38-3272
Research Associates	Yuki Moriya	moriya@kuicr	0774-38-3092
	Zenichi Nakagawa	nakagawa@kuicr	0774-38-3092
	Xin An Huang	xahuang@kuicr	0774-38-3296
Research Fellow	Masataka Takarabe	takarabe@kuicr	0774-38-3273
	Tadahiko Sakiyama	sakiyama@kuicr	0774-38-3273
Graduate Students	Ritsuko Onuki	onuki@ims	03-5449-5467
	Wataru Honda	honda@kuicr	0774-38-3273
	Ai Muto	muto@kuicr	0774-38-3272
	Akitsugu Suga	suga@kuicr	0774-38-3273
	Yugo Shimizu	shimizu@kuicr	0774-38-3273
	Dong Hong Ju	dong@kuicr	0774-38-3273
	Sayaka Mizutani	smizutan@kuicr	0774-38-3273
	Kazushi Hiranuka	hiranuka@kuicr	0774-38-3273
	Megumi Fujita	megumif@kuicr	0774-38-3272
	Yosuke Nishimura	yosuke@kuicr	0774-38-3272
	Bekir Erguner	bekir@kuicr	0774-38-3272
	Tomoko Mihara	mihara@kuicr	0774-38-3272
	Yuki Okada	okada@kuicr	0774-38-3273
	Kong Xiang Shuo	andykong@kuicr	0774-38-3273
	Zhao Jin	jin@kuicr	0774-38-3273

http://kanehisa.kuicr.kyoto-u.ac.jp/people.html

Minoru Kanehisa



Minoru Kanehisa

Director and Professor, Bioinformatics Center, Institute for Chemical Research, Kyoto University [Lab Home]

Professor, Human Genome Center, Institute of Medical Science, University of Tokyo [Lab Home]

After receiving D.Sc. in physics from the University of Tokyo in 1976, Minoru Kanehisa worked in the Johns Hopkins University School of Medicine, the Los Alamos National Laboratory where he was one of the cofounders of GenBank, and the National Cancer Institute of the National Institutes of Health. Since 1987 he is Professor in the Institute for Chemical Research, Kyoto University, and since April 2001 he is Director of the newly established Bioinformatics Center of Kyoto University. He has also been professor at the Human Genome Center, Institute of Medical Science, University of Tokyo (1991-1995 and 2002-present). Other activities include: concurrent professorship in Kyoto University Graduate Schools of Biological Sciences (1987-present) and Pharmaceutical Sciences (2003-present), visiting professorship in National Institute for Basic Biology in Okazaki (1999-2001), Institute for Advanced Biosciences of Keio University (2001-2003), and the Boston University Bioinformatics Program (2005-present), presidents of the Japanese Society for Bioinformatics (1999-2003) and NPO Bioinformatics Japan (2009-present), principal investigator of the KEGG database project (1995-present), and many more (see the archive).

Additional information in English
Additional information in Japanese

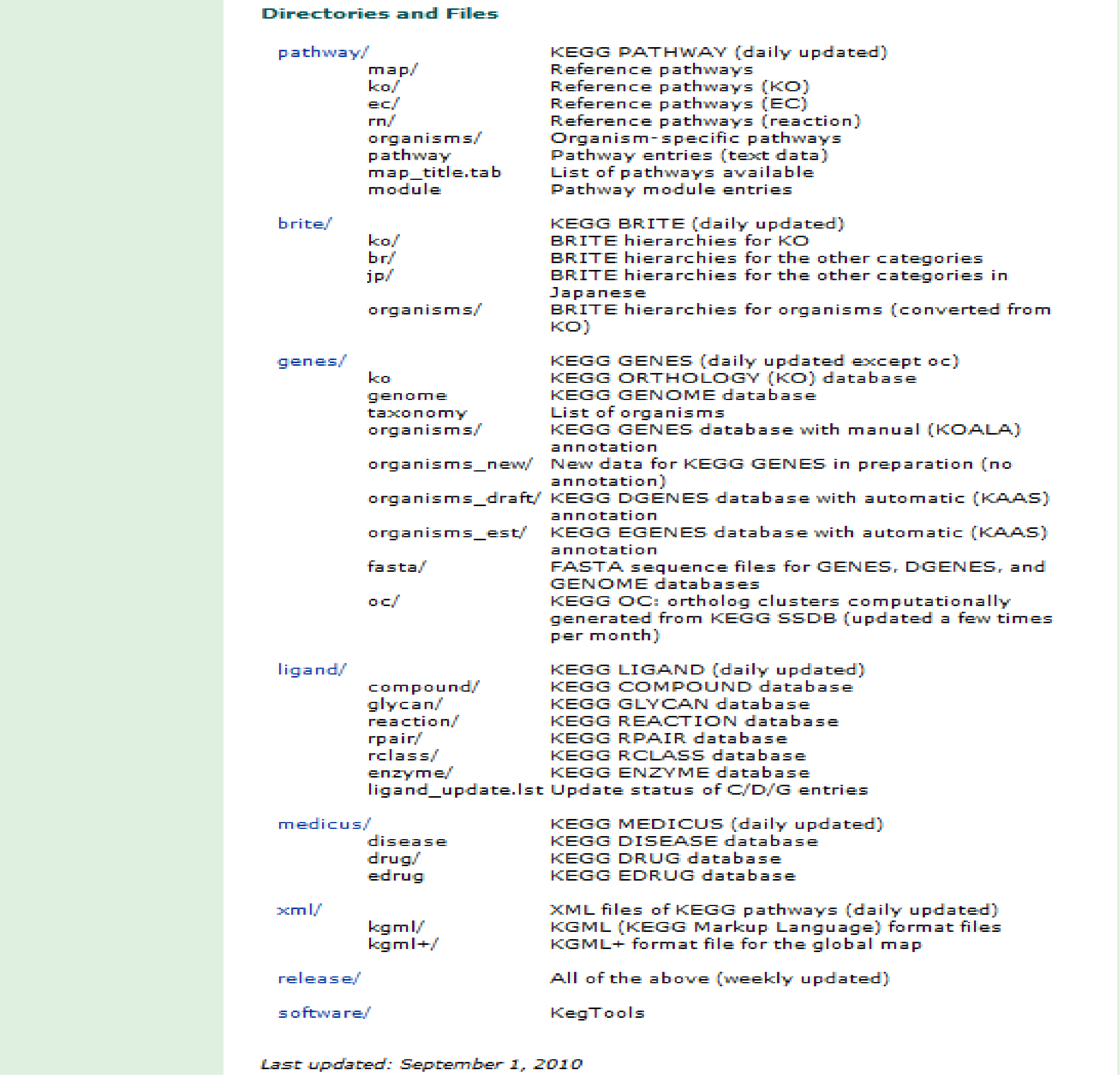
List of publications
Invited talks

KEGG @GenomeNet
NPO Bioinformatics Japan

« Back

http://kanehisa.kuicr.kyoto-u.ac.jp/People/kanehisa.html

数据类型

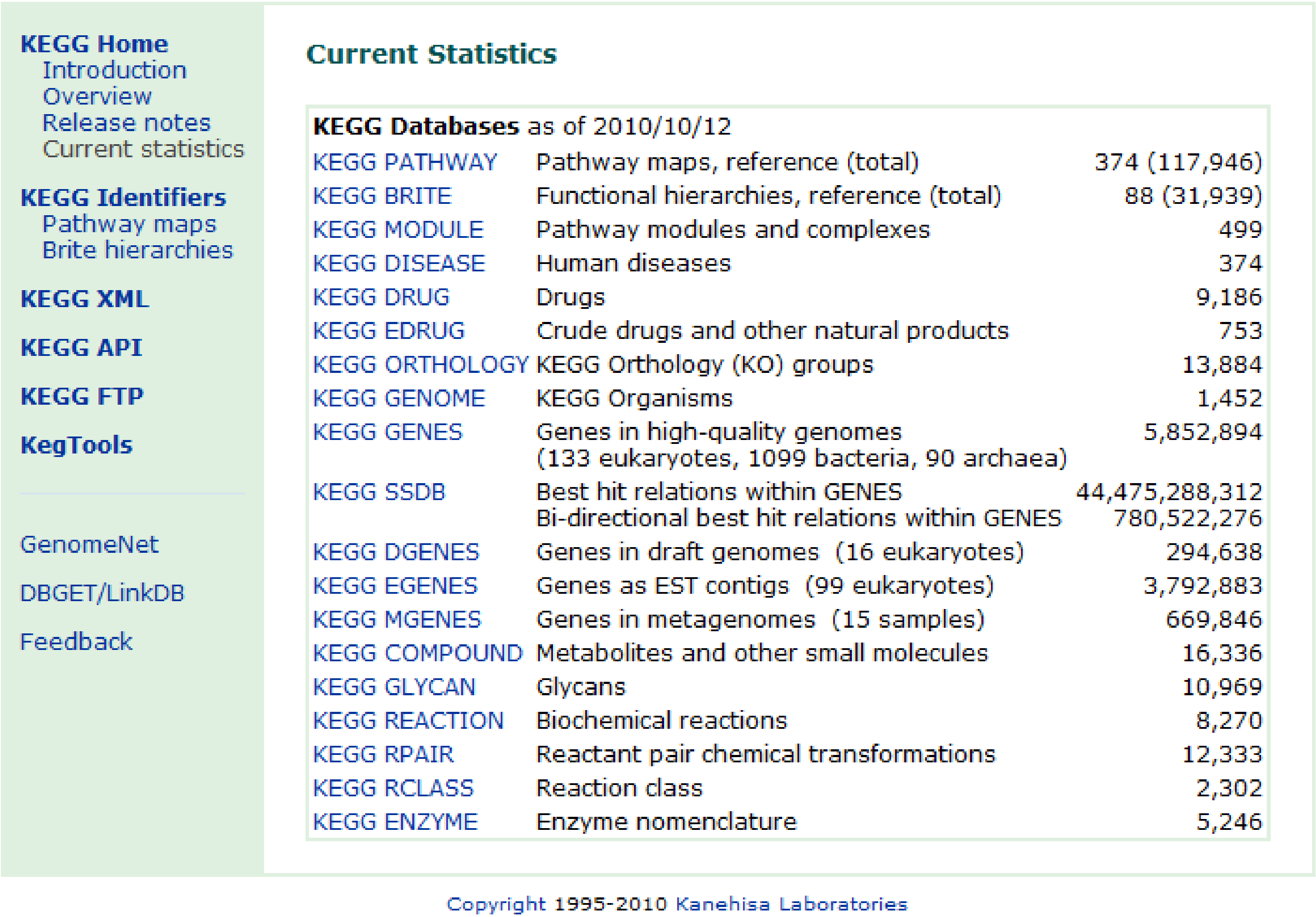


Directories and Files

pathway/		KEGG PATHWAY (daily updated)
	map/	Reference pathways
	ko/	Reference pathways (KO)
	ec/	Reference pathways (EC)
	rn/	Reference pathways (reaction)
	organisms/	Organism-specific pathways
	pathway	Pathway entries (text data)
	map_title.tab	List of pathways available
	module	Pathway module entries
brite/		KEGG BRITE (daily updated)
	ko/	BRITE hierarchies for KO
	br/	BRITE hierarchies for the other categories
	jp/	BRITE hierarchies for the other categories in Japanese
	organisms/	BRITE hierarchies for organisms (converted from KO)
genes/		KEGG GENES (daily updated except oc)
	ko	KEGG ORTHOLOGY (KO) database
	genome	KEGG GENOME database
	taxonomy	List of organisms
	organisms/	KEGG GENES database with manual (KOALA) annotation
	organisms_new/	New data for KEGG GENES in preparation (no annotation)
	organisms_draft/	KEGG DGENES database with automatic (KAAS) annotation
	organisms_est/	KEGG EGENES database with automatic (KAAS) annotation
	fasta/	FASTA sequence files for GENES, DGENES, and GENOME databases
	oc/	KEGG OC: ortholog clusters computationally generated from KEGG SSDB (updated a few times per month)
ligand/		KEGG LIGAND (daily updated)
	compound/	KEGG COMPOUND database
	glycan/	KEGG GLYCAN database
	reaction/	KEGG REACTION database
	rpair/	KEGG RPAIR database
	rclass/	KEGG RCLASS database
	enzyme/	KEGG ENZYME database
	ligand_update.lst	Update status of C/D/G entries
medicus/		KEGG MEDICUS (daily updated)
	disease	KEGG DISEASE database
	drug/	KEGG DRUG database
	edrug	KEGG EDRUG database
xml/		XML files of KEGG pathways (daily updated)
	kgml/	KGML (KEGG Markup Language) format files
	kgml+/	KGML+ format file for the global map
release/		All of the above (weekly updated)
software/		KegTools

Last updated: September 1, 2010

数据库容量



Current Statistics

KEGG Databases as of 2010/10/12		
KEGG PATHWAY	Pathway maps, reference (total)	374 (117,946)
KEGG BRITE	Functional hierarchies, reference (total)	88 (31,939)
KEGG MODULE	Pathway modules and complexes	499
KEGG DISEASE	Human diseases	374
KEGG DRUG	Drugs	9,186
KEGG EDRUG	Crude drugs and other natural products	753
KEGG ORTHOLOGY	KEGG Orthology (KO) groups	13,884
KEGG GENOME	KEGG Organisms	1,452
KEGG GENES	Genes in high-quality genomes (133 eukaryotes, 1099 bacteria, 90 archaea)	5,852,894
KEGG SSDB	Best hit relations within GENES Bi-directional best hit relations within GENES	44,475,288,312 780,522,276
KEGG DGENES	Genes in draft genomes (16 eukaryotes)	294,638
KEGG EGENES	Genes as EST contigs (99 eukaryotes)	3,792,883
KEGG MGENES	Genes in metagenomes (15 samples)	669,846
KEGG COMPOUND	Metabolites and other small molecules	16,336
KEGG GLYCAN	Glycans	10,969
KEGG REACTION	Biochemical reactions	8,270
KEGG RPAIR	Reactant pair chemical transformations	12,333
KEGG RCLASS	Reaction class	2,302
KEGG ENZYME	Enzyme nomenclature	5,246

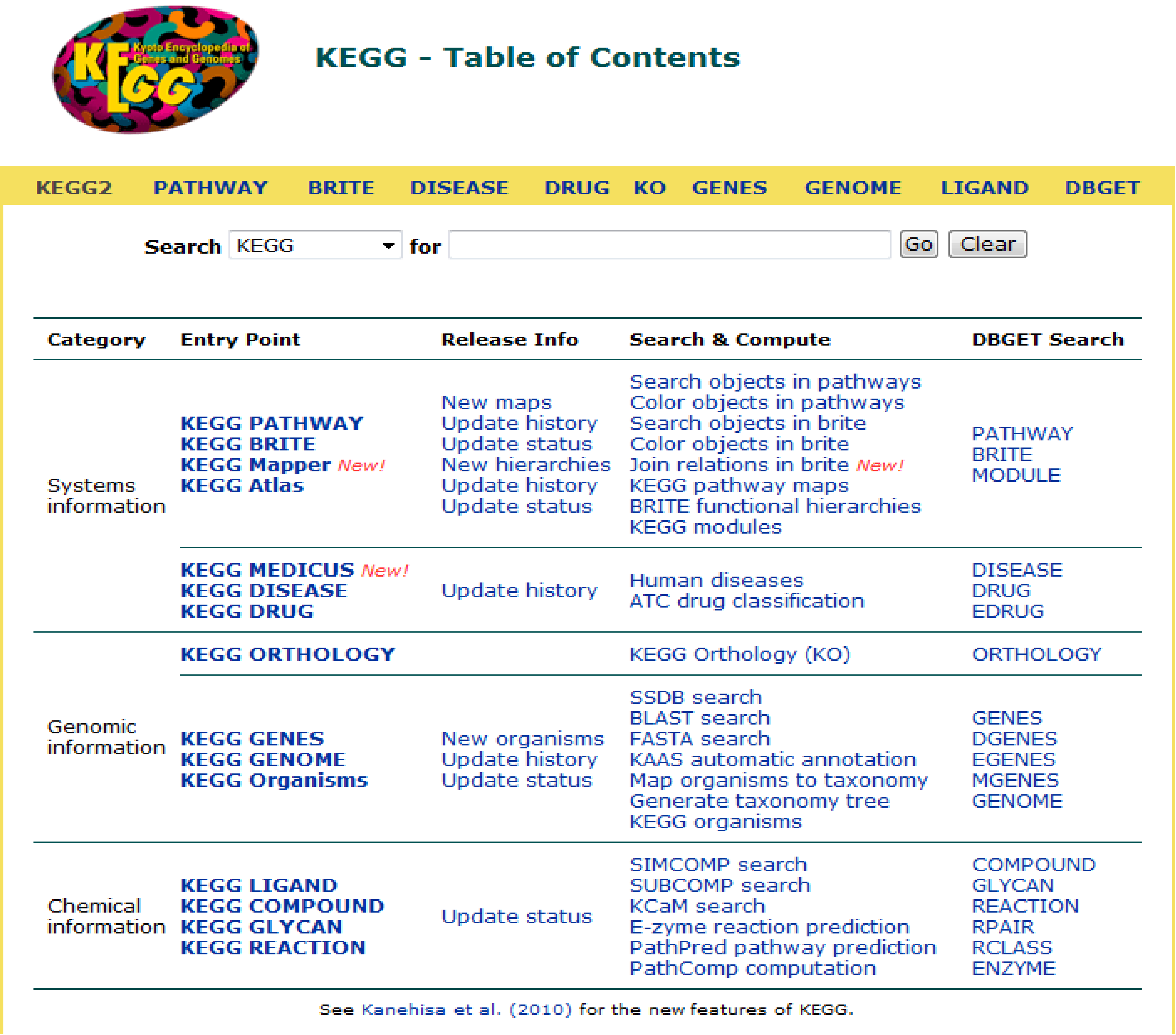
Copyright 1995-2010 Kanehisa Laboratories

Current statistics

http://www.genome.jp/kegg/docs/statistics.html

更新频率

- KEGG数据库更新频繁，几乎每1~2天更新一次



KEGG - Table of Contents

KEGG2 PATHWAY BRITE DISEASE DRUG KO GENES GENOME LIGAND DBGET

Search KEGG ▾ **for** [Go] [Clear]

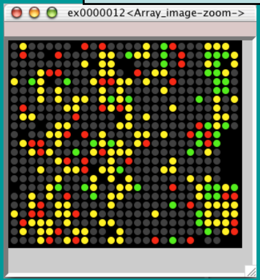
Category	Entry Point	Release Info	Search & Compute	DBGET Search
Systems information	**KEGG PATHWAY** **KEGG BRITE** **KEGG Mapper** *New!* **KEGG Atlas**	New maps Update history Update status New hierarchies Update history Update status	Search objects in pathways Color objects in pathways Search objects in brite Color objects in brite Join relations in brite *New!* KEGG pathway maps BRITE functional hierarchies KEGG modules	PATHWAY BRITE MODULE
	KEGG MEDICUS *New!* **KEGG DISEASE** **KEGG DRUG**	Update history	Human diseases ATC drug classification	DISEASE DRUG EDRUG
Genomic information	**KEGG ORTHOLOGY**		KEGG Orthology (KO)	ORTHOLOGY
	KEGG GENES **KEGG GENOME** **KEGG Organisms**	New organisms Update history Update status	SSDB search BLAST search FASTA search KAAS automatic annotation Map organisms to taxonomy Generate taxonomy tree KEGG organisms	GENES DGENES EGENES MGENES GENOME
Chemical information	**KEGG LIGAND** **KEGG COMPOUND** **KEGG GLYCAN** **KEGG REACTION**	Update status	SIMCOMP search SUBCOMP search KCaM search E-zyme reaction prediction PathPred pathway prediction PathComp computation	COMPOUND GLYCAN REACTION RPAIR RCLASS ENZYME

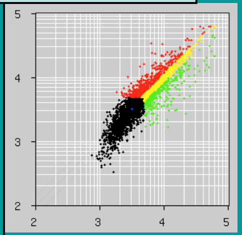
See Kanehisa et al. (2010) for the new features of KEGG.

KEGG 包含的信息大致可以分为三类：
系统信息，基因组信息和化学信息。

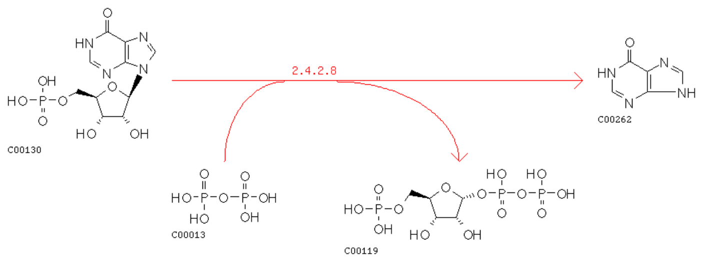
Databases in KEGG

Expression

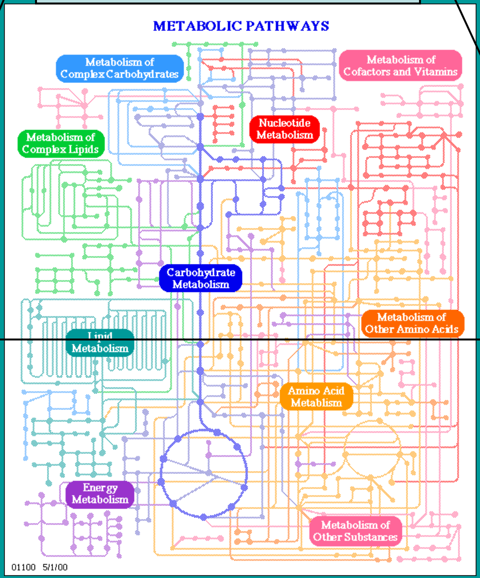




Chemicals and their reactions



Pathway



Genomes
Genes

Orthologs

Sequence similarity

Information on relations between molecules

KEGG的特点

与其他数据库相比，KEGG 的一个显著特点就是具有强大的图形功能，它利用图形而不是繁缛的文字来介绍众多的代谢途径以及各途径之间的关系，这样可以使研究者能够对其所要研究的代谢途径有一个直观全面的了解。此外KEGG 还具有检索系统全面、界面友好等特点。另一方面通过与世界上其它一些大型生物信息学数据库的连接，KEGG可以为研究者提供更为丰富的生物学信息。

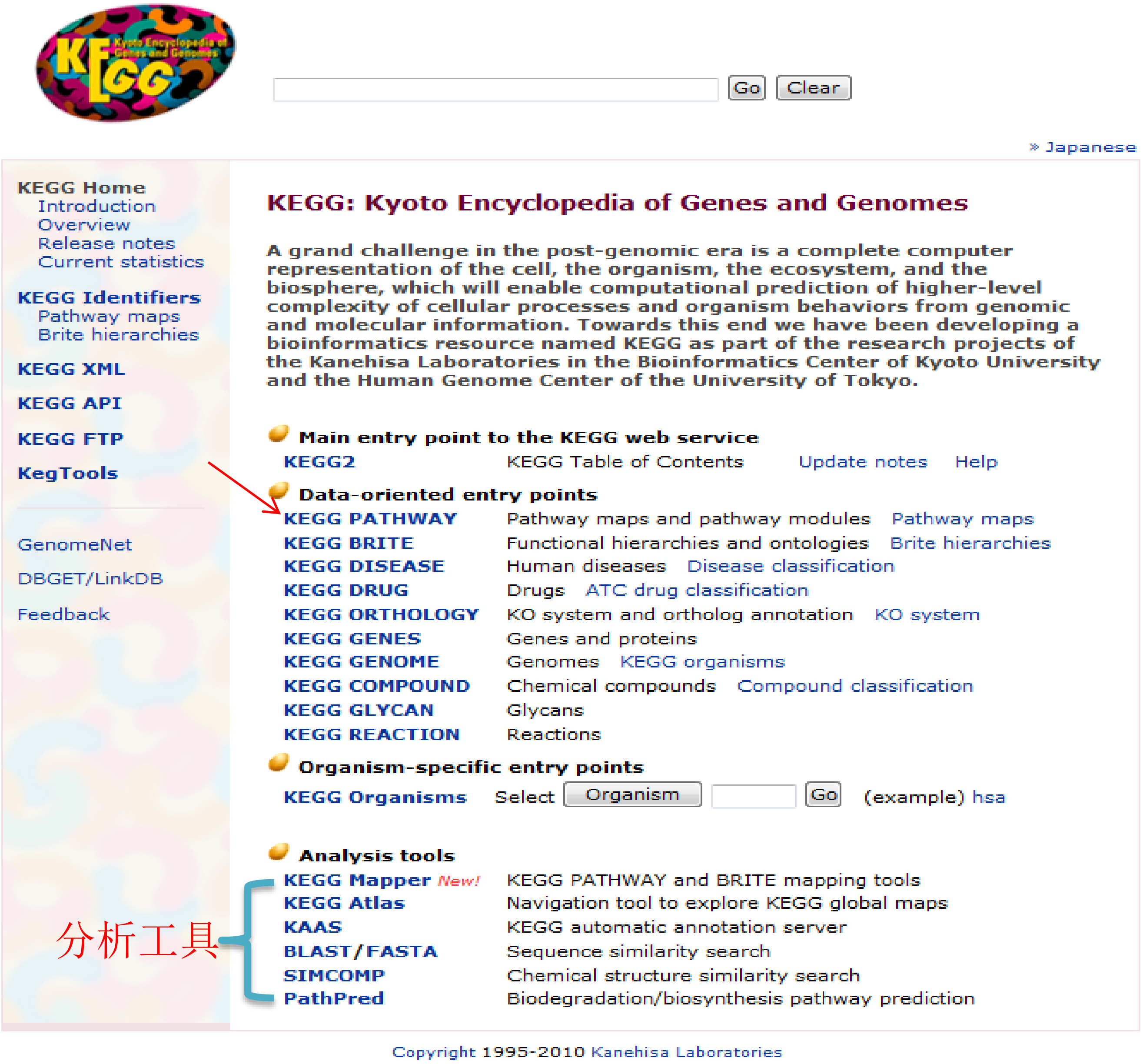
KEGG的用途

KEGG是系统分析基因功能，联系基因组信息和功能信息的知识库。基因组信息存储在GENES数据库里，包括完整和部分测序的基因组序列；更高级的功能信息存储在PATHWAY数据库里，包括图解的细胞生化过程如代谢、膜转运、信号传递、细胞周期，还包括同系保守的子通路等信息；KEGG的另一个数据库LIGAND，包含关于化学物质、酶分子、酶反应等信息。KEGG提供了Java的图形工具来访问基因组图谱，比较基因组图谱和操作表达图谱，以及其它序列比较、图形比较和通路计算的工具，可以免费获取。

由于KEGG中包括含的数据库比较多下面将选取下面6个比较常用的子数据库，简单演示如何进行数据查询。

KEGG PATHWAY Database(代谢通路数据库)
KEGG BRITE Database(分层分类数据库)
KEGG GENES Database(基因数据库)
KEGG GENOME Database(基因组数据库)
KEGG LIGAND Database(配体数据库)
KEGG EXPRESSION Database(基因芯片数据库)

一、代谢通路数据库



KEGG Home
- Introduction
- Overview
- Release notes
- Current statistics

KEGG Identifiers
- Pathway maps
- Brite hierarchies

KEGG XML

KEGG API

KEGG FTP

KegTools

GenomeNet

DBGET/LinkDB

Feedback

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.

Main entry point to the KEGG web service

KEGG2	KEGG Table of Contents	Update notes Help

Data-oriented entry points

KEGG PATHWAY	Pathway maps and pathway modules	Pathway maps
KEGG BRITE	Functional hierarchies and ontologies	Brite hierarchies
KEGG DISEASE	Human diseases	Disease classification
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 entry points

KEGG Organisms   Select [Organism] [ ] [Go]   (example) hsa

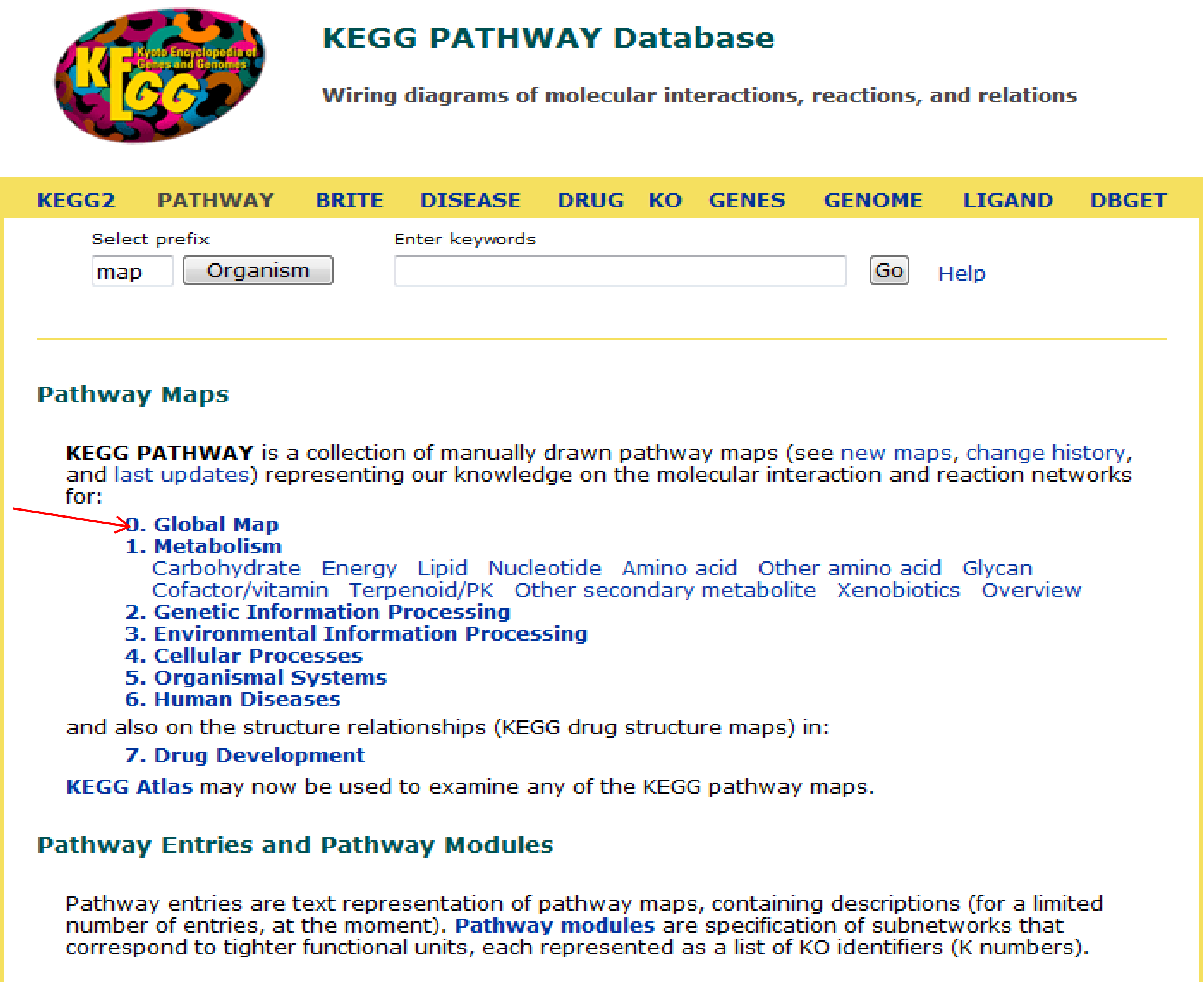
Analysis tools

分析工具

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

Copyright 1995-2010 Kanehisa Laboratories

http://www.genome.jp/kegg/



KEGG PATHWAY Database

Wiring diagrams of molecular interactions, reactions, and relations

Pathway Maps

KEGG PATHWAY is a collection of manually drawn pathway maps (see new maps, change history, and last updates) representing our knowledge on the molecular interaction and reaction networks for:

0. Global Map
1. Metabolism
  Carbohydrate Energy Lipid Nucleotide Amino acid Other amino acid Glycan
  Cofactor/vitamin Terpenoid/PK Other secondary metabolite Xenobiotics Overview
2. Genetic Information Processing
3. Environmental Information Processing
4. Cellular Processes
5. Organismal Systems
6. Human Diseases

and also on the structure relationships (KEGG drug structure maps) in:

7. Drug Development

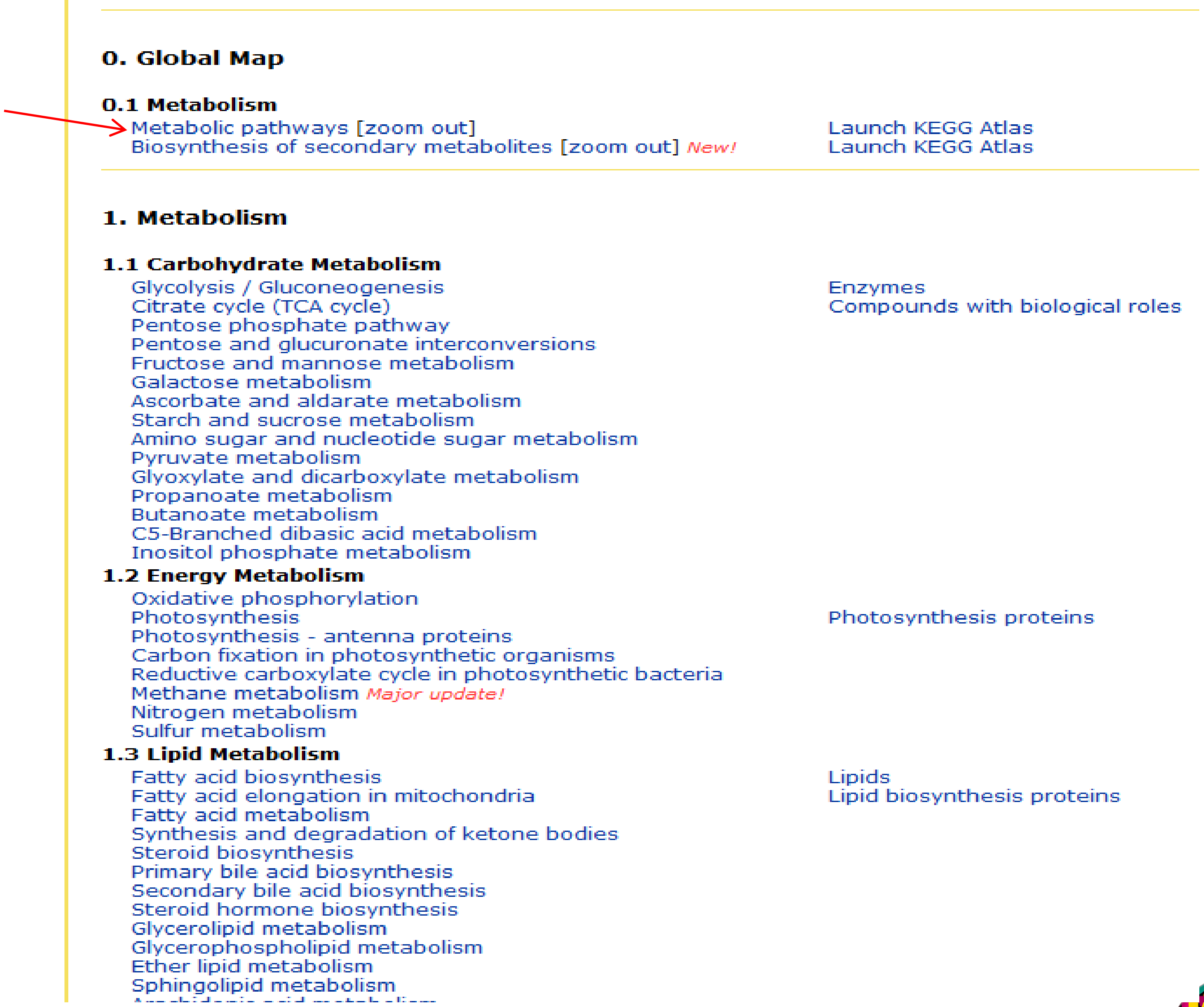
KEGG Atlas may now be used to examine any of the KEGG pathway maps.

Pathway Entries and Pathway Modules

Pathway entries are text representation of pathway maps, containing descriptions (for a limited number of entries, at the moment). **Pathway modules** are specification of subnetworks that correspond to tighter functional units, each represented as a list of KO identifiers (K numbers).

http://www.genome.jp/kegg/pathway.html

Metabolism



0. Global Map

0.1 Metabolism

Metabolic pathways [zoom out] — Launch KEGG Atlas
Biosynthesis of secondary metabolites [zoom out] *New!* — Launch KEGG Atlas

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
- Pyruvate metabolism
- Glyoxylate and dicarboxylate metabolism
- Propanoate metabolism
- Butanoate metabolism
- C5-Branched dibasic acid metabolism
- Inositol phosphate metabolism

Enzymes
Compounds with biological roles

1.2 Energy Metabolism

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

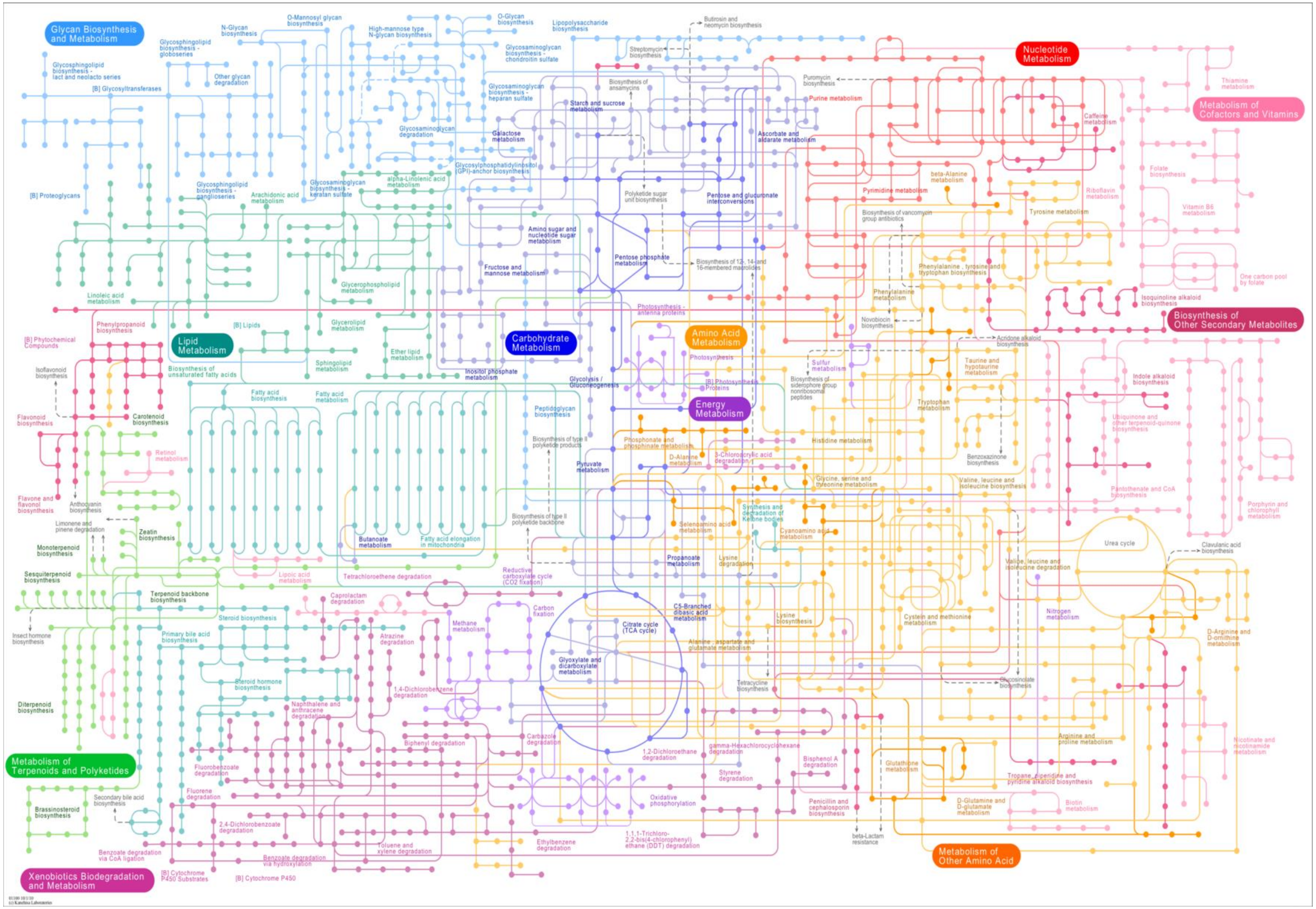
Photosynthesis proteins

1.3 Lipid Metabolism

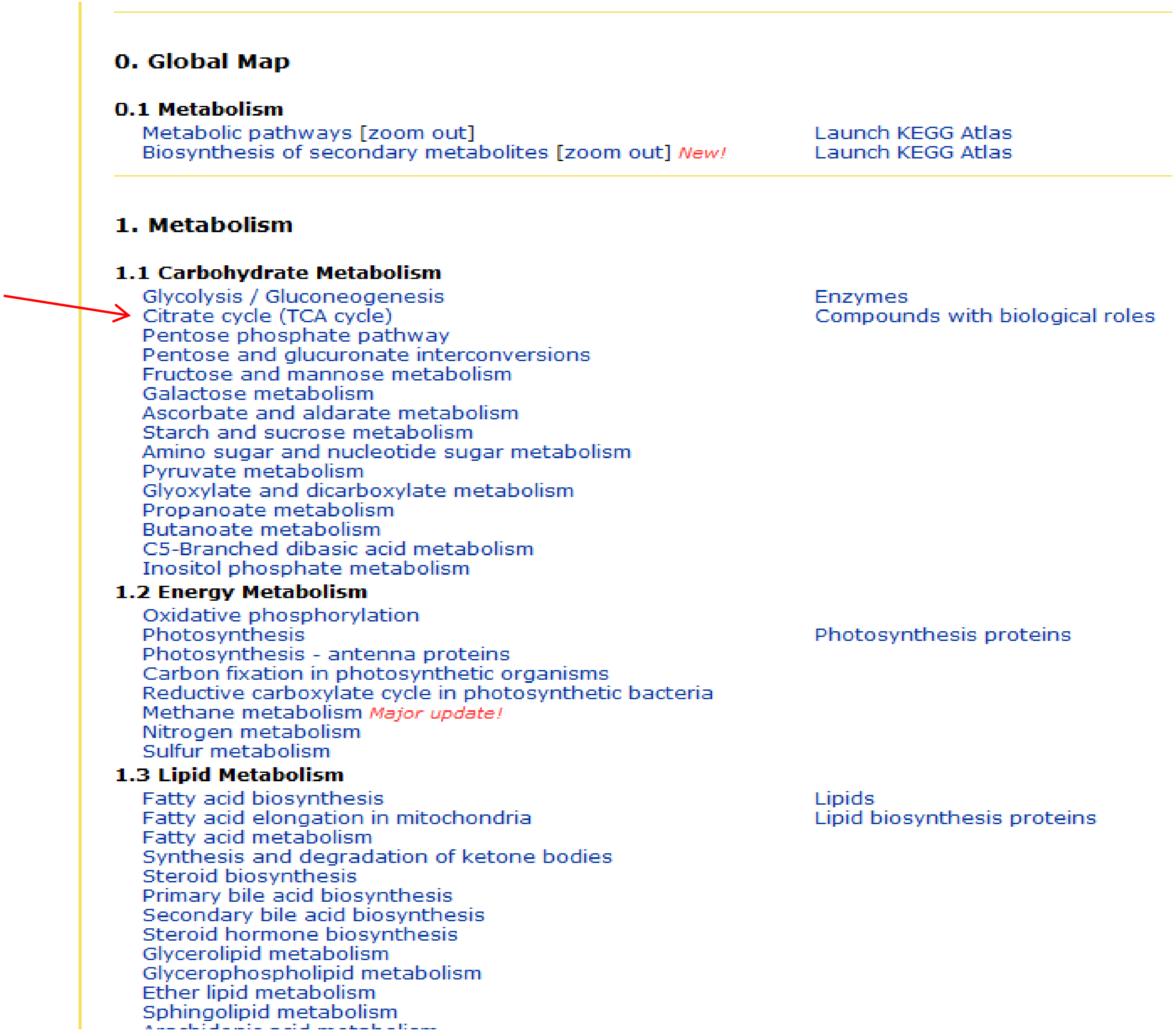
- Fatty acid biosynthesis
- Fatty acid elongation in mitochondria
- Fatty acid metabolism
- Synthesis and degradation of ketone bodies
- Steroid biosynthesis
- Primary bile acid biosynthesis
- Secondary bile acid biosynthesis
- Steroid hormone biosynthesis
- Glycerolipid metabolism
- Glycerophospholipid metabolism
- Ether lipid metabolism
- Sphingolipid metabolism

Lipids
Lipid biosynthesis proteins

http://www.genome.jp/kegg/pathway.html#global



Metabolic pathways（KEGG全局通路图）



0. Global Map

0.1 Metabolism

- Metabolic pathways [zoom out]
- Biosynthesis of secondary metabolites [zoom out] *New!*

- Launch KEGG Atlas
- Launch KEGG Atlas

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
- Pyruvate metabolism
- Glyoxylate and dicarboxylate metabolism
- Propanoate metabolism
- Butanoate metabolism
- C5-Branched dibasic acid metabolism
- Inositol phosphate metabolism

- Enzymes
- Compounds with biological roles

1.2 Energy Metabolism

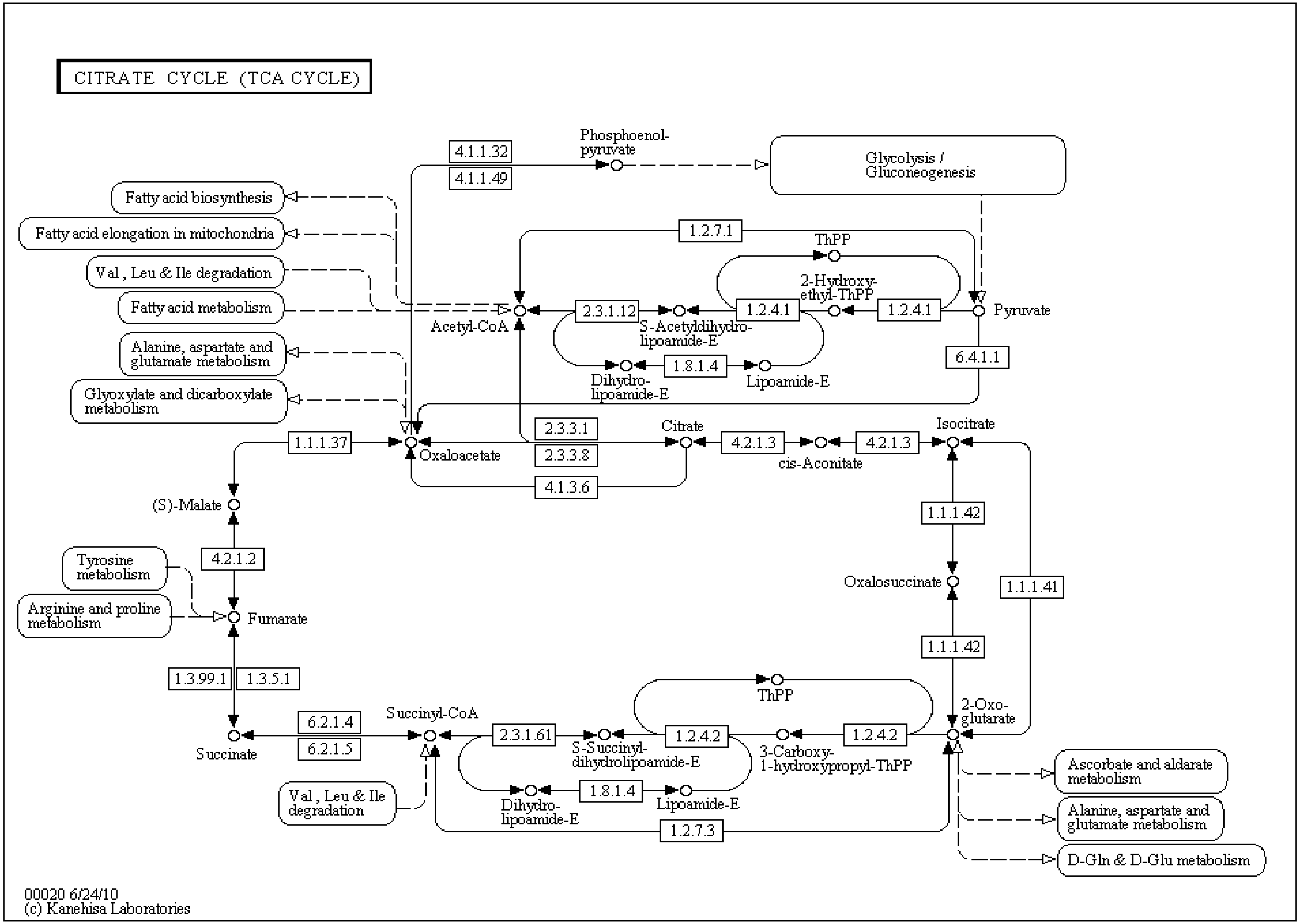
- Oxidative phosphorylation
- Photosynthesis
- Photosynthesis - antenna proteins
- Carbon fixation in photosynthetic organisms
- Reductive carboxylate cycle in photosynthetic bacteria
- Methane metabolism *Major update!*
- Nitrogen metabolism
- Sulfur metabolism

- Photosynthesis proteins

1.3 Lipid Metabolism

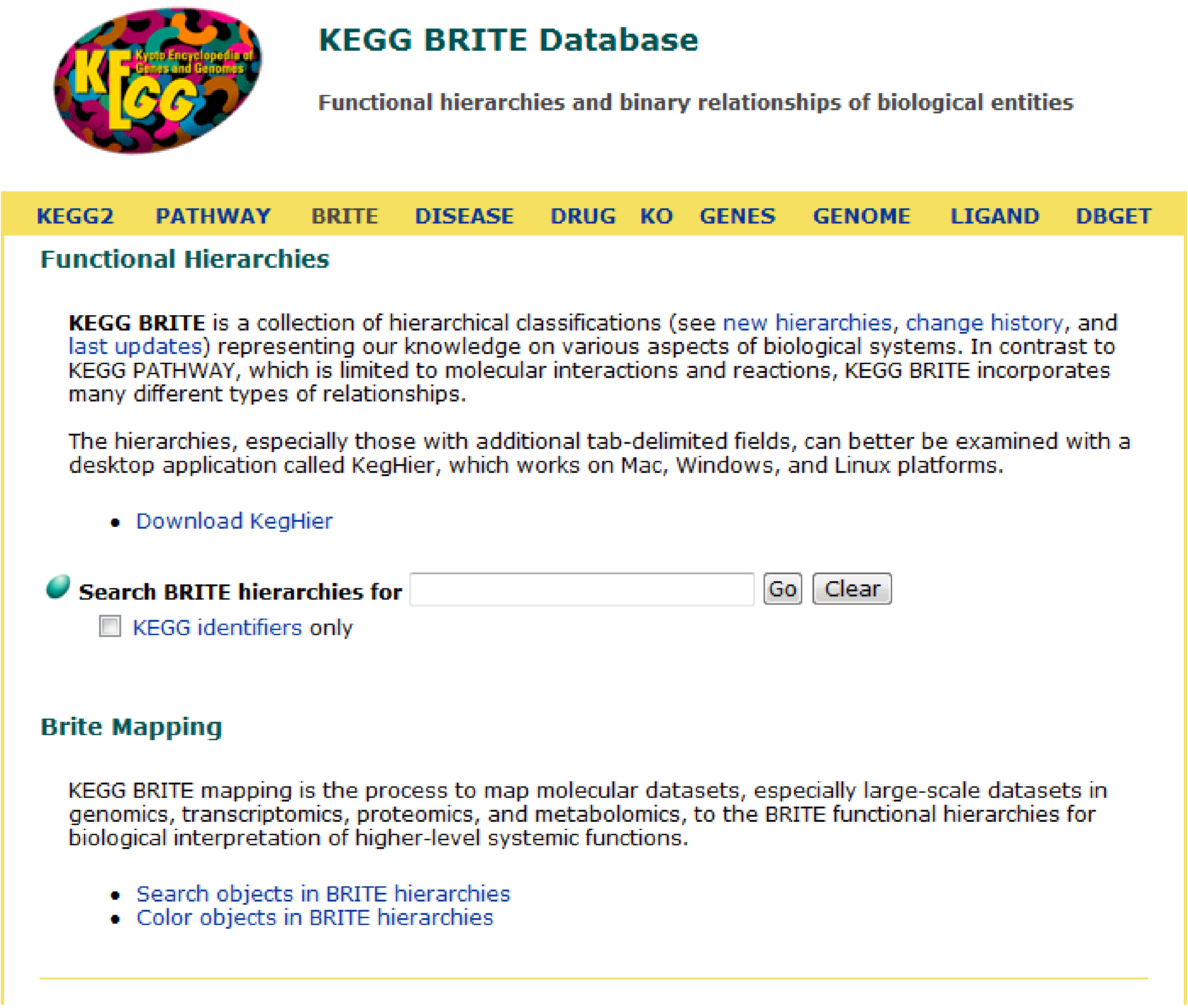
- Fatty acid biosynthesis
- Fatty acid elongation in mitochondria
- Fatty acid metabolism
- Synthesis and degradation of ketone bodies
- Steroid biosynthesis
- Primary bile acid biosynthesis
- Secondary bile acid biosynthesis
- Steroid hormone biosynthesis
- Glycerolipid metabolism
- Glycerophospholipid metabolism
- Ether lipid metabolism
- Sphingolipid metabolism

- Lipids
- Lipid biosynthesis proteins



TCA循环的代谢通路(Reference pathway)

二、分层分类数据库



KEGG BRITE Database

Functional hierarchies and binary relationships of biological entities

Functional Hierarchies

KEGG BRITE is a collection of hierarchical classifications (see new hierarchies, change history, and last updates) representing our knowledge on various aspects of biological systems. In contrast to KEGG PATHWAY, which is limited to molecular interactions and reactions, KEGG BRITE incorporates many different types of relationships.

The hierarchies, especially those with additional tab-delimited fields, can better be examined with a desktop application called KegHier, which works on Mac, Windows, and Linux platforms.

- Download KegHier

Search BRITE hierarchies for

☐ KEGG identifiers only

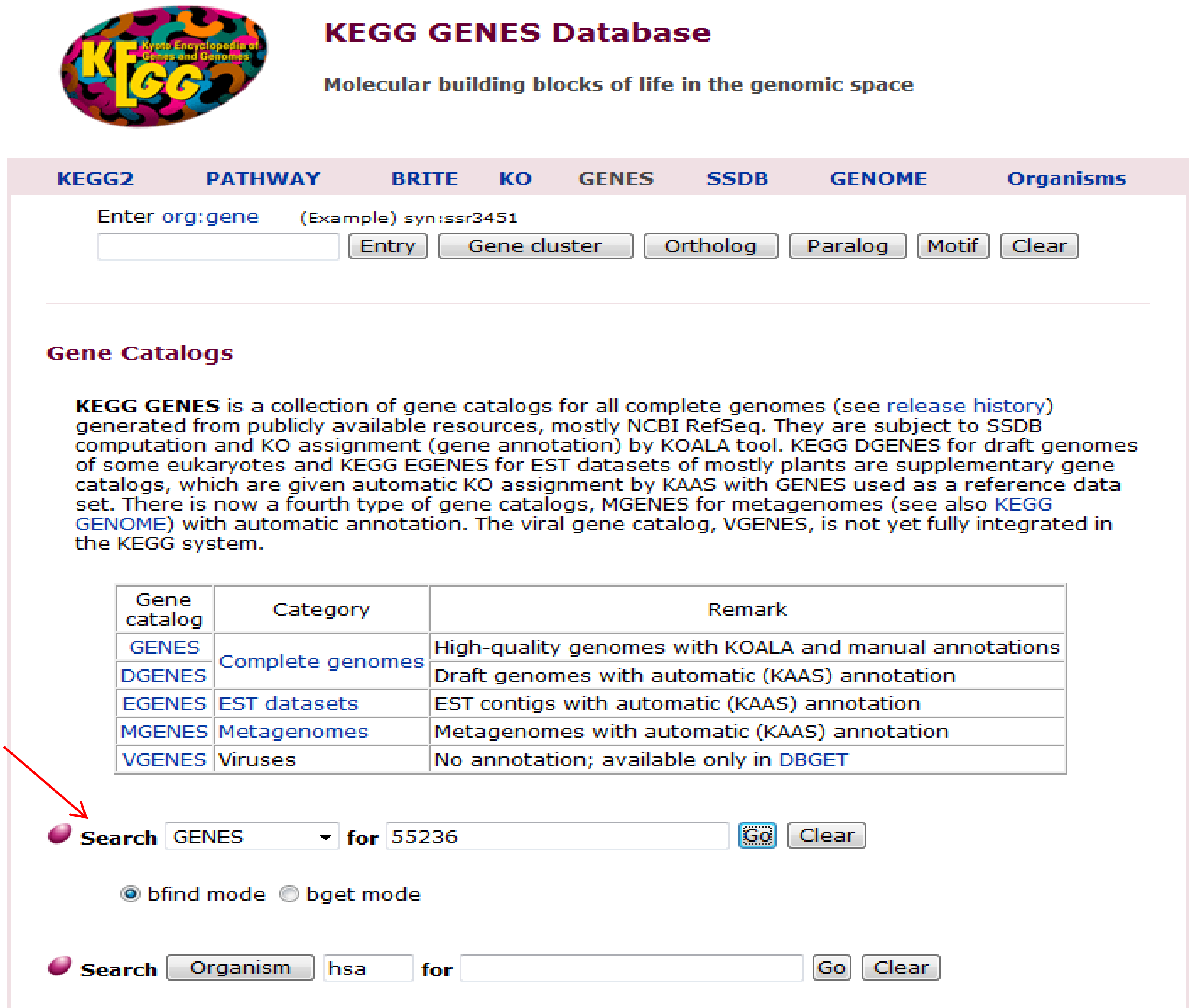
Brite Mapping

KEGG BRITE mapping is the process to map molecular datasets, especially large-scale datasets in genomics, transcriptomics, proteomics, and metabolomics, to the BRITE functional hierarchies for biological interpretation of higher-level systemic functions.

- Search objects in BRITE hierarchies
- Color objects in BRITE hierarchies

http://www.genome.jp/kegg/brite.html

三、基因数据库



KEGG GENES Database

Molecular building blocks of life in the genomic space

Enter org:gene (Example) syn:ssr3451

Gene Catalogs

KEGG GENES is a collection of gene catalogs for all complete genomes (see release history) generated from publicly available resources, mostly NCBI RefSeq. They are subject to SSDB computation and KO assignment (gene annotation) by KOALA tool. KEGG DGENES for draft genomes of some eukaryotes and KEGG EGENES for EST datasets of mostly plants are supplementary gene catalogs, which are given automatic KO assignment by KAAS with GENES used as a reference data set. There is now a fourth type of gene catalogs, MGENES for metagenomes (see also KEGG GENOME) with automatic annotation. The viral gene catalog, VGENES, is not yet fully integrated in the KEGG system.

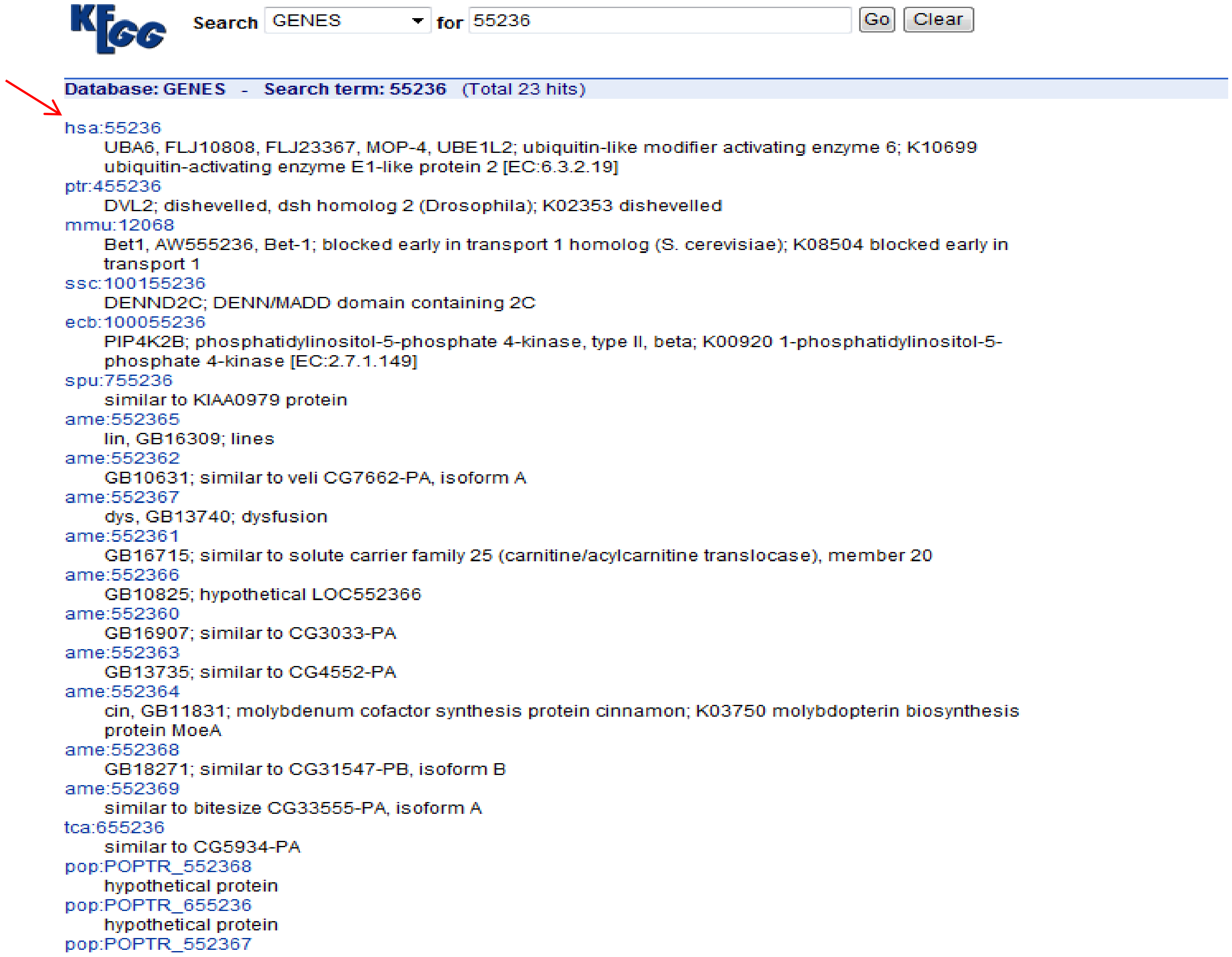
Gene catalog	Category	Remark
GENES	Complete genomes	High-quality genomes with KOALA and manual annotations
DGENES		Draft genomes with automatic (KAAS) annotation
EGENES	EST datasets	EST contigs with automatic (KAAS) annotation
MGENES	Metagenomes	Metagenomes with automatic (KAAS) annotation
VGENES	Viruses	No annotation; available only in DBGET

Search GENES for 55236

● bfind mode ○ bget mode

Search Organism hsa **for**

http://www.genome.jp/kegg/genes.html



Search GENES for 55236

Database: GENES - Search term: 55236 (Total 23 hits)

hsa:55236
UBA6, FLJ10808, FLJ23367, MOP-4, UBE1L2; ubiquitin-like modifier activating enzyme 6; K10699 ubiquitin-activating enzyme E1-like protein 2 [EC:6.3.2.19]

ptr:455236
DVL2; dishevelled, dsh homolog 2 (Drosophila); K02353 dishevelled

mmu:12068
Bet1, AW555236, Bet-1; blocked early in transport 1 homolog (S. cerevisiae); K08504 blocked early in transport 1

ssc:100155236
DENND2C; DENN/MADD domain containing 2C

ecb:100055236
PIP4K2B; phosphatidylinositol-5-phosphate 4-kinase, type II, beta; K00920 1-phosphatidylinositol-5-phosphate 4-kinase [EC:2.7.1.149]

spu:755236
similar to KIAA0979 protein

ame:552365
lin, GB16309; lines

ame:552362
GB10631; similar to veli CG7662-PA, isoform A

ame:552367
dys, GB13740; dysfusion

ame:552361
GB16715; similar to solute carrier family 25 (carnitine/acylcarnitine translocase), member 20

ame:552366
GB10825; hypothetical LOC552366

ame:552360
GB16907; similar to CG3033-PA

ame:552363
GB13735; similar to CG4552-PA

ame:552364
cin, GB11831; molybdenum cofactor synthesis protein cinnamon; K03750 molybdopterin biosynthesis protein MoeA

ame:552368
GB18271; similar to CG31547-PB, isoform B

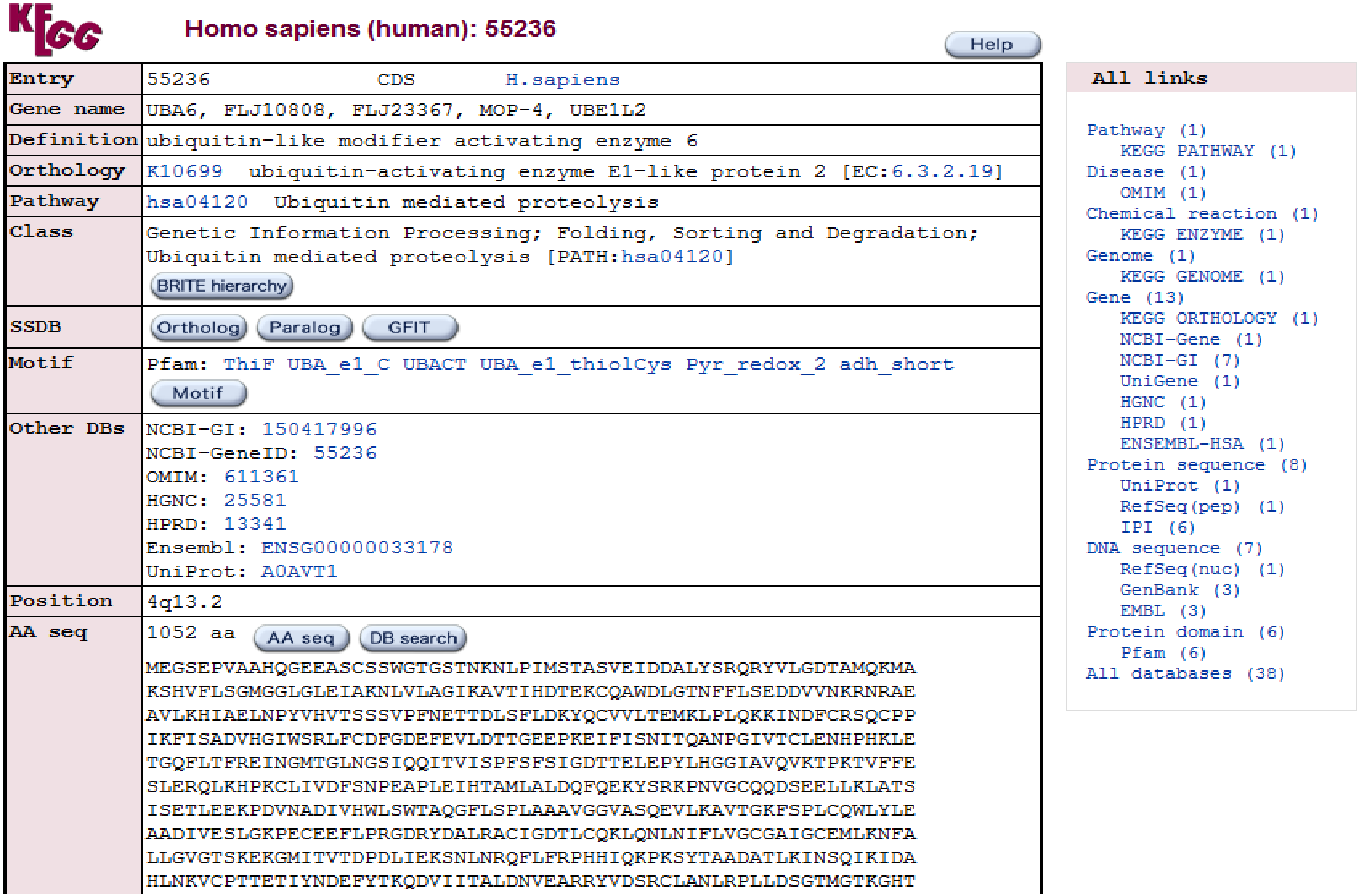
ame:552369
similar to bitesize CG33555-PA, isoform A

tca:655236
similar to CG5934-PA

pop:POPTR_552368
hypothetical protein

pop:POPTR_655236
hypothetical protein

pop:POPTR_552367



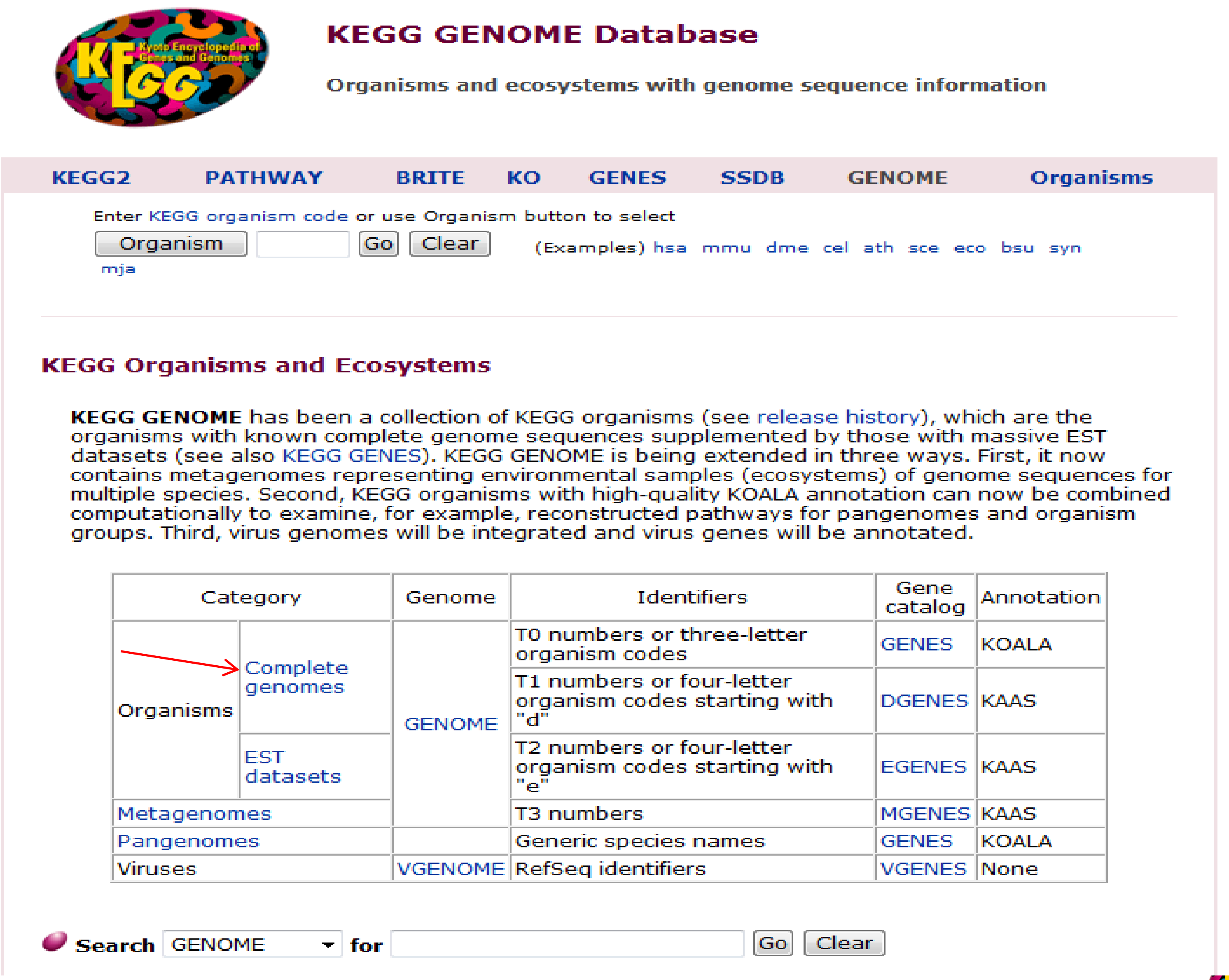
Homo sapiens (human): 55236

Entry	55236 CDS H.sapiens
Gene name	UBA6, FLJ10808, FLJ23367, MOP-4, UBE1L2
Definition	ubiquitin-like modifier activating enzyme 6
Orthology	K10699 ubiquitin-activating enzyme E1-like protein 2 [EC:6.3.2.19]
Pathway	hsa04120 Ubiquitin mediated proteolysis
Class	Genetic Information Processing; Folding, Sorting and Degradation; Ubiquitin mediated proteolysis [PATH:hsa04120]
SSDB	
Motif	Pfam: ThiF UBA_e1_C UBACT UBA_e1_thiolCys Pyr_redox_2 adh_short
Other DBs	NCBI-GI: 150417996 NCBI-GeneID: 55236 OMIM: 611361 HGNC: 25581 HPRD: 13341 Ensembl: ENSG00000033178 UniProt: A0AVT1
Position	4q13.2
AA seq	1052 aa MEGSEPVAAHQGEEASCSSWGTGSTNKNLPIMSTASVEIDDALYSRQRYVLGDTAMQKMA KSHVFLSGMGGLGLEIAKNLVLAGIKAVTIHDTEKCQAWDLGTNFFLSEDDVVNKRNRAE AVLKHIAELNPYVHVTSSSVPFNETTDLSFLDKYQCVVLTEMKLPLQKKINDFCRSQCPP IKFISADVHGIWSRLFCDFGDEFEVLDTTGEEPKEIFISNITQANPGIVTCLENHPHKLE TGQFLTFREINGMTGLNGSIQQITVISPFSFSIGDTTELEPYLHGGIAVQVKTPKTVFFE SLERQLKHPKCLIVDFSNPEAPLEIHTAMLALDQFQEKYSRKPNVGCQQDSEELLKLATS ISETLEEKPDVNADIVHWLSWTAQGFLSPLAAAVGGVASQEVLKAVTGKFSPLCQWLYLE AADIVESLGKPECEEFLPRGDRYDALRACIGDTLCQKLQNLNIFLVGCGAIGCEMLKNFA LLGVGTSKEKGMITVTDPDLIEKSNLNRQFLFRPHHIQKPKSYTAADATLKINSQIKIDA HLNKVCPTTETIYNDEFYTKQDVIITALDNVEARRYVDSRCLANLRPLLDSGTMGTKGHT

All links

- Pathway (1)
    - KEGG PATHWAY (1)
- Disease (1)
    - OMIM (1)
- Chemical reaction (1)
    - KEGG ENZYME (1)
- Genome (1)
    - KEGG GENOME (1)
- Gene (13)
    - KEGG ORTHOLOGY (1)
    - NCBI-Gene (1)
    - NCBI-GI (7)
    - UniGene (1)
    - HGNC (1)
    - HPRD (1)
    - ENSEMBL-HSA (1)
- Protein sequence (8)
    - UniProt (1)
    - RefSeq(pep) (1)
    - IPI (6)
- DNA sequence (7)
    - RefSeq(nuc) (1)
    - GenBank (3)
    - EMBL (3)
- Protein domain (6)
    - Pfam (6)
- All databases (38)

四、基因组数据库



KEGG GENOME Database

Organisms and ecosystems with genome sequence information

Enter KEGG organism code or use Organism button to select

(Examples) hsa mmu dme cel ath sce eco bsu syn mja

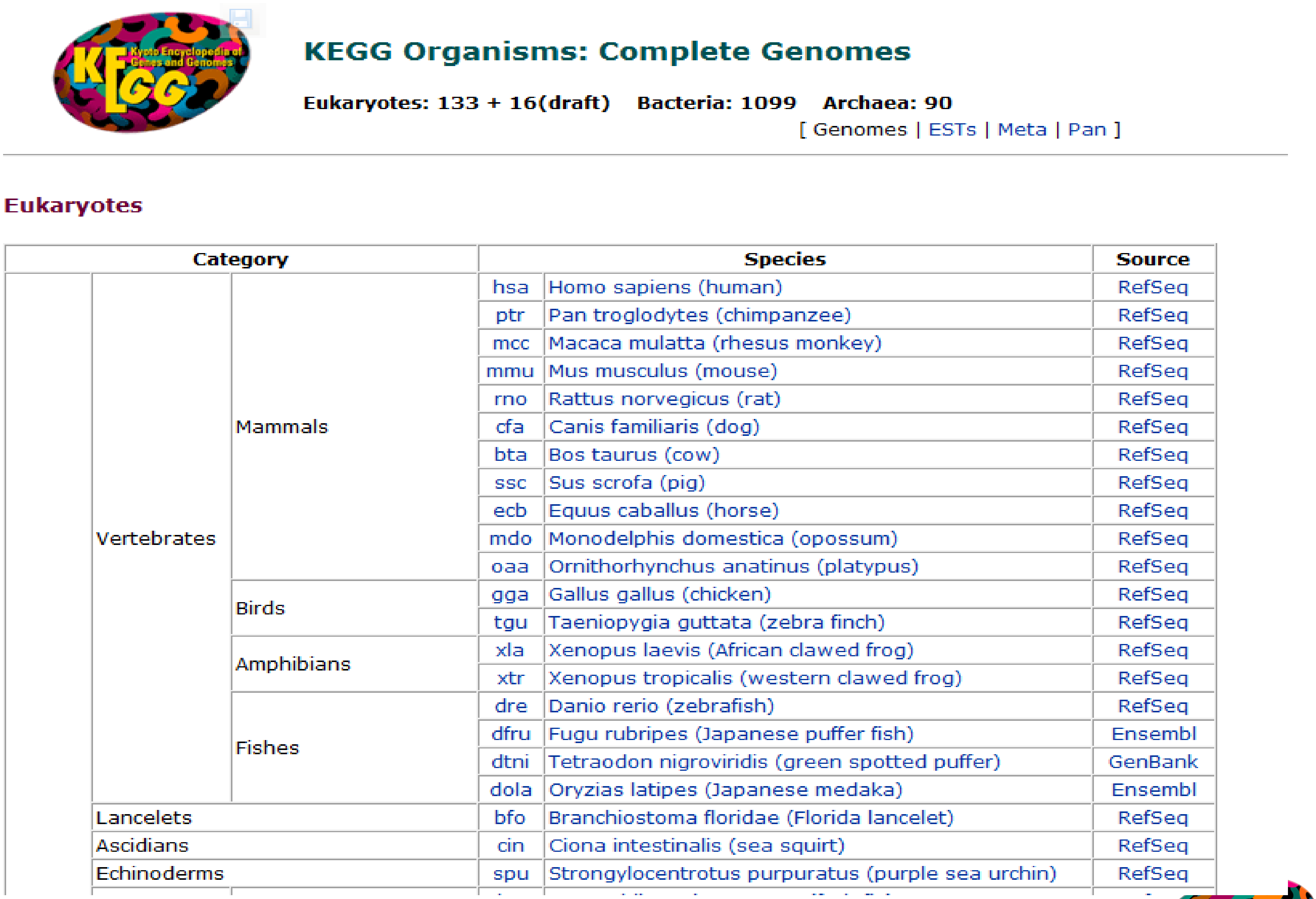
KEGG Organisms and Ecosystems

KEGG GENOME has been a collection of KEGG organisms (see release history), which are the organisms with known complete genome sequences supplemented by those with massive EST datasets (see also KEGG GENES). KEGG GENOME is being extended in three ways. First, it now contains metagenomes representing environmental samples (ecosystems) of genome sequences for multiple species. Second, KEGG organisms with high-quality KOALA annotation can now be combined computationally to examine, for example, reconstructed pathways for pangenomes and organism groups. Third, virus genomes will be integrated and virus genes will be annotated.

Category		Genome	Identifiers	Gene catalog	Annotation
Organisms	Complete genomes	GENOME	T0 numbers or three-letter organism codes	GENES	KOALA
			T1 numbers or four-letter organism codes starting with "d"	DGENES	KAAS
	EST datasets		T2 numbers or four-letter organism codes starting with "e"	EGENES	KAAS
Metagenomes			T3 numbers	MGENES	KAAS
Pangenomes			Generic species names	GENES	KOALA
Viruses		VGENOME	RefSeq identifiers	VGENES	None

Search GENOME for

http://www.genome.jp/kegg/genome.html



KEGG Organisms: Complete Genomes

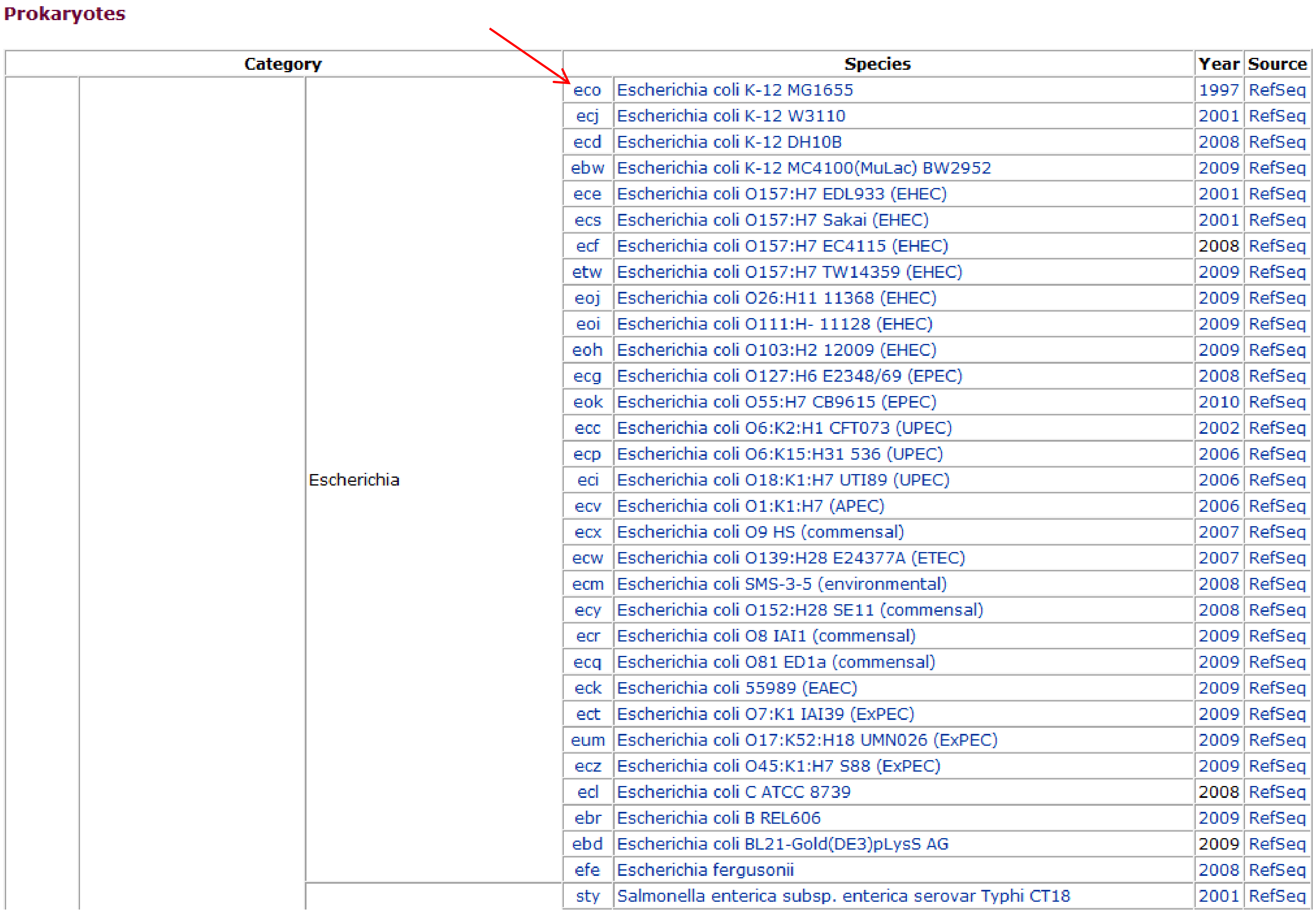
Eukaryotes: 133 + 16(draft) **Bacteria: 1099** **Archaea: 90**

[Genomes | ESTs | Meta | Pan]

Eukaryotes

Category			Species		Source
	Vertebrates	Mammals	hsa	Homo sapiens (human)	RefSeq
			ptr	Pan troglodytes (chimpanzee)	RefSeq
			mcc	Macaca mulatta (rhesus monkey)	RefSeq
			mmu	Mus musculus (mouse)	RefSeq
			rno	Rattus norvegicus (rat)	RefSeq
			cfa	Canis familiaris (dog)	RefSeq
			bta	Bos taurus (cow)	RefSeq
			ssc	Sus scrofa (pig)	RefSeq
			ecb	Equus caballus (horse)	RefSeq
			mdo	Monodelphis domestica (opossum)	RefSeq
			oaa	Ornithorhynchus anatinus (platypus)	RefSeq
		Birds	gga	Gallus gallus (chicken)	RefSeq
			tgu	Taeniopygia guttata (zebra finch)	RefSeq
		Amphibians	xla	Xenopus laevis (African clawed frog)	RefSeq
			xtr	Xenopus tropicalis (western clawed frog)	RefSeq
		Fishes	dre	Danio rerio (zebrafish)	RefSeq
			dfru	Fugu rubripes (Japanese puffer fish)	Ensembl
			dtni	Tetraodon nigroviridis (green spotted puffer)	GenBank
			dola	Oryzias latipes (Japanese medaka)	Ensembl
	Lancelets		bfo	Branchiostoma floridae (Florida lancelet)	RefSeq
	Ascidians		cin	Ciona intestinalis (sea squirt)	RefSeq
	Echinoderms		spu	Strongylocentrotus purpuratus (purple sea urchin)	RefSeq

http://www.genome.jp/kegg/catalog/org_list.html



Prokaryotes

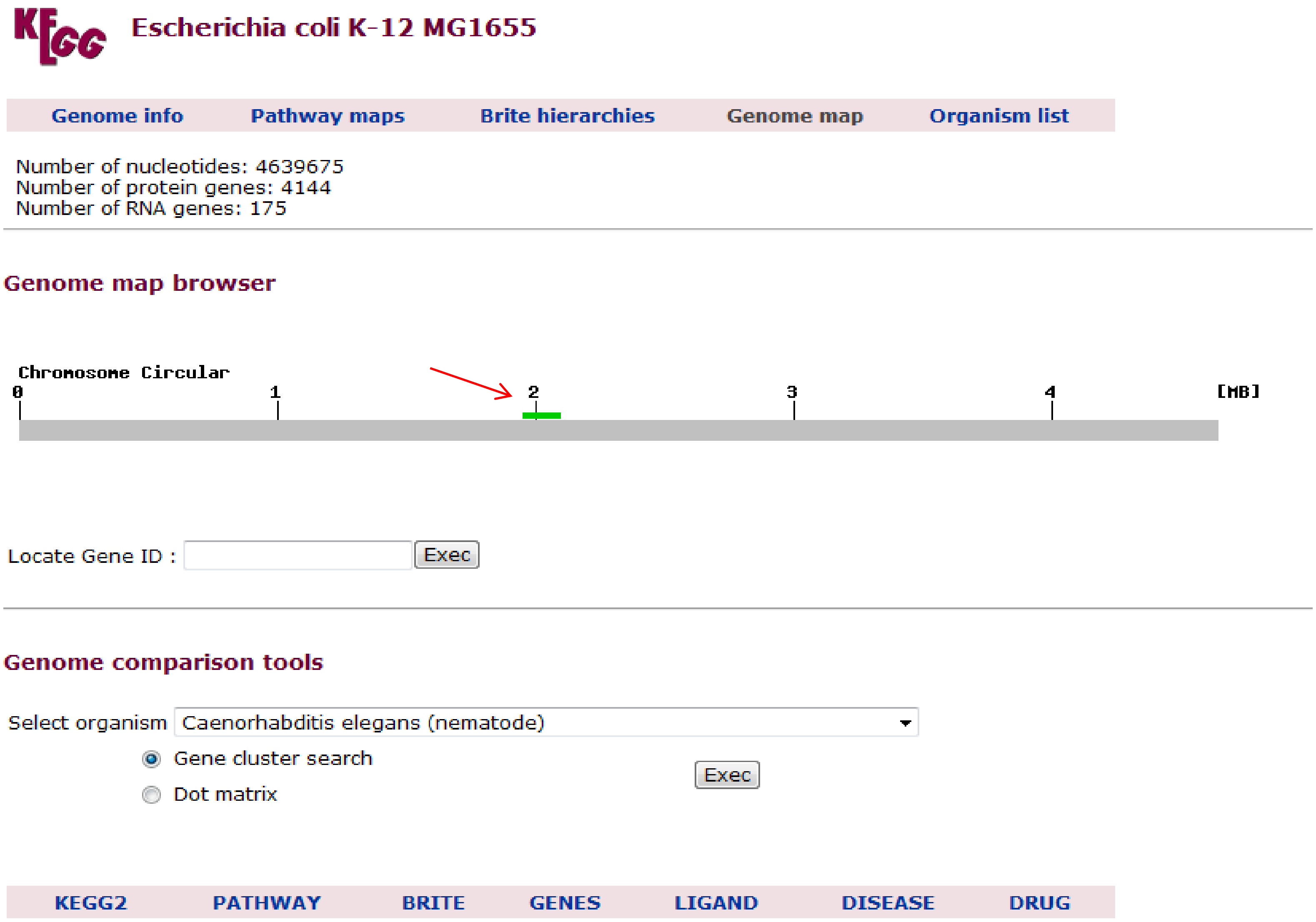
Category			Species		Year	Source
		Escherichia	eco	Escherichia coli K-12 MG1655	1997	RefSeq
			ecj	Escherichia coli K-12 W3110	2001	RefSeq
			ecd	Escherichia coli K-12 DH10B	2008	RefSeq
			ebw	Escherichia coli K-12 MC4100(MuLac) BW2952	2009	RefSeq
			ece	Escherichia coli O157:H7 EDL933 (EHEC)	2001	RefSeq
			ecs	Escherichia coli O157:H7 Sakai (EHEC)	2001	RefSeq
			ecf	Escherichia coli O157:H7 EC4115 (EHEC)	2008	RefSeq
			etw	Escherichia coli O157:H7 TW14359 (EHEC)	2009	RefSeq
			eoj	Escherichia coli O26:H11 11368 (EHEC)	2009	RefSeq
			eoi	Escherichia coli O111:H- 11128 (EHEC)	2009	RefSeq
			eoh	Escherichia coli O103:H2 12009 (EHEC)	2009	RefSeq
			ecg	Escherichia coli O127:H6 E2348/69 (EPEC)	2008	RefSeq
			eok	Escherichia coli O55:H7 CB9615 (EPEC)	2010	RefSeq
			ecc	Escherichia coli O6:K2:H1 CFT073 (UPEC)	2002	RefSeq
			ecp	Escherichia coli O6:K15:H31 536 (UPEC)	2006	RefSeq
			eci	Escherichia coli O18:K1:H7 UTI89 (UPEC)	2006	RefSeq
			ecv	Escherichia coli O1:K1:H7 (APEC)	2006	RefSeq
			ecx	Escherichia coli O9 HS (commensal)	2007	RefSeq
			ecw	Escherichia coli O139:H28 E24377A (ETEC)	2007	RefSeq
			ecm	Escherichia coli SMS-3-5 (environmental)	2008	RefSeq
			ecy	Escherichia coli O152:H28 SE11 (commensal)	2008	RefSeq
			ecr	Escherichia coli O8 IAI1 (commensal)	2009	RefSeq
			ecq	Escherichia coli O81 ED1a (commensal)	2009	RefSeq
			eck	Escherichia coli 55989 (EAEC)	2009	RefSeq
			ect	Escherichia coli O7:K1 IAI39 (ExPEC)	2009	RefSeq
			eum	Escherichia coli O17:K52:H18 UMN026 (ExPEC)	2009	RefSeq
			ecz	Escherichia coli O45:K1:H7 S88 (ExPEC)	2009	RefSeq
			ecl	Escherichia coli C ATCC 8739	2008	RefSeq
			ebr	Escherichia coli B REL606	2009	RefSeq
			ebd	Escherichia coli BL21-Gold(DE3)pLysS AG	2009	RefSeq
			efe	Escherichia fergusonii	2008	RefSeq
			sty	Salmonella enterica subsp. enterica serovar Typhi CT18	2001	RefSeq

选Genome map

Escherichia coli K-12 MG1655

Genome information

T number	T00007
Org code	eco
Aliases	E.coli, ECOLI, 511145
Full name	Escherichia coli K-12 MG1655
Definition	Escherichia coli K-12 MG1655
Annotation	manual
Taxonomy	TAX: 511145
Lineage	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia
Data source	RefSeq (Project:225)
Original DB	Wisconsin, Pasteur, RegulonDB, EcoGene, ECOCYC
Chromosome	Circular
Sequence	RS: NC_000913
Length	4639675
Statistics	Number of nucleotides: 4639675 Number of protein genes: 4144 Number of RNA genes: 175
Reference	PMID: 9278503



Escherichia coli K-12 MG1655

Number of nucleotides: 4639675
Number of protein genes: 4144
Number of RNA genes: 175

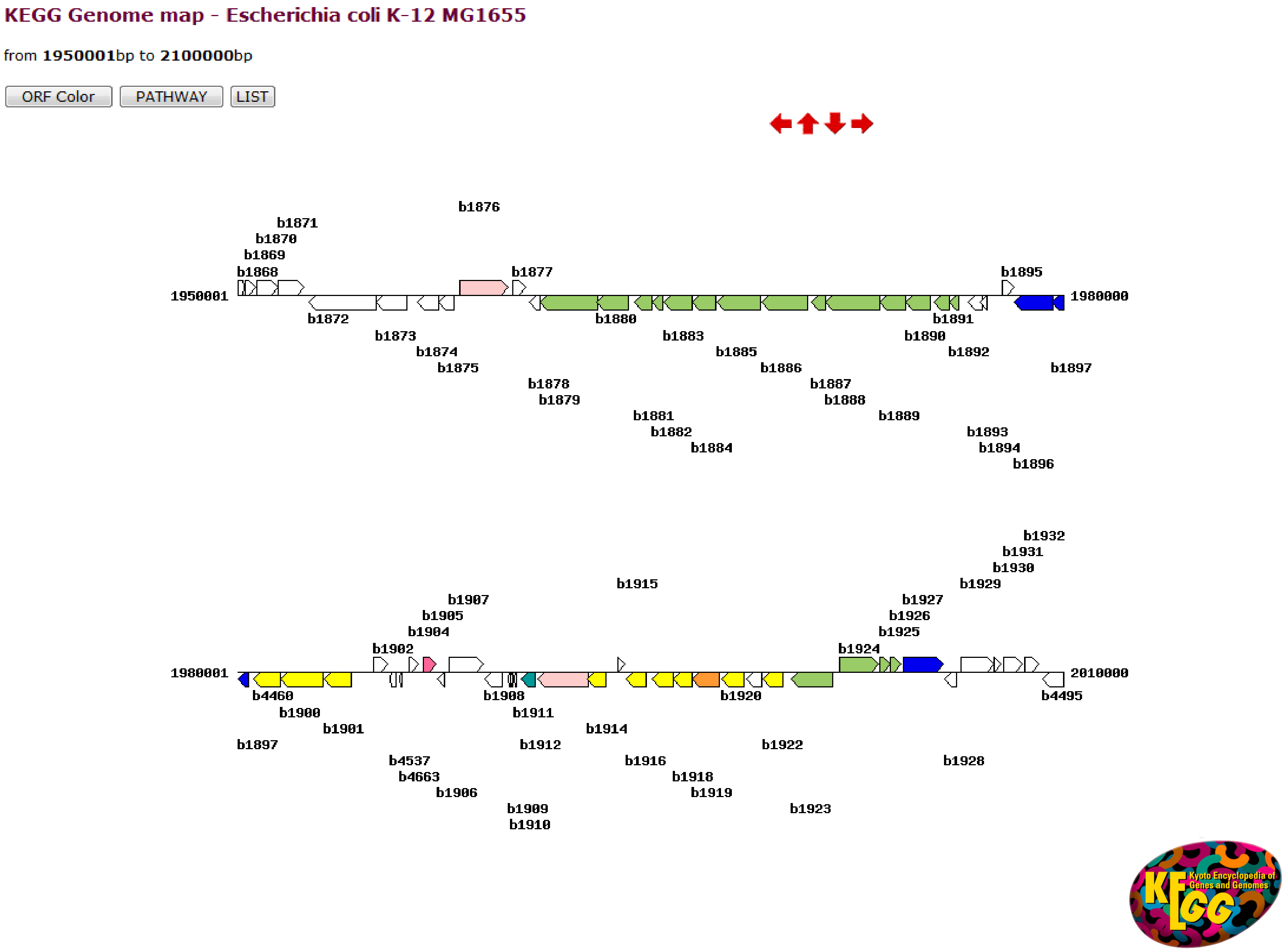
Genome map browser

Locate Gene ID :

Genome comparison tools

Select organism: Caenorhabditis elegans (nematode)

● Gene cluster search
○ Dot matrix



KEGG Genome map - Escherichia coli K-12 MG1655

from **1950001**bp to **2100000**bp

五、 配体数据库

KEGG LIGAND Database

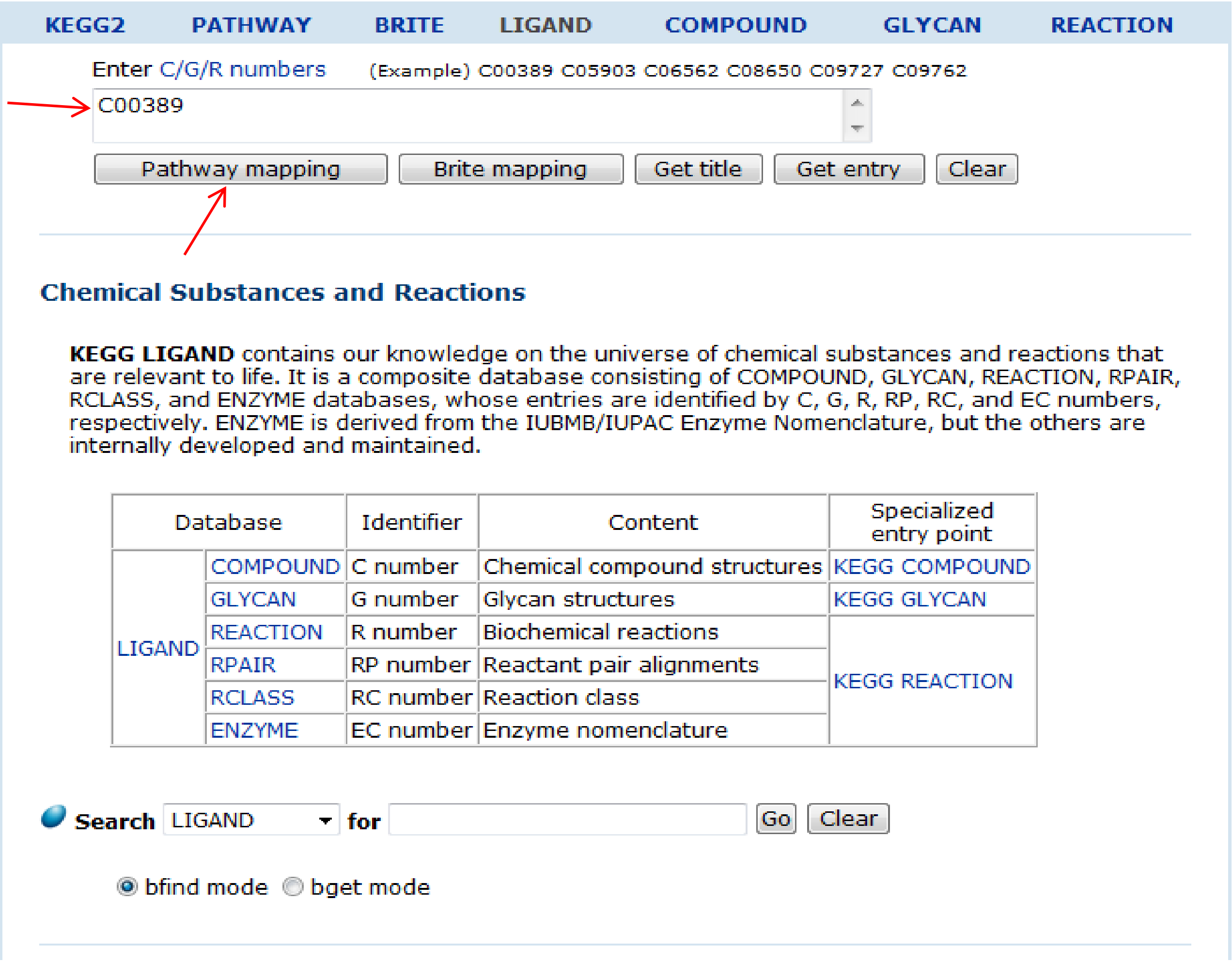
Molecular building blocks of life in the chemical space

小分子编号C00389

或名称

Quercetin

(橡黄素)



Enter C/G/R numbers (Example) C00389 C05903 C06562 C08650 C09727 C09762

C00389

Chemical Substances and Reactions

KEGG LIGAND contains our knowledge on the universe of chemical substances and reactions that are relevant to life. It is a composite database consisting of COMPOUND, GLYCAN, REACTION, RPAIR, RCLASS, and ENZYME databases, whose entries are identified by C, G, R, RP, RC, and EC numbers, respectively. ENZYME is derived from the IUBMB/IUPAC Enzyme Nomenclature, but the others are internally developed and maintained.

Database		Identifier	Content	Specialized entry point
LIGAND	COMPOUND	C number	Chemical compound structures	KEGG COMPOUND
	GLYCAN	G number	Glycan structures	KEGG GLYCAN
	REACTION	R number	Biochemical reactions	KEGG REACTION
	RPAIR	RP number	Reactant pair alignments	
	RCLASS	RC number	Reaction class	
	ENZYME	EC number	Enzyme nomenclature	

Search LIGAND for

● bfind mode ○ bget mode

http://www.genome.jp/kegg/ligand.html

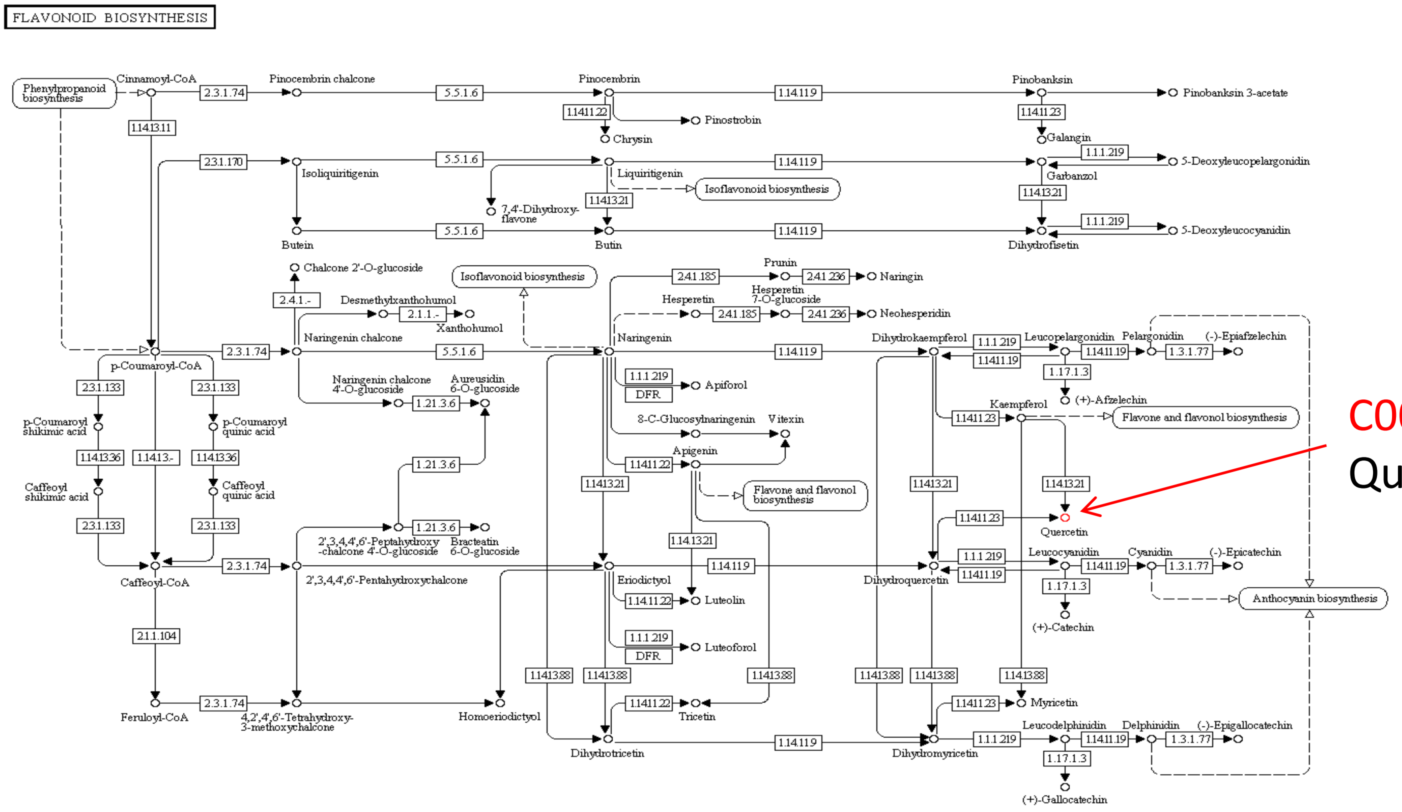
Pathway Search Result

Sort by the pathway list

Show all objects

黄酮合成途径

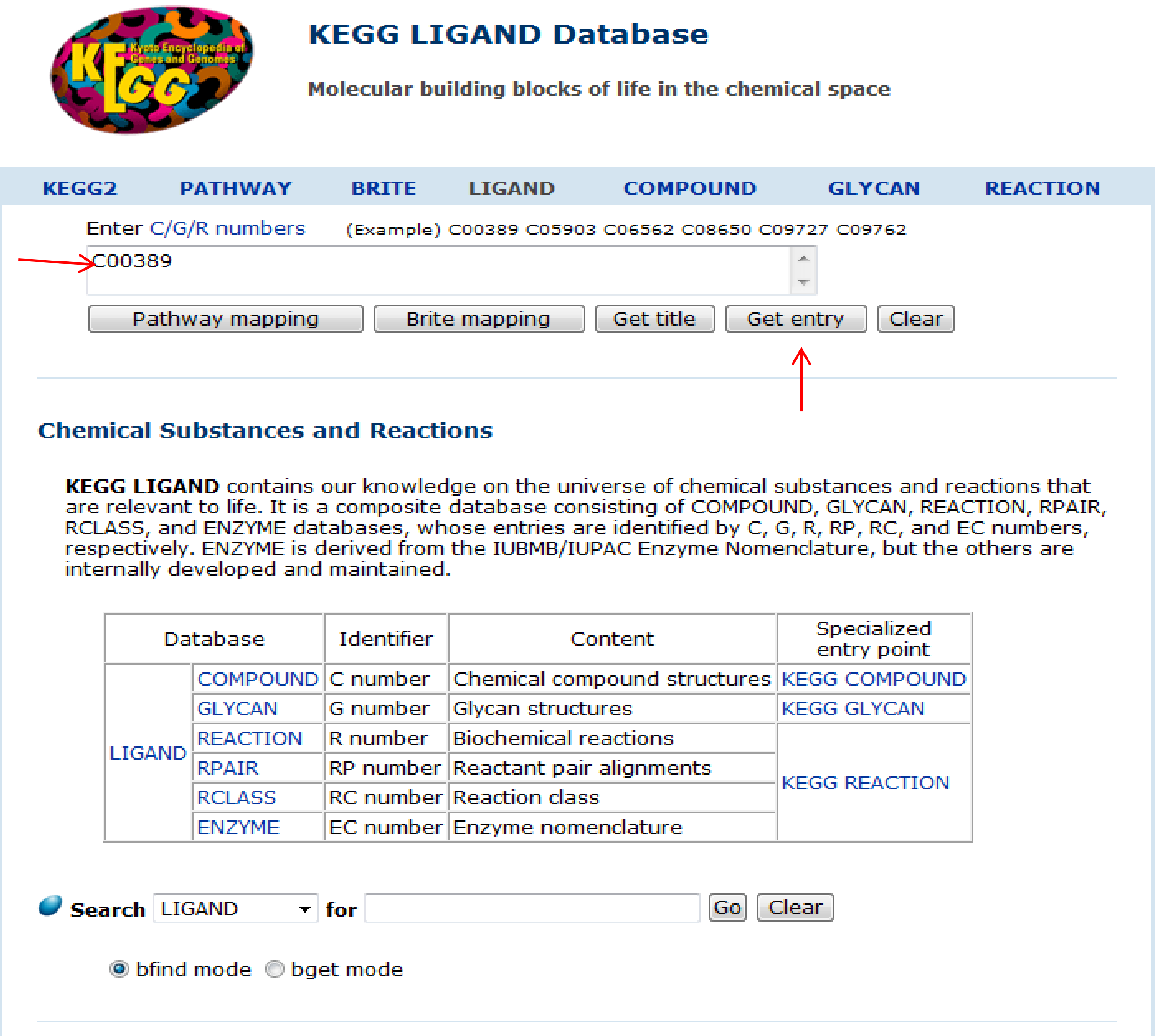
- map00941 Flavonoid biosynthesis (1)
- map00944 Flavone and flavonol biosynthesis (1)
- map01100 Metabolic pathways (1)
- map01110 Biosynthesis of secondary metabolites (1)



C00389

Quercetin

小分子编号C00389
或名称
Quercetin
(橡黄素)



KEGG LIGAND Database

Molecular building blocks of life in the chemical space

Enter C/G/R numbers (Example) C00389 C05903 C06562 C08650 C09727 C09762

C00389

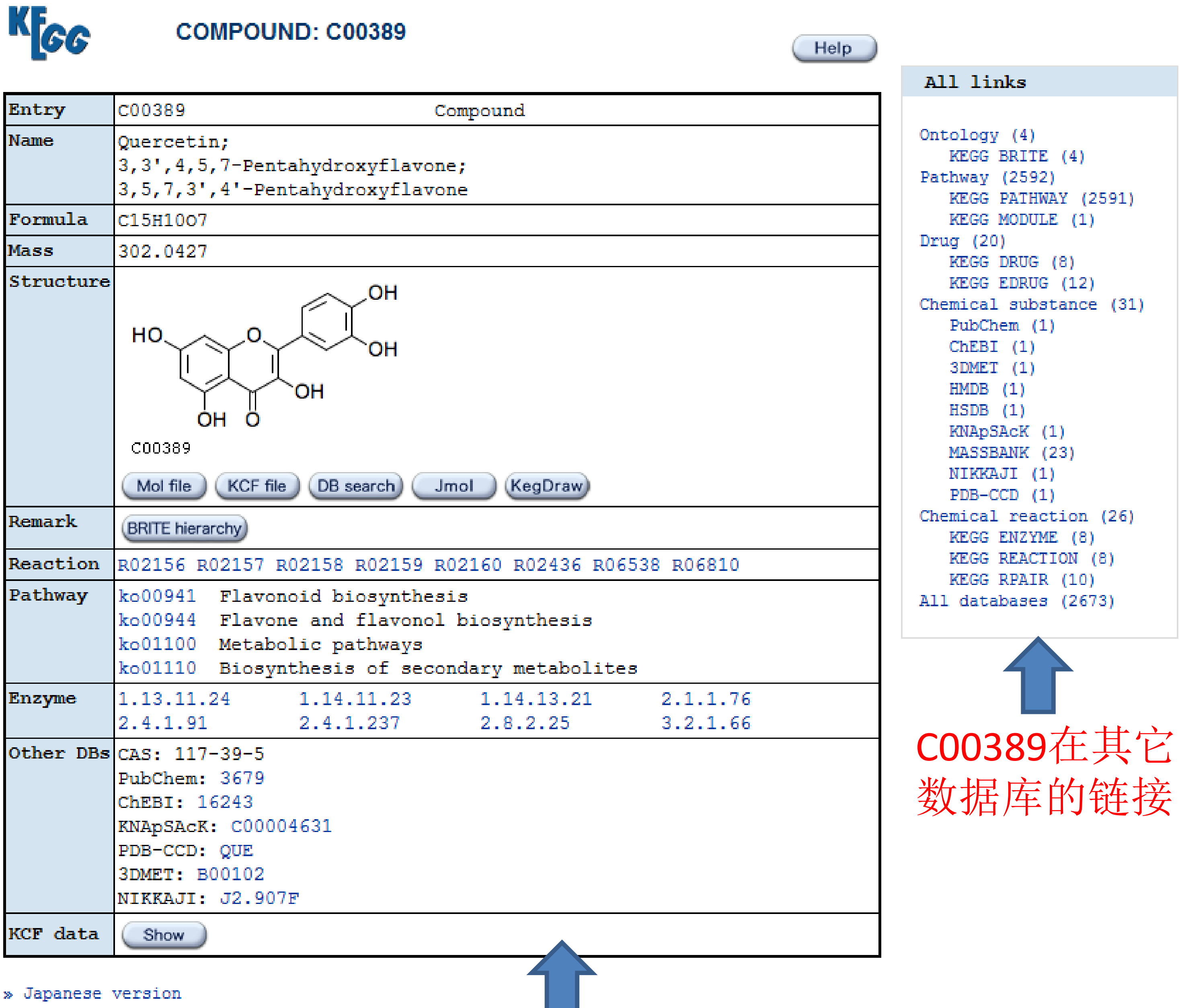
Chemical Substances and Reactions

KEGG LIGAND contains our knowledge on the universe of chemical substances and reactions that are relevant to life. It is a composite database consisting of COMPOUND, GLYCAN, REACTION, RPAIR, RCLASS, and ENZYME databases, whose entries are identified by C, G, R, RP, RC, and EC numbers, respectively. ENZYME is derived from the IUBMB/IUPAC Enzyme Nomenclature, but the others are internally developed and maintained.

Database		Identifier	Content	Specialized entry point
LIGAND	COMPOUND	C number	Chemical compound structures	KEGG COMPOUND
	GLYCAN	G number	Glycan structures	KEGG GLYCAN
	REACTION	R number	Biochemical reactions	KEGG REACTION
	RPAIR	RP number	Reactant pair alignments	
	RCLASS	RC number	Reaction class	
	ENZYME	EC number	Enzyme nomenclature	

Search LIGAND **for**

● bfind mode ○ bget mode



COMPOUND: C00389

Entry	C00389 Compound
Name	Quercetin; 3,3',4,5,7-Pentahydroxyflavone; 3,5,7,3',4'-Pentahydroxyflavone
Formula	C15H10O7
Mass	302.0427
Structure	C00389 Mol file KCF file DB search Jmol KegDraw
Remark	BRITE hierarchy
Reaction	R02156 R02157 R02158 R02159 R02160 R02436 R06538 R06810
Pathway	ko00941 Flavonoid biosynthesis ko00944 Flavone and flavonol biosynthesis ko01100 Metabolic pathways ko01110 Biosynthesis of secondary metabolites
Enzyme	1.13.11.24 1.14.11.23 1.14.13.21 2.1.1.76 2.4.1.91 2.4.1.237 2.8.2.25 3.2.1.66
Other DBs	CAS: 117-39-5 PubChem: 3679 ChEBI: 16243 KNApSAcK: C00004631 PDB-CCD: QUE 3DMET: B00102 NIKKAJI: J2.907F
KCF data	Show

» Japanese version

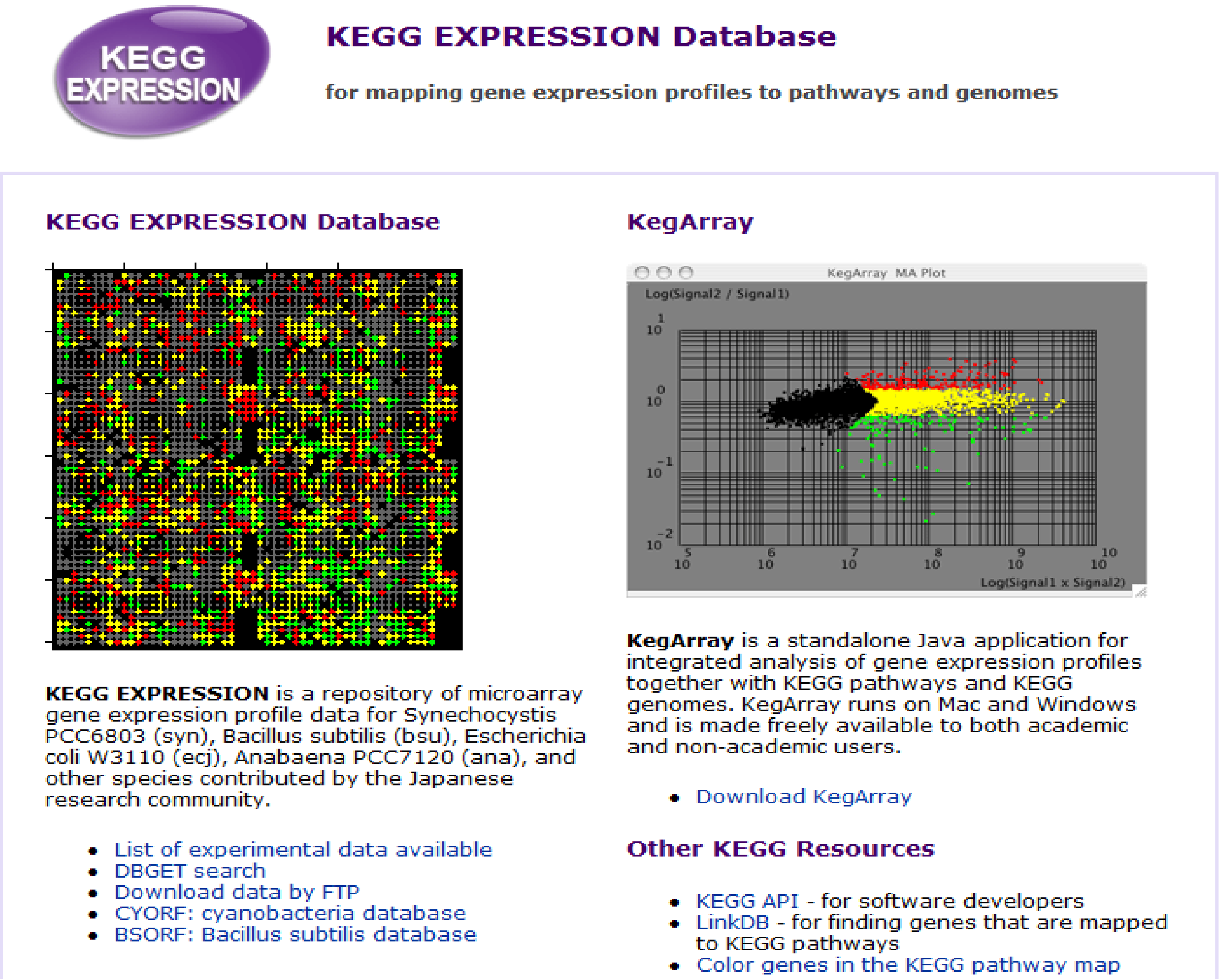
All links

- Ontology (4)
  - KEGG BRITE (4)
- Pathway (2592)
  - KEGG PATHWAY (2591)
  - KEGG MODULE (1)
- Drug (20)
  - KEGG DRUG (8)
  - KEGG EDRUG (12)
- Chemical substance (31)
  - PubChem (1)
  - ChEBI (1)
  - 3DMET (1)
  - HMDB (1)
  - HSDB (1)
  - KNApSAcK (1)
  - MASSBANK (23)
  - NIKKAJI (1)
  - PDB-CCD (1)
- Chemical reaction (26)
  - KEGG ENZYME (8)
  - KEGG REACTION (8)
  - KEGG RPAIR (10)
- All databases (2673)

C00389在其它数据库的链接

小分子C00389的详细信息

六、基因芯片数据查询



KEGG EXPRESSION Database

for mapping gene expression profiles to pathways and genomes

KEGG EXPRESSION Database

KEGG EXPRESSION is a repository of microarray gene expression profile data for Synechocystis PCC6803 (syn), Bacillus subtilis (bsu), Escherichia coli W3110 (ecj), Anabaena PCC7120 (ana), and other species contributed by the Japanese research community.

- List of experimental data available
- DBGET search
- Download data by FTP
- CYORF: cyanobacteria database
- BSORF: Bacillus subtilis database

KegArray

KegArray is a standalone Java application for integrated analysis of gene expression profiles together with KEGG pathways and KEGG genomes. KegArray runs on Mac and Windows and is made freely available to both academic and non-academic users.

- Download KegArray

Other KEGG Resources

- KEGG API - for software developers
- LinkDB - for finding genes that are mapped to KEGG pathways
- Color genes in the KEGG pathway map

http://www.genome.jp/kegg/expression/

KEGG Kyoto Encyclopedia of Genes and Genomes

Thank You !