Introduction to KEGG

Susumu Goto, Masahiro Hattori, Wataru Honda, Junko Yabuzaki

Kyoto University, Bioinformatics Center

Today's Menu
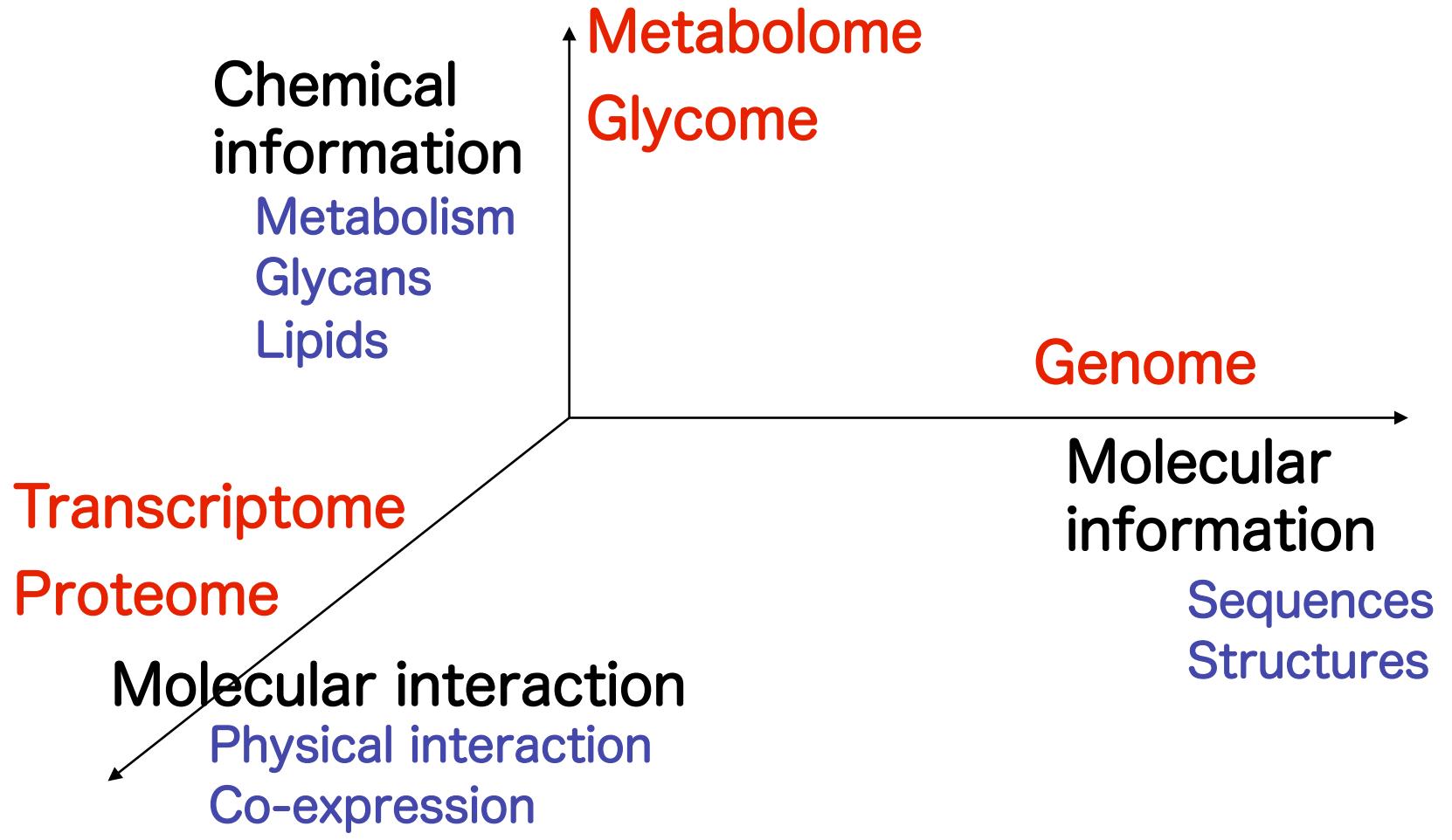
Morning session

- Brief history and overview of KEGG and GenomeNet (Goto)
- KEGG resources related to genomic information (Goto)
- KEGG resources related to systems information (Honda)
- KEGG resources related to chemical information (Hattori)

Afternoon session (laboratory work)

- Searching genomic information in KEGG (Yabuzaki)
- Assembling cDNA sequences and annotating functions (Goto/Yabuzaki)
- Handling microarray data for mapping KEGG pathways (Goto/Honda)
- Searching and computing pathways and chemical information in KEGG (Honda/Hattori)

Background

Metabolome
Glycome

Chemical
information
Metabolism
Glycans
Lipids

Genome

Molecular
information
Sequences
Structures

Transcriptome

Proteome

Molecular interaction
Physical interaction
Co-expression

Various types of omics data are now available

Overview
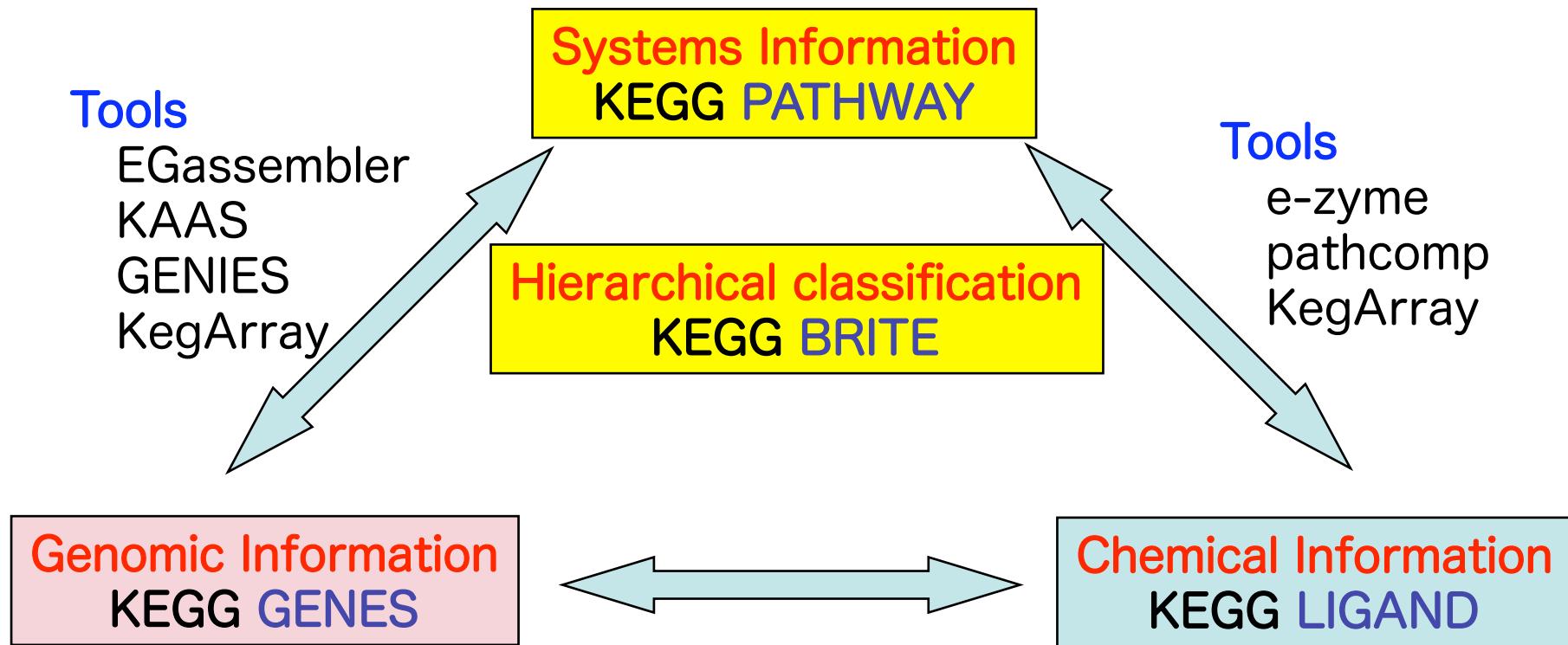
KEGG (http://www.genome.jp/kegg/)

- Kyoto Encyclopedia of Genes and Genomes
- Integrated database of biological systems, genetic building blocks and chemical building blocks

References

- Kanehisa, M., *et al*. KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 36, D480 (2008)
- Kanehisa, M., *et al*. From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.* 34, D354 (2006)

Four Major Components of KEGG

Systems Information
KEGG PATHWAY

Tools
EGassembler
KAAS
GENIES
KegArray

Hierarchical classification
KEGG BRITE

Tools
e-zyme
pathcomp
KegArray

Genomic Information
KEGG GENES

Chemical Information
KEGG LIGAND

Reconstructing biological phenomenon with various omics data and researcher's knowledge

KEGG history with ID system

Release	Database	Object identifier
1995	KEGG PATHWAY	map number
	KEGG GENOME	organism code (T number)
	KEGG GENES	locus_tag / NCBI GeneID
	KEGG ENZYME	EC number
	KEGG COMPOUND	C number
2001	KEGG REACTION	R number
2002	KEGG ORTHOLOGY	K number
2003	KEGG GLYCAN	G number
2004	KEGG RPAIR	A number
2005	KEGG BRITE	br number
	KEGG DRUG	D number
2008	KEGG MODULE	M number
	KEGG DISEASE	H number

Kanehisa, M., *et al*. *Nucleic Acids Res.* 36, D480 (2008)

Databases in GenomeNet

Established in 1991 under the Japanese Human
Genome Project
http://www.genome.jp/
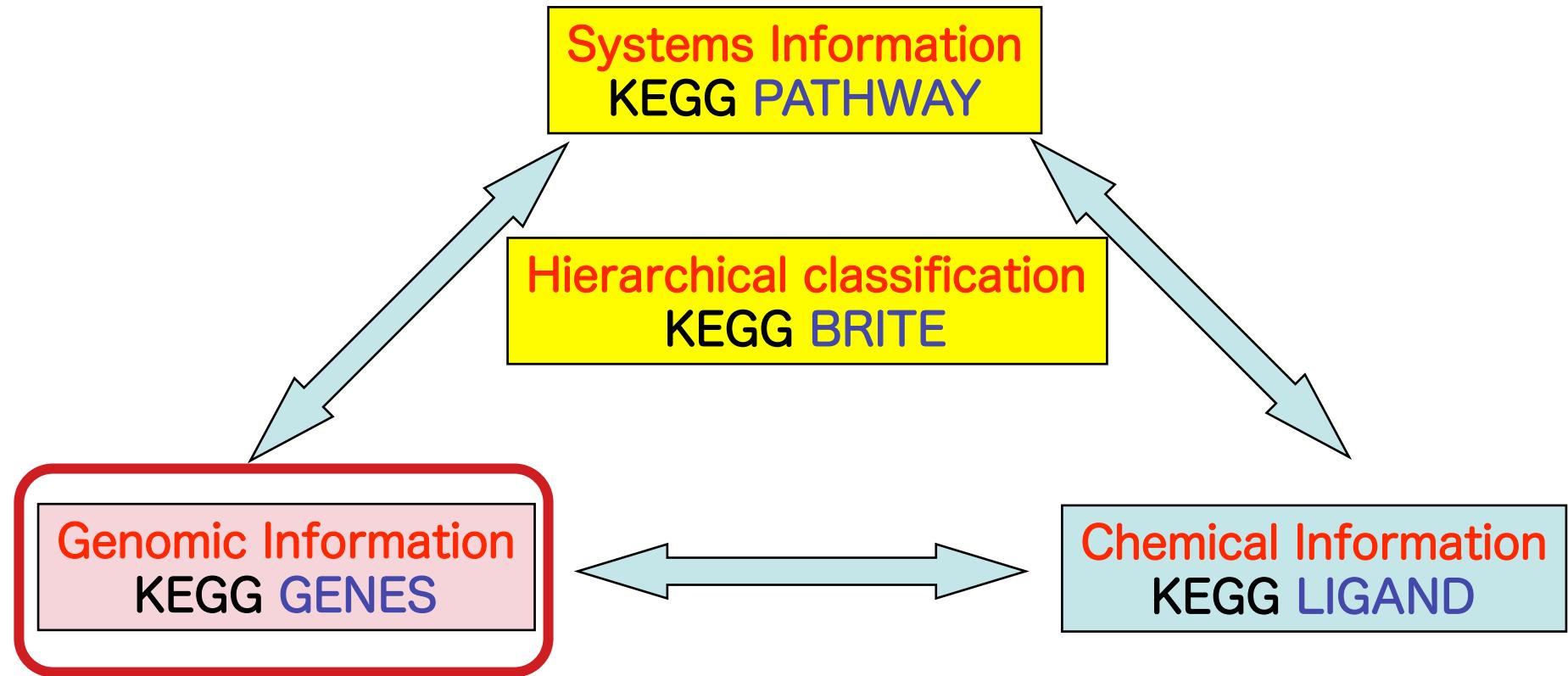
A network of databases and computational
services for genome research and related
research areas in biomedical sciences

>30 databases including KEGG, GenBank,
UniProt, PDB, …

>500 databases linking from/to GenomeNet

Unique ID for each entry *database:entry*
e.g. compound:C00002 for ATP

Four Major Components of KEGG

Systems Information
KEGG PATHWAY

Hierarchical classification
KEGG BRITE

Genomic Information
KEGG GENES

Chemical Information
KEGG LIGAND

Reconstructing biological phenomenon with various omics data and researcher's knowledge

KEGG GENES Database

Molecular building blocks of life
in the genomic space

KEGG GENES

A collection of gene catalogs for all complete genomes and some partial genomes generated from publicly available resources, mostly NCBI RefSeq

Genomes in the taxonomy	Gene catalog	Genome	Remark
KEGG organisms	GENES	GENOME	High-quality genomes
	DGENES		Draft genomes
	EGENES		EST contigs
Complete viral genomes	VGENES	VGENOME	
Complete mitochondrial genomes Complete plastid genomes Complete nucleomorph genomes	OGENES		Available only in DBGET

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

KEGG GENES main components

KEGG Orthology (KO)	Ortholog groups linked to PATHWAY and BRITE
GENES	Gene catalogs of complete genomes with manual functional annotation
DGENES	Gene catalogs of draft genomes with automatic functional annotation
EGENES	Consensus contigs of EST data with automatic functional annotation
SSDB	Sequence similarity with best-hit information for identifying ortholog/paralogs

GENES, DGENES, EGENES: Genomic information

KEGG Orthology, SSDB: Relationship

Organisms in KEGG GENES

KEGG Organisms

Eukaryotes: 49+20+29   Bacteria: 521   Archaea: 45
1995: 2   1996: 2   1997: 8   1998: 7   1999: 6   2000: 18   2001: 28
2002: 38   2003: 46   2004: 67   2005: 90   2006: 138   2007: 133

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

Eukaryotes

Category			Species		Completed	Source
Animals	Vertebrates	Mammals	Homo sapiens (human)	hsa		RefSeq
			Pan troglodytes (chimpanzee)	ptr		RefSeq
			Macaca mulatta (rhesus monkey)	mcc		RefSeq
			Mus musculus (mouse)	mmu		RefSeq
			Rattus norvegicus (rat)	rno		RefSeq
			Canis familiaris (dog)	cfa		RefSeq
			Bos taurus (cow)	bta		RefSeq
			Sus scrofa (pig)	ssc		RefSeq
			Monodelphis domestica (opossum)	mdo		RefSeq
		Bird	Gallus gallus (chicken)	gga		RefSeq
		Amphibians	Xenopus laevis (African clawed frog)	xla		GenBank
			Xenopus tropicalis (western clawed frog)	xtr		RefSeq
		Fishes	Danio rerio (zebrafish)	dre		RefSeq
			Fugu rubripes (Japanese puffer fish)	dfru		Ensembl
			Tetraodon nigroviridis (green spotted puffer)	dtni		GenBank
	Ascidians		Ciona intestinalis (sea squirt)	dcin		JGI
			Ciona savignyi (EST)	ecsv		dbEST
			Molgula tectiformis (EST)	emte		dbEST
	Echinoderms		Strongylocentrotus purpuratus (purple sea urchin)	spu		RefSeq
			Strongylocentrotus purpuratus (purple sea urchin) (EST)	espu		dbEST

Organisms in KEGG GENES

141 Eukaryotes

    43 Animals (Complete / Draft Genomes)
    54 Plants (Mostly EST Consensus Contigs)
    27 Fungi (Mostly Complete Genomes)
    17 Protists (Mostly Complete Genomes)

633 Eubacteria (All Complete Genomes)

52 Archea (All Complete Genomes)

Total: **826** organisms (30 May 2008)

Organisms in KEGG GENES

- Hierarchical classification of organisms is based on Taxonomy database in NCBI

- 3-4 letter code / T number

$$
\begin{array}{rcll}
\textit{Homo sapiens} & \Rightarrow & \text{hsa} & \\
\textit{Escherichia coli K-12 MG1655} & \Rightarrow & \text{eco} & \textbf{\textit{GENES}} \\
\textit{Escherichia coli K-12 W3110} & \Rightarrow & \text{ecj} & \\
\textit{Fugu rubripes} & \Rightarrow & \text{dfru} & \textbf{\textit{DGENES}} \\
\textit{Hordeum vulgare (barley) (EST)} & \Rightarrow & \text{ehvu} & \textbf{\textit{EGENES}} \\
\end{array}
$$

Entry of GENES

KEGG

Escherichia coli K-12 MG1655: b2843

[Help]

Entry	b2843　　　　　CDS　　　　E.coli
Gene name	kduI, yqeE
Definition	predicted 5-keto 4-deoxyuronate isomerase [EC:5.3.1.17]
Orthology	KO: K01815　4-deoxy-L-threo-5-hexosulose-uronate ketol-isomerase
Pathway	PATH: eco00040　Pentose and glucuronate interconversions
Class	[BRITE hierarchy]
SSDB	[Ortholog] [Paralog] [Gene cluster]
Motif	Pfam: KduI IolB [Motif]
Other DBs	Wisconsin: b2843 Colibri: kduI RegulonDB: B2843 NCBI-GI: 16130747 NCBI-GeneID: 947319 UniProt: Q46938
LinkDB	[All DBs]
Structure	PDB: 1XRU 1X8M [Thumbnails] [Jmol]
Position	complement(2981310..2982146) [Genome map]
AA seq	278 aa　[AA seq] [DB search] MDVRQSIHSAHAKTLDTQGLRNEFLVEKVFVADEYTMVYSHIDRIIVGGIMPITKTVSVG GEVGKQLGVSYFLERRELGVINIGGAGTITVDGQCYEIGHRDALYVGKGAKEVVFASIDT

Entry field ➤

Entry ID
Entry types
 (CDS, RNA, Contig, etc. ···)
Organism name

Escherichia coli K-12 MG1655: b2843

Help

Entry	b2843 CDS E.coli
Gene name	kduI, yqeE
Definition	predicted 5-keto 4-deoxyuronate isomerase [EC:5.3.1.17]
Orthology	KO: K01815 4-deoxy-L-threo-5-hexosulose-uronate ketol-isomerase
Pathway	PATH: eco00040 Pentose and glucuronate interconversions
Class	BRITE hierarchy
SSDB	Ortholog Paralog Gene cluster
Motif	Pfam: KduI IolB Motif
Other DBs	Wisconsin: b2843 Colibri: kduI RegulonDB: B2843 NCBI-GI: 16130747 NCBI-GeneID: 947319 UniProt: Q46938
LinkDB	All DBs
Structure	PDB: 1XRU 1X8M Thumbnails Jmol
Position	complement(2981310..2982146) Genome map
AA seq	278 aa AA seq DB search MDVRQSIHSAHAKTLDTQGLRNEFLVEKVFVADEYTMVYSHIDRIIVGGIMPITKTVSVG GEVGKQLGVSYFLERRELGVINIGGAGTITVDGQCYEIGHRDALYVGKGAKEVVFASIDT GTPAKFYYNCAPAHTTYPTKKVTPDEVSPVTLGDNLTSNRRTINKYFVPDVLETCQLSMG LTELAPGNLWNTMPCHTHERRMEVYFYFNMDDDACVFHMMGQPQETRHIVMHNEQAVISP SWSIHSGVGTKAYTFIWGMVGENQVFDDMDHVAVKDLR
NT seq	837 nt NT seq +upstream 0 nt +downstream 0 nt gtggacgtaagacagagcatccacagtgcgcacgcaaaaacgctggatacccaagggctg

KEGG   **Escherichia coli K-12 MG1655: b2843**

Gene name field ▶

Names and synonyms
of geens and/or
proteins

Entry	b2843	CDS	E.coli
Gene name	kduI, yqeE		
Definition	predicted 5-keto 4-deoxyuronate isomerase [EC:5.3.1.17]		
Orthology	KO: K01815 4-deoxy-L-threo-5-hexosulose-uronate ketol-isomerase		
Pathway	PATH: eco00040 Pentose and glucuronate interconversions		
Class	BRITE hierarchy		
SSDB	Ortholog Paralog Gene cluster		
Motif	Pfam: KduI IolB Motif		
Other DBs	Wisconsin: b2843 Colibri: kduI RegulonDB: B2843 NCBI-GI: 16130747 NCBI-GeneID: 947319 UniProt: Q46938		
LinkDB	All DBs		
Structure	PDB: 1XRU 1X8M Thumbnails Jmol		
Position	complement(2981310..2982146) Genome map		
AA seq	278 aa AA seq DB search MDVRQSIHSAHAKTLDTQGLRNEFLVEKVFVADEYTMVYSHIDRIIVGGIMPITKTVSVG GEVGKQLGVSYFLERRELGVINIGGAGTITVDGQCYEIGHRDALYVGKGAKEVVFASIDT GTPAKFYYNCAPAHTTYPTKKVTPDEVSPVTLGDNLTSNRRTINKYFVPDVLETCQLSMG LTELAPGNLWNTMPCHTHERRMEVYFYFNMDDDACVFHMMGQPQETRHIVMHNEQAVISP SWSIHSGVGTKAYTFIWGMVGENQVFDDMDHVAVKDLR		
NT seq	837 nt NT seq +upstream 0 nt +downstream 0 nt gtggacgtaagacagagcatccacagtgcgcacgcaaaaacgctggatacccaagggctg		

Definition field

Functional annotation assigned by original genome project

KEGG Escherichia coli K-12 MG1655: b2843

Help

Entry	b2843 CDS E.coli
Gene name	kduI, yqeE
Definition	predicted 5-keto 4-deoxyuronate isomerase [EC:5.3.1.17]
Orthology	KO: K01815 4-deoxy-L-threo-5-hexosulose-uronate ketol-isomerase
Pathway	PATH: eco00040 Pentose and glucuronate interconversions
Class	BRITE hierarchy
SSDB	Ortholog Paralog Gene cluster
Motif	Pfam: KduI IolB Motif
Other DBs	Wisconsin: b2843 Colibri: kduI RegulonDB: B2843 NCBI-GI: 16130747 NCBI-GeneID: 947319 UniProt: Q46938
LinkDB	All DBs
Structure	PDB: 1XRU 1X8M Thumbnails Jmol
Position	complement(2981310..2982146) Genome map
AA seq	278 aa AA seq DB search MDVRQSIHSAHAKTLDTQGLRNEFLVEKVFVADEYTMVYSHIDRIIVGGIMPITKTVSVG GEVGKQLGVSYFLERRELGVINIGGAGTITVDGQCYEIGHRDALYVGKGAKEVVFASIDT GTPAKFYYNCAPAHTTYPTKKVTPDEVSPVTLGDNLTSNRRTINKYFVPDVLETCQLSMG LTELAPGNLWNTMPCHTHERRMEVYFYFNMDDDACVFHMMGQPQETRHIVMHNEQAVISP SWSIHSGVGTKAYTFIWGMVGENQVFDDMDHVAVKDLR
NT seq	837 nt NT seq +upstream 0 nt +downstream 0 nt gtggacgtaagacagagcatccacagtgcgcacgcaaaaacgctggatacccaagggctg

KEGG

Escherichia coli K-12 MG1655: b2843

Entry	b2843　　　　　　　CDS　　　　E.coli
Gene name	kduI, yqeE
Definition	predicted 5-keto 4-deoxyuronate isomerase [EC:5.3.1.17]
Orthology	KO: K01815　4-deoxy-L-threo-5-hexosulose-uronate ketol-isomerase
Pathway	PATH: eco00040　Pentose and glucuronate interconversions
Class	BRITE hierarchy
SSDB	Ortholog　Paralog　Gene cluster
Motif	Pfam: KduI IolB Motif
Other DBs	Wisconsin: b2843 Colibri: kduI RegulonDB: B2843 NCBI-GI: 16130747 NCBI-GeneID: 947319 UniProt: Q46938
LinkDB	All DBs
Structure	PDB: 1XRU 1X8M Thumbnails Jmol
Position	complement(2981310..2982146)　Genome map
AA seq	278 aa　AA seq　DB search MDVRQSIHSAHAKTLDTQGLRNEFLVEKVFVADEYTMVYSHIDRIIVGGIMPITKTVSVG GEVGKQLGVSYFLERRELGVINIGGAGTITVDGQCYEIGHRDALYVGKGAKEVVFASIDT GTPAKFYYNCAPAHTTYPTKKVTPDEVSPVTLGDNLTSNRRTINKYFVPDVLETCQLSMG LTELAPGNLWNTMPCHTHERRMEVYFYFNMDDDACVFHMMGQPQETRHIVMHNEQAVISP SWSIHSGVGTKAYTFIWGMVGENQVFDDMDHVAVKDLR
NT seq	837 nt　NT seq　+upstream 0 nt　+downstream 0 nt gtggacgtaagacagagcatccacagtgcgcacgcaaaaacgctggatacccaagggctg

Orthology field

Ortholog annotation assigned by KEGG project

KEGG

Escherichia coli K-12 MG1655: b2843

Help

Entry	b2843 CDS E.coli
Gene name	kduI, yqeE
Definition	predicted 5-keto 4-deoxyuronate isomerase [EC:5.3.1.17]
Orthology	KO: K01815 4-deoxy-L-threo-5-hexosulose-uronate ketol-isomerase
Pathway	PATH: eco00040 Pentose and glucuronate interconversions
Class	BRITE hierarchy
SSDB	Ortholog Paralog Gene cluster
Motif	Pfam: KduI IolB Motif
Other DBs	Wisconsin: b2843 Colibri: kduI RegulonDB: B2843 NCBI-GI: 16130747 NCBI-GeneID: 947319 UniProt: Q46938
LinkDB	All DBs
Structure	PDB: 1XRU 1X8M Thumbnails Jmol
Position	complement(2981310..2982146) Genome map
AA seq	278 aa AA seq DB search MDVRQSIHSAHAKTLDTQGLRNEFLVEKVFVADEYTMVYSHIDRIIVGGIMPITKTVSVG GEVGKQLGVSYFLERRELGVINIGGAGTITVDGQCYEIGHRDALYVGKGAKEVVFASIDT GTPAKFYYNCAPAHTTYPTKKVTPDEVSPVTLGDNLTSNRRTINKYFVPDVLETCQLSMG LTELAPGNLWNTMPCHTHERRMEVYFYFNMDDDACVFHMMGQPQETRHIVMHNEQAVISP SWSIHSGVGTKAYTFIWGMVGENQVFDDMDHVAVKDLR
NT seq	837 nt NT seq +upstream [0] nt +downstream [0] nt gtggacgtaagacagagcatccacagtgcgcacgcaaaaacgctggatacccaagggctg

Pathway field

Links to pathway maps including this gene

PENTOSE AND GLUCURONATE INTERCONVERSIONS

KEGG   Escherichia coli K-12 MG1655: b2843

Help

Entry	b2843 CDS E.coli
Gene name	kduI, yqeE
Definition	predicted 5-keto 4-deoxyuronate isomerase [EC:5.3.1.17]
Orthology	KO: K01815 4-deoxy-L-threo-5-hexosulose-uronate ketol-isomerase
Pathway	PATH: eco00040 Pentose and glucuronate interconversions
Class	[BRITE hierarchy]
SSDB	[Ortholog] [Paralog] [Gene cluster]
Motif	Pfam: KduI IolB [Motif]
Other DBs	Wisconsin: b2843 Colibri: kduI RegulonDB: B2843

Class field

➤

Link to BRITE
functional categories

Search BRITE hierarchies

```
KEGG Orthology (KO) [BR:eco00001]
  01100 Metabolism
    01110 Carbohydrate Metabolism
      00040 Pentose and glucuronate interconversions [PATH:eco00040]
        b2843 kduI, ECK2841, JW2811, yqeE; predicted 5-keto 4-deoxyuronate isomerase; K01815 4-deoxy-L-threo-5-h

Enzymes [BR:eco01000]
  5. Isomerases
    5.3  Intramolecular oxidoreductases
      5.3.1  Interconverting aldoses and ketoses
        5.3.1.17  4-deoxy-L-threo-5-hexosulose-uronate ketol-isomerase
          b2843 kduI, ECK2841, JW2811, yqeE; predicted 5-keto 4-deoxyuronate isomerase; K01815 4-deoxy-L-threo-5
```

Position	complement(2981310..2982146) [Genome map]
AA seq	278 aa [AA seq] [DB search] MDVRQSIHSAHAKTLDTQGLRNEFLVEKVFVADEYTMVYSHIDRIIVGGIMPITKTVSVG GEVGKQLGVSYFLERRELGVINIGGAGTITVDGQCYEIGHRDALYVGKGAKEVVFASIDT GTPAKFYYNCAPAHTTYPTKKVTPDEVSPVTLGDNLTSNRRTINKYFVPDVLETCQLSMG LTELAPGNLWNTMPCHTHERRMEVYFYFNMDDDACVFHMMGQPQETRHIVMHNEQAVISP SWSIHSGVGTKAYTFIWGMVGENQVFDDMDHVAVKDLR
NT seq	837 nt [NT seq] +upstream [0] nt +downstream [0] nt gtggacgtaagacagagcatccacagtgcgcacgcaaaaacgctggatacccaagggctg

Another Example of BRITE

Search BRITE hierarchies

```
KEGG Orthology (KO) [BR:hsa00001]
  01300 Environmental Information Processing
    01320 Signal Transduction
      04020 Calcium signaling pathway [PATH:hsa04020]
        1128 CHRM1; cholinergic receptor, muscarinic 1; K04129 cholinergic receptor, muscarinic 1
    01330 Signaling Molecules and Interaction
      04080 Neuroactive ligand-receptor interaction [PATH:hsa04080]
        1128 CHRM1; cholinergic receptor, muscarinic 1; K04129 cholinergic receptor, muscarinic 1
  01400 Cellular Processes
    01410 Cell Motility
      04810 Regulation of actin cytoskeleton [PATH:hsa04810]
        1128 CHRM1; cholinergic receptor, muscarinic 1; K04129 cholinergic receptor, muscarinic 1

Receptors and channels [BR:hsa04000]
  G Protein-Coupled Receptors
    Rhodopsin family: amine receptors
      Acetylcholine (muscarinic)
        1128 CHRM1; cholinergic receptor, muscarinic 1; K04129 cholinergic receptor, muscarinic 1
```

SSDB field

Link to SSDB for obtaining
- Orthologs:
  - BBH pairs
- Paralogs:
  - homologs in the same organism
- Conserved gene clusters

KEGG

Escherichia coli K-12 MG1655: b2843

Help

Entry	b2843 CDS E.coli
Gene name	kduI, yqeE
Definition	predicted 5-keto 4-deoxyuronate isomerase [EC:5.3.1.17]
Orthology	KO: K01815 4-deoxy-L-threo-5-hexosulose-uronate ketol-isomerase
Pathway	PATH: eco00040 Pentose and glucuronate interconversions
Class	[BRITE hierarchy]
SSDB	[Ortholog] [Paralog] [Gene cluster]
Motif	Pfam: KduI IolB [Motif]
Other DBs	Wisconsin: b2843 Colibri: kduI RegulonDB: B2843 NCBI-GI: 16130747 NCBI-GeneID: 947319 UniProt: Q46938
LinkDB	[All DBs]
Structure	PDB: 1XRU 1X8M [Thumbnails] [Jmol]
Position	complement(2981310..2982146) [Genome map]
AA seq	278 aa [AA seq] [DB search] MDVRQSIHSAHAKTLDTQGLRNEFLVEKVFVADEYTMVYSHIDRIIVGGIMPITKTVSVG GEVGKQLGVSYFLERRELGVINIGGAGTITVDGQCYEIGHRDALYVGKGAKEVVFASIDT GTPAKFYYNCAPAHTTYPTKKVTPDEVSPVTLGDNLTSNRRTINKYFVPDVLETCQLSMG LTELAPGNLWNTMPCHTHERRMEVYFYFNMDDDACVFHMMGQPQETRHIVMHNEQAVISP SWSIHSGVGTKAYTFIWGMVGENQVFDDMDHVAVKDLR
NT seq	837 nt [NT seq] +upstream [0] nt +downstream [0] nt gtggacgtaagacagagcatccacagtgcgcacgcaaaaacgctggatacccaagggctg

SSDB Gene Cluster Search Result

KEGG ID : eco:b2843 (278 a.a.)

Definition: predicted 5-keto 4-deoxyuronate isomerase

Include: ☑ best hits   **Gap size:** 3
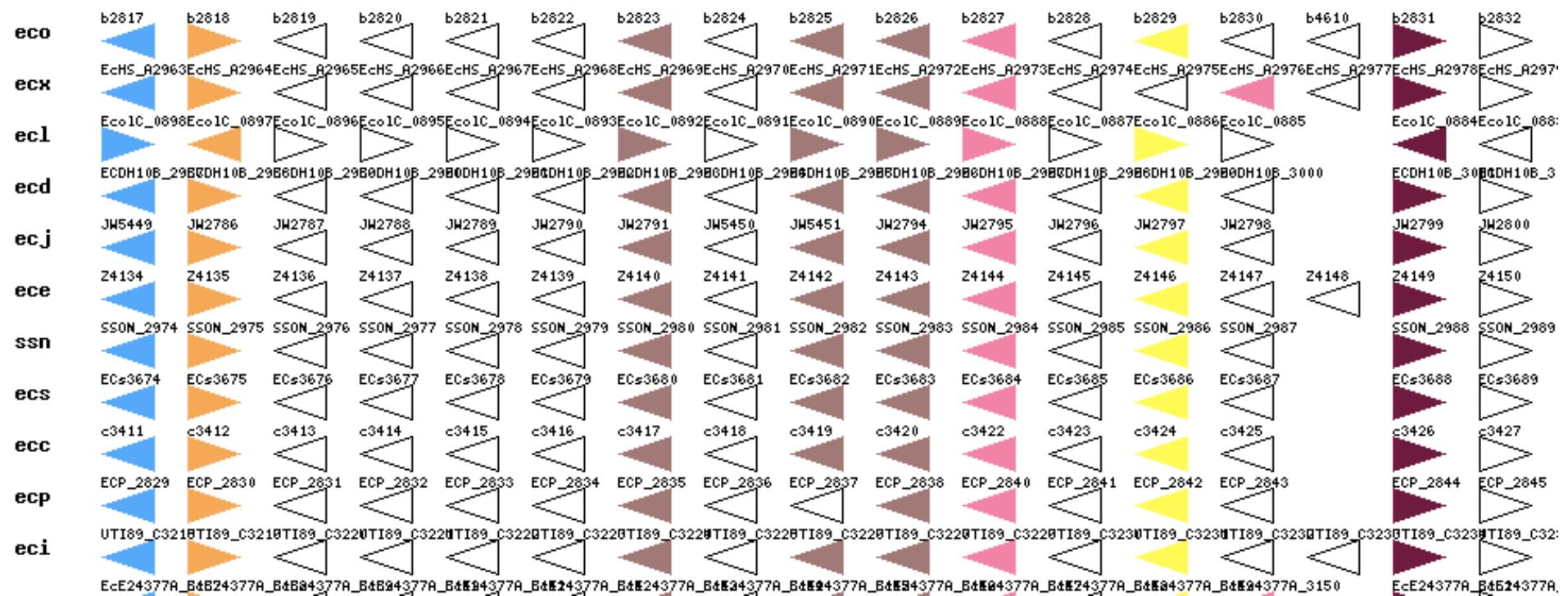
Sort by: ⦿ SW-score ◯ KEGG-species
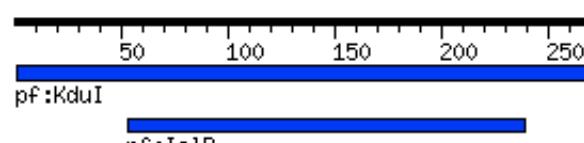
Search against: ⦿ All organisms ◯ Selected organism group Eukaryotes

Threshold: 100 (Go)

	b2817	b2818	b2819	b2820	b2821	b2822	b2823	b2824	b2825	b2826	b2827	b2828	b2829	b2830	b4610	b2831	b2832
eco																	

ecx: EcHS_A2963 EcHS_A2964 EcHS_A2965 EcHS_A2966 EcHS_A2967 EcHS_A2968 EcHS_A2969 EcHS_A2970 EcHS_A2971 EcHS_A2972 EcHS_A2973 EcHS_A2974 EcHS_A2975 EcHS_A2976 EcHS_A2977 EcHS_A2978 EcHS_A297?

ecl: Eco1C_0898 Eco1C_0897 Eco1C_0896 Eco1C_0895 Eco1C_0894 Eco1C_0893 Eco1C_0892 Eco1C_0891 Eco1C_0890 Eco1C_0889 Eco1C_0888 Eco1C_0887 Eco1C_0886 Eco1C_0885 Eco1C_0884 Eco1C_088?

ecd: ECDH10B_2962 ECDH10B_2968 DH10B_2980 DH10B_2980 DH10B_2986 DH10B_2986 DH10B_2986 DH10B_2986 DH10B_2986 DH10B_2987 DH10B_2986 DH10B_2980 DH10B_3000 ECDH10B_3006 DH10B_3?

ecj: JW5449 JW2786 JW2787 JW2788 JW2789 JW2790 JW2791 JW5450 JW5451 JW2794 JW2795 JW2796 JW2797 JW2798 JW2799 JW2800

ece: Z4134 Z4135 Z4136 Z4137 Z4138 Z4139 Z4140 Z4141 Z4142 Z4143 Z4144 Z4145 Z4146 Z4147 Z4148 Z4149 Z4150

ssn: SSON_2974 SSON_2975 SSON_2976 SSON_2977 SSON_2978 SSON_2979 SSON_2980 SSON_2981 SSON_2982 SSON_2983 SSON_2984 SSON_2985 SSON_2986 SSON_2987 SSON_2988 SSON_2989

ecs: ECs3674 ECs3675 ECs3676 ECs3677 ECs3678 ECs3679 ECs3680 ECs3681 ECs3682 ECs3683 ECs3684 ECs3685 ECs3686 ECs3687 ECs3688 ECs3689

ecc: c3411 c3412 c3413 c3414 c3415 c3416 c3417 c3418 c3419 c3420 c3422 c3423 c3424 c3425 c3426 c3427

ecp: ECP_2829 ECP_2830 ECP_2831 ECP_2832 ECP_2833 ECP_2834 ECP_2835 ECP_2836 ECP_2837 ECP_2838 ECP_2840 ECP_2841 ECP_2842 ECP_2843 ECP_2844 ECP_2845

eci: UTI89_C3218 UTI89_C3218 UTI89_C3220 UTI89_C3220 UTI89_C3222 UTI89_C3222 UTI89_C3222 UTI89_C3222 UTI89_C3222 UTI89_C3223 UTI89_C3230 UTI89_C3231 UTI89_C3232 UTI89_C3233 UTI89_C3234 UTI89_C3?

EcE24377A_B4624377A_B4624377A_B4624377A_B4624377A_B4624377A_B4624377A_B4624377A_B4624377A_B4624377A_B4624377A_B4624377A_B4624377A_3150 EcE24377A_B4624377A_

Motif field

Domains and motifs found in the protein sequence

KEGG

Help

Entry	b2843 CDS E.coli
Gene name	kduI, yqeE
Definition	predicted 5-keto 4-deoxyuronate isomerase [EC:5.3.1.17]
Orthology	KO: K01815 4-deoxy-L-threo-5-hexosulose-uronate ketol-isomerase
Pathway	PATH: eco00040 Pentose and glucuronate interconversions
Class	BRITE hierarchy
SSDB	Ortholog Paralog Gene cluster
Motif	Pfam: KduI IolB Motif
Other DBs	Wisconsin: b2843

SSDB Motif Search Result

Organism : E.coli
Gene :   b2843
Definition : predicted 5-keto 4-deoxyuronate isomerase

Motif id	From	To	Definition	E value	Score
pf:KduI	1	278	5-keto 4-deoxyuronate isomerase	0	-
pf:IolB	53	239	Myo-inositol catabolism protein IolB	0.12	-

Search GENES with the same motifs

eco:b2843

```
             50    100   150   200   250    (278)
pf:KduI
             pf:IolB
```

View sequence

KTVSVG
FASIDT
CQLSMG
QAVISP

SWSIHSGVGTKAYTFIWGMVGENQVFDDMDHVAVKDLR

NT seq	837 nt NT seq +upstream 0 nt +downstream 0 nt

gtggacgtaagacagagcatccacagtgcgcacgcaaaaacgctggatacccaagggctg

KEGG

Escherichia coli K-12 MG1655: b2843

Help

Entry	b2843 CDS E.coli
Gene name	kduI, yqeE
Definition	predicted 5-keto 4-deoxyuronate isomerase [EC:5.3.1.17]
Orthology	KO: K01815 4-deoxy-L-threo-5-hexosulose-uronate ketol-isomerase
Pathway	PATH: eco00040 Pentose and glucuronate interconversions
Class	BRITE hierarchy
SSDB	Ortholog Paralog Gene cluster
Motif	Pfam: KduI IolB Motif
Other DBs	Wisconsin: b2843 Colibri: kduI RegulonDB: B2843 NCBI-GI: 16130747 NCBI-GeneID: 947319 UniProt: Q46938
LinkDB	All DBs
Structure	PDB: 1XRU 1X8M Thumbnails Jmol
Position	complement(2981310..2982146) Genome map
AA seq	278 aa AA seq DB search MDVRQSIHSAHAKTLDTQGLRNEFLVEKVFVADEYTMVYSHIDRIIVGGIMPITKTVSVG GEVGKQLGVSYFLERRELGVINIGGAGTITVDGQCYEIGHRDALYVGKGAKEVVFASIDT GTPAKFYYNCAPAHTTYPTKKVTPDEVSPVTLGDNLTSNRRTINKYFVPDVLETCQLSMG LTELAPGNLWNTMPCHTHERRMEVYFYFNMDDDACVFHMMGQPQETRHIVMHNEQAVISP SWSIHSGVGTKAYTFIWGMVGENQVFDDMDHVAVKDLR
NT seq	837 nt NT seq +upstream 0 nt +downstream 0 nt gtggacgtaagacagagcatccacagtgcgcacgcaaaaacgctggatacccaagggctg

Other DBs field ▶

LinkDB field ▶
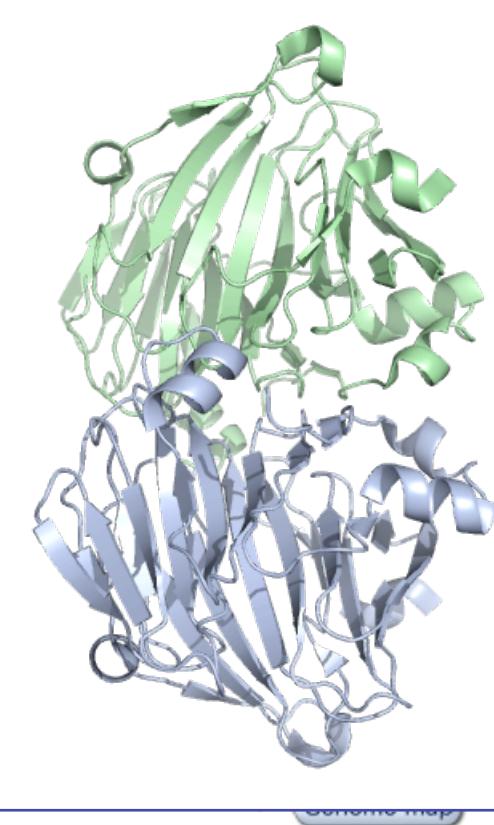
Links to other databases

KEGG **Escherichia coli K-12 MG1655: b2843**

Help

Entry	b2843　　　　　CDS　　　E.coli
Gene name	kduI, yqeE
Definition	predicted 5-keto 4-deoxyuronate isomerase [EC:5.3.1.17]
Orthology	KO: K01815　4-deoxy-L-threo-5-hexosulose-uronate ketol-isomerase
Pathway	PATH: eco00040　Pentose and glucuronate interconversions
Class	BRI
SSDB	Ort
Motif	Pfa M
Other DBs	Wis Col Reg NCB NCB Uni
LinkDB	All
Structure	PDB
Position	com
AA seq	278 aa　AA seq　DB search
	MDVRQSIHSAHAKTLDTQGLRNEFLVEKVFVADEYTMVYSHIDRIIVGGIMPITKTVSVG GEVGKQLGVSYFLERRELGVINIGGAGTITVDGQCYEIGHRDALYVGKGAKEVVFASIDT GTPAKFYYNCAPAHTTYPTKKVTPDEVSPVTLGDNLTSNRRTINKYFVPDVLETCQLSMG LTELAPGNLWNTMPCHTHERRMEVYFYFNMDDDACVFHMMGQPQETRHIVMHNEQAVISP SWSIHSGVGTKAYTFIWGMVGENQVFDDMDHVAVKDLR
NT seq	837 nt　NT seq　+upstream 0 nt　+downstream 0 nt
	gtggacgtaagacagagcatccacagtgcgcacgcaaaaacgctggatacccaagggctg

Structure field ➤

Thumbnails

Jmol

Position field ➤

Locus on the genome sequence contained in the KEGG GENOME database
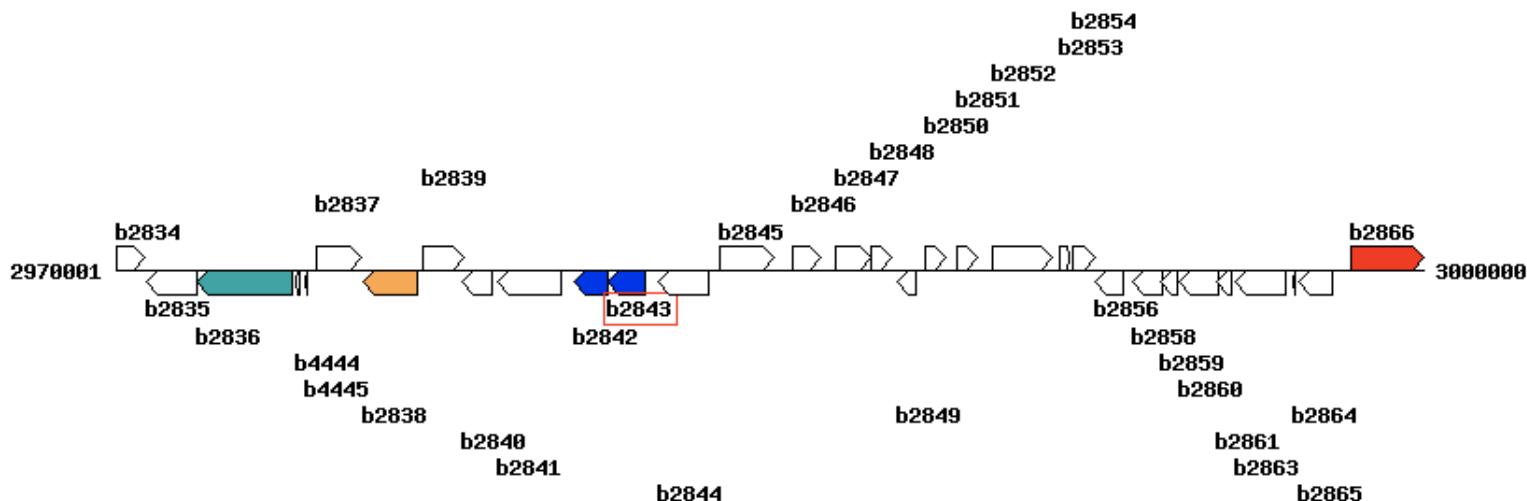
Position	complement(2981310..2982146) (Genome map)

AA seq	278 aa (AA seq) (DB search)
	MDVRQSIHSAHAKTLDTQGLRNEFLVEKVFVADEYTMVYSHIDRIIVGGIMPITKTVSVG GEVGKQLGVSYFLERRELGVINIGGAGTITVDGQCYEIGHRDALYVGKGAKEVVFASIDT GTPAKFYYNCAPAHTTYPTKKVTPDEVSPVTLGDNLTSNRRTINKYFVPDVLETCQLSMG LTELAPGNLWNTMPCHTHERRMEVYFYFNMDDDACVFHMMGQPQETRHIVMHNEQAVISP SWSIHSGVGTKAYTFIWGMVGENQVFDDMDHVAVKDLR

NT seq

KEGG Genome map - Escherichia coli K-12 MG1655

from **2970001**bp to **3120000**bp

(ORF Color) (PATHWAY) (LIST)

AA seq field ➤

NT seq field ➤

Amino acid and
nucleotide sequence
in FASTA format and
link to BLAST search
in GenomeNet

	Thumbnails [structure image] Jmol
Position	complement(2981310..2982146) Genome map
AA seq	278 aa AA seq DB search MDVRQSIHSAHAKTLDTQGLRNEFLVEKVFVADEYTMVYSHIDRIIVGGIMPITKTVSVG GEVGKQLGVSYFLERRELGVINIGGAGTITVDGQCYEIGHRDALYVGKGAKEVVFASIDT GTPAKFYYNCAPAHTTYPTKKVTPDEVSPVTLGDNLTSNRRTINKYFVPDVLETCQLSMG LTELAPGNLWNTMPCHTHERRMEVYFYFNMDDDACVFHMMGQPQETRHIVMHNEQAVISP SWSIHSGVGTKAYTFIWGMVGENQVFDDMDHVAVKDLR
NT seq	837 nt NT seq +upstream 0 nt +downstream 0 nt gtggacgtaagacagagcatccacagtgcgcacgcaaaaacgctggatacccaagggctg cgcaatgaatttttggttgaaaaggtatttgtcgccgatgagtacaccatggtttacagc cacattgaccgaattattgttggcggcattatgccgataactaaaacggtttccgttggc ggggaagttggtaaacaactcggcgtaagctatttccttgaacgtcgcgagttaggtgtt atcaatattggcggtgccggtacgattactgtcgatggccaatgctatgaaatcggtcac cgcgacgccctgtatgttggtaaaggtgcaaaagaagttgtctttgccagtattgatacc ggcactccggcgaagtttttattacaattgcgcacccgcgcatacgacgtatcccaccaaa aaagtcacaccggacgaagtatctccagtcacgttaggcgataacctcaccagtaaccgt cgcacgattaacaaatattttgtcccggatgtactggaaacctgccaattgagtatgggg ctgacggagctggctccgggtaacttgtggaacaccatgccgtgtcacacccacgagcgc cggatggaagtttatttctatttcaatatggatgatgacgcctgcgtttttccacatgatg gggcagccgcaagaaacgcgtcatattgtgatgcataacgagcaggcggtgatctcccccg agctggtcgatccattccggtgtcggaaccaaagcttataccttatctggggcatggtc ggtgaaaaccaggtctttgatgatatggaccatgtggccgttaaagatttgcgctag

KEGG GENES main components

KEGG Orthology (KO)	Ortholog groups linked to PATHWAY and BRITE
GENES	Gene catalogs of complete genomes with manual functional annotation
DGENES	Gene catalogs of draft genomes with automatic functional annotation
EGENES	Consensus contigs of EST data with automatic functional annotation
SSDB	Sequence similarity with best-hit information for identifying ortholog/paralogs

GENES, DGENES, EGENES: Genomic information

KEGG Orthology, SSDB: Relationship

KEGG Orthology Database

Ortholog groups bridging the genomic space
and systems space
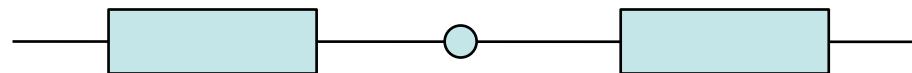
Linking GENES to PATHWAY

- KEGG PATHWAY database
  - Important component of the KEGG databases
  - Golden standard of metabolic pathways for bioinformatics
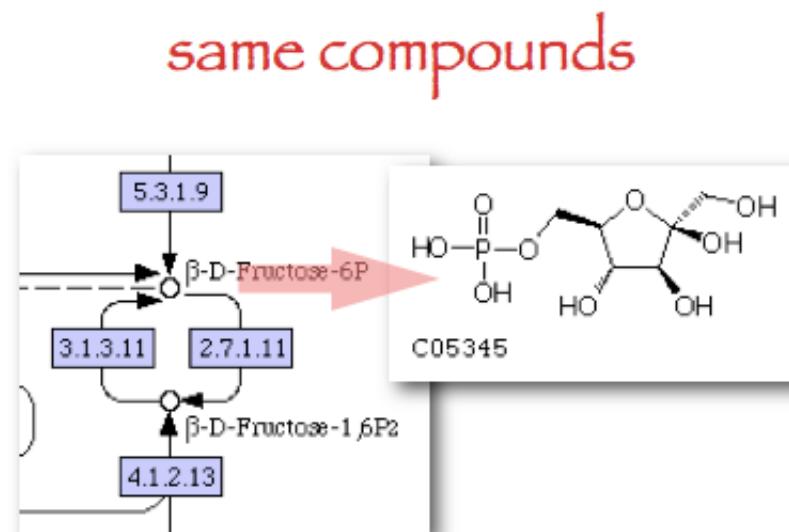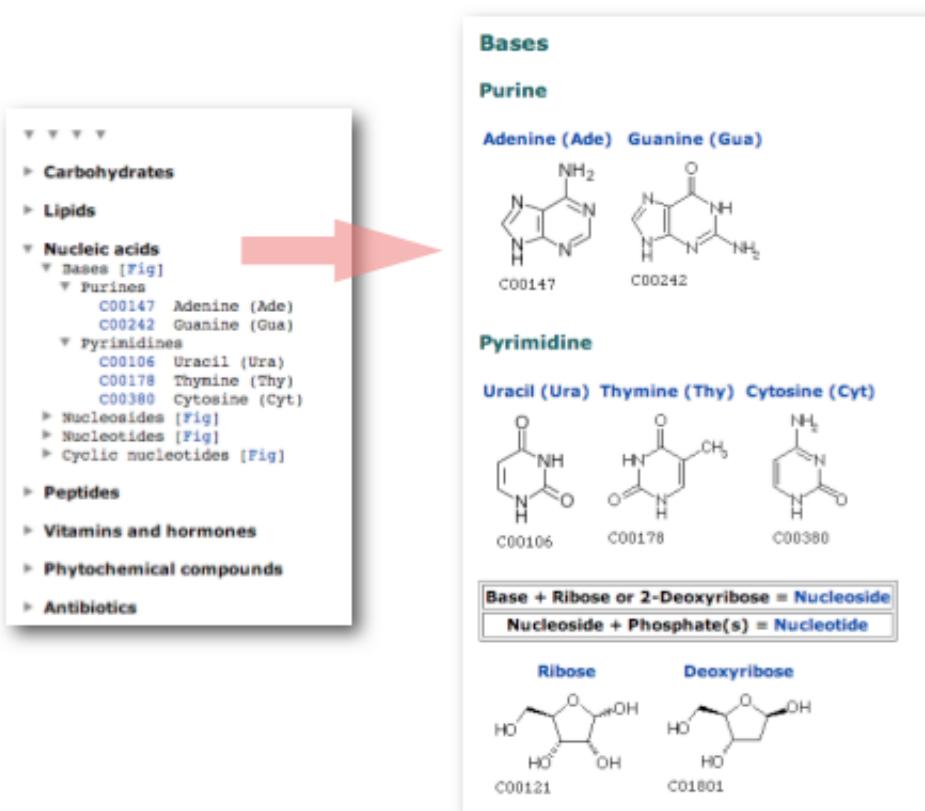
Node: genes and compounds

Edge: reactions and interactions
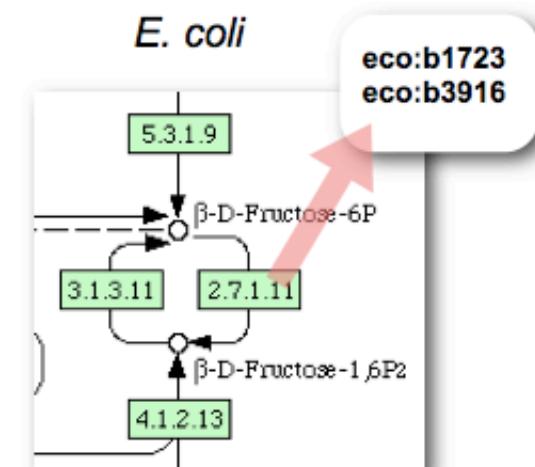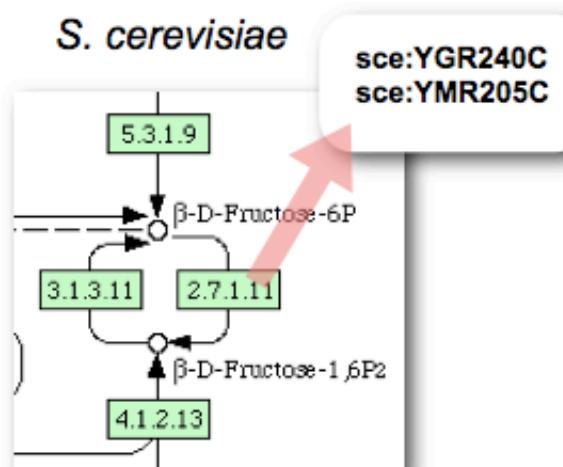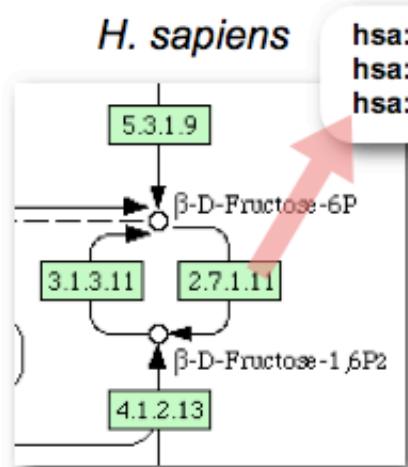
Linking GENES to PATHWAY

Compound nodes in the pathway networks

Linking GENES to PATHWAY

Gene/protein nodes in the pahtway networks
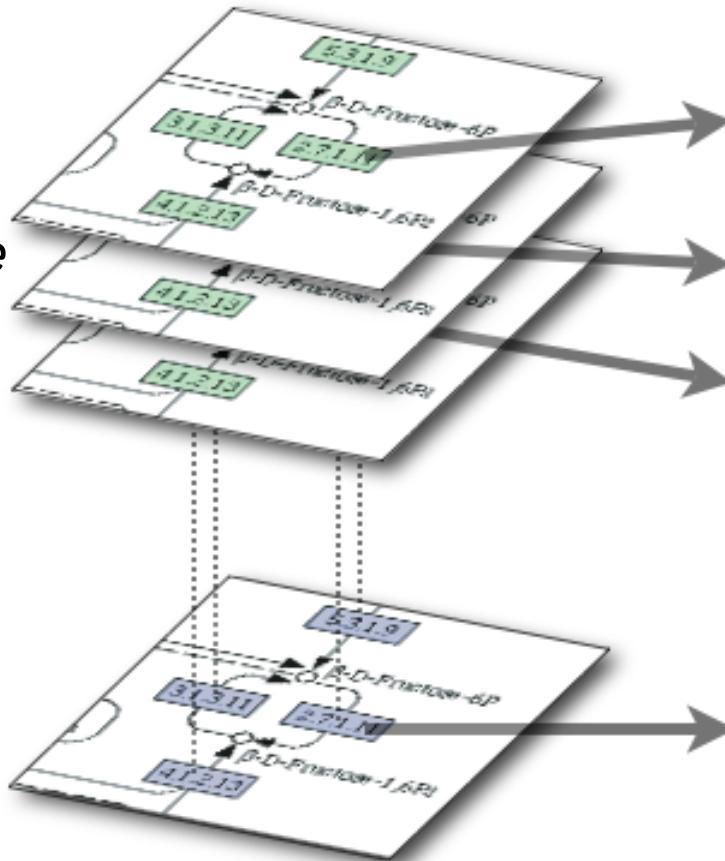
Different sequences

Linking GENES to PATHWAY
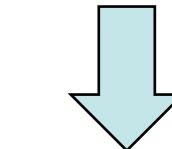
Thus, grouping by the orthologous relationships

→ KEGG Orthology

H. sapiens

S. cerevisiae

E. coli

hsa:5211
hsa:5213
hsa:5214

sce:YGR240C
sce:YMR205C

eco:b1723
eco:b3916

K00850

KEGG Orthology

Nodes in the reference pathways

Reference nodes for genes of all organisms

Pathways in
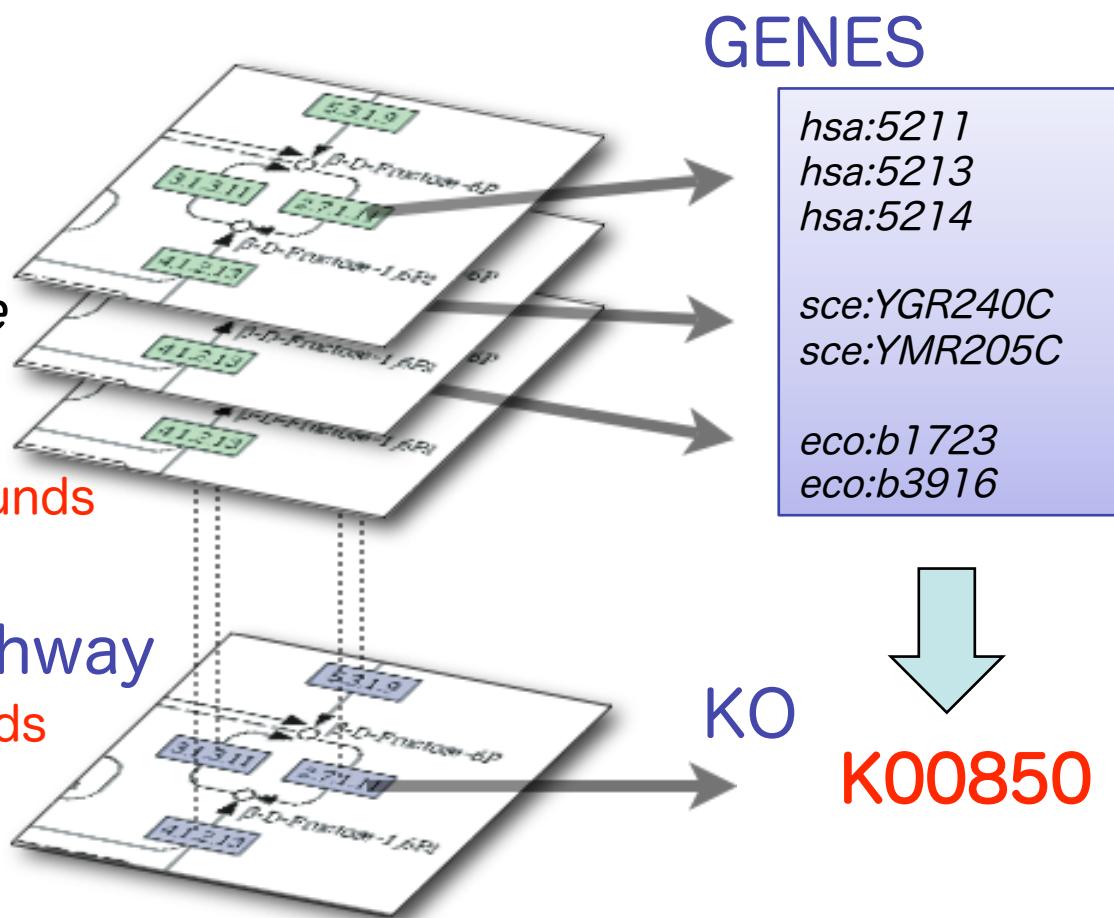
H. sapiens

S. cerevisiae

E. coli

Genes and compounds

GENES

hsa:5211
hsa:5213
hsa:5214

sce:YGR240C
sce:YMR205C

eco:b1723
eco:b3916
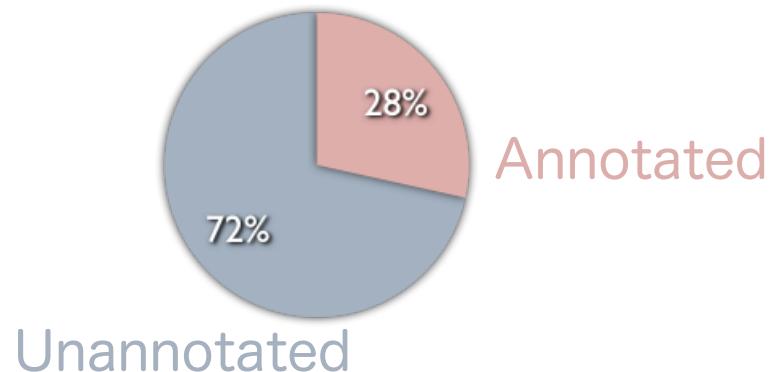
Reference pathway

KOs and compounds

KO

K00850

KEGG Orthology

GENES

28% Annotated

72% Unannotated
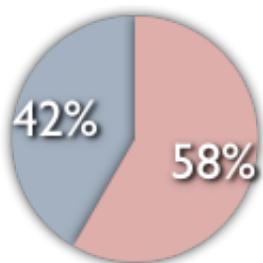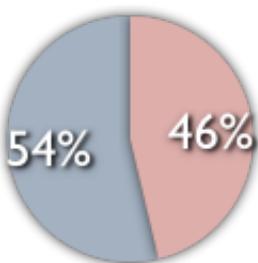
10,890 ortholog groups contain 28% of KEGG GENES entries
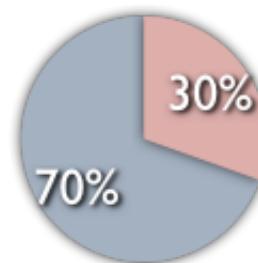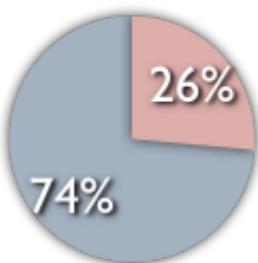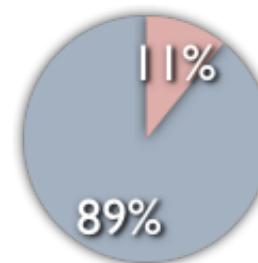
E. coli   *B. subtilis*   *S. cerevisiae*   *H. sapiens*   *A. thaliana*
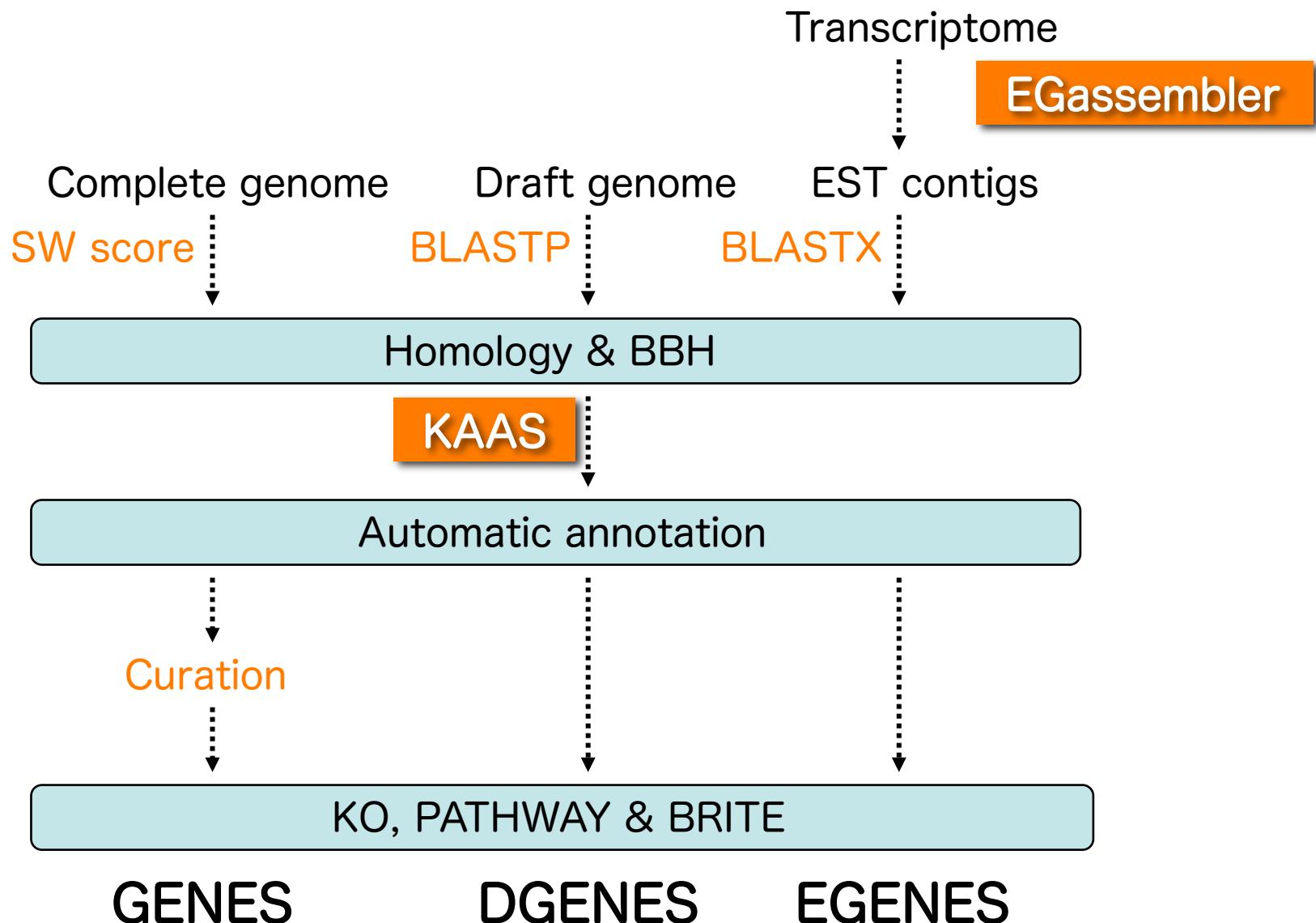
42%   58%

54%   46%

70%   30%

74%   26%

89%   11%

Annotation via KEGG Orthology

Transcriptome

EGassembler

Complete genome   Draft genome   EST contigs

SW score   BLASTP   BLASTX

Homology & BBH

KAAS

Automatic annotation

Curation

KO, PATHWAY & BRITE

GENES   DGENES   EGENES

KEGG Automatic Annotation Server

Ortholog assignment and pathway mapping

KAAS

- KEGG Automatic Annotation Server

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

- Automatic annotation system for KO

    - Using GENES as a template set

    - More than 90% accuracy

- Reconstruct PATHWAY by using your own data set

Functional Annotation in KAAS

1. **Query gene**

   ⬇ BLAST

2. **Homologs**

   ⬇ Cut off by
   bi-directional best hit rate

3. **Ortholog candidates**
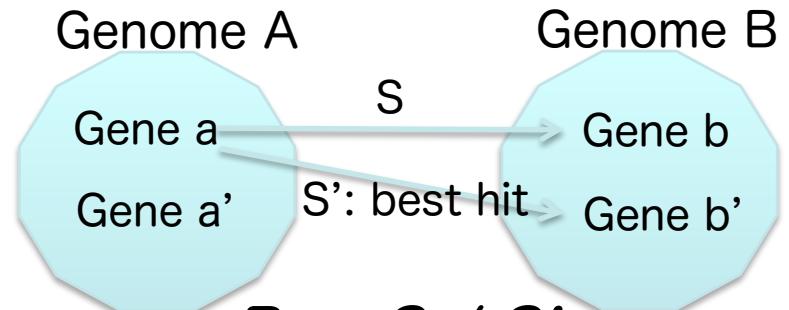
   ⬇ Grouping by KEGG Orthology

4. **KEGG Orthology groups**

   ⬇ Scoring by probability
   and heuristics

5. **Ranking of KEGG Orthology**

Bi-directional best hit rate

$$BHR_{ab} = R_f \times R_r$$

Genome A   Genome B

Gene a ——— S ——→ Gene b

Gene a' S': best hit   Gene b'

$$R_f = S / S'$$

$$S_{KO} = S_h - \log_2(mn) - \log_2\left(\sum_{k=N}^{x} {}_x C_k p^k (1-p)^{x-k}\right)$$
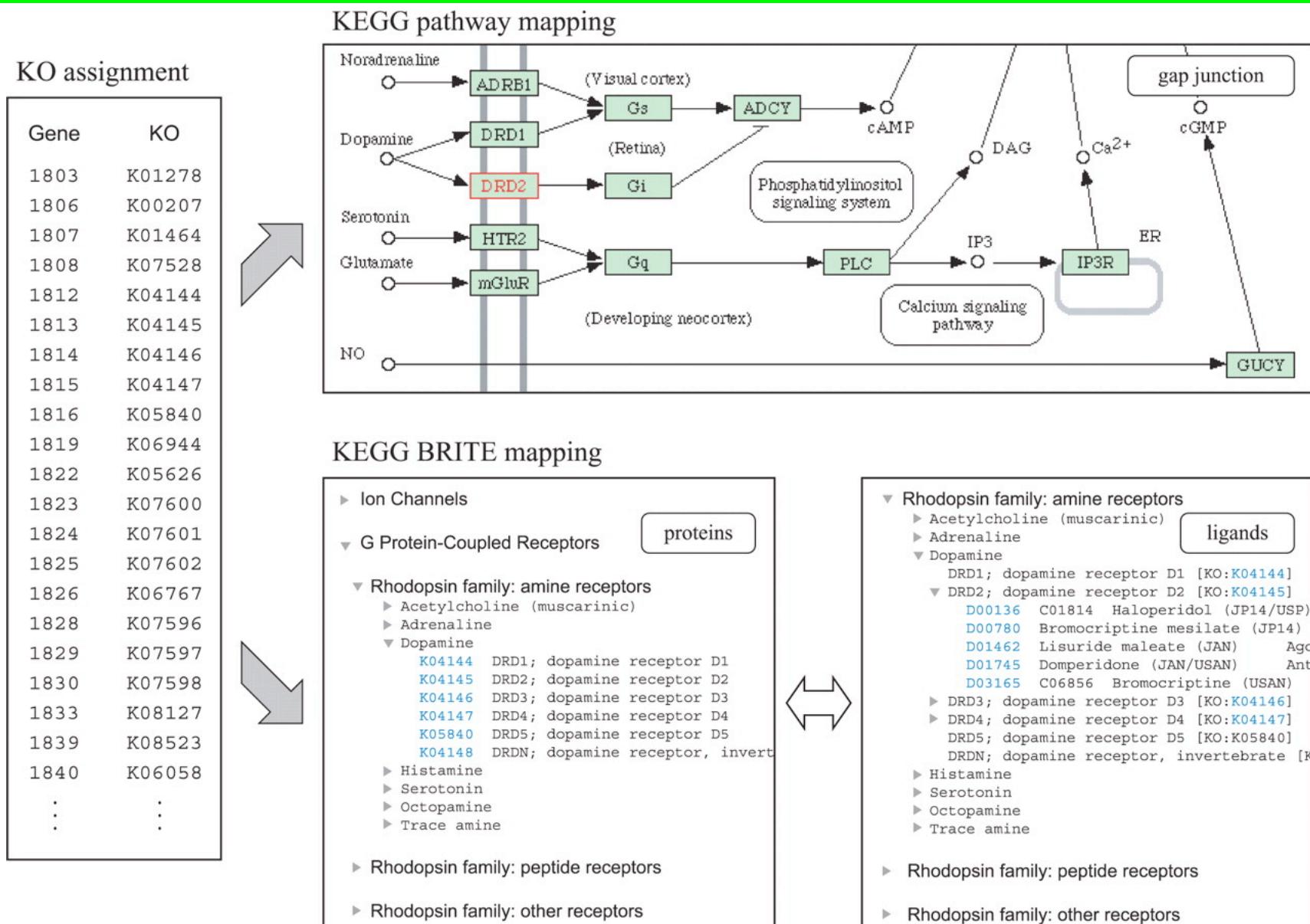
Scoring and Ranking of KO

$$S_{KO} = \underset{1.}{\underline{S_h}} - \underset{2.}{\underline{log_2(mn)}} - \underset{3.}{\underline{log_2\left(\sum_{k=N}^{x} {}_{x}C_k p^k (1-p)^{x-k}\right)}}$$

Your query genome sequences	KEGG GENES entries with KO									
	＞ seq 1	＞ seq 2	＞ seq 3	＞ seq 4	＞ seq 5	＞ seq 6	＞ seq 7	…	…	…
＞ Your query sequences 1		Ortholog1		Homolog		Ortholog2			Ortholog3	
＞ Your query sequences 2										
＞ Your query sequences 3										
＞ Your query sequences 4		KO3				KO1			KO2	
＞ Your query sequences 5										
…										
…		Score 3				Score 1			Score 2	
…										
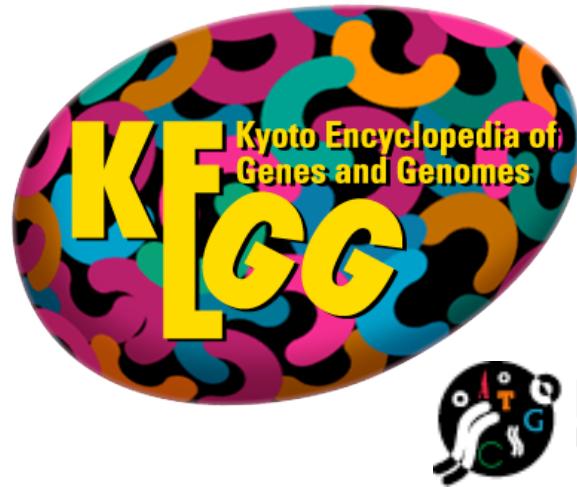
1. The highest score
2. Normalization by the sequence lengths
3. Weighting factor of the number of ortholog candidates

Pathway mapping in KAAS

Moriya, Y. et al. Nucl. Acids Res. 2007 35:W182-W185

EGassembler

Easy assembling nucleotide sequences
for pathway reconstruction
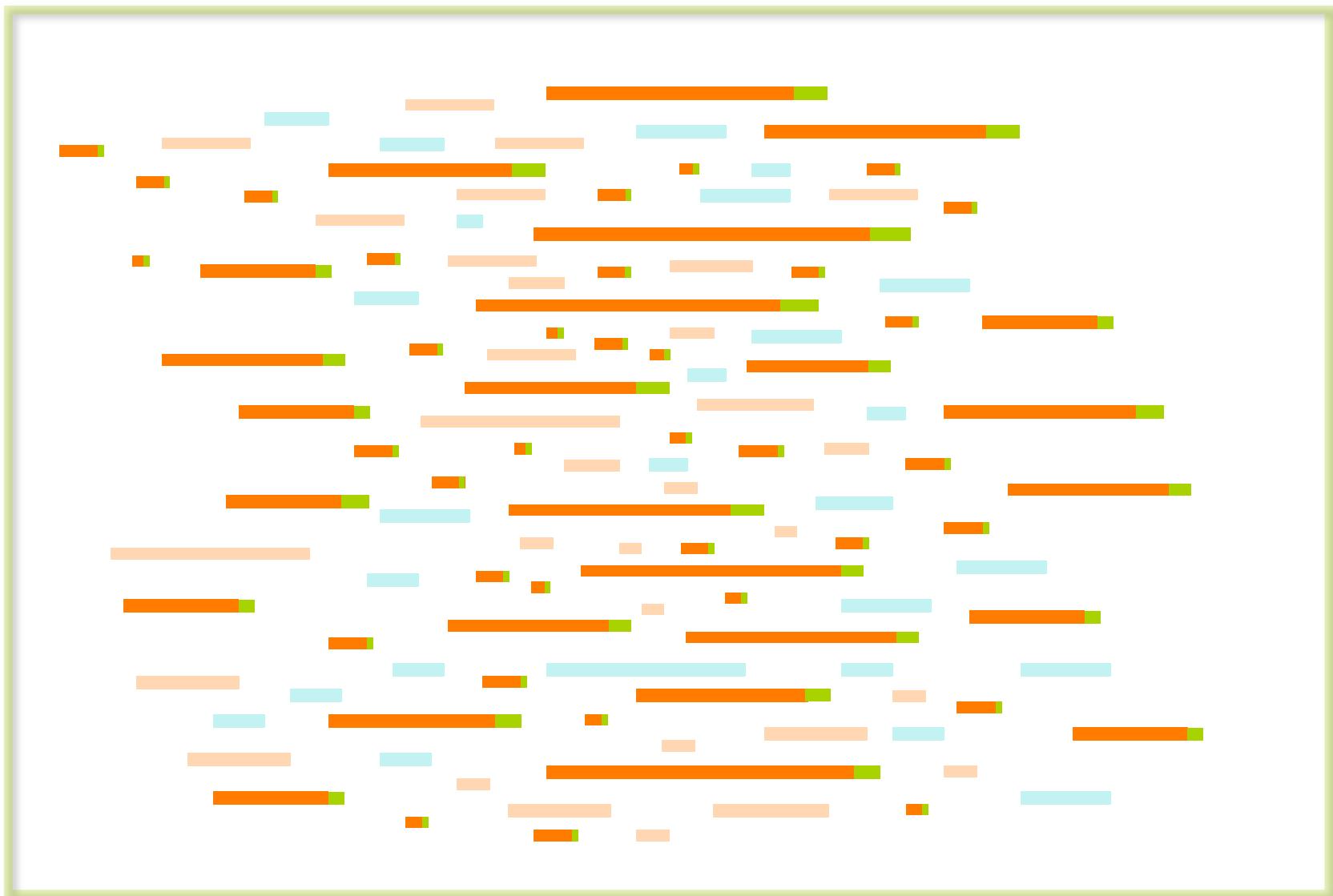
EGassembler
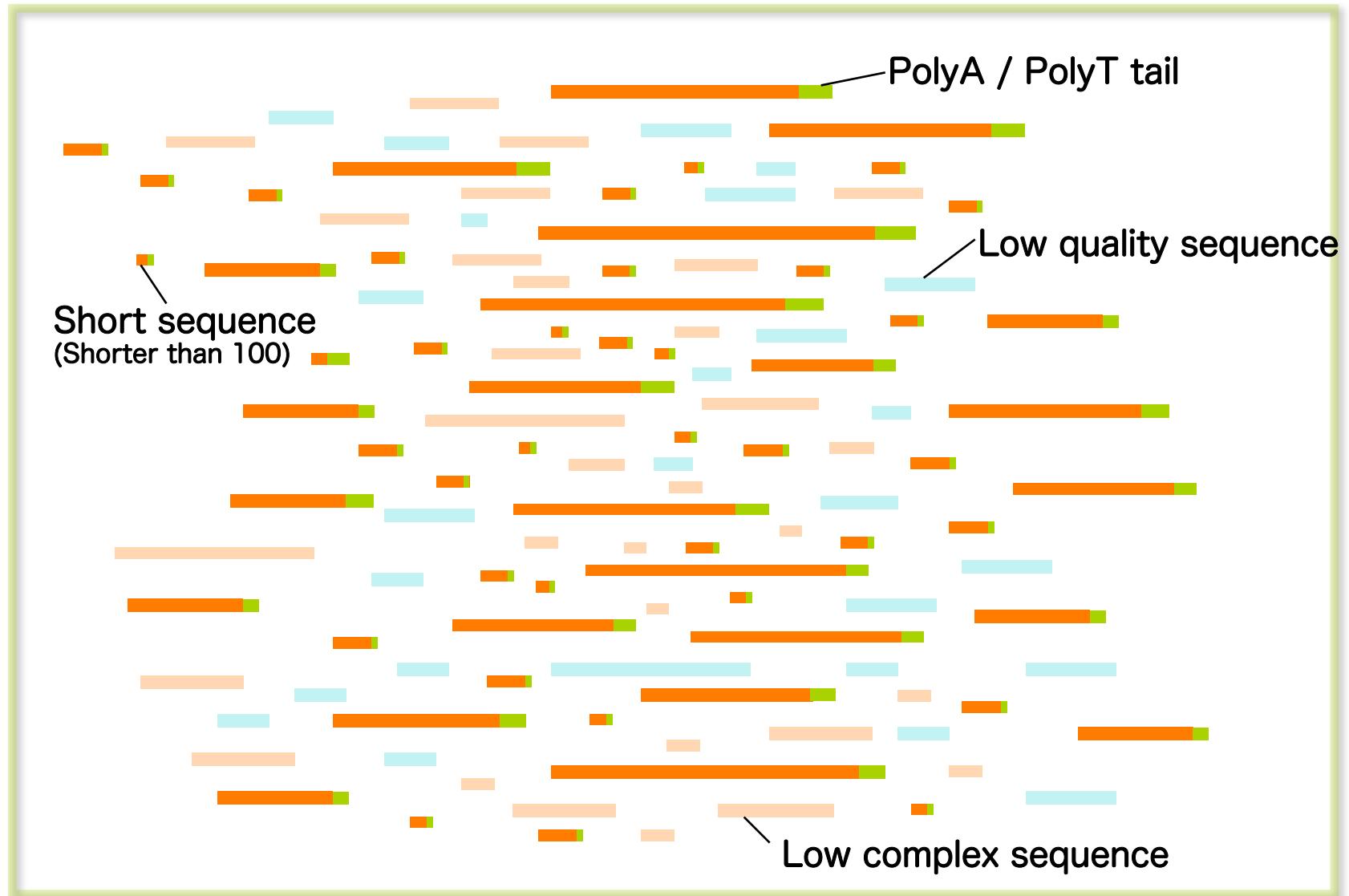
1. EST (GSS, cDNA, gDNA) sequences

2. Sequence cleaning
   PolyA/PolyT, Low-complexity, Low-quality filtering
   Short sequence removal (seqclean)

3. Repeat masking
   RepBase, TIGR, TREP, User's database (RepeatMasker)

4. Vector masking
   UniVec, emvec, User's database (CrossMatch)

5. Organelle masking
   NCBI organelle database

6. Sequence assembling (CAP3)

Automatic all-in-one pipeline

Masoudi-Nejad, A. et al. Nucl. Acids Res. 2006 34:W459-W462

2. Sequence Cleaning

PolyA / PolyT tail

Low quality sequence

Short sequence
(Shorter than 100)

Low complex sequence
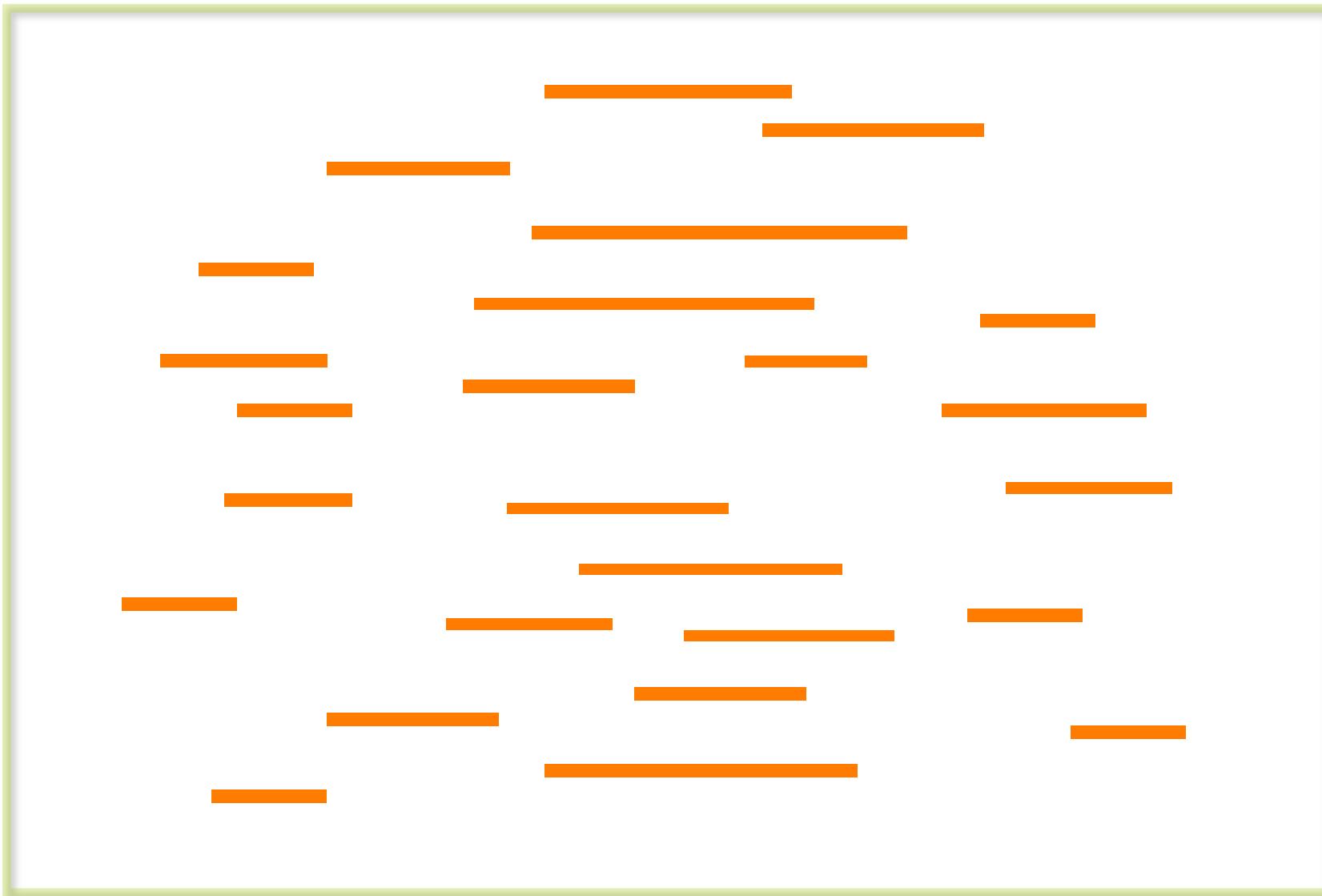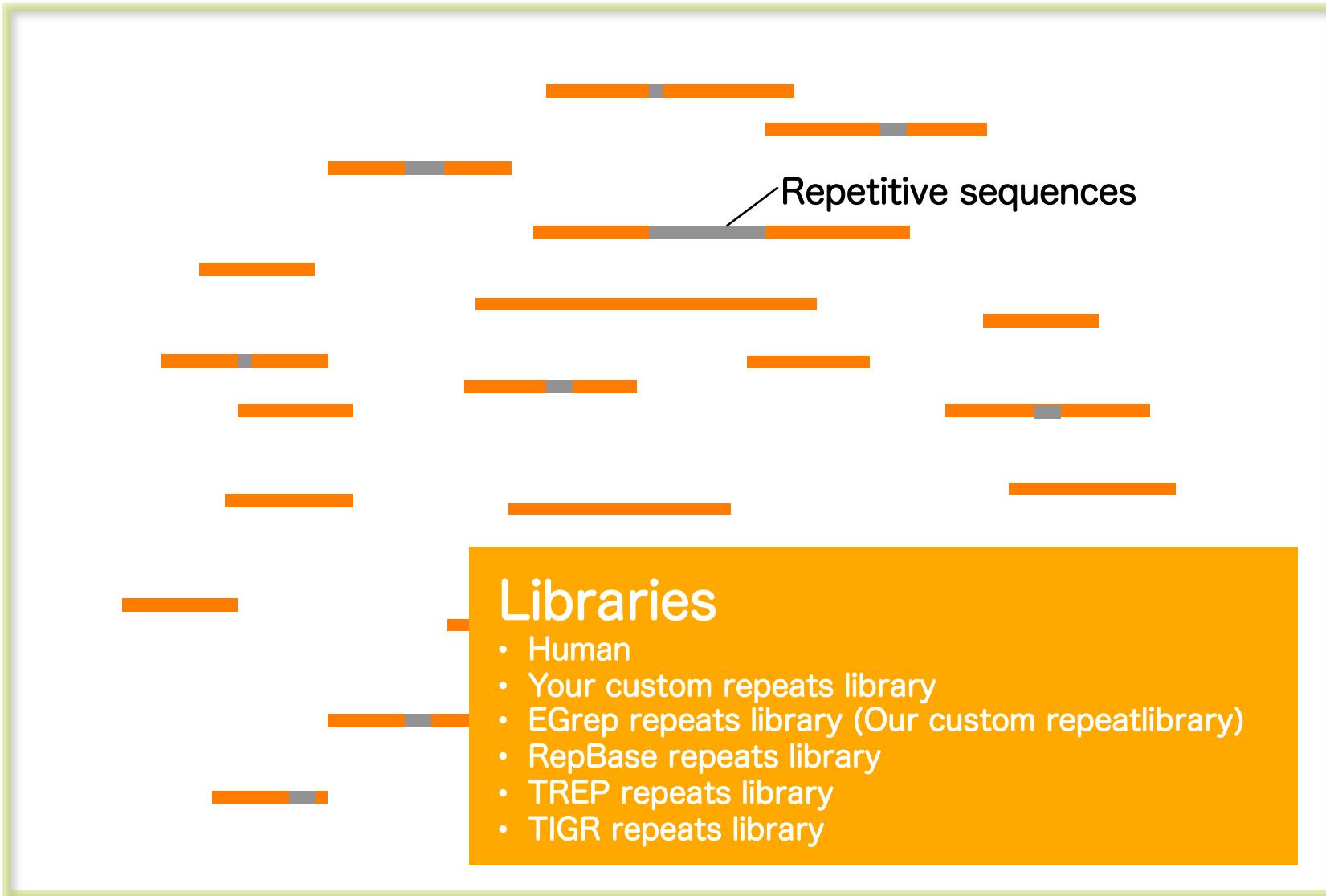
Removed

3. Repeat Masking

Repetitive sequences

Libraries
- Human
- Your custom repeats library
- EGrep repeats library (Our custom repeatlibrary)
- RepBase repeats library
- TREP repeats library
- TIGR repeats library

Masked

Masked as "XXXXXX…"

4. Vector Masking

Vector sequences

Libraries
- Your custom vector library
- EGvec vector library (Our custom vectors library)
- NCBI's vector library CoreRedundant
- EMBL vector library
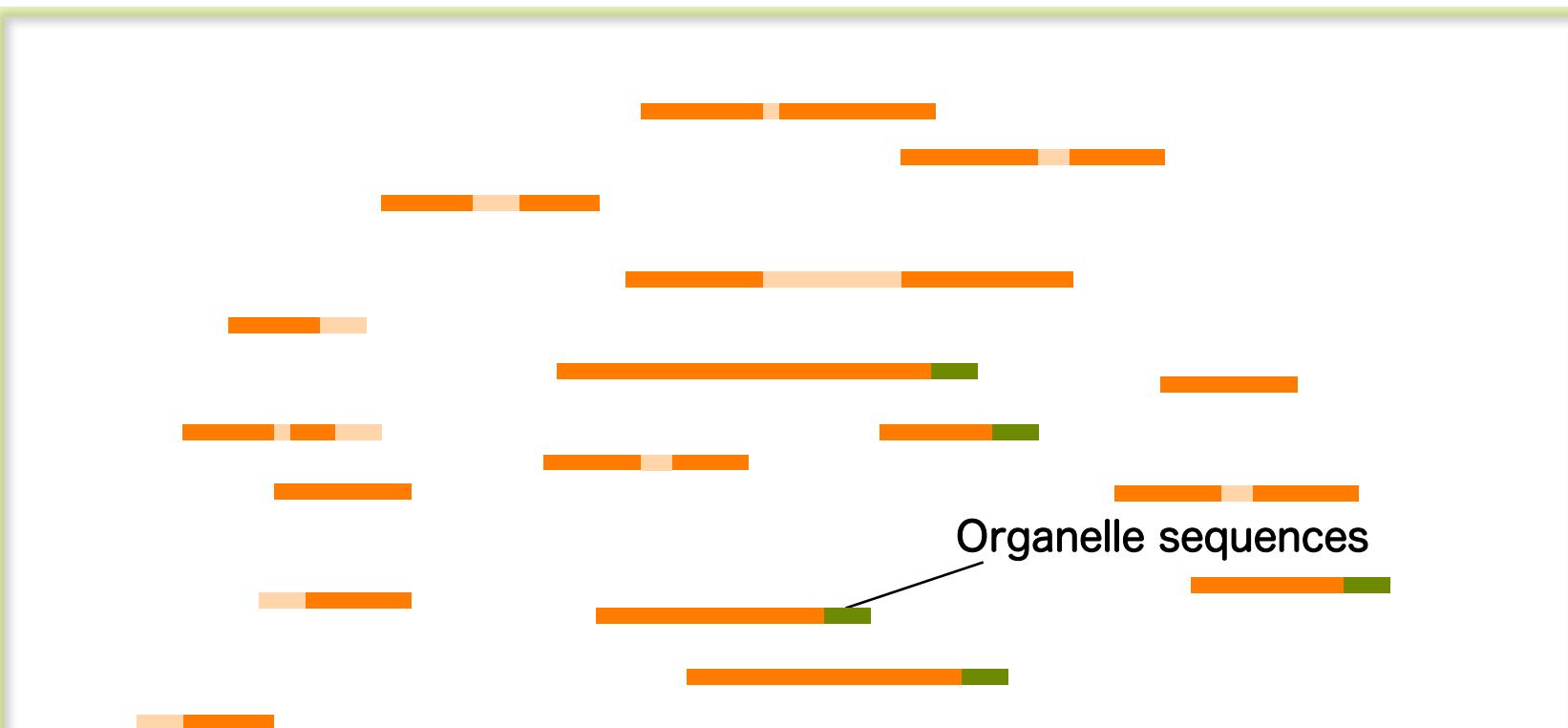
Masked

Masked as "XXXXXX…"

5. Organelle Masking

Organelle sequences
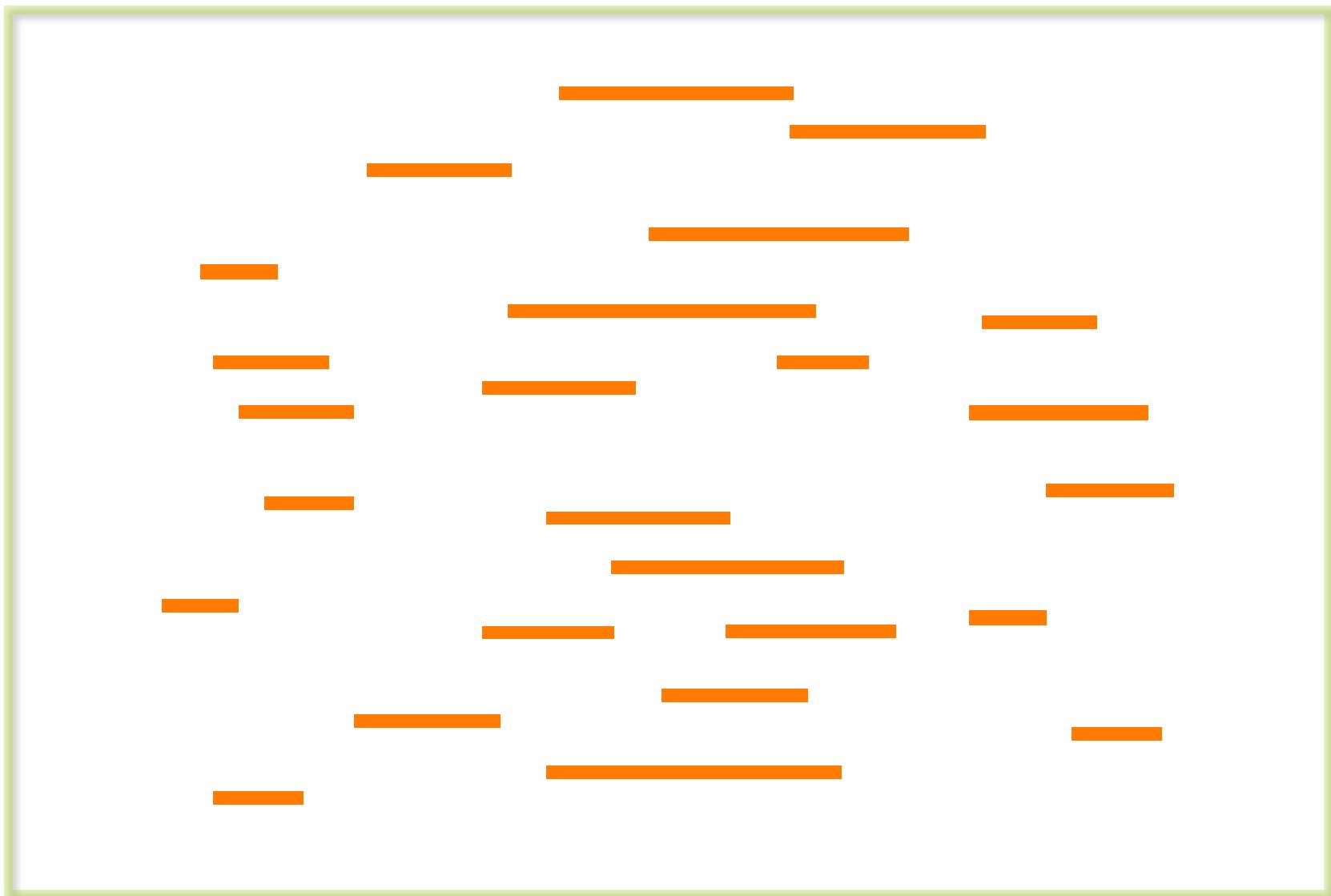
Libraries

- Your custom organelle library
- Plastids library (44 species including Arabidopsis, Chlamydomonas, Lotus, Zea mays, etc.)
- Mitochondria library (Fungi, Metazoa, Plants, Plasmid, etc.)

Masked

Masked as "XXXXXX…"

6. Assembly by CAP3
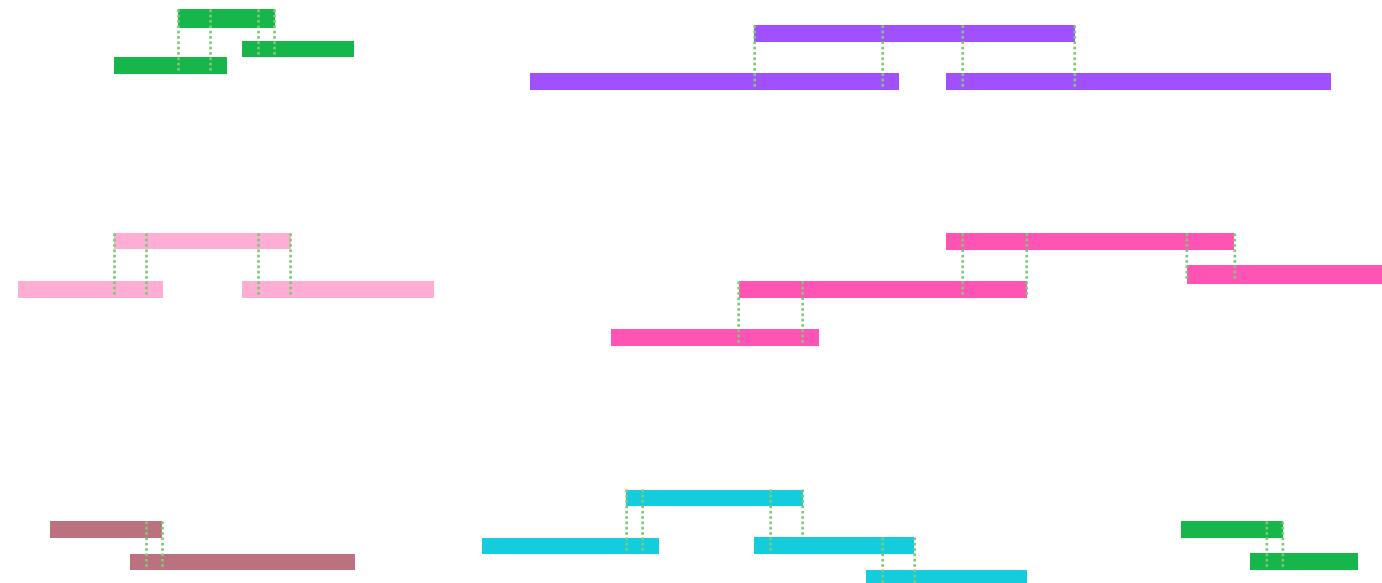
Potential overlap determination by BLAST-like technique
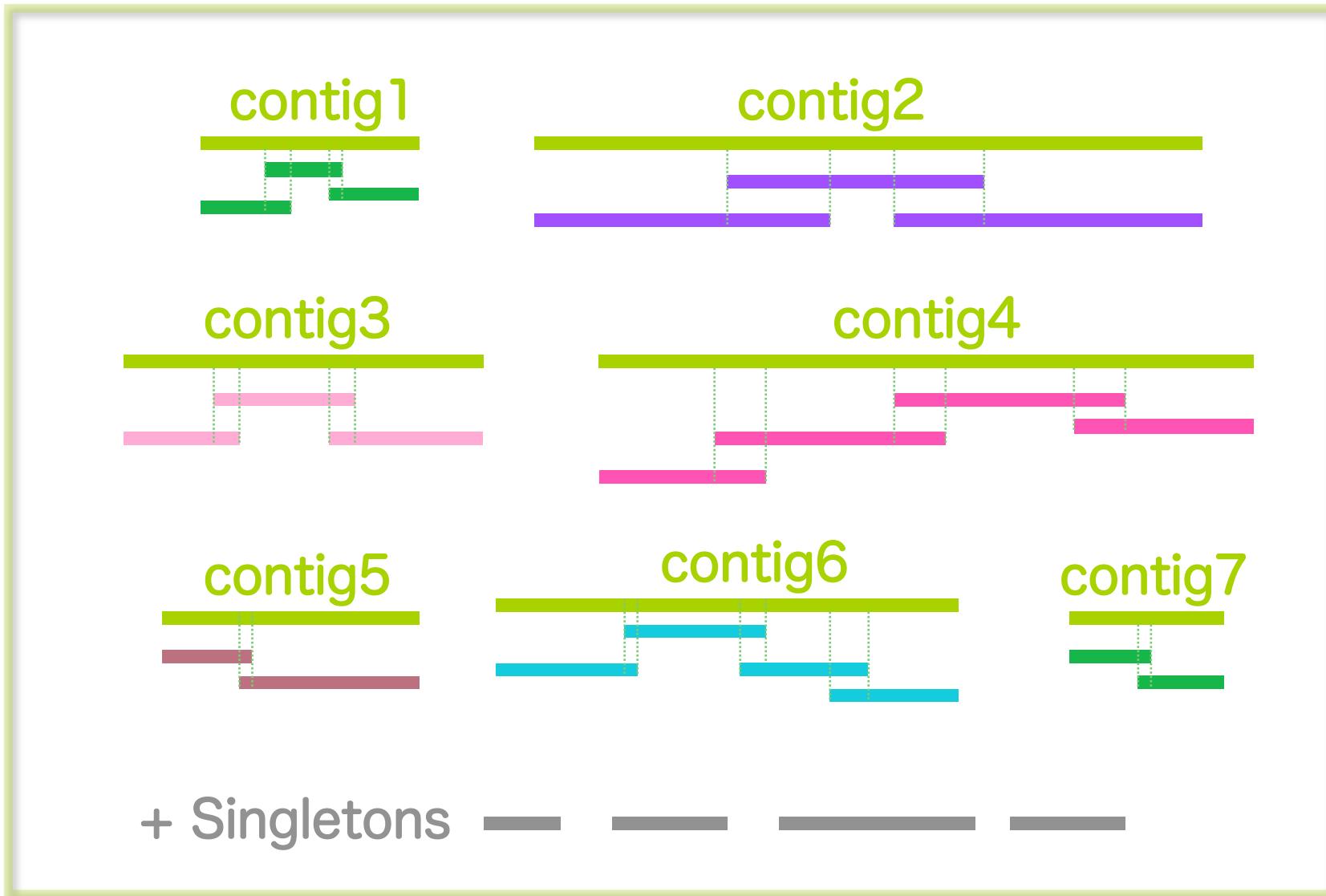
Local alignment by Smith-Waterman algorithm

+ Singletons

Assembly

contig1

contig2

contig3

contig4

contig5

contig6

contig7

+ Singletons

EST consensus contigs and Singletons in FASTA format

>contig1

>contig2

>contig3

>contig4

>contig5

>contig6

>contig7

+ Singletons

>1

>2

>3

>4

EGENES Example

Solanum lycopersicum (Tomato) : Taxonomy TAX: 4081

Number of accepted ESTs at NCBI : 260126 (2007/5/22)

↓ clean

Number of cleaned ESTs : 246118 (~96% left)

↓ assembly

Assembled sequences: Contigs 19479 + Singletons 18010
(Total : 37507)

An example of an alignment result

>Contig1
GGCACGAGGAAAAAAAAAAAAAAAGAATGGATTCCAAAGCATTTCTATTTTTTGGTCTTT
TTTTGGCTATTTTCCTAATGATAAGCTCTGAGATTTTAGCTACTGAGTTGGCTGAGAACT
CTAAGAAATCTGAAAACAAGAATGAAGTACATGAAGCCCAATACGGTGGATATCCTGGTG
GTGGTGGTGGATATGGACGCGGTGGTGGTGGTGGATATGGACGCGGTGGTGGTGGAGGAT
ATGGACGTGGTGGTGGATACGGACATGGTGGTGGTGGTGGATATGGACATGGTGGTGGTG
GTGGATATGGACACGGTGGTGGTGGATACGGACACCGTGGTGGCGGTGGTGGTGGACGAC
GTGGTGGATACTGCCAGTATGGTTGCTGTGGCCATGGTGACAATGGTTGCTATAGGTGTT
GCTCCTATAAAGGTGAGGCAATGGACAAAGTTACTCAAGCTAAGCCACACAATTAATTAA
TTATGTGTGGAGTACGTAGTATATTATATTAAAACTTTTGTAATGGCAATTATGTAATAT
TATTAGCAATGCTCTTTCTACTTTAGAGCTTGCTATAATATTACTAAAAATGTATTGAAT
AAAAGCCATGTTGTAGTAATTTTATTATCAATATTTTATCATGATATTCATATTTACTGT
ATTTCACTAATTATACCAAAAAGTTTTAGTGC

...

>Contig13317
GATTTGGGAAGACCATATGGTAGAGTTAGTAGAAAGAAGCTGAAGTTGAAATTGTTGAAT
AGTTGTGTTGAGAACAGAGTGAAGTTTTATAAAGCTAAGGTTTGGAAAGTGGAACATGAA
GAATTTGAGTCTTCAATTGTTTGTGATGATGGTAAGAAGATAAGAGGTAGTTTGGTTGTG
GATGCAAGTGGTTTTGCTAGTGATTTTATAGAGTATGACAGGCCAAGAAACCATGGTTAT
CAAATTGCTCATGGGGTTTTAGTAGAAGTTGATAATCATCCATTTGATTTGGATAAAATG
GTGCTTATGGATTGGAGGGATTCTCATTTGGGTAATGAGCCATATTTAAGGGTGAATAAT
GCTAAAGAACCAACATTCTTGTATGCAATGCCATTTGATAGAGATTTGGTTTTCTTGGAA
GAGACTTCTTTGGTGAGTCGTCCTGTTTTATCGTATATGGAAGTAAAAAGAAGGATGGTG
GCAAGATTAAGGCATTTGGGGATCAAAGTGAAAAGTGTTATTGAGGAAGAGAAATGTGTG
ATCCCTATGGGAGGACCACTTCCGCGGATTCCTCAAAATGTTATGGCTATTGGTGGGAAT
TCAGGGATAGTTCATCCATCAACAGGGTACATGGTGGCTAGGAGCATGGCTTTAGCACCA
GTACTAGCTGAAGCCATCGTCGAGGGGCTTGGCTCAACAAGAATGATAAGAGGGTCTCAA
CTTTACCATAGAGTTTGGAATGGTTTGTGGCCTTTGGATAGAAGATGTGTTAGAGAATGT
TATTCATTTGGGATGGAGACATTGTTGAAGCTTGATTTGAAAGGGACTAGGAGATTGTTT
GACGCTTTCTTTGATCTTGATCCTAAATACTGGCAAGGGTTCCTTTCTTCAAGATTGTCT
GTCAAAGAACTTGGTTTACTCAGCTTGTGTCTTTTCGGACATGGCTCAAACATGACTAGG
TTGGATATTGTTACAAAATGTCCTCTTCCTTTGGTTAGACTGATTGGCAATCTAGCAATA
GAGAGCCTTTGAATGTGAAAAGTTTGAATCATTTTCTTCATTTTAAA

...

>Contig19479
...

Total : 19479 contigs

```
>gi|56121773|gb|CV970760.1|CV970760  SlE8 SlE
Solanum lycopersicum cDNA, mRNA sequence
TGATAACCTAGCTTTGTGATCATAATCATGGGTTCTTCTTCTCAAAAGAGGGTTGTAGGG
TTACTTCTTGTCTTGAGTATTTTCTTGGAATTGAGTGCCATCACTTTTGGTGATGATAAG
TTTGAAGAGTCGAGGTGGGGTAATGATTATGGTTGTGGACGATTTGGTAGAAGGGGATGT
AGTGGACGCGATCAAGGTGGTGGTAGAGGTGTTGGAGGAGGGTTTGGAGGCGGAGCTGGT
GGAGGAGGAGGTC
>gi|56121770|gb|CV970757.1|CV970757  SlE5 SlE Solanum lycopersicum cDNA, mRNA sequence
CTTAATTCCAAAGAGTCTTGTTGTACATCCCCCAAATTTGAGGAACTGGTGGTTGGTGCA
TTATCCCTGCTAGTAATGAGGCTTTTTGCTGGCGGCCCCGGCCGGGGTCGTCATTATTTG
CCACATATCTTAGTGGCCATTTTGGCTCTTGTTGATATTGTTTCTGCTGACCCTTACATA
TATGCCTCTCCACCACCACCGTATGTGTACAAATCCCCACCACCTCCTTCTCCTTCTCCA
CCACCACCATACGTGTACAAGTCCCCGCCACCTCCTTCTCCCTCTCCTCCACCACCGTAC
GTGTACAAGT
>gi|117725616|gb|DB705376.1|DB705376  DB705376
Solanum lycopersicum cv. Micro-Tom leaf Solanum ly
copersicum cDNA clone LEFL1088BH06 5', mRNA sequence
TTTTAAAATGAATGCATATAGTGGCGAAGCATGTTCTGTAGTTAATAGGGCTTGTTGTTG
TATTATAGGTATAAGAAAGATACATTTTTGCACTTAGATCCACTATGATGTGAGTTATTC
AACTTGGATTTGAGTGTAAAGTGTATATATAGTTGAGGTCTTCAATCTATTACAATGTTA
CGACAAATTGGATCTACAGTGGGTTCTTGGACTCACAAGAAGATCGTTGATCCTCTTCTC
CAAATCCTTCGTAGGGGTGCAGAGCCAAAACAATTGGCATTCTCTGGGGCTCTTGGTGCT
ACATTGGGTCTCTTTCCCATCTGTGGGGTTGCTGTGTTTCTATGTGGCGTAGCTATTGTA
GTACTTGGATCCTTGTGTCATGCACCAACTGTGTTGTTGGTCAACTTCATTGTTACTCCC
ATTGAGCTGAGTTTGGTGATTCCTTTCCTACGTTTAGGTGAATATGTGAGTGGTGGACCT
CATTTTGCTTTGACCTCAGATGCATTAAAGAGGGTCTTCACTGGTAAAGCTTCGTGGGAA
GTCTTGCTGAGCATTTACCATGCGTTGCTGGGCTGGCTTGTTGCTGTACCATTCATC
…
```

+ 18010 singletons

EGENES Example

Solanum lycopersicum (tomato) (EST)

Genome info	Pathway maps	Gene catalogs	Genome map	Organism list

Organism	esly
Name	S.lycopersicum_est, SOLTU, 4081
Full name	Solanum lycopersicum (tomato) (EST)
Definition	Solanum lycopersicum (tomato) (EST)
Annotation	kaas-blast
Taxonomy	TAX: 4081
Lineage	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; asterids; lamiids; Solanales; Solanaceae; Solanoideae; Solaneae; Solanum; Lycopersicon
Data source	dbEST

KEGG2	PATHWAY	BRITE	GENES	LIGAND	DBGET

KEGG Pathway maps

01110 Carbohydrate Metabolism

00010 Glycolysis / Gluconeogenesis
00020 Citrate cycle (TCA cycle)
00030 Pentose phosphate pathway
00040 Pentose and glucuronate interconversions
00051 Fructose and mannose metabolism
00052 Galactose metabolism
00053 Ascorbate and aldarate metabolism
00500 Starch and sucrose metabolism
00530 Aminosugars metabolism
00520 Nucleotide sugars metabolism
00620 Pyruvate metabolism
00630 Glyoxylate and dicarboxylate metabolism
00640 Propanoate metabolism
00650 Butanoate metabolism
00031 Inositol metabolism
00562 Inositol phosphate metabolism

01120 Energy Metabolism

00190 Oxidative phosphorylation
00195 Photosynthesis
00196 Photosynthesis - antenna proteins
00710 Carbon fixation
00720 Reductive carboxylate cycle (CO2 fixation)
00680 Methane metabolism
00910 Nitrogen metabolism
00920 Sulfur metabolism

01130 Lipid Metabolism

00061 Fatty acid biosynthesis
00062 Fatty acid elongation in mitochondria
00071 Fatty acid metabolism
00072 Synthesis and degradation of ketone bodies
00100 Biosynthesis of steroids
00120 Bile acid biosynthesis
00140 C21-Steroid hormone metabolism
00150 Androgen and estrogen metabolism
00561 Glycerolipid metabolism
00564 Glycerophospholipid metabolism

01180 Biosynthesis of Polyketides and Nonribosomal Peptides

00523 Polyketide sugar unit biosynthesis
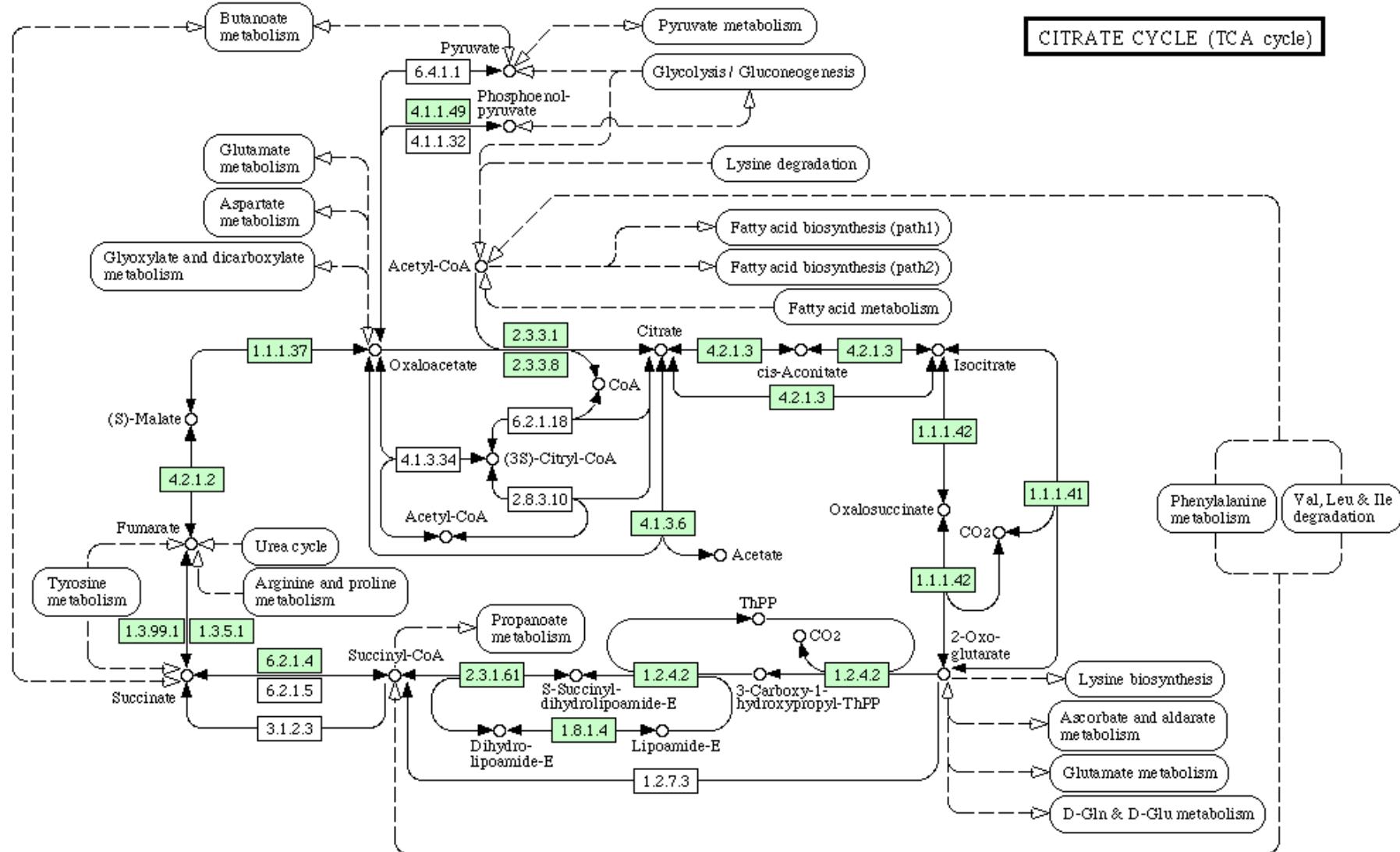
01190 Metabolism of Cofactors and Vitamins

00730 Thiamine metabolism
00740 Riboflavin metabolism
00750 Vitamin B6 metabolism
00760 Nicotinate and nicotinamide metabolism
00770 Pantothenate and CoA biosynthesis
00780 Biotin metabolism
00785 Lipoic acid metabolism
00790 Folate biosynthesis
00670 One carbon pool by folate
00860 Porphyrin and chlorophyll metabolism
00130 Ubiquinone biosynthesis

01195 Biosynthesis of Secondary Metabolites

00900 Terpenoid biosynthesis
00904 Diterpenoid biosynthesis
00902 Monoterpenoid biosynthesis
00903 Limonene and pinene degradation
00901 Indole and ipecac alkaloid biosynthesis
00908 Zeatin biosynthesis
00906 Carotenoid biosynthesis
00905 Brassinosteroid biosynthesis
00940 Phenylpropanoid biosynthesis
00941 Flavonoid biosynthesis
00950 Alkaloid biosynthesis I
00960 Alkaloid biosynthesis II
00232 Caffeine metabolism
00521 Streptomycin biosynthesis
00401 Novobiocin biosynthesis

EGENES Example

CITRATE CYCLE (TCA cycle)
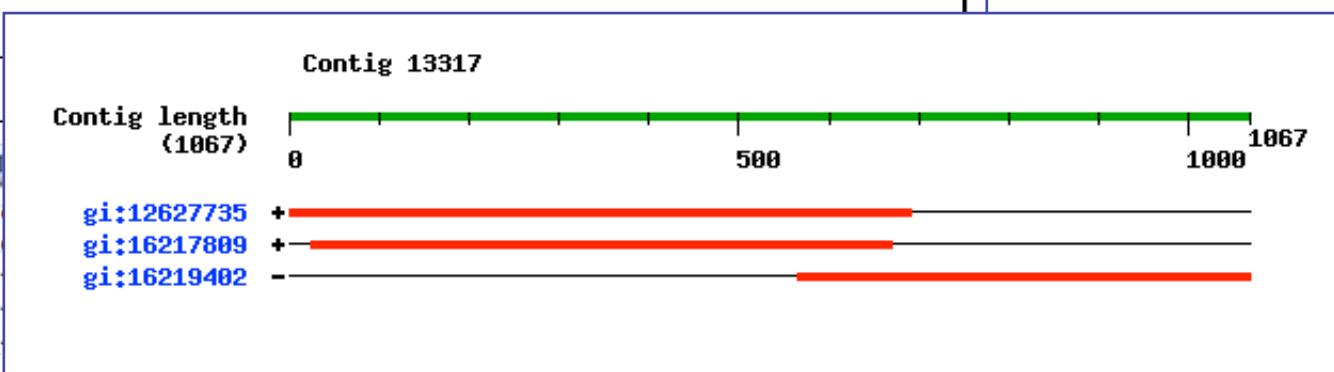
![KEGG] **Solanum lycopersicum (tomato) (EST): 13317**

(Help)

Entry	13317 Contig S.lycopersicum_est
Orthology	KO: K06443 lycopene beta cyclase
Pathway	PATH: esly00906 Carotenoid biosynthesis – General
Class	(BRITE hierarchy)
Other DBs	NCBI-GI: 12627735 16217809 16219402 (Alignment)
LinkDB	(All DBs)
NT seq	1067 nt (NT

Contig 13317

Contig length
(1067)

0   500   1000   1067

gi:12627735 +

gi:16217809 +

gi:16219402 –

```
gatttgggaag
agttgtgttga
gaatttgagtc
gatgcaagtgg
caaattgctca
gtgcttatggattggagggattctcatttgggtaatgagccatatttaagggtgaataat
gctaaagaaccaacattcttgtatgcaatgccatttgatagagatttggttttcttggaa
gagacttctttggtgagtcgtcctgtttttatcgtatatggaagtaaaaagaaggatggtg
gcaagattaaggcatttggggatcaaagtgaaaagtgttattgaggaagagaaatgtgtg
atccctatgggaggaccacttccgcggattcctcaaaatgttatggctattggtgggaat
tcagggatagttcatccatcaacagggtacatggtggctaggagcatggctttagcacca
gtactagctgaagccatcgtcgaggggcttggctcaacaagaatgataagagggtctcaa
ctttaccatagagtttggaatggtttgtggcctttggatagaagatgtgttagagaatgt
tattcatttgggatggagacattgttgaagcttgatttgaaagggactaggagattgttt
gacgctttctttgatcttgatcctaaatactggcaagggttcctttcttcaagattgtct
gtcaaagaacttggtttactcagcttgtgtcttttcggacatggctcaaacatgactagg
ttggatattgttacaaaatgtcctcttcctttggttagactgattggcaatctagcaata
gagagcctttgaatgtgaaaagtttgaatcattttcttcattttaaa
```

KegArray

Mapping microarray expression data
to KEGG pathway data

KegArray

- Java application for handling expression data
  - http://www.genome.jp/download/

- Microarray data with KEGG Gene IDs
  - Coloring based on expression levels
  - Mapping the coloring to pathway data
  - ID conversion available: NCBI-GeneID, IPI, …

- Metabolome data can be also mapped

Other tools

- GECS: Gene Expression to Chemical Structure
  - Microarray data -> Carbohydrate structures
  - http://www.genome.jp/tools/gecs/

- GENIES
  - Gene Network Inference Engine based on Supervised Analysis
  - Integration of multiple omics data to infer gene functions
 - Using pathway data for supervised data
  - http://www.genome.jp/tools/genies/
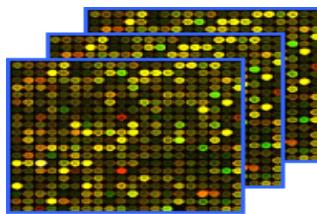
Summary
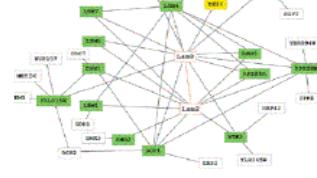
- KEGG is 13 years old database for genomic, chemical, and systems information.

- Genomic information (GENES section) includes complete genomes, draft genomes and EST contigs

- KEGG Orthology plays a key role in connecting genomic and systems information

- Users can input their own sequences (genomes or EST collections) for reconstructing pathway data using KAAS and EGassembler

- KegArray is a Java application for expression data

- GECS and GENIES can be used to infer glycan structures and gene functions, respectively

Network inference from multiple data
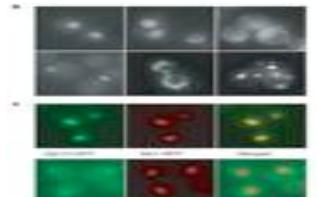
Microarray
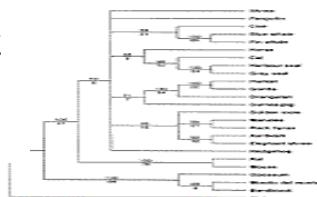
Yeast two-hybrid
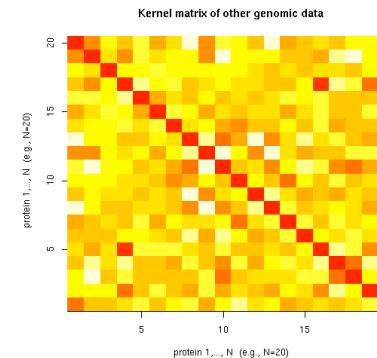
Subcellular localizaton

Phylogenetic profile

Similarity matrix

Network inference