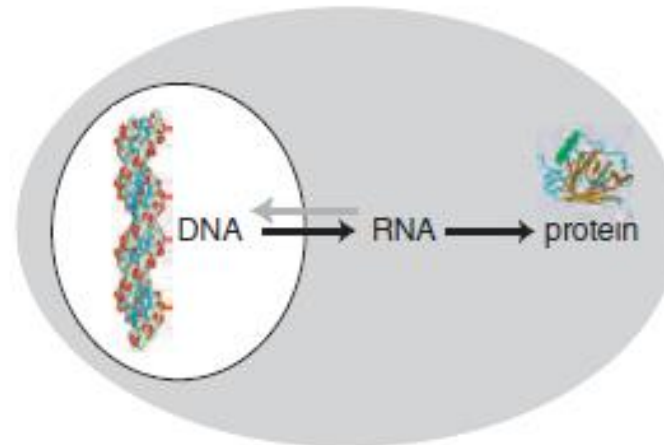
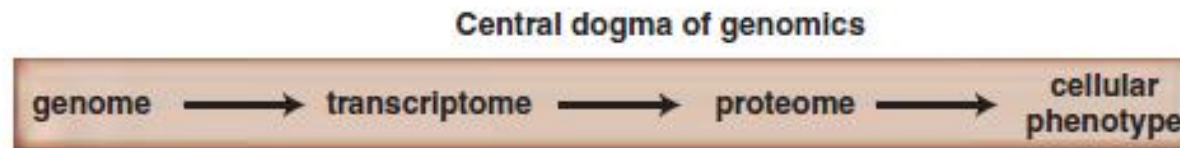
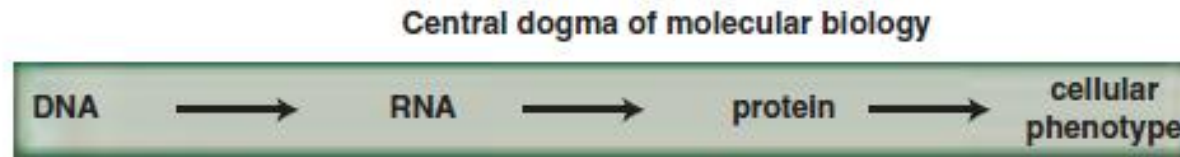


BIOL312

Practical 1

Introduction to Bioinformatics/Types of
Homology/ENTREZ

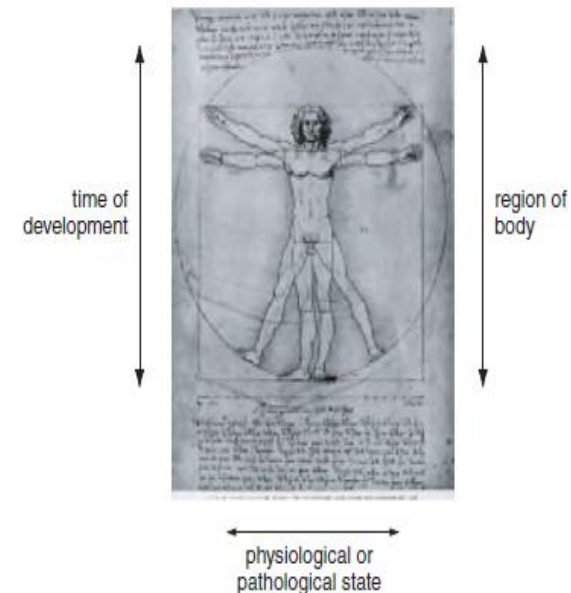
- A first perspective of the field of bioinformatics is the **cell**.
- Bioinformatics has emerged as a discipline as biology has become transformed by the emergence of molecular sequence data. Databases such as the European Molecular Biology Laboratory



While the discipline of bioinformatics focuses on the **analysis of molecular sequences**, genomics and functional genomics are two closely related disciplines.

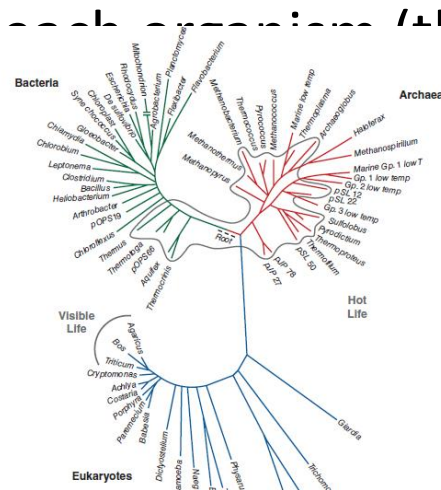
- The goal of genomics is to determine and analyze the complete DNA sequence of an organism, that is, its genome. The DNA encodes genes can be expressed as ribonucleic acid (RNA) transcripts and then, in many cases, further translated into protein.
- Functional genomics describes the use of genome-wide assays to study gene and protein function. For humans and other species, it is now possible to characterize an individual's genome, collection of RNA (transcriptome), proteome and even the collections of metabolites and epigenetic changes, and the catalog of organisms inhabiting the body

- A second perspective of bioinformatics is the **organism**.
- Broadening our view from the level of the cell to the organism, we can consider the individual's **genome (collection of genes)**, including the genes that are expressed as RNA transcripts and the protein products.
- For an individual organism, bioinformatics tools can therefore be applied to describe changes through
 - developmental time,
 - changes across body regions, and
 - changes in a variety of physiological or pathological states.



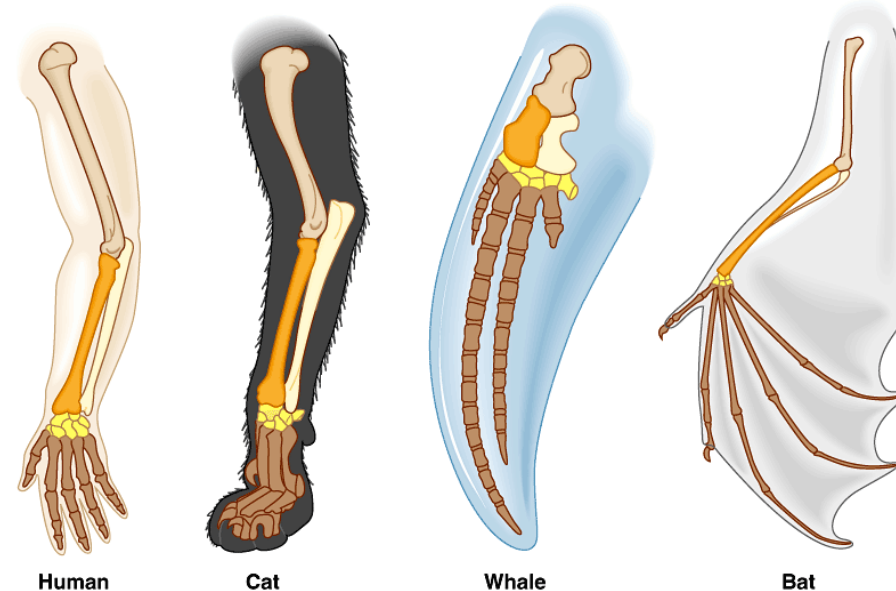
- A third perspective of the field of bioinformatics is represented by the **tree of life**.
- The scope of bioinformatics includes all of life on Earth, including the three major branches of bacteria, archaea, and eukaryotes.
- Viruses, which exist on the borderline of the definition of life, are not depicted here.
- For all species, the collection and analysis of molecular sequence data allow us to describe the complete collection of DNA that comprises each organism (the genome).
- We can further learn about relationships between species and among groups of species, and we can deduce the evolution of life on Earth.

Hugenholtz and Pace (1996), and Pace (1997).



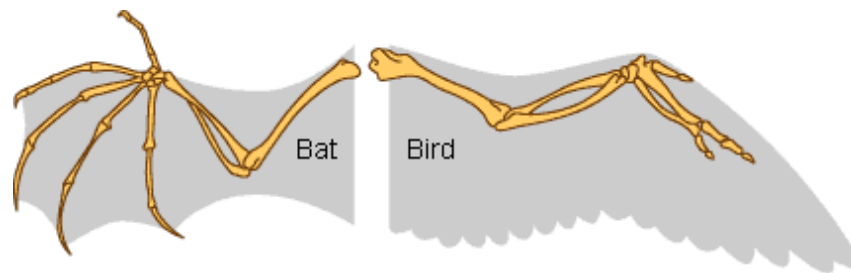
t occur between
ies, and we can
Earth. Adapted from Barns *et al.* (1996),

Homolog structure: Similar structure and position, but different function



Copyright © Pearson Education, Inc., publishing as Benjamin Cummings.

Courtesy of Prof. Ken Sytsma

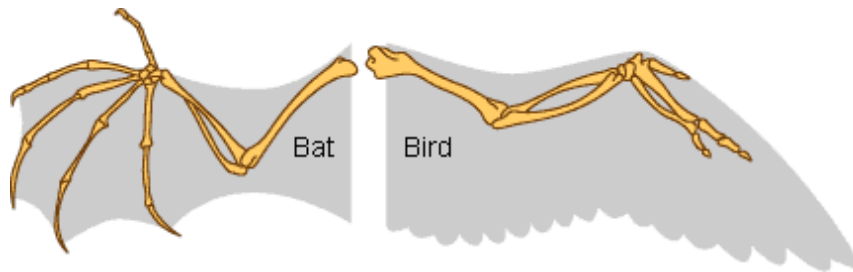


<http://evolution.berkeley.edu>

Analog structure:
Similar function, but
different origin

Analogy

- Different structures which perform the same function (Owen, 1843)
- Convergence: Similarities between organisms that evolved independently.



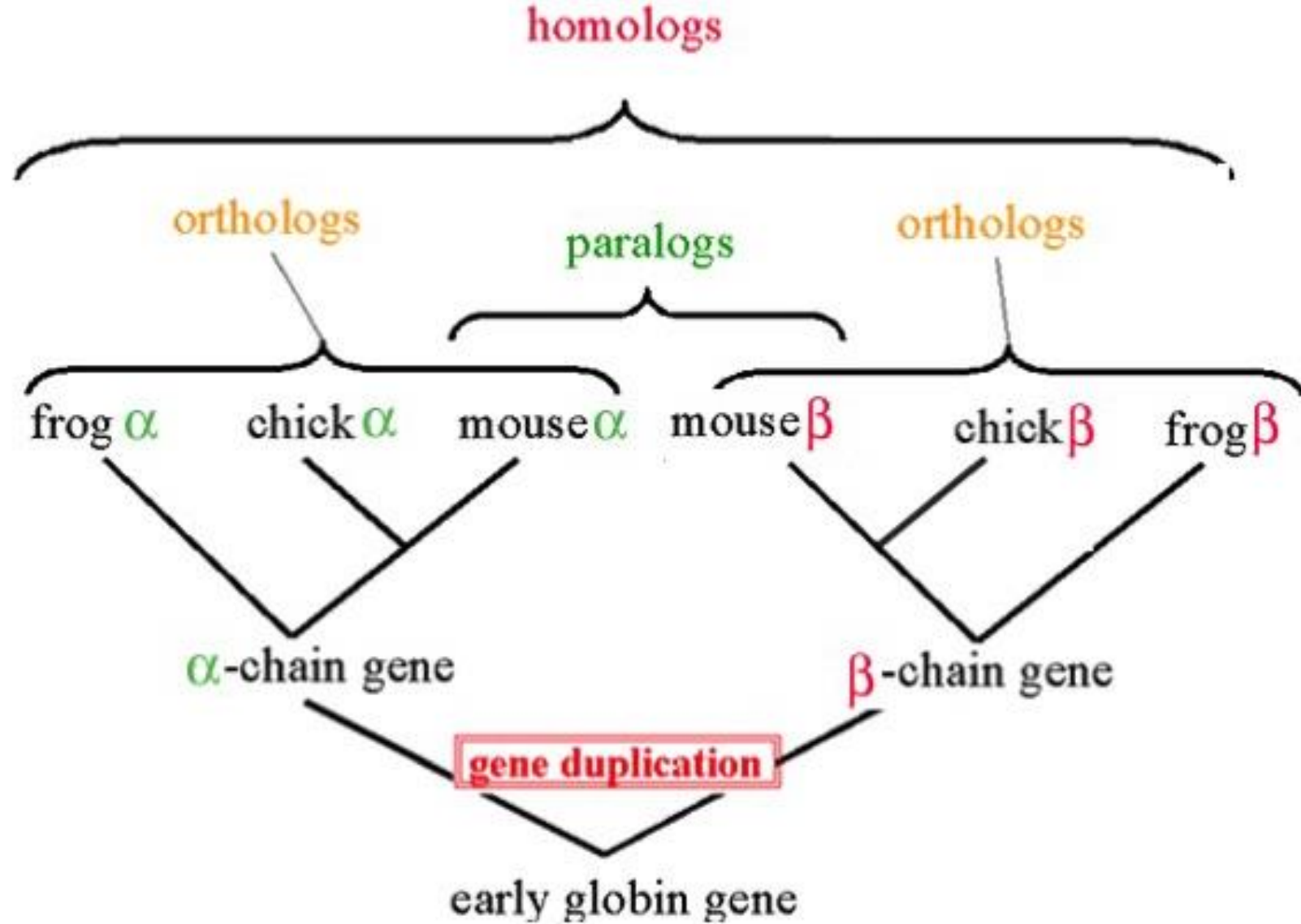
Types of homology

- **Homologues (homologous genes)** are genes that derive from a common ancestor-gene
- **Orthologues** (orthologs) are homologous genes in different species
- **Paralogues** (paralogs) are homologous genes in one species that derive from gene duplication

When one gene is duplicated, the duplication event results in two paralogous genes (paralogues)

Studies of paralogs have found that one paralogue of a pair often retains the ancestral gene's function, while the other paralogue is free to evolve and adopt new functions





http://bioweb.uwlax.edu/GenWeb/Molecular/Bioinformatics/Unit_4/Lab_4-2/lab_4-2.htm

ENTREZ/GQUERY

www.ncbi.nlm.nih.gov/gquery/

NCBI Resources How To Sign in to NCBI

Search NCBI databases [Help](#)

Literature

| | |
|--------------------------------|---|
| Books | books and reports |
| MeSH | ontology used for PubMed indexing |
| NLM Catalog | books, journals and more in the NLM Collections |
| PubMed | scientific & medical abstracts/citations |
| PubMed Central | full-text journal articles |

Health

| | |
|-------------------------------|--|
| ClinVar | human variations of clinical significance |
| dbGaP | genotype/phenotype interaction studies |
| GTR | genetic testing registry |
| MedGen | medical genetics literature and links |
| OMIM | online mendelian inheritance in man |
| PubMed Health | clinical effectiveness, disease and drug reports |

Genomes

| | |
|----------------------------|---|
| Assembly | genome assembly information |
| BioProject | biological projects providing data to NCBI |
| BioSample | descriptions of biological source materials |
| Clone | genomic and cDNA clones |
| dbVar | genome structural variation studies |

Genes

| | |
|------------------------------|--|
| EST | expressed sequence tag sequences |
| Gene | collected information about gene loci |
| GEO DataSets | functional genomics studies |
| GEO Profiles | gene expression and molecular abundance profiles |
| HomoloGene | homologous gene sets for selected organisms |
| PopSet | sequence sets from phylogenetic and population studies |
| UniGene | clusters of expressed transcripts |

Proteins

| | |
|-----------------------------------|---|
| Conserved Domains | conserved protein domains |
| Protein | protein sequences |
| Protein Clusters | sequence similarity-based protein clusters |
| Structure | experimentally-determined biomolecular structures |

Chemicals

| | |
|----------------------------------|--|
| BioSystems | molecular pathways with links to genes, proteins and chemicals |
| PubChem BioAssay | bioactivity screening studies |
| PubChem Compound | chemical information with structures, information and links |

Genomes

| | |
|-----------------------------------|---|
| Assembly | genome assembly information |
| BioProject | biological projects providing data to NCBI |
| BioSample | descriptions of biological source materials |
| Clone | genomic and cDNA clones |
| dbVar | genome structural variation studies |
| Epigenomics | epigenomic studies and display tools |
| Genome | genome sequencing projects by organism |
| GSS | genome survey sequences |
| Nucleotide | DNA and RNA sequences |
| Probe | sequence-based probes and primers |
| SNP | short genetic variations |
| SRA | high-throughput DNA and RNA sequence read archive |
| Taxonomy | taxonomic classification and nomenclature catalog |

Chemicals

BioSystems

PubChem BioAssay

PubChem Compound

PubChem Substance

You are here: NCBI > GQuery

GETTING STARTED

- [NCBI Education](#)
- [NCBI Help Manual](#)
- [NCBI Handbook](#)
- [Training & Tutorials](#)

RESOURCES

- [Chemicals & Bioassays](#)
- [Data & Software](#)
- [DNA & RNA](#)
- [Domains & Structures](#)
- [Genes & Expression](#)
- [Genetics & Medicine](#)

POPULAR

- [PubMed](#)
- [Bookshelf](#)
- [PubMed Central](#)
- [PubMed Health](#)
- [BLAST](#)
- [Nucleotide](#)

FEATURED

- [Genetic Testing](#)
- [PubMed](#)
- [GenBank](#)
- [Reference](#)
- [Gene Expression](#)
- [Map Viewer](#)

Search

- glucokinase NOT similar

18)

- 1. [Xenopus \(Silurana\) tropicalis glucokinase \(hexokinase 4\) \(gck\), mRNA](#)
2,442 bp linear mRNA
Accession: NM_001102851.1 GI: 156717561
[GenBank](#) [FASTA](#) [Graphics](#)

- 2. [Trypanosoma grayi glucokinase partial mRNA](#)
1,131 bp linear mRNA
Accession: XM_009310115.1 GI: 686631888
[GenBank](#) [FASTA](#) [Graphics](#)

- 3. [Human liver glucokinase \(ATP:D-hexose 6-phosphotransferase\) mRNA, complete cds](#)
2,550 bp linear mRNA
Accession: M69051.1 GI: 183226
[GenBank](#) [FASTA](#) [Graphics](#)

- 4. [Z.mobilis glucose transport protein \(glf\), glucose-6-phosphate dehydrogenase \(zwf\), 6-phosphogluconate dehydratase \(edd\) and glucokinase \(glk\) genes, complete cds](#)
6,440 bp linear DNA
Accession: M60615.1 GI: 155589
[GenBank](#) [FASTA](#) [Graphics](#)

- 5. [Homo sapiens glucokinase \(GCK\) gene, 5' end of coding regions from three isoforms](#)
4,261 bp linear DNA
Accession: M90298.1 GI: 183276
[GenBank](#) [FASTA](#) [Graphics](#)

- 6. [Brucella sp. BO1 partial glk gene for glucokinase, strain BO1](#)
475 bp linear DNA
Accession: AM884814.1 GI: 170015682
[GenBank](#) [FASTA](#) [Graphics](#)

- 7. [Mus musculus glucokinase mRNA, exon 1](#)
146 bp linear mRNA
Accession: L38988.1 GI: 886340
[GenBank](#) [FASTA](#) [Graphics](#)

Escherichia coli (107)
Streptococcus equi subsp.
zooepidemicus (104)
Yersinia pestis (67)
All other taxa (4033)
More...

Find related data

Database:

Search details

GLUCOKINASE[All Fields] NOT SIMILAR[All Fields]

[See more...](#)

Recent activity

[Turn Off](#) [Clear](#)

- Nucleotide
- Nucleotide
- Protein
- Protein
- Nucleotide

Nucleotide

Nucleotide

Search

Advanced

Help

Display Settings: GenBank

Xenopus (Silurana) tropicalis glucokinase (hexokinase 4) (gck), m

NCBI Reference Sequence: NM_001102851.1

[FASTA](#) [Graphics](#)

Go to:

LOCUS NM_001102851 2442 bp mRNA linear VRT 21-APR-2014
 DEFINITION Xenopus (Silurana) tropicalis glucokinase (hexokinase 4) (gck), mRNA.
 ACCESSION NM_001102851
 VERSION NM_001102851.1 GI:156717561
 KEYWORDS RefSeq.
 SOURCE Xenopus (Silurana) tropicalis (western clawed frog)
 ORGANISM [Xenopus \(Silurana\) tropicalis](#)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Pipoidea; Pipidae; Xenopodinae; Xenopus; Silurana.
 REFERENCE 1 (bases 1 to 2442)
 AUTHORS Klein SL, Strausberg RL, Wagner L, Pontius J, Clifton SW and Richardson P.
 TITLE Genetic and genomic tools for Xenopus research: The NIH Xenopus initiative
 JOURNAL Dev. Dyn. 225 (4), 384-391 (2002)
 PUBMED [12454917](#)
 COMMENT PROVISIONAL [REFSEQ](#): This record has not yet been subject to final NCBI review. The reference sequence was derived from [BC135716.1](#).

Send:

- Complete Record
- Coding Sequences
- Gene Features

Choose Destination

- File
- Clipboard
- Collections
- Analysis Tool

Add 1 items.

[Add to Clipboard](#)

SAVED FOR 8 HOURS

Articles about the gck gene

- Generation and initial analysis of more than 15,000 full-length [Proc Natl Acad Sci U S A. 2002]
- Genetic and genomic tools for Xenopus research: The NIH Xenopus initi; [Dev Dyn. 2002]

See all...

Pathways for the gck gene

- Metabolism
- SLC-mediated transmembrane transport

Saving the DNA Sequence

```
1   aaaggactga gtgttcatta aaaaaatgac tacatcccag aaatactttt aatcaccata
61  gatacctggg caggacatt tagacacttc ccacaagatc tactaaaaga tacactgaca
121 taaaagccct tctataaacc cttcattgag ctgcatctgc tgcggaagat tccctttcaa
181 gcgagcatgg aaaagtttga acagaacgag gtggatcaaa tcctgtctga gttccatctc
241 caggaggaag atttgcatgt tctgatgtgt aggatgcagg ctgagatgga gaggggtcct
301 cacctggaaa ccaatgagga gccagtggtt aaaatgctgc ccacctatgt tcgttctaca
361 cctgatggct ccgaggtggg tgacttccta gccctggacc ttgggggaac caactttcga
421 gtcatgttgg taaaggttgg ggaggatttg gaagggcagt ggaaagtgga gacaaaacac
481 aaaatgtatt caattccaga ggatgccatg actggaacag cggagatgct attcgactac
541 attgcagaat gcatttctga ttacctggac cagcaaaaaca tgaaacacaa gaaactgcc
601 ctgggattta cttctctttt ccagttagg catgaggata ttgacaaggg tattcttctc
661 aactggacca aaggattcaa agcatctgga gcagaaggaa acaatgttgt tggactcctt
721 agagatgcta tcaagagacg tggtgatttt gagatggatg tggttgccat ggtgaatgat
781 acagtggcca ctatgatatc ttgttactat gaagaccatc actgtgaggt tggcctgac
841 gtagggacag gttgtaacgc ctgctacatg gaggagatga gtaatgtgga gttggtagaa
901 ggagaggagg gacgtatgtg tgtgaacaca gagtggggag catttggcga cacaggggaa
961 ctggaggatt tccgtctgga gtatgaccga gtggtggatg aggcattctc taatcctggt
1021 cagcaactct atgaaaagat gattggtgga aaatacatgg gagagctggt gagactagtc
1081 ctgataaaga tgtgfaatga gaatctgata tttggaggag aatcttctga aaagctgaag
1141 acccgaggag ctttgacac acaatttgtg tcccagatag agggggacac ctcagatttt
1201 aagcagactc tcaacatcct gcggactcta ggggttcagg ccacaattgg agactgtcat
1261 gctgtgcgct tagcatgtga gagtgtatct actcagagctg ccgtcatgtg ttcctctggc
1321 cttgctgcca tattgaaccg catgcaccaa agcagaagag gggaaactgtc tcgaattaca
1381 gtgggagtcg atggatctgt gtacaaacta catcctagtt tcaaagacaa gttccatgcc
1441 actgtgcaca aactgacttc aggctgtgaa attaccttca ttcagtcaga ggaaggaagc
1501 gggcgtggtg ctgcccta atcagcagtg gcgtataaga tggctgtcct gattggtcac
1561 tagtgagaaa tgcatacatg acatacagac ccagaagcca ctcttaggtt tgcagtgcat
1621 tccacagatc tgaagagatt cttcatatt tacaatgtag aacatctgac aaaacattct
1681 gcttctaagt gcaattactt actaagggga atcttactct aaaacagctt tataaaatgc
1741 acagagagct aatttcgctg ttgcagtggt cctctggata ctgccataag gtgtactgaa
1801 gatatatcca gaaatcagta ggcaagagat ccaaagacca tttctatggt tacacaataa
1861 atggctttca caaacatggt ctacagagca cagatccacc aaacattgga gtgtccacta
1921 accaagcgaa atgaattcaa cactggaaaa taactacaga agacatttgc aaaatgctaa
1981 gtcaatagca actaacagat aaggggttcc aaagataatg gctattatta gaccctaagg
2041 agattgagtt atccacaaa aactgccact aataagtgtt caacctggtg tttcttagac
2101 atcatatcaa atcttgacaa aagttatgtg gcggagcatt caacctatga caatttacca
2161 ctagcatttg ctgcttacgc cttcaccgg ttttaagagtt acttagagct tacacctcaa
2221 gtggagaacc accaggcagc acttctgtga tccagatcaa tgaccattcg tcatgtatat
2281 tacacagcca tcaaaatgtg agaatgaact aaatgtatca atagccacac cagcagccac
2341 tgttctctct aaacttttca agcatcaaca ttaaatcatc taaaagcctc ccatcacggg
2401 aacatttgtg aatgtcaaaa ggctgtcaaaa aaaaaaaaaa aa
```






AUTHORS Barnell,W.O., Yi,K.C. and Conway,T.
TITLE Sequence and genetic organization of a Zymomonas mobilis gene cluster that encodes several enzymes of glucose metabolism
JOURNAL J. Bacteriol. 172 (12), 7227-7240 (1990)
PUBMED [2254282](#)

FEATURES Location/Qualifiers
source 1..6440
/organism="Zymomonas mobilis subsp. mobilis str. CP4 =
NRRL B-14023"
/mol_type="genomic DNA"
/strain="Cp4"
/db_xref="taxon:[627343](#)"
gene 185..1606
/gene="glf"
CDS 185..1606
/gene="glf"
/codon_start=1
/transl_table=11
Coding sequence /product="glucose transport protein"
/protein_id="[AAA27691.1](#)"
/db_xref="GI:155590"
/translation="MSESSQGLVTRLALIAAIGLLFGYDSAVIAAIGTPVDIHFA
PRHLSATAAASLSGMVVAVLVGCVTGSLLSGWIGIRFGRRGLLMSSICFVAAGFGA
ALTEKLFGTGGSALQIFCFRFLAGLGIGVVSTLTPTYIAEIRPPDKRGQMVSGQQMA
IVTGALTGYIFTWLLAHFGSIDWNASGWCWSPASEGLIGIAFLLLLLTAPDTPHWLV
MKGRHSEASKILARLEPQADPNLTIQKIKAGFDKAMDKSSAGLFAFGITVVFAGVSA
AFQQLVGINAVLYAPQMFQNLGFGADTALLQTISIGVVNFIFTMIASRVDRFGRKP
LLIWGALGMAAMMAVLGCCFWFKVGGVPLASVLLYIAVFGMSWGPVCWVWVLSMFPS
SIKGAAMPIAVTGQWLANILVNFLFKVADGSPALNQTFNHGF SYLVFAALSILGGLIV
ARFVPETKGRSLDEIEEMWRSQK"
gene 1752..3209
/gene="zwf"
CDS 1752..3209
/gene="zwf"
/codon_start=1
/transl_table=11
/product="glucose-6-phosphate dehydrogenase"
/protein_id="[AAA27692.1](#)"
/db_xref="GI:155591"
/translation="MTNTVSTMILFGSTGDLSQRMLLPSLYGLDADGLLADDLRIVCT

Taxonomy

Recent activity

[Tur](#)

-  [Z.mobilis glucose transport prot glucose-6-phosphate dehydrog](#)
-  [gck glucokinase \(hexokinase 4\) \(Silurana\) tropicalis\]](#)
-  [Xenopus \(Silurana\) tropicalis gl \(hexokinase 4\) \(gck\), mRNA](#)
-  [GLUCOKINASE NOT SIMILAR \(-](#)
-  [glycokinase \(0\)](#)

[Region](#) /db_xref="CDD:182225"
15..454
/region_name="MFS"
/note="The Major Facilitator Superfamily (MFS) is a large and diverse group of secondary transporters that includes uniporters, symporters, and antiporters. MFS proteins facilitate the transport across cytoplasmic or internal membranes of a variety of...; cd06174"
[Site](#) /db_xref="CDD:119392"
order(28,31..33,35..36,53,56..57,60,72..73,75..76,130..131,134..135,138..139,142,154..155,158..159,165,281,284..285,288..289,292,305,309,313,316,320,370,373..374,377,383..384,387,398..399,402..403,406..407,410)
/site_type="other"
/note="putative substrate translocation pore"
[CDS](#) /db_xref="CDD:119392"
1..473
/gene="glf"
/coded_by="M60615.1:185..1606"
/transl_table=11






ORIGIN

```
1 mssessqglv trlaliaaig gllfgydsav iaaigtvdi hfiaphlsa taaaslgmv
61 vvavlvgcvt gllsgwigi rfgrrgllm ssicfvaagf gaalteklfg tggسالqifc
121 ffrflaglgi gvvstltpty iaairppdkr gqmvsgqma ivtgaltgyi ftwllahfgs
181 idwvnasgwc wspaseglig iafl111lta pdtphwlvmk grhseaskil arlepqadpn
241 ltiqkikagf dkamdksag lfafgitvvf agvsvaafqq lvginavlyy apqmfqnlgf
301 gadtallqti sigvvnfift miasrvvdrf grkp11iwa lgmaammavl gccfwfkvvgg
361 vlplaslly iavfgmswgp vcwvvlsemf pssikgaamp iavtgqwan ilvnflfkva
421 dgspalnqtf nhgfsylvfa alsilggliv arfvpetkgr sldeieemwr sqk
```

//

[Nucleotide](#)
[PubMed](#)
[Related Structures \(List\)](#)
[Related Structures \(Summary\)](#)
[Taxonomy](#)

Recent activity

-  [glucose transport prote mobilis subsp. mobilis](#)
-  [Z.mobilis glucose trans glucose-6-phosphate c](#)
-  [gck glucokinase \(hexo \(Silurana\) tropicalis\]](#)
-  [Xenopus \(Silurana\) tro \(hexokinase 4\) \(gck\), n](#)
-  [GLUCOKINASE NOT S](#)

Save amino acid sequence

You are here: [NCBI](#) > [Proteins](#) > [Protein Database](#)

GETTING STARTED

[NCBI Education](#)
[NCBI Help Manual](#)
[NCBI Handbook](#)
[Training & Tutorial](#)

RESOURCES

[Chemicals & Bioassays](#)
[Data & Software](#)
[DNA & RNA](#)
[Domains & Structures](#)

POPULAR

[PubMed](#)
[Bookshelf](#)
[PubMed Central](#)
[PubMed Health](#)

FEATURED

[Genetic Testing Registry](#)
[PubMed Health](#)
[GenBank](#)
[Reference Sequences](#)

NCBI INFORMATION

[About NCBI](#)
[Research at NCBI](#)
[NCBI News](#)
[NCBI FTP Site](#)