

# *The Pathway Tools Genome Browser*

*Markus Krummenacker*  
*Bioinformatics Research Group*  
*SRI International*  
*kr@ai.sri.com*

# *Author Crediting System*

- **For Pathways & Enzymes**
- **Types of Credit: Created, Reviewed, Revised**
- **Home Pages for Authors and Organizations**
  
- **CREDITS slot lists author and organization frames**
- **CREDITS:**
  - SRI**
  - CREATED: 3357505001**
  - |caspi|**
  - CREATED: 3357505001**

# *General Features*

- **Visual integration of several annotation types**
- **Shows multiple lines for more genome context**
- **Intuitive navigational controls**
- **Red cursor on whole chromosome shows zoomed region**
- **Semantic zoom levels**
- **Whole chromosome Poster printing (in 10.5)**

# *Prokaryotic-Friendly Features*

- Operons shown with grey background
- Shows transcription starts and terminators
- Navigation through origin of circular chromosomes
- Visual distinction between gene types:
  - Protein
  - RNA
  - Pseudo genes / Phantom genes
- Overlapping genes shown at a higher level

# *Mouse-over Labels*

- Labels show names of gene and (protein) product
- On WWW, pathway(s) are mentioned
- Transcriptional regulators of operons
- Intergenic distances (in bp):
  - $\text{xdhA} < +11 \text{ xdhB} -3 > \text{xdhC}$

# *Comparative Genome Browser*

- **One master organism that anchors the comparison**
- **Set of selected organisms**
- **One gene of master organism directs the alignment**
- **Ortholog dblinks are followed to the selected organisms**
- **Ortholog links can be bulk loaded (UG II 2.4.10.5)**
- **BioCyc uses best bi-directional BLAST hits (from CMR), stored in a MySQL db**
- **Genes connected by ortholog links have same color**

## *Bulk Loading Ortholog DB links*

- Ortholog links can be bulk loaded (UG II 2.4.10.5)
- 2 Tab-delimited files needed per PGDB
- Database definition file:

FRANTCYC FrantCyc <http://BioCyc.org/...~A>

- Link file:

\$ORGANISM ECOLI

EG11024 FRANTCYC TRPA Ortholog

- Load files:

pathway-tools -org ECOLI -dbdef file1 -linkdef file2

# Gene Sequence Retrieval Dialog

**Request Nucleotide Sequence from Chromosome**

This panel allows specifying a sequence region from the chromosome.  
In addition to the gene, flanking upstream and downstream regions can be requested.

All numbers are basepair positions.

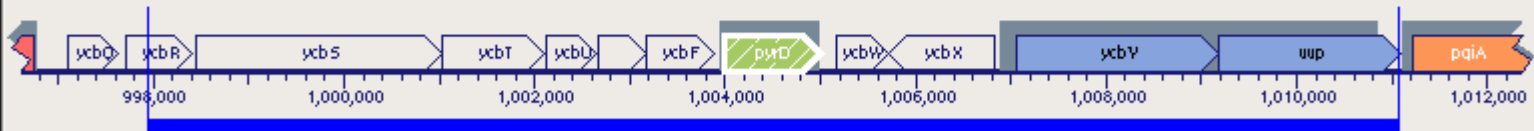
pyrD Start:	1,003,991	1,005,001	pyrD End
-------------	-----------	-----------	----------

Additional Flanking Regions:	Upstream: -	6,060	+ 6,060	Downstream
------------------------------	-------------	-------	---------	------------

<b>Requested Region:</b>	Left End Position:	997,931	1,011,061	Right End Position
--------------------------	--------------------	---------	-----------	--------------------

Requested Sequence Length: 13,130

Return Reverse Complement?:  The current setting will return the coding strand.



AAAATCAGCAGACTTTATTATTACACCACCGCTATTTGTGCTGAATCCGGCAAATGAGAATCTGTTACGCATTATGTACAT. . .  
... CAGCCGCATGAGCAGACGCCAAAAAGTGCTTTGCTGATATGGCTGCTGCAGAGCAGGAGCTGGAGCAAAGCCTTTGAACGCTGG



## *Comparison to GBrowse*

- **GBrowse is a popular open source browser**
- **GBrowse is only intended for WWW**
- **GBrowse is eukaryote-oriented:**
  - **Depicts introns / exons (Ptools do not yet)**
  - **Does not support circular chromosomes**
- **GBrowse does not line-wrap to show more context**
- **GBrowse displays are spread out among tracks, not compact (separate tracks for operons, pseudogenes, etc.)**
- **GBrowse is user-customizable, by adding arbitrary tracks, using GFFx file format**

# *ToDo List*

- **WWW: autom. determining browser size**
- **Adding tracks, in GFF2 file format**
- **Coloring genes according to expression data score values in GFF2 ?**
- **Low-res whole chromosome overview ?**
- **Adding DNA sequence and 6 translation frames**
- **Depicting introns / exons**
- **Popup of transcription factor binding sites**
- **Frameshift locations for pseudogenes**