



# Use of Pathway Tools to Aid Comparative and Functional Genomics

**Ian T. Paulsen**

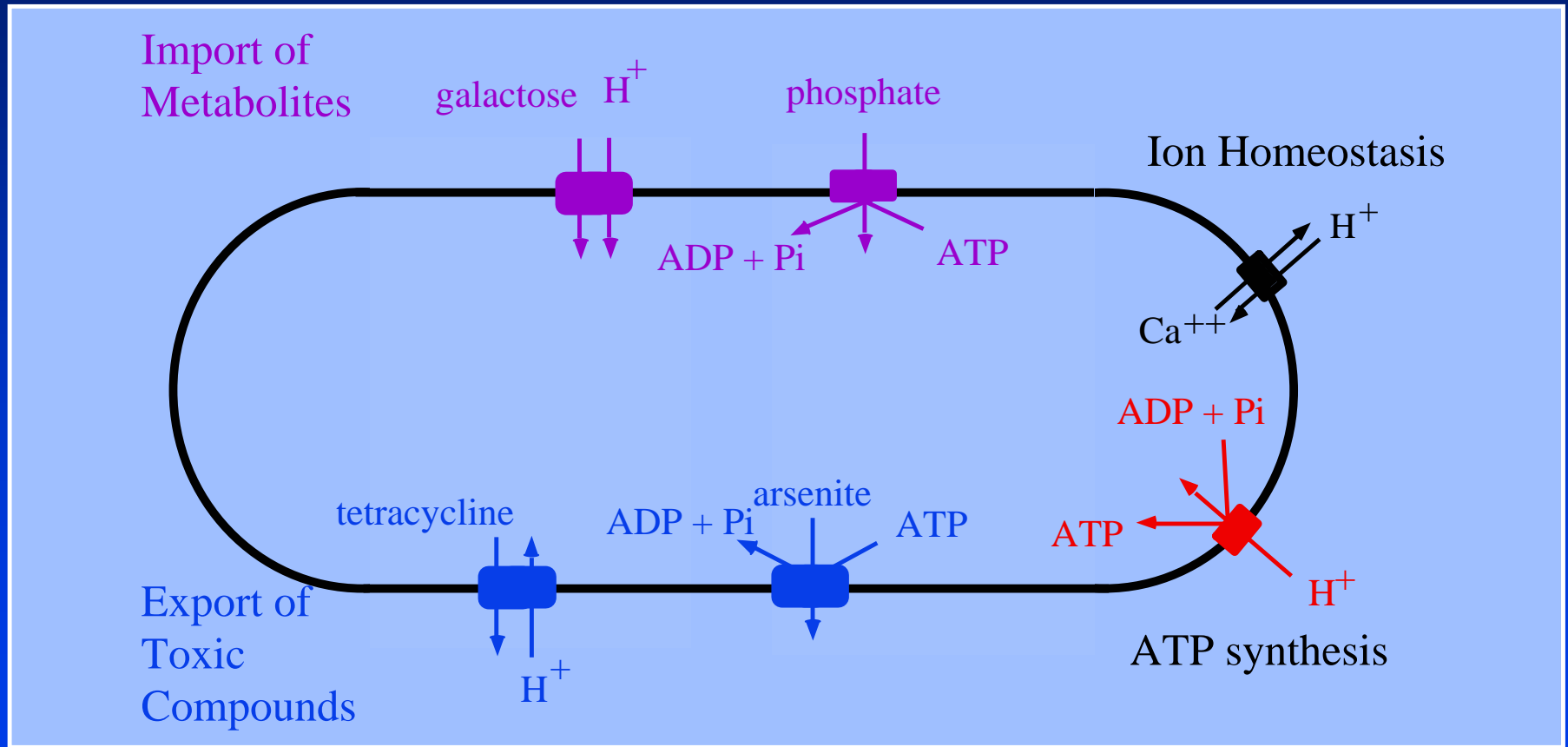
**The Institute for Genomic Research**



**TIGR**

THE INSTITUTE FOR GENOMIC RESEARCH

# Representative *E. coli* transport systems

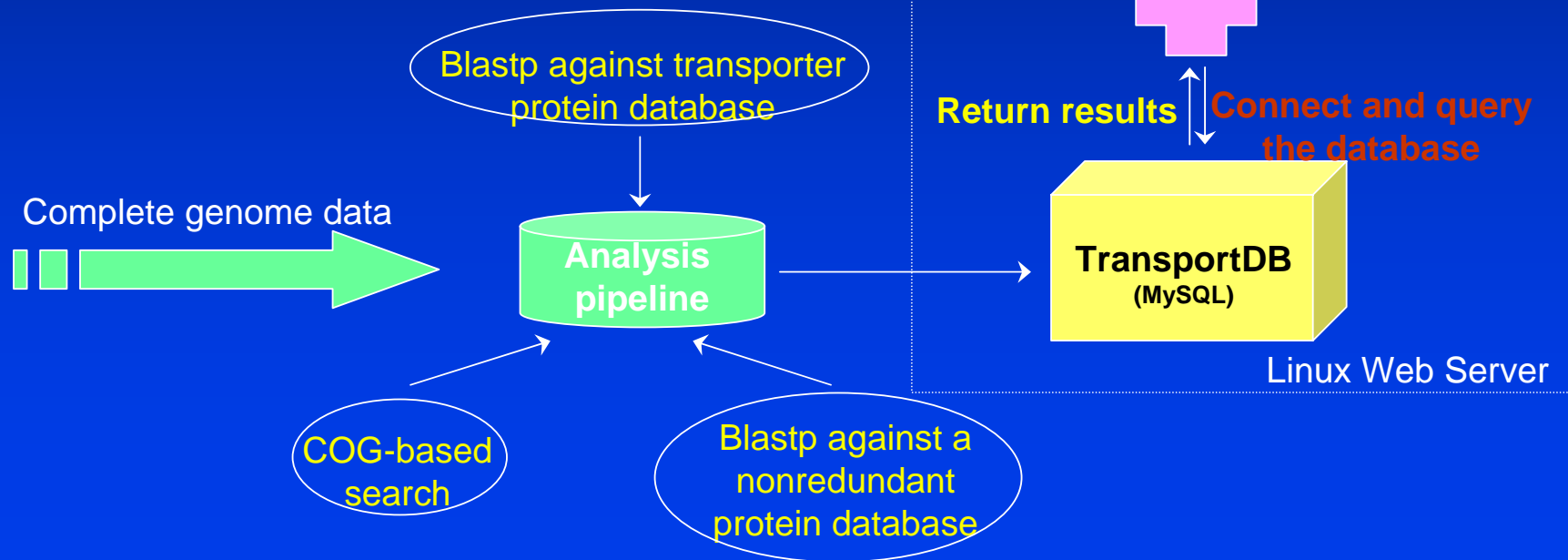


# Transporter Genome Analysis

## Transporter Protein Data Sheet

- Transporter Name
- Transporter Family
- Possible Substrate/Function
- Genomic Locus
- COG
- Protein Sequence

<http://www.membranetransport.org>



# TransportDB

## Database Statistics

As of May, 2006

Total transport proteins: 85,234

Total families: 144

Total organisms: 244

Major Families:	ABC	37,031	
	MFS	8,886	
	DMT	2,572	
	P-ATPase	2,172	
	APC	1,837	
	RND	1,670	
	MOP	1,664	
	TRAP-T	1,363	
	VIC	1,075	
	<b>Total</b>	<b>51,970</b>	<b>(61%)</b>

# Transporter Proteins in *Escherichia coli* K12-MG1655

ATP-Dependent    Ion Channel    Phosphotransferase System    Secondary Transporter    Unclassified Transporters

print all transporters

< home >  
All Organisms (224 total)  
E. coli K12-MG1655

Bacteria (179 total)  
-- Select Bacteria --  
Archaea (23 total)  
-- Select Archaea --  
Eukaryota (22 total)  
L. major Friedlin

Organisms List **NEW!**  
Structure **NEW!**  
Search  
Compare Organisms  
Blast  
Phylogenetics

Genome Size(Mb):  
Total Transporter Protein  
No. of Transporters per Mb genome:

TRANSPORTE  
ATP-Dependent

ABC Family\*  
F-ATPase Family\*  
P-ATPase Family\*

### The ATP-binding Cassette (ABC) Superfamily

PROTEIN	ABC	membrane	binding protein	SUBSTRATE	
AraG (b1900) <b>ZPAS 0</b>	AraH (b4460) <b>ZPAS 10</b>	AraF (b1901) <b>ZPAS 0</b>	arabinose	3.A.1.2.2	●
HisP (b2306) <b>ZPAS 0</b>	HisQ (b2308) <b>ZPAS 5</b>	ArgT (b2310) <b>ZPAS 0</b>	lysine/arginine/ornithine	3.A.1.3.1	●
ArtP (b0864) <b>ZPAS 0</b>					
BtuD (b1709) <b>ZPAS 0</b>					

### Transporter Protein AraH

Transporter Function    Genome Locus    Transmembrane Segment    Sequence    Publication

	<i>B. melitensis</i>	<i>B. subtilis</i>	<i>B. suis</i>	<i>E. coli</i>
Genome Size(Mb):	3.3	4.2	3.31	4.6
Total Transporter Proteins:	366	419	358	477
No. of Transporters per 100 Kb genome:	11.1	10	10.8	10.4

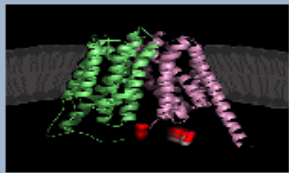
	<i>B. melitensis</i>	<i>B. subtilis</i>	<i>B. suis</i>	<i>E. coli</i>
ATP-Dependent	283 (77%)	209 (50%)	286 (80%)	216 (45%)
ABC family	268	196	280	203
ArsAB family	0	0	0	0
F-ATPase family	10	9	0	9

#### Transport Function

Transporter Name:	AraH
Transporter Type:	ATP-Dependent
Transporter Family:	ABC (TC# 3.A.1) The ATP-binding Cassette (ABC) Superfamily
Transporter Subfamily:	membrane
Substrate/Function:	arabinose
TC#:	3.A.1.2.2

#### Genome Locus

PID:	49176167 <a href="#">Blast</a>
Source:	Escherichia coli K12-MG1655
Chromosome:	-
Location:	1980578..1981567
Gene:	b4460
Length:	329
Strand:	-
Code:	-
COG:	-
Product:	high-affinity L-arabinose transport protein (ABC superfamily, membrane)



< home >

All Organisms  
(224 total)

-- Select Organism --

Bacteria (179 total)

-- Select Bacteria --

Archaea (23 total)

-- Select Archaea --

Eukaryota (22 total)

-- Select Eukaryota --

Organisms List

Structure

Search

Compare Organism

Blast

Phylogenetics

Links



46318

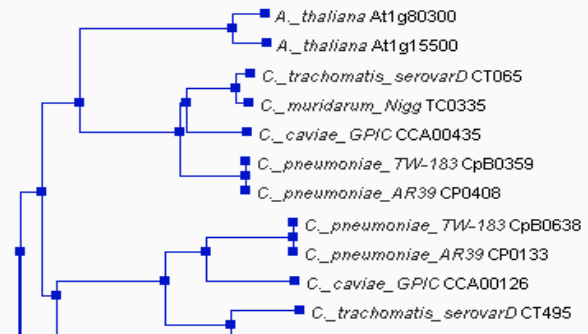


# Genomic Comparisons of Membrane Transporters

## TransportDB

Please choose the search option and

Transporter Type:



## Crystal Structure of Membrane Transporters

**T**ransportDB is a relational database complement for organisms whose membrane transport complement was determined, and functional predictions are made from the pull down menus on this page.

**F**or each organism, a summary page lists possible transporter families and individual transporter families. Possible substrates are shown with links to information. You can also compare the

**A** search engine is set up for easy search and their substrates. You can also blast

**W**ith the rapid development of sequencing are available for the analysis of their

### Secondary Transporter

**AAA:** The ATP:ADP

[FASTA](#)

**AAAP:** The Amino Ac

[FASTA](#)

**AAE:** The Aspartate:Alanine Exchanger (AAE) Family

[FASTA](#)

[ALN](#)

[DND](#)

[Tree](#)

**AbgT:** The p-Aminobenzoyl-glutamate Transporter (AbgT) Family

[FASTA](#)

[ALN](#)

[DND](#)

[Tree](#)

<a href="#">ABC</a>	<a href="#">Amt</a>	<a href="#">Annexin</a>	<a href="#">Bcl-2</a>	<a href="#">CIC</a>	<a href="#">CPA1</a>	<a href="#">DAACS</a>	<a href="#">DMT</a>	<a href="#">F-ATPase</a>
<a href="#">GPTS</a>	<a href="#">Hsp70</a>	<a href="#">IISP</a>	<a href="#">LIC</a>	<a href="#">MFS</a>	<a href="#">MIP</a>	<a href="#">MIT</a>	<a href="#">MscL</a>	<a href="#">MscS</a>
<a href="#">NSS</a>	<a href="#">O-CIC</a>	<a href="#">P-ATPase</a>	<a href="#">PCC</a>	<a href="#">PIT</a>	<a href="#">RND</a>	<a href="#">SSPTS</a>	<a href="#">Tat</a>	<a href="#">VIC</a>

### ATP-Dependent

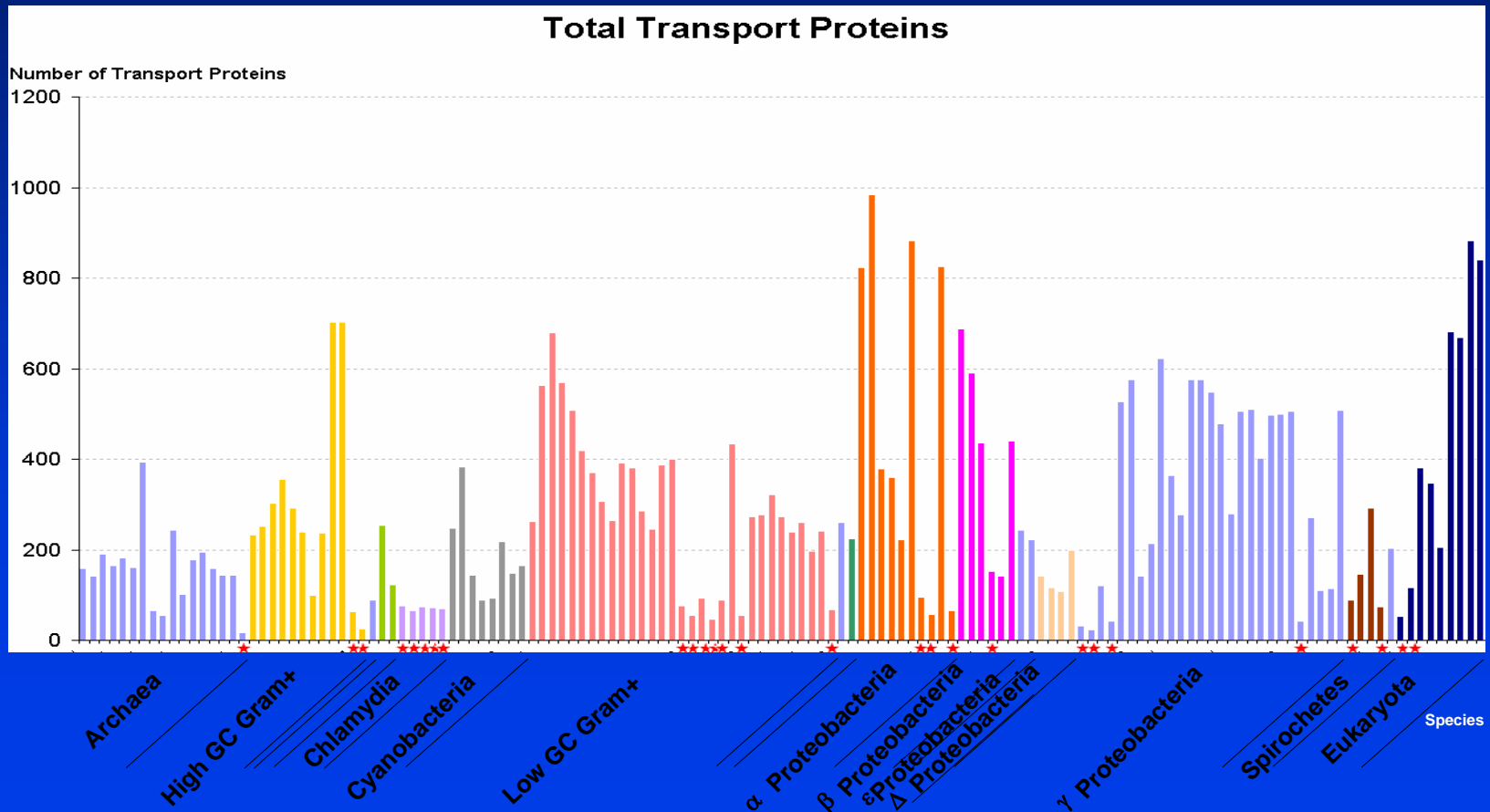
#### ABC: The ATP-binding Cassette (ABC) Superfamily

	PDB	Description	Method	Resolution	Links
	1W0D	Crystal Structure Of ModA, A Molybdate Protein, Complexed With Tungstate <b>ModA</b> [ <i>Escherichia coli</i> K12-MG1655]	X-Ray Diffraction	1.75	<a href="#">PDB</a> <a href="#">MMDB</a> <a href="#">Entrez</a> <a href="#">PubMed</a> <a href="#">TransportDB</a>
	10GC	The Structure Of Bacillus Subtilis RbsD Complexed With D-Ribose <b>RbsD</b> [ <i>Bacillus subtilis</i> 168]	X-Ray Diffraction	2	<a href="#">PDB</a> <a href="#">MMDB</a> <a href="#">Entrez</a> <a href="#">PubMed</a> <a href="#">TransportDB</a>

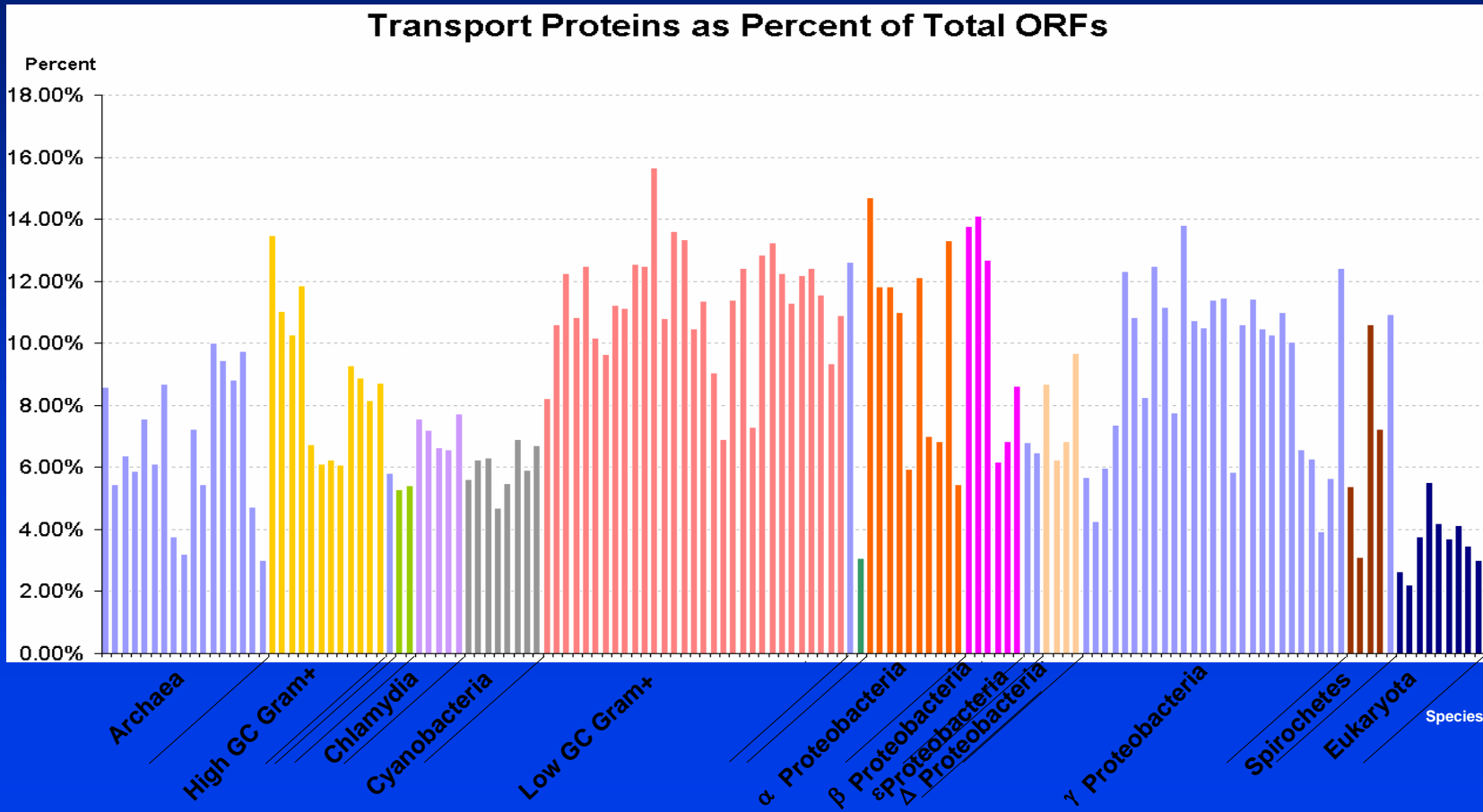
TIGR

THE INSTITUTE FOR GENOMIC RESEARCH

# Number of Transport Proteins



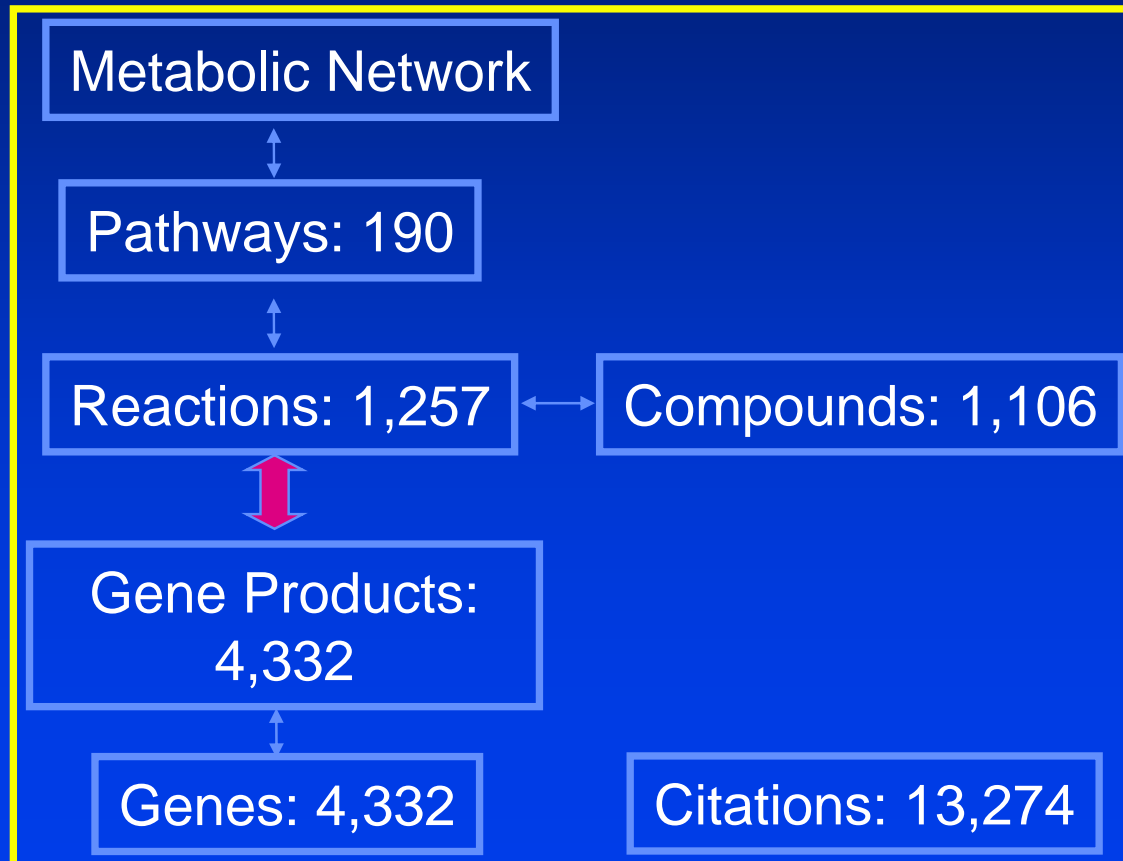
# Transporters vs Genome Size





# EcoCyc Project Overview

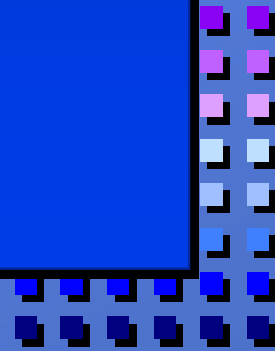
- *E. coli* Encyclopedia
- *E. coli* functional-bioinformatics database





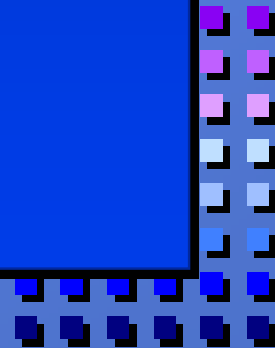
# Inclusion and annotation of membrane transporters in EcoCyc

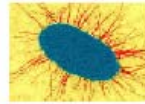
- Starting point- the set of known and putative *E. coli* membrane transporters from our transporter database
- Modified the schema for enzyme reactions in EcoCyc to facilitate representation of membrane transport reactions
- Annotated the set of known and predicted cytoplasmic membrane transporters
  - ◆ Bioinformatic predictions
  - ◆ Medline searches
  - ◆ Other leads from original journal articles
  - ◆ Other online resources, e.g., SwissProt



# Information captured on membrane transporters

- Substrate specificity and affinity
- Mechanism of transport
  - ◆ ATP-dependent, secondary transport, channels, PTS, unknown
- Experimental evidence
  - ◆ Gene knockouts, purified proteins, whole cell or vesicle transport assays, growth phenotypes
- Protein family
- Physiological Role
- Protein structure





## *E. coli* K-12 Transporter: XylIE xylose MFS transporter

Protein Sequence

Synonyms: B4031 , XylIE

### Comment:

XylIE is a D-xylose/proton symporter, one of two systems in *E. coli* responsible for the uptake of D-xylose. The other being the ATP-dependent ABC transporter XylFGH. The cloned *xylE* gene has been shown to complement *xylE* mutants *in vivo* [ [Davis87](#) ]. XylIE-mediated transport in whole cells is inhibited by protonophores and elicits an alkaline pH change [ [Lam80](#) ]. Experiments using *xylE* and *xylF* mutants have established that XylIE has a  $K_m$  of 63-169  $\mu\text{M}$  for D-xylose [ [Sumiya95](#) ]. XylIE is a member of the major facilitator superfamily (MFS) of transporters [ [Griffith92](#) ] and appears to function as a xylose/proton symporter. The *xylE* gene probably constitutes a monocistronic operon whose expression is inducible by D-xylose. Imported xylose is catabolised to xylulose-5-phosphate by the action of the XylA and XylB enzymes.

Gene: [xylIE](#)

Locations: inner membrane

Sequence Length: 491 AAs

Molecular Weight of Polypeptide (from nucleotide sequence): 53.608 kD

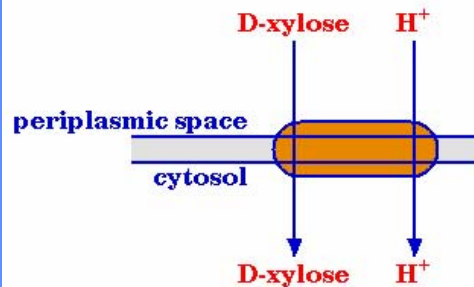
Unification Links: [MODBASE:P0AGF4](#) , [RefSeq:NP\\_418455](#) , [Swiss-Model:P0AGF4](#) , [UniProtKB:P0AGF4](#)

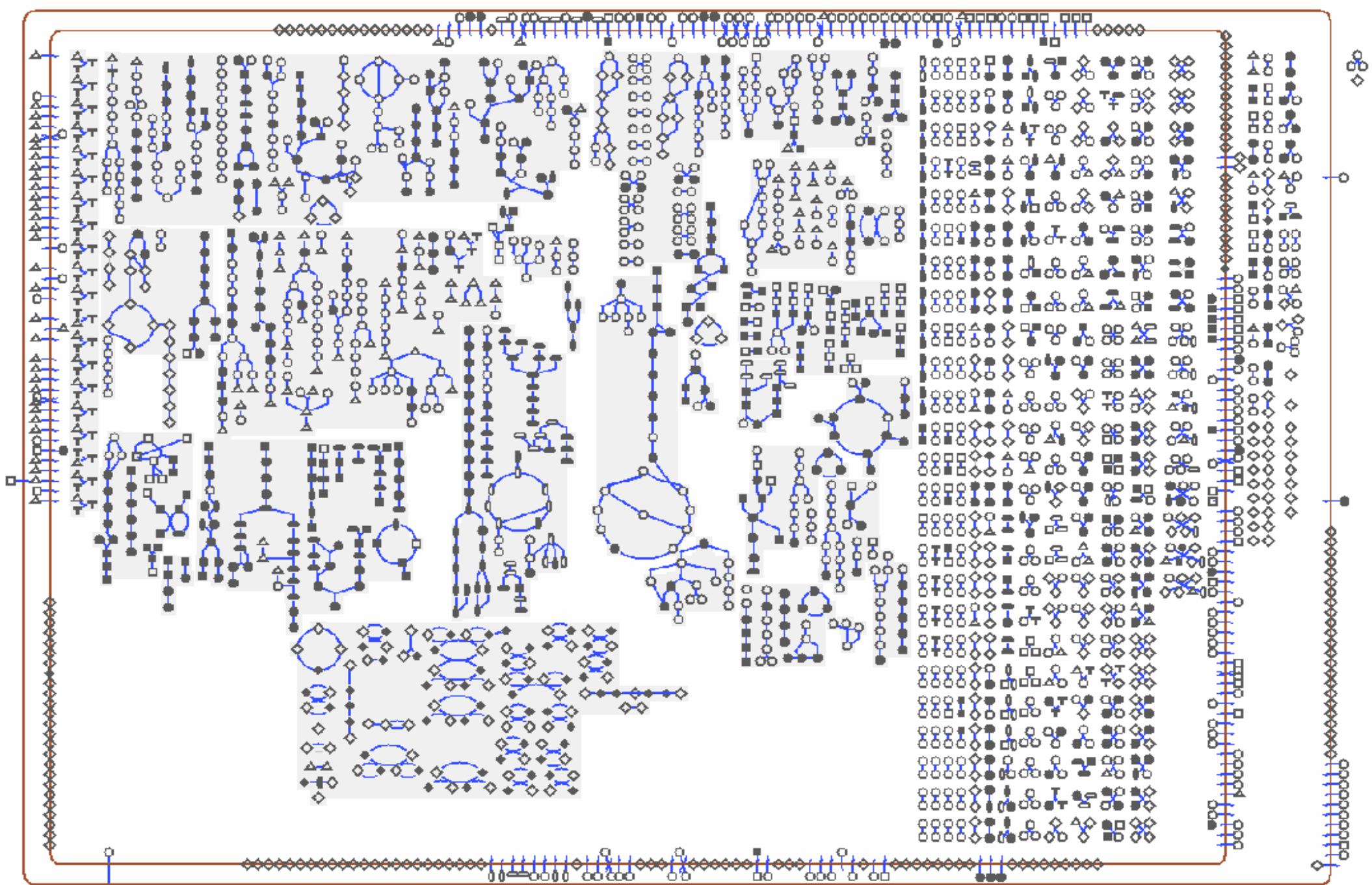
Relationship Links: [Pfam:IN-FAMILY:PF00083](#)

Gene-Reaction Schematic: [?](#)



### Enzymatic reaction of: XylIE xylose MFS transporter



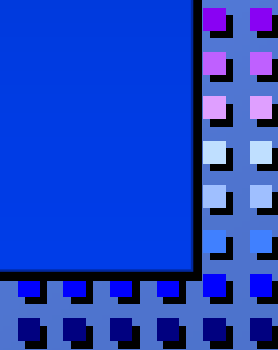
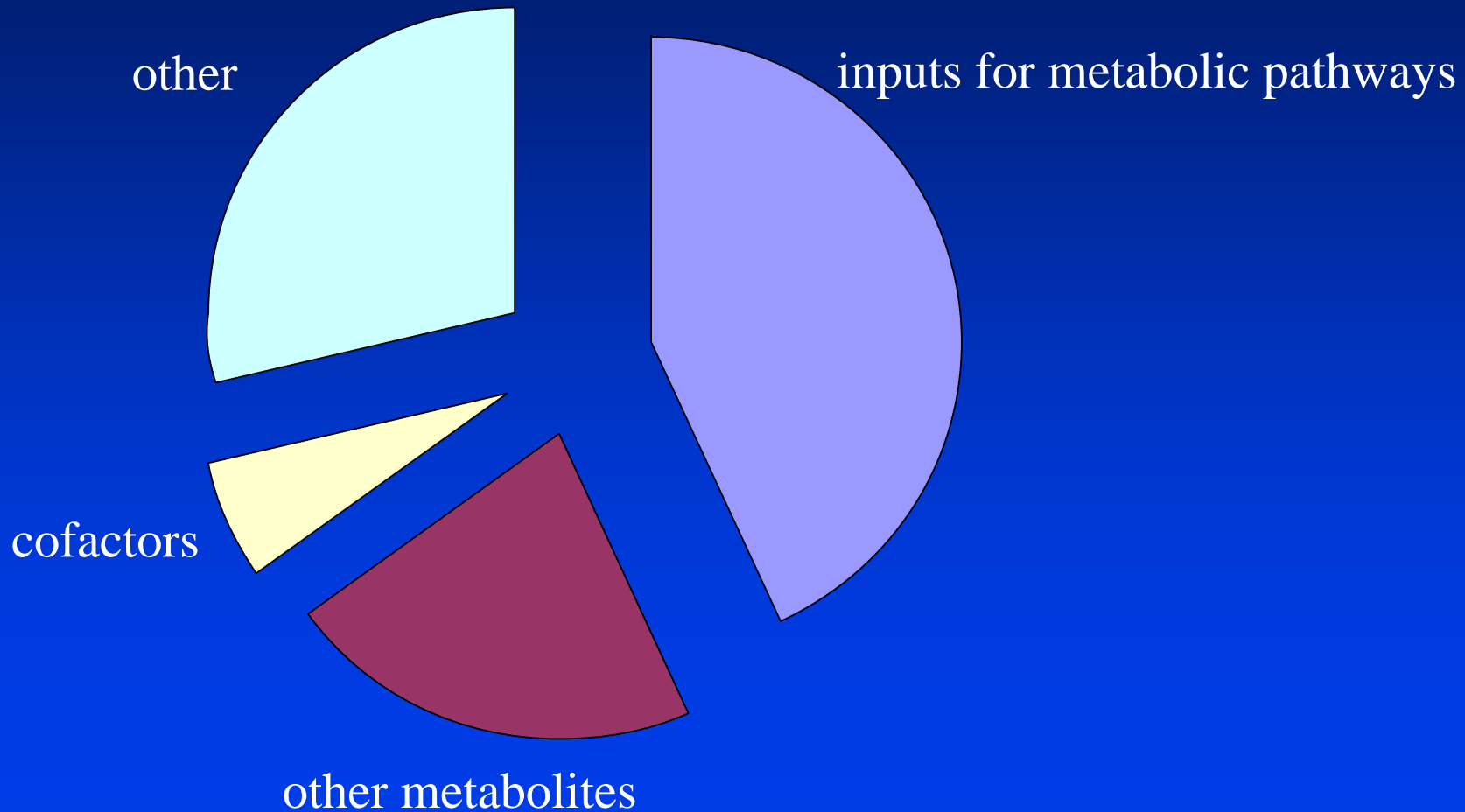


# Ecocyc- Transport statistics

- 310 cytoplasmic membrane transport systems
  - 681 genes classified as transporter genes
  - 217 transporter reactions defined
  - 146 transported compounds
- 
- 60% of transport genes have experimental evidence for their function
  - 17% transporters are co-encoded in an operon with an enzyme with the same substrate



# Breakdown of known transported substrates



# Overview of Paulsen group curation efforts

- Originally- incorporation of membrane transport systems into EcoCyc
- expanded our annotation efforts focused on membrane biology
  - ◆ protein secretion
  - ◆ flagella, pilus, fimbriae biogenesis
  - ◆ lipoproteins
  - ◆ outer membrane proteins
  - ◆ lipopolysaccharide biosynthesis
  - ◆ other membrane proteins
- Other areas, e.g., chaperones, and DNA repair
- Curation of non-K12 *E. coli* – CFT073 and EDL933 (O157)





# Transporter predictions using Pathway Tools

- Collaborated with Tom Lee and Peter Karp at SRI on including transporter predictions in Pathway Tools
- Used an analogous approach as they used for metabolic systems
- Iterated a series of predictions for several genomes- compared with TransportDB
- Issues in unambiguously assigning:
  - ◆ Substrate specificity
  - ◆ Transport direction
  - ◆ Energy coupling mechanisms
  - ◆ Identifying parts of multi-component systems, eg., ABC, PTS transporters

# Pathway Tools Prediction of *Z. mobilis* transporters

## *Z. mobilis* Reactions Class: Secondary Transporters

Comment:

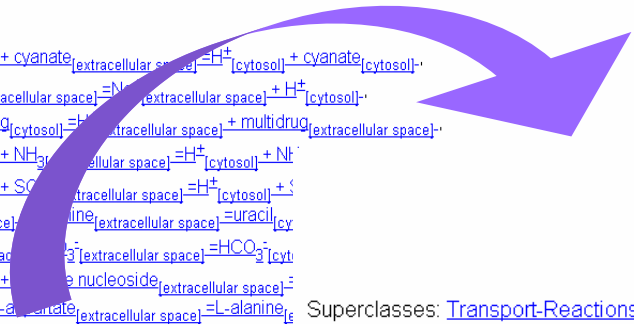
Secondary Transporters use chemiosmotic energy in the form of a proton, sodium or other ion or solute gradient to drive uptake or export of substrates. Secondary transporters can catalyze uniport (a single species is transported by facilitated diffusion), antiport (two or more species are transported in opposite directions) and/or symport (two or more species are transported together in the same direction).

Parent Classes:

[Transport-Reactions](#)

Instances:

$H^+_{[extracellular\ space]} + cyanate_{[extracellular\ space]} = H^+_{[cytosol]} + cyanate_{[cytosol]}$   
 $Na^+_{[cytosol]} + H^+_{[extracellular\ space]} = Na^+_{[extracellular\ space]} + H^+_{[cytosol]}$   
 $H^+_{[cytosol]} + multidrug_{[cytosol]} = H^+_{[extracellular\ space]} + multidrug_{[extracellular\ space]}$   
 $H^+_{[extracellular\ space]} + NH_3_{[extracellular\ space]} = H^+_{[cytosol]} + NH^+_{[cytosol]}$   
 $H^+_{[extracellular\ space]} + SO_4^{2-}_{[extracellular\ space]} = H^+_{[cytosol]} + S^{2-}_{[cytosol]}$   
 $uracil_{[extracellular\ space]} + uridine_{[extracellular\ space]} = uracil_{[cytosol]} + uridine_{[cytosol]}$   
 $HCO_3^-_{[extracellular\ space]} + H^+_{[extracellular\ space]} = HCO_3^-_{[cytosol]} + H^+_{[cytosol]}$   
 $H^+_{[extracellular\ space]} + nucleoside_{[extracellular\ space]} = H^+_{[cytosol]} + nucleoside_{[cytosol]}$   
 $L-alanine_{[cytosol]} + L-alanine_{[extracellular\ space]} = L-alanine_{[extracellular\ space]} + L-alanine_{[cytosol]}$   
 $H^+_{[extracellular\ space]} + gluconate_{[extracellular\ space]} = H^+_{[cytosol]} + gluconate_{[cytosol]}$   
 $H^+_{[extracellular\ space]} + Fe^{3+}_{[extracellular\ space]} = H^+_{[cytosol]} + Fe^{3+}_{[cytosol]}$



## *Z. mobilis* Reaction: Transport of gluconate

Cross-Species Comparison

Superclasses: [Transport-Reactions](#) -> [Secondary Transporters](#)

H<sup>+</sup>/gluconate symporter : [gntP](#)

### *Z. mobilis* Reaction

Comment:

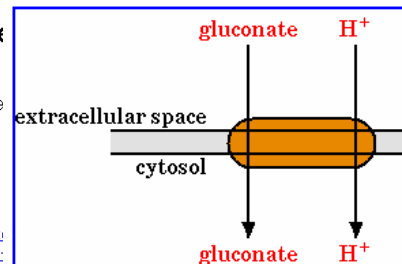
ATP-driven transporters use the hydrolysis of ATP to provide energy.

Parent Classes:

[Transport-Reactions](#)

Instances:

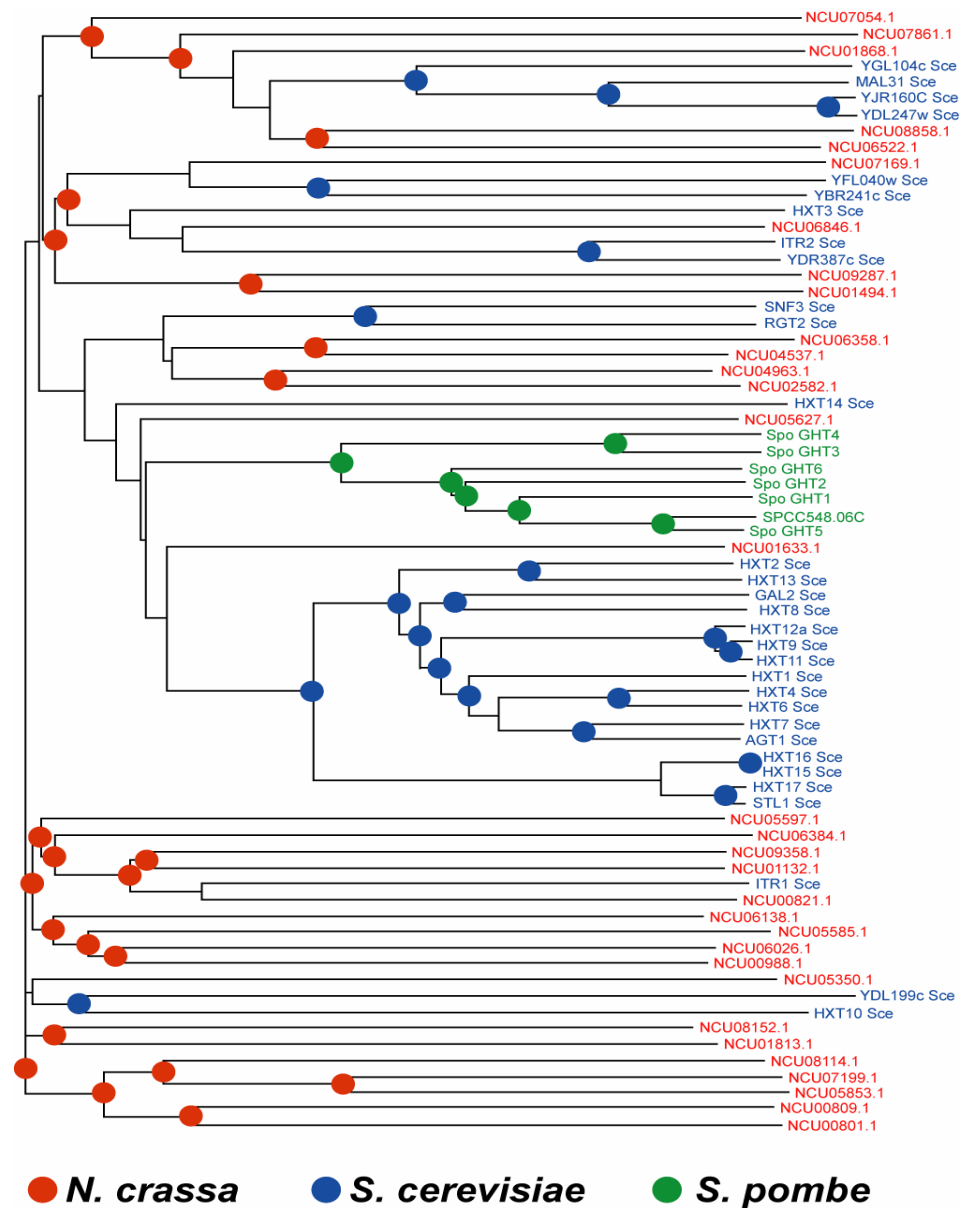
$H_2O + ATP + Fe^{3+}_{[extracellular\ space]} = phosphate + ADP + Fe^{3+}_{[cytosol]}$   
 $H_2O + ATP + phosphate_{[extracellular\ space]} = phosphate_{[cytosol]} + ADP + H_2O$   
 $H_2O + ATP + protohematin_{[cytosol]} = phosphate + ADP + protohematin_{[cytosol]}$   
 $H_2O + ATP + NO_3^-_{[extracellular\ space]} = phosphate + ADP + NO_3^-_{[cytosol]}$   
 $H_2O + ATP + HCO_3^-_{[extracellular\ space]} = phosphate + ADP + H_2CO_3$   
 $H_2O + ATP + multidrug_{[extracellular\ space]} = phosphate + ADP + multidrug_{[cytosol]}$   
 $H_2O + ATP + Fe^{3+}_{[extracellular\ space]} = phosphate + ADP + Fe^{3+}_{[cytosol]}$



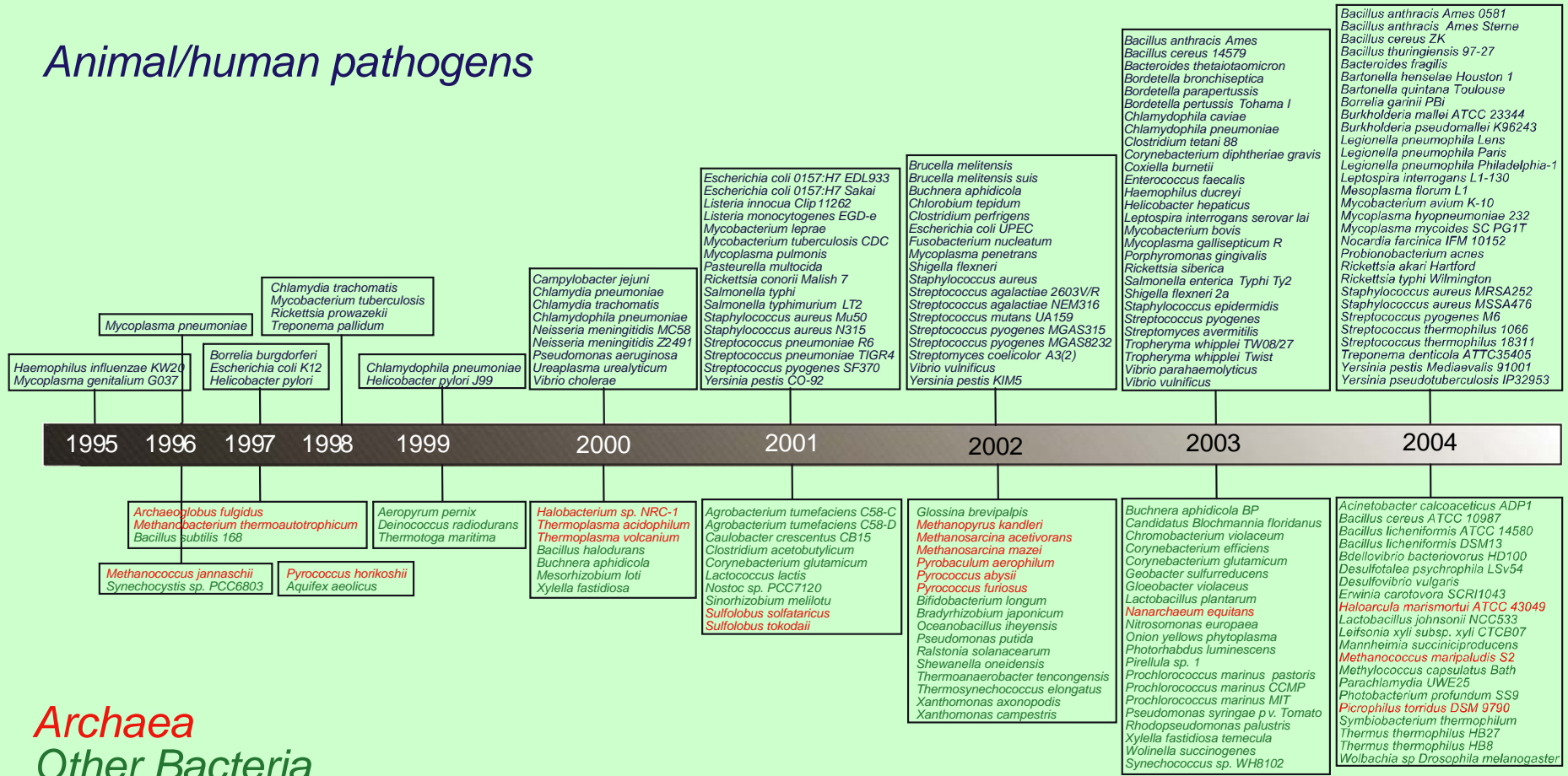
Gene-Reaction Schematic: [?](#)



# Fungal sugar transporters



# Animal/human pathogens

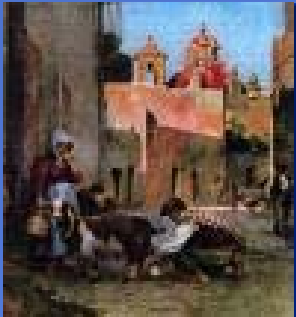


Archaea  
Other Bacteria

# Genome Sequencing projects

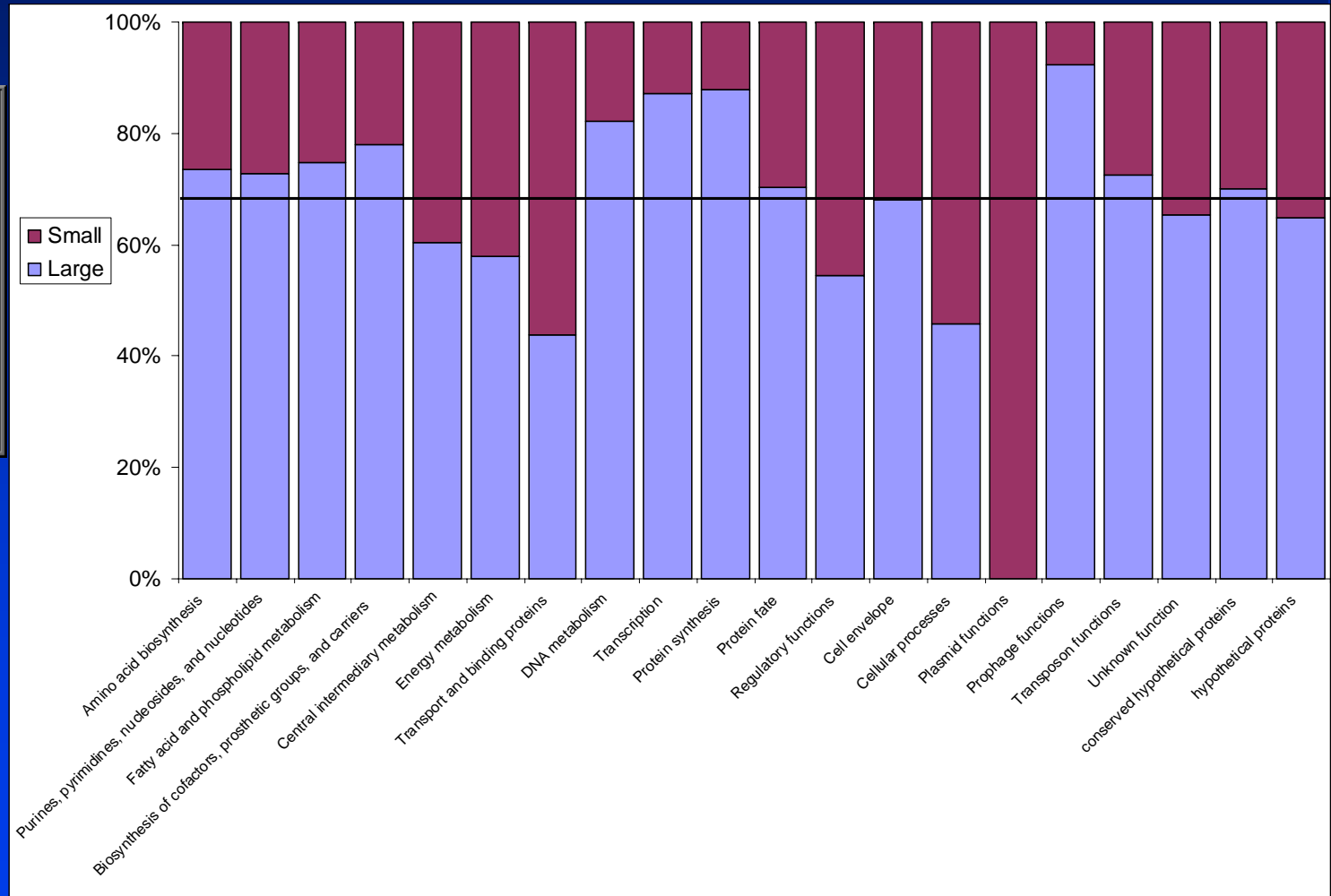
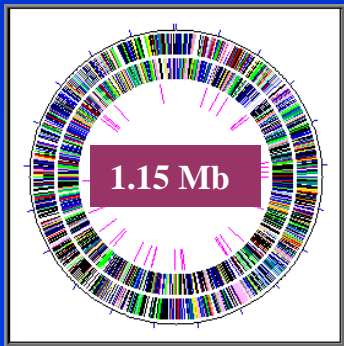
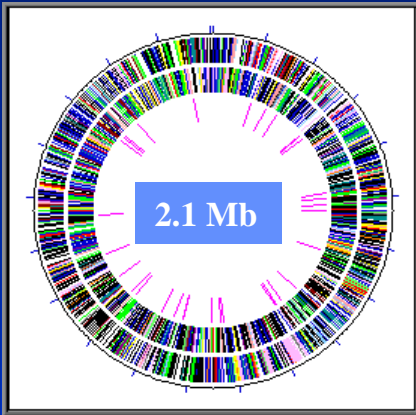
- ◆ *Brucella suis*- human/animal pathogen- 3.3 Mb Paulsen et al. PNAS (2002)
- ◆ *Enterococcus faecalis*- human pathogen- 3.0 Mb Paulsen et al. Science (2003)
- ◆ *Treponema denticola*- dental pathogen- 3.0 Mb Seshadri et al. PNAS (2004)
- ◆ *Pseudomonas fluorescens*- biological control- 6.5 Mb Paulsen et al. Nature Biotech (2005)
- ◆ *Clostridium perfringens*- gas gangrene- 3.3 Mb Myers et al. Genome Research (in press)
- ◆ *Clostridium perfringens*- food poisoning- 3.0 Mb Myers et al. Genome Research (in press)
- ◆ *Synechococcus* sp- coastal isolate- 2.6 Mb Palenik et al. PNAS (in press)
- ◆ *Toxoplasma gondii*- 3 strains- toxoplasmosis- 63 Mb Paulsen et al. Science (in prep)
- ◆ *Brucella ovis*- animal pathogen- 3.3 Mb CLOSED and ANNOTATED
- ◆ *Dichelobacter nodosus*- sheep footrot- 1.6 Mb CLOSED and ANNOTATED
- ◆ *Pseudomonas aeruginosa*- opportunistic pathogen- 6.7 Mb CLOSURE

# *Brucella suis*

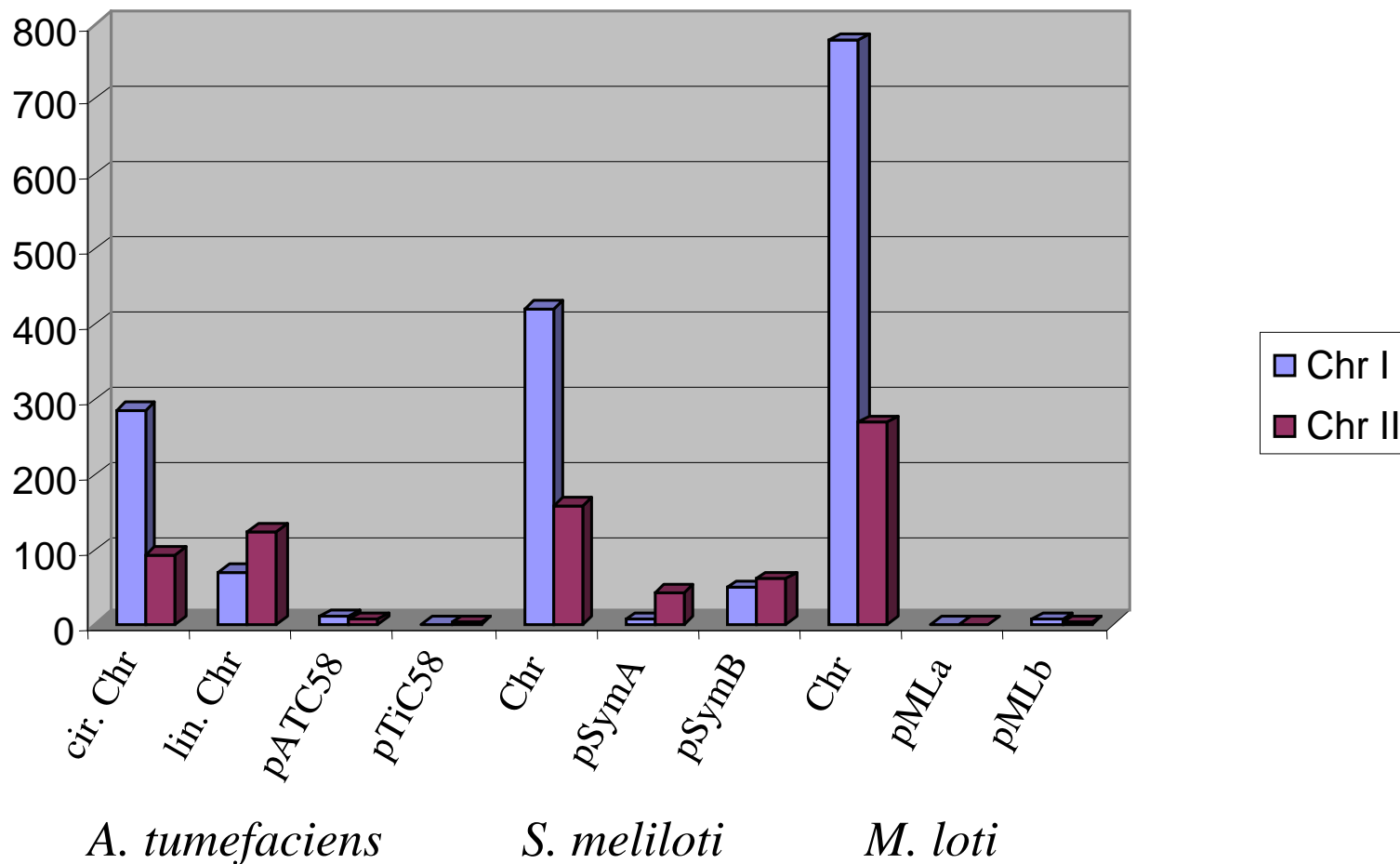


- *Brucella* spp- causative agents of brucellosis, a zoonotic disease of humans and animals
- Infective via inhalation, contact with diseased animals or consumption of contaminated foods
- Potential biowarfare agent- *B. suis* - first agent weaponized by the US military
- *Brucella melitensis*, *B. suis*, and *B. abortus* are the species most frequently associated with pathogenicity in humans
- Intracellular pathogen that invades macrophages

# Chromosome- role category breakdown



# *Brucella suis*- best hits to rhizobial replicons





# Generation of suisCyc

Pathway Tools version 10.0

File Overview Pathway Reaction Protein RNA Gene Compound Chromosome Tools Help

Brucella suis 1330 Home Back Forward History Next Answer Clone Save DB

**Brucella suis**  
Strain: 1330 version 1.0  
[Summarize Pathway Evidence](#)

**Authors:**  
Aaron Johnson, TIGR

<u>Replicon</u>	<u>Total Genes</u>	<u>Protein Genes</u>	<u>RNA Genes</u>	<u>Pseudogenes</u>	<u>Size (bp)</u>
Chromosome 1	2123	2123	0	0	2,107,794
Chromosome 2	1150	1150	0	0	1,207,381
Total:	3273	3273	0	0	3,315,175

**Pathways:** 242  
**Enzymatic Reactions:** 1030  
**Transport Reactions:** 3

**Polypeptides:** 3286  
**Protein Complexes:** 4  
**Enzymes:** 601  
**Transporters:** 13

**Compounds:** 888

**Transcription Units:** 0  
**tRNAs:** 0

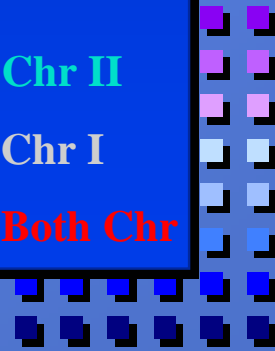
# *Brucella suis* Carbon and Nitrogen Sources

- Erythritol
  - Formate
  - Fructose
  - Galactonate
  - Galactose
  - **Glucose**
  - **Glycerol**
  - Glycolate
  - Maltose/Trehalose
  - Malate
  - Mannose
  - Mannitol
  - Ribitol
  - Ribose
  - Sorbose
  - Xylose
  - Xylulose
  - Acetoin
  - Alanine
  - Arginine
  - Asparagine
  - Aspartate
  - Glutamate
  - Glycine
  - Histidine
  - Isoleucine/Leucine/Valine
  - Proline
  - Serine
  - Threonine
  - Sarcosine
  - N-acetyl-glucosamine
  - D-glucoseamine
  - Choline
  - Ornithine
  - Putrescine
  - Spermidine
- Chr II  
■ Chr I  
■ Both Chr

# “Unexpected metabolism”

- **Protocatechuate pathway**
- **Beta-ketoadipate pathway**
- **Hydroxybenzoate and benzoate metabolism**
- **4 monooxygenases of unclear specificity**
- **Exopolysaccharide biosynthesis**
- **Nitrous oxide, nitric oxide, nitrate and nitrite reduction**

■ Chr II  
■ Chr I  
■ Both Chr



# *Brucella suis*- transport

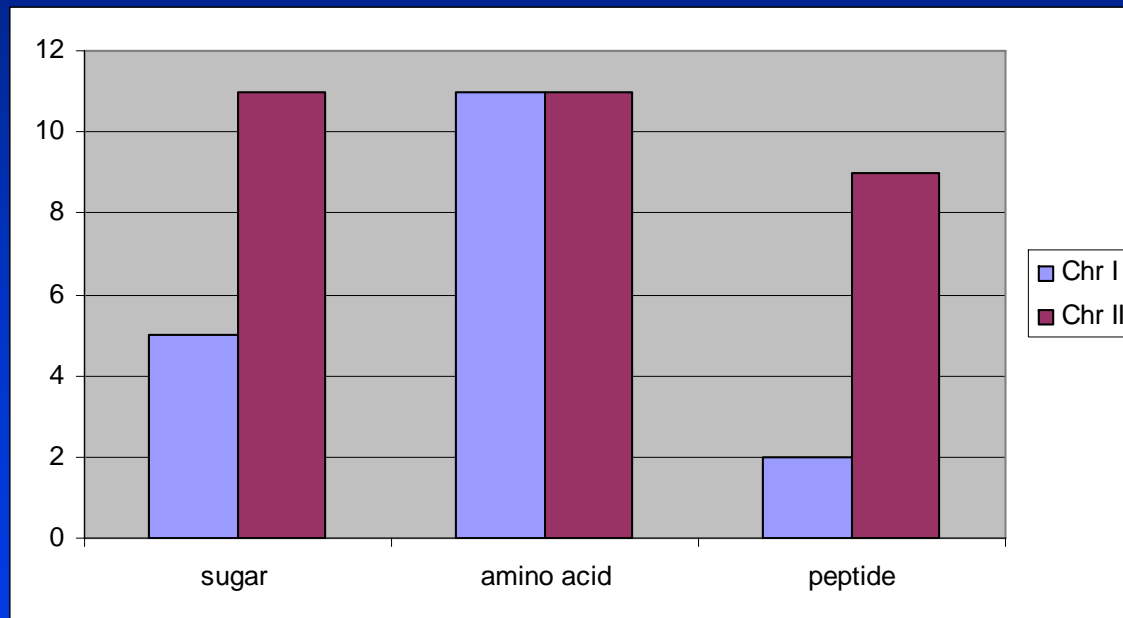
## ■ Emphasis on ABC –type transporters (~90 systems)

sugar transporters (16)

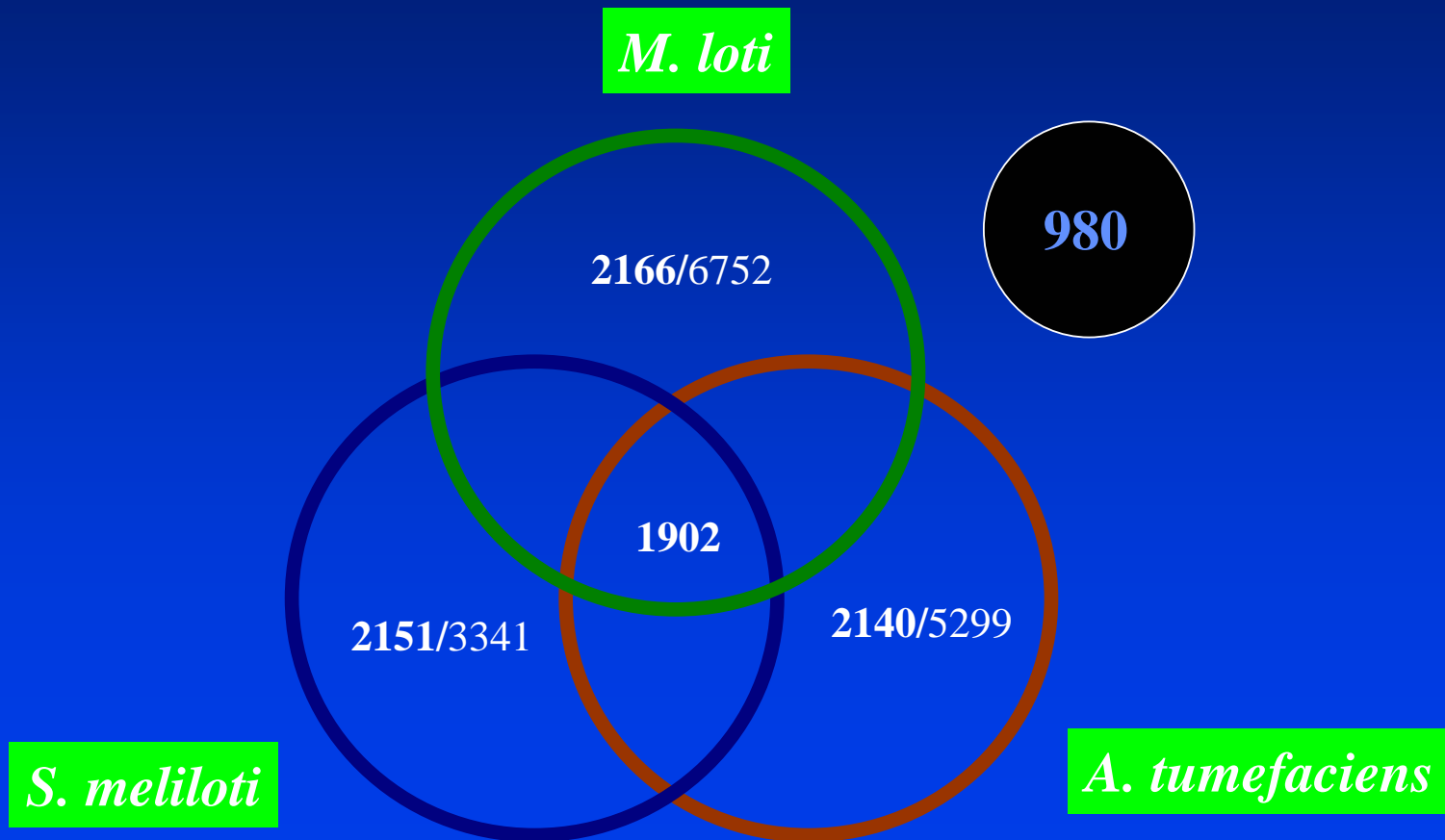
amino acid transporters (22)

peptide transporters (11)

iron chelate uptake (5)

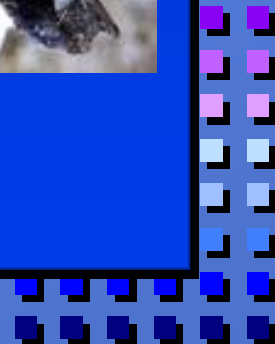


# Gene conservation with other $\alpha$ -proteobacteria



# *Dichelobacter nodosus*: sheep footrot genome

- Causative agent of footrot in ruminants, especially sheep
- Severe economic losses to the wool and meat industries
  - ◆ lameness, loss of body condition and poor wool growth
- Gram-negative anaerobe- small genome (1.38 Mb)
- Type IV fimbriae and extracellular proteases- key virulence determinants



# Construction of nodoCyc

Pathway Tools version 10.0

File Overview Pathway Reaction Protein RNA Gene Compound Chromosome Tools Help

Dichelobacter nodosus Home Back Forward History Next Answer Clone Save DB

### Dichelobacter nodosus

Strain: VCS1703A version 1.0

Summarize Pathway Evidence

**Authors:**  
Aaron Johnson, TIGR

<u>Replicon</u>	<u>Total Genes</u>	<u>Protein Genes</u>	<u>RNA Genes</u>	<u>Pseudogenes</u>	<u>Size (bp)</u>
Chromosome	1303	1303	0	0	1,389,350

**Pathways:** 148  
**Enzymatic Reactions:** 698  
**Transport Reactions:** 2

**Polypeptides:** 1317  
**Protein Complexes:** 0  
**Enzymes:** 350  
**Transporters:** 9

**Compounds:** 639

**Transcription Units:** 643  
**tRNAs:** 0

Command: ''

# Unexpected pathways present in *D. nodosus*

## ■ Sugar utilization

- ◆ glycolysis, pentose phosphate pathway, fructose PTS, glycerol transporter and kinase

## ■ Electron transport

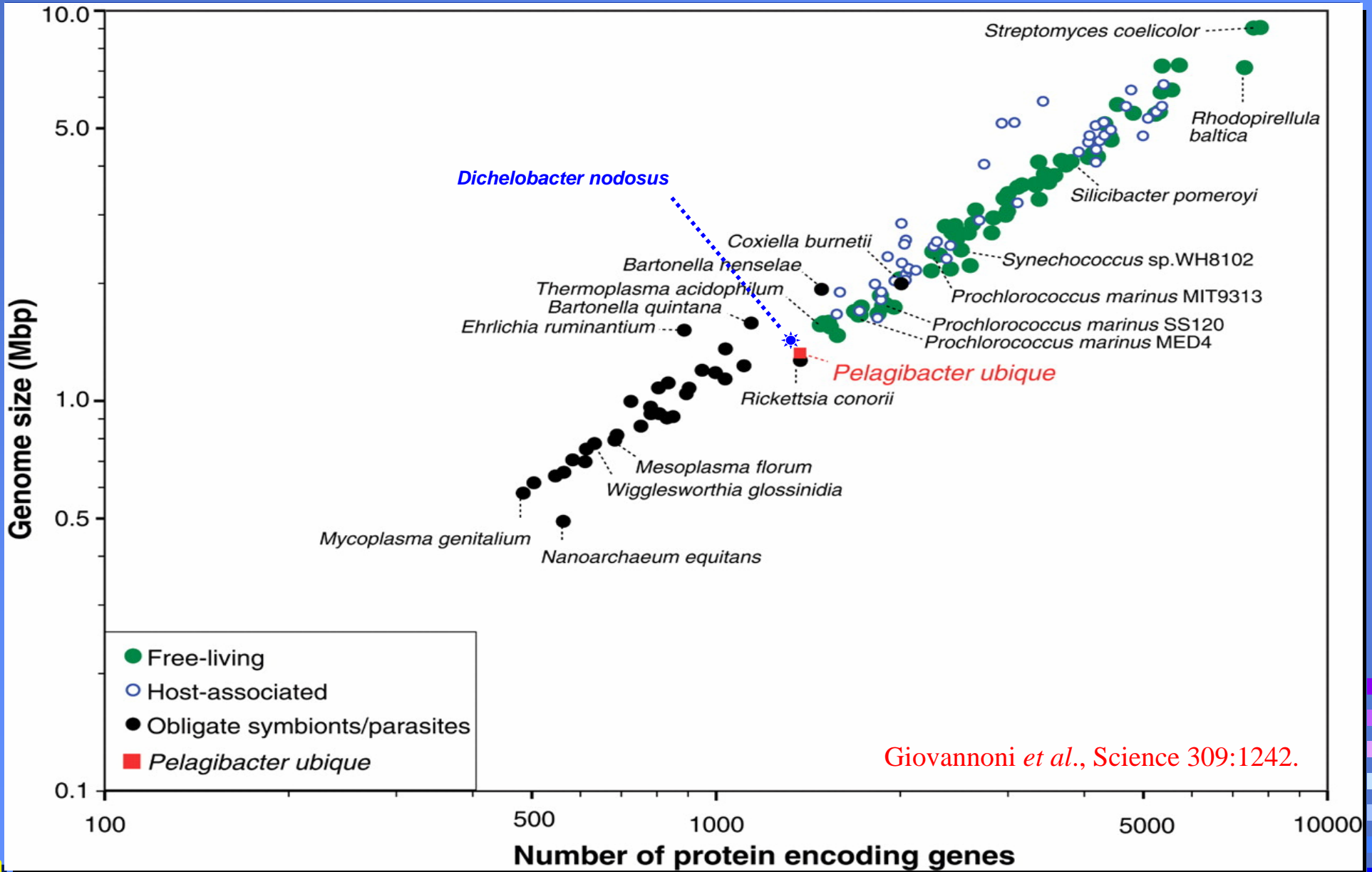
- ◆ NADH ubiquinone oxidoreductase, ubiquinone biosynthesis, fumarate reductase

## ■ Aerobic metabolism/oxidative stress resistance

- ◆ superoxide dismutase (Mn), alkyl hydroperoxide reductase, methionine sulfoxide reductase
- ◆ class I (aerobic) ribonucleotide reductase, cytochrome (quinone) oxidase, bd type.  
(aerobic) ubiquinone biosynthesis







# Pathway conservation in small genome bacteria

## Cross-Species Comparison: cobalamin biosynthesis I

Note: In addition to reflecting differences in biology among different organisms, these statistics and data will also reflect differences in the level of curation and data availability and completeness among different organisms.

This table compares a single pathway across the requested organisms. The Evidence Glyph is a cartoon of the pathway diagram showing which steps have identified enzymes, which are pathway holes, and which steps are unique to this pathway (a color key is included at the bottom of this page). For each reaction step, we list the identified enzymes and genes. A pathway may not be present in an organism even if enzymes have been identified for one or more of its reactions, and we indicate if this is the case. Finally, we include diagrams showing all the operons that contain the genes for this pathway. Genes that are actually involved in the pathway are colored dark purple, whereas other genes in the same operon which are not involved in the pathway are shaded a lighter color. Clicking on an organism name will take you to the display page for the pathway in that organism.

Organism	Evidence Glyph	Enzymes and Genes for cobalamin biosynthesis I	Operons
<a href="#">B. henselae Houston-1</a>		<p><a href="#">EC#2.5.1.-</a> None</p> <p><a href="#">EC#2.7.-.-</a> None</p> <p><a href="#">EC#2.7.7.62</a> None</p> <p><a href="#">EC#2.4.2.21</a> <a href="#">Nicotinate-nucleotide--dimethylbenzimidazole phosphoribosyltransferase : cobT1</a></p> <p><a href="#">EC#3.1.3.-</a> <a href="#">Phosphoglycerate mutase / alpha-ribazole-5'-P phosphatase / phosphoglycerate mutase 1 : gpmA</a></p> <p><a href="#">EC#2.-.-</a> None</p> <p><a href="#">EC#2.5.1.17</a> None</p>	<p><a href="#">cobT1</a></p> <p><a href="#">dapB</a> <a href="#">gpmA</a></p>
<a href="#">B. quintana Toulouse</a>		<p><b>This pathway is not marked as present in this organism.</b></p> <p>No Enzymes or Genes have been identified for this pathway</p>	
<a href="#">B. burgdorferi B31</a>		<p><b>This pathway is not marked as present in this organism.</b></p> <p>No Enzymes or Genes have been identified for this pathway</p>	
<a href="#">B. garinii PBI</a>		<p><b>This pathway is not marked as present in this organism.</b></p> <p>No Enzymes or Genes have been identified for this pathway</p>	
<a href="#">B. aphidicola (Baizongia pistaciae)</a>		<p><b>This pathway is not marked as present in this organism.</b></p> <p>No Enzymes or Genes have been identified for this pathway</p>	
<a href="#">B. sp. APS</a>		<p><b>This pathway is not marked as present in this organism.</b></p>	

# Pathway conservation in small genome bacteria

## Biosynthesis

fatty acid	-	-	+/-	+/-	+	+	+	+/-	+	+
heme	-	-	+	+	+/-	+	+	+/-	+/-	+
ubiquinone	-	-	+	+	+	+	+	+/-	+	+
NAD	-	-	-	-	-	+	+	+	+/-	+
glutathione	-	-	-	-	+	+	+	+	+	+
pantothenate	-	-	-	-	+	-	+	-	+	+
riboflavin	-	-	-	+	-	+	+	-	+	+
folate	-	-	+/-	-	+/-	+	+	+/-	+/-	+/-
	<i>N. equitans</i>	<i>M. genitalium</i>	<i>R. prowazekii</i>	<i>C. pneumoniae</i>	<i>W. pipientis</i>	<i>P. ubique</i>	<i>D. nodosus</i>	<i>E. rumantium</i>	<i>B. henselae</i>	<i>C. burnetii</i>

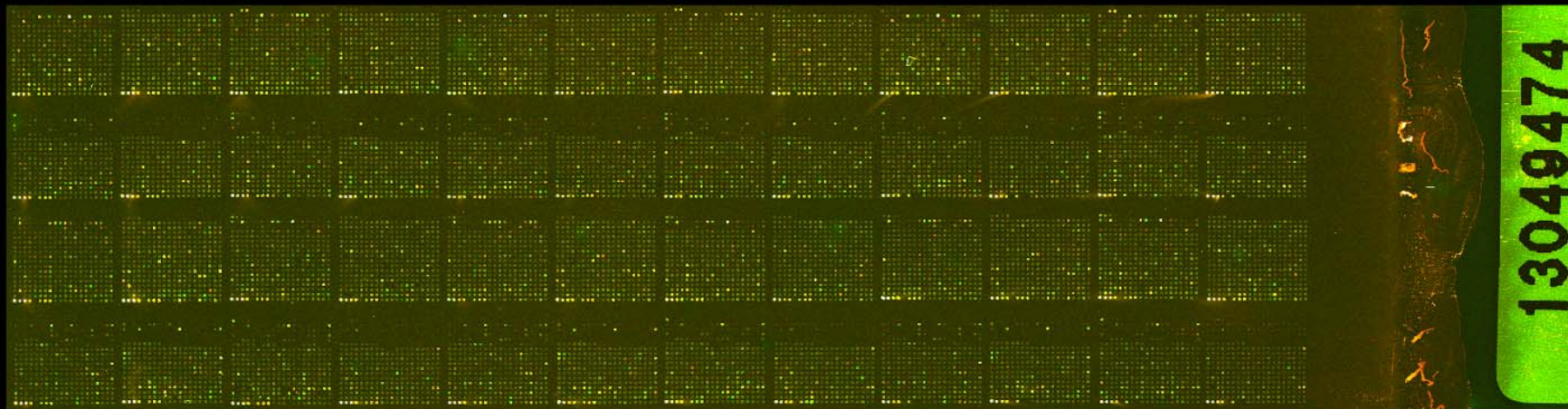
# Coastal and open ocean *Synechococcus* sp.

- Unicellular cyanobacteria constitute 20-40% of marine chlorophyll biomass and carbon fixation
- *Synechococcus* found both in oligotrophic ocean and in coastal waters
- Utilizing a genomic approach to investigate adaptations between coastal and open ocean *Synechococcus* isolates
- Coastal environments differ from oligotrophic ocean
  - Higher nutrient concentrations
  - More dynamic
  - Light attenuation from debris



# WH8102 Microarray expression analysis

- Utilizing a complete genome array of WH8102 to investigate
  - how cyanobacteria perceive and respond to different conditions in the marine environment
  - Develop an overall model of the regulatory and metabolic networks of WH8102

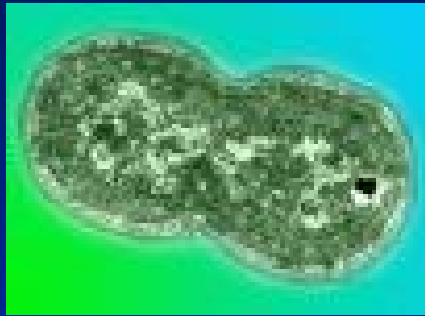


TIGR

THE INSTITUTE FOR GENOMIC RESEARCH



# Whole genome microarray studies of *Synechococcus*



- Major Nutrients
  - Nitrogen- Ammonia, Urea, Nitrate
  - Phosphorus
- Trace Nutrients
  - Nickel, Iron
- Co-cultivation with other marine bacteria
- Effects of Shock
  - Osmoregulation- Salt Shock
  - Toxic Compounds- Ethidium, Mitomycin C
- Mutants
  - Regulatory mutants- two component systems
  - Other mutants

# Construction of CyanoCyc

The screenshot shows the Pathway Tools version 10.0 interface. The title bar reads "Pathway Tools version 10.0". The menu bar includes "File", "Overview", "Pathway", "Reaction", "Protein", "RNA", "Gene", "Compound", "Chromosome", "Tools", and "Help". The breadcrumb trail shows "Synechococcus sp" with a dropdown arrow, followed by navigation buttons: "Home", "Back", "Forward", "History", "Next Answer", and "Clone". A "Save DB" button is in the top right corner.

The main content area displays the following information:

**Synechococcus sp**  
Strain: WH8102 version 1.0  
[Summarize Pathway Evidence](#)

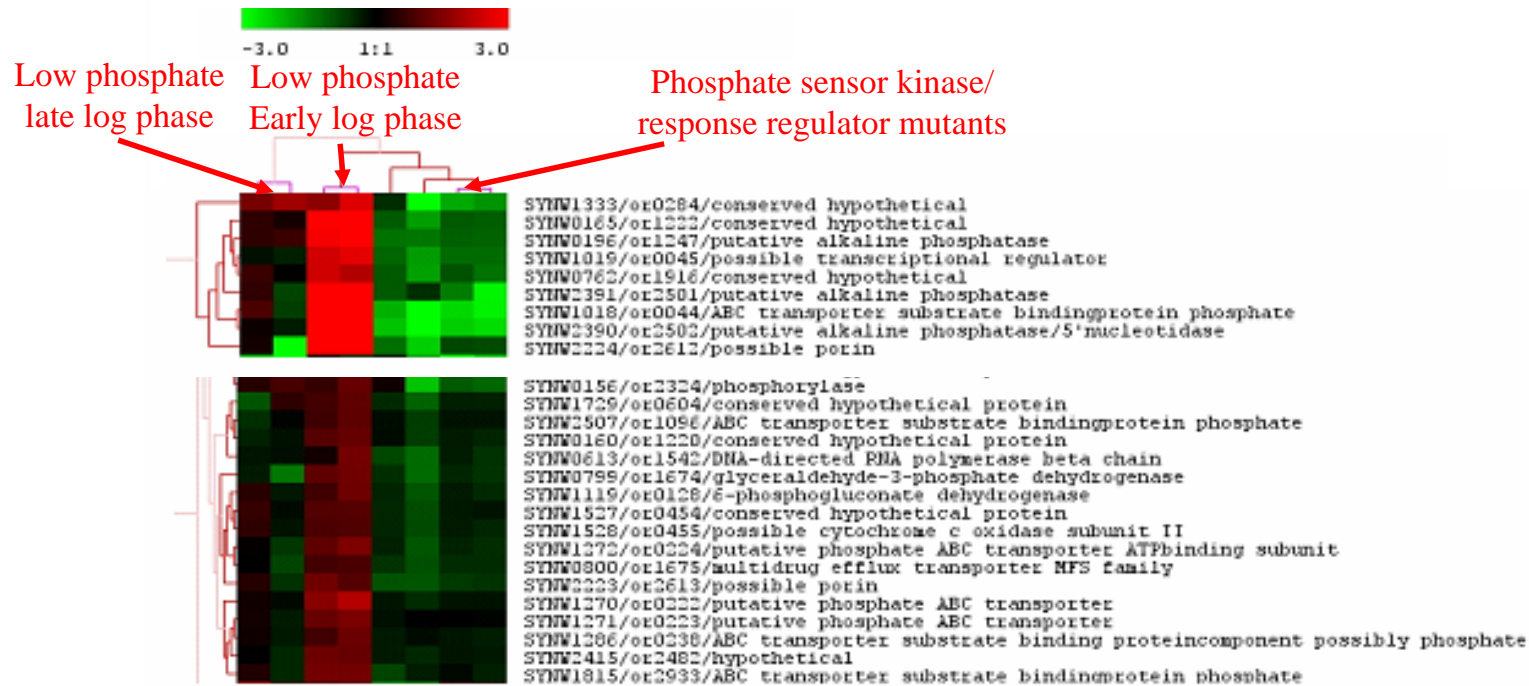
**Authors:**  
Aaron Johnson, TIGR

<u>Replicon</u>	<u>Total Genes</u>	<u>Protein Genes</u>	<u>RNA Genes</u>	<u>Pseudogenes</u>	<u>Size (bp)</u>
Chromosome	2581	2526	55	9	2,434,428

Summary statistics:

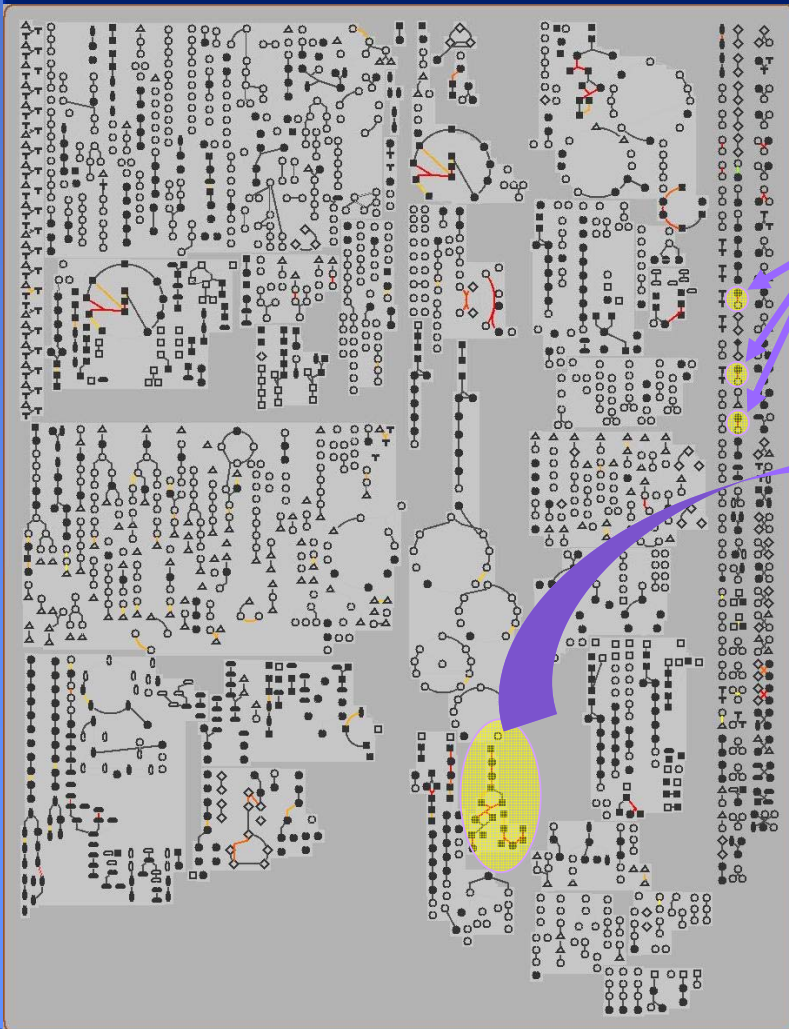
- Pathways:** 231
- Enzymatic Reactions:** 1028
- Transport Reactions:** 1
- Polypeptides:** 2527
- Protein Complexes:** 0
- Enzymes:** 556
- Transporters:** 9
- Compounds:** 888
- Transcription Units:** 0
- tRNAs:** 44

# Phosphate regulation in *Synechococcus*

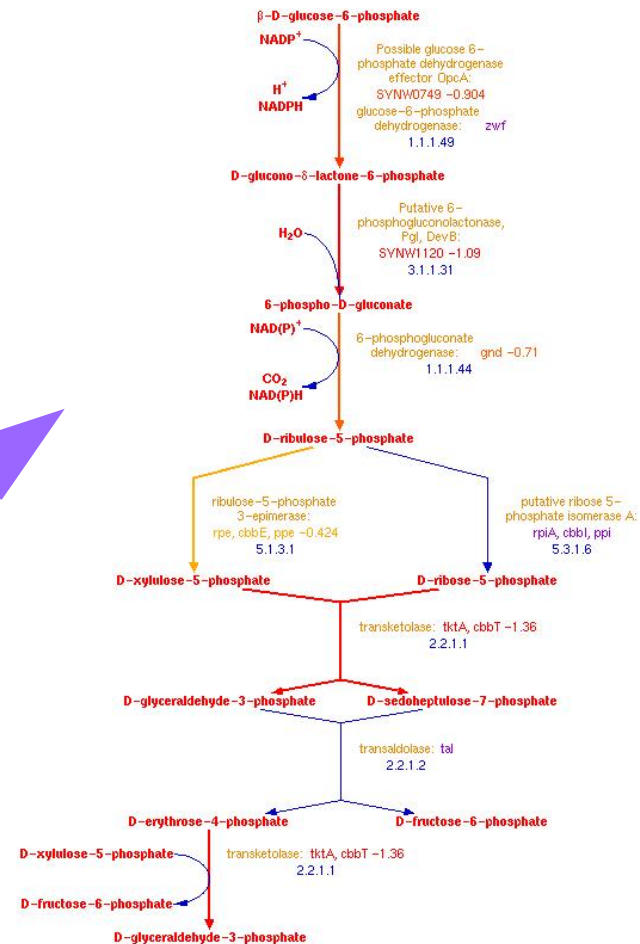




# Mapping array data onto the metabolic overview of *Synechococcus* WH8102



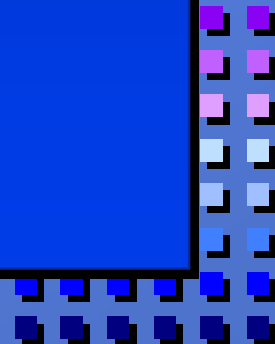
*S. sp* Pathway: superpathway of oxidative and non-oxidative branches of pentose phosphate pathway





# Functional Genomics of Microbial Membrane Transporters

- Bioinformatics-> predicted/putative functions of many many genes- how accurate?
- Detailed characterization of interesting/novel transporters
- Large scale analysis of *P.aeruginosa* transporter gene mutants



# *P. aeruginosa* transposon mutants

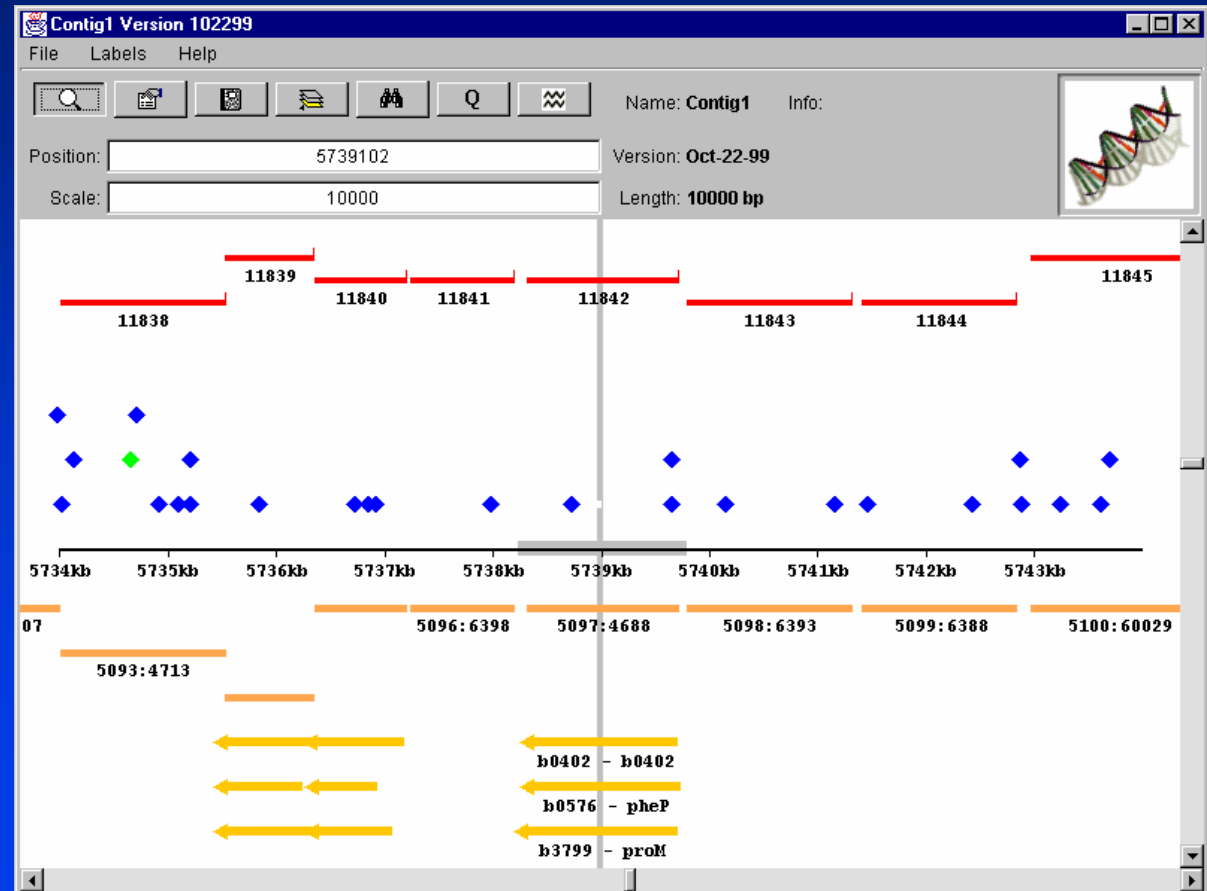
	<u>Total</u>	<u>% Total</u>
Number of predicted PAO1 Genes	5570	
Tn5 Insertions hitting PAO1 sequence	10149	
Hits in predicted coding regions	9199	91
Intergenic Tn5 Hits?	950	9
Unique Tn5 insertions	9870	97
Total genes hit by at least one Tn5	3581	64
<b>Total genes not yet hit by Tn5</b>	<b>1989</b>	<b>36</b>



# Transporter Mutant Collection

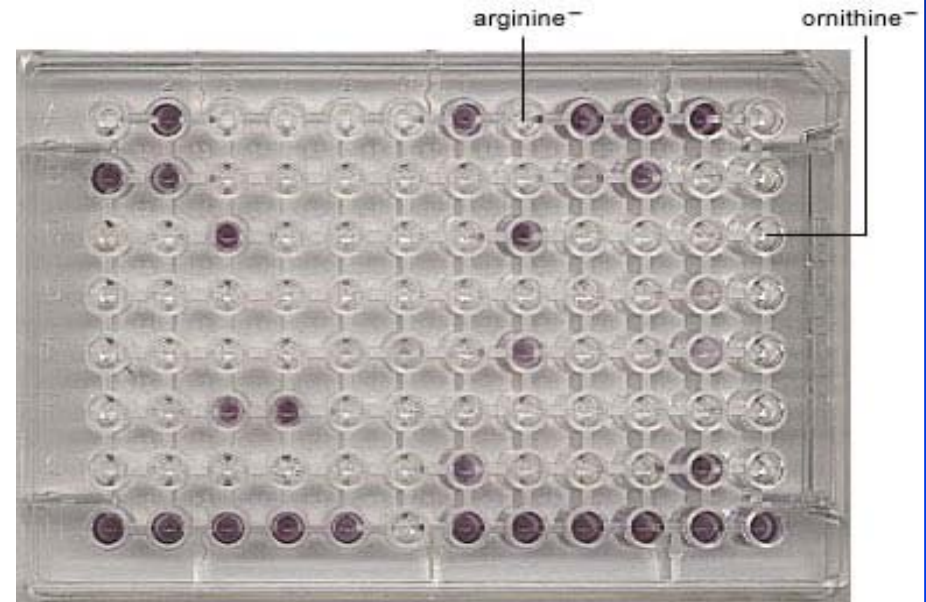
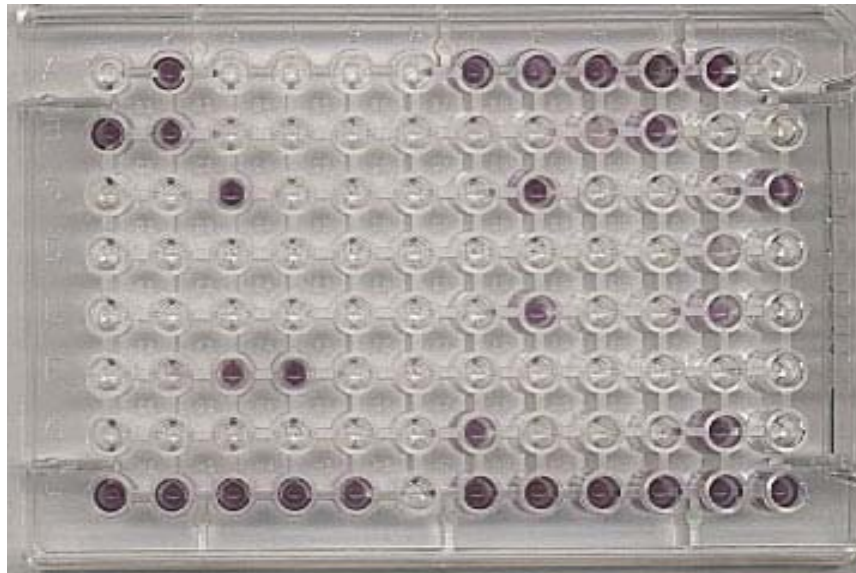
384 mutants in different transporter genes

85% transporter genes-non-essential?



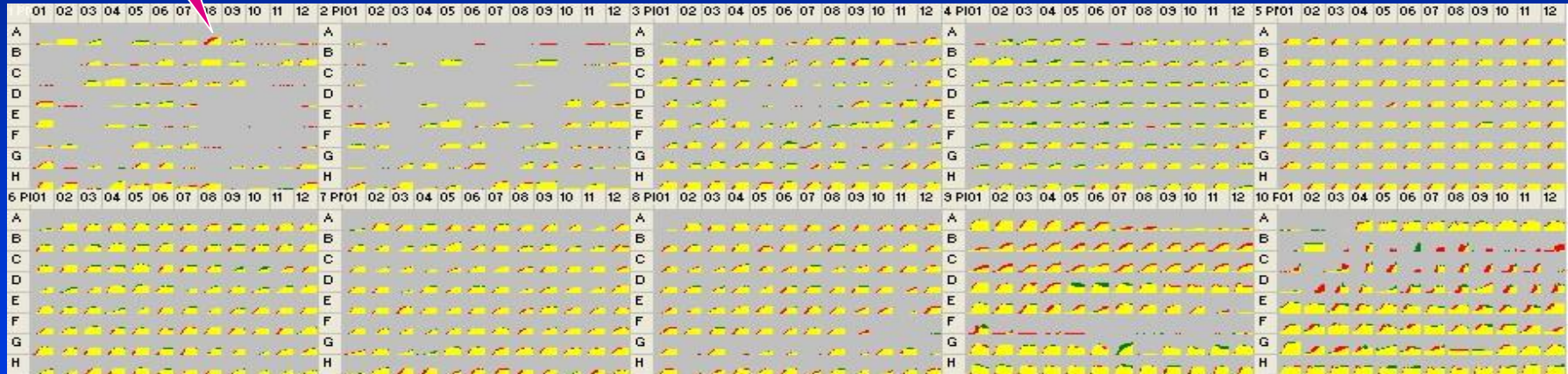
# Method I: BIOLOG MicroPlates

- BIOLOG MicroPlates test the ability of a microorganism to utilize or oxidize a preselected panel of different carbon sources.



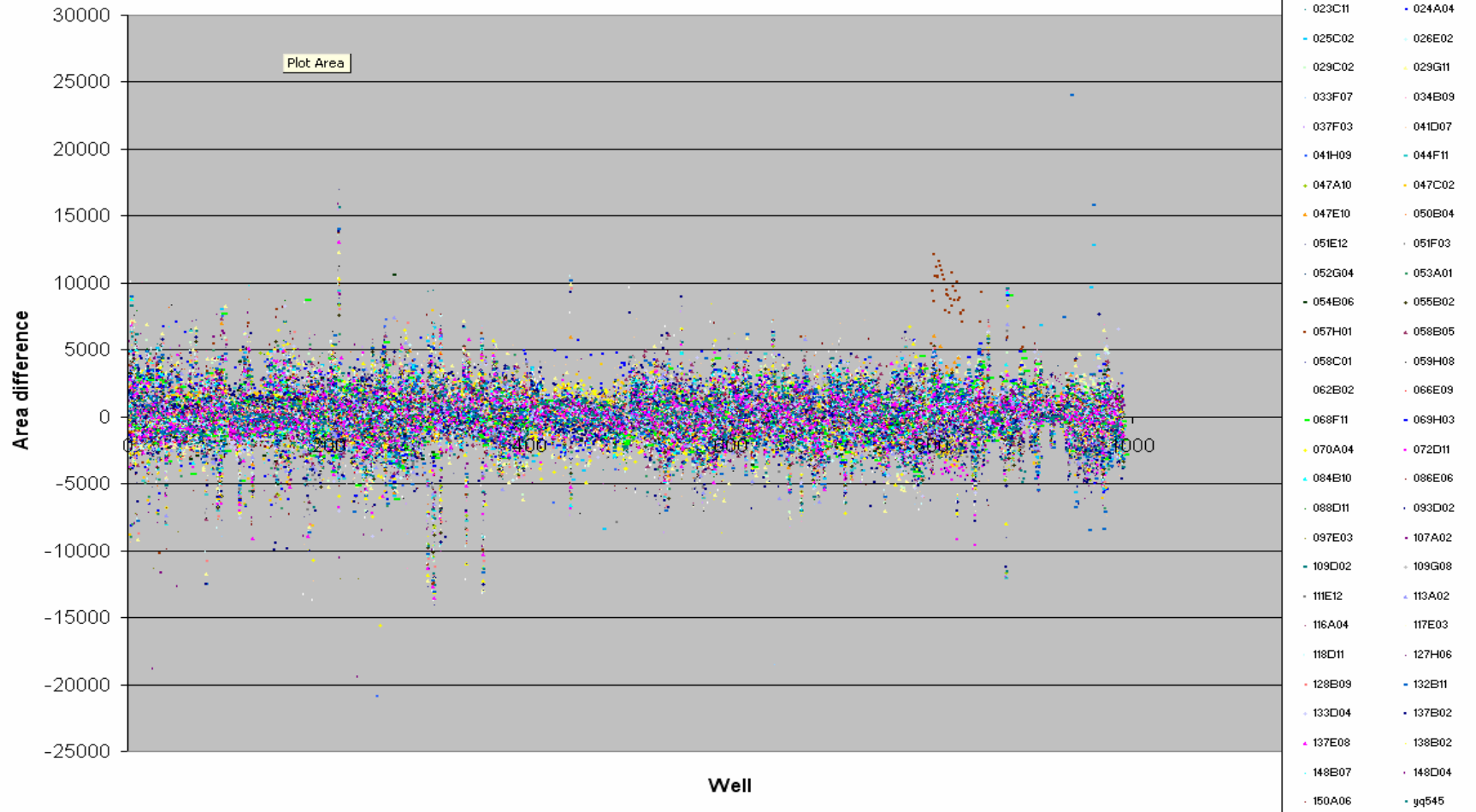
# 086E06: predicted proline/sodium symporter PM1-10 plates

Proline



# Global comparison of phenotype microarray data

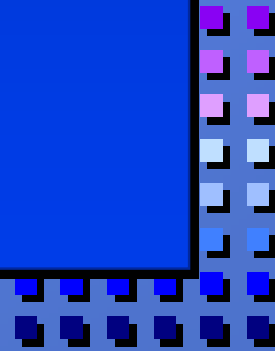
Plates 1-8 mean centered, 48 hr; Plates 9-10 24 hr





# Confirmation of Biolog-based data

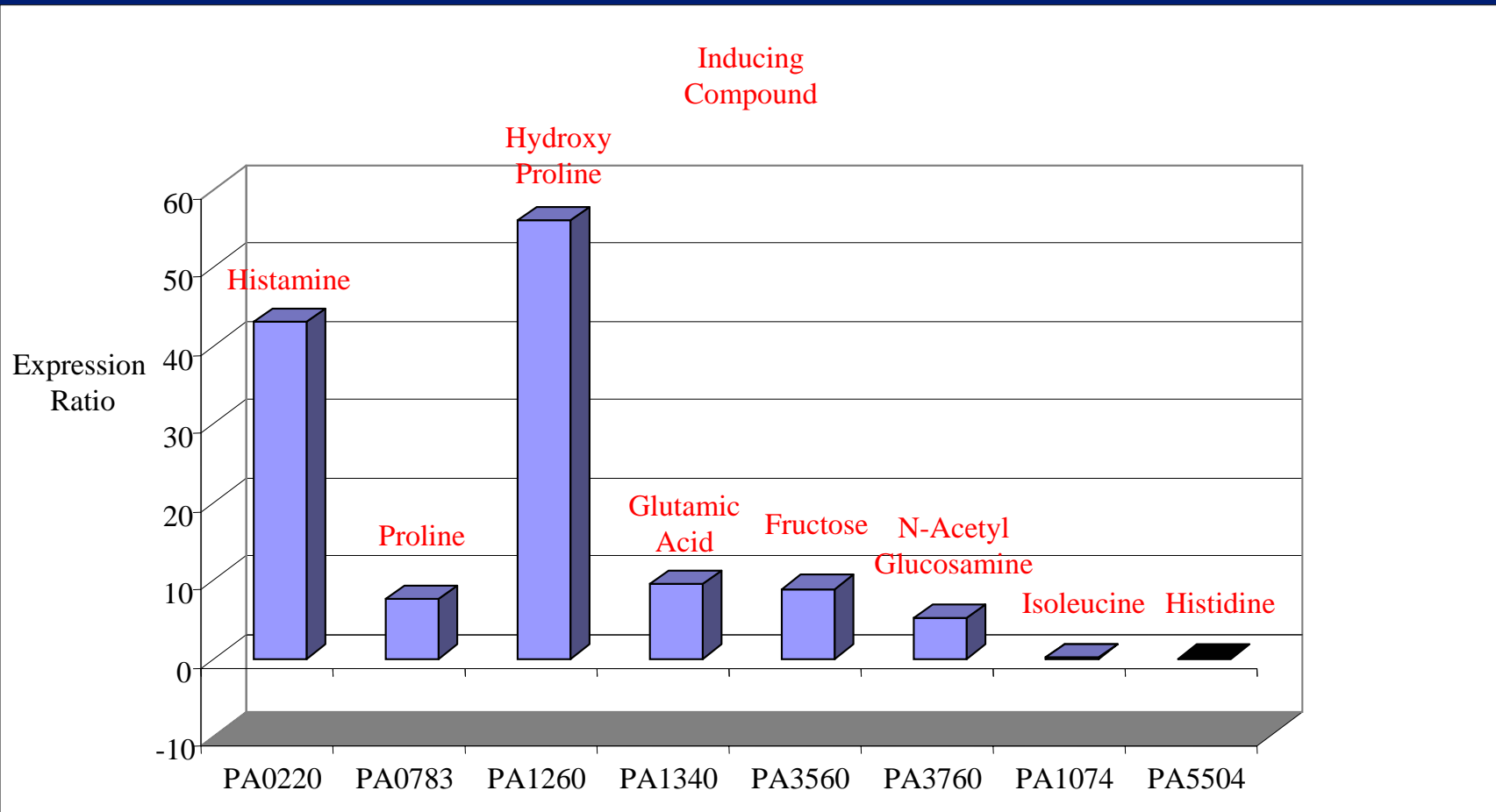
76 mutants tested on Biolog- putative amino acid/sugar transporters

- Growth on minimal media agar plates
  - Complementation studies with cloned genes
  - RT-PCR studies of gene expression
  - Membrane transport assays
- 27/76 mutants confirmed phenotypes
- 





# QRT-PCR analysis of *P. aeruginosa* transport gene expression



# Transport genes identified by Phenotype Arrays

- 041H09 – amino acid?
  - ◆ Histamine
- 086E06 – proline/sodium
  - ◆ L-Proline
- 062B02, 203A09, 109G08, and 041D07 – arginine/ornithine
  - ◆ L-Arginine and L-Ornithine
- 111E12, 188H02, 138B02 – amino acid
  - ◆ Hydroxy-L-Proline
- 137E08 - probable ABC transporter
  - ◆ L-Aspartic Acid
- 128B09 – branched chain amino acid
  - ◆ L-Alanine and D-Alanine

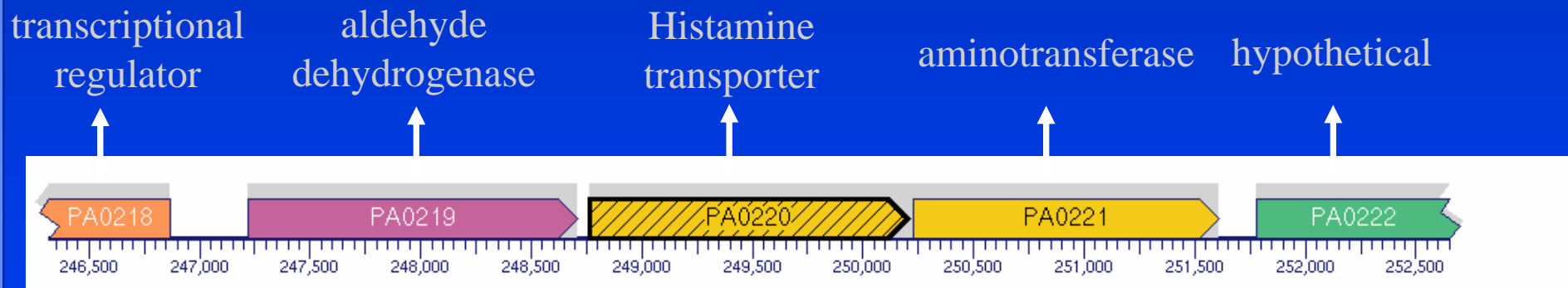


# Mapping functional transporter data onto PseudoCyc



# Novel pathway for histamine utilization?

- 27 novel transporter genes- 16 substrates
- All of these substrates except for one mapped onto predicted pathways in PseudoCyc
- One exception- histamine

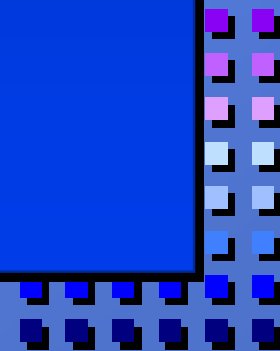


# Conclusions

- Incorporation and curation of membrane transporters in EcoCyc
- Enabling prediction of transport reactions using Pathway Tools
- Using Pathway Tools to assist in analysis of completed genomes
  - ◆ examples- *Brucella suis*, *Dichelobacter nodosus*
- Using Pathway Tools to assist in functional genomics analyses
  - ◆ Biolog analysis of transporter gene knockouts in *Pseudomonas aeruginosa*
  - ◆ Microarray expression analysis of *Synechococcus* WH8102



# Acknowledgments



# Acknowledgments

## TIGR

Kathy Phillippy

Aaron Johnson

TIGR faculty, sequencing and annotation core

Garry Myers

Qinghu Ren

## SRI International

Peter Karp

Suzanne Paley

Tom Lee

## Scripps Institute of Oceanography

Brian Palenik

Bianca Brahamsha Chris DuPont

## Virginia Tech

Steve Boyle

## Monash University

Julian Rood

Dane Parker

## Biolog Corporation

Barry Bochner

Jeffrey Carlson

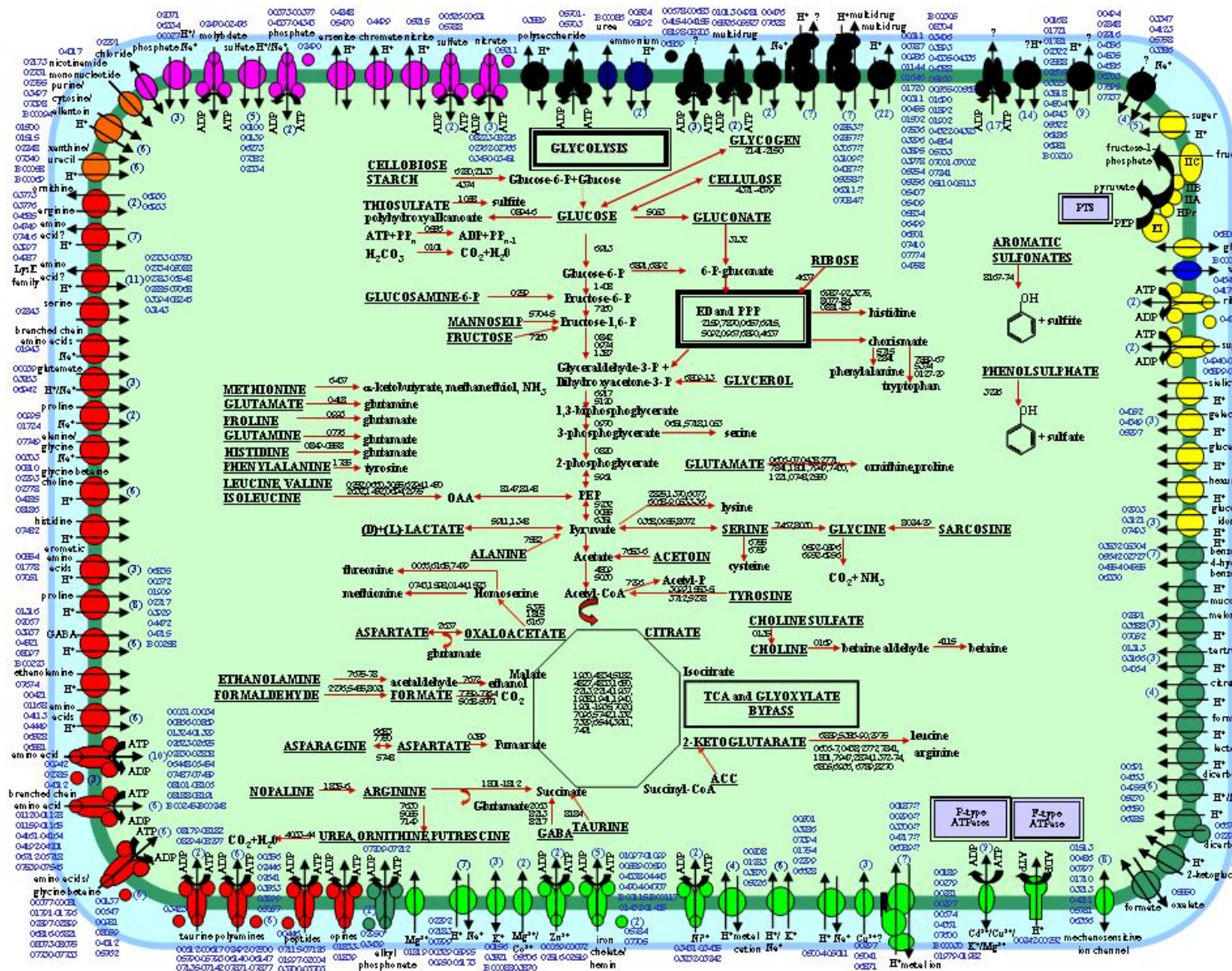
## NIH, DOE, USDA, NSF



TIGR

THE INSTITUTE FOR GENOMIC RESEARCH







# Current metabolic reconstruction approaches

