Reconstructing Metabolism by Comparative Genomics and Metabolite Analysis

Morgan N. Price
Arkin lab
August 2008

The Diversity of Metabolism

- Metabolism mostly worked out in a few organisms (*E. coli, S. cerevisiae, B. subtilis,* rat)
- Sequenced >1,000 genomes
  - Often don't use “standard” pathways
- Cannot predict a bacterium's physiology
  - What nutrients does this bacterium need?
  - What carbon sources can it utilize?
- New enzymes for metabolic engineering?

Towards Better Models of Metabolism

- Problems with gene annotations
  - Incorrect gene annotations & missing steps
  - Contradicted by growth requirements or capabilities
  - New biology?
- Metabolite data to further constrain pathways
- Use this data, and comparative genomics, to get better annotations

Outline

- Tutorial on how to annotate metabolism
- Reconstructing metabolism in *Desulfovibrio vulgaris* Hildenborough

Tutorial:
Annotating Pathways with Comparative Genomics

Lies, damn lies, and gene annotations

- Sequence similarity
- Adjacent genes tend to have related functions
- Look at entire pathway, not just individual genes

Similarity to Characterized Enzymes

- Known enzyme families (COG, PFam, etc.)
  - Enzymes within a family often prefer different substrates

COG0119 includes 2-IPM, homocitrate, citramalate, Re-citrate synthases, glutaryl-CH2OH lyase

PF00682 "HMGL-like: various aldolases & a region of pyruvate carboxylase"

VIMSS207383: **2-isopropylmalate synthase/homocitrate synthase family protein (TIGR), 538 a.a.** [Desulfovibrio vulgaris Hildenborough]

Description	Domain ID	Range	Ident	E-value	Start	End	Tree
Aldolase	SSF51569		--	0	1	297	T
Isopropylmalate/homocitrate/citramalate synthases	COG0119		--	1e-85	3	429	T
LYASE	PDB:2cw6A		23.3%	1e-06	3	284	
Aldolase-type TIM barrel	G3DSA:3.20.20.70		--	0	3	283	T
LYASE / OXIDOREDUCTASE	PDB:1nvmA		24.6%	5e-09	3	254	
2-isopropylmalate synthase/homocitrate synthase related	TIGR00977		--	0	4	535	T
Pyruvate carboxyltransferase	PS50991		--	--	5	270	
Alpha-isopropylmalate/homocitrate synthase	PS00815		--	--	12	28	
Pyruvate carboxyltransferase	PF00682		--	8.4e-31	13	285	T
[low-complexity (repetitive) sequence]	seg		--	--	92	102	
ISOPROPYLMALATE SYNTHASE RELATED	PTHR10277		--	0	109	527	
CARBOXYLASE:PYRUVATE/ACETYL-COA/PROPIONYL-COA CARB	PTHR18866		--	0.00098	154	227	
[low-complexity (repetitive) sequence]	seg		--	--	221	232	
LeuA allosteric (dimerisation) domain	PF08502		--	2.3e-29	378	526	T
[low-complexity (repetitive) sequence]	seg		--	--	378	389	
2-isopropylmalate synthase LeuA, allosteric (dimer	SSF110921		--	3e-19	401	526	T

"Domains & families" tab from MicrobesOnline.org

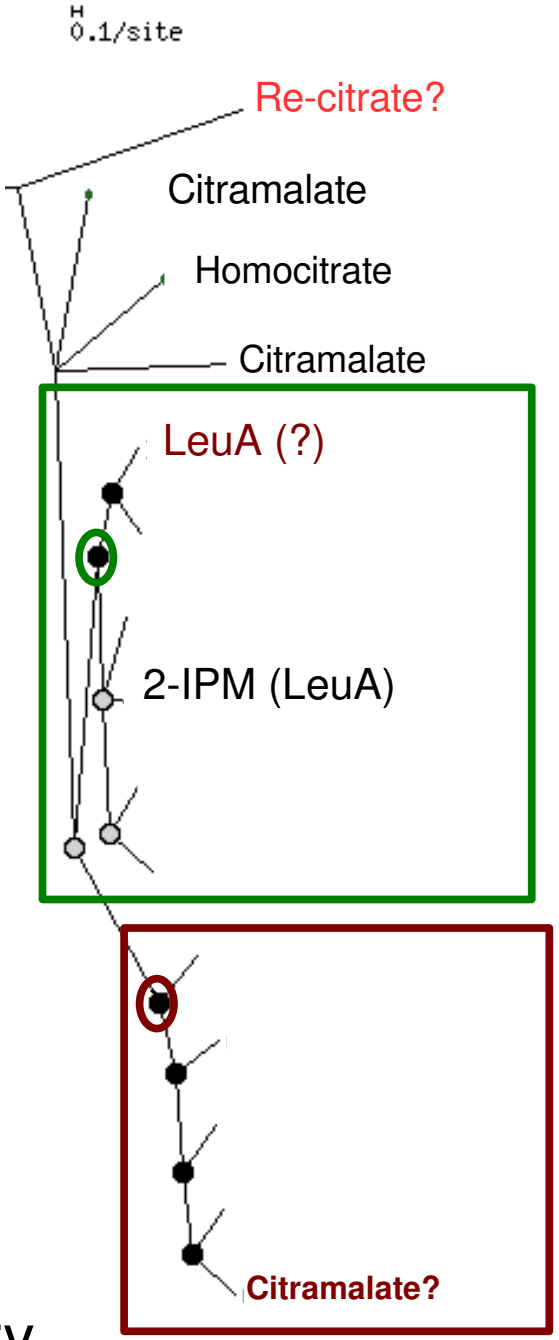
Similarity to Characterized Enzymes

- BLAST hits
  - Below 50% identity, function often changes
  - Which homolog is most relevant?

	28.57%	2-isopropylmalate synthase (leuA)	*Methanocaldococcus jannaschii*	Cart
	27.81%	2-isopropylmalate synthase (leuA) (see papers)	*Methanocaldococcus jannaschii*	Cart
P	28.78%	2-isopropylmalate synthase	*Desulfovibrio vulgaris Hildenborough*	Cart
	25.38%	2-isopropylmalate synthase	*Leptospira interrogans, L1-130*	Cart
	25.38%	alpha-isopropylmalate synthase	*Leptospira interrogans, 56601*	Cart
	31.28%	2-isopropylmalate synthase (leuA)	*Methanocaldococcus jannaschii*	Cart
	27.85%	alpha-isopropylmalate synthase	*Leptospira interrogans, 56601*	Cart
	27.85%	2-isopropylmalate synthase	*Leptospira interrogans, L1-130*	Cart
	22.43%	alpha-isopropylmalate synthase	*Leptospira interrogans, 56601*	Cart
	22.43%	2-isopropylmalate synthase 2	*Leptospira interrogans, L1-130*	Cart
P	26.36%	homocitrate synthase	*Desulfovibrio vulgaris Hildenborough*	Cart
P	24.80%	HMGL-like domain protein	*Desulfovibrio vulgaris Hildenborough*	Cart
	23.55%	Hydroxymethylglutaryl-CoA lyase	*Leptospira interrogans, 56601*	Cart

Phylogenetic Trees

- Protein function is conserved by evolution
- Phylogenetic groups (clades) often have consistent functions



Rooting is arbitrary

x axis is evolutionary distance

Functional Residues

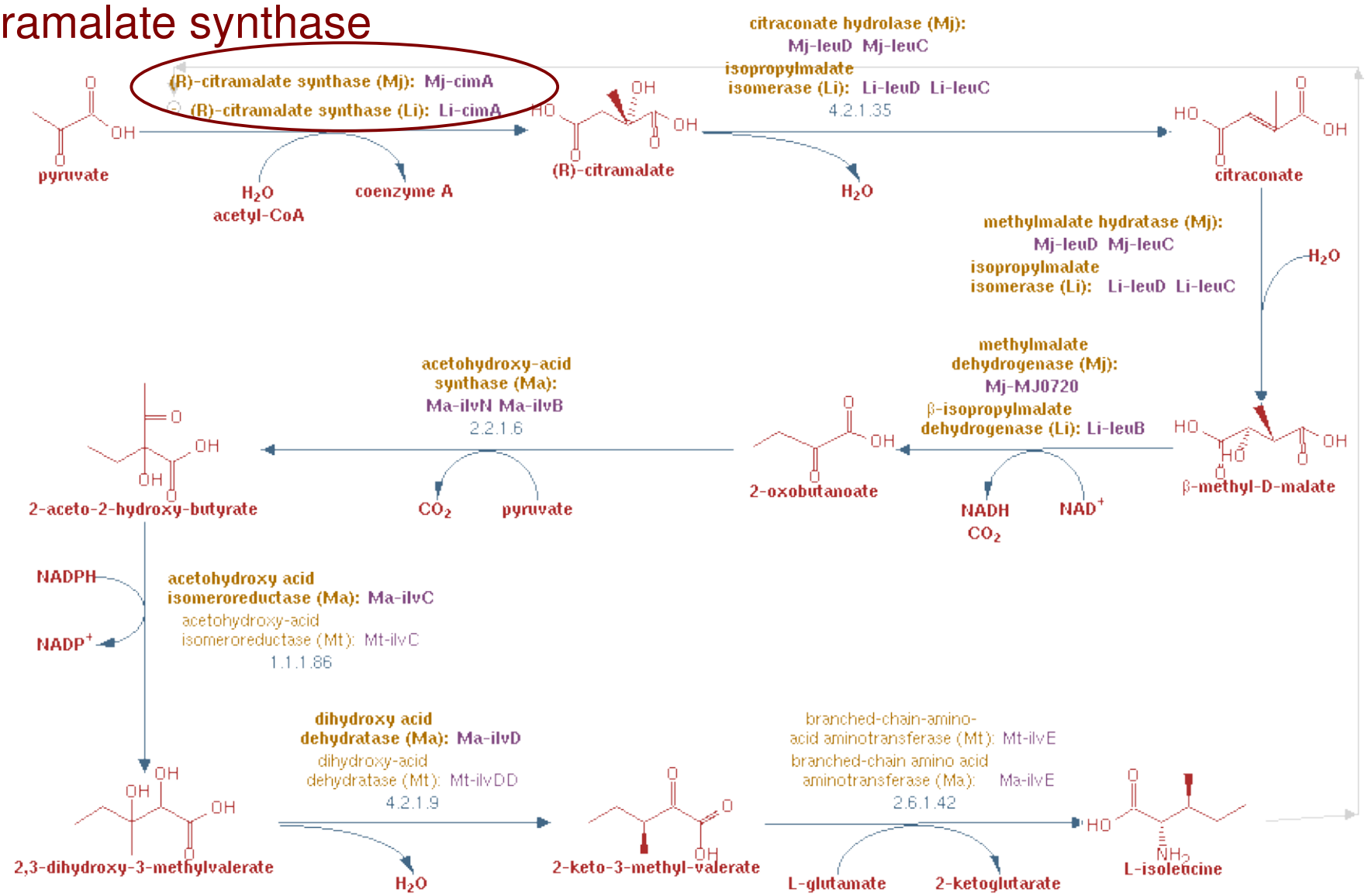
- E.g. catalytic residues in the active site
- Confirm that gene is likely to be an enzyme, but
  - Specificity-determining a. a. usually not known
  - Variant amino acids can function

Sequence-HMM alignment, with Pfam annotation, from MicrobesOnline.org

More Characterized Enzymes

Pathways in MetaCyc, e.g. **Isoleucine synthesis II**

Citramalate synthase



Similarity to Characterized Enzymes

Homologs from M. jannaschii & L. interrogans (MicrobesOnline.org)

VIMSS207383: **2-isopropylmalate synthase/homocitrate synthase family protein (TIGR), 538 a.a.** [Desulfovibrio vulgaris Hildenborough]

BLASTp Report: 200 | homologs in selected genomes | change

Annotation: 2-isopropylmalate synthase

J. Bacteriology: citramalate synthase

Best COG	33.33%	COG119 (E) Isopropylmalate/homoc		
Best KEGG	100%	dvu:DVU1914 2-isopropylmalate syn 2-isopropylmalate synthase		
Best UniProt	50.57%	O66682|LY41_AQUAE Putative AIPM/Hcit synthase family transferase aq_356 (EC 2.3.3.-). Aquifex aeolicus. (O66682)		
	28.57%	2-isopropylmalate synthase (leuA)	*Methanocaldococcus jannaschii*	Cart
	27.81%	2-isopropylmalate synthase (leuA) (see papers)	*Methanocaldococcus jannaschii*	Cart
P	28.78%	2-isopropylmalate synthase	*Desulfovibrio vulgaris Hildenborough*	Cart
	25.38%	2-isopropylmalate synthase	*Leptospira interrogans, L1-130*	Cart
	25.38%			
	31.28%			
	27.85%			
	27.85%			
	22.43%			
	22.43%			
P	26.36%			
P	24.80%			
	23.55%			
	23.55%			

Done

NCBI PubMed www.pubmed.gov

A service of the U.S. National Library of Medicine and the National Institutes of Health

My NCBI — Welcome mnprice. [Sign Out]

All Databases | PubMed | Nucleotide | Protein | Genome | Structure | OMIM | PMC | Journals | Books

Search PubMed for Go Clear Advanced Search (beta)

Limits | Preview/Index | History | Clipboard | Details

About Entrez
Text Version

Entrez PubMed
Overview
Help | FAQ
Tutorials
New/Noteworthy
E-Utilities

Display Abstract Show 20 Sort By Send to

All: 1 Review: 0

1: J Bacteriol. 1999 Jan;181(1):331-3. Related Articles, Links

Final Version FREE J Bacteriol

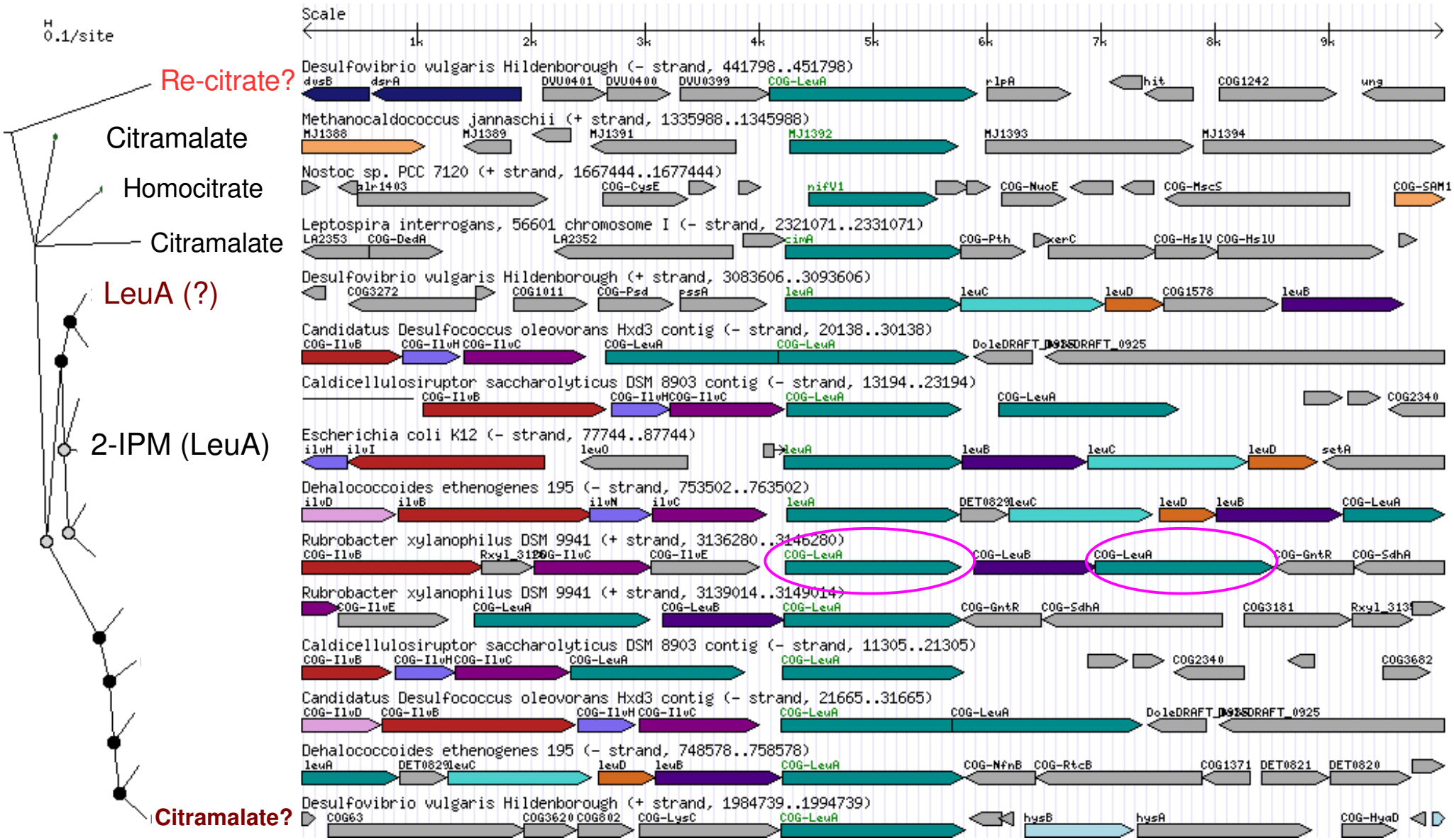
FREE full text article in PubMed Central

(R)-citramalate synthase in methanogenic archaea.

Gene Proximity

- In bacteria, adjacent genes tend to have related functions
  - Because of operons & horizontal gene transfer
- Helps predict a gene's function
  - which member of a family has a given function
  - candidates for missing steps
- Look at close homologs, even if the gene itself has no relevant neighbors

The 2-Isopropylmalate/Homocitrate/Citramalate/Re-citrate Synthase Family



rooting is arbitrary

Tree-browser from MicrobesOnline.org

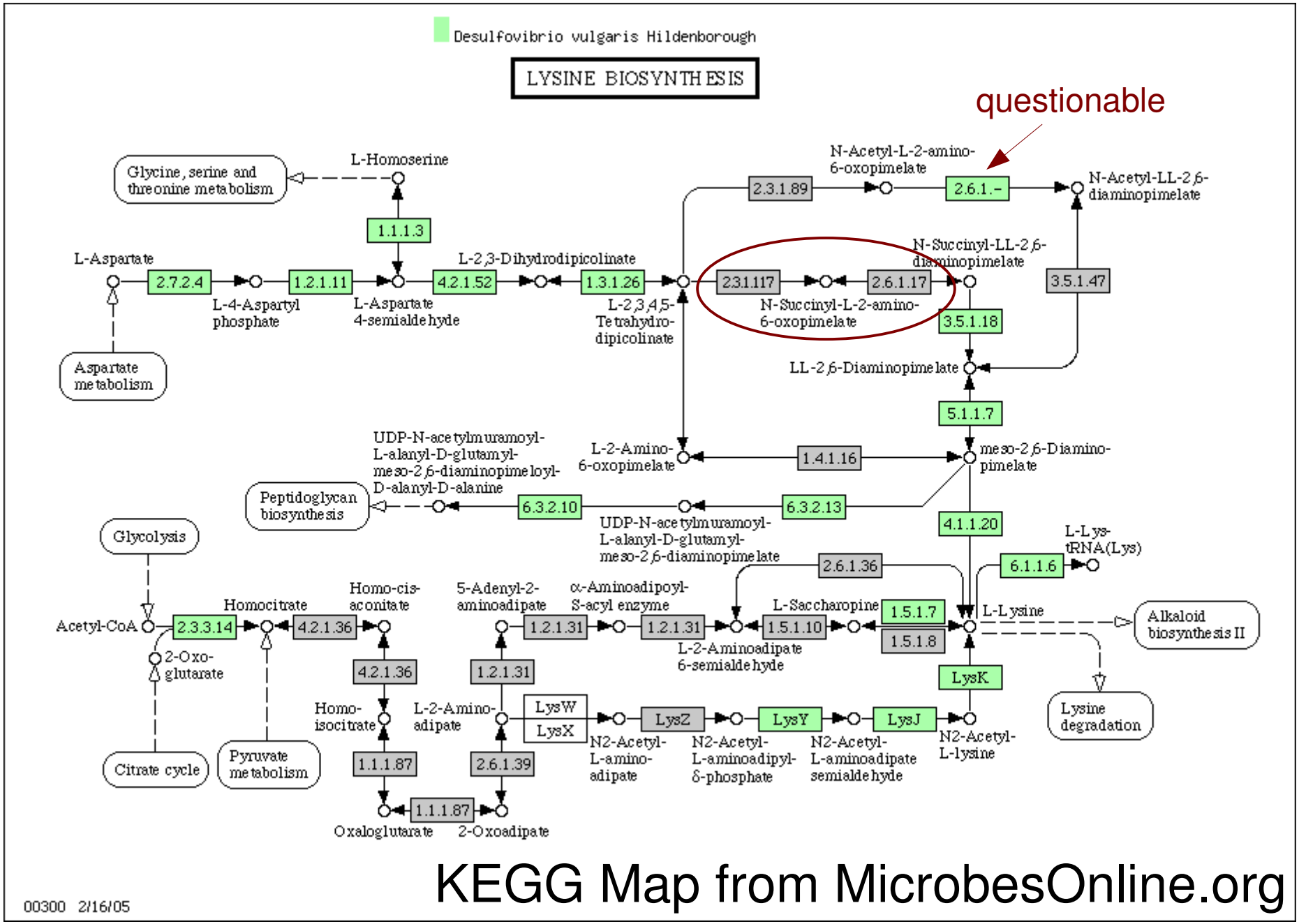
Reactions of the IPM Synthase Family

- oxalacetate + acetyl-CoA + H_2O <=> citrate + CoA
- pyruvate + acetyl-CoA + H_2O <=> citramalate + CoA
- 3-methyl-2-oxobutanoate + acetyl-CoA + H_2O
  <=> 2-isopropylmalate + CoA
- α-ketoglutarate + acetyl-CoA + H_2O <=> homocitrate + CoA
- acetoacetate + acetyl-CoA <=> HMG-CoA

Pathway Analysis

- Pathway analysis
  - if genes for 9/10 steps are present, 10th step is probably there too
  - Conversely, if only 1 step of a pathway is annotated, it's probably wrong

Missing Steps in Lysine Synthesis



KEGG Map from MicrobesOnline.org

Summary of Tutorial

- Lies, damn lies, and gene annotations
- Annotating pathways relies on
  - Similarity to known pathways
  - Gene proximity
  - Pathway analysis
- Questions?

Identifying Pathways in Desulfovibrio

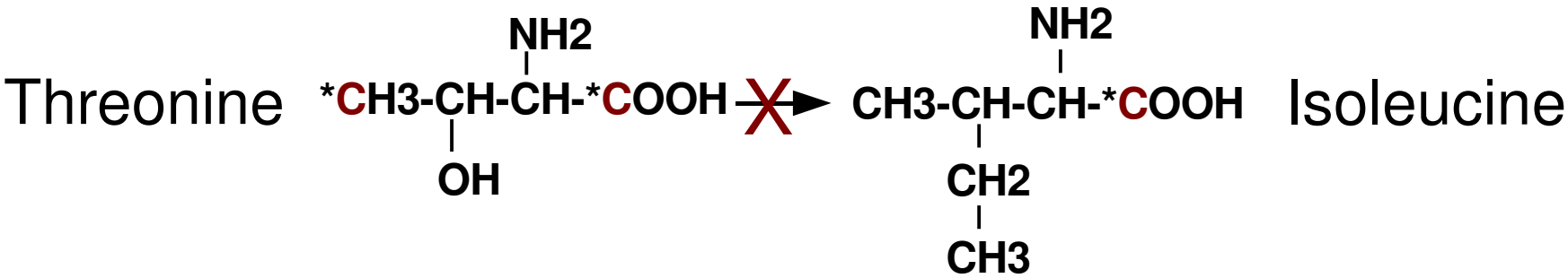
- *D. vulgaris* Hildenborough (DvH) is a model sulfate-reducing bacterium (δ-Proteobacteria)
  - oxidizes lactate to acetate
- No annotated pathways for synthesis of Ile, Met, Lys
  - but it grows in defined media (no amino acids)
- Questions about the TCA cycle and folate synthesis
- I've predicted a bunch of pathways...

Experimental Approaches

- Metabolite labeling (Yinjie Tang)
- Presence/absence of metabolites (Peter Benke, Edward Baidoo, and Aindrila Mukhopadhyay)
- Gene knockouts (Grant Zane, Judy Wall's lab)
- *In vitro* assays (Swapnil Chhabra, Keasling lab)
- Complementation assays (Swapnil Chhabra)

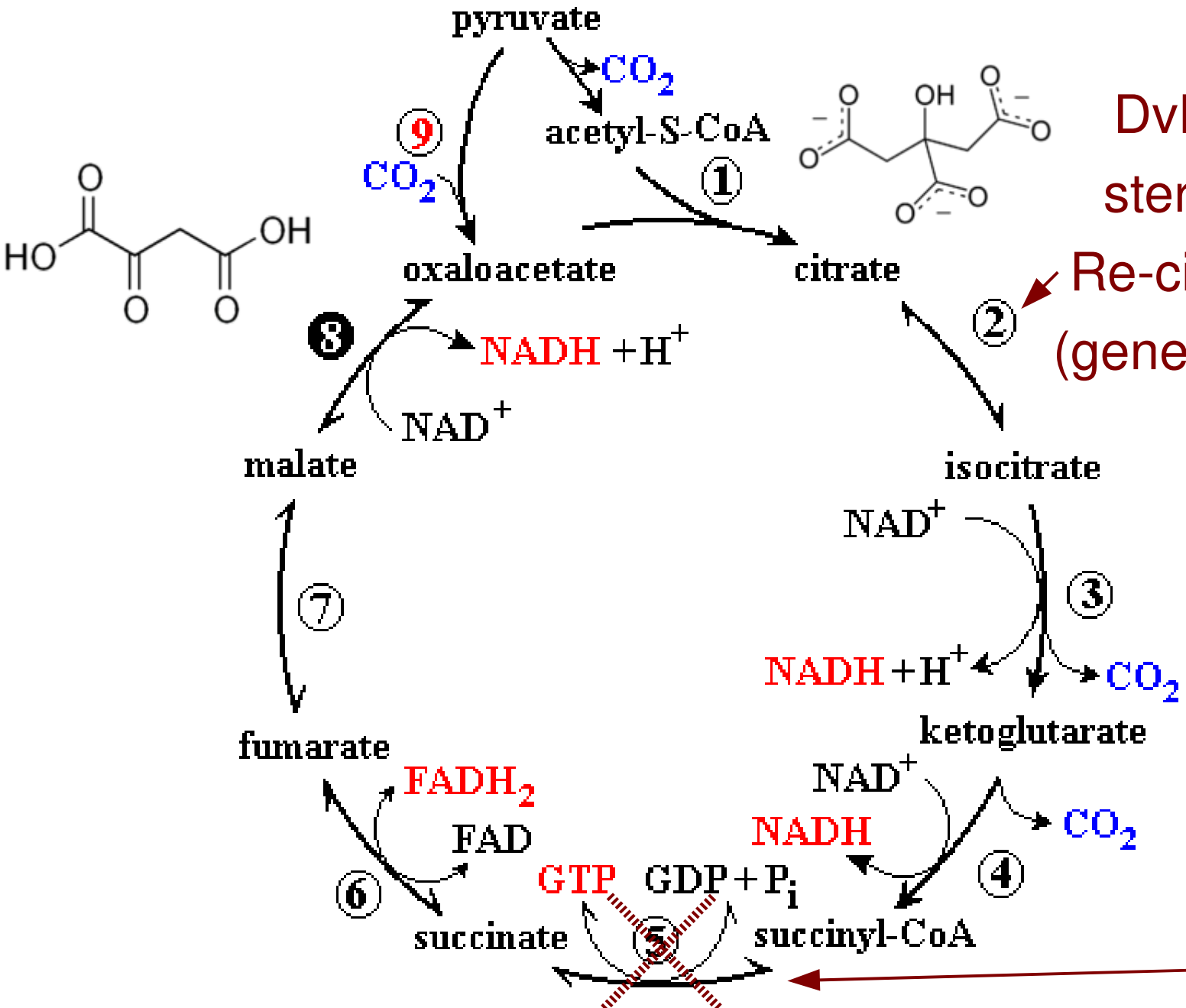
Isotopic Labeling Analysis

- Grow DvH on 1-^{13}C lactate
- Analyze labeling of amino acids (GC/MS, FT-ICR/MS)
- Infer sources of amino acids & metabolic fluxes
  - TCA cycle ends at α-ketoglutarate & succinate
    - labeling of glutamate confirms Re-citrate synthase
  - isoleucine does not originate from threonine + acteyl-CoA



Y. Tang et al. 2007, J Bacteriol. 189:940-9

Tricarboxylic Acid (Krebs) Cycle



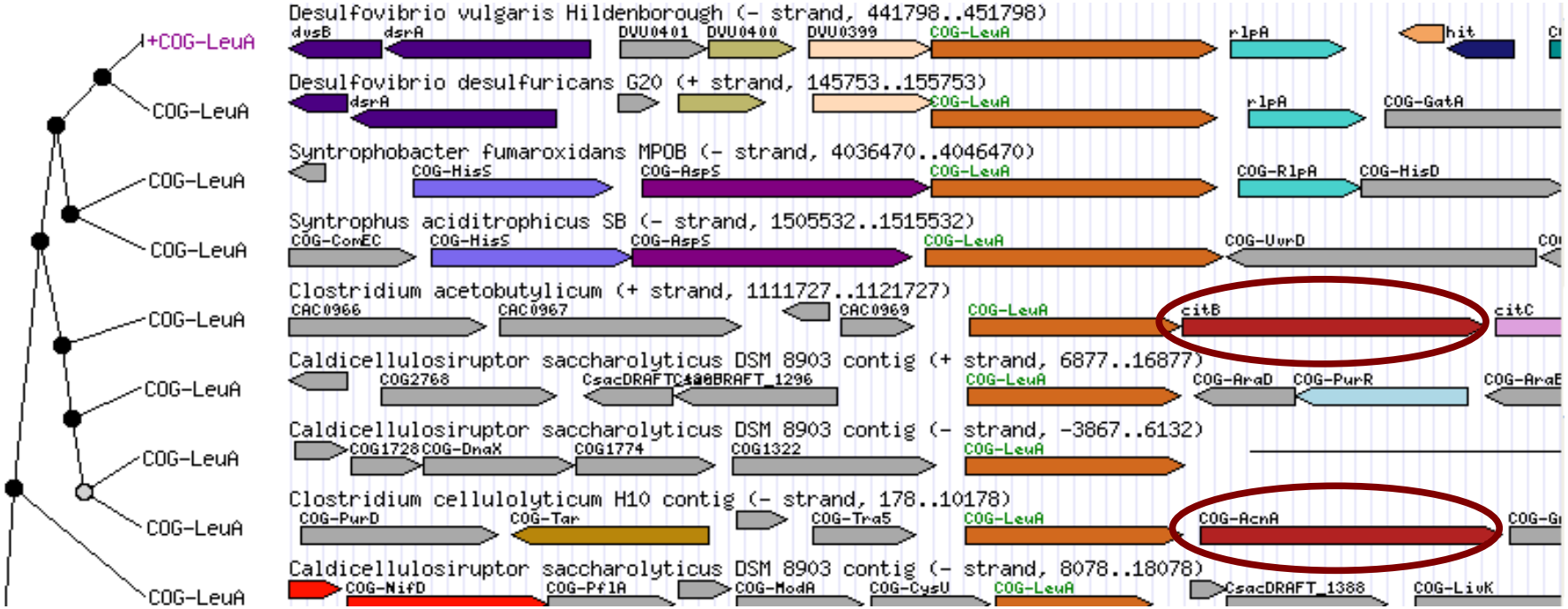
DvH as atypical stereochemistry: Re-citrate synthase (gene was unknown)

DvH genome has possible α-ketoglutarate synthases

Finding Re-citrate Synthase

Might be a member of the IPM synthase family?

Adjacent to aconitase in Clostridia!



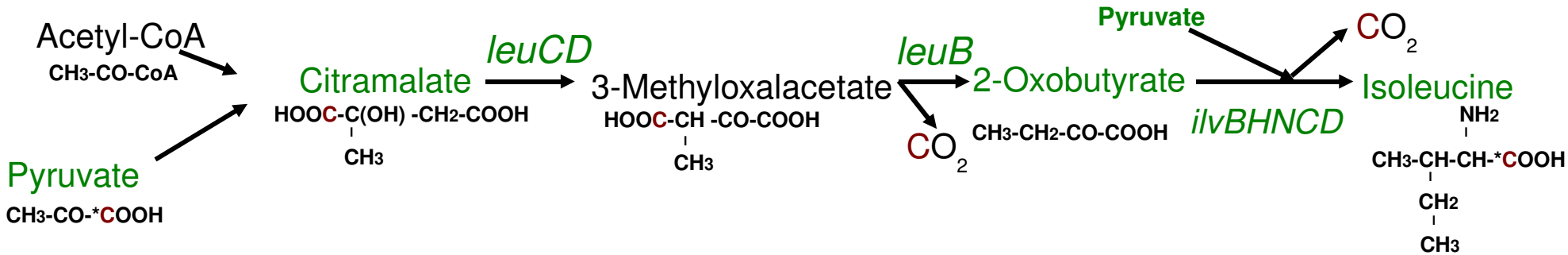
E-mailed Gottschalk, the discoverer of Re-citrate synthase

tested the Clostridial enzyme *in vitro* (F. Li et al. JB2007)

DVU0398 is Re-Citrate Synthase

- *In vitro*, cleaves acetyl-CoA in presence of oxaloacetate
  - requires Mn^{2+} for activity; inactivated by O_2
  - Swapnil Chhabra & Sam Meyers, Keasling lab
- *DVU0398* knockout requires α-ketoglutarate, glutamate, or glutamine to grow
  - confirms TCA cycle is broken at α-ketoglutarate dehydr.
  - Grant Zane, Judy Wall's lab

Predicted Citramalate Synthase DVU1914



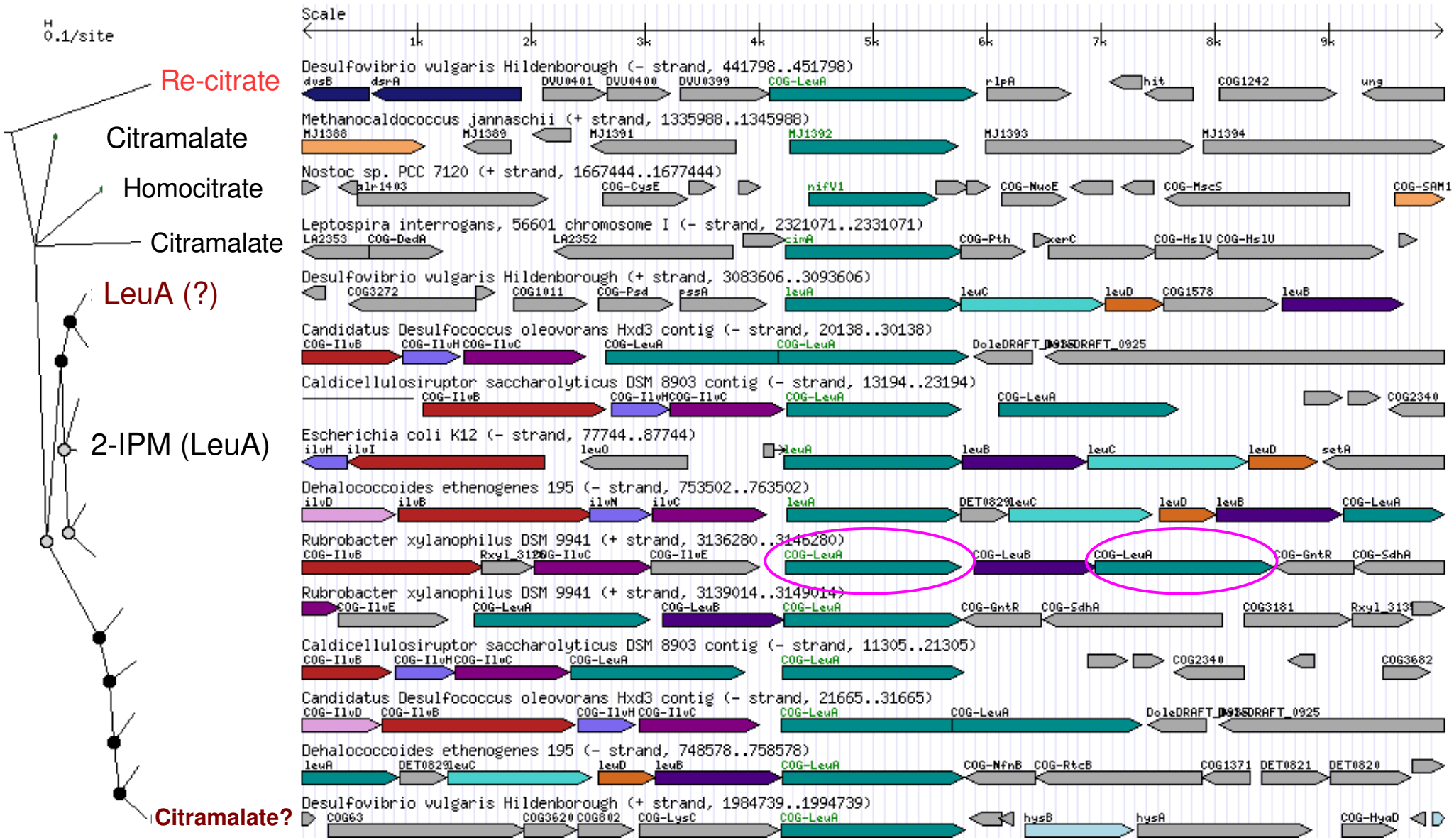
50% label of Ile could be due to reversible decarboxylation

Green: identified by metabolite analysis (Edward Baidoo & Peter Benke, Keasling lab)

DVU1914 is a member of the IPM synthase family

- close homologs of *DVU1914* are often near *leuA*, *ilvBCDHN*
- >50% identical to recently identified citramalate synthase in *G. sulfurreducens* (C. Risso et al. JB2008)
- Cloning into *E. coli* ilvA$^-$ strain (Swap)

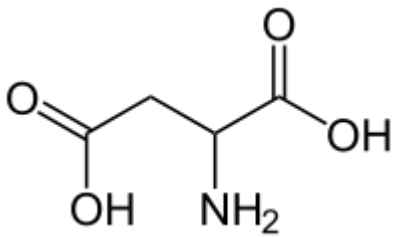
The 2-Isopropylmalate/Homocitrate/Citramalate/Re-citrate Synthase Family



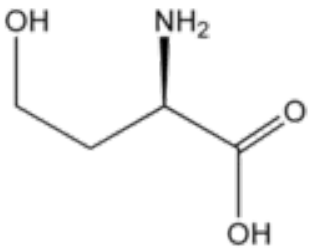
rooting is arbitrary

Tree-browser from MicrobesOnline.org

Methionine Synthesis



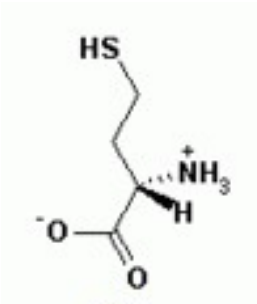
Aspartate → → → Homoserine



Activation ✗

O-modified-homoserine

Sulfhydrylation ✗



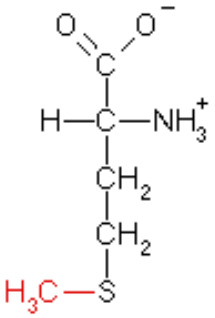
S-adenosylhomocysteine (SAH) —adenosine→ Homocysteine

methyl donation

methyl-THF → THF

S-adenosylmethionine (SAM) ← Methionine

Pi, PPi ATP



but we know DvH & G20 can do it

Activation & Sulfhydrylation

Homoserine

metA

metX

thrB

Threonine

O-succinyl-homoserine

O-acetyl-homoserine

O-phospho-homoserine

(to cystathionine in plants)

cysteine

metB

succinate

metW?

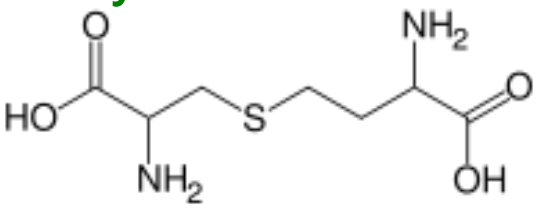
cysteine

sulfide

cystathionine

acetate

metY or metZ or met17

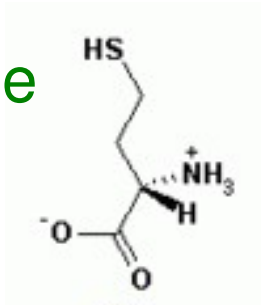


metC, malY or **patB?**

pyruvate

ammonia

homocysteine



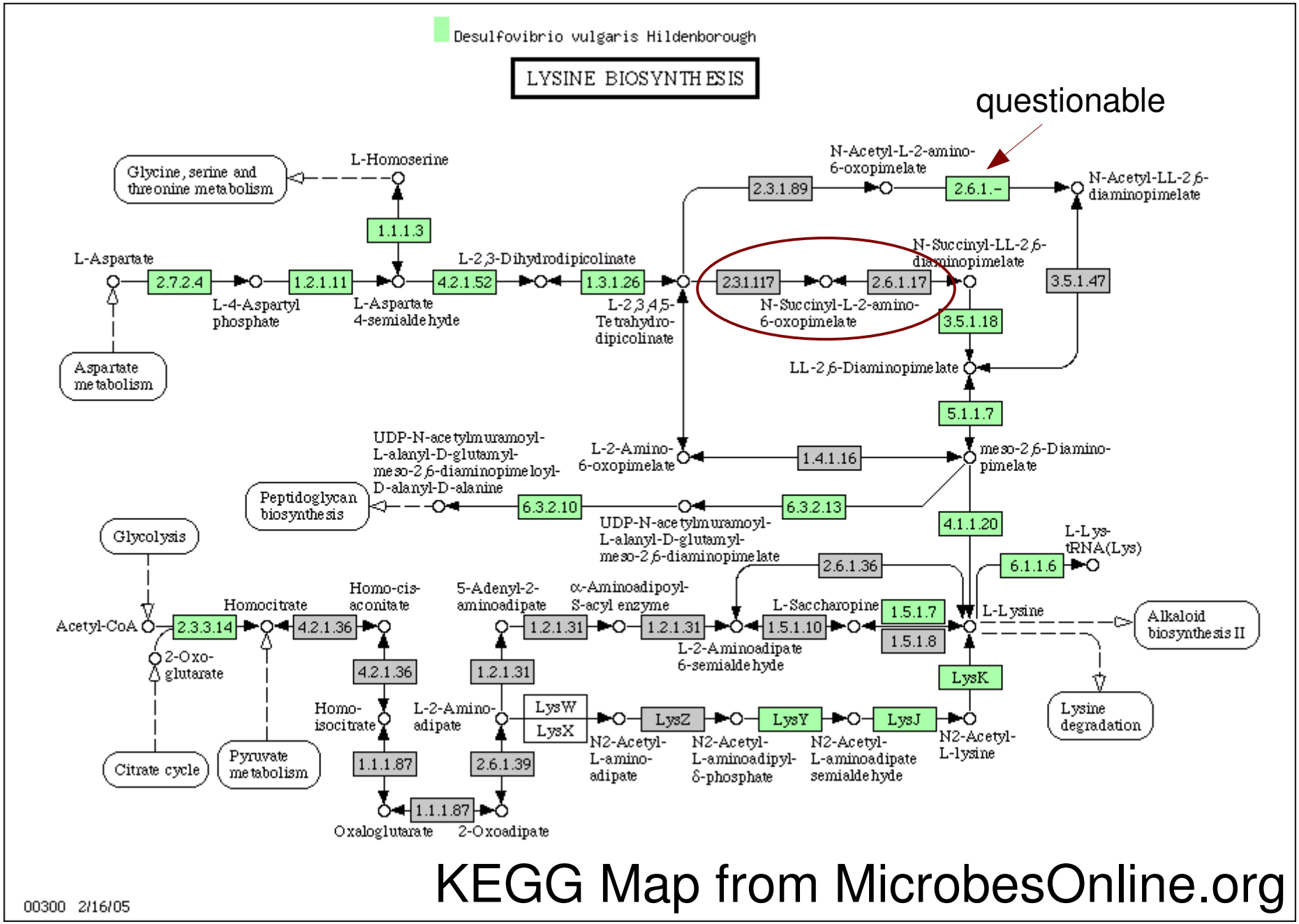
Methionine Synthesis Mystery

- Homoserine not seen by metabolomics
  - Thr synthesis expected to require it
  - Has the enzyme to make it
  - Present at low levels??
- Knockouts still grow w/o external methionine
  - "*metW*", *DVU3369*; "*patB*"; *DVU0171*
  - Grant Zane, Judy Wall's lab

More Methionine Experiments

- Gene expression with added methionine (Joe Zhou lab)
  - Not helpful yet
- Inhibitors of the known pathways (Aindrila)
- Protein complexes of adjacent enzymes (Swap)
- *D. alaskensis* G20 has a similar mystery
  - Systematic mutagenesis (Adam D. & Jen K.)

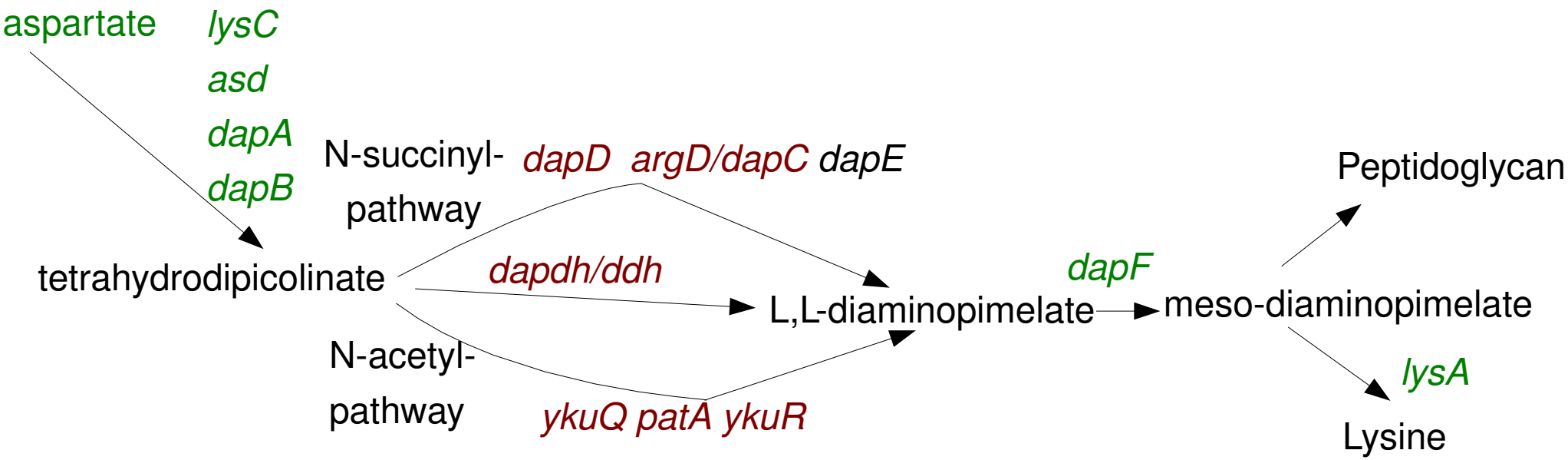
Missing Steps in Lysine Synthesis



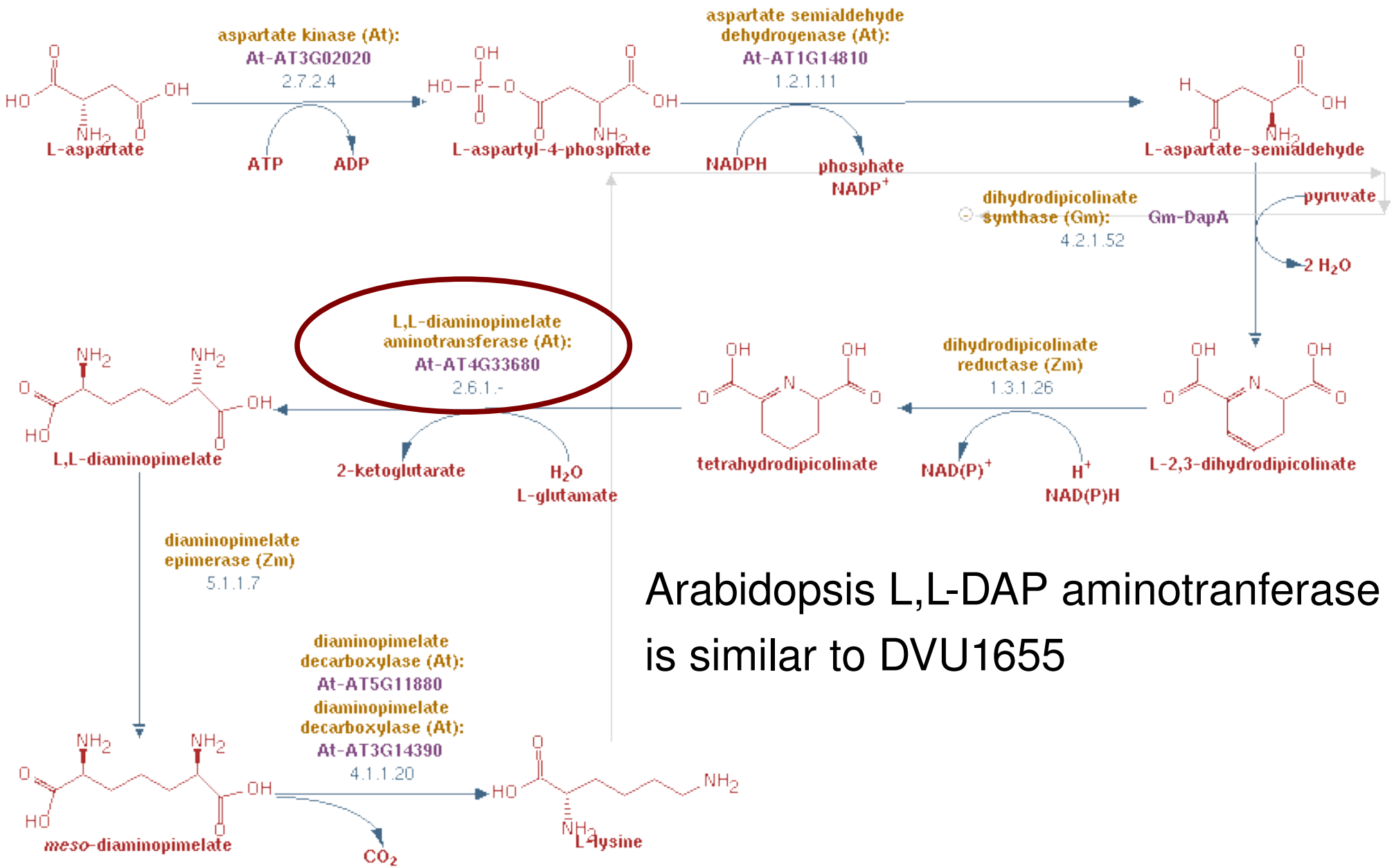
KEGG Map from MicrobesOnline.org

Diaminopimelate Synthesis

DvH missing steps from tetrahydrodipicolinate to L,L-DAP

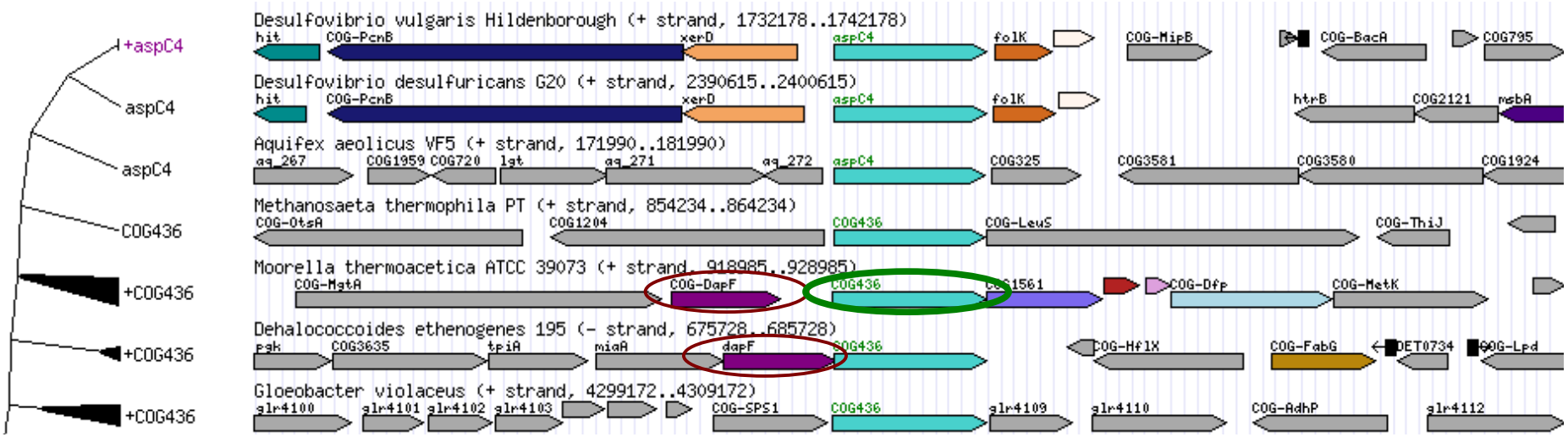


Direct Amination (Arabidopsis)



Arabidopsis L,L-DAP aminotranferase is similar to DVU1655

Bacterial L,L-DAP Aminotransferases



Homologs of DVU1655 adjacent to dapF (the next step)

Some were recently shown to complement *E. coli* dap$^-$

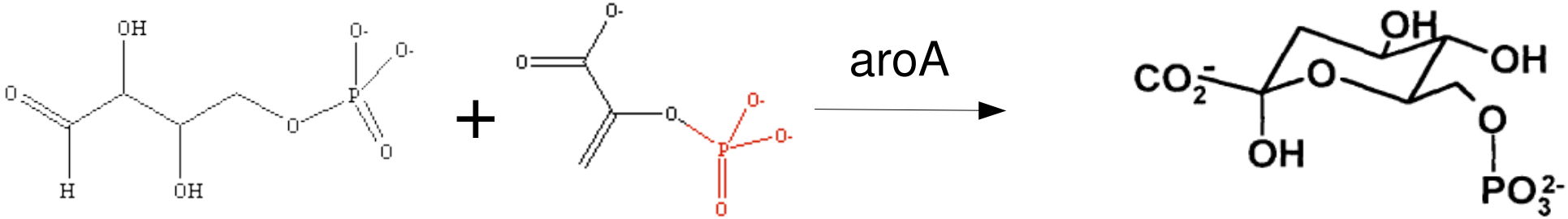
A. Hudson et al., JB2008

Cloning DVU1655 into *E. coli* dapD-/dapE- strain (Swap)

Chorismate Synthesis

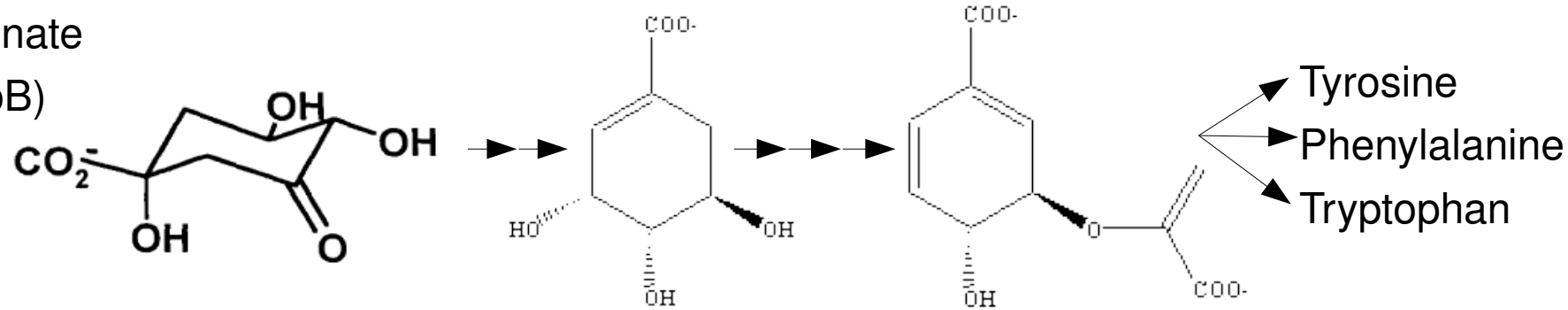
Required for synthesis of aromatic amino acids

E4P from the pentose phosphate pathway



D-Erythrose-4-P PEP 7-Phospho-2-dehydro-3-deoxy-D-arabinoheptulosonate (DAHP)

3-Dehydroquinate synthase (aroB)

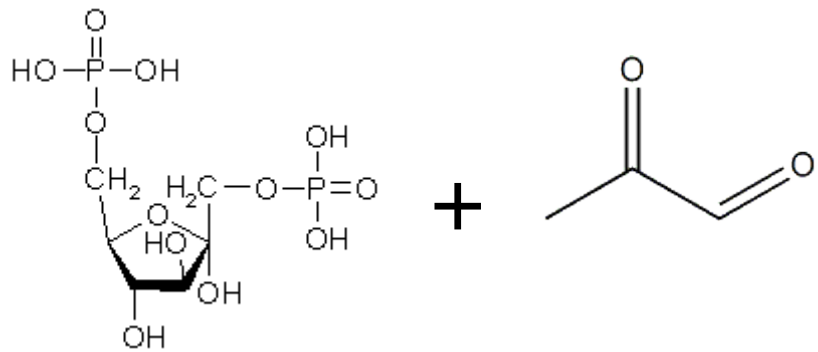


3-dehydroquinate shikimate chorismate

Archaeal-Type Pathway

In *Methanococcus* & *Methanocaldococcus*

I. Porat et al. Mol. Micro. 2006; R.H. White & H. Xu Biochemistry 2006

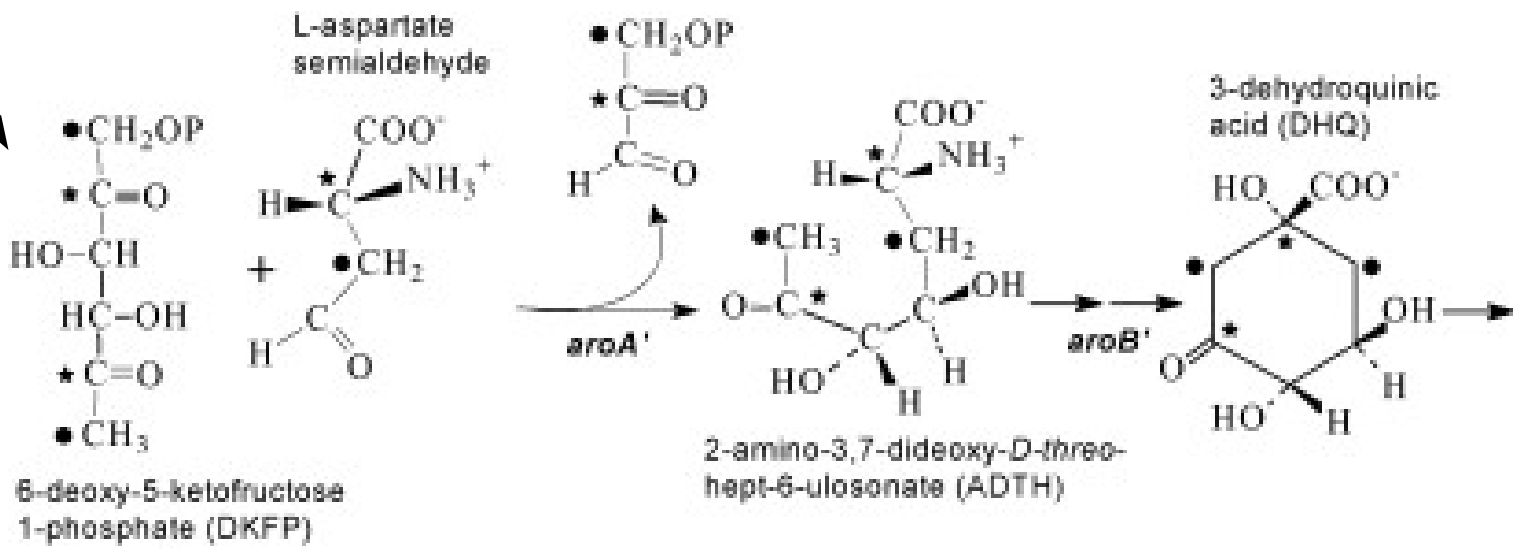


+

Methylglyoxal
(from side reactions of triose-phosphate isomerase?)

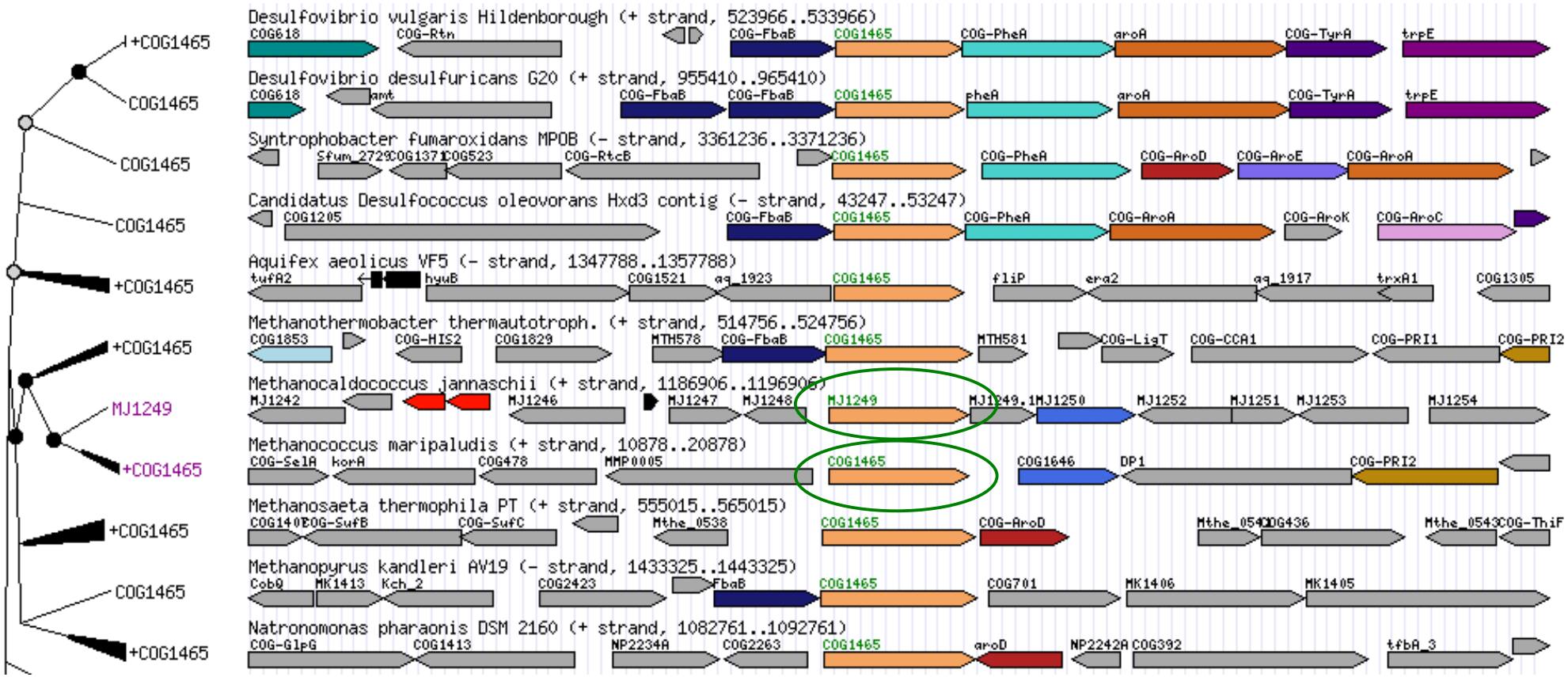
Fructose-1,6,-bisphosphate

DKFP aldolase



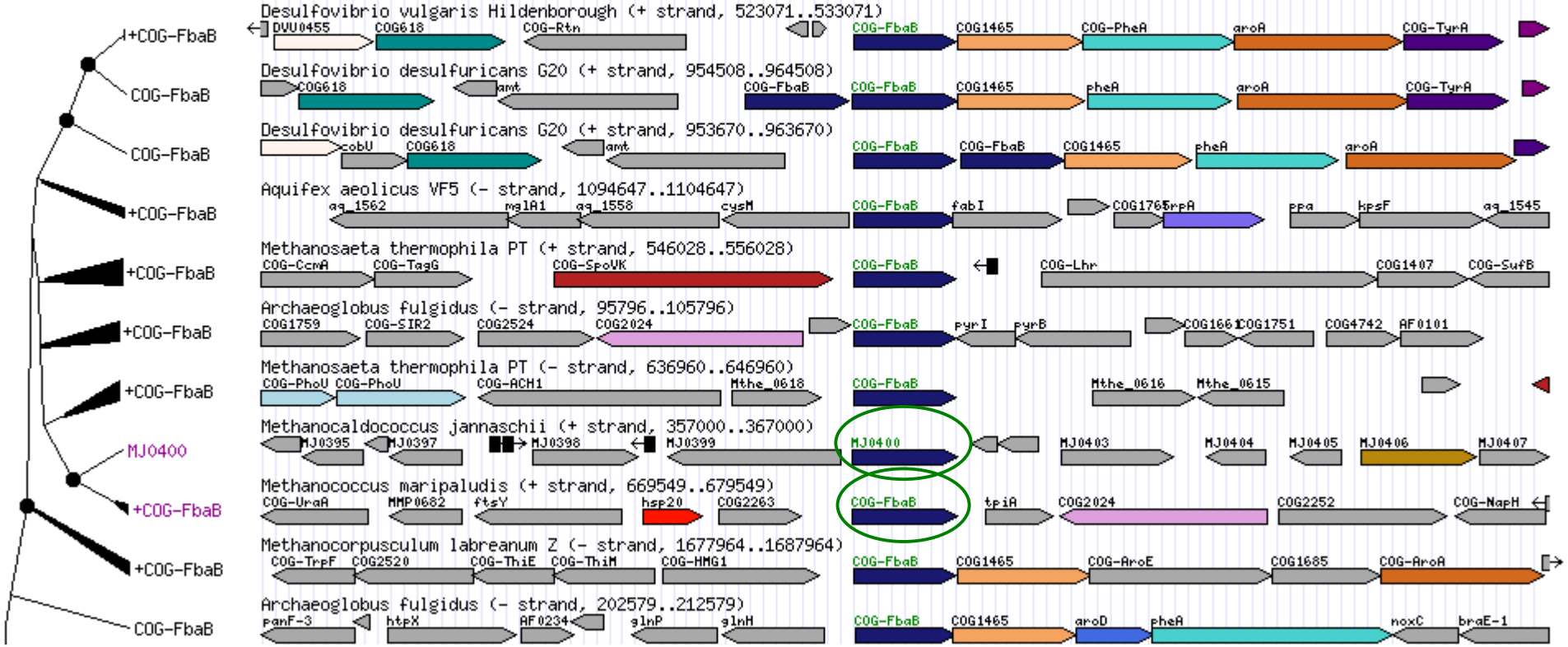
shikimate, etc.

D. vulgaris has archaeal-type pathway?



3-dehydroquinate synthase in an aromatic a.a. synthesis operon, next to

D. vulgaris has archaeal-type pathway?



ADTH aldolase. But where does the DKFP come from?

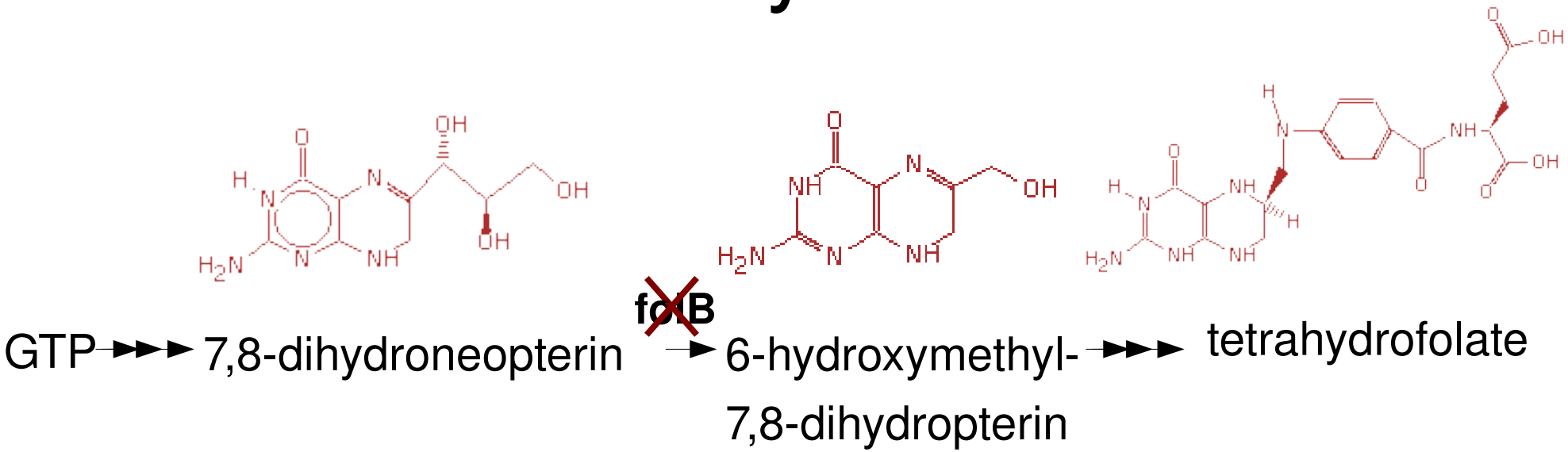
DvH has no close homolog of the DKFP aldolase

Does DvH even have methylglyoxal? (No mgsA)

Chorismate Synthesis, Next Steps

- Cloning DvH aroA' and aroB' into *E. coli* aroB^{-} (Swap)
  - Should complement if aroA' is a bi-functional aldolase or if the starting substrate is in *E. coli*
- Metabolite measurements, but
  - standards are not readily available
  - key compounds are uncharged or phosphorylated

Folate Synthesis



- DVU1658 predicted Dihydroneopterin aldolase
  - based on gene order & absence of folB/folX
  - V. Crecy-Lagard et al., BMC Genomics 2007
- Jennifer K. is knocking it out

Summary:
Revisions to DvH Metabolic Annotation

- Incomplete TCA cycle with an atypical Re-citrate synthase
- Synthesis of isoleucine via citramalate synthase
- Synthesis of methionine via cystathionine & homocysteine
- Synthesis of lysine via L,L-DAP aminotransferase
- Synthesis of chorismate via archaeal-type aldolase & 3-dehydroquinate cyclase