

*A graph theoretic method for
comparing TCA cycles of
some species*

Losiana Nayak and Rajat K. De

{losiana_t,rajat}@isical.ac.in

Machine Intelligence Unit

Indian Statistical Institute

203 B. T. Road

Kolkata-700108, India

Objective

- *Comparison of metabolic pathways*

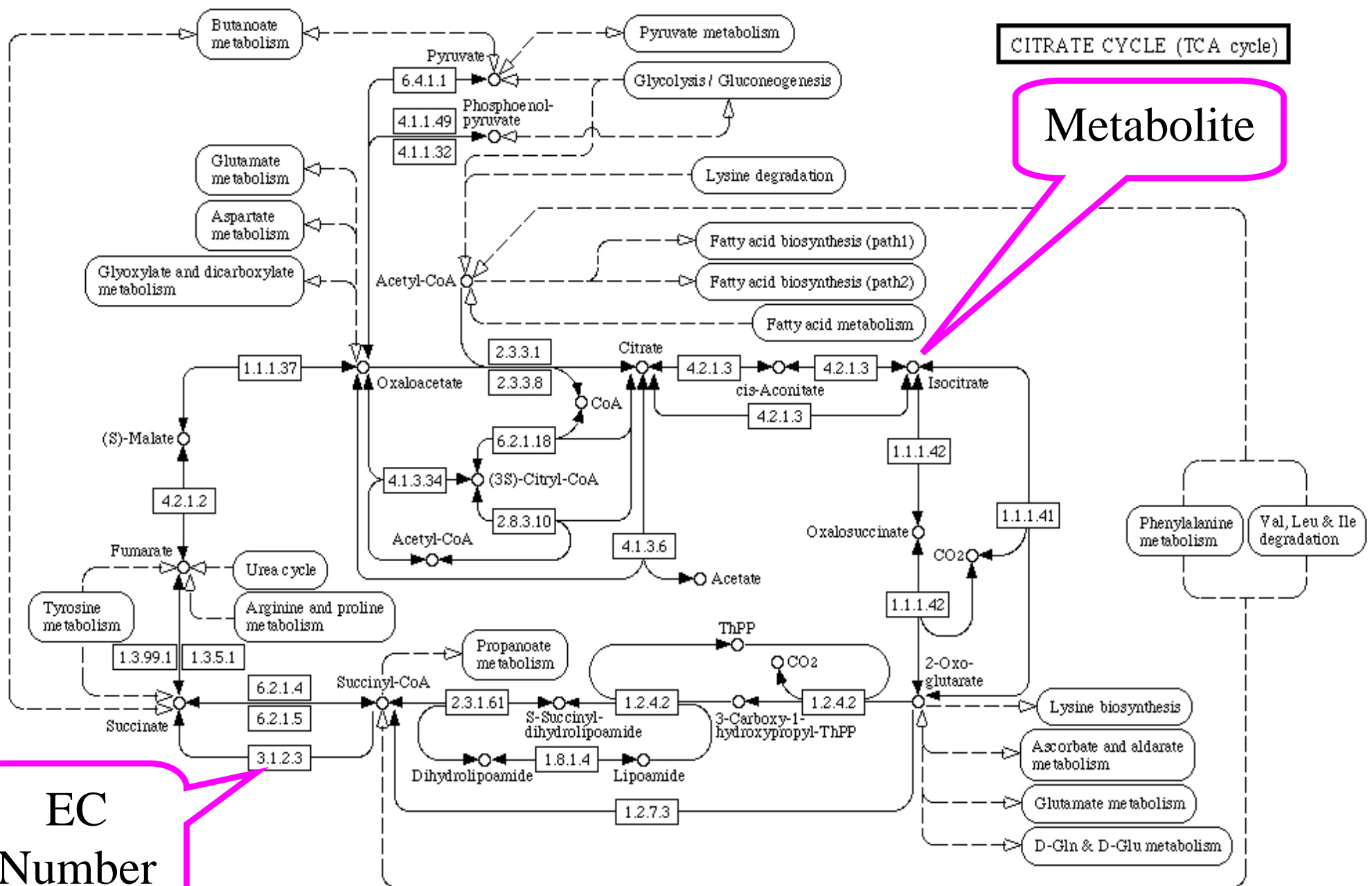
Future use

- *Evolution/ Development of metabolic pathways*
- *Its similarity to conventional evolution*
- *Patterns in development of metabolic pathways*
- *Stages of major change*

We took TCA Cycle for a certain set of species from KEGG/Pathway database

<http://www.genome.jp/kegg/pathway/map/map00020.html>

KEGG/PATHWAY: Generalized TCA Cycle



ENZYME COMMISSION NUMBERS

EC numbers (Enzyme Commission numbers) are a numerical classification scheme for enzymes based on the chemical reactions they catalyze. According to IUPAC-IUBMB every enzyme code consists of the letters "EC" followed by four numbers separated by periods ('.'). Those numbers represent a progressively finer classification of the enzyme.

The first number shows to which of the six main divisions (classes) the enzyme belongs.

The second number indicates the subclass.

The third one gives the sub-subclass.

The fourth one is the serial number of the enzyme in its sub-subclass.

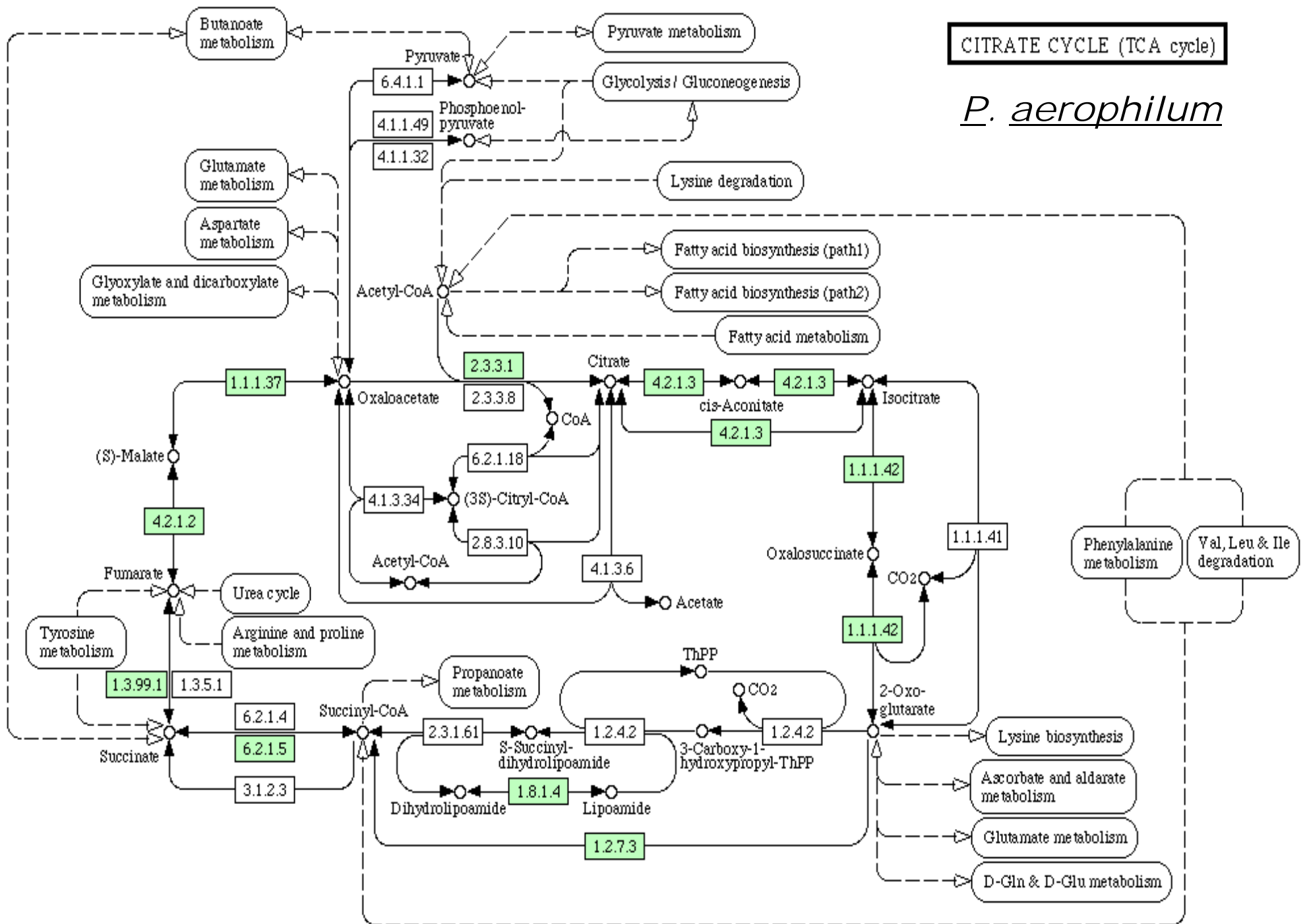
ENZYME GRAPH?

An enzyme graph derived from a metabolic pathway can be represented as $G = (V, E)$

- V is the set of nodes representing EC numbers**
- $E \subseteq V \times V$ is the set of edges representing connections between two EC numbers of two successive reactions**

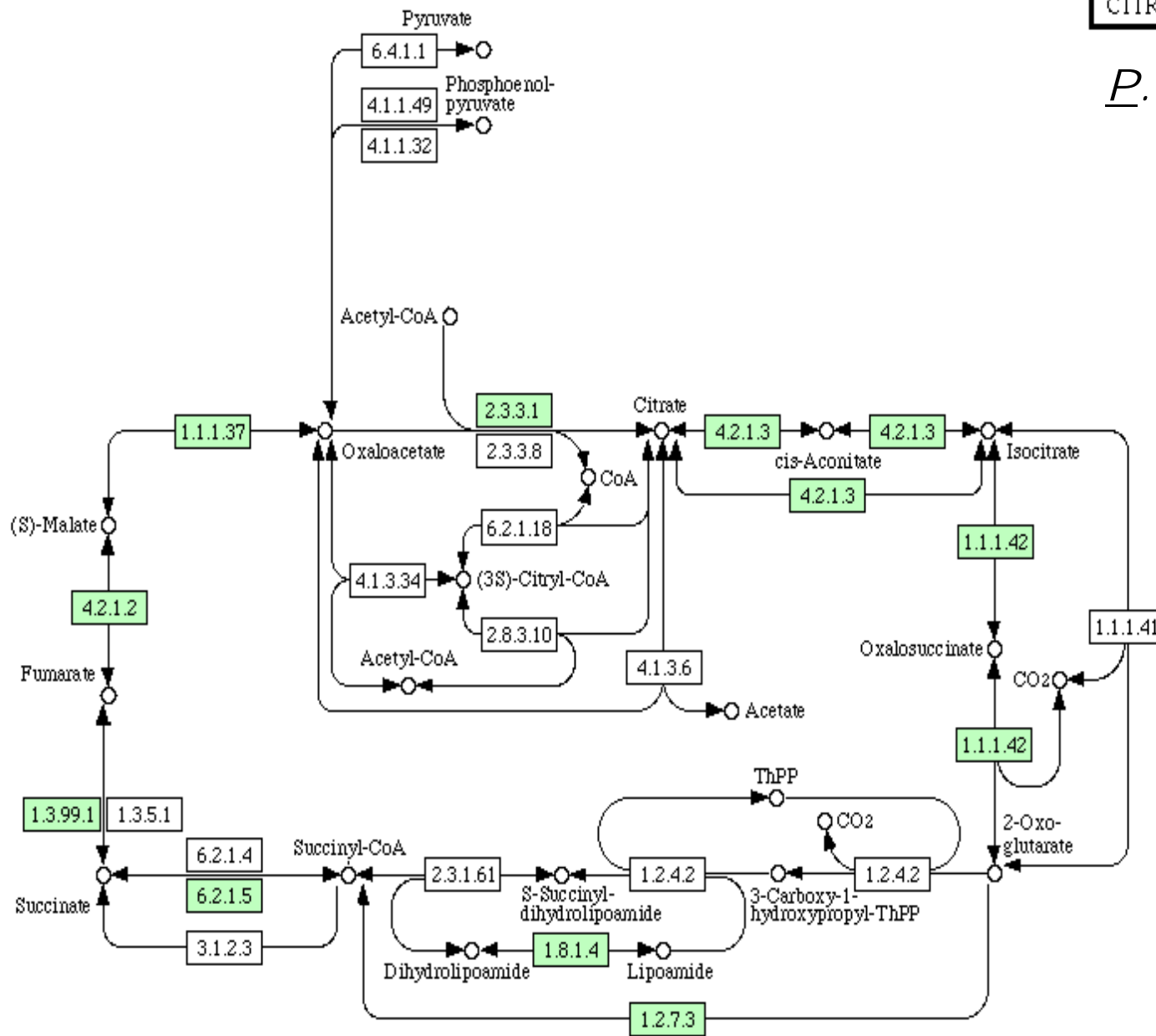
CITRATE CYCLE (TCA cycle)

P. aerophilum



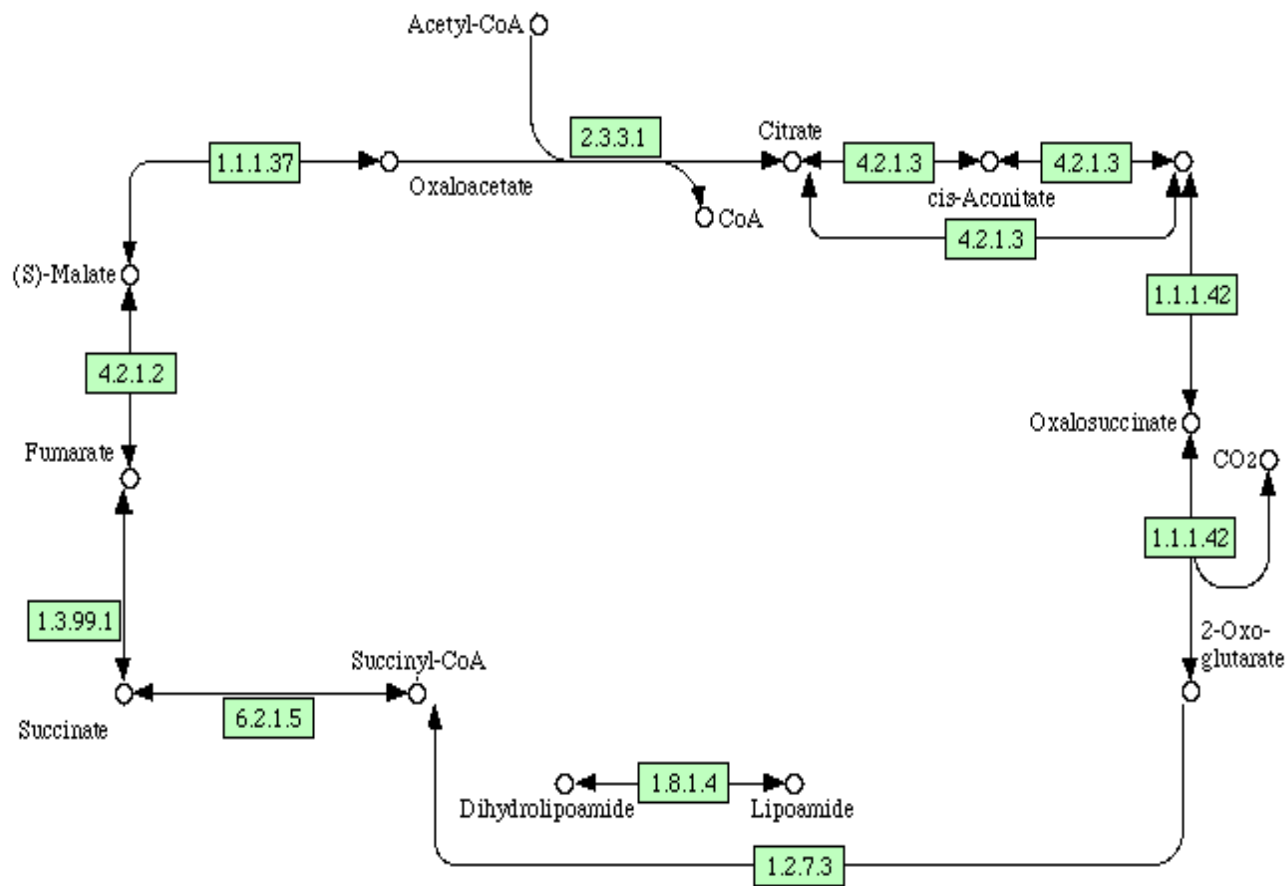
CITRATE CYCLE (TCA cycle)

P. aerophilum



CITRATE CYCLE (TCA cycle)

P. aerophilum



Symbols For Metabolites

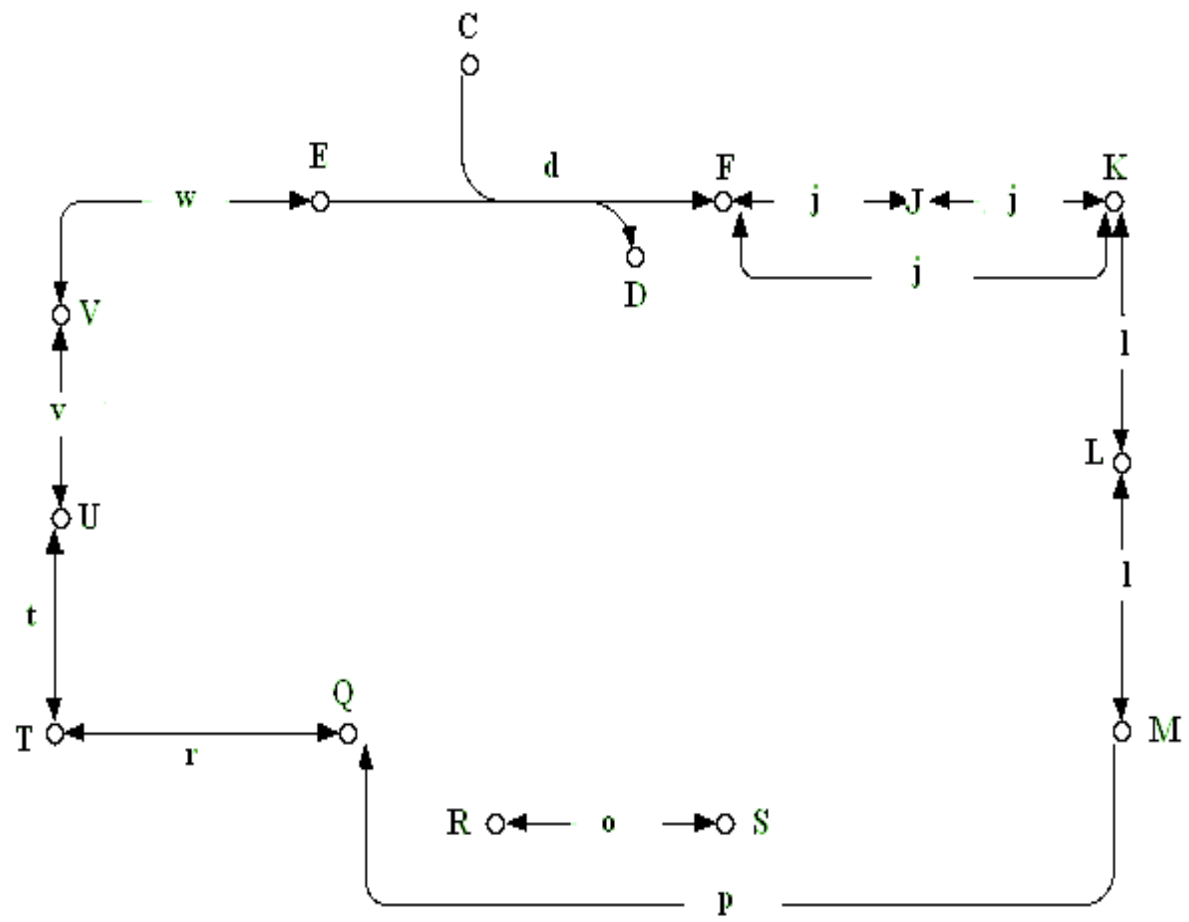
A... Pyruvate,	M... 2-Oxo-glutarate,
B... Phosphoenol pyruvate,	N... 3-carboxy-1-
C... Acetyl Co A,	hydroxypropyl-Thpp
D... Co A,	O... S-succinyl Co A
E... Oxaloacetate,	P... ThPP
F... Citrate,	Q... Succinyl Co A,
G... (3S)-Citryl Co A,	R... Dihydrolipoamide,
I... Acetate,	S... Lipoamide,
J... Cisaconitate,	T... Succinate,
K... Isocitrate,	U... Fumerate,
L... Oxaloacetate,	V... (S)-Malate.

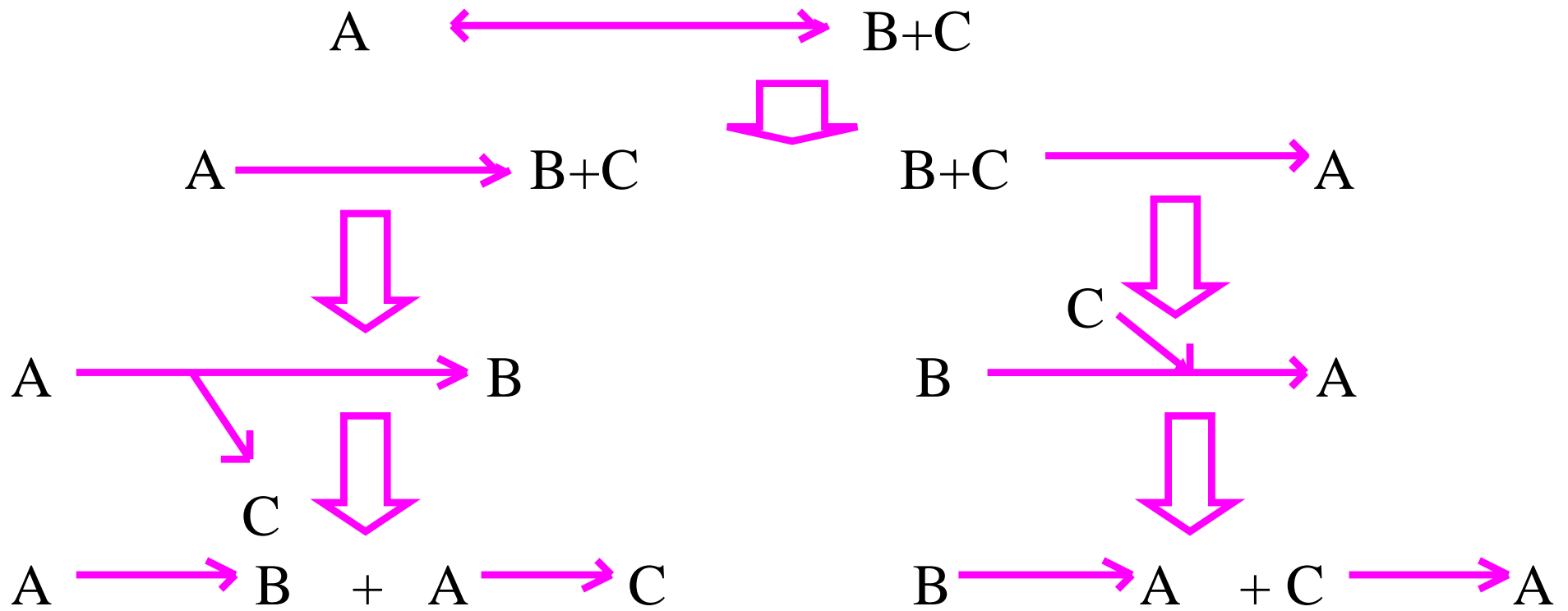
Symbols For EC Numbers

a... 6.4.1.1	l... 1.1.1.42
b... 4.1.1.49	m... 1.2.4.2
c... 4.1.1.32	n... 2.3.1.61
d... 2.3.3.1	o... 1.8.1.4
e... 2.3.3.8	p... 1.2.7.3
f... 4.1.3.34	q... 6.2.1.4
g... 2.8.3.10	r... 6.2.1.5
h... 6.2.1.18	s... 3.1.2.3
i... 4.1.3.6	t... 1.3.99.1
j... 4.2.1.3	u... 1.3.5.1
k... 1.1.1.41	v... 4.2.1.2
	w... 1.1.1.37

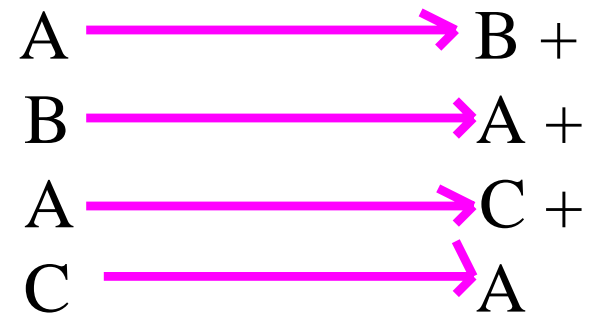
CITRATE CYCLE (TCA cycle)

Pyrobaculum aerophilum





So $A \longleftrightarrow B + C$



By no means B and C are connected.

(Individual Member Reactions of TCA)

1. C \longrightarrow D(d)

2. C \longrightarrow F(d)

3. E \longrightarrow F(d)

4. F \longrightarrow J(j)

5. J \longrightarrow F(j)

6. F \longrightarrow K(j)

7. K \longrightarrow F(j)

8. J \longrightarrow K(j)

9. K \longrightarrow J(j)

10. K \longrightarrow L(l)

11. L \longrightarrow K(l)

12. L \longrightarrow M(l)

13. M \longrightarrow L(l)

14. M \longrightarrow Q(p)

15. R \longrightarrow S(o)

16. S \longrightarrow R(o)

17. Q \longrightarrow T(r)

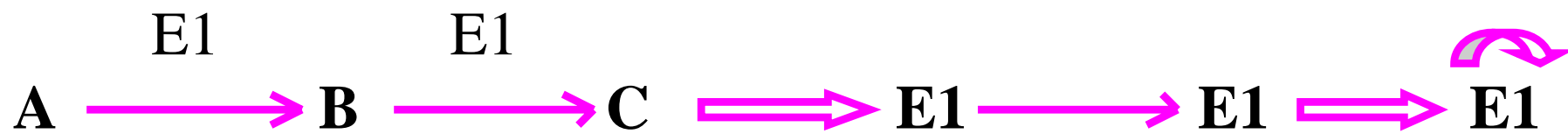
18. T \longrightarrow Q(r)

19. T \longrightarrow U(t)

20. U \longrightarrow T(t)

21. U \longrightarrow V(v)

22. V \longrightarrow U(v)

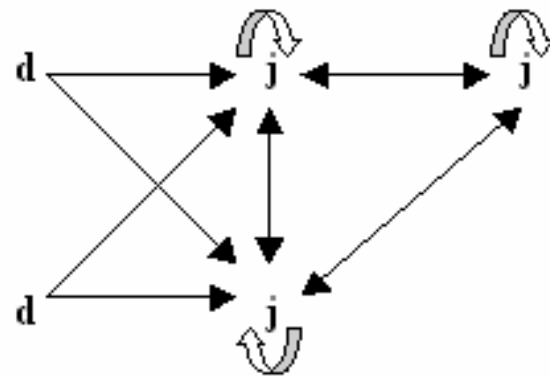


d

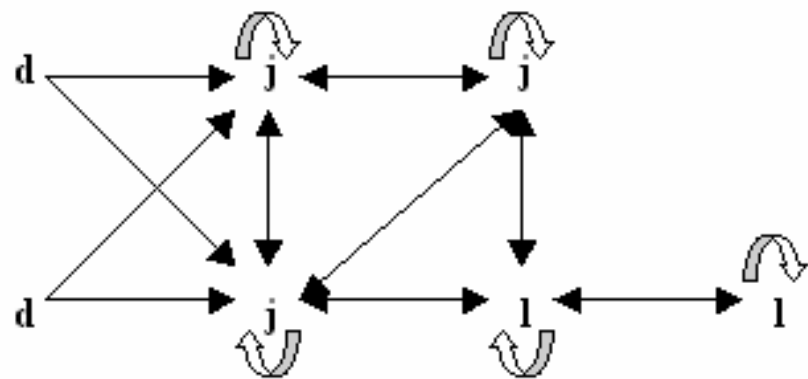
d

d

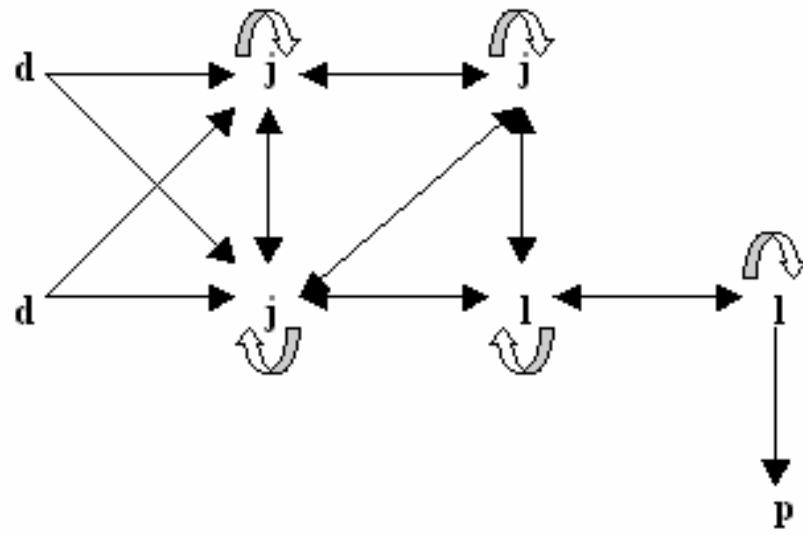
d



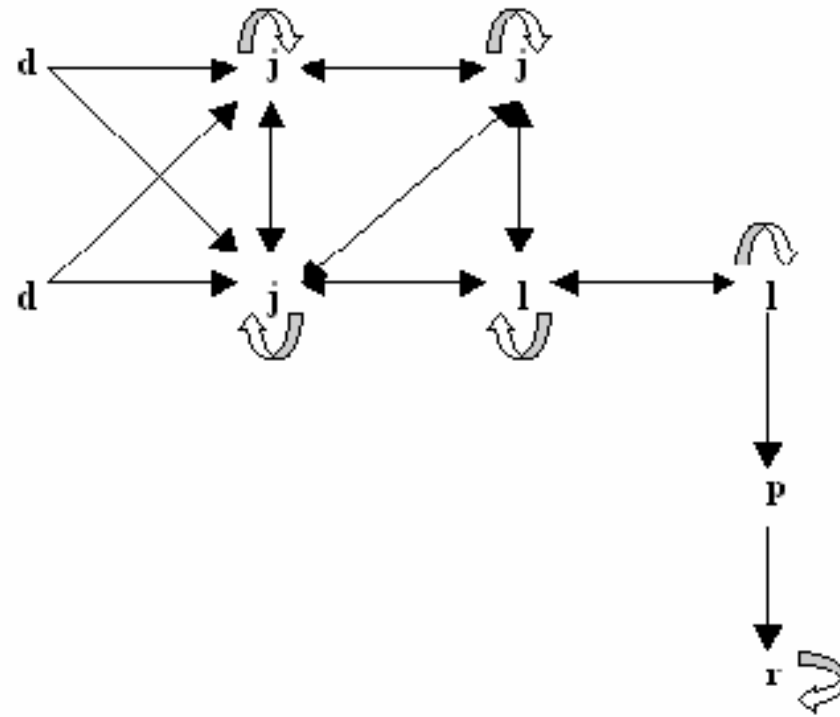
d



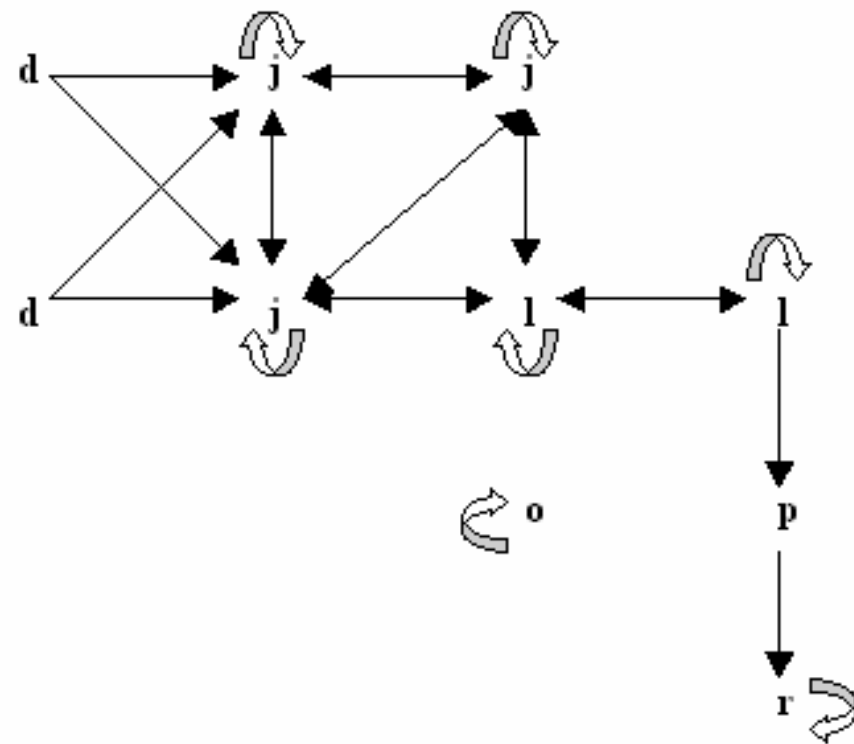
d



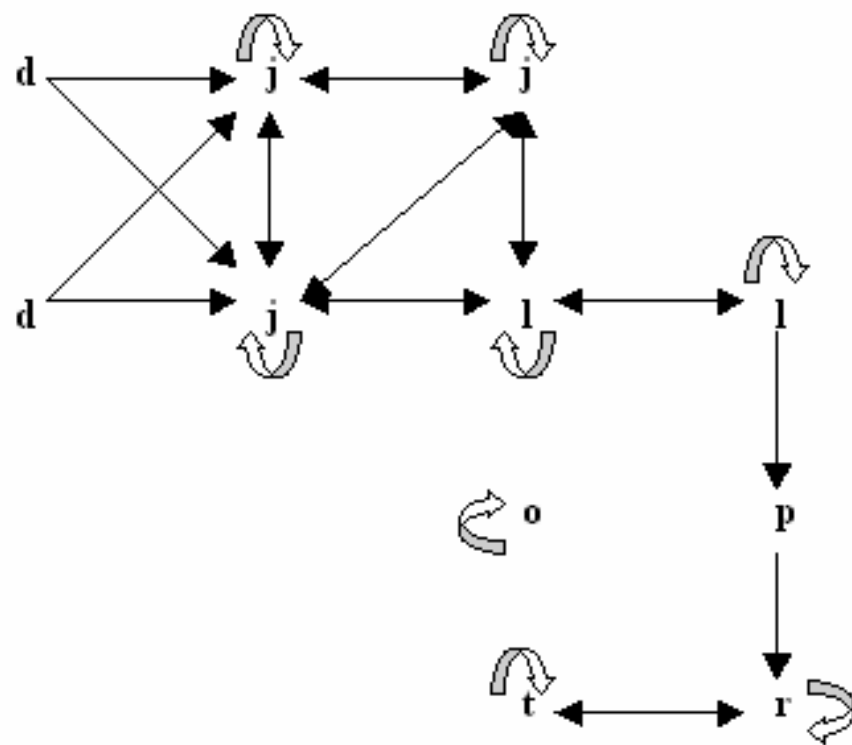
d



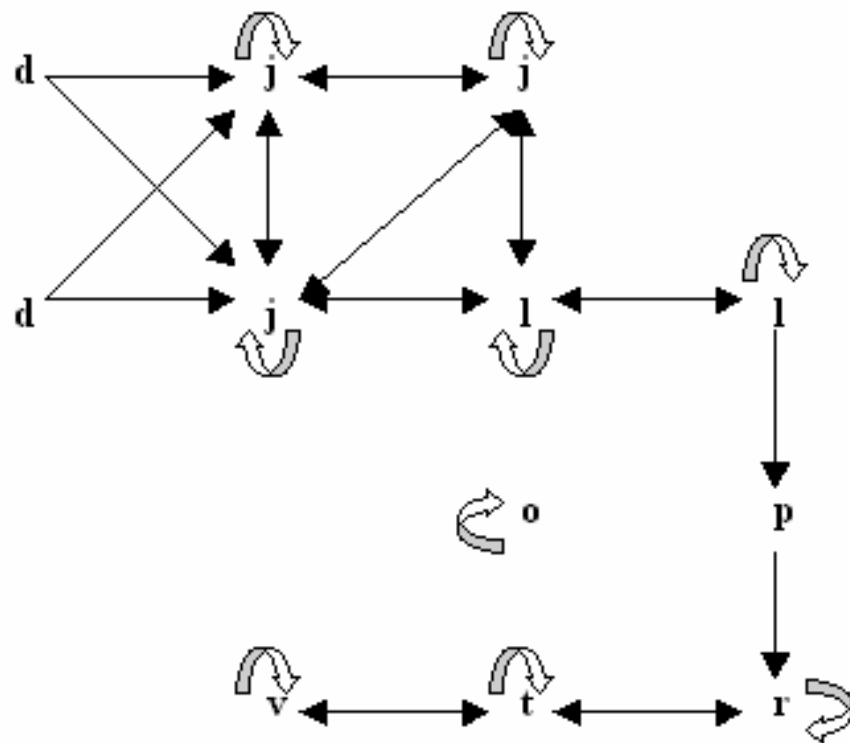
d



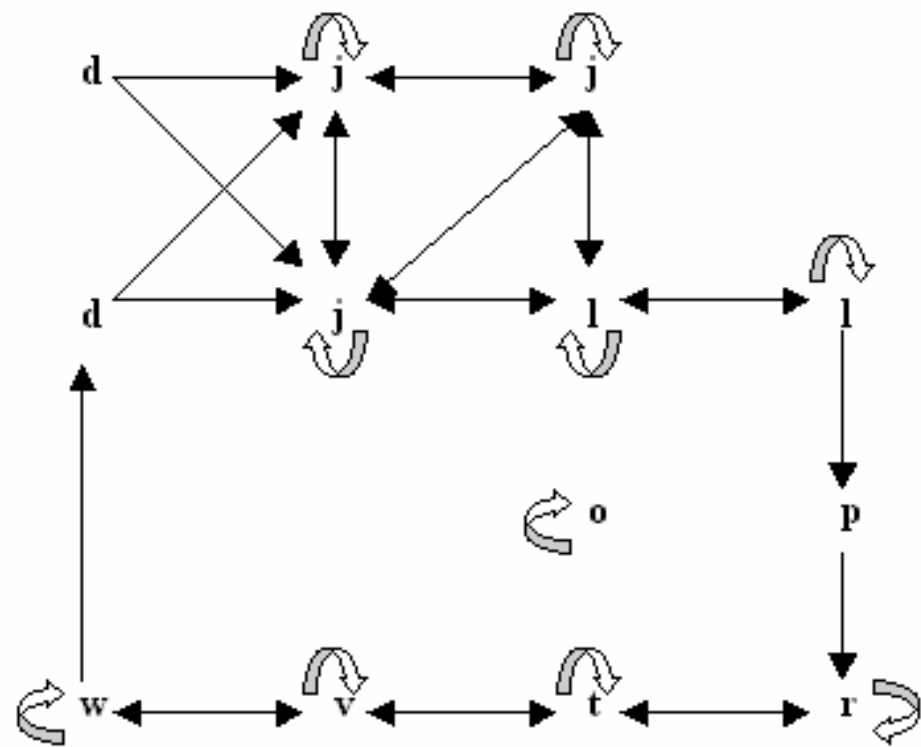
d

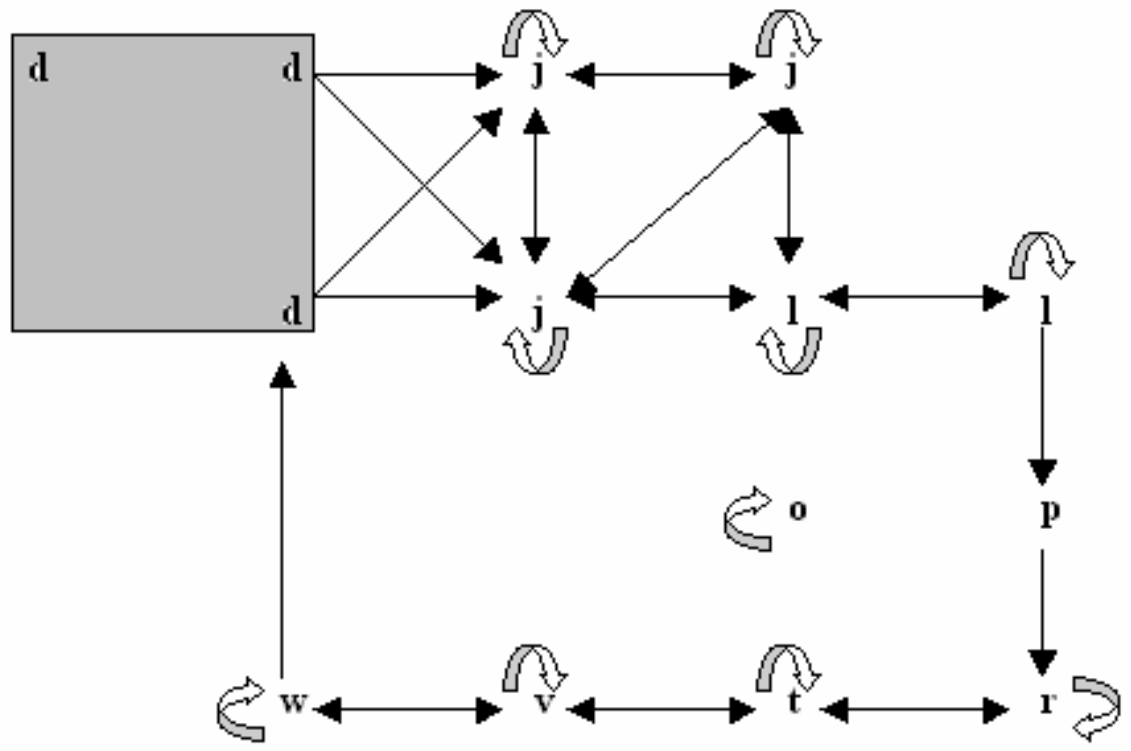


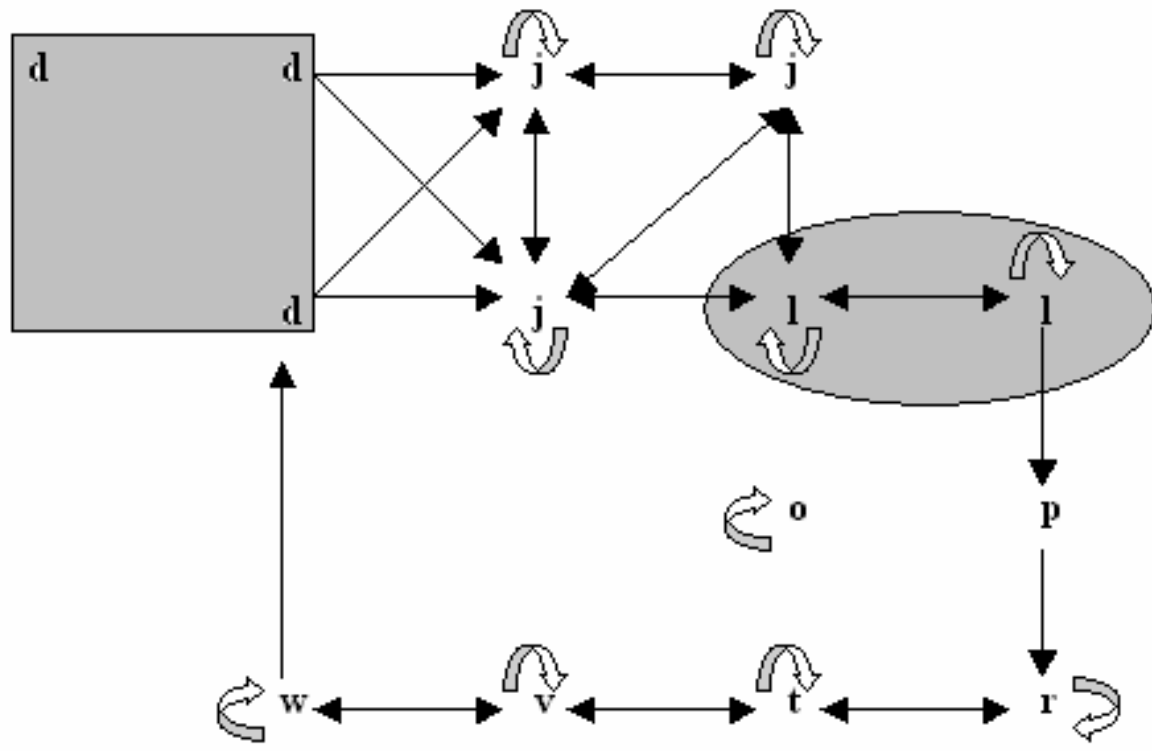
d

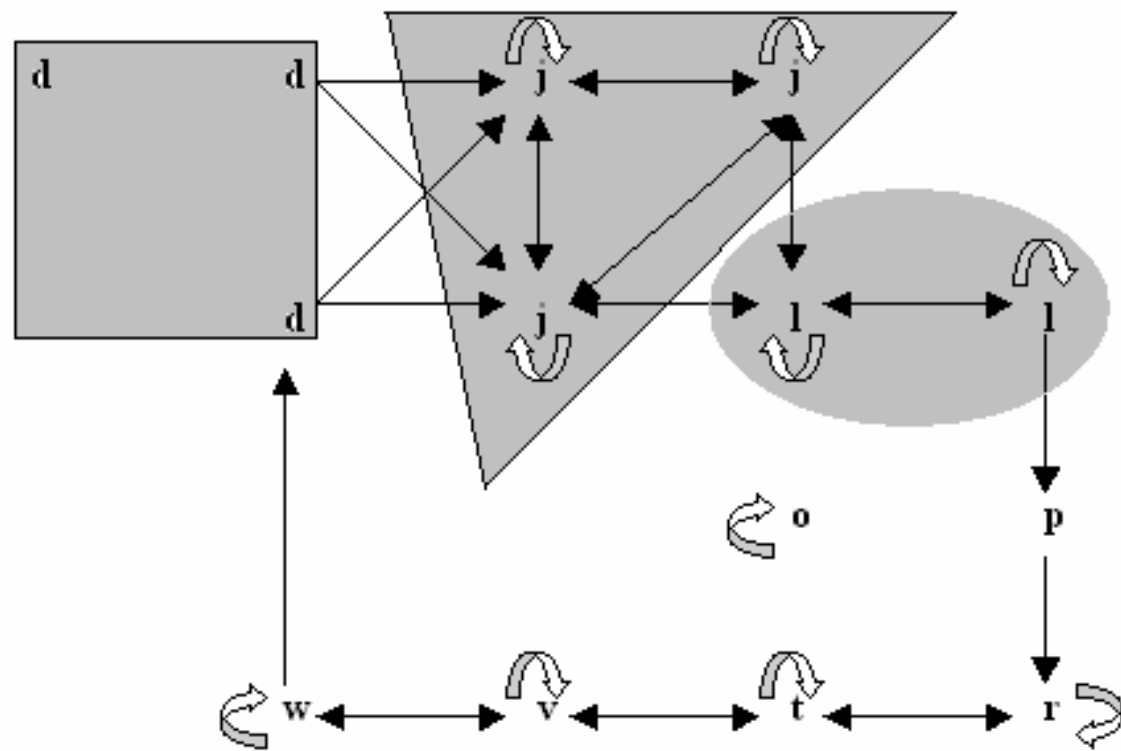


d

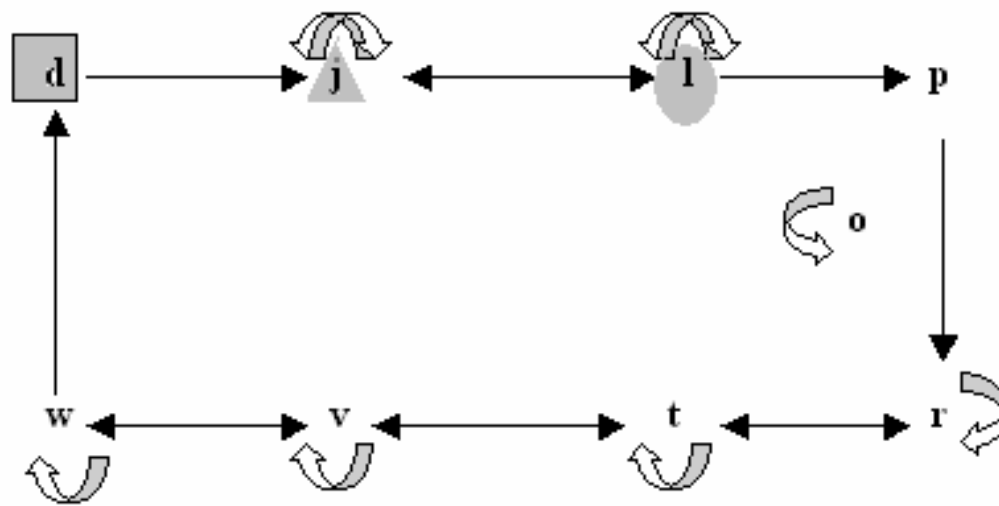




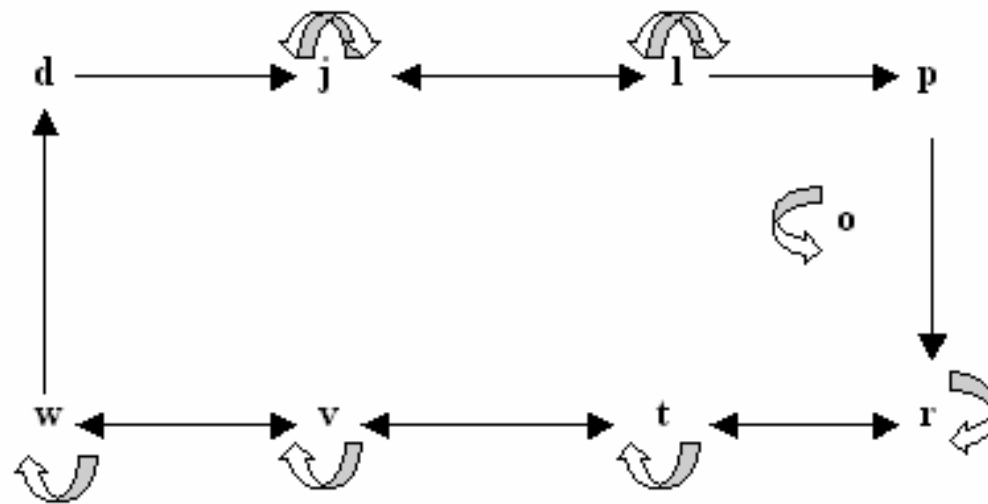




TCA Cycle (*Pyrobaculum aerophilum*)



TCA Cycle (*Pyrobaculum aerophilum*)



CONVERSION OF GRAPHS TO ADJACENCY MATRICES

If graph G has n vertices, then the adjacency matrix is an $n \times n$ matrix A defined as

$$\begin{aligned} A(x, y) &= 1, \text{ if } x \longrightarrow y \text{ in } G \\ &= 0, \text{ otherwise} \end{aligned}$$

*Adjacency matrix representing TCA Cycle
enzyme graph of P. aerophilum*

x \ y	d	j	l	o	p	r	t	v	w
d	0	1	0	0	0	0	0	0	0
j	0	1	1	0	0	0	0	0	0
l	0	1	1	0	1	0	0	0	0
o	0	0	0	1	0	0	0	0	0
p	0	0	0	0	0	1	0	0	0
r	0	0	0	0	0	1	1	0	0
t	0	0	0	0	0	1	1	1	0
v	0	0	0	0	0	0	1	1	1
w	1	0	0	0	0	0	0	1	1

*Here a to z
alphabets are
taken as
abbreviation
of nodes
which are
actually
enzyme
commission
numbers*

Two adjacency matrices representing two metabolic pathways is our input data for graph comparison

*The algorithm we follow is a modification of that by **Heymans and Singh** (Bioinformatics, 2003, vol. 19, suppl. 2, pages i138- i146)*

Instead of Bipartite graph matching we simply took the pair of nodes (x,y) having maximum similarity, where $x \in G_1$ and $y \in G_2$

Initialization:

$$S_0(a, b) = \text{Sim}(a, b)$$

Iterative step:

$$S_{(k+1)}(a, b) = ((A_{k1}(a, b) + A_{k2}(a, b) + A_{k3}(a, b) + A_{k4}(a, b) - D_{k1}(a, b) + D_{k2}(a, b) + D_{k3}(a, b) + D_{k4}(a, b)) / 4) \times \text{Sim}(a, b)$$

Normalization:

$$S \leftarrow \frac{S}{\|S\|_2}$$

Computing similarity between two graphs:

$$S_{(G1, G2)} = \frac{\sum_{a \in G1, b \in G2, M(a,b)=1} S(a, b)}{\sqrt{n1 \cdot n2}}$$

Species we have considered

Y. pestis

A. sp. CC7120

X. campestris

M. janaschii

P. luminescens

A. fulgidus

M. leprae

T. acidophilum

H. influenzae

C. jejuni

S. solfataricus

M. mazei

R. norvegicus

P. furiosus

E. coli K12MG

C. acetobutylicum

M. tuberculosis

C. pneumonia

M. musculus

C. muridarum

H. sapiens

W. brevipalpis

C. elegans

P. horikoshii

A. aeolicus

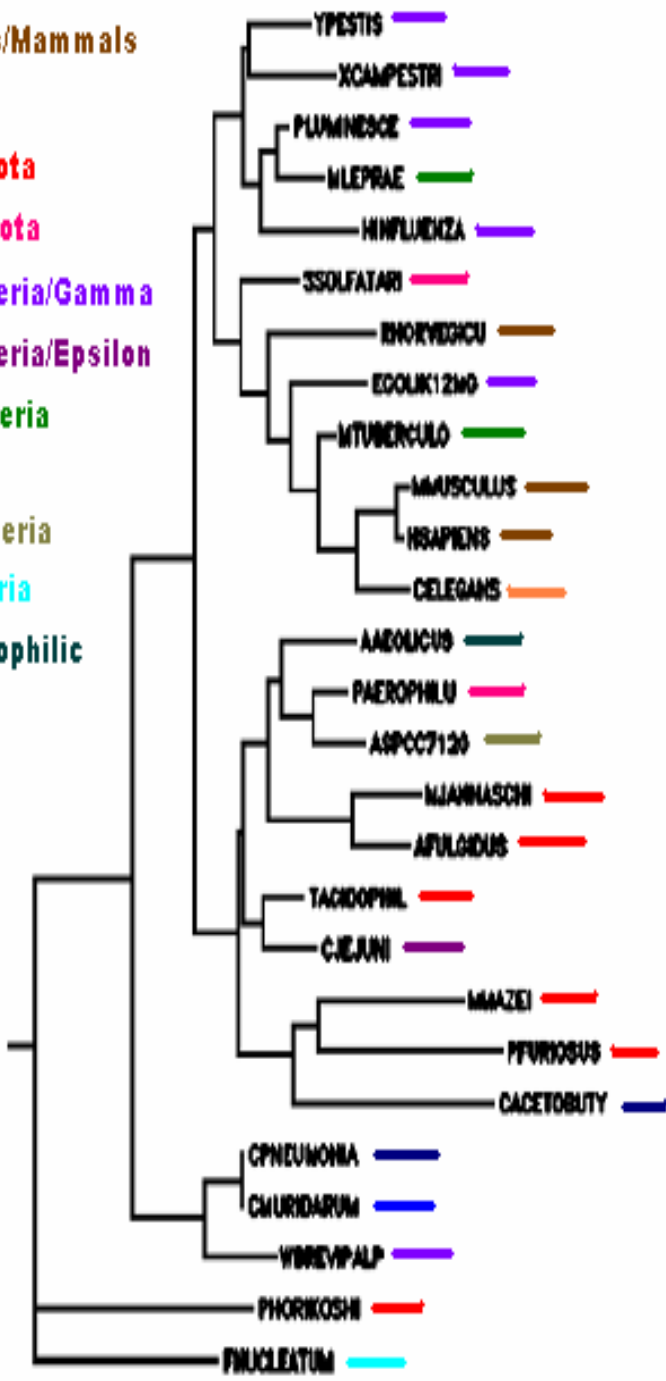
F. nucleatum

P. aerophilum

Phylogenetic Tree of Life

- Animals/Vertebrates/Mammals
- Animals/Nematodes
- Archaea/Euryarchaeota
- Archaea/Crenarchaeota
- Bacteria/Proteobacteria/Gamma
- Bacteria/Proteobacteria/Epsilon
- Bacteria/Actinobacteria
- Bacteria/Chlamydia
- Bacteria/Cyanobacteria
- Bacteria/Fusobacteria
- Bacteria/Hypothermophilic bacteria
- Bacteria/Firmicutes

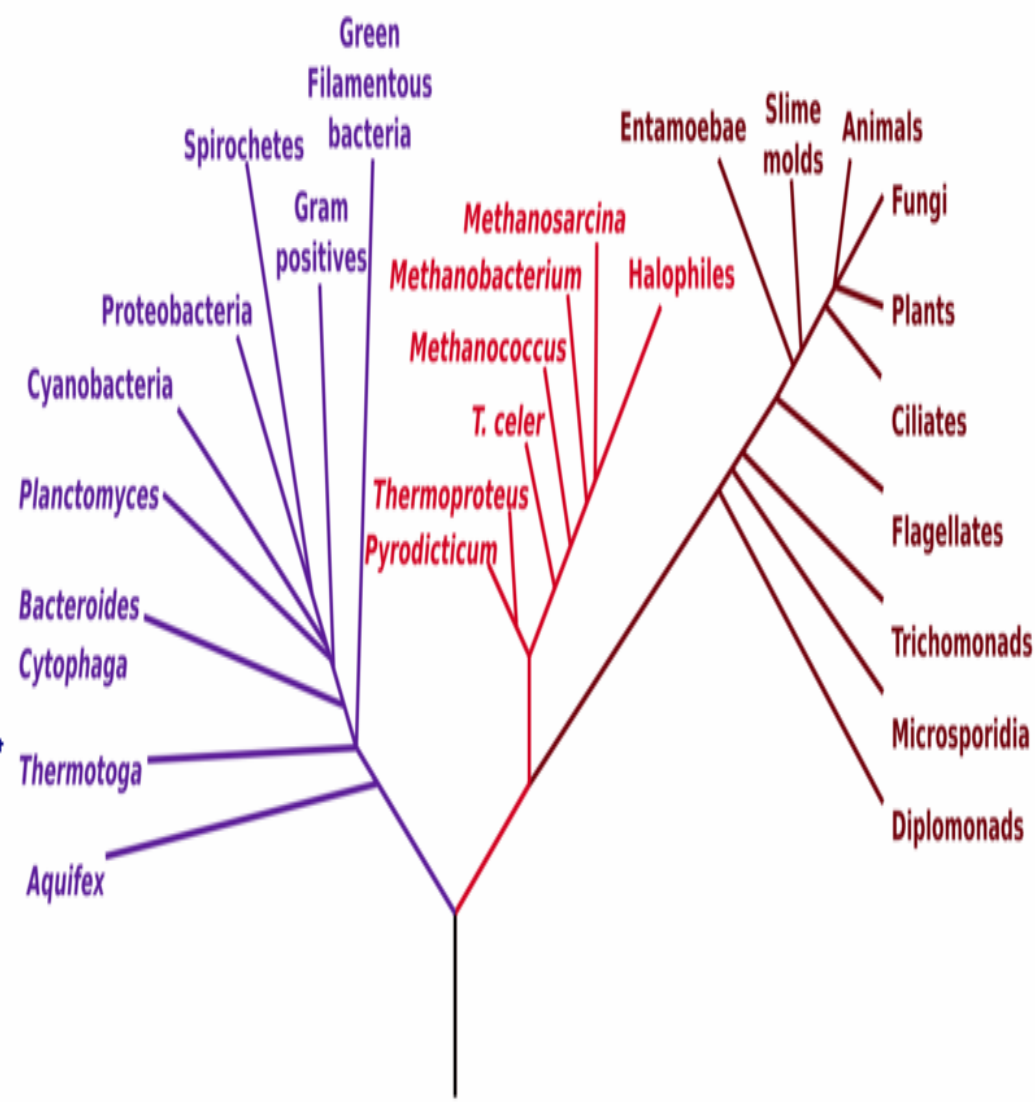
TCA Cycle Tree



Bacteria

Archaea

Eucarya



*TCA cycle tree is not in
accordance with Phylogenetic
tree of life.*

*From evolution point of view
closest species may have
different kind of metabolism.*

Some Useful References

1. *Maureen H., Ambuj, S. K. (2003) "Deriving phylogenetic trees from the similarity analysis of metabolic pathways", Bioinformatics, vol. 19(supl. 2), pp i138- i146*
2. *Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K. F., Itoh, M., Kawashima, S., Katayama, T., Araki, M. and Hirakawa, M. (2006), "From genomics to chemical genomics: New developments in KEGG", Nucleic Acids Res., 34, pp D354-357*
3. *Clemente, J., C., Kenji, S., Valiente, G. (2005), "Reconstruction of Phylogenetic Relationships from Metabolic Pathways Based on the Enzyme Hierarchy and the Gene Ontology", Genome Informatics 16(2), pp 45-55*

Thank you!