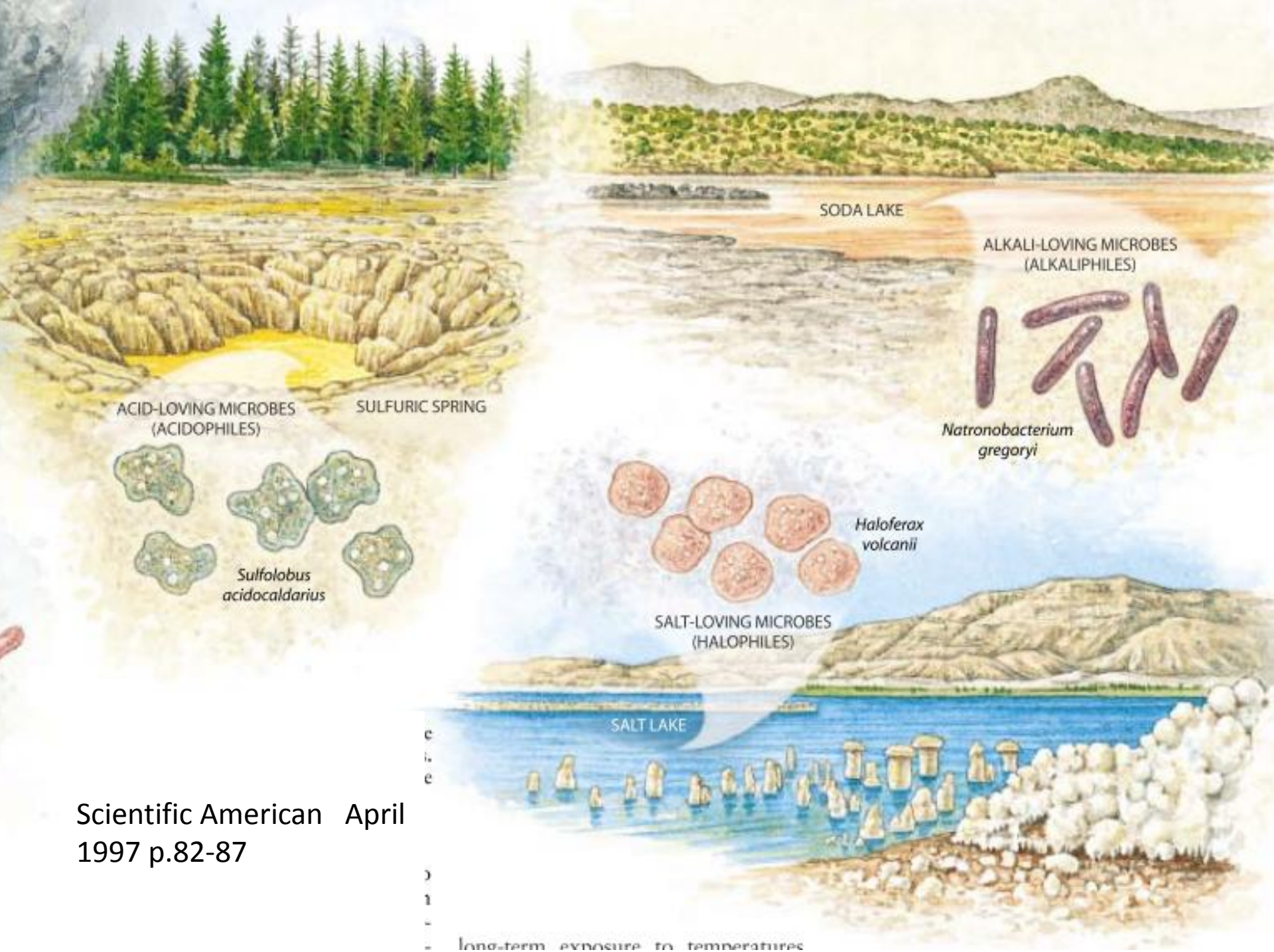
Correlation of metabolic network structure of extremophiles to optimal growth temperature

C Suguna
Centre for Cellular and Molecular Biology, Hyderabad

ICMSat , Hyderabad
16-18 Aug. 2010



Scientific American April 1997 p.82-87

long-term exposure to temperatures

Types of Extremophiles

- Acidophile- Low pH values
- Alkaliphile - High pH values.
- Anaerobe - Absence of oxygen.
- Endolith - Inside rocks or grains.
- Halophile - High salt concentrations.
- Oligotroph - limited nutrient conditions
- Barophiles - High hydrostatic pressure.
- Xerophile - Low water activity
- Psychrophiles – Low temperatures
- Thermophiles & hyperthermophiles – High temperatures

Optimal Growth Temperature

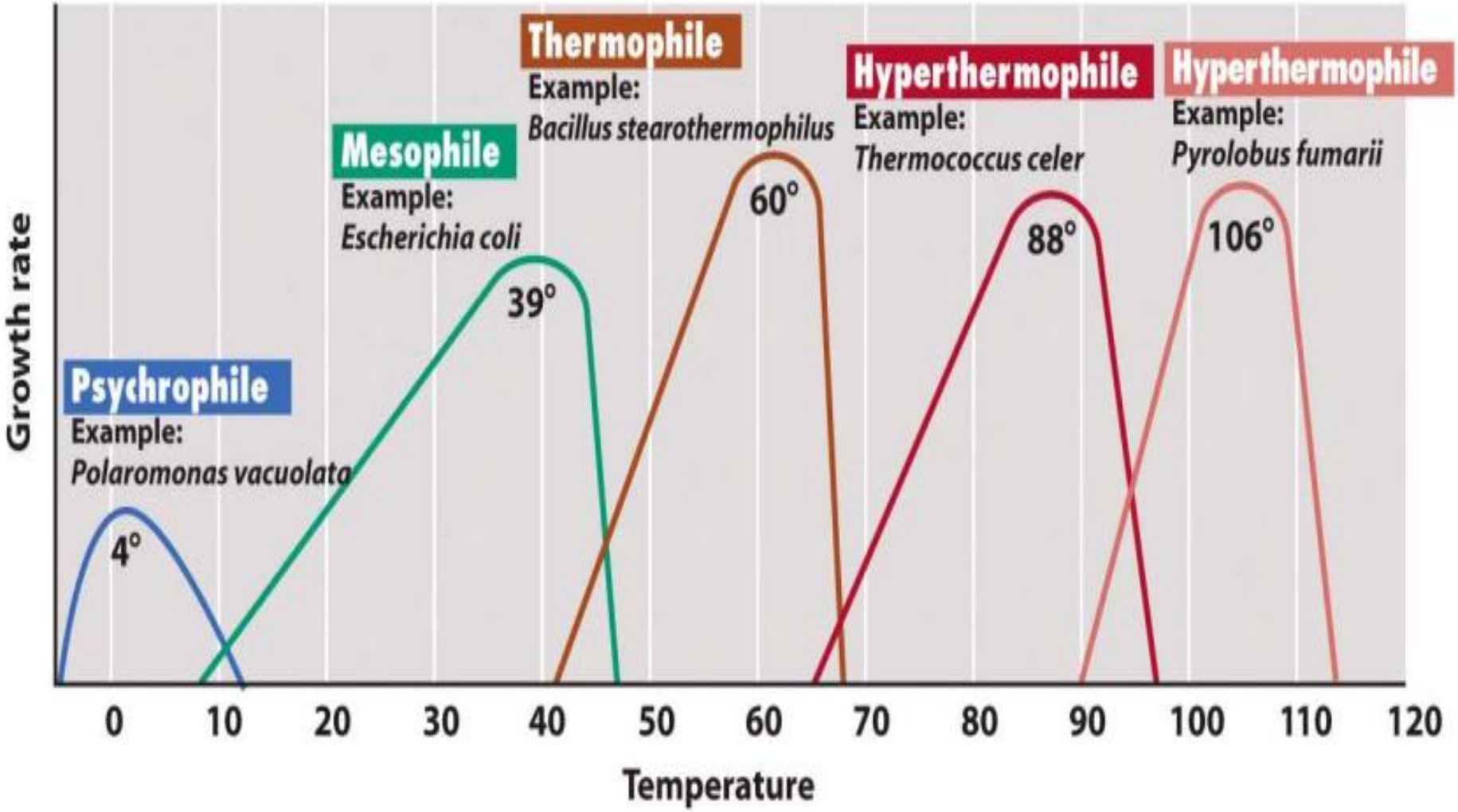


Figure 6-17 Brock Biology of Microorganisms 11/e

NCBI classification based on life-style

1. Aquatic - live in fresh or seawater and not associated with hosts.
2. Host Associated
   a) Obligate - associated with host (intracellular/ extracellular) has little contact with outside world.
   b) Facultative - free living bacteria such as *E. coli* that often associate with a host.
3. Multiple - multiple different kinds of environments Ex: wide host range or different environments
4. Specialized - live in specialized environments ex: marine thermal vents.
5. Terrestrial - live in the soil

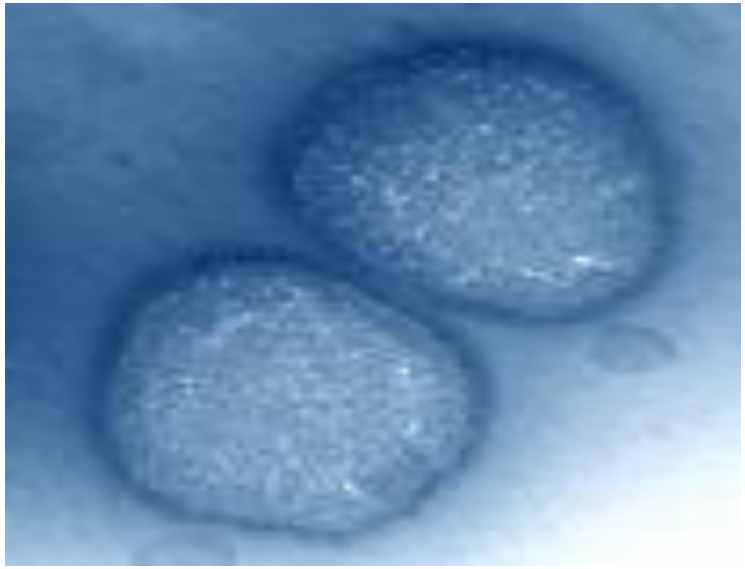
Organisms studied

165 prokaryotes (includes bacteria & archea) living in varying environmental conditions

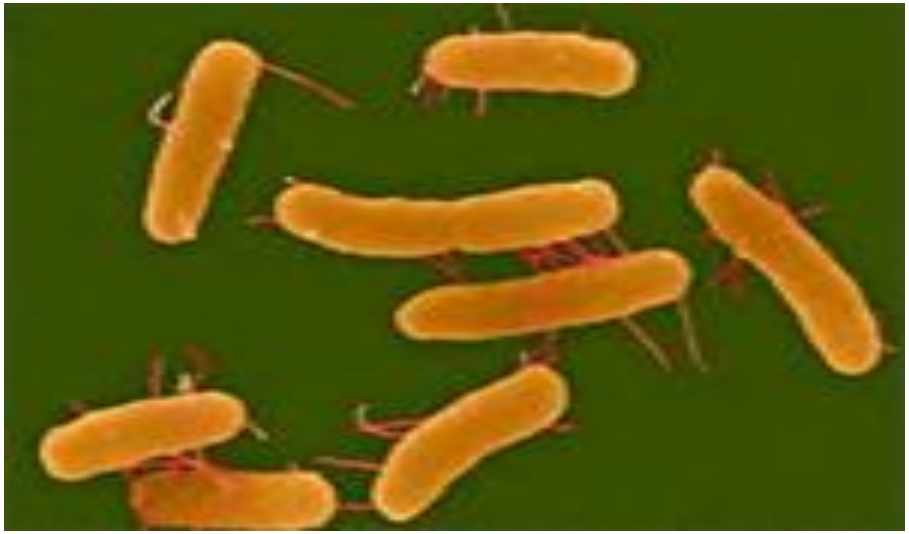
Life style	Number of Organisms	Range of optimal growth temperature			
		Psychro	Meso	Thermo	Hyperthermo
A	18	2	8	4	4
HA	50	2	47	0	1
M	50	1	43	2	0
S	31	3	7	15	6
T	16	0	14	4	2
Total	165	8	119	25	13

Construct metabolic network from genome information & analyse using graph theory

Methanococcoides burtonii



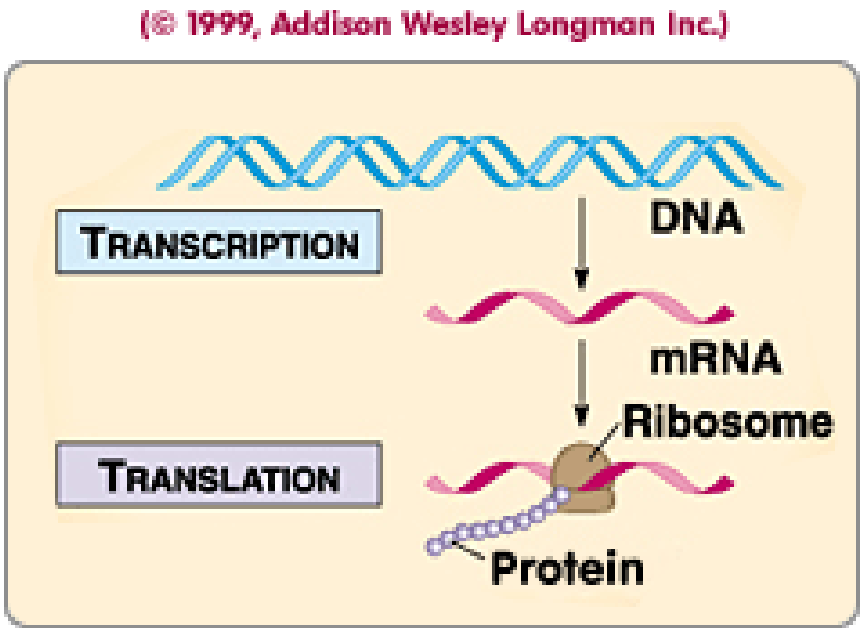
Salmonella typhimurium



- Irrespective of the shape, size or environmental conditions all organisms are optimized to function robustly and efficiently.
- Aim to understand the factors which give rise to cellular economy & robustness in presence of noise and wide variety of environmental conditions.

Control and Optimzation in Cells

Pathways controlled through enzymatic regulation, through positive & negative feedback
Only enzymes required for survival synthesized
Only in required amounts and only when required



Structural proteins

•basic building blocks (ex: collagen)

Transmembrane

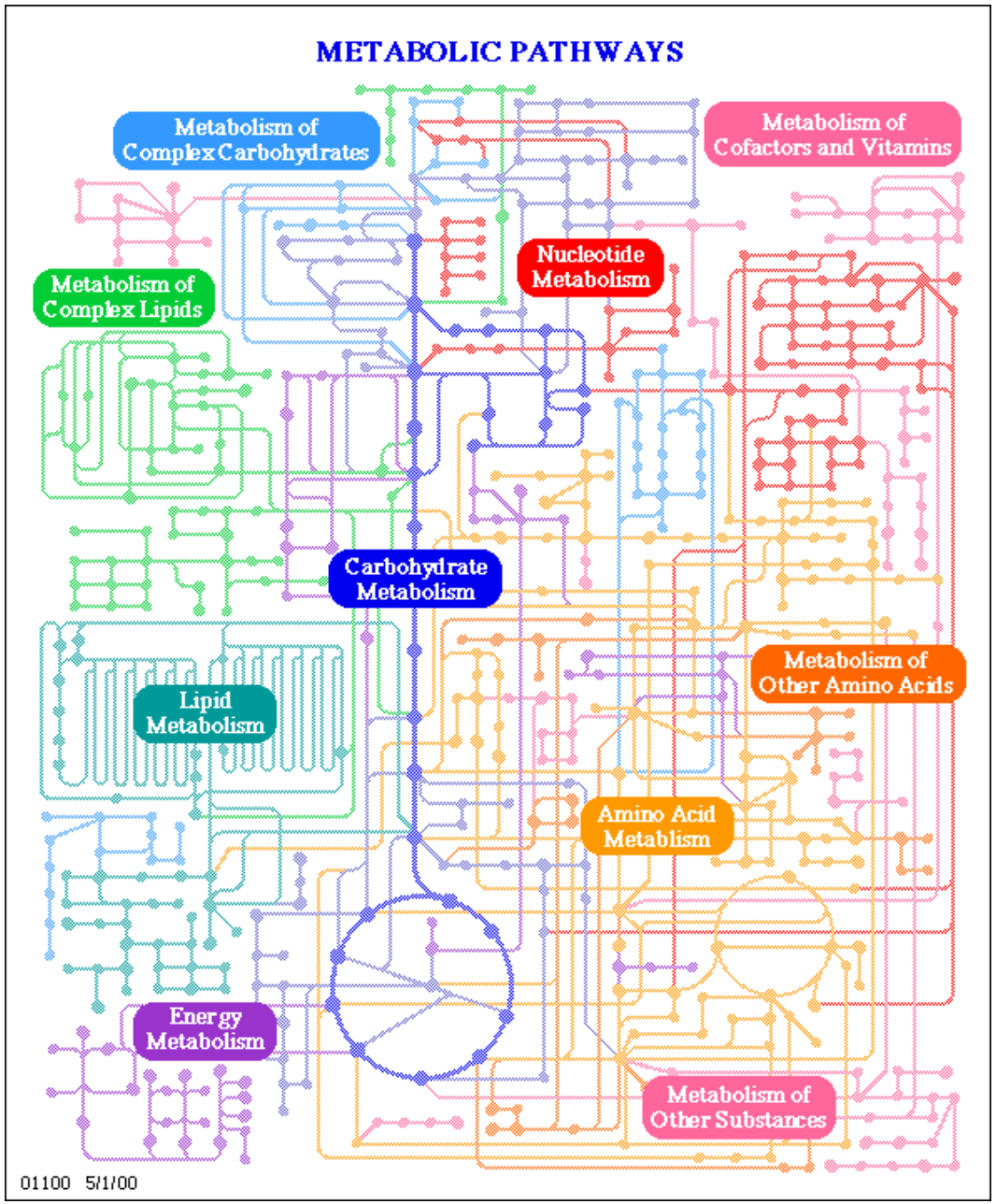
- regulating cell volume,
- generation of ionic gradients

(ex: sodium/potassium pump).

Enzymes

•Catalyze biochemical reactions of different metabolic pathways (ex: glycolysis pathway).
•Very specific and generally catalyze only a single type of reaction.
•Same enzyme can play role in more than one pathway.

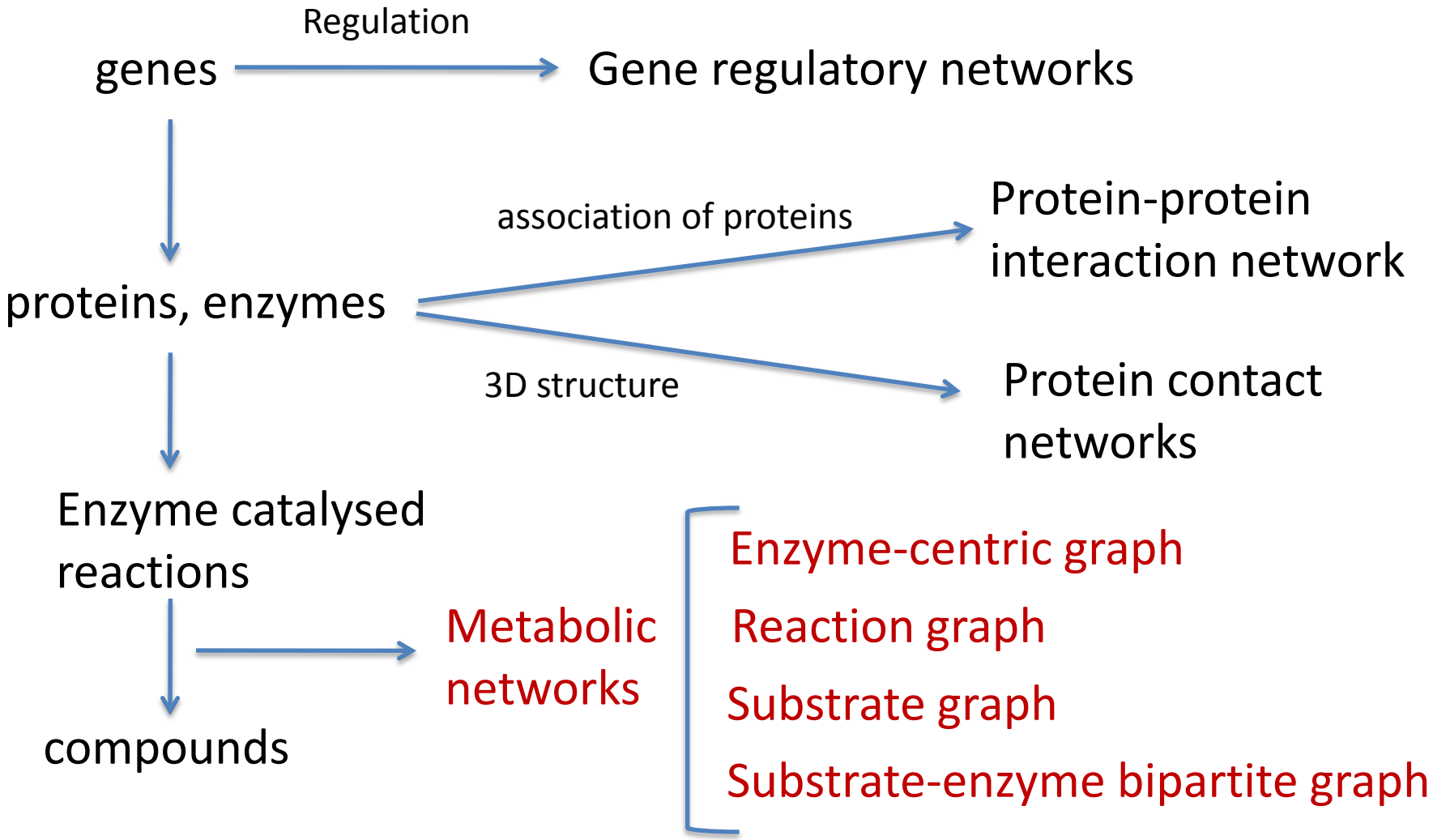
METABOLIC PATHWAYS



1.Carbohydrate Metabolism
2.Lipid Metabolism
3.Nucleotide Metabolism
4.Metabolism of Other Amino acids
5.Energy Metabolism
6. Metabolism of Complex Lipids.
7. Metabolism of Other substances.
8. Metabolism of cofactors and vitamins.
9.Amino Acid Metabolism

Highly interconnected with complex regulatory processes

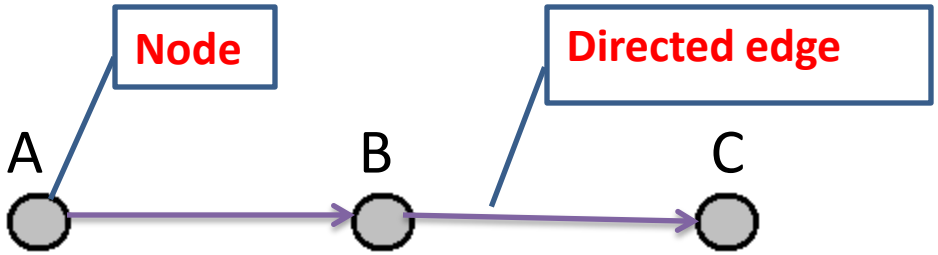
Information in Genome Databases Used to Construct Cellular Networks of Different Types



Graphical Representation of Metabolic Reactions

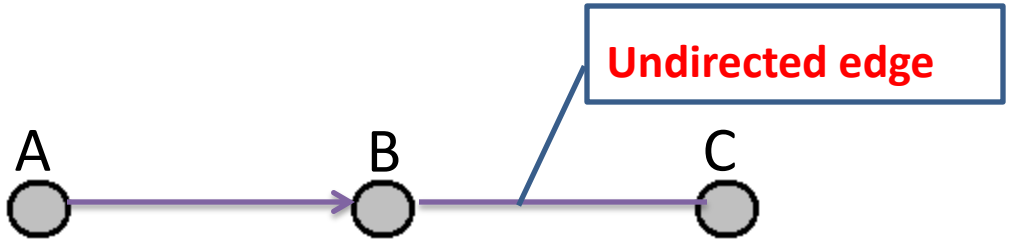
R1: A → B

R2: B → C



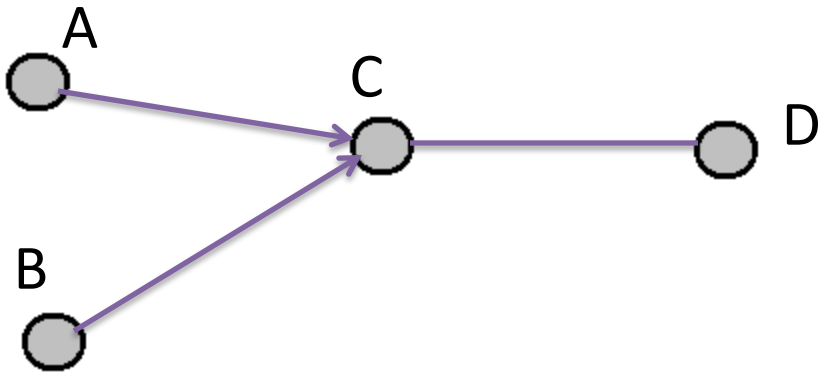
R3: A → B

R4: B ⇌ C



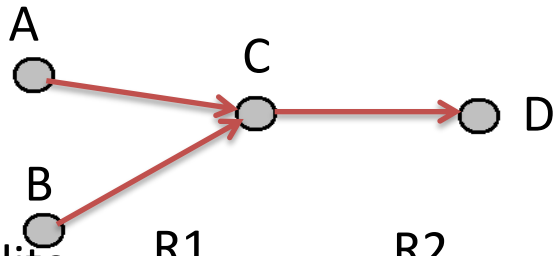
R5: A + B → C

R6: C ↔ D



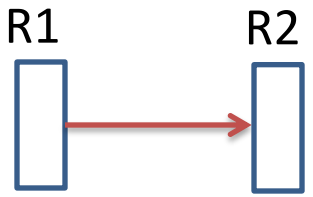
Metabolic Pathway Graph Representations

$$R1: A + B \xrightarrow{E1} C \qquad R2: C \xrightarrow{E2} D$$

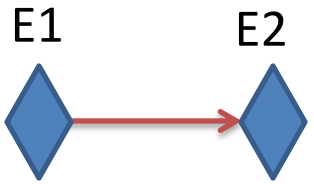


Substrate graph: Nodes – compounds
Edges between compound & product (directed)

Reaction graph: nodes – reactions share common metabolite
Two reactions connected if they share compound that is product of one reaction& substrate of other.

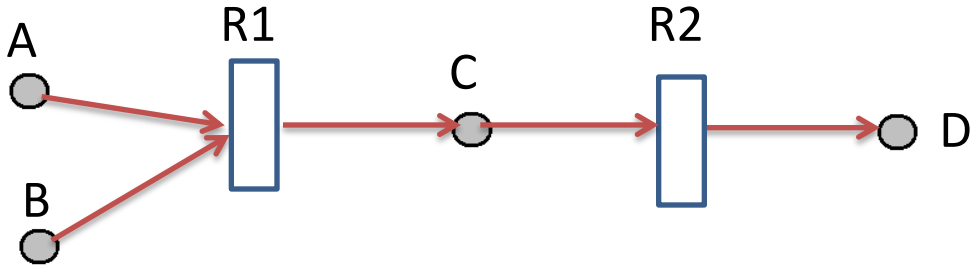


Enzyme-centric graph: nodes - enzymes.
Two enzymes connected if they share compound that is product of one enzyme catalyzed reaction and substrate of reaction catalyzed by the other enzyme.



Substrate-enzyme bipartite graph: two sets of nodes one corresponds to compounds, while the other corresponds to reactions

A reaction is connected with its substrates and products - only nodes in different sets can be connected



DATABASES

1.Kyoto Encyclopedia of Genes and Genomes(KEGG) Rel. 40.0

2.*Hongwu Ma, and* An-Ping Zeng *Bioreaction Database.* ***Reconstruction of metabolic networks from genomeReconstruction of metabolic networks from genome data and analysis of their global structure for various organism, BMC Bioinformatics, 19, 2003, 270(requested)***

KEGG Home
Introduction
Overview
Release notes
Current statistics
KEGG Identifiers
KEGG XML
KEGG API
KEGG FTP
KegTools

3 - Organism

HABITAT

Temperature

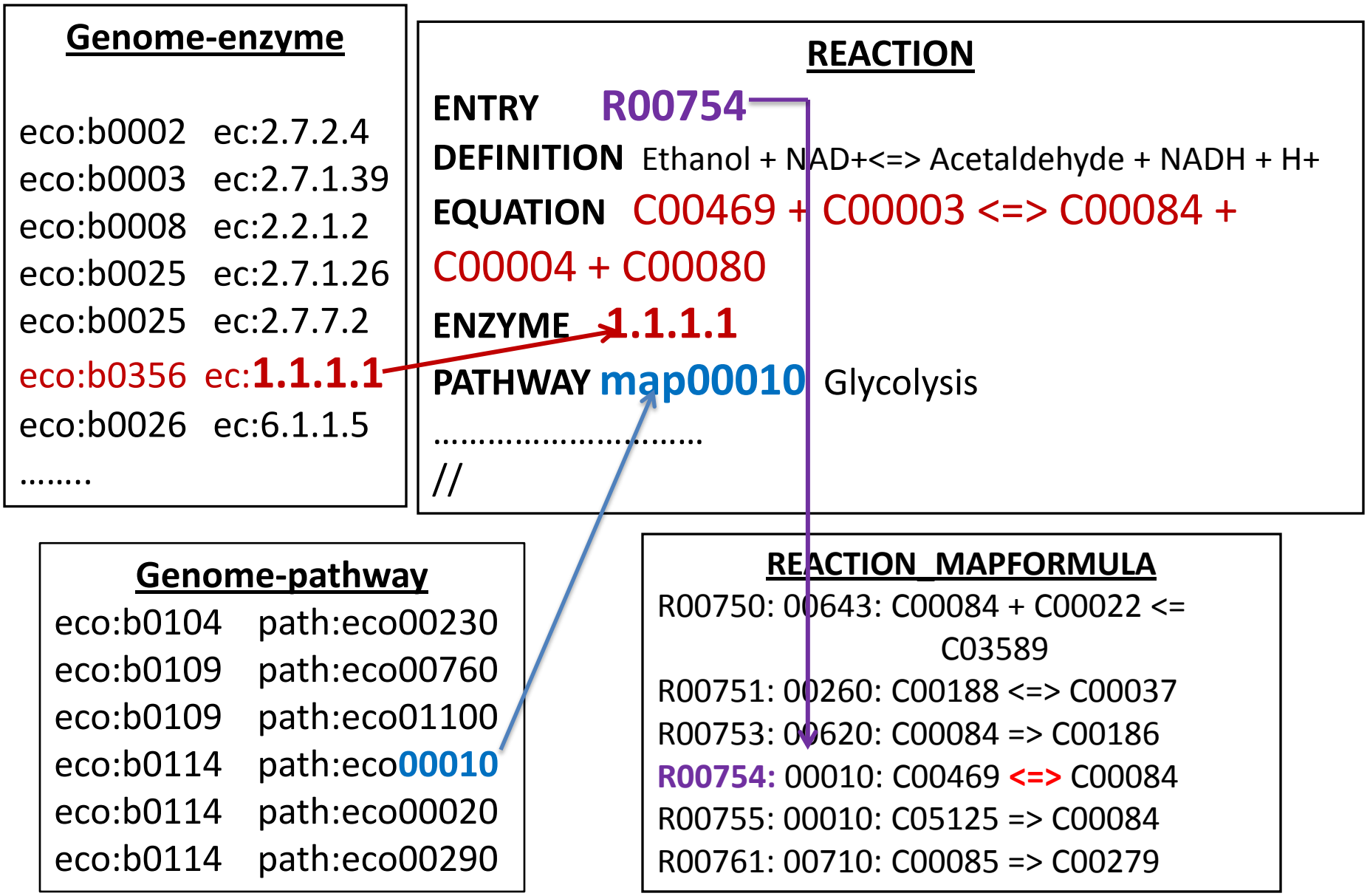
NCBI

PGTdb
Prokaryotic Growth Temperature

GOLD


KEGG Encyclopedia - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://www.genome.jp/kegg/kegg2.html
metacyc
KEGG Encyclopedia
MetaCyc Encyclopedia of Metabolic Pat...
The MetaCyc Database. [Nucleic Acids ...
KEGG - Table of Contents
KEGG2 ATLAS PATHWAY BRITE GENES SSDB LIGAND DBGET
Search KEGG for Go Clear
Category Entry Point Release Info Search & Compute DBGET Search
Systems information
KEGG PATHWAY
KEGG BRITE
New maps
Update status
New hierarchies
Update status
Search objects in pathways
Color objects in pathways
Map relations to hierarchies
KEGG Orthology (KO)
PATHWAY
BRITE
MODULE
DISEASE
Genomic information
KEGG GENES
New organisms
Update status
SSDB search
BLAST search
FASTA search
EGassembler for ESTs
KAAS automatic annotation
GENIES network prediction
ORTHOLOGY
GENES
GENOME
EGENES
VGENES / OGENES
VGENOME
Enzymes & pathways in each organism
Chemical information
KEGG LIGAND
Update status
SIMCOMP compound search
KCaM glycan search
e-zyme reaction prediction
PathComp computation
COMPOUND
DRUG
GLYCAN
REACTION
RPAIR
ENZYME
start
2 Windows Explorer
3 Microsoft Office ...
2 Firefox
Microsoft PowerPoin...
Book1
12:43 PM

Extraction of Enzymatic/Nonenzymatic Reactions



KEGG Data curated using Ma's Bioreaction database(2003) –based on Kegg rel. 20.0

OLD Reaction Id	New Reaction Id
R00007	R00008
R00035	R00068
R00040	R01022

MA's Data used Kegg Ligand database rel 20.0 and at present Kegg release **42.0**

SOME PATHWAY IDS CHANGED
ex: **Map00272** → **Map00270**
(Cysteine and methionine metabolism)
[not mentioned in KEGG]

Changes	No of Reactions
(1) Changed Reaction Id	148
(2) Changed Compound Id	900
(3) Current Metabolites	1200

CURRENCY METABOLITE

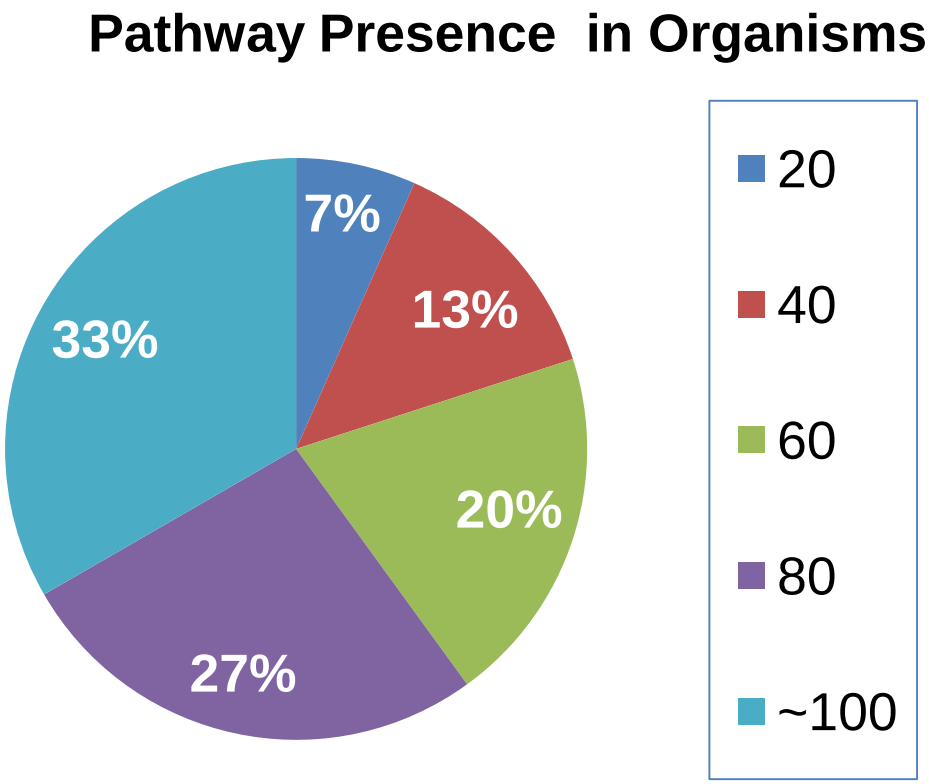
Transfer electrons and other functional groups

- Phosphate group transfer- ATP/ADP/AMP.
- Hydrogen transfer - NADH/NAD, NADPH/NADP, FADH/FAD .
- Acetyl group transfer - Acetyl-Coenzyme-A.
- Others - CO_2,NH_3,O_2,H_2O_2, H_2CO_3 H_2S, H_2SO_3, NO_2, Sulfate etc.

Currency metabolites are abundant in cell and present in many reactions

Remove current metabolites since connections through them produces biologically meaningless results

Results :



Pathways present in almost all organisms

- Glycolysis / Gluconeogenesis
- Pentose phosphate pathway
- Oxidative phosphorylation
- Purine metabolism
- Pyrimidine metabolism
- Cysteine and methionine metabolism
- Valine, leucine and isoleucine biosynthesis

Exception: *Nanoarecheum equitans depends entirely on host for all its nutritional requirements and has only DNA processing pathways*

Reconstructed metabolic network

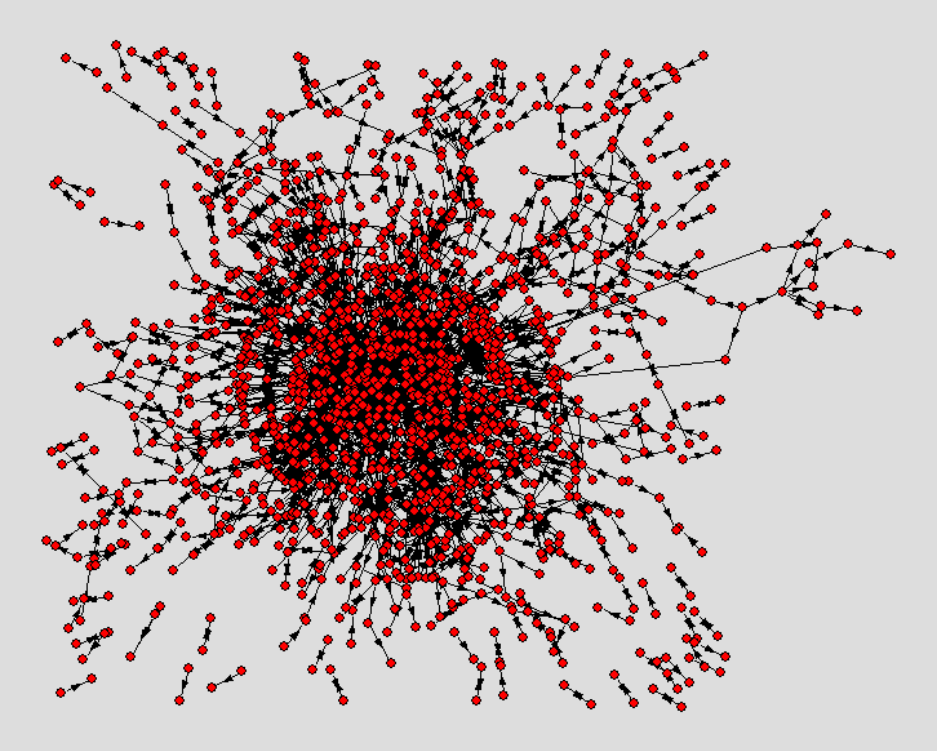
Shewanella baltica

Substrate Graph of *Shewanella baltica* metabolic pathway

Nodes – metabolites

Edges – educt-product pairs

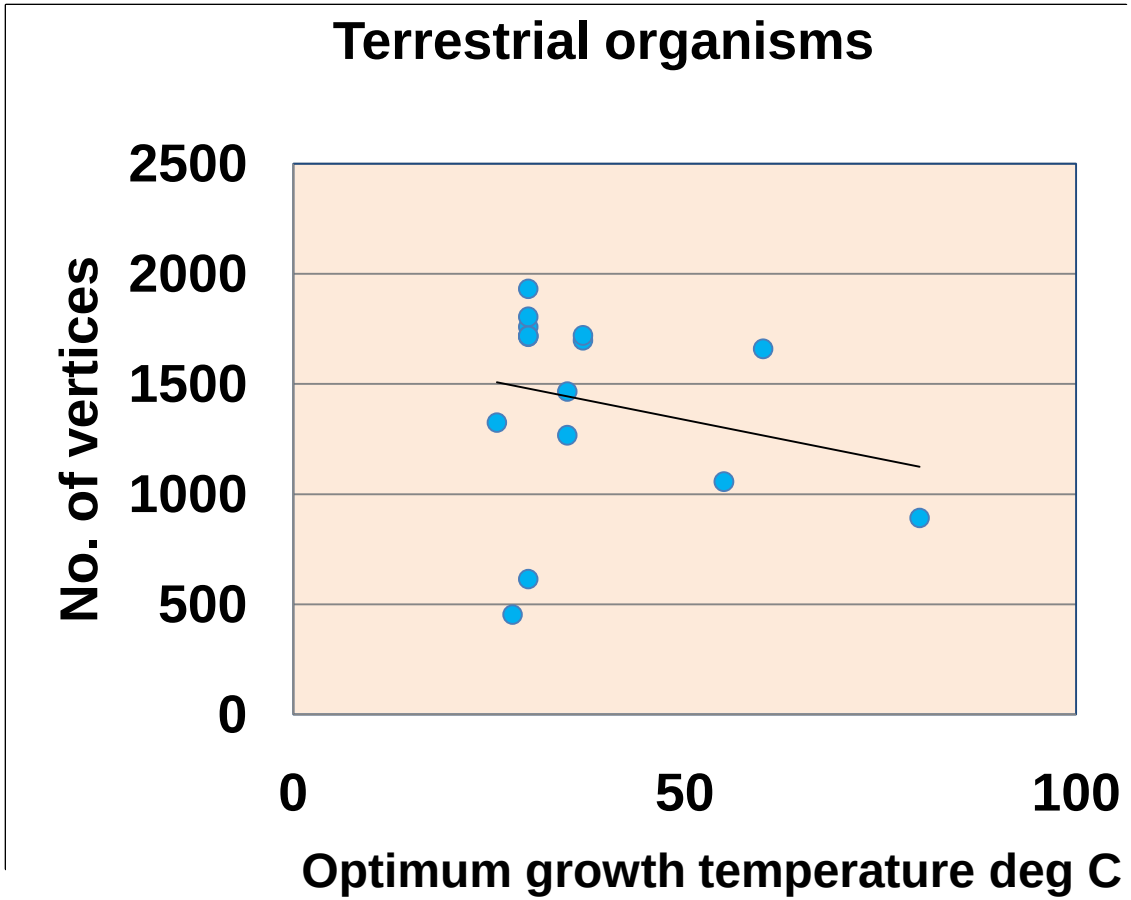
- ***arrows indicate direction of reaction***
- ***Orphan clusters : nodes not connected to main cluster***



Analysis

- Important nodes - Degree “hubs” (most highly connected nodes)
- Degree distribution & scale-free exponent
- Clustering coefficient
- K-cores and k-core hubs

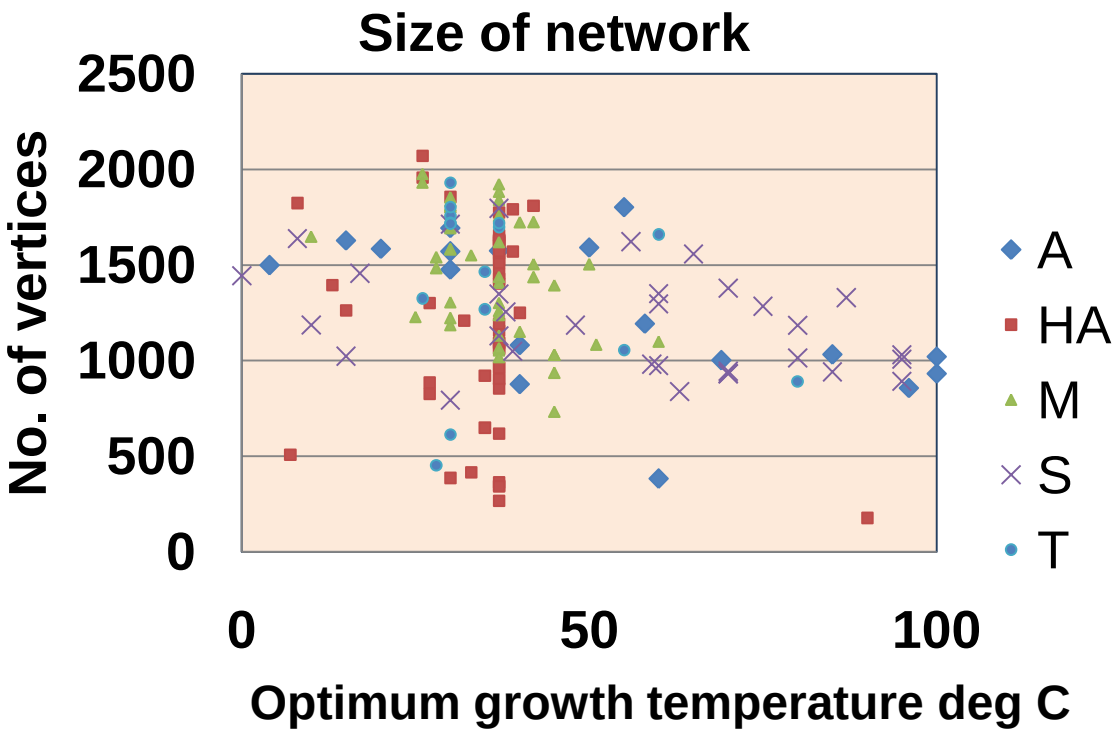
Size of metabolic networks



Organisms at normal temperature range (30-37 ^{0} C) show most variability in size

Mostly Host associated (both pathogen & symbiont)

Other factors also affect the network structure



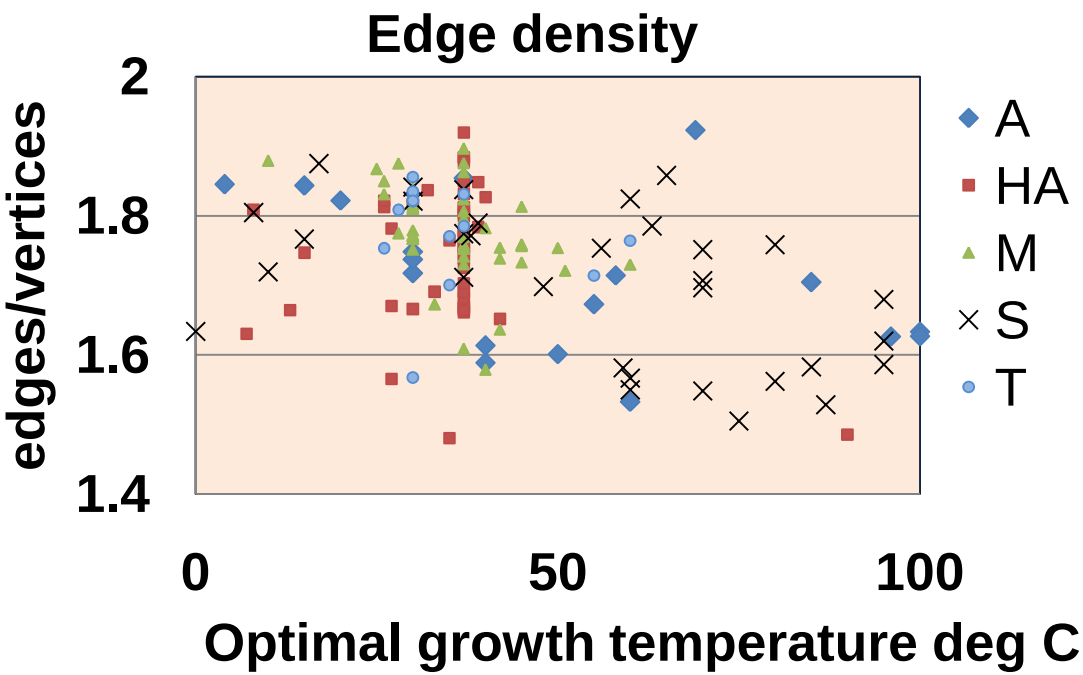
Life-style	Organisms	*r*	*p-val*
A	**18**	-0.61	0.007
HA	**50**	-0.22	0.12
M	**50**	-0.49	0.003
S	**31**	-0.38	0.03
T	**16**	-0.23	0.38
Total	**165**	-0.32	0.00003

Negatively correlated to growth temperature

Aquatic, multiple & specialized organisms show better correlation

Host-associated & terrestrial organisms show less correlation to growth temperature

Edge density

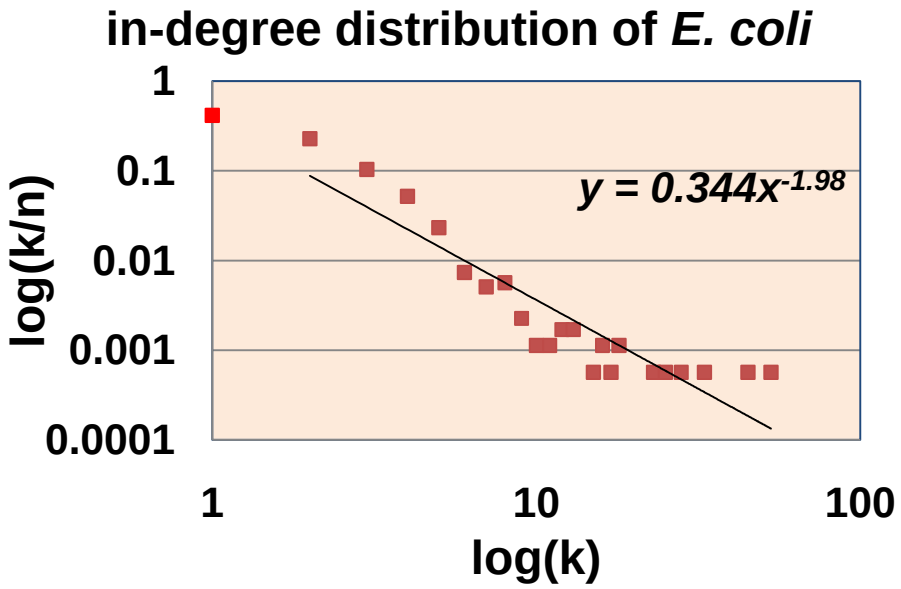


Negatively correlated to growth temperature

Host associated mesophiles have relatively higher densities

Life-style	Organisms	r	p-val
A	18	-0.46	0.05
HA	50	-0.17	0.23
M	50	-0.45	0.0009
S	31	-0.51	0.003
T	16	0.43	0.09
Total	165	-0.37	0.000001

Degree Distribution

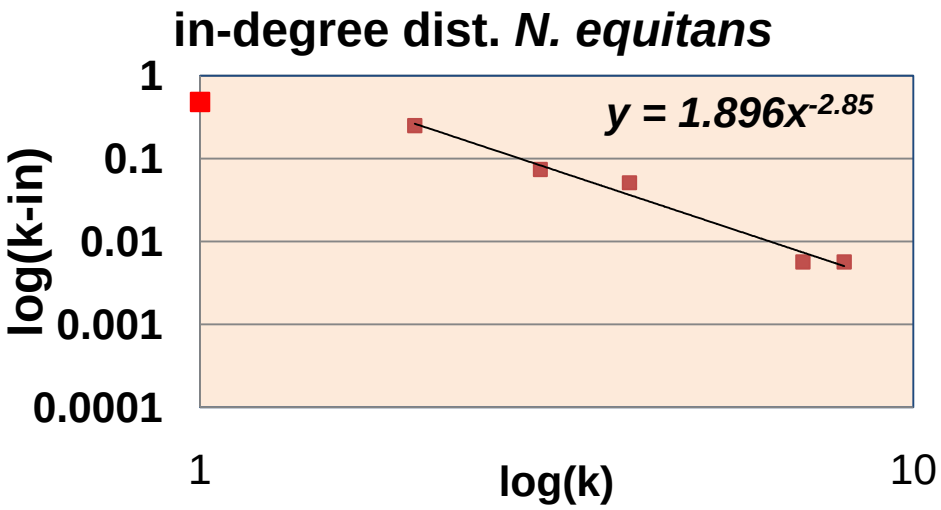


Many low-degree nodes
Few high-degree nodes

$$P(k) \sim k^{-\gamma}$$

γ - *degree exponent*

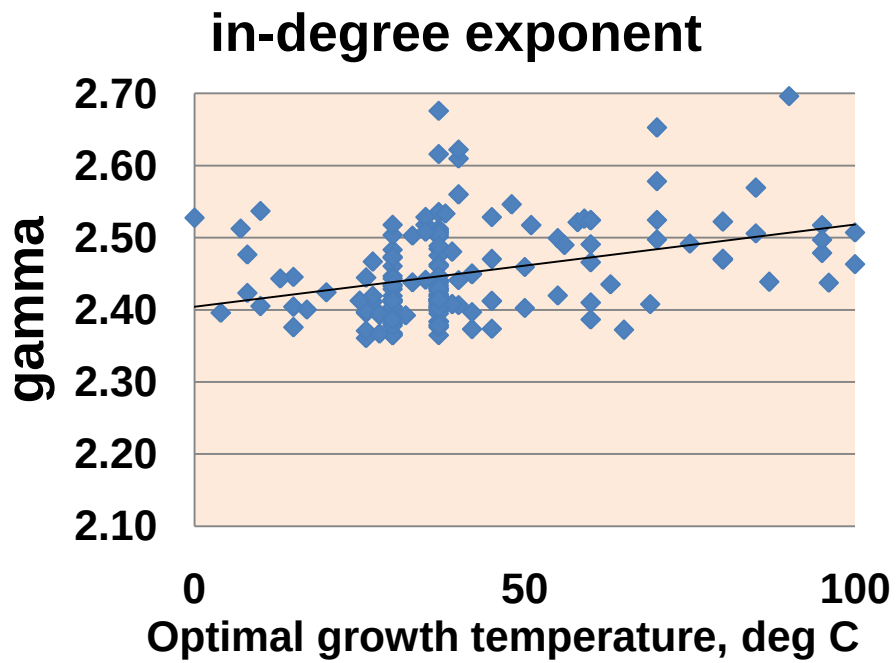
Similar out-degree distribution



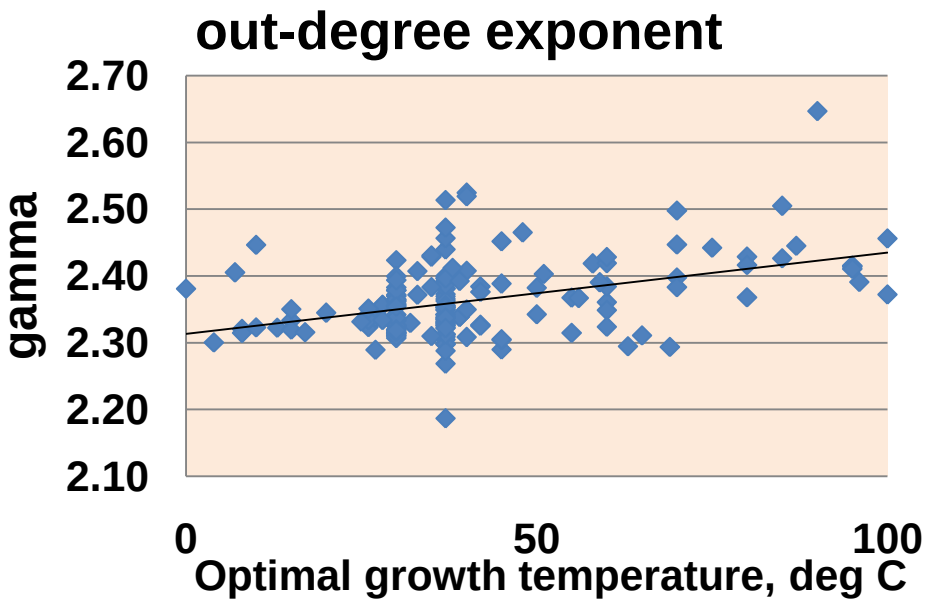
Metabolic networks have evolved in accordance with requirement of organism & not randomly

Growth of network through preferential attachment

Robust to random deletion of nodes



Life style	Org	r (gamma-in)	p-val
A	18	0.19	0.45
HA	50	0.39	0.004
M	50	0.35	0.01
S	31	0.25	0.17
T	16	0.38	0.14
Total	165	0.34	0.000006

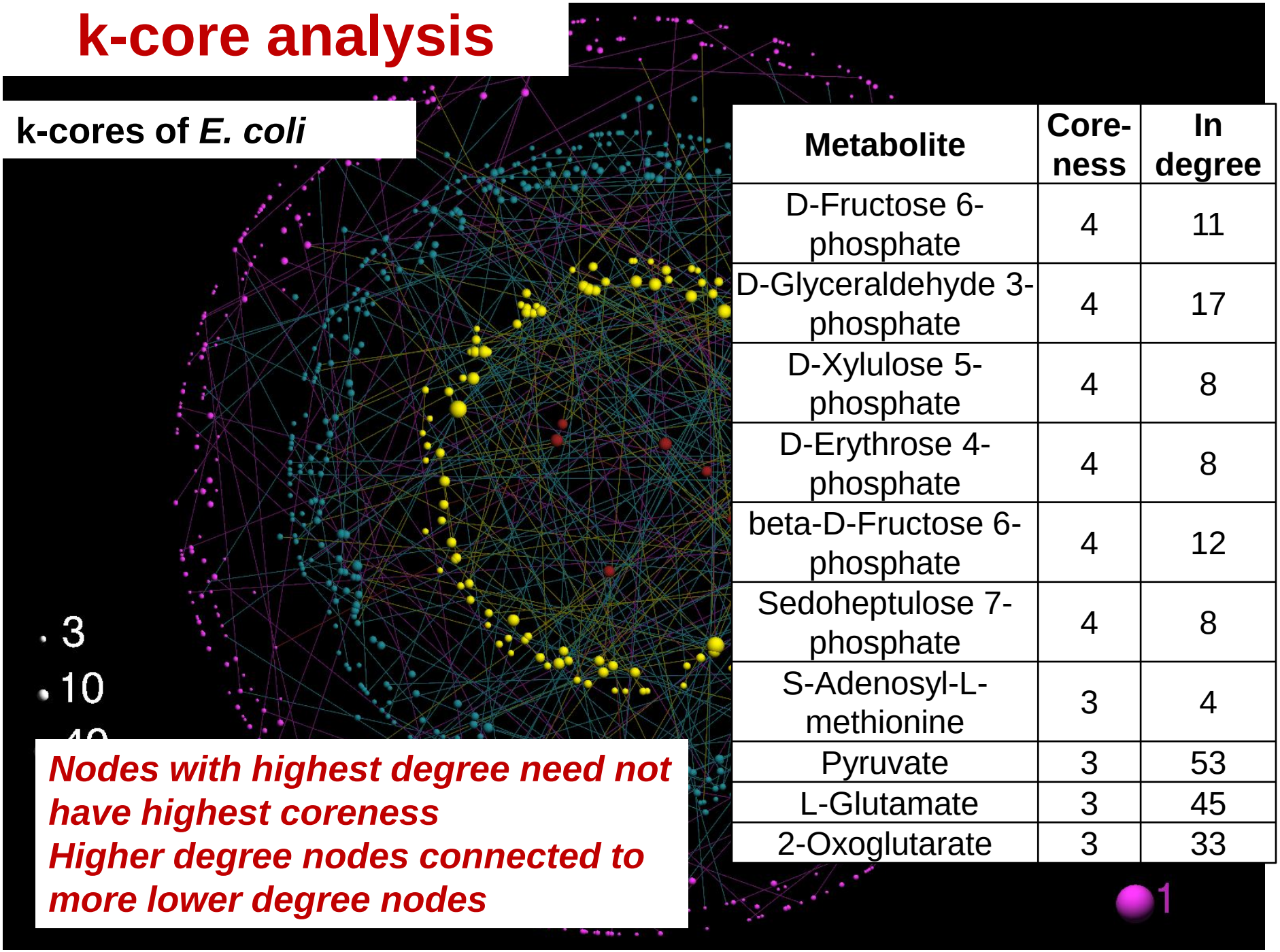


Life style	Org	r (out)	p-val
A	18	0.35	0.154
HA	50	0.50	0.0002
M	50	0.18	0.22
S	31	0.41	0.02
T	16	0.42	0.101
Total	165	0.46	<0.000001

Networks of organisms at lower optimal growth temperatures have more heterogenous structure

k-core analysis

k-cores of *E. coli*



Metabolite	Core-ness	In degree
D-Fructose 6-phosphate	4	11
D-Glyceraldehyde 3-phosphate	4	17
D-Xylulose 5-phosphate	4	8
D-Erythrose 4-phosphate	4	8
beta-D-Fructose 6-phosphate	4	12
Sedoheptulose 7-phosphate	4	8
S-Adenosyl-L-methionine	3	4
Pyruvate	3	53
L-Glutamate	3	45
2-Oxoglutarate	3	33

Nodes with highest degree need not have highest coreness
Higher degree nodes connected to more lower degree nodes

Inferences

Metabolic network topology of prokaryotes

- Correlated to environmetal conditions.
- Has evolved in accordance to their environmental and nutrional requirements.
- Scale free – robust to random deletion of nodes while targeted attack on several nodes can lead to breakdown in network structure.
- Network structure of prokaryotes at lower temperatures are more heterogeneous than those of thermophiles & hyperthermophiles.
- Nodes which have high degree centrality may have lower coreness; they are connected to lower degree nodes.

Acknowledgement

Radhika Khandelwal
MBN Chakravarthy
Roopali Upadhyay