Correlated evolutionary scenarios of metabolic functions

Murray Patterson[1], Thomas Bernard[1] and Daniel Kahn[1,2]

[1]LBBE, Université de Lyon 1 and [2]Département MIA, INRA

MCEB, L'île de Porquerolles, Jun 2015

Metabolic functions

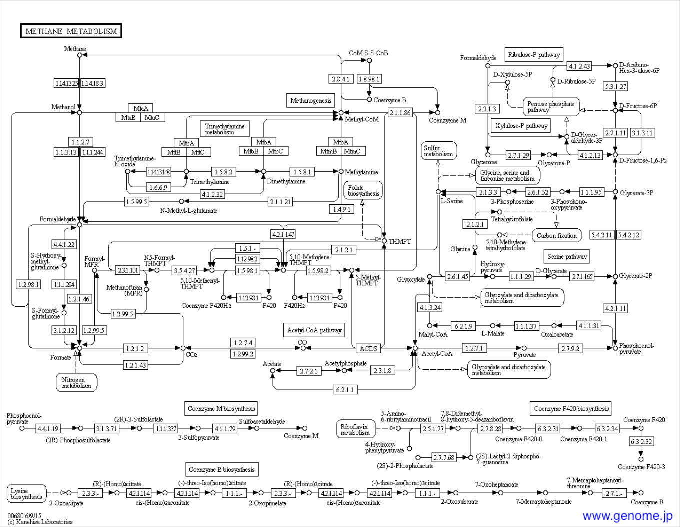
Determined by the enzymes (ECs) present – enzymes catalyze reactions

- many enzymes can be present at the same time, creating a complex metabolic network

Metabolic functions

Determined by the enzymes (ECs) present – enzymes catalyze reactions

- many enzymes can be present at the same time, creating a complex metabolic network



www.genome.jp

Metabolic functions, cont'd

Inferring metabolic functions

In principle, this can be inferred from the nucleic acid sequence content, e.g., as encoded by a gene or a genome

[1] PRIAM: PRofils pour l'Identification Automatique du Métabolisme (priam.prabi.fr)

[2] pbil.univ-lyon1.fr/databases/hogenom

Metabolic functions, cont'd

Inferring metabolic functions

In principle, this can be inferred from the nucleic acid sequence content, e.g., as encoded by a gene or a genome

- we do this systematically using PRIAM[1] – a set of profiles (PSSMs) for protein modules covering the Swiss-Prot database

[1] PRIAM: PRofils pour l'Identification Automatique du Métabolisme (priam.prabi.fr)

[2] pbil.univ-lyon1.fr/databases/hogenom

Metabolic functions, cont'd

Inferring metabolic functions

In principle, this can be inferred from the nucleic acid sequence content, e.g., as encoded by a gene or a genome

- we do this systematically using PRIAM[1] – a set of profiles (PSSMs) for protein modules covering the Swiss-Prot database
- idea : an rps-blast of the profiles against a protein sequence delivers a set of enzymes (ECs) with associated (presence) probabilities

[1] PRIAM: PRofils pour l'Identification Automatique du Métabolisme (priam.prabi.fr)

[2] pbil.univ-lyon1.fr/databases/hogenom

Metabolic functions, cont'd

Inferring metabolic functions

In principle, this can be inferred from the nucleic acid sequence content, e.g., as encoded by a gene or a genome

- we do this systematically using PRIAM[1] – a set of profiles (PSSMs) for protein modules covering the Swiss-Prot database
- idea : an rps-blast of the profiles against a protein sequence delivers a set of enzymes (ECs) with associated (presence) probabilities

The evolution of metabolic functions

We then want to explore evolutionary scenarios of these functions in order to understand the dependencies between them

- we do a study on HOGENOM 6[2] – a database of homologous gene families

[1] PRIAM: PRofils pour l'Identification Automatique du Métabolisme (priam.prabi.fr)

[2] pbil.univ-lyon1.fr/databases/hogenom

Hogenom Families

There are 296 917 families in HOGENOM 6

[3] note that we restrict to families with a tree on at least 3 sequences

Hogenom Families

There are 296 917 families in HOGENOM 6

- of which 10 699 (3.60% of the 296 917) families[3] have an EC assignment ($P(\text{EC}) > 0.5$) , i.e., at least one of its sequences has an EC assigned

[3] note that we restrict to families with a tree on at least 3 sequences

Hogenom Families

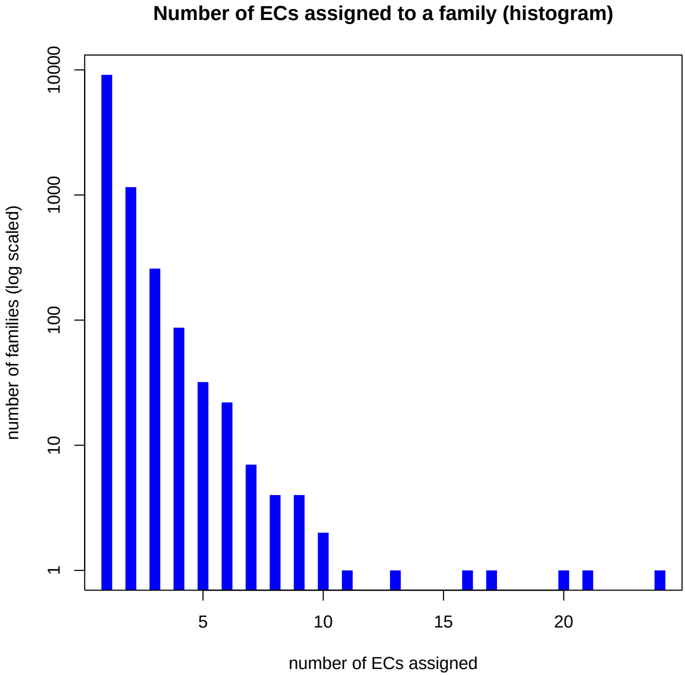
There are 296 917 families in HOGENOM 6

- of which 10 699 (3.60% of the 296 917) families[3] have an EC assignment ($P(\mathrm{EC}) > 0.5$) , i.e., at least one of its sequences has an EC assigned

First question :

How is EC assignment distributed amongst these 10 699 families?

[3] note that we restrict to families with a tree on at least 3 sequences


Number of ECs assigned to a family (histogram)
number of families (log scaled)
1
10
100
1000
10000
5
10
15
20
number of ECs assigned

Families with one and two EC assignements

Of the 10 699 families with at least one EC assignment, no family has more than 24 ECs assigned to it, while

Families with one and two EC assignements

Of the 10 699 families with at least one EC assignment, no family has more than 24 ECs assigned to it, while

- 9120 : 85.2% of the families have a single EC assignment, and

Families with one and two EC assignements

Of the 10 699 families with at least one EC assignment, no family has more than 24 ECs assigned to it, while

- 9120 : 85.2% of the families have a single EC assignment, and
- 1156 : 10.8% of the families have exactly two EC assignments

Families with one and two EC assignements

Of the 10 699 families with at least one EC assignment, no family has more than 24 ECs assigned to it, while

- 9120 : 85.2% of the families have a single EC assignment, and
- 1156 : 10.8% of the families have exactly two EC assignments

that is, 10 278 : 96.0% of the (10 699) families with any EC assignment have either one or two EC assignments

Families with one and two EC assignements

Of the 10 699 families with at least one EC assignment, no family has more than 24 ECs assigned to it, while

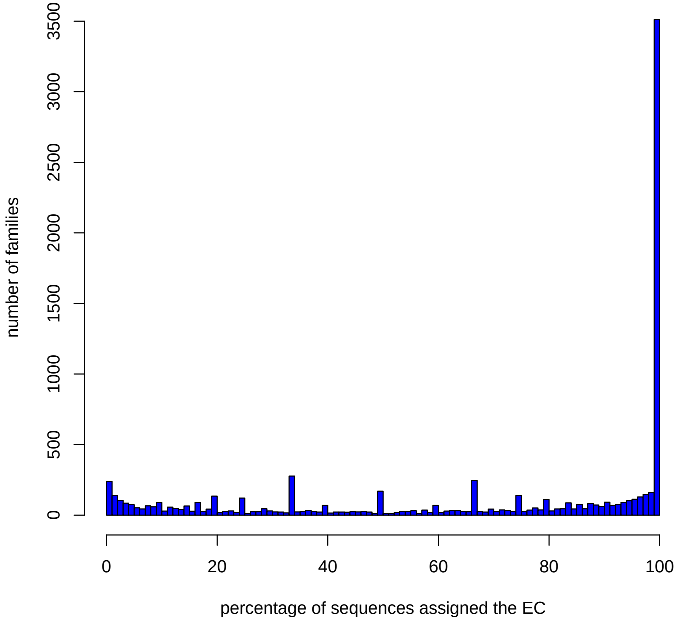
- 9120 : 85.2% of the families have a single EC assignment, and
- 1156 : 10.8% of the families have exactly two EC assignments

that is, 10 278 : 96.0% of the (10 699) families with any EC assignment have either one or two EC assignments

Families with one EC assignment :

What is the distribution of the percentage of sequences assigned the EC?

Families with one EC assignment (histogram)



Families with one and two EC assignements, cont'd

Families with one EC assignment :

In fact, of the 9120 families with one EC assignment, 3249 : 35.6% of these families are completely assigned with its EC

Families with one and two EC assignements, cont'd

Families with one EC assignment :

In fact, of the 9120 families with one EC assignment, 3249 : 35.6% of these families are completely assigned with its EC

- this is 30.4% of the (10 699) families with any EC assignment

Families with one and two EC assignements, cont'd

Families with one EC assignment :

In fact, of the 9120 families with one EC assignment, 3249 : 35.6% of these families are completely assigned with its EC

- this is 30.4% of the (10 699) families with any EC assignment

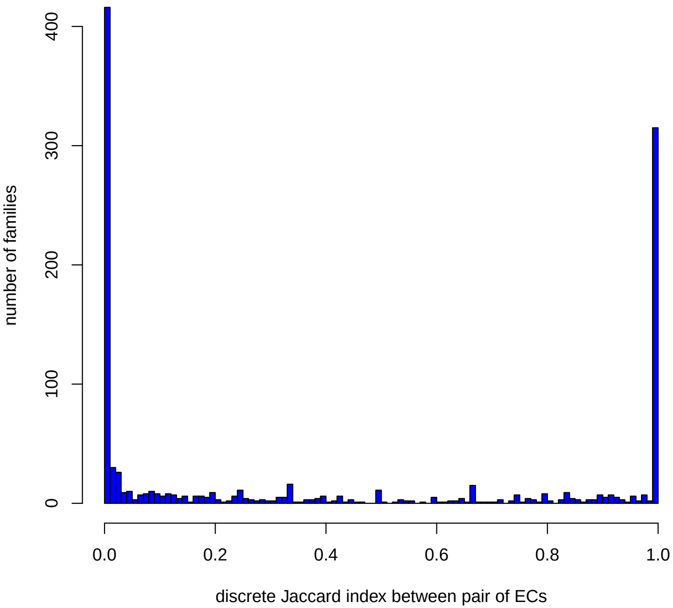
Families with two EC assignments :

What is the distribution of the Jaccard index,

$$\frac{|A \cap B|}{|A \cup B|}$$

of the pair of EC assignments for a family, where A (resp., B) is the set of sequences assigned with one (resp., the other) EC?

Families with two EC assignments (histogram)



Families with two EC assignments

In fact, of the 1156 families with two EC assignments, 318 : 27.5% (resp., 309 : 26.7%) have Jaccard index 0 (resp., 1)

Families with two EC assignments

In fact, of the 1156 families with two EC assignments, 318 : 27.5% (resp., 309 : 26.7%) have Jaccard index 0 (resp., 1)

- What about the families in the middle?

Hogenom Organisms

The gene families of HOGENOM 6 fall into 1471 organisms (compose 1471 genome sequences)

Hogenom Organisms

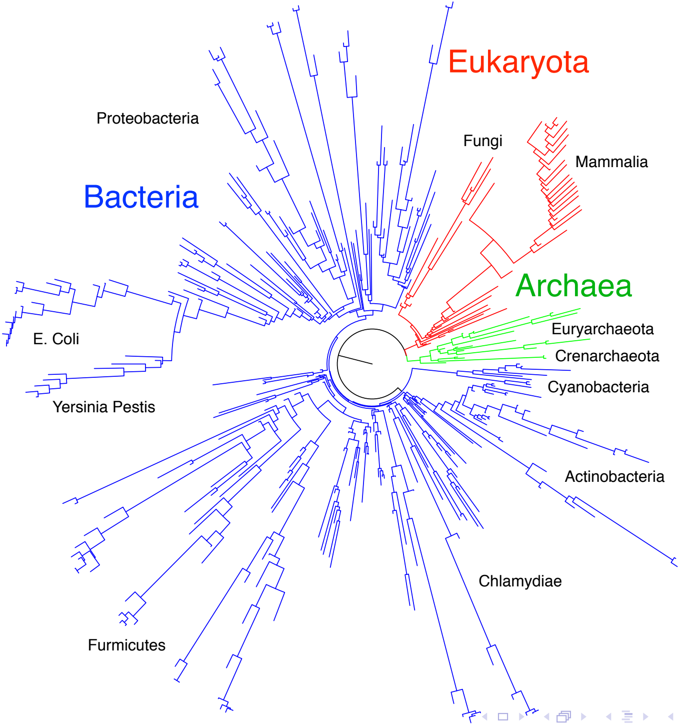
The gene families of HOGENOM 6 fall into 1471 organisms (compose 1471 genome sequences)

- for which there is a (an ultrametric) species tree on 1460 of these organisms

Hogenom Organisms

The gene families of HOGENOM 6 fall into 1471 organisms (compose 1471 genome sequences)

- for which there is a (an ultrametric) species tree on 1460 of these organisms
- from which we consider a subtree on 1452 organisms, due to incomplete genome sequences, or non-species (i.e., mitochondria, plasmids or nucleomorphs)


Eukaryota
Proteobacteria
Fungi
Mammalia
Bacteria
Archaea
E. Coli
Euryarchaeota
Crenarchaeota
Cyanobacteria
Yersinia Pestis
Actinobacteria
Chlamydiae
Furmicutes

Hogenom Organisms

The gene families of HOGENOM 6 fall into 1471 organisms (compose 1471 genome sequences)

- for which there is a (an ultrametric) species tree on 1460 of these organisms
- from which we consider a subtree on 1452 organisms, due to incomplete genome sequences, or non-species (i.e., mitochondria, plasmids or nucleomorphs)

Hogenom Organisms

The gene families of HOGENOM 6 fall into 1471 organisms (compose 1471 genome sequences)

- for which there is a (an ultrametric) species tree on 1460 of these organisms
- from which we consider a subtree on 1452 organisms, due to incomplete genome sequences, or non-species (i.e., mitochondria, plasmids or nucleomorphs)

For a collection of protein sequences (as encoded by the genome of an organism), the PRIAM search tool

Hogenom Organisms

The gene families of HOGENOM 6 fall into 1471 organisms (compose 1471 genome sequences)

- for which there is a (an ultrametric) species tree on 1460 of these organisms
- from which we consider a subtree on 1452 organisms, due to incomplete genome sequences, or non-species (i.e., mitochondria, plasmids or nucleomorphs)

For a collection of protein sequences (as encoded by the genome of an organism), the PRIAM search tool

- screens the sequences with the PRIAM profiles (using rps-blast) to get a set of hits, with associated proabilities, for each profile

Hogenom Organisms

The gene families of HOGENOM 6 fall into 1471 organisms (compose 1471 genome sequences)

- for which there is a (an ultrametric) species tree on 1460 of these organisms
- from which we consider a subtree on 1452 organisms, due to incomplete genome sequences, or non-species (i.e., mitochondria, plasmids or nucleomorphs)

For a collection of protein sequences (as encoded by the genome of an organism), the PRIAM search tool

- screens the sequences with the PRIAM profiles (using rps-blast) to get a set of hits, with associated proabilities, for each profile
- applies a set of EC-specific logical rules to the (sets of) hits of all profiles that concern the given EC to infer an overall probability for each enzyme in the collection (i.e., organism)

Hogenom Organisms

The gene families of HOGENOM 6 fall into 1471 organisms (compose 1471 genome sequences)

- for which there is a (an ultrametric) species tree on 1460 of these organisms
- from which we consider a subtree on 1452 organisms, due to incomplete genome sequences, or non-species (i.e., mitochondria, plasmids or nucleomorphs)

For a collection of protein sequences (as encoded by the genome of an organism), the PRIAM search tool

- screens the sequences with the PRIAM profiles (using rps-blast) to get a set of hits, with associated proabilities, for each profile
- applies a set of EC-specific logical rules to the (sets of) hits of all profiles that concern the given EC to infer an overall probability for each enzyme in the collection (i.e., organism)

Here we apply PRIAM search to the organisms of HOGENOM 6

Evolutionary scenarios of metabolic functions

Construction

A given EC x is at each leaf of the species tree with a certain probability

[4] biopp.univ-montp2.fr/forge/testnh

Evolutionary scenarios of metabolic functions

Construction

A given EC x is at each leaf of the species tree with a certain probability

- use MapNH[4] to infer gain (resp., loss) probabilities $p_b(x), b \in \mathcal{B}$ of this EC x on the branches $\mathcal{B}$ of the species tree

[4] biopp.univ-montp2.fr/forge/testnh

Evolutionary scenarios of metabolic functions

Construction

A given EC x is at each leaf of the species tree with a certain probability

- use MapNH[4] to infer gain (resp., loss) probabilities $p_b(x), b \in \mathcal{B}$ of this EC x on the branches $\mathcal{B}$ of the species tree

Analysis

We combine the probabilities on the branches of the tree to obtain

[4] biopp.univ-montp2.fr/forge/testnh

Evolutionary scenarios of metabolic functions

Construction

A given EC x is at each leaf of the species tree with a certain probability

- use MapNH[4] to infer gain (resp., loss) probabilities $p_b(x), b \in \mathcal{B}$ of this EC x on the branches $\mathcal{B}$ of the species tree

Analysis

We combine the probabilities on the branches of the tree to obtain

- $p(x) = \sum_{b \in \mathcal{B}} p_b(x)$: the overall gain (resp., loss) probability of EC x on the tree,

[4] biopp.univ-montp2.fr/forge/testnh

Evolutionary scenarios of metabolic functions

Construction

A given EC x is at each leaf of the species tree with a certain probability

- use MapNH[4] to infer gain (resp., loss) probabilities $p_b(x), b \in \mathcal{B}$ of this EC x on the branches $\mathcal{B}$ of the species tree

Analysis

We combine the probabilities on the branches of the tree to obtain

- $p(x) = \sum_{b \in \mathcal{B}} p_b(x)$: the overall gain (resp., loss) probability of EC x on the tree,
- $p(x, y) = \sum_{b \in \mathcal{B}} p_b(x) \cdot p_b(y)$: the overall *joint* gain (resp., loss) of ECs x and y on the tree

[4] biopp.univ-montp2.fr/forge/testnh

Evolutionary scenarios of metabolic functions

Construction

A given EC x is at each leaf of the species tree with a certain probability

- use MapNH[4] to infer gain (resp., loss) probabilities $p_b(x), b \in \mathcal{B}$ of this EC x on the branches $\mathcal{B}$ of the species tree

Analysis

We combine the probabilities on the branches of the tree to obtain

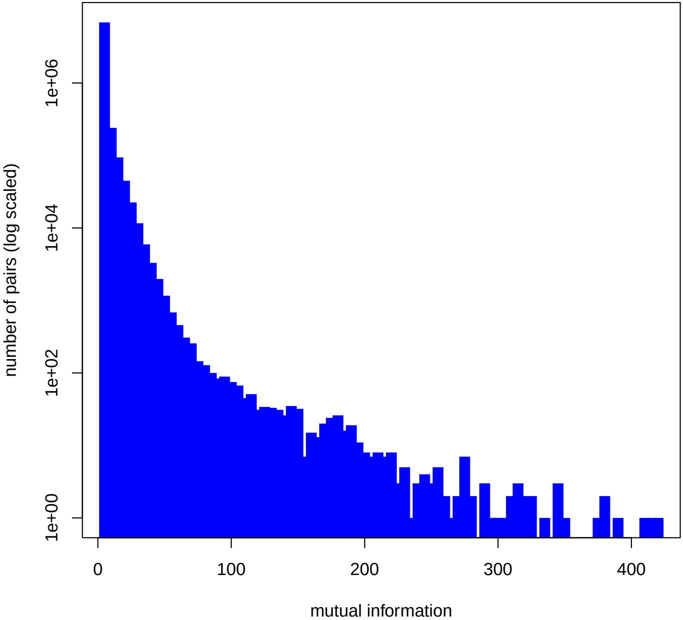
- $p(x) = \sum_{b \in \mathcal{B}} p_b(x)$: the overall gain (resp., loss) probability of EC x on the tree,
- $p(x, y) = \sum_{b \in \mathcal{B}} p_b(x) \cdot p_b(y)$: the overall *joint* gain (resp., loss) of ECs x and y on the tree

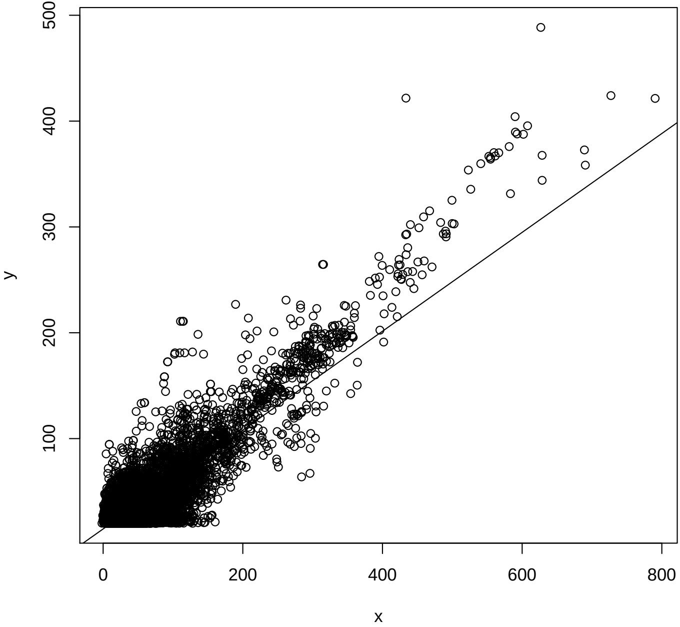
Compute the mutual information (MI) for each pair of ECs x and y :

$$\mathrm{MI}(x, y) = \sum_{x' \in \{x, \overline{x}\}} \sum_{y' \in \{y, \overline{y}\}} p(x', y') \log \left(\frac{p(x', y')}{p(x')p(y')} \right)$$

[4] biopp.univ-montp2.fr/forge/testnh

Mutual information for a pair of ECs (histogram)




500
400
300
200
100
y
0
200
400
600
800
x

Community Analysis of Mutual Information

The Graph

threshold	nodes (isolated)	edges	avg. degree
5.0E-3	1124 (+ 1608 = 2732)	22005	39.2
5.0E-4	974 (+ 1758 = 2732)	11792	24.2
5.0E-5	836 (+ 1896 = 2732)	6962	16.7

Community Analysis of Mutual Information

The Graph

threshold	nodes (isolated)	edges	avg. degree
5.0E-3	1124 (+ 1608 = 2732)	22005	39.2
5.0E-4	974 (+ 1758 = 2732)	11792	24.2
5.0E-5	836 (+ 1896 = 2732)	6962	16.7

Node Communities (Blondel et al., 2008)

threshold	number	# /w 2 ECs	# /w 3 ECs	mean size	max size
5.0E-3	40	22 (55%)	6 (15%)	28.1	335
5.0E-4	43	26 (60%)	7 (16%)	22.7	268
5.0E-5	53	29 (55%)	6 (11%)	15.8	223

Community Analysis of Mutual Information

The Graph

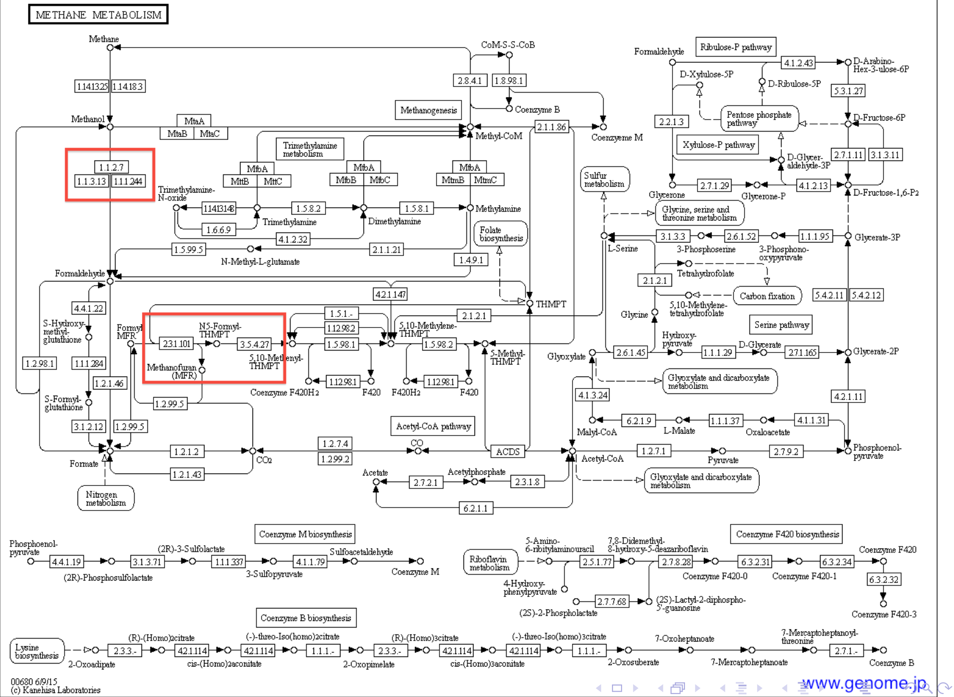
threshold	nodes (isolated)	edges	avg. degree
5.0E-3	1124 (+ 1608 = 2732)	22005	39.2
5.0E-4	974 (+ 1758 = 2732)	11792	24.2
5.0E-5	836 (+ 1896 = 2732)	6962	16.7

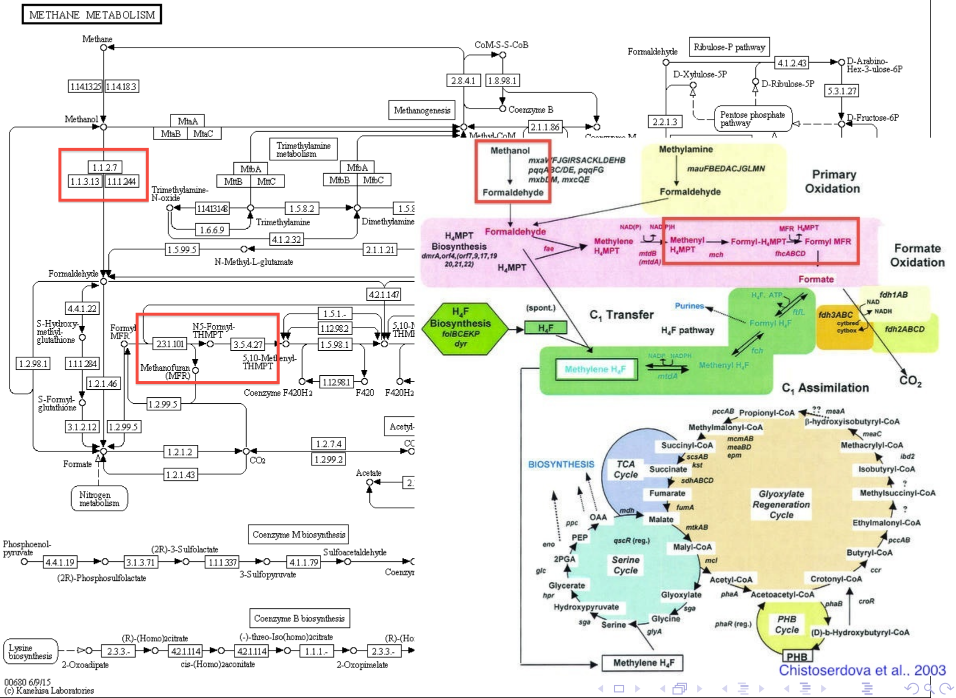
Node Communities (Blondel et al., 2008)

threshold	number	# /w 2 ECs	# /w 3 ECs	mean size	max size
5.0E-3	40	22 (55%)	6 (15%)	28.1	335
5.0E-4	43	26 (60%)	7 (16%)	22.7	268
5.0E-5	53	29 (55%)	6 (11%)	15.8	223

Link Communities (Ahn et al., 2010)

threshold	number	# /w 2 ECs	# /w 3 ECs	mean size	max size
5.0E-3	5571	4480 (80%)	473 (8%)	3.1	129
5.0E-4	3084	2381 (77%)	318 (10%)	3.0	107
5.0E-5	2043	1554 (76%)	265 (13%)	2.7	93


METHANE METABOLISM
Methane
1.14.13.25
1.14.18.3
Methanol
MtaA
MtaB
MtaC
CoM-S-S-CoB
2.8.4.1
1.8.98.1
Coenzyme B
Methanogenesis
Methyl-CoM
2.1.1.86
Coenzyeme M
1.1.2.7
1.1.3.13
1.1.1.244
Trimethylamine metabolism
MtbA
MtbB
MtbC
MttB
MttC
MtmB
MtmC
Trimethylamine-N-oxide
1.14.13.148
1.5.8.2
1.5.8.1
Methylamine
1.6.6.9
Trimethylamine
Dimethylamine
4.1.2.32
1.5.99.5
N-Methyl-L-glutamate
2.1.1.21
Folate biosynthesis
1.4.9.1
Formaldehyde
4.4.1.22
S-Hydroxy-methyl-glutathione
1.2.98.1
1.1.1.284
S-Formyl-glutathione
1.2.1.46
3.1.2.12
1.2.99.5
Formate
Nitrogen metabolism
Formyl-MFR
2.3.1.101
N5-Formyl-THMPT
3.5.4.27
5,10-Methenyl-THMPT
Methanofuran (MFR)
1.2.99.5
4.2.1.147
1.5.1.-
1.12.98.2
1.5.98.1
2.1.2.1
5,10-Methylene-THMPT
1.5.98.2
5-Methyl-THMPT
THMPT
1.12.98.1
Coenzyme F420H2
F420
F420H2
1.2.1.2
1.2.1.43
CO2
1.2.7.4
1.2.99.2
CO
Acetyl-CoA pathway
ACDS
Acetate
2.7.2.1
Acetylphosphate
2.3.1.8
6.2.1.1
Acetyl-CoA
1.2.7.1
Pyruvate
2.7.9.2
Phosphoenol-pyruvate
Glyoxylate and dicarboxylate metabolism
Malyl-CoA
6.2.1.9
L-Malate
1.1.1.37
Oxaloacetate
4.1.1.31
4.1.3.24
Glyoxylate
2.6.1.45
Hydroxy-pyruvate
1.1.1.29
D-Glycerate
2.7.1.165
Glycerate-2P
4.2.1.11
Glyoxylate and dicarboxylate metabolism
Serine pathway
Glycine
2.1.2.1
Tetrahydrofolate
5,10-Methylene-tetrahydrofolate
Carbon fixation
5.4.2.11
5.4.2.12
L-Serine
3.1.3.3
3-Phosphoserine
2.6.1.52
3-Phosphono-oxypyruvate
1.1.1.95
Glycerate-3P
Sulfur metabolism
Glycine, serine and threonine metabolism
Formaldehyde
Ribulose-P pathway
4.1.2.43
D-Arabino-Hex-3-ulose-6P
D-Xylulose-5P
D-Ribulose-5P
5.3.1.27
2.2.1.3
Pentose phosphate pathway
D-Fructose-6P
Xylulose-P pathway
D-Glycer-aldehyde-3P
2.7.1.11
3.1.3.11
Glycerone
2.7.1.29
Glycerone-P
4.1.2.13
D-Fructose-1,6-P2
Coenzyme M biosynthesis
Phosphoenol-pyruvate
4.4.1.19
(2R)-Phosphosulfolactate
3.1.3.71
(2R)-3-Sulfolactate
1.1.1.337
3-Sulfopyruvate
4.1.1.79
Sulfoacetaldehyde
Coenzyme M
Coenzyme F420 biosynthesis
Riboflavin metabolism
5-Amino-6-ribitylaminouracil
2.5.1.77
4-Hydroxy-phenylpyruvate
7,8-Didemethyl-8-hydroxy-5-deazariboflavin
2.7.8.28
Coenzyme F420-0
6.3.2.31
Coenzyme F420-1
6.3.2.34
Coenzyme F420
6.3.2.32
Coenzyme F420-3
2.7.7.68
(2S)-2-Phospholactate
(2S)-Lactyl-2-diphospho-5'-guanosine
Coenzyme B biosynthesis
Lysine biosynthesis
2-Oxoadipate
2.3.3.-
(R)-(Homo)2citrate
4.2.1.114
cis-(Homo)2aconitate
4.2.1.114
(-)-threo-Iso(homo)2citrate
1.1.1.-
2-Oxopimelate
2.3.3.-
(R)-(Homo)3citrate
4.2.1.114
cis-(Homo)3aconitate
4.2.1.114
(-)-threo-Iso(homo)3citrate
1.1.1.-
2-Oxosuberate
7-Oxoheptanoate
7-Mercaptoheptanoate
7-Mercaptoheptanoyl-threonine
2.7.1.-
Coenzyme B
00680 6/9/15
(c) Kanehisa Laboratories
www.genome.jp



Chistoserdova et al., 2003

So, what's next?

Reconciled Trees

Perform the same study on a Cyanobacteria dataset reconciled by the method of Szöllősi et al., (2013)

So, what's next?

Reconciled Trees

Perform the same study on a Cyanobacteria dataset reconciled by the method of Szöllősi et al., (2013)

- we will have (more detailed) evolutionary scenarios for ECs,

So, what's next?

Reconciled Trees

Perform the same study on a Cyanobacteria dataset reconciled by the method of Szöllősi et al., (2013)

- we will have (more detailed) evolutionary scenarios for ECs,
- but also ancestral sequences (idea : run PRIAM on these and compare to the ECs scenarios)

So, what's next?

Reconciled Trees

Perform the same study on a Cyanobacteria dataset reconciled by the method of Szöllősi et al., (2013)

- we will have (more detailed) evolutionary scenarios for ECs,
- but also ancestral sequences (idea : run PRIAM on these and compare to the ECs scenarios)

Stoichiometry Analysis

Given a collection of enzymes (ECs), PRIAM can construct a draft metabolic network, using Kegg pathway data

So, what's next?

Reconciled Trees

Perform the same study on a Cyanobacteria dataset reconciled by the method of Szöllősi et al., (2013)

- we will have (more detailed) evolutionary scenarios for ECs,
- but also ancestral sequences (idea : run PRIAM on these and compare to the ECs scenarios)

Stoichiometry Analysis

Given a collection of enzymes (ECs), PRIAM can construct a draft metabolic network, using Kegg pathway data

- construct extant, and ancestral networks

So, what's next?

Reconciled Trees

Perform the same study on a Cyanobacteria dataset reconciled by the method of Szöllősi et al., (2013)

- we will have (more detailed) evolutionary scenarios for ECs,
- but also ancestral sequences (idea : run PRIAM on these and compare to the ECs scenarios)

Stoichiometry Analysis

Given a collection of enzymes (ECs), PRIAM can construct a draft metabolic network, using Kegg pathway data

- construct extant, and ancestral networks
- perform a comparison from a reaction stoichiometry point of view (Poolman et al., 2007)

Acknowledgements

- Ancestrome Project and UCBL (funding, etc.)
- Simon Penel and Vincent Daubin (hogenom)
- Laurent Guéguen and Julien Dutheil (bio++)
- Dominique Guyot (paraload)
- Gergely Szöllősi (reconciled trees)

Thank you!

Any questions?

The goals of this project

We outline a few of the main goals of this project :

The goals of this project

We outline a few of the main goals of this project :

1. to investigate principles of metabolic network evolution : how is function related to evolution? How is evolution determined by function?

The goals of this project

We outline a few of the main goals of this project :

1. to investigate principles of metabolic network evolution : how is function related to evolution? How is evolution determined by function?
2. in the context of the Ancestrome project : to introduce functional dependencies into the likelihood calculations

The goals of this project

We outline a few of the main goals of this project :

1. to investigate principles of metabolic network evolution : how is function related to evolution? How is evolution determined by function?
2. in the context of the Ancestrome project : to introduce functional dependencies into the likelihood calculations
3. evaluating hypotheses about ancestral environments : metabolism sheds light on environmental factors, which could provide clues on the events associated with the emergence of ancestral phyla

Dependencies between metabolic functions

Our goal is to find (and to understand) the dependencies between different metabolic functions within these evolutionary scenarios

Dependencies between metabolic functions

Our goal is to find (and to understand) the dependencies between different metabolic functions within these evolutionary scenarios

From the evolutionary tree

Determine the (pairwise) relationships between ECs using mutual information

Dependencies between metabolic functions

Our goal is to find (and to understand) the dependencies between different metabolic functions within these evolutionary scenarios

From the evolutionary tree

Determine the (pairwise) relationships between ECs using mutual information

- the mutual information between two (discrete) random variables X and Y is defined as :

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right)$$

Dependencies between metabolic functions

Our goal is to find (and to understand) the dependencies between different metabolic functions within these evolutionary scenarios

From the evolutionary tree

Determine the (pairwise) relationships between ECs using mutual information

- the mutual information between two (discrete) random variables X and Y is defined as :

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right)$$

- from here we can build a large graph that represents these pairwise dependencies, and try to find modules within this graph

Dependencies between metabolic functions

Our goal is to find (and to understand) the dependencies between different metabolic functions within these evolutionary scenarios

Dependencies between metabolic functions

Our goal is to find (and to understand) the dependencies between different metabolic functions within these evolutionary scenarios

From the structure of the networks

Conversely, we may detect functional dependencies purely from the structure of a metabolic network (i.e., at a given node of the species tree)

Dependencies between metabolic functions

Our goal is to find (and to understand) the dependencies between different metabolic functions within these evolutionary scenarios

From the structure of the networks

Conversely, we may detect functional dependencies purely from the structure of a metabolic network (i.e., at a given node of the species tree)

- from a reaction stoichiometry there exist methods for finding correlations among subsets of reactions (Poolman et al., 2007)

Dependencies between metabolic functions

Our goal is to find (and to understand) the dependencies between different metabolic functions within these evolutionary scenarios

From the structure of the networks

Conversely, we may detect functional dependencies purely from the structure of a metabolic network (i.e., at a given node of the species tree)

- from a reaction stoichiometry there exist methods for finding correlations among subsets of reactions (Poolman et al., 2007)
- from this, we get dependencies between sets of reactions, functions (their ECs)

Reconstructing evolutionary scenarios

For each of the individual family and whole genome viewpoints, there are again two viewpoints on the reconstruction of evolutionary scenarios – in terms of either

Reconstructing evolutionary scenarios

For each of the individual family and whole genome viewpoints, there are again two viewpoints on the reconstruction of evolutionary scenarios – in terms of either

ECs

Apply a method (parsimony or ML) to propagate the ECs to the ancestral nodes of the tree

- the collection of ECs at the ancestral nodes will then determine the functions active at these nodes

Reconstructing evolutionary scenarios

For each of the individual family and whole genome viewpoints, there are again two viewpoints on the reconstruction of evolutionary scenarios – in terms of either

ECs

Apply a method (parsimony or ML) to propagate the ECs to the ancestral nodes of the tree

- the collection of ECs at the ancestral nodes will then determine the functions active at these nodes

Nucelic acid sequences

Apply a method (parsimony or ML) to propagate the sequences to the ancestral nodes of the tree

- we can apply PRIAM search on these ancestral sequences (just like we did for the extant ones) to get collections of ancestral ECs at the ancestral nodes

reaction stoichimetry

	s_1	s_2
e_a	0.76	0.45
e_b	0.39	0.87
e_c	0.68	0.23

reaction stoichimetry

	s_1	s_2
e_a	0.76	0.45
e_b	0.39	0.87
e_c	0.68	0.23

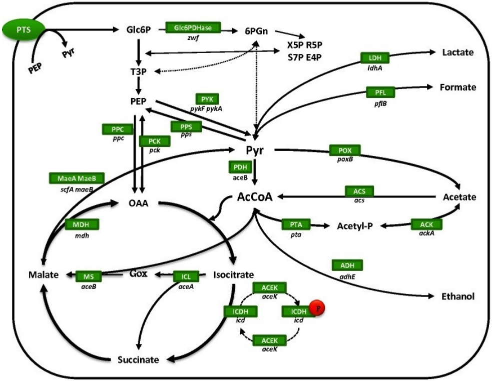
➡

EC 1.	A+O → AO
EC 2.	AB+C → A+BC
EC 3.	AB+H_2O → AOH+BH
EC 6.	X+Y+ATP → XY+ADP+Pi

reaction stoichimetry

	s_1	s_2
e_a	0.76	0.45
e_b	0.39	0.87
e_c	0.68	0.23

EC 1.	$A+O \rightarrow AO$
EC 2.	$AB+C \rightarrow A+BC$
EC 3.	$AB+H_2O \rightarrow AOH+BH$
EC 6.	$X+Y+ATP \rightarrow XY+ADP+Pi$



PTS
PEP
Pyr
Glc6P
Glc6PDHase
zwf
6PGn
X5P R5P
S7P E4P
T3P
PEP
PYK
pykF pykA
PPS
pps
PPC
ppc
PCK
pck
LDH
ldhA
Lactate
PFL
pflB
Formate
Pyr
POX
poxB
PDH
aceB
AcCoA
ACS
acs
Acetate
MaeA MaeB
scfA maeB
OAA
MDH
mdh
PTA
pta
Acetyl-P
ACK
ackA
ADH
adhE
Ethanol
Malate
MS
aceB
Gox
ICL
aceA
Isocitrate
ACEK
aceK
ICDH
icd
ICDH
icd
P
ACEK
aceK
Succinate