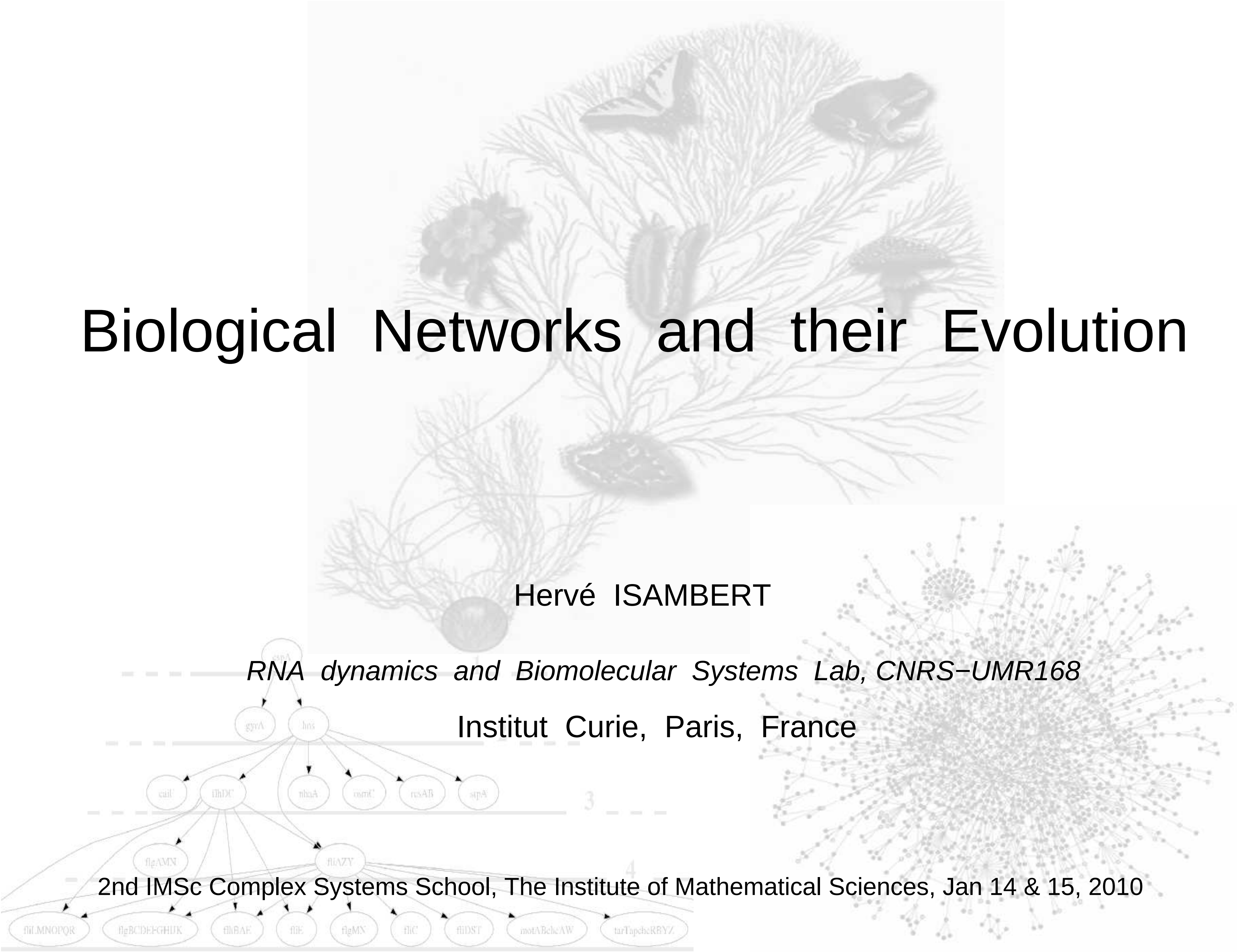


Biological Networks and their Evolution

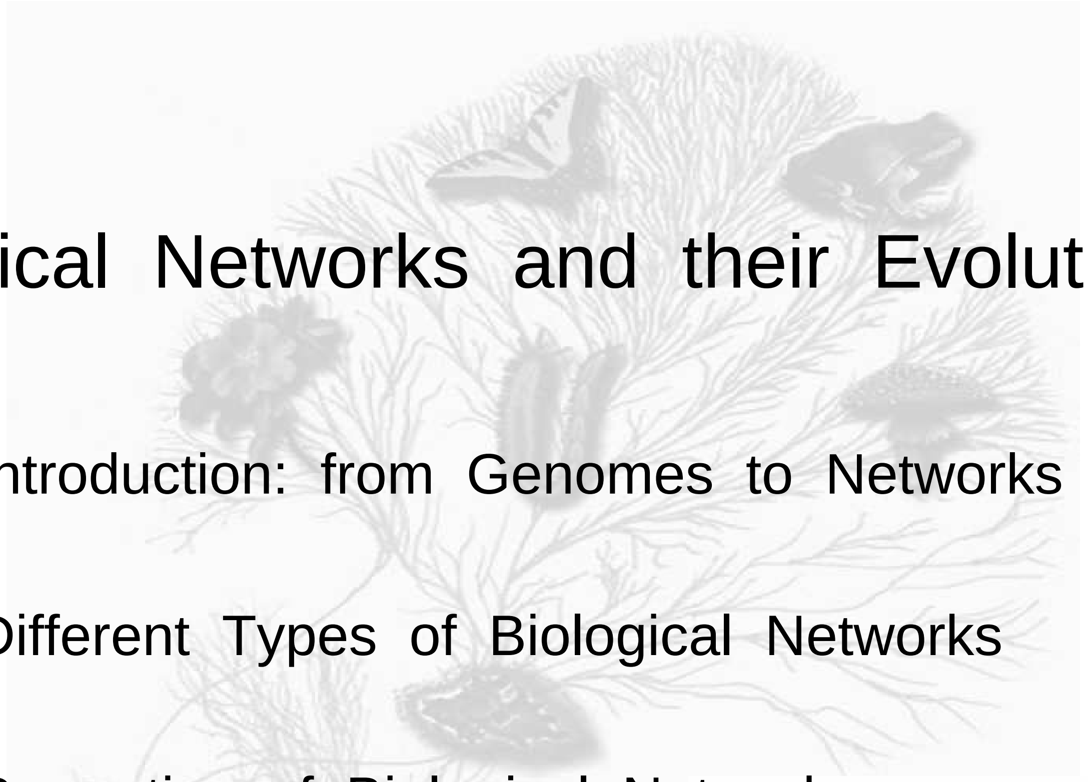
Hervé ISAMBERT

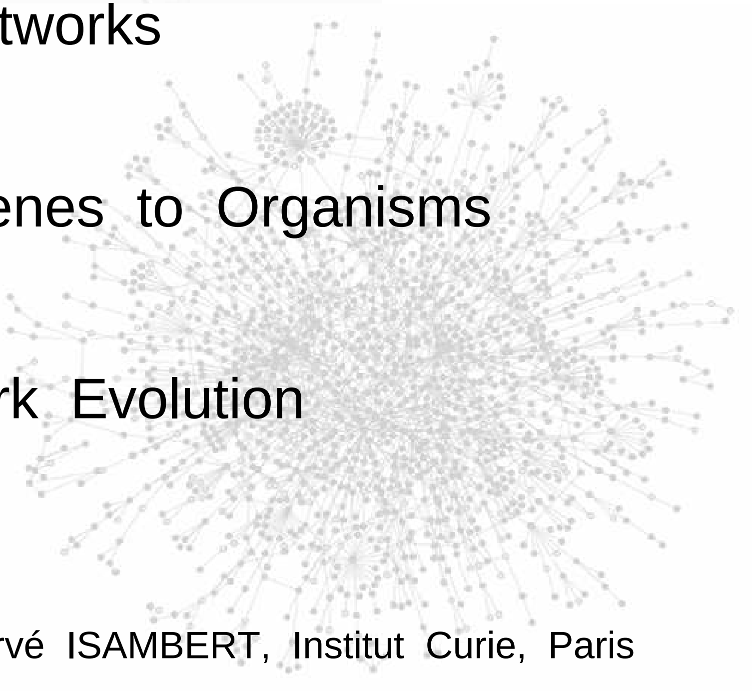
RNA dynamics and Biomolecular Systems Lab, CNRS−UMR168

Institut Curie, Paris, France

2nd IMSc Complex Systems School, The Institute of Mathematical Sciences, Jan 14 & 15, 2010

Biological Networks and their Evolution





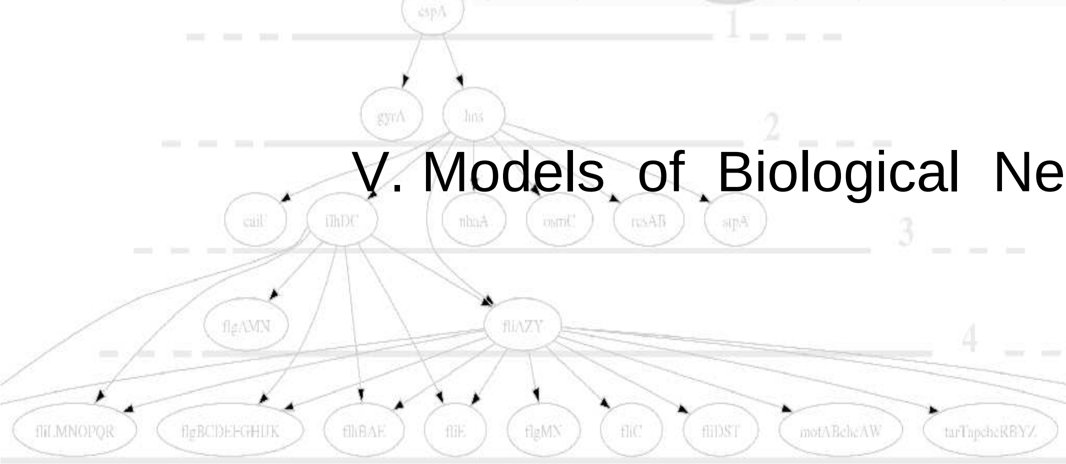
I. Introduction: from Genomes to Networks

II. Different Types of Biological Networks

III. Properties of Biological Networks

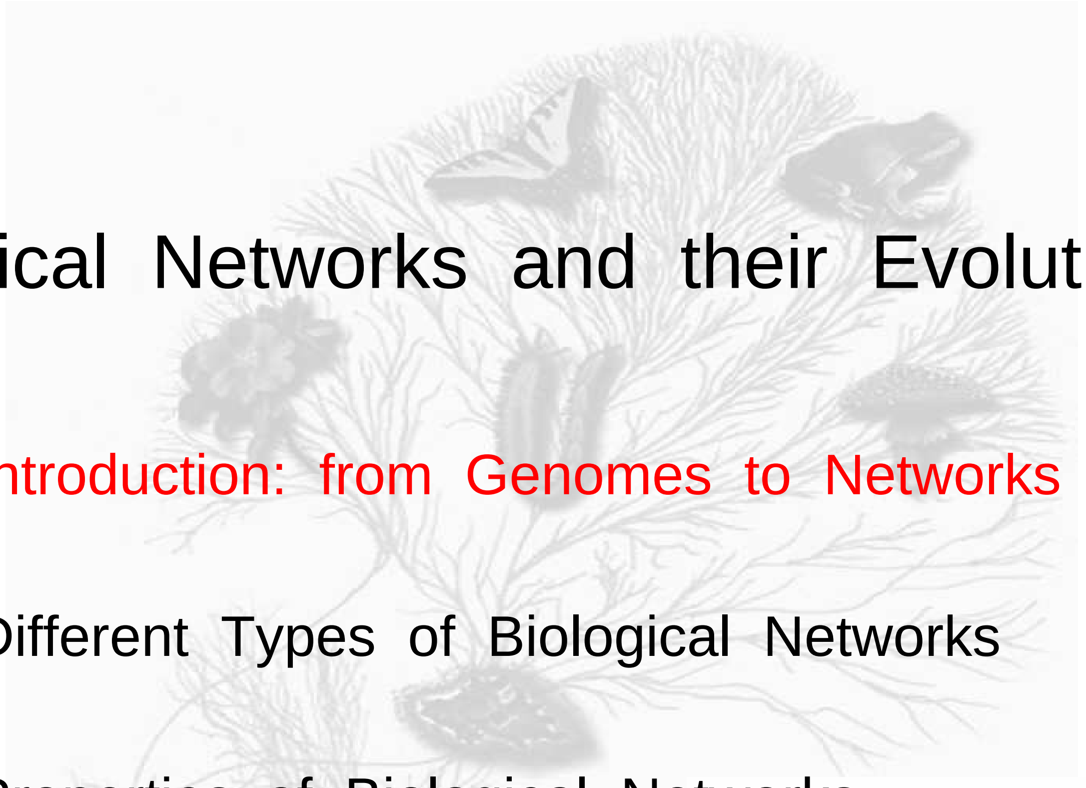
IV. Network Evolution: from Genes to Organisms

V. Models of Biological Network Evolution



Hervé ISAMBERT, Institut Curie, Paris

Biological Networks and their Evolution



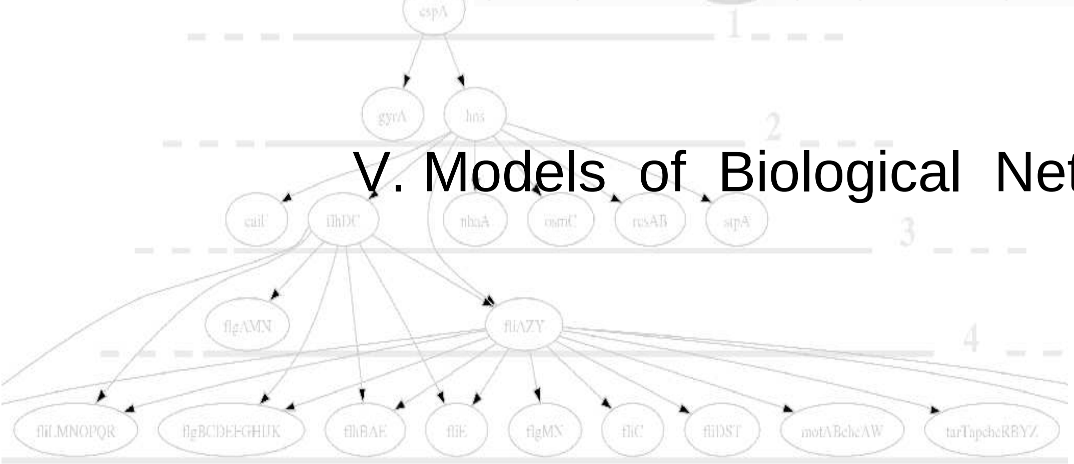
I. Introduction: from Genomes to Networks

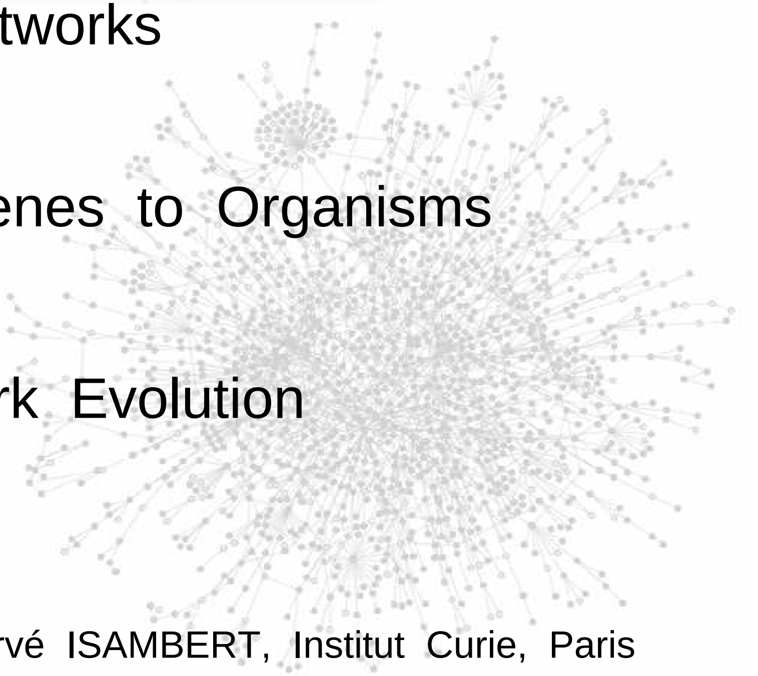
II. Different Types of Biological Networks

III. Properties of Biological Networks

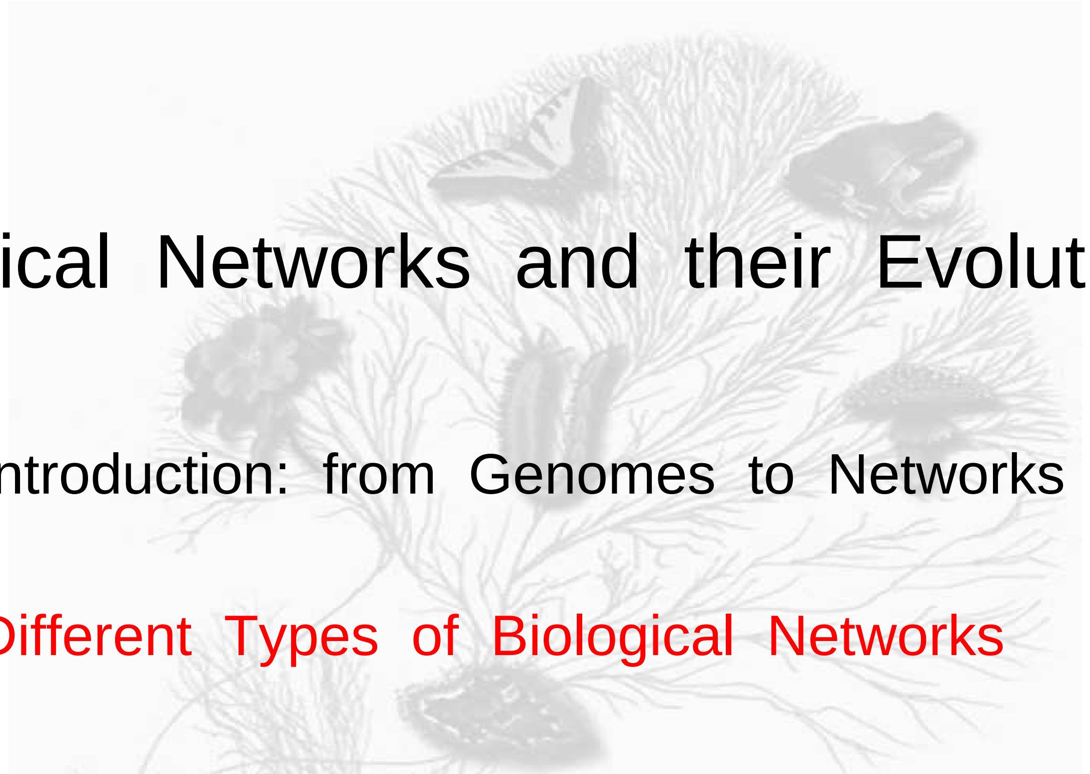
IV. Network Evolution: from Genes to Organisms

V. Models of Biological Network Evolution





Hervé ISAMBERT, Institut Curie, Paris



Biological Networks and their Evolution

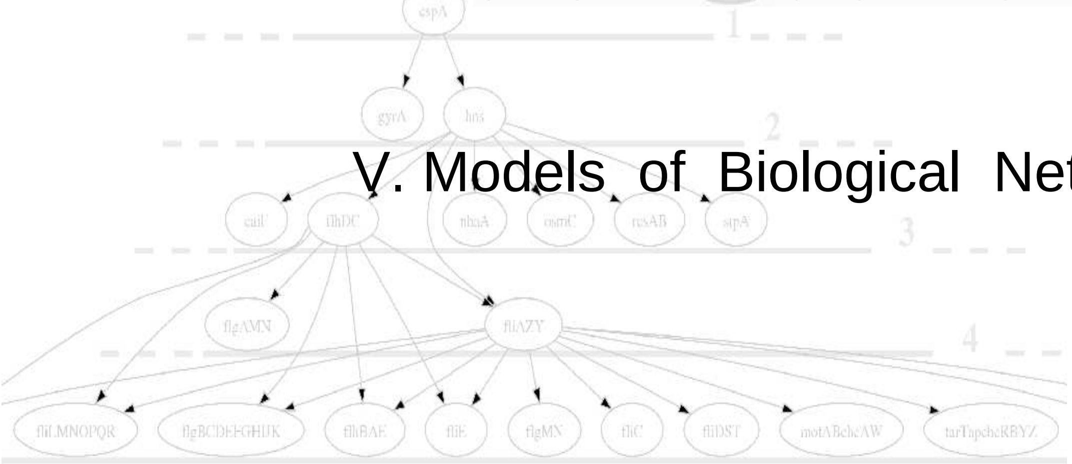
I. Introduction: from Genomes to Networks

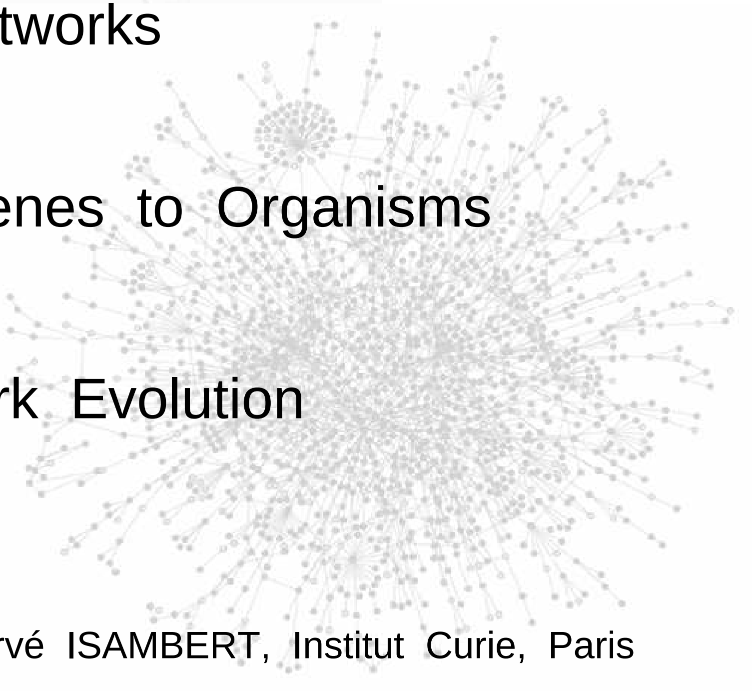
II. Different Types of Biological Networks

III. Properties of Biological Networks

IV. Network Evolution: from Genes to Organisms

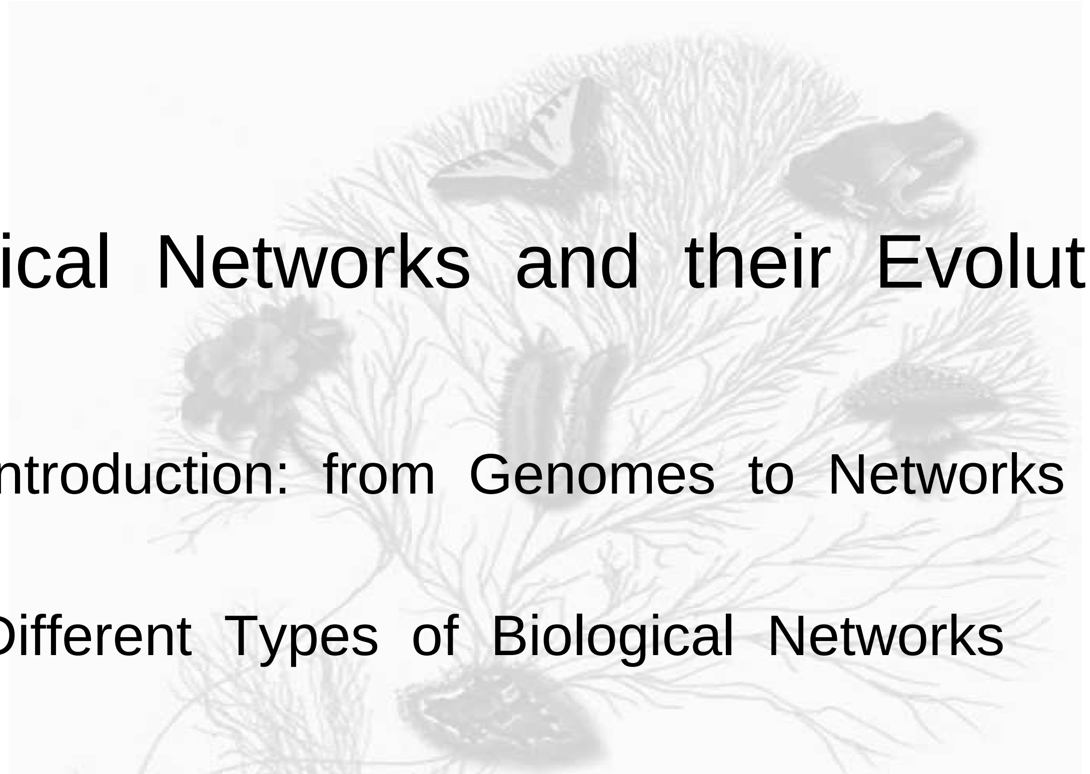
V. Models of Biological Network Evolution

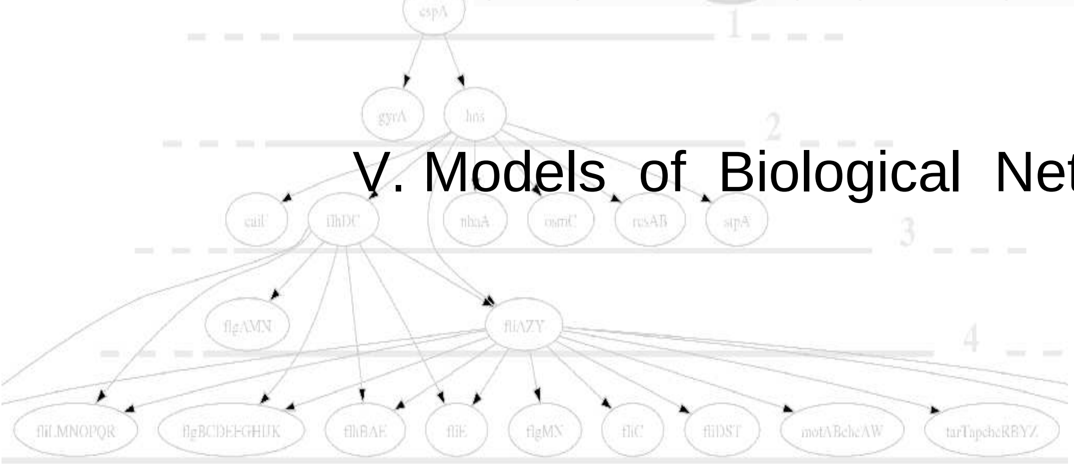


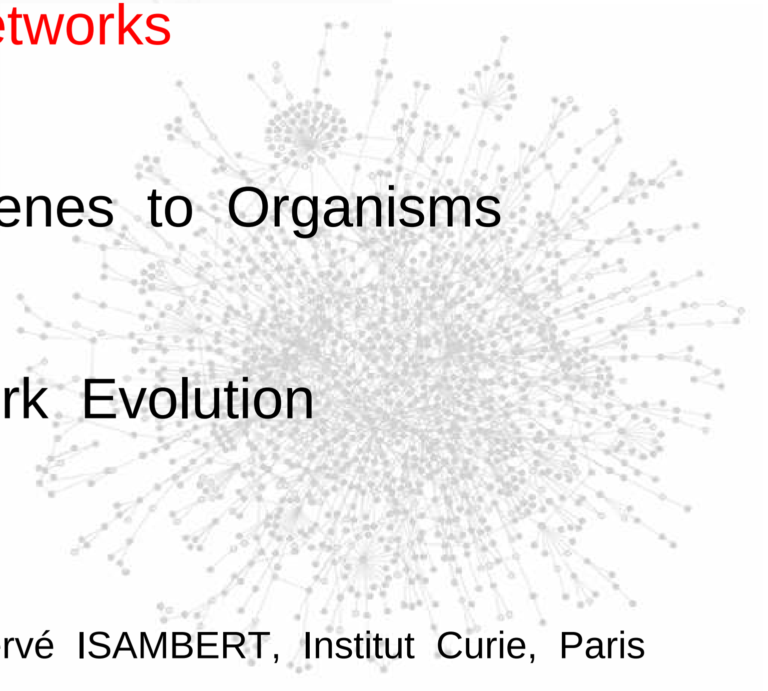


Hervé ISAMBERT, Institut Curie, Paris

Biological Networks and their Evolution







I. Introduction: from Genomes to Networks

II. Different Types of Biological Networks

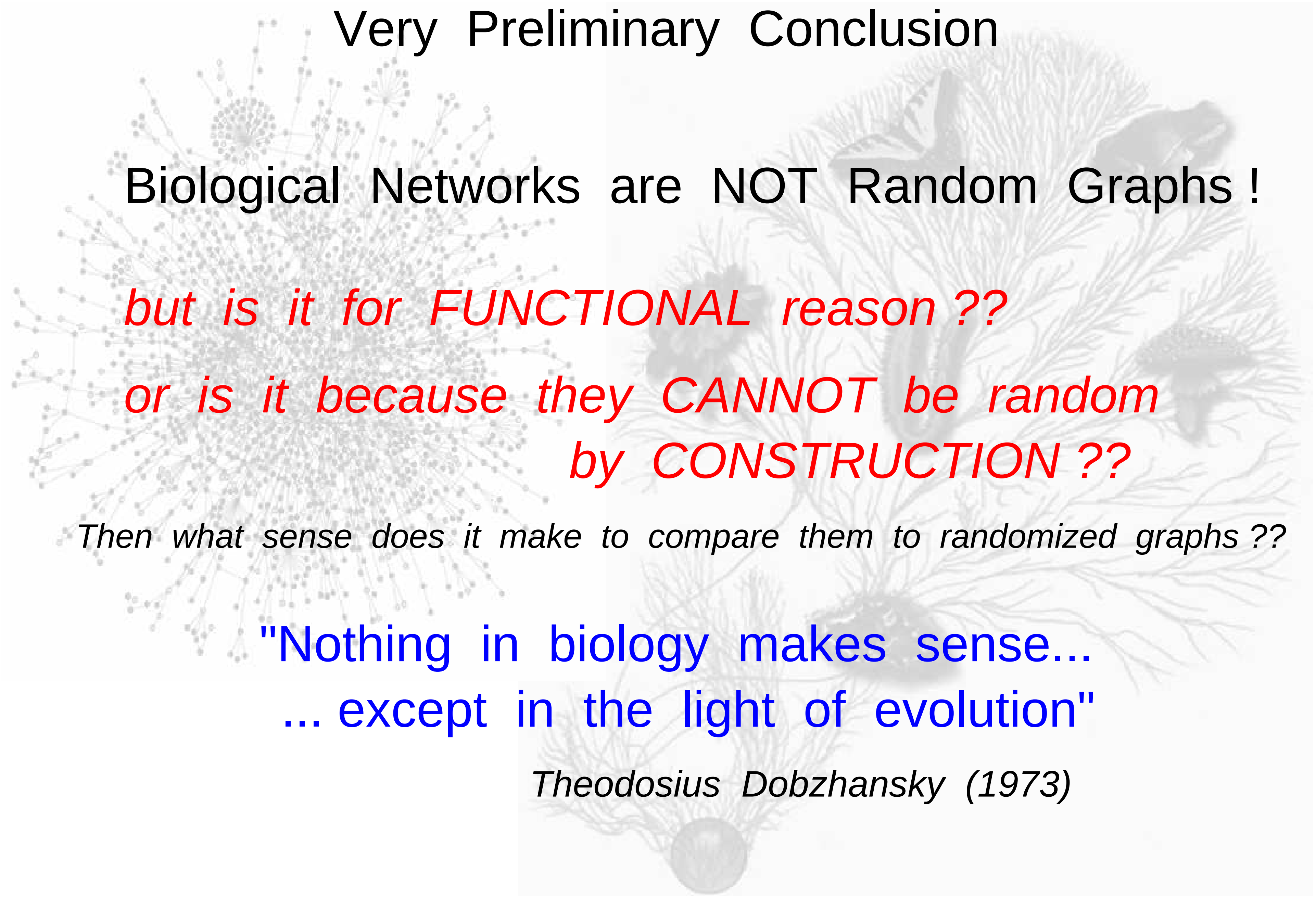
III. Properties of Biological Networks

IV. Network Evolution: from Genes to Organisms

V. Models of Biological Network Evolution

Hervé ISAMBERT, Institut Curie, Paris

Very Preliminary Conclusion



Biological Networks are NOT Random Graphs !

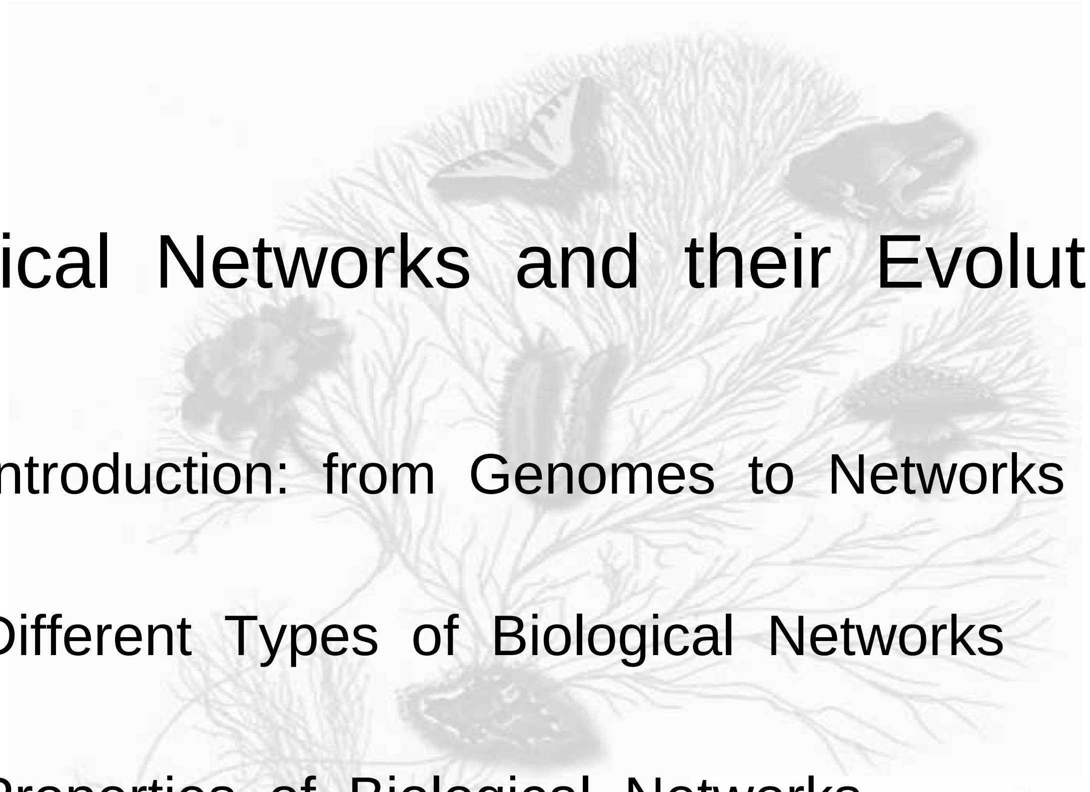
but is it for FUNCTIONAL reason ??

or is it because they CANNOT be random
by CONSTRUCTION ??

Then what sense does it make to compare them to randomized graphs ??

"Nothing in biology makes sense...
... except in the light of evolution"

Theodosius Dobzhansky (1973)



Biological Networks and their Evolution

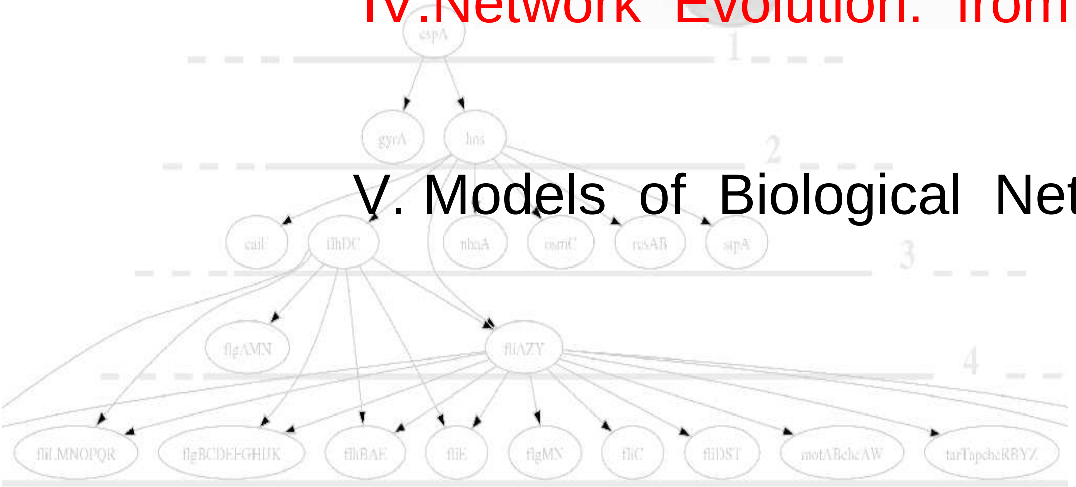
I. Introduction: from Genomes to Networks

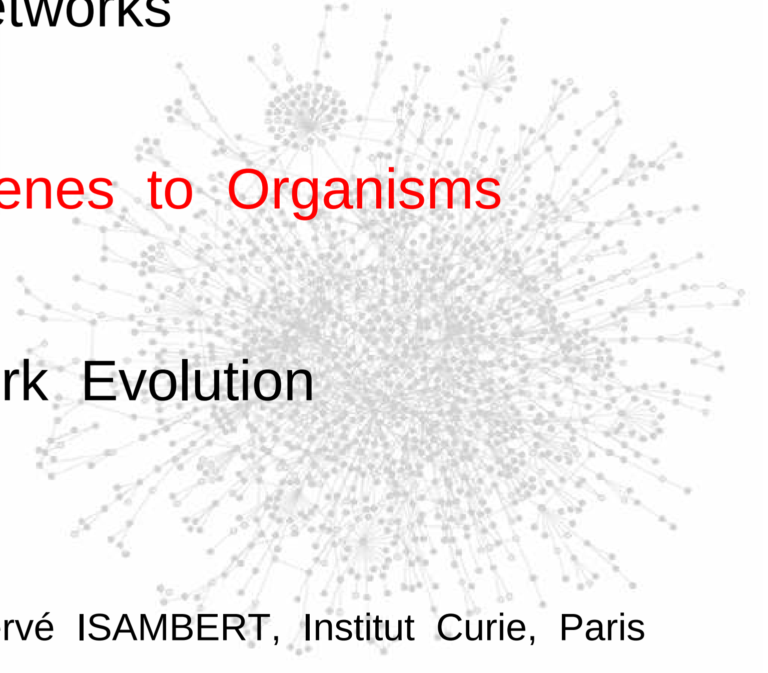
II. Different Types of Biological Networks

III. Properties of Biological Networks

IV. Network Evolution: from Genes to Organisms

V. Models of Biological Network Evolution





Hervé ISAMBERT, Institut Curie, Paris

Mechanisms of Genome Evolution

- Random mutations (nucleotide substitutions)
  - synonymous (the same aminoacid)
  - nonsynonymous (a different aminoacid)
  - nonsense codon modifications (Stop)
- Recombinations – crossing over/gene conversion
  - exchange of homologous sequences between homologous chromosomes
  - non reciprocal exchange (loss of one of the variant sequences)
- Deletions and Insertions – unequal crossing over
  gene deletion and local gene duplication → *shuffling of protein domains*
- Global gene duplication – whole genome duplication
  - replication of chromosomes without segregation

Long suspected... recently proved!

Mechanisms of Genome Evolution

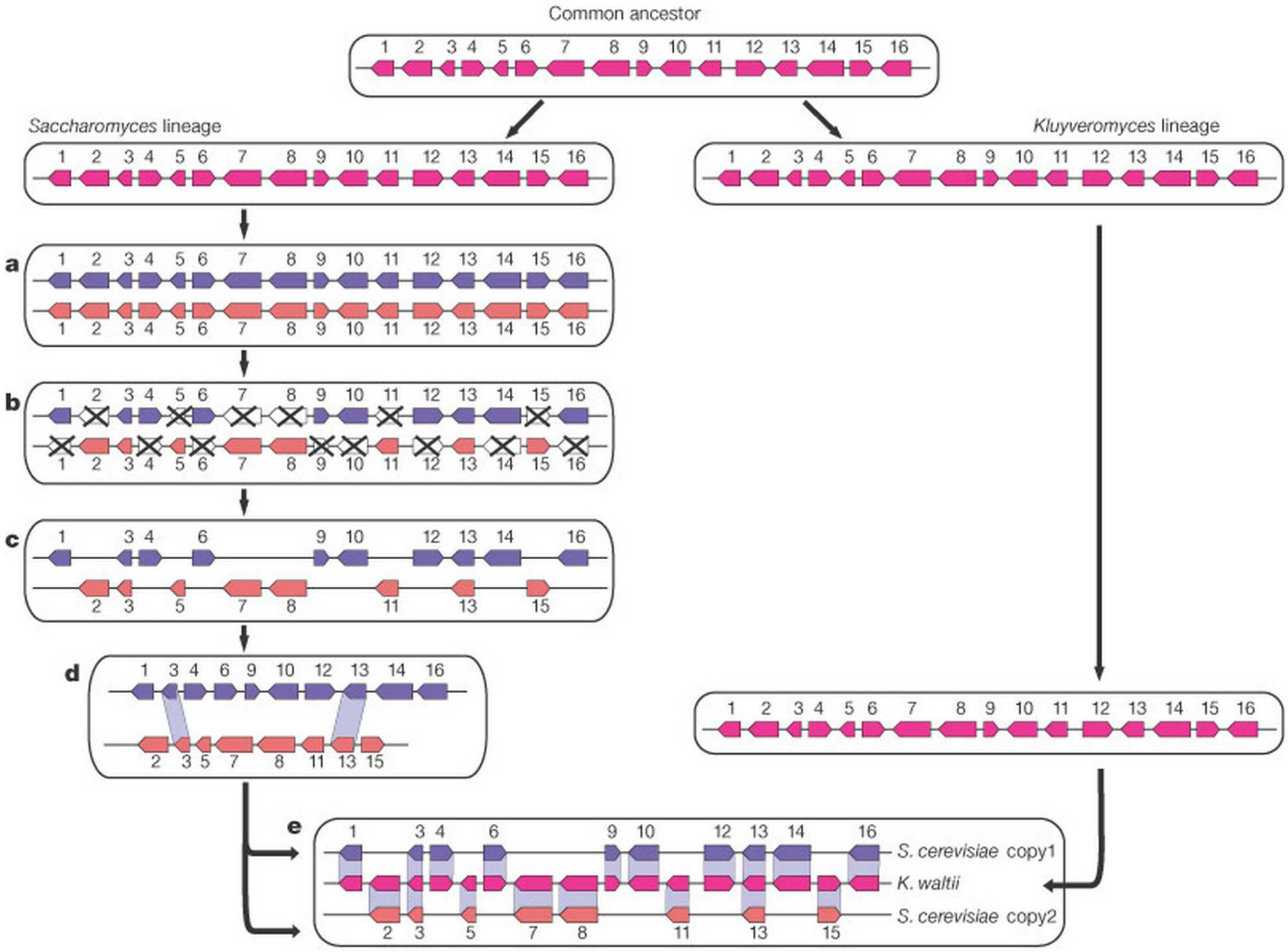
- Random mutations (nucleotide substitutions)
  - synonymous (the same aminoacid)
  - nonsynonymous (a different aminoacid)
  - nonsense codon modifications (Stop)
- Recombinations – crossing over/gene conversion
  - exchange of homologous sequences between homologous chromosomes
  - non reciprocal exchange (loss of one of the variant sequences)
- Deletions and Insertions – unequal crossing over
  gene deletion and local gene duplication → *shuffling of protein domains*
- Global gene duplication – whole genome duplication
  - replication of chromosomes without segregation

Recently proved!!!

Implies important genetic modifications!!

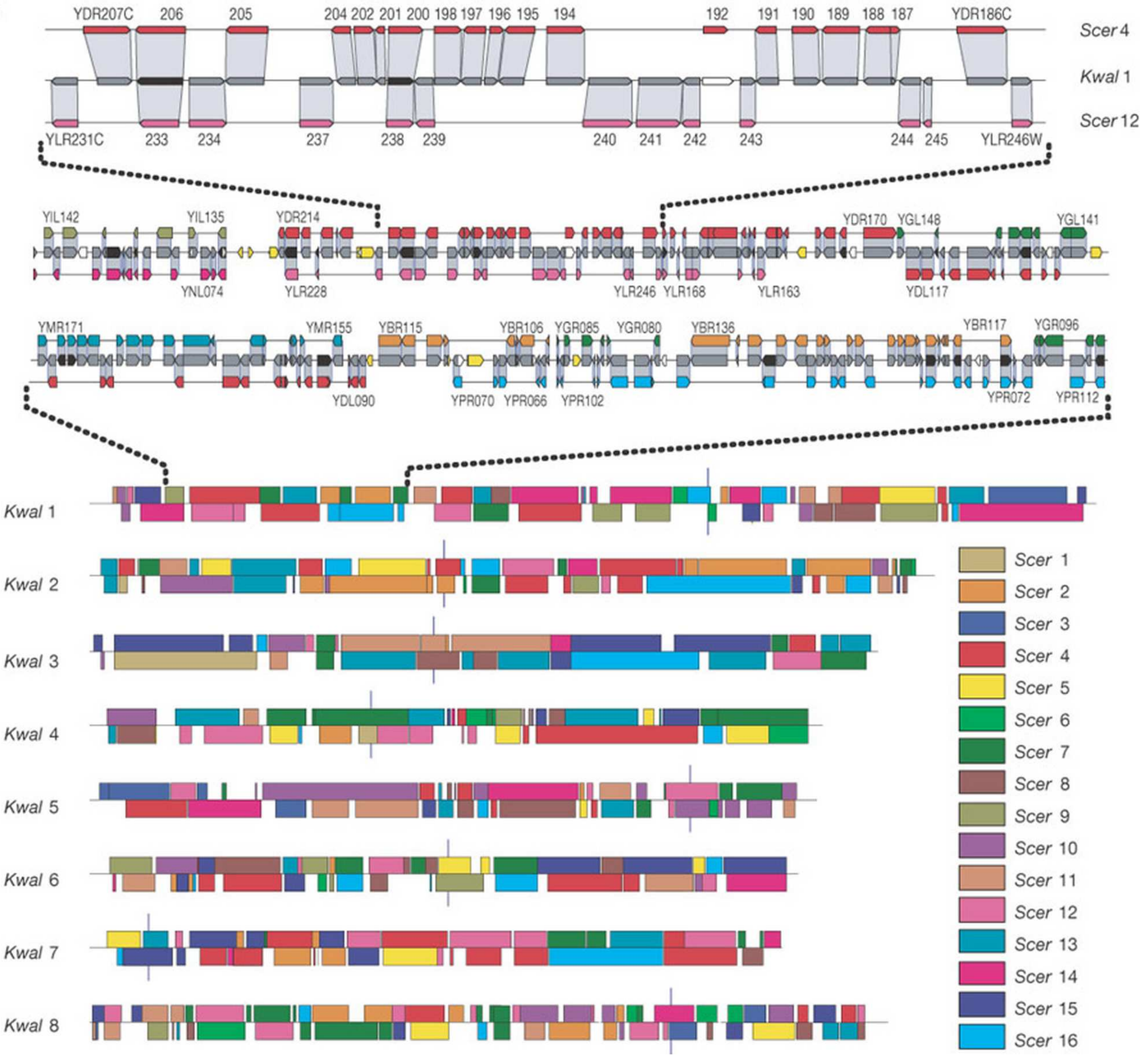
Whole Genome Duplication in Yeast Genome

Kellis et al. 2004

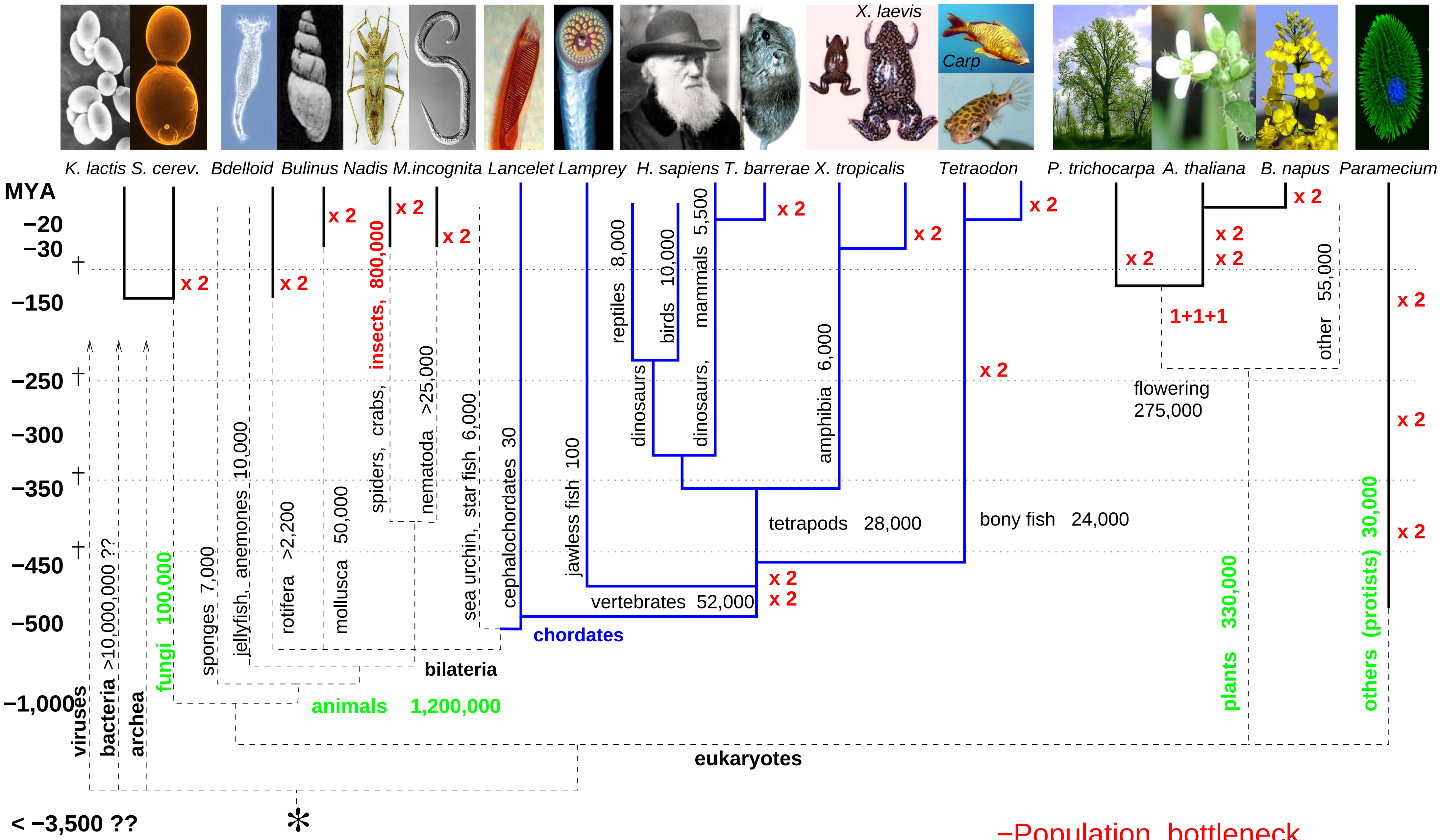


Whole Genome Duplication in Yeast Genome

Kellis et al. 2004



Whole Genome Duplications in Evolution

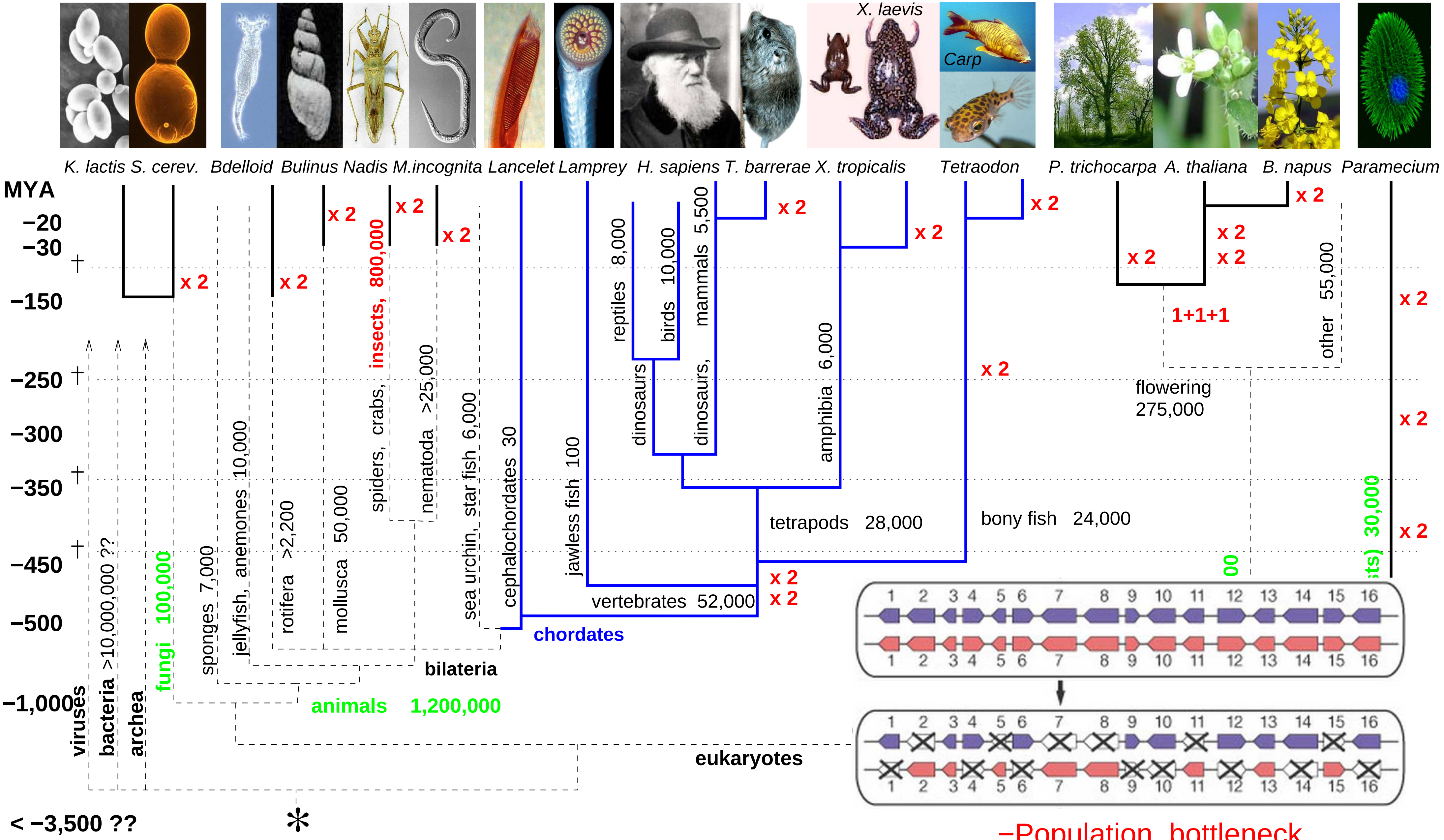


Whole Genome Duplications promote:

−Population bottleneck

−Speciation events

Whole Genome Duplications in Evolution

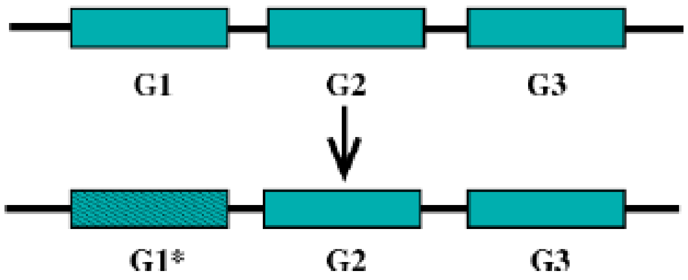


Whole Genome Duplications promote:

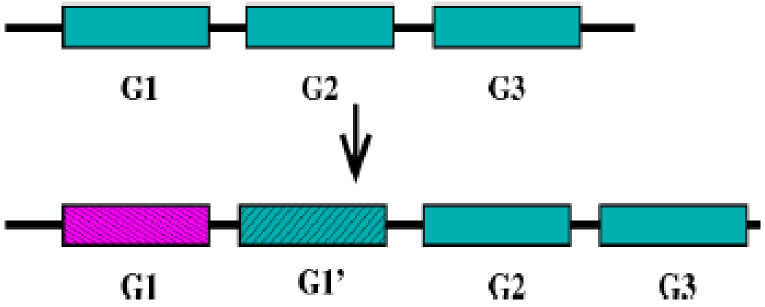
−Population bottleneck
−Speciation events

Evolution by Duplication Divergence

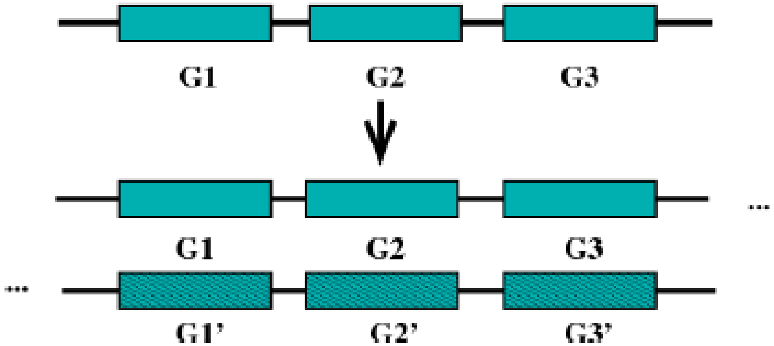
- Point mutations (substitutions)
  continuously occuring
  10^{-3} fixed events/ nucleotide / MY



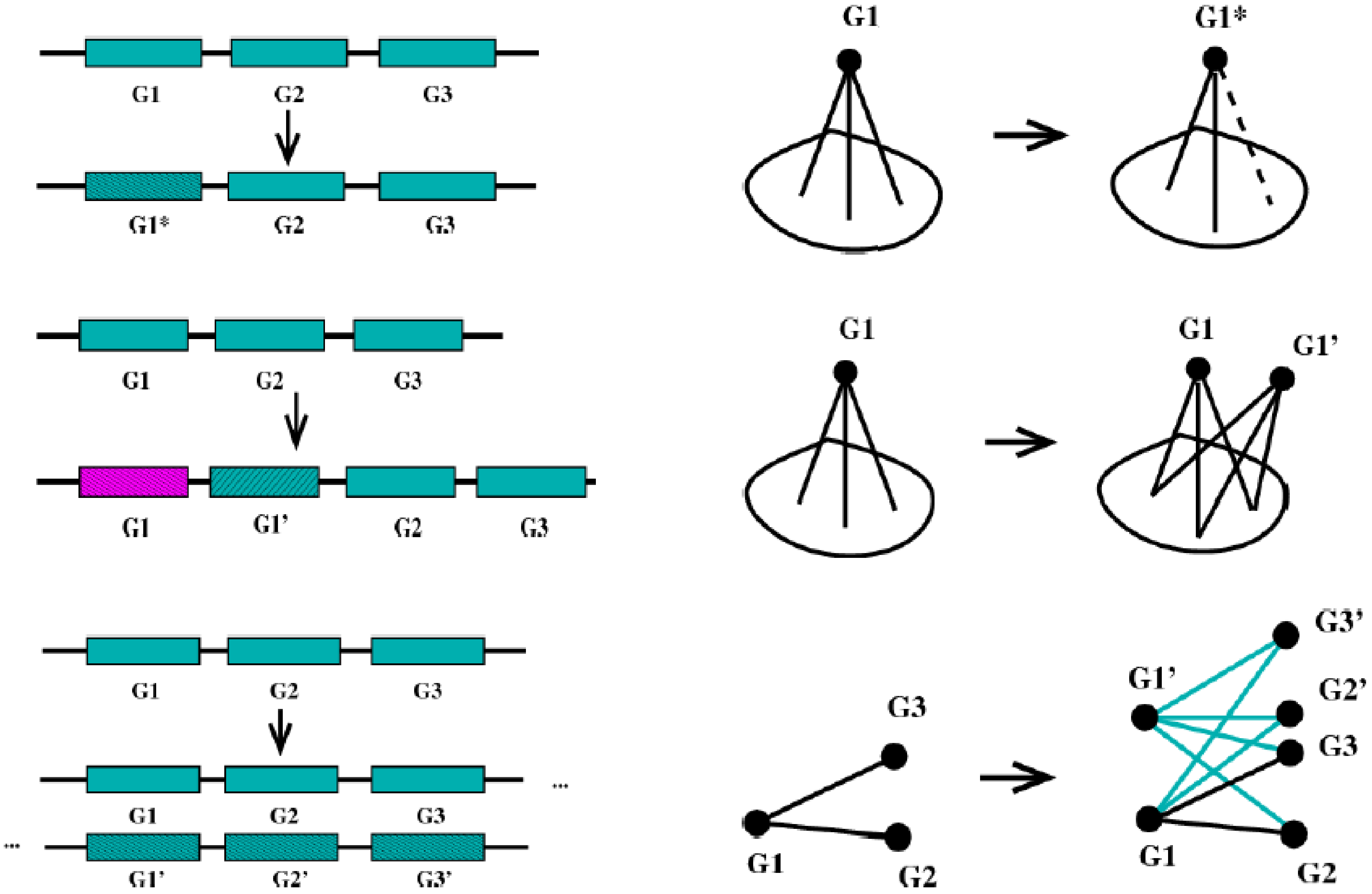
- Duplication of single genes
  and groupes of genes
  high frequency at
  evolutionary timescale
  10^{-3} fixed events/ gene /MY



- Duplication of chromosomes
  and of whole genome
  rare events
  1 fixed event / 100-200 MY=10^{-3} events/ gene /MY



Evolution by Duplication Divergence

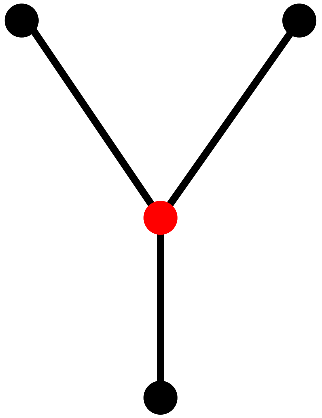


Evidence of Genome Duplication on PPI network

Yeast 6000 genes : 4000 in PPI network

500 duplicates from WGD : 250 with both duplicates in available PPI network

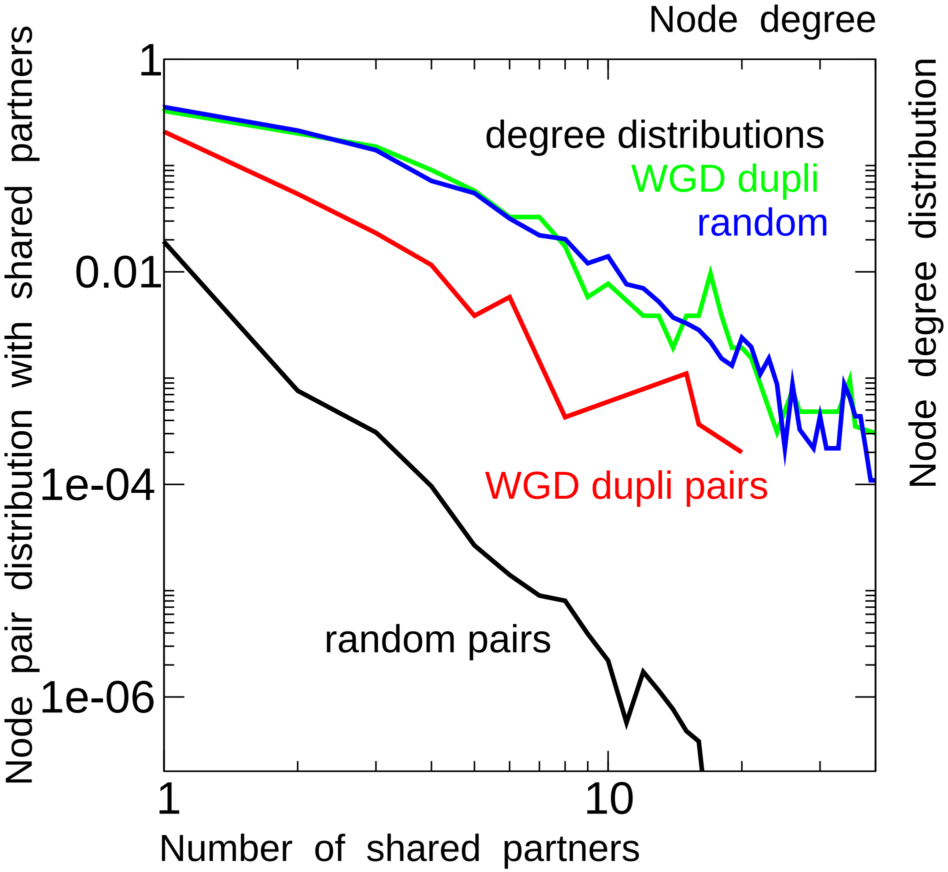
PPI network



Proba to share k+ partners :

k=1 WGD dupli > 20 x random pairs

k=10 WGD dupli > 1,000 x random pairs

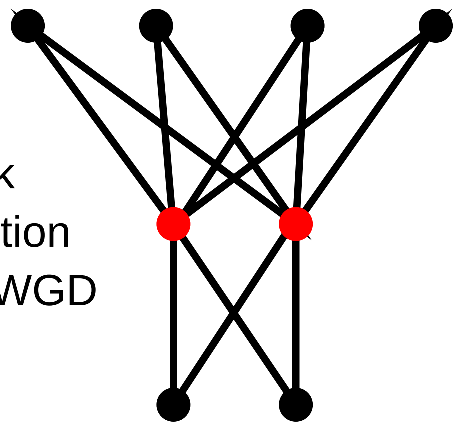


Evidence of Genome Duplication on PPI network

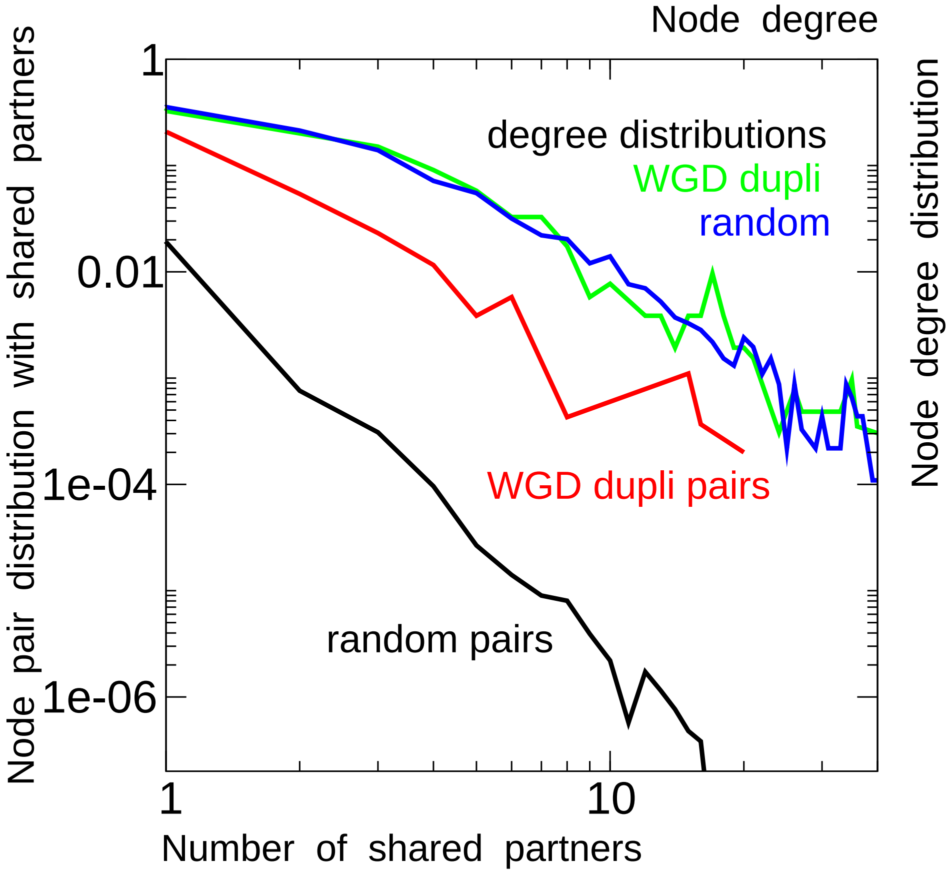
Yeast 6000 genes : 4000 in PPI network

500 duplicates from WGD : 250 with both duplicates in available PPI network

PPI network



Network
Duplication
under WGD



Proba to share k+ partners :

k=1 WGD dupli > 20 x random pairs

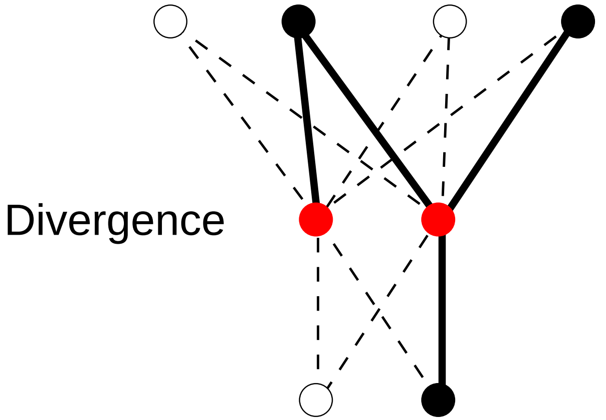
k=10 WGD dupli > 1,000 x random pairs

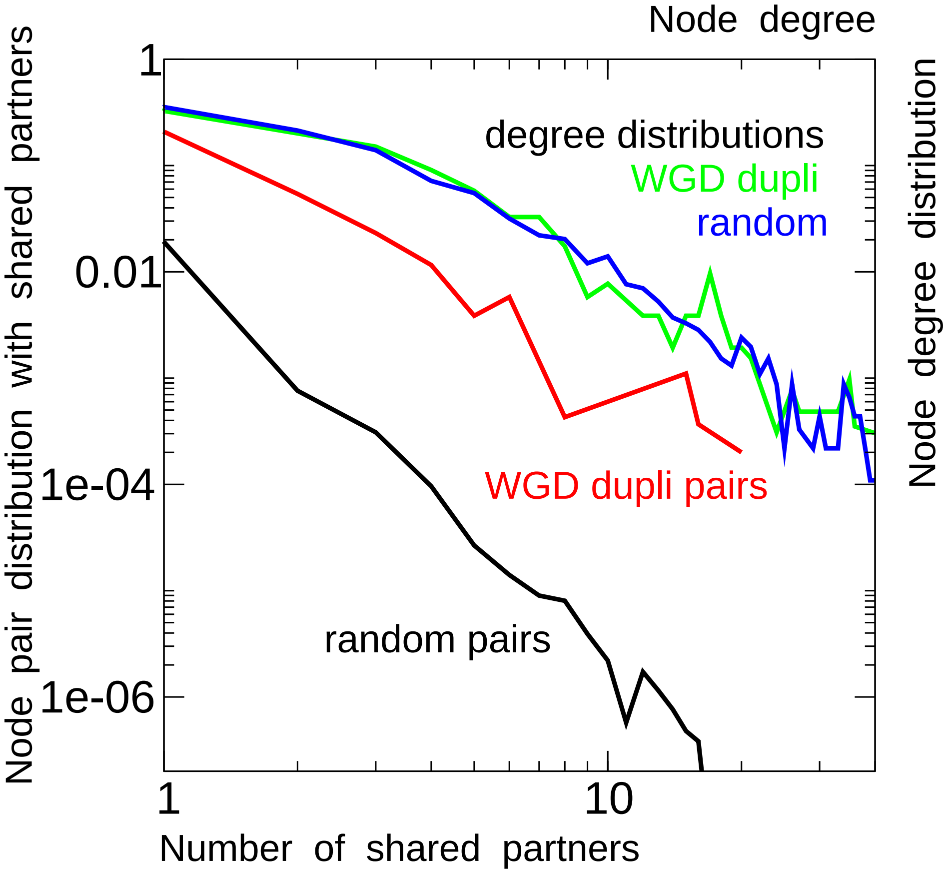
Evidence of Genome Duplication on PPI network

Yeast 6000 genes : 4000 in PPI network

500 duplicates from WGD : 250 with both duplicates in available PPI network

PPI network





Proba to share k+ partners :

k=1 WGD dupli > 20 x random pairs

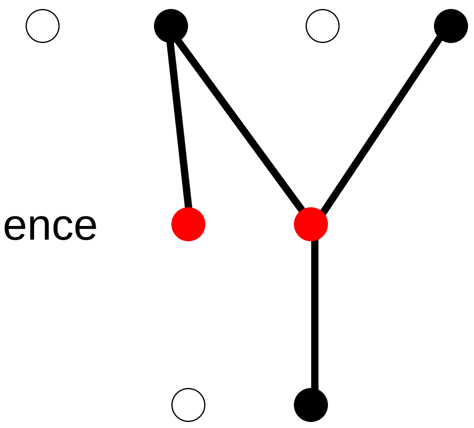
k=10 WGD dupli > 1,000 x random pairs

Evidence of Genome Duplication on PPI network

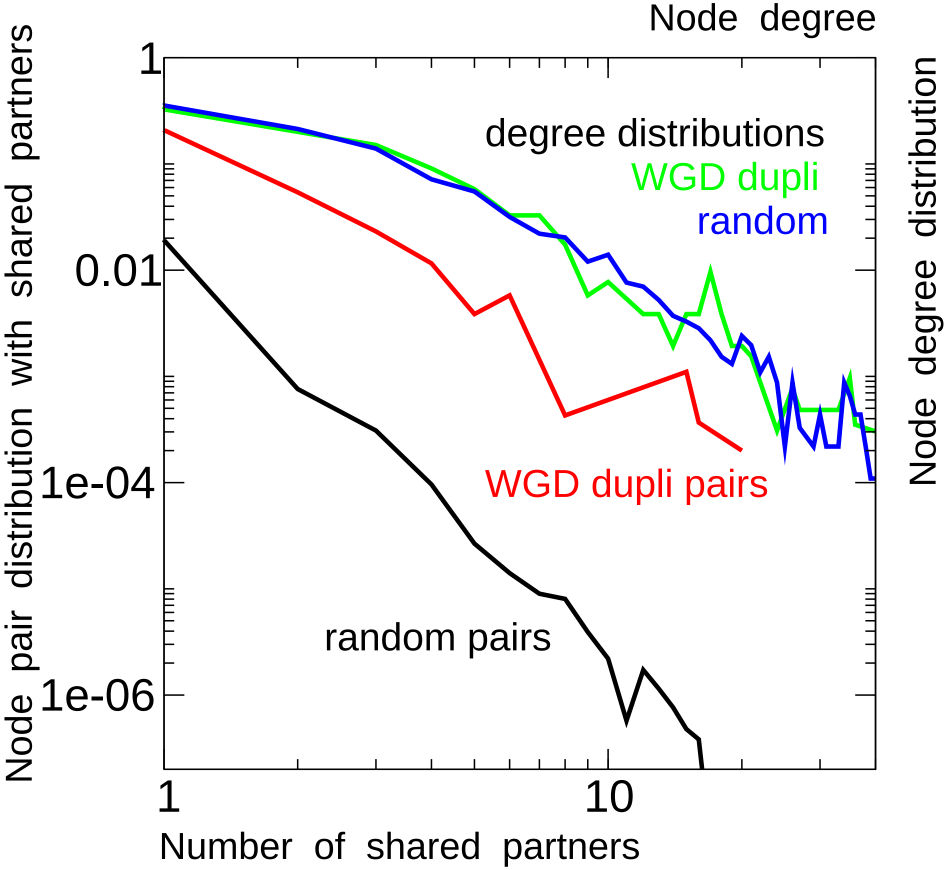
Yeast 6000 genes : 4000 in PPI network

500 duplicates from WGD : 250 with both duplicates in available PPI network

PPI network



Divergence



Proba to share k+ partners :

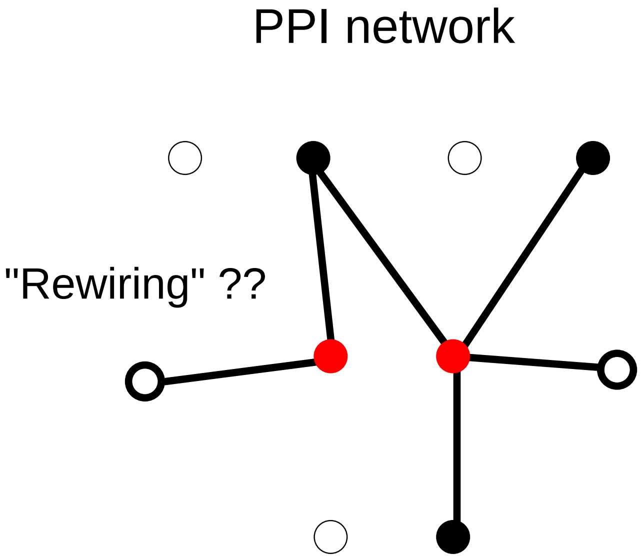
k=1 WGD dupli > 20 x random pairs

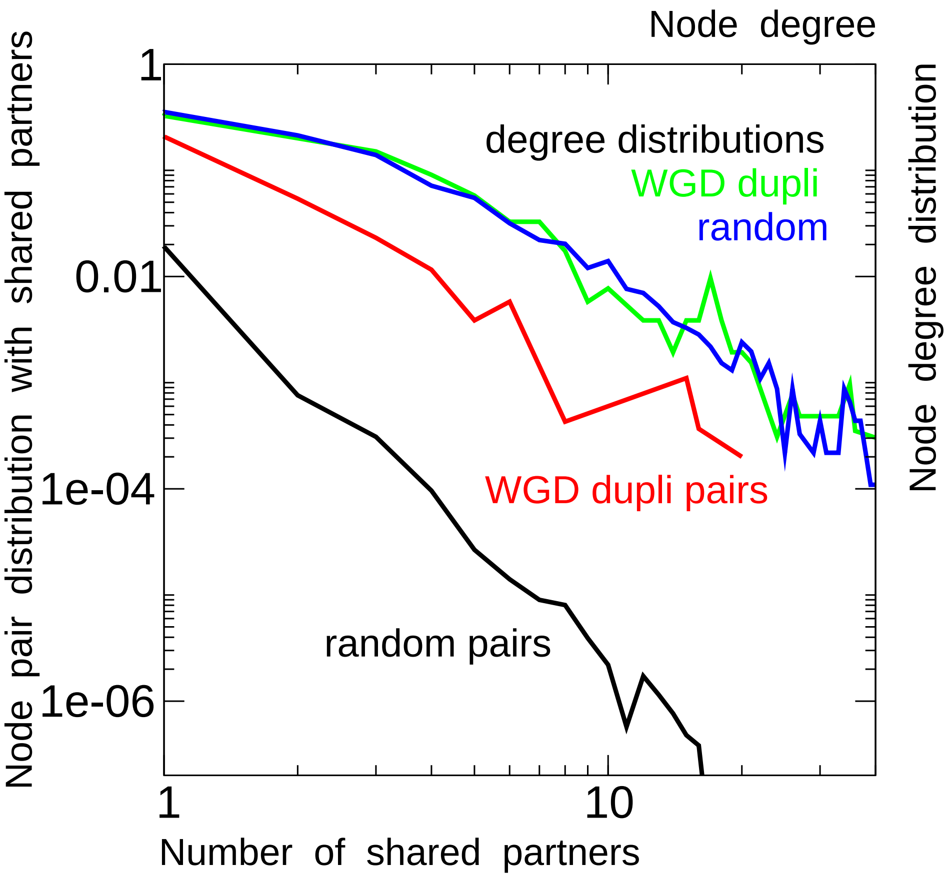
k=10 WGD dupli > 1,000 x random pairs

Evidence of Genome Duplication on PPI network

Yeast 6000 genes : 4000 in PPI network

500 duplicates from WGD : 250 with both duplicates in available PPI network





Proba to share k+ partners :

k=1 WGD dupli > 20 x random pairs

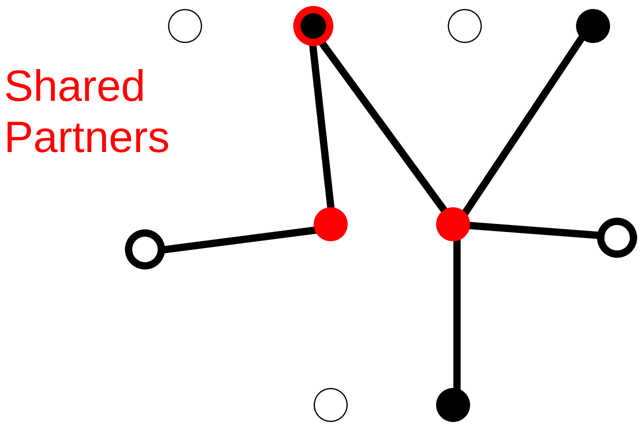
k=10 WGD dupli > 1,000 x random pairs

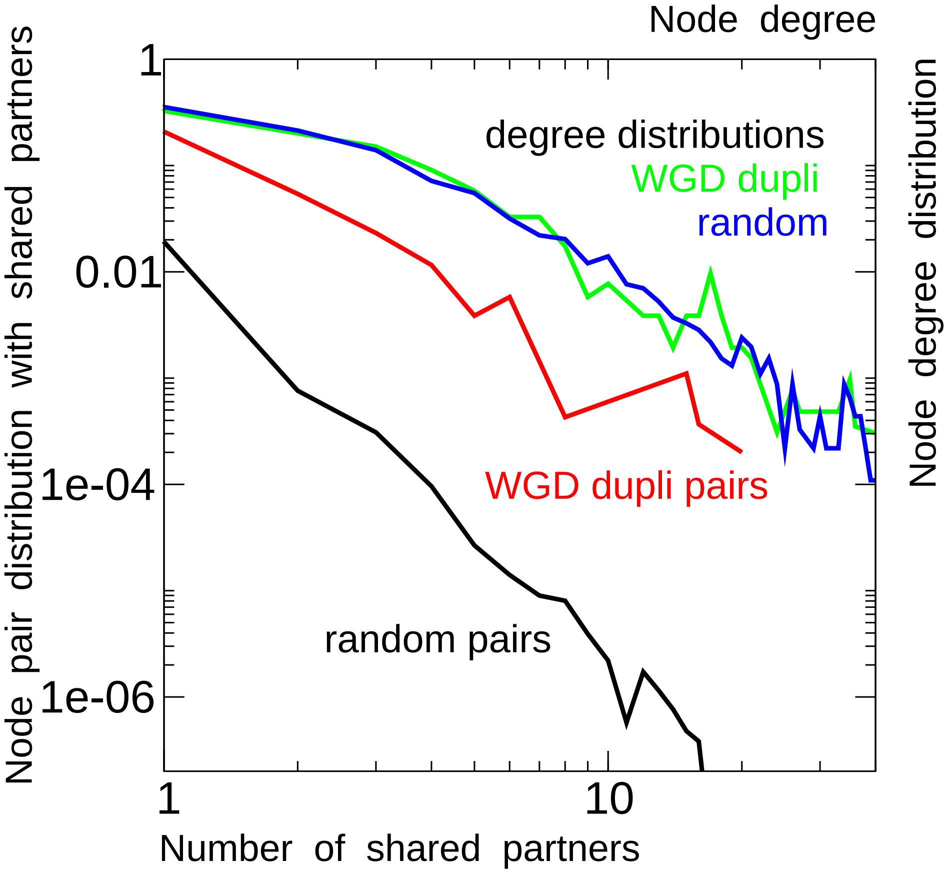
Evidence of Genome Duplication on PPI network

Yeast 6000 genes : 4000 in PPI network

500 duplicates from WGD : 250 with both duplicates in available PPI network

PPI network



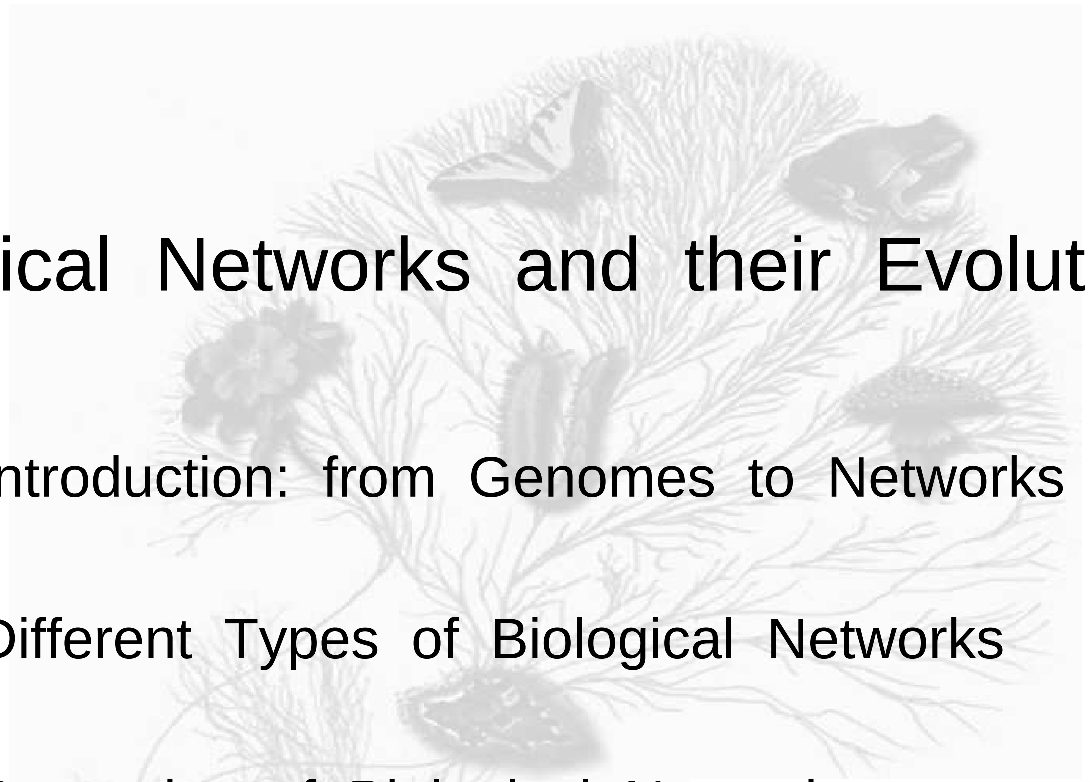


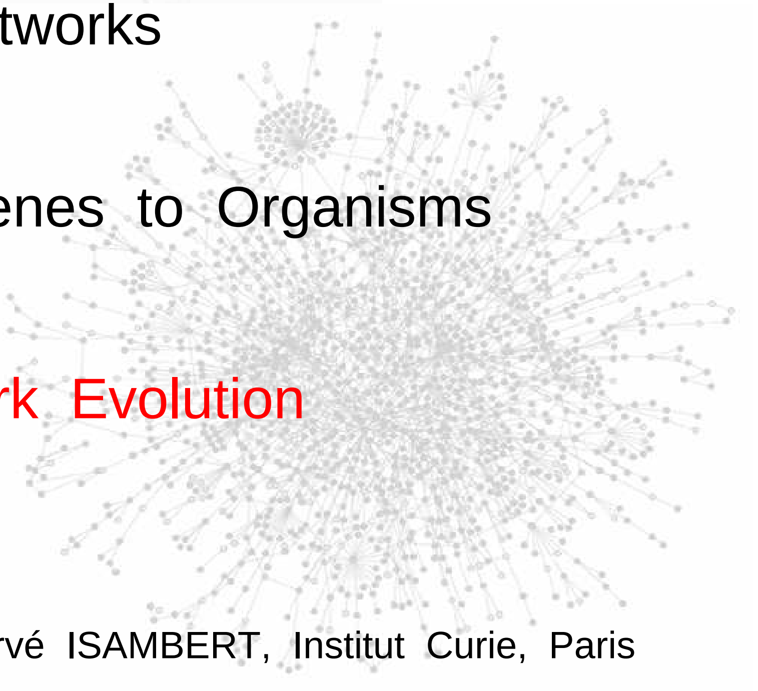
Proba to share k+ partners :

k=1 WGD dupli > 20 x random pairs

k=10 WGD dupli > 1,000 x random pairs

Biological Networks and their Evolution





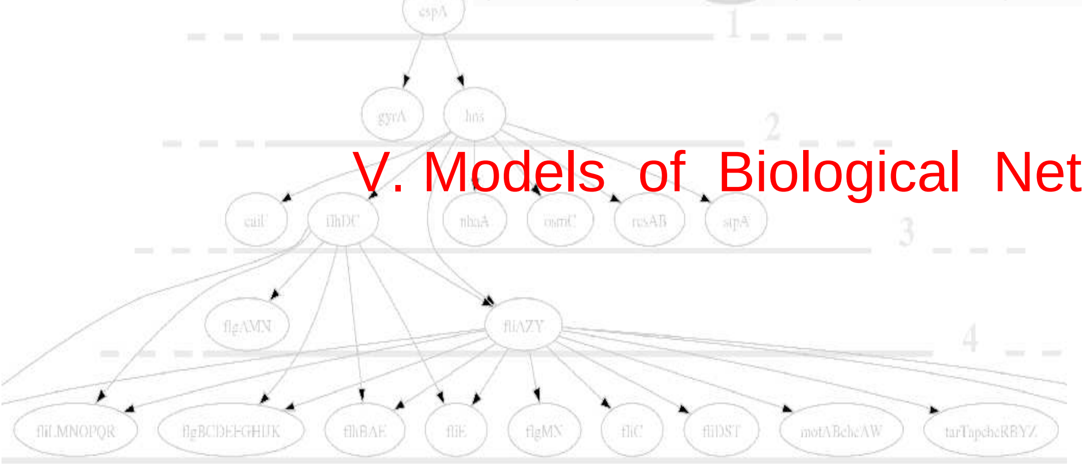
I. Introduction: from Genomes to Networks

II. Different Types of Biological Networks

III. Properties of Biological Networks

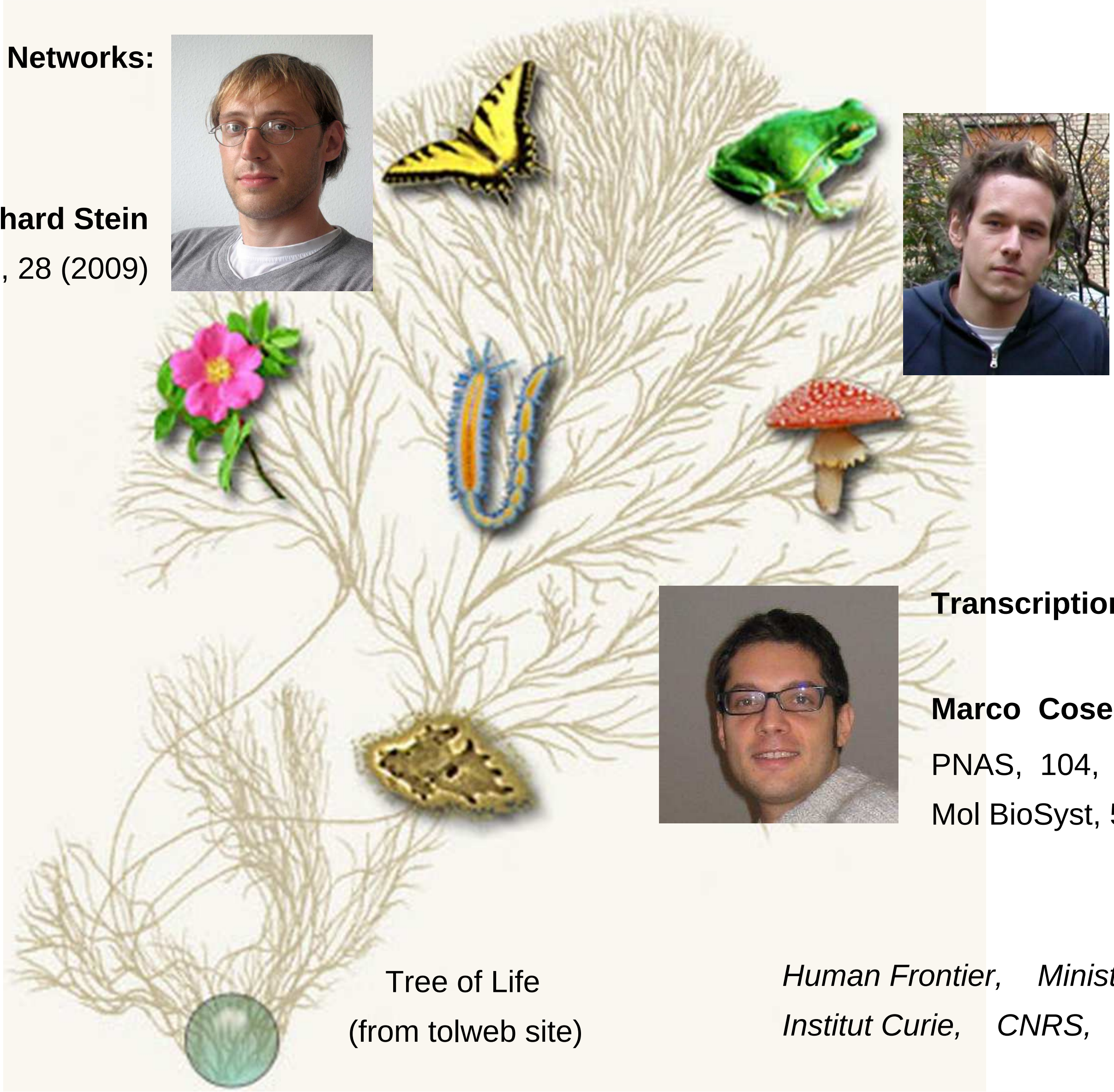
IV. Network Evolution: from Genes to Organisms

V. Models of Biological Network Evolution



Hervé ISAMBERT, Institut Curie, Paris

Evolution and Topology of Large Biological Networks



Signaling Networks:

Richard Stein
Biol Direct, 4, 28 (2009)

Prot−Prot Interaction Networks:

Kirill Evlampiev
PNAS, 105, 9863−9868 (2008)
BMC Syst Biol 1:49 (2007)

Transcription Networks:

Marco Cosentino Lagomarsino
PNAS, 104, 5516−5520 (2007)
Mol BioSyst, 5, 170−179 (2009)

Tree of Life
(from tolweb site)

Human Frontier, Ministere de la Recherche, ANR
Institut Curie, CNRS, Fondation PG de Gennes

Evolution of PPI Networks

Evolution of Protein-Protein Interaction Networks reflects evolution at the level of Genome

random mutations, local gene duplications, global gene duplications

→ Different models of PPI Network Evolution

local → evolution by local gene duplication-divergence process

Ispolatov, Krapivsky, Yuryev, *Phys. Rev. E* **71**, 061911 (2005)

→ Barabasi-Albert like models (preferential attachment)

Barabasi, Albert, *Science* **286** 509-512 (1999)

time−linear growth

global → evolution by whole genome duplication

Evlampiev, Isambert BMC Syst. Bio. (2007)

global+local → general duplication-divergence model

Evlampiev, Isambert PNAS USA (2008)

exponential growth

Evolution of PPI Networks

Evolution of Protein-Protein Interaction Networks reflects
evolution at the level of Genome

random mutations, local gene duplications, global gene duplications

→ Different models of PPI Network Evolution

local → evolution by local gene duplication-divergence process

Ispolatov, Krapivsky, Yuryev, *Phys. Rev. E* **71**, 061911 (2005)

→ Barabasi-Albert like models (preferential attachment)

Barabasi, Albert, *Science* **286** 509-512 (1999)

time−linear growth

global → evolution by whole genome duplication

Evlampiev, Isambert BMC Syst. Bio. (2007)

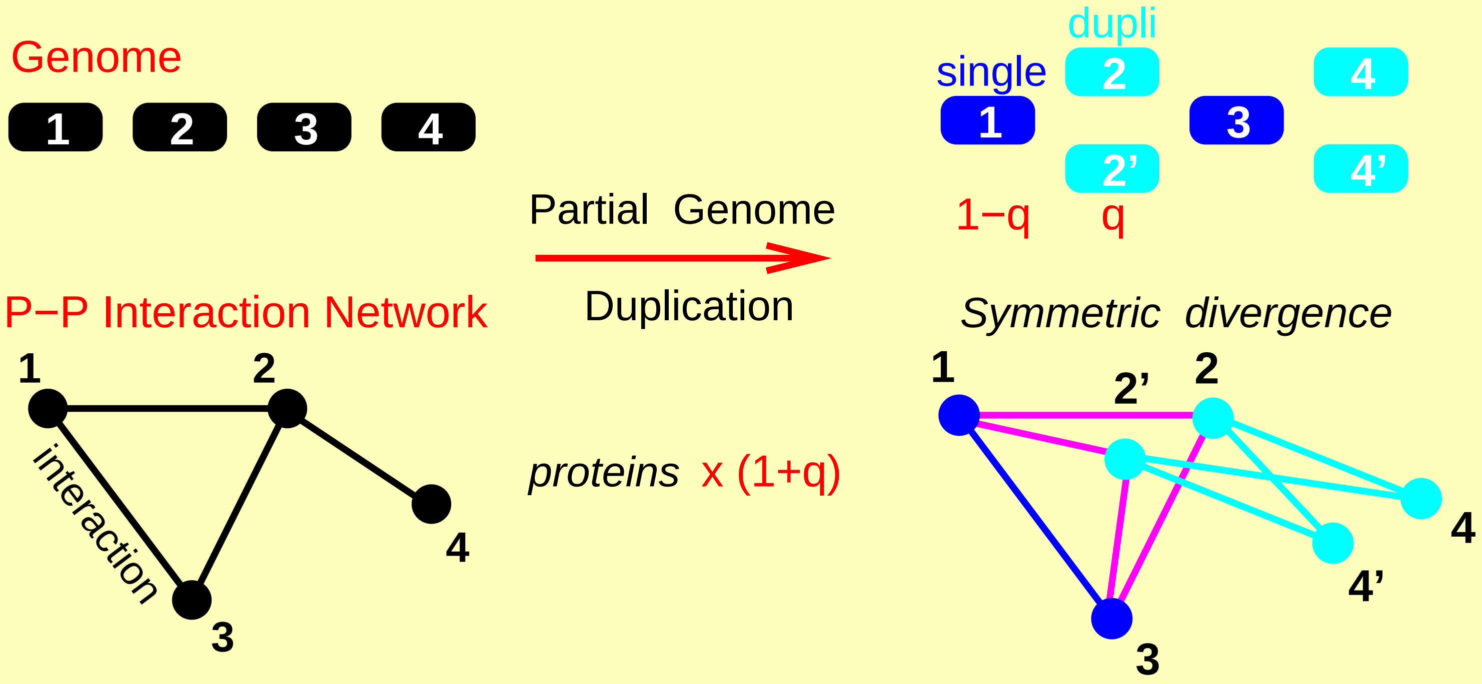
global+local → general duplication-divergence model

Evlampiev, Isambert PNAS USA (2008)

exponential growth

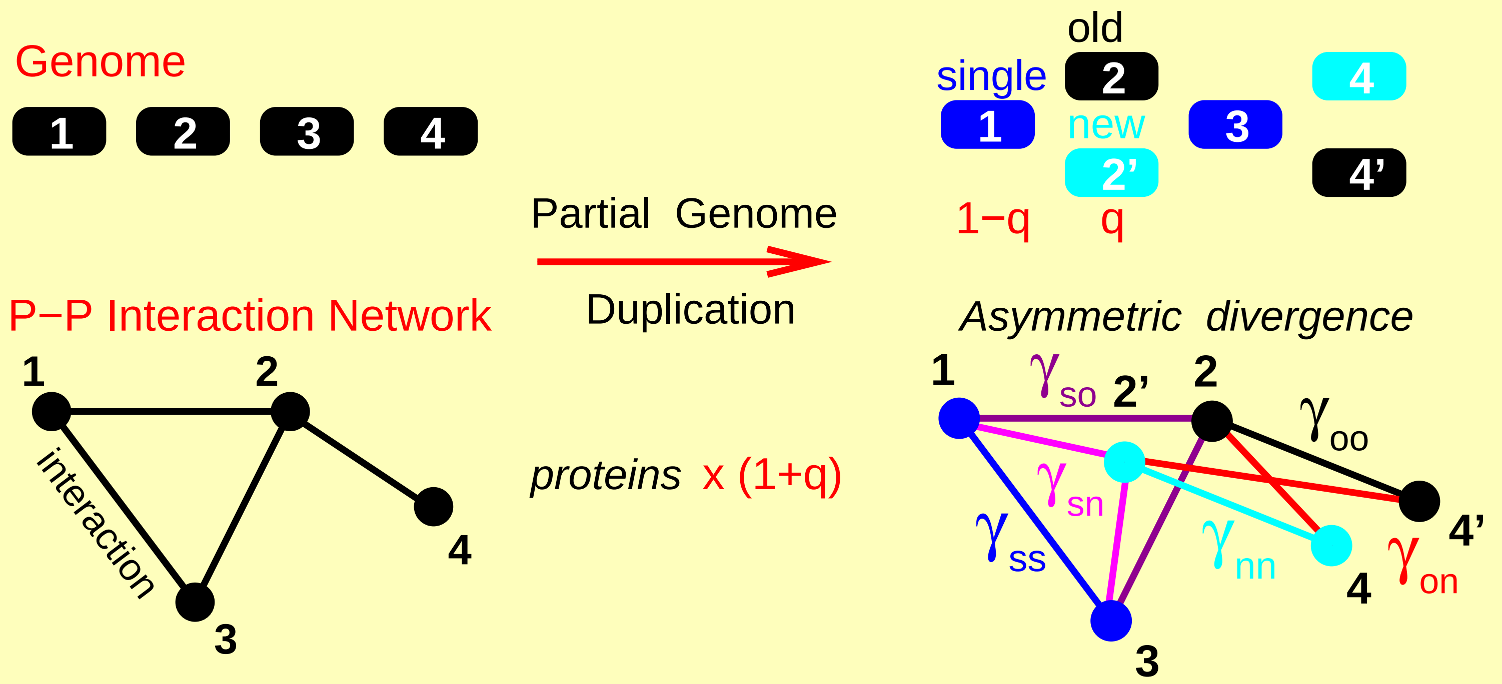
General Duplication−Divergence Model of PPI Network Evolution

GDD Model of PPI Network Evolution



General Duplication−Divergence Model of PPI Network Evolution

GDD Model of PPI Network Evolution



Genome

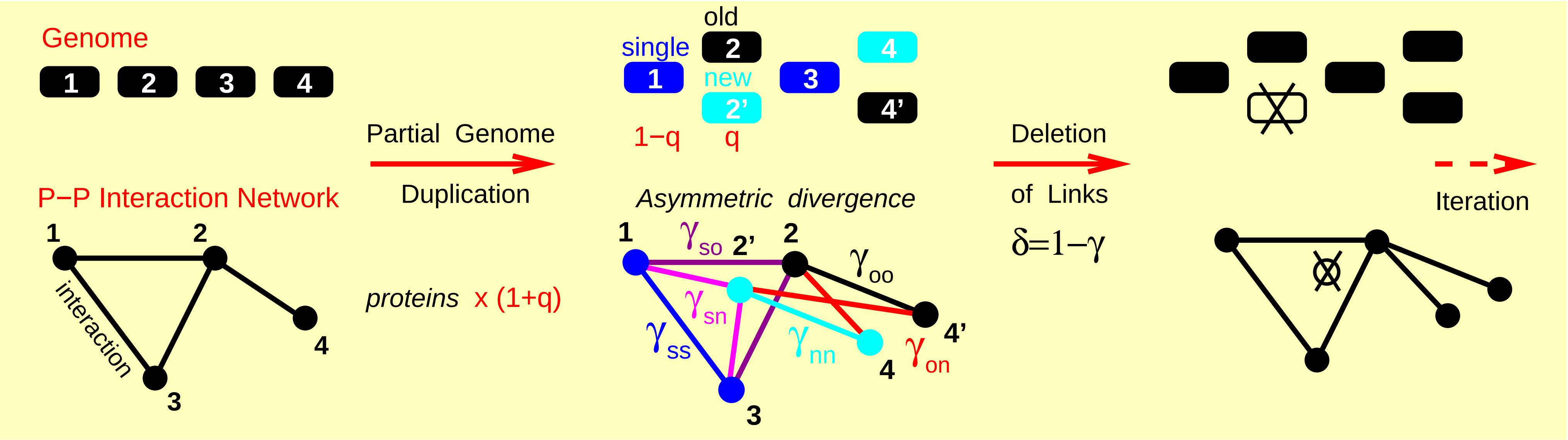
Partial Genome Duplication

P−P Interaction Network

proteins x (1+q)

Asymmetric divergence

General Duplication−Divergence Model of PPI Network Evolution



Genome

P−P Interaction Network

Partial Genome Duplication

proteins x (1+q)

Asymmetric divergence

Deletion of Links

$\delta=1-\gamma$

Iteration

Overview of the Evolutionary Regimes

Model can be solved theoretically in the asymptotic limit

Network evolutionary dynamics

vanishing

growing

Overview of the Evolutionary Regimes

Model can be solved theoretically in the asymptotic limit

- Network evolutionary dynamics
  - **vanishing**
  - **growing**
    - nonstationary
    - stationary

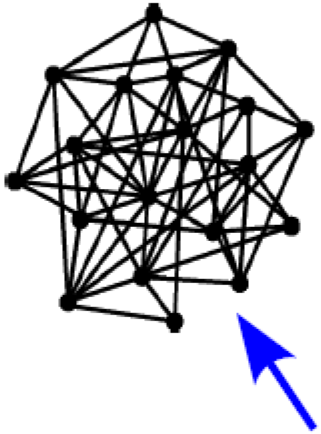
with respect to p(k)

Overview of the Evolutionary Regimes

Model can be solved theoretically in the asymptotic limit

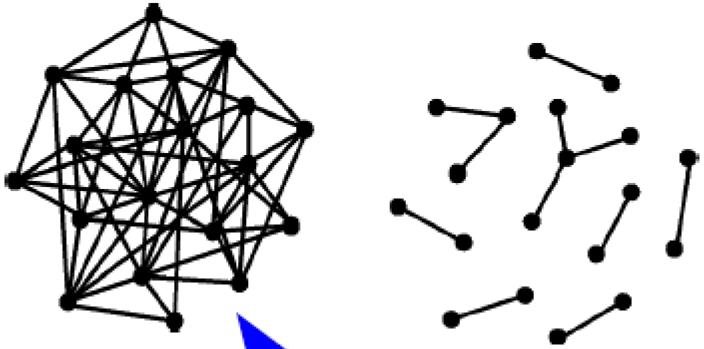
Network
evolutionary
dynamics

- **vanishing**
- **growing**
  - nonstationary → **dense**
  - stationary



Overview of the Evolutionary Regimes

Model can be solved theoretically in the asymptotic limit



Network evolutionary dynamics

vanishing

growing

nonstationary → **dense**

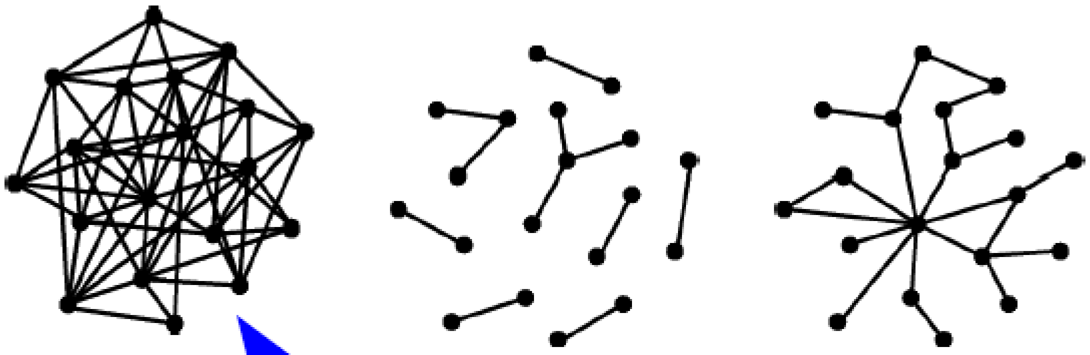
stationary

exponential distribution

lowly connected node

Overview of the Evolutionary Regimes

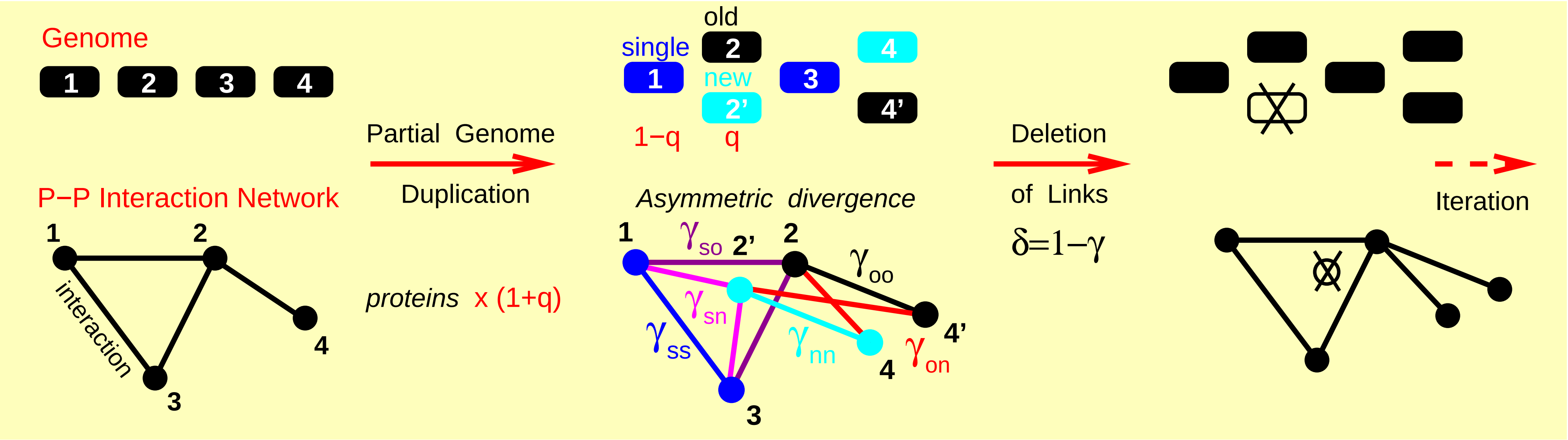
Model can be solved theoretically in the asymptotic limit



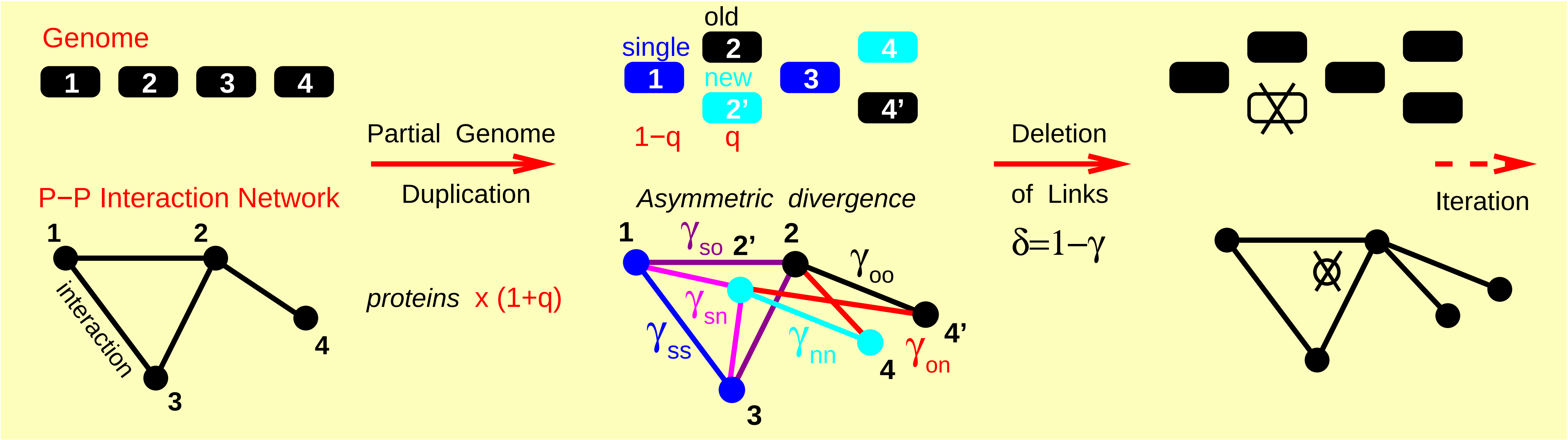
Network evolutionary dynamics
- **vanishing**
- **growing**
  - nonstationary → **dense**
  - stationary
    - **exponential distribution**
      lowly connected nodes
    - **scale free distribution**
      some highly connected nodes

Model is complex but can be described with few «global» parameters

General Duplication−Divergence Model of PPI Network Evolution

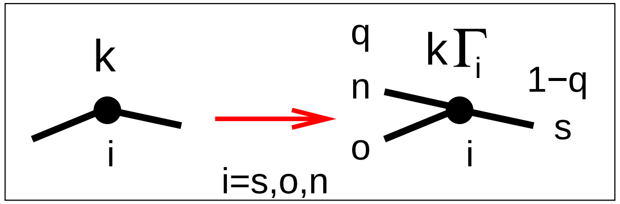


General Duplication−Divergence Model of PPI Network Evolution



Node Degree Growth Rate

$$\Gamma_i = (1-q)\gamma_{is} + q(\gamma_{io} + \gamma_{in})$$



Network Link Growth Rate

$$\Gamma = (1-q)\Gamma_s + q\,\Gamma_o + q\,\Gamma_n$$

1

$\Gamma > 1$ Growing Network

$\Gamma < 1$ Vanishing Network

Protein Conservation Index

$$M = (1-q)\Gamma_s + q\,\Gamma_o \quad < \quad \Gamma$$

2

$M > 1$ Conserved Network

$M < 1$ Nonconserved Network

Network Topology Index

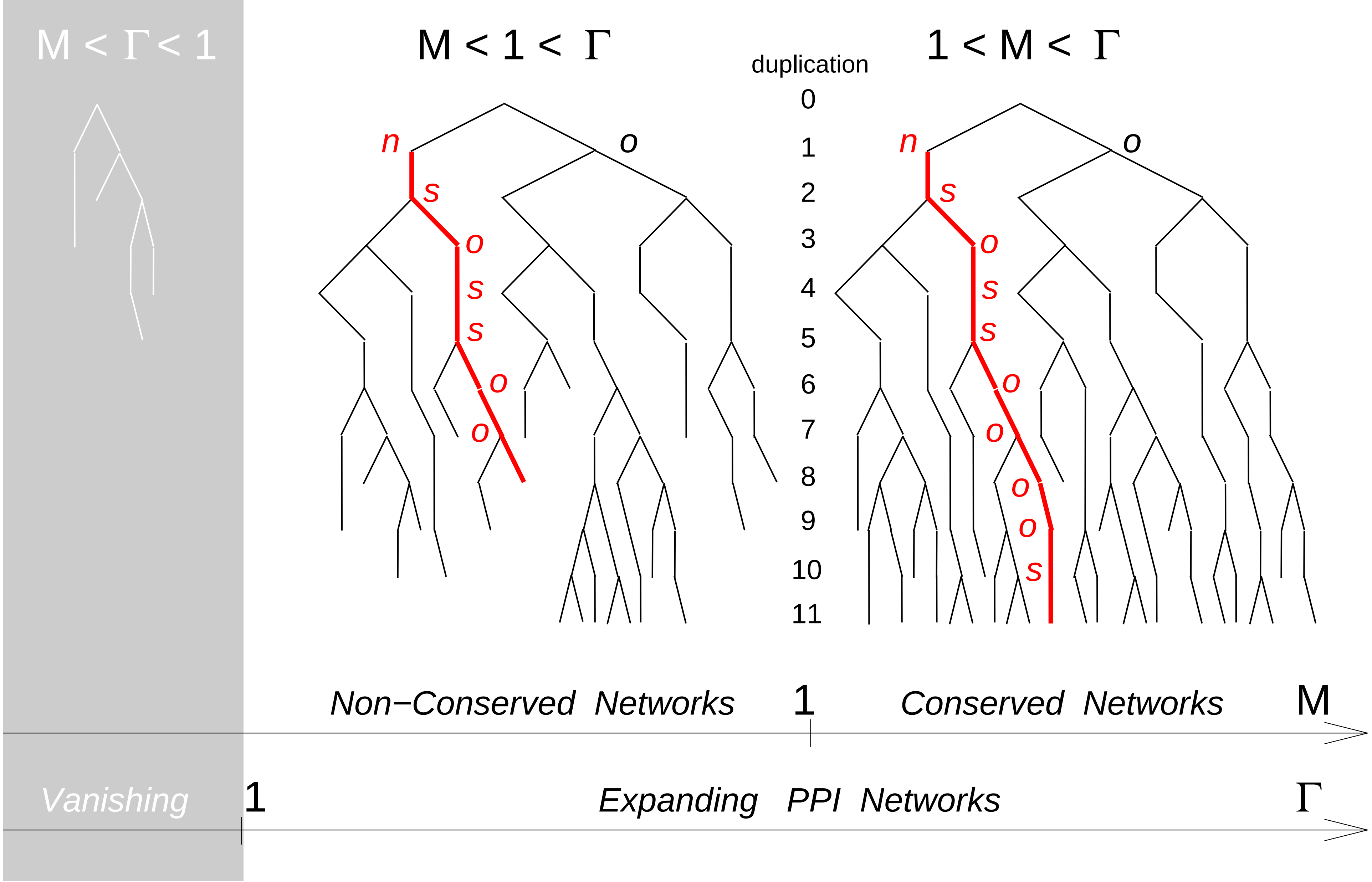
$$M' = \max_{i=s,o,n}(\Gamma_i) \quad \gtrsim M$$

3

$M' > 1$ Scale−free

$M' < 1$ Exponential

Network Conservation (M) and Expansion (Γ)



$M < \Gamma < 1$

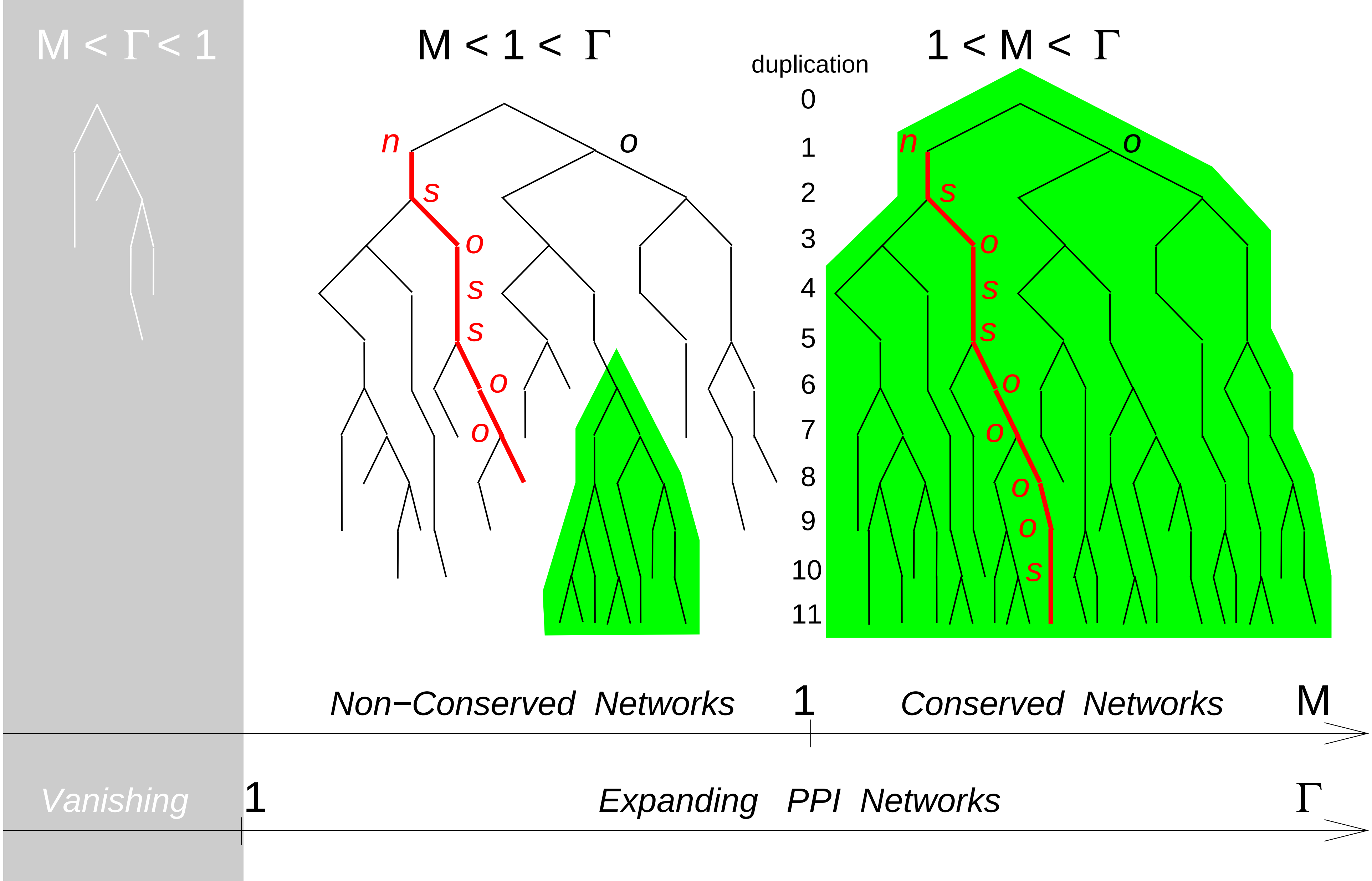
$M < 1 < \Gamma$

$1 < M < \Gamma$

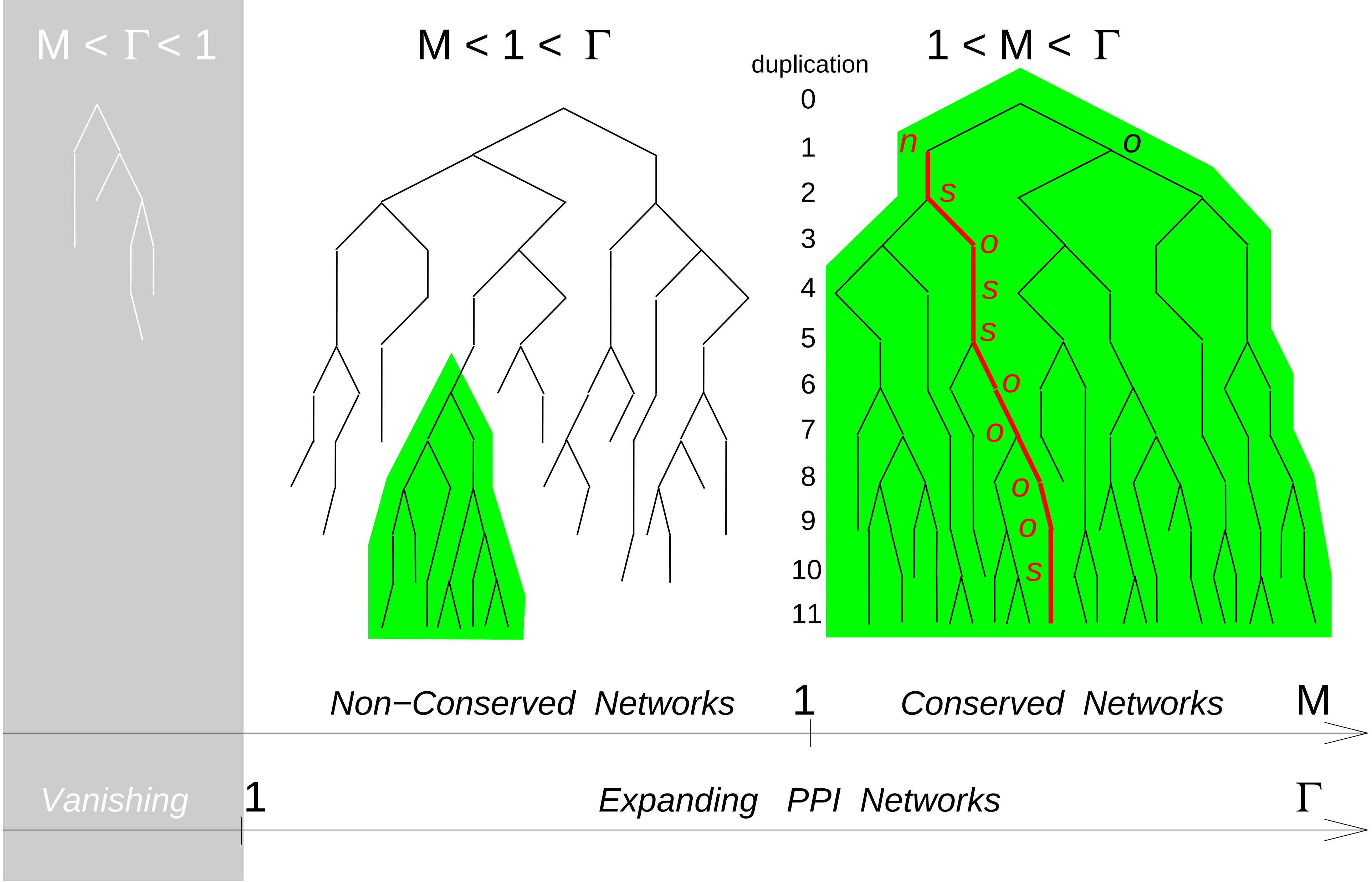
Non−Conserved Networks 1 *Conserved Networks* M

Vanishing 1 *Expanding PPI Networks* Γ

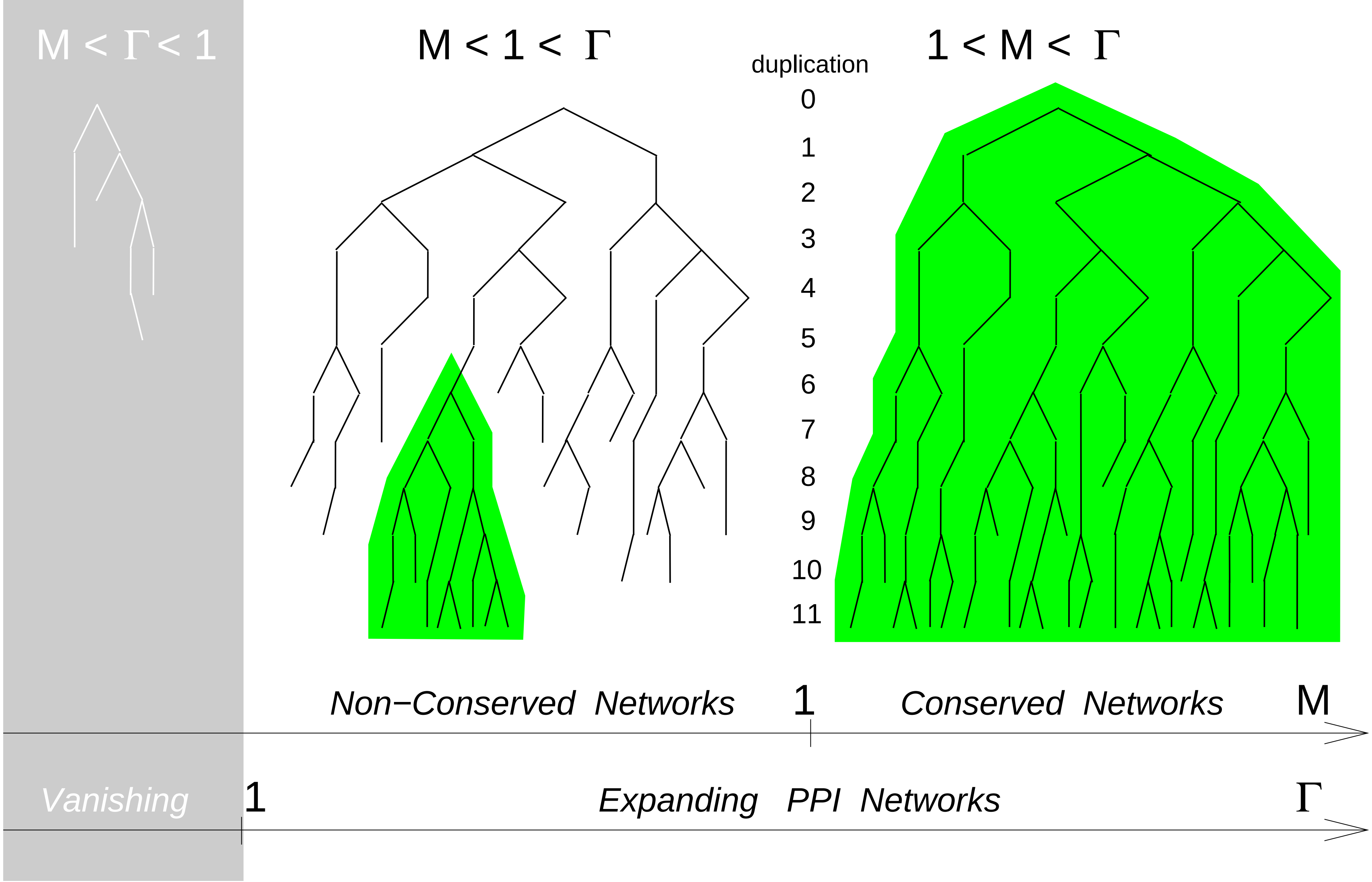
Network Conservation (M) and Expansion (Γ)



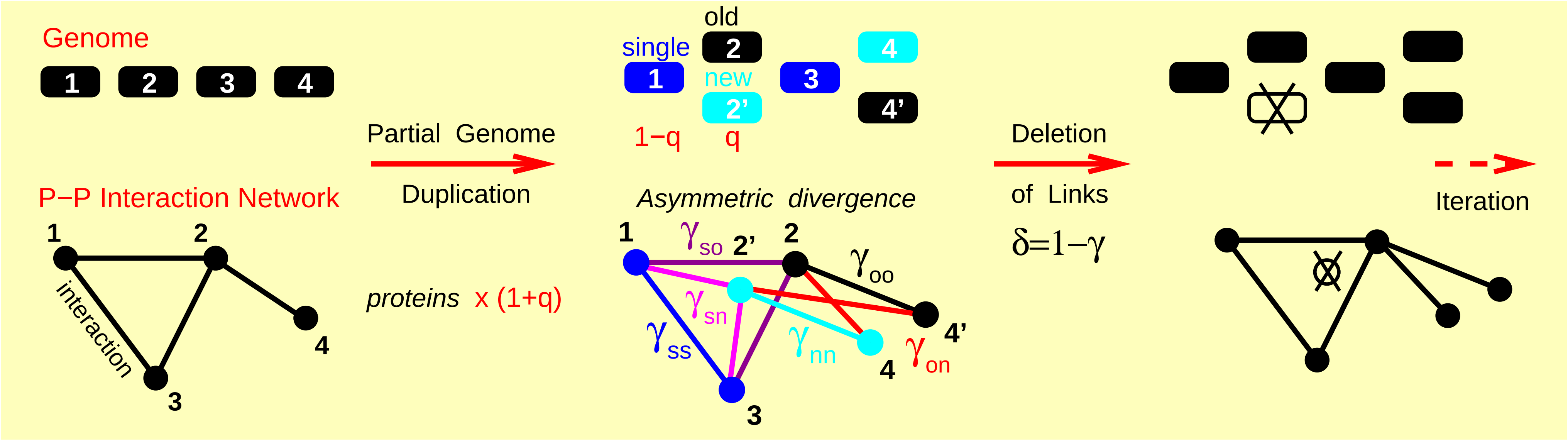
Network Conservation (M) and Expansion (Γ)



Network Conservation (M) and Expansion (Γ)



General Duplication−Divergence Model of PPI Network Evolution



Node Degree Growth Rate

$$\Gamma_i = (1-q)\gamma_{is} + q(\gamma_{io} + \gamma_{in})$$

Network Link Growth Rate

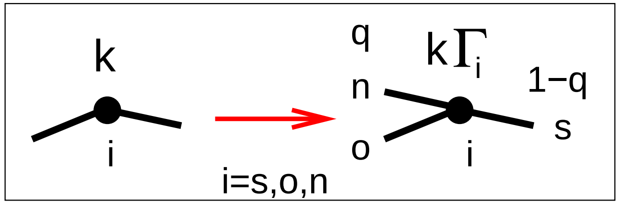
$$\Gamma = (1-q)\Gamma_s + q\,\Gamma_o + q\,\Gamma_n$$

Protein Conservation Index

$$M = (1-q)\Gamma_s + q\,\Gamma_o \quad < \quad \Gamma$$

Network Topology Index

$$M' = \max_{i=s,o,n}(\Gamma_i) \quad \gtrsim M$$



1

$\Gamma > 1$ Growing Network

$\Gamma < 1$ Vanishing Network

2

$M > 1$ Conserved Network

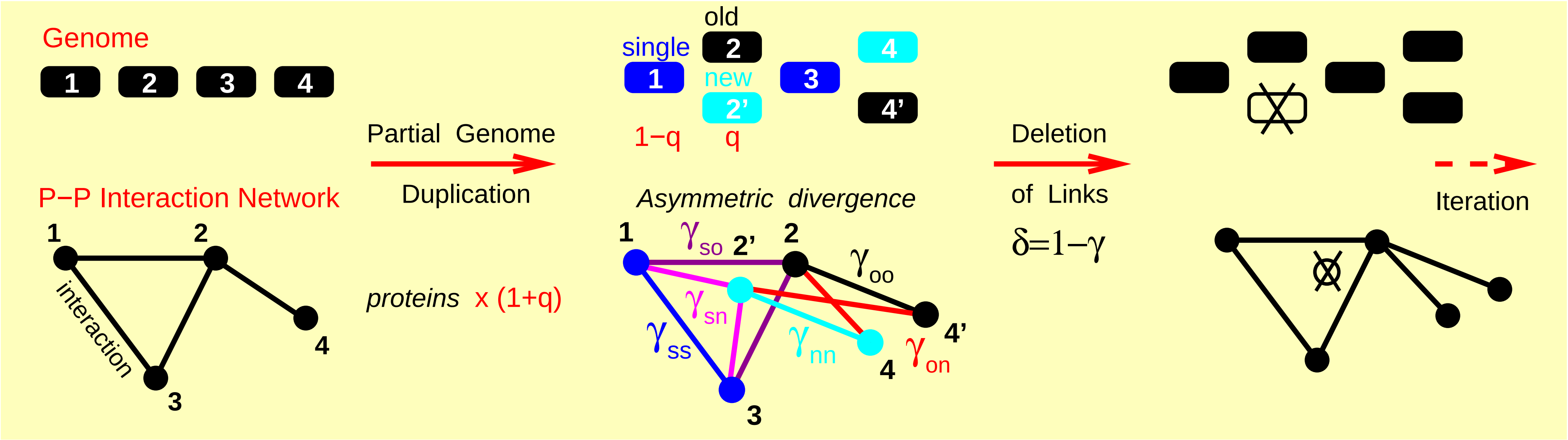
$M < 1$ Nonconserved Network

3

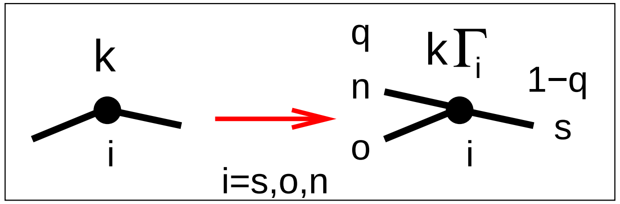
$M' > 1$ Scale−free

$M' < 1$ Exponential

General Duplication−Divergence Model of PPI Network Evolution

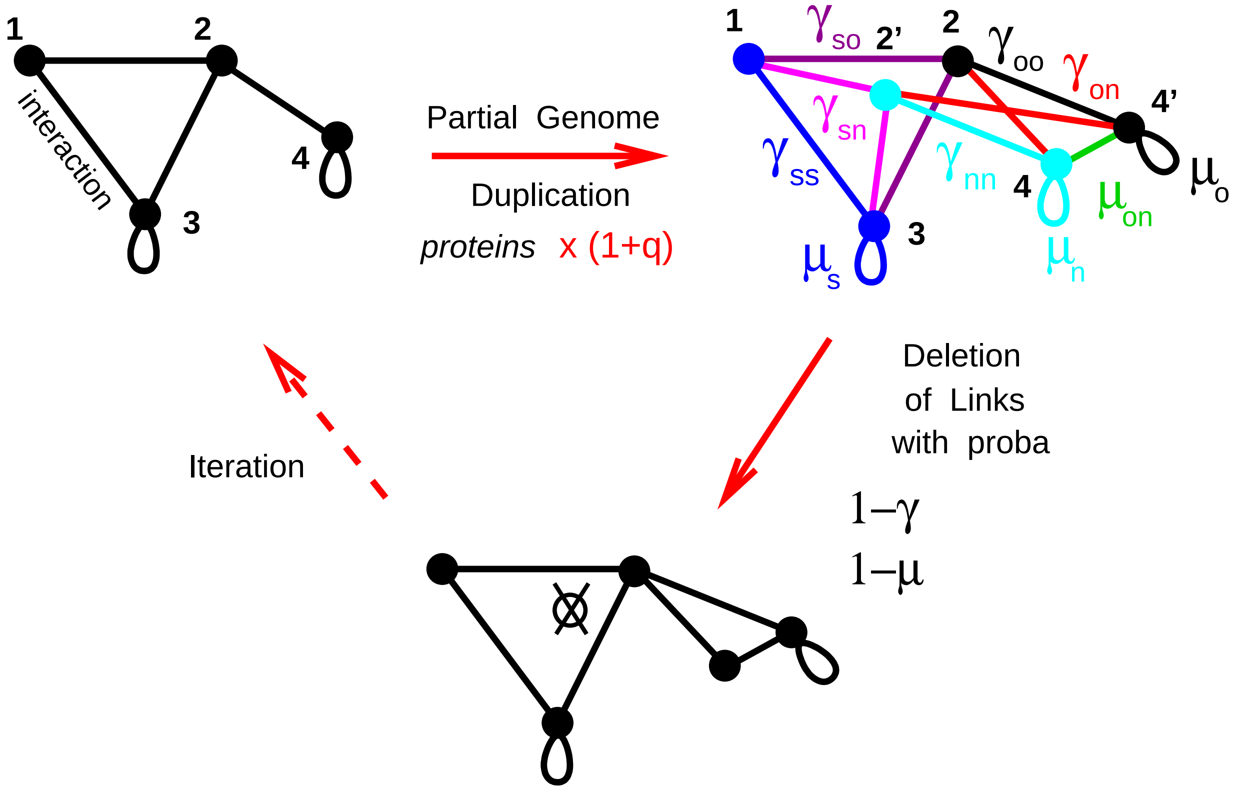


Node Degree Growth Rate	$\Gamma_i = (1-q)\gamma_{is} + q(\gamma_{io} + \gamma_{in})$	
Network Link Growth Rate	$\Gamma = (1-q)\Gamma_s + q\,\Gamma_o + q\,\Gamma_n$	1 $\Gamma > 1$ Growing Network $\Gamma < 1$ Vanishing Network
Protein Conservation Index	$M = (1-q)\Gamma_s + q\,\Gamma_o < \Gamma$	2 $M > 1$ Conserved Network $M < 1$ Nonconserved Network
Network Topology Index	$M' = \max_{i=s,o,n}(\Gamma_i) \gtrsim M$	3 $M' > 1$ Scale−free $M' < 1$ Exponential



→ Conserved Networks are also *necessary* Scale−free $(M' \gtrsim M > 1)$

Duplication−Divergence Models Including Self−links



Partial Genome

Duplication

proteins x (1+q)

Deletion
of Links
with proba

$1-\gamma$

$1-\mu$

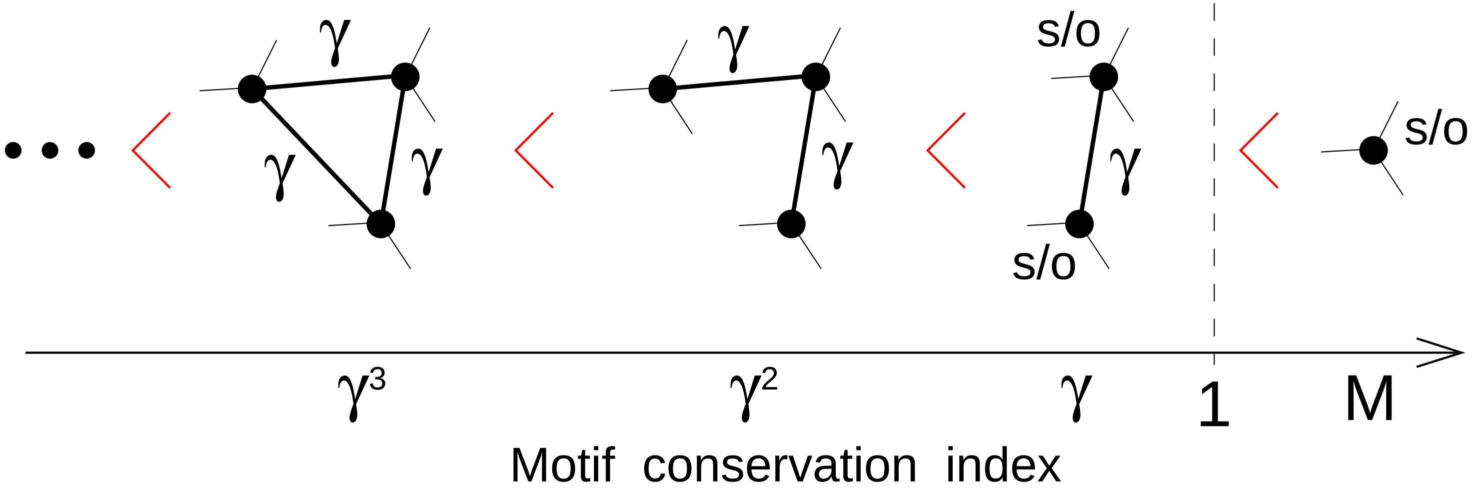
Iteration

Self−links leads to time−linear (not exponential) connectivity expansion

$k \longrightarrow k+1$ $\qquad$ $k \longrightarrow k\,\Gamma$

→ Self−links do not affect evolutionary regimes but are essential contributors to certain network motifs

Conservation of Network Motifs

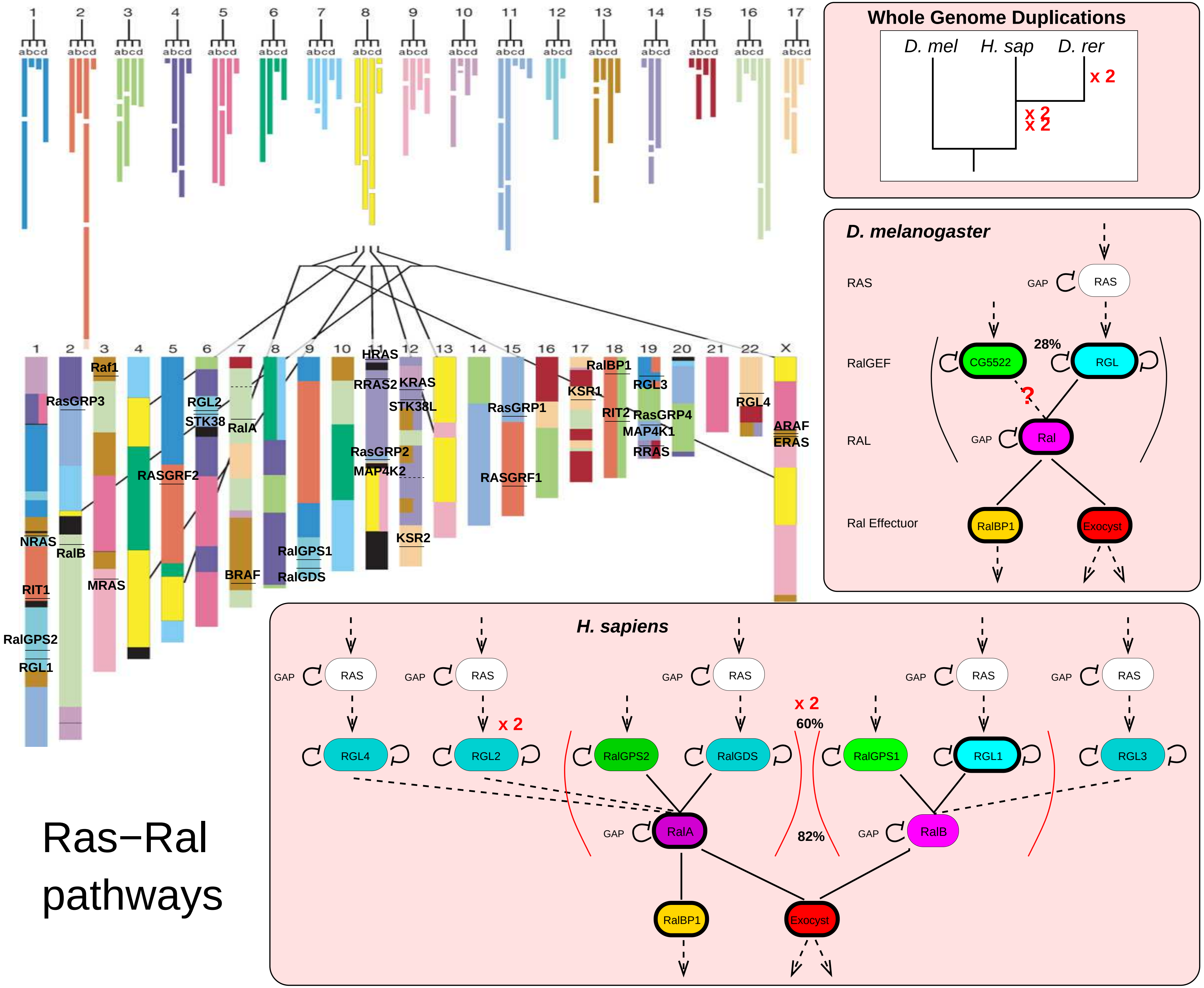


While Proteins are typically Conserved (if M>1)
Network Motifs are NOT Conserved under
Long−term Duplication−Divergence Evolution

Eventually: structural orthology $\not\longrightarrow$ functional orthology

This is in broad agreement with empirical observations !

Ras−Ral pathways



Generating Function Formalism

Kirill Evlampiev

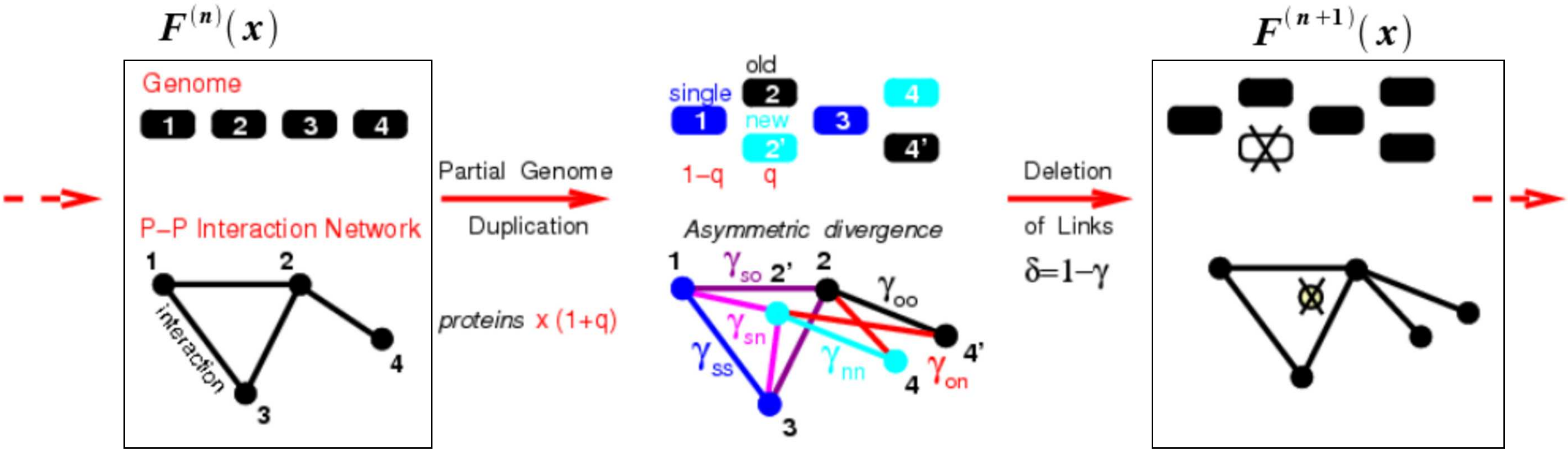
- At each step generating function of number of nodes

$$F^{(n)}(x)=\sum_{k\geq 0}\langle N_k^{(n)}\rangle x^k$$

- For two consecutive steps

$$\boxed{F^{(n+1)}(x)=(1-q)\,F^{(n)}(A_s(x))+qF^{(n)}(A_o(x))+qF^{(n)}(A_n(x)),}$$

$$A_i(x)=(1-q)(\delta_{is}+\gamma_{is}x)+q(\delta_{io}+\gamma_{io}x)(\delta_{in}+\gamma_{in}x),\ \delta_{ij}=1-\gamma_{ij},\ \ \Gamma_i=A'_i(1)$$



Generating Function Formalism

The real degree distribution and its generating function

$$p_k^{(n)} = \frac{\langle N_k^{(n)} \rangle}{\langle N^{(n)} \rangle}, k \geq 1 \qquad p^{(n)}(x) = \sum_{k \geq 1} p_k^{(n)} x^k - 1$$

Network's dynamics (get rid of k=0)

$$p^{(n+1)}(x) = \frac{(1-q)\, p^{(n)}(A_s(x)) + q\, p^{(n)}(A_o(x)) + q\, p^{(n)}(A_n(x))}{\Delta^{(n)}}$$

$$\Delta^{(n)} = \frac{\langle N^{(n+1)} \rangle}{\langle N^{(n)} \rangle} = -(1-q)\, p^{(n)}(A_s(0)) - q\, p^{(n)}(A_o(0)) - q\, p^{(n)}(A_n(0))$$

Stationary solution? $\lim_{n \to \infty} p^{(n)}(x) = p(x)$ → functional equation on *p(x)*

It can be analysed using asymptotic methods

$$p(x) = 1 - A_1(1-x) + A_2(1-x)^2 + \ldots + A_\alpha (1-x)^\alpha + o((1-x)^\alpha), \quad x \to 1$$

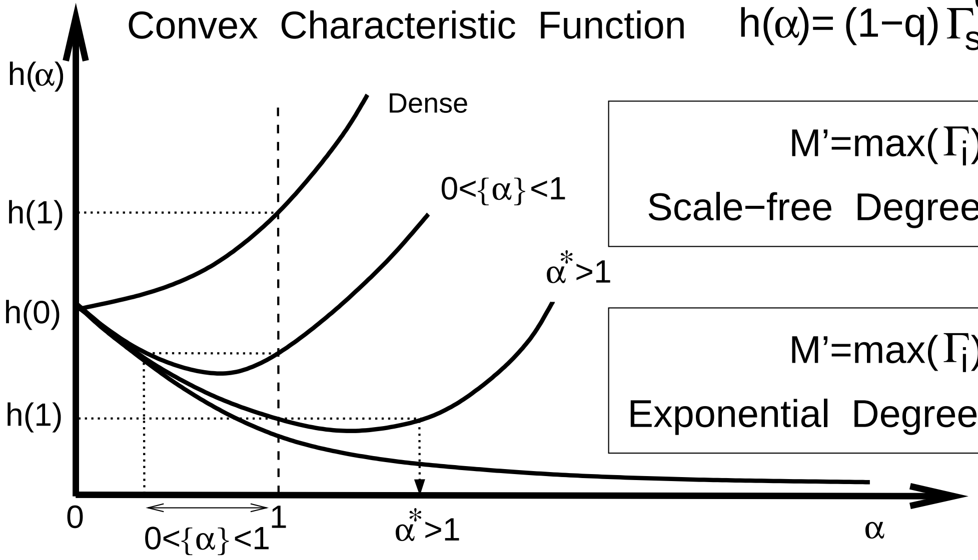
Asymptotic Topology of PPI Networks

Network Node Growth Rate

$$\langle N^{(n+1)}\rangle / \langle N^{(n)}\rangle = \Delta^{(n)} \longrightarrow \Delta = h(\alpha) = (1-q)\,\Gamma_s^{\alpha} + q\,\Gamma_o^{\alpha} + q\,\Gamma_n^{\alpha}$$

Asymptotic "Characteristic Equation"

Convex Characteristic Function $h(\alpha) = (1-q)\,\Gamma_s^{\alpha} + q\,\Gamma_o^{\alpha} + q\,\Gamma_n^{\alpha}$



$M' = \max(\Gamma_i) > 1$

Scale−free Degree Distribution

$p_k \sim 1/k^{\alpha+1}$

$M' = \max(\Gamma_i) < 1$

Exponential Degree Distribution

$p_k \sim \exp(-\mu k)$

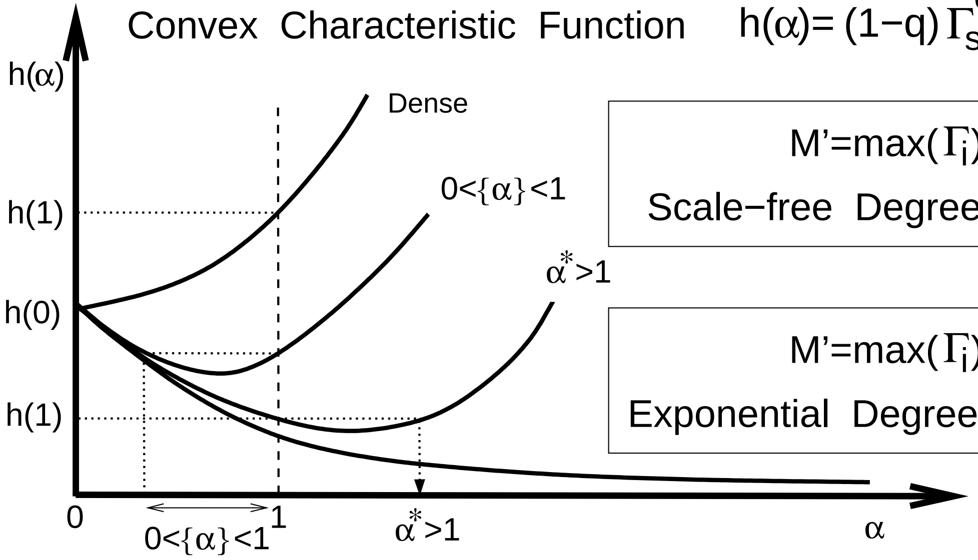
Asymptotic Topology of PPI Networks

Network Node Growth Rate

$$<N^{(n+1)}>/<N^{(n)}> = \Delta^{(n)} \longrightarrow \Delta = h(\alpha) = (1-q)\,\Gamma_s^{\alpha} + q\,\Gamma_o^{\alpha} + q\,\Gamma_n^{\alpha}$$

Asymptotic "Characteristic Equation"

Convex Characteristic Function $h(\alpha) = (1-q)\,\Gamma_s^{\alpha} + q\,\Gamma_o^{\alpha} + q\,\Gamma_n^{\alpha}$



$M' = \max(\Gamma_i) > 1$

Scale−free Degree Distribution

$p_k \sim 1/k^{\alpha+1}$

$M' = \max(\Gamma_i) < 1$

Exponential Degree Distribution

$p_k \sim \exp(-\mu k)$

⟶ Asymptotic Network Topology $M' = \max_{i=s,o,n}(\Gamma_i)$

$M' > 1$ Scale−free

$M' < 1$ Exponential

GDD Model + Fluctuations

$$M' = \prod_{n=1}^{R} \max(\Gamma_i^{(n)}) \geqslant \prod_{n=1}^{R} \left[(1-q^{(n)})\,\Gamma_s^{(n)} + q^{(n)}\,\Gamma_o^{(n)}\right] = M$$

Conserved Networks are also *necessary* Scale−free $(M' \geqslant M > 1)$

Statistical Properties of the Model

- Network evolution is described in terms of ensemble averages <Q>
- This description is meaningfull since fluctuations are not large

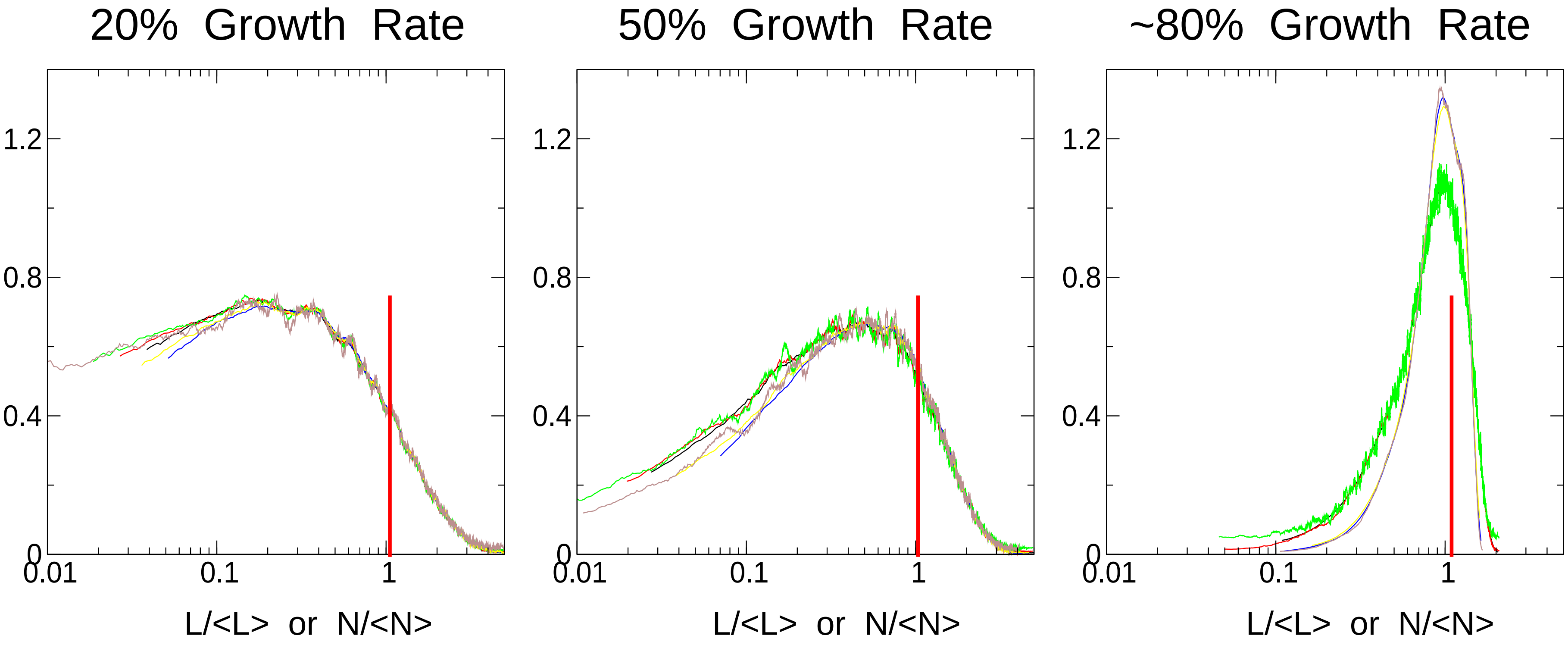
distribution function of ***L***

$$\boxed{\boldsymbol{P}^{(n)}(L) \sim \frac{1}{\langle L^{(n)} \rangle} \boldsymbol{F}\left[\frac{L}{\langle L^{(n)} \rangle}\right]}$$

the pth moment is proportional to the pth power of the first one!

- Due to $N_k \leq N \leq 2L$ in linear regimes $\langle N \rangle \sim \langle L \rangle$ Exp & Scale Free this holds also for distributions of N and N_k
- For nonlinear regime this probably holds too (numerical simulations)

Distribution of Network Sizes


20% Growth Rate

50% Growth Rate

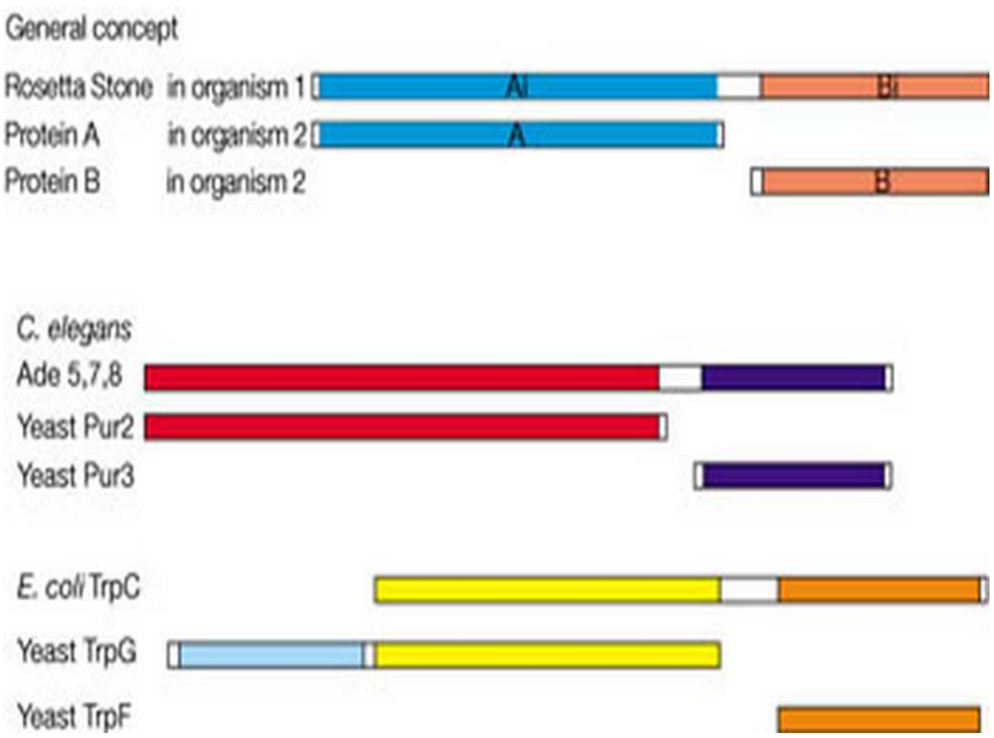
~80% Growth Rate

Linear regime N ~ L

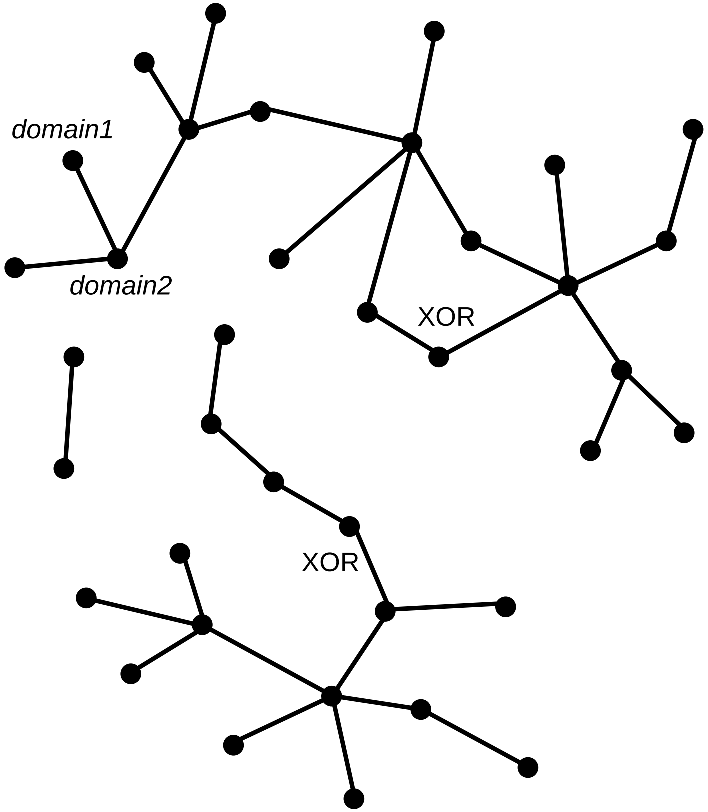
NL regime N < L

Redefining PPI Networks as Domain Interaction Networks

Genome and PPI Network Evolution under Protein Domain Shuffling



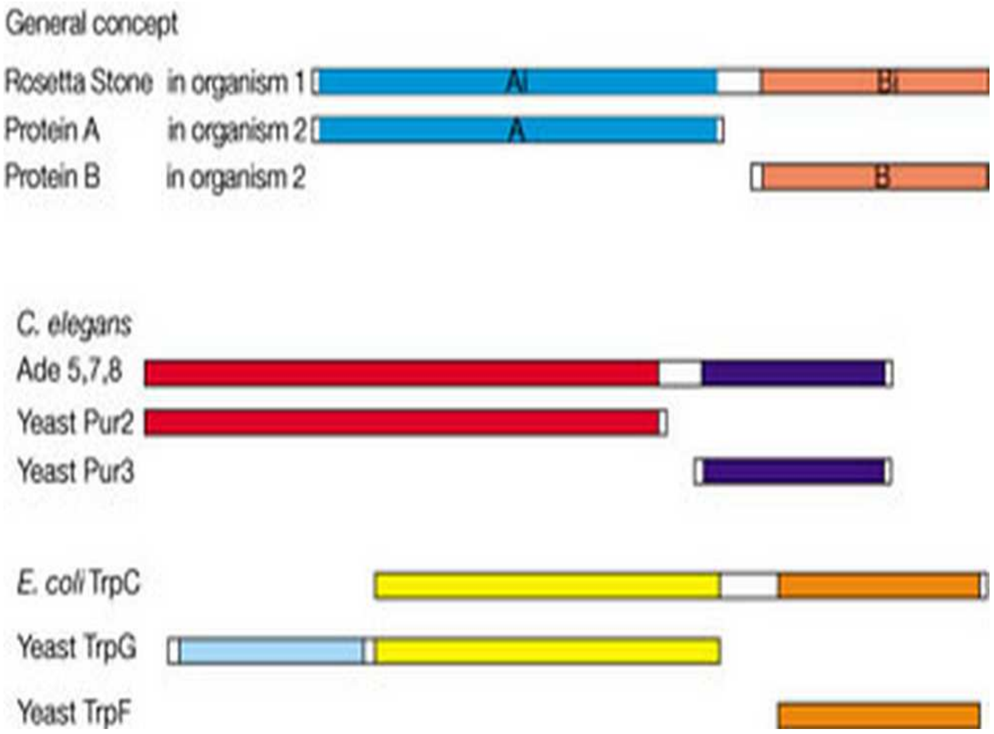
Domain−Domain Interaction Network



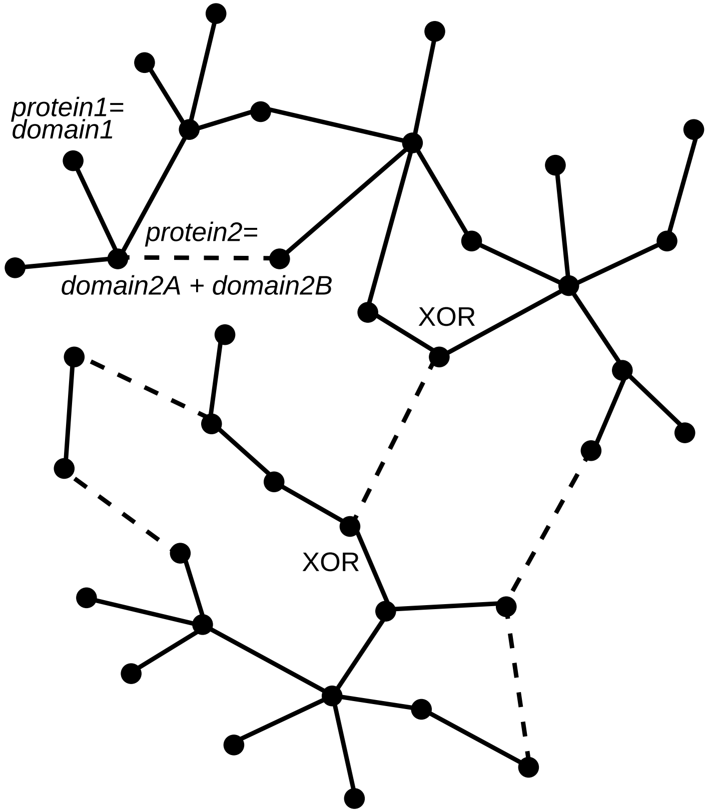
— Direct interaction bwn protein domains

Redefining PPI Networks as Domain Interaction Networks

Genome and PPI Network Evolution under Protein Domain Shuffling



Domain−Domain Interaction Network

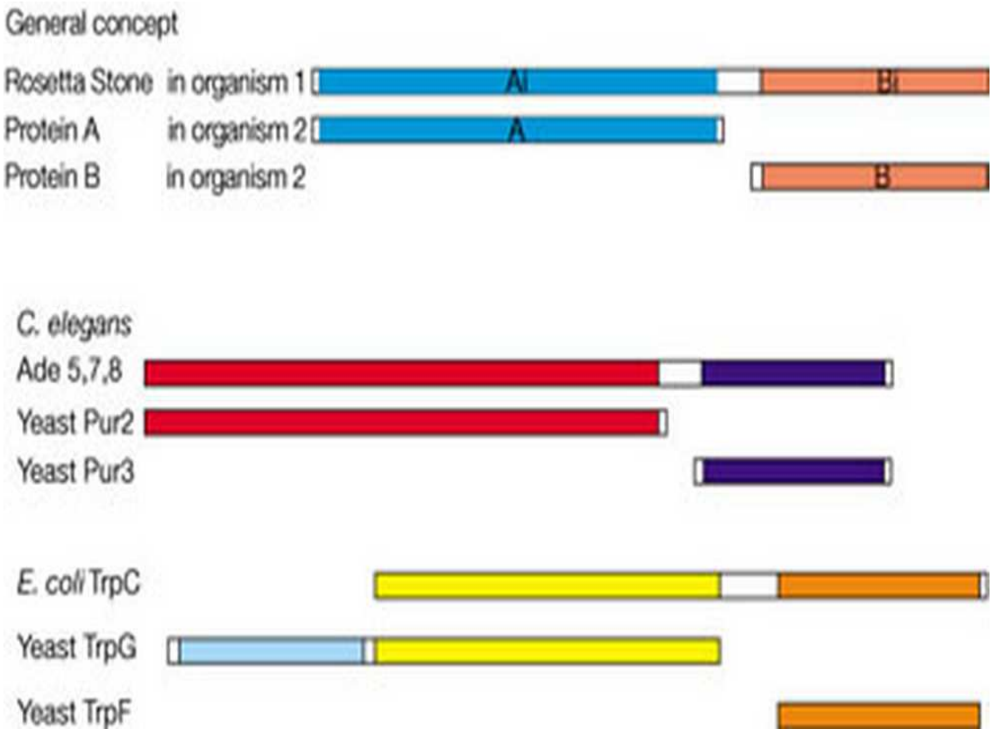


—— Direct interaction bwn protein domains

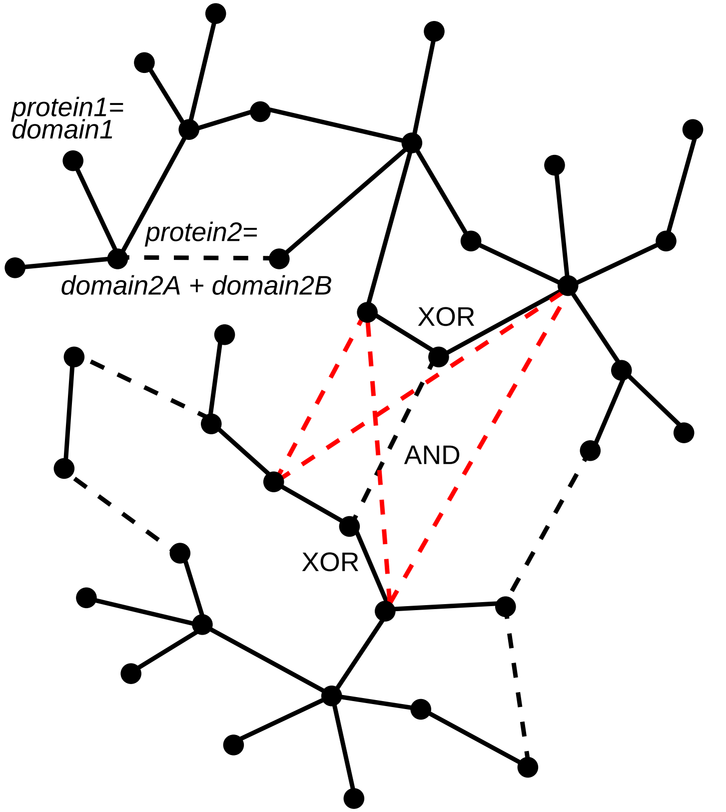
- - - Multidomain proteins (random shuffling)

Redefining PPI Networks as Domain Interaction Networks

Genome and PPI Network Evolution under Protein Domain Shuffling



Domain−Domain Interaction Network



—— Direct interaction bwn protein domains

- - - Multidomain proteins (random shuffling)

- - - Indirect interaction within complexes

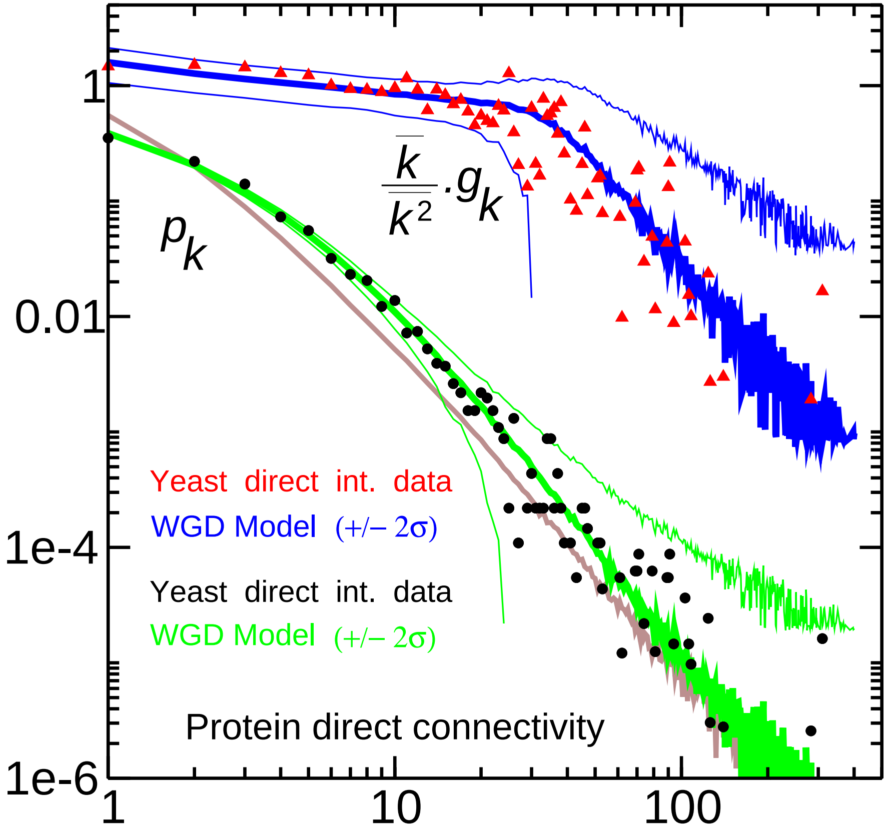
→ Adding some "combinatorial logic" through multidomain proteins

Comparison with Yeast Direct Interaction Data

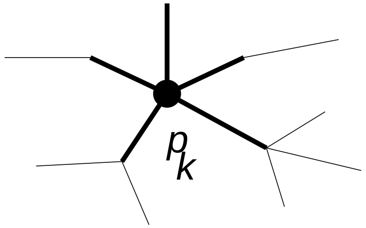
A two−parameter Whole Genome Duplication Model
with Random Protein Domain Shuffling γ_{on} λ

$\gamma_{on}=0.1$ $(\gamma_{oo}=1$ $\gamma_{nn}=0)$
$\lambda=1.5$ prot. bind. domains per protein

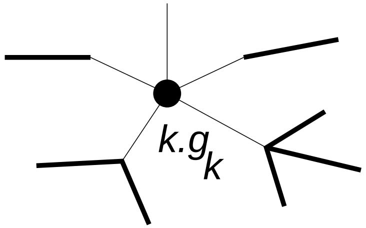
Topology of direct interaction network



Node Degree Distribution



Average Connectivity of Neighbours

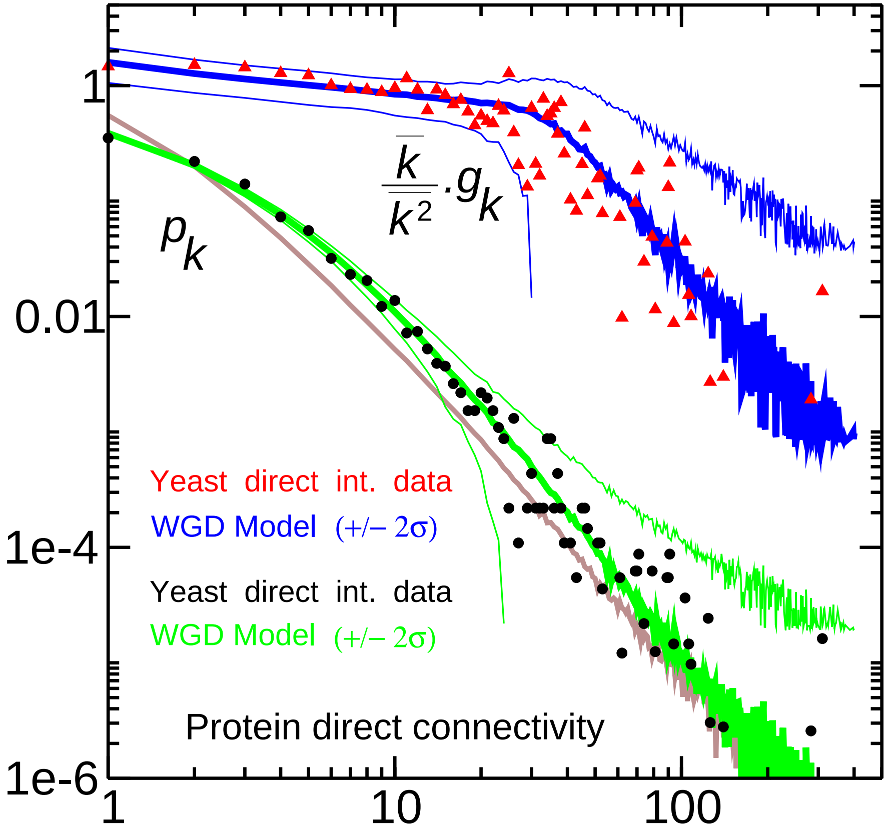


Comparison with Yeast Direct Interaction Data

A two−parameter Whole Genome Duplication Model
with Random Protein Domain Shuffling γ_{on} λ

γ_{on} =0.1 (γ_{oo}=1 γ_{nn}=0)
λ=1.5 prot. bind. domains per protein

Topology of direct interaction network



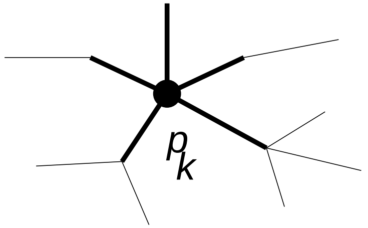
PPI Network growth rate

$$\Gamma = \Gamma_o + \Gamma_n = 1 + 2\,\gamma_{on} = 20\%$$

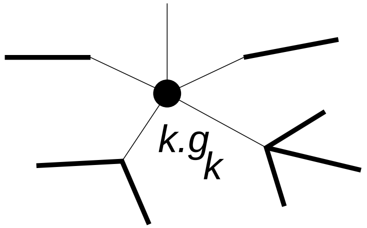
Seasquirt−Human (2WGD): 25%

Seasquirt−Fugu (3WGD): 11%

Node Degree
Distribution



Average Connectivity
of Neighbours

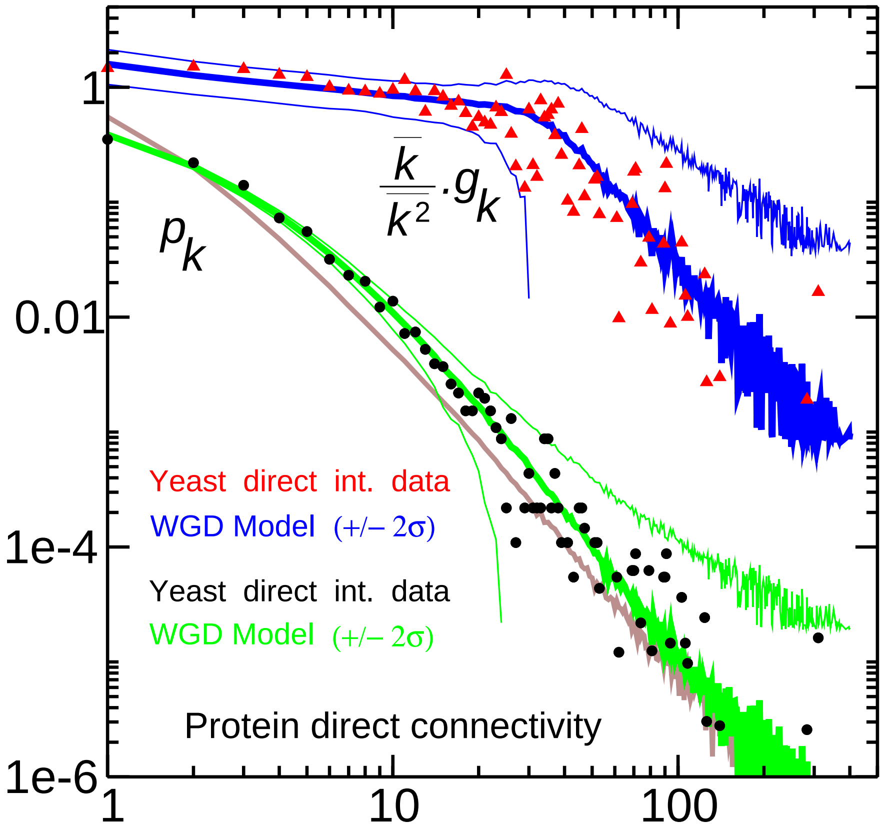


Comparison with Yeast Direct + Indirect Interaction Data

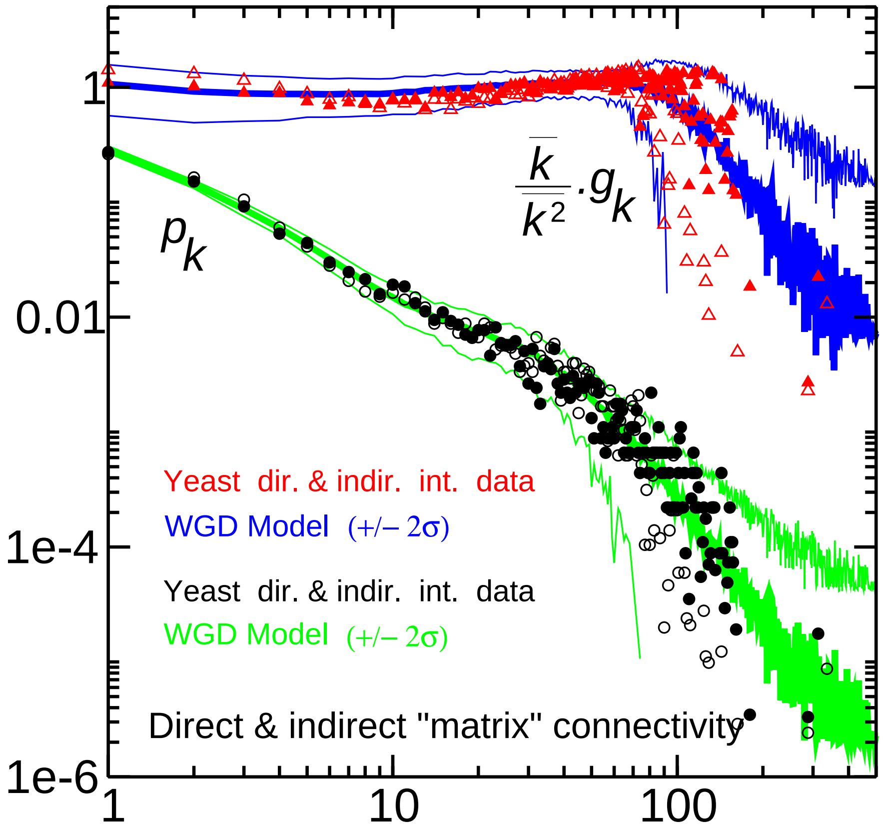
A two−parameter Whole Genome Duplication Model
with Random Protein Domain Shuffling γ_{on} λ

γ_{on} =0.1 (γ_{oo}=1 γ_{nn}=0)
λ=1.5 prot. bind. domains per protein

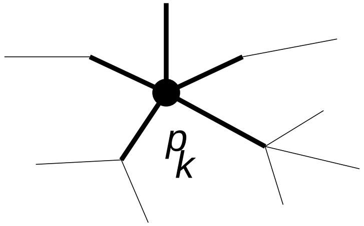
Topology of direct interaction network



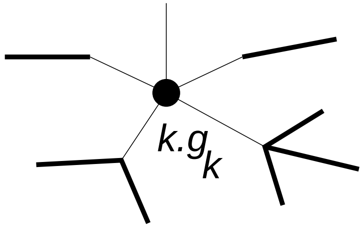
Topology of direct & indirect interaction network



Node Degree
Distribution



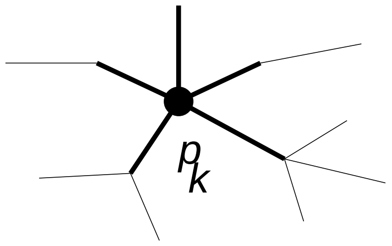
Average Connectivity
of Neighbours

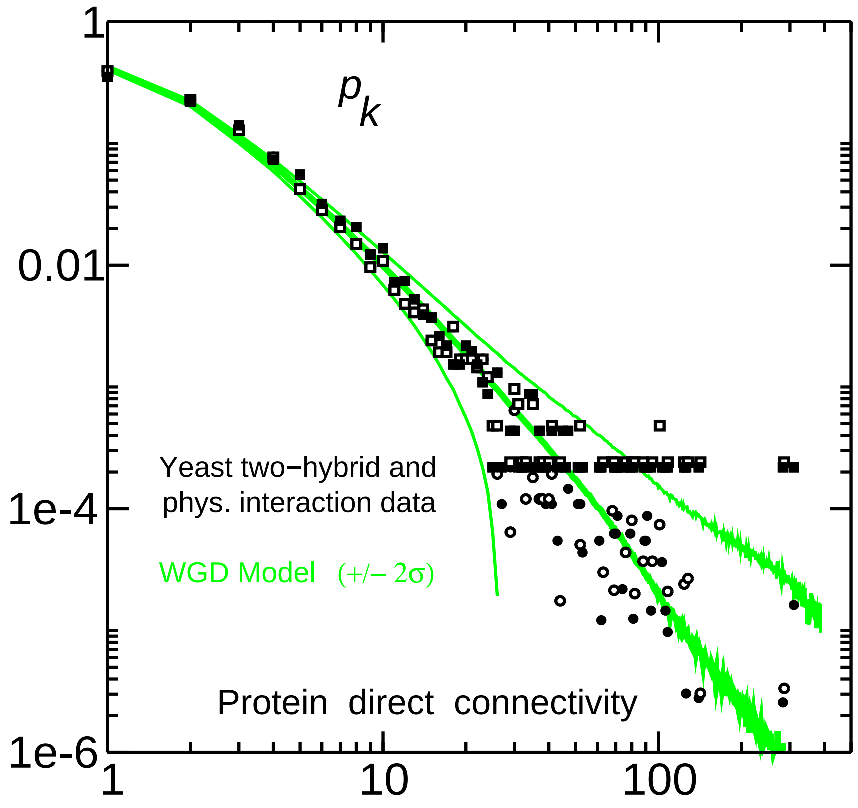


Comparison with Yeast Direct Interaction Data

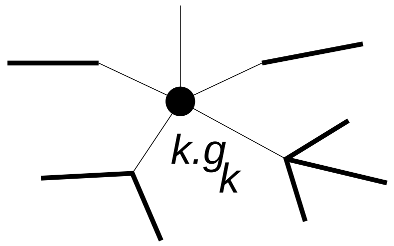
Fit of Yeast Data with a one−parameter WGD Model γ_{on}=0.26 (γ_{oo}=1 γ_{nn}=0)

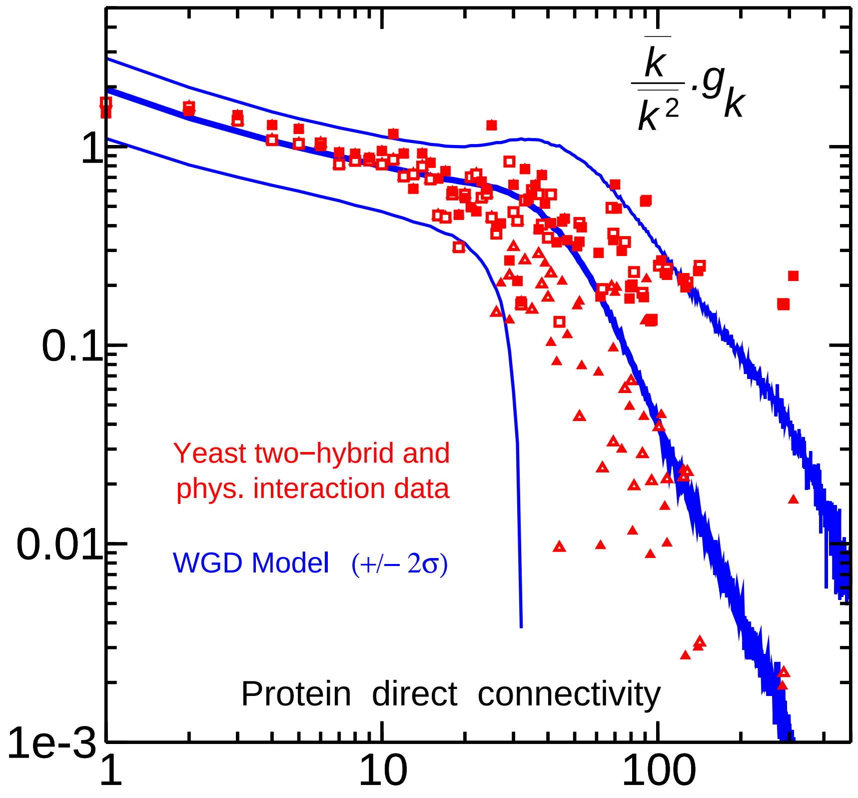
Node Degree Distribution





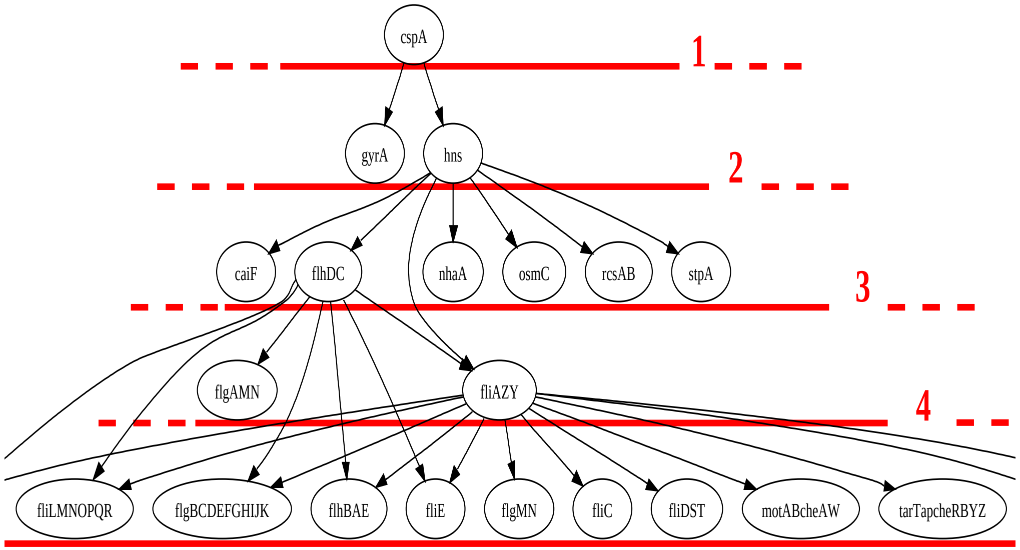
Average Connectivity of Neighbours



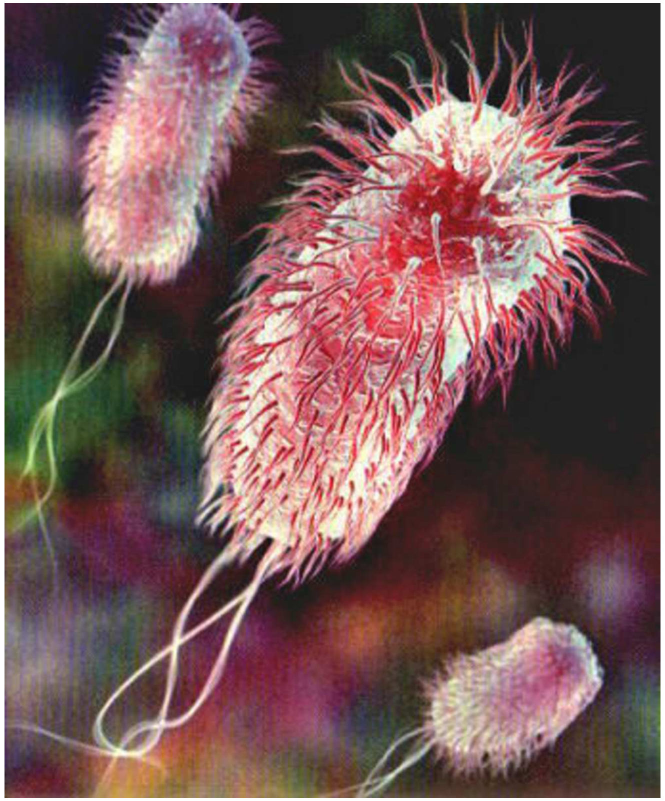


Evolution and Topology of Transcription Networks

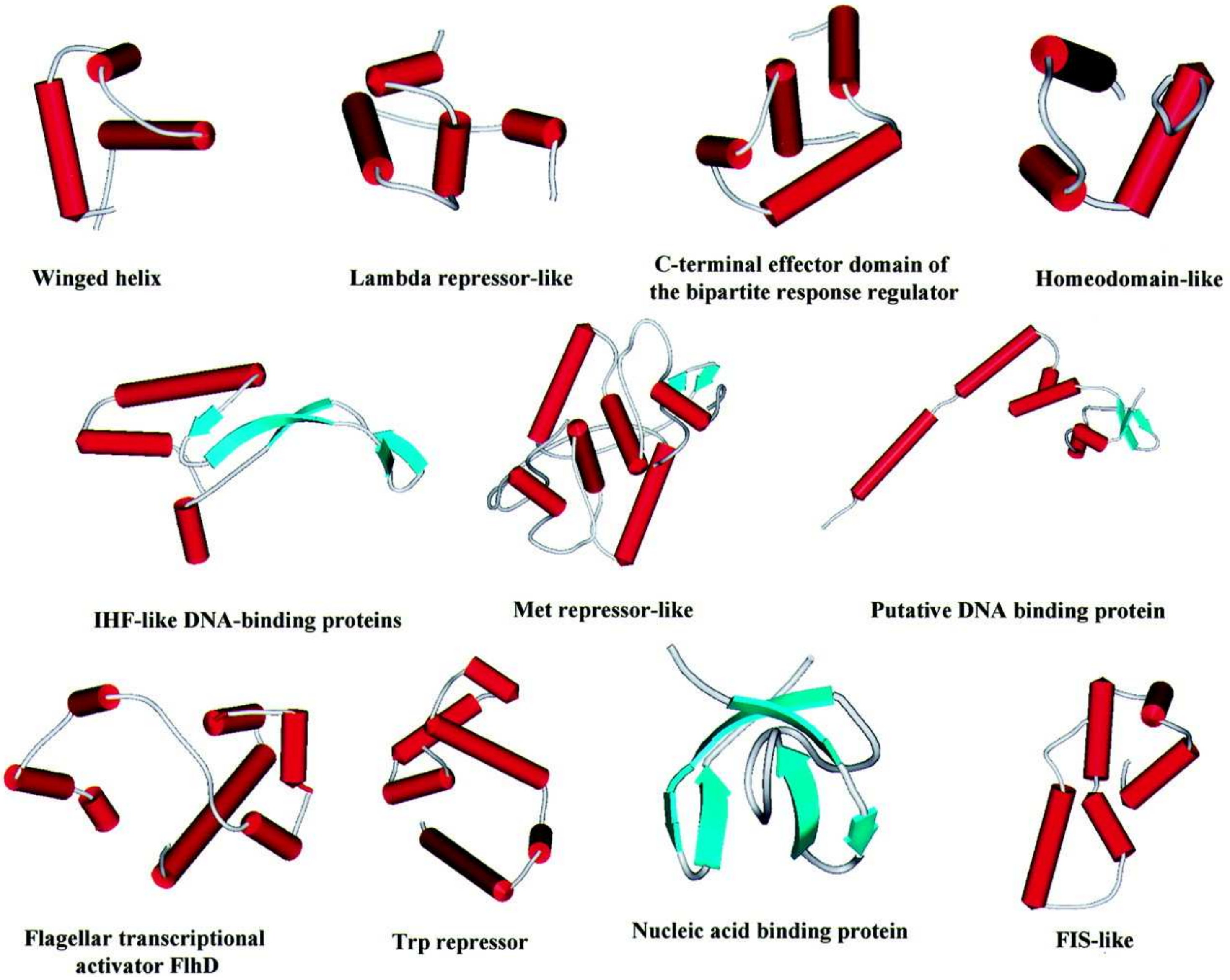
E.coli shallow Transcription Network



No large scale feedbacks
but 60% autoregulators (AR)

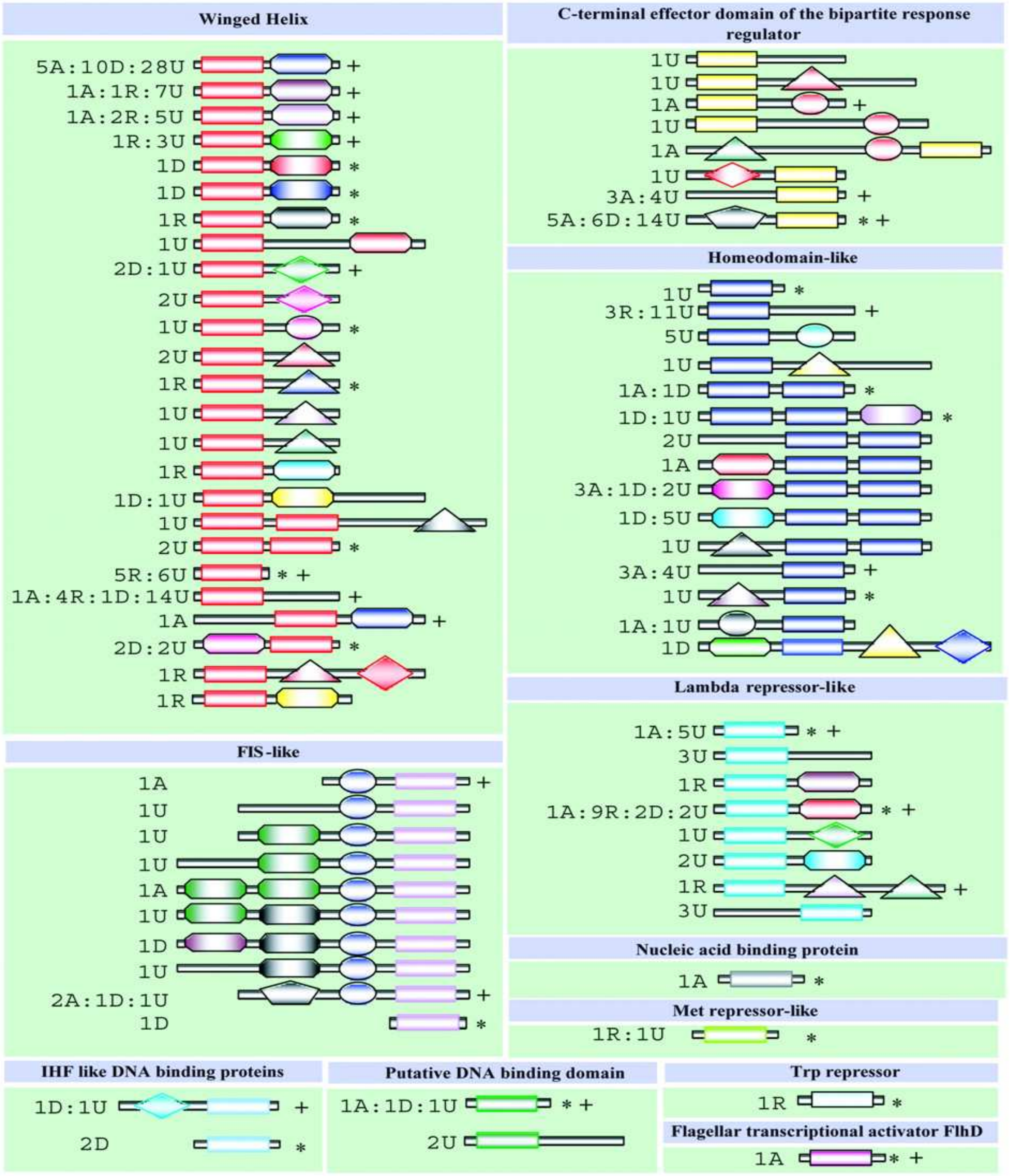


Transcription Factor Families



Babu et al., NAR 2006

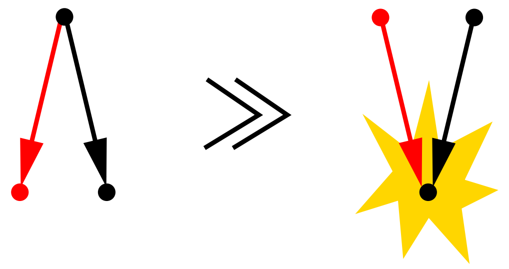
TF Family Expansion by Duplication



Babu et al., NAR 2006

Hierarchy and Feedback in the Evolution of Bacterial Transcription Networks

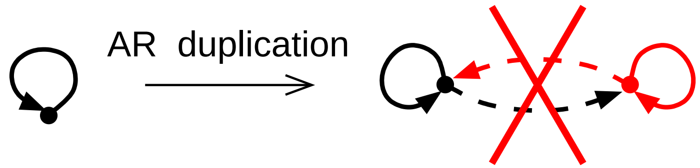
Duplication / Divergence Asymmetry



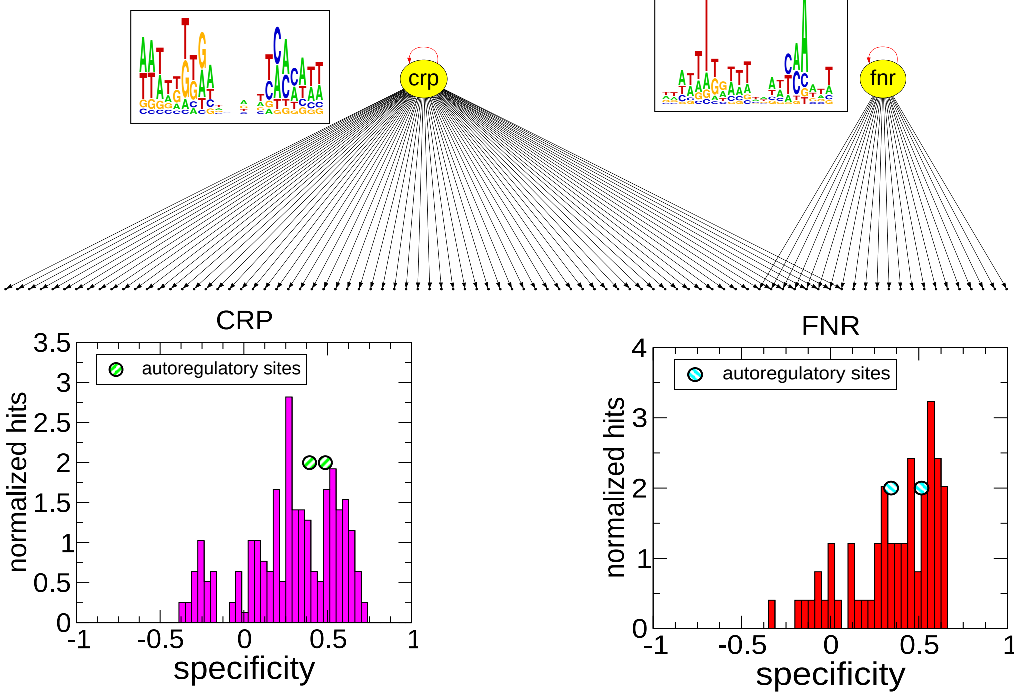
Regulatory conflict !

Evolutionary decoupling of duplicated ARs?

elimination of crosstalks

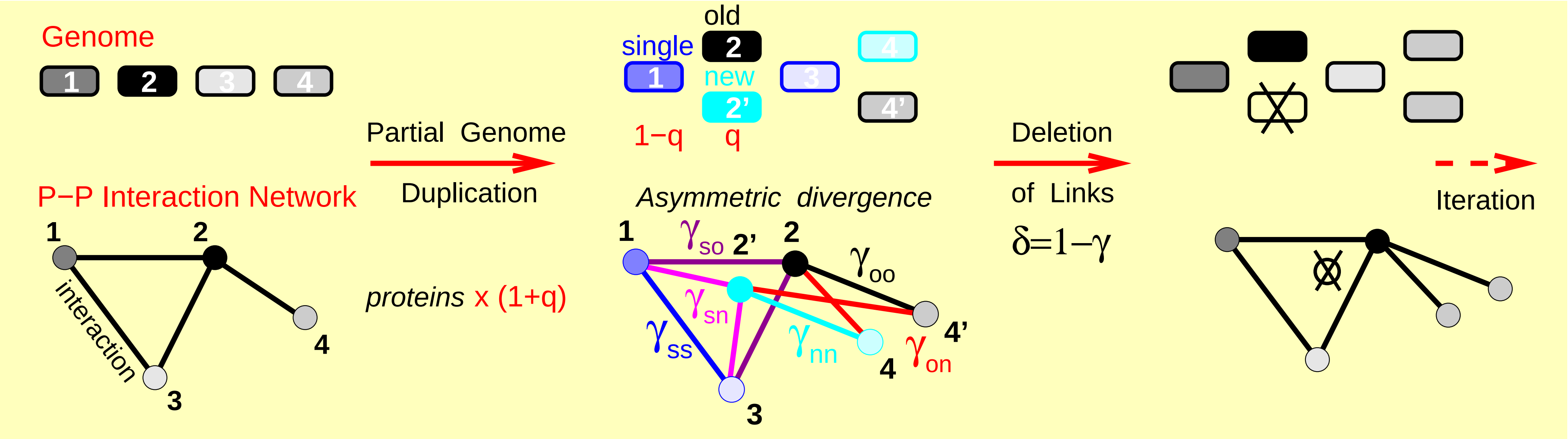


crp / fnr example



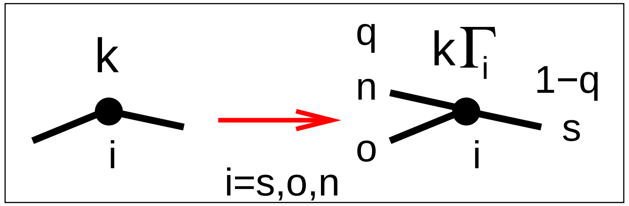
Cosentino−Lagomarsino, Jona, Bassetti & Isambert,
PNAS, 104, 5516 (2007) arxiv.org/abs/q.bio/0701035

Inhomogeneous Models of Duplication−Divergence Evolution



While the general "multicolor" inhomogeneous model cannot be solved exactly, one can use a posteriori choices of evolutionary parameters' values

$$q^{(n)} \quad \text{and} \quad \Gamma_i^{(n)} = (1-q)\gamma_{is}^{(n)} + q(\gamma_{io}^{(n)} + \gamma_{in}^{(n)})$$



to mimick positive / negative selection of combinatorial features of networks (approximation)

→ This approach can model for instance the adaptative expansion of gene families

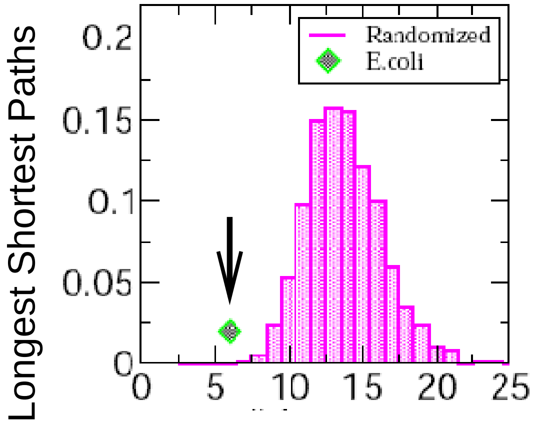
→ The approach becomes exact for two−color bipartite networks or oriented networks

Bacteria versus Eukaryote Transcriptional Hierarchy

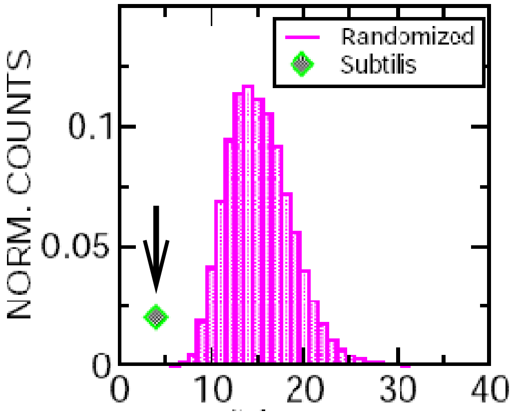
S. cerevisiae has long transcriptional cascades

while bacteria have short ones

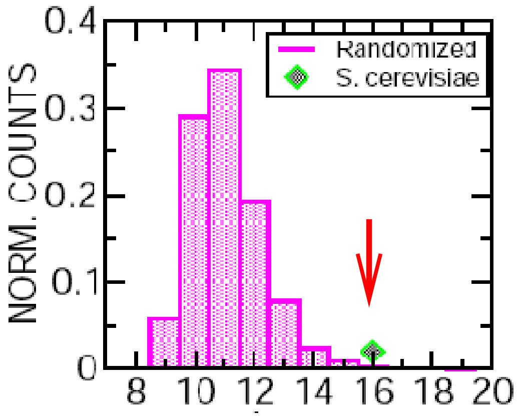
Escherichia coli

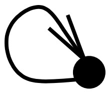


Bacillus subtilis

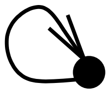


S. cerevisiae





> 50% of TF

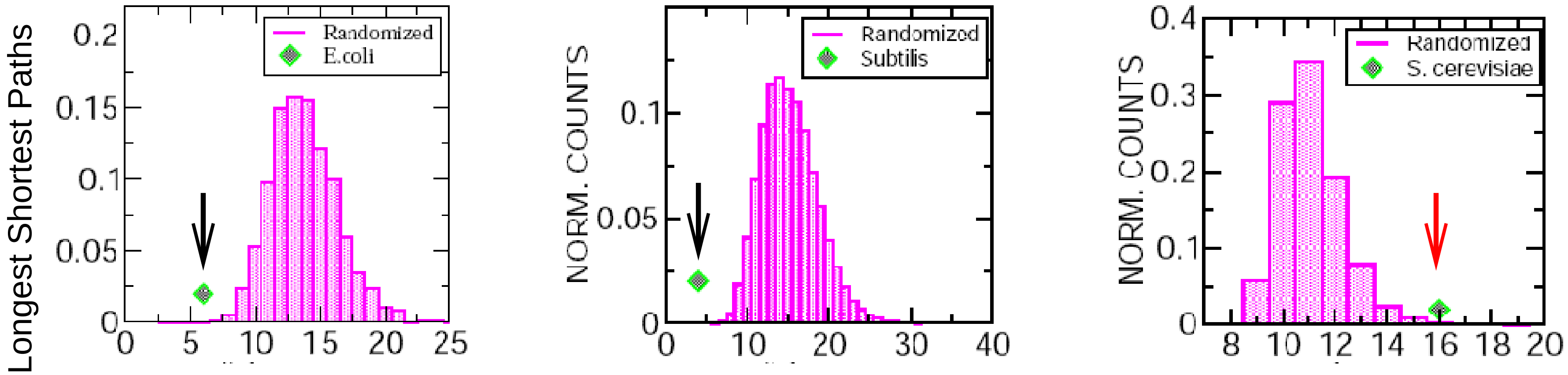


~ 10% of TF

Bacteria versus Eukaryote Transcriptional Hierarchy

S. cerevisiae has long transcriptional cascades

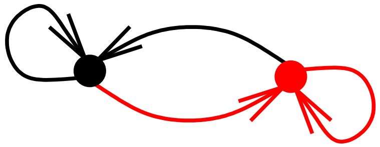
while bacteria have short ones

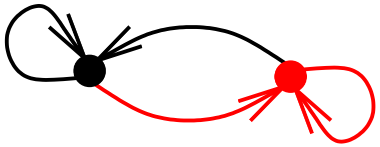


Escherichia coli

Bacillus subtilis

S. cerevisiae

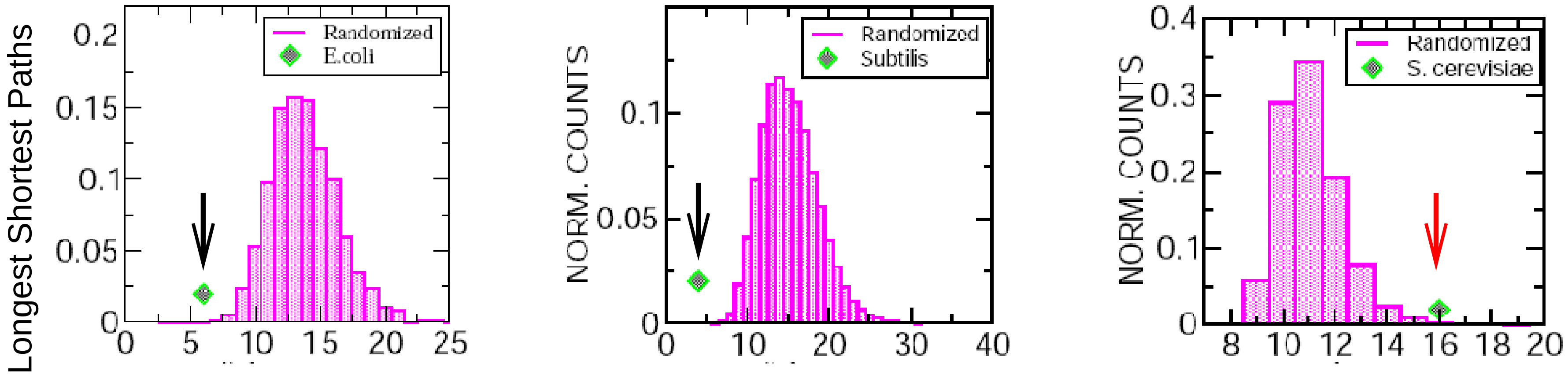




Bacteria versus Eukaryote Transcriptional Hierarchy

S. cerevisiae has long transcriptional cascades

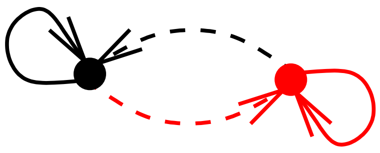
while bacteria have short ones

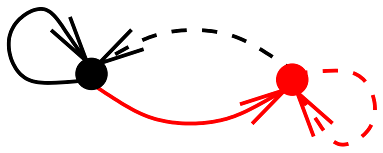


Escherichia coli

Bacillus subtilis

S. cerevisiae



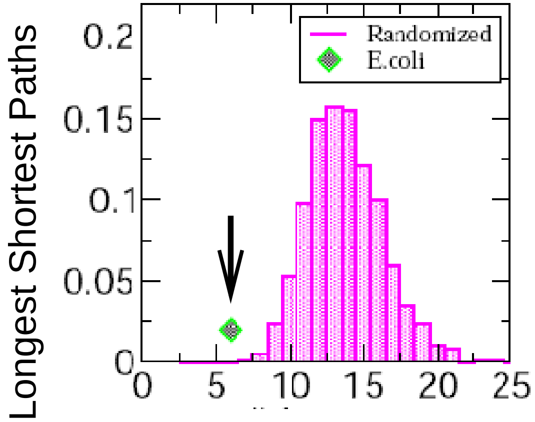


Bacteria versus Eukaryote Transcriptional Hierarchy

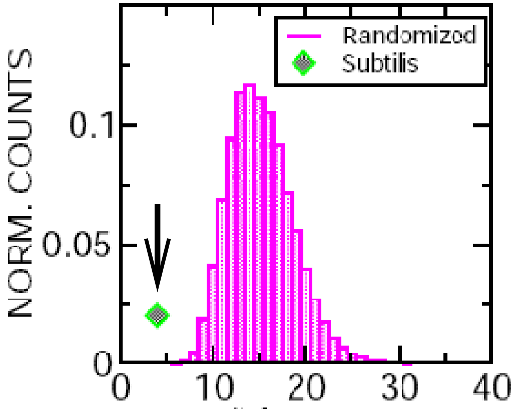
S. cerevisiae has long transcriptional cascades

while bacteria have short ones

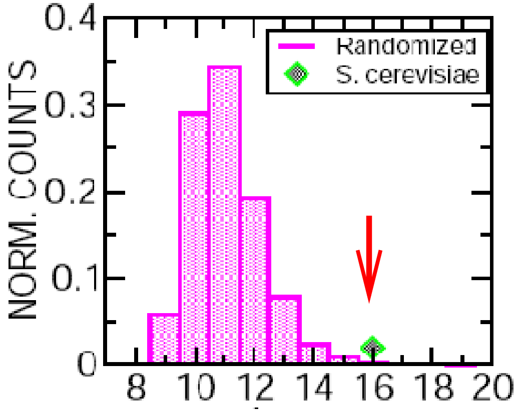
Escherichia coli



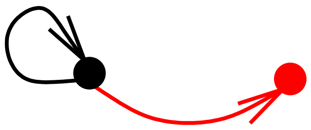
Bacillus subtilis



S. cerevisiae



Adaptative Selection ?



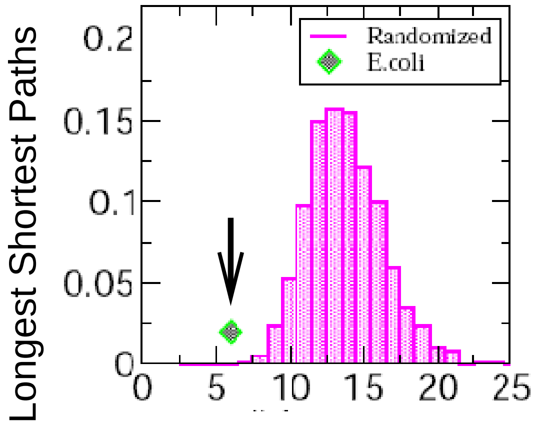
Random drift ?

Bacteria versus Eukaryote Transcriptional Hierarchy

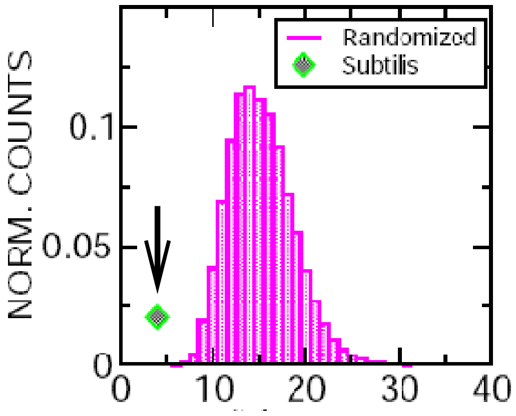
S. cerevisiae has long transcriptional cascades

while bacteria have short ones

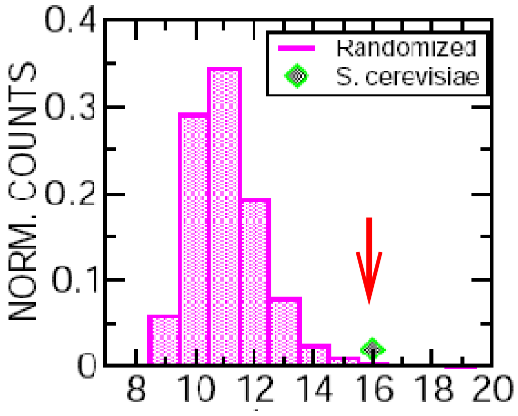
Escherichia coli

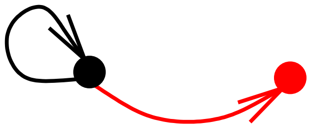


Bacillus subtilis



S. cerevisiae





Population Dynamics

Adaptative Selection ?

N.s > 1

N population size

Random drift ?

N.s << 1

OR non−adaptative constraints on bacterial genomes ?? Isambert & Stein, Biol Direct (2009)

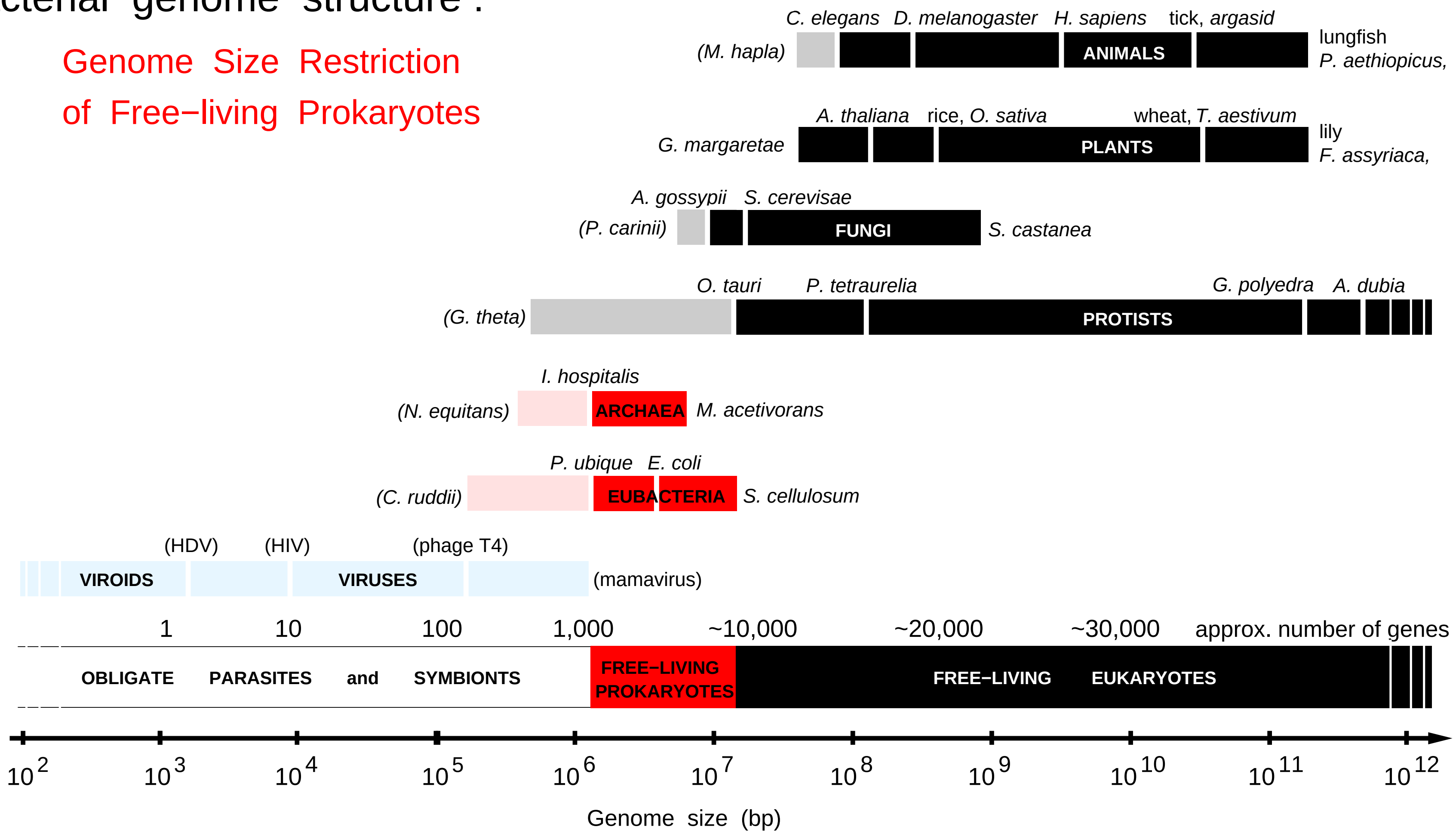
Alternatively... Specific Evolutionary Constraints of Bacteria?

- Transcription / translation coupling versus mRNA export out of the nucleus

Alternatively... Specific Evolutionary Constraints of Bacteria?

- Transcription / translation coupling versus mRNA export out of the nucleus
- Bacterial genome structure :

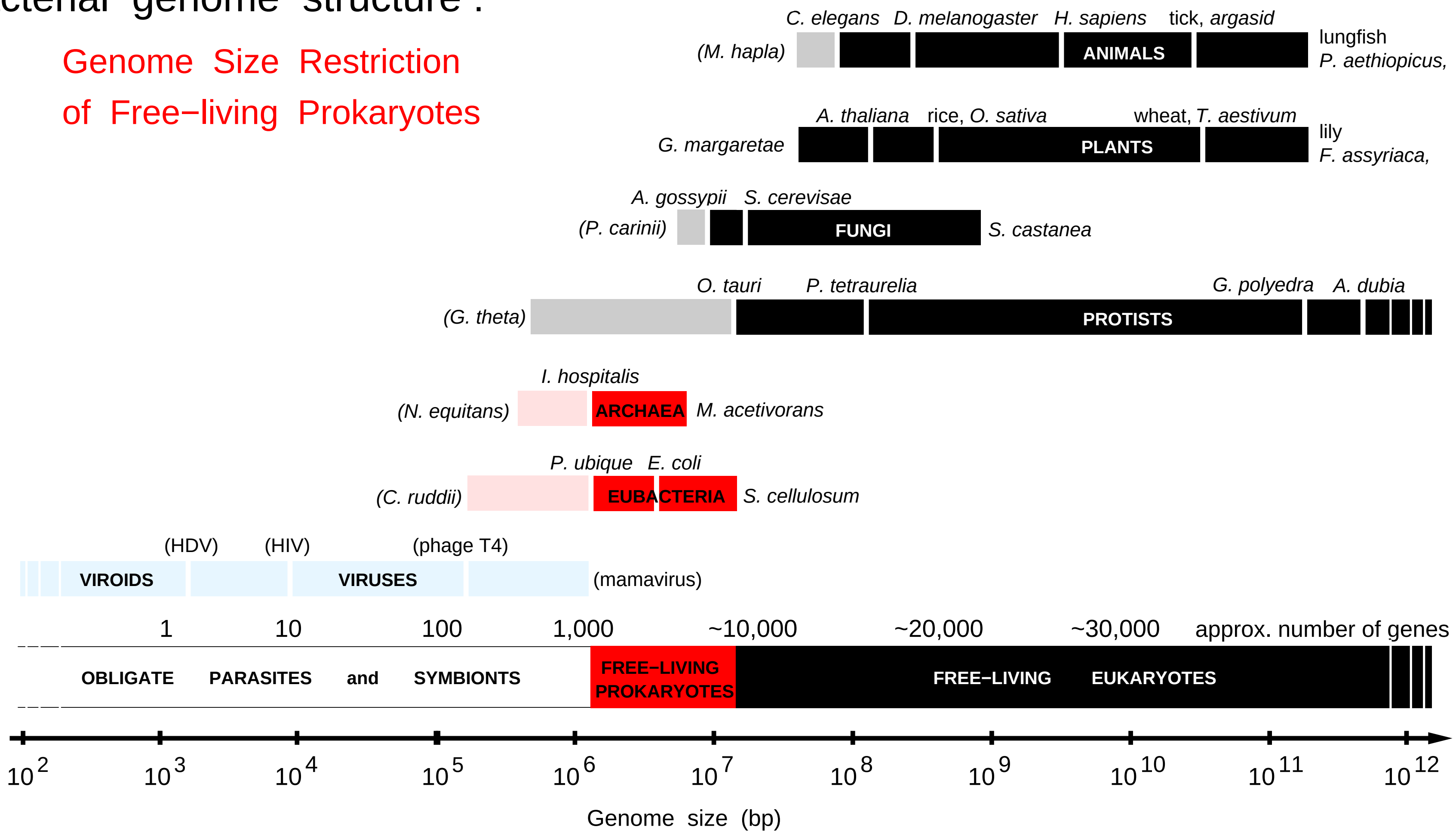
Genome Size Restriction of Free−living Prokaryotes



Alternatively... Specific Evolutionary Constraints of Bacteria?

- Transcription / translation coupling versus mRNA export out of the nucleus
- Bacterial genome structure :

Genome Size Restriction
of Free−living Prokaryotes



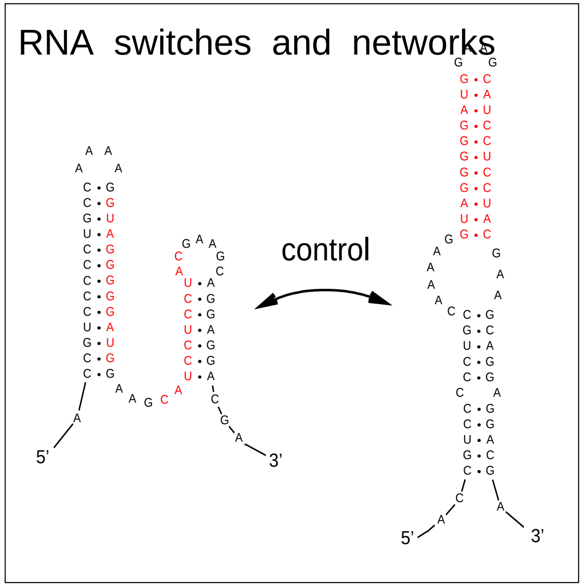
Gene duplications + genome size constraint → gene turnover

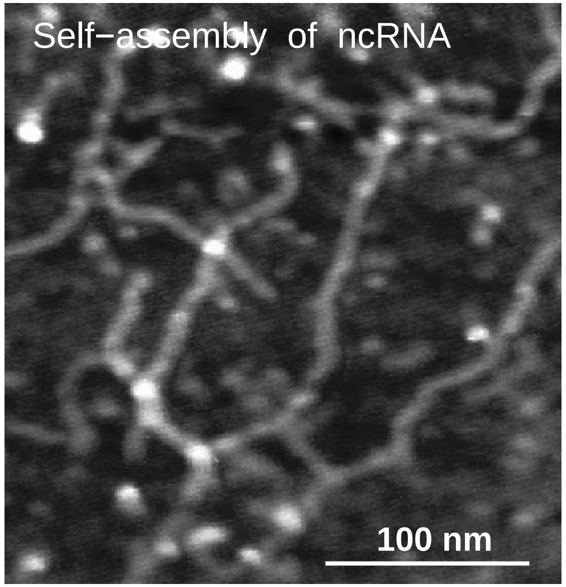
but gene conservation required → "mutualized" gene pool + gene transfers

RNA Dynamics and Biomolecular Systems Lab

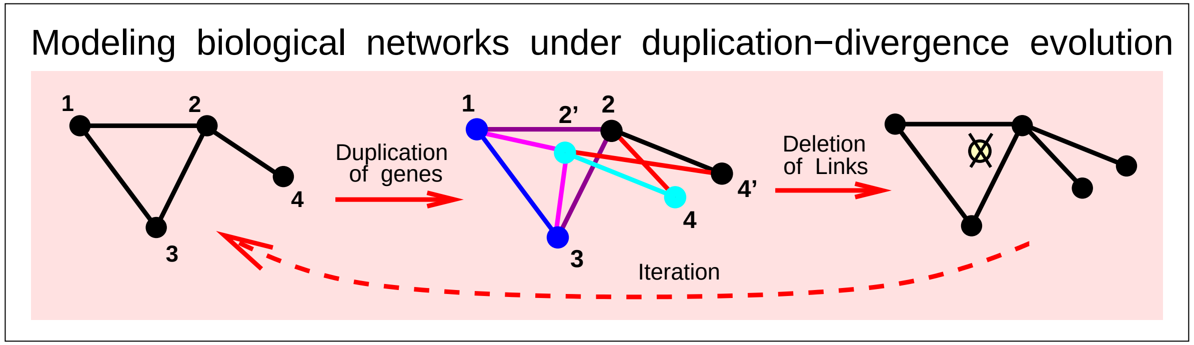
Institut Curie, CNRS, Paris

I. RNA Regulatory and Physical Networks


RNA switches and networks


Self−assembly of ncRNA

II. Evolution of Biological Networks


Modeling biological networks under duplication−divergence evolution

$$. Human Frontier, ANR, Ministère de la Recherche, Institut Curie, CNRS, Fondation PG de Gennes