

# RNA dynamics and Biomolecular Systems

Hervé Isambert, *RNA dynamics and Biomolecular Systems Lab*, Institut Curie, UMR 168, Paris.

## I. RNA Dynamics

Alexandre Dawid  
Bastien Cayrol  
Cédrine Lemoine  
Alain Xayaphoummine

Thomas Buchert  
*Veronique Arluison, IBPC, Paris*  
*Harald Putzer, IBPC, Paris*  
*Irit Sagi, Weizmann*



## II. DNA Nanotechnology

Virgile Viasnoff  
*Amit Meller, Rowland Institute, Harvard*



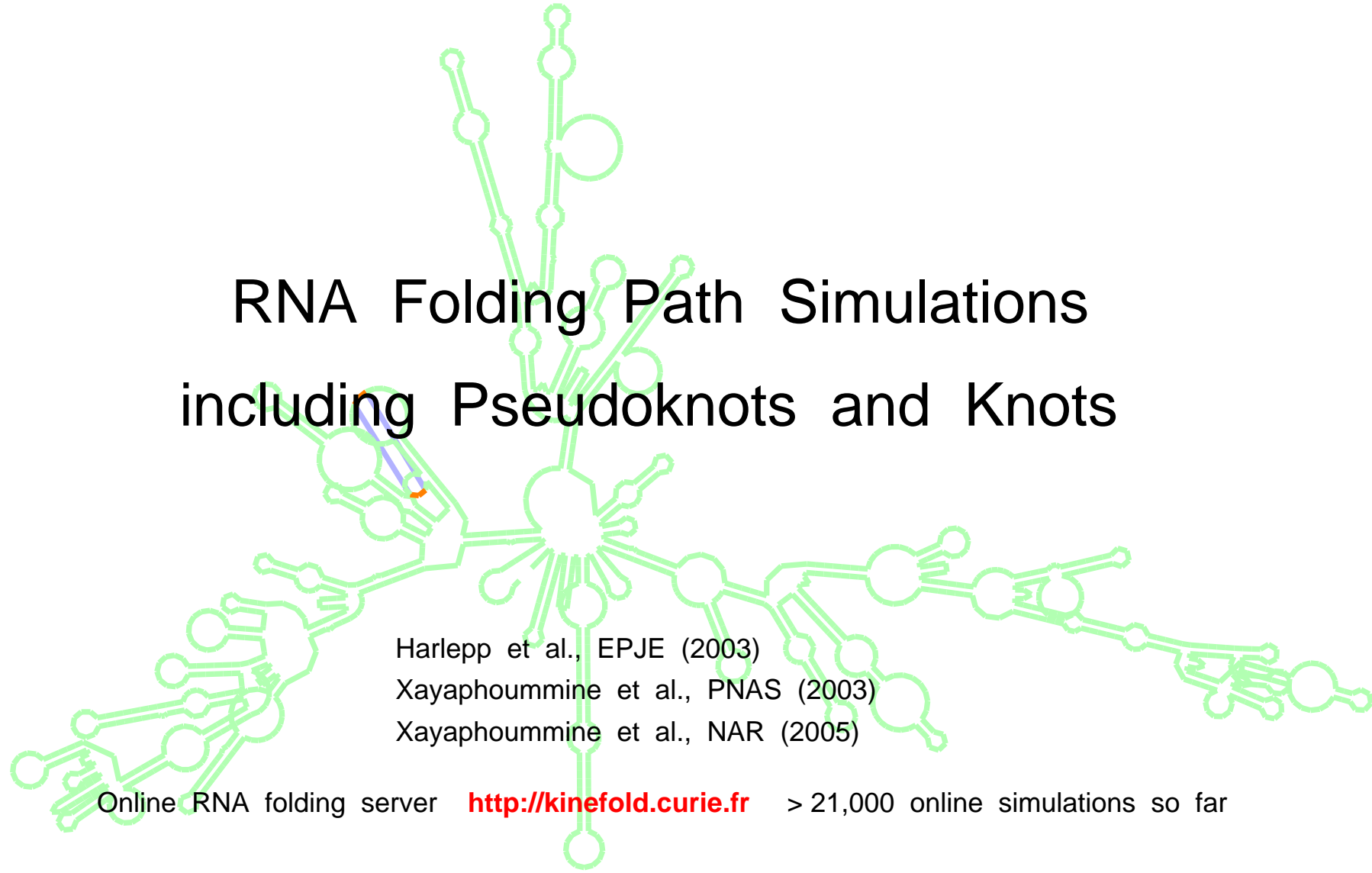
## III. Biological Networks

Kirill Evlampiev  
Marco Cosentino Lagomarsino

Thomas Rolland



\$ HFSP Foundation CNRS Institut Curie  
Ministère de la Recherche



# RNA Folding Path Simulations including Pseudoknots and Knots

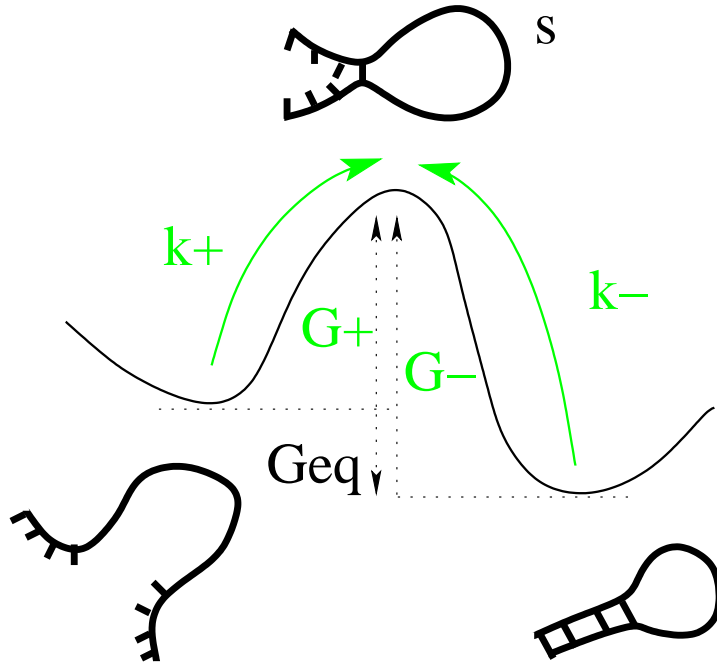
Harlepp et al., EPJE (2003)

Xayaphoummine et al., PNAS (2003)

Xayaphoummine et al., NAR (2005)

Online RNA folding server <http://kinfold.curie.fr> > 21,000 online simulations so far

# DNA / RNA Folding / Unfolding Kinetics



$$k_+ = k_0 e^{-\frac{G_+}{kT}} = k_0 \sum_s \frac{1}{s^{3\nu}} \quad \nu = 0.5-0.6$$

$$k_+^{-1} = 10^{-5} - 10^{-4} \text{ sec} \quad (s = 4 - 20 \text{ bases})$$

(helix zipping  $10^{-7}$  sec)

$$k_- = k_0 e^{-\frac{G_-}{kT}}$$

$$k_-^{-1} = \text{sec, min, hours, days, weeks...}$$

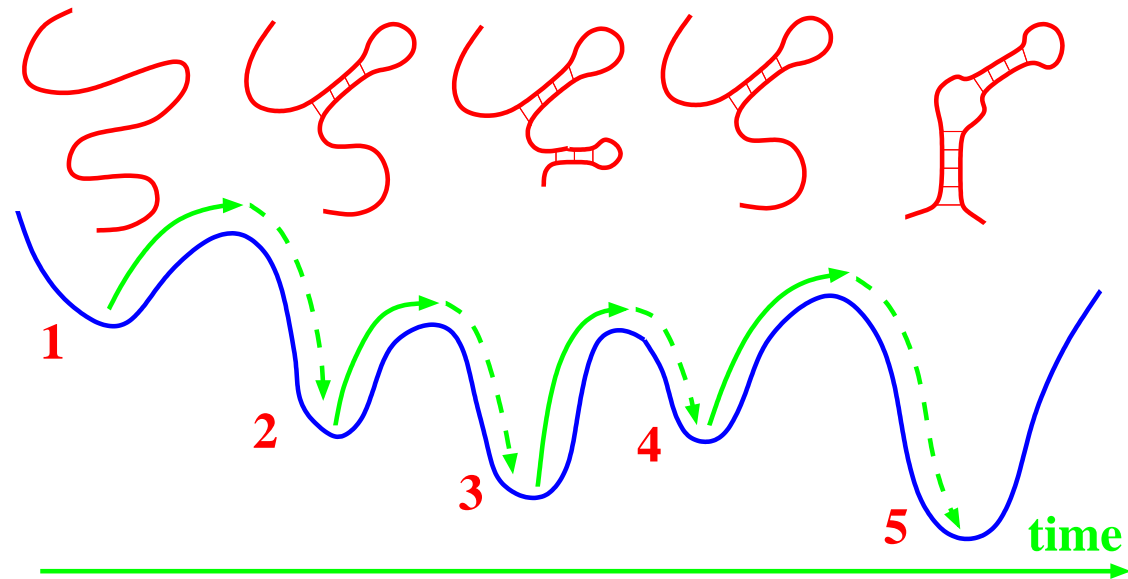
$k_+$  is slow compare to typical molecular time scales, local heat diffusion, etc  
 → DNA switch under folding kinetics control

$k_+$  is fast compare to transcription rate  $< 1$  base/ms  
 → RNA switch under co-transcriptional folding control

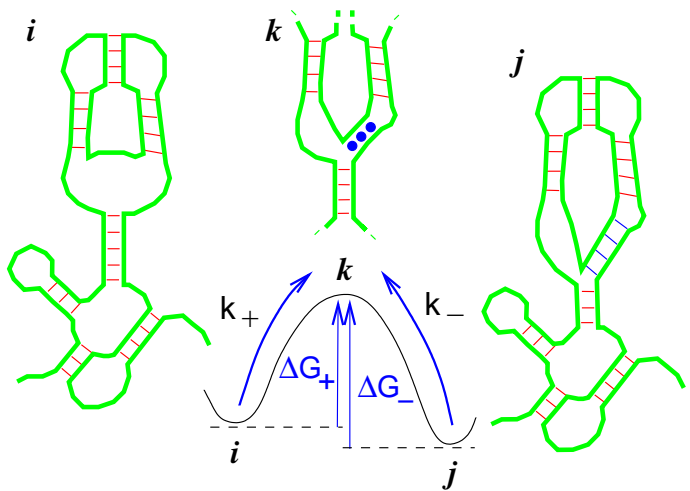
$k_-$  can be arbitrary slow → long-lived out-of-equilibrium folds

# Modeling RNA Folding Kinetics

by following the *stochastic* formation and dissociation of *entire* helices



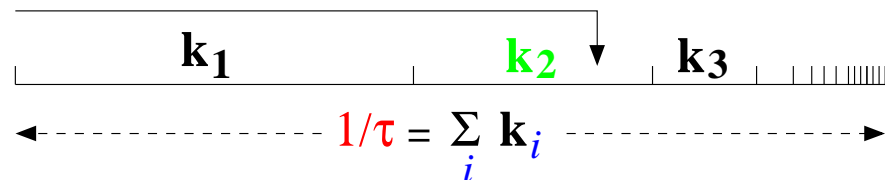
Elementary stochastic transitions :



PNAS 97, 6515 (2000)

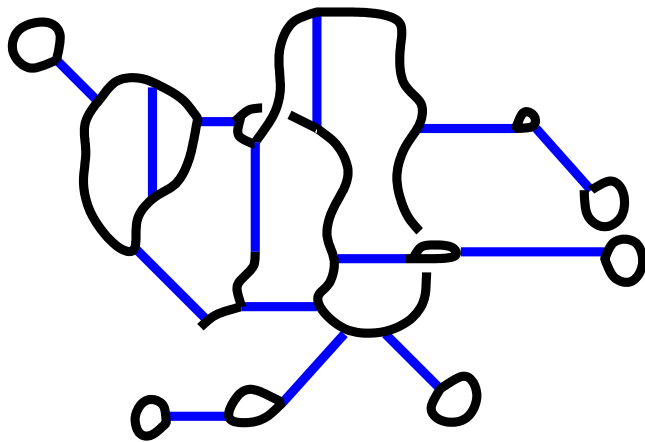
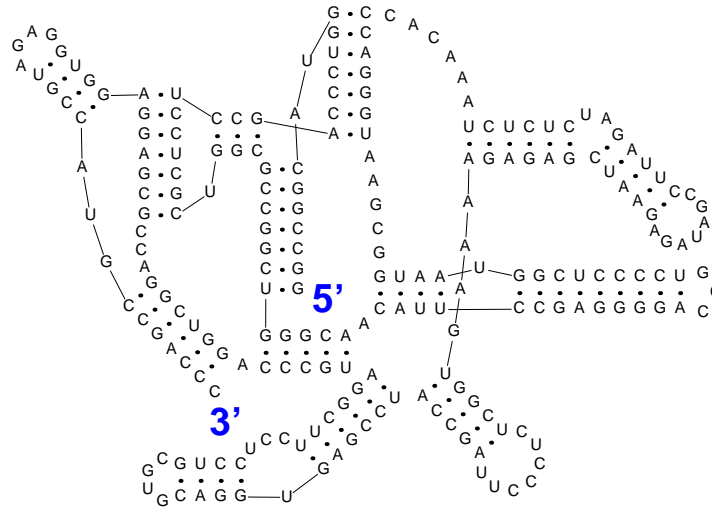
At **each configuration** along the folding path :

- calculate *all* rates :  $k_i = k^0 \exp(-\Delta G_i / kT)$
- choose *one transition* stochastically :



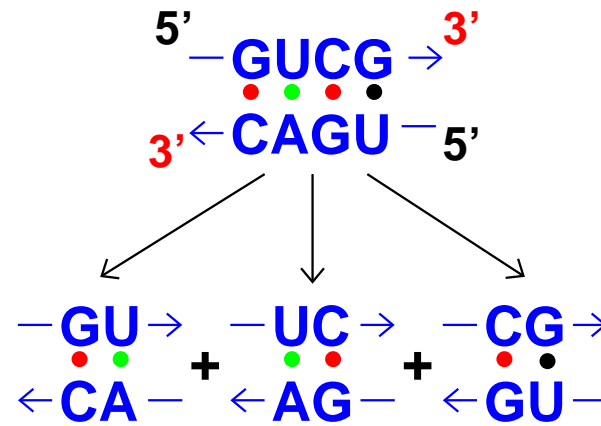
- Efficient speed-up using "*Exactly Clustered Stochastic Simulations*" PNAS 100, 15310 (2003)

# Modeling RNA Secondary Structures



Conformational Entropy

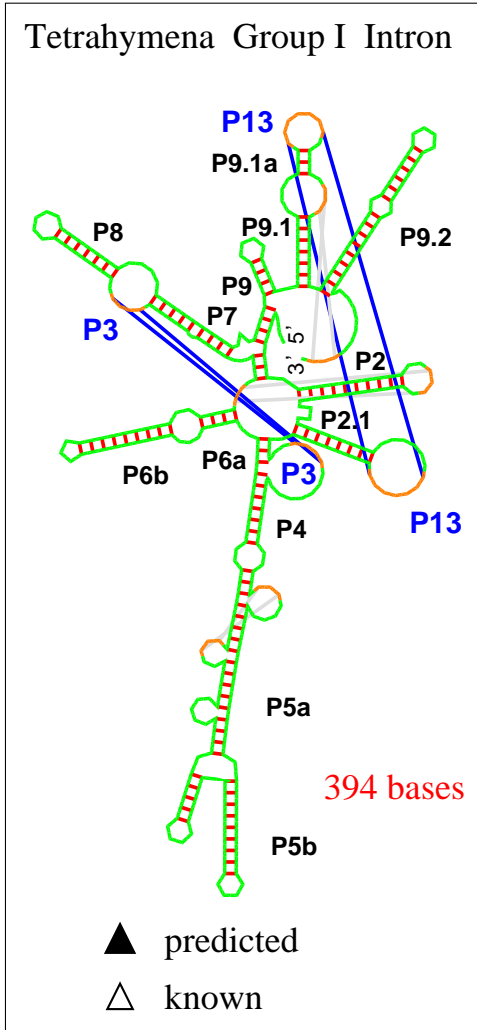
+



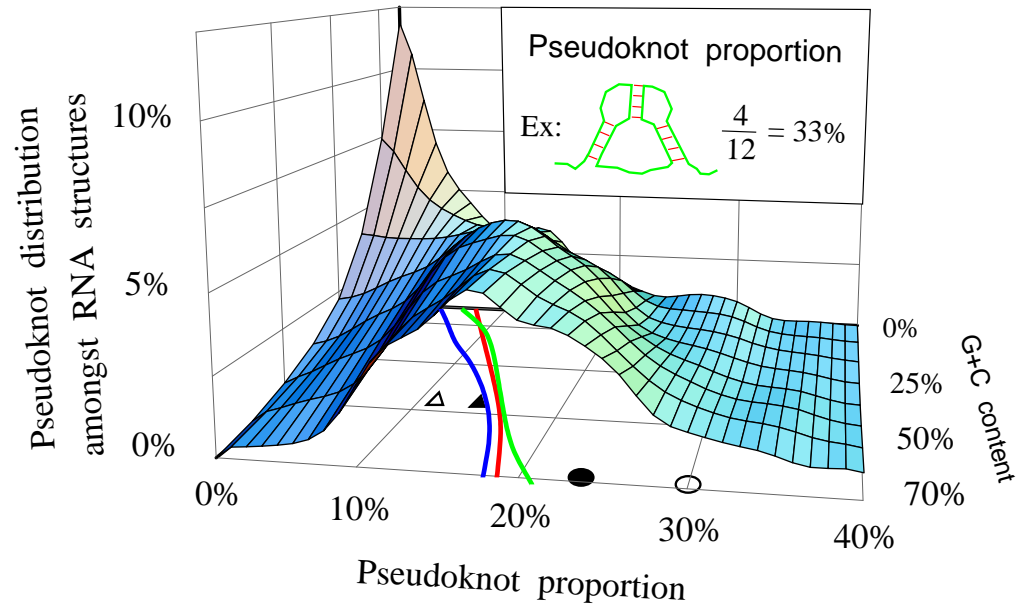
Helix Thermodynamics

# Pseudoknot Prevalence in RNA Structures

PNAS, 100, 15310 (2003)

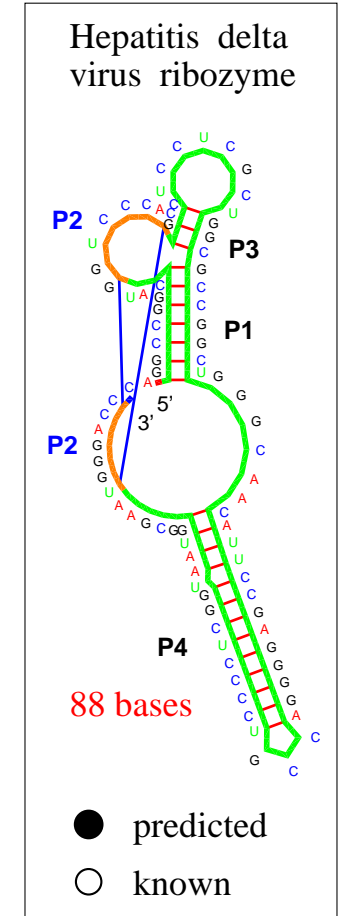


## 100-base-long Random Sequences

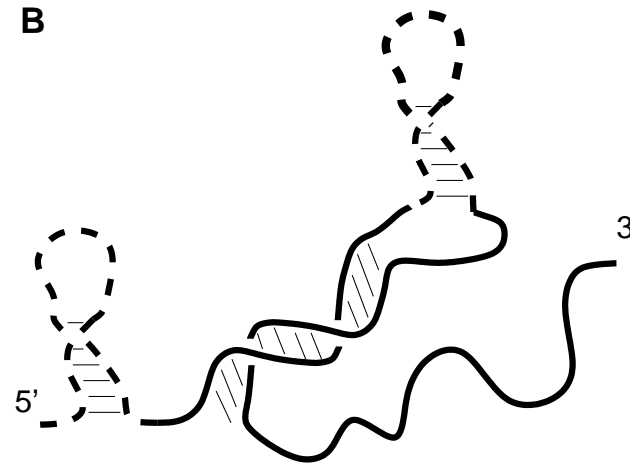
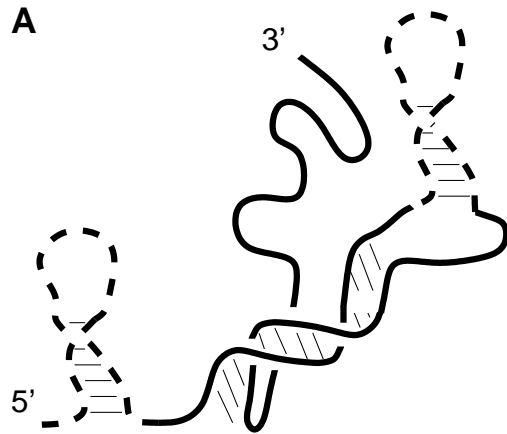


Average Pseudoknot Proportion :

- 50-base-long random sequence
- 100-base-long random sequence
- 150-base-long random sequence

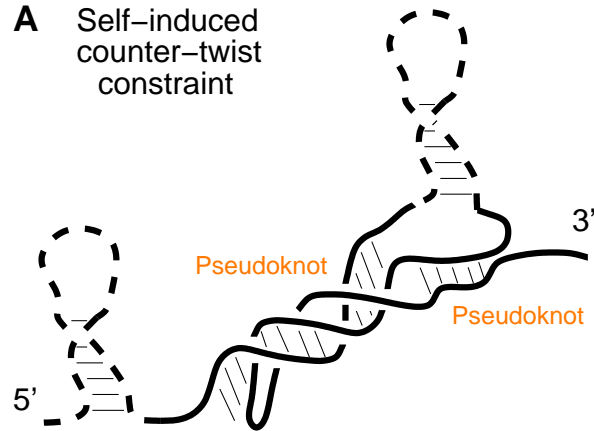


# Pseudoknots versus knots ('entangled helices')

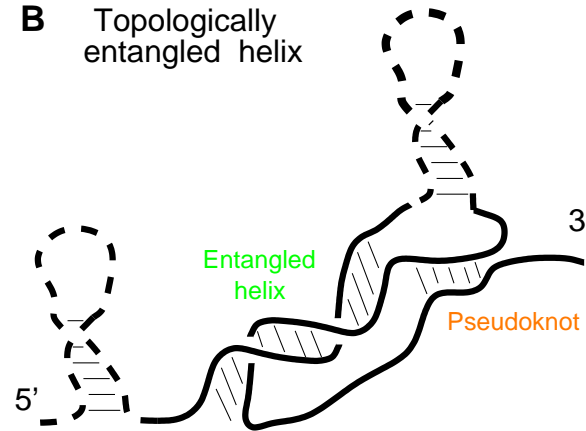


# Pseudoknots versus knots ('entangled helices')

**A** Self-induced counter-twist constraint



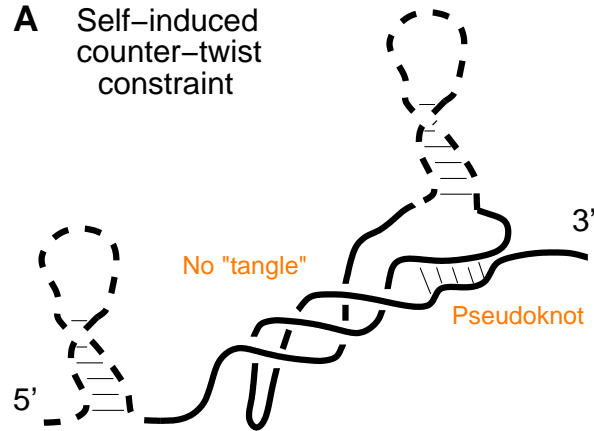
**B** Topologically entangled helix



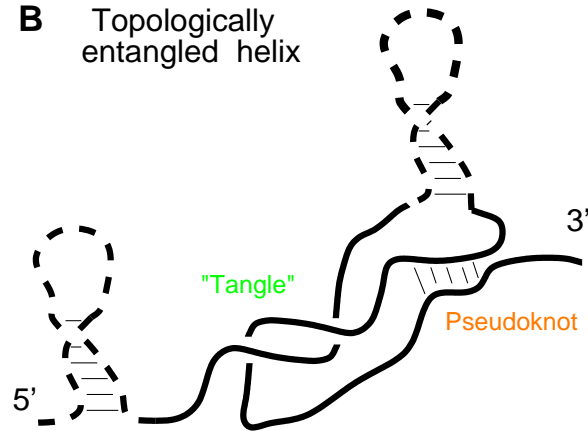


# Pseudoknots versus knots ('entangled helices')

**A** Self-induced counter-twist constraint

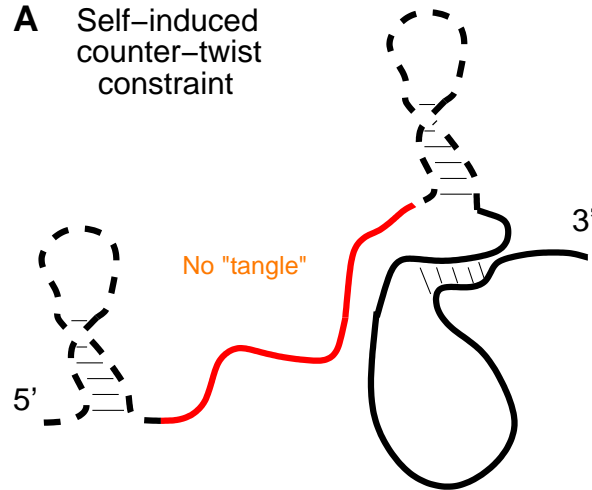


**B** Topologically entangled helix

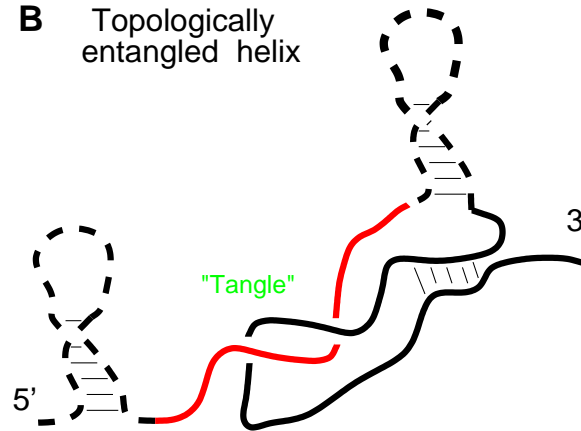


# Pseudoknots versus knots ('entangled helices')

**A** Self-induced counter-twist constraint

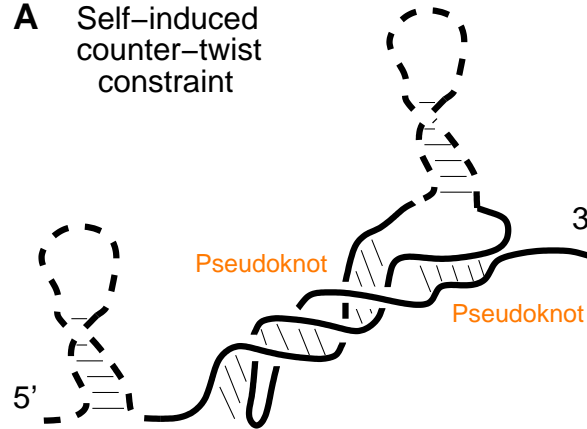


**B** Topologically entangled helix

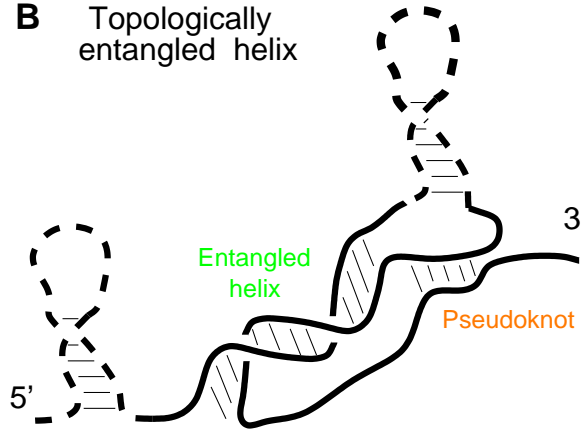


# Pseudoknots versus knots ('entangled helices')

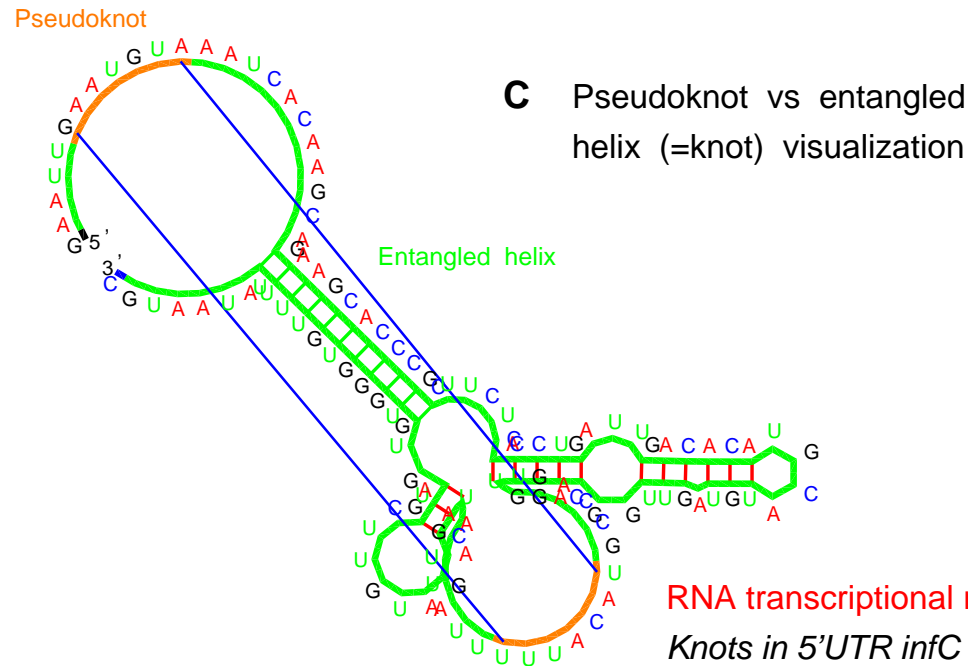
**A** Self-induced counter-twist constraint



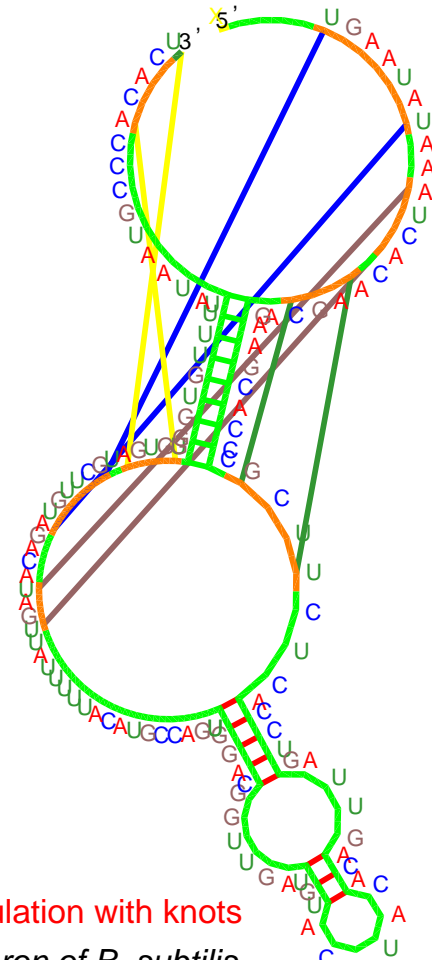
**B** Topologically entangled helix



Pseudoknot



**C** Pseudoknot vs entangled helix (=knot) visualization



RNA transcriptional regulation with knots  
*Knots in 5'UTR infC operon of B. subtilis*  
(exp by N. Choonee & H. Putzer, IBPC)



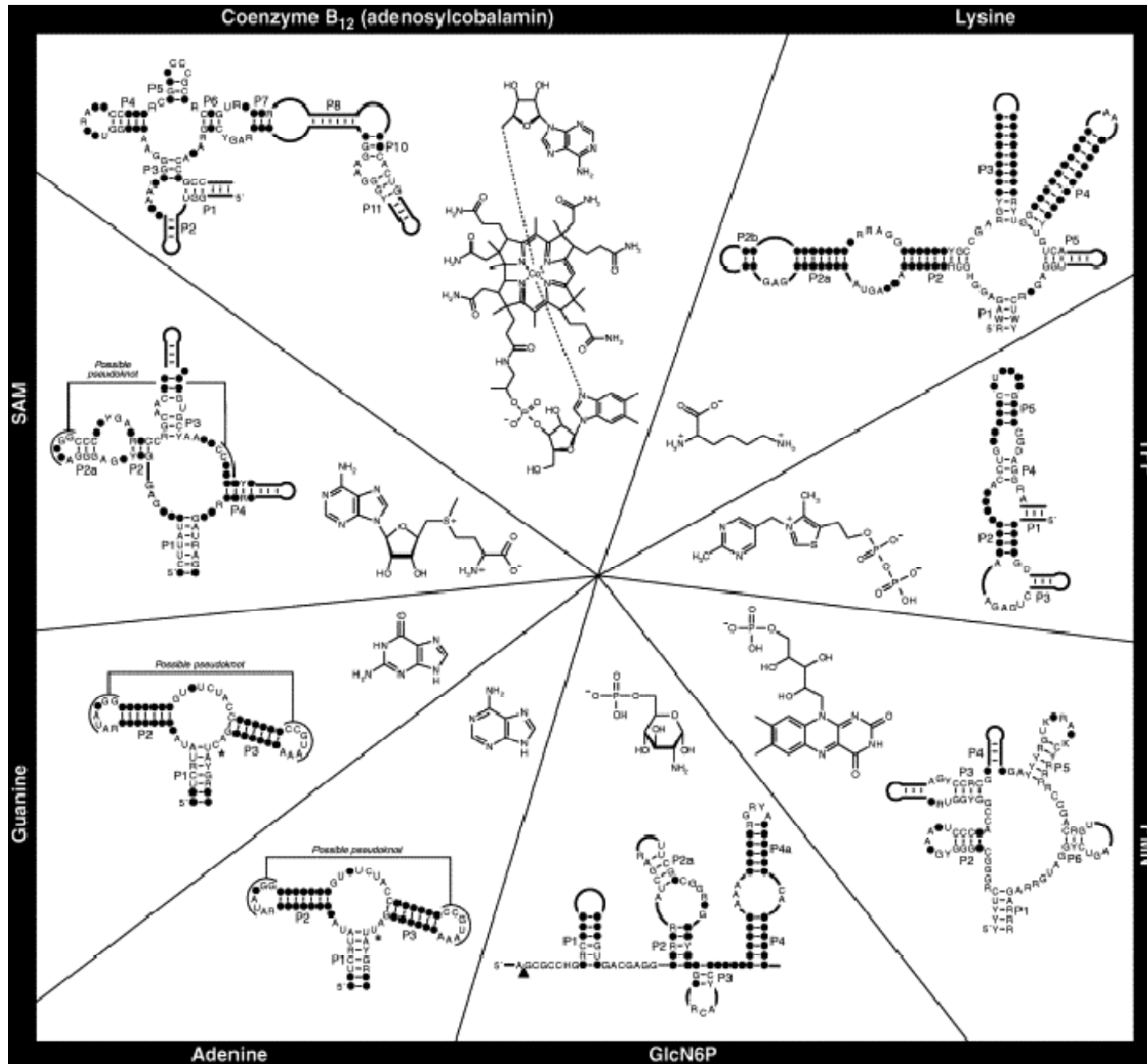
# Encoding Folding Paths of RNA Switches under Antisense Regulation

Viasnoff et al., NanoLett. (2006)

Xayaphoummine et al., NAR (2007)

Dawid, Cayrol, Arluison

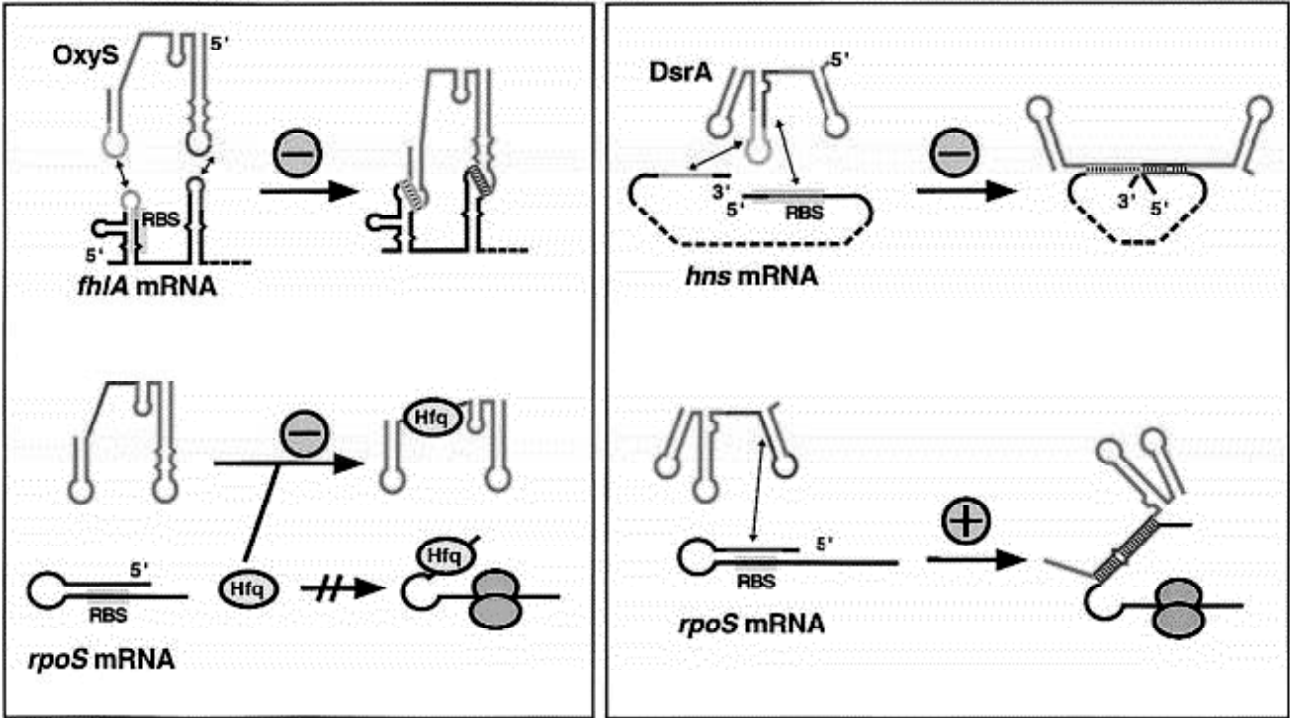
# Riboswitches (metabolite-induced conformational change)



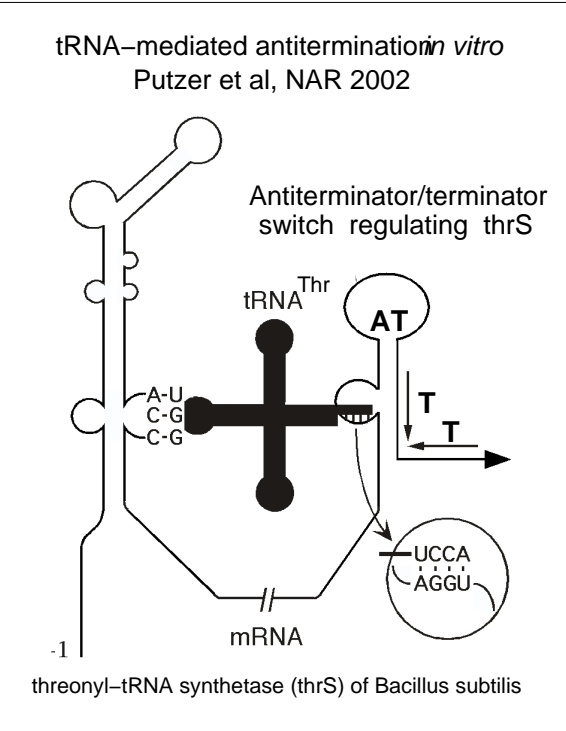
Breaker et al.  
2002–2004

# (Post)-transcriptional regulation through RNA switch control

*Escherichia coli*



*Bacillus subtilis*

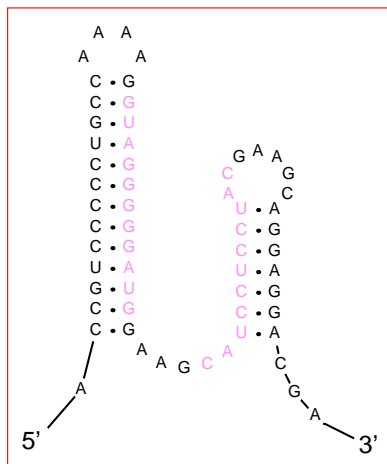


→ combine : antisense interaction + structural switching

# Transcription-guided vs equilibrium folding

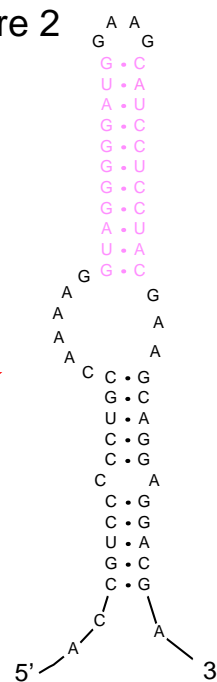
Example of simple RNA switch design (both structures have equivalent energy here)

structure 1



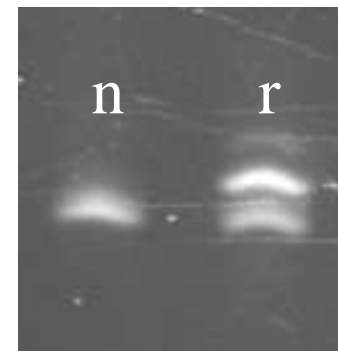
Structure after "transcription"

structure 2



One "native" structure  
**time < 1 day at 37C**  
 after "transcription"

Two "relaxed" structures  
**time > 1 day at 37C**  
 or time > 5 min at 85C

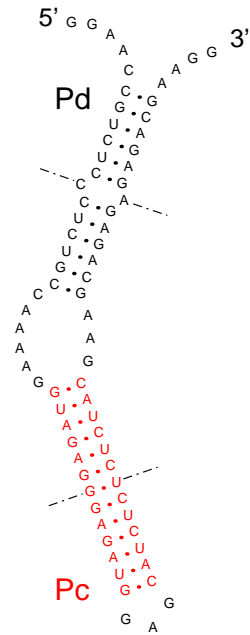


— structure 2  
 — structure 1

"Direct" RNA switch (D)

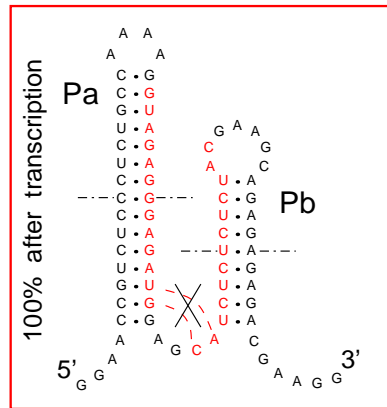
(5'-ABCD-3')

structure 2D



co-transcriptional folding at 37°C ~~↕~~ 0%

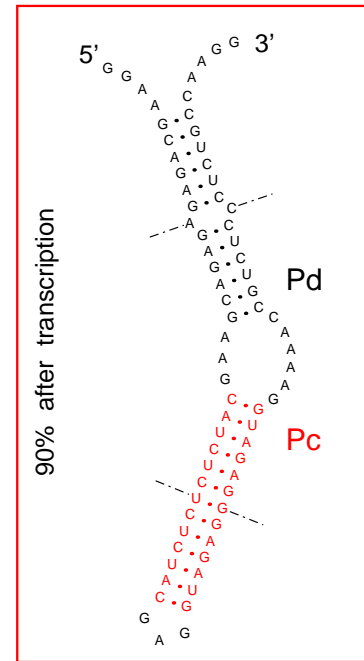
structure 1D



"Reverse" RNA switch (R)

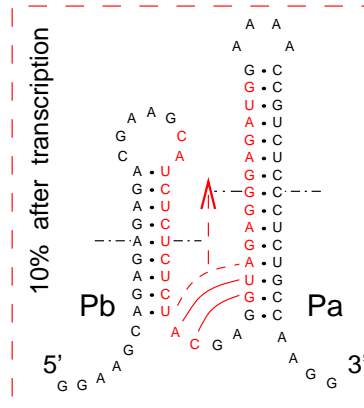
(5'-DCBA-3')

structure 2R



90% ↕ co-transcriptional folding at 37°C

structure 1R



D renat.

D natif

R natif

R renat.

structure 2D

structure 1D



structure 2R

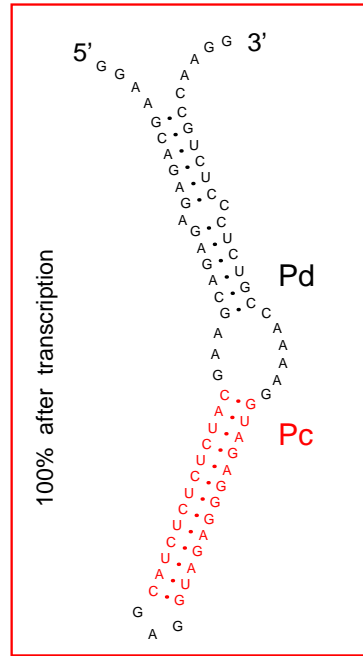
structure 1R



"Reverse" RNA switch (R)

slow transcription (25°C)

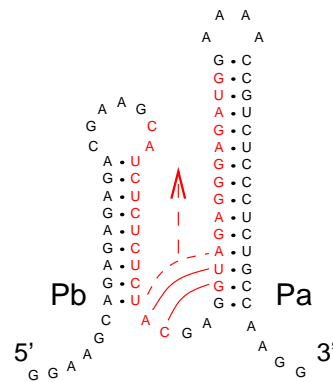
structure 2R



co-transcriptional folding at 25°C

100 %

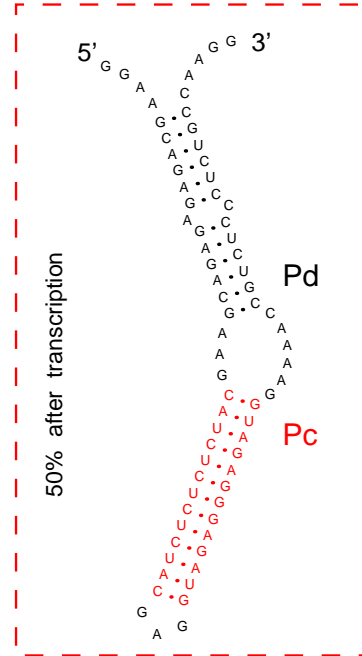
structure 1R



"Reverse" RNA switch (R)

with antisense 5'-CCTCTAC-3'

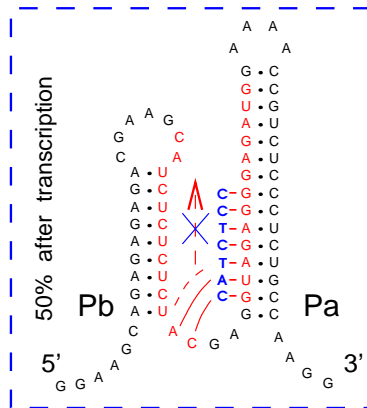
structure 2R



50 %

co-transcriptional folding at 37°C with 5'-CCTCTAC-3'

structure 1R



R renat.

R native

R+oligo native

R+oligo renat.

structure 2R  
structure 1R

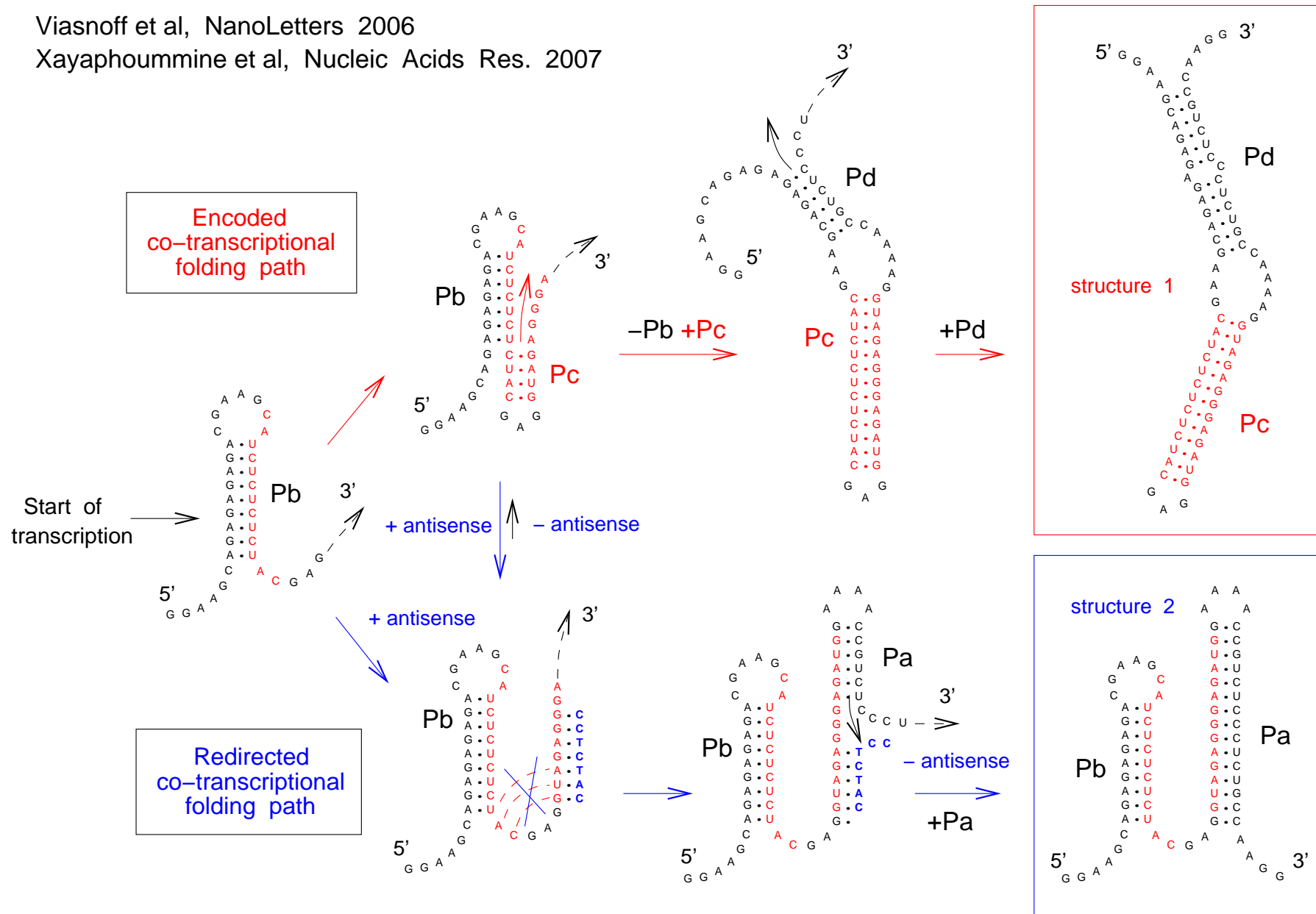


structure 2R  
structure 1R

# Encoded and regulated folding paths of RNA switches

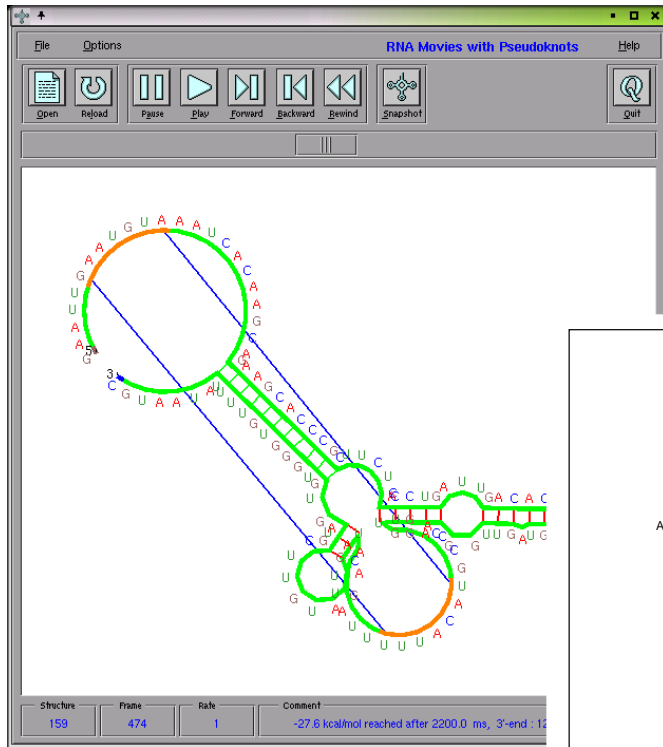
Viasnoff et al, NanoLetters 2006

Xayaphoummine et al, Nucleic Acids Res. 2007

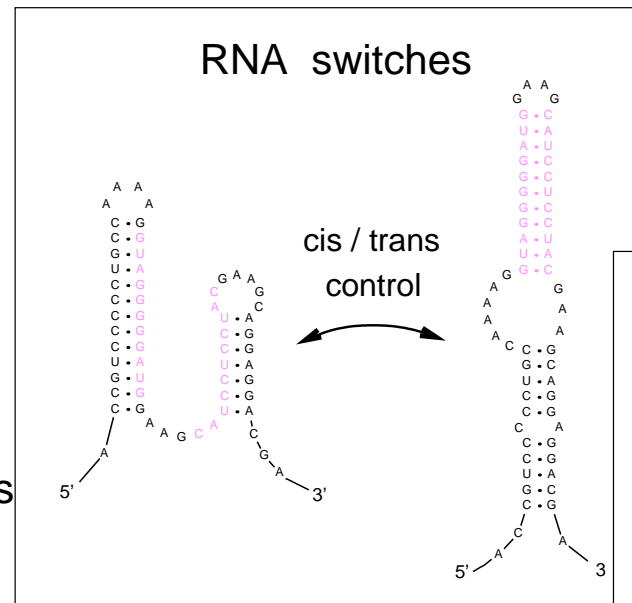


➔ Constitutive coupling between RNA synthesis and folding regulation

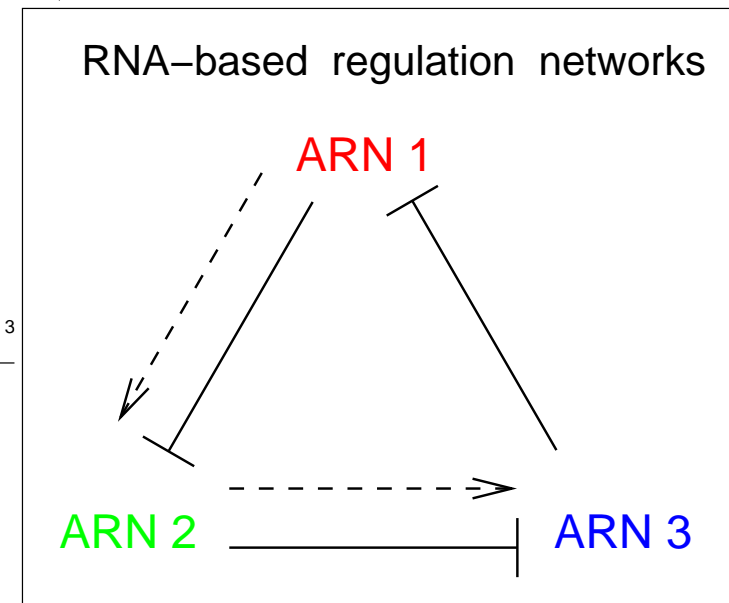
# Exploring the principles of RNA-based regulation networks



folding/interaction simulations



RNA network components



In vitro/in vivo RNA-based networks



# Evolution of Large Biological Networks under General Duplication–Divergence Models

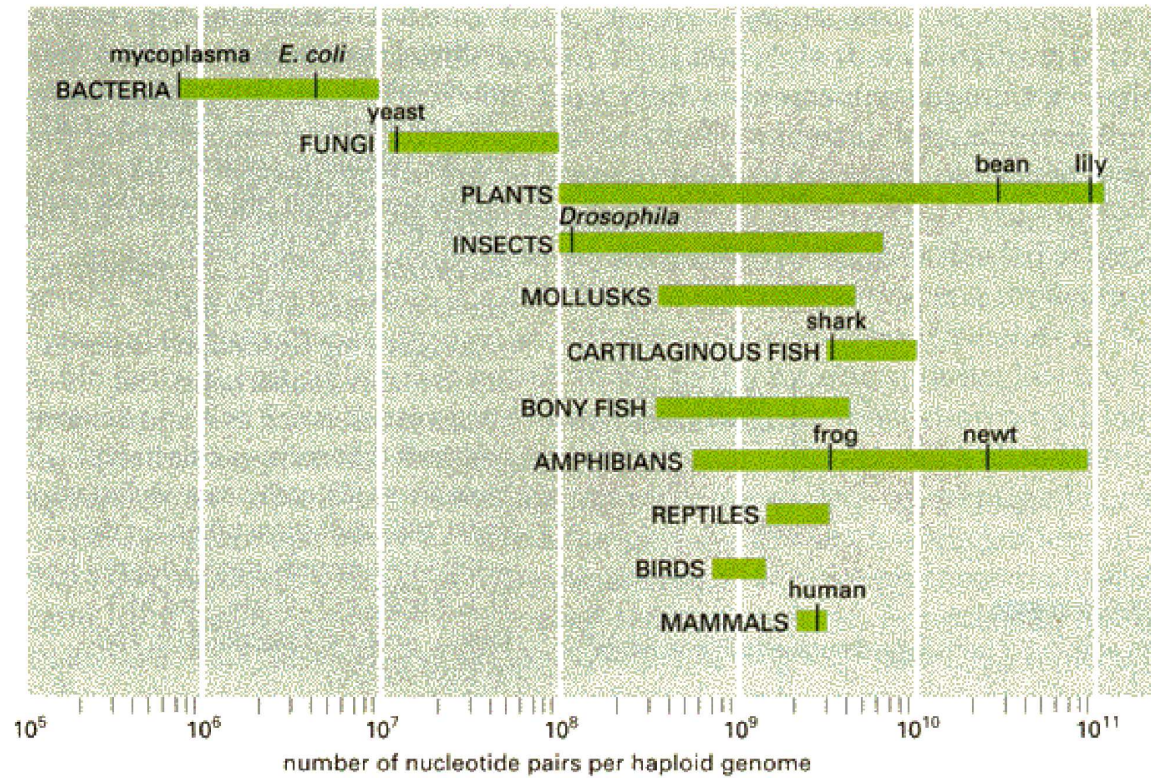
Evlampiev & Isambert, [arxiv.org/abs/q-bio.MN/0606036](https://arxiv.org/abs/q-bio.MN/0606036)

Evlampiev & Isambert, [arxiv.org/abs/q-bio.MN/0611070](https://arxiv.org/abs/q-bio.MN/0611070)

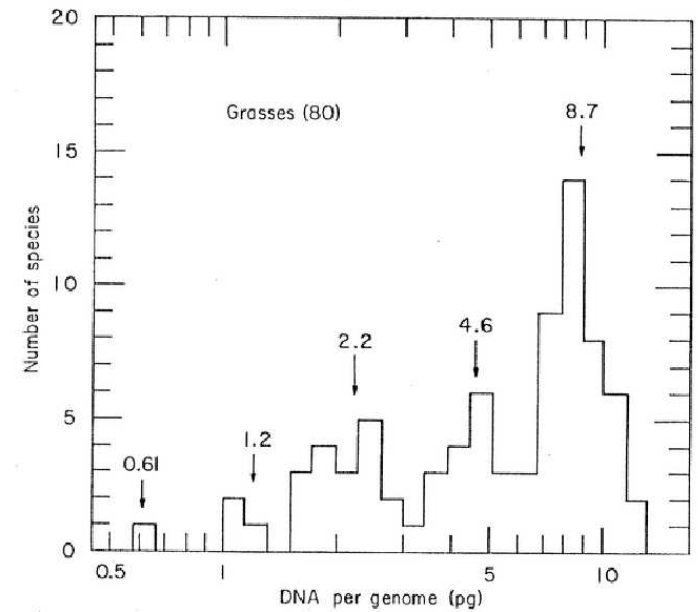
# Genome size distribution

what was known before sequencing

5 decades (8 including viruses)

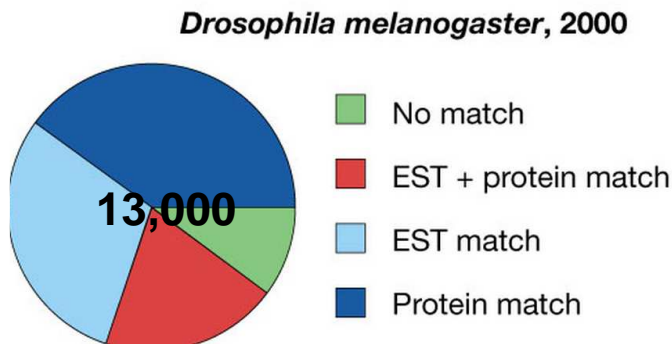
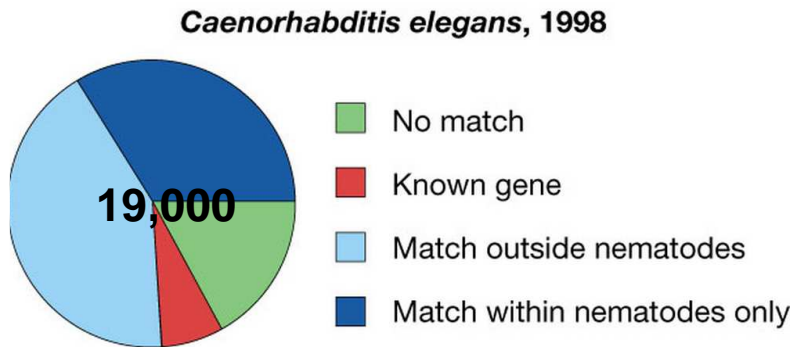
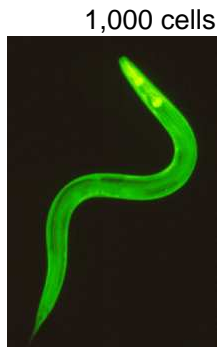
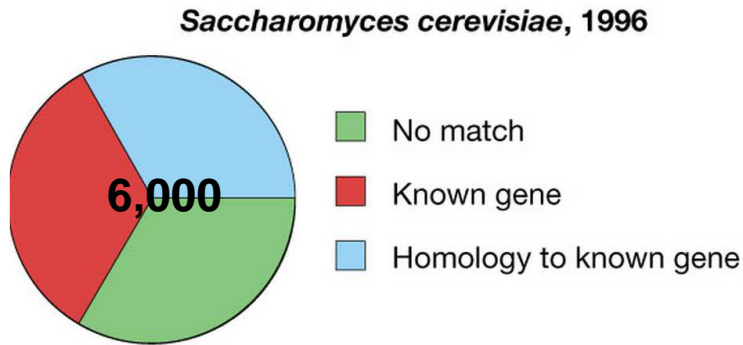
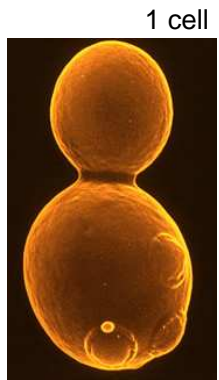


Even multimodal distribution within some species families



Sparrow 1976

# Ten Years of Genome Sequencing: *few genes... same genes...*



*Tetraodon* **22,000**



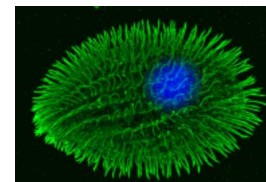
*Human* **20,000–25,000**



*Arabidopsis* **29,000**



*Gallus* **20,000–23,000**



*Paramecium* **39,000**

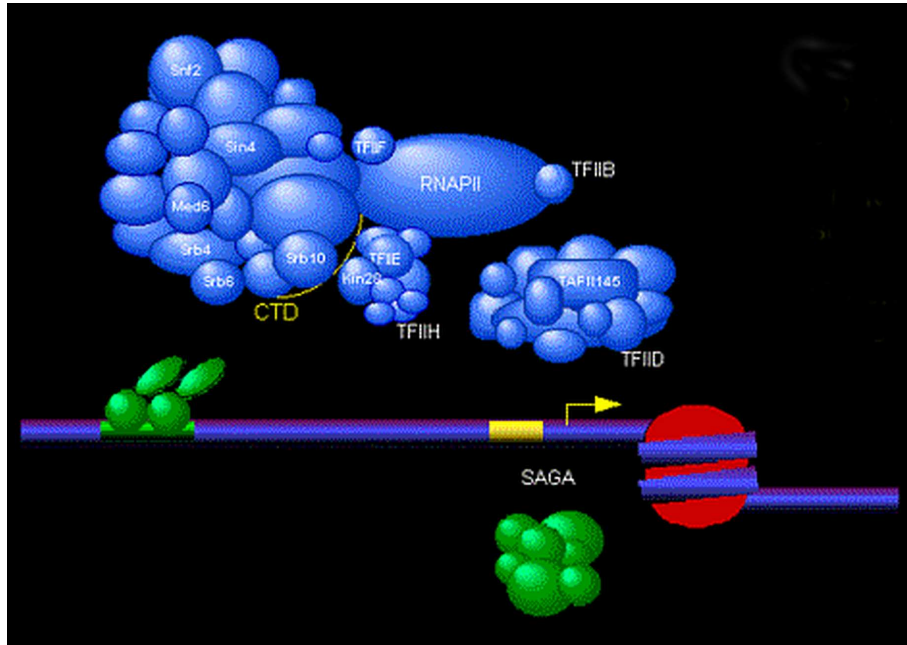


*Wheat* **75,000 ?**  
(allohexaploid)

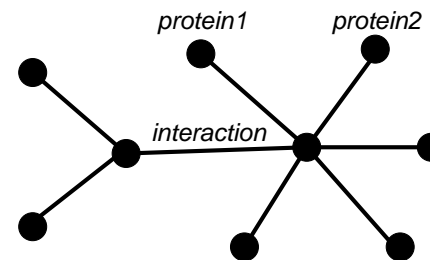
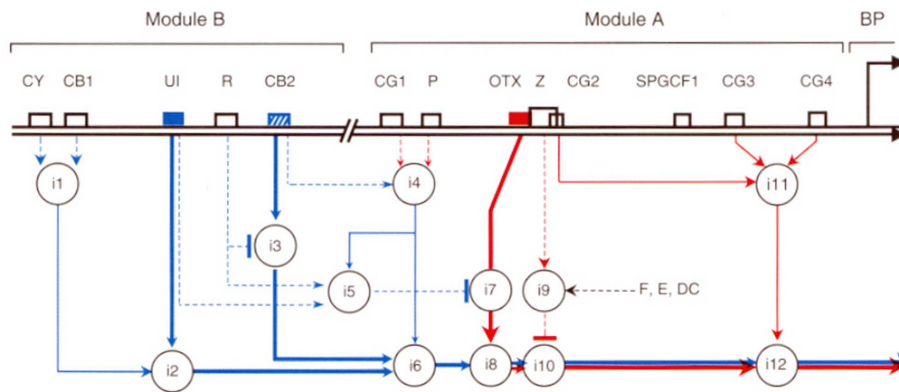
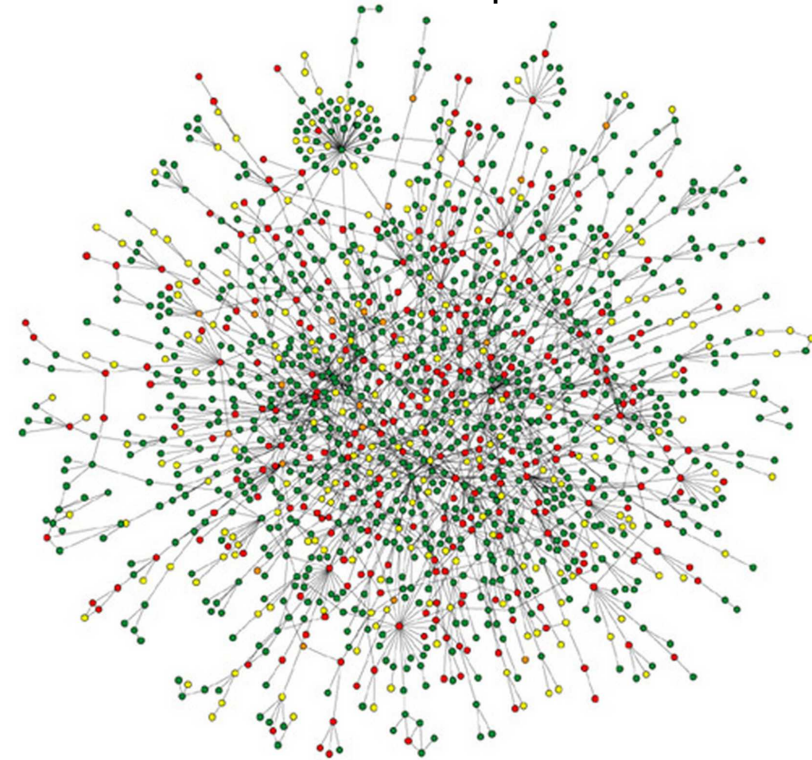
➔ Evolution through Duplication–Divergence of Genes and Genomes

# Combinatorial Gene Expression

Detailed interaction logic (AND/OR/NOT)

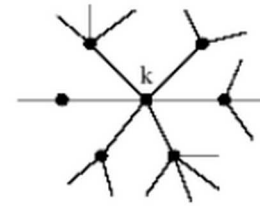


Network-like schematic representations :



**Combinatorics**  
without logic !!

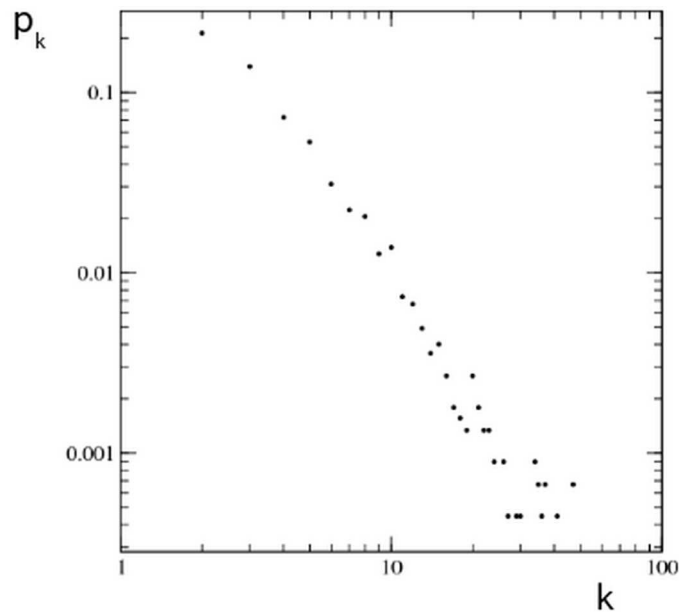
# PPI Networks: Properties



- $N$  number of nodes,  $L$  number of links  
 $N_k$  number of nodes with  $k$  neighbours  
 $p_k$  **degree distribution**,  $\langle k \rangle$  mean degree

$$p_k = \frac{N_k}{N}, \langle k \rangle = \sum_{k \geq 1} k p_k = \frac{2L}{N}$$

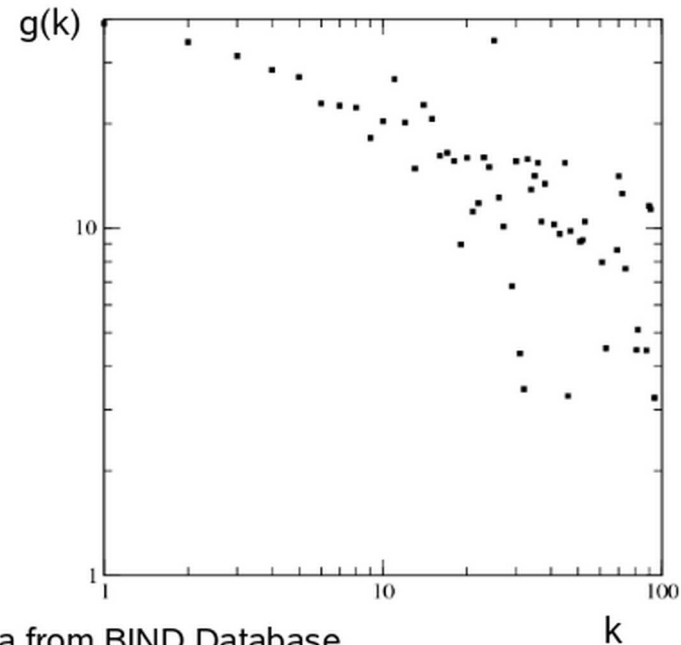
- $p_k$  is a **local** characteristics  $p_k \propto \frac{1}{k^\alpha}, k \gg 1$   
 $\langle k \rangle = 3.93$ ,  $\alpha$  between 2 and 3



- $g(k)$  **average connectivity of neighbours**

$$g(k) = \sum_{(n: d_n=k)} \sum_{i \text{ n.n.}} \frac{N_i}{k N_k}$$

the simplest **non local** distribution



Yeast Data from BIND Database



# 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

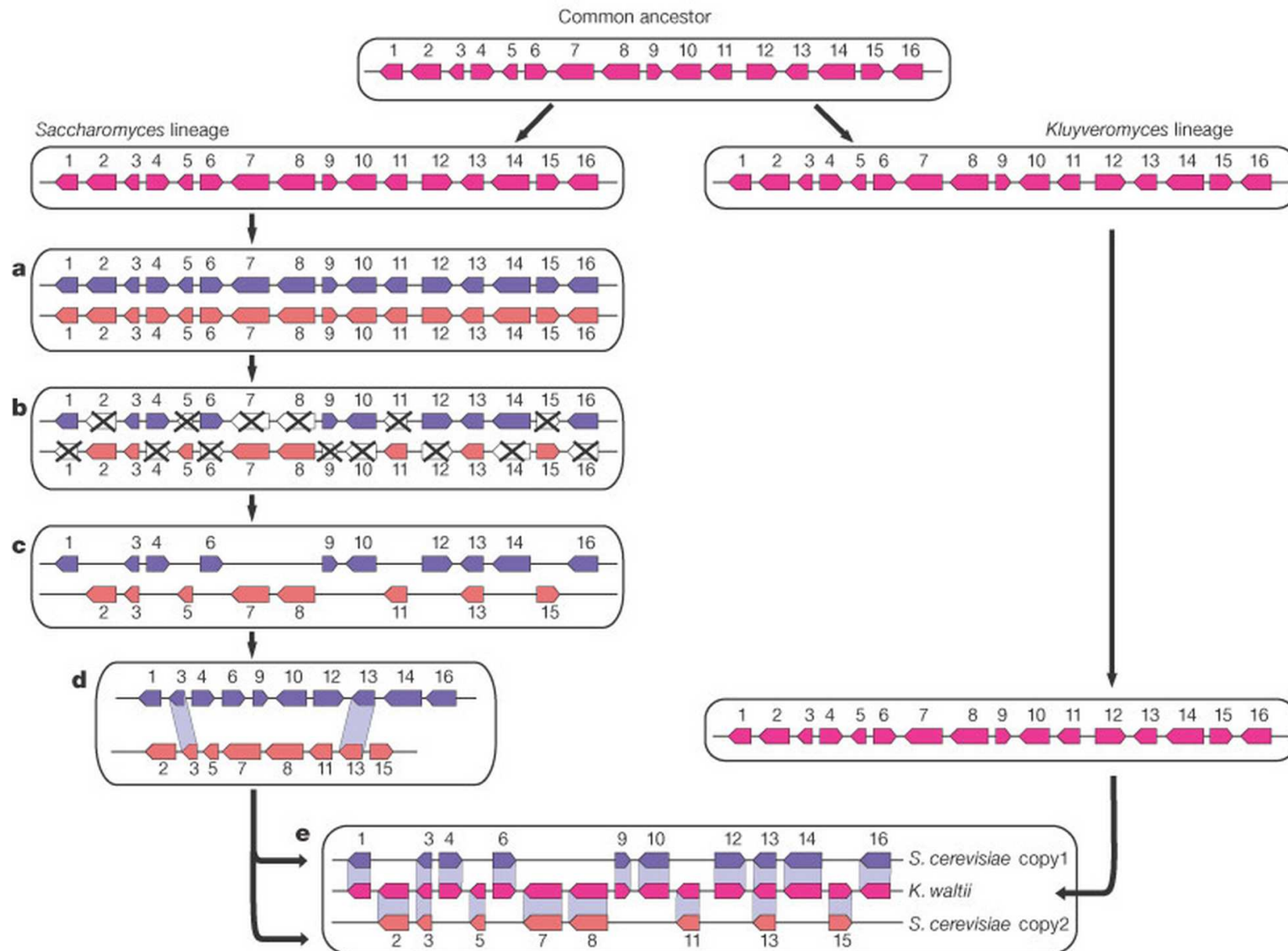
- Random mutations (nucleotide substitutions)
  - synonymous (the same aminoacid)
  - nonsynonymous (a different aminoacid)
  - nonsense codon modifications (Stop)
- Recombinations – crossing over/gene conversion
  - exchange of homologous sequences between homologous chromosomes
  - non reciprocal exchange (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!!**

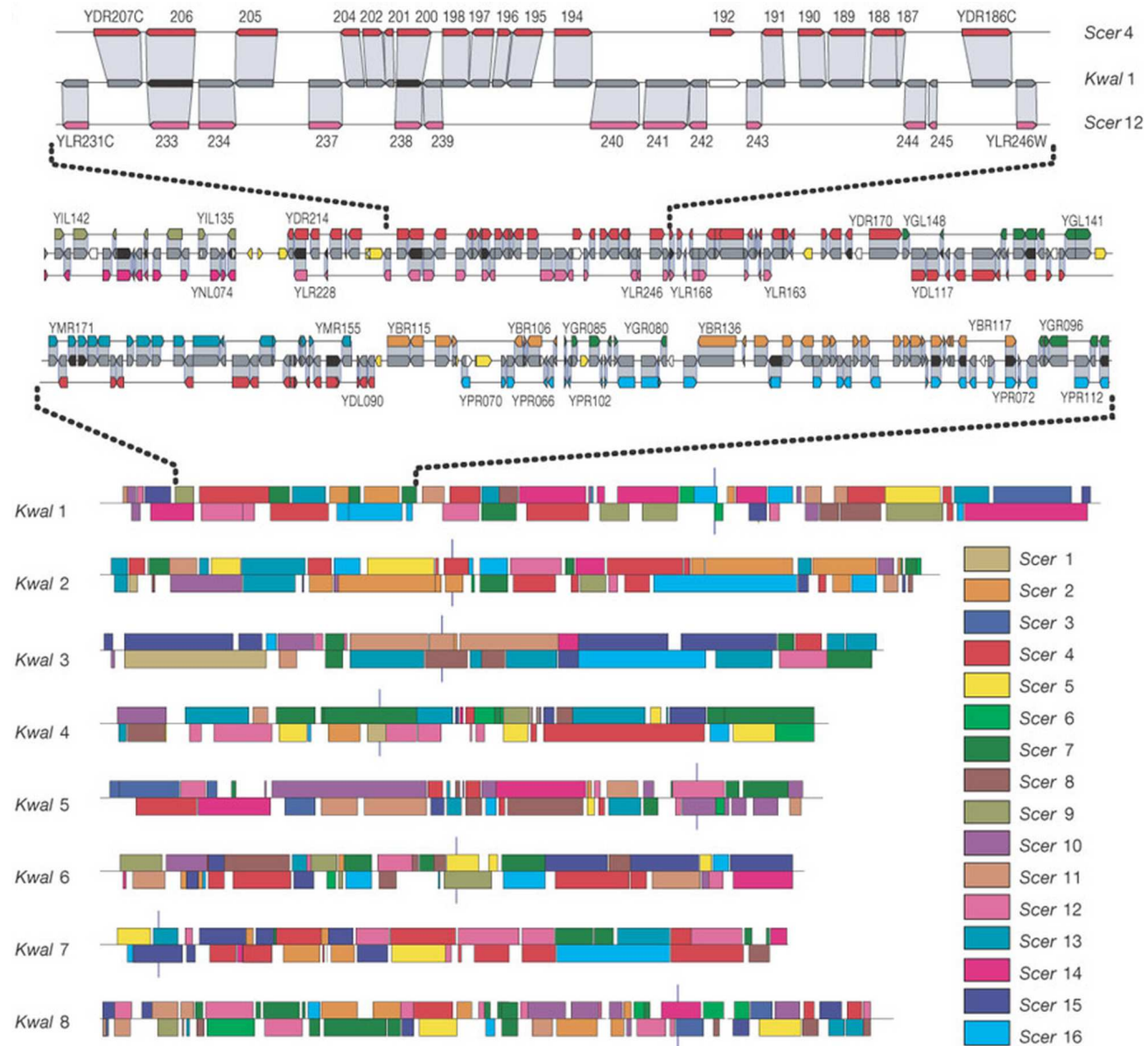
# Whole Genome Duplication in Yeast Genome

## Kellis et al. 2004

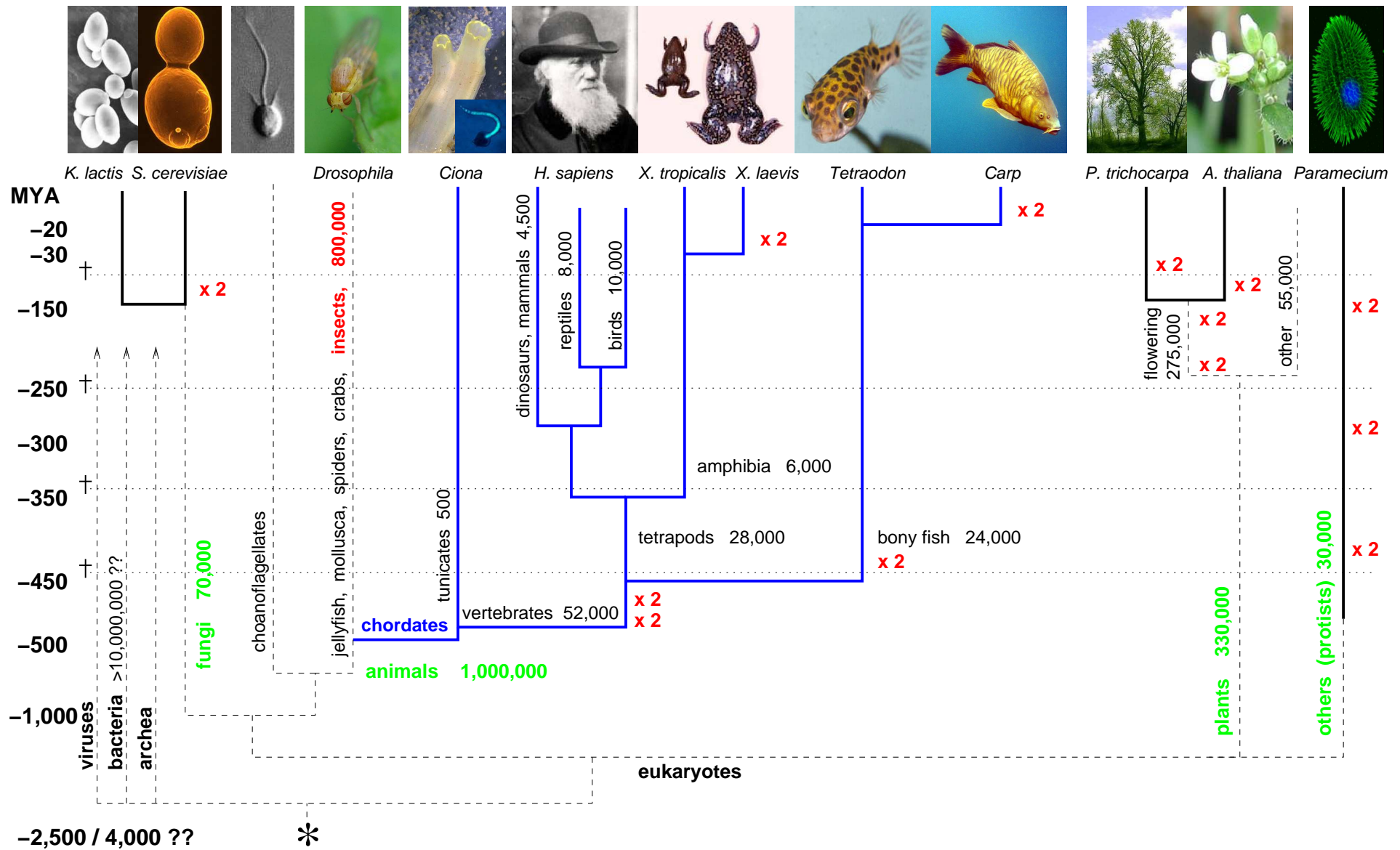


# Whole Genome Duplication in Yeast Genome

## Kellis et al. 2004



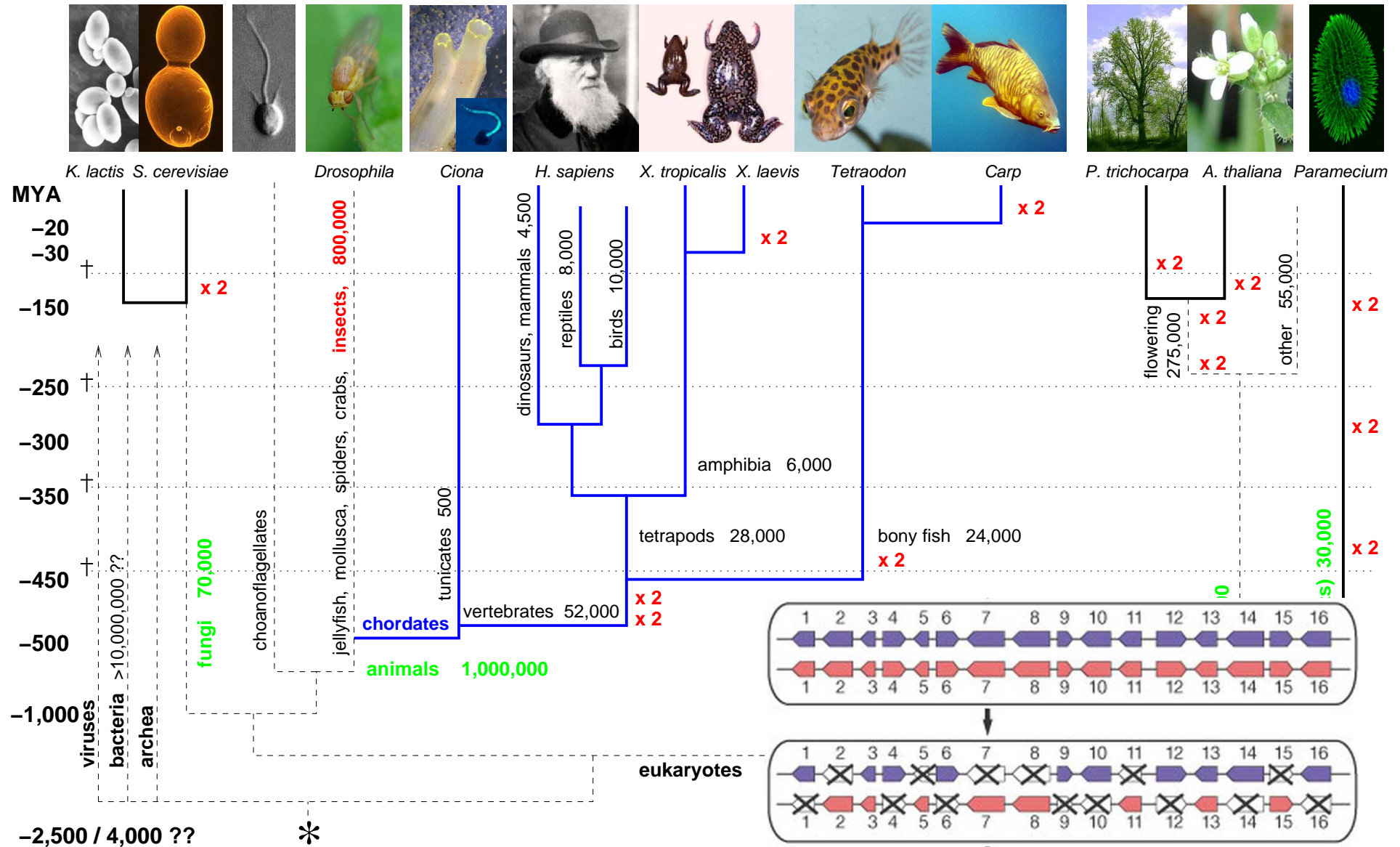
# Whole Genome Duplications in Evolution



Whole Genome Duplications promote:

- Speciation events
- Shuffling of protein domains

# Whole Genome Duplications in Evolution



Whole Genome Duplications promote:

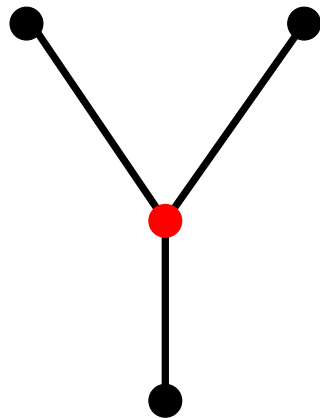
- Speciation events
- Shuffling of protein domains

# 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

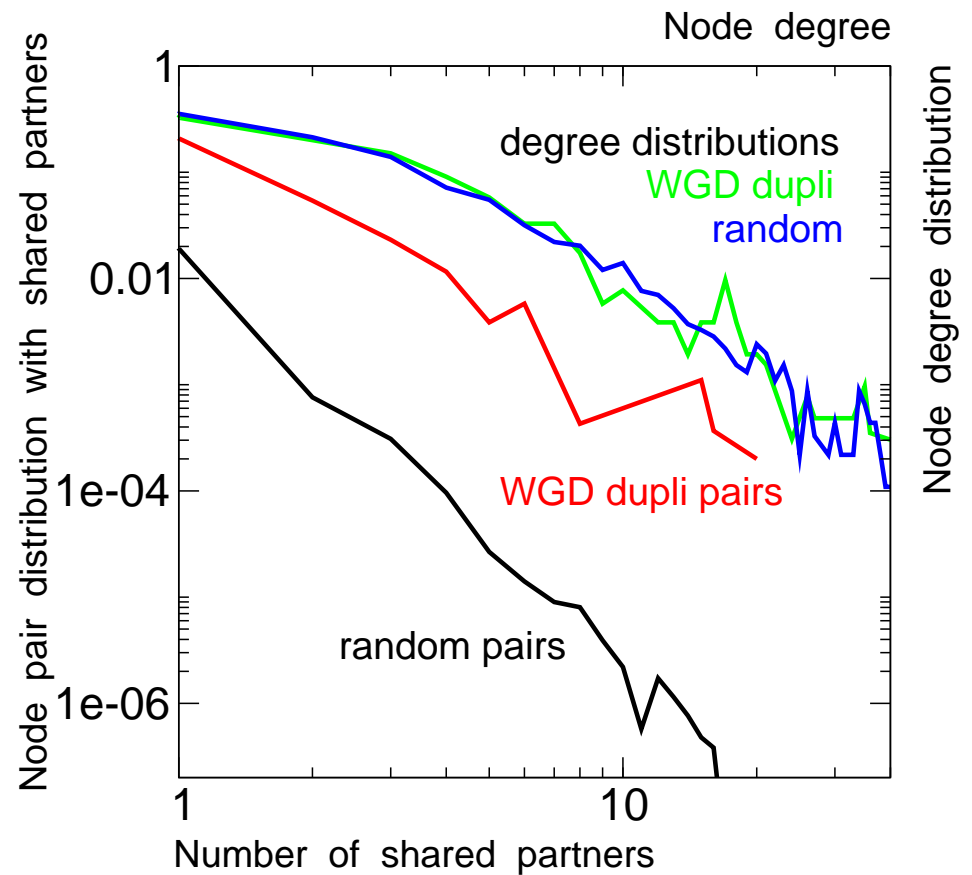
PPI network



Proba to share  $k+$  partners :

$k=1$  WGD dupli > 20 x random pairs

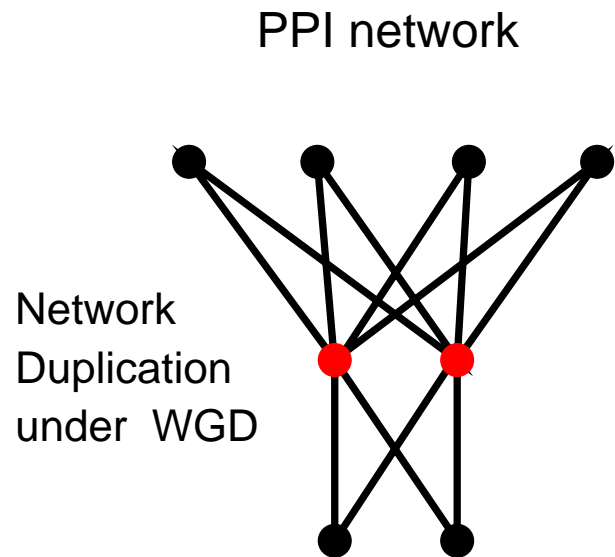
$k=10$  WGD dupli > 1,000 x random pairs



# Evidence of Genome Duplication on PPI network

Yeast 6000 genes : 4000 in PPI network

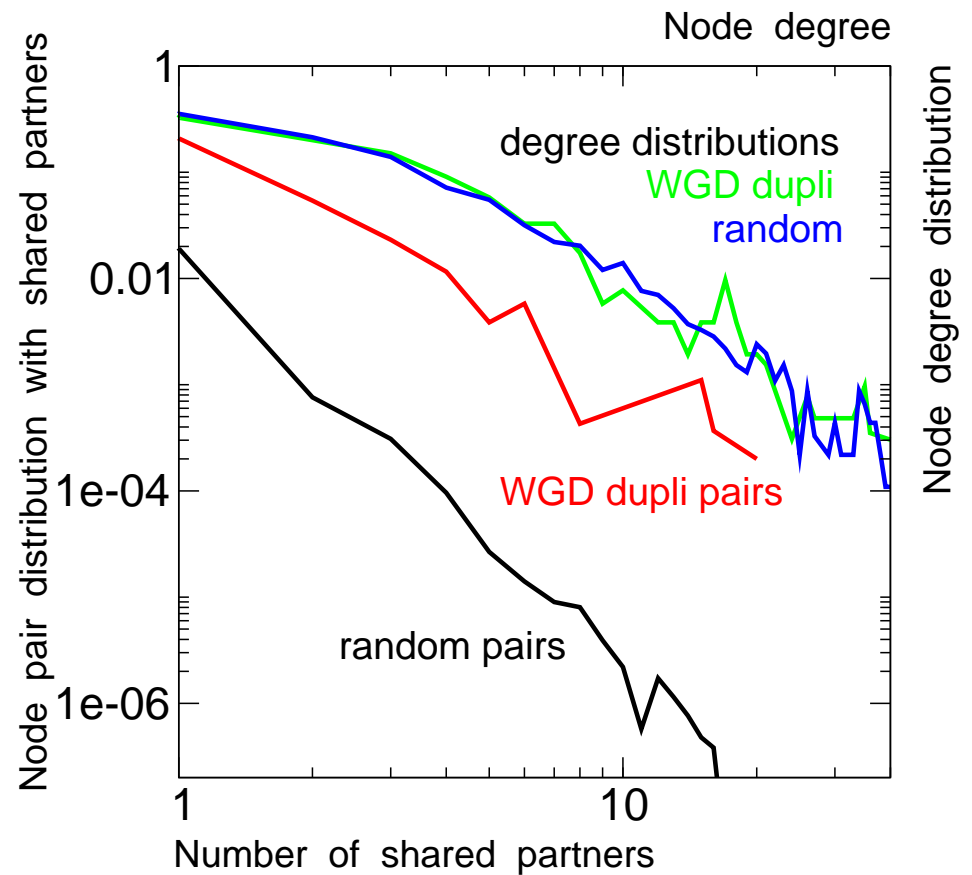
500 duplicates from WGD : 250 with both duplicates in available PPI network



Proba to share  $k+$  partners :

$k=1$  WGD dupli > 20 x random pairs

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

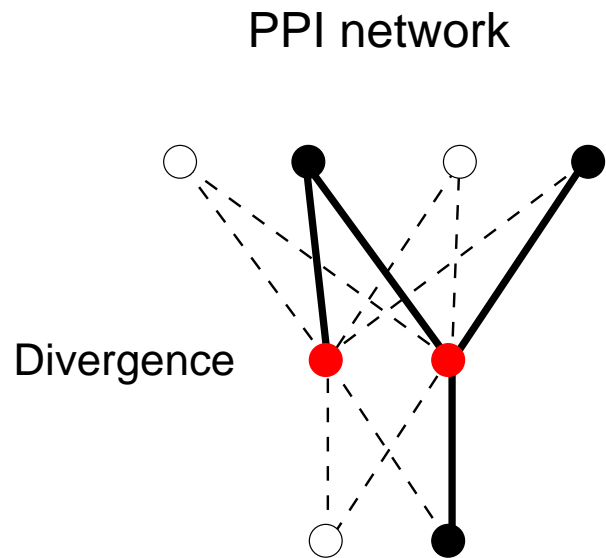




# Evidence of Genome Duplication on PPI network

Yeast 6000 genes : 4000 in PPI network

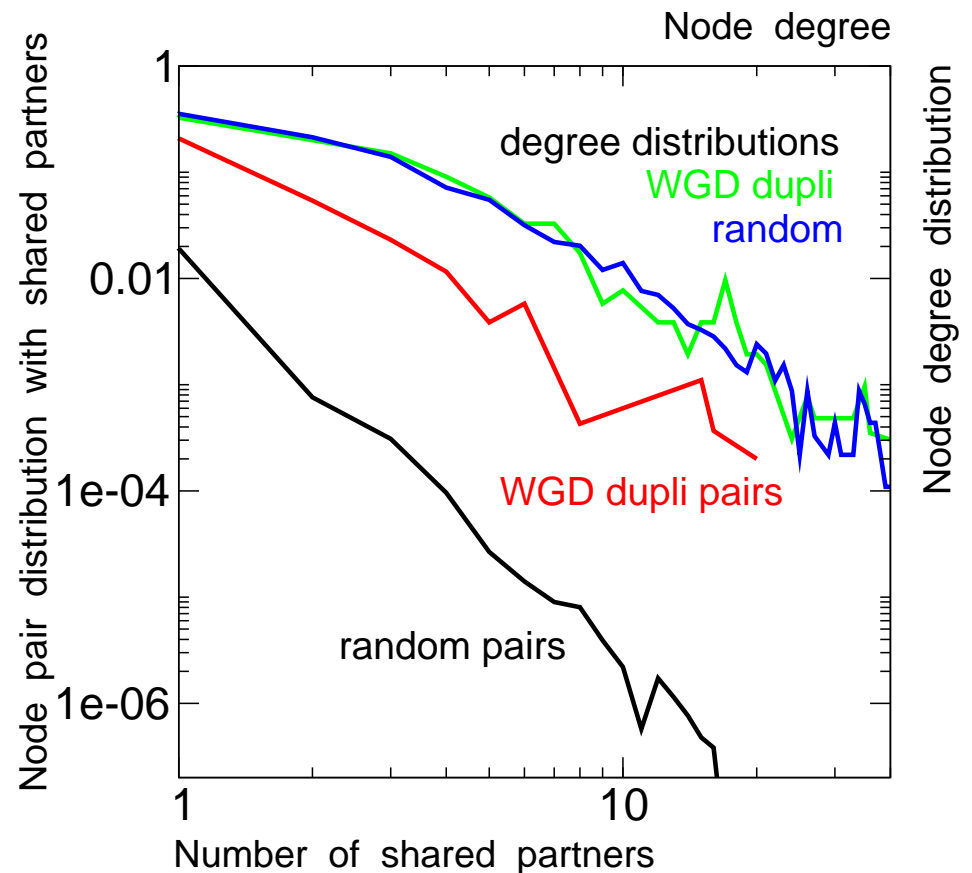
500 duplicates from WGD : 250 with both duplicates in available PPI network



Proba to share  $k+$  partners :

$k=1$  WGD dupli > 20 x random pairs

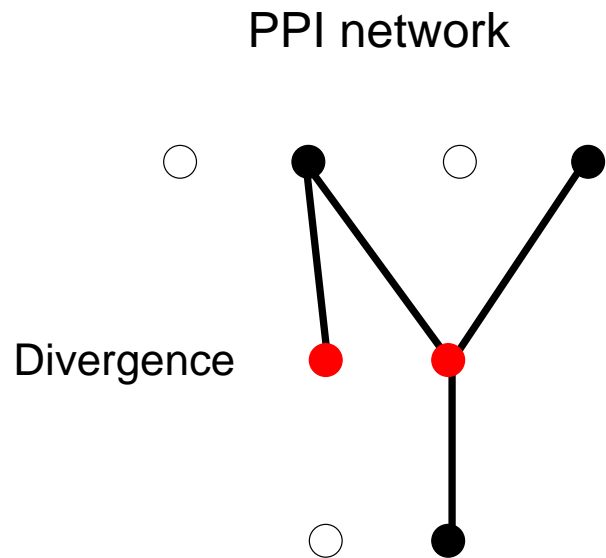
$k=10$  WGD dupli > 1,000 x random pairs



# Evidence of Genome Duplication on PPI network

Yeast 6000 genes : 4000 in PPI network

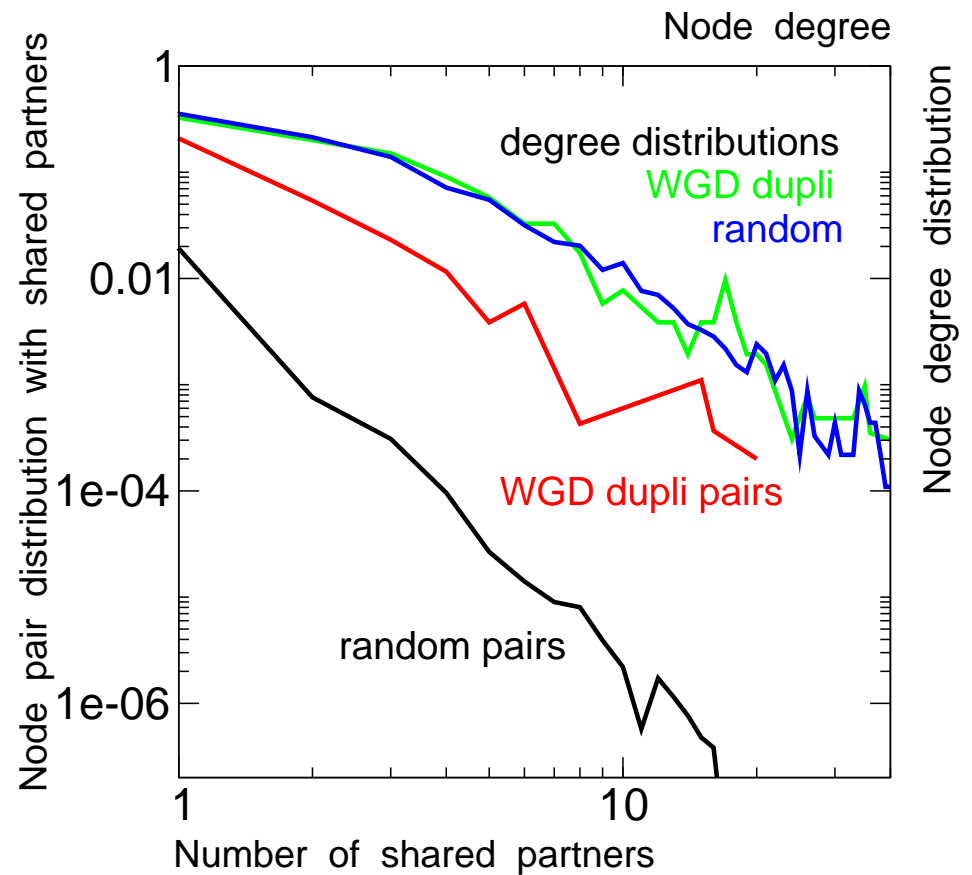
500 duplicates from WGD : 250 with both duplicates in available PPI network



Proba to share  $k+$  partners :

$k=1$  WGD dupli > 20 x random pairs

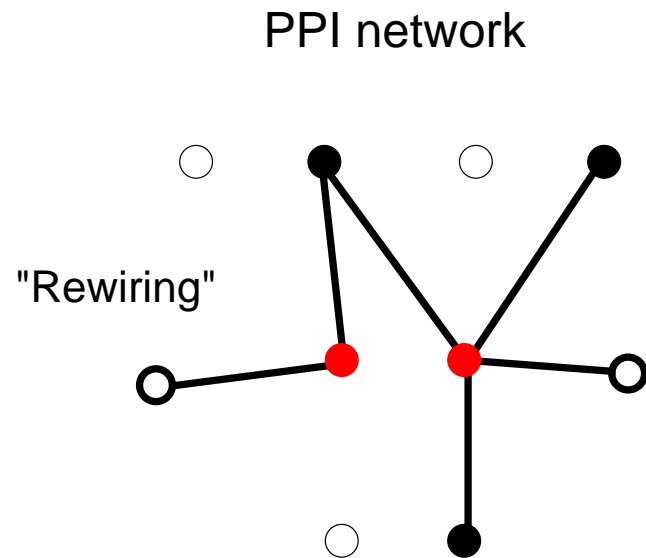
$k=10$  WGD dupli > 1,000 x random pairs



# Evidence of Genome Duplication on PPI network

Yeast 6000 genes : 4000 in PPI network

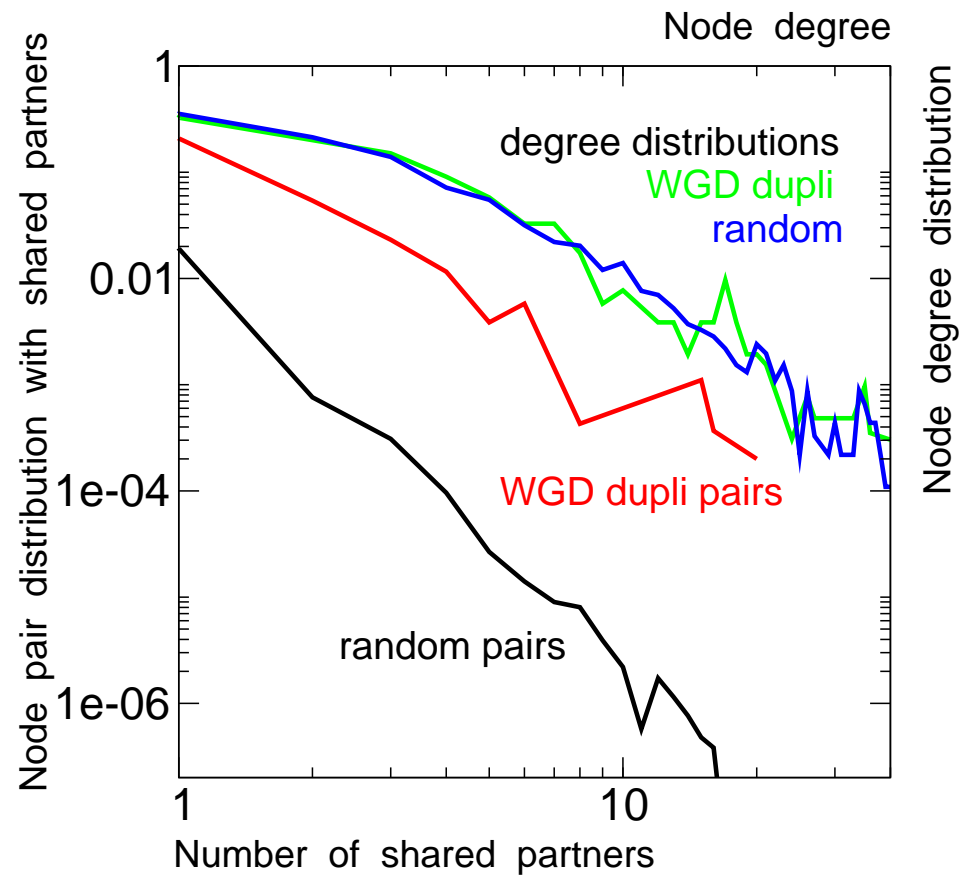
500 duplicates from WGD : 250 with both duplicates in available PPI network



Proba to share  $k+$  partners :

$k=1$  WGD dupli > 20 x random pairs

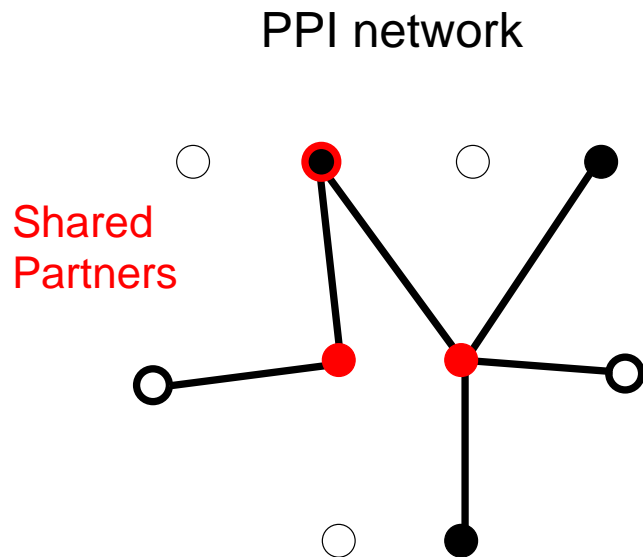
$k=10$  WGD dupli > 1,000 x random pairs



# Evidence of Genome Duplication on PPI network

Yeast 6000 genes : 4000 in PPI network

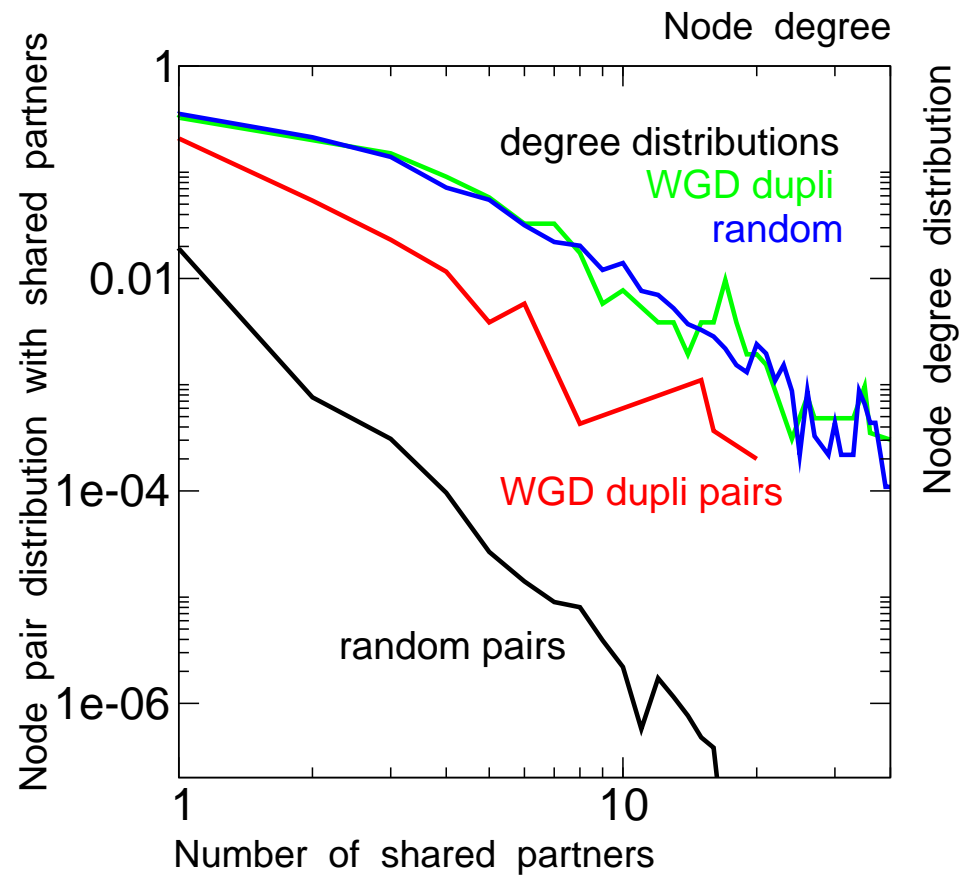
500 duplicates from WGD : 250 with both duplicates in available PPI network



Proba to share  $k+$  partners :

$k=1$  WGD dupli > 20 x random pairs

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



# 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 [q-bio.MN/0606036](#)

**global+local** → general duplication-divergence model

Evlampiev, Isambert [q-bio.MN/06011070](#)

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 q-bio.MN/0606036**

global+local → general duplication-divergence model

**Evlampiev, Isambert q-bio.MN/06011070**

exponential  
growth

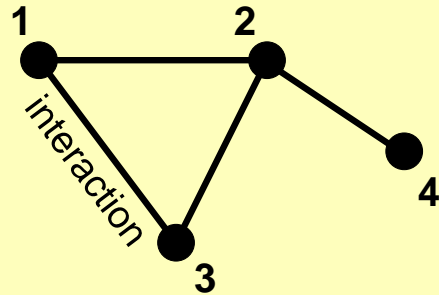
# General Duplication–Divergence Model of PPI Network Evolution

## GDD Model of PPI Network Evolution

Genome



P–P Interaction Network

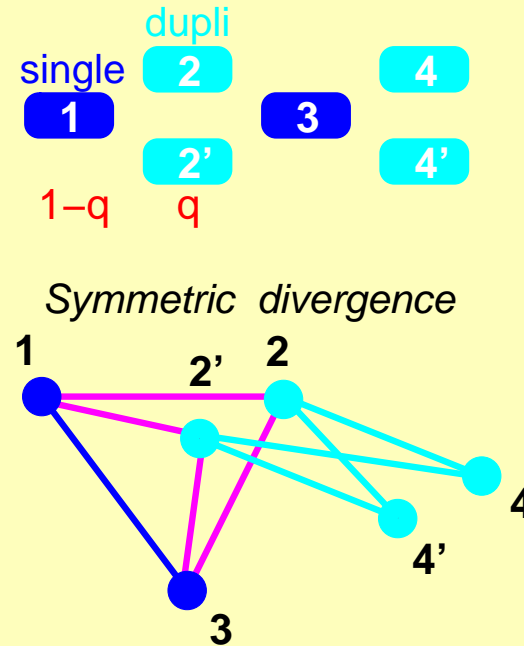


Partial Genome



Duplication

proteins  $\times (1+q)$



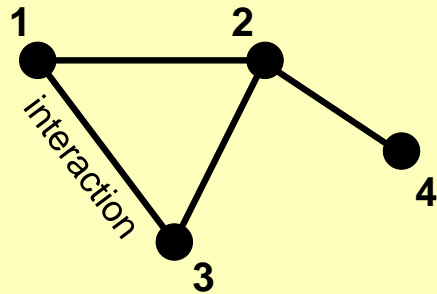
# General Duplication–Divergence Model of PPI Network Evolution

## GDD Model of PPI Network Evolution

Genome



P–P Interaction Network

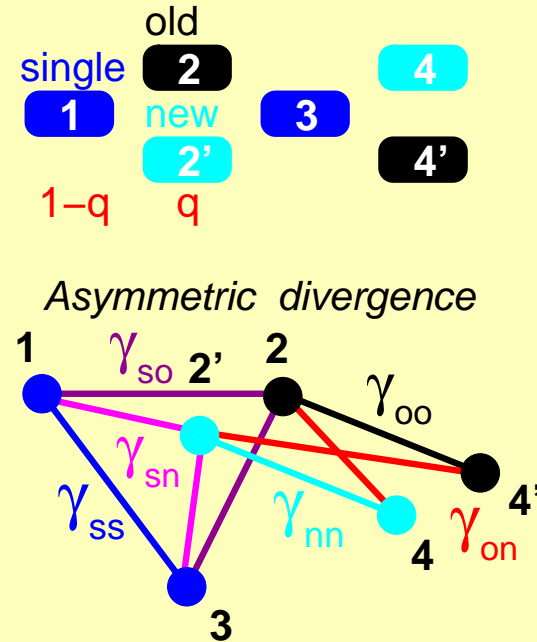


Partial Genome



Duplication

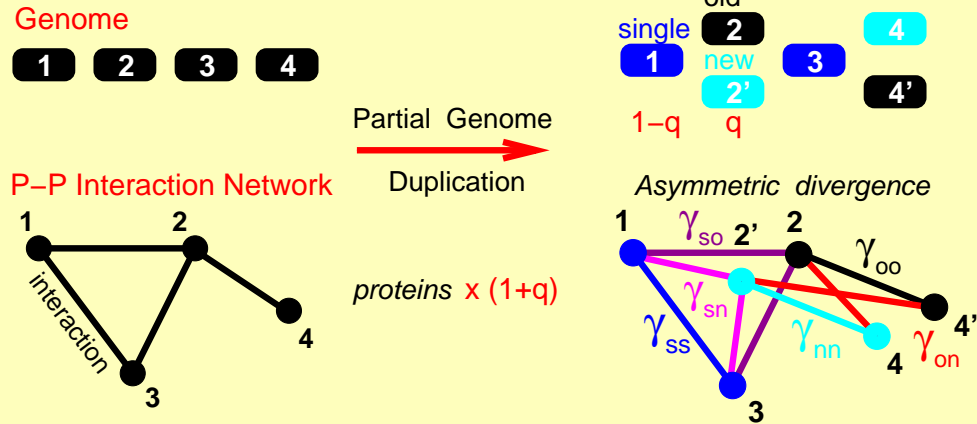
proteins  $\times (1+q)$





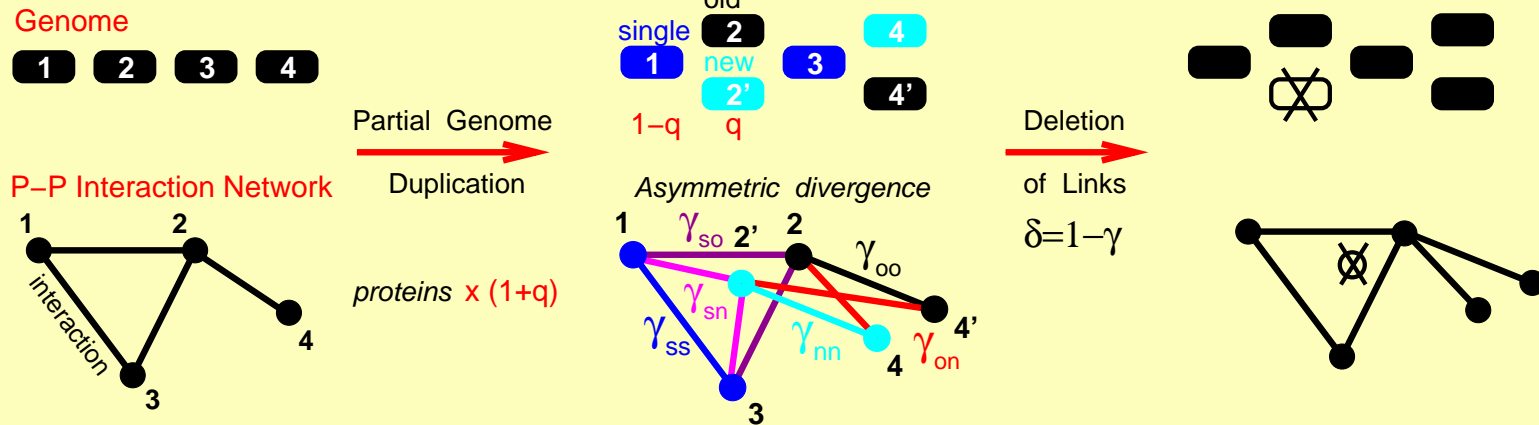
# General Duplication–Divergence Model of PPI Network Evolution

## GDD Model of PPI Network Evolution



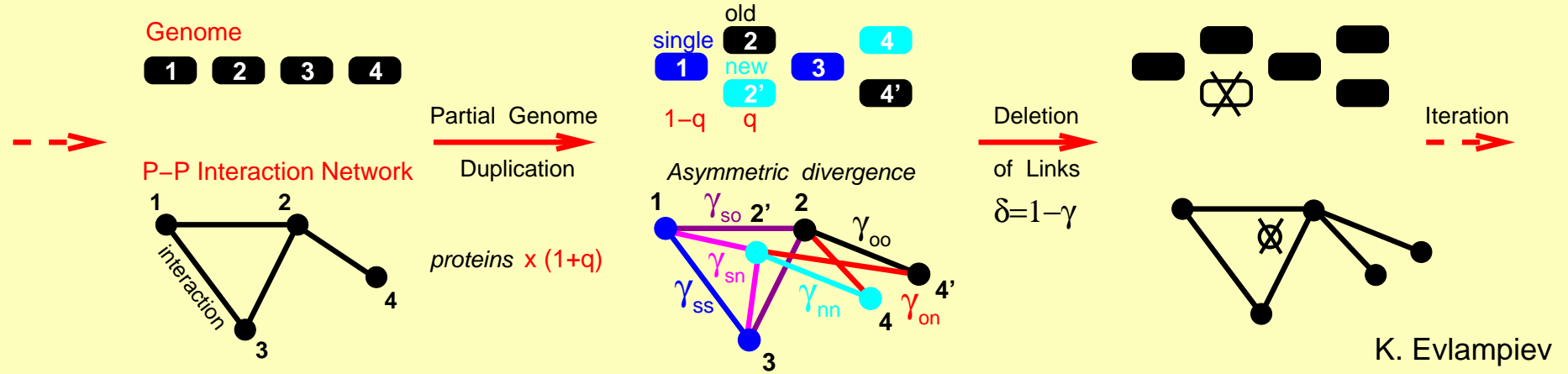
# General Duplication–Divergence Model of PPI Network Evolution

## GDD Model of PPI Network Evolution



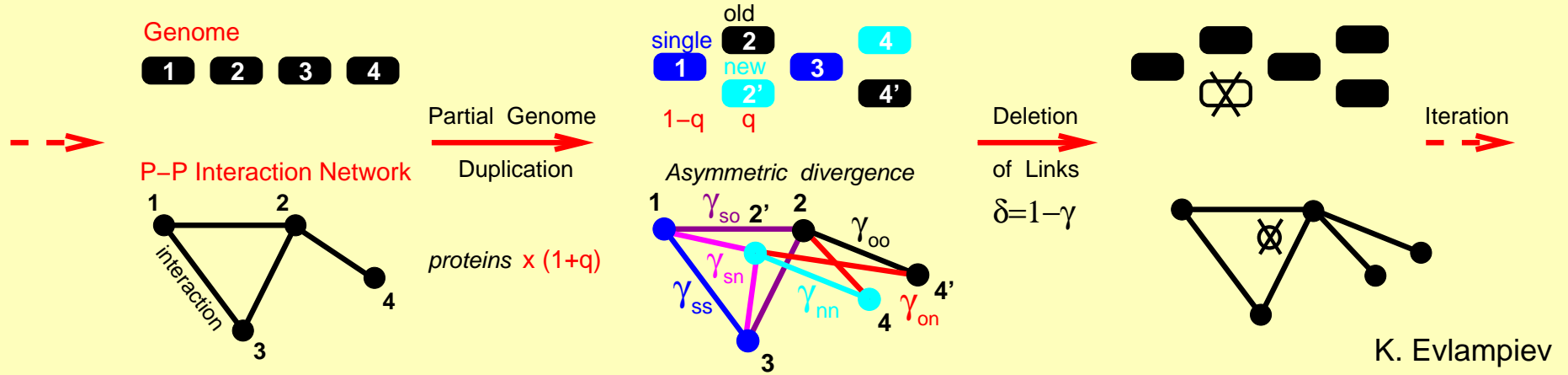
# General Duplication–Divergence Model of PPI Network Evolution

## GDD Model of PPI Network Evolution



# General Duplication–Divergence Model of PPI Network Evolution

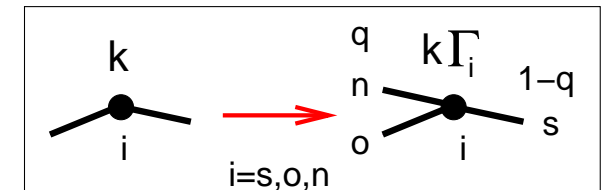
GDD Model of PPI Network Evolution



Exponential Dynamics of Genome and PPI Network Evolution

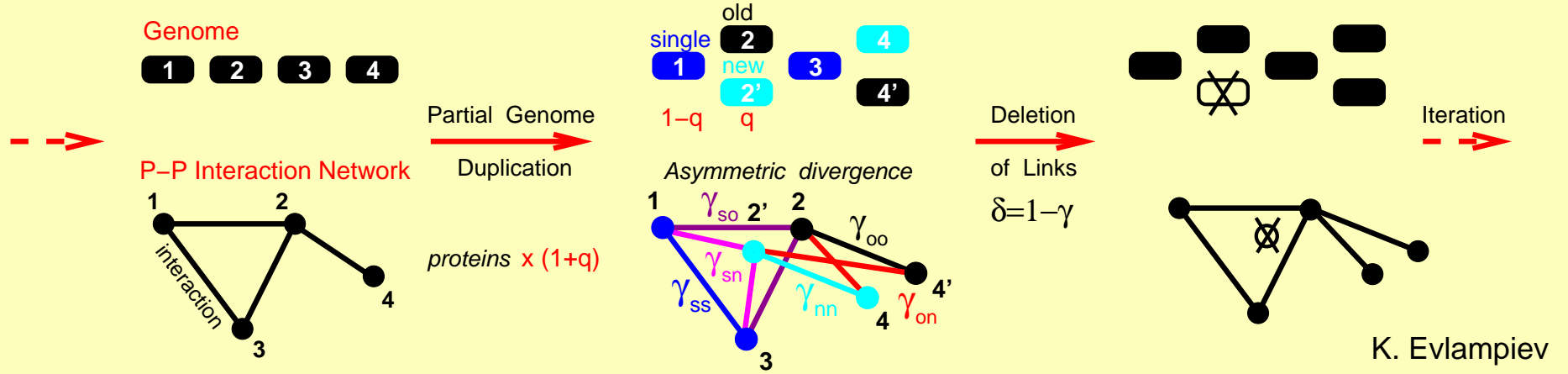
Node Degree Growth Rate

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



# General Duplication–Divergence Model of PPI Network Evolution

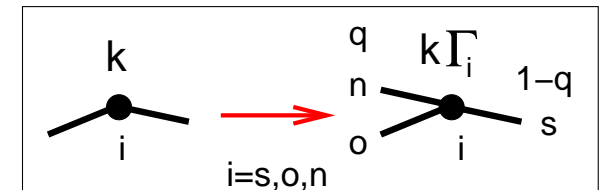
## GDD Model of PPI Network Evolution



## Exponential Dynamics of Genome and PPI Network Evolution

Node Degree Growth Rate

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



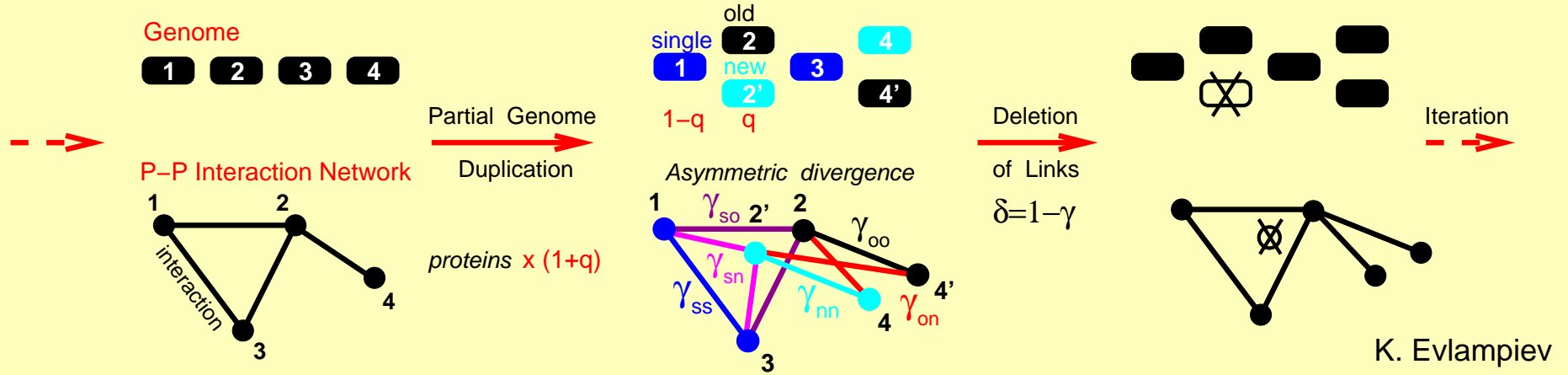
Network Link Growth Rate

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

$\Gamma > 1$  Growing Network  
 $\Gamma < 1$  Vanishing Network

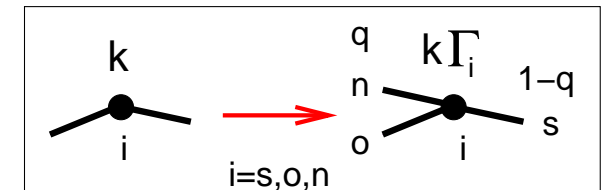
# General Duplication–Divergence Model of PPI Network Evolution

## GDD Model of PPI Network Evolution



## Exponential Dynamics of Genome and PPI Network Evolution

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



Network Link Growth Rate  $\Gamma = (1-q)\Gamma_s + q\Gamma_o + q\Gamma_n$

$\Gamma > 1$  Growing Network

$\Gamma < 1$  Vanishing Network

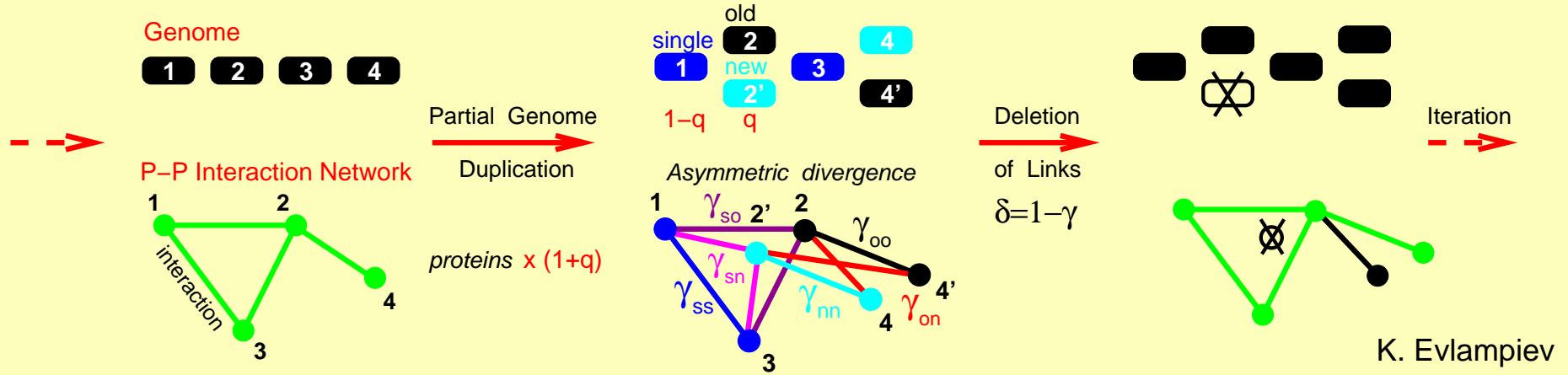
Network Conservation Index  $M = (1-q)\Gamma_s + q\Gamma_o$

$M > 1$  Conserved Network

$M < 1$  Nonconserved Network

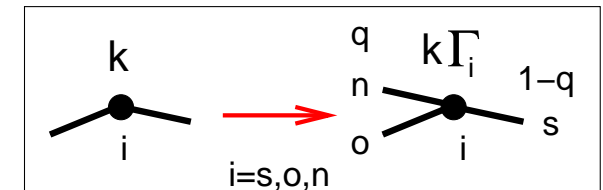
# General Duplication–Divergence Model of PPI Network Evolution

## GDD Model of PPI Network Evolution



## Exponential Dynamics of Genome and PPI Network Evolution

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



Network Link Growth Rate  $\Gamma = (1-q)\Gamma_s + q\Gamma_o + q\Gamma_n$

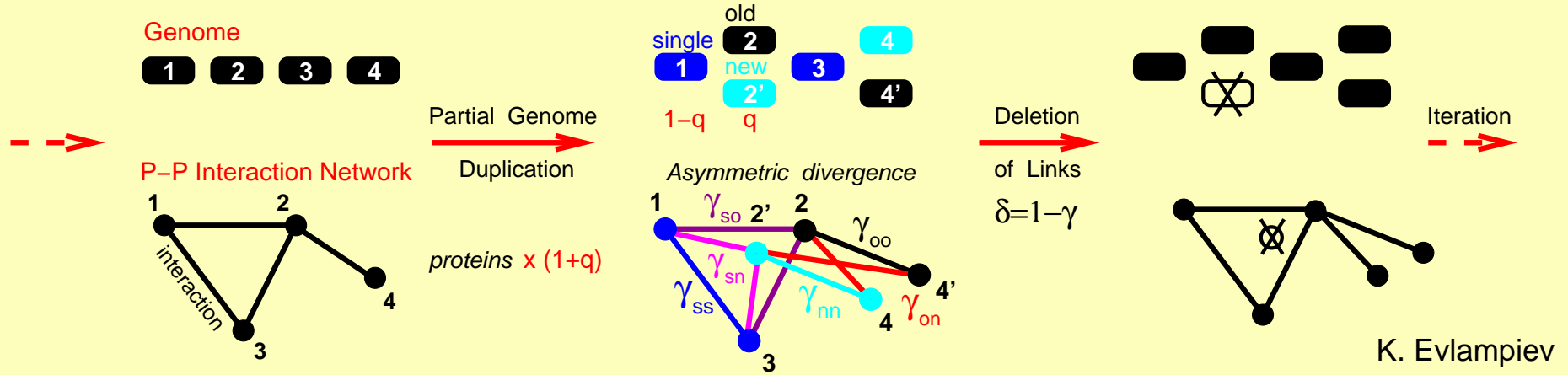
$\Gamma > 1$  Growing Network  
 $\Gamma < 1$  Vanishing Network

Network Conservation Index  $M = (1-q)\Gamma_s + q\Gamma_o$

$M > 1$  Conserved Network  
 $M < 1$  Nonconserved Network

# General Duplication–Divergence Model of PPI Network Evolution

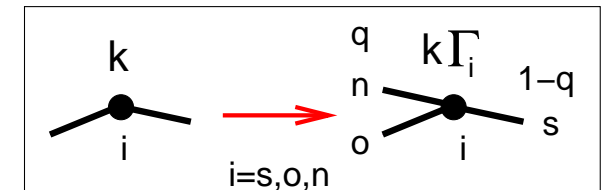
## GDD Model of PPI Network Evolution



## Exponential Dynamics of Genome and PPI Network Evolution

Node Degree Growth Rate

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



Network Link Growth Rate

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

$\Gamma > 1$  Growing Network  
 $\Gamma < 1$  Vanishing Network

Network Conservation Index

$$M = (1-q)\Gamma_s + q\Gamma_o$$

$M > 1$  Conserved Network  
 $M < 1$  Nonconserved Network

GDD Model + Fluctuations

$$M = \prod_{n=1}^R [(1-q^{(n)})\Gamma_s^{(n)} + q^{(n)}\Gamma_o^{(n)}]$$

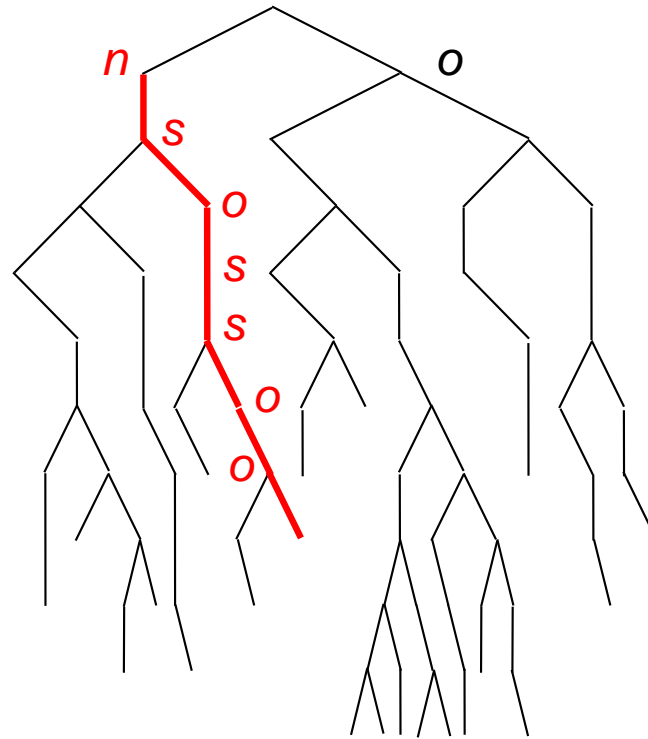


# Network Conservation ( $M$ ) and Expansion ( $\Gamma$ )

$M < \Gamma < 1$



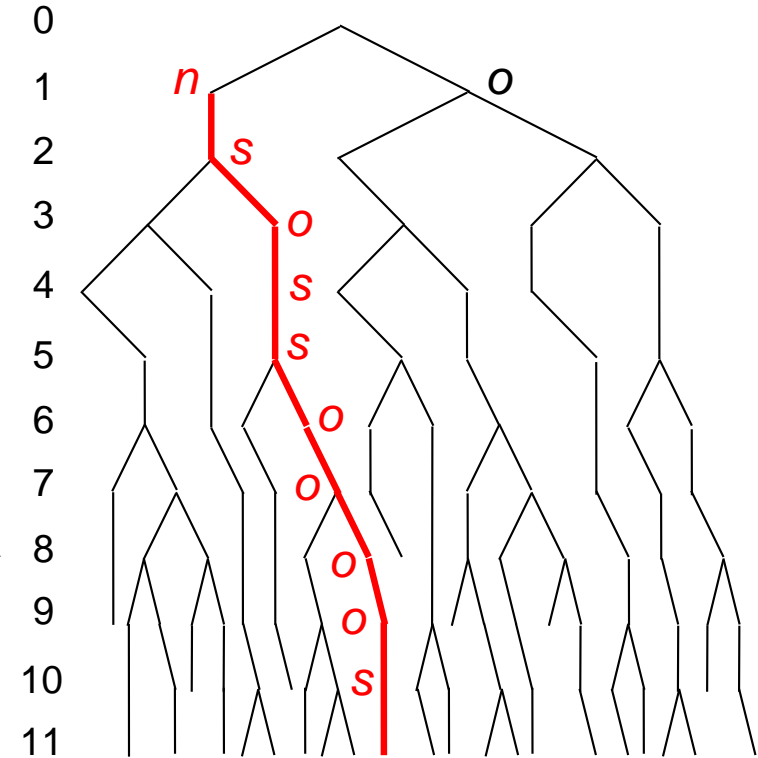
$M < 1 < \Gamma$



*Non-Conserved Networks*

duplication

$1 < M < \Gamma$



0  
1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11

*Conserved Networks*

$M$

*Vanishing*

1

*Expanding PPI Networks*

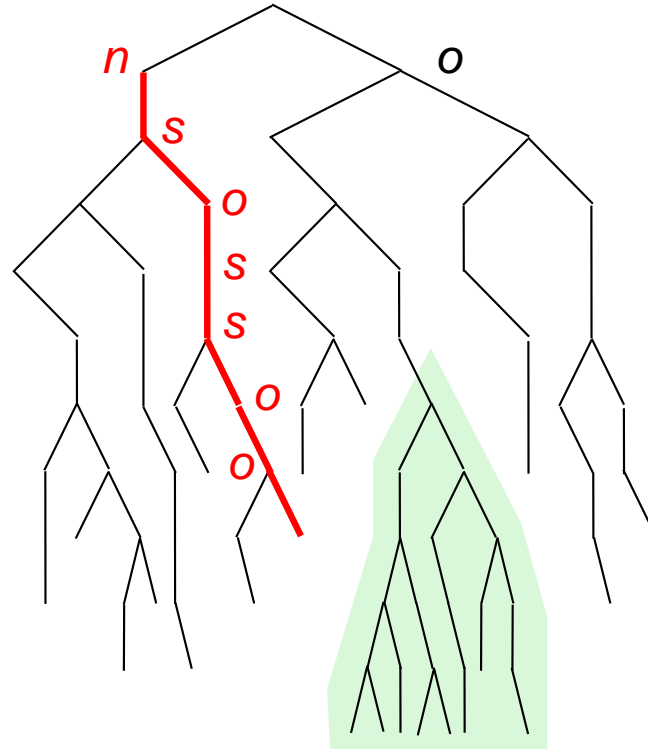
$\Gamma$

# Network Conservation ( $M$ ) and Expansion ( $\Gamma$ )

$M < \Gamma < 1$



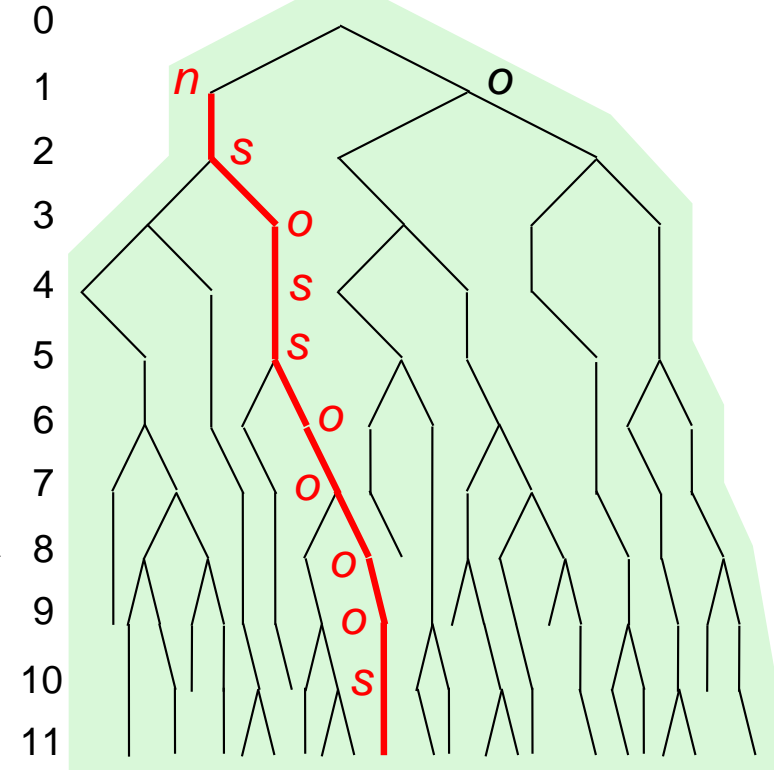
$M < 1 < \Gamma$



*Non-Conserved Networks*

duplication

$1 < M < \Gamma$



*Conserved Networks*

*Vanishing*

1

*Expanding PPI Networks*

$\Gamma$

$M$

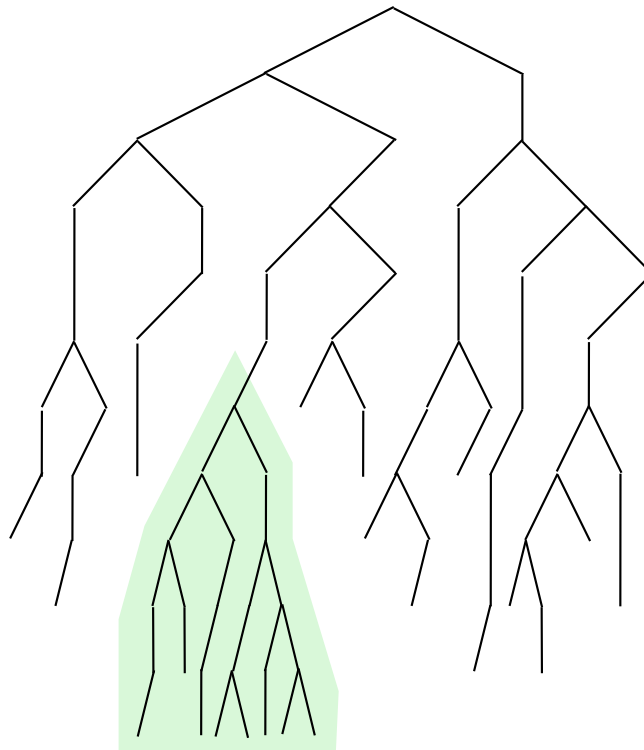
1

# Network Conservation (M) and Expansion ( $\Gamma$ )

$M < \Gamma < 1$



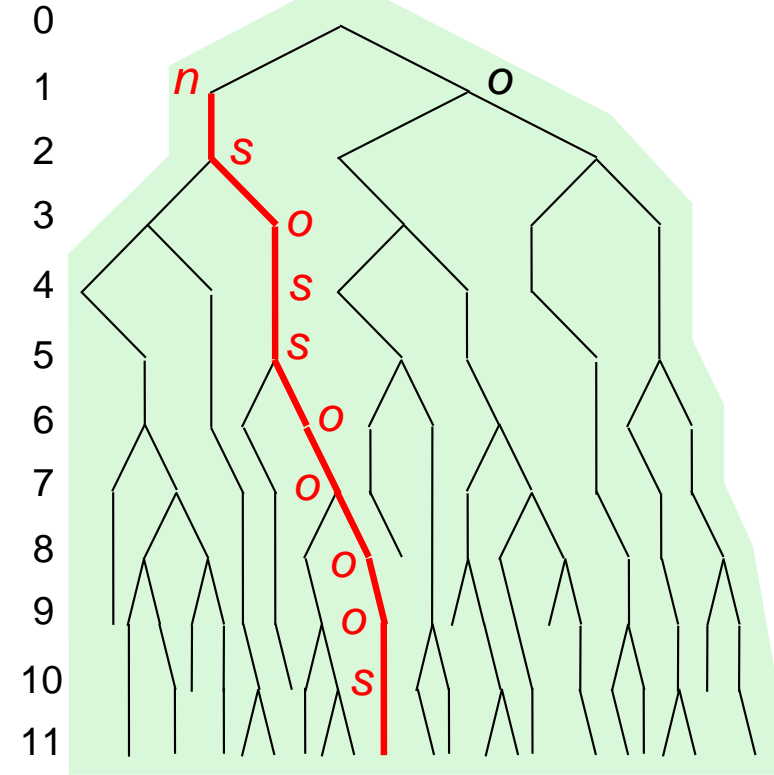
$M < 1 < \Gamma$



*Non-Conserved Networks*

duplication

$1 < M < \Gamma$



*Conserved Networks*

*Vanishing*

1

*Expanding PPI Networks*

$\Gamma$

1

M

0  
1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11

n

s

o

s

s

s

o

o

o

o

s

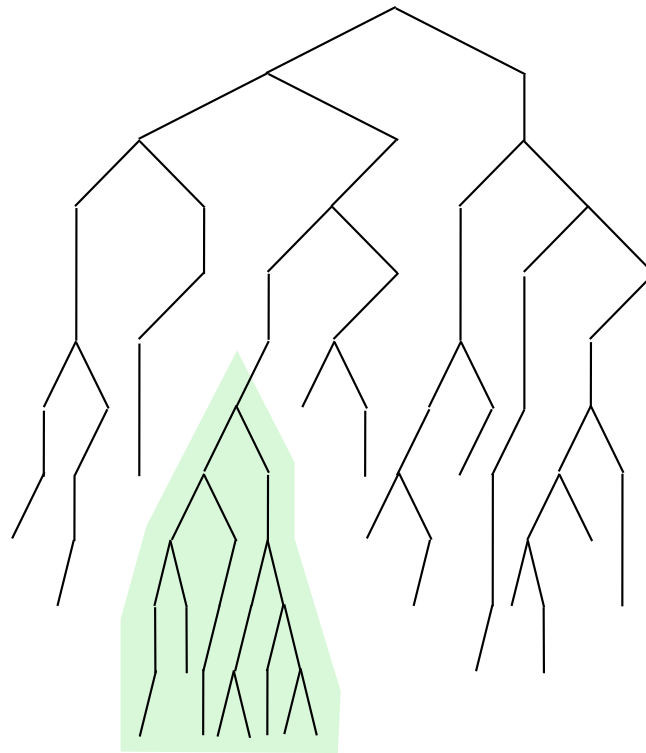
o

# Network Conservation (M) and Expansion ( $\Gamma$ )

$M < \Gamma < 1$



$M < 1 < \Gamma$

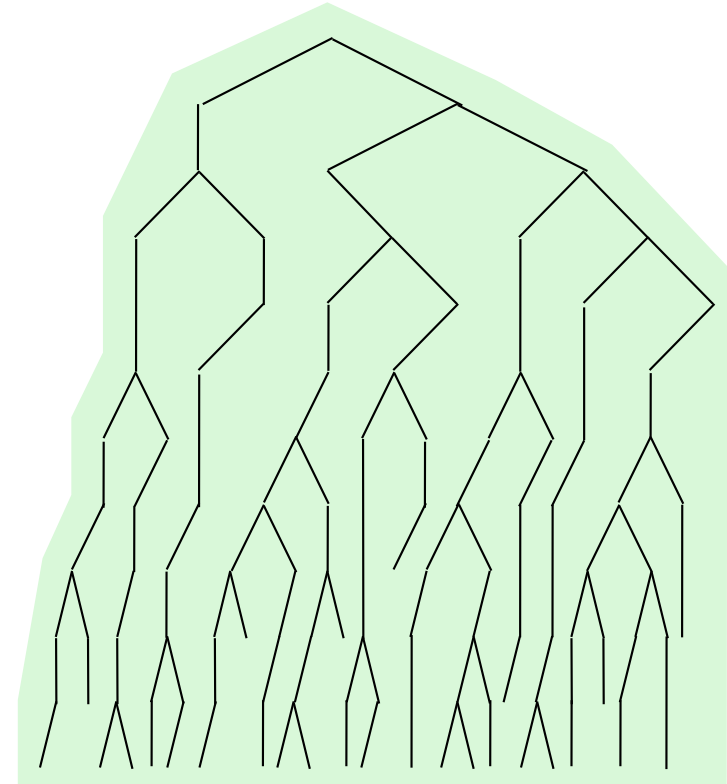


*Non-Conserved Networks*

duplication

0  
1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11

$1 < M < \Gamma$



*Conserved Networks*

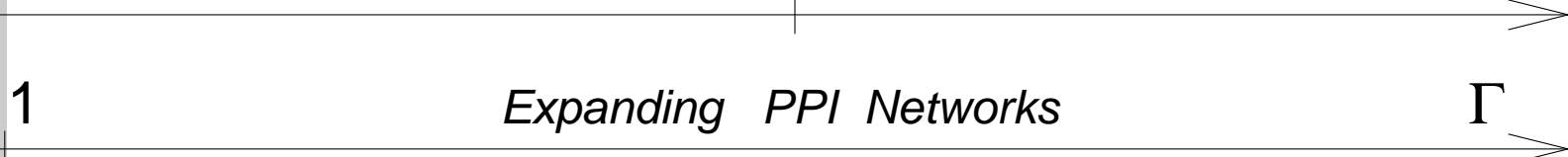
M

*Vanishing*

1

*Expanding PPI Networks*

$\Gamma$



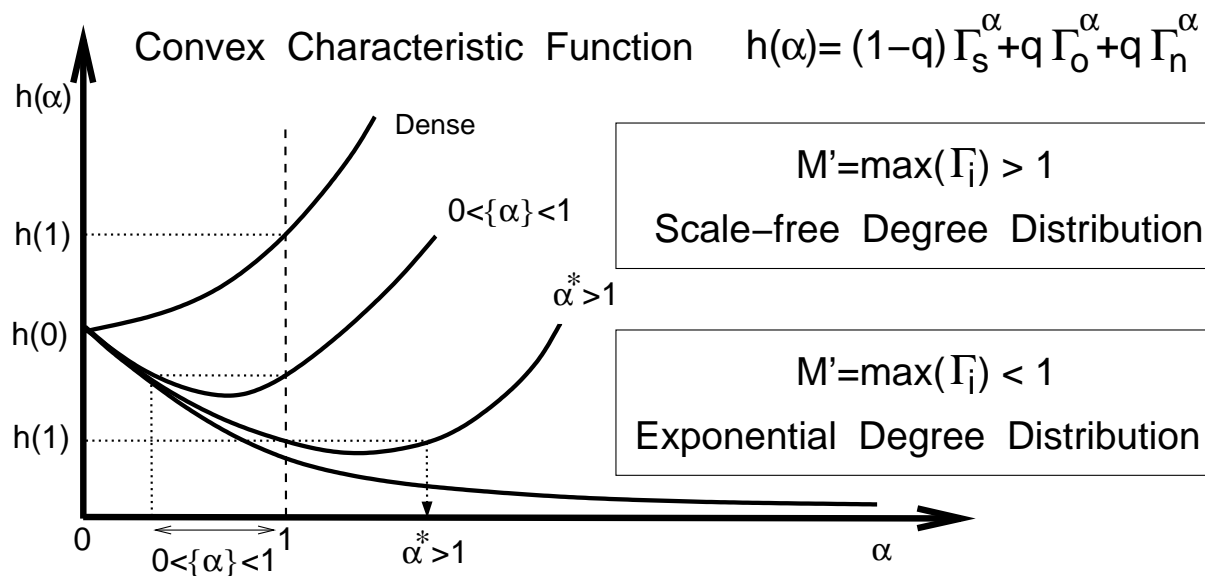
# Asymptotic Topology of PPI Networks

Network Node Growth Rate

$$\langle N^{(n+1)} \rangle / \langle N^{(n)} \rangle = \Delta^{(n)}$$

Asymptotic "Characteristic Equation"

$$\Delta = h(\alpha) = (1-q)\Gamma_S^\alpha + q\Gamma_O^\alpha + q\Gamma_n^\alpha$$



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

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

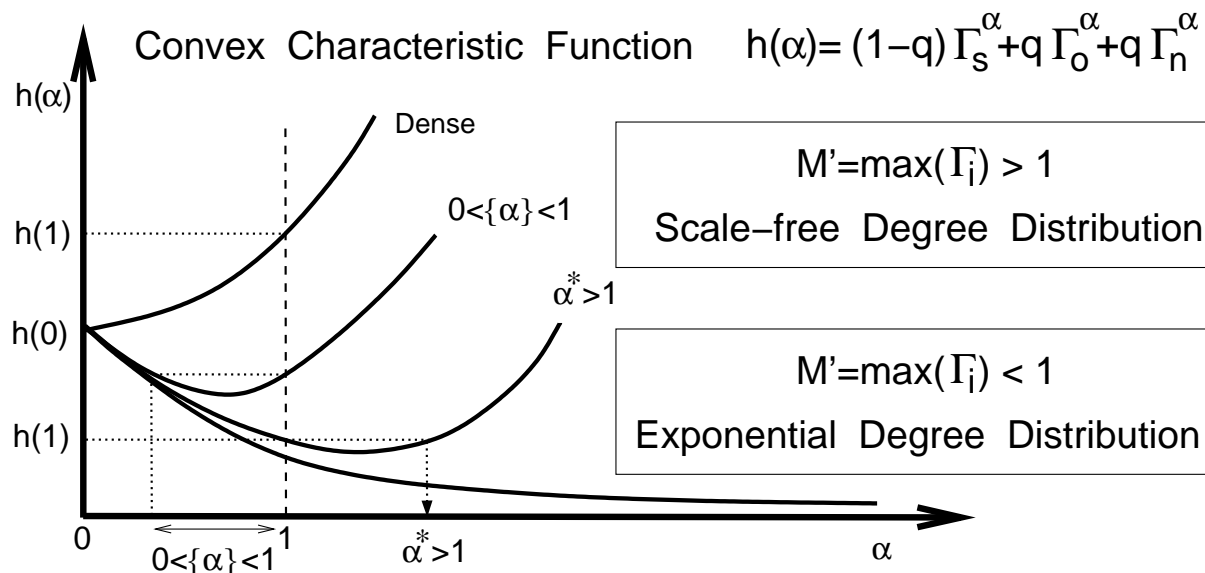
# Asymptotic Topology of PPI Networks

Network Node Growth Rate

$$\langle N^{(n+1)} \rangle / \langle N^{(n)} \rangle = \Delta^{(n)}$$

Asymptotic "Characteristic Equation"

$$\Delta = h(\alpha) = (1-q)\Gamma_s^\alpha + q\Gamma_o^\alpha + q\Gamma_n^\alpha$$



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

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

→ Asymptotic Network Topology

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

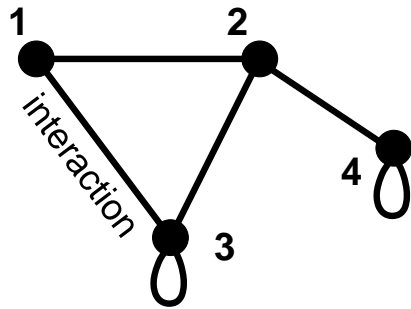
$M' > 1$  Scale-free **3**  
 $M' < 1$  Exponential

GDD Model + Fluctuations  $M' = \prod_{n=1}^R \max(\Gamma_i^{(n)}) \geq \prod_{n=1}^R [(1-q^{(n)})\Gamma_s^{(n)} + q^{(n)}\Gamma_o^{(n)}] = M$

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

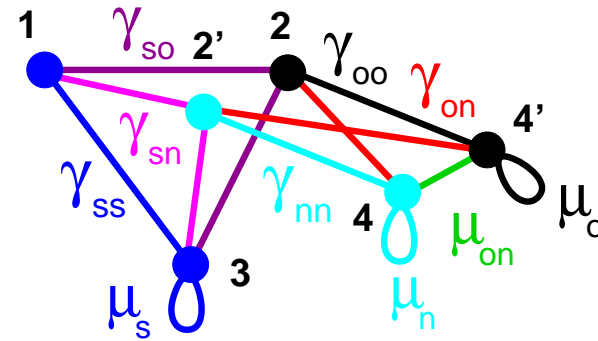
# Duplication–Divergence Models Including Self–links

P–P Interaction Network



Partial Genome  
 Duplication  
 proteins  $\times (1+q)$

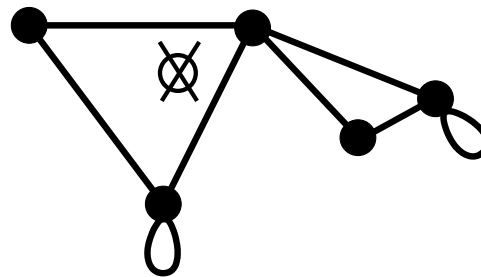
Asymmetric divergence



Iteration

Deletion  
 of Links  
 with proba

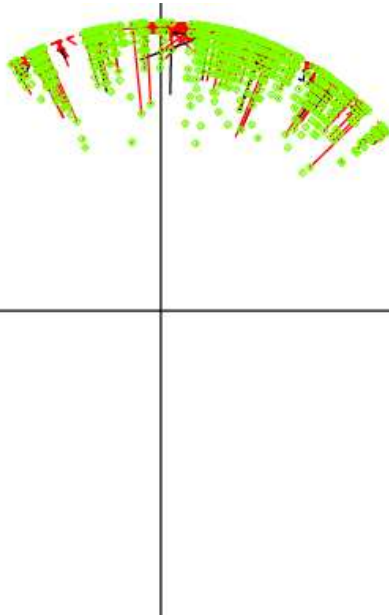
$1-\gamma$   
 $1-\mu$



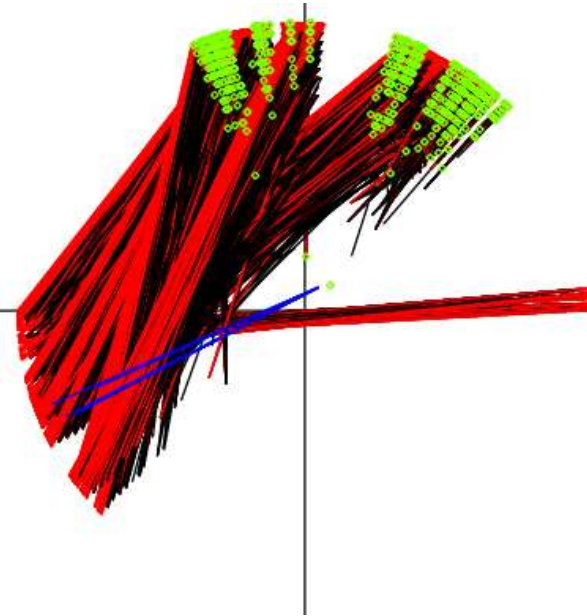
→ Self–links do not affect evolutionary regimes (just exponents)

# Network Conservation (M) and Topology (M')

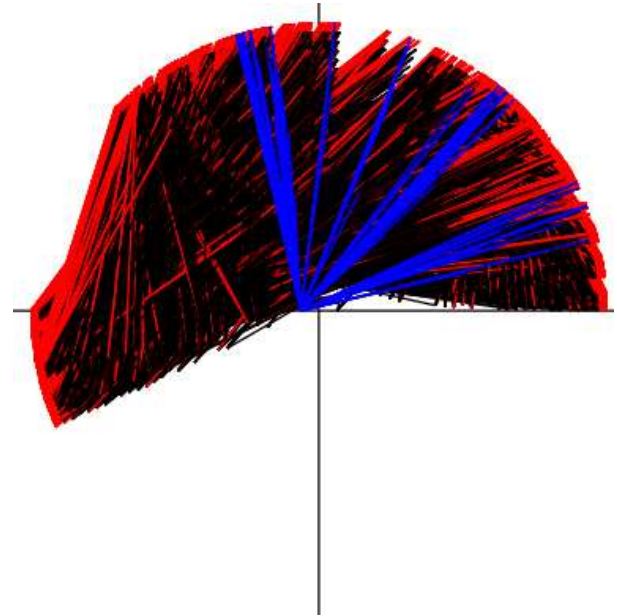
$$M < M' < 1$$



$$M < 1 < M'$$



$$1 < M < M'$$



*Non-Conserved*

1

*Conserved*

M

*Exponential*

1

*Scale-Free*

M'



# Generating Function Formalism

Kirill Evlampiev

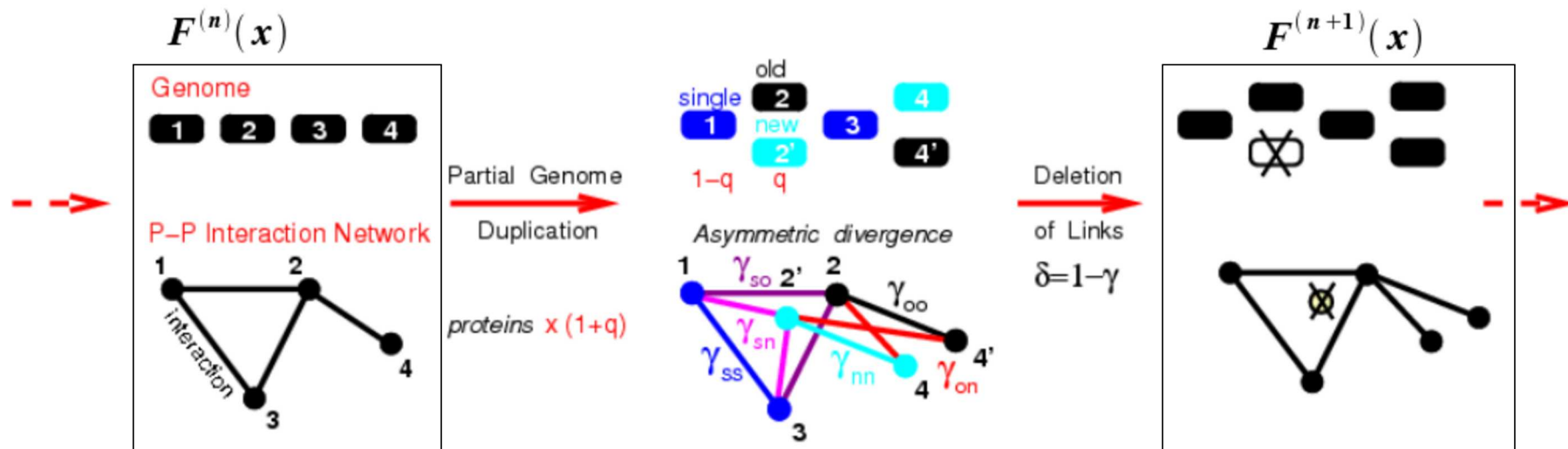
- At each step **generating function** of number of nodes

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

- For two consecutive steps

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

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



# 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  $k=0$ )

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

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

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

It can be analysed using asymptotic methods

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

## Statistical Properties of the Model

- Network evolution is described in terms of ensemble averages  $\langle Q \rangle$
- This description is meaningful since fluctuations are not large

distribution function of  $L$

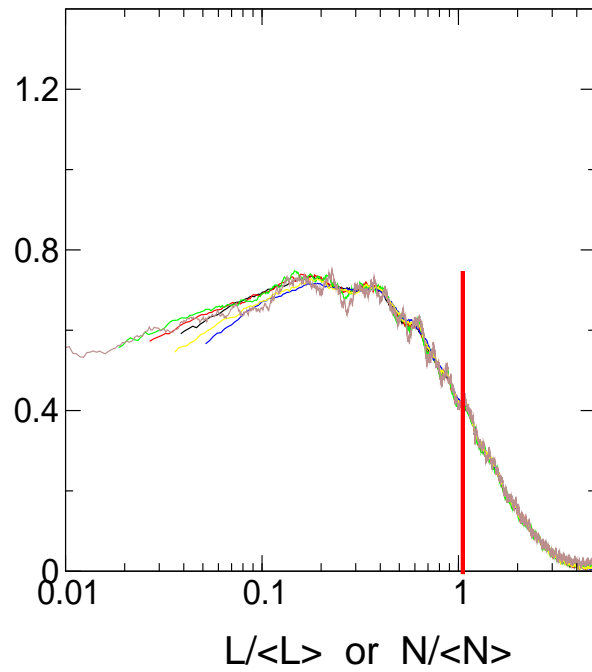
$$P^{(n)}(L) \sim \frac{1}{\langle L^{(n)} \rangle} F \left[ \frac{L}{\langle L^{(n)} \rangle} \right]$$

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

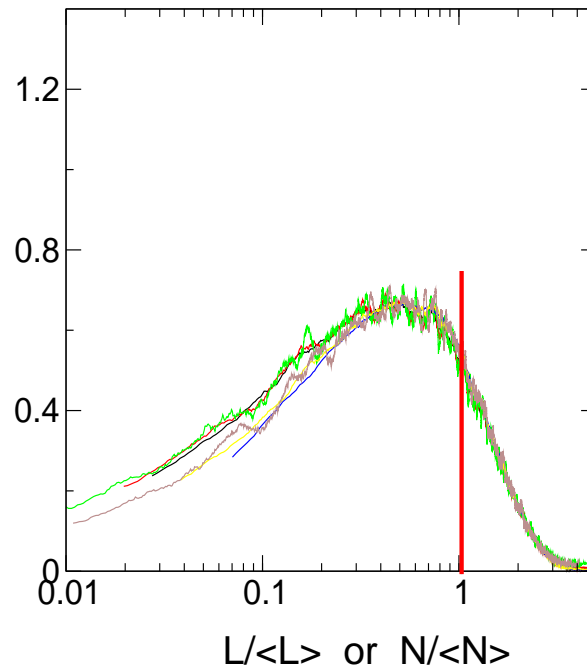
- Due to  $N_k \leq N \leq 2L$  in linear regimes  $\langle N \rangle \sim \langle L \rangle$  **Exp & Scale Free**  
this holds also for distributions of  $N$  and  $N_k$
- For nonlinear regime this probably holds too (numerical simulations)

# Distribution of Network Sizes

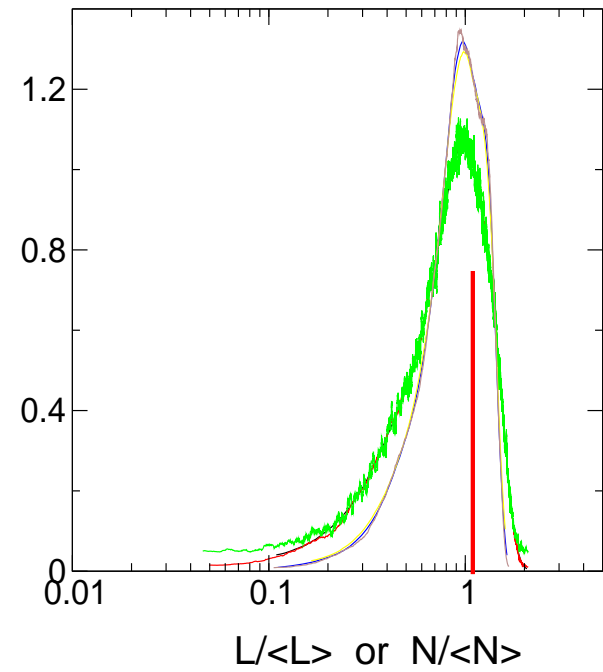
20% Growth Rate



50% Growth Rate



~80% Growth Rate



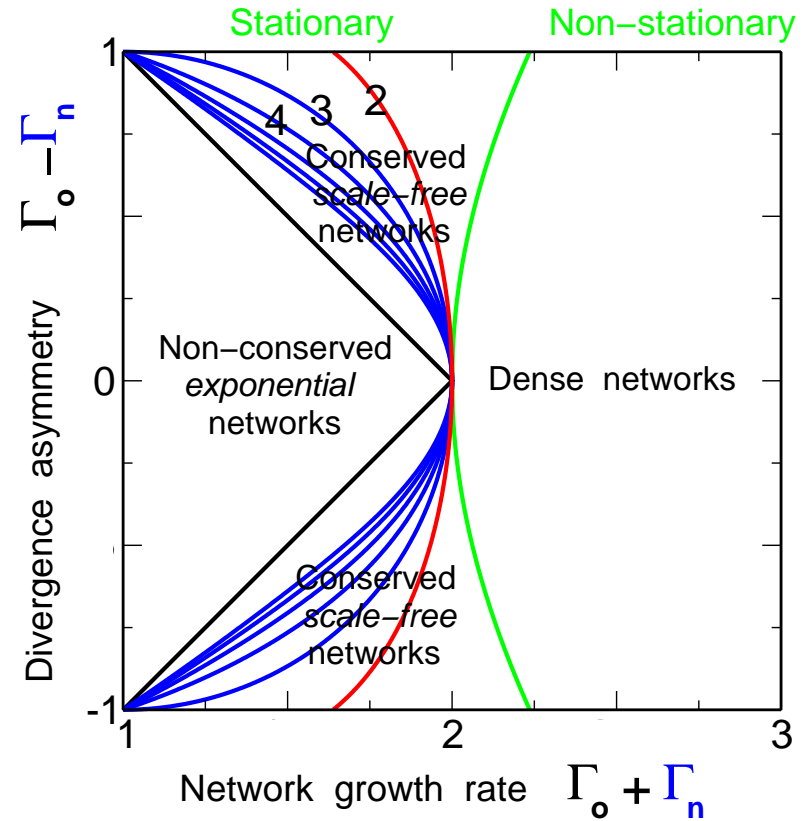
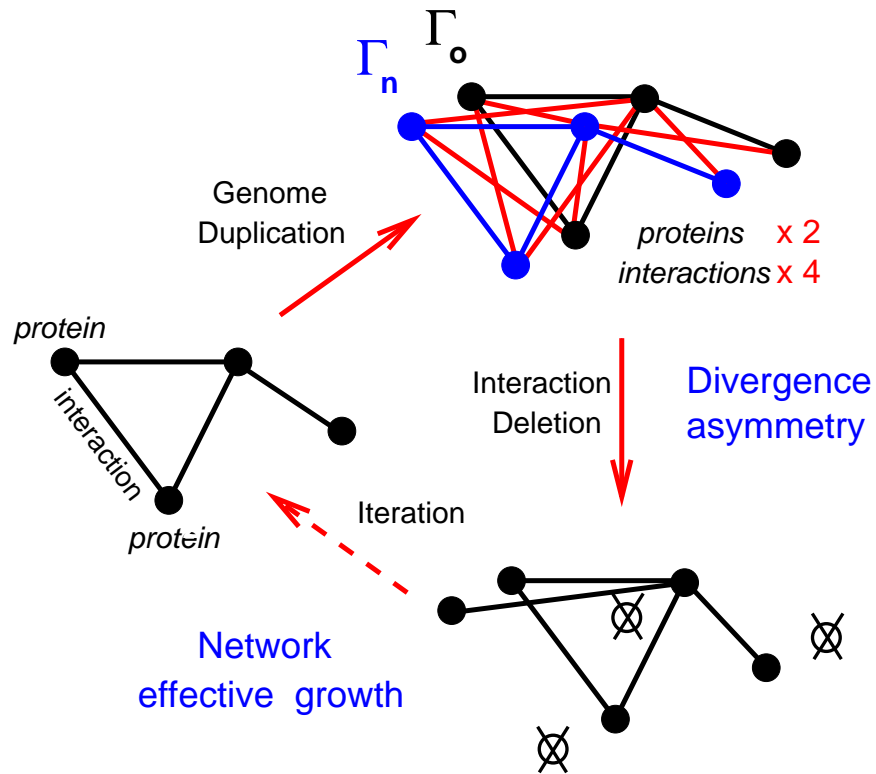
Linear regime  $N \sim L$

NL regime  $N < L$



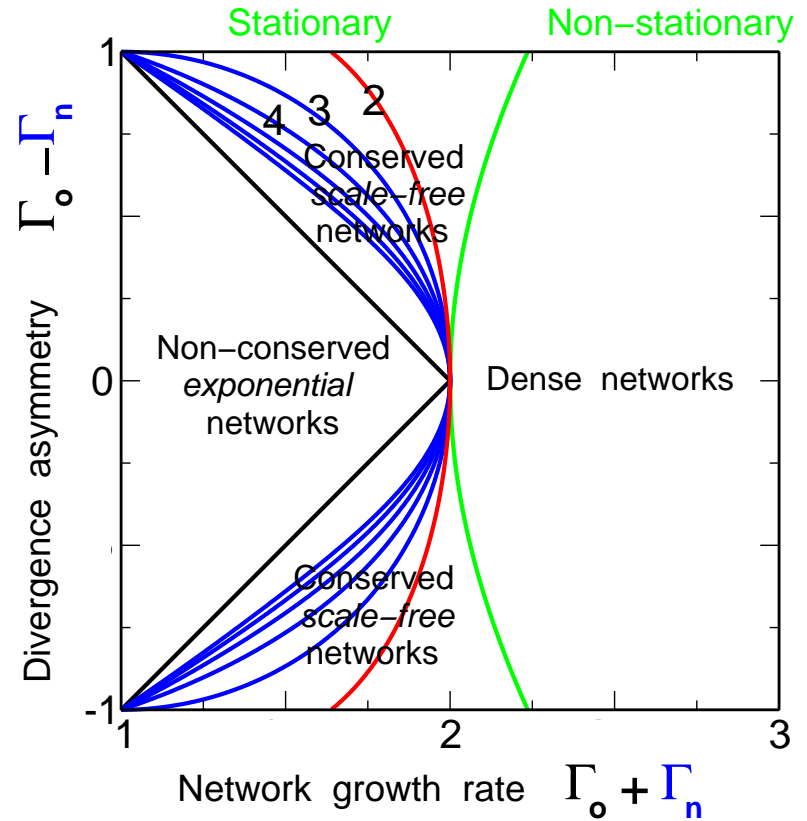
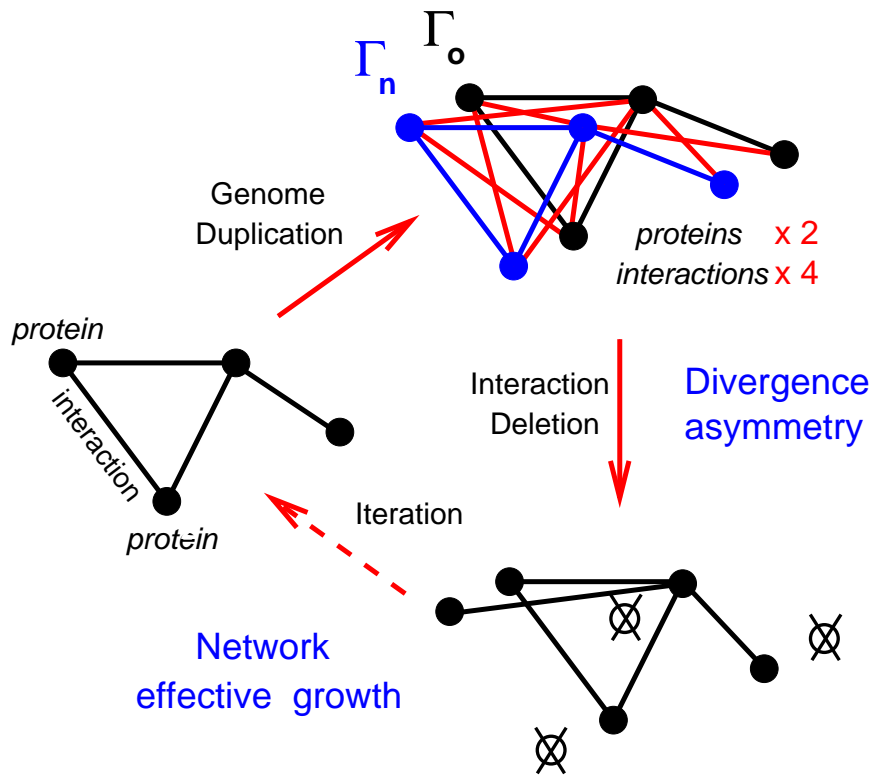
# Asymmetric Divergence of Duplicates is Required

Whole Genome Duplication Limit ( $q=1$ )



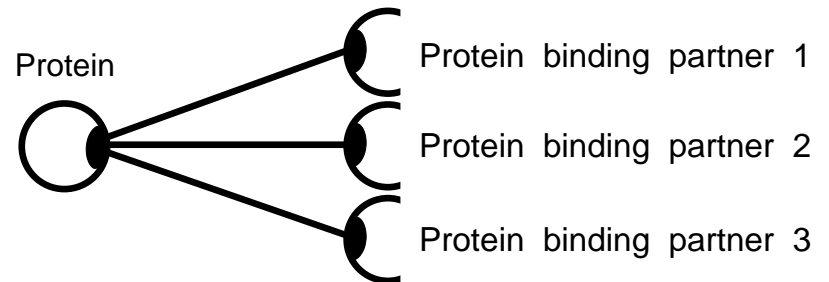
# Asymmetric Divergence of Duplicates is Required

Whole Genome Duplication Limit ( $q=1$ )



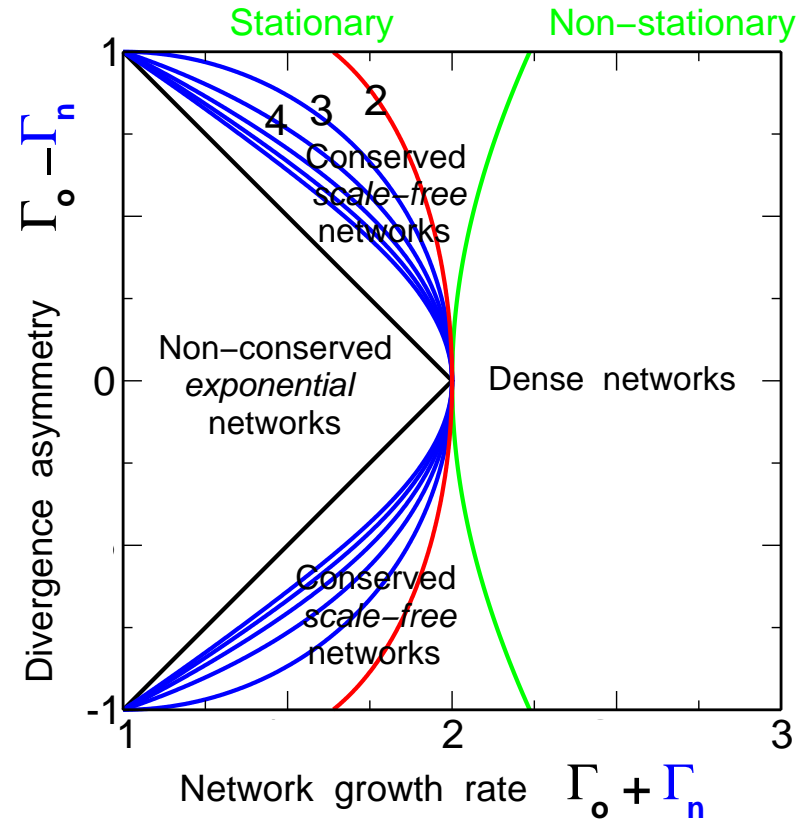
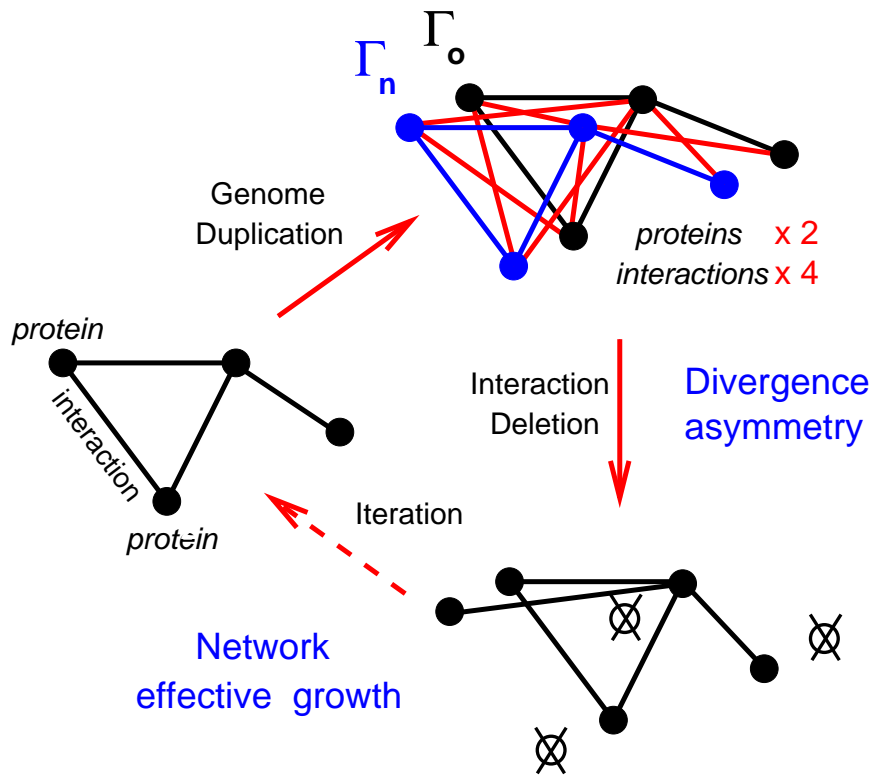
Origin of Divergence Asymmetry ?

*"Spontaneous Symmetry Breaking"*  
between duplicated binding sites



# Asymmetric Divergence of Duplicates is Required

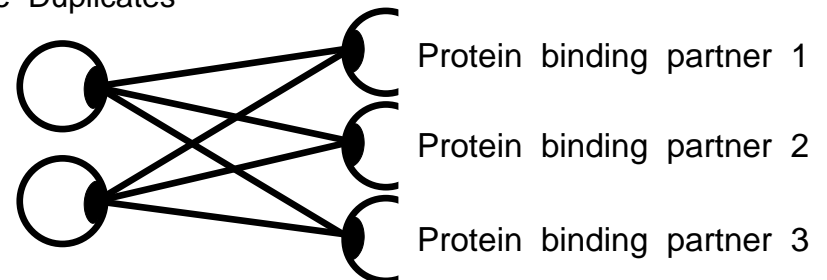
Whole Genome Duplication Limit ( $q=1$ )



Origin of Divergence Asymmetry ?

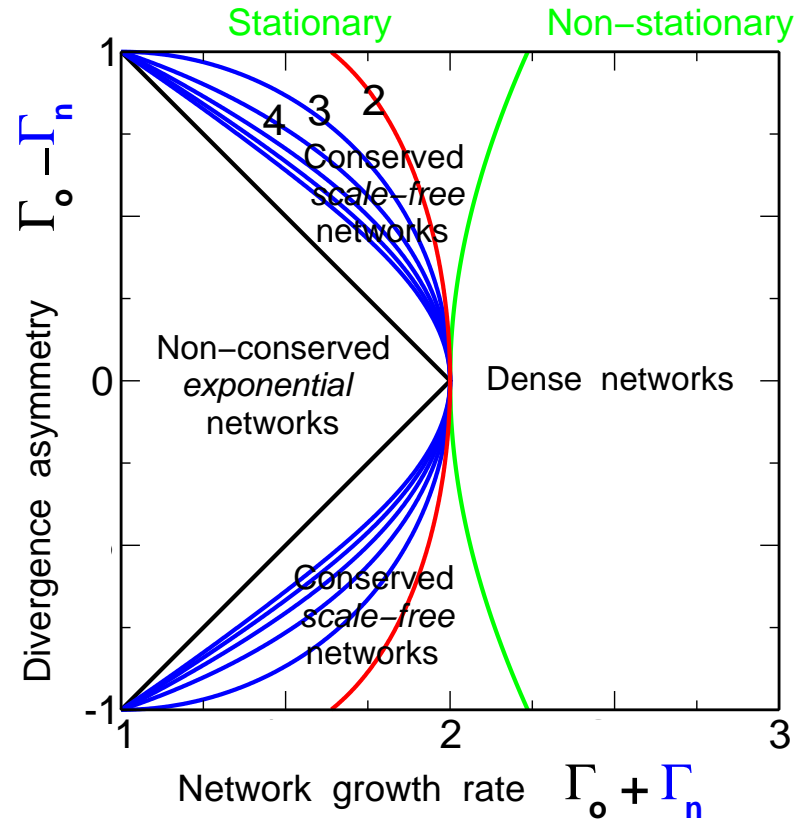
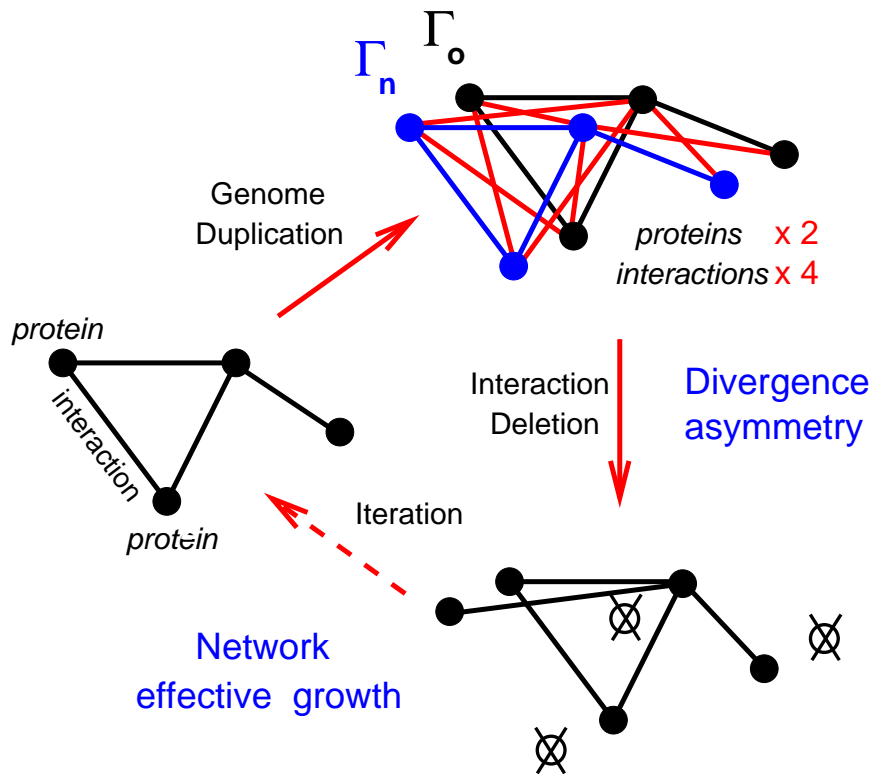
*"Spontaneous Symmetry Breaking"*  
between duplicated binding sites

Gene Duplicates



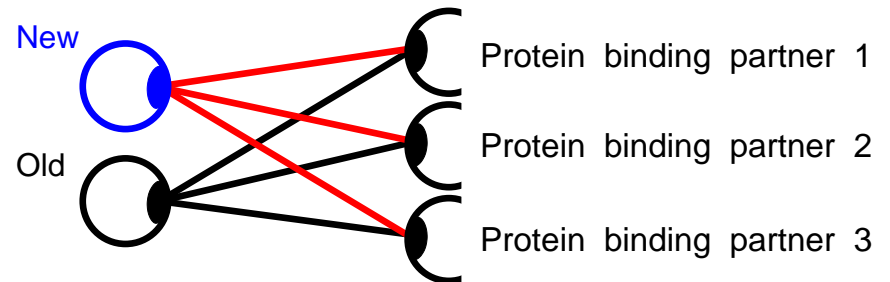
# Asymmetric Divergence of Duplicates is Required

Whole Genome Duplication Limit ( $q=1$ )



Origin of Divergence Asymmetry ?

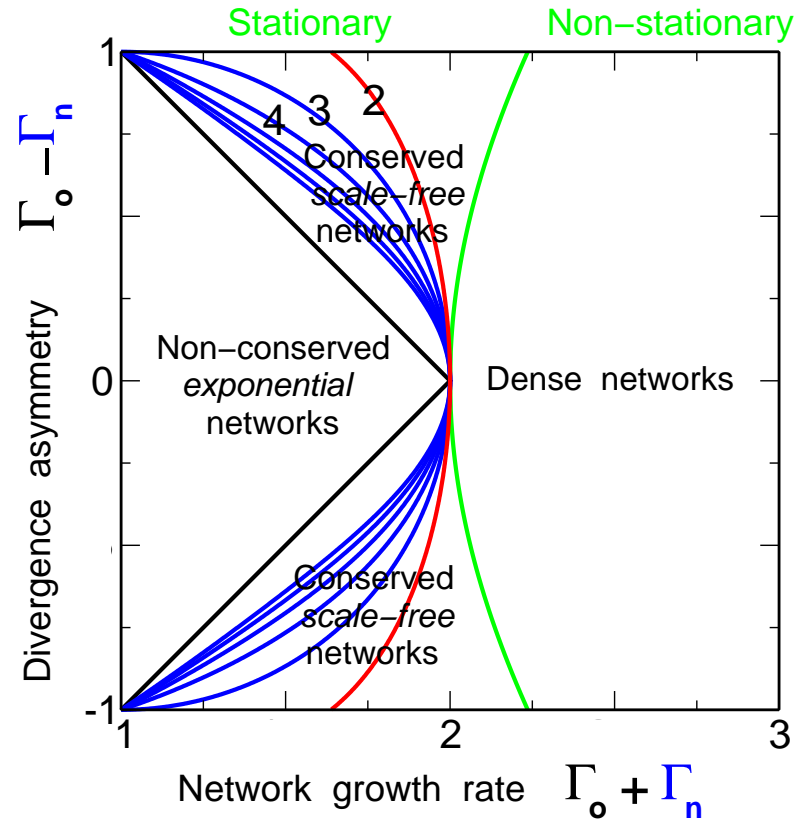
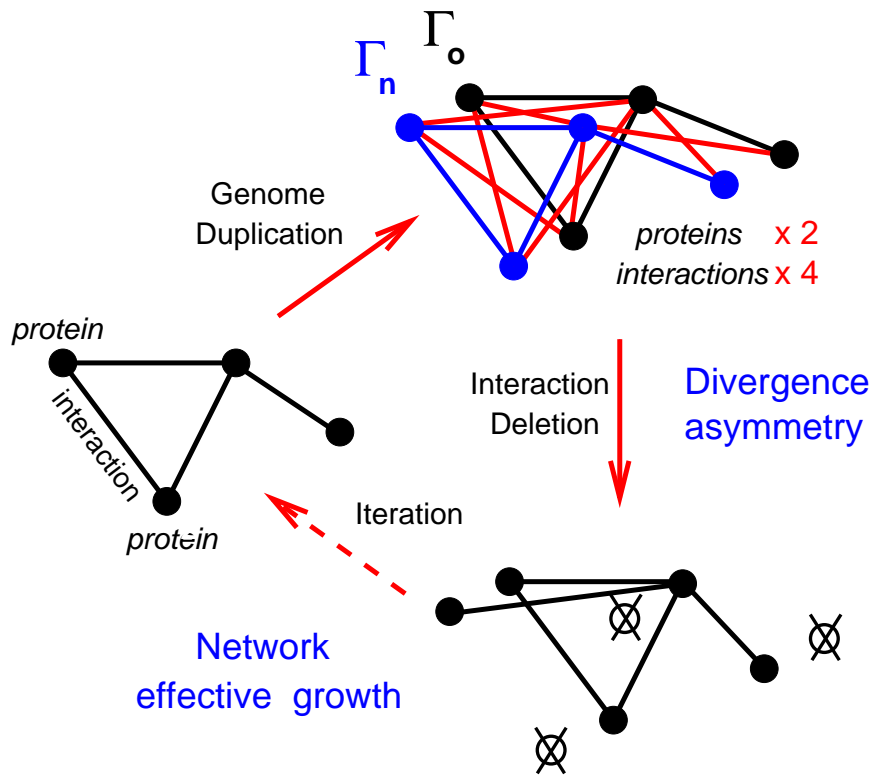
*"Spontaneous Symmetry Breaking"*  
*between duplicated binding sites*





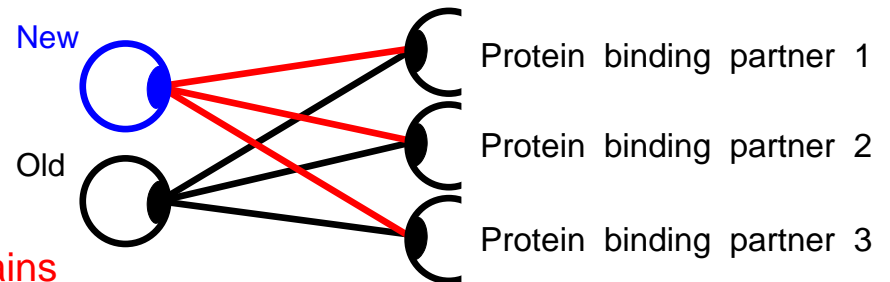
# Asymmetric Divergence of Duplicates is Required

Whole Genome Duplication Limit ( $q=1$ )



Origin of Divergence Asymmetry ?

*"Spontaneous Symmetry Breaking"  
between duplicated binding sites*



→ PPI Network Evolution in terms of protein domains

# Redefining PPI Networks as Domain Interaction Networks

## Genome and PPI Network Evolution under Protein Domain Shuffling

General concept



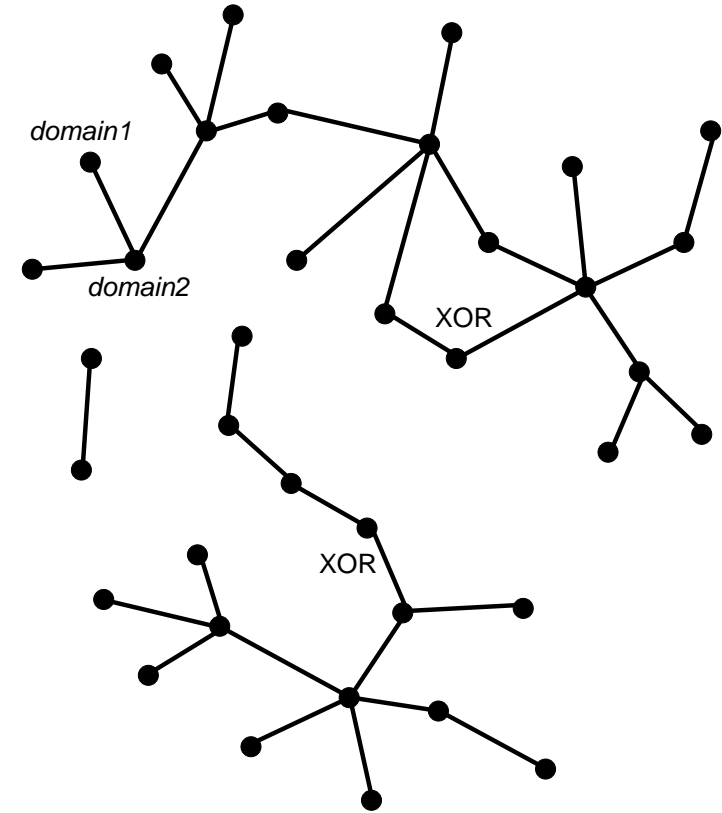
*C. elegans*



*E. coli* TrpC



## 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

General concept



*C. elegans*



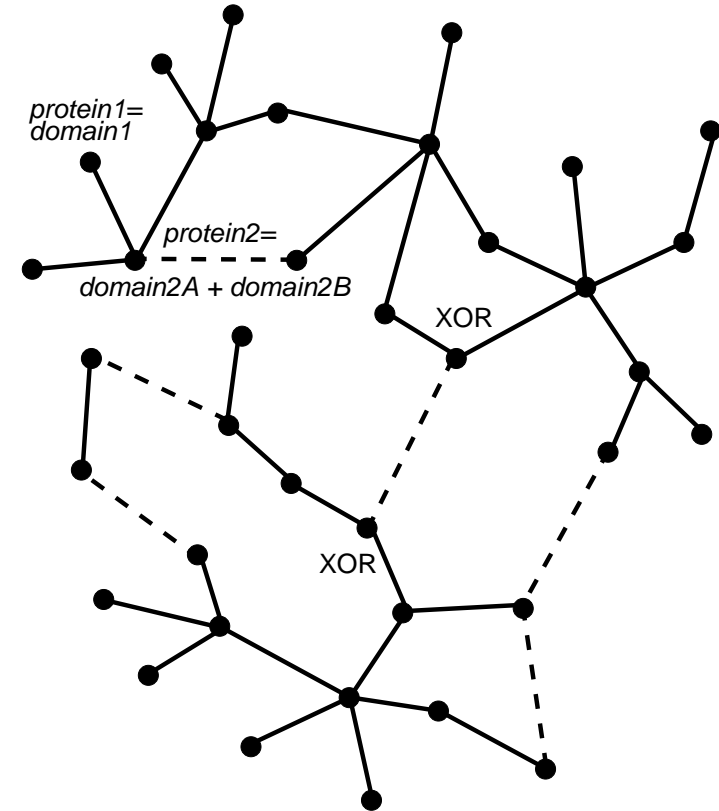
*E. coli* TrpC

Yeast TrpG

Yeast TrpF



## Domain–Domain Interaction Network



- Direct interaction bwn protein domains
- - - Multidomain proteins (random shuffling)

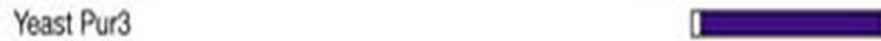
# Redefining PPI Networks as Domain Interaction Networks

## Genome and PPI Network Evolution under Protein Domain Shuffling

General concept



*C. elegans*



*E. coli* TrpC



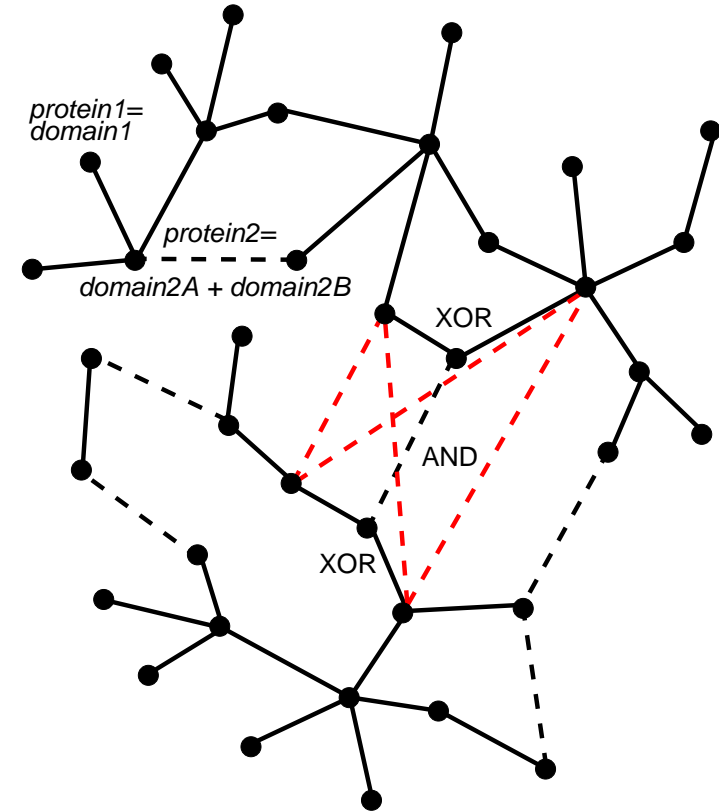
Yeast TrpG



Yeast TrpF



## Domain–Domain Interaction Network



- Direct interaction bwn protein domains
- - - Multidomain proteins (random shuffling)
- - - Indirect interaction within complexes

→ Adding some "combinatorial logic" through multidomain proteins

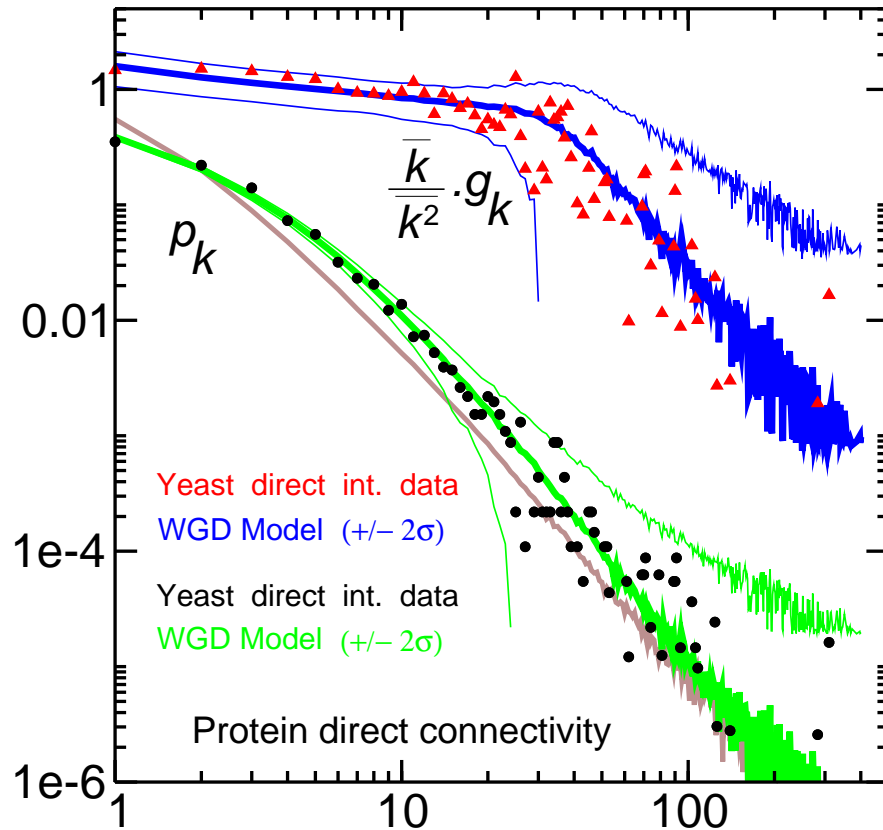
# Comparison with Yeast Direct Interaction Data

A two-parameter Whole Genome Duplication Model with Random Protein Domain Shuffling  $\gamma_{on}$   $\lambda$

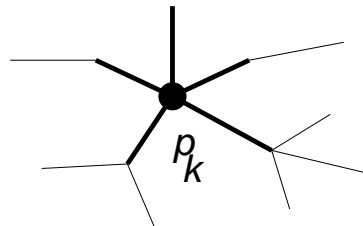
$$\gamma_{on}=0.1 \quad (\gamma_{oo}=1 \quad \gamma_{nn}=0)$$

$$\lambda=1.5 \text{ prot. bind. domains per protein}$$

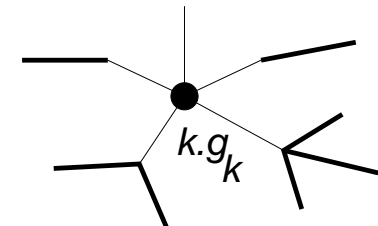
## Topology of direct interaction network



Node Degree Distribution



Average Connectivity of Neighbours



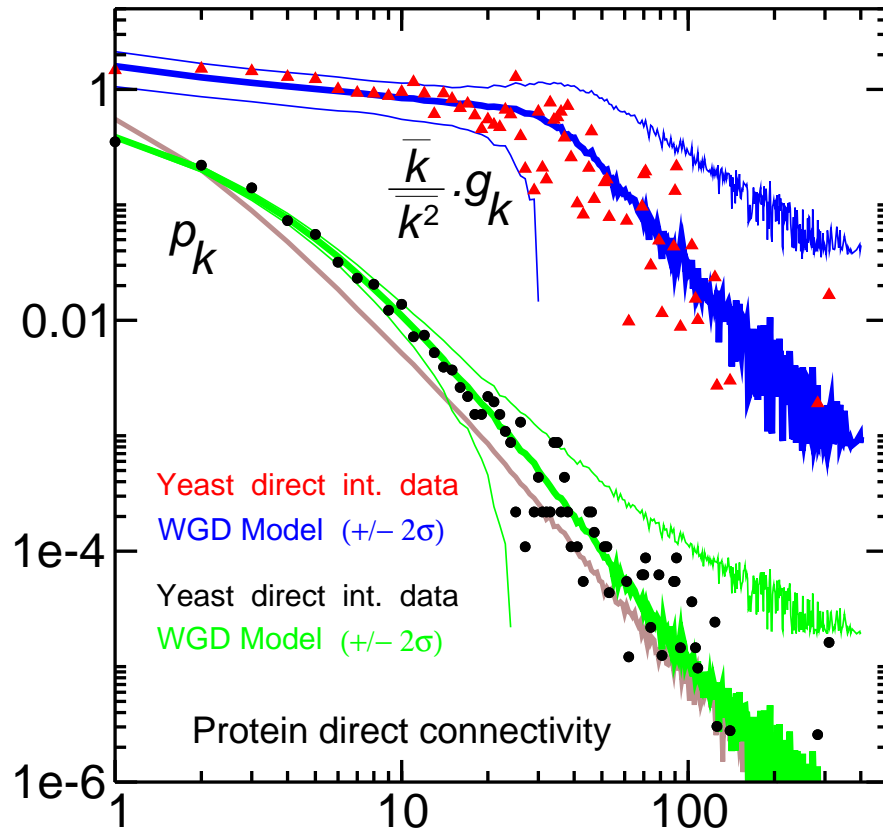
# Comparison with Yeast Direct Interaction Data

A two-parameter Whole Genome Duplication Model with Random Protein Domain Shuffling  $\gamma_{on}$   $\lambda$

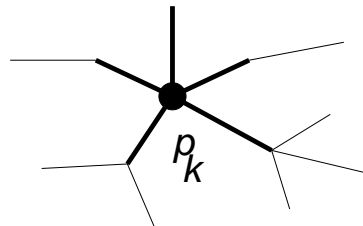
$$\gamma_{on} = 0.1 \quad (\gamma_{oo} = 1 \quad \gamma_{nn} = 0)$$

$$\lambda = 1.5 \text{ prot. bind. domains per protein}$$

Topology of direct interaction network



Node Degree Distribution



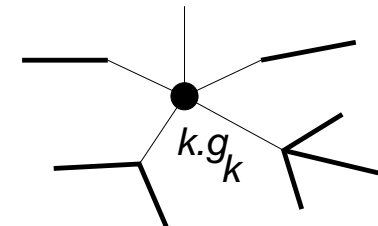
PPI Network growth rate

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

Seasquirt-Human (2WGD): 25%

Seasquirt-Fugu (3WGD): 11%

Average Connectivity of Neighbours



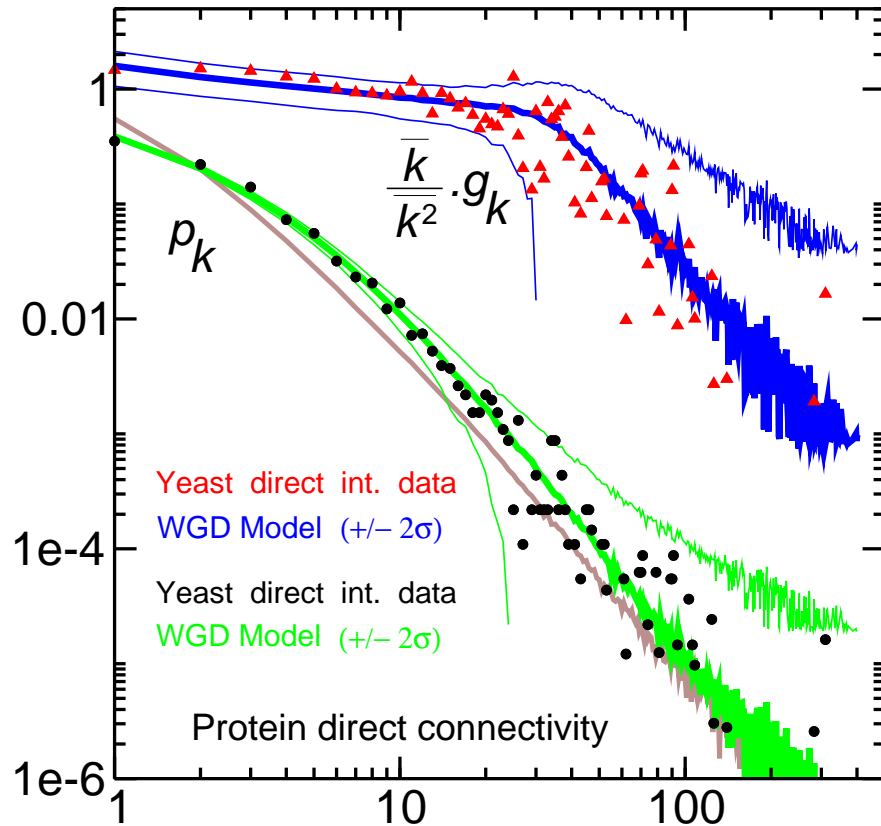
# Comparison with Yeast Direct + Indirect Interaction Data

A two-parameter Whole Genome Duplication Model with Random Protein Domain Shuffling  $\gamma_{on}$   $\lambda$

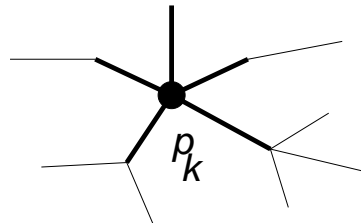
$$\gamma_{on} = 0.1 \quad (\gamma_{oo} = 1 \quad \gamma_{nn} = 0)$$

$$\lambda = 1.5 \text{ prot. bind. domains per protein}$$

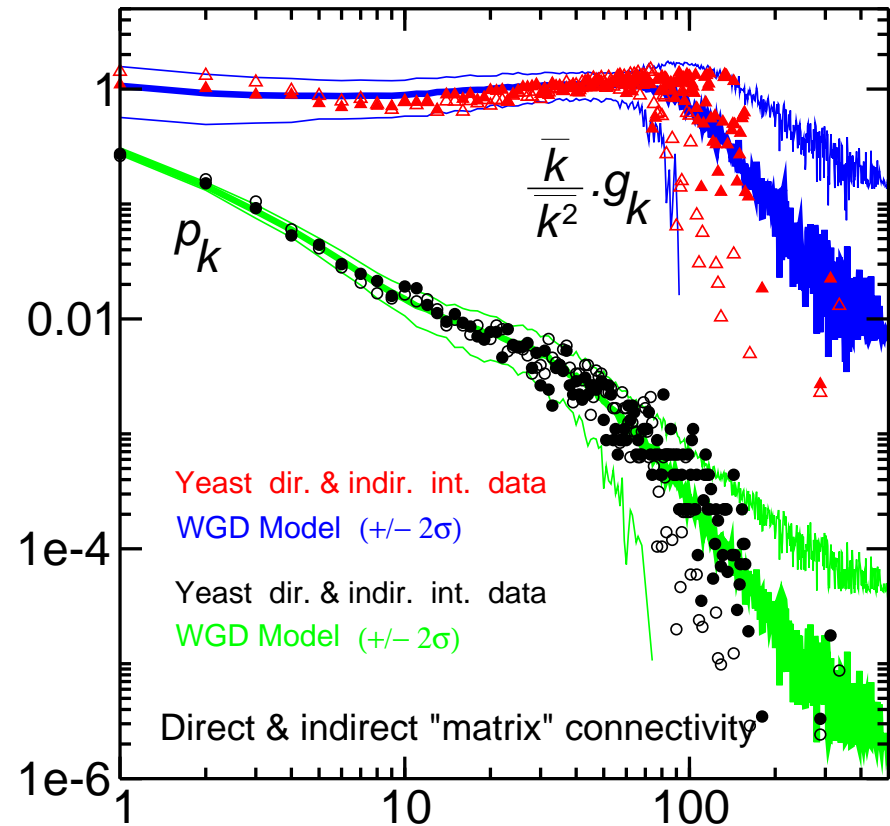
Topology of direct interaction network



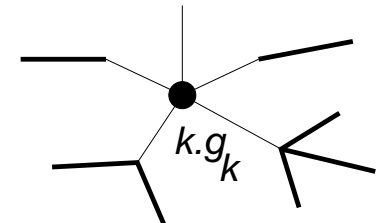
Node Degree Distribution



Topology of direct & indirect interaction network



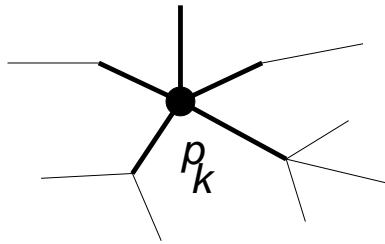
Average Connectivity of Neighbours



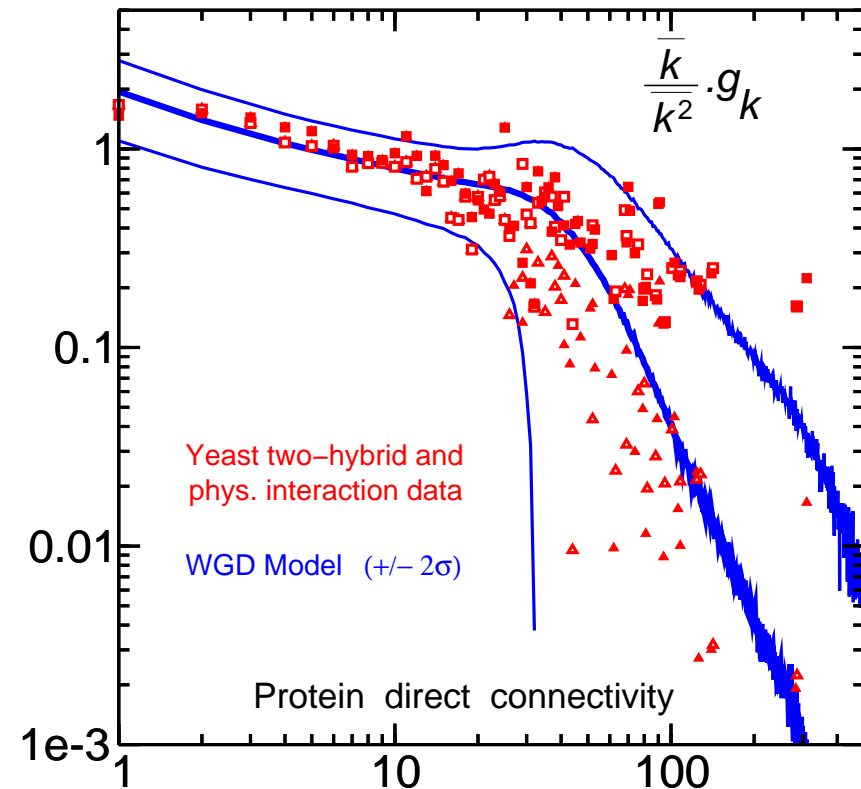
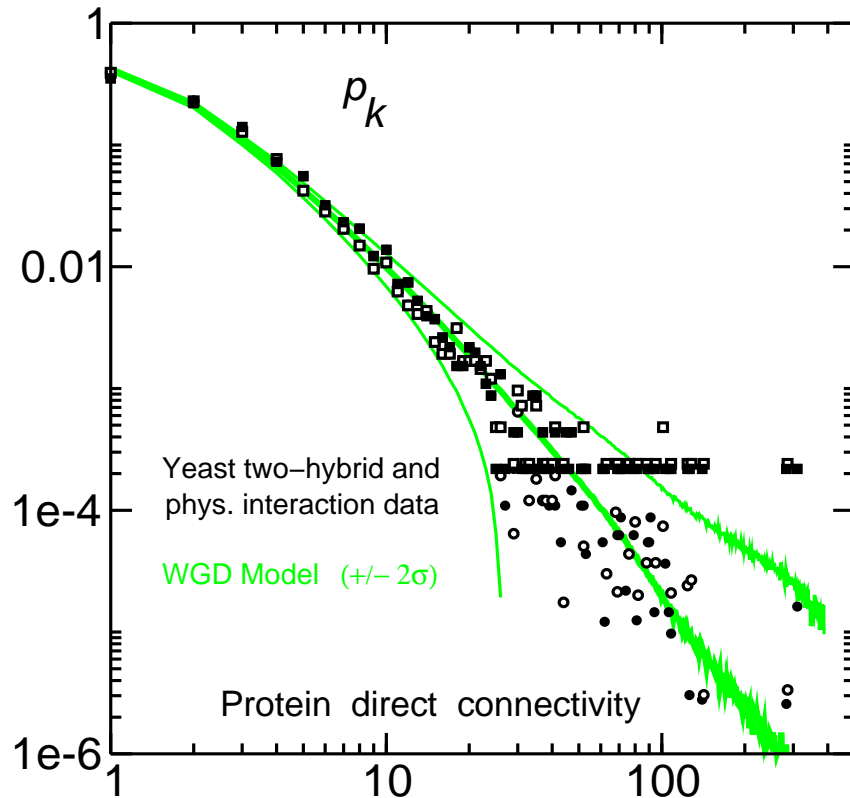
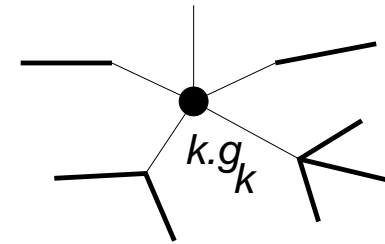
# Comparison with Yeast Direct Interaction Data

Fit of Yeast Data with a one-parameter WGD Model  $\gamma_{on}=0.26$  ( $\gamma_{oo}=1$   $\gamma_{nn}=0$ )

Node Degree Distribution

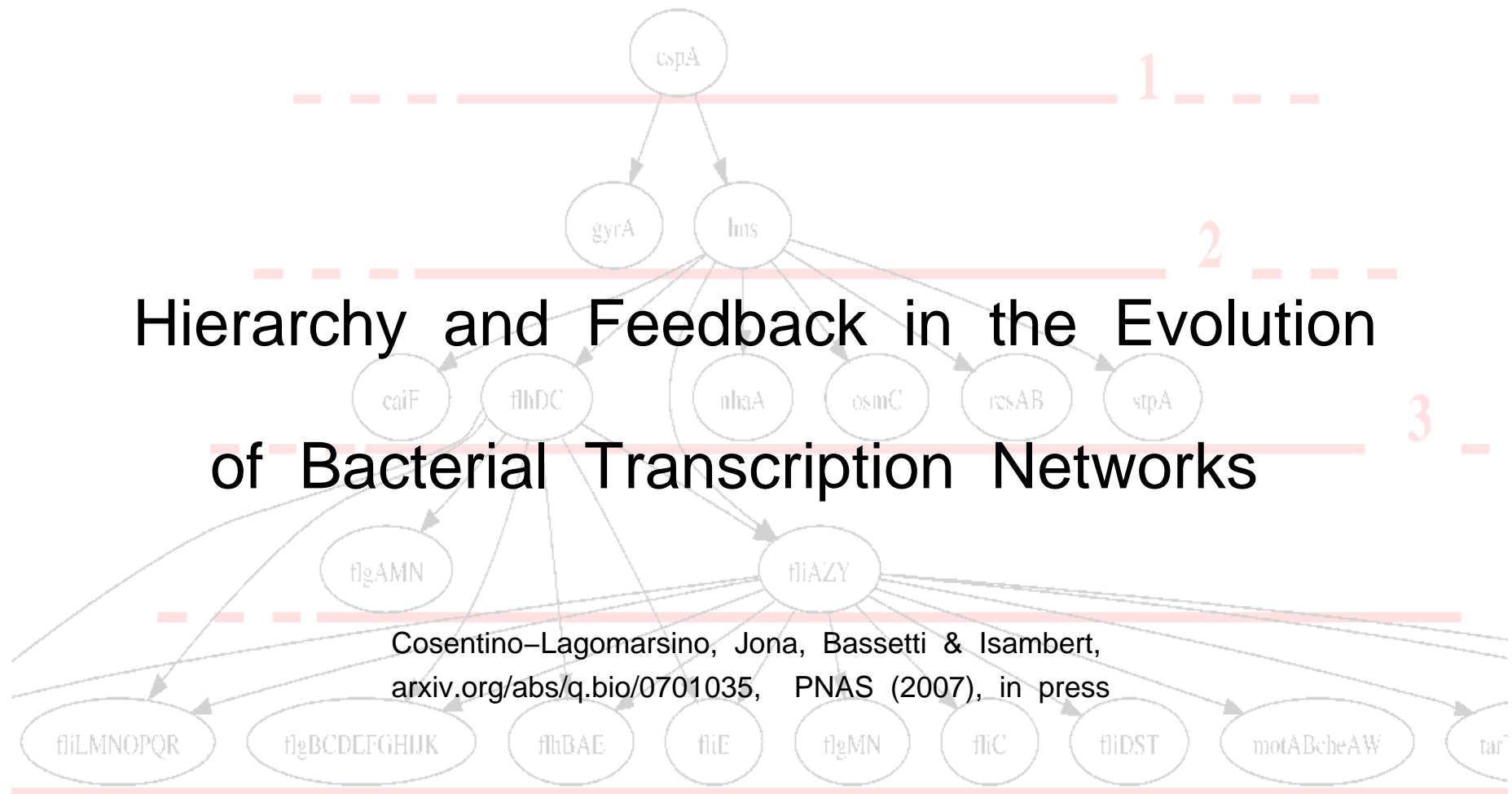


Average Connectivity of Neighbours



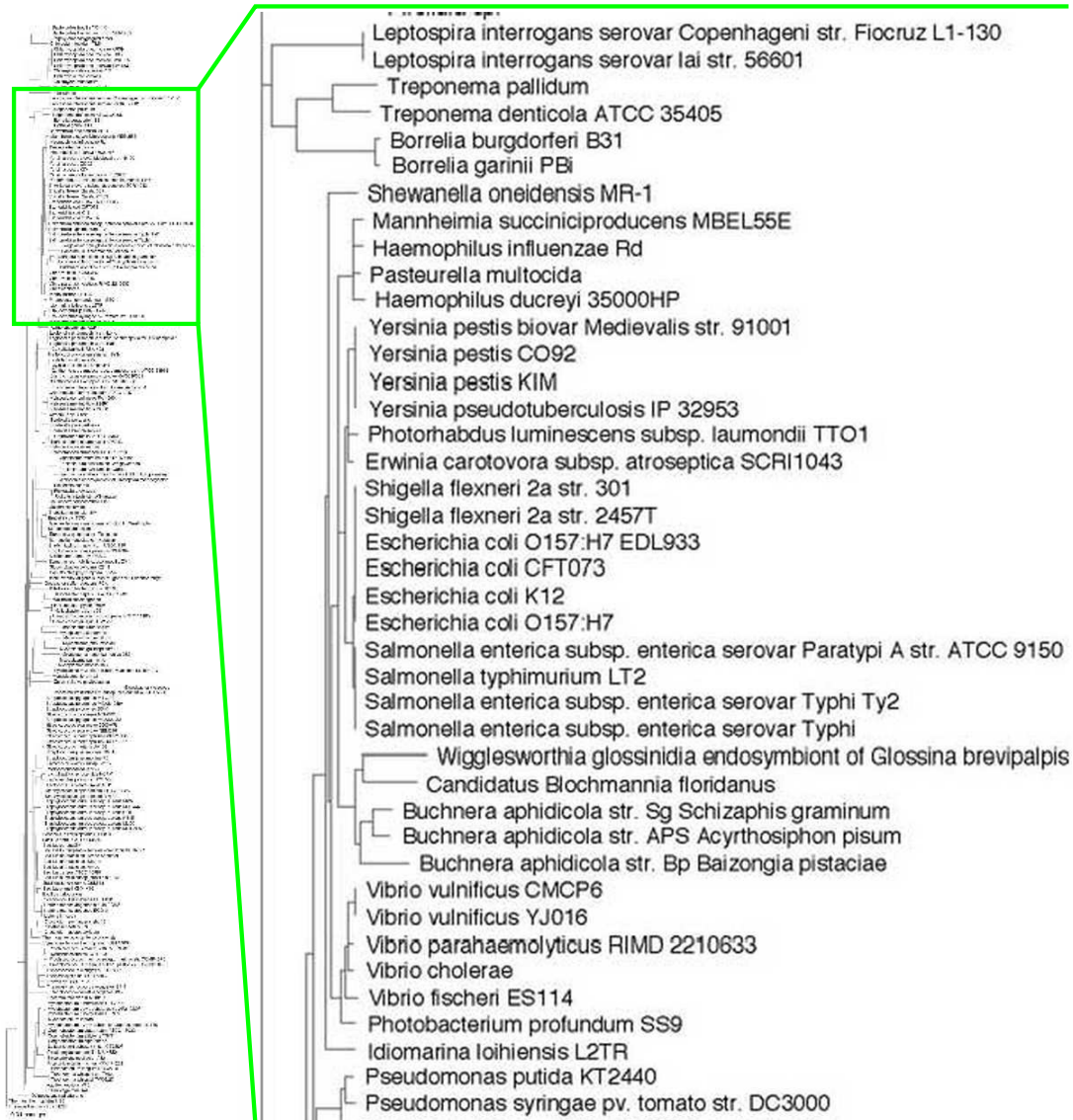


# Hierarchy and Feedback in the Evolution of Bacterial Transcription Networks



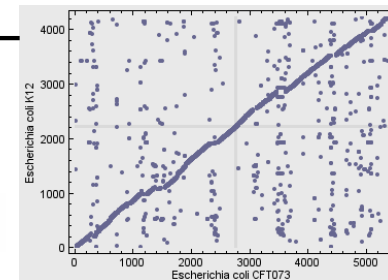
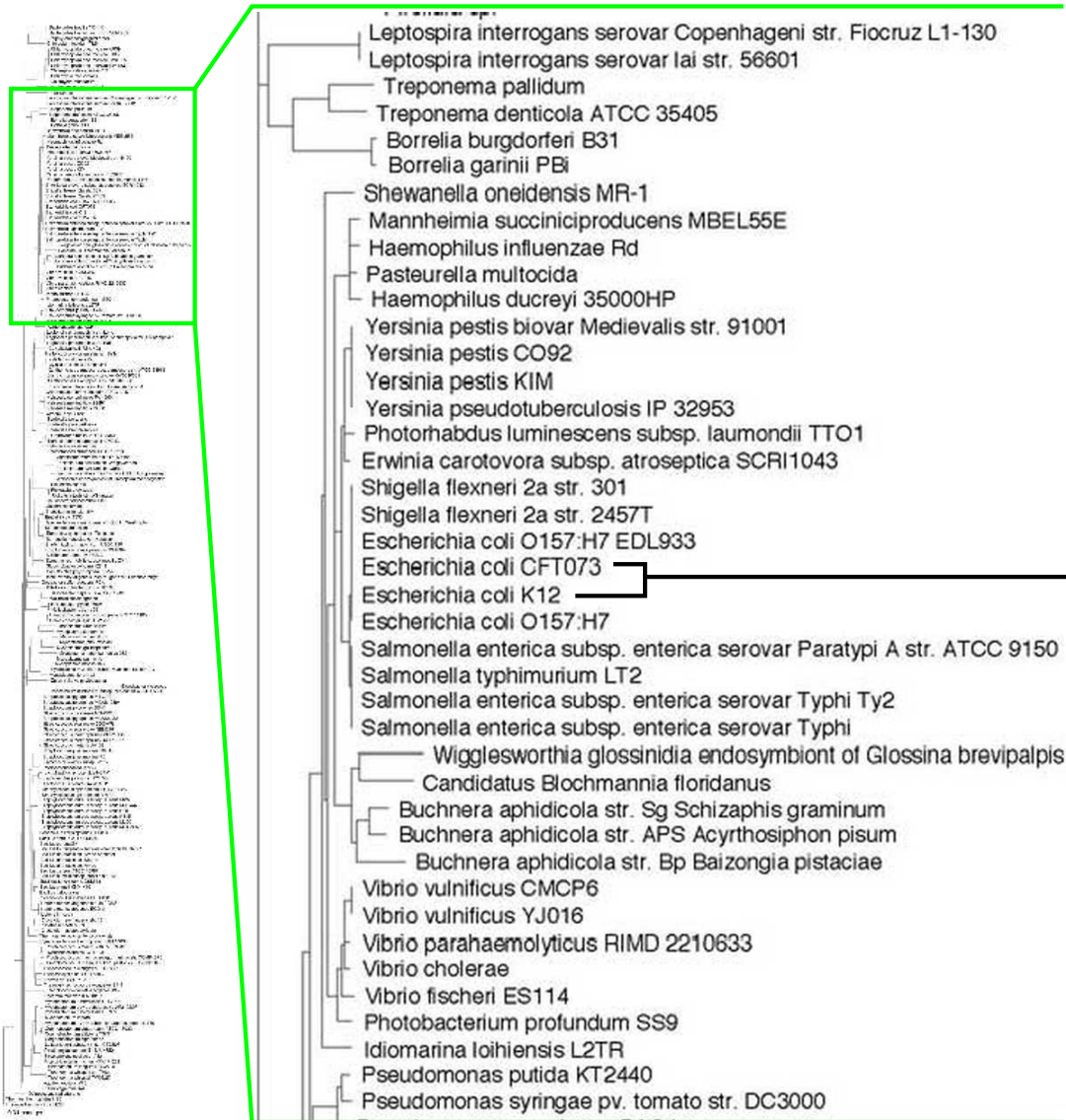
# Prokaryote Genome Evolution

300 sequenced genomes



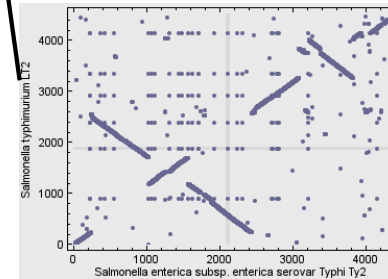
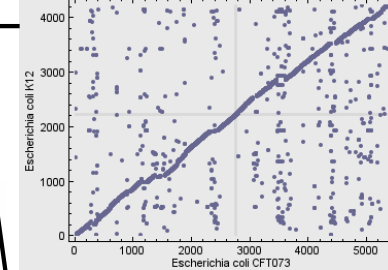
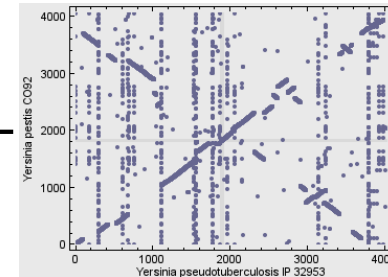
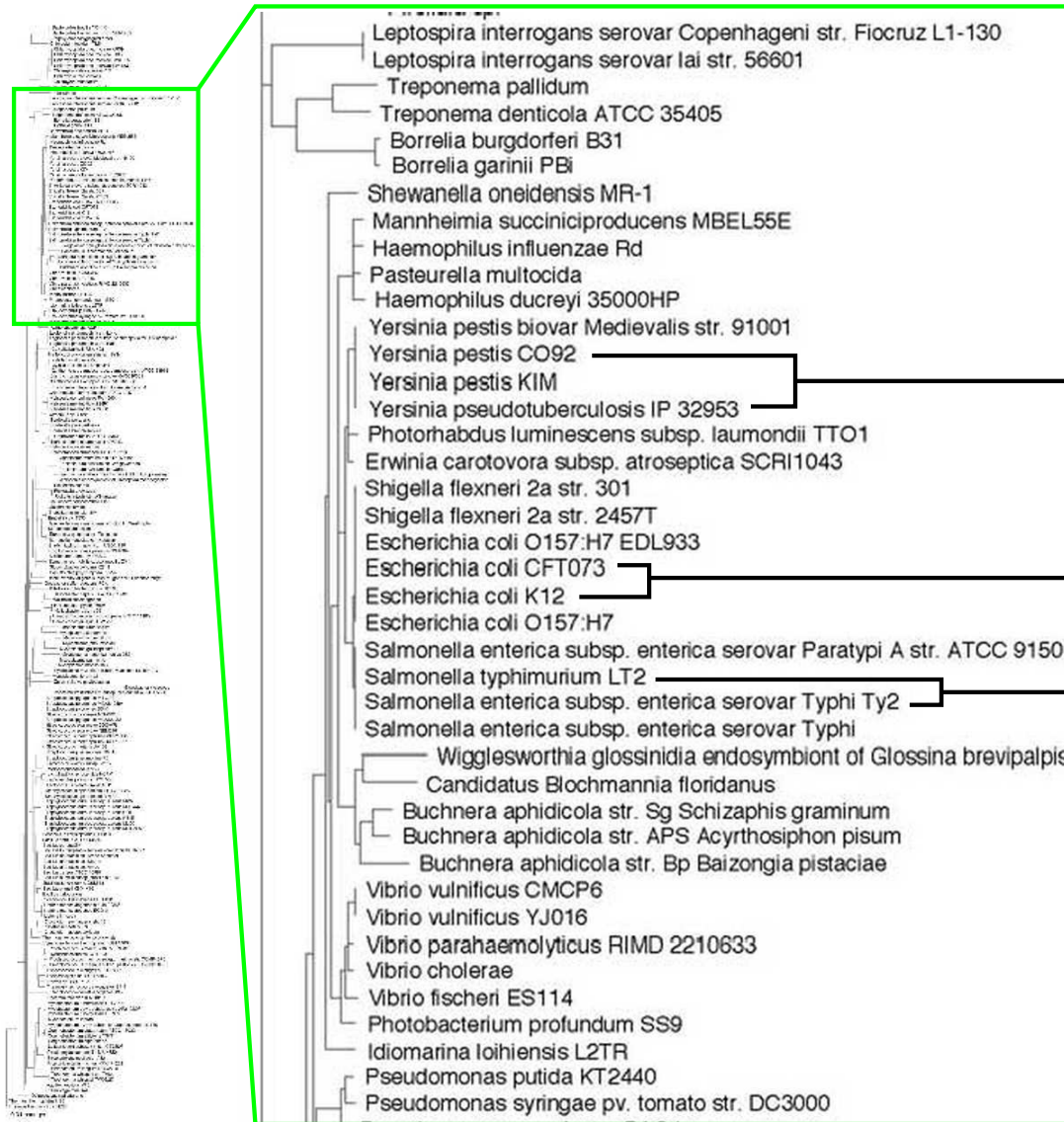
# Prokaryote Genome Evolution

300 sequenced genomes



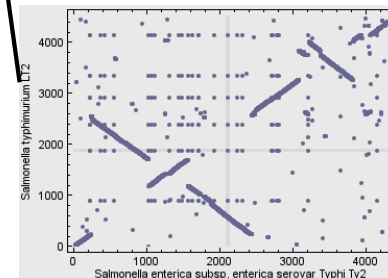
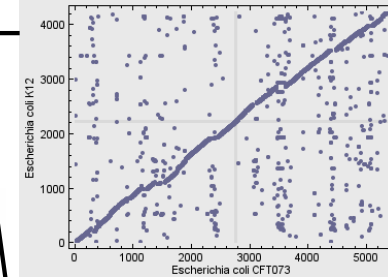
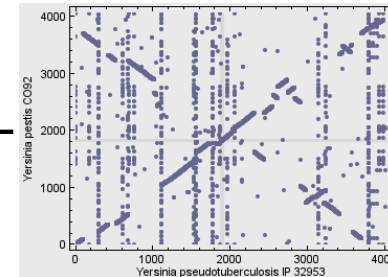
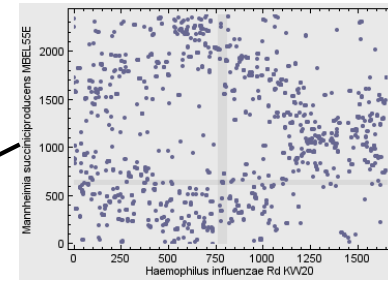
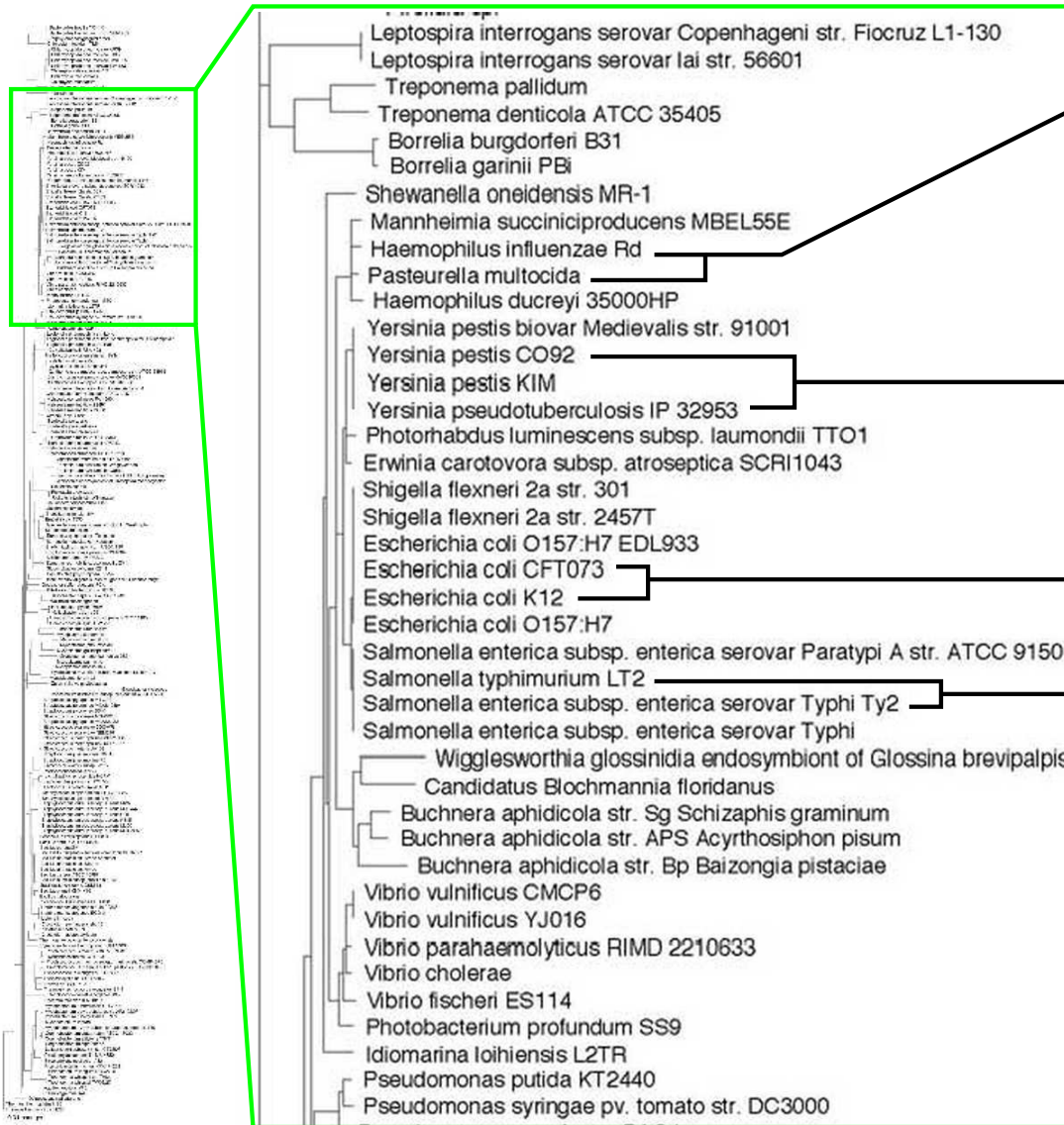
# Prokaryote Genome Evolution

300 sequenced genomes



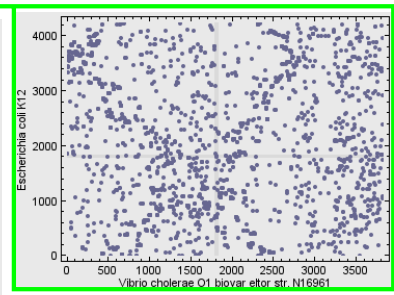
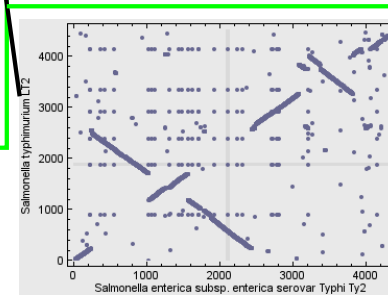
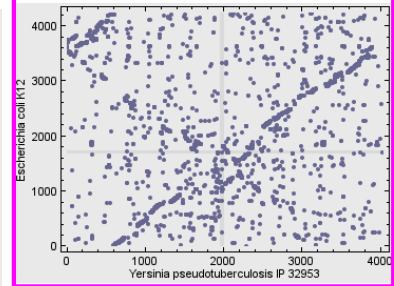
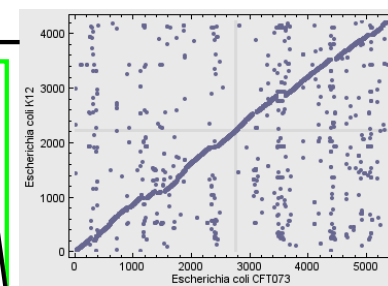
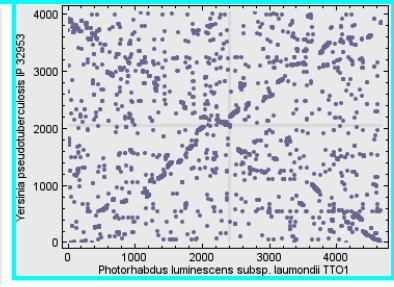
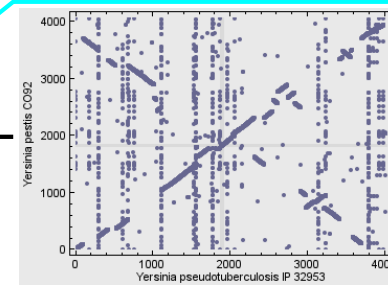
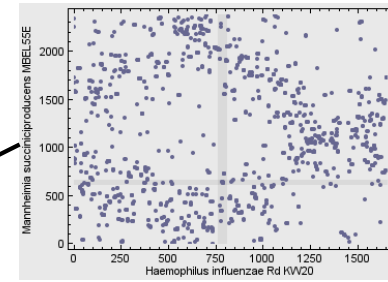
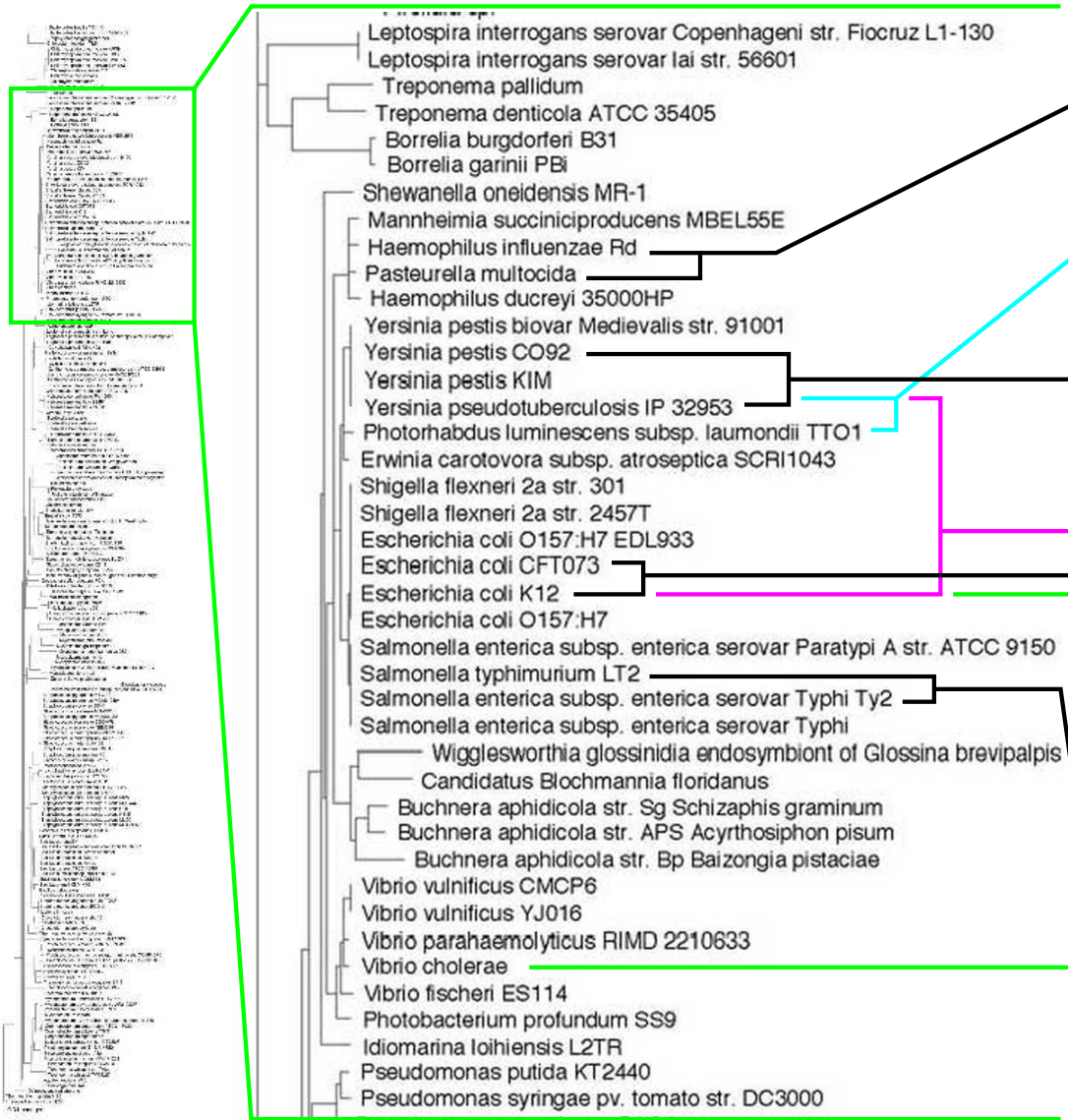
# Prokaryote Genome Evolution

300 sequenced genomes



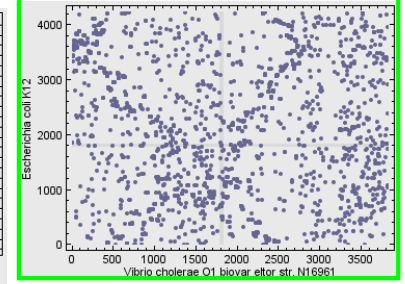
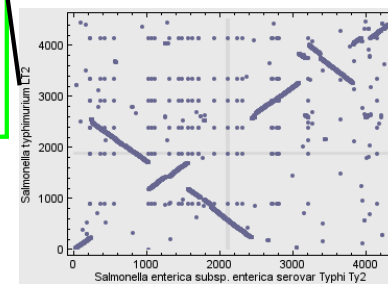
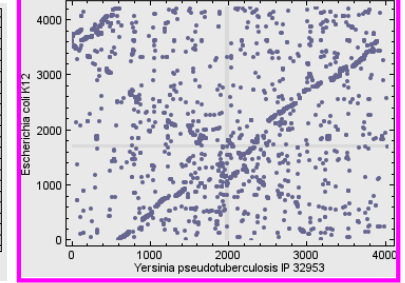
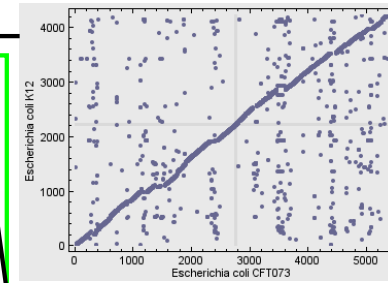
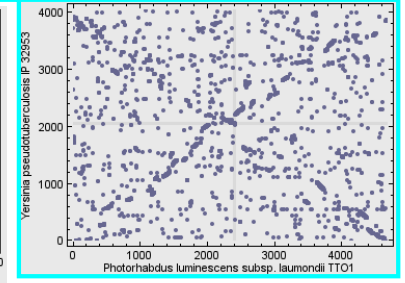
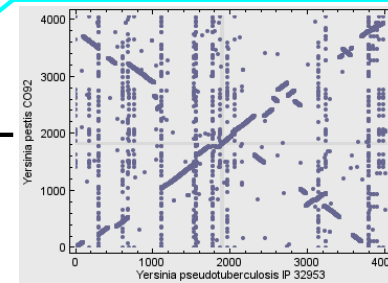
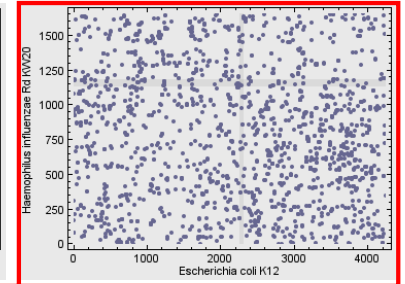
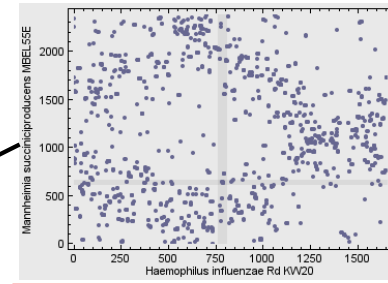
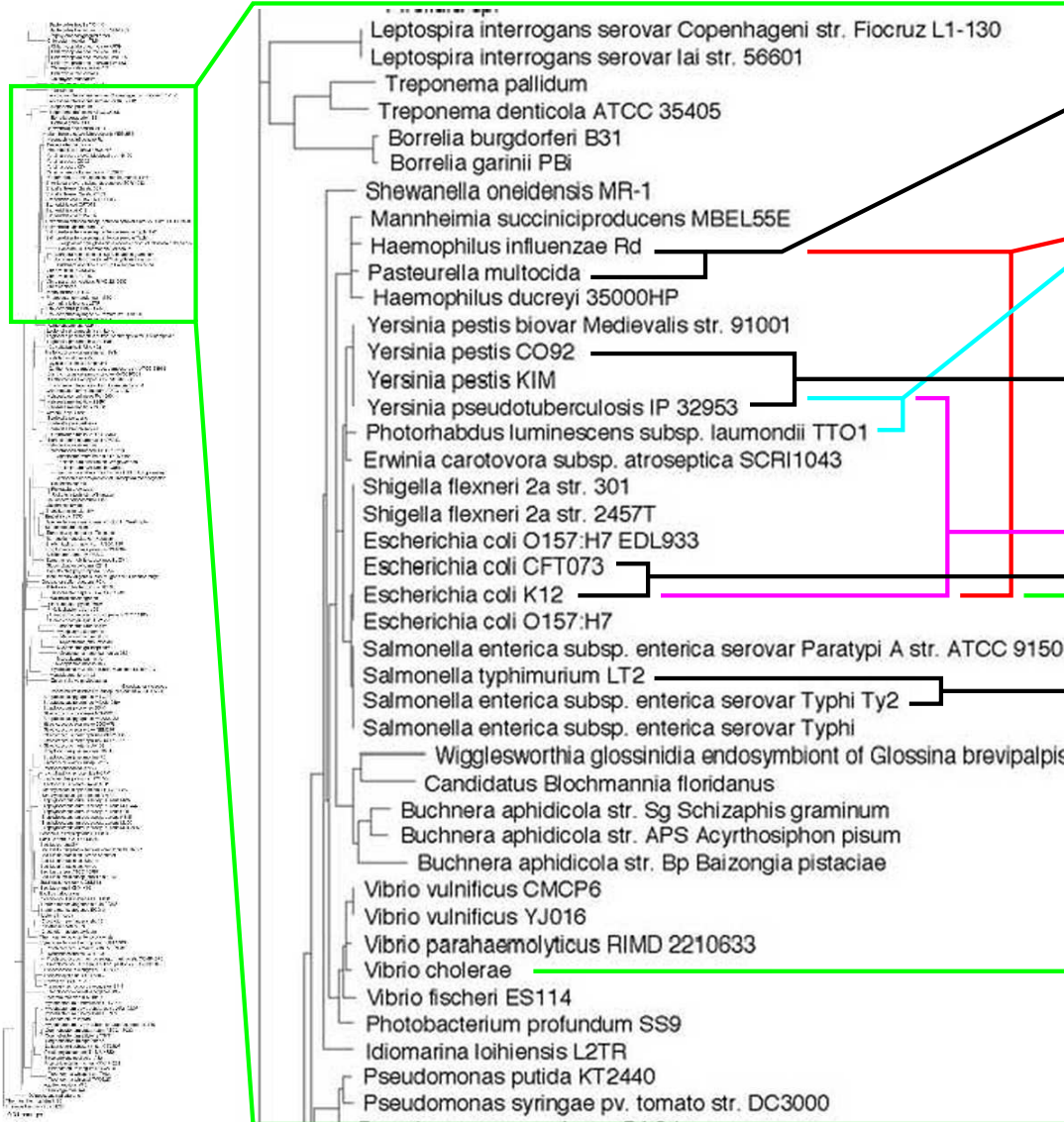
# Prokaryote Genome Evolution

300 sequenced genomes

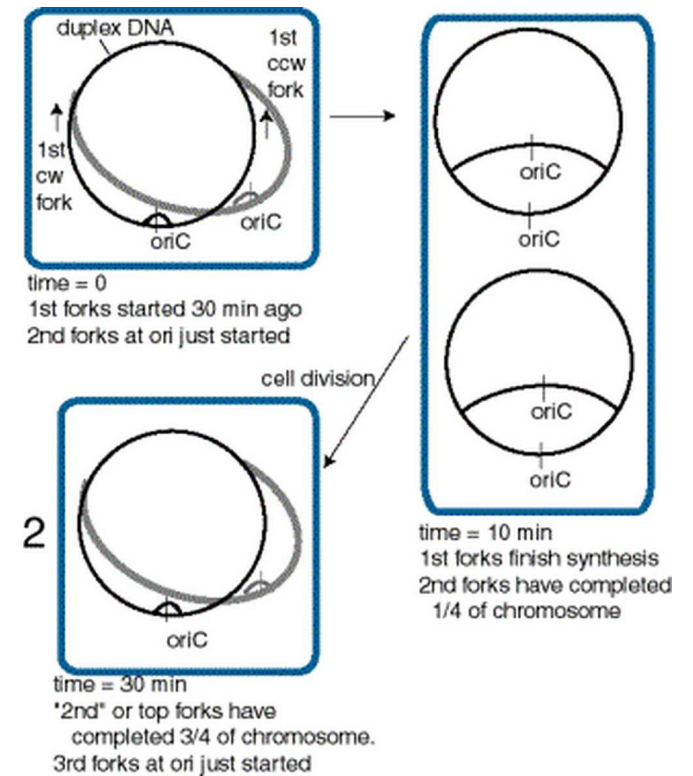
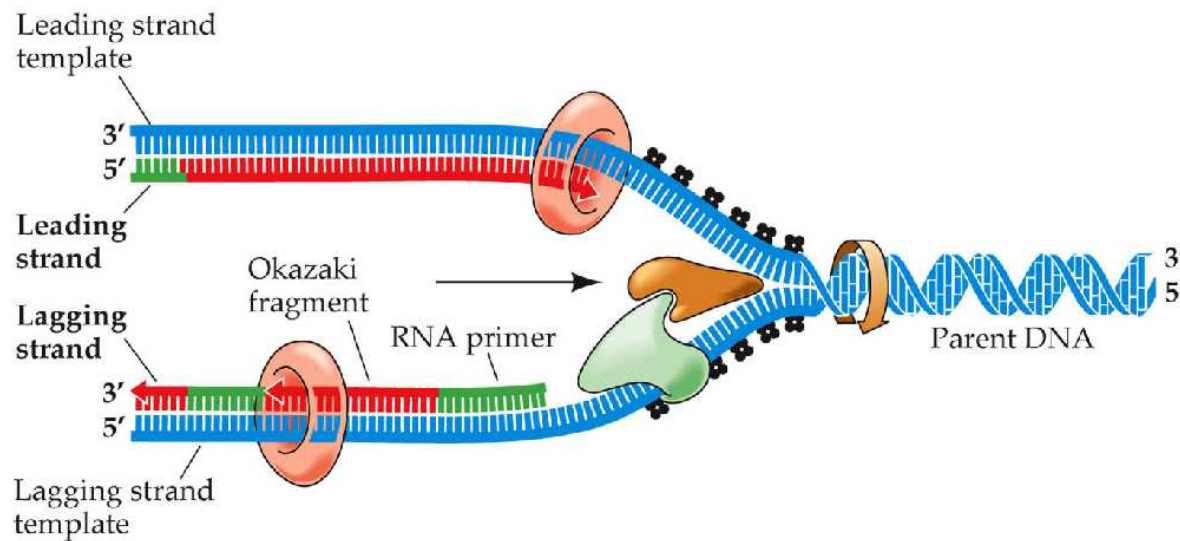


# Prokaryote Genome Evolution

300 sequenced genomes

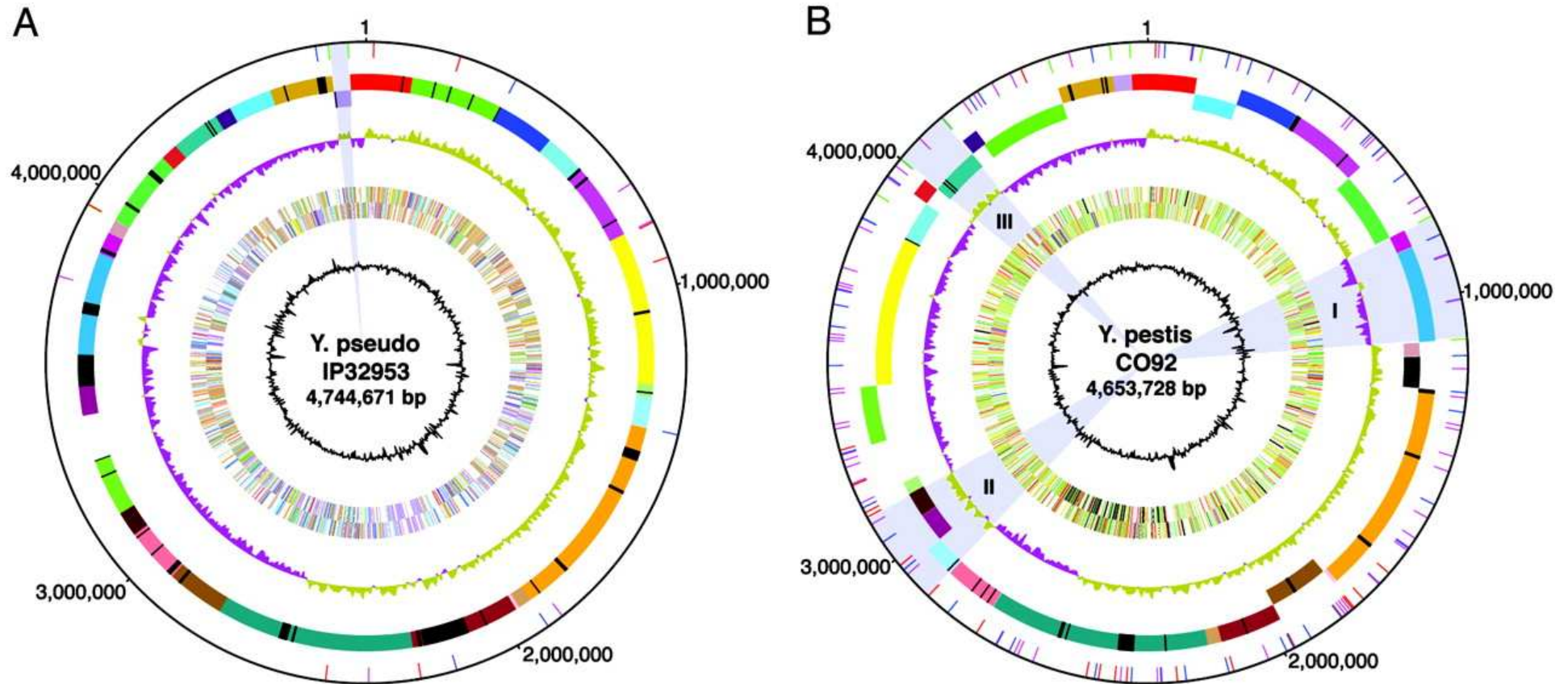


# Replication pressure on bacterial genome evolution

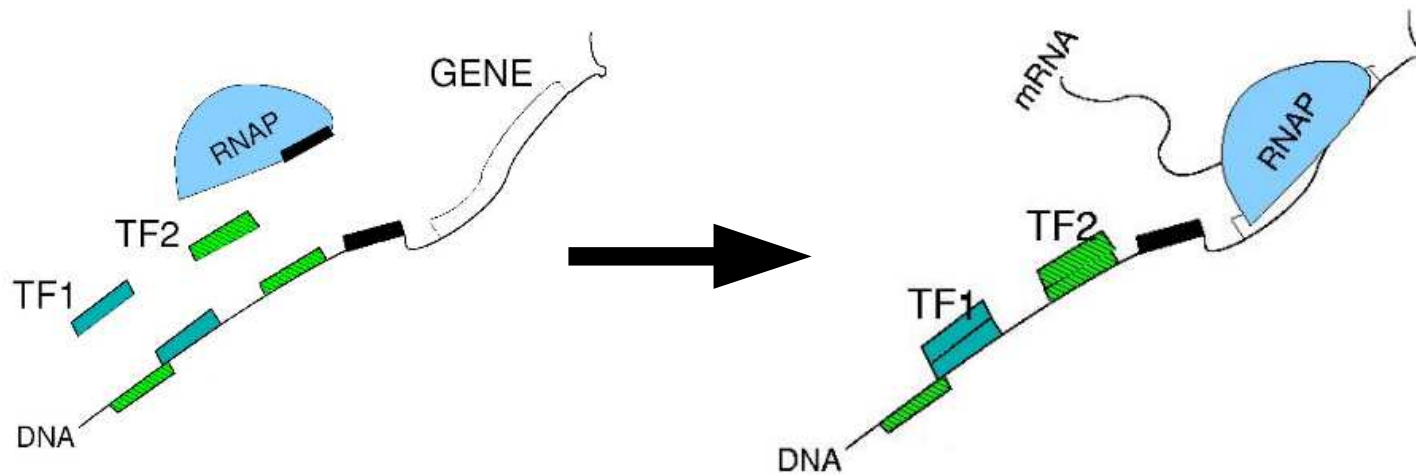




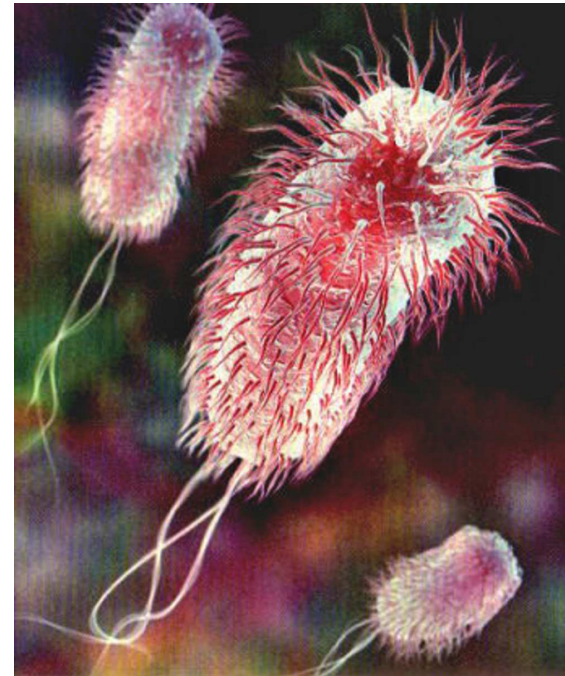
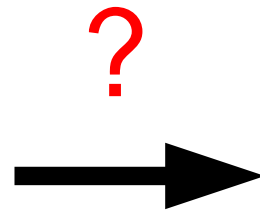
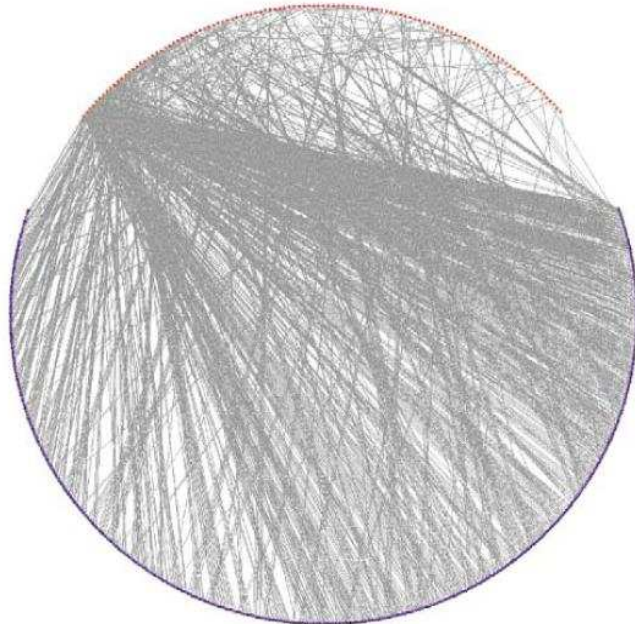
# *Yersinia pseudotuberculosis* versus *Yersinia pestis*



# Transcription in Bacteria



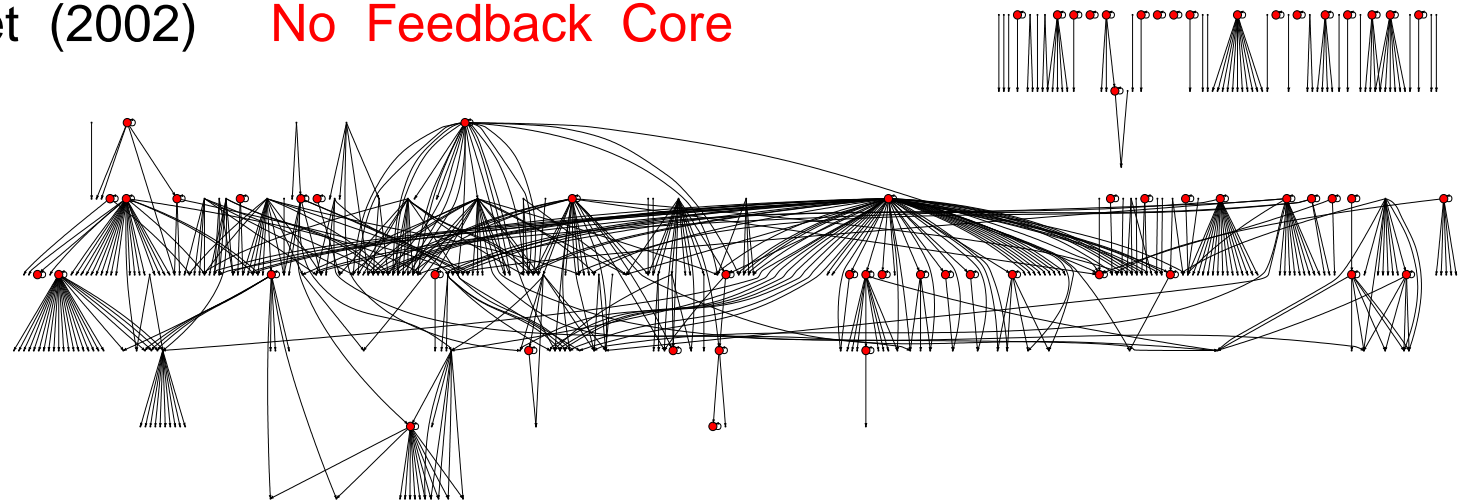
## Regulatory Interactions



# Hierarchy with Mostly Self-Regulatory Feedback

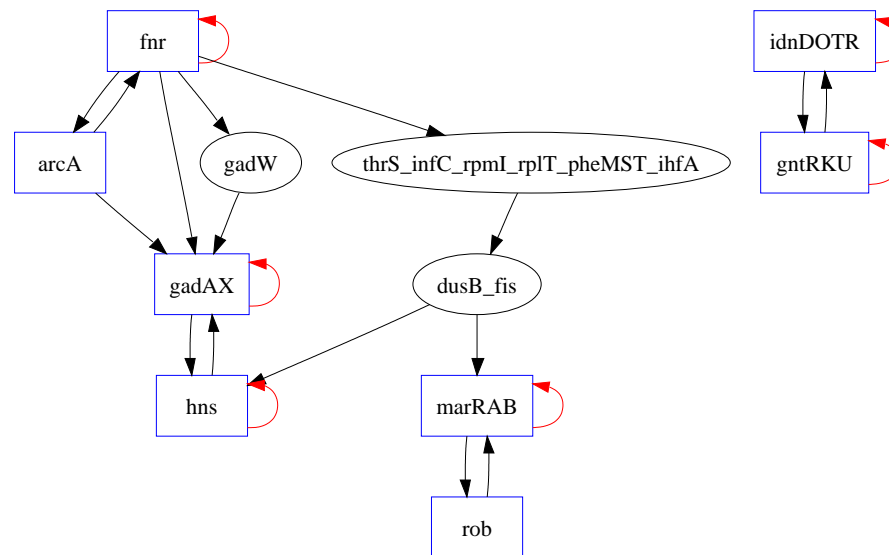
Shen-Orr Dataset (2002) **No Feedback Core**

Operons : 423  
 TFs : 117  
**ARs : 59 (50% TFs)**  
 Reg Int : 578

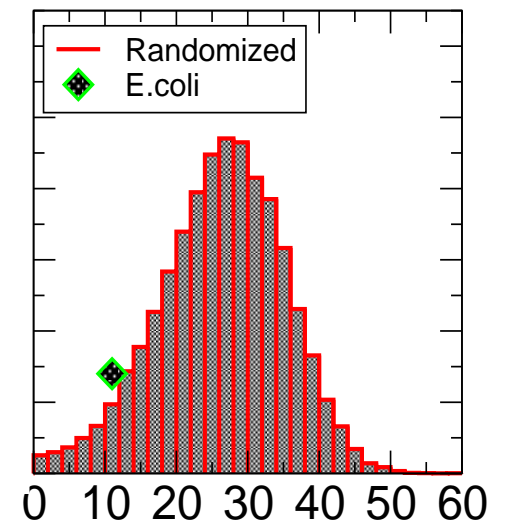


RegulonDB 5.5 (2006) **Small Feedback Core**

Operons : 648  
 TFs : 147  
**ARs : 85 (58% TFs)**  
 Reg Int : 1170



Feedback Core Sizes



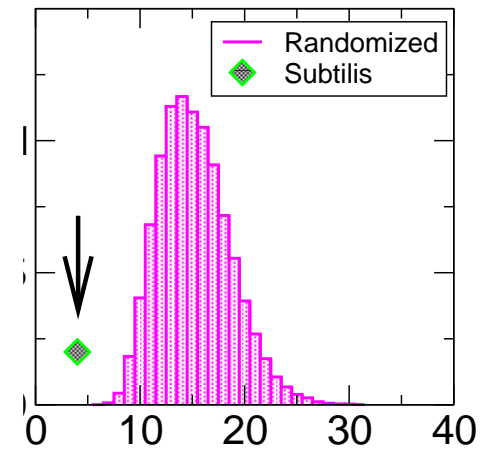
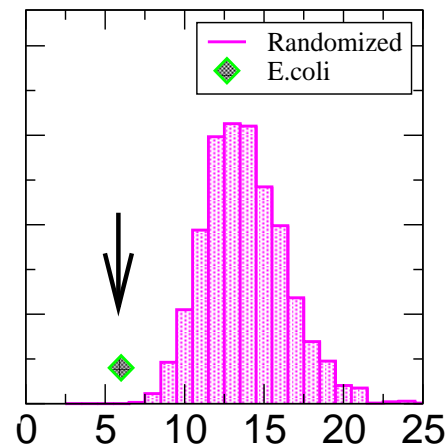
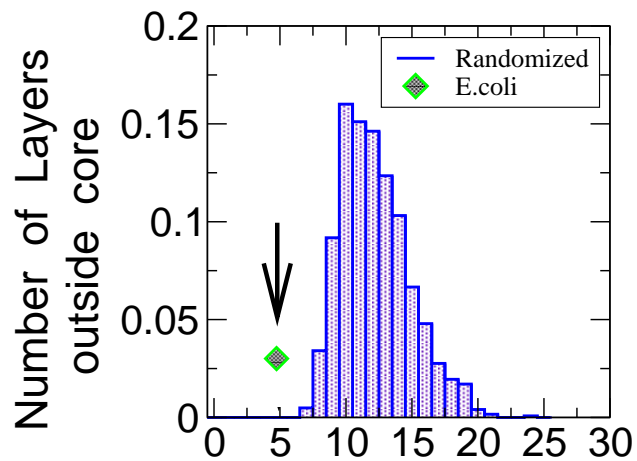
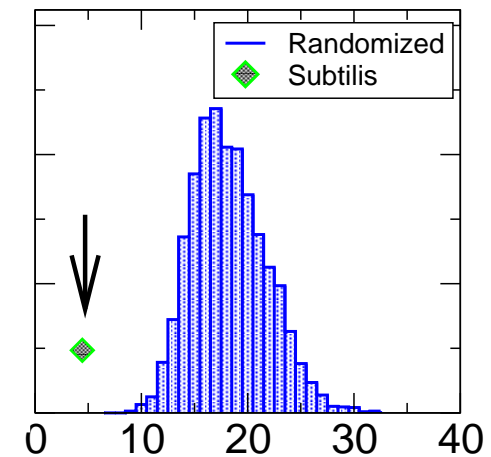
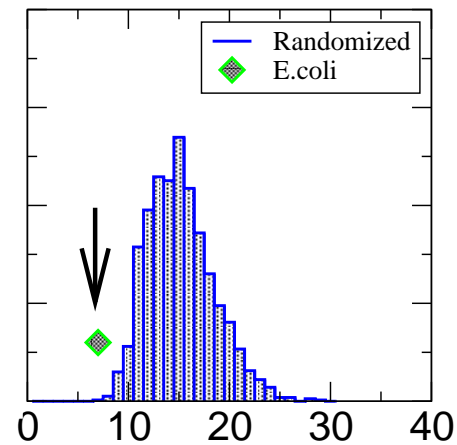
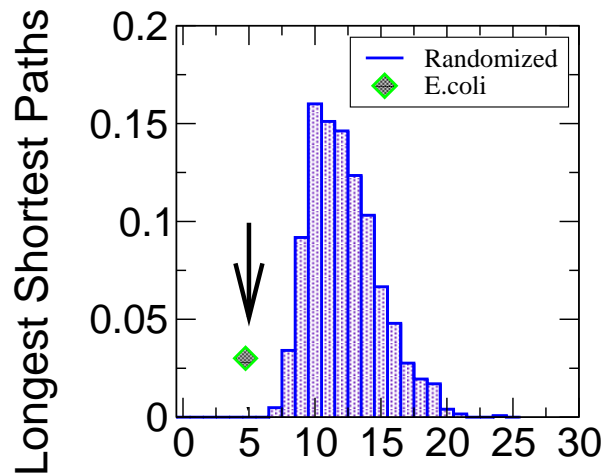
# Hierarchy with few Transcriptional Layers

*Escherichia coli*

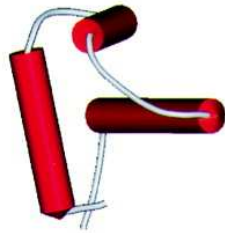
*Bacillus subtilis*

Shen Orr et al (2002)

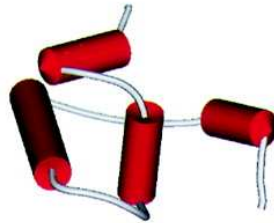
RegulonDB 5.5 (2006)



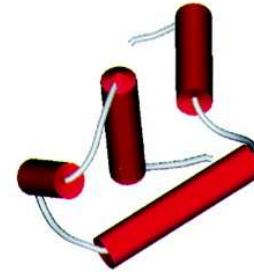
# Transcription Factor Families



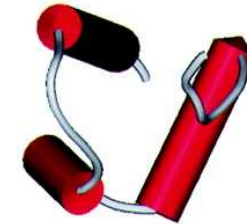
**Winged helix**



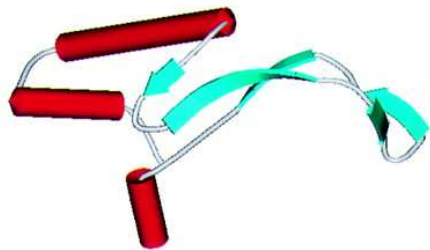
**Lambda repressor-like**



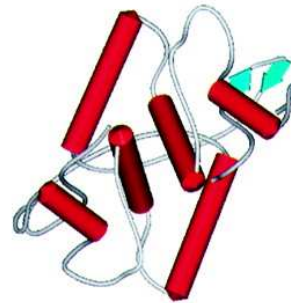
**C-terminal effector domain of  
the bipartite response regulator**



**Homeodomain-like**



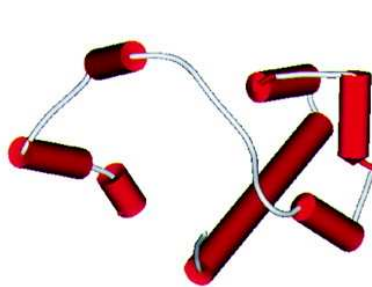
**IHF-like DNA-binding proteins**



**Met repressor-like**



**Putative DNA binding protein**



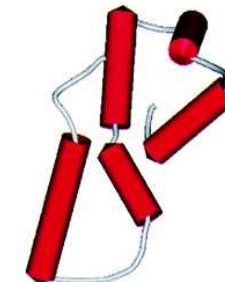
**Flagellar transcriptional  
activator FlhD**



**Trp repressor**

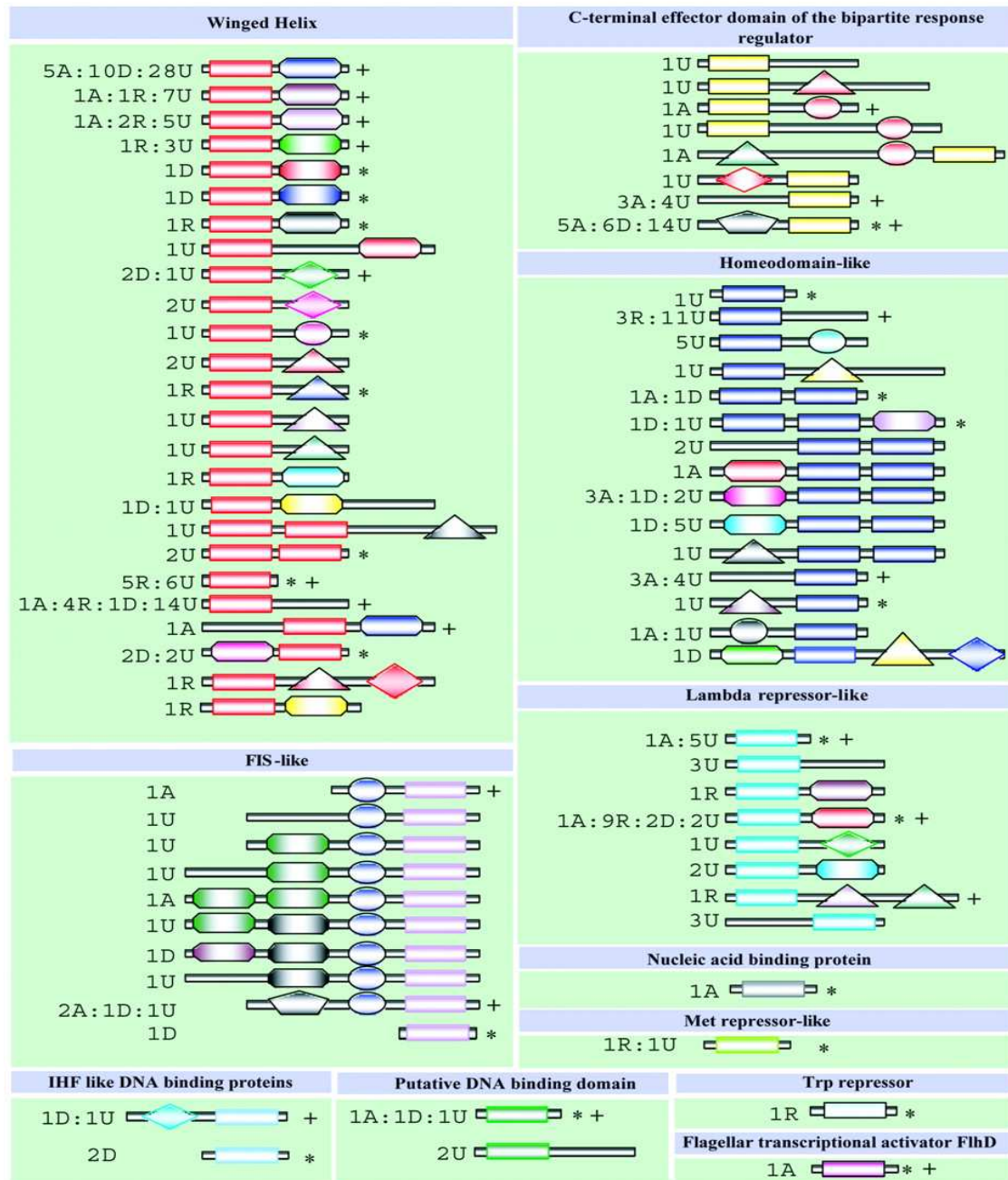


**Nucleic acid binding protein**

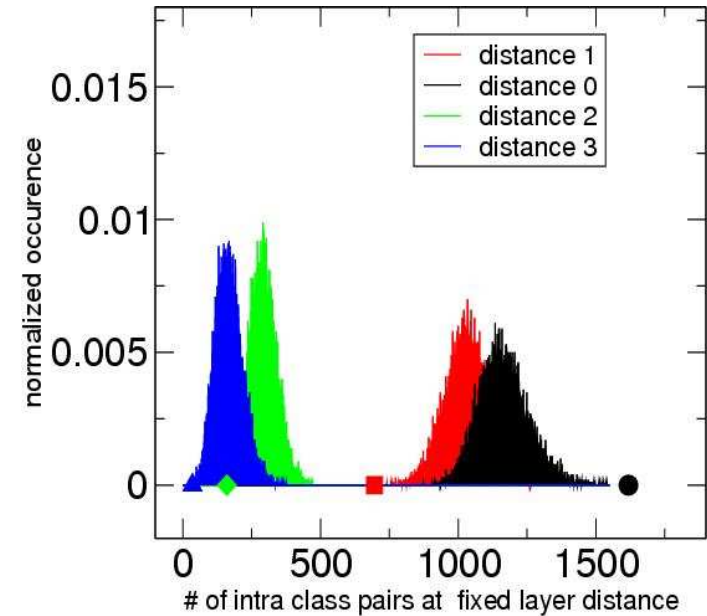
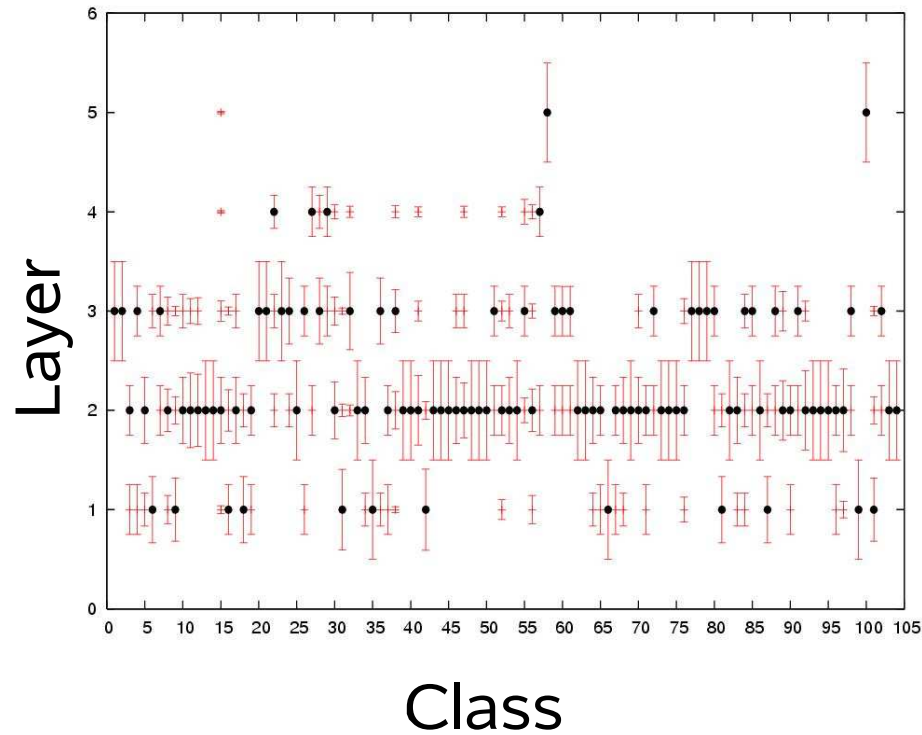


**FIS-like**

# TF Family Expansion by Duplication



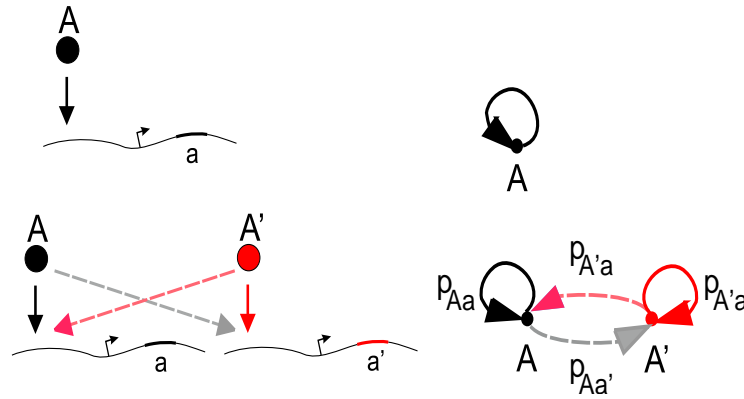
# Duplication and layers



Classes of nodes tend to occupy the same layer: “lateral” growth.

Shuffling of links between TFs exists but preserves layer structure.

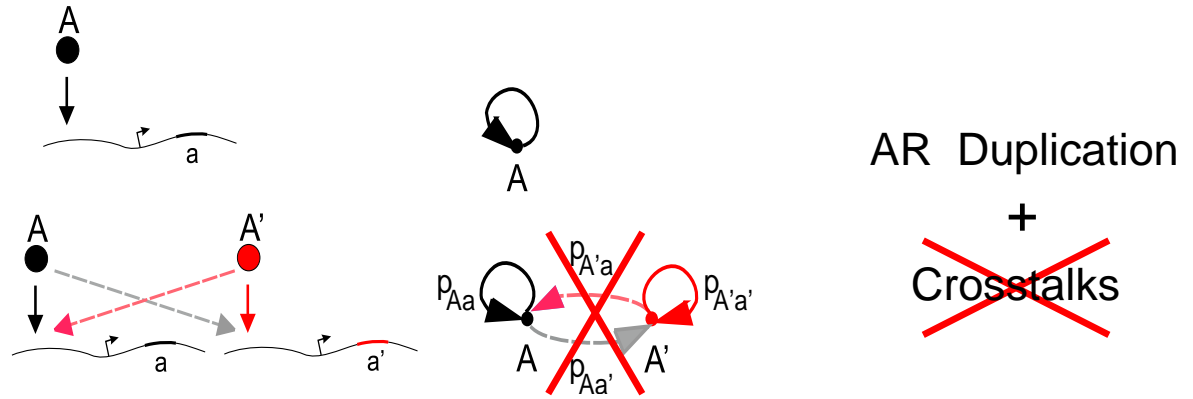
# Evolutionary Decoupling of Duplicated ARs



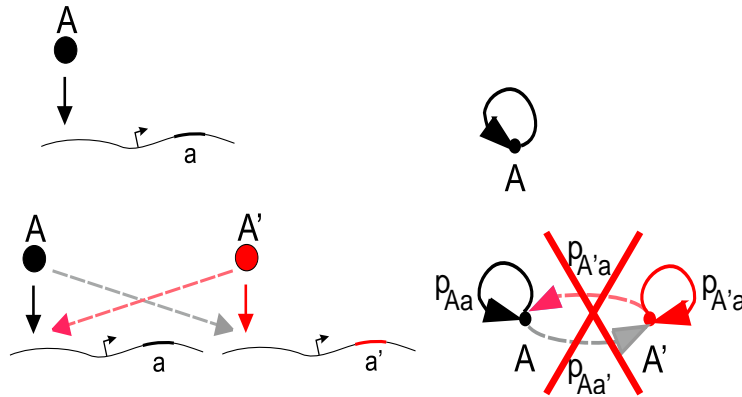
AR Duplication  
+  
Crosstalks



# Evolutionary Decoupling of Duplicated ARs



# Evolutionary Decoupling of Duplicated ARs

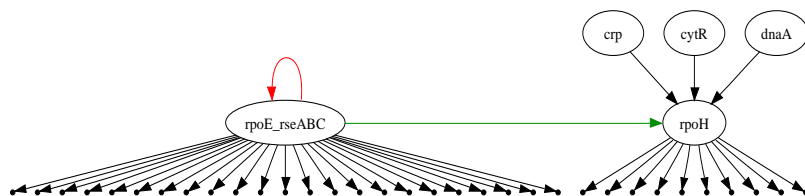
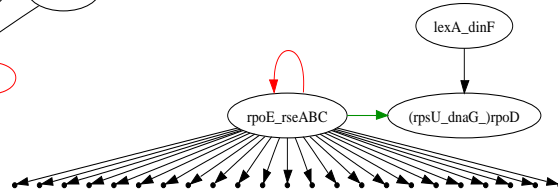
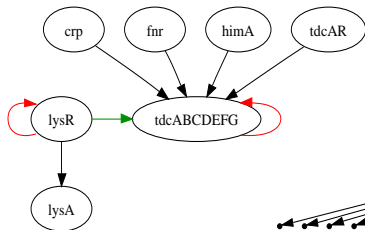
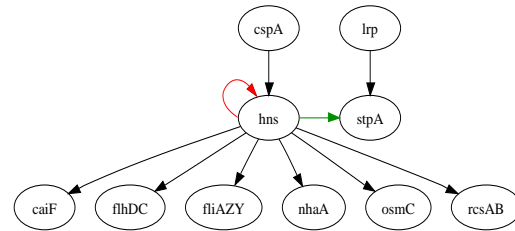
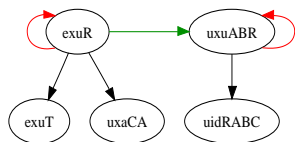


AR Duplication

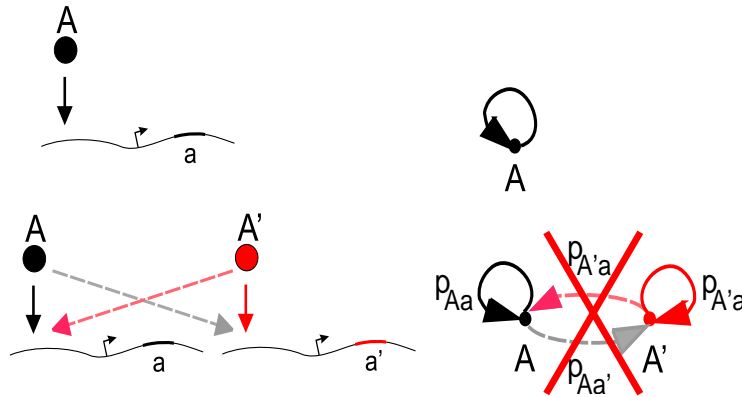
+

~~Crosstalks~~

A few partial decoupling

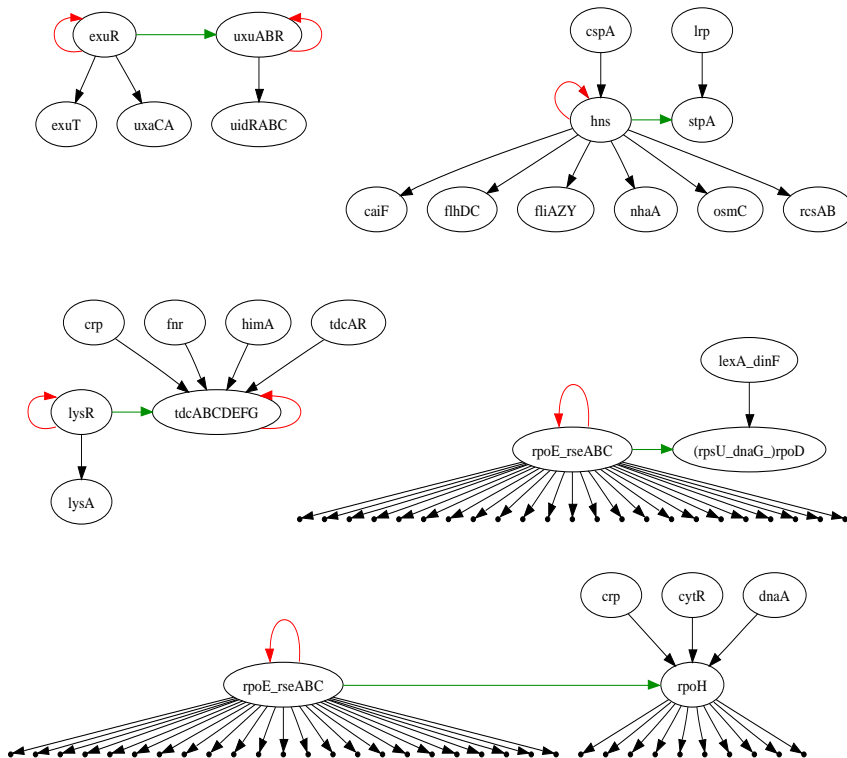


# Evolutionary Decoupling of Duplicated ARs

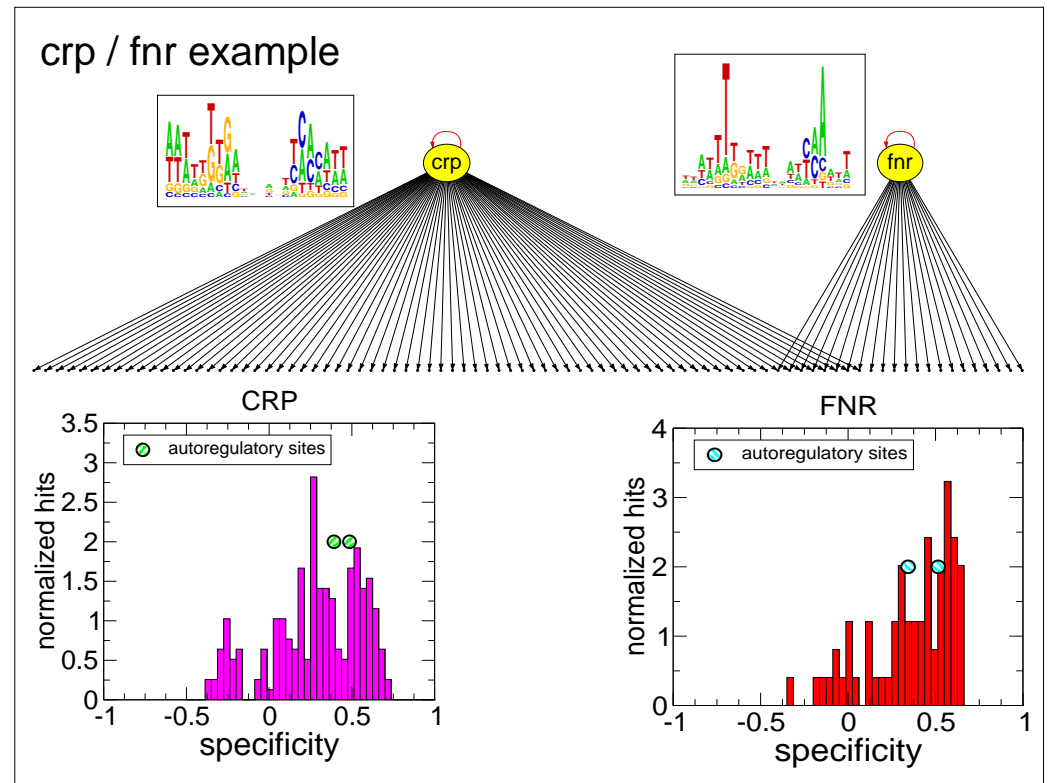


AR Duplication  
+  
~~Crosstalks~~

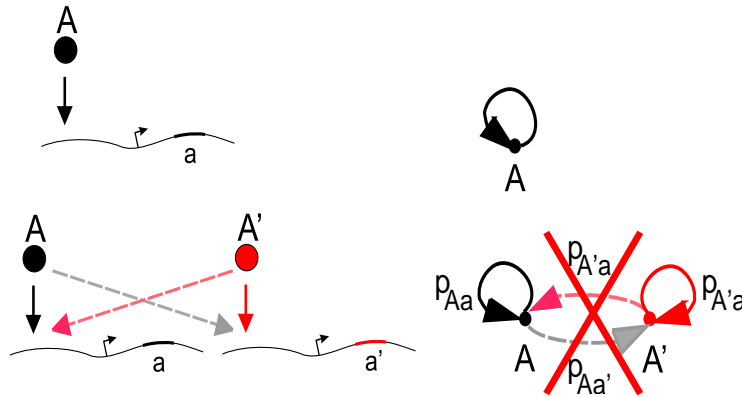
A few partial decoupling



Most fully decoupled



# Evolutionary Decoupling of Duplicated ARs

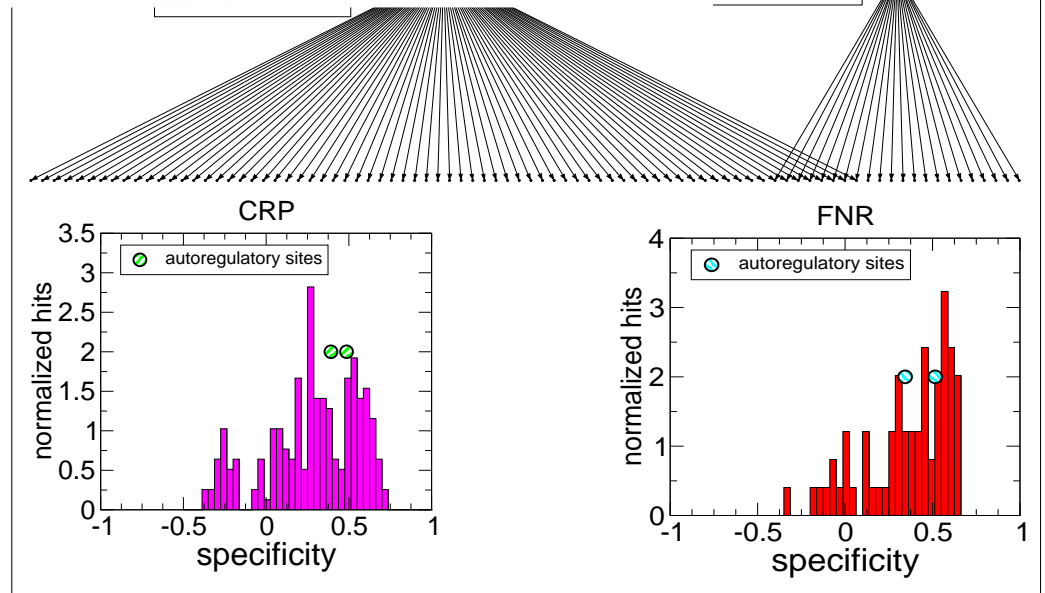
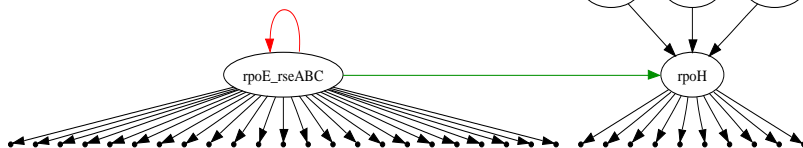
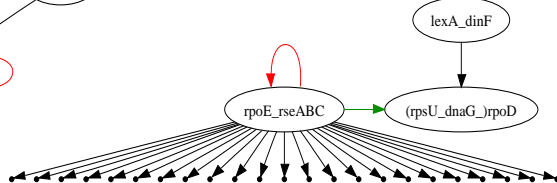
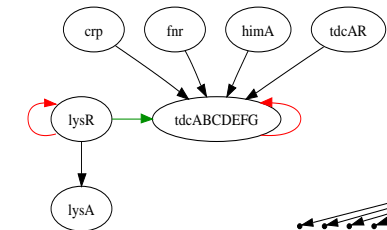
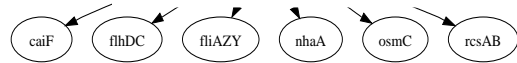
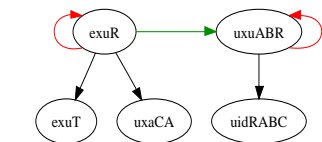


AR Duplication  
+  
~~Crosstalks~~

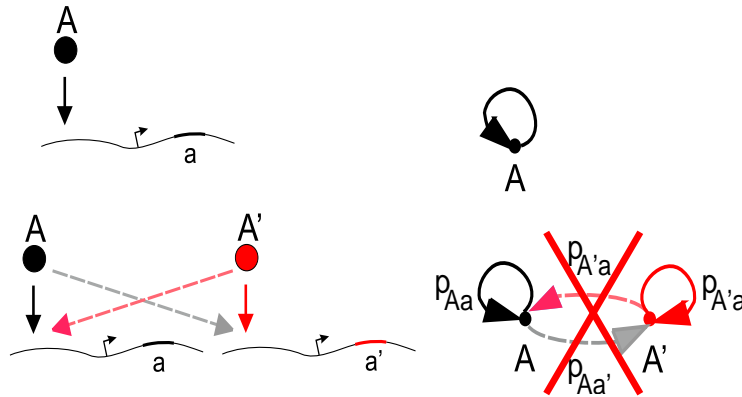
A few partia

# WHY?

ed



# Evolutionary Decoupling of Duplicated ARs



AR Duplication  
+  
~~Crosstalks~~

A few partia

# HOW?

ed

