Neutral evolution and the acceleration of the molecular clock

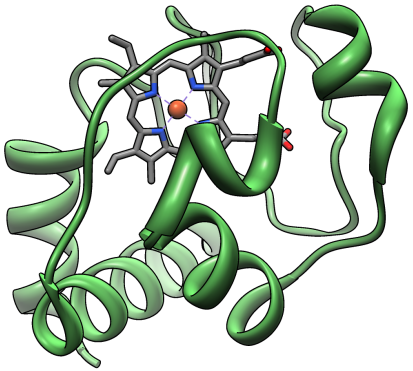
José A. Cuesta & Susanna Manrubia

Grupo Interdisciplinar de Sistemas Complejos

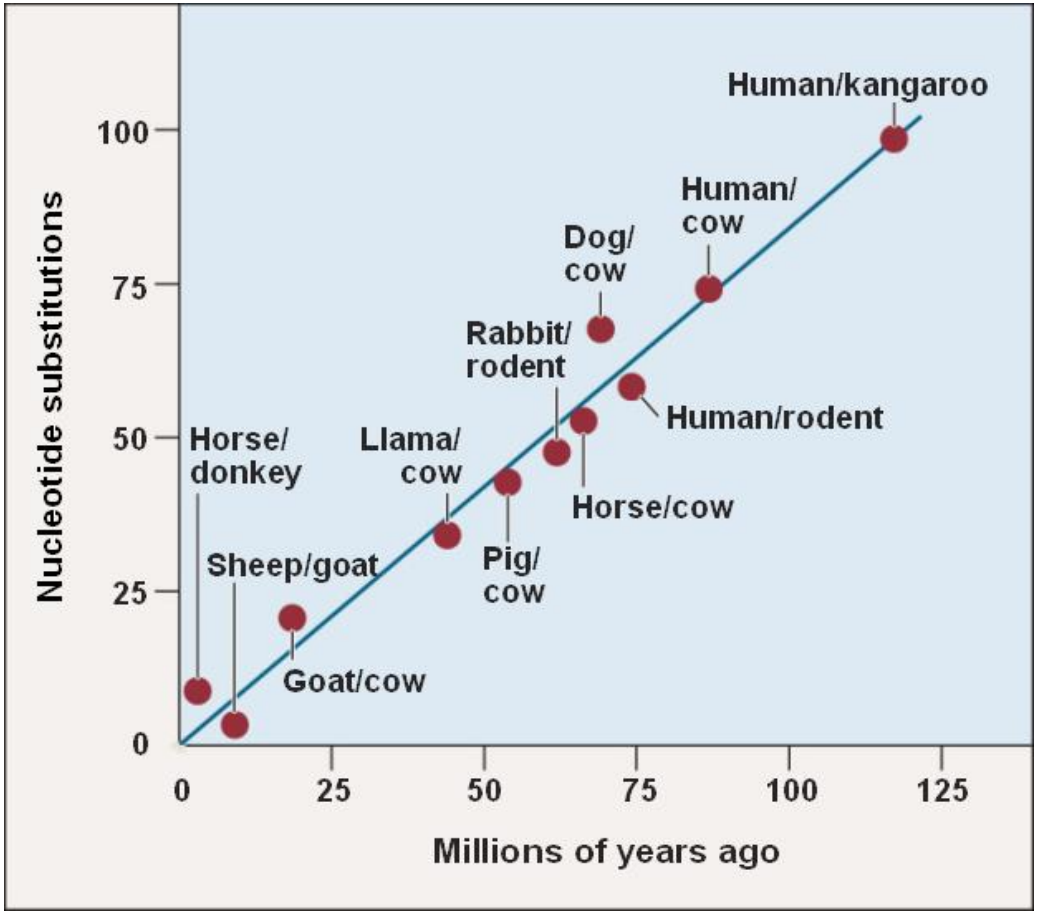
math UC3M & CNB-CSIC

gisc

empirical evidence of a molecular clock

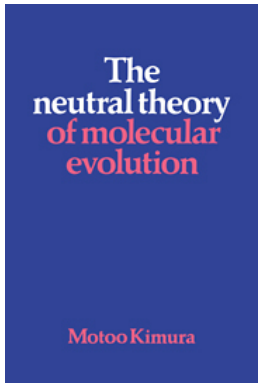


cytochrome c gene



Zuckerkandl & Pauling, *J. Teor. Biol.* (1965)

neutralist explanation

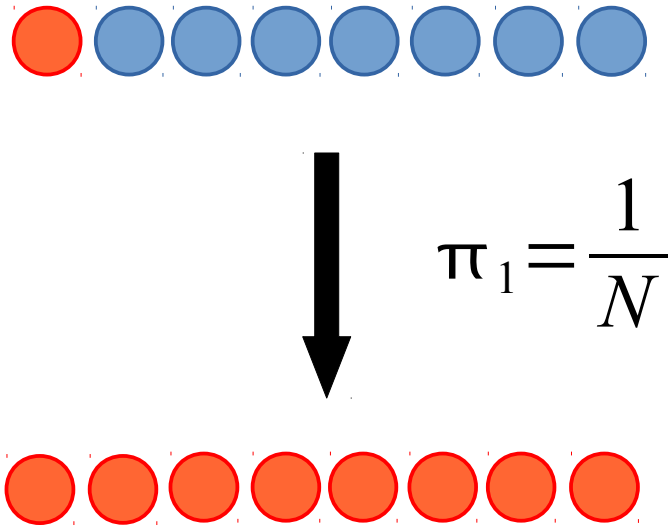


(1983)

Motoo Kimura
(1924-1994)

genetic drift

neutralist explanation



$$\pi_1 = \frac{1}{N}$$

population size

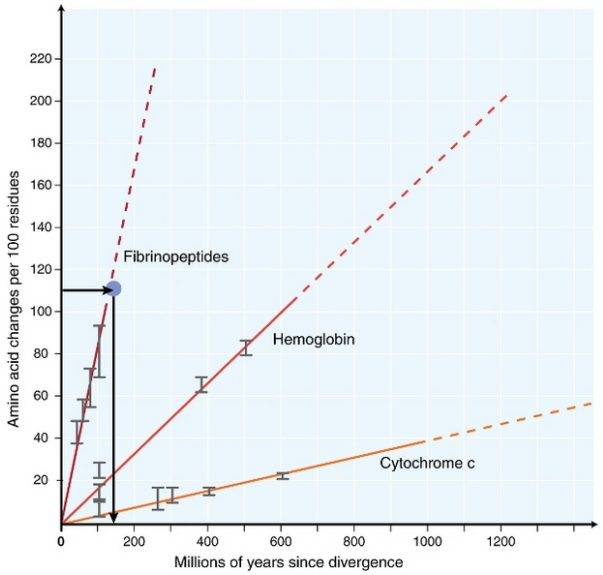
μN mutants/generation x $1/N$ = μ mutations/generation go to fixation

neutral mutation rate

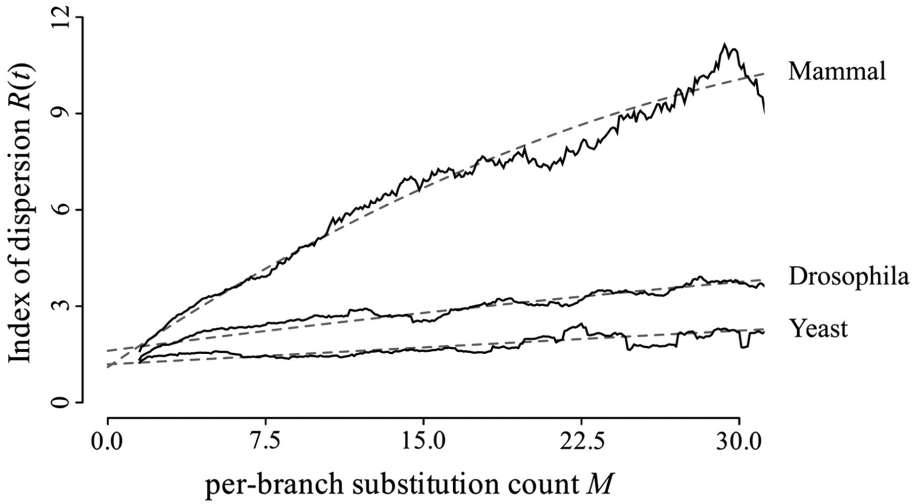
fixation probability

irregularities of the molecular clock

rate differences across lineages



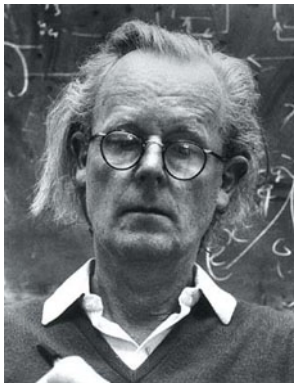
overdispersion



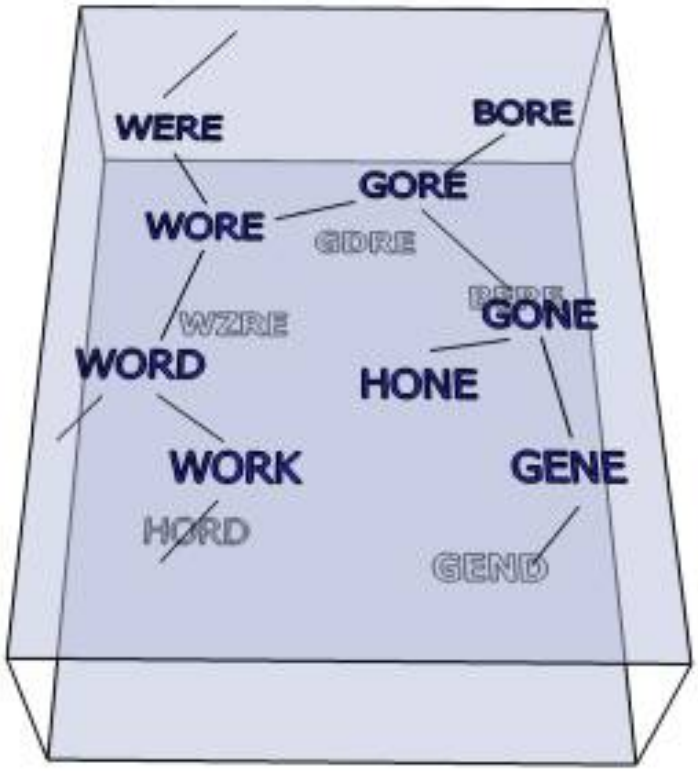
$$R(t) = \frac{\sigma^2}{\mu} = 1$$

Poisson process

neutral networks



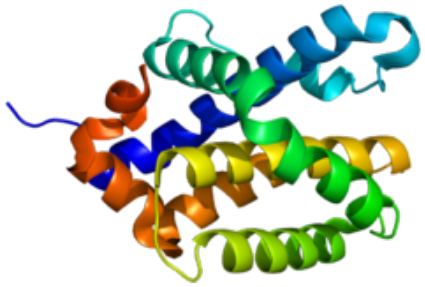
John Maynard Smith
(1920-2004)



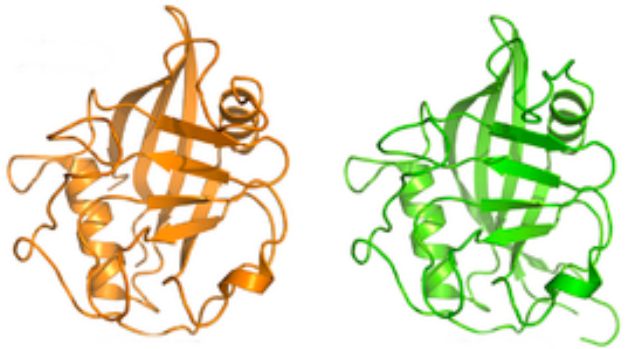
Maynard Smith, “Natural selection and the concept of a protein space”, *Nature* (1970)

redundancy

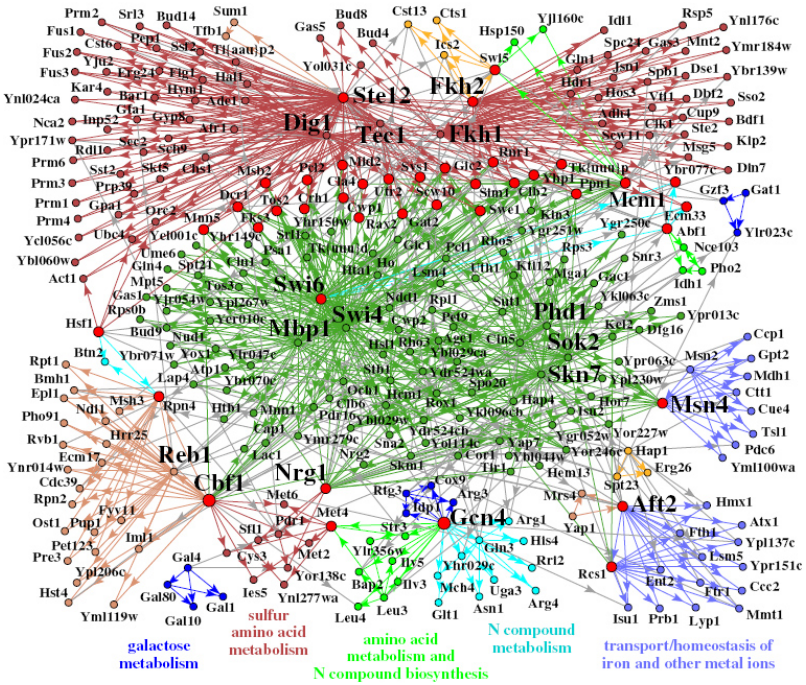
synonimous mutations



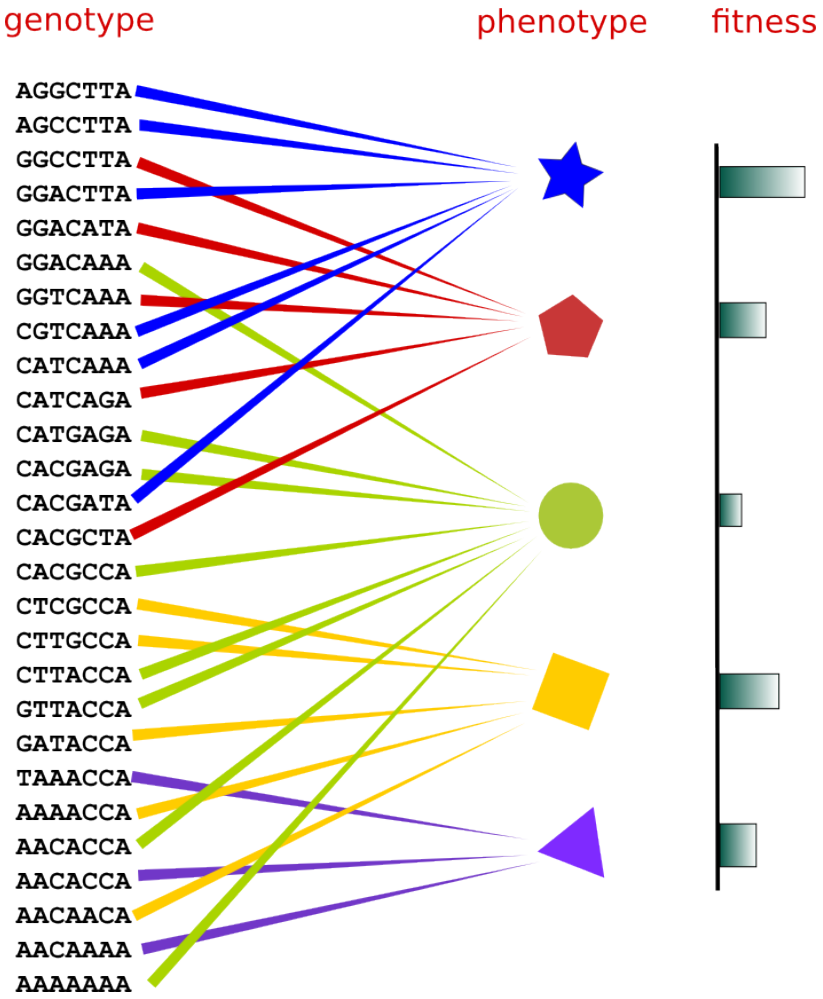
irrelevant structural changes

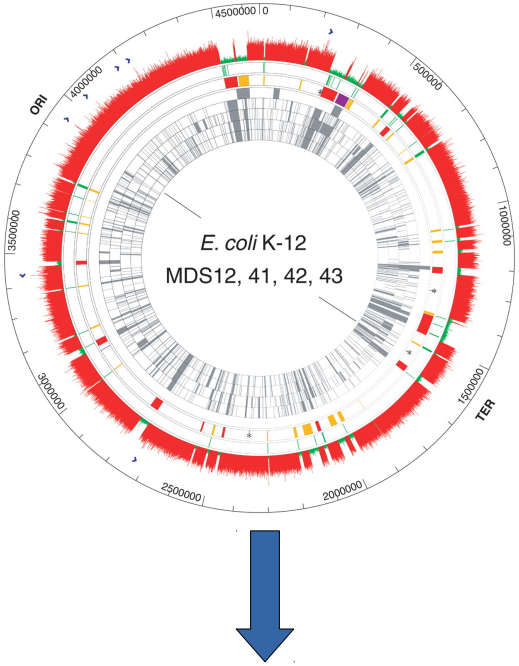


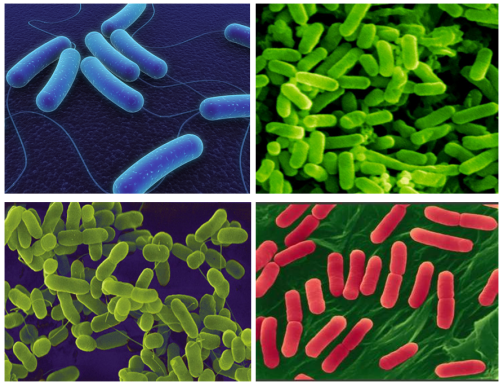
gene knock-out



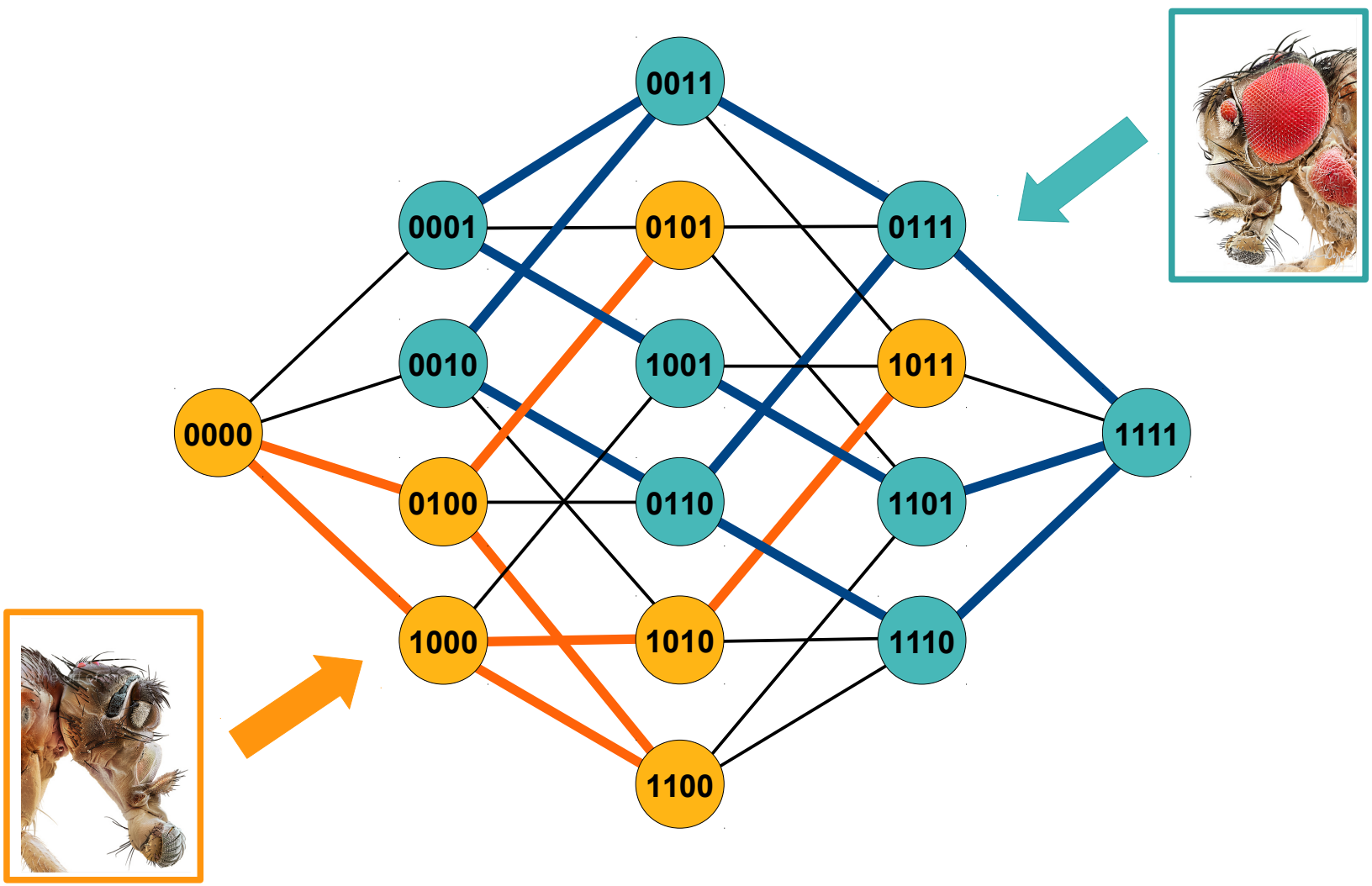
genotype-phenotype mapping





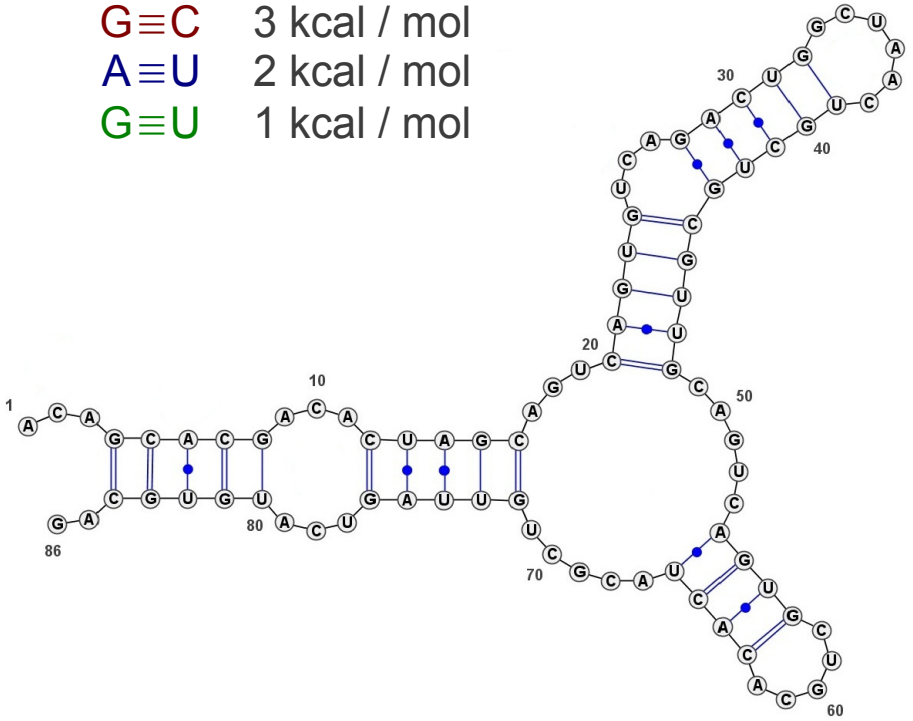


genotype networks



RNA secondary structure

G≡C 3 kcal / mol
A≡U 2 kcal / mol
G≡U 1 kcal / mol



mean number of sequences of length *n* folding into the same secondary structure:

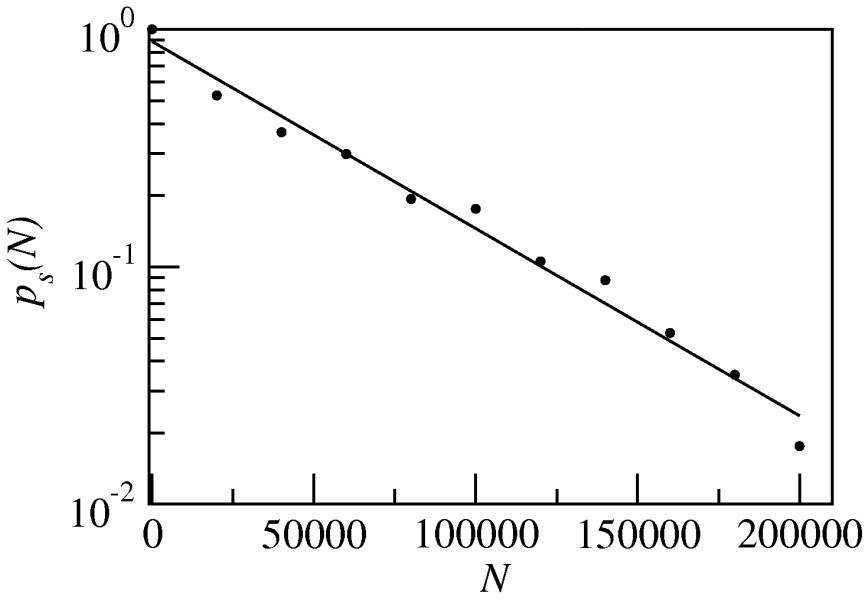
$$\sim 0.6735\, n^{3/2} (2.1635)^n$$

Fontana, *BioEssays* **24**, 1164 (2002)

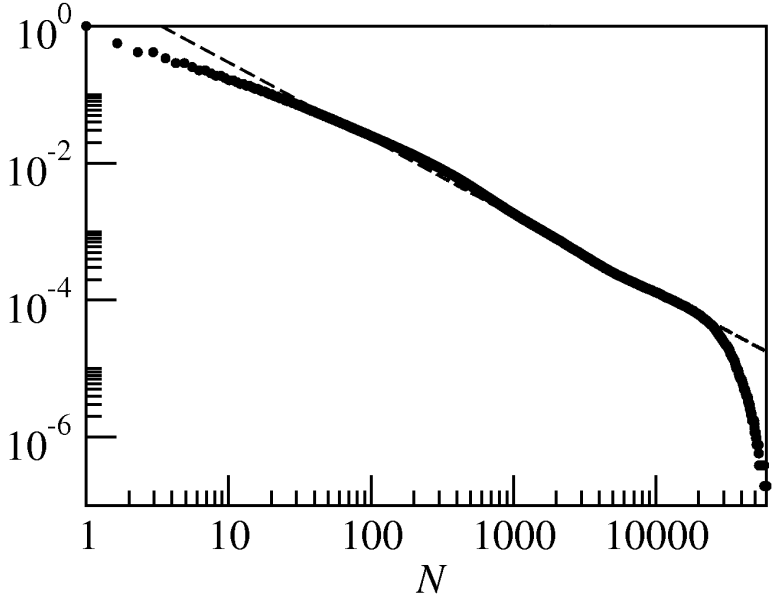
```
1          10         20         30         40         50         60         70         80       86
ACAGCACGACACUAGCAGUCAGUGUCAGACUGGCUAACUGCUGCGUUGCAGUCAGUGCUGCACACUACGCUGUUAGUCAUGUGCAG
...(((((...(((((...(((((...(((((......)))))))))).....(((.....)))).....)))))...)))))..
```

RNA secondary structure

network-size distribution



$L = 12$

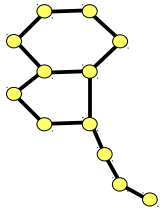


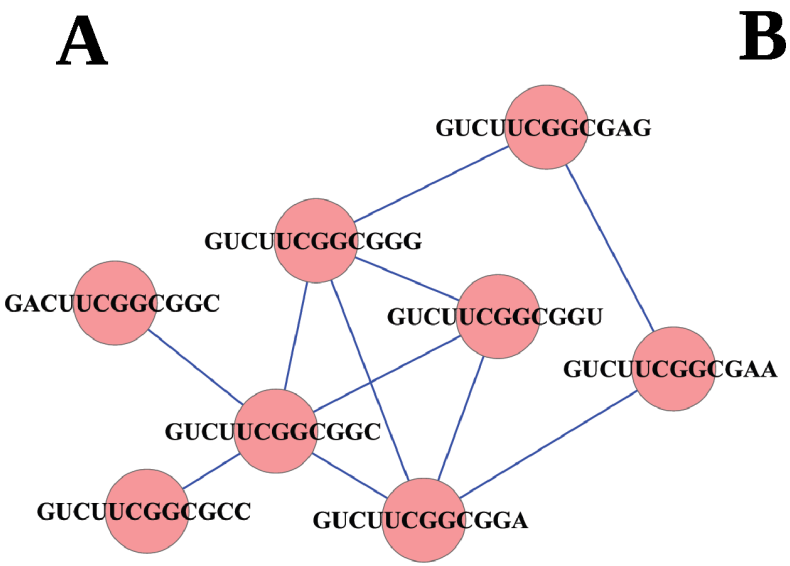
$L = 35$

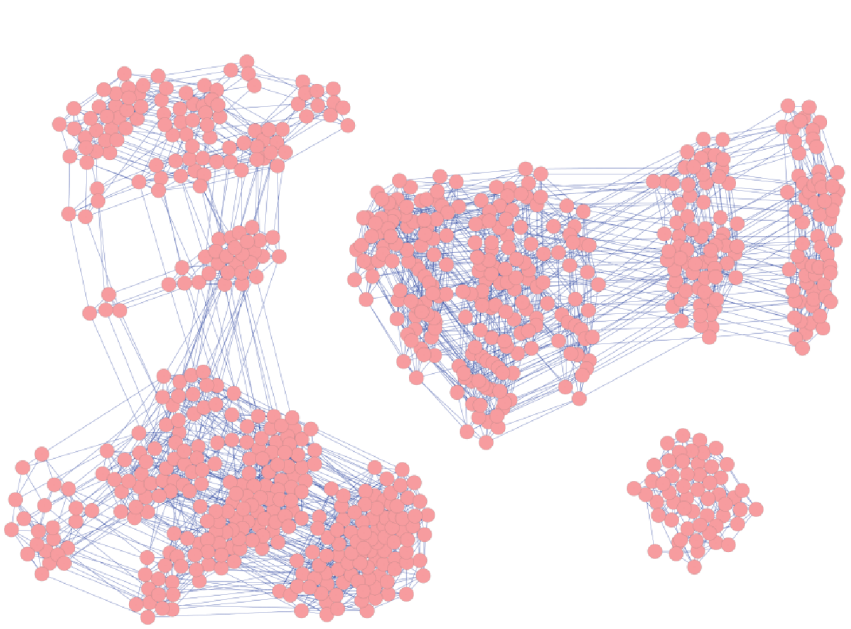
Aguirre et al., *PLoS ONE* **6**, e26324 (2011)

RNA secondary structure

networks topology ($L = 12$)



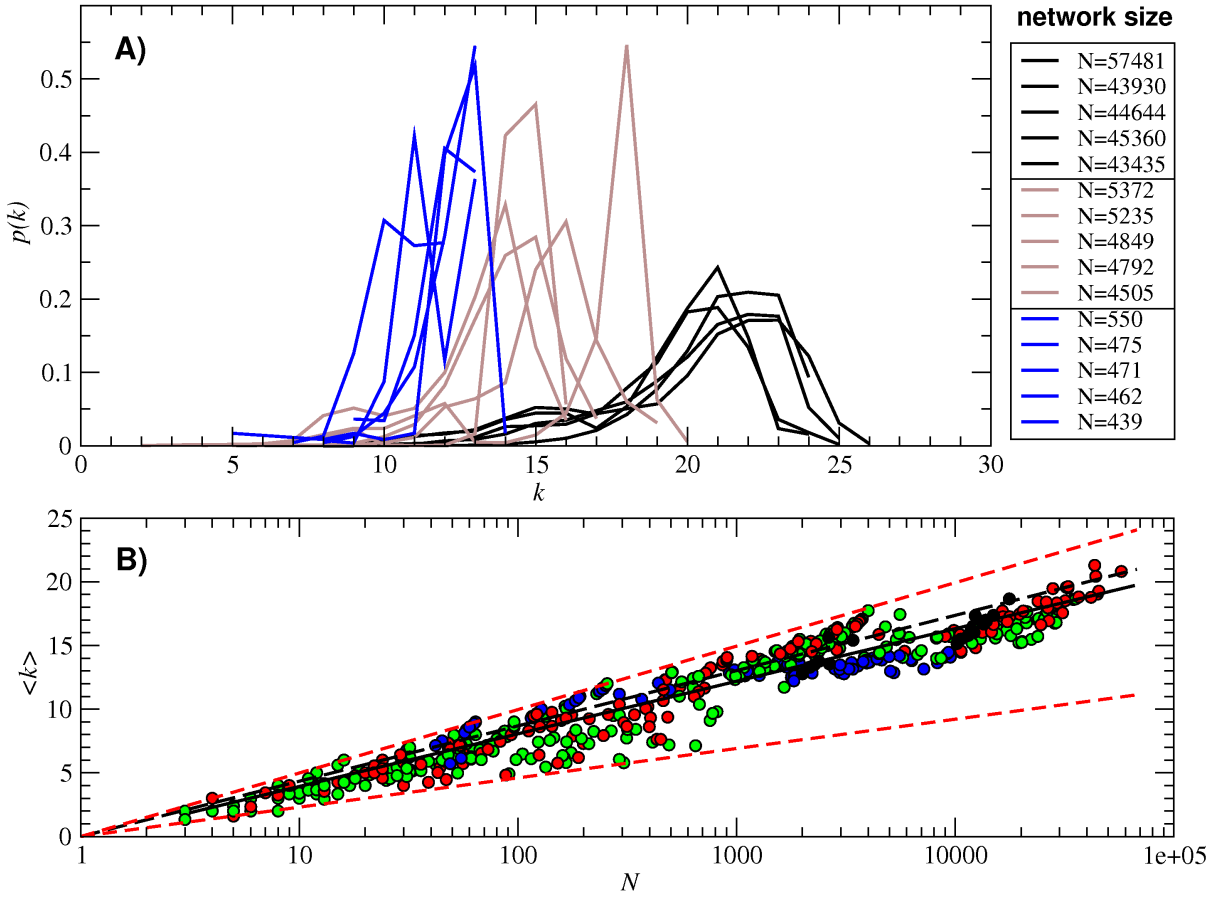




Aguirre et al., *PLoS ONE* **6**, e26324 (2011)

RNA secondary structure

networks topology ($L = 12$)



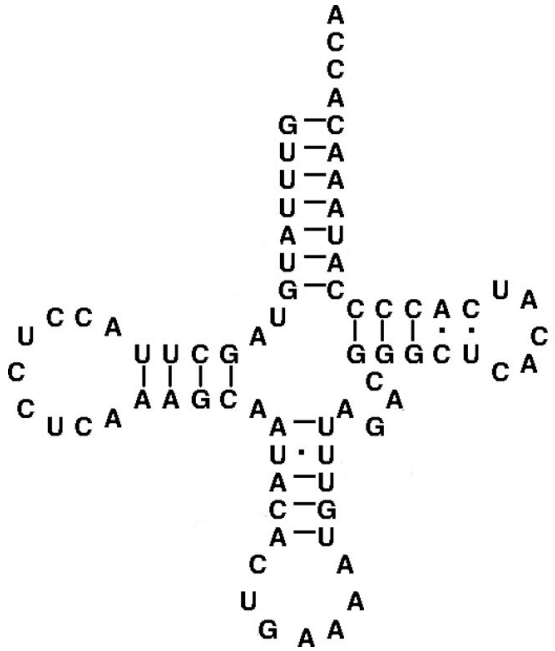
Aguirre et al., *PLoS ONE* **6**, e26324 (2011)

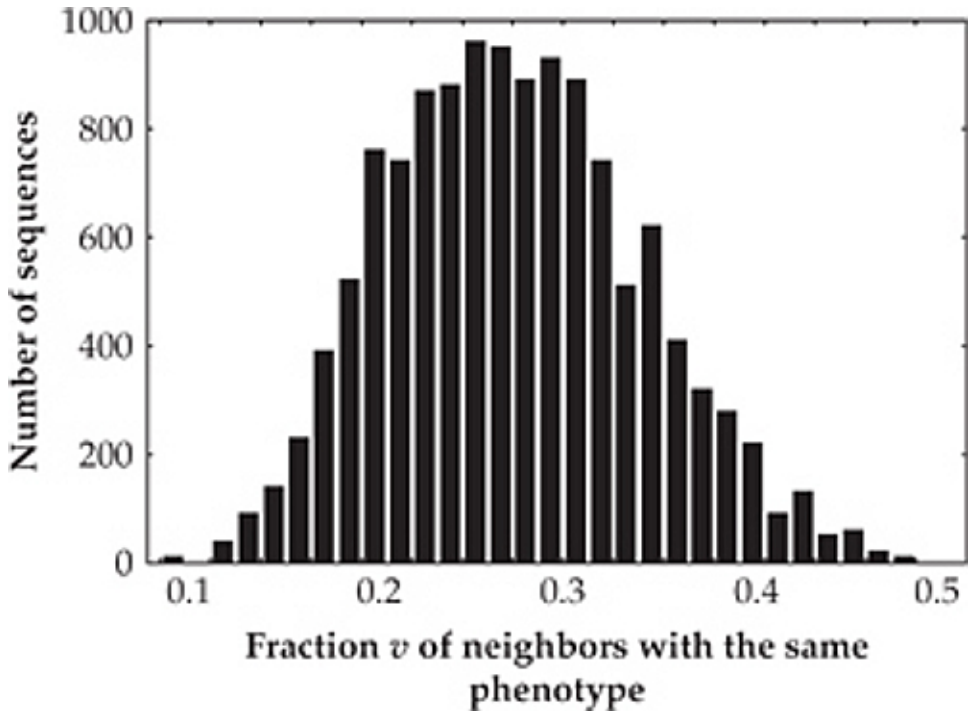
RNA secondary structure

degree distribution

tRNA$^{\text{Phe}}$

$(L=76)$



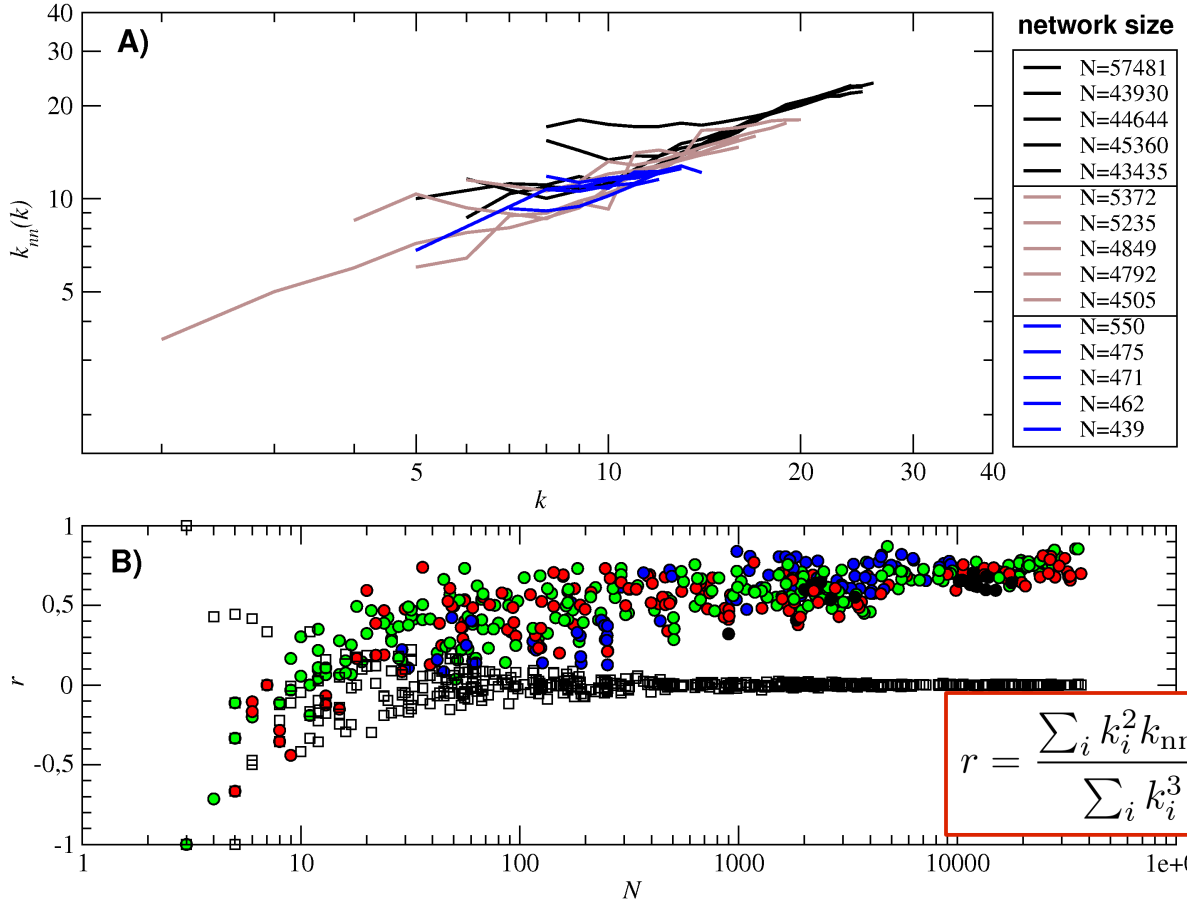


(degree distribution of the genotype network)

Sumedhaa, Martin, Wagner, *BioSystems* **90,** 475–485 (2007)

RNA secondary structure

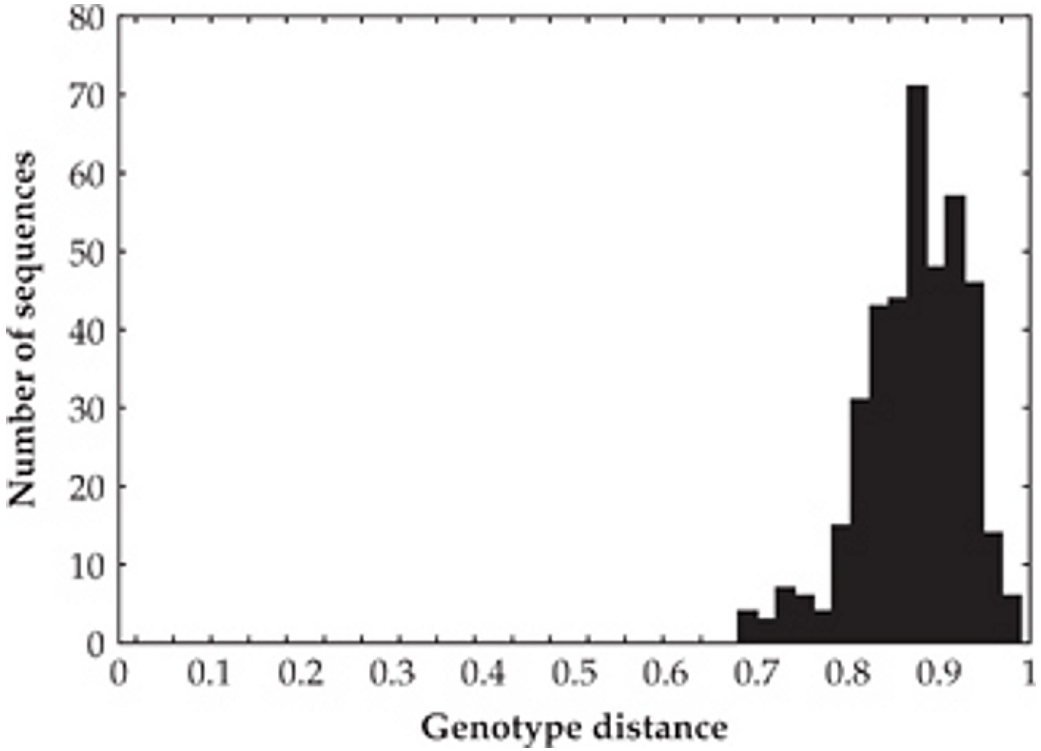
networks topology ($L = 12$)



$$r = \frac{\sum_i k_i^2 k_{\mathrm{nn},i} - (2\ell)^{-1}\left[\sum_i k_i^2\right]^2}{\sum_i k_i^3 - (2\ell)^{-1}\left[\sum_i k_i^2\right]^2}$$

Aguirre et al., *PLoS ONE* **6**, e26324 (2011)

genomic variability



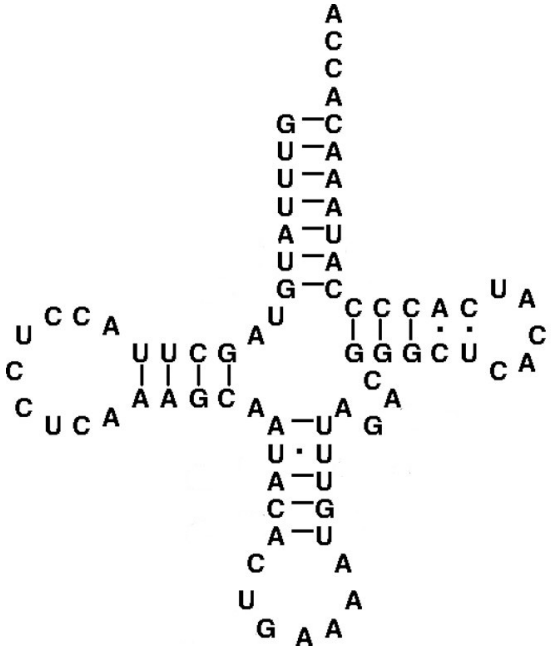
Distance between pairs of genotypes of the same “typical” RNA neutral network, 5000 mutations away of each other (L=100; aver. over 466 pairs)

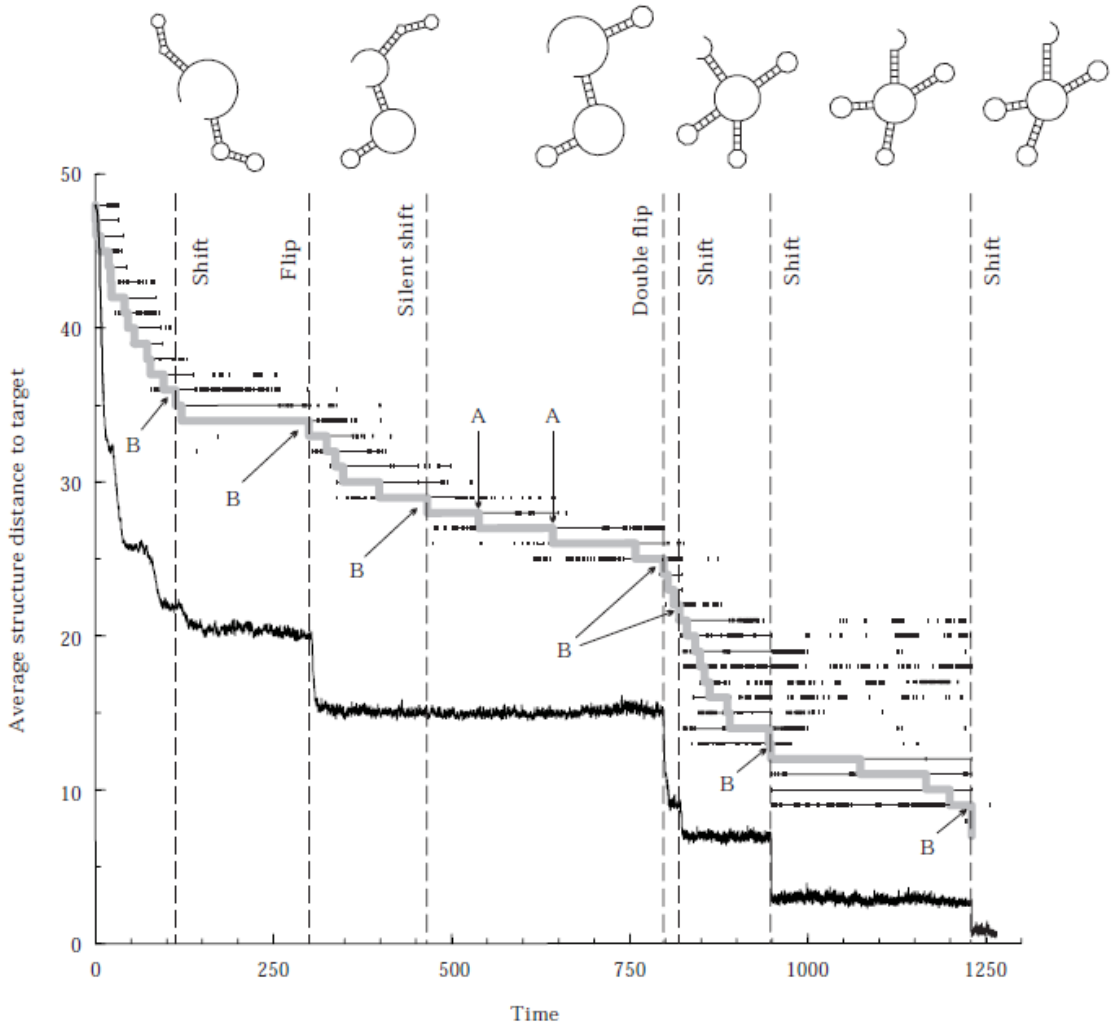
Wagner, *The origins of evolutionary innovations* (Oxford, 2011)

punctuated evolution

tRNA^{Phe}

(L=76)



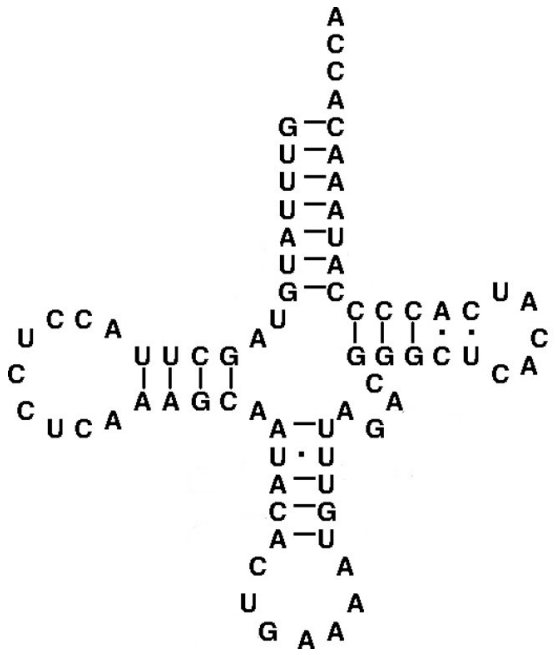


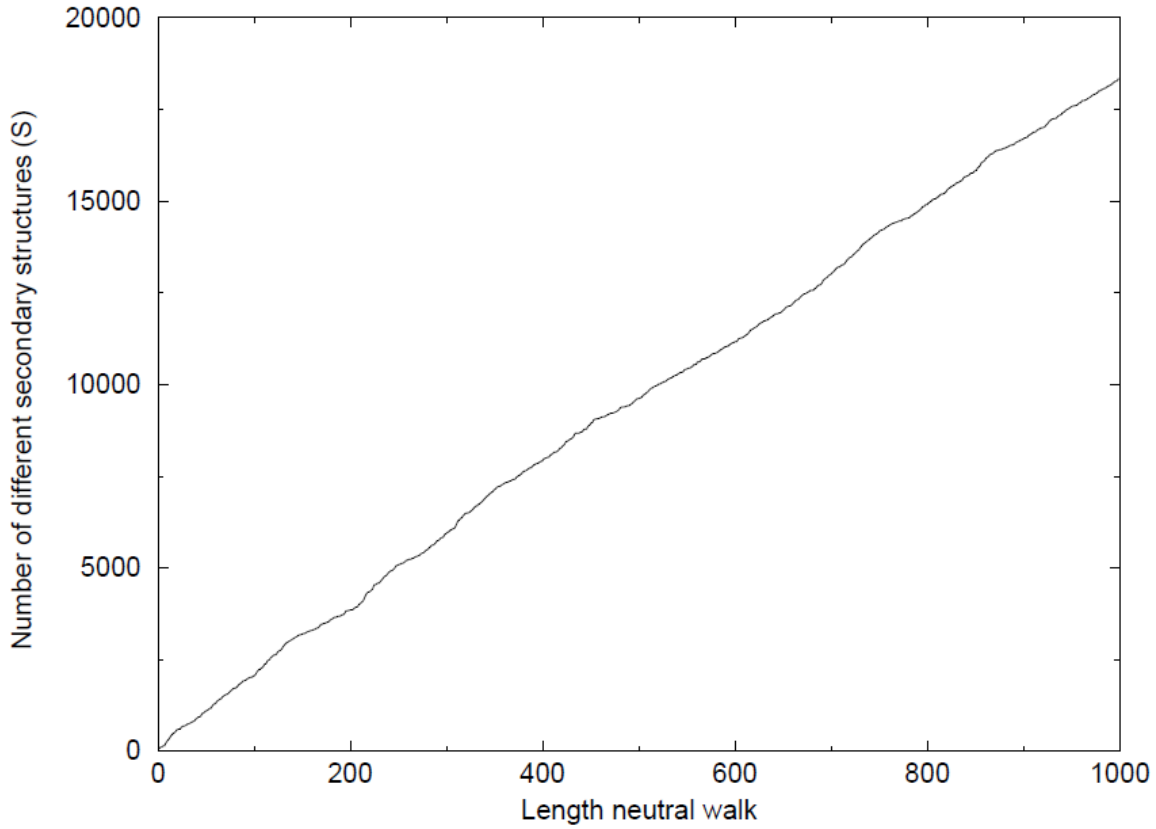
Fontana, Schuster, *J. Theor. Biol.* **194,** 495–515 (1998)

closeness of phenotypes

tRNA^{Phe}

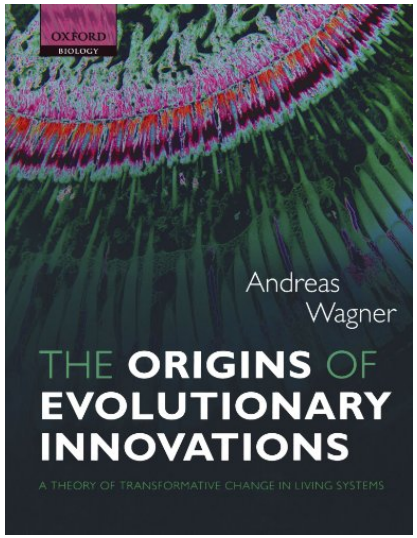
(L=76)





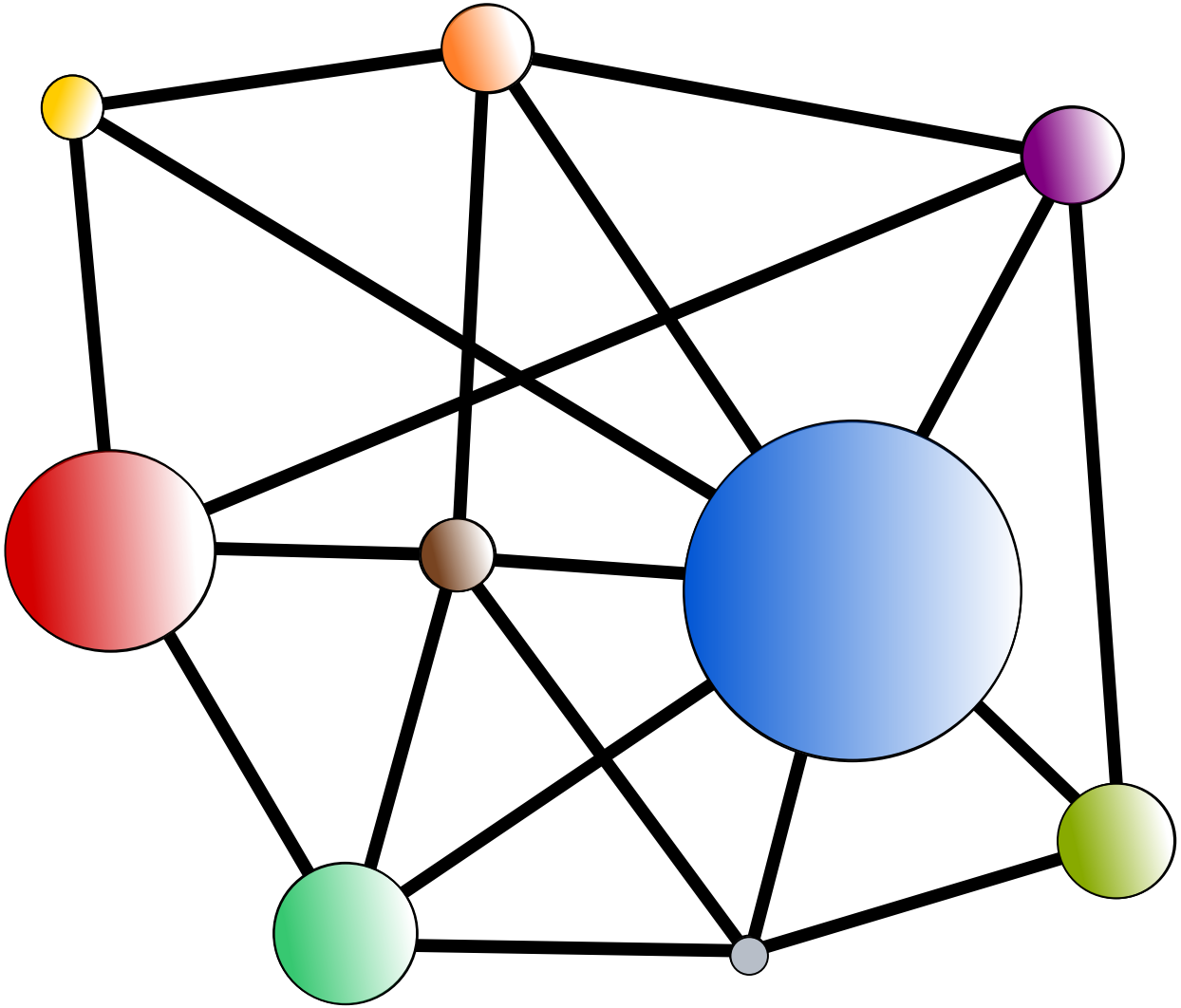
Huynen, *J. Mol. Evol.* **43,** 165–169 (1996)

genotype networks

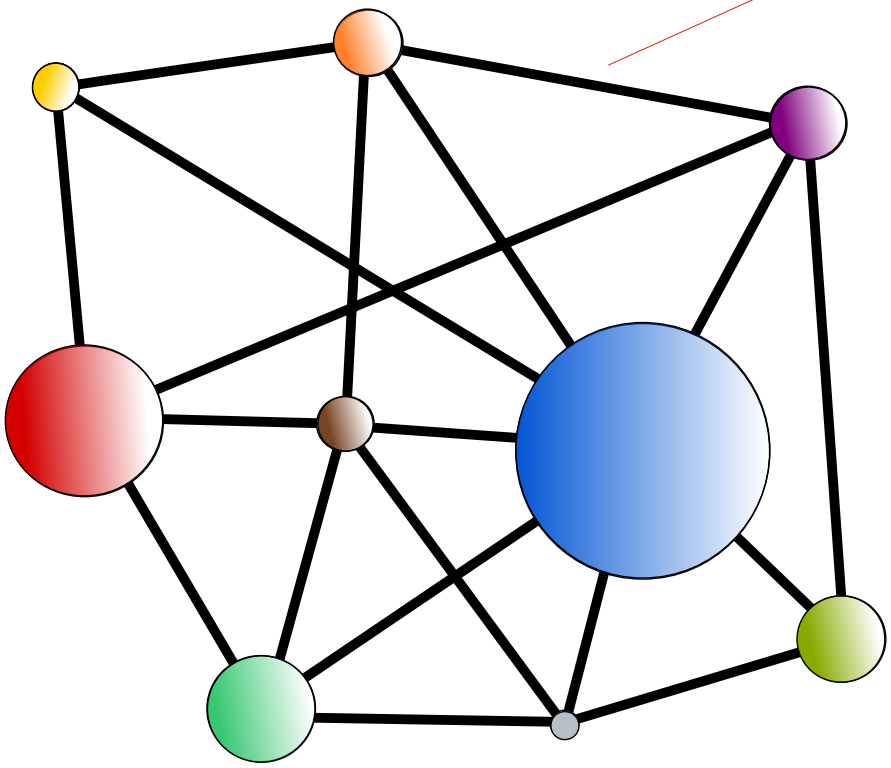


- Huge (but tiny compared to genotype space)
- Percolate genotype space
- Broad (possibly scale-free) size distribution
- Heterogeneous in degree
- Assortative
- Highly interwoven (each phenotype is neighbor of virtually any other one)

phenotype landscape



phenotypic mutations

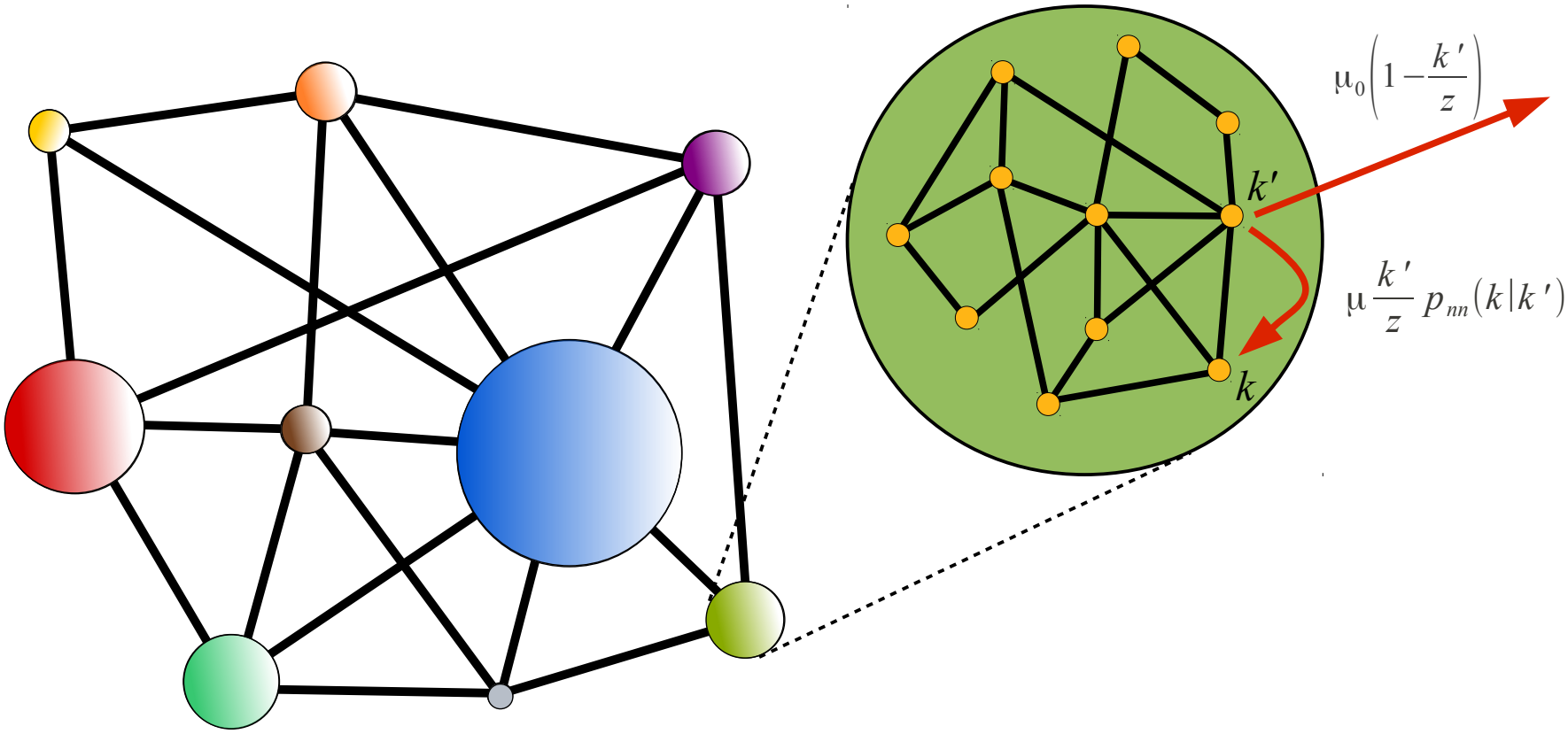


$$w_{ij} = f(d)\,\Omega(N_i)\,\Omega(N_j)$$

$$f(d) \sim d^{-\alpha} \qquad \alpha \approx 1$$

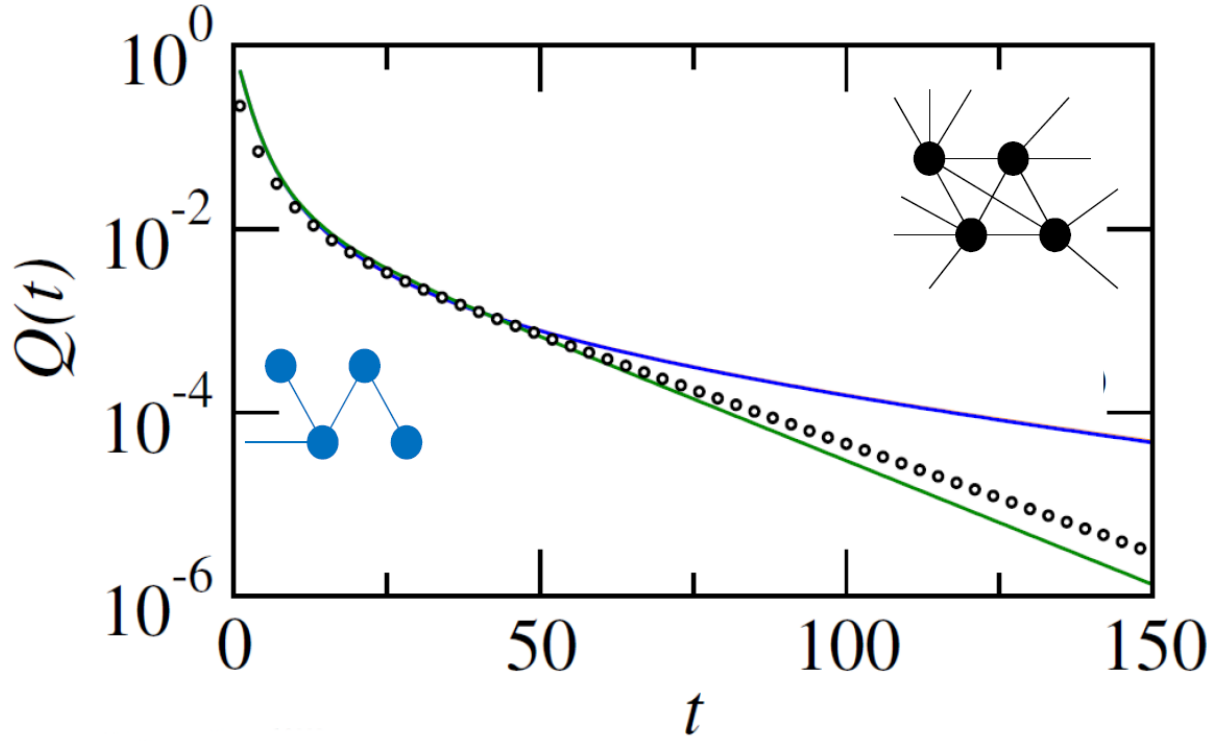
Manrubia & Sanjuan, Adv. Compl. Sys. **6**, 1250052 (2013)

intra-node dynamics



intra-node life-time distribution

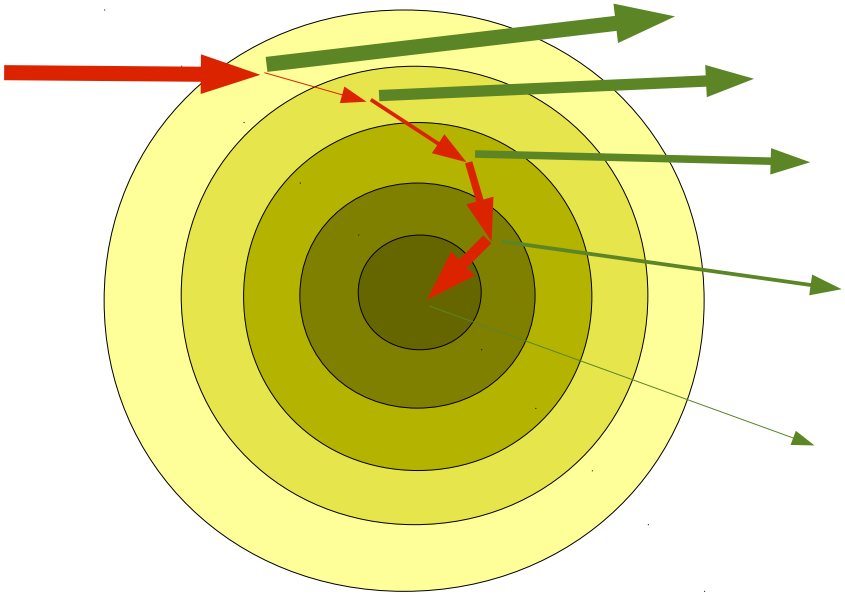
$$Q(t)=\sum_{k=0}^{z} c_k\, e^{-\alpha_k t}=e^{-\mu_{eff}(t)\,t} \qquad 0<\alpha_z<\alpha_{z-1}<\cdots<\alpha_1<\alpha_0=\mu$$

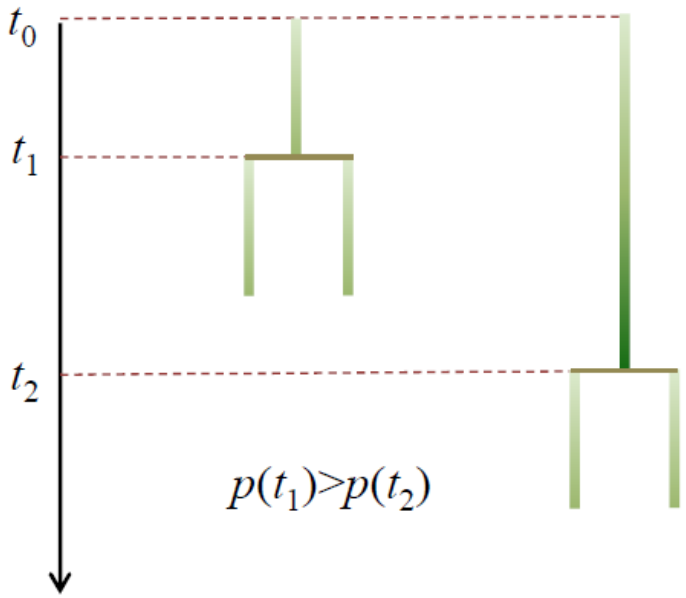


($\mu_0=\mu$, all phenotypes equally fit)

Manrubia & Cuesta, J. R.. Soc. Interface **102**, 20141010 (2015)

phenotypic entrapment





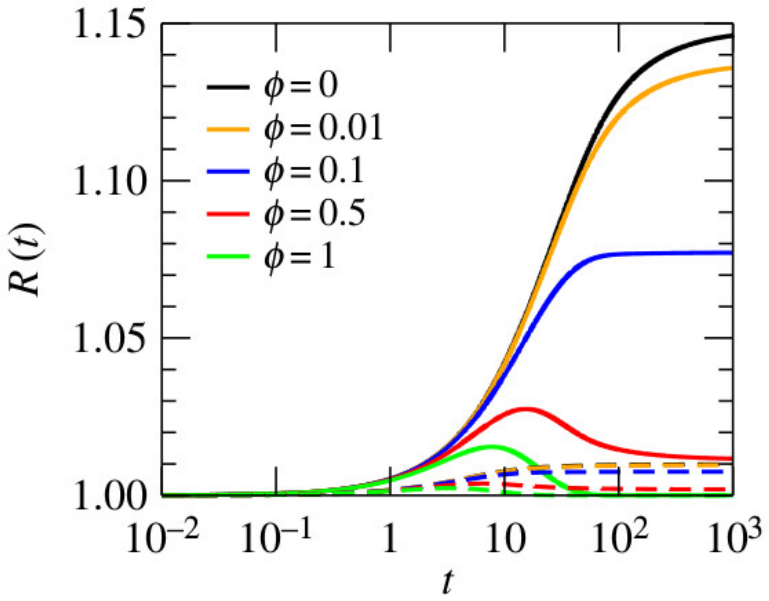
branching probability decreases with branch length

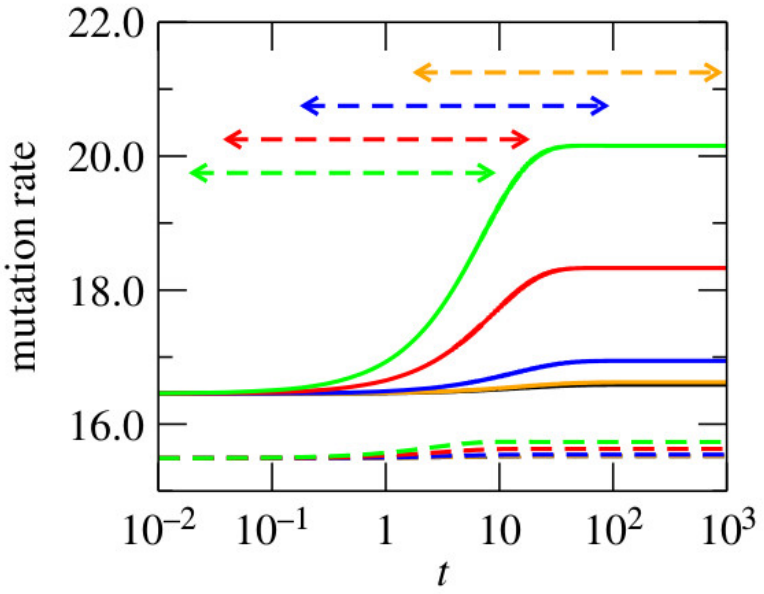
evolution in a phenotype landscape is not markovian!

Manrubia & Cuesta, J. R.. Soc. Interface **102**, 20141010 (2015)

molecular clock accelerates

RNA ($L = 12$)



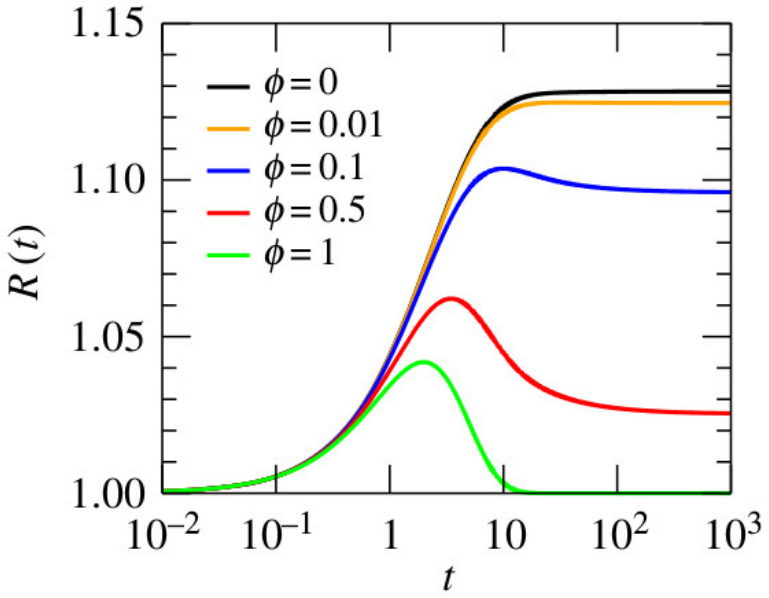


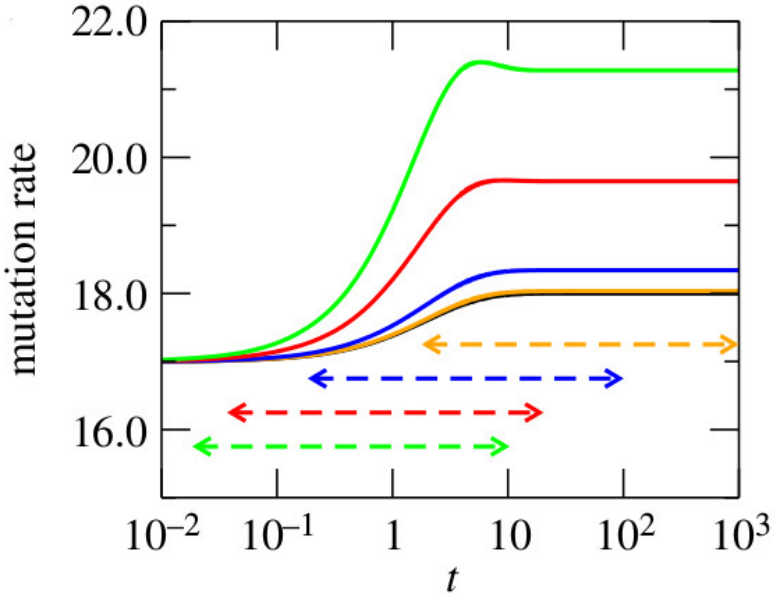
$$\Phi \equiv \frac{\text{mean fitness of neighbour phenotypes}}{\text{fitness of focal phenotype}}$$

Manrubia & Cuesta, J. R.. Soc. Interface **102**, 20141010 (2015)

molecular clock accelerates

random (Erdős-Renyi)



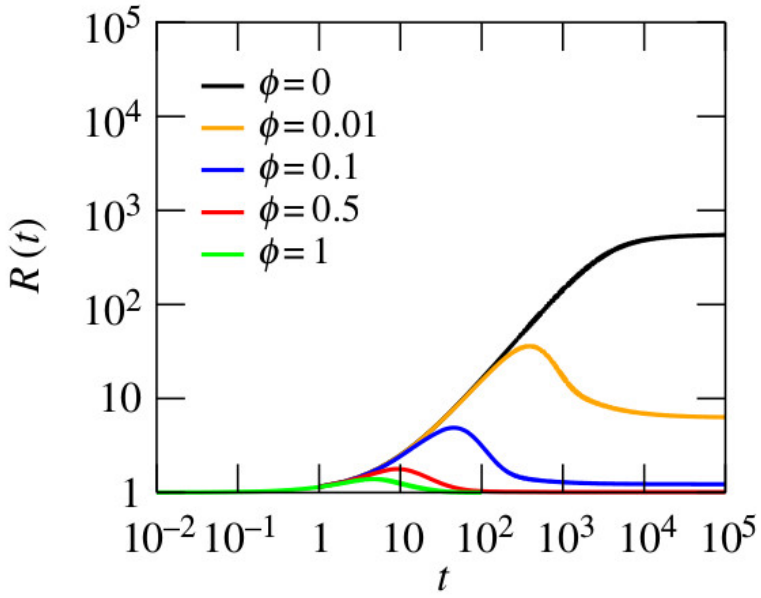


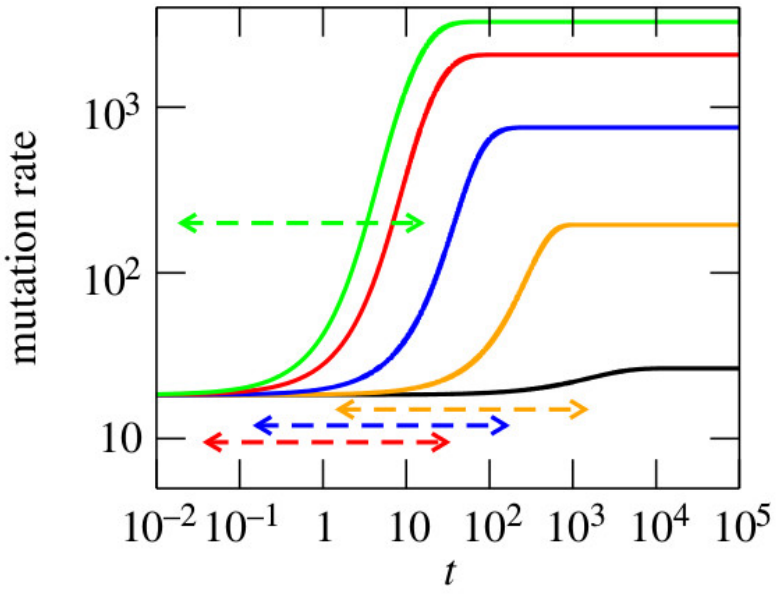
$$\Phi \equiv \frac{\text{mean fitness of neighbour phenotypes}}{\text{fitness of focal phenotype}}$$

Manrubia & Cuesta, J. R.. Soc. Interface **102**, 20141010 (2015)

molecular clock accelerates

uniform degree distribution, high assortativity ($r \approx 0.9997$)





$$\Phi \equiv \frac{\text{mean fitness of neighbour phenotypes}}{\text{fitness of focal phenotype}}$$

Manrubia & Cuesta, J. R.. Soc. Interface **102**, 20141010 (2015)

Conclusions

- Molecular clock relies on neutral mutation being a Poisson process
- Neutral mutations are described through neutral networks (high redundancy of the genotype-phenotype map)
- Network heterogeneity and assortativity cause mutational dynamics to deviate from Poisson process
- There is always overdispersion and acceleration of the molecular clock
- Long branches in phylogenetic trees tend to be longer

Take-home message

Beware of the molecular clock!