Genotype networks shape evolution

Steffen Schaper and Ard A. Louis

*Rudolf Peierls Centre for Theoretical Physics, University of Oxford,
1 Keble Road, Oxford OX1 3NP, UK*

University of Oxford — Oxford Physics

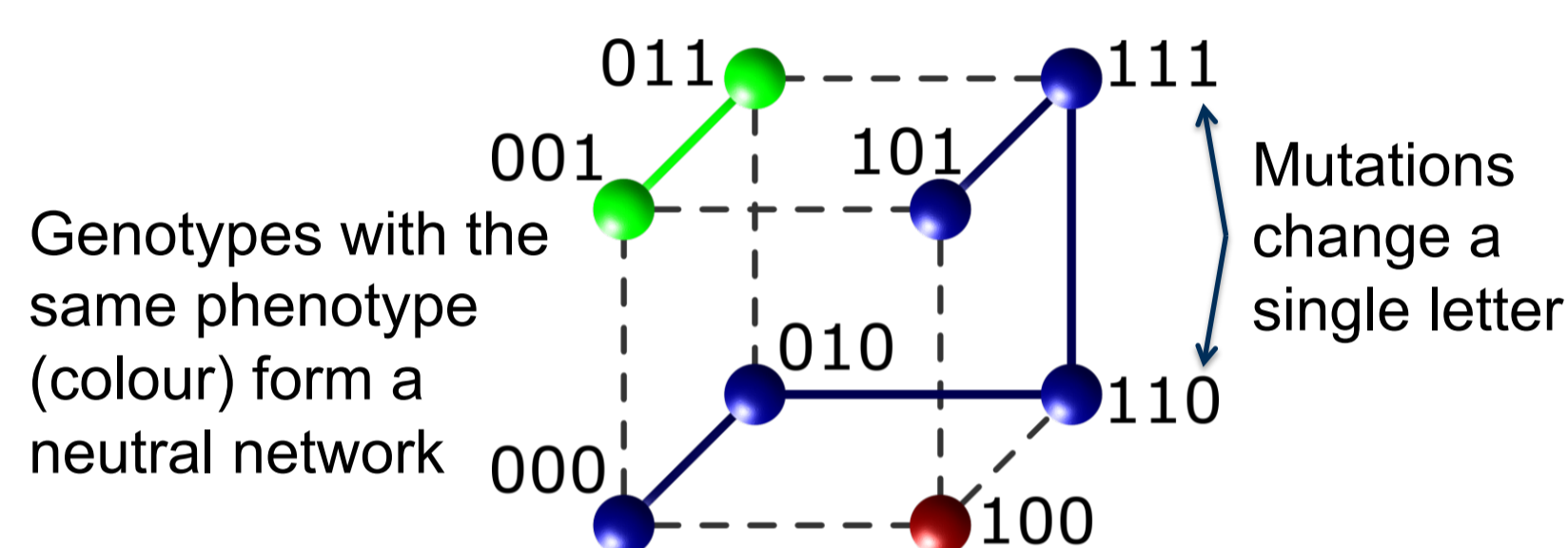**EPSRC — Engineering and Physical Sciences Research Council**

Introduction

"Natural selection may explain the survival of the fittest, but it cannot explain the arrival of the fittest."
Hugo de Vries, Species and Varieties – Their Origin by Mutation, 1904

Innovations occur in evolution when mutations change genetic information (the *genotype*) and these changes have an effect on the physical appearance (*phenotype*) and fitness of an organism.
The arrival of the fittest depends on how phenotypes arise from genotypes. This relation is characterized by the genotype-phenotype (GP) map.

Genotype networks and the exploration of genotype space

Phenotype-preserving ('neutral') mutations can change the set of accessible phenotypes and increase the potential for innovation.

Generally, genotypes are strings of length L over an alphabet of K letters ($K=4$ for DNA and RNA). There are K^L genotypes in total. Under point mutations each genotype has $(K-1)L$ neighbours. For each phenotype, the GP map induces a **neutral network** containing all genotypes that realize this phenotype.
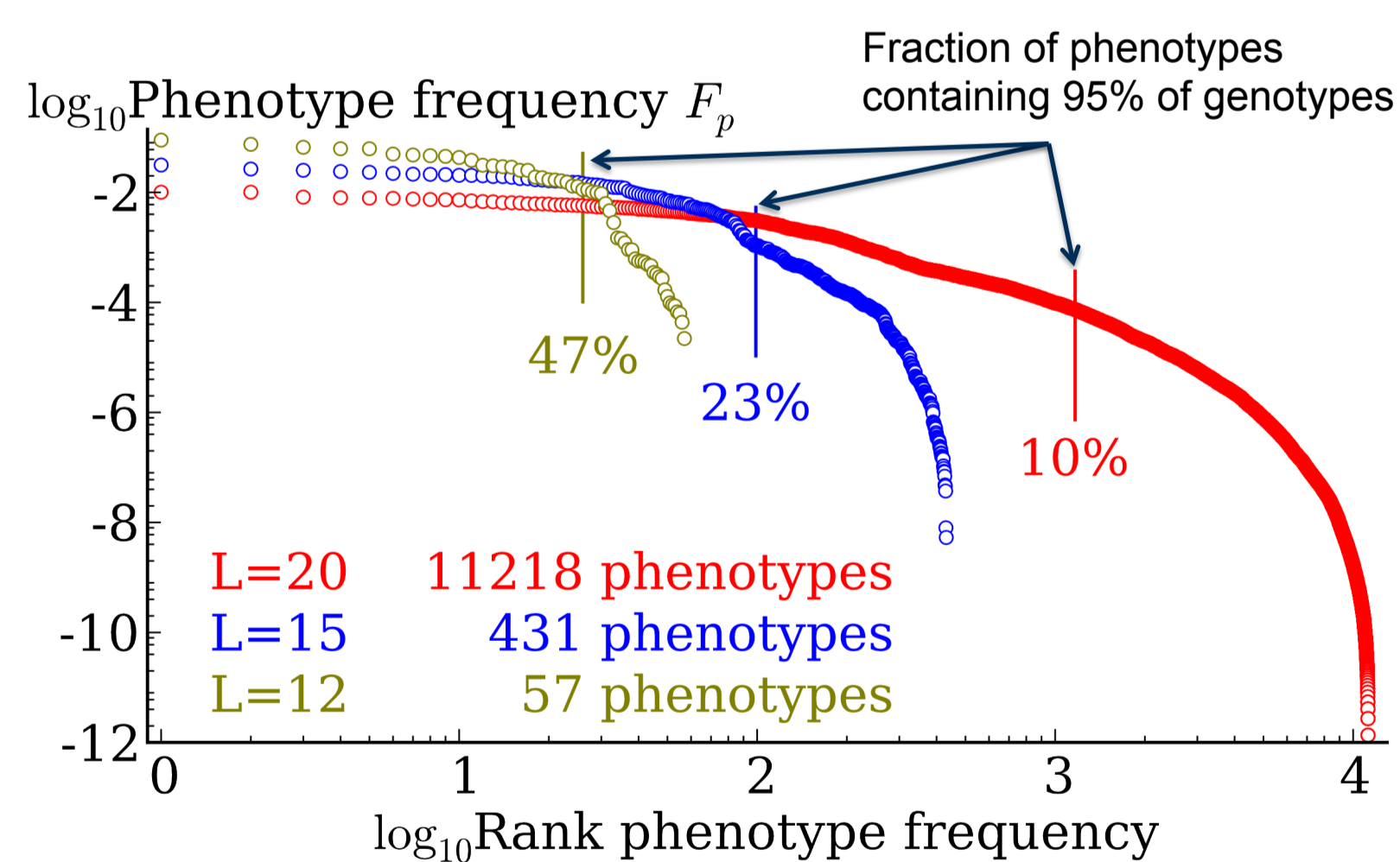The picture below illustrates the case $L=3$, $K=2$.

Genotypes with the same phenotype (colour) form a neutral network

011, 001, 101, 111, 010, 000, 110, 100

Mutations change a single letter

A neutral mutation from A to B can make different phenotypes accessible without the need to cross a fitness valley.

Neutral mutation
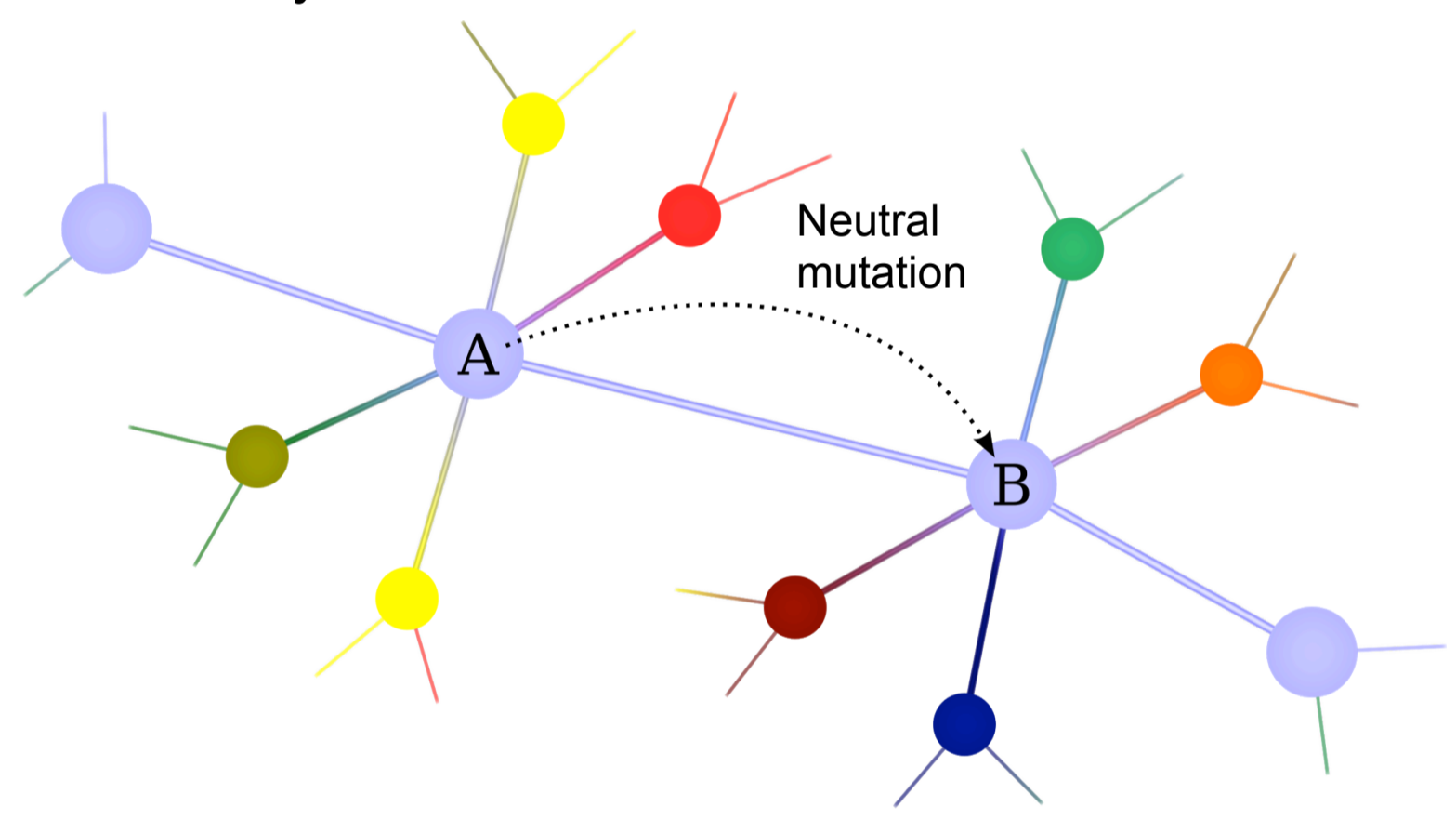
Distribution of phenotype frequencies

The majority of genotypes maps into a small fraction of phenotypes.

In many biological systems, the number of different genotypes is much greater than the number of distinct phenotypes. But the assignment of genotypes to different phenotypes is often highly biased:

$\log_{10}$ Phenotype frequency F_p

Fraction of phenotypes containing 95% of genotypes

47%   23%   10%

L=20   11218 phenotypes
L=15   431 phenotypes
L=12   57 phenotypes

$\log_{10}$ Rank phenotype frequency

The frequency F_p of a phenotype p is the probability that a random genotype maps into p. The diagram above shows phenotype frequencies of RNA secondary structures, folded using the Vienna package [1], obtained by exhaustive enumeration.
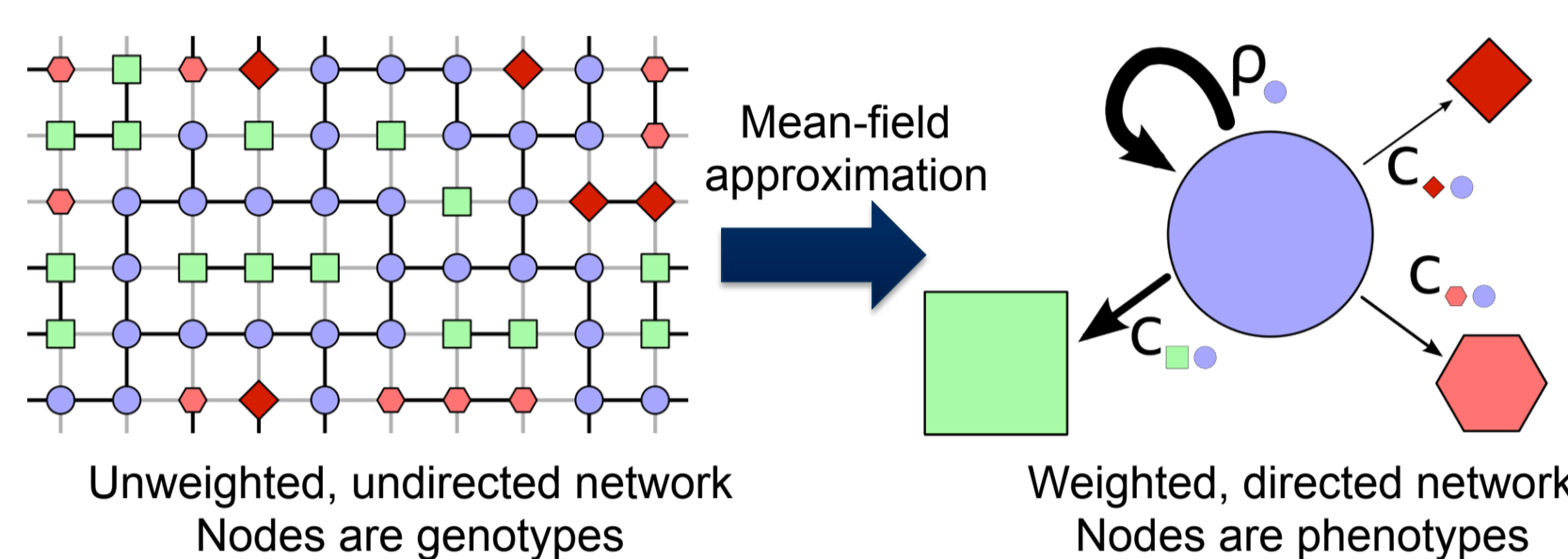
Similar bias has been observed in other systems, including models of protein folding [2], self-assembly [3] and gene regulatory networks [4].

Mean-field model of exploration

We view evolution as a two-stage process: Neutral exploration *within* a neutral network, and adaptive transitions *between* networks.

Suppose a population of N individuals has adapted to a phenotype q, when the environment suddenly changes and a different phenotype p has greater fitness. Given a mutation rate μ per base, when is p first discovered?

In general, the discovery time depends on the detailed connectivity of the neutral networks of p and q. We simplify the problem by a *mean-field* approximation that ignores local correlations:

Mean-field approximation

Unweighted, undirected network
Nodes are genotypes

Weighted, directed network
Nodes are phenotypes

The outcome of mutations depends on the *robustness* ρ_q (fraction of mutations from q going back to q) and the *weighted connectivities* c_{pq} (fraction of mutations away from q that lead to p, with normalization $\sum_{p \neq q} c_{pq} = 1$).

Analysis and simulations

Frequent phenotypes are discovered earlier and produced more often than rare ones.

If c_{pq} is large, p is always accessible and there is no need for neutral exploration. The median discovery time is then

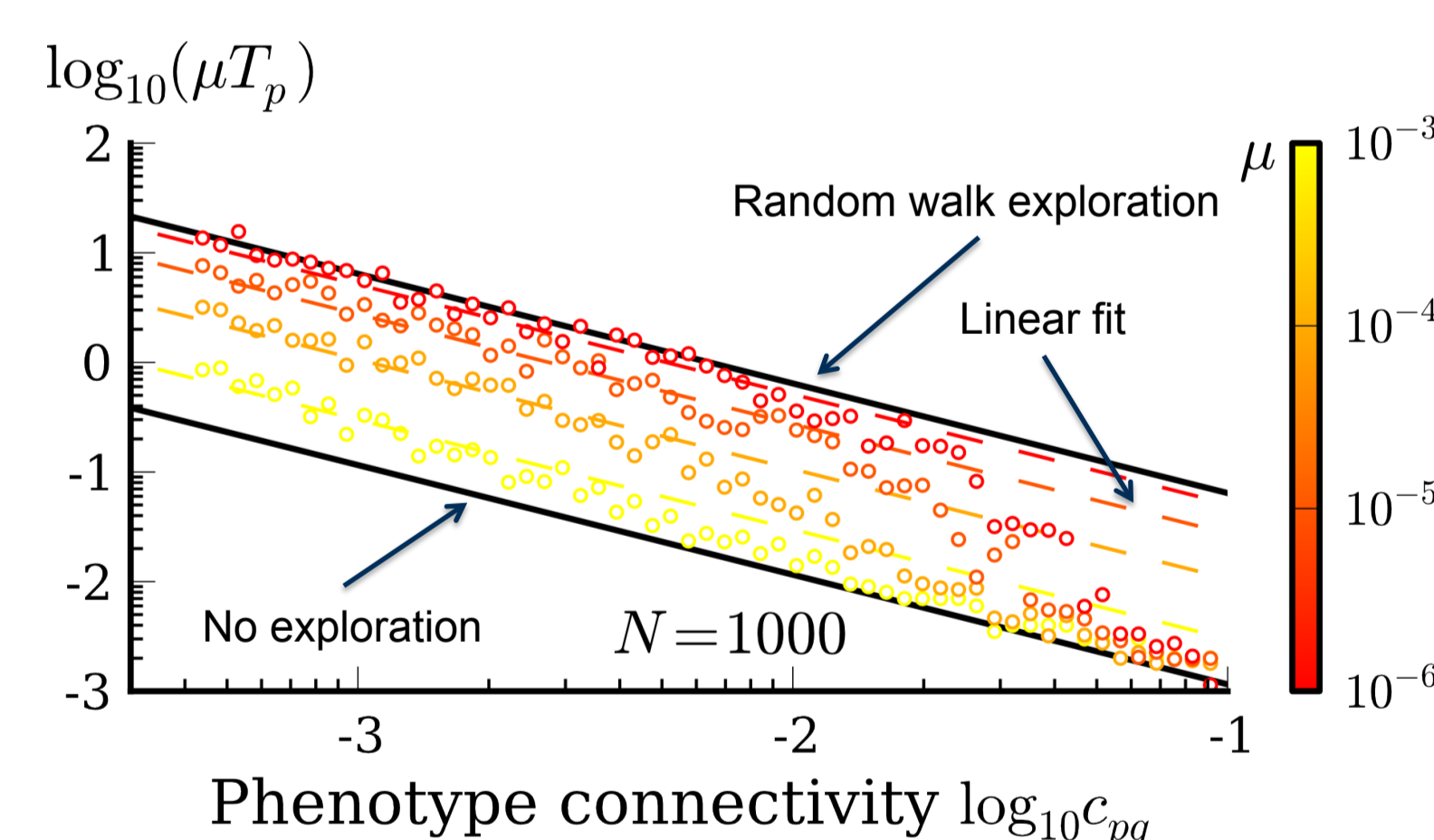$$T_p = \frac{\log 2}{NL\mu(1-\rho_q)c_{pq}}$$

But most phenotypes are rare; generally, they are only discovered through neutral exploration.

The dynamics of exploration are easiest to understand in the limit $NL\mu \ll 1$, when the entire population performs a random walk on the neutral space [5]. The time between steps is distributed exponentially with mean $\tau_f = 1/(L\mu\rho_q)$ and we find
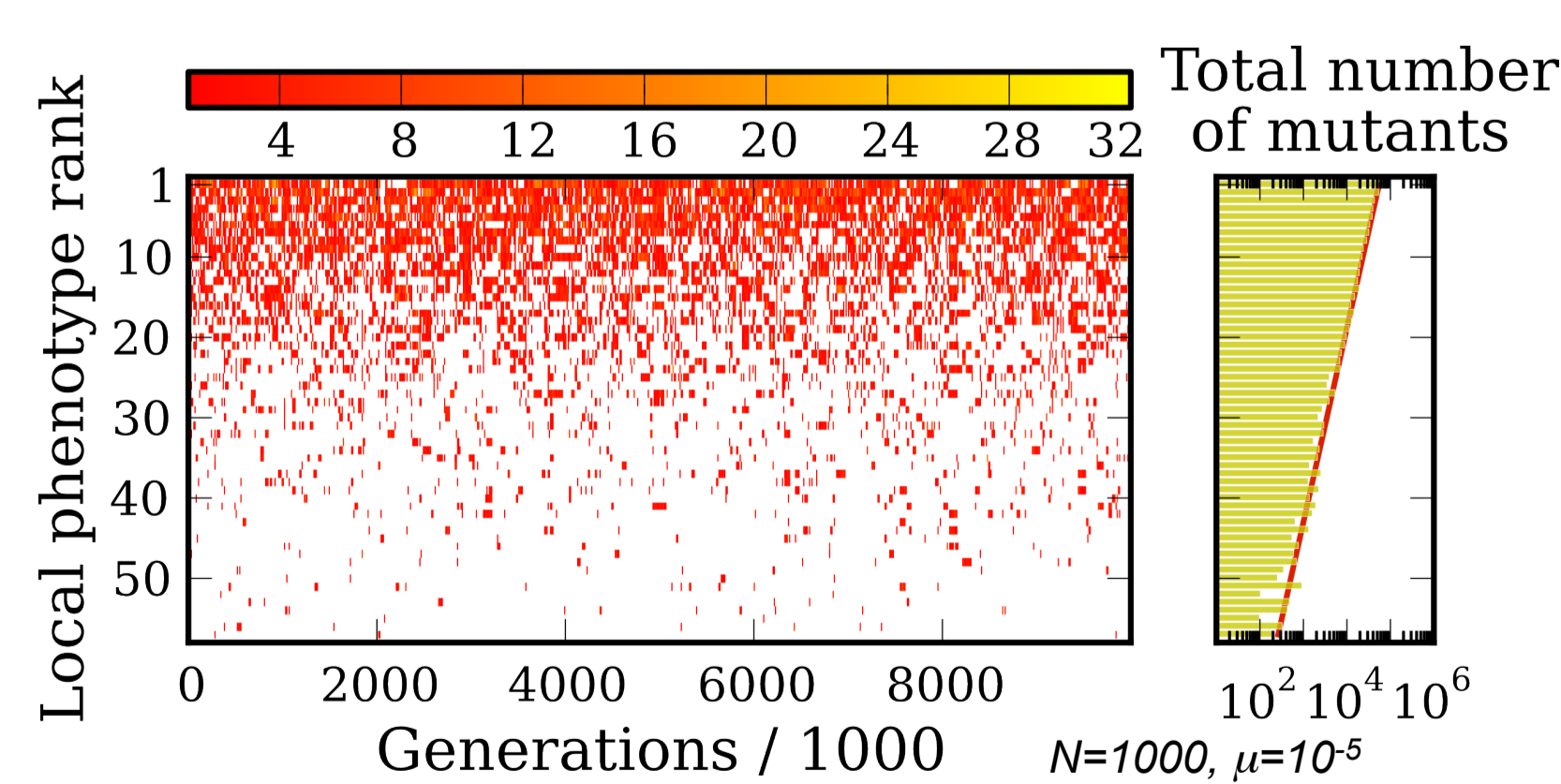
$$T_p = \frac{\log 2}{(K-1)L^2\mu\rho_q(1-\rho_q)c_{pq}}$$

Note the non-monotonic dependence on ρ_q: More robust neutral networks are explored faster, but alternative phenotypes are produced less often [6].

To test these results, we performed simulations under a GP map that randomly assigns genotypes to phenotypes, so that $c_{pq} \propto F_p$ (using $L=12$, $K=4$). For intermediate $NL\mu$, we observe a smooth transition between the extremes.

$\log_{10}(\mu T_p)$

Random walk exploration
Linear fit
No exploration
$N=1000$
μ: 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}

Phenotype connectivity $\log_{10} c_{pq}$

Over long times, the network is explored uniformly so that number of p-mutants scales with c_{pq}. But if the population is monomorphic, there are strong correlations over times on the order of τ_f: The currently accessible phenotypes are produced in bursts.

Local phenotype rank
Total number of mutants
4   8   12   16   20   24   28   32
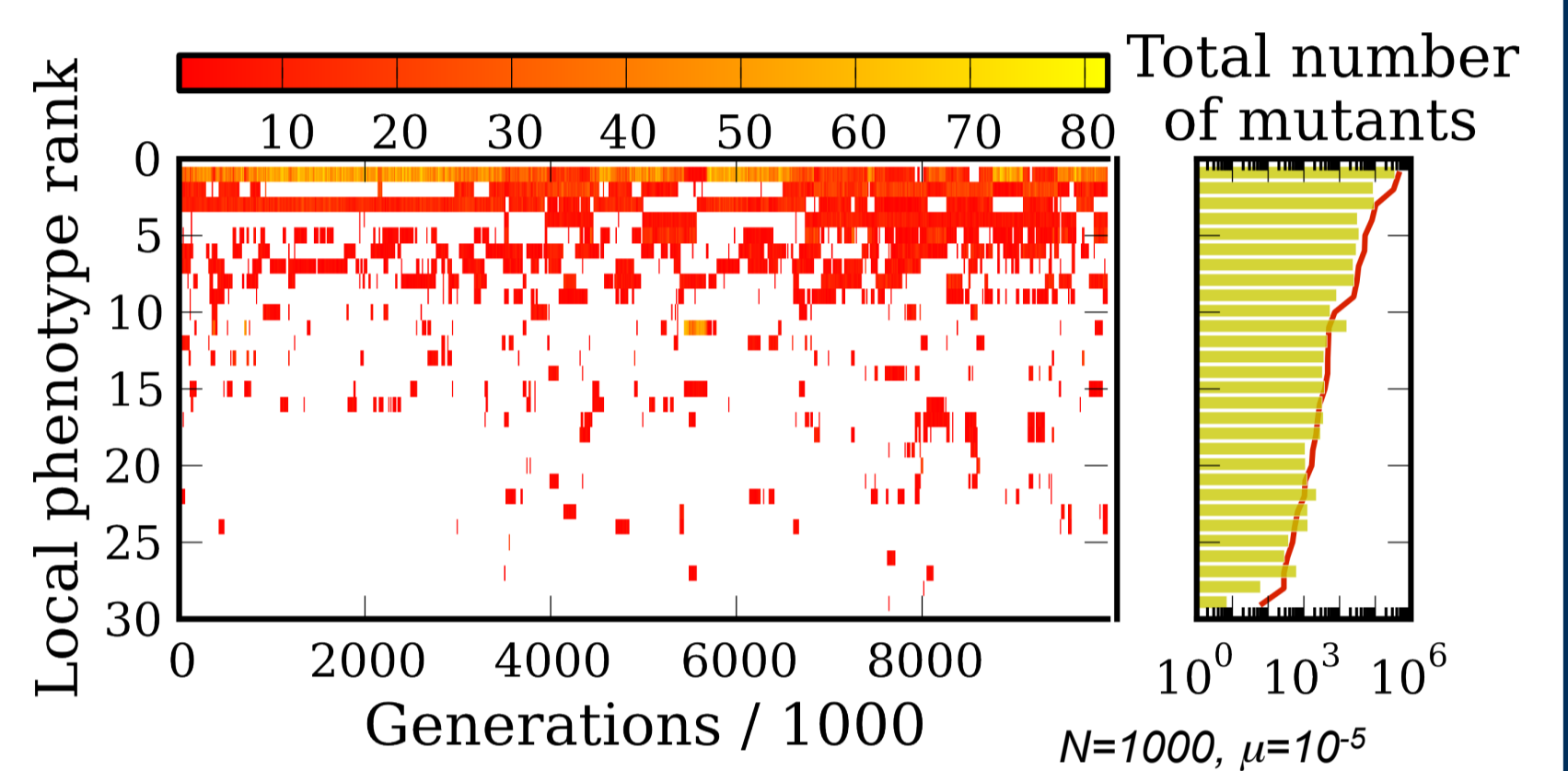Generations / 1000
$N=1000$, $\mu=10^{-5}$

Correlations in neutral networks

More realistic GP maps lead to correlations between genotypes. We study the impact of these correlations using RNA folding as the GP map.
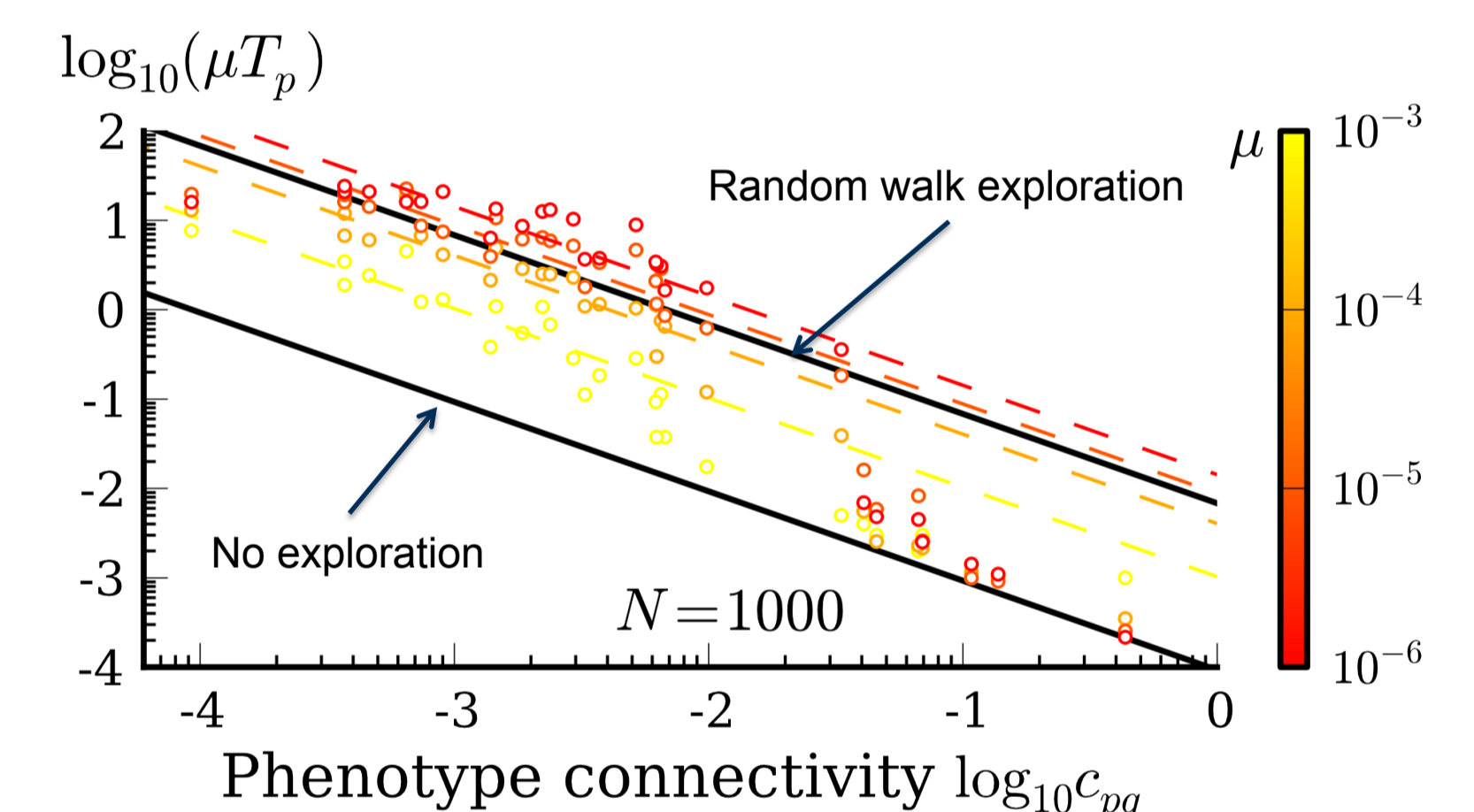
Neutral networks in this system are generally not fully connected [6], and the mutational connectivities are different from the global phenotype frequencies.

$\log_{10} c_{pq}$
Local connectivity
$\rho = 0.38$
Even frequent phenotypes not accessible
Global frequency $\log_{10} F_p$
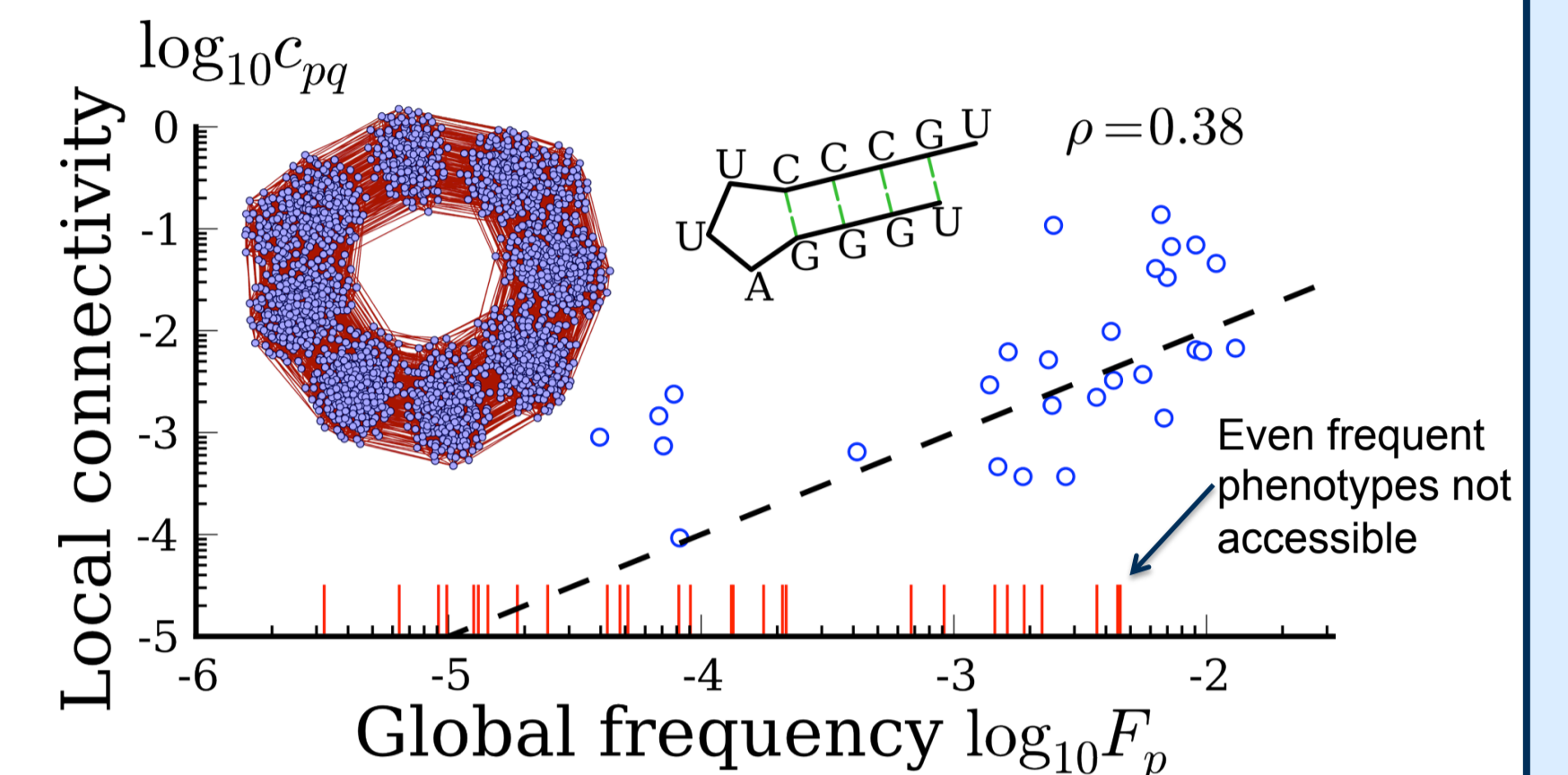
Realistic GP maps can induce complex internal structure in neutral networks. Moving from one genotype to a neighbour may not completely change the spectrum of accessible phenotypes, leading to more pronounced bursts of mutants:

Local phenotype rank
Total number of mutants
10   20   30   40   50   60   70   80
Generations / 1000
$N=1000$, $\mu=10^{-5}$

If it takes several neutral steps to get access to new phenotypes, the discovery of rare phenotypes is delayed even further. The inverse dependence on c_{pq} remains.

$\log_{10}(\mu T_p)$
Random walk exploration
No exploration
$N=1000$
μ: 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}
Phenotype connectivity $\log_{10} c_{pq}$

Conclusions and Outlook

Even though mutations may cause *uniformly* random genotypic change, their effect on phenotypes can be biased by the GP map. We have shown how such bias emerges from the connections between neutral networks that arise through the GP map.

Our framework presents a microscopic null-model for the dynamics of evolving populations on complex networks. As a reward, we can understand important scaling relations (such as the impact of robustness on the discovery of new phenotypes) in intuitive ways.

Mutational bias can drive populations onto suboptimal peaks in the fitness landscape [8]. This bias can arise naturally from the GP map if we view connectivities between neutral networks as effective mutation rates.
In the future, including adaptive transitions into our model promises further insight into the factors that shape the course of evolution.

Questions about the poster?

Please just ask me, or email me at
steffen.schaper@physics.ox.ac.uk

References

1 Hofacker, IL, et al. (1994) *Monatsh. Chemie* **125**:167-188.
2 Li H, Helling R, Tang C, Wingreen N (1996) *Science* **273**:666-669.
3 Johnston IG, Ahnert SE, Doye JPK, Louis AA (2011) *Phys. Rev. E* **83**:066105.
4 Borenstein E, Krakauer DC (2008) *PLoS Comput Biol* **4**:e1000202.
5 van Nimwegen E, Crutchfield JP, Huynen M (1999) *Proc. Nat. Acad. Sci. USA* **96**:9716-9720.
6 Draghi JA, Parsons TL, Wagner GP, Plotkin JB (2010) *Nature* **463**:353-355.
7 Schaper S, Johnston IG, Louis AA (2012) *Proc. Roy. Soc. B.* **279**:1777-1783.
8 Yampolsky LY, Stoltzfus A (2001) *Evo. and Dev.* **3**:73–83.