Simple Models of Complex Biological Systems

Sergei Maslov
Department of Bioengineering,
Carl R. Woese Institute for Genomic Biology,
National Center for Supercomputing Applications,
and Department of Physics,
University of Illinois at Urbana-Champaign
Visiting appointments
Argonne National Laboratory
Brookhaven National Laboratory

International Society of Bottom-Down Modeling

- You all heard about Top-Down and Bottom-Up models
- Time to try **Bottom-Down models**
- Start with a clear idea on what empirical facts to capture
- Simplify, simplify, simplify…
- Everything Should Be Made as Simple as Possible, But Not Simpler (A. Einstein)
- Grades:
  - 0 parameters – A+ & Per Bak award
  - 1 parameter – A
  - 2 parameters – B
  - 3 parameters –C
  - >=4 parameters – F:
    You don't have a bottom down model
    Please return your badge to the society!

I do modeling at all timescales

- Fast: Intra-cellular dynamics
  - Law-of-mass-action protein dynamics in networks & pathways
  - Co-expression networks: core- and pan-network approach
- Medium: Population dynamics in ecosystems
  - **Kill-the-Winner population collapses in saturated environments**
  - Kelly-optimal bet-hedging strategies of viruses infecting bacteria
- Slow: Evolutionary dynamics

  Models of bacterial genome evolution by Horizontal Gene Transfer:
  - **"Copy-replace" vs. "Copy-insert" in genome evolution**
  - How many bureaucrats does a genome need?
    Or "Parkinson's law" in biological systems
  - Why bacteria & archaea run Linux, while eukaryotes run Windows?

  Origins of life: autocatalysis in information-coding polymers
- No multi-scale models!

Evolution of bacterial genomes by Horizontal Gene Transfer and Homologous Recombination

Studier FW, Daegelen P, Lenski RE**, Maslov S,** Kim JF, J. Mol Biol. (2009)

Dixit P*, Pang TY*, Studier FW, **Maslov S** PNAS (2015); arXiv:1405.2548

Dixit P*, Pang TY*, McCorkle S, Amgalan A, **Maslov S,** in preparation

How bacterial genomes evolve

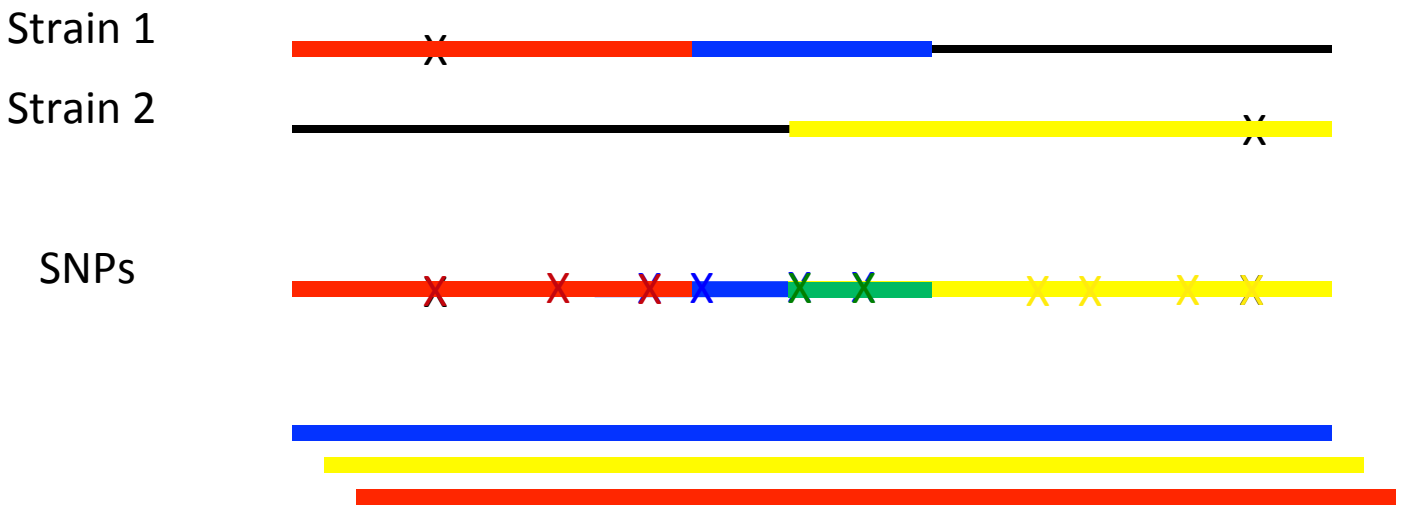
- Bacterial genomes are highly variable: genomes of the same species with ~4000 genes can differ by up to ~1500 horizontally transferred genes
- New genes can be "installed" by Horizontal Gene Transfer (copy-insert)
- Existing genes can be "updated" by Homologous Recombination (copy-replace)
- Here I will talk about copy-replace. For copy-insert see:
  - **S. Maslov**, TY Pang, K. Sneppen, S. Krishna, PNAS (2009)
  - TY Pang, **S. Maslov**, PLoS Comp Bio (2011)
  - J Grilli, B Bassetti, **S Maslov**, M Cosentino Lagomarsino, NAR (2012)
  - TY Pang, **S. Maslov**, PNAS (2013)

Simple neutral evolutionary model

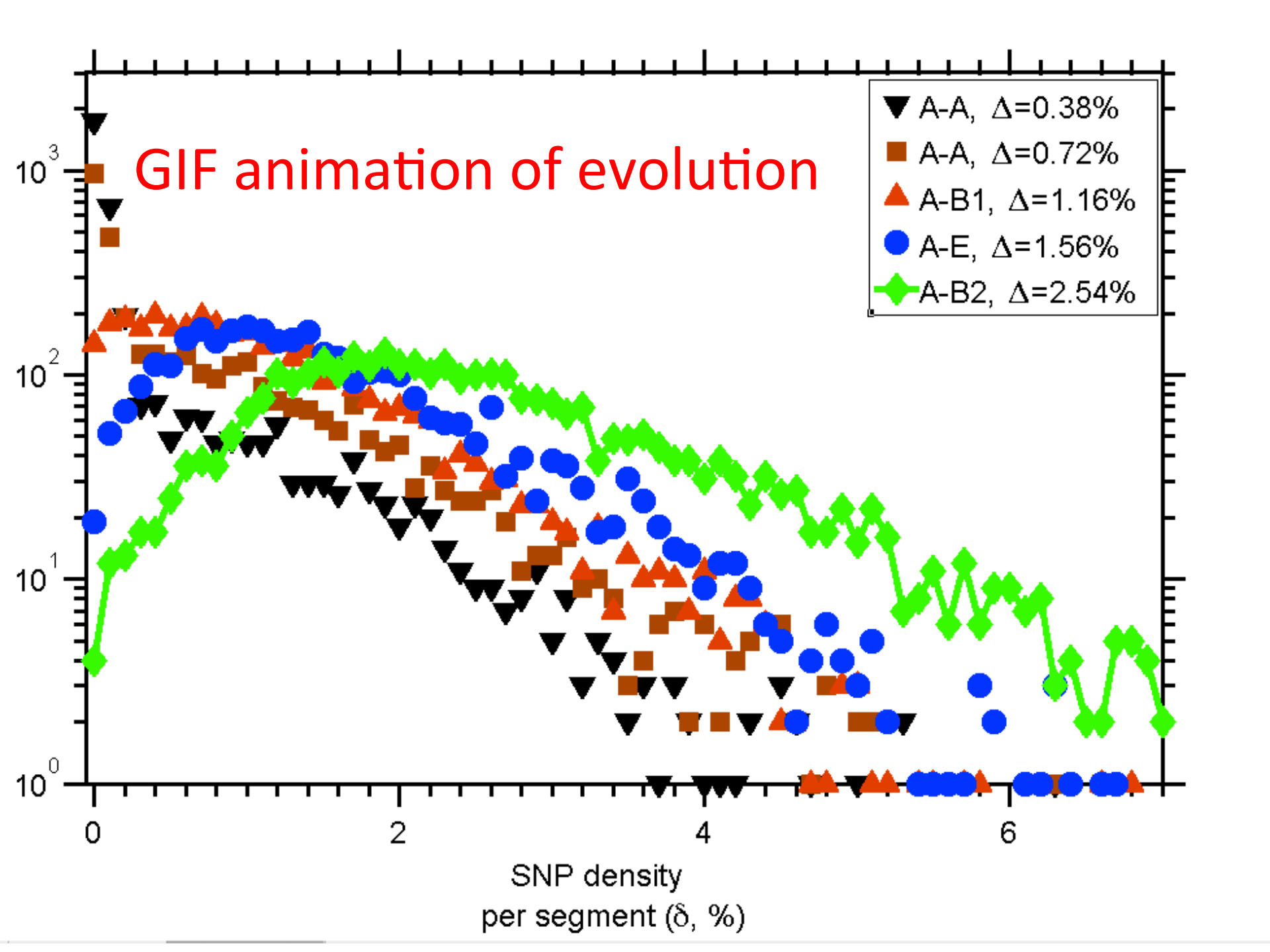
- Mutation rate $\mu = 8.9 \times 10^{-11}$ /bp/generation
- Recombination rate $\rho = 2.8 \times 10^{-11}$/bp/generation
- l_R=3kb - average length of recombined segments
- δ_{TE}=0.8% transfer efficiency = Prob(successful transfer + recombination: ~ $\exp(-\delta/\delta_{TE})$)
- N_e – (effective) population size

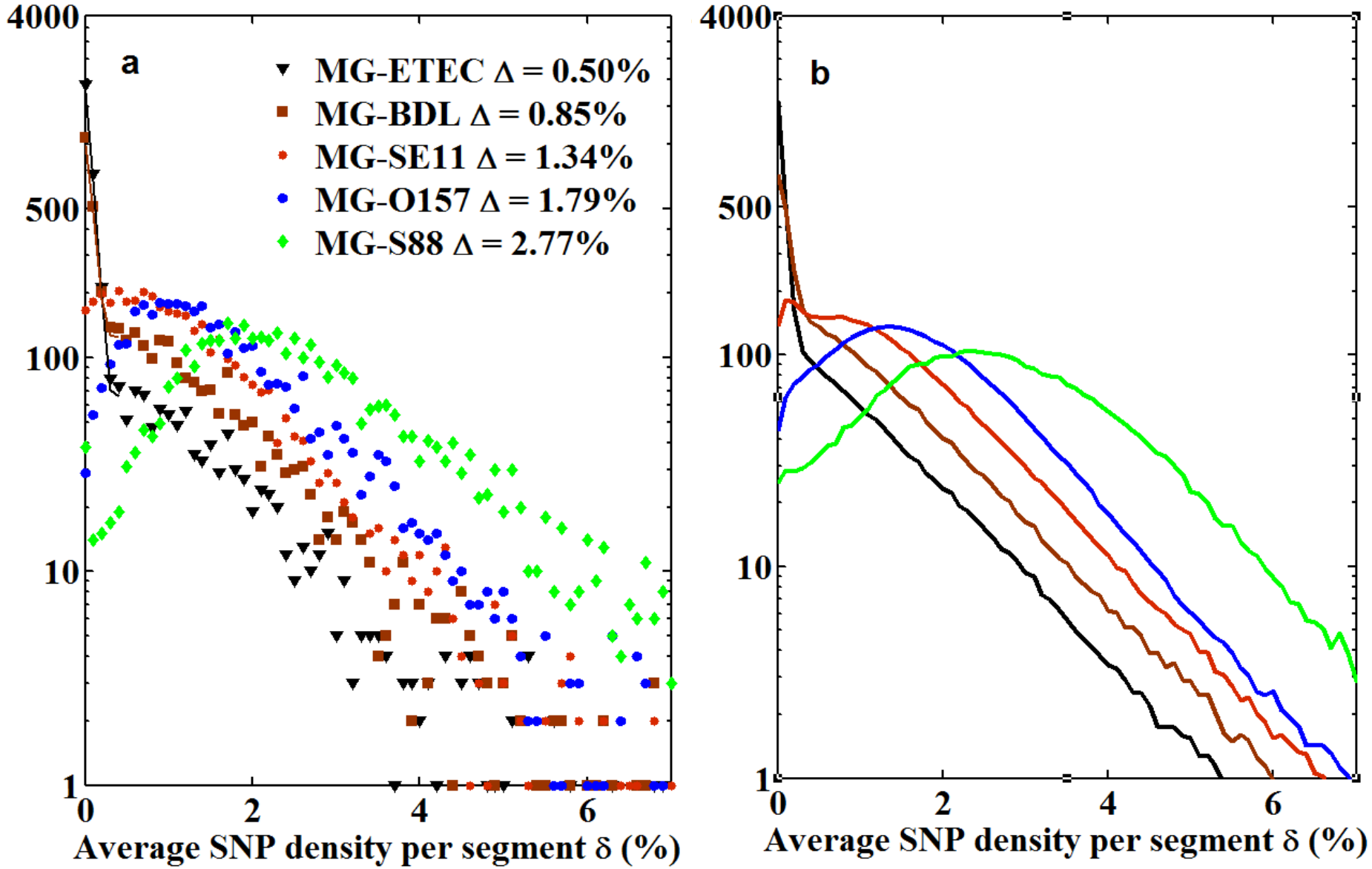
  Many parameters but only two combinations matter: grade B

- **r/m**= $\mu/\rho \times l_R \times \Delta_T(\delta_{TE}, N_e)$=11.2: Ratio= (# of horizontally)/(# of vertically) acquired SNPs
- **θ/δ_{TE}** =$2\mu N_e//\delta_{TE}$: Competition between population diversity and homogenizing transfers



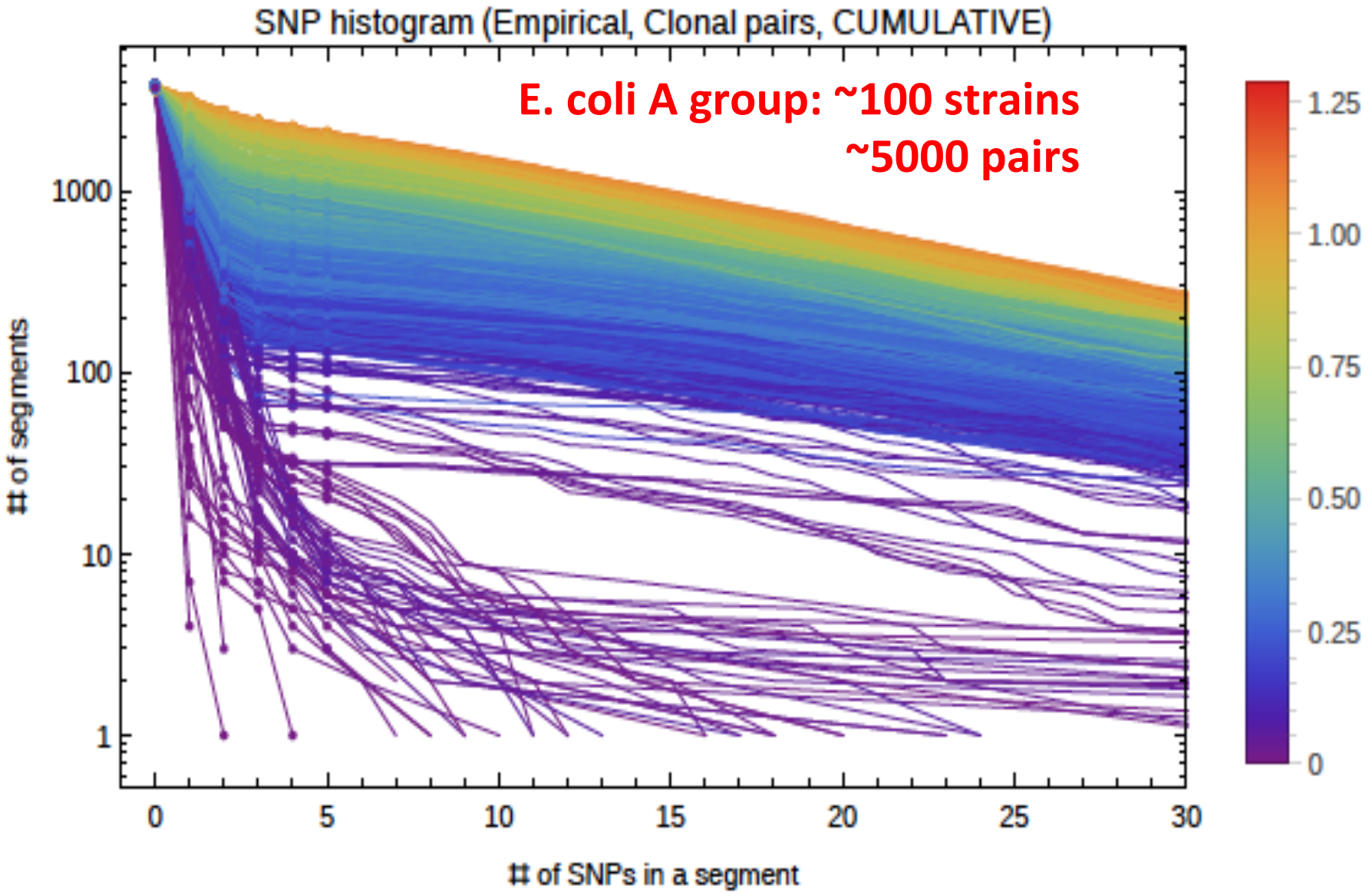
Dixit P, Pang TY, Studier FW, **Maslov S,** PNAS (2015); arXiv:1405.2548


GIF animation of evolution
A-A, Δ=0.38%
A-A, Δ=0.72%
A-B1, Δ=1.16%
A-E, Δ=1.56%
A-B2, Δ=2.54%
10^3
10^2
10^1
10^0
0
2
4
6
SNP density
per segment (δ, %)



Dixit P, Pang TY, Studier FW, **Maslov S,** PNAS (2015); arXiv:1405.2548

HD movie of bacterial evolution



Dixit P, Pang TY, McCorkle S, Amgalan A, **Maslov S,** in preparation (2015)

Pros and cons of working with biologists vs. physicists

- Biologists know a lot about their favorite organism/pathway/protein
  (but are rarely interested in others)
- Biologists now need data analysis and models
  (but the hate equations)
- A solution: write a biological paper
  with equations in SI,
  followed by an expanded physics paper

Pros and cons of working with biologists vs. physicists

PNAS

Heavy use of equations impedes communication among biologists

Tim W. Fawcett[1] and Andrew D. Higginson

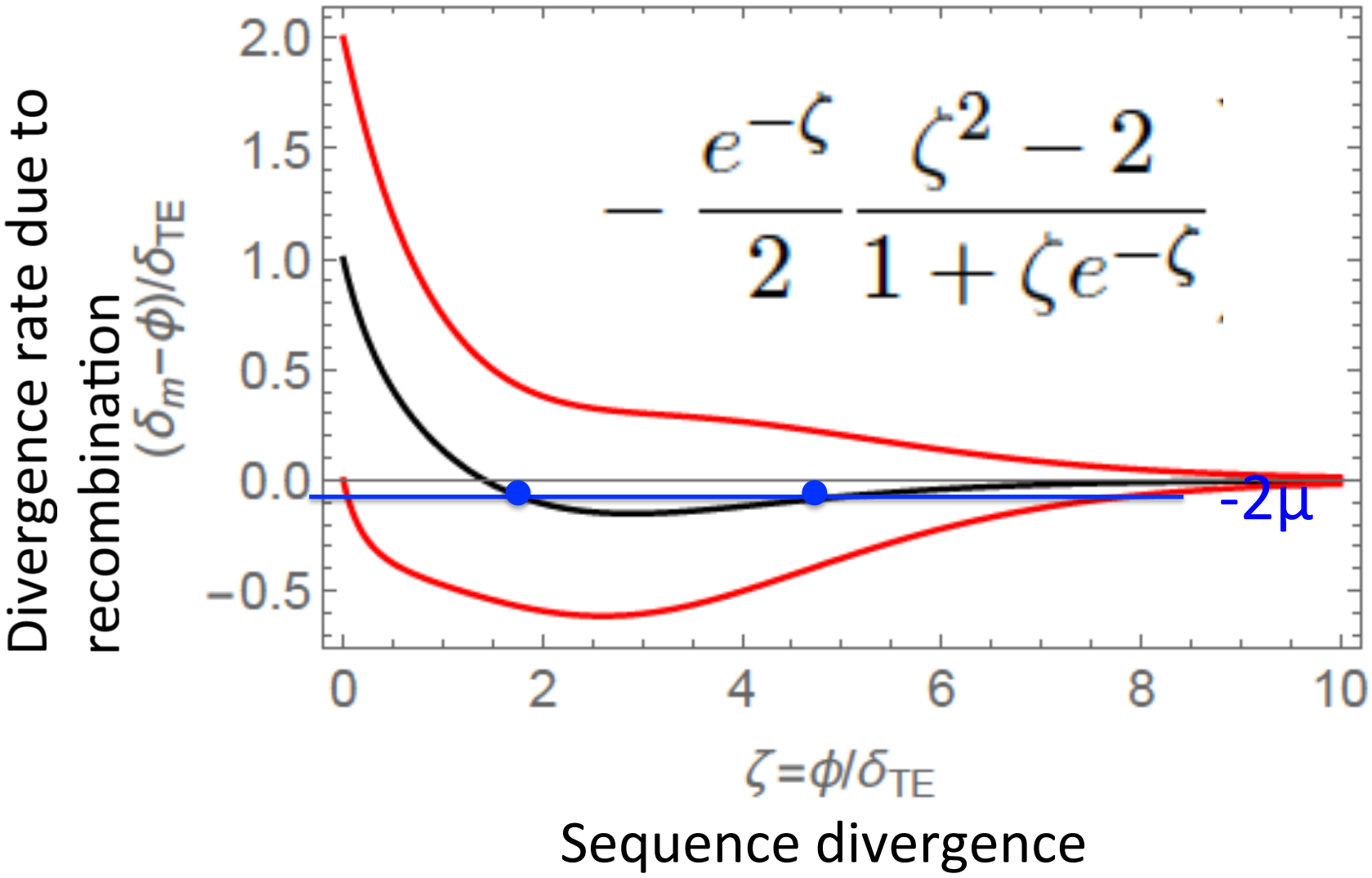
School of Biological Sciences, University of Bristol, Bristol BS8 1UG, United Kingdom

Edited[†] by Robert M. May, University of Oxford, Oxford, United Kingdom, ar

Most research in biology is empirical, yet empirical studies rely fundamentally on theoretical work for generating testable predictions and interpreting observations. Despite this interdependence, many empirical studies build largely on other empirical studies with little direct reference to relevant theory, suggesting a failure of communication that may hinder scientific progress. To investigate the extent of this problem, we analyzed how the use of mathematical equations affects the scientific impact of studies in ecology and evolution. The density of equations in an article has a significant negative impact on citation rates, with papers receiving 28% fewer citations overall for each additional equation per page in the main text. Long, equation-dense papers tend to be more frequently cited

10 equations take a paper with 100 citations down to $100*0.72^{10}$ ~4 citations

Metastability in genome evolution



Population growth in collapsing environments

- There is no unlimited exponential growth: (it assumed for fitness, GDP, house prices). Impossible in the long term (Malthus and Co)
  - We will model species in a saturated environment of finite carrying capacity: average growth rate =0
- Life is uncertain and populations of individual species can rapidly collapse
  - Growth due to resources vacated by collapsed populations. Changes are too fast for ODEs → Cellular automaton with discrete time steps.

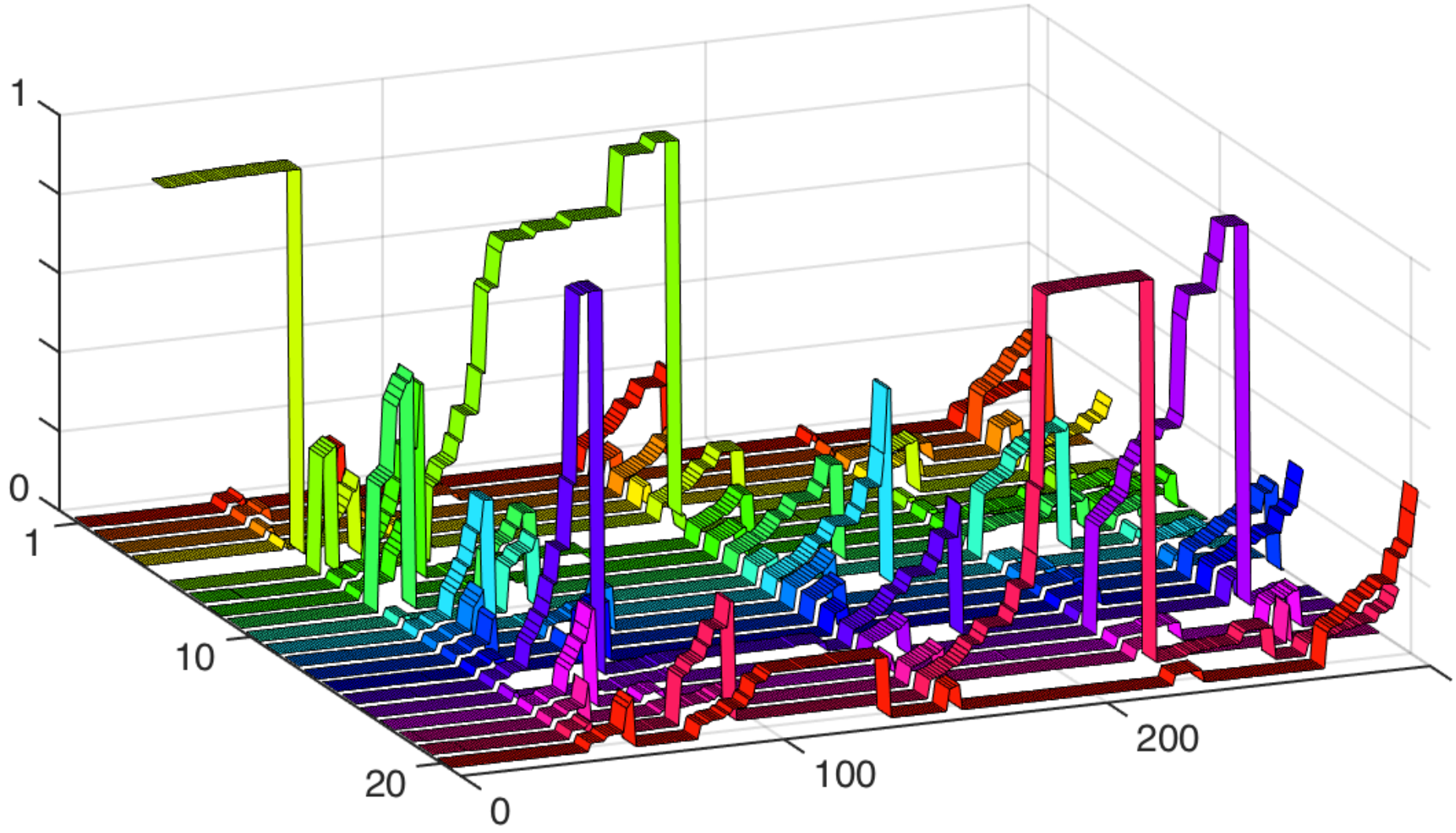
Making a bottom-down model

$$\frac{dP_i}{dt} = \Omega \cdot P_i \cdot \left(1 - \frac{\sum_j P_j}{K} \right) - \eta_i(t) \cdot P_i$$

- Competition for shared resource + collapses
- Ω and K are not needed. Make them 1.
- N competing populations
- Random population collapses down to $\gamma << 1$
- All populations grow by the same factor up to carrying capacity: $\sum_j P_j = 1$. No need for ODE.

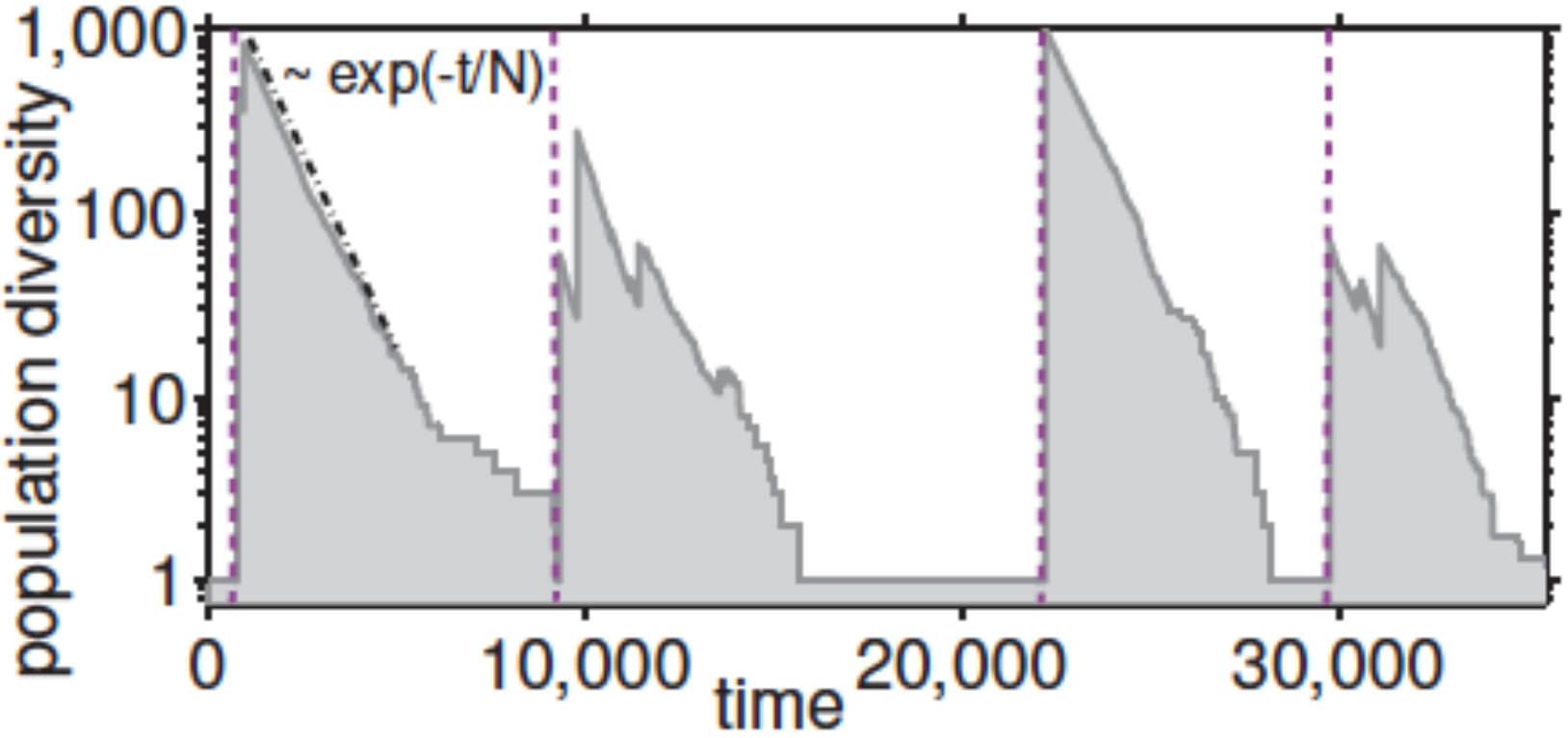
SM, K. Sneppen, PLoS Comp Bio (2015)

Growth is fuelled by collapses



SM, K. Sneppen, PLoS Comp Bio (2015)

Emergent property: diversity waves



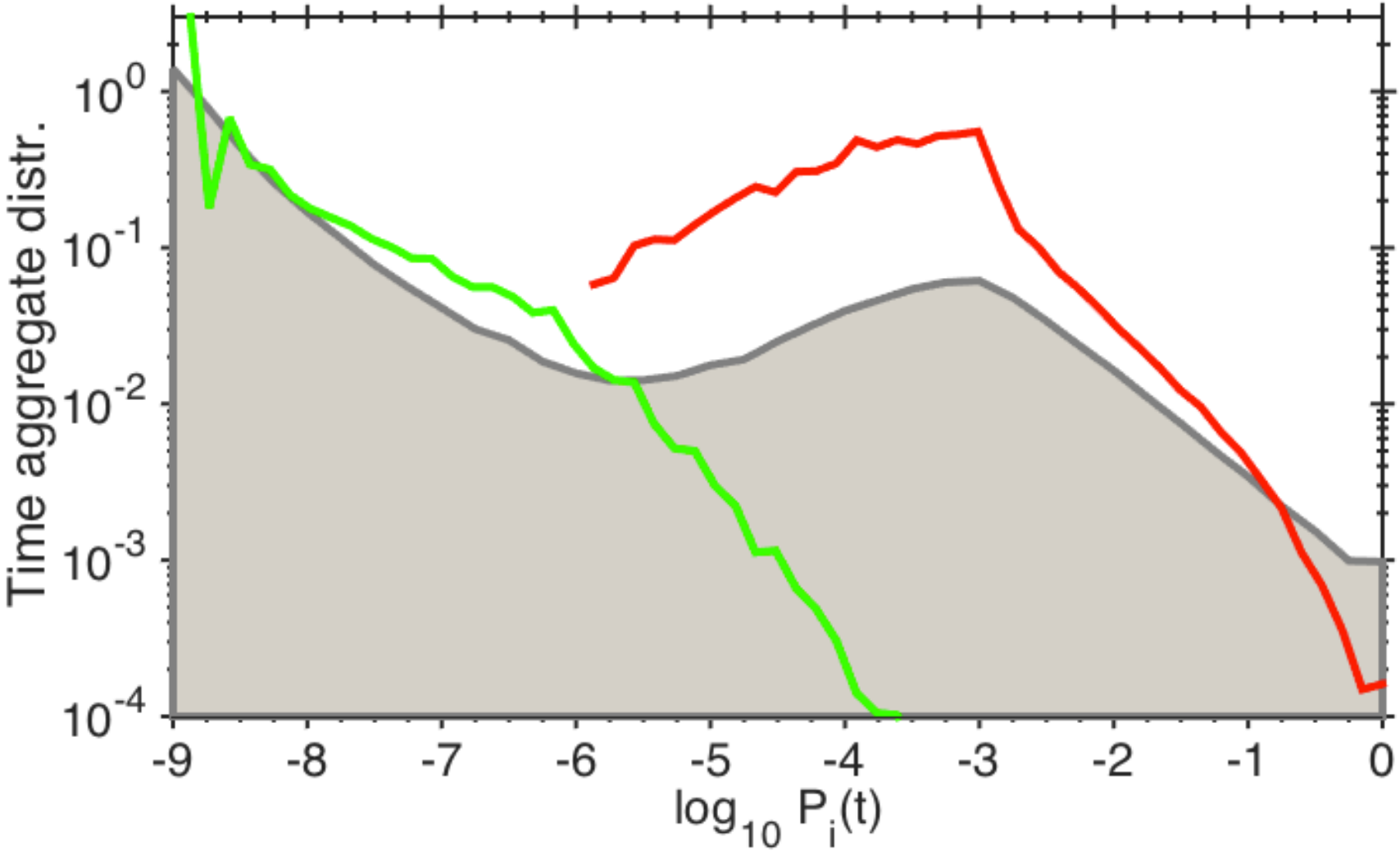
$$D(t) = \frac{1}{\sum_j P_j(t)^2}$$

$$N \cdot \exp(-t_{wave} / N) \sim 1$$

$$t_{wave} \sim N \cdot \log_e N$$

SM, K. Sneppen, PLoS Comp Bio (2015)

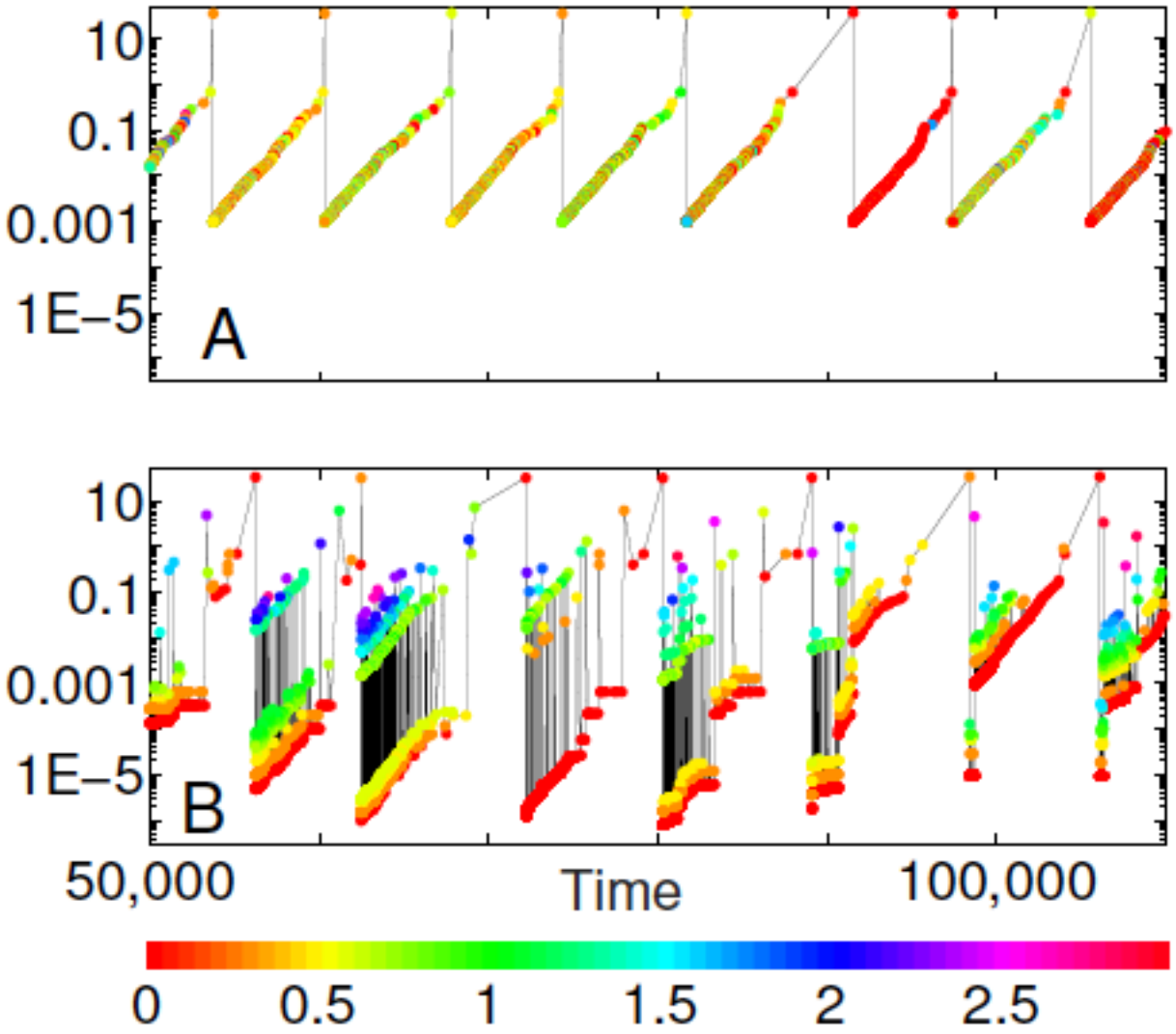
Bimodal Species Abundance Distribution



N=1000, γ=10-9

SM, K. Sneppen, PLoS Comp Bio (2015)

Can we simplify it even more?



Start new wave with equal populations 1/N: boring

Start with populations in the lower wave: exciting long-term memory

SM, K. Sneppen, PLoS Comp Bio (2015)

THOSE ARE MY PRINCIPLES, AND IF YOU DON'T LIKE THEM... WELL, I HAVE OTHERS.

GROUCHO MARX

Adopted as
Center for Models of Life
motto by Kim Sneppen

Bottom-Down Model variants (can have > 3 parameters)

1. Neutral drift model – allows neutral population redistribution between collapses
2. Fitness model – species have individual growth and collapse rates
3. Diffusion model – instead of species we have N environments connected by diffusion
4. "Kill-the-Loser" – collapse of large populations is less likely relevant for companies: collapse~ $P^{-0.2}$ (
5. "Kill-the-Winner" – collapse of large populations is more likely. Relevant for phage-triggered collapses
6. "Kill-the-King" – the largest population collapses first. Relevant for monarchies ☺ .

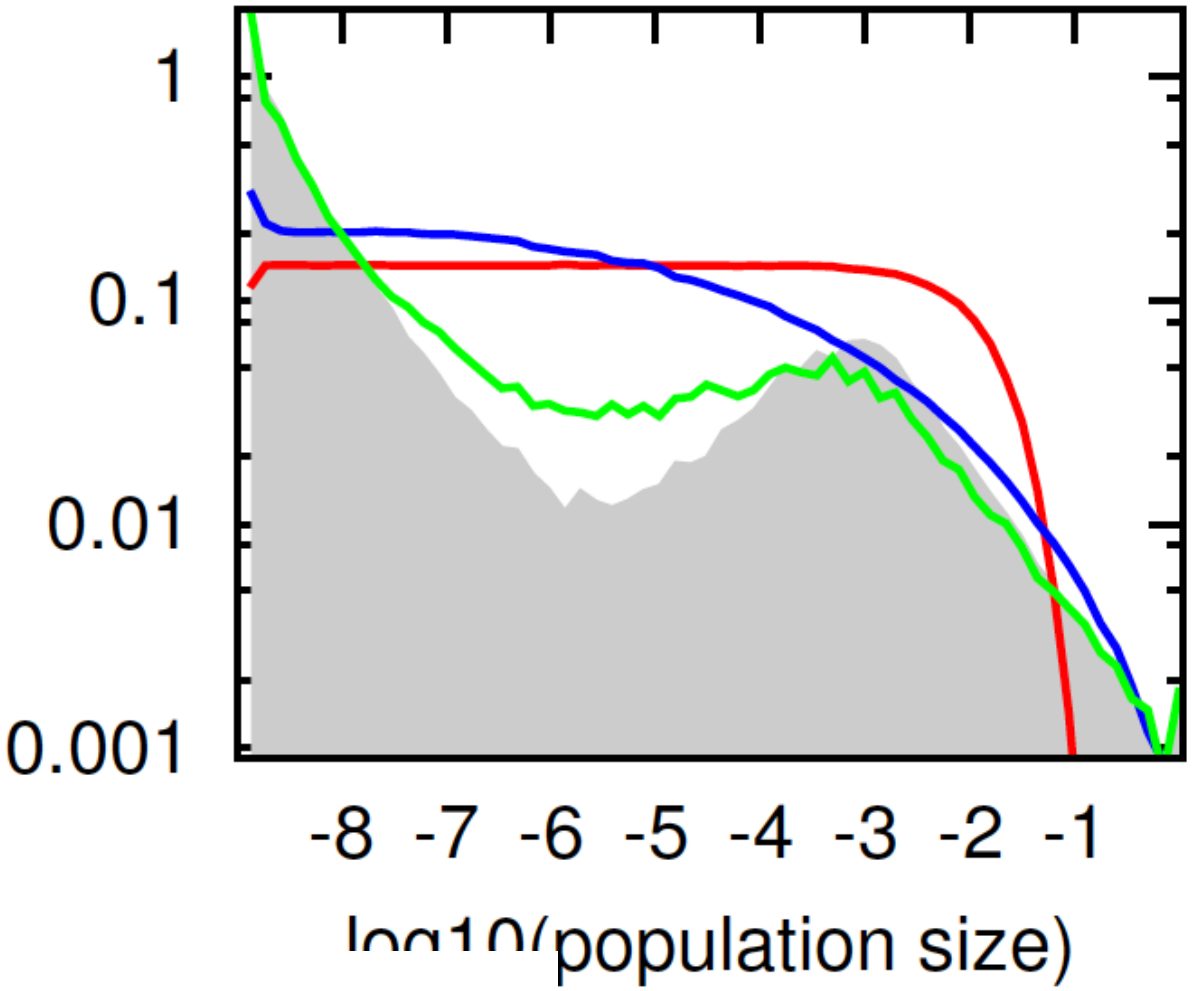
SM, K. Sneppen, PLoS Comp Bio (2015)

“Kill-the-Winner” model

α=0.01

α=0.2

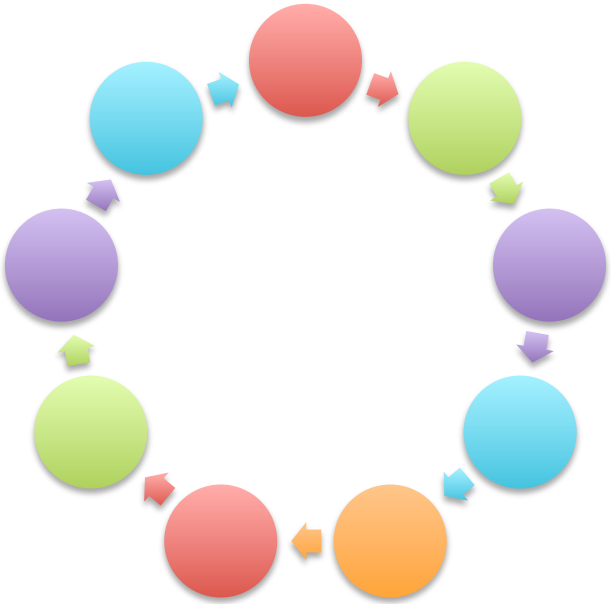
α=1

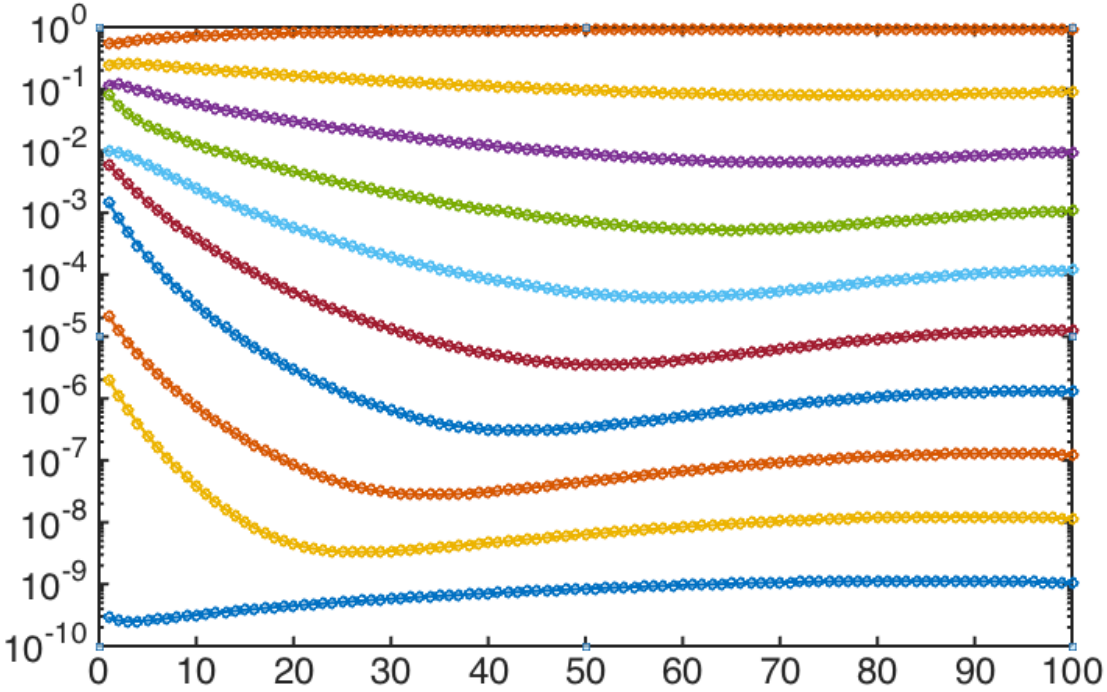


$Collapse_i \sim P_i^{\alpha}$

SM, K. Sneppen, PLoS Comp Bio (2015)

“Kill-the-King” model





$$\delta_i(t) = P_{i+1}(t) / P_i(t)$$

“Hamburger” equation
(anisotropic Burgers’ equation)

$$\delta_i(t+1) - \delta_i(t) = \delta_i(t) \cdot (\delta_{i+1}(t) - \delta_i(t))$$

$$\frac{\partial \delta}{\partial t} = \delta \cdot \frac{\partial \delta}{\partial x} + \frac{\delta}{2} \cdot \frac{\partial^2 \delta}{\partial x^2}$$

SM, K. Sneppen, in preparation (2015)

Collaborators & Funding




- **Kim Sneppen (NBI, Copenhagen)**
- **Tin Yau Pang (CEPLAS Dusselforf)**
- **Purushottam Dixit (Columbia U.)**
- **Bill Studier (BNL)**
- Rich Lenski (Michigan State)
- Patrick Daegelen (France)
- Jinhyun Kim (Korea)

DOE Office of Biological and Environmental Research

DOE Systems Biology Knowledgebase (KBase)

Studier FW, Daegelen P, Lenski RE**, Maslov S,** Kim JF, J. Mol Biol. (2009)
Dixit P*, Pang TY*, Studier FW, **Maslov S** PNAS (2015); arXiv:1405.2548
Dixit P*, Pang TY*, McCorkle S, Amgalan A, **Maslov S,** in preparation (2015)
Maslov S, Sneppen K, PNAS (2015); arXiv:1503.00529

I am looking for physics PhD students

- Have funding for at least 2 PhD students in theoretical statistical physics
- If interested contact me: maslov@illinois.edu
- (217) 265-5705 (o) (631) 327-8222 (cell)
- http://www.cmth.bnl.gov/~maslov
- Physical location: 3406 Carl R. Woese Institute for Genomic Biology
  (office right of Nigel Goldenfeld)