Genetic diversity in changing environments

Kavita Jain

(jain@jncasr.ac.in)

J. Nehru Centre, Bangalore

[with Sachin Kaushik (UNC), Parul Johri (UNC)]

Moran process (Moran 1958)

- Start with a single mutant (1), $N-1$ wild type (0)
- No further mutations are allowed
- Mutant fitness $= 1 + s$, wild type fitness $= 1$
- Prob(birth) $\propto$ fitness, Prob(death) $\propto 1$

t	0	1	0	0	0	1
t+1	0	1 1	0	0	~~0~~	1

$$p_{i\to i+1} \propto (1+s)i \times (N-i)$$

$$p_{i\to i-1} \propto (N-i) \times i$$

- Two absorbing states: eventually none or all mutants

Fixation probability

- Starting with single mutant, prob eventually all mutants?
- Backward Fokker-Planck equation for $P(p,t;x=1,t'>t)$,

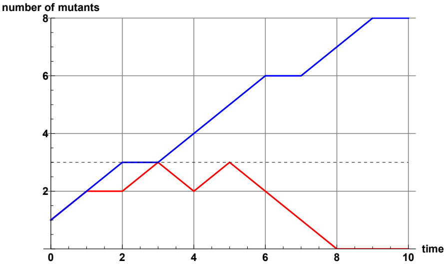
$$-\frac{\partial P(p,t)}{\partial t} = \frac{sp(1-p)}{2}\frac{\partial P(p,t)}{\partial p} + \frac{p(1-p)}{2N}\frac{\partial^2 P(p,t)}{\partial p^2}$$

- For boundary conditions, $P(0,t)=0, P(1,t)=1$,

$$P(p=\frac{1}{N}, t\to\infty) = \frac{1-e^{-s}}{1-e^{-Ns}} \approx \begin{cases} s & , \quad Ns \gg 1 \\ \frac{1}{N} & , \quad s = 0 \\ e^{-N|s|} & , \quad Ns \ll -1 \end{cases}$$

Mean sojourn time

- Mean # of visits to a site before eventual absorption



- Mean time spent between x and $x + dx$, starting from p, before eventual absorption,

$$G(p;x) = \int_{-\infty}^{0} P(p,t;x,0)dt\ ,\ \ 0 < x < 1$$

where, $P(p,t;x,0)$ obeys the backward equation

Mean sojourn time

- Since

$$-\frac{\partial P(p,t)}{\partial t} = a(p)\frac{\partial P(p,t)}{\partial p} + b(p)\frac{\partial^2 P(p,t)}{\partial p^2}$$

- Mean sojourn time is the Green's function,

$$a(p)\frac{\partial G(p;x)}{\partial p} + b(p)\frac{\partial^2 G(p;x)}{\partial p^2} = -\delta(x-p)$$

with $G(0;x) = G(1;x) = 0$

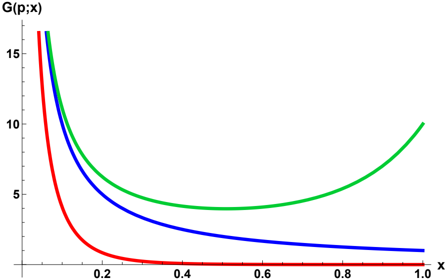
- Mean absorption time, starting from p,

$$\int_0^1 G(p;x)dx$$

Moran process: mean sojourn time

- Starting with single mutant, the mean sojourn time is

$$G(p;x) \overset{x>p\to 0}{=} \frac{1}{x(1-x)} \frac{1-e^{-Ns(1-x)}}{1-e^{-Ns}} \propto \begin{cases} \frac{1}{x(1-x)} & , \quad Ns \gg 1 \\ \frac{1}{x} & , \quad s = 0 \\ e^{-N|s|(1-x)} & , \quad Ns \ll -1 \end{cases}$$



Genetic diversity

- Data from 6 individuals, large number of sequenced loci

Samples \ Loci	1	2	3	4	5	6	7	8	...
1	0	1	0	0	0	0	1	0	...
2	1	0	1	0	0	0	1	0	...
3	0	1	1	0	0	1	1	0	...
4	0	0	0	0	1	0	1	1	...
5	0	0	1	0	0	0	1	0	...
6	0	0	0	1	0	1	1	0	...

- E.g., how many 'diverse' loci?
- What evolutionary forces shaped the diversity?

Site frequency spectrum

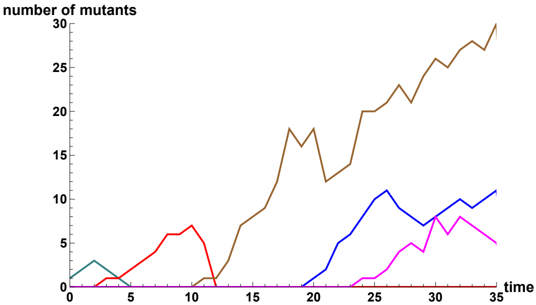
- $f(j,t)$=Mean # of loci with $0 < j < N$ mutants at time t?

Samples \ Loci	1	2	3	4	5	6	7	8	...
1	0	1	0	0	0	0	1	0	...
2	1	0	1	0	0	0	1	0	...
3	0	1	1	0	0	1	1	0	...
4	0	0	0	0	1	0	1	1	...
5	0	0	1	0	0	0	1	0	...
6	0	0	0	1	0	1	1	0	...
# of 1's	1	2	3	1	1	2	6	1	...

- Measurable from data; $\sum_{j=1}^{N-1} f(j,t) = \#$ of diverse loci; ...

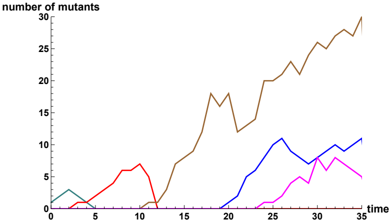
Modeling genetic diversity (Sawyer+Hartl 1992)

- Assume: independent evolution at each locus
- Stochastic (say, Moran) trajectories start with single mutant that arrive at different instants with rate $2N\mu$
- No more mutations (assuming infinite loci)



Stationary state

- Trajecs lost due to absorption; created via new mutations



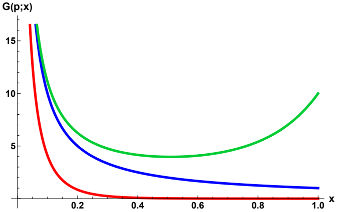
- Mean # of loci with freq $0 < x < 1$ at large times?

$$f(x, t \to \infty) = 2N\mu \int_{-\infty}^{0} P(p \to 0, t; x, 0) dt \propto G(p \to 0; x)$$

Moran process: Stationary state

- Assuming stationary state, SFS used to infer selection

  E.g., U-shaped? Suggests $s > 0$



- Mean number of 'diverse' loci,

$$\int_{\frac{1}{N}}^{1-\frac{1}{N}} f(x)dx \approx 2N\mu \times \begin{cases} 2\ln N & , \quad Ns \gg 1 \\ \ln N & , \quad s = 0 \\ const & , \quad Ns \ll -1 \end{cases}$$

Larger populations are more diverse

Diversity in nonequilibrium situations

- Dynamics of $f(x, t)$ in constant environments

  How diversity varies with time? Relaxation to equilibrium?

  (Evans et al. 2007; ...; Götsch+Bürger 2023)

- In time-inhomogeneous environments?

  Effect of changing population size (Williamson et al. 2005 ...), changing selection (Huerta-Sanchez et al. 2008; Kaushik+KJ 2021), both (KJ+Kaushik 2022; Balick 2023)

Fokker-Planck equation with time-dependent rates

- Starting from $p \to 0$, we have

$$f(x,t) = \int_0^t 2N(t')\mu \times P(x,t;p,t')dt'$$

- The forward Fokker-Planck equation for $f(x,t)$ is

(Evans, Shvets, Slatkin 2007)

$$\frac{\partial f(x,t)}{\partial t} = -s(t)\frac{\partial}{\partial x}\left[\frac{x(1-x)f(x,t)}{2}\right] + \frac{\partial^2}{\partial x^2}\left[\frac{x(1-x)f(x,t)}{2N(t)}\right]$$

- Mutational input modeled by a boundary condition:

$$\mathrm{Lim}_{x\to 0} f(x,t) = \frac{2N(t)\mu}{x}, \quad f(1,t) = \text{finite}$$

Fokker-Planck equation with time-dependent rates

- Since inhomogeneous boundary condition, work with

$$v(x,t) = x(1-x)f(x,t) - 2N(t)\mu(1-x)$$

- Expand in an orthonormal basis (that obey bdry condns) with time-dependent coefficients,

$$v(x,t) = \sum_m a_m(t)\psi_m(x)$$

- Due to selection term, in general, $a_m(t)$ obeys a three-term recursion (Kimura 1964; KJ+Devi 2020)

$$\frac{da_m}{dt} = c_+(m)a_{m+1} + c_-(m)a_{m-1} + c_0(m)a_m$$

Neutral case: exactly solvable (Evans et al. 2007)

- For $s = 0$, we have

$$\frac{\partial f(x,t)}{\partial t} = \frac{\partial^2}{\partial x^2}\left[\frac{x(1-x)f(x,t)}{2N(t)}\right]$$

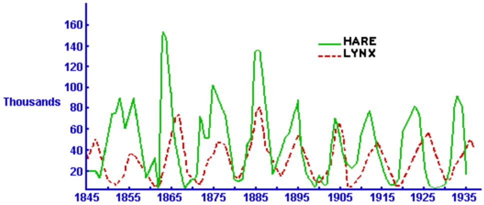
- Expand in eigenfunctions,

$$\frac{\partial^2}{\partial x^2}\left[\frac{x(1-x)\psi}{2}\right] = -\lambda\psi(x)$$

given by Gauss hypergeometric function (Kimura 1955)

I. Periodically changing environment (KJ+Kaushik 2022)

- e.g., seasonal variations can affect fitness in plants
- demography due to, for e.g., prey-predator dynamics



(Odum 1953)

Model parameters

$$\frac{\partial f(x,t)}{\partial t} = -s(t)\frac{\partial}{\partial x}\left[\frac{x(1-x)f(x,t)}{2}\right] + \frac{\partial^2}{\partial x^2}\left[\frac{x(1-x)f(x,t)}{2N(t)}\right]$$

- In general,

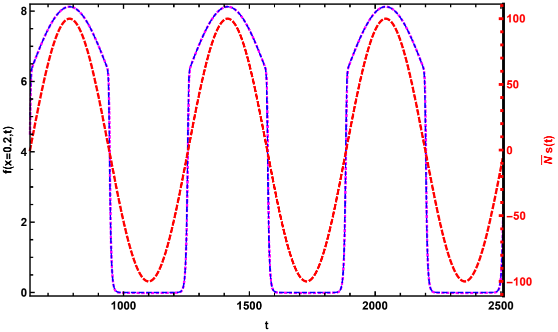
$$\begin{aligned} s(t) &= \bar{s} + \sigma \sin(\omega t + \phi) \\ N(t) &= \bar{N}[1 + \nu \sin(\Omega t + \Phi)] \end{aligned}$$

- Time scales: $\bar{N}, 1/\bar{s}, \omega^{-1} = \Omega^{-1}$
  - Slowly changing environment, $\omega^{-1} \gg \bar{N}, 1/\bar{s}$
  - Rapidly changing environment, $\omega^{-1} \ll \bar{N}, 1/\bar{s}$
- At late times, $f(x,t)$ changes periodically; on averaging

$$\bar{f}(x) = \frac{\omega}{2\pi}\int_0^{2\pi/\omega} f(x,t)dt$$

Slowly changing environment

- Adiabatic approx: $s \to s(t), N \to N(t)$ in stationary result



- In the absence of selection,

$$f(x,t) = \frac{2N(t)\mu}{x} \;,\; \bar{f}(x) = \frac{2\bar{N}\mu}{x}$$

Slowly changing environment

- But with selection, nonlinear dependence on N, s:

$$f(x,t) \approx \begin{cases} \dfrac{2N(t)\mu}{x(1-x)} & , \quad N(t)s(t) \gg 1 \\ \dfrac{2N(t)\mu}{x(1-x)} e^{-N(t)|s(t)|x} & , \quad N(t)s(t) \ll -1 \end{cases}$$

- Only positive part of cycle contributes,

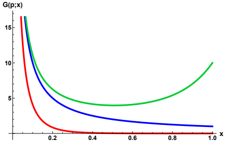
$$\bar{f}(x) = \frac{1}{x(1-x)} \times \int_0^{2\pi/\omega} 2\mu N(t')\ \Theta_H[s(t')]\ \frac{dt'}{2\pi/\omega}$$

Implications

$$\bar{f}(x) = \frac{1}{x(1-x)} \times \int_0^{2\pi/\omega} 2\mu N(t')\ \Theta_H[s(t')]\ \frac{dt'}{2\pi/\omega}$$

- Even if selection zero on average, $\bar{f}(x)$ still U-shaped

  $\implies$ misinfer parameters if assume constant selection



- Lewontin's paradox (1974): Observed (neutral) diversity smaller than predicted using census pop size. Effective pop size captures joint effect of changing N and s (KJ+Kaushik 2022); lower diversity than average pop size

Rapidly varying environment

- In the absence of selection, using the exact solution,

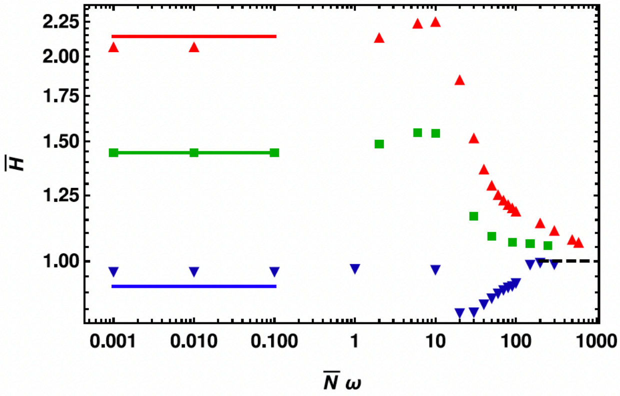
$$\bar{f}(x) = \frac{1}{x} \times \left[\int_0^{2\pi/\omega} \frac{2\mu}{N(t')} \frac{dt'}{2\pi/\omega} \right]^{-1}$$

so that effective pop size is the harmonic mean

- With selection, numerical analyses suggest that stationary state results with average parameters $(\bar{N}, \bar{s})$ hold (but not always true)

Diversity in changing environment

- Varies non-monotonically with environmental frequency



- Max/min depends on other parameters (dominance coeff)

II. Selective sweep (Maynard Smith+Haigh 1974)

- Motivated by Lewontin's paradox of low diversity
- Consider Moran process for 2 physically linked loci

  W0, W1 have fitness 1; $S0$, $S1$ have $1 + s, s > 0$

1	W0	W0	W0	S1
2	W0	W0	W0	S1
3	W1	W1	S1	S1
4	W1	S1	S1	S1

- Due to selection at special site (provided S not lost), initial diversity in 0s and 1s is lowered

Selective sweep in asexuals (with Kaushik+Johri)

- As before: large number of loci; single 1 arises at new loci at different instants
- But now "interacting" loci as they are physically linked

1	W 0 0...	W 0 0...	W 0 0 0...	S 1 0 0...
2	W 0 0...	W 0 0...	W 0 0 0...	S 1 0 0...
3	W 0 0...	S 1 0...	S 1 0 0...	S 1 1 0...
4	S 1 0...	S 1 0...	S 1 1 0...	S 1 1 0...

- Process stops when S is fixed in the population

Diversity in growing population

- Interested in diversity in S-subpopulation
- Full model has selection (W vs. S), fixed population size but within S-subpop, no selection but growing size, $N(t)$

			S 1 0 0...
			S 1 1 0...
	S 1 0...	S 1 0 0...	S 1 0 0...
S 1 0...	S 1 0...	S 1 1 0...	S 1 1 0...

Moran process with growing size

- Assume: independent evolution at each locus
- When S-subpop size remains same

t	0	1	0	0	0	1
t+1	0	1 1	0	0	~~0~~	1

- If increases, either type equally likely to be added

t	0	1	0	0	0	1
t+1	0	1 1	0	0	0	1

- If decreases, either type equally likely to be removed

t	0	1	0	0	0	1
t+1	0	1	0	0	~~0~~	1

Fokker-Planck equation

- Change in frequency requires taking care of not only change in mutant number but also population size
  E.g., $n_{t+1} = n_t$, $N_{t+1} = N_t + 1$,...

- The effective pop size is derived,

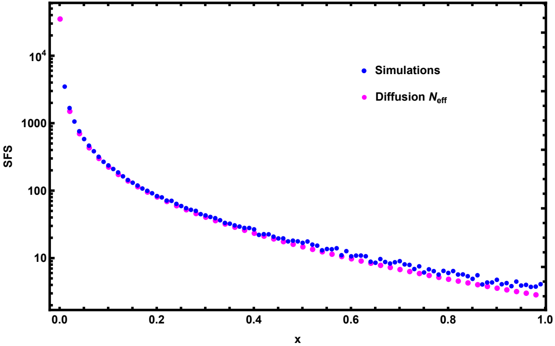
$$\frac{\partial f(x,t)}{\partial t} = \frac{\partial^2}{\partial x^2}\left[\frac{x(1-x)f(x,t)}{2N_e(t)}\right]$$

  $N_e(t)$ is smaller than the naïve expectation, $N(t)$

On-fixation diversity

- At the end of the process when S has fixed,

$$f(x, t_{fix}) \sim \begin{cases} 1/x \,, & x \to 0 \\ 1/x^2 \,, & x \to 1 \end{cases}$$



- Dynamics under study

Summary

- Long history of fruitful exchange of ideas between population genetics, statistical physics, probability theory (de Vladar+Barton 2011)
- Resolving Lewontin's paradox: <u>joint</u> effect of several factors including demography, fluctuating selection, sweeps, ...; <u>nonequilibrium</u> population, consider dynamics (Charlesworth+Jensen 2022)
- Considered stochastic models with time-dependent rates