Controls on genotype entropy in dispersive environments

Burcu Tepekule, Gregory Britten

June 2017

Abstract

Motivated by models of marine microbial diversity, we investigate the forces acting on the distribution of genotypes in a dispersive environment. Integrating patch dynamics and mutator-replicator theory, we demonstrate the opposing forces of dispersal, selection, and mutation controlling the distribution of genotypes across connected environments of variable fitness.

Hypotheses: local Shannon entropy is controlled by the following forces:

- **Selection acting to reduce local entropy via competitive exclusion**. At zero dispersal with no mutation, species in isolated environments approach competitive exclusion with the most fit genotypes approaching fixation.

- **Mutation generates entropy within-environments**. Mutation leads to survival of the flattest.

- **Dispersal mixes entropy across environments**. Infinite dispersal with or without mutation collapses the meta-environmental structure and leads to competitive exclusion with respect to the mean environment, or also complete neutrality in the 'no free lunch' scenario. At intermediate dispersal, a mutation-selection-dispersal balance emerges wherein suboptimal genotype distributions arise due to mixing of environments. High dispersal leads to survival of the flattest.

Throughout the paper we characterize genotype distributions in terms of the Shannon entropy within and across environments.

Questions

- **mutation-selection balance** How does the entropy distribution change as a function of 1) mutation rate, 2) dispersal rate, 3) selection gradient ($\alpha\delta r$), 4) dispersal network structure (degree distribution),

- **Species interactions** Is 'few strong many weak' robust to dispersal network structure? Are other interaction networks more robust for particular dispersal networks? Mutualism vs. antagonism; conjugate interaction-dispersal matrix pairs

- **Complexity versus stability**

Binary sequences

$$X_{000} = [0, 0, 0]$$
$$X_{001} = [0, 0, 1]$$
$$\vdots$$
$$X_{111} = [1, 1, 1]$$

With three bits we have 2^n sequences.

Governing ordinary differential equation

The governing ODE for the logistic quasispecies on a dispersal network is as follows

$$\frac{dx_i^l}{dt} = r_i^l x_i^l \left(1 - \frac{\Sigma x^l}{k_i^l} \right) + \sum_{\substack{j \\ j \neq i}}^{N} q^{h(i,j)} x_j^l - \sum_{\substack{j \\ j \neq i}}^{N} q^{h(i,j)} x_i^l + \sum_{\substack{k \\ k \neq l}}^{P} m_i^{k \to l} x_i^k - \sum_{\substack{k \\ l \neq k}}^{P} m_i^{l \to k} x_i^l$$

with parameters and state variable symbols defined in Table X.

Table 1: Parameter and state variance definitions for the governing ODE

Symbol	Definition
r_i^l	intrinsic growth rate
x_i^l	concentration of sequence i in environment l
Σx^l	total sequence abundance in environment l
k_i^l	carrying capacity of species i in environment l
q	mutation probability per bit
$h(i, j)$	Hamming distance between sequences i and j
N	Number of unique sequences per patch
$m_i^{h \to l}$	dispersal rate of species i from patch k to l
P	number of patches

We write the system of ODEs in matrix-vector notation

$$\frac{d\mathbf{x}}{dt} = \mathbf{R}\Sigma\mathbf{x} + \mathbf{Q}\mathbf{x} + \mathbf{M}\mathbf{x}$$

where

$$
\mathbf{R} =
\begin{bmatrix}
r_1^1 & 0 & \dots & \dots & \dots & 0 \\
0 & r_2^1 & \ddots & \ddots & \ddots & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & r_1^2 & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & \dots & \dots & \dots & r_N^P
\end{bmatrix}
$$

$$
\mathbf{\Sigma} =
\begin{bmatrix}
\left(1 - \frac{\Sigma x^1}{k_1^1}\right) & 0 & \dots & \dots & \dots & 0 \\
0 & \left(1 - \frac{\Sigma x^1}{k_2^1}\right) & \ddots & \ddots & \ddots & 0 \\
\vdots & & \ddots & \ddots & \ddots & \vdots \\
\vdots & & \ddots & \ddots & \left(1 - \frac{\Sigma x^2}{k_1^2}\right) & \ddots & \vdots \\
\vdots & & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & \dots & \dots & \dots & \left(1 - \frac{\Sigma x^P}{k_N^P}\right)
\end{bmatrix}
$$

$$
\mathbf{Q} =
\begin{bmatrix}
0 & q^{h(1,2)} & \dots & q^{h(1,N)} & 0 & \dots & \dots & 0 & \dots & 0 \\
q^{h(2,1)} & 0 & \dots & q^{h(2,N)} & 0 & \dots & \dots & 0 & \dots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\
q^{h(N,1)} & q^{h(N,2)} & \dots & 0 & 0 & \dots & \dots & 0 & \dots & 0 \\
0 & 0 & \dots & 0 & 0 & q^{h(1,2)} & \dots & q^{h(1,N)} & \dots & 0 \\
0 & 0 & \dots & 0 & q^{h(2,1)} & 0 & \dots & q^{h(2,N)} & \dots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \dots & \vdots \\
0 & 0 & \dots & 0 & q^{h(N,1)} & q^{h(N,2)} & \dots & 0 & \dots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & 0
\end{bmatrix}
$$

with $h(i,j) = h(j,i)$, and

$$
\mathbf{M} =
\begin{bmatrix}
-m_1^{1\to2} & 0 & \dots & m_1^{2\to1} & 0 & \dots & m_1^{3\to1} & 0 & \dots & 0 \\
0 & -m_2^{1\to2} & 0 & \dots & m_2^{2\to1} & 0 & \dots & m_2^{3\to1} & \dots & 0 \\
\vdots & 0 & \ddots & \ddots & \dots & \vdots & & & & \\
m_1^{1\to2} & \vdots & \ddots & -m & \ddots & \vdots & & & & \\
0 & \ddots & \ddots & \ddots & \ddots & \vdots & & & & \\
0 & 0 & \dots & \dots & \dots & \left(1 - \Sigma x^P\right) & & & &
\end{bmatrix}
$$

Steady state analysis

For the purpose of notation, we define

$$\frac{dx_i^l}{dt} \equiv f(x_i^l).$$

The Jacobian matrix

$$J_{i,j}^l = \frac{\partial f(x_i^l)}{\partial x_j^l}\bigg|_{\substack{x_i^l=x_i^{l*} \\ x_j^l=x_j^{l*}}}$$

The general Jacobian matrix is then

$$
\mathbf{J} = \begin{bmatrix}
\frac{\partial f(x_1^1)}{\partial x_1^1} & \frac{\partial f(x_1^1)}{\partial x_2^1} & \cdots & \frac{\partial f(x_1^1)}{\partial x_N^1} & \frac{\partial f(x_1^1)}{\partial x_1^2} & \frac{\partial f(x_1^1)}{\partial x_2^2} & \cdots & \frac{\partial f(x_1^1)}{\partial x_N^P} \\
\frac{\partial f(x_2^1)}{\partial x_1^1} & \frac{\partial f(x_2^1)}{\partial x_2^1} & \cdots & \cdots & \cdots & \cdots & \cdots & \frac{\partial f(x_2^1)}{\partial X_N^P} \\
\vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
\frac{\partial f(x_N^1)}{\partial x_1^1} & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
\frac{\partial f(x_1^2)}{\partial x_1^1} & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
\frac{\partial f(x_2^2)}{\partial x_1^1} & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
\vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \frac{\partial f(x_{N-1}^P)}{\partial x_N^P} \\
\frac{\partial f(x_N^P)}{\partial x_1^1} & \frac{\partial f(x_N^P)}{\partial x_1^2} & \cdots & \cdots & \cdots & \cdots & \frac{\partial f(x_N^P)}{\partial x_{N-1}^P} & \frac{\partial f(x_N^P)}{\partial x_N^P}
\end{bmatrix}
$$

where species $1:N$ in patch l are stacked as the top rows, followed by species $1:N$ in patch $l+1$, and so on.

Equations for the case $N = 2, P = 2$

$$\frac{dx_1^1}{dt} = r_1^1 x_1^1 \left(1 - \frac{\Sigma x^1}{k_1^1}\right) + qx_2^1 - qx_1^1 + m_1^{2\to1}x_1^2 - m_1^{1\to2}x_1^1$$

$$\frac{dx_2^1}{dt} = r_2^1 x_2^1 \left(1 - \frac{\Sigma x^1}{k_2^1}\right) + qx_1^1 - qx_2^1 + m_2^{2\to1}x_2^2 - m_2^{1\to2}x_2^1$$

$$\frac{dx_1^2}{dt} = r_1^2 x_1^2 \left(1 - \frac{\Sigma x^2}{k_2^2}\right) + qx_2^2 - qx_1^2 + m_1^{1\to2}x_1^1 - m_1^{2\to1}x_1^2$$

$$\frac{dx_2^2}{dt} = r_2^2 x_2^2 \left(1 - \frac{\Sigma x^2}{k_2^2}\right) + qx_1^2 - qx_2^2 + m_2^{1\to2}x_2^1 - m_2^{2\to1}x_2^2$$

So the Jacobian is

$$
\mathbf{J} =
\begin{bmatrix}
\frac{\partial f(x_1^1)}{\partial x_1^1} & \frac{\partial f(x_1^1)}{\partial x_2^1} & \frac{\partial f(x_1^1)}{\partial x_1^2} & \frac{\partial f(x_1^1)}{\partial x_2^2} \\
\frac{\partial f(x_2^1)}{\partial x_1^1} & \frac{\partial f(x_2^1)}{\partial x_2^1} & \frac{\partial f(x_2^1)}{\partial x_1^2} & \frac{\partial f(x_2^1)}{\partial x_2^2} \\
\frac{\partial f(x_1^2)}{\partial x_1^1} & \frac{\partial f(x_1^2)}{\partial x_2^1} & \frac{\partial f(x_1^2)}{\partial x_1^2} & \frac{\partial f(x_1^2)}{\partial x_2^2} \\
\frac{\partial f(x_2^2)}{\partial x_1^1} & \frac{\partial f(x_2^2)}{\partial x_2^1} & \frac{\partial f(x_2^2)}{\partial x_1^2} & \frac{\partial f(x_2^2)}{\partial x_2^2}
\end{bmatrix}
$$

$$
=
\begin{bmatrix}
r_1^1 - \frac{r_1^1}{k_1^1}\Sigma x^1 - q - m_1^{1\to2} & q & m_1^{2\to1} & 0 \\
q & r_2^1 - \frac{r_2^1}{k_2^1}\Sigma x^1 - q - m_2^{1\to2} & 0 & m_2^{2\to1} \\
m_1^{1\to2} & 0 & r_1^2 - \frac{r_1^2}{k_1^2}\Sigma x^2 - q - m_1^{2\to1} & q \\
0 & m_2^{1\to2} & q & r_2^2 - \frac{r_2^2}{k_2^2}\Sigma x^2 - q - m_2^{2\to1}
\end{bmatrix}
$$

We partition the Jacobian into intra-community interaction and meta-community dispersal components

$$
\mathbf{J} = \mathbf{J_{R\Sigma}} + \mathbf{J_Q} + \mathbf{J_M}
$$

where

$$
\mathbf{J_{R\Sigma}} =
\begin{bmatrix}
r_1^1 - \frac{r_1}{k_1}^1\Sigma x^1 & 0 & 0 & 0 \\
0 & r_2^1 - \frac{r_2^1}{k_2^1}\Sigma x^1 & 0 & 0 \\
0 & 0 & r_1^2 - \frac{r_1^2}{k_1^2}\Sigma x^2 & 0 \\
0 & 0 & 0 & r_2^2 - \frac{r_2^2}{k_2^2}\Sigma x^2
\end{bmatrix}
$$

$$
\mathbf{J_Q} =
\begin{bmatrix}
0 & q^{h(1,2)} & 0 & 0 \\
q^{h(2,1)} & 0 & 0 & 0 \\
0 & 0 & 0 & q^{h(1,2)} \\
0 & 0 & q^{h(2,1)} & 0
\end{bmatrix}
$$

$$
\mathbf{J_M} =
\begin{bmatrix}
-m_1^{1\to2} & 0 & m_1^{2\to1} & 0 \\
0 & -m_2^{1\to2} & 0 & m_2^{2\to1} \\
m_1^{1\to2} & 0 & -m_1^{2\to1} & 0 \\
0 & m_2^{1\to2} & 0 & -m_2^{2\to1}
\end{bmatrix}
$$

Shannon diversity

Shannon diversity is simply the Shannon information entropy applied to the distribution of genotypes in the environment

$$
H = -\sum_{l\in P}\sum_{i\in l} p_i^l \log p_i^l - sum
$$

where the $p_i^l \log p_i^l$ sum of is taken per genotype within environment l and then taken with respect to the environmental mean and the

We consider the disaggegared Shannon diversity for within-environment and across-environment, respectively

$$H_w = -\sum_{i \in N} \langle p_i^l \rangle_l \log \langle p_i^l \rangle_l$$

$$H_b = -\sum_{l \in P} \langle p_i^l \rangle_i \log \langle p_i^l \rangle_i$$

Classical ANOVA decomposition

The within-between decomposition of entropy is made analogously to the classical ANOVA decomposition from statistics

$$
\begin{aligned}
SS_{total} &= \sum_{l \in P} \sum_{i \in l} \left(x_i^l - \langle\langle x \rangle_i \rangle_l \right)^2 \\
&= \sum_{i \in l} \left(x_i^l - \langle x^l \rangle_i \right)^2 + \sum_{i \in N} \left(\langle x^l \rangle_i - \langle\langle x \rangle_i \rangle_l \right)^2 \\
&= SS_{within} + SS_{between}
\end{aligned}
$$

Similar to entropy, the hypothesis is that mutation, dispersal, and selection all contribute uniquely to these terms. Via generation of within-environment variance, mutation positively drives SS_{within} upward, while selection directly opposes mutation via mutation-selection balance. Dispersal, on the other hand, positively drives SS_{within} by effectively converting $SS_{between}$ to SS_{within}. Note that $SS_{between}$ can only be generated by in this model by variation in selection among environments.

Genetic drift and neutral mutations

Genetic drift is not included in this model, not the capacity for gene mutation to be selectively neutral and therefore contribute to genetic entropy.

Competitive exclusion

In a deterministic environment with no dispersal or mutation, the species with the highest carrying capacity (lowest R*)

$$\frac{dx_1}{dt} = rx_1\left(1 - \frac{\Sigma x}{k_1}\right)$$

$$\frac{dx_2}{dt} = rx_2\left(1 - \frac{\Sigma x}{k_2}\right)$$

$$\frac{dx_3}{dt} = rx_3\left(1 - \frac{\Sigma x}{k_3}\right)$$

in which case the genotype with the largest k will outcompete the others and go to fixation

$$t \to \infty \implies \frac{x_{\max k}}{\Sigma x} \to 1$$

and accordingly

$$\frac{x_{\max k}}{\Sigma x} \to 1 \implies H \to 0$$

Competitive exclusion dynamics for the 3 genotype system with no mutation and no dispersal is given in Figure X.
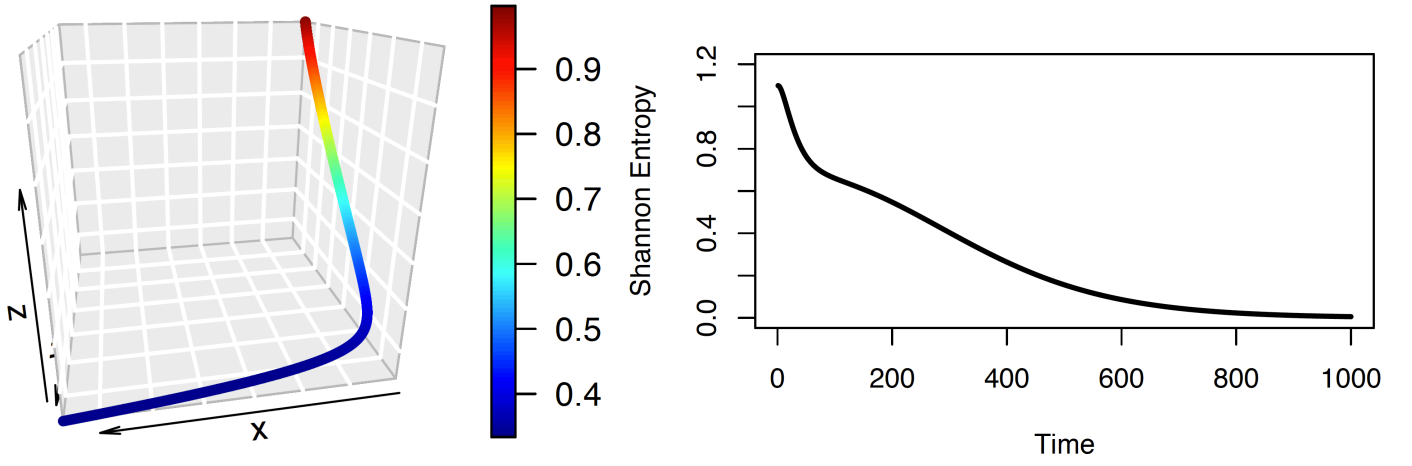
Evolutionary tradeoffs

Figure 1: Competitive exclusion among three genotypes in an environment with no dispersal and no mutation.
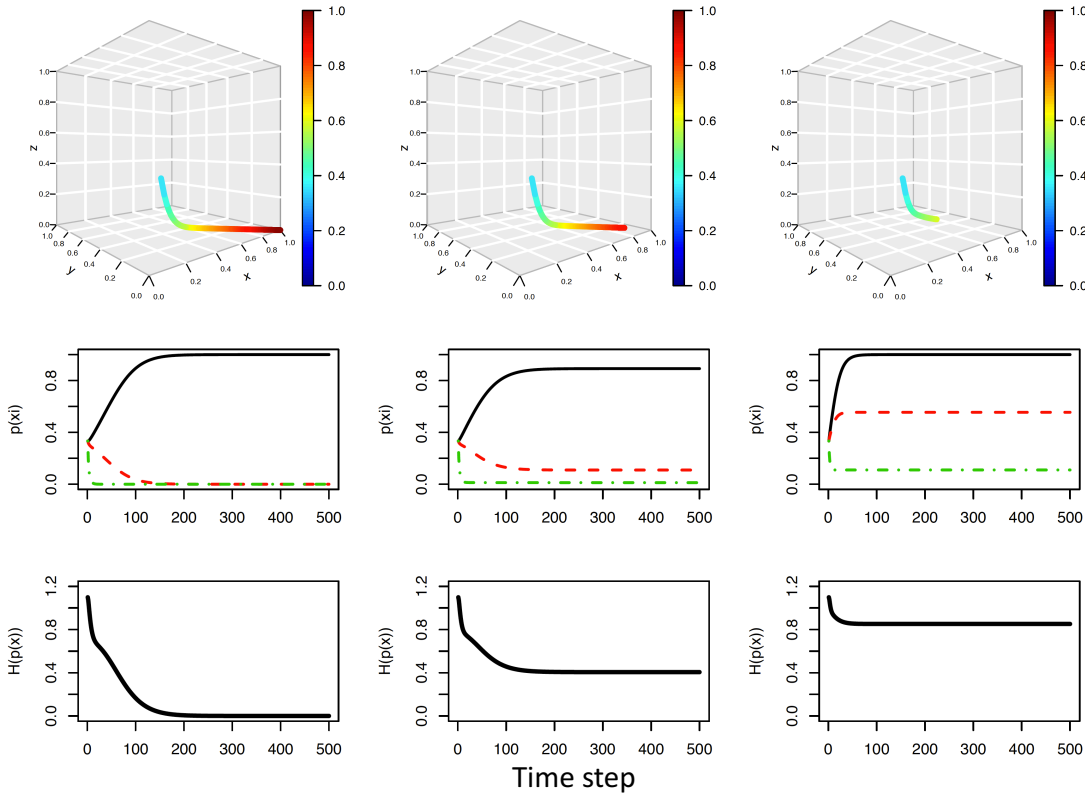

Figure 2: Competitive exclusion among three genotypes with and without mutation and no dispersal. Left column is with 0 mutation, while the middle and right columns give the steady state with $q = 0.1$ and $q = 1$, respectively.

8