Are there species smaller than 1mm?

Alan McKane

Theoretical Physics Division, University of Manchester

Warwick, September 9, 2013

OUTLINE

- Introduction

- Model

- Analysis and Results

- Conclusions

with

Tim Rogers and Axel Rossberg

1. Introduction

The question of how species form is a perennial one in biology. But an even more fundamental question is "What exactly constitutes a species?"

An undergraduate textbook by Ridley (Blackwell, 1993) discusses seven different definitions of species (phenetic, biological, recognition, ecological, cladistic, pluralistic and evolutionary), and concludes that a combination of four (biological, recognition, ecological and cladistic) is ideal.

King (CUP, 1993) has eight (morphological, biological, recognition, cohesion, evolutionary, cladistic, ecological and phylogenetic).

The pub quiz (or university challenge) answer to the question "What is a species?" would probably be the "biological species" concept — popularised by Mayr — in sexually reproducing populations, biological species are distinct groups of organisms that can mate successfully only with other members of the same group.

But there are situations where members of group A can mate with group B, and B with C, but not A with C, as is the case for so-called ring species.

Famously, in the *Origin of Species* Darwin did not explain the origin of species — he argued for evolution through natural selection, and evolution is not concerned with species, but with individuals.

The survival and reproduction of those individuals who are best adapted to their environment determines the characteristics of subsequent populations, but neither the process nor the theory require that these populations be organised into species.

The formation of species, when it occurs, is a phenomenon that needs to be explained. It may be that it was historically a useful concept for humans who used it to classify different types of animals and plants they saw around them, but in the realm of meio- or microfauna there may be no distinct gaps between groups of individuals which allow for the definition of species.

In fact Darwin himself was of the opinion that "In short, we shall have to treat species in the same manner as those naturalists treat genera, who admit that genera are merely artificial combinations made for convenience. This may not be a cheering prospect, but we shall at least be freed from the vain search for the undiscovered and undiscoverable essence of the term species".

Here I'm going to suggest that the clustering of individuals in trait or gene space that we recognise as species is a pattern that emerges from underlying ecological and evolutionary mechanisms.

Just as mixtures of chemical constituents which react and diffuse may create patterns (for instance, spots and stripes), so individuals which react (for example, compete) and diffuse (for example, mutate in trait or gene space) may create patterns (clusters).

My approach will be that of a theoretical physicist: construct a simple model to see if the effect appears, and see if we can understand it from the model.

The model will have two different variants, to ensure that the results aren't specific to a particular situation. The species that emerge will be what are referred to as ecological species and genetic species.

An ecological species or eco-species is one localised in niche space; its position determined by a particular value of a phenotypic trait.

A genetic species is a collection of organisms whose genomes are very similar to each other, but distinct from those of organisms belonging to other species. The genetic separation of species is frequently called the DNA barcoding gap.

Variant I of the model will have individuals defined by particular trait and variant II by the a particular genome.

The effect of noise coming from the discreteness of the individuals will be crucial, so the model will be formulated as an individual based model (IBM) and then studied using analytic methods and simulation.

2. Model

Organisms.

• In the phenotypic variant I of the model, organisms have a single trait, so that each individual in the model is entirely characterised by an ecotype x, where x is a real number.

• In the genotypic variant II, each individual is characterised by an N-dimensional vector of ones and zeros, with each entry of the vector being interpreted as distinguishing two viable alleles.

Population. At time t there are $M(t)$ organisms in the community.

• In variant I they are specified by their ecotypes $x_1, \ldots, x_{M(t)}$.

• In variant II they are specified by N-dimensional vectors $\mathbf{x}_1, \ldots, \mathbf{x}_{M(t)}$.

In order to define the processes of competition and reproduction we need a criterion which specifies how close two individuals are to each other.

This 'distance' between individuals i and j will be taken to be:

• The distance between their ecotypes, $D_{ij} = |x_i - x_j|$ in variant I.

• In variant II, D_{ij} will be the number of sites at which the genomes of the two individual differ (the Hamming distance).

Birth. Each organism reproduces asexually at the same rate, and we choose units of time so that this rate equals unity. The characteristics of the offspring are chosen probabilistically according to a distribution centred on the parent, but falling off with distance D_{ij}.

• In variant I the ecotype of the offspring is chosen from a normal distribution with variance μ centred on the ecotype of the parent.

• In variant II, each of the N loci will independently mutate with probability μ.

Competition. The strength of competition between two individuals i and j is taken to be a function $g(D_{ij})$ (the competition kernel) which decreases with D_{ij} to reflect increased competition between similar individuals.

	Variant I	**Variant II**
Genotype/Phenotype	Real number, x	Binary vector, e.g. $\mathbf{x} = (0, 1, 1, 1, 1, 0, 0, 1)$
Similarity measure	$D_{ij} = \|x_i - x_j\|$	D_{ij} = number of elements $\mathbf{x}_i$ and $\mathbf{x}_j$ have in common
Birth rate	$b(x_i) = 1$ for all i	$b(\mathbf{x}_i) = 1$ for all i
Effect of Mutation	Add a Gaussian random number of variance μ	Each bit independently flipped with probability μ
Death rate	$d(x_i) = \frac{1}{K} \sum_j g(D_{ij})$	$d(\mathbf{x}_i) = \frac{1}{K} \sum_j g(D_{ij})$

Summary of the model definitions for variant I and II. In both cases the only parameters are the carrying capacity K, mutation strength μ and competition kernel g.

Various functional forms for the dependence of competition on distance D_{ij} have been studied, for instance a normal distribution with standard deviation w or a 'top-hat' which is a non-zero constant only when $|D_{ij}| < w$.

The stochastic fluctuations in this model will have an important role to play, but for variant I we can take the limit $K \to \infty$ and find the deterministic analogue of the model.

The population density at point x in niche space at time t, $\phi(x, t)$, obeys the deterministic equation

$$\frac{\partial}{\partial t}\phi(x, t) = \phi(x, t) + \mu\, \frac{\partial^2}{\partial x^2}\phi(x, t) - \int \phi(x, t)\phi(y, t)g(x - y)\, dy$$

The homogeneous state is a fixed point of this equation. A typical analysis to search for patterns considers the effect of small perturbations to the homogeneous state. The perturbations may die out and the population remains evenly spread, or they may grow, signalling the possible existence of stable patterns.

3. Analysis and Results

The mathematical analysis proceeds as follows:

• Write down a master equation for the probability of finding a particular distribution of individuals. For variant I this distribution is

$$\phi(x) = \frac{1}{K} \sum_{i=1}^{M(t)} \delta\left(x - x_i\right),$$

at a given time t.

• Expand in powers of K^{-1} and keep only the first two terms in the expansion. This gives a Fokker-Planck equation.

The Fokker-Planck equations which are obtained are:

• For variant I :
$$\frac{\partial P(\phi, t)}{\partial t} = -\int dx \, \frac{\delta}{\delta\phi(x)} \left[\mathcal{A}_x\left(\phi\right) P(\phi, t)\right]$$
$$+ \frac{1}{2K} \int \int dx \, dx' \, \frac{\delta^2}{\delta\phi(x)\delta\phi(x')} \left[\mathcal{B}_{x,x'}\left(\phi\right) P(\phi, t)\right],$$

where

$$
\begin{aligned}
\mathcal{A}_x(\phi) &= \int dy \, [r(x-y)\phi(y) - \phi(x)g(x-y)\phi(y)] \\
\mathcal{B}_{x,x'}(\phi) &= \int dy \, [r(x-y)\phi(y) + \phi(x)g(x-y)\phi(y)] \, \delta(x-x').
\end{aligned}
$$

- For variant II :
$$
\frac{\partial P(\mathbf{x}, t)}{\partial t} = - \sum_I \frac{\partial}{\partial x_I} \left[\mathcal{A}_I(\mathbf{x}) \, P(\mathbf{x}, t) \right]
$$
$$
+ \frac{\kappa}{2} \sum_{I,I'} \frac{\partial^2}{\partial x_I \partial x_{I'}} \left[\mathcal{B}_{I,I'}(\mathbf{x}) \, P(\mathbf{x}, t) \right],
$$

where

$$
\begin{aligned}
\mathcal{A}_I(\mathbf{x}) &= \sum_J \left[r_{IJ} x_J - x_I g_{IJ} x_J \right] \\
\mathcal{B}_{I,I'}(\mathbf{x}) &= \sum_J \left[r_{IJ} x_J + x_I g_{IJ} x_J \right] \delta_{I,I'}.
\end{aligned}
$$

• Write down the stochastic differential equation which is equivalent to these Fokker-Planck equations.

For variant II this is (in the Itō sense)

$$\frac{dx_I}{dt} = \sum_J \left(r_{IJ}x_J - x_I g_{IJ}x_J \right) + \left[\kappa \sum_J \left(r_{IJ}x_J + x_I g_{IJ}x_J \right) \right]^{1/2} \eta_I(t),$$

where the $\eta_I(t)$ are independent Gaussian white noise variables with zero mean and unit variance, that is,

$$\langle \eta_I(t)\eta_J(t') \rangle = \delta(t - t')\delta_{I,J}.$$

• It is easier to use transformed variables (a standard Fourier transform for variant I; a Fourier transform on the space of binary sequences for variant II). So transform to this system of variables.

The stochastic differential equations in these transformed variables for the two variants look very similar.

In addition to simulations of the original model, we have performed two types of analysis:

1) We have performed a linear noise approximation, by writing (in the case of variant I, for instance)

$$\phi(x, t) = \phi^* + \frac{1}{\sqrt{K}} \zeta(x, t),$$

where ϕ^* is the homogeneous fixed point of the deterministic differential equation.

Only linear terms in the noise are retained, and so the resulting stochastic differential equation can be solved. This approximation assumes moderate to large K.

2) If the competition kernel is flat ($g(x) = 1$), then we may adiabatically eliminate the fast mode corresponding to the mean population size, and obtain a set of corresponding stochastic differential equations for the remaining modes.

In the limit $K \to \infty$ and $\mu \to 0$, but with $\tau \equiv \mu K$ finite, we can analyse these equations.

For example, we find for the covariance in trait space for variant I:

$$\Xi(x) \equiv \int \langle \phi(y)\phi(y - x) \rangle_{\text{st}} \, dy = \sum_{n=-\infty}^{\infty} \frac{1}{1 + n^2\tau} \, e^{inx},$$

with an analogous expression for variant II.

The simulations and analytic methods all show that

- The macroscopic limit (for example, the differential equation in the case of variant I) does not give a correct description of the system — the effects of noise cause completely different behaviour.
- Instead we find that when τ is small, tight clusters form. As τ increases, clusters are still seen, but they become less well defined. Finally, at large τ the homogeneous state is stable to the effects of demographic noise.
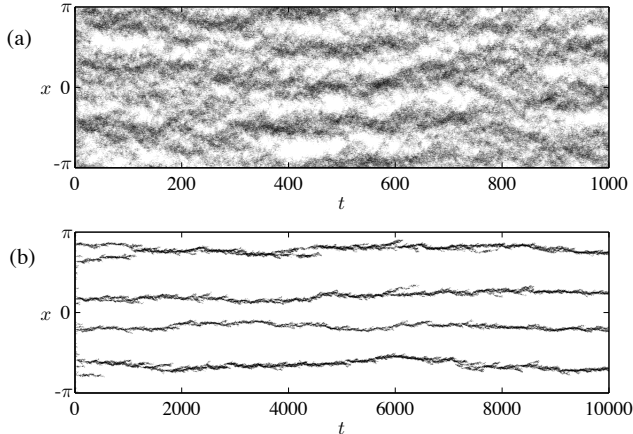
Noise-induced pattern formation in simulations of variant I of the model with $\mu = 10^{-3}$ and $K = 10^3$ in (a) and $\mu = 10^{-5}$ and $K = 10^3$ in (b). In both cases the macroscopic theory predicts that patterns should not form.
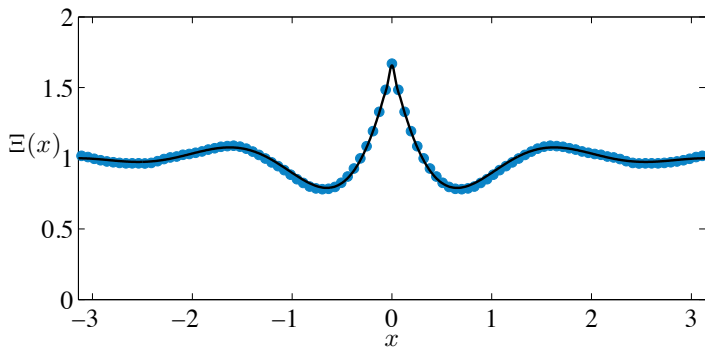
Covariance with $\mu = 10^{-3}$ and $K = 10^3$ as in (a) above. Blue circles show the results of simulations (100 runs and 100 bins) and the black line the theoretical prediction of the linear noise approximation.
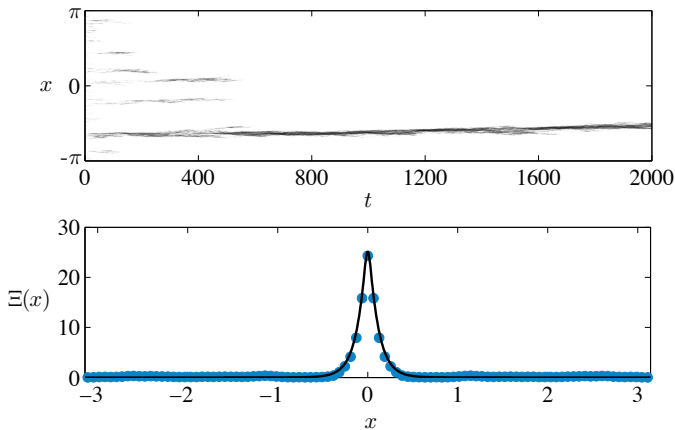
Covariance with $\mu = 10^{-5}$ and $K = 10^3$ as in (b) above, with a flat competition kernel. Blue circles show the results of simulations (100 runs and 100 bins) and the black line the theoretical prediction from the adiabatic elimination procedure.

4. Conclusions

Demographic noise can lead to the spontaneous formation of species in simple models, whether the individuals are characterised by a continuous trait or by a binary sequence.

Species are formed if $\tau = \mu K$ is small, and gradually become less defined as τ increases. For macrofauna (size >1mm) the carrying capacity, K, is sufficiently small that we expect $\tau \ll 1$, so the prediction would be that species would form.

For meio- and microfauna (size <1mm), K is typically large enough to expect that τ will be greater than 1 and the prediction is that species will not form.

Obviously more complex models need to be investigated to see if the effect persists.

However, ultimately the issue will only be settled through experiment. Luckily, the rapid advance of genetic sequencing technology should allow the question to be answered.

In a recent paper we have carried out a preliminary investigation from published data which indicates organisms with a size less than about 1mm may not form distinct species.

Publications. *Emergence of species*:

"Demographic noise can lead to the spontaneous formation of species" Europhys. Lett. **97**, 40008 (2012)

"Spontaneous genetic clustering in populations of competitive organisms" Phys. Biol. **9**, 066002 (2012)

"Are there species smaller than 1mm?" Proc. Roy. Soc. B **280**, 20131248 (2013)

General (non-mathematical) review of stochastic formulation of models:

"Stochastic formulation of ecological models and their applications". Trends Ecol. Evol. **27**, 337 (2012)