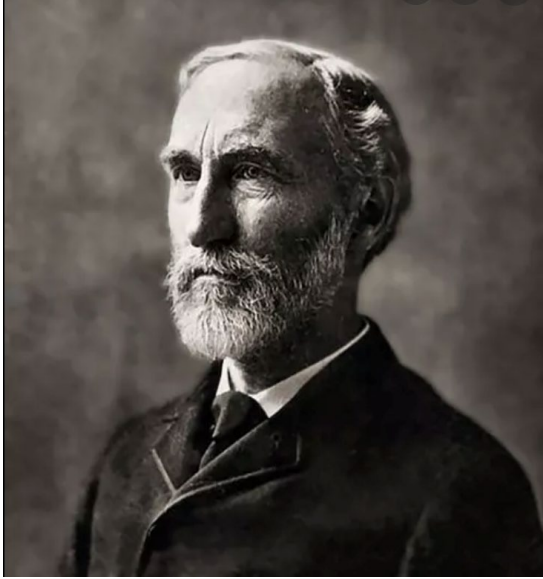




Integrations of Theories in Ecology: From the replicator equation to macroecology.

Pablo A. Marquet
Departamento de Ecología,
P. Universidad Católica de Chile

Why theory and why integration of theories?



Pierre Duhem

...there are no experiments without instruments, and no instruments without theory "without theory it is impossible to regulate a single instrument or to interpret a single reading".



Willard Van Orman Quine (Two dogmas of empiricism, 1951)

"theoretical-character" of the observation. Without theory there is no observation. All observations are created and interpreted within the framework of a pre-existing theory, so the observation has no life of its own. ("Philosophy of science is philosophy enough")

Integration or unification?

The current development of ecology and biogeography is far from unification, we still have many alternative theories, and hypothesis for empirical facts.

Integration of theories seems desirable, and more likely to be achieved in the short term.

It a humble objective for a relatively young and complex science.

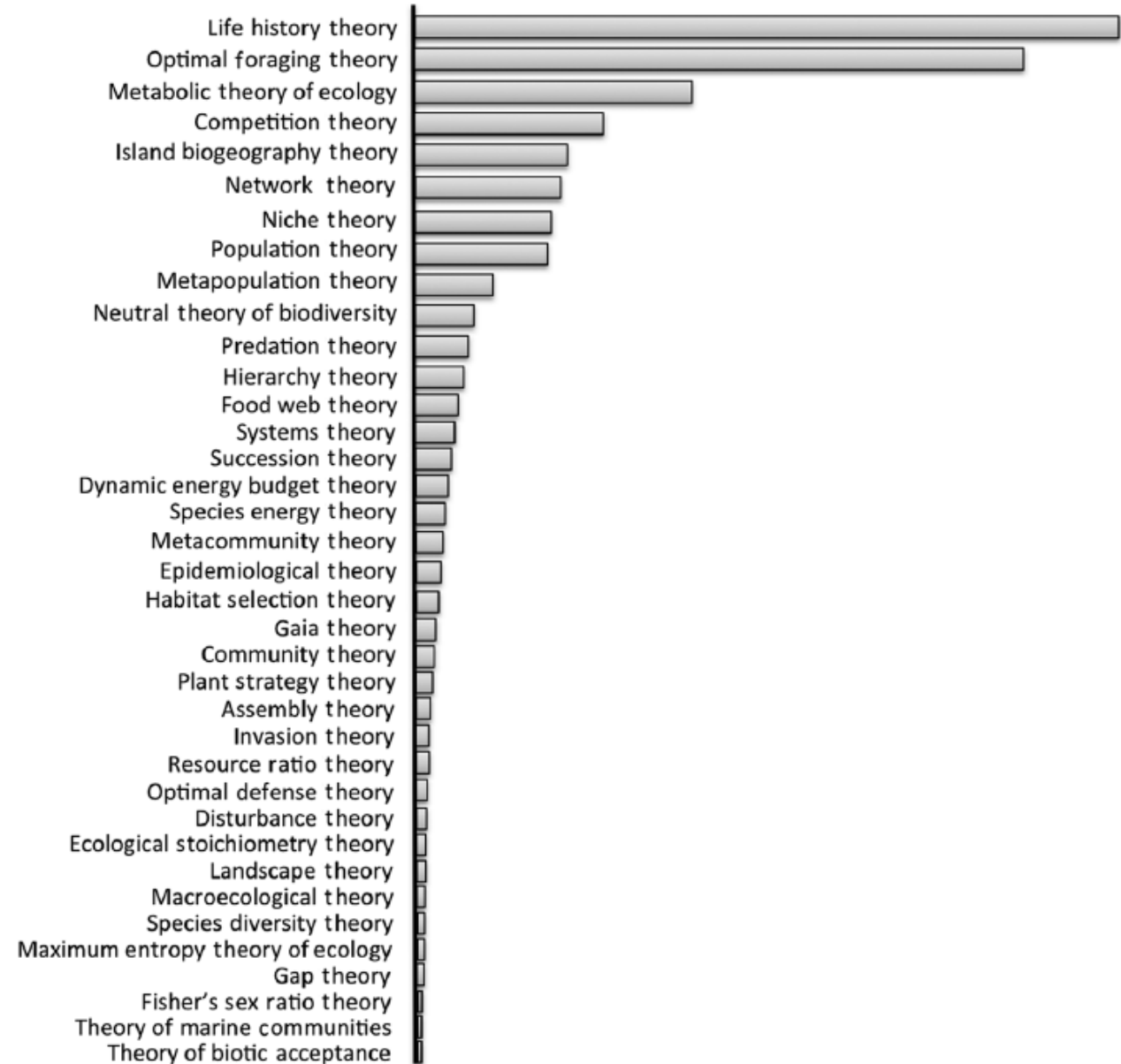
Integration is not innocuous. It can bring into light contradictions between theories.

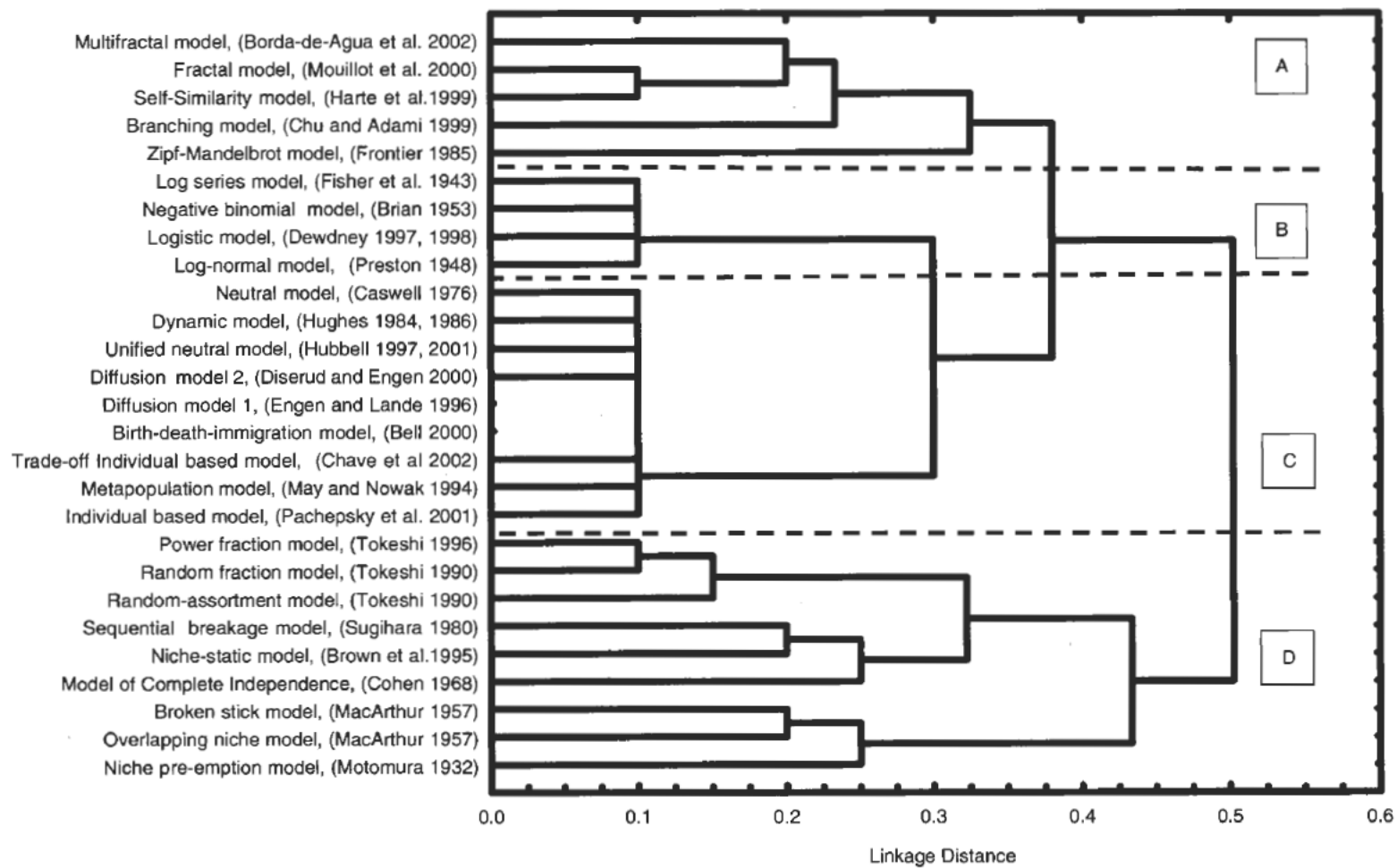
Thus, integration is a first step to eventual unification. It is a proof of concept sort to speak.

Ecology is awash in theories

*“Ecologists have been reluctant to place their observations
And their findings in the frame of a general theory”*

Margalef (1963)

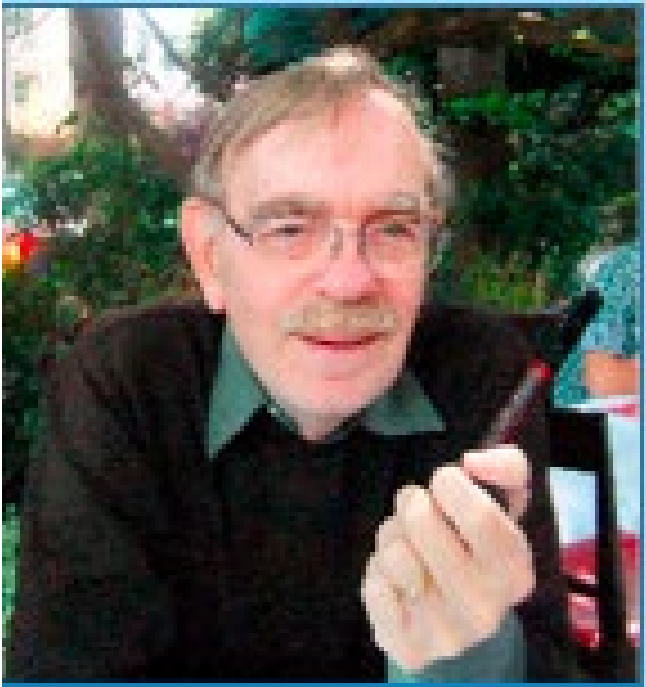




The underdetermination of theories (The Duhem-Quine Thesis)

About the limitations of empirical evidence and the rules of scientific method as a constraint on our acceptance or rejection of scientific theories.

- **It holds that:** *for any theory, T , and any given body of evidence supporting T , there is at least one rival (i.e. contrary) to T that is as well supported as T*
- **In sum, the physicist can never subject an isolated hypothesis to experimental test, but only a whole group of hypotheses; when the experiment is in disagreement with his predictions, what he learns is that at least one of the hypotheses, constituting this group is unacceptable and ought to be modified; but the experiment does not designate which one should be changed.**



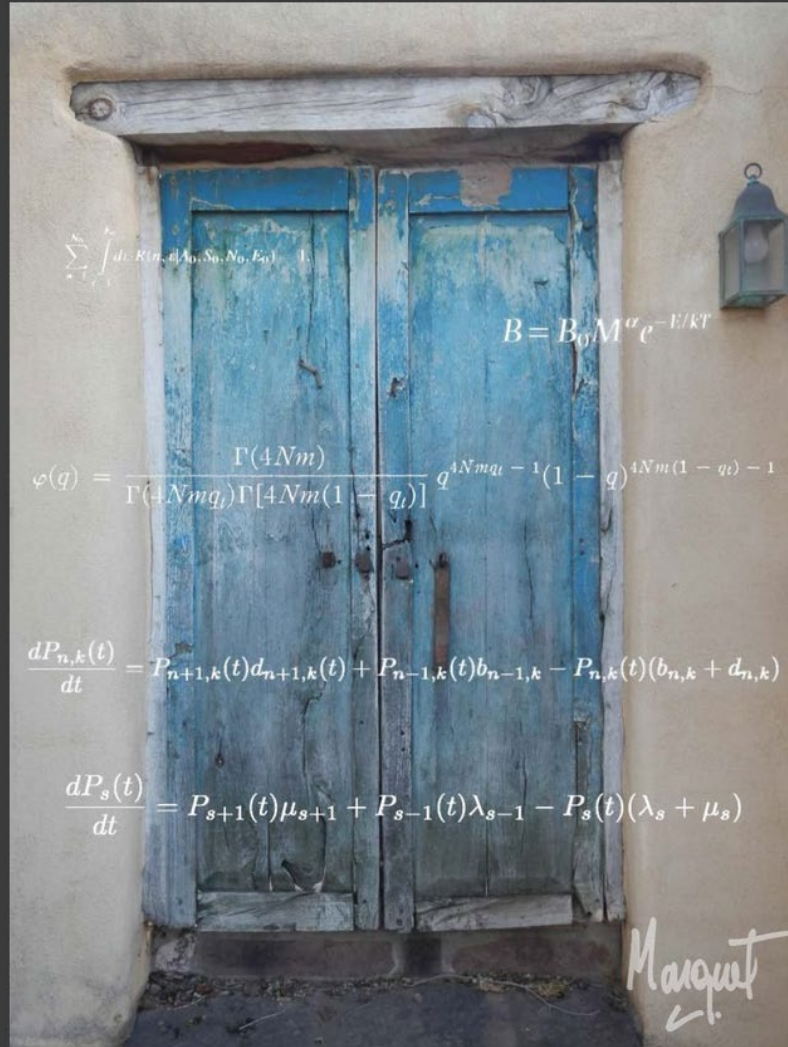
A good theory is more efficient than its rivals. It produces more and better explanations and predictions with the same number of free parameters or constructs (Laudan 1977).

The evaluation of theories is a comparative matter (Laudan 1977), and an important criterion for comparison is efficacy.

Welcome to Neti



Second NETI meeting this coming September at the Santa Fe Institute in Santa Fe New Mexico



On Theory in Ecology

PABLO A. MARQUET, ANDREW P. ALLEN, JAMES H. BROWN, JENNIFER A. DUNNE, BRIAN J. ENQUIST, JAMES F. GILLOOLY, PATRICIA A. GOWATY, JESSICA L. GREEN, JOHN HARTE, STEVE P. HUBBELL, JAMES O'DWYER, JORDAN G. OKIE, ANNETTE OSTLING, MARK RITCHIE, DAVID STORCH, AND GEOFFREY B. WEST

We argue for expanding the role of theory in ecology to accelerate scientific progress, enhance the ability to address environmental challenges, foster the development of synthesis and unification, and improve the design of experiments and large-scale environmental-monitoring programs. To achieve these goals, it is essential to foster the development of what we call efficient theories, which have several key attributes. Efficient theories are grounded in first principles, are usually expressed in the language of mathematics, make few assumptions and generate a large number of predictions per free parameter, are approximate, and entail predictions that provide well-understood standards for comparison with empirical data. We contend that the development and successive refinement of efficient theories provide a solid foundation for advancing environmental science in the era of big data.

Keywords: theory unification, metabolic theory, neutral theory of biodiversity, maximum entropy theory of ecology, big data

BioScience 64: 701–710. (2014)



Efficient theories are effective!!

“Effective scientific theories magnify understanding, help supply legitimate explanations, and assist in formulating predictions.”

Efficient theories

- Grounded in first principles
- Expressed in mathematical language
- Explain, predict and agree with empirical evidence
- Number of predictions \ggg free parameters

Science is built up of facts, as a house is built of stones; but an accumulation of facts is no more a science than a heap of stones is a house.

Poincaré

First principles

“quantitative law-like postulates about processes underlying a given class of phenomena in the natural world with well-established validity, both theoretical and empirical (i.e., core knowledge)” Marquet et al. (2015)
(e.g. Pauli exclusion principle, laws of thermodynamics etc)”

- Openness. When we define a system in order to analyze its dynamics, or make any other “ontological commitment”, we inevitably left something out. All systems are open systems. The right way to analyze open systems is by incorporating stochastic fluctuations.
- Kinetics. Biological rates are a function of temperature.

Openness

MacArthur & Wilson (1963)

$$S = bA^k$$

AN EQUILIBRIUM THEORY OF INSULAR ZOOGEOGRAPHY

ROBERT H. MACARTHUR¹ AND EDWARD O. WILSON²

THE THEORY OF

Island Biogeography

ROBERT H. MACARTHUR

EDWARD O. WILSON

MONOGRAPHS IN POPULATION BIOLOGY - 1

$$(3-4) \quad \frac{dP_s(t)}{dt} = -(\lambda_s + \mu_s)P_s(t) + \lambda_{s-1}P_{s-1}(t) + \mu_{s+1}P_{s+1}(t).$$

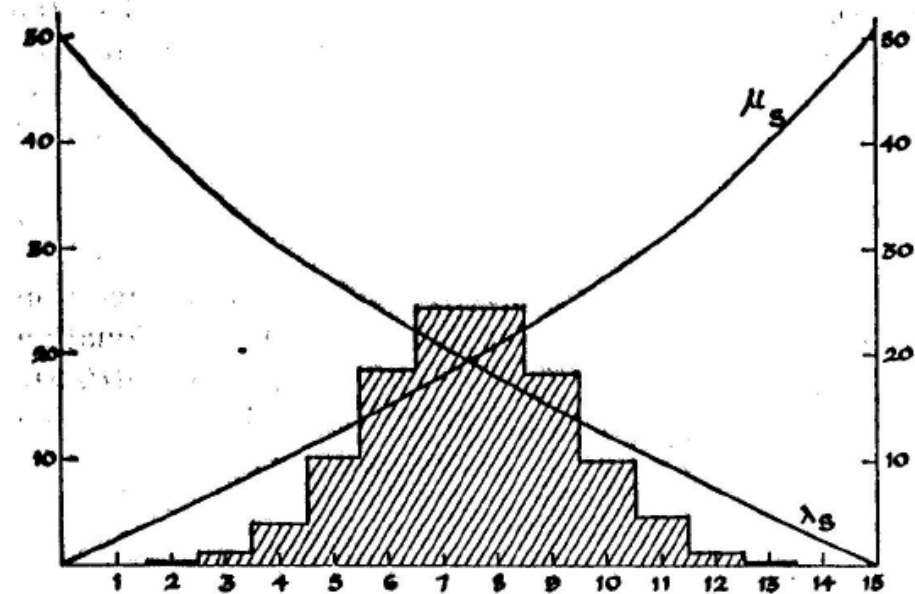


FIGURE 19. A particular case of a predicted distribution of numbers of species on a family of island biotas all with identical extinction and immigration curves and all having had time to reach equilibrium. The histogram represents the number of islands with each number of resident species in an equilibrium situation. The species pool from which the biotas were assembled contained 15 species. If the immigration and extinction curves were straighter, the variance of equilibril species numbers would be even greater; yet this large variance is still consistent with the equilibril condition.

"In principle one could solve eq. 3-4....for our purpose is more useful to find the mean $M(t)$ and the variance, $var(t)$, of the number of species at time t . These can be estimated in nature by measuring the mean and variance in number of species of a series of islands of about the same distance and are and hence of the same λ and μ ." pp.33-34



Master equation for the probability of observing k species with n individuals

$$\frac{dP_{n,k}(t)}{dt} = P_{n+1,k}(t)d_{n+1,k}(t) + P_{n-1,k}(t)b_{n-1,k} - P_{n,k}(t)(b_{n,k} + d_{n,k})$$

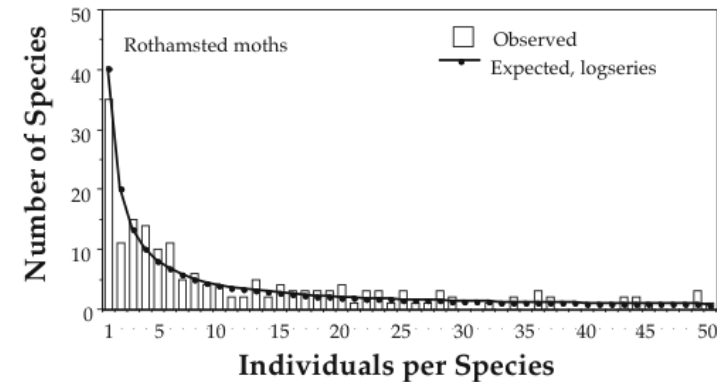
Assumptions:

- i) The species are assumed to be demographically identical, i.e. $b_{n,k} = b_n$ and $d_{n,k} = d_n$.
- ii) Density independent case, i.e. $b_n = b * n$ and $d_n = d * n$ ($n > 0$)

Fisher's Log-series distribution

$$\langle \Phi_n \rangle = \theta \frac{x^n}{n}$$

where $x = b/d$ and $\theta = SP_0\nu/d$ biodiversity parameter.



Neutral theory and relative species abundance in ecology

ME for the number of species within communities (islands)

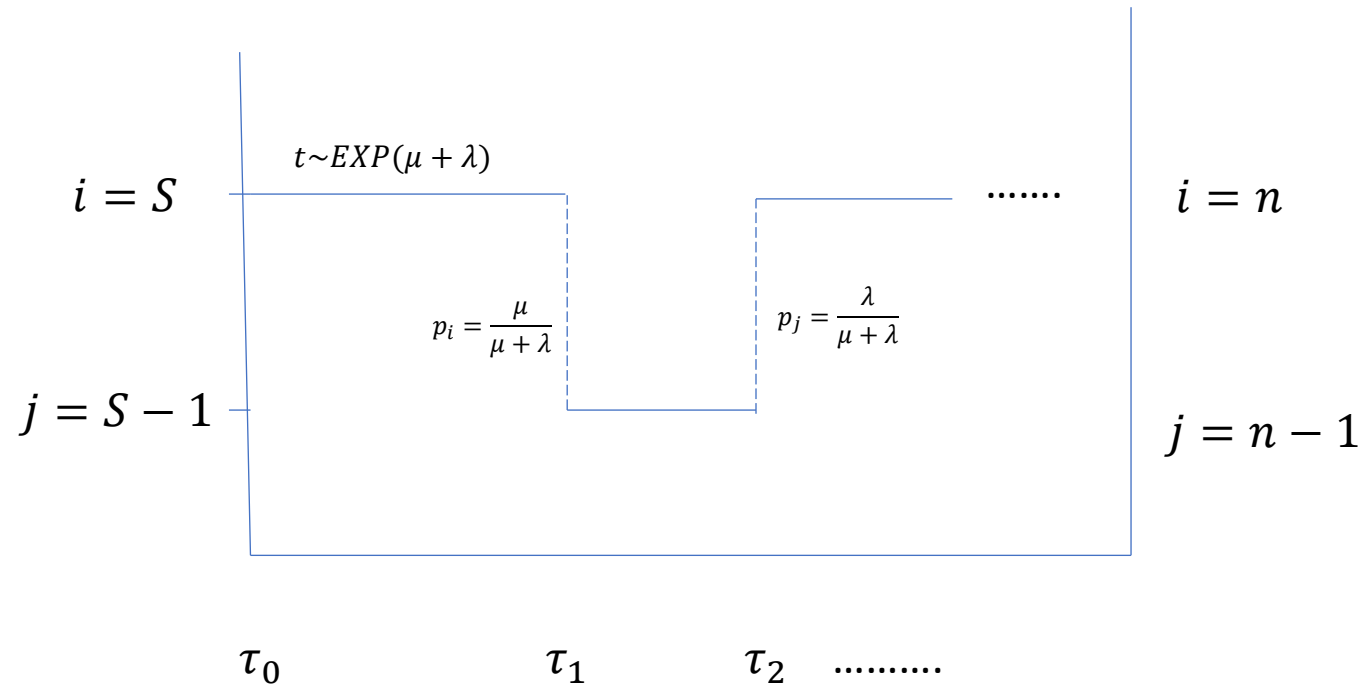
$$\frac{dP_s(t)}{dt} = P_{s+1}(t)\mu_{s+1} + P_{s-1}(t)\lambda_{s-1} - P_s(t)(\lambda_s + \mu_s),$$

ME for the number of individuals within species

$$\frac{dP_{n,k}(t)}{dt} = P_{n+1,k}(t)d_{n+1,k}(t) + P_{n-1,k}(t)b_{n-1,k} - P_{n,k}(t)(b_{n,k} + d_{n,k})$$

They cannot be true at the same time!

$$\frac{dp_{s,k}}{dt} = p_{s+1,k}(t)\mu_{s+1,k} + p_{s-1,k}(t)\lambda_{s-1,k} - p_{s,k}(t)(\lambda_{s,k} + \mu_{s,k})$$



The Markov process starts with $X(0) = S$. It waits there an exponential time of parameter

$(\mu + \lambda)$ and then jumps at time τ_1 to the new state $j = S - 1$ with probability $p_i = \frac{\mu}{\mu + \lambda}$ so that $X(\tau_1) = j$

for any $t \in \mathbb{R}_+$, let $(N_1, \dots, N_K)(t)$ be the abundance vector of the local community, i.e., $\{N_k(t)\}_{t \in \mathbb{R}_+}$ is the stochastic process accounting for the number of individuals of the species k present in the focal island through time.

let $\{S(t)\}_{t \in \mathbb{R}_+}$ be the stochastic process accounting for the number of species present in a focal island A , say, and consider $K = \text{pool}$.

Then, $S(t) = \sum_{k \in A} \mathbf{1}_{[1, \infty[}(N_k(t))$, where $\mathbf{1}_{[0, \infty[}$ denotes the indicator function of $[1, \infty[$ describes the number of species living (or dying) inside the island A .

Then, the event $\{s \text{ species at time } t\}$ now depends upon a large number of possibilities of the abundance vector $(N_1, \dots, N_K)(t)$:

$$\{S(t) = s\} = \bigcup_{k_1, \dots, k_s} \{(N_{k_1}(t), \dots, N_{k_s}(t)) \in [1, \infty[^s\}$$

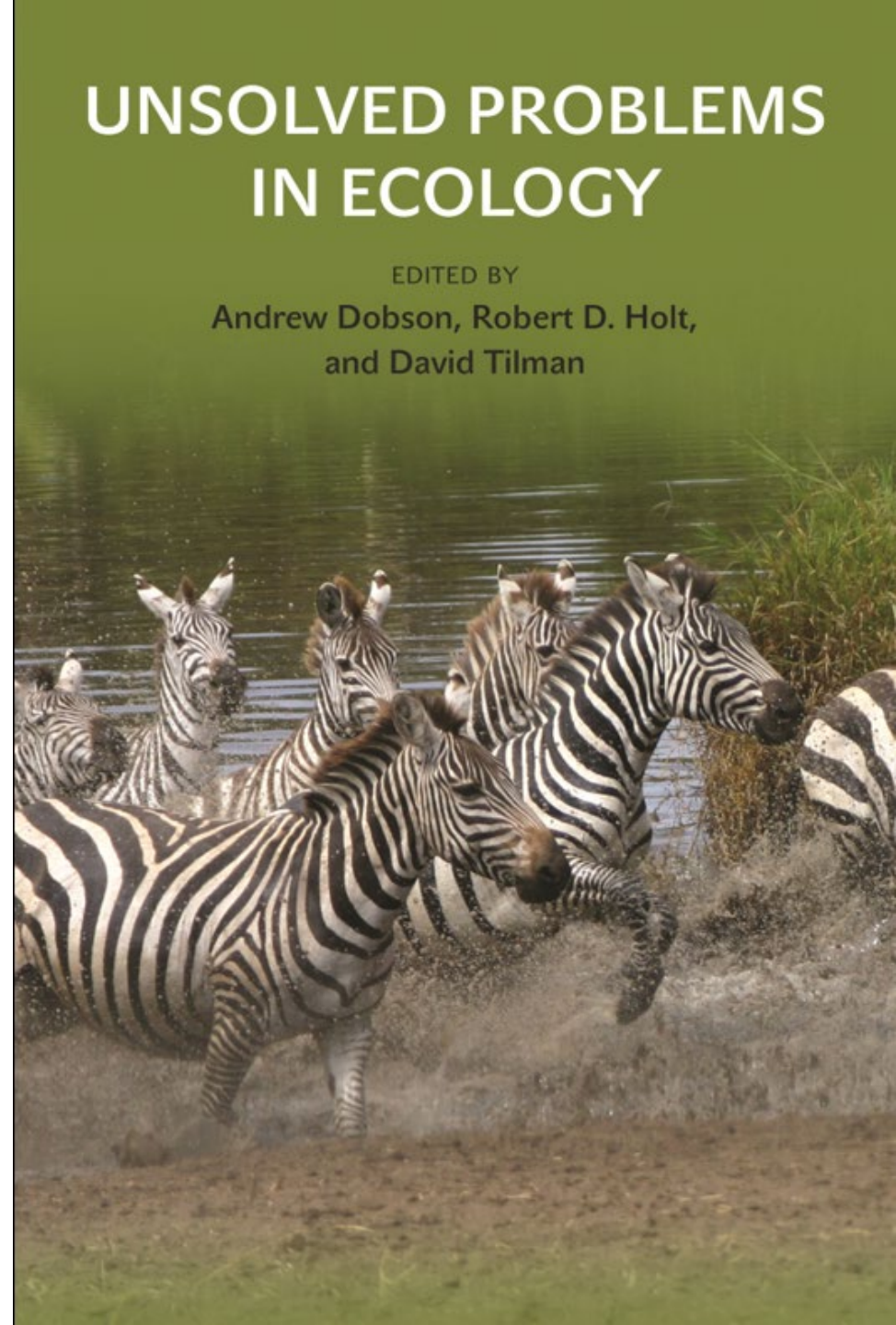
so that times between transitions $S \rightarrow S \pm 1$ are not longer exponential distributed.

That is, $S(t)$ is a function is a function of a Markov process, but it is not Markov itself.

UNSOLVED PROBLEMS IN ECOLOGY

EDITED BY

Andrew Dobson, Robert D. Holt,
and David Tilman



Diffusion processes



Sewall Wright



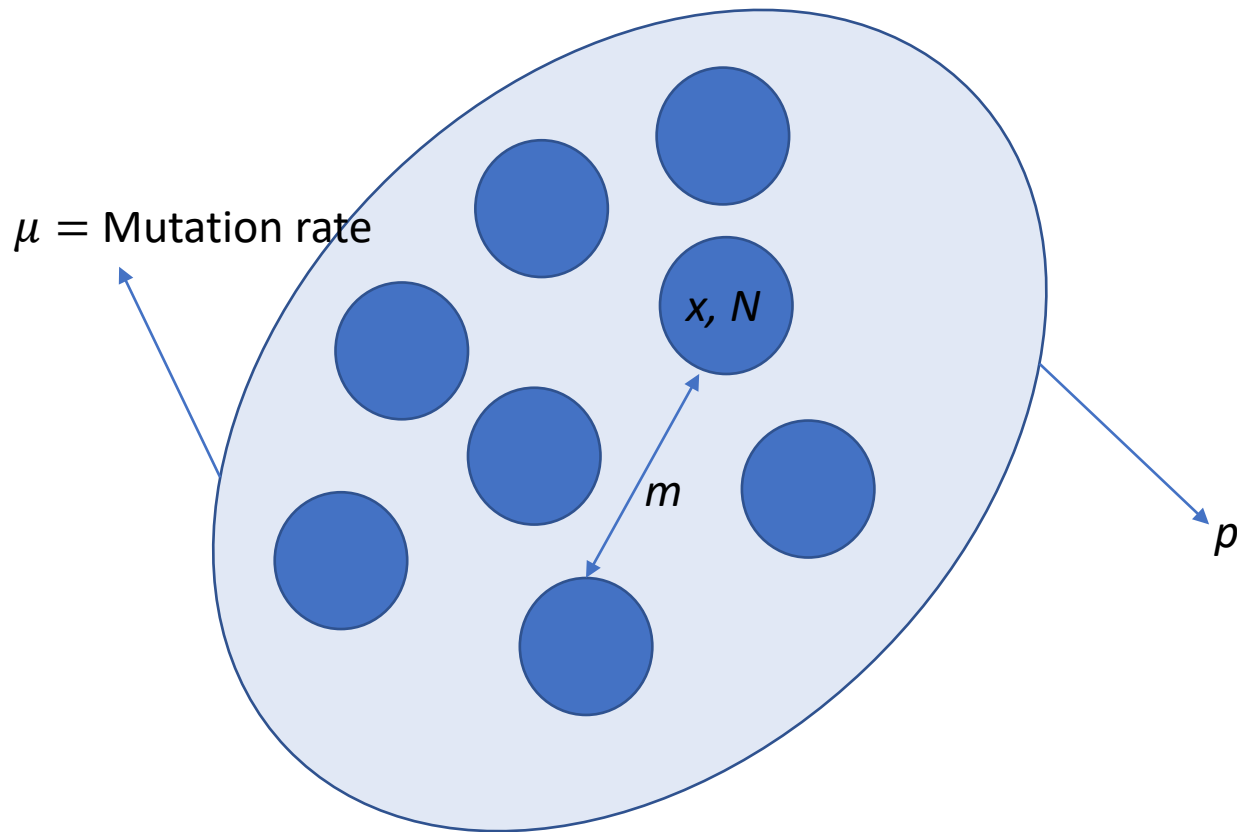
Andrey Kolmogorov



Ronald Fisher

- The frequency of genes in a structured population

$$\rho_{\infty}(x) = \frac{\Gamma(4Nm)}{\Gamma(4Nmp)\Gamma(4Nmq)} x^{4Nmq-1} (1-x)^{4Nmp-1}.$$



x = Frequency of a given allele in a local population

N = Effective population size

m = Proportion of migrating individuals among population each generation.

p = Frequency of a given allele in the total population

Kolmogorov (1935) Dokl. Akad. Nauk SSSR 3 (1935), 129-132.

Large population of N individuals consisting of s partial populations with n individuals each ($N=sn$)

Each generation k individuals disperse randomly across the partial populations.

\bar{p} = gene frequency in the large population

p = gen frequency in a local population

Δp = change in the frequency of p over one generation

$$q = 1 - p$$

Following Wright and Fisher:

$$M(p) = E(\Delta p) = \frac{k}{n}(\bar{p} - p) \quad \text{Var}(p) = E(\Delta p)^2 = \frac{pq}{2n}$$

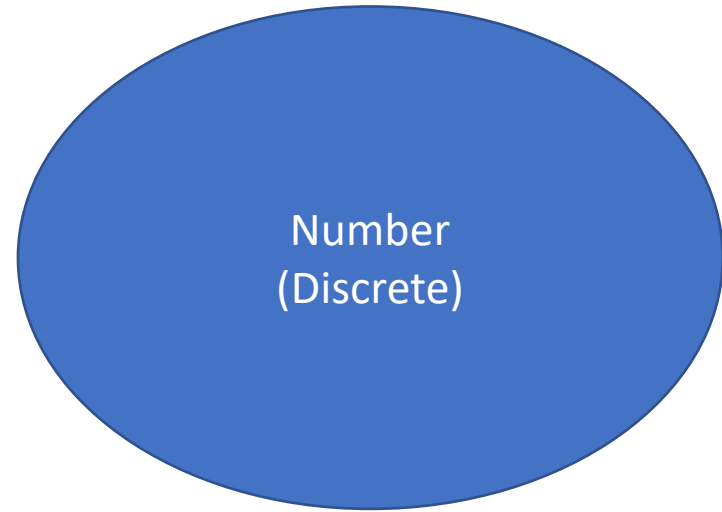
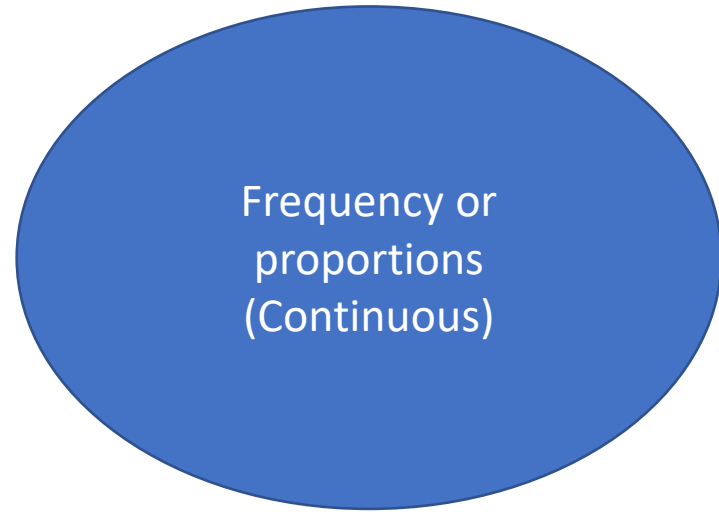
“Since s is large, the variation of the total concentration \bar{p} will proceed much more slowly than those of the partial concentrations p . Therefore, \bar{p} can temporarily be taken constant. The concentrations p in partial populations deviate from \bar{p} in either direction. After sufficiently long time the fluctuations of p around \bar{p} result in a certain stationary probability distribution for the concentrations p .” (Kolmogorov 1935)

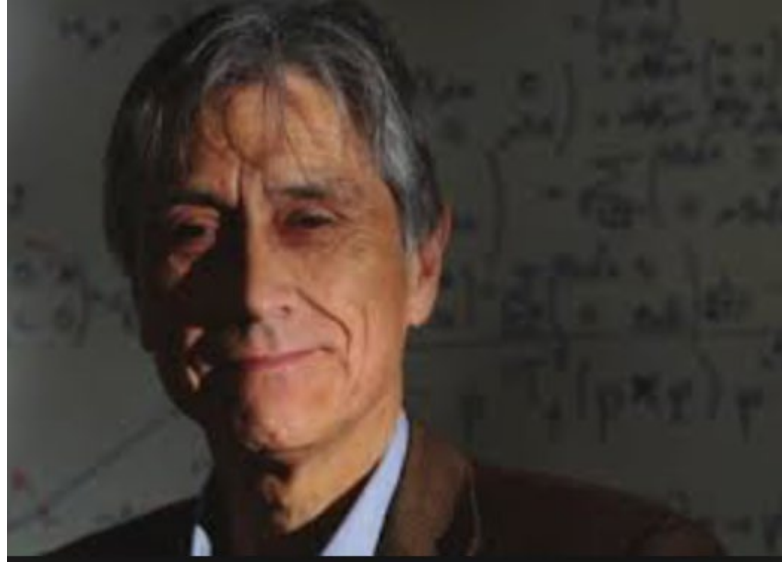
This stationary distribution satisfies the Kolmogorov forward or Fokker-Planck equation:

$$\frac{1}{2} \frac{\partial^2}{\partial p^2} (\text{Var}(p)\mu) - \frac{\partial}{\partial p} (M(p)\mu) = 0$$

Whose solution $\mu(p)$ is:

$$\mu(p) = \frac{1}{B(4k\bar{p}, 4k\bar{q})} p^{4k\bar{p}-1} q^{4k\bar{q}-1}$$





Prof. Rolando Rebolledo

**La méthode des martingales appliquée à l'étude de
la convergence en loi de processus**

Mémoires de la S. M. F., tome 62 (1979), p. I-V+1-125.

http://www.numdam.org/item?id=MSMF_1979__62__R1_0

$$B_j(n) = b_j(n) + c_j(n)$$

$$D_j(n) = d_j(n) + c_j(n).$$

$$b(x) = b_0 + b_1x$$

$$d(x) = d_0 + d_1x$$

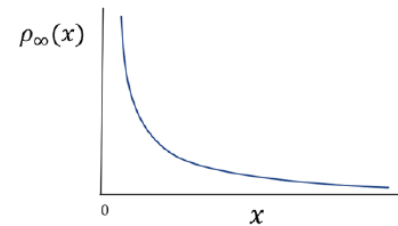
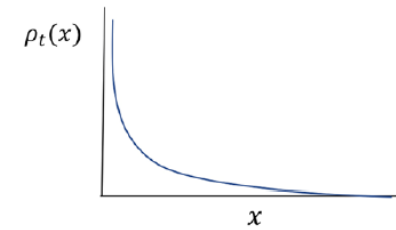
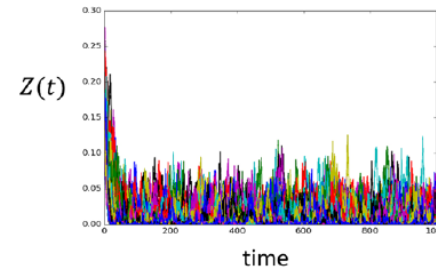
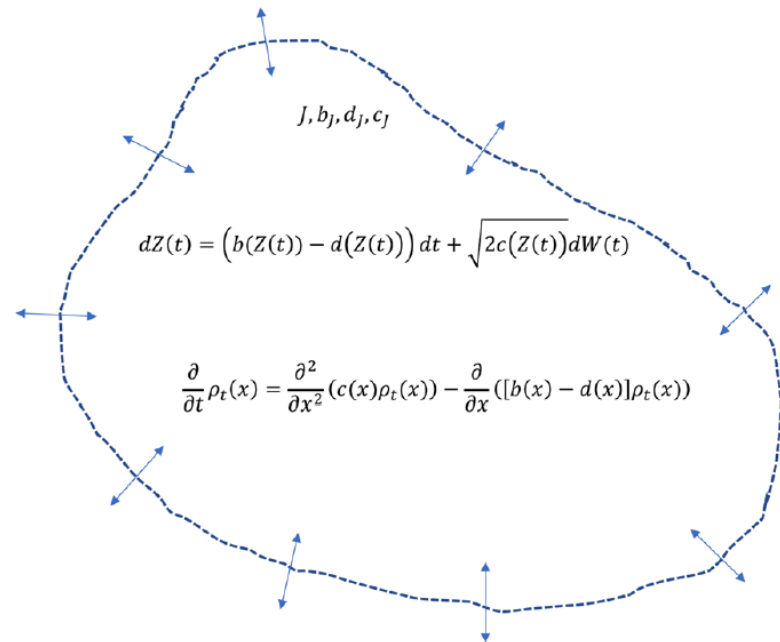
$$c(x) = \gamma x(1-x),$$

SCIENTIFIC REPORTS

OPEN On the proportional abundance of species: Integrating population genetics and community ecology

sd: 24 March 2017
sd: 21 November 2017

Pablo A. Marquet^{1,2,3,4,5}, Guillermo Espinoza¹, Sebastian R. Abades⁵, Angela Ganz⁷ & Rolando Rebolledo^{7,8}



$$\rho_\infty(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) + \Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}$$

$$\rho_{\infty}(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) + \Gamma(\beta)} x^{\alpha-1} (1 - x)^{\beta-1}$$

The Proportional Species Abundance Distribution (PSAD)

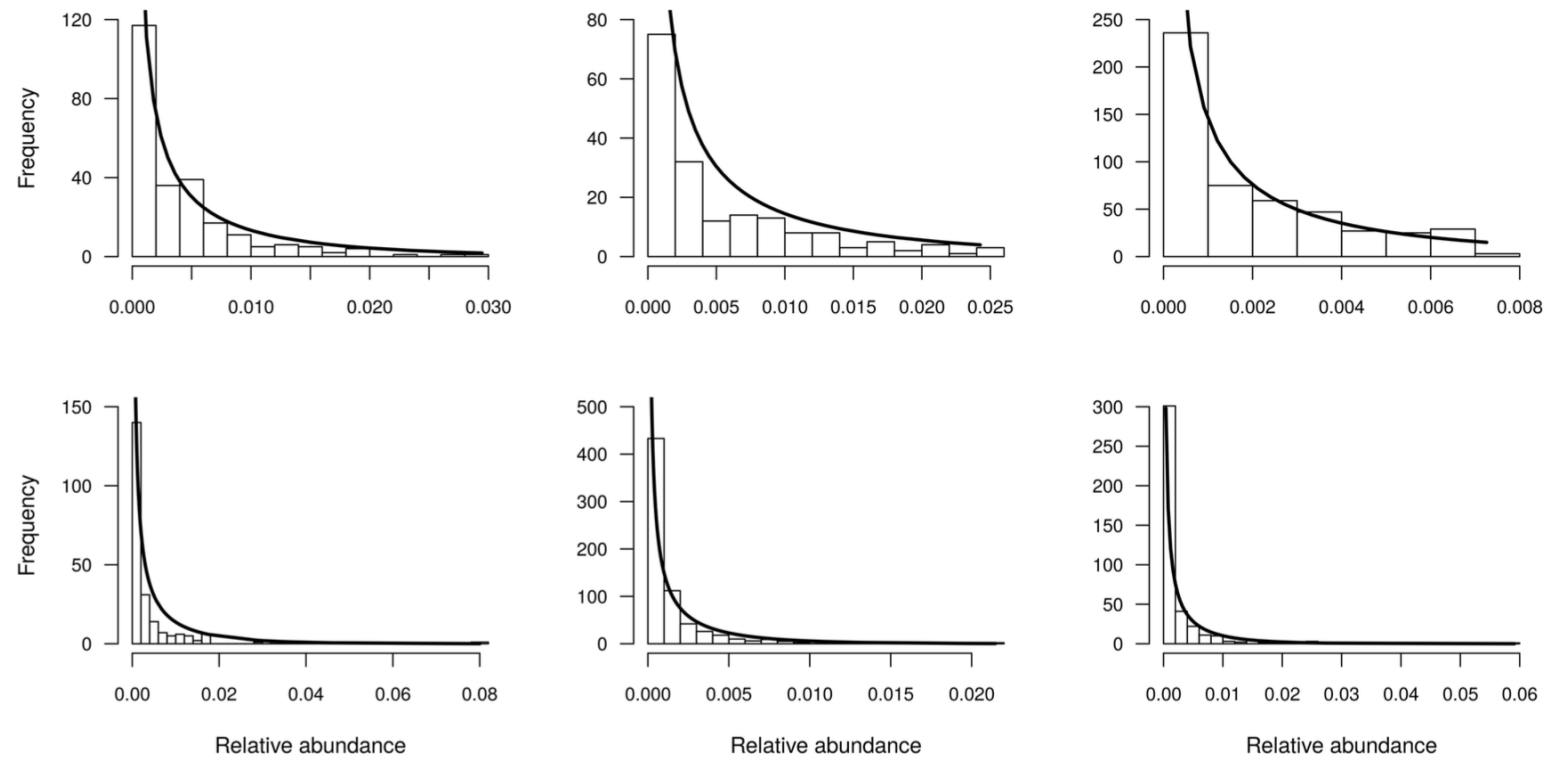
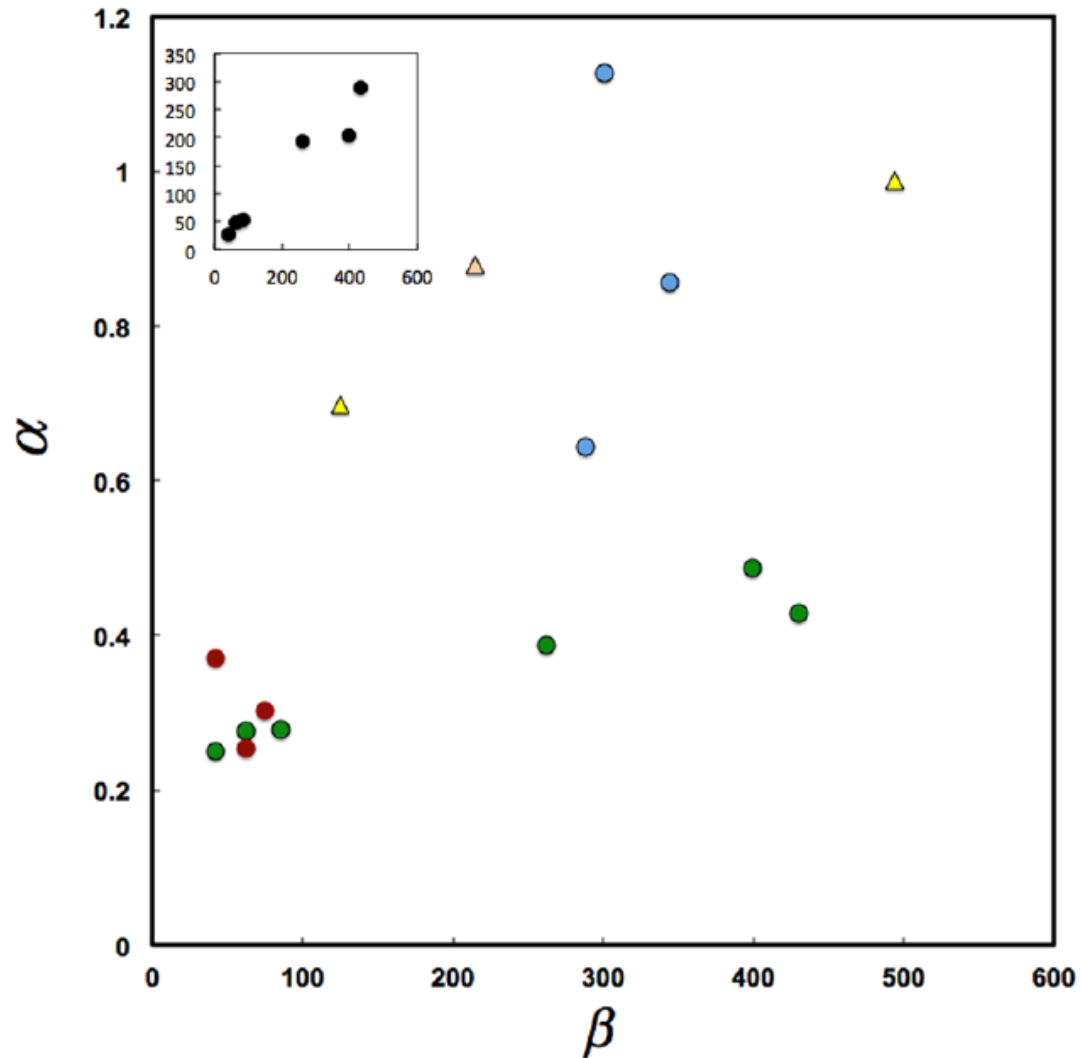


Figure 1. Fit of the Beta distribution to different animal and plant communities. First row, from left to right Amazon birds (community 10 in Table 1), Lepidoptera (12 in Table 1), butterflies (11 in Table 1), second row from left to right Tropical trees (6 in Table 1), Tropical trees (2 in Table 1) and Coral reefs (14 in Table 1)

$$\rho_{\infty}(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) + \Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}$$



- Marine communities
- Shrublands
- Tropical Forest
- ▲ Bird communities

$\alpha \sim$ Dispersal

$\beta \sim$ Speciation



"Forces maintaining species diversity and genetic diversity are similar. An understanding of community structure will come from considering how these kind of diversity interact." Janis Antonovics

THE THEORY OF
Island Biogeography
ROBERT H. MACARTHUR
EDWARD O. WILSON

$$(3-4) \quad \frac{dP_s(t)}{dt} = -(\lambda_s + \mu_s)P_s(t) + \lambda_{s-1}P_{s-1}(t) + \mu_{s+1}P_{s+1}(t).$$

$$Z(t) = Z(0) + \int_0^t \theta(Z(s))ds + \int_0^t \sigma(Z(s))dW_s, \quad (t \geq 0).$$

MONOGRAPHS IN POPULATION BIOLOGY - 1

$$\rho_\infty(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) + \Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}$$

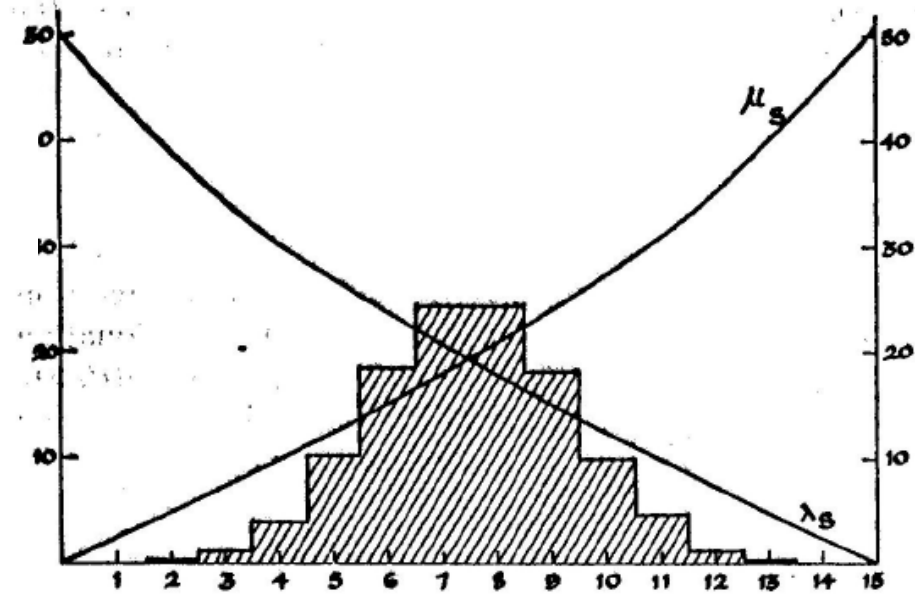
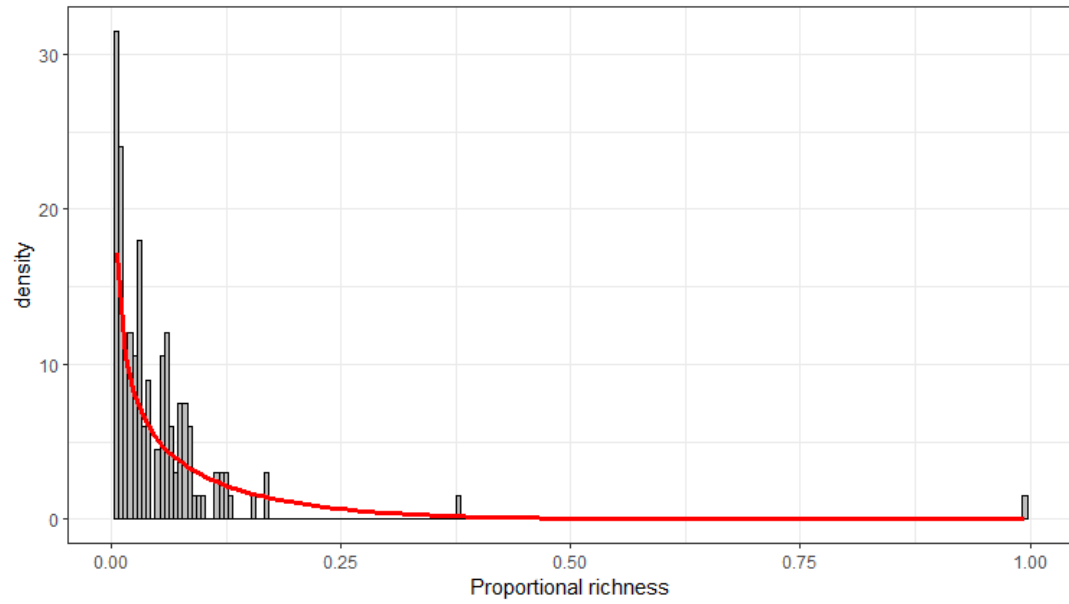


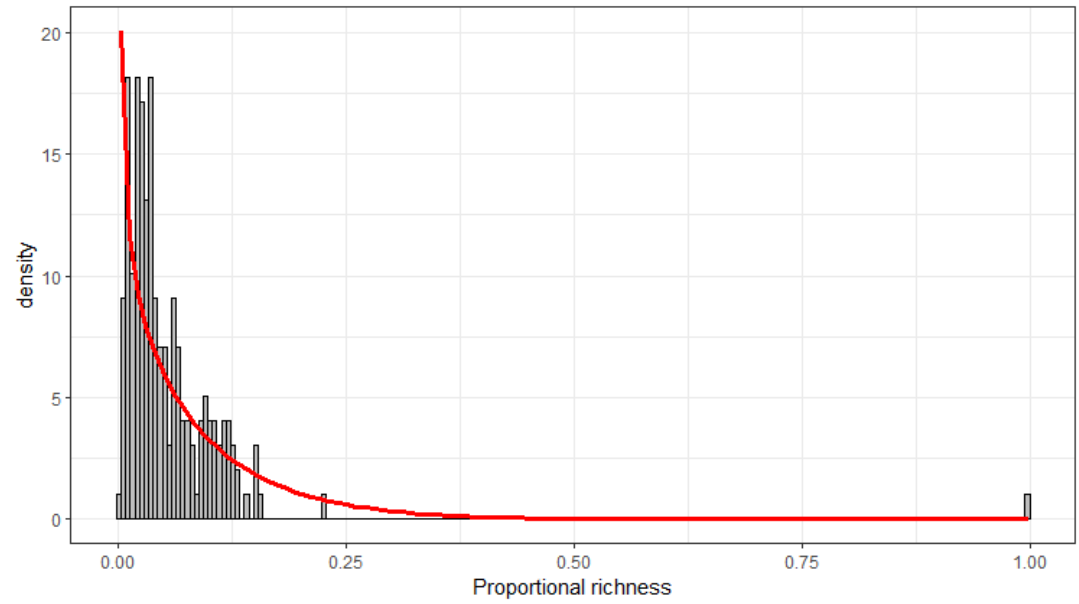
FIGURE 19. A particular case of a predicted distribution of numbers of species on a family of island biotas all with identical extinction and immigration curves and all having had time to reach equilibrium. The histogram represents the number of islands with each number of resident species in an equilibrium situation. The species pool from which the biotas were assembled contained 15 species. If the immigration and extinction curves were straighter, the variance of equilibril species numbers would be even greater; yet this large variance is still consistent with the equilibril condition.

"In principle one could solve eq. 3-4...for our purpose is more useful to find the mean $M(t)$ and the variance, $var(t)$, of the number of species at time t . These can be estimated in nature by measuring the mean and variance in number of species of a series of islands of about the same distance and are hence of the same λ and μ ." pp.33-34

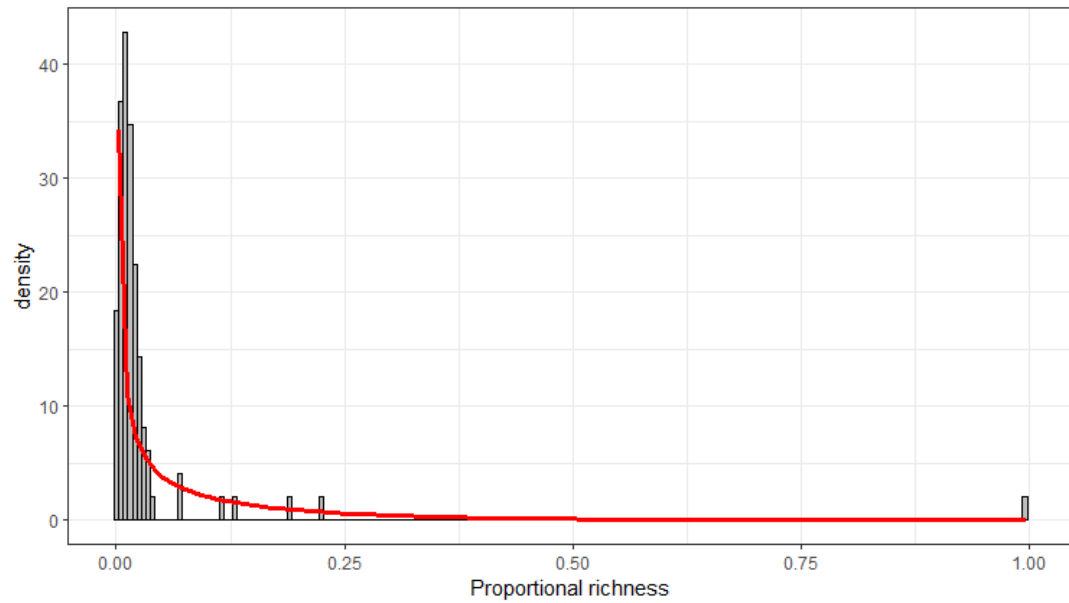
Mammals



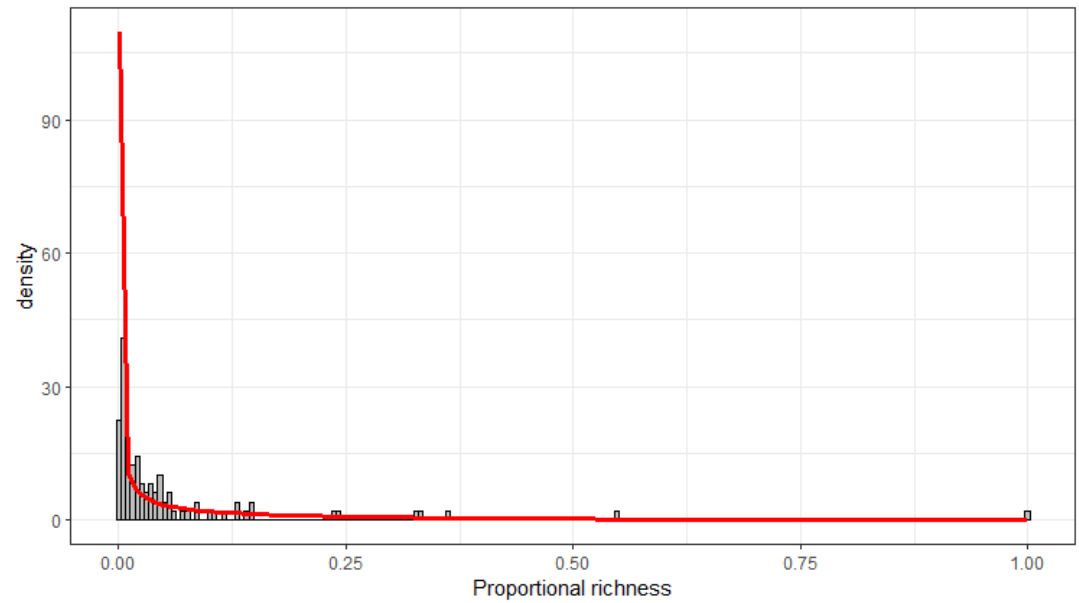
Birds



Reptiles



Plantas



Some Sources of Stochasticity at Different Levels of Organization

Frequency of Gene families
within Genomes

- Mutation
- Lateral gene transfer
- Gene duplication

Frequency of genes within
and among population

- Mutation
- Drift
- Migration
- Loss

Frequency of individuals
among species

- Birth
- Death
- Migration

Frequency of species
Among communities

- Speciation
- Extinction
- Migration

Abundance of gene families within genomes

- Birth-death processes are applicable to understand genome evolution
- Driving processes are similar

Duplication

Mutation

Lateral gene transfer and retroviral infection

Plasmids

Gene extinction

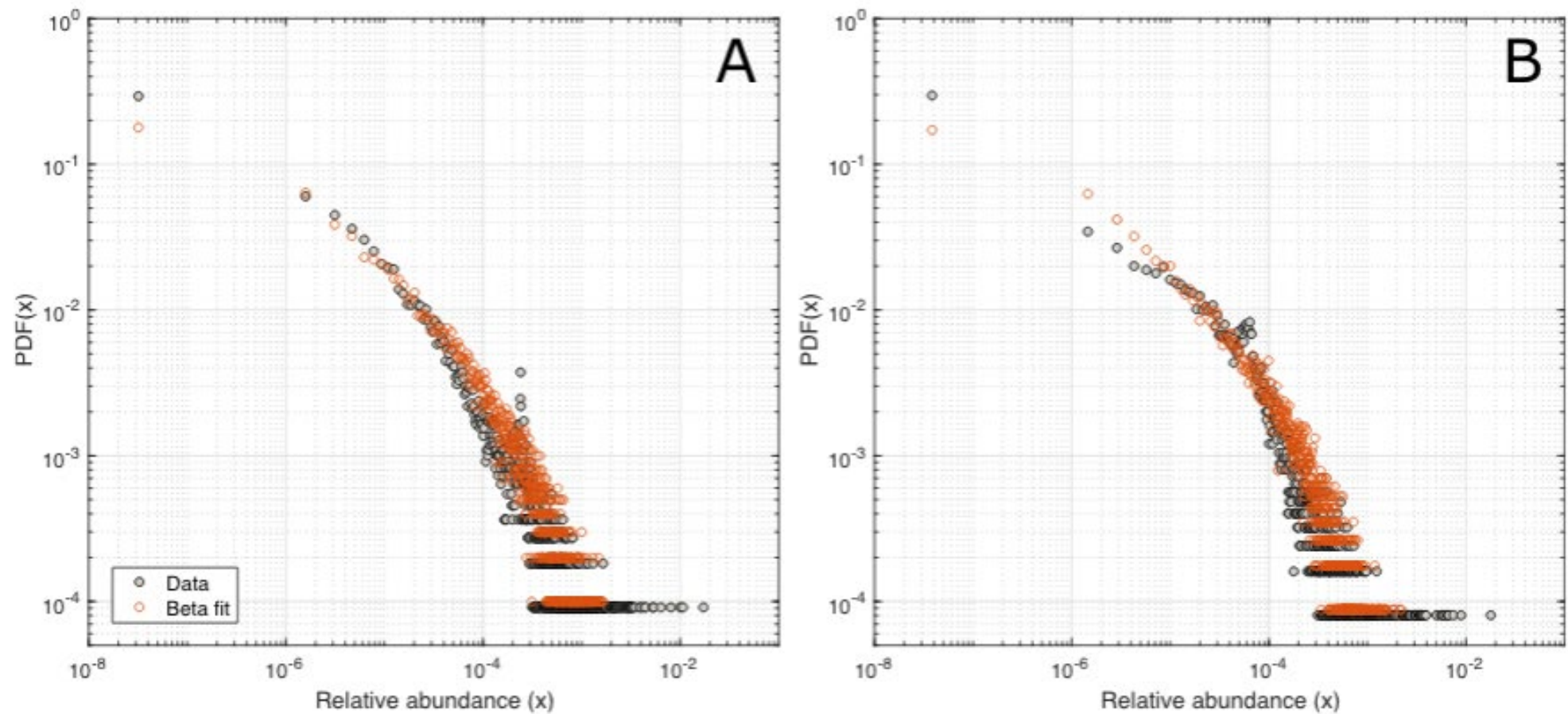


Figure 1. The relative abundance of protein families follows a Beta distribution. The relative abundance of protein families in bacteria and eukarya according to the Pfam classification. **A.** Data from 7,694 bacterial proteomes; 11,106 families; and 30,743,438 genes. Best fit beta distribution parameters: $\alpha=0.273$ (0.264, 0.283), $\beta=3,031.9$ (2823.4, 3240.4). **B.** Data from 1,496 eukaryotic proteomes; 12,579 families; and 25,625,956 genes. Best fit beta distribution parameters: $\alpha=0.291$ (0.281, 0.300), $\beta=3,655.2$ (3424.3, 3886.2).

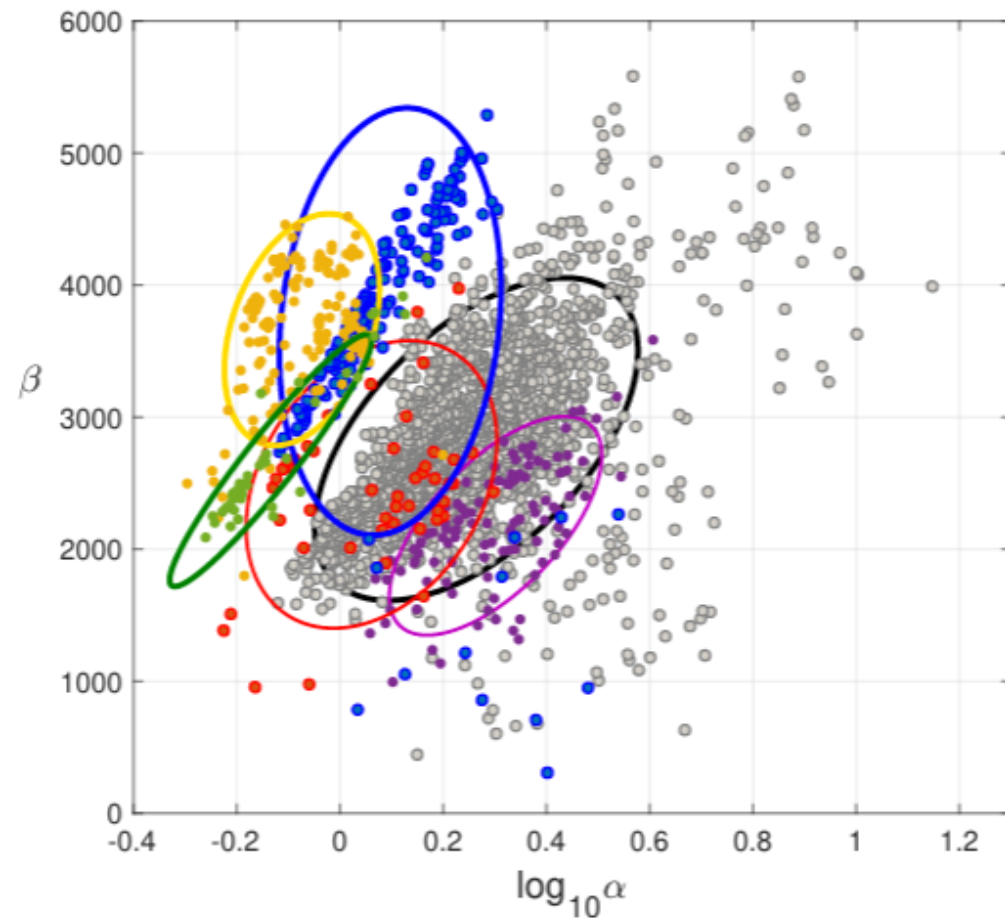


Figure 2. Shape parameters of Beta distributions arising from the relative abundance of gene families in genomes. Annotations of protein families (Pfam v32) in Uniprot reference proteomes (n=17,543). Reference proteomes include Archea (n=285, magenta), Bacteria (n=6,554, grey); unicellular eukarya (n=1,202, red); plants (n=, green); metazoa (n=, yellow); fungi (n=, blue). The total number of protein families classified by Pfam is 15,964.

The replicator equation

Following Hofbauer and Sigmund (1998), consider a population divided into n types E_1 to E_n with frequencies x_1 to x_n . The fitness f_i of E_i will be a function of the composition of the population, or state \mathbf{x} . Such that the rate of increase \dot{x}_i/x_i of type E_i is a measure of its evolutionary success, which can be expressed as the difference between the $f_i(\mathbf{x})$ of E_i and the average fitness $\bar{f}(\mathbf{x}) = \sum x_i f(\mathbf{x})$ of the population.

$$\frac{\dot{x}_i}{x_i} = \text{fitness of } E_i - \text{average fitness},$$

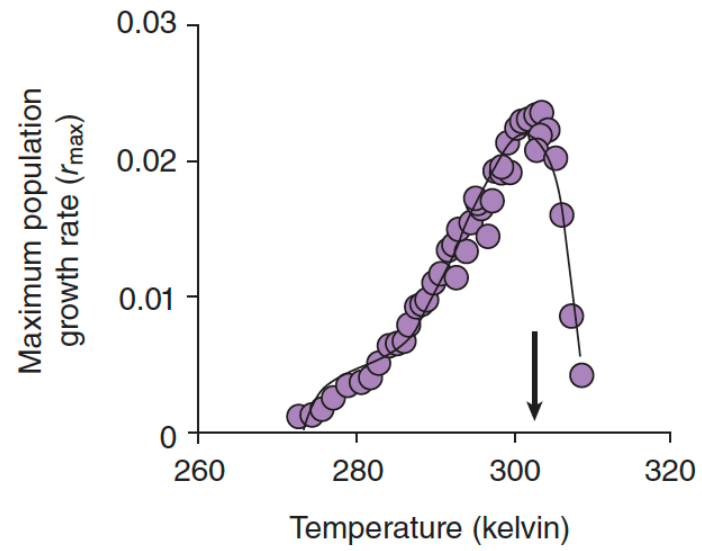
which yields the *replicator equation*

$$\dot{x}_i = x_i(f_i(\mathbf{x}) - \bar{f}(\mathbf{x})) \quad i = 1, \dots, n .$$

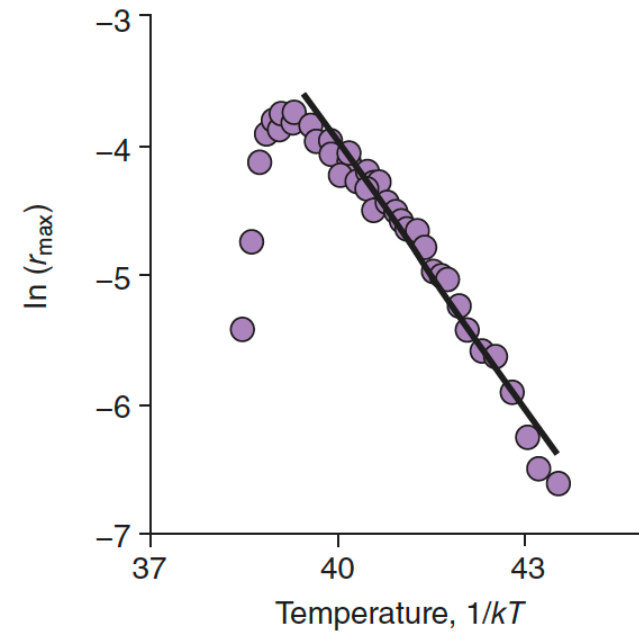
Kinetics

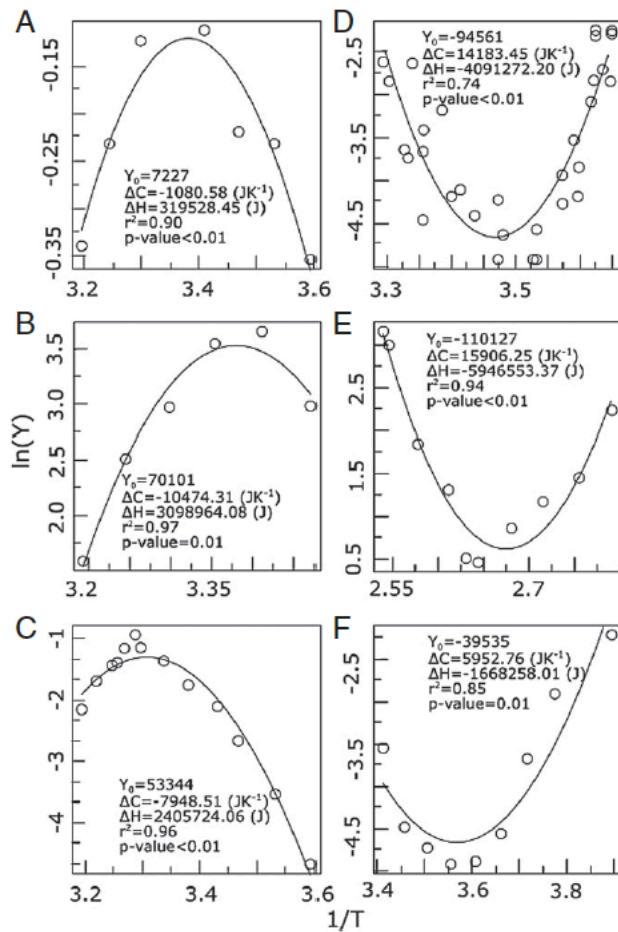
$$k \propto e^{\frac{-E}{KT}}$$

A



B

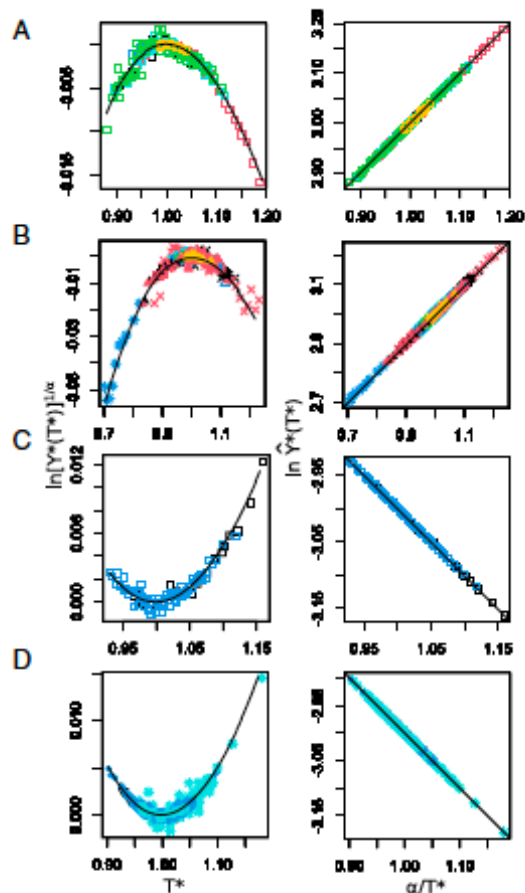




The Eyring–Evans–Polanyi (EEP) transition state theory (TST)

$$k = \frac{k_B}{h} T e^{\Delta S/R} e^{-\Delta H/RT}$$

A general theory for temperature dependence in biology



- Enzyme activity parameters
- Body size
- △ Carbon stock
- + Developmental rate
- × Ecosystem flux
- ◇ Functional response (Handling time, feeding rate)
- ▽ Generation time
- ⊠ Maximum germination
- * Metabolic rate
- ⊕ Mutation rate
- ◆ Performance
- ⊠ Population density
- ⊠ Population flux
- ⊠ Population growth rate
- ⊠ Energy use
- Species richness
- Latency period
- ▲ Community abundance
- ◆ Mortality rate
- Fecundity
- Mass-specific metabolic rate
- Archaea
- Bacteria
- Unicellular eukaryotes
- Ectotherm
- Endotherm
- Virus
- Multicellular eukaryotes

$$Y(T) \approx Y_0 \left(\frac{1}{T} \right)^{\frac{-\overline{\Delta C}}{R} - \alpha} e^{\frac{-\overline{\Delta H}}{RT}}$$

$$Y^{*1/a} = T^* e^{1/T^* - 1}$$

$$\hat{Y}^*(T^*) \equiv (e/T^*)^a Y^*(T^*) = e^{a/T^*}$$

- The development of theories based on first principles and using the language of mathematics can help to improve scientific enquiry in ecology by identifying inconsistencies in auxiliary hypothesis (Markovianity) and by fostering integration across phenomena occurring at different levels of organization.
- Genes, individuals and species seem to obey the same probability law.

Special thanks to:

Rolando Rebolledo

Mauricio Tejo

Evandro Ferrada

Octavio Reyes

Simón Castillo

Ignacio Arroyo

Lucas Gallart

- Thanks
- Obrigado
- Gracias
- *Ckantatur*
- Chaltu may
- Yuspagara

