Unexpected connections

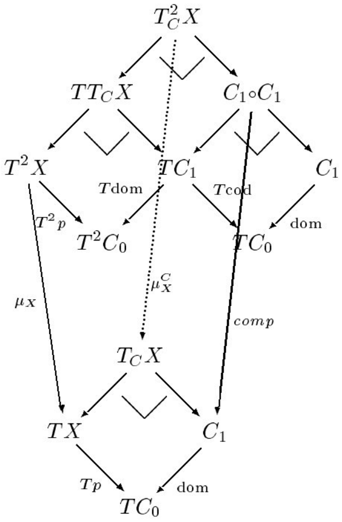
Tom Leinster

School of Mathematics
University of Edinburgh

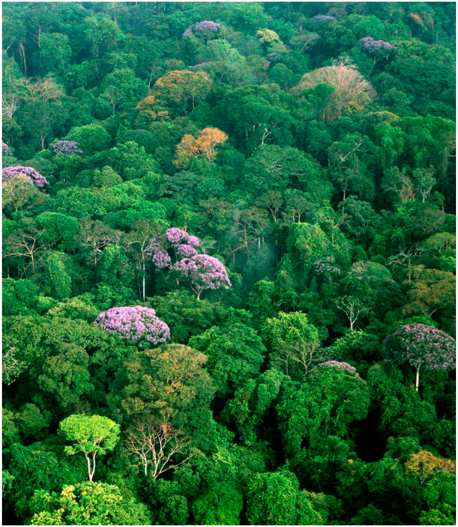
Boyd Orr Centre
for Population and Ecosystem Health
University of Glasgow

We will go on a journey. . .

. . . from here. . .



. . . to here.



But first

Inclusivity

More obvious

Gender

Sexuality

Ethnicity

Disability

Age

Less obvious

Prestige of university

Access to funds

Country

Native language

Even less obvious

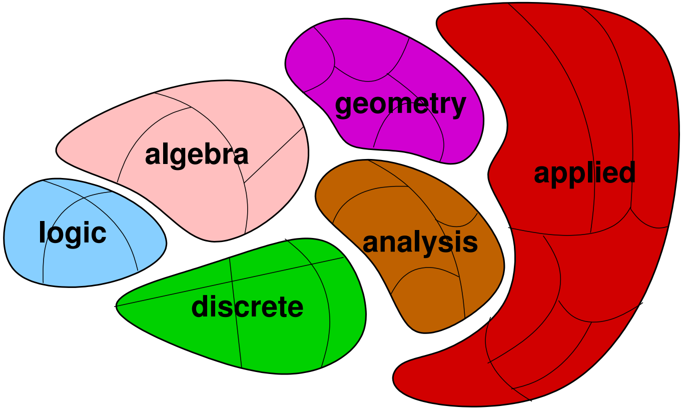
Character traits

Ways of thinking

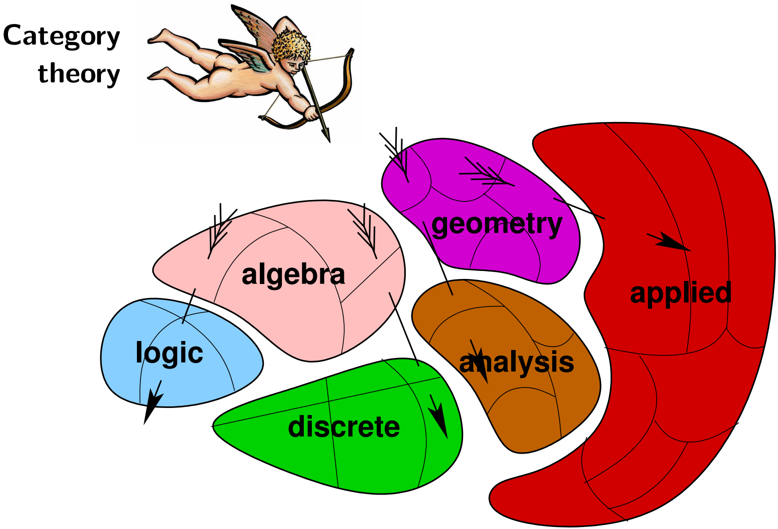
Preference for working alone or in groups

Some mathematics
that looks *pure*

A not-too-serious map of mathematics



A not-too-serious map of mathematics



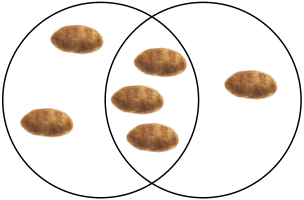
- Category theory is also *part* of mathematics.
- The map is constantly being redrawn.

Counting potatoes

We're going to look at various measures of size.

The simplest notion of 'size' is the number of things.

Basic rule: the inclusion-exclusion principle:

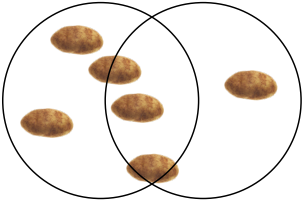


On left: 5 In middle: 3 On right: 4

Total: $5 + 4 - 3 = 6$ ✓

Counting potatoes

Less obvious fact: the inclusion-exclusion principle still holds if we allow the potatoes to cross the lines!

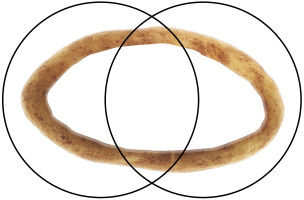


On left: 5 In middle: 3 On right: 4

Total: $5 + 4 - 3 = 6$ ✓

Counting potatoes

A potato with an unusual shape:



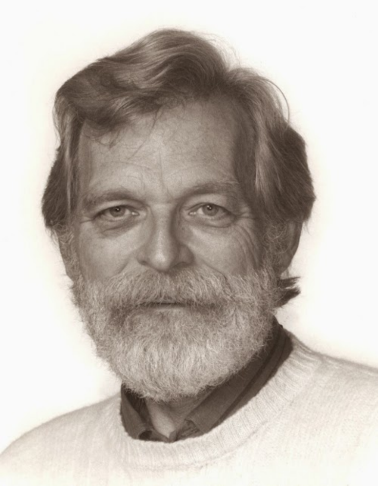
On left: 1 In middle: 2 On right: 1

Total: $1 + 1 - 2 = 0$?!?

Can that really be right?

Yes! It's the story of Euler characteristic. . . which I won't tell.

Schanuel on Euler characteristic



Stephen Schanuel, 1991:

Euler's analysis, which demonstrated that in counting suitably 'finite' spaces one can get well-defined negative integers, was a revolutionary advance in the idea of cardinal number—perhaps even more important than Cantor's extension to infinite sets, if we judge by the number of areas in mathematics where the impact is pervasive.

Other ways to measure a potato

- Its volume is 84cm^3.
- Its surface area is 91cm^2.
- Its mean width is 7cm.

Or simply:

- It's 1 potato.

All four measures obey the inclusion-exclusion principle.

A general notion of size

What happens if we keep developing these counting ideas?

Using some category theory, it's possible to give a very general definition of 'size', called magnitude, which:

- makes sense in lots of different branches of mathematics (geometry, algebra, ...)
- connects various old notions of size (volume, cardinality, dimension, ...)
- also produces some interesting new quantities.

The magnitude of a collection of points

The magnitude of a collection of points can be thought of as the ‘effective number of points’.

Magnitude: 1

The magnitude of a collection of points

The magnitude of a collection of points can be thought of as the ‘effective number of points’.

Magnitude: 1.01

The magnitude of a collection of points

The magnitude of a collection of points can be thought of as the 'effective number of points'.

Magnitude: 1.2

The magnitude of a collection of points

The magnitude of a collection of points can be thought of as the 'effective number of points'.

Magnitude: 1.6

The magnitude of a collection of points

The magnitude of a collection of points can be thought of as the ‘effective number of points’.

Magnitude: 2.3

The magnitude of a collection of points

The magnitude of a collection of points can be thought of as the 'effective number of points'.

Magnitude: 2.3

As the points get further apart, the magnitude gets closer to 3.

If you want to try it at home...

Here's the definition of the magnitude of a set of points $x_1, \ldots, x_n$ in $\mathbb{R}^k$.

Write d_{ij} for the distance between x_i and x_j.

Let Z be the $n \times n$ matrix whose (i, j)-entry is $e^{-d_{ij}}$.

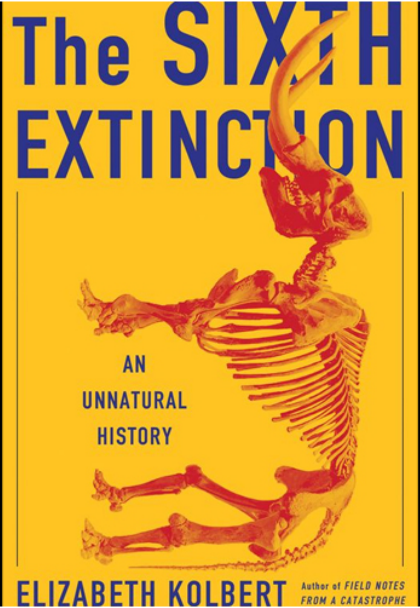
Fact: Z is invertible.

The magnitude of $\{x_1, \ldots, x_n\}$ is the sum of all n^2 entries of Z^{-1}.

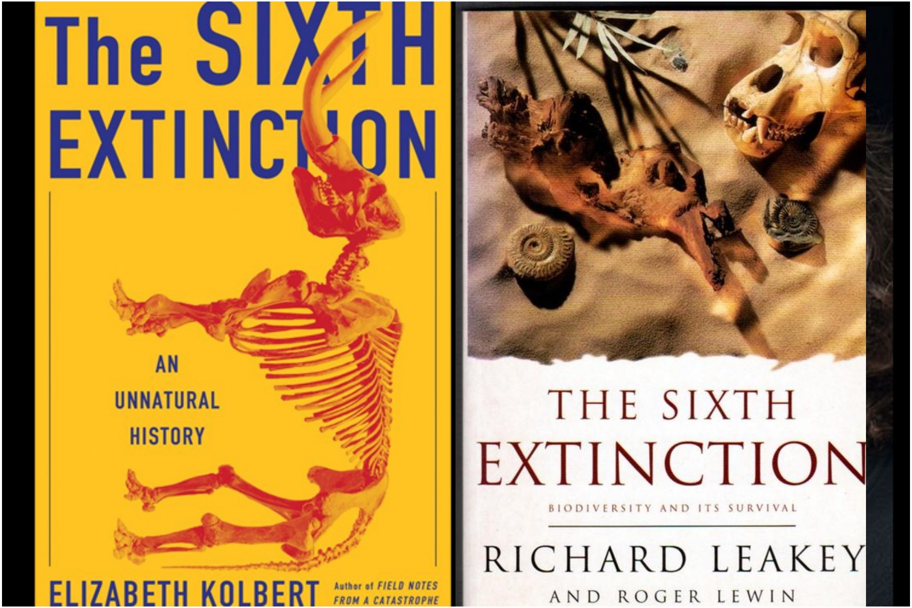
It's not obvious that this is an interesting definition... but it is!

Biological diversity

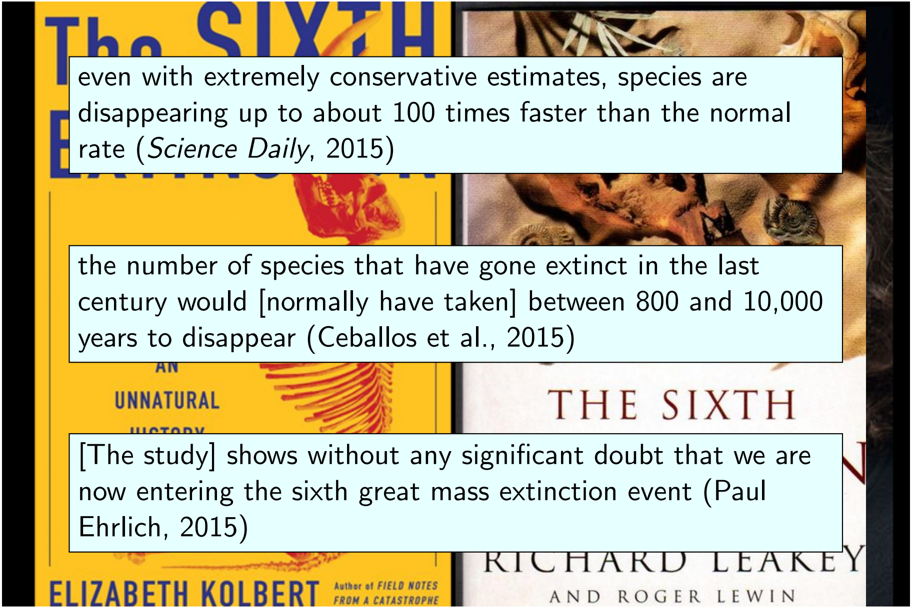
The global biodiversity crisis



The global biodiversity crisis



The global biodiversity crisis



even with extremely conservative estimates, species are disappearing up to about 100 times faster than the normal rate (*Science Daily*, 2015)

the number of species that have gone extinct in the last century would [normally have taken] between 800 and 10,000 years to disappear (Ceballos et al., 2015)

[The study] shows without any significant doubt that we are now entering the sixth great mass extinction event (Paul Ehrlich, 2015)

The global biodiversity crisis

Photo by Ray Wiltshire/Rex Features

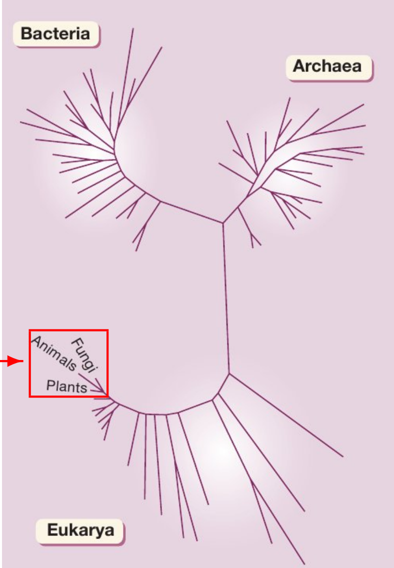
The global biodiversity crisis

Photo by Arthur Anker/Animal Earth/Thames & Hudson

The global biodiversity crisis

The tree of life

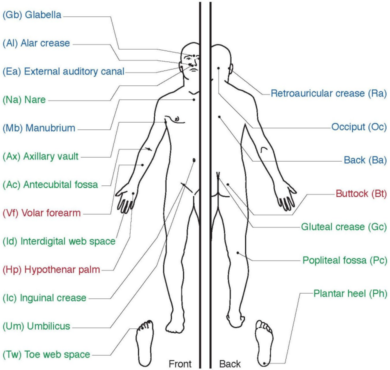
Most of visible life



Almost all living organisms are invisible to us!

You are not alone

Within the body of a healthy adult, microbial cells are estimated to outnumber human cells by a factor of ten to one. —BEI Resources



⟵These are just the ones on your skin

The diversity of this ecosystem matters! E.g. experiments suggest:

more diverse intestinal bacteria $\leftrightsquigarrow$ less likely to be overweight.

Viruses and vaccines

Consider the flu. . .

Q. Why can't we have a single flu vaccination that lasts a lifetime?

A. Because the flu virus evolves *fast* — different types appear every year.

The more diverse this year's collection of flu types is, the more different vaccines you need.

Here, diversity is bad news (for us!)

What's the best measure of diversity?

This question has been debated for > 50 years, leading to *many* proposed diversity measures...

Whittaker's index of association
Percentage difference (*alias* Bray–Curtis)
Wishart coefficient = (1−similarity ratio)
D = (1−Kulczynski coefficient)
Abundance-based Jaccard
Abundance-based Sørensen
Abundance-based Ochiai

Species richness $x \equiv \sum_{i=1}^{S} p_i^0$

Shannon entropy $x \equiv -\sum_{i=1}^{S} p_i \ln p_i$

Simpson concentration $x \equiv \sum_{i=1}^{S} p_i^2$

Gini–Simpson index $x \equiv 1 - \sum_{i=1}^{S} p_i^2$

HCDT entropy $x \equiv \left(1 - \sum_{i=1}^{s} p_i^q\right)/(q-1)$

Renyi entropy $x \equiv \left(-\ln \sum_{i=1}^{s} p_i^q\right)/(q-1)$

What's the best measure of diversity?

This question has been debated for $>$ 50 years, leading to *many* proposed diversity measures. . .

Whittaker's index of association

$^qD_{\gamma j}$	$1/\bar{p}_{(i\|j)j}$	$\gamma_j = {}^q\lambda_{\gamma j}^{1/(1-q)}$	gamma unit j (
$\bar{p}_{(i\|j)\mathrm{all}}$	$\sqrt[q-1]{\sum_{j=1}^{N}\sum_{i=1}^{S} p_{ij}p_{i\|j}^{q-1}}$		mean p general
α_t	$^q\bar{D}_{\gamma j} = \bar{\gamma}_j$	$1/\bar{p}_{(i\|j)\mathrm{all}}$	mean s mean v weights
α_d	$^qD_\alpha$	α_t/CU	true alp sampli (measu
α_R	$^qD_{\cdot\gamma\omega'/\omega}$	$^qD_{\cdot\gamma\omega'}/{}^qD_\omega$	effectiv samplin
β_{Md}	$^qD_\beta = {}^qD_\gamma/{}^qD_\alpha$	γ/α_d	true be (measu
β_{Mt}	$^qD_{\gamma/\bar\gamma_j} = {}^qD_\gamma/{}^q\bar{D}_{\gamma j}$	γ/α_t	regiona
β_R	$^qD_\gamma\ {}^qD_\omega/{}^qD_{\cdot\gamma\omega'}$	γ/α_R	two-wa
β_{At}	$^qD_\gamma - {}^q\bar{D}_{\gamma j}$	$\gamma - \alpha_t$	regiona (measu
β_{Mt-1}	$\gamma/\alpha_t - 1$	$(\gamma-\alpha_t)/\alpha_t$	Whitta multipl (measu
β_{Pt}	$1 - \alpha_t/\gamma$	$(\gamma-\alpha_t)/\gamma$	proport as a pr
H'_β	$H'_\gamma - H'_\alpha$	$\log(^1\beta_{Md}) = \log(\gamma) - \log(\alpha_d)$	beta Shannon entropy (measurement unit: depends on the base of the logarithm)
$\bar{H}'_{\gamma-\gamma j}$	$H'_\gamma - \bar{H}'_{\gamma j}$	$\log(^1\beta_{Mt}) = \log(\gamma) - \log(\alpha_t)$	regional Shannon entropy excess (measurement unit: depends on the base of the logarithm)
$^2\bar\lambda_{\gamma j-\gamma}$	$^2\bar\lambda_{\gamma j} - {}^2\lambda_\gamma$	$(\gamma-\alpha_t)/\gamma\alpha_t$	regional variance excess (measurement unit: sp_E/sp_E^2)

^1E	Sheldon 1969, Buzas and Gibson 1969, Buzas and H McCarthy 2002, Camargo 2008
^2E	Weiher and Keddy 1999, Wilsey and Potvin 2000, Mo 2003, Ma 2005, Martin et al. 2005, Bock et al. 200 Camargo 2008
$D_{q/0}$	Alatalo 1981, Taillie 1979, Patil and Taillie 1982, Ricc
$D_{2/1}$	Rotenberry 1978, Alatalo 1981, Ricotta and Avena 20
$J'_{1/\max}$ or $J'_{1/0}$ or $^1J'$	Sheldon 1969, Tramer 1969, Kricher 1972, DeBenedic Wills et al. 1997, Rex et al. 2000, Wilsey and Potvi Miranda et al. 2002, Woodd-Walker et al. 2002, Ols et al. 2005, Kimbro and Grosholz 2006, Wilsey and Anticamara et al. 2010, Castro et al. 2010, Kardol e
H'_{1-0}	Hill 1973, Ricotta 2003, Kindt et al. 2006
$T_{2/\max}$ or PIE'	Hurlbert 1971, Fager 1972, DeJong 1975, Smith and
* H'	Jabot and Chave 2009, Anticamara et al. 2010
* ^2T or PIE	Weiher and Keddy 1999, Stevens and Willig 2002, D Vellend 2005, Ulrich and Zalewski 2007, Jarvis et a
* $^2H'$	Walker and Cyr 2007
* ^2D or ^1D	Gardezi and Gonzales 2008, Anticamara et al. 2010
# O	Mouillot and Wilson 2002, Stevens and Willig 2002
# E' or Gini coefficient	Camargo 1992a, 1993, Drobner et al. 1998, Mouillot 2002, Mouillot and Wilson 2002, Stevens and Will Ghersa 2011
# E_{var}	Drobner et al. 1998, Weiher and Keddy 1999, Mouill Symonds and Johnson 2008, Bernhardt-Römerman
# NHC	Weiher and Keddy 1999
# E_Q	Drobner et al. 1998, Mouillot and Wilson 2002, Ma 2

$\sum_{i=1} p_i \Big) / (q-1)$

What's the best measure of diversity?

The simplest notion of diversity is the number of species.

But sometimes things aren't so simple.

Question: which of these two communities is more diverse?

more species, less balanced

fewer species, more balanced

The question is impossible to answer objectively.

There are simply different viewpoints.

The mathematics of biodiversity

The most exciting phrase to hear in science,
the one that heralds new discoveries,
is not 'Eureka' but 'That's funny. . . '
—Isaac Asimov

Making a connection, part 1

Barcelona, 2008: I gave a talk on magnitude (‘effective number of points’). . .

André Joyal: *Don’t you think magnitude is a bit like entropy?*

Me: *Huh?*

Entropy

Take numbers $p_1, p_2, \ldots, p_n$ between 0 and 1, adding up to 1.

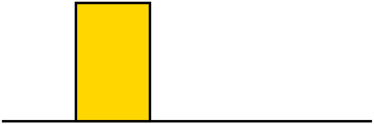
(That is, take a probability distribution on $\{1, \ldots, n\}$.)

Its entropy is the logarithm of

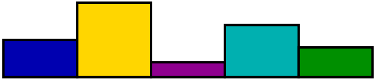
$$D(p_1, \ldots, p_n) = \frac{1}{p_1^{p_1} p_2^{p_2} \cdots p_n^{p_n}}.$$

Examples:

$D(p_1, \ldots, p_5) = 5$ (the maximum)



$D(p_1, \ldots, p_5) = 1$ (the minimum)



$D(p_1, \ldots, p_5) \approx 3.8$ (in between)

Making a connection, part 2

Back home in Glasgow. . .

Christina Cobbold: *That's funny. Ecologists sometimes use entropy to measure biodiversity.*

Me: *Huh?*

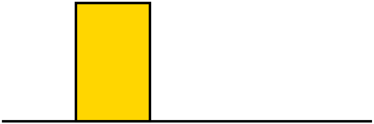
Entropy as a measure of biodiversity

Suppose we have a community consisting of n species, in proportions $p_1, p_2, \ldots, p_n$.

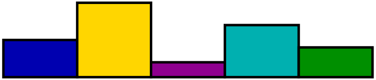
You can think of D, the exponential of entropy, as the 'effective number of species'.

Examples:

Effectively 5 species.

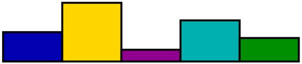


Effectively 1 species.



Effectively $\sim$3.8 species.

What's going on?



Ecologists sometimes consider 'effective numbers of species', taking into account how abundant they are.

I had been considering 'effective numbers of points', taking into account how far apart they are.

Mathematically:

- the first is about a probability distribution on a finite set
- the second is about a metric on a finite set.

Bringing it all together

A good measure of biodiversity should take two things into account:

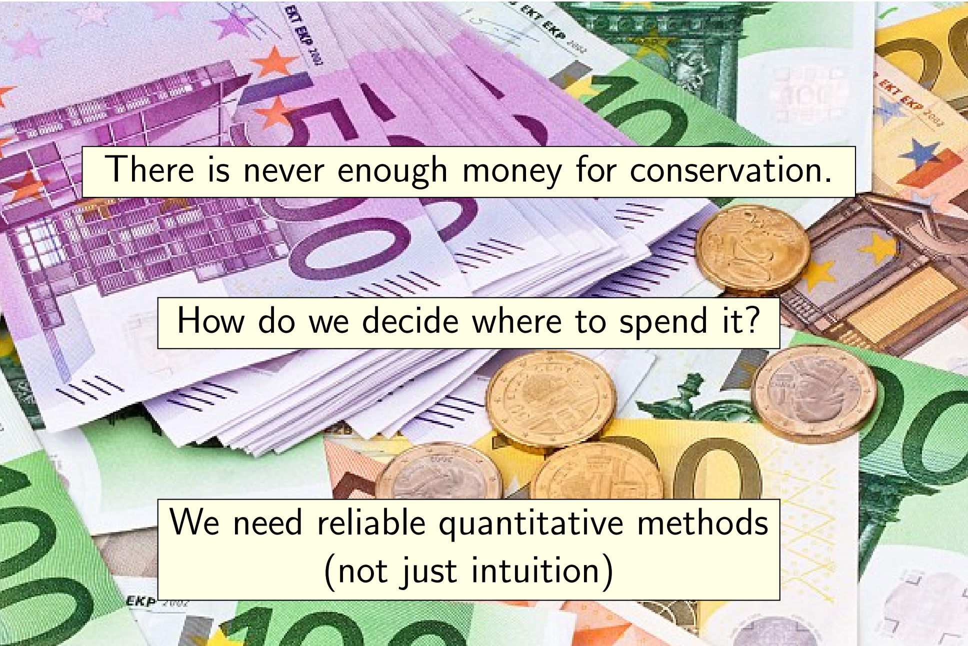
- how abundant the various species are; but also
- how different the species are.

Almost no existing measures of diversity did that!

Christina Cobbold and I combined two ideas — entropy and magnitude — to make a diversity measure that reflects both abundance and difference. It has been applied in contexts from bacteria to polar bears.

Real-life ecology

An enormous problem in conservation



There is never enough money for conservation.

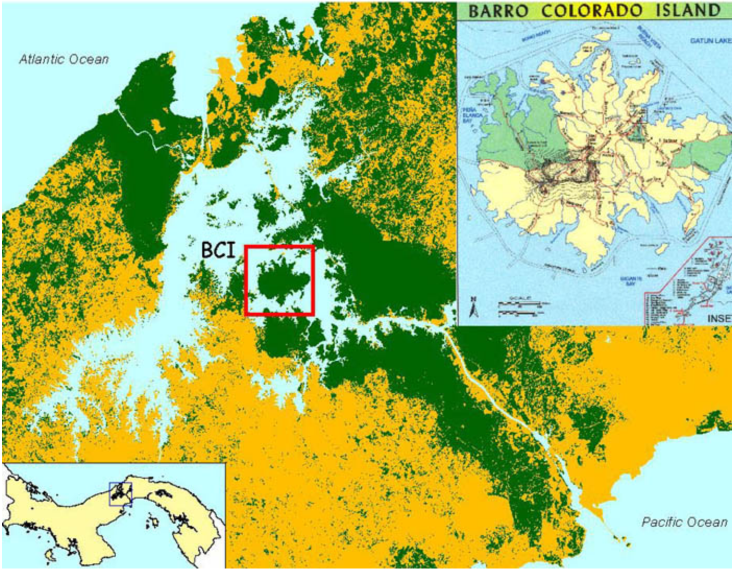
How do we decide where to spend it?

We need reliable quantitative methods
(not just intuition)

Fast-forwarding through lots of theory...

We made quantitative tools to locate areas that are highly diverse—and also to locate areas that are highly unusual.

Barro Colorado Island, Panama

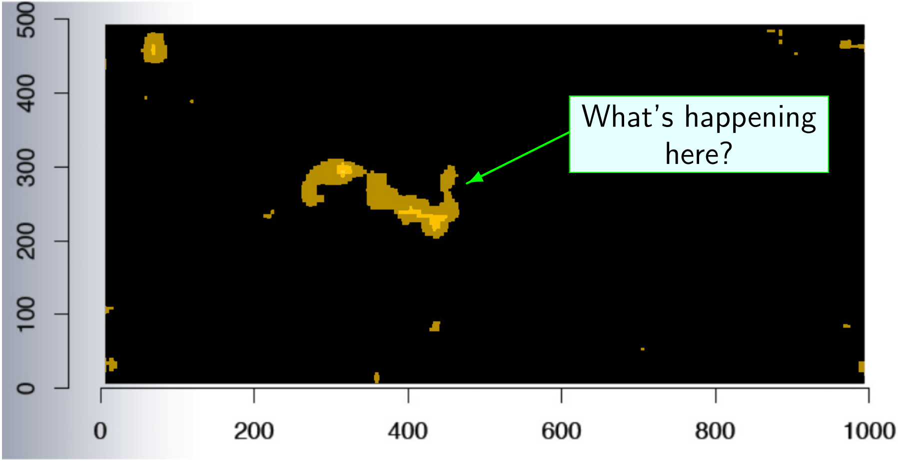


The island is mostly tropical forest, and contains a fully-censused study site.

Example result from using our tools on the study site

Yellow: areas most different from rest of forest.

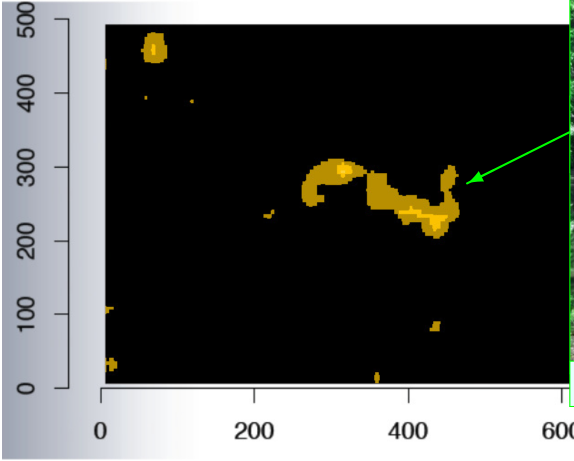
Black: areas most similar to rest of forest.



Example result from using our tools on the study site

Yellow: areas most different from rest of forest.

Black: areas most similar to rest of forest.



It's a swamp.

Thanks


Neil Brummitt


Christina Cobbold


André Joyal


Louise Matthews


Sonia Mitchell


Richard Reeve


Jill Thompson


Simon Willerton


The Barro Colorado Island project


You

(The BCI forest dynamics research project was made possible by National Science Foundation grants to Stephen P. Hubbell: DEB-0640386, DEB-0425651, DEB- 0346488, DEB-0129874, DEB-00753102, DEB-9909347, DEB-9615226, DEB- 9615226, DEB-9405933, DEB-9221033, DEB-9100058, DEB-8906869, DEB-8605042, DEB-8206992, DEB-7922197, support from the Center for Tropical Forest Science, the Smithsonian Tropical Research Institute, the John D. and Catherine T. MacArthur Foundation, the Mellon Foundation, the Small World Institute Fund, and numerous private individuals, and through the hard work of over 100 people from 10 countries over the past two decades. The plot project is part the Center for Tropical Forest Science, a global network of large-scale demographic tree plots.)