Limit Theorems for the Unified Neutral Theory of Biodiversity and Their Application to the Human Gut Microbiome

Todd Parsons

Laboratoire de Probabilités et Modèles Aléatoires, UPMC
Center for Interdisciplinary Research in Biology, Collège de France

Rencontre de la Chaire Modélisation Mathématique et Biodiversité
École Polytechnique
29 janvier 2014

JOINT WORK WITH

Keith Harris
University of Sheffield

Ian Holmes
University of California, Berkeley

Umer Zeeshan Ijaz
University of Glasgow

Christopher Quince
University of Glasgow

Outline

Outline

NEUTRAL *vs.* NICHE MODELS IN ECOLOGY

"When we look at the plants and bushes clothing an entangled bank, we are tempted to attribute their proportional numbers and kinds to what we call chance. But how false a view is this!"

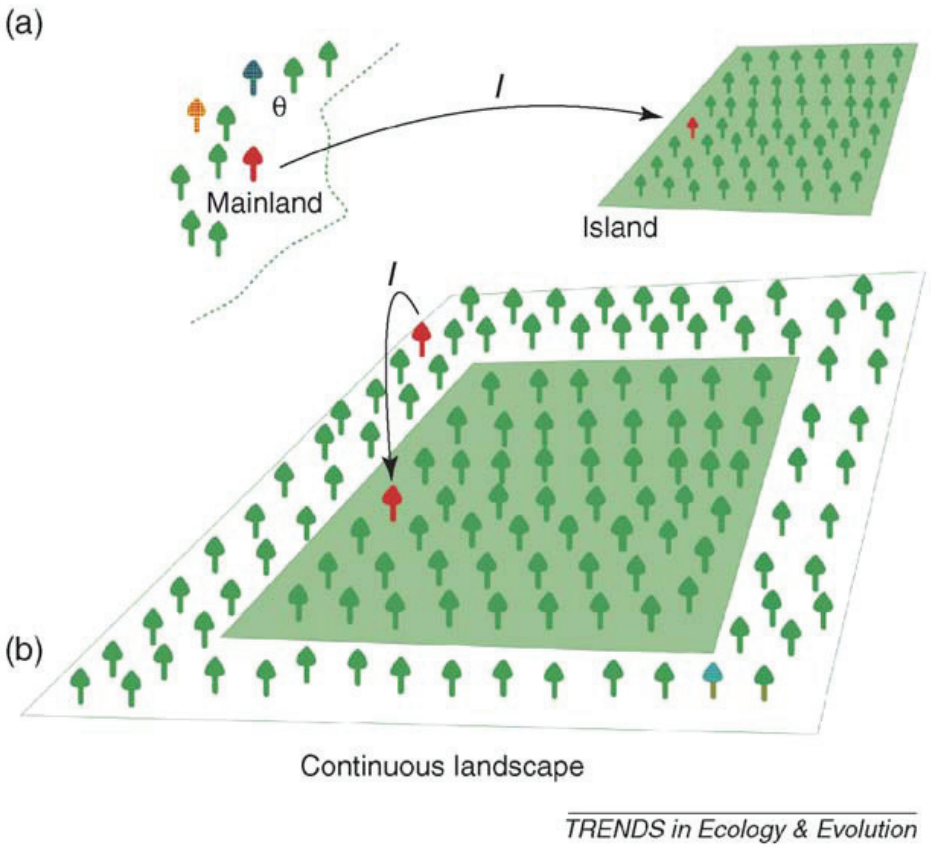
– Charles Darwin, *The Origin of Species*

NEUTRAL *vs.* NICHE MODELS IN ECOLOGY

After more than 25 years of research on the tropical forests of Barro Colorado Island, however, Stephen Hubbell controversially proposed exactly that. . . That random chance may in fact be the best explanation of the observed biodiversity.

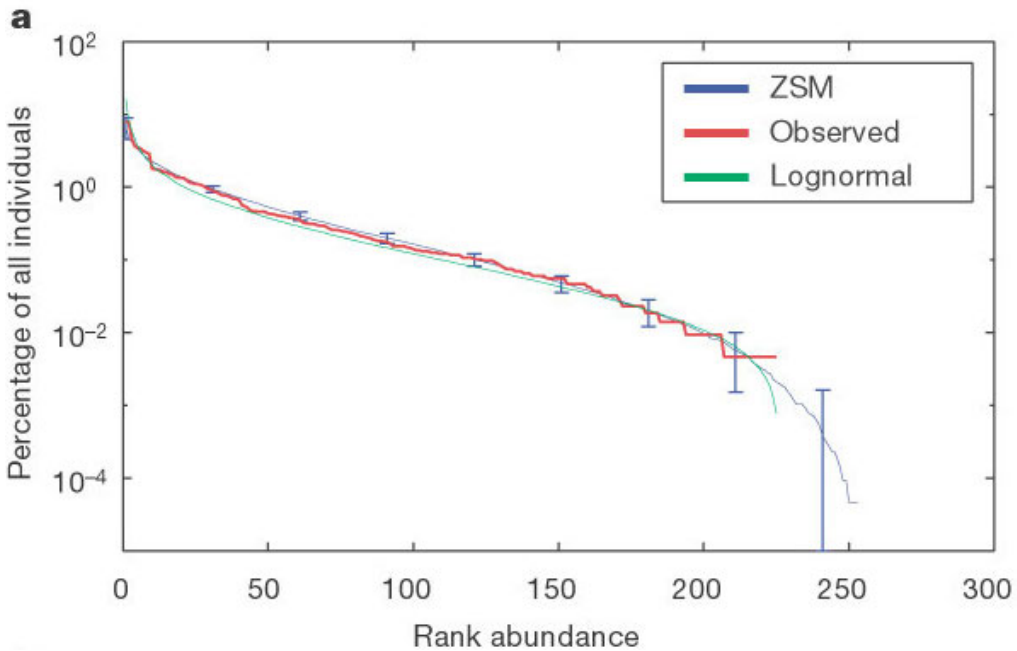
Gewin (2006)

HUBBELL'S UNIFIED NEUTRAL THEORY OF BIODIVERSITY AND BIOGEOGRAPHY (UNTB)



Alonso, Etienne & McKane (2006).

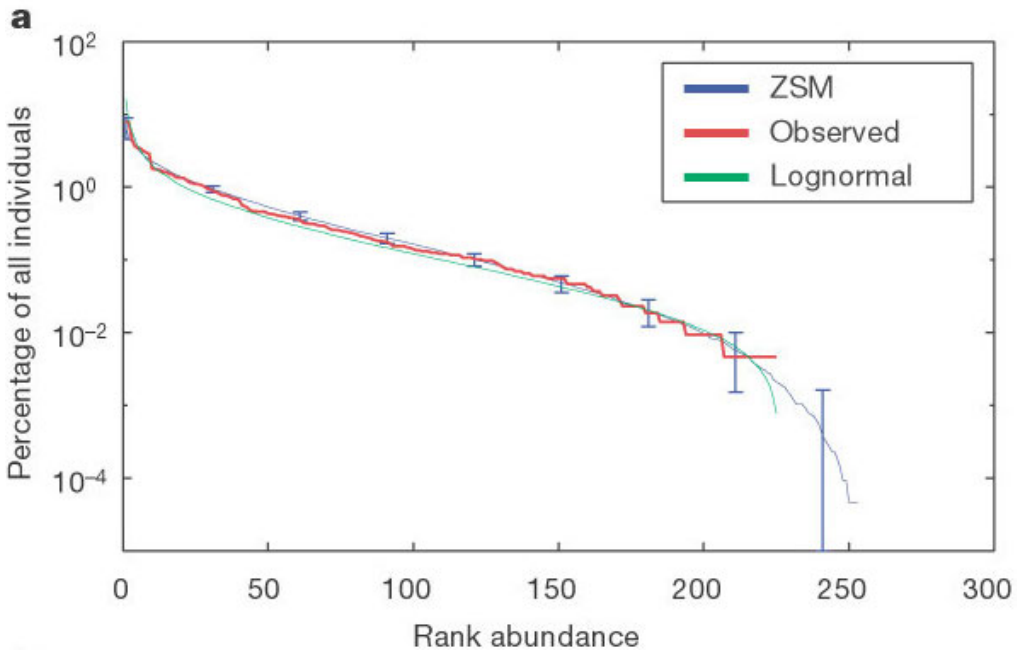
RANK ABUNDANCE CURVES: EVIDENCE FOR THE UNTB?



McGill (2003).

- Hubbell saw evidence for neutral assembly in a neutral model's accuracy in predicting species abundance distributions.

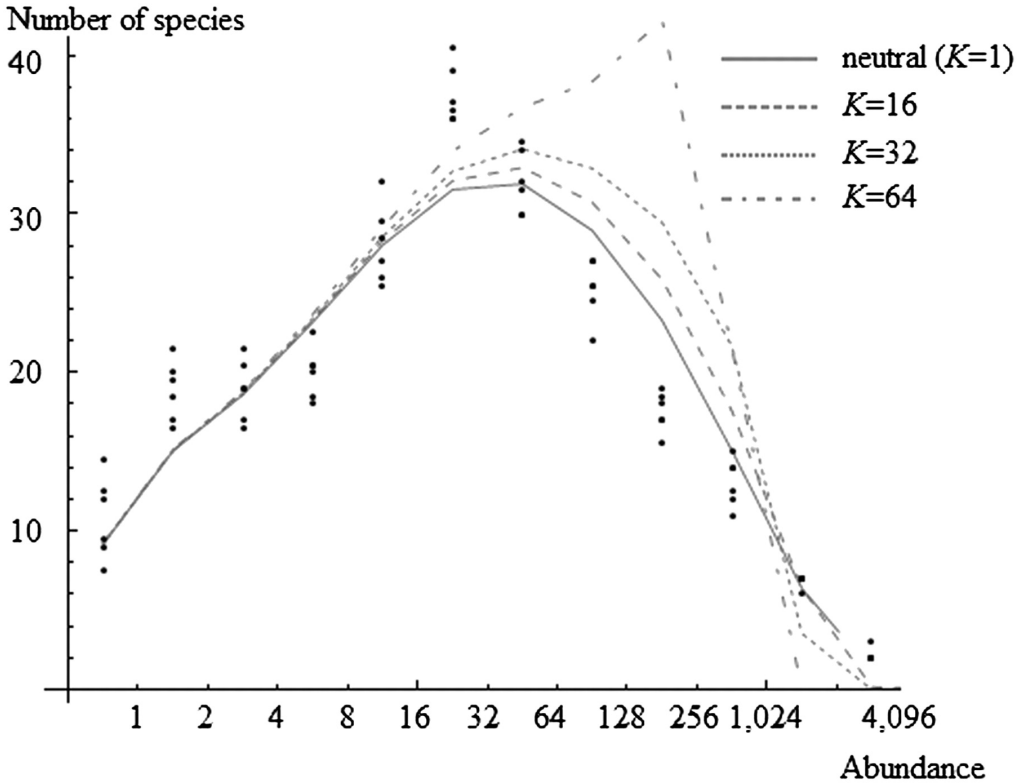
RANK ABUNDANCE CURVES: EVIDENCE FOR THE UNTB?



McGill (2003).

- Hubbell saw evidence for neutral assembly in a neutral model's accuracy in predicting species abundance distributions.
- Others took up the neutral theory, but in the same manner as neutral models in genetics, as a null model.

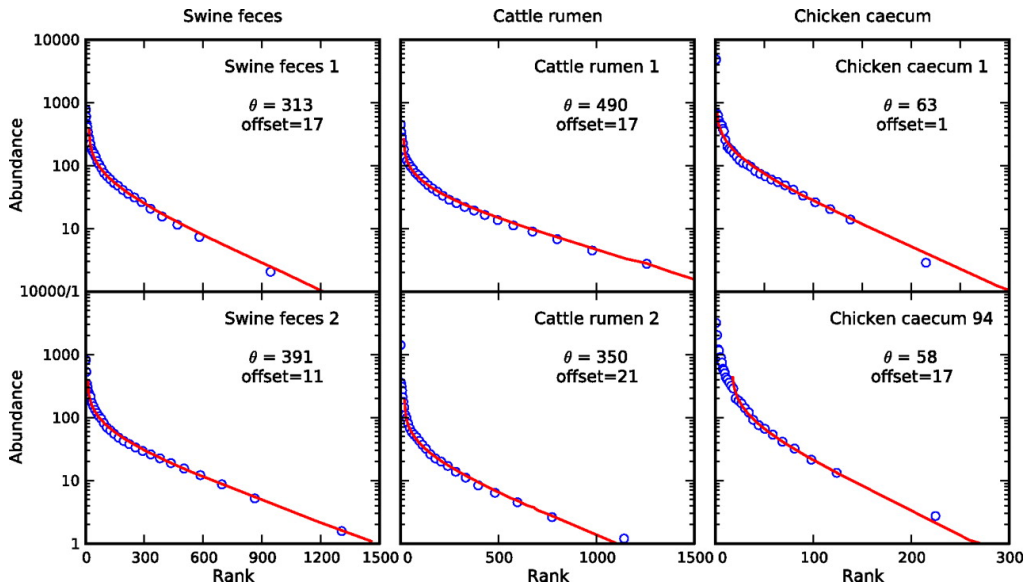
A REJECTABLE NULL?



Chisholme & Pacala (2010).

However, further modelling showed that niche-based models could produce species abundance curves almost identical to the neutral model.

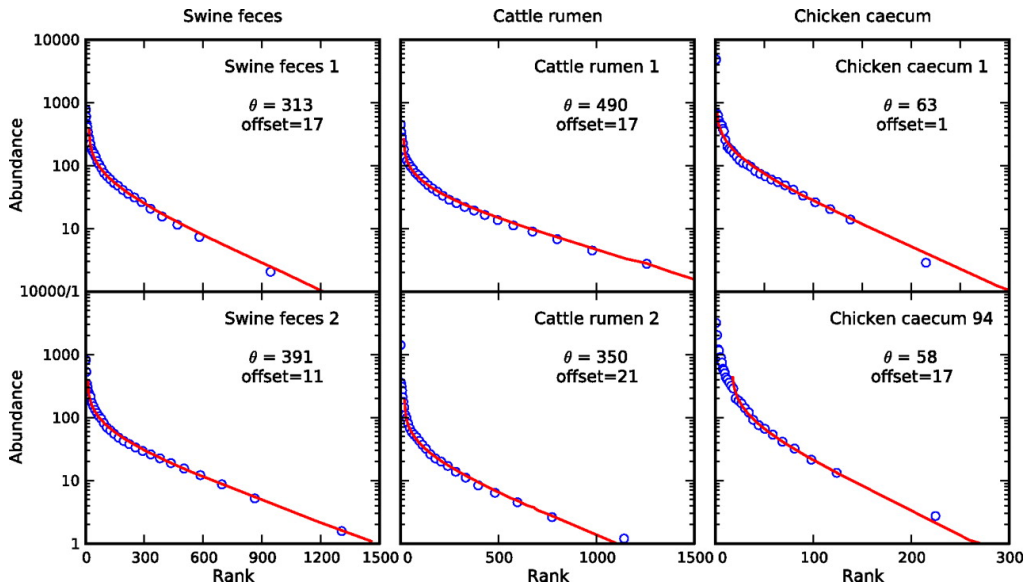
A REJECTABLE NULL?



Jeraldo *et. al.* (2012).

- Moreover, species abundance curves fit from the neutral theory provide excellent matches to communities with known niches.

A REJECTABLE NULL?



Jeraldo *et. al.* (2012).

- Moreover, species abundance curves fit from the neutral theory provide excellent matches to communities with known niches.
- Does this mean we need to reject the UNTB as even a null model?

OBJECTIVES

- To find a robust approximation to a wide class of neutral models, and understand its domain of applicability

OBJECTIVES

- To find a robust approximation to a wide class of neutral models, and understand its domain of applicability
  - We will take inspiration from Cannings' models in posing this class.

OBJECTIVES

- To find a robust approximation to a wide class of neutral models, and understand its domain of applicability
  - We will take inspiration from Cannings' models in posing this class.
- Use this approximation to create a fast and accurate means of simultaneously fit the neutral model to species abundance counts in samples from islands in an island-mainland metacommunity

OBJECTIVES

- To find a robust approximation to a wide class of neutral models, and understand its domain of applicability
  - We will take inspiration from Cannings' models in posing this class.
- Use this approximation to create a fast and accurate means of simultaneously fit the neutral model to species abundance counts in samples from islands in an island-mainland metacommunity
  - multiple islands with distinct immigration rates

OBJECTIVES

- To find a robust approximation to a wide class of neutral models, and understand its domain of applicability
  - We will take inspiration from Cannings' models in posing this class.
- Use this approximation to create a fast and accurate means of simultaneously fit the neutral model to species abundance counts in samples from islands in an island-mainland metacommunity
  - multiple islands with distinct immigration rates
  - potentially large numbers of samples (100s of sites) and large samples (1000s of individuals per site).

OBJECTIVES

- To find a robust approximation to a wide class of neutral models, and understand its domain of applicability
  - We will take inspiration from Cannings' models in posing this class.
- Use this approximation to create a fast and accurate means of simultaneously fit the neutral model to species abundance counts in samples from islands in an island-mainland metacommunity
  - multiple islands with distinct immigration rates
  - potentially large numbers of samples (100s of sites) and large samples (1000s of individuals per site).
- To apply these tools to testing and possibly rejecting the neutral hypothesis for the assembly of the human gut microbiome.

Outline

CANNINGS' MODELS

- In 1974, Chris Cannings proposed a class of haploid population genetic models, which use exchangeability as a general mathematical formulation of neutrality;

CANNINGS' MODELS

- In 1974, Chris Cannings proposed a class of haploid population genetic models, which use exchangeability as a general mathematical formulation of neutrality;
- Random variables $\nu_1, \ldots, \nu_N$ are *exchangeable* if the random vectors $(\nu_{\pi(1)}, \ldots, \nu_{\pi(N)})$ are equal in distribution for all permutations π of $\{1, \ldots, N\}$.

CANNINGS' MODELS

- In 1974, Chris Cannings proposed a class of haploid population genetic models, which use exchangeability as a general mathematical formulation of neutrality;
- Random variables $\nu_1, \ldots, \nu_N$ are *exchangeable* if the random vectors $(\nu_{\pi(1)}, \ldots, \nu_{\pi(N)})$ are equal in distribution for all permutations π of $\{1, \ldots, N\}$.
- Informally, the labels $1, \ldots, N$ are arbitrary, and can be changed without essentially changing the process.

CANNINGS' MODELS

- In 1974, Chris Cannings proposed a class of haploid population genetic models, which use exchangeability as a general mathematical formulation of neutrality;
- Random variables $\nu_1, \ldots, \nu_N$ are *exchangeable* if the random vectors $(\nu_{\pi(1)}, \ldots, \nu_{\pi(N)})$ are equal in distribution for all permutations π of $\{1, \ldots, N\}$.
- Informally, the labels $1, \ldots, N$ are arbitrary, and can be changed without essentially changing the process.
- In a Cannings' model, population size is fixed at N and generations are discrete.

CANNINGS' MODELS

- In 1974, Chris Cannings proposed a class of haploid population genetic models, which use exchangeability as a general mathematical formulation of neutrality;
- Random variables $\nu_1, \ldots, \nu_N$ are *exchangeable* if the random vectors $(\nu_{\pi(1)}, \ldots, \nu_{\pi(N)})$ are equal in distribution for all permutations π of $\{1, \ldots, N\}$.
- Informally, the labels $1, \ldots, N$ are arbitrary, and can be changed without essentially changing the process.
- In a Cannings' model, population size is fixed at N and generations are discrete.
- the i^{th} individual of the n^{th} generation has $\nu_i(n)$ offspring.

CANNINGS' MODELS

- In 1974, Chris Cannings proposed a class of haploid population genetic models, which use exchangeability as a general mathematical formulation of neutrality;
- Random variables $\nu_1, \ldots, \nu_N$ are *exchangeable* if the random vectors $(\nu_{\pi(1)}, \ldots, \nu_{\pi(N)})$ are equal in distribution for all permutations π of $\{1, \ldots, N\}$.
- Informally, the labels $1, \ldots, N$ are arbitrary, and can be changed without essentially changing the process.
- In a Cannings' model, population size is fixed at N and generations are discrete.
- the i^{th} individual of the n^{th} generation has $\nu_i(n)$ offspring.
- $(\nu_1, \ldots, \nu_N)$ is exchangeable.

CANNINGS' MODELS

- In 1974, Chris Cannings proposed a class of haploid population genetic models, which use exchangeability as a general mathematical formulation of neutrality;
- Random variables $\nu_1, \ldots, \nu_N$ are *exchangeable* if the random vectors $(\nu_{\pi(1)}, \ldots, \nu_{\pi(N)})$ are equal in distribution for all permutations π of $\{1, \ldots, N\}$.
- Informally, the labels $1, \ldots, N$ are arbitrary, and can be changed without essentially changing the process.
- In a Cannings' model, population size is fixed at N and generations are discrete.
- the i^{th} individual of the n^{th} generation has $\nu_i(n)$ offspring.
- $(\nu_1, \ldots, \nu_N)$ is exchangeable.
- $\sum_{i=1}^{N} \nu_i = N$.

CANNINGS' MODELS

- In 1974, Chris Cannings proposed a class of haploid population genetic models, which use exchangeability as a general mathematical formulation of neutrality;
- Random variables $\nu_1, \ldots, \nu_N$ are *exchangeable* if the random vectors $(\nu_{\pi(1)}, \ldots, \nu_{\pi(N)})$ are equal in distribution for all permutations π of $\{1, \ldots, N\}$.
- Informally, the labels $1, \ldots, N$ are arbitrary, and can be changed without essentially changing the process.
- In a Cannings' model, population size is fixed at N and generations are discrete.
- the i^{th} individual of the n^{th} generation has $\nu_i(n)$ offspring.
- $(\nu_1, \ldots, \nu_N)$ is exchangeable.
- $\sum_{i=1}^{N} \nu_i = N$.
- *e.g.*, $(\nu_1, \ldots, \nu_N) \sim \text{Multinomial}\left(N, \frac{1}{N}\right)$ in the Wright-Fisher model.

A "CANNINGS" NEUTRAL COMMUNITY MODEL

- Want a broad class of neutral island-mainland metacommunity models that includes the UNTB.

A "CANNINGS" NEUTRAL COMMUNITY MODEL

- Want a broad class of neutral island-mainland metacommunity models that includes the UNTB.
- Use Cannings' models as basis:

A "CANNINGS" NEUTRAL COMMUNITY MODEL

- Want a broad class of neutral island-mainland metacommunity models that includes the UNTB.
- Use Cannings' models as basis:
  - We assume the mainland supports a population of size $N_0 = N$

A "CANNINGS" NEUTRAL COMMUNITY MODEL

- Want a broad class of neutral island-mainland metacommunity models that includes the UNTB.
- Use Cannings' models as basis:
  - We assume the mainland supports a population of size $N_0 = N$
  - Islands $1, \ldots, M$; island i supports a population of size N_i.

A "CANNINGS" NEUTRAL COMMUNITY MODEL

- Want a broad class of neutral island-mainland metacommunity models that includes the UNTB.
- Use Cannings' models as basis:
  - We assume the mainland supports a population of size $N_0 = N$
  - Islands $1, \ldots, M$; island i supports a population of size N_i.
  - We will assume that the islands are all approximately the same size, and asymptotically smaller than the mainland.

A "CANNINGS" NEUTRAL COMMUNITY MODEL

- Want a broad class of neutral island-mainland metacommunity models that includes the UNTB.
- Use Cannings' models as basis:
  - We assume the mainland supports a population of size $N_0 = N$
  - Islands $1, \ldots, M$; island i supports a population of size N_i.
  - We will assume that the islands are all approximately the same size, and asymptotically smaller than the mainland.
  - The j^{th} individual on the i^{th} island (or mainland if $i = 0$) has $\nu_{ij}^{(N)}(n)$ offspring in the n^{th} time step.

A "CANNINGS" NEUTRAL COMMUNITY MODEL

- Want a broad class of neutral island-mainland metacommunity models that includes the UNTB.
- Use Cannings' models as basis:
  - We assume the mainland supports a population of size $N_0 = N$
  - Islands $1, \ldots, M$; island i supports a population of size N_i.
  - We will assume that the islands are all approximately the same size, and asymptotically smaller than the mainland.
  - The j^{th} individual on the i^{th} island (or mainland if $i = 0$) has $\nu_{ij}^{(N)}(n)$ offspring in the n^{th} time step.
  - $(\nu_{i1}^{(N)}(n), \ldots, \nu_{iN_i}^{(N)}(n))$ is exchangeable.

A "CANNINGS" NEUTRAL COMMUNITY MODEL

- Want a broad class of neutral island-mainland metacommunity models that includes the UNTB.
- Use Cannings' models as basis:
  - We assume the mainland supports a population of size $N_0 = N$
  - Islands $1, \ldots, M$; island i supports a population of size N_i.
  - We will assume that the islands are all approximately the same size, and asymptotically smaller than the mainland.
  - The j^{th} individual on the i^{th} island (or mainland if $i = 0$) has $\nu_{ij}^{(N)}(n)$ offspring in the n^{th} time step.
  - $(\nu_{i1}^{(N)}(n), \ldots, \nu_{iN_i}^{(N)}(n))$ is exchangeable.
  - $(\nu_{i1}^{(N)}(n), \ldots, \nu_{iN_i}^{(N)}(n))$ is independent of $(\nu_{j1}^{(N)}(m), \ldots, \nu_{jN_j}^{(N)}(m))$ unless $i = j$ and $m = n$

A "CANNINGS" NEUTRAL COMMUNITY MODEL

- Want a broad class of neutral island-mainland metacommunity models that includes the UNTB.
- Use Cannings' models as basis:
  - We assume the mainland supports a population of size $N_0 = N$
  - Islands $1, \ldots, M$; island i supports a population of size N_i.
  - We will assume that the islands are all approximately the same size, and asymptotically smaller than the mainland.
  - The j^{th} individual on the i^{th} island (or mainland if $i = 0$) has $\nu_{ij}^{(N)}(n)$ offspring in the n^{th} time step.
  - $(\nu_{i1}^{(N)}(n), \ldots, \nu_{iN_i}^{(N)}(n))$ is exchangeable.
  - $(\nu_{i1}^{(N)}(n), \ldots, \nu_{iN_i}^{(N)}(n))$ is independent of $(\nu_{j1}^{(N)}(m), \ldots, \nu_{jN_j}^{(N)}(m))$ unless $i = j$ and $m = n$
- $c_{N_i} := \frac{\mathbb{E}[(\nu_{i1})_2]}{N_i - 1}$; $c_{N_i}^{-1}$ is the *coalescent effective population size* of the i^{th} deme

A "CANNINGS" NEUTRAL COMMUNITY MODEL

- Want a broad class of neutral island-mainland metacommunity models that includes the UNTB.
- Use Cannings' models as basis:
  - We assume the mainland supports a population of size $N_0 = N$
  - Islands $1, \ldots, M$; island i supports a population of size N_i.
  - We will assume that the islands are all approximately the same size, and asymptotically smaller than the mainland.
  - The j^{th} individual on the i^{th} island (or mainland if $i = 0$) has $\nu_{ij}^{(N)}(n)$ offspring in the n^{th} time step.
  - $(\nu_{i1}^{(N)}(n), \ldots, \nu_{iN_i}^{(N)}(n))$ is exchangeable.
  - $(\nu_{i1}^{(N)}(n), \ldots, \nu_{iN_i}^{(N)}(n))$ is independent of $(\nu_{j1}^{(N)}(m), \ldots, \nu_{jN_j}^{(N)}(m))$ unless $i = j$ and $m = n$
- $c_{N_i} := \frac{\mathbb{E}[(\nu_{i1})_2]}{N_i - 1}$; $c_{N_i}^{-1}$ is the *coalescent effective population size* of the i^{th} deme – we will assume that $c_{N_i}^{-1} \ll c_{N_0}^{-1}$ for all $i \geq 1$.

AN INFINTE ALLELES "CANNINGS" NEUTRAL COMMUNITY MODEL

- Still need to incorporate mutation and migration:

AN INFINTE ALLELES "CANNINGS" NEUTRAL COMMUNITY MODEL

- Still need to incorporate mutation and migration:
- Migration:

AN INFINTE ALLELES "CANNINGS" NEUTRAL COMMUNITY MODEL

- Still need to incorporate mutation and migration:
- Migration:
  - With probability $c_{N_i} \frac{I_i}{2} + o(c_{N_i})$ per generation, an individual in the i^{th} deme is replaced by a migrant offspring from another deme.

AN INFINTE ALLELES "CANNINGS" NEUTRAL COMMUNITY MODEL

- Still need to incorporate mutation and migration:
- Migration:
  - With probability $c_{N_i} \frac{I_i}{2} + o(c_{N_i})$ per generation, an individual in the i^{th} deme is replaced by a migrant offspring from another deme.
  - The parent of the migrant is chosen uniformly at random from all parents in all demes.

AN INFINTE ALLELES "CANNINGS" NEUTRAL COMMUNITY MODEL

- Still need to incorporate mutation and migration:
- Migration:
  - With probability $c_{N_i} \frac{I_i}{2} + o(c_{N_i})$ per generation, an individual in the i^{th} deme is replaced by a migrant offspring from another deme.
  - The parent of the migrant is chosen uniformly at random from all parents in all demes.
  - Scaling by c_{N_i} means that the number of migrants is $\mathcal{O}(1)$ (in N).

AN INFINTE ALLELES "CANNINGS" NEUTRAL COMMUNITY MODEL

- Still need to incorporate mutation and migration:
- Migration:
  - With probability $c_{N_i} \frac{I_i}{2} + o(c_{N_i})$ per generation, an individual in the i^{th} deme is replaced by a migrant offspring from another deme.
  - The parent of the migrant is chosen uniformly at random from all parents in all demes.
  - Scaling by c_{N_i} means that the number of migrants is $\mathcal{O}(1)$ (in N).
- Mutation:

AN INFINTE ALLELES "CANNINGS" NEUTRAL COMMUNITY MODEL

- Still need to incorporate mutation and migration:
- Migration:
  - With probability $c_{N_i} \frac{I_i}{2} + o(c_{N_i})$ per generation, an individual in the i^{th} deme is replaced by a migrant offspring from another deme.
  - The parent of the migrant is chosen uniformly at random from all parents in all demes.
  - Scaling by c_{N_i} means that the number of migrants is $\mathcal{O}(1)$ (in N).
- Mutation:
  - Each genotype is given an arbitrary (i.i.d.) label from $[0, 1]$.

AN INFINTE ALLELES "CANNINGS" NEUTRAL COMMUNITY MODEL

- Still need to incorporate mutation and migration:
- Migration:
  - With probability $c_{N_i}\frac{I_i}{2} + o(c_{N_i})$ per generation, an individual in the i^{th} deme is replaced by a migrant offspring from another deme.
  - The parent of the migrant is chosen uniformly at random from all parents in all demes.
  - Scaling by c_{N_i} means that the number of migrants is $\mathcal{O}(1)$ (in N).
- Mutation:
  - Each genotype is given an arbitrary (i.i.d.) label from $[0, 1]$.
  - $P^{(N)}(x, A)$ is the probability that the offspring of an individual of type x has a type in $A \subseteq [0, 1]$.

AN INFINTE ALLELES "CANNINGS" NEUTRAL COMMUNITY MODEL

- Still need to incorporate mutation and migration:
- Migration:
  - With probability $c_{N_i} \frac{I_i}{2} + o(c_{N_i})$ per generation, an individual in the i^{th} deme is replaced by a migrant offspring from another deme.
  - The parent of the migrant is chosen uniformly at random from all parents in all demes.
  - Scaling by c_{N_i} means that the number of migrants is $\mathcal{O}(1)$ (in N).
- Mutation:
  - Each genotype is given an arbitrary (i.i.d.) label from $[0, 1]$.
  - $P^{(N)}(x, A)$ is the probability that the offspring of an individual of type x has a type in $A \subseteq [0, 1]$.
  - $\int f(y)\, P^{(N)}(x, dy) = \left(1 - c_{N_0} \frac{\theta}{2}\right) f(x) + c_{N_0} \frac{\theta}{2} \int f(y)\, dy + o(c_{N_0})$.

EXAMPLE: HUBBELL'S UNTB

- In Hubbell's original model, only a single individual is replaced in each deme at at each time step: $\nu_{ij} \in \{0, 1, 2\}$, with

$$(\nu_{i1}, \ldots, \nu_{iN_i}) = (1, \ldots, 1, 0, 1, \ldots, 1, 2, 1, \ldots, 1)$$

(the vector with i^{th} entry 0 and j^{th} entry 2 for some $i \neq j$) with probability $\frac{2}{N_i(N_i-1)}$

EXAMPLE: HUBBELL'S UNTB

- In Hubbell's original model, only a single individual is replaced in each deme at at each time step: $\nu_{ij} \in \{0, 1, 2\}$, with

$$(\nu_{i1}, \ldots, \nu_{iN_i}) = (1, \ldots, 1, 0, 1, \ldots, 1, 2, 1, \ldots, 1)$$

  (the vector with i^{th} entry 0 and j^{th} entry 2 for some $i \neq j$) with probability $\frac{2}{N_i(N_i-1)}$
- Thus the ν_{ij} are exchangeable.

EXAMPLE: HUBBELL'S UNTB

- In Hubbell's original model, only a single individual is replaced in each deme at at each time step: $\nu_{ij} \in \{0, 1, 2\}$, with

$$(\nu_{i1}, \ldots, \nu_{iN_i}) = (1, \ldots, 1, 0, 1, \ldots, 1, 2, 1, \ldots, 1)$$

  (the vector with i^{th} entry 0 and j^{th} entry 2 for some $i \neq j$) with probability $\frac{2}{N_i(N_i-1)}$
- Thus the ν_{ij} are exchangeable.
- For the UNTB we have $c_{N_i} = \frac{2}{N_i(N_i-1)}$.

EXAMPLE: HUBBELL'S UNTB

- In Hubbell's original model, only a single individual is replaced in each deme at at each time step: $\nu_{ij} \in \{0, 1, 2\}$, with

$$(\nu_{i1}, \ldots, \nu_{iN_i}) = (1, \ldots, 1, 0, 1, \ldots, 1, 2, 1, \ldots, 1)$$

  (the vector with i^{th} entry 0 and j^{th} entry 2 for some $i \neq j$) with probability $\frac{2}{N_i(N_i-1)}$
- Thus the ν_{ij} are exchangeable.
- For the UNTB we have $c_{N_i} = \frac{2}{N_i(N_i-1)}$.
- In Hubbell's model, immigrants are always from the mainland, which is assumed to have a fixed, stationary distribution (so that samples are distributed according to Ewens formula), and no mutations are assumed to occur on the islands.

EXAMPLE: HUBBELL'S UNTB

- In Hubbell's original model, only a single individual is replaced in each deme at at each time step: $\nu_{ij} \in \{0, 1, 2\}$, with

$$(\nu_{i1}, \ldots, \nu_{iN_i}) = (1, \ldots, 1, 0, 1, \ldots, 1, 2, 1, \ldots, 1)$$

  (the vector with i^{th} entry 0 and j^{th} entry 2 for some $i \neq j$) with probability $\frac{2}{N_i(N_i-1)}$
- Thus the ν_{ij} are exchangeable.
- For the UNTB we have $c_{N_i} = \frac{2}{N_i(N_i-1)}$.
- In Hubbell's model, immigrants are always from the mainland, which is assumed to have a fixed, stationary distribution (so that samples are distributed according to Ewens formula), and no mutations are assumed to occur on the islands.
- We will not need to make these assumptions, but will instead derive them (in the limit as $N \to \infty$) as a consequence of the relative size of the mainland and the islands.

Outline

ROBUSTNESS OF THE WRIGHT-FISHER DIFFUSION

- Möhle (2001) showed that if time is rescaled by c_N^{-1}, *i.e.*, by the effective population size, then the frequencies of types in a Cannings' model converge to those given by the Wright-Fisher diffusion, if and only if

$$\lim_{N\to\infty} \frac{\mathbb{E}\left[\nu_1(\nu_1-1)(\nu_1-2)\right]}{N\mathbb{E}\left[\nu_1(\nu_1-1)\right]} = 0.$$

ROBUSTNESS OF THE WRIGHT-FISHER DIFFUSION

- Möhle (2001) showed that if time is rescaled by c_N^{-1}, *i.e.*, by the effective population size, then the frequencies of types in a Cannings' model converge to those given by the Wright-Fisher diffusion, if and only if

$$\lim_{N\to\infty} \frac{\mathbb{E}\left[\nu_1(\nu_1-1)(\nu_1-2)\right]}{N\mathbb{E}\left[\nu_1(\nu_1-1)\right]} = 0.$$

- We can thus say that the Wright-Fisher diffusion is a robust approximation to a broad class of neutral population genetic processes.

ROBUSTNESS OF THE WRIGHT-FISHER DIFFUSION

- Möhle (2001) showed that if time is rescaled by c_N^{-1}, *i.e.*, by the effective population size, then the frequencies of types in a Cannings' model converge to those given by the Wright-Fisher diffusion, if and only if

$$\lim_{N\to\infty} \frac{\mathbb{E}\left[\nu_1(\nu_1-1)(\nu_1-2)\right]}{N\mathbb{E}\left[\nu_1(\nu_1-1)\right]} = 0.$$

- We can thus say that the Wright-Fisher diffusion is a robust approximation to a broad class of neutral population genetic processes.
- We will see that a similar invariance principle exists for the class of Cannings' neutral community models, and moreover its stationary distribution is the Hierarchical Dirichlet Process (Teh, 2006)

INVARIANCE PRINCIPLE I: ISLANDS

- Let $X_{ij}(n) \in [0,1]$ be the type of the j^{th} individual in the i^{th} deme in the n^{th} generation, and let

$$G_i^{(N)}(n) = \frac{1}{N_i} \sum_{j=1}^{N_i} \delta_{X_{ij}(n)}.$$

INVARIANCE PRINCIPLE I: ISLANDS

- Let $X_{ij}(n) \in [0,1]$ be the type of the j^{th} individual in the i^{th} deme in the n^{th} generation, and let

$$G_i^{(N)}(n) = \frac{1}{N_i} \sum_{j=1}^{N_i} \delta_{X_{ij}(n)}.$$

- Suppose that there exists a_N such that

$$\lim_{N\to\infty} \frac{c_{N_i}}{a_N} = \begin{cases} \gamma_i & \text{if } i \geq 1, \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

  exists for all i.

INVARIANCE PRINCIPLE I: ISLANDS

- Let $X_{ij}(n) \in [0,1]$ be the type of the j^{th} individual in the i^{th} deme in the n^{th} generation, and let

$$G_i^{(N)}(n) = \frac{1}{N_i} \sum_{j=1}^{N_i} \delta_{X_{ij}(n)}.$$

- Suppose that there exists a_N such that

$$\lim_{N\to\infty} \frac{c_{N_i}}{a_N} = \begin{cases} \gamma_i & \text{if } i \geq 1, \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

  exists for all i.
- Then, if $G_i^{(N)}(0) \overset{w}{\Longrightarrow} G_i(0)$, a probability measure on $[0,1]$, then

$$G_i^{(N)}(\lfloor a_N^{-1} t \rfloor) \overset{w}{\Longrightarrow} G_i(\gamma_i t),$$

  where $G_0(t) \equiv G_0(0)$ for all $t \geq 0$ and $G_i(t)$ satisfies the *infinite alleles model* with base measure $G_0(0)$ and "mutation rate" I_i.

INVARIANCE PRINCIPLE I: ISLANDS

- Let $X_{ij}(n) \in [0,1]$ be the type of the j^{th} individual in the i^{th} deme in the n^{th} generation, and let

$$G_i^{(N)}(n) = \frac{1}{N_i} \sum_{j=1}^{N_i} \delta_{X_{ij}(n)}.$$

- Suppose that there exists a_N such that

$$\lim_{N\to\infty} \frac{c_{N_i}}{a_N} = \begin{cases} \gamma_i & \text{if } i \geq 1, \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

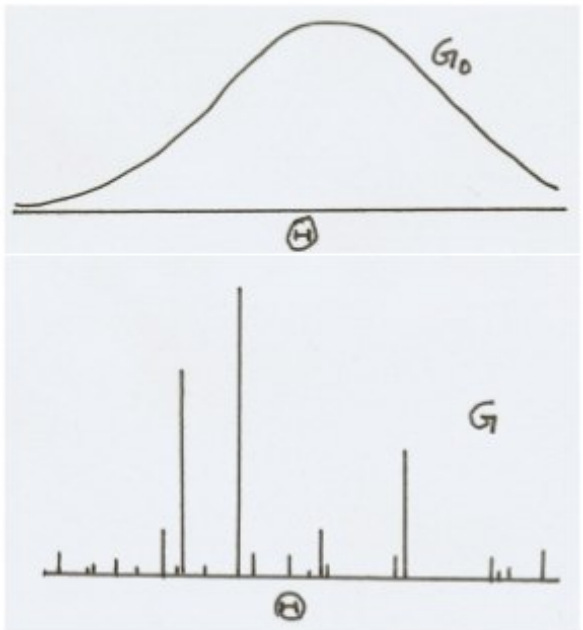
  exists for all i.
- Then, if $G_i^{(N)}(0) \overset{w}{\Longrightarrow} G_i(0)$, a probability measure on $[0,1]$, then

$$G_i^{(N)}(\lfloor a_N^{-1} t \rfloor) \overset{w}{\Longrightarrow} G_i(\gamma_i t),$$

  where $G_0(t) \equiv G_0(0)$ for all $t \geq 0$ and $G_i(t)$ satisfies the *infinite alleles model* with base measure $G_0(0)$ and "mutation rate" I_i.
- Further, conditional on $G_0(0)$, the $G_i^{(N)}(0)$ are independent.

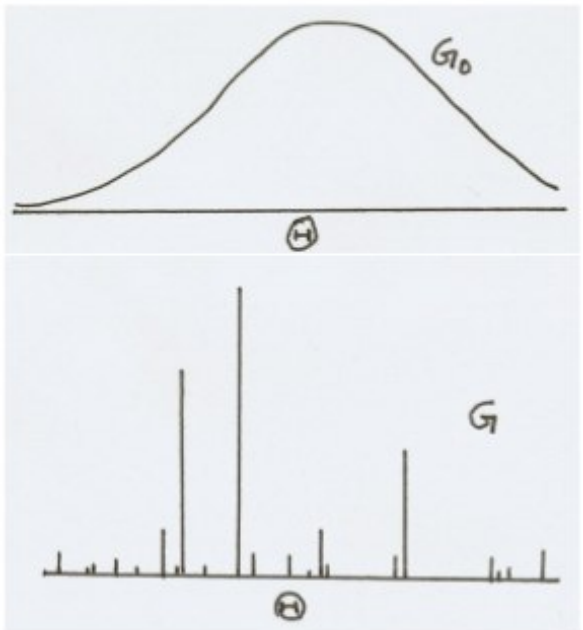
INVARIANCE PRINCIPLE I: STATIONARY DIRICHLET PROCESS

- As $t \to \infty$, the infinite alleles model with mutation rate θ and base measure μ tends to a stationary distribution with law $DP(\theta, \mu)$



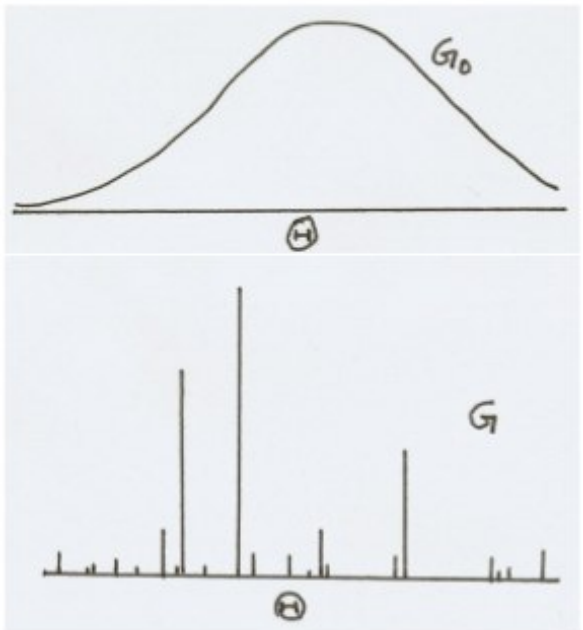
INVARIANCE PRINCIPLE I: STATIONARY DIRICHLET PROCESS

- As $t \to \infty$, the infinite alleles model with mutation rate θ and base measure μ tends to a stationary distribution with law $DP(\theta, \mu)$
- A Dirichlet Process with law $DP(\theta, \mu)$ can be constructed via stick breaking:
  - Draw $\beta_i' \sim \text{Beta}(1, \theta)$.
  - Set $\beta_i = \beta_i' \prod_{j=1}^{i-1}(1 - \beta_j')$.
  - Let μ be a probability measure on a space Θ and let $X_i \sim \mu$ be i.i.d.
  - $\sum_{i=1}^{\infty} \beta_i \delta_{X_i}$ is a random variable with law $\text{DP}(\theta, \mu)$
  - $\mathbb{E}\left[\sum_{i=1}^{\infty} \beta_i \delta_{X_i}\right] = \mu$.



INVARIANCE PRINCIPLE I: STATIONARY DIRICHLET PROCESS

- As $t \to \infty$, the infinite alleles model with mutation rate θ and base measure μ tends to a stationary distribution with law $DP(\theta, \mu)$
- A Dirichlet Process with law $DP(\theta, \mu)$ can be constructed via stick breaking:
  - Draw $\beta_i' \sim \text{Beta}(1, \theta)$.
  - Set $\beta_i = \beta_i' \prod_{j=1}^{i-1}(1 - \beta_j')$.
  - Let μ be a probability measure on a space Θ and let $X_i \sim \mu$ be i.i.d.
  - $\sum_{i=1}^{\infty} \beta_i \delta_{X_i}$ is a random variable with law $\text{DP}(\theta, \mu)$
  - $\mathbb{E}\left[\sum_{i=1}^{\infty} \beta_i \delta_{X_i}\right] = \mu$.
- In particular, samples from a $\text{DP}(\theta, \mu)$ r.v. can be generated using Aldous' *Chinese restaurant process* and are distributed according to *Ewens' sampling formula*.



INVARIANCE PRINCIPLE II: MAINLAND

- Further, if we assume that $G_i^{(N)}(0) \stackrel{w}{\Longrightarrow} G_i(0)$, where $G_i(0) \sim \mathrm{DP}(I_i, G_0(0))$, *i.e.*, we assume that the islands are already at their stationary state, then

$$G_0^{(N)}(\lfloor c_{N_0}^{-1} t \rfloor) \stackrel{w}{\Longrightarrow} G_0(t),$$

where G_0 is the infinite alleles process with Lebesgue measure on $[0, 1]$, λ, as base measure and mutation rate θ.

INVARIANCE PRINCIPLE II: MAINLAND

- Further, if we assume that $G_i^{(N)}(0) \stackrel{w}{\Longrightarrow} G_i(0)$, where $G_i(0) \sim \mathrm{DP}(I_i, G_0(0))$, *i.e.*, we assume that the islands are already at their stationary state, then

$$G_0^{(N)}(\lfloor c_{N_0}^{-1} t \rfloor) \stackrel{w}{\Longrightarrow} G_0(t),$$

  where G_0 is the infinite alleles process with Lebesgue measure on $[0, 1]$, λ, as base measure and mutation rate θ.
- This tends, as $t \to \infty$, to a stationary distribution $\mathrm{DP}(\theta, \lambda)$.

INVARIANCE PRINCIPLE II: MAINLAND

- Further, if we assume that $G_i^{(N)}(0) \overset{w}{\Longrightarrow} G_i(0)$, where $G_i(0) \sim \mathrm{DP}(I_i, G_0(0))$, *i.e.*, we assume that the islands are already at their stationary state, then

$$G_0^{(N)}(\lfloor c_{N_0}^{-1} t \rfloor) \overset{w}{\Longrightarrow} G_0(t),$$

  where G_0 is the infinite alleles process with Lebesgue measure on $[0, 1]$, λ, as base measure and mutation rate θ.
- This tends, as $t \to \infty$, to a stationary distribution $\mathrm{DP}(\theta, \lambda)$.
- In particular, if we assume that $G_0(0) \sim \mathrm{DP}(\theta, \lambda)$, then the process is stationary, and the islands to *hierarchical Dirichlet processes* (HDP) with laws $\mathrm{DP}(I_i, G_0(0))$.

INVARIANCE PRINCIPLE II: MAINLAND

- Further, if we assume that $G_i^{(N)}(0) \overset{w}{\Longrightarrow} G_i(0)$, where $G_i(0) \sim \mathrm{DP}(I_i, G_0(0))$, *i.e.*, we assume that the islands are already at their stationary state, then

$$G_0^{(N)}(\lfloor c_{N_0}^{-1} t \rfloor) \overset{w}{\Longrightarrow} G_0(t),$$

  where G_0 is the infinite alleles process with Lebesgue measure on $[0, 1]$, λ, as base measure and mutation rate θ.
- This tends, as $t \to \infty$, to a stationary distribution $\mathrm{DP}(\theta, \lambda)$.
- In particular, if we assume that $G_0(0) \sim \mathrm{DP}(\theta, \lambda)$, then the process is stationary, and the islands to *hierarchical Dirichlet processes* (HDP) with laws $\mathrm{DP}(I_i, G_0(0))$.
- We can thus apply the extensive statistical machinery for the HDP developed in the machine learning literature.

Outline

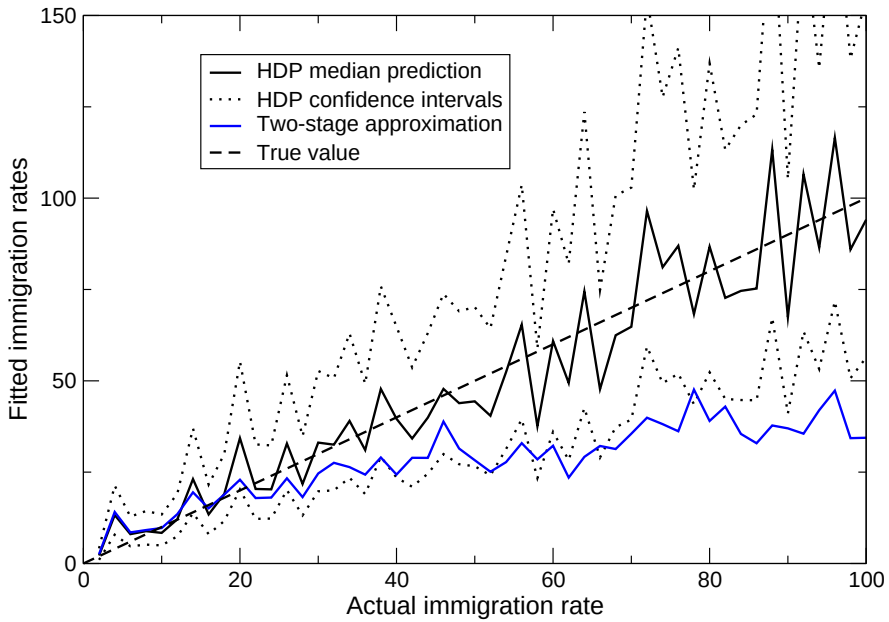
An Efficient MCMC Algorithm

- The HDP lends itself very well to inference via a Gibbs sampler:
  - Gibbs sampling is a special case of Metropolis–Hastings.
  - Construct an ergodic discrete time Markov chain such that the joint probability distribution of the parameters, conditional on the observed data, is its stationary distribtution.
  - Parameters are the rescaled mutation and migration rates, θ and the I_i.
  - Data are counts X_{ij} of incidence of species j in community i.
  - Use the auxiliary variables approach of Escobar & West (1995),
  - Gamma priors for our parameters.

An Efficient MCMC Algorithm

- The HDP lends itself very well to inference via a Gibbs sampler:
  - Gibbs sampling is a special case of Metropolis–Hastings.
  - Construct an ergodic discrete time Markov chain such that the joint probability distribution of the parameters, conditional on the observed data, is its stationary distribtution.
  - Parameters are the rescaled mutation and migration rates, θ and the I_i.
  - Data are counts X_{ij} of incidence of species j in community i.
  - Use the auxiliary variables approach of Escobar & West (1995),
  - Gamma priors for our parameters.
- Ergodicity of the Markov chain allows us to average over repeated samples to obtain expectations of arbitrary functions of the parameters.
  - The expected value of the parameters gives the Bayes' minimum square estimators (MSE), $\hat{\theta}$, $\hat{I}_i$.
  - We can easily determine the full posterior distribution or any related summary statistics.

TESTING THE GIBBS SAMPLER I: SIMULATED DATA



Estimated immigration rates vs. true values for the UNTB-HDP model fit to a neutral model simulation. Predictions are medians (solid line) from 25,000 posterior samples together with lower (2.5%) and upper (97.5%) Bayesian confidence intervals (dotted lines). The predictions from the two-stage approximation are also given (blue line).

TESTING THE GIBBS SAMPLER II: PANAMA CANAL ZONE DATASETS

Method	θ	I_{BCI}	I_C	I_S
Etienne fixed I	259	44.2	44.2	44.2
Etienne approx	342	53.7	30.8	33.9
Etienne exact	235 ± 23	65.3 ± 5.9	31.5 ± 3.9	35.7 ± 3.9
HDP approx	231 ± 22	65.5 ± 5.9	31.6 ± 3.8	35.8 ± 3.9

- Neutral parameter estimates for samples from three local tree communities (Sherman, BCI and Cocoli) in the Panama Canal Zone using previous approaches and the hierarchical Dirichlet process approximation. Standard errors are given for the methods where they are available.

TESTING THE GIBBS SAMPLER II: PANAMA CANAL ZONE DATASETS

Method	θ	I_{BCI}	I_C	I_S
Etienne fixed I	259	44.2	44.2	44.2
Etienne approx	342	53.7	30.8	33.9
Etienne exact	235 ± 23	65.3 ± 5.9	31.5 ± 3.9	35.7 ± 3.9
HDP approx	231 ± 22	65.5 ± 5.9	31.6 ± 3.8	35.8 ± 3.9

- Neutral parameter estimates for samples from three local tree communities (Sherman, BCI and Cocoli) in the Panama Canal Zone using previous approaches and the hierarchical Dirichlet process approximation. Standard errors are given for the methods where they are available.
- HDP gives good agreement with exact methods.

TESTING THE GIBBS SAMPLER II: PANAMA CANAL ZONE DATASETS

Method	θ	I_{BCI}	I_C	I_S
Etienne fixed I	259	44.2	44.2	44.2
Etienne approx	342	53.7	30.8	33.9
Etienne exact	235 ± 23	65.3 ± 5.9	31.5 ± 3.9	35.7 ± 3.9
HDP approx	231 ± 22	65.5 ± 5.9	31.6 ± 3.8	35.8 ± 3.9

- Neutral parameter estimates for samples from three local tree communities (Sherman, BCI and Cocoli) in the Panama Canal Zone using previous approaches and the hierarchical Dirichlet process approximation. Standard errors are given for the methods where they are available.
- HDP gives good agreement with exact methods.
- When we applied our HDP approach to 29 Panamanian tropical tree communities, we found a significant negative correlation between distance and estimated immigration rate – the latter are informative.

Outline

CAN NEUTRALITY EXPLAIN DIVERSITY IN THE GUT MICROBIOME?

- Microbial communities play functionally important roles in many ecosystems yet are rich in diversity.

CAN NEUTRALITY EXPLAIN DIVERSITY IN THE GUT MICROBIOME?

- Microbial communities play functionally important roles in many ecosystems yet are rich in diversity.
- Such systems might, a priori, be expected to contain at least subpopulations shaped primarily by stochastic forces.

CAN NEUTRALITY EXPLAIN DIVERSITY IN THE GUT MICROBIOME?

- Microbial communities play functionally important roles in many ecosystems yet are rich in diversity.
- Such systems might, a priori, be expected to contain at least subpopulations shaped primarily by stochastic forces.
- Jeraldo, *et. al.* (2012) found that neutrality could not be rejected using species abundance curves.

CAN NEUTRALITY EXPLAIN DIVERSITY IN THE GUT MICROBIOME?

- Microbial communities play functionally important roles in many ecosystems yet are rich in diversity.
- Such systems might, a priori, be expected to contain at least subpopulations shaped primarily by stochastic forces.
- Jeraldo, *et. al.* (2012) found that neutrality could not be rejected using species abundance curves.
- However, evidence of clustering of gut microbiota into different enterotypes (Arumugam et al., 2011; Holmes et al., 2012), which implies non- neutral structuring at the whole community level.

Can Neutrality Explain Diversity in the Gut Microbiome?

- Microbial communities play functionally important roles in many ecosystems yet are rich in diversity.
- Such systems might, a priori, be expected to contain at least subpopulations shaped primarily by stochastic forces.
- Jeraldo, *et. al.* (2012) found that neutrality could not be rejected using species abundance curves.
- However, evidence of clustering of gut microbiota into different enterotypes (Arumugam et al., 2011; Holmes et al., 2012), which implies non- neutral structuring at the whole community level.
- We explored this by subdividing the species according to their taxa at multiple taxonomic levels; it should be increasingly the case that species in smaller clades occupy similar community roles and for neutrality we may only require this in the broadest sense, e.g., methanogens vs. sulphate reducers.

METHODS

- Used gut microbiome data from twins and their mothers (Turnbaugh et al., 2009): faecal samples from 154 individuals, characterized by family and body mass index (BMI), at two time points two months apart.
- The V2 hypervariable region of the 16S rRNA gene was amplified by PCR and then sequenced using 454 and de-noised using the AmpliconNoise pipeline (Quince et al., 2009, 2011).
- 570,851 reads split over 278 samples – sample size 53 to 10,580 with a median of 1,598.
- 19,647 unique sequences following noise removal, taxonomically classified using the RDP stand-alone classifier of Wang et al. (2007), split by phylum, using a cut-off of 70% bootstrap confidence
- Constructed 7,238 Operational Taxonomic Units (OTUs) at 3% sequence difference using average linkage clustering (Youssef et al., 2009)
- Fitted UNTB-HDP to each phylum, family and genus separately
- Only samples with > 100 representatives from a taxa were included, only fit to taxa with > 50 such samples.
- Used Monte Carlo significance test for neutrality from Etienne (2007): p_N metacommunity p-value; p_L 'local' p-value.

FITTING THE UNTB-HDP MODEL TO HUMAN GUT MICROBIOTA

Taxa	%age	#samples	#3% OTU	$\hat{\theta}$	$\hat{I}_i$	p_N	p_L
Bacteroidetes	29.9	249	585	151.32	1.46-5.57-13.52	0.0	0.0
Bacteroidaceae	23.6	221	224	50.79	0.85-3.33-7.70	0.0	0.05
Bacteroides	23.6	238	227	51.01	0.71-3.33-7.80	0.0	0.04
Rikenellaceae							
Alistipes	2.22	66	40	8.72	0.33-2.38-11.10	0.02	0.77
Firmicutes	66.1	277	4771	1383.38	21.44-44.82-80.81	0.0	0.0
Incertae Sedis XIV	7.56	124	217	47.75	2.16-9.88-27.42	0.0	0.06
Blautia	7.55	197	252	52.01	2.26-10.62-34.18	0.0	0.14
Lachnospiraceae	12.4	230	1076	314.66	6.57-13.28-23.91	0.0	0.0
Roseburia	2.61	87	124	38.21	0.40-2.41-7.12	0.0	0.18
Ruminococcaceae	24.2	257	1489	412.68	4.19-15.73-38.20	0.0	0.0
Faecalibacterium	12.1	236	369	84.27	0.66-6.87-21.46	0.0	0.01
Oscillibacter	1.90	58	72	19.33	0.84-3.23-8.16	0.068	0.34
Ruminococcus	1.74	60	35	10.99	0.00-0.38-1.94	0.013	0.65
Subdoligranulum	2.95	94	86	23.14	0.22-1.67-8.50	0.0	0.32

- $\hat{\theta}$: median over 25,000 Gibbs samples
- $\hat{I}_i$: the fitted immigration rates (lower 2.5% – median – upper 97.5% quantiles) over 25,000 Gibbs samples

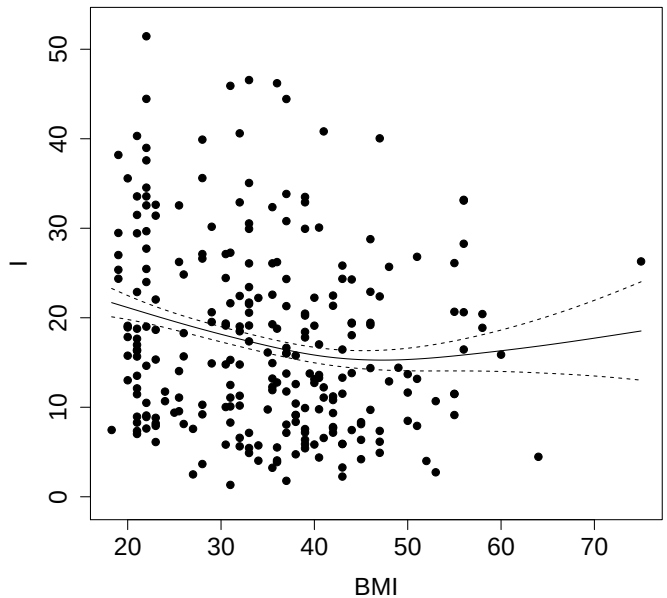
FITTING THE UNTB-HDP MODEL TO HUMAN GUT MICROBIOTA

Taxa	%age	#samples	#3% OTU	$\hat{\theta}$	$\hat{I}_i$	p_N	p_L
Bacteroidetes	29.9	249	585	151.32	1.46-5.57-13.52	0.0	0.0
Bacteroidaceae	23.6	221	224	50.79	0.85-3.33-7.70	0.0	0.05
Bacteroides	23.6	238	227	51.01	0.71-3.33-7.80	0.0	0.04
Rikenellaceae							
Alistipes	2.22	66	40	8.72	0.33-2.38-11.10	0.02	0.77
Firmicutes	66.1	277	4771	1383.38	21.44-44.82-80.81	0.0	0.0
Incertae Sedis XIV	7.56	124	217	47.75	2.16-9.88-27.42	0.0	0.06
Blautia	7.55	197	252	52.01	2.26-10.62-34.18	0.0	0.14
Lachnospiraceae	12.4	230	1076	314.66	6.57-13.28-23.91	0.0	0.0
Roseburia	2.61	87	124	38.21	0.40-2.41-7.12	0.0	0.18
Ruminococcaceae	24.2	257	1489	412.68	4.19-15.73-38.20	0.0	0.0
Faecalibacterium	12.1	236	369	84.27	0.66-6.87-21.46	0.0	0.01
Oscillibacter	1.90	58	72	19.33	0.84-3.23-8.16	0.068	0.34
Ruminococcus	1.74	60	35	10.99	0.00-0.38-1.94	0.013	0.65
Subdoligranulum	2.95	94	86	23.14	0.22-1.67-8.50	0.0	0.32

- $\hat{\theta}$: median over 25,000 Gibbs samples
- $\hat{I}_i$: the fitted immigration rates (lower 2.5% – median – upper 97.5% quantiles) over 25,000 Gibbs samples
- The immigration rates are again informative: they are much lower for the Bacteroides than the Firmicutes, probably reflecting the fact that the latter are spore-forming.

IMMIGRATION RATES *vs.* BMI

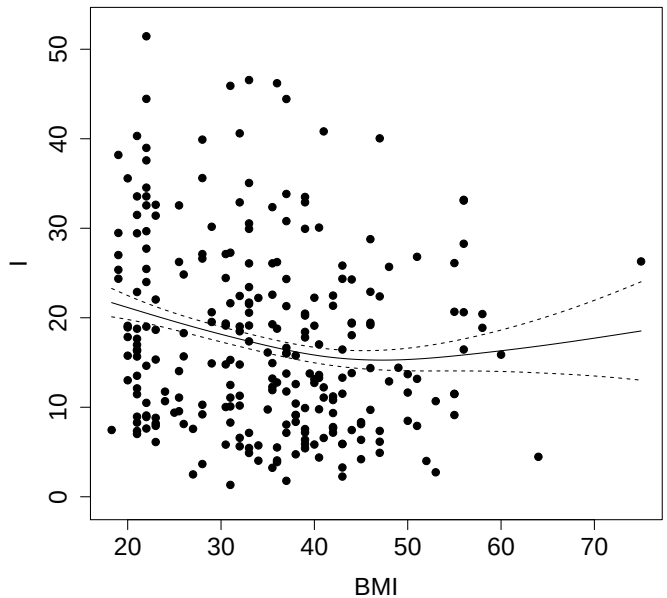
.



- Moreover, when we compared fitted migration rates against BMI, a significant negative correlation is observed (p-value = 0.006776 - Pearson's correlation) in Ruminococcaceae and Firmicutes.

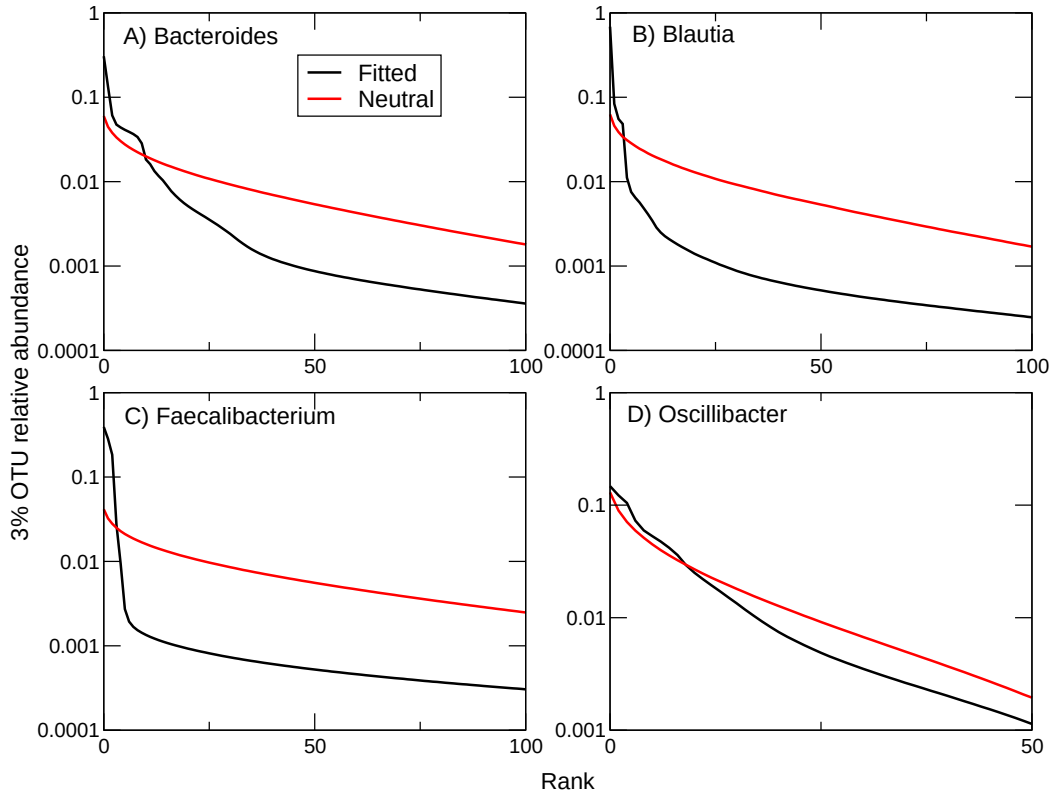
IMMIGRATION RATES *vs.* BMI

.



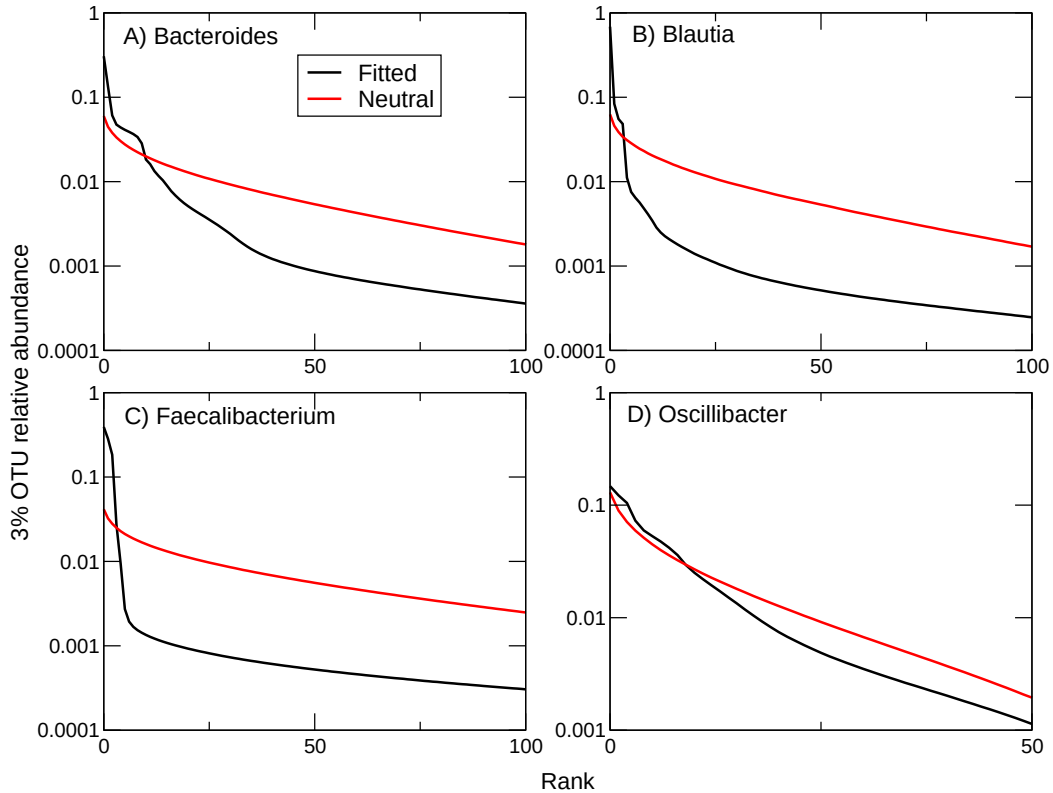
- Moreover, when we compared fitted migration rates against BMI, a significant negative correlation is observed (p-value = 0.006776 - Pearson's correlation) in Ruminococcaceae and Firmicutes.
- Increasing nutrient supply counterintuitively reduces local biodiversity.

FITTED RANK-ABUNDANCE CURVES



- Empricial metacommunity distributions (black line) and neutral metacommunity predictions (red line) for four genera: A) Bacteroides, B) Blautia, C) Faecalibacterium and D) Oscillibacter

FITTED RANK-ABUNDANCE CURVES



- Empricial metacommunity distributions (black line) and neutral metacommunity predictions (red line) for four genera: A) Bacteroides, B) Blautia, C) Faecalibacterium and D) Oscillibacter
- Unlike Jeraldo, *et. al.* (2012), when we fit the whole model, as opposed to the rank-abundance curve, we can reject the neutral hypothesis.

SUMMARY

- The UNTB-HDP Gibbs sampler, can fit large multi-sample data sets in a way that can detect deviations of the metacommunity from neutrality whilst still inferring informalive immigration rates.

Summary

- The UNTB-HDP Gibbs sampler, can fit large multi-sample data sets in a way that can detect deviations of the metacommunity from neutrality whilst still inferring informalive immigration rates.
- Under the assumption of large samples, it is invariant under a variety of neutral models, and is thus a good test of the neutral hypothesis, rather than the specifics of the model.
- The resulting significance tests and fitted parameters reveal a great deal about the ecology of the human gut microbiota:

Summary

- The UNTB-HDP Gibbs sampler, can fit large multi-sample data sets in a way that can detect deviations of the metacommunity from neutrality whilst still inferring informalive immigration rates.
- Under the assumption of large samples, it is invariant under a variety of neutral models, and is thus a good test of the neutral hypothesis, rather than the specifics of the model.
- The resulting significance tests and fitted parameters reveal a great deal about the ecology of the human gut microbiota:
  - Only at the genus level do we consistently see evidence of neutral local community assembly in the gut; however, neutral local community assembly may be operating within the species occupying those roles, and that neutral processes may be responsible for maintaining some of the vast diversity that is observed in the human microbiota.

Summary

- The UNTB-HDP Gibbs sampler, can fit large multi-sample data sets in a way that can detect deviations of the metacommunity from neutrality whilst still inferring informalive immigration rates.
- Under the assumption of large samples, it is invariant under a variety of neutral models, and is thus a good test of the neutral hypothesis, rather than the specifics of the model.
- The resulting significance tests and fitted parameters reveal a great deal about the ecology of the human gut microbiota:
  - Only at the genus level do we consistently see evidence of neutral local community assembly in the gut; however, neutral local community assembly may be operating within the species occupying those roles, and that neutral processes may be responsible for maintaining some of the vast diversity that is observed in the human microbiota.
  - The immigration rates, are also informative; for the family Ruminococcaceae and phylum Firmicutes, they correlated negatively with body mass index; they are much lower for the Bacteroides than the Firmicutes, reflecting the much higher tendency for the latter to be spore-forming.

Summary

- The UNTB-HDP Gibbs sampler, can fit large multi-sample data sets in a way that can detect deviations of the metacommunity from neutrality whilst still inferring informalive immigration rates.
- Under the assumption of large samples, it is invariant under a variety of neutral models, and is thus a good test of the neutral hypothesis, rather than the specifics of the model.
- The resulting significance tests and fitted parameters reveal a great deal about the ecology of the human gut microbiota:
  - Only at the genus level do we consistently see evidence of neutral local community assembly in the gut; however, neutral local community assembly may be operating within the species occupying those roles, and that neutral processes may be responsible for maintaining some of the vast diversity that is observed in the human microbiota.
  - The immigration rates, are also informative; for the family Ruminococcaceae and phylum Firmicutes, they correlated negatively with body mass index; they are much lower for the Bacteroides than the Firmicutes, reflecting the much higher tendency for the latter to be spore-forming.
- Moreover, we have formally linked a model from ecology, the UNTB, with a highly flexible model from machine learning, the hierarchical Dirichlet process; we hope that the connection we have made here will lead to further hierarchical Bayesian modelling in ecology.

Thank you!

Outline

A Community Model With Tradeoffs, No Fixed N

$$X_i^N(t) = \text{the number of individuals of species } i \text{ at time } t$$

Event	Rate
$X_i^N(t) \to X_i^N(t) + 1$	$\left(\beta_i X_i^N(t) + m_i\right)\left(\frac{N - \sum_{j=1}^K X_j^N(t)}{N} + \sum_{j=1}^K \kappa_{ij} \frac{X_j^N(t)}{N}\right)$
$X_i^N(t) \to X_i^N(t) - 1$	$\delta_i X_i^N(t)$

$$X_i^N(t) = \text{the number of individuals of species } i \text{ at time } t$$

Event	Rate
$X_i^N(t) \to X_i^N(t) + 1$	$\left(\beta_i X_i^N(t) + m_i\right)\left(\frac{N - \sum_{j=1}^K X_j^N(t)}{N} + \sum_{j=1}^K \kappa_{ij} \frac{X_j^N(t)}{N}\right)$
$X_i^N(t) \to X_i^N(t) - 1$	$\delta_i X_i^N(t)$

- Individuals of species i give birth at per-capita rate β_i.

$$X_i^N(t) = \text{the number of individuals of species } i \text{ at time } t$$

Event	Rate
$X_i^N(t) \to X_i^N(t) + 1$	$\left(\beta_i X_i^N(t) + m_i\right)\left(\frac{N - \sum_{j=1}^K X_j^N(t)}{N} + \sum_{j=1}^K \kappa_{ij} \frac{X_j^N(t)}{N}\right)$
$X_i^N(t) \to X_i^N(t) - 1$	$\delta_i X_i^N(t)$

- Individuals of species i give birth at per-capita rate β_i.
- They die at rate δ_i.

$$X_i^N(t) = \text{the number of individuals of species } i \text{ at time } t$$

Event	Rate
$X_i^N(t) \to X_i^N(t) + 1$	$\left(\beta_i X_i^N(t) + m_i\right)\left(\frac{N - \sum_{j=1}^K X_j^N(t)}{N} + \sum_{j=1}^K \kappa_{ij} \frac{X_j^N(t)}{N}\right)$
$X_i^N(t) \to X_i^N(t) - 1$	$\delta_i X_i^N(t)$

- Individuals of species i give birth at per-capita rate β_i.
- They die at rate δ_i.
- Individuals of species i immigrate from the mainland at rate $m_i = \frac{\varpi_i}{N}$.

$$X_i^N(t) = \text{the number of individuals of species } i \text{ at time } t$$

Event	Rate
$X_i^N(t) \to X_i^N(t) + 1$	$\left(\beta_i X_i^N(t) + m_i\right)\left(\frac{N - \sum_{j=1}^K X_j^N(t)}{N} + \sum_{j=1}^K \kappa_{ij} \frac{X_j^N(t)}{N}\right)$
$X_i^N(t) \to X_i^N(t) - 1$	$\delta_i X_i^N(t)$

- Individuals of species i give birth at per-capita rate β_i.
- They die at rate δ_i.
- Individuals of species i immigrate from the mainland at rate $m_i = \frac{\varpi_i}{N}$.
- Individuals survive if they find an empty patch or if they out-compete an individual in an occupied patch.

For the community model with tradeoffs, long-term coexistence is only possible if $\kappa_{ij} \equiv \kappa$ and

$$\frac{\beta_i}{\delta_i} \equiv 1 - \nu.$$

For $t \in (\delta, \infty)$ for any fixed $\delta > 0$, the relative frequency process, $\mathbf{P}(t) = \frac{1}{\sum_{i=1}^{K} X_i(t)} \mathbf{X}(Nt)$ is a diffusion on the standard simplex with generator:

$$\begin{aligned}
\mathcal{L}f = \sum_{i=1}^{K-1} \frac{(1-\nu)(1-\kappa)}{\nu} \Bigg[\mu_i - \frac{\mu_. \beta_i p_i}{\sum_{k=1}^{K} \beta_k p_k} \\
+ \frac{\beta_i p_i}{\left(\frac{1}{\alpha} - 1\right) \left(\sum_{k=1}^{K} \beta_k p_k\right)_2} \sum_{k=1}^{K} (\beta_k - \beta_i) \beta_k p_k \Bigg] \partial_{p_i} f \\
+ \sum_{i=1}^{K-1} \sum_{j=1}^{K-1} \frac{(1-\nu)(1-\kappa)}{\nu} \beta_i p_i \left(\delta_{ij} - \frac{\beta_j p_j}{\sum_{k=1}^{K} \beta_k p_k} \right) \partial_{p_i} \partial_{p_j} f.
\end{aligned}$$

- The probability of having relative species abundances $(p_1, \ldots, p_K)$ is Dirichlet distributed:

$$\frac{\Gamma\left(\sum_{i=1}^{K} \frac{\varpi_i}{\beta_i}\right)}{\prod_{i=1}^{K} \Gamma\left(\frac{\varpi_i}{\beta_i}\right)} \prod_{i=1}^{K} p_i^{\frac{\varpi_i}{\beta_i} - 1}$$

- The probability of having relative species abundances $(p_1, \ldots, p_K)$ is Dirichlet distributed:

$$\frac{\Gamma\left(\sum_{i=1}^{K} \frac{\varpi_i}{\beta_i}\right)}{\prod_{i=1}^{K} \Gamma\left(\frac{\varpi_i}{\beta_i}\right)} \prod_{i=1}^{K} p_i^{\frac{\varpi_i}{\beta_i} - 1}$$

- This offers one potential explanation for the effectiveness of the UNTB in fitting real species abundance data: if we assume migration rates are proportional to birth rates ($\varpi_i = \beta_i \theta$,) then the abundance distribution under trade-offs is type independent.