

SI Appendix: 1) Large Population Limits for a Neutral Metacommunity and 2) Gibbs Sampling for the UNTB-HDP

Keith Harris¹, Todd L Parsons², Umer Ijaz³, Leo Lahti⁴, Ian Holmes⁵ and Christopher Quince^{6,*}

¹ School of Mathematics and Statistics, University of Sheffield, Sheffield, UK

² Laboratoire de Probabilités et Modèles Aléatoires, CNRS UMR 7599, UPMC Univ Paris 06, Paris, France.

³ Infrastructure and Environment Research Division, School of Engineering, University of Glasgow, Glasgow, G12 8LT, UK.

⁴ Department of Veterinary Biosciences, University of Helsinki, Helsinki, Finland & Laboratory of Microbiology, Wageningen University, Wageningen, Netherlands.

⁵ Department of Bioengineering, University of California, Berkeley, California, USA.

⁶ Warwick Medical School, University of Warwick, Coventry, CV4 7AL, UK.

*Corresponding author

1 Large Population Limits for a Neutral Metacommunity

1.1 Summary and Outline

Given the length and technical nature of this supplement, we will begin with a summary that outlines the results herein. Our intent is to formulate a class of models that generalize Hubbell’s formulation of the Unified Neutral Theory of Biodiversity and Biogeography (UNTB) and a number of variants that have appeared in the community ecology literature, whilst retaining the essential feature of neutrality. Our inspiration in this are Cannings’ models [1], which have become the standard in theoretical population genetics. We discuss coalescent theory and these models in detail below, but in brief, a Cannings’ model allows any reproduction law with discrete generations that keeps the total population size fixed, provided that relabeling the parents leaves the distribution of offspring unchanged. More generally, we could consider models replacing fixed population sizes with density dependent population dynamics, as in [2], [3, 4] and [5], but this would have further lengthened and complicated this supplement.

We formulate a mainland-island Cannings’ model, in which the mainland has size $N_0 = N$ and the islands have size N_i that grow with N , but are approximately equal. We allow migration between any pair of island and mainland, and further allows mutations to give rise to new types on both island and mainland. After collecting a few results regarding the reproduction law for a Cannings’ model, we show in Section 1.4, provided that:

- the islands are asymptotically smaller than the mainland (in both census and effective population size; see the discussion below),
- migration between demes is rare (we assume that the probability that a migrant arrives in a local community is inversely proportional to the size of that community), and
- the probability of multiple mergers is asymptotically smaller in N than the rate of pairwise coalescence,

then Proposition 1 shows that if we rescale time proportionally to the effective population size of the islands (*i.e.*, we measure time so that one time step corresponds to N_e generations) for large values of N , the population dynamics on the islands converge to the dynamics of Moran’s infinitely many alleles model, with the migration rate from the mainland taking the place of the mutation rate in the population genetic model, and such that the type of all new mutants/migrants is drawn from the initial type distribution for the mainland (*i.e.*, the probability of migration between islands or novel mutations appearing on an island becomes vanishingly small as N grows large, and can be completely ignored in the limit), and moreover, the composition of the mainland remains constant on this timescale - the dynamics are sufficiently slow that one cannot see changes when time is scaled according to the effective population size of the islands. Moreover, this limit is independent of the specific reproduction law for the islands, provided it satisfies Cannings’ conditions - indeed, we don’t even need to assume the same law between islands. As a consequence of the

identification of the islands’ dynamics as a variation of the infinitely many alleles model, we can use previous results from theoretical population genetics to conclude that the stationary distribution for the islands is a Dirichlet Process, and that the composition of a sample is distributed according to Ewens’ sampling formula.

In Section 1.5 we turn our attention to the mainland. We first observe that for large values of time, the species distribution on the islands converge onto stationary processes governed by the Dirichlet process above. We can then apply this with results from [6] to obtain Proposition 3, which tells us that we need to rescale time according to the effective population size of the mainland (again, so that one time step corresponds to N_e generations, but now N_e for the mainland, which is substantially larger). On this slow scale, the islands will essentially instantaneously arrive at their stationary state (an instant in this “slow” time scale is in fact an extremely long time in the natural “intermediate” time scale for the islands), whilst now the population on mainland follows the “real” infinitely many alleles model (with the actual mutation rate), and again, migrations from an island to the mainland become vanishingly rare as N becomes large, and, as before, the stationary distribution is again a Dirichlet process, where each newly appearing genotype is assigned a label chosen uniformly at random from $[0, 1]$ (thus the probability of two distinct mutations giving rise to the same type is 0). In particular, the islands have the Hierarchical Dirichlet Process for their stationary distribution: they are Dirichlet Processes in which the types are drawn from the underlying Dirichlet Process that describes the mainland.

1.2 A Mainland-Island “Cannings’ Model”

We begin by formulating a broad class of haploid models that includes Hubbell’s Unified Neutral Theory of Biodiversity and Biogeography (UNTB) [7]. Our inspiration are Cannings’ population genetic models [1], which use exchangeability as a general mathematical formulation of neutrality: random variables ν_1, \dots, ν_N are *exchangeable* if the random vectors $(\nu_{\pi(1)}, \dots, \nu_{\pi(N)})$ are equal in distribution for all permutations π of $\{1, \dots, N\}$. Informally, the labels $1, \dots, N$ are arbitrary, and can be changed without essentially changing the process. In a Cannings’ model, one assumes a fixed population of size N and discrete generations; $\nu_i(n)$ is the number of offspring in the $n + 1^{\text{st}}$ generation of the i^{th} individual of the n^{th} generation. (ν_1, \dots, ν_N) is assumed to be exchangeable and must satisfy

$$\sum_{i=1}^N \nu_i = N.$$

Under suitable conditions on the higher moments (*n.b.*, as a consequence of exchangeability, we must have $\mathbb{E}[\nu_i] = 1$ for all i), one can show [8] that as $N \rightarrow \infty$ the frequency of types (here, the type of an individual is inherited from its ancestor in the initial population) and the genealogical process converge to the Wright-Fisher diffusion and Kingman’s coalescent, respectively (relaxing the moment conditions leads to a Λ -coalescent limit for the genealogical process). In particular, if $X_i^{(N)}(n)$ is the number of descendants alive in the n^{th} generation of the i^{th} ancestral individual in the 0^{th} generation, and c_{N_i} is the coalescence probability, *i.e.*, the probability two individuals sampled without replacement from deme i have the same parent,

$$c_N := \frac{\mathbb{E}[(\nu_1)_2]}{N - 1},$$

where

$$(x)_k := x(x - 1) \cdots (x - k + 1)$$

is the *falling factorial* or *Pochhammer symbol*. Then, [8] shows that

$$\lim_{N \rightarrow \infty} \frac{\mathbb{E}[(\nu_1)_3]}{N \mathbb{E}[(\nu_1)_2]} = 0$$

is a necessary and sufficient condition for $X_i^{(N)}(\lfloor c_N^{-1}t \rfloor)$ to converge weakly¹ as $N \rightarrow \infty$ to a Wright-Fisher diffusion, *i.e.*, to a diffusion process with probability density

$$p(\mathbf{y}, t | \mathbf{x}) := \mathbb{P} \{ \mathbf{X}(t) \in \mathbf{y} + d\mathbf{y} | \mathbf{X}(0) = \mathbf{x} \}$$

satisfying the Kolmogorov backward equation

$$\frac{\partial p}{\partial t} = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N x_i (\delta_{ij} - x_j) \frac{\partial^2 p}{\partial x_i \partial x_j}.$$

The quantity c_N^{-1} has been referred to as the coalescent effective population size, and can be shown to generalize previously defined notions of an effective population size [9].

Here, we take our cues from the discussion of infinite-alleles models in [6], which we will closely follow, in formulating a ‘‘Cannings’ UNTB’’ with migration and mutation. As in previous models, we will assume a mainland, which supports a population of size $N_0 = N$, together with a collection of islands labelled $i = 1, \dots, M$ which support populations of size N_i . We will assume that the islands are all approximately the same size, and substantially smaller than the mainland; for Section 1.4, we will require $N_i \ll N_0^2$, whereas we will need to impose sharper estimates of the relative sizes in Section 1.5. In what follows, we will refer to the mainland and each of the islands as having N_0 or N_i niches respectively, we will use the term deme when we are referring to a local community that can be either an island or the mainland, and will refer to *e.g.*, the individual in the j^{th} niche in the i^{th} deme.

We will assume discrete generations, and that at each time step the current residents reproduce and are replaced by their offspring. The j^{th} individual has $\nu_{ij}^{(N)}$ offspring so that

$$\sum_{j=1}^{N_i} \nu_{ij} = N_i,$$

and model neutrality by assuming that each random vector $(\nu_{i1}^{(N)}(n), \dots, \nu_{iN_i}^{(N)}(n))$ is exchangeable. We further assume that $(\nu_{i1}^{(N)}(n), \dots, \nu_{iN_i}^{(N)}(n))$ is independent of $(\nu_{j1}^{(N)}(m), \dots, \nu_{jN_j}^{(N)}(m))$ unless $i = j$ and all $m = n$. Following [8], we define

$$c_{N_i} := \frac{\mathbb{E}[(\nu_{i1})_2]}{N_i - 1},$$

for $i = 0, \dots, M$, and assume the analogue of Möhle’s condition:

$$\lim_{N_i \rightarrow \infty} \frac{\mathbb{E}[(\nu_{i1})_3]}{N_i \mathbb{E}[(\nu_{i1})_2]} = 0, \tag{1}$$

which has the following consequence [10]:

¹A family of random variables $\{X^{(N)}\}$ taking values in a space S is said to *converge weakly* to X if

$$\lim_{N \rightarrow \infty} \mathbb{E}[f(X^{(N)})] = \mathbb{E}[f(X)]$$

for all $f \in C(S)$; the values $\mathbb{E}[f(X)]$ completely characterize the distribution of X . Weak convergence is denoted by

$$X^{(N)} \Rightarrow X.$$

²We will write $a_N = o(b_N)$, or $a_N \ll b_N$, if

$$\lim_{N \rightarrow \infty} \frac{a_N}{b_N} = 0,$$

and use $a_N \asymp_N b_N$ to indicate that

$$\lim_{N \rightarrow \infty} \frac{a_N}{b_N} = 1.$$

We will also write $a_N = \mathcal{O}(b_N)$ if there exists a constant C such that

$$a_N \leq C b_N,$$

for all N .

Lemma 1. *Assume (1). Then,*

$$\lim_{N_i \rightarrow \infty} c_{N_i} = 0,$$

and

$$\lim_{N_i \rightarrow \infty} \frac{\mathbb{E}[(\nu_1)_2(\nu_2)_2]}{c_{N_i}} = 0.$$

We will further assume that there exists a_N such that

$$\lim_{N \rightarrow \infty} \frac{c_{N_i}}{a_N} = \begin{cases} \gamma_i & \text{if } i > 0, \text{ and} \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

which formalises the notion that the populations on the islands are all of the same order of magnitude (their effective population sizes are asymptotically proportional $c_{N_i} \sim \gamma_i a_N$) and asymptotically smaller than the mainland ($a_N \ll c_N$).

We will further assume that each individual has a type, which is a label in $[0, 1]$, which we think of as a probability space with the uniform (Lebesgue) measure λ . The labels are more of a mathematical convenience for tracking ancestries, and have no effect on fitness, so we could equally well take labels in any compact Polish space \mathfrak{X} that is equipped with a probability measure $\gamma(dx)$. We write $X_{ij}(n) \in [0, 1]$ for the type of the individual in the j^{th} niche of the i^{th} deme in generation n – the labels are inherited from the parent, except when an individual is subject to mutation at birth. We discuss the processes of reproduction and mutation below. The state of the i^{th} deme in the n^{th} generation is conveniently represented by an atomic probability measure on $[0, 1]$,

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

where $\delta_{X_{ij}(t)}$ is the Dirac point mass at $X_{ij}(t)$, and the superscript (N) emphasizes the dependence on the “system size” N , *i.e.*, for any subset $A \subseteq [0, 1]$, $G_i^{(N)}(n)(A)$ is the number of individuals in the i^{th} deme with a type in the set A . We write $\mathbf{G}^{(N)}(n) = G_0^{(N)}(n) \otimes \cdots \otimes G_M^{(N)}(n)$ for the product measure,

$$\mathbf{G}^{(N)}(n)(A) = G_0^{(N)}(n)(A) \cdots G_M^{(N)}(n)(A).$$

Given a measure μ and a continuous function f on $[0, 1]$, we will use the shorthand

$$\langle f, \mu \rangle := \int f(x) \mu(dx)$$

for the integral. More generally, if $f \in C([0, 1]^{M+1})$, then

$$\langle f, \mu_0 \otimes \cdots \otimes \mu_M \rangle := \int f(x_0, \dots, x_M) \mu_0(dx_0) \cdots \mu_M(dx_M).$$

By definition, we have

$$\langle f, G_i^{(N)}(n) \rangle = \frac{1}{N_i} \sum_{j=1}^{N_i} f(X_{ij}(n)).$$

We model migration by assuming that with probability $c_{N_i} \frac{\varpi_i}{2}$ (the factor of $\frac{1}{2}$ is to maintain consistency of notation with the cited population genetics literature), a given individual in the $n+1^{\text{st}}$ generation is replaced by the migrant offspring of a parent chosen uniformly at random from the entire metapopulation, *i.e.*, we assume a parent of type $X_{pq}(n)$, where the p and q are drawn uniformly from $\{0, \dots, M\}$ and $\{1, \dots, N_p\}$, respectively. Thus, the average number of migrants to a given island is asymptotically independent of N ; this is a weak migration limit. Equivalently, the parent is drawn from the *metapopulation* measure,

$$G^{(N)}(n) := \frac{1}{\sum_{k=0}^M N_k} \sum_{i=0}^M N_i G_i^{(N)}(n). \quad (3)$$

Finally, we allow for the possibility that individuals mutate after birth; we assume that there is a probability measure $P^{(N)}$ such that the offspring of a parent with type $x \in [0, 1]$ mutates to a type in $A \subseteq [0, 1]$ with probability $P^{(N)}(x, A)$. Define an operator $Q^{(N)}$ on $C([0, 1])$ by

$$(Q^{(N)}f)(x) = \int_0^1 f(y)P^{(N)}(x, dy).$$

Then, for all $f \in C([0, 1])$, we define

$$\begin{aligned} (Q_i^{(N)}f)(x) &:= \mathbb{E} \left[f(X_{ij}) \middle| \mathbf{G}^{(N)}(n), \text{parent of type } x \right] \\ &= (1 - c_{N_i} \frac{\varpi_i}{2})(Q^{(N)}f)(x) + c_{N_i} \frac{\varpi_i}{2} \int (Q^{(N)}f)(y)G^{(N)}(n)(dy) \end{aligned} \quad (4)$$

and

$$(B_i^{(N)}f)(x) := \frac{\varpi_i}{2} \left(\int f(y)G^{(N)}(n)(dy) - f(x) \right). \quad (5)$$

While it may at first appear unusual, this notation will greatly simplify subsequent calculations.

We will assume mutation is weak:

$$B := \lim_{N \rightarrow \infty} c_N^{-1}(I - Q^{(N)})$$

exists and B is a bounded operator. Thus, for any set $A \subseteq [0, 1]$, the probability that the offspring has a type in A approaches 1 as $N \rightarrow \infty$, if the parent has a type in A , and approaches 0 otherwise. Here, c_N is the coalescent effective population size for the mainland, and we are making the standard assumption that mutation rates scale like the reciprocal of the effective population size. For the sake of clarity in the arguments that follow, we emphasize that our assumptions entail that

$$Q_i^{(N)} = I + c_{N_i} B_i^{(N)} + c_N B + o(c_N).$$

One can consider many forms for the operator B ; the operator

$$(B^{(L)}f) \left(\frac{i}{L} \right) = \frac{\theta}{L-1} \sum_{j=1}^L \left(f \left(\frac{j}{L} \right) - f \left(\frac{i}{L} \right) \right)$$

corresponds to the classical population genetic models, in which the number of possible types is discrete and finite (here, there are L) and mutation is symmetric (*i.e.*, the offspring of an individual have the same type as their parent with probability $1 - \frac{\theta}{N}$, and mutate to any other type with probability $\frac{\theta}{N(L-1)}$). Since the labels are arbitrary, they can be assumed to be chosen from the set $\{\frac{1}{L}, \frac{2}{L}, \dots, 1\}$. Now, as $L \rightarrow \infty$, $B^{(L)}$ converges to the operator

$$(Bf)(x) = \frac{\theta}{2} \int_0^1 f(y) dy - f(x) = \theta(\langle f, \lambda \rangle - f(x)),$$

which corresponds to the infinitely many alleles model; the probability that two mutations give rise to the same type is 0. We will henceforth assume B is of this form.

Remark 1. Although we have formulated the community dynamics in discrete time, we could equally well consider a continuous time Markov process $\tilde{G}_i^{(N)}(t)$ in which disturbances happen at some rate D ; in the latter case, we consider the embedded Markov chain: if disturbances happen at random times τ_1, τ_2, \dots , then the embedded chain is the process $G_i^{(N)}(n) := \tilde{G}_i^{(N)}(\tau_n)$. The limiting (continuous time) process as $N \rightarrow \infty$ is the same for both $G^{(N)}$ and $\tilde{G}_i^{(N)}$.

In the next section, we will consider the limiting behaviour as first N and then L are taken to infinity. We will see that under moment assumptions corresponding to those in [8], all of these models converge to the same limiting process. First, however, we illustrate how Hubbell's original UNTB is an example of our class of models.

Example 1 (Hubbell's UNTB). In Hubbell's original model [7], only a single individual is replaced in each deme at each time step. We then have ν_{0i} takes values in $\{0, 1\}$, with

$$\mathbb{P}\{\nu_{0i} = 1\} = m.$$

We then have that the remaining offspring numbers are either

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

(the vector with i^{th} entry 0, and all others 1), if $\nu_{0i} = 1$, and is

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

(the vector with i^{th} entry 0 and j^{th} entry 2 for some $i \neq j$), if $\nu_{0i} = 0$, with conditional probabilities equal to $\frac{1}{N_i}$ and $\frac{1}{N_i(N_i-1)}$, respectively (and thus the ν_{ij} are exchangeable, given ν_{i0}).

For this model, we have

$$c_{N_i} = \frac{2}{N_i(N_i - 1)},$$

whereas by definition, $(\nu_{i1})_3 = 0$, so (1) holds.

In Hubbell's model, immigrants are always from the mainland, which is assumed to have a fixed, stationary distribution (usually taken so that samples from the mainland are distributed according to Ewens' sampling formula [11]), 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 $N \rightarrow \infty$) as a consequence of the relative size of the mainland and the islands.

Example 2 ("Wright-Fisher" UNTB). We can regard Hubbell's UNTB as a community analogue of the discrete Moran model. We could similarly define a community analogue to the Wright-Fisher model by assuming that the vector $(\nu_{i1}, \dots, \nu_{iN_i})$ follows a multinomial distribution with parameters N_i and $(\frac{1}{N_i}, \dots, \frac{1}{N_i})$, *i.e.*, for each i :

$$\mathbb{P}\{(\nu_{i1}, \dots, \nu_{iN_i}) = (k_1, \dots, k_{N_i})\} = \frac{N_i!}{k_1! \cdots k_{N_i}!} \left(\frac{1}{N_i}\right)^{k_1} \cdots \left(\frac{1}{N_i}\right)^{k_{N_i}}.$$

Here, $c_{N_i} = \frac{1}{N_i}$, whereas $\mathbb{E}[(\nu_{i1})_3] = \frac{1}{N_i}^2$.

Example 3. We briefly note that it is possible to have $c_{N_i} \equiv 0$, by assuming that $(\nu_{i1}, \dots, \nu_{iN_i}) = (1, \dots, 1)$ with probability 1 (a trivial case that we will ignore), whereas it need not be the case that

$$\lim_{N_i \rightarrow \infty} c_{N_i} = 0$$

if (1) is violated: if we assume that with probability $\frac{1}{N_i}$, $\nu_{ij} = N_i$ and $\nu_{ik} = 0$ for all $k \neq j$, then $c_{N_i} \equiv 1$.

1.3 Preliminaries Considering Exchangeable Variables

It is well known [12] that

$$\frac{\binom{N_i}{j}}{\binom{N_i}{k}} \mathbb{E}[(\nu_{i1})_{k_1} \cdots (\nu_{ij})_{k_j}],$$

where $j, k_1, \dots, k_j \in \mathbb{N}$ and $k := k_1 + \dots + k_j$, is the probability that k individuals, sampled uniformly at random without replacement from the i^{th} deme have exactly j parents in the previous generation, *n.b.*, exchangeability implies that

$$\frac{\binom{N_i}{j}}{\binom{N_i}{k}} \mathbb{E} [(\nu_{i1})_{k_1} \cdots (\nu_{ij})_{k_j}] = \frac{\binom{N_i}{j}}{\binom{N_i}{k}} \mathbb{E} [(\nu_{i\pi(1)})_{k_1} \cdots (\nu_{i\pi(j)})_{k_j}]$$

for any permutation π of $\{1, \dots, N_i\}$, so that these probabilities only depend on j, k , and the unordered list of values k_1, \dots, k_j . In [8], we find the following monotonicity result for these probabilities:

Lemma 2. *Let $j \geq l$, $k_1 \geq m_1, \dots, k_l \geq m_l$, and $m := m_1 + \dots + m_l$. Then,*

$$\frac{\binom{N_i}{j}}{\binom{N_i}{k}} \mathbb{E} [(\nu_{i1})_{k_1} \cdots (\nu_{ij})_{k_j}] \leq \frac{\binom{N_i}{l}}{\binom{N_i}{m}} \mathbb{E} [(\nu_{i1})_{m_1} \cdots (\nu_{il})_{m_l}].$$

Remark 2. In particular, in conjunction with Lemma 2, we have

$$\frac{\binom{N_i}{j-1}}{\binom{N_i}{j}} \mathbb{E} [(\nu_{i1})_2 \nu_{i2} \cdots \nu_{ij-1}] \leq c_{N_i}, \quad (6)$$

(and, by exchangeability, whenever at least one $k_i \geq 2$) and

$$\frac{\binom{N_i}{j}}{\binom{N_i}{k}} \mathbb{E} [(\nu_{i1})_{k_1} \cdots (\nu_{ij})_{k_j}] = o(c_{N_i}) \quad (7)$$

whenever $k_q, k_r \geq 2$ for at least two distinct indices q, r or $k_q \geq 3$ for some index q .

Remark 3. In particular, in (7), $\frac{\binom{N_i}{j}}{\binom{N_i}{k}} \mathbb{E} [(\nu_{i1})_{k_1} \cdots (\nu_{ij})_{k_j}]$ is always smaller than one of

$$\frac{\binom{N_i}{1}}{\binom{N_i}{3}} \mathbb{E} [(\nu_{i1})_3]$$

or

$$\frac{\binom{N_i}{2}}{\binom{N_i}{4}} \mathbb{E} [(\nu_{i1})_2 (\nu_{i2})_2].$$

In what follows, all terms $o(c_{N_i})$ will be of order at most equal to one of these two quantities (which are the probability of three individuals sampled at random having the same parent in the previous generation, or a sample of four individuals consisting of two pairs of descendants of two distinct parents, respectively) or will be of order less than or equal to $\frac{c_{N_i}}{N_i}$. This will be very important when we consider the long timescale.

We will have use of some general relations between exchangeable random variables in the sequel:

Lemma 3. *For all $j > 1$*

$$\mathbb{E} [\nu_{i1} \cdots \nu_{ij-1}] - \mathbb{E} [\nu_{i1} \cdots \nu_{ij}] = (j-1) \frac{\binom{N_i}{j-1}}{\binom{N_i}{j}} \mathbb{E} [(\nu_{i1})_2 \cdots \nu_{ij-1}].$$

Proof. We begin by observing that

$$\begin{aligned}
N_i \mathbb{E} [\nu_{i1} \cdots \nu_{ij-1}] &= \mathbb{E} [N_i \nu_{i1} \cdots \nu_{ij-1}] \\
&= \mathbb{E} [(\nu_{i1} + \cdots + \nu_{iN_i}) \nu_{i1} \cdots \nu_{ij-1}] \\
&= \mathbb{E} \left[\sum_{k=1}^{N_i} \nu_{i1} \cdots \nu_{ij-1} \nu_{ik} \right] \\
&= \mathbb{E} \left[\sum_{k=1}^{j-1} \nu_{i1} \cdots \nu_{ij-1} \nu_{ik} + \sum_{k=j}^{N_i} \nu_{i1} \cdots \nu_{ij-1} \nu_{ik} \right] \\
&= \sum_{k=1}^{j-1} \mathbb{E} [\nu_{i1} \cdots \nu_{ij-1} \nu_{ik}] + \sum_{k=j}^{N_i} \mathbb{E} [\nu_{i1} \cdots \nu_{ij-1} \nu_{ik}] \\
&= \sum_{k=1}^{j-1} \mathbb{E} \left[\nu_{ik}^2 \prod_{\substack{l=1 \\ l \neq k}}^{j-1} \nu_{il} \right] + \sum_{k=j}^{N_i} \mathbb{E} [\nu_{i1} \cdots \nu_{ij-1} \nu_{ik}]
\end{aligned}$$

and thus, exploiting the exchangeability of the ν_{ij} ,

$$= (j-1) \mathbb{E} [\nu_{i1}^2 \cdots \nu_{ij-1}] + (N_i - j + 1) \mathbb{E} [\nu_{i1} \cdots \nu_{ij}].$$

On the other hand,

$$N_i \mathbb{E} [\nu_{i1} \cdots \nu_{ij-1}] = (j-1) \mathbb{E} [\nu_{i1} \cdots \nu_{ij-1}] + (N_i - j + 1) \mathbb{E} [\nu_{i1} \cdots \nu_{ij-1}].$$

Equating the two sides and subtracting, we get

$$(N_i - j + 1) (\mathbb{E} [\nu_{i1} \cdots \nu_{ij-1}] - \mathbb{E} [\nu_{i1} \cdots \nu_{ij}]) = (j-1) (\mathbb{E} [\nu_{i1}^2 \cdots \nu_{ij-1}] - \mathbb{E} [\nu_{i1} \cdots \nu_{ij-1}]).$$

The result follows. □

Remark 4. In conjunction with (6), the lemma tells us that for all $j > 1$,

$$\mathbb{E} [\nu_{i1} \cdots \nu_{ij-1}] - \mathbb{E} [\nu_{i1} \cdots \nu_{ij}] = \mathcal{O}(c_{N_i}),$$

and thus,

$$\mathbb{E} [\nu_{i1} \cdots \nu_{iq}] - \mathbb{E} [\nu_{i1} \cdots \nu_{ir}] = \mathcal{O}(c_{N_i}),$$

for any $q < r$.

Next, we observe that

Lemma 4. *For all j ,*

$$\mathbb{E} [\nu_{i1} \cdots \nu_{ij}] = 1 - \binom{j}{2} \frac{(N_i)_{j-1}}{(N_i)_j} \mathbb{E} [(\nu_{i1})_2 \cdots \nu_{ij-1}] - o(c_{N_i}).$$

Proof. This is a consequence of the identity

$$(N_i)_j = (\nu_{i1} + \cdots + \nu_{iN_i})_j = \sum_{j_1 + \cdots + j_{N_i} = j} \frac{j!}{j_1! \cdots j_{N_i}!} (\nu_{i1})_{j_1} \cdots (\nu_{iN_i})_{j_{N_i}},$$

where we assume $0! = 1$ for ease of notation, and we assume that most of the j_i are equal to zero. Equivalently, if we only consider non-zero values,

$$(N_i)_j = \sum_{m=1}^j \sum_{\substack{n_1, \dots, n_m \\ \text{distinct}}} \sum_{k_1 + \cdots + k_m = j} \frac{j!}{k_1! \cdots k_m!} (\nu_{in_1})_{k_1} \cdots (\nu_{in_m})_{k_m}.$$

Taking expectations on both sides, and using the exchangeability of $(\nu_{i1}, \dots, \nu_{iN_i})$, we have

$$(N_i)_j = \sum_{m=1}^j \sum_{\substack{n_1, \dots, n_m \\ \text{distinct}}} \sum_{k_1 + \dots + k_m = k} \frac{j!}{k_1! \dots k_m!} \mathbb{E}[(\nu_{i1})_{k_1} \dots (\nu_{im})_{k_m}].$$

Now, observe that the expected value of the summand is independent of the choice of the values n_1, \dots, n_m , that can be chosen in $\binom{N}{m}$ ways. Moreover, the expectation $\mathbb{E}[(\nu_{i1})_{k_1} \dots (\nu_{im})_{k_m}]$ remains unchanged under permutations, and thus are all equal to

$$\mathbb{E}[(\nu_{i1})_{\tilde{k}_1} \dots (\nu_{im})_{\tilde{k}_m}],$$

where $\tilde{k}_1 \geq \tilde{k}_2 \geq \dots \geq \tilde{k}_m$ are the values k_1, \dots, k_m listed in decreasing order. If we let a_p be the number of indices q such that $k_q = p$,

$$a_p = \#\{q : k_q = p\},$$

then

$$\begin{aligned} \sum_{m=1}^j \sum_{\substack{n_1, \dots, n_m \\ \text{distinct}}} \sum_{k_1 + \dots + k_m = k} \frac{j!}{k_1! \dots k_m!} \mathbb{E}[(\nu_{i1})_{k_1} \dots (\nu_{im})_{k_m}] \\ = \sum_{m=1}^j \sum_{\substack{\tilde{k}_1 + \dots + \tilde{k}_m = k \\ \tilde{k}_1 \geq \tilde{k}_2 \geq \dots \geq \tilde{k}_m}} \frac{j!}{\tilde{k}_1! \dots \tilde{k}_m!} \frac{m!}{a_1! \dots a_j!} \binom{N}{m} \mathbb{E}[(\nu_{i1})_{\tilde{k}_1} \dots (\nu_{im})_{\tilde{k}_m}], \end{aligned}$$

so that, simplifying and dividing through by $(N)_j$, we have

$$\begin{aligned} 1 &= \sum_{m=1}^j \sum_{\substack{\tilde{k}_1 + \dots + \tilde{k}_m = k \\ \tilde{k}_1 \geq \tilde{k}_2 \geq \dots \geq \tilde{k}_m}} \frac{j!}{\tilde{k}_1! \dots \tilde{k}_m!} \frac{1}{a_1! \dots a_j!} \frac{(N)_m}{(N)_j} \mathbb{E}[(\nu_{i1})_{\tilde{k}_1} \dots (\nu_{im})_{\tilde{k}_m}] \\ &= \mathbb{E}[\nu_{i1} \dots \nu_{ij}] + \binom{j}{2} \frac{(N_i)_{j-1}}{(N_i)_j} \mathbb{E}[(\nu_{i1})_2 \dots \nu_{ij-1}] + o(c_{N_i}), \end{aligned}$$

where, using (7), we have truncated after the two highest order terms in the sum. \square

We conclude this section with a final observation,

Lemma 5. *Let $j > 1$. Then,*

$$\frac{(N_i)_j}{(N_i)_{j+1}} \mathbb{E}[(\nu_{i1})_2 \nu_{i2} \dots \nu_{ij}] = \frac{(N_i)_{j-1}}{(N_i)_j} \mathbb{E}[(\nu_{i1})_2 \nu_{i2} \dots \nu_{ij-1}] + o(c_{N_i}).$$

Proof. Again exploiting exchangeability, we see that

$$\begin{aligned} (N_i - j + 1) \mathbb{E}[(\nu_{i1})_2 \nu_{i2} \dots \nu_{ij}] \\ &= \mathbb{E}[(\nu_{i1})_2 \nu_{i2} \dots \nu_{ij-1} \nu_{ij}] + \mathbb{E}[(\nu_{i1})_2 \nu_{i2} \dots \nu_{ij-1} \nu_{ij+1}] + \dots + \mathbb{E}[(\nu_{i1})_2 \nu_{i2} \dots \nu_{ij-1} \nu_{iN_i}] \\ &= \mathbb{E}[(\nu_{i1})_2 \nu_{i2} \dots \nu_{ij-1} (\nu_{ij} + \dots + \nu_{iN_i})] \\ &= \mathbb{E}[(\nu_{i1})_2 \nu_{i2} \dots \nu_{ij-1} (N_i - \nu_{i1} - \dots - \nu_{ij-1})] \\ &= \mathbb{E}[(\nu_{i1})_2 \nu_{i2} \dots \nu_{ij-1} (N_i - j + 2 - (\nu_{i1} - 2) - (\nu_{i2} - 1) - \dots - (\nu_{ij-1} - 1))] \\ &= (N_i - j + 2) \mathbb{E}[(\nu_{i1})_2 \nu_{i2} \dots \nu_{ij-1}] - \mathbb{E}[(\nu_{i1})_3 \nu_{i2} \dots \nu_{ij-1}] \\ &\quad - \mathbb{E}[(\nu_{i1})_2 (\nu_{i2})_2 \nu_{i3} \dots \nu_{ij-1}] - \dots - \mathbb{E}[(\nu_{i1})_2 \nu_{i2} \dots (\nu_{ij-1})_2]. \end{aligned}$$

In particular, dividing both sides by $(N_i - j + 1)(N_i - j + 2)$, we have

$$\begin{aligned} \frac{\binom{N_i}{j}}{\binom{N_i}{j+1}} \mathbb{E}[(\nu_{i1})_2 \nu_{i2} \cdots \nu_{ij}] &= \frac{\binom{N_i}{j-1}}{\binom{N_i}{j}} \mathbb{E}[(\nu_{i1})_2 \nu_{i2} \cdots \nu_{ij-1}] - \frac{\binom{N_i}{j-1}}{\binom{N_i}{j+1}} \mathbb{E}[(\nu_{i1})_3 \nu_{i2} \cdots \nu_{ij-1}] \\ &\quad - \frac{\binom{N_i}{j-1}}{\binom{N_i}{j+1}} \mathbb{E}[(\nu_{i1})_2 (\nu_{i2})_2 \nu_{i3} \cdots \nu_{ij-1}] - \cdots - \frac{\binom{N_i}{j-1}}{\binom{N_i}{j+1}} \mathbb{E}[(\nu_{i1})_2 \nu_{i2} \cdots (\nu_{ij-1})_2] \end{aligned}$$

and the result again follows by (7). \square

Remark 5. Iterating the previous lemma, we see that

$$\frac{\binom{N_i}{j}}{\binom{N_i}{j+1}} \mathbb{E}[(\nu_{i1})_2 \cdots \nu_{ij}] = \cdots = \frac{\mathbb{E}[(\nu_{i1})_2]}{N_i - 1} + o(c_{N_i}) = c_{N_i} + o(c_{N_i}).$$

1.4 Convergence to a Limit

We will be interested in weak limits of the random measures $\mathbf{G}^{(N)}(n)$ in two time-scales determined by N , a “slow-time” process, $\mathbf{G}^{(N)}(\lfloor c_N^{-1}t \rfloor)$, and an “intermediate-time” process $\mathbf{G}^{(N)}(\lfloor a_N^{-1}t \rfloor)$, where $t > 0$ is a continuous time variable, and we will consider the limits as $N \rightarrow \infty$.

Our principal tool in doing this is the generator of $\mathbf{G}^{(N)}(n)$, an operator on $C(\mathcal{P}([0, 1])^{M+1})$ defined by

$$(\mathcal{G}_N F)(\boldsymbol{\mu}) = \mathbb{E} \left[F(\mathbf{G}^{(N)}(n+1)) \middle| \mathbf{G}^{(N)}(n) = \boldsymbol{\mu} \right] - F(\boldsymbol{\mu}).$$

Knowing $(\mathcal{G}_N F)(\boldsymbol{\mu})$ for all $F \in C(\mathcal{P}([0, 1])^{M+1})$ and all $\boldsymbol{\mu} \in \mathcal{P}([0, 1])^{M+1}$ completely characterizes the transition probabilities of $\mathbf{G}^{(N)}$, and thus, together with the initial value $\mathbf{G}^{(N)}(0)$, allow us to characterize the process (although not necessarily the limit, see *e.g.*, [6]).

Our limiting processes are continuous, rather than discrete time random variables, but also have associated generators; in general, if $\mathbf{H}(t)$ is a continuous time process taking values in $\mathcal{P}([0, 1])^{M+1}$ and $F \in C(\mathcal{P}([0, 1])^{M+1})$, then $\mathbf{H}(t)$ has generator \mathcal{H} :

$$(\mathcal{H}F)(\boldsymbol{\mu}) = \lim_{h \rightarrow 0^+} \frac{\mathbb{E}[F(\mathbf{H}(t+h)) | \mathbf{H}(t) = \boldsymbol{\mu}] - F(\boldsymbol{\mu})}{h},$$

with domain $\mathcal{D}(\mathcal{H})$, consisting of all functions F for which the limit exists.

The notion of a generator simultaneously generalizes the transition matrix, master equation, and diffusion equations of classical probability. The typical proof of convergence proceeds by first showing that a limit exists, then characterizing the limit by first determining the limit of the generators, and finally showing that given the initial conditions (via a distribution from which they are drawn), there is a unique process with that generator (*e.g.*, [6] is a standard reference).

Remark 6. Note that $(\mathcal{H}F)(\boldsymbol{\mu})$ is the right-hand derivative of $\mathbb{E}[F(\mathbf{H}(t+h)) | \mathbf{H}(t) = \boldsymbol{\mu}]$ at $t = 0$. In particular, if the generator vanishes, then $\mathbb{E}[F(\mathbf{H}(t)) | \mathbf{H}(0) = \boldsymbol{\mu}] = F(\boldsymbol{\mu})$ for all $t > 0$, and all F , and the process $\mathbf{H}(t) \equiv \boldsymbol{\mu}$ is constant. This will be important when we come to consider the limit on the intermediate time scale.

We will make use of the fact that the set of functions

$$\mathcal{C} := \left\{ F(\boldsymbol{\mu}) = \prod_{i=0}^M \prod_{k=1}^{K_i} \langle f_{ik}, \mu_i \rangle \middle| K_i \in \mathbb{N}_0, f_{ik} \in C([0, 1]) \right\}$$

is separating, and convergence determining [6], so that for the purpose of characterizing our process and its limits, we need only compute the value the generator takes on functions $F \in \mathcal{C}$ and its limits.

We will evaluate the generator on this class of functions, but we first begin with a pair of lemmas. We will use \coprod to indicate the disjoint union of sets and, for all integers $M > 0$, we use the shorthand $[M] = \{1, \dots, M\}$.

Lemma 6. Let $\mu_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \delta_{x_{ij}}$ for $x_{ij} \in [0, 1]$. Then,

$$\begin{aligned} \prod_{k=1}^{K_i} \langle f_{ik}, \mu_i \rangle &= \frac{1}{N_i^{K_i}} \sum_{m=1}^{K_i} \sum_{\substack{j_1, \dots, j_m \\ \text{distinct}}} \sum_{A_1 \amalg \dots \amalg A_m = [K_i]} \prod_{q=1}^m \prod_{r \in A_q} f_{ir}(x_{ij_q}) \\ &= \frac{1}{N_i^{K_i}} \sum_{\substack{j_1, \dots, j_{K_i} \\ \text{distinct}}} \prod_{k=1}^{K_i} f_{ik}(x_{ij_k}) + \mathcal{O}(N^{-1}), \end{aligned}$$

where the sum is over all partitions of $[K_i]$ into m disjoint sets.

Proof. The first statement is simply a matter of collecting terms according to the number of distinct values j_k :

$$\begin{aligned} N_i^{K_i} \prod_{k=1}^{K_i} \langle f_{ik}, \mu_i \rangle &= N_i^{K_i} \prod_{k=1}^{K_i} \left(\frac{1}{N_i} \sum_{j=1}^{N_i} f_{ik}(x_{ij}) \right) = \sum_{j_1=1}^{N_i} \dots \sum_{j_{K_i}=1}^{N_i} \prod_{k=1}^{K_i} f_{ik}(x_{ij_k}) \\ &= \sum_{m=1}^{K_i} \sum_{\substack{j_1, \dots, j_m \\ \text{distinct}}} \sum_{A_1 \amalg \dots \amalg A_m = [K_i]} \prod_{q=1}^m \prod_{r \in A_q} f_{ir}(x_{ij_q}). \end{aligned}$$

Now, for the final term, we have $m = 1$, $A_1 = [K_i]$, so it takes the form:

$$\sum_{j_1=1}^{N_i} \prod_{k=1}^{K_i} f_{ik}(x_{ij_1}) = N_i \langle \prod_{k=1}^{K_i} f_{ik}, \mu_i \rangle,$$

whilst for $m = 2$, we have:

$$\begin{aligned} &= \sum_{j_1=1}^{N_i} \sum_{j_2 \neq j_1} \sum_{A_1 \amalg A_2 = [K_i]} \prod_{k \in A_1} f_{ik}(x_{ij_1}) \prod_{k \in A_2} f_{ik}(x_{ij_2}) \\ &= \sum_{A_1 \amalg A_2 = [K_i]} \sum_{j_1=1}^{N_i} \prod_{k \in A_1} f_{ik}(x_{ij_1}) \sum_{j_2=1}^{N_i} \prod_{k \in A_2} f_{ik}(x_{ij_2}) - \sum_{A_1 \amalg A_2 = [K_i]} \sum_{j_1=1}^{N_i} \prod_{k \in A_1} f_{ik}(x_{ij_1}) \prod_{k \in A_2} f_{ik}(x_{ij_1}) \\ &= N_i^2 \sum_{A_1 \amalg A_2 = [K_i]} \langle \prod_{k \in A_1} f_{ik}, \mu_i \rangle \langle \prod_{k \in A_2} f_{ik}, \mu_i \rangle - S(K_i, 2) N_i \langle \prod_{k \in A_1} f_{ik}, \mu_i \rangle, \end{aligned}$$

where $S(K_i, 2)$ is a Stirling number of the second kind [13] and gives the number of distinct partitions of K_i elements into 2 sets.

Proceeding inductively in this manner completes the proof of the lemma. \square

The previous lemma shows we will be interested in products over distinct indices j_1, \dots, j_m . In particular, we have the result of Lemma 7.

Lemma 7. For distinct values j_1, \dots, j_{K_i} in $\{1, \dots, N_i\}$,

$$\begin{aligned} \mathbb{E} \left[\prod_{k=1}^{K_i} f_{ik}(X_{ij_k}(n+1)) \middle| \{X_{ij}(n) = x_{ij}\} \right] &= \frac{\mathbb{E}[\nu_{i1} \dots \nu_{iK_i}]}{(N_i)^{K_i}} \sum_{\substack{p_1, \dots, p_{K_i} \\ \text{distinct}}} \prod_{k=1}^{K_i} (Q_i^{(N)} f_{ik})(x_{ip_k}) \\ &+ \frac{\mathbb{E}[(\nu_{i1})_2 \nu_{i2} \dots \nu_{iK_i-1}]}{(N_i)^{K_i}} \sum_{q < r} \sum_{\substack{p_1, \dots, p_{K_i} \\ p_q = p_r}} \prod_{\substack{k=1 \\ k \neq q, r}}^{K_i} (Q_i^{(N)} f_{ik})(x_{ip_k}) (Q_i^{(N)} f_{iq} Q_i^{(N)} f_{ir})(x_{ip_q}) + o(c_{N_i}). \end{aligned}$$

Proof. We begin by recalling that conditional on an individual's parent having type x , its type is independently distributed according to the probability measure $P(x, \cdot)$, *i.e.*,

$$\mathbb{E} \left[f(X_{ij}(n+1)) \middle| \mathbf{G}^{(N)}(n), \text{parent of type } x \right] = (Q_i^{(N)} f)(x).$$

We can thus, similar to the previous lemma, write:

$$\begin{aligned} \mathbb{E} \left[\prod_{k=1}^{K_i} f_{ik}(X_{ij_k}(n+1)) \middle| \{X_{ij}(n) = x_{ij}\} \right] \\ = \sum_{m=1}^{K_i} \sum_{\substack{p_1, \dots, p_m \\ \text{distinct}}} \sum_{A_1 \amalg \dots \amalg A_m = [K_i]} \frac{\mathbb{E} [(\nu_{ip_1})_{|A_1|} \cdots (\nu_{ip_m})_{|A_m|}]}{(N_i)_{K_i}} \prod_{q=1}^m \prod_{r \in A_q} (Q_i^{(N)} f_{ir})(x_{ip_q}), \end{aligned}$$

where

$$\frac{\mathbb{E} [(\nu_{ip_1})_{|A_1|} \cdots (\nu_{ip_m})_{|A_m|}]}{(N_i)_{K_i}} = \frac{\mathbb{E} [(\nu_{i1})_{|A_1|} \cdots (\nu_{im})_{|A_m|}]}{(N_i)_{K_i}}$$

is the probability that the K_i distinct individuals have m ancestors p_1, \dots, p_m (with types $x_{ip_1}, \dots, x_{ip_m}$), and that the individuals in A_q had parent p_q .

Next, we observe that since $\|Q_i^{(N)}\| \leq 1$,

$$\begin{aligned} \left| \sum_{\substack{p_1, \dots, p_m \\ \text{distinct}}} \sum_{A_1 \amalg \dots \amalg A_m = [K_i]} \frac{\mathbb{E} [(\nu_{ip_1})_{|A_1|} \cdots (\nu_{ip_m})_{|A_m|}]}{(N_i)_{K_i}} \prod_{q=1}^m \prod_{r \in A_q} (Q_i^{(N)} f_{ir})(x_{ip_q}) \right| \\ \leq \sum_{A_1 \amalg \dots \amalg A_m = [K_i]} \frac{(N_i)_m}{(N_i)_{K_i}} \mathbb{E} [(\nu_{i1})_{|A_1|} \cdots (\nu_{im})_{|A_m|}] \prod_{k=1}^{K_i} \|f_{ik}\| \end{aligned}$$

and is thus $o(c_{N_i})$ whenever $|A_q| \geq 3$ for some q or $|A_q|$ and $|A_r|$ are both ≥ 2 for distinct indices q, r by (7). The result follows. \square

We now turn to the main result of this section:

Proposition 1. *Let $\mu_i^{(N)} = \frac{1}{N_i} \sum_{j=1}^{N_i} \delta_{x_{ij}}$, for $x_{ij} \in [0, 1]$ and let $\boldsymbol{\mu}^{(N)} = \mu_1^{(N)} \otimes \cdots \otimes \mu_M^{(N)}$ converge weakly to a measure $\boldsymbol{\mu} \in \mathcal{P}([0, 1])^{M+1}$.*

Let $F(\boldsymbol{\mu}) = \prod_{i=0}^M \prod_{k=1}^{K_i} \langle f_{ik}, \mu_i \rangle \in \mathcal{C}$ and, for $i = 1, \dots, M$, let

$$\begin{aligned} (\mathcal{G}_i F)(\boldsymbol{\mu}) = \prod_{\substack{j=0 \\ j \neq i}}^M \prod_{k=1}^{K_j} \langle f_{jk}, \mu_j \rangle \left(\sum_{q=1}^{K_i} \frac{\varpi_i}{2} \langle f_{iq}, \mu_i - \mu_0 \rangle \prod_{\substack{k=1 \\ k \neq q}}^{K_i} \langle f_{ik}, \mu_i \rangle \right. \\ \left. + \frac{1}{2} \sum_{q \neq r} \prod_{\substack{k=1 \\ k \neq q, r}}^{K_i} \langle f_{ik}, \mu_i \rangle (\langle f_{iq} f_{ir}, \mu_i \rangle - \langle f_{iq}, \mu_i \rangle \langle f_{ir}, \mu_i \rangle) \right) \quad (8) \end{aligned}$$

define an operator on $C(\mathcal{P}([0, 1])^{M+1})$.

Then,

$$\lim_{N \rightarrow \infty} a_N^{-1} (\mathcal{G}_N F)(\boldsymbol{\mu}^{(N)}) = (\mathcal{G} F)(\boldsymbol{\mu}) := \sum_{i=1}^M \gamma_i (\mathcal{G}_i F)(\boldsymbol{\mu}).$$

Moreover, given $\tilde{\mu}_i \in \mathcal{P}(\mathcal{P}([0, 1]))$, there exist unique independent processes $G_i(t)$ with generators \mathcal{G}_i , such that $G_i(0)$ is distributed according to $\tilde{\mu}_i$ and such that

$$\mathbf{G}^{(N)}(\lfloor a_N^{-1}t \rfloor) \Rightarrow \mathbf{G}(t) := G_0(0) \otimes G_1(\gamma_1 t) \otimes \cdots \otimes G_M(\gamma_M t),$$

for all $t > 0$, where convergence is in the space of càdlàg functions endowed with the Skorokhod topology, $\mathbb{D}_{\mathcal{P}([0,1])^{M+1}}[0, \infty)$ (see e.g., [6]).

Remark 7. Because

$$\lim_{N \rightarrow \infty} \frac{c_N}{a_N} = 0,$$

the component of the generator acting on the mainland vanishes in the limit; if

$$\mathcal{C}_0 := \left\{ F \in \mathcal{C} \left| F(\boldsymbol{\mu}) = \prod_{k=1}^{K_0} \langle f_{ik}, \mu_0 \rangle \right. \right\},$$

then $\mathcal{G}_i F \equiv 0$ for all $F \in \mathcal{C}_0$ and thus the generator vanishes on this set. Equivalently, the process $G_0(t) \equiv \mu_0$.

Remark 8. Recall from Equation 2 that the effective population size of the i^{th} island is $c_{N_i} \sim \gamma_i a_N$; since we have rescaled time by a_N rather than the individual effective population sizes, the factors γ_i appear in the generator and in the components G_i . These reflect the fact that the different effective population sizes on the different islands result in their population dynamics having different rates (*i.e.*, different expected inter-event times), which are given by the γ_i .

Remark 9. This theorem tells us that on the intermediate time scale, the islands have essentially independent dynamics, coupled only by immigration from a mainland which remains unchanged on the intermediate timescale. The generator of the dynamics on the island is identical to that in the infinite population limit for the infinitely many alleles model, with the rescaled migration rate, $\frac{\varpi_i}{2}$ taking the place of the rescaled mutation rate θ , and the mainland density measure μ_0 taking the place of Lebesgue measure.

Proof. Applying Lemmas 6 and 7, we have

$$\begin{aligned} \mathbb{E} \left[F(\mathbf{G}^{(N)}(n+1)) \middle| \mathbf{G}^{(N)}(n) = \boldsymbol{\mu} \right] &= \mathbb{E} \left[\prod_{i=0}^M \prod_{k=1}^{K_i} \langle f_{ik}, G_i^{(N)}(n+1) \rangle \middle| \{X_{ij}(n) = x_{ij}\} \right] \\ &= \prod_{i=0}^M \mathbb{E} \left[\prod_{k=1}^{K_i} \langle f_{ik}, G_i^{(N)}(n+1) \rangle \middle| \{X_{ij}(n) = x_{ij}\} \right] \\ &= \prod_{i=0}^M \frac{1}{N_i^{K_i}} \sum_{m=1}^{K_i} \sum_{\substack{j_1, \dots, j_m \\ \text{distinct}}} \sum_{A_1 \prod \cdots \prod A_m = [K_i]} \mathbb{E} \left[\prod_{q=1}^m \prod_{r \in A_q} f_{ir}(x_{ij_q}) \middle| \{X_{ij}(n) = x_{ij}\} \right] \\ &= \prod_{i=0}^M \frac{1}{N_i^{K_i}} \sum_{m=1}^{K_i} \sum_{\substack{j_1, \dots, j_m \\ \text{distinct}}} \sum_{A_1 \prod \cdots \prod A_m = [K_i]} \left(\frac{\mathbb{E}[\nu_{i1} \cdots \nu_{im}]}{(N_i)_m} \sum_{\substack{p_1, \dots, p_m \\ \text{distinct}}} \prod_{k=1}^m (Q_i^{(N)} \prod_{l \in A_k} f_{il})(x_{ip_k}) \right. \\ &\quad \left. + \frac{\mathbb{E}[(\nu_{i1})_2 \nu_{i2} \cdots \nu_{im-1}]}{(N_i)_m} \sum_{q < r} \sum_{\substack{p_1, \dots, p_m \\ p_q = p_r}} \prod_{\substack{k=1 \\ k \neq q, r}}^m (Q_i^{(N)} \prod_{l \in A_k} f_{il})(x_{ip_k}) ((Q_i^{(N)} \prod_{l \in A_q} f_{il})(Q_i^{(N)} \prod_{l \in A_r} f_{il}))(x_{ip_q}) + o(c_{N_i}) \right). \end{aligned}$$

Now, observing that the term in brackets is independent of the values j_k , we note that j_1, \dots, j_m can be

chosen in $(N_i)_m$ ways, and we are left with a product over sums of the form:

$$\begin{aligned} & \frac{1}{N_i^{K_i}} \sum_{m=1}^{K_i} \mathbb{E}[\nu_{i1} \cdots \nu_{im}] \sum_{A_1 \amalg \cdots \amalg A_m = [K_i]} \sum_{\substack{p_1, \dots, p_m \\ \text{distinct}}} \prod_{k=1}^m (Q_i^{(N)} \prod_{l \in A_k} f_{il})(x_{ip_k}) \\ & + \frac{1}{N_i^{K_i}} \sum_{m=1}^{K_i} \mathbb{E}[(\nu_{i1})_2 \nu_{i2} \cdots \nu_{im-1}] \sum_{A_1 \amalg \cdots \amalg A_m = [K_i]} \sum_{q < r} \sum_{\substack{p_1, \dots, p_m \\ p_q = p_r}} \\ & \quad \prod_{\substack{k=1 \\ k \neq q, r}}^m (Q_i^{(N)} \prod_{l \in A_k} f_{il})(x_{ip_k}) ((Q_i^{(N)} \prod_{l \in A_q} f_{il})(Q_i^{(N)} \prod_{l \in A_r} f_{il}))(x_{ip_q}) + o(c_{N_i}). \end{aligned}$$

We will focus our attention on the first sum in the first line. Using Lemma 6 in reverse, we have

$$\begin{aligned} \frac{1}{N_i^{K_i}} \sum_{\substack{p_1, \dots, p_{K_i} \\ \text{distinct}}} \prod_{k=1}^{K_i} (Q_i^{(N)} f_{ik})(x_{ip_k}) &= \prod_{k=1}^{K_i} \langle Q_i^{(N)} f_{ik}, \mu_i \rangle \\ & - \frac{1}{N_i^{K_i}} \sum_{m=1}^{K_i-1} \sum_{A_1 \amalg \cdots \amalg A_m = [K_i]} \sum_{\substack{p_1, \dots, p_m \\ \text{distinct}}} \prod_{k=1}^m (Q_i^{(N)} \prod_{l \in A_k} f_{il})(x_{ip_k}), \end{aligned}$$

where the terms on the second line are $\mathcal{O}\left(\frac{1}{N_i}\right)$. Thus,

$$\begin{aligned} & \frac{1}{N_i^{K_i}} \sum_{m=1}^{K_i} \mathbb{E}[\nu_{i1} \cdots \nu_{im}] \sum_{A_1 \amalg \cdots \amalg A_m = [K_i]} \sum_{\substack{p_1, \dots, p_m \\ \text{distinct}}} \prod_{k=1}^m (Q_i^{(N)} \prod_{l \in A_k} f_{il})(x_{ip_k}) \\ & = \mathbb{E}[\nu_{i1} \cdots \nu_{iK_i}] \prod_{k=1}^{K_i} \langle Q_i^{(N)} f_{ik}, \mu_i \rangle \\ & + \frac{1}{N_i^{K_i}} \sum_{m=1}^{K_i-1} (\mathbb{E}[\nu_{i1} \cdots \nu_{im}] - \mathbb{E}[\nu_{i1} \cdots \nu_{iK_i}]) \sum_{A_1 \amalg \cdots \amalg A_m = [K_i]} \sum_{\substack{p_1, \dots, p_m \\ \text{distinct}}} \prod_{k=1}^m (Q_i^{(N)} \prod_{l \in A_k} f_{il})(x_{ip_k}). \end{aligned}$$

Further, we observed in Remark 4 that the differences $\mathbb{E}[\nu_{i1} \cdots \nu_{im}] - \mathbb{E}[\nu_{i1} \cdots \nu_{iK_i}]$ are $\mathcal{O}(c_{N_i})$, so that the first sum reduces to

$$\mathbb{E}[\nu_{i1} \cdots \nu_{iK_i}] \prod_{k=1}^{K_i} \langle Q_i^{(N)} f_{ik}, \mu_i \rangle + o(c_{N_i}).$$

Proceeding similarly, applying Lemma 6 with the set of $K_i - 1$ distinct functions $\{Q_i^{(N)} f_{ik}\}_{k \neq q, r} \cup \{(Q_i^{(N)} f_{iq})(Q_i^{(N)} f_{ir})\}$, we see that

$$\begin{aligned} & \frac{1}{N_i^{K_i}} \sum_{m=1}^{K_i} \mathbb{E}[(\nu_{i1})_2 \nu_{i2} \cdots \nu_{im-1}] \sum_{A_1 \amalg \cdots \amalg A_m = [K_i]} \sum_{q < r} \sum_{\substack{p_1, \dots, p_m \\ p_q = p_r}} \\ & \quad \prod_{\substack{k=1 \\ k \neq q, r}}^m (Q_i^{(N)} \prod_{l \in A_k} f_{il})(x_{ip_k}) ((Q_i^{(N)} \prod_{l \in A_q} f_{il})(Q_i^{(N)} \prod_{l \in A_r} f_{il}))(x_{ip_q}) \\ & = \frac{1}{N_i} \mathbb{E}[(\nu_{i1})_2 \nu_{i2} \cdots \nu_{iK_i-1}] \sum_{q < r} \prod_{\substack{k=1 \\ k \neq q, r}}^{K_i} \langle Q_i^{(N)} f_{ik}, \mu_i \rangle \langle (Q_i^{(N)} f_{iq})(Q_i^{(N)} f_{ir}), \mu_i \rangle + o(c_{N_i}), \end{aligned}$$

where we have used the fact that $\frac{1}{N_i} \mathbb{E}[(\nu_{i1})_2 \nu_{i2} \cdots \nu_{iK_i-1}] = \mathcal{O}(c_{N_i})$ in bounding the higher order terms. Thus,

$$\begin{aligned}
\mathbb{E} \left[F(\mathbf{G}^{(N)}(n+1)) \middle| \mathbf{G}^{(N)}(n) = \boldsymbol{\mu} \right] &= \prod_{i=0}^M \left(\mathbb{E} [\nu_{i1} \cdots \nu_{iK_i}] \prod_{k=1}^{K_i} \langle Q_i^{(N)} f_{ik}, \mu_i \rangle \right. \\
&\quad \left. + \frac{1}{N_i} \mathbb{E}[(\nu_{i1})_2 \nu_{i2} \cdots \nu_{iK_i-1}] \sum_{q < r} \prod_{\substack{k=1 \\ k \neq q, r}}^{K_i} \langle Q_i^{(N)} f_{ik}, \mu_i \rangle \langle (Q_i^{(N)} f_{iq})(Q_i^{(N)} f_{ir}), \mu_i \rangle + o(c_{N_i}) \right) \\
&= \prod_{i=0}^M \mathbb{E} [\nu_{i1} \cdots \nu_{iK_i}] \prod_{k=1}^{K_i} \langle Q_i^{(N)} f_{ik}, \mu_i \rangle + \sum_{i=0}^M \prod_{\substack{j=0 \\ j \neq i}}^M \mathbb{E} [\nu_{j1} \cdots \nu_{jK_j}] \prod_{k=1}^{K_j} \langle Q_i^{(N)} f_{jk}, \mu_i \rangle \\
&\quad \times \frac{1}{N_i} \mathbb{E}[(\nu_{i1})_2 \nu_{i2} \cdots \nu_{iK_i-1}] \sum_{q < r} \prod_{\substack{k=1 \\ k \neq q, r}}^{K_i} \langle Q_i^{(N)} f_{ik}, \mu_i \rangle \langle (Q_i^{(N)} f_{iq})(Q_i^{(N)} f_{ir}), \mu_i \rangle + o(c_{N_i}).
\end{aligned}$$

Further, recalling (4), by assumption

$$Q_i^{(N)} = I + c_{N_i} B_i^{(N)} + \mathcal{O}(c_{N_i}),$$

we have

$$\begin{aligned}
\mathbb{E} \left[F(\mathbf{G}^{(N)}(n+1)) \middle| \mathbf{G}^{(N)}(n) = \boldsymbol{\mu} \right] &= \prod_{i=0}^M \mathbb{E} [\nu_{i1} \cdots \nu_{iK_i}] \prod_{k=1}^{K_i} \langle f_{ik}, \mu_i \rangle \\
&\quad + \sum_{i=0}^M c_{N_i} \mathbb{E} [\nu_{i1} \cdots \nu_{iK_i}] \prod_{\substack{j=0 \\ j \neq i}}^M \mathbb{E} [\nu_{j1} \cdots \nu_{jK_j}] \prod_{k=1}^{K_j} \langle f_{jk}, \mu_i \rangle \sum_{q=1}^{K_i} \langle B_i^{(N)} f_{iq}, \mu_i \rangle \prod_{\substack{k=1 \\ k \neq q}}^{K_i} \langle f_{ik}, \mu_i \rangle \\
&\quad + \sum_{i=0}^M \prod_{\substack{j=0 \\ j \neq i}}^M \mathbb{E} [\nu_{j1} \cdots \nu_{jK_j}] \prod_{k=1}^{K_j} \langle f_{jk}, \mu_i \rangle \frac{1}{N_i} \mathbb{E}[(\nu_{i1})_2 \nu_{i2} \cdots \nu_{iK_i-1}] \sum_{q < r} \prod_{\substack{k=1 \\ k \neq q, r}}^{K_i} \langle f_{ik}, \mu_i \rangle \langle f_{iq} f_{ir}, \mu_i \rangle \\
&\quad + o(c_{N_i}). \tag{9}
\end{aligned}$$

Now recall,

$$(B_i^{(N)} f)(x) := \frac{\varpi_i}{2} \left(\int f(y) G^{(N)}(n)(dy) - f(x) \right),$$

and, from (3), we have

$$G^{(N)}(n) = \frac{1}{\sum_{k=0}^M N_k} \sum_{i=0}^M N_i G_i^{(N)}(n) = \frac{1}{\sum_{k=0}^M N_k} \sum_{i=0}^M N_i \mu_i = \mu_0 + \mathcal{O}\left(\frac{N_i}{N_0}\right),$$

so that

$$(B_i^{(N)} f)(x) = \frac{\varpi_i}{2} \left(\int f(y) \mu_0(dy) - f(x) \right) + o(1).$$

Now, recalling Lemma 4, we have

$$\begin{aligned}
\mathbb{E}[\nu_{i1} \cdots \nu_{iK_i}] &= 1 - \binom{K_i}{2} \frac{(N_i)_{K_i-1}}{(N_i)_{K_i}} \mathbb{E}[(\nu_{i1})_2 \nu_{i2} \cdots \nu_{iK_i-1}] - o(c_{N_i}) \\
&= 1 - \binom{K_i}{2} \frac{1}{N_i - K_i + 1} \mathbb{E}[(\nu_{i1})_2 \nu_{i2} \cdots \nu_{iK_i-1}] - o(c_{N_i}) \\
&= 1 - \binom{K_i}{2} \left(\frac{1}{N_i} + \frac{K_i - 1}{N_i(N_i - K_i + 1)} \right) \mathbb{E}[(\nu_{i1})_2 \nu_{i2} \cdots \nu_{iK_i-1}] - o(c_{N_i}) \\
&= 1 - \binom{K_i}{2} \frac{1}{N_i} \mathbb{E}[(\nu_{i1})_2 \nu_{i2} \cdots \nu_{iK_i-1}] - o(c_{N_i}),
\end{aligned}$$

so that

$$\begin{aligned}
F(\boldsymbol{\mu}) &= \prod_{i=0}^M \prod_{k=1}^{K_i} \langle f_{ik}, \mu_i \rangle \\
&= \prod_{i=0}^M \left(\mathbb{E}[\nu_{i1} \cdots \nu_{iK_i}] \prod_{k=1}^{K_i} \langle f_{ik}, \mu_i \rangle + \binom{K_i}{2} \frac{1}{N_i} \mathbb{E}[(\nu_{i1})_2 \nu_{i2} \cdots \nu_{iK_i-1}] \prod_{k=1}^{K_i} \langle f_{ik}, \mu_i \rangle + o(c_{N_i}) \right) \\
&= \prod_{i=0}^M \mathbb{E}[\nu_{i1} \cdots \nu_{iK_i}] \prod_{k=1}^{K_i} \langle f_{ik}, \mu_i \rangle \\
&\quad + \sum_{i=0}^M \prod_{\substack{j=0 \\ j \neq i}}^M \mathbb{E}[\nu_{j1} \cdots \nu_{jK_j}] \prod_{k=1}^{K_j} \langle f_{jk}, \mu_j \rangle \frac{1}{N_i} \mathbb{E}[(\nu_{i1})_2 \nu_{i2} \cdots \nu_{iK_i-1}] \sum_{q < r} \prod_{k=1}^{K_i} \langle f_{ik}, \mu_i \rangle + o(c_{N_i}). \quad (10)
\end{aligned}$$

Thus, taking the difference of (9) and (10) and using Lemmas 4 and 5 respectively to replace $\mathbb{E}[\nu_{i1} \cdots \nu_{iK_i}]$ and $\frac{1}{N_i} \mathbb{E}[(\nu_{i1})_2 \nu_{i2} \cdots \nu_{iK_i-1}]$ by $1 - \mathcal{O}(c_{N_i})$ and $c_{N_i} + o(c_{N_i})$, we see that

$$\mathbb{E} \left[F(\mathbf{G}^{(N)}(n+1)) \middle| \mathbf{G}^{(N)}(n) = \boldsymbol{\mu} \right] - F(\boldsymbol{\mu}) = c_{N_i} \sum_{i=0}^M (\mathcal{G}_i F)(\boldsymbol{\mu}) + o(c_{N_i}).$$

The first assertion follows directly.

We now observe that, restricted to the space of functions

$$\mathcal{C}_i := \left\{ F(\boldsymbol{\mu}) = \prod_{k=1}^{K_i} \langle f_{ik}, \mu_i \rangle \middle| K_i \in \mathbb{N}_0, f_{ij} \in C([0, 1]) \right\} \subseteq C(\mathcal{P}([0, 1])),$$

\mathcal{G}_i is exactly the generator (4.4) of the infinitely many alleles model of Chapter 10 of [6]. In particular, Theorem 4.1 of the same chapter tells us that given a fixed initial measure $\tilde{\mu}_i \in \mathcal{P}(\mathcal{P}([0, 1]))$, the martingale problem for $(\mathcal{G}_i, \tilde{\mu}_i)$ is well posed, *i.e.*, there exists a unique in distribution process $G_i(t)$ with initial value $G_i(0)$ distributed according to $\tilde{\mu}_i$ with generator \mathcal{G}_i . Moreover, using Theorem 1.1 of Chapter 6 of [6], we see that $G_i(\gamma_i t)$ is the unique process with generator $\gamma_i \mathcal{G}_i$. We can thus appeal to Theorem 10.1 in [6] to conclude that given an initial measure $\tilde{\boldsymbol{\mu}} = \tilde{\mu}_0 \otimes \cdots \otimes \tilde{\mu}_M$, then the martingale problem for

$$\sum_{i=1}^M \gamma_i \mathcal{G}_i$$

is well posed and has solution $G_0(0) \otimes G_1(\gamma_1 t) \otimes \cdots \otimes G_M(\gamma_M t)$.

Given convergence of the generators, and well-posedness of the limiting generator, the second assertion then follows by Lemma 5.1 in Chapter 4 of [6]. \square

Finally, we conclude this section by observing that our characterization of the limiting generator in terms of the generator of the infinitely many alleles diffusion model also allows us to characterize the stationary distribution:

Corollary 1. *The stationary process for the islands is the joint law of M independent Dirichlet processes with scaling parameters ϖ_i and base probability measure μ_0 , $DP(\varpi_i, \mu_0)$.*

Proof. This is immediate from the result for a single copy of the infinitely many alleles model. See e.g., Theorem 4.1, Chapter 9 in [6]. \square

1.5 Long-Term Behaviour

In the previous section, we simply assumed that the mainlands were asymptotically smaller in size (as measured by the coalescence probability of two randomly selected individuals) than the mainland, in order to show that the Cannings' UNTB converged to a sum of independent copies of the infinitely many alleles diffusion process, with migration from the mainland playing the role of mutation. In this section, we will show that in a slow timescale, the dynamics on the mainland converge to the standard infinitely many alleles model as well, from which we can conclude, as before, that the stationary distribution of the mainland is that of the Dirichlet process $DP(\theta, \lambda)$, where λ is the Lebesgue measure on $[0, 1]$. Thus, after a transient period, the mainland will approach a measure $\mu_0 \sim DP(\theta, \lambda)$, whereas the islands will converge on Hierarchical Dirichlet Processes $DP(\varpi_i, \mu_0)$ [14].

Let $\tilde{\nu}_i$, $i = 1, \dots, n$, be the law of the stationary process $DP(\varpi_i, \mu_0)$ from Corollary 1 above, and let $\tilde{\nu} = \tilde{\nu}_1 \otimes \dots \otimes \tilde{\nu}_M$, i.e., given a function $F \in C(\mathcal{P}([0, 1])^M)$,

$$\int F(\boldsymbol{\mu}) \tilde{\nu}(d\boldsymbol{\mu}) = \int \dots \int F(\mu_1, \dots, \mu_M) \tilde{\nu}_1(d\mu_1) \dots \tilde{\nu}_M(d\mu_M),$$

then $\tilde{\nu}$ is a stationary distribution for $\mathbf{G}(t)$: we have

$$\int (\mathcal{G}F)(\boldsymbol{\mu}) \tilde{\nu}(d\boldsymbol{\mu}) = 0$$

for all $F \in C(\mathcal{P}([0, 1])^{M+1})$, or equivalently, writing $\mathcal{T}(t)$ for the semi-group generated by \mathcal{G} , (i.e.,

$$(\mathcal{T}(t)F)(\boldsymbol{\mu}) = \mathbb{E}[F(\mathbf{G}(t)) | \mathbf{G}(0) = \boldsymbol{\mu}],$$

where $\mathbf{G}(t)$ is the process with generator \mathcal{G} of Proposition 1) we have

$$\int (\mathcal{T}(t)F)(\boldsymbol{\mu}) \tilde{\nu}(d\boldsymbol{\mu}) = \int F(\boldsymbol{\mu}) \tilde{\nu}(d\boldsymbol{\mu})$$

for all $F \in C(\mathcal{P}([0, 1])^M)$.

We start by showing that as $t \rightarrow \infty$, $\mathbf{G}(t)$ converges to a stationary process \mathbf{G}^* distributed according to $\tilde{\nu}$, (i.e.,

$$\mathbb{P}\{\mathbf{G}^*(t) \in A | \mathbf{G}^*(0) \sim \tilde{\nu}\} = \tilde{\nu}(A)$$

for all subsets $A \subseteq \mathcal{P}([0, 1])^{M+1}$). To this end, we begin with a series of lemmas, which are essentially the same as results appearing in [15]:

Lemma 8. *Let $\boldsymbol{\mu} = \mu_0 \otimes \dots \otimes \mu_M \in \mathcal{P}([0, 1])^{M+1}$ and let $F(\boldsymbol{\mu}) = \prod_{i=0}^M \prod_{k=1}^{K_i} \langle f_{ik}, \mu_i \rangle \in \mathcal{C}$. Let $K = \sum_{i=0}^M K_i$ be the degree of F . If $K \geq 1$, there exists a scalar $\lambda > 0$ and a function ψ , which is a sum of functions of the same form as F , but of degree $K - 1$, such that*

$$\mathcal{G}F = -\lambda F + \psi.$$

Thus,

$$(\mathcal{T}(t)F)(\boldsymbol{\mu}) = e^{-\lambda t} F + \int_0^t e^{-\lambda(t-s)} \mathcal{T}(s)\psi ds.$$

Proof. Recalling Equation (8), we have

$$\begin{aligned}
(\mathcal{G}_i F)(\boldsymbol{\mu}) &= \sum_{i=0}^M \sum_{1 \leq j \neq k \leq K_i} (\langle f_{ij} f_{ik}, \mu_i \rangle - \langle f_{ij}, \mu_i \rangle \langle f_{ik}, \mu_i \rangle) \prod_{l \neq j, k} \langle f_{il}, \mu_i \rangle \\
&\quad + \sum_{i=0}^M \sum_{j=1}^{K_i} \frac{\varpi_i}{2} \langle f_{ij}, x_0 - \mu_i \rangle \prod_{k \neq j} \langle f_{ik}, \mu_i \rangle \\
&= - \left(\sum_{i=0}^M \frac{K_i(K_i-1)}{2} + \frac{\varpi_i}{2} \right) F + \sum_{i=0}^M \left\langle \sum_{1 \leq j \neq k \leq K_i} f_{ij} f_{ik}, \mu_i \right\rangle \prod_{l \neq j, k} \langle f_{il}, \mu_i \rangle \\
&\quad + \sum_{i=0}^M \left\langle \sum_{j=1}^{K_i} \frac{\varpi_i}{2} f_{ij}, x_0 \right\rangle \prod_{k \neq j} \langle f_{ik}, \mu_i \rangle,
\end{aligned}$$

giving the first statement. In particular, if $K = 1$, say $K_i = 1$, we have

$$\mathcal{G}F = -\frac{\gamma_i \varpi_i}{2} F + \left\langle \frac{\gamma_i \varpi_i}{2} f_{i1}, x_0 \right\rangle.$$

For the second statement, we observe that

$$\frac{d}{dt} e^{\lambda t} \mathcal{T}(t)F = e^{\lambda t} (\lambda \mathcal{T}(t)F + \mathcal{T}(t)\mathcal{G}F) = e^{\lambda t} \mathcal{T}(t)\psi.$$

The result follows by integrating both sides over $(0, t)$. \square

With this lemma, we can show that the process $\mathbf{G}(t)$ is ergodic, *i.e.*, the distribution of $\mathbf{G}(t)$ converges on $\tilde{\nu}$, independently of the initial condition.

Proposition 2. *Let $F \in C(\mathcal{P}([0, 1])^M)$. As $t \rightarrow \infty$,*

$$\lim_{t \rightarrow \infty} \left\| \mathcal{T}(t)F - \int F(\boldsymbol{\mu}) \tilde{\nu}(d\boldsymbol{\mu}) \right\| = 0.$$

Proof. Since they are convergence-determining, it suffices to show the result for functions of the form $F \in \mathcal{C}$. We then have

$$(\mathcal{T}(t)F) = e^{-\lambda t} F + \int_0^t e^{-\lambda(t-s)} \mathcal{T}(s)\psi ds$$

for $\lambda > 0$ and ψ of degree $K - 1$. Integrating both sides, and recalling that

$$\int (\mathcal{T}(t)F)(\boldsymbol{\mu}) \tilde{\nu}(d\boldsymbol{\mu}) = \int F(\boldsymbol{\mu}) \tilde{\nu}(d\boldsymbol{\mu})$$

we have

$$\int F(\boldsymbol{\mu}) \tilde{\nu}(d\boldsymbol{\mu}) = e^{-\lambda t} \int F(\boldsymbol{\mu}) \tilde{\nu}(d\boldsymbol{\mu}) + \int_0^t e^{-\lambda(t-s)} \mathcal{T}(s) \int \psi(\mathbf{x}) \tilde{\nu}(d\boldsymbol{\mu}) ds,$$

so that

$$\left\| \mathcal{T}(t)F - \int F(\boldsymbol{\mu}) \tilde{\nu}(d\boldsymbol{\mu}) \right\| \leq e^{-\lambda t} \int F(\boldsymbol{\mu}) \tilde{\nu}(d\boldsymbol{\mu}) + \int_0^t e^{-\lambda(t-s)} \left\| \mathcal{T}(s)\psi - \int \psi(\boldsymbol{\mu}) \tilde{\nu}(d\boldsymbol{\mu}) \right\| ds.$$

The first term on the right hand side clearly vanishes as $t \rightarrow \infty$; for the latter, we can iterate the above inequality, relying on the fact that the process will eventually terminate when the degree reaches 1; when $K = 1$, say $\psi(\boldsymbol{\mu}) = \langle f_{i1}, \mu_i \rangle$, we have

$$(\mathcal{T}(t)\psi) = e^{-\frac{\omega_i}{2}t} \psi + \int_0^t e^{-\frac{\omega_i}{2}(t-s)} \mathcal{T}(s) \left\langle \frac{\omega_i}{2} f_{i1}, x_0 \right\rangle ds = e^{-\frac{\omega_i}{2}t} \psi + \int_0^t e^{-\lambda(t-s)} \left\langle \frac{\omega_i}{2} f_{i1}, x_0 \right\rangle ds$$

whereas

$$\int \psi(\boldsymbol{\mu}) \tilde{\nu}(d\boldsymbol{\mu}) = e^{-\frac{\omega_i}{2}t} \int \psi(\boldsymbol{\mu}) \tilde{\nu}(d\boldsymbol{\mu}) + \int_0^t e^{-\lambda(t-s)} \langle \frac{\omega_i}{2} f_{i1}, x_0 \rangle ds,$$

so that

$$\left\| \mathcal{T}(t)\psi - \int \psi(\boldsymbol{\mu}) \tilde{\nu}(d\boldsymbol{\mu}) \right\| = e^{-\frac{\omega_i}{2}t} \left\| \psi - \int \psi(\boldsymbol{\mu}) \tilde{\nu}(d\boldsymbol{\mu}) \right\| \rightarrow 0$$

as $t \rightarrow \infty$. □

Define a linear map \mathcal{P} on $C(\mathcal{P}([0, 1])^M)$ by

$$\mathcal{P}F = \int F(\boldsymbol{\mu}) \tilde{\nu}(d\boldsymbol{\mu}),$$

i.e., \mathcal{P} sends $F \in C(\mathcal{P}([0, 1])^M)$ to a constant function; more generally, if $F \in C(\mathcal{P}([0, 1])^{M+1})$, $\mathcal{P}F$ is a function of μ_0 alone. In particular,

$$(\mathcal{P}F)(\mu_0) = \mathbb{E}[F(\mu_0, G_1(t), \dots, G_M(t)) | G_i(0) \sim \tilde{\nu}_i],$$

so that applying the operator \mathcal{P} is equivalent to conditioning on the islands being at their stationary state.

Note that $\mathcal{P}^2 = \mathcal{P}$, so that \mathcal{P} is a projection. Moreover,

$$\mathcal{P}(\mathcal{G}F) = \int (\mathcal{G}F) \tilde{\nu}(d\boldsymbol{\mu}) = 0,$$

so the range of \mathcal{G} is contained in the null space of \mathcal{P} , $\mathcal{R}(\mathcal{G}) \subseteq \mathcal{N}(\mathcal{P})$, whereas $\mathcal{G}1 = 0$, so that $\mathcal{R}(\mathcal{P}) \subseteq \mathcal{N}(\mathcal{G})$.

In fact, we have:

Lemma 9. \mathcal{P} is the spectral projection onto $\mathcal{N}(\mathcal{G})$.

Proof. By definition, the spectral projection onto $\mathcal{N}(\mathcal{G})$, \mathcal{Q} , is the residue of the resolvent of \mathcal{G} at $\lambda = 0$:

$$\mathcal{Q} = \lim_{\lambda \rightarrow 0^+} \lambda(\lambda - \mathcal{G})^{-1} = \lim_{\lambda \rightarrow 0^+} \lambda \int_0^\infty e^{-\lambda t} \mathcal{T}(t) dt.$$

Now, fix $\varepsilon > 0$ and choose $t_0 > 0$ so that $\|\mathcal{T}(t) - \mathcal{P}\| < \varepsilon$ for $t > t_0$. Then, for $\lambda > 0$,

$$\begin{aligned} \left\| \lambda \int_0^\infty e^{-\lambda t} \mathcal{T}(t) dt - \mathcal{P} \right\| &= \left\| \lambda \int_0^\infty e^{-\lambda t} (\mathcal{T}(t) - \mathcal{P}) dt \right\| \\ &\leq \lambda \int_0^\infty e^{-\lambda t} \|\mathcal{T}(t) - \mathcal{P}\| dt \\ &= \lambda \int_0^{t_0} e^{-\lambda t} \|\mathcal{T}(t) - \mathcal{P}\| dt + \lambda \int_{t_0}^\infty e^{-\lambda t} \|\mathcal{T}(t) - \mathcal{P}\| dt \\ &\leq \lambda t_0 \sup_{t \leq t_0} \|\mathcal{T}(t) - \mathcal{P}\| + \varepsilon. \end{aligned}$$

$\|\mathcal{T}(t) - \mathcal{P}\|$ is a continuous function, and thus bounded on $[0, t_0]$. Thus the first term vanishes as $\lambda \rightarrow 0^+$, whereas ε can be chosen arbitrarily small. We conclude $\mathcal{Q} = \mathcal{P}$. □

With this, we are able to obtain our final result.

Proposition 3. Assume, as before, that

$$\lim_{N \rightarrow \infty} \frac{c_N}{a_N} = 0.$$

Let \mathcal{P} be the projection defined above. Define an operator \mathcal{G}_0 on \mathcal{C}_0 by

$$(\mathcal{G}_0 F)(\boldsymbol{\mu}) = \left(\sum_{q=1}^{K_0} \frac{\theta}{2} \langle f_{0q}, \lambda - \mu_0 \rangle \prod_{\substack{k=1 \\ k \neq q}}^{K_0} \langle f_{0k}, \mu_0 \rangle + \frac{1}{2} \sum_{q \neq r} \prod_{\substack{k=1 \\ k \neq q, r}}^{K_0} \langle f_{0k}, \mu_0 \rangle (\langle f_{0q} f_{0r}, \mu_0 \rangle - \langle f_{0q}, \mu_0 \rangle \langle f_{0r}, \mu_0 \rangle) \right), \quad (11)$$

and let $\mathcal{T}_0(t)$ be the semigroup generated by $\mathcal{P}\mathcal{G}_0$. Then, for all $F \in C(\mathcal{P}([0, 1]^M))$, and all $\delta \in (0, 1)$ we have

$$\left(I + \mathcal{G}^{(N)} \right)^{\lfloor c_N^{-1} t \rfloor} F \rightarrow \mathcal{T}_0(t) \mathcal{P} F$$

uniformly in $\delta \leq t \leq \delta^{-1}$. If in addition, we assume that $G_i(0) \sim \tilde{v}_i$ for all $i = 1, \dots, M$, and $G_0(t)$ is a stochastic process with generator \mathcal{G}_0 , then

$$\mathbf{G}^{(N)}(\lfloor c_N^{-1} t \rfloor) \Rightarrow \mathbf{G}(t) = G_0(t) \otimes G_1(t) \cdots \otimes G_M(t),$$

where the processes $G_i(t)$ are stationary for all $i = 1, \dots, M$.

Remark 10. The heuristic understanding of Proposition 3 is that

$$\mathcal{G}^{(N)} = c_N^{-1} a_N \mathcal{G} + c_N^{-1} \mathcal{G}_0 + \text{lower order terms}$$

where $\mathcal{H}\mathcal{P} \equiv 0$. Now $c_N^{-1} a_N \rightarrow \infty$ as $N \rightarrow \infty$, so the first term dominates. $c_N^{-1} a_N$ is essentially the rate at which the first term shapes the dynamics of the process, and so as N grows large, the first term, which acts only on the islands, causes them to rapidly approach their equilibrium state (which, as we have already seen, corresponds to projection by \mathcal{P}). The first term, however, has no effect on the mainland. Moreover, the mainland only changes at the slower rate c_N^{-1} . Thus, the first term has already forced the faster terms to equilibrium, and we can assume that they are at equilibrium when we consider the mainland. Finally, the first two terms completely specify the limit, so what remains can only contribute a higher order correction. This is essentially the infinite dimensional analogue of the following simple dynamical system:

$$\begin{aligned} \dot{x} &= -Nax + f(x, y), \\ \dot{y} &= -\sqrt{N}by + g(x, y), \end{aligned}$$

for $a, b > 0$. Using variation of constants, we have

$$\begin{aligned} x(t) &= e^{-Nat} x(0) + \int_0^t e^{-Na(t-s)} f(x(s), y(s)) ds, \\ y(t) &= e^{-\sqrt{N}bt} y(0) + \int_0^t e^{-\sqrt{N}b(t-s)} g(x(s), y(s)) ds. \end{aligned}$$

Thus, provided f and g are bounded,

$$\int_0^t e^{-Na(t-s)} f(x(s), y(s)) ds \leq \frac{1}{Na} \|f\|,$$

and

$$\int_0^t e^{-\sqrt{N}b(t-s)} g(x(s), y(s)) ds \leq \frac{1}{\sqrt{N}b} \|g\|,$$

so that as $N \rightarrow \infty$, we have $x(t) = 0 + \mathcal{O}(\frac{1}{N})$. We can thus substitute this back into the equation for $y(t)$ to conclude that

$$y(t) = e^{-\sqrt{N}bt}y(0) + \int_0^t e^{-\sqrt{N}b(t-s)}g(0, y(s)) ds + \mathcal{O}(\frac{1}{N}),$$

(setting $x(t) \equiv 0$ is equivalent to the action of the projection \mathcal{P}). Thus, similarly, $y(t) = 0 + \mathcal{O}(\frac{1}{\sqrt{N}})$.

Remark 11. It is necessary to assume $G_i(0) \sim \tilde{\nu}_i$ to obtain continuity of $\mathcal{T}_0(t)\mathcal{P}$ at $t = 0$, which in turn is required to ensure weak convergence. More generally, Proposition 3 tell us that in the slow timescale, the island demes instantaneously jump to their stationary states, and henceforth evolve as stationary processes; see [16] and [5] for more detailed discussions of processes with this behaviour.

Proof. Calculations essentially identical to those in Proposition 1 show that, when restricted to \mathcal{C}_0 , $c_N^{-1}\mathcal{G}^{(N)} = \mathcal{G}_0 + o(c_N)$, with the primary difference being with the operator $Q_0^{(N)}$. Here,

$$Q_0^{(N)} = I + c_N B_0^{(N)} + c_N B + o(c_N),$$

where, as before

$$(B_0^{(N)}f)(x) = \frac{\varpi_i}{2} (\langle f, \mu_0 \rangle - f(x)) + o(1),$$

but now

$$(Bf)(x) = \frac{\theta}{2} \int_0^1 f(y) dy - f(x) = \theta(\langle f, \lambda \rangle - f(x))$$

(recall that λ is Lebesgue measure, $\lambda(dx) = dx$) is of the same asymptotic order. Moreover, we now only consider terms of the form $\langle Q_0^{(N)}f_{0k}, \mu_0 \rangle$, and

$$\langle B_0^{(N)}f_{0k}, \mu_0 \rangle = \frac{\varpi_i}{2} (\langle f, \mu_0 \rangle - \langle f, \mu_0 \rangle) + o(1),$$

which vanishes in the limit. Thus,

$$c_N \langle Q_0^{(N)}f_{0k}, \mu_0 \rangle - \langle f_{0k}, \mu_0 \rangle = \frac{\theta}{2} \int_0^1 f(y) dy - f(x) + o(1) = \theta(\langle f, \lambda \rangle - f(x)) + o(1),$$

giving the corresponding terms in the generator (11).

The first statement is then a restatement of Corollary 7.7, Chapter 1 of [6]; translating our notation into theirs, we have

$$\begin{aligned} \varepsilon_N &= c_N, \\ \alpha_N &= c_N^{-1}a_N, \\ A_N &= c_N^{-1}\mathcal{G}^N, \end{aligned}$$

$B = \mathcal{G}$, and $A = \mathcal{G}_0$. That \mathcal{G}_0 generates a strongly continuous semigroup is Theorem 4.1, Chapter 10 of [6], which we used previously.

The second statement is a consequence of Corollary 8.9, Chapter 4, [6], where our initial condition ensures continuity of the semigroup $\mathcal{T}_0(t)$ at $t = 0$. \square

2 Gibbs Sampling for the UNTB-HDP

2.1 Observed abundances

The observed data takes the form of an $N \times S$ matrix of counts \mathbf{X} whose elements x_{ij} are the observed frequency of species j in community sample i . Here, N denotes the total number of communities and S the total number of different species found in those communities. We will also denote the row vectors of \mathbf{X} , which give the observed frequency distribution of species in each individual sample, by \bar{X}_i , $i = 1, \dots, N$. The size of each sample is simply $J_i = \sum_{j=1}^S x_{ij}$.

2.2 Neutral-HDP model

$$\bar{\beta}|\theta \sim \text{Stick}(\theta), \quad (12)$$

$$\bar{\pi}_i|I_i, \bar{\beta} \sim \text{DP}(I_i, \bar{\beta}), \quad (13)$$

$$\bar{X}_i|\bar{\pi}_i, J_i \sim \text{MN}(J_i, \bar{\pi}_i). \quad (14)$$

This model for the observed frequencies can be interpreted as the generation of an infinite dimensional meta-community distribution $\bar{\beta}$ which is obtained from a stick-breaking or GEM distribution with concentration parameter θ . From this, for each community i we sample using the Dirichlet process a vector of taxa probabilities $\bar{\pi}_i$ which has concentration I_i , the immigration rate for that site, and base distribution $\bar{\beta}$. Finally, we sample the observed frequencies for each community \bar{X}_i from $\bar{\pi}_i$ using the multinomial distribution. We also include gamma hyper-priors for θ and the I_i :

$$\theta|\alpha, \zeta \sim \text{Gamma}(\alpha, \zeta), \quad (15)$$

$$I_i|\eta, \kappa \sim \text{Gamma}(\eta, \kappa), \quad (16)$$

where α, ζ, η and κ are all constants. This completes the definition of our model.

2.3 Finite dimensional representation

In any given sample although the potential number of species is infinite we only observe S different types. It is convenient therefore to represent the model in terms of these finite dimensional number of types and one further class corresponding to all unobserved species. We will derive this as the limit of L total types as $L \rightarrow \infty$. We will represent the proportions of the S observed species explicitly as β_k with $k = 1, \dots, S$ and the unrepresented component as $\beta_u = \sum_{k=S+1}^L \beta_k$. Let $\theta_r = \theta/L$ and $\theta_u = \theta(L - S)/L$, then we will have a Dirichlet prior on $\bar{\beta} \sim \text{Dir}(\theta_r, \dots, \theta_r, \theta_u)$. In this finite dimensional representation we can also determine the distributions in the local communities:

$$\bar{\pi}_i \sim \text{Dir}(I_i\beta_1, \dots, I_i\beta_S, I_i\beta_u). \quad (17)$$

We can then marginalise the local community distributions and derive the probability of the observed frequencies given $\bar{\beta}$:

$$P(\mathbf{X}|\bar{\beta}, I_1, \dots, I_N) = \prod_{i=1}^N \frac{J_i!}{X_{i1}! \dots X_{iS}!} \frac{\Gamma(I_i)}{\Gamma(J_i + I_i)} \prod_{j=1}^S \frac{\Gamma(x_{ij} + I_i\beta_j)}{\Gamma(I_i\beta_j)}. \quad (18)$$

2.4 Gibbs sampling

To devise a Gibbs sampling strategy we need to determine the full conditional distributions of the parameters we wish to sample, θ and I_i , for $i = 1, \dots, N$. Our starting point will be the joint distribution of these parameters and the data, that is, Equation 18 multiplied by the prior distributions for $\bar{\beta}$, θ and I_i , marginalised over $\bar{\beta}$:

$$P(\theta, I_1, \dots, I_N, \mathbf{X}) = \int_{\bar{\beta}} P(\mathbf{X}|\bar{\beta}, I_1, \dots, I_N) P(\bar{\beta}|\theta) d\bar{\beta} \text{Gamma}(\theta|\alpha, \zeta) \prod_{i=1}^N \text{Gamma}(I_i|\eta, \nu). \quad (19)$$

The key to simplifying this expression is to expand the terms $\Gamma(x_{ij} + I_i\beta_j)/\Gamma(I_i\beta_j)$ in Equation 18 as polynomials [14]:

$$\frac{\Gamma(x_{ij} + I_i\beta_j)}{\Gamma(I_i\beta_j)} = \sum_{T_{ij}=0}^{T_{ij}=x_{ij}} s(x_{ij}, T_{ij})(I_i\beta_j)^{T_{ij}}, \quad (20)$$

where the coefficients $s(x_{ij}, T_{ij})$ are unsigned Stirling numbers of the first kind. We substitute these sums into Equation 19 and then introduce the T_{ij} and $\bar{\beta}$ as auxiliary variables to give:

$$Q(\theta, \bar{\beta}, I_1, \dots, I_N, T_{ij}) \propto \left(\prod_{i=1}^N \frac{J_i!}{X_{i1}! \cdots X_{iS}!} \frac{\Gamma(I_i)}{\Gamma(J_i + I_i)} \prod_{j=1}^S s(x_{ij}, T_{ij}) (I_i \beta_j)^{T_{ij}} \right) P(\bar{\beta} | \theta) \text{Gamma}(\theta | \alpha, \zeta) \prod_{i=1}^N \text{Gamma}(I_i | \eta, \nu). \quad (21)$$

2.4.1 Full conditional for the ancestral states

From Equation 21, we see that the full conditional distribution for the number of ancestors (tables in the Chinese restaurant franchise analogy) of species j in sample i is given by:

$$P(T_{ij} | x_{ij}, I_i, \beta_j) \propto s(x_{ij}, T_{ij}) (I_i \beta_j)^{T_{ij}}. \quad (22)$$

The reciprocal of Equation 20 is the normalising constant of this probability distribution and thus:

$$P(T_{ij} | x_{ij}, I_i, \beta_j) = \frac{\Gamma(I_i \beta_j)}{\Gamma(x_{ij} + I_i \beta_j)} s(x_{ij}, T_{ij}) (I_i \beta_j)^{T_{ij}}. \quad (23)$$

2.4.2 Full conditional for the metapopulation

In their derivation of a posterior sampling scheme for the hierarchical Dirichlet process mixture model using an augmented Chinese restaurant franchise representation, [14] showed that the full conditional distribution for the metapopulation vector $\bar{\beta}$ was:

$$\bar{\beta} = (\beta_1, \beta_2, \dots, \beta_S, \beta_u) \sim \text{Dir}(T_{.1}, T_{.2}, \dots, T_{.S}, \theta), \quad (24)$$

where $T_{.j} = \sum_{i=1}^N T_{ij}$.

2.4.3 Full conditional for the immigration rates

To derive the full conditional distribution of each I_i given the other parameters we simply pull out all terms that depend on I_i from Equation 21. This gives:

$$P(I_i | T_{ij}) \propto \frac{\Gamma(I_i)}{\Gamma(J_i + I_i)} I_i^{T_{i.}} \text{Gamma}(I_i | \eta, \nu), \quad (25)$$

where $T_{i.} = \sum_{j=1}^S T_{ij}$. We can use the auxiliary variable approach of [17] to develop a Gibbs sampling update for I_i , $i = 1, \dots, N$. Here, for each i , we can write:

$$\frac{\Gamma(I_i)}{\Gamma(J_i + I_i)} = \frac{1}{\Gamma(J_i)} \int_0^1 w_i^{I_i} (1 - w_i)^{J_i - 1} \left(1 + \frac{J_i}{I_i} \right) dw_i \quad (26)$$

(cf. with equation (A.2) of [14]). We now define auxiliary variables $\bar{w} = (w_i)_{i=1}^N$ and $\bar{s} = (s_i)_{i=1}^N$, where each w_i is a variable taking on values in $[0, 1]$ and each s_i is a binary $\{0, 1\}$ variable, and define the following distribution:

$$q(I_i, \bar{w}, \bar{s}) \propto \prod_{i=1}^N I_i^{\eta - 1 + T_{i.}} e^{-\nu I_i} w_i^{I_i} (1 - w_i)^{J_i - 1} \left(\frac{J_i}{I_i} \right)^{s_i} \quad (27)$$

(cf. with equation (A.3) of [14]). Now marginalising q to I_i gives the desired conditional distribution for I_i . Hence q defines an auxiliary variable sampling scheme for I_i . Given \bar{w} and \bar{s} , we have:

$$q(I_i | \bar{w}, \bar{s}) \propto I_i^{\eta - 1 + T_{i.} - s_i} e^{-I_i(\nu - \log w_i)}, \quad (28)$$

which is a Gamma distribution with parameters $\eta + T_i - s_i$ and $\nu - \log w_i$ (cf. with equation (A.4) of [14]). Given I_i , the w_i and s_i are conditionally independent, with distributions:

$$q(w_i|I_i) \propto w_i^{I_i}(1 - w_i)^{J_i-1} \quad (29)$$

and

$$q(s_i|I_i) \propto \left(\frac{J_i}{I_i}\right)^{s_i}, \quad (30)$$

which are Beta($I_i + 1, J_i$) and Bernoulli($\frac{J_i}{I_i}$), respectively (cf. with equations (A.5) and (A.6) of [14]).

2.4.4 Full conditional for the biodiversity parameter

A direct consequence of the stick-breaking prior for $\bar{\beta}$ is that the probability of observing S species from a total number of $T = \sum_{i=1}^N \sum_{j=1}^S T_{ij}$ ancestors is given by:

$$P(S|\theta, T) = s(T, S)\theta^S \frac{\Gamma(\theta)}{\Gamma(\theta + T)} \quad (31)$$

(cf. with equation (A.7) of [14]). The biodiversity parameter θ does not govern any other aspects of the joint distribution in Equation 21, hence Equation 31, along with the prior for θ in Equation 15, is all that is needed to derive a Gibbs sampling update for θ . The auxiliary variable approach of [17] can also be applied here, which leads to the following auxiliary variable sampling scheme for θ :

$$\theta|\rho, \phi, S \sim \text{Gamma}(\alpha + S - \rho, \zeta - \log \phi), \quad (32)$$

$$\rho|\theta, T \sim \text{Bernoulli}\left(\frac{T}{T + \theta}\right), \quad (33)$$

$$\phi|\theta, T \sim \text{Beta}(\theta + 1, T). \quad (34)$$

2.5 Results

In order to examine how well our HDP estimation approach performed in comparison with existing methods [18–20], we used a combination of simulated data and real data that had been analysed before. Firstly, we generated 1,000 simulated data sets of three local samples with 1,000 individuals each for the eight parameter combinations given in Table 1. Note that the migration probability is simply $m_i = I_i/(I_i + J_i - 1)$. These data sets were generated using the PARI/GP code provided in [18], which is an urn algorithm based on coalescence theory. We then estimated the parameters using the Gibbs sampling approach based on the HDP approximation and the approximate two stage approach of [19]. Tables 2 and 3 gives the means, coefficients of variation and mean absolute deviations from the true values of our approach and Etienne’s two stage approximate method, respectively, across the 1,000 data sets for each parameter combination.

For all parameter combinations considered the HDP approximation outperforms Etienne’s approximation as an estimator of θ , as in each case the overall means are closer to the true values and the coefficients of variations and mean absolute deviations from the true values are considerably smaller. The HDP approximation provides a less biased and more reliable estimator of θ than Etienne’s approximation.

A similar pattern is observed with the estimates of the immigration probabilities m_i , as for the parameter combinations considered our approach gives lower coefficients of variation and mean absolute deviations from the true value than Etienne’s approximate method. Both approximations break down when the immigration rate I is significantly larger than the fundamental biodiversity parameter θ (for example, see the estimates of m_3 for synthetic data sets 1-5 in Tables 2 and 3), but in different ways. Our method underestimates the immigration probability m in such cases, but the standard deviation around that estimate remains low, and thus our estimator for m is biased when $I > \theta$, but as this bias is consistent it would be possible to correct for it. On the other hand, Etienne’s approximate approach gives an overall mean over the 1,000 simulated data sets that is much closer to the true value in such a case than our method does. However, the variability

Data set	J_i	θ	I_1	I_2	I_3	m_1	m_2	m_3
1	1000	5	111	249.75	666	0.1	0.2	0.4
2	1000	50	111	249.75	666	0.1	0.2	0.4
3	1000	500	111	249.75	666	0.1	0.2	0.4
4	1000	5	10.0909	52.5789	333	0.01	0.05	0.25
5	1000	50	10.0909	52.5789	333	0.01	0.05	0.25
6	1000	500	10.0909	52.5789	333	0.01	0.05	0.25
7	1000	5	1	2.002	4.012	0.001	0.002	0.004
8	1000	50	1	2.002	4.012	0.001	0.002	0.004

Table 1: The parameter values chosen for the synthetic neutral model data sets that composed our simulation study.

Data set	$\hat{\theta}$	CV	MAD	\hat{m}_1	CV	MAD	\hat{m}_2	CV	MAD	\hat{m}_3	CV	MAD
1	5.4092	0.20	0.8950	0.0934	0.29	0.0232	0.1508	0.23	0.0522	0.2002	0.19	0.1998
2	51.5476	0.09	3.9993	0.0990	0.14	0.0114	0.1923	0.15	0.0242	0.3262	0.12	0.0749
3	498.8622	0.07	25.8993	0.0999	0.08	0.0067	0.1982	0.07	0.0119	0.3836	0.07	0.0252
4	5.4477	0.22	1.0088	0.0110	0.42	0.0032	0.0526	0.36	0.0144	0.1417	0.26	0.1083
5	51.7504	0.12	4.8836	0.0101	0.21	0.0017	0.0504	0.17	0.0065	0.2211	0.16	0.0387
6	488.8805	0.10	40.7537	0.0100	0.17	0.0014	0.0503	0.10	0.0040	0.2495	0.09	0.0171
7	5.3388	0.46	1.8189	0.0014	0.96	0.0007	0.0030	0.98	0.0015	0.0066	0.95	0.0035
8	55.0994	0.43	17.2483	0.0010	0.44	0.0004	0.0022	0.34	0.0006	0.0043	0.29	0.0009

Table 2: Estimates of θ and m_i from the various scenarios of simulated data sets of Table 1 using the hierarchical Dirichlet process approximation. The values reported are the means, coefficients of variation and mean absolute deviations from the true value of the parameter estimates over 1000 such data sets.

Data set	$\hat{\theta}$	CV	MAD	\hat{m}_1	CV	MAD	\hat{m}_2	CV	MAD	\hat{m}_3	CV	MAD
1	5.9130	0.40	1.9880	0.1899	1.45	0.1621	0.2763	1.14	0.2300	0.4057	1.10	0.3260
2	51.9033	0.20	8.2626	0.1071	0.44	0.0274	0.2239	0.56	0.0776	0.4231	0.48	0.1556
3	507.2382	0.12	50.4488	0.1006	0.09	0.0070	0.2010	0.09	0.0138	0.4032	0.12	0.0356
4	6.0710	0.45	2.1911	0.0410	3.62	0.0356	0.1177	1.88	0.1042	0.3086	1.11	0.2666
5	54.2026	0.29	12.6540	0.0102	0.55	0.0020	0.0580	0.83	0.0190	0.2897	0.72	0.1440
6	578.4131	0.36	166.5742	0.0100	0.18	0.0014	0.0503	0.13	0.0048	0.2601	0.34	0.0503
7	9.9517	1.41	6.5506	0.0164	7.03	0.0158	0.0348	4.69	0.0338	0.0473	3.88	0.0450
8	860.1590	7.00	824.9333	0.0011	1.61	0.0004	0.0022	0.73	0.0007	0.0075	6.32	0.0045

Table 3: Estimates of θ and m_i from the various scenarios of simulated data sets of Table 1 using Etienne's approximate method. The values reported are the means, coefficients of variation and mean absolute deviations from the true value of the parameter estimates over 1000 such data sets.

around Etienne’s approximate estimate of m is much higher because the algorithm often converges to an immigration probability of 1, even when the true value is much lower. It is also worth noting that Etienne’s approximate method also breaks down badly for data sets 7 and 8 where the immigration probabilities are very low, whereas the HDP approximation copes much better in such scenarios. Thus, we conclude that the HDP approximation is a better estimator of the neutral model’s parameters than Etienne’s approximation unless $I \gg \theta$ and the immigration probabilities are close to 1.

In Table 4, we present the average times in seconds of Etienne’s approximate method using the code given in [19] and PARI/GP’s default settings, and our Gibbs sampling approach coded in C++ when it was run for 50,000 iterations with half of these being conservatively discarded as burn-in. Under these settings, for all but one of the simulated data set scenarios of Table 1, Etienne’s approximate method is two to three times faster than our approach. However, we are being very conservative with sample number and equivalent results could be achieved with as little as 10,000 iterations when the two methods would be of comparable speed.

We were unable to replicate these results using Etienne’s ‘exact’ maximum likelihood method, so instead we quote those that he gave in a similar simulation study [20] in Table 5. We see that Etienne’s ‘exact’ method slightly outperforms the HDP approximation as an estimator of θ , as although the coefficients of variation are broadly similar, the overall means are generally closer to their true values and thus Etienne’s ‘exact’ method is less biased for this parameter. Regarding the estimation of immigration probabilities, the results are comparable when $\theta \leq I$. When $\theta > I$, there is a tendency for Etienne’s ‘exact’ method to overestimate the immigration probability, but not as badly as the HDP approximation underestimates it. The advantage of the HDP approximation is that our code is easier to implement than Etienne’s ‘exact’ method’s PARI/GP algorithm, it is much faster, and our approach can handle the large data sets often encountered in microbiomics.

As an example of how the methods compare on real data, we reanalysed the tropical tree data set used as an example in [18–20]. The data consists of three forest plots in Panama called Barro Colorado Island (50 ha), Cocoli (4 ha) and Sherman (5.96 ha), which lie along a precipitation gradient [21]. Table 6 shows the results of the parameter estimation for Etienne’s three methods and our HDP approach. We see that in this case the results from the HDP approximation closely match Etienne’s ‘exact’ method, while his approximate method overestimates θ and underestimates the immigration rates. The matching results of our approach and Etienne’s ‘exact’ method is unsurprising as in this case $\theta \gg I_i$.

References

1. Cannings C (1974) The latent roots of certain markov chains arising in genetics: A new approach, I. haploid models. *Adv Appl Prob* 6: 260–290.
2. Parsons TL, Quince C (2007) Fixation in haploid populations exhibiting density dependence II: The quasi-neutral case. *Theor Popul Biol* 72: 468–479.
3. Parsons TL, Quince C, Plotkin JB (2008) Expected times to absorption and fixation for quasi-neutral and neutral haploid populations exhibiting density dependence. *Theor Popul Biol* 74: 302–310.
4. Parsons TL, Quince C, Plotkin JB (2010) Some consequences of demographic stochasticity in population genetics. *Genetics* 185: 1345–1354.
5. Parsons TL (2012) Asymptotic Analysis of Some Stochastic Models from Population Dynamics and Population Genetics. Ph.D. thesis, University of Toronto.
6. Ethier SN, Kurtz TG (1986) *Markov Processes: Characterization and Convergence*. New York: John Wiley and Sons.
7. Hubbell SP (2001) *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press.

Data set	Etienne’s approximation	HDP approximation
1	13.8583	40.6223
2	21.5615	41.1254
3	208.6595	41.5881
4	14.9588	41.8532
5	14.9767	40.6765
6	27.3442	42.4084
7	20.0091	56.1613
8	17.8649	57.5658

Table 4: Average time in seconds that Etienne’s approximate method and the HDP approximation took to run on the various scenarios of simulated data sets of Table 1. Note that the HDP approximation was run for 50,000 iterations and half of these were conservatively discarded as burn-in.

Data set	$\hat{\theta}$	CV	\hat{m}_1	CV	\hat{m}_2	CV	\hat{m}_3	CV
1	4.9689	0.21	0.1119	0.44	0.2353	0.49	0.4727	0.50
2	49.9838	0.10	0.1022	0.16	0.2041	0.16	0.4105	0.18
3	501.5142	0.07	0.1005	0.08	0.2009	0.08	0.4007	0.08
4	4.8982	0.25	0.0108	0.43	0.0572	0.46	0.3658	0.70
5	49.9892	0.12	0.0103	0.21	0.0513	0.16	0.2643	0.25
6	504.0792	0.11	0.0101	0.17	0.0504	0.11	0.2521	0.09
7	5.0388	0.45	0.0012	0.67	0.0027	1.27	0.0066	4.85
8	56.0378	0.55	0.0010	0.42	0.0020	0.35	0.0042	0.30

Table 5: Estimates of θ and m_i from the various scenarios of simulated data sets of Table 1 using Etienne’s ‘exact’ maximum likelihood method. The values reported are the means and coefficients of variation over 1000 such data sets, and were obtained from [20].

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

Table 6: Neutral parameter estimates for samples from three local tree communities (Sherman, BCI and Cocoli) in the Panama Canal Zone using Etienne’s approaches and the hierarchical Dirichlet process approximation. Standard errors are given for the methods where they are available.

8. Möhle M (2001) Forward and backward diffusion approximations for haploid exchangeable population models. *Stochastic Processes Appl* 95: 133–149.
9. Sjödin P, Kaj I, Krone S, Lascoux M, Nordborg M (2005) On the meaning and existence of an effective population size. *Genetics* 169: 1061–70.
10. Möhle M, Sagitov S (2003) Coalescent patterns in diploid exchangeable population models. *J Math Biol* 47: 337–352.
11. Ewens WJ (1972) The sampling theory of selectively neutral alleles. *Theor Popul Biol* 3: 87–112.
12. Kingman JFC (1982) On the genealogy of large populations. *J Appl Prob* 19A: 27–43.
13. Abramowitz M, Stegun IA (1964) *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York: Dover.
14. Teh YW, Jordan MI, Beal MJ, Blei DM (2006) Hierarchical Dirichlet processes. *Journal of the American Statistical Association* 101: 1566–1581.
15. Ethier SN, Kurtz TG (1981) The infinitely-many-neutral-alleles diffusion model. *Advances in Applied Probability* : 429–452.
16. Katzenberger GS (1991) Solutions of a stochastic differential equation forced onto a manifold by a large drift. *Ann Probab* 19: 1587–1628.
17. Escobar MD, West M (1995) Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* 90: 577–588.
18. Etienne RS (2007) A neutral sampling formula for multiple samples and an ‘exact’ test of neutrality. *Ecology Letters* 10: 608–618.
19. Etienne RS (2009) Improved estimation of neutral model parameters for multiple samples with different degrees of dispersal limitation. *Ecology* 90: 847–852.
20. Etienne RS (2009) Maximum likelihood estimation of neutral model parameters for multiple samples with different degrees of dispersal limitation. *Journal of Theoretical Biology* 257: 510–514.
21. Condit R, Pitman N, Leigh Jr E, Chave J, Terborgh J, et al. (2002) Beta-diversity in tropical forest trees. *Science* 295: 666–669.