1 Two-island literature review

The n-island model was developed by Sewall Wright Slatkin (1985) and models the interactions between n populations that have a social structure. An n-island structure means that a total population is divided in to n disjoint groups, and these groups of individuals (or genes) can interact with one another through migration. It has been studied by various researchers since its development, including ... who has studied the n-island model? What were they researching specifically? Arredondo et al. (2021) where it has been used to infer demographic history of genes.

Since the island model assumes social structure, it is more accurately representative of a true population in comparison to the Wright-Fisher population model which assumes no population structure. Most species display population structure to some extent, which is why the n-island model is more useful in such instances. That is, genes reproduce with other individuals that are in physical proximity. In humans, for example, the probability of mating within the same continent is larger than mating between them. On the other hand, for plant species, the probability of pollination is larger for individual plants close together than for individual plants at opposite ends of the field Hein et al. (2004).

Wright's infinite island model has also been used to study evolutionary game dynamics (cite Ohtsuki, 2010).

"The simplest model is that in which the total population is assumed to be divided into subgroups, each breeding at random within itself, except for a certain proportion of migrants drawn at random from the whole (Wright, 1943)."

2 My work

We seek to understand the ancestral genealogy of a population by starting with a current day sample and looking backwards in time. Coalescent theory aims to answer questions about this ancestry and the behaviour of populations. Genealogical quantities of interest are introduced, including tree height and total branch length, and their biological interpretation is discussed. One of the simplest models to describe the genealogical relationship among genes over a series of generations was introduced by Wright (1931) and Fisher (1930). This reproductive model describes the evolution of a population and hence the transmission of genes through subsequent genes in discrete time (see HERE for more information about this model). The continuous time coalescent... From this Wright-Fisher model the continuous time coalescent is derived, where quantities of interest are

- Expected tree height and

- Total branch length.

Throughout the following section we will consider a coalescent process such that the underlying genealogy follows the two-island model. By looking backwards-in-time we therefore consider coalescent events as well as migration events. We will

denote our populations as Island 1 and Island 2. We define the following rate parameters:

- M, rate at which a single gene migrates from Island 1 to Island 2,

- R, the rate at which a single gene migrates from Island 2 to Island 1,

- α, the rate at which a single gene can coalesce on Island 1,

- β, the rate at which a single gene can coalesce on Island 2,

and we suppose $i = I_1 + I_2$ represents total population size, where I_1 and I_2 are population sizes of Island 1 and Island 2, respectively. Figure 2 shows the dynamics of such a population.
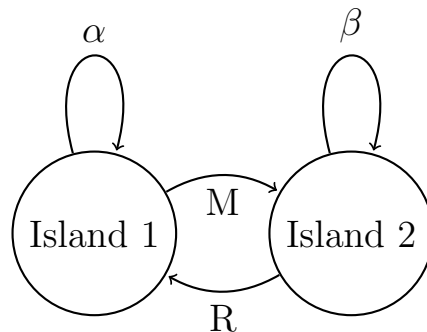

Figure 1: Fix this diagram to make it more intuitive!!!

What are some descriptors of the two-island model?

The Island Model introduced by *Sewall Wright (1951)* has proven to be a useful construction for studying the interaction of genetic drift, population subdivision, and mutation.

2.1 Coalescence and the island model

Coalescent models have been studied in the context of the island-model/models of segregation. Wakeley (2001) investigated the coalescence in an island model of population subdivision (with variation amongst Demes/groups).

Wilkinson-Herbots 1998: The structured coalescent is used to calculate some quantities relating to the genealogy of a pair of homologous genes and to the degree of subpopulation differentiation, under a range of models of subdivided populations and assuming the infinite alleles model of neutral mutation. In this paper they list theoretical values of coalescence times for a range of models of population structure.

Wilkinson-Herbots 1998: If $c_1 = c_2$ (note $c = c_1 + c_2$ where the rate of coalescence between any two genes in sub-population i is $\frac{1}{c_i}$) "c can be considered as the relative population size". In this paper we have listed the theoretical values of coalescence times and FST for a range of models of population structure. They show that in

the two-population model, the mean coalescence time of a pair of genes from the larger subpopulation is larger than c (the total population size) (is this measured in generations?), while the mean coalescence time of two genes from the smaller subpopulation is smaller than c.

Wakeley 2001: The structured coalescent provides a basis for studying the genealogical history of a sample from a population that is divided into discrete demes (groups) among which there is some pattern of migration. Wilkinson-Herbots (1998) gives an excellent discussion of this model, which is a generalization of Kingman's (1982a, b) "n-coalescent" (Wakeley, 2001).

Alcala et al. 2019: In this study, we investigate all small network motifs: the set of all possible migration networks among populations subdivided into at most four subpopulations. For each motif, we use coalescent theory to derive expectations for three quantities that describe genetic variation: nucleotide diversity, FST (a measure of population differentiation due to genetic structure), and half-time to equilibrium diversity. So three main quantities that people are interested in when looking at the island model are nucleotide diversity, FST and half-time to equilibrium distribution?

3 Embedding it in a Markov jump process

The time epochs in betwen coalescent events are exponentially distributed units of time, denoted $T_k \sim \text{Exp}(\lambda_k)$. For the case of the Kingman's coalescent, we have that any two genes merge (to find a common ancestor) at rate $T_k \sim \text{Exp}(\binom{k}{2})$. This process can hence be represented as a continuous time Markov jump process, where the jump rates are T_k, and the jump from state $n-1$ to n represents the jump into our absorption state whereby the n initial genes have found their common ancestor n generations back in time. The underlying CTMC can only transition forward from state k to state $(k+1)$, for all $k = 1, 2, \ldots, n-1$, since we are counting the tree height one generation at a time. This is why the time until absorption is the sum of the individual occupation times, T_k, for $k = n, n-1, \ldots, 2$. The transition matrix for this CTMC is represented as

$$
A = \begin{pmatrix}
-\binom{n}{2} & \binom{n}{2} & & & & \\
& -\binom{n-1}{2} & \binom{n-1}{2} & & & \\
& & -\binom{n-2}{2} & \binom{n-2}{2} & & \\
& & & \ddots & \ddots & \\
& & & & & -\binom{2}{2}
\end{pmatrix}.
\tag{1}
$$

We are interested in the time until extinction, T_n, representing the time at which we find the MRCA of the n initial genes. See this citation for the derivation from

the discrete-time coalescence.

We will now discuss more thoroughly the intensities in (1). Recall that the time between coalescent events are exponentially distributed with a rate which depends on the number of genes in the current population. For the case of the Kingman's coalescent, we have that $T_k \sim \text{Exp}\left(\binom{k}{2}\right)$. That is, any 2 of the k genes in the current population can coalesce. Thus, we have

1. The first exit intensity is
$$\binom{n}{2} = \frac{n(n-1)}{2},$$
representing the rate at which the first event of coalescence occurs whereby the population decreases from n to $n-1$.

2. Similarly, the second exit intensity is
$$\binom{n-1}{2} = \frac{(n-1)(n-2)}{2},$$
representing the rate at which the second event of coalescence occurs, whereby the population size decreases from $n-1$ to $n-2$.

$\vdots$

k. In general, the kth event of coalescence occurs at rate
$$\binom{n-k+1}{2} = \frac{(n-k+1)(n-k)}{2},$$
which represents the rate at which the population decreases from size $n-k+1$ to $n-k$.

3.1   The 2-island model

We are interested in a 2-island population whereby the initial population is separated into two distinct populations whereby migration can occur. A continuous-time Markov jump process, $\{X_t, t \geq 0\}$, is used to describe the genealogical process in such a model. We suppose that $i = I_1 + i_2$ represents the total population size, where I_1 and I_2 are the number of genes on both I_1 and I_2 at any given time, respectively. We consider a population whereby genes can migrate and events of coalescence occur on each island according to the dynamics of Kingman's coalescent (cite). The transition rate for any event is

$$\lambda_{i,I_2} = \binom{i-I_2}{2}\alpha + \binom{I_2}{2}\beta + (i-I_2)M + I_2 R. \tag{2}$$

Suppose we begin with population size $i = n$. The first event, whether it be an event of migration or coalescence, occurs at rate

$$\lambda_{n,I_2} = \binom{n-I_2}{2}\alpha + \binom{I_2}{2}\beta + (n-I_2)M + I_2 R,$$

where

- any two of the $n - I_2$ genes on I_1 will coalesce with probability

$$\frac{\binom{n-I_2}{2}}{\lambda_{n,I_2}}$$

such that $n - I_2 \geq 2$,

- any two of the I_2 genes on I_2 will coalesce with probability

$$\frac{\binom{I_2}{2}}{\lambda_{n,I_2}}$$

- one of the genes on Island 1 will migrate to Island 2 with probability

$$\frac{(n - I_2)M}{\lambda_{n,I_2}}, and$$

- one of the genes on Island 2 will migrate to Island 1 with probability

$$\frac{(I_2 R)}{\lambda_{n,I_2}}.$$

We denote the states of this CTMC as $S = (I_1, I_2)$ and show these possible states in the table:

Now, to construct our sub-intensity matrix for tree height, we firstly define the $(i + 1) \times (i + 1)$ matrix

$$\Gamma(i) = \begin{pmatrix} -\lambda_{i,0} & iM & & & & \\ R & -\lambda_{i,1} & (i-1)M & & & \\ & 2R & -\lambda_{i,2} & (i-2)M & & \\ & & \ddots & \ddots & \ddots & \\ & & & & -\lambda_{i,i-1} & M \\ & & & & iR & -\lambda_{i,i} \end{pmatrix} \tag{3}$$
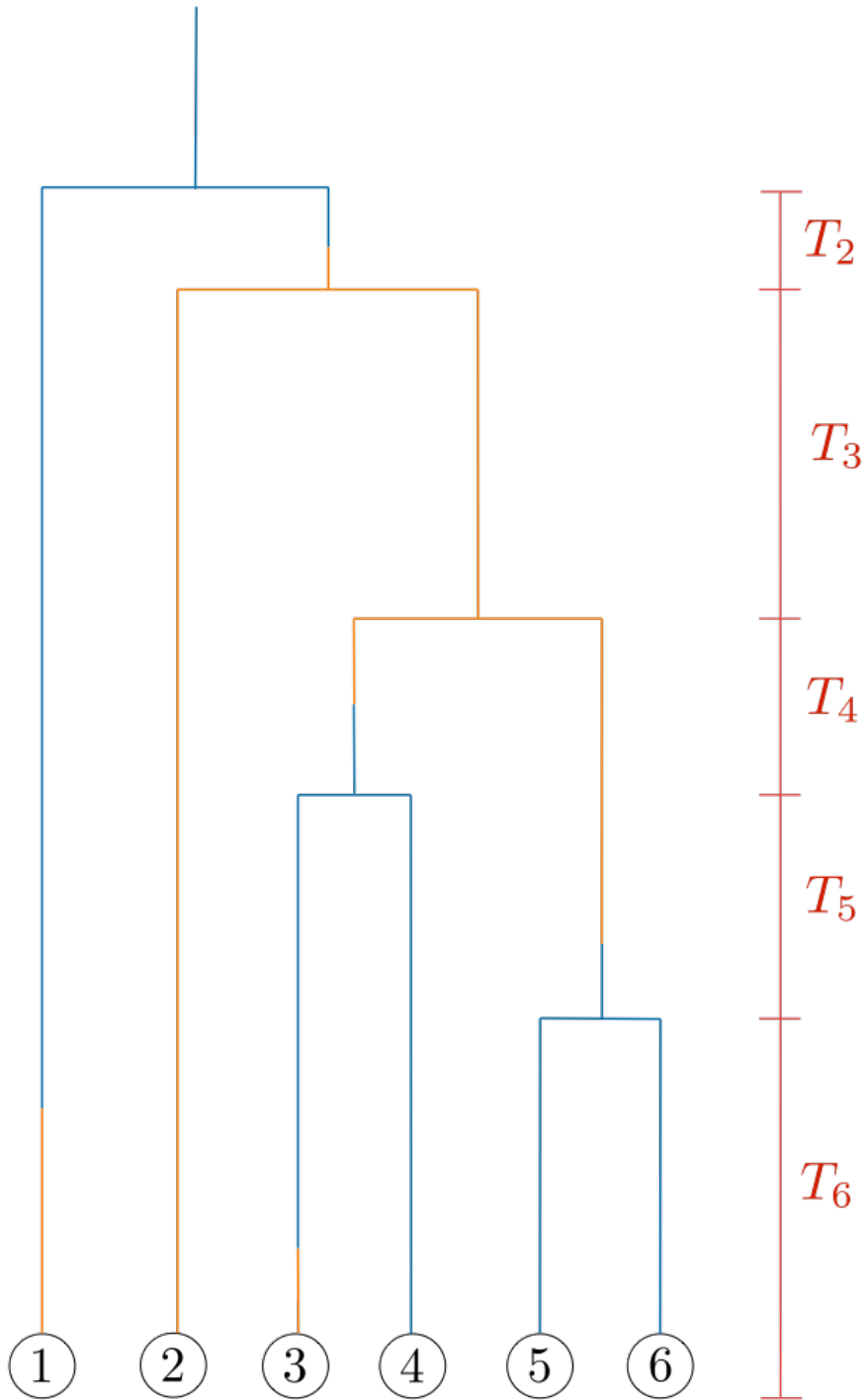
Figure 2: A coalescent tree representing a genealogical process described above.......
Initially, $i = 6$, $I_1 = 3$ and $I_2 = 3$.

Population size	State
$i = n$	$(n, 0)$
	$(n-1, 1)$
	$(n-2, 2)$
	$\vdots$
	$(1, n-1)$
	$(0, n)$
$i = n-1$	$(n-1, 1)$
	$(n-2, 2)$
	$(n-3, 3)$
	$\vdots$
	$(2, n-2)$
	$(1, n-1)$
$i = n-2$	$(n-2, 2)$
	$(n-3, 3)$
	$(n-4, 4)$
	$\vdots$
	$(3, n-3)$
	$(2, n-2)$
	$\vdots$
	$\vdots$
$i = 2$	$(2, 0)$
	$(1, 1)$
	$(0, 2)$

Table 1: States of the two-island model and their corresponding rewards.

along with the $(i+1) \times i$ matrix

$$
D(i) = \begin{pmatrix}
\alpha\binom{i}{2} & & & & & \\
& \alpha\binom{i-1}{2} & & & & \\
& \beta\binom{2}{2} & \alpha\binom{i-2}{2} & & & \\
& & \beta\binom{3}{2} & \alpha\binom{i-3}{2} & & \\
& & & \beta\binom{4}{2} & \ddots & \\
& & & & \ddots & \\
& & & & & 0
\end{pmatrix} \tag{4}
$$

where the parameters M, R, α and β are as defined above. We fit our $\Gamma(i)$ and $D(i)$ matrices together to represent the sub-intensity matrix for tree height in the same

way as for the seed-bank coalescent. We thus have that

$$
B(i) = \begin{pmatrix}
\Gamma(i) & D(i) & & & & \\
& \Gamma(i-1) & D(i-1) & & & \\
& & \Gamma(i-2) & D(i-2) & & \\
& & & \ddots & \ddots & \\
& & & & & D(3) \\
& & & & & \Gamma(2)
\end{pmatrix}
$$

where

(1) Row 1 corresponds to the states whereby there are $i = n$ genes in the total population,

(2) Row 2 corresponds to the states whereby there are $i = n - 1$ genes in the total population, and

$$\vdots$$

(k) Row k corresponds to the state whereby there are $i = n - k$ genes in the total population.

References

Arredondo, A., Mourato, B., Nguyen, K., Boitard, S., Rodríguez, W., Noûs, C., ...
Chikhi, L. (2021). Inferring number of populations and changes in connectivity
under the n-island model. *Heredity*, 1–17.

Hein, J., Schierup, M., & Wiuf, C. (2004). *Gene genealogies, variation and evolution:
a primer in coalescent theory.* Oxford University Press, USA.

Slatkin, M. (1985). Gene flow in natural populations. *Annual review of ecology and
systematics*, *16*(1), 393–430.

Wakeley, J. (2001). The coalescent in an island model of population subdivision
with variation among demes. *Theoretical population biology*, *59*(2), 133–144.

Wright, S. (1943). Isolation by distance. *Genetics*, *28*(2), 114.