0.1 Singlet Model of SiCloneFit

0.1.1 Model Description

We assume that we have measurements from m single cells. For each cell, n somatic single nucleotide variant count have been measured.. We denote the variant-by-cell matrix of alternate read counts by $D_{n\times m}$ and the variant-by-cell matrix of total read counts(sum of reference and alternate read counts) by $N_{n\times m}$. Entries in D and N matrices are non-negative integers, with missing entries in the matrix N indicating zero read coverage for a given cell and variant. Let g_t be the set of possible true genotype values for the SNVs, and g_o be the set of observable values for the SNVs. For binary measurements for SNVs, $g_t = \{0, 1\}$, whereas $g_o = \{0, 1, X\}$, where 0, 1 and X denote the absence of mutation, presence of mutation, and missing value respectively. If ternary measurements are available for SNVs, $g_t = \{0, 1, 2\}$ and $g_o = \{0, 1, 2, X\}$, where 0 denotes homozygous reference genotype, 1 and 2 denote heterozygous, and homozygous non-reference genotypes, respectively, and X denotes missing data.

We assume that there is a set of K clonal populations from which m single cells are sampled and the clonal populations can be placed at the leaves of a clonal phylogeny, $\mathcal{T}$. Each clonal population consists of a set of cells that have identical genotype (with respect to the set of mutations in consideration) and a common ancestor. The genotype vector associated with a clone c is called clonal genotype (denoted by G_c) and it records the genotype values for all n sites for the corresponding clone. The true genotype vector of each cell is identical to the clonal genotype of the clonal population where it belongs to. The clonal genotype matrix, $G_{K\times n}$, represents the clonal genotypes of K clones. It is important to note that, K, the number of clones is unknown. To automatically infer the number of clones and assign the cells to clones, we introduce a tree-structured infinite mixture model. [**?**] describes a nonparametric Bayesian prior over trees similar to mixture models using a Chinese restaurant process (CRP) [**?**] prior. For this tree-structured CRP, each node of the tree represents a cluster. In our model, we extend this idea to define a nonparametric Bayesian prior over binary trees, leaves of which represent the mixture components (clonal clusters). A Chinese restaurant process defines a distribution for partitioning customers

into different tables. In our problem, single cells are analogous to customers and clonal clusters are analogous to tables. Let c_j denote the cluster assignment for cell j and assume that cells $1:j-1$ have already been assigned to clonal clusters $\{1,\ldots,|\boldsymbol{c}_{1:j-1}|\}$, where $|\boldsymbol{c}_{1:j-1}|$ denotes the number of clusters induced by the cluster indicators of $j-1$ cells. The cluster assignment of cell j, c_j is based on the distribution defined by a Chinese restaurant process is given by

$$\begin{aligned} p(c_j = c|c_{1:(j-1)},\alpha_0) &= \frac{n_c}{j-1+\alpha_0} \\ p(c_j \neq c_k \forall k < j|c_{1:(j-1)},\alpha_0) &= \frac{\alpha_0}{j-1+\alpha_0} \end{aligned} \tag{1}$$

where n_c denotes the number of cells already assigned (excluding cell j) to cluster c. α_0 is the concentration parameter for the CRP model.

The clonal phylogeny, $\mathcal{T}$, is a rooted directed binary tree whose number of leaves is equal to the number of clonal clusters, $K = |\boldsymbol{c}|$ defined by the assignment of m cells to different clusters by the CRP. The root of $\mathcal{T}$ represents normal (unmutated) genotype and somatic mutations are accumulated along the branches of the phylogeny. Each leaf in the clonal phylogeny corresponds to a clonal cluster, $c \in \{1,\ldots,K\}$ and is associated with a clonal genotype G_c that records the set of mutations accumulated along the branches from the root. To model the evolution of the clonal genotypes, we employ a finite-site model of evolution, $\mathcal{M}_\lambda$, that accounts for the effects of point mutations, deletion and loss of heterozygosity on the clonal genotypes. The model of evolution assigns transition probabilities to different genotype transitions along the branches of the clonal phylogeny. The true genotype of each cell is identical to the clonal genotype of the clonal cluster where it is assigned. However, observed genotypes of single cells differ from their true genotype due to amplification errors introduced during the single-cell sequencing work flow. The effect of amplification errors is modeled using an error model distribution parameterized by FP error rate, α and FN error rates, β. Here β is different and independent for every mutation. The generative process can be described as follows:

1. Draw $\alpha_0 \sim Gamma(a,b)$, $\alpha \sim Beta(a_\alpha, b_\alpha)$ and for $i \in \{1,2,\ldots,n\}$, $\beta_i \sim Beta(a_\beta, b_\beta)$

2. For $j \in \{1, 2, \ldots, m\}$, draw $c_j \sim CRP(\alpha_0)$.
   From this, derive $K = |\boldsymbol{c}|$, the total number of clusters (or clones) implicitly defined by $\boldsymbol{c}$.

3. draw $\mathcal{T} \sim T_{prior}(K)$.

4. For $\lambda \in \mathcal{M}_\lambda$, draw $\lambda \sim Beta(a_{M_\lambda}, b_{M_\lambda})$

5. For $k \in \{1, 2, \ldots, K\}$, draw $G_k \sim F(G_k|\mathcal{T}, \mathcal{M}_\lambda)$.

6. For $j \in \{1, 2, \ldots, m\}$ and $i \in \{1, 2, \ldots, n\}$, draw $D_{ij} \sim E(D_{ij}|G_{c_j i}, N_{ij}, \alpha, \beta_i)$.

$\boldsymbol{c}$ denotes the clonal assignments of all cells. T_{prior} is the prior distribution on phylogenetic trees for a fixed number of leaves. $\mathcal{M}_\lambda$ denotes the set of parameters in the finite-sites model of evolution. F denotes a distribution on the genotypes at the leaves of a phylogenetic tree and can be computed using Felsenstein's pruning algorithm [**?**] given the phylogeny and a finite-site model of evolution. E is the error model distribution that relates the observed alternate read count at locus i for cell j, D_{ij} to clonal genotype $G_{c_j i}$ and to N_{ij}, the total read count at locus i for cell j . $a, b, a_\alpha, b_\alpha, a_\beta, b_\beta, a_M, b_M$ denote different hyperparameters used in this model.

0.1.2 Model of Evolution

To capture the effect of point mutations, LOH and deletion on the clonal genotypes along the branches of clonal phylogeny, we employ a finite-site model of evolution similar to the one introduced in SiFit [**?**]. The finite-site model of evolution, $\mathcal{M}_\lambda$, is modeled using a continuous-time Markov chain that assigns a probability with each possible transition of genotypes. The branches of clonal phylogeny $\mathcal{T}$, have associated branch lengths that represent expected number of mutations per locus. We assume that the genomic loci evolve identically and independently. For ternary genotype, $g_t = \{0, 1, 2\}$, a 3×3 transition probability matrix describes the model of evolution. The transition probability matrix, P_t, along a branch of length t is given by $P_t = exp(Qt)$, where, Q denotes the transition rate matrix of the Markov chain. The transition rate matrix consists of the infinitesimal rates (during infinitesimally small time, Δt) for switching between genotype states for the

continuous-time Markov chain. As in SiFit, we assume that only one event can occur at a site during Δt, the smallest unit of time. The parameter λ_r accounts for the effect of recurrent mutation and the parameter λ_l captures mutation loss due to deletion and LOH. The product of the transition rate matrix and the branch length (t) is given by:

$$Qt = \begin{bmatrix} -t & t & 0 \\ \frac{(\lambda_r+\lambda_l)\times t}{2} & -(\lambda_r+\lambda_l)\times t & \frac{(\lambda_r+\lambda_l)\times t}{2} \\ 0 & \lambda_r \times t & -\lambda_r \times t \end{bmatrix} \tag{2}$$

In Eq. (2), $Qt(i,j)$ denotes the rate of genotype i changing to genotype j along a branch of length t, $i,j \in \{0,1,2\}$. We assume that the parameters λ_r and λ_l are Beta distributed as they represent relative rates with value between 0 and 1. $P_t(i,j)$ denotes the probabality of transition of genotype i to genotype j along a branch of length t. Each entry of P_t is a function of t, λ_r and λ_l.

For binary genotype states, the product of transition rate matrix and branch length is given by:

$$Qt = \begin{bmatrix} -t & t \\ \frac{(\lambda_r+\lambda_l)\times t}{2} & -\frac{(\lambda_r+\lambda_l)\times t}{2} \end{bmatrix} \tag{3}$$

0.1.3 Single-cell Error Model

The FP and FN errors in single-cell SNV profiles have been modeled using two parameters α and β respectively as in SiFit [**?**]. The error model distribution, $E(D_{ij}|G_{c_j i}, N_{ij}, \alpha, \beta_i)$, gives the probability of observing alternate genotype for locus i in cell j, given the true clonal genotype $G_{c_j i}$ and **??** shows it for ternary genotype. **??** shows the error model distribution for binary genotype. α and β_i are assumed to be Beta distributed variables as they represent probability of FP and FN errors respectively.

0.1.4 Posterior Distribution

The SiCloneFit model has several hidden variables as well as some observed variables. The posterior distribution, $\mathcal{P}$ over the latent variables is given by

$$\begin{aligned}
\mathcal{P}(\boldsymbol{c}, \boldsymbol{G}, \mathcal{T}, \mathcal{M}_\lambda, \alpha, \boldsymbol{\beta}, \alpha_0 | \boldsymbol{D}, \boldsymbol{N}, a_\alpha, b_\alpha, a_\beta, b_\beta, a_M, b_M, a, b) \propto \\
P(\boldsymbol{D}|\boldsymbol{N}, \boldsymbol{c}, \boldsymbol{G}, \mathcal{T}, \mathcal{M}_\lambda, \alpha, \boldsymbol{\beta}, \alpha_0, a_\alpha, b_\alpha, a_\beta, b_\beta, a_M, b_M, a, b) \times \\
P(\boldsymbol{c}, \boldsymbol{G}, \mathcal{T}, \mathcal{M}_\lambda, \alpha, \boldsymbol{\beta}, \alpha_0 | a_\alpha, b_\alpha, a_\beta, b_\beta, a_M, b_M, a, b) \\
= E(\boldsymbol{D}|\boldsymbol{N}, \boldsymbol{c}, \boldsymbol{G}, \alpha, \boldsymbol{\beta}) F(\boldsymbol{G}|\mathcal{T}, \mathcal{M}_\lambda) P(\boldsymbol{c}|\alpha_0) P(\mathcal{T}) \\
P(\alpha|a_\alpha, b_\alpha) P(\beta|a_\beta, b_\beta) P(\mathcal{M}_\lambda|a_M, b_M) P(\alpha_0|a, b) \quad (4)
\end{aligned}$$

The hidden variables that we want to estimate from this model are

1. $\boldsymbol{c}$, a vector containing the cluster assignment for all cells,
2. $\boldsymbol{G}$, a $K \times n$ clonal genotype matrix, where G_k denotes the genotype of clone k, $K = |\boldsymbol{c}|$, the number of clusters defined by $\boldsymbol{c}$,
3. $\mathcal{T}$, the clonal phylogeny, representing the genealogical relationships between the clones,
4. $\mathcal{M}_\lambda$, parameters of the model of evolution,
5. α, false positive rate, and
6. $\boldsymbol{\beta}$, false negative rate for each mutation.

The number of clones is implicitly defined by the vector $\boldsymbol{c}$. The posterior probability is a product of likelihood function and prior. These are described in the following.

0.1.5 Likelihood Function

The likelihood function employed by SiCloneFit is given by

$$P(\boldsymbol{D}|\boldsymbol{N},\boldsymbol{c},\boldsymbol{G},\mathcal{T},\mathcal{M}_\lambda,\alpha,\boldsymbol{\beta},\alpha_0,a_\alpha,b_\alpha,a_\beta,b_\beta,a_M,b_M,a,b) = E(\boldsymbol{D}|\boldsymbol{N},\boldsymbol{c},\boldsymbol{G},\alpha,\boldsymbol{\beta}) = \prod_{i=1}^{n}\prod_{j=1}^{m} E(D_{ij}|N_{ij},G_{c_j i},\alpha,\beta_i) \tag{5}$$

In Eq. (5), $E(D_{ij}|N_{ij},G_{c_j i},\alpha,\beta_i)$ is obtained from the error model distribution for binary and ternary genotype as defined in **??** and **??** respectively.

0.1.6 Prior Distributions

The SiCloneFit model incorporates a compound prior given by

$$P(\boldsymbol{c},\boldsymbol{G},\mathcal{T},\mathcal{M}_\lambda,\alpha,\boldsymbol{\beta},\alpha_0|a_\alpha,b_\alpha,a_\beta,b_\beta,a_M,b_M,a,b) = F(\boldsymbol{G}|\mathcal{T},\mathcal{M}_\lambda)P(\boldsymbol{c}|\alpha_0)P(\mathcal{T})P(\alpha|a_\alpha,b_\alpha)P(\boldsymbol{\beta}|a_\beta,b_\beta)P(\mathcal{M}_\lambda|a_M,b_M)P(\alpha_0|a,b) \tag{6}$$

Below we describe each prior distribution.

0.1.6.1 Prior on Clonal Genotypes

$F(\boldsymbol{G}|\mathcal{T},\mathcal{M}_\lambda)$ denotes the prior distribution on the clonal genotype matrix keeping the clonal phylogeny and parameters of model of evolution fixed. $F(\boldsymbol{G}|\mathcal{T},\mathcal{M}_\lambda)$ can be efficiently calculated using Felsenstein's pruning algorithm [**?**] as

$$F(\boldsymbol{G}|\mathcal{T},\mathcal{M}_\lambda) = \prod_{i=1}^{n} F(\boldsymbol{G}_{*i}|\mathcal{T},\mathcal{M}_\lambda) \tag{7}$$

Here, $\boldsymbol{G}_{*i}$ denotes the genotype of all clones at i^{th} site. The prior probability for site i, $F(\boldsymbol{G}_{*i}|\mathcal{T},\mathcal{M}_\lambda)$ is given by the partial likelihood of the root r of clonal phylogeny $\mathcal{T}$ for genotype 0 and is computed using Felsenstein's pruning algorithm, a dynamic programming on clonal phylogeny that marginalizes over all possible mutational histories along the branches of the phylogeny.

0.1.6.2 Prior on Partition of Cells into Clonal Clusters

$P(\boldsymbol{c}|\alpha_0)$ denotes the prior probability of partitioning m single cells into $|\boldsymbol{c}|$ clusters under a CRP with concentration parameter α_0 and is given by

$$P(\boldsymbol{c}|\alpha_0) = \frac{\Gamma(\alpha_0)\alpha_0^{|\boldsymbol{c}|}}{\Gamma(\alpha_0 + m)} \prod_{k \in \boldsymbol{c}} \Gamma(n_k) \tag{8}$$

In Eq. (8), Γ denotes Gamma function, which is defined as $\Gamma(N) = (N-1)!$ for a positive integer N. n_k denotes the number of cells assigned to a clonal cluster k in the current cluster assignment $\boldsymbol{c}$.

0.1.6.3 Prior on Phylogeny

$P(\mathcal{T})$ denotes the prior probability on the clonal phylogeny. This is a product of prior on topology and prior on branch length. We consider uniform distribution for the prior on topology and exponential distribution for the prior on branch lengths. The overall prior probability for the branches is given by a product over the branches in the phylogeny.

0.1.6.4 Prior on Other Parameters

The values of the parameters α, $\boldsymbol{\beta}$, $\mathcal{M}_\lambda = \{\lambda_r, \lambda_l\}$ lie between 0 and 1. So, we use Beta prior for these parameters. The hyper parameters for α and $\boldsymbol{\beta}$ are computed from the mean and standard deviation of these prior distribution and are kept fixed. The mean is computed from a simple estimation of α and β from the observed genotype matrix assuming usual rate for these parameters and wide standard deviation is used to cover a wide range of values.

For the concentration parameter α_0, we assume a Gamma prior as suggested in [**?**]. We set the value of hyper parameters for the Gamma distribution to $a = 1, b = 1$ for all the analyses performed, but this is also a configurable parameter in the software.

0.1.7 Error Model Distribution

The calculation for the error model distribution $E(\boldsymbol{D}|\boldsymbol{N}, \boldsymbol{c}, \boldsymbol{G}, \alpha, \boldsymbol{\beta})$ is done using a binomial model. For a given site in a given cell, there are two possibilities: the variant is

"absent" in the clone i.e $G_{c_j i} = 0$ or the variant is "present" in the clone i.e. $G_{c_j i} = 1$ denoted by the modelled matrix $\boldsymbol{G}$. When considering the "success probability" for the binomial model, where here a success is defined as observing an alternate read, we consider two alternative (sets of) parameters for each of these settings: α for the cases of false positive(variant absence) and $\beta = \beta_1, \beta_2..., \beta_N$ for the cases of false negative(variant allele present). Therefore, the error model distribution combining the two cases can be written in the following binomial distributions:-

$$E(D_{ij}|N_{ij}, c_j, G_{c_j i}, \alpha, \beta_i) = \begin{cases} Binom(D_{ij}|N_{ij}, \alpha), & if G_{c_j i} = 0 \\ Binom(D_{ij}|N_{ij}, \beta_i), & if G_{c_j i} = 1 \end{cases} \tag{9}$$
$$= Binom(D_{ij}|N_{ij}, \beta_i)^{G_{c_j i}} \times Binom(D_{ij}|N_{ij}, \alpha)^{1-G_{c_j i}}$$

where $G_{c_j i}$ denotes the variant presence of i^{th} mutation of the clonal genotype of cell j^{th}.

0.1.8 Inference

As analytically computing the posterior distribution given by Eq. (4) is computationally intractable, we implemented a Markov chain Monte Carlo (MCMC) sampling procedure based on the Gibbs sampling algorithm. Different classes of Gibbs sampling algorithm have been designed to infer from infinite mixture models based on conjugate as well as non-conjugate prior distributions [**?**, **?**]. Our algorithm is inspired by a partial Metropolis-Hastings partial Gibbs Sampling algorithm described in [**?**]. In our case, while performing the partial Metropolis-Hastings steps, the dimensionality of the sample may change due to addition of a new cluster (resulting in addition of new edge in the clonal phylogeny) or removal of an existing singleton cluster (resulting in removal of edges from the clonal phylogeny). In case the dimensionality changes, the absolute value of the determinant of the Jacobian matrix is also taken into account, which results in partial reversible-jump MCMC [**?**] updates. The resulting algorithm is a partial reversible-jump MCMC partial Gibbs sampling algorithm.

Our sampling algorithm samples the hidden variables from their corresponding condi-

tional posterior distributions. In each iteration, it first samples the cluster indices for each cell, then the parameters of the model of evolution and the clonal phylogeny (on a number of leaves equal to the number of clones defined by cluster indices vector) is sampled. After that the clonal genotypes are sampled followed by sampling of α and β. Finally, the concentration parameter α_0 is sampled. The sampling algorithm is outlined below.

0.1.9 Partial Reversible-jump MCMC Partial Gibbs Sampling Algorithm

Given $\alpha_0^{(t-1)}$, $\{c_j^{(t-1)}\}_{j=1}^m$, $\{G_k^{(t-1)}\}_{k=1}^{|\boldsymbol{c}|}$, $\mathcal{T}^{(t-1)}$, $\mathcal{M}_\lambda^{(t-1)}$, $\alpha^{(t-1)}$, and $\{\beta_i^{(t-1)}\}_{i=1}^n$ from the previous iteration, we need to sample a new set of these parameters. $t-1$ denotes the previous iteration.

Set

- $\boldsymbol{c} = \boldsymbol{c}^{(t-1)}$, $\alpha_0 = \alpha_0^{(t-1)}$
- $\boldsymbol{G} = \{G_k^{(t-1)}\}_{k=1}^{|\boldsymbol{c}|}$
- $\mathcal{T} = \mathcal{T}^{(t-1)}$, $\mathcal{M}_\lambda = \mathcal{M}_\lambda^{(t-1)}$
- $\alpha = \alpha^{(t-1)}$, $\beta = \beta^{(t-1)}$

Sample cluster indicators:

1. For $j = 1, \ldots, m$, update c_j as follows:
   - If c_j is not a singleton (i.e., $c_j = c_l$ for some $l \neq j$)
     (a) let c_j^* be a newly created clone.
     (b) propose a new clonal tree, $\mathcal{T}^* \sim q_T(\mathcal{T}^*|\mathcal{T})$, by adding the new clone c_j^* to $\mathcal{T}$. q_T is the proposal distribution that adds a new leaf to the clonal phylogeny.
     (c) Sample genotype vector for the new clone, $G_{c_j^*} \sim \mathcal{F}(G_{c_j^*}|\mathcal{T}^*, \boldsymbol{G}^*_{\setminus c_j^*}, \mathcal{M}_\lambda)$. $\boldsymbol{G}^*_{\setminus c_j^*}$ is the clonal genotype matrix excluding the genotype vector for clone c_j^*. New clonal genotype matrix after sampling $G_{c_j^*}$ is denoted by $\boldsymbol{G}^*$.
     (d) compute acceptance ratio $a(c_j^*, c_j)$ as follows:

$$a(c_j^*, c_j) =$$
$$min\left[1, \frac{\alpha_0}{m-1}\frac{E(D[j]|N[j], G_{c_j^*}, \alpha, \boldsymbol{\beta})}{E(D[j]|N[j], G_{c_j}, \alpha, \boldsymbol{\beta})}\frac{F(\boldsymbol{G}^*|\mathcal{T}^*, \mathcal{M}_\lambda)}{F(\boldsymbol{G}|\mathcal{T}, \mathcal{M}_\lambda)}\frac{P(\boldsymbol{c}^*|\alpha_0)}{P(\boldsymbol{c}|\alpha_0)}\frac{T_{prior}(\mathcal{T}^*)}{T_{prior}(\mathcal{T})}\frac{q_T(\mathcal{T}|\mathcal{T}^*)}{q_T(\mathcal{T}^*|\mathcal{T})}J_q\right] \tag{10}$$

J_q is the jacobian. $D[j]$ and $N[j]$ are the j^{th} column of observed genotype count matrices respectively.

(e) Set the new c_j to this c_j^* with probability $a(c_j^*, c_j)$

(f) If new c_j is set to c_j^*,

- Set $\boldsymbol{G} = \boldsymbol{G}^*$, $\mathcal{T} = \mathcal{T}^*$

- Otherwise, when c_j is a singleton,

(a) Sample c_j^* from $\boldsymbol{c}_{-j}$, choosing $c_j^* = c$ with probability $\frac{n_c}{m-1}$.

(b) Propose a new clonal tree, $\mathcal{T}^* \sim q_T(\mathcal{T}^*|\mathcal{T})$, by removing the clone c_j from $\mathcal{T}$.

(c) Propose new clonal genotype matrix $\boldsymbol{G}^*$, by removing G_{c_j} from $\boldsymbol{G}$.

(d) compute acceptance ratio $a(c_j^*, c_j)$ as follows:

$$a(c_j^*, c_j) =$$
$$min\left[1, \frac{m-1}{\alpha_0}\frac{E(D[j]|N[j], G_{c_j^*}, \alpha, \boldsymbol{\beta})}{E(D[j]|N[j], G_{c_j}, \alpha, \boldsymbol{\beta})}\frac{F(\boldsymbol{G}^*|\mathcal{T}^*, M)}{F(\boldsymbol{G}|\mathcal{T}, M)}\frac{P(\boldsymbol{c}^*|\alpha_0)}{P(\boldsymbol{c}|\alpha_0)}\frac{T_{prior}(\mathcal{T}^*)}{T_{prior}(\mathcal{T})}\frac{q_T(\mathcal{T}|\mathcal{T}^*)}{q_T(\mathcal{T}^*|\mathcal{T})}J_q\right] \tag{11}$$

(e) Set the new c_j to this c_j^* with probability $a(c_j^*, c_j)$.

(f) If new c_j is set to c_j^*,

- Set $\boldsymbol{G} = \boldsymbol{G}^*$, $\mathcal{T} = \mathcal{T}^*$

- If the new c_j is not set to c_j^*, it is the same as the old c_j. $\boldsymbol{G}$ and $\mathcal{T}$ remains same.

2. For $j = 1, \ldots, m$, update c_j as follows:

- If c_j is a singleton, do nothing.

- Otherwise, choose a new value for c_j from $\{c_1, \ldots, c_m\}$ using the following probabilities:

$$P(c_j = c|\boldsymbol{c}_{-j}, D[j], N[j], \boldsymbol{G}, \alpha, \boldsymbol{\beta}) \propto \frac{n_c}{m-1} E(D[j]|N[j], G_c, \alpha, \boldsymbol{\beta})$$

Sample clonal phylogeny and evolution model parameters:

Sample new clonal phylogeny $\mathcal{T}^*$ and new set of values for parameters of model of evolution, $\mathcal{M}_\lambda^*$ from the joint conditional posterior distribution, $\mathcal{P}_{\mathcal{T},\mathcal{M}_\lambda}(\mathcal{T}^*, \mathcal{M}_\lambda^*|\mathcal{T}, \mathcal{M}_\lambda, \boldsymbol{G}, a_M, b_M)$

$$\mathcal{T}^*, \mathcal{M}_\lambda^* \sim \mathcal{P}_{\mathcal{T},\mathcal{M}_\lambda}(\mathcal{T}^*, \mathcal{M}_\lambda^*|\mathcal{T}, \mathcal{M}_\lambda, \boldsymbol{G}, a_M, b_M)$$

Sample clonal genotypes:

For $k = 1, ..., |\boldsymbol{c}|$

- Sample clonal genotype G_k for each clone as follows:
  For $i = 1, ..., n$, sample G_{ki} from the following distribution

$$G_{ki} \propto \mathcal{F}(G_{ki}|T, \boldsymbol{G}_{-ki}, M) \times \prod_{j|c_j=k} E(D_{ij}|N_{ij}, G_{ki}, \alpha, \beta_i)$$

Sample error rates:

1. Sample $\alpha \sim \mathcal{P}_\alpha(\alpha|\boldsymbol{D}, \boldsymbol{N}, \boldsymbol{c}, \boldsymbol{G}, \boldsymbol{\beta}, a_\alpha, b_\alpha) \sim E(\boldsymbol{D}|\boldsymbol{N}, \boldsymbol{c}, \boldsymbol{G}, \boldsymbol{\beta}, \alpha)P(\alpha|a_\alpha, b_\alpha)$ using rejection sampling.

2. For $i = 1, ..., n$, sample $\beta_i \sim \mathcal{P}_{\beta_i}(\boldsymbol{\beta}|\boldsymbol{D}, \boldsymbol{N}, \boldsymbol{c}, \boldsymbol{G}, \alpha, a_\beta, b_\beta) \sim E(\boldsymbol{D}|\boldsymbol{N}, \boldsymbol{c}, \boldsymbol{G}, \boldsymbol{\beta}, \alpha)P(\beta_i|a_\beta, b_\beta)$ using rejection sampling.

Sample concentration parameter:

Sample $\alpha_0^t \sim p(\alpha_0|m, |\boldsymbol{c}|, a, b)$ based on the method described in **[?]** assuming the prior distribution for α_0 is $Gamma(a, b)$.

0.1.9.1 Algorithm For Sampling Cluster Indicators

Partial reversible-jump MCMC partial Gibbs updates are used for sampling the cluster indicators for cells as outlined above. In the partial reversible-jump MCMC steps, new clusters are assigned to cells based on an acceptance ratio. The calculation of acceptance ratio involves the calculation of likelihood ratio, prior ratio, proposal ratio and jacobian. Below, we describe how each of these terms are computed.

0.1.9.1.1 Likelihood Ratio

The likelihood ratio, L_r is defined by:

$$L_r = \frac{E(D[j]|N[j], G_{c_j^*}, \alpha, \boldsymbol{\beta})}{E(D[j]|N[j], G_{c_j}, \alpha, \boldsymbol{\beta})} \tag{12}$$

In Eq. (12), c_j^* and c_j are the new and old cluster indicators for cell j respectively. The values in the numerator and the denominator can be calculated by:

$$E(D[j]|N[j], G_{c_j=c}, \alpha, \boldsymbol{\beta}) = \prod_{i=1}^{n} E(D_{ij}|N_{ij}, G_{ci}, \alpha, \beta_i) \tag{13}$$

$E(D_{ij}|N_{ij}, G_{ci}, \alpha, \beta_i)$ is given by the error model distribution as shown in **??** or **??**.

0.1.9.1.2 Prior Ratio

The prior ratio, P_r is given by:

$$P_r = \frac{F(\boldsymbol{G}^*|\mathcal{T}^*, \mathcal{M}_\lambda)}{F(\boldsymbol{G}|\mathcal{T}, \mathcal{M}_\lambda)} \frac{P(\boldsymbol{c}^*|\alpha_0)}{P(\boldsymbol{c}|\alpha_0)} \frac{T_{prior}(\mathcal{T}^*)}{T_{prior}(\mathcal{T})} \tag{14}$$

and is a product of three ratios from three prior distributions. The first ratio, $\frac{F(\boldsymbol{G}^*|\mathcal{T}^*, \mathcal{M}_\lambda)}{F(\boldsymbol{G}|\mathcal{T}, \mathcal{M}_\lambda)}$ can be computed using Eq. (7). The second ratio, $\frac{P(\boldsymbol{c}^*|\alpha_0)}{P(\boldsymbol{c}|\alpha_0)}$ can be computed using Eq. (8). The third ratio is the ratio of prior probabilities on clonal phylogeny. Let us assume, the number of clones based on the new set of cluster indicators is, $|\boldsymbol{c}^*| = K$. For non-singleton cells (i.e., $c_j = c_l$ for some $l \neq j$), when a new leaf is added to the clonal phylogeny, the

third ratio is defined by

$$\frac{T_{prior}(\mathcal{T}^*)}{T_{prior}(\mathcal{T})} = \frac{K-1}{(K-2)(2K-3)} \frac{f(\nu_1)f(\nu_2)f(\nu^*)}{f(\nu_1+\nu_2)} \tag{15}$$

In Eq. (15), ν_1 and ν_2 are the new branch lengths created by adding a new leaf to the branch of length $\nu = \nu_1 + \nu_2$ and ν^* is the branch length assigned to the branch connected to the new leaf. $f(\nu)$ is the edge length prior density evaluated at any branch of length ν. All other edge lengths maintain the same values before and after adding the new leaf, so all other terms in the prior ratio cancel each other.

For singleton cells, when an existing leaf is removed from the clonal phylogeny, the third ratio is defined by

$$\frac{T_{prior}(\mathcal{T}^*)}{T_{prior}(\mathcal{T})} = \frac{(K-1)(2K-1)}{K} \frac{f(\nu_1+\nu_2)}{f(\nu_1)f(\nu_2)f(\nu^*)} \tag{16}$$

In Eq. (16), $\nu = \nu_1 + \nu_2$ is the branch length of the new branch after removing the leaf associated with branch of length ν^*. As a result of the removal of this leaf, two branches of length ν_1 and ν_2 are merged into one branch of length $\nu_1 + \nu_2$. All other edge lengths maintain the same values before and after removal of the leaf, so all other terms in the prior ratio cancel each other. For the distribution on branch lengths (f), we use exponential distribution.

0.1.9.1.3 Proposal Ratio and Jacobian

The proposal or Hastings ratio, Q_r, is given by:

$$Q_r = \frac{q_T(\mathcal{T}|\mathcal{T}^*)}{q_T(\mathcal{T}^*|\mathcal{T})} \tag{17}$$

where q_T is the proposal distribution. We have two moves corresponding to adding a new leaf and removing an existing leaf respectively. The moves and their corresponding proposal ratio are described below.

Add Clone: This move is performed when a new clonal cluster is created for a cell. This results in adding a new leaf to the existing clonal phylogeny and the new leaf corresponds to the new cluster. As a result, this move adds new parameters to the model. One branch of the existing phylogeny is chosen at random. Let us assume that the length of the chosen branch is ν. A new node is created on this branch which serves as the parent of the new clone/leaf to be added. As a result, the existing branch gets divided into two new branches of lengths ν_1 and ν_2. To choose the lengths of these new branches, we generate a uniformly random number, w_1 between 0 and 1, $w_1 \sim U(0,1)$ and the branch lengths are set as $\nu_1 = \nu * w_1$ and $\nu_2 = \nu * (1 - w_1)$. To propose the length of the branch that connects the new leaf to its parent, we generate another uniform random number, $w_2 \sim U(0,1)$ and it is transformed into a random deviate from the edge length prior distribution, $\nu^* = -\frac{1}{\theta}\ln(1 - w_2)$.

The Hastings ratio for adding a new clone to the clonal phylogeny is the probability of proposing a remove clone move that exactly reverses the proposed add clone move, divided by the probability of proposing the add clone move itself. Proposing an add clone move involves the following steps:

1. Choose to perform the add clone move
2. Choose an existing branch of the phylogeny
3. Divide the branch into two branches
4. Choose a length for the newly created edge

The probability of the first step is $\frac{\alpha_0}{m+\alpha_0-1}$, as the new clone is created with this probability. The probability of the second step is $\frac{1}{n_e}$, where n_e is the number of branches in the phylogeny before the move. If we assume that the number of clones based on the new set of cluster indicators is, $|\boldsymbol{c}^*| = K$, then $n_e = 2K - 3$. To divide the branch into two branches, we generate a uniform random variate w_1, so the third step has no effect on the probability of Add Clone move because the value w_1 has Uniform probability density 1.0, similarly the fourth move does not have any effect on the probability of Add Clone move as we generate another uniform random deviate w_2.

Proposing the corresponding Remove Clone move involves two steps:

1. Choose to perform Remove Clone move

2. Choose the leaf in the phylogeny to remove to restore the phylogeny that existed before the Add Clone move.

The probability of the first step is $\frac{1}{m+\alpha_0-1}$, as size of the new clone is 1. The probability of the second step is $\frac{1}{K}$, where K is the number of leaves in the phylogeny after Add Clone move. Therefore, the Hastings ratio is given by:

$$\begin{aligned} \text{Hastings ratio for Add Clone move} &= \frac{(\frac{1}{m+\alpha_0-1})(\frac{1}{K})}{(\frac{\alpha_0}{m+\alpha_0-1})(\frac{1}{2K-3})} \\ &= \frac{2K-3}{\alpha_0 * K} \end{aligned} \tag{18}$$

The Jacobian term for this move is given by:

$$\begin{aligned} J_q &= \begin{vmatrix} \frac{\partial \nu_1}{\partial \nu} & \frac{\partial \nu_1}{\partial w_1} & \frac{\partial \nu_1}{\partial w_2} \\ \frac{\partial \nu_2}{\partial \nu} & \frac{\partial \nu_2}{\partial w_1} & \frac{\partial \nu_2}{\partial w_2} \\ \frac{\partial \nu^*}{\partial \nu} & \frac{\partial \nu^*}{\partial w_1} & \frac{\partial \nu^*}{\partial w_2} \end{vmatrix} \\ &= \begin{vmatrix} w_1 & \nu & 0 \\ 1-w_1 & -\nu & 0 \\ 0 & 0 & \frac{1}{1-w_2} \end{vmatrix} \\ &= \frac{\nu}{1-w_2} \end{aligned} \tag{19}$$

Remove Clone: This move is performed when an existing clonal cluster is removed. This results in removing a leaf from the existing clonal phylogeny. As a result, this move removes some parameters from the model. The leaf to be removed is chosen and removed from the phylogeny, the associated branch is also removed. The parent node of the leaf is also removed, as a result two branches of lengths ν_1 and ν_2 get merged into a single branch of length $\nu = \nu_1 + \nu_2$.

Hastings ratio for the Remove Clone move is given by the probability of proposing an Add Clone move divided by the probability of the Remove Clone move and is calculated

as follows:

$$\begin{aligned}\text{Hastings ratio for Remove Clone move} &= \frac{(\frac{\alpha_0}{m+\alpha_0-1})(\frac{1}{2K-1})}{(\frac{1}{m+\alpha_0-1})(\frac{1}{K+1})} \\ &= \frac{\alpha_0 * (K+1)}{2K-1}\end{aligned} \tag{20}$$

The Jacobian term for this move is given by:

$$\begin{aligned}J_q &= \begin{vmatrix} \frac{\partial \nu}{\partial \nu_1} & \frac{\partial \nu}{\partial \nu_2} & \frac{\partial \nu}{\partial \nu^*} \\ \frac{\partial w_1}{\partial \nu_1} & \frac{\partial w_1}{\partial \nu_2} & \frac{\partial w_1}{\partial \nu^*} \\ \frac{\partial w_2}{\partial \nu_1} & \frac{\partial w_2}{\partial \nu_2} & \frac{\partial w_2}{\partial \nu^*} \end{vmatrix} \\ &= \begin{vmatrix} 1 & 1 & 0 \\ \frac{1}{\nu} & -\frac{1}{\nu} & 0 \\ 0 & 0 & e^{-\nu^*} \end{vmatrix} \\ &= \frac{e^{-\nu^*}}{\nu}\end{aligned} \tag{21}$$

0.1.9.2 Algorithm For Sampling Clonal Phylogeny and Evolution Model Parameters

We designed a Metropolis-Hastings [**?**] sampler for sampling the clonal phylogeny and evolution model parameters from the joint conditional posterior given by:

$$\mathcal{P}_{\mathcal{T},\mathcal{M}_\lambda}(\mathcal{T}^*, \mathcal{M}_\lambda^* | \boldsymbol{G}, a_M, b_M) \propto F(\boldsymbol{G}|\mathcal{T}^*, \mathcal{M}_\lambda^*) p(\mathcal{T}^*) p(\mathcal{M}_\lambda^* | a_M, b_M) \tag{22}$$

We consider two different types of moves to explore the joint $\mathcal{T}, \mathcal{M}_\lambda$ space. In tree changing moves, a new clonal phylogenetic tree, $\mathcal{T}^*$ is proposed from current state $\mathcal{T}$. In parameter changing moves, a new value of the parameter, $\mathcal{M}_\lambda^*$ is proposed from the current parameter value $\mathcal{M}_\lambda$. The proposed configuration is accepted or rejected based on an acceptance ratio. The acceptance ratio for proposing a new clonal phylogenetic tree is given by:

$$\rho_T = min \left\{ 1, \frac{F(\boldsymbol{G}|\mathcal{T}^*, \mathcal{M}_\lambda) p(\mathcal{T}^*) q_T(\mathcal{T}|\mathcal{T}^*)}{F(\boldsymbol{G}|\mathcal{T}, \mathcal{M}_\lambda) p(\mathcal{T}) q_T(\mathcal{T}^*|\mathcal{T})} \right\} \tag{23}$$

In Eq. (23), the likelihood ratio, $\frac{F(\boldsymbol{G}|\mathcal{T}^*,\mathcal{M}_\lambda)}{F(\boldsymbol{G}|\mathcal{T},\mathcal{M}_\lambda)}$ is computed using Felsenstein's pruning algorithm [**?**]. q_T denotes the proposal distribution for proposing a new phylogeny from the current phylogeny. Here, we use a combination of branch change (alter branch lengths) and branch-rearrangement (alter the tree topology) proposals as used in [**?**]. The prior ratio is computed using uniform prior for topology and exponential prior for branch lengths.

The acceptance ratio for proposing a new parameter value is given by:

$$\rho_{\mathcal{M}_\lambda} = min\left\{1, \frac{F(\boldsymbol{G}|\mathcal{T},\mathcal{M}^*_\lambda)p(\mathcal{M}^*_\lambda|a_M,b_M)q_{\mathcal{M}_\lambda}(\mathcal{M}_\lambda|\mathcal{M}^*_\lambda)}{F(\boldsymbol{G}|\mathcal{T},\mathcal{M}_\lambda)p(\mathcal{M}^*_\lambda|a_M,b_M)q_{\mathcal{M}_\lambda}(\mathcal{M}^*_\lambda|\mathcal{M}_\lambda)}\right\} \tag{24}$$

In Eq. (24), the likelihood is calculated in the same way as for Eq. (23). $q_{\mathcal{M}_\lambda}$ is the proposal distribution. The parameters, λ_r and λ_l are beta distributed variables. For each of these parameters, the next value is proposed from a normal distribution centered at the current value. The standard deviation is chosen so that a wide range of values are covered. The algorithm is shown in Algorithm 1.

Algorithm 1: Algorithm for sampling clonal phylogeny and evolution model parameters. $\mathcal{T}^s$ is the starting tree. $\mathcal{M}^s_\lambda$ is the starting value of model parameters. The algorithm runs for n_{iter} iterations. With probability p_λ, model parameters are updated.

Input: $\boldsymbol{G}, \mathcal{T}^s, \mathcal{M}^s_\lambda, n_{iter}, p_\lambda$
Output: $\mathcal{T}^*, \mathcal{M}^*_\lambda$
Initialization: $\mathcal{T}^0 \leftarrow \mathcal{T}^s, \mathcal{M}^0_\lambda \leftarrow \mathcal{M}^s_\lambda$
for $i = 1...n_{iter}$ **do**
  $\mathcal{T} \leftarrow \mathcal{T}^{i-1}, \mathcal{M}_\lambda \leftarrow \mathcal{M}^{i-1}_\lambda$
  Sample $r \sim U(0,1)$
  if $r \leq p_\lambda$ **then**
    Sample $\mathcal{M}'_\lambda \sim q_{\mathcal{M}_\lambda}(\mathcal{M}'_\lambda|\mathcal{M}_\lambda)$
    Compute $\rho_{\mathcal{M}_\lambda} = min\left\{1, \frac{F(\boldsymbol{G}|\mathcal{T},\mathcal{M}'_\lambda)p(\mathcal{M}'_\lambda|a_M,b_M)q_{\mathcal{M}_\lambda}(\mathcal{M}_\lambda|\mathcal{M}'_\lambda)}{F(\boldsymbol{G}|\mathcal{T},\mathcal{M}_\lambda)p(\mathcal{M}'_\lambda|a_M,b_M)q_{\mathcal{M}_\lambda}(\mathcal{M}'_\lambda|\mathcal{M}_\lambda)}\right\}$
    Accept $\mathcal{M}'_\lambda$ with probability $\rho_{\mathcal{M}_\lambda}$
    $\mathcal{M}^i_\lambda \leftarrow \mathcal{M}'_\lambda, \mathcal{T}^i \leftarrow \mathcal{T}$
  else
    Sample $\mathcal{T}' \sim q_T(\mathcal{T}'|\mathcal{T})$
    Compute $\rho_T = min\left\{1, \frac{F(\boldsymbol{G}|\mathcal{T}',\mathcal{M}_\lambda)p(\mathcal{T}')q_T(\mathcal{T}|\mathcal{T}')}{F(\boldsymbol{G}|\mathcal{T},\mathcal{M}_\lambda)p(\mathcal{T})q_T(\mathcal{T}'|\mathcal{T})}\right\}$
    Accept $\mathcal{T}'$ with probability ρ_T
    $\mathcal{M}^i_\lambda \leftarrow \mathcal{M}_\lambda, \mathcal{T}^i \leftarrow \mathcal{T}'$
$\mathcal{T}^* \leftarrow \mathcal{T}^{n_{iter}}, \mathcal{M}^*_\lambda \leftarrow \mathcal{M}^{n_{iter}}_\lambda$
return $\mathcal{T}^*, \mathcal{M}^*_\lambda$

0.1.9.3 Algorithm For Sampling Clonal Genotypes

The genotype of each clone is sampled by keeping the genotypes of other clones fixed. Genotype of each position can be sampled independently. The clonal genotype for clone k, G_k, where $k \in \{1, \ldots, |\boldsymbol{c}|\}$ is sampled from the conditional posterior distribution given by:

$$G_k \sim \mathcal{P}_G(G_k | D_{j|c_j=k}, N_{j|c_j=k}, \boldsymbol{G}_{\setminus k}, \mathcal{T}, \mathcal{M}_\lambda, \alpha, \boldsymbol{\beta}) \tag{25}$$

In Eq. (25), $\boldsymbol{G}_{\setminus k}$ denotes the genotypes of other clones. Clonal genotype G_k is a vector of length n and records the genotype state for n mutation loci. Genotype for locus i is sampled from a categorical distribution defined by

$$G_{ki} \propto \mathcal{F}(G_{ki} | \mathcal{T}, \boldsymbol{G}_{-ki}, \mathcal{M}_\lambda) \times \prod_{j|c_j=k} E(D_{ij} | N_{ij}, G_{ki}, \alpha, \beta_i) \tag{26}$$

For $G_{ki} \in g_t$, $\mathcal{F}(G_{ki} | \mathcal{T}, \boldsymbol{G}_{-ki}, \mathcal{M}_\lambda)$ is calculated using Felsenstein's pruning algorithm and $E(D_{ij} | Ni, j, G_{ki}, \alpha, \beta_i)$ is given by the error model distribution as shown in **??** or **??**.

0.1.9.4 Algorithm For Sampling Error Rates

Rejection sampling [**?**] is used for sampling the value of error rates α and β from their corresponding conditional posterior distributions.

0.1.9.4.1 False Positive Rate

The conditional posterior distribution from which α is sampled, is given by:

$$\alpha \sim \mathcal{P}_\alpha(\alpha | \boldsymbol{D}, \boldsymbol{N}, \boldsymbol{c}, \boldsymbol{G}, \boldsymbol{\beta}, a_\alpha, b_\alpha) \sim E(\boldsymbol{D} | \boldsymbol{N}, \boldsymbol{c}, \boldsymbol{G}, \boldsymbol{\beta}, \alpha) P(\alpha | a_\alpha, b_\alpha) \tag{27}$$

By varying α for a grid of values between 0.001 to 1, we first compute the maximum of the posterior distribution. Based on this maximum value, we create an envelope function for the range of values of α and this serves as the proposal distribution using which we sample a new value of α using rejection sampling.

0.1.9.4.2 False Negative Rate

The conditional posterior distribution from which β_i is sampled, is given by:

$$\beta_i \sim \mathcal{P}_\beta(\boldsymbol{\beta}|\boldsymbol{D}, \boldsymbol{N}, \boldsymbol{c}, \boldsymbol{G}, \alpha, a_\beta, b_\beta) \sim E(\boldsymbol{D}|\boldsymbol{N}, \boldsymbol{c}, \boldsymbol{G}, \boldsymbol{\beta}, \alpha) P(\beta_i|a_\beta, b_\beta) \tag{28}$$

By varying β_i for a grid of values between 0.001 to 1, we first compute the maximum of the posterior distribution. Based on this maximum value, we create an envelope function for the range of values of β_i and this serves as the proposal distribution using which we sample a new value of β using rejection sampling.