Highlights

Gain-loss-duplication models on a phylogeny: exact algorithms for computing the likelihood and its gradient

Miklós Csűrös

- In evolutionary genomics, one of the most useful characteristics of a gene family is its phyletic profile that gives how many copies the family has in extant genomes.

- Birth-and-death processes model copy number evolution by gain (lateral transfer), duplication, and loss within a species phylogeny.

- We decompose the model into a probabilistic network of ancestral and conserved copy numbers.

- The decomposition leads to clean, fast algorithms for ancestral inference and model parameter optimization.

Gain-loss-duplication models on a phylogeny: exact algorithms for computing the likelihood and its gradient

Miklós Csűrös

Department of Computer Science and Operations Research, Université de Montréal;
C.P. 6128 succursale Centre-Ville, Montréal, Québec H3C 3J7, Canada

Abstract

Gene gain-loss-duplication models are commonly based on continuous-time birth-death processes. Employed in a phylogenetic context, such models have been increasingly popular in studies of gene content evolution across multiple genomes. While the applications are becoming more varied and demanding, bioinformatics methods for probabilistic inference on copy numbers (or integer-valued evolutionary characters, in general) are scarce.

We describe a flexible probabilistic framework for phylogenetic gene-loss-duplication models. The framework is based on a novel elementary representation by dependent random variables with well-characterized conditional distributions: binomial, Pólya (negative binomial), and Poisson.

The corresponding graphical model yields exact numerical procedures for computing the likelihood and the posterior distribution of ancestral copy numbers. The resulting algorithms take quadratic time in the total number of copies. In addition, we show how the likelihood gradient can be computed by a linear-time algorithm.

Keywords: genome evolution, gene content, birth-death process, maximum likelihood, phyletic profile

Email address: `csuros@iro.umontreal.ca` (Miklós Csűrös)

1. Introduction

Homology between two genes is the equivalence relation of descent from the same ancestral gene [1], defining the corresponding equivalence classes of gene *families*. During genome annotation, homologies are routinely recognized by sequence similarity, and annotated genes are assigned to families [2]. The *copy number* for a family is the number of family representatives in the genome, a non-negative integer. The *profile* of a family comprises the copy numbers across different genomes. Family profiles are used in evolutionary genomics to infer ancestral gene content [3], and in functional genomics to recognize associations between families [4]. Probabilistic approaches to copy number evolution are based on continuous-time birth-death processes [5]. Such processes are also fundamental in queuing theory [6], epidemiology and population growth models [7]. In the context of genome evolution, the process captures fixation events eliminating a gene ("death") by segmental loss and pseudogenization, or adding a gene ("birth") to the genome, either by duplication within the same genome, or by lateral transfer from another genome.

Bioinformatics problems for copy numbers differ fundamentally from molecular sequence evolution problems with a finite character set, and porting standard methods to an unbounded character domain is generally not possible. A simple, but unsatisfactory fix is to impose a limit on maximum copy number, and usual techniques like Felsenstein's peeling method [8] apply. Indeed, pioneer applications of birth-death processes for gene content evolution by Hahn et al. [9] and by Iwasaki and Takagi [10] employ the same workaround. The first algorithm for computing the profile likelihood for unbounded copy numbers with a gain-loss-duplication model is by Csűrös and Miklós [11, 3], and the corresponding methods are implemented in the software package Count [12], which has been used in hundreds of studies.

The algorithmics of phylogenetic birth-death models is difficult mainly because of unobserved empty profiles (gene families "waiting to be discovered"), and of the intricate transition probabilities in birth-death processes without known closed expressions. We introduce a fresh mathematical framework for linear gain-loss-duplication (and gain-loss without duplication, or duplication-loss without gain) models on a phylogeny. The novel formulation is based on our pivotal Theorem 2 giving the transition probabilities in a closed form that involves only basic discrete distributions. The theorem suggests a fundamental dependency network of random variables along the phylogeny, representing ancestral copy numbers and conservation. The elementary decomposition yields relatively simple algorithms to compute' the likelihood of a family profile (Theorem 4). While the likelihood computation algorithm is a simpler and more insightful version of an existing method (Theorem 17), the network factorization also leads to an algorithm for posterior probabilities of ancestral copy numbers (Theorem 6 and Corollary 7), and to our main result, a hitherto elusive algorithm for exactly computing the gradient of the log-likelihood with respect to model parameters (Corollary 10 and Theorem 11).

2. Theory

A **phylogeny** is a rooted binary tree with nodes numbered $u \in [R] = \{1, 2, \ldots, R\}$. Every node either has two non-null child nodes, or is a terminal node (a **leaf**) with two null children. For ease of notation, we assume that the nodes are indexed respecting postfix order, with every child's index being less than the parent's, so that the last one is the root. The tree is identified by its root R and its edges $T \subset [R] \times [R]$ directed from parent to child. The edges in the subtree rooted at a node u are denoted by T_u, including $T_R = T$. The set of leaves is denoted by $\mathcal{L}$, and the leaf set for T_u by $\mathcal{L}_u$; in particular, $\mathcal{L} = \mathcal{L}_R$. For simplicity, start the indices with the leaves respecting the postfix order, so that $\mathcal{L} = [L]$ and every subset $\mathcal{L}_u$ comprises consecutive integers.

Consider the problem of **copy number evolution**: each node u has an associated random variable ξ_u, called the copy number, taking non-negative integer values, and the joint distribution is determined by dependencies along the phylogeny:

$$\mathbb{P}\{\xi_1 = n_1, \ldots, \xi_R = n_t\} = \mathbb{P}\{\xi_R = n_R\} \prod_{uv \in T} \underbrace{\mathbb{P}\Big\{\xi_v = n_v \ \Big|\ \xi_u = n_u\Big\}}_{\text{transition on edge } uv}. \qquad (1)$$

The leaf variables are observable, corresponding to extant species, forming the **profile** $\Xi = \{\xi_v\}_{v \in \mathcal{L}}$. Non-leaf nodes are (hypothetical) ancestors with unobserved copy numbers. The **ancestral inference** problem is that of estimating $\{\xi_u\}_{u \notin \mathcal{L}}$ for ancestral nodes, knowing the distribution of Eq. (1) and Ξ.

Suppose that we observe the leaf variables across sample profiles called **families** $f = 1, \ldots, F$, with independent and identically distributed (**iid**) copy number vectors $(\xi_{1,1}, \ldots, \xi_{1,R}), \ldots, (\xi_{F,1}, \ldots, \xi_{F,R})$. The **model inference** problem is that of deducing the distribution of (1) from an iid sample $(\Xi_1, \ldots, \Xi_F)$.

A linear birth-death model defines the probabilities $\mathbb{P}\Big\{\xi_v = m \ \Big|\ \xi_u = n\Big\}$ along every edge uv by a continuous-time Markov process $\{\xi(t) \colon 0 \le t \le t_{uv}\}$ via $\xi_u = \xi(0)$ and $\xi_v = \xi(t_{uv})$ during some time $t_{uv} \ge 0$ (the **edge length**). The process is characterized by the constant instantaneous rates for loss $\mu > 0$, duplication $\lambda \ge 0$ and $\kappa \ge 0$, so that $n \to (n-1)$ death events arrive with a rate of μn, and $n \to (n+1)$ birth events arrive either with a rate of $\lambda(n + \kappa)$. In particular, for $p_n(t) = \mathbb{P}\{\xi(t) = n\}$, the Kolmogorov backward equations are

$$p_n'(t) = \{n > 0\}\lambda(n - 1 + \kappa)p_{n-1}(t) + \mu(n+1)p_{n+1}(t) - (\lambda(n + \kappa) + \mu n)p_n(t)$$

with $p_n'(t) = \frac{\partial p_n(t)}{\partial t}$. The no-duplication model is the limit for $\kappa\lambda \to \gamma\mu$ while $\lambda \to 0$:

$$p_n'(t) = \{n > 0\}\mu\gamma p_{n-1}(t) + \mu(n+1)p_{n+1}(t) - \mu(\gamma + n)p_n(t),$$

characterized by loss rate μ and the relative gain rate γ. The model's rate parameters have convenient biological interpretations. The components μ and λ are the per-copy instantaneous rates of loss and duplication. The κ and γ

parameters represent the propensity for gene acquisition from external sources, acting as an *environmental* fractional copy that contributes $\kappa\lambda$ (if $\lambda > 0$) or $\gamma\mu$ (if $\lambda = 0$) to the birth rate. In other words, the constant *gain* rate component serves as an abstraction of horizontal gene transfer from all sources in the organism's environment. In contrast, *duplication* originates from copies within the genome, each contributing λ to the birth rate. Duplication-loss rates can be embedded in a population-genetic model of genome size evolution, so that they are determined by a family-specific selection coefficient, and (constant) population size [13]. The particular case $\mu, \lambda > 0$ and $\kappa = 1$ is the process of gene length evolution in the Thorne-Kishino-Felsenstein model [14], where the *immortal link* plays the same role as the environmental copy here.

The gene copies evolve independently, forming a set of Galton-Watson trees on each edge [11]. (Every copy at the ancestor and every gained copy defines the root of a phylogeny over the copies with time-annotated nodes; loss events create terminal nodes and duplication events create bifurcations.) Our interest lies not in inferring the trees (i.e., in *reconciling* the gene histories with the species phylogeny), but rather in designing a model for the copy numbers without explaining about each copy where they originate.

The transition probabilities for an arbitrary starting value $\xi(0) = n$ are given in Theorem 2. The basic transition probabilities are well understood [7, 15]:

$$h_n(t) = \mathbb{P}\Big\{\xi(t) = n \ \Big| \ \xi(0) = 0\Big\} = \binom{\kappa + n - 1}{n}(1-q)^{\kappa}q^n \quad \text{if } \lambda, \kappa > 0 \qquad (2a)$$

$$h_n(t) = \mathbb{P}\Big\{\xi(t) = n \ \Big| \ \xi(0) = 0\Big\} = e^{-r}\frac{r^n}{n!} \quad \text{if } \lambda = 0, \kappa > 0 \qquad (2b)$$

$$g_n(t) = \mathbb{P}\Big\{\xi(t) = n \ \Big| \ \xi(0) = 1\Big\} = \begin{cases} p & \{n = 0\} \\ (1-p)(1-q)q^{n-1} & \{n > 0\} \end{cases} \quad \text{if } \kappa = 0, \lambda > 0$$

with the parameters

$$p = \frac{\mu - \mu e^{-(\mu-\lambda)t}}{\mu - \lambda e^{-(\mu-\lambda)t}} \qquad\qquad (3a)$$

$$q = \frac{\lambda - \lambda e^{-(\mu-\lambda)t}}{\mu - \lambda e^{-(\mu-\lambda)t}} \quad \text{if } \lambda > 0 \qquad r = \gamma(1 - e^{-\mu t}) \quad \text{if } \lambda = 0 \qquad (3b)$$

assuming $\lambda \neq \mu$; or if $\lambda = \mu$,

$$p = q = \frac{\mu t}{1 + \mu t}. \qquad\qquad (3c)$$

The Pólya distribution of (2a) is the generalized version of the negative binomial, allowing for non-integer κ parameter. Recall that the generalized binomial coefficient $\binom{\theta}{k}$ for all $\theta \in \mathbb{R}$ and nonnegative integer $k \in \mathbb{N}$ is defined by

$$\binom{\theta}{k} = \frac{(\theta)_k}{(k)_k} \quad \text{with} \quad (\theta)_k = \begin{cases} 1 & \{k = 0\} \\ \theta \times (\theta - 1)_{k-1} = \theta(\theta - 1)\cdots(\theta - k + 1) & \{k > 0\} \end{cases}$$

So, the point mass function for the Pólya distribution with parameters (κ, q) is

$$h_n(t) = \begin{cases} (1-q)^\kappa & \{n=0\} \\ \frac{\kappa(\kappa+1)\cdots(\kappa+n-1)}{n!}(1-q)^\kappa q^n & \{n>0\} \end{cases}$$

The rates and the edge length can be rescaled simultaneously without affecting the distributions. Dissecting into scale-independent parameters (assuming $q \neq p$):

$$p = \frac{1 - e^{-\delta(\mu t)}}{1 - (1-\delta)e^{-\delta(\mu t)}} \qquad\qquad 1 - p = \frac{\delta e^{-\delta(\mu t)}}{1 - (1-\delta)e^{-\delta(\mu t)}}$$

$$q = \frac{(1-\delta)\left(1 - e^{-\delta(\mu t)}\right)}{1 - (1-\delta)e^{-\delta(\mu t)}} \qquad\qquad 1 - q = \frac{\delta}{1 - (1-\delta)e^{-\delta(\mu t)}}$$

with $\delta = 1 - \frac{\lambda}{\mu} = 1 - \frac{q}{p}$. The formulas are invertible: for a given $0 < p, q < 1$ we can find δ and the scaled edge length (μt).

Theorem 1 (Unicity of distribution parameters). *Let $0 < t$ be fixed. For any given $0 < p, q < 1$ and $0 < \kappa$, or with $q = 0$, for any given $0 < p < 1$ and $0 < r$, there exist valid rate settings $0 < \mu, 0 \leq \lambda$ that yield those distribution parameters as in Eq. (3)*

Proof. If $q = 0$, then $\lambda = 0$, and by $p = 1 - e^{-\mu t}$ and $r = \gamma p$, we can set μt and γ to match p and r. If $0 < q = p$, then set $(\mu t) = p/(1+p)$ and $\lambda = \mu$. Otherwise, since $q/p = 1 - \delta$ and $(1-q)/(1-p) = e^{\delta \mu t}$, set $\delta = 1 - \frac{q}{p}$, $(\mu t) = \frac{\ln \frac{1-q}{1-p}}{1 - \frac{q}{p}}$ and $\lambda = \mu(1 - \delta)$. $\qquad \square$

Note that even if the birth-death process has a stationary distribution only when $\lambda \leq \mu$ or $\delta \geq 0$, the formulas remain valid for all transient probabilities $(t < \infty)$ even when $\lambda > \mu$.

3. Results and discussion

3.1. Transient probabilities in the general case

First, suppose that duplications are allowed, and $\lambda_v > 0$ on all edges $uv \in T$. If there are $\xi_u = n$ copies at an ancestral node u, then they evolve independently along each child edge uv:

$$\xi_v = \zeta_0 + \zeta_1 + \cdots + \zeta_n \tag{4}$$

where ζ_0 denotes the *xenolog copies*, and ζ_i denote iid variables for the descendant *inparalog* copies from each ancestral instance $i = 1, \ldots, n$. The ζ_i variables follow the basic transition probabilities

$$\mathbb{P}\{\zeta_0 = k\} = h_k(t_{uv}) \quad \text{and} \quad \mathbb{P}\{\zeta_i = k\} = g_k(t_{uv}) \quad \text{for all } i > 0.$$

The key observation for calculating $\mathbb{P}\Big\{\xi_v = m \ \Big| \ \xi_u = n\Big\} = \mathbb{P}\{\zeta_0 + \zeta_1 + \cdots + \zeta_n = m\}$ is that $\zeta_i | \zeta_i > 0$ has the same geometric tail as the Pólya distribution of ζ_0. Since the distributions with the same tail parameter can be summed at ease, $\xi_v - s$ has a Pólya distribution with parameter $(\kappa + s)$, where $s = \sum_{i=1}^{n} \{\zeta_i > 0\}$ is the number of conserved copies. (The shorthand notation $\{\zeta_i > 0\}$ denotes indicator variable that takes the value 1 whenever ζ_i is positive, and the value 0 when $\zeta_i = 0$.)

Theorem 2 (Transient probabilities in the general case). *For a linear birth-death process with parameters* $\kappa, \lambda, \mu > 0$,

$$\mathbb{P}\Big\{\xi(t) = m \ \Big| \ \xi(0) = n\Big\}$$
$$= \sum_{s=0}^{\min\{n,m\}} \binom{\kappa + m - 1}{m - s}(1 - q)^{\kappa + s} q^{m-s} \binom{n}{s} p^{n-s}(1 - p)^s \quad (5)$$

with the parameters p, q *defined in Eqs.* (3).
    For a linear birth-death process with parameters $\lambda = 0$ *and* $\mu, \gamma > 0$,

$$\mathbb{P}\Big\{\xi(t) = m \ \Big| \ \xi(0) = n\Big\} = \sum_{s=0}^{\min\{n,m\}} e^{-r} \frac{r^{m-s}}{(m - s)!} \binom{n}{s}(1 - p)^s p^{n-s} \quad (6)$$

with the parameters p, r *defined in Eqs.* (3).

3.2. Phylogenetic model with conserved copies

    We amend the phylogenetic model by explicitly inserting a hidden random variable η_v of conserved ancestral copies between the copy numbers ξ_u and ξ_v on every edge uv. For the ease of presentation, we continue with $\lambda_v > 0$ at every node v, and return to the no-duplication model afterwards. Using Theorem 2,

$$\mathbb{P}\Big\{\eta_v = s \ \Big| \ \xi_u = n\Big\} = \binom{n}{s}(1 - p_v)^s (p_v)^{n-s} \qquad \{s \le n\} \quad (7a)$$

$$\mathbb{P}\Big\{\xi_v = m \ \Big| \ \eta_v = s\Big\} = \binom{\kappa_v + m - 1}{m - s}(1 - q_v)^{\kappa_v + s}(q_v)^{m-s} \quad \{s \le m\} \quad (7b)$$

with edge-specific loss, duplication, and gain parameters p_v, q_v, κ_v.
    A complete history fixes all counts ξ_u and η_u: $\{\xi_1 = n_1, \ldots, \xi_R = n_R, \eta_1 = s_1, \ldots, \eta_{R-1} = s_{R-1}\}$. The joint distribution of our phylogenetically linked

6

random variables is written explicitly as

$$
\begin{aligned}
\mathbb{P}\{\xi_1 = n_1, \ldots, \xi_R = n_R, \eta_1 = s_1, \ldots, \eta_{R-1} = s_{R-1}\} \\
= \mathbb{P}\{\xi_R = n_R\} \times \prod_{uv \in T} \Bigg(\underbrace{\binom{n_u}{s_v}(1 - p_v)^{s_v}(p_v)^{n_u - s_v}}_{\mathbb{P}\left\{\eta_v = s_v \mid \xi_u = n_u\right\}} \\
\times \underbrace{\binom{\kappa_v + n_v - 1}{n_v - s_v}(1 - q_v)^{\kappa_v + s_v}(q_v)^{n_v - s_v}}_{\mathbb{P}\left\{\xi_v = n_v \mid \eta_v = s_v\right\}} \Bigg),
\end{aligned}
\tag{8}
$$

All histories satisfying $s_v \leq \min\{n_u, n_v\}$ on every edge $uv \in T$ and $\mathbb{P}\{\xi_R = n_R\} \neq 0$ have positive probability if p_u, q_u are bounded away from 0 and 1.

Let $\Xi = \{n_v : v \in \mathcal{L}\}$ be a profile comprising the observed copy numbers. The **profile likelihood** is the sum of all history probabilities from (8) for the same profile:

$$
L(\Xi) = \mathbb{P}\{\Xi\} = \sum_{n_u, s_u \,:\, x \notin \mathcal{L}} \mathbb{P}\{\xi_1 = n_1, \ldots, \xi_R = n_R, \eta_1 = s_1, \ldots, \eta_{R-1} = s_{R-1}\},
$$

with infinitely many terms. Define the **partial profile** within every subtree as $\Xi_u = \{\forall v \in \mathcal{L}_u : \xi_v = n_v\}$ where $\mathcal{L}_u$ denotes the leaves in the subtree rooted at u, including the singleton $\mathcal{L}_u = \{u\}$ whenever u is a leaf. Define the likelihood of the partial profiles conditioned on ξ_u or η_u:

$$
C_u(n) = \mathbb{P}\Big\{\Xi_u \,\Big|\, \xi_u = n\Big\} \qquad \text{and} \qquad K_u(s) = \mathbb{P}\Big\{\Xi_u \,\Big|\, \eta_u = s\Big\}.
$$

At a leaf u, we have $C_u(n) = 1$ if $n = n_u$, the observed count, or $C_u(n) = 0$ if $n \neq n_u$. All other conditional likelihoods can be expressed using Equations (7a) and (7b) about the conditional distributions $\xi_u \mid \eta_u$ and $\eta_v \mid \xi_u$. At all nodes u,

$$
K_u(s) = \sum_{k=0}^{\infty} \binom{\kappa_u + s + (k-1)}{k}(1 - q_u)^{\kappa_u + s}(q_u)^k \times C_u(s + k); \tag{9a}
$$

and at every ancestral node u,

$$
C_u(n) = \prod_{uv \in T} \left(\sum_{s=0}^{n} \binom{n}{s}(1 - p_v)^s (p_v)^{n-s} \times K_v(s) \right). \tag{9b}
$$

The family distribution at the root R is needed to sum across the likelihoods $C_R(n)$ to get the profile likelihood

$$
L(\Xi) = \mathbb{P}\{\Xi\} = \sum_{n=0}^{\infty} \mathbb{P}\{\xi_R = n\}\mathbb{P}\{\Xi_R \mid \xi_R = n\} = \sum_{n=0}^{\infty} \mathbb{P}\{\xi_R = n\}C_R(n).
$$

Assume that the root copy number follows a Pólya distribution with some parameters $\kappa_R, q_R > 0$:

$$L(\Xi) = \binom{\kappa_R + n - 1}{n}(1 - q_R)^r (q_R)^n C_R(n). \tag{10}$$

After defining $\eta_R = 0$, Eq. (10) is the same formula for the likelihoods K_R as on the edges, and $L(\Xi) = K_R(0)$.

3.3. Empty profile likelihood

Typically, the input sample does not include families with an **empty profile** that has $\xi_v = 0$ at all leaves v. The model defines the probability of such a profile.

Theorem 3 (Empty profile likelihood). *Define $\epsilon_u = 0$ for all leaves u, and for every non-leaf u, $\epsilon_u = \prod_{uv \in T} \tilde{p}_v$ with*

$$\tilde{p}_v = \big(p_v + (1 - p_v)\epsilon_v(1 - \tilde{q}_v)\big) = \frac{p_v(1 - \epsilon_v) + \epsilon_v(1 - q_v)}{1 - q_v \epsilon_v}$$

at every non-root v, and

$$\tilde{q}_u = q_u \frac{1 - \epsilon_u}{1 - q_u \epsilon_u}$$

at every node u. The probability of the empty profile is

$$L(0) = \prod_{u=1}^{R}(1 - \tilde{q}_R)^{\kappa_R} = \prod_{u=1}^{R}\left(\frac{1 - q_u}{1 - q_u \epsilon_u}\right)^{\kappa_u}.$$

Let the input sample consist of the observed profiles for families $f = 1, \ldots, F$: $\Xi_f = \big\{\xi_u = n_{f,u} \colon u \in \mathcal{L}\big\}$. If the empty profiles are unobservable, then the likelihood of a single family profile is conditioned on the fact that at least one copy number is positive:

$$L^*(\Xi_f) = \mathbb{P}\Big\{\forall u \in \mathcal{L}_R \colon \xi_u = n_{f,u} \,\Big|\, \exists u \in \mathcal{L}_R \colon \xi_u \neq 0\Big\} = \frac{L(\Xi_f)}{1 - L(0)},$$

using the uncorrected likelihoods $L(\Xi)$ without conditioning on being empty, and in particular the empty profile likelihood $L(0)$ from Theorem 3. Applying the correction to the entire sample:

$$L^* = \prod_{f=1}^{F} L^*(\Xi_f) = \frac{\prod_{f=1}^{F} L(\Xi_f)}{(1 - L(0))^F}. \tag{11}$$

The correction of Equation (11) is akin to Felsenstein's likelihood correction formula for restriction site evolution [16].

3.4. Computing the profile likelihood

Since the ancestors' copy number $\{\xi_u = n\}$ may be possible for all nonnegative integers n, the likelihood recurrences of (9) involve infinite sums for K_u, and infinitely many $C_u(n)$. We can, however, factor out the histories with parallel losses for a finite calculation. Define $\tilde{\xi}_u$ at every ancestral node u as the number of copies that are not lost simultaneously in all descendant lineages to $\mathcal{L}_u$. Let $\tilde{\eta}_u$ denote the number of ancestral copies that are not lost either on the edge leading to u or in the subtree T_u. In other words, $\tilde{\eta}$ and $\tilde{\eta}$ count only the progenitors of copies at the leaves. (Note that the *ancestral* copy numbers $\tilde{\xi}, \tilde{\eta}$ count the ancestral genes of extant copies, as opposed to the *ancestors'* copy numbers ξ, η that count all homologs in the ancestors' genomes.) Define ϵ_u, $\tilde{p}_v$ and $\tilde{q}_u$ as in Theorem 3. Since ancestral copies are lost independently with probability ϵ_u, for $0 \leq \ell \leq n$, $\mathbb{P}\{\tilde{\xi}_u = \ell \mid \xi_u = n\} = \binom{n}{\ell}(1 - \epsilon_u)^\ell (\epsilon_u)^{n-\ell}$ and, for all $0 \leq s \leq t$, $\mathbb{P}\{\tilde{\eta}_u = s \mid \eta_u = t\} = \binom{t}{s}(1 - \epsilon_u)^s (\epsilon_u)^{t-s}$.

Theorem 4 (Likelihood computation). *Given a profile Ξ, define the conditional likelihoods*

$$\tilde{K}_u(s) = \mathbb{P}\{\Xi_u \mid \tilde{\eta}_u = s\} \quad and \quad \tilde{C}_u(\ell) = \mathbb{P}\{\Xi_u \mid \tilde{\xi}_u = \ell\}$$

at all nodes u. In particular, the profile likelihood is $L(\Xi) = \tilde{K}_R(0)$ at the root R. Define the sum of observed leaf copy numbers within every subtree: $m_u = \sum_{v \in \mathcal{L}_u} n_v$.

(i) *For all $s > m_u$, $\tilde{K}_u(s) = 0$ and for all $\ell > m_u$, $\tilde{C}_u(\ell) = 0$.*

(ii) *At every node u, for all $0 \leq s \leq m_u$,*

$$\tilde{K}_u(s) = \sum_{\ell=s}^{m_u} \tilde{C}_u(\ell) \times \binom{\kappa_u + \ell - 1}{\ell - s}(1 - \tilde{q}_u)^{\kappa_u + s}(\tilde{q}_u)^{\ell - s}. \tag{12}$$

(iii) *If u is a leaf, then $\tilde{C}_u(\ell) = \{\ell = n_u\}$. If u is an ancestral node with children $uv, uw \in T$, then for all $0 \leq \ell \leq m_u = m_v + m_w$,*

$$\tilde{C}_u(\ell) = \sum_{s=0}^{\min\{\ell, m_v\}} \tilde{K}_v(s) \times \tilde{K}_w^\ell(\ell - s)$$

$$\times \binom{\ell}{s}\left(\frac{1 - \tilde{p}_v}{1 - \tilde{p}_v \tilde{p}_w}\right)^s \left(\frac{\tilde{p}_v - \tilde{p}_v \tilde{p}_w}{1 - \tilde{p}_v \tilde{p}_w}\right)^{\ell - s} \tag{13}$$

with $\tilde{K}_w^\ell(\ell) = \tilde{K}_w(\ell)$, and, for all $0 \leq d < \ell$,

$$\tilde{K}_w^\ell(d) = (1 - \tilde{p}_w)\tilde{K}_w^\ell(d + 1) + \tilde{p}_w \tilde{K}_w^{\ell-1}(d). \tag{14}$$

Note that Equation (12) also applies to a duplication-loss ($\lambda_u, \mu_u > 0$) model with no gain($\kappa_u = 0$). Since $\tilde{\xi}_u$ is the sum of $s = \tilde{\eta}_u$ geometric distributions,

9

it has a negative binomial distribution with parameters s and $\tilde{q}$. So, $\tilde{K}_u(0) = \tilde{C}_u(0)$, and for all $1 \leq s \leq m_u$,

$$\tilde{K}_u(s) = \sum_{\ell=0}^{m_u} \binom{\ell-1}{\ell-s} (1-\tilde{q}_u)^s \tilde{q}_u^{\ell-s}.$$

3.5. Multifurcations, missing data and partial genomes

The graphical model, as presented, assumes (1) a binary phylogeny, (2) unambiguous observation of the copy numbers ξ_u at the leaves, and (3) a complete annotated genome. All three assumptions can be relaxed.

Non-binary phylogeny

A *degenerate* phylogeny T represents the parent-child relationships in a non-binary rooted tree. In such a phylogeny, the ancestral nodes may have 2 or more children. In practice, it makes sense to put multifurcating nodes at deep ancestors to represent the ambiguity of resolving short edges, and a ternary root is common if the phylogeny was derived from an unrooted tree. The likelihood recurrences of Theorem 4 can accommodate any d-ary node, by considering survival in $1, 2, 3, \ldots d$ child lineages incrementally (for any child ordering).

Theorem 5 (Likelihood recurrences for multifurcating node)**.** *Let u be a node in a degenerate phylogeny with $d \geq 2$ distinct children $uv_1, \ldots, uv_d \in T$ enumerated in any order. Let $\epsilon_{u,-i} = \prod_{j=i}^{d} \tilde{p}_{v_j}$ for $i = 1, \ldots, d$, so that $\epsilon_u = \epsilon_{u,-1}$. Define the likelihoods $\tilde{C}_u^{-i}(\ell)$ and $\tilde{K}_{v_{i..d}}^{\ell}(s)$ conditioned on s surviving copies in the subtrees of $v_i, v_{i+1}, \ldots, v_d$:*

$$\tilde{C}_u^{-d}(\ell) = K_{v_d}(\ell) \qquad \{0 \leq \ell \leq m_{v_d}\} \tag{15a}$$

and, for all $0 < i \leq d$ and for all $0 \leq \ell \leq m_{v_{i-1}} + \cdots + m_{v_d}$

$$\tilde{C}_u^{-(i-1)}(\ell) = \sum_{s=0}^{\min\{\ell, m_{v_{i-1}}\}} \tilde{K}_{v_{i-1}}(s) \times \tilde{K}_{v_{i..d}}^{\ell}(\ell-s)$$

$$\times \binom{\ell}{s} \left(\frac{1-\tilde{p}_{v_{i-1}}}{1-\epsilon_{u,-(i-1)}}\right)^s \left(\frac{\tilde{p}_{v_{i-1}} - \tilde{p}_{v_{i-1}}\epsilon_{u,-i}}{1-\epsilon_{u,-(i-1)}}\right)^{\ell-s}, \tag{15b}$$

with

$$\tilde{K}_{v_{i..d}}^{\ell}(\ell) = \tilde{C}_u^{-i}(\ell) \tag{15c}$$

$$\tilde{K}_{v_{i..d}}^{\ell}(k) = (1-\epsilon_{u,-i})\tilde{K}_{v_{i..d}}^{\ell}(k+1) + \epsilon_{u,-i}\tilde{K}_{v_{i..d}}^{\ell-1}(k) \quad \{0 \leq k < \ell\} \tag{15d}$$

Then $\tilde{C}_u(\ell) = \tilde{C}_u^{-1}(\ell)$.

Missing copy numbers

Instead of setting 0, a copy number can be declared unobserved, or ambiguous. In the likelihood computations, such a profile can be accommodated by imposing $\tilde{K}_u(s) = 1$ and $\tilde{C}_u(\ell) = 1$ at all nodes u for which the partial profile is ambiguous at every leaf $\mathcal{L}_u$. Equivalently, truncate the phylogeny by clipping all edges leading to unobserved copy numbers in a postorder traversal.

Partial genomes

An incomplete genome at a leaf u is characterized by the fraction $(1 - \epsilon_u)$ of the genome that is annotated. Assuming a simple model of randomly missing copies, we have

$$\mathbb{P}\left\{\tilde{\xi}_u = k \;\middle|\; \xi_u = n\right\} = \binom{n}{k}(1 - \epsilon_u)^k(\epsilon_u)^{n-k},$$

where $\tilde{\xi}$ is the number of annotated copies, and ξ is the true copy number in the complete genome. In other words, the recurrences of Theorems 4 for the likelihood and 3 for the empty profile remain the same, with the only change that $\tilde{q}_u \neq q_u$ at such a leaf with $\epsilon_u > 0$. Without constraints, however, the trio (p_u, q_u, ϵ_u) is not identifiable: the distribution parameters

$$p = p_u + (1 - p_u)\epsilon_u\frac{1 - q_u}{1 - q_u\epsilon_u} \qquad q = q_u\frac{1 - \epsilon_u}{1 - q_u\epsilon_u} \qquad \epsilon = 0,$$

produce the exact same distribution at u as (p_u, q_u, ϵ_u).

3.6. Posterior probabilities for ancestral copy numbers

Let $\Xi = \{\xi_v = n_v \colon v \in \mathcal{L}\}$ be an arbitrary profile of copy numbers observed at the leaves. Theorems 4 and 5 provide the recurrences for the conditional likelihoods $\tilde{C}_u$ and $\tilde{K}_u$ of the partial profile Ξ_u conditioned on the surviving copies $\tilde{\xi}_u$ and $\tilde{\eta}_u$, respectively. Define the complementary **outside likelihoods**

$$B_u(\ell) = \mathbb{P}\{\Xi - \Xi_u, \tilde{\xi}_u = \ell\} \quad \text{and} \quad J_u(s) = \mathbb{P}\{\Xi - \Xi_u, \tilde{\eta}_u = s\}, \qquad (16)$$

where $\Xi - \Xi_u = \{\xi_v = n_v \colon v \in \mathcal{L} - \mathcal{L}_u\}$ denotes the profile outside the subtree rooted at node u.

Theorem 6 (Outside likelihoods). *Let $\Xi = \{\xi_v = n_v \colon v \in \mathcal{L}\}$ be an arbitrary profile, and define the outside likelihoods as in Equation (16). The following recurrences hold.*

(i) *At the root, $J_R(0) = 1$ and $J_R(s) = 0$ for $s > 0$.*

(ii) *At any node u, for all $0 \leq \ell \leq m_u$,*

$$B_u(\ell) = \sum_{s=0}^{\ell} J_u(s) \times \binom{\kappa_u + \ell - 1}{\ell - s}(1 - \tilde{q}_u)^{\kappa_u + s}(\tilde{q}_u)^{\ell - s}. \qquad (17)$$

11

(iii) At every non-root node v with parent u and sibling w (i.e, $uv, uw \in T$), for all $0 \le s \le m_v$,

$$J_v(s) = \sum_{\ell=s}^{m_u} B_u(\ell) \times \tilde{K}_w^\ell(\ell - s)$$

$$\times \binom{\ell}{s} \Big(\frac{1 - \tilde{p}_v}{1 - \tilde{p}_v \tilde{p}_w} \Big)^s \Big(\frac{\tilde{p}_v - \tilde{p}_v \tilde{p}_w}{1 - \tilde{p}_v \tilde{p}_w} \Big)^{\ell-s}. \quad (18)$$

Theorem 6 with Theorem 4 deliver the posterior probabilities in computable formulas.

Corollary 7 (Posterior probabilities). *Fix an arbitrary profile Ξ and let $\tilde{C}_u, \tilde{K}_u, B_u, Ju$ denote the inside and outside likelihoods at every node u.*

(i) The profile likelihood can be computed by either formulas

$$L(\Xi) = \mathbb{P}\{\Xi\} = \sum_{\ell=0}^{m_u} B_u(\ell) \times \tilde{C}_u(\ell) = \sum_{s=0}^{m_u} J_u(s) \times \tilde{K}_u(s). \quad (19)$$

(ii) The posterior distribution of $\tilde{\xi}_u$ is $\mathbb{P}\{\tilde{\xi}_u = \ell \mid \Xi\} = \frac{B_u(\ell) \times \tilde{C}_u(\ell)}{L(\Xi)}$.

(iii) The posterior distribution of $\tilde{\eta}_u$ is $\mathbb{P}\{\tilde{\eta}_u = s \mid \Xi\} = \frac{J_u(s) \times \tilde{K}_u(s)}{L(\Xi)}$.

3.7. Partial derivatives of the likelihood

Suppose that we are interested in the corrected likelihood for a sample of family profiles $\{\Xi_f : f = 1, \ldots, F\}$. By Equation (11), the derivative of the corrected log-likelihood, with respect to any distribution parameter θ is

$$\frac{\partial}{\partial \theta} (\ln L^*) = \Big(\sum_{f=1}^{F} \frac{L'(\Xi_f)}{L(\Xi_f)} \Big) + F \frac{L'(0)}{1 - L(0)}, \quad (20)$$

where $L'(\Xi) = \frac{\partial L(\Xi)}{\partial \theta}$ denotes the derivative of the uncorrected profile likelihood.

It is tempting to choose the optimized distribution parameters directly as κ_u and the survival parameters $\tilde{p}_u, \tilde{q}_u$ for the maximization of the corrected log-likelihood $\ln L^*$. They uniquely determine the parameters p_u, q_u, and, consequently, the edge-specific rate parameters. The values of $\tilde{p}$ and $\tilde{q}$ are, however, not arbitrary across the tree.

Theorem 8 (Unicity of survival parameters). *Let T be a phylogeny equipped with arbitrary gain rates $0 < \kappa_u$ and arbitrary survival parameters $0 < \tilde{p}_u, \tilde{q}_u < 1$ at every node u. If, at every non-root ancestral node u,*

$$\tilde{p}_u > (1 - \tilde{q}_u) \prod_{uv \in T} \tilde{p}_v, \quad (*)$$

then there exists a phylogenetic birth-death model on the same phylogeny with valid distribution parameters $0 < p_u, q_u < 1$ and same gain rates κ_u. Otherwise, no solution exists with positive p_u on every edge.

In light of Theorem 8, we should aim at using the partial derivatives with respect to $\tilde{p}$ and $\tilde{q}$ as an intermediate step toward inferring the dependance on p and q. Using Corollary 7, we can determine the partial derivatives with respect to the survival distribution parameters.

Theorem 9. *(i) At every node $1 \leq u \leq R$,*

$$\frac{\partial L(\Xi)}{\partial \tilde{q}_u} = L(\Xi) \times \left(\frac{1}{\tilde{q}_u} \mathbb{E}\big[\tilde{\xi}_u \mid \Xi\big] - \left(\frac{1}{\tilde{q}_u} + \frac{1}{1-\tilde{q}_u}\right) \mathbb{E}\big[\tilde{\eta}_u \mid \Xi\big] - \frac{1}{1-\tilde{q}_u}\kappa_u \right).$$

(ii) At every non-root node $1 \leq v < R$,

$$\frac{\partial L(\Xi)}{\partial \tilde{p}_v} = L(\Xi) \times \left(\frac{1}{\tilde{p}_v} \mathbb{E}\big[\tilde{\xi}_u \mid \Xi\big] - \left(\frac{1}{\tilde{p}_v} + \frac{1-\epsilon}{1-\tilde{p}_v}\right) \mathbb{E}\big[\tilde{\eta}_v \mid \Xi\big] \right),$$

where $\epsilon = \frac{\epsilon_u}{\tilde{p}_v} = \frac{\prod_{uw \in T} \tilde{p}_w}{\tilde{p}_v}$ is the product of $\tilde{p}_w$ across the siblings with the same parent $uv, uw \in T$.

(iii) The partial derivatives with respect to κ_u are, for all $1 \leq u \leq R$,

$$\frac{\partial L(\Xi)}{\partial \kappa_u} = L(\Xi) \times \left(\ln(1-\tilde{q}_u) + \sum_{i=0}^{m_u-1} \frac{\mathbb{P}\{\tilde{\xi}_u > i \mid \Xi\} - \mathbb{P}\{\tilde{\eta}_u > i \mid \Xi\}}{\kappa_u + i} \right).$$

(iv) The partial derivatives for the empty profile $\Xi = 0$ are

$$\frac{\partial L(0)}{\partial \tilde{q}_u} = -L(0)\frac{\kappa_u}{1-\tilde{q}_u}, \quad \frac{\partial L(0)}{\partial \tilde{p}_u} = 0 \quad and \quad \frac{\partial L(0)}{\partial \kappa_u} = L(0) \times \ln(1-\tilde{q}_u).$$

Note that using the posterior distributions from Corollary 7, we readily obtain the posterior expectations

$$\mathbb{E}\big[\tilde{\xi}_u \mid \Xi\big] = \sum_{\ell=0}^{m_u} \ell \times \mathbb{P}\{\tilde{\xi}_u = \ell \mid \Xi\} \quad \text{and} \quad \mathbb{E}\big[\tilde{\eta}_u \mid \Xi\big] = \sum_{s=0}^{m_u} s \times \mathbb{P}\{\tilde{\eta}_\xi \mid \Xi\},$$

as well as the distribution tails $\mathbb{P}\{\tilde{\xi}_u > i \mid \Xi\} = \sum_{\ell=i+1}^{m_u} \mathbb{P}\{\tilde{\xi}_u = \ell \mid \Xi\}$ and $\mathbb{P}\{\tilde{\eta}_u > i \mid \Xi\} = \sum_{\ell=i+1}^{m_u} \mathbb{P}\{\tilde{\eta}_u = \ell \mid \Xi\}$ which are needed in Theorem 9 and the following Corollary 10, which combines Theorem 9 with Equation (20).

Corollary 10. *Let $\Phi = \ln L^*$ denote the corrected log-likelihood for a sample of family profiles $\{\Xi_f \colon f = 1, \ldots, F\}$. Define the posterior expected counts across the sample*

$$\tilde{N}_u = \sum_{f=1}^{F} \mathbb{E}\big[\tilde{\xi}_u \mid \Xi_f\big] \qquad\qquad \tilde{S}_u = \sum_{f=1}^{F} \mathbb{E}\big[\tilde{\eta}_u \mid \Xi_f\big]$$

$$\tilde{N}_u^{>i} = \sum_{f=1}^{F} \mathbb{P}\{\tilde{\xi}_u > i \mid \Xi_f\} \qquad\qquad \tilde{S}_u^{>i} = \sum_{f=1}^{F} \mathbb{P}\{\tilde{\eta}_u > i \mid \Xi_f\}$$

at every node $1 \leq u \leq R$.

(i) *At every node* $1 \leq u \leq R$,

$$\frac{\partial(\ln L^*)}{\partial \tilde{q}_u} = \frac{\tilde{N}_u - \tilde{S}_u}{\tilde{q}_u} - \frac{\tilde{S}_u + \kappa_u/\big(1 - L(0)\big)}{1 - \tilde{q}_u}$$

(ii) *For a non-root node* $1 \leq v < R$, *let* u *be its parent:*

$$\frac{\partial(\ln L^*)}{\partial \tilde{p}_u} = \frac{\tilde{N}_u}{\tilde{p}_u} - \frac{(1 - \epsilon)\tilde{S}_v}{1 - \tilde{p}_u}$$

with $\epsilon = \big(\prod_{uw \in T} \tilde{p}_w\big)/\tilde{p}_v$.

(iii) *At every node* $1 \leq u \leq R$,

$$\frac{\partial(\ln L^*)}{\partial \kappa_u} = F\frac{\ln(1 - \tilde{q}_u)}{1 - L(0)} + \sum_{i=0}^{m_u - 1} \frac{\tilde{N}_u^{>i} - \tilde{S}_u^{>i}}{\kappa_u + i}$$

Powerful numerical algorithms for function maximization (conjugate gradient and variable metric methods like Broyden-Fletcher-Goldfarb-Shanno) exploit the gradient for quick convergence to optimum. The likelihood optimization for a phylogenetic birth-death model can rely on the computation of both the likelihood (Theorem 4), and the gradient with respect to the parameters κ_u, p_u, q_u across the tree. By Theorem 1, the probabilistic model is uniquely determined by the parameter set, up to equivalent rate scalings. Maximizing the likelihood with respect to the survival distribution parameters $\tilde{p}$ and $\tilde{q}$ from Theorem 9 is not straightforward because Theorem 8 imposes monotonicity constraints between parameters on adjoining edges. Let u be an arbitrary node at some depth d (root is at depth 0). For a distribution parameter such as $\theta_v = p_v$ or $\theta_v = q_v$,

$$\frac{\partial L(\Xi)}{\partial \theta_v} = \sum_{u=1}^{R} \frac{\partial L(\Xi)}{\partial \tilde{q}_u} \frac{\partial \tilde{q}_u}{\partial \theta_v} + \sum_{u=1}^{R} \frac{\partial L(\Xi)}{\partial \tilde{p}_u} \frac{\partial \tilde{p}_u}{\partial \theta_v},$$

by the chain rule. In particular, p_v and q_v influence $\tilde{p}_u$ and $\tilde{q}_u$ at nodes u along the path between the root and v. Consequently, the above sums include only the ancestors of v, and the partial derivatives can be computed in a preorder traversal. We state the procedure in a generic theorem about recovering the derivatives of any function Φ of the distribution parameters.

Theorem 11 (Gradient computation). *Let* Φ *be an arbitrary differentiable function of the distribution parameters* $\{\tilde{p}_u, \tilde{q}_u\}_{u=1}^{R}$. *Let* $\Phi^{(\theta_v)} = \frac{\partial f}{\partial \theta_v}$ *denote the partial derivative with respect to any distribution parameter* θ_v. *The partial derivatives* $\Phi^{(p_v)}$ *(for non-root* v*),* $\Phi^{(q_v)}$ *(for any* v*) and* $\Phi^{(\epsilon_v)}$ *(for non-leaf* v*) can be computed in a preorder traversal by the following recurrences.*

(i) *At the root* $v = R$,

$$\Phi^{(q_R)} = \frac{1 - \epsilon_R}{(1 - q_R \epsilon_R)^2} \Phi^{(\tilde{q}_R)} \tag{21a}$$

$$\Phi^{(\epsilon_R)} = \frac{1 - q_R}{(1 - q_R \epsilon_R)^2} \Big(-q_R \Phi^{(\tilde{q}_R)}\Big). \tag{21b}$$

14

(ii) At every non-root node $1 \le v < R$,

$$\Phi^{(p_v)} = \frac{1 - \epsilon_v}{1 - q_v \epsilon_v} \left(\Phi^{(\tilde{p}_v)} + \epsilon \Phi^{(\epsilon_u)} \right) \tag{22a}$$

$$\Phi^{(q_v)} = \frac{1 - \epsilon_v}{(1 - q_v \epsilon_v)^2} \left(\Phi^{(\tilde{q}_v)} - (1 - p_v) \epsilon_v \left(\Phi^{(\tilde{p}_v)} + \epsilon \Phi^{(\epsilon_u)} \right) \right) \tag{22b}$$

and, if v is not a leaf,

$$\Phi^{(\epsilon_v)} = \frac{1 - q_v}{(1 - q_v \epsilon_v)^2} \left((1 - p_v) \left(\Phi^{(\tilde{p}_v)} + \epsilon \Phi^{(\epsilon_u)} \right) - q_v \Phi^{(\tilde{q}_v)} \right) \tag{22c}$$

with the parent u and

$$\epsilon = \prod_{w:\, uw \in T} \{w \ne v\} \tilde{p}_w = \frac{\epsilon_u}{\tilde{p}_v}.$$

Theorem 11 can be employed with the individual family profiles using $\Phi = L(\Xi_f)$ and plugging $L'(\Xi_f) = \Phi^{(\theta)}$ into the corrected log-likelihood formula of (20) for each $f = 1, \dots, F$ in the sum, as well as for $L'(0)$. But it is more efficient to carry out the procedure only once at the end, using $\Phi = \ln L^*$ directly with its partial derivatives from Corollary 10.

For the purposes of likelihood maximization, use a parametrization with the logistic and exponential functions as

$$p_u = \frac{1}{1 + e^{-\alpha_u}} \quad q_u = \frac{1}{1 + e^{-\beta_u}} \quad \kappa_u = e^{\gamma_u}$$

with unconstrained real-valued parameters

$$\alpha_u = \ln \frac{p_u}{1 - p_u} \quad \beta_u = \ln \frac{q_u}{1 - q_u} \quad \gamma_u = \ln \kappa_u.$$

The partial derivatives of $\Phi = \ln L^*$ (or of a single-profile likelihood $\Phi = L(\Xi)$) are computed by the chain rule as

$$\Phi^{(\alpha_u)} = \Phi^{(p_u)} \frac{\partial p_u}{\partial \alpha_u} = p_u (1 - p_u) \Phi^{(p_u)} \qquad \{0 < u < R\}$$

$$\Phi^{(\beta_u)} = \Phi^{(q_u)} \frac{\partial q_u}{\partial \beta_u} = q_u (1 - q_u) \Phi^{(q_u)} \qquad \{0 < u \le R\}$$

$$\Phi^{(\gamma_u)} = \Phi^{(\kappa_u)} \frac{\partial \kappa_u}{\partial \gamma_u} = \kappa_u \Phi^{(\kappa_u)} \qquad \{0 < u \le R\}$$

3.8. Likelihoods in the no-duplication model

In the case of $\lambda_v = 0$ on all edges $uv \in T$, the joint distribution of the random variables multiplies Poisson and binomial masses:

$$\mathbb{P}\{\xi_1 = n_1, \dots, \xi_R = n_R, \eta_1 = s_1, \dots, \eta_{R-1} = s_{R-1}\}$$

$$= \mathbb{P}\{\xi_R = n_R\} \times \prod_{uv \in T} \left(\binom{n_u}{s_v} (1 - p_v)^{s_v} (p_v)^{n_u - s_v} \times e^{-r_v} \frac{(r_v)^{n_v - s_v}}{(n_v - s_v)!} \right).$$

15

This time we assume a Poisson distribution at the root: $\mathbb{P}\{\xi_R = n\} = e^{-r}r^n/(n!)$, and, as before $s_R = 0$ for retrieving the likelihood $L(\Xi) = K_R(0)$. The recurrences for the likelihood and the empty profile are adjusted accordingly. In particular,

$$K_u(s) = \sum_{k=0}^{\infty} e^{-r_u} \frac{(r_u)^k}{k!} \times C_u(s+k), \qquad (23)$$

but the recurrence for C_u stays the same.

Theorem 12 (Empty profile in the no-duplication model). *Define ϵ_u as in Theorem 3, with $\tilde{p}_v = \big(p_v + (1-p_v)\epsilon_v\big)$ at every non-root node v, and $\tilde{r}_u = r_u\big(1 - \epsilon_u\big)$ at every node u. The probability of the empty profile is*

$$L(0) = \prod_{u=1}^{R} e^{-\tilde{r}_u} = \prod_{u=1}^{R} \exp\Big(-r_u(1 - \epsilon_u)\Big).$$

The likelihood computations of Theorems 4 and 6 adapt easily to the no-duplication model, with $\tilde{\xi}$ and $\tilde{\eta}$ defined as before. Two recurrences change: at every node u, and for all $0 \le s \le m_u$,

$$\tilde{K}_u(s) = \sum_{\ell=s}^{m_u} \tilde{C}_u(\ell) \times e^{-\tilde{r}_u} \frac{(\tilde{r}_u)^{\ell-s}}{(\ell-s)!},$$

and, for all $0 \le \ell \le m_u$,

$$B_u(\ell) = \sum_{s=0}^{\ell} J_u(s) \times e^{-\tilde{r}_u} \frac{(\tilde{r}_u)^{\ell-s}}{(\ell-s)!}.$$

Consequently, the derivatives of the profile likelihood are

$$\frac{\partial L(\Xi)}{\partial \tilde{r}_u} = \frac{\partial}{\partial \tilde{r}_u} \Big(\sum_{\ell=0}^{m_u} B_u(\ell) \times \tilde{C}_u(\ell)\Big)$$

$$= \sum_{0 \le s \le \ell \le m_u} J_{u(s)} \times \tilde{C}_u(\ell) \times e^{-\tilde{r}_u} \frac{(\tilde{r}_u)^{\ell-s}}{(\ell-s)!} \left(\frac{\ell-s}{\tilde{r}_u} - 1\right)$$

$$= L(\Xi) \times \left(\frac{\mathbb{E}\big[\tilde{\xi}_u \mid \Xi\big] - \mathbb{E}\big[\tilde{\eta}_u \mid \Xi\big]}{\tilde{r}_u} - 1\right).$$

at every node $1 \le u \le R$. In particular, for the empty profile $\Xi = 0$,

$$\frac{\partial L(0)}{\partial \tilde{r}_u} = \frac{\partial}{\partial \tilde{r}_u} \Big(\prod_{v=1}^{R} e^{-\tilde{r}_v}\Big) = -L(0).$$

by Theorem 12. Substituting into Equation (20) for the derivatives of corrected log-likelihood on a sample of family profiles gives

$$\frac{\partial(\ln L^*)}{\partial \tilde{r}_v} = \frac{\tilde{N}_u - \tilde{S}_u}{\tilde{r}_u} - \frac{F}{1 - L(0)}.$$

The analogue of Theorem 11 is the following claim.

Theorem 13 (Gradient in the no-duplication model). *Let Φ be an arbitrary differentiable function of the distribution parameters $\left\{\tilde{p}_u, \tilde{r}_u\right\}_{u=1}^R$ in a no-duplication model. Let $\Phi^{(\theta_v)} = \frac{\partial f}{\partial \theta_v}$ denote the partial derivative with respect to any distribution parameter θ_v. The partial derivatives $\Phi^{(p_v)}$ (for non-root v), $\Phi^{(r_v)}$ (for any v) and $\Phi^{(\epsilon_v)}$ (for non-leaf v) can be computed in a preorder traversal by the following recurrences.*

(i) *At the root $v = R$,*

$$\Phi^{(r_R)} = \left(1 - \epsilon_R\right)\Phi^{(\tilde{r}_R)} \tag{24a}$$

$$\Phi^{(\epsilon_R)} = -r_R \Phi^{(\tilde{r}_R)}. \tag{24b}$$

(ii) *At every non-root node $1 \le v < R$,*

$$\Phi^{(p_v)} = \left(1 - \epsilon_v\right)\left(\Phi^{(\tilde{p}_v)} + \epsilon \Phi^{(\epsilon_u)}\right) \tag{25a}$$

$$\Phi^{(r_v)} = \left(1 - \epsilon_v\right)\Phi^{(\tilde{r}_v)}. \tag{25b}$$

and, if v is not a leaf,

$$\Phi^{(\epsilon_v)} = \left(1 - p_v\right)\left(\Phi^{(\tilde{p}_v)} + \epsilon \Phi^{(\epsilon_u)}\right) - r_v \Phi^{(\tilde{r}_v)} \tag{25c}$$

with the parent u and

$$\epsilon = \prod_{w:\, uw \in T} \{w \ne v\}\tilde{p}_w = \frac{\epsilon_u}{\tilde{p}_v}.$$

Note that the different duplications models can be used in the same tree: some edges can have $\lambda = 0$, and some $\lambda > 0$. In the recurrences for $\tilde{K}_v$ and B_v, either the Poisson (if $\lambda_v = 0$) or the Pólya (if $\lambda_v > 0$) formulas apply, and the computed derivatives are ∂r_v or $\partial \kappa_v$, respectively.

3.9. Algorithmic complexity

The set of conditional likelihoods $\tilde{C}_u(\ell)$ and $\tilde{K}_u(s)$ for a given profile Ξ can be computed in a postorder traversal of the phylogeny using Theorem 4. The recurrences for $\tilde{K}_u(s)$ from (12) are straigthforward to implement by embedded loops over $0 \le s \le \ell \le m_u$. Define

$$\mathsf{h}_u(s, t) = \binom{\kappa_u + s + t - 1}{t}(1 - \tilde{q}_u)^{\kappa_u + s}(\tilde{q}_u)^t.$$

```
      // Computing K̃_u(s) for all s
 1  for ℓ ← 0, 1, …, m_u
 2      for s ← 0, 1, …, ℓ
 3          K̃_u(s) ← K̃_u(s) + C̃_u(ℓ) × h_u(s, ℓ − s)
```

For the recurrence of (12), compute $K_w^{s+t}(s)$ looping over t and s in the opposite direction. Let

$$\mathsf{g}_v(s,t) = \binom{s+t}{s}\left(\frac{1-\tilde{p}_v}{1-\tilde{p}_v\tilde{p}_w}\right)^s\left(\frac{\tilde{p}_v-\tilde{p}_v\tilde{p}_w}{1-\tilde{p}_v\tilde{p}_w}\right)^t.$$

    // Computing $\tilde{C}_u(\ell)$ for all ℓ at u with children $uv, uw \in T$
1 **for** $t \leftarrow m_w, m_w - 1, \ldots, 0$
2     $K_w^t(t) \leftarrow K_w(t)$
3   **for** $s \leftarrow 0, 1, \ldots, m_v$
4        $\tilde{C}_u(s+t) \leftarrow \tilde{C}_u(s+t) + \tilde{K}_v(s) \times \tilde{K}_w^{s+t}(t) \times \mathsf{g}_v(s,t)$
5        $K_w^{(s+1)+t}(t) \leftarrow (1-\tilde{p}_w)\tilde{K}_w^{(s+1)}(t+1) + \tilde{p}_w\tilde{K}_w^{s+t}(t)$

Note that $\mathsf{g}_u(s,t)$ and $\mathsf{h}_u(s,t)$ can be computed in constant time. For instance, $\ln \mathsf{h}_u(s,0) = (\kappa_u + s)\ln(1-\tilde{q}_u)$, and for $t > 0$,

$$\ln \mathsf{h}_u(s,t) = (\kappa_u + s)\ln(1-\tilde{q}_u) + t\ln\tilde{q}_u$$
$$+ \ln\,`\Gamma(\kappa_u + s + t) - \ln\Gamma(\kappa_u + s) - \ln\Gamma(t+1)$$

with the Gamma function $\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}\,dx$ (so that $\Gamma(t+1) = t!$).

The outside likelihoods $B_u(\ell)$ and $J_u(s)$ from Theorem 6 are computed in a preorder traversal. Concomitantly, the posterior distributions for $\tilde{\xi}_u$ and $\tilde{\eta}_u$ are obtained by Corollary 7 in the same traversal. In addition, during the same preorder traversal, the partial derivatives can be computed with respect to all $\tilde{p}_u, \tilde{q}_u, \kappa_u$ parameters. The running time is quadratic in the total number of observed copies.

Theorem 14 (Running time for likelihood computation). *Let $\Xi = \{\xi_u = n_u\colon u \in \mathcal{L}\}$ be an arbitrary profile across $L = |\mathcal{L}|$ leaves. The profile likelihood and all posterior distributions for $\tilde{\xi}_u$ and $\tilde{\eta}_u$ for all ancestral nodes u can be computed in $O\big(hL(L\bar{n}^2 + 1)\big)$ time, where $\bar{n} = \frac{1}{L}\sum_{u=1}^{L} n_u$ is the average of the copy numbers at the leaves, and h is the phylogeny's height.*

In order to get the gradient of the corrected log-likelihood $\Phi = \ln L^*$ over a sample of F families, first compute the partial derivatives $\Phi^{(\tilde{p}_u)}$, $\Phi^{(\tilde{q}_u)}$, and $\Phi^{(\kappa_u)}$ of the corrected log-likelihood from the derivatives for the individual profile likelihoods using Equation (20). Subsequently, the recurrences of Theorem 11 compute all $\Phi^{(p_u)}$ and $\Phi^{(q_u)}$ in a single preorder traversal.

4. Conclusion

The mathematical framework for phylogenetic gain-loss-duplication models provides the clean decomposition of Equation (8), involving a network of dependent random variables. The elementary decomposition can be employed with standard Bayesian and likelihood methods, leading to efficient algorithms for a notoriously hard bioinformatics problem. A case in point is the fast gradient computation algorithm reported here.

5. Calculation

5.1. *Proof of Theorem 2*

Proof. First, suppose that $\lambda > 0$. Decompose $\xi(t)$ as in (4):

$$\xi(t) = \zeta_0(t) + \sum_{i=1}^{n} \zeta_i(t)$$

where ζ_0 follows Pólya with parameters (κ, q), and ζ_i are iid shifted geometric with parameters (p, q). Now define the random variable $\eta(t)$ as the number of conserved copies

$$\eta(t) = \sum_{i=1}^{n} \{\zeta_i(t) > 0\}$$

Since ζ_i are independent with $\mathbb{P}\{\zeta_i(t) = 0\} = p$,

$$\mathbb{P}\Big\{\eta(t) = s \ \Big| \ \xi(0) = n\Big\} = \binom{n}{s}(1-p)^s p^{n-s}. \tag{26}$$

Since the $\zeta_i = 0$ are immaterial in the sum, we can condition on $\eta(t)$:

$$\mathbb{P}\Big\{\xi(t) = m \ \Big| \ \eta(t) = s, \xi(0) = n\Big\} = \mathbb{P}\Big\{\xi(t) = m \ \Big| \ \eta(t) = s\Big\}$$
$$= \mathbb{P}\{\zeta_0 + \zeta_1' + \cdots + \zeta_s' = m\},$$

where $\zeta_1' - 1$ are iid random variables following a Pólya distribution with parameter 1:

$$\mathbb{P}\{\zeta_i' - 1 = k\} = (1-q)q^k = \binom{1+k-1}{k}(1-q)^1 q^k.$$

Looking specifically at the generator functions:

$$F_0(z) = \sum_{i=0}^{\infty} \mathbb{P}\{\zeta_0 = i\}z^i = \Big(\frac{1-q}{1-qz}\Big)^{\kappa}$$

$$F_i(z) = \sum_{i=0}^{\infty} \mathbb{P}\{\zeta_i' = i\}z^i = \frac{z(1-q)}{1-qz}$$

so

$$F(z) = \sum_{m=s}^{\infty} \mathbb{P}\Big\{\xi(t) = m \ \Big| \ \eta(t) = s\Big\}z^m = F_0(z)\prod_{i=1}^{s} F_i(z) = z^s \Big(\frac{1-q}{1-qz}\Big)^{\kappa+s}.$$

19

Hence, $\xi(t) - \eta(t)$ follows a Pólya distribution with parameter $(\kappa + \eta(t))$, and the same tail parameter q. Now,

$$\mathbb{P}\Big\{\xi(t) = m \,\Big|\, \xi(0) = n\Big\}$$
$$= \sum_s \mathbb{P}\Big\{\xi(t) = m \,\Big|\, \eta(t) = s\Big\}\mathbb{P}\Big\{\eta(t) = s \,\Big|\, \xi(0) = n\Big\}$$
$$= \sum_s \mathbb{P}\Big\{\xi(t) - \eta(t) = m - s \,\Big|\, \eta(t) = s\Big\}\mathbb{P}\Big\{\eta(t) = s \,\Big|\, \xi(0) = n\Big\}$$
$$= \sum_{s=0}^{\min\{n,m\}} \binom{n}{s}(1-p)^s p^{n-s}\binom{(\kappa+s)+(m-s)-1}{m-s}(1-q)^{\kappa+s}q^{m-s},$$

as claimed.

When $\lambda = 0$, define $\zeta_i(t)$ for xenologs and inparalogs:

$$\mathbb{P}\{\zeta_0(t) = k\} = e^{-r}\frac{r^k}{k!},$$

and $\zeta_i(t)$ for $i > 0$ are Bernoulli random variables with

$$\mathbb{P}\{\zeta_i(t) = 0\} = p \qquad \mathbb{P}\{\zeta_i(t) = 1\} = 1 - p.$$

We condition on $\eta(t) = \sum_{i=1}^n \{\zeta_i(t) > 0\} = \sum_{i=1}^n \zeta_i(t)$ with the same binomial distribution as in (26): now $\xi(t) - \eta(t)$ has a Poisson distribution. $\qquad\square$

5.2. Proof Theorem 3

Proof. Let C_u, K_u denote the likelihoods for the empty profile:

$$C_u(n) = \mathbb{P}\Big\{\forall v \in \mathcal{L}_u \colon \xi_v = 0 \,\Big|\, \xi_u = n\Big\}$$
$$K_u(s) = \mathbb{P}\Big\{\forall v \in \mathcal{L}_u \colon \xi_v = 0 \,\Big|\, \eta_u = s\Big\}.$$

Let Q_u denote the product of $(1 - \tilde{q}_v)^{\kappa_v}$ across all edges in the subtree of u: $Q_u = 1$ at a leaf, and at an ancestral node u with children v, w

$$Q_u = \big(Q_v(1 - \tilde{q}_v)^{\kappa_v}\big)\big(Q_w(1 - \tilde{q}_w)^{\kappa_w}\big).$$

We prove that for all nodes u,

$$K_u(s) = Q_u \times (\epsilon_u)^s(1 - \tilde{q}_u)^{\kappa_u+s} \qquad C_u(n) = Q_u \times (\epsilon_u)^n.$$

(With $0^0 = 1$ and $0^n = 0$ for $n > 0$.) In particular, at the root R, the probability of the empty profile is

$$L(0) = K_R(0) = Q_R(1 - \tilde{q}_R)^{\kappa_R} = \prod_{u=1}^R \left(\frac{1 - q_u}{1 - q_u\epsilon_u}\right)^{\kappa_u}.$$

We prove the claim by induction in the node height, starting with the leaves.

Base case.. At a leaf u (height 0), we have $C_u(0) = 1$ and $C_u(n) = 0$ for $n > 0$. Since $\epsilon_u = 0$, $C_u(n) = 0^n = \epsilon_u^n$ holds at all n.

At any node v, with $\kappa = \kappa_v, q = q_v, \epsilon = \epsilon_v, Q = Q_v$ and $\tilde{q} = q\frac{1-\epsilon}{1-q\epsilon}$:

$$
\begin{aligned}
K_v(s) &= \sum_{n=s}^{\infty} \binom{\kappa + n - 1}{n - s}(1 - q)^{\kappa+s}(q)^{n-s}C_v(n) \\
&= Q\sum_{n=s}^{\infty} \binom{\kappa + n - 1}{n - s}(1 - q)^{\kappa+s}(q)^{n-s}\epsilon^n \\
&= Q\epsilon^s \left(\frac{1 - q}{1 - q\epsilon}\right)^{\kappa+s} \\
&= Q\epsilon^s(1 - \tilde{q})^{\kappa+s}.
\end{aligned}
$$

Induction.. Suppose u is an ancestral node with two non-null children v and w. The height of u is $(h + 1)$ for some $h \geq 0$: suppose that the induction claim holds for all nodes at heights up to h. Both children have heights at most h, so

$$K_v(s) = Q_v(1 - \tilde{q}_v)^{\kappa_v}(\epsilon_v(1 - \tilde{q}_v))^s \qquad K_w(s) = Q_w(1 - \tilde{q}_w)^{\kappa_w}(\epsilon_w(1 - \tilde{q}_w))^s.$$

Therefore,

$$
\begin{aligned}
C_u(n) &= \left(\sum_{s=0}^{n} \binom{n}{s}(1 - p_v)^s(p_v)^{n-s}K_v(s)\right) \times \left(\sum_{s=0}^{n} \binom{n}{s}(1 - p_w)^s(p_w)^{n-s}K_w(s)\right) \\
&= \left(Q_v(1 - \tilde{q}_v)^{\kappa_v}\right)\left(p_v + (1 - p_v)\epsilon_v(1 - \tilde{q}_v)\right)^n \\
&\quad \times \left(Q_w(1 - \tilde{q}_w)^{\kappa_w}\right)\left(p_w + (1 - p_w)\epsilon_w(1 - \tilde{q}_w)\right)^n \\
&= Q_u(\epsilon_u)^n.
\end{aligned}
$$

$\square$

5.3. Proof of Theorem 4

Proof.   (i) Given the definition of $\tilde{\eta}_u$ and $\tilde{\xi}_u$, the Pigeonhole Principle implies that their maximal value is $m_u = \sum_{v \in \mathcal{L}_u} n_v$, the sum of copy numbers at the leaves descending from u.

(ii) By Equation (7b), the generating function for the conditional distribution of $\tilde{\xi}_u \mid \tilde{\eta}_u$ is

$$
\begin{aligned}
\tilde{F}_s(z) &= \sum_{\ell=0}^{\infty} \mathbb{P}\left\{\tilde{\xi}_u = \ell \mid \tilde{\eta}_u = s\right\}z^{\ell} \\
&= \sum_{n=s}^{\infty} \binom{\kappa_u + n - 1}{n - s}(1 - q_u)^{\kappa_u+s}(q_u)^{n-s} \sum_{i=0}^{n-s} \binom{n - s}{i}(1 - \epsilon_u)^i(\epsilon_u)^{n-s-i}z^{s+i} \\
&= z^s \sum_{k=0}^{\infty} \binom{\kappa_u + s + k - 1}{k}(1 - q_u)^{\kappa_u+s}(q_u)^k(\epsilon_u + (1 - \epsilon_u)z)^k \\
&= z^s \left(\frac{1 - \tilde{q}_u}{1 - \tilde{q}_u z}\right)^{\kappa_u+s},
\end{aligned}
$$

21

where we used $1 - \tilde{q}_u = \frac{1-q_u}{1-q_u\epsilon_u}$. Hence, $(\tilde{\xi}_u - \tilde{\eta}_u)$ has a Pólya distribution with parameters $(\kappa_u + \tilde{\eta}_u)$ and $\tilde{q}_u$:

$$\mathbb{P}\Big\{\tilde{\xi}_u = \ell \,\Big|\, \tilde{\eta}_u = s\Big\} = \binom{\kappa_i + \ell - 1}{\ell - s}(1 - \tilde{q}_u)^{\kappa_u + s}(\tilde{q}_u)^{\ell - s}. \qquad (27)$$

Now we have the recurrences for $\tilde{K}_u$:

$$\tilde{K}_u(s) = \mathbb{P}\Big\{\Xi_u \,\Big|\, \tilde{\eta}_u = s\Big\} = \sum_{\ell \geq s} \tilde{C}_u(\ell) \times \binom{\kappa_u + \ell - 1}{\ell - s}(1 - \tilde{q}_u)^{\kappa_u + s}(\tilde{q}_u)^{\ell - s},$$

as claimed.

(iii) The $\tilde{\xi}_u = \ell$ ancestral copies get sorted in the two child lineages with probabilities $(1-\tilde{p}_v)\tilde{p}_w/(1-\tilde{p}_v\tilde{p}_w)$, $(1-\tilde{p}_w)\tilde{p}_v/(1-\tilde{p}_v\tilde{p}_w)$, and $(1-\tilde{p}_v)(1-\tilde{p}_w)/(1-\tilde{p}_v\tilde{p}_w)$ as conserved only on the left v, only on the right w, or on both sides. Hence,

$$\mathbb{P}\Big\{\tilde{\eta}_v = s \,\Big|\, \tilde{\xi}_u = \ell\Big\} = \binom{\ell}{s}\Big(\frac{1-\tilde{p}_v}{1-\tilde{p}_v\tilde{p}_w}\Big)^s\Big(\tilde{p}_v\frac{1-\tilde{p}_w}{1-\tilde{p}_v\tilde{p}_w}\Big)^{\ell-s} \qquad (28)$$

$$\mathbb{P}\Big\{\tilde{\eta}_v = s \,\Big|\, \tilde{\xi}_u = \ell\Big\} = \binom{\ell}{s}\Big(\frac{1-\tilde{p}_w}{1-\tilde{p}_v\tilde{p}_w}\Big)^s\Big(\tilde{p}_w\frac{1-\tilde{p}_v}{1-\tilde{p}_v\tilde{p}_w}\Big)^{\ell-s}$$

for $0 \leq s \leq \ell$. Define $\tilde{\psi}_u$ as the ancestral copies from $\tilde{\xi}_u$ that survive in both child lineages:

$$\mathbb{P}\Big\{\tilde{\psi}_u = k \,\Big|\, \tilde{\eta}_v = s\Big\} = \binom{s}{k}(1 - \tilde{p}_w)^k(\tilde{p}_w)^{s-k} \qquad (29)$$

$$\mathbb{P}\Big\{\tilde{\psi}_u = k \,\Big|\, \tilde{\eta}_w = s\Big\} = \binom{s}{k}(1 - \tilde{p}_v)^k(\tilde{p}_v)^{s-k}$$

for $0 \leq k \leq s$. The two random variables $\tilde{\eta}_v, \tilde{\eta}_w$ are not independent when conditioned on $\tilde{\xi}_u$, since $\tilde{\eta}_w = \tilde{\xi}_u - \tilde{\eta}_v + \tilde{\psi}_u$:

$$\mathbb{P}\Big\{\tilde{\eta}_v = s, \tilde{\eta}_w = t \,\Big|\, \tilde{\xi}_u = \ell\Big\} = \mathbb{P}\Big\{\tilde{\eta}_v = s \,\Big|\, \tilde{\xi}_u = \ell\Big\}\mathbb{P}\Big\{\tilde{\psi}_u = (s+t) - \ell \,\Big|\, \tilde{\eta}_v = s\Big\}.$$

Combining (28) and (29) gives us the recurrence for $\tilde{C}_u$:

$$\tilde{C}_u(\ell) = \mathbb{P}\Big\{\Xi_u \,\Big|\, \tilde{\xi}_u = \ell\Big\}$$

$$= \sum_{s+t\geq\ell}^{s,t\leq\ell} \mathbb{P}\Big\{\Xi_v \,\Big|\, \tilde{\eta}_v = s\Big\}\mathbb{P}\Big\{\Xi_w \,\Big|\, \tilde{\eta}_w = t\Big\}\mathbb{P}\Big\{\tilde{\eta}_v = s, \tilde{\eta}_w = t \,\Big|\, \tilde{\xi}_u = \ell\Big\}$$

$$= \sum_{s=0}^{\ell}\bigg(\ \tilde{K}_v(s) \times \binom{\ell}{s}\Big(\frac{1-\tilde{p}_v}{1-\tilde{p}_v\tilde{p}_w}\Big)^s\Big(\tilde{p}_v\frac{1-\tilde{p}_w}{1-\tilde{p}_v\tilde{p}_w}\Big)^{\ell-s}$$

$$\times \sum_{k=0}^{s}\tilde{K}_w(\ell - s + k) \times \binom{s}{k}(1 - \tilde{p}_w)^k(\tilde{p}_w)^{s-k}\bigg).$$

The inner sum can be computed in $O(1)$ amortized time by dynamic programming: for all $0 \le d \le \ell$, let

$$\tilde{K}_w^\ell(d) = \sum_{k=0}^{\ell-d} \tilde{K}_w(d+k) \times \binom{\ell-d}{k}(1-\tilde{p})^k \tilde{p}^{\ell-d-k},$$

with $\tilde{p} = \tilde{p}_w$, so that

$$\tilde{C}_u(\ell) = \sum_{s=0}^{\ell} \tilde{K}_v(s) \times \tilde{K}_w^\ell(\ell-s) \times \binom{\ell}{s}\left(\frac{1-\tilde{p}_v}{1-\tilde{p}_v\tilde{p}_w}\right)^s \left(\frac{\tilde{p}_v - \tilde{p}_v\tilde{p}_w}{1-\tilde{p}_v\tilde{p}_w}\right)^{\ell-s}$$

The initial values are
$$\tilde{K}_w^\ell(\ell) = \tilde{K}_w(\ell).$$

Let $s = \ell - d$. Since

$$\binom{s}{k}(1-\tilde{p})^k \tilde{p}^{s-k} = \{k < s\}\tilde{p}\binom{s-1}{k}(1-\tilde{p})^k \tilde{p}^{(s-1)-k}$$
$$+ \{0 < k\}(1-\tilde{p})\binom{s-1}{k-1}(1-\tilde{p})^{k-1}\tilde{p}^{(s-1)-(k-1)},$$

we have the recursions for $d < \ell$:

$$\tilde{K}_w^\ell(d) = \sum_{k=0}^{\ell-d} \tilde{K}_w(d+k) \times \binom{\ell-d}{k}(1-\tilde{p})^k \tilde{p}^{\ell-d-k}$$
$$= \tilde{p} \sum_{k=0}^{\ell-d-1} \tilde{K}_w(d+k)\binom{\ell-d-1}{k}(1-\tilde{p})^k \tilde{p}^{\ell-d-1-k}$$
$$+ (1-\tilde{p}) \sum_{k=1}^{\ell-d} \tilde{K}_w(d+k)\binom{\ell-d-1}{k-1}(1-\tilde{p})^{k-1}\tilde{p}^{(\ell-d-1)-(k-1)}.$$

By setting $d + k = (d+1) + (k-1)$ in the second term,

$$\tilde{K}_w^\ell(d) = \tilde{p}\tilde{K}_w^{\ell-1}(d) + (1-\tilde{p})\tilde{K}_w^\ell(d+1).$$

$\square$

5.4. Proof of Theorem 5

Consider first a resolution into a left-leaning binary tree, and survival in 1, 2, ..., d child subtrees incrementally. The extinction probabilities for general arity are

$$\epsilon_{u,i} = \prod_{j=1}^{i} \underbrace{\left(p_{v_j} + (1-p_{v_j})\epsilon_{v_j}(1-\tilde{q}_{v_j})\right)}_{=\tilde{p}_{v_j}} \quad \text{for all } 1 \le i \le d, \text{ and} \quad \epsilon_u = \epsilon_{u,d}$$

at an ancestral node u with children $uv_1, \ldots, uv_d \in T$ in an arbitrary order.

Theorem 15 (Likelihood recurrence for multifurcating node). *Let u be a node in a degenerate phylogeny with $d \geq 2$ distinct children $uv_1, \ldots, uv_d \in T$ indexed in any order. Define the likelihoods $\tilde{C}_u^i(\ell)$ conditioned on ℓ surviving copies in the subtrees of $v_1, \ldots, v_i$:*

$$\tilde{C}_u^1(\ell) = K_{v_1}(\ell) \qquad \{0 \leq \ell \leq m_{v_1}\} \tag{30a}$$

and, for all $0 \leq i < d$ and for all $0 \leq \ell \leq m_{v_1} + \cdots + m_{v_{i+1}}$

$$\tilde{C}_u^{i+1}(\ell) = \sum_{s=0}^{\min\{\ell, m_{v_1} + \cdots + m_{v_i}\}} \tilde{C}_u^i(s) \times \tilde{K}_{v_{i+1}}^\ell(\ell - s) \tag{30b}$$

$$\times \binom{\ell}{s} \left(\frac{1 - \epsilon_{u,i}}{1 - \epsilon_{u,i+1}}\right)^s \left(\frac{\epsilon_{u,i} - \epsilon_{u,i+1}}{1 - \epsilon_{u,i+1}}\right)^{\ell - s}.$$

Then

$$\tilde{C}_u(\ell) = \tilde{C}_u^d(\ell).$$

Proof. In order to accommodate a multifurcation at u, imagine a resolution of u into $(d-1)$ binary nodes $\{u_2, \ldots, u_d\}$ with $u_d = u$, set $u_1 = v_1$, and define the edges

$$T' = T - \{uv \colon uv \in T\} \cup \{u_i u_{i-1}, u_i v_i\}_{i=2}^d.$$

The corresponding random variables are $\tilde{\xi}_{u_1}, \ldots, \tilde{\xi}_{u_d}$, so that each $\tilde{\xi}_i$ denotes survival in lineages $uv_1, \ldots, uv_i$. Edges $u_i u_{i-1}$ have length 0, so that their distribution parameters are $p = 0$ and $q = 0$. Applying the recurrences of Theorem 4 to the resolved nodes $\tilde{u}_i$ in T' give the recurrences for the multifurcating node (with $\tilde{C}_{u_i} = \tilde{C}_u^i$ in Equations 30). $\square$

Theorem 15 is based on resolving a multifurcation into a left-leaning binary tree. Alternatively, the node can be resolved into a right-leaning binary tree with $(d-2)$ right edges of length 0, giving Theorem 5.

Proof of Theorem 5. The Theorem combines the techniques of Theorems 4 and 15. For Equation (15b), note that $\epsilon_{u,-(i-1)} = \tilde{p}_{v_{i-1}} \epsilon_{u,-i}$ for all $i < d$. $\square$

5.5. Proof of Theorem 6

Proof. At the root, $J_R(s) = \mathbb{P}\{\tilde{\eta}_R = s\} = \{s = 0\}$ by our model. Let u be an arbitrary node, and let $\kappa = \kappa_u, \tilde{q} = \tilde{q}_u$. Using Equation (27) for $\tilde{\xi}_u | \tilde{\eta}_u$,

$$B_u(\ell) = \mathbb{P}\{\Xi - \Xi_u, \tilde{\xi}_u = \ell\}$$

$$= \sum_s \mathbb{P}\left\{\Xi - \Xi_u, \tilde{\xi}_u = \ell \,\middle|\, \tilde{\eta}_u = s\right\} \mathbb{P}\{\tilde{\eta}_u = s\}$$

$$= \sum_s \mathbb{P}\{\Xi - \Xi_u, \tilde{\eta}_u = s\} \mathbb{P}\left\{\tilde{\xi}_u = \ell \,\middle|\, \tilde{\eta}_u = s\right\}$$

$$= \sum_s J_u(s) \binom{\kappa + \ell - 1}{\ell - s} (1 - \tilde{q})^{\kappa + s} (\tilde{q})^{\ell - s}.$$

Now let $uv \in T$ be a parent-child pair, and let $uw \in T$ be the sibling lineage (with $v \neq w$). Since $\Xi - \Xi_v = (\Xi - \Xi_u) \cup \Xi_w$,

$$
\begin{aligned}
J_v(s) &= \mathbb{P}\{\Xi - \Xi_v, \tilde{\eta}_v = s\} \\
&= \sum_\ell \mathbb{P}\Big\{\Xi - \Xi_v, \tilde{\eta}_v = s \;\Big|\; \tilde{\xi}_u = \ell\Big\}\mathbb{P}\{\tilde{\xi}_u = \ell\} \\
&= \sum_\ell \; \mathbb{P}\{\Xi - \Xi_u, \tilde{\xi}_u = \ell\} \times \mathbb{P}\Big\{\tilde{\eta}_v = s \;\Big|\; \tilde{\xi}_u = \ell\Big\} \\
&\qquad \times \sum_t \mathbb{P}\Big\{\tilde{\eta}_w = t \;\Big|\; \tilde{\xi}_u = \ell, \tilde{\eta}_v = s\Big\}\mathbb{P}\Big\{\Xi_w \;\Big|\; \tilde{\eta}_w = t\Big\} \\
&= \sum_\ell B_u(\ell) \times \binom{\ell}{s}\Big(\frac{1 - \tilde{p}_v}{1 - \tilde{p}_v \tilde{p}_w}\Big)^s \Big(\frac{\tilde{p}_v - \tilde{p}_v \tilde{p}_w}{1 - \tilde{p}_v \tilde{p}_w}\Big)^{\ell - s} \times \tilde{K}_w^\ell(\ell - s),
\end{aligned}
$$

where we used Equations (28) and (29) as in the proof of Theorem 4. □

5.6. Proof of Theorem 8

Proof. First, define ϵ_u at all nodes using $\tilde{p}$: if u is a leaf, then $\epsilon_u = 0$, and at an ancestral node u, $\epsilon_u = \prod_{uv \in T} \tilde{p}_v$. Since all $\tilde{p}_u$ are positive, $\epsilon_u > 0$ at every ancestral node u.

Let u be an arbitrary node and let $\tilde{q} = \tilde{q}_u, \tilde{p} = \tilde{p}_u$. Since $0 < (1 - \tilde{q})\epsilon_u < 1$, the equation $\tilde{q} = q_u \frac{1 - \epsilon_u}{1 - q_u \epsilon_u}$ has a unique positive solution

$$
q_u = \frac{\tilde{q}}{1 - (1 - \tilde{q})\epsilon_u} = \frac{\tilde{q}}{\tilde{q} + (1 - \tilde{q})(1 - \epsilon_u)} < 1.
$$

Furthermore, the equation $\tilde{p} = p_u + (1 - p_u)\epsilon_u(1 - \tilde{q})$ has a unique solution

$$
p_u = \frac{\tilde{p} - \epsilon_u(1 - \tilde{q})}{1 - \epsilon_u(1 - \tilde{q})} < 1.
$$

Since $\epsilon_u = \prod_{uv \in T} \tilde{p}_v$, by the assumption of (*), $p_u > 0$. If the assumption is violated by $\tilde{p} < \epsilon_u(1 - \tilde{q})$, then $p_u < 0$, which is illegal.

Since $0 < p_u < 1$ and $0 < q_u < 1$ can be selected at every node, Theorem 1 implies that a corresponding phylogenetic birth-death model exists that is unique up to equivalent rate scalings. □

5.7. Proof of Theorem 9

Proof. By Equations (19) and (17),

$$
\begin{aligned}
\frac{\partial L(\Xi)}{\partial \tilde{q}_u} &= \sum_{\ell=0}^{m_u} \tilde{C}_u(\ell)\frac{\partial B_u(\ell)}{\partial \tilde{q}_u} \\
&= \sum_{\ell=0}^{m_u} \tilde{C}_u(\ell)\frac{\partial}{\partial \tilde{q}_u}\Big(\sum_{s=0}^\ell J_u(s) \times \binom{\kappa_u + \ell - 1}{\ell - s}(1 - \tilde{q}_u)^{\kappa_u + s}(\tilde{q}_u)^{\ell - s}\Big).
\end{aligned}
$$

25

So,

$$\frac{\partial L(\Xi)}{\partial \tilde{q}_u} = \sum_{0 \leq s \leq \ell \leq m_u} \tilde{C}_u(\ell) \times J_u(s)$$

$$\times \binom{\kappa_u + \ell - 1}{\ell - s}(1 - \tilde{q}_u)^{\kappa_u + s}(\tilde{q}_u)^{\ell - s}\left(\frac{\ell - s}{\tilde{q}_u} - \frac{\kappa_u + s}{1 - \tilde{q}_u}\right). \quad (31)$$

By Theorem 3, the empty profile likelihood is $L(0) = \prod_{u=1}^{R}(1 - \tilde{q}_u)^{\kappa_u}$, so

$$\frac{\partial L(0)}{\partial \tilde{q}_u} = -L(0)\frac{\kappa_u}{1 - \tilde{q}_u}.$$

For derivatives with respect to $\tilde{p}_v$ on an edge between a non-root node v and its parent $uv \in T$, consider the recurrences of Theorems 4 and 5. Both can be written as

$$\tilde{C}_u(\ell) = \sum_{s=0}^{\ell} \tilde{K}_v(s) \times \tilde{K}_{-v}^{\ell}(\ell - s) \times \binom{\ell}{s}\left(\frac{1 - \tilde{p}_v}{1 - \tilde{p}_v\epsilon}\right)^s\left(\frac{\tilde{p}_v - \tilde{p}_v\epsilon}{1 - \tilde{p}_v\epsilon}\right)^{\ell - s}.$$

At a binary node u (Theorem 4), $\epsilon = \tilde{p}_w$ with the sibling $uw \in T$, and $\tilde{K}_{-v}^{\ell}(k) = \tilde{K}_w^{\ell}(k)$. If u has more than 2 children $v_1, \ldots, v_d$, then order them so that v is the first, and apply Theorem 5: $\epsilon = \epsilon_{u,-2} = \prod_{j=2}^{d}\tilde{p}_{v_j}$ and $\tilde{K}_{-v}^{\ell}(k) = \tilde{K}_{v_{2..d}}^{\ell}(k)$ from Equation (15b). Hence, using Corollary 7,

$$\frac{\partial L(\Xi)}{\partial \tilde{p}_v} = \sum_{\ell=0}^{m_u} B_u(\ell)\frac{\partial \tilde{C}_u(\ell)}{\partial \tilde{p}_v}$$

$$= \sum_{\ell=0}^{m_u} B_u(\ell) \sum_{s=0}^{\min\{\ell, m_u\}} \tilde{K}_v(s) \times \tilde{K}_{-v}^{\ell}(\ell - s)$$

$$\times \frac{\partial}{\partial \tilde{p}_v}\left(\binom{\ell}{s}\left(\frac{1 - \tilde{p}_v}{1 - \tilde{p}_v\epsilon}\right)^s\left(\frac{\tilde{p}_v - \tilde{p}_v\epsilon}{1 - \tilde{p}_v\epsilon}\right)^{\ell - s}\right).$$

Therefore,

$$\frac{\partial L(\Xi)}{\partial \tilde{p}_v} = \sum_{0 \leq s \leq \ell \leq m_u} B_u(\ell) \times \tilde{K}_v(s) \times \tilde{K}_{-v}^{\ell}(\ell - s)$$

$$\times \binom{\ell}{s}\left(\frac{1 - \tilde{p}_v}{1 - \tilde{p}_v\epsilon}\right)^s\left(\frac{\tilde{p}_v - \tilde{p}_v\epsilon}{1 - \tilde{p}_v\epsilon}\right)^{\ell - s}\left(\frac{\ell - s}{\tilde{p}_v} - \frac{s(1 - \epsilon)}{1 - \tilde{p}_v}\right). \quad (32)$$

The derivatives for the empty profile likelihood are trivial, since $L(0)$ does not depend on any of the $\tilde{p}_v$.

By Corollary 7 and Theorem 6,

$$\frac{\partial L(\Xi)}{\partial \kappa_u} = \sum_{\ell=0}^{m_u} \tilde{C}u(\ell) \sum_{s=0}^{\ell} J_u(s)\frac{\partial}{\partial \kappa_u}\left((1 - \tilde{q}_u)^{\kappa_u + s}(\tilde{q}_u)^{\ell - s}\binom{\kappa_u + \ell - 1}{\ell - s}\right).$$

Since

$$\frac{\partial}{\partial \kappa}\left(\ln\binom{\kappa + \ell - 1}{\ell - s}\right) = \frac{\frac{\partial\binom{\kappa + \ell - 1}{\ell - s}}{\partial \kappa}}{\binom{\kappa + \ell - 1}{\ell - s}},$$

and

$$\frac{\partial}{\partial \kappa}\left(\ln\binom{\kappa+\ell-1}{\ell-s}\right) = \sum_{i=0}^{\ell-s-1}\frac{\partial \ln(\kappa+s+i)}{\partial \kappa} = \sum_{i=s}^{\ell-1}\frac{1}{\kappa+i},$$

we have

$$\frac{\partial\, L(\Xi)}{\partial\, \kappa_u} = \ln(1-\tilde{q}_u) \times L(\Xi)$$

$$+ \sum_{0\le s\le \ell\le m_u} \tilde{C}_u(\ell) \times J_u(s)$$

$$\times \binom{\kappa_u+\ell-1}{\ell-s}(1-\tilde{q}_u)^{\kappa_u+s}(\tilde{q}_u)^{\ell-s}\left(\sum_{i=s}^{\ell-1}\frac{1}{\kappa_u+i}\right).$$

$$= \ln(1-\tilde{q}_u) \times L(\Xi) + \sum_{i=0}^{m_u-1}\frac{1}{\kappa_u+i}\sum_{s=0}^{i}\sum_{\ell=i+1}^{m_u}\tilde{C}_u(\ell)\times J_u(s)\times\binom{\kappa_u+\ell-1}{\ell-s}(1-\tilde{q}_u)^{\kappa_u+s}(\tilde{q}_u)^{\ell-s}$$

$$= \ln(1-\tilde{q}_u) \times L(\Xi) + \sum_{i=0}^{m_u-1}\frac{1}{\kappa_u+i}\left(\mathbb{P}\{\tilde{\xi}_u > i;\Xi\} - \mathbb{P}\{\tilde{\eta}_u > i;\Xi\}\right).$$

For the empty profile,

$$\frac{\partial\, L(0)}{\partial\, \kappa_u} = \frac{\partial}{\partial\, \kappa_u}\left(\prod_{v=1}^{R}(1-\tilde{q}_v)^{\kappa_v}\right) = L(0)\times\ln(1-\tilde{q}_u)$$

$\square$

5.8. Proof of Theorem 11

Proof. Let $u_0 u_1 \cdots u_{d-1}$ denote the path between $u_d = v$ and the root $u_0 = R$ with edges $u_i u_{i+1} \in T$. Since p_v and q_v influence $\tilde{p}_u$ and $\tilde{q}_u$ at $u = v$ and at all the other ancestors $u = u_i$, but not at any other node,

$$\Phi^{(\theta_v)} = \sum_{i=0}^{d}\left(\Phi^{(\tilde{q}_{u_i})}\frac{\partial \tilde{q}_{u_i}}{\partial \theta_v} + \Phi^{(\tilde{p}_{u_i})}\frac{\partial \tilde{p}_{u_i}}{\partial \theta_v}\right).$$

Recall the definitions $\tilde{q}_v = q_v\frac{1-\epsilon_v}{1-q_v\epsilon_v}$ and $\tilde{p}_v = \frac{p_v(1-\epsilon_v)+\epsilon_v(1-q_v)}{1-q_v\epsilon_v}$ (substituting $p_R = 0$ at the root). We have thus

$$\frac{\partial \tilde{p}_v}{\partial p_v} = \frac{1-\epsilon_v}{1-q_v\epsilon_v} \qquad\qquad \frac{\partial \tilde{q}_v}{\partial p_v} = 0$$

$$\frac{\partial \tilde{p}_v}{\partial q_v} = \frac{-(1-p_v)\epsilon_v(1-\epsilon_v)}{(1-q_v\epsilon_v)^2} \qquad\qquad \frac{\partial \tilde{q}_v}{\partial q_v} = \frac{1-\epsilon_v}{(1-q_v\epsilon_v)^2}$$

$$\frac{\partial \tilde{p}_v}{\partial \epsilon_v} = \frac{(1-p_v)(1-q_v)}{(1-q_v\epsilon_v)^2} \qquad\qquad \frac{\partial \tilde{q}_v}{\partial \epsilon_v} = \frac{-q_v(1-q_v)}{(1-q_v\epsilon_v)^2}.$$

(i) If $v = R$ is the root, then

$$\Phi^{(q_R)} = \Phi^{(\tilde{p}_R)}\frac{\partial \tilde{p}_R}{\partial q_R} + \Phi^{(\tilde{q}_R)}\frac{\partial \tilde{q}_R}{\partial q_R} = -\Phi^{(\tilde{p}_R)}\frac{\epsilon_R(1-\epsilon_R)}{(1-q_R\epsilon_R)^2} + \Phi^{(\tilde{q}_R)}\frac{1-\epsilon_R}{(1-q_R\epsilon_R)^2},$$

27

and, for $R > 1$,

$$\Phi^{(\epsilon_R)} = \Phi^{(\tilde{p}_R)} \frac{\partial \tilde{p}_R}{\partial \epsilon_R} + \Phi^{(\tilde{q}_R)} \frac{\partial \tilde{q}_R}{\partial \epsilon_R} = \Phi^{(\tilde{p}_R)} \frac{1 - q_R}{(1 - q_R \epsilon_R)^2} - \Phi^{(\tilde{q}_R)} \frac{q_R(1 - q_R)}{(1 - q_R \epsilon_R)^2},$$

as claimed in (21).

(ii) Now suppose that v is not the root. At any ancestor u_i with $i < d$, the distribution parameters of v affect the extinction probability ϵ_{u_i}. For a distribution parameter $\theta_v = p_v$, $\theta_v = q_v$, or $\theta_v = \epsilon_v$,

$$\frac{\partial \tilde{p}_{u_i}}{\partial \theta_v} = \frac{\partial \tilde{p}_{u_i}}{\partial \epsilon_{u_i}} \frac{\partial \epsilon_{u_i}}{\partial \tilde{p}_{u_{i+1}}} \frac{\partial \tilde{p}_{u_{i+1}}}{\partial \theta_v} \quad \text{and} \quad \frac{\partial \tilde{q}_{u_i}}{\partial \theta_v} = \frac{\partial \tilde{q}_{u_i}}{\partial \epsilon_{u_i}} \frac{\partial \epsilon_{u_i}}{\partial \tilde{p}_{u_{i+1}}} \frac{\partial \tilde{p}_{u_{i+1}}}{\partial \theta_v},$$

with

$$\frac{\partial \epsilon_{u_i}}{\partial \tilde{p}_{u_{i+1}}} = \frac{\partial}{\partial \tilde{p}_{u_{i+1}}} \prod_{u_i w \in T} \tilde{p}_w = \frac{\epsilon_{u_i}}{\tilde{p}_{u_{i+1}}}$$

Let $u = u_{d-1}$ be the parent of $v = u_d$. Since $\frac{\partial \tilde{q}_v}{p_v} = 0$,

$$\Phi^{(p_v)} = \left(\Phi^{(\tilde{p}_v)} + \Phi^{(\epsilon_u)} \frac{\partial \epsilon_u}{\partial \tilde{p}_v} \right) \frac{\partial \tilde{p}_v}{\partial p_v} = \left(\Phi^{(\tilde{p}_v)} + \epsilon \Phi^{(\epsilon_u)} \right) \frac{1 - \epsilon_v}{1 - q_v \epsilon_v}$$

with $\epsilon = \frac{\epsilon_u}{\tilde{p}_v} = \frac{\epsilon_{u_{d-1}}}{\tilde{p}_{u_d}}$. The other two recurrences include $\Phi^{(\tilde{q}_v)}$, as well:

$$\Phi^{(q_v)} = \left(\Phi^{(\tilde{p}_v)} + \Phi^{(\epsilon_u)} \frac{\partial \epsilon_u}{\partial \tilde{p}_v} \right) \frac{\partial \tilde{p}_v}{\partial q_v} + \Phi^{(\tilde{q}_v)} \frac{\partial \tilde{q}_v}{\partial q_v}$$

$$= \left(\Phi^{(\tilde{p}_v)} + \epsilon \Phi^{(\epsilon_u)} \right) \frac{-(1 - p_v) \epsilon_v (1 - \epsilon_v)}{(1 - q_v \epsilon_v)^2} + \Phi^{(\tilde{q}_v)} \frac{1 - \epsilon_v}{(1 - q_v \epsilon_v)^2};$$

$$\Phi^{(\epsilon_v)} = \left(\Phi^{(\tilde{p}_v)} + \epsilon \Phi^{(\epsilon_u)} \right) \frac{\partial \tilde{p}_v}{\partial \epsilon_v} + \Phi^{(\tilde{q}_v)} \frac{\partial \tilde{q}_v}{\partial \epsilon_v}$$

$$= \left(\Phi^{(\tilde{p}_v)} + \epsilon \Phi^{(\epsilon_u)} \right) \frac{(1 - p_v)(1 - q_v)}{(1 - q_v \epsilon_v)^2} - \Phi^{(\tilde{q}_v)} \frac{q_v(1 - q_v)}{(1 - q_v \epsilon_v)^2},$$

as shown in (22).

$\square$

5.9. No-duplication model

Proof of Theorem 12. Let C_u, K_u denote the likelihoods for the empty profile:

$$C_u(n) = \mathbb{P}\left\{ \forall v \in \mathcal{L}_u \colon \xi_v = 0 \,\Big|\, \xi_u = n \right\}$$

$$K_u(s) = \mathbb{P}\left\{ \forall v \in \mathcal{L}_u \colon \xi_v = 0 \,\Big|\, \eta_u = s \right\}.$$

Let Q_u denote the product of $e^{-\tilde{r}_v}$ across all edges in the subtree of u: $Q_u = 1$ at a leaf, and at an ancestral node u with children v, w

$$Q_u = \left(Q_v e^{-\tilde{r}_v} \right) \left(Q_w e^{-\tilde{r}_w} \right).$$

28

We prove that for all nodes u,

$$K_u(s) = Q_u \times (\epsilon_u)^s \times e^{-\tilde{r}_u} \qquad C_u(n) = Q_u \times (\epsilon_u)^n.$$

(With $0^0 = 1$ and $0^n = 0$ for $n > 0$.) In particular, at the root R, $L(0) = \tilde{K}_R(0)$.

We adjust the induction proof of Theorem 3. At any node u, with $r = r_u, \epsilon = \epsilon_u$ and $Q = Q_u$, by Equation (23),

$$K_u(s) = \sum_{k=0}^{\infty} e^{-r} \frac{r^k}{k!} Q\epsilon^{s+k} = Q\epsilon^s e^{-r} \sum_{k=0}^{\infty} \frac{(r\epsilon)^k}{k!}$$
$$= Q\epsilon^s e^{-r(1-\epsilon)} = Q\epsilon^s e^{-\tilde{r}}$$

with $\tilde{r} = r(1 - \epsilon)$. The inductive case for C_u is adjusted:

$$C_u(n) = \prod_{uv \in T} \left(Q_v e^{-\tilde{r}_v} \big(p_v + (1 - p_v)\epsilon_v \big)^n \right) = Q_u(\epsilon_u)^n.$$

$\square$

Proof of Theorem 13. Since $\tilde{p}_u = p_u + (1 - p_u)\epsilon_u$ and $\tilde{r}_u = r_u(1 - \epsilon_u)$,

$$\frac{\partial \tilde{p}_u}{\partial p_u} = 1 - \epsilon_u \qquad\qquad \frac{\partial \tilde{r}_u}{\partial r_u} = 1 - \epsilon_u$$

$$\frac{\partial \tilde{p}_u}{\partial \epsilon_u} = 1 - p_u \qquad\qquad \frac{\partial \tilde{r}_u}{\partial \epsilon_u} = -r_u.$$

The rest of the proof is based on applications of the chain rule as in the proof of Theorem 11. $\square$

5.10. Proof of Theorem 14

Proof. Let $N = \sum_{w \in \mathcal{L}} n_w = m_R$ be the sum of copy numbers across the leaves. At an ancestral node u, the calculations of $\tilde{C}_u(\ell)$ for all $0 \le \ell \le m_u$ and of $J_u(s)$ for all $0 \le s \le m_u$ take $(1 + m_v)(1 + m_w)$ iterations. Calculating $\tilde{K}(s)$ for all $0 \le s \le m_u$ and $B_u(\ell)$ for all $0 \le \ell \le m_u$ is done in $(1+m_u)(2+m_u)/2$ iterations. The total running time can be thus bounded asymptotically as $O(\sum_{u=1}^{R} m_u^2)$, or as $O(R) = O(L)$ if $N^2 < R$, the number of nodes. Summing by the height of the nodes $h(u)$,

$$\sum_{u=1}^{R} m_u^2 = \sum_{i=0}^{h-1} \sum_{u\,:\,h(u)=i} (m_u)^2 = \sum_{i=0}^{h-1} \sum_{u\,:\,h(u)=i} \left(\sum_{v \in \mathcal{L}_u} n_v \right)^2$$
$$\le \sum_{i=0}^{h-1} \left(\sum_{u\,:\,h(u)=i} \sum_{v \in \mathcal{L}_u} n_v \right)^2$$
$$\le \sum_{i=0}^{h-1} \left(\sum_{v \in \mathcal{L}} n_v \right)^2 = hN^2.$$

For the last inequality, note that if $h(v) = h(w)$ then their subtrees do not intersect and $\mathcal{L}_v \cap \mathcal{L}_w = \emptyset$. $\square$

5.11. An old algorithm for computing the profile likelihood

The recursive algorithm of [3, 12] for computing the profile likelihood uses the basic birth-death transitions from (2), and arrives at a set of recurrences by combinatorial principles. We can infer the same method algebraically in the present framework — the resulting formulas are not useful beyond serving up $\tilde{C}$. As a warmup, we extract the recurrences for transition probabilities from Theorem 2.

Corollary 16 (Transition probability recurrences)**.** *Let $uv \in T$ be any edge and* $w(m \mid n) = \mathbb{P}\Big\{\xi_v = m \ \Big| \ \xi_u = n\Big\}$ *denote the transition probabilities.*

For $\lambda_v > 0$, let $p = p_v, q = q_v, \kappa = \kappa_v$ denote the applicable distribution parameters from Equation (3). Then

$$w(m \mid 0) = \binom{\kappa + m - 1}{m}(1 - q)^{\kappa}q^m$$

$$\begin{aligned} w(m \mid n) = \quad &qw(m \mid n - 1) &\{n > 0\} \\ &+\{m > 0\}(1 - p - q)w(m - 1 \mid n - 1) \\ &+\{m > 0\}qw(m - 1 \mid n). \end{aligned}$$

For $\lambda = 0$, let $p = p_v, r = r_v$ denote the applicable distribution parameters from Equation (3). Then

$$w(m \mid 0) = e^{-r}\frac{r^m}{m!}$$

$$w(m \mid n) = pw(m \mid n - 1) + \{m > 0\}(1 - p)w(m - 1 \mid n - 1) \quad \{n > 0\}$$

Proof. First, let $\lambda > 0$. By Theorem 2, the generating function for the transition probabilities is

$$\begin{aligned} G_n(z) &= \sum_{m=0}^{\infty} w(m \mid n)z^m \\ &= \Big(\frac{1 - q}{1 - qz}\Big)^{\kappa}\Big(p + (1 - p)\frac{(1 - q)z}{1 - qz}\Big)^n \\ &= \Big(\frac{1 - q}{1 - qz}\Big)^{\kappa}\Big(\frac{p + z(1 - p - q)}{1 - qz}\Big)^n. \end{aligned}$$

The generating function satisfies

$$G_n(z) \times (1 - qz) = G_{n-1}(z) \times (p + z(1 - p - q)).$$

Noting that $zG_n(z) = \sum_{m=1}^{\infty} w(m - 1 \mid n)z^m$, the equality of the coefficients implies that

$$w(m \mid n) - qw(m - 1 \mid n) = pw(m \mid n - 1) + (1 - p - q)w(m - 1 \mid n - 1),$$

which is the Theorem's recurrence.

For $\lambda = 0$, the generating function is

$$G_n(z) = e^{-r(z-1)}\big(p + (1 - p)z\big)^n,$$

so $G_n(z) = G_{n-1}(z)\big(p + (1 - p)z\big)$, giving the recurrence. $\qquad\square$

The profile likelihood algorithm of [3, 12] combines the recurrences of Theorem 5, bypassing the explicit representation of conserved ancestral copies $\tilde{\eta}$. Consider an ancestral node u with children $v_1, \ldots, v_d$, and the step for computing $\tilde{C}_u^{-(i-1)}(\ell)$ for some $0 < i \leq d$ from Equation (15). Let $\tilde{p} = \tilde{p}_{v_{i-1}}$, and $\epsilon = \epsilon_{u,-(i)} = \prod_{j=i}^{d} \tilde{p}_{v_j}$:

$$\tilde{C}_u^{-(i-1)}(\ell) = \sum_{j+k=\ell} \sum_{b=0}^{j} \tilde{K}_{v_{i-1}}(j) \times \tilde{C}_u^{-(i)}(k+b)$$

$$\times \binom{j+k}{j} \left(\frac{1-\tilde{p}}{1-\tilde{p}\epsilon}\right)^j \left(\frac{\tilde{p}-\tilde{p}\epsilon}{1-\tilde{p}\epsilon}\right)^k \binom{j}{b} (1-\epsilon)^b \epsilon^{j-b}.$$

Set $t = k + b$ and $s = j - b$ with $\binom{j+k}{j}\binom{j}{b} = \binom{s+t}{t}\binom{t}{b}$:

$$\tilde{C}_u^{-(i-1)}(\ell) = (1-\tilde{p}\epsilon)^{-\ell} \sum_{s+t=\ell} \tilde{C}_u^{-(i)}(t) \times \binom{\ell}{t}(1-\epsilon)^t \epsilon^s$$

$$\times \underbrace{\left(\sum_{b=0}^{t} \tilde{K}_{v_{i-1}}(s+b) \times \binom{t}{b}(1-\tilde{p})^{s+b}\tilde{p}^{t-b}\right)}_{=D_{v_{i-1}}^{s+t}(s)}.$$

The inner sum resembles $\tilde{K}^{s+t}(t)$ in Theorems 4 and 5, but the mixing coefficients differ ($\tilde{K}$ uses ϵ, and D uses $\tilde{p}$). The recurrence for $D_{v_{i-1}}^{s+t}(s) = \sum_{b=0}^{t} \tilde{K}_{v_i}(s+b) \times \binom{t}{b}(1-\tilde{p}_{v_i})^{s+b}(\tilde{p}_{v_i})^{t-b}$ with $t > 0$ is

$$D_{v_{i-1}}^{\ell}(s) = D_{v_{i-1}}^{\ell}(s+1) + \tilde{p}D_{v_{i-1}}^{\ell-1}(s). \tag{33a}$$

The starting values are $D_{v_{i-1}}^{s}(s) = (1-\tilde{p})^s \tilde{K}_{v_{i-1}}(s)$, which, by (12), further expand into

$$D_{v_i}^{s}(s) = \sum_{\ell \geq s} \tilde{C}_{v_i}(\ell) \times \underbrace{(1-\tilde{p})^s \binom{\kappa+\ell-1}{\ell-s}(1-\tilde{q})^{\kappa+s}\tilde{q}^{\ell-s}}_{=w^*(\ell|s)}$$

with $\tilde{q} = \tilde{q}_{v_i}$ and $\kappa = \kappa_{v_i}$. The transition weights are

$$w^*(\ell \mid s) = \mathbb{P}\left\{\tilde{\xi}_{v_i} = \ell, \tilde{\eta}_{v_i} = s \,\middle|\, \tilde{\xi}_u = s\right\},$$

satisfying

$$w^*(\ell \mid 0) = \binom{\kappa+\ell-1}{\ell}(1-\tilde{q})^{\kappa}\tilde{q}^{\ell} = h_{\ell}(t_{uv_i}), \tag{33b}$$

and, for all $1 \leq s \leq \ell$,

$$w^*(\ell \mid s) = \{s < \ell\}\tilde{q}w^*(\ell-1 \mid s) + (1-\tilde{p})(1-\tilde{q})w^*(\ell-1 \mid s-1) \tag{33c}$$

31

with the basic gain transitions h_ℓ from (2).

$$D_{v_i}^s(s) = \sum_{\ell \geq s} \tilde{C}_{v_i}(\ell) \times w^*(\ell \mid s). \tag{33d}$$

We have thus the formulas relating $\tilde{C}$ between themselves:

$$\tilde{C}_u^{-(i-1)}(\ell) = (1 - \tilde{p}\epsilon)^{-\ell} \sum_{t=0}^{\ell} \tilde{C}_u^{-i}(t) \times D_{v_i}^\ell(\ell - t) \times \binom{\ell}{t}(1 - \epsilon)^t \epsilon^{\ell - t}. \tag{33e}$$

Equations (33) form the basis of the algorithm reported in [3].

Theorem 17 (An old algorithm for the profile likelihood). *Let $\Xi = \{\xi_u = n_u \mid u \in \mathcal{L}\}$ be an arbitrary profile with an average of $\bar{n} = \frac{1}{L}\sum_{u \in \mathcal{L}} n_u$ copies across $L = |\mathcal{L}|$ leaves. The profile likelihood can be computed in a postorder traversal of the phylogeny, using Equations (33a)–(33e), as*

$$L(\Xi) = \sum_{\ell \geq 0} \tilde{C}_R(\ell) \times \binom{\kappa_R + \ell - 1}{\ell}(1 - \tilde{q}_R)^{\kappa_R}(\tilde{q}_R)^\ell.$$

with at the root R. The computations take $O(hL(L\bar{n}^2 + 1))$ time if the phylogeny height is h.

Acknowledgments

References

[1] W. M. Fitch, Homology a personal view on some of the problems, Trends in Genetics 16 (5) (2000) 227–231.

[2] R. K. Aziz, et al., The RAST server: Rapid Annotations using Subsystems Technology, BMC Genomics 9 (2008) 75. `doi:10.1186/1471-2164-9-75`.

[3] M. Csűrös, I. Miklós, Streamlining and large ancestral genomes in Archaea inferred with a phylogenetic birth-and-death model, Molecular Biology and Evolution 26 (9) (2009) 2087–2095. `doi:10.1093/molbev/msp123`.

[4] G. Dey, T. Meyer, Phylogenetic profiling for probing the modular architecture of the human genome, Cell Systems 1 (2015) 106–115. `doi: 10.1016/j.cels.2015.08.006`.

[5] T. M. W. Nye, Modelling the evolution of multi-gene families, Statistical Methods in Medical Research 18 (2009) 487–504. `doi:10.1177/ 0962280208099450`.

[6] L. Takács, Introduction to the Theory of Queues, Oxford University Press, New York, 1962.

[7] D. G. Kendall, Stochastic processes and population growth, Journal of the Royal Statistical Society Series B 11 (2) (1949) 230–282.

[8] J. Felsenstein, Maximum likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters, Systematic Zoology 22 (3) (1973) 240–249.

[9] M. W. Hahn, T. De Bie, J. E. Stajich, C. Nguyen, N. Cristianini, Estimating the tempo and mode of gene family evolution from comparative genomic data, Genome Research 15 (2005) 1153–1160. `doi:10.1101/gr.3567505`.

[10] W. Iwasaki, T. Takagi, Reconstruction of highly heterogeneous gene-content evolution across the three domains of life, Bioinformatics 23 (13) (2007) i230–i239. `doi:10.1093/bioinformatics/btm165`.

[11] M. Csűrös, I. Miklós, A probabilistic model for gene content evolution with duplication, loss, and horizontal transfer, Springer Lecture Notes in Bioinformatics 3909 (2006) 206–220, proc. Tenth Annual International Conference on Research in Computational Molecular Biology (RECOMB). `doi:10.1007/11732990_18`.

[12] M. Csűrös, Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood, Bioinformatics 26 (15) (2010) 1910–1912. `doi:10.1093/bioinformatics/btq315`.

[13] I. Sela, Y. I. Wolf, E. V. Koonin, Theory of prokaryotic evolution, Proceedings of the National Academy of Sciences of the USA 113 (2016) 11399–11407. `doi:10.1073/pnas.1614083113`.

[14] J. L. Thorne, H. Kishino, J. Felsenstein, An evolutionary model for maximum likelihod alignment of DNA sequences, Journal of Molecular Evolution 33 (1991) 114–124.

[15] S. Karlin, J. McGregor, Linear growth, birth, and death processes, Journal of Mathematics and Mechanics 7 (4) (1958) 643–662.

[16] J. Felsenstein, Phylogenies from restriction sites, a maximum likelihood approach, Evolution 46 (1992) 159–173.

Contents