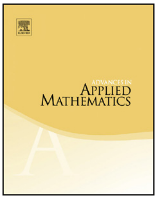
Advances in Applied Mathematics 153 (2024) 102612

Contents lists available at ScienceDirect

Advances in Applied Mathematics

journal homepage: www.elsevier.com/locate/yaama



Bijections between the multifurcating unlabeled rooted trees and the positive integers

Alessandra Rister Portinari Maranca [a], Noah A. Rosenberg [b,*]

[a] *Department of Mathematics, Stanford University, United States of America*
[b] *Department of Biology, Stanford University, United States of America*

ARTICLE INFO

Article history:
Received 3 April 2023
Received in revised form 24 August 2023
Accepted 27 August 2023
Available online xxxx

MSC:
05C05
05C30
92D15

Keywords:
Bifurcation
Multifurcation
Phylogenetics
Recursion
Unlabeled trees

ABSTRACT

Colijn and Plazzotta (2018) [1] described a bijective scheme for associating the unlabeled bifurcating rooted trees with the positive integers. In mathematical and biological applications of unlabeled rooted trees, however, nodes of rooted trees are sometimes multifurcating rather than bifurcating. Building on the bijection between the unlabeled bifurcating rooted trees and the positive integers, we describe bijective schemes for associating the unlabeled multifurcating rooted trees with the positive integers. We devise bijections with the positive integers for a set of trees in which each non-leaf node has exactly k child nodes, and for a set of trees in which each non-leaf node has at most k child nodes. The calculations make use of Macaulay's binomial expansion formula. The generalization to multifurcating trees can assist with the use of unlabeled trees for applications in evolutionary biology, such as the measurement of phylogenetic patterns of genetic lineages in pathogens.

* Corresponding author.
E-mail addresses: arpm@stanford.edu (A.R.P. Maranca), noahr@stanford.edu (N.A. Rosenberg).

https://doi.org/10.1016/j.aam.2023.102612
0196-8858/© 2023 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

In mathematical and statistical phylogenetics, the properties of evolutionary trees are used to make inferences about the processes that have given rise to those trees [2,11]. Each tree examined in a data set is an element of some class of trees, and the mathematics pertaining to that class suggests quantities that can be informatively measured for the trees contained in the class. Often, for a label set X containing n labels, the class of trees of interest is the set of bifurcating *labeled topologies* associated with the label set. For example, if an investigator seeks to understand the evolutionary relationships among n named species in a taxonomic group, the phylogenetic problem of interest is to identify from data the appropriate labeled topology among the $(2n-3)!!$ possibilities.

For many phylogenetic problems, it is not the labels of specific lineages that are of interest, but rather, the *shape* of the evolutionary tree [6,9]. For example, does the tree shape fit a model in which each lineage is equally likely to be the next to split into descendant lineages? If not, what biological features of lineages produce differences in the rates at which lineages split? Are the rates affected by abiotic factors, such as features of the environments that the lineages inhabit? For such questions, it is not the labeled topologies that are of interest, but rather, the *unlabeled* topologies.

The importance of the unlabeled topologies has grown with the proliferation of phylodynamic studies—in which fast-evolving genetic sequences from pathogens are investigated using evolutionary trees [4,7]. When many genetic sequences are collected across large numbers of hosts on influenza, SARS-CoV-2, or other pathogens, the placement of specific sequences in the tree is of less interest than what the tree shape reveals about transmission chains, clusters of epidemiologically related cases, and the processes that spread the pathogen. Hence, although unlabeled topologies have long been examined as mathematical objects [2,11], phylodynamic studies have given rise to new quantities that can be computed from them as statistics for measuring their properties [1,5].

Colijn & Plazzotta [1] introduced a variety of statistics for unlabeled topologies — bifurcating unlabeled rooted trees — focusing on metrics for comparing pairs of unlabeled topologies. Underlying some of their statistics is a bijection they devised between the set of unlabeled topologies with $n \geqslant 1$ leaves and the positive integers. Each unlabeled topology t is associated with a positive integer that is computed recursively from the integers associated with the immediate subtrees of its root. In reverse, given an integer, the left and right subtrees associated with that integer are calculated, uniquely identifying the associated unlabeled topology. The "distance" between unlabeled topologies is the absolute difference between their associated integers. Rosenberg [8] then studied the bijection between unlabeled topologies and positive integers, characterizing the unlabeled topologies associated with the smallest and largest values among the integers associated with unlabeled topologies of n leaves, and investigating the asymptotic growth of those quantities.

The Colijn–Plazzotta bijection and distance metric assume bifurcating trees. However, in scenarios in which the rate of divergence of evolutionary lineages is large compared to the rates of evolution along lineages individually, it is natural to instead consider multifurcating trees. In such cases, a sequence of bifurcations might be indistinguishable from a multifurcation event if insufficient time has occurred between bifurcations for mutations to accumulate. Additionally, in cases in which multiple genealogical lineages of interest trace to a single parent — such as in an instantaneous diversification of a lineage into multiple descendants — a tree has a genuine multifurcation.

Here, we introduce bijections between unlabeled *multifurcating* topologies and positive integers. We consider multifurcating trees in two different ways. First, for fixed $k \geqslant 2$, we consider the class of unlabeled multifurcating topologies in which each internal node possesses exactly k immediate descendants (*strict* k-furcation). Next, for fixed $k \geqslant 2$, we consider the class of unlabeled multifurcating topologies in which internal nodes can vary in their numbers of immediate descendants, with each node possessing at least two and *at most* k immediate descendants. The bijections between unlabeled multifurcating topologies and positive integers can be used in comparing trees in a manner analogous to the use by Colijn & Plazzotta [1] of the bijection for bifurcating topologies.

In Section 2, we recall the Colijn-Plazzotta ranking scheme for bifurcating trees. Next, in Section 3, as a prelude to the general case of strictly k-furcating trees, we extend the scheme to strictly trifurcating trees; in Section 4, we complete the generalization. In Section 5, we consider trees that are at-most-k-furcating. We conclude with a discussion in Section 6.

2. Bifurcating trees

We begin by recalling the Colijn–Plazzotta scheme for assigning ranks to bifurcating unlabeled rooted trees [1,8]. Let B_n be the set of bifurcating unlabeled rooted trees with exactly n leaves. For $n = 1$, B_n has a single unlabeled tree with one leaf. Let $B = \cup_{n=1}^{\infty} B_n$ be the set of bifurcating unlabeled rooted trees, considering all possible numbers of leaves. We write $B^* = B \setminus B_1$.

For a tree $b \in B$, let $m : B \to \mathbb{Z}^+$ be the function that yields the number of leaves of b. Let $s : B^* \to B \times B$ be a function that extracts a vector containing the immediate subtrees of the root of a tree (in a canonical order that we describe shortly). We abbreviate by b_1 and b_2 the first and second coordinates of $s(b)$.

Definition 2.1. The Colijn–Plazzotta ranking $f : B \to \mathbb{Z}^+$ for bifurcating unlabeled rooted trees is a function that satisfies

$$f(b) = \begin{cases} 1 & \text{if } m(b) = 1 \\ \frac{1}{2} f(b_1)[f(b_1) - 1] + f(b_2) + 1 & \text{if } m(b) > 1. \end{cases}$$

We shall abbreviate the Colijn–Plazzotta ranking as the CP ranking. To determine the CP rank of a tree, we require it to be written in a canonical form in which $f(b_1) \geqslant f(b_2)$. The 1-leaf tree has CP rank 1; hence, if it is an immediate subtree of the root of b and

$m(b) \geqslant 3$, then because it has the smallest CP rank among all trees, it is necessarily in the second coordinate of $s(b)$. For convenience, we draw all trees b such that b_1 is the left-hand subtree and b_2 is the right-hand subtree. This notation will be generalized for larger trees.

That Definition 2.1 describes a bijection and yields an inverse function is proven by Colijn & Plazzotta [1] and Rosenberg [8]. For $v = 1$, the inverse function $f^{-1} : \mathbb{Z}^+ \to B$ is the tree with one leaf, and for $v \geqslant 2$, $f^{-1}(v)$ is the tree in B whose two subtrees have ranks

$$\Big(k_1(v), k_2(v)\Big) = \left(\left\lceil \frac{\sqrt{8v-7}-1}{2} \right\rceil, v - \frac{k_1(v)[k_1(v)-1]}{2} - 1\right). \tag{2.1}$$

We can also examine the function $m : B \to \mathbb{Z}^+$ that gives the number of leaves in the tree with specified CP rank. The function m satisfies $m\big(f^{-1}(1)\big) = 1$ and, for $v \geqslant 2$,

$$\begin{aligned} m\big(f^{-1}(v)\big) = m\left(f^{-1}\left(\left\lceil \frac{\sqrt{8v-7}-1}{2} \right\rceil\right)\right) \\ + m\left(f^{-1}\left(v - \left\lceil \frac{\sqrt{8v-7}-1}{2} \right\rceil \left\lceil \frac{\sqrt{8v-7}-3}{2} \right\rceil \Big/ 2 - 1\right)\right). \end{aligned}$$

For bifurcating trees with small ranks, Fig. 1 displays the trees along with their ranks and the ranks of their subtrees.

3. Trifurcating trees

To prepare for the general k-furcating case, we now consider trifurcating trees, in which each internal node possesses exactly three immediate descendant nodes. Let T_n be the set of trifurcating unlabeled rooted trees with exactly n leaves. As was true for the bifurcating case, for the trivial tree with $n = 1$, T_n consists of a single unlabeled tree with one leaf. Let $T = \cup_{n=1}^{\infty} T_n$ be the set of trifurcating unlabeled rooted trees, considering all possible numbers of leaves, and let $T^* = T \setminus T_1$. Note that for even values of n, T_n is empty, as no strictly trifurcating tree can possess an even number of leaves.

We continue to use the notation m and s for concepts analogous to those of the bifurcating case. For a tree $t \in T$, let $m : T \to \mathbb{Z}^+$ count the number of leaves of t. Let $s : T^* \to T \times T \times T$ denote the vector that extracts the immediate subtrees of the root of a tree. We abbreviate by t_1, t_2, t_3 the first, second, and third coordinates of $s(t)$, in a canonical order discussed below.

3.1. Ranking scheme

We seek to find a bijection $f : T \to \mathbb{Z}^+$ between trifurcating trees and positive integers. We accomplish this task in a manner analogous to that used in the bifurcating case. In

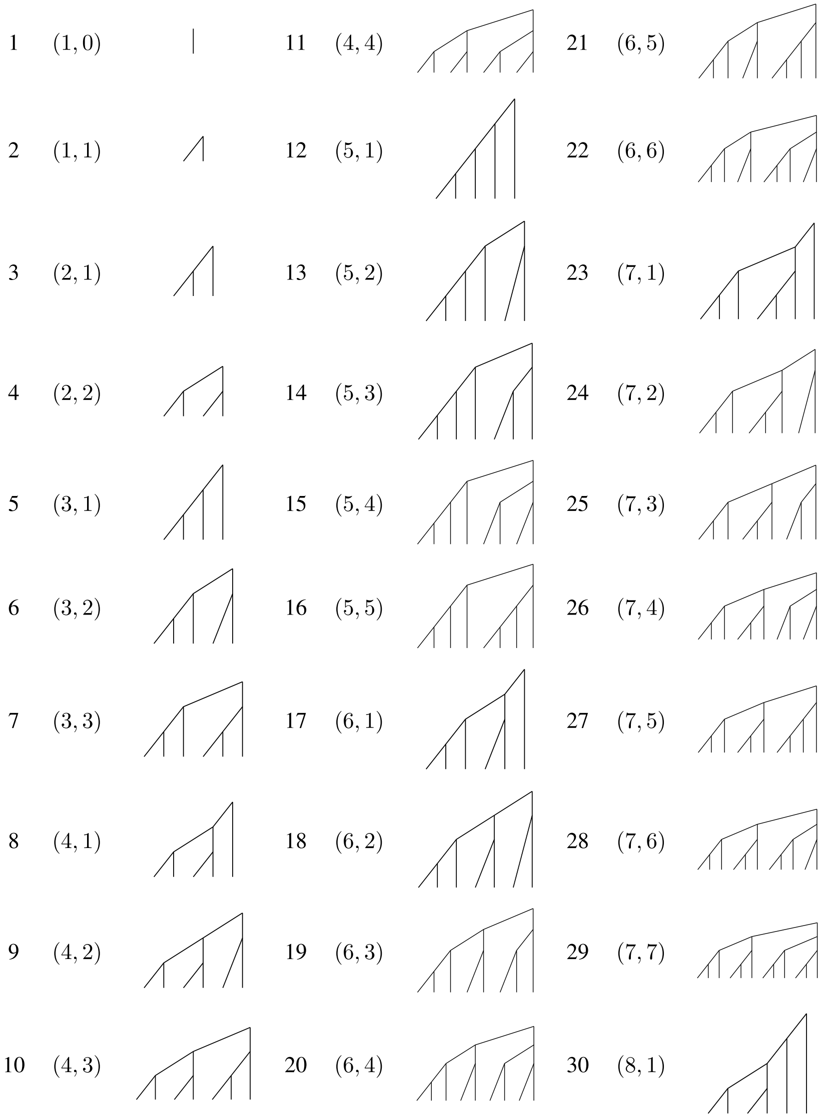


Fig. 1. The Colijn–Plazzotta ranking for bifurcating rooted trees for small ranks. For ranks v from 1 to 30, the tree b with CP rank v is shown, as is the ordered pair of CP ranks of the subtrees, $(f(b_1), f(b_2))$. Each tree is drawn in canonical form, so that the rank associated with the left-hand subtree is greater than or equal to the rank associated with the right-hand subtree. The ordered pair $\big(k_1(v), k_2(v)\big)$ is obtained from eq. (2.1), and the tree is obtained by recursive application of eq. (2.1).

that case, we devised a polynomial that is quadratic in the rank of the first subtree and linear in that of the second subtree. With three subtrees, we increase the degree of the polynomial by one, and compute a cubic term for the first subtree, a quadratic term for the second, and a linear term for the third.

Consider a tree $t = (t_1, t_2, t_3)$. We assume a canonical form in which $f(t_1) \geqslant f(t_2) \geqslant f(t_3)$; this form places a dictionary order on trees. To assign the rank $f(t)$, we must sum three quantities: (1) the number of nontrivial trees in T whose first subtree has rank less than $f(t_1)$; (2) the number of nontrivial trees in T whose first subtree has rank $f(t_1)$ and second subtree has rank less than $f(t_2)$; and (3) the number of nontrivial trees in T whose first subtree has rank $f(t_1)$, second subtree has rank $f(t_2)$, and third subtree has rank less than $f(t_3)$. We then assign t the rank equal to the sum of the quantities in (1), (2), and (3), plus 2. The +2 assigns the next available rank to t, accounting for the trivial tree with $n = 1$; we can view the trivial tree as having subtrees with ranks $(1, 0, 0)$.

For quantity (3), the number of nontrivial trees in T whose first subtree has rank $f(t_1)$, whose second tree has rank $f(t_2)$, and whose third subtree has rank less than $f(t_3)$, we count all trees with subtrees (t_1, t_2, y_3), where y_3 ranges from 1 to $f(t_3) - 1$. The number of such trees is

$$\sum_{y_3=1}^{f(t_3)-1} 1 = f(t_3) - 1. \tag{3.1}$$

For quantity (2), we count nontrivial trees in T whose first subtree has rank $f(t_1)$ and whose second tree has rank less than $f(t_2)$. The number of such trees is

$$\sum_{y_2=1}^{f(t_2)-1} \sum_{y_3=1}^{y_2} 1 = \sum_{y_2=1}^{f(t_2)-1} y_2 = \frac{[f(t_2) - 1]f(t_2)}{2}. \tag{3.2}$$

Finally, for quantity (1), we count nontrivial trees in T with first subtree rank less than $f(t_1)$:

$$\sum_{y_1=1}^{f(t_1)-1} \sum_{y_2=1}^{y_1} \sum_{y_3=1}^{y_2} 1 = \sum_{y_1=1}^{f(t_1)-1} \sum_{y_2=1}^{y_1} y_2 = \sum_{y_1=1}^{f(t_1)-1} \frac{y_1(y_1+1)}{2} = \frac{[f(t_1) - 1]f(t_1)[f(t_1) + 1]}{6}. \tag{3.3}$$

We sum eqs. (3.3), (3.2), and (3.1), plus 2, and we obtain the index for tree t:

$$f(t) = \frac{1}{6}[f(t_1) - 1]f(t_1)[f(t_1) + 1] + \frac{1}{2}[f(t_2) - 1]f(t_2) + f(t_3) + 1. \tag{3.4}$$

As an example, consider tree $t = (t_1, t_2, t_3)$ with $\big(f(t_1), f(t_2), f(t_3)\big) = (4, 3, 3)$. With eq. (3.4), we obtain $f(t) = \frac{1}{6}(4 - 1)4(5) + \frac{1}{2}(3 - 1)3 + 3 + 1 = 17$. Hence, the tree with subtree ranks $(4, 3, 3)$ is the tree with rank 17.

Definition 3.1. The ranking $f : T \to \mathbb{Z}^+$ for trifurcating unlabeled rooted trees is a function that satisfies

$$f(t) = \begin{cases} 1 & \text{if } m(t) = 1 \\ \frac{1}{6}[f(t_1) - 1]f(t_1)[f(t_1) + 1] + \frac{1}{2}[f(t_2) - 1]f(t_2) + f(t_3) + 1 & \text{if } m(t) > 1. \end{cases}$$

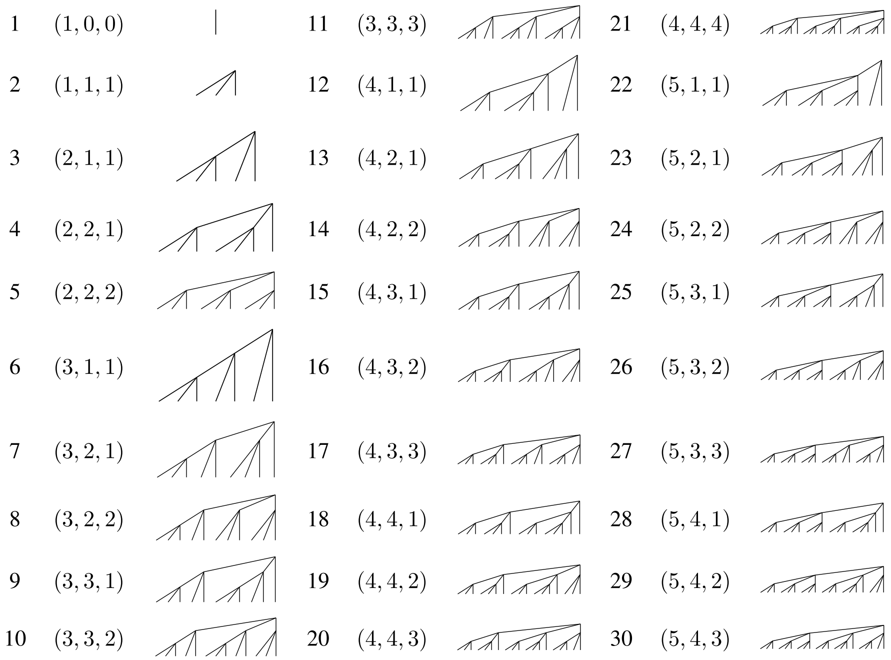


Fig. 2. Trifurcating trees associated with specified ranks. For each rank v from 1 to 30, the ranks $\big(k_1(v), k_2(v), k_3(v)\big)$ of its three subtrees appear, followed by the trifurcating tree associated with rank v. The ordered triple $\big(k_1(v), k_2(v), k_3(v)\big)$ is obtained from Theorem 3.3, and the tree is obtained by recursive application of Theorem 3.3.

For trifurcating trees with small ranks, Fig. 2 displays the trees along with their ranks and the ranks of their subtrees.

3.2. Bijectiveness of ranking scheme

That our ranking for trifurcating trees bijectively associates T with the natural numbers $\mathbb{Z}^+$ is clear by construction. As in the Colijn-Plazzotta ranking for bifurcating trees, underlying the construction is the idea that given a tree t, the integer assigned to t is obtained by counting trees whose ranks are less than those of t and adding 1 to yield the rank of t. Trees are ordered lexicographically, as illustrated in Fig. 2. Each positive integer is trivially assigned a rank (surjectivity). That two distinct trees t and t' possess different ranks is likewise trivial, as one of the two trees must be enumerated among the trees with lower ranks than the other.

Nevertheless, an algebraic proof that our ranking for trifurcating trees bijectively associates T with the natural numbers $\mathbb{Z}^+$ is instructive. The proof illustrates the way in which the three subtree ranks are analogous to the elementary concept of "place value." Each subtree entry is associated with a polynomial: cubic for the first, quadratic for the second, and linear for the third. As in "place value," the contribution to the rank from each of the latter two entries is bounded above by the increase in rank caused by the smallest possible increment to its predecessor.

Theorem 3.2. *The function $f : T \to \mathbb{Z}^+$ is a bijection.*

Proof. To prove injectivity, first note that for the trivial tree, $f\big((1,0,0)\big) = 1$, and for any tree $t \neq (1,0,0)$, $f(t) > 1$. Next, we must show that for two distinct nontrivial trees $t = (t_1, t_2, t_3)$ and $y = (y_1, y_2, y_3)$, $f(t) \neq f(y)$. We can separate the problem into three cases: (i) $t_1 = y_1$, $t_2 = y_2$, and $t_3 \neq y_3$; (ii) $t_1 = y_1$ and $t_2 \neq y_2$; (iii) $t_1 \neq y_1$.

Case (i). Suppose $t_1 = y_1$ and $t_2 = y_2$; without loss of generality, assume $f(t_3) > f(y_3)$. Then

$$f(t) - f(y) = f(t_3) - f(y_3) > 0,$$

so that $f(t) \neq f(y)$.

Case (ii). Suppose $t_1 = y_1$, and without loss of generality, assume $f(t_2) > f(y_2)$. Then

$$f(t) - f(y) = \frac{[f(t_2) - 1]f(t_2)}{2} - \frac{[f(y_2) - 1]f(y_2)}{2} + f(t_3) - f(y_3).$$

Because $f(t_2) > f(y_2)$ and both $f(t_2)$ and $f(y_2)$ are integers, $f(t_2) \geqslant f(y_2) + 1$, and

$$\frac{[f(t_2) - 1]f(t_2)}{2} - \frac{[f(y_2) - 1]f(y_2)}{2} \geqslant \frac{f(y_2)[f(y_2) + 1]}{2} - \frac{[f(y_2) - 1]f(y_2)}{2} = f(y_2).$$

Next, note that because t is not the trivial tree $(1,0,0)$, $f(t_3) \geqslant 1$. Because a tree is required to be in canonical form, $f(y_3) \leqslant f(y_2)$. Hence $f(t_3) - f(y_3) \geqslant 1 - f(y_2)$. Then $f(t) - f(y) \geqslant f(y_2) + \big(1 - f(y_2)\big) = 1$, and $f(t) \neq f(y)$.

Case (iii). Suppose $t_1 \neq y_1$. We can assume without loss of generality that $f(t_1) > f(y_1)$. Then

$$\begin{aligned} f(t) - f(y) = {} & \frac{[f(t_1) - 1]f(t_1)[f(t_1) + 1]}{6} - \frac{[f(y_1) - 1]f(y_1)[f(y_1) + 1]}{6} \\ & + \frac{[f(t_2) - 1]f(t_2)}{2} - \frac{[f(y_2) - 1]f(y_2)}{2} + f(t_3) - f(y_3). \end{aligned}$$

Because $f(t_1) > f(y_1)$, and both $f(t_1)$ and $f(y_1)$ are integers, $f(t_1) \geqslant f(y_1) + 1$, and

$$\begin{aligned} & \frac{[f(t_1) - 1]f(t_1)[f(t_1) + 1]}{6} - \frac{[f(y_1) - 1]f(y_1)[f(y_1) + 1]}{6} \\ \geqslant {} & \frac{f(y_1)[f(y_1) + 1][f(y_1) + 2]}{6} - \frac{[f(y_1) - 1]f(y_1)[f(y_1) + 1]}{6} \\ & = \frac{f(y_1)[f(y_1) + 1]}{2}. \end{aligned} \tag{3.5}$$

We also know that because $f(y_2) \leqslant f(y_1)$,

$$\frac{[f(t_2)-1]f(t_2)}{2} - \frac{[f(y_2)-1]f(y_2)}{2} \geqslant 0 - \frac{[f(y_1)-1]f(y_1)}{2}. \tag{3.6}$$

As in Case (ii), we have

$$f(t_3) - f(y_3) \geqslant 1 - f(y_1). \tag{3.7}$$

Combining inequalities (3.5), (3.6), and (3.7), we see that

$$f(t) - f(y) \geqslant \frac{f(y_1)[f(y_1)+1]}{2} - \frac{[f(y_1)-1]f(y_1)}{2} + 1 - f(y_1) = 1,$$

and $f(t) \neq f(y)$.

With all three cases demonstrated, we conclude that if $(t_1, t_2, t_3) \neq (y_1, y_2, y_3)$, then $f(t) \neq f(y)$, so that f is injective.

For surjectivity, each positive integer $v \geqslant 2$ has a unique representation as a decomposition

$$v = \frac{(k_1-1)k_1(k_1+1)}{6} + \frac{(k_2-1)k_2}{2} + k_3 + 1, \tag{3.8}$$

with k_1, k_2, k_3 positive integers and $k_1 \geqslant k_2 \geqslant k_3$. To understand why such a decomposition exists, note that as k_3 ranges from 1 to k_2, $(k_2-1)k_2/2 + k_3$ ranges from $(k_2-1)k_2/2+1$ to $k_2(k_2+1)/2$, so that the ordered pairs (k_2, k_3) with k_2 and k_3 variable, $k_1 \geqslant k_2 \geqslant k_3 \geqslant 1$, enumerate all positive integers from 1 to $k_1(k_1+1)/2$.

Next, as k_2 ranges from 1 to k_1 and k_3 ranges from 1 to k_2, $(k_1-1)k_1(k_1+1)/6 + (k_2-1)k_2/2 + k_3$ ranges from $(k_1-1)k_1(k_1+1)/6 + 1$ to $k_1(k_1+1)(k_1+2)/6$. Hence, the ordered pairs (k_1, k_2, k_3) with $c \geqslant k_1 \geqslant k_2 \geqslant k_3$ enumerate all positive integers from 1 to $c(c+1)(c+2)/6$.

Noting that 1 is added in the decomposition in eq. (3.8), eq. (3.8) traverses all the positive integers greater than or equal to 2. □

3.3. *The inverse function that recursively converts an integer to a tree*

Theorem 3.3. *The function $f^{-1} : \mathbb{Z}^+ \to T$ gives the three coordinates of the tree whose rank is v, and it satisfies*

(a) *$f^{-1}(1)$ is the tree with one leaf, and*
(b) *for $v \geqslant 2$, $f^{-1}(v)$ is the tree $t \in T$ whose subtrees have the ranks:*

$$k_1(v) = \left\lfloor \frac{3^{1/3} + \left(27v - 54 + \sqrt{3(243v^2 - 972v + 971)}\right)^{2/3}}{3^{2/3}\left(27v - 54 + \sqrt{3(243v^2 - 972v + 971)}\right)^{1/3}} \right\rfloor \tag{3.9}$$

$$k_2(v) = \left\lfloor \frac{1}{6}\left(\sqrt{3}\sqrt{-4k_1^3(v) + 4k_1(v) + 24v - 45} + 3\right)\right\rfloor \tag{3.10}$$

$$k_3(v) = v - \frac{[k_1(v) - 1]k_1(v)[k_1(v) + 1]}{6} - \frac{[k_2(v) - 1]k_2(v)}{2} - 1. \tag{3.11}$$

Proof. For $v \geqslant 2$, we find the unique $\big(f(t_1), f(t_2), f(t_3)\big)$ with $f(t_1) \geqslant f(t_2) \geqslant f(t_3) \geqslant 1$ that solves

$$v = \frac{1}{6}[f(t_1) - 1]f(t_1)[f(t_1) + 1] + \frac{1}{2}[f(t_2) - 1]f(t_2) + f(t_3) + 1.$$

First, $f(t_1)$ is the largest integer satisfying

$$\frac{1}{6}[f(t_1) - 1]f(t_1)[f(t_1) + 1] + 2 \leqslant v.$$

Solving the inequality, the first subtree has rank as in eq. (3.9).

Next, $f(t_2)$ is the largest integer satisfying

$$\frac{1}{2}[f(t_2) - 1]f(t_2) + 2 \leqslant v - \frac{1}{6}[k_1(v) - 1]k_1(v)[k_1(v) + 1].$$

Solving, the second subtree has rank as in eq. (3.10). □

Using the inverse function f^{-1}, Fig. 2 gives the trifurcating trees for ranks 1 to 30.

3.4. Number of leaves associated with the tree of a given integer

With the bijection between trifurcating trees and positive integers in Theorem 3.3, we quickly obtain a recursion for the number of leaves possessed by a trifurcating tree with specified rank.

Theorem 3.4. *The function $m : B \to \mathbb{Z}^+$ that gives the number of leaves in the tree $f^{-1}(v)$ with specified rank satisfies $m\big(f^{-1}(1)\big) = 1$, and for $v \geqslant 2$,*

$$\begin{aligned} m\big(f^{-1}(v)\big) = m&\left(f^{-1}\left(\left\lfloor \frac{3^{1/3} + \big(27v - 54 + \sqrt{3(243v^2 - 972v + 971)}\big)^{2/3}}{3^{2/3}\big(27v - 54 + \sqrt{3(243v^2 - 972v + 971)}\big)^{1/3}} \right\rfloor\right)\right) \\ &+m\left(f^{-1}\left(\left\lfloor \frac{1}{6}\left(\sqrt{3}\sqrt{-4k_1^3(v) + 4k_1(v) + 24v - 45} + 3\right)\right\rfloor\right)\right) \\ &+m\left(f^{-1}\left(v - \frac{[k_1(v) - 1]k_1(v)[k_1(v) + 1]}{6} - \frac{[k_2(v) - 1]k_2(v)}{2} - 1\right)\right). \end{aligned}$$

Proof. The number of leaves in the tree of rank $v \geqslant 2$, or $m\big(f^{-1}(v)\big)$, is the sum of the numbers of leaves in its first, second, and third subtrees, or $m\big(k_1(v)\big) + m\big(k_2(v)\big) + m\big(k_3(v)\big)$. □

4. k-furcating trees

The results for bifurcating and trifurcating trees generalize. Let U_n be the set of k-*furcating* unlabeled rooted trees with n leaves. For $n = 1$, U_n has a single unlabeled tree with one leaf. Suppose each non-leaf node of a tree with $n \geqslant k$ leaves has exactly k descendants. Let $U = \cup_{n=1}^{\infty} U_n$ be the set of k-furcating unlabeled rooted trees, considering all possible numbers of leaves. Let $U^* = U \setminus U_1$. We describe a bijection between unlabeled k-furcating rooted trees and positive integers.

For a tree $u \in U$, $m : U \to \mathbb{Z}^+$ yields the number of leaves of u, and $s : U^* \to U \times U \times \cdots \times U$ extracts a vector containing the immediate subtrees of the root of a tree. We abbreviate by $u_1, u_2, \ldots, u_k$ the k coordinates of $s(u)$. The case of $k = 2$ is the Colijn-Plazzotta ranking. The case of $k = 3$ is the case described in Section 3. We now consider arbitrary $k \geqslant 2$.

4.1. Ranking scheme

We seek to derive the rank of a tree $u = (u_1, u_2, \ldots, u_k)$. We must sum k quantities: (1) the number of nontrivial trees in U whose first subtree has rank less than $f(u_1)$; (2) the number of nontrivial trees in U whose first subtree has rank $f(u_1)$ and whose second subtree has rank less than $f(u_2)$; … (k) the number of nontrivial trees whose first subtree has rank $f(u_1)$, whose second subtree has rank $f(u_2)$, …, whose $(k-1)$-th subtree has rank $f(u_{k-1})$, and whose kth subtree has rank less than $f(u_k)$. We assign to u the next available rank, equal to the sum of the quantities in (1), (2), …, (k) plus 2, accounting for the trivial tree with $n = 1$.

The number of trees whose first term is less than $f(u_1)$ is the number of trees satisfying $f(u_1) > f(u_2) \geqslant f(u_3) \geqslant \ldots \geqslant f(u_k)$. Coordinate $f(u_2)$ is strictly less than $f(u_1)$, $f(u_3)$ can be at most $f(u_2)$, and so on. Because all coordinates take integer values, the desired number of trees is

$$\sum_{y_1=1}^{f(u_1)-1} \sum_{y_2=1}^{y_1} \cdots \sum_{y_k=1}^{y_{k-1}} 1.$$

More generally, considering only the "end" of a vector of subtrees, from coordinate $n \leqslant k$ to coordinate k, the number of trees whose nth term is less than $f(u_n)$ follows the same logic. We must consider all possible values $(u_n, u_{n+1}, \ldots, u_k)$ with $f(u_{n+1}) \leqslant f(u_n) - 1$, $f(u_{n+2}) \leqslant f(u_{n+1})$, and so on, so that the desired number of trees is

$$\sum_{y_n=1}^{f(u_n)-1} \sum_{y_{n+1}=1}^{y_n} \cdots \sum_{y_k=1}^{y_{k-1}} 1.$$

From this expression, we devise a formula for the rank of a tree u.

We have the following lemma.

Lemma 4.1. *The number of nonincreasing sequences of positive integers* $(y_n, y_{n+1}, \dots, y_k)$ *in which all entries are bounded above by* y_{n-1} *is*

$$\sum_{y_n=1}^{y_{n-1}} \sum_{y_{n+1}=1}^{y_n} \cdots \sum_{y_k=1}^{y_{k-1}} 1 = \binom{y_{n-1}+k-n}{k-n+1}.$$

Proof. We proceed by induction, working from inner sums outward. For the inner sum, with $n = k$,

$$\sum_{y_k=1}^{y_{k-1}} 1 = \binom{y_{k-1}}{1}.$$

Now assume that our statement holds for the sum indexed by y_{n+1}. That is, assume that

$$\sum_{y_{n+1}=1}^{y_n} \sum_{y_{n+2}=1}^{y_{n+1}} \cdots \sum_{y_k=1}^{y_{k-1}} 1 = \binom{y_n+k-n-1}{k-n}.$$

Therefore,

$$\sum_{y_n=1}^{y_{n-1}} \sum_{y_{n+1}=1}^{y_n} \cdots \sum_{y_k=1}^{y_{k-1}} 1 = \sum_{y_n=1}^{y_{n-1}} \binom{y_n+k-n-1}{k-n}.$$

Using the standard combinatorial identity for positive integers m, n,

$$\sum_{k=1}^{n} \binom{k+m-1}{m} = \binom{n+m}{m+1},$$

we see that, as desired,

$$\sum_{y_n=1}^{y_{n-1}} \binom{y_n+k-n-1}{k-n} = \binom{y_{n-1}+k-n}{k-n+1}. \qquad \square$$

Definition 4.2. Consider a k-furcating tree $u = (u_1, u_2, \dots, u_k)$. The ranking $f : U \to \mathbb{Z}^+$ for k-furcating unlabeled rooted trees is a function that satisfies $f(u) = 1$ for the one-leaf tree with $m(u) = 1$, and for $m(u) > 1$,

$$\begin{aligned} f(u) &= \frac{1}{k!}[f(u_1)-1]f(u_1)[f(u_1)+1]\cdots[f(u_1)+k-2] + \dots \\ &\quad + \frac{1}{2!}[f(u_{k-1})-1]f(u_{k-1}) + f(u_k) + 1 \qquad (4.1) \\ &= 2 + \sum_{i=1}^{k} \binom{f(u_i)+k-i-1}{k-i+1} \end{aligned}$$

$$= 2 + \sum_{i=1}^{k} \binom{f(u_{k-i+1}) + i - 2}{i}. \tag{4.2}$$

To obtain this definition, we have applied Lemma 4.1 term-wise to each of the k entries of the vector of subtrees associated with tree u. If $f(u_k)+1$ is more instructively written $[f(u_k)-1]+2$, then the term with index i in eq. (4.1) counts the number of trees with fixed rank in preceding positions and rank less than $f(u_i)$ in position i; we adjust the sum by $+2$ to assign the next available rank.

For $k = 3$, Definition 4.2 reduces to Definition 3.1; for $k = 2$, it reduces to Definition 2.1.

4.2. Bijectiveness of ranking scheme

As in the bifurcating and trifurcating cases, that Definition 4.2 bijectively relates multifurcating rooted trees and positive integers is clear from the construction. An algebraic argument demonstrating the bijection can make use of Macaulay's binomial expansion theorem [10,12].

Theorem 4.3 *(Macaulay's binomial expansion theorem [10]). Each positive integer can be decomposed as a certain sum of binomial coefficients. In particular, let h and i be positive integers. Then h can be uniquely written in the form*

$$h = \binom{d_i}{i} + \binom{d_{i-1}}{i-1} + \ldots + \binom{d_j}{j}, \tag{4.3}$$

where d_k takes integer values for all k and $d_i > d_{i-1} > \ldots > d_j \geqslant j \geqslant 1$. The expression in eq. (4.3) is called the i-binomial expansion of the integer h.

We can make a slight adaptation of Macaulay's binomial expansion theorem. By the theorem, for fixed $i > 0$, each integer $h > 0$ can be uniquely written in the form

$$h = \binom{d_i}{i} + \binom{d_{i-1}}{i-1} + \ldots + \binom{d_j}{j}, \tag{4.4}$$

where d_k takes integer values for all k and $d_i > d_{i-1} > \ldots > d_{j-1} \geqslant j \geqslant 1$, and $d_j \geqslant 0$. We restate this adaptation as a corollary to Theorem 4.3.

Corollary 4.4. *If the last entry in the decreasing sequence $(d_i, d_{i-1}, \ldots, d_j)$ is permitted to equal 0, then the decomposition of h in the form of eq. (4.4) is still unique.*

Proof. We take the i-binomial expansion of the integer h according to Theorem 4.3. If it does not contain a term $\binom{d_j}{1}$, then we append a term $\binom{0}{1}$ to make a modified i-binomial expansion. □

Theorem 4.5. *The function $f : U \to \mathbb{Z}^+$ described in Definition 4.2 is bijective.*

Proof. Definition 4.2 associates the 1-leaf tree with $f(u) = 1$ and the k-furcating tree of k leaves with $f(u) = 2$. We wish to show that for all other k-furcating trees u, the value produced by Definition 4.2 for tree u is greater than or equal to 3 and is associated only with tree u (injectivity). We also wish to show that each positive integer greater than or equal to 3 is associated with a tree u (surjectivity).

For injectivity, in our Definition 4.2, for each i from 1 to k, let

$$d_i = f(u_{k-i+1}) + i - 2.$$

Then eq. (4.2) can be rewritten

$$f(u) - 2 = \binom{d_k}{k} + \binom{d_{k-1}}{k-1} + \ldots + \binom{d_1}{1}. \tag{4.5}$$

We have $d_{i+1} > d_i$; we see this by noting that $f(u_{k-(i+1)+1}) \geqslant f(u_{k-i+1})$, so that

$$d_{i+1} = f(u_{k-(i+1)+1}) + (i+1) - 2 > f(u_{k-i+1}) + i - 2 = d_i.$$

We can now make use of the modification of Macaulay's binomial expansion theorem in Corollary 4.4, applied with the integer $f(u) - 2$ in the role of h and k in the role of i. The modified theorem states that $f(u) - 2$ has a unique modified k-binomial decomposition; as eq. (4.5) describes such a decomposition, that decomposition is unique.

For surjectivity, each positive integer $f(u) \geqslant 3$ can be applied in eq. (4.5) to recover the associated $\big(f(u_1), f(u_2), \ldots, f(u_k)\big)$. □

4.3. The inverse function that recursively converts an integer to a tree

We can apply the bijective representation of k-furcating trees with positive integers to identify the tree associated with a specific integer. With the exception of the 1-leaf tree, associated with the integer 1, a k-furcating tree has the form $(u_1, u_2, \ldots, u_k)$.

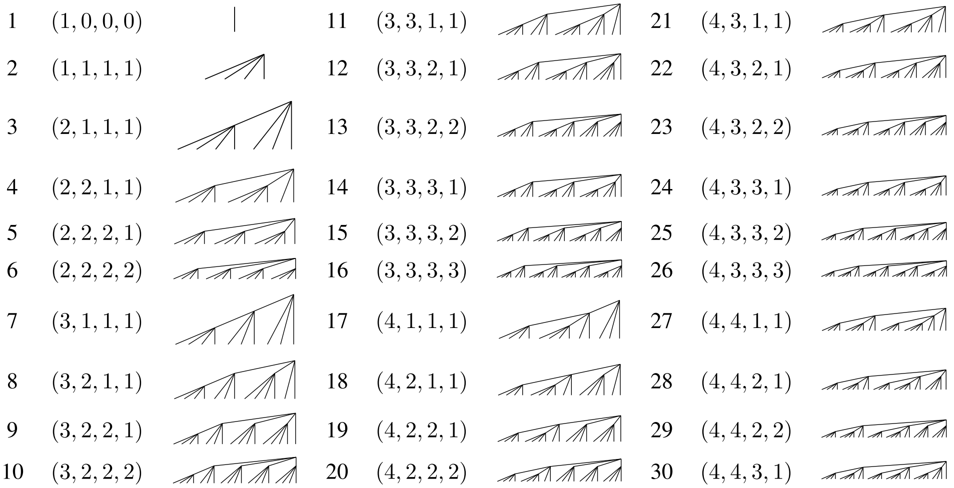
The function $f^{-1} : \mathbb{Z}^+ \to U$ gives a tree; to obtain $f^{-1}(v)$ for a tree with rank v, we first compute the vector of ranks of the immediate subtrees of the root of the tree with rank v:

(a) $f^{-1}(1)$ is the tree with one leaf.

(b) For $v \geqslant 2$, $f^{-1}(v)$ is the tree $u \in U$ whose subtrees have the ranks given by the algorithm below.

1. To find the first term, $f(u_1)$ is the largest integer satisfying

$$\binom{f(u_1) + k - 2}{k} + 2 \leqslant v.$$



1	$(1,0,0,0)$	11	$(3,3,1,1)$	21	$(4,3,1,1)$
2	$(1,1,1,1)$	12	$(3,3,2,1)$	22	$(4,3,2,1)$
3	$(2,1,1,1)$	13	$(3,3,2,2)$	23	$(4,3,2,2)$
4	$(2,2,1,1)$	14	$(3,3,3,1)$	24	$(4,3,3,1)$
5	$(2,2,2,1)$	15	$(3,3,3,2)$	25	$(4,3,3,2)$
6	$(2,2,2,2)$	16	$(3,3,3,3)$	26	$(4,3,3,3)$
7	$(3,1,1,1)$	17	$(4,1,1,1)$	27	$(4,4,1,1)$
8	$(3,2,1,1)$	18	$(4,2,1,1)$	28	$(4,4,2,1)$
9	$(3,2,2,1)$	19	$(4,2,2,1)$	29	$(4,4,2,2)$
10	$(3,2,2,2)$	20	$(4,2,2,2)$	30	$(4,4,3,1)$

Fig. 3. 4-furcating trees associated with specified ranks. For each rank v from 1 to 30, the ranks $\big(k_1(v), k_2(v), k_3(v), k_4(v)\big)$ of its four subtrees appear, followed by the 4-furcating tree associated with rank v. The ordered quadruple $\big(k_1(v), k_2(v), k_3(v), k_4(v)\big)$ is obtained from the algorithm in Section 4.3, and the tree is obtained by recursive application of the algorithm.

2. As in the 3-furcating case, we remove the contributions to the total rank of previous elements. Proceeding sequentially from $j = 2$ until $j = k$, the jth term in $\big(f(u_1), f(u_2), \dots, f(u_k)\big)$ is obtained from terms 1 to $j-1$ by computing the largest integer satisfying

$$\binom{f(u_j)+k-j-1}{k-j+1} + 2 \leqslant v - \sum_{i=1}^{j-1} \binom{f(u_i)+k-i-1}{k-i+1}.$$

The algorithm gives the ranks of the k immediate subtrees of the root of the tree with rank v; to obtain the *entire* tree for rank v, we apply the algorithm recursively to ranks $f(u_1), f(u_2), \dots, f(u_k)$.

Note that in order to calculate the number of leaves associated with the tree of rank v, we recursively proceed by identifying the k subtrees of tree $f^{-1}(v)$, summing their numbers of leaves. In other words, the function $m : U \to \mathbb{Z}^+$ that gives the number of leaves in the tree with specified rank v satisfies $m\big(f^{-1}(1)\big) = 1$, and for $v \geqslant 2$,

$$m\big(f^{-1}(v)\big) = \sum_{j=1}^{k} m\Big(f^{-1}\big(k_j(v)\big)\Big),$$

with $k_j(v)$ representing the rank of the jth coordinate in the tree with rank v, as obtained when the algorithm proceeds through index j.

As an example, Fig. 3 gives the bijection for 4-furcating trees associated with ranks 1 to 30.

5. At-most-k-furcating trees

We next consider *at-most-k-furcating* trees, trees for which each non-leaf node possesses *at most* k descendants. Let A_n be the set of unlabeled rooted trees with n leaves, such that each non-leaf node possesses at most k descendants. Let $A = \cup_{n=1}^{\infty} A_n$ be the set of all at-most-k-furcating unlabeled rooted trees. We include the single-leaf tree for $n = 1$, and write $A^* = A \setminus A_1$.

For a tree $a \in A$, let $m : A \to \mathbb{Z}^+$ denote the number of leaves of a. The function $s : A^* \to A \times A \times (A \cup \emptyset) \times \ldots \times (A \cup \emptyset)$ contains the immediate subtrees of the root of a tree. Because trees in A are *at-most* k-furcating, each non-leaf node possesses at least two and at most k descendants. In a canonical ordering of the subtrees of a non-leaf node, we allow subtrees $3, 4, \ldots, k$ to be empty—that is, we use the empty set to indicate non-existent subtrees if a node has fewer than k immediate subtrees. The number of descendant subtrees of the node is its number of nonempty subtrees.

Note that a bijection of at-most-k-furcating trees with positive integers was considered by Colijn & Plazzotta [1]. For Colijn & Plazzotta [1], however, an internal node was permitted to possess exactly one descendant subtree, whereas we require non-leaf nodes to possess at least two descendant subtrees. We regard a node with exactly one descendant subtree as somewhat unnatural in biological settings, where such a node would often be identified with its unique child node. Disallowing this possibility in order to align with the more common biological scenario requires additional bookkeeping, but the bijective construction is similar to that of Colijn & Plazzotta [1].

We now describe the bijection between unlabeled at-most-k-furcating rooted trees and positive integers. Note that as before, the case of $k = 2$ is the Colijn-Plazzotta ranking; $k = 3$ is the smallest case for which the sets of strictly k-furcating trees and at-most-k-furcating trees differ.

5.1. Ranking scheme

Our goal is to obtain a rank $f(a)$ for a tree $a = (a_1, a_2, \ldots, a_k)$. We make use of a standard combinatorial identity. In particular, for integers $m, n \geqslant 0$

$$\sum_{k=0}^{n} \binom{k+m}{m} = \binom{n+m+1}{m+1}. \tag{5.1}$$

We will also use the following lemma.

Lemma 5.1. *The number of non-increasing sequences of nonnegative integers $(y_n, y_{n+1}, \ldots, y_k)$ in which all entries are bounded above by y_{n-1} is*

$$\sum_{y_n=0}^{y_{n-1}} \sum_{y_{n+1}=0}^{y_n} \cdots \sum_{y_k=0}^{y_{k-1}} 1 = \binom{y_{n-1}+k-n+1}{k-n+1}.$$

Proof. We proceed by induction, working from the innermost sum outward. First, consider the innermost sum, with $n = k$:

$$\sum_{y_k=0}^{y_{k-1}} 1 = \binom{y_{k-1}+1}{1}.$$

Now assume that our statement holds for the sum indexed by y_{n+1}. That is, assume that

$$\sum_{y_{n+1}=0}^{y_n} \sum_{y_{n+2}=0}^{y_{n+1}} \cdots \sum_{y_k=0}^{y_{k-1}} 1 = \binom{y_n + k - n}{k - n}.$$

Therefore,

$$\sum_{y_n=0}^{y_{n-1}} \sum_{y_{n+1}=0}^{y_n} \cdots \sum_{y_k=0}^{y_{k-1}} 1 = \sum_{y_n=0}^{y_{n-1}} \binom{y_n + k - n}{k - n}.$$

Using the combinatorial identity in eq. (5.1),

$$\sum_{y_n=0}^{y_{n-1}} \binom{y_n + k - n}{k - n} = \binom{y_{n-1} + k - n + 1}{k - n + 1},$$

as desired. □

As in Section 4.2, to find the rank of a, we must sum (1) the number of nontrivial trees in A whose first subtree has rank less than $f(a_1)$; (2) the number of nontrivial trees in A whose first subtree has rank $f(a_1)$ and whose second subtree has rank less than $f(a_2)$; (3) the number of nontrivial trees in A whose first subtree has rank $f(a_1)$, second subtree has rank $f(a_2)$, and third subtree has rank less than $f(a_3)$. Here, unlike in the strictly k-furcating case, we allow the third subtree to be empty. Continuing with subsequent subtrees, the last item in the sum is the number of nontrivial trees in A whose first subtree has rank $f(a_1)$, second subtree has rank $f(a_2)$, ..., (possibly empty) $(k-1)$th subtree has rank $f(a_{k-1})$ and (possibly empty) kth subtree has rank less than $f(a_k)$. We assign a the next available rank, which, accounting for the trivial tree with $n = 1$, is 2 more than the quantities we have summed. The required sum is

$$\sum_{y_1=1}^{f(a_1)-1} \sum_{y_2=1}^{y_1} \sum_{y_3=0}^{y_2} \cdots \sum_{y_k=0}^{y_{k-1}} 1.$$

Note that the two outermost sums begin from 1, as these terms correspond to nonempty subtrees; the remaining $k-2$ indices are permitted to equal 0.

Using eq. (5.1) and Lemma 5.1, the number of trees with second subtree rank less than a fixed second term $f(a_2)$ is given by

$$\begin{aligned}\sum_{y_2=1}^{f(a_2)-1}\sum_{y_3=0}^{y_2}\cdots\sum_{y_k=0}^{y_{k-1}} 1 &= \sum_{y_2=1}^{f(a_2)-1}\binom{y_2+k-2}{k-2}\\ &= \left(\sum_{y_2=0}^{f(a_2)-1}\binom{y_2+k-2}{k-2}\right)-\binom{k-2}{k-2}\\ &= \binom{f(a_2)-1+k-1}{k-1}-1. \end{aligned} \tag{5.2}$$

Next, to find the number of trees with rank less than the first term $f(a_1)$, we have

$$\begin{aligned}\sum_{y_1=1}^{f(a_1)-1}\left(\binom{y_1+k-1}{k-1}-1\right) &= -\left[f(a_1)-1\right]+\sum_{y_1=1}^{f(a_1)-1}\binom{y_1+k-1}{k-1}\\ &= -\left[f(a_1)-1\right]+\binom{f(a_1)+k-1}{k}-1. \end{aligned} \tag{5.3}$$

Finally, for a general fixed term $f(a_n)$, we have, analogously to the k-furcating case, the number of trees with nth subtree rank less than the fixed term $f(a_n)$ is given by

$$\sum_{y_n=0}^{f(a_n)-1}\sum_{y_{n+1}=0}^{y_n}\cdots\sum_{y_k=0}^{y_{k-1}} 1 = \binom{f(a_n)+k-n}{k-n+1}. \tag{5.4}$$

Eqs. (5.2), (5.3) and (5.4) yield our desired ranking, as described in the following definition.

Definition 5.2. Consider an at-most-k-furcating tree $a = (a_1, a_2, \ldots, a_k)$. The ranking $f : A \to \mathbb{Z}^+$ for at-most-k-furcating unlabeled rooted trees is a function that satisfies $f(a) = 1$ for the one-leaf tree with $m(a) = 1$, and, for $m(a) > 1$,

$$\begin{aligned} f(a) = & -f(a_1)+\binom{f(a_1)+k-1}{k}+\binom{f(a_2)+k-2}{k-1}-1\\ & +\left[\sum_{i=1}^{k-2}\binom{f(a_{k-i+1})+i-1}{i}\right]+2\\ = & -f(a_1)+1+\sum_{i=1}^{k}\binom{f(a_{k-i+1})+i-1}{i}. \end{aligned} \tag{5.5}$$

Here, we have applied eq. (5.1) to each of the k entries of the vector of subtrees for tree a. For $k = 2$, Definition 5.2 reduces to Definition 2.1.

5.2. Bijectiveness of ranking scheme

As in the case of strictly k-furcating trees, the construction that gives rise to Definition 5.2 provides a bijective ranking scheme between at-most-k-furcating trees and positive integers. We state the result for completeness but omit a detailed algebraic argument.

Theorem 5.3. *The function $f : A \to \mathbb{Z}^+$ as defined in Definition 5.2 is a bijection.*

5.3. The inverse function that recursively converts an integer to a tree

As in the strictly k-furcating case, we can obtain the at-most-k-furcating tree associated with a specified integer. An at-most-k-furcating tree has the form $(a_1, a_2, \dots, a_k)$, with terms after a_2 possibly empty. The function $f^{-1} : \mathbb{Z}^+ \to T$ gives the tree whose rank is v:
(a) $f^{-1}(1)$ is the tree with one leaf, and
(b) for $v \geqslant 2$, $f^{-1}(v)$ is the tree $t \in T$ whose subtrees have the ranks given by the algorithm below.

1. To find the first term, $f(a_1)$ is the largest integer satisfying

$$-f(a_1) + 1 + \binom{f(a_1) + k - 1}{k} \leqslant v.$$

2. Proceeding sequentially from $j = 2$ until $j = k$, the jth term in $\big(f(a_1), f(a_2), \dots, f(a_k)\big)$ is obtained from terms 1 to $j - 1$ by computing the largest integer satisfying

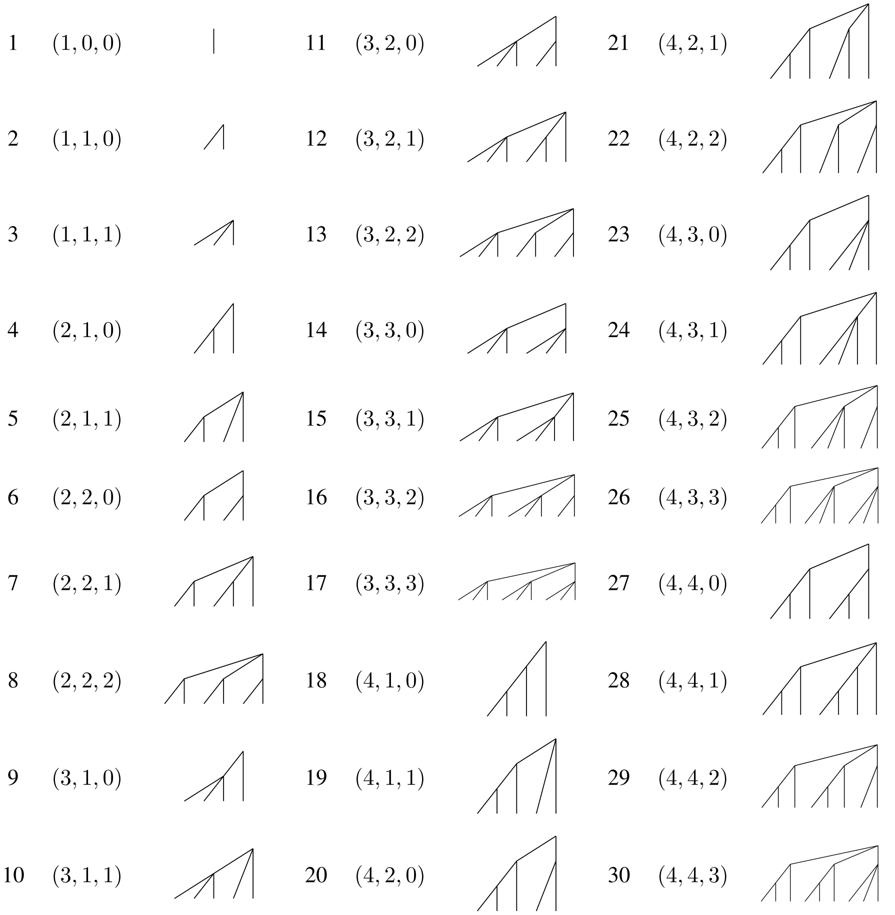
$$\binom{f(a_j) + k - j}{k - j + 1} + 1 \leqslant v + f(a_1) - \sum_{i=1}^{j-1} \binom{f(a_i) + k - i}{k - i + 1}.$$

After step 2, all k coordinates are obtained for the rank v. Note that step 2 can recover values of 0 for coordinates after the first two. We then have that the function $m : A \to \mathbb{Z}^+$ that gives the number of leaves in the tree with specified rank satisfies $m\big(f^{-1}(1)\big) = 1$, and for $v \geqslant 2$,

$$m\big(f^{-1}(v)\big) = \sum_{j=1}^{k} m\Big(f^{-1}\big(k_j(v)\big)\Big),$$

with $k_j(v)$ representing the rank of the jth coordinate in the tree with rank v, as obtained when the algorithm proceeds through index j.

Fig. 4 provides the bijection for at-most-trifurcating trees associated with ranks 1 to 30, and Fig. 3 gives the corresponding bijection for at-most-4-furcating trees.



1	$(1,0,0)$	11	$(3,2,0)$	21	$(4,2,1)$
2	$(1,1,0)$	12	$(3,2,1)$	22	$(4,2,2)$
3	$(1,1,1)$	13	$(3,2,2)$	23	$(4,3,0)$
4	$(2,1,0)$	14	$(3,3,0)$	24	$(4,3,1)$
5	$(2,1,1)$	15	$(3,3,1)$	25	$(4,3,2)$
6	$(2,2,0)$	16	$(3,3,2)$	26	$(4,3,3)$
7	$(2,2,1)$	17	$(3,3,3)$	27	$(4,4,0)$
8	$(2,2,2)$	18	$(4,1,0)$	28	$(4,4,1)$
9	$(3,1,0)$	19	$(4,1,1)$	29	$(4,4,2)$
10	$(3,1,1)$	20	$(4,2,0)$	30	$(4,4,3)$

Fig. 4. At-most-trifurcating trees associated with specified ranks. For each rank v from 1 to 30, the ranks $\big(k_1(v), k_2(v), k_3(v)\big)$ of its three subtrees appear, followed by the trifurcating tree associated with rank v. The ordered triple $\big(k_1(v), k_2(v), k_3(v)\big)$ is obtained from the algorithm in Section 5.3, and the tree is obtained by recursive application of the algorithm.

5.4. Special case: at-most-trifurcating trees

As an example of an at-most-k-furcating ranking, we consider *at-most-trifurcating* trees, in which internal nodes are permitted to bifurcate or trifurcate. A_n becomes the set of at-most-trifurcating unlabeled rooted trees with n leaves. For $n = 1$, A_n contains the unlabeled tree with one leaf. For $n = 2$, A_n contains the unique unlabeled tree with two leaves. We consider $k = 3$ in Definition 5.2.

Definition 5.4. The ranking $f : A \to \mathbb{Z}^+$ for at-most-trifurcating unlabeled rooted trees is a function that satisfies

$$f(a) = \begin{cases} 1 & \text{if } m(a) = 1 \\ \frac{1}{6}[f(a_1) - 1]f(a_1)[f(a_1) + 4] + \frac{1}{2}f(a_2)[f(a_2) + 1] + f(a_3) + 1 & \text{if } m(a) > 1. \end{cases} \tag{5.6}$$

Theorem 5.5. *The function $f^{-1} : \mathbb{Z}^+ \to A$ gives the three coordinates of the tree whose rank is v, and it satisfies*

(a) *$f^{-1}(1)$ is the tree with one leaf, and*
(b) *for $v \geqslant 2$, $f^{-1}(v)$ is the tree $t \in T$ whose subtrees have the ranks:*

$$k_1(v) = \left\lfloor \frac{7\sqrt[3]{3} + \left(27v - 81 + \sqrt{3(243v^2 - 1458v + 1844)}\right)^{2/3}}{3^{2/3}\left(27v - 81 + \sqrt{3(243v^2 - 1458v + 1844)}\right)^{1/3}} - 1 \right\rfloor \tag{5.7}$$

$$k_2(v) = \left\lfloor \frac{1}{6}\left(\sqrt{3}\sqrt{-4k_1(v)^3 - 12k_1(v)^2 + 16k_1(v) + 24v - 21} - 3\right) \right\rfloor \tag{5.8}$$

$$k_3(v) = v - \frac{[k_1(v) - 1]k_1(v)[k_1(v) + 4]}{6} - \frac{k_2(v)[k_2(v) + 1]}{2} - 1. \tag{5.9}$$

Proof. Given $v \geqslant 2$, we seek to find the unique $\big(f(a_1), f(a_2), f(a_3)\big)$ with $f(a_1) \geqslant f(a_2) \geqslant f(a_3)$, $f(a_1) \geqslant 1$, $f(a_2) \geqslant 1$, and $f(a_3) \geqslant 0$, that solves

$$v = \frac{1}{6}[f(a_1) - 1]f(a_1)[f(a_1) + 4] + \frac{1}{2}f(a_2)[f(a_2) + 1] + f(a_3) + 1.$$

The solution for $f(a_1)$ in eq. (5.7) is obtained by solving the inequality

$$\frac{1}{6}[f(a_1) - 1]f(a_1)[f(a_1) + 4] + 2 \leqslant v.$$

Next, the equation for $f(a_2)$ in eq. (5.8) is obtained by solving

$$\frac{1}{2}f(a_2)[f(a_2) + 1] + 1 \leqslant v - \frac{1}{6}[k_1(v) - 1]k_1(v)[k_1(v) + 4].$$

Finally, the inverse function for the third coordinate is given by eq. (5.9). □

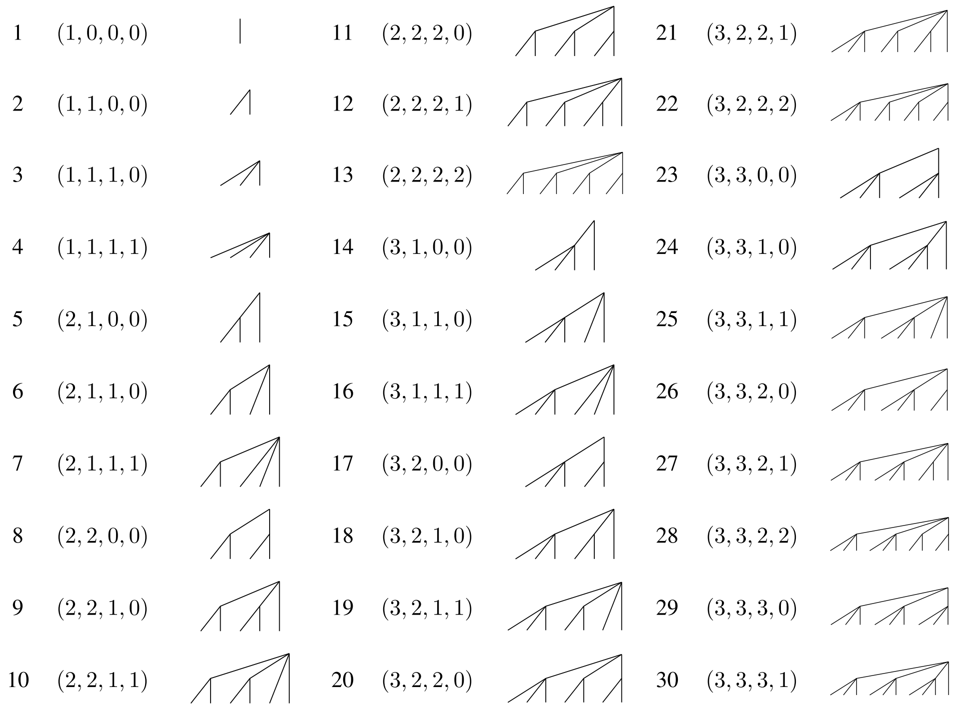
6. Discussion

We have devised ranking systems for multifurcating unlabeled trees, considering trees for which each internal node contains a fixed number of child nodes k as well as trees for which each internal node contains *at most* k child nodes. The general rankings extend a ranking for bifurcating trees, relying on lexicographic orders for subtrees, an analogy between subtree ranks and "place value," and Macaulay's binomial expansion. The main results are summarized in Table 1.

For the k-furcating trees with $k \geqslant 3$, as the rank v increases, in comparison with the bifurcating case ($k = 2$), we observe faster growth in the number of leaves of k-furcating trees associated with rank v (Figs. 1–3). Because each internal node necessarily has k descendants, relatively few multifurcating trees possess small numbers of leaves. Thus, the number of leaves of k-furcating trees increases quickly with v. This effect is magnified

Table 1
Summary of main results.

Description	Result	Equation
Trifurcating tree rank	$f(x_1,x_2,x_3)=\frac{1}{6}(x_1-1)x_1(x_1+1)+\frac{1}{2}(x_2-1)x_2+x_3+1$	(3.4)
k-furcating tree rank	$f(x_1,x_2,\ldots,x_k)=2+\sum_{i=1}^{k}\binom{x_{k-i+1}+i-2}{i}$	(4.2)
At-most-trifurcating tree rank	$f(x_1,x_2,x_3)=\frac{1}{6}(x_1-1)x_1(x_1+4)+\frac{1}{2}x_2(x_2+1)+x_3+1$	(5.6)
At-most-k-furcating tree rank	$f(x_1,x_2,\ldots,x_k)=-x_1+1+\sum_{i=1}^{k}\binom{x_{k-i+1}+i-1}{i}$	(5.5)



1	$(1,0,0,0)$	11	$(2,2,2,0)$	21	$(3,2,2,1)$
2	$(1,1,0,0)$	12	$(2,2,2,1)$	22	$(3,2,2,2)$
3	$(1,1,1,0)$	13	$(2,2,2,2)$	23	$(3,3,0,0)$
4	$(1,1,1,1)$	14	$(3,1,0,0)$	24	$(3,3,1,0)$
5	$(2,1,0,0)$	15	$(3,1,1,0)$	25	$(3,3,1,1)$
6	$(2,1,1,0)$	16	$(3,1,1,1)$	26	$(3,3,2,0)$
7	$(2,1,1,1)$	17	$(3,2,0,0)$	27	$(3,3,2,1)$
8	$(2,2,0,0)$	18	$(3,2,1,0)$	28	$(3,3,2,2)$
9	$(2,2,1,0)$	19	$(3,2,1,1)$	29	$(3,3,3,0)$
10	$(2,2,1,1)$	20	$(3,2,2,0)$	30	$(3,3,3,1)$

Fig. 5. At-most-4-furcating trees associated with specified ranks. For each rank v from 1 to 30, the ranks $\big(k_1(v),k_2(v),k_3(v),k_4(v)\big)$ of its four subtrees appear, followed by the 4-furcating tree associated with rank v. The ordered quadruple $\big(k_1(v),k_2(v),k_3(v),k_4(v)\big)$ is obtained from the algorithm in Section 5.3, and the tree is obtained by recursive application of the algorithm.

as k increases, so that for large k, the number of leaves in the tree with rank v is generally larger than for small k.

For the at-most-k-furcating trees with $k \geqslant 3$, the growth in the number of leaves is not as fast as for k-furcating trees (Figs. 1, 4, and 5). We can explain this observation by noting that for at-most-k-furcating trees, the flexibility in the number of child nodes that could descend from an internal node increases the number of possible trees for a given number of leaves, in comparison with the k-furcating case. As k increases for at-most-k-furcating trees, more possible trees with a given number of leaves exist in the lexicographical ordering, so that the trees in a lexicographical ordering associated with a lower k value are contained in the ordering for a higher k value. Hence, as k increases, the number of leaves in the tree with fixed rank v tends to decrease.

The rankings for multifurcating unlabeled trees have potential uses in phylogenetic studies. In the analysis of phylogenetic trees, particularly in the context of the dynamics of pathogen sequences during epidemics, the unlabeled shape of a tree can provide insight about features of an evolutionary process [1,5]. For rapidly spreading epidemics, a pathogen sequence in one infected individual can branch into multiple sequences in multiple subsequently infected individuals before mutations begin to accumulate in the subsequent sequences. Hence, phylogenetic trees during epidemics often appear to possess a multifurcating structure. The ranking schemes for multifurcating trees can potentially be used in statistics that are informative about the epidemic process; as is the case for bifurcating trees, they may be possible to use in "tree balance" concepts for multifurcating trees [3].

We note that our conceptualization of multifurcating trees differs from some that have been previously considered. Colijn & Plazzotta [1] discussed an extension of their enumeration from bifurcating to at-most-k-furcating trees, permitting internal nodes with only a single immediate subtree. Although it is sensible to consider an internal node with exactly one subtree when that node and its immediate descendant represent known fossil species or pathogen sequences in a known epidemic transmission chain, our assumption that internal nodes possess at least two descendant subtrees reflects a more typical phylogenetic scenario. In another enumeration, Felsenstein [2, p. 33] fixed the number of leaves at n and counted multifurcating trees with at most n leaves; this calculation corresponds to the enumeration of at-most-n-furcating trees with at most n leaves. However, in our ranking, many trees with *more than* n leaves would be assigned ranks smaller than the total number of at-most-n-furcating trees with at most n leaves—so that Felsenstein's set of trees does not correspond to a set of trees tabulated by any of our schemes. As mathematical phylogenetic analysis continues to examine multifurcating trees from the rapidly diversifying sequences that arise during the spread of pathogenic organisms, it will be of interest to clarify which definitions for sets of multifurcating trees give rise to the greatest potential for meaningful biological interpretation.

Acknowledgments

We acknowledge National Science Foundation grant BCS-2116322 for support.

References

[1] C. Colijn, G. Plazzotta, A metric on phylogenetic tree shapes, Syst. Biol. 67 (2018) 113–126.

[2] J. Felsenstein, Inferring Phylogenies, Sinauer, Sunderland, MA, 2004.

[3] M. Fischer, L. Herbst, S. Kersting, L. Kühn, K. Wicke, Tree Balance Indices: a Comprehensive Survey, Springer, Berlin, 2023.

[4] B.T. Grenfell, O.G. Pybus, J.R. Gog, et al., Unifying the epidemiological and evolutionary dynamics of pathogens, Science 303 (2004) 327–332.

[5] J. Kim, N.A. Rosenberg, J.A. Palacios, Distance metrics for ranked evolutionary trees, Proc. Natl. Acad. Sci. USA 117 (2020) 28 876–28 886.

[6] A. Mooers, S. Heard, Inferring evolutionary process from phylogenetic tree shape, Q. Rev. Biol. 72 (1997) 31–54.

[7] O.G. Pybus, A. Rambaut, Evolutionary analysis of the dynamics of viral infectious disease, Nat. Rev. Genet. 10 (2009) 540–550.

[8] N.A. Rosenberg, On the Colijn–Plazzotta numbering scheme for unlabeled binary rooted trees, Discrete Appl. Math. 291 (2021) 88–98.

[9] T. Stadler, Recovering speciation and extinction dynamics based on phylogenies, J. Evol. Biol. 26 (2013) 1203–1219.

[10] R.P. Stanley, Hilbert functions of graded algebras, Adv. Math. 28 (1978) 57–83.

[11] M. Steel, Phylogeny, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2016.

[12] B. Sury, Macaulay expansion, Am. Math. Mon. 121 (2014) 359–360.