THE ALDOUS CHAIN ON CLADOGRAMS IN THE DIFFUSION LIMIT

BY WOLFGANG LÖHR[1,*], LEONID MYTNIK[2] AND ANITA WINTER[1,†]

[1]*Fakultät Mathematik, Universität Duisburg-Essen,* *wolfgang.loehr@uni-due.de;* †*anita.winter@uni-due.de*

[2]*William Davidson Faculty of Industrial Engineering and Management, Technion—Israel Institute of Technology,*
leonid@ie.technion.ac.il

In (*Combin. Probab. Comput.* **9** (2000) 191–204), Aldous investigates a symmetric Markov chain on cladograms and gives bounds on its mixing and relaxation times. The latter bound was sharpened in (*Random Structures Algorithms* **20** (2002) 59–70). In the present paper, we encode cladograms as binary, algebraic measure trees and show that this Markov chain on cladograms with a fixed number of leaves converges in distribution as the number of leaves tends to infinity. We give a rigorous construction of the limit as the solution of a well-posed martingale problem. The existence of a continuum limit diffusion was conjectured by Aldous, and we therefore refer to it as Aldous diffusion. We show that the Aldous diffusion is a Feller process with continuous paths, and the algebraic measure Brownian CRT is its unique invariant distribution.

Furthermore, we consider the vector of the masses of the three subtrees connected to a sampled branch point. In the Brownian CRT, its annealed law is known to be the Dirichlet distribution. Here, we give an explicit expression for the infinitesimal evolution of its quenched law under the Aldous diffusion.

1. Introduction. An *N-cladogram* is a semi-labelled, unrooted, binary tree with $N \geq 2$ *leaves* labelled $\{1, 2, \ldots, N\}$ and with $N - 2$ unlabelled internal vertices. Cladograms are particular phylogenetic trees for which no information on the edge lengths is available, and which therefore only capture the tree structure. Reconstructing cladograms from DNA data is of major interest in population genetics. An important ingredient for several algorithms are Markov chains that move through a space of finite trees (see, e.g., [12] for a survey on Markov chain Monte Carlo algorithms in maximum likelihood tree reconstruction). Usually, such chains are based on a set of simple rearrangements that transform a tree into a "neighboring" tree (see, e.g., [4, 6, 7, 12]).

The present paper considers (a continuous-time version of) the *Aldous chain* on cladograms, which is a Markov chain on the space $\mathfrak{C}_N$ of all N-cladograms. It has the following transition rates: for each pair (u, e) consisting of a leaf and an edge, at rate 1, the Markov chain jumps from its current state $\mathfrak{t}$ to $\mathfrak{t}^{(u,e)}$, where that latter is obtained as follows (see Figures 1 and 2). If u is not incident to e, then:

- Erase the unique edge (including the incident vertices) incident to u,
- split the remaining subtree at the edge e into two pieces, and
- reintroduce the above edge (including u and the branch point) at the split point.

Otherwise, if u is incident to e, we set $\mathfrak{t}^{(u,e)} = \mathfrak{t}$. In total, these so-called *Aldous moves* from $\mathfrak{t}$ to $\mathfrak{t}^{(u,e)}$ happen at rate $N(2N-3)$, and the rate of actual jumps of the Markov chain (where $\mathfrak{t}^{(u,e)} \neq \mathfrak{t}$) is $N(2N-6)$.
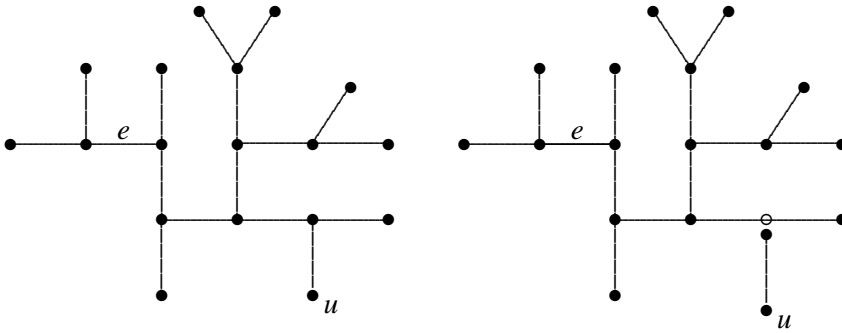
FIG. 1. *At rate $N(2N - 3)$, (a) a leaf u and an edge e are picked at random, and if e and u are not adjacent, (b) the edge incident to u is taken away, leaving behind a branch point of degree 2. (continued in Figure 2).*

This Markov chain has the generator Ω_N, acting on all functions $\phi \colon \mathfrak{C}_N \to \mathbb{R}$ as follows:

$$(1.1) \qquad \Omega_N \phi(\mathfrak{t}) = \sum_{(u,e)} \big(\phi(\mathfrak{t}^{(u,e)}) - \phi(\mathfrak{t}) \big),$$

where the sum runs over all pairs (u, e) consisting of a leaf and an edge, and $\mathfrak{t} \in \mathfrak{C}_N$. Obviously, the Aldous chain is reversible, and the uniform distribution on $\mathfrak{C}_N$ is the stationary distribution. It was shown in [3] that both mixing and relaxation time of the discrete-time chain are of order at least $\mathcal{O}(N^2)$, but at most of order $\mathcal{O}(N^3)$. [27] verified that the relaxation time is of order $\mathcal{O}(N^2)$. Therefore, our continuous-time version has relaxation time of order 1.

As [1] shows that a random N-cladogram with uniform edge lengths $\frac{1}{\sqrt{N}}$ converges weakly to the Brownian Continuum Random Tree (CRT), Aldous conjectured the existence of a CRT-symmetric diffusion limit of the Aldous chain on N-cladograms observed at time scale of order $\mathcal{O}(N^2)$ as $N \to \infty$. This conjecture was presented in a talk in March 1999 given at the Fields Institute, and is supported by the following calculation: suppose we start the Markov chain in some initial N-cladogram, fix a branch point, and consider the relative sizes (η_1, η_2, η_3) of the three subtrees attached to this branch point. Then, as the Markov chain runs, these proportions change as a certain Markov chain, until the branch point disappears. On the proposed time-rescaling of N^2, the $N \to \infty$ limit is the diffusion with generator

$$(1.2) \qquad \Omega f(\underline{\eta}) = \sum_{1 \le i,j \le 3} \eta_i (\delta_{i,j} - \eta_i) \partial^2_{i,j} f(\underline{\eta}) - \frac{1}{2} \sum_{i=1}^3 (1 - 3\eta_i) \partial_i f(\underline{\eta})$$
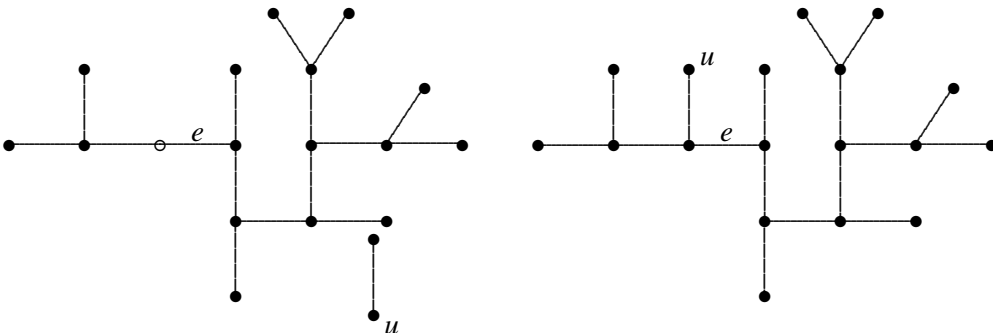

FIG. 2. *(c) The two edges containing the branch point of degree 2 are identified, the edge e is opened, and (d) the free edge is reattached there.*

which records certain aspects of a diffusion on the continuum tree. Aldous raised the question of how this diffusion should be constructed rigourously and what more can we calculate from there? On Aldous's open problem website, the construction was rated as straightforward provided the right set-up is chosen.

The present paper is demonstrating that indeed a straightforward construction can be given once we choose the right state space. A classical starting point would be to think of continuum trees as real trees which are particular metric spaces. A metric space is called a real tree if it is path connected and satisfies the so-called four-point condition. For convergence results one would like to be in a position to treat the approximating discrete trees and their path-connected scaling limits in a unified way. One therefore also considers *metric trees* (introduced in [5]), which are metric spaces differing from a real tree by not necessarily being path connected. A metric space is a metric tree if it can be embedded isometrically into a real tree in such a way that for every choice of three points in the metric tree, the corresponding branch point (defined in the real tree) belongs to the metric tree.

In many applications, it is useful to have metric trees equipped with a probability measure as, for example, the definition of the discrete Aldous chain dynamics requires to sample leaves according to some probability measure. One therefore considers the space $\mathbb{M}$ of isometry classes of *metric measure spaces* and equips it with the Gromov-weak topology. In fact, Aldous's CRT arises as the Gromov-weak scaling limit of uniformly chosen N-cladograms with the uniform distribution on the leaves and edge lengths scaled down by the factor $\frac{1}{\sqrt{N}}$.

One of the equivalent definitions of the *Gromov-weak topology* is by convergence of the distance matrix distributions, that is, a sequence $(\chi_N)_{N \in \mathbb{N}}$ of metric measure spaces converges to a metric measure space $\chi \in \mathbb{M}$ if and only if $\Phi(\chi_N) \longrightarrow_{N \to \infty} \Phi(\chi)$ for all test functions $\Phi \colon \mathbb{M} \to \mathbb{R}$ of the form

$$(1.3) \qquad \Phi(\chi) := \int \phi\big((r(u_i, u_j))_{1 \leq i < j \leq m}\big) \mu^{\otimes m}(\mathrm{d}\underline{u}),$$

where $\chi = (X, r, \mu)$, $m \in \mathbb{N}$, and $\phi \in \mathcal{C}_b(\mathbb{R}_+^{\binom{m}{2}})$ (see [18, 22]).

In this set-up, many tree-valued Markov processes have been constructed and in some cases also the convergence of approximating discrete tree-valued dynamics has been established (see, for example, [8, 11, 19, 21, 23]). One could think that metric (measure) trees are the natural framework for rescaling the Aldous chain as well. However, the Aldous chain resists this approach. An easy calculation shows that the quadratic variation of the averaged distance process rescales at time scale $N^{\frac{3}{2}}$. But how does it relate to the conjecture that the Aldous chain rescales on the time scale N^2? One reason might be that distances behave too wildly for tightness on that time scale to hold. Which in turn might be a hint that the naively used graph distance is not the notion of distance intrinsic to the Aldous chain dynamics. And indeed, one can argue that two points are close if the mass branching off the line segment connecting them is small rather than if the length of that line segment is small. The idea for our new state space is to overcome the metric issue by focusing on the tree structure only.

In what follows, we refer to (T, c) as an *algebraic tree* if $T \neq \varnothing$ is a set equipped with a *branch point map* $c \colon T^3 \to T$ satisfying consistency conditions (see Definition 2.1). Even though algebraic trees can be seen as metric trees where one has "forgotten" the metric (i.e., equivalence classes of metric trees), the branch point map is defined such that the notions of leaves, branch points, degree, subtrees, line segments, etc. can be formalized without reference to a metric (and agree with the corresponding notions in the metric tree). The Aldous diffusion takes values in the new state space $\mathbb{T}$ of (equivalence classes of) *algebraic measure trees* introduced in [24] (see Section 2 for algebraic trees as topological spaces and for equivalence classes of algebraic measure trees). An algebraic measure tree (T, c, μ) consists of an algebraic tree (T, c) satisfying a separability condition, together with a probability measure

μ on it (see Definition 2.2). For a notion of convergence in $\mathbb{T}$, we first introduce the *branch point distribution* on T,

$$\nu_{(T,c,\mu)} := \mu^{\otimes 3} \circ c^{-1}, \tag{1.4}$$

and then associate an algebraic measure tree $\chi = (T, c, \mu) \in \mathbb{T}$ with the metric measure tree $(T, r_\mu, \mu) \in \mathbb{M}$. To this end, define the pseudometric

$$r_\mu(x, y) := \nu_\chi([x, y]) - \frac{1}{2}\nu_\chi(\{x\}) - \frac{1}{2}\nu_\chi(\{y\}), \tag{1.5}$$

where $x, y \in T$, and $[x, y]$ is the interval ("line segment") from x to y. We define convergence of the algebraic measure trees in $\mathbb{T}$ as Gromov-weak convergence of these associated metric measure trees, that is, we say

$$(1.6) \qquad (T_N, c_N, \mu_N)_{N \in \mathbb{N}} \text{ converges in } \mathbb{T} \quad \text{iff} \quad (T_N, r_{\mu_N}, \mu_N)_{N \in \mathbb{N}} \text{ converges in } \mathbb{M}.$$

The space $\mathbb{T}$ equipped with the so-called *branch point distribution distance* Gromov-weak topology (or, for short, bpdd-Gromov-weak topology) is introduced and further studied in [24]. Because cladograms are by definition binary, it is for the purpose of the present paper enough to consider the subspace of $\mathbb{T}$ consisting of binary trees. More precisely, we consider the subspaces

$$\mathbb{T}_2 := \{(T, c, \mu) \in \mathbb{T} : \text{degrees at most } 3, \text{ atoms of } \mu \text{ only at leaves}\} \tag{1.7}$$

of binary trees with no atoms on the skeleton, and

$$\mathbb{T}_2^{\text{cont}} := \{(T, c, \mu) \in \mathbb{T}_2 : \mu \text{ nonatomic}\} \tag{1.8}$$

of continuum binary trees. It is shown in [24], Theorem 3, that both $\mathbb{T}_2$ and $\mathbb{T}_2^{\text{cont}}$ are compact, which is very convenient for showing tightness of the approximating processes. Furthermore, on $\mathbb{T}_2$, we have equivalent formulations of bpdd-Gromov-weak convergence which we can use to prove our limit statements (see Section 2 for more details).

Let $\mathfrak{C}_m$ denotes the set of m-cladograms (see (2.9)). For an algebraic tree (T, c) and $\underline{u} = (u_1, \ldots, u_m) \in T^m$, let $\mathfrak{s}_{(T,c)}(\underline{u})$ denote the m-cladogram generated by the points $u_1, \ldots, u_m$ in (T, c) (see Definition 2.5 for a precise definition). For $m \in \mathbb{N}$ and $\mathfrak{t} \in \mathfrak{C}_m$, let $\Phi^{m,\mathfrak{t}}$ be the function which sends an algebraic measure tree to the probability that m points sampled independently with μ generate the cladogram $\mathfrak{t}$, that is,

$$\Phi^{m,\mathfrak{t}}(T, c, \mu) := \mu^{\otimes m}(\mathfrak{s}_{(T,c)}^{-1}(\mathfrak{t})), \tag{1.9}$$

where $(T, c, \mu) \in \mathbb{T}_2$. We refer to $\mu^{\otimes m} \circ \mathfrak{s}_{(T,c)}^{-1}$ as m-sample shape distribution, and to functions in the linear span of functions of the form (1.9) as *shape polynomials*. One of the main results of [24] is that $\Phi^{m,\mathfrak{t}} \in \mathcal{C}_b(\mathbb{T}_2)$, and moreover, the set of shape polynomials is convergence determining for measures on $\mathbb{T}_2^{\text{cont}}$. Therefore, it is a convenient set of test functions.

We characterize the Aldous diffusion analytically as the unique solution of a martingale problem. We use the following terminology (see Sections 4.3 and 4.4 of [9]). Let E be a polish space, $B(E)$ be the set of bounded, measurable, real-valued functions on E.

DEFINITION 1.1 (Martingale problem).   Let $A: \mathcal{D}(A) \to B(E)$ with $\mathcal{D}(A) \subseteq B(E)$ be a linear operator, and P a probability measure on E.

1. A *solution of* the $(A, \mathcal{D}(A), P)$-martingale problem is an E-valued, measurable stochastic process $X = (X_t)_{t \geq 0}$ such that P is the law of X_0 and, for all $\Phi \in \mathcal{D}(A)$, the process $M := (M_t)_{t \geq 0}$ given by

$$M_t := \Phi(X_t) - \int_0^t A\Phi(X_s) \, ds \tag{1.10}$$

is a martingale (w.r.t. the natural, augmented filtration of X).

2. The $(A, \mathcal{D}(A), P)$-martingale problem is *well-posed* if there exists a unique (in finite dimensional distribution) solution of it.

If the $(A, \mathcal{D}(A), P)$-martingale problem is well-posed for every probability measure P on E, the solution X is necessarily a Markov process by [9], Theorem 4.4.2. We sometimes call the operator A *pre-generator* of X, because it is the restriction of the full generator to $\mathcal{D}(A)$. As pre-generator of the Aldous diffusion, we introduce the operator Ω_{Ald} acting on functions of the form (1.9) as follows:

$$(1.11) \qquad \Omega_{\mathrm{Ald}}\Phi^{m,\mathfrak{t}}(T, c, \mu) := \int \Omega_m \mathbb{1}_{\mathfrak{t}}\big(\mathfrak{s}_{(T,c)}(\underline{u})\big)\mu^{\otimes m}(\mathrm{d}\underline{u}).$$

Obviously, Ω_{Ald} can be extended linearly to the set of shape polynomials, that is, to

$$(1.12) \qquad \mathcal{D}(\Omega_{\mathrm{Ald}}) := \mathrm{span}\{\text{functions } \Phi^{m,\mathfrak{t}} \text{ of the form (1.9) with } m \in \mathbb{N}, \mathfrak{t} \in \mathfrak{C}_m\},$$

where span denotes the linear span of a set of functions. Our first main result is the following.

THEOREM 1 (The well-posed martingale problem). *For all probability measures P_0 on $\mathbb{T}_2^{\mathrm{cont}}$, the $(\Omega_{\mathrm{Ald}}, \mathcal{D}(\Omega_{\mathrm{Ald}}), P_0)$-martingale problem is well-posed. Its unique solution is a Feller process with continuous paths, taking values in the compact state space $\mathbb{T}_2^{\mathrm{cont}}$. In particular, it is a strong Markov process. Moreover, this solution is ergodic with the algebraic measure Brownian CRT as unique invariant distribution.*

We refer to the process from Theorem 1 as *Aldous diffusion*.

DEFINITION 1.2 (Aldous diffusion on binary algebraic measure trees). The unique solution of the $(\Omega_{\mathrm{Ald}}, \mathcal{D}(\Omega_{\mathrm{Ald}}), P_0)$-martingale problem is called *Aldous diffusion on binary algebraic nonatomic measure trees*, or simply *Aldous diffusion*, started in P_0.

It is important to mention that the Aldous diffusion is dual to the Aldous chain, as for all $m \in \mathbb{N}$ and m-cladograms $\mathfrak{t}$, the Aldous diffusion $X_t = (T_t, c_t, \mu_t)$ started in $X_0 = (T, c, \mu) \in \mathbb{T}_2^{\mathrm{cont}}$ satisfies

$$(1.13) \qquad \mathbb{E}_{(T,c,\mu)}\big[\mu_t^{\otimes m}\{\underline{u} \in T_t^m : \mathfrak{s}_{(T_t,c_t)}(\underline{u}) = \mathfrak{t}\}\big] = \mathbb{E}_{\mathfrak{t}}\big[\mu^{\otimes m}\{\underline{u} \in T^m : \mathfrak{s}_{(T,c)}(\underline{u}) = \mathcal{T}_t\}\big],$$

where $(\mathcal{T}_t)_{t \geq 0}$ denotes the Aldous chain on m-cladograms started in $\mathfrak{t}$.

The name *Aldous diffusion* is justified by the following convergence result. Here, we identify the $\mathfrak{C}_N$-valued Aldous chain on N-cladograms with the $\mathbb{T}_2$-valued Markov chain obtained by forgetting the labels of the cladograms and equipping it with the uniform distribution on the leaves.

THEOREM 2 (Diffusion approximation). *For each $N \in \mathbb{N}$, let χ_N be an N-cladogram with the uniform distribution on the leaves. Assume that $\chi_N \to \chi \in \mathbb{T}_2^{\mathrm{cont}}$. Then the Aldous chain X^N starting in $X_0^N = \chi_N$ converges weakly in Skorokhod path space w.r.t. the bpdd-Gromov-weak topology to the Aldous diffusion starting in χ.*

Our last result makes a connection to Aldous's original calculation (1.2) of the evolution of the relative sizes of the three subtrees attached to a fixed branch point until that branch point disappears. Instead of fixing a branch point in the beginning, we take the average over branch points w.r.t. the branch point distribution (1.4). Our topology on $\mathbb{T}_2$ turns out to be strong enough for us to use the diffusion approximation from Theorem 2 to extend the martingale problem for the Aldous diffusion to the corresponding test functions. Thus, we can do explicit

calculations which show the missing term compensating for the disappearance of branch points.

To state the result, we need some notation. For a branch point $v \in \mathrm{br}(T)$, consider the three subtrees (components) attached to v, and denote by $\mathcal{S}_v(u)$ the one containing $u \in T$ with $u \neq v$ (see (2.2) below for a precise definition). For $\underline{u} = (u_1, u_2, u_3) \in T^3$, let

$$\tag{1.14} \underline{\eta}(\underline{u}) := (\eta_i(\underline{u}))_{i=1,2,3} := (\mu(\mathcal{S}_{c(\underline{u})}(u_i)))_{i=1,2,3}$$

be the vector of the three masses of the components connected to the branch point $c(\underline{u})$ of $\underline{u}$. We consider test functions of the following form, called *mass polynomials of degree* 3: For $f : [0, 1]^3 \to \mathbb{R}$ continuous define

$$\tag{1.15} \Phi^f(T, c, \mu) := \int f(\underline{\eta}(\underline{u})) \mu^{\otimes 3}(\mathrm{d}\underline{u}),$$

where $(T, c, \mu) \in \mathbb{T}_2$. One of the main results of [24] is that $\Phi^f \in \mathcal{C}(\mathbb{T}_2)$.

We can extend the domain of the pre-generator Ω_{Ald} to the set of those mass polynomials Φ^f of degree 3 with $f : [0, 1]^3 \to \mathbb{R}$ twice continuously differentiable. To this end, consider the migration operators $\Theta_{i,j} : \mathcal{C}^2([0, 1]^3) \to \mathcal{C}^1([0, 1]^3)$, $i, j \in \{1, 2, 3\}$, $i \neq j$,

$$\tag{1.16} \Theta_{1,2} f(\underline{x}) := \frac{\mathbb{1}_{x_1 > 0}}{x_1} (f(0, x_2 + x_1, x_3) - f(\underline{x})) + \mathbb{1}_{x_1 = 0}(\partial_2 f(\underline{x}) - \partial_1 f(\underline{x})),$$

and $\Theta_{i,j} f$ defined analogously with the indices 1 and 2 replaced by i and j, respectively. Let $e_i = (\delta_{ij})_{j=1,2,3}$ be the ith unit vector and define

$$\tag{1.17} \Omega_{\mathrm{Ald}} \Phi^f(T, c, \mu) := \int_{T^3} \mathrm{d}\mu^{\otimes 3} \Bigg(2 \sum_{i,j=1}^3 \eta_i(\delta_{ij} - \eta_j) \partial_{ij}^2 f(\underline{\eta}) + 3 \sum_{i=1}^3 (1 - 3\eta_i) \partial_i f(\underline{\eta})$$
$$+ \frac{1}{2} \sum_{i,j=1,i\neq j}^3 \Theta_{i,j} f(\underline{\eta}) + \sum_{i=1}^3 (f(e_i) - f(\underline{\eta})) \Bigg).$$

THEOREM 3 (Extended martingale problem for subtree masses). *Let* $X = (X_t)_{t \geq 0}$ *be the Aldous diffusion on* $\mathbb{T}_2^{\mathrm{cont}}$. *Then for all test functions* Φ^f *of the form* (1.15) *with* $f \in \mathcal{C}^2([0, 1]^3)$, *the process* $M^f := (M_t^f)_{t \geq 0}$ *given by*

$$\tag{1.18} M_t^f := \Phi^f(X_t) - \Phi^f(X_0) - \int_0^t \Omega_{\mathrm{Ald}} \Phi^f(X_s) \, \mathrm{d}s$$

is a martingale.

Related work. We note that a construction related to the Aldous diffusion has been recently established independently in a sequence of papers [14–17]. A discussion of the differences is therefore in order. Their construction was first sketched in [26]. Pal suggests to first take a finite number of branch points, consider the cladogram spanned by them, and decompose the lines connecting any two neighboring branch points of this cladogram into subtrees. Then study the suitably rescaled subtree masses as partitions of an interval of random length while relaxing the constrain that the total number of vertices must be preserved by letting removing and inserting of edges happen independently. When applying the time change which reverses the described Poissonization, on the proposed time scale the masses converge to an evolving interval partition described by a family of diffusions indexed by $\mathbb{N}$. However, if this chain runs, then the mass branching off one of the external edges of the cladogram gets exhausted. When this happens, the dynamics breaks down, and one needs to find a slightly different set of branch points to proceed. To resolve the problem of disappearing vertices [15] suggests a

smart way of swapping labels of the cladograms in such a way that the resulting dynamics preserves stationarity when one starts from the uniform distribution.

Our construction is related in spirit but differs in some important aspects. First, rather than sampling cladograms and describing their dynamics under the Aldous chain, we describe the behavior of the *average* of the quantities of interest over uniformly sampled cladograms. This allows us to give a nice characterization of the Aldous diffusion as a unique solution of some martingale problem. As a consequence, we do not require the initial distribution for the Aldous diffusion to be uniform but can rather let it start in any deterministic continuum tree. We can show that the Aldous chain converges weakly in path space to the Aldous diffusion, and that the latter is a Feller process. Note that [17] never states explicitly that the $\mathbb{R}$-tree-valued diffusion constructed is a strong Markov process. Furthermore, we are also able to state a duality relation, which allows us to conclude convergence to the uniform distribution for all starting points as time tends to infinity. In [24], we put some effort in establishing with the space of algebraic measure trees a new state space and invested in a detailed study of topological aspects. As a result, we obtained equivalent formulations of our notion of convergence on the subspace of binary trees which made martingale convergence statements very much straightforward. Finally, the framework provided is not restricted to the construction of the continuum limit of the Aldous Markov chain. It can also be applied to other (not necessarily symmetric) sampling consistent tree dynamics. For example, in [25] a tree-valued dynamics is constructed which has the algebraic measure Kingman tree as its stationary distribution.

Other approaches of encoding relatives of binary algebraic trees can be found in [13] and [10]. The *Rémy chain* considered in [10] is a Markov chain of growing (ordered) trees that is somewhat related to the Aldous chain: it is the process obtained by successively inserting new leaves at randomly chosen edges without removing a leaf before.

Outline. The rest of the paper is organized as follows. In Section 2, we introduce our state space of algebraic measure trees and recall its most important properties from [24].

In Section 3, we show tightness of the Aldous chains and existence of solutions of the martingale problem from Theorem 1. We do so by using and proving uniform convergence of (pre-)generators. In Section 4, we obtain the duality for the Aldous diffusion (Proposition 4.1), and use it to show uniqueness of solutions of the martingale problem. In Section 5, we show that the Aldous diffusion has a unique invariant measure, namely the algebraic measure Brownian CRT, and that the Aldous diffusion converges to it in law as time goes to infinity (Proposition 5.3). We also finish the proof of Theorems 1 and 2. In Section 6, we prove Theorem 3 and apply it to calculate the annealed average distance of two points in the Brownian CRT with respect to our intrinsic metric.

2. The state space of binary, algebraic measure trees.

In this section, we introduce the state space. The goal is to overcome the metric issue raised in the introduction by focusing on the algebraic tree structure only. We encode the cladograms as binary, algebraic trees and use the space of these trees together with the bpdd-Gromov-weak topology studied in [24]. All proofs can be found there.

DEFINITION 2.1 (Algebraic tree). An *algebraic tree* is a nonempty set T together with a symmetric map $c: T^3 \to T$ satisfying the following:

(2pc) For all $x_1, x_2 \in T$, $c(x_1, x_2, x_2) = x_2$.
(3pc) For all $x_1, x_2, x_3 \in T$, $c(x_1, x_2, c(x_1, x_2, x_3)) = c(x_1, x_2, x_3)$.
(4pc) For all $x_1, x_2, x_3, x_4 \in T$,

$$(2.1) \qquad c(x_1, x_2, x_3) \in \{c(x_1, x_2, x_4), c(x_1, x_3, x_4), c(x_2, x_3, x_4)\}.$$

We refer to the map c as *branch point map*. A *tree isomorphism* between two algebraic trees (T_i, c_i), $i = 1, 2$, is a bijective map $\phi \colon T_1 \to T_2$ with $\phi(c_1(x_1, x_2, x_3)) = c_2(\phi(x_1), \phi(x_2), \phi(x_3))$ for all $x_1, x_2, x_3 \in T_1$.

For each point $x \in T$, we define an equivalence relation $\sim_x$ on $T \setminus \{x\}$ such that for all $y, z \in T \setminus \{x\}$, $y \sim_x z$ iff $c(x, y, z) \neq x$. For $y \in T \setminus \{x\}$, we denote by

$$(2.2) \qquad\qquad \mathcal{S}_x(y) := \{z \in T \setminus \{x\} : z \sim_x y\}$$

the equivalence class w.r.t. $x \in T$ which contains y. We also call $\mathcal{S}_x(y)$ the *component* of $T \setminus \{x\}$ containing y. An algebraic tree (T, c) allows for all kinds of notions which capture the tree structure, for example:

- we say that $S \subseteq T$ is a *subtree* of T iff $c(S^3) = S$,
- we call the number of components of $T \setminus \{x\}$ the *degree* of $x \in T$ and write $\deg(x) = \#\{\mathcal{S}_x(y) : y \in T \setminus \{x\}\}$,
- we say that $u \in T$ is a *leaf* iff $\deg(u) = 1$, and write $\mathrm{lf}(T)$ for the set of leaves,
- we say that $v \in T$ is a *branch point* iff $\deg(v) \geq 3$, or equivalently, $v = c(x_1, x_2, x_3)$ for some $x_1, x_2, x_3 \in T \setminus \{v\}$, and write $\mathrm{br}(T)$ for the set of branch points,
- we write $[x, y]$, $x, y \in T$, for the *interval*

$$(2.3) \qquad\qquad [x, y] := \{z \in T : c(x, y, z) = z\},$$

- and we say that $\{x, y\}$ is an edge iff $x \neq y$ and $[x, y] = \{x, y\}$.

There is a natural topology on a given algebraic tree, namely the *component topology* generated by the set of all components $\mathcal{S}_x(y)$ as defined in (2.2) with $x \neq y$, $x, y \in T$. In what follows, we refer to an algebraic tree (T, c) as *order separable* if it is separable w.r.t. this topology and has at most countably many edges. We further equip order separable algebraic trees with a probability measure on the Borel σ-algebra $\mathcal{B}(T)$ of the component topology. This so-called *sampling measure* allows to sample vertices from the tree.

DEFINITION 2.2 (Algebraic measure trees). A (separable) *algebraic measure tree* (T, c, μ) is an order separable algebraic tree (T, c) together with a probability measure μ on $\mathcal{B}(T)$.

In what follows, we call two algebraic measure trees (T_i, c_i, μ_i), $i = 1, 2$, *equivalent* if there exist subtrees $S_i \subseteq T_i$ with $\mu_i(S_i) = 1$, $i = 1, 2$, and a measure preserving tree isomorphism ϕ from S_1 onto S_2, that is, $c_2(\phi(x), \phi(y), \phi(z)) = \phi(c_1(x, y, z))$ for all $x, y, z \in S_1$, and $\mu_1 \circ \phi^{-1} = \mu_2$. We define

$$(2.4) \qquad\qquad \mathbb{T} := \text{set of equivalence classes of algebraic measure trees.}$$

With a slight abuse of notation, we will write $\chi = (T, c, \mu)$ for the algebraic tree as well as the equivalence class. Note that $\deg(x)$, $\mathrm{lf}(T)$, $\mathrm{edge}(T), \dots$ are properties of the particular representative and not preserved under equivalence, because we do not require the whole trees to be isomorphic. For instance, every equivalence class contains a representative without edges (informally, we can replace edges by line segments carrying no measure).

Recall the branch point distribution $\nu = \nu_\chi = \mu^{\otimes 3} \circ c^{-1}$ and the pseudometric r_μ from Equations (1.4) and (1.5), respectively. For every equivalence class of algebraic measure trees, a representative (T, c, μ) can be chosen such that r_μ is a metric (by identifying points of distance zero in any representative). One can check that in this case, r_μ induces the component topology, and (T, r_μ) is a separable metric tree in the sense of [5] (i.e., isometric to a subset of an $\mathbb{R}$-tree containing all branch points) satisfying for all $x, y, z \in T$,

$$(2.5) \qquad\qquad [x, y]_{r_\mu} \cap [x, z]_{r_\mu} \cap [y, z]_{r_\mu} = \{c(x, y, z)\},$$

where $[x, y]_{r_\mu} = \{v \in T : r_\mu(x, y) = r_\mu(x, v) + r_\mu(v, y)\}$ denotes the interval in (T, r_μ). In particular, $[x, y]_{r_\mu} = [x, y]$. Note that any point which carries positive mass is an isolated point in the metric space (T, r_μ).

As in any metric tree, we can define for a fixed reference point (root) $\rho \in T$ a unique measure $\ell^{(T,c,\mu,\rho)}$ on $(T, \mathcal{B}(T))$ which is characterized by the two properties $\ell^{(T,c,\mu,\rho)}((\rho, y]) := r_\mu(\rho, y)$ and $\ell^{(T,c,\mu,\rho)}(\mathrm{lf}(T) \setminus \mathrm{at}(\mu)) = 0$, where $\mathrm{at}(\mu)$ denotes the set of atoms of μ. The measure $\ell^{(T,c,\mu,\rho)}$ is referred to as *length measure* w.r.t. ρ. Note that it depends on the choice of the distinguished point ρ. However, the total mass of the length measure does not depend on the choice of ρ and equals

$$(2.6) \qquad \|\ell^{(T,c,\mu,\rho)}\| := \ell^{(T,c,\mu,\rho)}(T) = \frac{1}{2} \int_T \deg(v) \nu(dv).$$

We define convergence in $\mathbb{T}$ as follows.

DEFINITION 2.3 (Bpdd-Gromov-weak topology). We say that a sequence $(\chi_n)_{n \in \mathbb{N}}$ of (equivalence classes of) algebraic measure trees $\chi_n = (T_n, c_n, \mu_n) \in \mathbb{T}$ converges *branch point distribution distance Gromov-weakly* (*bpdd-Gromov-weakly*) to the algebraic measure tree $(T, c, \mu) \in \mathbb{T}$ iff the sequence $(\tilde{\chi}_n)_{n \in \mathbb{N}}$ of (equivalence classes of) metric measure trees $\tilde{\chi}_n := (T_n, r_{\mu_n}, \mu_n) \in \mathbb{M}$ converges to the metric measure tree $(T, r_\mu, \mu) \in \mathbb{M}$ Gromov-weakly, that is, if for $U_1^n, U_2^n, \ldots$ independent and μ_n-distributed, and $U_1, U_2, \ldots$ independent and μ-distributed, for all $m \in \mathbb{N}$,

$$(2.7) \qquad \left(r_{\mu_n}(U_i^n, U_j^n)\right)_{1 \le i < j \le m} \underset{n \to \infty}{\Longrightarrow} \left(r_\mu(U_i, U_j)\right)_{1 \le i < j \le m}.$$

In this paper, we are only considering binary algebraic measure trees with the property that the measure has atoms only (if at all) on the leaves of the tree, that is, the subspace of $\mathbb{T}$ given by

$$(2.8) \qquad \mathbb{T}_2 = \left\{(T, c, \mu) \in \mathbb{T} : \deg(v) \le 3 \ \forall v \in T, \mathrm{at}(\mu) \subseteq \mathrm{lf}(T)\right\}$$

(compare (1.7)). Even though the equivalence class $\chi \in \mathbb{T}_2$ contains algebraic measure trees which are not binary, we will implicitly assume that the chosen representative (T, c, μ) satisfies $\deg(v) \le 3$. In this subspace, it turns out that bpdd-Gromov-weak convergence is equivalent to another very useful notion of convergence, namely the so-called sample shape convergence, which we introduce next.

DEFINITION 2.4 (*m*-labelled cladogram). For $m \in \mathbb{N}$, an *m-labelled cladogram* is a binary, finite tree (C, c) consisting only of leaves and branch points, together with a surjective labelling map $\zeta : \{1, \ldots, m\} \to \mathrm{lf}(C)$.

Note that an m-labelled cladogram has at most m leaves (and $m - 2$ branch points), but can have less if a leaf has multiple labels. An m-labelled cladogram (C, c, ζ) is an m-cladogram if and only if ζ is injective. We call two m-labelled cladograms (C_1, c_1, ζ_1) and (C_2, c_2, ζ_2) *isomorphic* if there exists a tree isomorphism ϕ from (C_1, c_1) onto (C_2, c_2) such that $\zeta_2 = \phi \circ \zeta_1$. Furthermore, we denote the sets of isomorphism classes of m-labelled cladogram and m-cladograms by $\overline{\mathfrak{C}}_m$ and $\mathfrak{C}_m$, respectively, that is,

$$(2.9) \qquad \overline{\mathfrak{C}}_m := \{\text{isomorphism classes of } m\text{-labelled cladograms}\}$$

and

$$(2.10) \qquad \mathfrak{C}_m := \{(C, c, \zeta) \in \overline{\mathfrak{C}}_m : \zeta \text{ injective}\}.$$
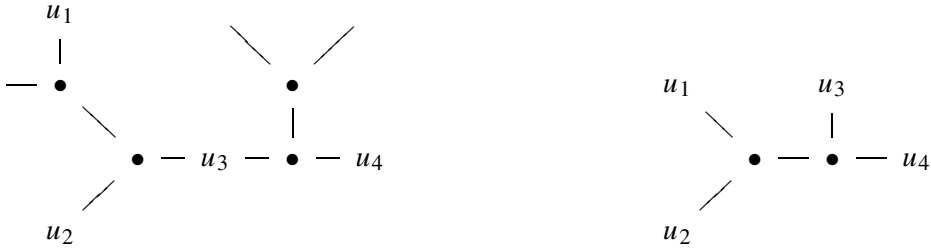
FIG. 3. *A tree T and the shape $\mathfrak{s}_{(T,c)}(u_1, u_2, u_3, u_4)$. The cladogram is not isomorphic to the subtree $c(\{u_1, u_2, u_3, u_4\}^3)$ because $u_3 \in]u_1, u_4[$.*

DEFINITION 2.5 (The shape function). For a binary algebraic tree (T, c), $m \in \mathbb{N}$, and $u_1, \ldots, u_m \in T \setminus \mathrm{br}(T)$ (not necessarily distinct), there exists a unique (up to isomorphism) m-labelled cladogram

$$(2.11) \qquad \mathfrak{s}_{(T,c)}(u_1, \ldots, u_m) = (C, c_C, \zeta)$$

with $\mathrm{lf}(C) = \{u_1, \ldots, u_m\}$ and $\zeta(i) = u_i$, such that the identity on $\mathrm{lf}(C)$ extends to a tree homomorphism π from C onto $c(\{u_1, \ldots, u_m\}^3)$, that is, for all $i, j, k = 1, \ldots, m$,

$$(2.12) \qquad \pi\big(c_C(u_i, u_j, u_k)\big) = c(u_i, u_j, u_k).$$

We refer to $\mathfrak{s}_{(T,c)}(u_1, \ldots, u_m) \in \overline{\mathfrak{C}}_m$ as the *shape* of $u_1, \ldots, u_m$ in (T, c).

Note that the shape need not be isomorphic to the generated subtree. This situation is illustrated in Figure 3.

DEFINITION 2.6 (Sample shape convergence). We say that a sequence $(\chi_n)_{n \in \mathbb{N}}$ of (equivalence classes of) binary algebraic measure trees (T_n, c_n, μ_n) *converges in sample shape* to the (equivalence class of the) algebraic measure tree (T, c, μ) iff for $U_1^n, U_2^n, \ldots$ independent and μ_n-distributed, and $U_1, U_2, \ldots$ independent and μ-distributed, for all $m \in \mathbb{N}$,

$$(2.13) \qquad \mathfrak{s}_{(T,c)}\big(U_1^n, \ldots, U_m^n\big) \underset{n \to \infty}{\Longrightarrow} \mathfrak{s}_{(T,c)}(U_1, \ldots, U_m).$$

To be later in a position to recover the calculations of Aldous and others concerning the dynamics of subtree masses, we introduce yet another notion of convergence.

DEFINITION 2.7 (Sample subtree mass convergence). We say that a sequence $(\chi_n)_{n \in \mathbb{N}}$ of (equivalence classes of) algebraic measure trees (T_n, c_n, μ_n) *converges in sample subtree mass* to the (equivalence class of the) algebraic measure tree (T, c, μ) iff for $U_1^n, U_2^n, \ldots$ independent and μ_n-distributed, and $U_1, U_2, \ldots$ independent and μ-distributed, for all $m \in \mathbb{N}$,

$$(2.14) \qquad \big(\mu_n\big(\mathcal{S}_{c_n(U_i^n, U_j^n, U_k^n)}(U_i^n)\big)\big)_{i,j,k=1,\ldots,m} \underset{n \to \infty}{\Longrightarrow} \big(\mu\big(\mathcal{S}_{c(U_i, U_j, U_k)}(U_i)\big)\big)_{i,j,k=1,\ldots,m}.$$

The following results are crucial for the construction of Aldous diffusion and stated in [24], Proposition 2.32, Theorem 3, Corollary 5.21. On $\mathbb{T}_2$, all of the above notions of convergence are equivalent. By (2.6), the total length of binary algebraic measure trees is uniformly bounded by 3, and one can show that the space $\mathbb{T}_2$ is compact.

PROPOSITION 2.8. *Let $(\chi_N = (T_N, c_N, \mu_N))_{N \in \mathbb{N}}$ and $\chi = (T, c, \mu)$ be in $\mathbb{T}_2$. The following are equivalent*:

(1) $\chi_N \longrightarrow_{n \to \infty} \chi$ *w.r.t. sample shape convergence.*

(2) *For all $m \in \mathbb{N}$ and $\mathfrak{t} \in \overline{\mathfrak{C}}_m$,*

$$(2.15) \qquad \mu_N^{\otimes m}\{(u_1, \ldots, u_m) : \mathfrak{s}_{(T_N, c_N)}(\underline{u}) = \mathfrak{t}\} \xrightarrow[n \to \infty]{} \mu^{\otimes m}\{(u_1, \ldots, u_m) : \mathfrak{s}_{(T, c)}(\underline{u}) = \mathfrak{t}\}.$$

(3) $\chi_N \xrightarrow{}_{n \to \infty} \chi$ *Gromov-weakly w.r.t. the branch point distribution distance.*

(4) *For all $m \in \mathbb{N}$ and $\phi \in \mathcal{C}_b(\mathbb{R}_+^{m \times m})$,*

$$(2.16) \qquad \int \mu_N^{\otimes m}(d\underline{u})\phi\big((r_{\mu_N}(u_i, u_j))_{1 \le i,j \le m}\big) \xrightarrow[n \to \infty]{} \int \mu^{\otimes m}(d\underline{u})\phi\big((r_\mu(u_i, u_j))_{1 \le i,j \le m}\big).$$

(5) $\chi_N \xrightarrow{}_{n \to \infty} \chi$ *w.r.t. sample subtree mass convergence.*

(6) *For all $m \in \mathbb{N}$ with $m \ge 3$ and $f \in \mathcal{C}_b([0, 1]^{m^3})$,*

$$(2.17)$$
$$\int \mu_N^{\otimes m}(d\underline{u})f\big((\mu_N(\mathcal{S}_{c_N(u_i, u_j, u_k)}(u_i)))_{i,j,k=1,\ldots,m}\big)$$
$$\xrightarrow[n \to \infty]{} \int \mu^{\otimes m}(d\underline{u})f\big((\mu(\mathcal{S}_{c(u_i, u_j, u_k)}(u_i)))_{i,j,k=1,\ldots,m}\big).$$

In what follows, we will need the following two subspaces of $\mathbb{T}_2$. Let for each $N \in \mathbb{N}$,

$$(2.18) \qquad \mathbb{T}_2^N := \left\{(T, c, \mu) \in \mathbb{T}_2 : \#\mathrm{lf}(T) = N \text{ and } \mu = \frac{1}{N}\sum_{u \in \mathrm{lf}(T)} \delta_u\right\},$$

and

$$(2.19) \qquad \mathbb{T}_2^{\mathrm{cont}} := \{(T, c, \mu) \in \mathbb{T}_2 : \mathrm{at}(\mu) = \varnothing\}.$$

The Aldous chain on N-cladograms is naturally defined on $\mathbb{T}_2^N$: for $\chi \in \mathbb{T}_2^N$, there is a unique (up to measure preserving tree isomorphism) minimal representative (T, c, μ) of χ (i.e., no subset with the restrictions of c and μ is an algebraic measure tree) with $2N - 2$ vertices and $2N - 3$ edges. We identify $\chi \in \mathbb{T}_2^N$ with this minimal representative and interpret it as "N-caldogram without labels" with uniform distribution on the leaves. We define the Aldous chain on $\mathbb{T}_2^N$ in the same way as the one on $\mathfrak{C}_N$, via its generator Ω_N in (1.1). With a slight abuse of notation, we use the same notation for the generators of the $\mathbb{T}_2^N$-valued and of the $\mathfrak{C}_N$-valued chain. We will define the Aldous diffusion on the space $\mathbb{T}_2^{\mathrm{cont}}$ in view of the following approximation result.

PROPOSITION 2.9 (Approximations with $\mathbb{T}_2^N$). *Let $\chi \in \mathbb{T}_2$. Then $\chi \in \mathbb{T}_2^{\mathrm{cont}}$ if and only if there exists for each $N \in \mathbb{N}$ an $\chi_N \in \mathbb{T}_2^N$ such that $\chi_N \to \chi$ in one (and thus all) of the equivalent notions of convergence on $\mathbb{T}_2$ given above.*

PROPOSITION 2.10 (Compactness and metrizability). *$\mathbb{T}_2$ is a compact, metrizable space. Both $\mathbb{T}_2^N$ and $\mathbb{T}_2^{\mathrm{cont}}$ are closed subspaces of $\mathbb{T}_2$, and thus compact as well.*

To deal with the Aldous chain and diffusions, it is convenient to introduce the following set of test functions on $\mathbb{T}_2$.

DEFINITION 2.11 (Shape polynomials). A *shape polynomial* is a linear combination of functions $\Phi^{m,\mathfrak{t}} \colon \mathbb{T}_2 \to \mathbb{R}$ of the form

$$(2.20) \qquad \Phi^{m,\mathfrak{t}}(\chi) := \mu^{\otimes m}\big(\mathfrak{s}_{(T,c)}^{-1}(\mathfrak{t})\big),$$

where $\chi = (T, c, \mu)$, $m \in \mathbb{N}$ and $\mathfrak{t} \in \overline{\mathfrak{C}}_m$. Let $\Pi_{\mathfrak{s}}$ be the set of all shape polynomials.

Apart from its combinatorial nature, the usefulness of shape polynomials stems from the fact that every real-valued continuous function on $\mathbb{T}_2$ can be approximated by them.

LEMMA 2.12. *The set $\Pi_{\mathfrak{s}}$ of shape polynomials is a uniformly dense sub-algebra of* $\mathcal{C}(\mathbb{T}_2)$.

PROOF. It is immediate from Propositions 2.8 and 2.10 that $\Pi_{\mathfrak{s}}$ is contained in the space $\mathcal{C}(\mathbb{T}_2)$ of continuous functions and separates the points of $\mathbb{T}_2$. To see that $\Pi_{\mathfrak{s}}$ is multiplicatively closed, consider for $k, m \in \mathbb{N}$, $k < m$ the projection $\pi_{m,k} \colon \overline{\mathfrak{C}}_m \to \overline{\mathfrak{C}}_k$, mapping $\mathfrak{t} = (C, c, \zeta) \in \overline{\mathfrak{C}}_m$ to the subcladogram spanned by $\zeta(1), \ldots, \zeta(k)$, as well as the projection $\tilde{\pi}_{m,k} \colon \overline{\mathfrak{C}}_m \to \overline{\mathfrak{C}}_{m-k}$, mapping to the subcladogram spanned by $\zeta(k+1), \ldots, \zeta(m)$ (relabelled to have labels in $\{1, \ldots, m-k\}$). Because μ is a probability and the product measure is used in (2.20), we have for $m, n \in \mathbb{N}$, $\mathfrak{t} \in \overline{\mathfrak{C}}_m$, $\tilde{\mathfrak{t}} \in \overline{\mathfrak{C}}_n$

$$(2.21) \qquad \Phi^{m,\mathfrak{t}} \cdot \Phi^{n,\tilde{\mathfrak{t}}} = \sum_{\mathfrak{t}' \in \pi_{n+m,m}^{-1}(\mathfrak{t}) \cap \tilde{\pi}_{n+m,n}^{-1}(\tilde{\mathfrak{t}})} \Phi^{m+n,\mathfrak{t}'} \in \Pi_{\mathfrak{s}}.$$

Because $\mathbb{T}_2$ is compact by Proposition 2.10, $\Pi_{\mathfrak{s}}$ is dense in $\mathcal{C}(\mathbb{T}_2)$ by the Stone–Weierstrass theorem. □

Consider $m \in \mathbb{N}$ and $\mathfrak{t} = (C, c, \zeta) \in \overline{\mathfrak{C}}_m \setminus \mathfrak{C}_m$, that is, there is at least one leaf in the m-labelled cladogram $\mathfrak{t}$ with multiple labels. Then $\mathfrak{s}_{(T,c)}(u_1, \ldots, u_m) = \mathfrak{t}$ implies that the $u_1, \ldots, u_m$ are not distinct. Hence, $\Phi^{m,\mathfrak{t}}(\chi) = 0$ for all $\chi \in \mathbb{T}_2^{\mathrm{cont}}$. This is in fact the reason why we restricted the domain of the pre-generator of the Aldous diffusion $\mathcal{D}(\Omega_{\mathrm{Ald}})$ to shape polynomials using m-cladograms instead of m-labelled cladograms (see (1.12)). Note that the set of restrictions to $\mathbb{T}_2^{\mathrm{cont}}$ of functions in $\mathcal{D}(\Omega_{\mathrm{Ald}})$ is dense in $\mathcal{C}(\mathbb{T}_2^{\mathrm{cont}})$.

3. Convergence of generators, tightness and existence.

In this section, we prepare the proofs of our main results by showing the uniform convergence of the generators of the discrete chains to the pre-generator $(\mathcal{D}(\Omega_{\mathrm{Ald}}), \Omega_{\mathrm{Ald}})$, and deduce tightness of the Aldous chains (provided tightness of initial conditions) as well as existence of solutions of the limiting martingale problem by general theory. We also obtain continuous paths of all limit processes.

A first simple observation about the pre-generator is that it maps $\mathcal{D}(\Omega_{\mathrm{Ald}})$ into itself.

LEMMA 3.1. *For every $\Phi \in \mathcal{D}(\Omega_{\mathrm{Ald}})$, we have $\Omega_{\mathrm{Ald}}\Phi \in \mathcal{D}(\Omega_{\mathrm{Ald}})$. In particular,* $(\Phi, \Omega_{\mathrm{Ald}}\Phi) \in \mathcal{C}(\mathbb{T}_2) \times \mathcal{C}(\mathbb{T}_2)$.

PROOF. Both Φ and $\Omega_{\mathrm{Ald}}\Phi$ are shape polynomials, hence continuous by definition of sample shape convergence. □

For $N \in \mathbb{N}$, recall from (1.1) the generator Ω_N of the Aldous chain on N-cladograms and from (2.18) the space $\mathbb{T}_2^N$ of algebraic measure trees with N leaves and uniform distribution on the leaves.

PROPOSITION 3.2 (Convergence of generators). *For all $\Phi \in \mathcal{D}(\Omega_{\mathrm{Ald}})$, we have*

$$(3.1) \qquad \lim_{N \to \infty} \sup_{\chi \in \mathbb{T}_2^N} |\Omega_N \Phi(\chi) - \Omega_{\mathrm{Ald}}\Phi(\chi)| = 0.$$

PROOF. Consider $\Phi \in \mathcal{D}(\Omega_{\mathrm{Ald}})$. By linearity, we may assume w.l.o.g. that $\Phi = \Phi^{m,\mathfrak{t}}$ for some $m \in \mathbb{N}$ and $\mathfrak{t} \in \mathfrak{C}_m$. In particular, $\mathfrak{t}$ is such that no leaf has multiple labels, and consequently for $\underline{u} \in T^m$, $\mathfrak{s}_{(T,c)}(\underline{u}) = \mathfrak{t}$ implies that $u_1, \ldots, u_m$ are distinct.

Fix $N \in \mathbb{N}$ and $\chi \in \mathbb{T}_2^N$, and let (T, c, μ) be the unique (up to measure preserving tree isomorphism) minimal representative (i.e., $\#T = 2N - 2$). Then $\#\mathrm{lf}(T) = N$ and $\#\mathrm{edge}(T) =$
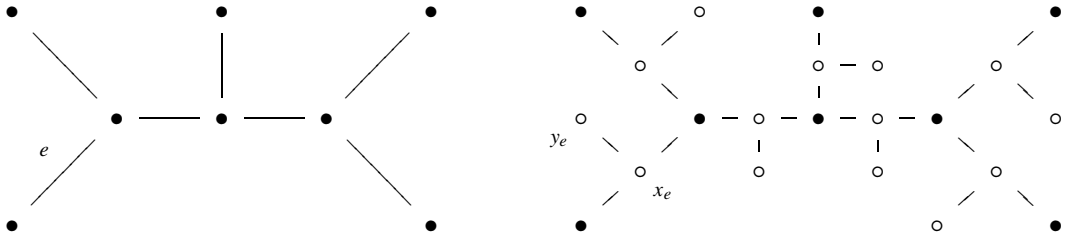
FIG. 4. *A finite algebraic tree (T, c) and the extended tree $(\overline{T}, \bar{c})$.*

$2N - 3$). In the following we abbreviate the inverse numbers of leaves and edges respectively, by

$$(3.2) \qquad \epsilon = \epsilon_N := \frac{1}{N}, \quad \text{and} \quad \delta = \delta_N := \frac{1}{2N - 3}.$$

We extend the algebraic tree to allow for potential new branch points (due to inserting an edge) and new leaves. To this end, for every edge $e \in \text{edge}(T)$, we introduce two additional points x_e, y_e. Informally, x_e is inserted in the middle of e, and y_e is attached to x_e as a leaf. More precisely, we consider

$$(3.3) \qquad \overline{T} = T \uplus \biguplus_{e \in \text{edge}(T)} \{x_e, y_e\},$$

and extend c to $\bar{c} \colon \overline{T}^3 \to \overline{T}$ which is uniquely defined by the following. $(\overline{T}, \bar{c})$ is an algebraic tree such that for $e = \{u, v\} \in \text{edge}(T)$, we have $x_e \in [u, v]$ in $(\overline{T}, \bar{c})$, and

$$(3.4) \qquad \bar{c}(y_e, x_e, z) = x_e \quad \forall z \in \overline{T} \setminus \{y_e\}.$$

The construction is illustrated in Figure 4. Note that $(\overline{T}, \bar{c}, \mu)$ is a binary algebraic measure tree equivalent to (T, c, μ).

For $k \in \{1, \ldots, m\}$ and $x \in \overline{T}$, let $\theta_{k,x} \colon \overline{T}^m \to \overline{T}^m$ be the *replacement operator* which replaces the kth-coordinate by x, that is,

$$(3.5) \qquad \theta_{k,x}(u_1, \ldots, u_m) = (u_1, \ldots, u_{k-1}, x, u_{k+1}, \ldots, u_m).$$

For $z = (x, e) \in \text{lf}(T) \times \text{edge}(T)$, let

$$(3.6) \qquad \chi_z := (\overline{T}, \bar{c}, \mu + \epsilon \delta_{y_e} - \epsilon \delta_x)$$

be the binary algebraic measure tree after the Aldous move with z. The difference between sampling with the new and old measure is given by

$$(\mu + \epsilon \delta_{y_e} - \epsilon \delta_x)^{\otimes m} - \mu^{\otimes m}$$

$$= \epsilon \sum_{k=1}^{m} \mu^{\otimes(k-1)} \otimes (\delta_{y_e} - \delta_x) \otimes \mu^{\otimes(m-k)}$$

$$(3.7) \qquad + \epsilon^2 \sum_{1 \leq k < j \leq m} \mu^{\otimes(k-1)} \otimes (\delta_{y_e} - \delta_x) \otimes \mu^{\otimes(j-k-1)} \otimes (\delta_{y_e} - \delta_x) \otimes \mu^{\otimes(m-j)} + \tilde{\mu}$$

$$= \epsilon \sum_{k=1}^{m} (\mu^{\otimes m} \circ \theta_{k, y_e}^{-1} - \mu^{\otimes m} \circ \theta_{k,x}^{-1}) - \epsilon^2 \sum_{j,k=1, j \neq k}^{m} \mu^{\otimes m} \circ \theta_{k, y_e}^{-1} \circ \theta_{j,x}^{-1} + \tilde{\mu}',$$

where $\tilde{\mu}$, $\tilde{\mu}'$ are signed measures on $\overline{T}^m$ vanishing on $\{(u_1, \ldots, u_m) : u_1, \ldots, u_m \text{ distinct}\}$. Thus

$$(3.8) \qquad \Omega_N \Phi^{m,\mathsf{t}}(\chi) = \sum_{z \in \mathrm{lf}(T) \times \mathrm{edge}(T)} (\Phi^{m,\mathsf{t}}(\chi_z) - \Phi^{m,\mathsf{t}}(\chi)) = \sum_{k=1}^m A_k - \sum_{j,k=1, j \neq k}^m B_{k,j},$$

with

$$(3.9) \quad A_k := \epsilon \sum_{(x,e) \in \mathrm{lf}(T) \times \mathrm{edge}(T)} \int_{T^m} \mu^{\otimes m}(d\underline{u}) \big(\mathbb{1}_{\mathsf{t}}(\mathfrak{s}_{(\overline{T},\overline{c})}(\theta_{k,y_e}\underline{u})) - \mathbb{1}_{\mathsf{t}}(\mathfrak{s}_{(\overline{T},\overline{c})}(\theta_{k,x}\underline{u})) \big),$$

and

$$(3.10) \qquad B_{k,j} := \epsilon^2 \sum_{(x,e) \in \mathrm{lf}(T) \times \mathrm{edge}(T)} \int_{T^m} \mu^{\otimes m}(d\underline{u}) \mathbb{1}_{\mathsf{t}}(\mathfrak{s}_{(\overline{T},\overline{c})}(\theta_{k,y_e} \circ \theta_{j,x}\underline{u})).$$

We use the notation $\mathsf{t}_{\wedge k} \in \mathfrak{C}_{m-1}$ for the $(m-1)$-cladogram obtained from t by deleting the leaf with label k (and relabelling the labels $j > k$ to $j - 1$), i.e., if $\mathsf{t} = \mathfrak{s}_{(\overline{T},\overline{c})}(\underline{u})$, then $\mathsf{t}_{\wedge k} = \mathfrak{s}_{(\overline{T},\overline{c})}(\underline{u}_{\wedge k})$ with $\underline{u}_{\wedge k} := (u_1, \ldots, u_{k-1}, u_{k+1}, \ldots, u_m)$. Furthermore, for $\underline{u} \in T^m$, we define

$$(3.11) \qquad E_{\mathsf{t},k}(\underline{u}) := \{v \in \overline{T} : \mathfrak{s}_{(\overline{T},\overline{c})}(\theta_{k,v}\underline{u}) = \mathsf{t}\}.$$

Note that $E_{\mathsf{t},k}(\underline{u})$ does not depend on u_k, contains no u_j for $j \neq k$, and that $E_{\mathsf{t},k}(\underline{u}) \neq \varnothing$ only if $\mathfrak{s}_{(T,c)}(\underline{u}_{\wedge k}) = \mathsf{t}_{\wedge k}$. In this case, $E_{\mathsf{t},k}(\underline{u})$ "corresponds to" an edge of $\mathsf{t}_{\wedge k}$. Let $\ell := \delta \sum_{e \in \mathrm{edge}(T)} \delta_{y_e}$ be the uniform distribution on $\{y_e : e \in \mathrm{edge}(T)\}$. By Fubini's theorem and using that $\epsilon \delta \sum_{(x,e) \in \mathrm{lf} \times \mathrm{edge}} \delta_x \otimes \delta_{y_e} = \mu \otimes \ell$, we obtain

$$
\begin{aligned}
(3.12) \qquad A_k &= \delta^{-1} \int_{\overline{T}^m} \mu^{\otimes m}(d\underline{u}) \big(\ell(E_{\mathsf{t},k}(\underline{u})) - \mu(E_{\mathsf{t},k}(\underline{u})) \big) \\
&= \int_{T^m} \mu^{\otimes m}(d\underline{u}) \mathbb{1}_{\mathsf{t}_{\wedge k}}((\mathfrak{s}_{(\overline{T},\overline{c})}\underline{u}_{\wedge k})) \cdot (3\mu(E_{\mathsf{t},k}(\underline{u})) + 1) \\
&= 3\Phi^{m,\mathsf{t}}(\chi) + \Phi^{m-1,\mathsf{t}_{\wedge k}}(\chi),
\end{aligned}
$$

where we have used in the second step that, because (T, c) is binary,

$$(3.13) \qquad \#\{e \in \mathrm{edge}(T) : c(y_e, z, z') \in (z, z')\} = 2\#\{x \in \mathrm{lf}(T) : c(x, z, z') \in (z, z')\} + 1$$

for every $z, z' \in T$, and hence $\delta^{-1}\ell(E_{\mathsf{t},k}(\underline{u})) = 2N\mu(E_{\mathsf{t},k}(\underline{u})) + 1$ if $\mathfrak{s}_{(\overline{T},\overline{c})}(\underline{u}_{\wedge k}) = \mathsf{t}_{\wedge k}$. Similarly,

$$
\begin{aligned}
(3.14) \qquad B_{k,j} &= \frac{\epsilon}{\delta} \int_{\overline{T}} \ell(dy) \int_T \mu(dx) \int_{T^m} \mu^{\otimes m}(d\underline{u}) \mathbb{1}_{\mathsf{t}}(\mathfrak{s}_{(\overline{T},\overline{c})} \circ \theta_{k,y} \circ \theta_{j,x}(\underline{u})) \\
&= \frac{\epsilon}{\delta} \Phi^{m,\mathsf{t}}(\chi) + \epsilon A_k \\
&= 2\Phi^{m,\mathsf{t}}(\chi) + \epsilon A_k + 3\epsilon \Phi^{m,\mathsf{t}}(\chi).
\end{aligned}
$$

Combining (3.8), (3.12) and (3.14), we obtain that

$$
\begin{aligned}
(3.15) \qquad \Omega_N \Phi^{m,\mathsf{t}}(\chi) &= \sum_{k=1}^m \Phi^{m-1,\mathsf{t}_{\wedge k}}(\chi) + (3m - 2m(m-1))\Phi^{m,\mathsf{t}}(\chi) - \epsilon(m-1)\sum_{k=1}^m A_k \\
&\quad - 3\epsilon m(m-1)\Phi^{m,\mathsf{t}}(\chi) \\
&= \sum_{k=1}^m \Phi^{m-1,\mathsf{t}_{\wedge k}}(\chi) - m(2m-5)\Phi^{m,\mathsf{t}}(\chi) - \epsilon(m-1)\sum_{k=1}^m \Phi^{m-1,\mathsf{t}_{\wedge k}}(\chi) \\
&\quad - 6\epsilon m(m-1)\Phi^{m,\mathsf{t}}(\chi).
\end{aligned}
$$

For an edge e of $t_{\wedge k}$, denote by $t^{(k,e)}$ the cladogram obtained by inserting a leaf labelled k in $t_{\wedge k}$ at the edge e (and relabelling the labels $j \geq k$ to $j + 1$). In particular, $t^{(k,e)}$ is the cladogram obtained from t by the Aldous move (k, e). For $\underline{u} \in T^m$, we have $\mathfrak{s}_{(T,c)}(\underline{u}_{\wedge k}) = t_{\wedge k}$ if and only if there is an edge e of $t_{\wedge k}$ such that $\mathfrak{s}_{(T,c)}(\underline{u}) = t^{(k,e)}$, and this e is unique. Hence,

$$\sum_{k=1}^{m} \Phi^{m-1, t_{\wedge k}}(\chi) = \int_{T^m} \mu^{\otimes m}(d\underline{u}) \sum_{k=1}^{m} \mathbb{1}_{t_{\wedge k}}\big(\mathfrak{s}_{(T,c)}(\underline{u}_{\wedge k})\big)$$

$$(3.16) \qquad = \int_{T^m} \mu^{\otimes m}(d\underline{u}) \sum_{k=1}^{m} \sum_{e \in \text{edge}(t_{\wedge k})} \mathbb{1}_{t^{(k,e)}}\big(\mathfrak{s}_{(T,c)}(\underline{u})\big)$$

$$= \int_{T^m} \mu^{\otimes m}(d\underline{u}) \Omega_m \mathbb{1}_t\big(\mathfrak{s}_{(T,c)}(\underline{u})\big) + m \# \text{edge}(t_{\wedge k}) \Phi^{m,t}(\chi).$$

Inserting this into (3.15) and using that $\# \text{edge}(t_{\wedge k}) = 2m - 5$, we see that

$$(3.17) \qquad \left| \Omega_N \Phi^{m,t}(\chi) - \int_{T^m} (\Omega_m \mathbb{1}_t) \circ \mathfrak{s}_{(T,c)} \, d\mu^{\otimes m} \right| \leq 7m(m-1)\epsilon,$$

which gives the claim. $\quad\square$

As $\mathbb{T}_2$ is compact by Proposition 2.10, we can immediately conclude the following from the convergence of the generators.

COROLLARY 3.3 (The limiting martingale problem). *Let $(\chi_N)_{N \in \mathbb{N}}$ be a sequence of random binary algebraic measure trees with $\chi_N \in \mathbb{T}_2^N$, such that*

$$(3.18) \qquad \chi_N \Rightarrow \chi \quad \text{as } N \to \infty,$$

where χ is distributed according to P_0 on $\mathbb{T}_2^{\text{cont}}$. Let $X^N = (X_t^N)_{t \geq 0}$ be the Aldous chain started in χ_N. Then the sequence $(X^N)_{N \in \mathbb{N}}$ is tight in Skorokhod path space. Any limit process $(X_t)_{t \geq 0}$ has continuous, $\mathbb{T}_2^{\text{cont}}$-valued paths, and satisfies the $(\Omega_{\text{Ald}}, \mathcal{D}(\Omega_{\text{Ald}}), P_0)$-martingale problem.

PROOF. *Tightness.* Tightness follows, in view of the approximation result Proposition 2.9 and Lemma 3.1, with the exactly same proof as Theorems 3.9.1 and 3.9.4 in [9] (see also [9], Remark 4.5.2).

Continuous paths. For $\Phi \in \mathcal{D}(\Omega_{\text{Ald}})$, let $\Phi(X^N) = (\Phi(X_t^N))_{t \geq 0}$. By definition, $\mathcal{D}(\Omega_{\text{Ald}})$ induces the topology of sample-shape convergence on $\mathbb{T}_2^{\text{cont}}$. Hence, continuity of the paths of the limit process $X = (X_t)_{t \geq 0}$ in $\mathbb{T}_2^{\text{cont}}$ is equivalent to path-continuity of $\Phi(X)$ for all $\Phi \in \mathcal{D}(\Omega_{\text{Ald}})$. Because $\Phi(X)$ is the limit of $\Phi(X^N)$, this follows from the obvious estimate $|\Phi(X_t^N) - \Phi(X_{t-}^N)| \leq \frac{m}{N}$ for $\Phi = \Phi^{m,t}$.

Values in $\mathbb{T}_2^{\text{cont}}$. That any limit point has $\mathbb{T}_2^{\text{cont}}$-valued paths follows directly from the fact that X^N is $\mathbb{T}_2^N$-valued, together with the approximation result Proposition 2.9.

Martingale problem. That all limit points satisfy the martingale problem follows with the same proof as Lemma 4.5.1 in [9]. $\quad\square$

The following corollary is immediate from the previous corollary and the approximation result Proposition 2.9.

COROLLARY 3.4 (Existence). *For any probability measure P_0 on $\mathbb{T}_2^{\text{cont}}$ there exists a solution in $\mathcal{C}_{\mathbb{T}_2^{\text{cont}}}(\mathbb{R}_+)$ of the $(\Omega_{\text{Ald}}, \mathcal{D}(\Omega_{\text{Ald}}), P_0)$-martingale problem.*

4. Duality, uniqueness and convergence.   In this section, we first obtain a duality result that in turn allows to conclude the uniqueness of the martingale problem. We also use duality to show that the Aldous diffusion is a Feller process on $\mathbb{T}_2^{\mathrm{cont}}$. For $m \in \mathbb{N}$ let $Y^m := (Y_t^m)_{t \geq 0}$ be the $\mathfrak{C}_m$-valued Aldous chain with generator Ω_m from (1.1). If $Y_0^m = \mathfrak{t} \in \mathfrak{C}_m$, then $\mathbb{E}_{\mathfrak{t}}^Y$ denotes the corresponding expectation.

PROPOSITION 4.1 (Duality).   *Let P_0 be an arbitrary probability measure on $\mathbb{T}_2^{\mathrm{cont}}$ and let $X := ((T_t, c_t, \mu_t))_{t \geq 0}$ be a solution of the $(\Omega_{\mathrm{Ald}}, \mathcal{D}(\Omega_{\mathrm{Ald}}), P_0)$-martingale problem in $\mathcal{D}_{\mathbb{T}_2^{\mathrm{cont}}}(\mathbb{R}_+)$. For arbitrary $m \in \mathbb{N}$ and $\mathfrak{t} \in \mathfrak{C}_m$, let $Y^m := (Y_t^m)_{t \geq 0}$ be the $\mathfrak{C}_m$-valued Aldous chain with $Y_0^m = \mathfrak{t}$. Assume that Y^m is independent of X. Then*

(4.1)
$$
\begin{aligned}
&\mathbb{E}_{P_0}^X\big[\mu_t^{\otimes m}\{\underline{u} \in T_t^m : \mathfrak{s}_{(T_t, c_t)}(\underline{u}) = \mathfrak{t}\}\big] \\
&\qquad = \int_{\mathbb{T}_2^{\mathrm{cont}}} \mathbb{E}_{\mathfrak{t}}^Y\big[\mu^{\otimes m}\{\underline{u} \in T^m : \mathfrak{s}_{(T,c)}(\underline{u}) = Y_t^m\}\big] P_0\big(\mathrm{d}(T, c, \mu)\big).
\end{aligned}
$$

PROOF.   Let $m \in \mathbb{N}$. For $\chi = (T, c, \mu) \in \mathbb{T}_2^{\mathrm{cont}}$ and $\mathfrak{t} \in \mathfrak{C}_m$, we define $H(\chi, \mathfrak{t}) := \mu^{\otimes m}\{\underline{u} \in T^m : \mathfrak{s}_{(T,c)}(\underline{u}) = \mathfrak{t}\}$. Then

(4.2)
$$
\Omega_{\mathrm{Ald}} H(\cdot, \mathfrak{t})(\chi) = \int_{T^m} \mu^{\otimes m}(\mathrm{d}\underline{u}) \Omega_m \mathbb{1}_{\mathfrak{t}}\big(\mathfrak{s}_{(T,c)}(\underline{u})\big) = \Omega_m H(\chi, \cdot)(\mathfrak{t}).
$$

By our assumptions on the test functions H and definitions of Ω_{Ald} and Ω_m, the result follows by [9], Lemma 4.4.11, Corollary 4.4.13. $\square$

COROLLARY 4.2 (Uniqueness of the martingale problem).   *Let P_0 be an arbitrary probability measure on $\mathbb{T}_2^{\mathrm{cont}}$. Then uniqueness holds for the $(\Omega_{\mathrm{Ald}}, \mathcal{D}(\Omega_{\mathrm{Ald}}), P_0)$-martingale problem in $\mathcal{D}_{\mathbb{T}_2^{\mathrm{cont}}}(\mathbb{R}_+)$.*

PROOF.   As the set of all shape polynomials is separating (for probability measures), the result is immediate by the previous proposition and Proposition 4.4.7 from [9]. $\square$

COROLLARY 4.3 (Feller process).   *For $F \in \mathcal{C}(\mathbb{T}_2^{\mathrm{cont}})$, $t \geq 0$, and $\chi \in \mathbb{T}_2^{\mathrm{cont}}$, let*

(4.3)
$$
S_t F(\chi) := \mathbb{E}_\chi\big(F(X_t)\big),
$$

where, under $\mathbb{E}_\chi$, $X = (X_t)_{t \geq 0}$ is the Aldous diffusion on $\mathbb{T}_2^{\mathrm{cont}}$ started in χ. Then $(S_t)_{t \geq 0}$ is a Feller semi-group. In particular, the Aldous diffusion is a strong Markov process.

PROOF. $(S_t)_{t \geq 0}$ is well-defined by existence and uniqueness shown in Corollaries 3.4 and 4.2. It is a semi-group on the set of bounded measurable functions on $\mathbb{T}_2^{\mathrm{cont}}$ by the Markov property of X, which in turn follows from uniqueness and Theorem 4.4.2(a) in [9]. Recall from Proposition 2.10 and Lemma 2.12 that the state space $\mathbb{T}_2^{\mathrm{cont}}$ is compact, and the set $\mathcal{D}(\Omega_{\mathrm{Ald}})$ of shape polynomials is uniformly dense in $\mathcal{C}(\mathbb{T}_2^{\mathrm{cont}})$. Hence, in order to show that S_t maps $\mathcal{C}(\mathbb{T}_2^{\mathrm{cont}})$ into itself, it is enough to show $S_t F \in \mathcal{C}(\mathbb{T}_2^{\mathrm{cont}})$ for $F = \Phi^{m,\mathfrak{t}} \in \mathcal{D}(\Omega_{\mathrm{Ald}})$. Using duality, we have $S_t \Phi^{m,\mathfrak{t}}(\chi) = \mathbb{E}[\Phi^{m,Y_t}(\chi)]$ for the $\mathfrak{C}_m$-valued Aldous chain started in $\mathfrak{t}$. Thus $S_t \Phi^{m,\mathfrak{t}} \in \mathcal{C}(\mathbb{T}_2^{\mathrm{cont}})$, because it is a finite linear combination of the continuous functions $\Phi^{m,\mathfrak{t}'}$ for different $\mathfrak{t}' \in \mathfrak{C}_m$.

We have shown that $(S_t)_{t \geq 0}$ is a contraction semigroup on $\mathcal{C}(\mathbb{T}_2^{\mathrm{cont}})$. Its weak continuity follows directly from continuity of the sample paths of X. Weak continuity implies strong continuity, for example, by Theorem 19.6 of [20]. Therefore, X is a Feller process. This also implies the strong Markov property (see, e.g., [9], Theorem 4.2.7). $\square$

5. Long term behavior and the Brownian CRT. In this section, we define the algebraic measure Brownian CRT, and provide the joint density of the cladogram shape spanned by a sample of finite size together with the vector of subtree masses branching off the edges of the cladogram. Moreover, we show that the algebraic measure Brownian CRT is invariant under the Aldous diffusion and that for any initial $\chi \in \mathbb{T}_2^N$, the Aldous diffusion converges in law to the algebraic measure Brownian CRT as time goes to infinity.

Recall the definition of the set $\mathfrak{C}_m$ of m-cladograms (after Definition 2.4) and the shape $\mathfrak{s}_{(T,c)}(u_1, \ldots, u_m)$ spanned by the vector of m points $u_1, \ldots, u_m \in T$ (Definition 2.5). We define the following.

DEFINITION 5.1 (Algebraic measure Brownian CRT). The algebraic measure Brownian CRT is the unique (in distribution) random binary algebraic measure tree $X_{\mathrm{CRT}} = (T, c, \mu)$ with uniform annealed sample shape distribution, that is, for all $m \in \mathbb{N}$, for all $\mathfrak{t} \in \mathfrak{C}_m$,

$$(5.1) \qquad \mathbb{E}_{\mathrm{CRT}}\big[\mu^{\otimes m}\{(u_1, \ldots, u_m) : \mathfrak{s}_{(T,c)}(u_1, \ldots, u_m) = \mathfrak{t}\}\big] = \frac{1}{\#\mathfrak{C}_m}.$$

Note that there is a unique law on $\mathbb{T}_2$ satisfying (5.1) because the sample shape distribution separates probability measures on $\mathbb{T}_2$, and it is realized through the well-known Brownian CRT once we ignore the distances (compare, [1], Theorem 23).

Now we provide the analog of [1], Theorem 23, by considering, together with the sample shape, the vector of masses of the subtrees branching off the edges of the shape cladogram. As expected, under the annealed law of the Brownian CRT, we obtain that this vector is Dirichlet distributed and independent of the shape. To state the result more precisely, for $\underline{u} = (u_1, \ldots, u_m) \in T$ let $T(\underline{u}) = c(\{u_1, \ldots, u_m\}^3)$ be the generated subtree, and for $e = \{e_x, e_y\} \in \mathrm{edge}(T(\underline{u}))$ let

$$(5.2) \qquad \eta_{(T,c,\mu)}(\underline{u}, e) := \mu\big(\{v \in T : c(v, e_x, e_y) \in (e_x, e_y) \cup (\{e_x, e_y\} \cap \mathrm{lf}(T(\underline{u})))\}\big).$$

Let $\eta_{(T,c,\mu)}(\underline{u}) = (\eta_{(T,c,\mu)}(\underline{u}, e))_{e \in \mathrm{edge}(T(\underline{u}))}$ be the vector of these $2m - 3$ masses (assuming $u_1, \ldots, u_m$ are distinct). We obtain the following, which is proven in the special case $m = 3$ in [2], Theorem 2.

PROPOSITION 5.2 (Brownian CRT and $\mathrm{Dir}(\frac{1}{2}, \ldots, \frac{1}{2})$). *Let X_{CRT} be the Brownian CRT, $m \in \mathbb{N}$, $\mathfrak{t} \in \mathfrak{C}_m$ and $f : \Delta_{2m-3} \to \mathbb{R}$ bounded measurable, where Δ_k is the k-simplex for $k \in \mathbb{N}$, that is,*

$$(5.3) \qquad \Delta_k := \{x \in [0, 1]^k : x_1 + \cdots + x_k = 1\}.$$

Then the following holds:

$$\mathbb{E}_{\mathrm{CRT}}\bigg[\int \mu^{\otimes m}(\mathrm{d}\underline{u}) \mathbb{1}_{\mathfrak{t}}\big(\mathfrak{s}_{(T,c)}(\underline{u})\big) f\big(\eta_{(T,c,\mu)}(\underline{u})\big)\bigg]$$

$$(5.4) \qquad = \frac{1}{\#\mathfrak{C}_m} \int_{\Delta_{2m-3}} f(\underline{x}) \mathrm{Dir}\Big(\frac{1}{2}, \ldots, \frac{1}{2}\Big)(\mathrm{d}\underline{x})$$

$$= \frac{\Gamma(m - \frac{3}{2})}{\#\mathfrak{C}_m \Gamma(\frac{1}{2})^{2m-3}} \int_{\Delta_{2m-3}} f(\underline{x})(x_1 \cdot \ldots \cdot x_{2m-3})^{-\frac{1}{2}} \, \mathrm{d}\underline{x},$$

where $\mathrm{Dir}(\frac{1}{2}, \ldots, \frac{1}{2})$ is the Dirichlet distribution and the last integral in (5.4) is w.r.t. Lebesgue measure on Δ_{2m-3}.

PROOF.   We follow Aldous's proof of [2], Theorem 2, and show a local limit theorem for the subtree mass distribution of a uniform N-cladogram (with uniform distribution on the leaves) as $N \to \infty$. Because the uniform N-cladogram converges to the Brownian CRT, this implies the claim.

Fix $m \in \mathbb{N}$ and $\mathfrak{t} \in \mathfrak{C}_m$. Let $N \geq m$. Denote by $\pi_{N,m} : \mathfrak{C}_N \to \mathfrak{C}_m$ the projection map which sends an N-cladogram (T, c, ζ) to the m-cladogram spanned by the first m leaves, that is, for $T_m := c(\zeta(\{1, \ldots, m\})^3)$,

$$(5.5) \qquad \pi_{N,m}(T, c, \zeta) := \left(T_m, c{\restriction}_{T_m}, \zeta{\restriction}_{\{1, \ldots, m\}} \right).$$

For $n = (n_e)_{e \in \text{edge}(\mathfrak{t})}$ with $\sum_e n_e = N$, let $q_N(n)$ be the probability that the first m leaves of a uniform N-cladogram span the m-cladogram $\mathfrak{t}$, and the numbers of leaves of the N-cladogram in the subtrees corresponding to the edges of $\mathfrak{t}$ are given by the vector n, that is,

$$
(5.6) \quad
\begin{aligned}
q_N(n) &:= \frac{1}{\#\mathfrak{C}_N} \#\left\{ (T, c, \zeta) \in \pi_{N,m}^{-1}(\mathfrak{t}) : N \cdot \eta_{(T,c,\zeta)}(\zeta(1), \ldots, \zeta(m)) = n \right\} \\
&= \frac{1}{\#\mathfrak{C}_N} \cdot \frac{(N - m)! \prod_{e \in \text{ex-edge}(\mathfrak{t})} \#\mathfrak{C}_{n_e+1} \prod_{e \in \text{in-edge}(\mathfrak{t})} \#\mathfrak{C}_{n_e+2}}{\prod_{e \in \text{ex-edge}(\mathfrak{t})} (n_e - 1)! \prod_{e \in \text{in-edge}(\mathfrak{t})} n_e!}.
\end{aligned}
$$

The first factor in the numerator together with the denominator counts the number of possibilities to distribute the $N - m$ remaining leaves to the edges of the m-cladogram $\mathfrak{t}$ (with quantities specified by n), and the products in the numerator count the possibilities to give cladogram structure to the leaves associated to every edge of $\mathfrak{t}$. For an external edge $e \in \text{ex-edge}(\mathfrak{t})$, this is the number of $(n_e + 1)$-cladograms (identify the additional leaf with the branch point of $\mathfrak{t}$ it is attached to). For an internal edge, we need two additional leaves. Recall that

$$(5.7) \qquad \#\mathfrak{C}_N = (2N - 5)!! = \frac{(2N - 4)!}{2^{N-2}(N - 2)!} \approx (N - 2)! \cdot 2^{(N-2)} (\pi(N - 2))^{-\frac{1}{2}},$$

where $\approx$ means that the quotient tends to 1 as $N \to \infty$, and we have applied the Stirling formula.

Fix $\eta = (\eta_1, \ldots, \eta_{2m-3}) \in \Delta_{2m-3}$ with $\eta_i > 0$ for $i = 1, \ldots, 2m - 3$. For $n_i = n_i(N)$ with $\sum_{i=1}^{2m-3} n_i = N$ and $N^{-1} n_i \to \eta_i$ (in particular $n_i \to \infty$), as $N \to \infty$, we obtain (using the convention that the first m edges are external)

$$
\begin{aligned}
q_N(n) &= \frac{1}{\#\mathfrak{C}_N} \cdot (N - m)! \cdot \prod_{i=1}^{m} \frac{\#\mathfrak{C}_{n_i+1}}{(n_i - 1)!} \prod_{i=m+1}^{2m-3} \frac{\#\mathfrak{C}_{n_i+2}}{n_i!} \\
&\approx \frac{\sqrt{\pi(N - 2)}}{(N - 2)! \cdot 2^{N-2}} \cdot (N - m)! \prod_{i=1}^{m} 2^{n_i - 1} (\pi(n_i - 1))^{-\frac{1}{2}} \prod_{i=m+1}^{2m-3} 2^{n_i} (\pi n_i)^{-\frac{1}{2}} \\
(5.8) \qquad &= \sqrt{(N - 2)} \cdot \frac{(N - m)!}{(N - 2)!} \cdot 2^{2-m} \pi^{\frac{1}{2} - \frac{2m-3}{2}} \prod_{i=1}^{m} (n_i - 1)^{-\frac{1}{2}} \prod_{i=m+1}^{2m-3} n_i^{-\frac{1}{2}} \\
&\approx N^{-(m-\frac{5}{2})} \cdot (2\pi)^{2-m} \cdot N^{-(m-\frac{3}{2})} \cdot (\eta_1 \eta_2 \cdot \ldots \cdot \eta_{2m-3})^{-\frac{1}{2}} \\
&= N^{-(2m-4)} \cdot \frac{1}{\#\mathfrak{C}_m} \frac{\Gamma(\frac{2m-3}{2})}{\Gamma(\frac{1}{2})^{2m-3}} (\eta_1 \eta_2 \cdot \ldots \cdot \eta_{2m-3})^{-\frac{1}{2}}.
\end{aligned}
$$

This gives the claimed Dirichlet density on the $(2m - 3)$-simplex in the limit. $\square$

PROPOSITION 5.3 (Convergence to the CRT). *Let $(X_t)_{t \geq 0}$ be the Aldous diffusion started in $\chi \in \mathbb{T}_2^{\text{cont}}$, and $\mathcal{X}_{\text{CRT}}$ the algebraic measure Brownian CRT. Then*

$$(5.9) \qquad\qquad X_t \underset{t \to \infty}{\Longrightarrow} \mathcal{X}_{\text{CRT}}.$$

In particular, the algebraic measure Brownian CRT is the unique invariant distribution of the Aldous diffusion.

PROOF. Fix $m \in \mathbb{N}$ and $\mathfrak{t} \in \mathfrak{C}_m$. Let $(Y_t)_{t \geq 0}$ be the Aldous chain on m-caldograms started in $Y_0 = \mathfrak{t}$. Then, for $\Phi^{m,\mathfrak{t}} \in \mathcal{D}(\Omega_{\text{Ald}})$ as in (2.20), we have by duality (Proposition 4.1)

$$(5.10) \qquad\qquad \mathbb{E}\big(\Phi^{m,\mathfrak{t}}(X_t)\big) = \sum_{\mathfrak{t}' \in \mathfrak{C}_m} \mathbb{P}\{Y_t = \mathfrak{t}'\} \Phi^{m,\mathfrak{t}'}(\chi).$$

Because the uniform distribution on $\mathfrak{C}_m$ is the unique reversible distribution of the Aldous chain (see [3]), $\lim_{t \to \infty} \mathbb{P}\{Y_t = \mathfrak{t}'\} = \frac{1}{\#\mathfrak{C}_m}$ for every $\mathfrak{t}' \in \mathfrak{C}_m$. Because $\sum_{\mathfrak{t}' \in \mathfrak{C}_m} \Phi^{m,\mathfrak{t}'} = 1$ on $\mathbb{T}_2^{\text{cont}}$, this means

$$(5.11) \qquad\qquad \lim_{t \to \infty} \mathbb{E}\big(\Phi^{m,\mathfrak{t}}(X_t)\big) = \frac{1}{\#\mathfrak{C}_m} = \mathbb{E}\big(\Phi^{m,\mathfrak{t}}(\mathcal{X}_{\text{CRT}})\big).$$

Because $\mathcal{D}(\Omega_{\text{Ald}})$ is convergence determining for probability measures on $\mathbb{T}_2^{\text{cont}}$, this proves (5.9). Invariance of the law of $\mathcal{X}_{\text{CRT}}$ follows from the convergence (5.9) together with the Feller property (Corollary 4.3). $\square$

In summary, we have now proven the first two theorems.

PROOF OF THEOREMS 1 AND 2. Well-posedness of the martingale problem is shown in Corollaries 3.4 and 4.2. Continuous paths and tightness of the sequence of Aldous chains are shown in Corollary 3.3. Furthermore, every limit process satisfies the martingale problem (Corollary 3.3) and this implies convergence because of the uniqueness shown in Corollary 4.2. The Feller property is shown in Corollary 4.3, and unique ergodicity with the algebraic measure Brownian CRT as invariant distribution is shown in Proposition 5.3. $\square$

6. On the dynamics of the sample subtree mass vector. In this section, we further study the Aldous diffusion on binary, algebraic nonatomic measure trees and prove Theorem 3. Recall from Proposition 5.2 that under the annealed law of the Brownian CRT, the sample tree shape is uniform and independent of the vector of subtree masses branching of the cladogram spanned by the sample. Furthermore, the vector of subtree masses is Dirichlet distributed. Next, we study the infinitesimal evolution of the quenched law of this vector under the dynamics of the Aldous diffusion in the case of sample size $m = 3$.

Recall the definition of the components $\mathcal{S}_v(u)$, $u, v \in T$, from (2.2), and from (1.14) that $\underline{\eta}(\underline{u})$ with $\underline{u} = (u_1, u_2, u_3) \in T^3$ denotes the vector of the three masses of the components connected to $c(\underline{u})$, that is,

$$(6.1) \qquad\qquad \underline{\eta}(\underline{u}) = \big(\eta_i(\underline{u})\big)_{i=1,2,3} = \big(\mu(\mathcal{S}_{c(\underline{u})}(u_i))\big)_{i=1,2,3}.$$

With a slight abuse of notation, we also denote for $v \in \text{br}(T)$ by $\underline{\eta}(v)$ the μ-masses of the three components of $T \setminus \{v\}$ (ordered decreasingly for definiteness), so that $\underline{\eta}(\underline{u}) = \underline{\eta}(c(\underline{u}))$ up to a permutation of the entries of the vector. Also recall that for mass polynomials

$$(6.2) \qquad\qquad \Phi^f(\chi) = \int f\big(\underline{\eta}(c(\underline{u}))\big) \mu^{\otimes 3}(\mathrm{d}\underline{u})$$

with $f \in \mathcal{C}^2([0, 1]^3)$ and $\chi = (T, c, \mu) \in \mathbb{T}_2$, we define

$$
\Omega_{\mathrm{Ald}}\Phi^f(T, c, \mu) = \int_{T^3} \mathrm{d}\mu^{\otimes 3}\bigg(2 \sum_{i,j=1}^{3} \eta_i(\delta_{ij} - \eta_j)\partial_{ij}^2 f(\underline{\eta}) + 3 \sum_{i=1}^{3}(1 - 3\eta_i)\partial_i f(\underline{\eta})
$$

(6.3)

$$
+ \frac{1}{2} \sum_{i,j=1, i \neq j}^{3} \Theta_{i,j} f(\underline{\eta}) + \sum_{i=1}^{3}(f(e_i) - f(\underline{\eta}))\bigg);
$$

see (1.15) and (1.17), and the definition of migration operators $\Theta_{i,j}$ in (1.16). Note that $\Phi^f \in \mathcal{C}(\mathbb{T}_2)$ by Proposition 2.8, and because the functions $\Theta_{ij} f$ are continuous due to continuous differentiability of f, $\Omega_{\mathrm{Ald}}\Phi^f$ is also a mass-polynomial. In particular, $\Omega_{\mathrm{Ald}}\Phi^f \in \mathcal{C}(\mathbb{T}_2^{\mathrm{cont}})$.

REMARK 6.1.

1. If $f \in \mathcal{C}^2([0, 1]^3)$ is symmetric, we can use the symmetry of the sampling procedure and rewrite (6.3) as

(6.4)
$$
\Omega_{\mathrm{Ald}}\Phi^f(\chi) = 3 \int_{T^3} \mathrm{d}\mu^{\otimes 3}(2\eta_1(1 - \eta_1)\partial_{11}^2 f(\underline{\eta}) - 4\eta_1\eta_2\partial_{12}^2 f(\underline{\eta}) + 3(1 - 3\eta_1)\partial_1 f(\underline{\eta})
$$
$$
+ \Theta_{1,2} f(\underline{\eta}) + f(1, 0, 0) - f(\underline{\eta})).
$$

This helps to reduce the number of terms in explicit calculations.

2. We can often assume f to be symmetric. If f is not necessarily symmetric, we use that $\Phi^f = \Phi^{\tilde{f}}$ for the symmetrization $\tilde{f}$ of f defined as follows. Given a permutation π of $\{1, 2, 3\}$, define $f_\pi(x_1, x_2, x_3) := f(x_{\pi(1)}, x_{\pi(2)}, x_{\pi(3)})$. Then $\tilde{f} = \frac{1}{6} \sum_{\pi \in S_3} f_\pi$.

For the proof of Theorem 3, we do not use the martingale problem of Theorem 1, because approximating the mass polynomial of degree three by shape polynomials, though possible in theory, seems difficult in praxis. Instead, we show that uniform convergence of generators holds also for mass polynomials of degree three, and use the diffusion approximation of Theorem 2. For $N \in \mathbb{N}$, recall the state space $\mathbb{T}_2^N$ from (2.18) and the Aldous chain with generator Ω_N from (1.1).

PROPOSITION 6.2 (Subtree mass under the Aldous diffusion). *For all test functions Φ^f of the form* (1.15) *with $f : [0, 1]^3 \to \mathbb{R}$ twice continuously differentiable,*

(6.5)
$$
\lim_{N \to \infty} \sup_{\chi \in \mathbb{T}_2^N} |\Omega_N\Phi^f(\chi) - \Omega_{\mathrm{Ald}}\Phi^f(\chi)| = 0.
$$

From here, we can prove Theorem 3 by standard arguments.

PROOF OF THEOREM 3. Let $X = (X_t)_{t \geq 0}$ be the Aldous diffusion on $\mathbb{T}_2^{\mathrm{cont}}$. Due to Proposition 2.9, there exist $\mathbb{T}_2^N$-valued random variables X_0^N such that X_0^N converges in law to X_0. By Theorem 2, the Aldous chains X^N started in X_0^N converge in law to X. Furthermore, Ω_{Ald} maps into $\mathcal{C}(\mathbb{T}_2)$. Hence, the same proof as Lemma 4.5.1 in [9] shows that (1.18) follows from Proposition 6.2. $\square$

Recall the intrinsic metric r_μ on an algebraic measure tree (T, c, μ), as defined in (1.5). Before proving Proposition 6.2, we show how to use the extended martingale problem from Theorem 3 to calculate the annealed average r_μ-distance in the algebraic measure Brownian CRT.

COROLLARY 6.3 (Mean average distance of the Brownian CRT). *Let $\mathcal{X}_{CRT} = (T, c, \mu)$ be the algebraic measure Brownian CRT. Then*

$$(6.6) \qquad \mathbb{E}_{CRT}\left[\int_{T^2} r_\mu(x, y)\mu^{\otimes 2}(\mathrm{d}(x, y))\right] = \frac{2}{5}.$$

PROOF. First, we express the average distance $\Phi(\chi) := \int r_\mu(x, y)\mu^{\otimes 2}(\mathrm{d}(x, y))$ for $\chi = (T, c, \mu) \in \mathbb{T}_2^{cont}$ as mass polynomial of degree three. Recall the branch point distribution $\nu = \mu^{\otimes 3} \circ c^{-1}$ used in the definition of r_μ.

$$
\begin{aligned}
(6.7) \qquad \Phi(\chi) &= \int \sum_{v \in \mathrm{br}(T) \cap (x,y)} \nu(\{v\})\mu^{\otimes 2}(\mathrm{d}(x, y)) \\
&= \sum_{v \in \mathrm{br}(T)} \nu(\{v\})\mu^{\otimes 2}(\{(x, y) : v \in (x, y)\}) \\
&= 2 \sum_{v \in \mathrm{br}(T)} \nu(\{v\})\big(\eta_1(v)\eta_2(v) + \eta_2(v)\eta_3(v) + \eta_1(v)\eta_3(v)\big) \\
&= 2 \int \big(\eta_1(\underline{u})\eta_2(\underline{u}) + \eta_2(\underline{u})\eta_3(\underline{u}) + \eta_1(\underline{u})\eta_3(\underline{u})\big)\mu^{\otimes 3}(\mathrm{d}\underline{u}),
\end{aligned}
$$

that is, $\Phi = 2\Phi^f$ (on $\mathbb{T}_2^{cont}$) with $f(x, y, z) := xy + yz + xz$. From here, we could obtain (6.6) by a direct computation using Proposition 5.2. But it also follows easily from the invariance of the CRT under the Aldous diffusion: Because f is symmetric, we obtain from Remark 6.1

$$
\begin{aligned}
(6.8) \qquad \Omega_{Ald}\Phi^f(\chi) &= 3\int\left(-4\eta_1\eta_2 + 3(1 - 3\eta_1)(\eta_2 + \eta_3) + \frac{(\eta_1 + \eta_2)\eta_3 - f(\underline{\eta})}{\eta_1} - 3\eta_1\eta_2\right)\mathrm{d}\mu^{\otimes 3} \\
&= 3\int(5\eta_1 - 25\eta_1\eta_2)\,\mathrm{d}\mu^{\otimes 3} = 5 - 25\Phi^f(\chi).
\end{aligned}
$$

Thus, $\mathbb{E}_{CRT}[\Omega_{Ald}\Phi^f] = 0$ implies $\mathbb{E}_{CRT}[\Phi^f(\chi)] = \frac{1}{5}$. $\square$

To prove Proposition 6.2, fix $N \in \mathbb{N}$ and $\chi = (T, c, \mu) \in \mathbb{T}_2^N$. We use some notation from the proof of Proposition 3.2, in particular recall from (3.2) that ϵ and δ denote the inverse numbers of leaves and edges, respectively, and the extended tree $(\overline{T}, \overline{c})$ which allows to represent one Aldous move on the same tree (see Figure 4). We consider μ, ν, and η to be defined on $(\overline{T}, \overline{c})$ and, for $z \in \mathrm{lf}(T) \times \mathrm{edge}(T)$, we denote by μ_z, ν_z, and η^z the corresponding objects after the Aldous move z. Because our trees are binary, the relation between the fraction of leaves and the fraction of edges in a subtree can be easily related as follows.

LEMMA 6.4 (Proportion of leaves versus edges). *Let $\chi = (T, c, \mu) \in \mathbb{T}_2^N$, and $S = \mathcal{S}_v(u)$ for some $v \in \mathrm{br}(T)$, $u \in T \setminus \{v\}$ a component. Let $\ell(S)$ be the fraction of the edges contained in S (including the edge to v). Then*

$$(6.9) \qquad \ell(S) = \mu(S) \cdot (1 + 3\delta) - \delta.$$

Recall the branch point distribution $\nu = \mu^{\otimes 3} \circ c^{-1}$. The next lemma shows the effect we would see if the branch point distribution remained unchanged. This corresponds exactly to Aldous's original calculation (compare with (1.2) but notice that our chain runs at total rate $N(2N - 3)$ rather than N^2 as in Aldous's case). In what follows, we write $O(\epsilon)$ for terms which divided by $\epsilon = \epsilon_N$ are bounded uniformly in the tree (and N), while the bound may depend on f, and similarly for $o(\epsilon)$ and so on.

LEMMA 6.5 (Wright–Fisher term with negative drift). *Let f be as in Proposition* 6.2. *Then*

$$\sum_{z\in\mathrm{lf}(T)\times\mathrm{edge}(T)}\sum_{v\in\mathrm{br}(T)}v\{v\}\big(f\big(\underline{\eta}^z(v)\big)-f\big(\underline{\eta}(v)\big)\big)$$

(6.10)

$$=\int\bigg(2\sum_{i,j=1}^3\eta_i(\delta_{ij}-\eta_j)\partial_{ij}^2 f(\underline{\eta})-\sum_{i=1}^3(1-3\eta_i)\partial_i f(\underline{\eta})\bigg)d\mu^{\otimes 3}+o(1),$$

and the $o(1)$-term tends to zero as $N\to\infty$ uniformly in the binary tree with N leaves.

PROOF. According to Remark 6.1, we may and do assume w.l.o.g. that f is symmetric so that $f(\underline{\eta}(\underline{u}))$ depends on $\underline{u}\in T^3$ only through $v=c(\underline{\eta}(\underline{u}))$. Fix $v\in\mathrm{br}(T)$. To make the calculation more readable, we abbreviate $\eta_i=\eta_i(v)$ (ordered decreasingly) in the following equations as long as v is fixed. Denote the three components of $T\setminus\{v\}$ by $S_i(v)$, $i=1,2,3$, ordered such that $\eta_i=\mu(S_i)$. For all $z=(u,e)\in\mathrm{lf}(T)\times\mathrm{edge}(T)$ with $u\in S_i(v)$ and $e\in S_j(v)$, a Taylor expansion yields

(6.11)
$$f\big(\underline{\eta}^z(v)\big)=\bigg(1-\epsilon(\partial_i-\partial_j)+\frac{1}{2}\epsilon^2(\partial_i-\partial_j)(\partial_i-\partial_j)\bigg)f(\underline{\eta})+o(\epsilon^2).$$

Using first this expansion and then Lemma 6.4, we obtain

$$A_v:=\sum_{z\in\mathrm{lf}(T)\times\mathrm{edge}(T)}\big(f\big(\underline{\eta}^z\big)-f(\underline{\eta})\big)$$

(6.12)
$$=\sum_{i,j=1,i\neq j}^3\frac{\eta_i}{\epsilon}\frac{\ell_j}{\delta}\epsilon\bigg(\bigg(\partial_j-\partial_i+\frac{\epsilon}{2}(\partial_{ii}+\partial_{jj}-2\partial_{ij})\bigg)f(\underline{\eta})+o(\epsilon)\bigg)$$

$$=\sum_{i,j=1,i\neq j}^3\eta_i\frac{\eta_j(1+3\delta)-\delta}{\delta}\bigg(\bigg(\partial_j-\partial_i+\frac{\epsilon}{2}(\partial_{ii}+\partial_{jj}-2\partial_{ij})\bigg)f(\underline{\eta})+o(\epsilon)\bigg).$$

As the highest order term is anti-symmetric in $i\neq j$, that is, $\sum_{i\neq j=1}^3\eta_i\eta_j(\partial_j-\partial_i)f(\underline{\eta})=0$, and $\frac{\epsilon}{\delta}=2+O(\epsilon)$, we obtain

$$A_v=-\sum_{i,j=1,i\neq j}^3\eta_i(\partial_j-\partial_i)f(\underline{\eta})+\frac{\epsilon}{\delta}\sum_{i,j=1,i\neq j}^3\eta_i\eta_j(\partial_{ii}-\partial_{ij})f(\underline{\eta})+o(1)$$

$$=-\sum_{i=1}^3(1-3\eta_i)\partial_i f(\underline{\eta})+2\sum_{i=1}^3\eta_i(1-\eta_i)\partial_{ii}f(\underline{\eta})$$

(6.13)

$$-2\sum_{i,j=1,i\neq j}^3\eta_i\eta_j\partial_{ij}f(\underline{\eta})+o(1)$$

$$=-\sum_{i=1}^3(1-3\eta_i)\partial_i f(\underline{\eta})+2\sum_{i,j=1}^3\eta_i(\delta_{ij}-\eta_j)\partial_{ij}f(\underline{\eta})+o(1),$$

and the claim follows by (weighted) summation over v and Fubini's theorem. □

Recall that, for $e\in\mathrm{edge}(T)$, we introduced $x_e,y_e\in\overline{T}\setminus T$, where x_e is "in the middle" of e and y_e is a leaf attached to x_e. Because μ is supported by T and $\eta(x_e)$ is ordered decreasingly, we always have $\eta_3(x_e)=0$ and $\eta_1(x_e)+\eta_2(x_2)=1$. The following lemma is easily obtained by associating a branch point to its three adjacent edges.

LEMMA 6.6 (Matching lemma).   *Let $g: [0, 1]^2 \to \mathbb{R}$ be symmetric. Then*

$$(6.14) \qquad \sum_{e \in \text{edge}(T)} g(\eta_1(x_e), \eta_2(x_e)) = \frac{1}{2} \sum_{v \in \text{br}(T)} \sum_{i=1}^{3} g(\eta_i(v), 1 - \eta_i(v)) + \frac{1}{2} N g(1 - \epsilon, \epsilon).$$

PROOF.   If $e \in \text{edge}(T)$ is adjacent to $v \in \text{br}(T)$, there is $i \in \{1, 2, 3\}$, $j \in \{1, 2\}$ with $\eta_j(x_e) = \eta_i(v)$ and $\eta_{3-j}(x_e) = 1 - \eta_i(v)$. The edge e is either adjacent to precisely two branch points, or it is an external edge, that is, adjacent to a leaf of T. In the latter case, we have $(\eta_1(x_e), \eta_2(x_e)) = (1 - \epsilon, \epsilon)$, and there are N external edges. Therefore,

$$(6.15) \qquad 2 \sum_{e \in \text{edge}(T)} g(\eta_1(x_e), \eta_2(x_e)) - N g(1 - \epsilon, \epsilon) = \sum_{v \in \text{br}(T)} \sum_{i=1}^{3} g(\eta_i(v), 1 - \eta_i(v))$$

as claimed.   □

PROOF OF PROPOSITION 6.2.   We assume w.l.o.g. that f is symmetric (see Remark 6.1).
Step 1. Recall that for $z = (u, e) \in \text{lf}(T) \times \text{edge}(T)$, ν_z and η^z are the branch point distribution and mass vector after the Aldous move z, respectively. In the first step, we calculate the effect of the branch point "created" by the Aldous move z due to the fact that $\nu_z(\{x_e\})$ might be nonzero, whereas $\nu(\{x_e\}) = 0$ for all $e \in \text{edge}(T)$. To this end, set

$$(6.16) \qquad C_{x_e} := \sum_{z \in \text{lf}(T) \times \text{edge}(T)} \nu_z \{x_e\} f(\underline{\eta}^z(x_e)).$$

Recall that we order the entries of $\underline{\eta}$ decreasingly, so that $\eta_1(x_e) + \eta_2(x_e) = 1$ and $\eta_3(x_e) = 0$. For $(x, y, z) \in \Delta_3$, let

$$(6.17) \qquad \begin{aligned} h_\epsilon(x, y, z) &:= \big(1 - \epsilon(2 + x\partial_1 + y\partial_2 - \partial_3)\big) f(x, y, z), \\ g_\epsilon(x, y) &:= 6xy \cdot h_\epsilon(x, y, 0). \end{aligned}$$

Then g_ϵ is a symmetric function. Let $e \in \text{edge}(T)$. For $z = (u, e') \in \text{lf}(T) \times \text{edge}(T)$, we have $\nu_z\{x_e\} \neq 0$ if and only if $e = e'$, and hence, using symmetry of f,

$$\begin{aligned} C_{x_e} &= \sum_{u \in \text{lf}(T)} \nu_{(u,e)}\{x_e\} f(\underline{\eta}^{(u,e)}(x_e)) \\ (6.18) \qquad &= \sum_{i=1}^{2} N \eta_i(x_e) \cdot 6\epsilon(\eta_i(x_e) - \epsilon)\eta_{3-i}(x_e) \cdot f(\eta_i(x_e) - \epsilon, \eta_{3-i}(x_e), \epsilon) \\ &= g_\epsilon(\eta_1(x_e), \eta_2(x_e)) + O(\epsilon^2), \end{aligned}$$

where we used, in the last equality, a first order Taylor expansion of f and the identity $\eta_1(x_e) + \eta_2(x_e) = 1$. For $v \in \text{br}(\overline{T}) \setminus T$, there is a unique edge $e \in \text{edge}(T)$ with $v = x_e$. Thus, using Lemma 6.6,

$$\begin{aligned} \sum_{v \in \text{br}(\overline{T}) \setminus T} C_v &= \sum_{e \in \text{edge}(T)} g_\epsilon(\eta_1(x_e), \eta_2(x_e)) + O(\epsilon) \\ (6.19) \qquad &= \frac{1}{2} \sum_{v \in \text{br}(T)} \sum_{i=1}^{3} g_\epsilon(\eta_i(v), 1 - \eta_i(v)) + \frac{1}{2} N g_\epsilon(1 - \epsilon, \epsilon) + O(\epsilon). \end{aligned}$$

Now we use that $\nu\{v\} = 6\eta_1(v)\eta_2(v)\eta_3(v)$ for $v \in \text{br}(T)$, and hence for any permutation $(i, j, k) \in S_3$, we have $g_\epsilon(\eta_i(v), 1 - \eta_i(v)) = \nu\{v\}(\frac{1}{\eta_j(v)} + \frac{1}{\eta_k(v)})h_\epsilon(\eta_i, 1 - \eta_i, 0)$, and obtain

$$(6.20) \qquad \sum_{v \in \text{br}(\overline{T}) \setminus T} C_v = \frac{1}{2} \sum_{v \in \text{br}(T)} \nu\{v\} \sum_{i,j=1, i \neq j}^{3} \frac{h_\epsilon(\eta_i(v), 1 - \eta_i(v), 0)}{\eta_j(v)} + 3f(1, 0, 0) + O(\epsilon).$$

Step 2. In the second step, we calculate the effect of the change in $v\{v\}$ for the "old" branch points. Fix $v \in \mathrm{br}(T)$. To make the calculation more readable, we abbreviate $\eta_i = \eta_i(v)$ as long as v is fixed. We use Lemma 6.4 in the first transformation, and a first-order Taylor expansion of f in the second:

$$
\begin{aligned}
B_v := & \sum_{z \in \mathrm{lf}(T) \times \mathrm{edge}(T)} (v_z\{v\} - v\{v\}) f(\underline{\eta}^z(v)) \\
= & \sum_{(i,j,k) \in S_3} \frac{\eta_i}{\epsilon} \frac{\eta_j(1+3\delta) - \delta}{\delta} \cdot v\{v\} \left(\frac{(\eta_i - \epsilon)(\eta_j + \epsilon)\eta_k}{\eta_i \eta_j \eta_k} - 1 \right) \\
& \times f(\eta_i - \epsilon, \eta_j + \epsilon, \eta_k) \\
= & \, v\{v\} \sum_{(i,j,k) \in S_3} \left(\frac{1}{\delta} \eta_j(1+3\delta) - 1 \right) \frac{\eta_i - \eta_j - \epsilon}{\eta_j} \\
& times \big(f(\underline{\eta}) + \epsilon(\partial_j - \partial_i) f(\underline{\eta}) + O(\epsilon^2) \big).
\end{aligned}
$$

(6.21)

Cancelling all terms which are anti-symmetric in (i, j), we obtain

$$
\begin{aligned}
B_v = & \, v\{v\} \sum_{(i,j,k) \in S_3} \left(\frac{\epsilon}{\delta}(\eta_i - \eta_j)(\partial_j - \partial_i) f(\underline{\eta}) - \left(\frac{\eta_i - \epsilon}{\eta_j} - 1 + \frac{\epsilon}{\delta} \right) f(\underline{\eta}) \right. \\
& \left. - \epsilon \frac{\eta_i}{\eta_j}(\partial_j - \partial_i) f(\underline{\eta}) + O(\epsilon) \right).
\end{aligned}
$$

(6.22)

Using $\epsilon/\delta = 2 + O(\epsilon)$, that $\sum_{(i,j,k) \in S_3} \frac{\eta_i}{\eta_j} = \sum_{(i,j,k) \in S_3} (\frac{1}{2\eta_j} - \frac{1}{2})$, and

$$
\begin{aligned}
\sum_{(i,j,k) \in S_3} \frac{\eta_i}{\eta_j}(\partial_j - \partial_i) f = & \sum_{j=1}^3 \left(\frac{1 - \eta_j}{\eta_j} \partial_j - \frac{\eta_i}{\eta_j} \partial_i - \frac{\eta_k}{\eta_j} \partial_k \right) f \\
= & \sum_{(i,j,k) \in S_3} \frac{1}{2\eta_j}(\partial_j - \eta_i \partial_i - \eta_k \partial_k) f + O(1),
\end{aligned}
$$

(6.23)

we continue

$$
\begin{aligned}
B_v = & \, v\{v\} \sum_{(i,j,k) \in S_3} \left(4(\eta_j - \eta_i) \partial_i f(\underline{\eta}) \right. \\
& \left. - \left(\frac{1}{2\eta_j} + \frac{1}{2} - \frac{\epsilon}{\eta_j} + \frac{\epsilon}{2\eta_j}(\partial_j - \eta_i \partial_i - \eta_k \partial_k) \right) f(\underline{\eta}) \right) \\
& + v\{v\} O(\epsilon) \\
= & \, 4v\{v\} \sum_{i=1}^3 (1 - 3\eta_i) \partial_i f(\underline{\eta}) - v\{v\} \sum_{(i,j,k) \in S_3} \frac{h_\epsilon(\eta_i, \eta_j, \eta_k)}{2\eta_k} \\
& - v\{v\} \big(3f(\underline{\eta}) + O(\epsilon) \big).
\end{aligned}
$$

(6.24)

Step 3. Because f is twice continuously differentiable, $\frac{1}{z}(2 + x\partial_1 + y\partial_2 - \partial_3)(f(x, y + z, 0) - f(x, y, z))$ is bounded and hence

$$
\begin{aligned}
\frac{1}{z} \big(h_\epsilon(x, y + z, 0) - h_\epsilon(x, y, z) \big) = & \frac{1}{z} \big(f(x, y + z, 0) - f(x, y, z) \big) + O(\epsilon) \\
= & \, \Theta_{3,2} f(x, y, z) + O(\epsilon).
\end{aligned}
$$

(6.25)

Therefore, using Fubini's theorem and combining (6.19) with (6.24) yields

$$\sum_{z\in\mathrm{lf}(T)\times\mathrm{edge}(T)}\sum_{v\in\mathrm{br}(\overline{T})}\left(\nu_z\{v\}-\nu\{v\}\right)f\left(\underline{\eta}^z(v)\right)$$

$$=\sum_{v\in\mathrm{br}(T)}B_v+\sum_{v\in\mathrm{br}(\overline{T})\setminus T}C_v$$

(6.26)

$$=\sum_{v\in\mathrm{br}(T)}\nu\{v\}\left(4\sum_{i=1}^{3}(1-3\eta_i)\partial_i f\left(\underline{\eta}(v)\right)+\frac{1}{2}\sum_{i,j=1,i\neq j}^{3}\Theta_{i,j}f\left(\underline{\eta}(v)\right)+3f(1,0,0)\right.$$

$$\left.-3f\left(\underline{\eta}(v)\right)\right)+O(\epsilon).$$

Together with Lemma 6.5 (and using symmetry of f), we have obtained for $\chi\in\mathbb{T}_2^N$

$$\Omega_N\Phi^f(\chi)=\sum_{z\in\mathrm{lf}(T)\times\mathrm{edge}(T)}\sum_{v\in\mathrm{br}(\overline{T})}\left(\nu_z\{v\}f\left(\underline{\eta}^z(v)\right)-\nu\{v\}f\left(\underline{\eta}(v)\right)\right)$$

(6.27)

$$=\Omega_{\mathrm{Ald}}\Phi^f(\chi)+o(1),$$

which shows the claim of Proposition 6.2. □

REFERENCES

[1] ALDOUS, D. (1993). The continuum random tree. III. *Ann. Probab.* **21** 248–289. MR1207226

[2] ALDOUS, D. (1994). Recursive self-similarity for random trees, random triangulations and Brownian excursion. *Ann. Probab.* **22** 527–545. MR1288122

[3] ALDOUS, D. J. (2000). Mixing time for a Markov chain on cladograms. *Combin. Probab. Comput.* **9** 191–204. MR1774749 https://doi.org/10.1017/S096354830000417X

[4] ALLEN, B. L. and STEEL, M. (2001). Subtree transfer operations and their induced metrics on evolutionary trees. *Ann. Comb.* **5** 1–15. MR1841949 https://doi.org/10.1007/s00026-001-8006-8

[5] ATHREYA, S., LÖHR, W. and WINTER, A. (2017). Invariance principle for variable speed random walks on trees. *Ann. Probab.* **45** 625–667. MR3630284 https://doi.org/10.1214/15-AOP1071

[6] BASTERT, O., ROCKMORE, D., STADLER, P. F. and TINHOFER, G. (2002). Landscapes on spaces of trees. *Appl. Math. Comput.* **131** 439–459. MR1920237 https://doi.org/10.1016/S0096-3003(01)00164-3

[7] BILLERA, L. J., HOLMES, S. P. and VOGTMANN, K. (2001). Geometry of the space of phylogenetic trees. *Adv. in Appl. Math.* **27** 733–767. MR1867931 https://doi.org/10.1006/aama.2001.0759

[8] DEPPERSCHMIDT, A., GREVEN, A. and PFAFFELHUBER, P. (2012). Tree-valued Fleming–Viot dynamics with mutation and selection. *Ann. Appl. Probab.* **22** 2560–2615. MR3024977 https://doi.org/10.1214/11-AAP831

[9] ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes: Characterization and Convergence. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics.* Wiley, New York. MR0838085 https://doi.org/10.1002/9780470316658

[10] EVANS, S. N., GRÜBEL, R. and WAKOLBINGER, A. (2017). Doob–Martin boundary of Rémy's tree growth chain. *Ann. Probab.* **45** 225–277. MR3601650 https://doi.org/10.1214/16-AOP1112

[11] EVANS, S. N. and WINTER, A. (2006). Subtree prune and regraft: A reversible real tree-valued Markov process. *Ann. Probab.* **34** 918–961. MR2243874 https://doi.org/10.1214/009117906000000034

[12] FELSENSTEIN, J. (2003). *Inferring Phylogenies.* Sinauer, Sunderland.

[13] FORMAN, N. (2018). Mass structure of weighted real trees. Available at arXiv:1801.02700v1.

[14] FORMAN, N., PAL, S., RIZZOLO, D. and WINKEL, M. (2019). Diffusions on a space of interval partitions with Poisson–Dirichlet stationary distributions. Available at arXiv:1609.06707v2.

[15] FORMAN, N., PAL, S., RIZZOLO, D. and WINKEL, M. (2018). Projections of the Aldous chain on binary trees: Intertwining and consistency. Available at arXiv:1802.00862v1.

[16] FORMAN, N., PAL, S., RIZZOLO, D. and WINKEL, M. (2018). Interval partition evolutions with emigration related to the Aldous diffusion. Available at arXiv:1804.01205v1.

[17] FORMAN, N., PAL, S., RIZZOLO, D. and WINKEL, M. (2018). Aldous diffusion I: A projective system of continuum k-tree evolutions. Available at arXiv:1809.07756.

[18] GREVEN, A., PFAFFELHUBER, P. and WINTER, A. (2009). Convergence in distribution of random metric measure spaces (Λ-coalescent measure trees). *Probab. Theory Related Fields* **145** 285–322. MR2520129 https://doi.org/10.1007/s00440-008-0169-3

[19] GREVEN, A., PFAFFELHUBER, P. and WINTER, A. (2013). Tree-valued resampling dynamics martingale problems and applications. *Probab. Theory Related Fields* **155** 789–838. MR3034793 https://doi.org/10.1007/s00440-012-0413-8

[20] KALLENBERG, O. (2002). *Foundations of Modern Probability*, 2nd ed. *Probability and Its Applications* (*New York*). Springer, New York. MR1876169 https://doi.org/10.1007/978-1-4757-4015-8

[21] KLIEM, S. and LÖHR, W. (2015). Existence of mark functions in marked metric measure spaces. *Electron. J. Probab.* **20** no. 73, 24. MR3371432 https://doi.org/10.1214/EJP.v20-3969

[22] LÖHR, W. (2013). Equivalence of Gromov–Prohorov- and Gromov's $\square_\lambda$-metric on the space of metric measure spaces. *Electron. Commun. Probab.* **18** no. 17, 10. MR3037215 https://doi.org/10.1214/ecp.v18-2268

[23] LÖHR, W., VOISIN, G. and WINTER, A. (2015). Convergence of bi-measure $\mathbb{R}$-trees and the pruning process. *Ann. Inst. Henri Poincaré Probab. Stat.* **51** 1342–1368. MR3414450 https://doi.org/10.1214/14-AIHP628

[24] LÖHR, W. and WINTER, A. (2018). Spaces of algebraic measure trees and triangulations of the circle. Available at arXiv:1811.11734v2.

[25] NUSSBAUMER, J. (2018). The Kingman coalescent: Old and new aspects. Master thesis. Available at http://www-stud.uni-due.de/~snjonuss/master_thesis_kingman_coalescent.pdf.

[26] PAL, S. (2011). On the Aldous diffusion on continuum tree, I. Available at arXiv:1104.4186v1.

[27] SCHWEINSBERG, J. (2002). An $O(n^2)$ bound for the relaxation time of a Markov chain on cladograms. *Random Structures Algorithms* **20** 59–70. MR1871950 https://doi.org/10.1002/rsa.1029