Probab. Theory Relat. Fields (2013) 155:789–838
DOI 10.1007/s00440-012-0413-8

Tree-valued resampling dynamics Martingale problems and applications

Andreas Greven · Peter Pfaffelhuber · Anita Winter

Received: 20 May 2010 / Revised: 14 December 2011 / Published online: 3 February 2012

Abstract The measure-valued Fleming–Viot process is a diffusion which models the evolution of allele frequencies in a multi-type population. In the neutral setting the Kingman coalescent is known to generate the genealogies of the "individuals" in the population at a fixed time. The goal of the present paper is to replace this static point of view on the genealogies by an analysis of the evolution of genealogies. We encode the genealogy of the population as an (isometry class of an) ultra-metric space which is equipped with a probability measure. The space of ultra-metric measure spaces together with the Gromov-weak topology serves as state space for tree-valued processes. We use well-posed martingale problems to construct the tree-valued resampling dynamics of the evolving genealogies for both the finite population Moran model and the infinite population Fleming–Viot diffusion. We show that sufficient

All authors were supported in part by the DFG-Forschergruppe 498 through grant GR 876/13-1,2,3.
P. Pfaffelhuber was supported in part by the BMBF, Germany, through FRISYS (Freiburg Initiative for Systems biology), Kennzeichen 0313921.
A. Winter was supported in part at the Technion by a fellowship from the Aly Kaufman Foundation.

A. Greven
Department Mathematik, University of Erlangen, Cauerstrasse 11,
91058 Erlangen, Germany
e-mail: greven@mi.uni-erlangen.de

P. Pfaffelhuber (✉)
Abteilung für Mathematische Stochastik, Albert-Ludwigs University of Freiburg,
Eckerstrasse 1, 79104 Freiburg, Germany
e-mail: p.p@stochastik.uni-freiburg.de

A. Winter
Fakultät für Mathematik, Universität Duisburg-Essen, Universitätsstr. 2,
45141 Essen, Germany
e-mail: anita.winter@uni-due.de

information about any ultra-metric measure space is contained in the distribution of the vector of subtree lengths obtained by sequentially sampled “individuals”. We give explicit formulas for the evolution of the Laplace transform of the distribution of finite subtrees under the tree-valued Fleming–Viot dynamics.

Keywords Tree-valued Markov process · Fleming–Viot process · Moran model · Genealogical tree · Martingale problem · Duality · (ultra-)Metric measure space · Gromov-weak topology

Mathematics Subject Classification (2000) Primary: 60K35 · 60J25; Secondary: 60J70 · 92D10

1 Introduction

The evolution of a population is commonly modeled using branching or resampling dynamics. In both cases the analysis of the genealogical relationships of individuals leads to a deeper understanding of the underlying dynamics and is crucial in applications in population genetics. An observation which is fundamental for the present paper is that genealogical relationships between individuals change as the population evolves. We here want to construct and study the evolution of the *genealogical structure* for the neutral Fleming–Viot process which arises as a large population limit of various finite resampling models [5,8,17,22,27,28].

A basic finite resampling model is the Moran model, which can be described as follows: Each pair of individuals, taken from a finite population of fixed size, *resamples* at constant rate. Resampling means that one individual is chosen at random from the pair, the pair dies and is replaced by two new individuals which are both offspring of the chosen individual.

In resampling models genealogical trees can be generated by coalescent processes. The equilibrium genealogy of the Fleming–Viot diffusion, for example, is generated by the Kingman coalescent [2,23,29,34]. More general resampling dynamics which allow for an infinite offspring variance are studied in [4]. Their genealogical trees are described by Λ-coalescents [30,39]. Genealogical trees are also considered for branching models which allow for a varying population size. Prominent examples are the Kallenberg tree [33], the Yule tree [19], the Brownian continuum random tree [1] and the Brownian snake [37]. More general branching mechanisms lead e.g. to Lévy trees [12], which are the infinite variance offspring distribution counterpart of the Brownian continuum random trees, and trees arising in catalytic branching systems [31].

Coalescent trees describe the genealogy of a population at a fixed time and give therefore a static picture only. The main goal of the present paper is to give with the *tree-valued Fleming–Viot dynamics* a dynamic picture which describes the evolution of genealogies. Evolving genealogies in exchangeable population models have already been described by *look-down processes* [3,9–11]; see also Remark 2.20. For neutral evolution, look-down processes contain—though in an implicit way—all information about the genealogies. The depth of the tree [6,21,40] and the total tree length [41] are examples of functionals of a genealogy which are constructed and studied via the look-down construction. The crucial point in the construction of look-down processes is the use of labels as coordinates. This information is often not needed and constraints the construction of tree-valued processes in selective (unequal chances of producing offspring) and spatial (only pairs in the same location may resample) settings.

A first approach in the direction of a coordinate-free description has already been established for spatially structured populations via *historical processes* [14,29]. Here, however, the coding of the genealogical relationships requires that different "offspring" immediately follow different spatial paths, almost surely. Only then the genealogy can be read off from the spatial paths of the "individuals". Therefore, in non-spatial situations or if space is discrete, additional structure would be required for an investigation of genealogies via historical processes.

A different and more canonical coding of trees is therefore needed. In this paper we rely on the fact that genealogical distances between individuals define a metric. To take the individuals' contribution to the population into account we equip the resulting metric space with the (weak limit of the) empirical distribution of the individuals. We then follow the theory of *metric measure spaces* equipped with the *Gromov-weak topology* as developed in [30]. We show weak convergence of tree-valued Moran models and construct the limiting tree-valued Fleming–Viot dynamics. Such weak convergence results are best treated by using well-posed martingale problems, which allow—in contrast to other techniques such as Dirichlet forms—for statements concerning convergence of infinitesimal characteristics. In order to define these characteristics, we require a suitably large class of continuous functions which are easy to manipulate. For tree-valued processes such an approach is novel. We make use of general theory in order to establish well-posedness of the limiting martingale problem (Theorem 1), weak convergence of tree-valued Moran models (Theorem 2) and the long-time behavior (Theorem 3).

Another useful consequence of a well-posed martingale problem is that it allows to study the evolution of continuous functionals of these processes and to characterize the functionals which are strong Markov processes. Of particular importance is the vector of tree lengths for subsequently sampled "individuals". An important result (Theorem 4) is that the resulting *subtree length distribution* characterizes the ultrametric measure tree uniquely. From a theoretical point of view this can be considered as a generalization of the moment problem for bounded real-valued random variables to metric measure spaces. It is also of interest in statistical applications since it states that all sufficient information about genealogies is contained in the lengths of subtrees spanned by a finite sample. Under the Fleming–Viot dynamics we construct the evolution of the tree length distribution via a well-posed martingale problem (Theorem 5). Moreover, we derive explicit formulas for the evolution of the Laplace transform of finite subtrees.

Markov dynamics with values in the space of continuum trees have been constructed only recently. Examples include excursion path-valued Markov processes with continuous sample paths—which can therefore be thought of as tree-valued diffusions—as investigated in [45–47], and dynamics working with real-trees, for example, the so-called *root growth with re-grafting* [20], the so-called *subtree prune and re-graft move* [25] and the limiting *random mapping* [18]. The present construction is extended to Fleming–Viot processes with selection in [7].

2 Main results (Theorems 1, 2 and 3)

In this section we state our main results. In Sect. 2.1 we recall concepts and terminology used to define the state space which consists of (ultra-)metric measure spaces equipped with the Gromov-weak topology. In Sect. 2.2 we state the tree-valued Fleming–Viot martingale problem and its well-posedness (Theorem 1), and present the approximation by tree-valued Moran dynamics in Sect. 2.3 (Theorem 2). In Sect. 2.4 we identify a unique equilibrium and state that it will be approached as time tends to infinity (Theorem 3).

2.1 State space: metric measure spaces

To define the state space we consider trees as metric spaces. Moreover, to allow for a topology which discards atypical points in the tree, we will equip these metric spaces with a probability measure on the leaves. (Compare also with Remark 2.15). We then equip the space of metric measure spaces with the Gromov-weak topology which combines the concept of weak convergence of probability measures in a fixed metric space with Gromov's idea of comparing different metric spaces. In [30] topological aspects of the space of metric measure spaces equipped with the Gromov-weak topology are investigated. In this subsection we recall basic facts and notation.

As usual, given a topological space $(X, \mathcal{O})$ we denote by $\mathcal{M}_1(X)$ the space of all probability measures defined on the Borel-σ-algebra of X, and by $\Rightarrow$ weak convergence in $\mathcal{M}_1(X)$. Recall that the support $\mathrm{supp}(\mu)$ of $\mu \in \mathcal{M}_1(X)$ is the smallest closed set $X_0 \subseteq X$ such that $\mu(X_0) = 1$. The push forward of μ under a measurable map φ from X into another topological space Z is the probability measure $\varphi_*\mu \in \mathcal{M}_1(Z)$ defined by

$$\varphi_*\mu(A) := \mu\big(\varphi^{-1}(A)\big), \tag{2.1}$$

for all Borel subsets $A \subseteq Z$. We denote by $\mathcal{B}(X)$ and $\mathcal{C}_b(X)$ the bounded real-valued functions on X which are measurable and continuous, respectively.

A *metric measure space* is a triple (X, r, μ) where (X, r) is a metric space equipped with $\mu \in \mathcal{M}_1(X)$ such that $(\mathrm{supp}(\mu), r)$ is complete and separable. Two metric measure spaces (X, r, μ) and (X', r', μ') are *measure-preserving isometric* or *equivalent* if there exists an isometry φ between $\mathrm{supp}(\mu)$ and $\mathrm{supp}(\mu')$ such that $\mu' = \varphi_*\mu$. It is clear that the property of being measure-preserving isometric is an equivalence relation. We write $\overline{(X, r, \mu)}$ for the equivalence class of a metric measure space (X, r, μ). Define the set of (equivalence classes of) metric measure spaces

$$\mathbb{M} := \big\{\chi = \overline{(X, r, \mu)} : \ (X, r, \mu) \text{ metric measure space}\big\}. \tag{2.2}$$

If (X, r, μ) is such that r is only a pseudo-metric on X, (i.e. $r(x, y) = 0$ is possible for $x \neq y$) we can still define its measure-preserving isometry class. Since this class contains also metric measure spaces, there is a bijection between the set of pseudo-metric measure spaces and the set of metric measure spaces and we use both notions interchangeably.

For a metric space (X, r) we define by

$$R^{(X,r)}: \begin{cases} X^{\mathbb{N}} \to \mathbb{R}_+^{\binom{\mathbb{N}}{2}} \\ \big((x_i)_{i\geq 1}\big) \mapsto \big(r(x_i, x_j)\big)_{1\leq i<j} \end{cases} \tag{2.3}$$

the map which sends a sequence of points in X to its (infinite) distance matrix, and denote, for a metric measure space (X, r, μ), the *distance matrix distribution* of (X, r, μ) by

$$\nu^{(X,r,\mu)} := \big(R^{(X,r)}\big)_* \mu^{\otimes\mathbb{N}} \in \mathcal{M}_1\left(\mathbb{R}_+^{\binom{\mathbb{N}}{2}}\right). \tag{2.4}$$

Obviously, $\nu^{(X,r,\mu)}$ depends on (X, r, μ) only through its measure-preserving isometry class $\mathfrak{x} = \overline{(X, r, \mu)}$. We can therefore define:

Definition 2.1 (*Distance matrix distribution*) The distance matrix distribution $\nu^{\mathfrak{x}}$ of $\mathfrak{x} \in \mathbb{M}$ is the distance matrix distribution $\nu^{(X,r,\mu)}$ of an arbitrary representative $(X, r, \mu) \in \mathfrak{x}$.

By Gromov's reconstruction theorem metric measure spaces are uniquely determined by their distance matrix distribution (see Section $3\frac{1}{2}.5$ in [32] and Proposition 2.6 in [30]). We therefore base our notion of convergence in $\mathbb{M}$ on the convergence of distance matrix distributions. In [30] we introduced the *Gromov-weak topology* in which a sequence $(\mathfrak{x}_n)_{n\in\mathbb{N}}$ converges to $\mathfrak{x}$ if and only if

$$\nu^{\mathfrak{x}_n} \underset{n\to\infty}{\Longrightarrow} \nu^{\mathfrak{x}} \tag{2.5}$$

in the weak topology on $\mathcal{M}_1(\mathbb{R}_+^{\binom{\mathbb{N}}{2}})$ (and, as usual, $\mathbb{R}_+^{\binom{\mathbb{N}}{2}}$ equipped with the product topology); compare Theorem 5 of [30]. Notice that possible limits $\nu \in \mathcal{M}_1(\mathbb{R}_+^{\binom{\mathbb{N}}{2}})$ are not necessarily of the form $\nu = \nu^{\mathfrak{x}}$ for some $\mathfrak{x} \in \mathbb{M}$. Although $\{\nu^{\mathfrak{x}};\ \mathfrak{x} \in \mathbb{M}\}$ is not closed, we could show that $\mathbb{M}$ equipped with the Gromov-weak topology is Polish (compare, Theorem 1 in [30]).

Several sub-spaces of $\mathbb{M}$ are of special interest throughout the paper. Above all, these are the *ultra-metric* and *compact* metric measure spaces.

(The equivalence class of) a metric measure space (X, r, μ) is called *ultra-metric* iff

$$r(u, w) \leq r(u, v) \vee r(v, w), \tag{2.6}$$

for μ-almost all $u, v, w \in X$. Define

$$\mathbb{U} := \big\{\mathfrak{u} \in \mathbb{M} :\ \mathfrak{u} \text{ is ultra-metric}\big\}. \tag{2.7}$$

Remark 2.2 (Ultra-metric spaces are trees) Notice that there is a close connection between ultra-metric spaces and $\mathbb{R}$-trees, i.e., complete path-connected metric spaces (X, r_X) which satisfy the four-point condition

$$\begin{aligned} &r_X(x_1, x_2) + r_X(x_3, x_4) \\ &\quad \le \max\left\{r_X(x_1, x_3) + r_X(x_2, x_4), r_X(x_1, x_4) + r_X(x_2, x_3)\right\}, \end{aligned} \tag{2.8}$$

for all $x_1, x_2, x_3, x_4 \in X$ (see, for example, [13,15,43]). On the one hand, every complete ultra-metric space (U, r_U) *spans* a path-connected complete metric space (X, r_X) which satisfies the *four point condition*, such that (U, r_U) is isometric to the set of leaves $X\backslash X^o$. On the other hand, given an $\mathbb{R}$-tree (X, r_X) and a distinguished point $\rho_X \in X$ which is often referred to as the *root* of (X, r_X), the level sets $X^t := \{x \in X : r(\rho_X, x) = t\}$, for $t \geq 0$, form ultra-metric sub-spaces of (X, r_X). For more details, see [24, Theorem 3.38].

Because of this connection between ultra-metric spaces and real trees, ultra-metric spaces are often (especially in phylogenetic analysis) referred to as *ultra-metric trees*. □

The next lemma implies that $\mathbb{U}$ equipped with the Gromov-weak topology is again Polish.

Lemma 2.3 *The sub-space $\mathbb{U} \subset \mathbb{M}$ is closed.*

Proof Let $(\mathfrak{u}_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{U}$ and $\mathfrak{x} \in \mathbb{M}$ such that $\mathfrak{u}_n \to \mathfrak{x}$ in the Gromov-weak topology, as $n \to \infty$. Equivalently, by (2.5), $\nu^{\mathfrak{u}_n} \Rightarrow \nu^{\mathfrak{x}}$ in the weak topology on $\mathcal{M}_1(\mathbb{R}_+^{\binom{\mathbb{N}}{2}})$, as $n \to \infty$. Consider the open set

$$A := \left\{(r_{i,j})_{1\le i<j} : r_{1,2} > r_{23} \vee r_{1,3} \text{ or } r_{2,3} > r_{1,2} \vee r_{1,3} \text{ or } r_{1,3} > r_{1,2} \vee r_{2,3}\right\}. \tag{2.9}$$

By the Portmanteau Theorem, $\nu^{\mathfrak{x}}(A) \le \liminf_{n\to\infty} \nu^{\mathfrak{u}_n}(A) = 0$. Thus, (2.6) holds for $\mu^{\otimes 3}$-all triples $(u, v, w) \in X^3$. In other words, $\mathfrak{x}$ is ultra-metric. □

(The equivalence class of) a metric measure space (X, r, μ) is called *compact* if and only if the metric space $(\text{supp}(\mu), r)$ is compact. Define

$$\mathbb{M}_c := \left\{\mathfrak{x} \in \mathbb{M} : \mathfrak{x} \text{ is compact}\right\}. \tag{2.10}$$

Moreover, we set

$$\mathbb{U}_c := \mathbb{U} \cap \mathbb{M}_c. \tag{2.11}$$

Remark 2.4 ($\mathbb{M}_c$ is not a closed subset of $\mathbb{M}$)

(i) If $\mathfrak{x} = \overline{(X, r, \mu)}$ is a *finite* metric measure space, i.e, $\#\text{supp}(\mu) < \infty$, then $\mathfrak{x} \in \mathbb{M}_c$.

(ii) Since elements of $\mathbb{M}$ can be approximated by a sequence of finite metric measure spaces (see the proof of Proposition 5.3 in [30]), the sub-space $\mathbb{M}_c$ is not closed. A similar argument shows that $\mathbb{U}_c$ is not closed.

(iii) In order to establish convergence within the space of compact metric measure spaces, we provide a *relative compactness criterion* in $\mathbb{M}_c$ in Proposition 6.2.

2.2 The martingale problem (Theorem 1)

In this subsection we define the tree-valued Fleming–Viot dynamics as the solution of a well-posed martingale problem. We start by recalling the terminology. All proofs are given in Sect. 8.

Definition 2.5 (*Martingale problem*) Let $(E, \mathcal{O})$ be a Polish space, $\mathbf{P}_0 \in \mathcal{M}_1(E)$, $\mathcal{F}$ a subspace of the space $\mathcal{B}(E)$ of bounded measurable functions on E and Ω a linear operator on $\mathcal{B}(E)$ with domain $\mathcal{F}$.

The law $\mathbf{P}$ of an E-valued stochastic process $X = (X_t)_{t\geq 0}$ is called a solution of the $(\mathbf{P}_0, \Omega, \mathcal{F})$-martingale problem if X_0 has distribution $\mathbf{P}_0$, X has paths in the space $\mathcal{D}_E([0,\infty))$ of E-valued càdlàg functions, almost surely (where $\mathcal{D}_E([0,\infty)$ is equipped with the Skorohod topology) and for all $F \in \mathcal{F}$,

$$\left(F(X_t) - \int_0^t \mathrm{d}s\, \Omega F(X_s) \right)_{t\geq 0} \tag{2.12}$$

is a $\mathbf{P}$-martingale with respect to the canonical filtration.

Moreover, the $(\mathbf{P}_0, \Omega, \mathcal{F})$-martingale problem is said to be well-posed if there is a unique solution $\mathbf{P}$.

Recall that the classical measure-valued Fleming–Viot process $\zeta = (\zeta_t)_{t\geq 0}$ is a probability measure-valued diffusion process, which describes the evolution of allelic frequencies; see e.g. [5,22]. In particular, for a fixed time t, the state $\zeta_t \in \mathcal{M}_1(K)$ records the current distributions of allelic types on some (Polish) type space K. This process is defined as the unique solution of the martingale problem corresponding to the following operator $\widehat{\Omega}^{\uparrow}$ (see [17]): for functions $\widehat{\Phi} : \mathcal{M}_1(K) \to \mathbb{R}$ of the form

$$\widehat{\Phi}(\zeta) = \left\langle \zeta^{\otimes\mathbb{N}}, \widehat{\phi} \right\rangle := \int_{K^{\mathbb{N}}} \zeta^{\otimes\mathbb{N}}(\mathrm{d}\underline{u})\, \widehat{\phi}\big(\underline{u}\big) \tag{2.13}$$

with $\underline{u} = (u_1, u_2, \ldots) \in K^{\mathbb{N}}$ and $\widehat{\phi} \in \mathcal{C}_b(K^{\mathbb{N}})$ depending only on finitely many coordinates, set

$$\widehat{\Omega}^{\uparrow}\widehat{\Phi}(\zeta) = \frac{\gamma}{2} \sum_{k,l\geq 1} \Big(\langle \zeta^{\otimes\mathbb{N}}, \widehat{\phi} \circ \widehat{\theta}_{k,l} \rangle - \langle \zeta^{\otimes\mathbb{N}}, \widehat{\phi} \rangle \Big) \tag{2.14}$$

where the *replacement operator* $\widehat{\theta}_{k,l} : K^{\mathbb{N}} \to K^{\mathbb{N}}$ is the map which replaces the l^{th} component of an infinite sequence of types by the kth:

$$\widehat{\theta}_{k,l}(u_1, u_2, \ldots, u_{l-1}, u_l, u_{l+1}, \ldots) := (u_1, u_2, \ldots, u_{l-1}, u_k, u_{l+1}, \ldots). \tag{2.15}$$

Here and in the following $\gamma \in (0, \infty)$ is referred to as the *resampling rate*.

In order to state the martingale problem for the tree-valued Fleming–Viot dynamics we need the notion of *polynomials* on $\mathbb{M}$.

Definition 2.6 (*Polynomials*) A function $\Phi : \mathbb{M} \to \mathbb{R}$ is called a *polynomial* if there exists a bounded, measurable *test function* $\phi : \mathbb{R}_+^{\binom{\mathbb{N}}{2}} \to \mathbb{R}$, depending only on finitely many variables such that

$$\Phi\big(\chi\big) = \big\langle \nu^{\chi}, \phi\big\rangle := \int_{\mathbb{R}_+^{\binom{\mathbb{N}}{2}}} \nu^{\chi}(\mathrm{d}\underline{\underline{r}})\, \phi\big(\underline{\underline{r}}\big), \tag{2.16}$$

where $\underline{\underline{r}} := (r_{i,j})_{1 \le i < j}$.

Remark 2.7 (Properties of polynomials)

(i) Let Φ and ϕ be as in Definition 2.6. If $\chi = \overline{(X, r, \mu)}$, then

$$\Phi\big(\chi\big) = \int_{X^{\mathbb{N}}} \mu^{\otimes \mathbb{N}}(\mathrm{d}(x_1, x_2, \ldots))\, \phi\big((r(x_i, x_j))_{1 \le i < j}\big), \tag{2.17}$$

where $\mu^{\otimes \mathbb{N}}$ is the $\mathbb{N}$-fold product measure of μ.

(ii) If $n \in \mathbb{N}$ is the minimal number such that there exists $\phi \in \mathcal{B}(\mathbb{R}_+^{\binom{\mathbb{N}}{2}})$, depending only on $(r_{i,j})_{1 \le i < j \le n}$ such that (2.16) holds, n is referred to as *degree* and ϕ as a *minimal test function* of Φ. We write $\Phi = \Phi^{n,\phi}$.

(iii) For $m \in \mathbb{N}$, let Σ_m be the set of permutations of $\mathbb{N}$ which leave $m+1, m+2, \ldots$ fixed. For $\sigma \in \Sigma_\infty := \bigcup_{m \in \mathbb{N}} \Sigma_m$, define

$$\widetilde{\sigma}\big((r_{i,j})_{1 \le i < j}\big) := \big(r_{\sigma(i) \wedge \sigma(j), \sigma(i) \vee \sigma(j)}\big)_{1 \le i < j}. \tag{2.18}$$

The *symmetrization* of ϕ is given by

$$\bar{\phi} = \frac{1}{n!} \sum_{\sigma \in \Sigma_n} \phi \circ \widetilde{\sigma}. \tag{2.19}$$

By symmetry of ν^{χ}, $\langle \nu^{\chi}, \phi \rangle = \langle \nu^{\chi}, \bar{\phi} \rangle$, or equivalently, $\Phi^{n,\phi} = \Phi^{n,\bar{\phi}}$. □

Recall from Sect. 2.1 the space $\mathcal{B}(\mathbb{R}_+^{\binom{\mathbb{N}}{2}})$ of bounded measurable real-valued functions on $\mathbb{R}_+^{\binom{\mathbb{N}}{2}}$. An element $\phi \in \mathcal{B}(\mathbb{R}_+^{\binom{\mathbb{N}}{2}})$ is said to be differentiable if for all $1 \le i < j$ the partial derivatives $\frac{\partial \phi}{\partial r_{i,j}}$ exist and if $\sum_{1 \le i < j} |\frac{\partial \phi}{\partial r_{i,j}}| < \infty$. In this case we put

$$\langle \nabla \phi, \underline{\underline{2}} \rangle := 2 \sum_{1 \le i < j} \frac{\partial \phi}{\partial r_{i,j}} = \sum_{\substack{1 \le i, j \\ i \ne j}} \frac{\partial \phi}{\partial r_{i \wedge j, i \vee j}}. \tag{2.20}$$

Denote by $\mathcal{C}_b^1(\mathbb{R}_+^{\binom{\mathbb{N}}{2}})$ the space of all bounded and continuously differentiable real-valued functions ϕ on $\mathbb{R}_+^{\binom{\mathbb{N}}{2}}$ with bounded derivatives. The function spaces we use in the sequel are the space of polynomials

$$\Pi := \big\{ \Phi^{n,\phi} \text{ as in Remark 2.7(ii)} : n \in \mathbb{N}, \phi \in \mathcal{B}(\mathbb{R}_+^{\binom{\mathbb{N}}{2}}) \big\}, \tag{2.21}$$

and its sub-spaces

$$\Pi^0 := \big\{ \Phi^{n,\phi} \in \Pi : n \in \mathbb{N}, \phi \in \mathcal{C}_b(\mathbb{R}_+^{\binom{\mathbb{N}}{2}}) \big\}, \tag{2.22}$$

and

$$\Pi^1 := \big\{ \Phi^{n,\phi} \in \Pi : n \in \mathbb{N}, \phi \in \mathcal{C}_b^1(\mathbb{R}_+^{\binom{\mathbb{N}}{2}}) \big\}. \tag{2.23}$$

Remark 2.8 (Polynomials form an algebra that separates points)

(i) Observe that Π, Π^0 and Π^1 are algebras of functions. Specifically, given $\Phi^{n,\phi}$, and $\Phi^{m,\psi} \in \Pi$,

$$\Phi^{n,\phi} \cdot \Psi^{m,\psi} = \Phi^{n+m,(\phi,\psi)_n} = \Phi^{n+m,(\psi,\phi)_m} \tag{2.24}$$

where for $\phi, \psi \in \mathcal{B}(\mathbb{R}_+^{\binom{\mathbb{N}}{2}})$ and $\ell \in \mathbb{N}$,

$$(\phi, \psi)_\ell(\underline{\underline{r}}) := \phi(\underline{\underline{r}}) \cdot \psi(\tau_\ell \underline{\underline{r}}), \tag{2.25}$$

with $\tau_\ell\big((r_{i,j})_{1 \le i < j}\big) = (r_{\ell+i,\ell+j})_{1 \le i < j}$.

(ii) By Proposition 2.6 in [30], Π and Π^0 separate points in $\mathbb{M}$. Since $\mathcal{C}_b^1(\mathbb{R}_+^{\binom{\mathbb{N}}{2}})$ is dense in the topology of point-wise convergence in $\mathcal{C}_b(\mathbb{R}_+^{\binom{\mathbb{N}}{2}})$, Π^1 separates points as well. □

Remark 2.9 (The Gromov-weak topology) Let $\mathcal{X}, \mathcal{X}_1, \mathcal{X}_2, \ldots \in \mathbb{M}$. Recall from (2.5) that $\mathcal{X}_n \to \mathcal{X}$, as $n \to \infty$, in the Gromov-weak topology iff $\nu^{\mathcal{X}_n} \Longrightarrow \nu^{\mathcal{X}}$, as $n \to \infty$. Equivalently, $\Phi(\mathcal{X}_n) \to \Phi(\mathcal{X})$, as $n \to \infty$, for all $\Phi \in \Pi^0$ (see Theorem 5 in [30]). Notice that $\mathcal{X}_n \to \mathcal{X}$, as $n \to \infty$, if we restrict to Π^1 or to the set $\{\Phi^{n,\bar{\phi}} : \Phi^{n,\phi} \in \Pi\}$ of symmetric test functions. (Compare with Remark 2.7(iii)).

To lift the measure-valued Fleming–Viot process to the level of trees and thereby construct the tree-valued Fleming–Viot dynamics, we consider the martingale problem associated with the operator $\Omega^\uparrow$ on Π with domain $\mathcal{D}(\Omega^\uparrow) = \Pi^1$. To define $\Omega^\uparrow$ we let for $\Phi = \Phi^{n,\phi} \in \Pi^1$,

$$\Omega^\uparrow \Phi := \Omega^{\uparrow,\text{grow}} \Phi + \Omega^{\uparrow,\text{res}} \Phi. \tag{2.26}$$

The *growth operator* $\Omega^{\uparrow,\text{grow}}$ reflects the fact that the population gets older and therefore the genealogical distances grow at speed 2 as time goes on. We therefore put

$$\Omega^{\uparrow,\text{grow}}\Phi\big(u\big) := \big\langle \nu^u, \langle \nabla\phi, \underline{\underline{2}}\rangle\big\rangle. \tag{2.27}$$

For the *resampling operator* let

$$\Omega^{\uparrow,\text{res}}\Phi\big(u\big) := \frac{\gamma}{2}\sum_{1\le k,l\le n}\Big(\big\langle \nu^u, \phi\circ\theta_{k,l}\big\rangle - \big\langle \nu^u, \phi\big\rangle\Big), \tag{2.28}$$

where we put $r_{k,k} = 0$ for all $k \ge 1$, and

$$\big(\theta_{k,l}\big((r_{i',j'})_{1\le i'<j'}\big)\big)_{i,j} := \begin{cases} r_{i,j}, & \text{if } i, j \ne l \\ r_{i\wedge k, i\vee k}, & \text{if } j = l, \\ r_{j\wedge k, j\vee k}, & \text{if } i = l. \end{cases} \tag{2.29}$$

Note that $\Omega^{\uparrow}\Phi \in \Pi$ for all $\Phi \in \Pi^1$.

Our first main result states that the martingale problem associated with $(\Omega^{\uparrow}, \Pi^1)$ is well-posed.

Theorem 1 (Well-posed martingale problem) *For all* $\mathbf{P}_0 \in \mathcal{M}_1(\mathbb{U})$*, the* $(\mathbf{P}_0, \Omega^{\uparrow}, \Pi^1)$*-martingale problem is well-posed.*

This leads to the following definition.

Definition 2.10 (*The tree-valued Fleming–Viot dynamics*) Fix $\mathbf{P}_0 \in \mathcal{M}_1(\mathbb{U})$. The tree-valued Fleming–Viot dynamics with initial distribution $\mathbf{P}_0$ is a stochastic process with distribution $\mathbf{P}$, the unique solution of the $(\mathbf{P}_0, \Omega^{\uparrow}, \Pi^1)$-martingale problem.

Proposition 2.11 (Sample path properties) *The tree-valued Fleming–Viot dynamics* $\mathcal{U}$ *has the following properties.*

(i) $\mathcal{U}$ *has sample paths in* $\mathcal{C}_{\mathbb{U}}([0,\infty))$*,* $\mathbf{P}$*-almost surely.*
(ii) $\mathcal{U}_t \in \mathbb{U}_c$*, for all* $t > 0$*,* $\mathbf{P}$*-almost surely.*

Proposition 2.12 (Feller property) *The tree-valued Fleming–Viot dynamics* $\mathcal{U}$ *is a strong Markov process. Moreover, it has the Feller property, i.e.,* $u \mapsto \mathbf{E}[f(\mathcal{U}_t)|\mathcal{U}_0 = u]$ *is continuous if* $f \in \mathcal{C}_b(\mathbb{U})$.

Corollary 2.13 (Quadratic variation) *Let* $\mathcal{U} = (\mathcal{U}_t)_{t\ge 0}$ *be the tree-valued Fleming–Viot dynamics with initial distribution* $\mathbf{P}_0 \in \mathcal{M}_1(\mathbb{U})$ *and* $\Phi = \Phi^{n,\phi} \in \Pi^1$*, Then* $\Phi(\mathcal{U}) := (\Phi(\mathcal{U}_t))_{t\ge 0}$ *is a continuous* $\mathbf{P}$*-semi-martingale with quadratic variation*

$$\langle \Phi(\mathcal{U})\rangle_t = \gamma n^2 \int_0^t \mathrm{d}s\, \big\langle \nu^{\mathcal{U}_s}, (\bar{\phi}, \bar{\phi})_n \circ \theta_{1,n+1} - (\bar{\phi}, \bar{\phi})_n\big\rangle. \tag{2.30}$$

Remark 2.14 (Quadratic variation for a representative) Assume that for all $t > 0$, $\mathcal{U}_t = \overline{(U_t, r_t, \mu_t)}$. Then the quadratic variation of $\Phi(\mathcal{U})$ can be expressed as

$$\langle \Phi(\mathcal{U})\rangle_t = \gamma n^2 \int_0^t \mathrm{d}s\, \langle \mu_s, \big(\chi_s - \langle \mu_s, \chi_s\rangle\big)^2\rangle, \tag{2.31}$$

where $\chi_s = \chi_s^\phi : U_s \to \mathbb{R}$ is defined as

$$\chi_s(u_1) := \int \mu_s^{\otimes \mathbb{N}}(\mathrm{d}(u_2, u_3, \ldots))\, \bar{\phi}((r_s(u_i, u_j))_{1\le i<j}). \tag{2.32}$$

Remark 2.15 (The rôle of μ) Throughout the paper we encode trees as metric measure spaces rather than just metric spaces. In the context of resampling, given $u = \overline{(U, r, \mu)}$, the measure μ can be understood as the weak limit of empirical distribution of the individuals in the population (which are associated with points in (U, r)). This observation is in analogy to the measure-valued Fleming–Viot processes which arises as the large population limit of empirical distributions on type space. Moreover, the additional structure of a probability measure μ allows for defining polynomials and is therefore very helpful to come up with a suitably large class of generic functions on equivalence classes of measure metric spaces.

Remark 2.16 (Extended martingale problem) The martingale approach characterizes a Markov process through a separating class of martingales. Here, for example, the operator $(\Omega^\uparrow, \Pi^1)$ extends to an operator on the algebra

$$\mathcal{F} = \{f \circ \Phi : f \in \mathcal{B}(\mathbb{R}), \Phi \in \Pi\} \tag{2.33}$$

with domain

$$\mathcal{F}^{2,1} := \big\{f \circ \Phi : f \in \mathcal{C}_b^2(\mathbb{R}), \Phi \in \Pi^1\big\} \tag{2.34}$$

as follows (see e.g. [26, Corollary 1.2]):

$$\begin{aligned}\Omega^\uparrow(f\circ\Phi)(u) = {} & f'\big(\Phi(u)\big)\cdot \Omega^\uparrow\Phi(u) \\ & +\tfrac{1}{2} f''\big(\Phi(u)\big)\cdot \gamma n^2 \cdot \big\langle \nu^u, (\bar{\phi}, \bar{\phi})_n \circ \theta_{1,n+1} - (\bar{\phi}, \bar{\phi})_n\big\rangle.\end{aligned} \tag{2.35}$$

In particular, the tree-valued Fleming–Viot dynamics is the unique solution of the $(\Omega^\uparrow, \mathcal{F}^{2,1})$-martingale problem.

Remark 2.17 (Reduced martingale problem) In view of Remark 2.16 one is interested in finding a preferably minimal class of functions such that the martingales (2.12) uniquely determine the process. Here, for example, we can use the class of *prime* polynomials, where we want to refer to to $\Phi \in \Pi$ as *prime* if Φ is not of the form $\Phi \neq \widehat{\Phi}\cdot\widetilde{\Phi}$ for non-constant $\widehat{\Phi}, \widetilde{\Phi} \in \Pi$. Indeed by (2.35) together with Corollary 2.13

it is easy to see that an $\mathbb{U}$-valued process $\mathcal{U} = (\mathcal{U}_t)_{t\geq 0}$ is the unique solution of the $(\Omega^{\uparrow}, \Pi^1)$-martingale problem iff

$$\left(\Phi(\mathcal{U}_t) - \Phi(\mathcal{U}_0) - \int_0^t \mathrm{d}s\, \Omega^{\uparrow}\Phi(\mathcal{U}_s)\right)_{t\geq 0} \tag{2.36}$$

is a martingale for all prime $\Phi \in \Pi^1$ with quadratic variation given by (2.30).

2.3 Particle approximation (Theorem 2)

A classical result in population genetics gives the approximation of the measure-valued Fleming–Viot process by a finite population model—the so called Moran model—in the limit of large population size (see e.g. [5,22]). In this model, ordered pairs of individuals are replaced by new pairs in a way that the "children" choose a parent—which then becomes their common ancestor—independently at random from the parent pair. In this subsection we state that also the tree-valued Fleming–Viot dynamics can be approximated by tree-valued resampling dynamics which correspond to the Moran model.

We will proceed as follows. For further reference, we provide with Proposition 2.22 a condition for the compact containment condition for finite population models in a general setting. For example, the population size in Definition 2.18 and Proposition 2.22 is not assumed anymore to be constant, and τ denotes the time when the population goes extinct. We use Proposition 2.22 for the convergence of the tree-valued Moran dynamics to the tree-valued Fleming–Viot dynamics in the proof of Theorem 2 (where we have a constant population size and $\tau = \infty$.) Our compact containment condition will be applicable also in the construction of evolving Λ-coalescents or branching trees.

Definition 2.18 (*Finite population dynamics*) Let $(\Omega, (\mathcal{A}_t)_{t\geq 0}, \mathbf{P})$ be a filtered probability space. Let $\mathcal{I} = (\mathcal{I}_t)_{t\in\mathbb{R}}$ be an adapted process with values in $\{\{1,\ldots,n\} : n \in \mathbb{N}_0\}$. For each $t \in \mathbb{R}$, we refer to $\mathcal{I}_t$ as the population at time t. Furthermore, let $\preceq = (\preceq_t)_{t\geq 0}$ be an adapted process such that for all $t \geq 0$, $\preceq_t$ is a partial order on $\{(i,s) : s \in (-\infty, t],\ i \in \mathcal{I}_s\}$ which defines the genealogical relationships at all times before t and satisfies the following:

- (i) for all $r, s, t \in \mathbb{R}$ with $0, r \leq s \leq t$, $i_r \in \mathcal{I}_r$, and $i_s \in \mathcal{I}_s$, $(i_r, r) \preceq_s (i_s, s)$ implies that $(i_r, r) \preceq_t (i_s, s)$, i.e., order relations from earlier times are preserved,
- (ii) for all $i \in \mathcal{I}_t$ and $s \leq t$ there is a unique $A_s(i,t) \in \mathcal{I}_s$ such that $(A_s(i,t), s) \preceq_t (i,t)$. We say that $A_s(i,t)$ is the ancestor of i at time s,
- (iii) for all $i, j \in \mathcal{I}_0$ there is an almost surely finite time T^0_{ij} such that $A_{T^0_{ij}}(i,t) = A_{T^0_{ij}}(j,t)$, i.e., all individuals at time $t = 0$ are related.

Let $\tau := \inf\{s \geq 0 : \mathcal{I}_s = \emptyset\}$ be the lifetime of the population. Put then for all $t \leq \tau$ and $i, j \in \mathcal{I}_t$,

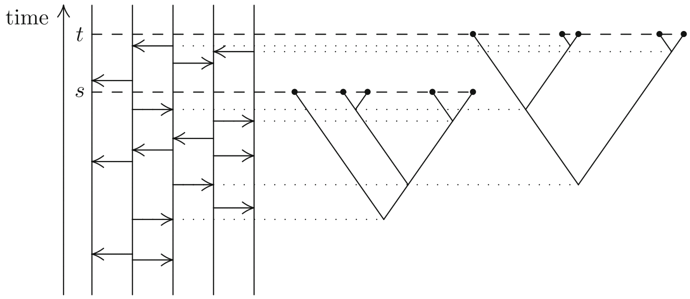

Fig. 1 The graphical representation of a Moran model of size $N = 5$. By resampling the genealogical relationships between individuals change. *Arrows* between *lines* indicate resampling events. The individual at the tip dies and the other one reproduces. At any time, genealogical relationships of individuals $\bullet$, which are currently alive, can be read from this graphical representation

$$r_t(i, j) := 2\big(t - \sup\big\{s \le t : A_s(i, t) = A_s(j, t)\big\}\big). \tag{2.37}$$

The tree-valued population dynamics $(\mathcal{U}_t)_{t\in[0,\tau)}$ read off from $(\mathcal{I}, \preceq)$ and is defined as follows: for all $t \in [0, \tau)$,

$$\mathcal{U}_t := \overline{\left(\mathcal{I}_t, r_t, \tfrac{1}{|\mathcal{I}_t|}\sum_{i\in\mathcal{I}_t}\delta_i\right)} \in \mathbb{U}. \tag{2.38}$$

For a particular choice of $(\mathcal{I}, \preceq)$ we obtain the Moran dynamics. (Compare also with Fig. 1).

Definition 2.19 (*Tree-valued Moran dynamics of population size N*) Fix $N \in \mathbb{N}$. The tree-valued population Moran dynamics with population size N is the tree-valued population dynamics read off from $(\mathcal{I}, \preceq)$ as follows: Put $\mathcal{I} = (\mathcal{I}_t)_{t\in\mathbb{R}}$ with $\mathcal{I}_t := \mathcal{I}^N := \{1, 2, \dots, N\}$ for all $t \in \mathbb{R}$. Let $\preceq_0$ be a random partial order on $(-\infty, 0] \times \mathcal{I}^N$ which satisfies (iii) in Definition 2.18, almost surely. Consider an independent family of rate $\frac{\gamma}{2}$-Poisson processes $\eta := \{\eta^{i,j};\ i, j \in \mathcal{I}^N\}$. (Note that at time $\eta^{i,j}$ an arrows from i to j appears in the graphical representation, Fig. 1.)

For any $s, t \in \mathbb{R}$ with $0 \le s \le t$ and $i_s, i_t \in \mathcal{I}^N$, we say that $(i_s, s) \preceq_t (i_t, t)$ iff there is a path of descent from (i_s, s) to (i_t, t), i.e., if there exist $n \in \mathbb{N}$, $s =: u_0 \le u_1 < u_2 < \cdots < u_n := t$ and $j_1 := i_s, j_n := i_t, j_1, \dots, j_{n-1} \in \{1, 2, \dots, N\}$ such that for all $k \in \{1, \dots, n\}$, $\eta^{j_{k-1}, j_k}\{u_k\} = 1$ and $\eta^{m, j_{k-1}}(u_{k-1}, u_k) = 0$ for all $m \in \mathcal{I}^N$.

In empirical population genetics models for finite populations rather than infinite populations are of primary interest. The next result states that the known convergence of Moran to Fleming–Viot dynamics holds also on the level of trees.

Theorem 2 (Convergence of Moran to Fleming–Viot dynamics) *For $N \in \mathbb{N}$, let $\mathcal{U}^N := (\mathcal{U}_t)_{t\ge 0}$ be the tree-valued Moran dynamics of population size N, and let $\mathcal{U} = (\mathcal{U}_t)_{t\ge 0}$ be the tree-valued Fleming–Viot dynamics. If $\mathcal{U}_0^N \Rightarrow \mathcal{U}_0$ weakly with respect to the Gromov-weak topology, as $N \to \infty$, then*

$$\mathcal{U}^N \underset{N\to\infty}{\Longrightarrow} \mathcal{U}, \tag{2.39}$$

weakly with respect to the Skorohod topology on $\mathcal{D}_{\mathbb{U}}([0,\infty))$.

Remark 2.20 (Connection with the look-down process) Since all the information about trees seems to be contained in the look-down construction of [9], one might wonder whether one could read off the tree-valued Fleming–Viot dynamics from there. This works for the well-posedness of the Fleming–Viot martingale problem as we want to sketch here shortly. Recall that the look-down construction contains the tree-valued Moran dynamics for different population sizes on the same probability space as follows: Put $\mathcal{I} \equiv \mathbb{N}$. Choose a partial order $\preceq_0$ on $(-\infty,0]\times\mathbb{N}$ which satisfies (iii) in Definition 2.18. Consider an independent family of rate γ-Poisson processes $\eta := \{\eta^{i,j};\ 1\le i<j\}$. As in Definition 2.19, let for any $0\le s\le t$ and $i_s, i_t \in \mathbb{N}$, $(i_s,s)\preceq_t (i_t,t)$ iff there is a path of descent from (i_s,s) to (i_t,t). As in (2.37) we can define a process $(\underline{\underline{R}}_t)_{t\ge 0}$ with $\underline{\underline{R}}_t := (R_t(i,j))_{1\le i<j}$ which satisfy for all $1\le i<j$,

$$\begin{aligned} &R_t(i,j) \quad \text{grows linearly at speed } 2,\\ &R_t(i,j)=0 \quad \text{if } t\in\eta^{i,j},\\ &R_t(i,j)=R_{t-}(k,j) \quad \text{if } t\in\eta^{k,i} \text{ for some } k<i,\\ &R_t(i,j)=R_{t-}(i\wedge k, i\vee k) \quad \text{if } t\in\eta^{k,j} \text{ for some } i\neq k<j. \end{aligned} \tag{2.40}$$

If $\preceq_0$ is exchangeable, the tree-valued population dynamics $(\mathcal{U}_t^N)_{t\ge0}$ read off from the restricted graphical representation $(\{1,2,3,\dots,N\},\preceq)$ equals the tree-valued Moran dynamics, for each $N\in\mathbb{N}$. Moreover, the almost sure limit

$$\mathcal{U}_t^\infty := \lim_{N\to\infty} \mathcal{U}_t^N \tag{2.41}$$

exists for all $t\ge 0$. (Compare with Theorem 4 in [30]). This limit easily extends to finitely many time points. By the Kolmogorov extension theorem, existence of a process $(\mathcal{U}_t^\infty)_{t\ge0}$ with these finite dimensional distributions follows, as well as convergence of finite Moran models in finite dimensional distributions. In addition, with a bit more effort it is possible to show that there is a modification of $(\mathcal{U}_t^\infty)_{t\ge0}$ with continuous sample paths, as an estimate of $\mathbb{E}[(\Phi(\mathcal{U}_t)-\Phi(\mathcal{U}_s))^4]$ for $\Phi\in\Pi^1$ reveals.

The process $(\mathcal{U}_t^\infty)_{t\ge0}$ solves the martingale problem for $\Omega^\uparrow$. Indeed, if $\mathbf{P}_0\in\mathcal{M}_1(\mathbb{U})$ is independent of the Poisson processes and its first moment measure equals the distribution of $(R_0(i,j))_{1\le i<j}$, then the process $(\underline{\underline{R}}_t)_{t\ge0}$ is the unique strong Markov process with generator $\widetilde{\Omega}$ acting on functions $\phi\in\mathcal{C}_1^b\big(\mathbb{R}_+^{\binom{\mathbb{N}}{2}}\big)$ which depend only on finitely many coordinates given by

$$\widetilde{\Omega}\phi := \langle \nabla\phi, \underline{\underline{2}}\rangle + \gamma \sum_{1\le k<l} \big(\phi\circ\theta_{k,l}-\phi\big) \tag{2.42}$$

with $\theta_{k,l}$ as in (2.29). That is, for $\Phi = \Phi^{\phi,n} \in \Pi^1$,

$$\Omega^{\uparrow}\Phi^{\phi,n}(u) = \Phi^{\widetilde{\Omega}\phi,n}(u) = \langle \nu^u, \widetilde{\Omega}\phi \rangle \tag{2.43}$$

and therefore by exchangeability, for all $\phi \in \mathcal{C}_b(\mathbb{R}^{\binom{\mathbb{N}}{2}})$,

$$\mathbb{E}\big[\langle \nu^{\mathcal{U}_t^\infty}, \phi \rangle\big] = \mathbb{E}\big[\phi\big(\underline{\underline{R}}_t\big)\big]. \tag{2.44}$$

Since distance matrix distributions are determined uniquely by their first moment measure (this follows since polynomials are separating, see Remark 2.8), the process $(\mathcal{U}_t^\infty)_{t\geq 0}$ solves the $(\mathbf{P}_0, \Omega^{\uparrow}, \Pi^1)$-martingale problem.

However, the above arguments establish convergence of Moran models to the tree-valued Fleming–Viot dynamics only in finite-dimensional distributions. A proof of tightness of Moran models in $\mathcal{D}_{\mathbb{U}}([0,\infty))$ must be carried out to obtain a full convergence result as stated in Theorem 2. We therefore follow a different route, which also has the advantage of not explicitly relying on an exchangeable population model. Hence our approach allows also for the construction of tree-valued dynamics coming from population models with selection and recombination, or more generally, also from tree-valued Markov chains arising outside the context of population models.

Remark 2.21 (Universality) The measure-valued Fleming–Viot process is universal in the sense that it is the limit point of frequency paths of various exchangeable population models of constant size. (A precise condition is found in [38].) We conjecture that the same universality holds on the level of trees, i.e., the tree-valued Fleming–Viot dynamics is the point of attraction of various exchangeable tree-valued dynamics. The crucial step for convergence of tree-valued processes is tightness of the finite models; see Sect. 6.3 in the case of the tree-valued Moran dynamics.

The proof of Theorem 2 relies on a criterion for the compact containment condition in $\mathbb{U}$ to hold. We state it here for the class of population dynamics given in Definition 2.18. It is based on the number of ancestors and descendants.

For $t \in [0,\tau)$ and $\varepsilon > 0$, denote by

$$S_{2\varepsilon}(\mathcal{U}_t) := \#\big\{A_{t-\varepsilon}(i,t) : i \in \mathcal{I}_t\big\} \tag{2.45}$$

the *number of ancestors of $\mathcal{I}_t$ at time $t-\varepsilon$*, and by

$$\widetilde{S}_{2\varepsilon}(\mathcal{U}_t) := \inf_{\mathcal{J}\subseteq\mathcal{I}_t:\, \#\mathcal{J}\leq 2\varepsilon\#\mathcal{I}_t} \#\big\{A_{t-\varepsilon}(i,t) :\ i \in \mathcal{I}_t \backslash \mathcal{J}\big\} \tag{2.46}$$

the *minimal number of ancestors at time $t-\varepsilon$ whose descendants cover a fraction of at least $1-2\varepsilon$ of the time-t-population*. For $t \geq \tau$ and $\varepsilon > 0$, set $\widetilde{S}_{2\varepsilon}(\mathcal{U}_t) = S_{2\varepsilon}(\mathcal{U}_t) = 0$. Moreover, for $\mathcal{J} \subseteq \mathcal{I}_s$ and $s \leq t$, let

$$D_t(s,\mathcal{J}) := \#\big\{i \in \mathcal{I}_t :\ A_s(i,t) \in \mathcal{J}\big\} \tag{2.47}$$

denote the *number of descendants of the set $\mathcal{J}$ at time t*.

The following criterion for a compact containment condition will be proved in Sect. 6. It uses the setting of finite population models from Definition 2.18. Recall that the population size is in general not constant and τ refers to the time the population goes extinct.

Proposition 2.22 (Compact containment for population dynamics) *For each* $N \in \mathbb{N}$, *let* $(\Omega^N, (\mathcal{A}_t^N)_{t\in\mathbb{R}}, \mathbf{P}^N)$, $(\mathcal{I}^N, \preceq^N)$, *and* τ^N *be as in Definition 2.18. Let* $\mathcal{U}^N = (\mathcal{U}_t^N)_{t\in[0,\tau^N)}$ *be the tree-valued population dynamics read off from* $(\mathcal{I}^N, \preceq^N)$.

Assume that the family $\{\mathcal{U}_0^N;\ N \in \mathbb{N}\}$ *is tight in* $\mathbb{U}$. *Furthermore fix* $T > 0$, *and consider the following assumptions:*

(i) *For all* $0 < \varepsilon < T$ *there exists a* $\delta = \delta(\varepsilon) > 0$ *such that for all* $s \in [0, T)$, $N \in \mathbb{N}$ *and* $\mathcal{A}_s^N$*-measurable random subsets* $\mathcal{J}^N \subset \mathcal{I}_s^N$ *with* $\#\mathcal{J}^N \le \delta \cdot \#\mathcal{I}_s^N$,

$$\limsup_{N\in\mathbb{N}} \mathbf{P}^N \left\{ \sup_{t\in[s,T\wedge\tau^N)} \frac{D_t^N(\mathcal{J}^N,s)}{\#\mathcal{I}_t^N} > \varepsilon \right\} \le \varepsilon. \tag{2.48}$$

(ii.i) *For all* $0 < \varepsilon \le t < T$, *the family* $\{S_{2\varepsilon}(\mathcal{U}_t^N) : N \in \mathbb{N}\}$ *is tight.*
(ii.ii) *For all* $0 < \varepsilon \le t < T$, *the family* $\{\widetilde{S}_{2\varepsilon}(\mathcal{U}_t^N) : N \in \mathbb{N}\}$ *is tight.*

Then, the following compact containment conditions hold:

(a) *Under (i) and (ii.i), for all* $\varepsilon > 0$ *there exists a set* $\Gamma_{\varepsilon,T} \subseteq \mathbb{U}_c$ *which is compact in* $\mathbb{U}_c$ *such that*

$$\inf_{N\in\mathbb{N}} \mathbf{P}^N \big\{ \mathcal{U}_t^N \in \Gamma_{\varepsilon,T} \textit{ for all } t \in [\varepsilon, T \wedge \tau^N) \big\} > 1 - \varepsilon. \tag{2.49}$$

(b) *Under (i) and (ii.ii), for all* $\varepsilon > 0$ *there exist a set* $\widetilde{\Gamma}_{\varepsilon,T} \subseteq \mathbb{U}$ *which is compact in* $\mathbb{U}$ *such that*

$$\inf_{N\in\mathbb{N}} \mathbf{P}^N \big\{ \mathcal{U}_t^N \in \widetilde{\Gamma}_{\varepsilon,T} \textit{ for all } t \in [0, T \wedge \tau^N) \big\} > 1 - \varepsilon. \tag{2.50}$$

2.4 Long-term behavior (Theorem 3)

Genealogical relationships in neutral models are frequently studied since the introduction of the Kingman coalescent in [34]. This stochastic process describes the genealogy of a Moran population in equilibrium and its projective limit as the population size tends to infinity. In this section we formulate the related convergence result for the tree-valued resampling dynamics.

Recall that a partition of $\mathbb{N}$ is a collection $\mathfrak{p} = \{\pi_1, \pi_2, \ldots\}$ of pairwise disjoint subsets of $\mathbb{N}$, also called *blocks*, such that $\mathbb{N} = \cup_i \pi_i$. The partition $\mathfrak{p}$ defines an equivalence relation $\sim_{\mathfrak{p}}$ on $\mathbb{N}$ by $i \sim_{\mathfrak{p}} j$ if and only if there exists a partition element $\pi \in \mathfrak{p}$ with $i, j \in \pi$. We denote by $\mathbb{S}$ the set of partitions of $\mathbb{N}$ and define for each $k \in \mathbb{N}$ the restriction ρ_k on $\mathbb{S}$ to the set $\mathbb{S}_k$ of partitions of $\{1, 2, \ldots, k\}$ by $\rho_k \circ \mathfrak{p} := \{\pi_i \cap \{1, \ldots, k\} : \pi_i \in \mathfrak{p}\}$. Each $\mathfrak{p} \in \mathbb{S}$ can be identified with the sequence $(\rho_1 \circ \mathfrak{p}, \rho_2 \circ \mathfrak{p}, \ldots) \in \mathbb{S}_1 \times \mathbb{S}_2 \times \cdots$. Give $\mathbb{S}$ the topology it inherits as a subset of

$\mathbb{S}_1 \times \mathbb{S}_2 \times \cdots$ with the product of discrete topologies. So $\mathbb{S}$ is compact and metrizable and hence Polish.

Starting in $\mathcal{P}_0 = p \in \mathbb{S}$, the *Kingman coalescent* is the unique $\mathbb{S}$-valued strong Markov process $\mathcal{K} = (\mathcal{K}_s)_{s\geq 0}$ such that any pair of blocks merges at rate γ (see, for example, [35,39]). Every realization $\kappa = (\kappa_s)_{s\geq 0}$ of $\mathcal{K}$ gives a pseudo-metric r^{κ} on $\mathbb{N}$ defined by

$$r^{\kappa}\big(i, j\big) := 2 \cdot \inf\big\{s \geq 0 : i \sim_{\kappa_s} j\big\}, \tag{2.51}$$

i.e., $r^{\kappa}\big(i, j\big)$ is proportional to the time needed for i and j to coalesce. Note that $(\mathbb{N}, r^{\kappa})$ is ultra-metric and that $r^{\kappa}(i, j)$ can be thought of as a genealogical distance. Denote then by (L^{κ}, r^{κ}) the completion of $(\mathbb{N}, r^{\kappa})$. Clearly, (L^{κ}, r^{κ}) is also ultra-metric. Define H_N to be the map which takes a realization of the $\mathbb{S}$-valued coalescent and maps it to (an equivalence class of) a pseudo-metric measure space by

$$H_N : \kappa \mapsto \overline{\Big(L^{\kappa}, r^{\kappa}, \mu_N^{\mathcal{K}} := \tfrac{1}{N}\sum\nolimits_{i=1}^{N} \delta_i\Big)}. \tag{2.52}$$

Notice that for each N, the map H_N is continuous.

By Theorem 4 in [30], there exists a $\mathbb{U}$-valued random variable $\mathcal{U}_\infty$ such that

$$H_N(\mathcal{K}) \underset{N\to\infty}{\Longrightarrow} \mathcal{U}_\infty, \tag{2.53}$$

weakly with respect to the Gromov-weak topology. The limit object $\mathcal{U}_\infty$ is called the *Kingman measure tree*. Since the Kingman coalescent comes immediately down from ∞, the Kingman measure tree is compact (see [23]).

Theorem 3 (Convergence to the Kingman measure tree) *Let $\mathcal{U} = (\mathcal{U}_t)_{t\geq 0}$ be the tree-valued Fleming–Viot dynamics starting in $\mathcal{U}_0$ and $\mathcal{U}_\infty$ the Kingman coalescent measure tree. Then*

$$\mathcal{U}_t \underset{t\to\infty}{\Longrightarrow} \mathcal{U}_\infty. \tag{2.54}$$

In particular, the distribution of $\mathcal{U}_\infty$ is the unique equilibrium distribution of the tree-valued Fleming–Viot dynamics.

Remark 2.23 (Exchange of limits) Recall from Definition 2.19 the tree-valued Moran dynamics $\{\mathcal{U}^N = (\mathcal{U}_t^N)_{t\geq 0}; N \in \mathbb{N}\}$. It is straightforward to check that for all $N \in \mathbb{N}$ and for all possible initial states, $\mathcal{U}_t^N \underset{t\to\infty}{\Longrightarrow} H_N(\mathcal{K})$, and therefore the limits $N \to \infty$ (see Theorem 2) and $t \to \infty$ (see Theorem 3) can be exchanged due to (2.53).

Outline The rest of the paper is organized as follows. As an application we study the evolution of subtree length distributions in Sect. 3. A duality relation of the tree-valued Fleming–Viot dynamics to the tree-valued Kingman coalescent is given in Sect. 4. In Sect. 5 we give a formal construction of tree-valued Moran dynamics using well-posed martingale problems. The Moran models build, as shown in Sect. 6, a tight sequence. Duality and tightness provide the tools necessary for the proof of Theorems 1 through 3, which are carried out in Sect. 8. In Sect. 9 we give the proofs of the applications of Sect. 3.

3 Application: Subtree length distribution (Theorems 4 and 5)

In this section we investigate the distribution of the vector containing the lengths of the subtrees spanned by subsequently sampled points, which is referred to as the *subtree length distribution*. All proofs are given in Sect. 9.

The main result in Sect. 3.1 is that the subtree length distribution uniquely determines ultra-metric measure spaces (Theorem 4). In Sect. 3.2 the corresponding martingale problem and its well-posedness is stated (Theorem 5). In Sect. 3.3 we study with the mean sample Laplace transform a special functional of the subtree length distribution.

3.1 The subtree length distribution (Theorem 4)

Recall from Remark 2.2 that we can isometrically embed any ultra-metric space (U, r_U) via a function φ into a path-connected space (X, r_X) which satisfies the four-point condition (2.8) such that $X\backslash X^\circ$ is isometric to (U, r_U). For a sequence $u_1, \ldots, u_n \in U$ with $n \in \mathbb{N}$, let

$$\begin{aligned} &L_n^{(U,r_U)}\big(\{u_1, \ldots, u_n\}\big) := L_n^{(X,r_X)}\big(\{\varphi(u_1), \ldots, \varphi(u_n)\}\big) \\ &:= \text{length of the subtree of } (X, r) \text{ spanned by } \{\varphi(u_1), \ldots, \varphi(u_n)\}, \end{aligned} \tag{3.1}$$

where for an $\mathbb{R}$-tree (X, r_X) with finitely many leaves the length of the tree is defined as the total mass of the one-dimensional Hausdorff measure on $(X, \mathcal{B}(X))$.

Note that the length of the tree spanned by a finite sample is a function of their mutual distances as we state next.

Lemma 3.1 (Total length of a sub-tree spanned by a finite subset) *For a metric space (X, r_X) satisfying the four point condition (2.8) and for $x_1, \ldots, x_n \in X$,*

$$L_n^{(X,r_X)}\big(\{x_1, \ldots, x_n\}\big) = \frac{1}{2} \inf \left\{ \sum_{i=1}^n r(x_i, x_{\sigma(i)});\ \sigma \in \Sigma_n^1 \right\}, \tag{3.2}$$

where $\Sigma_n^1 :=$ {permutations of $\{1, \ldots, n\}$ with one cycle}.

To specify the distribution of the length of the subtrees of subsequently sampled points we consider the map

$$\underline{\ell} : \begin{cases} \mathbb{R}_+^{\binom{\mathbb{N}}{2}} \to \mathbb{R}_+^{\mathbb{N}} \\ \underline{\underline{r}} \mapsto (0, \ell_2(\underline{\underline{r}}), \ell_3(\underline{\underline{r}}), \ldots), \end{cases} \tag{3.3}$$

where for each $n \in \mathbb{N}$,

$$\ell_n(\underline{\underline{r}}) := \frac{1}{2} \inf \left\{ \sum_{i=1}^{n} r_{i,\sigma(i)};\ \sigma \in \Sigma_n^1 \right\}. \tag{3.4}$$

We then define the *subtree length distribution* of $u \in \mathbb{U}$ by

$$\xi(u) := \underline{\ell}_* \nu^u \in \mathcal{M}_1(\mathbb{R}_+^{\mathbb{N}}). \tag{3.5}$$

The first key result states that the subtree length distribution uniquely characterizes ultra-metric measure spaces.

Theorem 4 (Uniqueness and continuity of tree lengths distribution) *The map* $\xi : \mathbb{U} \to \mathcal{M}_1(\mathbb{R}_+^{\mathbb{N}})$ *from* (3.5) *is injective. Let* $\xi(\mathbb{U}) \subseteq \mathcal{M}_1(\mathbb{R}_+^{\mathbb{N}})$ *be equipped with the weak topology and* $\mathbb{R}_+^{\mathbb{N}}$ *with the product topology. Then,* ξ *and* $\xi^{-1} : \xi(\mathbb{U}) \to \mathbb{U}$ *are continuous.*

Remark 3.2 ($\xi(\mathbb{U})$ is Polish) Take a complete metric $d^{\mathbb{U}}$ on $\mathbb{U}$. Using the injectivity of ξ, we define a metric $d^{\xi(\mathbb{U})}$ on $\xi(\mathbb{U})$ by setting

$$d^{\xi(\mathbb{U})}(\lambda_1, \lambda_2) := d^{\mathbb{U}}(\xi^{-1}(\lambda_1), \xi^{-1}(\lambda_2)), \qquad \lambda_1, \lambda_2 \in \xi(\mathbb{U}). \tag{3.6}$$

Since both, ξ and ξ^{-1}, are continuous (with respect to the weak topology on $\mathcal{M}_1(\mathbb{U})$), we see that $d^{\xi(\mathbb{U})}$ generates the weak topology on $\xi(\mathbb{U})$. Since $\xi(\mathbb{U})$ inherits the separability from $\mathbb{U}$, we conclude that $\xi(\mathbb{U})$ is Polish. □

Remark 3.3 (Conjecture about general tree spaces) Theorem 4 shows uniqueness of the tree length distribution on the space of ultra-metric spaces. We conjecture that uniqueness still holds on the space of metric measure spaces satisfying the four-point condition (2.8).

3.2 Martingale problem of subtree length distribution (Theorem 5)

We investigate the evolution of the subtree length distribution under the tree-valued Fleming–Viot dynamics. That is, given the tree-valued Fleming–Viot dynamics $\mathcal{U} = (\mathcal{U}_t)_{t \geq 0}$, we consider

$$\Xi = (\Xi_t)_{t \geq 0}, \qquad \Xi_t := \xi(\mathcal{U}_t). \tag{3.7}$$

To describe the process Ξ via a martingale problem, we define the operator $\Omega^{\uparrow,\Xi}$ on the algebra $\Pi^{\Xi} := \{\Phi \circ \xi^{-1} : \Phi \in \Pi\}$ with domain $\Pi^{1,\Xi} := \{\Phi \circ \xi^{-1} : \Phi \in \Pi^1\}$ by

$$\Omega^{\uparrow,\Xi}(\Phi \circ \xi^{-1})(\lambda) := \Omega^{\uparrow}\Phi(\xi^{-1}(\lambda)), \tag{3.8}$$

for all $\lambda \in \xi(\mathbb{U})$.

In $\Pi^{1,\Xi}$ we find, in particular, functions $\Psi \in \Pi^{1,\Xi}$ which are of the form

$$\Psi^{\psi}(\lambda) = \langle \lambda, \psi \rangle := \int_{\mathbb{R}_+^{\mathbb{N}}} \lambda(\mathrm{d}\underline{l})\, \psi(\underline{l}), \tag{3.9}$$

for a test function $\psi \in \mathcal{C}_b^1(\mathbb{R}_+^{\mathbb{N}})$ which depends on finitely many entries only. Indeed, if ψ depends only on the first k entries, then $\Psi^{\psi} = \Phi^{k,\psi\circ\ell} \circ \xi^{-1}$.

The main result of the section is the following.

Theorem 5 (The subtree lengths distribution process) *For $\mathbf{P}_0 \in \mathcal{M}_1(\mathbb{U})$, let $\mathcal{U}$ be the tree-valued Fleming–Viot dynamics with initial distribution $\mathbf{P}_0$.*

(i) *The $(\xi_*\mathbf{P}_0, \Omega^{\uparrow,\Xi}, \Pi^{1,\Xi})$-martingale problem is well-posed. Its unique solution is given by $\Xi = (\Xi_t)_{t\geq 0}$ with $\Xi_t = \xi(\mathcal{U}_t)$, for $t \geq 0$. The process Ξ has the Feller property. In addition, $\mathbf{P}$-almost surely, it has continuous sample paths, where $\xi(\mathbb{U}) \subseteq \mathcal{M}_1(\mathbb{R}_+^{\mathbb{N}})$ is equipped with the weak topology.*

(ii) *The action of $\Omega^{\uparrow,\Xi}$ on a function Ψ^{ψ} of the form (3.9) is given by*

$$\Omega^{\uparrow,\Xi}\Psi^{\psi}(\lambda) = \sum_{n\geq 2} n\Big\langle \lambda, \frac{\partial}{\partial l_n}\psi\Big\rangle + \gamma \sum_{n\geq 1} n\big\langle \lambda, \psi\circ\beta_n - \psi\big\rangle \tag{3.10}$$

where
$\beta_n : \{0\}\times\mathbb{R}_+^{\mathbb{N}} \to \{0\}\times\mathbb{R}_+^{\mathbb{N}}$ is given by

$$\beta_n : (l_1 = 0, l_2, l_3, \ldots) \mapsto (l_1 = 0, l_2, \ldots, l_{n-1}, l_n, l_n, l_{n+1}, \ldots). \tag{3.11}$$

3.3 Explicit calculations

We consider in this section the *mean sample Laplace transforms*, i.e., functions of the form (3.9) with test functions

$$\psi(\underline{l}) = e^{-\sigma l_n} \tag{3.12}$$

for some $n \in \mathbb{N}$ and $\sigma \in \mathbb{R}_+$ in (3.9) for each $n \in \mathbb{N}$. Using (3.10) we obtain the following explicit expressions.

Corollary 3.4 (Mean sample Laplace transforms) *Let $\Xi = (\Xi_t)_{t\geq 0}$ be the solution of the $(\xi_*\mathbf{P}_0, \Omega^{\uparrow,\Xi}, \Pi^{1,\Xi})$ martingale problem. For all $\sigma \in \mathbb{R}_+$ and $n \geq 2$, set*

$$g^n(t,\sigma) := \mathbf{E}\Big[\int \Xi_t(\mathrm{d}\underline{l})\, e^{-\sigma l_n}\Big]. \tag{3.13}$$

Then,

$$g^n(t,\sigma) = \frac{\Gamma(n)\Gamma\left(\frac{2}{\gamma}\sigma+1\right)}{\Gamma\left(\frac{2}{\gamma}\sigma+n\right)} + n!\sum_{k=2}^{n} \frac{\binom{n-1}{k-1}(-1)^k\left(\frac{2}{\gamma}\sigma+2k-1\right)}{\Gamma\left(\frac{2}{\gamma}\sigma+n+k\right)} \cdot e^{-k(\sigma+\frac{\gamma}{2}(k-1))t}$$

$$\cdot \left\{ \left(\sum_{m=2}^{k} \frac{\binom{k-1}{m-1}(-1)^m \Gamma\left(\frac{2}{\gamma}\sigma + k + m - 1\right)}{m!} g^m(0;\sigma) \right) - \frac{k-1}{k\left(\frac{2}{\gamma}\sigma + k - 1\right)} \Gamma\left(\frac{2}{\gamma}\sigma + k + 1\right) \right\}. \tag{3.14}$$

In particular, if $g^n(\sigma) = \lim_{t\to\infty} g^n(t;\sigma)$ *then*

$$g^n(\sigma) = \mathbf{E}\big[e^{-\sigma \sum_{k=2}^n \mathcal{E}^k}\big], \tag{3.15}$$

where $\{\mathcal{E}^k;\ k = 2,\ldots,n\}$ *are independent and* $\mathcal{E}^k$ *is exponentially distributed with mean* $\frac{2}{\gamma(k-1)}$, $k = 2,\ldots,n$.

Remark 3.5 (Length of n-Kingman coalescent) Consider the Kingman coalescent started with n individuals, and let L_n denote the total length of the corresponding genealogical tree. Note that (3.15), together with Theorem 3 implies the well-known fact (implicitly stated already in [44]) that

$$L_n \stackrel{d}{=} \sum_{k=2}^{n} \mathcal{E}^k. \tag{3.16}$$

□

4 Duality

Duality is an extremely useful technique in the study of Markov processes. It is well-known that the Kingman coalescent is dual to the neutral measure-valued Fleming–Viot process (see, for example, [5,22]). In this section this duality is lifted to the tree-valued Fleming–Viot dynamics. We apply the duality to show uniqueness of the martingale problem for the tree-valued Fleming–Viot process and its relaxation to the equilibrium Kingman measure tree in Sect. 8.

The dual process Recall from Sect. 2.4 the Kingman coalescent $\mathcal{K} = (\mathcal{K}_s)_{s\geq 0}$ and its state space $\mathbb{S}$ of partitions of $\mathbb{N}$. Since we are constructing a dual to the $\mathbb{U}$-valued dynamics, we add a component which measures genealogical distances. The state space of the dual tree-valued Kingman coalescent therefore is

$$\mathbb{K} := \mathbb{S} \times \mathbb{R}_+^{\binom{\mathbb{N}}{2}}, \tag{4.1}$$

equipped with the product topology. In particular, since $\mathbb{S}$ and $\mathbb{R}_+^{\binom{\mathbb{N}}{2}}$ are Polish, $\mathbb{K}$ is Polish as well.

In the following we call the $\mathbb{K}$-valued stochastic process $\mathcal{K} = (\mathcal{K}_s)_{s\geq 0}$, with

$$\mathcal{K}_s = (\mathcal{K}_s, \underline{\underline{r}}'_s) \tag{4.2}$$

the *tree-valued Kingman coalescent*, if it follows the dynamics:

Coalescence $\mathcal{K} = (\mathcal{K}_s)_{s\geq 0}$ is the $\mathbb{S}$-valued Kingman coalescent with pair coalescence rate γ.
Distance growth At time t, for all $1 \leq i < j$ with $i \not\sim_{\mathcal{K}_s} j$, the genealogical distance $r'_\cdot(i, j)$ grows with constant speed 2.

To state the duality relation it is necessary to associate a martingale problem with the tree-valued Kingman coalescent. Consider for $p \in \mathbb{S}$, the *coalescent operator* $\kappa_p : p^2 \to \mathbb{S}$ such that for $\pi, \pi' \in p$,

$$\kappa_p(\pi, \pi') := \big(p \backslash \{\pi, \pi'\}\big) \cup \big\{\pi \cup \pi'\big\}, \tag{4.3}$$

i.e., κ_p sends two partition elements of the partition p to the new partition obtained by coalescence of the two partition elements into one.

We consider the space (recall ρ_k from Sect. 2.4)

$$\mathcal{G} := \Big\{G \in \mathcal{B}(\mathbb{K}) : G(\cdot, \underline{\underline{r}}') \in \mathcal{C}(\mathbb{S}), G(\cdot, \underline{\underline{r}}') \text{ depends on } p \text{ only through } \rho_k \circ p \text{ for some } k \in \mathbb{N};\ \forall \underline{\underline{r}}' \in \mathbb{R}_+^{\binom{\mathbb{N}}{2}}\Big\} \tag{4.4}$$

and the domain

$$\mathcal{G}^{1,0} := \Big\{G \in \mathcal{G} : \langle \nabla_p^{r'} G, \underline{\underline{2}}\rangle \text{ exists}, \forall p \in \mathbb{S}\Big\} \tag{4.5}$$

with

$$\langle \nabla_p^{r'} \cdot, \underline{\underline{2}}\rangle := 2 \sum_{i \not\sim_p j, i<j} \frac{\partial}{\partial r'_{i,j}} = \sum_{i \not\sim_p j} \frac{\partial}{\partial r'_{i\wedge j, i\vee j}}. \tag{4.6}$$

We then consider the martingale problem associated with the operator $\Omega^{\downarrow}$ on $\mathcal{G}$ with domain $\mathcal{G}^{1,0}$, where $\Omega^{\downarrow} := \Omega^{\downarrow,\text{grow}} + \Omega^{\downarrow,\text{coal}}$, with

$$\Omega^{\downarrow,\text{grow}} G(p, \underline{\underline{r}}') := \langle \nabla_p^{r'} G, \underline{\underline{2}}\rangle (p, \underline{\underline{r}}') \tag{4.7}$$

and

$$\Omega^{\downarrow,\text{coal}} G(p, \underline{\underline{r}}') := \gamma \sum_{\substack{\{\pi,\pi'\}\subseteq p \\ \pi\neq\pi'}} \big(G(\kappa_p(\pi, \pi'), \underline{\underline{r}}') - G(p, \underline{\underline{r}}')\big). \tag{4.8}$$

Fix $\mathbf{P}_0 \in \mathcal{M}_1(\mathbb{K})$. By construction, the tree-valued Kingman coalescent solves the $(\mathbf{P}_0, \Omega^{\downarrow}, \mathcal{G}^{1,0})$-martingale problem.

The duality relation We are ready to state a duality relation between the tree-valued Fleming–Viot dynamics and the tree-valued Kingman coalescent.

To introduce a class $\mathcal{H}$ of *duality functions*, we identify every partition $p \in \mathbb{S}$ with the map p which sends $i \in \mathbb{N}$ to the block $\pi \in p$ iff $i \in \pi$, and put for $p \in \mathbb{S}$,

$$(\underline{\underline{r}})^p := \big(r_{\min p(i),\min p(j)}\big)_{1\le i<j}. \tag{4.9}$$

Let then for each $n \in \mathbb{N}$ and $\phi \in \mathcal{C}_b^1(\mathbb{R}_+^{\binom{\mathbb{N}}{2}})$ depending on the coordinates $(r_{i,j})_{1\le i<j\le n}$ only, the function $H^{n,\phi} : \mathbb{U} \times \mathbb{K} \to \mathbb{R}$ be defined as

$$H^{n,\phi}\big(u, (p, \underline{\underline{r}}')\big) := \int \nu^u(\mathrm{d}\underline{\underline{r}})\, \phi\big((\underline{\underline{r}})^p + \underline{\underline{r}}'\big). \tag{4.10}$$

Notice that then the collection of functions

$$\mathcal{H} = \big\{H^{n,\phi}(\cdot, k) : n \in \mathbb{N}, k \in \mathbb{K}, \phi \in \mathcal{C}_b^1(\mathbb{R}^{\binom{\mathbb{N}}{2}})\big\} \tag{4.11}$$

is equal to Π^1, and thus separates points in $\mathcal{M}_1(\mathbb{U})$, see Remark 2.8.

Proposition 4.1 (Duality relation) *For* $\mathbf{P}_0 \in \mathcal{M}_1(\mathbb{U})$ *and* $k \in \mathbb{K}$, *let* $\mathcal{U} = (\mathcal{U}_t)_{t\ge 0}$ *and* $\mathcal{K} = (\mathcal{K}_t)_{t\ge 0}$ *be solutions of the* $(\mathbf{P}_0, \Omega^\uparrow, \Pi^1)$ *and* $(\delta_k, \Omega^\downarrow, \mathcal{G}^{1,0})$*-martingale problem, respectively. Then, if* $\mathcal{U}$ *and* $\mathcal{K}$ *are independent,*

$$\mathbf{E}\big[H(\mathcal{U}_t, k)\big] = \mathbf{E}\big[H(\mathcal{U}_0, \mathcal{K}_t)\big], \tag{4.12}$$

for all $t \ge 0$ *and* $H \in \mathcal{H}$.

Proof We shall establish that for $H^{n,\phi} \in \mathcal{H}$,

$$\Omega^\uparrow H^{n,\phi}(\cdot, k)(u) = \Omega^\downarrow H^{n,\phi}\phi(u, \cdot)(k). \tag{4.13}$$

Using the fact that $H^{n,\phi}$ is bounded the assertion then follows from Theorem 4.4.11 (with $\alpha = \beta = 0$) in [16].

We verify (4.13) for the two components of the dynamics separately. Observe first that by (2.27) and (4.7),

$$\begin{aligned}\Omega^{\uparrow,\mathrm{grow}} H^{n,\phi}(\cdot, (p, \underline{\underline{r}}'))(u) &= 2 \cdot \int \nu^u(\mathrm{d}\underline{\underline{r}}) \sum_{1\le i<j} \frac{\partial}{\partial r_{i,j}} \phi\big((\underline{\underline{r}})^p + \underline{\underline{r}}'\big)\\ &= \int \nu^u(\mathrm{d}\underline{\underline{r}})\, \langle \nabla_p^{r'} \phi, \underline{\underline{2}}\rangle\big((\underline{\underline{r}})^p + \underline{\underline{r}}'\big)\\ &= \Omega^{\downarrow,\mathrm{grow}} H^{n,\phi}\big(u, \cdot\big)(p, \underline{\underline{r}}'),\end{aligned} \tag{4.14}$$

where we have used in the second equality that $\frac{\partial}{\partial r_{i,j}} \phi((\underline{\underline{r}})^p + \underline{\underline{r}}') = 0$, whenever $i \sim_p j$.

Similarly, using $\theta_{k,l}$ from (2.29),

$$\begin{aligned}&\Omega^{\uparrow,\mathrm{res}} H^{n,\phi}\big(\cdot, (p, \underline{\underline{r}}')\big)(u)\\ &\quad = \frac{\gamma}{2} \int \nu^u(\mathrm{d}\underline{\underline{r}}) \sum_{1\le k,l} \big(\phi\big(\theta_{k,l}(\underline{\underline{r}})^p + \underline{\underline{r}}'\big) - \phi\big((\underline{\underline{r}})^p + \underline{\underline{r}}'\big)\big)\end{aligned}$$

$$= \gamma \int \nu^u(\mathrm{d}\underline{\underline{r}}) \sum_{\substack{\{\pi,\pi'\}\subseteq p \\ \pi\neq\pi'}} \left(\phi\left((\underline{\underline{r}})^{\kappa_p(\pi,\pi')} + \underline{\underline{r}}'\right) - \phi\left((\underline{\underline{r}})^p + \underline{\underline{r}}'\right)\right)$$

$$= \Omega^{\downarrow,\mathrm{coal}} H^{n,\phi}\left(u, \cdot\right)(p, \underline{\underline{r}}'). \tag{4.15}$$

Combining (4.14) with (4.15) yields (4.13) and thereby completes the proof. □

5 Martingale problems for tree-valued Moran dynamics

Fix $N \in \mathbb{N}$, and recall from Definition 2.19 the tree-valued Moran dynamics $\mathcal{U}^N = (\mathcal{U}^N_t)_{t\geq 0}$ of population size N. In this section we characterize the tree-valued Moran dynamics as unique solutions of a martingale problem in Subsection 5.1. We then use an approximation argument to establish the existence of the solution to the Fleming–Viot martingale problem in Sect. 5.2. Section 5.3 establishes a coupling of tree-valued Moran models needed to establish the Feller property of the tree-valued Fleming–Viot dynamics.

Notice that the states of the tree-valued Moran dynamics with population size N are restricted to

$$\mathbb{U}_N := \left\{u = \overline{(U, r, \mu)} \in \mathbb{U} : \; N\mu \in \mathcal{N}(U)\right\} \subset \mathbb{U}_c, \tag{5.1}$$

where $\mathcal{N}(U)$ is the set of integer-valued measures on U. Moreover, if $u \in \mathbb{U}_N$, then u can be represented by the pseudo-metric measure space

$$\left(\{1, 2, \ldots, N\}, r', N^{-1} \sum_{i=1}^{N} \delta_i\right), \tag{5.2}$$

for some pseudo-metric r' on $\{1, \ldots, N\}$. In the following we refer to the elements $i \in \{1, 2, \ldots, N\}$ as the *individuals* of the population of size N.

By construction, the tree-valued Moran dynamics are derived from the following particle dynamics on the representative (5.2):

Resampling At rate $\frac{\gamma}{2} > 0$, a resampling event occurs between two individuals k, l such that distances to l are replaced by distances to k. This implies, in particular, that the genealogical distance between k and l is set to be zero. Equivalently, the measure changes from μ to $\mu + \frac{1}{N}\delta_k - \frac{1}{N}\delta_l$.

Distance growth The distance between any two different individuals i, j grows at speed 2.

5.1 The Martingale problem for a fixed population size N

In this subsection we characterize the resampling and distance growth dynamics by a martingale problem.

Fix $N \in \mathbb{N}$. Similarly as in (2.3), for a metric space (U, r), define a map which sends a sequence of N points to the matrix of mutual distances

$$R^{N,(U,r)} : \begin{cases} U^N \to \mathbb{R}^{\binom{N}{2}} \\ (x_1, \ldots, x_N) \mapsto (r(x_i, x_j))_{1 \le i < j \le N} \end{cases} . \tag{5.3}$$

For a pseudo-metric measure space (U, r, μ) with $N\mu \in \mathcal{N}(U)$, let

$$\begin{aligned} &\mu^{\otimes_\downarrow N}(\mathrm{d}(u_1, \ldots, u_N)) \\ &:= \mu(\mathrm{d}u_1) \otimes \frac{\mu - \frac{1}{N}\delta_{u_1}}{1 - \frac{1}{N}}(\mathrm{d}u_2) \otimes \cdots \otimes \frac{\mu - \frac{1}{N}\sum_{k=1}^{N-1}\delta_{u_k}}{1 - \frac{(N-1)}{N}}(\mathrm{d}u_N), \end{aligned} \tag{5.4}$$

the *sampling* (*without replacement*) *measure* and define the N *distance matrix distribution* (*without replacement*) $\nu^{N,(U,r,\mu)}$ of $\mathfrak{u} = \overline{(U, r, \mu)} \in \mathbb{U}_N$ by

$$\nu^{N,\mathfrak{u}} := \left(R^{N,(U,r)}\right)_* \mu^{\otimes_\downarrow N} \in \mathcal{M}_1(\mathbb{R}_+^{\binom{N}{2}}). \tag{5.5}$$

Observe that $\mathfrak{u} \in \mathbb{U}_N$ is uniquely characterized by its N distance matrix distribution.

Once more, it is obvious that $\nu^{N,(U,r,\mu)}$ depends on (U, r, μ) only through its equivalence class $\overline{(U, r, \mu)} \in \mathbb{U}_N$ leading to the following definition.

Definition 5.1 (*N-distance matrix distribution*) For $N \in \mathbb{N}$, the N distance matrix distribution $\nu^{N,\mathfrak{u}}$ (without replacement) of $\mathfrak{u} \in \mathbb{U}_N$ is defined as the N distance matrix distribution $\nu^{N,(U,r,\mu)}$ of an arbitrary representative (U, r, μ) of the equivalence class $\mathfrak{u} = \overline{(U, r, \mu)}$.

For a measurable, bounded $\phi : \mathbb{R}_+^{\binom{N}{2}} \to \mathbb{R}$, introduce the polynomial Φ_N^ϕ by

$$\Phi_N^\phi(\mathfrak{u}) = \langle \nu^{N,\mathfrak{u}}, \phi \rangle := \int_{\mathbb{R}_+^{\binom{N}{2}}} \nu^{N,\mathfrak{u}}(\mathrm{d}\underline{\underline{r}})\, \phi(\underline{\underline{r}}) \tag{5.6}$$

and set

$$\Pi_N := \left\{ \Phi_N^\phi : \phi \in \mathcal{B}(\mathbb{R}_+^{\binom{N}{2}}) \right\}, \tag{5.7}$$

and

$$\Pi_N^1 := \left\{ \Phi_N^\phi : \phi \in \mathcal{C}_b^1(\mathbb{R}_+^{\binom{N}{2}}) \right\}. \tag{5.8}$$

In contrast to Π^1, the space Π_N^1 does not form an algebra. However, we only require that Π_N^1 is separating on $\mathbb{U}_N$, which can easily be shown.

We define an operator $\Omega^{\uparrow,N} := \Omega^{\uparrow,\mathrm{grow},N} + \Omega^{\uparrow,\mathrm{res},N}$ on Π_N with domain Π_N^1 by independent superposition of *resampling* and *distance growth*.

We begin with the *distance growth* operator $\Omega^{\uparrow,\mathrm{grow},N}$. Since distances of any pair of distinct points grow at speed 2 in periods without resampling, we put

$$\Omega^{\uparrow,\mathrm{grow},N}\Phi_N^\phi := \left\langle \nu^{N,\mathfrak{u}}, \langle \nabla\phi, \underline{\underline{2}} \rangle \right\rangle, \tag{5.9}$$

with $\langle \nabla\phi, \underline{\underline{2}} \rangle$ from (2.20).

For the *resampling operator* $\Omega^{\uparrow,\mathrm{res},N}$, consider first the action on a representative (U, r, μ) of the form (5.2). Any resampling event in which the individual l is replaced by a copy of the individual k changes the measure from μ to $\mu + \frac{1}{N}\delta_k - \frac{1}{N}\delta_l$.

Therefore, since

$$\sum_{1\le k,l\le N} (R^{N,(U,r)})_*\big(\mu + \tfrac{1}{N}\delta_k - \tfrac{1}{N}\delta_l\big)^{\otimes_\downarrow N} = \sum_{1\le k,l\le N} (\theta_{k,l})_*\nu^{N,u} \tag{5.10}$$

we obtain for $u = \overline{(U, r, \mu)}$ that

$$\begin{aligned}
&\Omega^{\uparrow,\mathrm{res},N}\Phi_N^{\phi}(u)\\
&\quad= \frac{\gamma}{2}\sum_{1\le k,l\le N}\Big(\langle (R^{N,(U,r)})_*\big(\mu + \tfrac{1}{N}\delta_k - \tfrac{1}{N}\delta_l\big)^{\otimes_\downarrow N}, \phi\rangle - \langle (R^{N,(U,r)})_*\mu^{\otimes_\downarrow N}, \phi\rangle\Big)\\
&\quad= \frac{\gamma}{2}\sum_{1\le k,l\le N}\big(\langle \nu^{N,u}, \phi\circ\theta_{k,l}\rangle - \langle \nu^{N,u}, \phi\rangle\big).
\end{aligned}\tag{5.11}$$

It is easy to see that for given $N \in \mathbb{N}$, Π^1_N is separating in $\mathbb{U}_N$. We can therefore use the operator $(\Omega^{\uparrow,N}, \Pi^1_N)$ to characterize the tree-valued Moran models analytically.

Proposition 5.2 (Tree-valued Moran dynamics) *For all $N \in \mathbb{N}$ and $\mathbf{P}_0^N \in \mathcal{M}_1(\mathbb{U}_N)$, the $(\mathbf{P}_0^N, \Omega^{\uparrow,N}, \Pi^1_N)$-martingale problem is well-posed.*

Proof Let $(\mathcal{I}^N, \preceq^N)$ in Definition 2.18 be such that the law of $\mathcal{U}_0^N$ equals $\mathbf{P}_0^N$. Then the tree-valued Moran dynamics given by Definition 2.19 solve the $(\mathbf{P}_0^N, \Omega^{\uparrow,N}, \Pi^1_N)$-martingale problem, by construction. This proves *existence*.

For *uniqueness*—following the same line of argument as given in Sect. 4—one can check that the $(\mathbf{P}_0^N, \Omega^{\uparrow,N}, \Pi^1_N)$-martingale problem is dual to the tree-valued Kingman coalescent where the duality functions $\Phi \in \Pi^1_N$ are smooth polynomials that involve sampling without replacement (see, for example, Corollary 3.7 in [29] where a similar duality is proved on the level of the measure-valued processes). □

5.2 Convergence to the Fleming–Viot generator

The goal of this subsection is to show that the operator for the tree-valued Fleming–Viot martingale problem is the limit of the operators for the tree-valued Moran martingale problems. This is one ingredient for the proof of Theorem 2 given in Sect. 8.

Proposition 5.3 *Let $\Phi \in \Pi^1$. There exist $\Phi_1 \in \Pi^1_1, \Phi_2 \in \Pi^1_2, \ldots$ such that*

$$\lim_{N\to\infty}\sup_{u\in\mathbb{U}_N}\big|\Phi_N(u) - \Phi(u)\big| = 0, \tag{5.12}$$

and

$$\lim_{N\to\infty}\sup_{u\in\mathbb{U}_N}\big|\Omega^{\uparrow,N}\Phi_N(u) - \Omega^{\uparrow}\Phi(u)\big| = 0. \tag{5.13}$$

Proof First, define the extension operator

$$\iota_N : \begin{cases} \mathbb{R}^{\binom{N}{2}} \mapsto \mathbb{R}^{\binom{\mathbb{N}}{2}} \\ (r_{i,j})_{1\le i<j\le N} \mapsto (r_{i_{\simeq N}\wedge j_{\simeq N}}, r_{i_{\simeq N}\vee j_{\simeq N}})_{1\le i<j}, \end{cases} \tag{5.14}$$

where $i_{\simeq N} := 1 + ((i-1) \bmod N)$. Fix $\Phi = \Phi^{n,\phi} \in \Pi^1$ for $n \in \mathbb{N}$, $\phi \in \mathcal{C}_b^1(\mathbb{R}_+^{\binom{\mathbb{N}}{2}})$. For $N \ge n$ set $\Phi_N := \Phi_N^{\phi\circ\iota_N} \in \Pi_N^1$. By the definition of the N-distance matrix distribution of a representative (5.5), there is a $C > 0$ such that

$$\begin{aligned} \sup_{u\in\mathbb{U}_N} \big|\Phi_N(u) - \Phi(u)\big| &= \sup_{u\in\mathbb{U}_N} \big|\langle \nu^{N,u}, \phi\circ\iota_N\rangle - \langle \nu^u, \phi\rangle\big| \\ &= \sup_{u\in\mathbb{U}_N} \big|\langle (\iota_N)_*\nu^{N,u} - \nu^u, \phi\rangle\big| \\ &\le \frac{C}{N}||\phi|| \end{aligned} \tag{5.15}$$

for all $N \ge n$. This shows (5.12). For (5.13) observe that $\Omega^{\uparrow}\Phi(u) = \langle \nu^u, \psi\rangle$ and $\Omega^{\uparrow,N}\Phi_N(u) = \langle \nu^{N,u}, \widetilde{\psi}\rangle$ for continuous, bounded functions ψ and $\widetilde{\psi}$ satisfying $\widetilde{\psi} = \psi\circ\iota_N$. Hence, (5.13) follows from (5.15). □

5.3 Coupling tree-valued Moran dynamics

In this section we show how to couple two tree-valued Moran dynamics. In particular, using a metric on ultra-metric measure spaces introduced in [30], we show that the coupled processes become closer as time evolves (Proposition 5.8). This will be an important ingredient in showing the Feller property of the tree-valued Fleming–Viot dynamics stated in Theorem 1.

We fix $N \in \mathbb{N}$ and $\mathcal{I}^N := \{1, \dots, N\}$. Informally, we couple two tree-valued Moran dynamics by using the same resampling events. For this, recall the Poisson processes $\eta = \{\eta^{i,j}; i, j \in \mathcal{I}^N\}$ from Definition 2.19 which determine resampling events. Recall from Definition 2.18 the notion of ancestors $A_s(i,t)$, $i \in \mathcal{I}^N$ and $s \le t$.

In order to be in a position to compare coupled Moran models, we use the following metric on $\mathbb{U}$ introduced in [30, Section 10].

Definition 5.4 (*Modified Eurandom metric*) The modified Eurandom distance between $u_1 = \overline{(U_1, r_1, \mu_1)}$ and $u_2 = \overline{(U_2, r_2, \mu_2)} \in \mathbb{U}$ is given by

$$\begin{aligned} &d'_{\mathrm{Eur}}(u_1, u_2) \\ &\quad := \inf_{\widetilde{\mu}} \int_{U_1^2}\int_{U_2^2} \widetilde{\mu}(d(i_1,i_2))\widetilde{\mu}(d(j_1,j_2))\,\big|r_1(i_1,j_1) - r_2(i_2,j_2)\big| \wedge 1 \end{aligned} \tag{5.16}$$

where the infimum is taken over all couplings of μ_1 and μ_2, i.e.,

$$\widetilde{\mu} \in \big\{\widetilde{\mu}' \in \mathcal{M}_1(U_1\times U_2) : (\pi_k)_*\widetilde{\mu}' = \mu_k, k = 1, 2\big\}, \tag{5.17}$$

with $\pi_k : U_1 \times U_2 \to U_k$ denoting the projection on the kth coordinate, $k = 1, 2$.

Remark 5.5 (Connection to the Gromov-weak topology) By Proposition 10.5 in [30], the distance d'_{Eur} is indeed a metric and generates the Gromov-weak topology but is not complete. In particular, for $\mathbb{U}$-valued random variables $\mathcal{U}, \mathcal{U}_1, \mathcal{U}_2, \dots$ which are all defined on the same probability space, we find that $\mathcal{U}_n \Rightarrow \mathcal{U}$, as $n \to \infty$, iff $\mathbf{E}[d'_{\mathrm{Eur}}(\mathcal{U}_n, \mathcal{U})] \to 0$, as $n \to \infty$. □

Remark 5.6 (Modified Eurandom metric on $\mathbb{U}_N$) Recall $\mathbb{U}_N$ from (5.1), and let $u_k = (\mathcal{I}^N, \widetilde{r}_k, \mu_k)$, $k = 1, 2$, be two $\mathbb{U}_N$-valued random variables. Since μ_k has atoms of size $1/N$, $k = 1, 2$, the modified Eurandom metric is given by

$$d'_{\mathrm{Eur}}(u_1, u_2) = \inf_{\sigma \in \Sigma_{\mathcal{I}^N}} \frac{1}{N^2} \sum_{i,j \in \mathcal{I}^N} |\widetilde{r}_1(i, j) - \widetilde{r}_2(\sigma(i), \sigma(j))| \wedge 1, \qquad (5.18)$$

where $\Sigma_{\mathcal{I}^N}$ is the set of permutations of $\mathcal{I}^N$. Moreover, there exist $(\mathcal{I}^N, r_k, \mu_k) \in u_k$, $k = 1, 2$ such that

$$d'_{\mathrm{Eur}}(u_1, u_2) = \frac{1}{N^2} \sum_{i,j \in \mathcal{I}^N} |r_1(i, j) - r_2(i, j)| \wedge 1. \qquad (5.19)$$

Definition 5.7 (*Coupled tree-valued Moran dynamics*) For $\mathcal{I} = (\mathcal{I}_t)_{t \geq 0}$ and $\mathcal{I}_t := \mathcal{I}^N := \{1, \dots, N\}$, let $\preceq_0^1, \preceq_0^2$ be two partial orders on $(-\infty, 0] \times \mathcal{I}^N$, both satisfying (iii) in Definition 2.18. Moreover, let η be a realization of the Poisson processes given in Definition 2.19, defining the processes $\preceq^1 := (\preceq_t^1)_{t \geq 0}$ and $\preceq^2 := (\preceq_t^2)_{t \geq 0}$ as in Definition 2.19. Then, for $(\mathcal{U}_t^{N,k})_{t \geq 0}$, read off from $(\mathcal{I}, \preceq^k)$, $k = 1, 2$, the process $(\mathcal{U}_t^{N,1}, \mathcal{U}_t^{N,2})_{t \geq 0}$ is referred to as *the coupled tree-valued Moran dynamics* started in $(\mathcal{U}_0^{N,1}, \mathcal{U}_0^{N,2})$.

Proposition 5.8 (Contraction of coupled tree-valued *Moran dynamics) Let* $(\mathcal{U}_t^{N,1}, \mathcal{U}_t^{N,2})_{t \geq 0}$ *be the coupled tree-valued Moran dynamics started in* $(\mathcal{U}^{N,1}, \mathcal{U}^{N,2})$*. Then for all* $t > 0$*,*

$$\mathbf{E}[d'_{Eur}(\mathcal{U}_t^{N,1}, \mathcal{U}_t^{N,2})] = e^{-\gamma t} d'_{Eur}(\mathcal{U}_0^{N,1}, \mathcal{U}_0^{N,2}). \qquad (5.20)$$

Proof Recall from Definition 2.18 that $A_s(i, t)$ is the ancestor of (i, t) by time s and from (2.37) that r_t^1, r_t^2 are the metrics given by the coupled Moran dynamics by time $t \geq 0$.

By the definition of the coupled tree-valued Moran dynamics, for $i, j \in \mathcal{I}_N$,

$$\left|r_t^1(i, j) - r_t^2(i, j)\right| = \left|r_0^1(A_0(i, t), A_0(j, t)) - r_0^2(A_0(i, t), A_0(j, t))\right|. \quad (5.21)$$

Let I, J be independent, uniformly distributed on $\mathcal{I}_N$ and independent of all other random variables. Given $I \neq J$, we distinguish two cases: (i) $s \in \eta^{A_s(I,t), A_s(J,t)} \cup \eta^{A_s(J,t), A_s(I,t)}$ for some $0 \leq s \leq t$. Here, the ancestral lines of I and J were affected by a joint resampling event, resulting in $A_0(I, t) = A_0(J, t)$. This event happens with probability $1 - e^{-\gamma t}$. (ii) In the other case, occurring with probability $e^{-\gamma t}$, we find that $A_0(I,t)$ and $A_0(J,t)$ are again distributed as I and J. Hence, for all $t \geq 0$, by (5.19),

$$\begin{aligned}\mathbf{E}[d'_{\mathrm{Eur}}(\mathcal{U}_t^{N,1},\mathcal{U}_t^{N,2})] &= \mathbf{E}[|r_t^1(I,J) - r_t^2(I,J)| \wedge 1] \\ &= \mathbf{E}[|r_0^1(A_0(I,t),A_0(J,t)) - r_0^2(A_0(I,t),A_0(J,t))| \wedge 1] \\ &= e^{-\gamma t}\mathbf{E}[|r_0^1(I,J) - r_0^2(I,J)| \wedge 1] \\ &= e^{-\gamma t} d'_{\mathrm{Eur}}(\mathcal{U}^{N,1},\mathcal{U}^{N,2}), \end{aligned} \tag{5.22}$$

as claimed. □

6 Limit points are compact

Recall from Definition 2.19 the tree-valued Moran dynamics $\mathcal{U}^N$ with population size $N \in \mathbb{N}$. In this section we show that potential limit points of the sequence $\{\mathcal{U}^N;\ N \in \mathbb{N}\}$ have càdlàg sample paths in $\mathbb{U}$ and take values in the space $\mathbb{U}_c$ of *compact* ultra-metric measure spaces for $t > 0$. In Sect. 6.1 we state a sufficient condition for relative compactness in $\mathbb{M}_c$ and give in Sect. 6.2 a criterion for a sequence of population models to satisfy the compact containment condition. In Sect. 6.3 we apply this criterion to show that the sequence of tree-valued Moran dynamics $\mathcal{U}^N$ satisfies the compact containment condition.

6.1 Relative compactness in $\mathbb{M}_c$

We give a criterion for a set to be *relatively compact* in $\mathbb{M}_c$. In this subsection we are dealing with general (not necessarily ultra-) metric measure spaces. We define for $\chi \in \mathbb{M}$ the *distance distribution* $w_\chi \in \mathcal{M}_1(\mathbb{R}_+)$ by

$$w_\chi(A) := \nu^\chi \left\{ \underline{\underline{r}} \in \mathbb{R}_+^{\binom{\mathbb{N}}{2}} : r_{1,2} \in A \right\}, \tag{6.1}$$

for all $A \in \mathcal{B}(\mathbb{R}_+)$. Recall from [30, Proposition 7.1] the following characterization of relative compactness.

Proposition 6.1 (Characterization of relative compactness in $\mathbb{M}$) *A set $\Gamma \subseteq \mathbb{M}$ is relatively compact in the Gromov-weak topology iff the following two conditions hold:*

(i) *$\{w_\chi : \chi \in \Gamma\}$ is tight in $\mathcal{M}_1(\mathbb{R}_+)$.*
(ii) *For all $\varepsilon > 0$ there exists $C_\varepsilon > 0$ such that* $\sup_{\chi\in\Gamma} \widetilde{S}_\varepsilon(\chi) \leq C_\varepsilon$,

where for $(X,r,\mu) \in \chi \in \mathbb{M}$

$$\widetilde{S}_\varepsilon(\chi) := \min\left\{ K : \exists x_1,\dots,x_K \in X : \mu\left(\bigcup_{k=1}^K B_\varepsilon(x_i)\right) > 1-\varepsilon \right\}. \tag{6.2}$$

The relative compactness criterion in $\mathbb{M}_c$ reads as follows:

Proposition 6.2 (Criterion for relative compactness in $\mathbb{M}_c$) *A set $\Gamma \subseteq \mathbb{M}_c$ is relatively compact in the Gromov-weak topology on $\mathbb{M}_c$ if the following two conditions are satisfied.*

(i) $\{w_{\mathcal{X}} : \mathcal{X} \in \Gamma\}$ *is tight in* $\mathcal{M}_1(\mathbb{R}_+)$.
(ii) *For all* $\varepsilon > 0$ *there exists* $N_\varepsilon \in \mathbb{N}$ *such that* $\sup_{\mathcal{X}\in\Gamma} S_\varepsilon(u) \leq N_\varepsilon$, *where* S_ε *is the minimal number of open* ε*-balls needed to cover* $(supp(\mu), r)$ *for* $(X, r, \mu) \in \mathcal{X} \in \mathbb{M}$.

Remark 6.3 (Relative compactness criterion is only sufficient) By Proposition 6.1, (i) is a necessary condition for relative compactness in $\mathbb{M}$. Note that (ii) is not necessary for relative compactness in $\mathbb{M}_c$: Consider, for example,

$$\Gamma = \{\mathcal{X}_n = \overline{(\{0, 1, \ldots, n\}, r_{\text{eucl}}, \text{Bin}(n, \tfrac{1}{n^2}))} : n \in \mathbb{N}\} \subset \mathbb{M}_c. \tag{6.3}$$

Since $\mathcal{X}_n \to \overline{(\mathbb{N}, r_{\text{eucl}}, \delta_0)}$, as $n \to \infty$, the set Γ is relatively compact, but (ii) does not hold. □

The proof of Proposition 6.2 is based on two Lemmata. Recall that for a metric space (X, r) an *ε-separated set* is a subset $X' \subseteq X$ such that $r(x', y') > \varepsilon$, for all $x', y' \in X'$ with $x' \neq y'$.

Lemma 6.4 (Relation between ε-balls and ε-separated nets) *Fix* $N \in \mathbb{N}$, *a metric space* (X, r) *with* $\#X \geq N + 1$ *and* $\varepsilon > 0$. *The following hold.*

(i) *If* (X, r) *can be covered by* N *open balls of radius* ε, *then* (X, r) *has no* 2ε*-separated sets of cardinality* $k \geq N + 1$.
(ii) *If* (X, r) *has no* ε*-separated set of cardinality* $N + 1$, *then* (X, r) *can be covered by* N *closed balls of radius* ε.

Proof (i) Assume that $x_1, \ldots, x_N \in X$ are such that $X = \bigcup_{i=1}^N B_\varepsilon(x_i)$, where we denote by $B_\varepsilon(x)$ the open ball around $x \in X$ of radius $\varepsilon > 0$. Choose $(N + 1)$ distinct points $y_1, \ldots, y_{N+1} \in X$. By the pigeonhole principle, two of the points must fall into the same ball $B_\varepsilon(x_i)$, for some $i = 1, \ldots, N$, and are therefore in distance smaller than 2ε. Hence $\{y_1, \ldots, y_{N+1}\}$ is not 2ε-separated. Since $y_1, \ldots, y_{N+1} \in X$ were chosen arbitrarily, the claim follows.

(ii) Again, we proceed by contradiction. Let K be the maximal possible cardinality of an ε-separated set in (X, r). By assumption, $K \leq N$. Assume that $S_\varepsilon^K := \{x_1, \ldots, x_K\}$ is an ε-separated set in (X, r). We claim that $X = \bigcup_{i=1}^K \overline{B}_\varepsilon(x_i)$ with $\overline{B}_\varepsilon(x) := \{x' \in X : r(x, x') \leq \varepsilon\}$. Indeed, assume, to the contrary, that $y \in X$ is such that $r(y, x_i) > \varepsilon$, for all $i = 1, \ldots, K$, then $S_\varepsilon^K \cup \{y\}$ is an ε-separated set of cardinality $K + 1$, which gives the contradiction. □

Lemma 6.5 (Bounds on the number of balls to cover a limit point) *Fix* $\varepsilon > 0$ *and* $N \in \mathbb{N}$. *Let* $\mathcal{X} = \overline{(X, r, \mu)}$, $\mathcal{X}_1 = \overline{(X_1, r_1, \mu_1)}$, $\mathcal{X}_2 = \overline{(X_2, r_2, \mu_2)}$, *…be elements of* $\mathbb{M}$ *such that* $\mathcal{X}_n \to \mathcal{X}$ *in the Gromov-weak topology, as* $n \to \infty$. *If* $(\text{supp}(\mu_1), r_1)$, $(\text{supp}(\mu_2), r_2)$, *…can be covered by* N *open balls of radius* ε *then* $(\text{supp}(\mu), r)$ *can be covered by* N *closed balls of radius* 2ε.

Proof Define the restriction operator $\rho_N\big((r_{i,j})_{1\le i<j}\big) := (r_{i,j})_{1\le i<j\le N}$. By Lemma 6.4(i), there is no $n \in \mathbb{N}$ for which $(\mathrm{supp}(\mu_n), r_n)$ has a 2ε-separated set of cardinality $N+1$. Set $B_{2\varepsilon} := (2\varepsilon,\infty)^{\binom{N+1}{2}}$. Notice that $\rho_{N+1}^{-1}(B_{2\varepsilon})$ is open. Moreover, $(\mathrm{supp}(\mu), r)$ has a 2ε-separated set of cardinality $N+1$ if and only if $(\rho_{N+1})_*\nu^{\mathcal{X}}(B_{2\varepsilon}) > 0$. However,

$$0 \le (\rho_{N+1})_*\nu^{\mathcal{X}}(B_{2\varepsilon}) \le \liminf_{n\to\infty}(\rho_{N+1})_*\nu^{\mathcal{X}_n}(B_{2\varepsilon}) = 0 \tag{6.4}$$

by Theorem 5(b) in [30] together with the Portmanteau theorem, therefore $(\rho_{N+1})_*\nu^{\mathcal{X}}(B_{2\varepsilon}) = 0$. By Lemma 6.4(ii), $(\mathrm{supp}(\mu), r)$ can therefore be covered by N closed balls of radius 2ε. □

Proof of Proposition 6.2 Assume (i) and (ii) hold for a set $\Gamma \subseteq \mathbb{M}_c$. First note that by Theorem 2 in [30] the set Γ is relatively compact in $\mathbb{M}$. It remains to show that every limit point of Γ is compact. To see this take $\mathcal{X} \in \mathbb{M}$ and $\mathcal{X}_1, \mathcal{X}_2, \ldots \in \Gamma$ such that $\mathcal{X}_n \to \mathcal{X}$ in the Gromov-weak topology, as $n\to\infty$, and let $\varepsilon > 0$. By Assumption(ii) together with Lemma 6.5, $(\mathrm{supp}(\mu), r)$ can be covered by $N_{\varepsilon/2}$ closed balls of radius ε. Therefore, $\mathcal{X}$ is totally bounded which implies $\mathcal{X} \in \mathbb{M}_c$, and we are done. □

6.2 Compact containment (Proof of Proposition 2.22)

Recall that Proposition 2.22 is based on the general notion of a finite population model; see Definition 2.18. In particular $\preceq = (\preceq_t)_{t\ge0}$ is the process of partial orderings connected to genealogical relationships, $\mathcal{I}^N = (\mathcal{I}_t^N)_{t\ge0}$ is the process of population sizes and τ is the lifetime of the population. Moreover, let for each $N \in \mathbb{N}$, $(\mathcal{U}_t^N)_{t\in[0,\tau^N)}$ be the tree-valued population dynamics read off from $(\mathcal{I}^N, \preceq^N)$.

As a preparation we show two auxiliary lemmata which discuss the consequences of the assumptions made in Proposition 2.22. Recall the distance distribution w_u from (6.1).

Lemma 6.6 (Bounds on the distance distribution under Assumption (i)) *Fix $T > 0$, and assume that $\{\mathcal{U}_0^N : N \in \mathbb{N}\}$ is tight in $\mathbb{U}$. If condition (i) of Proposition 2.22 holds, then for all $\varepsilon > 0$ there is a $C_\varepsilon > 0$ such that*

$$\limsup_{N\to\infty} \mathbf{P}^N\left\{\sup_{t\in[0,T\wedge\tau^N)} w_{\mathcal{U}_t^N}([C_\varepsilon,\infty)) > \varepsilon\right\} \le \varepsilon. \tag{6.5}$$

Proof Let $\varepsilon > 0$. Choose $\delta = \delta(\frac{\varepsilon}{4}) > 0$ such that (2.48) holds for and $N \in \mathbb{N}$ and $\varepsilon/4$ (instead of ε) and any $\mathcal{J}^N$ such that $\mathcal{J}^N \subseteq \mathcal{I}_0^N$ is $\mathcal{A}_0^N$-measurable with $\mu^N(\mathcal{J}^N) \le \delta$. Since $\{\mathcal{U}_0^N : N \in \mathbb{N}\}$ is tight in $\mathbb{U}$, we can find such $\mathcal{A}_0^N$-measurable $\mathcal{J}^N \subseteq \mathcal{I}_0^N$, $N \in \mathbb{N}$, and a constant $\widetilde{C}_\varepsilon > 0$ such that $\mu_0^N(\mathcal{J}^N) \le \delta$, almost surely, and

$$\liminf_{N\to\infty} \mathbf{P}^N\left\{\mathcal{I}_0^N\backslash\mathcal{J}^N \text{ has diameter at most } \widetilde{C}_\varepsilon\right\} > 1 - \tfrac{\varepsilon}{2} \tag{6.6}$$

(see (i) in Proposition 6.1). Clearly, on the event that $\mathcal{I}_0^N \backslash \mathcal{J}^N$ has diameter at most $\widetilde{C}_\varepsilon$, the set $D_t(0, \mathcal{I}_0^n \backslash \mathcal{J}^N)$ of descendants of $\mathcal{I}_0^n \backslash \mathcal{J}^N$ at time t has diameter at most $\widetilde{C}_\varepsilon + 2t$. Hence

$$\begin{aligned}
&\limsup_{N\to\infty} \mathbf{P}^N \left\{ \sup_{t\in[0,T\wedge\tau^N)} w_{\mathcal{U}_t^N}([\widetilde{C}_\varepsilon + 2T, \infty)) > \varepsilon \right\} \\
&\quad \le \limsup_{N\to\infty} \mathbf{P}^N \left\{ \sup_{t\in[0,T\wedge\tau^N)} \mu_t^N(D_t(0, \mathcal{J}^N)) > \tfrac{\varepsilon}{4} \right\} \\
&\qquad + \mathbf{P}^N \left\{ \mathcal{I}_0^N \backslash \mathcal{J}^N \text{ has diameter at least } \widetilde{C}_\varepsilon \right\} \\
&\le \varepsilon
\end{aligned} \tag{6.7}$$

by Assumption (i). The claim follows. □

For the next lemma recall that for all $\varepsilon > 0$ and $\mathcal{U}_t = u = \overline{(U, r, \mu)} \in \mathbb{U}$, $S_{2\varepsilon}(\mathcal{U}_t)$ from (2.45) and $\widetilde{S}_{2\varepsilon}(\mathcal{U}_t)$ from (2.46) denote the minimal numbers of 2ε-balls needed to cover $\mathrm{supp}(\mu)$ or to cover $\mathrm{supp}(\mu)\backslash V$ where the exceptional set $V \subseteq U$ satisfies $\mu(V) \le 2\varepsilon$. In particular, these definitions coincide with the same notions introduced in Propositions 6.1 and 6.2.

Lemma 6.7 (Uniform bounds on $S_{2\varepsilon}$ and $\widetilde{S}_{2\varepsilon}$) *Fix* $T > 0$.

(a) *Assume Condition* (ii.i) *from Proposition* 2.22. *Then for all* $\varepsilon > 0$ *we can find* $C_\varepsilon > 0$ *such that*

$$\limsup_{N\to\infty} \mathbf{P}^N \left\{ \sup_{t\in[\varepsilon,T)} S_{2\varepsilon}(\mathcal{U}_t^N) > C_\varepsilon \right\} \le 2\varepsilon. \tag{6.8}$$

(b) *Assume that the family* $\{\mathcal{U}_0^N;\ N \in \mathbb{N}\}$ *is tight in* $\mathbb{U}$ *and Conditions* (i) *and* (ii.ii) *from Proposition* 2.22. *Then for all* $\varepsilon > 0$ *we can find* $C_\varepsilon > 0$ *such that*

$$\limsup_{N\to\infty} \mathbf{P}^N \left\{ \sup_{t\in[0,T)} \widetilde{S}_{2\varepsilon}(\mathcal{U}_t^N) > C_\varepsilon \right\} \le 2\varepsilon. \tag{6.9}$$

Proof (a) The proof relies heavily on the fact that for all t, t', ε, and ε' such that $[t-\varepsilon, t] \subseteq [t'-\varepsilon', t']$

$$S_{2\varepsilon}(\mathcal{U}_t^N) \ge S_{2\varepsilon'}(\mathcal{U}_{t'}^N). \tag{6.10}$$

Fix $\varepsilon > 0$. Without loss of generality, we assume that $T = k\varepsilon$ for some $k \in \mathbb{N}$. Since for all $t \in [\varepsilon, T)$, the family $\{S_{2\varepsilon}(\mathcal{U}_t^N);\ N \in \mathbb{N}\}$ is tight by assumption, there exists a $C_\varepsilon > 0$ such that for all $N \in \mathbb{N}$,

$$\sum_{i=2}^{2k-1} \mathbf{P}^N \left\{ S_\varepsilon(\mathcal{U}_{i\frac{\varepsilon}{2}}^N) > C_\varepsilon \right\} \le 2\varepsilon. \tag{6.11}$$

Applying (6.10) therefore yields that for all $N \in \mathbb{N}$,

$$\mathbf{P}^N\left\{\sup_{t\in[\varepsilon,T)} S_{2\varepsilon}(\mathcal{U}_t^N) > C_\varepsilon\right\} \le \sum_{i=2}^{2k-1} \mathbf{P}^N\left\{\sup_{t\in[i\frac{\varepsilon}{2},(i+1)\frac{\varepsilon}{2})} S_{2\varepsilon}(\mathcal{U}_t^N) > C_\varepsilon\right\}$$
$$\le \sum_{i=2}^{2k-1} \mathbf{P}^N\left\{S_\varepsilon(\mathcal{U}_{i\frac{\varepsilon}{2}}^N) > C_\varepsilon\right\} \le 2\varepsilon \tag{6.12}$$

and the assertion follows.

(b) We extend the notion introduced in (2.46) by setting for $\varepsilon > 0$ and $0 < \zeta < 1$,

$$\widetilde{S}_{2\varepsilon,\zeta}(\mathcal{U}_t^N) := \inf_{\mathcal{J}\subseteq\mathcal{I}_t:\ \mu_t^N(\mathcal{J})\le\zeta} \#\{A_{t-\varepsilon}(i,t) : i \in \mathcal{I}\backslash\mathcal{J}\}. \tag{6.13}$$

In particular, $\widetilde{S}_{2\varepsilon}(\mathcal{U}_t^N) = \widetilde{S}_{2\varepsilon,2\varepsilon}(\mathcal{U}_t^N)$, and thus for all $0 < \zeta < 1$ and $t \in [\varepsilon, T)$, the family $\{\widetilde{S}_{2\varepsilon,\zeta}(\mathcal{U}_t^N) : N \in \mathbb{N}\}$ is tight by Assumption (ii.ii).

Let t, t', δ and δ' be such that $[t-\delta, t] \subseteq [t'-\delta', t']$. By definition of $\widetilde{S}_{2\delta,\zeta}(\mathcal{U}_t^N)$, for all $0 < \zeta < 1$, $t < \tau^N$ and $N \in \mathbb{N}$ there is a $\mathcal{A}_t^N$-measurable subset $\mathcal{J}^{N,\zeta,t} \subseteq \mathcal{I}_t^N$ such that $\mu_t^N(\mathcal{J}^{N,\zeta,t}) \le \zeta$ and $\mathcal{I}_t^N\backslash\mathcal{J}^{N,\zeta,t}$ can be covered by $\widetilde{S}_{2\delta,\zeta}(\mathcal{U}_t^N)$ balls of radius 2δ. Moreover, for all $\zeta, \zeta' \in (0, 1)$,

$$\left\{\widetilde{S}_{2\delta,\zeta}(\mathcal{U}_t^N) < \widetilde{S}_{2\delta',\zeta'}(\mathcal{U}_{t'}^N)\right\} \subseteq \left\{\mu_{t'}^N\big(D_{t'}(\mathcal{J}^{N,\zeta,t}, t)\big) > \zeta'\right\}, \tag{6.14}$$

and hence

$$\mathbf{P}^N\left\{\widetilde{S}_{2\delta,\zeta}(\mathcal{U}_t^N) < \sup_{t'\in[t,(t-\delta)+\delta')} \widetilde{S}_{2\delta',\zeta'}(\mathcal{U}_{t'}^N)\right\}$$
$$\le \mathbf{P}^N\left\{\sup_{t'\in[t,(t-\delta)+\delta')} \mu_{t'}^N\big(D_{t'}(\mathcal{J}^{N,\zeta,t}, t)\big) > \zeta'\right\}. \tag{6.15}$$

Fix $T > 0$ and $\varepsilon > 0$. Without loss of generality, we assume that $T = k\varepsilon$ for some $k \in \mathbb{N}$ as well as $\tau^N \ge T$. By Condition (i) of Proposition 2.22 applied ($2k$ times) with $s := \frac{1}{2}i\varepsilon$, $i = 0, \ldots, 2k-1$, we can choose a $\zeta = \zeta(\varepsilon, T)$ suitably small such that for each $i = 0, \ldots, 2k-1$ and for all $\mathcal{A}_{i\frac{\varepsilon}{2}}^N$-measurable sets $\mathcal{J}^{N,\zeta,\frac{1}{2}i\varepsilon} \subseteq \mathcal{I}_{\frac{1}{2}i\varepsilon}^N$ with $\mu_s^N(\mathcal{J}^{N,\zeta,\frac{1}{2}i\varepsilon}) \le \zeta$,

$$\limsup_{N\to\infty} \mathbf{P}^N\left\{\sup_{t\in[\frac{1}{2}i\varepsilon,\frac{1}{2}(i+1)\varepsilon)} \mu_t^N(D_t(\mathcal{J}^{N,\zeta,\frac{1}{2}i\varepsilon}, \tfrac{1}{2}i\varepsilon)) > \varepsilon\right\} \le \tfrac{\varepsilon}{2k}. \tag{6.16}$$

Thus, inserting (6.16) into (6.15) applied with $t = \frac{1}{2}i\varepsilon$, $\delta = \frac{\varepsilon}{2}$, $\delta' := \zeta' := \varepsilon$, and ζ from (6.16),

$$\limsup_{N\to\infty} \mathbf{P}^N \left\{ \widetilde{S}_{\varepsilon,\zeta}(\mathcal{U}^N_{\frac{1}{2}i\varepsilon}) < \sup_{t\in[\frac{1}{2}i\varepsilon,\frac{1}{2}(i+1)\varepsilon)} \widetilde{S}_{2\varepsilon,2\varepsilon}(\mathcal{U}^N_t) \right\} \le \tfrac{\varepsilon}{2k}. \tag{6.17}$$

Since for all $\zeta \in (0,1)$, $t \in [\varepsilon, T)$, the family $\{\widetilde{S}_{2\varepsilon,\zeta}(\mathcal{U}^N_t);\ N \in \mathbb{N}\}$ is tight by assumption (ii.ii), and $\{\mathcal{U}^N_0 : N \in \mathbb{N}\}$ is assumed to be tight as well, there exists a $C_\varepsilon > 0$ such that for all $N \in \mathbb{N}$,

$$\sum_{i=0}^{2k-1} \mathbf{P}^N \left\{ \widetilde{S}_{\varepsilon,\zeta}(\mathcal{U}^N_{\frac{1}{2}i\varepsilon}) > C_\varepsilon \right\} \le \varepsilon. \tag{6.18}$$

Therefore

$$\begin{aligned}
&\limsup_{N\to\infty} \mathbf{P}^N \left\{ \sup_{t\in[0,T)} \widetilde{S}_{2\varepsilon}(\mathcal{U}^N_t) > C_\varepsilon \right\} \\
&\le \limsup_{N\to\infty} \sum_{i=0}^{2k-1} \mathbf{P}^N \left\{ \sup_{t\in[\frac{1}{2}i\varepsilon,\frac{1}{2}(i+1)\varepsilon)} \widetilde{S}_{2\varepsilon}(\mathcal{U}^N_t) > C_\varepsilon \right\} \\
&\le \limsup_{N\to\infty} \sum_{i=0}^{2k-1} \mathbf{P}^N \left\{ \widetilde{S}_{\varepsilon,\zeta}(\mathcal{U}^N_{\frac{1}{2}i\varepsilon}) < \sup_{t\in[\frac{1}{2}i\varepsilon,\frac{1}{2}(i+1)\varepsilon)} \widetilde{S}_{2\varepsilon}(\mathcal{U}^N_t) \right\} \\
&\quad + \sum_{i=0}^{2k-1} \mathbf{P}^N \left\{ \widetilde{S}_{\varepsilon,\zeta}(\mathcal{U}^N_{\frac{1}{2}i\varepsilon}) > C_\varepsilon \right\} \\
&\le 2\varepsilon,
\end{aligned} \tag{6.19}$$

which finally shows the assertion. □

Proof of Proposition 2.22 Fix $T > 0$ and $\delta > 0$.

(a) Since Conditions (i) and (ii.i) from Proposition 2.22 hold, we find for all $n \in \mathbb{N}$ a $C_{\delta 2^{-n}} > 0$ such that (6.5) and (6.8) hold with $\varepsilon = \delta 2^{-n}$. Put

$$\Gamma_{1,\delta} := \left\{ u \in \mathbb{U} : \ w_u([C_{\delta 2^{-n}}, \infty)) \le \delta 2^{-n}, \quad \text{for all } n \in \mathbb{N} \right\}, \tag{6.20}$$

and

$$\Gamma_{2,\delta} := \left\{ u \in \mathbb{U}_c : \ S_{2\delta 2^{-n}}(u) \le C_{\delta 2^{-n}}, \quad \text{for all } n \in \mathbb{N} \right\}, \tag{6.21}$$

where we denote by $S_{2\delta 2^{-n}}(u)$ the number of balls of radius $\delta 2^{-n}$ needed to cover u. Then $\Gamma_{1,\delta} \cap \Gamma_{2,\delta}$ is relatively compact in $\mathbb{U}_c$ by Proposition 6.2. Moreover, by Lemma 6.6,

$$\begin{aligned}\inf_{N\in\mathbb{N}} \mathbf{P}^N\big\{\mathcal{U}_t^N \in \Gamma_{1,\delta}, \quad \text{for all } t \in [0, T\wedge\tau^N)\big\} \\ \geq 1 - \sum_{n=1}^{\infty} \sup_{N\in\mathbb{N}} \mathbf{P}^N \left\{ \sup_{t\in[0,T\wedge\tau^N)} w_{\mathcal{U}_t^N}([C_{2^{-n}\delta}, \infty)) > 2^{-n}\delta \right\} \\ \geq 1 - \sum_{n=1}^{\infty} 2^{-n}\delta = 1-\delta. \end{aligned} \tag{6.22}$$

Similar calculations based on Lemma 6.7 show that

$$\inf_{N\in\mathbb{N}} \mathbf{P}^N\big\{\mathcal{U}_t^N \in \Gamma_{2,\delta}, \ \text{for all } t\in[\delta, T)\big\} \geq 1-2\delta. \tag{6.23}$$

Hence

$$\inf_{N\in\mathbb{N}} \mathbf{P}^N(\mathcal{U}_t^N \in \Gamma_{1,\delta}\cap\Gamma_{2,\delta}, \ \text{for all } \delta\in[t, T\wedge\tau^N)) \geq 1-3\delta, \tag{6.24}$$

and (2.49) follows.

(b) Assume the conditions (i) and (ii.ii) from Proposition 2.22. Then for all $n \in \mathbb{N}$ there is a $C_{\delta 2^{-n}} > 0$ such that (6.7) and (6.9) hold with $\varepsilon = \delta 2^{-n}$. Put

$$\Gamma_{3,\delta} := \big\{u\in\mathbb{U}_c : \widetilde{S}_{2\delta 2^{-n}}(u) \leq C_{\delta 2^{-n}}, \quad \text{for all } n\in\mathbb{N}\big\}, \tag{6.25}$$

where $\widetilde{S}_{2\delta 2^{-n}}(u)$ denotes the number of $2\delta 2^{-n}$-balls needed to cover a frequency of $(1-2\delta 2^{-n})$ of u. By Proposition 6.1, $\Gamma_{1,\delta}\cap\Gamma_{3,\delta}$ is compact in $\mathbb{U}$. Moreover, by a similar argument as above we find that

$$\inf_{N\in\mathbb{N}} \mathbf{P}^N\big\{\mathcal{U}_t^N \in \Gamma_{3,\delta}, \quad \text{for all } t\in[0, T\wedge\tau^N)\big\} \geq 1-2\delta, \tag{6.26}$$

which gives (2.50). □

6.3 The compact containment condition for Moran models

The following result is an important step in the proof of tightness of the family of tree-valued Moran dynamics. Recall the distance distribution w_χ from (6.1). The next result states that the family $\{\mathcal{U}^N;\ N\in\mathbb{N}\}$ satisfies all assumptions from Proposition 2.22.

Proposition 6.8 (Compact containment) *Let for each $N \in \mathbb{N}$, $\mathcal{U}^N = (\mathcal{U}_t^N)_{t\geq 0}$ be the tree-valued Moran dynamics of population size N. Assume that the family $(\mathcal{U}_0^N)_{N=1,2,\dots}$ is tight in $\mathcal{M}_1(\mathbb{U})$. Then, the family $\{\mathcal{U}^N : N\in\mathbb{N}\}$ satisfies the Conditions* (i), (ii.i) *and* (ii.ii) *from Proposition* 2.22.

Proof Fix $T>0$, $\varepsilon\in(0,T)$ and $N\in\mathbb{N}$, and note that $\tau^N=\infty$. As for *Condition (i)*, let $s\in[0,T)$ and consider a $\mathcal{A}_s^N$-measurable sequence $(\mathcal{J}^N)_{N\in\mathbb{N}}$ with $\mathcal{J}^N\subseteq\mathcal{I}$. Then the process $Y^N := (Y_t^N)_{t\in[s,T)}$, defined for $t\in[s,T)$ as $Y_t^N := \frac{\#D_t(s,\mathcal{J}^N)}{\#\mathcal{I}}$, is a $\{0, \frac{1}{N}, \ldots, 1\}$-valued birth-death process with transitions $y \mapsto y \pm \frac{1}{N}$ (each) with rate $\frac{1}{2}N^2\gamma y(1-y)$. In particular, Y^N is a martingale, and therefore the claim follows by Doob's maximum inequality.

To verify *Condition* (ii.i)*, notice that* the family $\{S^N_{2\varepsilon}(t);\ N \in \mathbb{N}\}$ is stochastically uniformly bounded by K_ε, where $K = (K_t)_{t\geq 0}$ denotes the process for the number of lines in a rate γ Kingman coalescent. In particular, the family $\{S^N_\varepsilon(t);\ N \in \mathbb{N}\}$ is tight.

Condition (ii.ii) directly follows from Condition (ii.i). □

7 Limit points have continuous paths

It is well-known that the measure-valued Fleming–Viot process has continuous paths (e.g., [5]). In this section we show that the same is true for the tree-valued Fleming–Viot dynamics by controlling the jump sizes in the approximating sequence of Moran models.

Recall from Definition 2.19 the tree-valued Moran model $\mathcal{U}^N$ of population size $N \in \mathbb{N}$.

Proposition 7.1 (Limit points have continuous paths) *If* $\mathcal{U}^N \underset{N\to\infty}{\Longrightarrow} \mathcal{U}$ *for some process* $\mathcal{U}$ *with sample paths in the Skorohod space,* $\mathcal{D}_{\mathbb{U}}([0,\infty))$*, of càdlàg functions from* $[0,\infty)$ *to* $\mathbb{U}$*, then* $\mathcal{U} \in \mathcal{C}_{\mathbb{U}}([0,\infty))$*, almost surely.*

Proof Recall from Sect. 2.3 the construction of the tree-valued Moran dynamics $\mathcal{U}^N = (\mathcal{U}^N_t)_{t\geq 0}, \mathcal{U}^N_t = \overline{(\mathcal{I}, r^N_t, \frac{1}{N}\sum \delta_i)}$ with $\mathcal{I} = \{1, \ldots, N\}$ based on Poisson point processes $\{\eta^{i,j};\ 1 \leq i, j \leq N\}$. (Compare also with Fig. 1). In addition, recall the modified Eurandom metric from Definition 5.4. Note that the tree-valued Moran dynamics has paths in $\mathcal{D}_{\mathbb{U}_c}(\mathbb{R}_+)$, almost surely.

If $\eta^{k,l}\{t\} = 0$ for all $k, l \in \mathcal{I}$, then $\mathcal{U}^N_{t-} = \mathcal{U}^N_t$. Otherwise, if $\eta^{k,l}\{t\} = 1$, for some $k, l \in \mathcal{I}$, then

$$\begin{aligned} d'_{\text{Eur}}\big(\mathcal{U}^N_{t-}, \mathcal{U}^N_t\big) &\leq \frac{1}{N^2}\sum_{i,j} |r^N_{t-}(i,j) - r^N_t(i,j)| \wedge 1 \\ &= \frac{1}{N^2}\sum_{i=l \text{ or } j=l} |r^N_{t-}(i,j) - r^N_t(i,j)| \wedge 1 \\ &\leq \frac{2}{N} \end{aligned} \tag{7.1}$$

and therefore

$$\int_0^\infty \mathrm{d}T\, e^{-T} \sup_{t\in[0,T]} d'_{\text{Eur}}\big(\mathcal{U}^N_{t-}, \mathcal{U}^N_t\big) \leq \frac{2}{N}, \tag{7.2}$$

for all $T > 0$ and almost all sample paths $\mathcal{U}^N$. Hence the assertion follows by Theorem 3.10.2 in [16]. □

8 Proofs of the main results (Theorems 1, 2, 3)

In this section we give the proof of the main results stated in Sect. 2. Theorems 1 and 2 are proved simultaneously.

Proof of Theorems 1 and 2 Recall, for each $N \in \mathbb{N}$, the state-space $\mathbb{U}_N$, and the $\mathbb{U}_N$-valued Moran dynamics, $\mathcal{U}^N = (\mathcal{U}_t^N)_{t\geq 0}$, from (5.1) and Definition 2.19, respectively. Let $\mathbf{P}_0 \in \mathcal{M}_1(\mathbb{U})$ be the distribution of $\mathcal{U}_0$ and $\mathbf{P}_0^N \in \mathcal{M}_1(\mathbb{U}_N)$ be the distribution of $\mathcal{U}_0^N$ such that $\mathcal{U}_0^N \Longrightarrow \mathcal{U}_0$ as $N \to \infty$.

By Proposition 5.2, the $(\mathbf{P}_0^N, \Omega^{\uparrow,N}, \Pi_N^1)$-martingale problem is well-posed, and is solved by $\mathcal{U}^N$. Proposition 5.3 implies with a standard argument (see, for example, Lemma 4.5.1 in [16]) that if $\mathcal{U}^N \Rightarrow \mathcal{U}$, for some $\mathcal{U} \in \mathcal{D}_{\mathbb{U}}([0,\infty))$, as $N \to \infty$, then $\mathcal{U}$ solves the $(\mathbf{P}_0, \Omega^{\uparrow}, \Pi^1)$-martingale problem. Hence for *existence* we need to show that the sequence $\{\mathcal{U}^N;\ N \in \mathbb{N}\}$ is tight, or equivalently by Remark 2.8 combined with Remark 4.5.2 in [16] that the compact containment condition in $\mathbb{U}$ holds. However, the latter follows directly from Propositions 6.8 and 2.22.

By standard theory (see, for example, Theorem 4.4.2 in [16]), *uniqueness* of the $(\mathbf{P}_0, \Omega^{\uparrow}, \Pi^1)$-martingale problem follows from uniqueness of the one-dimensional distributions of solutions of the $(\mathbf{P}_0, \Omega^{\uparrow}, \Pi^1)$-martingale problem. The latter can be verified using the duality of the tree-valued Fleming–Viot dynamics to the tree-valued Kingman coalescent, $\mathcal{K} := (\mathcal{K}_t)_{t\geq 0}$, as defined in (4.2). That is, if $\mathcal{U} = (\mathcal{U}_t)_{t\geq 0}$ is a solution of the $(\mathbf{P}_0, \Omega^{\uparrow}, \Pi^1)$-martingale problem, then (4.12) holds for all $\kappa \in \mathbb{K}$, $t \geq 0$ and $H \in \mathcal{H}$. Since $\mathcal{H}$ is separating in $\mathcal{M}_1(\mathbb{U})$ by Proposition 4.1(i), uniqueness of the one-dimensional distributions follows.

So far we have shown that the $(\mathbf{P}_0, \Omega^{\uparrow}, \Pi_1)$-martingale problem is well-posed and its solution arises as the weak limit of the solutions of the $(\mathbf{P}_0^N, \Omega^{\uparrow,N}, \Pi_N^1)$-martingale problems. In particular, the tree-valued Moran dynamics converge to the tree-valued Fleming–Viot dynamics. Hence we have shown Theorem 1 and Theorem 2. □

Proof of Proposition 2.11 (i), (ii) The tree-valued Fleming–Viot dynamics is the weak limit of tree-valued Moran dynamics. Hence, Propositions 6.8 and 2.22 imply that the tree-valued Fleming–Viot dynamics have values in the space of compact ultra-metric measure spaces for each $t > 0$, almost surely. In addition, the tree-valued Fleming–Viot dynamics has continuous paths by Proposition 7.1, almost surely. □

Proof of Proposition 2.12 Note that the strong Markov property follows from the Feller property, [16, Theorem 4.2.7]. (By completeness, we can assume the filtration generated by the tree-valued Fleming–Viot dynamics is right-continuous, as needed in this Theorem.) Let $\mathcal{U}^{u} = (\mathcal{U}_t^{u})_{t\geq 0}$ be the solution of the $(\delta_{u}, \Omega^{\uparrow}, \Pi^1)$-martingale problem, i.e. the tree-valued Fleming–Viot dynamics, started in $\mathcal{U}_0 = u$. For the Feller property, it suffices to show that $u' \to u$ implies that $\mathcal{U}_t^{u'} \Longrightarrow \mathcal{U}_t^{u}$ for all $u \in \mathbb{U}$ and $t > 0$. Recall the coupled tree-valued Moran dynamics from Section 5.3. For $u, u' \in \mathbb{U}$, take $u_N, u'_N \in \mathbb{U}_N$ with $u_N \to u$ and $u'_N \to u'$ in the Gromov-weak topology. Let $(\mathcal{U}_t^{N,1}, \mathcal{U}_t^{N,2})_{t\geq 0}$ be the coupled tree-valued Moran dynamics, started in (u_N, u'_N). Since $\{(\mathcal{U}_t^{N,k})_{t\geq 0}, N \in \mathbb{N}\}$ is tight in $\mathbb{U}$ by Theorem 2, $k = 1, 2$, $\{(\mathcal{U}_t^{N,1}, \mathcal{U}_t^{N,2})_{t\geq 0} : N \in \mathbb{N}\}$ is tight in $\mathbb{U} \times \mathbb{U}$. Let $(\mathcal{U}_t^{u}, \mathcal{U}_t^{u'})_{t\geq 0}$ be a weak limit point which must be a coupling of tree-valued Fleming–Viot dynamics by construction. Moreover, since the modified Eurandom metric (see Definition 5.4) is continuous in the Gromov-weak topology and bounded

$$\begin{aligned}\mathbf{E}[d'_{\mathrm{Eur}}(\mathcal{U}_t^{u},\mathcal{U}_t^{u'})] &= \lim_{N\to\infty}\mathbf{E}[d'_{\mathrm{Eur}}(\mathcal{U}_t^{N,1},\mathcal{U}_t^{N,2})]\\ &\le \lim_{N\to\infty} d'_{\mathrm{Eur}}(u_N, u'_N)\\ &= d'_{\mathrm{Eur}}(u, u')\end{aligned} \tag{8.1}$$

by Proposition 5.8. In particular, $u'_n \to u$, as $n \to \infty$, implies that

$$\mathbf{E}[d'_{\mathrm{Eur}}(\mathcal{U}_t^{u},\mathcal{U}_t^{u'_n})]\underset{n\to\infty}{\longrightarrow} 0,$$

which in turn implies $\mathcal{U}_t^{u'_n} \Longrightarrow \mathcal{U}_t^{u}$, as $n \to \infty$, by Remark 5.5. □

Proof of Corollary 2.13 For $\Phi = \Phi^{n,\phi}$ as in the Corollary, observe that $\langle \nu^u, \phi\rangle^2 = \langle \nu^u, (\phi,\phi)_n\rangle$ with $(\phi,\phi)_n$ from (2.25). Therefore, given $\mathcal{U}_t = u$, we compute (compare with [26, Proof of Theorem 1.1])

$$\begin{aligned}\frac{\mathrm{d}\langle\Phi(\mathcal{U})\rangle_t}{\mathrm{d}t} &= \Omega^{\uparrow}\Phi^2(u) - 2\Phi(u)\Omega^{\uparrow}\Phi(u)\\ &= \langle \nu^u, \langle\nabla(\phi,\phi)_n, \underline{\underline{2}}\rangle - 2(\phi, \langle\nabla\phi, \underline{\underline{2}}\rangle)_n\rangle\\ &\quad +\frac{\gamma}{2}\sum_{k,l=1}^{n}\langle \nu^u, (\phi\circ\theta_{k,l},\phi)_n + (\phi,\phi\circ\theta_{k,l})_n - 2(\phi,\phi\circ\theta_{k,l})_n\rangle\\ &\quad +\gamma\sum_{k,l=1}^{n}\big(\langle \nu^u, (\phi,\phi)_n\circ\theta_{k,n+l}\rangle - \langle\nu^u,(\phi,\phi)_n\rangle\big)\end{aligned} \tag{8.2}$$

and the result follows from the first two terms vanishing and

$$\begin{aligned}\sum_{k,l=1}^{n}\langle \nu^u, (\phi,\phi)_n\circ\theta_{k,n+l}\rangle &= \sum_{k,l=1}^{n}\langle \nu^u, (\bar\phi,\bar\phi)_n\circ\theta_{k,n+l}\rangle\\ &= n^2\langle \nu^u, (\bar\phi,\bar\phi)_n\circ\theta_{1,n+1}\rangle\end{aligned} \tag{8.3}$$

with the symmetrization $\bar\phi$ introduced in Remark 2.7(iii) and $\Phi^{n,\phi} = \Phi^{n,\bar\phi}$. □

Proof of Theorem 3 In order to prove Theorem 3 we need two ingredients:

- The family $\{\mathcal{U}_t;\ t > 1\}$ is tight.
- $\mathbf{E}^{\delta_u}[\Phi(\mathcal{U}_t)] \to \mathbf{E}[\Phi(\mathcal{U}_\infty)]$, as $t \to \infty$, for all $\Phi \in \Pi^1$ and $u \in \mathbb{U}$.

Then, Theorem 3 follows from Lemma 3.4.3 together with Theorem 3.4.5 of [16].

We show tightness of $\{\mathcal{U}_t;\ t > 1\}$ in $\mathbb{U}$ using Theorem 3 and (3.3) of [30]. First, recalling (6.1) and when $\mathbf{E}[w_{\mathcal{U}_t}]$ is the first moment measure of $w_{\mathcal{U}_t} \in \mathcal{M}_1(\mathcal{M}_1(\mathbb{R}_+))$, for t and $C > 0$,

$$\mathbf{E}[w_{\mathcal{U}_t}]([C,\infty)) = \begin{cases} e^{-\gamma t}\mathbf{E}[w_{\mathcal{U}_0}]([C-t,\infty)), & C \geq t \\ e^{-\gamma C}, & C < t. \end{cases} \tag{8.4}$$

Indeed, by exchangeability $\mathbf{E}[w_{\mathcal{U}_t}]([C,\infty))$ equals the probability that a "typical" pair of individuals drawn from the population at time t has distance at least C, if $t \leq C$ then this event equals the event that their ancestral lines do not coalesce in the time window $[0,t]$ and that the distance of their ancestors at time 0 is at least C. This event has probability $e^{-\gamma t}$ (no coalescence for at least time t) times $\mathbf{E}[w_{\mathcal{U}_0}]([C-2t,\infty))$. If $t > C$ then the distance between a "typical" pair of individuals to be at least C is equivalent to that their ancestral lines do not coalesce in the time window $[0,C]$ which has probability $e^{-\gamma C}$.

So, for given $\varepsilon > 0$, choose $C > 0$ large enough such that $\mathbf{E}[w_{\mathcal{U}_0}]([C,\infty)) < \varepsilon$ and $e^{-\gamma C} < \varepsilon$. Then, $\mathbf{E}[w_{\mathcal{U}_t}]([2C,\infty)) < \varepsilon$ for all $t > 0$ and so, $\{\mathbf{E}[w_{\mathcal{U}_t}], t > 1\}$ is tight.

Secondly, for $\mathcal{U}_t = \overline{(U_t, r_t, \mu_t)}$, we have to show that for $0 < \varepsilon < 1$ there is $\delta > 0$ with

$$\sup_{t>1} \mathbf{E}[\mu_t\{u : \mu_t(B_\varepsilon(u)) \leq \delta\}] < \varepsilon. \tag{8.5}$$

Note that the expectation on the left hand side does not depend on t. Using that $\mathcal{U}_\infty$ is determined by $\Lambda = \gamma \cdot \delta_0$ in (4.7) of [30] we find

$$\lim_{\delta\to 0} \sup_{t>1} \mathbf{E}[\mu_t\{u : \mu_t(B_\varepsilon(x)) \leq \delta\}] = \lim_{\delta\to 0} \mathbf{E}[\mu_\infty\{u : \mu_\infty(B_\varepsilon(x)) \leq \delta\}] = 0 \tag{8.6}$$

by (4.9) and (4.11) of [30]. So, tightness follows.

The fact that the Kingman tree is a unique equilibrium distribution is an application of the duality relation from Proposition 4.1. Fix $\phi \in \mathcal{C}_b^1(\mathbb{R}_+^{\binom{\mathbb{N}}{2}})$. We apply the duality relation (4.12) between the tree-valued Fleming–Viot dynamics and the tree-valued Kingman coalescent which starts in $\mathfrak{k}_0 = (\mathfrak{p}_0, \underline{\underline{r}}'_0)$ with $\mathfrak{p}_0 := \{\{n\}, n \in \mathbb{N}\}$ and $\underline{\underline{r_0}}' \equiv 0$. By construction of the dual process $\mathcal{K}$, $\mathbf{E}^{\delta_{\mathfrak{k}_0}}[\phi(\underline{\underline{r}}_t)] \to \mathbf{E}[\langle \nu^{\mathcal{U}_\infty}, \phi\rangle]$ and $\mathcal{P}_t \to \{\mathbb{N}\}$, as $t \to \infty$ where $\mathcal{U}_\infty$ is the (rate γ) Kingman measure tree from (2.53). Hence, by (4.12),

$$\begin{aligned} \lim_{t\to\infty} \mathbf{E}^{\delta_u}\big[\langle \nu^{\mathcal{U}_t}, \phi\rangle\big] &= \lim_{t\to\infty} \mathbf{E}^{\delta_{\mathfrak{k}_0}}\left[\int_{\mathbb{R}_+^{\binom{\mathbb{N}}{2}}} \nu^u(\mathrm{d}\underline{\underline{r}})\, \phi\big((\underline{\underline{r}})^{\mathcal{P}_t} + \underline{\underline{r_t}}'\big)\right] \\ &= \lim_{t\to\infty} \mathbf{E}^{\delta_{\mathfrak{k}_0}}\big[\phi\big(\underline{\underline{r}}'_t\big)\big] \\ &= \mathbf{E}\big[\langle \nu^{\mathcal{U}_\infty}, \phi\rangle\big]. \end{aligned} \tag{8.7}$$

Since $\phi \in \mathcal{C}_b^1(\mathbb{R}_+^{\binom{\mathbb{N}}{2}})$ was chosen arbitrarily, (ii) follows and we are done. □

9 Proof of the applications (Proof of Theorems 4 and 5)

In this section we prove the results stated in Sect. 3.

Proof of Lemma 3.1 Consider the *traveling salesperson problem* for a salesperson who must visit all $x_1, \ldots, x_n$ and who starts at one x_i to which she comes back at the end of the trip. It is easy to see that such a path must pass all edges of the subtree spanned by $x_1, \ldots, x_n$ in both directions, so the length of the path is at least twice the tree length. It is also easy to see that taking an optimal path and leaving out x_i gives an optimal path for the remaining leaves $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$.

We claim that there is one path connecting the set of leaves such that each edge in the tree is passed exactly twice, which is equivalent to the assertion of the Lemma. Assume to the contrary that such an order does not exist. We take a path of minimal length. There must be one edge which is visited at least four times. W.l.o.g. we assume that this edge is internal, i.e. not adjacent to any x_i. So there are four points $x_i, x_j, x_k, x_l \in X$, visited in the order x_i, x_j, x_k, x_l, x_i, such that $[x_i, x_j] \cap [x_k, x_l]$ is visited at least four times, where $[x, y]$ is the path from x to y in X. Since leaving out leaves gives again an optimal path, leaving out all leaves except x_i, x_j, x_k, x_l must lead to an optimal path connecting these four points. However, this optimal path must be x_i, x_j, x_l, x_k, x_i (or its reverse), since this path passes all edges only twice. Hence, we have a contradiction and the assertion is proved. □

Proof of Theorem 4 We first show *injectivity* of ξ. Assume we are given a compact ultra-metric measure space (U_0, r_0, μ_0) and its equivalence class $\mathfrak{u}_0 = \overline{(U_0, r_0, \mu_0)}$. We show that if $\lambda := \xi(\mathfrak{u}_0)$, then $\xi^{-1}(\{\lambda\}) = \{\mathfrak{u}_0\}$. We do this by explicitly reconstructing $\mathfrak{u}_0$ from λ.

We proceed in three steps. In the first two steps we consider the case where μ_0 is supported by finitely many atoms. In Step 1 we follow an argument provided to us by Steve Evans which explains how to recover the isometry class of $(\text{supp}(\mu_0), r_0)$ from λ. In Step 2 we then recover the measure μ_0. Finally, the case of a general element in $\mathbb{U}$ is obtained by approximation via finite ultra-metric measure spaces in Step 3.

Step 1 (*Evans's reconstruction procedure for finite trees*) Assume that $\mathfrak{u} \in \xi^{-1}(\{\lambda\})$ and that $\mathfrak{u} = \overline{(U, r, \mu)}$ with $\#\text{supp}(\mu) < \infty$. Put

$$A_N := \big\{(l_1 := 0, l_2, \ldots) : l_k > l_{k-1} \text{ for exactly } N-1 \text{ different } k\big\}. \tag{9.1}$$

First observe that $\#\text{supp}(\mu) = N$ if and only if λ is supported on A_N. That is, we can recover $\#\text{supp}(\mu)$ from λ. So, assume that μ has N atoms and w.l.o.g. $U := \{1, \ldots, N\}$. We now recover $\underline{\underline{r}} = (r_{i,j})_{1 \le i < j \le N}$ from λ.

For that purpose, introduce on $\mathbb{R}_+^{\mathbb{N}}$ the lexicographic ordering $\prec$, i.e., $\underline{l} \prec \underline{l}'$ iff for $k^* := \min\{k : l_k \neq l'_k\}$ we have $l_{k^*} < l'_{k^*}$. Let

$$B := \big\{\underline{l} \in \text{supp}(\lambda) : l_1 < \cdots < l_N\big\} \tag{9.2}$$

be the space of all vectors $\underline{l}$ which are accessible by sequentially sampling the N different points of U and evaluating subsequently the lengths of the sub-trees spanned by them. Moreover, let

$$\underline{l}^* := \min_{\prec} B, \tag{9.3}$$

i.e., $\underline{l}^* := (l_k^*)_{k\in\mathbb{N}}$ is the minimal element in B with respect to the order relation $\prec$.

W.l.o.g. we assume that $U = \{1, \dots, N\}$ and that for all $n \in \{1, \dots, N\}$,

$$l_n^* := L_n^{(U,r)}(\{1, \dots, n\}). \tag{9.4}$$

Notice that if d_n^* denotes the depth of the sub-tree spanned by $\{1, \dots, n\}$, i.e., $d_n^* := \frac{1}{2}\max\{r(i, j); 1 \le i, j \le n\}$, for $n \in \mathbb{N}$, then $d_1^* = 0$ and the recursion

$$d_n^* = \frac{1}{2}\big(d_{n-1}^* + (l_n^* - l_{n-1}^*) \vee d_{n-1}^*\big). \tag{9.5}$$

holds for $n \ge 2$.

We claim that we can even recover $(r_{i,j})_{1\le i<j\le N}$ from $(l_n^*)_{n=1,\dots,N}$. In fact, for all $n \in \mathbb{N}$,

$$r_{n-1,n} = \min_{1\le k\le n-1} r_{k,n}, \tag{9.6}$$

To see this, assume to the contrary that there is a minimal $n \in \mathbb{N}$ for which we find a $k < n-1$ such that $r_{k,n}$ is minimal and $r_{k,n} < r_{n-1,n}$. Choose the minimal i with $k < i \le n-1$ and $r_{k,n} < r_{i,n}$. Then, sampling the i points $1, 2, \dots, k, \dots, i-1, n$ (in that order) leads to the sequence of tree lengths $l_1^*, l_2^*, \dots, l_{i-1}^*, l_{i-1}^* + \frac{1}{2}r_{k,n}$. However, by the minimality of i we have that $r_{k,n} \ge r_{i-1,n}$ and by the ultra-metric property $r_{k,n} < r_{i,n} \vee r_{i-1,n} = r_{i-1,i}$. Hence, the above tree lengths are smaller (with respect to $\prec$) than $l_1^*, l_2^*, \dots, l_{i-1}^*, l_i^*$ since $l_i^* \ge l_{i-1}^* + \frac{1}{2}r_{i-1,i}$. So, assuming that (9.6) does not hold contradicts the assumption that l^* is minimal.

However, from (9.6) we conclude the following recursion: for all $n \in \{2, \dots, N\}$ and $1 \le k \le n-1$,

$$r_{k,n} = r_{k,n-1} \vee 2\big(l_n^* - l_{n-1}^* - (d_n^* - d_{n-1}^*)\big). \tag{9.7}$$

The latter together with the necessary requirements that $r_{n,n} := 0$ and $r_{1,2} := \frac{1}{2}l_2^*$ determines the metric on U uniquely.

Step 2 (*Reconstruction of weights in finite trees*) In this step we reconstruct weights $(p_1, \dots, p_N)$ on $(\{1, \dots, N\}, r)$ from the given λ. Denote by $\Gamma \subseteq \Sigma_N$ the set of permutations of $\{1, \dots, N\}$ for which the metric r given in Step 1 satisfies $r_{i,j} = r_{\sigma(i),\sigma(j)}$, for all $1 \le i, j \le N$. Since we are interested in measure-preserving isometry classes only, we need to show that $(p_1, \dots, p_N)$ are uniquely determined up to permutations $\sigma \in \Gamma$.

For all $\underline{k} = (k_1, \dots, k_{N-1}, k_N) \in \{0, 1, \dots\}^{N-1} \times \{\infty\}$, define

$$\underline{l}_{\underline{k}}^* := \left(l_1^* = 0, \underbrace{l_1^*, \dots, l_1^*}_{k_1\text{-times}}, l_2^*, \underbrace{l_2^*, \dots, l_2^*}_{k_2\text{-times}}, l_3^*, \underbrace{l_3^*, \dots, l_3^*}_{k_3\text{-times}}, \dots\right) \tag{9.8}$$

where $\underline{l}^*$ is the minimal subtree length vector in the support of λ from Step 1. Observe that sampling from the subtree length distribution first the point 1 a number of k_1+1 times, then the point 2, then one of the points in $\{1,2\}$ a number of k_2 times, and so on, results exactly in the vector $\underline{l}^*_{\underline{k}}$. Hence, taking all possible permutations $\sigma\in\Gamma$ into account, and since $\lambda(\{\underline{l}^*\}) = |\Gamma|\cdot\prod_{i=1}^N p_i$,

$$\begin{aligned}\lambda(\{\underline{l}^*_{\underline{k}}\}) &= \Big(\prod_{i=1}^N p_i\Big)\cdot\sum_{\sigma\in\Gamma}\prod_{i=1}^{N-1}\Big(\sum_{1\le j\le i} p_{\sigma(j)}\Big)^{k_i}\\ &= \frac{1}{|\Gamma|}\lambda(\{\underline{l}^*\})\cdot\sum_{\sigma\in\Gamma}\prod_{i=1}^{N-1}\Big(\sum_{1\le j\le i} p_{\sigma(j)}\Big)^{k_i}.\end{aligned} \tag{9.9}$$

We claim that (9.9) determines $(p_1,\dots,p_N)$ uniquely up to permutations $\sigma\in\Gamma$.

To see this, observe that the algebra of functions on the $N-1$-dimensional simplex S_N, generated by the functions

$$\Big\{ f((p_1,\dots,p_N)) := \prod_{i=1}^{N-1}\Big(\sum_{1\le j\le i} p_j\Big)^{k_i}\ ;\ k_1,\dots,k_{N-1}\in\mathbb{N}_0\Big\} \tag{9.10}$$

separates points. Hence, $f\in\mathcal{C}_b(S_N)$ can be approximated uniformly by functions in this algebra by the Stone-Weierstrass Theorem. Hence, by knowing $\lambda(\{\underline{l}^*_{\underline{k}}\})/\lambda(\{\underline{l}^*\})$ for all $\underline{k}$, using (9.9), we also know the values of

$$\frac{1}{|\Gamma|}\sum_{\sigma\in\Gamma} f((p_{\sigma(1)},\dots,p_{\sigma(n)})) \tag{9.11}$$

by an approximation argument.

In particular, we can find the set $A := \{(p_{\sigma(1)},\dots,p_{\sigma(N)}) : \sigma\in\Gamma\}$. By setting $\mu\{i\} = p_i$ for an arbitrary $(p_1,\dots,p_N)\in A$ we have recovered μ uniquely up to isometries such that $\xi^{-1}(\{\lambda\}) = \{\mathfrak{u}\}$ by construction.

Step 3 (*General ultra-metric measure spaces*) Let $\mathfrak{u} = \overline{(U,r,\mu)}\in\xi^{-1}(\{\mathfrak{u}_0\})$ not necessarily finite anymore. We shall approximate $\mathfrak{u}$ by finite ultra-metric measure spaces which we then treat as described in the first two steps.

For that purpose, let for all $\varepsilon>0$, the (ε-shrunken) pseudo-metric r_ε on U by putting

$$r_\varepsilon := 0\vee(r-\varepsilon). \tag{9.12}$$

Notice that since $\overline{(\mathrm{supp}(\mu), r)}$ is ultra-pseudo-metric, $\overline{(\mathrm{supp}(\mu), r_\varepsilon)}$ is ultra-pseudo-metric as well, for all ε.

Moreover, for all $\varepsilon > 0$ there is a covering of U of disjunct balls $B_1, B_2, \ldots \subseteq U$ of radius ε with $\mu(B_1) \geq \mu(B_2) \geq \ldots$. Take N_ε large enough such that for $B_\varepsilon = \bigcup_{i=1}^{N_\varepsilon} B_i$ we have $\mu(B_\varepsilon) > 1 - \varepsilon$. Set $\mu_\varepsilon(\cdot) := \mu(\cdot|B_\varepsilon)$ and

$$\mathfrak{u}^\varepsilon := \overline{(U, r_\varepsilon, \mu_\varepsilon)}. \tag{9.13}$$

Then $\mathfrak{u}^\varepsilon$ is a finite metric measure space and $\mathfrak{u}^\varepsilon \to \mathfrak{u}$ in the Gromov-weak topology, as $\varepsilon \to 0$.

Given $u_1, u_2, \ldots \in U$, set $l_n := L_n^{(U,r)}(\{u_1, \ldots, u_n\})$ leading to the subtree length vector $(l_1, l_2, \ldots) \in \mathbb{R}_+^{\mathbb{N}}$. We define the map $\underline{\ell}^\varepsilon : \mathbb{R}_+^{\mathbb{N}} \to \mathbb{R}_+^{\mathbb{N}}$ given by

$$\underline{\ell}^\varepsilon : (l_1, l_2, \ldots) \mapsto (l_1^\varepsilon, l_2^\varepsilon, \ldots) \tag{9.14}$$

with $l_1^\varepsilon = 0$ and $l_2^\varepsilon = 0 \vee (l_2 - \varepsilon)$ and for $n \geq 3$, recursively,

$$l_n^\varepsilon := l_{n-1}^\varepsilon + \big(l_n - l_{n-1} - \tfrac{1}{2}\varepsilon\big)^+. \tag{9.15}$$

Moreover, set

$$A_{\varepsilon,n} := \{(l_1 = 0, l_2, \ldots) : l_i > l_{i-1} \text{ for exactly } N_\varepsilon - 1 \text{ different } i \in \{1, 2, \ldots, n\}\}, \tag{9.16}$$

and we observe that

$$\xi(\mathfrak{u}^\varepsilon)(\cdot) = (\underline{\ell})_* \nu^{\mathfrak{u}^\varepsilon} = \lim_{n\to\infty} (\underline{\ell}_*^\varepsilon \lambda)\big(\,\cdot\,|A_{\varepsilon,n}\big). \tag{9.17}$$

Now, take $\mathfrak{u}, \widetilde{\mathfrak{u}} \in \xi^{-1}(\{\lambda\})$. Observe that $\widetilde{\mathfrak{u}}^\varepsilon \to \widetilde{\mathfrak{u}}$ and $\mathfrak{u}^\varepsilon \to \mathfrak{u}$ in the Gromov-weak topology, as $\varepsilon > 0$. Hence we are in a position to apply Steps 1 and 2 to find that $\widetilde{\mathfrak{u}}^\varepsilon \in \xi^{-1}(\lim_{n\to\infty} \underline{\ell}_*^\varepsilon \lambda(\cdot|A_{\varepsilon,n})) = \{\mathfrak{u}^\varepsilon\}$, for all $\varepsilon > 0$. This shows that $\mathfrak{u} = \lim_{\varepsilon\to 0} \mathfrak{u}^\varepsilon = \lim_{\varepsilon\to 0} \widetilde{\mathfrak{u}}^\varepsilon = \widetilde{\mathfrak{u}}$.

As for *continuity of* ξ, assume that $(\mathfrak{u}_k)_{k\in\mathbb{N}}$ is a sequence in $\mathbb{U}$ such that $\mathfrak{u}_k \to \mathfrak{u}$, for some $\mathfrak{u} \in \mathbb{U}$, in the Gromov-weak topology, as $k \to \infty$. Then by definition, $\Phi(\mathfrak{u}_k) \to \Phi(\mathfrak{u})$, for all $\Phi \in \Pi^0$, as $k \to \infty$. In particular, since the map $\underline{\underline{r}} \mapsto \ell_n(\underline{\underline{r}})$ is continuous as it is the minimum of finitely many continuous functions, for all $n \in \mathbb{N}$, $\langle \xi(\mathfrak{u}_k), \psi \rangle \to \langle \xi(\mathfrak{u}), \psi \rangle$, for all $\psi \in \mathcal{C}_b(\mathbb{R}_+^{\mathbb{N}})$, or equivalently, $\xi(\mathfrak{u}_k) \Rightarrow \xi(\mathfrak{u})$ in the weak topology on $\mathcal{M}_1(\mathbb{R}_+^{\mathbb{N}})$, as $k \to \infty$.

In order to show *continuity of* ξ^{-1}, we take λ, λ_1, λ_2,…in $\xi(\mathbb{U})$ such that $\lambda_m \Rightarrow \lambda$, as $m \to \infty$. We have to show that $\mathfrak{u}_m := \xi^{-1}(\lambda_m) \to \xi^{-1}(\lambda) =: \mathfrak{u}$ in the Gromov-weak topology, as $m \to \infty$. For this, we need to show that the three steps in the proof of injectivity of ξ hold under weak limits.

For Steps 1 and 2, assume that $\mathfrak{u}$ is finite with $\#\text{supp}(\mu) = N$. Then the same holds for all large $m \in \mathbb{N}$. Define for all $m \in \mathbb{N}$ (based on $\mathfrak{u}^m$) sets $B^m \subseteq \mathbb{R}_+^{\mathbb{N}}$, minimal elements $\underline{l}^{*,m} \in B^m$, $(d_n^{*,m})_{n\geq 1}$ and $\underline{\underline{r}}^m$ as in (9.2), (9.3), (9.5) and (9.7), respectively. Then we can clearly recover that the mutual distances $\underline{\underline{r}}$ in $\mathfrak{u}$ as the limit of $\underline{\underline{r}}^m$, as $m \to \infty$. Moreover, note that the set of functions (9.10) is not only separating, but also convergence determining. Hence since all metric measure spaces are finite, we find that $u_m \to u$.

For the general case considered in Step 3, recall the notions u^ε, $\underline{\ell}^\varepsilon$ and $A_{\varepsilon,n}$ from (9.13), (9.14) and (9.16). Note then that $u_m \to u$ as $m \to \infty$ if and only if $u_m^\varepsilon \to u^\varepsilon$ as $m \to \infty$ for all $\varepsilon > 0$. Moreover, for all $\varepsilon > 0$,

$$u_m^\varepsilon = \underline{\ell}_* \nu^{u_m^\varepsilon}(\cdot) = \lim_{n\to\infty} \underline{\ell}_*^\varepsilon \lambda_m(\cdot|A_{\varepsilon,n}) \to \lim_{n\to\infty} \underline{\ell}_*^\varepsilon \lambda(\cdot|A_{\varepsilon,n}) = u^\varepsilon. \tag{9.18}$$

The interchange of limits is justified, because $\underline{\ell}_*^\varepsilon \lambda_m(\cdot|A_{\varepsilon,n}) \Rightarrow \underline{\ell}_*^\varepsilon \lambda(\cdot|A_{\varepsilon,n})$ as $m \to \infty$, if n is large enough, and we have shown continuity of ξ^{-1}. □

Proof of Theorem 5 (i) Since ξ is bijective on $\xi(\mathbb{U})$, it is a consequence of Theorem 3.2 in [36] that the martingale problem for $(\xi_*\mathbf{P}(\mathbb{U}), \Omega^{\uparrow,\Xi}, \Pi^{1,\Xi})$ is well-posed. Moreover, by construction, $(\xi(\mathcal{U}_t))_{t\geq 0}$ solves the martingale problem. In addition, since $\mathcal{U}$ has the Feller property and ξ and also ξ^{-1} (see Theorem 4) are continuous, Ξ is Feller, too. The last assertion follows from the continuity of the sample paths of the tree-valued Fleming–Viot dynamics and the continuity of ξ.

(ii) With $\underline{\ell}$ from (3.3),

$$\begin{aligned}
&\Omega^{\uparrow}(\Psi \circ \xi)\big(u\big)\\
&= \langle \nu^u, \langle \nabla\psi \circ \underline{\ell}, \underline{\underline{2}}\rangle\rangle + \gamma \sum_{1\leq k<l} \langle \nu^u, \psi \circ \underline{\ell} \circ \theta_{k,l} - \psi \circ \underline{\ell}\big\rangle\\
&= \sum_{n\geq 2} n\langle \nu^u, \frac{\partial}{\partial \ell_n}(\psi \circ \underline{\ell})\rangle + \gamma \sum_{n\geq 2}(n-1)\langle \nu^u, \psi \circ \beta_{n-1} \circ \underline{\ell} - \psi \circ \underline{\ell}\rangle\\
&= \sum_{n\geq 2} n\langle \xi(u), \frac{\partial}{\partial \ell_n}\psi\rangle + \gamma \sum_{n\geq 1} n\langle \xi(u), \psi \circ \beta_n - \psi\big\rangle
\end{aligned} \tag{9.19}$$

and we are done. □

To prepare the proof of Corollary 3.4 we investigate for each time $t \geq 0$ the *mean sample Laplace transform*,

$$g(t; \underline{\sigma}) := \mathbf{E}\big[\Psi^{\underline{\sigma}}(\Xi_t)\big], \tag{9.20}$$

of the subtree lengths distribution Ξ_t, where for $\underline{\sigma} \in \mathbb{R}_+^{\mathbb{N}}$,

$$\Psi^{\underline{\sigma}}(\Xi) := \int_{\mathbb{R}_+^{\mathbb{N}}} \Xi(\mathrm{d}\underline{l})\psi^{\underline{\sigma}}(\underline{l}) \tag{9.21}$$

with the test function

$$\psi^{\underline{\sigma}}(\underline{l}) := \exp(-\langle \underline{\sigma}, \underline{l}\rangle). \tag{9.22}$$

As usual, $\langle \cdot, \cdot \rangle$ denotes the inner product.

Lemma 9.1 (ODE system for the mean sample Laplace transforms) *For* $\underline{\sigma} \in \mathbb{R}_+^{\mathbb{N}}$ *having only finitely many non-zero entries, the functions* $g(.;\underline{\sigma})$ *satisfy the following system of differential equations:*

$$\frac{\mathrm{d}}{\mathrm{d}t} g(t;\underline{\sigma}) = -\left(\sum_{k=2}^{\infty} k\sigma_k\right) g(t;\underline{\sigma}) + \gamma \sum_{k=1}^{\infty} k\big(g(t;\tau_k\underline{\sigma}) - g(t;\underline{\sigma})\big) \tag{9.23}$$

with the merging operator

$$\tau_k : (\sigma_1, \ldots, \sigma_{k-1}, \sigma_k, \sigma_{k+1}, \sigma_{k+2}, \ldots) \mapsto (\sigma_1, \ldots, \sigma_{k-1}, \sigma_k + \sigma_{k+1}, \sigma_{k+2}, \ldots). \tag{9.24}$$

Proof By standard arguments, $\Psi^{\underline{\sigma}} \in \Pi^{\Xi}$ and

$$\frac{\mathrm{d}}{\mathrm{d}t} g(t;\underline{\sigma}) = \mathbb{E}\big[\Omega^{\uparrow,\Xi}\Psi^{\underline{\sigma}}(\Xi_t)\big]. \tag{9.25}$$

Hence, inserting (9.19), and using $\psi^{\underline{\sigma}}(\beta_k \underline{l}) = \psi^{\tau_k\underline{\sigma}}(\underline{l})$ for all $k = 1, 2, \ldots$, with β_k from (3.11) and τ_k from (9.24), we find

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t} g(t,\underline{\sigma}) \\
&= \mathbf{E}\left[-\int \Xi_t(\mathrm{d}\underline{l}) \sum_{k=2}^{\infty} k\sigma_k \psi^{\underline{\sigma}}(\underline{l}) + \gamma \int \Xi_t(\mathrm{d}\underline{l}) \sum_{k=1}^{\infty} k\big(\psi^{\underline{\sigma}}(\beta_k \underline{l}) - \psi^{\underline{\sigma}}(\underline{l})\big)\right] \\
&= -\left(\sum_{k=2}^{\infty} k\sigma_k\right) g(t,\underline{\sigma}) + \gamma \sum_{k=1}^{\infty} k\big(g(t,\tau_k\underline{\sigma}) - g(t,\underline{\sigma})\big),
\end{aligned} \tag{9.26}$$

as claimed. □

Remark 9.2 Recall, for each $n \in \mathbb{N}$, the function g^n from (3.13). For each $n \geq 2$ and $\sigma \geq 0$, applying (9.23) to $\underline{\sigma} = (\sigma\delta_{k,n})_{k\geq 2}$ yields, setting $g^1(t;\sigma) := 1$,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} g^n(t;\sigma) &= -n\sigma g^n(t;\sigma) + \gamma\binom{n}{2}\big(g^{n-1}(t;\sigma) - g^n(t;\sigma)\big) \\
&= \frac{\gamma}{2} n(n-1) g^{n-1}(t;\sigma) - \frac{\gamma}{2} n\big(\tfrac{2}{\gamma}\sigma + n - 1\big) g^n(t;\sigma),
\end{aligned} \tag{9.27}$$

i.e.,

$$\frac{\mathrm{d}}{\mathrm{d}t}\big(g^2(t;\sigma), g^3(t;\sigma), \ldots\big) = \frac{\gamma}{2}\Big[A\big(\tfrac{2}{\gamma}\sigma\big)\big(g^2(t;\sigma), g^3(t;\sigma), \ldots\big)^{\top} + b^{\top}\Big], \tag{9.28}$$

where

$$b^{\top} := (2, 0, \ldots)^{\top} \tag{9.29}$$

and for $\widetilde{\sigma} \geq 0$ the matrix $A := A(\widetilde{\sigma})$ is defined by

$$A_{k,l} := \begin{cases} k(k-1), & \text{if } k = l+1, \\ -k(\widetilde{\sigma}+k-1), & \text{if } k = l, \\ 0, & \text{else,} \end{cases} \tag{9.30}$$

for all $k, l \geq 2$. □

The proof of Corollary 3.4 uses the following preparatory lemma.

Lemma 9.3 *Fix $\widetilde{\sigma} \geq 0$. Let $B = (B_{k,l})_{k,l\geq 2}$ and $B^{-1} = (B^{-1}_{k,l})_{k,l\geq 2}$ be matrices defined by*

$$B_{k,l} := \frac{\frac{k!}{l!}\binom{k-1}{l-1}\Gamma(\widetilde{\sigma}+2l)}{\Gamma(\widetilde{\sigma}+k+l)}, \quad \text{and} \quad B^{-1}_{k,l} = \frac{(-1)^{k+l}\frac{k!}{l!}\binom{k-1}{l-1}\Gamma(\widetilde{\sigma}+k+l-1)}{\Gamma(\widetilde{\sigma}+2k-1)}. \tag{9.31}$$

(i) *The matrices B and B^{-1} are inverse to each other.*
(ii) *The matrix $A = A(\widetilde{\sigma}) = (A_{k,l})_{k,l\geq 2}$ has eigenvalues*

$$\lambda_k := -k(\widetilde{\sigma}+k-1), \qquad k \geq 2. \tag{9.32}$$

(iii) *If $D = (\lambda_k \delta_{k,l})_{k,l\geq 2}$ then*

$$f(A) = Bf(D)B^{-1} \tag{9.33}$$

for all analytical functions $f : \mathbb{R}^{\mathbb{N}^2} \to \mathbb{R}^{\mathbb{N}^2}$. Specifically, $A^{-1} = BD^{-1}B^{-1}$ and $e^{At} = Be^{Dt}B^{-1}$ for all $t \geq 0$.
(iv) *For $\widetilde{\sigma} > 0$, let $A^{-1}(\widetilde{\sigma}) = (A^{-1}_{k,l})_{k,l\geq 2}$ be given by $A^{-1}_{k,l} = 0$ for $k < l$ and*

$$A^{-1}_{k,l} := -\frac{(k-1)!\Gamma(\widetilde{\sigma}+l-1)}{l!\Gamma(\widetilde{\sigma}+k)}, \qquad k \geq l. \tag{9.34}$$

Then A^{-1} and A are inverse to each other.

Proof First, we note that A, A^{-1}, B, B^{-1} are lower triangular infinite matrices. This implies that the domain of the maps induced by these matrices is $\mathbb{R}^{\mathbb{N}}$. In particular, we do not have to distinguish between left- and right inverse matrices of A and B.

(i) We need to show that

$$(B \cdot B^{-1})_{k,l} = \delta_{k,l} \tag{9.35}$$

for $k \geq l \geq 2$. This is clear in the case where $k \leq l$. For $k > l \geq 2$, with constants C changing from line to line, and using the abbreviations $\widehat{k} := k-l$ and $\widehat{\sigma} := \widetilde{\sigma}+2l-1$,

$$
\begin{aligned}
(B\cdot B^{-1})_{k,l} &= \sum_{m=l}^{k} B_{k,m}B^{-1}_{m,l}\\
&= \sum_{m=l}^{k} \frac{\frac{k!}{m!}\binom{k-1}{m-1}\Gamma(\widetilde{\sigma}+2m)}{\Gamma(\widetilde{\sigma}+k+m)}\cdot\frac{(-1)^{m+l}\frac{m!}{l!}\binom{m-1}{l-1}\Gamma(\widetilde{\sigma}+m+l-1)}{\Gamma(\widetilde{\sigma}+2m-1)}\\
&= C\sum_{m=l}^{k}(-1)^{m+l}\frac{(\widetilde{\sigma}+2m-1)\Gamma(\widetilde{\sigma}+m+l-1)}{(k-m)!(m-l)!\Gamma(\widetilde{\sigma}+k+m)}\\
&= C\sum_{m=0}^{\widehat{k}}(-1)^{m}\frac{(\widehat{\sigma}+2m)\Gamma(\widehat{\sigma}+m)}{\Gamma(\widehat{k}-m+1)\Gamma(m+1)\Gamma(\widehat{\sigma}+\widehat{k}+m+1)}\\
&= C\sum_{m=0}^{\widehat{k}}(-1)^{m}\frac{(\widehat{\sigma}+2m)\Gamma(\widehat{\sigma}+m)}{\Gamma(m+1)}\cdot\frac{\Gamma(\widehat{\sigma}+2\widehat{k}+1)}{\Gamma(\widehat{\sigma}+\widehat{k}+m+1)\Gamma(\widehat{k}-m+1)}\\
&= 0,
\end{aligned}
\tag{9.36}
$$

where we have used that

$$
C\cdot\frac{(\widehat{\sigma}+2m)\Gamma(\widehat{\sigma}+m)}{\Gamma(m+1)} = \frac{\Gamma(\widehat{\sigma}+m+1)}{\Gamma(m+1)\Gamma(\widehat{\sigma}+1)} + \frac{\Gamma(\widehat{\sigma}+m)}{\Gamma(m)\Gamma(\widehat{\sigma}+1)} \tag{9.37}
$$

and then applied Formula (5d) on page 10 in [42].

(ii) Since A is lower triangular, this is obvious.

(iii) Note that

$$
(A\cdot B)_{2,l} - \lambda_l B_{2,l} = 0 \tag{9.38}
$$

and

$$
\begin{aligned}
\lambda_l - \lambda_k &= \widetilde{\sigma}(k-l) + (k^2-k-l^2+l)\\
&= (k-l)\big(\widetilde{\sigma}+k+l-1\big).
\end{aligned}
\tag{9.39}
$$

Thus for all $k\ge 3$ and $l\ge 2$,

$$
B_{k,l} = \frac{k(k-1)}{(k-l)(\widetilde{\sigma}+k+l-1)}B_{k-1,l}, \tag{9.40}
$$

and since $A_{k,k}=\lambda_k$,

$$
\begin{aligned}
\big(A\cdot B\big)_{k,l} - \lambda_l B_{k,l} &= A_{k,k-1}B_{k-1,l} + (\lambda_k-\lambda_l)B_{k,l}\\
&= \big(k(k-1)-k(k-1)\big)B_{k-1,l}\\
&= 0,
\end{aligned}
\tag{9.41}
$$

which proves that B contains all eigenvectors of A. Hence the claim follows by standard linear algebra.

(iv) It is clear that $(A \cdot A^{-1})_{k,k} = 1$, while for $k \neq l$,

$$\begin{aligned}(A \cdot A^{-1})_{k,l} &= A_{k,k-1} \cdot A^{-1}_{k-1,l} + A_{k,k} \cdot A^{-1}_{k,l} \\ &= k(k-1)\frac{(k-2)!\Gamma(\widetilde{\sigma}+l-1)}{l!\Gamma(\widetilde{\sigma}+k-1)} - k(\widetilde{\sigma}+k-1)\frac{(k-1)!\Gamma(\widetilde{\sigma}+l-1)}{l!\Gamma(\widetilde{\sigma}+k)} \\ &= 0. \end{aligned} \tag{9.42}$$

□

Proof of Corollary 3.4 Fix $n \in \mathbb{N}$ and $\sigma \geq 0$. Put

$$h^{\sigma,n}(t) := g^n\big(\tfrac{2t}{\gamma}; \sigma\big). \tag{9.43}$$

By (9.27), the vector $\underline{h} := (h^{\sigma,2}, h^{\sigma,3}, \ldots)^\top$ satisfies the linear system of ordinary differential equations

$$\frac{\mathrm{d}}{\mathrm{d}t}\underline{h} = A\underline{h} + b, \tag{9.44}$$

or equivalently,

$$h(t) = e^{At}h(0) + e^{At}A^{-1}b - A^{-1}b, \tag{9.45}$$

with $b = (2, 0, 0, \ldots)^\top$ and $A = (A_{k,l})_{k,l\geq 2}$ as defined in (9.30). Consequently, if B, B^{-1} and D are as in Lemma 9.3, then

$$h(t) = -A^{-1}b + Be^{Dt}\big(B^{-1}h(0) + D^{-1}B^{-1}b\big). \tag{9.46}$$

Combining (9.43) with (9.46) yield the explicit expressions given in (3.14).

Finally, by (9.46),

$$\begin{aligned} g^n\big(t;\sigma\big) &\underset{t\to\infty}{\longrightarrow} -2\Big(A\big(\tfrac{2}{\gamma}\sigma\big)^{-1}\Big)_{n,2} \\ &= \frac{\Gamma(n)\Gamma(\widetilde{\sigma}+1)}{\Gamma(n+\widetilde{\sigma})} \\ &= \mathbf{E}\big[e^{-\sigma\sum_{k=2}^{n}\mathcal{E}^k}\big]. \end{aligned} \tag{9.47}$$

□

Acknowledgments We thank David Aldous, Steve Evans, Patric Glöde, Pleuni Pennings and Sven Piotrowiak for helpful discussions. We are particularly grateful to Steve Evans who provided the key argument in the proof of Theorem 4.

References

1. Aldous, D.: The continuum random tree. II: An overview. In: Stochastic Analysis. Proc. Symp., Durham/UK 1990. Lond. Math. Soc. Lect. Note Ser., vol. 167, pp. 23–70 (1991)
2. Aldous, D.: The continuum random tree III. Ann. Probab. **21**(1), 248–289 (1993)
3. Birkner, M., Blath, J., Möhle, M., Steinrücken, M., Tams, J.: A modified lookdown construction for the xi-fleming-viot process with mutation and populations with recurrent bottlenecks. ALEA **6**, 25–61 (2009)
4. Bertoin, J., Le Gall, J.-F.: Stochastic flows associated to coalescent processes. Probab. Theory Relat. Fields **126**, 261–288 (2003)
5. Dawson, D.A.: Measure-valued Markov processes. In: Hennequin, P.L. (ed.) École d'Été de Probabilités de Saint-Flour XXI–1991. Lecture Notes in Mathematics, vol. 1541, pp. 1–260. Springer, Berlin (1993)
6. Delmas, J.-F., Dhersin, J.-S., Siri-Jegousse, A.: On the two oldest families for the Wright-Fisher process. Electron. J. Probab. **15**, 776–800 (2010)
7. Depperschmidt, A., Greven, A., Pfaffelhuber, P.: Tree-valued fleming-viot dynamics with mutation and selection. Ann. Appl. Probab. (2011, to appear)
8. Dawson, D.A., Greven, A., Vaillancourt, J.: Equilibria and quasi-equilibria for infinite systems of Fleming-Viot processes. Trans. Mem. Am. Math. Soc. **347**(7), 2277–2360 (1995)
9. Donnelly, P., Kurtz, T.G.: A countable representation of the Fleming-Viot processes. Ann. Probab. **24**(2), 698–742 (1996)
10. Donnelly, P., Kurtz, T.G.: Genealogical processes for Fleming-Viot models with selection and recombination. Ann. Appl. Probab. **9**, 1091–1148 (1999)
11. Donnelly, P., Kurtz, T.G.: Particle representation for measure-valued population models. Ann. Probab. **27**(1), 166–205 (1999)
12. Duquesne, T., Le Gall, J.-F.: Random trees, Lévy processes and spatial branching processes. Astérisque **281** (2002)
13. Dress, A., Moulton, V., Terhalle, W.: T-theory. Eur. J. Combin. **17**, 161–175 (1996)
14. Dawson, D.A., Perkins, E.A.: Historical processes. Mem. Am. Math. Soc. **93**(454) (1991)
15. Dress, A.: Trees, tight extensions of metric spaces, and the cohomological dimension of certain groups: a note on combinatorical properties of metric spaces. Adv. Math. **53**, 321–402 (1984)
16. Ethier, S.N., Kurtz, T.G.: Markov Processes. Characterization and Convergence. Wiley, New York (1986)
17. Ethier, S.N., Kurtz, T.G.: Fleming-Viot processes in population genetics. SIAM J. Contr. Optim. **31**, 345–386 (1993)
18. Evans, S.N., Lidman, T.: Asymptotic evolution of acyclic random mappings. Electron. J. Probab. **12**(42), 1151–1180 (2008)
19. Evans, S.N., O'Connell, N.: Weighted occupation time for branching particle systems and a representation for the supercritical superprocess. Can. Math. Bull. **37**(2), 187–196 (1994)
20. Evans, S.N., Pitman, J., Winter, A.: Rayleigh processes, real trees, and root growth with re-grafting. Probab. Theory Relat. Fields **134**(1), 81–126 (2006)
21. Evans, S.N., Ralph, P.L.: Dynamics of the time to the most recent common ancestor in a large branching population. Ann. Appl. Probab. **20**(1), 1–25 (2010)
22. Etheridge, A.: An Introduction to Superprocesses. American Mathematical Society, New York (2001)
23. Evans, S.: Kingman's coalescent as a random metric space. In: Gorostiza, L.G., B.G. Ivanoff, B.G. (eds.) Stochastic Models: Proceedings of the International Conference on Stochastic Models in Honour of Professor Donald A. Dawson, Ottawa, Canada, June 10–13, 1998. Canad. Math. Soc. (2000)
24. Evans, S.N.: Probability and real trees. In: Picard, J. (ed.) École d'Été de Probabilités de Saint-Flour XXXV–2005. Lecture Notes in Mathematics, vol. 1920, pp. 1–193. Springer, Berlin (2007)
25. Evans, S.N., Winter, A.: Subtree prune and re-graft: a reversible real-tree valued Markov chain. Ann. Probab. **34**(3), 918–961 (2006)
26. Fukushima, M., Stroock, D.: Reversibility of solutions to martingale problems. Adv. Math. (Supp. Studies) **9**, 107–123 (1986)
27. Fleming, W.H., Viot, M.: Some measure-valued population processes. In: Stochastic Analysis (Proc. Internat. Conf., Northwestern Univ., Evanston, Ill., 1978), pp. 97–108. Academic Press, New York (1978)

28. Fleming, W.H., Viot, M.: Some measure-valued Markov processes in population genetics theory. Indiana Univ. Math. J. **28**(5), 817–843 (1979)
29. Greven, A., Limic, V., Winter, A.: Representation theorems for interacting Moran models, interacting Fisher-Wright diffusions and applications. Electron. J. Probab. **10**(39), 1286–1358 (2005)
30. Greven, A., Pfaffelhuber, P., Winter, A.: Convergence in distribution of random metric measure spaces (The Λ-coalescent measure tree). Probab. Theory Relat. Fields **145**, 285–322 (2009)
31. Greven, A., Popovic, L., Winter, A.: Genealogy of catalytic branching models. Ann. Appl. Probab. **19**(3), 1232–1272 (2009)
32. Gromov, M.: Metric structures for Riemannian and non-Riemannian spaces. Progress in Mathematics, vol. 152. Birkhäuser, Boston (1999)
33. Kallenberg, O.: Stability of critical cluster fields. Math. Nachr. **77**, 7–43 (1977)
34. Kingman, J.F.C.: The coalescent. Stoch. Process. Appl. **13**(3), 235–248 (1982)
35. Kingman, J.F.C.: Exchangeability and the evolution of large populations. In: Proceedings of the International Conference on Exchangeability in Probability and Statistics, Rome, 6–9 April, 1981, in honour of Professor Bruno de Finetti, pp. 97–112. North-Holland, Elsevier, Amsterdam (1982)
36. Kurtz, T.G.: Martingale problems for conditional distributions of Markov processes. Electron. J. Probab. **3**(9), 1–29 (1998)
37. Le Gall, J.-F.: Spatial branching processes, random snakes and partial differential equations. Lectures in Mathematics ETH Zürich. Birkhäuser, Basel (1999)
38. Möhle, M., Sagitov, S.: A classification of coalescent processes for haploid exchangeable models. Ann. Prob. **29**, 1547–1562 (2001)
39. Pitman, J.: Coalescents with multiple collisions. Ann. Probab. **27**(4), 1870–1902 (1999)
40. Pfaffelhuber, P., Wakolbinger, A.: The process of most recent common ancestors in an evolving coalescent. Stoch. Process. Appl. **116**, 1836–1859 (2006)
41. Pfaffelhuber, P., Wakolbinger, A., Weisshaupt, H.: The tree length of an evolving coalescent. Probab. Theory Relat. Fields **151**, 529–557 (2011)
42. Riordan, J.: Combinatorial Identities. Wiley, New York-London-Sydney (1968)
43. Terhalle, W.F.: R-trees and symmetric differences of sets. Eur. J. Combin. **18**, 825–833 (1997)
44. Watterson, G.A.: On the number of segregating sites in genetical models without recombination. Theory Popul. Biol. **7**, 256–276 (1975)
45. Zambotti, L.: A reflected stochastic heat equation as symmetric dynamics with respect to the 3-d Bessel bridge. J. Funct. Anal. **180**(1), 195–209 (2001)
46. Zambotti, L.: Integration by parts on Bessel bridges and related SPDEs. C. R. Math. Acad. Sci. Paris **334**(3), 209–212 (2002)
47. Zambotti, L.: Integration by parts on δ-Bessel bridges, $\delta > 3$ and related SPDEs. Ann. Probab. **31**(1), 323–348 (2003)