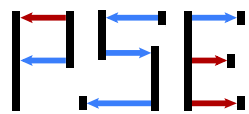
Modeling evolving phylogenies by means of marked metric measure spaces

Sandra Kliem

University Duisburg-Essen



PROBABILISTIC STRUCTURES
IN EVOLUTION

DFG SPP 1590

This is joint work with Anita Winter

Singapore - IMS - August 7, 2017
Workshop on Genealogies of Interacting Particle Systems

Introduction

PopGen Motivation: Phylogenetic pathogen patterns within one host and within the host population.

Goal: model pathogen phylogenies and their evolution in time.

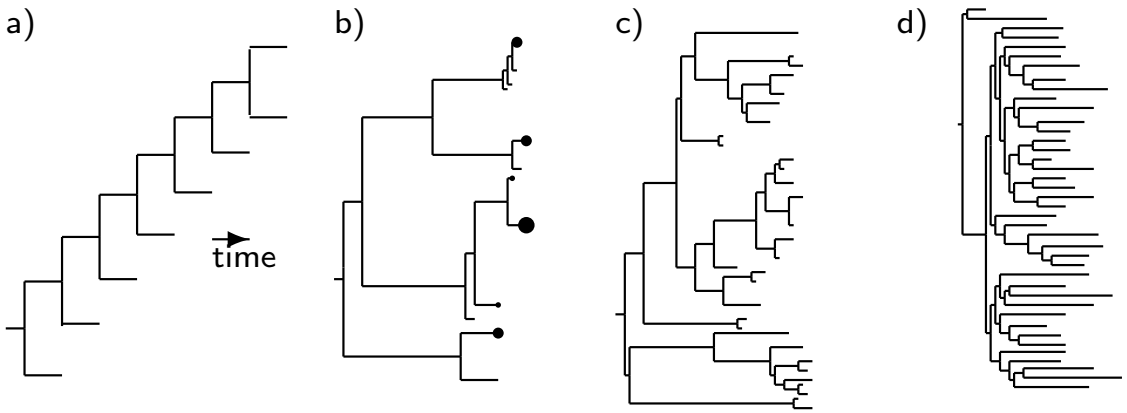
1. high mutation and replication rates cause viral variability;
2. the *genetic variation* is further affected by
   - the strength of *cross-immunity* (= the ability of a host's immune system to fight a certain strain or related strains of a virus),
   - transmission (susceptibility-infection-recovery-times),
   - the size of the "risk group" in the population which can be infected,
   - effects of migration etc.

Resulting Model:

1. a Branching model with Selection, Mutation and Competition.
2. Consider virus which *evolves very fast*, i.e. regime of
   - large population sizes,
   - high mutation rates and
   - short generation times.

References: [GPGWDMH04] and [LoH07].

First attempt at a Classification



Branch lengths are according to the expected number of substitutions.

(a) one dominating strain,
Influenza A on population level; HIV over time on host level

(b) a bounded number of coexisting strains,
Serotypes of Dengue-virus on population level

(c) an unbounded number of coexisting strains with proper frequencies,
Measles on population level

(d) an unbounded number of coexisting strains without proper frequencies.
HIV or HCV on population level

Goal

Universal Model to

1. model pathogen phylogenies and their evolution in time;
2. in particular, model mechanisms such as transmission (susceptibility-infection-recovery-times), influence of cross-immunity and effects of migration by means of evolution rates.
3. give conditions on the evolution rates to decide which class a given phylogeny belongs to
4. and thereby establish a link between the above mentioned mechanisms and classes;
5. present statistics (functions of sample / set of data) which allow for a classification.

Approximating (discrete) Particle Systems

In the N^{th}-approximation step:

Initial population:

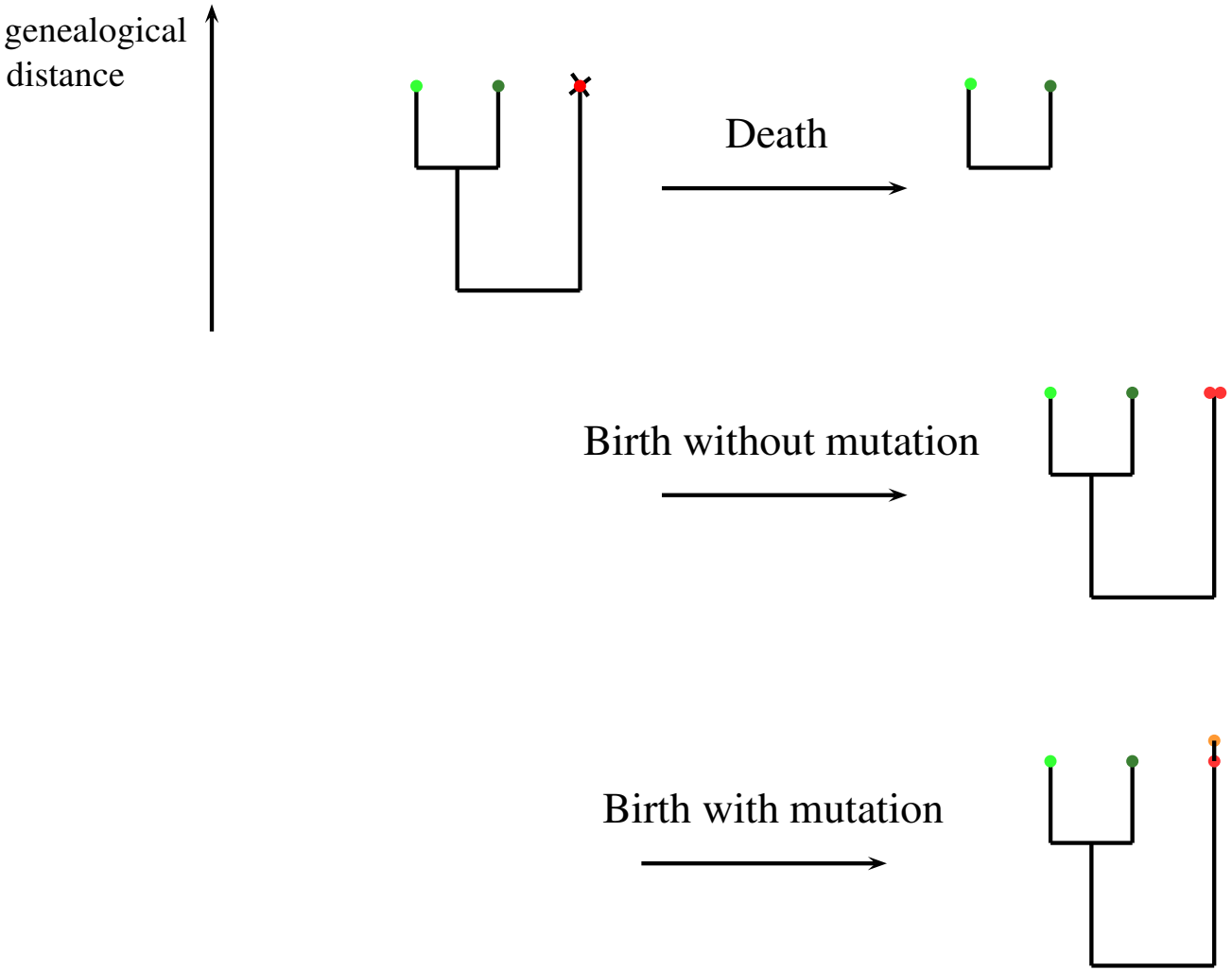
- Each individual has mass $\frac{1}{N}$,
- the overall population has mass $m_0^N \sim m_0$
- $\Rightarrow$ the overall number of individuals alive at time 0 is thus $N \cdot m_0^N \sim N \cdot m_0$.
- Assumption: The initial population, encoded as a marked metric measure space $\mathcal{X}_0^N \in \mathbb{M}^K$, converges weakly to $\mathcal{X}_0$ in $\mathbb{M}^K$ with $\mathbb{E}[(m_0)^3] < \infty$.

Evolution over time:

- consider an asexual population;
- At time t, **traits/types** and **mutual genealogical distances** of individuals are recorded.
- They remain constant during an individual's life.
- Individuals die or give birth to one new individual at a random point in time.
- Death- and birth-rates depend on the traits of the parent as well as the traits of and genealogical distances to other individuals.

- At birth, **mutation** occurs with constant probability $p > 0$ independent of N.
  - ▷ *no mutation* with probability $1 - p$:
    1. child has trait of its parent,
    2. genealogical distance to parent is 0,
    3. genealogical distance to other individuals as for parent ($\Rightarrow$ they are "**clones**" and are part of one "**clan**").
  - ▷ *mutation* with probability p:
    1. child has trait according to a transition matrix (average distance to trait of its parent is of order $1/N$),
    2. genealogical distance to parent is $1/N$,
    3. genealogical distance to other individuals is: as for parent $+1/N$ ($\Rightarrow$ the child constitutes a new clan).

Types of Branching



Previous work: [MT12] (Méléard and Tran (2012)) consider historical branching processes, where rates depend on time and the whole history of the traits over time.

- instead of full history up to time t, use genealogical distances and traits at time t,
- generalize trait space and mutation generator,
- add competition in birth-term.

Marked metric measure spaces

State space for evolving genealogies: marked metric measure spaces (*References:* [DGP11] (Depperschmidt, Greven and Pfaffelhuber (2011)), [GPW13] (Greven, Pfaffelhuber and Winter (2013)))

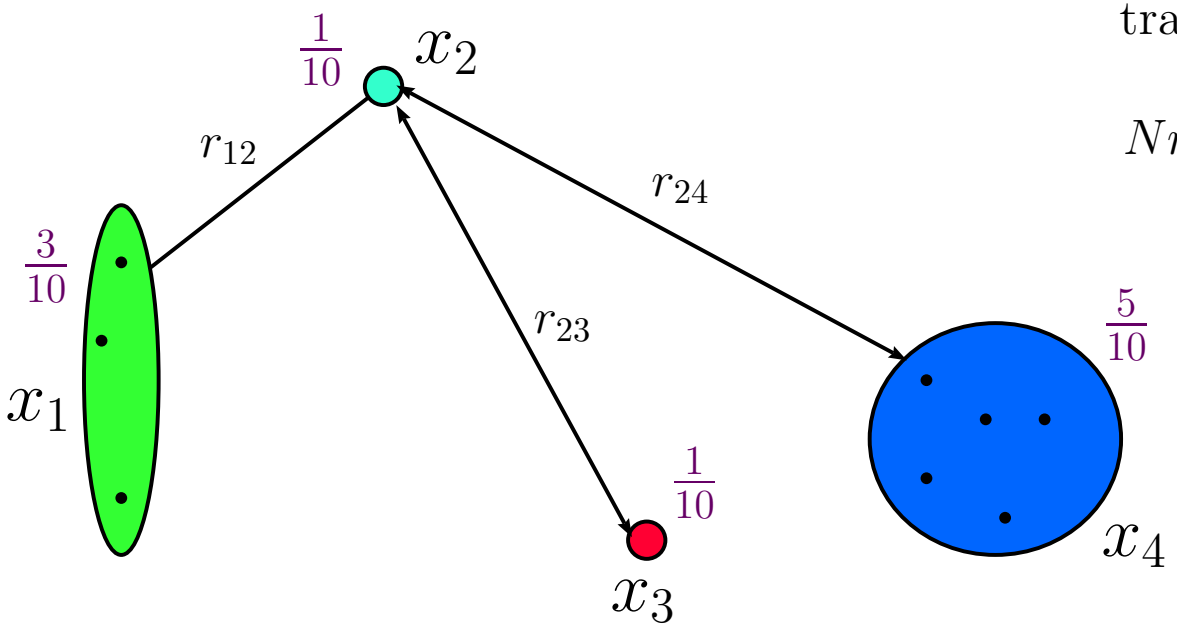
Definition

1. *Let K denote the type/trait space.*
   We assume that K is a complete and separable metric space.
2. *A K-marked metric measure space, or mmm-space, can be written (X, r, μ), where*
   - *(X, r) is a complete and separable metric space, (clans and mutual distances)*
   - *$\mu \in \mathcal{M}_f(X \times K)$,*
   - *$m := \mu(X \times K) \in \mathbb{R}_+$ is the mass of the population (each individual has a certain biomass) and hence*
   - $$\bar{\mu} := \begin{cases} \mu/m, & m \neq 0, \\ \text{arbitrary in } \mathcal{M}_1(X \times K), & m = 0 \end{cases}$$
     samples elements of X and their traits.
   - $\mathbb{M}^K := \{\mathfrak{X} = \overline{(X, r, \mu)} : (X, r, \mu)$ *K-marked metric measure space*$\}$.

$X =$ set of clans $= \{x_1, x_2, x_3, x_4\}$

traits of clans $= \{\kappa_1, \kappa_2, \kappa_3, \kappa_4\}$

$Nm = \#$ individuals $= 10$



$$\bar{\mu} = \frac{3}{10} \cdot \delta_{(x_1,\kappa_1)} + \frac{1}{10} \cdot \delta_{(x_2,\kappa_2)} + \frac{1}{10} \cdot \delta_{(x_3,\kappa_3)} + \frac{5}{10} \cdot \delta_{(x_4,\kappa_4)}$$

Equivalence of mmm-spaces

Definition

$(X, r_X, \mu_X), (Y, r_Y, \mu_Y) \in \mathbb{M}^K$ *are equivalent if*
they are measure- and mark-preserving isometric, i.e.
there is a measurable $\varphi : supp((\pi_X)_*\mu_X) \to supp((\pi_Y)_*\mu_Y)$ *such that*

$$r_X(x, x') = r_Y(\varphi(x), \varphi(x')) \text{ for all } x, x' \in supp((\pi_X)_*\mu_X)$$

$$\text{and} \qquad \tilde{\varphi}_*\mu_X = \mu_Y \text{ for } \tilde{\varphi}(x, u) = (\varphi(x), u).$$

We denote the equivalence class of (X, r, μ) *by* $\overline{(X, r, \mu)}$.

Remark (Nice consequences)

1. *If a particle* z *of type* κ_z *and weight* $1/N$ *dies at time* t*:*
   keep $X_t \equiv X_{t-}, \underline{r}_t \equiv \underline{r}_{t-}$ *but change* $\mu_t \equiv \mu_{t-} - \frac{1}{N}\delta_{(z,\kappa_z)}$.
2. *If particle* x *of type* κ_x *gives birth to a particle* z *of type* κ_z*:*
   - *mutation:* $X_t \equiv X_{t-} \uplus \{z\}$ *and* $\mu_t \equiv \mu_{t-} + \frac{1}{N}\delta_{(z,\kappa_z)}$,
   - *no mutation:* $X_t \equiv X_{t-} \uplus \{z\}$ *and* $\mu_t \equiv \mu_{t-} + \frac{1}{N}\delta_{(x,\kappa_x)}$.

Reproduction and Death:

At time t an individual of clan x and trait $\kappa(x)$ gives birth at rate

$$N\beta(\kappa(x)) + \sum_{z \in X} \bar{\mu}(\{z\} \times K) \cdot \gamma^{\text{birth}}(m, r(z,x), \kappa(z), \kappa(x))$$

$\beta(\cdot) \rightsquigarrow$ *branching rate of an individual of a clan*
is a function of the trait of its clan x

$\gamma^{\text{birth}}(\cdot,\cdot,\cdot,\cdot) \rightsquigarrow$ a function of the overall mass, the genetic distance of clan z and x, the trait of clan z and the trait of clan x

$\sum_{z \in X} \bar{\mu}(\{z\} \times K) \rightsquigarrow$ weighted average over all clans; weights according to number of individuals per clan; $(\bar{\mu}(X \times K) = 1)$

and at time t an individual of clan x and trait $\kappa(x)$ dies at rate

$$N\beta(\kappa(x)) + \sum_{z \in X} \bar{\mu}(\{z\} \times K) \cdot \gamma^{\text{death}}(m, r(z,x), \kappa(z), \kappa(x)).$$

Assumptions include: $0 < \underline{\beta} \leq \beta(\kappa) \leq \overline{\beta}$,

$$0 \leq \gamma^{\text{birth}}(m, r, \kappa_1, \kappa_2) \leq \overline{\gamma}_b, \quad \exists n \in \mathbb{N} : 0 \leq \gamma^{\text{death}}(m, r, \kappa_1, \kappa_2) \leq (1 \vee m)\overline{\gamma}_d.$$

At time t an individual of clan x and trait $\kappa(x)$ gives birth at rate

$$N\beta(\kappa(x)) + \sum_{z \in X} \bar{\mu}(\{z\} \times K) \cdot \gamma^{\text{birth}}(m, r(z,x), \kappa(z), \kappa(x)),$$

where $0 < \underline{\beta} \leq \inf_\kappa \beta(\kappa) \leq \sup_\kappa \beta(\kappa) \leq \bar{\beta} < \infty$ and $\sup_{m,r,\kappa,\kappa'} \gamma^{(\text{birth})}(m, r, \kappa, \kappa') \leq \bar{\gamma}_b < \infty$.

Remark (A coupling)

For later purposes (compact containment), we need a statement of the form: *Couple for each $N \in \mathbb{N}$, $\mathcal{X}^N$ with birth-enhancement rate γ^{birth}, to a process $\mathcal{Y}^N$ with*

$$\gamma^{\text{birth}}(m, r(z,x), \kappa(z), \kappa(x)) \leq C\beta(\kappa(x)) \equiv \gamma^{\text{birth},Y}(\kappa(x)) \ (*)$$

such that

$$X_t^N \subseteq Y_t^N, r_t^N = r_t^{N,Y}|_{X_t^N} \text{ and } \mu_t^N \leq \mu_t^{N,Y}. \tag{1}$$

Idea:

$$\frac{\sup_{m,r,\kappa} \gamma^{\text{birth}}(m, r, \kappa, \kappa')}{\beta(\kappa')} \leq \frac{\bar{\gamma}_b}{\underline{\beta}} = C.$$

Note: It is important, that the quantity $(*)$ is independent of the weights $\bar{\mu}(\{z\} \times K)$. E.g., $\gamma^{\text{birth}}(m, r(z,x), \kappa(z), \kappa(x)) \leq \gamma^{\text{birth},Y}(m, r(z,x), \kappa(z), \kappa(x))$ does not ensure (1).

Result

Theorem

The family $(\chi^N)_{N\in\mathbb{N}}$ of approximating particle systems (+ assumptions on initial masses, branching rates, competition rates, mutation operator) is tight. Any limit process χ satisfies "the" $(\Omega, \mathcal{D}(\Omega))$-martingale problem.

Proof includes:

- apply test-functions F to χ_t^N,
- $\Rightarrow$ χ_t^N solves martingale problem characterized by generator Ω_N,
- convergence of $\Omega_N F$ to a generator ΩF,
- **existence** limit: use Jakubowski's criterion for tightness.
  - this includes: show **compact containment condition**: $\exists N_0 \in \mathbb{N}$ such that $\forall T, \epsilon_0 > 0$ $\exists K_{T,\epsilon_0} \subset \mathbb{M}^K$ compact such that

$$\inf_{N\geq N_0} \mathbb{P}(\{\chi_t^N \in K_{T,\epsilon_0} \text{ for all } t \in [0,T]\}) > 1-\epsilon_0$$

Remark

Work in progress: *Uniqueness of solutions to the martingale problem. The latter would imply that χ^N converges to χ in law in $\mathbb{D}(\mathbb{R}_+, \mathbb{M}^K)$.*

Relative compactness in mmm-spaces

(see [DGP11] and [GPW09] (Greven, Pfaffelhuber and Winter (2009)) in case $\mu(X \times K) = 1$)

Proposition

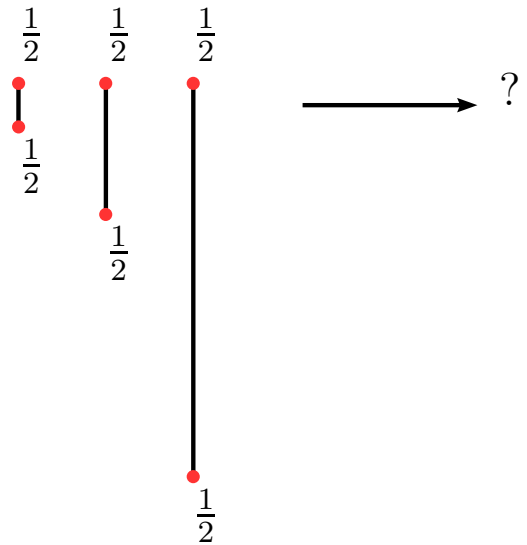
A family $\Gamma \subset \mathbb{M}^K$ is relatively compact iff
for all $\epsilon > 0$ there exists $N_\epsilon \in \mathbb{N}$ and a compact subset $K_\epsilon \subset K$ such that for all $\mathcal{X} = \overline{(X, r, \mu)} \in \Gamma$:

- (i) $m \leq N_\epsilon$,
- (ii) $\mu(X \times K_\epsilon^c) \leq \epsilon$, *(i.e., distribution of traits tight)*
- (iii) *there exists a subset $X_\epsilon \subset X$ with*
  - $(iii-a)$ $\mu({X_\epsilon}^c \times K) \leq \epsilon$,
  - $(iii-b)$ X_ϵ *has diameter at most N_ϵ,*
  - $(iii-c)$ X_ϵ *can be covered by at most N_ϵ balls of radius ϵ.*

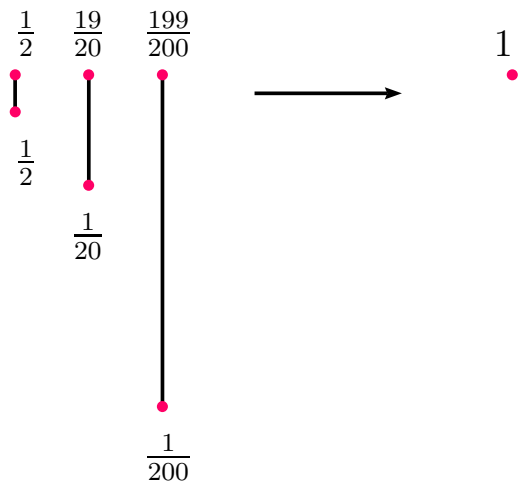
$(iii-a)$ $\mu(X_\epsilon{}^c \times K) \leq \epsilon$,

$(iii-b)$ X_ϵ has diameter at most N_ϵ

Example where relative compactness fails:

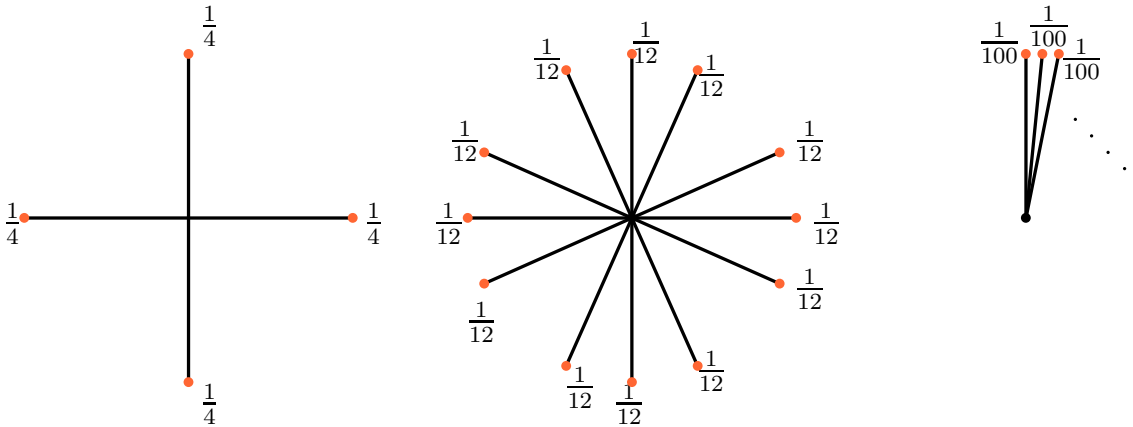


Example where relative compactness holds:

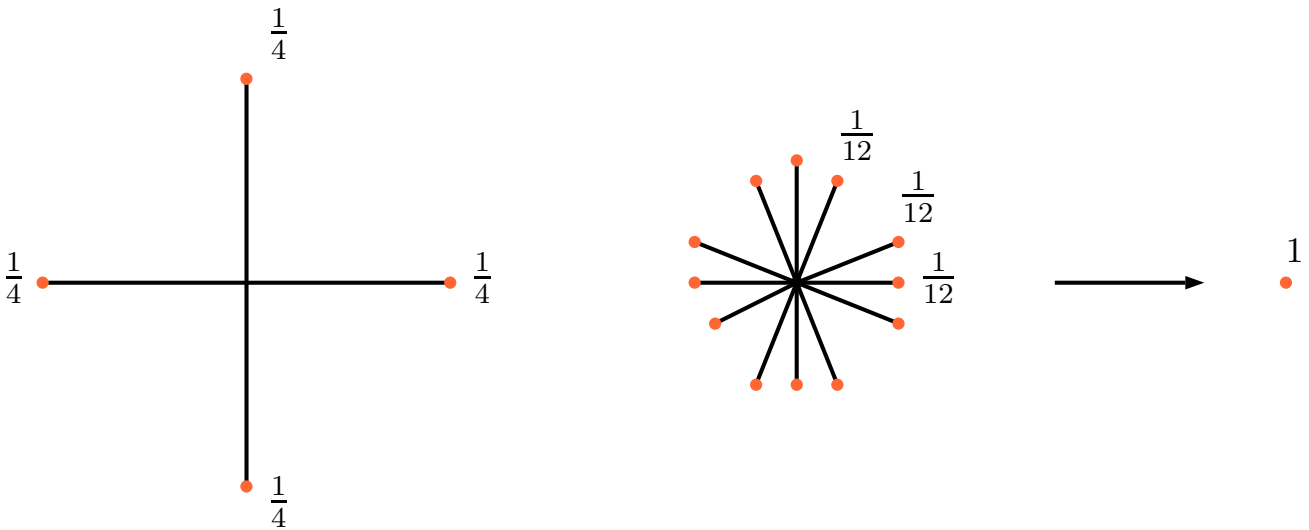


$(iii - c)$ X_ϵ can be covered by at most N_ϵ balls of radius ϵ

Example where relative compactness fails:



Example where relative compactness holds:



1. **The mass.**
   (i) $m \leq N_\epsilon$
   Assumption on initial condition: $\sup_{N\in\mathbb{N}} \mathbb{E}[(m_0^N)^3] < \infty$ implies $\sup_{N\in\mathbb{N}} \mathbb{E}[\sup_{t\in[0,T]} m_t^N] < \infty$. Now use Chebyshev's inequality.

2. **The trait and the diameter.**

   (ii) $\mu(X \times K_\epsilon^c) \leq \epsilon$,
   (iii) $\exists X_\epsilon \subset X$ with

   $(iii-a)$ $\mu({X_\epsilon}^c \times K) \leq \epsilon$,
   $(iii-b)$ X_ϵ has diameter at most N_ϵ.

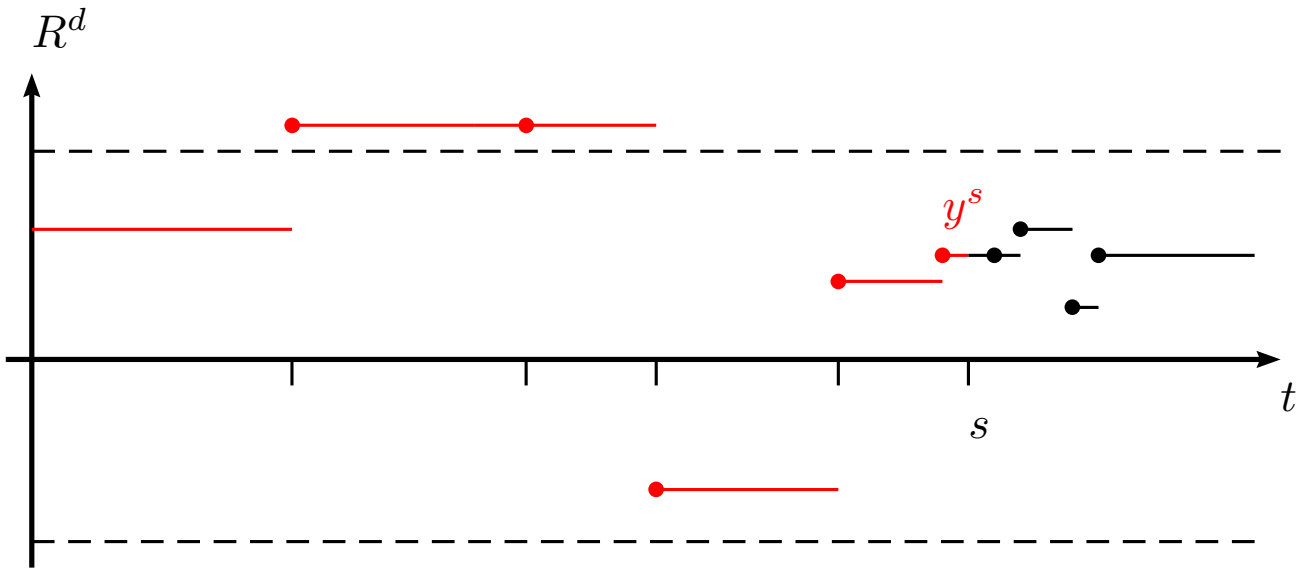
   Apply results of [MT12] (Méléard and Tran (2012), [K14] (Kliem (2014)). For (iii-b), add age to trait-space.

[MT12], [K14] - historical particle systems

The **population** is represented by a point measure as follows:

$$X_t^N := \frac{1}{N} \sum_{i=1}^{Nm_t^N} \delta_{y^i_{\cdot \wedge t}} \in \mathcal{M}_P^N(\mathbb{D}_{\mathbb{R}^d}),$$

where Nm_t^N is the number of individuals alive at time t.

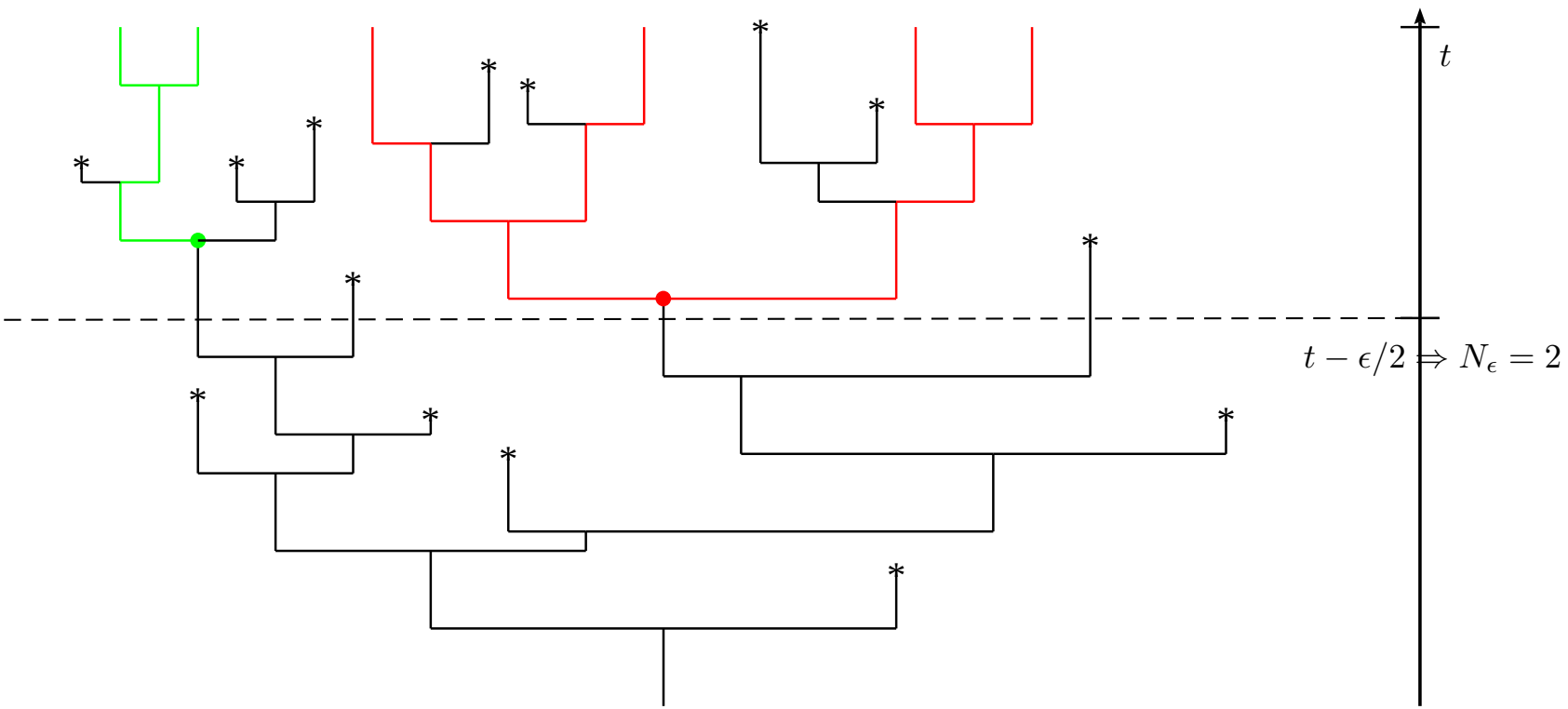


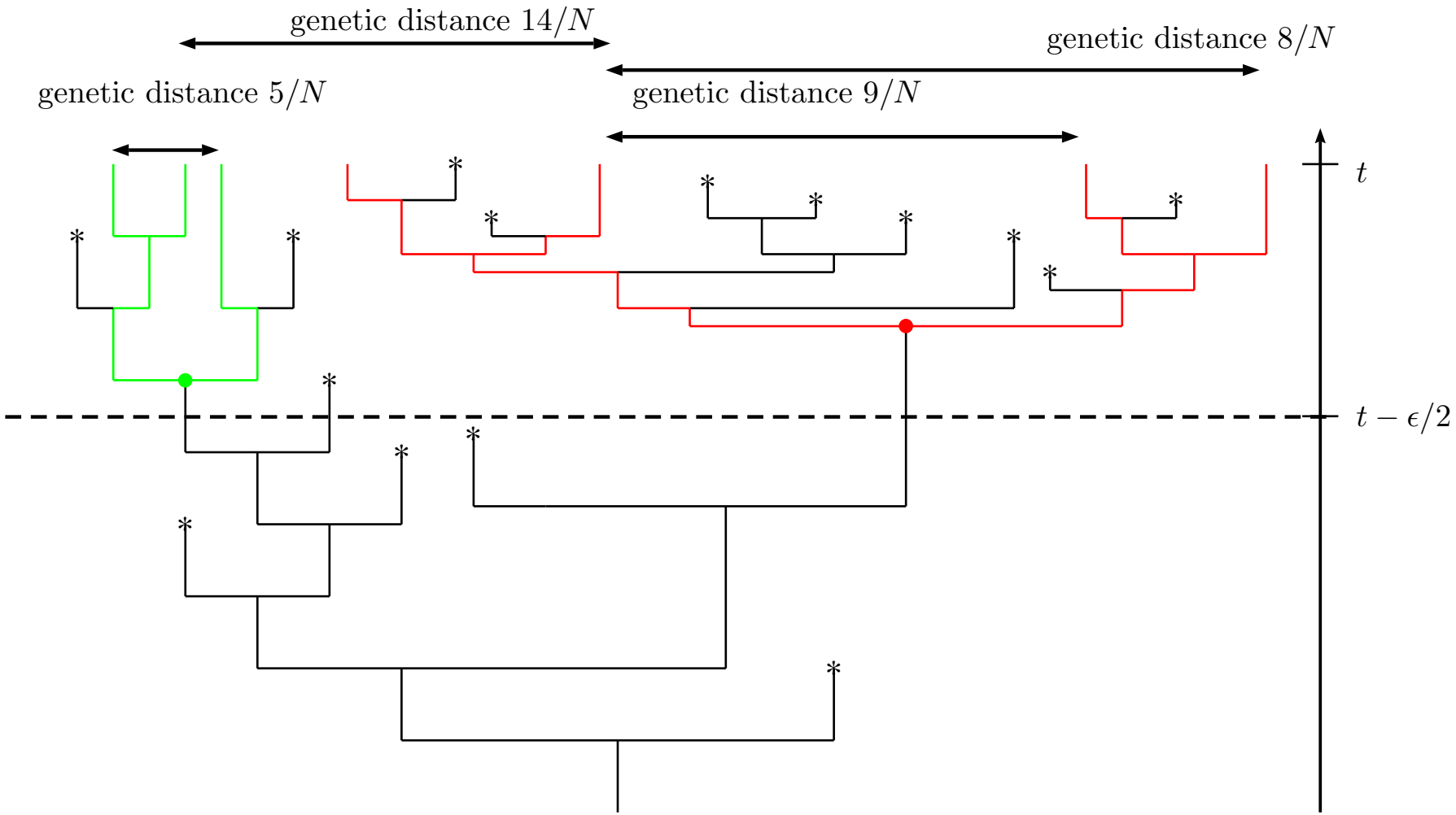
Relative compactness in $\mathbb{D}_{\mathbb{R}^d}$ (cf. [EK05]) (Ethier and Kurtz (2005)): for each finite time-interval, traits stay in a compact set **and** there exists a uniform modulus of continuity.

1.

2.

3. **Coverage by ϵ-balls.**
   (iii − c) X_ϵ can be covered by at most N_ϵ balls of radius ϵ.
   Idea in the **ultrametric case**: ([GPW13] applied it to tree-valued Moran dynamics)
   1. *ultrametric* $\Rightarrow$ time = genetic distance/2 = time to MRCA
   2. *1. Case:* $t \leq \epsilon/2$, then $N_\epsilon = 1$,
   3. *2. Case:* $t > \epsilon/2$, then $N_\epsilon = \#$ ancestors at time $t - \epsilon/2$.

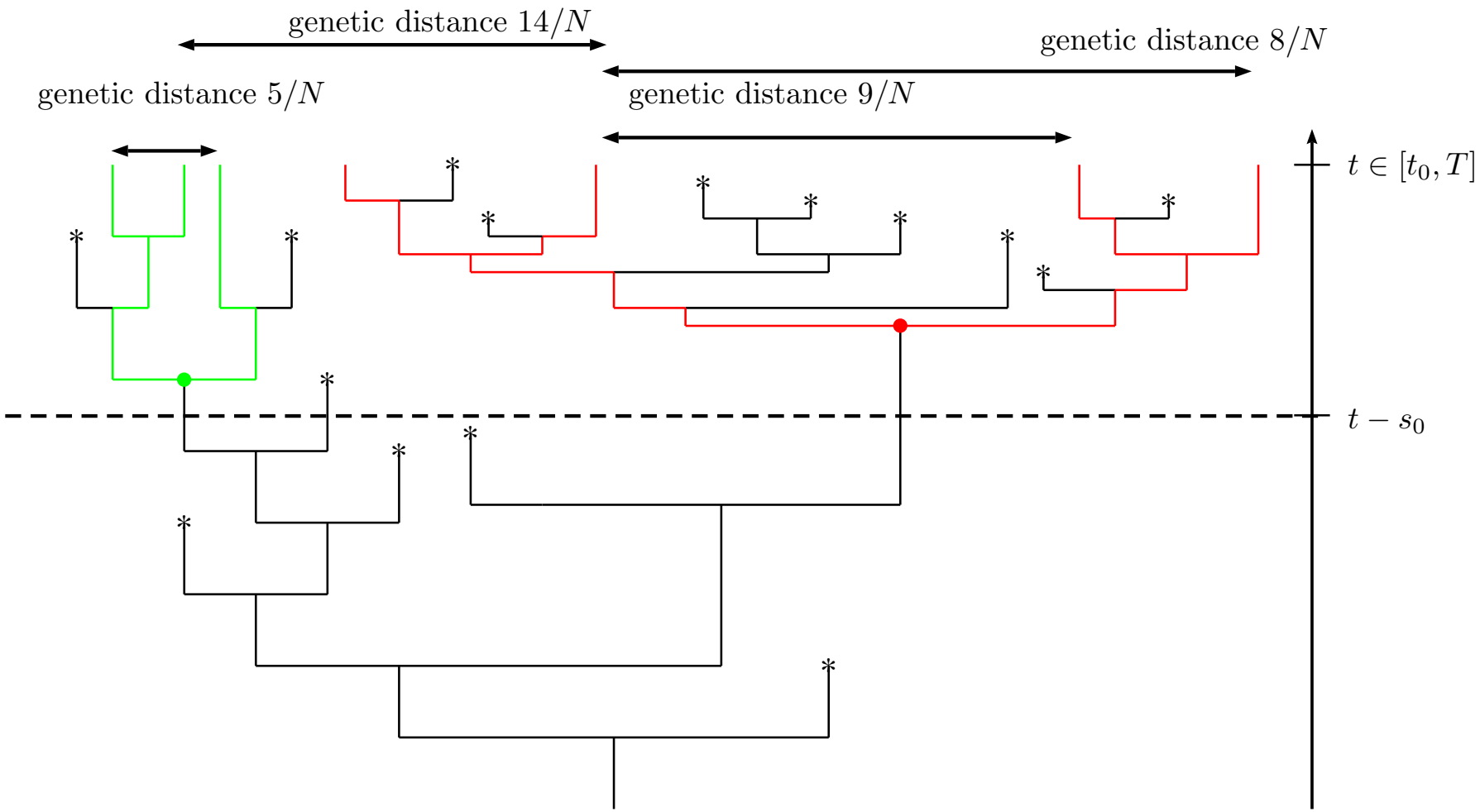



genetic distance $14/N$
genetic distance $8/N$
genetic distance $5/N$
genetic distance $9/N$
t
$t - \epsilon/2$

$\exists X_\epsilon \subset X$ with $(iii-a)$ $\mu(X_\epsilon{}^c \times K) \leq \epsilon$,
$(iii-c)$ X_ϵ can be covered by at most N_ϵ balls of radius ϵ.

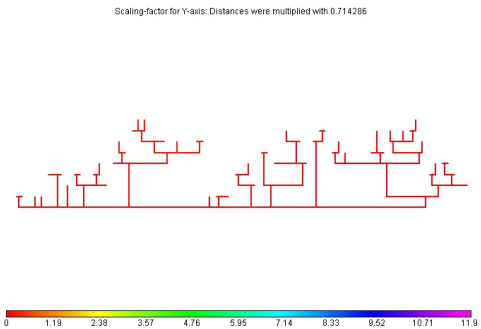
Idea in the **non-ultrametric case**:

- On a small enough time-intervall $[0, t_0(\epsilon)]$, the genetic distances of a large enough proportion of particles alive at time $t \in [0, t_0]$ to their ancestors at time 0 are at most $\epsilon/2$ with high probability. Combine this with tightness-assumption at time $t = 0$.
- For the remaining time-interval $[t_0, T]$: $\exists 0 < s_0 < t_0/2$ such that
  - the # of ancestors at time $t - s_0$ can be uniformly bounded in $t \in [t_0, T]$ and $N \geq N_0$ **and**
  - the "main part" of the progeny at time t has genealogical distance less than ϵ from its respective ancestor at time $t - s_0$.

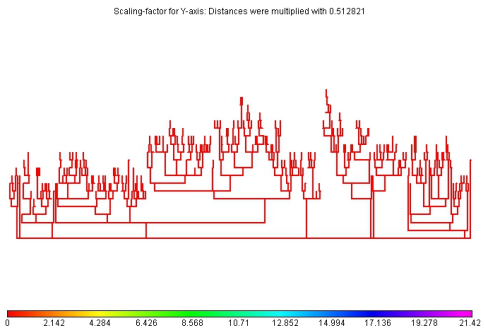

genetic distance $14/N$
genetic distance $8/N$
genetic distance $5/N$
genetic distance $9/N$
$t \in [t_0, T]$
$t - s_0$

Simulations with 

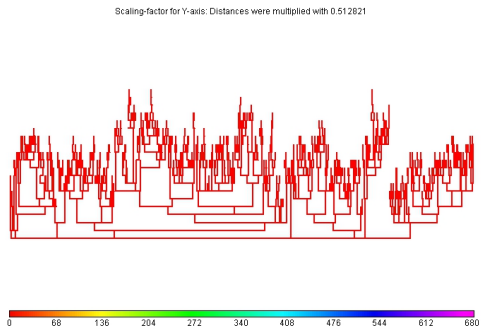
Different mutation speed:



p=0.5



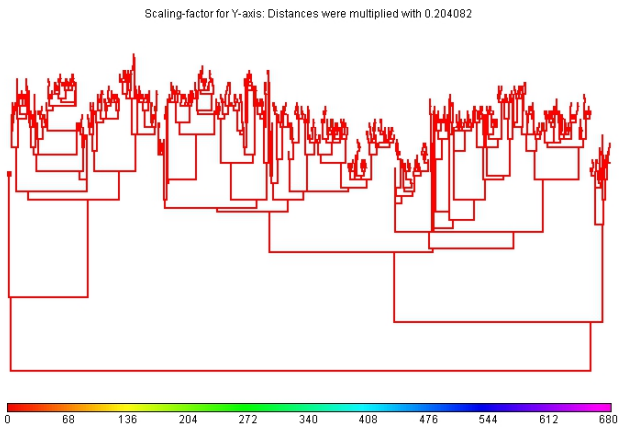
p=0.9



p=1.0

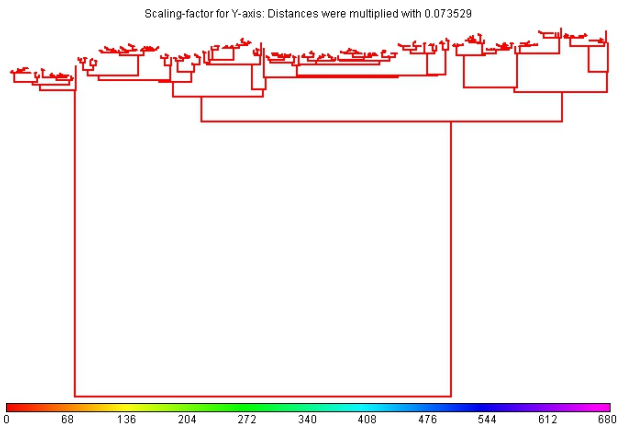
$$\beta = 1.,\ \gamma^{\text{birth}} = 10.,\ \gamma^{\text{death}} = e^{-20 \cdot r},\ T = .35$$

Different birth- and death-rates:



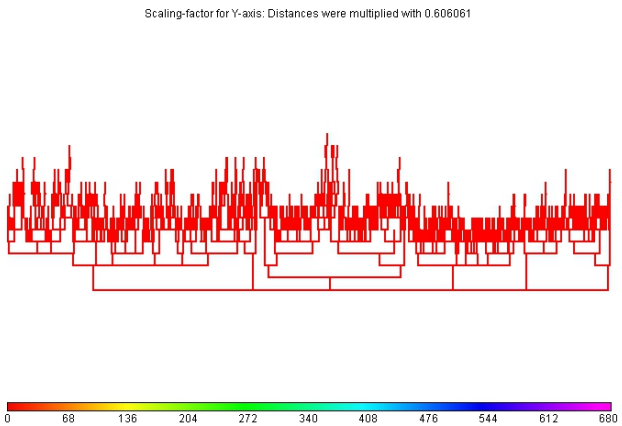
$T = 2.7, \quad \gamma^{\text{birth}} = 1.$

$\gamma^{\text{death}}(r) = e^{-200 \cdot r}$



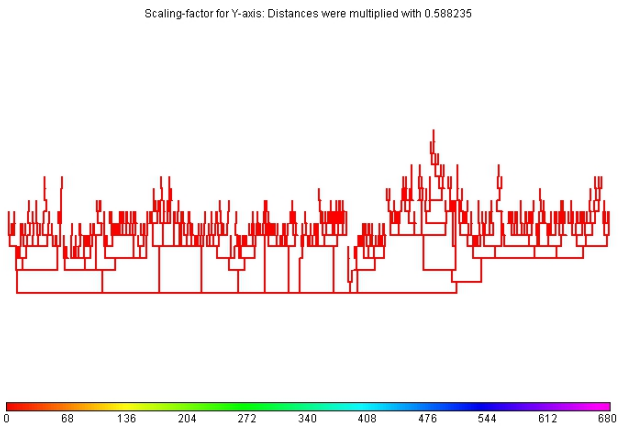
$T = 10., \quad \gamma^{\text{birth}} = 1.$

$\gamma^{\text{death}}(r) = e^{-200 \cdot r} + e^{-200 \cdot |r-1|}$



$T = .27, \quad \gamma^{\text{birth}}(r) = 1. + 2e^{-200 \cdot |r-0.5|}$

$\gamma^{\text{death}}(r) = e^{-200 \cdot r} + e^{-200 \cdot |r-1|}$



$T = .2, \quad \gamma^{\text{birth}}(r) = 2e^{-200 \cdot |r-0.5|}$

$\gamma^{\text{death}}(r) = e^{-200 \cdot r} + e^{-200 \cdot |r-1|}$

References


Depperschmidt, A., Greven, A. and Pfaffelhuber, P.
Marked metric measure spaces.
Elect. Comm. in Probab. **16** *(2011), 174–188.*


Dawson, D.A. and Perkins, E.A.
Historical processes.
Memoirs of the AMS **93** *(1991).*


Ethier, S.N. and Kurtz, T.G.
Markov Processes: Characterization and Convergence.
Wiley and Sons, Inc., Hoboken, New Jersey (2005).


Greven, A., Pfaffelhuber, P. and Winter, A.
Convergence in distribution of random metric measure spaces (Λ-coalescent measure trees).
Probab. Theory Relat. Fields **145** *(2009), no. 1-2, 285–322.*

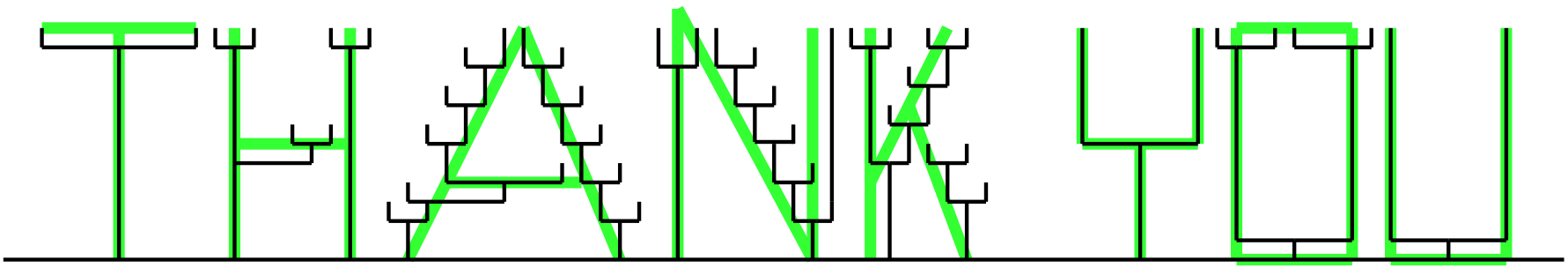

Greven, A., Pfaffelhuber, P. and Winter, A.
Tree-valued resampling dynamics martingale problems and applications.
Probab. Theory Relat. Fields **155** *(2013), no. 3-4, 789–838.*


Grenfell, B.T., Pybus, O.G., Gog, J.R., Wood, J.L.N., Daly, J.M., Mumford, J.A. and Holmes, E.C.
Unifying the Epidemiological and Evolutionary Dynamics of Pathogens.
Science **303** *(2004), 327–332.*

Kliem, S.
A compact containment result for nonlinear historical superprocess approximations for population models with trait-dependence.
Electron. J. Probab. **19** *(2014), no. 97, 1–13.*

Lipsitch, M. and O'Hagan, J.J.
Patterns of antigenic diversity and the mechanisms that maintain them.
J. R. Soc. Interface **4** *(2007), 787–802.*

Méléard, S. and Tran, V.C.
Nonlinear historical superprocess approximations for population models with past dependence.
Electron. J. Probab. **17** *(2012), no. 47, 1–32.*


THANK YOU