

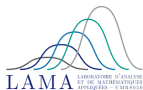
# Evolving genealogies for branching populations under selection and competition

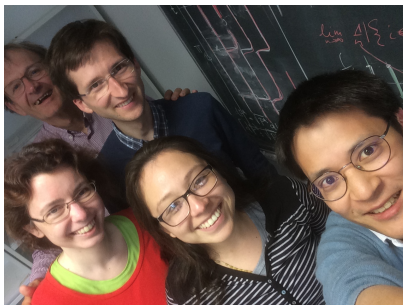
TRAN Viet Chi

LAMA, Université Gustave Eiffel, France

CIRM

February 13, 2020



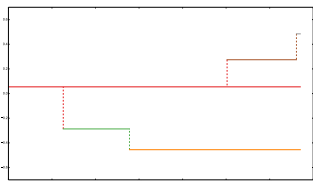
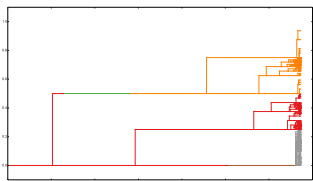


Airam Blancas Benitez, Stephan Gufler, Sandra Kliem and Anton Wakolbinger

# Motivations

★ Trait structure, fluctuating size due to interactions defined at the individual level, in particular competition.

★ Forward-Backward coalescent:



**But what happens to phylogenies if we look at them in the ecological time-scale?**

1. Billiard, Ferrière, Méléard & **Tran**, *J. of Math. Biol.*, 2015
2. Lepers, Méléard, Porte, Billiard, **Tran**, in prep.
3. Neuhauser & Krone, *Genetics*, 1996
4. Kaj & Krone, *JAP*, 2003

# Donnelly and Kurtz

★ Two seminal papers by Donnelly and Kurtz (1999)

DK99a: monotype systems with fluctuating population size

DK99b: multitype systems with constant population size

★ Our program:

1. A **look-down** representation of the evolving genealogy (the tree-valued process) that underlies a multitype continuous state branching process with **competition** and **fecundity selection**.
2. The identification of 1. as the scaling limit of the evolving genealogies of the approximating finite population system

---

1. Donnelly & Kurtz, *AAP*, 1999

2. Donnelly & Kurtz, *AP*, 1999

# Toy model with 2 coexisting traits

★ 2 traits:  $\mathbb{I} = \{A, B\}$ .

★ Rescaled birth-death process:  $\tilde{N}_t^{(K)} = \frac{1}{K} (\tilde{N}_t^{K,A} + \tilde{N}_t^{K,B})$ ,

Birth rate:  $\frac{K}{2} + b\mathbf{1}_A$

Death rate:  $\frac{K}{2} + c \frac{\tilde{N}_t^{K,A}}{K} \mathbf{1}_B + c \frac{\tilde{N}_t^{K,B}}{K} \mathbf{1}_A$

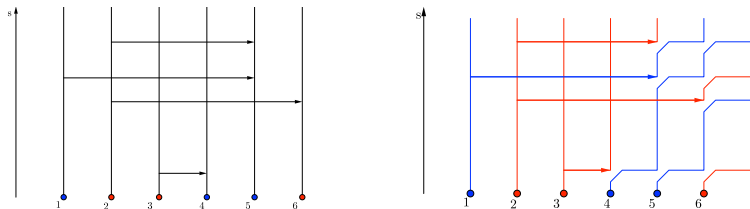
★ Large population limit:  $K \rightarrow +\infty$

$$d\xi_t^A = b\xi_t^A dt - c\xi_t^A \xi_t^B dt + \sqrt{\xi_t^A} dW_t^A$$

$$d\xi_t^B = -c\xi_t^A \xi_t^B dt + \sqrt{\xi_t^B} dW_t^B.$$

$$d\xi_t = (b\mu_t^A \xi_t - 2c\mu_t^A (1 - \mu_t^A) \xi_t^2) dt + \sqrt{\xi_t} dW_t.$$

# Look-down representation



★ Here:

- ▶ **Natural births:**  $i \rightarrow j$ ,  $1 \leq i < j \leq N_{s-}^K$   
rate 1, levels  $\geq j$  are shifted to the right.
- ▶ **Natural deaths:** remove individual at level  $N_{s-}^K$   
rate  $\frac{N_{s-}^K (1 + N_{s-}^K)}{2}$
- ▶ **Selective births:**  $i \rightarrow j$ , if  $i$  is of type  $A$   
rate  $\frac{b}{K}$ , copies level  $j$  at level  $i$ .
- ▶ **Competitive deaths:**  $i \in \{1, \dots, N_{s-}^K\}$   
rate  $c \frac{N_{s-}^{K,A} (N_{s-}^K + 1)}{K^2}$ , replaces level  $i$  by level  $N_{s-}^K$ .

# Marked Metric Measure space (mmm)

★ We want to keep track of **population size**, **types** of individuals and **genealogical distances** between them.

★ Two time measures: **ecological time**  $t$  and **Look-down time**  $s$ .

$(N_s^K, R_s^K, G_s^K : s \geq 0)$ : pop. size, matrix of distances, vector of types.  
 $(\tilde{N}_t^K, \tilde{R}_t^K, \tilde{G}_t^K : t \geq 0)$

★ Time change  $dt = \frac{N_s^K + 1}{K} ds$

★ Unlabelling: for  $N \in \mathbb{N}$ ,  $\nu = (N, r, g)$  and for a permutation  $\pi \in \mathcal{S}_N$ ,  
 $\nu^\pi := (N, r^\pi, g^\pi) = (N, (r(\pi_i, \pi_j), 1 \leq i, j \leq N), (g(\pi_i), 1 \leq i \leq N))$ .

$$\nu^\nu := \frac{1}{N!} \sum_{\pi \in \mathcal{S}_N} \delta_{\nu^\pi}$$

★ **Th:** The mmm associated with the unlabelled BPS and the mmm corresponding to the unlabelled look-down changed of time have the same distribution.

# Infinite Look-down

★ Limit when  $K \rightarrow +\infty$ :  $(\zeta_s, R_s, G_s)$

★ Time change:  $ds = \frac{1}{\xi_t} dt$ ,  $\zeta_s = \xi_t$ .

$$d\zeta_s = (b\mu_s^A \zeta_s^2 - 2c\mu_s^A(1 - \mu_s^A)\zeta_s^3) ds + \zeta_s dW_s.$$

★ Look-down:

- ▶ **Natural births:**  $i \rightarrow j$ ,  $1 \leq i < j$   
rate  $\mathbf{1}$ , levels  $\geq j$  are shifted to the right.
- ▶ **Selective births:**  $i$   
rate  $b\mu_s^A \zeta_s$ , copies at level  $i$  an  $A$  individual randomly sampled.
- ▶ **Competitive deaths:**  $i$   
rate  $c\mu_s^A \zeta_s^2$ , replaces level  $i$  by a randomly sampled individual.



# Type frequencies

★ Question: do we have

$$\mu_s^A \stackrel{?}{=} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{A\}}(g_s(i))$$

★ a type configuration  $g \in \{A, B\}^{\mathbb{N}}$  admits type frequencies if

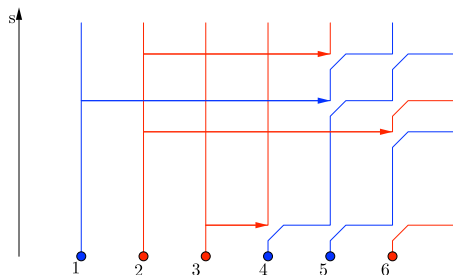
$$\mu^g := \lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{i=1}^n \delta_{g(i)}$$

exists.

# Initial conditions

- ★ a (total) **population size**  $v_0 \in \mathbb{R}_+$   
(thought of as the limit of  $N_0^K$  when  $K \rightarrow +\infty$ ),
- ★ a **matrix**  $(R_0(i, j))_{i, j \in \mathbb{N}}$  of **ultrametric distances**  
(the genealogical distances of pairs of individuals in an infinite random sample from the population, visualised as a coalescent tree)
- ★ a **type configuration**  $(G_0(i))_{i \in \mathbb{N}}$   
(the types of the drawn individuals)
- ★ The couple  $(R_0, G_0)$  is jointly exchangeable.

# Neutral framework



★ One of the classical results of Donnelly and Kurtz:

**Th:** Let  $(G_s^{(0)})_{s \geq 0}$  be the type transport along the neutral LD graph. Then (because of preservation of exchangeability)  $G_s^{(0)}$  admits type frequencies, and  $(\mu^{G_s^{(0)}}\{A\})_{s \geq 0}$  is a standard Wright-Fisher diffusion.

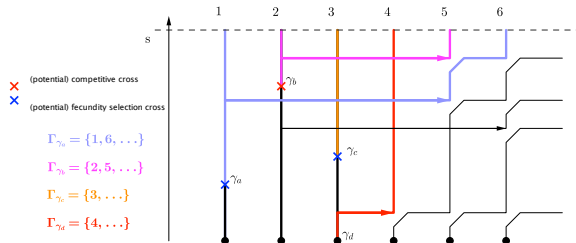
★ Neutral lookdown space: on the completion of  $\mathbb{R}_+ \times \mathbb{N}$ , we can define

$$m_s^{(0)} = \lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{i=1}^n \delta_{(s,i)}$$

and extend to include type distributions:

# SDE for the type of $(s, j)$

$$\begin{aligned}
 G_s(j) = & g_0(j) + \sum_{i=1}^{j-1} \int_{[0, s]} (G_{u-}(i) - G_{u-}(j)) dL_{ij}(u) \\
 & + \sum_{1 \leq i < k < j} \int_{[0, s]} (G_{u-}(j-1) - G_{u-}(j)) dL_{ik}(u) \\
 & + \int_{[0, s] \times \mathbb{R}_+ \times [0, 1] \times \{\beta, \delta\}} (q(G_{u-}(j), G_{u-}, \zeta_{u-}, z, w, \omega) - G_{u-}(j)) dK_j(u, z, w, \omega).
 \end{aligned}$$



## SDE for the type of $(s, j)$

$$\begin{aligned}G_s(j) &= g_0(j) + \sum_{i=1}^{j-1} \int_{[0, s]} (G_{u-}(i) - G_{u-}(j)) dL_{ij}(u) \\ &+ \sum_{1 \leq i < k < j} \int_{[0, s]} (G_{u-}(j-1) - G_{u-}(j)) dL_{ik}(u) \\ &+ \int_{[0, s] \times \mathbb{R}_+ \times [0, 1] \times \{\beta, \delta\}} (q(G_{u-}(j), G_{u-}, \zeta_{u-}, z, w, \omega) - G_{u-}(j)) dK_j(u, z, w, \omega).\end{aligned}$$

where

$$q(h, g, \zeta, z, w, \beta) := \begin{cases} A & \text{if } z \leq b \mu^g \{A\} \zeta, \\ h & \text{otherwise,} \end{cases}$$
$$q(h, g, \zeta, z, w, \delta) := \begin{cases} \kappa(\mu^g, w) & \text{if } z \leq c \mu^g \{A\} \zeta^2 \text{ and } h = B, \\ h & \text{otherwise.} \end{cases}$$

with  $\mu^g \{A\} = \lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{g(i)=A}$ .

# Joint evolution of genealogies and population size

★ Recall:

$$d\zeta_s = (b\mu_s^A \zeta_s^2 - 2c\mu_s^A(1 - \mu_s^A)\zeta_s^3) ds + \zeta_s dW_s.$$

$$\begin{aligned} G_s(j) &= g_0(j) + \sum_{i=1}^{j-1} \int_{[0,s]} (G_{u-}(i) - G_{u-}(j)) dL_{ij}(u) \\ &+ \sum_{1 \leq i < k < j} \int_{[0,s]} (G_{u-}(j-1) - G_{u-}(j)) dL_{ik}(u) \\ &+ \int_{[0,s] \times \mathbb{R}_+ \times [0,1] \times \{\beta, \delta\}} (q(G_{u-}(j), G_{u-}, \zeta_{u-}, z, w, \omega) - G_{u-}(j)) d\mathcal{K}_j(u, z, w, \omega). \end{aligned}$$

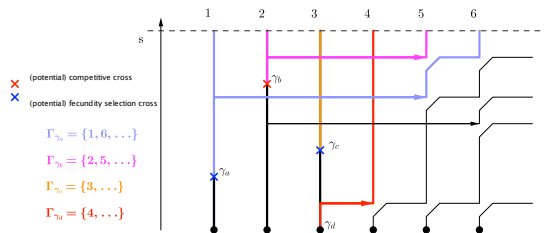
**Th:** There exists a unique strong solution  $(\zeta_s, G_s(j), j \in \mathbb{N} : s \geq 0)$  (up to extinction or explosion) such that  $G_s$  admits type frequencies.

# Sampling measure with competition

★ In the neutral case,

$$m_s^{(0)} = w - \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \delta_{((s,i), G_s^0(i))}$$

★ We partition  $\mathbb{R} \times \mathbb{N}$  into blocks  $\Gamma_\gamma$  determined by the events  $\gamma \in \mathbb{R} \times \mathbb{N}$  corresponding to selective births and competition death.



★ For all root  $\gamma$ , we have a.s.

$$m_s^{(0)}(\Gamma_\gamma) = m_s^{(0)}(\Gamma_\gamma \times \mathbb{I}) = \lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{(s,i) \in \Gamma_\gamma}.$$

# Existence of the infinite look down

★ Consider a sequence  $(G_s^{(\ell)}(j), \zeta_s^{(\ell)}, \mu_s^{(\ell)}\{A\})$ ,  $\ell \in \{0, 1, \dots\}$ .

Plug in the SDEs for  $G_s^{(\ell)}(j)$  the mass  $\zeta_s^{(\ell-1)}$  and proportion  $\mu_s^{(\ell-1)}\{A\}$  of the previous iteration.

$$\mu_s^{(\ell-1)}\{A\} = \sum_{\gamma=(u,i)/G_u^{(k)}(i)=A} m_s^{(0)}(\Gamma_\gamma).$$

**Prop:** for each iteration step, we have a.s.

$$\mu_s^{(\ell)} = \lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{G_s^{(\ell)}(i)}.$$

★ Convergence with Picard's like argument.

★ Description of the genealogies.



# Unlabelling the leaves

★ We can write  $\mathbf{A}$  the generator of our infinite LD.

★ Define:

$$\mathbb{A}\Phi_F(v, \chi) = \int \frac{1}{v} \mathbf{A}F(v, r, g) \nu^\chi(dr, dg).$$

**Th:** The process  $(\xi, Y)$  is the unique solution of a martingale problem with generator  $(\mathbb{A}, D_0)$ .

★ We have a decoupling of population size and local phylogenies in the limit:

$$d\mu_s\{A\} = b\mu_s\{A\}(1 - \mu_s\{A\})\zeta_s ds + \sqrt{\mu_s\{A\}(1 - \mu_s\{A\})} dB_s^A, \quad (1)$$

with  $\langle B^A, \mathcal{W} \rangle_s = 0$ .

# Thank You

