

# Asymptotic genealogies of non-neutral populations

**Suzie Brown**

University of Warwick, U.K.

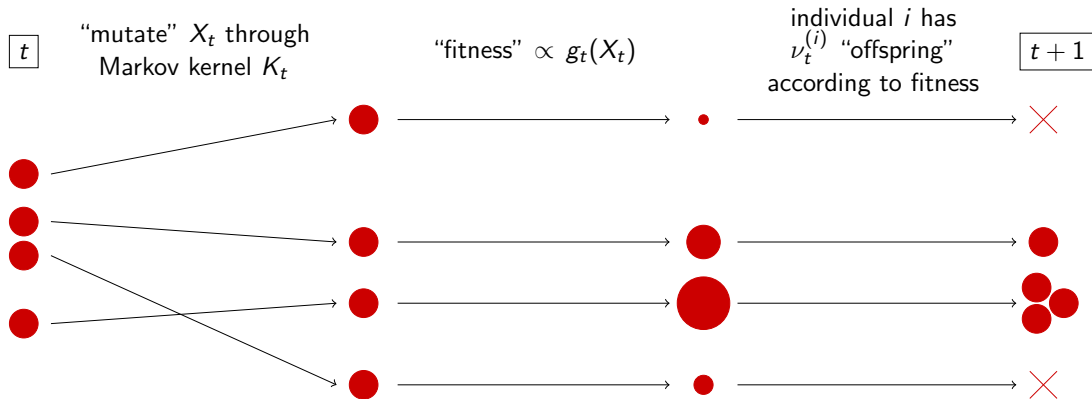
with Paul Jenkins, Adam Johansen & Jere Koskela

13 May 2021

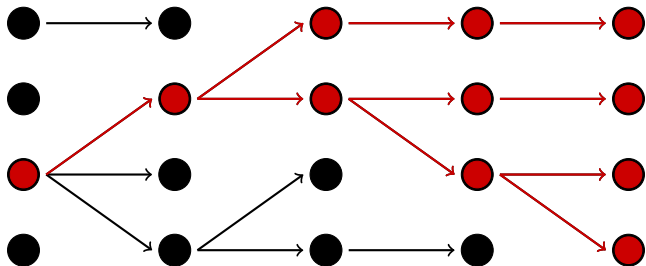
# Interacting particle system

- ▶ Constant population size  $N$
- ▶ “Genotype”  $X_t^{(i)} \in \mathcal{X} \subseteq \mathbb{R}^d$
- ▶ Initial genotypes  $X_0 \sim \mu$

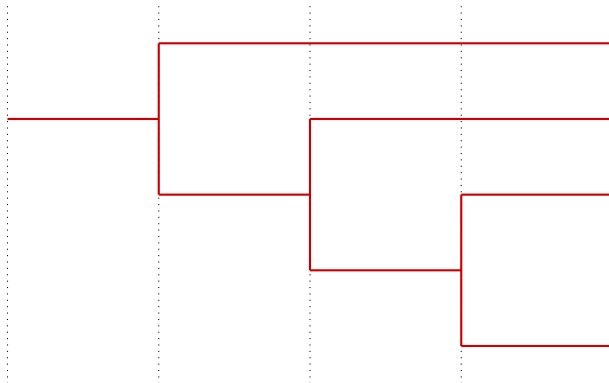
# Interacting particle system



# Genealogies

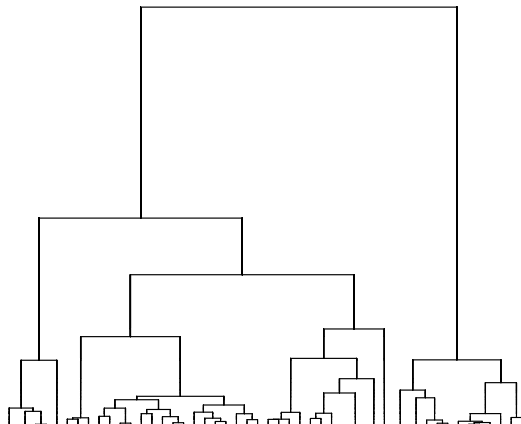


# Genealogies



# Kingman's $n$ -coalescent

- ▶ Continuous-time Markov chain on the space of partitions of  $\{1, \dots, n\}$
- ▶ Single pair mergers only
- ▶ Each pair merges independently at rate 1 (total merge rate  $\binom{k}{2}$  while there are  $k$  distinct lineages)



# Scenario

- ▶ Population size  $N$
- ▶ Discrete generations
- ▶ Sample  $n \leq N$  individuals from the terminal generation
- ▶ Rescale time appropriately
- ▶ Let  $N \rightarrow \infty$

## Sufficient conditions, neutral case

### Theorem (Kingman 1982)

- ▶ *Individuals are exchangeable*
- ▶ *Offspring counts  $\nu^{(1:N)}$  are i.i.d. across generations*
- ▶  $\text{Var}[\nu^{(1)}] = \sigma_N^2 \rightarrow \sigma^2 \in (0, \infty)$
- ▶  $\sup_N \mathbb{E}[(\nu^{(1)})^k] < \infty$  for all  $k \geq 3$

*Then the rescaled genealogy of  $n$  individuals converges weakly to the  $n$ -coalescent as  $N \rightarrow \infty$ .*



## Sufficient conditions, neutral case

- ▶ Exchangeability = neutrality (genotype does not affect number of offspring)
- ▶ System completely specified by distribution of  $\nu^{(1:N)}$
- ▶ **Wright-Fisher model:**  $\nu^{(1:N)} \sim \text{Multinomial}(N, (\frac{1}{N}, \dots, \frac{1}{N}))$
- ▶ **Moran model:**  $\nu^{(1:N)}$  uniform over permutations of  $(2, 0, 1, \dots, 1)$
- ▶ Since  $\sum \nu^{(i)} = N$  and individuals are exchangeable,  $\mathbb{E}[\nu^{(i)}] = 1$ .
- ▶ Case  $\sigma^2 = 0$  would mean no coalescences in the limit

## Necessary and sufficient conditions, neutral case

Theorem (Möhle Sagitov 2001, 2003)

- ▶ *Individuals are exchangeable*
- ▶ *Offspring counts  $\nu^{(1:N)}$  are i.i.d. across generations*
- ▶  *$c_N > 0$  for all  $N < \infty$*
- ▶  *$c_N \rightarrow 0$*
- ▶  *$d_N/c_N \rightarrow 0$*

*If and only if the rescaled genealogy of  $n$  individuals converges weakly to the  $n$ -coalescent as  $N \rightarrow \infty$ .*

$$d_N := \frac{N\mathbb{E}[(\nu^{(1)})_3]}{(N)_3}, \quad c_N := \frac{N\mathbb{E}[(\nu^{(1)})_2]}{(N)_2}$$

## Necessary and sufficient conditions, neutral case

- ▶ The condition  $c_N > 0$  plays the same role as Kingman's condition  $\sigma^2 > 0$
- ▶  $c_N = \frac{\text{Var}[\nu^{(1)}]}{N-1}$ , so  $c_N \rightarrow 0$  is less restrictive than Kingman's condition  $\text{Var}[\nu^{(1)}] \rightarrow \sigma^2$
- ▶ Only requires control up to 3rd moment, cf. Kingman requires *all* moments finite

## Sufficient conditions, non-neutral case

### Theorem (B Koskela Jenkins Johansen 2021)

- ▶ Given  $\nu_t^{(1:N)}$ , assignment of offspring to parents is uniform over all valid assignments
- ▶ Time scale is almost surely finite
- ▶  $\exists$  deterministic sequence  $b_N \rightarrow 0$  such that  $\forall N, t$

$$\frac{1}{(N)_3} \sum_{i=1}^N \mathbb{E}_t[(\nu_t^{(i)})_3] \leq b_N \frac{1}{(N)_2} \sum_{i=1}^N \mathbb{E}_t[(\nu_t^{(i)})_2]$$

*Then the rescaled genealogy of  $n$  individuals converges weakly to the  $n$ -coalescent as  $N \rightarrow \infty$ .*

# Open questions

- ▶ Are the conditions necessary and sufficient in the non-neutral case?
- ▶ What does the random time scale look like?
- ▶ Rate of convergence?
- ▶ Can the conditions be verified for some interesting population genetic models?

## References

1. JFC Kingman (1982) *The coalescent*. Stochastic Processes and Their Applications 13:235–248.
2. JFC Kingman (1982) *On the genealogy of large populations*. Journal of Applied Probability 19A:27–43.
3. M Möhle, S Sagitov (2001) *A classification of coalescent processes for haploid exchangeable population models*. The Annals of Probability 29(4):1547–1562.
4. M Möhle, S Sagitov (2003) *Coalescent patterns in diploid exchangeable population models*. Journal of Mathematical Biology 47(4):337–352.
5. S Brown, PA Jenkins, AM Johansen, J Koskela (2021) *Simple conditions for convergence of sequential Monte Carlo genealogies with applications*. Electronic Journal of Probability 26:1–22.