A null model for the observed transposition rate of transposons in the genome

Workshop "Probability meets Biology", WIAS Berlin, 27-29/11/2024
With help from Luzie Wingen (TU Munich) and Quentin Cordeau (BioSP, INRAE Avignon)

Apolline Louvet

Junior researcher at BioSP unit, INRAE Avignon

Null models and probability theory

Null models for processes with a stochastic component

=

important question in biology and ecology

- Observations $\neq$ **intuition**
- Stochasticity **cannot be neglected**

EXAMPLE 1: Hubbell's "*unified neutral theory of biodiversity and biogeography*" [Hubbell (2001)] in community ecology

$\Longrightarrow$ **No need for selection/interaction between species to explain some of the diversity patterns observed in ecological communities**

Null models and probability theory

EXAMPLE 2: Genetic diversity in spatially expanding populations

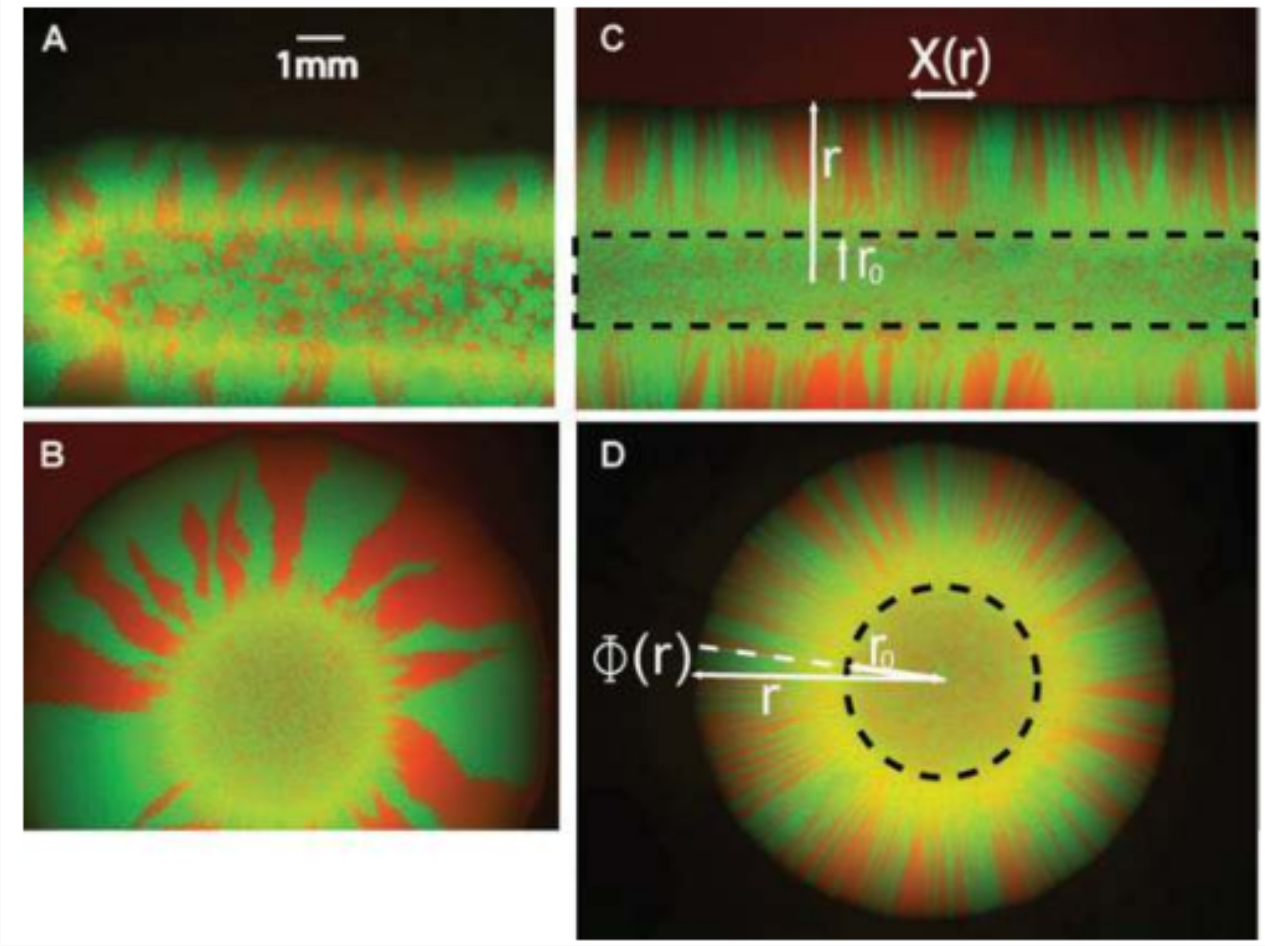


Image taken from [Hallatschek and Nelson (2010)].

Null models and probability theory

EXAMPLE 3: Ancestral reproductive bias in branching processes

- Consider a critical birth-death branching process, with birth rate = death rate = 1
- Let $T > 0$, assume that the process is still alive, and sample an individual uniformly at random
- [Cheek and Johnston (2023)]: Reproduction rate on the ancestral line ≈ 2 rather than 1

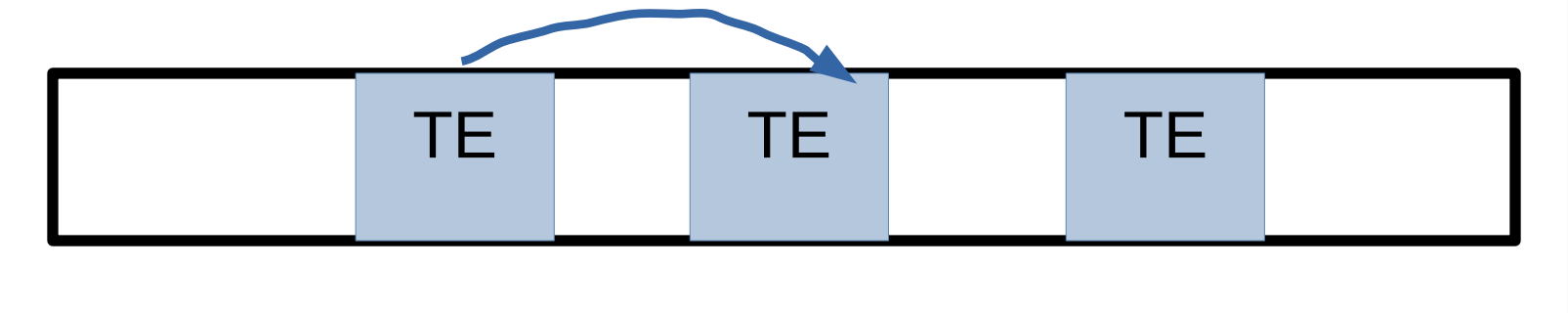
$\Longrightarrow$ **Observed** branching rate $\neq$ **actual** branching rate

What are transposons

transposons ≈ parasites of the genome

- Copied during cell division
- Can also self-duplicate (**"transpose"**) within the genome

What are transposons



Effect of transposition

- New TE in non-coding region ≈ neutral
- New TE in coding region = bad (generally)

Regulation of transposition rates

[Le Rouzic and Capy (2005)] : Optimal strategy for transposon dynamics :

- Initially, high transposition rate
- Reduce transposition rates when abundance increases

$\Longrightarrow$ Regulation of transposition rates (self or by external source)

- **Apparent regulation** of transposition rates observed experimentally
- Extensive literature on possible regulation mechanisms (see [Bourque et al. (2018)] for a review)

Regulation of transposition rates

How is apparent regulation observed

- Count TEs at time t, $2t$, $3t$,...
- Compute the difference in TE counts over time
- Deduce transposition rates

⚠Possible bias = only non-dead cells observed

Bias in observed transposition rates

ASSUMPTION

$$\text{Death rate of a cell} \sim \text{TE count}$$

- Cells with higher than average TE count = more likely to die
- Cells with lower than average TE count = less likely to die

$\Longrightarrow$ **We tend to only observe cells with lower than average TE count**

Is this sufficient to explain the temporal dynamics of the observed transposition rate ?

A first toy model

Model

"TE dynamics embedded in a neutral Wright-Fisher model"

- N haploid individuals
- Initially, each individual contains 1 TE

Step 1: "TE dynamics"

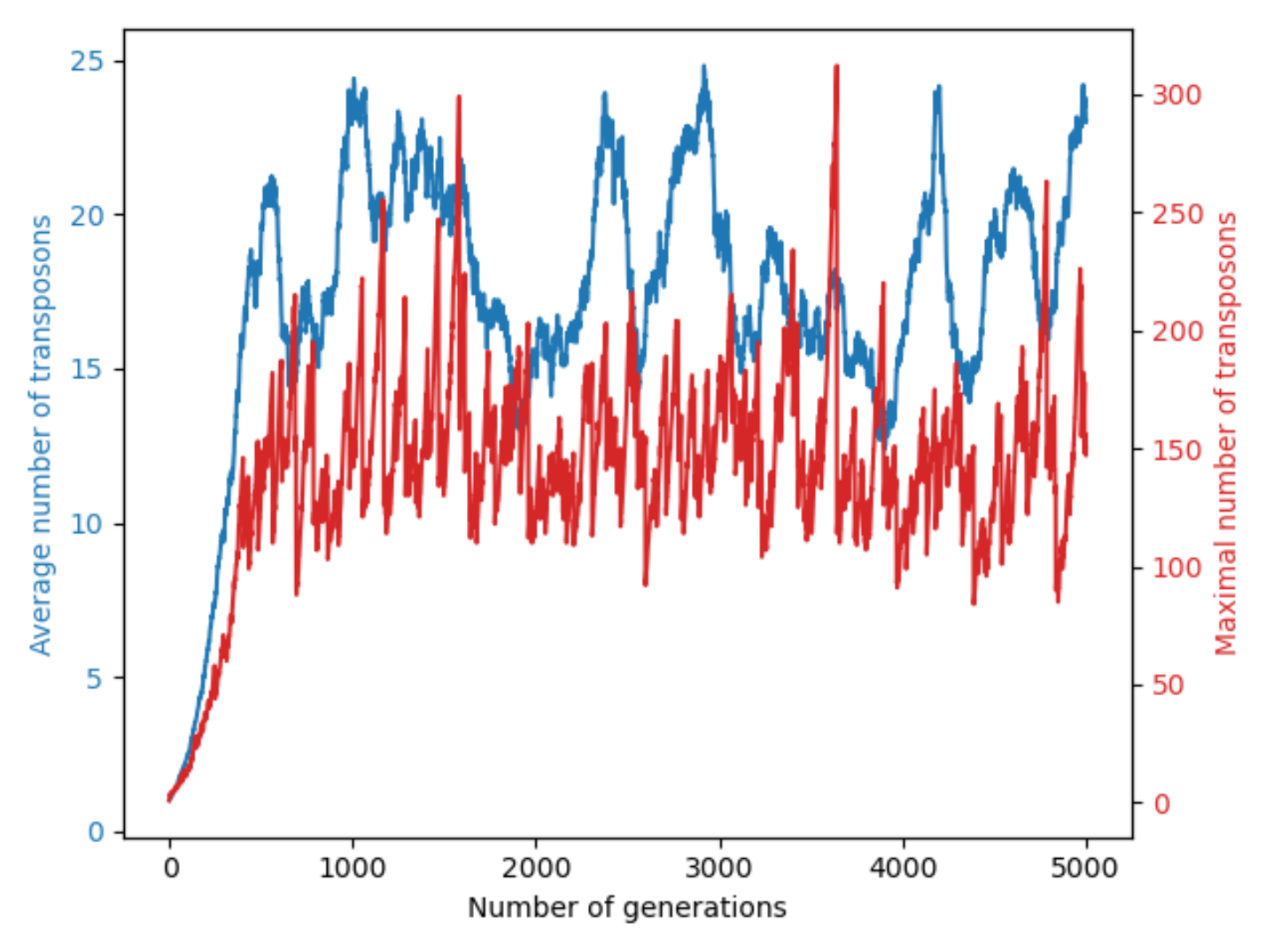
- Each TE attempts to transpose
- Three outcomes: no new TE, success, death of the cell

Step 2: "Neutral Wright-Fisher model"

- N new individuals produced from N' remaining individuals
- TE count conserved from parent to child

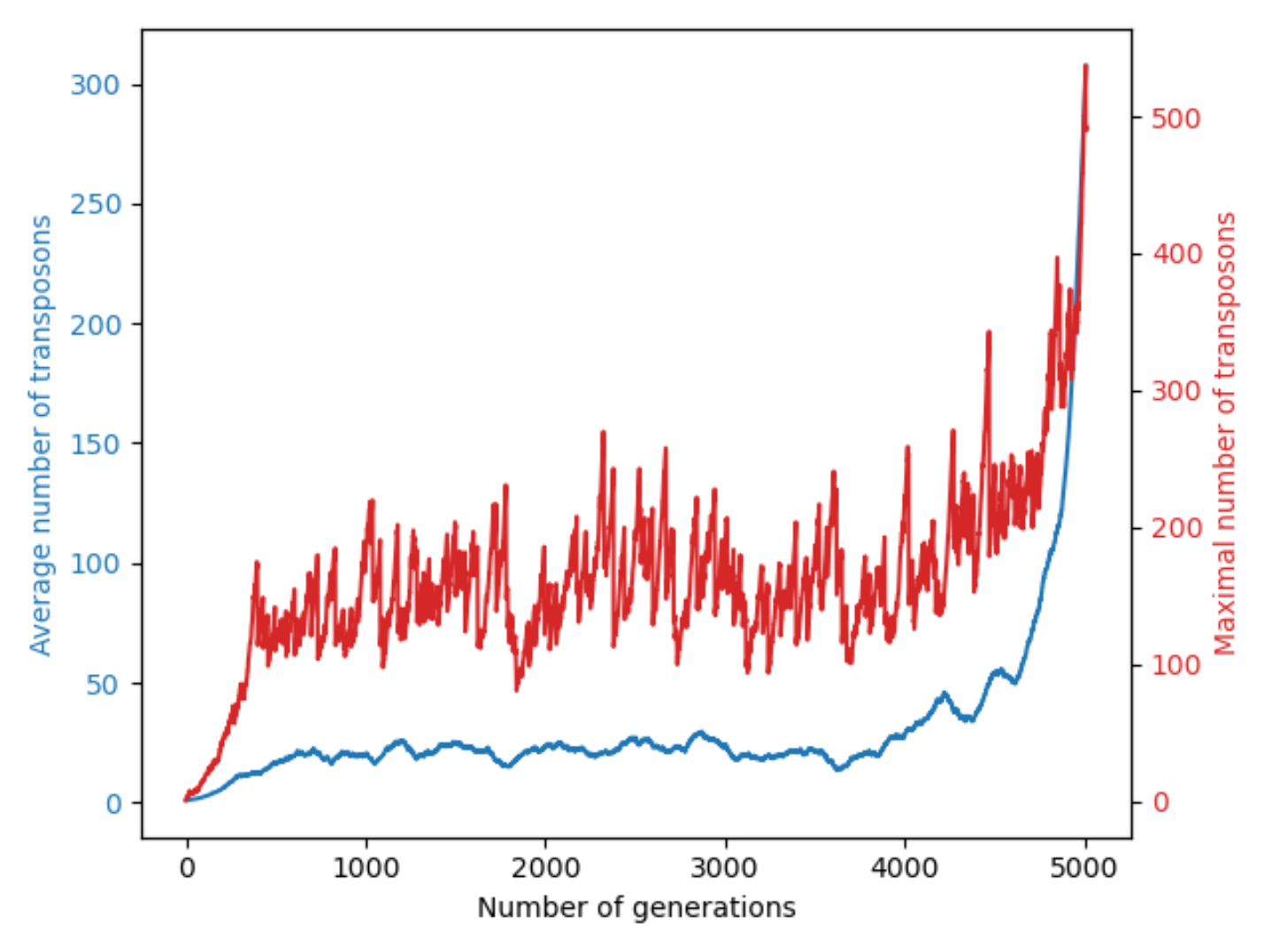
A first toy model

$\Longrightarrow$ **Possible to find parameter sets for which the TE count evolves as wanted for long durations**



A first toy model

$\Longrightarrow$ As no mechanism to reduce TE count in a cell, **eventual TE-driven population collapse** ($\approx$ Muller's ratchet)



A first toy model

Can we parametrize the model to generate biologically-relevant numbers of TEs ? Using biologically-relevant parameter values ?

- simulation study ($\Longrightarrow$ Quentin Cordeau)
- theoretical results on the probabilistic model

What are biologically-relevant values ? ($\Longrightarrow$ Luzie Wingen)

Ex: [Bourque et al. (2018)] LINE1 in humans = around 500.000 copies, but most are deactivated, only $\sim$ 100 active per individual

Towards a more complete model

In first model, **no mechanism** to reduce TE count in a cell

POSSIBLE WAYS

– "Soft reduction" by recombination (diploid individuals)

⚠**Need to take into account genome structure**

– Host-mediated deactivation

See LINE1 example

$\Longrightarrow$ "Dormant" and "active" TEs

Towards a more complete model

$\Longrightarrow$ More explicit model of transposition dynamics ?

⚠Huge variety of TEs

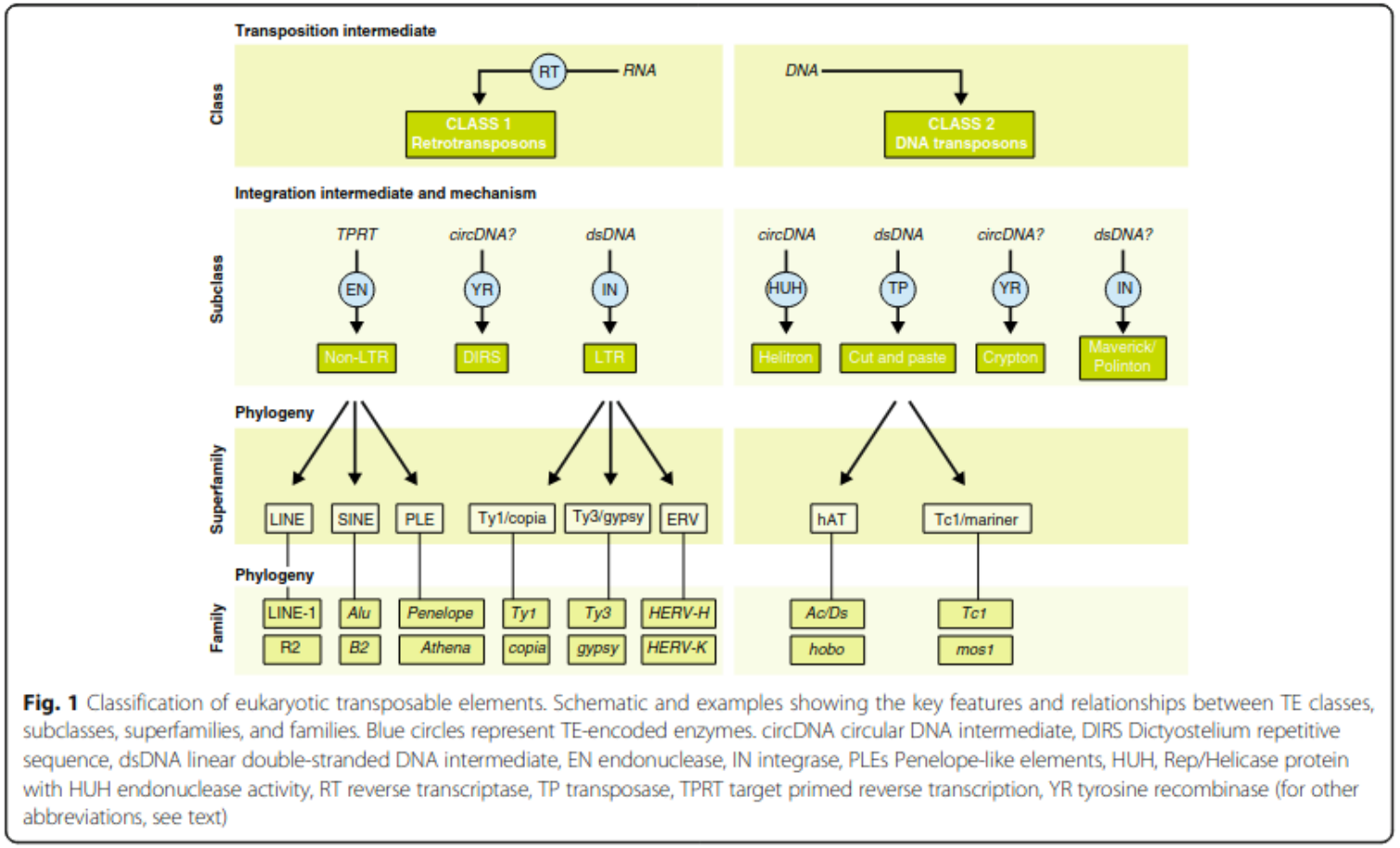


Fig. 1 Classification of eukaryotic transposable elements. Schematic and examples showing the key features and relationships between TE classes, subclasses, superfamilies, and families. Blue circles represent TE-encoded enzymes. circDNA circular DNA intermediate, DIRS Dictyostelium repetitive sequence, dsDNA linear double-stranded DNA intermediate, EN endonuclease, IN integrase, PLEs Penelope-like elements, HUH, Rep/Helicase protein with HUH endonuclease activity, RT reverse transcriptase, TP transposase, TPRT target primed reverse transcription, YR tyrosine recombinase (for other abbreviations, see text)

Image from [Bourque et al. (2018)]

Summary of the project

1. Construct a null model for TE dynamics without self-regulation
2. Parametrize it with biologically-relevant values
3. Study to what extent we can reproduce the documented observed transposition rate
    - Same overall shape, different values ?
    - Same overall shape AND values ?
    - None of the above ?

To learn more about transposons: [Bourque et al. (2018)]

References

Bourque, G., Burns, K., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., ... others (2018). Ten things you should know about transposable elements. *Genome biology*, *19*, 1–12.

Cheek, D., & Johnston, S. (2023). Ancestral reproductive bias in branching processes. *Journal of Mathematical Biology*, *86*(5), 70.

Hallatschek, O., & Nelson, D. (2010). Life at the front of an expanding population. *Evolution: International Journal of Organic Evolution*, *64*(1), 193–206.

References ii

Hubbell, S. (2001). *The unified neutral theory of biodiversity and biogeography*. Princeton: Princeton University Press.

Le Rouzic, A., & Capy, P. (2005). The first steps of transposable elements invasion: parasitic strategy vs. genetic drift. *Genetics*, *169*(2), 1033–1043.