



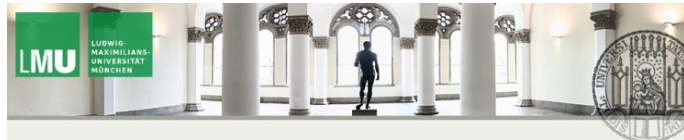
Plant ecology influences population genetics: the role of seed banks in structuring genetic diversity



Aurélien Tellier
Section of Population Genetics
Technische Universität München



Acknowledgements



Wolfgang Stephan
Daniel Živković
Katharina Böndel
Stefan Laurent
Mamadou Mboup
Tetyana Nosenko
Gwenna Breton
Andreas Wollstein
Laura Rose



Technische Universität München 

Johannes Müller

ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

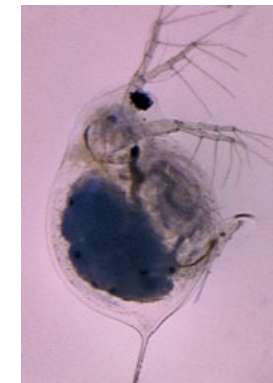
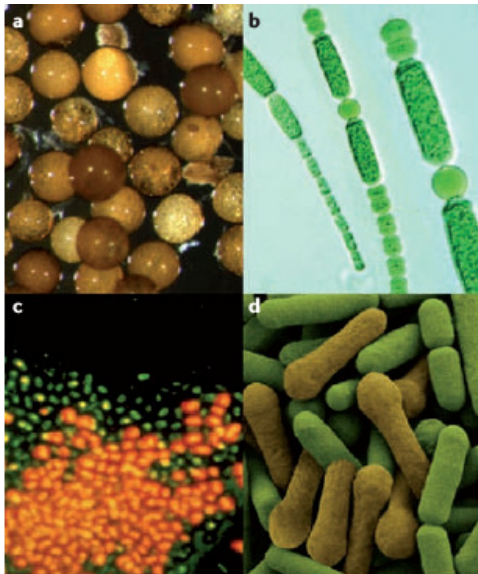
Thomas Städler

 **UNIVERSITÄT
SALZBURG**

Anja Hörger

- **Seed banks may evolve as bet-hedging strategies**
- If the environment is stochastically variable or competition between species generates temporal variability (*Cohen 1966 J Theor Biol; Snyder and Adler 2011 Am Nat*)
- Bet hedging strategies such as germ banking are ubiquitous to many species of:
 - **Bacteria** (*Jones and Lennon 2011 Nat Rev Microbiol*)
 - **Invertebrates**: diapause in insects, eggs banks in crustaceans (*Daphnia*)

a) Arbuscular mycorrhizal fungus, b) Cyanobacteria
c) Bacterial biofilm (*P. aeruginosa*), d) Bacteria *Viridibacillus arvi*



© J. Wolinska

Host-parasite 'Red Queen' dynamics archived in pond sediment

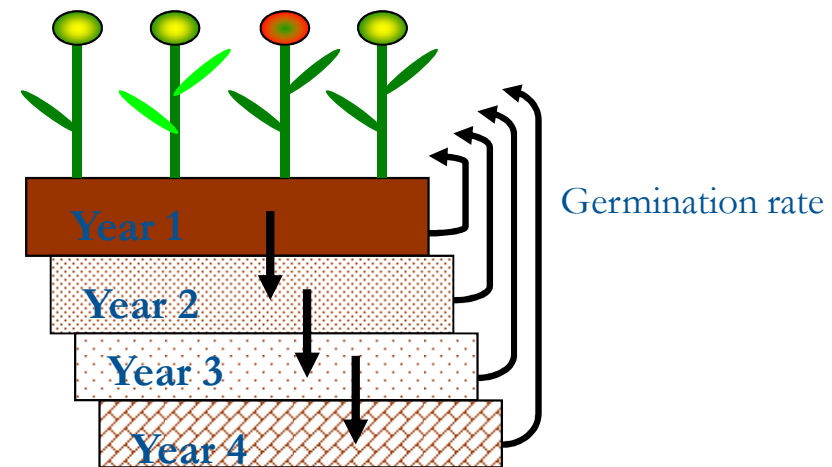
Ellen Decaestecker^{1,3}, Sabrina Gaba^{4,5}, Joost A. M. Raeymaekers^{1,2}, Robby Stoks¹, Liesbeth Van Kerckhoven¹, Dieter Ebert^{4*} & Luc De Meester^{1*}

Nature 2007

- **Seed banks are important for conservation biology**
- Promote the temporal rescue effect (*Brown and Kodric-Brown 1977 Ecology*)
- **Seed banks promote storage of diversity in the soil** => time lag between above-ground and seed banks
 - Selection is slower (*Hairston and De Stasio 1988 Nature*)
 - Balancing selection is favored (*Turelli et al. 2001 Evolution*)
 - Coevolutionary dynamics are stabilized (*Tellier and Brown 2009 Am Nat*)



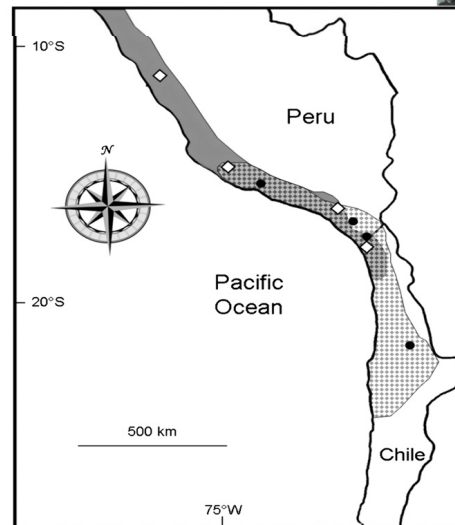
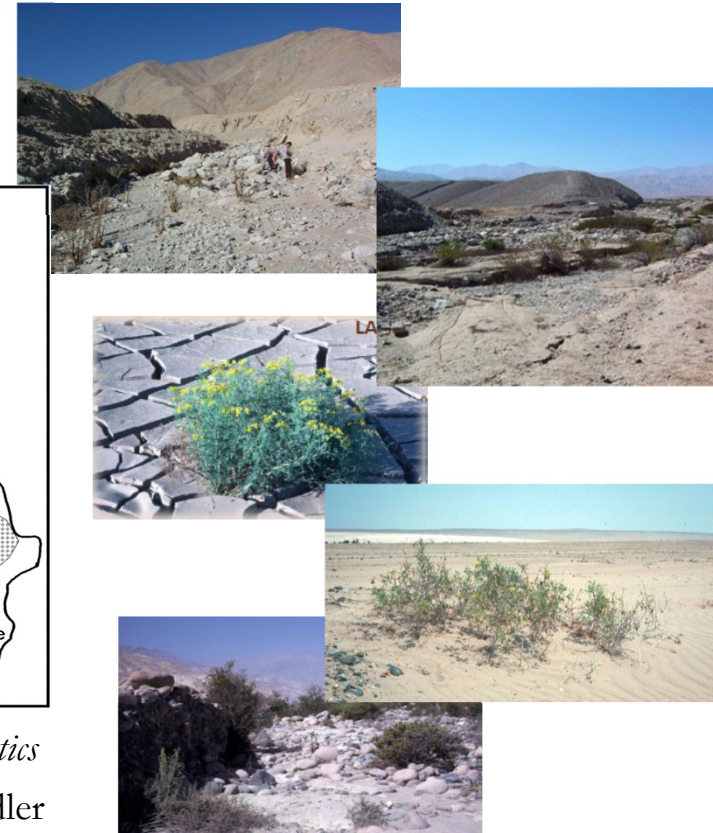
Linanthus parryae ©Bierzychudek lab



Solanum peruvianum



Solanum chilense



Städler et al. 2008 Genetics

Photos © TGRC. T. Städler

How to explain the very large genetic diversity observed despite small census sizes?



Overview

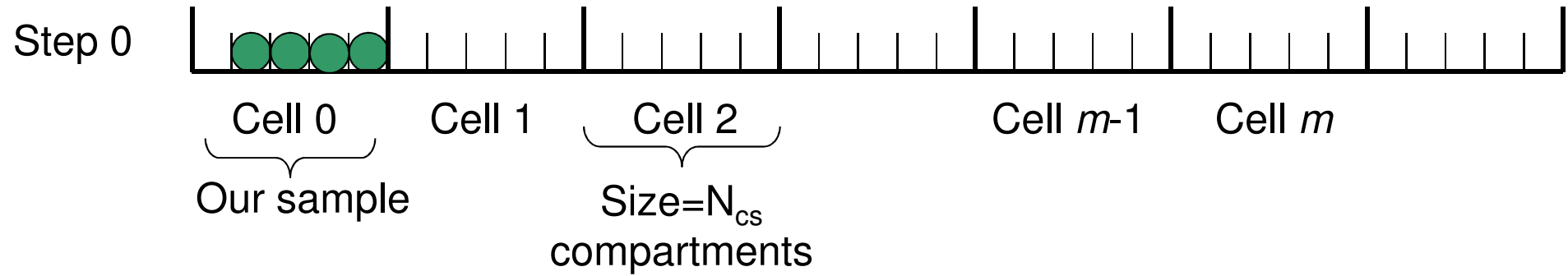
- 1) Seed banks models and evolutionary inference
 - 2) Heterogeneity in seed banks and adaptation
 - 3) Work in progress and open questions
-



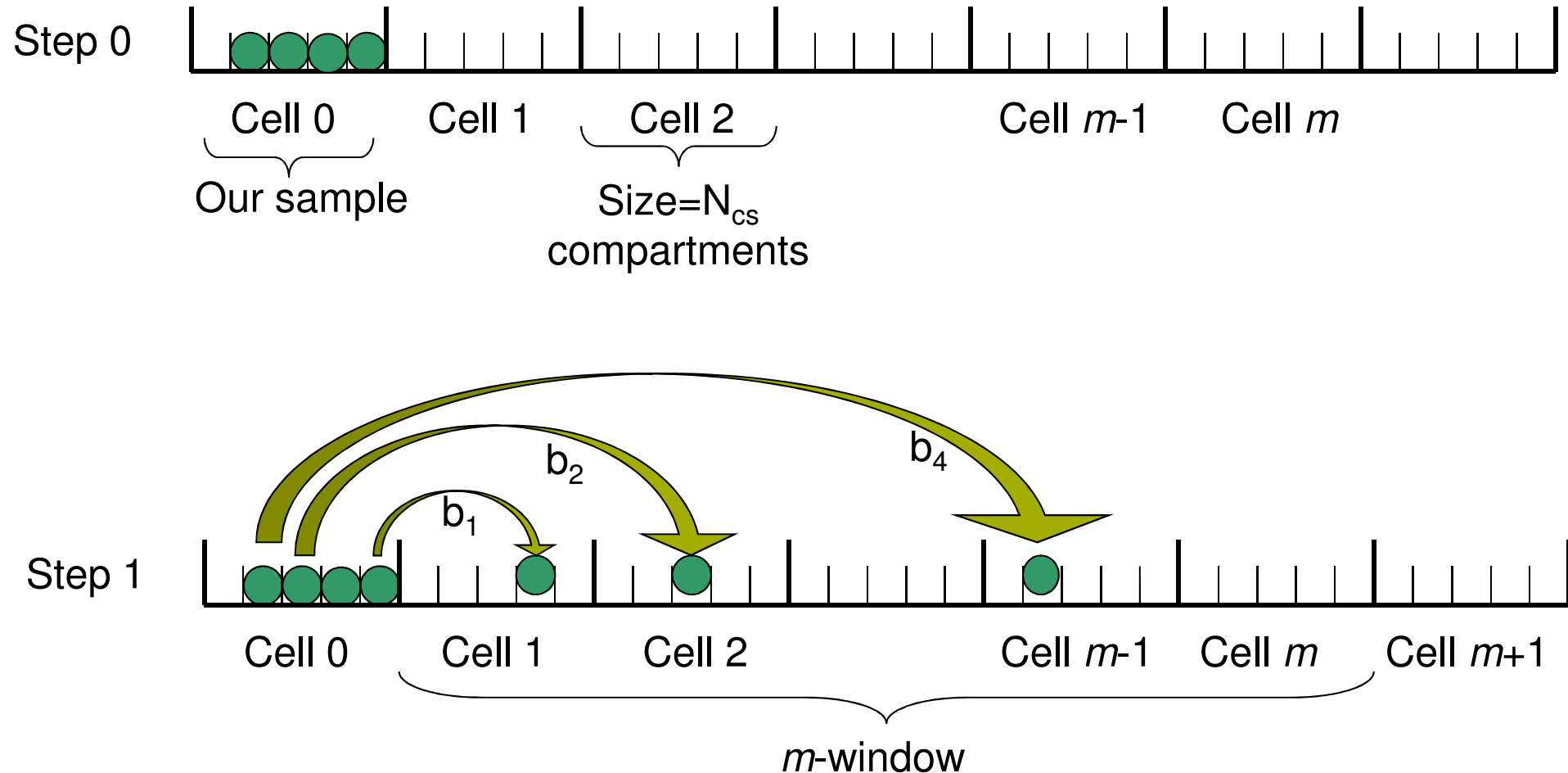
1) Seed banks model and inference



If we have no information about what lies in the seed bank. here $m = 5$

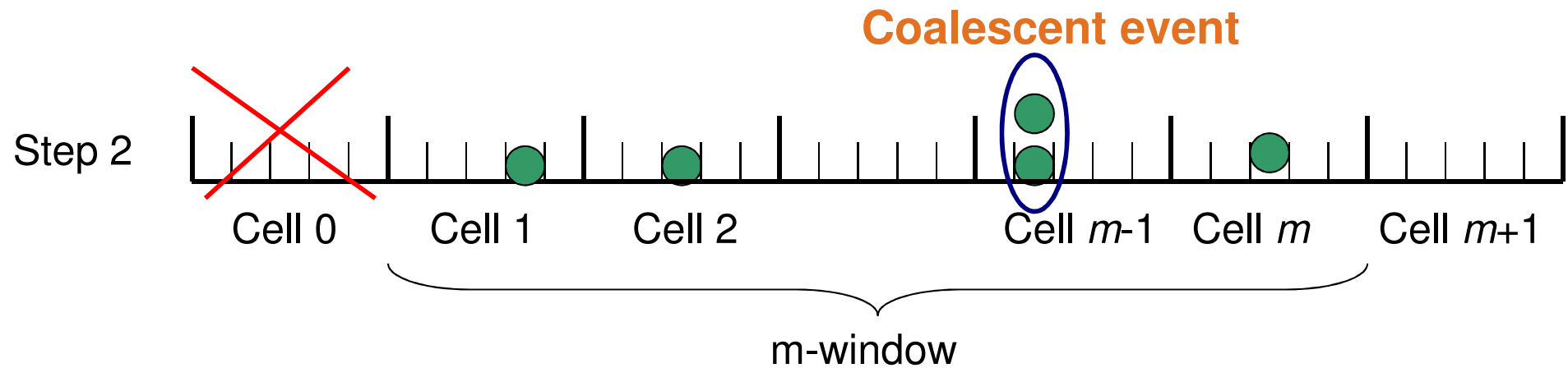


If we have no information about what lies in the seed bank. here $m = 5$



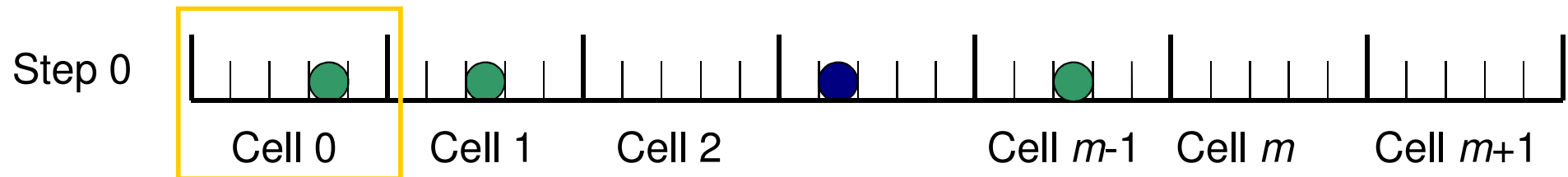
For all plants in cell 0

$m = 5$. if two seeds fall on the same compartment within a cell. there can be a coalescent event



At any moment. there are r lineages in the m -window

Then move every cell to the left. and start again step 0





Our main result states that the limit of the time-rescaled ancestral process, $A_N([Nt])$, as $N \rightarrow \infty$, is given by Kingman's n -coalescent run on a slower time-scale.

➤ *The coalescent rate when there are r ancestors is: $\beta^2 \binom{r}{2}$*

➤ *With $\beta = \frac{\sum_{i=1}^m b_i}{\sum_{i=1}^m i b_i}$*

➤ β is a composite seed bank parameter function of b_i and m

➤ *We assume small finite $m \Rightarrow$ weak seed bank from Blath et al.*



- The germination process is memoryless => the germination rate decreases geometrically with age of seeds
 - $b_1 = b$ = probability for a seed to germinate after one generation (= germination rate)
 - $b_i = b(1-b)^{i-1}$

 - The rate of coalescence is rescaled by β^2 (**the size of the genealogy is affected**)
 - Mutation does not increase with age of seeds
 - The scaled mutation rate is scaled by β along a given ancestral line
 - The recombination rate and migration rate between demes are also scaled by β
-



Result 1:

- **Show the existence of seed banks over evolutionary time scale without extensive sampling of seeds and above-ground populations**
- *e.g.* for species where sampling is complicated

Result 2:

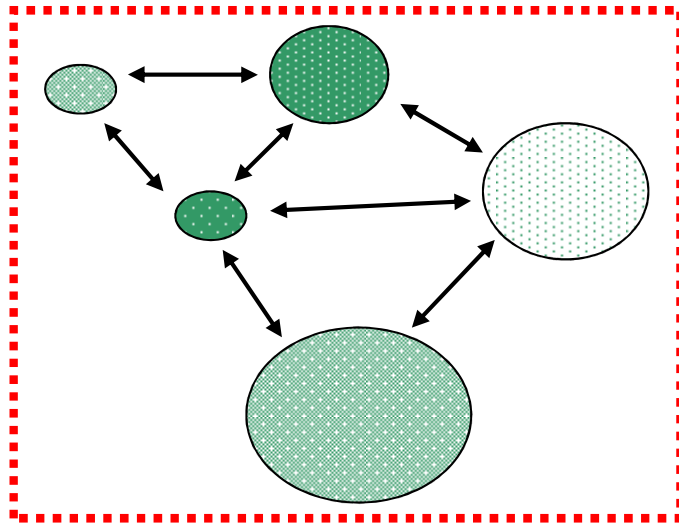
- **What are the consequences of seed banks on statistical inference of past demographic events?**
- Of interest in conservation biology where DNA sequences are increasingly used to detect past or recent crash of populations?

Result 3:

- **What are the consequences of seed banks on statistical inference for speciation models?**
 - Can seed banks affect the detection of introgression between closely related species?
-

Result 1

Spatial structuring of populations



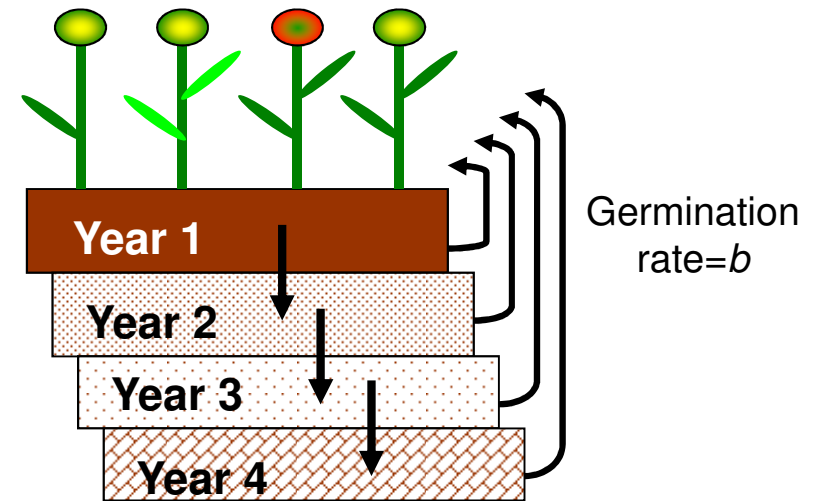
$$N_e = f(N_{cs}, n_d, mig)$$

Number of
individuals per
population

Number of
demes

Migration rate

Seed banks in plants



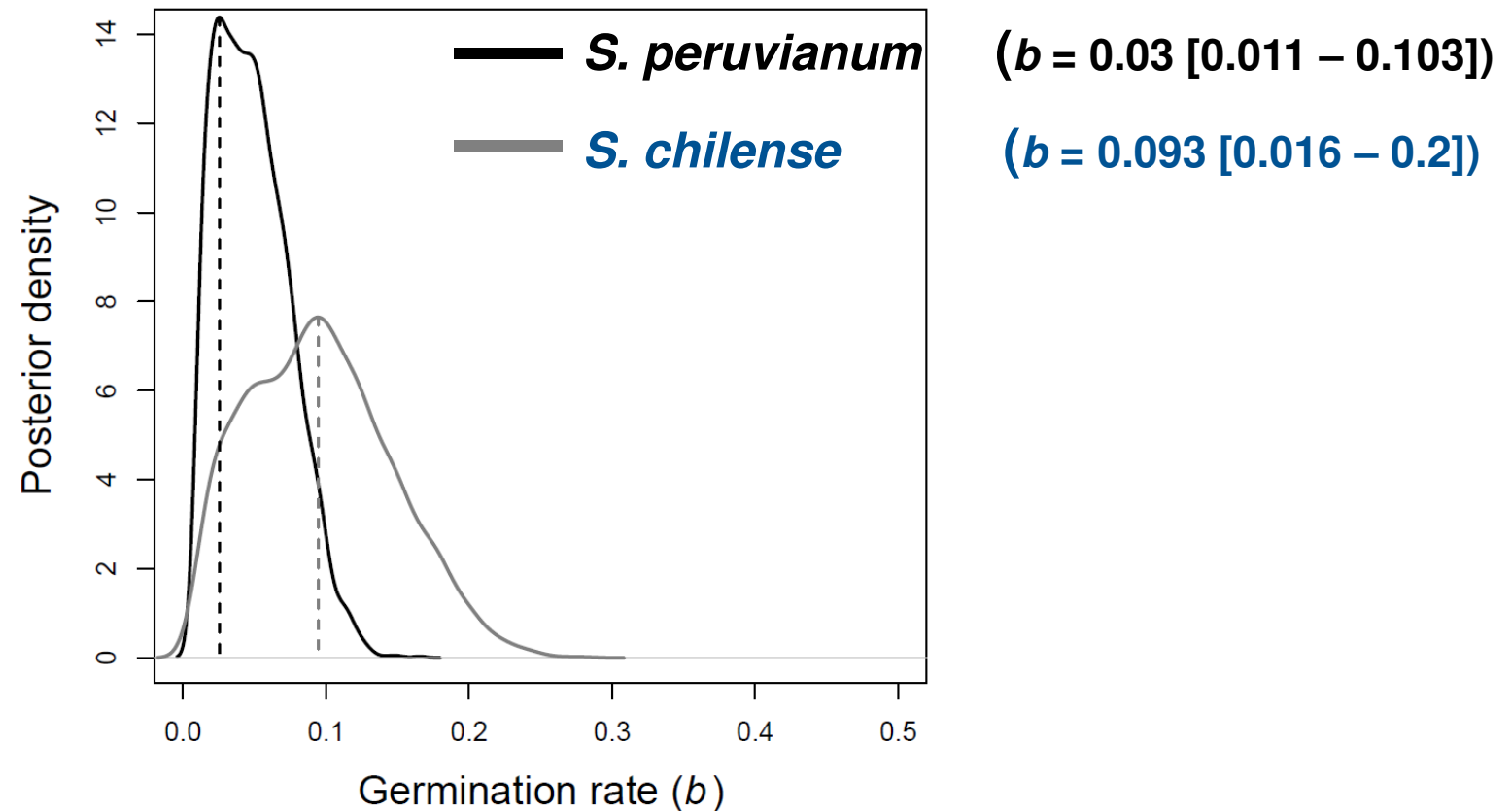
$$N_e = f(N_{cs}, b)$$

Number of
individuals per
population

Germination
rate

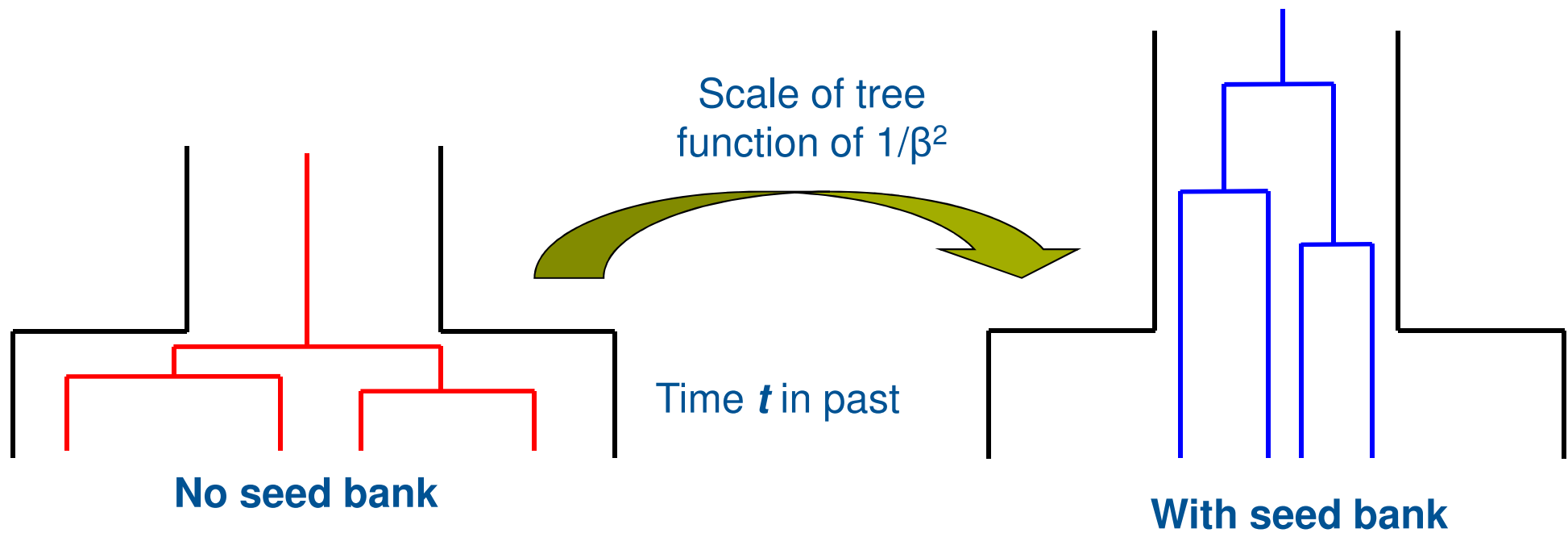
Result 1

Posterior density for the germination rate b



Estimate b for both species using priors on N_{cs}

Result 2



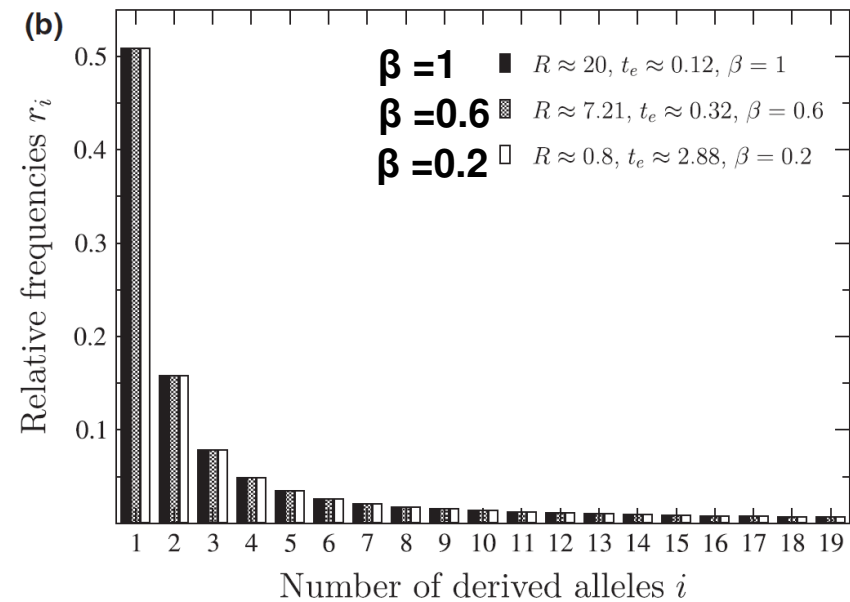
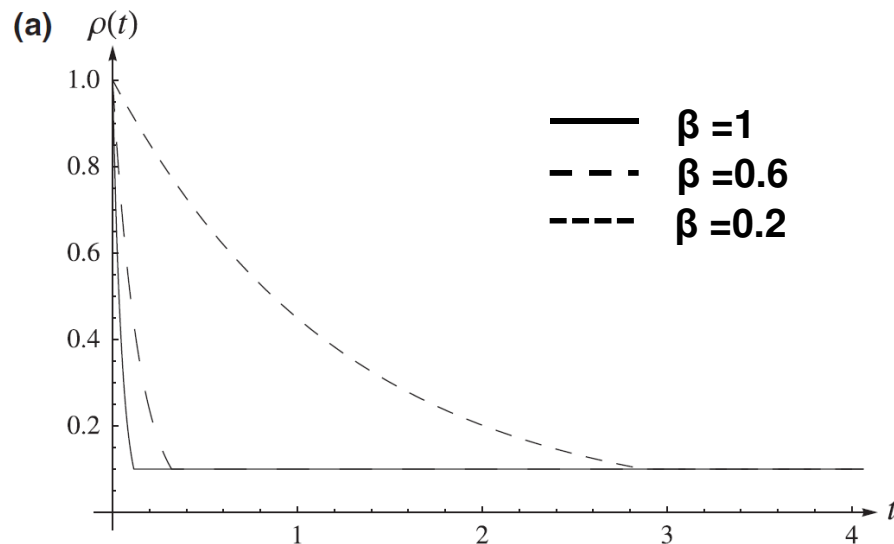
Result 2

- We compute the SFS for a model with seed bank and varying population size

$$P(A_n^\rho(t) = j) = \sum_{k=j}^n c_{nk} r_{kj} \exp\left(-\beta^2 \binom{k}{2} \int_0^t \rho(s)^{-1} ds\right)$$

$$E(T_j) = \int_0^\infty P(A_n^\rho(t) = j) dt$$

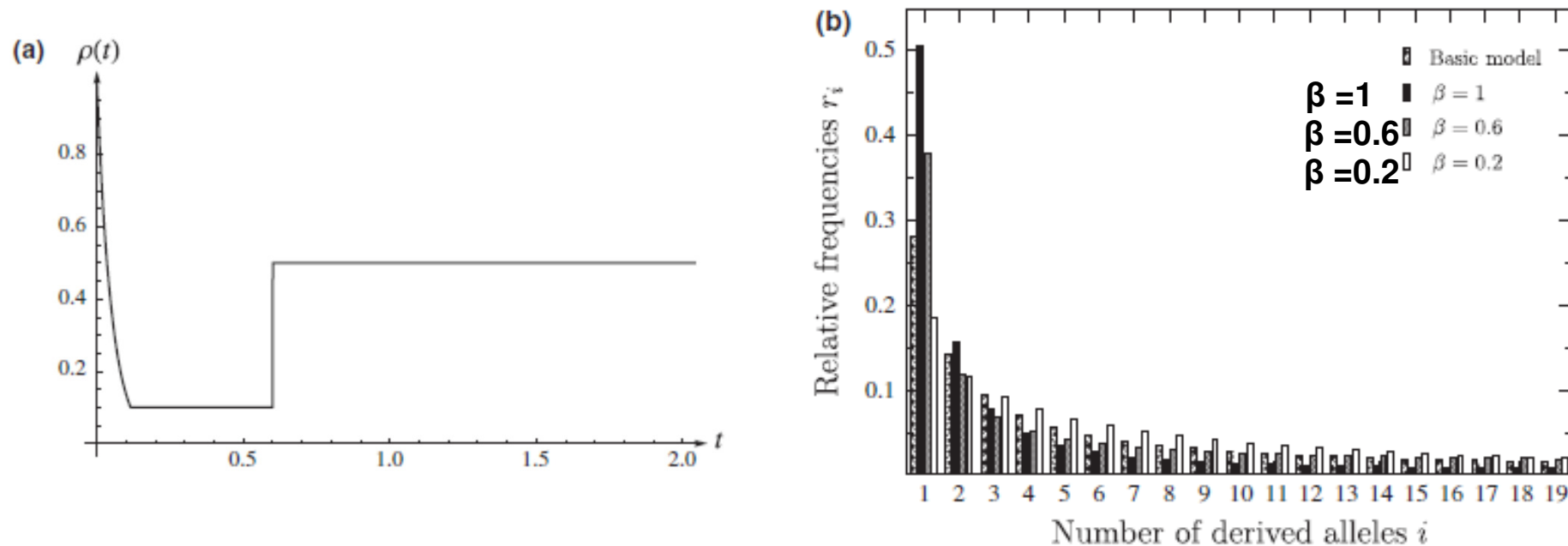
$$r_i = \sum_{k=2}^{n-i+1} k \frac{\binom{n-i-1}{k-2}}{\binom{n-1}{k-1}} E(T_k) / \sum_{k=2}^n k E(T_k)$$



Different past demography can result in the same allele frequency spectrum

Result 2

- For more complex demographic models, it may even be worst

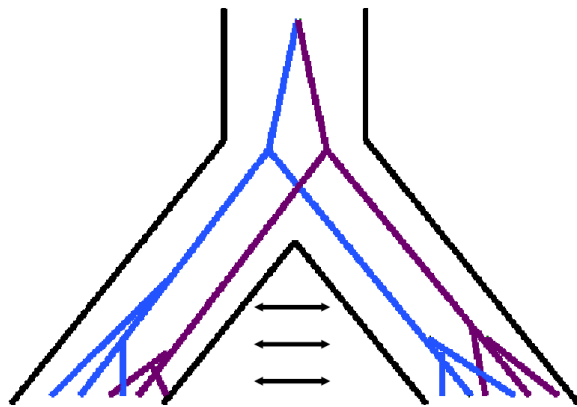


Good news: seed bank gives access to more ancient demographic events

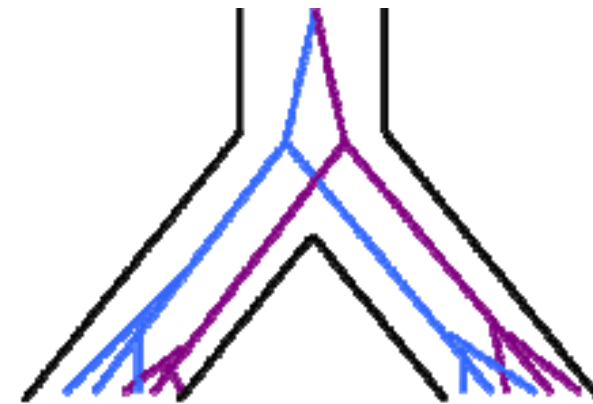
Bad news: we need to know the germination rate

Result 3

- Coevolution may generate balancing selection at resistance genes in plants (*Stahl et al. 1999 Nature; Tellier and Brown 2011 Annu Rev Phytopathol*)
- Under balancing selection with multiple alleles (*Castric et al. 2008 PLoS Genetics*) =>
 - Repeated adaptive introgression due to frequency-dependent selection?
 - Maintenance of ancestral polymorphism without gene flow?



Adaptive introgression model



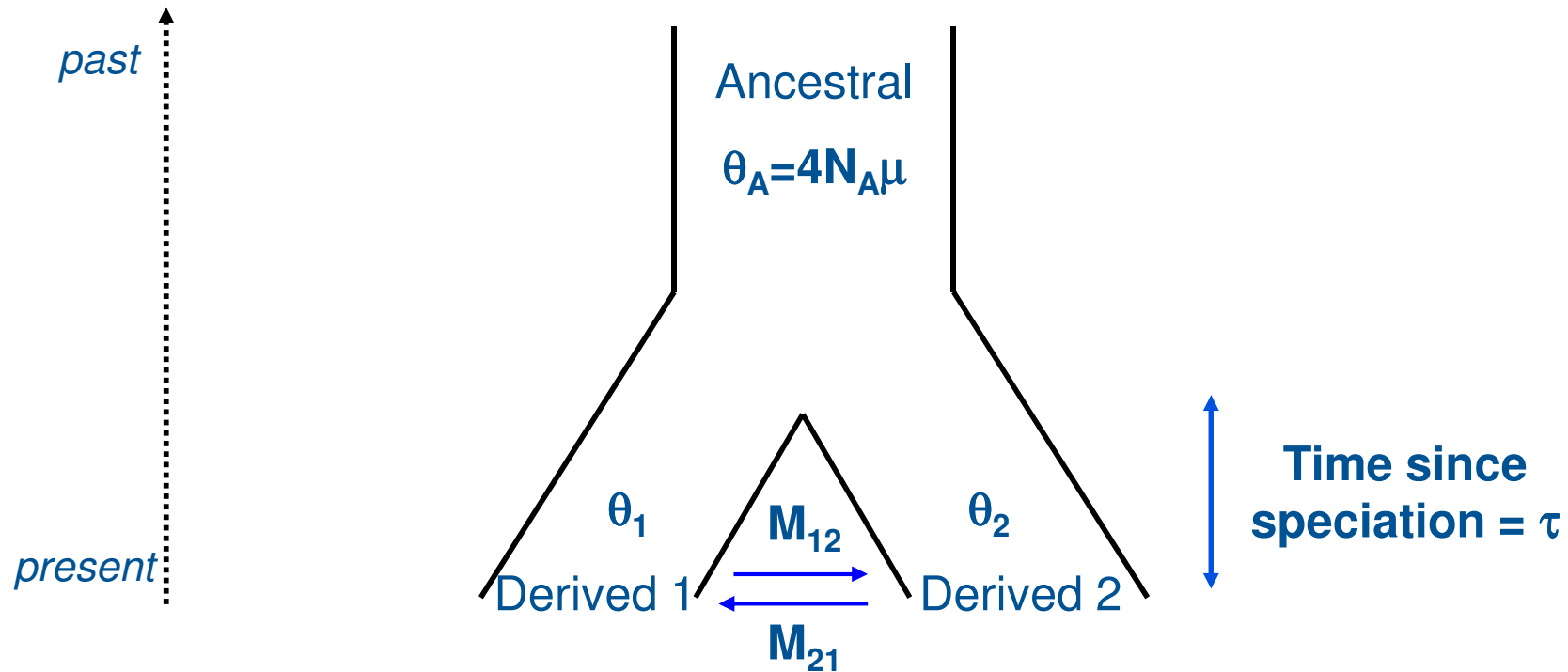
Ancestral polymorphism model

Result 3

- We study three resistance genes: *Pto*, *Pfi1*, *Rin4* (Rose et al. 2005 *Genetics*, Rose et al. 2011 *Mol Plant Pathol*)
- In two sister species *S. chilense* and *S. peruvianum* with history of gene flow
- 10 reference loci are sequenced (=980 SNPs)
- Species wide sampling: one individual per population



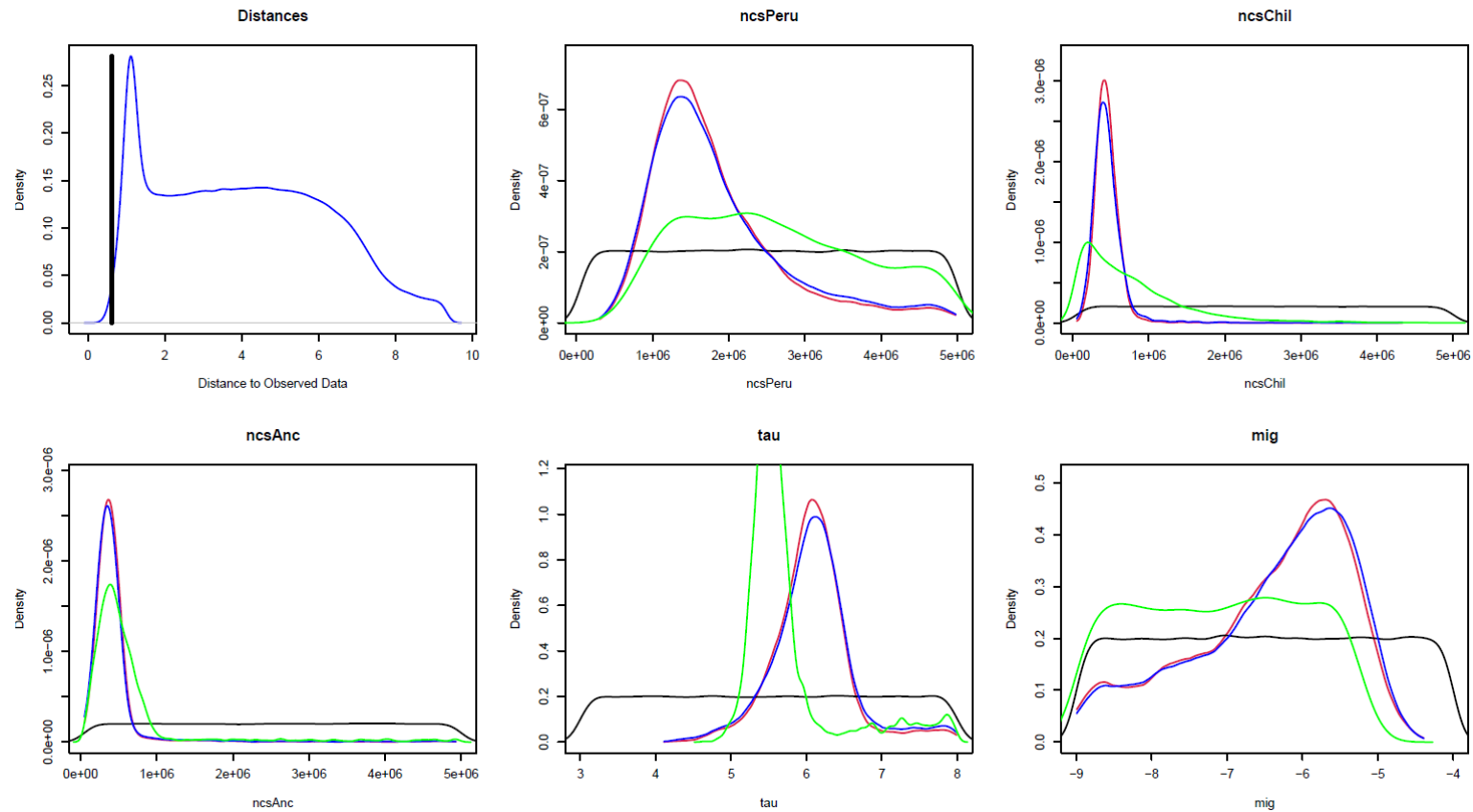
Result 3



The number of migrants per generation

- Use ABC + Joint-Site Frequency Spectrum frequency to summarize SNPs

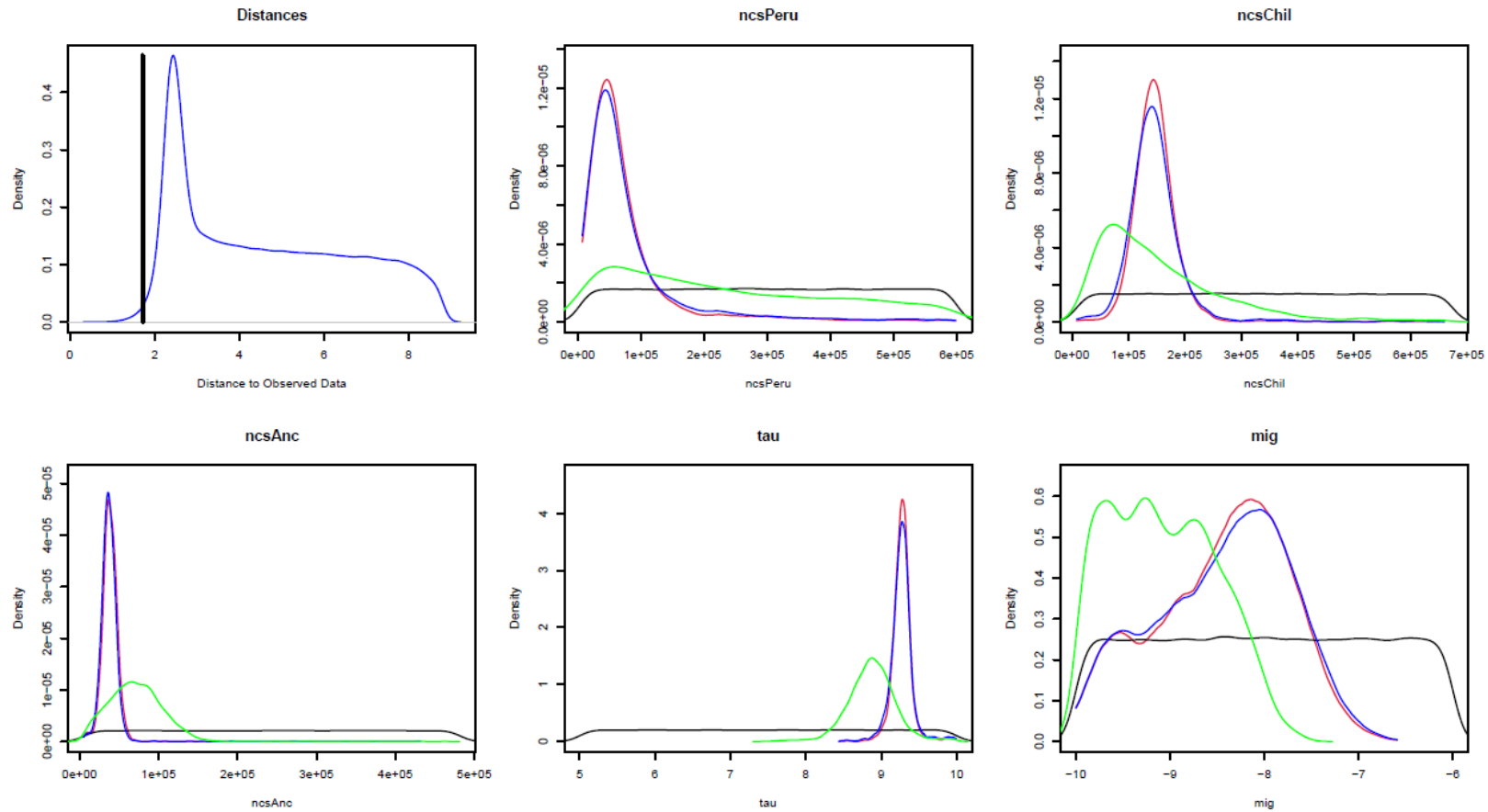
Result 3: without seed banks



Time divergence = 1.6×10^6
Migration estimate = 1.73×10^{-6}

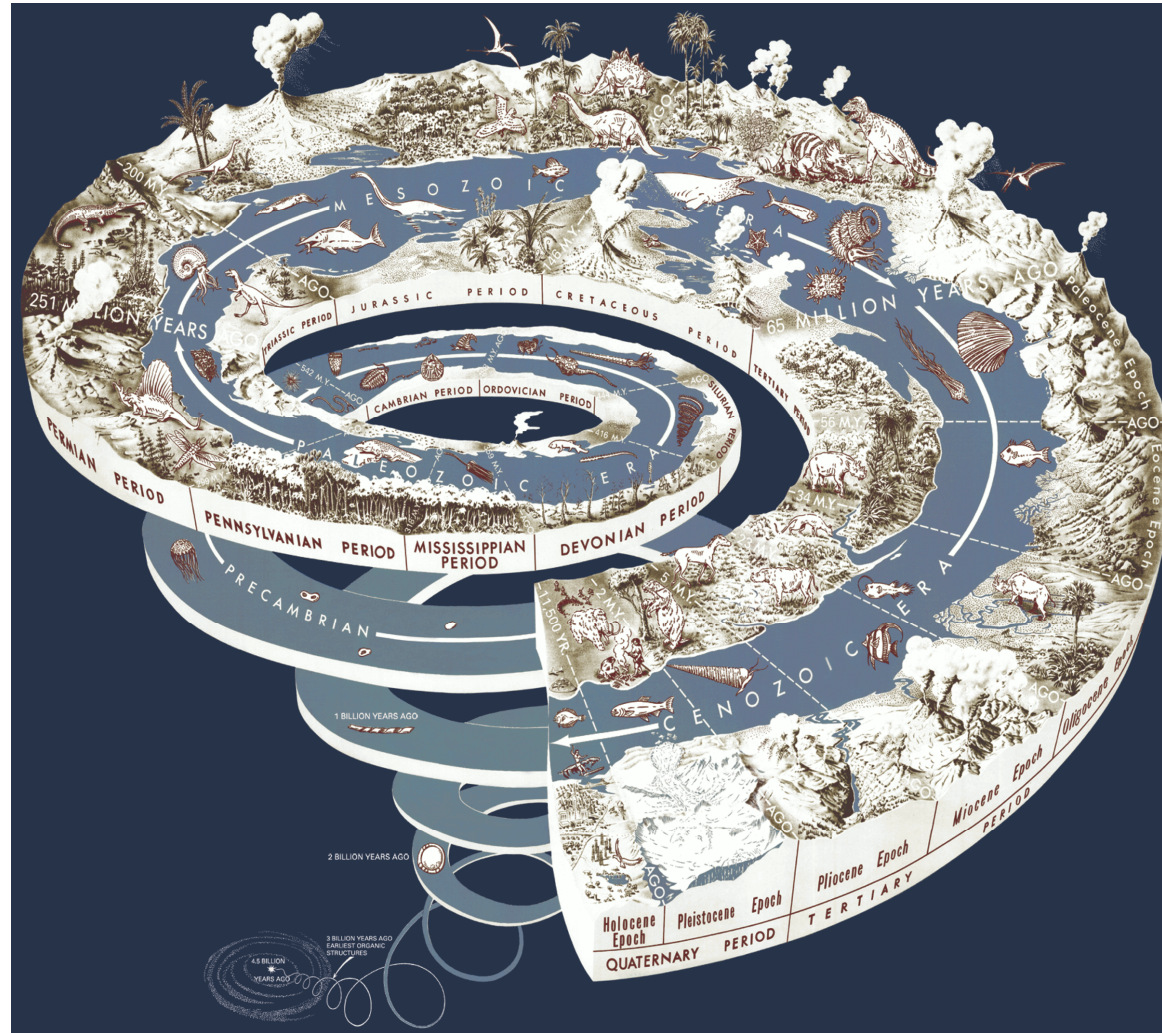
Similar to previous results

Result 3: with seed banks



Time divergence $> 10^9$
Migration estimate $< 10^{-8}$

Time divergence $> 10^9$
Migration estimate $< 10^{-8}$



Oups....

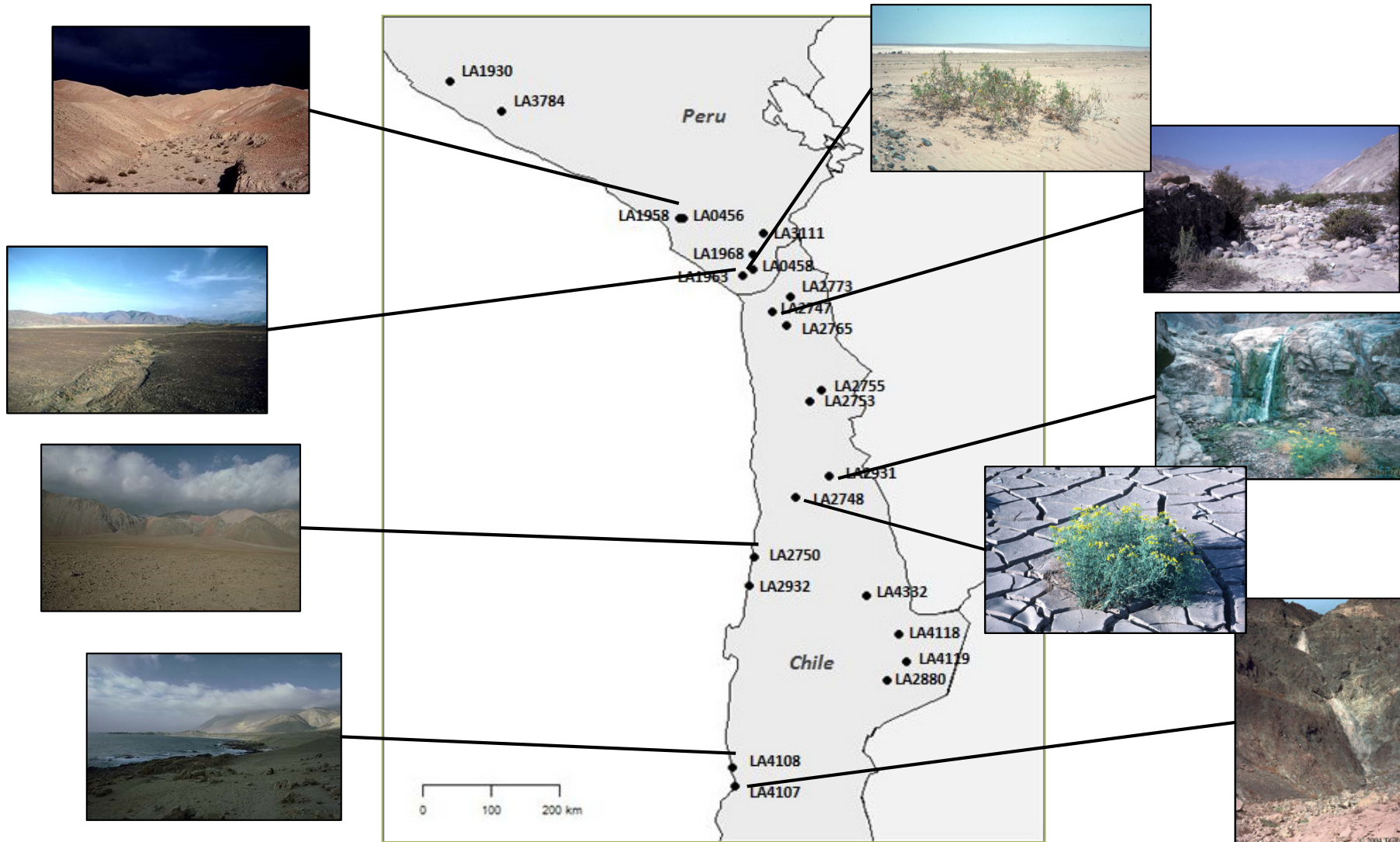


Part 2:

***Solanum chilense* demography and heterogeneity in seed banks**

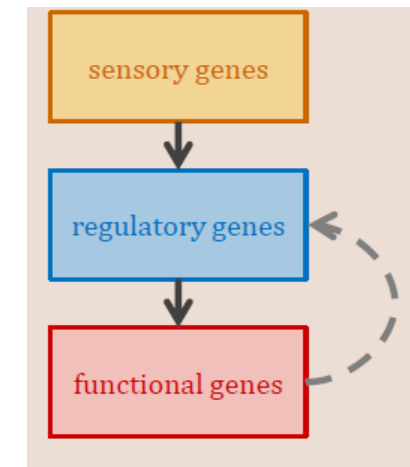
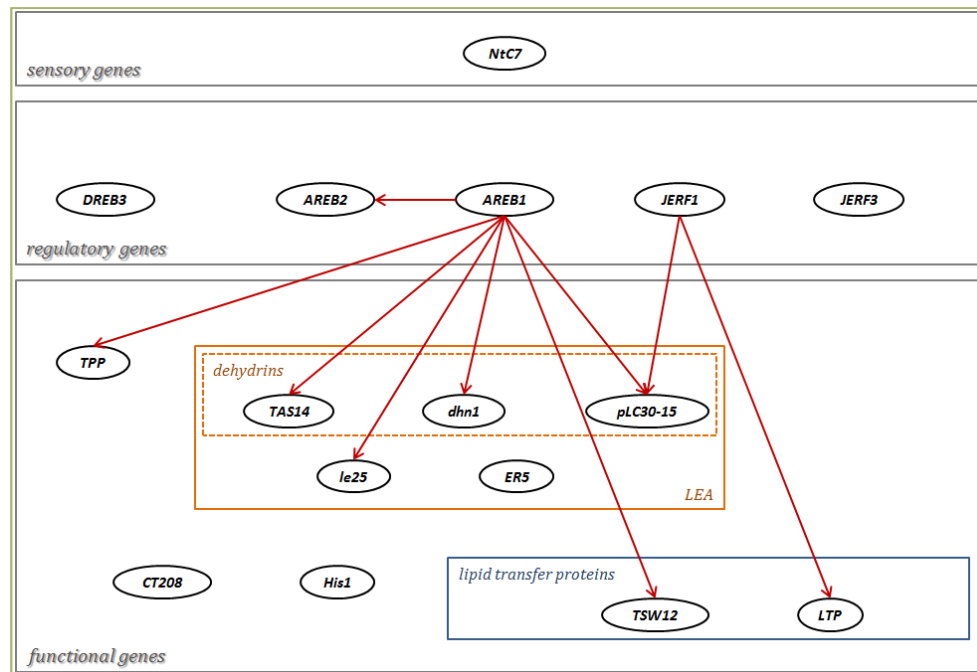
Böndel et al. submitted

S. chilense sampling: habitats



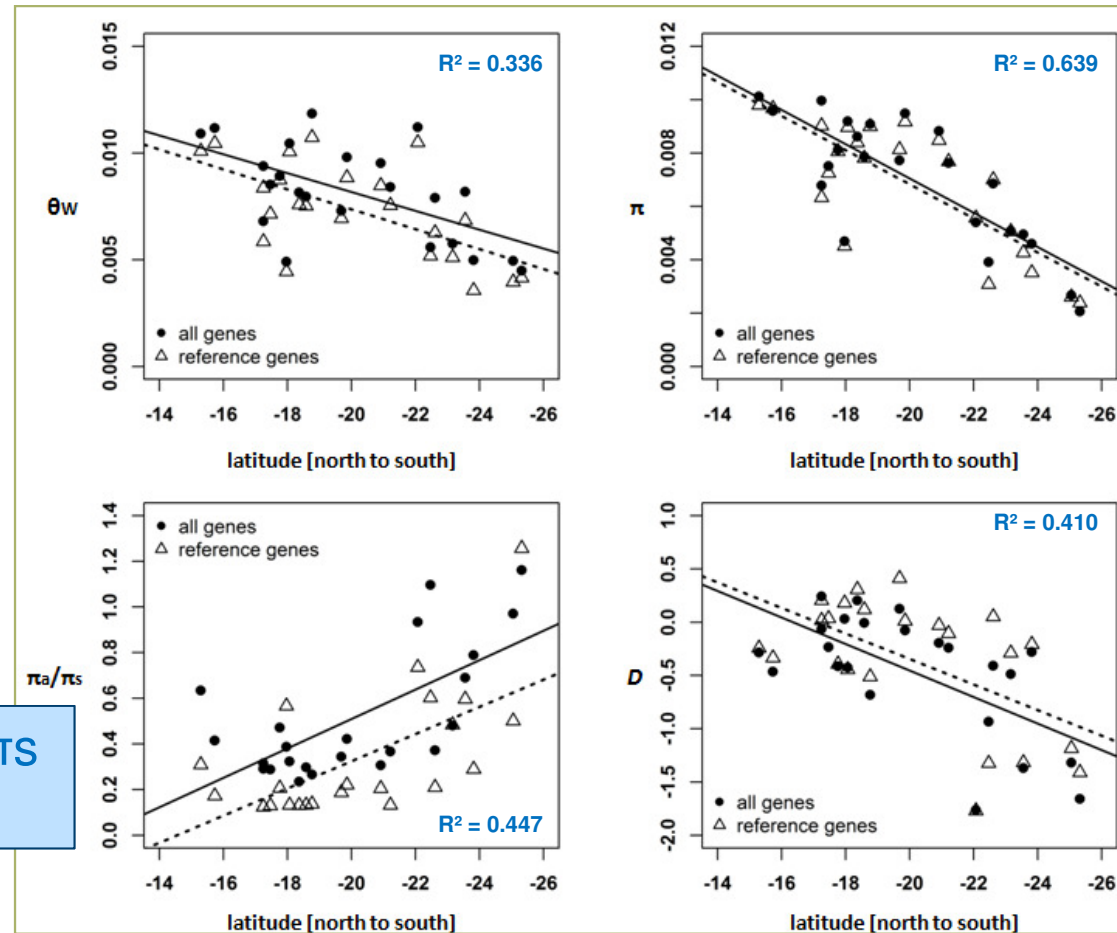
S. chilense data set

- 23 populations, 25 diploid plants per population
- 30 genes (16 candidate genes, 14 reference genes)
 - candidate genes: related to abiotic stress for selection/local adaptation



Results: North to South colonization

decrease of variation from N to S



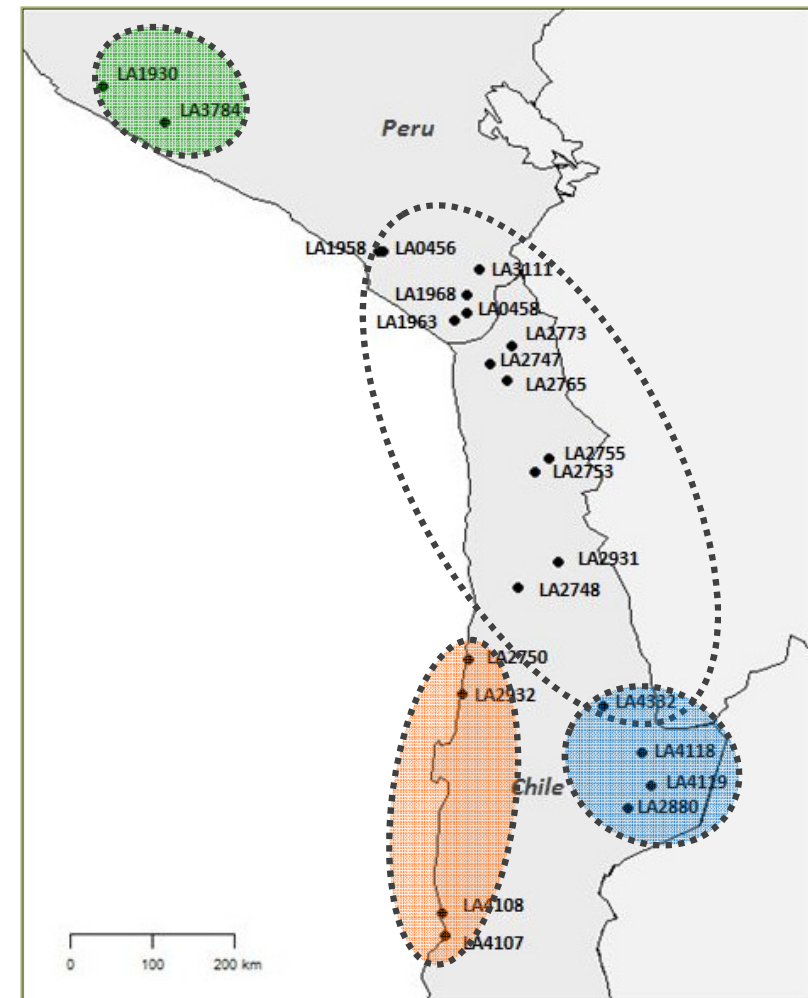
increase of π_a/π_s
from N to S

decrease of Tajima's D from N to S

Note: given R^2 values are for all genes

Results: North to South colonization

We define 4 different genetic groups
→ north-south colonization, possibly
through (at least) three bottlenecks



Results: Selection and adaptation

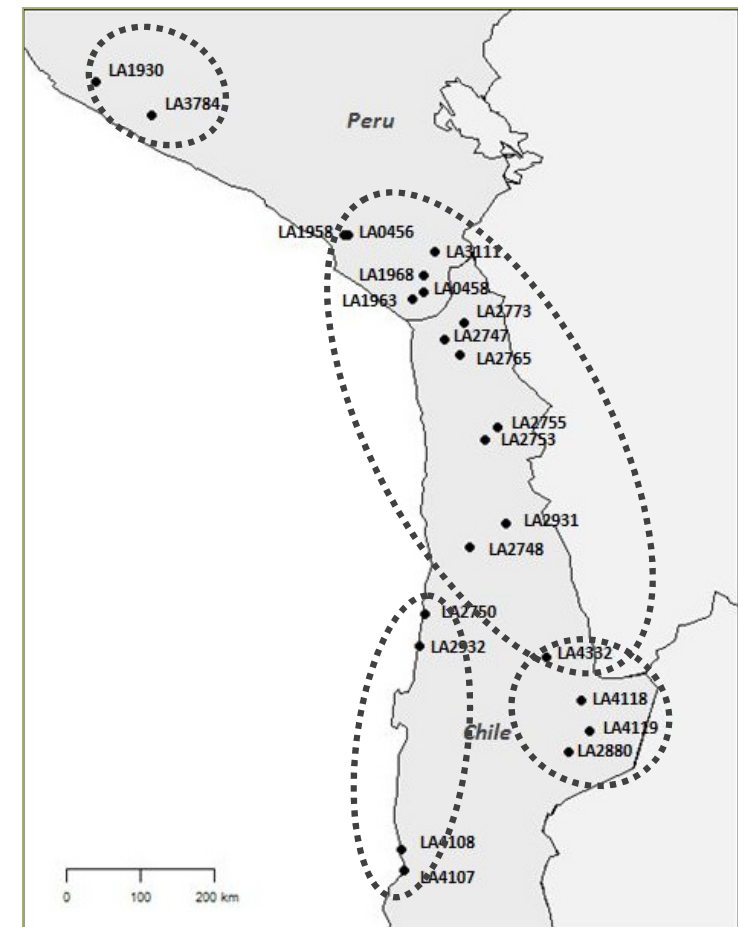
Methods to detect signatures of selection

- candidate gene approach (summary statistics)
- McDonald-Kreitman test (two outgroup species)
- BayeScan (F_{ST} outlier approach)

→ **Stronger purifying selection in North populations**

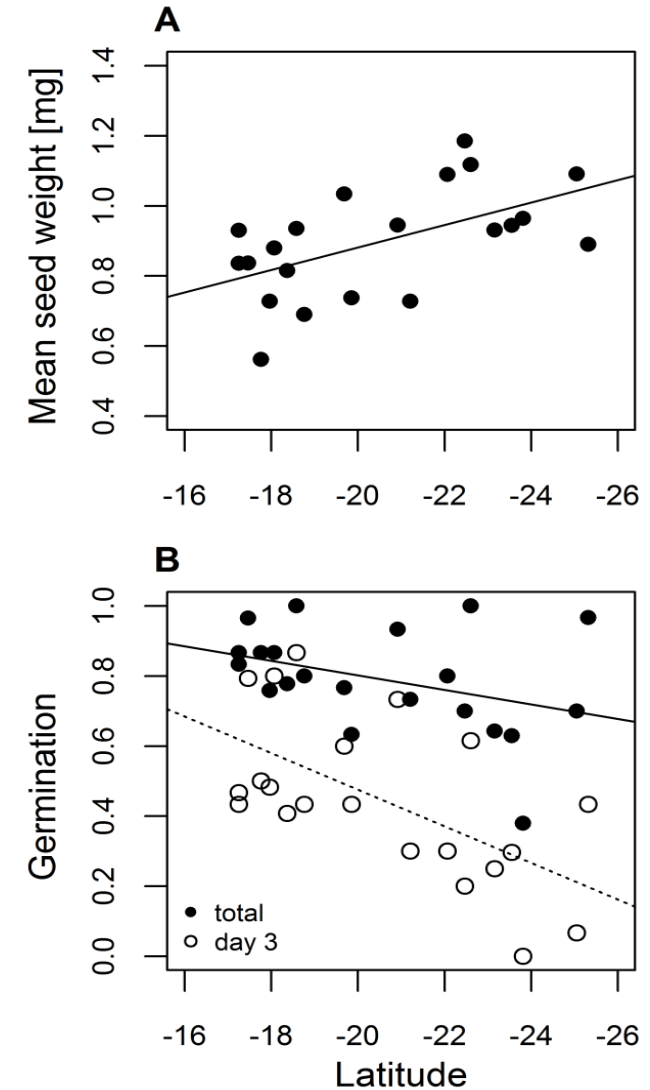
→ **Local adaptation more often in peripheral groups (Southern populations)**

→ Hypothesis that colonization of new territories is associated with an initial local adaptation event (e.g. Innan & Kim 2006)



Conclusion

- North to South colonization with adaptation to new habitats
- Role of seed banks? if North populations are more ancestral (with smaller seed size)
- Adaptation for bigger seeds in South => shorter seed bank (higher b) or decrease of m ?
- => stronger drift and smaller N ?



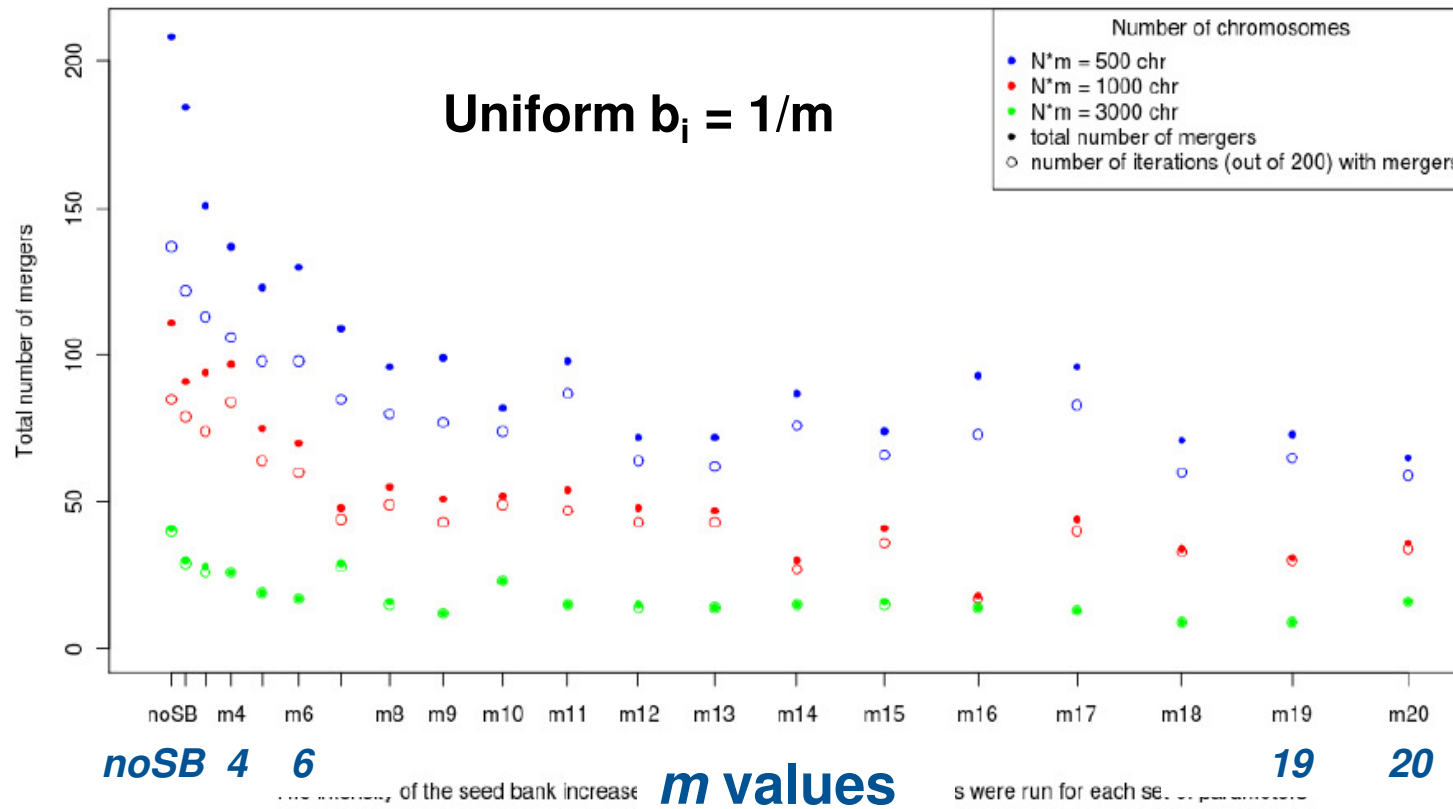


Part 3: Preliminary additional results

work in progress

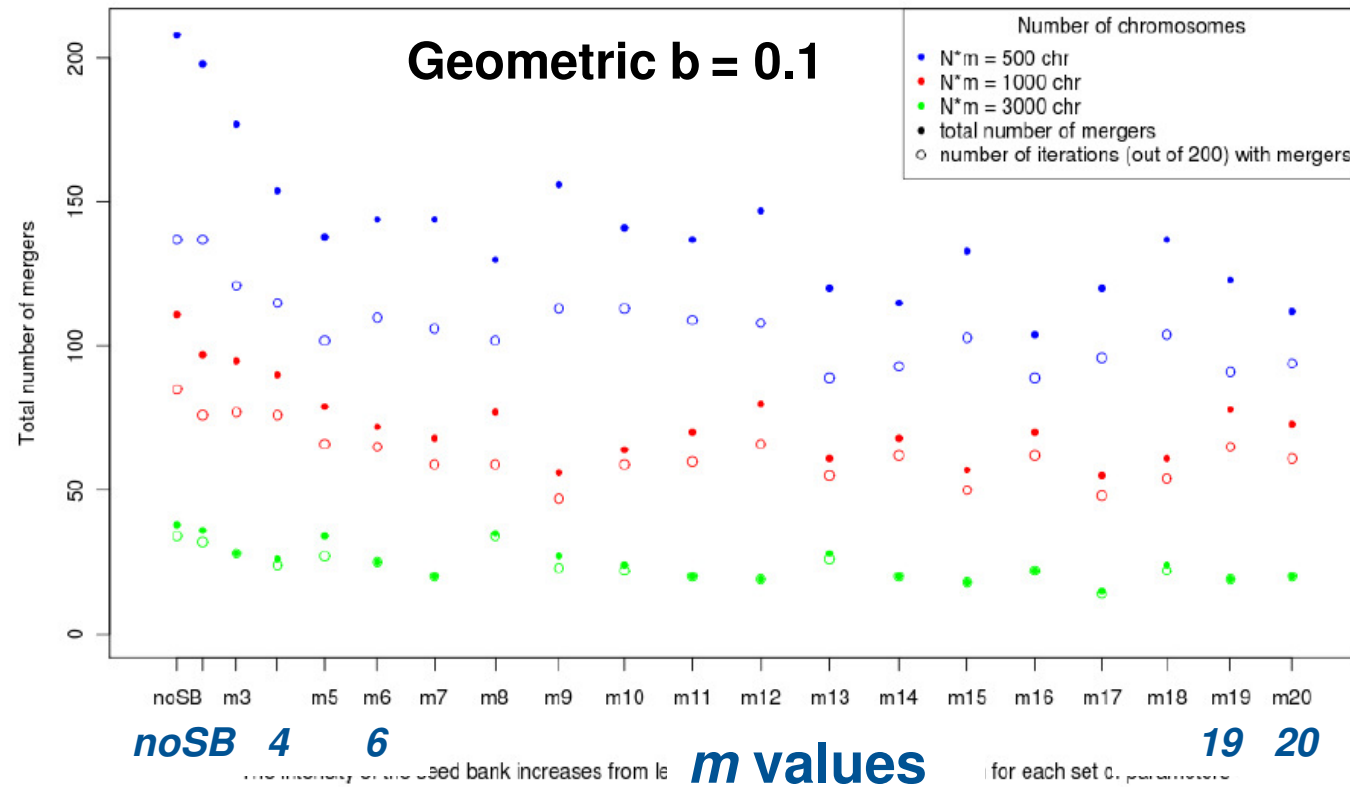
WIP 1: how do we know N_{CS} ?

- We develop a forward simulator of the Kaj et al. Model
- Number of multiple and simultaneous mergers $n=20$, $m \cdot N_{CS}=500, 1000$ or 3000



WIP 1: how do we know N_{CS} ?

- Number of multiple and simultaneous mergers $n=20$, $m \cdot N_{CS}=500, 1000$ or 3000
- With large sample size n , one may be able to disentangle m and N_{CS}



WIP 2: a forward model

3 Time Continuous Seedbank Model with Deterministic Seedbank

3.1 Individual based model

We have a population of N plants (population size is fixed). There are two genotypes, A and B. State of the model at time t :

(1) Plants: $X_t \in \{0, \dots, N\}$ denotes the number of A-type plants, $N - X_t$ those of B-type.

(2) Seeds: The number of (living) seeds in the seedbank is $A_t \in \mathbb{N}_0$, the number of seeds (type B) is $B_t \in \mathbb{N}_0$.

The seeds are described by their age at time t ; let

$$\mathcal{A}_t = \{a_i(t) \mid i = 1, \dots, A_t\}, \quad \mathcal{B}_t = \{a_i(t) \mid i = 1, \dots, B_t\}$$

the ages of the seeds (of genotype A resp. B).

The dynamics consists of (1) death and re-birth of a plant, (2) creation of seeds, and (3) death of seeds.

WIP 2: a forward model

(1) Death of a plant: At rate μ , plants die. They are immediately replaced by germinating seeds, s.t. the total number of plants is constant. The probability for a plant to be type A (or type B) depends on the seeds in the seedbank. Let $b(a)$ the ability of a seed to germinate. The total ability for germination of A-plants (of B-plants) reads

$$\sum_{a \in \mathcal{A}_t} b(a), \quad \sum_{a \in \mathcal{B}_t} b(a).$$

- Note that $b(a)$ is smooth, and has a bounded support.
- Also we do not remove the seed from the bank, assuming that plants produce enough seeds of one type, and few germinate.

(2) Production of seeds

Living plants produce seeds at rate β . That is, the “birthdays” of seeds of A-type are given by an inhomogeneous Poisson process with intensity βX_t . Similarly, that for B-seeds is given by $\beta(N - X_t)$. Please note already here, that the production of any seeds (of either type) follow a Poisson process with constant intensity βN , independent on X_t .

(3) Death of seeds.

Seeds die at rate $\mu_s(a)$.

WIP 2: a forward model

Now we assume that the seeds are numerous. That is, conditioned on the history of X_τ for $\tau \leq t$, we know that βX_τ seeds of *A*-type, resp. $\beta(N - X_\tau)$ seeds of *B*-type have been produced at time $t - \tau$. If we interpret $\tilde{b}(a)$ as the probability that a seed of age a is still alive (better: able to germinate), we know that we have $\int_0^\infty \tilde{b}(\tau) \beta X_{t-\tau} d\tau$ seeds of type A that can germinate, and $\int_0^\infty \tilde{b}(\tau) \beta(N - X_{t-\tau}) d\tau$ seeds of type B that can germinate. Hence, the amount of living seeds (type A plus type B) in the seedbank is given by $\int_0^\infty \tilde{b}(\tau) \beta N d\tau$.

If a plant dies, the probability that this plant is replaced by a type-A-plant

$$\frac{\int_0^\infty \tilde{b}(\tau) \beta X_{t-\tau} d\tau}{\int_0^\infty \tilde{b}(\tau) \beta N d\tau} = \int_0^\infty \frac{\tilde{b}(\tau)}{\int_0^\infty \tilde{b}(s) ds} X_{t-\tau} / N d\tau.$$

We define the distribution

$$b(a) = \tilde{b}(a) / \int_0^\infty \tilde{b}(s) ds$$

and find the probability for a type-A-plant given by

$$\int_0^\infty b(\tau) X_{t-\tau} / N d\tau.$$

WIP 2: a forward model

Summary: Our individual based model (that assumes a deterministic seedbank) reads

$$P(X_{t+\Delta t} = X_t + 1 | X_\tau \text{ for } \tau \leq t) = \Delta t \mu N (1 - X_t/N) \int_0^\infty b(\tau) X_{t-\tau}/N d\tau + \mathcal{O}(\Delta t)$$

$$P(X_{t+\Delta t} = X_t - 1 | X_\tau \text{ for } \tau \leq t) = \Delta t \mu N (X_t/N) \left(1 - \int_0^\infty b(\tau) X_{t-\tau}/N d\tau\right) + \mathcal{O}(\Delta t).$$

$$X_{t+\Delta t} = X_t + Y_{+,t} - Y_{-,t}$$

where we used $\int_0^\infty b(s) ds = 1$ in the last step. Hence, we are led to the SDDE (with $\varepsilon = 1/N$)

$$dx_t = \mu \left(\int_0^\infty b(s) x_{t-s} d\tau - x_t \right) dt + \left\{ \varepsilon \mu \int_0^\infty b(s) (x_t + x_{t-s} - 2x_t x_{t-s}) ds \right\}^{1/2} dW_t$$

WIP 2: a forward model

In order to simplify life we choose for $b(a)$ a uniform distribution on an interval $[0, A]$,

$$b(a) = \frac{1}{A} \chi_{[0,A]}(a).$$

In this case, the model equation reads

$$du_t^\varepsilon = \frac{\mu}{A} \left(\int_0^A (u_{t-s}^\varepsilon - u_t^\varepsilon) ds \right) dt + \left(\frac{\varepsilon \mu}{A} \int_0^A (u_t^\varepsilon + u_{t-s}^\varepsilon - 2 u_t^\varepsilon u_{t-s}^\varepsilon) ds \right)^{1/2} dW_t$$

- **Time scaling coefficient to apply = $(1 + A / 2)^{-1}$**
- **In the limit of $N \rightarrow \infty$ (i.e. $\varepsilon \rightarrow 0$) the drift term can be absorbed into the noise under appropriate scaling (*in contrast to Blath et al.*)**

WIP 2: a forward model

Now we replace $b(a) = \chi_{[0,A]}(a)/A$ by

$$b(a) = \chi_{[0,A]}(a) \varphi(a)$$

where $\int_0^A \varphi(a) = 1$, $\varphi(a) \geq 0$ and $\varphi \in C^1$.

Accordingly, we expect that the “right” approximation of the seedbank model in case that $b(a)$ has a compact support (and is smooth enough) by a time-rescaled Moran model reads

$$du_t = \frac{\{2u_t(1-u_t)\}^{1/2}}{1+E(B)} dW_t. \quad E(B) = \int_0^A x\varphi(x)dx$$

- Scaling coefficient is based on the expected time seeds spend in the seedbank ($E(B)$) (in contrast to Blath et al.)
- **Also extension for weak selection:** $s_S s_P = 1 + \frac{\sigma}{N}$.
- **The drift and diffusion coefficients are scaled by: $(1+E(B))^{-1}$**
- **So selection is indeed slower!!!**



Current group

Remco Stam

Amaryllis Vidali

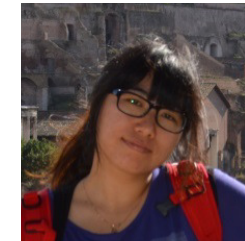
Florence Parat

Saurabh Pophaly

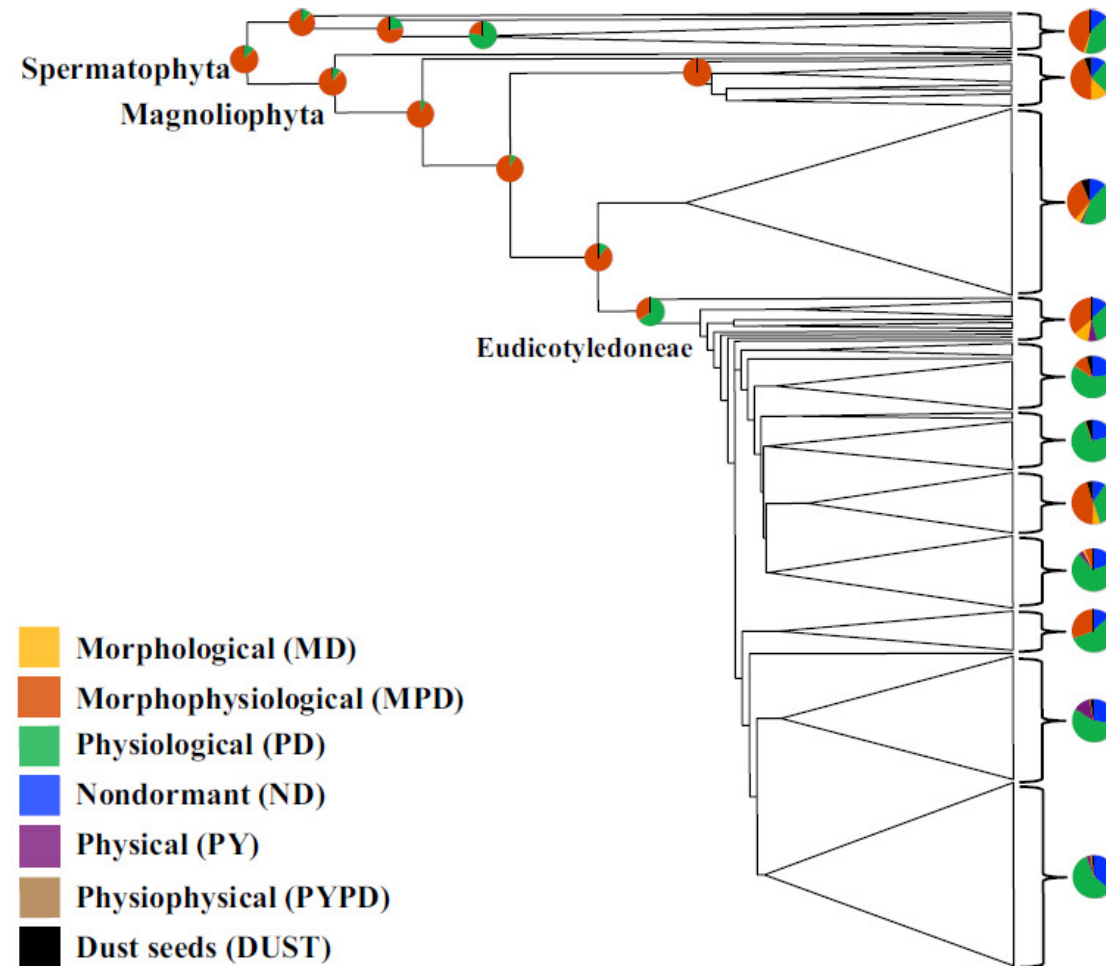
Di Zhang

Mélissa Vérin

Hanna Märkle



- **Seed banks are widespread in plants**
- **Phenotype for dormancy is necessary but not sufficient**



Plant populations are like iceberg

plants above ground = census size = the tip of the iceberg



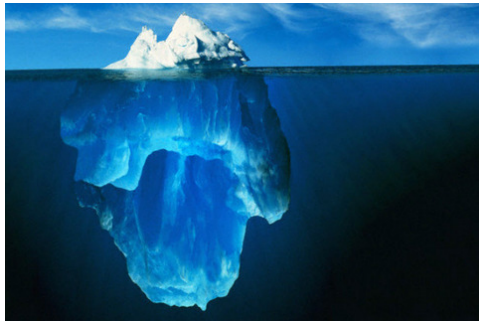
from the Tomato Genetics Resource Center
database (Davis, USA)

	<i>S. chilense</i>	<i>S. peruvianum</i>
total number populations in database	147	304
Mean number of plants above ground (N_{cs})	33 - 154	44 - 185
Maximum number of plants observed	400	600

total number of populations (From Nakazato et al. 2010 *Am J Bot*):

526 (*S. peruvianum*) and **428** (*S. chilense*)

The hidden part: Effective population size



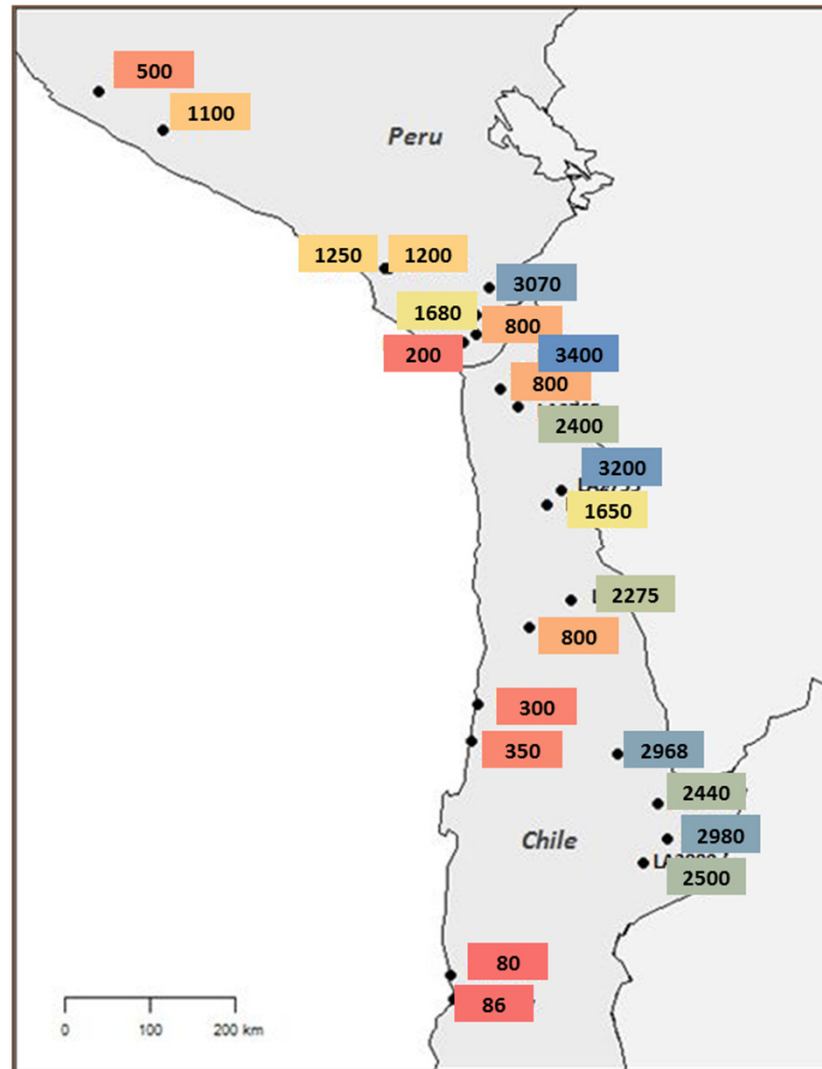
Effective population size (N_e) reflects genetic diversity

Key question

Why is N_e so different from N_{cs} ???

Species	Population	N_e	N_{cs} (census size)
<i>S. peruvianum</i>	Tarapaca	1.23×10^6	≈ 150
	Arequipa		9
	Nazca		10
	Canta		≈ 300
<i>S. chilense</i>	Antofagasta	1.06×10^6	≈ 300
	Tacna		42
	Moquegua		≈ 200
	Quicacha		≈ 40

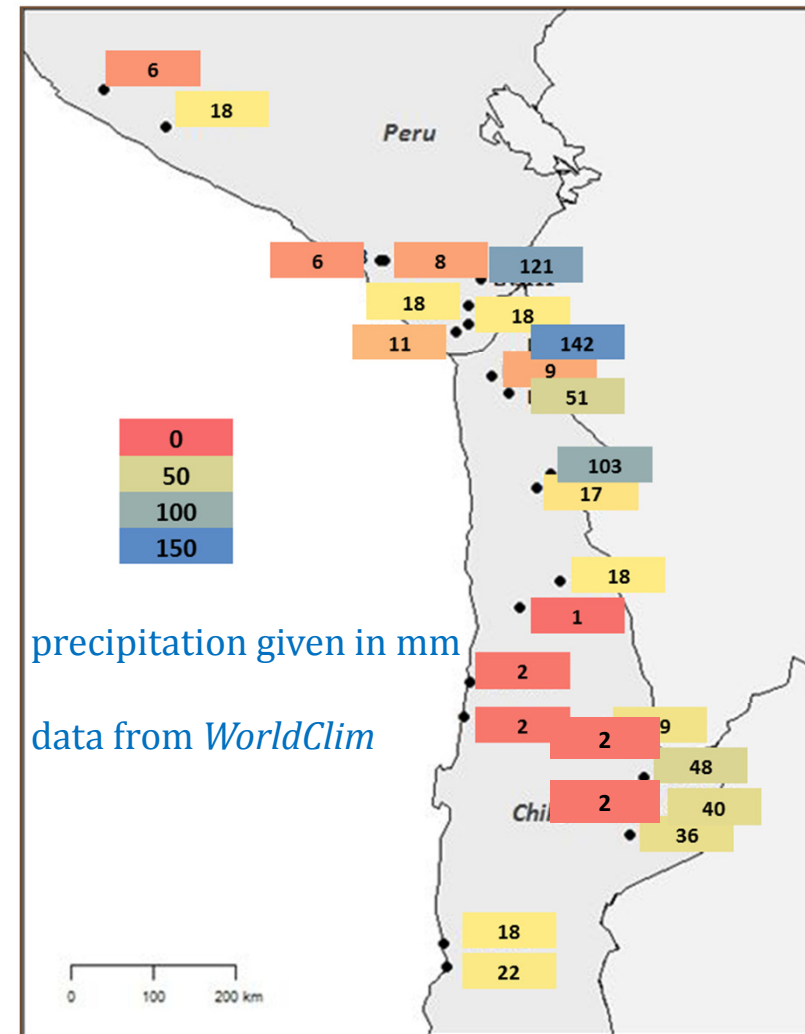
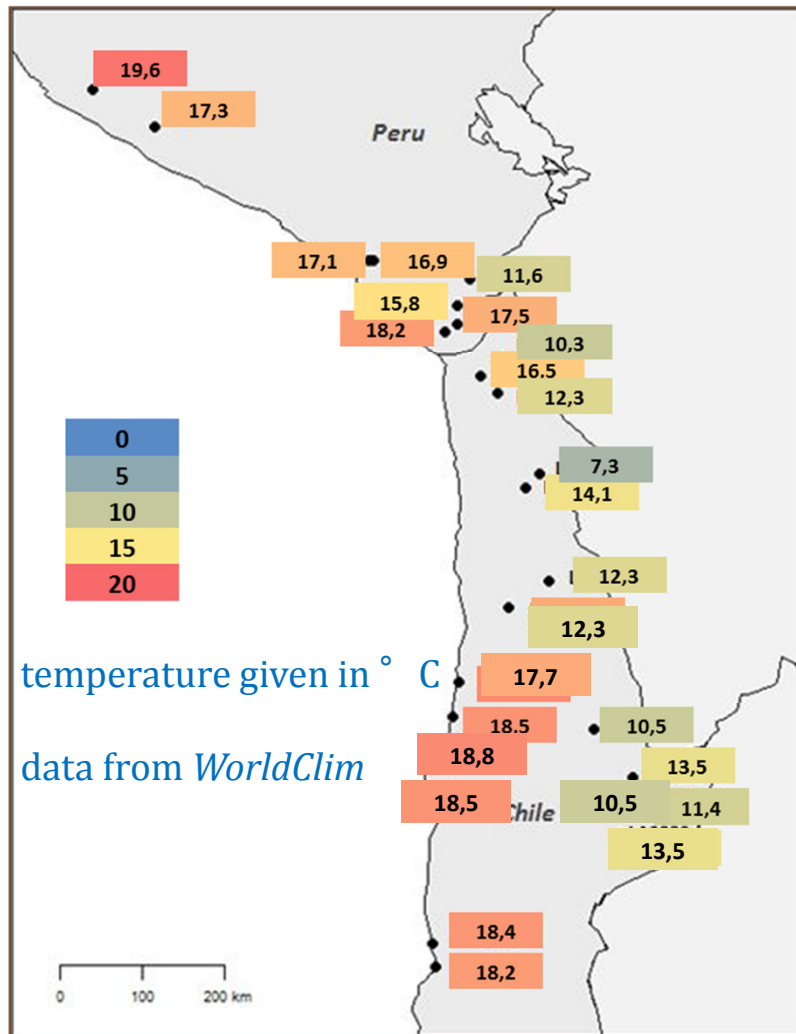
S. chilense sampling: altitudes



altitude given in
meter above sea level

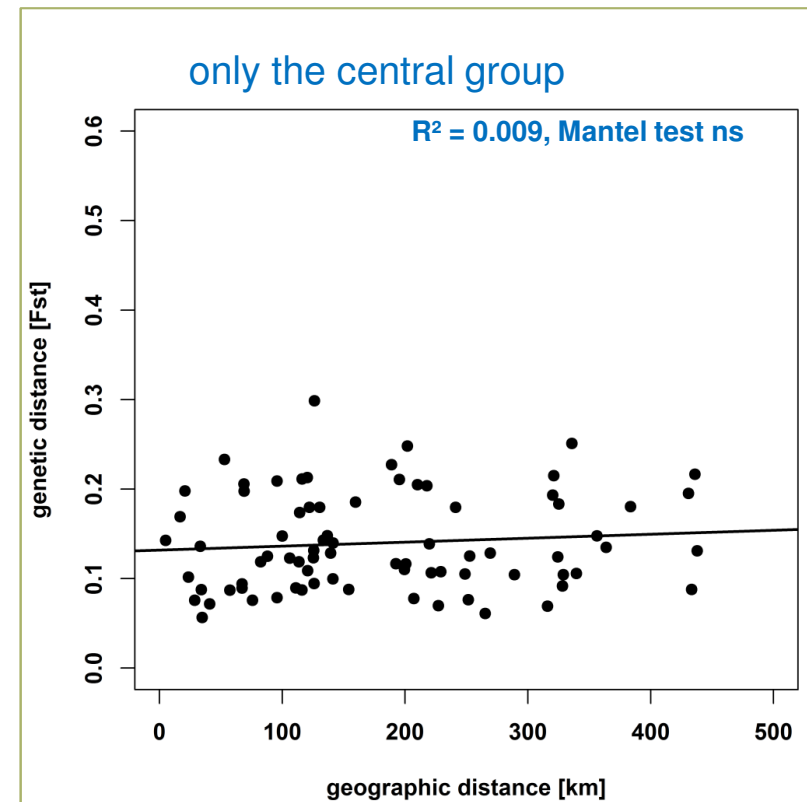
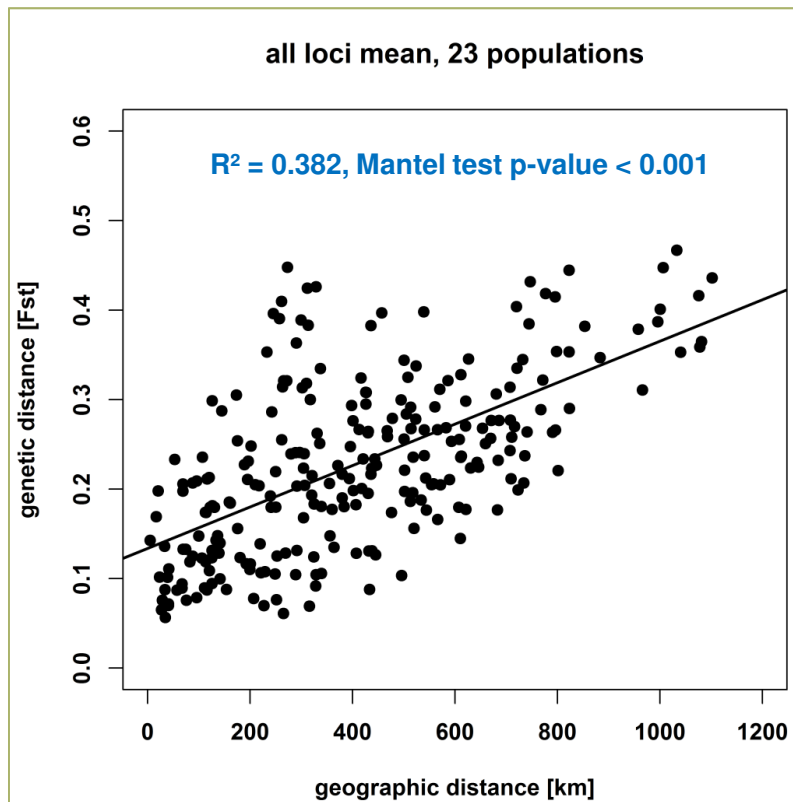
data from *TGRC.ucdavis.edu*

S. chilense sampling: climate



Results: North to South colonization

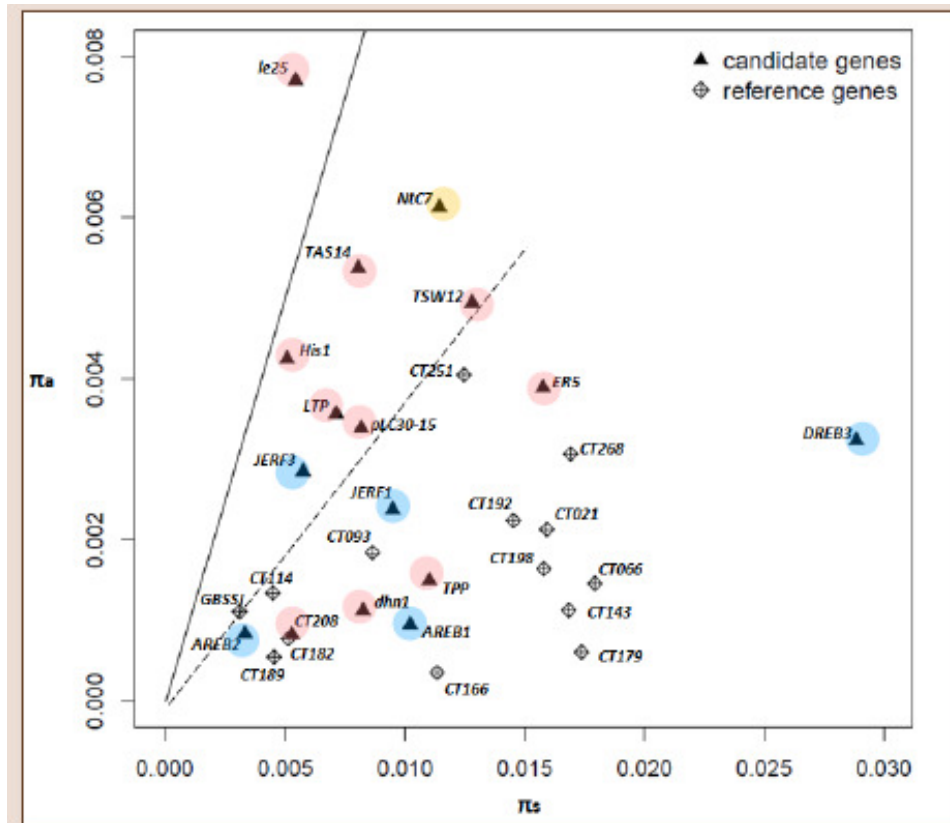
Genetic differentiation between populations and Isolation by Distance



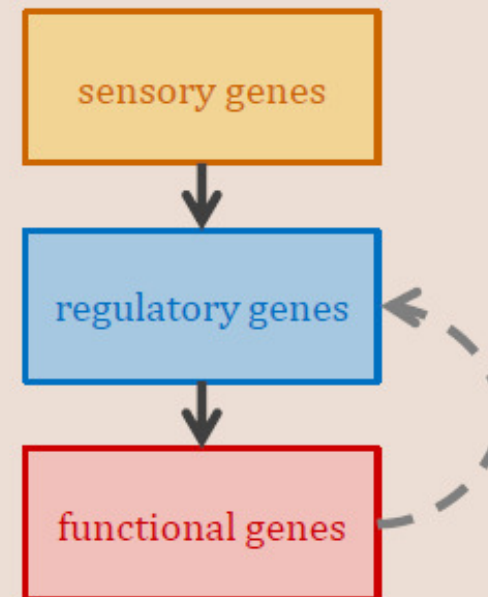
Note: given R^2 values are for all genes

Results: Selection and adaptation

mean π_a and π_s for all genes

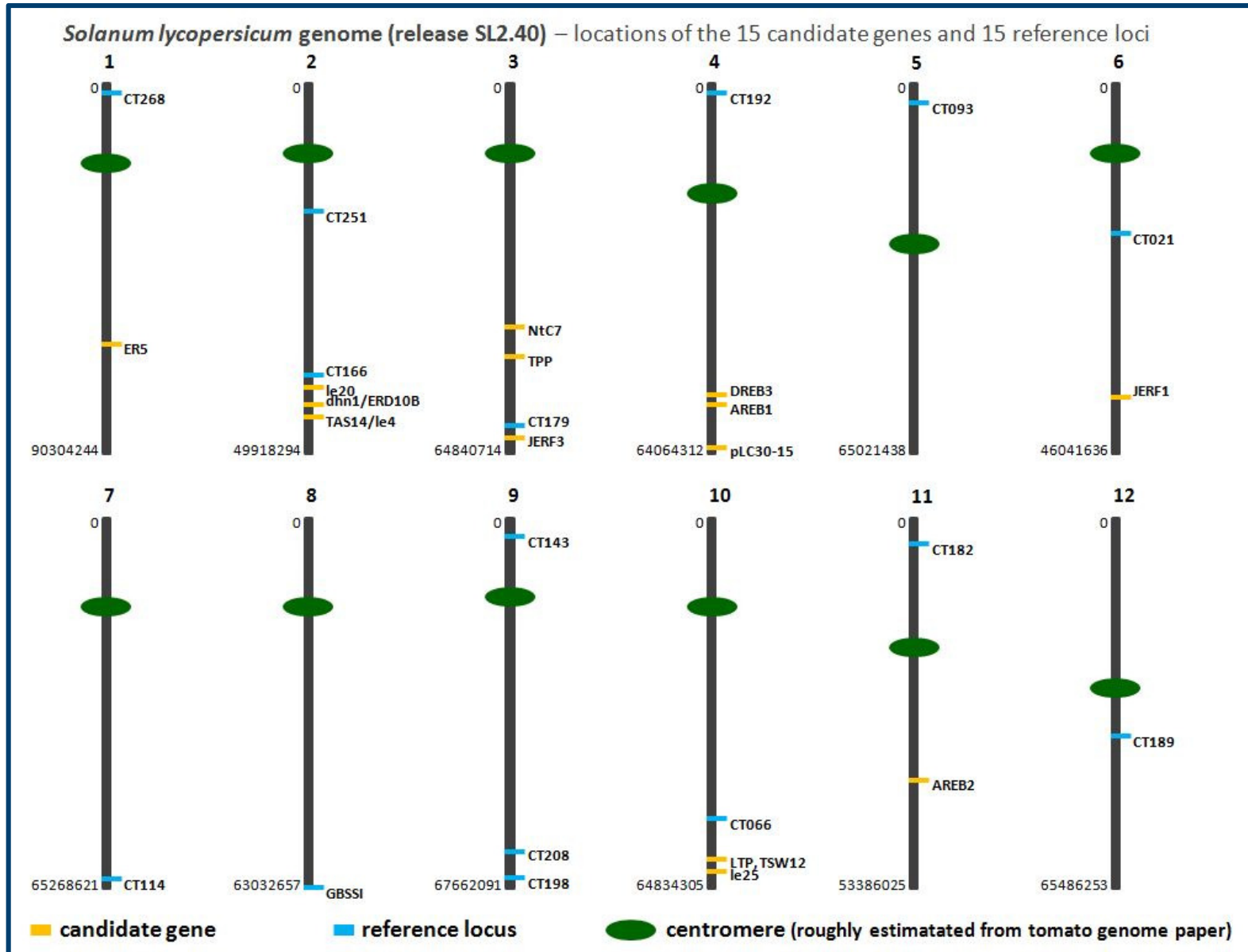


grouping of abiotic stress related genes according to
e.g. Shinozaki & Yamaguchi-Shinozaki 2007, J. Exp. Bot.



solid line: $\pi_a/\pi_s = 1$

dashed line: upper boundary of reference genes



Problem 2: a new forward model

2 Moran Model

As a reference, we briefly review the approach of the classical Moran model.

2.1 Individual based model

State: Out of N individuals, there are at time t X_t individuals with genotype A , and $N - X_t$ individuals with genotype B .

Dynamics: An individual dies at rate μ ; it is immediately replaced by a newborn. The probability for a type- A individual reads X_t/N .

Hence, we find

$$P(X_{t+\Delta t} = k + 1 | X_t = k) = (\Delta t \mu N) (1 - X_t/N)(X_t/N) + \mathcal{O}(\Delta t)$$

and

$$P(X_{t+\Delta t} = k - 1 | X_t = k) = (\Delta t \mu N) (X_t/N)(1 - X_t/N) + \mathcal{O}(\Delta t).$$

Problem 2: a new forward model

$$X_{t+\Delta t} = X_t + Y_{+,t} - Y_{-,t}$$

where approximately

$$\begin{aligned} Y_{+,t} &\sim \text{Pois}((\Delta t \mu N) (1 - X_t/N)(X_t/N)) \\ &\approx (\Delta t \mu N) (1 - X_t/N)(X_t/N) + [(\Delta t \mu N) (1 - X_t/N)(X_t/N)]^{1/2} N(0, 1) \\ Y_{-,t} &\sim \text{Pois}((\Delta t \mu N) (X_t/N)(1 - X_t/N)) \\ &\approx (\Delta t \mu N) (1 - X_t/N)(X_t/N) + [(\Delta t \mu N) (1 - X_t/N)(X_t/N)]^{1/2} N(0, 1) \end{aligned}$$

Now we use relative fraction of A -individuals instead of absolute numbers of A -individuals, $x_t = X_t/N$.
Then,

$$x_{t+\Delta} \approx x_t + N^{-1/2} [2\mu (1 - x_t)(x_t)]^{1/2} N(0, \Delta t)$$

For the time continuous process, we are led to a SDE

$$dx_t = \varepsilon^{1/2} [2\mu (1 - x_t) x_t]^{1/2} dW_t.$$

with $\varepsilon = 1/N$.

Problem 2: a new forward model

3 Time Continuous Seedbank Model with Deterministic Seedbank

3.1 Individual based model

We have a population of N plants (population size is fixed). There are two genotypes, A and B. State of the model at time t :

(1) Plants: $X_t \in \{0, \dots, N\}$ denotes the number of A-type plants, $N - X_t$ those of B-type.

(2) Seeds: The number of (living) seeds in the seedbank is $A_t \in \mathbb{N}_0$, the number of seeds (type B) is $B_t \in \mathbb{N}_0$.

The seeds are described by their age at time t ; let

$$\mathcal{A}_t = \{a_i(t) \mid i = 1, \dots, A_t\}, \quad \mathcal{B}_t = \{a_i(t) \mid i = 1, \dots, B_t\}$$

the ages of the seeds (of genotype A resp. B).

The dynamics consists of (1) death and re-birth of a plant, (2) creation of seeds, and (3) death of seeds.

Problem 2: a new forward model

(1) Death of a plant: At rate μ , plants die. They are immediately replaced by germinating seeds, s.t. the total number of plants is constant. The probability for a plant to be type A (or type B) depends on the seeds in the seedbank. Let $b(a)$ the ability of a seed to germinate. The total ability for germination of A-plants (of B-plants) reads

$$\sum_{a \in \mathcal{A}_t} b(a), \quad \sum_{a \in \mathcal{B}_t} b(a).$$

- Note that $b(a)$ is smooth and has a bounded support.
- Also we do not remove the seed from the bank, assuming that plants produce enough seeds of one type, and few germinate.

(2) Production of seeds

Living plants produce seeds at rate β . That is, the “birthdays” of seeds of A-type are given by an inhomogeneous Poisson process with intensity βX_t . Similarly, that for B-seeds is given by $\beta(N - X_t)$. Please note already here, that the production of any seeds (of either type) follow a Poisson process with constant intensity βN , independent on X_t .

(3) Death of seeds.

Seeds die at rate $\mu_s(a)$.

Problem 2: a new forward model

Now we assume that the seeds are numerous. That is, conditioned on the history of X_τ for $\tau \leq t$, we know that βX_τ seeds of *A*-type, resp. $\beta(N - X_\tau)$ seeds of *B*-type have been produced at time $t - \tau$. If we interpret $\tilde{b}(a)$ as the probability that a seed of age a is still alive (better: able to germinate), we know that we have $\int_0^\infty \tilde{b}(\tau) \beta X_{t-\tau} d\tau$ seeds of type A that can germinate, and $\int_0^\infty \tilde{b}(\tau) \beta(N - X_{t-\tau}) d\tau$ seeds of type B that can germinate. Hence, the amount of living seeds (type A plus type B) in the seedbank is given by $\int_0^\infty \tilde{b}(\tau) \beta N d\tau$.

If a plant dies, the probability that this plant is replaced by a type-A-plant

$$\frac{\int_0^\infty \tilde{b}(\tau) \beta X_{t-\tau} d\tau}{\int_0^\infty \tilde{b}(\tau) \beta N d\tau} = \int_0^\infty \frac{\tilde{b}(\tau)}{\int_0^\infty \tilde{b}(s) ds} X_{t-\tau} / N d\tau.$$

We define the distribution

$$b(a) = \tilde{b}(a) / \int_0^\infty \tilde{b}(s) ds$$

and find the probability for a type-A-plant given by

$$\int_0^\infty b(\tau) X_{t-\tau} / N d\tau.$$

Problem 2: a new forward model

Summary: Our individual based model (that assumes a deterministic seedbank) reads

$$P(X_{t+\Delta t} = X_t + 1 | X_\tau \text{ for } \tau \leq t) = \Delta t \mu N (1 - X_t/N) \int_0^\infty b(\tau) X_{t-\tau}/N d\tau + \mathcal{O}(\Delta t)$$

$$P(X_{t+\Delta t} = X_t - 1 | X_\tau \text{ for } \tau \leq t) = \Delta t \mu N (X_t/N) \left(1 - \int_0^\infty b(\tau) X_{t-\tau}/N d\tau\right) + \mathcal{O}(\Delta t).$$

$$X_{t+\Delta t} = X_t + Y_{+,t} - Y_{-,t}$$

Hence, the deterministic offset of $Y_{+,t} - Y_{-,t}$ yields

$$\begin{aligned} & \Delta t \mu (1 - x_t) \int_0^\infty b(\tau) x_{t-\tau} d\tau - \Delta t \mu x_t \left(1 - \int_0^\infty b(\tau) x_{t-\tau} d\tau\right) \\ &= \Delta t \mu \left(\int_0^\infty b(\tau) x_{t-\tau} d\tau - x_t \right), \end{aligned}$$

Problem 2: a new forward model

while the variance of the two $N(0, 1)$ distributed random variables amount to

$$\begin{aligned} & N^{-1} \mu \Delta t (1 - x_t) \int_0^\infty b(\tau) x_{t-\tau} d\tau + N^{-1} \Delta t \mu x_t \left(1 - \int_0^\infty b(\tau) x_{t-\tau} d\tau \right) \\ = & N^{-1} \mu \Delta t \left\{ x_t + \int_0^\infty b(t - \tau) x_{t-\tau} d\tau - 2x_t \int_0^\infty b(\tau) x_{t-\tau} d\tau \right\} \\ = & N^{-1} \mu \Delta t \left\{ \int_0^\infty b(s) (x_t + x_{t-s} d\tau - 2x_t x_{t-s}) ds \right\} \end{aligned}$$

where we used $\int_0^\infty b(s) ds = 1$ in the last step. Hence, we are led to the SDDE (with $\varepsilon = 1/N$)

$$dx_t = \mu \left(\int_0^\infty b(s) x_{t-s} d\tau - x_t \right) dt + \left\{ \varepsilon \mu \int_0^\infty b(s) (x_t + x_{t-s} - 2x_t x_{t-s}) ds \right\}^{1/2} dW_t$$