



**UNIVERSITÀ POLITECNICA DELLE MARCHE
DIPARTIMENTO SCIENZE DELLA VITA E DELL'AMBIENTE**

Corso di Laurea Magistrale
Biologia Molecolare e applicata – Curriculum Biotecnologie

**STUDIO DELLA RELAZIONE TRA MITO-NUCLEAR
DISCORDANCE E SEX BIASED DISPERSAL UTILIZZANDO UN
METODO DI FORWARD SIMULATION (SLIM)**

**STUDYING THE RELATIONSHIP BETWEEN MITO-NUCLEAR
DISCORDANCE AND SEX BIASED DISPERSAL USING A
FORWARD SIMULATION METHOD (SLIM)**

Tesi di Laurea Magistrale
di:
Francesco Giannelli

Relatore
Chiar.mo Prof.
Emiliano Trucchi

**Sessione estiva
Anno Accademico 2020/2021**

Studio della relazione tra mito-nuclear discordance e sex biased dispersal utilizzando un metodo di forward simulation (SLiM).

Riassunto esteso della tesi di laurea di Francesco Giannelli

1. INTRODUZIONE

Negli ultimi anni, numerosi studi basati sull'analisi della distribuzione geografica dei marcatori mitocondriali e nucleari, hanno rilevato un'elevata incidenza di discordanza nei pattern di distribuzione di questi due marcatori. Tale discordanza prende il nome di *mito-nuclear discordance*. Le ipotesi avanzate per spiegare il fenomeno sono varie. Tra queste, una delle più accreditate è il *sex biased dispersal*. Con questo termine ci si riferisce alla presenza di differenti pattern di dispersione tra individui di diverso sesso all'interno della stessa specie. Tale comportamento è stato rilevato in numerosi taxa: in particolare nei mammiferi, che in linea di massima prenderemo come esempio in questo lavoro, l'elevata dispersione geografica del sesso maschile e la concomitante filopatria femminile sono eventi riportati con elevata frequenza.

2. OBIETTIVI E ASPETTATIVE

Lo scopo di questo lavoro è verificare se processi di sex biased dispersal possono produrre pattern di discordanza geografica nella distribuzione della

diversità genetica tra marcatori nucleari e mitocondriali in popolazioni naturali a dimensioni stazionarie. Ci aspettiamo di osservare che processi di sex biased dispersal, caratterizzati da una scarsa dispersione femminile, siano sempre associati ad un aumento medio della diversità e differenziazione genetica mitocondriale (aumento della struttura genetica) non bilanciata da un aumento equivalente della diversità genetica nucleare e, inoltre che la filopatria femminile comporti un aumento della varianza nei valori di differenziazione genetica mitocondriale.

3. METODI

Per validare le nostre ipotesi, lo strumento utilizzato è SLiM, un software di simulazione in forward ottimizzato per la creazione di modelli evolutivi sofisticati. All'interno dell'ambiente simulativo di SLiM abbiamo generato individui caratterizzati da genomi composti da una porzione nucleare e da una porzione mitocondriale, che si disperdono e interagiscono in un'area bidimensionale. Per lo scopo sono stati disegnati e testati cinque modelli, che prevedono differenti coefficienti di dispersione per i due sessi:

Modello	1	2	3	4	5
Dispersione maschile	2Km	3.3333 Km	5 Km	6.66667 Km	2Km
Dispersione femminile	8 Km	6.66667 Km	5 Km	3.3333 Km	8 Km
Pattern di sex biased dispersal	0.25	0.5	1	2	4

Ogni modello viene lanciato 10 volte e il campionamento di individui, effettuato utilizzando 3 diverse distanze di campionamento (50 Km, 100 Km, 150 Km), inizia solo una volta che la popolazione raggiunge un equilibrio definito da una sufficiente stabilizzazione dei valori di eterozigotà nucleare e mitocondriale.

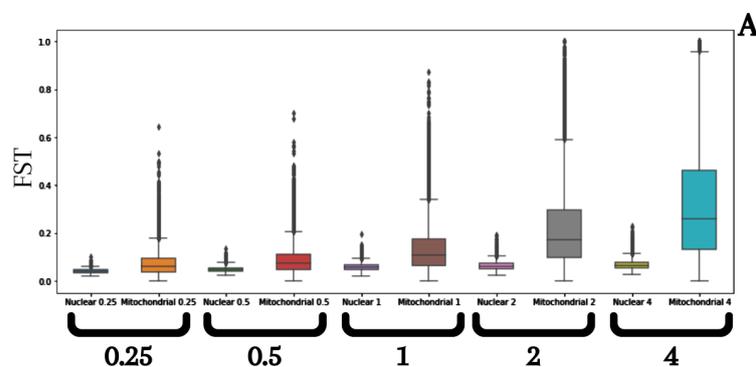
Le statistiche calcolate per questi modelli sono:

- F_{ST} : misura la differenziazione genetica tra campioni di individui
- Eterozigotà: misura la frequenza di eterozigoti, stimata per ogni locus e rapportata alla lunghezza del genoma.

Entrambe le statistiche vengono calcolate sia per il genoma nucleare che per il genoma mitocondriale.

4. RISULTATI

4.1 Effetto dei pattern di sex biased dispersal sulla differenziazione genetica



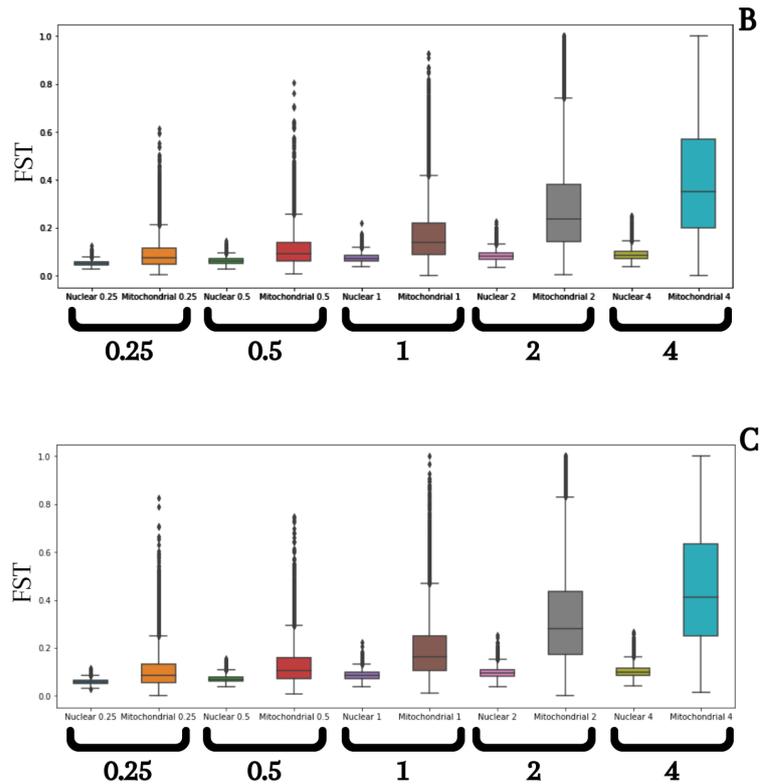


Figura 1. **Effetto di diversi pattern di sex biased dispersal sulla differenziazione genetica nucleare e mitocondriale.** I tre boxplot mostrano i valori F_{ST} , per il genoma nucleare e mitocondriale, relativi a tutti i campioni dei cinque diversi modelli, con diversi pattern di sex biased dispersal: 0,25, 0,5, 1, 2, 4. I valori sono suddivisi nelle tre immagini in base alla rispettiva distanza di campionamento. **A:** 50 Km, **B:** 100 Km, **C:** 150 Km.

Come si evince da ogni confronto in figura 1, i valori di differenziazione mitocondriale sono, in media, sempre maggiori di quelli nucleari. Ciò è dovuto al minor flusso genico del genoma mitocondriale, la cui dimensione di popolazione effettiva è di base ridotta (i.e., eredità uniparentale e aploidia) rispetto a quello nucleare. È invece particolarmente interessante vedere come i diversi pattern di sex biased dispersal agiscano sulla differenziazione genetica mitocondriale: possiamo infatti apprezzare come il valore della F_{ST} mitocondriale tra individui diminuisca leggermente quando il rapporto di

dispersione è sbilanciato in favore della femmina, mentre aumenta notevolmente quando il rapporto di dispersione è sbilanciato a favore del maschio.

4.2 Distribuzione comparativa dei valori F_{ST}

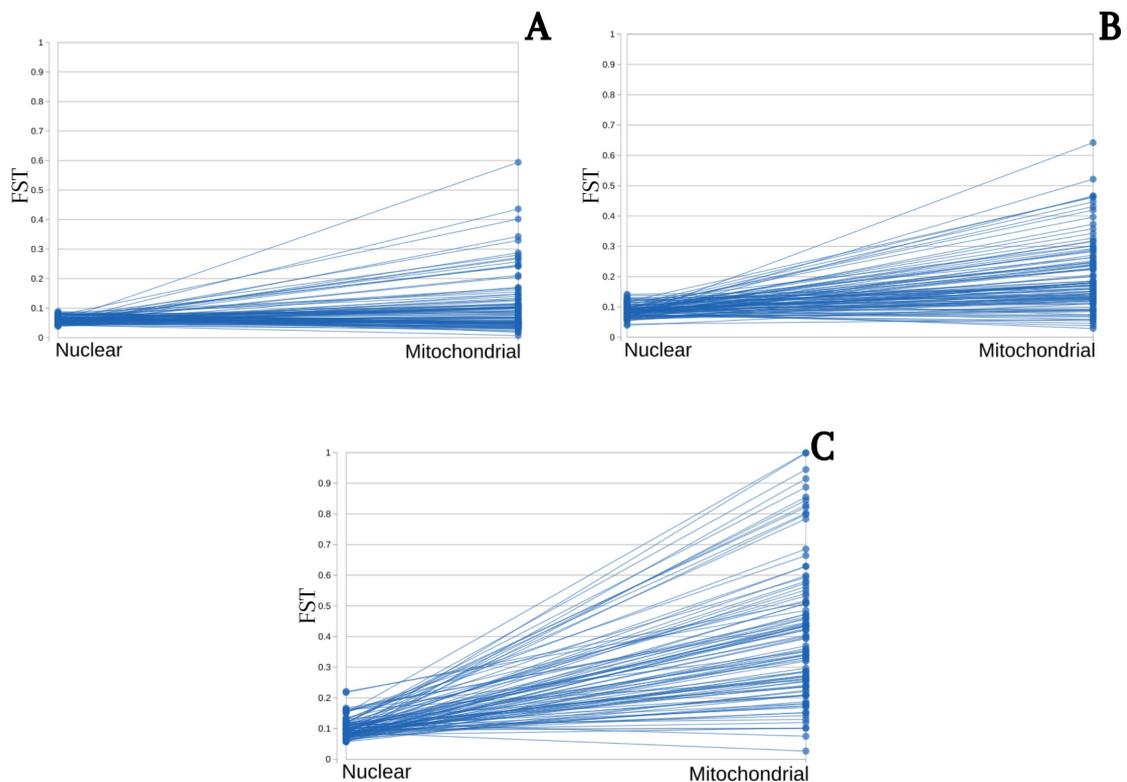


Figura 2. **Relazione tra differenziazione genetica nucleare e mitocondriale per coppie di campioni per diversi modelli di dispersione influenzata dal sesso (0.25, 1, 4).** I grafici mostrano i valori F_{ST} per i genomi nucleari e mitocondriali calcolati tra coppie di campioni prelevati alla generazione 100.000 utilizzando 150 Km come distanza di campionamento. Le immagini mostrano i risultati di tre diversi modelli con diversi pattern di dispersione basati sul sesso: **A:** 0,25, **B:** 1, **C:** 4.

Risulta inoltre evidente come i pattern di sex biased dispersal in cui la dispersione femminile è ridotta (Figura 2, immagine C) portano ad una distribuzione molto ampia dei valori di F_{ST} mitocondriale. Ciò evidenzia che

nei modelli caratterizzati da un'elevata dispersione maschile e filopatria femminile, prelevando pochi campioni casuali, si possono osservare situazioni in cui la differenziazione genetica mitocondriale tra gruppi di individui è molto alta e discordante con la differenziazione nucleare o, al contrario, piuttosto bassa e più coerente con i valori di differenziazione genetica nucleare, che rimangono invece costanti.

4.3 Impatto della distanza geografica

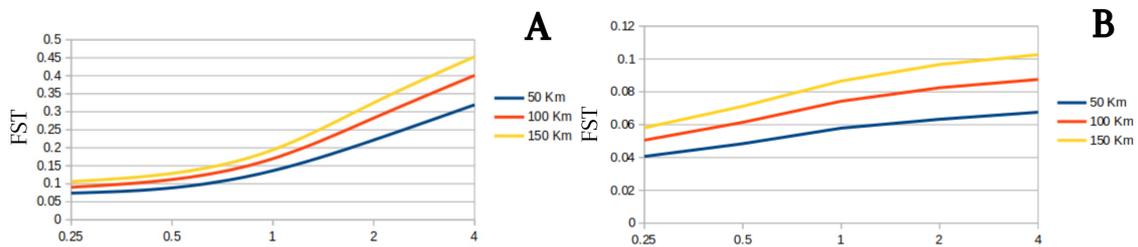


Figura 3. **Differenziazione genetica media per diversi modelli di dispersione influenzata dal sesso in base alla distanza di campionamento.** L'immagine A mostra i risultati del genoma mitocondriale, dove è possibile vedere come, nei pattern di sex biased dispersal 2 e 4 (asse X), il valore medio della differenziazione genetica (F_{ST}) aumenta notevolmente. L'immagine B mostra i risultati del genoma nucleare. Notare la diversa scala sull'asse Y tra i due pannelli.

I risultati suggeriscono che, in ogni modello, la variazione dei valori di F_{ST} , sia mitocondriale (Figura 3, immagine B) che nucleare (Figura 3, immagine A), è sempre direttamente proporzionale alla distanza. Questo pattern è coerente con un effetto di isolation by distance, dove una distanza geografica crescente è sempre proporzionalmente correlata a una distanza genetica maggiore tra gli individui.

4.4 Dinamiche dell'eterozigotità

L'eterozigotità è un parametro di cui abbiamo tenuto conto per definire il tempo di attesa necessario affinché la popolazione simulata raggiungesse un equilibrio sufficiente per iniziare le nostre analisi. Nel tener traccia di questa statistica, tuttavia, ci siamo resi conto che nelle simulazioni il pattern di eterozigotità mitocondriale aveva un andamento atipico e inatteso, caratterizzato da picchi e successivi crolli (Figura 4).

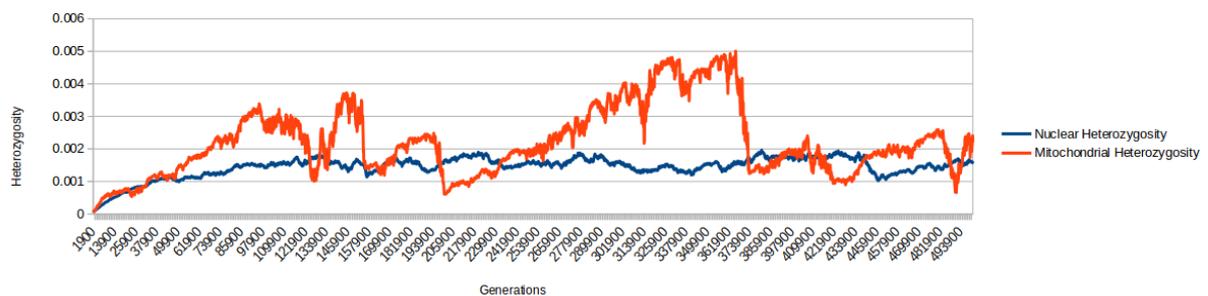


Figura 4. **Variazione dell'eterozigosi nucleare e mitocondriale in 500.000 generazioni.** Nel modello utilizzato per creare questo grafico la costante di dispersione maschile è 4 volte la costante di dispersione femminile (modello $SBD = 4$).

Le cause di questi cicli di quasi-fissazione seguiti da nuova differenziazione degli aplotipi potrebbero essere ricondotte alla mancanza di ricombinazione dei marcatori mitocondriali, insieme al più intenso effetto della deriva su questo genoma, ma sicuramente la dinamica instabile dell'eterozigosi mitocondriale merita di essere ulteriormente approfondita.

5. Discussione e conclusioni

Le nostre simulazioni rivelano chiaramente che i pattern di sex biased dispersal, caratterizzati da una ridotta dispersione femminile e quindi da un flusso genico mitocondriale ridotto, sono sempre associati a una differenziazione genetica mitocondriale (F_{ST}) media, significativamente più alta e con una maggiore variabilità nel suo intervallo di valori (Figura 1). Allo stesso tempo, a causa della stocasticità intrinseca del processo di dispersione, possiamo osservare che a valori costanti di F_{ST} nucleari possono essere associati sia valori alti che bassi di F_{ST} mitocondriale (Figura 2, immagine C). I nostri risultati mostrano come, semplicemente diminuendo l'entità del flusso genico mitocondriale, si ha un forte impatto sulla struttura della diversità genetica di questo genoma, suggerendo una forte correlazione tra sex biased dispersal e mito-nuclear discordance.

SUMMARY

ABSTRACT.....	15
1. INTRODUCTION	16
1.1 - Mito-nuclear discordance	16
1.2 - Sex biased dispersal	21
2. AIM AND OBJECTIVES	26
2.1 - Aims of the project.....	26
2.2 - Project objectives	26
2.3 - Expectations draw from population genetics theory.....	27
3. METHODS	30
3.1 - Simulations	30
3.2 - SLiM.....	32
3.3 - non-Wright Fisher model	34
3.4 - Continuous space model.....	36
3.5 - Neutral conditions	37
3.6 - Reference species	37
3.7 - Model design and basic parameters.....	38
3.8 - The genome.....	38
3.9 - Mitochondrial and nuclear inheritance models.....	40
3.10 - The simulated area	43
3.11 - Sex biased dispersal simulation.....	44
3.12 - Fitness and population size.....	47
3.13 - Spatial interactions	48
3.13.1 Reproduction interaction	48
3.13.2 Competitive interaction	49
3.14 - Simulation burn-in to reach stationarity	51
3.15 - Sampling	52
3.16 - Summary statistics	54
3.17 - Running the simulations.....	56

4. RESULTS	58
4.1 - Effect of sex biased dispersal patterns on genetic differentiation	58
4.2 - Comparative distribution of F_{ST} values	61
4.3 - The impact of geographic distance	64
4.4 - Heterozygosity dynamics across generations	66
5. DISCUSSION	69
5.1 - Discussion	69
5.2 - Future directions	73
5.2.1 More realistic scenarios	73
5.2.2 Use of phylogenetic trees as summary statistics	74
5.2.3 Application to a real biological system	75
5.2.4 Building a predictive model.....	76
6. CONCLUSIONS.....	77
BIBLIOGRAPHY	80
SUPPLEMENTARY MATERIAL.....	87

ABSTRACT

The number of papers reporting mito-nuclear discordance in natural populations is continuously increasing. One of the most accredited hypotheses to explain mito-nuclear discordance is sex biased dispersal, a phenomenon that is common in mammals in which male biased dispersal and female philopatry are frequently reported.

To test the hypothesis that these two phenomena are correlated, using the forward simulation software SLiM, we created 5 simulation models, implementing a different sex biased dispersal pattern in each model.

The results clearly reveal that sex biased dispersal patterns, characterized by reduced female dispersion, are always associated with mean, significantly higher mitochondrial genetic differentiation and with greater variability in its range of values, while the nuclear genetic differentiation remains largely unaltered. These results suggest a strong correlation between sex biased dispersal and mito-nuclear discordance.

1. INTRODUCTION

1.1 - Mito-nuclear discordance

In the context of population genetics, the study of the geographical distribution of variation at different genetic markers is fundamental to understand the dynamics that occur within and among populations. A genetic marker is a DNA sequence with a known position in the genome. Genetic markers are tools widely used in various fields of genetics (Sunnucks et al., 2000). In population genetics, mitochondrial and nuclear markers are used to evaluate how they vary between individuals or groups of individuals in order to infer evolutionary processes.

The most used approaches for reconstructing genetic relationships between individuals have always taken greater account of the use of mitochondrial markers only, but in the past years, thanks also to the increased availability of whole genome (i.e., nuclear) data, a large number of studies have highlighted that this approach can lead to incorrect conclusions (Zink & Barrowclough 2008).

Mitochondrial DNA has many specific characteristics, some of which make this genome very suitable for phylogenetic analysis, which can be summarized as follows even though some exceptions exist:

- it is haploid
- it is not subject to recombination or segregation
- it does not present the problem of duplications
- it is present in multiple equal copies
- it is maternally inherited
- it has a high rate of evolution due to the fact that it does not have reparation mechanisms

However, conducting analyses based only on the mitochondrial genome and therefore ignoring the nuclear genome data, in many cases diploid and inherited from both parents, introduces important limitations and does not guarantee us a complete description of the population from a genetic point of view. At the same time analyses conducted by combining data from mitochondrial and nuclear genomes can lead to contrasting results (Zink & Barrowclough 2008; Folt et al., 2019), as the mitochondrial genome, due to its maternal inheritance and its haploidy, has an effective population size four times smaller than the nuclear one (Hudson & Turelli 2003; Zink & Barrowclough 2008). It follows that mitochondrial DNA is more susceptible to the accumulation of mutations due to genetic drift, and also to reduced gene flow compared to the nuclear genome. Taken together, these characteristics cause the mitochondrial genome

to usually show different patterns of genetic diversity than the nuclear genome, heavily conditioning this type of analysis.

This implies that analyses based only on mitochondrial data or on both nuclear and mitochondrial data can lead to results that do not completely describe the evolutionary dynamics occurring within the population or, even worse, that can support incorrect hypotheses (Folt et al., 2019).

The suggested solution to avoid these kinds of problems is to conduct analyses on the nuclear genome and the mitochondrial genome separately. However, in the past years, this approach has brought to light a high frequency of papers reporting a discrepancy between the geographical distribution pattern of nuclear and mitochondrial markers (Toews et al., 2012). This so-called mito-nuclear discordance can be defined as the different geographic distribution of diversity between mitochondrial and nuclear markers within a natural population.

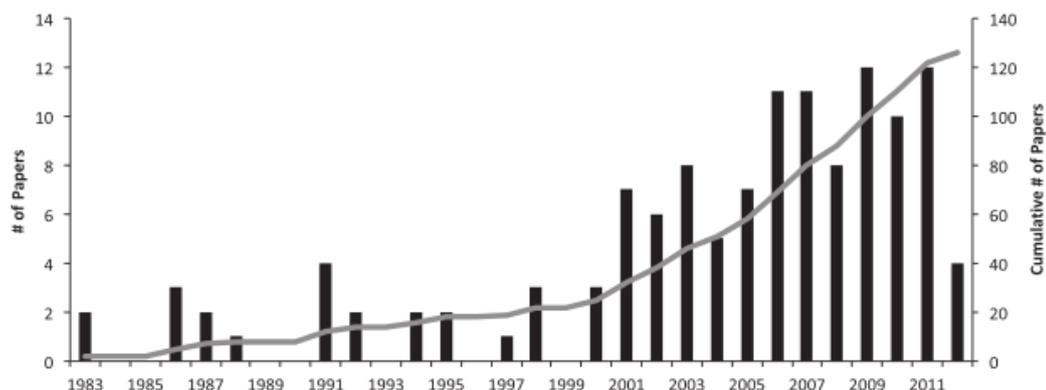


Figure 1: **Number of papers reporting mito-nuclear discordance:** cases from 1983 to 2011. (From Toews et. al. 2012)

The diagram in Figure 1 shows how many papers, reporting mito-nuclear discordance, have been published from 1983 to 2011. It clearly highlights how the number of papers has increased over the years, suggesting mito-nuclear discordance as an important phenomenon in nature and so a real problem for researchers attempting to reconstruct the evolutionary dynamics of populations and species (Figure 2).

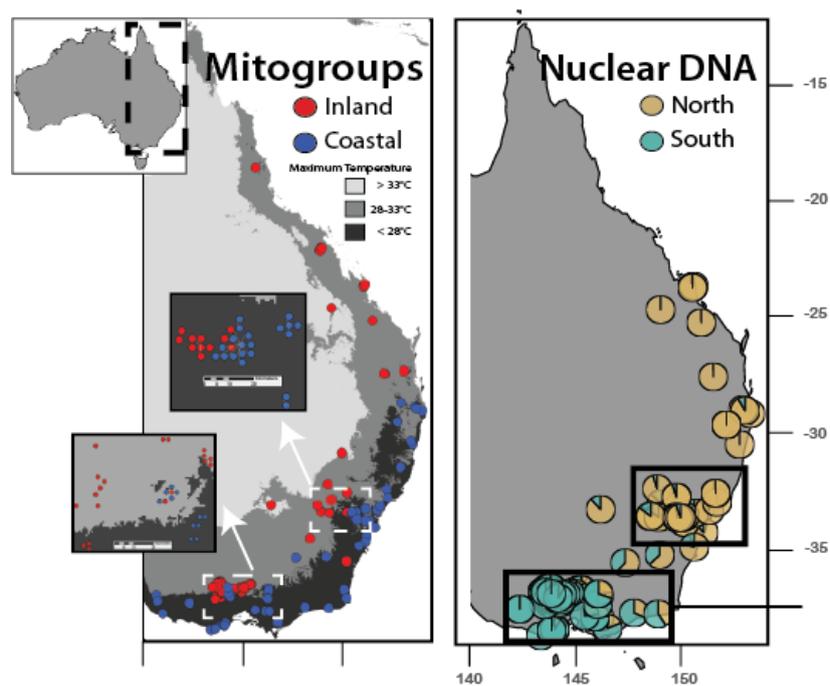


Figure 2. **Example of mito-nuclear discordance** (from Morales et al 2017).

In the past years many works have investigated the causes of this discrepancy, trying to give an explanation to this phenomenon. These works suggest that the

most probable causes for the arising of mito-nuclear discordance within a population are:

- Demographic characteristics and demographic events that can enhance genetic drift: genetic drift alone can produce geographically discordant patterns between nuclear and mitochondrial DNA (Petit & Excoffier, 2009) and if the demographic history of the population is characterized by bottlenecks the effect of drift is more pronounced.
- Introgression: if the phenomena of introgression mainly or totally concern only one between the nuclear and mitochondrial genome (Cahill et al., 2013; Morales et al., 2017).
- Incomplete lineage sorting: describes the situation in which copies of genes found on individuals of different species coalesce into a common ancestral copy in a time prior to the speciation event between the two species. Therefore, the gene tree produced by the gene differs from the tree produced by the species under examination, producing discordant results (Maddison 1997; Suh et al., 2015).
- Selective processes: if in the population there are differences in how selection acts between mitochondrial and nuclear genome, mito-nuclear discordance can arise (Toews et al., 2012; Morales et al., 2017).

- Sex-biased asymmetries: in this context these asymmetries may be caused by mating behaviour, sex-biased offspring production or sex biased dispersal (Toews et al., 2012). In particular, sex biased dispersal will be the process on which the present work is focused.

1.2 - Sex biased dispersal

Sex biased dispersal refers to a particular type of sex biased asymmetry that consists in different patterns of dispersion between individuals of different sexes within a population. Sex biased dispersal is a common behaviour in animal species as it is often reported in mammals, birds and reptiles (Trochet et al., 2016).

Particularly, in mammals, male biased dispersal (i.e., males disperse more than the females) and female philopatry are the most frequently reported phenomena (Trochet et al. 2016; Dobson et al., 1982).

Taxonomic groups	Methods of sex-biased dispersal identification					Both	Proportion of male- and female-biased dispersal	Total (species)
	Direct methods		Indirect methods					
	Mark-release-recapture	Radio-tracking	Nuclear DNA	Mitochondrial DNA				
Amphibians	1	0	3	0	0		4	
Arthropods	6	0	1	0	0		7	
Birds	109	2	5	1	1		118	
Actinopterygian fishes	0	0	2	0	2		4	
Mammals	56	11	21	6	16		110	
Nonavian reptiles	5	0	6	1	2		14	
Total	177	13	38	8	21		257	

Table 1. **Proportion of sex-biased dispersal:** the black part represents male-biased dispersal while the grey part represents female biased dispersal. (From Trochet et al 2016)

The hypothesized causes to explain this phenomenon are:

- Inbreeding avoidance: due to the greater parental investment in the offspring by females, these pay a higher price in case of inbreeding, so females prefer immigrants to residents, causing male biased dispersal (Lehmann et al., 2003).
- Mate competition: the sex in which there is greater intra-sexual competition for reproduction has a higher dispersion coefficient (Dobson, 1982).

- Resource competition: suggested to cause sex biased dispersal when the competition for resources is high and a portion of individuals is prevented from breeding (Greenwood, 1980).

As the geographical dispersion of individuals is the basis of the gene flow within and among populations, sex-biased asymmetries in dispersal may have profound effects on the distribution of genetic diversity (Bohonak et al., 1999). Nuclear gene flow depends on both male and female dispersal while mitochondrial gene flow depends on female dispersal only. So, even when both sexes disperse at the same rate, mitochondrial gene flow is lower than the nuclear gene flow. Hence, if females disperse less than males, mitochondrial gene flow is highly reduced while nuclear gene flow is only partially affected by the lower female dispersion rate (Figure 3).

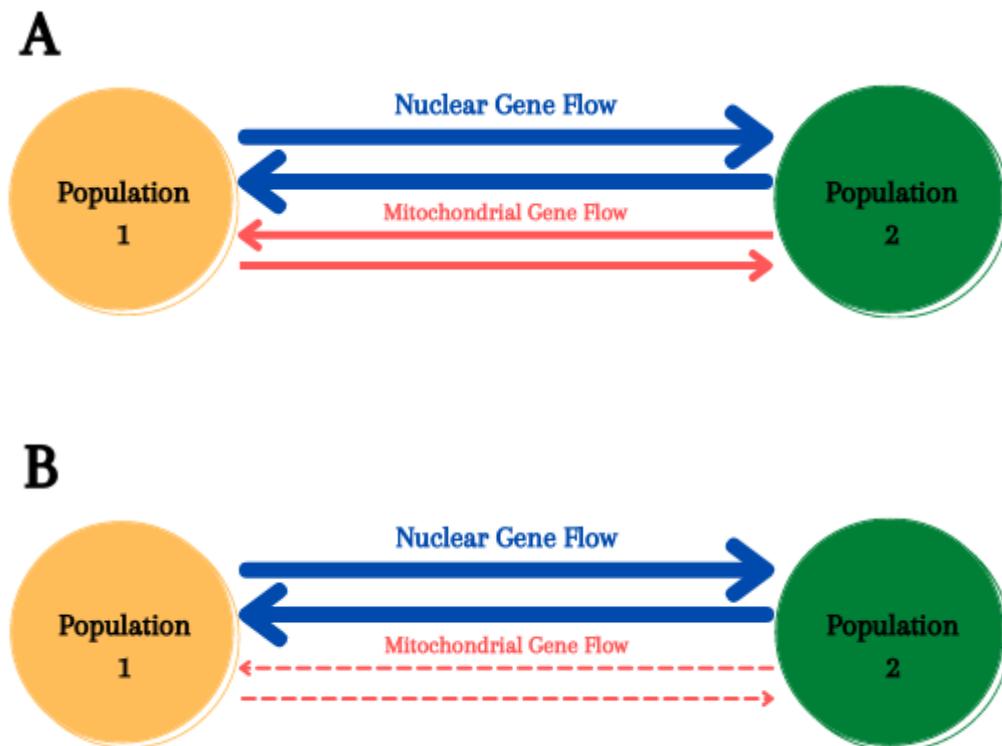


Figure 3. **Nuclear and mitochondrial gene flow between two populations.** Panel A shows the gene flow between two populations in which there is no sex biased dispersal. In Panel B is represented the gene flow between two populations in which there is sex biased dispersal.

Different dispersion rates between sexes creates a significant difference between gene flow at the different genomes (mitochondrial and nuclear) which can lead to marked differences in the geographic distribution of genetic diversity calculated between nuclear and mitochondrial markers.

In particular, analyses using mitochondrial markers could highlight the presence of a clear structure among populations, whereas analyses using nuclear markers do not reveal any population's structure.

A population is considered structured when it deviates from a model of complete panmixia and it is instead composed of local subpopulations which are distinguishable from each other at the genetic level. These subpopulations are characterized by similar (usually low) levels of genetic diversity among all the individuals that compose each of them, but at the same time a certain level of genetic diversity when comparing individuals from different subpopulations (Wright, 1949).

Thus, we have that sex biased dispersal can then significantly contribute to creating a mitochondrial genetic structure within natural populations that is not balanced by an equally marked nuclear genetic structure.

This lack of consistency in the patterns of diversity between the two genomes is defined as mito-nuclear discordance.

2. AIM AND OBJECTIVES

2.1 - Aims of the project

The purpose of this work is to test whether sex-biased dispersal processes can produce patterns of geographic discordance in the distribution of genetic diversity between nuclear markers and mitochondrial markers in natural populations at stationary population size.

On a more methodological perspective, another important target of this work is to test whether a newly developed forward simulation approach can be useful to create sophisticated evolutionary scenarios in order to explore mito-nuclear interactions in a spatial model.

2.2 - Project objectives

We can summarize the objectives of this work in four main points:

1. Learn how to use an advanced bioinformatic tool for forward in time simulations of natural populations to develop a spatially-explicit two-dimensional simulation model.

2. Design a simulation system which allows to simulate mitochondrial genome alongside nuclear genome with the appropriate inheritance of both DNAs and implement spatial simulation dynamics in the system such as dispersion and spatial interactions among individuals.
3. Use this simulation system to develop a biologically realistic model to test different patterns of sex biased dispersal and their effect on both nuclear and mitochondrial genetic diversity while producing the necessary output for estimating summary statistics of interest.
4. Analyse the output of the different simulation scenarios in the context of the basic population genetics theoretical expectations to verify the likelihood of our simulation design and understand the role of the sex-biased dispersal on mito-nuclear discordance patterns.

2.3 - Expectations draw from population genetics theory

Due to the haploidy and the maternal inheritance of the genome, mitochondrial effective population size is expected to be four times smaller than the nuclear one. The effective population size has a strong impact on the genetic diversity of the population being inversely correlated with the intensity of genetic drift. Because of this difference in the effective size of the populations, we expect

that the genetic diversity of the two genomes will be already different from one another.

In addition, nuclear gene flow derives from both male and female dispersal while mitochondrial gene flow derives from female dispersal only. This means that even when male and females disperse at the same rate nuclear gene flow is higher than mitochondrial gene flow.

Reduced gene flow means again a more intense effect of divergence by drift and, hence, higher genetic diversity between distant individuals. So, if in the population females disperse less than the males, nuclear gene flow will be partially reduced, due to the fact that the female contribution is lower than in conditions of equal dispersion, while mitochondrial gene flow will be instead highly reduced. So, we can presume that sex biased dispersal differently affects mitochondrial genetic diversity and nuclear genetic diversity leading to discordant patterns of genetic diversity distribution between the two genomes.

We expect that by increasing the difference in dispersion between the males and the females (male biased dispersal) in our population, we will observe an increased mitochondrial genetic diversity and differentiation (i.e., increased genetic structure) not balanced by an equivalent increase in nuclear genetic diversity.

In order to have meaningful insights of the correlation between sex biased dispersal and mito-nuclear discordance, we also expect to detect a broader distribution of mitochondrial genetic differentiation, characterized by values that can be in total contrast with nuclear differentiation, and by values that are consistent with the latter.

On a broader perspective, investigating the correlation between mito nuclear discordance and sex biased dispersal will help us understand how these processes can affect the results of scientific studies and, for example, how they can bias populations and species conservation programs. A better understanding of the effect of sex biased dispersal on the genetic diversity can be fundamental to delineate more accurate expectations and to correctly set up studies on natural populations.

3. METHODS

3.1 - Simulations

Simulation software has proven to be a very useful tool to investigate the evolutionary dynamics that occur in natural populations. Simulation approaches allow the user to create sophisticated ecological and life-history evolutionary models which are suited to inferring natural processes or demographic characteristics of populations. In the last decades, thanks to technological progress and the greater availability of computing power, simulations have become more powerful and accurate, ensuring greater predictive capacity, and allowing these methods to gain an increasingly important role in many scientific fields (Hoban et al., 2012; Yuan et al., 2012). Moreover, thanks to the greater use of this software by the scientific community, the developers of these programs have worked to make these methods increasingly accessible, even to researchers less specialized in computer science, also implementing simple graphical user interfaces (gui). In population genetics simulations software are mainly used to predict the effects of interacting evolutionary forces, to infer historical processes or to understand the properties of newly developed methods (Hoban et al., 2012).

In genetics, simulation of evolutionary models can be done in forward or in backward (or coalescent-based).

Backward simulations generate a gene tree that describes the history of different copies of a gene until the most recent common ancestor (MRCA) is reached. The probability of two copies to coalesce into a common ancestor is influenced by their demographic history. These approaches are based on the coalescent theory (Kingman, 1982). Backward simulations look back in time, starting from a final condition in which we have several copies of a single gene and inferring the coalescence events between the samples, to an initial condition in which all copies of the target gene reach a common ancestor. This approach only considers DNA fragments and not individuals, this feature makes these methods much faster and less complex than forward simulations.

Forward simulations are focused on individuals and their genomes (or parts of genome); every individual has different characteristics and follows its own life cycle. Individual's features can be determined by the users to reach their purpose or can be emergent from the simulation model. Forward simulations, as the name suggests, look forward in time, starting from an initial population to a final population. Since the simulation must consider a large number of

individuals and therefore a large number of DNA fragments and individual's features, this approach requires a much higher computational effort than the backward approach.

Between these two approaches there is no one better than the other, both are valid methods, the choice between the two depends only on the application for which they are required. Forward simulation methods allow to simulate more complex evolutionary scenarios than the backward simulations methods, but the latter are more suitable for inferring processes that occur on large time scales.

3.2 - SLiM

To test the effect of sex biased dispersal on mitochondrial and nuclear genetic diversity, we used SLiM (Selection on Linked Mutations): an optimized forward simulation software (Haller et. al., 2019). SLiM allows to create sophisticated evolutionary models thanks to its unique features:

- It is a highly scriptable software, meaning that every mechanic of our scenario can be extensively modified, and it does not require any high work of scripting, as even complex models can be expressed in a page of code or less.

- The software is highly optimized in order to be more efficient and to require less computing power.

These characteristics make this software very suitable for our work and more generally make SLiM a very interesting method for carrying out population genetic analyses.

As proof of this, since 2013, the year of release of the software, the number of users and scientific articles based on the use of this software has continuously grown as well as the range of its applications.

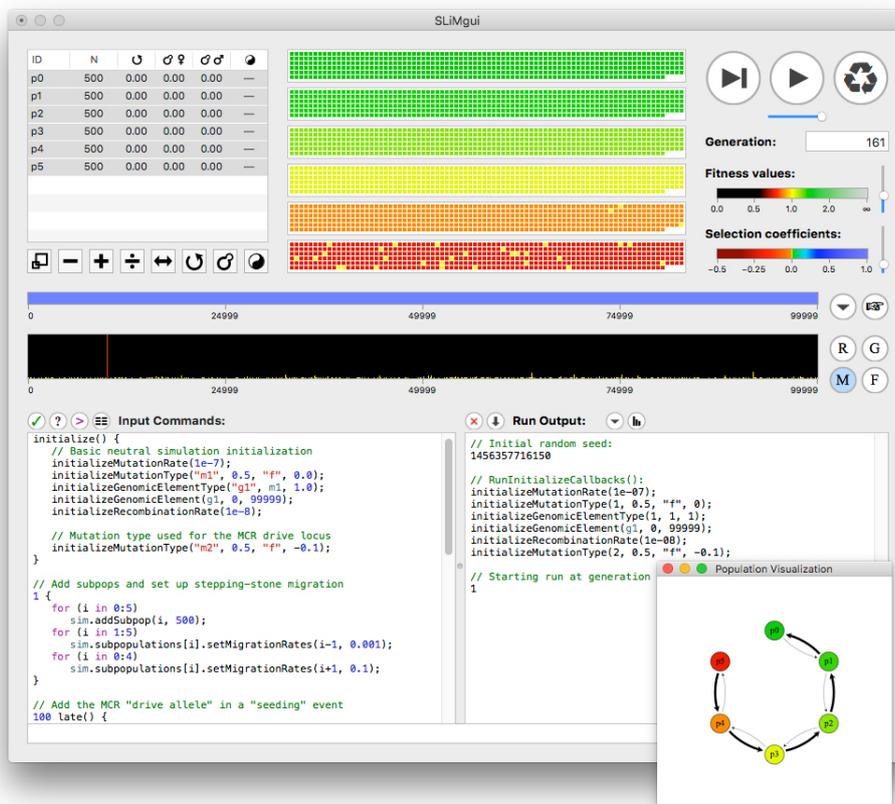


Figure 4. SLiMgui: SLiM graphical user interface - <https://messerlab.org/slim/>

3.3 - non-Wright Fisher model

In population genetics, a widely used model to make predictions or infer demographic processes is the Wright-Fisher model (Fisher 1922; Wright 1931). In SLiM, the Wright-Fisher system is the default setting for creating evolutionary scenarios, and it is designed as follows:

- Non overlapping generations, as each individual is born, reproduces and dies in the same generation,
- Fitness affects the probability of every individual to reproduce and not mortality,
- Population size is a fixed parameter pre-set by the user and not a value emerging from the characteristics of the population.

In addition to this simple model, SLiM offers the opportunity to implement a different nonWright-Fisher model, also known as the Moran model (Moran, 1962). In this nonWF model, generations are overlapping, meaning that every generation is not an actual generation of individuals but it is just an opportunity for every individual to be born, reproduce or die.

Age is considered as an individual's property, and it affects the probability of each individual to die. In our work, we decided to use a U-shaped age based mortality (Tidière et al., 2016; Lahdenperä et al. 2018), scaled for a small-medium sized mammal (Nowak 1991; Witting, 1997), meaning that the

probability of each individual to die is higher for the first year, very low for the years after the first and starts increasing after the fifth year of life, until it reaches 100% in the eighth year of life. (Figure 5).

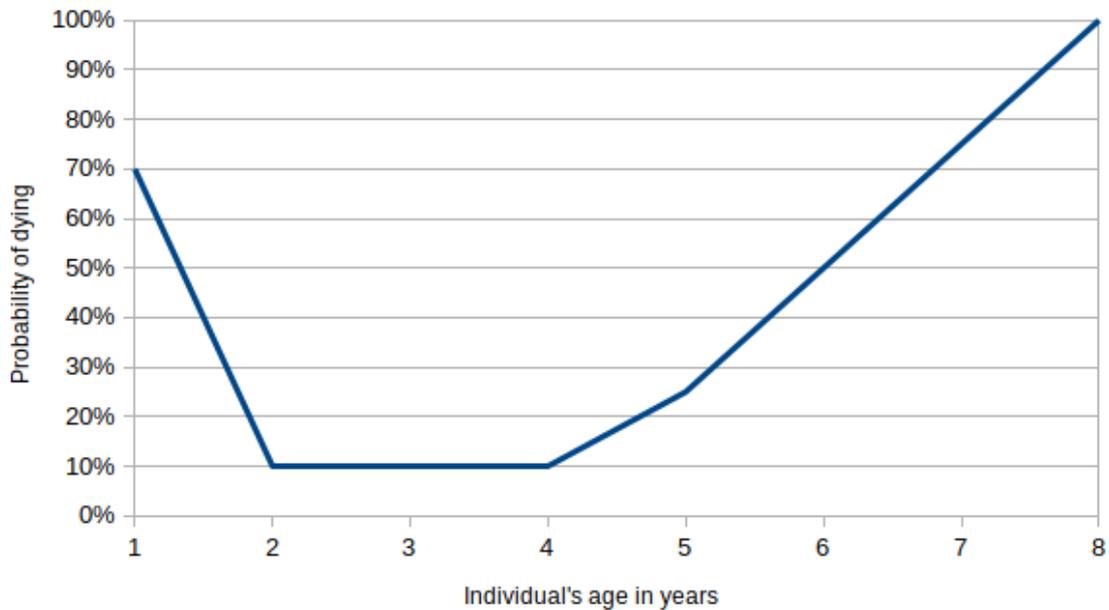


Figure 5. **Age based mortality.** The graph shows the probability of each individual to die in the current generation according to its age.

In nonWF models, population size and sex ratio are emergent from the simulation, and they are not parameters imposed by the user, it follows that population size is the result of how many offspring are created versus how many individuals die in each generation. This feature contributes to creating a more realistic and more dynamic evolutionary scenario.

In addition, fitness does not affect the probability of each individual to reproduce, but it only affects each individual's likelihood of surviving

To investigate the correlation between sexual dispersion and mito-nuclear discordance, we implemented our simulation using a non-Wright-Fisher model.

3.4 - Continuous space model

In SLiM, it is possible to create continuous spatial models, which allow the user to simulate one-dimensional, two-dimensional or three-dimensional geographic landscapes in which individuals can disperse and interact with each other. In these models, spatial coordinates are considered as properties of each individual and affect how individuals reproduce, disperse, and compete with each other. In continuous spatial models, it is possible to create complex spatial landscapes by including numerous environmental variables in order to evaluate their impact on the studied population.

In our model, we designed a simple homogeneous two-dimensional landscape to minimize the number of variables that can interfere with our system.

3.5 - Neutral conditions

All of the models we implemented in our experiment have been defined to work in neutral conditions, which means that all the mutations are neutral and in no way affect the probability of individuals to reproduce or survive. Selective processes are not something we want to investigate in this work, so in defining our model we have decided not to take these phenomena into consideration.

3.6 - Reference species

In this work, we used mammals as a reference biological system to set up our simulations. Sex biased dispersal is a common behaviour in mammalian species. In particular, male biased dispersal and female philopatry are commonly reported in this taxon (Dobson, 1982; Trochet et al., 2016). We then used dispersion rates and age-based mortality which could be broadly consistent with a small-medium size mammal species (Nowak 1991; Witting, 1997).

3.7 - Model design and basic parameters

In the first part of every SLiM's script the most important parameters of the simulations are defined: simulation options, mutation rate, recombination rate, length of the different components of the “chromosome”, simulation constants and so forth.

We start setting up a nonWright-Fisher, two-dimensional spatial model, and defining a carrying capacity of 15.000. The carrying capacity is a constant parameter that affects the population size (see below for further details). We define constant values to regulate the dispersal of the offspring and the age-based mortality. We also define that we want our model to include sexual reproduction and so different sexes enabling the “sex” parameter.

3.8 - The genome

In SLiM is it possible to work with different elements of the genome at the same time, but the software manages all the DNA as a single chromosome. For our work we wanted to create ten nuclear genomic elements alongside one mitochondrial genome.

To do that we created two different genomic elements, g1 and g2, that represent the nuclear portion and the mitochondrial portion of the genome respectively.

Our chromosome is composed of ten g1 nuclear elements of 16.000 bp each and one g2 mitochondrial element of 16.000 bp, for a total of 176.000 bp (Figure 6).

The recombination rate is set as 1×10^{-8} within the nuclear elements and as 0.5 between them, this last parameter allows us to create recombination breakpoints and to work with independent elements as we wanted to simulate unlinked nuclear regions. For the mitochondrial genome we are using a 0 recombination rate within the genome and of course a 0.5 recombination rate between the mitochondrial genome and the rest of the chromosome.

After that we set up the mutation rate as 1×10^{-7} for the mitochondrial DNA and 1×10^{-7} for the nuclear DNA. The latter is rather high as compared to the reference literature (Drake et al., 1998), but it is a shortcut to reduce the simulation time, as we are working on a little portion of the nuclear genome and we need enough mutations to work with. It is basically the same as working with 1.600.000 bp of nuclear DNA if the mutation rate is 1×10^{-8} .

Lastly, we create three different kinds of mutations, m1, m2 and m3:

- m1 is the nuclear mutation that occurs only in the nuclear portion of the genome, this mutation has 0.5 as dominance coefficient and a selection coefficient of 0.0.

- m2 mutation is the mitochondrial mutation that only occurs within the mitochondrial portion of DNA, it has a dominance coefficient of 1.0 and a selection coefficient of 0.0.
- m3 mutation is a marker mutation that allows us to simulate the maternally inheritance of the mitochondrial DNA.

Every mutation in this simulation is neutral as we do not want to take into account selective processes.



Figure 6. **Representation of the simulated genome from SLiM.** One copy of the genome is coloured in blue, while the other one is shown in purple. The black portion of the genome represents the mitochondrial genome (haploid). Yellow bars represent the frequency of the mutations in the population.

3.9 - Mitochondrial and nuclear inheritance models

SLiM works only with diploid genomes, so every deviation from this standard has to be implemented in the simulation. In our model, we need to simulate mitochondrial DNA alongside nuclear DNA. To do so we need to define that only a part of the genome is maternally inherited and haploid while the other part is inherited from both parents and it is diploid.

To simulate the haploidy of the mitochondrial genome, in every generation every m2 mutation, appearing on the second copy of the mitochondrial genome, is removed, allowing only one copy of Mt DNA accumulating mutations.

To simulate the maternal inheritance of the mitochondrial genome one important thing to know is that in SLiM the first copy of the genome of every individual is always inherited from the mother while the second one is always inherited from the father. When simulating the maternal inheritance of the mitochondrial DNA, we took this mechanism into consideration as we only needed to avoid the inheritance of the mother's second (disabled and without mutations) mitochondrial genome. We set up the model so that at each generation the second genome of each female is marked with the marker mutation m3 and, at the same time, every individual that has this mutation on its first genome is discarded. This prevents individuals with an incorrect inheritance of the mitochondrial genome from being born in the first place (Figure 7).

Another important thing to do is to remove mutations when they fix in the population. SLiM does it automatically, but for the mitochondrial genome we need to change this function in order to remove mutations when they reach fixation in a haploid system, meaning that we implemented a function that

removes every mitochondrial mutation which reaches a frequency equal or greater than 0.5.

For the nuclear genome, the standard SLiM setting works fine, so we run it with the default options.

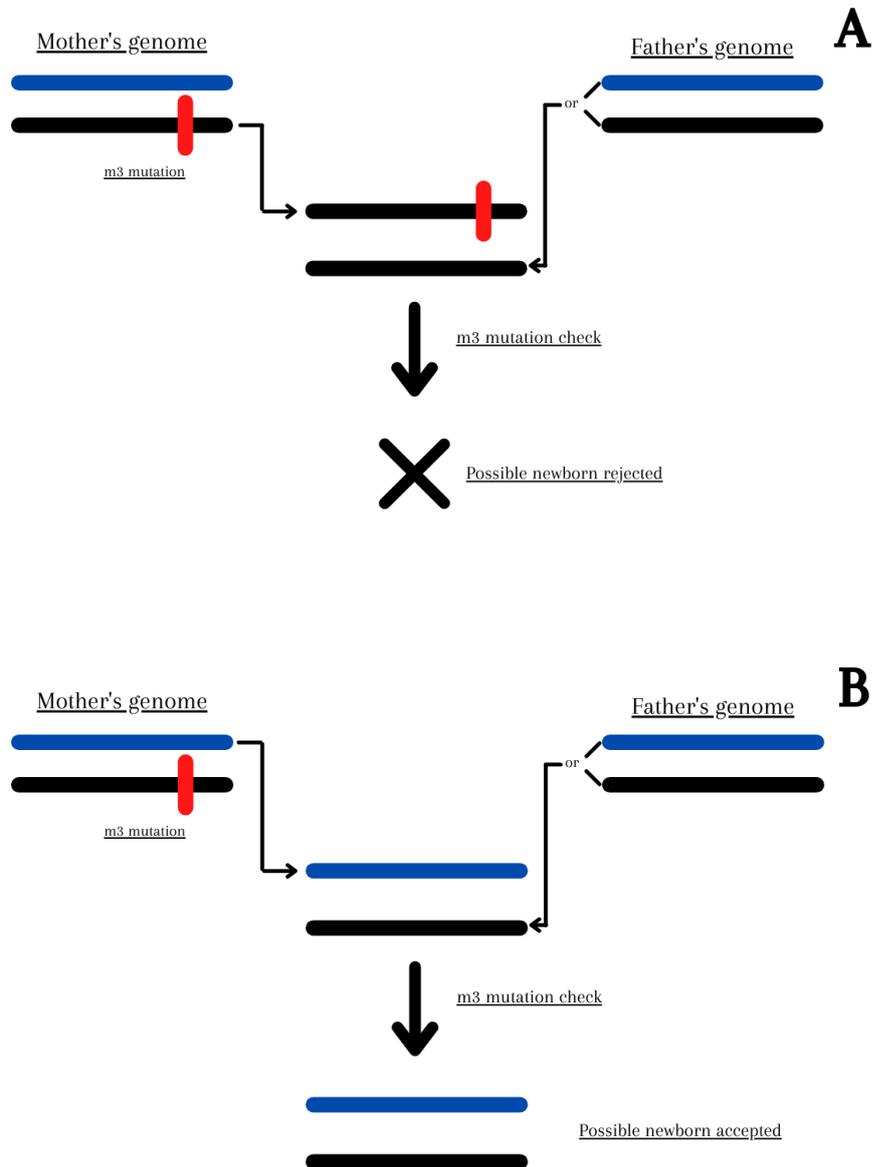


Figure 7. **Mitochondrial inheritance.** In the figure we can see the method used to ensure the correct inheritance of the mitochondrial genome. In both figures (A and B) we see that copy 2 of the genome

of the newborn is inherited from the father and can be copy 1 or 2 of the latter. This is irrelevant, as in the newborn copy 2 will in any case be deactivated. Copy 1 is inherited from the mother and the mechanism foresees, that only when copy 1 is inherited from the newborn, this is accepted, as in the m3 mutation check we make sure that the m3 mutation, generated on all copies 2 of the females, is not found on copy 1 of the newborn. Not shown in the picture: if the newborn is female, a new m3 mutation will be generated on copy 2 of its genome.

3.10 - *The simulated area*

As said above, we want to simulate individuals that disperse over a two-dimensional area, so we defined a homogeneous landscape map of 200 x 400 pixels. As we interpret a single pixel as a kilometer, our simulated area corresponds to 80.000 square kilometers. The so-defined landscape has boundaries at the edges, which means that the population is confined within the area.

To start the simulation with a population that presents the characteristics of an equilibrium population, we define the initial spatial coordinates of the single individuals by setting random values as X and Y coordinates that ensure that the spatial position of each individual is included within a square of 100 km x 100 km in the centre of the area (Image 8, panel A). From this initial position, with the generations progress, the population expands until it homogeneously colonizes the entire area (Image 8, panel B).

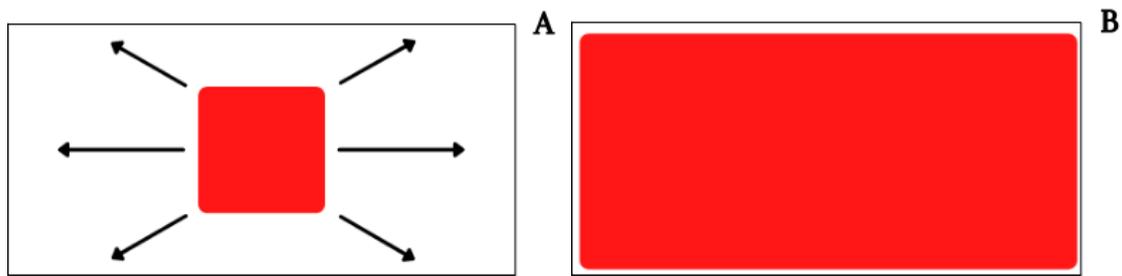


Figure 8. **Colonization of the area by the population.** At the beginning of the simulation individuals are randomly generated within a square area in the center of the landscape (panel A). After several generations, the population grows in number and expands to colonize the entire area (panel B).

3.11 - Sex biased dispersal simulation

To simulate different pattern of dispersal for individuals of different sexes, we define the dispersion function for males and females as constant values before the simulation's start. The sum of the two constants is always 10 kilometers, so that in every simulation we choose a value between 2, ~3.3, 5, ~6.6 or 8, and we assign this constant value to the male and calculate the female's dispersal constant as ten minus the male's dispersal constant. Using this approach, we make sure that the ratio between these two constants is always 0.25, 0.5, 1, 2, or 4. These values define the sex biased dispersal patterns that we will use in our five models (Table 2).

Model	1	2	3	4	5
Male dispersal	2Km	3,3333 Km	5 Km	6,66667 Km	2Km
Female dispersal	8 Km	6,66667 Km	5 Km	3,3333 Km	8 Km
Sex biased dispersal pattern	0,25	0,5	1	2	4

Table 2. **Dispersal constant values for male and female individuals.**

An important aspect of our dispersion simulation system is that the nuclear gene flow is expected to be rather constant across all models: as both sexes contribute to it and the sum of the male and female constants is always 10, the magnitude of the global movement of nuclear genes is fixed. This helps us further simplify the system so that we can focus on mitochondrial genetic diversity.

We apply the dispersion function on the offspring so that individuals move to a different location starting from their mother's spatial position in their first year of life only, while the extent of their movement is based on their sex.

The coordinates of the mother are defined by X and Y values. To generate the newborn's coordinates, two values are added to the X and Y coordinates of the mother, these values are random numbers taken from a normal distribution

function which has as 0 as mean and the sex dispersal constant assigned of the given model as standard deviation (Figure 9).

In this way, the newly generated child will move to a new position which is the result of the coordinates of the mother and of the two random values generated by the normal distributions for the X and Y axes.

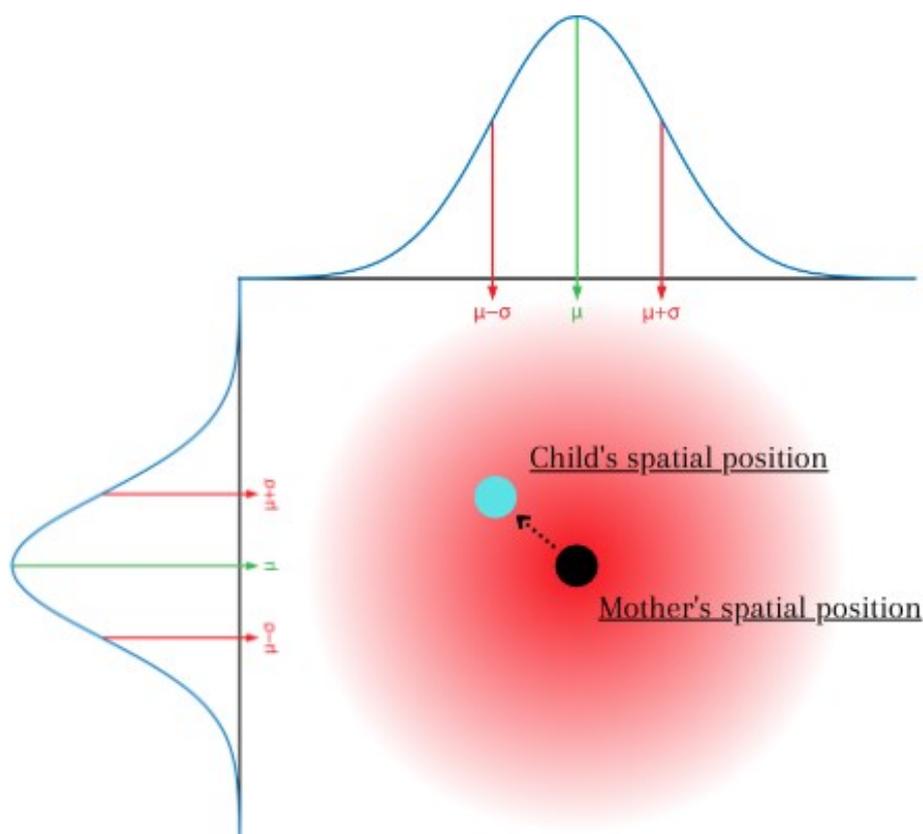


Figure 9. **Offspring's spatial dispersion function.** The scatter function assigns spatial coordinates to the child, using the mother's spatial coordinates as a reference and two random functions based on a normal distribution with mean 0 and standard deviation as the model's sex specific dispersal constant. The functions shown in the figure are for illustrative purposes only and do not derive from functions used in the simulation.

3.12 - *Fitness and population size*

In nonWright-Fisher models, fitness is defined as hard fitness, meaning that this individual's feature does not affect the probability of each individual to reproduce, but it affects its chances of surviving until the next generation. So, every individual that is alive during the reproduction event has the same probability to reproduce as the others. In our experiment this setting is fine as we are not working with selective processes. However, in our models, fitness is also important to regulate the population size.

In section 3.7, we defined a constant value for the carrying capacity allowing us to define a value around which the size of our population will fluctuate.

So, the population size in our simulations depends on three factors:

- Carrying capacity: a constant value defined by the user;
- Total density of the population in the area: as we want our population size to remain stable during all the simulation time and not indefinitely growing;
- Age based mortality: we define a probability of every individual to die during each generation, this probability is based on the individual's age and follows an "U shaped" distribution (Figure 5).

These three parameters together define the total population size. The carrying capacity of our model is set as a constant number of 15.000, causing the size of

our population to fluctuate around 10.000 individuals as population size is an emergent parameter in the simulation.

3.13 - *Spatial interactions*

To simulate more realistic interactions between individuals in our simulation we set up two spatial interactions: the first one is a positive interaction that regulates mate choice, while the second one is a negative interaction that regulates the competition between individuals which are in the same area.

3.13.1 *Reproduction interaction*

In our nonWright-Fisher models every individual has the same probability of reproducing as any other one without any effect of selection on reproduction. However, in our model individuals need to reproduce only with potential mates which are located nearby. We set up an interaction function which defines a circular area with radius 5 kilometers around every female; inside this area, the female “looks for” the nearest interacting male neighbor, using a specific SLiM’s function called *nearestinteractingneighbor* (Figure 10). If the function

finds a potential mate for the female, reproduction happens and the child is generated, otherwise the female does not reproduce.

Importantly, in nonWright-Fisher models population size is emergent from the carrying capacity, population density and age-based mortality, so most of the new individuals are only born if old individuals die. So, every generation there is a different number of newborns based on how many individuals died in the same generation.

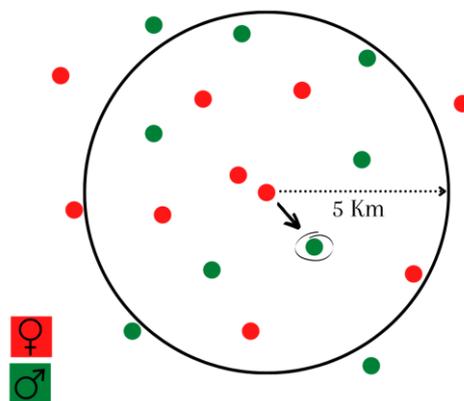


Figure 10. **Reproduction interaction.** In the image females are coloured in red while males are coloured in green. This image shows how females choose their potential mates.

3.13.2 Competitive interaction

In SLiM, individuals have no dimensionality, meaning that they can overlap on the same spatial coordinates and that there is no disadvantage for the individuals to live all gathered in the same area. As we want to simulate

mammal-like organisms dispersing across an area, we have to take into account the competition between individuals. To create a more realistic scenario we set a negative interaction function between newborns and other individuals. This function ensures that newborns have a lower probability of surviving in the first year of their life if they share the same area with too many potential competitors. Every generation for every newborn we define an area of 2 kilometers radius around it, then we look for other individuals that can compete with it, if there are five competitors or more within the defined area the probability of the newborn surviving to the next generation is highly reduced (Figure 11). Every individual in the simulation is considered a possible competitor.

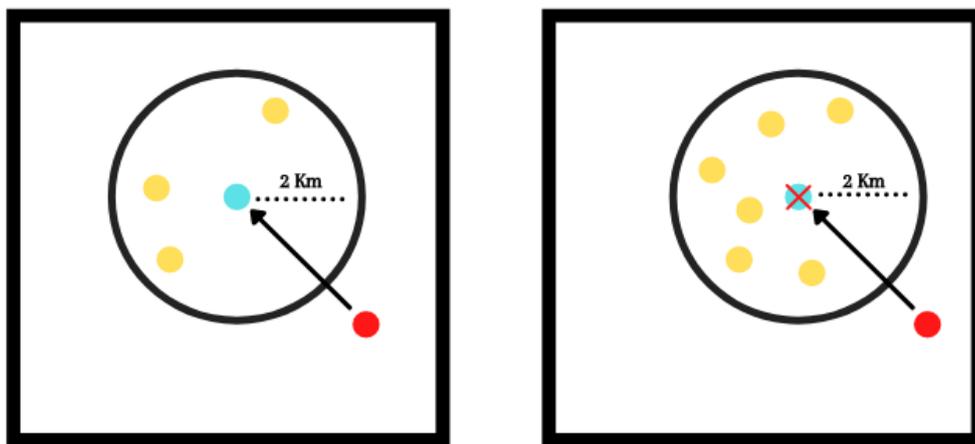


Figure 11. **Competitive interaction.** The mother's spatial position is coloured in red, while the newborn's spatial position is coloured in light blue and the potential competitors are coloured in yellow. The left image represents a situation in which the newborn moves, starting from the mother's spatial position to a place that is not sufficiently crowded to affect its fitness. On the right image the area is too crowded.

3.14 - *Simulation burn-in to reach stationarity*

In every simulation individuals start from a square area in the middle of the landscape and colonize all the area dispersing according to their sex. At the same time mutations start to occur, increase (or decrease afterwards) in frequency and, eventually, get fixed (or lost) in the population.

To obtain reliable estimates of the parameters of our modelled scenario, we set up our simulation with an initial burn-in to let the population reach an equilibrium. As summary statistics, we used the total population heterozygosity, calculated over the entire length of the nuclear genome and the mitochondrial genome, to understand how many generations to wait until the equilibrium is reached. In fact, in a population at its equilibrium, without any interfering force, average heterozygosity across many different loci is expected to stabilize around a value depending on population size.

Starting from generation 1.000 and every 500 generations, we used all the individuals in the population to calculate the total nuclear heterozygosity and the total mitochondrial heterozygosity (Figure 12). As we are working with a low number of mutations due to computational limitations, even after several thousands generations, heterozygosity can still fluctuate. In particular the mitochondrial heterozygosity shows a peculiar pattern that we discuss below.

To find the average generation at which all our models reach equilibrium, meaning that the heterozygosity value stabilizes, we ran 10 simulations for each of the five possible sex biased dispersal patterns (0.25, 0.5, 1, 2, 4) and observed the average heterozygosity through time. When the average value stabilizes for both genomes, we can say that the equilibrium is reached and considered the previous as the burn-in time. For our simulation, the estimated average burn-in time is 50.000 generations.

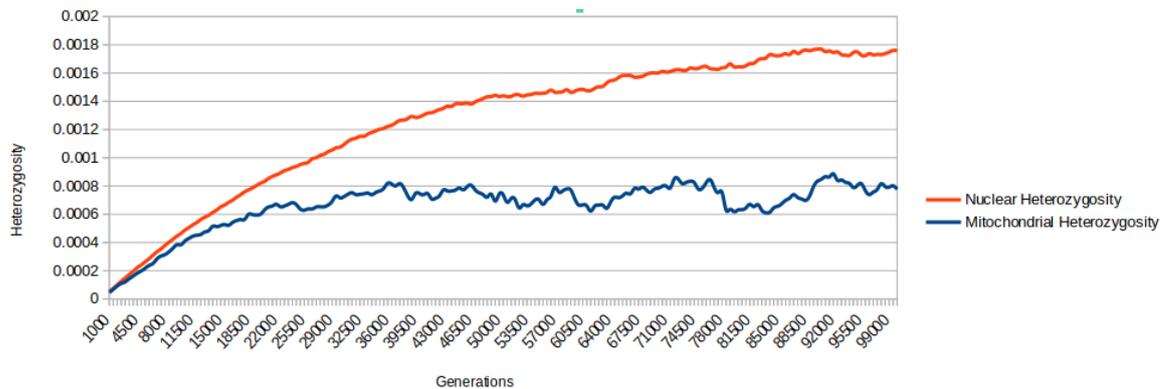


Figure 12. **Heterozygosity through generations.** Nuclear and mitochondrial total heterozygosity variation during all the simulation time. This graph shows the values related to the sex biased dispersal pattern 1.

3.15 - Sampling

As our task is to investigate how nuclear and mitochondrial genetic diversity change at different sampling distances across the five models, with different patterns of dispersal between the sexes, we sampled groups of individuals at

different predefined distances: 50, 100 and 150 kilometers apart. The sampling function, created specifically for this simulation, takes 10 pairs of samples to compare for each defined sampling distance. The algorithm is set to select an area of 14 x 14 km (minimum size to always sample at least 20 individuals), within which 20 individuals are sampled. At a predefined distance (among those pre-set), a second area of equal size to the first one is selected and 20 individuals are sampled within it (Figure 13).

The procedure is automatically performed 10 times for each defined sampling distance, so that we have 30 pairs of groups of individuals to compare for our analyses, 10 for each sampling distance.

This procedure is repeated every 500 generations starting from the end of the burn-in time, when the population has reached its equilibrium.

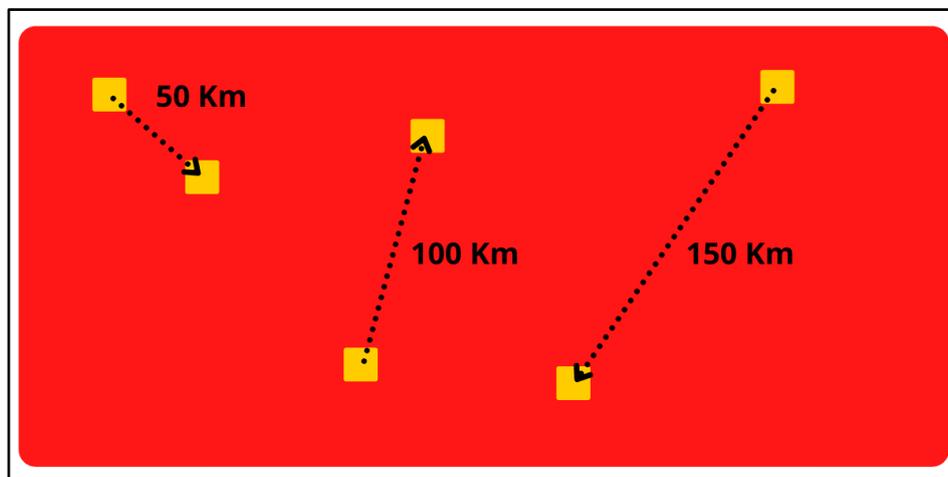


Figure 13. **Sampling distances.** The three different distances used to sample groups of individuals to compare.

3.16 - Summary statistics

This entire simulation system was designed to evaluate how different patterns of dispersal between the sexes can affect the mitochondrial and nuclear genetic diversity between individuals of the same population. One of the most widely used statistics for estimating genetic differentiation between individuals is F_{ST} (Wright, 1949).

$$F_{ST} = \frac{H_T - H_S}{H_T}$$

Equation 1. **Calculation of F_{ST} .** H_T refers to the total expected heterozygosity, while H_S refers to the average heterozygosity across the two subpopulations.

F_{ST} is used to reveal population's differentiation caused by population's structure (Weir & Cockerham, 1984) and in our case is calculated for both mitochondrial and nuclear genome, so we can have an estimation of the genetic diversity calculated for both genomes at the same time, giving us the possibility to compare these two parameters.

The reference values for this statistic are (Wright, 1978; Hartl e Clark, 1997):

- <0.05 = low genetic differentiation
- $0.05-0.15$ = moderate genetic differentiation
- $0.15-0.25$ = high genetic differentiation
- >0.25 = very high genetic differentiation

In our simulation, F_{ST} between pairs of samples is calculated every sampling event to estimate how different these two sampled groups are between them.

Another statistic that we calculate is the total heterozygosity. Heterozygosity consists in the condition of having two different alleles in one locus. It is a fundamental parameter for measuring the genetic variability between individuals, from which a great deal of information on the structure and history of a population can be inferred. In our model, heterozygosity is given by the sum of the frequency of heterozygotes for each polymorphic site in the genomes, calculated as $2pq$ (frequency of heterozygotes from the Equation 2), normalized for the length of the genome.

$$p^2 + 2pq + q^2 = 1$$

Equation 2. **Hardy-Weinberg equilibrium equation (Hardy, 1908; Weinberg, 1908).** In the equation a biallelic locus is considered, in which p and q are respectively the frequencies of one or the other allele, it follows that $2pq$ is the frequency of heterozygotes for that locus.

As said before we used this statistic to calculate the burn-in time and to better understand how our model works. As the F_{ST} , we calculate heterozygosity for both mitochondrial and nuclear genome, but we calculate this statistic for the whole population and not for groups of individuals.

3.17 - Running the simulations

To obtain sufficiently robust results we need to run multiple simulations, to have a sufficient number of simulations for every different sex biased dispersal pattern. Every simulation takes at least two hours on a single CPU and the time increases for highly different dispersion rates (model 1 and 5). To get enough data in a relatively short time is then necessary to work on computer clusters. Every simulation (Supplementary material, 1) can be run on a single CPU, so the only way to accelerate all the process is to run multiple simulations on multiple CPUs at the same time. To do that we created a script in Bash with the instructions to run the simulations (Supplementary material, 2) and then a for-loop script that reads and executes the instruction file multiple times (Supplementary material, 3).

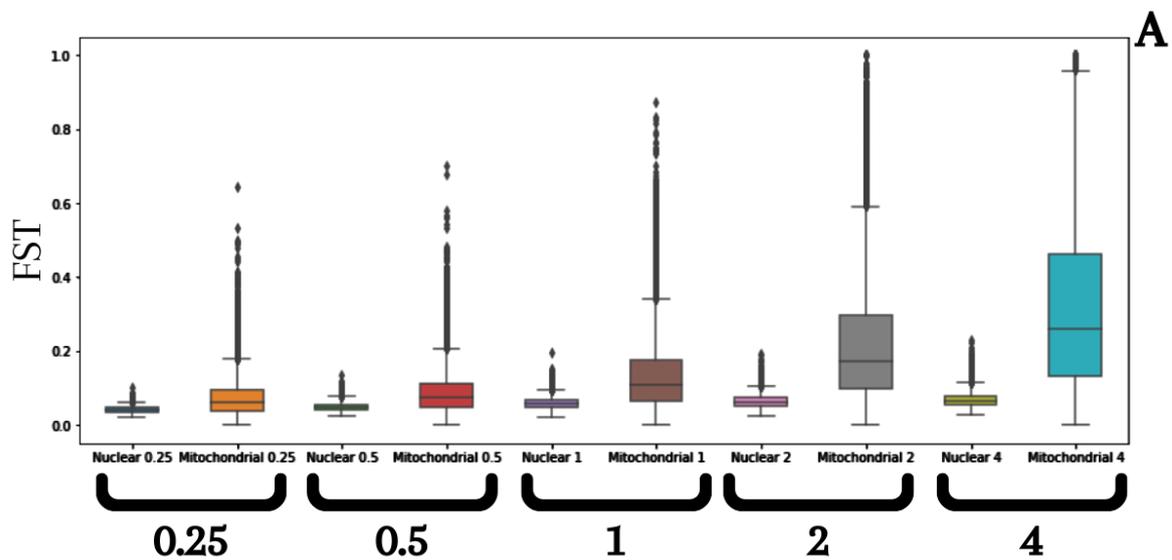
We created five different instruction files, every file used different dispersion rates for the sexes, so that we can have the same number of simulations for

every dispersion ratio. Usually, we ran from 10 to 20 simulations at the same time.

4. RESULTS

4.1 - Effect of sex biased dispersal patterns on genetic differentiation

The main objective of this work is to understand how mitochondrial genetic differentiation and nuclear genetic differentiation, calculated between pairs of samples, are influenced by different sex biased dispersal patterns. We used groups of individuals taken at different distances from each other to explore the role played by geographic distance on genetic differentiation, for different sex biased dispersal patterns.



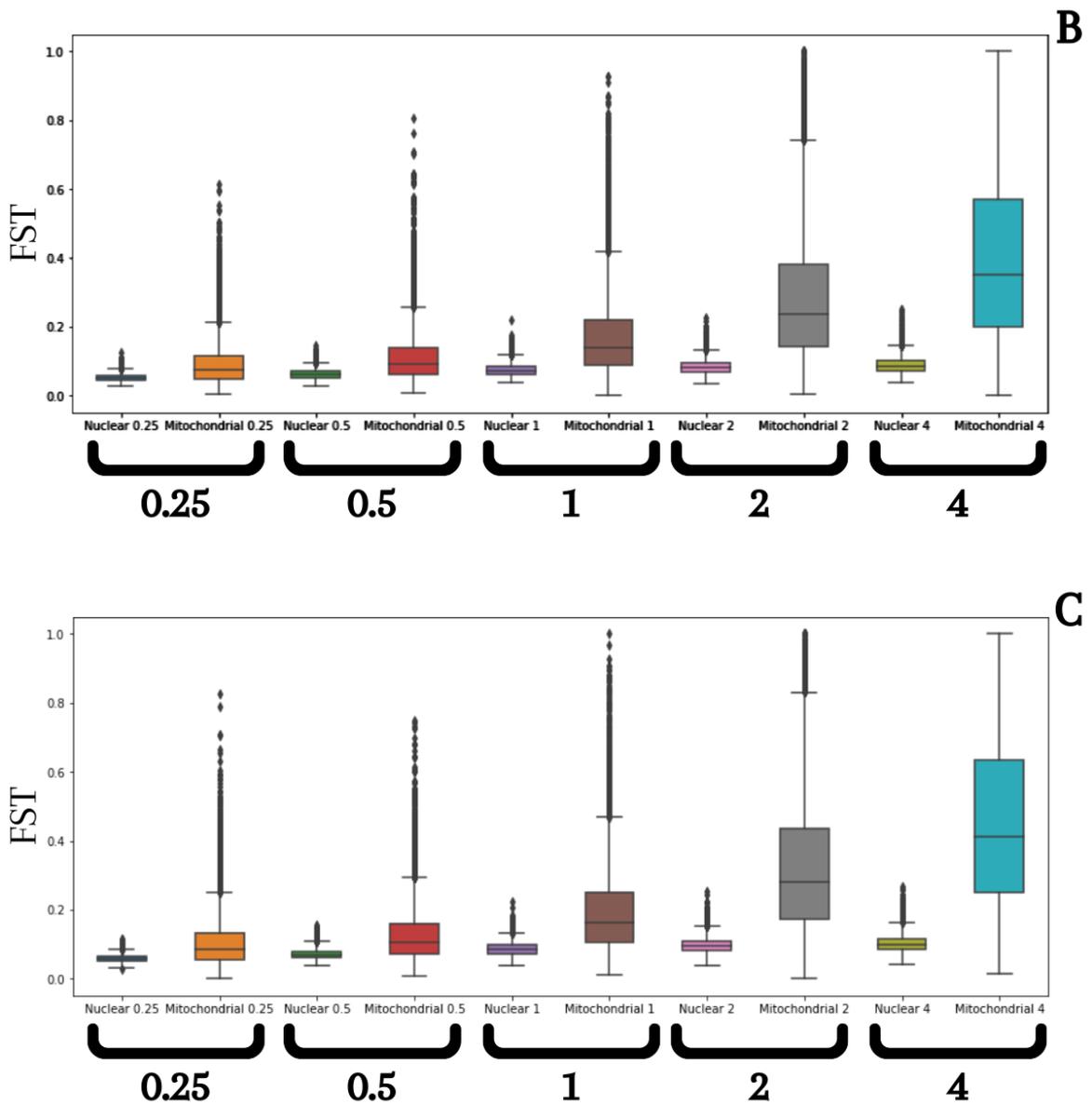


figure 14. **Effect of different patterns of sex biased dispersal on nuclear and mitochondrial genetic differentiation.** The three boxplots show the F_{ST} values, for the nuclear and mitochondrial genome, related to all of the samples from the five different models, with different patterns of sex biased dispersal: 0.25, 0.5, 1, 2, 4. The values are divided into the three panels according to the respective sampling distance. **A:** 50 Km, **B:** 100 Km, **C:** 150 Km

As clear in every comparison in figure 14, the mitochondrial differentiation values are, on average, always greater than the nuclear one. This is expected

due to the inherent smaller effective gene flow of the mitochondrial genome compared to the nuclear one.

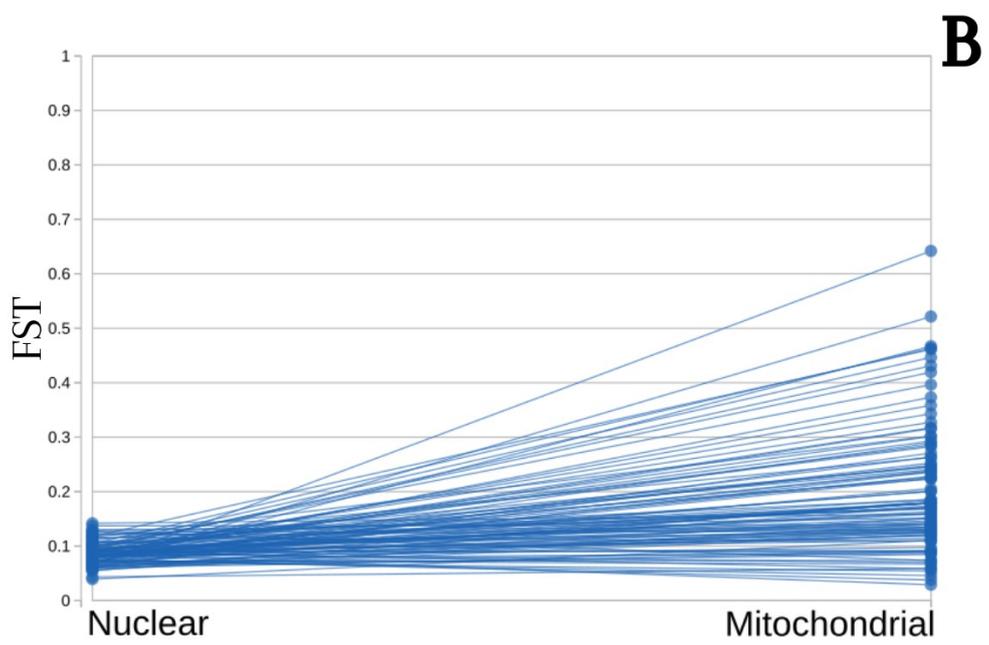
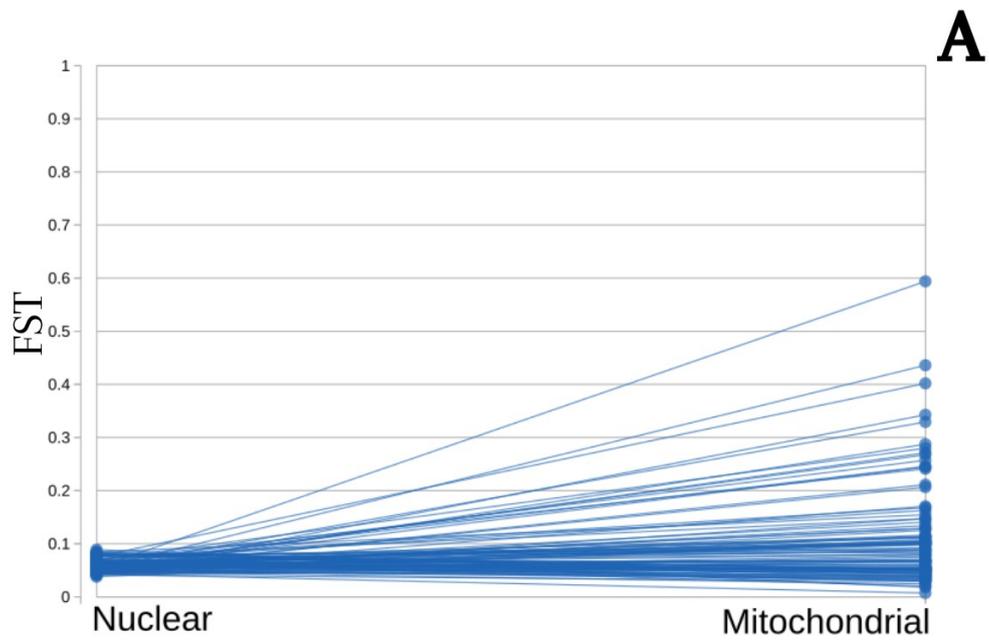
What is more interesting is how the different patterns of sex biased dispersal act on mitochondrial genetic differentiation: taking as a reference the 1:1 ratio between male and female dispersion (sex biased dispersal pattern = 1), we can appreciate how the mitochondrial F_{ST} between individuals decreases slightly when the dispersion ratio is unbalanced in favour of the female, while it increases considerably when the dispersion ratio is unbalanced in favour of the male. Our simulation results clearly show how a sex biased dispersal pattern characterized by a reduced rate of female dispersion, considerably accentuates mitochondrial genetic differentiation, leading to a situation where analyses based on the mitochondrial genome alone would, on average, reveal the presence of a strong structure within the population. Moreover, with a male to female dispersion ratio of 2 and 4, the variance of the mitochondrial genetic differentiation increases significantly for all three sampling distances used, meaning that there is the possibility to observe very high and moderately low F_{ST} values using mitochondrial markers.

Regarding the nuclear genetic differentiation, it is evident that the system used to determine the dispersion coefficients of the two sexes, with a fixed global

movement of nuclear genes, ensures that nuclear diversity remains largely unaltered by the different sex biased dispersal models (figure 14) and that in no model is it possible to observe the presence of a structured nuclear population. Nevertheless, it is possible to detect a slight increase of nuclear F_{ST} values in the models with lower female dispersal. This effect could be due to the fact that all the newborns move starting from a point of origin which is defined by the spatial coordinates of the mother. Due to this characteristic of the simulations, in these models the female could play a more important role than the male in determining the extent of nuclear gene flow. However, this hypothesis remains to be investigated by further simulations.

4.2 - Comparative distribution of F_{ST} values

To observe how different patterns of sex biased dispersal act on the relationship between mitochondrial and nuclear genetic differentiation in the same pair of samples, we directly compared the nuclear and mitochondrial F_{ST} values calculated for each pair of samples (Figure 15).



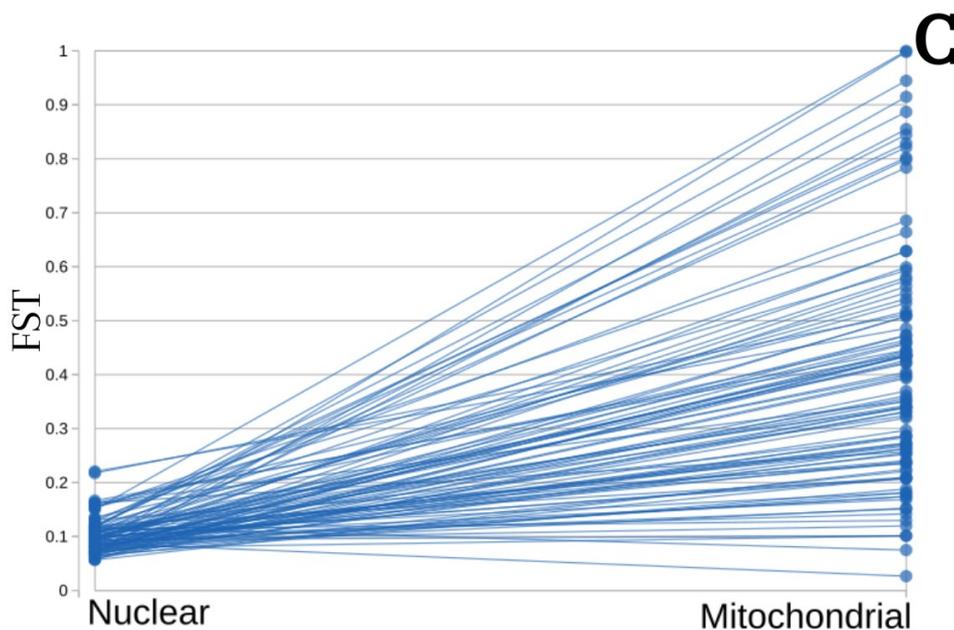


Figure 15. **Relationship between nuclear and mitochondrial genetic differentiation for pairs of samples for different patterns of sex biased dispersal (0.25, 1, 4).** The plots show the F_{ST} values for nuclear and mitochondrial genomes calculated between pairs of samples taken at the 100.000 generation using 150 Km as sampling distance. The panels show the results of three different models with different sex biased dispersal patterns: **A:** 0.25, **B:** 1, **C:** 4.

The most important result is the distribution of mitochondrial F_{ST} values for the different dispersion models, given that the distribution of nuclear F_{ST} values remains rather constant.

Observing these graphs, it is evident that the sex biased dispersal pattern in which female dispersion is reduced (Figure 15, Panel C) leads to a very wide distribution of mitochondrial F_{ST} values, with very high or extreme values of differentiation (which are not paralleled by an increase in the nuclear

differentiation), and cases in which it remains much more consistent with the nuclear F_{ST} values.

This highlights even more the previous observation: in models characterized by male biased dispersal and female philopatry, taking a few random samples, we can have situations in which the mitochondrial genetic differentiation between groups of individuals is very high and discordant with the nuclear differentiation or, on the contrary, rather low and more consistent with it.

This increased variance in the mitochondrial differentiation is, in our opinion, a strong indication of the possible correlation between sex biased dispersal and mito-nuclear discordance.

4.3 - The impact of geographic distance

To evaluate whether the geographical distance between samples had different effects according to the sex biased dispersal patterns, we compared the three sampling distances across the mean genetic differentiation for the different models.

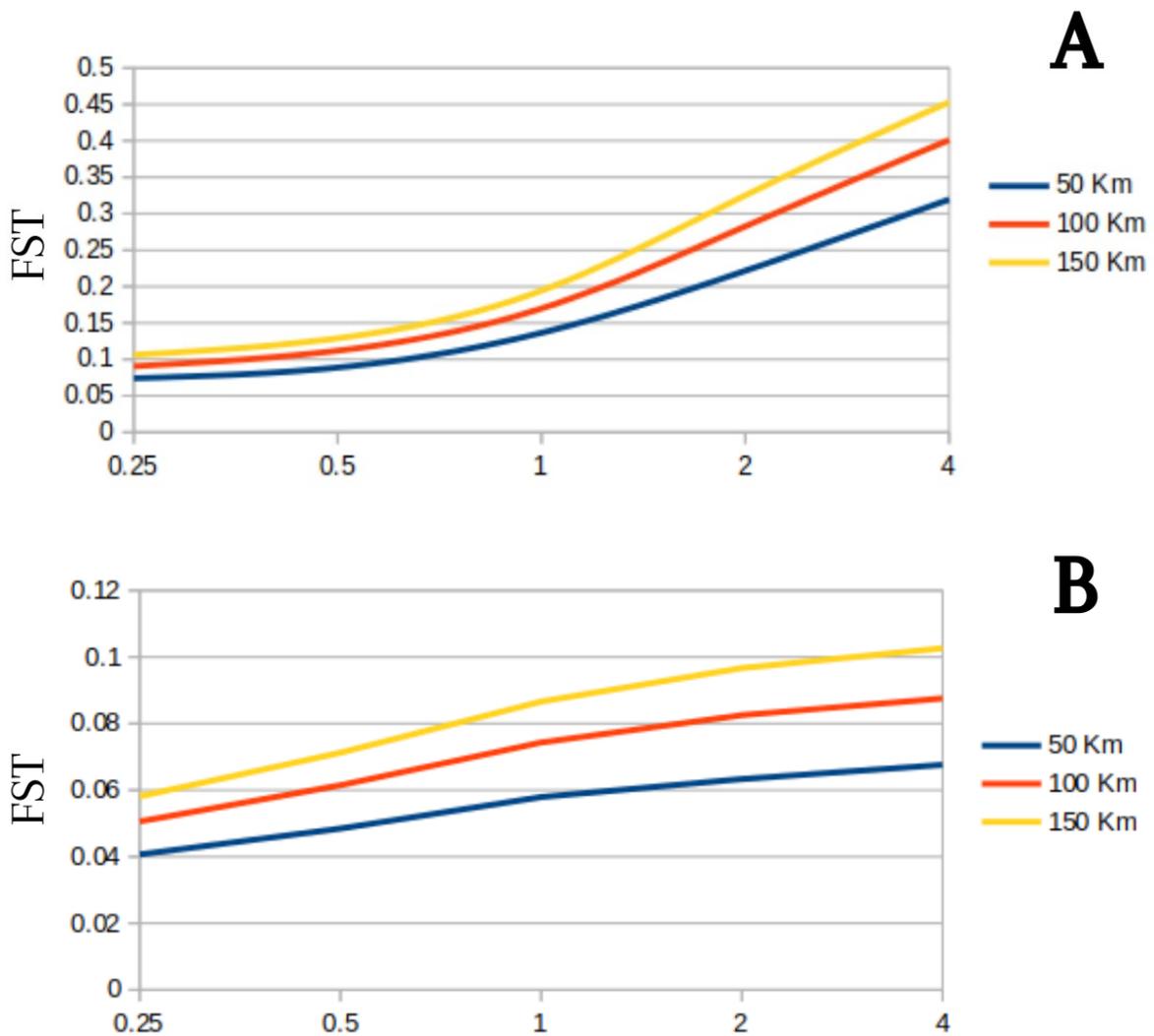


Figure 16. **Average genetic differentiation for different patterns of sex biased dispersal according to the sampling distance.** Panel A shows the results of the mitochondrial genome, where it is possible to see how, in the sex biased dispersal ratio 2 and 4 (X axis), the mean value of genetic differentiation (F_{ST}) increases considerably. Panel B shows the results of the nuclear genome. Note the different scale on the Y-axis between the two panels.

The results suggest that the variation in the values of F_{ST} , both mitochondrial (Figure 16, panel B) and nuclear (Figure 16, panel A), is always proportional to the distance in every model. Even if it is not so intense at this distance range, this pattern is consistent with an isolation by distance effect, where an

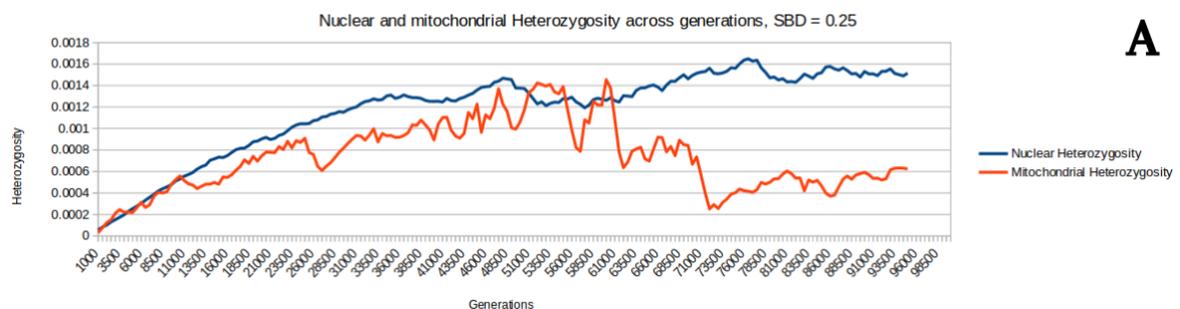
increasing geographic distance is always proportionally correlated to an increased genetic distance between individuals.

4.4 - Heterozygosity dynamics across generations

The heterozygosity calculated for the mitochondrial and nuclear genome for all individuals are two parameters that we have taken into consideration during all the simulations to calculate the burn-in time and to check that the simulation was proceeding correctly.

The results show that at about 50.000 generations, in all models, the average nuclear and mitochondrial heterozygosity across replicated runs reach a level of stability which we deemed sufficient to start the analytical phase.

However, in each individual replicated run, the mitochondrial heterozygosity showed rather unstable dynamics.



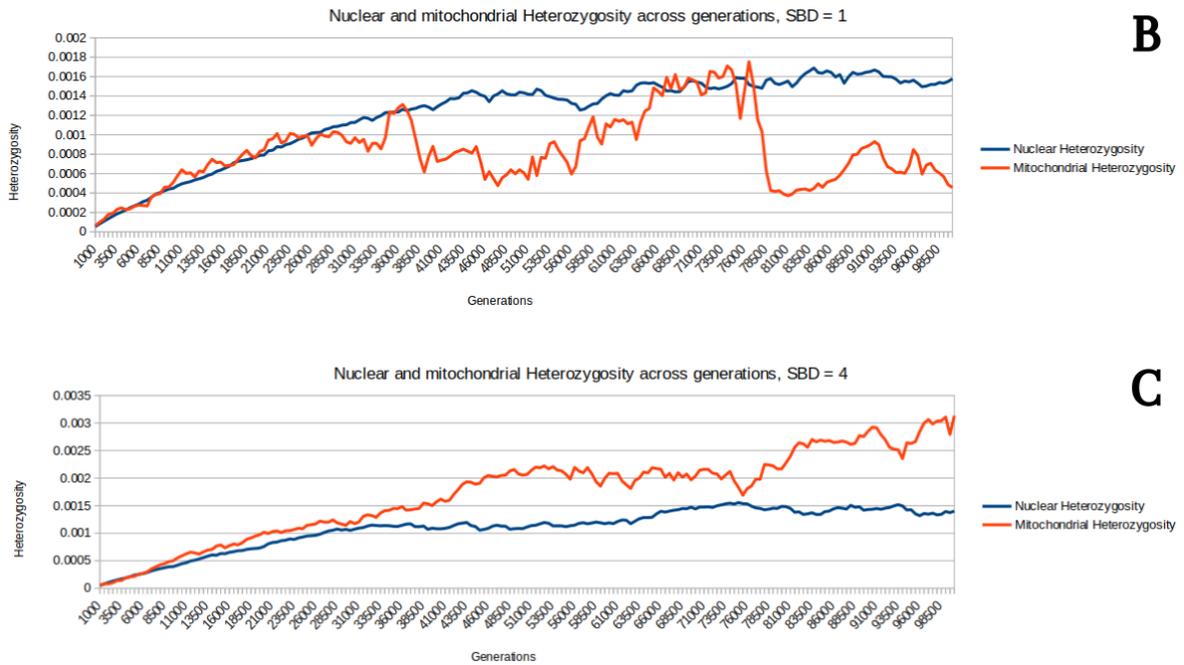


Figure 17. **Heterozygosity through generations for different sex biased dispersal patterns.** The graphs show Nuclear (blue) and mitochondrial (red) heterozygosity calculated for all the simulated individuals across 100.000 generations for sex biased dispersal patterns of 0.25 (panel A), 1 (panel B) and 4 (panel C). We used different scales to show the trend of the curves more clearly, as the total value is not something that interests us now.

In fact, regardless of the sex biased dispersal pattern imposed, the mitochondrial heterozygosity never seems to reach a stable equilibrium, but it is rather characterized by important variations across the generations (Figure 17). What is interesting is that when the dispersion ratio is 4 (Figure 17, panel C) we observe that the value of mitochondrial heterozygosity exceeds that of nuclear heterozygosity. Further analysing this process by running the same simulation for 500.000 generations and sampling individuals every 100 generations rather than 500 we can observe that the heterozygosity dynamics

is even more intense, resulting in very high peaks of mitochondrial heterozygosity followed by sudden collapses, while the nuclear heterozygosity remains stable (Figure 18).

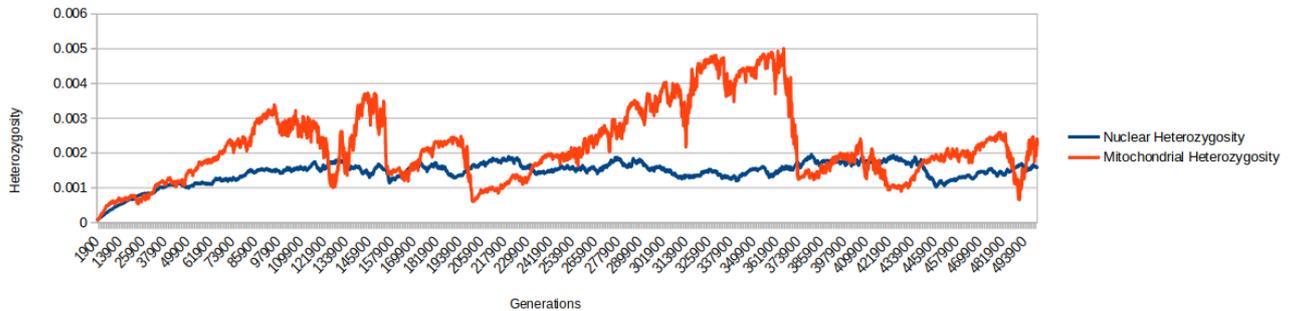


Figure 18. **Nuclear and mitochondrial heterozygosity variation across 500.000 generations.** In the model used to create this graph the male dispersal constant is 4 times the female's dispersal constant (SBD pattern = 4).

While the lack of recombination in the mitochondrial markers together with the more intense effect of drift could be the causes of these cycles of haplotypes quasi-fixation followed by new differentiation, the unstable dynamics of mitochondrial heterozygosity certainly deserves to be further investigated.

5. DISCUSSION

5.1 - Discussion

Our simulations clearly reveal that sex biased dispersal patterns, characterized by reduced female dispersion and therefore by a reduced mitochondrial gene flow, are always associated with mean, significantly higher mitochondrial genetic differentiation (F_{ST}) and with a greater variability in its range of values (Figure 14; Figure 15 panel C).

As a consequence of the reduced mitochondrial gene flow in models with lower female dispersal, the mitochondrial genetic diversity is on average significantly more structured than in the other models. At the same time, due to the inherent stochasticity of the dispersal process, we can observe either high or low values of mitochondrial F_{ST} (Figure 15, panel C).

The same does not hold for models in which the sex biased dispersal pattern is characterized by a female dispersion rate which is greater than or equal to that of males. In these models, in fact, we find that the mitochondrial genetic differentiation values are on average lower, indicating an absence of structure within the mitochondrial population and a distribution of F_{ST} values that appears to be consistent with the nuclear F_{ST} values (Figure 15, panels A and

B). Nuclear F_{ST} values never significantly increase and instead, always suggest the absence of structure within the nuclear diversity. This was expected as total nuclear gene flow is always constant in our models (Figure 14).

Our results show how by acting on a single variable (i.e. by decreasing the entity of the mitochondrial gene flow) there is a strong impact on the structure of the mitochondrial genetic diversity suggesting a strong correlation between sex biased dispersal and mito-nuclear discordance.

Minimizing the variables which were explored, our simulation design makes simplicity one of its strongest points, even if introducing important limitations. Defining a constant total nuclear gene flow certainly helps us to better understand the dynamics concerning the mitochondrial genome, but it does not make us fully understand the full contribution of the nuclear genome in these processes. Furthermore, the conditions of neutrality and the use of a completely homogeneous area where our simulated population thrives mean that our models are far from describing a completely realistic scenario.

These limitations are aspects that we have taken into account when developing our models, but our initial target was to design a simple but clearly functioning simulation system before adding more variables.

The simulation approach used proved to be a method lived up to the task assigned, showing a high efficiency and flexibility of use. Furthermore, it is evident that SLiM (Haller et al., 2019) is potentially able to respond to much more complex tasks than those it has already solved.

However, in order to convincingly prove the correlation between sex biased dispersal and mito-nuclear discordance, this work needs to implement other methods and other statistics. We know that the F_{ST} is a widely used statistics to estimate the structure of a population (Weir & Cockerham, 1984), but, concerning the mito-nuclear discordance, it can only give us some general indications. The main methods used in the literature to detect mito-nuclear discordance (Cahill et al., 2013; Folt et al., 2019; Firreno et al., 2019; Rakotoarivelo et al., 2019; Prous et al., 2019) preferentially used maximum likelihood gene tree topology comparison or bayesian clustering analysis (BCA) methods such as *Structure* (Falush, Stephens & Pritchard, 2003) and *Geneland* (Guillot et al., 2005). These systems will necessarily have to be implemented in our work to reach more clear-cut insights.

The investigation of the causes of mito-nuclear discordance in natural populations has been indicated as a priority for future studies on this subject (Toews et al., 2012). In recent years, the number of works that have set themselves the goal of detecting the phenomenon in nature and investigating its causes has increased. Most of the analyses carried out in recent works have mainly tried to correlate the genetic data of a certain population with its ecological and historical data (Cahill et al., 2013; Folt et al., 2019; Firreno et al., 2019).

In this context, our work is innovative, as it proposes a method never used before for this type of investigation. Although our system could be further improved to provide more relevant results, it is certainly a good starting point for future developments. Having designed and tested a fully controllable simulation system is certainly innovative and potentially very useful for investigating the correlation between mito-nuclear discordance and sex biased dispersal.

5.2 - Future directions

The models created for this project have a great potential for future investigations in this field, as we can imagine our simulation scenarios as working basic models in which we can implement new functions to further explore the mito-nuclear interactions in a two-dimensional area.

5.2.1 More realistic scenarios

One of the future objectives of the project is certainly to include new variables in the models and to evaluate their impact on the whole system. Among the variables evaluated so far, the most interesting to be implemented in the model are:

- Heterogeneous environment: by simulating natural barriers in our system we can evaluate how they affect the dispersion of different genetic markers.
- Different demographic histories: by simulating a more dynamic demographic history, with bottleneck events we will be able to understand how demography affects mito-nuclear interactions.
- Different patterns of dispersion: the system used for this project deliberately presents numerous simplifications. A good starting point for

the future development of this work is to set up a system in which the nuclear genetic flow is not always constant or in which individuals move also after the first year of life.

5.2.2 Use of phylogenetic trees as summary statistics

A common way to highlight the presence of mito-nuclear discordance in natural populations is to compare phylogenetic trees constructed using nuclear markers with those constructed using mitochondrial markers (Rakotoarivelo et al., 2019; Prous et al., 2019). In this way we can understand if the two trees are consistent with each other and with the geographical distribution of the individuals. For this reason, a possible implementation of our approach is to build phylogenetic trees, based on nuclear and mitochondrial genomes separately, using individuals sampled within the simulated area. This will allow us to test whether the discordance between nuclear and mitochondrial phylogenetic trees obtained from natural samples can also be observed by comparing trees obtained through simulations, and to evaluate the impact of different patterns of sex biased dispersal on this discrepancy.

5.2.3 Application to a real biological system

Once models that are capable of producing robust data, consistent with expectations, have been obtained, it will be possible to proceed with the application on a real study system. In our case, the target species on which to test the application is the African crested porcupine (*Hystrix cristata*). Previous studies (Trucchi et al., 2016) have pointed out that within the African population of this rodent there is an important discordance between the geographic distribution of nuclear and mitochondrial markers, particularly by observing the phylogenetic trees produced using the two different genomes (Trucchi, unpublished).

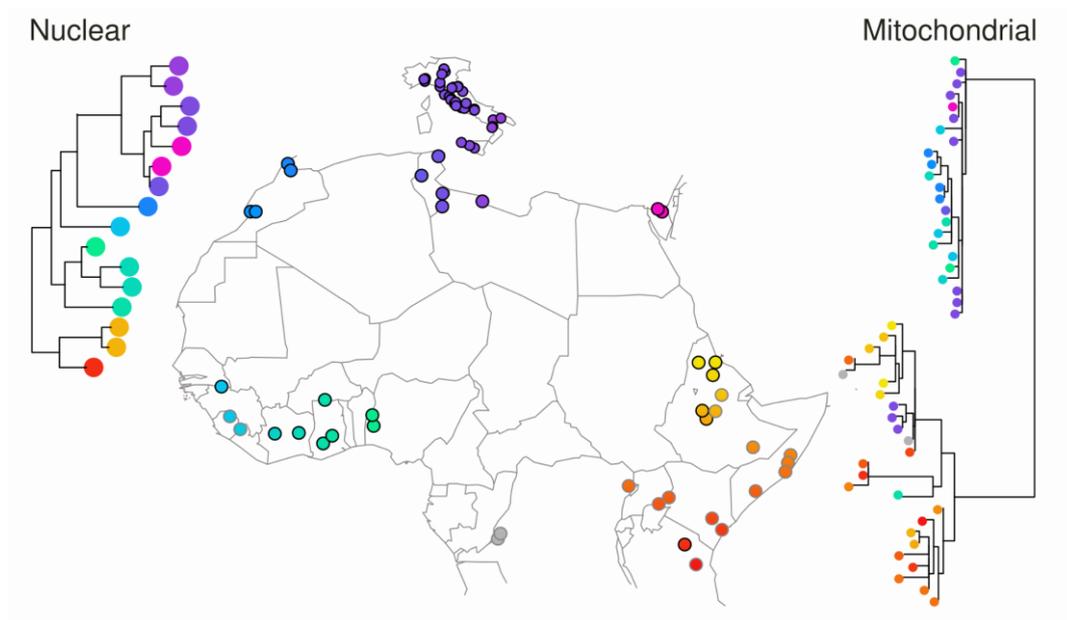


Figure 19. **Geographic distribution of African crested porcupine's nuclear and mitochondrial markers.** On the left is represented the phylogenetic tree built using nuclear markers, while on the right is represented the phylogenetic tree built using mitochondrial markers (Trucchi, unpublished).

Looking at Figure 19 it is clear that the phylogenetic tree based on nuclear markers seems to be coherent with the geographic distribution of individuals, while the phylogenetic tree based on mitochondrial markers is totally inconsistent with it.

The application on African crested porcupine should be conducted using specific simulation parameters able to describe the ecology and behaviour of this species.

5.2.4 Building a predictive model

A long-term objective of this project is to build a predictive model: a pre-set simulation model, capable of inferring a demographic parameter starting from a genetic parameter, or inferring a genetic parameter starting from a demographic parameter, simply by adapting the initial parameters of the simulation to the species to be studied. It could therefore be possible to have an estimate of the extent of sex biased dispersal starting from an estimation of the discrepancy between nuclear genome and nuclear genome or vice versa. Indications on how to calibrate this system could come from the application on the porcupine.

6. CONCLUSIONS

The main result that we can infer from the data obtained from the simulations is that a reduced female dispersion greatly affects the mitochondrial genetic differentiation between individuals within the population, leading more often to the formation of a highly structured mitochondrial population. The same is not applicable for the nuclear genome, as in these models it is minimally influenced by the different patterns of sex biased dispersal, and in all the simulations F_{ST} values are never high enough to suggest the presence of a structured nuclear population.

This suggests that, even if not considering all the variables at stake in a real population, the mitochondrial and the nuclear genome are subject to totally different dynamics simply due to their characteristics of inheritance and ploidy so that if sex biased dispersal arise in the species, this phenomenon has a very different impact on the two genomes leading to major inconsistencies between the two samples. These observations, however, are not yet sufficient to prove that patterns of sex biased dispersal may be at the basis of the observed cases of mito-nuclear discordance, but they are nevertheless excellent indications that there is indeed an important correlation between these two phenomena.

In the models implemented in this study, the sampling distance does not seem to be a critical parameter, as the increase in genetic differentiation is always consistent with the increase in the geographical distance between individuals. However, we think that exploring a different range of sampling distance can turn out to be important in future developments of this work.

The target of this project was to create a model as simple and controllable as possible in order to fully understand how it works and how the different patterns of sex biased dispersal affect the genetic distance between individuals calculated on both nuclear and mitochondrial genomes. Our results are therefore a fundamental step of a project that does not end with this work, but that aims to continue the investigation by including a growing number of variables to create an increasingly realistic evolutionary scenario.

BIBLIOGRAPHY

1. Bohonak, A. J. (1999). Dispersal, gene flow, and population structure. *The Quarterly review of biology*, 74(1), 21-45.
2. Cahill, J. A., Green, R. E., Fulton, T. L., Stiller, M., Jay, F., Ovsyanikov, N., ... & Shapiro, B. (2013). Genomic evidence for island population conversion resolves conflicting theories of polar bear evolution. *PLoS genetics*, 9(3), e1003345.
3. Dobson, F. S. (1982). Competition for mates and predominant juvenile male dispersal in mammals. *Animal behaviour*, 30(4), 1183-1192.
4. Drake, J. W., Charlesworth, B., Charlesworth, D., & Crow, J. F. (1998). Rates of spontaneous mutation. *Genetics*, 148(4), 1667-1686.
5. Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, 164(4), 1567-1587.
6. Fisher, R. A. (1923). XXI.—On the dominance ratio. *Proceedings of the royal society of Edinburgh*, 42, 321-341.
7. Firneno Jr, T. J., O'Neill, J. R., Portik, D. M., Emery, A. H., Townsend, J. H., & Fujita, M. K. (2020). Finding complexity in complexes: Assessing the causes of mitonuclear discordance in a problematic

- species complex of Mesoamerican toads. Molecular Ecology, 29(18), 3543-3559*
8. Folt, B., Bauder, J., Spear, S., Stevenson, D., Hoffman, M., Oaks, J. R., ... & Guyer, C. (2019). Taxonomic and conservation implications of population genetic admixture, mito-nuclear discordance, and male-biased dispersal of a large endangered snake, *Drymarchon couperi*. *PloS one, 14(3)*, e0214439
 9. Greenwood, P. J. (1980). Mating systems, philopatry and dispersal in birds and mammals. *Animal behaviour, 28(4)*, 1140-1162.
 10. Guillot, G., Mortier, F., & Estoup, A. (2005). GENELAND: a computer package for landscape genetics. *Molecular ecology notes, 5(3)*, 712-715.
 11. Haller, B. C., & Messer, P. W. (2019). SLiM 3: Forward genetic simulations beyond the Wright–Fisher model. *Molecular biology and evolution, 36(3)*, 632-637.
 12. Hardy, G. H. (1908). Mendelian proportions in a mixed population. *Science, 28(706)*, 49-50.
 13. Hartl, D. L., Clark, A. G., & Clark, A. G. (1997). *Principles of population genetics (Vol. 116)*. Sunderland, MA: Sinauer associates

14. Hoban, S., Bertorelle, G., & Gaggiotti, O. E. (2012). Computer simulations: tools for population and evolutionary genetics. *Nature Reviews Genetics*, *13*(2), 110-122.
15. Hudson, R. R., & Turelli, M. (2003). Stochasticity overrules the “three-times rule”: genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. *Evolution*, *57*(1), 182-190.
16. Kingman, J. F. C. (1982). The coalescent. *Stochastic processes and their applications*, *13*(3), 235-248.
17. Lahdenperä, M., Mar, K. U., Courtiol, A., & Lummaa, V. (2018). Differences in age-specific mortality between wild-caught and captive-born Asian elephants. *Nature communications*, *9*(1), 1-10
18. Lehmann, L., & Perrin, N. (2003). Inbreeding avoidance through kin recognition: choosy females boost male dispersal. *The American Naturalist*, *162*(5), 638-652.
19. Maddison, W. P. (1997). Gene trees in species trees. *Systematic biology*, *46*(3), 523-536.
20. Morales, H. E., Sunnucks, P., Joseph, L., & Pavlova, A. (2017). Perpendicular axes of differentiation generated by mitochondrial introgression. *Molecular Ecology*, *26*(12), 3241-3255.

21. Moran, P. A. P. (1962). The statistical processes of evolutionary theory.
The statistical processes of evolutionary theory.
22. Nowak, R. M., & Paradiso, J. L. (1991). *Mammals of the World*. John Hopkins Press, Baltimore.
23. Petit, R. J., & Excoffier, L. (2009). Gene flow and species delimitation.
Trends in Ecology & evolution, 24(7), 386-393.
24. Prous, M., Lee, K. M., & Mutanen, M. (2020). Cross-contamination and strong mitonuclear discordance in *Empria* sawflies (Hymenoptera, Tenthredinidae) in the light of phylogenomic data. *Molecular phylogenetics and evolution*, 143, 106670.
25. Rakotoarivelo, A. R., O'Donoghue, P., Bruford, M. W., & Moodley, Y. (2019). An ancient hybridization event reconciles mito-nuclear discordance among spiral-horned antelopes. *Journal of Mammalogy*, 100(4), 1144-1155.
26. Suh, A., Smeds, L., & Ellegren, H. (2015). The dynamics of incomplete lineage sorting across the ancient adaptive radiation of neoavian birds. *PLoS biology*, 13(8), e1002224.
27. Sunnucks, P. (2000). Efficient genetic markers for population biology. *Trends in ecology & evolution*, 15(5), 199-203.

28. Tidière, M., Gaillard, J. M., Berger, V., Müller, D. W., Lackey, L. B., Gimenez, O., ... & Lemaître, J. F. (2016). Comparative analyses of longevity and senescence reveal variable survival benefits of living in zoos across mammals. *Scientific reports*, 6(1), 1-7
29. Toews, D. P., & Brelsford, A. (2012). The biogeography of mitochondrial and nuclear discordance in animals. *Molecular ecology*, 21(16), 3907-3930.
30. Trochet, A., Courtois, E. A., Stevens, V. M., Baguette, M., Chaine, A., Schmeller, D. S., ... & Wiens, J. J. (2016). Evolution of sex-biased dispersal. *The Quarterly Review of Biology*, 91(3), 297-320.
31. Trucchi, E., Facon, B., Gratton, P., Mori, E., Stenseth, N. C., & Jentoft, S. (2016). Long live the alien: is high genetic diversity a pivotal aspect of crested porcupine (*Hystrix cristata*) long-lasting and successful invasion?. *Molecular Ecology*, 25(15), 3527-3539
32. Weinberg, W. (1908). *ber den Nachweis der Vererbung beim Menschen. Jahres. Wiertt. Ver. Vaterl. Natkd.*, 64, 369-382.
33. Weir, B. S., & Cockerham, C. C. (1984). Estimating *F*-statistics for the analysis of population structure. *evolution*, 1358-1370.
34. Witting, L. (1997). *A general theory of evolution. By means of selection by density dependent competitive interactions*, 330.

35. Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, 16(2), 97.
36. Wright, S. (1949). The genetical structure of populations. *Annals of eugenics*, 15(1), 323-354.
37. Wright, S. (1984). *Evolution and the genetics of populations, volume 4: variability within and among natural populations (Vol. 4)*. University of Chicago press.
38. Yuan, X., Miller, D. J., Zhang, J., Herrington, D., & Wang, Y. (2012). An overview of population genetic data simulation. *Journal of Computational Biology*, 19(1), 42-54.
39. Zink, R. M., & Barrowclough, G. F. (2008). Mitochondrial DNA under siege in avian phylogeography. *Molecular ecology*, 17(9), 2107-2121

SUPPLEMENTARY MATERIAL

Supplementary material 1. SLiM's script.

```
initialize() {
initializeSLiMModelType("nonWF");
initializeSLiMOptions(dimensionality="xy");
defineConstant("K", 15000); // carrying capacity
defineConstant("L", c(0.7, 0.1, 0.1, 0.1, 0.25, 0.5, 0.75, 1.0));

if (exists("slimgui"))
defineConstant("D", c(2, 3.333333333, 5, 6.666666667, 8));

defineConstant("Sm", sample(D, 1));
defineConstant("Sf", (10 - Sm));

defineConstant("d1", 50);
initializeSex("A");

initializeMutationRate(c(1e-7, 1e-7), c(160008, 176009));
initializeMutationType("m1", 0.5, "f", 0.0); //nuclear mutation
m1.convertToSubstitution = T;
initializeMutationType("m2", 1.0, "f", 0.0); //mitochondrial mutation
m2.convertToSubstitution = T;
initializeMutationType("m3", 1.0, "f", 0.0); //mitochondrial MARKER mutation

initializeGenomicElementType("g1", m1, 1.0); //nuclear genome
initializeGenomicElementType("g2", m2, 1.0); //mitochondrial genome

initializeGenomicElement(g2, 160009, 176009);
for (i in 1:10){
initializeGenomicElement(g1, ((i-1)*16000)+(i-1), (i*16000)+(i-2) );
}

initializeRecombinationRate(c(1e-8, 0.5, 1e-8, 0.5, 1e-8, 0.5, 1e-8, 0.5, 1e-8, 0.5,
1e-8, 0.5, 1e-8, 0.5, 1e-8, 0.5, 1e-8, 0.5, 1e-8, 0.5, 0), c(15999, 16000, 32000,
32001, 48001, 48002, 64002, 64003, 80003, 80004, 96004, 96005, 112005,
112006, 128006, 128007, 144007, 144008, 160008, 160009, 176009));

// spatial mate choice
initializeInteractionType(1, "xy", maxDistance = 5, sexSegregation= "FM");
i1.setInteractionFunction("I", 1.0);
initializeInteractionType(2, "xy", maxDistance= 2);
```

```

i2.setInteractionFunction("f", 1);
}

1 early() {
sim.addSubpop("p1", 500);
p1.individuals.age = rdunif(500, min=0, max=7);
for (ind in p1.individuals) {
px = runif(1, 150, 250);
py = runif(1, 50, 150);
p = (c(px, py));
ind.setSpatialPosition(p);}
p1.setSpatialBounds(c(0.0, 0.0, 400.0, 200.0));
mapImage = Image("Landscape2.png");
p1.defineSpatialMap("world", "xy", (1.0) - mapImage.floatG,
valueRange=c(0.0, 1.0), colors=c("#682C09", "#1BA81B"));
}

//Reproduction happens only between individuals that are close to each other
reproduction(NULL, "F") {
male = i1.nearestInteractingNeighbors(individual, 1);
if (male.size() > 0)
subpop.addCrossed(individual, male);
else
return;
}

modifyChild() {
//Female dispersal
if (child.sex == "F"){
do child.x = parent1.x + rnorm(1, 0, Sf);
while ((child.x < 0.0) | (child.x > 400.0));
do child.y = parent1.y + rnorm(1, 0, Sf);
while ((child.y < 0.0) | (child.y > 200.0));}

//Male dispersal
if (child.sex == "M"){
do child.x = parent1.x + rnorm(1, 0, Sm);
while ((child.x < 0.0) | (child.x > 400.0));
do child.y = parent1.y + rnorm(1, 0, Sm);
while ((child.y < 0.0) | (child.y > 200.0));}

//Mito-Nuclear simulation
if (childGenome1.containsMarkerMutation(m3, 175000))
return F;
if (childIsFemale)
child.genome2.addNewMutation(m3, 0.0, 175000);

```

```

return T;
}

1:100000 early() {
// life table based individual mortality
inds = p1.individuals;
ages = inds.age;
mortality = L[ages];
survival = 1 - mortality;
inds.fitnessScaling = survival;
// density-dependence, factoring in individual mortality
p1.fitnessScaling = K / (p1.individualCount * mean(survival));

i2.evaluate();

//If a newborn moves into a crowded area has an high probability of dying
juveniles = p1.sampleIndividuals(10000, maxAge=1);
for (ind in juveniles){
competitors = i2.nearestNeighbors(ind, 100);
if (competitors.size() >= 5){
ind.fitnessScaling = 0.005;}}

late () {
Mmut = sim.mutationsOfType(m2);

// remove any new mutations added to the disabled diploid genomes
sim.subpopulations.individuals.genome2.removeMutations(Mmut);

// remove mutations in the haploid genomes that have fixed
freqs = sim.mutationFrequencies(p1, Mmut);
if (any(freqs >= 0.5))
sim.subpopulations.genomes.removeMutations(Mmut[freqs >= 0.5], T);

i1.evaluate();}

1 late (){
  catn("Sf= " + Sf + " " + "Sm= " + Sm);}

1000:100000 late () {
if (sim.generation % 500 != 0)
return;
//creates the areas of sampling
p1select =
p1.individuals[p1.individuals.x>0&p1.individuals.x<100&p1.individuals.y>50&p1.in
dividuals.y<150];

```

```

p2select =
p1.individuals[p1.individuals.x>175&p1.individuals.x<225&p1.individuals.y>50&p1.
individuals.y<150];
p3select =
p1.individuals[p1.individuals.x>300&p1.individuals.x<400&p1.individuals.y>50&p1.
individuals.y<150];
if (p1select.size() >= 40 & p2select.size() >= 40 & p3select.size() >= 40){

sp1 = sample(p1select, 40);
sp2 = sample(p2select, 40);
sp3 = sample(p3select, 40);

sp1n = sp1.genomes;
sp2n = sp2.genomes;
sp3n = sp3.genomes;

sp1m = sp1.genome1;
sp2m = sp2.genome1;
sp3m = sp3.genome1;

Nmuts = sim.mutationsOfType(m1);
Mmuts = sim.mutationsOfType(m2);

H1n = calcHeterozygosity(sp1n, Nmuts, start= 0, end= 160008);
H2n = calcHeterozygosity(sp2n, Nmuts, start= 0, end= 160008);
H3n = calcHeterozygosity(sp3n, Nmuts, start= 0, end= 160008);

H1m = calcHeterozygosity(sp1m, Mmuts, start= 160008, end= 176009);
H2m = calcHeterozygosity(sp2m, Mmuts, start= 160008, end= 176009);
H3m = calcHeterozygosity(sp3m, Mmuts, start= 160008, end= 176009);

p1inds = p1.sampleIndividuals(10000);
p1genome1 = p1inds.genome1;

HTotn = calcHeterozygosity(p1.genomes, Nmuts, start= 0, end= 160008);
HTotm = calcHeterozygosity(p1genome1, Mmuts, start= 160008, end= 176009);

FST12n = calcFST(sp1n, sp2n, Nmuts, start= 0, end= 160008);
FST13n = calcFST(sp1n, sp3n, Nmuts, start= 0, end= 160008);
FST23n = calcFST(sp2n, sp3n, Nmuts, start= 0, end= 160008);

FST12m = calcFST(sp1m, sp2m, Mmuts, start= 160008, end= 176009);
FST13m = calcFST(sp1m, sp3m, Mmuts, start= 160008, end= 176009);
FST23m = calcFST(sp2m, sp3m, Mmuts, start= 160008, end= 176009);

```

```
MNnump1 = sp1n.countOfMutationsOfType(m1);
MMnump1 = sp1m.countOfMutationsOfType(m2);
```

```
MNnump2 = sp2n.countOfMutationsOfType(m1);
MMnump2 = sp2m.countOfMutationsOfType(m2);
```

```
MNnump3 = sp3n.countOfMutationsOfType(m1);
MMnump3 = sp3m.countOfMutationsOfType(m2);
```

```
Di = (Sm/Sf);
```

```
catn (" ");
catn ("Generation" + " " + sim.generation + " " + "Di: " + Di);
catn("Heterozygosity: " + sim.generation + " " + H1n + " " + " " + H1m + " " + H2n +
" " + " " + H2m + " " + H3n + " " + " " + H3m + " " + HTotn + " " + HTotm);
cat("Mutation count ");
catn(" ");
catn(" " + MNnump1 + " " + MNnump2 + " " + MNnump3);
catn(" ");
catn(" " + MMnump1 + " " + MMnump2 + " " + MMnump3);
catn(" ");
catn("FST12: " + "Nu: " + FST12n + " " + "Mt: " + FST12m);
catn("FST13: " + "Nu: " + FST13n + " " + "Mt: " + FST13m);
catn("FST23: " + "Nu: " + FST23n + " " + "Mt: " + FST23m);}}
```

```
50000:100000 late () {
if (sim.generation % 500 != 0)
return;
```

```
p1inds = p1.sampleIndividuals(10000);
p1genome1 = p1inds.genome1;
```

```
Nmut = sim.mutationsOfType(m1);
Mmut = sim.mutationsOfType(m2);
```

```
HTotn = calcHeterozygosity(p1.genomes, Nmut, start= 0, end= 160008);
HTotm = calcHeterozygosity(p1genome1, Mmut, start= 160008, end= 176009);
```

```
//Sampling
for (i in 1:3){
```

```
c = 0;
```

```
do{
```

```

ind = p1.sampleIndividuals(1);

p4select = p1.individuals[(p1.individuals.x <= (ind.x + 7) & p1.individuals.x >=
(ind.x - 7)) & (p1.individuals.y >= (ind.y - 7) & p1.individuals.y <= (ind.y + 7))];

d1a = (d1 * i);
d1b = (d1a + 5);

if (p4select.size() >=20){

paselect = p1.individuals[(p1.individuals.x > ind.x + d1a & p1.individuals.x < ind.x
+ d1b & p1.individuals.y < ind.y + d1b & p1.individuals.y > ind.y - d1a)];
pbselect= p1.individuals[(p1.individuals.x < ind.x - d1a & p1.individuals.x > ind.x -
d1b & p1.individuals.y < ind.y + d1b & p1.individuals.y > ind.y - d1a)];

pcselect= p1.individuals[(p1.individuals.y > ind.y + d1a & p1.individuals.y < ind.y +
d1b & p1.individuals.x < ind.x + d1b & p1.individuals.x > ind.x -d1a)];

pdselect= p1.individuals[(p1.individuals.y < ind.y - d1a & p1.individuals.y > ind.y -
d1b & p1.individuals.x < ind.x + d1b & p1.individuals.x > ind.x - d1a)];

p5select = c(paselect, pbselect, pcselect, pdselect);
if (p5select.size() >= 1){
samp = sample(p5select, 1);
p6select = p1.individuals[(p1.individuals.x <= (samp.x + 7) & p1.individuals.x >=
(samp.x - 7)) & (p1.individuals.y >= (samp.y - 7) & p1.individuals.y <= (samp.y +
7))];}

if (p6select.size() >= 20){
p4sample = sample(p4select, 20);
p6sample = sample(p6select, 20);
p4samplegenomes = p4sample.genomes;
p4samplegenome1 = p4sample.genome1;
p6samplegenomes = p6sample.genomes;
p6samplegenome1 = p6sample.genome1;

Mmuts = sim.mutationsOfType(m2);
Nmuts = sim.mutationsOfType(m1);

p4nuclearThetaW = calcWattersonsTheta(p4samplegenomes, muts= Nmuts,
start= 0, end= 160008);
p6nuclearThetaW = calcWattersonsTheta(p6samplegenomes, muts= Nmuts,
start= 0, end= 160008);
p4mitoThetaW = calcWattersonsTheta(p4samplegenome1, muts= Mmuts, start=
160008, end= 176009);
p6mitoThetaW = calcWattersonsTheta(p6samplegenome1, muts= Mmuts, start=
160008, end= 176009);

```

```

FSTnu = calcFST(p4samplegenomes, p6samplegenomes, start= 0, end=
160008);
FSTmt = calcFST(p4samplegenome1, p6samplegenome1, start= 160008, end=
176009);
Di = (Sm/Sf);

if (isNULL(FSTnu) | isNULL(FSTmt))
c = c+0;
else{
c= c+1;
catn("Seed: " + getSeed() + " " + sim.generation + " " + d1a + " " + Di + " " +
FSTnu + " " + FSTmt + " " + HTotn + " " + HTotm);}}}}
while (c < 10);}}

```

Supplementary material 2. Bash script to run the simulation.

```

#!/bin/bash
#PBS -l nodes=node2-gen:ppn=1
#PBS -l walltime=1000:00:00
#PBS -q long

#where all the steps will be performed:
cd /giorgio/data1/stud/fgiannelli/simul

/giorgio/data1/stud/fgiannelli/bin/build/slim -s $seed -d D=8
Modellodefinitivocluster.slim > 4_"$seed"_out

```

Supplementary material 3. Bash for loop. used to launch the simulations multiple times.

```

for i in {1..10};do qsub -v seed=$i slim.lanc ;done

```

