



**HAL**  
open science

# Speciation and organellar genome evolution in lineages of *Silene nutans* (Caryophyllaceae)

Zoé Postel

► **To cite this version:**

Zoé Postel. Speciation and organellar genome evolution in lineages of *Silene nutans* (Caryophyllaceae).  
Vegetal Biology. Université de Lille, 2022. English. NNT : 2022ULILR080 . tel-04318939

**HAL Id: tel-04318939**

**<https://theses.hal.science/tel-04318939>**

Submitted on 2 Dec 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Ecole doctorale Science de la Matière, du Rayonnement et de l'Environnement – SMRE ED 104  
UMR Evolution, Ecologie, Paléontologie – CNRS/Université de Lille – UMR 8198

# Speciation and organellar genome evolution in lineages of *Silene nutans* (Caryophyllaceae)

## Spéciation et évolution des génomes organellaire chez des lignées de *Silene nutans* (Caryophyllaceae)

Par Zoé Postel

Soutenue le 8 décembre 2022 en vue d'obtenir le grade de docteur en biologie de l'environnement, des organismes et des populations, écologie, devant le jury composé de :

Rapporteurs

**Sophie Karrenberg** (Professeur - *Uppsala University – Sweden*)  
**Alexander Papadopoulos** (Senior lecturer - *Bangor University – Wales*)

Examineurs

**Jacqui Shykoff** (DR CNRS - *CNRS, Université de Paris-Saclay – France*)  
**Fabienne Van Rossum** (Senior researcher - *Meise Botanical Garden – Belgium*)  
**Xavier Vekemans** (Professeur – *Université de Lille – France*) - Président

Directeur de thèse

**Pascal Touzet** (Professeur – *Université de Lille – France*)





## Remerciements / Greetings

Par où commencer ? Ou plutôt par qui ?

J'avais d'abord écrit trois pages, mais il paraît qu'il faut faire court... Alors, en une page à peine - et en ayant conscience que j'en oublierai certains et certaines et en les priant par avance de m'en excuser - je dois aller à l'essentiel !

Commençons par les institutions : j'adresse tous mes remerciements à la Région Haut de France et à l'Université de Lille qui m'ont fait confiance, m'ont accueillie et ont financé ce doctorat. Merci aussi au laboratoire EEP, indispensable cadre de mes travaux ainsi qu'à l'IREPSE et à l'I-site ULNE pour avoir financé mes mobilités. J'espère que le résultat sera à la hauteur de leurs attentes.

Mais les remerciements s'entendent mieux encore pour les êtres humains bien réels qui ont tant compté durant ces trois années :

Merci Pascal Touzet pour ta confiance et ton encadrement ! Je n'aurais pas pu rêver mieux comme guide. Merci pour ta bienveillance !

Merci Etienne Meyer, Céline Poux, Camille Roux, Fabienne Van Rossum, Jean-Stéphane Varré et Xavier Vekemans pour votre participation et votre aide sur mes différents chapitres de thèse, qui en ont été grandement améliorés.

Merci à Marc Lensink & Théo Mauri d'avoir permis cette approche interdisciplinaire sur l'analyse des ribosomes.

Thank you, Dan Sloan, for welcoming me in your lab for two months. What a great experience! And thanks for collaborating with us on the chapter 3.

Thanks also to the members of my committee for your advices and expertise : Tatiana Giraud, Sophie Karrenberg, Camille Roux, Fabienne Van Rossum. Thanks to the members of my PhD jury for agreeing to review and discuss this work : Sophie Karrenberg, Alexander Papadopoulos, Jackie Shykoff, Fabienne Van Rossum, Xavier Vekemans.

Merci à Sophie Galina pour m'avoir guidé dans les méandres de la bio-info que je ne connaissais pas du tout ! Merci aussi au reste de l'équipe Bio-Info du labo, Mathieu Genète, Clément Mazoyer pour votre aide. Merci à Cécile Godé pour m'avoir appuyé sur les manipulations de bio-moléculaire ! Merci à toute l'équipe de Bio-Mol, Christelle, Laurence, Anne-Cat pour votre bonne humeur ! Merci aussi Eric Schmitt et l'équipe de la serre pour votre travail d'entretien des collections de *Silene* et les bons moments rempotages!

Merci aussi aux autres membres du labo, qui ont rendu ces trois ans de thèse si stimulants intellectuellement ! Merci à l'équipe de doctorant.e.s / post-doctorant.e.s qui ont siégé au labo pendant mes 3 années ici, avec qui j'ai rigolé, chanté, dansé (et même escaladé !). Merci à Chloé Beaumont et Magshoud qui ont supporté de partager avec moi leur bureau pendant 2 ans ! Merci aussi aux stagiaires que j'ai eu la chance d'encadrer et qui m'ont bien aidée.

Et puis, sans ma famille, proche et lointaine, mais aussi sans mes amis et amies, je n'aurais pas pu aller au bout de ce travail de doctorat.

Merci à mes parents de m'avoir toujours soutenue, de m'avoir rassurée et de-stressée. Merci à mes deux sœurs (et à mes beaux-frères) – Luna et Noémie – pour leur présence, leur écoute, leur joie... pour les rires et la musique... et merci aussi à Noémie pour les magnifiques illustrations des en-têtes de chapitre, où l'art vient à l'appui de la science.

Merci à mes grands-parents, celles qui sont là - Françoise et Mireille (Amatxi et Mamour !) - ceux qui sont partis - Francis et Jean - et à tout le reste de ma grande et belle famille, oncles et tantes, cousins et cousines pour vos questions, votre intérêt quant à mes travaux de recherches, mais aussi votre capacité à, parfois, souvent, - et très efficacement – m'en distraire !

Merci les copains et les copines... Merci Jue, Fab, Louis, Astrid, Mitch, Tibse et à tous-tes les autres pour ces moments de détente, de rire et de défoulement, de solidarité, d'écoute, de soutien.

Enfin, last but not least (car j'ai toujours préféré garder le meilleur pour la fin), merci à mon tout, mon âme, mon roc :

merci Nicolas Lee



# Table of contents

## INTRODUCTION

<b>Speciation, Organellar genomes and Cytonuclear incompatibilities</b> .....	1
<b>I. Speciation and evolution of reproductive barriers</b> .....	3
<b>II. Cytonuclear genetic incompatibilities in plant speciation</b> .....	11
<b>1. Introduction</b> .....	11
<b>2. How can cytonuclear incompatibilities be involved in speciation?</b> .....	12
<i>2.1. Coevolution between nuclear and organellar genomes</i> .....	12
2.1.1. <u>Detection of coevolution</u> .....	13
<i>2.2. Cytonuclear incompatibilities</i> .....	16
2.2.1. <u>CNIs are the result of disrupted coadaptation between organellar and nuclear genes</u> .....	16
2.2.2. <u>Acceleration of organellar genome evolution enhances the propensity of CNIs</u> .....	18
2.2.3. <u>CNIs as the result of intergenomic conflict – the CMS case</u> .....	18
<i>2.3. Cytonuclear coevolution and environment</i> .....	20
<i>2.4. CNIs can contribute to speciation</i> .....	22
<b>3. Can pattern of organellar inheritance influence CNIs?</b> .....	23
3.1. <i>Inheritance pattern of the organellar genomes</i> .....	23
3.2. <i>Heteroplasmy: evidences and consequences</i> .....	24
3.3. <i>Paternal leakage rescues from CNIs</i> .....	25
<b>4. How could mating systems favor or limit CNIs?</b> .....	26
<b>5. Glossary</b> .....	29
<b>6. References</b> .....	30
<b>III. Project and problematic</b> .....	38
<b>1. Chapter 1</b> .....	39
<b>2. Chapter 2</b> .....	41
<b>3. Chapter 3</b> .....	41
<b>4. Chapter 4</b> .....	42
<b>5. References</b> .....	42

## CHAPTER 1

<b>Reproductive isolation among lineages of <i>Silene nutans</i> (Caryophyllaceae): a potential involvement of plastid-nuclear incompatibilities</b> .....	45
<b>1. Introduction</b> .....	50

<b>2. Material and methods</b> .....	51
2.1. <i>Plastid genomic data</i> .....	51
2.2. <i>Nuclear data</i> .....	52
2.3. <i>Variant detection</i> .....	54
2.4. <i>Conservation status analysis</i> .....	55
2.5. <i>Analysis of the pattern of selection</i> .....	55
2.6. <i>Identification of contact positions in plastid complexes</i> .....	56
<b>3. Results</b> .....	57
3.1. <i>Genomic and transcriptomic data</i> .....	57
3.2. <i>Plastid genome diversity in <i>Silene nutans</i></i> .....	57
3.3. <i>Analysis of nuclear genes encoding subunits of plastid complexes</i> .....	60
3.4. <i>Patterns of selection and coevolution</i> .....	61
3.5. <i>Toward the identification of plastid-nuclear gene pairs candidates for PNI</i> .....	63
<b>4. Discussion and conclusion</b> .....	65
4.1. <i>Evolution of the plastid genome in <i>Silene nutans</i></i> .....	65
4.2. <i>Plastid-nuclear coevolution in <i>Silene nutans</i> lineages</i> .....	66
4.3. <i>Are cytochrome b6/f and plastid ribosome possibly involved in PNI?</i> .....	67
4.4. <i>A link with gynodioecy?</i> .....	69
<b>5. References</b> .....	70
<b>6. Annexes</b> .....	78

## SUB-CHAPTER 1

<b>What is the potential impact of genetic divergence of ribosomal genes between <i>Silene nutans</i> lineages in hybrids? An <i>in-silico</i> approach</b> .....	93
<b>1. Introduction</b> .....	97
<b>2. Material and methods</b> .....	99
2.1. <i>Identification of mutations</i> .....	99
2.2. <i>Identification of impacted interactions between subunits</i> .....	99
2.3. <i>Identification of the type of interaction and potential modifications</i> .....	99
2.4. <i>Creation of the different models</i> .....	100
2.5. <i>Creation of the Residue Interaction Networks (RIN) from the models and centrality analyses</i> .....	100
2.6. <i>Principal component analysis (PCA) on centrality measures</i> .....	100
<b>3. Results and discussion</b> .....	102
3.1. <i>Modification of the interactions due to the identified mutations</i> .....	102
3.2. <i>RINs analysis of mutations on the <i>rps11-rps21</i> gene pair</i> .....	102
3.3. <i>Principal Component Analysis (PCA)</i> .....	104
<b>4. Conclusion</b> .....	107

5. References .....	108
6. Annexes .....	112

## CHAPTER 2

<b>Paternal leakage of plastid rescues inter-lineage hybrids in <i>Silene nutans</i></b> .....	115
1. Introduction .....	120
2. Material and methods .....	121
2.1. <i>Silene nutans</i> .....	121
2.2. Plant material .....	122
2.3. DNA extraction and plastid genotyping .....	124
3. Results .....	124
4. Discussion .....	127
4.1. Could paternal leakage slow down the speciation process? .....	127
4.2. Breeding system and paternal leakage .....	129
5. References .....	130
6. Annexes .....	134

## CHAPTER 3

<b>The decoupled evolution of the organellar genomes of <i>Silene nutans</i> leads to distinct roles in the speciation process</b> .....	136
1. Introduction .....	141
2. Material and methods .....	142
2.1. Genomic data acquisition and assemblies .....	142
2.2. Variant detection for both mt and pt data .....	143
2.3. Alignment construction analysis .....	143
2.4. McDonald Kreitman tests .....	144
2.5. Recombination .....	145
2.6. Divergence ( <i>Ks</i> ) with <i>Silene latifolia</i> .....	145
2.7. Checking for numts .....	145
3. Results .....	146
3.1. Data quality .....	146
3.2. Polymorphism .....	147
3.3. Selection .....	147
3.4. Recombination .....	149
3.5. Divergence .....	149
3.6. Are mt sequences contaminated by numts? .....	149
4. Discussion .....	152

4.1. <i>Mt and pt genomes differ in their pattern of diversity</i> .....	152
4.2. <i>Mt genetic diversity link to gynodioecy</i> .....	153
4.3. <i>The decoupled evolutionary pathways of the pt and mt genomes in <i>Silene nutans</i></i> .....	154
<b>5. References</b> .....	155
<b>6. Annexes</b> .....	161

## CHAPTER 4

<b>Complete isolation without gene flow between four lineages of <i>Silene nutans</i></b> .....	172
<b>1. Introduction</b> .....	176
<b>2. Material and methods</b> .....	177
2.1. <i>Transcriptomic data</i> .....	177
2.2. <i>Demographic inference using DILS ABC framework</i> .....	179
2.2.1. <u>Model comparison: 2 populations</u> .....	179
2.2.2. <u>Model comparison: 4 populations</u> .....	180
2.3. <i>Statistical analysis</i> .....	182
<b>3. Results</b> .....	183
3.1. <i>Population genetic statistics</i> .....	183
3.2. <i>Evo-demographic Model selection</i> .....	185
3.3. <i>Parameters estimation</i> .....	187
<b>4. Discussion</b> .....	189
4.1. <i>Allopatric speciation for all lineages of <i>S. nutans</i></i> .....	189
4.2. <i>Rapid speciation between lineages of <i>S. nutans</i></i> .....	191
4.3. <i>Outliers of differentiation unlikely represent speciation islands</i> .....	193
<b>5. References</b> .....	195
<b>6. Annexes</b> .....	199

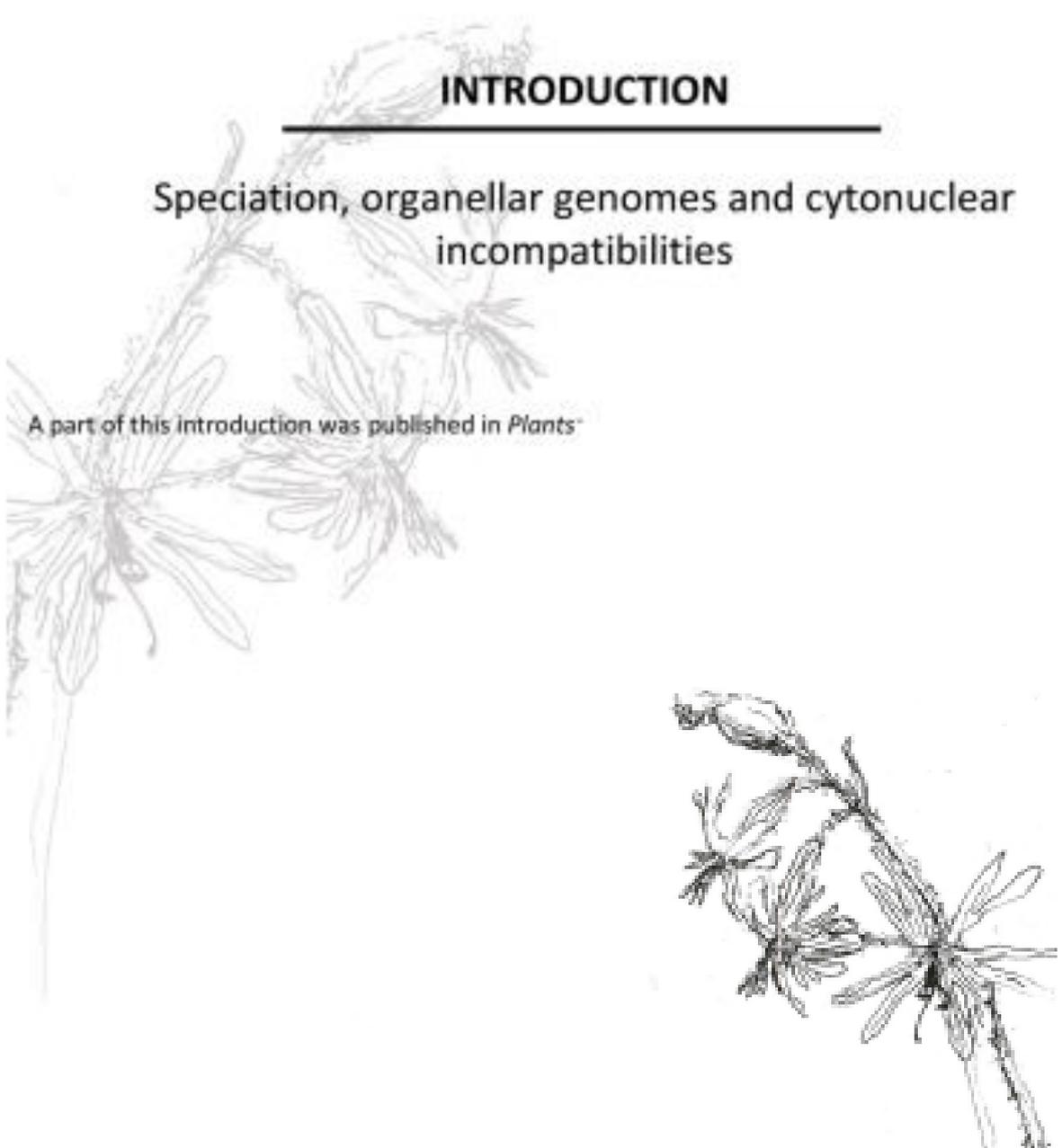
## DISCUSSION

<b>Opening on speciation and organellar genomes evolution</b> .....	208
<b>I. Summary of the results</b> .....	210
<b>II. Perspectives</b> .....	211
<b>1. Digging further into the analyses of the PNIs</b> .....	212
1.1. <i>Outstanding issues with the PNIs in <i>Silene nutans</i></i> .....	212
1.1.1. <u>The missing plastid genes</u> .....	212
1.1.2. <u>Strong, numerous or minor incompatibilities?</u> .....	214
1.1.3. <u>Dominance relationship at nuclear loci</u> .....	215
1.1.4. <u>Paternal leakage of the plastid genome: selection or not? That is the question</u> .....	216

1.2. Methodologies perspectives to answer the above questions.....	217
1.2.1. <u>Which hybrids to work with?</u> .....	217
1.2.2. <u>Sugar hybrids rescuing</u> .....	218
1.2.3. <u>Genetic mapping on backcross</u> .....	218
1.2.4. <u>Long-read sequencing and plastid genome assemblies</u> .....	219
2. The elevated rate of plastid genome evolution in <i>S. nutans</i> .....	219
3. What about the mito-nuclear coevolution in <i>S. nutans</i> ? .....	221
III. Evolutionary processes of the organellar genomes .....	223
1. Inheritance pattern of organellar genomes.....	223
1.2. <i>Uniparental inheritance: is it still the rule?</i> .....	223
1.3. <i>Benefits of biparental transmission</i> .....	224
1.4. <i>Keeping only one organellar type</i> .....	224
1.5. <i>Maternal exclusion: sometimes the father is the best</i> .....	227
1.6. <i>Why do some species group have the potential and other don't?</i> .....	228
2. Pace of evolution of the organellar genomes .....	229
2.1. <i>Organellar genome acceleration of evolutionary rate</i> .....	229
2.2. <i>What factors drove gene-specific acceleration?</i> .....	231
2.3. <i>Acceleration=speciation?</i> .....	233
IV. References .....	236







## INTRODUCTION

---

### Speciation, organellar genomes and cytonuclear incompatibilities

A part of this introduction was published in *Plants*<sup>1</sup>

<sup>1</sup> Univ. Lille, CNRS, UMR 8198 - Evo-Eco-Paleo, F-59000 Lille, France

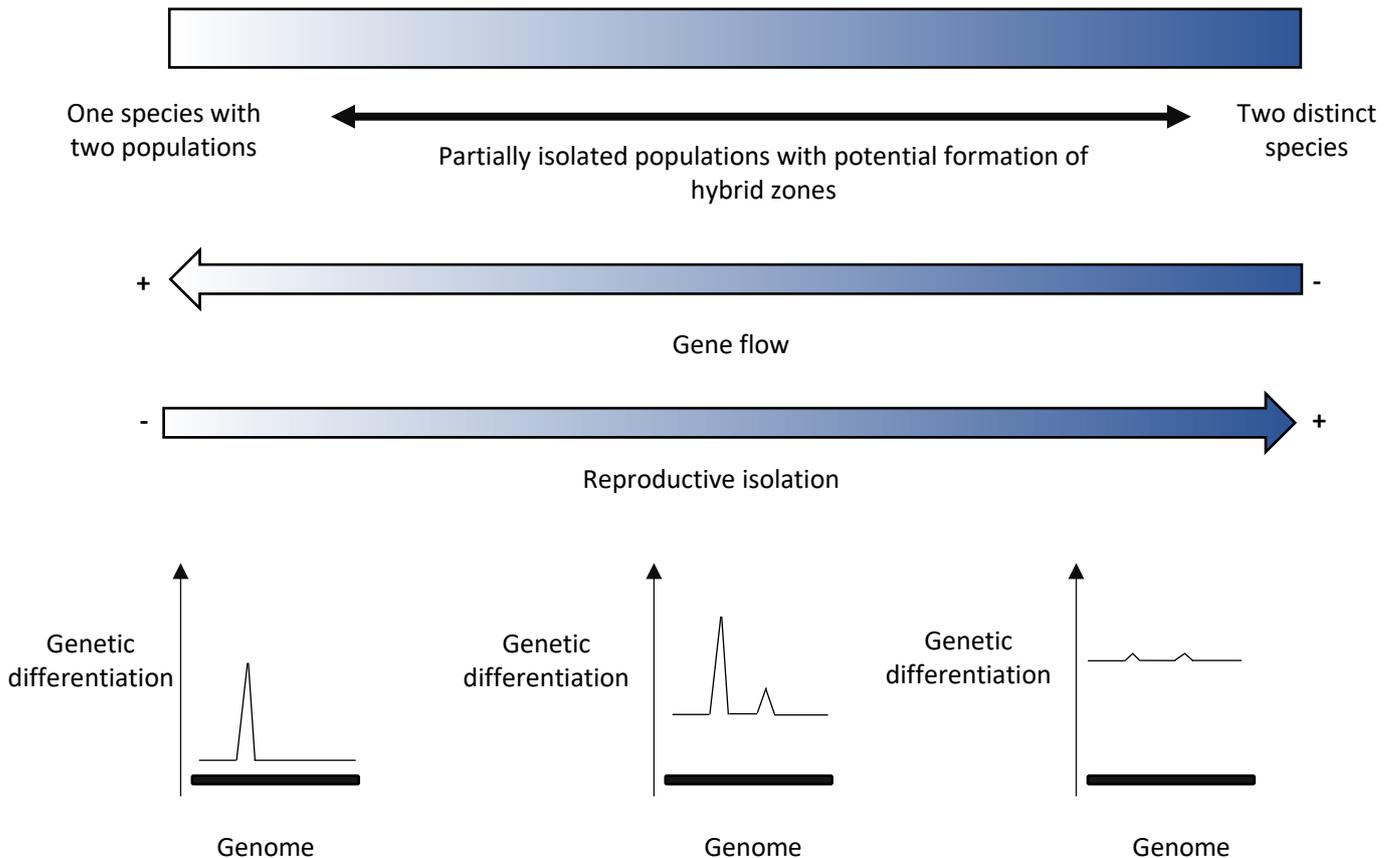
<sup>1</sup> Postel Z., Touzet P. (2020) Cytonuclear Genetic Incompatibilities in Plant Speciation. *Plants*, 9, 287



## I. Speciation and evolution of reproductive barriers

Speciation can be defined as the process that leads to the separation of populations into distinct evolutionary units: the species. But how do we define a species? Several criteria have been used over time, from the morphological criteria to the currently widely accepted biological concept of species (BSC) (Mallet, 2013). The biological concept of species currently states that species are defined as interbreeding groups of populations, regardless of their geographical isolation (Butlin & Stankowski, 2020). Reproductive isolation (RI) can emerge between populations of a species through the emergence of barriers preventing reproduction between them (Coyne & Orr, 2004; Kulmuni *et al.*, 2020). Over time, RI between these populations may increase and ultimately result in two distinct species. Speciation takes time and is thought to be a continuous process and is commonly referred to as the “speciation continuum” (Stankowski & Ravinet, 2021) (Figure 1). This term reflects that various levels of RI, that can be observed *in natura*, with species being more or less reproductively isolated in a reversible fashion until completion of speciation is reached (Barton, 2020; Stankowski & Ravinet, 2021): pairs of populations can evolve from being considered as the same species to two distinct ones when RI is complete (Seehausen *et al.*, 2014) (Figure 1). Reversibility becomes increasingly harder as barriers shaping RI become more complex (Coughlan & Matute, 2020). Along the speciation continuum, hybrid zones may appear where diverging populations exchange gene flow (Abbott, 2017; Pickup *et al.*, 2019) (Figure 1). Although the BSC states that species must be reproductively isolated from one another to be defined as such, along the speciation continuum there is an intermediate where RI is still incomplete with some level of interbreeding between populations of these species: this is called the “grey zone” of the speciation continuum (Roux *et al.*, 2016). Speciation and RI are the result of various evolutionary forces that ultimately drive the divergence of populations that can lead to speciation (Ravinet *et al.*, 2017). Mutation and recombination are among them, inducing genetic changes in DNA sequences and generating genetic variability that can become fixed in populations (Presgraves, 2010; Baack *et al.*, 2015). Genetic variability induced by mutation would not drive evolution of RI alone. Presence or absence of migration between populations also shape the evolution of RI. Varying levels of gene flow between populations can (i) introduce genetic variability between and within populations and (ii) homogenize the genetic content of populations (Ravinet *et al.*, 2017). Conversely, absence of gene flow and its homogenizing effect may further increase genetic differentiation between populations (Feder *et al.*, 2012; Suarez-gonzalez *et al.*, 2018). Genetic drift and natural selection are the two other main forces influencing the outcome of these genetic changes: either because of loss of allelic variants through random genetic drift or because of their active elimination or fixation through selection (Baack *et al.*, 2015). The extent of selection and genetic drift and their subsequent influence on genetic variability and evolution of divergence between

## SPECIATION CONTINUUM



**Fig.1 Illustration of the speciation continuum – adapted from Stankowski and Ravinet 2021, Evolution.** Along the speciation continuum, RI evolves toward complete isolation. Gene flow decreases until complete cessation, when RI is complete and populations became two distinct species. Between these two extremes, populations are partially isolated and can continue to exchange through gene flow, potentially forming hybrid zones when populations are in parapatry. With increasing level of RI and decreasing level of gene flow, genetic differentiation between population evolves from being restricted to small parts of the genome to the whole genome.

populations, will vary depending on intrinsic (e.g. demographic events such as bottleneck reducing population effective size and increasing genetic drift) and extrinsic factors (e.g. environmental conditions that may favor or not part of the genetic variation introduced) (Ravinet *et al.*, 2017).

Several modes of speciation have been described (Coyne & Orr, 2004). When RI between populations occurs because of divergence of these populations due to extrinsic barriers to gene flow (i.e. geographical distance or obstacles to migration between populations), we speak of allopatric speciation. Allopatric populations can meet again, through secondary contact where gene flow may occur between populations previously geographically isolated (Coyne & Orr, 2004). Secondary contact is one of the two processes that may lead to creation of hybrid zone (Abbott, 2017). Hybrid zones also appear through intergradation, which occurs when populations are connected through gene flow, with overlap of geographical distribution, during sympatric speciation (Coyne & Orr, 2004; Abbott, 2017).

When populations have distributions close to each other, with occasional gene flow occurring, parapatric speciation is invoked (Coyne & Orr, 2004). Although this classification of modes of speciation may be debated (Butlin *et al.*, 2008), it is still widely used by evolutionary biologists but with a more relaxed way : we generally talk about ‘phase of allopatry’ rather than just ‘allopatric speciation’ (Butlin *et al.*, 2008). All of these different scenarios will leave distinct genomic footprints (Liu *et al.*, 2020). When population divergence arises in allopatry, evolutionary forces such as genetic drift and selection for local adaptation will act independently in each lineage and shape genetic divergence between them (Matute & Cooper, 2021). When speciation is sympatric, divergence between populations might come from locally adapted alleles and selection against migrant alleles (Rieseberg & Blackman, 2010; Ravinet *et al.*, 2017; Shang *et al.*, 2020).

When speciation is complete, genetic divergence between populations may be genome-wide but initially during RI evolution, it is often only restricted to small parts of the genome exhibiting peaks of high genetic differentiation (Burri, 2017; Wolf & Ellegren, 2017) (Figure 1). These peaks of differentiation might be the result of two different mechanisms: (i) either because of linked selection in genomic regions of low recombination, where  $N_e$  is reduced (Burri *et al.*, 2015; Burri, 2017) or (ii) because they contain loci involved in RI (the so-called barrier loci) , thus being genomic regions impermeable to gene flow (Ravinet *et al.*, 2017) (Figure 1). In the former case, evolution of high differentiation will be mainly driven by background selection on slightly deleterious alleles (Burri, 2017). In the latter case, these regions of high differentiation containing barrier loci are commonly called “speciation islands” (Turner *et al.*, 2005; Burri, 2017; Ravinet *et al.*, 2017). Distinguishing whether genomic regions of high divergence include barrier loci and the genomic localization of such barrier loci is still a challenge (Burri, 2017). This relates to a more general question asking whether genetic bases of RI are mainly driven by adaptive or neutrally processes (Suarez-gonzalez *et al.*, 2018; Schluter & Rieseberg, 2022).

Some answers might be found when looking at factors responsible for the evolution of RI. RI cannot evolve without the establishment of barriers to reproduction. These barriers can act before reproduction between individuals of two different populations (i.e. pre-zygotic) or after it (i.e. post-zygotic) (Seehausen *et al.*, 2014; Baack *et al.*, 2015). The pre-zygotic barriers can also be divided as pre-mating and post-mating (Lowry *et al.*, 2008). Evolution of reproductive barriers (RBs) can be caused by intense selection on a trait with a simple genetic control, local adaptation, genetic drift or genomic conflict (Crespi & Nosil, 2013; Seehausen *et al.*, 2014; Baack *et al.*, 2015). Most of the pre-zygotic barriers emerge as an adaptive response to natural or sexual selection (Seehausen *et al.*, 2014; Baack *et al.*, 2015). For example, in plants, flowering time or floral traits can diverge between populations of the same species as the result of adaptation to more or less warm/cold habitat or pollinators. This will ultimately lead to divergence between populations as they will not

**Table 1**

Classification and non-exhaustive examples of reproductive barriers in plants. Adapted from Baack et al, 2015.

<b>Timeline of reproductive barrier action</b>	<b>Example of reproductive barriers</b>	<b>Effect of the reproductive barrier</b>
<b>PRE-ZYGOTIC</b>	<i>Pre-pollination</i> Niche differentiation (ecogeographic isolation) Phenological isolation Pollinator specialization / behaviour Immigrant inviability Mating type	Prevent crossing between individuals of different species/diverging populations
	<i>Post-pollination</i> Pollen-stigma interactions Pollen competition	
<b>POST-ZYGOTIC</b>	Genetic incompatibilities (genomic conflict, BDMIs) polyploidy Chromosomal rearrangement (translocation, inversion) Differential silencing	Affect hybrid viability and fertility (F1 or subsequent hybrid generations)

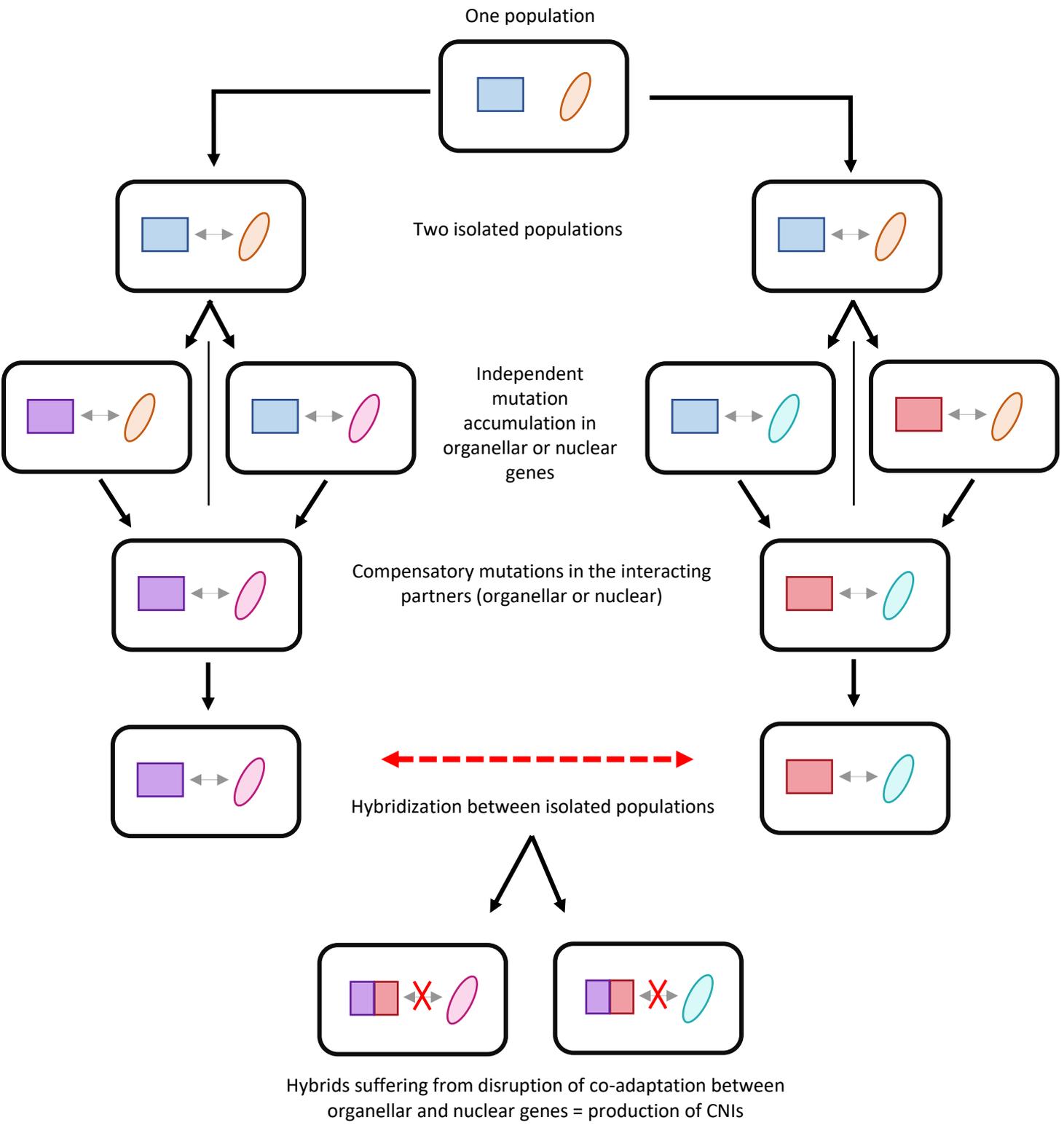
flower at the time or be pollinated by the same pollinators, being unable to mate with one another. Other examples of pre-zygotic barriers in plants are listed in Table 1. Post-zygotic barriers generally involve genetic incompatibilities in hybrids, which result in decreases of their viability and/or fertility because of negative interaction between divergent alleles (Presgraves, 2010; Baack *et al.*, 2015; Fishman & Sweigart, 2018). Post-zygotic RBs often evolve as a by-product of divergence between populations at two or more loci following the Bateson-Dobzhansky-Muller model of genetic incompatibility (BDMIs) (Coyne & Orr, 2004). There are other causes of post-zygotic RBs such as changes in ploidy, especially important in plants where it is common (Lowry *et al.*, 2008), structural genomic changes, meiosis defect due to sequence divergence, consequence of gene regulatory divergence (Coughlan & Matute, 2020) (Table 1). BDMI can involve both incompatibility between nuclear loci or between organellar and nuclear ones (Greiner *et al.*, 2011; Sloan *et al.*, 2017; Fishman & Sweigart, 2018). Three different stages of incompatibilities are expected: (i) the polymorphic stage which involves segregating alleles within species that can become fixed ; (ii) the simple stage which involves a few interacting alleles and (iii) the genetically complex stage where many interacting incompatible alleles are involved, when incompatible alleles continue to accumulate in a snowball

manner with potential for genetic redundancy (Coughlan & Matute, 2020). The more incompatible alleles are involved, the more genetic redundancy can prevent introgression and species collapse (Coughlan & Matute, 2020). There is still much to be done to disentangle what amount of the BDMIs are the results of mismatch between locally adapted alleles or neutrally derived ones. More generally, many questions are still open regarding the part of adaptation in driving evolution of RB and RI, i.e. how much of the genetic and phenotypic variation observed between diverging populations or different lineages is the result of selection and local adaptation (Schluter & Rieseberg, 2022).

Genetic basis of RI is of central importance in understanding the speciation process. For example, in the case of speciation with gene flow (i.e. sympatric), emergence of RB and evolution toward completion of speciation will require either strong divergent selection or coupling of multiple reproductive barriers, as if not, the homogenizing effect of gene flow will prevent divergence between populations (Coyne & Orr, 2004; Seehausen *et al.*, 2014; Baack *et al.*, 2015; Shang *et al.*, 2020). If RBs are controlled by a few loci of large effect or clustered loci of small effect, evolution of RBs can be very fast, even in the presence of gene flow as RBs either will have strong effect on fitness or represent a big island of speciation impermeable to gene flow (Stankowski & Ravinet, 2021). Speciation is thought to be particularly fast when involving ploidy changes or chromosomal rearrangement. More generally, accumulation of RBs is slow and the evolution of RI may not be monotonic (Nosil *et al.*, 2017; Kulmuni *et al.*, 2020).

All RBs should restrict introgression between diverging populations and RBs can vary in the rate at which they prevent it (Coughlan & Matute, 2020). Along RI evolution, the higher the barrier complexity is (i.e. the number of different RBs acting), the higher the overall decrease in gene flow will be through coupling of existing RBs (Kulmuni *et al.*, 2020; Shang *et al.*, 2020). Prevalence of pre-zygotic and post-zygotic barriers in speciation can depend on several factors and is still widely debated (Matute & Cooper, 2021). Pre-zygotic barriers are thought to be of main importance in plant speciation and the primary form of RI evolving during early stages (Lowry *et al.*, 2008; Baack *et al.*, 2015). Yet, strong intrinsic post-zygotic isolation is essential to the speciation process (Coughlan & Matute, 2020). Importance of post-zygotic RB and genetic incompatibilities to impede introgression theoretically depends on three aspects of the barrier loci involved : their genetic architecture (i.e. the number and genomic location of the incompatible loci, their dominance status and whether or not they are sex-linked), the genetic context they are evolving in (e.g. the amount of recombination) and the mechanism of evolution acting (i.e. is their evolution driven by selection/adaptation or is it neutral) (Coughlan & Matute, 2020). One type of BDMI has received some interest in particular these last years, with growing evidence of their importance in driving speciation as they might be one of the first post-zygotic barriers to evolve early in speciation (Greiner & Bock, 2013; Barnard-Kubow *et al.*, 2017) : cytonuclear incompatibilities (Figure 2).

 *Organellar*    
  *Nuclear genes*    
  *Population*    
  *Cytosuclear co-adaptation*



**Fig.2 Schematic representation of accumulation of BDMIs adapted for CNIs – inspired from Greiner et al (2011).** Mutations accumulate independently in isolated lineages, either in the nuclear or the organellar genes. This will impose selective pressure on the interacting partner (the nuclear or organellar one depending on which components initially mutated) for the fixation of compensatory mutations and maintenance of cytonuclear co-adaptation. If these isolated lineages reproduce, cytonuclear co-adaptation will be disrupted in the resulting hybrids leading to the formation of cytonuclear incompatibilities.

## References

- Abbott RJ. 2017.** Plant speciation across environmental gradients and the occurrence and nature of hybrid zones. *Journal of Systematics and Evolution* **55**: 238–258.
- Baack E, Melo MC, Rieseberg LH, Ortiz-Barrientos D. 2015.** The origins of reproductive isolation in plants. *New Phytologist* **207**: 968–984.
- Barnard-Kubow KB, McCoy MA, Galloway LF. 2017.** Biparental chloroplast inheritance leads to rescue from cytonuclear incompatibility. *New Phytologist* **213**: 1466–1476.
- Barton NH. 2020.** On the completion of speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences* **375**: 20190530.
- Burri R. 2017.** Interpreting differentiation landscapes in the light of long-term linked selection. *Evolution Letters* **1**: 118–131.
- Burri R, Nater A, Kawakami T, Mugal CF, Olason PI, Smeds L, Suh A, Dutoit L, Bure S, Garamszegi LZ, et al. 2015.** Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. *Genome Research* **25**: 1656–1665.
- Butlin RK, Galindo J, Grahame JW, Sheffield S. 2008.** Sympatric , parapatric or allopatric : the most important way to classify speciation ? *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**: 2997–3007.
- Butlin RK, Stankowski S. 2020.** Is it time to abandon the biological species concept ? No. *Molecular Biology & Genetics* **7**: 1400–1401.
- Coughlan JM, Matute DR. 2020.** The importance of intrinsic postzygotic barriers throughout the speciation process. *Philosophical Transactions of the Royal Society B: Biological Sciences* **375**: 20190533.
- Coyne JA, Orr HA. 2004.** *Speciation*. (Sinauer Associates, Ed.). Oxford University Press.
- Crespi B, Nosil P. 2013.** Conflictual speciation : species formation via genomic conflict. *Trends in Ecology & Evolution* **28**: 48–57.
- Feder JL, Egan SP, Nosil P. 2012.** The genomics of speciation-with- gene-flow. *Trends in Genetics* **28**: 342–350.
- Fishman L, Sweigart AL. 2018.** When Two Rights Make a Wrong : The Evolutionary Genetics of Plant Hybrid Incompatibilities. *Annual Review of Plant Biology* **69**: 707–731.
- Greiner S, Bock R. 2013.** Tuning a ménage à trois: Co-evolution and co-adaptation of nuclear and organellar genomes in plants. *BioEssays* **35**: 354–365.
- Greiner S, Rauwolf U, Meurer J, Herrmann RG. 2011.** The role of plastids in plant speciation. *Molecular Ecology* **20**: 671–691.

**Kulmuni J, Butlin RK, Lucek K, Savolainen V, Westram AM. 2020.** Towards the completion of speciation : the evolution of reproductive isolation beyond the first barriers. *Philosophical Transactions of the Royal Society B: Biological Sciences* **375**: 20190528.

**Liu X, Glémin S, Karrenberg S. 2020.** Evolution of putative barrier loci at an intermediate stage of speciation with gene flow in campions (*Silene*). *Molecular Ecology* **29**: 3511–3525.

**Lowry DB, Modliszewski JL, Wright KM, Wu CA, Willis JH. 2008.** The strength and genetic basis of reproductive isolating barriers in flowering plants. *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**: 3009–3021.

**Mallet J. 2013.** Species, Concepts of. *Encyclopedia of Biodiversity* **6**: 679–691.

**Matute DR, Cooper BS. 2021.** Comparative studies on speciation : 30 years since Coyne and Orr. *Evolution* **75**: 764–778.

**Nosil P, Feder JL, Flaxman SM, Gompert Z. 2017.** Tipping points in the dynamics of speciation. *Nature Ecology & Evolution* **1**: 0001.

**Pickup M, Brandvain Y, Fraïsse C, Yakimowski S, Barton NH, Dixit T, Lexer C, Cereghetti E, Field DL. 2019.** Mating system variation in hybrid zones: facilitation, barriers and asymmetries to gene flow. *New Phytologist* **224**: 1035–1047.

**Presgraves DC. 2010.** The molecular evolutionary basis of species formation. *Nature Reviews Genetics* **11**: 175–180.

**Ravinet M, Faria R, Butlin RK, Galindo J, Bierne N, Rafajlovic M, Noor MAF, Mehlig B, Westram AM. 2017.** Interpreting the genomic landscape of speciation : a road map for finding barriers to gene flow. *Journal of Evolutionary Biology* **30**: 1450–1477.

**Rieseberg LH, Blackman BK. 2010.** Speciation genes in plants. *Annals of Botany* **106**: 439–455.

**Roux C, Fraïsse C, Romiguier J, Anciaux Y, Galtier N, Bierne N. 2016.** Shedding Light on the Grey Zone of Speciation along a Continuum of Genomic Divergence. *PLoS Biology* **14**: 1–22.

**Schluter D, Rieseberg LH. 2022.** Three problems in the genetics of speciation by selection. *PNAS* **119**: e2122153119.

**Seehausen O, Butlin RK, Keller I, Wagner CE, Boughman JW, Hohenlohe PA, Peichel CL, Saetre G. 2014.** Genomics and the origin of species. *Nature* **15**: 176–192.

**Shang H, Hess J, Pickup M, Field DL, Ingvarsson PK, Liu J, Lexer C, Field DL. 2020.** Evolution of strong reproductive isolation in plants : broad-scale patterns and lessons from a perennial model group. *Philosophical Transactions of the Royal Society B: Biological Sciences* **375**: 20190544.

**Sloan DB, Havird JC, Sharbrough J. 2017.** The on-again, off-again relationship between mitochondrial genomes and species boundaries. *Molecular Ecology* **26**: 2212–2236.

**Stankowski S, Ravinet M. 2021.** Defining the speciation continuum. *Evolution* **75**: 1–18.

**Suarez-gonzalez A, Lexer C, Cronk QCB. 2018.** Adaptive introgression : a plant perspective.

Turner TL, Hahn MW, Nuzhdin S V. 2005. Genomic Islands of Speciation in *Anopheles gambiae*. *PLoS Biology* 3: e285.

Wolf JBW, Ellegren H. 2017. Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics* 18: 87–100.

## II. Cytonuclear Genetic Incompatibilities in Plant Speciation

This section was published in *Plants*<sup>2</sup>

### 1. Introduction

Speciation is the evolutionary process that leads to the differentiation of distinct species from an ancestral population. The delimitation of a species can be assessed by the level of barriers to reproduction to another species, i.e., the possibility to mate and produce viable and fertile hybrids. In plants, pre-zygotic barriers can involve differences in phenology, pollinator guild or habitat and pollen-stigma compatibility. Post-zygotic barriers expressed in hybrids can affect germination rate, survival or reproductive traits such as pollen quantity or quality and seed production [1–3]. The longer the species have diverged, the more they are expected to be reproductively isolated due to genetic differences, i.e., different fixed substitutions [4]. These mutations might have been directly selected by natural selection, for example if they confer a better adaptation to a given habitat, rendering the hybrid maladapted to both parental habitats. However, more generally, the reduced fitness of hybrids is believed to be due to genetic incompatibilities, i.e., mutations fixed independently by the species, that will interact negatively in the hybrid (Bateson–Dobzhansky–Muller incompatibilities—BDMIs) [5,6]. BDMIs can arise via different evolutionary processes. A given mutation might be neutral or nearly neutral in a given species, fixed by chance (genetic drift), but acting negatively on genes from the other species in the hybrid. If, on the contrary, this first mutation is not neutral and impacts, for instance, possible interactions with other genes, these interactions might be re-established through positive selection. A third possibility is that BDMIs result from a genetic conflict between a selfish element that distorts segregation to increase its own transmission, inducing an arms-race to counteract its effect [5,7,8].

In this review, we will assess how organellar genomes and their interaction with the nucleus can be involved in the process of speciation and RI. Interestingly, hybrids from reciprocal crosses often reveal an asymmetry in the level of RI. This phenomenon called Darwin’s corollary, points to the possible

---

<sup>2</sup> Postel Z., Touzet P. (2020) Cytonuclear Genetic Incompatibilities in Plant Speciation. *Plants*, 9, 287

cytonuclear origin of genetic incompatibilities (cytonuclear incompatibilities, CNIs) [9,10]. In addition, recent studies provide accumulative evidence of the frequent role of cytoplasmic genomes in adaptation and speciation in plants (reviewed in [9,11,12]).

## **2. How Can Cytonuclear Incompatibilities Be Involved in Speciation?**

### *2.1. Coevolution between Nuclear and Organellar Genomes*

Plastids and mitochondria are essential components of plant cells: mitochondria are responsible for cellular respiration and plastids have an essential role in photosynthesis and seed storage lipid synthesis [12]. The organellar protein complexes are encoded by nuclear and organellar genes, leading to genetic interactions between the nucleus and the organelles [12–15]. Such a pattern of cytonuclear (CN) interactions is the result of a long-term evolutionary history of organellar endosymbiotic origins [16]. Indeed, both organellar genomes originate from free-living bacteria, integrated into the host cell as endosymbiont, leading to the current eukaryotic cell as we know it [14,16,17]. These endosymbiotic events were followed by a massive functional gene transfer from the endosymbiont genome (i.e., plastidic and mitochondrial) to the host cell (i.e., the nucleus) and a subsequent gene loss of the redundant organellar function, leading to a reduction in size of organellar genomes [16,17]. Thus, organellar genomes do not encode the vast majority of the proteins they need for proper function (e.g., more than 90% of the plastid proteins are encoded by the nucleus [14]). Still, organellar genomes have conserved some key genes from their initial set of genes, encoding subunits (SUs) of enzyme complexes of essential eukaryotic pathways, such as respiration and photosynthesis [16–18]. However, for the proper cell performance, these organellar complexes also need gene products from nuclear genes, often derived from former organellar genes transferred to the nucleus [19]. Coordination between organellar and nuclear genomes is thus essential for eukaryotic cells [20] and represents a case of coevolution where organellar and nuclear genomes impose selection pressure on one another for the proper function of plant cells [12,17].

Differences exist between organellar and nuclear genomes. For example, due to the uniparental mode of inheritance resulting in a lack of recombination between parental genomes, organellar genomes have a reduced effective population size ( $N_e$ ) which reduces the efficacy of natural selection and increases the impact of genetic drift [17]. As an outcome, organellar genomes should be more prone than the nuclear genome to evolve under Muller's ratchet, i.e., the irreversible accumulation of deleterious mutations due to the lack of recombination [12]. Another contrasting characteristic is the mutation rate, which is much lower in the organellar genomes compared to

the nuclear one, at least in land plants [21]. Given the specific features of organellar genomes and the accumulation of deleterious mutations, it is expected that compensatory mutations on the nucleus may be positively selected to maintain coadaptation between interacting genomes and cellular performance [9,12,16,17]. However, the evolutionary dynamics of cytoplasmic genomes and the subsequent cytonuclear co-evolution might be more complex than expected. Analyses of mitochondrial and nuclear genomes in *Drosophila melanogaster* and *Homo sapiens* [22], mammal mitochondrial and proteobacteria genomes [23], suggest that despite their low  $N_e$ , mitochondrial genomes exhibit similar efficacy of purifying selection as nuclear ones. This could be explained by a stronger constraint on organellar complexes, due to the high number of protein-protein interactions and/or the additional layer of selection acting on mitochondrial genomes at the individual level, through the reduction of the number of mitochondrial genomes that occurs in the transmission from mother to offspring. A recent theoretical study shows that the way cytoplasmic genomes are transmitted not only slows down the accumulation of deleterious mutations, but also favors the fixation of beneficial ones [24].

#### Detection of Coevolution

Coevolution between two genes can be detected by looking at the encoded polypeptides. When mutual selective pressure exists between these genes, changes of amino acid in one polypeptide will lead to corresponding changes in the other one, resulting in correlated evolutionary rates [25]. Coevolution between genomes can be studied using different methods, based on phylogenies combined with the analysis of the rate of non-synonymous (NS) substitutions (reviewed in [26]). Under the hypothesis of coevolution, NS substitutions of interacting proteins should occur concurrently or sequentially in the phylogeny [15,25]. The generally low rate of evolution of the plastid genome makes it difficult to select a proper set of genes or taxa to study CN coevolution [25,27]. Nevertheless, in some independent lineages, accelerated plastid genome evolution has been detected, leading to faster CN coevolution [16,20,28–30]. For example, Geraniaceae exhibit accelerated rates of sequence evolution in plastid genes encoding SUs of the RNA polymerase complex [25]. Evolutionary rates of nuclear genes encoding SUs of this complex were thus estimated [25]. A correlation of the rates of non-synonymous substitutions ( $d_N$ ) was found between two nuclear genes and all the four plastid genes of this complex [25]. Conversely, no correlation was identified for the rate of synonymous substitution ( $d_S$ ), suggesting that this correlation is not due to the background mutation rate and very likely reflects coevolution signatures. Sloan et al. (2014) observed fast evolution of the nuclear genes encoding SUs of the organellar ribosomes in *Arabidopsis*, despite a low mutation rate of the organelle genomes [18]. This pattern was even more evident for *Silene* species with an accelerated rate of organellar genome evolution [18]. Interestingly, in these species,

nuclear-encoded cytosolic ribosomal proteins, as well as control nuclear genes not involved in organellar function, did not show any sign of accelerated evolution. This suggests that the accelerated rate of organellar genome evolution and the subsequent compensatory mutation in the nuclear compartment resulted in accelerated rates of sequence evolution of the nuclear genes involved in organellar protein complexes [18].

Evidence of coevolution can also be assessed by searching for signatures of positive selection on one of the interactors, i.e., the genes that are involved in the interaction. In response to mutations in the sequence of one interactor (generally, the organellar one), compensatory mutations will arise in the other (generally the nuclear one) and will be positively selected, as it allows the maintenance of the coadaptation and proper cell function. Thus, detecting positive selection in the sequence of nuclear genes encoding SUs of organellar protein complexes can be a sign of coevolution. The detection of positive selection is possible by analyzing the ratio  $d_N/d_S$ , which is used as a measure of selection (Box 1). The study of coevolution between plastid and nuclear genes was conducted on two plastid complexes (not involved in the photosynthetic process): Clp and ACCase. Clp complex is required for proper plastid function as it stabilizes the plastome (i.e., plastid genome), but its precise function remains unclear. ACCase is involved in fatty acid biosynthesis. Both complexes are composed of one plastid-encoded SU and several nuclear-encoded SUs [13,31]. Generally, the gene sequences of *accD* and *clpP1* (plastid genes) are relatively conserved among angiosperms but, in some independent lineages, evolution rates were found to be accelerated (quoted by [13]). Nearly all the nuclear encoded SUs of the Clp complex of the fast-evolving *Silene* species showed signatures of positive selection [13,20]. The greatest increase in substitution rate for the Clp complex was observed in the nuclear-encoded SU that is in direct interaction with the plastid SU. There was an enrichment of substitutions in the protein domain intimately interacting with this SU [13]. This pattern was less striking for ACCase, as *accD* does not evolve as fast as *clpP1* in the fast-evolving lineages. Thus, coevolution between nuclear-encoded and plastid-encoded SUs of Clp and ACCase complexes is suggested by signature of positive selection for compensatory mutations in *Silene* nuclear genes of these complexes [13]. Williams et al. (2019) conducted a study on the Clp complex across a broad range of angiosperms and observed correlated rates of accelerated evolution in the plastid-encoded and nuclear-encoded SUs [31]. Given these results, nuclear genes encoding SUs of plastid complex seem to experience positive selection to maintain coadaptation with their fast-evolving plastid genes.

A study on plastid ribosomes in Geraniaceae species found that signs of positive selection were detected on plastid genes, while the nuclear genes exhibited relaxed purifying selection [15]. Both nuclear and plastid genes exhibited accelerated  $d_N$  compared to the rest of the genes studied (i.e., nuclear genes encoding other functions, including cytosolic ribosomes), but no sign of

accelerated  $d_s$  suggesting again that increased substitution rate was independent from the background mutation rate [15]. Here, the coevolution pattern seems to work in the opposite way and is unlikely to be due to the compensatory evolution of the nuclear interacting genes. This highlights the diversity of cytonuclear coevolution patterns [15,32]. Coevolution signals in plants were mostly found in protein complexes involving plastid genes with accelerated evolution, such as genes encoding essential plastidic factors (*ycf1* and *ycf2* [29]), RNA polymerase SU, ribosomal proteins [18], *accD* [13] and *clpP1* [28,31]. Concerning the genes encoding components of the photosynthetic apparatus, this coevolution pattern was less evident, due to their high sequence conservation and low rate of sequence evolution [11–13,28].

Similar studies on the possible interaction between mitochondrial and nuclear genomes are scarce, due mainly to the low mutation rate of the mitochondrial genome in plants [21]. However, a study on Oxidative Phosphorylation (OXPHOS) complexes, for which SUs are encoded by both mitochondrial and nuclear genes, was conducted on *Silene* by taking advantage of the occurrence of species with fast-evolving mitochondrial genomes. These fast-evolving species exhibited faster sequence evolution for the nuclear genes encoding SUs of OXPHOS complexes, in response to higher mitochondrial mutation rates, indicating a strong coevolution signal between nuclear and mitochondrial genes of these complexes [33]. Positive selection was also detected in the nuclear genes encoding SUs of the OXPHOS complexes, while none was detected for the control genes (i.e., nuclear genes not targeted to mitochondria) [34]. In addition, the strength of this signal depended on the complex analyzed: the strongest compensatory signal was observed for the complexes with the highest proportion of mtDNA proteins (i.e., OXPHOS complexes III & IV) [33]. Mitochondrial mutation rate seems to play a role in determining the rate of sequence evolution of the nuclear interactors, imposing strong selection pressure for compensatory changes on nuclear-encoded proteins interacting with mitochondrial gene products [34]. This was further confirmed through the analysis of the sequence evolution of mitochondrial and nuclear OXPHOS genes of 84 eukaryotes with various mitochondrial mutation rates [35]. Interestingly, the structural analysis of the mito-nuclear pairs in *Silene* systems showed that the nuclear substitutions in fast-evolving *Silene* species were preferentially found at residues in direct contact with mutated mitochondrial residue, suggesting a structurally mediated coevolution [33]. Other studies identified a potential influence of protein residues contact on the coevolution pattern [13], but this influence was not always detected, like in the Geraniaceae studies [15,25]. Further analyses, including the structural modelling of organellar complexes, should be carried out to better understand the involvement of direct contact in shaping patterns of coevolution and rates of sequence evolution. CN coadaptation can also be revealed when crosses disrupt it, as in the case of F2 hybrids from *Arabidopsis*

*thaliana* crosses, where reduced germination capacity was observed for certain CN combinations [36].

## 2.2. Cytonuclear Incompatibilities

In isolated populations or different species, the divergence of gene sequences (in nuclear or organellar genomes) occurs through the independent accumulation of mutations. Coadaptation and coevolution are lineage-specific: each lineage will have nuclear and organellar combination with coadapted gene sequences. When hybridization occurs between these lineages, organelles are exchanged, breaking the intergenomic combinations and disrupting the coadaptation between genomes. These mismatches between genomes may lead to CN incompatibilities (CNIs) [16].

CNIs appear to be widespread among taxa and genera, between and within species [12, 37]. They often lead to altered phenotypes due to impaired organellar functions [38] and affect several fitness-related traits, such as viability and fertility of the hybrids [39]. These CNIs are often expressed directly in F1 hybrids [9,10,37] and often lead to asymmetrical decrease in fitness in reciprocal crosses, i.e., hybrids having the same hybrid nuclear background but different cytoplasms [40]. To further confirm CNI, backcrosses are often conducted (i) to partially restore or further disturb the parental CN combination and generate striking phenotypes [41,42], but also (ii) to disentangle the cytoplasm effect from the maternal one [43].

### 2.2.1 CNIs Are the Result of Disrupted Coadaptation between Organellar and Nuclear Genes

Several studies of hybrid incompatibilities and CNIs provide insight on the mechanisms at play. Most of them concern plastid-nuclear incompatibilities (PNIs), which have been frequently observed in flowering plants and often result in chlorosis/virescence or variegation [41], due to a decrease in photosynthetic function [37,44].

In *Pisum*, PNI was identified in hybrids from crosses between *P. sativum elatius* (wild type—VIR320) and cultivated peas lines [45,46]. Anomalies in leaf pigmentation and pollen inactivation, among others, resulted in a reduction of hybrid fitness [45]. This reduction was asymmetric: almost all of the F1 hybrids were sterile and displayed chlorophyll deficiency when VIR320 was the maternal parent, while hybrids had normal phenotypes when VIR320 was the paternal parent [45]. Bogdanova et al. (2009) identified two unlinked loci potentially involved in this PNI: *Scs1* and *Scs2* [46]. Thus, for *Pisum* species, PNI was likely due to plastid complex malfunction (i.e., hybrid bleaching) resulting from a disrupted coadaptation.

PNIs can also emerge from other mechanisms. Many of the organellar genes require RNA editing (RNAe) of their transcripts which usually concerns C to U conversion [47]. RNAe sites can be

species-specific [47,48] and the majority of the RNAe factors are pentatricopeptide repeat (PPR) proteins, encoded by nuclear genes and targeted to edit specific organellar genes [49]. Albino cybrids (i.e., 'artificial' hybrids generated by protoplast fusion and combining only one parental nuclear genome with only one plastid genome of another parent [44]) containing the nuclear genome of *Atropa belladonna* and the plastid genome of *Nicotiana tabacum* (i.e., Ab(Nt) cybrids) were reported [48]. Albinism seems to be due to a defect in RNAe of the transcript of *atpA* (plastid-gene encoding SU of the ATP synthase complex). In the Ab(Nt) cybrids, RNAe does not occur, due to the lack of specific RNAe factors in the *A. belladonna* nuclear genome, leading to impaired function of this complex [48]. CNI through editing disruption could also involve mitochondrial complexes, as editing is thought to be an important feature of the mitochondrial transcriptome [50].

Coevolution between organellar and nuclear genomes can also be disrupted in polyploids. Polyploidy results from nuclear whole genome doubling, leading to an increased number of nuclear genome copies. This phenomenon could disrupt CN interactions due to stoichiometry imbalance [51]. Allopolyploidization could make the maintenance of CN interaction even more challenging, as nuclear genome doubling results from hybridization. The allopolyploid individuals will thus contain biparental nuclear chromosomes and uniparental organellar genome inherited from different species, potentially resulting in CNI [51]. Several mechanisms are hypothesized to maintain coordination between organellar and nuclear genomes (reviewed in [51]): down-regulating the expression of the nuclear genes targeted to organelles (of both nuclear genomes or by the preferential expression of the organellar donor), or up-regulating the expression of organellar genes [52]. The study of allopolyploidization influence on CN interactions has mostly been studied on the Rubisco-encoding genes that revealed that maternal genes were preferentially expressed (reviewed in [51]). More recently, De Carvalho et al. (2019) studied the effect of recent and past allopolyploidization on the cytonuclear coadaptation in *Brassica*. In this study, no preferential transcription was identified. Rather, nuclear genes encoding SUs involved in plastid complexes were retained in duplicate or triplicate copies, while the genes encoding for cytosolic proteins were mostly found in single copy. Here the maintenance of CN coadaptation in allopolyploid *Brassica* seems due to the retention in multiple copies of the nuclear genes involved in plastid complexes, without any clear explanation of why the retention of multiple copies prevents CN maladaptation [14]. However, this study highlights the interest of conducting such approaches on a larger set of plant species, in order to have a better understanding of the consequence of polyploidization on CN interactions, polyploidization being an important feature of plant diversification [53].

### 2.2.2. Acceleration of Organellar Genome Evolution Enhances the Propensity of CNIs

Accelerated rate of plastid evolution increases the propensity for CNI as it leads to a faster coevolution of organellar and nuclear genomes within populations [20,54,55]. Indeed, as nuclear and plastid genomes are coadapted and in tight coevolution, an increased rate of nucleotide substitution accelerates the local coevolution of these two genomes [28]. The variation in evolution rates of organellar genomes results in asymmetries in CNI: CNI will be stronger when the organellar donor comes from the population with the highest relative rate of organelle evolution [10].

*Campanulastrum americanum* exhibits accelerated evolution of its plastid genome [28]. A former study identified a plastid gene of the small plastid ribosome SU as a good candidate for the generation of PNI, as it exhibited elevated levels of NS substitution and its interacting nuclear gene encoding a SU of the same complex showed evidence of compensatory substitutions [28]. Post-zygotic RI (RI) was observed between isolated populations, along with chlorotic hybrids (exhibiting chlorophyll deficiency), suggesting the presence of PNIs [28,41]. Further investigations of RI between populations of *C. americanum*, examining fitness-related traits in F1 hybrids from intra- and inter-cladescrosses and backcrosses, demonstrated a reduction in survival, germination and pollen viability for chlorotic plants [41]. RI was asymmetric and dependent on the direction of the crosses and backcrosses, for all the measured traits [41]. The plastid genetic distance seemed to determine the strength of the RI: even if the populations were in a narrow geographical area, hybrid fitness decreased more when hybrids originated from more genetically divergent populations [56]. The accelerated rate of plastome evolution, influencing plastid genetic distances between populations, drove the evolution and strength of PNI [28,41,55]. Male sterility was also observed among Mountain lineages of this species, with a probable cytoplasmic contribution [56]. This kind of cytoplasmic-induced male-sterility is a phenomenon known as cytoplasmic male-sterility (CMS) and is the result of a conflict between the nuclear and mitochondrial genomes [7,57,58].

### 2.2.3. CNIs as the Result of Intergenomic Conflict—the CMS Case

CMS is often the result of intergenomic conflict between the mitochondrial and the nuclear genomes [8,57,58]. Male-sterility mitochondrial genes will be selected as soon as they favor their own transmission via better seed production (their only way of transmission). This can be reached if the energy not devoted to pollen production is allocated towards seed production. Selection pressure will act on the nuclear genome for the emergence and fixation of the nuclear restorer of fertility (*Rf*) to re-establish the transmission through the pollen, by blocking the spread and expression of the CMS genes [59]. While this can lead to sexual polymorphism in populations, with the co-occurrence of hermaphrodites and females [60,61], CMS and *Rf* loci can also spread to fixation within a population,

leading to 'cryptic' CMS, all individuals being hermaphrodite [62,63]. In this case, CMS will only be revealed in crosses involving populations that do not contain the proper nuclear *Rf* locus [62,63]. Thus, in case of hybridization between populations/species, CNI can be the result of CMS-*Rf* systems disruption among hybrids containing mismatched CMS genes and nuclear *Rf* [12,64]. *Rf* genes are often part of the PPR gene family which encodes proteins targeted to organelles and involved in organellar biogenesis, transcription regulation and RNAe [65].

In flowering plants, the mitochondrial genome is very fluid in terms of structure as it varies in size, non-coding and repetitive DNA content, and structural rearrangement [66]. CMS often arises from these frequent structural rearrangements, which produce chimeric open-reading-frames (*orf*) co-transcribed with essential genes [67]. These ORFs will be directly toxic for pollen or lead to reduce mitochondrial function, compromising the high energy demands of pollen development [40]. Nuclear *Rf* acts by post-transcriptional processing of these chimeric *orf* transcripts or proteins [40]. As most of the flowering plants are hermaphrodite, male sterility seems to be rarely expressed, even if CMS genes are likely common (i.e., often observed in interspecific crosses) [40,65]. It results in various outcomes, ranging from complete failure to develop male floral organs to arrest of pollen development at different stages [68]. Several cases of cryptic CMS have been identified and in particular the well-studied case of the *Mimulus* species. In hybrids between species of this genus, one-fourth of F2 hybrids bearing *Mimulus guttatus* cytoplasm were male sterile, while F2 hybrids bearing the cytoplasm of *M. nasutus* were all male fertile [40]. Crosses revealed that male fertility could be restored by the dominant allele of *M. guttatus* at a single locus. Under the hypothesis that CMS genes are co-transcribed with essential genes and that *Rf* genes modify the sequence size and/or stability of mt transcripts associated with sterility, *nad6* was identified as a potential candidate [65]. Further analyses suggest that the *Rf* locus identified in *M. guttatus* is composed of two linked loci: *Rf1* and *Rf2* both occurring in a cluster of tandemly repeated PPR genes, both dominant in this species [69]. They could act in two ways: (i) cleave the transcript portion associated with the sterility and prevent its expression or (ii) alter the transcription of start sites to prevent the production of longer transcripts potentially containing the CMS genes [65]. The sterilizing cytoplasm is present in only one population of *M. guttatus*, suggesting that: (i) this CMS has a limited dispersion, (ii) this population is where the CMS arose (iii) the coevolution of CMS and *Rf* genes was local [65]. Consistent with this hypothesis, the two *Rf* loci identified exhibit several signs of localized selection (i.e., distinct haplotype structure, signs of selective sweep in the *Rf* region containing these loci . . . ) [62]. Mitochondrial CMS loci and their nuclear restorers seem to have coevolved, under strong positive selection and according to a model of conflictual coevolution, leading to a highly localized CMS-*Rf* system [62]. Another well-studied case of cryptic CMS concerns *Arabidopsis thaliana*, for which former studies identified CMS in crosses between two distant accessions: Sha and Mr-0 [70,71]. The

F1 hybrids plants containing Sha cytoplasm were sterile and unable to produce pollen, while F1s with the Mr-0 cytoplasm produced viable seeds [71]. The Sha lineages appeared to contain CMS factors and the nuclear *Rf* to restore male fertility [70].

CNI can arise through the disruption of coadaptation in major organellar complexes and intergenomic conflict between mitochondrial and nuclear genes. Much still needs to be done in order to understand the mechanisms of coevolution between genomes and which cellular functions are mainly affected by disruption of coadaptation. Databases on protein-protein but also RNA-protein or DNA-protein interactions could be useful to further investigate and identify nuclear and organellar interacting partners, coupled with the analysis of whole organellar genome diversity and functional validation [72]. The calculation of CN linkage disequilibrium (i.e., cnLD, representing the non-random association of organellar and nuclear alleles) could also help identify coadapted couples of nuclear and organellar genes [73].

### 2.3. Cytonuclear Coevolution and Environment

In animals, mitochondrial genomes can be involved in adaptation to climate or altitudinal variation [74]. A growing number of studies are focusing on the influence of cytoplasmic variation and its potential impact on environmental adaptation in plants (reviewed in [12,75]), and point out the potential role of organellar genomes in local adaptation.

Transplantation experiments were conducted with two *Helianthus* species, living in mesic (*H. annuus*) or xeric (*H. petiolaris*) habitats. They revealed that the cytoplasms of these species were adapted to each specific habitat, suggesting that variation in the organellar genome could contribute to local adaptation and ecological differentiation [76]. Potential cytoplasmic introgression driven by selection among *Helianthus* species was also reported, leading to widespread CN discordance in genealogies [77]. Selection might shape the pattern of organellar variation for some specific genes, resulting in adaptive introgression of the plastid genomes favoring local adaptation [77]. Cytoplasmic genomes can be important capacitors for the generation of novel phenotypes in specific environments, as shown in a study on *Arabidopsis thaliana* [42]. In addition, in this species, several adaptive traits seemed to be influenced and even shaped by CN interactions (around 80%) and organellar genome variation [38]. In *A. thaliana*, germination experiments on 64 cytotypes were conducted among populations from different geographical locations. A significant effect of the cytoplasm genotype on seed germination efficiency was identified, suggesting a role of the selection on the cytoplasmic geographical distribution and its role in *A. thaliana* populations adaptation [36]. Another study conducted on the same cytotypes concluded that cytonuclear interactions and coevolution had an impact on adaptive traits linked to seed vigor [78]. For example, the creation of novel CN combinations

in cytolines, leading to the disruption of CN coadaptation, had a deleterious effect on adaptive dormancy and germination, consistent with a contribution of CN co-adaptation and cytoplasmic variation on complex traits involved in plant adaptation [78].

Plastid genes encode components of essential functions in plant cells such as photosynthesis, a process that is influenced by many environmental factors and often associated with fitness differences [12,37,38]. The extreme degree of conservation of the genes encoding photosynthetic apparatus suggests that they evolve under genetic drift and/or strong purifying selection [11,12]. Yet, positive selection was identified for some specific genes and in particular lineages, pointing out a potential role of these genes on local adaptation [11,12,31]. For example, positive selection was found in the photosynthetic *rbcL* plastid gene, encoding a SU of the essential Rubisco enzyme [12,77]. Another study reported that the nuclear genes involved in photosystem I (PSI) had a very low absolute rate of substitution (as was expected given that this complex is slowly evolving), but also a significant excess of NS divergence between lineages. This demonstrates that some NS substitutions could be adaptive and spread under positive selection rather than fixation through genetic drift [13]. Positive selection for certain plastid genes was also discovered in the *Helianthus* species, suggesting the adaptive value of these genes [77]. In the peculiar case of anthropogenic environments, the plastid was directly involved in the adaptation to herbicide: a unique substitution in plastid *psbA* gene conferred resistance to Triazine in wild populations of *Arabidopsis thaliana* [79]. Selection could shape plastid sequence variation, depending on environmental factors and drive coevolution pattern between locally adapted plastid genes and the nuclear counterparts. These genes could also have adaptive value in relation to photosynthetic performance and colonization ability (see [76,79]).

If such a relation between organellar variation and local adaptation exists, then it could also drive the emergence of CNIs between locally adapted genes [76]. Indeed, specific environment can generate specific conditions leading to the selection for a particular mutation in the plastome and favor the compensatory mutation in the nuclear partner. This would lead to CN coadaptation driven by environmental conditions and the creation of CNIs among hybrids from populations coming from different environments [12]. In *C. americanum*, PNI could be due to local adaptation of one particular clade: reduced germination was observed in all crosses involving this clade, while there was none when it was absent from the crosses. This clade is geographically and environmentally distinct from the others and thus, its local adaptation could particularly contribute to disrupted CN coadaptation and PNI [41,56].

#### 2.4. CNI Can Contribute to Speciation

Several factors can impulse the generation of CNIs, such as an accelerated rate of organellar genome evolution or the involvement of organellar genomes in local adaptation [9,28,40]. As CNIs often result in reduced survival or germination and hybrid breakdown, it is very likely that both organellar genomes and CNI play a role in speciation through the development of barriers to gene flow between lineages [9,36,37]. CNIs are part of the genetic incompatibilities fitting the BDMI model [9,37,41] and are thought to be among the earliest genetic incompatibilities to arise in speciation, and play an important role in the emergence of post-zygotic barriers to reproduction [9,16,37,40,55,76]. Their contribution to such a process depends on several factors, such as the evolutionary history of the interaction (i.e., selfish cytoplasmic coevolution or neutral/adaptive coevolution), genetic characteristics of the parental species and the loci under consideration [32,40]. The strength of post-zygotic barriers due to CNIs also appears to be strongly associated with the degree of cytoplasmic divergence [9,37].

The possible involvement of CMS in post-zygotic RI is still unclear [5,7,12]. Due to its strong initial impact on individual fitness, CMS could represent a barrier to introgression, even more if the CMS-*Rf* system is geographically localized and when sterility is selected against in hybrids [40]. However, this effect will be short term, as the cost of male sterility might not be high enough to prevent the introgression of a CMS cytotype (conferring a better seed production to females) into a non-CMS population [62]. CMS participation in RI is even more unlikely if the matched nuclear restorer is also transmitted, as this system could increase introgression in secondary hybridizing populations [40]. Nevertheless, CMS-*Rf* systems can lead to a rapid genetic divergence between populations: a selective sweep due to CMS mutation will carry associated mitochondrial variation and impact mitochondrial divergence between populations, with consequences for organellar function. In the coadapted nuclear *Rf*, the responding sweep will alter the dynamic of linked nuclear variation and promote divergence between populations for relatively large regions of the nuclear genome [62]. Thus, local CMS-*Rf* dynamics is likely to generate species-wide incompatibilities and contribute to the establishing post-zygotic barriers during the early stage of speciation; even more so if its emergence is linked to specific ecological conditions [62,65]. Differences in mating system between the two hybridizing species and the fitness costs of the nuclear restorer alleles, depending on environment, could enhance the impact of this kind of CNI on the speciation process [62].

The nuclear component of CNIs could also play a role in speciation process. For example, the nuclear genes potentially involved in PNIs between *Pisum* exhibit high variability between lineages, even within the same geographical area [80]. This variability leads to different degrees of incompatibility and these genes could be viewed as 'speciation genes' [80], i.e., genes involved in RI as they contribute to the gene flow barrier between lineages [14].

The potential role of cytoplasmic variation and CN interactions in local adaptation could be an important driver of speciation, as it could maintain species-specific ecological differences [76]. This would generate associations between genes involved in ecological divergence and genes contributing to maladaptive CN interactions in hybrids. These associations would facilitate the origin and maintenance of species in presence of gene flow [76]. Ecological selection on a cytoplasm should limit its introgression and the introgression of the nuclear alleles interacting with it, maintaining the phenotype of the species despite hybridization [76].

Multiple genetic incompatibilities generally play a role in the RI for several traits, during the entire plant life cycle [41,81]. For example, in *C. americanum*, the number of incompatibilities and the strength of the RI is variable depending on the lineages crossed. The post-zygotic RI identified in *Mimulus* hybrids (*M. guttatus* × *M. nasutus*) seems to be controlled by different hybrid incompatibilities, also depending on the populations crossed [81]. Organellar genome pattern of inheritance could impact CN dynamics and the role of CNs in speciation. Indeed, biparental inheritance could be selected to avoid CNs and reduce the involvement of CNs in speciation.

### **3. Can Pattern of Organelle Inheritance Influence CNIs?**

#### *3.1. Inheritance Patterns of the Organellar Genomes*

Unlike the nuclear genome, inherited biparentally and following Mendelian segregation, the organellar genomes are mostly uniparentally inherited, usually maternally [19]. Uniparental inheritance can lead to severe evolutionary consequences (see Section 2.1), but also confer several evolutionary benefits. For example, it will favor the avoidance of deleterious interactions between co-existing organelle genomes potentially leading to disruption of CN coadaptation [12,19,37]. It will also limit within individual organellar diversity, since organellar genomes will go through a genetic bottleneck in the germline. This will limit the spread of selfish elements (mitochondrial) and lead to homoplasmy and the elimination of malfunctioning genotypes by selection [19,82]. The predominance of maternal inheritance remains largely unclear. It is likely due to the higher mutational load in the paternal gamete (the smaller one), more severe oxidative damages and more pronounced genetic drift, especially if the number of organellar DNA copies in the sperm cell is small. This mutational load will favor the evolution of gamete-controlled organelle exclusion mechanisms (“Killing one’s own cytoplasm”) and will result in maternal inheritance of organellar genomes [19].

Uniparental transmission of organelle genome has been repeatedly lost and restored over evolutionary timeframes. Mutational meltdown can be erased temporarily or over longer periods of sexual recombination between organelles through biparental transmission [19].

### 3.2. Heteroplasmy: Evidences and Consequences

Biparental transmission has arisen multiple times within the angiosperms and about 20% of the angiosperm species have the potential for biparental inheritance [55,64,83]. Accidental paternal transmission of organellar genomes (i.e., paternal leakage) can occur in crosses of divergent populations or species, due to a breakdown of mechanisms preventing biparental transmission [19,64,82]. In this case, paternal organellar DNA can endure different fates: it can be (i) destroyed within the sperm cell, (ii) physically excluded from the egg cell during fertilization or (iii) successfully replicated and maintained in the zygote after fertilization [39,64,82]. The latter case will lead to heteroplasmic individuals with two organellar haplotypes. Evolutionary consequences of heteroplasmy through paternal leakage will depend on the level of heteroplasmy (the ratio of paternal/maternal cytoplasms) at which the organellar genome has been paternally transmitted, but also the context in which it occurs. We can expect that, if the paternal organellar genome is in low frequency, it has a good chance of being lost during cell multiplication and organellar sorting out. Conversely, substantial levels of heteroplasmy may provide sufficient genetic variation for selection to act upon. In the case of mitochondrial genomes, it can even lead to genotypic novelty via recombination. In the case of paternal leakage of the plastid genome, the two plastid genomes will not fuse and thus will not undergo recombination. This will lead to competition between the two parental plastids. Their ability to compete against each other partially depends on the lipid composition of the plastid membrane, as it determines plastid stability and division rate [83].

In hybrids resulting from crosses between divergent population or species, biparental inheritance and the paternal leakage of plastids can result in leaf variegation. Hybrids will have variegated leaves or fully green/white ones, as a result of sorting-out the two plastid types during ontogenesis [83,84]. It indicates that one of the two parental plastids is unable to develop and undergo normal differentiation under the hybrid nuclear background, potentially because it is incompatible, while the other one is not [37,84–87], leading to chlorophyll-deficiency leaf sectors (i.e., white or yellowish sectors) [64,85].

Evidence of heteroplasmy and variegation exists in literature. For example, a study of Yao et al. (1995) identified that crosses between *Zantedeschia odorata* and two other species (*Z. elliottiana* and *Z. aethiopica*) resulted in hybrid variegation [88]. Heteroplasmy and hybrid variegation was also observed from crosses with *Pelargonium* species [84]. The analysis of the parental origin of the plastids in the green and white sectors of the variegated hybrids leaves revealed that in the green sectors, plastid DNA (ptDNA) of *Pelargonium zonale* 'Roseum' was present, while ptDNA of *P. zonale hort. 'Stadt Bern'* was present in the white tissue [84]. This suggests heteroplasmy and sorting out

of the plastid types, the 'Stadt Bern' one being incompatible with the hybrid nuclear background. Weihe et al. (2009) also analysed the inheritance pattern of both organellar genomes in the progeny of reciprocal crosses between *P. zonale* × *P. inquinans*. They observed biparental transmission of the two organellar genomes and hybrid plants exhibited variegation with the *P. inquinans* plastid bleaching out, potentially due to its incompatibility with the hybrid nuclear genome [85].

### 3.3. Paternal Leakage Rescues from Cytonuclear Incompatibilities

While hybridization between species can lead to biparental transmission of organelles and result in variegated hybrids, biparental transmission can also favor the restoration of compatibility between the plastid and nuclear genomes, leading to hybrid fitness recovery [64]. This rescuing could be due to several reasons. First, in crosses resulting in CNIs, occurrence of biparental inheritance could increase the likelihood that hybrids inherit an organellar genome compatible with the hybrid nuclear background. Second, as biparental transmission introduces genetic variation among organelles, selection can occur and could lead to the loss of the incompatible organellar genome [55]. Biparental transmission also seems to maintain the linkage disequilibrium between compatible cytoplasmic and nuclear genomes (cnLD) and thus favor the maintenance of adapted combinations between nuclear and organellar genomes [73]. Ramsey et al. (2019) showed that in gynodioecious species *Daucus carota*, hybrids from two distant populations exhibited lower fitness when parents were homoplasmic than when they were heteroplasmic [73]. Heteroplasmy could mitigate the negative effects of CNIs and maintain individual fitness, by providing cytoplasmic alternative allelic variants that potentially 'match' better with nuclear alleles [73].

Paternal leakage and heteroplasmy frequently occur in taxa that exhibit CNIs. Several studies showed that biparental inheritance could indeed increase the fitness of hybrids experiencing CNIs [55]. In crosses between *Pisum sativum* ssp. *elatius* wild species and cultivated peas forms, hybridization resulted in hybrid variegation. Paternal leakage occurred in crosses associated with PNI (i.e., chlorophyll deficiency): the fully green sectors of the variegated leaves contained paternally inherited plastids, suggesting that paternal plastid genome led to the recovery of normal photosynthetic performance [87–89]. Concerning the mitochondrial genome of pea, it appeared to be of maternal origin, indicating that paternal leakage was potentially driven by the presence of PNIs in hybrids [87].

Crosses in *C. americanum* revealed that biparental inheritance was constitutive in the species and did not depend on the level of genetic divergence between populations [55]. Biparental transmission led to an increased survival of F1 hybrids by enabling the selection against the incompatible plastid genome [55]. As F1 hybrids with heteroplasmy displayed better fitness, they

contributed to the majority of surviving F2 hybrids, which enforced the loss of the incompatible plastid. This phenomenon was stronger in crosses between divergent clades, suggesting that the level of CNI triggered the hybrid fitness recovery [55]. This phenomenon could lead to a counter-intuitive result: since CNI is stronger for hybrids from crosses between isolated populations, it could favour introgression and lead to the collapse of RI through heteroplasmy. Paternal leakage and heteroplasmy could thus slow down the speciation process, by rescuing from strong incompatibility, resulting in weaker reproductive barriers [55]. Given this, accelerated plastid genome evolution, which enhances the propensity for PNI, could also indirectly influence the level of paternal leakage [55]. Interestingly, several taxa exhibiting biparental inheritance also exhibit accelerated rates of plastid evolution [28]. CN interactions could regulate the paternal organellar DNA transmission and its selective replication in hybrids, in order to overcome CN dysfunction [39]. In a study of crosses between barley and wheat, Aksyonova et al. (2005) [39] reported the occurrence of a biparental transmission of organellar DNA and revealed a shift toward the paternal organelle DNA during repeated backcrosses. This shift indicated that paternal organellar DNA was transmitted through pollen and successfully replicated in the zygotes. Plastid transient heteroplasmy likely occurred as only paternal copies of wheat were detected in the stable self-fertile and vigorous lines obtained in the backcross generations. The increase in wheat paternal ptDNA content was correlated with fertility and restoration of vigor in these lines. Barley nuclear chromosomes were undetected and apparently replaced by wheat chromosomes. These results suggest that (i) the paternal wheat ptDNA was transmitted and selectively replicated, resulting in hybrid fitness recovery and (ii) that the wheat nuclear genome encoded for effective replication and retention of its corresponding ptDNA. Most likely, repeated backcrosses with wheat parents led to the replacement of the nuclear barley chromosomes and promoted the selective amplification of the paternal wheat organellar DNA copies [39].

#### **4. How Could Mating Systems Favor or Limit CNI?**

Mating systems are known to shape intra- and interspecific genetic variation [90], affect the efficacy of selection [91] and introduce selective forces in hybrid zones, where RI is incomplete [92]. The emergence of post-zygotic barriers to reproduction can be influenced by disruption of CN coadaptation. Studies are emerging about the impact of mating systems on the maintenance of coadaptation between interacting partners from different cellular compartments. Mating systems can influence the degree to which cytonuclear allele combinations are inherited [93]. For example, selfing rate increases the heritability of CN allele combinations. The higher the selfing rate, the more CN allele combinations will tend to be inherited as single units [93]. This will allow

direct selection of CN allele combinations, favoring the beneficial and eliminating the deleterious ones. Selfing will be more efficient than random mating or separate sex species [93]. Therefore, it is expected that crosses between distant populations of selfing species will disrupt such co-adapted CN units and thus generate strong CNIs. Studies on crosses in selfer *Arabidopsis thaliana* are in concordance with this expectation [36].

Transfer of organellar genes to the nuclear genome could also impact the level of CN coadaptation. Such transfers will not have the same probability of occurring, with regards to the mating system of the species. Selfing species will have an increased rate and probability of gene transfer compared to an outcrossing species, even more so if the transfer is adaptive [93,94]. In highly selfing species, the loss of the functional duplicated organellar copy will have no deleterious fitness consequences, as the nuclear functional gene copies will remain associated with the organellar genome. However, for outcrossing species, the organellar genome lacking the duplicate gene function will have a selective disadvantage compared to genomes with functional copies, since its function must be rescued by 'association' with a functional nuclear copy. In this context, as interspecies hybridization between an outcrosser and a selfer species will preferentially imply that the outcrosser is the pollen donor [94], it is expected that hybrids will exhibit CNIs, since the nucleus from the outcrosser will lack the gene(s) missing on the selfer organellar genome.

The presence of sex chromosomes may further reinforce genetic conflicts that might lead to segregation distortion [92]. Segregation distorters and their suppressors coevolve independently within a species and their interactions in hybrids could be a source of CNIs [92]. A shift in nuclear chromosomal distribution of the nuclear genes encoding gene products targeted to the organelles could also be observed, to maintain CN coadaptation (quoted in [95]). This chromosomal distribution is thought to be influenced by the inheritance pattern of organellar genes [95]. Male sex-chromosomes are often heterogametic (i.e., XY) and the female ones, homogametic (i.e., XX). Thus, the genes located on the X-chromosomes spend 2/3 time in females. Organellar DNAs are co-transmitted with 1/2 of the autosomal genes, 2/3 of the X-linked ones and none of the Y-linked ones [17]. As X-linked genes have a higher probability of co-transmission with organellar genes, selection for coadaptation could result in the overrepresentation of nuclear interacting genes on X chromosomes compared to the others [17,95]. In cases of sex-specific CN coadaptation disruption, this could induce chromosomal incompatibilities following Haldane's rule [96] (i.e., reduced fitness of the heterogametic sex in hybrids between species or divergent populations [10,97]). Another theory exists, named the sexual conflict hypothesis, for which the opposite chromosomal distribution is expected: more CN interactions involving autosomes to reduce the mutational load in males [98]. So far, in the only plant case studied (*Rumex hastatulus*), no evidence for over- or under- representation of genes interacting with organelles on X

chromosomes has been found [95]. This could be due to the relatively young age of its sex-chromosomes, leaving no time for selection to act on coadaptation or sexual antagonism [95]. The development of new protocols for sex-determination on non-model species should enable to tackle these hypotheses on a larger scale, with a gradient of sex chromosome age [99].

Last, CNIs could impact the evolution of mating systems. In particular, in certain conditions, CMS can generate gynodioecy (females and hermaphrodites in populations) that can be seen as a stable mating system over time through balancing selection, an evolutionary dynamic that maintains old cytoplasms. This could favor the accumulation of genetic incompatibilities on organellar genomes, as suggested in the gynodioecious *Silene nutans* which revealed cryptic speciation [100,101]. Interestingly, gynodioecy can also be a transitory step towards separate sexes (dioecy) [102,103].

### **Box 1. Acceleration pattern and selective forces acting on the organellar genomes**

Acceleration pattern is expressed as an increased rate of nucleotide substitution and an excess of non-synonymous (NS) substitutions compared to synonymous (S) ones, leading to an elevated  $dN/dS$  (annotated  $\omega$ , with  $dN$  as the NS substitution rate and  $dS$  as the S substitution rate). This acceleration can reveal either positive selection or relaxed purifying selection, due to the reduced  $N_e$  of the organellar genome [13,20,28].

Determination of selection forces generally relies on the estimation and analyses of  $\omega$ :  $\omega < 1$  indicates purifying selection,  $\omega=1$  neutral evolution,  $\omega>1$  indicates positive selection. It is an effective way to disentangle the effect of higher mutation rate or changes in selective pressure. S substitutions being considered as neutral,  $dS$  likely reflects the underlying mutation rate, while  $dN$  is impacted by the underlying mutation rate and selection [28]. Thus, changes in  $dN$  gives insight into changes in selection [28]. Other statistics exist for the detection of positive selection, such as the Tajima' D and the Fu's  $F_s$  which are useful to detect deviation from neutral variation or the McDonald and Kreitman test (MKT), which specifically tests for positive selection [77].

Distinguishing between positive selection and relaxed purifying selection is difficult, as they lead to the same pattern of increase of  $dN/dS$  ratio [31]. Looking at the variation of intra- and interspecific sequences can be a promising approach, since in the case of positive selection, the fixation of NS substitutions leads to an increase in  $\omega$  between species (i.e., interspecific divergence), compared to intraspecific polymorphism [18]. It must be noted that molecular tests for selection detection face several limits, for example limited power if the sequence variation is low and thus the inability to perform such tests on very closely related taxa [77]. Moreover, in the majority of these tests,  $dS$  is the reference against which selection is measured, but selection could also occur on S substitutions, leading to a potential contribution of these substitutions on differential fitness between individuals [77]. A molecular test can be used to detect relaxed selection on molecular sequence data: RELAX [105]. It has often been used in coevolution studies [15,31,34] and is based on a comparative phylogenetic framework, comparing gene-wide selection intensity across phylogenetic branches (see [105] for details).

**Author Contributions:** Z.P. and P.T. selected the bibliography and wrote the manuscript. Funding was obtained by PT. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Région Hauts-de-France, and the Ministère de l'Enseignement Supérieur et de la Recherche (CPER Climibio), and the European Fund for Regional Economic Development and the Région Hauts-de-France and the Ministère de l'Enseignement Supérieur et de la Recherche for ZP's PhD grant. The APC was funded by Univ. Lille and CNRS.

**Acknowledgments:** We acknowledge Eléonore Durand for comments on a previous version of the manuscript, and Licia Huffman-Touzet for English editing. **Conflicts of Interest:** The authors declare no conflict of interest.

## 5. Glossary

**Chlorosis :** loss of the green coloration of the plant leaves due to chlorophyll deficiency.

**Cytonuclear Linkage Disequilibrium (cnLD) :** represents the non-random association of nuclear and cytoplasmic alleles and is influenced by several factors, such as the inheritance pattern of organellar genome, hybridization, selective forces, mating system . . . Its evolutionary significance will depend on its magnitude and how stable CN interactions are [73,95].

**Effective size (Ne) :** the effective size of a population is a parameter that is used in population genetics to quantify the effect of genetic drift on the genetic diversity of a population. As the effective size decreases, the effect of genetic drift will be stronger and the efficacy of selection lower. This concept can be applied at the population level, but also at the genome or gene levels. For example, Ne is affected by the mode of inheritance: due to uniparental inheritance, Ne is lower for organellar genes than for autosomal ones which are transmitted biparentally [104].

**Functional gene transfer:** transfer of an organellar gene and its function to the nucleus. The new nuclear gene will be targeted back to the organelle and the redundant organellar gene will be eventually lost [94].

**Genetic drift:** Evolutionary process represented by the random sampling of alleles. It leads to changes in allele frequencies in a population over generation and influence population genetic diversity and divergence among populations.

**Heteroplasmy:** co-occurrence of two or more different organelle genotypes within an individual, which can be the result of biparental transmission.

**Intergenomic coevolution:** involves reciprocal effects of selection on interacting molecules from two genomes [16].

**Non-synonymous substitutions (NS) :** single-nucleotide mutations leading to changes of amino-acid in the encoded polypeptide.

**Paternal leakage:** occasional transmission of paternal organellar genome.

**Synonymous substitutions (S):** single-nucleotide mutations that do not result into changes of amino-acid in the encoded polypeptide.

**Variation:** presence of different colours, from green to white, on sectors on the same leaves or on different leaves of the same plant.

## 6. References

1. Sobel, J.M.; Chen, G.F.; Watt, L.R.; Schemske, D.W. The biology of speciation. *Evolution* **2010**, *64*, 295–315, doi:10.1111/j.1558-5646.2009.00877.x.
2. Lowry, D.B.; Modliszewski, J.L.; Wright, K.M.; Wu, C.A.; Willis, J.H. The strength and genetic basis of reproductive isolating barriers in flowering plants. *Philos. Trans. R. Soc. B Biol. Sci.* **2008**, *363*, 3009–3021, doi:10.1098/rstb.2008.0064.
3. Baack, E.; Melo, M.C.; Rieseberg, L.H.; Ortiz-Barrientos, D. The origins of reproductive isolation in plants. *New Phytol.* **2015**, *207*, 968–984.
4. Roux, C.; Fraïsse, C.; Romiguier, J.; Anciaux, Y.; Galtier, N.; Bierne, N. Shedding Light on the Grey Zone of Speciation along a Continuum of Genomic Divergence. *PLoS Biol.* **2016**, *14*, 1–22, doi:10.1371/journal.pbio.2000234.
5. Presgraves, D.C. The molecular evolutionary basis of species formation. *Nat. Rev. Genet.* **2010**, *11*, 175–180, doi:10.1038/nrg2718.
6. Fishman, L.; Sweigart, A.L. When Two Rights Make a Wrong : The Evolutionary Genetics of Plant Hybrid Incompatibilities. *Annu. Rev. Plant Biol.* **2018**, *69*, 707–731.
7. Crespi, B.; Nosil, P. Conflictual speciation : species formation via genomic conflict. *Trends Ecol. Evol.* **2013**, *28*, 48–57, doi:10.1016/j.tree.2012.08.015.
8. Havird, J.C.; Forsythe, E.S.; Williams, A.M.; Werren, J.H.; Dowling, D.K.; Sloan, D.B. Selfish Mitonuclear Conflict. *Curr. Biol.* **2019**, *29*, R496–R511, doi:10.1016/j.cub.2019.03.020.
9. Levin, D.A. The cytoplasmic factor in plant speciation. *Syst. Bot.* **2003**, *28*, 5–11, doi:10.1043/0363-6445-28.1.5.
10. Turelli, M.; Moyle, L.C. Asymmetric postmating isolation: Darwin’s corollary to Haldane’s rule. *Genetics* **2007**, *176*, 1059–1088, doi:10.1534/genetics.106.065979.
11. Budar, F.; Roux, F. The role of organelle genomes in plant adaptation: time to get to work! *Plant Signal. Behav.* **2011**, *6*, 635–639, doi:10.4161/psb.6.5.14524.
12. Greiner, S.; Bock, R. Tuning a ménage à trois: Co-evolution and co-adaptation of nuclear and organellar genomes in plants. *BioEssays* **2013**, *35*, 354–365, doi:10.1002/bies.201200137.
13. Rockenbach, K.; Havird, J.C.; Grey Monroe, J.; Triant, D.A.; Taylor, D.R.; Sloan, D.B. Positive

- selection in rapidly evolving plastid-nuclear enzyme complexes. *Genetics* **2016**, *204*, 1507–1522, doi:10.1534/genetics.116.188268.
14. Ferreira de Carvalho, J.; Lucas, J.; Deniot, G.; Falentin, C.; Filangi, O.; Gilet, M.; Legeai, F.; Lode, M.; Morice, J.; Trotoux, G.; et al. Cytonuclear interactions remain stable during allopolyploid evolution despite repeated whole-genome duplications in *Brassica*. *Plant J.* **2019**, *98*, 434–447, doi:10.1111/tpj.14228.
  15. Weng, M.L.; Ruhlman, T.A.; Jansen, R.K. Plastid–nuclear interaction and accelerated coevolution in plastid ribosomal genes in Geraniaceae. *Genome Biol. Evol.* **2016**, *8*, 1824–1838, doi:10.1093/gbe/evw115.
  16. Sloan, D.B.; Warren, J.M.; Williams, A.M.; Wu, Z.; Abdel-Ghany, S.E.; Chicco, A.J.; Havird, J.C. Cytonuclear integration and co-evolution. *Nat. Rev. Genet.* **2018**, *19*, 635–648, doi:10.1038/s41576-018-0035-9.
  17. Rand, D.M.; Haney, R.A.; Fry, A.J. Cytonuclear coevolution: The genomics of cooperation. *Trends Ecol. Evol.* **2004**, *19*, 645–653, doi:10.1016/j.tree.2004.10.003.
  18. Sloan, D.B.; Triant, D.A.; Wu, M.; Taylor, D.R. Cytonuclear interactions and relaxed selection accelerate sequence evolution in organelle ribosomes. *Mol. Biol. Evol.* **2014**, *31*, 673–682, doi:10.1093/molbev/mst259.
  19. Greiner, S.; Sobanski, J.; Bock, R. Why are most organelle genomes transmitted maternally? *BioEssays* **2014**, *37*, 80–94, doi:10.1002/bies.201400110.
  20. Sloan, D.B.; Triant, D.A.; Forrester, N.J.; Bergner, L.M.; Wu, M.; Taylor, D.R. A recurring syndrome of accelerated plastid genome evolution in the angiosperm tribe *Sileneae* (Caryophyllaceae). *Mol. Phylogenet. Evol.* **2014**, *72*, 82–89, doi:10.1016/j.ympev.2013.12.004.
  21. Smith, D.R. Mutation rates in plastid genomes: They are lower than you might think. *Genome Biol. Evol.* **2015**, *7*, 1227–1234, doi:10.1093/gbe/evv069.
  22. Cooper, B.S.; Burrus, C.R.; Ji, C.; Hahn, M.W.; Montooth, K.L. Similar Efficiencies of Selection Shape Mitochondrial and Nuclear Genes in Both *Drosophila melanogaster* and *Homo sapiens*. **2015**, *5*, 2165–2176, doi:10.1534/g3.114.016493.
  23. Mamirova, L.; Popadin, K.; Gelfand, M.S. Purifying selection in mitochondria , free-living and obligate intracellular proteobacteria. **2007**, *12*, 1–12, doi:10.1186/1471-2148-7-17.
  24. Christie, J.R.; Beekman, M. Uniparental inheritance promotes adaptive evolution in cytoplasmic genomes. *Mol. Biol. Evol.* **2016**, *34*, 677–691, doi:10.1093/molbev/msw266.
  25. Zhang, J.; Ruhlman, T.A.; Sabir, J.; Blazier, J.C.; Jansen, R.K. Coordinated rates of evolution between interacting plastid and nuclear genes in Geraniaceae. *Plant Cell* **2015**, *27*, 563–573, doi:10.1105/tpc.114.134353.
  26. De Juan, D.; Pazos, F.; Valencia, A. Emerging methods in protein co-evolution. *Nat. Rev. Genet.*

- 2013**, *14*, 249–261, doi:10.1038/nrg3414.
27. Zhang, J.; Ruhlman, T.A.; Sabir, J.S.M.; Blazier, J.C.; Weng, M.L.; Park, S.; Jansen, R.K. Coevolution between nuclear-encoded DNA replication, recombination, and repair genes and plastid genome complexity. *Genome Biol. Evol.* **2016**, *8*, 622–634, doi:10.1093/gbe/evw033.
  28. Barnard-Kubow, K.B.; Sloan, D.B.; Galloway, L.F. Correlation between sequence divergence and polymorphism reveals similar evolutionary mechanisms acting across multiple timescales in a rapidly evolving plastid genome. *BMC Evol. Biol.* **2014**, *14*, 1–10, doi:10.1186/s12862-014-0268-y.
  29. Sloan, D.B.; Alverson, A.J.; Wu, M.; Palmer, J.D.; Taylor, D.R. Recent acceleration of plastid sequence and structural evolution coincides with extreme mitochondrial divergence in the angiosperm genus *Silene*. *Genome Biol. Evol.* **2012**, *4*, 294–306, doi:10.1093/gbe/evs006.
  30. Jansen, R.K.; Cai, Z.; Raubeson, L.A.; Daniell, H.; Claude, W.; Leebens-mack, J.; Mu, K.F.; Guisinger-bellian, M.; Haberle, R.C.; Hansen, A.K.; et al. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. **2007**, *104*, 19369–19374.
  31. Williams, A.M.; Friso, G.; J. van Wijk, K.; Sloan, D.B. Extreme variation in rates of evolution in the plastid Clp protease complex. *Plant J.* **2019**, *98*, 243–259, doi:10.1111/tbj.14208.
  32. Sloan, D.B.; Havird, J.C.; Sharbrough, J. The on-again, off-again relationship between mitochondrial genomes and species boundaries. *Mol. Ecol.* **2017**, *26*, 2212–2236, doi:10.1111/mec.13959.
  33. Havird, J.C.; Whitehill, N.S.; Snow, C.D.; Sloan, D.B. Conservative and compensatory evolution in oxidative phosphorylation complexes of angiosperms with highly divergent rates of mitochondrial genome evolution. *Evolution* **2015**, *69*, 3069–3081, doi:10.1111/evo.12808.
  34. Havird, J.C.; Trapp, P.; Miller, C.M.; Bazos, I.; Sloan, D.B. Causes and Consequences of Rapidly Evolving mtDNA in a plant lineage. *Genome Biol. Evol.* **2017**, *9*, 323–336, doi:10.1093/gbe/evx010.
  35. Havird, J.C.; Sloan, D.B. The Roles of Mutation , Selection , and Expression in Determining Relative Rates of Evolution in Mitochondrial versus Nuclear Genomes. *Mol. Biol. Evol.* **2016**, *33*, 3042–3053, doi:10.1093/molbev/msw185.
  36. Moison, M.; Roux, F.; Quadrado, M.; Duval, R.; Ekovich, M.; Lê, D.H.; Verzaux, M.; Budar, F. Cytoplasmic phylogeny and evidence of cyto-nuclear co-adaptation in *Arabidopsis thaliana*. *Plant J.* **2010**, *63*, 728–738, doi:10.1111/j.1365-313X.2010.04275.x.
  37. Greiner, S.; Rauwolf, U.; Meurer, J.; Herrmann, R.G. The role of plastids in plant speciation. *Mol. Ecol.* **2011**, *20*, 671–691, doi:10.1111/j.1365-294X.2010.04984.x.
  38. Roux, F.; Mary-Huard, T.; Barillot, E.; Wenes, E.; Botran, L.; Durand, S.; Villoutreix, R.; Martin-

- Magniette, M.L.; Camillerif, C.; Budar, F. Cytonuclear interactions affect adaptive traits of the annual plant *Arabidopsis thaliana* in the field. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 3687–3692, doi:10.1073/pnas.1520687113.
39. Aksyonova, E.; Sinyavskaya, M.; Danilenko, N.; Pershina, L.; Nakamura, C.; Davydenko, O. Heteroplasmy and paternally oriented shift of the organellar DNA composition in barley-wheat hybrids during backcrosses with wheat parents. *Genome* **2005**, *48*, 761–769, doi:10.1139/G05-049.
40. Fishman, L.; Willis, J.H. Cytonuclear Incompatibility Causes Anther Sterility in *Mimulus* Hybrids. *Evolution* **2006**, *60*, 1372, doi:10.1554/05-708.1.
41. Barnard-Kubow, K.B.; So, N.; Galloway, L.F. Cytonuclear incompatibility contributes to the early stages of speciation. *Evolution* **2016**, *70*, 2752–2766, doi:10.1111/evo.13075.
42. Flood, P.J.; Theeuwen, T.P.J.M.; Schneeberger, K.; Keizer, P.; Kruijer, W.; Severing, E.; Kouklas, E.; Hageman, J.A.; Wijfjes, R.; Calvo-Baltanas, V.; et al. Reciprocal cybrids reveal how organellar genomes affect plant phenotypes. *Nat. Plants* **2020**, *6*, 13–21, doi:10.1038/s41477-019-0575-9.
43. Galloway, L.F.; Fenster, C.B. The Effect of Nuclear and Cytoplasmic Genes on Fitness and Local Adaptation in an Annual Legume, *Chamaecrista fasciculata*. *Evolution* **1999**, *53*, 1734, doi:10.2307/2640436.
44. Zubko, M.K.; Zubko, E.I.; Gleba, Y.Y. Self-fertile cybrids *Nicotiana tabacum* (+*Hyoscyamus aureus*) with a nucleo-plastome incompatibility. *Theor. Appl. Genet.* **2002**, *105*, 822–828, doi:10.1007/s00122-002-1037-7.
45. Bogdanova, V.S.; Berdnikov, V.A. Observation of a phenomenon resembling hybrid dysgenesis in a wild pea subspecies *Pisum sativum* ssp. *elatius*. *Pisum Genet.* **2001**, *33*, 5–9.
46. Bogdanova, V.S.; Galieva, E.R.; Kosterin, O.E. Genetic analysis of nuclear-cytoplasmic incompatibility in pea associated with cytoplasm of an accession of wild subspecies *Pisum sativum* subsp. *elatius* (Bieb.) Schmahl. *Theor. Appl. Genet.* **2009**, *118*, 801–809, doi:10.1007/s00122-008-0940-y.
47. Inada, M.; Sasaki, T.; Yukawa, M.; Tsudzuki, T.; Sugiura, M. A systematic search for RNA editing sites in pea chloroplasts: An editing event causes diversification from the evolutionarily conserved amino acid sequence. *Plant Cell Physiol.* **2004**, *45*, 1615–1622, doi:10.1093/pcp/pch191.
48. Schmitz-Linneweber, C.; Kushnir, S.; Babiyshuk, E.; Poltnigg, P.; Herrmann, R.G.; Maier, R.M. Pigment deficiency in nightshade/tobacco cybrids is caused by the failure to edit the plastid ATPase  $\alpha$ -subunit mRNA. *Plant Cell* **2005**, *17*, 1815–1828, doi:10.1105/tpc.105.032474.
49. Takenaka, M.; Zehrmann, A.; Verbitskiy, D.; Barbara, H.; Brennicke, A. RNA Editing in Plants and

- Its Evolution., doi:10.1146/annurev-genet-111212-133519.
50. Gray, M.W. RNA editing in plant mitochondria: 20 Years later. *IUBMB Life* **2009**, *61*, 1101–1104, doi:10.1002/iub.272.
  51. Sharbrough, J.; Conover, J.L.; Tate, J.A.; Wendel, J.F.; Sloan, D.B. Cytonuclear responses to genome doubling. *Am. J. Bot.* **2017**, *104*, 1277–1280, doi:10.3732/ajb.1700293.
  52. Coate, J.E.; Schreyer, W.M.; Kum, D. Robust Cytonuclear Coordination of Transcription in Nascent *Arabidopsis thaliana* Autopolyploids. **2020**.
  53. Peer, Y. Van De; Mizrahi, E.; Marchal, K. The evolutionary significance of polyploidy. *Nat. Publ. Gr.* **2017**, *18*, 411–424, doi:10.1038/nrg.2017.26.
  54. Burton, R.S.; Barreto, F.S. A disproportionate role for mtDNA in Dobzhansky-Muller incompatibilities? *Mol. Ecol.* **2012**, *21*, 4942–4957, doi:10.1111/mec.12006.
  55. Barnard-Kubow, K.B.; McCoy, M.A.; Galloway, L.F. Biparental chloroplast inheritance leads to rescue from cytonuclear incompatibility. *New Phytol.* **2017**, *213*, 1466–1476, doi:10.1111/nph.14222.
  56. Barnard-Kubow, K.B.; Galloway, L.F. Variation in reproductive isolation across a species range. *Ecol. Evol.* **2017**, *7*, 9347–9357, doi:10.1002/ece3.3400.
  57. Budar, F.; Touzet, P.; De Paepe, R. The nucleo-mitochondrial conflict in cytoplasmic male sterilities revisited. *Genetica* **2003**, *117*, 3–16, doi:10.1023/A.
  58. Cosmides, L.M.; Tobby, J. Cytoplasmic inheritance and intragenomic conflict. *J. Theor. Biol.* **1981**, *89*, 83–129.
  59. Delph, L.F.; Touzet, P.; Bailey, M.F. Merging theory and mechanism in studies of gynodioecy. *Trends Ecol. Evol.* **2007**, *22*, 17–24, doi:10.1016/j.tree.2006.09.013.
  60. Gouyon, P.H.; Vichot, F.; Van Damme, J.M.M. Nuclear-cytoplasmic male sterility: single-point equilibria versus limit cycles. *Am. Nat.* **1991**, *137*, 498–514, doi:10.1086/285179.
  61. Dufaÿ, M.; Touzet, P.; Maurice, S.; Cuguen, J. Modelling the maintenance of male-fertile cytoplasm in a gynodioecious population. *Heredity* **2007**, *99*, 349–356, doi:10.1038/sj.hdy.6801009.
  62. Case, A.L.; Finseth, F.R.; Barr, C.M.; Fishman, L. Selfish evolution of cytonuclear hybrid incompatibility in *Mimulus*. *Proc. R. Soc. B Biol. Sci.* **2016**, *283*, 20161493, doi:10.1098/rspb.2016.1493.
  63. Touzet, P. Mitochondrial Genome Evolution and Gynodioecy. *Adv. Bot. Res.* **2012**, *63*, 71–98, doi:10.1016/B978-0-12-394279-1.00004-1.
  64. Ramsey, A.J.; Mandel, J.R. When one genome is not enough: organellar heteroplasmy in plants. *Annu. Plant Rev.* **2019**, *2*, 1–40, doi:10.1002/9781119312994.apr0616.
  65. Case, A.L.; Willis, J.H. Hybrid male sterility in *Mimulus* (Phrymaceae) is associated with a

- geographically restricted mitochondrial rearrangement. *Evolution* **2008**, *62*, 1026–1039, doi:10.1111/j.1558-5646.2008.00360.x.
66. Gualberto, J.M.; Newton, K.J. Plant Mitochondrial Genomes: Dynamics and Mechanisms of Mutation. *Annu. Rev. Plant Biol.* **2017**, *68*, 225–252, doi:10.1146/annurev-arplant-043015-112232.
  67. Hanson, M.R.; Bentolila, S. Interactions of mitochondrial and nuclear genes that affect male gametophyte development. *Plant Cell* **2004**, *16*, doi:10.1105/tpc.015966.
  68. Chase, C.D. Cytoplasmic male sterility : a window to the world of plant mitochondrial – nuclear interactions. *Trends Genet.* **2007**, *23*, 81–90, doi:10.1016/j.tig.2006.12.004.
  69. Barr, C.M.; Fishman, L. The Nuclear Component of a Cytonuclear Hybrid Incompatibility in *Mimulus* Maps to a Cluster of Pentatricopeptide Repeat Gene. *Genetics* **2010**, *184*, 455–465, doi:10.1534/genetics.109.108175.
  70. Simon, M.; Durand, S.; Pluta, N.; Gobron, N.; Botran, L.; Ricou, A.; Camilleri, C.; Budar, F. Genomic conflicts that cause pollen mortality and raise reproductive barriers in *Arabidopsis thaliana*. *Genetics* **2016**, *203*, 1353–1367, doi:10.1534/genetics.115.183707.
  71. Gobron, N.; Waszczak, C.; Simon, M.; Hiard, S.; Boivin, S.; Charif, D.; Ducamp, A.; Wenes, E.; Budar, F. A Cryptic Cytoplasmic Male Sterility Unveils a Possible Gynodioecious Past for *Arabidopsis thaliana*. *PLoS One* **2013**, *8*, e62450, doi:10.1371/journal.pone.0062450.
  72. Burton, R.S.; Pereira, R.J.; Barreto, F.S. Cytonuclear Genomic Interactions and Hybrid Breakdown. *Annu. Rev. Ecol. Evol. Syst.* **2013**, *44*, 281–302, doi:10.1146/annurev-ecolsys-110512-135758.
  73. Ramsey, A.J.; McCauley, D.E.; Mandel, J.R. Heteroplasmy and Patterns of Cytonuclear Linkage Disequilibrium in Wild Carrot. *Integr. Comp. Biol.* **2019**, *59*, 1005–1015, doi:10.1093/icb/icz102.
  74. Hill, G.E. Mitonuclear coevolution as the genesis of speciation and the mitochondrial DNA barcode gap. *Ecol. Evol.* **2016**, *6*, 5831–5842, doi:10.1002/ece3.2338.
  75. Bock, D.G.; Andrew, R.L.; Rieseberg, L.H. On the adaptive value of cytoplasmic genomes in plants. *Mol. Ecol.* **2014**, *23*, 4899–4911, doi:10.1111/mec.12920.
  76. Sambatti, J.B.M.; Ortiz-Barrientos, D.; Baack, E.J.; Rieseberg, L.H. Ecological selection maintains cytonuclear incompatibilities in hybridizing sunflowers. *Ecol. Lett.* **2008**, *11*, 1082–1091, doi:10.1111/j.1461-0248.2008.01224.x.
  77. Lee-Yaw, J.A.; Grassa, C.J.; Joly, S.; Andrew, R.L.; Rieseberg, L.H. An evaluation of alternative explanations for widespread cytonuclear discordance in annual sunflowers (*Helianthus*). *New Phytol.* **2018**, *221*, 515–526, doi:10.1111/nph.15386.
  78. Boussardon, C.; Martin-Magniette, M.L.; Godin, B.; Benamar, A.; Vittrant, B.; Citerne, S.; Mary-Huard, T.; Macherel, D.; Rajjou, L.; Budar, F. Novel cytonuclear combinations modify

- Arabidopsis thaliana* seed physiology and vigor. *Front. Plant Sci.* **2019**, *10*, 1–15, doi:10.3389/fpls.2019.00032.
79. Flood, P.J.; Heerwaarden, J. van; Becker, F.; De Snoo, C.B.; Harbinson, J.; Aarts, M.G.M. Whole-Genome Hitchhiking on an Organelle Report Whole-Genome Hitchhiking on an Organelle Mutation. *Curr. Biol.* **2016**, *26*, 1306–1311, doi:10.1016/j.cub.2016.03.027.
  80. Bogdanova, V.S.; Kosterin, O.E.; Yadrikhinskiy, A.K. Wild peas vary in their cross-compatibility with cultivated pea (*Pisum sativum* subsp. *sativum* L.) depending on alleles of a nuclear–cytoplasmic incompatibility locus. *Theor. Appl. Genet.* **2014**, *127*, 1163–1172, doi:10.1007/s00122-014-2288-9.
  81. Martin, N.H.; Willis, J.H. Geographical variation in postzygotic isolation and its genetic basis within and between two *Mimulus* species. **2010**, 2469–2478, doi:10.1098/rstb.2010.0030.
  82. McCauley, D.E. Paternal leakage, heteroplasmy, and the evolution of plant mitochondrial genomes. *New Phytol.* **2013**, *200*, 966–977.
  83. Sobanski, J.; Giavalisco, P.; Fischer, A.; Kreiner, J.M.; Walther, D.; Schöttler, M.A.; Pellizzer, T.; Golczyk, H.; Obata, T.; Bock, R.; et al. Chloroplast competition is controlled by lipid biosynthesis in evening primroses. *PNAS* **2019**, *116*, 5665–5674, doi:10.1073/pnas.1811661116.
  84. Metzlaf, M.; Pohlheim, F.; Börner, T.; Hagemann, R. Hybrid variegation in the genus *Pelargonium*. *Curr. Genet.* **1982**, *5*, 245–249, doi:10.1007/BF00391813.
  85. Weihe, A.; Aplitz, J.; Pohlheim, F.; Salinas-Hartwig, A.; Börner, T. Biparental inheritance of plastidial and mitochondrial DNA and hybrid variegation in *Pelargonium*. *Mol. Genet. Genomics* **2009**, *282*, 587–593, doi:10.1007/s00438-009-0488-9.
  86. Sakamoto, W. Leaf-variegated mutations and their responsible genes in *Arabidopsis thaliana*. *Genes Genet. Syst.* **2003**, *78*, 1–9, doi:10.1266/ggs.78.1.
  87. Bogdanova, V.S. Inheritance of organelle DNA markers in a pea cross associated with nuclear–cytoplasmic incompatibility. *Theor. Appl. Genet.* **2007**, *114*, 333–339, doi:10.1007/s00122-006-0436-6.
  88. Yao, J.L.; Cohen, D.; Rowland, R.E. Interspecific albino and variegated hybrids in the genus *Zantedeschia*. *Plant Sci.* **1995**, *109*, 199–206, doi:10.1016/0168-9452(95)04163-O.
  89. Bogdanova, V.S.; Kosterin, O.E. A case of anomalous chloroplast inheritance in crosses of garden pea involving an accession of wild subspecies. *Dokl. Biol. Sci.* **2006**, *406*, 44–46, doi:10.1134/S0012496606010121.
  90. Charlesworth, D.; Wright, S.I. Breeding systems and genome evolution. **2001**, 685–690.
  91. Glémin, S.; Bazin, E.; Charlesworth, D. Impact of mating systems on patterns of sequence polymorphism in flowering plants. *Proc. R. Soc. B Biol. Sci.* **2006**, *273*, 3011–3019, doi:10.1098/rspb.2006.3657.

92. Pickup, M.; Brandvain, Y.; Fraïsse, C.; Yakimowski, S.; Barton, N.H.; Dixit, T.; Lexer, C.; Cereghetti, E.; Field, D.L. Mating system variation in hybrid zones: facilitation, barriers and asymmetries to gene flow. *New Phytol.* **2019**, *224*, 1035–1047, doi:10.1111/nph.16180.
93. Wade, M.J.; Goodnight, C.J. Cyto-Nuclear Epistasis: Two-Locus Random Genetic Drift in Hermaphroditic and Dioecious Species. *Evolution* **2006**, *60*, 643–659, doi:10.1111/j.0014-3820.2006.tb01146.x.
94. Brandvain, Y.; Wade, M.J. The functional transfer of genes from the mitochondria to the nucleus: The effects of selection, mutation, population size and rate of self-fertilization. *Genetics* **2009**, *182*, 1129–1139, doi:10.1534/genetics.108.100024.
95. Hough, J.; Ågren, J.A.; Barrett, S.C.H.; Wright, S.I. Chromosomal distribution of cytonuclear genes in a dioecious plant with sex chromosomes. *Genome Biol. Evol.* **2014**, *6*, 2439–2443, doi:10.1093/gbe/evu197.
96. Delph, L.F.; Demuth, J.P. Haldane’s rule: Genetic bases and their empirical support. *J. Hered.* **2016**, *107*, 383–391, doi:10.1093/jhered/esw026.
97. Brothers, A.N.; Delph, L.F. Haldane’s rule is extended to plants with sex chromosomes. *Evolution* **2010**, *64*, 3643–3648, doi:10.1111/j.1558-5646.2010.01095.x.
98. Drown, D.M.; Preuss, K.M.; Wade, M.J. Evidence of a Paucity of Genes That Interact with the Mitochondrion on the X in Mammals. *Genome Biol. Evol.* **2012**, *4*, 875–880, doi:10.1093/gbe/evs064.
99. Muyle, A.; Käfer, J.; Zemp, N.; Mousset, S.; Picard, F.; Marais, G.A.B. Sex-detector: A probabilistic approach to study sex chromosomes in non-model organisms. *Genome Biol. Evol.* **2016**, *8*, 2530–2543, doi:10.1093/gbe/evw172.
100. Martin, H.; Touzet, P.; Rossum, F. Van; Delalande, D.; Arnaud, J. Phylogeographic pattern of range expansion provides evidence for cryptic species lineages in *Silene nutans* in Western Europe. *Heredity* **2016**, *116*, 286–294, doi:10.1038/hdy.2015.100.
101. Martin, H.; Touzet, P.; Dufay, M.; Godé, C.; Schmitt, E.; Lahiani, E.; Delph, L.F.; Van Rossum, F. Lineages of *Silene nutans* developed rapid, strong, asymmetric postzygotic reproductive isolation in allopatry. *Evolution* **2017**, *71*, 1519–1531, doi:10.1111/evo.13245.
102. Maurice, S.; Belhassen, E.; Couvet, D.; Gouyon, P.H. Evolution of dioecy: Can nuclear-cytoplasmic interactions select for maleness? *Heredity* **1994**, *73*, 346–354, doi:10.1038/hdy.1994.181.
103. Barrett, S.C.H. The evolution of plant sexual diversity. *Nat. Rev. Genet.* **2002**, *3*, 274–284, doi:10.1038/nrg776.
104. Charlesworth, B. Effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* **2009**, *10*, 195–205, doi:10.1038/nrg2526.

105. Wertheim, J.O.; Murrell, B.; Smith, M.D.; Pond, S.L.K.; Scheffler, K. RELAX : Detecting Relaxed Selection in a Phylogenetic Framework. *Mol. Biol. Evol.* **2015**, *32*, 820–832, doi:10.1093/molbev/msu400.

### III. Project & problematic

*Silene nutans* (Caryophyllaceae) is a gynodioecious species, i.e. natural populations are composed of hermaphrodite and female individuals. This reproductive system relies on the presence of cytoplasmic male sterility factors (CMS) encoded in the mitochondrial genome, that produces male-sterile individuals (McCauley and Bailey, 2009; Touzet, 2012). Male fertility can be restored through the fixation of a nuclear-encoded restorer of fertility (*Rf*) (Budar et al., 2003; Chase, 2007). So far, two CMS factors and four *Rfs* were identified in *S. nutans*, with evidence for epistatic interactions between them (Garraud et al., 2011). Gynodioecy is thought to be maintained through two mechanisms: 1) balancing selection, which maintains mitochondrial polymorphism at CMS factors and *Rf* loci in populations, or 2) through episodic and regular invasion of CMS factors followed by *Rf*, via selective sweep (Delph et al., 2007). In *S. nutans*, a previous study analyzed the genetic diversity at some plastid and mitochondrial loci to assess which of these two mechanisms was maintaining gynodioecy in this species (Lahiani et al., 2013). This genetic diversity at some organellar loci was higher in *Silene nutans* compared to another gynodioecious species, *Silene vulgaris*, with evidence for balancing selection on the mitochondrial markers (Touzet and Delph, 2009; Lahiani et al., 2013).

Besides, *S. nutans* is composed of at least two strongly genetically differentiated lineages in France, based on analysis of plastid and nuclear markers, and associated with post-glacial refugia: an eastern one (E1) widespread in the north of Europe and a western one (W), composed of three sub-lineages: W1 distributed in England, France and Belgium ; W2 restricted in Spain and south-west France ; W3 in the Alps and Italy (Martin et al., 2016; Van Rossum et al., 2018). Lineages E1 and W1 also represent distinct ecotypes in Belgium (calcicolous and silicicolous) with edaphic specialization of the two lineages following their divergence in allopatry (De Bilde, 1973). At secondary contact zones between these two lineages, in southern England and southern Belgium, no hybridization events were detected suggesting the absence of gene flow between them (Martin et al., 2016). Some other contact zones exist, for example between W1 and W2 in south-eastern France and between W1 and W3 in the Alps, with apparently low degree of gene flow at these locations.

Strong and asymmetric post-zygotic reproductive isolation (RI) was observed between E1 and W1, expressed as a high proportion of chlorotic seedling and reduced hybrid fitness (Martin et al., 2017). This post-zygotic RI could be the result of the accumulation of genetic incompatibilities in allopatry prior to recolonization from the glacial refugia. A scenario of allopatric divergence was identified

between lineages E1 and W with strict isolation since their split around 300 000 years ago (Martin, 2016). Post-zygotic asymmetric RI was also observed when conducting diallelic cross between all four lineages, with a high proportion of chlorotic seedlings and mortality depending on the direction of the cross (Van Rossum et al, in prep) (Table 1). Concerning the existence of pre-zygotic isolation, studies have focused on the Belgian E1 and W1 ecotypes. Pre-zygotic isolation through pollinator isolation/specialization is unlikely between these lineages, as they share the same night pollinators and intra and inter ecotypic pollen flows are similar (Cornet et al., 2022). Pollen-stigma incompatibilities could also play a role as post-pollination pre-zygotic isolation between lineages E1 and W1 (Van Rossum et al., 1996) and these two lineages also exhibit differences in morphology and phenology that could play an additional role as pre-zygotic barrier to reproduction (De Bilde, 1973; Van Rossum, 2000).

## **1. Chapter 1**

Based on the previous results of post-zygotic RI between the four *S. nutans* lineages, expressed by high mortality of chlorotic seedlings (Table 1), we thought that cytonuclear incompatibilities might be involved. As mentioned above, these incompatibilities are the result of disruption of lineage-specific co-adaptation between nuclear and organellar genes, rendering organellar complexes nonfunctional and leading to asymmetric RI due to the uniparental mode of inheritance of organellar genomes (Levin, 2003; Turelli and Moyle, 2007; Greiner et al., 2011) (Table 1). Because the hybrids did not suffer from germination problems and chlorosis was observed at an early stage of development, we focused on plastid-nuclear (PNIs) rather than mito-nuclear incompatibilities. We attempted to identify pairs of plastid-nuclear genes that co-evolve within lineages and for which this co-adaptation might be disrupted in inter-lineage hybrids. To do so, we sought to identify signatures of plastid-nuclear co-evolution, such as the presence of differently fixed mutations in each lineage, signatures of positive selection as potential indications for fixation of compensatory mutations, the presence of mutations at residue contact positions between nuclear-encoded and plastid-encoded proteins. We used plastid genomic and nuclear transcriptomic data for several individuals from each of the four lineages to address this question.

## **2. Chapter 2**

RI between the four lineages is strong but incomplete as some hybrids survived depending on the cross type and direction (Table 1). These surviving hybrids exhibited either pale, fully green or variegated

**Table 1**

Summary results of the hybrid seedling phenotypes resulting from the diallele conducted by Fabienne Van Rossum in 2015 & 2018.

		PATERNAL PARENT			
		E	W1	W2	W3
MATERNAL PARENT	E	3,9%*	99,2%	99,8%	92,9%
	W1	78,2%	3,5%	48,2%	9,3%
	W2	99,9%	87,4%	0,1%	64,7%
	W3	58,9%	2,9%	0,5%	1,4%

Circles represent the proportion of chlorotic (white) / variegated (pale green) and fully green (green) hybrids ; The size of the circles is proportional to the number of each phenotype category ; \* : hybrids mortality percentage.

phenotypes. Such phenotypes can be the result of paternal leakage of the plastid genome (Weihe et al., 2009; Greiner et al., 2011). Indeed, even though organellar genomes are generally maternally transmitted, there are growing evidence for deviation from strict uniparental inheritance to biparental inheritance and/or occasional paternal leakage (i.e. episodic transmission of the organellar genome through the pollen) (Greiner et al., 2014; Ramsey and Mandel, 2019). When there is paternal leakage of the plastid genome in hybrids suffering of PNIs (such as it is the case here), variegation can be observed as a result of sorting out of the two transmitted plastid types (cf II/3.2). This suggests that one of the two types is highly incompatible with the hybrid nuclear background, resulting in chlorophyll deficiency sectors, while the other is not (i.e. resulting in fully green sectors) (Ramsey and Mandel, 2019; Postel and Touzet, 2020). Because of the presence of these fully green sectors, it has also been suggested that paternal leakage of the plastid genome could rescue inter-lineages hybrids suffering from PNIs (Barnard-Kubow et al., 2017).

In the second chapter of this thesis, we tested for the presence of the paternal plastid genome in surviving inter-lineage hybrids of *S. nutans*. We expected to observe the maternal plastid genome on the white leaves/sectors and the paternal one on the green sectors. To detect which of the two parental plastid types was present in these hybrids, we genotyped them with six plastid SNPs differentiating the four lineages.

### 3. Chapter 3

The plastid genome in this species seems to be involved in RI. Plastid and mitochondrial genomes are generally supposed to be co-transmitted, so tight linkage-disequilibrium is expected between them, with shared evolutionary forces shaping their genetic diversity (Olson and McCauley, 2000). Assuming co-transmission of the organellar genomes in *S. nutans*, if the plastid genome is involved in RI through PNIs in inter-lineages hybrids of *S. nutans*, then the mitochondrial genome could also be involved through mito-nuclear incompatibilities. In addition, as said before, this species is gynodioecious, which could imply particular evolutionary dynamics of the mitochondrial genome (Touzet, 2012). Balancing selection was previously identified as evolutionary acting on the mitochondrial genome in *S. nutans* (Lahiani et al., 2013). If so, it could prevent its involvement in RI between lineages through the maintenance of ancestral mitochondrial polymorphism and the absence of fixed genetic differences in mitochondrial genes between lineages.

In the third chapter of this PhD, we compared the evolutionary patterns of the organellar genomes in the four lineages of *S. nutans*, using gene capture data for both organellar genomes. We analyzed the different evolutionary forces that might shape their genetic diversity to assess whether mitochondrial genome was involved in the RI between lineages and how gynodioecy might impact it.

#### 4. Chapter 4

In the last chapter of this PhD thesis, we further examined the demographic history of the four lineages of *S. nutans*. As previously mentioned, allopatric speciation was identified between lineages E1 and all of the westerns ones, with a divergence time of around 300 000 years, consistent with the last glacial maximum (Martin, 2016). To go further, we sampled new individuals within the western lineages to have eleven individuals per lineage (i.e. equal number of individuals per lineages). To test for potential gene flow within the western lineages, we collected samples in close geographical proximity when it was possible. Using both the previously and newly collected RNAseq data, we infer the full evolutionary-demographic history of the four lineages of *S. nutans*. These inferences were done using DILS, a newly released program based on the ABC framework (Fraïsse et al., 2021).

#### 5. References

- Barnard-Kubow, K. B., McCoy, M. A., and Galloway, L. F. (2017). Biparental chloroplast inheritance – leads to rescue from cytonuclear incompatibility. *New Phytol.* 213, 1466–1476. doi: 10.1111/nph.14222.
- Budar, F., Touzet, P., and De Paepe, R. (2003). The nucleo-mitochondrial conflict in cytoplasmic male sterilities revisited. *Genetica* 117, 3–16. doi: 10.1023/A.
- Chase, C. D. (2007). Cytoplasmic male sterility : a window to the world of plant mitochondrial – nuclear interactions. *Trends Genet.* 23, 81–90. doi: 10.1016/j.tig.2006.12.004.
- Cornet, C., Noret, N., and Rossum, F. Van (2022). Pollinator sharing between reproductively isolated genetic lineages of *Silene nutans*. *Front. Plant Sci.* 13, 1–18. doi: 10.3389/fpls.2022.927498.
- De Bilde, J. (1973). Etude génécologique du *Silene nutans* L. en Belgique : Populations du *Silene nutans* L. sur substrats siliceux et calcaires. *Rev. Générale Bot.* 80, 161–176.
- Delph, L. F., Touzet, P., and Bailey, M. F. (2007). Merging theory and mechanism in studies of gynodioecy. *Trends Ecol. Evol.* 22, 17–24. doi: 10.1016/j.tree.2006.09.013.
- Fraïsse, C., Popovic, I., Mazoyer, C., Spataro, B., Delmotte, S., Romiguier, J., et al. (2021). DILS : Demographic inferences with linked selection by using ABC. *Mol. Ecol. Resources* 00, 1–16. doi: 10.1111/1755-0998.13323.
- Garraud, C., Brachi, B., Dufay, M., Touzet, P., and Shykoff, J. A. (2011). Genetic determination of male sterility in gynodioecious *Silene nutans*. *Heredity* 106, 757–764. doi: 10.1038/hdy.2010.116.
- Greiner, S., Rauwolf, U., Meurer, J., and Herrmann, R. G. (2011). The role of plastids in plant speciation. *Mol. Ecol.* 20, 671–691. doi: 10.1111/j.1365-294X.2010.04984.x.

- Greiner, S., Sobanski, J., and Bock, R. (2014). Why are most organelle genomes transmitted maternally? *BioEssays* 37, 80–94. doi: 10.1002/bies.201400110.
- Lahiani, E., Dufay, M., Castric, V., Le Cadre, S., Charlesworth, D., Van Rossum, F., et al. (2013). Disentangling the effects of mating systems and mutation rates on cytoplasmic diversity in gynodioecious *Silene nutans* and dioecious *Silene otites*. *Heredity* 111, 157–164. doi: 10.1038/hdy.2013.32.
- Levin, D. A. (2003). The cytoplasmic factor in plant speciation. *Syst. Bot.* 28, 5–11. doi: 10.1043/0363-6445-28.1.5.
- Martin, H. (2016). Processus de spéciation et impact des systèmes de reproduction dans le genre *Silene*.
- Martin, H., Touzet, P., Dufay, M., Godé, C., Schmitt, E., Lahiani, E., et al. (2017). Lineages of *Silene nutans* developed rapid, strong, asymmetric postzygotic reproductive isolation in allopatry. *Evolution* 71, 1519–1531. doi: 10.1111/evo.13245.
- Martin, H., Touzet, P., Rossum, F. Van, Delalande, D., and Arnaud, J. (2016). Phylogeographic pattern of range expansion provides evidence for cryptic species lineages in *Silene nutans* in Western Europe. *Heredity* 116, 286–294. doi: 10.1038/hdy.2015.100.
- McCauley, D. E., and Bailey, M. F. (2009). Recent advances in the study of gynodioecy : the interface of theory and empiricism. *Ann. Bot.* 104, 611–620. doi: 10.1093/aob/mcp141.
- Olson, M. S., and Mccauley, D. E. (2000). Linkage disequilibrium and phylogenetic congruence between chloroplast and mitochondrial haplotypes in *Silene vulgaris*. *Proc. R. Soc. London B Biol. Sci.* 267, 1801–1808. doi: 10.1098/rspb.2000.1213.
- Postel, Z., and Touzet, P. (2020). Cytonuclear Genetic Incompatibilities in Plant Speciation. *Plants* 9, 487.
- Ramsey, A. J., and Mandel, J. R. (2019). When one genome is not enough: organellar heteroplasmy in plants. *Annu. Plant Rev.* 2, 1–40. doi: 10.1002/9781119312994.apr0616.
- Touzet, P. (2012). Mitochondrial Genome Evolution and Gynodioecy. *Adv. Bot. Res.* 63, 71–98. doi: 10.1016/B978-0-12-394279-1.00004-1.
- Touzet, P., and Delph, L. F. (2009). The Effect of Breeding System on Polymorphism in Mitochondrial Genes of *Silene*. *Genetics* 644, 631–644. doi: 10.1534/genetics.108.092411.
- Turelli, M., and Moyle, L. C. (2007). Asymmetric postmating isolation: Darwin’s corollary to Haldane’s rule. *Genetics* 176, 1059–1088. doi: 10.1534/genetics.106.065979.
- Van Rossum, F. (2000). Amplitude synécologique de *Silene nutans* L. (Caryophyllaceae). *DUMORTIERA* 75, 11–24.
- Van Rossum, F., De Bilde, J., and Lefèbvre, C. (1996). Barriers to hybridization in calcicolous and silicolous populations of *Silene nutans* from Belgium. *R. Bot. Soc. Belgium* 129, 13–18.

Van Rossum, F., Martin, H., Le Cadre, S., Brachi, B., Christenhusz, M. J. M., and Touzet, P. (2018).

Phylogeography of a widely distributed species reveals a cryptic assemblage of distinct genetic lineages needing separate conservation strategies. *Perspect. Plant Ecol. Evol. Syst.* 35, 44–51.

doi: 10.1016/j.ppees.2018.10.003.

Weihe, A., Apitz, J., Pohlheim, F., Salinas-Hartwig, A., and Börner, T. (2009). Biparental inheritance of

plastidial and mitochondrial DNA and hybrid variegation in *Pelargonium*. *Mol. Genet. Genomics*

282, 587–593. doi: 10.1007/s00438-009-0488-9.





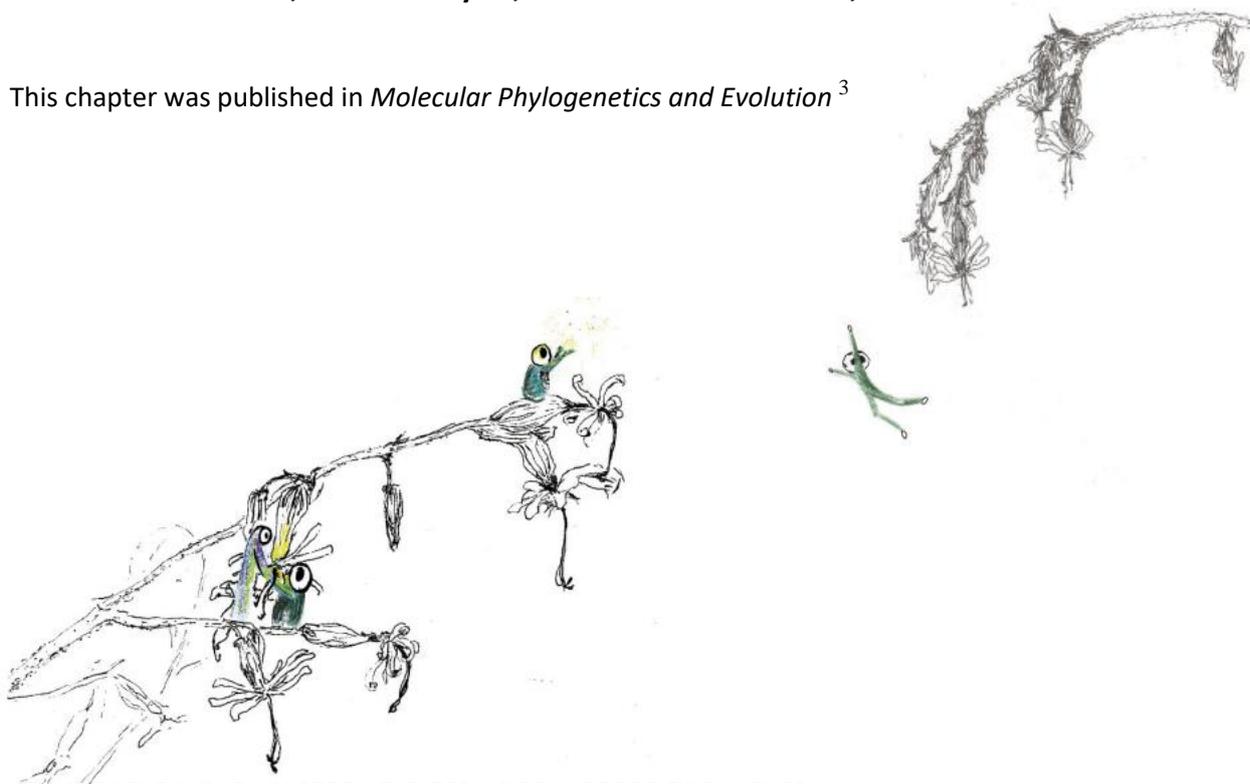
## CHAPTER 1

---

# Plastid-nuclear incompatibilities between lineages of *Silene nutans*

Zoé Postel<sup>1</sup>, Céline Poux<sup>1</sup>, Sophie Gallina<sup>1</sup>, Jean-Stéphane Varré<sup>2</sup>, Cécile Godé<sup>1</sup>, Eric Schmitt<sup>1</sup>, Etienne Meyer<sup>3</sup>, Fabienne Van Rossum<sup>3,4</sup>, Pascal Touzet<sup>1</sup>

This chapter was published in *Molecular Phylogenetics and Evolution*<sup>3</sup>



---

<sup>2</sup> Univ. Lille, Inria, UMR CNRS 9189 - CRISTAL Centre de Recherche en Informatique Signal et Automatique de Lille F-59000 Lille, France

<sup>3</sup> Institute of Plant Physiology, Martin-Luther-University, Halle-Wittenberg, Germany

<sup>4</sup> Meise Botanic Garden, Nieuwelaan 38, BE-1860 Meise, Belgium

<sup>5</sup> Service général de l'Enseignement supérieur et de la Recherche scientifique, Fédération Wallonie-Bruxelles, rue A. Lavallée 1, BE-1080 Brussels, Belgium

---

<sup>3</sup> Postel, Z., Poux C., Gallina S., Varré J. S., Godé C., Schmitt E., Meyer E., Van Rossum F., Touzet P. (2022) Reproductive isolation among lineages of *Silene nutans* (Caryophyllaceae): a potential involvement of plastid-nuclear incompatibilities. MPE, 169, 107436



The objective of this chapter was to search for candidate plastid-nuclear gene pairs responsible for plastid-nuclear incompatibilities in inter-lineage hybrids. Working on plastid genomic and nuclear transcriptomic data, we screened the quasi complete set of plastid genes and nuclear genes whose gene products are targeted to the plastid. We conducted molecular data analyses to assess pattern of genetic divergence between lineages, assess the functional impact of the identified mutations, search for potential selective signatures and test for structurally mediated plastid-nuclear co-evolution.

Plastid genomic data were acquired through gene capture before I started my PhD by Cécile Godé, Sophie Galina, Jean-Stéphane Varré and Pascal Touzet. Nuclear transcriptomic data were also acquired before my arrival, during Hélène Martin's PhD<sup>4</sup>. Assemblies of the plastid genomes were done by Sophie Galina. The rest of the analyses were done by myself, with the help of Sophie Galina for the bioinformatic parts of them. Interpretation of the results was done by myself with the participation of Pascal Touzet, Céline Poux, Fabienne Van Rossum and Etienne Meyer. Writing of the manuscript and figure elaboration was done by myself with the close participation of Pascal Touzet. Editing of the manuscript was done by Pascal Touzet, Céline Poux and Fabienne Van Rossum.

---

<sup>4</sup> Martin Hélène, 2016. Processus de spéciation et impact des systèmes de reproduction dans le genre *Silene* - chapitre 3



# Reproductive isolation among lineages of *Silene nutans* (*Caryophyllaceae*): a potential involvement of plastid-nuclear incompatibilities

## Abstract

Early stages of speciation in plants might involve genetic incompatibilities between plastid and nuclear genomes, leading to inter-lineage hybrid breakdown due to the disruption between co-adapted plastid and nuclear genes encoding subunits of the same plastid protein complexes. We tested this hypothesis in *Silene nutans*, a gynodioecious Caryophyllaceae, where four distinct genetic lineages exhibited strong reproductive isolation among each other, resulting in chlorotic or variegated hybrids. By sequencing the whole gene content of the four plastomes through gene capture, and a large part of the nuclear genes encoding plastid subunits from RNAseq data, we searched for non-synonymous substitutions fixed in each lineage on both genomes. Lineages of *S. nutans* exhibited a high level of dN/dS ratios for plastid and nuclear genes encoding most plastid complexes, with a strong pattern of coevolution for genes encoding the subunits of ribosome and cytochrome b6/f that could explain the chlorosis of hybrids. Overall, relaxation of selection due to past bottlenecks and also positive selection have driven the diversity pattern observed in *S. nutans* plastid complexes, leading to plastid-nuclear incompatibilities. We discuss a possible role of gynodioecy in the evolutionary dynamics of the plastomes through linked selection.

**Key words:** coevolution, cytochrome b6/f, plastid-nuclear incompatibilities, ribosome, plastome, reproductive isolation, *Silene nutans*

## 1. Introduction

Speciation, the process that leads populations to reproductive isolation is a main topic of investigation in evolutionary genetics (Matute and Cooper, 2021). The involvement of cytonuclear interactions in the emergence of reproductive barriers has long been underestimated (Levin, 2003). However, the asymmetry of reproductive isolation when hybrids from reciprocal crosses are compared, the so-called Darwin's corollary, suggests possible cytoplasmic origin of genetic incompatibilities (Burton et al., 2013; Turelli and Moyle, 2007). In fact, cytonuclear incompatibilities might contribute to early stages of speciation (Barnard-Kubow et al., 2017). Several studies in plant species have pointed out the role of plastid-nuclear incompatibilities in postzygotic isolation (Bogdanova, 2020; Postel and Touzet, 2020). Due to the dual origin of plastid protein complexes (plastid and nuclear), any mutation fixed in one genome through adaptive or non-adaptive processes, is expected to trigger the selection on partner genome to maintain co-adaptation and a functional plastid (Greiner and Bock, 2013). The pattern of coevolution between plastid and nuclear genes involved in the same plastid complexes is indeed strong (Forsythe et al., 2021) and could subsequently generate incompatibilities between divergent lineages revealed by hybrid breakdown. Hybrid breakdown can be associated with chlorosis, reduced plant fitness, fertility and survival (Greiner et al., 2011). For most cases, the molecular mechanism behind this breakdown is unknown, with the exception of evening primroses (*Oenothera* spp.), where a recent study has demonstrated the involvement of a photosynthesis operon, the expression of which is most likely involved in light acclimation (Zupok et al., 2021).

In the present study, we aimed to assess the possible involvement of plastid-nuclear incompatibilities in *Silene nutans* L. (Caryophyllaceae), a perennial, moth-pollinated herb species from xero-thermophilous habitats showing a wide continental Eurasian distribution. Population genetic studies have revealed the occurrence of several distinct genetic lineages based on plastid sequences and nuclear microsatellite markers, whose geographic distribution in Europe reflects colonization from past glacial refugia (Martin et al., 2016; Van Rossum et al., 2018). Diallelic crosses between four of these lineages, one from the eastern part of *S. nutans* distribution (E1) and three found in Western Europe (W1, W2 and W3), revealed strong reproductive isolation between them (Martin et al., 2017; Van Rossum et al., unpublished results). This reproductive isolation depends on the direction of the cross in reciprocal crosses (i.e. is asymmetrical) and results in chlorotic or variegated hybrids with a high level of mortality or reduced plant fitness at juvenile stage.

To assess whether hybrid breakdown between E1, W1, W2 and W3 genetic lineages of *S. nutans* involves plastid-nuclear incompatibilities, we searched for non-synonymous substitutions in plastid and nuclear genes encoding plastid protein subunit complexes, which are specifically fixed in

lineages. We did so by analysing the whole gene content of the four plastomes, and most of the nuclear genes encoding subunits of plastid complexes. We then estimated functional effect of these non-synonymous substitutions (through a conservation index and their role in contacts between subunits within complexes) as well as whether they exhibited a signature of relaxed or positive selection. Finally, we discuss the different scenarios that could have favoured the emergence of plastid-nuclear incompatibilities in *S. nutans*.

## 2. Material and Methods

The overall workflow is summarized in Figure 1.

### 2.1. Plastid genomic data

To acquire plastid genomic data of *S. nutans*, 47 individuals from 24 populations (1-2 individuals per population) from UK, France, Belgium, Luxemburg, Germany and Finland (Table S1) were sampled from the DNA collection of the unit Evo-Eco-Paleo (UMR 8198 – CNRS University of Lille; see Martin et al., 2016 for DNA extraction procedure) (Table S1). These populations covered four genetic lineages of *S. nutans* based on plastid SNP markers (Martin et al., 2016), with 12 individuals belonging to E1 and W1 lineages and 8 individuals to W2 and W3 lineages. Genomic sequences for each individual were obtained through gene capture with a myBaits® target capture kit (Daicel Arbor Biosciences, <https://arborbiosci.com/>). DNA probes were defined from the published sequence of the plastid genome of *Silene latifolia* (NCBI accession: CI\_001 – NC\_016730.1). In order to get 1 Mbp per sample with a theoretical coverage of 50X, 4,255 probes of 120 nucleotides long were defined for a total target length of 267,156 bp with a 2x density (i.e. 534,328 bp). Enriched libraries were pooled and sequenced on Illumina MiSeq in paired-end (2x150 of the 48 genomic libraries for our sample, in dual index kappa) at the LIGAN platform (UMR 8199 LIGAN-PM Genomics platform – Lille, France), resulting in a total of 38 million reads. De novo assemblies of the reads were generated using SPAdes (Bankevich et al., 2012). Their quality was assessed by YASS (Noé and Kucherov, 2005) and Blast analyses (Table S2).

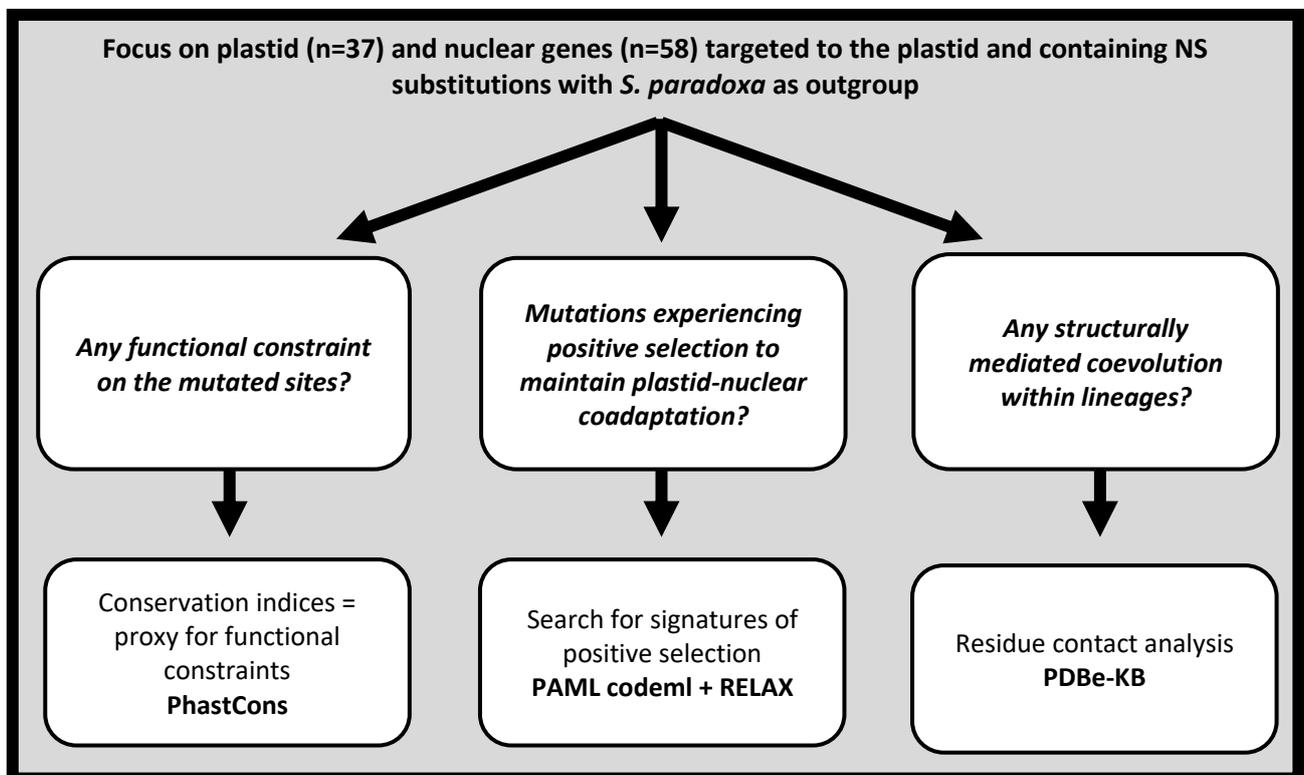
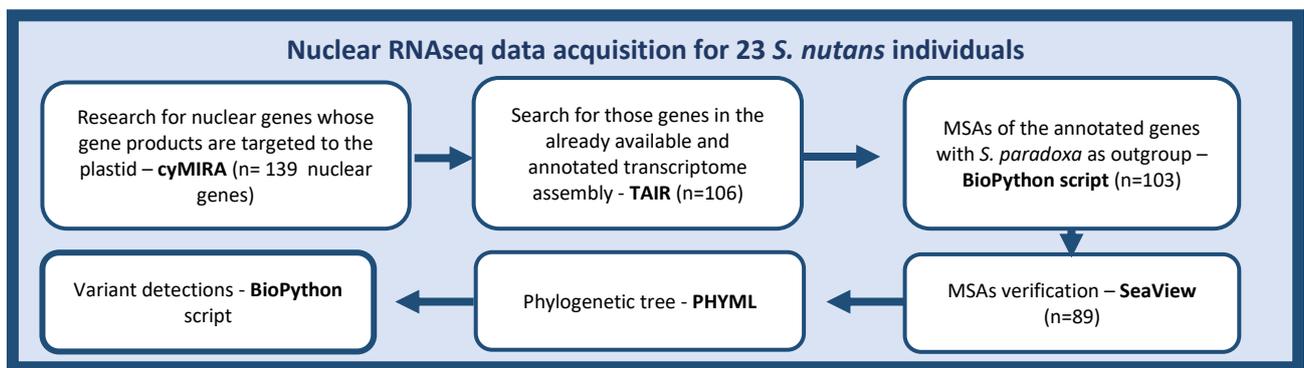
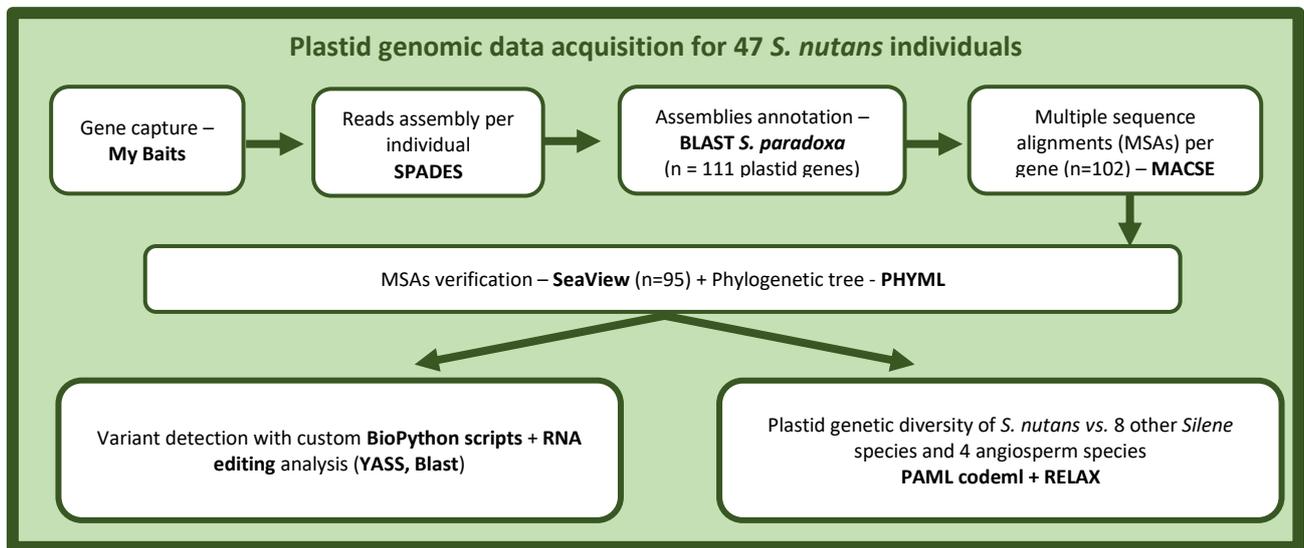
Multiple sequence alignments (MSAs) were generated using *Silene paradoxa* plastid genome assembly (NCBI accession: NC\_023360.1), which was not available when the baits were defined and was phylogenetically closer to *S. nutans* (Jafari et al., 2020). The 111 gene sequences of *S. paradoxa* were blasted against the assemblies of *S. nutans* individuals and the best hit for each gene was extracted. For two genes, *clpP* and *trnK-UUU*, no hits were detected. Blast results were then filtered for a percentage of identity of 90% and a length of 30 bp to also get the small *trn* genes. After filtration,

we conserved 103 genes out of the 111 annotated in *S. paradoxa* plastid genome. The gene *ycf1* was also excluded because it was missing in more than 50% of our samples. With the program MACSE v.2.00 (Ranwez et al., 2011), MSAs of the blasted plastid genes were generated and aligned, containing the sequences of *S. latifolia*, *S. paradoxa* and of the 47 sampled individuals of *S. nutans*. MSAs were manually cleaned and checked on SeaView v.4.7 (Gouy et al., 2010). Seven genes were excluded following this step, depending on the quality of the alignment and the correct frame of the ORFs (i.e. *accD*, *atpF*, *rpoC1*, *rps12*, *trnT-GGU*, *ycf2* and *ycf3*). Among the remaining 95 plastid genes, 23 were *trn* genes and 4 were *rrn* genes which were not analysed here. A phylogenetic tree was generated with PHYML v.3.0 (Guindon et al., 2010) with the concatenated set of the 68 genes, using the GTR + G model of nucleotide substitution selected by SMS (*Smart Model Selection*) (Lefort et al., 2017).

To compare the evolutionary trends of *S. nutans* plastid genomes, we constructed MSAs including 10 other angiosperm and *Silene* species for which plastid genome assemblies were available on NCBI : *Arabidopsis thaliana* (NC\_000932), *Nicotiana tabacum* (NC\_001879), *Oryza sativa* (NC\_001320) and *Zea mays* (NC\_001666), and for *Silene* species: *S. capitata* (NC\_035226.1), *S. chalcedonica* (NC\_023359.1), *S. conica* (NC\_016729.1), *S. conoidea* (NC\_023358.1), *S. noctiflora* (NC\_012728.1), and *S. vulgaris* (NC\_016727.1). Plastid gene sequences of *S. nutans* individuals were blasted against plastome assemblies of these species (percentage of identity of 98%) and the best blast hit was extracted. To avoid bias due to the larger amount of data for *S. nutans* lineages, MSAs were constructed for the 68 plastid genes with only one sequence per *S. nutans* lineage. To choose which individual to keep for this analysis, within each lineage we selected one individual which was the one for which the assembly was the least fragmented. MSAs with *Silene* and angiosperm species nucleotide sequences were aligned with MUSCLE and manually checked on SeaView. Concatenations were generated per plastid complex: photosystem I and II (PSI and PSII), ATP synthase, cytochrome b6/f, Rubisco, NDH, RNA polymerase, and large and small ribosomal subunits (LSU and SSU, respectively) - except for three genes (*ccsA*, *cemA* and *matK*). For all concatenations, phylogenetic trees were generated using PHYML v.3.0 with the GTR model of nucleotide substitutions selected by SMS.

## 2.2. Nuclear data

To identify nuclear genes potentially involved in plastid-nuclear incompatibilities (PNIs), we searched for nuclear genes encoding subunits of plastid complexes. We used the protein-protein interaction database cyMIRA (Forsythe et al., 2019), the Protein Data Bank in Europe-Knowledge Base



**Fig.1 Summary of the methods used for detecting plastid-nuclear incompatibilities between lineages of *S. nutans*.**

(PDBe-KB) (Varadi et al., 2020), SUBA4 (Hooper et al., 2017), UniProt (Bateman, 2019), STRING v.11 (Szkarczyk et al., 2019), and BioGrid (Oughtred et al., 2019) to establish a list of 139 putative nuclear interactors (Table S3).

A transcriptome assembly was already available as well as transcriptomic data for 22 individuals from the four genetic lineages of *S. nutans* (Table S1) (Muyle et al., 2021; PRJEB39526). RNAs were extracted from flower buds. An in-house BioPython script was used to align the reads of these 22 individuals on the transcriptome assembly. We used BamBam (SAMtools package) (Page et al., 2014) to generate one consensus sequence per individual from the reads. From this, we extracted the nuclear genes identified and annotated in the transcriptome, using their position in the transcriptome assembly and their TAIR identifiers. Only 106 out of the 139 nuclear genes were found in *S. nutans* transcriptomes and were subsequently analysed. MSAs per gene containing *S. nutans* individuals were blasted on *S. latifolia* and *S. paradoxa* transcriptomes (PRJEB39526), using the same criteria as for plastid genes, extracting the best hits for both species. Three nuclear genes for which no hit was identified in *S. latifolia* or *S. paradoxa* were excluded. The MSAs for the 103 remaining nuclear genes were then aligned with MUSCLE and checked on SeaView. Fourteen genes for which alignment with a correct ORF was not possible were excluded. A concatenated set of 89 nuclear genes was generated to construct a phylogenetic tree using PHYML v.3.0 with the GTR model of nucleotide substitution selected by SMS. As with the plastid data, we constructed nuclear gene concatenation per plastid complex.

To compare evolutionary patterns of nuclear genes encoding subunits of the plastid ribosome with the nuclear genes encoding subunits of the cytosolic ribosome (i.e. gene products not targeted to the plastid), RNAseq data for these genes were also extracted and aligned. We took the 204 nuclear genes used in Sloan, Triant, Wu, et al. (2014) and their available TAIR identifiers. The same steps as described above were followed, and MSAs containing the sequences of the 22 *S. nutans* individuals, *S. paradoxa* and *S. latifolia* were generated for 81 genes. Indeed, only 87 out of 204 genes were found in the transcriptomic data, 5 out of these 87 genes did not have any blast hits with *S. paradoxa* and *S. latifolia* transcriptomes and for one out of the remaining 82 we could not construct an alignment with a correct ORF.

### 2.3. Variant detection

We used an in-house BioPython script to identify the SNPs (single nucleotide polymorphisms) differing between the four lineages in the MSAs of the 95 plastid genes. As RNAseq data of *S. nutans* were available (Muyle et al. 2021), we also checked whether the identified non-synonymous

substitutions were “true non-synonymous substitutions” or RNA editing sites (T to C substitutions) using RNAseq data and confirmed that they were non edited (see Appendix S1).

For the nuclear genes, the diploid individual genotypes were determined at each position in the individual alignment with the program reads2spn (Gayral et al., 2013) and used to identify the nuclear variants. An in-house BioPython script was used to identify the SNPs only present in one lineage. Only SNPs for which at least two individuals per lineage were genotyped were kept. To determine non-synonymous or synonymous nature of the identified SNPs, we used the available assembly of *S. nutans* transcriptome, for which ORFs were set.

#### 2.4. Conservation status analysis

First, to improve the accuracy of the analysis, additional outgroups were added to the existing nuclear MSAs. We used the transcriptome assemblies of *Dianthus chinensis*, *S. diclinis*, *S. dioica*, *S. heuffelli*, *S. latifolia*, *S. marizii*, *S. otites*, *S. paradoxa*, *S. pseudotites*, *S. viscosa* and *S. vulgaris* (Muyle et al., 2021). Nuclear gene sequence data were extracted from these transcriptomes and new MSAs were constructed, integrating the sequences of those species. For the plastid genes, the analysis was run on the 12 species used for the plastid genetic diversity analysis. Then, for both nuclear and plastid data, a phylogenetic tree was constructed using the concatenated set of the nuclear and plastid genes, respectively, with PHYML – SMS, which assessed the best model of nucleotide substitutions to use.

We then estimated the conservation degree of the mutated amino-acids as a proxy for the functional constraint acting on the mutated sites of the plastid and nuclear genes using the PHAST program (Hubisz et al., 2011). For this purpose, we ran PhyloFit analyses to fit a phylogenetic model to the MSA by maximum likelihood (Siepel and Haussler, 2004), using HKY85 as model of nucleotide substitutions and the EM option. With the phylogenetic model output file of PhyloFit, we ran PhastCons a first time as a training step to estimate PhastCons free parameters as suggested in Hubisz *et al.* (2011) (option – *no-post-probs*). Then, we ran PhastCons a second time with the formerly estimated parameters, to compute the conservation scores for each nucleotide site in a given MSA, varying from 0 = poorly conserved to 1 = strongly conserved. These three steps were conducted for each MSA.

#### 2.5. Analysis of pattern of selection

For assessing the evolutionary dynamics and selection patterns on nuclear and plastid genes in *S. nutans*, we estimated the  $d_N/d_S$  ratios, with  $d_N$  = non-synonymous substitution rate and  $d_S$  = synonymous substitution rate using codeml implemented in PAML v.4.9 (Yang, 2007). We ran different

analyses with several models. For nuclear genes, as transit peptides of the proteins targeted to the plastid are highly variable and thus could increase  $d_N/d_S$  ratios (Christian et al., 2020), we used the TargetP software (Juan et al., 2019) to identify their locations in the nuclear sequences and remove them from the nuclear alignments.

Two types of analyses have been conducted with codeml, varying in terms of their assumptions about how  $d_N/d_S$  varies across the sequence (site models) or across branches of the phylogeny (branch models). Branch models were used to compare the  $d_N/d_S$  ratios of *S. nutans* lineages with those of other *Silene* and angiosperm species. These analyses were conducted on the plastid gene concatenations per plastid complex, with the angiosperms and *Silene* species as outgroups (Figure S1, Appendix S2). Compensatory mutations could be a target of positive selection for the maintenance of the interactions between plastid and nuclear gene partners. To detect signatures of positive selection, either on specific non-synonymous substitution sites or on particular genes, site models were run on plastid and nuclear genes alignments, with only *S. latifolia* and *S. paradoxa* as outgroups. Separated analyses were run for the plastid and nuclear ribosome genes and the 81 nuclear genes encoding the cytosolic ribosome (i.e. not targeted to the plastid, later called NuCyto) (Appendix S2). For each analysis, the  $d_N/d_S$  values of the best models were kept. To test for differences in  $d_N/d_S$  values on plastid vs. nuclear genes and on nuclear genes encoding the plastid ribosome vs. NuCyto, we performed Mann-Whitney *U* tests on R v. 1.3.1093 (package *stats4*).

A second analysis of selection with RELAX (Wertheim et al., 2015) was performed to assess selective pressures acting on nuclear and plastid genes in *S. nutans*. RELAX allows to distinguish positive selection from relaxed purifying selection in case of elevated  $d_N/d_S$  ratios. We used the program RELAX with default parameters values. We used the same dataset as for the codeml analyses : first the plastid gene concatenations with the angiosperm and *Silene* species and then the plastid and nuclear gene alignments, with only *S. latifolia* and *S. paradoxa*. For both datasets, the outgroups were annotated as references branches and *Silene nutans* individuals were again annotated as test branches, for both nuclear and plastid genes.

Finally, to test for parallel pattern of mutation accumulation between plastid and nuclear genes of *S. nutans* lineages, we performed a Spearman's rank correlation analysis between  $p_N/p_S$  ratios of plastid and nuclear gene concatenations per complex on *S. nutans* dataset. We used DNAsp v.6 (Rozas et al., 2017) to calculate  $p_N$  and  $p_S$  values, i.e. non-synonymous and synonymous divergences, respectively, between *S. nutans* lineages.

## 2.6. Identification of contact positions in plastid complexes

To assess whether some non-synonymous substitutions identified in plastid and nuclear genes were located at contact positions in the plastid complexes, we conducted a residue contact analysis, using crystallographic structures of the different plastid complexes. The PDBe-KB references of the plastid complexes were extracted from cyMIRA database (Forsythe et al., 2019). When the crystallographic structure of the plastid complex was available, we extracted the protein sequences of the organism with which the crystallographic structures had been defined and aligned them with the protein sequences of the *S. nutans* individuals. Data were not available for RNA polymerase and Rubisco.

### 3. Results

#### 3.1. Genomic and transcriptomic data

Plastid genome assemblies in *S. nutans* were relatively fragmented as they were composed of 6 nodes on average, with a minimum of 3 nodes for individuals AIG-1 and AIG-10 (W3 lineage) and a maximum of 23 for individual UK15-16 (W1 lineage). The data obtained represent between 86.25% (W2) and 96.63% (E1) of the complete *S. latifolia* plastid genome (Table S2). Compared with the plastid genome of *S. paradoxa*, most of the genes of the plastid complexes were recovered, with only a few plastid complexes lacking one gene (ATP synthase, NDH, RNA polymerase and large and small ribosomal subunits) (Table 1). As for the plastid genome, about 75% of the nuclear genes annotated in the *S. paradoxa* transcriptome whose gene products were targeted to the plastid found a blast hit in *S. nutans* transcriptomic assemblies (Table 1, Table S3). Nonetheless, compared to the plastid genomes, nuclear gene annotation was less complete, most likely depending on the level of gene expression in flower buds.

#### 3.2. Plastid genome diversity in *Silene nutans*

Variant analyses revealed that a high proportion of the annotated plastid genes were variable (72.1%), containing synonymous or non-synonymous substitutions (Table 1). The number of synonymous substitutions (66) was lower than the number of non-synonymous substitutions (139). All the plastid complexes contained plastid genes with mutated sites, especially the plastid genes encoding the expression machinery, i.e. the large and small ribosomal subunits and the RNA polymerase subunits (24.5%, 31.7% and 12.2% of the total number of non-synonymous substitutions were found in these three complexes, respectively). On average, half of the identified non-synonymous

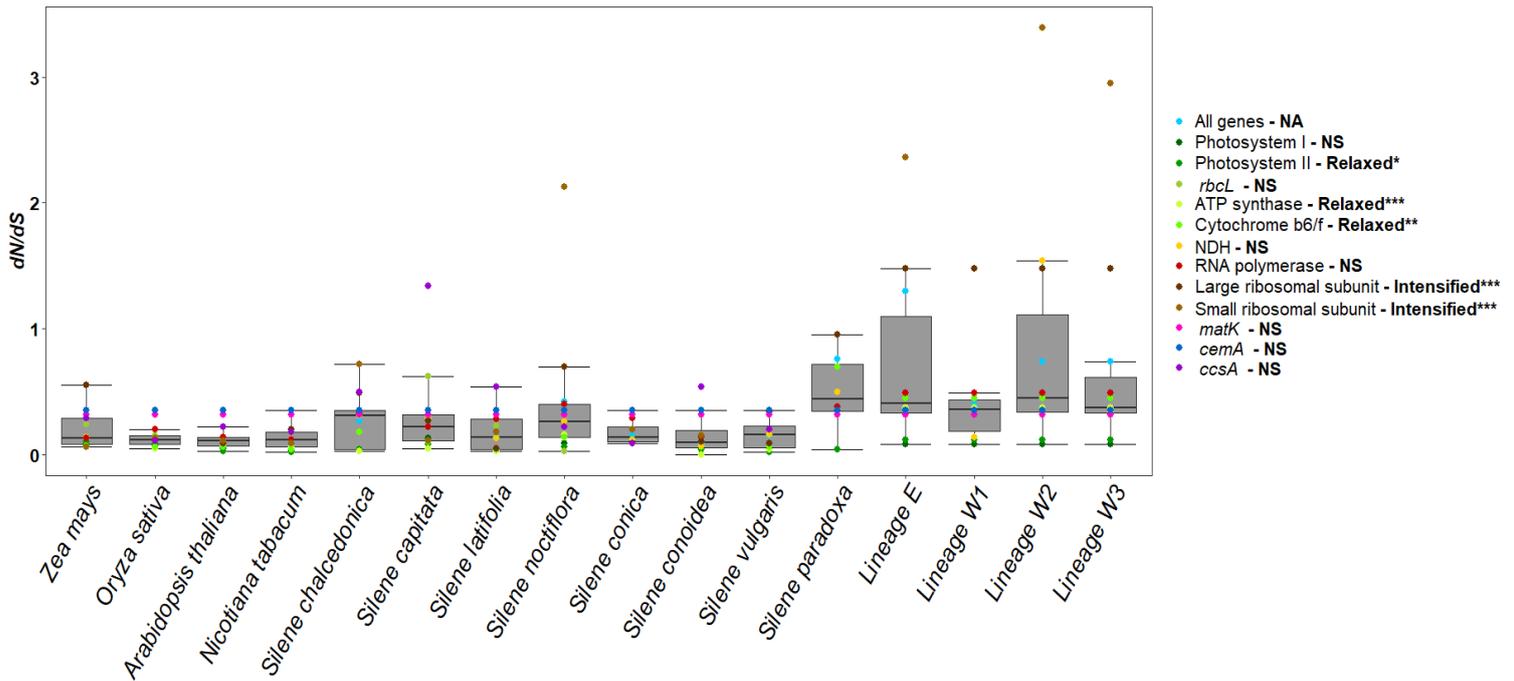
**Table 1**

Summary of the plastid and nuclear genetic diversity per plastid complex.

Plastid Complex	Genome	Number of genes				Gene concatenations length (in bp)	Number of synonymous substitutions	Number of non-synonymous substitutions			
		Total	Annotated	Analyzed	Variable			Total	Conserved positions	Positive selection	Residue contact
PSI	Plastid	5	5	5	2	5355	7	1	1	0	0
	Nuclear	20	16	14	8	10,624	17	10	0	0	1
PSII	Plastid	15	15	15	7	6672	18	5	5	0	0
	Nuclear	28	19	17	10	11,107	4	32	0	2	0
Rubisco	Plastid	1	1	1	1	1425	1	0	0	NA	NA
	Nuclear	4	1	1	1	1095	1	4	0	0	0
ATP synthase	Plastid	6	6	5	4	4375	3	6	3	0	0
	Nuclear	6	5	4	3	2858	3	6	1	0	0
Cyb6F	Plastid	6	6	6	2	2362	1	3	2	0	0
	Nuclear	4	4	4	4	3440	3	18	2	9	1
NDH	Plastid	11	11	11	7	10,370	8	17	10	1	1
	Nuclear	18	11	5	1	2472	1	1	0	0	0
RNA polymerase	Plastid	4	4	3	3	8279	11	17	9	2	0
	Nuclear	10	10	7	5	5654	1	8	1	0	2
LSU	Plastid	9	9	8	8	2338	6	34	19	1	8
	Nuclear	30	24	22	17	12,606	16	56	10	7	4
SSU	Plastid	12	11	10	7	4077	6	44	16	5	8
	Nuclear	12	10	10	8	6623	8	40	12	10	6
Others	Plastid	9	8	4	4	3157	5	12	7	0	NA
	Nuclear	NA	6	4	4	4449	5	9	1	2	NA
Total	Plastid	78	76	68	45	x	66	139	74	9	17
	Nuclear	139	106	88	61	x	63	180	27	30	14

Number of genes: Total: the total number of genes annotated in the plastid complex (references taken on PDBe-KB). Trn and rrn genes were not taking into account here ; Annotated: number of annotated genes in the *S. nutans* plastid genomes and transcriptomes ; Analysed: number of genes analyzed per genome and plastid complex ; Variable: number of variable genes per genome in each plastid complex ; Total: total number of NS substitutions per genome and plastid complex ; Conserved positions: number of NS substitutions located at strongly conserved positions (i.e. conservation indice > 0.8) ; Positive selection: number of NS substitutions identified under positive selection with PAML NSites model.

substitutions were located at strongly conserved positions (i.e. conservation index > 0.8), and most of these substitutions were found in the nucleotide sequences of the plastid genes encoding components of the plastid ribosome (Table 1). For these genes, some of the non-synonymous substitutions were located at strongly conserved positions (19/34 encoding the LSU and 16/44 encoding the SSU) and among them, some led to a change in the functional class of the encoded amino-acid (4/19 for the LSU and 8/16 for the SSU). These plastid genes also contained 94% of the non-synonymous substitutions located at residue contact positions (i.e. 8 substitutions for both large and small ribosomal subunit)

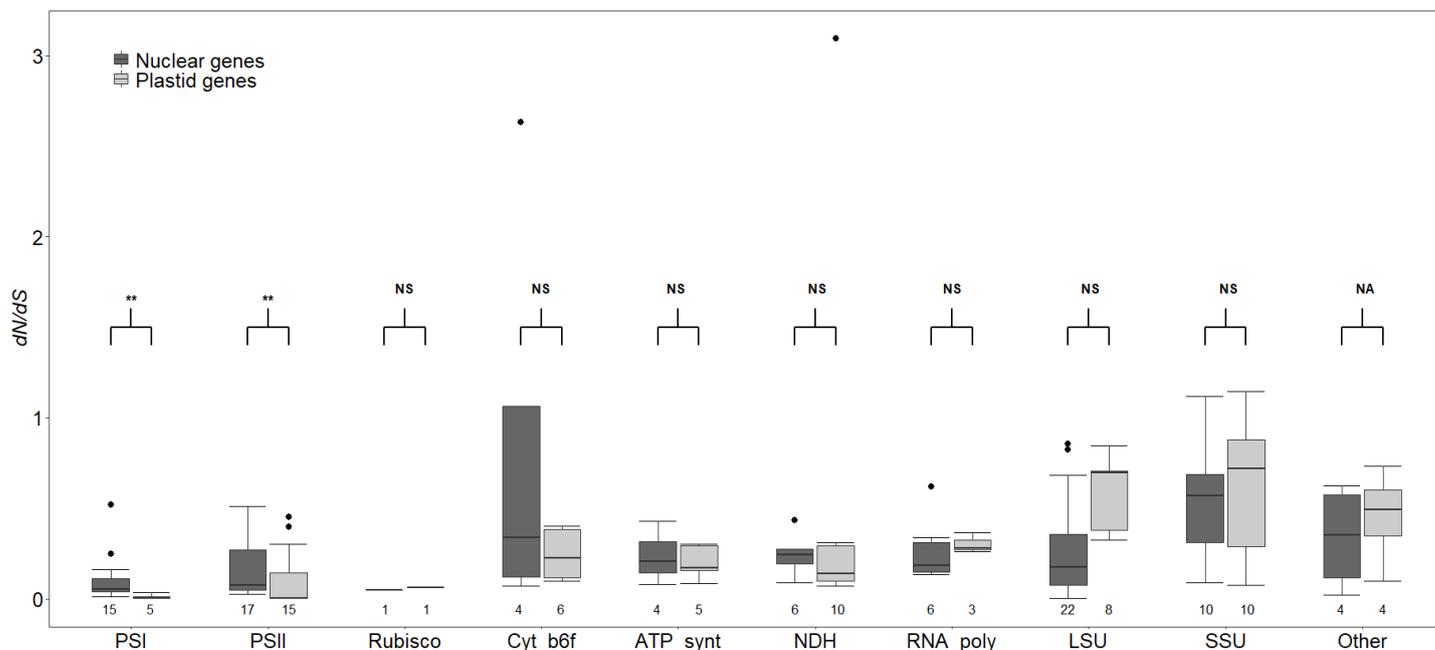


**Fig.2  $d_N/d_S$  ratios of plastid gene concatenations per plastid complex for each lineage of *Silene nutans* and the outgroups used.** The first species are the 4 angiosperm species, the next ones regroups the *Silene* species and the last four are the four lineages of *Silene nutans*. RELAX results as well as P value for the plastid gene concatenations of *S. nutans* are given in top right of each graph : ° < 0,1 ; \* < 0,05 ; \*\* < 0,01 ; \*\*\* < 0,001 ; NA = no result for this category due to RELAX own limitations.

(Table 1). The synonymous substitutions exhibited contrasting patterns of accumulation. An over represented portion of these substitutions (27.3%) was located in the genes encoding subunits of the photosystem II and very few substitutions were identified in the genes involved in the plastid gene expression machinery (Table 1).

Overall, *S. nutans* lineages exhibited the highest  $d_N/d_S$  ratios in the gene concatenations of plastid complexes compared with other *Silene* and angiosperm species (Figure 2), even with *Silene* species known to exhibit accelerated rates of organellar genome evolution (*S. conica*, *S. conoidea*, *S. noctiflora*). For these species, elevated  $d_N/d_S$  ratios were only observed in the large and small ribosomal subunits (Sloan et al., 2014). In *S. nutans*, not only gene concatenation of these two subunits had elevated  $d_N/d_S$  ratios (> 1), but this was also observed for all plastid complexes, especially in W2 and E1 lineages, and for the plastid genes encoding the cytochrome b6/f (Figure 2). Indeed, the latter exhibited the highest  $d_N/d_S$  ratio among the plastid genes encoding complexes involved in photosynthesis and compared to angiosperm and other *Silene* species. The particular case of *S. paradoxa* must be noted, with a median  $d_N/d_S$  ratio slightly higher than W1 and W3 lineages (Figure 2).

Using RELAX, we identified a tendency for relaxed purifying selection on the plastid genes encoding the plastid complexes of *S. nutans* (Figure 2). For large and small ribosomal subunits, showing  $d_N/d_S$  ratios > 1, RELAX results identified intensification of positive selection on their plastid genes (Figure 2).



**Fig.3**  $d_N/d_S$  ratios calculated on each *S. nutans* gene separately for plastid and nuclear genes encoding plastid complexes. In this analysis all four *S. nutans* lineages display the same  $d_N/d_S$  values: branch models analyse with different  $d_N/d_S$  ratios for each lineage were significant only for 3 plastid and 14 nuclear genes (data not shown) and exhibited aberrant values (999.99). We have therefore chosen to report the  $d_N/d_S$  ratios assessed with the best NSites model (one  $d_N/d_S$  ratio for all lineages of *S. nutans*). Overall,  $d_N/d_S$  ratios of *S. nutans* were similar between NSites and Branch-Model when a single  $d_N/d_S$  ratios was set for all four lineages. Mann-Whitney U tests P values: \*\* = < 0.01 ; NS = non-significant ; NA = no test conducted. The number of analysed plastid and nuclear genes are reported under each boxplot. PSI : Photosystem I ; PSII : Photosystem II ; Cyt-B6F : Cytochrome b6/f ; ATP\_synt : ATP synthase ; RNA\_poly : RNA polymerase ; LSU : Large ribosomal subunit ; SSU : Small ribosomal subunit ; Other : other functions. Numbers under the boxplots represent the number of genes in each category.

### 3.3. Analyses of nuclear genes encoding subunits of plastid complexes

As observed on plastid genes, more than half (57.5%) of the nuclear genes annotated in *S. nutans* lineage transcriptomes contained synonymous or non-synonymous substitutions. The amount of synonymous substitutions identified in nuclear genes was also lower than the amount of non-synonymous substitutions, and the difference between the number of synonymous and non-synonymous substitutions in nuclear genes was higher than for plastid genes (63 synonymous and 180 non-synonymous substitutions in nuclear genes vs. 66 synonymous and 139 non-synonymous substitutions in plastid genes; Table 1). Nuclear genes encoding products targeted to the plastid accumulated more non-synonymous substitutions than the plastid ones. Despite some specificities to each genome (i.e. substitutions in NDH complex for plastid genes vs. in cytochrome b6/f and photosystem II for nuclear genes), the accumulation patterns of the nuclear and plastid non-synonymous substitutions were similar. Most of the non-synonymous substitutions were located in the nuclear genes encoding subunits of the plastid ribosome (more than 50% of them) and the RNA polymerase (Table 1). For these nuclear genes, 22/96 non-synonymous substitutions led to changes of

amino-acid functional class (i.e. 17 for the genes encoding the SSU and 5 for the LSU). They also contained 71.4% of the substitutions located at residue contact positions (Table 1). For the cytochrome b6/f, non-synonymous substitutions were also identified, especially in the nuclear gene *petM* for which 11/13 substitutions led to changes of amino-acid functional class (Table 1, Table S4).

For non-synonymous substitutions, 15% were located at strongly conserved positions in the nuclear genes, which was less than for the plastid genes (53.2%) (Table 1). These mutations were mostly found in the nuclear genes encoding subunits of the plastid ribosome (81.5%), as observed for the plastid non-synonymous substitutions. The remaining non-synonymous substitutions at strongly conserved sites were identified in the nuclear genes encoding subunits of the cytochrome b6/f, ATP synthase, and RNA polymerase and in nuclear genes with other functions (Table 1).

### 3.4. Patterns of selection and coevolution

Sites models analyses in codeml pointed out non-synonymous substitutions in plastid and nuclear genes under positive selection in *S. nutans* genomes (Table 1). These positively selected mutations mostly concerned genes encoding the large and small ribosomal subunits for both plastid and nuclear compartments and the cytochrome b6/f for nuclear compartment (23.3%, 40% and 30% of the nuclear genes, respectively). Overall, nuclear genes encoding the plastid ribosome exhibited 56.7% of the nuclear sites under positive selection.  $d_N/d_S$  ratios between nuclear and plastid genes were significantly different for photosystem I (Mann-Whitney  $U$  tests = 3,  $P < 0.003$ ), photosystem II (Mann-Whitney  $U$  tests = 46,  $P < 0.03$ ), and not significantly different for the rest of the plastid complex gene pairs ( $P > 0.05$ ) (Figure 3). We did not find a significant correlation between  $d_N/d_S$  ratios of plastid and nuclear gene concatenations encoding the same complexes ( $r_s = 0.17$ ,  $P = 0.66$ ) but a significant one between  $p_N/p_S$  ratios, when focusing on divergence among lineages ( $r_s = 0.75$ ,  $P = 0.02$ ), suggesting a pattern of coevolution (Figure S2).

$d_N/d_S$  ratios were overall higher for nuclear genes (between 0.2 and 0.3) than for plastid genes (generally  $< 0.2$ ; Figure 3), except for the large ribosomal subunit where the reverse pattern was observed. For both nuclear and plastid genomes, elevated  $d_N/d_S$  ratios ( $> 1$ ) were observed for some genes encoding subunits of the plastid ribosome (Figure 3, Table 2). The nuclear genes of cytochrome b6/f also exhibited extremely high  $d_N/d_S$  ratios, with  $d_N/d_S \approx 2$  for *petM* (Figure 3, Table 2). Besides, RELAX results were significant and indicated intensification of positive selection for *petM*. The analysis was not decisive for most of the other genes (Table 2, Tables S5 and S6).

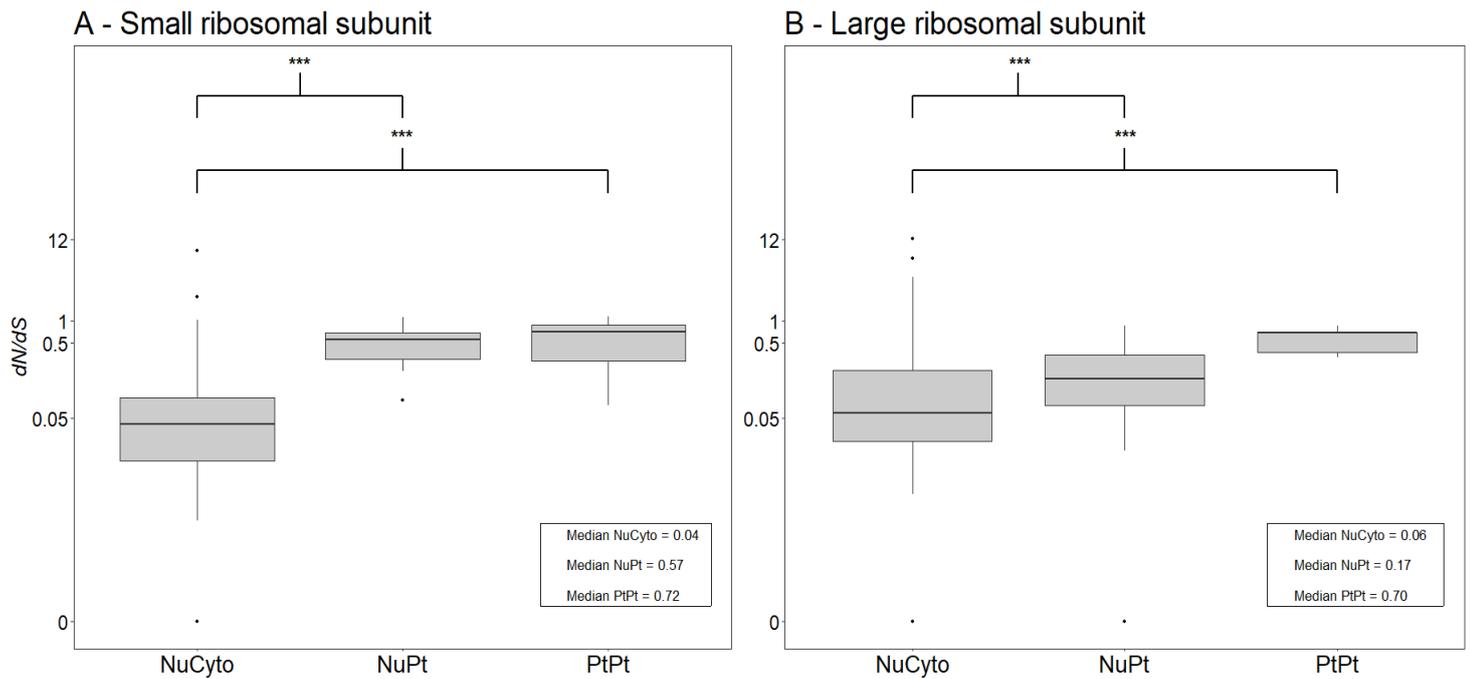
In order to formally test the hypothesis of plastid-nuclear coevolution, we calculated  $d_N/d_S$  ratios of nuclear genes encoding the subunits of the cytosolic ribosome, i.e. not targeted to the plastid. For

**Table 2**

$d_N/d_S$  ratios of the nuclear and plastid genes of the cytochrome b6/f, and of the large and small ribosomal subunits.

Complex	Genome	Gene	$d_N/d_S$	Complex	Genome	Gene	$d_N/d_S$
Cytochrome b6/f	Plastid	<i>petA</i>	0.22	Large ribosomal subunit	Plastid	<i>rpl2</i>	0.7
		<i>petB</i>	0.12			<i>rpl14</i>	0.33
		<i>petD</i>	0.18			<i>rpl16</i>	0.15
		<i>petG</i>	0.00			<i>rpl20</i>	0.51
		<i>petL</i>	NA			<i>rpl22</i>	1.01
	Nuclear	<i>DAC</i>	0.14		<i>rpl32</i>	0.50 <sup>°°</sup>	
		<i>petC</i>	0.07		<i>rpl33</i>	0.7	
		<i>petM_1</i>	2.95 <sup>**</sup>		<i>rpl36</i>	NA	
		<i>petM_2</i>	2.31		Nuclear	<i>EMB3105</i>	0.12
		<i>PPR</i>	0.54			<i>PRSP1</i>	0.68
Small ribosomal subunit	Plastid	<i>rps2</i>	0.64	<i>PSRP6</i>	0.14		
		<i>rps3</i>	0.26	<i>rpl1</i>	0.21		
		<i>rps4</i>	0.29	<i>rpl3</i>	0.1		
		<i>rps7</i>	NA	<i>rpl5</i>	0.32		
		<i>rps8</i>	0.08	<i>rpl6</i>	0.37		
		<i>rps11</i>	0.92	<i>rpl9</i>	0.47		
		<i>rps14</i>	0.21	<i>rpl10</i>	0.05		
		<i>rps15</i>	0.73	<i>rpl11</i>	0.1		
		<i>rps18</i>	NA	<i>rpl12</i>	0.03		
		<i>rps19</i>	0.73	<i>rpl13</i>	0.82		
		Nuclear	<i>PSRP3</i>	0.65	<i>rpl15</i>	0.22 <sup>°°</sup>	
			<i>rps1</i>	0.29	<i>rpl17</i>	0.02	
			<i>rps5</i>	0.62	<i>rpl18</i>	0.10	
			<i>rps6</i>	1.12 <sup>**</sup>	<i>rpl19</i>	0.21	
	<i>rps9</i>		0.09	<i>rpl21</i>	0.26		
	<i>rps10</i>		0.36	<i>rpl24</i>	0.07		
	<i>rps13</i>		0.7	<i>rpl27</i>	0.00		
	<i>rps17</i>	0.2	<i>rpl28</i>	0.04			
	<i>rps20</i>	0.52	<i>rpl29</i>	0.5			
	<i>rps21</i>	1.83	<i>rpl31</i>	0.85			

RELAX results : (\*)= intensification and (°) = relaxation. P value : \* or ° < 0.5 ; \*\* or °° < 0.01 ; \*\*\* or °°° < 0.001.

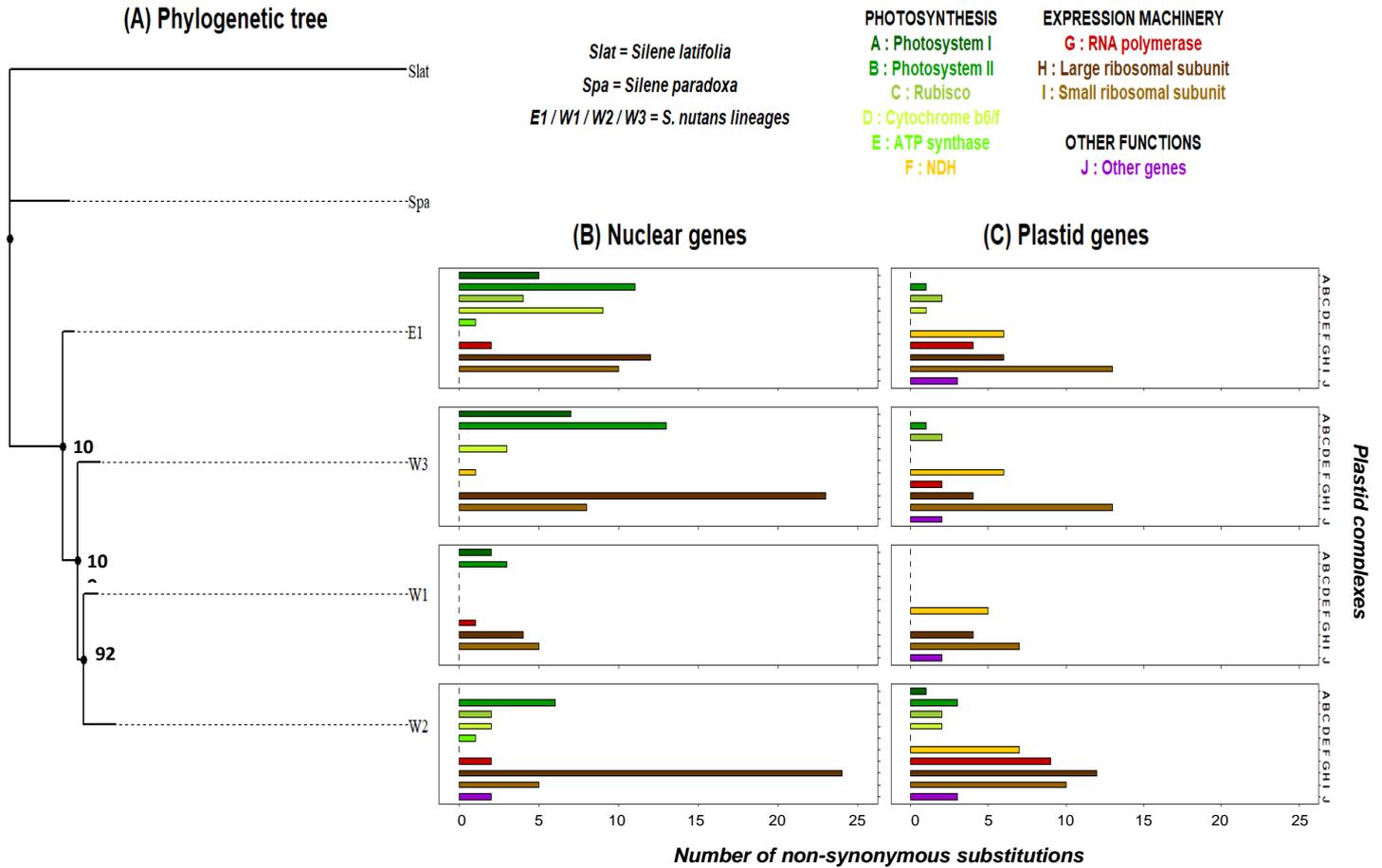


**Fig.4**  $d_N/d_S$  ratios calculated on each *S. nutans* gene separately for the plastid and nuclear genes encoding the large (A) and the small (B) SU of the plastid ribosome and for the nuclear genes encoding the cytosolic ribosomes in *S. nutans*. In this analysis all four *S. nutans* lineages display the same  $d_N/d_S$  values.  $d_N/d_S$  ratios were calculated with codeml (NSites models: one  $d_N/d_S$  ratio for all lineage of *S. nutans*). Significant differences in the distributions of  $d_N/d_S$  ratios were tested with non-parametric Mann-Whitney U tests. NuCyto: Nuclear genes encoding proteins of the cytosolic ribosome. NuPt : Nuclear genes encoding proteins of the plastid ribosome. PtPt : Plastid genes encoding proteins of the plastid ribosome.

the SSU, they showed a  $d_N/d_S$  ratio 10 times lower than the ratio of the nuclear genes encoding the SSU of the plastid ribosome (median value = 0.04 and 0.57, respectively; Figure 4), and almost 20 times lower than the ratio of the plastid genes encoding the SSU of the plastid ribosome (median value = 0.72). For the LSU of the plastid and cytosolic ribosome,  $d_N/d_S$  ratios of the nuclear genes encoding LSU of the cytosolic ribosome was around 6 and 10 times lower than nuclear and plastid genes encoding the LSU of the plastid ribosome, respectively (median value = 0.06, 0.17 and 0.70, respectively; Figure 4). These differences were significant (Mann-Whitney U tests = 173-534,  $P < 0.001$ ). Excess non-synonymous substitutions observed in nuclear genes encoding the plastid ribosome and elevated  $d_N/d_S$  ratio (in some case  $>1$ ) could thus be driven by plastid-nuclear coevolution within the lineages of *S. nutans*

### 3.5. Toward the identification of plastid-nuclear gene pairs candidates for PNI

As plastid-nuclear coevolution is expected to be lineage-specific, we looked at patterns of substitution accumulation within the four lineages of *S. nutans* in nuclear and plastid genes. As expected from former studies, phylogenetic relationships between *S. nutans* lineages showed that the western lineages (W1, W2 and W3) clustered together, with the eastern lineage (E1) diverging before



**Fig.5 Non-synonymous substitutions in each lineage of *S. nutans*, per plastid complex.** (A) Phylogeny constructed with PHYML based on the concatenated set of the 95 plastid genes, with *S. latifolia* and *S. paradoxa* as outgroups; number of substitutions per plastid complex for the nuclear (B) and plastid (C) genes.

the diversification of the western lineages (Martin et al., 2016; Van Rossum et al., 2018) (Figure 5A). E1 and W2 lineages contained a high number of non-synonymous substitutions in both plastid and nuclear genes: 36 and 55 for E1 and 49 and 44 for W2, respectively (Figure 5B and C). The W1 lineage exhibited fewer non-synonymous substitutions (17 and 15 for plastid and nuclear genes, respectively). The W3 lineage contained an intermediate number of substitutions in plastid genes (30), but in nuclear genes it contained as many substitutions as E1 (Figure 5B and C).

As PNI might be due to the disruption of co-adapted plastid-nuclear gene pairs specific to a given lineage, we focused on gene pairs with non-synonymous substitutions found in conserved sites and exhibiting high levels of  $d_N/d_S$  ratios (which can suggest positive selection). In all lineages and for both genomic compartments, the highest number of non-synonymous substitutions concerned the genes encoding the LSU and SSU of the plastid ribosome (ranging from 44.9% of substitutions in plastid genes for W2 to 56.4% of substitutions in nuclear genes for W3; Figure 5B and C). Most of the fixed substitutions of W1 and W2 lineages were also located in plastid and nuclear genes encoding

components of the plastid ribosome (Figure 5C). This concerned essential plastid-encoded ribosomal genes: *rpl20*, *rpl22*, *rpl32*, *rps2*, *rps3* and *rps18* and essential nuclear-encoded ones : *rpl3*, *rpl13*, *rpl21*, *rpl27*, *rps5*, *rps9*, *rps13* (Tiller et al., 2012; Tiller and Bock, 2014). Among the substitutions in these genes, around 50% were located in E1 lineage for the LSU and 30% in W2 lineage for SSU (Table 1, Tables S7 and S8). RELAX results reported intensification of positive selection in the plastid genes encoding these two complexes with especially elevated  $d_N/d_S$  ratios for E1, W2 and W3 lineages (Figure 2). Some non-synonymous substitutions were identified under positive selection in 1 and 5 plastid genes and in 9 and 7 nuclear genes encoding the LSU and SSU respectively. In particular in the genes *rps3*, *rps5*, *rps13* and *rpl13* which are considered as essential (Tiller et al., 2012; Tiller and Bock, 2014) (Table 1; and see Tables S7 and S8 for details). Among the substitutions under positive selection for the nuclear genes, 66.7% and 71.4% were located in E1 and W2 lineages, for LSU and SSU, respectively (Tables S7 and S8). Finally, some substitutions either in plastid or in nuclear genes, were identified at residue contact sites, again especially in the genes encoding subunits of the plastid ribosomes (Table 1, Tables S7 and S8).

Nuclear genes of cytochrome b6/f complex, involved in the electron transport chain, also exhibited a large number of mutations, especially in E1 lineage and in *petM*, and substitutions were only found in E1 and W2 for the plastid genes encoding subunits of this complex (Figure 6B-C; Table S4). In *petM*, 7/11 non-synonymous substitutions were under positive selection and among them 5/7 were located in lineages E1 or W2 (Table S4). One substitution was identified at strongly conserved positions in the plastid gene *petA*, in the W2 lineage. Interestingly, *petM* is known as a direct interactor of *petA* (PDBe-KB).

## 4. Discussion and conclusion

### 4.1. Evolution of the plastid genome in *Silene nutans*

In angiosperms, plastid gene sequences are generally strongly conserved between species (Jansen et al., 2007), partly due to their low mutation rate (Drouin et al., 2008; Wolfe et al., 1987). In the present study, we showed that genetic lineages of *S. nutans* accumulated non-synonymous substitutions in plastid genes. A large number of mutations were located at strongly conserved positions. These mutations could thus potentially have an important functional impact, being either deleterious or advantageous. As only a small portion of them showed signatures of positive selection, the vast majority of the non-synonymous substitutions located at strongly conserved positions could be at least slightly deleterious. Most plastid complexes exhibited elevated  $d_N/d_S$  ratios compared to other *Silene* and angiosperm species, even when considering *Silene* species for which an acceleration

of the plastid genome evolution rate has been described (Sloan *et al.*, 2014). The similar pattern of  $d_N/d_S$  ratios found in *S. paradoxa*, although weaker, suggests that similar evolutionary mechanisms could be at stake on the common ancestor of *S. nutans* and *S. paradoxa*. High  $d_N/d_S$  values can be the result of positive selection or of relaxed purifying selection. For the plastid genes encoding large and small ribosomal subunits, intensification of selection has been identified for all four genetic lineages of *S. nutans*. These results suggest that intensification of positive selection has played a major role in the excess of non-synonymous substitutions observed for these plastid genes. For the genes encoding other plastid complexes, a tendency for relaxed selection was detected (especially for the plastid genes encoding subunits of the PSII, ATP synthase, cytochrome b6/f and RNA polymerase).

Either due to positive selection or relaxation of purifying selection, this pattern of mutation accumulation in the plastid genes suggests an accelerated evolutionary rate of the plastid genome in *S. nutans*. Various evolutionary forces could have driven such pattern. On the one hand, as the plastid genome is a non-recombining genome due to its uniparental mode of inheritance, it has a reduced effective size compared with the nuclear genome (Christie and Beekman, 2017; Greiner *et al.*, 2015). This reduces the efficacy of natural selection and increases the impact of genetic drift on plastid gene sequence evolution (Burton *et al.*, 2013; Greiner and Bock, 2013; Rand *et al.*, 2004). Plastid genome is thus expected to evolve under Muller's ratchet, i.e., a process resulting in an irreversible accumulation of deleterious mutations due to a lack of recombination (Rand *et al.*, 2004). Therefore, accumulation of non-synonymous substitutions in the plastid genes might be the result of strong genetic drift effects that may have acted when the lineages were isolated from each other in separate refugia in southern and eastern Europe during the last glacial maximum and when populations experienced repeated bottlenecks during postglacial recolonization (Martin *et al.*, 2016; Van Rossum *et al.*, 2018). This might have led to reduced effective sizes and relaxation of purifying selection, increasing the fixation of weakly deleterious mutations (Rockenbach *et al.*, 2016). Conversely, any advantageous mutation occurring within these lineages might also have led to this pattern through linked selection (Cruickshank and Hahn, 2014). Indeed, experimental evidences suggest that cytoplasmic involvement in adaptation has been underestimated (Budar and Roux, 2011). High values of  $d_N/d_S$  ratios were also found in reproductively isolated populations of *Campanulastrum americanum*. However they are lower than the values observed in *S. nutans* lineages (Barnard-Kubow *et al.*, 2014; Barnard-Kubow and Galloway, 2017). As observed in *S. nutans*,  $d_N/d_S$  values appeared to be particularly high in plastid genes encoding subunits involved in plastid gene expression machinery.

#### 4.2. Plastid-nuclear coevolution in *Silene nutans* lineages

Excess of non-synonymous substitutions in plastid genes of *S. nutans*, mainly due to relaxed selection, could impose strong selective pressures on the nuclear genes within lineages and within genomic complexes to maintain plastid-nuclear co-adaptation (Sloan et al., 2018). Evidence of coevolution between plastid and nuclear genes has been reported in several plant species (Ferreira de Carvalho et al., 2019; Moison et al., 2010; Rockenbach et al., 2016; Schmitz-Linneweber et al., 2005; Sharbrough et al., 2017; Sloan et al., 2014b; Weng et al., 2016; Williams et al., 2019; Zhang et al., 2015). Generally, these coevolution rely on parallel patterns of substitution accumulation between plastid and nuclear genes whose gene products are targeted to the plastid, with a high number of substitutions in one of the two genomic compartments, and signatures of positive selection on the other one, to ensure the maintenance of co-adaptation between mutated genes and the correct functioning of the plastid complex (Postel and Touzet, 2020). We identified parallel patterns of non-synonymous substitution accumulation between nuclear and plastid genes in *S. nutans*, within each plastid complex and lineage.

Disruption of co-adaptation between nuclear and plastid genes has been previously described in other plant systems and generally leads to impaired hybrid phenotypes exhibiting chlorosis and variegation and/or hybrid mortality, depending on the lineage used as maternal parent (reviewed in Postel and Touzet, 2020). Asymmetric postzygotic reproductive isolation has been identified between W1 and E1 lineages of *S. nutans* using controlled reciprocal crosses (Martin et al., 2017). In addition, diallelic crosses conducted between the four studied lineages of *S. nutans* showed a large proportion of chlorotic or partially chlorotic hybrid progeny, and high hybrid seedling mortality, especially when E1 and W2 lineages were the maternal parents (Van Rossum *et al.* 2018, unpublished results). Hybrid juvenile mortality rate after five weeks of growth was 0.97 and 0.83 when E1 and W2 was the maternal lineage, and 0.50 and 0.23 when W1 and W3 was the mother, respectively (Van Rossum *et al.*, unpublished results). These findings are consistent with our observations that E1 and W2 showed accumulation of substitutions in the plastid and nuclear genes and nuclear substitutions under positive selection, suggesting that these lineages contain the most functionally divergent and incompatible plastid genomes, causing PNIs in their hybrid progeny. Barnard-Kubow *et al.* (2017) reported a similar pattern for between-population hybrids of *C. americanum*, with higher hybrid mortality and chlorosis when populations with highly divergent plastid genomes were used as maternal lineages. In this system, PNIs have been put forward to explain reproductive isolation (Barnard-Kubow *et al.*, 2017; Barnard-Kubow and Galloway, 2017). Asymmetric reproductive isolation likely due to cytonuclear incompatibilities has also been described in *Oenothera* (Greiner *et al.*, 2011), *Pelargonium* (Metzlaf *et al.*, 1982; Weihe *et al.*, 2009) and *Pisum* (Bogdanova *et al.*, 2018, 2012; Bogdanova and Kosterin, 2007).

#### 4.3. Are Cytochrome b6/f and plastid ribosome possibly involved in PNI?

Disruption of co-adaptation between nuclear and plastid genes within a plastid complex leads to malfunction of this complex. This malfunction often has physiological consequences, visible at the phenotypic level, and so hybrid phenotypes can give insight into the genetic location of the PNIs (Massouh et al., 2016; Yao and Cohen, 2000). For example, when leaves are almost entirely white, dysfunction of plastid expression machinery is suspected, while pale yellow leaves indicate a deficiency in chlorophyll and malfunction of the photosynthetic process (Liebers et al., 2017; Massouh et al., 2016; Yao and Cohen, 2000). Hybrid progeny between *S. nutans* lineages suffering from chlorosis exhibit pale yellow cotyledons rather than white ones (Martin et al., 2017 ; Van Rossum *et al.* unpublished data), suggesting a defect in one or several photosynthetic plastid complexes rather than a dysfunction of the plastid gene expression machinery (Li et al., 2019; Liebers et al., 2017).

Regarding our results, one good candidate for explaining the (partially) chlorotic hybrid phenotypes caused by PNIs is cytochrome b6/f. This plastid complex is directly involved in the photosynthetic process as it manages the electron transport chain between PSI and PSII (Malone et al., 2019). We found non-synonymous substitutions on the plastid genes encoding cytochrome b6/f (*petA* and *petD*) as well as on nuclear genes (mostly on *petM*). In particular, *petA* and *petM* exhibited parallel patterns of non-synonymous substitution accumulation in E1 and W2 lineages and our RELAX results highlighted a pattern of coevolution, at least between *petM* and *petA* within the cytochrome b6/f (Table S4). These two genes are known to be in physical interaction (Forsythe et al., 2019; Malone et al., 2019). Disruption of co-adaptation between these two genes, in hybrids with E1 or W2 lineages as maternal parent, could lead to the observed pale-yellow phenotypes.

Strong coevolution pattern was also identified between genes encoding the plastid ribosome of *S. nutans*. Knock-out or dysfunction of one of these genes can lead to impaired, potentially non-viable, mutant phenotypes (Tiller and Bock, 2014). The mutations identified in ribosomal plastid genes could have important functional consequences and reinforce the need for compensatory evolution of the nuclear genes to maintain function of the plastid ribosome. Whether the plastid genes or the nuclear genes evolved in response to mutation accumulation in plastid or nuclear genomes remains unclear. Signature of positive selection, often considered as an indication of a compensatory dynamic to maintain co-adaptation (Greiner and Bock, 2013; Levin, 2003; Sloan et al., 2018), was observed in both genomic compartments. Tiller and Bock (2014) hypothesized that environmental changes affect translational activity of the plastid and generate a signal to modify leaf morphology and therefore photosynthesis performance. Accordingly, a recent study on *Brassica campestris* ssp. *pekinensis* identified a missense mutation in a ribosomal plastid gene (*rps4*) causing aberrant rRNA processing, which affected plastid translation and resulted in chlorophyll deficiency and reduced plant growth (Tang et al., 2018). Through their indirect role in photosynthesis, plastid ribosomal genes could be a

target for positive selection which could lead to the fixation of adaptive mutations. Besides, a phylogenomic study across angiosperms found that genes involved in protein homeostasis exhibited pattern of coevolution between plastid and nuclear genomes, suggesting recurrent bouts of selection on the plastid proteostasis systems, e.g. ribosomes (Forsythe et al., 2021). Further analyses based on the crystallographic structure of the plastid ribosome of *S. nutans* could give clues to the impact of each mutation on the protein-protein interactions network and on the whole structure of the plastid ribosome (Sharma et al., 2007).

The present study focused on coding sequences of genes encoding subunits of plastid complexes. Recent studies have shown that PNI could be also due to variations occurring in regulating sequences of plastid genes (Schmitz-Linneweber et al., 2005; Zupok et al., 2021). In fact, we found SNPs at 500 bp upstream of coding sequences (data not shown), but further analyses will be needed to survey gene expression. To do so, non-viable chlorotic hybrids should be rescued, possibility using media supplemented by sugar (Kühn et al., 2015). We acknowledge that some plastid genes that were excluded from our analyses could play a role in PNI (see Material and Methods), especially if the fact that we could not analyse them was due to their fast evolutionary rate. For example it was observed that CLP and ACCase complexes with fast evolving plastid genes exhibited a strong pattern of coevolution with their nuclear counterparts, that could result in PNI (Forsythe et al., 2021; Rockenbach et al., 2016; Williams et al., 2019).

#### 4.4. A link with gynodioecy?

The use of genomic and transcriptomic data enabled us to assess the evolutionary dynamics of the plastid and nuclear genomes of *S. nutans* lineages, relying on the necessary coevolution between these two genomes for proper cell functioning. Plastid genomes of *S. nutans* exhibited high non-synonymous genetic diversity. It has been suggested that the gynodioecious mating system of *S. nutans* could have favoured the speciation process (Martin et al., 2017). This reproductive system implies the presence of sterilizing mitochondrial genomes, leading to male sterility, and to male-fertility restoration by nuclear genes (Chase, 2007; Gouyon et al., 1991). Plastid and mitochondrial genomes are in tight linkage disequilibrium, and so common evolutionary mechanisms may act on plastid and mitochondrial genomes (Olson and Mccauley, 2002). The specific evolutionary dynamics of the mitochondrial genome in this gynodioecious species could have impacted and dragged the co-transmitted plastid genome. Indeed, the selection of male-sterile mitochondrial genomes, with male-sterile plants exhibiting a higher female fitness (Frank, 1989), could have produced a pattern of linked selection on the co-transmitted plastid genomes. This would favour the fixation of mildly deleterious mutations in the plastid genome, subsequently launching the process of coevolution in the nuclear genome.

### **Credit authorship contribution statement**

Zoé Postel: Investigation, Data curation, Formal analysis, Visualization, Writing - original draft. Céline Poux: Formal analysis, Methodology, Supervision, Writing – review and editing. Sophie Gallina: Conceptualization, Resources, Software. Jean-Stéphane Varré: Conceptualization. Cécile Godé: Resources. Eric Schmitt: Resources. Etienne Meyer: Formal analysis, Writing – review and editing. Fabienne Van Rossum: Resources, Writing - review and editing. Pascal Touzet: Conceptualization, Funding acquisition, Project administration, Supervision, Writing - original draft.

### **Declaration of Competing Interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### **Acknowledgments**

This research was funded by the Agence Nationale de la Recherche (ANR-11-BSV7-013-03, TRANS), the Région Hauts-de-France, and the Ministère de l'Enseignement Supérieur et de la Recherche (CPER Climibio), and the European Fund for Regional Economic Development, and the Région Hauts-de-France and the Ministère de l'Enseignement Supérieur et de la Recherche for ZP's PhD grant. We wish to thank Dan Sloan for discussion about transit peptide analyses.

### **Data accessibility**

All plastid reads of *S. nutans* lineages were deposited in NCBI (PRJNA745523). Scripts are available at <https://github.com/ZoePos/Variants-detectons>.

## **5. References**

- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. <https://doi.org/10.1089/cmb.2012.0021>
- Barnard-Kubow, K.B., Galloway, L.F., 2017. Variation in reproductive isolation across a species range. *Ecol. Evol.* 7, 9347–9357. <https://doi.org/10.1002/ece3.3400>
- Barnard-Kubow, K.B., McCoy, M.A., Galloway, L.F., 2017. Biparental chloroplast inheritance leads to rescue from cytonuclear incompatibility. *New Phytol.* 213, 1466–1476. <https://doi.org/10.1111/nph.14222>

- Barnard-Kubow, K.B., Sloan, D.B., Galloway, L.F., 2014. Correlation between sequence divergence and polymorphism reveals similar evolutionary mechanisms acting across multiple timescales in a rapidly evolving plastid genome. *BMC Evol. Biol.* 14, 1–10. <https://doi.org/10.1186/s12862-014-0268-y>
- Bateman, A., 2019. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* 47, D506–D515. <https://doi.org/10.1093/nar/gky1049>
- Bogdanova, V.S., 2020. Genetic and Molecular Genetic Basis of Nuclear-Plastid Incompatibilities. *Plants* 9, 1–17.
- Bogdanova, V.S., Galieva, E.R., Yadrikhinskiy, A.K., Kosterin, O.E., 2012. Inheritance and genetic mapping of two nuclear genes involved in nuclear-cytoplasmic incompatibility in peas (*Pisum sativum* L.). *Theor. Appl. Genet.* 124, 1503–1512. <https://doi.org/10.1007/s00122-012-1804-z>
- Bogdanova, V.S., Kosterin, O.E., 2007. Hybridization barrier between *Pisum fulvum* Sibth. et Smith and *P. sativum* L. is partly due to nuclear-chloroplast incompatibility. *Pisum Genet.* 39, 8–9.
- Bogdanova, V.S., Mglinets, A. V., Shatskaya, N. V., Kosterin, O.E., Solovyev, V.I., Vasiliev, G. V., 2018. Cryptic divergences in the genus *Pisum* L. (peas), as revealed by phylogenetic analysis of plastid genomes. *Mol. Phylogenet. Evol.* 129, 280–290. <https://doi.org/10.1016/j.ympev.2018.09.002>
- Budar, F., Roux, F., 2011. The role of organelle genomes in plant adaptation: time to get to work! *Plant Signal. Behav.* 6, 635–639. <https://doi.org/10.4161/psb.6.5.14524>
- Burton, R.S., Pereira, R.J., Barreto, F.S., 2013. Cytonuclear Genomic Interactions and Hybrid Breakdown. *Annu. Rev. Ecol. Evol. Syst.* 44, 281–302. <https://doi.org/10.1146/annurev-ecolsys-110512-135758>
- Chase, C.D., 2007. Cytoplasmic male sterility: a window to the world of plant mitochondrial-nuclear interactions. *Trends Genet.* 23, 81–90. <https://doi.org/10.1016/j.tig.2006.12.004>
- Christian, R.W., Hewitt, S.L., Roalson, E.H., Dhingra, A., 2020. Genome-Scale Characterization of Predicted Plastid-Targeted Proteomes in Higher Plants. *Sci. Rep.* 10, 1–22. <https://doi.org/10.1038/s41598-020-64670-5>
- Christie, J.R., Beekman, M., 2017. Uniparental inheritance promotes adaptive evolution in cytoplasmic genomes. *Mol. Biol. Evol.* 34, 677–691. <https://doi.org/10.1093/molbev/msw266>
- Cruickshank, T.E., Hahn, M.W., 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity , not reduced gene flow. *Mol. Ecol.* 23, 3133–3157. <https://doi.org/10.1111/mec.12796>
- Drouin, G., Daoud, H., Xia, J., 2008. Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. *Mol. Phylogenet. Evol.* 49, 827–831. <https://doi.org/10.1016/j.ympev.2008.09.009>
- Ferreira de Carvalho, J., Lucas, J., Deniot, G., Falentin, C., Filangi, O., Gilet, M., Legeai, F., Lode, M.,

- Morice, J., Trotoux, G., Aury, J.M., Barbe, V., Keller, J., Snowdon, R., He, Z., Denoeud, F., Wincker, P., Bancroft, I., Chèvre, A.M., Rousseau-Gueutin, M., 2019. Cytonuclear interactions remain stable during allopolyploid evolution despite repeated whole-genome duplications in *Brassica*. *Plant J.* 98, 434–447. <https://doi.org/10.1111/tpj.14228>
- Forsythe, E., Williams, A.M., Sloan, D., 2021. Genome-wide signatures of plastid-nuclear coevolution point to repeated perturbations of plastid proteostasis systems across angiosperms. *Plant Cell* 33, 980–997. <https://doi.org/10.1093/plcell/koab021>
- Forsythe, E.S., Sharbrough, J., Havird, J.C., Warren, J.M., Sloan, D.B., Chaw, S.M., 2019. CyMIRA: The Cytonuclear Molecular Interactions Reference for *Arabidopsis*. *Genome Biol. Evol.* 11, 2194–2202. <https://doi.org/10.1093/gbe/evz144>
- Frank, S.A., 1989. The evolutionary dynamics of cytoplasmic male sterility. *Am. Nat.* 133, 345–376.
- Gayral, P., Melo-Ferreira, J., Glémin, S., Bierne, N., Carneiro, M., Nabholz, B., Lourenco, J.M., Alves, P.C., Ballenghien, M., Faivre, N., Belkhir, K., Cahais, V., Loire, E., Bernard, A., Galtier, N., 2013. Reference-Free Population Genomics from Next- Generation Transcriptome Data and the Vertebrate – Invertebrate Gap. *PLoS Genet.* 9. <https://doi.org/10.1371/journal.pgen.1003457>
- Gouy, M., Guindon, S., Gascuel, O., 2010. Sea view version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* 27, 221–224. <https://doi.org/10.1093/molbev/msp259>
- Gouyon, P.H., Vichot, F., Van Damme, J.M.M., 1991. Nuclear-cytoplasmic male sterility: single-point equilibria versus limit cycles. *Am. Nat.* 137, 498–514. <https://doi.org/10.1086/285179>
- Greiner, S., Bock, R., 2013. Tuning a ménage à trois: Co-evolution and co-adaptation of nuclear and organellar genomes in plants. *BioEssays* 35, 354–365. <https://doi.org/10.1002/bies.201200137>
- Greiner, S., Rauwolf, U., Meurer, J., Herrmann, R.G., 2011. The role of plastids in plant speciation. *Mol. Ecol.* 20, 671–691. <https://doi.org/10.1111/j.1365-294X.2010.04984.x>
- Greiner, S., Sobanski, J., Bock, R., 2015. Why are most organelle genomes transmitted maternally? *BioEssays* 37, 80–94. <https://doi.org/10.1002/bies.201400110>
- Guindon, S., Dufayard, J., Lefort, V., 2010. Guindon et al. - 2010 - New Algorithms and Methods to Estimate Maximim-Likelihood Phylogenies Assessing the Performance of PhyML 3.0. *Syst. Biol.* 59, 307–321.
- Hooper, C.M., Castleden, I.R., Tanz, S.K., Aryamanesh, N., Millar, A.H., 2017. SUBA4: The interactive data analysis centre for *Arabidopsis* subcellular protein locations. *Nucleic Acids Res.* 45, D1064–D1074. <https://doi.org/10.1093/nar/gkw1041>
- Hubisz, M.J., Pollard, K.S., Siepel, A., 2011. PHAST and RPHAST: Phylogenetic analysis with space/time models. *Brief. Bioinform.* 12, 41–51. <https://doi.org/10.1093/bib/bbq072>
- Jafari, F., Zarre, S., Gholipour, A., Eggens, F., Rabeler, R.K., Oxelman, B., 2020. A new taxonomic

- backbone for the infrageneric classification of the species-rich genus *Silene* (Caryophyllaceae). Syst. phylogeny 00, 1–32. <https://doi.org/10.1002/tax.12230>
- Jansen, R.K., Cai, Z., Raubeson, L.A., Daniell, H., Depamphilis, C.W., Leebens-Mack, J., Müller, K.F., Guisinger-Bellian, M., Haberle, R.C., Hansen, A.K., Chumley, T.W., Lee, S.B., Peery, R., McNeal, J.R., Kuehl, J. V., Boore, J.L., 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. PNAS 104, 19369–19374. <https://doi.org/10.1073/pnas.0709121104>
- Juan, J., Armenteros, A., Salvatore, M., Emanuelsson, O., Winther, O., Heijne, G. Von, Elofsson, A., Nielsen, H., 2019. Detecting sequence signals in targeting peptides using deep learning. e201900429 2, 1–14. <https://doi.org/10.26508/lsa.201900429>
- Kühn, K., Obata, T., Feher, K., Bock, R., Fernie, A.R., Meyer, E.H., 2015. Complete Mitochondrial Complex I Deficiency Induces an Up-Regulation of Respiratory Fluxes That Is Abolished by Traces of Functional Complex I. Plant Physiol. 168, 1537–1549. <https://doi.org/10.1104/pp.15.00589>
- Lefort, V., Longueville, J., Gascuel, O., 2017. SMS : Smart Model Selection in PhyML. Mol. Biol. Evol. 34, 2422–2424. <https://doi.org/10.1093/molbev/msx149>
- Levin, D.A., 2003. The cytoplasmic factor in plant speciation. Syst. Bot. 28, 5–11. <https://doi.org/10.1043/0363-6445-28.1.5>
- Li, M., Hensel, G., Mascher, M., Melzer, M., Budhagatapalli, N., Rutten, T., Himmelbach, A., Beier, S., Korzun, V., Kumlehn, J., Borner, T., Stein, N., 2019. Leaf Variegation and Impaired Chloroplast Development Caused by a Truncated CCT Domain Gene in *albostrians* Barley. Plant Cell 31, 1430–1445. <https://doi.org/10.1105/tpc.19.00132>
- Liebers, M., Grübler, B., Chevalier, F., Lerbs-mache, S., Merendino, L., Blanvillain, R., Pfannschmidt, T., 2017. Regulatory Shifts in Plastid Transcription Play a Key Role in Morphological Conversions of Plastids during Plant Development. Front. Plant Sci. 8, 1–8. <https://doi.org/10.3389/fpls.2017.00023>
- Malone, L.A., Qian, P., Mayneord, G.E., Hitchcock, A., Farmer, D.A., Thompson, R.F., Swainsbury, D.J.K., Ranson, N.A., Hunter, C.N., Johnson, M.P., 2019. Cryo-EM structure of the spinach cytochrome b6f complex at 3.6 Å resolution. Nature 575, 535–539. <https://doi.org/10.1038/s41586-019-1746-6>
- Martin, H., 2016. Processus de spéciation et impact des systèmes de reproduction dans le genre *Silene*.
- Martin, H., Touzet, P., Dufay, M., Godé, C., Schmitt, E., Lahiani, E., Delph, L.F., Van Rossum, F., 2017. Lineages of *Silene nutans* developed rapid, strong, asymmetric postzygotic reproductive isolation in allopatry. Evolution 71, 1519–1531. <https://doi.org/10.1111/evo.13245>
- Martin, H., Touzet, P., Van Rossum, F., Delalande, D., Arnaud, J.F., 2016. Phylogeographic pattern of range expansion provides evidence for cryptic species lineages in *Silene nutans* in Western

- Europe. *Heredity* 116, 286–294. <https://doi.org/10.1038/hdy.2015.100>
- Massouh, A., Schubert, J., Yaneva-Roder, L., Ulbricht-Jones, E.S., Zupok, A., Johnson, M.T.J., Wright, S.I., Pellizzer, T., Sobanski, J., Bock, R., Greiner, S., 2016. Spontaneous chloroplast mutants mostly occur by replication slippage and show a biased pattern in the plastome of *Oenothera*. *Plant Cell* 28, 911–929. <https://doi.org/10.1105/tpc.15.00879>
- Matute, D.R., Cooper, B.S., 2021. Comparative studies on speciation : 30 years since Coyne and Orr. *Evolution* 75, 764–778. <https://doi.org/10.1111/evo.14181>
- Metzlaf, M., Pohlheim, F., Börner, T., Hagemann, R., 1982. Hybrid variegation in the genus *Pelargonium*. *Curr. Genet.* 5, 245–249. <https://doi.org/10.1007/BF00391813>
- Moison, M., Roux, F., Quadrado, M., Duval, R., Ekovich, M., Lê, D.H., Verzaux, M., Budar, F., 2010. Cytoplasmic phylogeny and evidence of cyto-nuclear co-adaptation in *Arabidopsis thaliana*. *Plant J.* 63, 728–738. <https://doi.org/10.1111/j.1365-313X.2010.04275.x>
- Muyle, A., Zemp, N., Gallina, S., Tavares, R., Silva, A., Bataillon, T., Widmer, A., Gl, S., Touzet, P., Marais, G.A.B., 2021. Dioecy Is Associated with High Genetic Diversity and Adaptation Rates in the Plant Genus *Silene*. *Mol. Biol. Evol.* 38, 805–818. <https://doi.org/10.1093/molbev/msaa229>
- Noé, L., Kucherov, G., 2005. YASS : enhancing the sensitivity of DNA similarity search. *Nucleic Acids Res.* 33, 540–543. <https://doi.org/10.1093/nar/gki478>
- Olson, M.S., Mccauley, D.E., 2002. Mitochondrial DNA diversity, population structure, and gender association in the gynodioecious plant *Silene vulgaris* . *Evolution* 56, 253–262.
- Oughtred, R., Stark, C., Breitreutz, B.J., Rust, J., Boucher, L., Chang, C., Kolas, N., O'Donnell, L., Leung, G., McAdam, R., Zhang, F., Dolma, S., Willems, A., Coulombe-Huntington, J., Chatr-Aryamontri, A., Dolinski, K., Tyers, M., 2019. The BioGRID interaction database: 2019 update. *Nucleic Acids Res.* 47, D529–D541. <https://doi.org/10.1093/nar/gky1079>
- Page, J.T., Liechty, Z.S., Huynh, M.D., Udall, J.A., 2014. BamBam: Genome sequence analysis tools for biologists. *BMC Res. Notes* 7, 1–5. <https://doi.org/10.1186/1756-0500-7-829>
- Postel, Z., Touzet, P., 2020. Cytonuclear genetic incompatibilities in plant speciation. *Plants* 9, 1–21. <https://doi.org/10.3390/plants9040487>
- Rand, D.M., Haney, R.A., Fry, A.J., 2004. Cytonuclear coevolution: The genomics of cooperation. *Trends Ecol. Evol.* 19, 645–653. <https://doi.org/10.1016/j.tree.2004.10.003>
- Ranwez, V., Harispe, S., Delsuc, F., Douzery, E.J.P., 2011. MACSE: Multiple Alignment of Coding SEquences accounting for frameshifts and stop codons. *PLoS One* 6, e22594. <https://doi.org/10.1371/journal.pone.0022594>
- Rockenbach, K., Havird, J.C., Grey Monroe, J., Triant, D.A., Taylor, D.R., Sloan, D.B., 2016. Positive selection in rapidly evolving plastid-nuclear enzyme complexes. *Genetics* 204, 1507–1522. <https://doi.org/10.1534/genetics.116.188268>

- Rozas, J., Ferrer-mata, A., S, J.C., Guirao-rico, S., Librado, P., Ramos-onsins, E., Alejandro, S.-G., 2017. DnaSP 6 : DNA Sequence Polymorphism Analysis of Large Data Sets. *Mol. Biol. Evol.* 34, 3299–3302. <https://doi.org/10.1093/molbev/msx248>
- Schmitz-Linneweber, C., Kushnir, S., Babiychuk, E., Poltnigg, P., Herrmann, R.G., Maier, R.M., 2005. Pigment deficiency in nightshade/tobacco cybrids is caused by the failure to edit the plastid ATPase  $\alpha$ -subunit mRNA. *Plant Cell* 17, 1815–1828. <https://doi.org/10.1105/tpc.105.032474>
- Sharbrough, J., Conover, J.L., Tate, J.A., Wendel, J.F., Sloan, D.B., 2017. Cytonuclear responses to genome doubling. *Am. J. Bot.* 104, 1277–1280. <https://doi.org/10.3732/ajb.1700293>
- Sharma, M.R., Wilson, D.N., Datta, P.P., Barat, C., Schluenzen, F., Fucini, P., Agrawal, R.K., 2007. Cryo-EM study of the spinach chloroplast ribosome reveals the structural and functional roles of plastid-specific ribosomal proteins. *PNAS* 104, 19315–19320. <https://doi.org/10.1073/pnas.0709856104>
- Siepel, A., Haussler, D., 2004. Phylogenetic Estimation of Context-Dependent Substitution Rates by Maximum Likelihood. *Mol. Gen. Genet.* 21, 468–488. <https://doi.org/10.1093/molbev/msh039>
- Sloan, D.B., Triant, D.A., Forrester, N.J., Bergner, L.M., Wu, M., Taylor, D.R., 2014a. A recurring syndrome of accelerated plastid genome evolution in the angiosperm tribe *Sileneae* (Caryophyllaceae). *Mol. Phylogenet. Evol.* 72, 82–89. <https://doi.org/10.1016/j.ympev.2013.12.004>
- Sloan, D.B., Triant, D.A., Wu, M., Taylor, D.R., 2014b. Cytonuclear interactions and relaxed selection accelerate sequence evolution in organelle ribosomes. *Mol. Biol. Evol.* 31, 673–682. <https://doi.org/10.1093/molbev/mst259>
- Sloan, D.B., Warren, J.M., Williams, A.M., Wu, Z., Abdel-Ghany, S.E., Chicco, A.J., Havird, J.C., 2018. Cytonuclear integration and co-evolution. *Nat. Rev. Genet.* 19, 635–648. <https://doi.org/10.1038/s41576-018-0035-9>
- Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., Jensen, L.J., Von Mering, C., 2019. STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47, D607–D613. <https://doi.org/10.1093/nar/gky1131>
- Tang, X., Wang, Y., Zhang, Y., Huang, S., Liu, Z., Fei, D., Feng, H., 2018. A missense mutation of plastid RPS4 is associated with chlorophyll deficiency in Chinese cabbage (*Brassica campestris ssp. pekinensis*). *BMC Plant Biol.* 18, 1–11. <https://doi.org/10.1186/s12870-018-1353-y>
- Tiller, N., Bock, R., 2014. The translational apparatus of plastids and its role in plant development. *Mol. Plant* 7, 1105–1120. <https://doi.org/10.1093/mp/ssu022>
- Tiller, N., Weingartner, M., Thiele, W., Maximova, E., Schöttler, M.A., Bock, R., 2012. The plastid-specific ribosomal proteins of *Arabidopsis thaliana* can be divided into non-essential proteins and

- genuine ribosomal proteins. *Plant J.* 69, 302–316. <https://doi.org/10.1111/j.1365-313X.2011.04791.x>
- Turelli, M., Moyle, L.C., 2007. Asymmetric postmating isolation: Darwin's corollary to Haldane's rule. *Genetics* 176, 1059–1088. <https://doi.org/10.1534/genetics.106.065979>
- Van Rossum, F., Martin, H., Le Cadre, S., Brachi, B., Christenhusz, M.J.M., Touzet, P., 2018. Phylogeography of a widely distributed species reveals a cryptic assemblage of distinct genetic lineages needing separate conservation strategies. *Perspect. Plant Ecol. Evol. Syst.* 35, 44–51. <https://doi.org/10.1016/j.ppees.2018.10.003>
- Varadi, M., Berrisford, J., Deshpande, M., Nair, S.S., Gutmanas, A., Armstrong, D., Pravda, L., Al-Lazikani, B., Anyango, S., Barton, G.J., Berka, K., Blundell, T., Borkakoti, N., Dana, J., Das, S., Dey, S., Di Micco, P., Fraternali, F., Gibson, T., Helmer-Citterich, M., Hoksza, D., Huang, L.C., Jain, R., Jubb, H., Kannas, C., Kannan, N., Koca, J., Krivak, R., Kumar, M., Levy, E.D., Madeira, F., Madhusudhan, M.S., Martell, H.J., MacGowan, S., McGreig, J.E., Mir, S., Mukhopadhyay, A., Parca, L., Paysan-Lafosse, T., Radusky, L., Ribeiro, A., Serrano, L., Sillitoe, I., Singh, G., Skoda, P., Svobodova, R., Tyzack, J., Valencia, A., Fernandez, E.V., Vranken, W., Wass, M., Thornton, J., Sternberg, M., Orengo, C., Velankar, S., 2020. PDBE-KB: A community-driven resource for structural and functional annotations. *Nucleic Acids Res.* 48, D344–D353. <https://doi.org/10.1093/nar/gkz853>
- Weihe, A., Apitz, J., Pohlheim, F., Salinas-Hartwig, A., Börner, T., 2009. Biparental inheritance of plastidial and mitochondrial DNA and hybrid variegation in *Pelargonium*. *Mol. Genet. Genomics* 282, 587–593. <https://doi.org/10.1007/s00438-009-0488-9>
- Weng, M.L., Ruhlman, T.A., Jansen, R.K., 2016. Plastid–nuclear interaction and accelerated coevolution in plastid ribosomal genes in Geraniaceae. *Genome Biol. Evol.* 8, 1824–1838. <https://doi.org/10.1093/gbe/evw115>
- Wertheim, J.O., Murrell, B., Smith, M.D., Pond, S.L.K., Scheffler, K., 2015. RELAX : Detecting Relaxed Selection in a Phylogenetic Framework. *Mol. Biol. Evol.* 32, 820–832. <https://doi.org/10.1093/molbev/msu400>
- Williams, A.M., Friso, G., J. van Wijk, K., Sloan, D.B., 2019. Extreme variation in rates of evolution in the plastid Clp protease complex. *Plant J.* 98, 243–259. <https://doi.org/10.1111/tpj.14208>
- Wolfe, K.H., Li, W.-H., Sharp, P.M., 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *PNAS* 84, 9054–9058.
- Yang, Z., 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. <https://doi.org/10.1093/molbev/msm088>
- Yao, J.L., Cohen, D., 2000. Multiple gene control of plastome-genome incompatibility and plastid DNA inheritance in interspecific hybrids of *Zantedeschia*. *Theor. Appl. Genet.* 101, 400–406.

<https://doi.org/10.1007/s001220051496>

Zhang, J., Ruhlman, T.A., Sabir, J., Blazier, J.C., Jansen, R.K., 2015. Coordinated rates of evolution between interacting plastid and nuclear genes in Geraniaceae. *Plant Cell* 27, 563–573.

<https://doi.org/10.1105/tpc.114.134353>

Zupok, A., Kozul, D., Schöttler, M.A., Niehörster, J., Garbsch, F., Liere, K., Fischer, A., Zoschke, R., Malinova, I., Bock, R., Greiner, S., 2021. A photosynthesis operon in the chloroplast genome drives speciation in evening primroses. *Plant Cell* koab155.

<https://doi.org/https://doi.org/10.1093/plcell/koab155>

## 6. Annexes

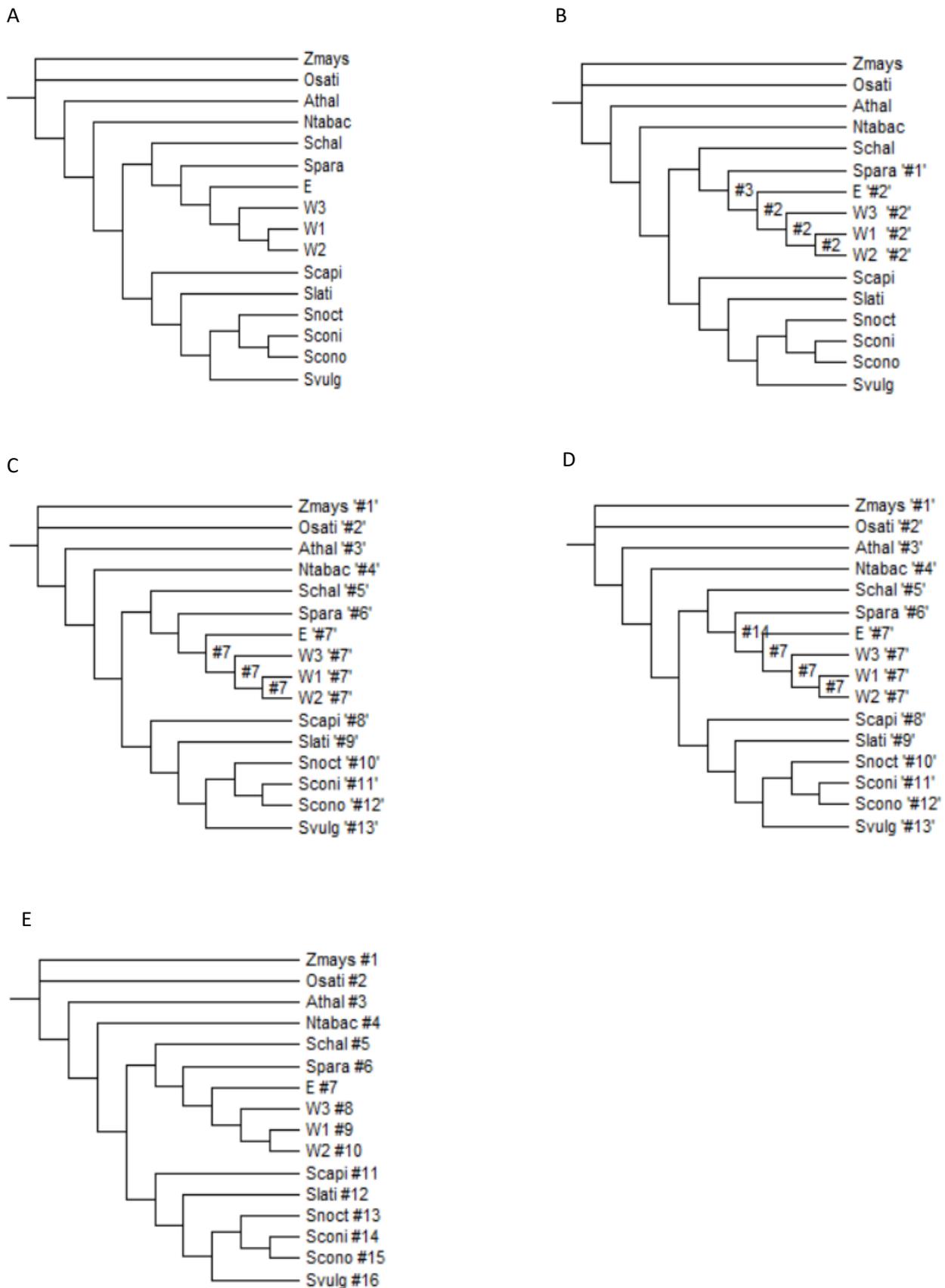
### Appendix S1 Search for potential editing sites

Eighteen nucleotide potentially edited nucleotide sites (T/C SNP in non-synonymous sites) were found to be located in 13 plastid genes (*rps19*, *rps15*, *rps11*, *rps18*, *rps2*, *rpl14*, *rpl16*, *rpl2*, *rpl32*, *rpl22*, *rpoB*, *matK*, and *ndhA*). The sequences of these genes were blasted against the assembly of the available transcriptome of *S. nutans*, with filter for percentage of identity of 98% and length of 170 pb (i.e. length of the smallest ORF). For the genes with a correct coverage of the length, the contigs of the RNAseq data were aligned with the DNA sequences of the individuals on SeaView v.4.7 and the nucleotide at the RNA level was manually determined, for each concerned position. As for 8 nucleotide sites blast was not concluding (i.e. no blast hits, poor coverage, and low pident), an additional method was used. The RNAseq reads of the two individuals used for the assembly of the transcriptome of *S. nutans* (Si18 and Si23 –E1 lineage) were aligned with the DNA sequences of the reference *S. paradoxa* using the program Bowtie2 v.2.3.5.1. Alignment of the reads of the two *S. nutans* individuals on *S. paradoxa* sequences were analysed on QualiMap v.2.2.1 (Okonechnikov *et al.*, 2016) and IGV (Thorvaldsdóttir *et al.*, 2013) to determine the presence of a C or a T at the read level for the position of interest.

### Appendix S2 – PAML – codeml models used on the different datasets

First, to compare the  $d_N/d_S$  ratio of *S. nutans* plastid genes with the different outgroup species, we ran five nested Branch-Models on the MSAs per plastid gene concatenations containing *S. nutans*, *Silene* and angiosperm nucleotide sequences. Branch-Models allowed the  $d_N/d_S$  ratio to vary along the branches of the phylogenetic tree (Yang and Nielsen, 1998). Each model differed according to the annotation of the phylogenetic tree (Figure S1). Codon frequencies were determined using F3 x 4 model. Transition/transversion ratio and omega were estimated with default setting of initial values 2 and 0.4, respectively. All other parameters were set as default. LRTs (likelihood ratio tests) were conducted to assess the best model between two nested models for each set of plastid concatenations, and the  $d_N/d_S$  values of the best model were kept (Yang, 1998; Yang and Nielsen, 2002).

Second, to search for traces of positive selection on the NS substitutions identified in plastid and nuclear genes, we used NSites models on MSA of the 72 plastid and 94 nuclear genes composed of *S. nutans*, *S. latifolia* and *S. paradoxa* nucleotide sequences. These models allowed the  $d_N/d_S$  to vary among sites and to detect positive selection on particular sites (Yang and Nielsen, 1998). We compared several nested models: neutral (M0), nearly neutral (M1) and positive selection (M2) according to Yang and Nielsen (2002). We conducted LRTs to assess the best model for each gene. When the LRT was significant, the result of the NEB (Naïve Empirical Bayes), which calculated posterior probabilities for site classes and identified the sites showing signs of positive selection, was taken into account (Yang, 2007). We reported the  $d_N/d_S$  values of the best NSites model for nuclear and plastid genes. We used the same parameters as above.



**Fig.S1 The different branch-models used to compare  $d_N/d_S$  ratio of *S. nutans* with other *Silene* and angiosperm species.** (A) M1 : common  $d_N/d_S$  for all species along the phylogenetic tree ; (B) M2 : variable  $d_N/d_S$  for *S. paradoxa*, the *S. nutans* lineages and the node leading to the divergence of *S. paradoxa* and *S. nutans* lineages ; (C) M3 : similar to M2 but allowing the  $d_N/d_S$  to vary for all species in the phylogenetic tree except for the four lineages of *S. nutans* for which only one  $d_N/d_S$  was estimated ; (D) M4 : allowed the  $d_N/d_S$  to vary for all species but without variation in  $d_N/d_S$  for the node leading to the differentiation of *S. paradoxa* and *S. nutans* lineages ; (E) M5 : similar to M4 but allowing the  $d_N/d_S$  to also vary between lineages of *S. nutans*.

**Table S1**

**List of the individuals for which we acquired plastid DNA and RNAseq data.** In orange: W1 lineage, red: W2 lineage, yellow: W3 lineage and in blue: E1 lineage (H. Martin and P. Touzet, unpublished data).

Lineage	Genome	Name	Ind in pool	Locality	Latitude	Longitude
W1	Plastid	F03	A4-7	Devèze, Auvergne, France	44.862	2.764
	Plastid	D01	AND-1	Les Andelys, Normandie, France	49.257	1.379
	Nuclear	Si25				
	Plastid	H03	AND-7			
	Nuclear	Si19				
	Plastid	C01	BZH1-1	Arzal, Morbihan, France	47.506	-2.404
	Nuclear	Si22				
	Plastid	A04	BZH1-4			
	Nuclear	Si17				
	Plastid	B01	CALV2-1	Thury-Harcourt, Normandie, France	48.997	0.489
	Plastid	B04	CALV2-11			
	Plastid	A07	DEV-1	Devèze, Cantal, France	44.862	2.764
	Plastid	C04	F1-1	Serre forest, Jura, France	47.167	5.557
	Plastid	D04	F1-10			
	Plastid	E04	OLL-210.4	Olloy-sur-Viroin, Belgium	50.069	4.606
	Nuclear	Si8				
Plastid	F04	OLL-C20				
Nuclear	Si9					
Plastid	A02	UK12-1	Gosport, South Hampshire, UK	50.789	-1.175	
Plastid	G04	UK12-10				
Plastid	H04	UK15-16	Dungeness, Kent, UK	50.933	0.959	
Nuclear	Si11					
W2	Plastid	B07	ARG-03	Argentine, Dordogne, France	45.472	0.379
	Plastid	A05	BAT-1	Castillon-la-Bataille, Gironde, France	44.865	-0.021
	Plastid	B05	BAT-10			
	Plastid	C05	BEN3-1	Saint Benoît, Poitou-Charente, France	46.542	0.338
	Plastid	D05	BEN3-10			
	Plastid	E05	D6-2	Bourdeilles, Dordogne, France	45.355	0.628
	Plastid	C07	PYR2-6.a	Aranvielle, Hautes Pyrénées, France	42.810	0.409
	Nuclear	Si24				
Plastid	H05	PYR2-9b				
Nuclear	Si26					
W3	Plastid	A06	AIG-1	Aiguebelle, Savoie, France	45.538	6.308
	Plastid	B06	AIG-10			
	Plastid	C06	D11-1	Stetten, Mülheim am Donau, Germany	48.017	8.861
	Plastid	D06	D11-10			
	Plastid	E06	F5-1	Les Granges, Jura, France	47.138	5.967
	Plastid	F06	F5-10			
	Plastid	G06	FQ3-6.2	Arvieux, Queyras, France	44.778	6.740
	Nuclear	Si15				
Plastid	H06	FQ3-7.2				
Nuclear	Si16					

<b>E1</b>	Plastid	H01	BUISA-12	Buis, Belgium	50.689	4.561
	Nuclear	Si21				
	Plastid	B02	BUISA-5			
	Nuclear	Si20				
	Plastid	G01	D19-15	Neckarzimmern, Germany	49.323	9.129
	Nuclear	Si2				
	Plastid	C02	D19-6			
	Nuclear	Si1				
	Plastid	D02	D2-1	Klein Rehberg, Germany	53.647	12.521
	Plastid	E02	D2-10			
	Plastid	F02	FIN1-3.1	Turku, Varsinais Suomi, Finland	60.458	22.257
	Nuclear	Si12				
	Plastid	F01	FIN1-4.1			
	Nuclear	Si13				
	Plastid	G02	LUX7-5	Brandenbourg, Oesling, Luxemburg	49.911	6.140
	Nuclear	Si4				
	Plastid	E01	LUX7-7			
	Nuclear	Si3				
	Plastid	H02	PLO-1	Plombières les Dijon, Bourgogne, France	47.332	4.936
	Plastid	A03	PLO-10			
Plastid	B03	UK16-10	Folkestone, Kent, UK	51.102	1.236	
Nuclear	Si18					
Plastid	C03	UK16-11				
Nuclear	Si23					
Nuclear	Si14	UK14-10	Littlehampton, West Sussex, UK	50.801	-0.558	
Plastid	D03	UK3-1	Buxton, Derbyshire, UK	53.252	-1.822	
Plastid	E03	UK3-10				

**Table S2**

Results of the YASS and the BLAST against *S. latifolia* genome assembly with mean values per lineage (E1, W1, W2 and W3).

		<b>E1</b>	<b>W1</b>	<b>W2</b>	<b>W3</b>
<b>YASS</b>	Number of Nodes	5.06	7.00	5.50	5.63
	% genome size	96.63	93.13	86.25	93.50
<b>BLAST</b>	Genes with blast hit	137	138	138	135
	% genes	91.56	91.27	91.38	91.13
	% gene size	84.75	81.13	83.88	86.63

Assemblies were described in two ways: (i) by conducting a YASS with *S. latifolia* as reference (Noé & Kucherov, 2005); and (ii) by blasting them against the sequences of *S. latifolia* collected on NCBI (which is composed of 149 plastid genes). We filtered the results with a percentage of identity (pident) of 90% and a total length of 35 pb. to get also the small *trn* genes.

**Table S3**

List of the 140 putative nuclear genes whose gene products were targeted to the plastid complexes.

Complexes	Name	TAIR identifier	Status
Photosystem I	<i>lhca6</i>	AT1G19150	Annotated (S)
	<i>psaG</i>	AT1G55670	Annotated (S)
	<i>psaE2</i>	AT2G20260	Annotated (2 S)
	<i>psaD2</i>	AT1G03130	Annotated
	<i>psaK</i>	AT1G30380	Annotated
	<i>lhca4</i>	AT3G47470	Annotated
	<i>lhca1</i>	AT3G54890	Annotated
	<i>psaO</i>	AT1G08380	Not variable
	<i>PYG7</i>	AT1G22700	Not variable
	<i>psaF</i>	AT1G31330	Not variable
	<i>psaH2</i>	AT1G52230	Not variable
	<i>lhca3</i>	AT1G61520	Not variable
	<i>lhca2</i>	AT3G61470	Not variable
	<i>psaN</i>	AT5G64040	Not variable
	<i>OP2</i>	AT1G34000	Not annotated
	<i>psaH1</i>	AT3G16140	Not annotated
	<i>psaD1</i>	AT4G02770	Not annotated
	<i>psaE1</i>	AT4G28750	Not annotated
	<i>LHCA5</i>	AT1G45474	No blast Spara
	<i>psaL</i>	AT4G12800	No blast Slati
Photosystem II	<i>ALB3</i>	AT2G28800	Annotated
	<i>DEGP1</i>	AT3G27925	Annotated
	<i>PPR2</i>	AT3G46610	Annotated
	<i>CRM3/PPR</i>	AT3G09650	Excluded
	<i>psbP1</i>	AT1G06680	Annotated (S)
	<i>lhcb6</i>	AT1G15820	Annotated (S)
	<i>LHCB4.2</i>	AT3G08940	Annotated (S)
	<i>psbY / ycf38</i>	AT1G67740	Annotated
	<i>lhcb2.1</i>	AT2G05100	Annotated
	<i>psbW</i>	AT2G30570	Annotated
	<i>lhb1B1</i>	AT2G34430	Annotated
	<i>psbO2</i>	AT3G50820	Annotated
	<i>psbQ2</i>	AT4G05180	Annotated
	<i>lhcb5</i>	AT4G10340	Annotated
	<i>psb27</i>	AT1G05385	Not variable
	<i>lhcb1.3</i>	AT1G29930	Not variable
	<i>psbR</i>	AT1G79040	Not variable
	<i>psbX</i>	AT2G06520	Not variable
	<i>lhcb4.3</i>	AT2G40100	Not variable
	<i>lhcb4.1</i>	AT5G01530	Not variable
	<i>LHCB1.1</i>	AT1G29910	Not annotated
	<i>LHCB1.2</i>	AT1G29920	Not annotated
	<i>lhcb2.2</i>	At2g05070	Not annotated
	<i>psbP2</i>	AT2G30790	Not annotated
	<i>Lhb1B2</i>	AT2G34420	Not annotated
	<i>psbTN</i>	AT3G21055	Not annotated
	<i>LHCB2.4</i>	At3g27690	Not annotated
	<i>psbQ1</i>	AT4G21280	Not annotated
	<i>psbO1</i>	AT5G66570	Not annotated

<b>Rubisco</b>	<i>RBCS1A</i>	AT1G67090	Annotated
	<i>RBCS3B</i>	AT5G38410	Not annotated
	<i>RBCS2B</i>	AT5G38420	Not annotated
	<i>RBCS1B</i>	AT5G38430	Not annotated
<b>Cytochrome b6/f</b>	<i>petM</i>	AT2G26500	Annotated
	<i>petC</i>	AT4G03280	Annotated
	<i>DAC</i>	AT3G17930	Annotated
	<i>PPR3</i>	AT5G42310	Annotated
<b>ATP synthase</b>	<i>atpG</i>	AT2G05620	Annotated
	<i>GL160</i>	AT2G31040	Annotated
	<i>PDE334</i>	AT4G32260	Annotated
	<i>atpC1</i>	AT4G04640	Not variable
	<i>ATPC2</i>	AT1G15700	Not annotated
	<i>atpD</i>	AT4G09650	Excluded
<b>N(A)DH</b>	<i>ndhS</i>	AT4G23890	Annotated
	<i>ndhV</i>	AT2G04039	Not variable
	<i>psnL1</i>	AT2G39470	Not variable
	<i>ndhM</i>	AT4G37925	Not variable
	<i>psnL5</i>	AT5G13120	Not variable
	<i>psnB5</i>	AT5G43750	Not variable
	<i>psnL2</i>	AT1G14150	Not annotated
	<i>PnsB2</i>	AT1G64770	Not annotated
	<i>ndhO</i>	AT1G74880	Not annotated
	<i>PNSL3</i>	AT3G01440	Not annotated
	<i>ndhT</i>	AT4G09350	Not annotated
	<i>PNSL4</i>	AT4G39710	Not annotated
	<i>ndhU</i>	AT5G21430	Not annotated
	<i>pnsB1</i>	AT1G15980	Excluded
	<i>pnsB4</i>	AT1G18730	Excluded
	<i>ndhL</i>	AT1G70760	Excluded
	<i>psnB3</i>	AT3G16250	Excluded
<i>ndhN</i>	AT5G58260	Excluded	
<b>RNA polymerase</b>	<i>PTAC12</i>	AT2G34640	Annotated
	<i>PTAC7</i>	AT5G24314	Annotated
	<i>GRF2</i>	AT1G78300	Annotated (S)
	<i>rpoT3</i>	AT2G24120	Annotated
	<i>hsp21</i>	AT4G27670	Annotated
	<i>MGP3</i>	AT1G68990	Not variable
	<i>patC3</i>	AT3G04260	Not variable
	<i>patC5</i>	AT4G13670	Not variable
	<i>AT5G15700</i>	AT5G15700	Not variable
<b>Large ribosomal subunit</b>	<i>rpl28</i>	AT2G33450	Annotated (S)
	<i>rpl10</i>	AT5G13510	Annotated (S)
	<i>rpl6</i>	AT1G05190	Annotated
	<i>rpl4</i>	AT1G07320	Annotated
	<i>rpl21</i>	AT1G35680	Annotated
	<i>EMB3105</i>	AT1G48350	Annotated
	<i>rpl31</i>	AT1G75350	Annotated
	<i>rpl13</i>	AT1G78630	Annotated
	<i>rpl3</i>	AT2G43030	Annotated
	<i>rpl18</i>	AT3G20230	Annotated
	<i>rpl15</i>	AT3G25920	Annotated
	<i>rpl17</i>	AT3G54210	Annotated
	<i>rpl1</i>	AT3G63490	Annotated

	<i>rpl5</i>	AT4G01310	Annotated
	<i>rpl19</i>	AT4G17560	Annotated
	<i>PRSP1</i>	AT5G24490	Annotated
	<i>rpl27</i>	AT5G40950	Annotated
	<i>rpl24</i>	AT5G54600	Annotated
	<i>rpl29</i>	AT5G65220	Annotated
	<i>rpl11.1</i>	AT1G32990	Not variable
	<i>rpl12.A</i>	AT3G27830	Not variable
	<i>rpl9</i>	AT3G44890	Not variable
	<i>PSRP6</i>	AT5G17870	Not variable
	<i>rpl34</i>	AT1G29070	Not annotated
	<i>PRSP4</i>	AT2G38140	Not annotated
	<i>rpl12B</i>	AT3G27840	Not annotated
	<i>rpl12C</i>	AT3G27850	Not annotated
	<i>PSRP2</i>	AT3G52150	Not annotated
	<i>PRSP5</i>	AT3G56910	Not annotated
	<i>rpl19</i>	AT5G47190	Not annotated
	<i>rpl11.2</i>	AT5G51610	Not annotated
	<i>rpl35</i>	AT2G24090	No blast
<b>Small ribosomal subunit</b>	<i>rps6</i>	AT1G64510	Annotated
	<i>rps9</i>	AT1G74970	Annotated
	<i>rps5</i>	AT2G33800	Annotated
	<i>rps10</i>	AT3G13120	Annotated
	<i>rps20</i>	AT3G15190	Annotated
	<i>rps21</i>	AT3G27160	Annotated
	<i>rps13</i>	AT5G14320	Annotated
	<i>rps1</i>	AT5G30510	Annotated
	<i>PSRP3</i>	AT1G68590	Not variable
	<i>rps17</i>	AT1G79850	Not variable
	<i>PRSP4</i>	AT2G38140	Not annotated
	<i>PSRP2</i>	AT3G52150	Not annotated
<b>Other</b>	<i>PDE338</i>	AT1G71720	Annotated
	<i>accA</i>	AT2G38040	Annotated
	<i>TRX</i>	AT3G06730	Annotated
	<i>UBQ3</i>	AT5G03240	Annotated
	<i>OHP2</i>	AT1G34000	Not annotated
	<i>PPR1</i>	AT4G31850	Excluded

S = genes containing only synonymous substitutions ; Annotated = the genes annotated and variable ; The genes highlighted in red are those found in the transcriptome assembly

**Table S4**

Characteristics of non-synonymous substitutions in the plastid and nuclear genes coding for interacting proteins in the cytochrome b6/f plastid complex.

Plastid Complex	Genes Name	Lineages	Nucleotide position	Amino-acid change?	Conservation Index	Residue contact site?	dN/dS	RELAX	Interactors	
Cytochrome b6/f	<i>petA</i>	E1	903	No	0.37	No	0.21	NS	<i>petM, petC</i>	
		W2	651	No	0.94	No				
	<i>petD</i>	W2	285	No	1.00	No	0.12	NS	<i>petc</i>	
	<i>petM_1</i>	E1		104	Yes	0.01	No	2.95	Intensification***	<i>petL, petG, petN, petA</i>
				232	Yes	0.00	No			
				263*	No	0.00	No			
				236	Yes	0.76	No			
			W2	205	Yes	0.00	No			
			W3	275*	No	0.01	No			
	<i>petM_2</i>	E1	W2	83*	Yes	x	NA	2.31	NS	
				255*	Yes	0.01	NA			
				314*	Yes	0.00	NA			
				323*	Yes	0.03	NA			
				324*	Yes	0.03	NA			
			W3 and E1	233*	Yes	0.00	NA			
	<i>DAC</i>	E1		68	Yes	0.01	NA	0.14	NS	Cytochrome b6F
		E1 - W3 / W1 - W2		156	No	0.00	NA			
			391	Yes	0.88	NA				

Genes in green = plastid genes – Genes in blue = nuclear genes; Nucleotide positions with a \* = non-synonymous substitutions under positive selection; RELAX P-value : 0.05 < \* ; 0.01 < \*\* , 0.001 < \*\*\* , NS = not significant

**Table S5**

dN/dS and RELAX results for the plastid genes containing non-synonymous substitutions.

Plastid complex	Plastid gene	dN/dS NSites	Pvalue RELAX	Kvalue RELAX	DoS RELAX
PHOTOSYSTEM I	<i>psaA</i>	0.03	0.35	0.57	Relaxation
PHOTOSYSTEM II	<i>psbB</i>	0.03	0.22	0.62	Relaxation
	<i>psbH</i>	1.00	0.99	1.00	Intensification
	<i>psbJ</i>	0.32	0.45	50.00	Intensification
CYTOCHROME B6/F	<i>petA</i>	0.21	0.12	0.00	Relaxation
	<i>petD</i>	0.12	0.69	0.35	Relaxation
ATP SYNTHASE	<i>atpA</i>	0.25	0.02	25.03	Intensification**
	<i>atpE</i>	0.17	0.73	1.36	Intensification
	<i>atpH</i>	0.30	0.21	7.95	Intensification
NDH	<i>ndhA</i>	3.09	0.05	1.62	Intensification*
	<i>ndhB</i>	0.31	0.75	0.40	Relaxation
	<i>ndhC</i>	0.08	0.06	0.00	Relaxation°
	<i>ndhD</i>	0.24	0.09	0.00	Relaxation°
	<i>ndhF</i>	0.30	0.14	0.14	Relaxation
	<i>ndhK</i>	0.13	0.12	50.00	Intensification
	<i>ndhH</i>	0.09	0.60	0.39	Relaxation
RNA POLYMERASE	<i>rpoB</i>	0.26	0.50	0.26	Relaxation
	<i>rpoC2</i>	0.37	0.24	0.22	Relaxation
SMALL RIBOSOMAL SUBUNIT	<i>rps11</i>	0.67	0.15	50.00	Intensification
	<i>rps15</i>	0.70	0.22	50.00	Intensification
	<i>rps18</i>	1.25	0.88	0.98	Relaxation
	<i>rps19</i>	0.72	0.68	50.00	Intensification
	<i>rps2</i>	0.53	0.34	39.07	Intensification
	<i>rps3</i>	1.34	0.14	0.15	Relaxation
	<i>rps7</i>	NA	0.90	1.07	Intensification
LARGE RIBOSOMAL SUBUNIT	<i>rpl14</i>	0.40	0.28	0.11	Relaxation
	<i>rpl16</i>	0.34	0.06	0.00	Relaxation°
	<i>rpl2</i>	0.73	0.82	0.62	Relaxation
	<i>rpl20</i>	0.71	0.22	0.00	Relaxation
	<i>rpl22</i>	0.85	0.07	6.13	Intensification°
	<i>rpl32</i>	0.42	0.02	0.00	Relaxation**
	<i>rpl33</i>	1.48	0.18	0.00	Relaxation
OTHER FUNCTIONS	<i>rpl36</i>	1.00	0.99	1.00	Intensification
	<i>ccsA</i>	0.66	0.15	5.45	Intensification
	<i>cemA</i>	0.43	0.62	0.24	Relaxation
	<i>matK</i>	0.55	0.06	0.38	Relaxation°
	<i>ycf4</i>	0.10	0.12	50.00	Intensification

P values: ° : &lt; 0.1 ; \* : &lt; 0.05 ; \*\* : &lt; 0.01 ; \*\*\* : &lt; 0.001

**Table S6 – dN/dS and RELAX results for the *S. nutans* nuclear genes containing non-synonymous substitutions.**

Kvalue\_RELAX is the selection intensity parameter, calculated by RELAX. When  $k > 1$ , it indicates that selection on test branches is intensified compared to reference branches. When  $k < 1$ , it indicates that selection on test branches is relaxed compared to reference branches. The DoS\_RELAX indicates the direction of selection depending on the K\_value calculated by RELAX: intensification if  $k > 1$  and relaxation if  $k < 1$ .

Plastid complex	TAIR symbol	dN/dS	Pvalue_RELAX	Kvalue_RELAX	DoS_RELAX
PSI	<i>ALB3</i>	0.30	0.49	1.24	NS
	<i>lhca1</i>	0.04	0.45	0.62	NS
	<i>lhca4</i>	0.11	0.17	2.31	NS
	<i>psaD2</i>	0.43	0.01	50.00	Intensification**
	<i>psaE2</i>	NA	0.87	0.95	NS
	<i>psaK</i>	0.01	0.14	11.69	NS
PSII	<i>DEGP1</i>	0.03	0.29	0.70	NS
	<i>lhb1B1</i>	0.07	0.68	1.27	NS
	<i>lhcb2.1</i>	0.06	0.50	0.66	NS
	<i>lhcb5</i>	0.11	0.35	0.75	NS
	<i>psbO2</i>	0.09	0.35	0.79	NS
	<i>psbQ2</i>	0.51	0.07	4.81	NS
	<i>psbW</i>	0.45	0.63	1.91	NS
	<i>PPR</i>	0.43	0.31	1.85	NS
RUBISCO	<i>rbc1A_1</i>	0.05	0.91	1.04	NS
	<i>rbc1A_2</i>	0.05	0.28	15.66	NS
ATP SYNTHASE	<i>atpG</i>	0.20	0.83	0.85	NS
CYTOCHROME B6/F	<i>petC</i>	0.07	0.93	0.95	NS
	<i>petM_1</i>	2.95	0.01	6.62	Intensification**
	<i>petM_2</i>	2.31	0.71	0.71	NS
	<i>DAC</i>	0.14	0.09	0	NS
NDH	<i>ndhS</i>	0.27	0.75	0.17	NS
RNA POLYMERASE	<i>HSP21</i>	0.13	0.00	5.01	Intensification***
	<i>ptaC12</i>	0.62	0.06	0.13	NS
	<i>ptaC7</i>	0.15	0.25	0.32	NS
	<i>rpoT3</i>	0.34	0.03	0.03	Relaxation*
LARGE RIBOSOMAL SUBUNIT	<i>emb2394</i>	0.37	0.63	0.67	NS
	<i>rpl1</i>	0.21	0.63	1.20	NS
	<i>rpl3</i>	0.10	0.15	0.57	NS
	<i>rpl5</i>	0.32	0.98	0.99	NS
	<i>rpl13</i>	0.82	0.10	2.84	NS
	<i>rpl15</i>	0.22	0.01	0.45	Relaxation**
	<i>rpl17</i>	0.02	0.43	0.42	NS
	<i>rpl18</i>	0.12	0.34	1.99	NS
	<i>rpl19</i>	0.21	0.62	0.65	NS
	<i>rpl21</i>	0.26	0.10	0.00	NS
	<i>rpl24</i>	0.07	0.78	0.89	NS

	<i>rpl27</i>	0.00	1.00	1.17	NS
	<i>rpl29</i>	0.50	0.11	2.03	NS
	<i>rpl31</i>	0.85	0.17	1.83	NS
<b>SMALL RIBOSOMAL SUBUNIT</b>	<i>rps1</i>	0.29	0.39	1.35	NS
	<i>rps5</i>	0.62	0.14	0.40	NS
	<i>rps6</i>	1.12	0.01	2.22	Intensification**
	<i>rps9</i>	0.09	0.11	0.42	NS
	<i>rps10</i>	0.36	0.62	1.34	NS
	<i>rps13</i>	0.70	0.16	1.53	NS
	<i>rps20</i>	0.52	0.12	7.24	NS
	<i>rps21</i>	0.75	0.08	0.48	NS
<b>OTHER</b>	<i>CPN60B2</i>	0.15	0.31	2.18	NS
	<i>PDE338</i>	0.56	0.11	0.38	NS

P-value: \* < 0.05; \*\* < 0.01; \*\*\* < 0.001; NS : not significant

**Table S7**

Details for the gene-pairs of the small ribosomal subunits.

Gene Name	Divergent Lineage	Nucleotide position	Changing nature of the encoded amino-acid ?	Conservation Index	NS located at residue contact?	dN/dS	RELAX	Interactors
<i>rps2</i>	W2	70	Yes	0.006	No	0.53	Intensification	<i>rps1, rps5</i>
<i>rps3</i>	W1	236	Yes	0.97	No	1.34	Relaxation	<i>rps5, rps10</i>
	W2	281	Yes	1	No			
		309	Yes	1	No			
	W3	348	No	0.98	No			
		636	No	0.99	No			
		348	No	0.98	No			
		636	No	0.99	No			
	W1-W3/W2-E1	372*	No	0.02	No			
<i>rps7</i>	W1-W3/W2-E1	357*	No	0	No	NA	Intensification	<i>rps9</i>
	W2	390*	No	0.01	No			
	E1-W3/W1-W2	453*	Yes	0.01	No			
<i>rps11</i>	E1	17	No	0.001	No	0.67	Intensification	<i>rps21</i>
		226	Yes	0	No			
		321	No	0.65	No			
		339	No	0.62	No			
	W2	232	Yes	0	No			
		244	Yes	0.05	No			
	W3	311	Yes	0.86	No - 103 <i>rps21</i>			
		345	No	0.11	<i>rps21</i>			
	E1-W3/W1-W2	291	Yes	0.99	No			
	E1 and W3	36	Yes	0	No			
	E1 and W3	37	Yes	0	No			
<i>rps18</i>	E1	49	No	0.99	No	1.25	Relaxation	<i>rps6, rps21</i>
	W1	237	Yes	0	<i>rps21</i>			
	W2	145	Yes	0.97	No			
		276	Yes	0	No			
	W3	38	Yes	NA	No			
		41	Yes	NA	No			
		262	No	NA	No			
		30	No	NA	No			
		32	No	NA	No			
		262	No	NA	No			
<i>rps19</i>	W3	57	Yes	0.01	No	0.72	Intensification	<i>rps13, rpl31</i>
		98	Yes	0.94	No			
		272	Yes	0.03	No			
	All	193*	Yes	0	<i>rps13</i>			
<i>rps1</i>	W2	66	No	0.99	No	0.37	Intensification	<i>rps2, rps8</i>

<u>rps2</u>	E1	106	No	0	<i>rps11</i>	0.51	Relaxation	<i>rps3, rps11, rps18</i>
		115	Yes	NA	No - 90 <i>rps11</i>			
		316	Yes	0	No			
	W1	241	Yes	0	No			
	W3	110	Yes	0.01	No - 90 <i>rps11</i>			
	E1	171	Yes	0.13	No	1.13	Intensification	
	W1	115	Yes	NA	No			
	W2	151	Yes	0.02	No – 120 <i>rps11</i>			
		254	Yes	0.84	No			
	E1	402	No	0.16	No	0.42	Relaxation	
	W2	137	Yes	0.19	No			
	E1-W3/W1-W2	206	No	0.04	No			
<u>rps5</u>	W2	547*	Yes	0.01	No	0.57	Intensification	<i>rps8, rps4, rps3, rps2</i>
	W3	200	Yes	0.05	No			
		392	No	0.8	No			
		625	Yes	0.42	No - 196 <i>rps3</i>			
	E1	590*	No	0	No			
<u>rps6</u>	E1	142	Yes	0.06	No	0.45	Intensification	<i>rps18</i>
		400	No	1	No			
		440	Yes	0.99	No			
	W3	425	Yes	0.84	RNA			
		E1-W3/W1-W2	166	Yes	0.01			
<u>rps9</u>	W3	440	No	0	<i>rps14</i>	0.3	Intensification	<i>rps3, rps14</i>
<u>rps10</u>		244	Yes	0	No			
<u>rps13</u>	W3	373*	Yes	0.08	<i>rps19</i>	0.37	Relaxation	<i>rps19</i>
<u>rps20</u>	E1-W3/W1-W2	313*	No	0.54	No	0.56	Intensification	No interactor identified
	W2	314*	No	0.01	interactor identified			
	E1-W3/W1-W2	315*	No	0.41				
	W1	419*	No	0.00				

Genes in green = plastid genes – Genes in blue = nuclear genes ; Underlined nuclear and plastid ribosomal genes represent essential plastid ribosomal genes as defined in (Tiller et al., 2012; Tiller and Bock, 2014). ; Nucleotide positions with a \* = NS substitutions under positive selection ; RELAX pvalue = 0,1 < ° ; 0,05 < \* ; 0,01 < \*\* , 0,001 < \*\*\*

**Table S8**

Details for the gene-pairs of the large ribosomal subunits.

Gene Name	Divergent Lineage	Nucleotide position	Changing nature of the encoded amino-acid ?	Conservation Index	NS located at residue contact?	dN/dS	RELAX	Interactors
<i>rpl14</i>	E1	144	Yes	0.22	No	0.4	Relaxation	<i>rpl19, rpl3</i>
		310	Yes	0	<i>rpl19</i>			
<i>rpl16</i>	W2	69	Yes	0.26	RNA	0.34	Relaxation	<i>rpl27</i>
		224	Yes	0	x			
		231	No	0.01	No			
<i>rpl22</i>	E1	279	Yes	0.99	RNA	0.85	Intensification°	<i>rpl13, rpl21</i>
		294	No	0.99	RNA			
	W1	16	Yes	0.96	No			
		276	Yes	0.99	No			
<i>rpl32</i>	W2	81	No	0.99	No	0.42	Relaxation**	<i>rpl17</i>
		65	Yes	0.99	No			
		144	No	0.99	No			
<i>rpl33</i>	W1-E1/W2-W3	9*	No	0.89	No	1.48	Relaxation	<i>rpl35</i>
<i>PSRP1</i>	W2 and W3	256*	No	0	No	0.58	Relaxation***	No interactor identified
	W2	319*	No	0	interactor identified			
<i>rpl3</i>	E1	95	Yes	0.03	No	0.24	Intensification	<i>rpl19, rpl13, rpl17, rpl14</i>
<i>rpl13</i>	W1 and W3	454*	Yes	0.39	No	0.6	Relaxation°	<i>rpl6, rpl20, rpl21, rpl22</i>
<i>rpl15</i>	W2 and W3	326*	No	0	No	0.28	Relaxation	<i>rpl14, rpl15</i>
<i>rpl17</i>	W2	61	No	0.08	No	0.90	Intensification	<i>rpl3, rpl32</i>
		163	No	0	No			
	E1 and W2	224*	Yes	0.35	No			
<i>rpl18</i>	W2	70*	No	0.19	No	0.04	Intensification	NA
<i>rpl19</i>	E	59	No	NA	No	0.23	Relaxation**	<i>rpl14, rpl3</i>
		702	Yes	0.02	No			
<i>rpl21</i>	E	68*	Yes	0	No	0.36	Relaxation	<i>rpl20, rpl6, rpl13, rpl22</i>
		182	No	0	No			
<i>rpl27</i>	E and W2	58	No	0	No	NA	Intensification*	<i>rpl16</i>
<i>rpl31</i>	W1	358*	No	0	<i>rpl5</i>	0.35	Intensification	<i>rps14, rps19, rps13, rpl5</i>
	W3	412*	No	0	No			

Genes in green = plastid genes – Genes in blue = nuclear genes ; Underlined nuclear and plastid ribosomal genes represent essential plastid ribosomal genes as defined in (Tiller et al., 2012; Tiller and Bock, 2014). ; Nucleotide positions with a \* = NS substitutions under positive selection ; RELAX pvalue = 0,1 < ° ; 0,05 < \* ; 0,01 < \*\* , 0,001 < \*\*\*





## SUB-CHAPTER 1

---

**What is the potential impact of genetic divergence of ribosomal genes between *Silene nutans* lineages in hybrids? An *in silico* approach.**

Theo Mauri<sup>1</sup>, Zoé Postel<sup>2</sup>, Andréa Bouanich<sup>12</sup>, Marion Liotier<sup>12</sup>, Zinara Lidamahasolo<sup>12</sup>, Pascal Touzet<sup>2</sup>, Marc Lensink<sup>1</sup>,

<sup>1</sup> CNRS UMR 8576-UGSF-Unité de Glycobiologie Structurale et Fonctionnelle, Université de Lille, 59000 Lille, France

<sup>2</sup> Univ. Lille, CNRS, UMR 8198 - Evo-Eco-Paleo, F-59000 Lille, France



Having identified potential candidates for plastid-nuclear incompatibilities in chapter 1, especially in the plastid ribosomal genes, we tried to further narrow the range of potential plastid-nuclear pairs. We used the 3D structure of the plastid ribosome of *Spinacia oleracea* available online in Protein-DataBase Knowledge to simulate the identified mutations in plastid and nuclear genes of the plastid ribosome. The idea was to try to assess the structural impact of these mutations on the interactions between plastid and nuclear genes within the plastid ribosome.

This sub-chapter was done in collaboration with Théo Mauri (PhD student) and Marc Lensink (CNRS), his supervisor. We also co-supervised a group of master students (Andréa Bouanich, Marion Liotier, Zinara Lidamahasolo) in bioinformatics who greatly helped us with the first analysis (results in table 2). Pascal Touzet and I formulated the biological questions about the plastid-nuclear interactions in lineages of *S. nutans*. Théo Mauri, with the help of Marc Lensink, conducted the analyses, made the figures and wrote the manuscript (Material and methods + results and discussion). I and Pascal Touzet participated to the interpretation of the results and writing of the manuscript (introduction + results and discussion).

This is a first draft.



## 1. Introduction

Plastids are ancient cyanobacteria that integrated the eukaryotic cells as endosymbionts roughly a billion years ago (1). After this integration, this organelle transferred a certain amount of its genes to the nucleus, ending up encoding only a few of the original gene set (2). These remaining 120 or so genes are involved in photosynthesis and housekeeping function in the plastid (3). Due to these transfers, the essential plastid protein complexes are encoded both by plastid and nuclear genes whose gene products are targeted to the plastid (later called ptNu). Plastid and ptNu genes need to interact with one another for correct protein complex function (4–6). Nuclear and plastid genomes have contrasting features, such as differences in mutation rate that is much lower in the plastid (7) or different inheritance patterns with biparental inheritance for the nuclear genome and maternal one for the plastid (8). As so, any mutation occurring in one of the two partners will generate strong selective pressure for fixation of compensatory mutation in the other one (4). Tight co-adaptation between interacting plastid and nuclear genes are then required and enforced (9). Independent accumulation of mutations in both plastid and nuclear genes can occur in isolated lineages or populations (10). If hybridization occurs between these isolated lineages, co-adaptation between nuclear and plastid genes will be disrupted in hybrids (2). Indeed, hybridization will bring together a plastid genome mismatched with a part of the hybrid nuclear background, leading to potential hybrid breakdown (i.e. decrease in fertility and survival) through creation of plastid-nuclear incompatibilities (PNIs) (11). These incompatibilities are thought to be part of the first post-zygotic reproductive barrier to emerge as they can lead to reproductive isolation between lineages through decrease in hybrid fitness (12). When such incompatibilities are involved in speciation (i.e. the process leading to reproductive isolation (13)), reproductive isolation is asymmetric in reciprocal crosses, depending on the lineage that is the plastid donor (14,15). Molecular mechanisms and identification of co-adapted pairs of genes is still largely missing, even though some studies have identified PNIs likely involved in reproductive isolation (16–19). Contrastingly, patterns of co-evolution between plastid and nuclear genes have been extensively studied especially in plant species exhibiting accelerated rates of plastid genome evolution (20–24).

PNIs were also potentially involved in reproductive isolation between lineages of *Silene nutans* (Caryophyllaceae) (25). This species is composed of several genetically differentiated lineages in France, based on plastid sequences and nuclear microsatellite markers and their geographic distribution in Europe reflects colonization from past glacial refugia (26,27). Diallelic crosses between four of these lineages, an eastern (E1) and three western (W1, W2, W3) revealed strong and asymmetric reproductive isolation between them (28)(Van Rossum et al., in prep). Analysis of plastid

genetic diversity and ptNu genes in these four lineages uncovered lineage specific coevolution patterns between plastid and nuclear genes that could result in PNIs in hybrids (25). Candidate gene pairs for PNIs were identified in the plastid ribosomes (25), a plastid complex composed of a large and a small subunit and encoded both by nuclear and plastid genes (29). Plastid and ptNu genes encoding this complex exhibited the largest amount of lineage specific non-synonymous (NS) mutations (i.e. mutation leading to a change of the encoded amino-acid) and elevated  $d_N/d_S$  (i.e. proportion of non-synonymous (N) and synonymous (S) mutations on the total number of N and S sites) (25). Elevated  $d_N/d_S$  was thought to be the result of positive selection on the plastid genes and on some nuclear genes (25). Regarding the nuclear genes,  $d_N/d_S$  was significantly higher compared to nuclear genes encoding the cytosolic ribosome (i.e. gene products not targeted to the plastid), suggesting this increase in number of NS mutations might be the result of plastid-nuclear coevolution (25). Some of the NS mutations identified in plastid and nuclear genes encoding the large and small plastid ribosomal subunit were directly located at protein residue contact position, suggesting structurally mediated coevolution between plastid and nuclear genes within the plastid ribosome (25,30).

Disruption of co-adaptation in the plastid ribosome can have dramatic consequences, especially if it concerns essential plastid ribosomal genes (31,32). For example, a missense mutation in the essential plastid gene *rps4* causes defaults in plant development and photosynthetic performances in *Brassica campestris* ssp. *pekinensis* (33). More generally, any mutation in plastid or ptNu essential ribosomal genes can have dramatic impact on photosynthesis, as all of the plastid-encoded photosynthetic genes are translated by the plastid ribosome (32). Many plastid-nuclear gene pairs encoding subunits of the plastid ribosome were identified as potential candidates for PNIs between lineages of *S. nutans* (25). To further identify which of these pairs could be responsible for PNIs, we used the crystallographic structure of the spinach plastid ribosome (34) to assess the potential impact of the NS mutations identified in each plastid and ptNu genes of these pairs on the residue contact interactions between plastid and nuclear proteins within the large and small plastid ribosomal subunit. To do this, we modeled the different NS mutations for each lineage and each nuclear and plastid candidate proteins on these subunits to further narrow down the list of PNIs candidates in the plastid ribosome. Models were then transformed into graphs called Residue Interactions Networks (RINs) and from these networks centralities of the residues were calculated (35). Centrality of a residue represents its amount of interactions with other residues. The centralities of the modified residue contact interactions can be used to see if the mutations of interest may have a role in modifying a central residue contact interaction between plastid/nuclear gene pairs and potentially disrupt plastid ribosome structure, resulting in PNIs in inter-lineages hybrids. This method has been shown to highlight residues important for protein structure and function (del Sol *et al.*, 2006; Hu *et al.*, 2014; Trouvilliez *et al.*, 2022) (36–38).

## 2. Material and methods

### 2.1. Identification of mutations

Mutation identification was previously done in (25). Briefly, we searched for all mutations differently fixed between lineages of *S. nutans*, in plastid and nuclear gene sequence alignments of the plastid ribosome, using an in-house biopython script (<https://github.com/ZoePos/Variants-detectations>) (25). We then aligned the plastid and nuclear gene sequences of *S. nutans* with the one of *S. oleracea*, used as reference. The spinach structure and the associated protein sequences were available in PDBe (European Protein Data Bank) (PDB id: 5MMM) and contains 60 chains corresponding to the different rps and rpl subunits and some RNA strands (39). After aligning *S. nutans* and *S. oleracea* sequences, we compared the encoded amino-acid between *S. nutans* and the spinach, at each position containing mutations between lineages of *S. nutans*. We reported the different mutations identified in lineages of *S. nutans* and the corresponding amino acid in the *S. oleracea* in Tables S1 and S2.

### 2.2. Identification of impacted interactions between subunits

To identify interactions between subunits inside the plastid ribosome, the structure was transformed into a RIN. The RIN is a graphical representation of the structure where nodes represent the residues and the edges the interactions between residues. To define an interaction, one atom of a residue A must be at a distance between 2.5 and 5 Ångström (Å) of one atom of the residue B. Detected interactions are then exported into a text file with the two amino acids involved and the minimal distance between these two. From this file, only interactions between plastid and nuclear genes within each of the large and small ribosomal subunits were analyzed to identify potential impacting mutations.

### 2.3. Identification of the type of interaction and potential modifications

To identify the type of interactions and the potential impact of the mutations on the interaction, mutations were modeled based on the spinach reference structure. To do that, the PyMOL software with the mutagenic tool was used (Delano *et al.*, 2002; Schrödinger *et al.* 2020) (40,41). This tool can replace an amino acid by another one by transforming the lateral chain. From this, an atomic point of view of the interaction can be deduced and the different types of interaction determined. The type of interaction can be one of the following: Hydrogen bond, hydrophobic, salt bridge and polar. The

mutation can lead to a change in the type of interaction, a creation of a new one or a loss of the interaction. The type of interaction has been determined manually on the PyMOL interface. Given the results, we chose to focus on one plastid-nuclear gene pair: *rps11* (plastid encoded) - *rps21* (nuclear encoded) (cf Results).

#### *2.4. Creation of the different models*

As there are 4 different lineages (E1, W1, W2 and W3), we created 16 different models called E1\_E1, E1\_W1, E1\_W2...W3\_W2 and W3\_W3 (*i.e.* one model per cross type and direction). Each model contained the associated mutations on the genes *rps11* and *rps21* described in table S2. The models were minimized using the YASARA software with YASARA minimization (Land *et al.* 2018) (42) (Figure 1).

#### *2.5. Creation of the Residue Interaction Networks (RINs) from the models and centrality analyses*

RINs were created for each model for a total of 16 RINs using ringraph, an in-house C program which calculates distances between amino acids as described above (Figure 1). From these networks, it is possible to calculate centralities of nodes thanks to graph theory. Centrality of a residue will represent residue that connect other residues together within a protein network (here within protein RPS11, RPS21 and their interactions). The more a residue contributes to residue connection within a structure, the more it is central and has an important structural role. Different types of centrality can be calculated by the same in-house program. In the present study, we calculated the five centralities listed in table 1. We then calculated a centrality score, (*i.e.* a Z-score) which is normalized with the size of the network. A residue with a high Z-score ( $\geq 2$ ) is considered as central. The results of all centralities for the 16 models were retrieved and imported into a csv file. To have an overall comprehension, the ten more central residues in the interaction network of *rps11* and *rps21* were represented and analysed (Figures 1 & 3).

#### *2.6. Principal Component Analysis (PCA) on centrality measures*

To see if modification of residue centralities associated with the cross type can explain the different outcomes of inter-lineages crosses and explain the differences in hybrids mortality, PCA were conducted using the centralities values of *rps11-rps21* residues for each cross type and the five different measure of centrality. Results being similar for the five centrality measures, we only reported results of the degree centrality measure. PCA has been run through an R script (R version 3.6.3) with



A



B

**Fig.1 Representation of the *rps11-rps21* genes as a structure (A) and as a RIN (B).** The blue chain is *rps21* and the green chain is *rps11*. The yellow residue corresponds to the yellow node in the networks and corresponds to a residue with a centrality  $\geq 2$ . Visualization of the network has been made with Cytoscape after running RINspector (Shannon *et al.*, 2003; Brysbaert *et al.*, 2018) (44,45).

**Table 1**

Summary of the types of centrality calculations used and their method.

CENTRALITY MEASURE	METHOD
<b>Betweenness Centrality Analysis (BCA)</b>	BCA highlights residues often found in the minimal path between every residue.
<b>Closeness Centrality Analysis (CCA)</b>	CCA is calculated as the reciprocal of the sum of the length of the shortest paths between the node and all other nodes in the graph
<b>Degree Centrality Analysis (DCA)</b>	DCA calculates centrality based on the number of nodes connected to the residue analysed.
<b>Eigenvector Centrality Analysis (ECA)</b>	ECA calculates the centrality of nodes based on the centrality of other nodes meaning that a node connected to a high centrality node will have a higher centrality .
<b>PageRank centrality Analysis (PRA)</b>	PCA was first an algorithm developed for Google and its output is a probability distribution used to represent the likelihood that a person randomly clicking on links will arrive at any particular page

RStudio and R packages (table.data V1.2.0 for data analysis and factoextra V1.0.7 for representation). PCA was calculated with the “prcomp” command (43).

### 3. Results and discussion

#### 3.1. Modification of the interactions due to the identified mutations

Lots of lineage-specific NS mutations have been identified in interacting plastid and nuclear genes encoding the plastid ribosomes within *S. nutans* lineages (25). Mutation selection leads to a subset of 28 mutations with the associated modified interaction (Table 2). In total, we observed 8 losses of interaction, 4 gains of interaction and 16 mutations without a change in interaction type (Table 2). Some of these changes of interaction might be responsible for inter-lineage hybrid breakdown through disruption of co-adaptation between these plastid-nuclear gene pairs within the subunits of the plastid ribosome. Indeed, these modifications of residue contact interaction between plastid and nuclear genes within the plastid ribosome could alter the whole structure of the plastid ribosome. A large majority of the mutations inducing a change of the interaction between plastid and nuclear genes were located on genes *rps11* (plastid encoded) and *rps21* (nuclear encoded). This gene pair is also the one that contained most of the mutations (i.e. 28 in total) (table 2, table S2). We focused on this gene pair in subsequent analyses. For each *S. nutans* lineage cross direction, we build a three-dimensional model based on the spinach ribosome structure complex resolved in 2007. In total, 16 models have been created with the associated lineage mutation then minimized.

#### 3.2. RINs analysis of mutations on the *rps11-rps21* gene pair

We looked at the centralities of the mutated residues in *rps11-rps21* genes for the lineages of *S. nutans* (E1, W1, W2 and W3) to see if the mutations could impact the stability of the different subunits of the ribosome. The mutations were simulated from a visualization tool called PyMOL with a mutagenesis tool allowing to change the lateral chain of amino acids. Since these *in silico* mutations can change the conformation of the complex or at least the interface area, we performed energy minimization of the models, this allows us to have more realistic models but these are still predictive models which can add a bias to our analyses. To be sure we did not miss information, we decided to calculate 5 different kinds of accessibility: betweenness, closeness, degree, eigenvector and PageRank. We looked at the difference of centrality for the mutated residues inducing a change of the interaction between *rps11* and *rps21* (i.e. residue marked with a \* in table 2). For each residue and each centrality measure, we retrieved a Z-score and looked at the difference of this score according to the different lineage cross

**Table 2**  
Detail of the interactions found between candidate gene pairs, with the impact of the mutation in one of the two partners on the interaction.

Interaction N°	Partner 1			Partner 2			Distance (Å)	Interaction type	
	Gene	Residue	Amino acid of E1 W1 W2 W3	Gene	Residue	Amino acid of E1 W1 W2 W3		Before mutation	After mutation
1	<i>rpl32</i>	Arg49	LLPL	<i>rpl17</i>	Tyr122	x	4.37	H bridge	∅
2	<i>rpl14</i>	Arg104	G___	<i>rpl19</i>	Val155	x	4.17	∅	∅
3					Glu157	x	4.85	∅	∅
4					Ser163	x	2.98	polar	∅
5					Tyr165	x	2.99	polar	∅
6	<i>rps3</i>	Lys146	x	<i>rps5</i>	Val198	___I	3.86	∅	∅
7	<i>rps11</i>	Pro98	_SS_	<i>rps21</i>	Ile113	x	3.47	∅	∅
8					Cys116	x	3.86	∅	∅
9		Leu116*	___V		Val90*	x	3.34	∅	hydrophobic
10					Glu94	x	3.76	∅	∅
11					Leu99	x	3.53	hydrophobic	hydrophobic
12		Ser117	x		Val88	I___	4.57	polar	polar
13					Leu89	F__S	3.41	∅	∅
14		Phe118*	x		Val88*	I___	3.73	∅	hydrophobic
15					Leu89*	F__S	3.44	hydrophobic	hydrophobic/∅
16		Val119	x		Val88	I___	3.03	polar	polar
17		Pro132	x		Tyr121*	HH_H	3.67	∅	∅
18		Pro133	x				3.49	∅	∅
19		Lys134*	x				3.51	∅	polar
20		Lys135*	x				3.79	∅	H bridge
21		Lys135	x		Glu127	_DDD	3.12	salt bridge	salt bridge
22	<i>rps11</i>	Arg136	x		Tyr121	HH_H	3.30	H bridge	H bridge
23	<i>rps13</i>	Arg124	x	<i>rps19</i>	Arg65	_HYD	2.64	H bridge	H bridge
24		Glu127	___Q				2.75	salt bridge	∅
25		Ile128	x				4.50	H bridge	∅
26	<i>rps18</i>	Arg50	__Q_	<i>rps21</i>	Arg139	x	3.60	H bridge	∅
27					Asn140	x	4.12	H bridge	H bridge
28					Arg143	x	3.36	H bridge	∅

In dark blue: no change of the type of the interaction with the mutation ; In light blue: loss of the interaction with the mutation ; In pink: creation of a new interaction with the mutation ; \* : residue of *rps11-rps21* gene pairs for which centrality was calculated ; "x" : no mutations

type (Figure 2). The ten more central residues in the interaction network of *rps11* and *rps21* (i.e. not only the mutated ones) are represented in figure 3.

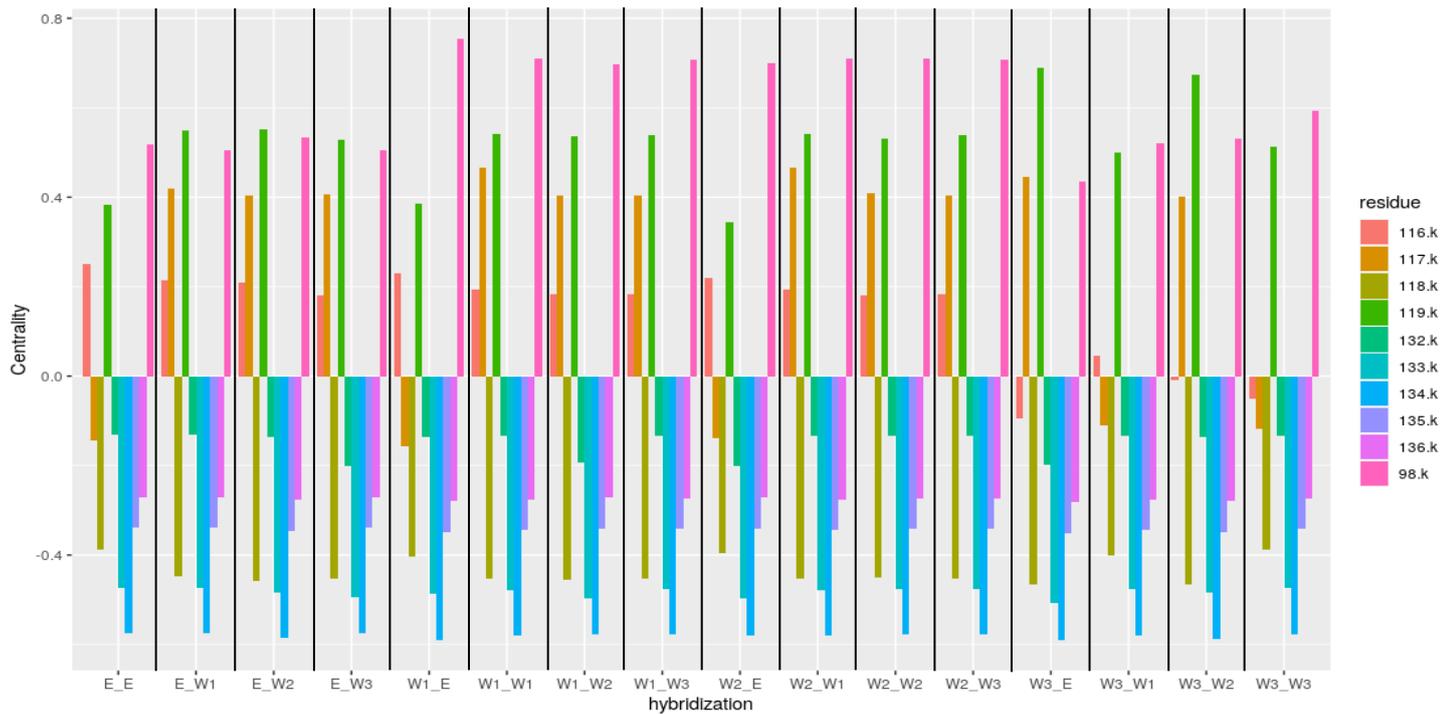
In most of the cases, the lineage which is the most impacting in terms of residue centrality is the lineage E1. Most of the time when the lineage E1 is involved in a cross, we can see a variation of the centralities of the residues notably with a change of centrality of the residue 117 (serine) (figure 2). When W3 is the maternal parent, we can see that the cysteine 116 loses its centrality (figure 2). We can also see with these results that the crosses between the lineages E1 and W3 impacted the centrality of the *rps11* alanine 34 (figure 3). Centrality of the residue Arginine 118 of *rps11* was also modified for the cross between lineages E1\_W2, W1\_E1, W3\_E1 and W3\_W2 (figure 3). For the same crosses, the residues glutamine 114 and acid glutamic 135 of the *rps21* exhibited a decrease in centrality (figure 3). Cross-specific modification of centrality is residue of genes *rps11* and *rps21* were identified. Especially when lineage E1 is involved and in plastid gene *rps11*. This lineage is also the one resulting in highest percentage of hybrid mortality. Though *rps21* does not seem to be an essential gene in the function of the plastid ribosome, *rps11* is (32). Modification of residues centralities in this essential gene in cross with lineage E1 might contribute to a modification of the protein network interaction and to an elevated amount of hybrid

mortality. We also observed differences in centrality when lineage W3 is the mother, again in gene *rps11*. Though W3 does not lead to high percentage of hybrid mortality when used in inter-lineages crosses, it could nevertheless impact the whole small subunit ribosomal structure and impact its function. Overall, the different centrality calculations showed a loss or a gain of centrality according to the different mutations and crosses, but there is no clear signal. Especially, this does not seem to follow the associated amount of hybrid mortality. One hypothesis to explain that might be that the centrality moves to another residue not

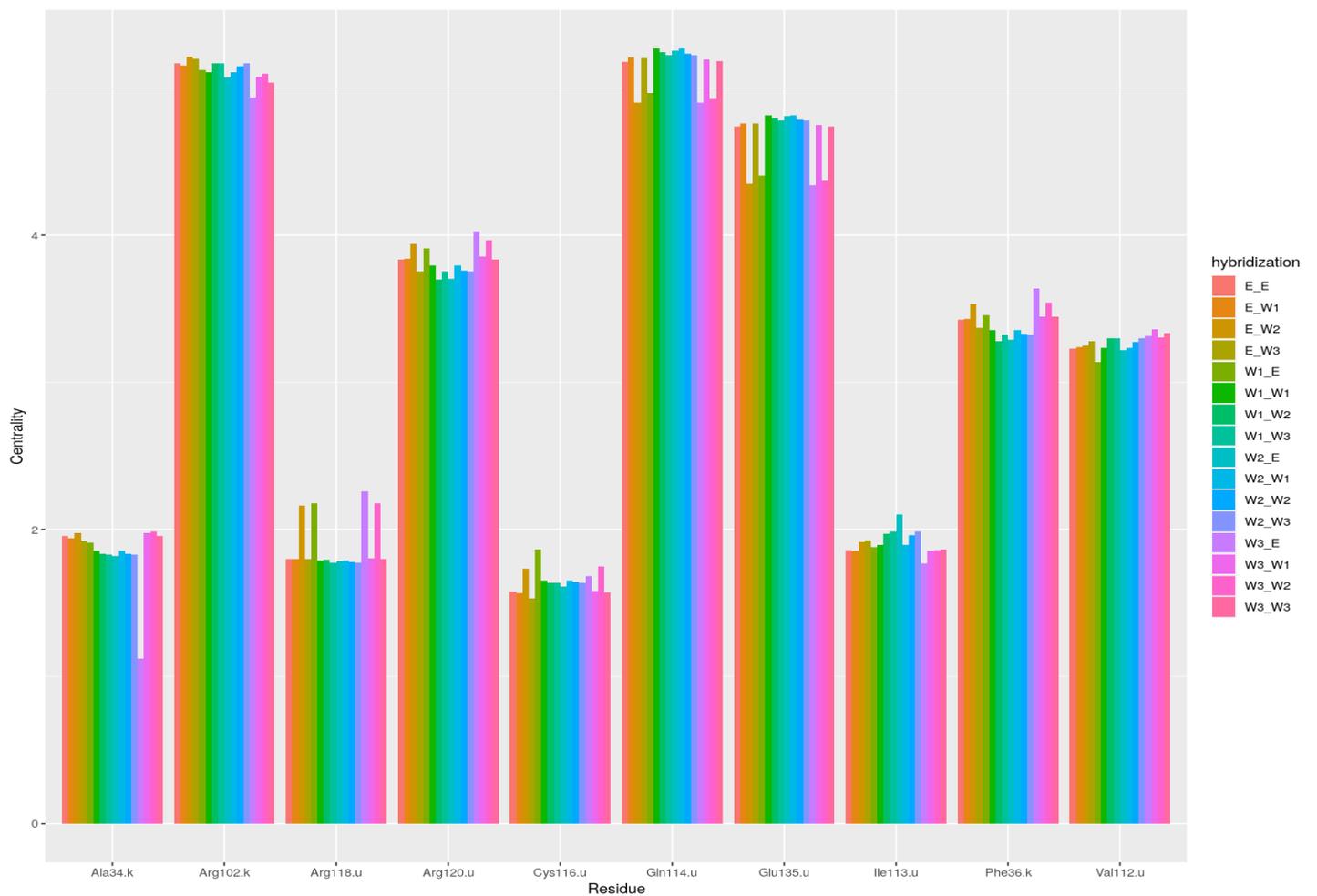
listed in the mutations identified here or that we only focused on one gene pair, the most mutated one while other plastid-nuclear gene pair, even though less mutated, could also generate modification of interaction and centrality that could impact the structure and function of the plastid ribosome. This might explain the lack of power and strong signals when looking at residues' centrality. Also, not all of these centrality measures are the most used in this kind of non-oriented networks (i.e. we do not consider the impact of one residue on another one, directionally but interactions between these two), in particular PageRank which is more used for oriented graphs.

### 3.3. Principal Component Analysis (PCA)

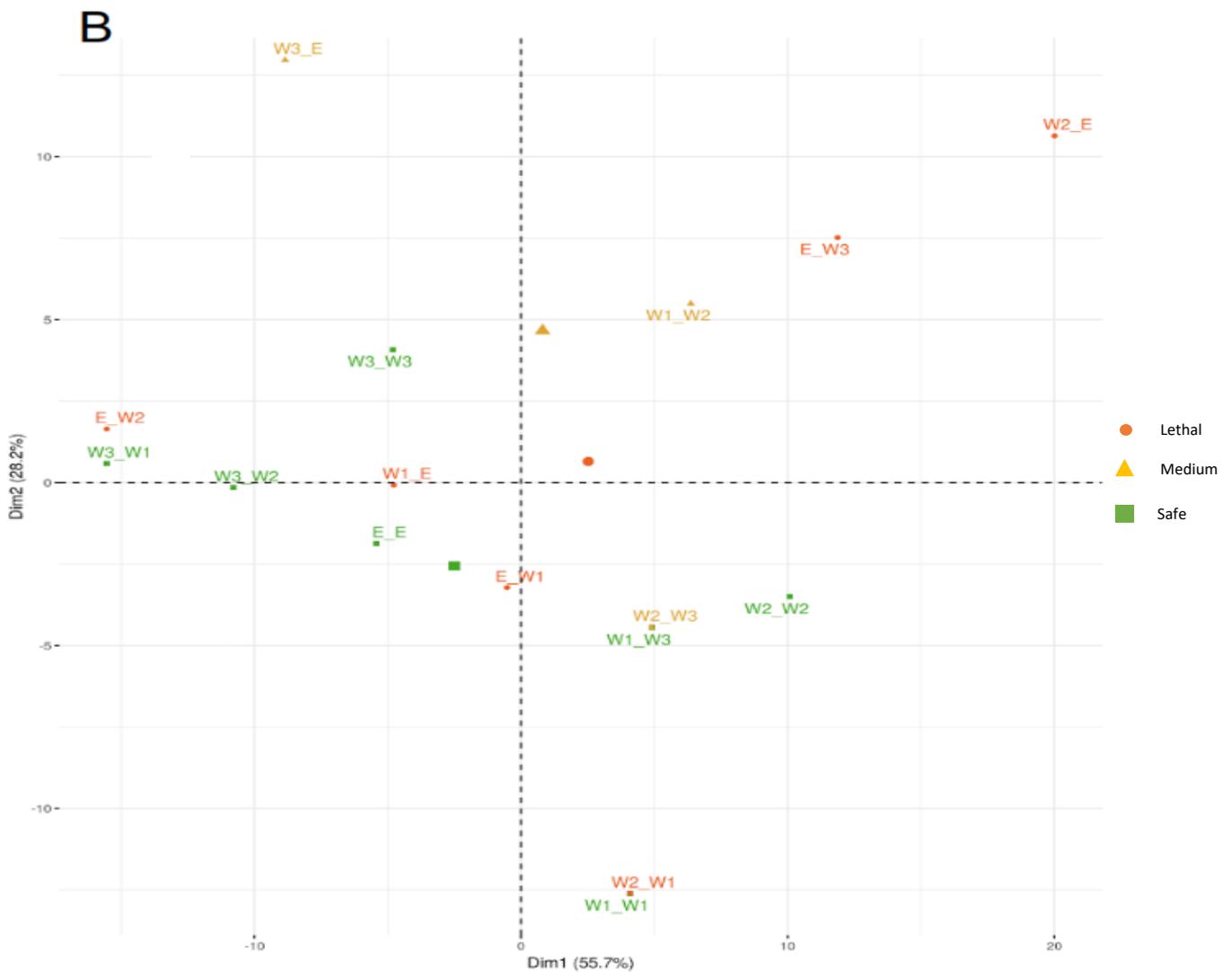
To go further and see whether changes in centrality in genes *rps11-rps21* might correlates with cross level of hybrid mortality, we looked at the distribution of the centrality of every residue of this gene



**Fig.2** BCA Centrality score of mutated residues of *rps11* according to the 16 different hybridizations. “k” corresponds to the *rps11*



**Fig.3** Betweenness centrality (BCA) of the ten most central residues in function of the different hybridization between the four lineages. “k” corresponds to the *rps11* and “u” to the *rps21*.



**Fig.4 Principal Component Analysis of Degree Centralities of residues in the *rps11-rps21* genes.** Representation of the 16 cross types on the two main dimensions. In green are cross outcomes with a low percent of hybrid mortality (<10%), in orange the one with a medium hybrid mortality (10% >= and <= 80 %) and in red the ones with high hybrid mortality (> 80%). The three biggest points represent the mean coordinates of a group.

pairs for the 16 models with Principal Component Analysis. We only show the results with one measure of centrality : degree centrality as the results are similar with the other four. The two first dimensions of the PCA explained between 51% and 83.9% of the variance which are good results (Suppl. Fig.1). Centralities associated with the different cross type can be discriminated on these two dimensions (figure 4). However, regarding the mean point for the three different classes (lethal, medium and non-lethal) we can see that they are near the center of the axes meaning that they are not well discriminating on these two PCA dimensions. If we look more closely, the amount of mortality associated with these crosses cannot be discriminated between lethal, medium or non-lethal hybridization on these two dimensions (figure 4). Three medium mortality and one high mortality can

be observed in the top right of the plot but there is some other elevated mortality cross type are also found among the low mortality crosses (figure 4). With the four other centrality measures, the strongly lethal cross types are also mixed with non-lethal ones. Yet, these two dimensions seem to discriminate reciprocal crosses. For example, when looking at cross between lineage E1 and W2 (E1/W2 and W2/E1), the centrality associated with these two directions are not found on the same sides of the PC1 and PC2: E1/W2 is on the top quarter left while W2/E1 is on the top quarter right. This is also true when looking at centrality associated with crosses W3/W1 and W1/W3: the former is near PC1 on top quarter left while the latter is near PC2 on bottom quarter right. More generally, we can observe this kind of asymmetry in location along the PC1 and PC2 for all the reciprocal crosses, suggesting modification of centrality differently depending on the cross direction.

Lack of signal to discriminate between strongly lethal and less lethal cross type might (1) suggest these two genes might not be the main driver of plastid ribosome non function in hybrids as modification of centrality associated with the cross type does not seem to correlate with the crosses' degree of hybrid mortality and/or (2) might come from the fact that we are only looking at one plastid-nuclear gene pairs and not the whole plastid ribosome, yet even though centrality modification associated with mutations in *rps11* and *rps21* is not discriminating how lethal are the crosses, it can influence the stability of the whole structure of the plastid ribosome and have an impact on its function.

#### 4. Conclusion

The results of this study showed that some mutations impacted the interactions between proteins in the plastid ribosome, potentially modifying the whole structure of the plastid ribosome and its function in inter-lineages hybrids. Several mutations modified the interactions between plastid and nuclear genes, either within the large or small subunits of the plastid ribosome. We focused on the most mutated gene pair: *rps11-rps21*. We showed that mutations associated with lineage E1 impacted the centrality of several residues potentially leading to a change of the interaction between these genes and driving the high hybrid mortality when use in inter-lineages cross. Some residues of *rps11* and *rps21* reacted the same way for four other cross type (E\_W2, W1\_E, W3\_E and W3\_W2). Modification of structure through mutations in one lineage could result in drastic changes of, if not the whole ribosome complex, interactions with its normally interacting nuclear gene *rps21*. If key interactions are disrupted, this could have subsequent consequences on the translation of photosynthetic proteins, which are essential to plastid function and plant development (3). Overall, centrality modification of residues association with each cross type cannot explain the differences in hybrid mortality observed for each cross.

In the present study, to gain time, we only focused on one plastid-nuclear gene pairs that contained most of the modified interactions. Yet the other mutations might also modify and disrupt the ribosomes structure, and the strength of the functional impact of a mutation and its associated structure modification might not follow a linear tendency: few mutations impacting essential or central residues might also have strong functional consequences. It could be interesting to perform the same analyses but using the whole set of ribosomal mutations, inducing changes of interactions between plastid-nuclear gene pairs identified in Table 1 and see if (i) these mutations change the centrality of highlighted residues in the whole structure and can explain better the lethality during hybridization and (ii) if they generate an impact on the interaction with the RNAs.

## 5. References

1. Gray MW. Evolution of organellar genomes. *Curr Opin Genet Dev.* 1999;9:678–87.
2. Sloan DB, Warren JM, Williams AM, Wu Z, Abdel-Ghany SE, Chicco AJ, et al. Cytonuclear integration and co-evolution. *Nat Rev Genet.* 2018;19(10):635–48.
3. Zoschke R, Bock R. Chloroplast Translation: Structural and Functional Organization, Operational Control, and Regulation. *Plant Cell.* 2018;30:745–70.
4. Greiner S, Bock R. Tuning a menage a trois: Co-evolution and co-adaptation of nuclear and organellar genomes in plants. *BioEssays.* 2013;35:354–65.
5. Zhang J, Ruhlman TA, Sabir J, Blazier JC, Jansen RK. Coordinated Rates of Evolution between Interacting Plastid and Nuclear Genes in Geraniaceae. *Plant Cell.* 2015;27(3):563–73.
6. Rand DM, Haney RA, Fry AJ. Cytonuclear coevolution: the genomics of cooperation. *Trends Ecol Evol.* 2004;19(12):645–53.
7. Smith DR. Mutation Rates in Plastid Genomes: They Are Lower than You Might Think. *Genome Biol Evol.* 2015;7(5):1227–34.
8. Greiner S, Sobanski J, Bock R. Why are most organelle genomes transmitted maternally? *Bioessays.* 2014;37:80–94.
9. Forsythe ES, Williams AM, Sloan DB. Genome-wide signatures of plastid-nuclear coevolution point to repeated perturbations of plastid proteostasis systems across angiosperms. *Plant Cell.* 2021;33(4):980–97.
10. Levin DA. The Cytoplasmic Factor in Plant Speciation. *Syst Bot.* 2003;28:8.
11. Greiner S, Rauwolf U, Meurer J, Herrmann RG. The role of plastids in plant speciation. *Mol Ecol.* 2011;20(4):671–91.
12. Barnard-Kubow KB, So N, Galloway LF. Cytonuclear incompatibility contributes to the early stages of speciation. *Evolution.* 2016;70(12):2752–66.

13. Matute DR, Cooper BS. Comparative studies on speciation: 30 years since Coyne and Orr. *Evolution*. 2021;75(4):764–78.
14. Burton RS, Barreto FS. A disproportionate role for mtDNA in Dobzhansky-Muller incompatibilities? *Mol Ecol*. 2012;21(20):4942–57.
15. Turelli M, Moyle LC. Asymmetric Postmating Isolation: Darwin’s Corollary to Haldane’s Rule. *Genetics*. 2007;176(2):1059–88.
16. Barnard-Kubow KB, McCoy MA, Galloway LF. Biparental chloroplast inheritance leads to rescue from cytonuclear incompatibility. *New Phytol*. 2017;213:1466–76.
17. Zupok A, Kozul D, Schöttler MA, Niehörster J, Garbsch F, Liere K, et al. A photosynthesis operon in the chloroplast genome drives speciation in evening primroses. *Plant Cell*. 2021;33(8):2583–601.
18. Ellison CK, Burton RS. DISRUPTION OF MITOCHONDRIAL FUNCTION IN INTERPOPULATION HYBRIDS OF *TIGRIOPUS CALIFORNICUS*. *Evolution*. 2006;60(7):1382–91.
19. Bogdanova VS, Galieva ER, Kosterin OE. Genetic analysis of nuclear-cytoplasmic incompatibility in pea associated with cytoplasm of an accession of wild subspecies *Pisum sativum* subsp. *elatius* (Bieb.) Schmahl. *Theor Appl Genet*. 2009;118:9.
20. Rockenbach K, Havird JC, Monroe JG, Triant DA, Taylor DR, Sloan DB. Positive Selection in Rapidly Evolving Plastid–Nuclear Enzyme Complexes. *Genetics*. 2016;204(4):1507–22.
21. Sloan DB, Triant DA, Forrester NJ, Bergner LM, Wu M, Taylor DR. A recurring syndrome of accelerated plastid genome evolution in the angiosperm tribe *Sileneae* (Caryophyllaceae). *Mol Phylogenet Evol*. 2014;72:82–9.
22. Postel Z, Touzet P. Cytonuclear Genetic Incompatibilities in Plant Speciation. *Plants*. 2020;9(487).
23. Sharbrough J, Conover JL, Tate JA, Wendel JF, Sloan DB. Cytonuclear responses to genome doubling. *Am J Bot*. 2017;104(9):1277–80.
24. Weng ML, Ruhlman TA, Jansen RK. Plastid–Nuclear Interaction and Accelerated Coevolution in Plastid Ribosomal Genes in Geraniaceae. *Genome Biol Evol*. 2016;8(6):1824–38.
25. Postel Z, Poux C, Gallina S, Varré JS, Godé C, Schmitt E, et al. Reproductive isolation among lineages of *Silene nutans* (Caryophyllaceae): A potential involvement of plastid-nuclear incompatibilities. *Mol Phylogenet Evol*. 2022;169:107436.
26. Martin H, Touzet P, Van Rossum F, Delalande D, Arnaud JF. Phylogeographic pattern of range expansion provides evidence for cryptic species lineages in *Silene nutans* in Western Europe. *Heredity*. 2016;116(3):286–94.
27. Van Rossum F, Martin H, Le Cadre S, Brachi B, Christenhusz MJM, Touzet P. Phylogeography of a widely distributed species reveals a cryptic assemblage of distinct genetic lineages needing separate conservation strategies. *Perspect Plant Ecol Evol Syst*. 2018;35:44–51.

28. Martin H, Touzet P, Dufay M, Godé C, Schmitt E, Lahiani E, et al. Lineages of *Silene nutans* developed rapid, strong, asymmetric postzygotic reproductive isolation in allopatry. *Evolution*. 2017;71(6):1519–31.
29. Bieri P, Leibundgut M, Saurer M, Boehringer D, Ban N. The complete structure of the chloroplast 70S ribosome in complex with translation factor pY. *EMBO J*. 2017;36(4):12.
30. Havird JC, Whitehill NS, Snow CD, Sloan DB. Conservative and compensatory evolution in oxidative phosphorylation complexes of angiosperms with highly divergent rates of mitochondrial genome evolution. *Evolution*. 2015;69(12):3069–81.
31. Tiller N, Weingartner M, Thiele W, Maximova E, Schöttler MA, Bock R. The plastid-specific ribosomal proteins of *Arabidopsis thaliana* can be divided into non-essential proteins and genuine ribosomal proteins. *Plant J*. 2012;69(2):302–16.
32. Tiller N, Bock R. The Translational Apparatus of Plastids and Its Role in Plant Development. *Mol Plant*. 2014;7(7):1105–20.
33. Tang X, Wang Y, Zhang Y, Huang S, Liu, Z, Fei D, et al. A missense mutation of plastid RPS4 is associated with chlorophyll deficiency in Chinese cabbage (*Brassica campestris* ssp. *pekinensis*). *BMC Plant Biol*. 2018;18(130):11.
34. Sharma MR, Wilson DN, Datta PP, Barat C, Schluenzen F, Fucini P, et al. Cryo-EM study of the spinach chloroplast ribosome reveals the structural and functional roles of plastid-specific ribosomal proteins. *Proc Natl Acad Sci*. 2007;104(49):19315–20.
35. Brysbaert G, Lorgouilloux K, Vranken WF, Lensink MF. RINspector: a Cytoscape app for centrality analyses and DynaMine flexibility prediction. *Bioinforma Oxf Engl*. 2018 Jan 15;34(2):294–6.
36. del Sol A, Fujihashi H, Amoros D, Nussinov R. Residue centrality, functionally important residues, and active site shape: analysis of enzyme and non-enzyme families. *Protein Sci Publ Protein Soc*. 2006 Sep;15(9):2120–8.
37. Hu G, Yan W, Zhou J, Shen B. Residue interaction network analysis of Dronpa and a DNA clamp. *J Theor Biol*. 2014 May 7;348:55–64.
38. Trouvilliez S, Cicero J, Lévêque R, Aubert L, Corbet C, Van Outryve A, et al. Direct interaction of TrkA/CD44v3 is essential for NGF-promoted aggressiveness of breast cancer cells. *J Exp Clin Cancer Res CR*. 2022 Mar 28;41(1):110.
39. Sharma MR, Wilson DN, Datta PP, Barat C, Schluenzen F, Fucini P, et al. Cryo-EM study of the spinach chloroplast ribosome reveals the structural and functional roles of plastid-specific ribosomal proteins. *Proc Natl Acad Sci U S A*. 2007 Dec 4;104(49):19315–20.
40. DeLano, W.L. The PyMOL Molecular Graphics System. Delano Scientific, San Carlos.; 2002.
41. Schrödinger, LLC, Warren DeLano. PyMOL [Internet]. 2020. Available from:

<http://www.pymol.org/pymol>

42. Land H, Humble MS. YASARA: A Tool to Obtain Structural Guidance in Biocatalytic Investigations. *Methods Mol Biol Clifton NJ*. 2018;1685:43–67.
43. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2018. Available from: <https://www.R-project.org/>
44. Brysbaert G, Mauri T, Lensink MF. Comparing protein structures with RINspecter automation in Cytoscape. *F1000Research*. 2018;7:563.
45. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res*. 2003 Nov 1;13(11):2498–504.

## 6. Annexes

**Table S1**

List of the nuclear and plastid genes encoding the small ribosomal subunit, selected as candidate for PNIs in Postel et al, 2022 and analyzed in the present study.

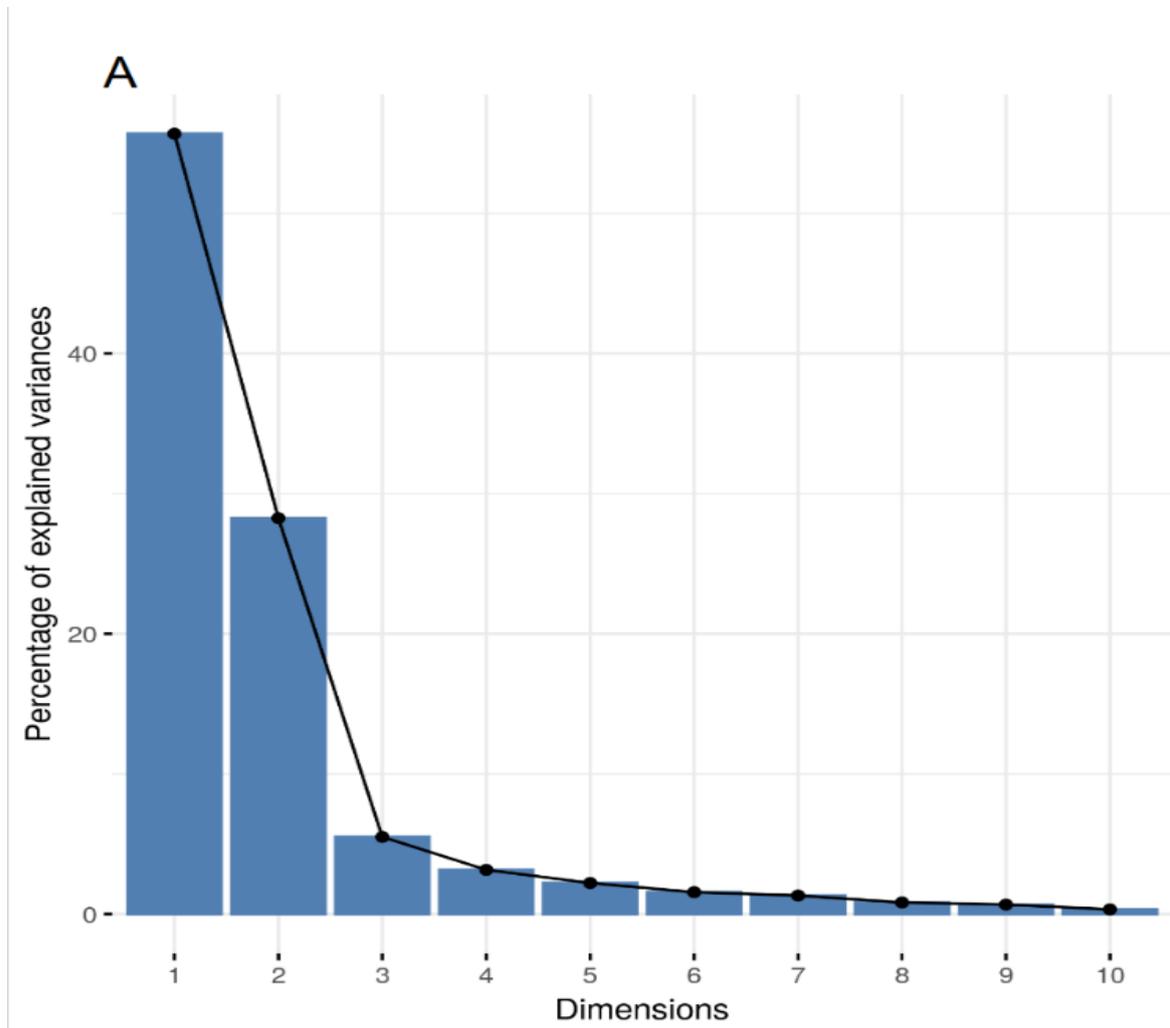
Genome	Gene name	UniProt identifier	Chain's name	Entitled	<i>Spinacia oleracea</i>		<i>S. nutans</i> lineages				
					Position	Amino Acid	Position	Amino acid			
								E1	W1	W2	W3
Nuclear	<i>rpl13</i>	P12629	K	S	132	A	152	A	S	A	A
	<i>rpl19</i>	P82413	Q	Y	229	L	235	F	L	L	L
	<i>rpl21</i>	P24613	S	AA	32	P	61	R	K	K	K
	<i>rpl27</i>	P82190	X	FA	20	L	20	L	L	V	L
	<i>rpl3</i>	P82191	D	L	35	S	32	S	F	F	F
Plastid	<i>rpl14</i>	P09596	L	T	49	N	49	N	H	H	H
					104	R	104	G	R	R	R
	<i>rpl16</i>	P17353	N	V	26	R	24	N	N	T	N
					78	P	76	P	P	S	P
					<i>rpl22</i>	P09594	T	BA	9	K	6
	114	V	92	L					F	L	L
	115	K	93	K					N	N	N
	<i>rpl32</i>	P28804	1	B	121	R	99	R	H	H	H
					22	K	22	K	K	M	K
					28	A	28	A	A	V	A
49					R	49	L	L	P	L	

In green: different amino-acid between *S. oleracea* and *S. nutans* lineages ; in purple: different amino-acid for one or more lineages of *S. nutans* compared to *S. oleracea* ; in orange: different amino-acid for one lineage of *S. nutans* compared to the others and to *S. oleracea* ; \* : mutations identified as under positive selection in Postel et al, 2022.

**Table S2**  
List of the nuclear and plastid genes encoding the large ribosomal subunit, selected as candidate for PNIs in Postel et al, 2022 and analyzed in the present study.

Genome	Gene name	UniProt identifier	Chain's name	Entitled	<i>Spinacia oleracea</i>		<i>S. nutans</i> lineages					
					Position	Amino Acid	Position	Amino acid				
								E1	W1	W2	W3	
Nuclear	<i>rps10</i>	P82162	j	TA	77	D	82*	E	E	E	K	
					142	Y	147	F	F	F	Y	
					127	E	125*	E	E	E	Q	
					88	V	36	I	V	V	V	
					89	L	37	F	L	L	S	
					91	Q	39	N	D	D	D	
					133	H	81	S	A	S	S	
					157	E	106	K	E	E	E	
					121	Y	51	H	H	Y	H	
					127	E	57	E	D	D	D	
					156	E	85	D	D	A	D	
					62	G	69	S	T	T	S	
	135	E	134	E	D	D	D					
	<i>rps5</i>	Q9ST69	e	OA	77	K	67	R	R	R	Q	
					141	S	131	T	T	T	S	
					172	M	183*	L	L	M	L	
	<i>rps6</i>	P82403	f	PA	198	V	209	V	V	V	I	
					62	A	48	A	T	T	T	
					149	V	134	L	V	V	V	
					157	K	142	N	N	N	I	
	162	A	147	A	E	E	E					
	Plastid	<i>rps11</i>	P06506	k	UA	6	P	6	L	P	P	P
						13	N	13	K	N	N	Y
						76	A	76	T	A	A	A
78						N	78	D	D	N	D	
82						T	82	T	T	P	T	
98						P	95	P	S	S	P	
98						P	96	P	S	S	P	
98						P	98	P	S	S	P	
104						A	104	A	A	A	G	
108						A	108	V	A	A	A	
114						I	114	I	L	L	L	
116						L	116	L	L	L	V	
<i>rps18</i>		Q9M3K7	r	BB	13	R	14	R	R	R	Q	
					18	R	17	H	R	R	R	
					50	R	49	R	R	Q	R	
					81	E	80	-	R	G	G	
					81	E	82	-	R	G	G	
					94	A	93	-	I	Q	I	
<i>rps19</i>		P06508	s	CB	19	I	19	M	M	M	I	
					33	T	33	T	T	T	N	
					65	R	65*	R	R	Y	D	
					91	R	91	R	R	R	Q	
<i>rps2</i>		P08242	b	LA	24	T	24	I	I	T	I	
					<i>rps3</i>	P09595	c	MA	79	G	79	G
94		D	94	D					D	A	D	
103		L	103	L	L	F	L					
117		I	117	I	I	I	L					
213		I	213	I	I	I	L					
<i>rps7</i>		P82129	g	QA	151	F	151*	F	L	L	F	

In green: different amino-acid between *S. oleracea* and *S. nutans* lineages ; in purple: different amino-acid for one or more lineages of *S. nutans* compared to *S. oleracea* ; in orange: different amino-acid for one lineage of *S. nutans* compared to the others and to *S. oleracea* ; \* : mutations identified as under positive selection in Postel et al, 2022.



**Suppl. Fig.1 – Principal Component Analysis of Degree Centralities of residues in the *rps11-rps21* genes.** Percentage of variance explained by the different dimensions.





## CHAPTER 2

---

# Paternal leakage of plastids rescues inter-lineage hybrids in *Silene nutans*

Zoé Postel<sup>1</sup>, Fabienne Van Rossum<sup>2,3</sup>, Cécile Godé<sup>1</sup>, Eric Schmitt<sup>1</sup>, Pascal Touzet<sup>1</sup>

Manuscript in preparation.



<sup>1</sup> Univ. Lille, CNRS, UMR 8198 - Evo-Eco-Paleo, F-59000 Lille, France

<sup>2</sup> Meise Botanic Garden, Nieuwelaan 38, BE-1860 Meise, Belgium

<sup>3</sup> Service général de l'Enseignement supérieur et de la Recherche scientifique, Fédération Wallonie-Bruxelles, rue A. Lavallée 1, BE-1080 Brussels, Belgium



The objective of this chapter was to assess whether paternal leakage of the plastid genome was occurring in inter-lineage hybrids of *S. nutans* and whether it could facilitate their rescue. We used molecular data analyses and genotyping through qPCR to do so. Paternal leakage was identified and rescued some of the surviving hybrids, through transmission of a less incompatible plastid genome.

Leaf fragments were collected and dried by Cécile Godé and Pascal Touzet. Weighting was done by myself. DNA extraction and genotyping were done by myself with the help and technical guiding of Cécile Godé. Data analysis was done by myself with under the experienced eye of Pascal Touzet. I wrote the original draft of the manuscript, again with the help of Pascal Touzet. Fabienne Van Rossum also participated to its editing, until we reached the current version.

I also supervised Zakia Sultana (MSc students – Univ. Lille) to try to apply the same methodology to mitochondrial diagnostic SNPs, but without success.



# Paternal leakage of plastids rescues inter-lineage hybrids in *Silene nutans*

## Abstract

Organelle genomes are usually maternally inherited in angiosperms. However, biparental inheritance has been observed in some species, especially within hybrids resulting from crosses between divergent genetic lineages. Focusing on the plastid genome, paternal leakage might rescue inter-lineage hybrids suffering from plastid-nuclear incompatibilities. *Silene nutans* is composed of at least four genetically differentiated lineages, exhibiting strong post-zygotic isolation highlighted by the occurrence of chlorotic inter-lineage hybrids with high juvenile mortality. Some of these hybrids survived and exhibited green or variegated leaves characterized by the occurrence of white and green leaf sectors, a phenotype thought to result from paternal leakage of the plastid genome in inter-lineages hybrids. Under the hypothesis that hybrid breakdown was due to plastid-nuclear incompatibility in *S. nutans* inter-lineages hybrids, we tested whether the surviving hybrids inherited the paternal plastid genome and survived thanks to this mechanism. By genotyping 504 surviving inter-lineage hybrids for six lineage-specific plastid SNPS, we demonstrated the presence of a substantial proportion of hybrids that inherited the paternal plastid genome, which in some cases allowed them to be rescued by providing a plastid genome potentially more compatible with the hybrid nuclear background, thus increasing their survival probability. We then discuss the effectiveness of this phenomena as a counter-barrier to speciation.

**Key words:** plastid genome, inter-lineages hybrids, paternal leakage, *Silene nutans*

## 1. Introduction

Plant eukaryotic cells are known to be composed of three major genomic compartments: the nucleus and two organelles, the mitochondrion, and the plastid. The two organellar genomes originate from free-living bacteria integrated into the host cell as endosymbionts (Greiner and Bock, 2013; Sloan et al., 2018). They differ from the nuclear genome in several characteristics, e.g. in mutation rate and inheritance pattern (Wolfe et al., 1987; Drouin et al., 2008; Greiner et al., 2014). One of the most striking difference lies in the inheritance pattern of these genomes. Plant organellar genomes are generally uniparentally inherited, and in angiosperms, mostly from the maternal parent (Greiner et al., 2014). Even though this mode of inheritance has severe consequences for species evolutionary potential, in particular a lack of recombination and evolution under Muller's ratchet (McCauley, 2013), it can also lead to some benefits (Greiner et al., 2014; Ramsey and Mandel, 2019; Postel and Touzet, 2020). For example, uniparental inheritance leads to a tight genetic interdependence of the nuclear and organellar genomes, reinforcing cytonuclear coevolution (Greiner et al., 2014). Uniparental transmission is also thought to have evolved because it limits the spread of selfish cytoplasmic elements, i.e. maladaptive fast replicating or aggressive elements (Sobanski et al., 2019; Radzvilavicius, 2021). By only transmitting one organellar haplotype, it limits organellar genetic diversity, within-cell selection and competition between transmitted haplotypes, which could otherwise lead to the emergence of these selfish organellar elements (McCauley, 2013; Christie and Beekman, 2016). More generally, uniparental inheritance improves purifying and adaptive selection of organellar haplotypes (Radzvilavicius, 2021).

Uniparental inheritance of organelles appears to be the rule, yet biparental transmission and paternal leakage of organellar genomes seem to have occurred multiple times, at least within the angiosperms (Yao et al., 1995; Zhang et al., 2003; Xu, 2005; Weihe et al., 2009). There are several hypotheses to explain such deviation from uniparental transmission. The first one is that, in crosses between isolated genetic lineages or different species, paternal leakage could be the result of a breakdown of the mechanisms preventing it in the first place (McCauley, 2013; Ramsey and Mandel, 2019). Other hypotheses rely on the beneficial effect of biparental transmission. For example, for mitochondrial genomes, even low levels of paternal leakage could counteract the effects of Muller's ratchet through recombination (Barr et al., 2005; Greiner et al., 2014; Parakatselaki and Ladoukakis, 2021). Paternal leakage could also lead to the rescue of hybrids suffering from cytonuclear incompatibilities (CNIs). Such incompatibilities are the result of disruption of co-adaptation between organellar and nuclear genes in hybrids, which can result in hybrid breakdown in cases of crosses between divergent or isolated lineages (Sloan et al., 2018; Postel and Touzet, 2020). When incompatibilities are due to the plastid, hybrids appear to be chlorotic and these might lead to a high rate of mortality. But if paternal leakage of the plastid genome

occurs, hybrids suffering from such incompatibilities would exhibit variegated leaves (from green to white leaves or leaf sectors on the same plant) or would even be fully green (Metzlaf et al., 1982; Greiner et al., 2011). Variegation would indicate that one of the two parental plastid genomes is unable to develop on the hybrid background, likely because it is incompatible, while the other is not (Sakamoto, 2003; Bogdanova and Kosterin, 2006; Weihe et al., 2009). In that sense, paternal leakage of the organellar genome might be a way of rescuing inter-lineages hybrids suffering from CNIs either (i) by increasing the likelihood of inheritance of a less incompatible organellar genome with the hybrid nuclear background, or (ii) because it introduces genetic variation among organellar haplotypes (Greiner et al., 2014; Barnard-Kubow et al., 2017). In both cases, it makes between-organellar selection possible and may lead to the loss of the incompatible organellar genome (Jansen and Ruhlman, 2012; Barnard-Kubow et al., 2017; Postel and Touzet, 2020).

In *Silene nutans* (Caryophyllaceae), several highly differentiated genetic lineages have been identified based on plastid and nuclear molecular markers (Martin *et al.*, 2016; Van Rossum *et al.*, 2018). A diallelic cross between four genetic lineages found in the western geographic range of the species, and exhibiting different plastid haplotypes has revealed a strong post-zygotic isolation among them: most of the inter-lineage crosses exhibit high hybrid mortality at juvenile stage, most likely due to chlorosis (Martin *et al.*, 2017; Van Rossum *et al.*, in prep.). We have recently shown that hybrid breakdown could be due to plastid-nuclear incompatibilities through disruption of co-adaptation between plastid and nuclear genes in inter-lineage hybrids (Postel *et al.*, 2022). Interestingly, some of the inter-lineage hybrids that survived exhibited a variegated leaf phenotype, suggesting the occurrence of paternal leakage of the plastid genome (Greiner *et al.*, 2014; Postel and Touzet, 2020). In the present study, we tested 1) for paternal transmission of the plastid genome of *S. nutans* and 2) whether paternal leakage could be a mean of rescuing these inter-lineages hybrids suffering from PNIs. To do so, we genotyped the 504 surviving hybrids using previously defined plastid SNP markers, analyzing separately white and green leaf samples from variegated hybrids. These hybrids suffering from PNIs, we expected to detect paternal plastid genome in some the green leaf samples or fully green individual, indicating that paternal leakage allow for the transmission and inheritance of a more compatible plastid genome with the hybrid nuclear background.

## **2. Material and methods**

### *2.1. Silene nutans*

*Silene nutans* L. is a Caryophyllaceae species, largely distributed in Europe. This species is gynodioecious, with a cytonuclear sex determination system (Garraud *et al.* 2011). Natural populations

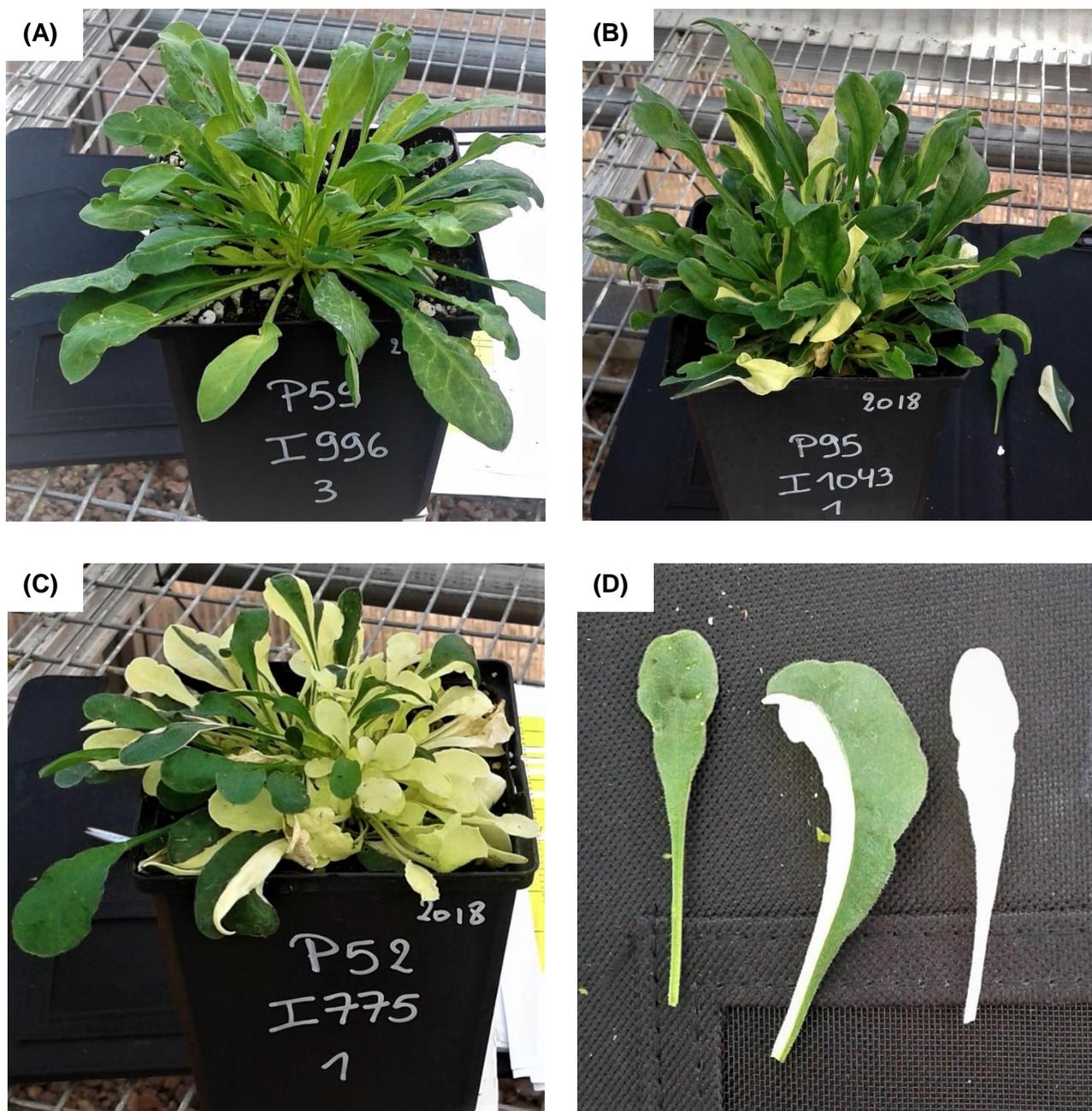
are composed of variable numbers of female and hermaphroditic individuals. Previous studies shed light on the existence of two strongly differentiated evolutionary groups, using nuclear (microsatellite) and plastid markers, in relation to past climatic events and postglacial recolonization:

an eastern one (E), widespread in northern and eastern Europe and reaching its western margin in France and UK, where some of the E1 lineage populations occur ; and a western one, composed of four lineages, in particular W1 distributed in England, France and Belgium, W2 restricted to Spain and south-western France, and W3 located in the Alps (including Jura) and Italy (Martin et al., 2016; Van Rossum et al., 2018). At secondary contact zones between E1 and W1 lineages in southern Belgium and southern England, and, in north-eastern France for E1, W1 and W3, no hybridization events were detected, suggesting the absence of gene flow at least between these two lineages (Martin et al., 2016). Diallelic crosses conducted between the four lineages (i.e. E1, W1, W2, and W3) highlighted the presence of strong and asymmetric post-zygotic barriers, not only between E1 and western lineages, but also among the three western lineages (Van Rossum *et al.*, in prep.). High and asymmetric proportions of hybrid mortality were observed, especially when the organelles were inherited from E1 or W2 lineage. Post-zygotic reproductive isolation between these lineages is likely due to the presence of plastid-nuclear incompatibilities, through disruption of plastid-nuclear co-adaptation in inter-lineages hybrids resulting in hybrid breakdown. Analyses of the plastid genes in the four lineages revealed potential candidate nuclear and plastid gene pairs involved in these plastid-nuclear incompatibilities (Postel et al., 2022). Plastid genome of *S. nutans* also exhibit a peculiar evolutionary pattern similar to the ones observed *Silene* species with “fast-evolving” plastid genome (Sloan et al., 2014): elevated number of mutations, high  $d_N/d_S$ , signatures of relaxed selection (Postel et al., 2022).

## 2.2. Plant material

To assess paternal leakage of the plastid genome in *S. nutans*, we investigated SNP patterns on inter-lineage hybrids (Table 1) obtained from a diallelic cross experiment conducted in 2016-2017, using four genetic lineages of *S. nutans*: E1, W1, W2, and W3 (Van Rossum *et al.*, in prep). The experiment involved the survey in April-May 2018 of 7409 juveniles resulting from intra- and inter-lineage reciprocal crosses, grown in a cold greenhouse. Juvenile growth and survival were followed for five weeks. Mortality was estimated for each cross type as the proportion of dead juveniles after five weeks. From the 4312 surviving juveniles after five weeks, 944 plants representative of the cross types (including 546 inter-lineage hybrids) were kept and repotted for estimating pollen viability after flowering in 2019.

We collected leaves from the 504 inter-lineage hybrids that had survived until autumn 2018 (Table 1), and dried them in silica gel. The following leaf phenotypes were identified: 426 fully green individuals,



**Fig.1 Pictures of the hybrids resulting from diallelic crosses in 2018.** For each cross type, the first lineage is the maternal one. (A) Fully green ExW3 hybrid. (B) Variegated W2xW1 hybrid. (C) Variegated W2xW3 hybrid. (D) White, green and variegated leaves from W2xW3 hybrid (Pictures: Pascal Touzet).

four light green individuals (i.e. with light green leaves) and 74 variegated individuals (i.e. with both green and white sectors on the same leaves or with white and green leaves on the same plant) (Figure 1). Concerning the variegated phenotype, we sampled both green and white leaves or sectors for 61 individuals, only green leaves or sectors for 12 individuals and only white leaves or sectors for one individual (Table 1). In total, 560 leaf samples were genotyped (Table 1).

**Table 1**

Summary of hybrid survival, number of hybrids genotyped and showing paternal leakage per phenotypic category, the proportion of hybrids exhibiting paternal leakage and effective paternal leakage rate per cross-type. Details results for the variegated individuals can be found in suppl. table S2.

CROSS EXPERIMENT			GENOTYPED				LEAKAGE				The proportion of genotyped hybrids exhibiting paternal leakage
Cross directions (mother x father)	Total	Mortality rate	G	V	LG	Total	G	V	LG	Total	
E1 x W1	507	1.00	0	0	0	0	-	-	-	0	0.00
W1 x E1	516	0.95	16	5	1	22	0	0	0	0	0.00
E1 x W2	447	1.00	1	0	0	1	0	-	-	0	0.00
W2 x E1	323	1.00	1	0	0	1	0	-	-	0	0.00
E1 x W3	215	0.91	10	5	0	15	9	5	-	14	0.93
W3 x E1	32	0.66	6	1	1	8	0	0	0	0	0.00
W1 x W2	699	0.42	109	19	0	128	0	13	-	13	0.10
W2 x W1	628	0.84	25	16	0	41	23	15	-	38	0.90
W1 x W3	153	0.12	75	3	0	78	20	3	-	23	0.28
W3 x W1	306	0.03	98	0	0	98	0	-	-	0	0.00
W2 x W3	393	0.66	32	25	2	59	32	24	2	58	0.98
W3 x W2	199	0.01	53	0	0	53	0	-	-	0	0.00
TOTAL	4418	0.30	426	74	4	504	84	60	2	146	0.29

**G:** green hybrids ; **V:** variegated hybrids ; **LG:** light green hybrids ; **Total:** Total number of surveyed hybrids during the cross experiment. ; **Mortality:** Number of hybrid seedlings that died after five weeks ; **Genotyped:** Number of hybrids for which leaf fragments were collected. per phenotypic category ; **Leakage:** Number of hybrids exhibiting paternal leakage. per phenotypic category

### 2.3. DNA extraction and plastid genotyping

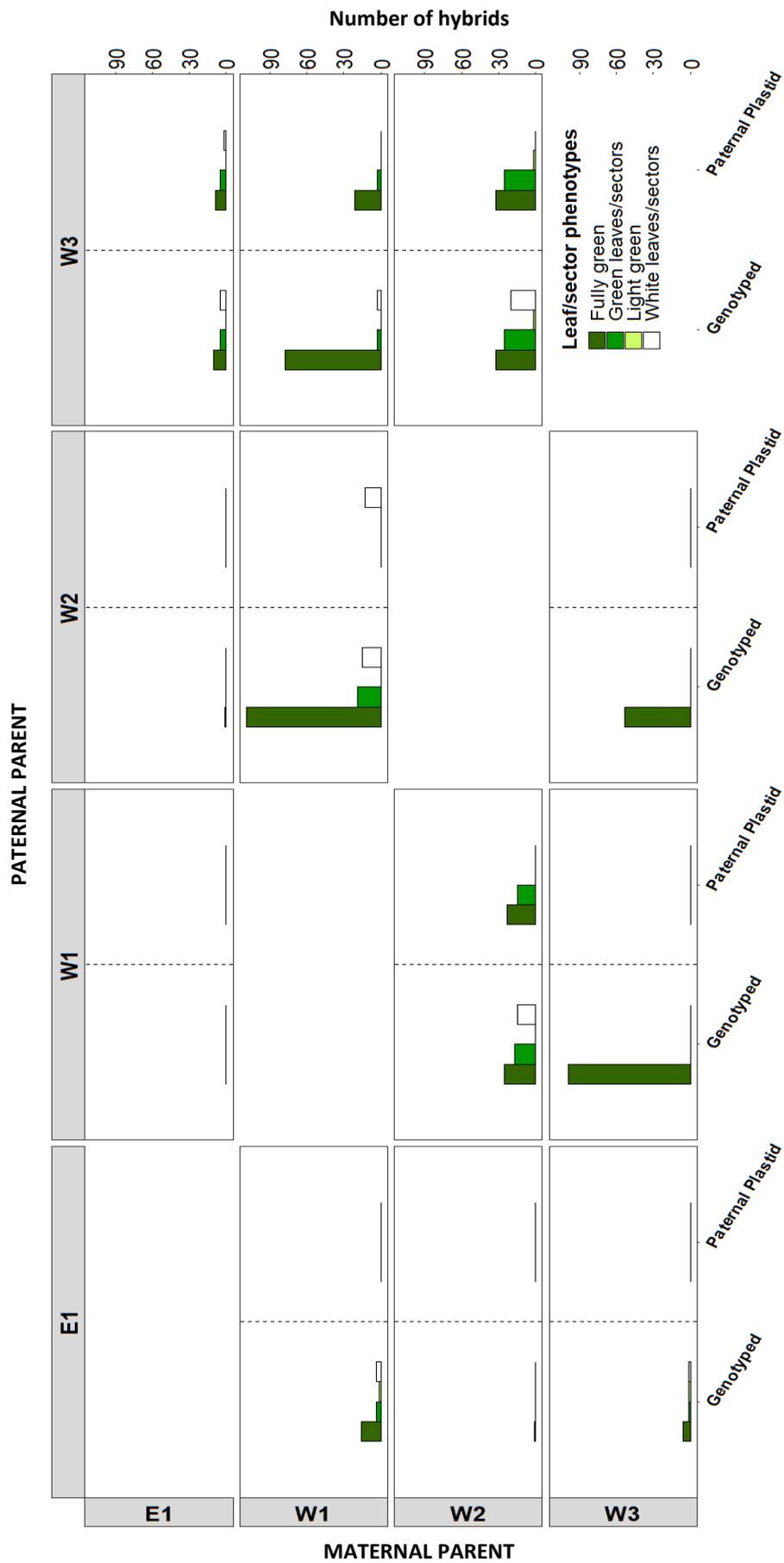
DNA extraction from leaf samples and plastid genotyping followed the protocol described in Martin *et al.* (2016). In brief, to identify the plastid genotype of the leaf samples, we used six diagnostic plastid SNPs that unambiguously enabled us to assign every leaf sample to one of the four lineages (E1, W1, W2, W3) and therefore the origin of the plastid genome (maternal or paternal) for each leaf sample (Suppl. Table S1).

## 3. Results

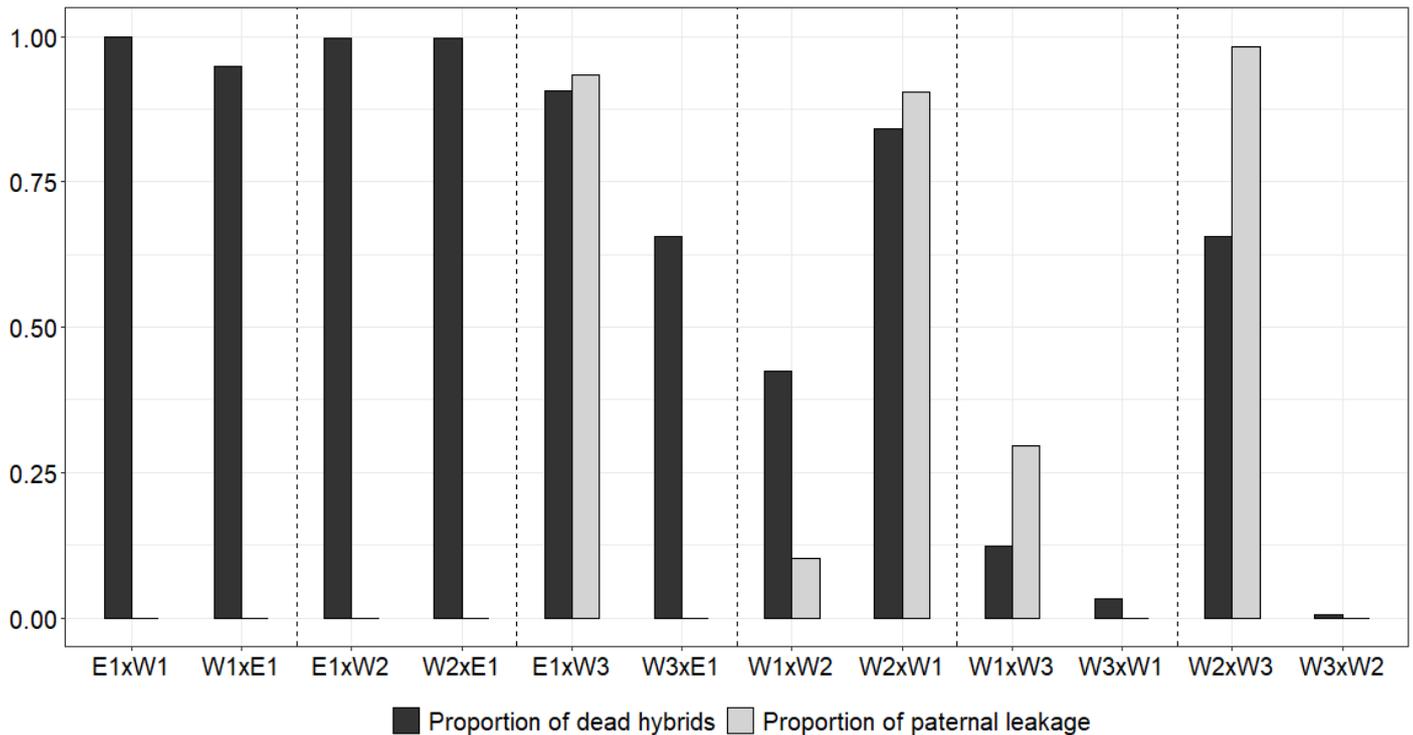
An uneven proportion of hybrids survived, depending on the type and direction of the cross: almost of the hybrids died for crosses involving lineage E1 (either as the mother or the father) and lineage W2 as the mother (Table 1). Among the surviving hybrids, the number of variegated ones was high for the crosses involving western lineages especially when W2 was involved: 87% of the variegated hybrids resulted from crosses involving W2 either as the mother or the father (Table 1, Figure 2, Suppl. Table S2). When E1 was involved, either as the mother or the father, rate of paternal leakage was near 0, likely because almost all resulted hybrid died (Table 1, Figure 2). Regarding the western lineages, rate of paternal leakage was the highest when W3 was crossed with W1, W2 or E1 as the paternal parent (i.e. 0.28, 0.98 and 0.93 respectively) (Table 1, Figure 2). W1 and W2 lineages exhibited intermediate levels of paternal leakage, with no paternal leakage when crossed with W3 as the mother and substantial levels of paternal transmission when crossing W1 and W2 together, in both directions of the cross (Table 1, Figure 2).

Based on the six plastid diagnostic SNPs, the presence of paternal plastid genomes was detected in 146 surviving inter-lineage hybrids, for five out of the 12 cross types, with a proportion of hybrids having inherited the paternal plastid varying from 0.10 to 0.98 (Table 1, Figure 2). As expected, paternal leakage was mostly detected for hybrids with green leaf phenotype and for green leaves/sectors of variegated hybrids, the white leaves or sectors containing the maternal plastid genome (Table 1, Figure 2). This pattern excluded one cross type, W1xW2, where the opposite was observed: all variegated hybrids exhibited paternal leakage on white leaves/sectors while green leaves/sectors contained the maternal plastid genome (Table 1).

The level of paternal leakage seemed to depend on 1) the cross type and 2) the cross direction. For example, hybrids resulting from crosses between E1 and W2 (in both direction) did not exhibited any paternal leakage (Table 1, Figure 2) while hybrids resulting from crosses between W3 and W1 (in both directions again), did so. Additionally, almost all of the few hybrids that survived when the maternal parent was either lineages E1 or W2 exhibited paternal leakage while none inherited the paternal plastid genome when W3 was the mother. Consistently, for the reciprocal cross between W1 and W3 lineages, when W1 was the maternal parent, a substantial proportion (0.28) of paternal leakage was observed on the green leaves/sectors while no paternal leakage was found when W3 was the mother (Figure 1). Results of figure 3 suggest that when one of the two cross direction leads to high amount of hybrid seedling mortality, the few hybrids that survived for this cross direction contained the paternal plastid genome (Figure 3). For example, taking the crosses between E1xW3 or W2xW3, when lineages E1 or W2 are the maternal parents, mortality is high (i.e. 0.91 and 0.66 vs 0.66 and 0.01 when W3 is the mother) (Table 1, Figure 3). Consistently, almost all of the few hybrids that survived with E1 or W2 as mother, exhibited paternal plastid genome (i.e. the plastid genome of lineage W3). This highlights the fact that when the maternal lineage is highly incompatible (i.e. high associated levels of



**Fig.2** Number of genotyped hybrids and hybrids exhibiting paternal leakage of the plastid genome for each cross and leaf phenotype. Green and white leaves sectors were collected from variegated hybrids



**Fig.3 Histograms representing the proportion of dead hybrids and surviving hybrids having inherited the paternal plastid genome per cross. Reciprocal crosses are represented together.**

hybrid mortality), the amount of paternal leakage among the few surviving hybrids is high too (Figure 3).

#### 4. Discussion

##### 4.1. Could paternal leakage slow down the speciation process?

In the present study, we tested the paternal transmission of the plastid genomes in inter-lineage hybrids of *S. nutans*. We detected a substantial proportion of hybrids that have inherited the paternal plastid genome. As such, it confirms the involvement of the plastid genome in the reproductive isolation between *S. nutans* lineages, as suggested indirectly by the pattern of plastid diversity described in a recent study (Postel *et al.*, 2022 – chapter 1).

Paternal leakage might be viewed as a way of rescuing hybrids suffering from cytonuclear incompatibilities (Hu *et al.*, 2008; Barnard-Kubow *et al.*, 2017). Having a more compatible plastid genome could lead at least to a partial recovery of plastid-nuclear co-adaptations and plastid-protein complex functions (Postel and Touzet, 2020). In inter-lineage hybrids of *C. americanum*, where plastid-nuclear incompatibilities were suspected to play a role in the reproductive isolation observed between

genetic lineages, paternal leakage allowed the rescue of some inter-lineages hybrids (Barnard-Kubow *et al.*, 2016; Barnard-Kubow *et al.*, 2017). In another set of studies on crosses of lineages of *Pisum sativum* ssp. *elatius*, plastid-nuclear incompatibilities were also likely involved in hybrid inviability and sterility, with paternal leakage of the plastid genome leading to hybrids fitness recovery (Bogdanova and Kosterin, 2006; Bogdanova, 2007; Bogdanova *et al.*, 2009). As paternal leakage can rescue inter-lineage hybrids, when reproductive isolation is incomplete (and even if it is strong), as it is the case in *S. nutans* and *C. americanum*, paternal leakage may slow down the speciation process by suppressing the reproductive barriers preventing the mixing of lineages (Barnard-Kubow *et al.*, 2017). Because of the presence of two plastid types, intracellular selection can occur and lead to the loss of the incompatible plastid genome, further facilitated by within-individual sorting-out of the plastids (Ramsey *et al.*, 2019). Subsequent generations might benefit from fitness recovery because of the loss of the incompatible plastid genome (Barnard-Kubow *et al.* (2017).

In our study, the proportion of hybrids which inherited the paternal plastid genome seemed to vary regarding the cross type and direction as suggested above. Firstly, we observed that depending on which lineage would be used as the maternal parent, the proportion of paternal leakage won't be equal for a given reciprocal cross. This might highlight the fact that paternal leakage might be "useful" in the rescuing process only when the maternal lineage contains a highly incompatible plastid genome giving the hybrid nuclear background. Among the four lineages of *S. nutans*, E1 and W2 contained the most highly divergent plastid genomes and leads to the highest levels of hybrids seedling mortality (Postel *et al.*, 2022). If we go back to the cross between E1xW3 and W3xE1: when E1 is the mother, we can observe high levels of hybrids mortality and almost 100% of paternal plastid genome inheritance among the few ones while when W3 is the mother, more hybrids survived and did so without paternal leakage. Similar results were observed for crosses between W2 and W3. This highlights the fact that when the maternal lineage already contains a "compatible" plastid genome, paternal leakage does not rescue hybrids and is not "selected" for.

Secondly, paternal leakage might not be enough to rescue inter-lineages hybrids resulting from crosses between highly divergent lineages. Again, looking at crosses involving for example lineage E1, very few if any hybrids survived, even when E1 lineage was not the mother. Concordantly, no hybrids were detected in the adult generation in contact zones between the two lineages (Martin *et al.*, 2017), despite evidence of inter-lineage pollen flow leading to hybrid seed production in the wild in southern Belgium (Cornet *et al.*, 2022). In this case, it is fair enough to consider that even if some paternal leakage occurs within *S. nutans*, the initial level of plastid divergence between the reproductive partner will initially set the possibility for paternal leakage to rescue the inter-lineages hybrids. This suggests that the capacity of paternal leakage to slow down the speciation process by allowing low levels of

gene flow between distant lineages will remain dependent on the initial level of divergence between the crossing lineages as well as how strong and numerous are the reproductive barriers already acting. So, even in the more favorable cases where paternal leakage rescued the hybrids, it is difficult to evaluate the impact of paternal leakage on the speciation process, as we do not know neither the fitness of the hybrids, e.g. their probability of survival in natural environments (xero-thermophilous vegetation), which can be more stressful than optimal plant care in greenhouse conditions, nor their reproductive ability. It would be worth to investigate the possible occurrence of hybrids in other situations of parapatry or sympatry, especially for the western lineages, e.g. in south-western France for W1 and W2, and eastern France for W1 and W3.

#### 4.2. Breeding system and paternal leakage

In the present study, we aimed to detect paternal transmission of the plastid genomes in inter-lineage hybrids of *S. nutans*. In this system, from to 98% of the surviving hybrids exhibited the paternal plastid genome, according to the cross type and direction. This rate is comparable, yet higher, to what was observed in other cases of inter-lineage or interspecific crosses. For instance, crosses of inter-lineage hybrids of *Campanulastrum americanum* exhibited 6% to 53% of non-maternal inheritance of the plastid genome, depending on the direction of the cross (Barnard-Kubow et al., 2017). In inter-specific hybrids of the genus *Passiflora*, up to 71% of paternal leakage of the plastid genome was detected in inter-specific hybrids (Hansen et al., 2007; Shrestha et al., 2021). Even though this level of paternal leakage observed might be due to the classical result of crosses between divergent lineages, leading to the breakdown of the mechanism preventing paternal transmission of the organellar genomes (Greiner et al. 2014), we cannot exclude that paternal leakage is naturally occurring in *S. nutans*, as observed in a sister species, *S. vulgaris*, with a rate ranging from 1.9% to 4.7% (McCauley et al., 2007). Indeed, for plastid genome to be paternally transmitted, presence of plastids within the male germinative cell is mandatory, regardless of whether paternal leakage is occurring because of breakdown of the mechanism ensuring uniparental inheritance or not. This is also observed in several species of Dipsacales, where presence of plastid genomes in the pollen germinative cell was observed, suggesting that these species have the potential for paternal leakage even though the paternal plastid genome is eliminated further during reproduction or zygote development (Hu et al., 2008). In *S. vulgaris* the mitochondrial genome also exhibits paternal leakage (McCauley et al., 2005), suggesting that plastids and mitochondria are co-transmitted through the pollen grains. In the case of the mitochondria, *S. vulgaris* breeding system, i.e. gynodioecy, might have favored the selection for paternal leakage in order to maintain mitochondrial polymorphism and limit female spread and thus avoid pollen limitation (Wade and McCauley, 2005). Since *S. nutans* is also gynodioecious, paternal

leakage of both organelles, not only the plastid, could be constitutive in the species. Indeed, the analysis of the diversity of the mitochondrial genomes in *S. nutans* also revealed a signature of paternal leakage of the mitochondria, i.e. heteroplasmy and reassortment events (Postel *et al.*, in prep.). In fact, the lack of fixed mutations differentiating the mitochondrial genomes of the four lineages prevented us to find specific mitochondrial SNPs to estimate the concomitant paternal leakage rate of the mitochondria in the present study.

## 5. References

- Barnard-Kubow, K. B., McCoy, M. A., and Galloway, L. F. (2017). Biparental chloroplast inheritance leads to rescue from cytonuclear incompatibility. *New Phytol.* 213, 1466–1476. doi: 10.1111/nph.14222.
- Barnard-Kubow, K. B., So, N., and Galloway, L. F. (2016). Cytonuclear incompatibility contributes to the early stages of speciation. *Evolution* 70, 2752–2766. doi: 10.1111/evo.13075.
- Barr, C. M., Barr, C. M., Neiman, M., and Taylor, D. R. (2005). Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. *New Phytol.* 168, 39–50.
- Bogdanova, V. S. (2007). Inheritance of organelle DNA markers in a pea cross associated with nuclear-cytoplasmic incompatibility. *Theor. Appl. Genet.* 114, 333–339. doi: 10.1007/s00122-006-0436-6.
- Bogdanova, V. S., Galieva, E. R., and Kosterin, O. E. (2009). Genetic analysis of nuclear-cytoplasmic incompatibility in pea associated with cytoplasm of an accession of wild subspecies *Pisum sativum* subsp. *elatius* (Bieb.) Schmahl. *Theor. Appl. Genet.* 118, 801–809. doi: 10.1007/s00122-008-0940-y.
- Bogdanova, V. S., and Kosterin, O. E. (2006). A case of anomalous chloroplast inheritance in crosses of garden pea involving an accession of wild subspecies. *Dokl. Biol. Sci.* 406, 44–46. doi: 10.1134/S0012496606010121.
- Christie, J. R., and Beekman, M. (2016). Uniparental inheritance promotes adaptive evolution in cytoplasmic genomes. *Mol. Biol. Evol.* 34, 677–691. doi: 10.1093/molbev/msw266.
- Cornet, C., Noret, N., and Rossum, F. Van (2022). Pollinator sharing between reproductively isolated genetic lineages of *Silene nutans*. *Front. Plant Sci.* 13, 1–18. doi: 10.3389/fpls.2022.927498.
- Drouin, G., Daoud, H., and Xia, J. (2008). Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. *Mol. Phylogenet. Evol.* 49, 827–831. doi: 10.1016/j.ympev.2008.09.009.
- Greiner, S., and Bock, R. (2013). Tuning a ménage à trois: Co-evolution and co-adaptation of nuclear and organellar genomes in plants. *BioEssays* 35, 354–365. doi: 10.1002/bies.201200137.

- Greiner, S., Rauwolf, U., Meurer, J., and Herrmann, R. G. (2011). The role of plastids in plant speciation. *Mol. Ecol.* 20, 671–691. doi: 10.1111/j.1365-294X.2010.04984.x.
- Greiner, S., Sobanski, J., and Bock, R. (2014). Why are most organelle genomes transmitted maternally? *BioEssays* 37, 80–94. doi: 10.1002/bies.201400110.
- Hansen, A. K., Escobar, L. K., Gilbert, L. E., and Jansen, R. K. (2007). Paternal, maternal, and biparental inheritance of the chloroplast genome in *Passiflora* (Passifloraceae): Implications for phylogenetic studies. *Am. J. Bot.* 94, 42–46. doi: 10.3732/ajb.94.1.42.
- Hu, Y., Zhang, Q., Rao, G., and Sodmergen (2008). Occurrence of Plastids in the Sperm Cells of Caprifoliaceae : Biparental Plastid Inheritance in Angiosperms is Unilaterally Derived from Maternal Inheritance. *Plant Cell Physiol.* 49, 958–968. doi: 10.1093/pcp/pcn069.
- Jansen, R. K., and Ruhlman, T. A. (2012). Plastid genomes of seed plants. *Genomics of chloroplasts and mitochondria* 35, 103–126. doi: 10.1007/978-94-007-2920-9.
- Martin, H., Touzet, P., Dufay, M., Godé, C., Schmitt, E., Lahiani, E., et al. (2017). Lineages of *Silene nutans* developed rapid, strong, asymmetric postzygotic reproductive isolation in allopatry. *Evolution* 71, 1519–1531. doi: 10.1111/evo.13245.
- Martin, H., Touzet, P., Rossum, F. Van, Delalande, D., and Arnaud, J. (2016). Phylogeographic pattern of range expansion provides evidence for cryptic species lineages in *Silene nutans* in Western Europe. *Heredity* 116, 286–294. doi: 10.1038/hdy.2015.100.
- McCauley, D. E. (2013). Paternal leakage, heteroplasmy, and the evolution of plant mitochondrial genomes. *New Phytol.* 200, 966–977.
- McCauley, D. E., Bailey, M. F., Sherman, N. A., and Darnell, M. Z. (2005). Evidence for paternal transmission and heteroplasmy in the mitochondrial genome of *Silene vulgaris*, a gynodioecious plant. *Heredity* 95, 50–58. doi: 10.1038/sj.hdy.6800676.
- McCauley, D. E., Sundby, A. K., Bailey, M. F., and Welch, M. E. (2007). Inheritance of chloroplast DNA is not strictly maternal in *Silene vulgaris* (Caryophyllaceae): evidence from experimental crosses and natural populations. *Am. J. Bot.* 94, 1333–1337.
- Metzlaf, M., Pohlheim, F., Börner, T., and Hagemann, R. (1982). Hybrid variegation in the genus *Pelargonium*. *Curr. Genet.* 5, 245–249. doi: 10.1007/BF00391813.
- Parakatselaki, M., and Ladoukakis, E. D. (2021). mtDNA Heteroplasmy : Origin, Detection, Significance, and Evolutionary Consequences. *Life* 11, 633.
- Postel, Z., Poux, C., Gallina, S., Varré, J.-S., Godé, C., Schmitt, E., et al. (2022). Reproductive isolation among lineages of *Silene nutans* (Caryophyllaceae): A potential involvement of plastid-nuclear incompatibilities. *Mol. Phylogenet. Evol.* 169, 107436.
- Postel, Z., and Touzet, P. (2020). Cytonuclear Genetic Incompatibilities in Plant Speciation. *Plants* 9, 487.

- Radzvilavicius, A. (2021). Beyond the “selfish mitochondrion” theory of uniparental inheritance : A unified theory based on mutational variance. *BioEssays* 43, 1–9. doi: 10.1002/bies.202100009.
- Ramsey, A. J., and Mandel, J. R. (2019). When one genome is not enough: organellar heteroplasmy in plants. *Annu. Plant Rev.* 2, 1–40. doi: 10.1002/9781119312994.apr0616.
- Ramsey, A. J., McCauley, D. E., and Mandel, J. R. (2019). Heteroplasmy and Patterns of Cytonuclear Linkage Disequilibrium in Wild Carrot. *Integr. Comp. Biol.* 59, 1005–1015. doi: 10.1093/icb/icz102.
- Sakamoto, W. (2003). Leaf-variegated mutations and their responsible genes in *Arabidopsis thaliana*. *Genes Genet. Syst.* 78, 1–9. doi: 10.1266/ggs.78.1.
- Shrestha, B., Gilbert, L. E., and Ruhlman, T. A. (2021). Clade-Specific Plastid Inheritance Patterns Including Frequent Biparental Inheritance in *Passiflora* Interspecific Crosses. *Int. J. Mol. Sci.* 22, 2278.
- Sloan, D. B., Triant, D. A., Forrester, N. J., Bergner, L. M., Wu, M., and Taylor, D. R. (2014). A recurring syndrome of accelerated plastid genome evolution in the angiosperm tribe *Sileneae* (Caryophyllaceae). *Mol. Phylogenet. Evol.* 72, 82–89. doi: 10.1016/j.ympev.2013.12.004.
- Sloan, D. B., Warren, J. M., Williams, A. M., Wu, Z., Abdel-Ghany, S. E., Chicco, A. J., et al. (2018). Cytonuclear integration and co-evolution. *Nat. Rev. Genet.* 19, 635–648. doi: 10.1038/s41576-018-0035-9.
- Sobanski, J., Giavalisco, P., Fischer, A., Kreiner, J. M., Walther, D., Schöttler, M. A., et al. (2019). Chloroplast competition is controlled by lipid biosynthesis in evening primroses. *PNAS* 116, 5665–5674. doi: 10.1073/pnas.1811661116.
- Van Rossum, F., Martin, H., Le Cadre, S., Brachi, B., Christenhusz, M. J. M., and Touzet, P. (2018). Phylogeography of a widely distributed species reveals a cryptic assemblage of distinct genetic lineages needing separate conservation strategies. *Perspect. Plant Ecol. Evol. Syst.* 35, 44–51. doi: 10.1016/j.ppees.2018.10.003.
- Wade, M. J., and McCauley, D. E. (2005). Paternal Leakage Sustains the Cytoplasmic Polymorphism Underlying Gynodioecy but Remains Invasible by Nuclear Restorers. *Am. Nat.* 166, 592–602.
- Weihe, A., Apitz, J., Pohlheim, F., Salinas-Hartwig, A., and Börner, T. (2009). Biparental inheritance of plastidial and mitochondrial DNA and hybrid variegation in *Pelargonium*. *Mol. Genet. Genomics* 282, 587–593. doi: 10.1007/s00438-009-0488-9.
- Wolfe, K. H., Li, W.-H., and Sharp, P. M. (1987). Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *PNAS* 84, 9054–9058.
- Xu, J. (2005). The inheritance of organelle genes and genomes : patterns and mechanisms. *Genome* 48, 951–958. doi: 10.1139/G05-082.
- Yao, J. L., Cohen, D., and Rowland, R. E. (1995). Interspecific albino and variegated hybrids in the

genus *Zantedeschia*. *Plant Sci.* 109, 199–206. doi: 10.1016/0168-9452(95)04163-O.

Zhang, Q., Liu, Y., and Sodmergen (2003). Examination of the Cytoplasmic DNA in Male Reproductive Cells to Determine the Potential for Cytoplasmic Inheritance in 295 Angiosperm Species. *Plant Cell Physiol.* 44, 941–951. doi: 10.1093/pcp/pcg121.

## 6. Annexes

**Table S1**

Nucleotide details for the diagnostic SNPs used for assessment of the plastid haplotypes.

Plastid sequence name	SNP name	SNPs	W1 Lineage	W2 Lineage	W3 Lineage	E1 Lineage
Intergenic spacer sequence <i>psbA-trnH</i>	Cp42	G/T	T	G	T	T
	Cp397	C/A	C	C	C	C
<i>matK</i>	Cp540	C/T	C	C	T	T
	Cp656	T/G	T	T	G	G
	Cp730	C/T	C	T	C	C
	Cp804	G/T	G	G	G	G
<b>Genotype</b>			<b>TCCTCG</b>	<b>GCCTTG</b>	<b>TCTGCG</b>	<b>TACTCT</b>

**Supplementary table S2**

Details of the results obtained for the variegated individuals on the green and the white sectors. For these individuals, when possible we sampled both DNA on the white and on the green sectors of leaves.

CROSSES	NUMBER OF GENOTYPED INDIVIDUALS				PL FOR WITHE SECTORS			PL FOR GREEN SECTORS		
	Total number	Both	White only	Green only	PL	no PL	NA	PL	no PL	NA
E1 x W1	0	0	x	x	0	0	0	0	0	0
W1 x E1	5	4	1	x	0	3	1	0	4	0
E1 x W2	0	0	x	x	0	0	0	0	0	0
W2 x E1	0	0	x	x	0	0	0	0	0	0
E1 x W3	5	5	x	x	2 (P)	2	1	5	0	0
W3 x E1	1	1	x	x	0	0	1	1	0	0
W1 x W2	19	15	x	4	13	2 (P)	0	0	18	1
W2 x W1	16	15	x	1	0	15	0	15	0	1
W1 x W3	3	3	x	x	0	3	0	3	0	0
W3 x W1	0	0	x	x	0	0	0	0	0	0
W2 x W3	25	20	x	5	0	20	0	24	0	1
W3 x W2	0	0	x	x	0	0	0	0	0	0

“**Total number**”: total number of variegated individuals genotyped for plastid SNPs ; “**Both**”: individuals for which we were able to sampled both the green and white sectors of the leaf ; “**White only**”: individuals for which we were able to sample only the white sector of the leaf ; “**Green only**”: individuals for which were able to sampled only the green sector of the leaf ; “**PL**”: number of leaves fragments containing the paternal plastid genome ; “**no PL**”: number of leaves fragments containing the maternal plastid genome ; “**NA**”: number of individuals for which qPCR results were not conclusive ; “**(P)**”: white sectors that were rather pale green that totally white.



## CHAPTER 3

---

### The decoupled evolution of the organellar genomes of *Silene nutans* leads to distinct roles in the speciation process

Zoé Postel<sup>1</sup>, Daniel B. Sloan<sup>2</sup>, Sophie Gallina<sup>1</sup>, Cécile Godé<sup>1</sup>, Eric Schmitt<sup>1</sup>, Sophie Mangenot<sup>3</sup>, Laurence Maréchal-Drouard<sup>4</sup>, Jean-Stéphane Varré<sup>5</sup>, Pascal Touzet<sup>1</sup>

This chapter was submitted to *New Phytologist*



<sup>1</sup> Univ. Lille, CNRS, UMR 8198 - Evo-Eco-Paleo, F-59000 Lille, France

<sup>2</sup> Department of Biology, Colorado State University, Fort Collins, Colorado, USA

<sup>3</sup> Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, 91057, Evry, France

<sup>4</sup> Institut de Biologie Moléculaire des Plantes-CNRS, Université de Strasbourg, Strasbourg, France

<sup>5</sup> Univ. Lille, Inria, UMR CNRS 9189 - CRISAL Centre de Recherche en Informatique Signal et Automatique de Lille F-59000 Lille, France



The objective of this chapter was to compare the evolutionary patterns of the mitochondrial and plastid genomes in *Silene nutans* lineages, to assess whether the mitochondrial genome could also be involved in the reproductive isolation between lineages through mito-nuclear incompatibilities. We worked with mitochondrial and plastid genomic data on individuals of the four lineages. We conducted molecular data analysis to assess patterns of genetic diversity, selective pressures, mutation and recombination rates in mitochondrial and plastid genes.

Genomic data for the organellar genomes in the four lineages were acquired before my arrival through gene capture by Cécile Godé, Sophie Galina, Jean-Stéphane Varré and Pascal Touzet. Plastid genome assemblies were done by Sophie Galina. Mitochondrial reads processing was also done by Sophie Galina. The rest of the molecular analyses were conducted by myself in close collaboration with Sophie Galina. Choices of the analyses and interpretation of the results were done by myself with Pascal Touzet and Dan Sloan. This collaboration benefited from a mobility on my part to Dan Sloan's lab in Fort Collins (Colorado – US) for two months in the beginning of my third year of PhD. Writing of the manuscript was done by myself and Pascal Touzet with Dan Sloan's editing.



# The decoupled evolution of the organellar genomes of *Silene nutans* leads to distinct roles in the speciation process

## SUMMARY

- There is growing evidence that cytonuclear incompatibilities (i.e. disruption of cytonuclear co-adaptation) might contribute to the speciation process. In a former study, we described the possible involvement of plastid-nuclear incompatibilities in the reproductive isolation between four lineages of *Silene nutans* (Caryophyllaceae). Because organellar genomes are usually co-transmitted, we assessed whether the mitochondrial genome could also be involved in the speciation process, knowing that the gynodioecious breeding system of *S. nutans* is expected to impact the evolutionary dynamics of this genome.
- Using hybrid capture and high-throughput DNA sequencing, we analyzed diversity patterns in the genic content of the organellar genomes in the four *S. nutans* lineages.
- Contrary to the plastid genome, which exhibited a large number of fixed substitutions between lineages, extensive sharing of polymorphisms between lineages was found in the mitochondrial genome. In addition, numerous recombination-like events were detected in the mitochondrial genome, loosening the linkage disequilibrium between the organellar genomes and leading to decoupled evolution.
- These results suggest that gynodioecy shaped mitochondrial diversity through balancing selection, maintaining ancestral polymorphism and, thus, limiting the involvement of the mitochondrial genome in evolution of hybrid inviability between *S. nutans* lineages.

## KEY WORDS

Organellar genomes ; balancing selection ; cytonuclear incompatibilities ; reproductive isolation ; gynodioecy ; genome evolution ; paternal leakage

## 1. Introduction

Plants cells are composed of three distinct genomic compartments: the nucleus, the mitochondrion (mt) and the plastid (pt). These three compartments exist in a tight relationship, as both cytoplasmic organelles depend upon proper import of nuclear-encoded proteins for organellar protein complex function (Rand *et al.*, 2004; Greiner & Bock, 2013; Sloan *et al.*, 2018). Nuclear and organellar genomes differ in characteristics such as mutation rates and inheritance patterns, with the nucleus being inherited from both parents while the organelles are generally inherited from the mother (Rand *et al.*, 2004; Greiner *et al.*, 2014; Smith, 2015; Ramsey & Mandel, 2019). Tight coordination and co-adaptation are required between organellar genomes and the nuclear genes whose gene products are targeted to the organelles, with mutations accumulation in one compartment generating selection for coevolutionary changes in the other (Osada & Akashi, 2012; Havird *et al.*, 2015). This coevolved relationship can be disrupted when crossing individuals of distant lineages, leading to cytonuclear incompatibilities and dysfunctional hybrid individuals (Yao & Cohen, 2000; Bogdanova *et al.*, 2015; Barnard-Kubow *et al.*, 2016; Zupok *et al.*, 2021). As such, cytonuclear incompatibilities are often considered as post-zygotic barriers that contributes to the first steps of speciation (Burton & Barreto, 2012; Postel & Touzet, 2020).

*Silene nutans* (Caryophyllaceae) is an angiosperm with at least four genetic lineages, including an eastern lineage (E1) widespread in the north of Europe (e.g. England, Belgium, North of France) and a western groups composed of three sub-lineages: W1 distributed in England, France and Belgium, W2 restricted to Spain and southwestern France, and W3 in the Alps and Italy (Martin, 2016; Martin *et al.*, 2016; Van Rossum *et al.*, 2018). A diallel cross reported a high percentage of inter-lineage hybrid mortality and chlorotic seedlings, suggesting strong reproductive isolation among these lineages (Martin *et al.*, 2017; Postel *et al.*, 2022). The level of reproductive isolation depended on the direction of the cross and which lineage was used as the maternal parent, suggesting that cytonuclear incompatibilities may be involved. We previously scanned pt and nuclear genetic variation in *S. nutans* to search for candidate gene pairs involved in a history of cytonuclear co-adaptation (Postel *et al.*, 2022). We focused on nuclear genes encoding gene products targeted to the pt (hereafter referred to as N-pt genes). We found that pt genes in *S. nutans* accumulated a large number of non-synonymous mutations (i.e. mutations leading to a change of the encoded amino acid) that were differentially fixed between lineages and that there was a mirrored accumulation of substitutions in N-pt genes encoding subunits within the corresponding protein complexes. Mutation accumulation in pt genes was inferred to be mainly driven by relaxed selection, although signatures of positive selection were also identified for some of the pt genes (Postel *et al.*, 2022).

Because of their typical mode of maternal inheritance, pt and mt genomes are expected to be in strong linkage disequilibrium (LD) (Olson & McCauley, 2000). However, this assumption is not always met. In some angiosperm species, evidence for mito-plastid incongruence and decay of LD between organellar genomes was identified, potentially due to paternal leakage of the organellar genomes and/or subsequent mt recombination (Houliston & Olson, 2006; Lahiani *et al.*, 2013; Govindarajulu *et al.*, 2015; Adhikari *et al.*, 2019; Ramsey *et al.*, 2019). Nevertheless, previous analysis of pt and mt genomes across *Silene* species showed correlated increases in evolutionary rates, suggesting that common evolutionary forces could shape organellar genome evolution (Sloan *et al.*, 2012). These observations raise questions about whether strong LD is observed between the organellar genomes of *Silene nutans* and whether mt genes are also involved in reproductive isolation and cytonuclear incompatibilities. In the present study, we compared the evolutionary patterns of the mt and pt genomes of *S. nutans* individuals from each of the four major lineages. We first looked at nucleotide genetic diversity in mt and pt genes to test whether mutation accumulation followed the same trend between the two organellar genomes. Then we conducted tests of selection to assess whether selective forces were equivalent in both organellar genomes. We also tested for a history of recombination/reassortment within and between the organellar genomes. The outcome of the tests shows that the organellar genomes exhibit striking differences, pointing to distinct evolutionary paths with implications for the speciation process of *Silene nutans*.

## 2. Material and methods

### 2.1. Genomic data acquisition and assemblies

We acquired genomic data for both the mt and the pt genes through gene capture (Suppl. Fig. 1; see sampling details in (Postel *et al.*, 2022)). Briefly, 47 individuals from 24 populations (1–2 individuals per population) from UK, France, Belgium, Luxembourg, Germany, and Finland were sampled from the DNA collection of the unit Evo-Eco-Paleo (UMR 8198 – CNRS University of Lille; see Martin *et al.*, 2016 for DNA extraction procedure). These populations covered four genetic lineages of *S. nutans* based on pt SNP markers (Martin *et al.*, 2016), with sixteen individuals belonging to E1, fifteen individuals to W1, and eight individuals each to W2 and W3. Genomic sequences for each individual were obtained through gene capture with a myBaits® target capture kit (Daicel Arbor Biosciences, <https://arborbiosci.com/>). DNA probes were defined from the published sequence of the organellar genomes of *Silene latifolia* (NCBI accessions NC\_016730.1 and NC\_014487.1) and previous Illumina data from one mt genome of *S. nutans* (Genoscope PRJEB54044). Enriched libraries were pooled and

sequenced on an Illumina MiSeq (2 × 150 with dual indexing) at the LIGAN platform (UMR 8199 LIGAN-PM Genomics platform – Lille, France).

Because of differences in read depth between mt (mean of 50×) and pt (mean of 600×) regions, we applied different strategies to obtain gene sequences from the two genomes. For the pt genes, we assembled the reads with SPAdes v.3.0.0 (Bankevich *et al.*, 2012) and then blasted *S. latifolia* reference sequences against the assembly as described in (Postel *et al.*, 2022). Mt sequence coverage was more fragmented due to the lower read depth, so we used HybPiper (Johnson *et al.*, 2016) to extract the genes sequences, after read cleaning with Trimmomatic using the following parameters: 'LEADING:10 TRAILING:10 - SLIDINGWINDOW:4:20 - MINLEN:36'. We first ran the HybPiper pipeline only on the reads from W1 individuals, which represented the least fragmented dataset, using the *S. latifolia* mt genome as the reference. Then we ran BamBam (SAMtools package) (Page *et al.*, 2014) on these read alignments to create a new reference. Finally, we ran HybPiper on the reads of all individuals of *S. nutans* using this new W1 reference.

Mt data quality was assessed using IGV (Thorvaldsson *et al.*, 2013), MultiQC (Suppl. Fig. 2), and FastQC (Ewels *et al.*, 2016; Andrews *et al.*, 2018). To assess how complete our mt dataset was, we also counted the number of mt genes annotated in *Silene nutans* compared to those annotated in *Silene latifolia* (Sloan *et al.*, 2010).

## 2.2. Variant detection for both mt and pt data

To detect pt and mt variants, we ran the Genome Analysis Toolkit (GATK) HaplotypeCaller and performed joint genotyping using GenotypeGVCFs (v.3.8) (McKenna *et al.*, 2010) with *S. latifolia* reference genomes and a min-cover of 10. Detected variants were filtered using vcfTools v.0.1.16 with the following options: (i) only keep SNPs (`--remove-indel`), (ii) remove variant sites with a minor allele frequency (*maf*) lower than 0.001 (`--maf 0.001`), (iii) exclude variant sites with a mean read depth value of less than 10 across all included individuals (`--min-meanDP 10`), (iv) exclude genotype calls from an individual if its read depth for the site is less than 5 (`--minDP 5`), (v) another round of filtering with `--maf 0.001`, and (vi) remove variants in resulting call set with a quality score of less than 100 (`--minQ 100`).

We then reported the number of polymorphic sites, per gene: (i) within individuals (i.e. apparent cases of heteroplasmy), (ii) within lineages, (iii) shared between lineages and (iv) fixed between lineages, using an in-house BioPython script to parse the final vcf files.

## 2.3. Alignment construction and analysis

For the pt genes, 68 previously constructed gene alignments for *S. nutans* individuals were already available (Postel *et al.*, 2022). These alignments also contained the reference sequences of both *S. paradoxa* and *S. latifolia*. As mt reference sequences for *S. paradoxa* were not available, we only worked with *S. latifolia* as an outgroup and removed *S. paradoxa* from the pt gene alignment using SeqKit (Shen *et al.*, 2016). We then realigned the sequences using MUSCLE (Edgar *et al.*, 2004) and checked the alignment for correct open reading frames (Suppl. Fig. 1).

For the mitochondrial genes, we took the reference sequence build using all W1 reads and for each individual, using the variant calling results, we put the corresponding nucleotide at variable positions, using an in-house BioPython script (Figure S1). Because we detected positions that were variable within individuals for both mt and pt DNA coding sequences, the ratio of alternative alleles varied among polymorphic sites within an individual. Haplotype reconstruction within these apparently heteroplasmic individuals was not feasible with our short-read data. Thus, we constructed three different alignment datasets to deal with this issue, regarding the file format required for each analysis (Suppl. Fig. 1). Details for each of the dataset are provided below and the different version of the BioPython script used to construct these sequences are available on GitHub ([https://github.com/ZoePos/Alignement\\_construction](https://github.com/ZoePos/Alignement_construction)).

#### 2.4. McDonald Kreitman tests

We applied the McDonald Kreitman test (MKT) (McDonald & Kreitman, 1991) to detect signatures of positive selection by comparing the number of synonymous/non-synonymous polymorphic sites within *S. nutans* ( $P_s$  and  $P_n$ ) and synonymous/non-synonymous divergent sites between *S. nutans* and *S. latifolia* ( $D_s$  and  $D_n$ ). Alignments constructions for this test is described in supplementary figure S1. We used these data to estimate the proportion of substitutions fixed by natural selection ( $\alpha$ ) and calculate the Neutrality Index (NI), which quantifies the direction and degree of departure from neutrality (i.e. equality of the two ratios) (Rand & Kann, 1996). NI values  $> 1$  indicate negative (purifying) selection, whereas values  $< 1$  indicate positive selection. The mt gene *rps13* appears to have been pseudogenized in *S. latifolia* (Sloan *et al.*, 2010), so this gene was excluded from this analysis. We used PopGenome R package (Pfeifer *et al.*, 2014) to run the MKT using the options `include.unknown = TRUE` when loading the data to allow for missing data. The MKT was run on each pt and mt gene separately and for concatenations of genes encoding subunits of the same pt or mt enzyme complex. Because *S. nutans* lineages exhibit strong reproductive isolation and could be considered incipient species, we also ran the MKT on each *S. nutans* lineage separately, with *S. latifolia* as an outgroup.

### 2.5. Recombination

To assess the number of reassortment events within and between organellar genomes, we conducted four gamete tests (FGTs) on the mt and pt data. Considering that homoplasmy (i.e. repeated mutations at the same position) is expected to be rare (infinite-sites assumption), the presence of four different allelic combinations between pairs of segregating sites most likely results from recombination. Alignments constructions for this test is described in supplementary figure S1

We implemented FGTs with a BioPython script (<https://github.com/ZoePos/Four-gamete-test>) on the concatenated gene sets of all organellar genes. Genotype calls with evidence of intra-individual polymorphism (heteroplasmy) were treated as missing data.

### 2.6. Divergence ( $K_s$ ) with *S. latifolia*

To assess whether the high level of polymorphism observed in the mt data was the result of recombination or elevated mutations rates, we calculated synonymous divergence ( $K_s$ ) between *S. nutans* and *S. latifolia*. As synonymous mutations are “silent” mutations (i.e. not leading to a change of the amino acid sequences), they are expected to experience weaker selective pressures and evolve relatively neutrally. Thus, the synonymous divergence between *S. nutans* and *S. latifolia* should reflect the mutation rate. Mean  $K_s$  was estimated for each mt and pt gene, using DNAsp 6.12.03 (Rozas *et al.*, 2017), first with all *S. nutans* individuals and then separately for each lineage. To test for significant differences between pt and mt genes, we conducted Mann-Whitney U tests in R v.4.1.2 (package stats4).

### 2.7. Checking for numts

Numts (nuclear mt DNA) are mt DNA sequences transferred to the nucleus (Parakatselaki & Ladoukakis, 2021). Such transfers are ongoing processes and the insertions can be very large (Fields *et al.*, 2022). Once integrated into the nuclear genome, they are generally non-functional and subject to degradation (Hazkani-covo *et al.*, 2010). Misidentifying numts as true mt DNA is a common cause of erroneous conclusions about phylogenetic relationships, biparental inheritance of mt genomes, and de novo mutations (Bensasson *et al.*, 2001; Wu *et al.*, 2020; Lutz-bonengel *et al.*, 2021). Therefore, we assessed whether our data were contaminated with numt sequences via additional analyses summarized in the supplementary Figure 2.

First, we checked the mt DNA data and excluded mt genes that exhibited any variant that created an early stop codon, as these can indicate sequence degradation and non-functionality of the mt genes.

**Table 1**

Number of genes annotated and analyzed in *S. nutans* individuals compared to other *Silene* species. The number of genes in other *Silene* species were taken from (Sloan *et al.* 2010)

COMPLEXES	<i>S. LATIFOLIA</i>	<i>S. NUTANS</i>	ANALYZED
OXPHOS complex I	9	9	9
OXPHOS complex II	1	2	0 (2 pseudogenes)
OXPHOS complex III	1	1	1
OXPHOS complex IV	3	3	3
OXPHOS complex V	5	5	4 (numts suspicion)
Cytochrome c biogenesis	4	4	4
<i>mttB</i>	1	1	1
<i>matR</i>	1	1	0 (numts suspicion)
Large ribosomal subunit	1	1	1
Small ribosomal subunit	4 (1 pseudogene)	4	2 (numts suspicion)
rRNA genes	3	2	0
tRNA genes	7	8	0
TOTAL	40	41	24

\*: mitochondrial genes excluded because of numts suspicion (*atp6*, *matR*, *rps4*, *rps14*) ; °: mitochondrial pseudogenes in *S. latifolia* and *S. nutans* (*rps13*, *sdh3*, *sdh4*)

Second, we assessed whether any variants were found exclusively in heteroplasmic individuals, as these would be candidates for derived changes in a numt sequences that, therefore, must always co-occur with the true mt allele. Finally, we randomly selected 100 nuclear loci in the transcriptome that was available for several individuals of the four lineages (Muyle *et al.*, 2021). We randomly sampled 100 nuclear loci excluded the one that might be interacting with mt and/or pt genes, to be sure that whatever dynamics is occurring in mitochondrial/pt genomes, we would be truly looking at nuclear genes experiencing evolutionary dynamics of the nuclear genome only. We constructed alignments for these 100 nuclear loci. We also blasted these alignments on *S. latifolia* transcriptome (PRJEB39526) to use it as an outgroup and add *S. latifolia* sequences to the *S. nutans* alignments. Finally, we checked the alignments to be sure that no early stop codons were presents. For nuclear alignments, we followed methods used in Postel *et al.* (2022). With these alignments, we concatenated mt genes and nuclear loci separately and constructed phylogenies, using RAxML with a GTR gamma model of nucleotide substitutions (Stamatakis, 2014) and bootstrap analysis for node support.

### 3. Results

#### 3.1. Data quality

For pt data quality, refer to (Postel *et al.*, 2022). The read depth was lower for the mt genome (~50×) than the pt genome (~600×). Most mt genes were found using *Silene latifolia* mt genome as reference, resulting in the analysis of 27 mt protein-coding genes (Table 1).

**Table 2**

Number of polymorphic sites (including indels and SNPs) in the mitochondrial and plastid gene concatenations: (1) within individuals (at the read level); (2) within lineage; (3) shared between all lineages; (4) differently fixed between lineages.

	LINEAGE	MITOCHONDRIAL GENES				PLASTID GENES			
		Intra-individual	Private (one lineage)	Shared between lineage	Fixed substitutions	Intra-individual	One lineage	Shared between lineage	Fixed substitutions
<b>Numbers of polymorphic sites for mitochondrial and plastid gene concatenations</b>	E1	45	74			7	3		
	W1	48	24			14	1		
	W2	4	17	477	0	3	10	7	139
	W3	15	26			9	2		
<b>Proportions of polymorphic sites for mitochondrial and plastid gene concatenations</b>	E1	1.3e-3	2.1e-3			4.6e-5	2.0e-5		
	W1	1.3e-3	6.7e-4			9.2e-4	6.6e-6		
	W2	1.1e-4	4.7e-4	1.3e-2	0	2.0e-5	6.6e-5	4.6e-5	9.1e-4
	W3	4.2e-4	7.3e-4			5.9e-4	1.3e-5		

Length of the concatenation of the mitochondrial genes: 35849 bp ; Length of the concatenation of the plastid genes: 151736 bp

### 3.2. Polymorphism

Overall, more polymorphic sites were identified in mt genes than in pt genes, and this was true for all four *S. nutans* lineages analyzed individually. At the lineage level, the proportion of polymorphic sites was approximately  $10^{-3}$  for the mt genes vs  $10^{-4}$  for the pt genes (Table 2). Polymorphism was also detected within individuals, which suggests heteroplasmy. Depending on the *S. nutans* lineage, there were 6 to 20 sites in the pt genome and 170 to 365 sites in the mt genome that exhibited apparent heteroplasmy in at least one individual (Table 2).

The two organellar genomes exhibited striking differences in the amount of shared polymorphism and fixed substitutions between lineages. In the pt genes, only a few polymorphic sites were shared between lineages. In other words, most of the polymorphic sites with *S. nutans* were fixed differences between the four lineages. In contrast, we did not detect any fixed substitutions between lineages in the mt genes; instead, there were numerous shared polymorphic sites even among E1 and the western lineages (Table 2, Suppl. Table S2).

### 3.3. Selection

**Table 3**

Results for the mitochondrial and plastid gene concatenations per protein complex.

Genome	Complexes	Neutrality index	alpha	Fisher Pvalue	DoS
<b>MITOCHONDRIAL</b>	All mitochondrial genes	0.70	0.30	0.06	-
	OXPHOS complex I	1.25	-0.25	0.78	-
	OXPHOS complex IV	0.25	0.75	0.00	Positive
	OXPHOS complex V	1.55	-0.55	0.17	-
	Cytochrome C Biogenesis	0.46	0.54	0.16	-
	Membrane protein	0.00	1.00	0.12	-
	Large ribosomal subunit	0.00	1.00	1.00	-
	Small ribosomal subunit	Inf	-Inf	0.55	NA
<b>PLASTID</b>	All plastid genes	2.95	-1.95	0.00	Negative
	Photosystem I	2.22	-1.22	0.46	-
	Photosystem II	3.65	-2.65	0.06	Negative
	ATP synthase	5.14	-4.14	0.05	Negative
	Cytochrome b6/f	4.80	-3.80	0.24	-
	Rubisco	0.00	1.00	1.00	-
	NDH	4.92	-3.92	0.00	Negative
	RNA polymerase	1.23	-0.23	0.68	-
	Large ribosomal subunit	4.92	-3.92	0.00	Negative
	Small ribosomal subunit	10.75	-9.75	8.93E-07	Negative
	Other functions	1.07	-0.07	1.00	-

DoS = Direction of Selection depending of the value of the NI index: if NI > 1 = negative selection; if NI < 1 = positive selection.

We used the MKT to search for signatures of selection in pt and mt genes at *Silene nutans* level (i.e. combining all lineages). Here too, patterns differed between the two organellar genomes. Significant signatures of negative selection were detected on several pt proteins complexes (e.g. photosystem II, ATP synthase, NDH), whereas no positive selection was detected (Table 3). On the mt side, we only detected positive selection on the cytochrome c oxidase protein complex (Table 3) and more specifically on *cox3*, as well as for the concatenation of all mt genes (Suppl. Table S3).

Because *S. nutans* lineages can be considered as four sub-species given the high level of reproductive isolation between them, we also conducted MKTs independently for each lineage. On the pt side, we did not detect any lineage-specific selection patterns, except for signatures of positive selection in W1 for the concatenation of all pt genes (Suppl. Table S4). For the mt data, we detected positive selection on the concatenation of all mt genes and on the cytochrome c oxidase gene set in each of the four lineages, as we found when considering all four *S. nutans* lineages combined (Suppl. Table S3 & S4).

### 3.4. Recombination

With the FGT, we detected possible recombination/reassortment events within and between genes in both pt and mt genomes (Figure 1). The numbers of between-gene positive results of the FGT were much higher for the mt genes compared to the pt ones (i.e. 542 between pt genes vs 10,401 for the mt ones). The same pattern was observed for within-gene positive FGTs: 22 within pt genes vs 521 within mt genes. Recombination/reassortment both within and between genes was detected for all mt genes except for *rp15* (Figure 1A), which is less variable compared to the other genes. For the pt data, almost all of the 46 pt genes with any variable sites exhibited reassortment events, either within or between genes, except for *ndhB* and *ndhK* (Figure 1B). Reassortment between pt and mt genes was also detected for the same genes that exhibited recombination within each organellar genome (Figure 1C).

### 3.5. Divergence

The mean divergence between *S. nutans* and *S. latifolia* was calculated using  $K_s$ , and was significantly lower for mt genes (0.0284) than for pt genes (0.0437) (Figure 2) (Mann-Whitney U tests = 1210.5, p-value =  $4.6 \times 10^{-5}$ ). Consistently, when comparing  $K_s$  between mt and pt genes within each lineage of *S. nutans*, the mutation rate was significantly higher for each lineage in pt genes (Suppl. Fig. 3).

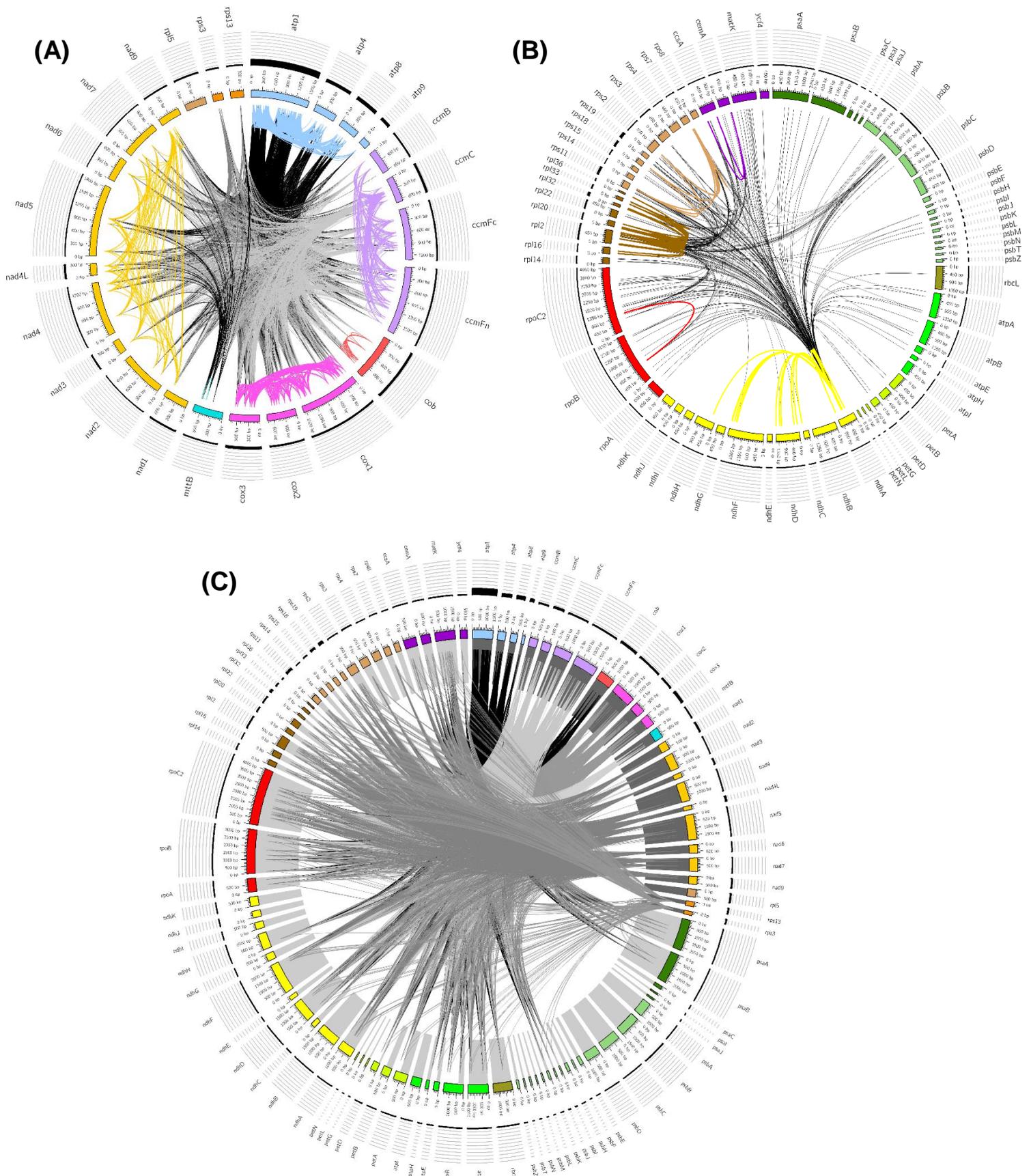
### 3.6. Are mt sequences contaminated by numts?

To assess whether the studied mt sequences were derived from numts or genuine mt genes, we asked the three following questions :

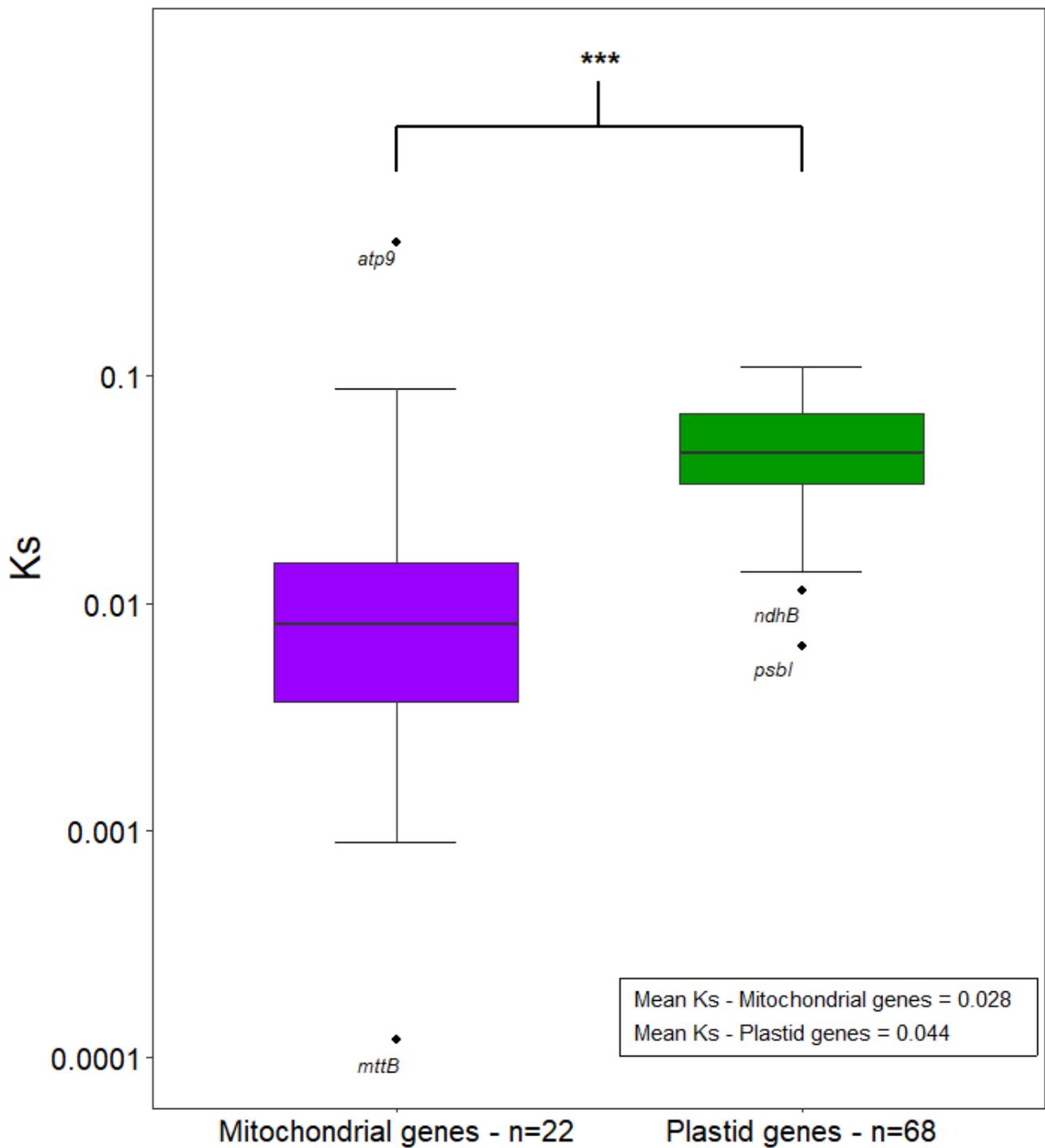
- Do we find early stop codons in genes that could be the signature of numts rather than mt genes?
- Are identified variants found exclusively in heteroplasmic individuals, as would be expected if one of the “alleles” was actually a derived change in a numt copy?
- Does the phylogeny of 100 nuclear genes sampled by chance generate a phylogeny as inconsistent with the lineage tree as the mt gene tree?

For the first question, we found only four mt genes in which early stop codons were identified in the nucleotide sequences: *atp6*, *matR*, *rsp4*, *rps14*. To be conservative, these genes were excluded from the analyses (Table 1).

For the second question, among the shared polymorphic variants, most of the heteroplasmic variants identified were also homoplasmic for other individuals: 0.2 of the shared variants were only



**Fig.1 CIRCOS figures to represent the number of recombination-like events. (A) Recombination within and between mitochondrial genes. (B) Recombination within and between plastid genes. (C) Recombination between plastid and mitochondrial genes. Black rectangle heights above each gene are proportional to the levels of polymorphism.**



**Fig.2**  $K_s$  between *S. nutans* and *S. latifolia* for plastid and mitochondrial genes. Values of  $K_s$  ( $\log_{10}$  transformation). Results of the Mann-Witney test are also shown: \*\*\* = p-value < 0.001.

heteroplasmic, the remaining 0.8 being either fully homoplasmic (proportion of 0.15) or both heteroplasmic for some individuals and homoplasmic for others (proportion of 0.65) (Suppl. Table S2). The fact that both alleles for most of the variants were found in at least some homoplasmic individuals implies that they are true mt polymorphisms. This observation does not rule out the possibility that some individual cases of apparent heteroplasmy are actually due to a numt sequence, but it does indicate that these variants arose in the mt genes and were not derived mutations that occurred in a nuclear copy. Meanwhile, the recombination data described above provides evidence that heteroplasmy is a real phenomenon in this system because it is a prerequisite for producing recombinant haplotypes.

For the final question, the nuclear gene phylogeny was highly supported (bootstrap nodes > 80%) and similar with the tree based on concatenated pt genes, with a strict clustering by lineage. Conversely, the mt phylogeny exhibited lower bootstraps values and no lineage clustering (Suppl. Fig.4). SNPS shared among lineages would be the signature of ancestral polymorphism, which does not concur with the idea of recent numts or the phylogenetic patterns recovered for known nuclear genes. Overall, even though we cannot rule out the possibility that some SNPs and apparent heteroplasmy are the result of numts, it appears that the identified variants are primarily from authentic mt sequences.

## 4. Discussion

### 4.1. Mt and pt genomes differ in their pattern of diversity

In the pt genes, most of the identified substitutions were differently fixed between lineages, whereas not a single mt variant differentiated the four lineages. Overall, there were 30-fold more shared polymorphic sites between the four lineages for mt genes than for pt genes.

Signatures of selection also differed between the organellar genomes. The MKT identified negative selection acting on most of the pt gene concatenations but not for the mt genes. These results are in accordance with previous studies comparing both genomes, where pt genes appeared to be under a stronger purifying selection than mt ones (Muse, 2000). Signatures of positive selection were identified for only one mt gene concatenation (cytochrome c oxidase) and only one mt gene (*cox3*).

The comparison of synonymous divergence ( $K_s$ ) with *S. latifolia* between pt and mt genes suggests that the pt mutation rate is higher than the mt one, which is consistent with previous findings (Drouin *et al.*, 2008). Yet, when looking at the polymorphism level within each lineage, it was higher for mt than pt genes. This could be due partly to a lower purifying selection pressure on mt genes, consistent with the result of the MKT. Furthermore, we detected many reassortment events in the mt genome, whereas they were scarce in the pt genome. Recombination might be one of the main factors shaping

the mt genome diversity in the species. Surprisingly, we also detected some recombination events between pt genes. These rare recombination event signatures in the pt genome could also be due to homoplasmy. True recombination in organelle genomes would imply the past co-occurrence of two haplotypes in a given individual, i.e. heteroplasmy, due to a shift from strict maternal inheritance of the organellar genomes (i.e. biparental inheritance or paternal "leakage") (Kmiec *et al.*, 2006; Ramsey & Mandel, 2019; Gonçalves *et al.*, 2020; Parakatselaki & Ladoukakis, 2021). Interestingly, paternal leakage of both organellar genomes has been documented in the congener *S. vulgaris* (McCauley *et al.*, 2005, 2007).

The consequence of paternal leakage and subsequent heteroplasmy might be different for pt and mt genomes. Indeed, recombination between two pt types is expected to be rare, as fusion of the plastids is uncommon (Chiu *et al.*, 1988; Birky, 2001; Sakamoto, 2003). Instead, the plastids compete with each other for fixation within the cells (Sobanski *et al.*, 2019). In contrast, mitochondria are known for frequent fusion events (Segui-Simarro *et al.*, 2008; McCauley, 2013; Greiner *et al.*, 2014; Ramsey *et al.*, 2019). Recombination between distinct mt haplotypes could then be the result of paternal leakage, increasing polymorphism within each lineage. We also detected polymorphic sites within individuals, suggesting heteroplasmy. Overall, the different outcomes for pt and mt genomes following biparental transmission could explain the distinct genetic patterns observed in the organellar genomes of *S. nutans*.

#### 4.2. Mt genetic diversity link to gynodioecy

*Silene nutans* is a gynodioecious species (i.e. presence of female and hermaphroditic individuals). Gynodioecious species often exhibit inter-genomic conflict resulting from cytoplasmic-male sterility (CMS) factors found in the mt genome (Budar *et al.*, 2003; Delph *et al.*, 2007). According to theoretical models, paternal leakage of the mt could maintain stable mt polymorphism and limit the risk of stochastic loss of CMS factors, which would be beneficial for the maintenance of gynodioecy (Wade & McCauley, 2005). Additionally, this inheritance pattern offers opportunity for heteroplasmy and recombination among mt haplotypes, which in turn can lead to the emergence of new CMS genes (McCauley & Olson, 2008). Finally, paternal mt genomes are more likely to carry fertile cytotype than maternal ones which would reduce the probability of local extinction due to pollen limitation (McCauley, 2013; Breton & Stewart, 2015; Ramsey & Mandel, 2019). Because of that, gynodioecy could select for paternal leakage of the mt genome (McCauley, 2013; Breton & Stewart, 2015; Ramsey & Mandel, 2019). We previously hypothesized that selection favouring sterilizing mt haplotypes might have led to hitchhiking and fixation of pt haplotypes (Postel *et al.*, 2022). However, the low observed LD between pt and mt genomes likely limits the effects of linked selection between the two genomes.

Maintenance of gynodioecy has been hypothesized to result from two different mechanisms: maintenance of CMS factor through balancing selection with negative frequency-dependence or recurrent invasion of CMS factors in population (epidemic-like dynamics) (Ingvarsson & Taylor, 2002; Stadler & Delph, 2002; Touzet & Delph, 2009). These mechanisms would result in distinct patterns of genetic diversity of the mt genome. The level of mt polymorphism, and in particular shared between lineages, corresponding to trans-specific polymorphism, is in favour of the hypothesis of balancing selection to explain the maintenance of gynodioecy in *S. nutans*, in accordance with previous studies that were conducted on two mt genes (Touzet & Delph, 2009; Lahiani *et al.*, 2013).

#### 4.3. The decoupled evolutionary pathways of the pt and mt genomes in *Silene nutans*

This study highlights that mt and pt genomes in *S. nutans* lineages do not share a common evolutionary history. LD between the organellar genomes seems to be loose enough to result in two decoupled evolutionary pathways. When considering pt genetic diversity at the lineage level, a large number of non-synonymous fixed mutations were documented, pointing to the involvement of plastid-nuclear incompatibilities in reproductive isolation between *S. nutans* lineages (Postel *et al.*, 2022). The low-recombining nature of the pt genome might favor the occurrence of speciation genes just as it is the case in any low recombining genomic regions such as for example chromosomal inversions (Schluter & Rieseberg, 2022). In contrast, no fixed mt variants were detected between lineages, with balancing selection most likely preventing the effect of drift observed on the pt genome. Balancing selection of the mt genome in a gynodioecious species might limit lineage specific co-evolution between the mt and the nuclear genomes and thus the accumulation of mito-nuclear incompatibilities. This could also imply that CMS mt genomes and their corresponding nuclear restorer genes might be distributed and maintained at the level of the *Silene nutans* species complex and would not contribute to post-zygotic isolation between lineages, in contrast to what was described in *Mimulus* (Fishman & Willis, 2006; Case *et al.*, 2016).

Thus, unlike the general case of strong LD between the organellar genomes, which can make it difficult to identify which genome might be involved in a given trait, the present study and prior work (Postel & Touzet, 2020) suggest a larger role for the plastid genome than the mt genome in post-zygotic reproductive isolation in the *S. nutans* species complex.

## ACKNOWLEDGEMENTS

This work has been performed using infrastructure and technical support of the Plateforme Serre, cultures et terrains expérimentaux – Université de Lille for the greenhouse/field facilities. The authors wish to thank Elodie Ubrig (IBMP) for the purification of mt DNA of *S. nutans* for the preliminary Illumina sequencing. This study was funded by Genoscope, the Agence Nationale de la Recherche (ANR-11-BSV7-013-03 ; TRANS) and CNRS via a PEPS-BMI grant. PT thanks the Région Hauts-de-France, and the Ministère de l'Enseignement Supérieur et de la Recherche (CPER Climibio), and the European Fund for Regional Economic Development for their financial support. ZP is the recipient of a PhD grant from the Ministère de l'Enseignement Supérieur et de la Recherche and the Région Hauts-de-France and thanks the I-Site Lille Nord Europe for a mobility grant to work in DBS's lab. DBS was supported by a grant from the US National Science Foundation (MCB-2048407).

## 5. References

- Adhikari B, Caruso CM, Case AL. 2019.** Beyond balancing selection: frequent mitochondrial recombination contributes to high-female frequencies in gynodioecious *Lobelia siphilitica* (Campanulaceae). *New Phytologist* **224**: 1381–1393.
- Andrews S, Wingett SW, Hamilton RS. 2018.** FastQ Screen : A tool for multi-genome mapping and quality control. *F1000Research* **7**: 1–13.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al. 2012.** SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* **19**: 455–477.
- Barnard-Kubow KB, So N, Galloway LF. 2016.** Cytonuclear incompatibility contributes to the early stages of speciation. *Evolution* **70**: 2752–2766.
- Bensasson D, Zhang D, Hartl DL, Hewitt GM. 2001.** Mitochondrial pseudogenes : evolution's misplaced witnesses. *Trends in Ecology & Evolution* **16**: 314–321.
- Bergero R, Levsen N, Wolff K, Charlesworth D. 2019.** Arms races with mitochondrial genome soft sweeps in a gynodioecious plant, *Plantago lanceolata*. *Molecular Ecology* **28**: 2772–2785.
- Birky CW. 2001.** The inheritance of genes in mitochondria and chloroplasts: Laws, Mechanisms and Models. *Annual Review of Genetics* **35**: 125–148.
- Bogdanova VS, Zaytseva OO, Mglinets A V, Shatskaya N V, Kosterin OE, Vasiliev G V. 2015.** Nuclear-Cytoplasmic Conflict in Pea (*Pisum sativum* L .) Is Associated with Nuclear and Plastidic Candidate Genes Encoding Acetyl- CoA Carboxylase Subunits. *PLOS ONE* **10**: 1–18.
- Breton S, Stewart DT. 2015.** Atypical mitochondrial inheritance patterns in eukaryotes. *Genome* **58**:

423–431.

**Burton RS, Barreto FS. 2012.** A disproportionate role for mtDNA in Dobzhansky-Muller incompatibilities? *Molecular Ecology* **21**: 4942–4957.

**Case AL, Finseth FR, Barr CM, Fishman L. 2016.** Selfish evolution of cytonuclear hybrid incompatibility in *Mimulus*. *Proceedings of the Royal Society B: Biological Sciences* **283**: 20161493.

**Chiu W-L, Stubbe W, Sears BB. 1988.** Plastid inheritance in *Oenothera*: organelle genome modifies the extent of biparental plastid transmission. *Current Genetics* **13**: 181–189.

**Drouin G, Daoud H, Xia J. 2008.** Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. *Molecular Phylogenetics and Evolution* **49**: 827–831.

**Edgar RC, Drive RM, Valley M. 2004.** MUSCLE : multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792–1797.

**Ewels P, Lundin S, Max K. 2016.** MultiQC : summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**: 3047–3048.

**Fields PD, Waneka G, Naish M, Schatz MC, Henderson IR, Daniel B. 2022.** Complete sequence of a 641-kb insertion of mitochondrial DNA in the *Arabidopsis thaliana* nuclear genome. *bioRxiv*.

**Fishman L, Willis JH. 2006.** Cytonuclear Incompatibility Causes Anther Sterility in *Mimulus* Hybrids. *Evolution* **60**: 1372.

**Gonçalves DJP, Jansen RK, Ruhlman TA, Mandel JR. 2020.** Under the rug : Abandoning persistent misconceptions that obfuscate organelle evolution. *Molecular Phylogenetics and Evolution* **151**: 106903.

**Govindarajulu R, Parks M, Tennessen JA, Liston A, Ashman T. 2015.** Comparison of nuclear, plastid and mitochondrial phylogenies and the origin of wild octoploid strawberry species. *American Journal of Botany* **102**: 544–554.

**Greiner S, Bock R. 2013.** Tuning a ménage à trois: Co-evolution and co-adaptation of nuclear and organellar genomes in plants. *BioEssays* **35**: 354–365.

**Greiner S, Sobanski J, Bock R. 2014.** Why are most organelle genomes transmitted maternally? *BioEssays* **37**: 80–94.

**Havird JC, Whitehill NS, Snow CD, Sloan DB. 2015.** Conservative and compensatory evolution in oxidative phosphorylation complexes of angiosperms with highly divergent rates of mitochondrial genome evolution. *Evolution* **69**: 3069–3081.

**Hazkani-covo E, Zeller RM, Martin W. 2010.** Molecular Poltergeists : Mitochondrial DNA Copies (*numts*) in Sequenced Nuclear Genomes. *PLoS Genetics* **6**: e1000834.

**Houliston GJ, Olson MS. 2006.** Nonneutral Evolution of Organelle Genes in *Silene vulgaris*. *Genetics* **174**: 1983–1994.

**Ingvarsson K, Taylor DR. 2002.** Genealogical evidence for epidemics of selfish genes. *PNAS* **99**: 11265–

11269.

**Johnson MG, Gardner EM, Liu Y, Medina R, Goffinet B, Shaw AJ, Zerega NJC, Wickett NJ. 2016.** HybPiPer: extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Applications in Plant Sciences* **4**: 1–7.

**Kmiec B, Woloszynska M, Janska H. 2006.** Heteroplasmy as a common state of mitochondrial genetic information in plants and animals. *Current Genetics* **50**: 149–159.

**Lahiani E, Dufay M, Castric V, Le Cadre S, Charlesworth D, Van Rossum F, Touzet P. 2013.** Disentangling the effects of mating systems and mutation rates on cytoplasmic diversity in gynodioecious *Silene nutans* and dioecious *Silene otites*. *Heredity* **111**: 157–164.

**Lutz-bonengel S, Niederst H, Naue J, Koziel R, Yang F, Timo S, Huber G, Berger C, Strobl C, Xavier C, et al. 2021.** Evidence for multi-copy Mega-NUMTs in the human genome. *Nucleic Acids Research* **49**: 1517–1531.

**Martin H. 2016.** Processus de spéciation et impact des systèmes de reproduction dans le genre *Silene*.

**Martin H, Touzet P, Dufay M, Godé C, Schmitt E, Lahiani E, Delph LF, Van Rossum F. 2017.** Lineages of *Silene nutans* developed rapid, strong, asymmetric postzygotic reproductive isolation in allopatry. *Evolution* **71**: 1519–1531.

**Martin H, Touzet P, Rossum F Van, Delalande D, Arnaud J. 2016.** Phylogeographic pattern of range expansion provides evidence for cryptic species lineages in *Silene nutans* in Western Europe. *Heredity* **116**: 286–294.

**McCauley DE. 2013.** Paternal leakage, heteroplasmy, and the evolution of plant mitochondrial genomes. *New Phytologist* **200**: 966–977.

**McCauley DE, Bailey MF, Sherman NA, Darnell MZ. 2005.** Evidence for paternal transmission and heteroplasmy in the mitochondrial genome of *Silene vulgaris*, a gynodioecious plant. *Heredity* **95**: 50–58.

**McCauley DE, Olson MS. 2008.** Do recent findings in plant mitochondrial molecular and population genetics have implications for the study of gynodioecy and cytonuclear conflict? *Evolution* **62**: 1013–1025.

**McCauley DE, Sundby AK, Bailey MF, Welch ME. 2007.** Inheritance of chloroplast DNA is not strictly maternal in *Silene vulgaris* (Caryophyllaceae): evidence from experimental crosses and natural populations. *American Journal of Botany* **94**: 1333–1337.

**McDonald JH, Kreitman M. 1991.** Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**: 652–654.

**Mckenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010.** The Genome Analysis Toolkit : A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**: 1297–1303.

**Muse S V. 2000.** Examining rates and patterns of nucleotide substitution in plants. *Plant Molecular Biology* **42**: 25–43.

**Muyle A, Martin H, Zemp N, Mollion M, Gallina S, Tavares R, Silva A, Bataillon T, Widmer A, Glémin S, et al. 2021.** Dioecy Is Associated with High Genetic Diversity and Adaptation Rates in the Plant Genus *Silene*. *Molecular Biology and Evolution* **38**: 805–818.

**Olson MS, McCauley DE. 2000.** Linkage disequilibrium and phylogenetic congruence between chloroplast and mitochondrial haplotypes in *Silene vulgaris*. *Proceedings of the Royal Society of London B: Biological Sciences* **267**: 1801–1808.

**Osada N, Akashi H. 2012.** Mitochondrial – Nuclear Interactions and Accelerated Compensatory Evolution : Evidence from the Primate Cytochrome c Oxidase Complex. *Molecular Biology and Evolution* **29**: 337–346.

**Page JT, Liechty ZS, Huynh MD, Udall JA. 2014.** BamBam: Genome sequence analysis tools for biologists. *BMC Research Notes* **7**: 1–5.

**Parakatselaki M, Ladoukakis ED. 2021.** mtDNA Heteroplasmy : Origin, Detection, Significance, and Evolutionary Consequences. *Life* **11**: 633.

**Pfeifer B, Wittelsbu U, Ramos-onsins SE, Lercher MJ. 2014.** PopGenome : An Efficient Swiss Army Knife for Population Genomic Analyses in R. *Molecular Biology and Evolution* **31**: 1929–1936.

**Postel Z, Poux C, Gallina S, Varré J-S, Godé C, Schmitt E, Meyer E, Van Rossum F, Touzet P. 2022.** Reproductive isolation among lineages of *Silene nutans* (Caryophyllaceae): A potential involvement of plastid-nuclear incompatibilities. *Molecular Phylogenetics and Evolution* **169**: 107436.

**Postel Z, Touzet P. 2020.** Cytonuclear Genetic Incompatibilities in Plant Speciation. *Plants* **9**: 487.

**Ramsey AJ, Mandel JR. 2019.** When one genome is not enough: organellar heteroplasmy in plants. *Annual Plant Reviews* **2**: 1–40.

**Ramsey AJ, McCauley DE, Mandel JR. 2019.** Heteroplasmy and Patterns of Cytonuclear Linkage Disequilibrium in Wild Carrot. *Integrative and Comparative Biology* **59**: 1005–1015.

**Rand DM, Haney RA, Fry AJ. 2004.** Cytonuclear coevolution: The genomics of cooperation. *Trends in Ecology and Evolution* **19**: 645–653.

**Rand DA, Kann LM. 1996.** Excess Amino Acid Polymorphism in Mitochondrial DNA: Contrasts Among Genes from *Drosophila*, Mice, and Humans. *Molecular Biology and Evolution* **13**: 735–748.

**Van Rossum F, Martin H, Le Cadre S, Brachi B, Christenhusz MJM, Touzet P. 2018.** Phylogeography of a widely distributed species reveals a cryptic assemblage of distinct genetic lineages needing separate conservation strategies. *Perspectives in Plant Ecology, Evolution and Systematics* **35**: 44–51.

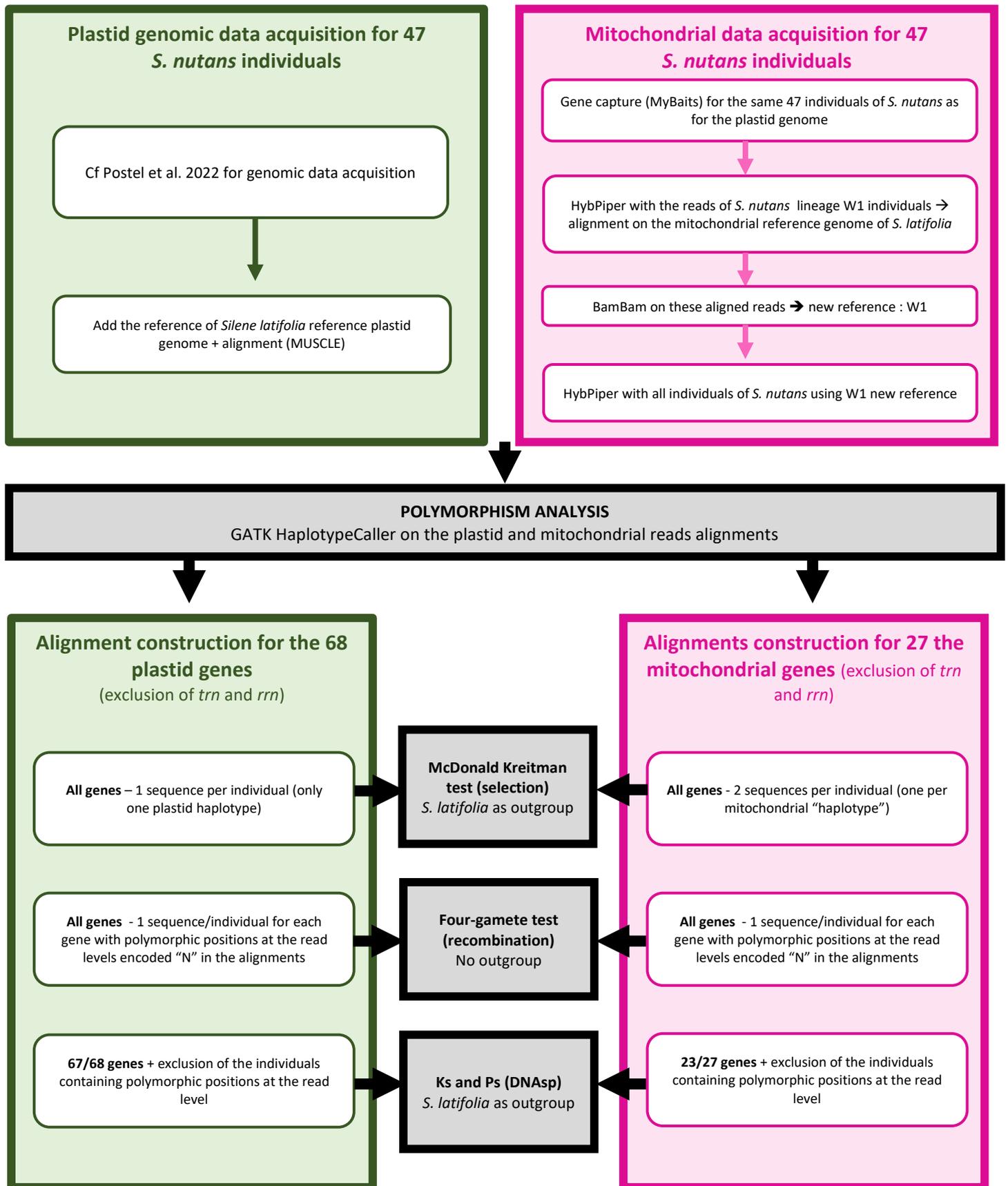
**Rozas J, Ferrer-mata A, S JC, Guirao-rico S, Librado P, Ramos-onsins E, Alejandro S-G. 2017.** DnaSP 6 : DNA Sequence Polymorphism Analysis of Large Data Sets. *Molecular Biology and Evolution* **34**: 3299–3302.

- Sakamoto W. 2003.** Leaf-variegated mutations and their responsible genes in *Arabidopsis thaliana*. *Genes and Genetic Systems* **78**: 1–9.
- Schluter D, Rieseberg LH. 2022.** Three problems in the genetics of speciation by selection. *PNAS* **119**: e2122153119.
- Segui-Simarro JM, Coronado MJ, Staehelin LA. 2008.** The Mitochondrial Cycle of Arabidopsis Shoot Apical Meristem and Leaf Primordium Meristematic Cells Is Defined by a Perinuclear Tentaculate/Cage-Like Mitochondrion. *Plant Physiology* **148**: 1380–1393.
- Shen W, Le S, Li Y, Hu F. 2016.** SeqKit : A Cross-Platform and Ultrafast Toolkit for FASTA / Q File Manipulation. *PLoS ONE* **11**: 1–10.
- Sloan DB, Alverson AJ, Helena Š, Palmer JD, Taylor DR. 2010.** Extensive loss of translational genes in the structurally dynamic mitochondrial genome of the angiosperm *Silene latifolia*. *BMC Evolutionary Biology* **10**: 274.
- Sloan DB, Alverson AJ, Wu M, Palmer JD, Taylor DR. 2012.** Recent acceleration of plastid sequence and structural evolution coincides with extreme mitochondrial divergence in the angiosperm genus *Silene*. *Genome Biology and Evolution* **4**: 294–306.
- Sloan DB, Warren JM, Williams AM, Wu Z, Abdel-Ghany SE, Chicco AJ, Havird JC. 2018.** Cytonuclear integration and co-evolution. *Nature Reviews Genetics* **19**: 635–648.
- Smith DR. 2015.** Mutation rates in plastid genomes: They are lower than you might think. *Genome Biology and Evolution* **7**: 1227–1234.
- Sobanski J, Giavalisco P, Fischer A, Kreiner JM, Walther D, Schöttler MA, Pellizzer T, Golczyk H, Obata T, Bock R, et al. 2019.** Chloroplast competition is controlled by lipid biosynthesis in evening primroses. *PNAS* **116**: 5665–5674.
- Stadler T, Delph LF. 2002.** Ancient mitochondrial haplotypes and evidence for intragenic recombination in a gynodioecious plant. *PNAS* **99**: 11730–11735.
- Stamatakis A. 2014.** RAxML version 8 : a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313.
- Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013.** Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Briefings in Bioinformatics* **14**: 178–192.
- Touzet P, Delph LF. 2009.** The Effect of Breeding System on Polymorphism in Mitochondrial Genes of *Silene*. *Genetics* **644**: 631–644.
- Wade MJ, McCauley DE. 2005.** Paternal Leakage Sustains the Cytoplasmic Polymorphism Underlying Gynodioecy but Remains Invasible by Nuclear Restorers. *The American Naturalist* **166**: 592–602.
- Wu Z, Waneka G, Broz AK, King CR, Sloan DB. 2020.** MSH1 is required for maintenance of the low mutation rates in plant mitochondrial and plastid genomes. *PNAS* **117**: 16448–16455.
- Yao JL, Cohen D. 2000.** Multiple gene control of plastome-genome incompatibility and plastid DNA

inheritance in interspecific hybrids of *Zantedeschia*. *Theoretical and Applied Genetics* **101**: 400–406.

**Zupok A, Kozul D, Schöttler MA, Niehörster J, Garbsch F, Liere K, Fischer A, Zoschke R, Malinova I, Bock R, et al. 2021.** A photosynthesis operon in the chloroplast genome drives speciation in evening primroses. *Plant Cell* **koab155**.

## 6. Annexes



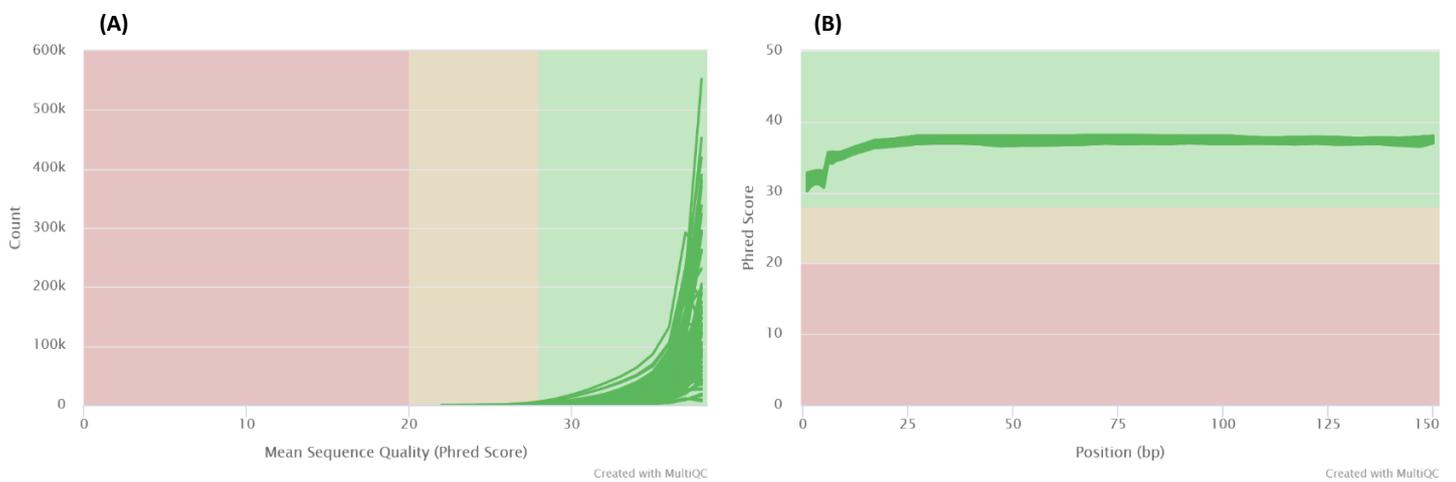
Suppl. Fig.1 Workflow of the analyses for the plastid and mitochondrial genes.

**Suppl. Table S1**

Number of individuals per gene for calculating  $K_s$

GENOME	COMPLEXES	GENES	TOTAL	E1	W1	W2	W3
MITOCHONDRIAL	OXPHOS complex I	<i>nad1</i>	27	8	7	5	7
		<i>nad2</i>	41	14	14	5	8
		<i>nad3</i>	33	11	11	4	7
		<i>nad4</i>	46	15	15	8	8
		<i>nad4L</i>	28	9	11	4	4
		<i>nad5</i>	42	13	15	6	8
		<i>nad6</i>	45	16	15	6	8
		<i>nad7</i>	36	13	12	5	6
		<i>nad9</i>	46	16	14	8	8
	OXPHOS complex III	<i>cob</i>	31	13	10	3	5
	OXPHOS complex V	<i>atp1</i>	10	0	3	1	6
		<i>atp4</i>	22	13	5	1	3
		<i>atp8</i>	37	10	12	7	8
		<i>atp9</i>	32	11	7	6	8
	OXPHOS complex IV	<i>cox1</i>	28	13	6	5	4
		<i>cox2</i>	37	6	15	8	8
		<i>cox3</i>	32	14	12	5	1
	Cytochrome biogenesis	<i>ccmB</i>	39	13	12	7	7
		<i>ccmC</i>	45	16	14	7	8
		<i>ccmFc</i>	43	16	12	7	8
		<i>ccmFn</i>	41	14	14	5	8
	Membrane protein	<i>mttB</i>	47	16	15	8	8
Large ribosomal subunit	<i>rpl5</i>	46	16	14	8	8	
Small ribosomal subunit	<i>rps13</i>	37	13	9	8	7	
	<i>rps3</i>	44	13	15	8	8	
PLASTID	Photosystem I	<i>psaA</i>	47	16	15	8	8
		<i>psaB</i>	47	16	15	8	8
		<i>psaC</i>	47	16	15	8	8
		<i>psaI</i>	47	16	15	8	8
		<i>psaJ</i>	47	16	15	8	8
	Photosystem II	<i>psbA</i>	47	16	15	8	8
		<i>psbB</i>	47	16	15	8	8
		<i>psbC</i>	47	16	15	8	8
		<i>psbD</i>	47	16	15	8	8
		<i>psbE</i>	47	16	15	8	8
		<i>psbF</i>	47	16	15	8	8
		<i>psbH</i>	47	16	15	8	8
		<i>psbI</i>	47	16	15	8	8
		<i>psbJ</i>	47	16	15	8	8
		<i>psbK</i>	47	16	15	8	8
		<i>psbL</i>	47	16	15	8	8
		<i>psbM</i>	47	16	15	8	8
		<i>psbN</i>	47	16	15	8	8
		<i>psbT</i>	47	16	15	8	8
		<i>psbZ</i>	47	16	15	8	8
	ATP synthase	<i>atpA</i>	47	16	15	8	8
		<i>atpB</i>	47	16	15	8	8

	<i>atpE</i>	47	16	15	8	8
	<i>atpH</i>	47	16	15	8	8
	<i>atpI</i>	47	16	15	8	8
Cytochrome b6/f	<i>petA</i>	47	16	15	8	8
	<i>petB</i>	47	16	15	8	8
	<i>petD</i>	47	16	15	8	8
	<i>petG</i>	47	16	15	8	8
	<i>petL</i>	47	16	15	8	8
	<i>petN</i>	47	16	15	8	8
Rubisco	<i>rbcl</i>	47	16	15	8	8
NADH	<i>ndhA</i>	47	16	15	8	8
	<i>ndhB</i>	47	16	15	8	8
	<i>ndhC</i>	47	16	15	8	8
	<i>ndhD</i>	47	16	15	8	8
	<i>ndhE</i>	47	16	15	8	8
	<i>ndhF</i>	47	16	15	8	8
	<i>ndhG</i>	47	16	15	8	8
	<i>ndhH</i>	46	16	15	8	7
	<i>ndhI</i>	47	16	15	8	8
	<i>ndhJ</i>	47	16	15	8	8
	<i>ndhK</i>	46	16	15	8	7
RNA polymerase	<i>rpoA</i>	47	16	15	8	8
	<i>rpoB</i>	47	16	15	8	8
	<i>rpoC2</i>	4	1	2	1	0
Large ribosomal subunit	<i>rpl14</i>	47	16	15	8	8
	<i>rpl16</i>	47	16	15	8	8
	<i>rpl2</i>	47	16	15	8	8
	<i>rpl20</i>	47	16	15	8	8
	<i>rpl22</i>	44	16	15	8	5
	<i>rpl32</i>	46	16	15	7	8
	<i>rpl33</i>	46	16	14	8	8
	<i>rpl36</i>	47	16	15	8	8
Small ribosomal subunit	<i>rps11</i>	47	16	15	8	8
	<i>rps14</i>	47	16	15	8	8
	<i>rps15</i>	47	16	15	8	8
	<i>rps18</i>	46	15	15	8	8
	<i>rps19</i>	46	16	15	8	7
	<i>rps2</i>	47	16	15	8	8
	<i>rps3</i>	45	16	14	8	7
	<i>rps4</i>	47	16	15	8	8
	<i>rps7</i>	47	16	15	8	8
	<i>rps8</i>	47	16	15	8	8
Other	<i>ccsA</i>	47	16	15	8	8
	<i>cemA</i>	45	16	13	8	8
	<i>matK</i>	47	16	15	8	8
	<i>ycf4</i>	47	16	15	8	8



**Suppl. Fig.2 Per sequence (A) and per base (B) quality scores plots, generated with MultiQC, after reads cleaning.** The paired reads (R1 and R2) for all individuals all *S. nutans* are displayed (i.e. one line = one individuals paired read). The different background colors of the plots refer to “very good quality calls” in green; reasonable quality in orange; “calls of poor quality” in red. Quality scores are calculated as the probability that the corresponding nucleotide was called.

**Suppl. Table S2**

Details for the shared polymorphism between lineages.

<b>GENOME</b>	<b>LINEAGES</b>	<b>Homoplasmic variants</b>	<b>Heteroplasmic variants</b>	<b>Homoplasmic + Heteroplasmic variants</b>	<b>Total number of shared variants</b>
<b>MITOCHONDRIAL</b>	<b>All lineages</b>	16	8	156	180
	<b>E1 + W1</b>	3	13	14	30
	<b>E1 + W2</b>	5	1	9	15
	<b>E1 + W3</b>	1	3	8	12
	<b>E1 + W1 + W2</b>	2	8	27	37
	<b>E1 + W1 + W3</b>	19	11	28	58
	<b>E1 + W2 + W3</b>	0	1	1	2
	<b>Western lineages</b>	6	11	23	40
	<b>W1 + W2</b>	6	31	15	52
	<b>W1 + W3</b>	9	14	22	45
	<b>W2 + W3</b>	3	1	2	6
	<b>TOTAL</b>		70	102	305
<b>PLASTID</b>	<b>All lineages</b>	0	1	0	1
	<b>E1 + W1</b>	0	2	0	2
	<b>E1 + W3</b>	0	1	0	1
	<b>W1 + W3</b>	0	2	1	3
	<b>TOTAL</b>		0	6	1

**Homoplasmic variants:** The two alternative alleles are only found in homoplasmic individuals; no heteroplasmic individuals identified (e.g. individuals have either an A or a T but not both) ; **Heteroplasmic variants:** One of the alleles is exclusively found in “heteroplasmic” individuals (e.g. individuals can be homoplasmic for A or heteroplasmic for A/T, but no individuals are homoplasmic for T). ; **Homoplasmic + heteroplasmic variants:** Both alleles are found in some homoplasmic and some heteroplasmic individuals (e.g. individuals with heteroplasmic A/T, homoplasmic A, and homoplasmic T are all present).

**Suppl. Table S3**

Results of the MK test calculated with all individuals of *S. nutans* for mt and pt genes per protein complex. Red and blue text indicates significant ( $p < .05$ ) and marginally significant ( $p < 0.1$ ) results, respectively.

Genome	Complexes	Genes	Neutrality index	DoS	alpha	Fisher Pvalue	
MITOCHONDRIAL	OXPHOS complex I	<i>nad1</i>	Inf	na	-Inf	1.00	
		<i>nad2</i>	3.00	-	-2.00	1.00	
		<i>nad3</i>	0.00	-	1.00	0.46	
		<i>nad4</i>	Inf	na	-Inf	1.00	
		<i>nad4L</i>	0.17	-	0.83	0.42	
		<i>nad5</i>	4.67	-	-3.67	0.31	
		<i>nad6</i>	na	na	na	1.00	
		<i>nad7</i>	na	na	na	1.00	
		<i>nad9</i>	0.00	-	1.00	1.00	
	OXPHOS complex III	<i>cob</i>	0.17	-	0.83	0.11	
	OXPHOS complex IV	<i>cox1</i>	1.08	-	-0.08	1.00	
		<i>cox2</i>	0.20	-	0.80	0.16	
		<i>cox3</i>	0.00	Positive	1.00	0.01	
	OXPHOS complex V	<i>atp1</i>	0.89	-	0.11	0.83	
		<i>atp4</i>	Inf	na	-Inf	0.41	
		<i>atp8</i>	0.71	-	0.29	1.00	
		<i>atp9</i>	1.36	-	-0.36	1.00	
	Cytochrome C Biogenesis	<i>ccmB</i>	0.60	-	0.40	1.00	
		<i>ccmC</i>	0.00	-	1.00	0.37	
		<i>ccmFc</i>	0.49	-	0.51	0.47	
		<i>ccmFn</i>	0.88	-	0.13	1.00	
	Membrane protein	<i>mttB</i>	0.00	-	1.00	0.12	
	Large ribosomal subunit	<i>rpl5</i>	0.00	-	1.00	1.00	
	Small ribosomal subunit	<i>rps3</i>	Inf	na	-Inf	0.55	
	PLASTID	Photosystem I	<i>psaA</i>	5.00	-	-4.00	0.33
			<i>psaB</i>	0.00	-	1.00	1.00
<i>psaC</i>			na	na	na	1.00	
<i>psaI</i>			na	na	na	na	
<i>psaJ</i>			na	na	na	1.00	
Photosystem II		<i>psbA</i>	na	na	na	1.00	
		<i>psbB</i>	5.50	-	-4.50	0.28	
		<i>psbC</i>	0.00	-	1.00	1.00	
		<i>psbD</i>	na	na	na	1.00	
		<i>psbE</i>	na	na	na	1.00	
		<i>psbF</i>	na	na	na	1.00	
		<i>psbH</i>	na	na	na	1.00	
		<i>psbI</i>	na	na	na	1.00	
		<i>psbJ</i>	Inf	na	-Inf	0.33	
		<i>psbK</i>	na	na	na	1.00	
		<i>psbL</i>	na	na	na	na	
		<i>psbM</i>	na	na	na	na	
		<i>psbN</i>	na	na	na	1.00	
		<i>psbT</i>	na	na	na	1.00	
		<i>psbZ</i>	na	na	na	1.00	
Rubisco		<i>rbcL</i>	0.00	-	1.00	1.00	

ATP synthase	<i>atpA</i>	Inf	na	-Inf	0.01
	<i>atpB</i>	0.00	-	1.00	0.53
	<i>atpE</i>	Inf	na	-Inf	0.38
	<i>atpH</i>	Inf	na	-Inf	0.33
	<i>atpI</i>	na	na	na	1.00
Cytochrome b6/f	<i>petA</i>	Inf	na	-Inf	0.38
	<i>petB</i>	na	na	na	1.00
	<i>petD</i>	4.00	-	-3.00	1.00
	<i>petG</i>	na	na	na	na
	<i>petL</i>	na	na	na	1.00
	<i>petN</i>	na	na	na	na
NDH	<i>ndhA</i>	5.00	-	-4.00	0.16
	<i>ndhB</i>	na	na	na	1.00
	<i>ndhC</i>	Inf	na	-Inf	0.25
	<i>ndhD</i>	4.80	-	-3.80	0.34
	<i>ndhE</i>	na	na	na	1.00
	<i>ndhF</i>	1.06	-	-0.06	1.00
	<i>ndhG</i>	na	na	na	1.00
	<i>ndhH</i>	3.00	-	0.48	na
	<i>ndhI</i>	na	na	na	1.00
	<i>ndhJ</i>	na	na	na	1.00
	<i>ndhK</i>	Inf	-	0.29	na
	RNA polymerase	<i>rpoA</i>	0.00	-	1.00
<i>rpoB</i>		1.55	-	-0.55	0.73
<i>rpoC2</i>		1.43	-	0.59	na
Large ribosomal subunit	<i>rpl14</i>	3.00	-	-2.00	1.00
	<i>rpl16</i>	Inf	na	-Inf	0.13
	<i>rpl2</i>	1.50	-	-0.50	1.00
	<i>rpl20</i>	2.86	-	-1.86	0.60
	<i>rpl22</i>	Inf	-Inf	0.24	na
	<i>rpl32</i>	Inf	na	-Inf	0.07
	<i>rpl33</i>	Inf	na	-Inf	0.46
	<i>rpl36</i>	na	na	NA	1.00
Small ribosomal subunit	<i>rps11</i>	Inf	na	-Inf	0.01
	<i>rps14</i>	NA	na	na	1.00
	<i>rps15</i>	Inf	na	-Inf	0.51
	<i>rps18</i>	Inf	-	0.14	na
	<i>rps19</i>	6.75	-	0.24	na
	<i>rps2</i>	4.29	-	-3.29	0.33
	<i>rps3</i>	Inf	-	0.20	na
	<i>rps4</i>	na	na	na	1.00
	<i>rps7</i>	na	na	na	1.00
	<i>rps8</i>	na	na	na	1.00
	Other	<i>ccsA</i>	Inf	na	-Inf
<i>cemA</i>		1.59	-	1.00	na
<i>matK</i>		0.23	-	0.78	0.17
<i>ycf4</i>		Inf	na	-Inf	0.25

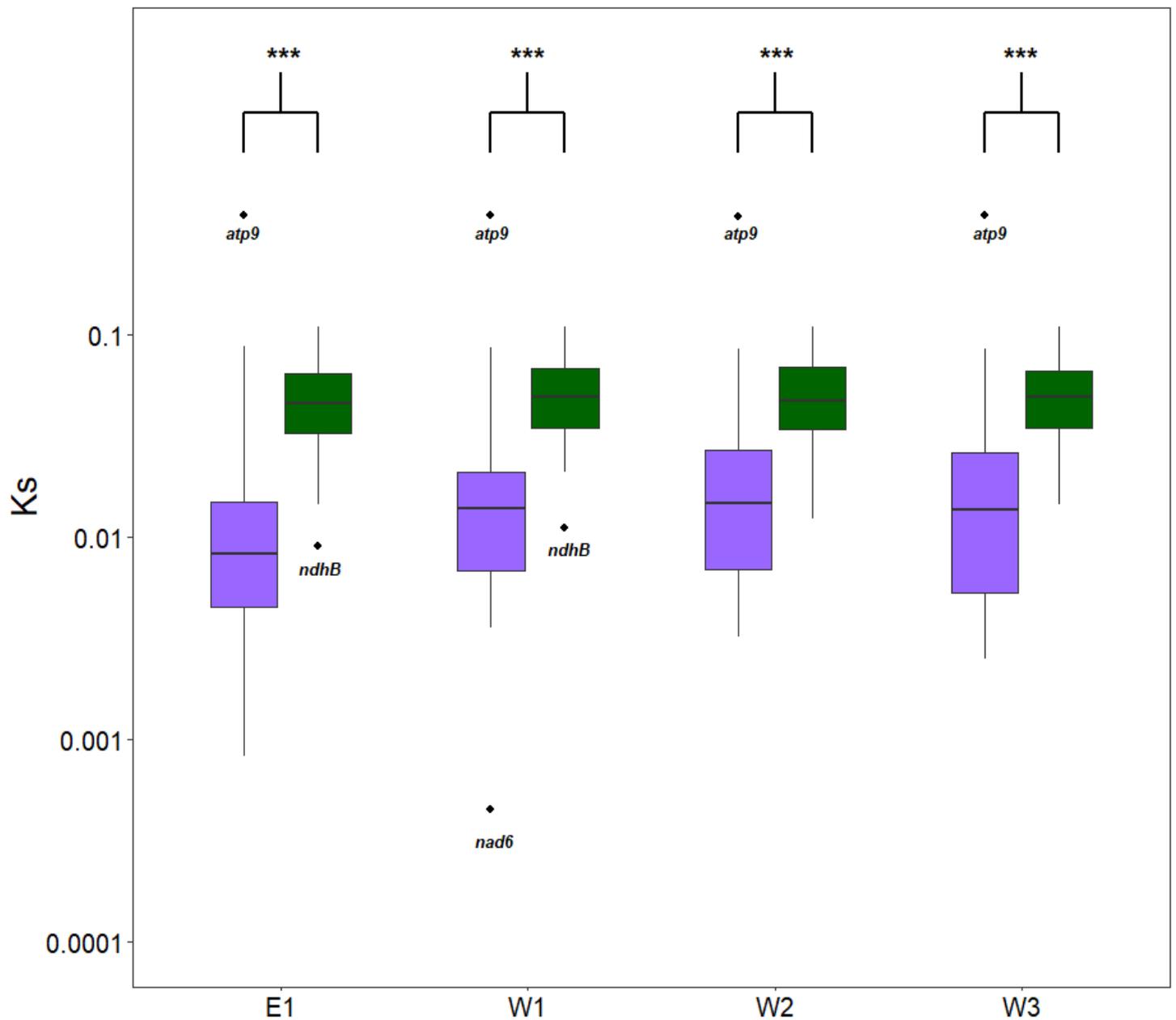
DoS = Direction of Selection ; NA = no result for the test ; - = results not significant

**Suppl. Table S4**

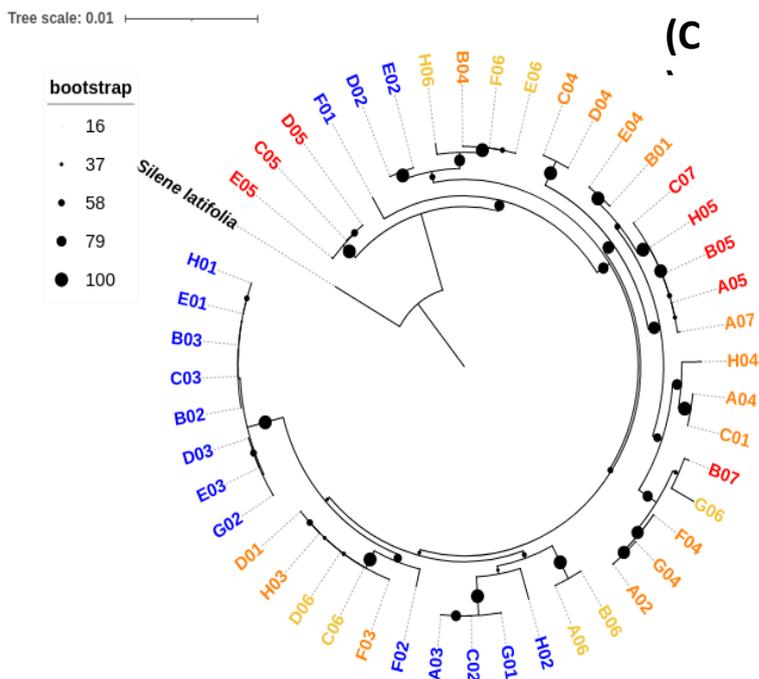
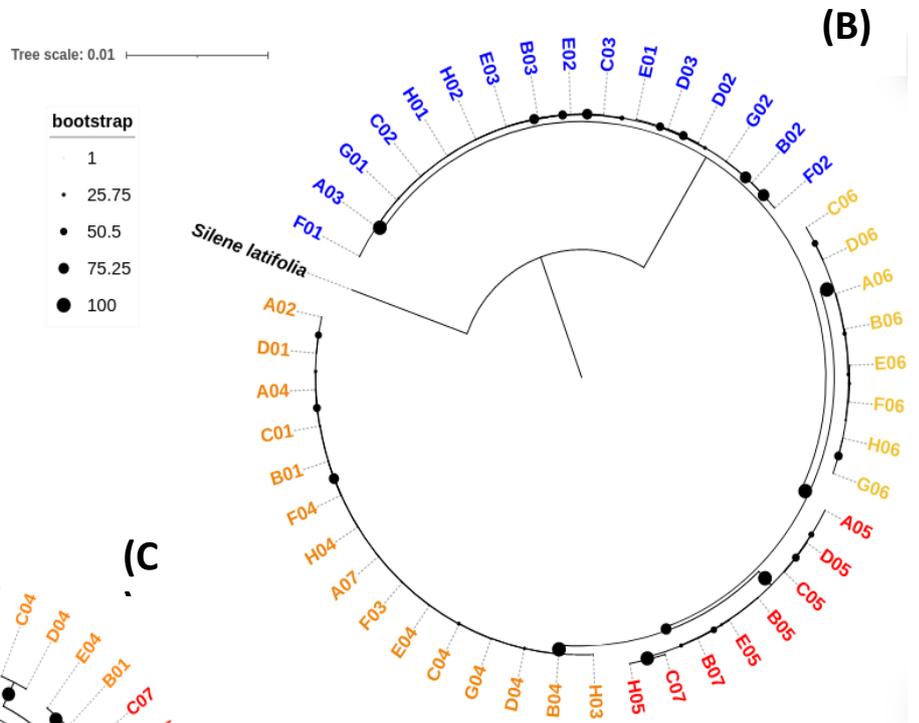
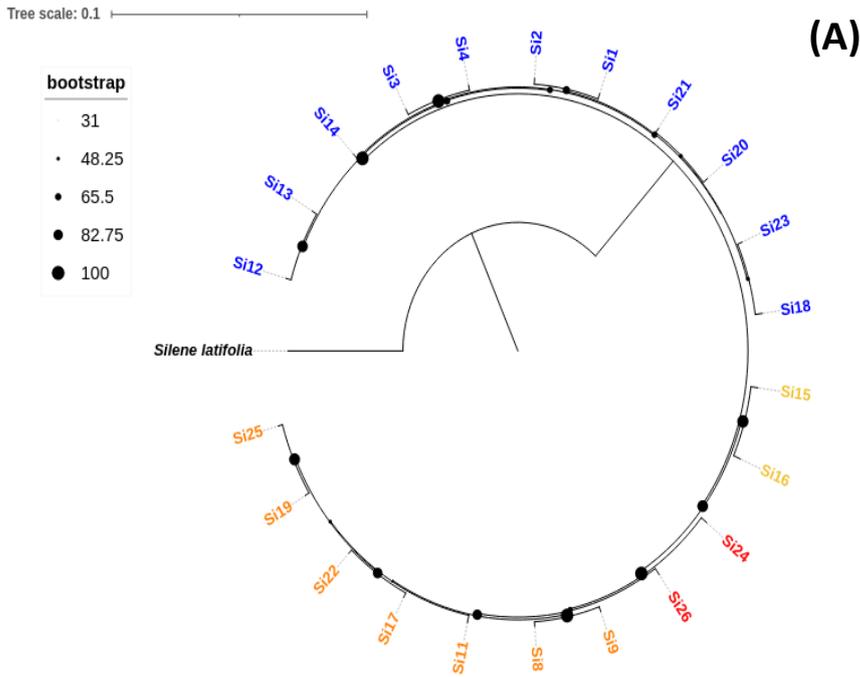
Results of the McDonald Kreitman test for mitochondrial and plastid gene concatenations per protein complex per lineage. Each lineage was compared to the outgroup *S. latifolia*. Red and blue text indicates significant ( $p < .05$ ) and marginally significant ( $p < 0.1$ ) results, respectively.

LINEAGES	GENOME	GENE	NEUTRALITY INDEX	ALPHA	FISHER PVALUE	DoS
<b>E1 vs <i>S. latifolia</i></b>	Mitochondrial	All mitochondrial genes	0.57	0.43	0.00	Positive
		OXPPOS complex I	1.29	-0.29	0.77	-
		OXPPOS complex III	0.22	0.78	0.18	-
		OXPPOS complex IV	0.21	0.79	0.00	Positive
		OXPPOS complex V	1.05	-0.05	1.00	-
		Cytochrome Biogenesis	0.51	0.49	0.23	-
		Membrane protein	0.00	1.00	0.08	Positive
		LSU	0.00	1.00	0.47	-
		SSU	Inf	-Inf	1.00	-
		Plastid	All plastid genes	2.20	1.20	0.30
	PSI		0.00	1.00	1.00	-
	PSII		na	na	1.00	-
	ATP synthase		na	na	1.00	-
	Cytochrome b6/f		na	na	1.00	-
	Rubisco		na	na	1.00	-
	NADH		5.16	-4.16	0.17	-
	RNA polymerase		0.90	0.10	1.00	-
	LSU		na	na	1.00	-
	SSU		na	na	1.00	-
	Other	na	na	1.00	-	
<b>W1 vs <i>S. latifolia</i></b>	Mitochondrial	All mitochondrial genes	0.61	0.39	0.01	Positive
		OXPPOS complex I	0.89	0.11	1.00	-
		OXPPOS complex III	0.17	0.83	0.18	-
		OXPPOS complex IV	0.34	0.66	0.07	Positive
		OXPPOS complex V	1.53	-0.53	0.22	-
		Cytochrome Biogenesis	0.36	0.64	0.06	Positive
		Membrane protein	0.00	1.00	0.19	-
		LSU	na	na	1.00	-
		SSU	Inf	-Inf	1.00	-
		Plastid	All plastid genes	6.62	-5.62	0.01
	PSI		na	na	1.00	-
	PSII		na	na	1.00	-
	ATP synthase		na	na	1.00	-
	Cytochrome b6/f		na	na	1.00	-
	Rubisco		na	na	1.00	-
	NADH		Inf	-Inf	0.08	Negative
	RNA polymerase		0.00	1.00	0.48	-
	LSU		2.64	-1.64	0.62	-
	SSU		Inf	-Inf	0.26	-
	Other	Inf	-Inf	1.00	-	
<b>W2 vs <i>S. latifolia</i></b>	Mitochondrial	All mitochondrial genes	0.53	0.47	0.00	Positive
		OXPPOS complex I	0.89	0.11	1.00	-
		OXPPOS complex III	1.33	-0.33	1.00	-
		OXPPOS complex IV	0.20	0.80	0.01	Positive

		OXPPOS complex V	0.92	0.08	0.86	-
		Cytochrome Biogenesis	0.41	0.59	0.13	-
		Membrane protein	0.00	1.00	0.06	Positive
		LSU	na	na	1.00	-
		SSU	Inf	-Inf	0.55	-
	Plastid	All plastid genes	1.91	-0.91	0.23	-
		PSI	na	na	1.00	-
		PSIILINEAGE	0.00	1.00	1.00	-
		ATP synthase	na	na	1.00	-
		Cytochrome b6/f	na	na	1.00	-
		Rubisco	na	na	1.00	-
		NADH	1.90	-0.90	0.66	-
		RNA polymerase	0.83	0.17	1.00	-
		LSU	Inf	-Inf	0.15	-
		SSU	na	na	1.00	-
		Other	na	na	1.00	-
<b>W3 vs <i>S. latifolia</i></b>	Mitochondrial	All mitochondrial genes	0.61	0.39	0.02	Positive
		OXPPOS complex I	0.92	0.08	1.00	-
		OXPPOS complex III	0.67	0.33	1.00	-
		OXPPOS complex IV	0.28	0.72	0.02	Positive
		OXPPOS complex V	1.14	-0.14	0.74	-
		Cytochrome Biogenesis	0.40	0.60	0.13	-
		Membrane protein	0.00	1.00	0.08	Positive
		LSU	na	na	1.00	-
		SSU	Inf	-Inf	1.00	-
	Plastid	All plastid genes	2.02	-1.02	0.26	-
		PSI	na	na	1.00	-
		PSII	na	na	1.00	-
		ATP synthase	na	na	1.00	-
		Cytochrome b6/f	na	na	1.00	-
		Rubisco	0.00	1.00	1.00	-
		NADH	2.62	-1.62	0.41	-
		RNA polymerase	0.00	1.00	0.22	-
		LSU	Inf	-Inf	0.25	-
		SSU	Inf	-Inf	1.00	-
		Other	na	na	1.00	-



Suppl. Fig. 3 -  $K_s$  calculated per lineage for mitochondrial (purple) and plastid (green) genes (n=23 and 68 genes respectively). Values of  $K_s$  with  $\log_{10}$  transformation for better visualisation.  $K_s$  was calculated using *S. nutans* individuals and *S. latifolia* reference genomes. Results of the Mann-Witney test are also shown : \*\*\* = pvalue < 0.001.



Suppl. Fig.4 Phylogenies constructed running RAxML – GTR Gamma model of nucleotide substitutions on the nuclear and mitochondrial genes concatenations. (A): nuclear phylogeny. (B): plastid phylogeny (C): mitochondrial phylogeny





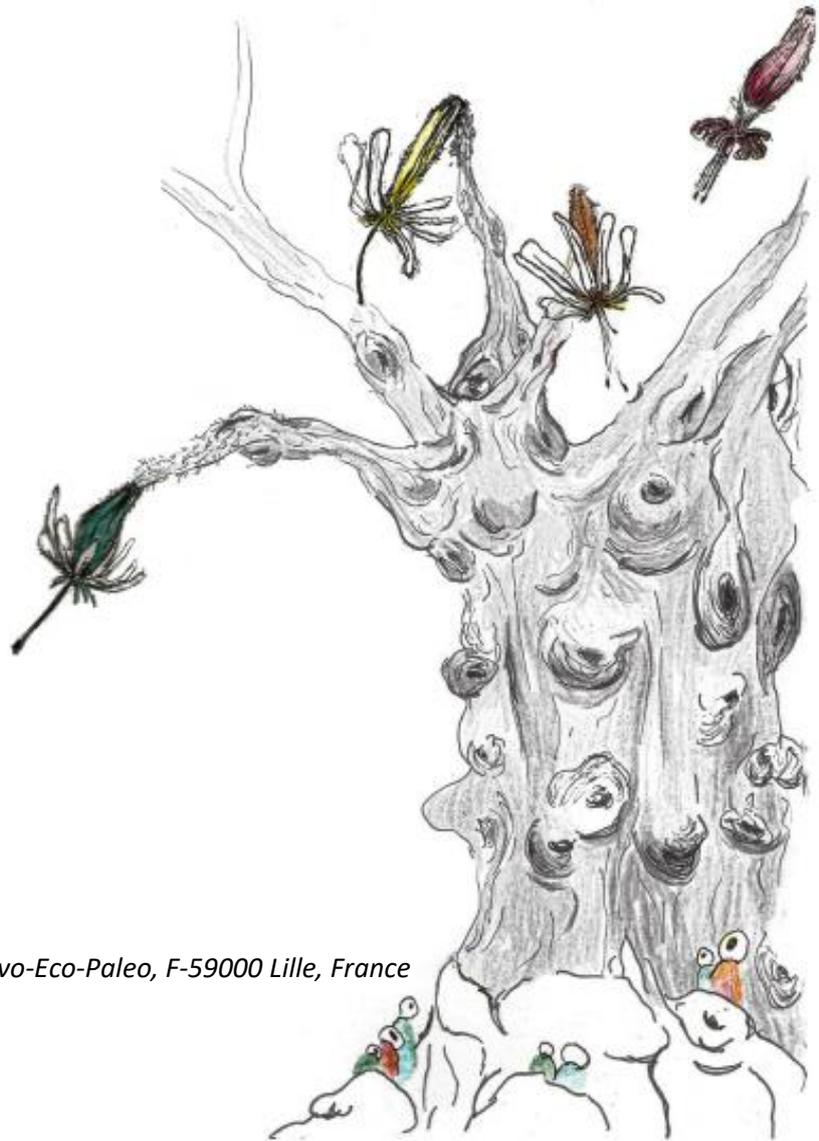
## CHAPTER 4

---

### Complete isolation without gene flow between four lineages of *Silene nutans*

Zoé Postel<sup>1</sup>, H  l  ne Martin<sup>1</sup>, Camille Roux<sup>1</sup>, C  cile God  <sup>1</sup>, Mathieu G  n  te<sup>1</sup>, Fran  ois Monnet<sup>1</sup>, Xavier Vekemans<sup>1</sup>, Pascal Touzet<sup>1</sup>

Manuscript in preparation.



<sup>1</sup> Univ. Lille, CNRS, UMR 8198 - Evo-Eco-Paleo, F-59000 Lille, France



In this chapter we tried to infer the evo-demographic scenario between the four lineages of *S. nutans*, using transcriptomic data for the lineages. Previous scenario of strict isolation was identified between lineages E1 and the western ones. Increasing the sampling effort for the western lineages compared to the previous study, we used DILS<sup>5</sup> to infer whether strict isolation was also identified between the four lineages of *S. nutans*. We also estimated the time of split between lineages and tested whether outliers of differentiation are the result of linked selection in low recombination genomic regions.

For this chapter, sampling strategy and choice of sampling populations were defined by myself and Pascal Touzet. Cécile Godé acquired the new transcriptomic data and Mathieu Génète did the quality check on these new data. With the help of François Monnet, I generated the input files and PCA prior running DILS. DILS with two populations was run online. Camille Roux developed the program to run DILS with four populations and ran it with the same input file. Result interpretation was done by myself with the help of Xavier Vekemans, Camille Roux, and Pascal Touzet. Figures were done by myself, with the help of Camille Roux for the one representing the grey zone of speciation. Writing of the chapter was done by myself except the material & methods for DILS with four populations which was written by Camille Roux. Editing of the manuscript was done by Pascal Touzet, Xavier Vekemans and Camille Roux.

This a first draft, it is not ready for publication yet. Additional analysis and editing of the manuscript still need to be done.

---

<sup>5</sup> Fraïsse C, Popovic I, Mazoyer C, Spataro B, Delmotte S, Romiguier J, Loire É, Simon A, Galtier N, Duret L, Bierne N, Vekemans X, Roux C. DILS: Demographic inferences with linked selection by using ABC. *Mol Ecol Resour.* 2021 Nov;21(8):2629-2644.

# **Complete isolation without gene flow between four lineages of *Silene nutans***

## **1. Introduction**

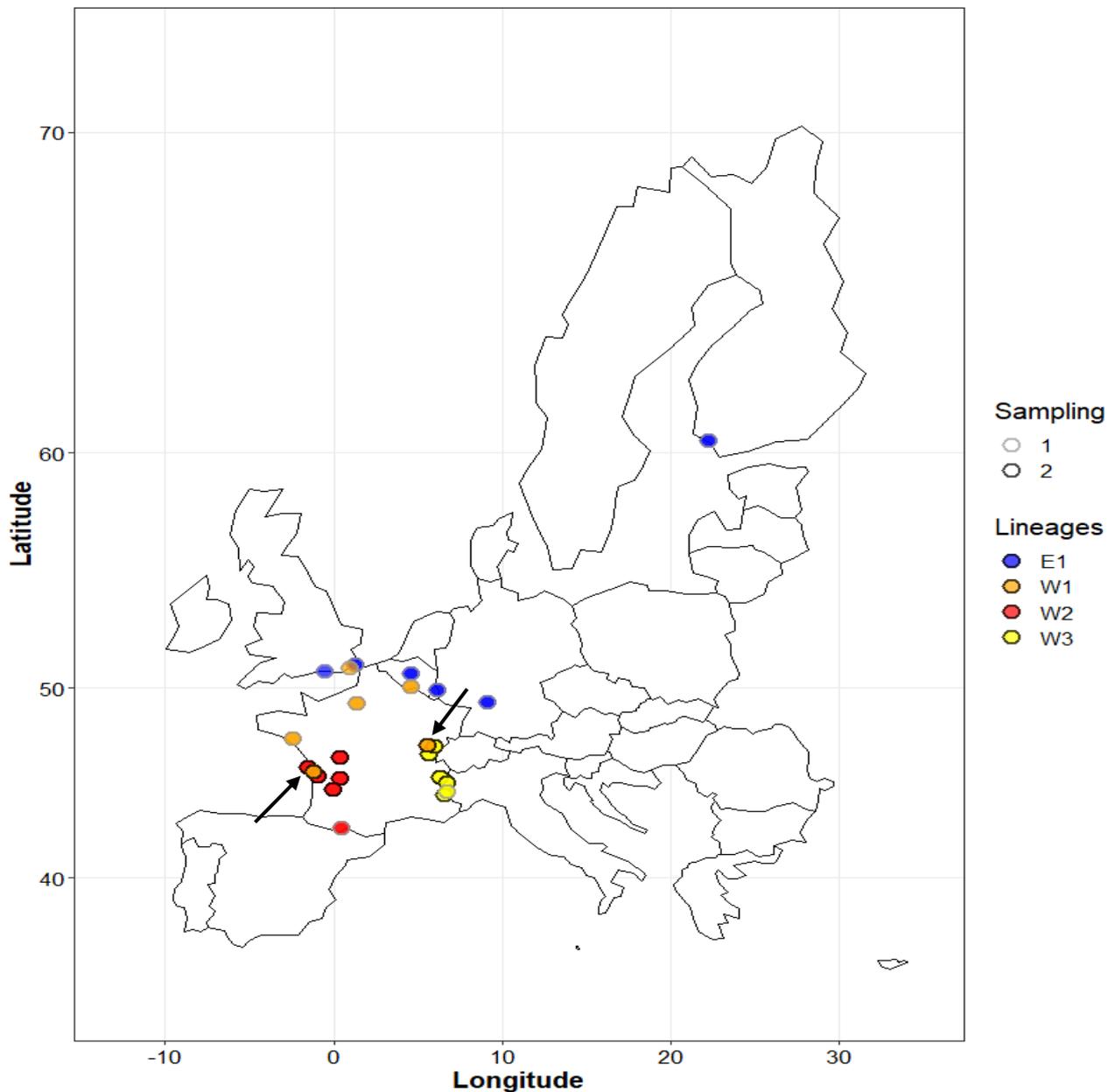
Speciation is the process by which one species composed of different populations splits into several new species (Coyne & Orr, 2004; Matute & Cooper, 2021). Species are defined as inter-reproductible units (Butlin & Stankowski, 2020). Speciation is supposed to be a gradual process (Seehausen *et al.*, 2014; Liu *et al.*, 2020) that involves many steps in its establishment (Feder *et al.*, 2012; Lowry, 2012; Kulmuni *et al.*, 2020). One of the first steps is the emergence of reproductive barriers that prevent gene flow between previously connected populations and may ultimately end up in reproductive isolation (RI) between these populations (Kulmuni *et al.*, 2020). These barriers can act before (pre-zygotic) or after (post-zygotic) reproduction between individuals of different populations occurs (Coughlan & Matute, 2020). Main modes of speciation have been identified and theorized throughout the years: (i) the allopatric speciation, where populations of one species split without exchanging gene flow (e.g. due to the emergence of a geographical obstacle to migration, enabling accumulation of divergence between the newly separated populations) ; (ii) sympatric speciation where populations diverged while continuously exchanging genes equally as there is no barrier to gene exchange ; (iii) parapatric speciation with populations having close geographical distributions and occasional gene flow (Coyne & Orr, 2004; Butlin *et al.*, 2008). These different modes of speciation result in distinct patterns of genetic diversity that enable us to distinguish among them using molecular data (Feder *et al.*, 2012; Seehausen *et al.*, 2014; Roux *et al.*, 2016b; Ravinet *et al.*, 2017). The reproductive barriers involved in the speciation process will also leave marks in the genomes. If speciation is primarily due to post-zygotic reproductive barrier, RI is thought to be the results of mismatch between independently evolved allelic combinations in isolated lineages, called Bateson-Dobzhansky Muller incompatibilities (BDMIs) (Presgraves, 2010; Fishman & Sweigart, 2018). Because of that, if at some point these lineages cross together, while the rest of the genome will introgress, these loci will not and will thus remain highly differentiated (Wu, 2001; Burri *et al.*, 2015). This would give rise to 'speciation islands' representing regions/loci in the genome where introgression does not occur and that can reach high levels of genetic differentiation (Turner *et al.*, 2005; Feder *et al.*, 2012; Seehausen *et al.*, 2014). But selection in low recombination zones induces high differentiation and thus can be mistaken as barriers to introgression (Cruickshank & Hahn, 2014; Shang *et al.*, 2020). Distinguishing between these two processes is thus a key point when studying speciation.

*Silene nutans* is a Caryophyllaceae species largely distributed in Europe. Previous studies identified two strongly differentiated evolutionary lineages within this species in relation to past climatic event and post-glacial recolonization: an eastern one (E1) widespread in the north of Europe (e.g England, western Europe, Great Britain) and a western one, composed of three sub-lineages: W1 distributed in England / France and Belgium, W2 restricted in Spain and south-western France and W3 in the Alp and Italy (Figure 1) (Martin *et al.*, 2016; Van Rossum *et al.*, 2018). These four lineages have specific plastid

haplotypes. At secondary contact zones between lineages E1 and W1, in France, southern Belgium and in south England, no hybridization events were detected suggesting the absence of gene flow between these lineages (Van Rossum *et al.*, 1997, 2018). These two lineages also show strong and complete post-zygotic reproductive isolation, expressed as high proportion of seedling chlorosis and reduced hybrid fitness. It could be the result of genetic incompatibilities accumulated in allopatry prior to the spread of lineages from their glacial refugia (Martin *et al.*, 2017). Result from diallelic crosses between the four lineages (i.e. E1/W1/W2/W3) also suggested the presence of a strong and asymmetric post-zygotic barriers not only between E1 and the Western sub-lineages but also between the western sub-lineages (Van Rossum *et al.* in prep). The result showed high and asymmetric proportion of hybrid mortality depending on which lineage was the cytoplasm donor and especially high when lineages E1 or W2 gave it (Van Rossum *et al.* in prep.). Plastid-nuclear incompatibilities (PNIs), a type of BDMIs, could be involved in the speciation process at stake between these four lineages (Postel *et al.*, 2022). *S. nutans* might then be composed of four potentially cryptic species. ABC approaches were already conducted on this system, but only using RNAseq data for lineages E1 and all of the western lineages (without distinction of the sub lineages). This study identified a scenario of allopatric speciation, without gene flow between these lineages since their split around 300 000 years ago (Martin, 2016). The aim of the current study was to increase the sampling effort for the western lineages in order to be able to infer the demographic history and patterns of differentiation among the four lineages of *S. nutans*. Specifically, we wanted to estimate the relative times of split of the lineages and test whether some lineages are currently connected by gene flow, despite the strong reproductive barriers identified previously among most lineages (Martin *et al.*, 2017, Van Rossum *et al.*, in prep). The aim was also to test whether loci exhibiting elevated differentiation between lineages might represent speciation islands (i.e. loci impermeable to gene flow) or genomic regions of low recombination experiencing linked selection (Cruickshank & Hahn, 2014; Seehausen *et al.*, 2014).

## **2. Material and methods**

### *2.1. Transcriptomic data*



**Fig.1 Geographic locations of all sampled and sequenced individuals for the four lineages.** Sampling time is given: (1)= Muyle et al. 2021 – (2)= 2022. The arcs represent populations in close geographical proximity for western lineages.

A transcriptome assembly was already available as well as transcriptomic data for 22 individuals from the four genetic lineages of *S. nutans* (Suppl. Table S1) ((Muyle *et al.*, 2021);PRJEB39526); but the number of individuals from each lineage was uneven: 11 individuals from E1, 7 from W1, 2 for W2 and 2 for W3. We increased the sampling effort to get a total of 11 individuals per lineage (i.e. +4 individuals for W1, 9 for W2 and W3). We followed the same sampling strategy (i.e. two individuals per population) (Suppl. Table S1, Figure 1). To test for potential gene flow within the western lineages, we sampled, when possible, individuals from population in close geographical proximity (Figure 1). RNAs were extracted from flower buds using NucleoSpin RNA plus kit from Macherey Nagel. Libraries were defined using NextFlex Rapid RNAseq kit. Sequencing for the 24 samples was done in paired-end

2x100bp. Libraries were sequenced on Illumina NextSeq500 HO in paired-end ( $2 \times 100$ ) at the LIGAN platform (UMR 8199 LIGAN-PM Genomics platform – Lille, France), resulting in a total of 44,436 GB and 21,654 GB after demultiplexing.

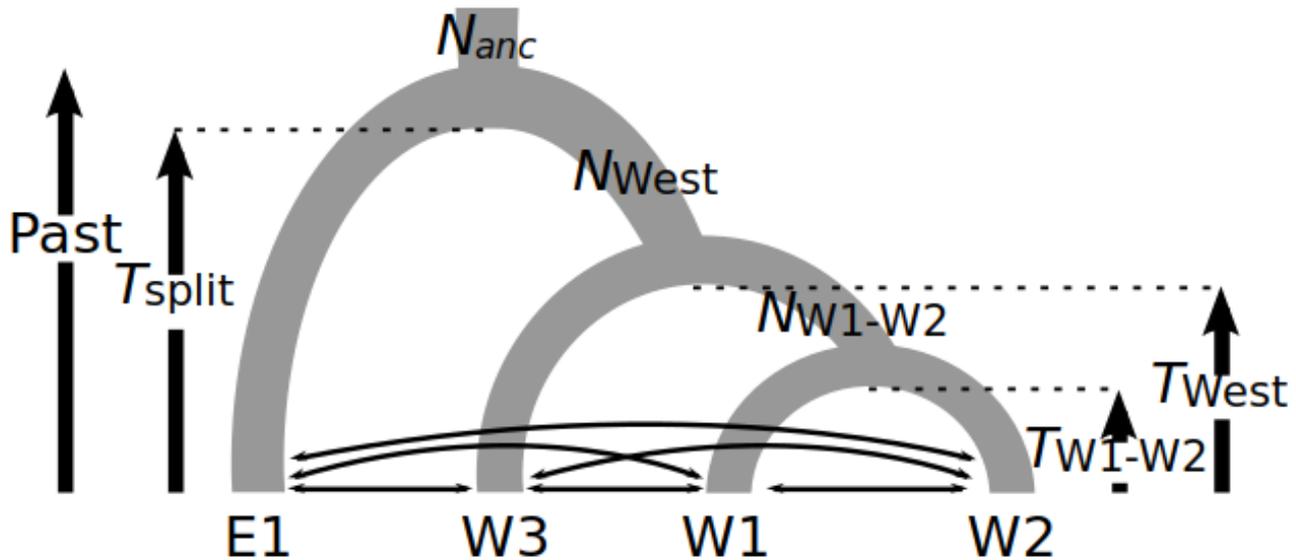
Newly acquired reads were aligned on the previously assembled transcriptome (Muyle *et al.*, 2021) using Bowtie v2.4.1 (Langmead & Salzberg, 2012) and duplicates eliminated using MarkDuplicate implemented in Picard v2.21.4 (“Picard Toolkit.” 2019. Broad Institute, GitHub Repository <https://broadinstitute.github.io/picard/>; Broad Institute). Reads were then cleaned, sorted and indexed using samtools v1.10 (Li *et al.*, 2009). To call the variants and produce the output format used for further analyses, we ran reads2snp v2.0 (Tsagkogeorga *et al.*, 2012) on these reads. The output multifasta format contains all biallelic nuclear loci for each individual.

## 2.2. Demographic inference using DILS ABC framework

### 2.2.1. Model comparison: 2 populations

Transcriptomic data can sometimes contain organellar gene sequences. The reference transcriptome was annotated using TAIR identifier, to easily find the organellar loci. Using seqkit (Shen *et al.*, 2016), we specifically removed those loci. Before running the ABC analysis using DILS (Fraïsse *et al.*, 2021), we checked whether individuals of the four lineages were indeed genetically differentiated. Using the fasta output of reads2snp we constructed a Principal Component Analysis (PCA) with popPhyl\_PCA ([https://github.com/popgenomics/popPhyl\\_PCA](https://github.com/popgenomics/popPhyl_PCA)) using all individuals from the four lineages.

DILS is a recently developed software that implements Approximate Bayesian Computational approaches to reconstruct evo-demographic scenarios using population genomic data (Fraïsse *et al.*, 2021). DILS analyzes a random sample of 1000 nuclear loci, from the overall dataset, to compute several summary statistics: a joint site frequency spectrum (jSFS), the number of private and shared polymorphic sites ( $S_X$  and  $S_S$  respectively), the number of fixed differences between pairs of lineages ( $S_F$ ), polymorphism statistics such as pairwise nucleotide diversity ( $\pi$ ), Watterson’s  $\theta$ , Tajima’s  $D$ , population divergence statistics  $D_{xy}$  (absolute divergence) and  $D_a$  (net divergence), and  $F_{ST}$  values of differentiation under different demographic scenarios. Additionally, DILS estimated the number of recombination events per locus running a four-gamete test. Giving a set of prior, simulated datasets are computed according to distinct alternative scenarios of speciation and divergence among the four lineages of *Silene nutans*, including scenarios with or without current gene flow. In the end, DILS compares values of the summary statistics from simulated and observed datasets to infer the most plausible scenario. It also estimates model parameters under the best scenario, such as the effective population size ( $N_e$ ) and the time of split ( $T_{split}$ ). We ran DILS for the 6 pairwise combinations of



**Fig.2 Four-population demographic model.** Four populations have been modelled here, each with independent current population sizes. The model describes three successive splitting events ( $T_{W1-W2}$ ,  $T_{West}$  and  $T_{split}$ ) down to the ancestral population. Each ancestral population is associated with an independent effective size of daughter populations. The six possible migration relationships are bidirectional and asymmetric secondary contacts.

lineages, using the same priors (Suppl. Table S2) for each combination and tested different demographic models of speciation: strict isolation (SI), ancient migration (AM), isolation with migration (IM) and secondary contact (SC). When searching for the best model, DILS also considered changes in population size along the genome (variation of the  $N_e$  among loci – genome homogeneous *versus* heterogeneous  $N_e$ ) and semipermeable barriers to gene flow (variation of the migration rate  $m_e$  among loci due to linkage to barriers). So, for each demographic scenario, fluctuating  $m_e$  and  $N_e$  were also testing through heterogeneity of these parameters along the genome. Because estimates of  $T_{split}$  can be tedious using only pairwise combinations, we also used a modified version of DILS performing model comparisons and parameters estimates on four-population models.

### 2.2.2. Model comparison: 4 populations

The four-population model explored assumes a topology separating the eastern lineage E1 from the western lineages (W3, (W1, W2)) (Figure 2). This topology was simulated using MSMS as below, where tbs means a value to be specified:

```
msms tbs 10000 -s tbs -r tbs tbs -l 4 tbs tbs tbs tbs 0 -n 1 tbs -n 2 tbs -n 3 tbs -n 4 tbs -m 1 3 tbs -m 3 1
tbs -m 2 4 tbs -m 4 2 tbs -m 1 2 tbs -m 2 1 tbs -m 2 3 tbs -m 3 2 tbs -m 1 4 tbs -m 4 1 tbs -m 3 4 tbs -m
4 3 tbs -em tbs 3 4 0 -em tbs 4 3 0 -em tbs 3 2 0 -em tbs 2 3 0 -em tbs 3 1 0 -em tbs 1 3 0 -em tbs 2 4 0
-em tbs 4 2 0 -em tbs 2 1 0 -em tbs 1 2 0 -em tbs 4 1 0 -em tbs 1 4 0 -ej tbs 3 4 -en tbs 4 tbs -ej tbs 4 2
-en tbs 2 tbs -ej tbs 2 1 -eN tbs tbs
```

In this model, the four current lineages E1, W1, W2 and W3 have the possibility to exchange alleles during six independent secondary contacts. These secondary contacts involve six possible lineage pairs: E1-W3 (at time  $TSC_{E1-W3} < T_{west}$ ), E1-W1 (at time  $TSC_{E1-W1} < T_{W1-W2}$ ), E1-W2 (at time  $TSC_{E1-W2} < T_{W1-W2}$ ), W3-W1 (at time  $TSC_{W3-W1} < T_{W1-W2}$ ), W3-W2 (at time  $TSC_{W3-W2} < T_{W1-W2}$ ) and W1-W2 (at time  $TSC_{W1-W2} < T_{W1-W2}$ ). These secondary contact constraints on split times are intended to test only recent gene flow involving only pairs in independent combinations. At each population split, the new ancestral population (backward in times) has an independent size from the daughter populations ( $N_{anc}$ ,  $N_{west}$  and  $N_{W1-W2}$ ). Under this model, we randomly simulated multi-locus datasets with properties corresponding to the observed sampling (number of loci, locus length, number of gametes per locus and per sequenced population). For each simulation, coalescent trees of loci were obtained for parameter values randomly drawn from prior distributions. The exact number of mutations corresponding to that observed for each locus was then randomly placed in the simulated tree, thus not following a molecular clock according to an assumed mutation rate. With this mutation model, the parameters are expressed in coalescent units and not demographic units, allowing only relative parameter estimates to be obtained.

Current and ancestral population sizes are randomly drawn in a uniform distribution between 0 and  $10.N_e$ ;  $T_{split}$  is drawn in a uniform distribution between 0 and  $20.N_e$ ;  $T_{west}$  is drawn between 0 and  $T_{split}$ ;  $T_{W1-W2}$  is drawn between 0 and  $T_{west}$ ; effective migration  $N_e.m$  is drawn in a uniform distribution between 0 and 10; migration is assumed to be heterogeneously distributed in the genome according to a Beta distribution of alpha and beta shape parameters drawn in a uniform distribution between 0 and 20.

Introgression between lineages was tested independently for each of the six possible pairs. For this, 64 alternative models were simulated, depending on whether the migration for a given pair is null ( $N_e.m = 0$  for all loci) or non-null ( $N_e.m$  drawn between 0 and 10). Each of the 64 alternative models was simulated 10,000 times. Model comparisons were then performed individually for each pairwise relationship. To test whether recent gene flow has shaped the genetic patterns for a pair involving E1 and W3, we label as "isolation" the 32 submodels for which there is no migration between E1 and W3 (but with or without migration between the other population combinations) and "migration" the 32 submodels for which there is migration. These six model comparisons were performed with the R package abcrf (Pudlo *et al.*, 2016) and using a forest of 1,000 trained trees. The trainings were conducted using the summary statistics described for the two-population models.

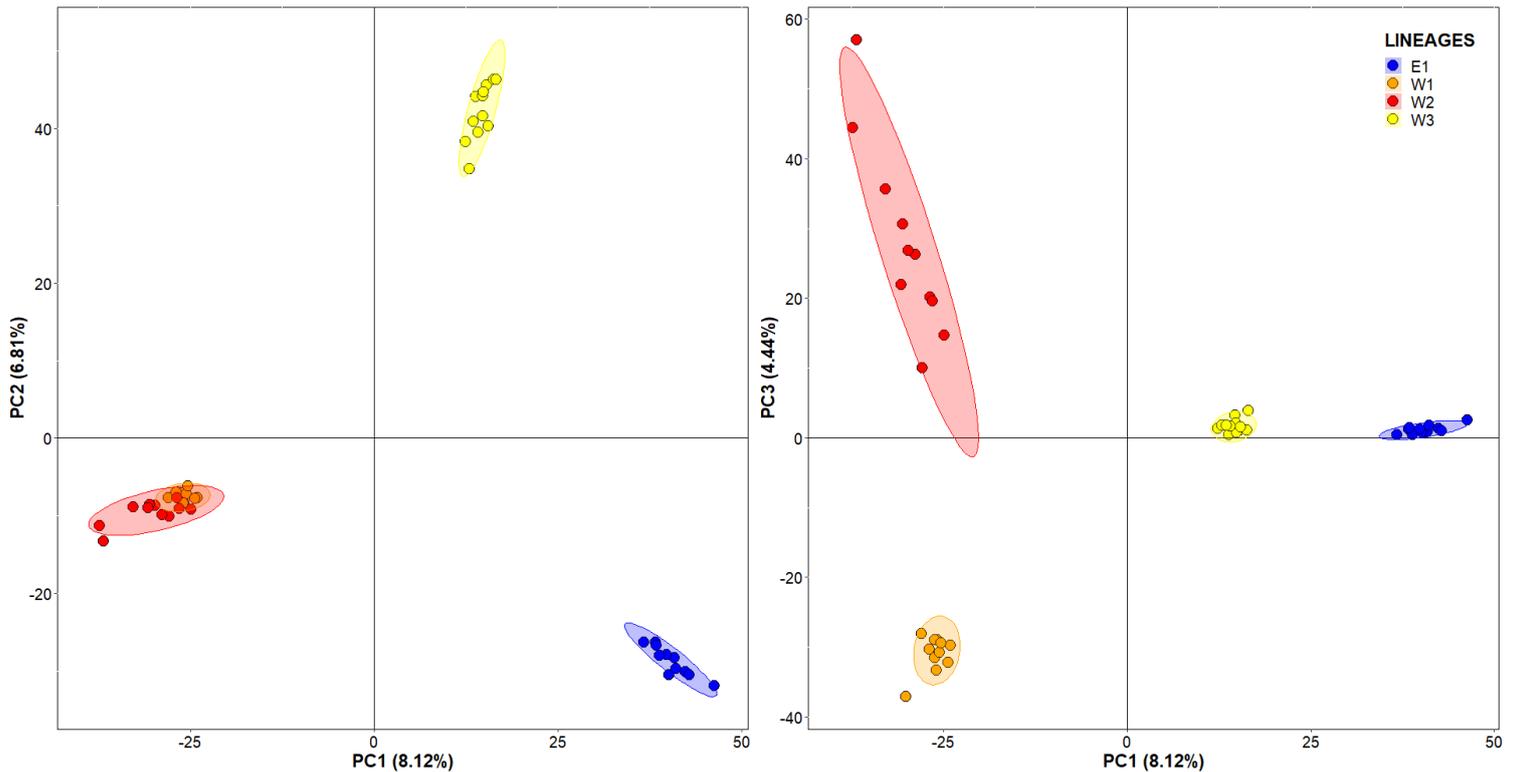
The parameters of the best-supported model among the 64 tested were also inferred by an ABC approach. Simulations similar to the previous step were performed, but using a molecular clock assuming  $\mu=7.31 \times 10^{-9}$  (Krasovec *et al.*, 2018). We calculated the split time in year considering that one

generation is equivalent to three years. Summary of parameters and prior use for this analysis can be found in suppl. table S2.

### 2.3. Statistical analysis

We compared values of some summary statistics among lineages or pairs of lineages using ? statistical tests. As DILS randomly samples 1,000 loci in the whole nuclear datasets before the data filtering procedures, we did not get the same number of loci for each pairwise combination nor the exact same loci. So, we first filtered the common loci to all pairwise combinations (n=3460). Then, on R v. 1.4.1717 we assessed whether levels of genetic differentiation ( $F_{ST}$ ), net and absolute divergence between two lineages ( $D_a$  and  $D_{xy}$  respectively), number of shared polymorphisms ( $S_S$ ) and fixed differences ( $S_F$ ) were significantly different between each pairwise combinations (e.g. does lineages E1 and W1 have the same level of divergence than lineages E1 and W2). As data were not normally distributed (Schapiro test p-value < 2.2e-16), we ran a Wilcoxon Mann-Whitney test for each comparison. We also assessed whether the number of private polymorphic sites ( $S_X$ ), the pairwise nucleotide diversity ( $\pi$ ), Watterson's  $\theta$  and Tajima's D were significantly different between lineages for each pairwise comparison. Data were not normally distributed either (Schapiro test p-value < 2.2e-16) except for Tajima's D. Statistical significance was then assessed using Student T-test for the normally distributed data and Wilcoxon-Mann Whitney for the non-normally distributed ones.

In regions of low recombination, linked selection can act and drive high values of  $F_{ST}$  (Cruickshank & Hahn, 2014). Consistently, high values of  $D_a$  should be identified in these regions and no differences in  $D_{xy}$  between these regions and the rest of the genome. We distinguished between the outliers of  $F_{ST}$  (FO), defined as the 5% highest values of  $F_{ST}$  for a pairwise combination and the non-outliers of  $F_{ST}$  (FNO) (i.e. the rest of the  $F_{ST}$  distribution) for the recombination rate, the nucleotide diversity within lineage  $\pi$  and  $D_{xy}$ . To assess whether significantly different values were observed between FO and FNO, for each lineage combination, we ran a Wilcoxon-Mann Whitney test. If FO had significantly lower recombination rates, lower within lineage diversity and lower genetic divergence between lineages, then it would suggest that these loci with high  $F_{ST}$  values are in fact in low recombination regions and subject to linked selection, driving high values of genetic differentiation between lineages. We also wanted to assess whether the FO concerned significantly more nuclear loci whose gene products are targeted to the plastid. Ninety-seven nuclear loci among the 3460 analyzed for all pairs of lineages were targeted to the plastid. We then calculated for each pair of lineages the number of loci targeted to the plastid among the FO and conducted a Fisher exact test to assess statistical significance.



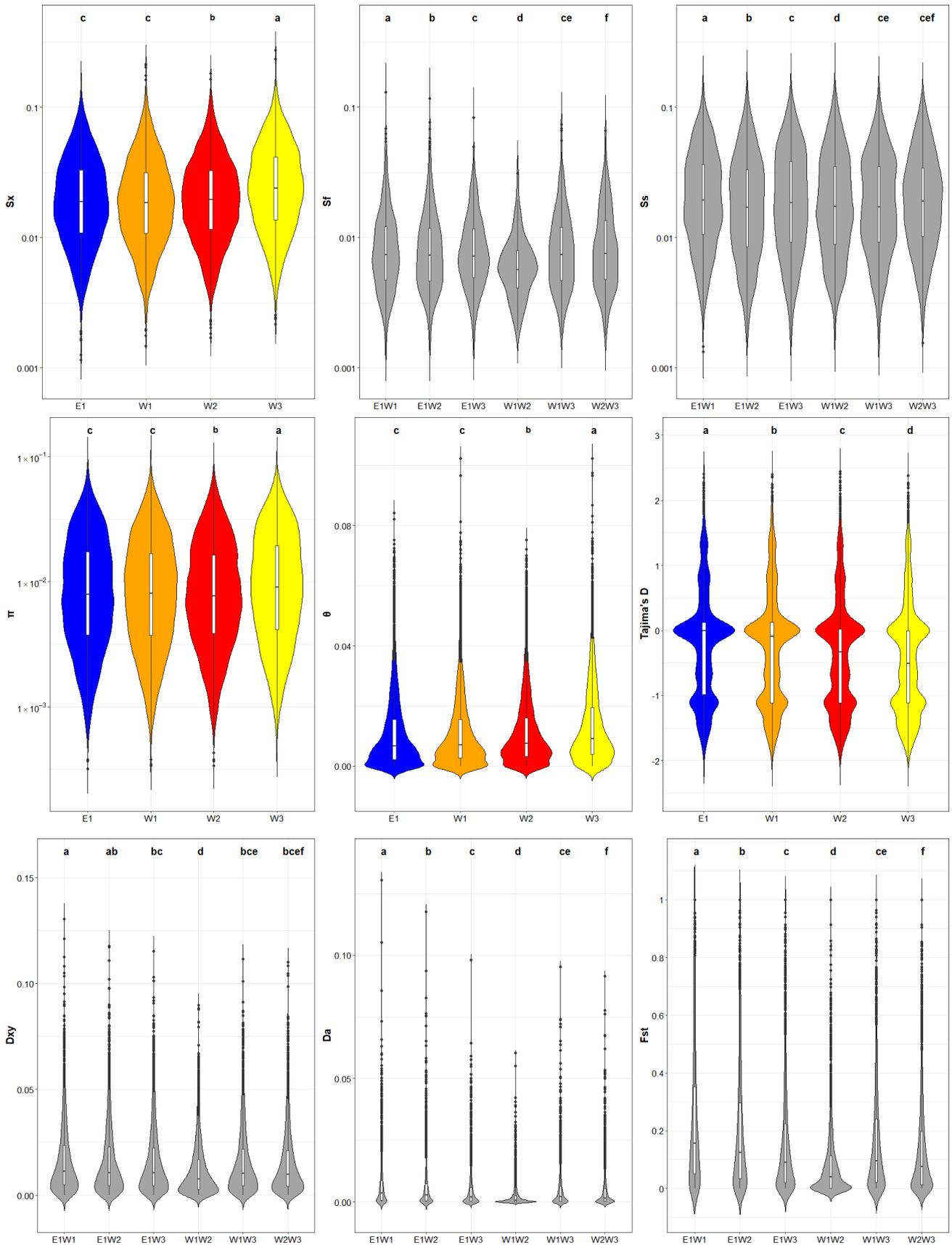
**Fig.3** Individuals coordinates on the PC1, PC2 and PC3. Eigen values for each PC axes are given.

### 3. Results

Prior to run ABC and subsequent analyses, we verified that the four lineages clustered separately in a PCA analysis. Overall, lineages clustered separately along PC1, PC2 and PC3 (Figure 3). Lineages W1 and W2 were very close to one another in a PC1-PC2 plot, with W1 individuals clustering within W2, but they appeared differentiated according to PC3 (Figure 3).

#### 3.1. Population genetic statistics

Using 5 individuals per lineage and a random sample of 1000 nuclear loci, DILS computed summary statistics. Statistical tests were run on the subset of commonly analyzed loci for each pairwise comparison. Pairwise nucleotide diversity ( $\pi$ ), Watterson's  $\theta$  measure of genetic diversity and the number of polymorphic sites specific to each lineage ( $S_X$ ) were significantly greater for lineage W3 than any other lineage (mean=0.0121 / 0.0133 / 0.027 for  $\pi$ ,  $\theta$  and  $S_X$  respectively, Figure 4, Suppl. Table S3). Lineage W2  $\pi$  value (mean = 0.0099) was significantly higher than that of W1 (mean = 0.0095) but significantly lower compared to E1 (mean = 0.0102) (Figure 4, Suppl. Table S3). Regarding the other measures ( $\theta$  and  $S_X$ ), values of lineage W2 (mean = 0.0109 / 0.0204 for  $\theta$  and  $S_X$  respectively) were significantly higher compared to both E1 and W1, for which differences were not significant (means = 0.0105 / 0.0197 for E1 and 0.0098 / 0.0167 for W1, for  $\theta$  and  $S_X$  respectively) (Figure 4, Suppl.



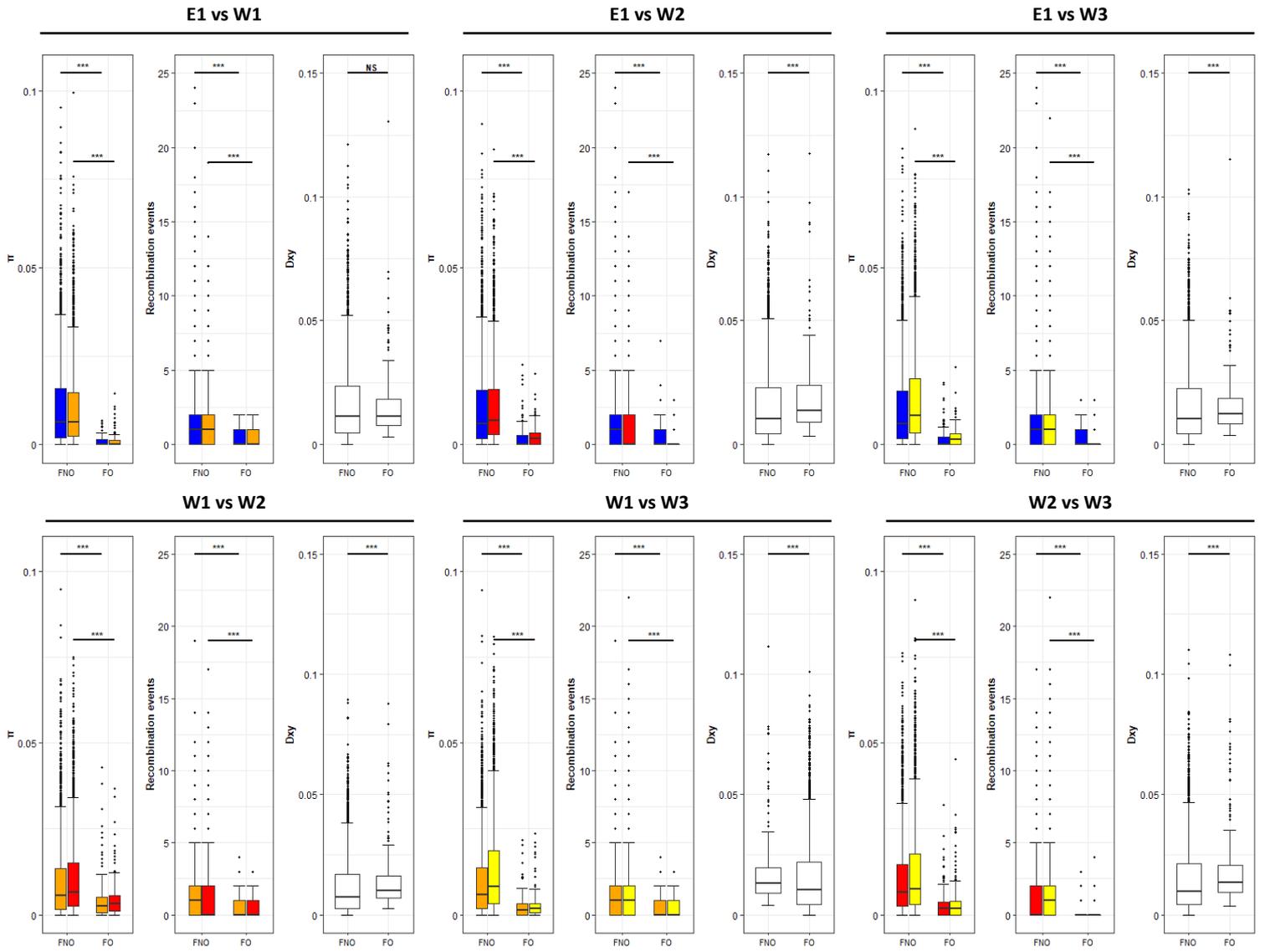
**Fig.4 Summary statistics values for each pairwise combinations and lineage.** Blue: lineage E1 ; orange = lineage W1 ; red = lineage W2 ; yellow = lineage W3.  $S_x$  = number of sites with a polymorphism specific to a lineage ;  $S_f$  = number of sites with fixed differences between the two lineages ;  $S_s$  = number of sites with a polymorphism shared between the two lineages. Results of the Wilcoxon Mann-Whitney and Student t-test are reported at the top of each violin plot. When letters are different, differences are significant with a pvalue < 0.05. For detailed about these results, see supplementary tables.

Table S3). Results of the Tajima's D values showed opposite patterns, i.e. each difference between lineages being statistically significant and with the following order : E1 (mean = -0.17) > W1 (mean = -0.20) > W2 (mean = -0.38) > W3 (mean = -0.47) (Figure 4, suppl. table S3). The number of fixed differences in each lineage compared to the others ( $S_F$ ) was the highest between lineages E1 and W1 and the lowest between lineages W1 and W2 (mean = 0.0018 and 0.0001 respectively) (Figure 4, Suppl. Table S4). Mean of  $S_F$  was not significantly different between lineage E1/W3 and W1/W3 (mean = 0.001 for both) which exhibit intermediate levels of fixed differences between all lineages. Consistently, level of net genetic divergence ( $D_a$ ) followed the same pattern, with mean  $D_a$  being the highest for E1/W1 > E1/W2 > E1/W3 > W1/W3 > W2/W3 > W1/W2 (Figure 4, Suppl. Table S4).  $D_a$  were significantly different for all combination's comparisons except between W1/W3 and E1/W3. The same pattern was also observed for the mean absolute divergence ( $D_{xy}$ ) (E1/W1 > E1/W2 > E1/W3 > W1/W3 > W2/W3 > W1/W2) (Figure 4). Significant different values were observed between combination E1/W2 and the rest of the combinations and also between combination E1/W1 and the rest of the combinations except E1/W2 for which  $D_{xy}$  values were not significantly different (Figure 4, Suppl. Table S4). Genetic differentiation between lineages also followed the exact same trend, with all differences between pairs of lineages being significant except again between W1/W3 and E1/W3 (Figure 4, Suppl. Table S4). Surprisingly, regarding the amount of shared polymorphism between lineages, lineages E1 and W1 seems to share more polymorphic sites than the others, and especially W1 and W2 which are the less genetically divergent pair of lineages (mean = 0.021 vs 0.013) (Figure 4, Suppl. Table S4).

Loci having the highest levels of FST (outliers, FO) exhibited significantly lower values of recombination events and  $\pi$  compared to non-outlier loci (FNO) for all pairwise combination (Figure 5, Suppl. Table S5). Regarding  $D_{xy}$  between lineages, FO exhibited significantly higher values compared to FNO for all pairs except E1/W1 where differences were not significant (Figure 5, Suppl. Table S5). FO were not significantly enriched in loci encoding nuclear genes targeted to the plastid, except for lineages W1 and W2 (Table 1).

### 3.2. *Evo-demographic Model selection*

We ran ABC analyses for each pair of lineages (n=6). For each, we selected the best evo-demographic model explaining the observed dataset looking at the posterior probabilities for each model and the goodness of fit produced as an output of DILS. For each of the 6 pairs, the best model was found to fit correctly to the data: on the PCA representing the goodness-of-fit for the best model, the observed



**Fig.5** Level of within lineage diversity  $\pi$ , number of recombination events and between lineages divergence ( $D_{XY}$ ) for each pair of lineages. For each measure, we distinguished between outliers of  $F_{ST}$  (FO) (i.e. the last 5% of the  $F_{ST}$  distribution for each pair of lineages) and non-outliers of  $F_{ST}$  (FNO) (i.e. the 95% other  $F_{ST}$  values). Results of the Wilcoxon-Mann Whitney test conducted for each lineage and pair of lineages between outliers and non-outliers are displayed: NS= non-significant ; \*\*\* = pvalue < 0.001

**Table 1**

Result of the Fisher test to assess whether  $F_{ST}$  outliers are enriched in loci targeted to the plastid. We only worked with the commonly analyzed loci for all pairs of lineages (n= 3460)

Pair of lineages	Number of $F_{ST}$ outliers targeted to the plastid	Number of $F_{ST}$ outliers not targeted to the plastid	Odds ratio	pvalue	IC 95
E1 vs W1	7	166	1.4618	0.344	0.563608 - 3.197023
E1 vs W2	9	164	1.9022	0.097	0.829497 - 3.852991
E1 vs W3	7	166	1.4618	0.344	0.563608 - 3.197023
W1 vs W2	10	163	2.1265	0.035	0.969985 - 4.182624
W1 vs W3	9	164	1.9022	0.097	0.829497 - 3.852991
W2 vs W3	8	165	1.6807	0.161	0.693831 - 3.524508

Total number of  $F_{ST}$  outliers targeted to the plastid = 97 ; Total number of  $F_{ST}$  outliers not targeted to the plastid = 3363

**Table 2**

Best demographic model assessed with DILS. Number under the name of the demographic model represent the posterior probabilities, from 0 to 1. Ancient Migration vs Strict Isolation was not tested with the DILS using four populations.

Pairs of lineages	ONGOING MIGRATION versus ISOLATION		ANCIENT MIGRATION versus STRICT ISOLATION	N-HOMO versus N-HETERO
	DILS 2 pops	DILS 4 pops	DILS 2 pops	DILS 2 pops
<b>E1 vs W1</b>	Isolation 0.885	Isolation 0.898	Ancient migration 0.711	N-hetero 0.997
<b>E1 vs W2</b>	Isolation 0.912	Isolation 0.929	Ancient migration 0.764	N-hetero 0.997
<b>E1 vs W3</b>	Isolation 0.915	Isolation 0.940	Ancient migration 0.859	N-hetero 0.992
<b>W1 vs W2</b>	Isolation 0.832	Isolation 0.857	Ancient migration 0.784	N-hetero 0.989
<b>W1 vs W3</b>	Isolation 0.892	Isolation 0.894	Ancient migration 0.750	N-hetero 0.979
<b>W2 vs W3</b>	Isolation 0.876	Isolation 0.890	Ancient migration 0.800	N-hetero 0.979

N-homo / N-hetero : heterogeneous or homogenous population effective size.

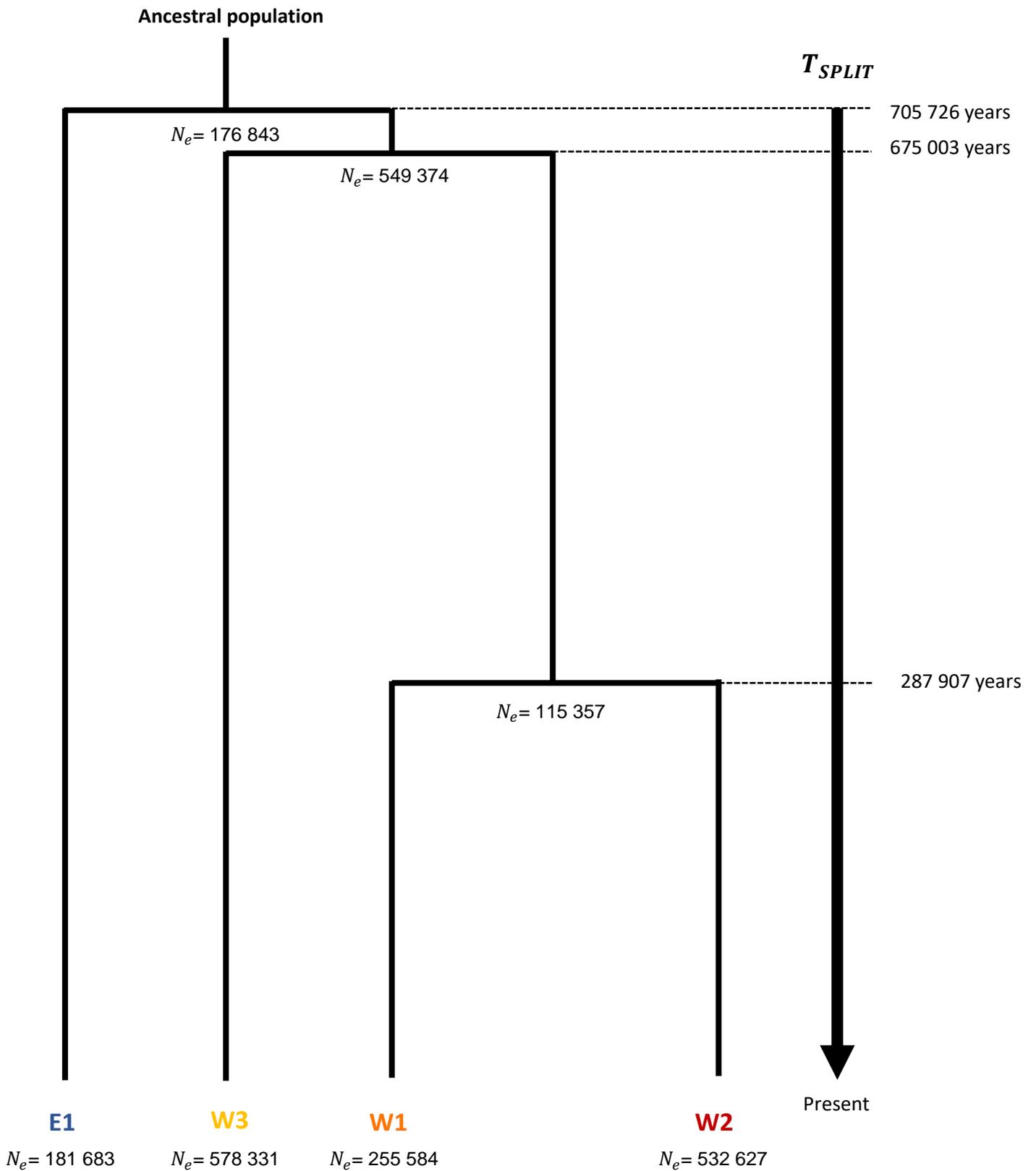
dataset was found roughly in the middle of the optimized posterior distribution (Figure S1) when looking at PC1, 2 and 3, meaning that values of the summary statistics obtained in simulations of the best model were similar to those of the observed dataset. Equivalent results were obtained for the analysis using DILS with a model of four population (Figure S2).

Posterior probabilities for all models and all pairs of lineages ranged from 0.711 to 0.997, indicating unambiguous support for the best model (Table 2). Overall, for all pairs of lineages, between strict isolation (SI) and isolation with migration (IM), the best model was SI. Between models of SI + ancient migration (AM) and SI + secondary contact (SC), SI+AM seemed to better explain the observed dataset. Finally, again for all pairs of lineages, demographic models of SI+AM with heterogeneity of the  $N_e$  best fitted the data. Results from DILS run with four populations were the same: scenario of strict isolation, with posterior probabilities ranging from 0.86 to 0.94 (Table 2)<sup>6</sup>.

### 3.3. Parameter estimation

Parameters were estimated using DILS directly using the four lineages model. Lineage W3 seems to exhibit the highest  $N_e$  followed by lineages W2 > W1 > E1 (Figure 6, Suppl. Table S6). The ancestral population of all lineages displays relatively low  $N_e$  compared to current lineages  $N_e$ , similarly to the  $N_e$  of the ancestral population of lineages W1 and W2 (Figure 6, Suppl. Table S6). The ancestral

<sup>6</sup> Ne-hetero vs Ne-homo was not tested yet but is planned.



**Fig.6 Phylogenetic relationship between lineages and evolutionary histories, with times of split.** Branch length are proportional to the time of separation between lineages (e.g. 705 726 = 15 cm). Effective sizes ( $N_e$ ) are also reported.

population of the western lineages exhibits one of the highest  $N_e$  (i.e. around 550 000 individuals). Regarding time of split ( $T_{split}$ ), eastern and western lineages split around 700 000 years ago (IC95= 283 695 – 1 363 743)

(Figure 6, Suppl. Table S6). Lineage W3 separated from lineages W1 and W2 soon after (around 680 000 years ago (IC95= 202 092 – 1 204 050)) (Figure 6, Suppl. Table S6). Finally, the split between lineages W1 and W2 occurred the most recently, around 300 000 years ago (IC95=82 248 – 782 862) (Figure 6, Suppl. Table S6).

## 4. Discussion

### 4.1. Allopatric speciation for all lineages of *S. nutans*

Regarding DILS results, either with two or four lineages, highest posterior probabilities were observed for model of divergence without gene flow for all four lineages. Lineage E1 separated first, followed shortly after by W3 and then, after a longer period of time, by W1 and W2. The estimated small split time between E1 and W3 might suggest either incomplete lineage sorting or use of an incorrect topology between the four lineages. Yet, when randomly sampling nuclear loci in chapter 3, the associated topology was correctly resolved: (E1,(W3,(W1,W2))). This still needs to be investigated. Compared to what was previously estimated in (Martin, 2016 - chapter 3), the split times are different, likely because we did not use the same mutation rate (i.e.  $7.31 \times 10^{-9}$  bp/generation here vs  $2.76 \times 10^{-8}$  bp/generation in Martin, 2016 – chapter 3). Consistently with what was found previously, these split times could match with glacial maxima during Quaternary glaciation cycles, which might have induced independent migration of each lineage to a glacial refugium of the four lineages as suggested in (Martin *et al.*, 2016; Van Rossum *et al.*, 2018). The first glacial events that could have led to a split between lineages E1 / W3 / W1-W2 is the Günz glaciation, that started around 760 000 years ago and ended around 530 000 years ago. Around 400 000 years after the first *S. nutans* lineages split, subsequent separation occurred between lineages W1 and W2 (i.e. around 300 000 years ago) which corresponds to the early stage of the Riss glacial period. This time is consistent with the geographical distribution of the lineages that reflects post-glacial recolonization pathways in Europe (Martin *et al.*, 2016). Refugia for lineages W2 and W3 seemed to be located in the Iberian Peninsula for the former and the Italian one for the latter, while lineage W1's refugium could be located in south-central France (the French Massif Central area) and E1's refugia in eastern Europe (Van Rossum *et al.*, 2018). Lineages W2 and W3 recolonization seemed to have been restricted to south-western France for W2 and the Alps and south-eastern Europe for W3 (Van Rossum *et al.*, 2018). W1 could have recolonized from the Massif Central area towards northern Europe, reaching its northern margins in southern England and

Belgium (Van Rossum *et al.*, 2018). Lastly, E1 followed a north-west expansion in Europe, having apparently the widest European geographical distribution as compared to the western sub-lineages (Van Rossum *et al.*, 2018).

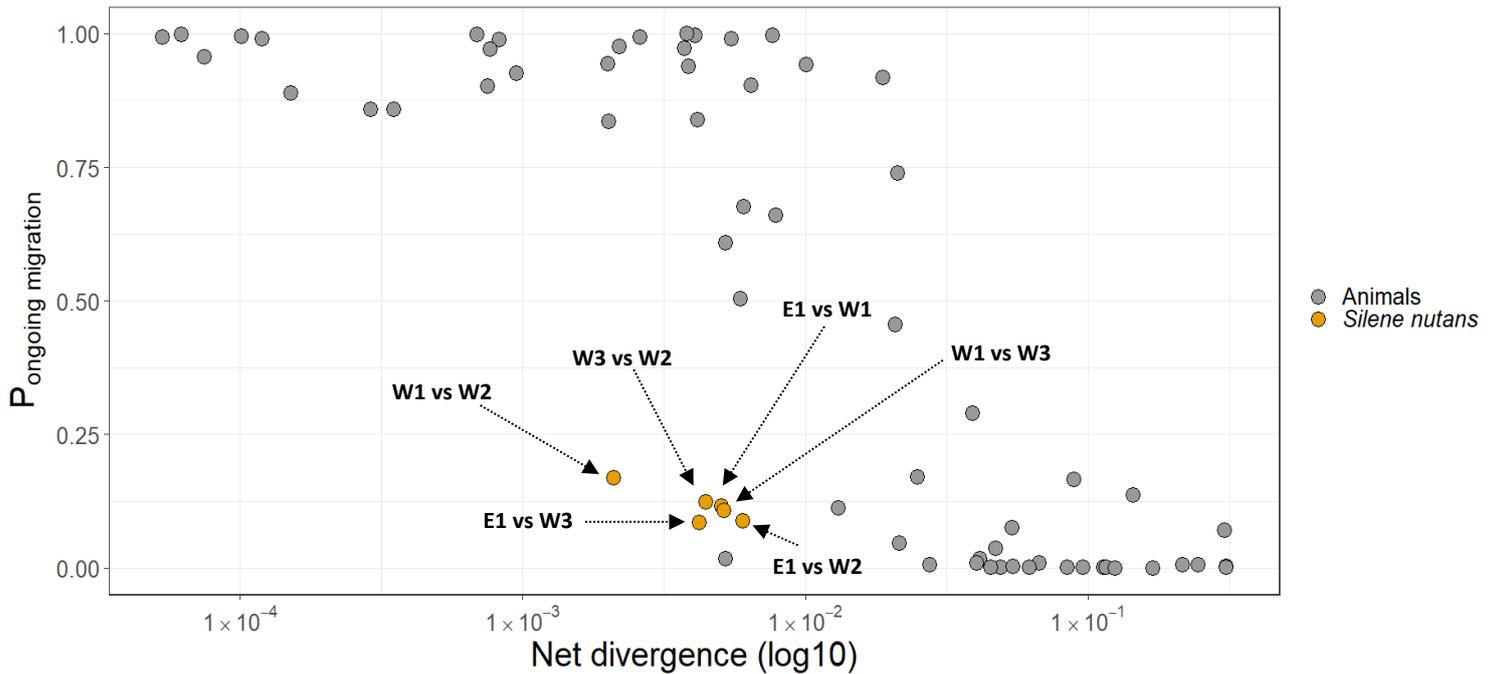
When looking at levels of genetic diversity and effective sizes ( $N_e$ ), they are the highest for lineages W2 and W3. These differences of  $N_e$  could be explained by the bottleneck effect in the glacial refugia that could have been more severe for W1 than W2 and W3.

This highlights also a discrepancy between the currently known geographical distribution of *S. nutans* lineages (E1 > W1 > W3 > W2) and their estimated  $N_e$  (W3 > W2 > W1 > E1). Such discrepancies between current populations size and  $N_e$  can be expected as  $N_e$  is calculated as the number of reproductive individuals in a population (and not the overall number of individuals). For example, varying selfing rates among populations can impact the estimated  $N_e$  with outcrossers generally having an higher  $N_e$  due to highest theoretical number of reproducing individuals (cited in (Muyle *et al.*, 2021)). *S. nutans* is mainly outcrossing, yet it is self-compatible with some occurrence of selfing (Van Rossum & Prentice, 2004; Vanderplanck *et al.*, 2020). We can thus speculate that varying selfing rates among populations could bias the estimation of  $N_e$  and lead to this discrepancy. Variation in our sampling strategy across lineages might also have influenced our estimates of  $N_e$ . Indeed, it is well known for European plant lineages that current diversity in regional samples may depend on their geographical distance from the last glacial refugia (Hewitt, 2000; Petit *et al.*, 2003). In the present study, irrespective of the lineage, we sampled individuals in the marginal range area of the lineages. However, their distribution is wider and goes beyond those sampled areas (Van Rossum *et al.*, 2018). For lineage W1, sampling was nearly exhaustive as we sampled individuals near the glacial refugia and beyond, spanning almost the whole geographical distribution of this lineage (Martin *et al.*, 2016; Van Rossum *et al.*, 2018). However, it is not the case for E1 as we only sampled populations at the margin of its geographical repartition and far from its glacial refugia, as highlighted with the relatively lower allelic richness of individuals from our sampled populations (Martin *et al.*, 2016; Van Rossum *et al.*, 2018). Hence, our estimate of  $N_e$  in E1 could have been biased downwards, due to restricted sampling. For W3, its geographical distribution extends in south eastern Europe and we only sampled individuals close to its glacial refugia in the Alps, with high allelic richness of the W3 populations (Martin *et al.*, 2016; Van Rossum *et al.*, 2018). This might be also true for W2 though we also sampled individuals more distant from the W2 glacial refugia in the Pyrenees (Van Rossum *et al.*, 2018). Alternatively, the high diversity observed for lineage W3 might suggest that its *census* size is higher than for the other lineages, or local population structure is higher, which is a feature often seen in current populations close to their glacial refugia (Heuertz *et al.*, 2004). The latter would be consistent with the geographical distribution of this lineage in mountain habitat which might increase local differentiation between populations.

The current distribution of the four lineages of *S. nutans* shows secondary contact zones at least between lineages E1 and W1 in south of England and south of Belgium, as well as potential overlap between geographical ranges of lineages W1 and W2 in central-western France (Figure 1). Yet, at these contact zones and overlapping areas no hybridizing events are detected from our analyses which is in agreement with the strong RI demonstrated from the diallelic crosses of Van Rossum et al (in prep). Several reproductive isolation mechanisms could have played a role in the overall pattern of ongoing speciation in the lineages of *S. nutans*. Regarding the pre-zygotic barriers, no evidence for pollinator isolation (pre-mating barrier) was identified between lineages E1 and W1 at secondary contact zone in south of Belgium : specialized pollinators were found on individuals of both lineages and pollen flow between/within lineages were similar (Cornet et al., 2022 in press). Pollen-stigma incompatibilities were identified between these Belgian lineages, potentially representing a post-pollination pre-zygotic reproductive barrier with pollen tubes having less probability to develop when resulting from inter-lineages pollination (Van Rossum *et al.*, 1996). Lineages E1 and W1 also exhibit morphological differences and different flowering time (De Bilde, 1973; Van Rossum, 2000), which might also generate pre-zygotic reproductive barrier (Baack *et al.*, 2015). These two lineages represent distinct edaphic ecotypes in Belgium, lineage E1 being specialized on calcicolous soil and W1 on siliceous ones (De Bilde, 1973; Van Rossum *et al.*, 1996; Van Rossum, 2000). This ecotypic specialization seems to have occurred following the divergence of both lineages in allopatry (De Bilde, 1973; Martin *et al.*, 2016; Van Rossum *et al.*, 2018). Soil specialization and habitat isolation might represent a pre-zygotic barrier, with local adaptation to different soil type playing a role in the creation of genetic incompatibilities between maladapted alleles in the hybrids, or alternatively association of soil specialization alleles with other genetic incompatibilities could have resulted from a coupling effect (Bierne *et al.*, 2011). Finally, post-zygotic reproductive isolation might also be the result of PNIs, with disruption of lineage-specific plastid-nuclear co-adaptation in hybrids (Postel *et al.*, 2022). Even though divergence between lineages of *S. nutans* seems to be recent given the split times, repeated bottlenecks or founding events experienced during post-glacial recolonization could have shaped the plastid genetic diversity independently in each lineage, increasing probability to observe lineage-specific plastid-nuclear co-adaptation and PNIs in hybrids (Postel *et al.*, 2022).

#### *4.2. Rapid speciation between lineages of S. nutans*

Speciation is supposed to be gradual but can sometimes be fast (Nosil *et al.*, 2017; Stankowski & Ravinet, 2021). Rapid speciation can be the results of various processes such as rapid environmental



**Fig.7 Grey zone of speciation representing data for animals and the pairs of lineages of *Silene nutans*.** Data for animals come from Roux et al. 2016.

turnover, genome rearrangement generating genetic incompatibilities, polyploidization, founder events... (Nosil *et al.*, 2017). Speciation seems to be quite sudden between lineages of *S. nutans*, as suggested with strong RI despite low overall levels of genetic divergence as compared to overall patterns in plants or animals (Roux *et al.*, 2016a). Indeed, levels of divergence between all pairs seem to be more or less the same, yet it is slightly higher between E1 and the western lineages, lower with W3 vs all the others and the lowest between W1 and W2. Yet probability of ongoing migration is very low. This pattern is different from what has been observed in a survey across many pairs of species (Figure 7 – (Roux *et al.*, 2016a)) and might represent an argument for a rapid speciation in *S. nutans*, although the survey was restricted to animal species and such a comparative analysis for plants is still missing.

When crossed in both directions, these lineages exhibit various levels of hybrid mortality and chlorosis (Van Rossum et al, 2022 in prep). Level of RI does not seem to reflect levels of genetic divergence between lineages. Except with E1 being the most divergent and leading to highest levels of RI when used in crosses, mortality is higher between W1 and W2 that are less divergent between each other than when W3 is used, while it is more genetically dissimilar. This highlights rapid evolution of reproductive barriers between lineages: they are slightly divergent from one another but strongly isolated. Especially, when represented in the grey zone of speciation, pairs of lineages are less divergent regarding the other organisms with a probability of ongoing migration (i.e. inter-breeding)

is almost equal to 0 (Figure 7) (Roux *et al.*, 2016a). Except for comparison between W1/W2 and E1/W1, where distinction between “true” species and semi-isolated one is ambiguous, lineages are considered as distinct species (Figure 7).

As discussed above, former studies using these lineages identified potential involvement of PNIs reducing hybrid viability and fertility (Postel *et al.*, 2022). There are growing evidences for their involvement in RI, especially as they might be one of the first post-zygotic reproductive barriers to evolve during speciation (Barnard-Kubow *et al.*, 2016; Postel & Touzet, 2020). Their study might be complicated by the fact that plastid genes are generally strongly conserved (Jansen *et al.*, 2007), so lineage specific co-adaptation between plastid and nuclear genes might not be evident except if time of isolation is high, with enough time to accumulate divergence between lineages. But in some angiosperm species, and especially *Silene* species, acceleration of the rate of evolution of the plastid genome has been identified (Barnard-Kubow *et al.*, 2014; Sloan *et al.*, 2014; Ruhlman & Jansen, 2018; Shrestha *et al.*, 2019). This acceleration might speed up plastid-nuclear co-adaptation and divergence at these nuclear and plastid genes in isolated lineages: if evolution of the plastid genes is higher, the interacting nuclear loci will follow up, increasing divergence between lineages but potentially only at these nuclear and plastid genes. In lineages of *S. nutans*, we observed high diversity pattern of the plastid genes, with numerous lineage specific mutations and potentially accelerated rate of plastid genome evolution (Postel *et al.*, 2022). If PNIs are one of the main barriers inducing RI between the lineages of *S. nutans*, then rapid evolution of reproductive barriers and strong RI despite low levels of divergence at nuclear loci might be driven by rapid evolution of the plastid genome in this species. In another angiosperm species, *Campanulastrum americanum*, similar levels of RI were observed between isolated clades of this species, with similar acceleration of the evolutionary rate of the plastid genome (Barnard-Kubow *et al.*, 2014, 2016; Barnard-Kubow & Galloway, 2017). Inferences of evo-demographic scenarios with the different clades of this species have not been done yet but we might expect a similar relationship between genetic divergence at nuclear loci and levels of RI between them. For the other angiosperm species with high rates of plastid genome evolution, independent lineages were not identified so far and levels of RI not assessed, but we could speculatively expect similar trend in genetic divergence and RI.

#### 4.3. Outliers of differentiation unlikely represent speciation islands

We don't know which of the above discussed reproductive barriers are mainly driving the fast-evolving RI between lineages. Also, we don't know whether genetic incompatibilities in inter-lineages hybrids of *S. nutans* are mostly the result of local adaptation (extrinsic barriers), e.g. to different soil conditions as observed for lineages E1 and W1, and/or of mutations accumulations during allopatric phase, mainly

through genetic drift during rapid post-glacial recolonization and repeated bottlenecks (Hewitt, 2000, 2011). It is possible that genetic drift acted in conjunction with some phases of local adaptation to shape the nuclear and plastid genetic patterns of diversity and nuclear levels of differentiation observed now.

Looking at outliers of genetic differentiation (outliers of  $F_{ST}$  – FO) might help answer the above question. FO can often be seen at early stages of speciation, where differentiation is restricted to small parts of the genome (Burri, 2017). They might be the result either of “speciation islands” or “barrier loci”, i.e. genomic regions resistant to gene flow, or the result of linked selection in low recombination regions (Cruickshank & Hahn, 2014; Ravinet *et al.*, 2017). In the former, these speciation islands are resulting from selection on locally adapted or coadapted alleles and selection against migrant ones (Liu *et al.*, 2020). In the latter, high differentiation might result from local genomic reduction in effective population size due to low recombination rate associated with either positive or background selection (Burri, 2017; Schluter & Rieseberg, 2022), but loci in these regions are not involved in RI. Supposing linked selection is the main driver of high differentiation, identification of FO and of the recombination context surrounding these loci might theoretically help distinguish whether local adaptation is responsible or not for emergence of divergence and potential generation of genetic incompatibilities. In the present study, FO seems to be mainly located in low recombination regions, and they are associated with lower values of  $\pi$  (Cruickshank & Hahn, 2014). Yet, even though levels of absolute divergence ( $D_{xy}$ ) should not be different between FO and FNO according to this hypothesis (Cruickshank & Hahn, 2014), they are here. Loci with high genetic differentiation between lineages of *S. nutans* might be essentially the results of linked selection in low recombination genomic regions, with associated reduction genetic diversity. But higher  $D_{xy}$  on average at these loci might suggest at least a partial role of adaptive differentiation, and possibly in relation to rapid evolution of barrier loci involved in an ongoing speciation process. Yet, identification of barrier loci is only possible when there is gene flow homogenizing the genome except at these loci, leading to heterogeneous levels of differentiation along the genome. Here, it seems that divergence between lineages occurred in complete allopatry, so without gene flow. As a result, the whole genome of the lineages might be at least slightly differentiated. To identify them if they exist, we would have to remove the effect due to the genomic position of the loci (and the subsequent recombination rate). Recombination rates is known to be heterogeneous along the genome and especially low in high gene density regions and near chromosomes centromeres (Brazier & Glémin, 2022). Having access to a fully annotated genome and a recombination map would also help distinguish whether high genetic differentiation is the result of barrier loci involved in RI or background selection in low recombination regions.

## 5. References

- Baack E, Melo MC, Rieseberg LH, Ortiz-Barrientos D. 2015.** The origins of reproductive isolation in plants. *New Phytologist* **207**: 968–984.
- Barnard-Kubow KB, Galloway LF. 2017.** Variation in reproductive isolation across a species range. *Ecology and Evolution* **7**: 9347–9357.
- Barnard-Kubow KB, Sloan DB, Galloway LF. 2014.** Correlation between sequence divergence and polymorphism reveals similar evolutionary mechanisms acting across multiple timescales in a rapidly evolving plastid genome. *BMC Evolutionary Biology* **14**: 1–10.
- Barnard-Kubow KB, So N, Galloway LF. 2016.** Cytonuclear incompatibility contributes to the early stages of speciation. *Evolution* **70**: 2752–2766.
- Bierne N, Welch J, Loire E, Bonhomme F, David P. 2011.** The coupling hypothesis : why genome scans may fail to map local adaptation genes. *Molecular Ecology* **20**: 2044–2072.
- De Bilde J. 1973.** Etude génécologique du *Silene nutans* L. en Belgique : Populations du *Silene nutans* L. sur substrats siliceux et calcaires. *Revue Générale de Botanique* **80**: 161–176.
- Brazier T, Glémin S. 2022.** Diversity and determinants of recombination landscapes in flowering plants. *PLoS Genetics* **18**: e1010141.
- Burri R. 2017.** Interpreting differentiation landscapes in the light of long-term linked selection. *Evolution Letters*: 118–131.
- Burri R, Nater A, Kawakami T, Mugal CF, Olason PI, Smeds L, Suh A, Dutoit L, Bure S, Garamszegi LZ, et al. 2015.** Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. *Genome Research* **25**: 1656–1665.
- Butlin RK, Galindo J, Grahame JW, Sheffield S. 2008.** Sympatric , parapatric or allopatric : the most important way to classify speciation ? *Philosophical Transactions of the Royal Society B: Biological Sciences*: 2997–3007.
- Butlin RK, Stankowski S. 2020.** Is it time to abandon the biological species concept ? No. *Molecular Biology & Genetics* **7**: 1400–1401.
- Coughlan JM, Matute DR. 2020.** The importance of intrinsic postzygotic barriers throughout the speciation process. *Philosophical Transactions of the Royal Society B: Biological Sciences* **375**: 20190533.
- Coyne JA, Orr HA. 2004.** *Speciation*. (Sinauer Associates, Ed.). Oxford University Press.
- Cruikshank TE, Hahn MW. 2014.** Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology* **23**: 3133–3157.
- Feder JL, Egan SP, Nosil P. 2012.** The genomics of speciation-with- gene-flow. *Trends in Genetics* **28**:

342–350.

**Fishman L, Sweigart AL. 2018.** When Two Rights Make a Wrong : The Evolutionary Genetics of Plant Hybrid Incompatibilities. *Annual Review of Plant Biology* **69**: 707–731.

**Fraïsse C, Popovic I, Mazoyer C, Spataro B, Delmotte S, Romiguier J, Loire É, Simon A, Galtier N, Duret L, et al. 2021.** DILS : Demographic inferences with linked selection by using ABC. *Molecular Ecology Ressources* **00**: 1–16.

**Heuertz M, Hausman J-F, Hardy OJ, Vendramin GG, Frascaria-Lacoste N, Vekemans X. 2004.** NUCLEAR MICROSATELLITES REVEAL CONTRASTING PATTERNS OF GENETIC STRUCTURE BETWEEN WESTERN AND SOUTHEASTERN EUROPEAN POPULATIONS OF THE COMMON ASH (*FRAXINUS EXCELSIOR* L.). *Evolution* **58**: 976–988.

**Hewitt G. 2000.** The genetic legacy of the Quaternary ice ages. *Nature* **405**: 907–913.

**Hewitt GM. 2011.** Quaternary phylogeography : the roots of hybrid zones. *Genetica* **139**: 617–638.

**Jansen RK, Cai Z, Raubeson LA, Daniell H, Claude W, Leebens-mack J, Mu KF, Guisinger-bellian M, Haberle RC, Hansen AK, et al. 2007.** Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. **104**: 19369–19374.

**Krasovec M, Chester M, Ridout K, Filatov DA. 2018.** The Mutation Rate and the Age of the Sex Chromosomes in *Silene latifolia*. *Current Biology* **28**: 1832–1838.

**Kulmuni J, Butlin RK, Lucek K, Savolainen V, Westram AM. 2020.** Towards the completion of speciation : the evolution of reproductive isolation beyond the first barriers. *Philosophical Transactions of the Royal Society B: Biological Sciences* **375**: 20190528.

**Langmead B, Salzberg SL. 2012.** Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**: 357–359.

**Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup 1000 genome project data processing. 2009.** The Sequence Alignment / Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

**Liu X, Glémin S, Karrenberg S. 2020.** Evolution of putative barrier loci at an intermediate stage of speciation with gene flow in champions (*Silene*). *Molecular Ecology* **29**: 3511–3525.

**Lowry DB. 2012.** Ecotypes and the controversy over stages in the formation of new species. *Biological Journal of the Linnean Society* **106**: 241–257.

**Martin H. 2016.** Processus de spéciation et impact des systèmes de reproduction dans le genre *Silene*.

**Martin H, Touzet P, Dufay M, Godé C, Schmitt E, Lahiani E, Delph LF, Van Rossum F. 2017.** Lineages of *Silene nutans* developed rapid, strong, asymmetric postzygotic reproductive isolation in allopatry. *Evolution* **71**: 1519–1531.

**Martin H, Touzet P, Rossum F Van, Delalande D, Arnaud J. 2016.** Phylogeographic pattern of range expansion provides evidence for cryptic species lineages in *Silene nutans* in Western Europe. *Heredity*

116: 286–294.

**Matute DR, Cooper BS. 2021.** Comparative studies on speciation : 30 years since Coyne and Orr. *Evolution* **75**: 764–778.

**Muyle A, Martin H, Zemp N, Mollion M, Gallina S, Tavares R, Silva A, Bataillon T, Widmer A, Glémin S, et al. 2021.** Dioecy Is Associated with High Genetic Diversity and Adaptation Rates in the Plant Genus *Silene*. *Molecular Biology and Evolution* **38**: 805–818.

**Nosil P, Feder JL, Flaxman SM, Gompert Z. 2017.** Tipping points in the dynamics of speciation. *Nature Ecology & Evolution* **1**: 0001.

**Petit RJ, Aguinalalde I, de Beaulieu J-L, Bittkau C, Brewer S, Cheddadi R, Ennos R, Fineschi S, Grivet D, Lascoux M, et al. 2003.** Glacial Refugia : Hotspots But Not. *SCIENCE* **300**: 1563–1566.

**Postel Z, Poux C, Gallina S, Varré J-S, Godé C, Schmitt E, Meyer E, Van Rossum F, Touzet P. 2022.** Reproductive isolation among lineages of *Silene nutans* (Caryophyllaceae): A potential involvement of plastid-nuclear incompatibilities. *Molecular Phylogenetics and Evolution* **169**: 107436.

**Postel Z, Touzet P. 2020.** Cytonuclear Genetic Incompatibilities in Plant Speciation. *Plants*.

**Presgraves DC. 2010.** The molecular evolutionary basis of species formation. *Nature Reviews Genetics* **11**: 175–180.

**Pudlo P, Marin J, Estoup A, Cornuet J, Gautier M, Robert CP. 2016.** Reliable ABC model choice via random forests. *Bioinformatics* **32**: 859–866.

**Ravinet M, Faria R, Butlin RK, Galindo J, Bierne N, Rafajlovic M, Noor MAF, Mehlig B, Westram AM. 2017.** Interpreting the genomic landscape of speciation : a road map for finding barriers to gene flow. *Journal of Evolutionary Biology* **30**: 1450–1477.

**Van Rossum F. 2000.** Amplitude synécologique de *Silene nutans* L. (Caryophyllaceae). *DUMORTIERA* **75**: 11–24.

**Van Rossum F, De Bilde J, Lefèbvre C. 1996.** Barriers to hybridization in calcicolous and silicicolous populations of *Silene nutans* from Belgium. *Royal Botanical Society of Belgium* **129**: 13–18.

**Van Rossum F, Martin H, Le Cadre S, Brachi B, Christenhusz MJM, Touzet P. 2018.** Phylogeography of a widely distributed species reveals a cryptic assemblage of distinct genetic lineages needing separate conservation strategies. *Perspectives in Plant Ecology, Evolution and Systematics* **35**: 44–51.

**Van Rossum F, Prentice HC. 2004.** Structure of allozyme variation in Nordic *Silene nutans* (Caryophyllaceae): Population size, geographical position and immigration history. *Biological Journal of the Linnean Society* **81**: 357–371.

**Van Rossum F, Vekemans X, Meerts P, Gratia E, Lefèbvre C. 1997.** Allozyme variation in relation to ecotypic differentiation and population size in marginal populations of *Silene nutans*. *Heredity* **78**: 552–560.

**Roux C, Fra C, Romiguier J, Anciaux Y, Galtier N, Bierne N. 2016a.** Shedding Light on the Grey Zone of

Speciation along a Continuum of Genomic Divergence. *PLoS Biology* **14**: e2000234.

**Roux C, Fraïsse C, Romiguier J, Anciaux Y, Galtier N, Bierne N. 2016b.** Shedding Light on the Grey Zone of Speciation along a Continuum of Genomic Divergence. *PLoS Biology* **14**: 1–22.

**Ruhlman TA, Jansen RK. 2018.** *Aberration or Analogy ? The Atypical Plastomes of Geraniaceae*. Elsevier Ltd.

**Schluter D, Rieseberg LH. 2022.** Three problems in the genetics of speciation by selection. *PNAS* **119**: e2122153119.

**Seehausen O, Butlin RK, Keller I, Wagner CE, Boughman JW, Hohenlohe PA, Peichel CL, Saetre G. 2014.** Genomics and the origin of species. *Nature* **15**: 176–192.

**Shang H, Hess J, Pickup M, Field DL, Ingvarsson PK, Liu J, Lexer C, Field DL. 2020.** Evolution of strong reproductive isolation in plants : broad-scale patterns and lessons from a perennial model group. *Philosophical Transactions of the Royal Society B: Biological Sciences* **375**: 20190544.

**Shen W, Le S, Li Y, Hu F. 2016.** SeqKit : A Cross-Platform and Ultrafast Toolkit for FASTA / Q File Manipulation. *PLoS ONE* **11**: 1–10.

**Shrestha B, Weng M, Theriot EC, Gilbert LE, Ruhlman TA, Krosnick SE, Jansen RK. 2019.** Highly accelerated rates of genomic rearrangements and nucleotide substitutions in plastid genomes of *Passiflora* subgenus *Decaloba*. *Molecular Phylogenetics and Evolution* **138**: 53–64.

**Sloan DB, Triant DA, Forrester NJ, Bergner LM, Wu M, Taylor DR. 2014.** A recurring syndrome of accelerated plastid genome evolution in the angiosperm tribe *Sileneae* (Caryophyllaceae). *Molecular Phylogenetics and Evolution* **72**: 82–89.

**Stankowski S, Ravinet M. 2021.** Defining the speciation continuum. *Evolution*: 1–18.

**Tsagkogeorga G, Cahais V, Galtier N. 2012.** The Population Genomics of a Fast Evolver : High Levels of Diversity, Functional Constraint, and Molecular Adaptation in the Tunicate *Ciona intestinalis*. *Genome Biology and Evolution* **4**: 852–861.

**Turner TL, Hahn MW, Nuzhdin S V. 2005.** Genomic Islands of Speciation in *Anopheles gambiae*. *PLoS Biology* **3**: e285.

**Vanderplanck M, Touzet P, Rossum F Van, Lahiani E, Cauwer I De, Dufaÿ M. 2020.** Does pollination syndrome reflect pollinator efficiency in *Silene nutans* ? *Acta Oecologica* **105**: 103557.

**Wu C. 2001.** The genic view of the process of speciation. *Journal of Evolutionary Biology* **14**: 851–865.

## 6. Annexes

**Suppl. Table S1**

Detailed of the sequenced individuals for all four lineages.

Lineage	Name	Identifier	Time of sequencing	Locality	Region	Country	Latitude	Longitude
<b>E1</b>	<b>Si1</b>	D19_6	Muyle et al. 2021	Neckarzellern	Bade-Wurtemberg	GERMANY	49.3227	9.1289
	<b>Si2</b>	D19_15	Muyle et al. 2021					
	<b>Si3</b>	LUX7_7	Muyle et al. 2021	Tandel	Vianden	LUXEMBURG	49.91056667	6.140283333
	<b>Si4</b>	LUX7_5	Muyle et al. 2021					
	<b>Si12</b>	FIN1_3	Muyle et al. 2021	Turku	Varsinais Suomi	FINLAND	60.4584	22.2565
	<b>Si13</b>	FIN1_4	Muyle et al. 2021					
	<b>Si14</b>	UK14_10	Muyle et al. 2021	Littlehampton, Sussex	Sussex	ENGLAND	50.80073333	-0.558416667
	<b>Si18</b>	UK_16.10	Muyle et al. 2021	Folkestone, Kent	Kent	ENGLAND	51.10206667	1.23645
	<b>Si23</b>	UK_16.11	Muyle et al. 2021					
	<b>Si20</b>	BUIS_A5	Muyle et al. 2021	Buis	Wallonia	BELGIUM	50.68863	4.56105
	<b>Si21</b>	BUIS_A12	Muyle et al. 2021					
<b>W1</b>	<b>Si5</b>	TROJ_09	2022	Saint Trojan les bains	Charente-Maritime	FRANCE	45.816257	-1.220005
	<b>Si6</b>	TROJ_07	2022					
	<b>Si7</b>	F1_05	2022	Serre forest	Jura	FRANCE	47.1666667	5.556944444
	<b>Si10</b>	F1_11	2022					
	<b>Si8</b>	OLL_210.4	Muyle et al. 2021	Olloy-sur-Viroin	Wallonia	BELGIUM	50.06888889	4.606111111
	<b>Si9</b>	OLL_C20	Muyle et al. 2021					
	<b>Si11</b>	UK15_16	Muyle et al. 2021	Dungeness, Kent	Kent	ENGLAND	50.9329	0.959083333
	<b>Si17</b>	BZH_1.4	Muyle et al. 2021	Arzal	Morbihan	FRANCE	47.506	-2.404222222
	<b>Si22</b>	BZH_1.1	Muyle et al. 2021					
	<b>Si19</b>	AND_7	Muyle et al. 2021	Les Andelys	Normandie	FRANCE	49.25722222	1.378611111
<b>Si25</b>	AND_1	Muyle et al. 2021						
<b>W2</b>	<b>Si24</b>	PYR_2.6a	Muyle et al. 2021	Aranvielle	Hte Pyrénées	FRANCE	42.81033	0.4092
	<b>Si26</b>	PYR_2.9b	Muyle et al. 2021					
	<b>Si27</b>	ARG_11	2022	Argentine	Dordogne	FRANCE	45.472166667	0.37922222
	<b>Si28</b>	ARG_06	2022					
	<b>Si29</b>	BAT_12	2022	Castillon-la-Bataille	Gironde	FRANCE	44.8653605446	-
	<b>Si30</b>	BAT_08	2022					0.0214855318
	<b>Si31</b>	YVE_04	2022	Yves	Charente-Maritime	FRANCE	46.010023	-1.552001
	<b>Si32</b>	YVE_21	2022					
	<b>Si33</b>	BEN3_09	2022	Saint Benoit	Poitou-Charente	FRANCE	46.54175	0.337666667
	<b>Si34</b>	BEN3_21	2022					
<b>Si35</b>	MES_11	2022	Mescher	Gironde	FRANCE	45.557308	-0.962765	
<b>Si36</b>	MES_04	2022						
<b>W3</b>	<b>Si15</b>	FQ_3.6.2	Muyle et al. 2021	Arvieux, Queyras	Queyras - Alpes	FRANCE	44.7776	6.739806
	<b>Si16</b>	FQ_3.7.2	Muyle et al. 2021					
	<b>Si37</b>	AIG_12	2022	Aiguebelle	Savoie - Alpes	FRANCE	45.53778	6.30761
	<b>Si38</b>	AIG_13	2022					
	<b>Si39</b>	AUS_17	2022	Aussois	Savoie - Alpes	FRANCE	45.23419	6.74638
	<b>Si40</b>	AUS_05	2022					
	<b>Si41</b>	BAU_04	2022	Granges-sur-Baume	Jura	FRANCE	46.712361	5.644472
	<b>Si42</b>	BAU_05	2022					
	<b>Si43</b>	SIG_08	2022	Siguret (Lac), Saint-André-d'Embrun	Alpes	FRANCE	44.6146	6.56161
	<b>Si44</b>	SIG_13	2022					
<b>Si45</b>	F5_08	2022	Les Granges	Jura	FRANCE	47.1375	5.966944444	
<b>Si46</b>	F5_04	2022						

**Suppl. Table S2**

Parameters and prior values for DILS analysis.

Type		DILS 2 pop	DILS 4 pops
<b>Parameters</b>	Genomic regions:	Coding	Coding
	max_N_tolerated :	0.2	0.5
	nMin :	10	8
	Lmin :	30	200
	use jSFS :	yes	0
	Change in population size :	constant	constant
<b>Priors</b>	$\mu$ (per base pair, per generation) :	0.000000003	0.00000000731
	Ratio $r/\mu$ :	0.1	1
	Time of split (in number of generation) :	min = 100 max = 10 000 000	min = 0 max = 2 000 000
	Populations size ( $N_e$ ) :	min = 100 max = 1 000 000	min = 0 max = 1 000 000
	Migration rates ( $4.N_e.m$ ) :	min = 0.4 max = 20	min = 0.4 max = 40
	Model for barriers :	bimodal	Beta

*max\_N\_tolerated* = maximum proportion of missing data in the sequence of a gene for an individual beyond which this sequence is excluded ; *nMin* = minimum number of individuals sequence within lineage for which the gene sequence is not excluded (regarding the number of individuals after filtering with *max\_N\_tolerated*) ; *Lmin* = minimum number of treatable sites below which a gene is excluded ; *jSFS* = jointed Site Frequency Spectrum – will be used as a summary statistic if set to *yes* ;  $\mu$  = mutation rate ; *Ratio  $r/\mu$*  = ratio of recombination (per base pair per generation) over mutation ; *Time of split* = speciation time ; *Population size* = number of diploid individuals within current and ancestral lineages ; *Migration rate* = calculated with  $4.N_e.m$  with *m* being the fraction of each subpopulation composed of new migrants at each generation

**Suppl. Table S3**

Details of the results of the Wilcoxon Mann-Whitney and Student T-test conducted on the number of private polymorphic sites ( $S_X$ ), the nucleotide diversity within lineage ( $\pi$ ), Watterson's  $\theta$  estimator of genetic diversity and Tajima's D.

**WILCOXON-MANN WHITNEY RESULTS**

Summary statistic	Pairwise comparison	p-value	W statistic	Interpretation
$S_X$	E1 vs W1	0.63	6 025 644	NS
	E1 vs W2	< 2.2e-16	5 288 840	W2 > E1
	E1 vs W3	< 2.2e-16	4 495 318	W3 > E1
	W1 vs W2	< 2.2e-16	5 022 038	W2 > W1
	W1 vs W3	< 2.2e-16	4 366 087	W3 > W1
	W2 vs W3	< 2.2e-16	5 180 602	W3 > W2
$\pi$	E1 vs W1	0.83	5 967 900	NS
	E1 vs W2	2.41e-07	5 557 701	W2 > E1
	E1 vs W3	< 2.2e-16	5 184 869	W3 > E1
	W1 vs W2	2.19e-08	5 521 866	W2 > W1
	W1 vs W3	< 2.2e-16	5 125 600	W3 > W1
	W2 vs W3	1.26e-07	5 547 223	W3 > W2
$\theta$	E1 vs W1	0.51	5 930 631	NS
	E1 vs W2	1.25e-12	5 397 235	W2 > E1
	E1 vs W3	< 2.2e-16	4 910 782	W3 > E1
	W1 vs W2	1.92e-14	5 351 184	W2 > W1
	W1 vs W3	< 2.2e-16	4 854 117	W3 > W1
	W2 vs W3	3.14e-11	5 434 808	W3 > W2

**STUDENT T-TEST RESULTS**

Summary statistic	Pairwise comparison	p-value	T-test	Df	Interpretation
Tajima's D	E1 vs W1	0.008	2.62	6895.2	W1 > E1
	E1 vs W2	< 2.2e-16	11.15	6910.6	W2 > E1
	E1 vs W3	< 2.2e-16	15.99	6913.4	W3 > E1
	W1 vs W2	< 2.2e-16	9.57	6917.4	W2 > W1
	W1 vs W3	< 2.2e-16	13.82	6916.6	W3 > W1
	W2 vs W3	4.87e-06	4.57	6918	W2 > W3

Df = degree of freedom

**Suppl. Table S4**

Result of the Wilcoxon Mann-Whitney test between each pairwise comparison of lineages for  $S_f$ ,  $S_s$ ,  $D_a$ ,  $D_{xy}$  and  $F_{ST}$  summary statistics.

Statistic	Combination n°1	Combination n°2	P-value	W-statistic	Direction of difference
$S_f$	E1 vs W1	E1 vs W2	0.00	6 176 224	E1 vs W1 > E1 vs W2
	E1 vs W1	E1 vs W3	0.00	6 444 962	E1 vs W1 > E1 vs W3
	E1 vs W1	W1 vs W2	0.00	6 954 188	E1 vs W1 > W1 vs W2
	E1 vs W1	W1 vs W3	0.00	6 477 030	E1 vs W1 > W1 vs W3
	E1 vs W1	W2 vs W3	0.00	6 608 629	E1 vs W1 > W2 vs W3
	E1 vs W2	E1 vs W3	0.00	6 254 323	E1 vs W2 > E1 vs W3
	E1 vs W2	W1 vs W2	0.00	6 762 926	E1 vs W2 > W1 vs W2
	E1 vs W2	W1 vs W3	0.00	6 286 702	E1 vs W2 > W1 vs W3
	E1 vs W2	W2 vs W3	0.00	6 418 588	E1 vs W2 > W2 vs W3
	E1 vs W3	W1 vs W2	0.00	6 493 760	E1 vs W3 > W1 vs W2
	E1 vs W3	W1 vs W3	0.45	6 018 271	NS
	E1 vs W3	W2 vs W3	0.00	6 150 584	E1 vs W3 > W2 vs W3
	W1 vs W2	W1 vs W3	0.00	5 510 933	W1 vs W3 > W1 vs W2
	W1 vs W2	W2 vs W3	0.00	5 644 658	W2 vs W3 > W1 vs W2
	W1 vs W3	W2 vs W3	0.00	6 118 084	W1 vs W3 > W2 vs W3
$S_s$	E1 vs W1	E1 vs W2	0.00	8 609 846	E1 vs W1 > E1 vs W2
	E1 vs W1	E1 vs W3	0.00	8 191 604	E1 vs W1 > E1 vs W3
	E1 vs W1	W1 vs W2	0.00	7 742 350	E1 vs W1 < W1 vs W2
	E1 vs W1	W1 vs W3	0.00	8 317 058	E1 vs W1 > W1 vs W3
	E1 vs W1	W2 vs W3	0.00	8 264 550	E1 vs W1 > W2 vs W3
	E1 vs W2	E1 vs W3	0.00	5 617 270	E1 vs W2 < E1 vs W3
	E1 vs W2	W1 vs W2	0.00	5 093 929	E1 vs W2 < W1 vs W2
	E1 vs W2	W1 vs W3	0.00	5 703 593	E1 vs W2 < W1 vs W3
	E1 vs W2	W2 vs W3	0.00	5 690 821	E1 vs W2 < W2 vs W3
	E1 vs W3	W1 vs W2	0.00	5 481 239	E1 vs W3 < W1 vs W2
	E1 vs W3	W1 vs W3	0.22	6 077 264	NS
	E1 vs W3	W2 vs W3	0.33	6 059 332	NS
	W1 vs W2	W1 vs W3	0.00	6 589 111	W1 vs W3 < W1 vs W2
	W1 vs W2	W2 vs W3	0.00	6 561 450	W2 vs W3 < W1 vs W2
	W1 vs W3	W2 vs W3	0.81	5 968 166	NS
$D_a$	E1 vs W1	E1 vs W2	0.00	6 421 876	E1 vs W1 > E1 vs W2
	E1 vs W1	E1 vs W3	0.00	6 930 183	E1 vs W1 > E1 vs W3
	E1 vs W1	W1 vs W2	0.00	8 446 608	E1 vs W1 > W1 vs W2
	E1 vs W1	W1 vs W3	0.00	6 825 450	E1 vs W1 > W1 vs W3
	E1 vs W1	W2 vs W3	0.00	7 207 554	E1 vs W1 > W2 vs W3
	E1 vs W2	E1 vs W3	0.00	6 482 626	E1 vs W2 > E1 vs W3
	E1 vs W2	W1 vs W2	0.00	8 024 432	E1 vs W2 > W1 vs W2
	E1 vs W2	W1 vs W3	0.00	6 388 034	E1 vs W2 > W1 vs W3
	E1 vs W2	W2 vs W3	0.00	6 769 410	E1 vs W2 > W2 vs W3
	E1 vs W3	W1 vs W2	0.00	7 590 361	E1 vs W3 > W1 vs W2
	E1 vs W3	W1 vs W3	0.28	5 896 506	NS
	E1 vs W3	W2 vs W3	0.00	6 287 825	E1 vs W3 > W2 vs W3
	W1 vs W2	W1 vs W3	0.00	4 330 128	W1 vs W3 > W1 vs W2
	W1 vs W2	W2 vs W3	0.00	4 697 148	W2 vs W3 > W1 vs W2
	W1 vs W3	W2 vs W3	0.00	6 368 396	W1 vs W3 > W2 vs W3
$D_{xy}$	E1 vs W1	E1 vs W2	0.07	6 135 554	NS
	E1 vs W1	E1 vs W3	0.03	6 167 663	E1 vs W1 > E1 vs W3
	E1 vs W1	W1 vs W2	0.00	7 067 334	E1 vs W1 > W1 vs W2
	E1 vs W1	W1 vs W3	0.01	6 188 421	E1 vs W1 > W1 vs W3
	E1 vs W1	W2 vs W3	0.00	6 296 360	E1 vs W1 > W2 vs W3
	E1 vs W2	E1 vs W3	0.71	6 017 026	NS

	E1 vs W2	W1 vs W2	0.00	6 915 088	E1 vs W2 > W1 vs W2
	E1 vs W2	W1 vs W3	0.54	6 036 394	NS
	E1 vs W2	W2 vs W3	0.06	6 142 991	NS
	E1 vs W3	W1 vs W2	0.00	6 892 944	E1 vs W3 > W1 vs W2
	E1 vs W3	W1 vs W3	0.81	6 006 304	NS
	E1 vs W3	W2 vs W3	0.13	6 112 896	NS
	W1 vs W2	W1 vs W3	0.00	5 095 293	W1 vs W3 > W1 vs W2
	W1 vs W2	W2 vs W3	0.00	5 191 570	W2 vs W3 > W1 vs W2
	W1 vs W3	W2 vs W3	0.20	6 092 728	NS
<b><math>F_{ST}</math></b>	E1 vs W1	E1 vs W2	0.00	6 533 284	E1 vs W1 > E1 vs W2
	E1 vs W1	E1 vs W3	0.00	7 159 242	E1 vs W1 > E1 vs W3
	E1 vs W1	W1 vs W2	0.00	8 502 081	E1 vs W1 > W1 vs W2
	E1 vs W1	W1 vs W3	0.00	7 041 376	E1 vs W1 > W1 vs W3
	E1 vs W1	W2 vs W3	0.00	7 494 225	E1 vs W1 > W2 vs W3
	E1 vs W2	E1 vs W3	0.00	6 621 844	E1 vs W2 > E1 vs W3
	E1 vs W2	W1 vs W2	0.00	8 013 098	E1 vs W2 > W1 vs W2
	E1 vs W2	W1 vs W3	0.00	6 499 877	E1 vs W2 > W1 vs W3
	E1 vs W2	W2 vs W3	0.00	6 962 564	E1 vs W2 > W2 vs W3
	E1 vs W3	W1 vs W2	0.00	7 420 640	E1 vs W3 > W1 vs W2
	E1 vs W3	W1 vs W3	0.13	5 861 188	NS
	E1 vs W3	W2 vs W3	0.00	6 324 582	E1 vs W3 > W2 vs W3
	W1 vs W2	W1 vs W3	0.00	4 437 226	W1 vs W3 > W1 vs W2
	W1 vs W2	W2 vs W3	0.00	4 876 443	W2 vs W3 > W1 vs W2
	W1 vs W3	W2 vs W3	0.00	6 450 560	W1 vs W3 > W2 vs W3

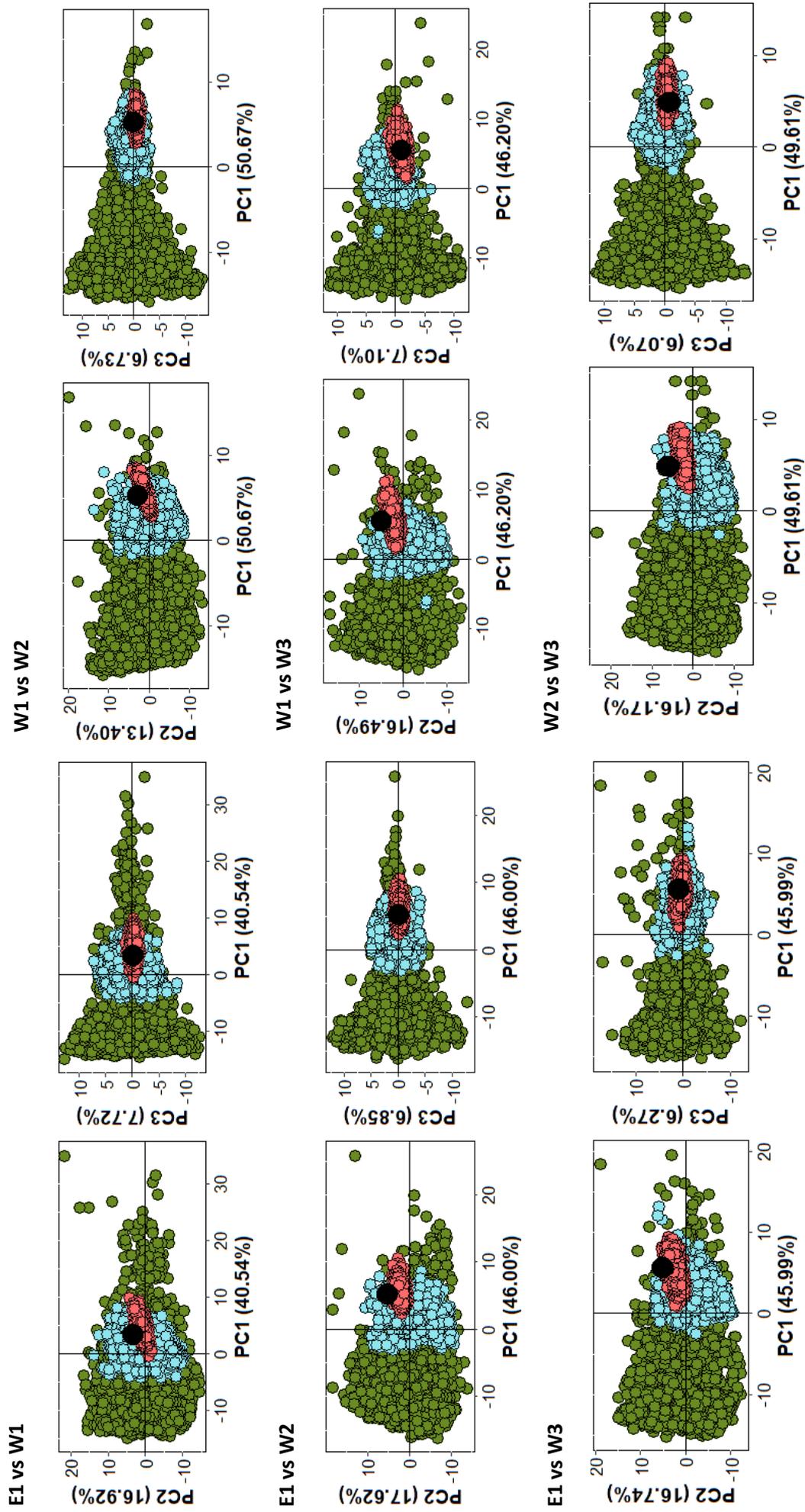
$S_f$ : Number of fixed polymorphic sites between lineage ;  $S_S$  : Number of shared polymorphic sites between lineages ;  $D_a$ : Net genetic divergence between lineages ;  $D_{xy}$ : Absolute genetic divergence between lineages ;  $F_{ST}$ : Genetic differentiation between lineages

**Suppl. Table S5**

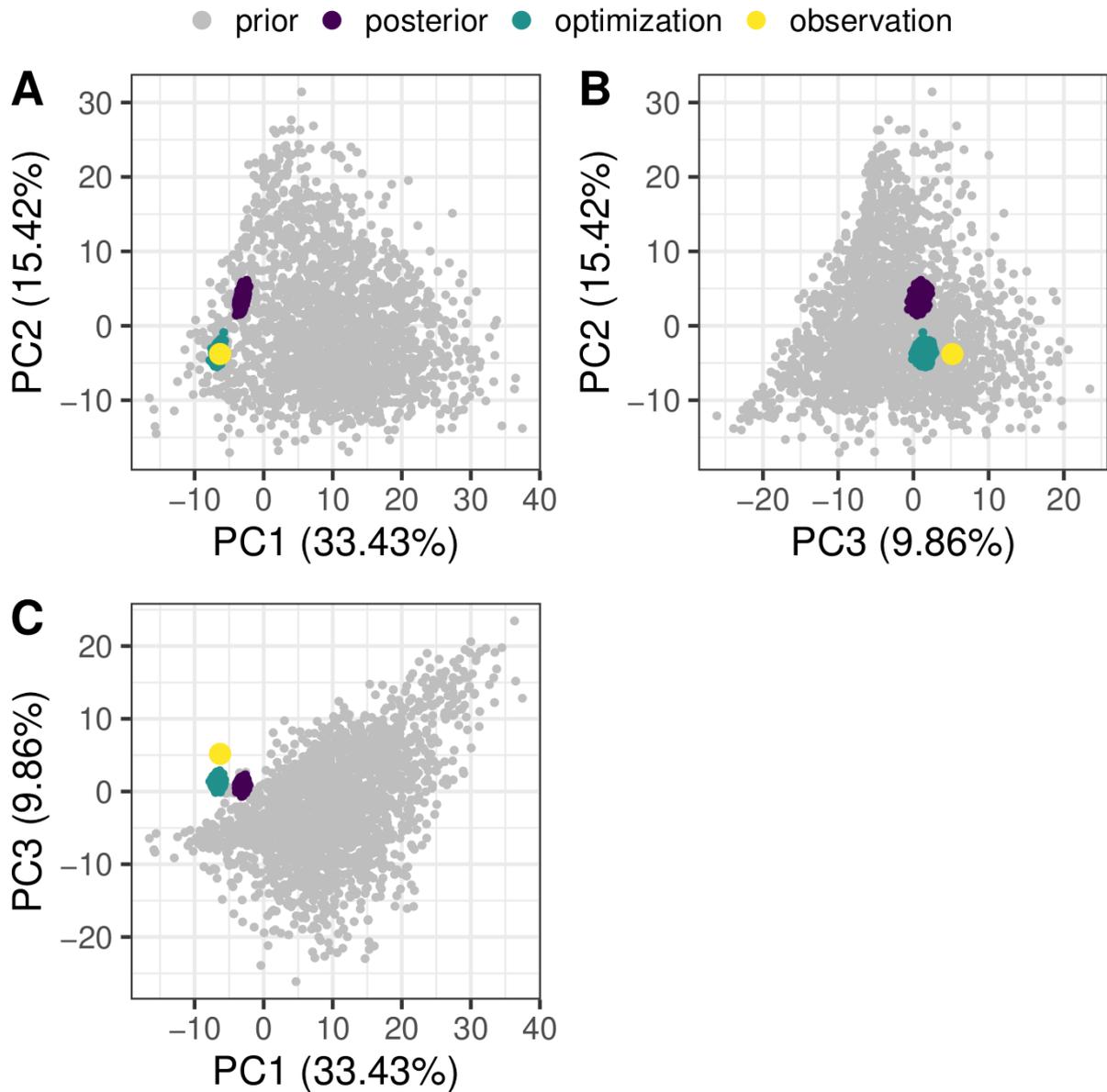
Means of  $\pi$ , number of recombination events,  $D_a$  and  $D_{xy}$  for outliers and non-outliers of  $F_{ST}$ , calculated for each pair of lineages.

Pairs of lineages	$F_{ST}$ type	$\pi$ A	$\pi$ B	Recombi A	Recombi B	$D_a$	$D_{xy}$
<b>E1 vs W1</b>	Non-outliers	0.011	0.010	1.608	1.206	6.04e-3	1.67e-2
	Outliers	0.001	0.001	0.368	0.284	1.59e-2	1.58e-2
<b>E1 vs W2</b>	Non-outliers	0.011	0.011	1.605	1.097	5.12e-3	1.59e-2
	Outliers	0.002	0.002	0.593	0.370	1.79e-2	2.01e-2
<b>E1 vs W3</b>	Non-outliers	0.011	0.013	1.608	1.514	4.00e-3	1.56e-2
	Outliers	0.002	0.002	0.517	0.178	1.40e-2	1.59e-2
<b>W1 vs W2</b>	Non-outliers	0.010	0.011	1.207	1.101	1.69e-3	1.18e-2
	Outliers	0.004	0.005	0.575	0.482	1.04e-2	1.50e-2
<b>W1 vs W3</b>	Non-outliers	0.010	0.013	1.203	1.512	4.27e-3	1.54e-2
	Outliers	0.002	0.003	0.538	0.375	1.60e-2	1.86e-2
<b>W2 vs W3</b>	Non-outliers	0.010	0.012	1.105	1.518	3.56e-3	1.48e-2
	Outliers	0.003	0.004	0.296	0.315	1.63e-2	1.99e-2

**Blue values** = significantly higher values in outliers/non-outliers' comparison ; **A** = first lineage in the pair (e.g. E1 vs W1 : E1 is A) ; **B** = second lineage in the pair (e.g. E1 vs W1 : W1 is B) ; **Recombi** = number of recombination events identified with the four-gamete test



**Fig.S1 PCA of the goodness of fit of the models for each pairwise comparison.** Green points: prior distribution – blue points: posterior distribution – pink points: observed dataset. For a model to correctly fit with the data, the black point must be more or less in the centre of the pink and blue ones. Eigen values for each principal component are given.



**Suppl. Fig.S2 Goodness-of-fit test for demographic inferences made under a four-population model.** This test is performed with a PCA of the calculated summary statistics. Only the best-supported of the 64 proposed models, the one without any migration, is shown here. In grey: 10,000 points under the best model with random combinations of parameters. In purple: 10,000 points under the best model with parameters estimated by Random Forest. In blue: 10,000 points under the best model with cyclic optimisation. In yellow: the observed point.

**Suppl. Table S6**

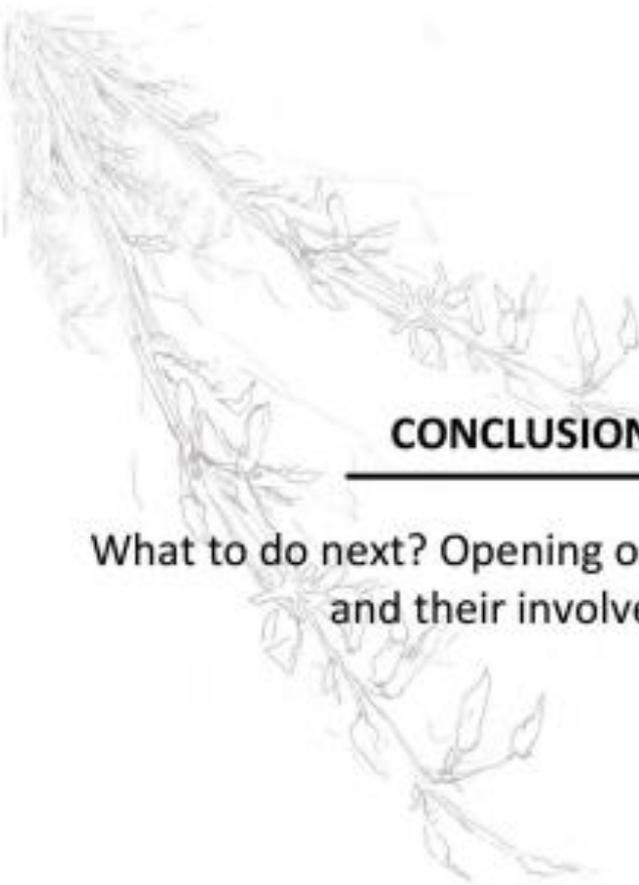
Parameters estimations calculated using DILS with four populations. Time of split in years was calculated considering that 1 generation equivalent 3 years.

<b>EFFECTIVE SIZE</b>	<b>Lineages</b>	<b>Estimations</b> (in number of effective individuals)		<b>IC 95</b>
	Ancestral population	176 843		[210 614 – 488 883]
	Western ancestor	549 374		[40 075 – 980 796]
	W1 & W2 ancestor	115 357		[13 245 – 921 587]
	E1	181 683		[60 687 – 704 350]
	W1	255 584		[74 469 – 904 304]
	W2	532 627		[147 897 – 964 900]
	W3	578 331		[200 036 – 957 062]
<b>T-SPLIT</b>	<b>Lineages split</b>	<b>Estimations + [IC 95]</b>		
		In generations		In years
	E1 - (W3,(W1,W2))	235 242 [94 655 – 454 581]		705 726 [283 695 – 1 363 743]
	W3 - (W1,W2)	225 001 [67 364 – 401 350]		675 003 [202 092 – 1 204 050]
	W1 - W2	95 969 [27 416 – 260 954]		287 907 [82 248 – 782 862]

For the Tsplitted, the notations represent the estimated time of split between lineages : e.g. W1 – W2 = time of split between lineages W1 and W2.







## **CONCLUSION & DISCUSSION**

What to do next? Opening on organellar genomes evolution and their involvement in speciation





## I. Summary of the results

*Silene nutans* is composed of four genetically differentiated lineages at least in France, associated with past glacial refugia : E1, W1, W2 and W3 (Martin *et al.*, 2016; Van Rossum *et al.*, 2018). Between lineages E1 and W1, strong and asymmetric reproductive isolation (RI) had been identified (Martin *et al.*, 2017). This pattern was extended to all lineages in reciprocal crosses, resulting in a high proportion of hybrid seedling mortality due to chlorosis, depending on the cross direction (Van Rossum *et al.*, in prep). Cytonuclear incompatibilities (CNIs) were suspected to play a role in this RI (Martin *et al.*, 2017). Because hybrids germinated at a normal rate and chlorosis was observed at an early stage of development, plastid-nuclear incompatibilities (PNIs) were most likely involved rather than mito-nuclear incompatibilities (Etienne Meyer, personal communication). In the **first chapter** of this PhD thesis, we identified potential plastid-nuclear gene pairs encoding proteins of the plastid ribosome and cytochrome b6/f for which co-evolution pattern was detected within lineages and that could result in PNIs in hybrids (Postel *et al.*, 2022). We also identified a peculiar evolutionary dynamic of the plastid genes. Accelerated rate of evolution was identified on these genes, with high number of non-synonymous substitutions fixed within each lineage, elevated  $d_N/d_S$  compared to other *Silene* and angiosperm species as well as evidences for relaxed selection on most of the plastid genes. Repeated bottlenecks during post glacial recolonization and genetic drift might have driven this evolutionary pattern in the plastid genome.

Some of the hybrids from the reciprocal crosses conducted by Fabienne Van Rossum survived and exhibited variegated and fully green phenotypes (Van Rossum *et al.*, in prep). Such phenotypes can be the result of paternal leakage (i.e. occasional paternal transmission and expression of the paternal organellar genomes of the plastid genome (Ramsey & Mandel, 2019; Greiner *et al.*, 2014). This mechanism of organellar genome transmission could sometimes rescue inter-lineages hybrids suffering from PNIs (Barnard-Kubow *et al.*, 2017). In the **second chapter** of this PhD thesis, we tested for paternal leakage of the plastid genome in *S. nutans* by genotyping the surviving hybrids. Occasional paternal transmission of the plastid genome was identified in some surviving hybrids and rescued them when the paternal genome was more compatible with the hybrid nuclear background. These results confirmed the involvement of the plastid in RI.

The mitochondrial genome could also be involved in the speciation process through mito-nuclear incompatibilities in hybrids (similarly as the plastid genome). Besides, *S. nutans* is also a gynodioecious species which might induce a peculiar evolutionary dynamics of the mitochondrial genome as it is a cytonuclear based reproductive system (Garraud *et al.*, 2011). In the **third chapter** of this PhD thesis, we analyzed the evolutionary pattern of the mitochondrial genes in the four lineages of *S. nutans* and compared it with the one observed in the plastid genes for the same individuals. Mitochondrial and

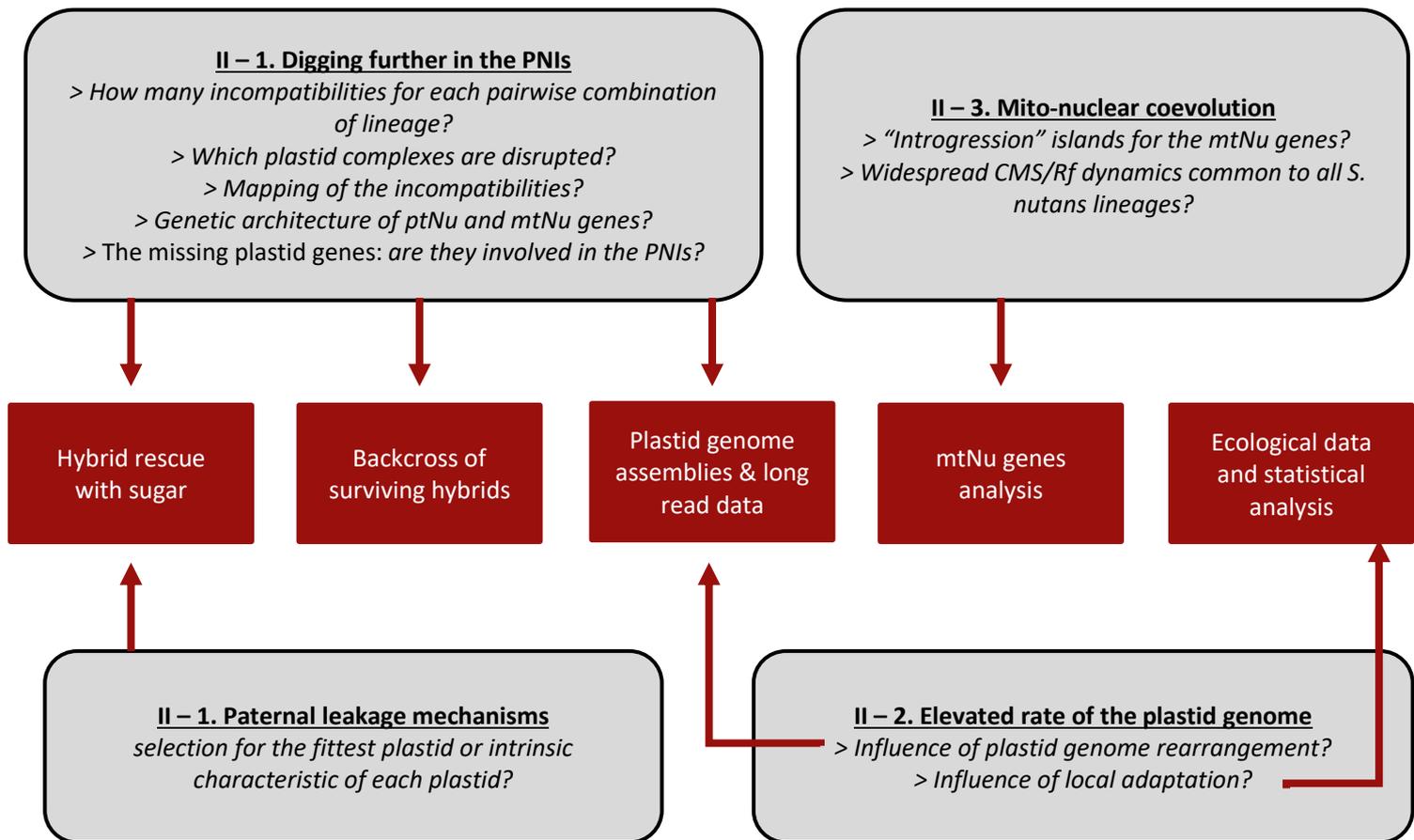
plastid genomes in *S. nutans* did not seem to be in complete linkage disequilibrium and exhibited contrasting evolutionary patterns: fixed genetic differences for all lineages in the plastid genes and shared polymorphism in the mitochondrial ones. This contrasting pattern suggests that the mitochondrial genome might not be involved in the RI observed between lineages, at least not through mito-nuclear incompatibilities. Presence of shared polymorphism could also confirm maintenance of gynodioecy and ancestral polymorphism through balancing selection as already pointed out in (Lahiani *et al.*, 2013). Besides, the identification of recombination events within and between mitochondrial genes as well as intra-individual polymorphic sites, might suggest the presence of paternal leakage of the mitochondrial genome. Gynodioecy might have played a role here, as it is thought to favor paternal leakage of the mitochondrial genome (McCauley, 2013; Breton & Stewart, 2015; Ramsey & Mandel, 2019).

Finally, among these four lineages, no hybridization events had been detected between lineages E1 and W1 in secondary contact zones in southern England, southern Belgium, north-eastern France (Martin *et al.*, 2016). Allopatric speciation was also identified between lineages E1 and all of the western lineages, with no gene flow since their split around 300 000 years ago (Martin, 2016). In the last and **fourth chapter** of this PhD thesis, we inferred the complete evo-demographic histories of the four lineages of *S. nutans*, by increasing the sampling effort for western lineages compared to the previous study. Allopatric speciation without gene flow was identified for all pairs of lineages. The split time was estimated around 700 000 years ago between lineages E1 and the western ones ; 380 000 years between lineages W3 and W1/W2 and 300 000 years ago between lineages W1 and W2, consistent with Glacial maxima.

To sum up, with the present work, we now know that all four lineages of *S. nutans* have diverged in allopatry (chapter 4). We also have a clearer idea of what might be responsible for strong post-zygotic RI between the four lineages of *S. nutans* (chapter 1), PNIs, likely resulting for accumulation of genetic divergence in allopatry. We also know that *S. nutans* have the potential for paternal transmission of its organellar genomes, which can rescue inter-lineages hybrids (chapter 2) and that the organellar genomes in this species exhibit different evolutionary histories (chapter 3).

## I. Perspectives

Regarding the above results, I propose future work and direction to further understand evolutionary dynamics of the organellar genomes in this species and how they impact the strong yet incomplete RI observed between the four lineages of *S. nutans*. Summary of the perspectives and methodologies can be found in Figure 1.



**Fig.1** Summary of the question raised following my PhD work and of the methodologies that could be used to study them. Number of each section is indicated.

## 1. Digging further into the PNIs

### 1.1. Outstanding issues with the PNIs in *S. nutans*

#### 1.1.1. The missing plastid genes

In the first chapter we identified potential candidates for the PNIs responsible for the strong RI between lineages of *S. nutans*. Regarding the first results on the 3D structure of the plastid ribosome, genes of this complexes are unlikely to be the only ones responsible. Cytochrome b6/f might represent a strong candidate. Yet, we did miss some of the plastid genes that are equally important for plastid function : *clpP1*, *accD*, *ycf1* and *ycf2*, among which two of them are the only plastid-encoded genes in their complexes : *accD* and *clpP1* (Rockenbach *et al.*, 2016). *accD* is part of an essential plastid cellular pathway – the acetyl-CoA carboxylase involved in lipid biosynthesis (Sobanski *et al.*, 2019). It generally exhibits elevated rate of evolution in plant species with high plastome evolutionary rate (Barnard-Kubow *et al.*, 2014; Sloan *et al.*, 2014a; Rockenbach *et al.*, 2016; Weng *et al.*, 2016) and it is known for frequently being transferred to the nucleus in other lineages (Magee *et al.*, 2010; Sloan *et al.*, 2012b;

Rockenbach *et al.*, 2016). In all lineages, one blast hit was identified. When aligned with the two reference sequences of *S. latifolia* and *S. paradoxa*, the sequences of lineages E1 and W1 contained early stop codon, suggesting pseudogenization while no early stop codons were identified in lineages W2 and W3. We might then argue that (1) lineages W2 and W3 contain a complete functional sequence of *accD* and (2) lineages E1 and W1 contained an *accD* sequence that does seem to be functional anymore. Pseudogenization of *accD* in lineages E1/W1 might be the result of its transfer to the nucleus for these lineages. If so, crossing between W2 or W3 individuals (which have kept a functional sequence of *accD*), and individuals from E1 or W1 (where *accD* might be nuclearly encoded and absent from the plastid genome), could generate incompatibility in hybrids: hybrids that have inherited the plastid genome of lineages E1/W1 would lack the plastid copy of *accD* and could also lack the nuclear copy of *accD* depending on allelic segregation during meiosis. Regarding *clpP1*, this gene is often not found in genomic data due to its high evolutionary rate and divergence between species (Magee *et al.*, 2010; Sloan *et al.*, 2012b; Rockenbach *et al.*, 2016; Williams *et al.*, 2019). It can either be composed of one or several exons depending on the species (e.g. one exon in *S. latifolia* and two in *S. paradoxa*) (Williams *et al.*, 2019). Using tblastn (Altschup *et al.*, 1990; Gertz *et al.*, 2006) and *S. latifolia* sequence used in chapter 1, we found *clpP1* in individuals of lineage W3. As *clpP1* evolved fast, we then used the sequence of W3 to find it in the rest of the lineages. Complete sequence of *clpP1* was only identified in lineage W3 ; partial sequences were found in lineages W1 & W2 and no sequences at all were found in lineage E1. We also searched for this gene in the transcriptomic data used in chapter 1 and 4. The reverse was found with complete sequence of *clpP1* in E1 but nothing for the western lineages. When the sequence of E1 found in the transcriptomic data was blasted against the E1 plastid genome assemblies (using tblastn again), no hit was identified either. In both cases, the *clpP1* blast resulted in only one hit, covering the whole length of *clpP1 S. latifolia* ortholog suggesting that *clpP1* might be composed of only one exon in *S. nutans*. The absence of sequence of *clpP1* in the lineage E1 might suggest either (1) loss of the gene after transfer to the nucleus or (2) that the sequence was too divergent to be found in lineage E1 using gene capture. The later might be the most likely giving that (1) no previous studies on this gene reported a transfer from *clpP1* to the nucleus (even though losses of exons have been reported (Williams *et al.*, 2019)) and that (2) plastid genome of lineage E1 is the most divergent one so baits to get this gene based on *S. latifolia* reference genome might be too divergent from the E1 sequence to effectively capture it. This would make sense as even using E1 transcriptomic sequence or W3 sequence (more closely related to E1 than *S. latifolia*), we did not find it in the E1 plastid genome assemblies. Yet, we observed similar levels sequence similarities of *clpP1* between E1 or W3 and *S. latifolia* (around 70% for E1 and 77% for W3) suggesting that level of divergence with *S. latifolia* baits might not be the only reason for *clpP1* absence in E1 plastid genome assemblies. So, the question of whether E1 *clpP1* plastid gene was not found because of bait

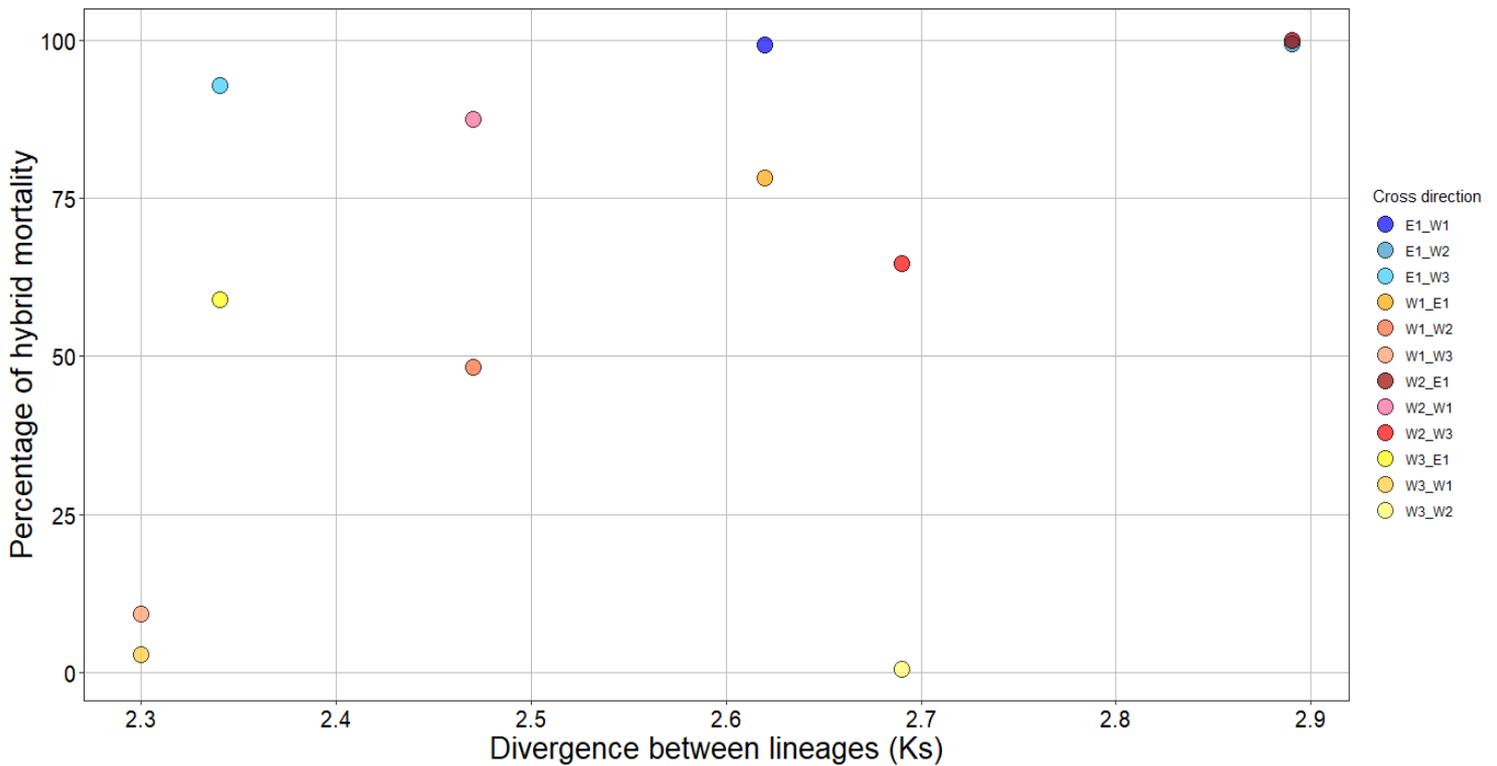
divergence or because it has been transferred to the nucleus remains open. Again, *clpP1* is part of the core complex CLP and is an essential gene for cell functioning (Zoschke & Bock, 2018). If the transfer hypothesis is true, then hybrid resulting from crosses between individuals without *clpP1* plastid copy and individuals that kept it might suffer from strong PNIs similarly to *accD* above. Two other genes were partially found or not found at all: *ycf1* and *ycf2*. These two genes also exhibit elevated rate of evolution (Sloan *et al.*, 2014a). Their cellular function and importance are not well understood but it could be worth trying to find them in the genomic and/or transcriptomic data.

#### 1.1.2. Strong, numerous, minor incompatibilities?

For chapter 1, we worked a bit at the lineage level but we have no idea, for each pair of lineages, whether major or minor effect incompatibilities are acting, and if strong RI is the result of minor incompatibilities accumulation or a few strong ones. Given the results of the diallelic cross, we could guess that few major effect incompatibilities are acting between lineages W1 and W2 as their level of RI is relatively high while they are only slightly divergent (Figure 2). Conversely, between lineages E1 and the rest of the western lineages, given the higher level of divergence, we would expect strong RI as a result of accumulation of minor effect incompatibilities (Figure 2). Following up on this, we could also wonder whether or not the PNIs responsible for hybrid breakdown are located in the same plastid complexes or not regarding the analyzed lineage pairs, which could give a clue about evolutionary convergence and preferential complexes responsible for PNIs and RI.

#### 1.1.3. Dominance relationship at nuclear loci

We also have no knowledge about the genetic architecture and nuclear hierarchy dominance at nuO loci (i.e. nuclear loci involved in cytonuclear interactions, i.e. nuclear genes whose gene products are targeted to the organelles) between lineages of *S. nutans*. Theoretically, nuclear background of inter-lineages hybrids is composed of one set of maternal alleles and one set of paternal ones. If no biparental inheritance of organelles is occurring, they will inherit the organellar genomes of their mother, co-adapted with the maternal nuclear alleles (Sloan *et al.*, 2018). NuO loci would have ½ of alleles co-adapted with the organelles and the other ½ less fitted. If the hybrid is unfit, this would mean that some of the mis-adapted alleles (i.e. the paternal ones) are expressed at almost all nuO or at nuO involved in essential organellar pathways. If the hybrid is fit, then either the maternal alleles are expressed at almost all nuO and/or paternal alleles are expressed at nuO not involved in essential organellar pathways. Paternal leakage of the organellar genome would further complexify the dynamic: if the hybrid is fit, additionally to the outcome listed above, it could also mean paternal leakage of the organellar genome along with paternal expression at one or more nuO or at essential organellar pathway. Going further very speculatively, dominance hierarchy at nuO loci might also



**Fig.2 Percentage of hybrid seedling mortality as a function of the level of plastid divergence (Ks) between lineages.** One point represents one cross direction. The level of mortality is not correlated with the level of divergence between lineages. For example, cross between lineage E1 and W3 resulted in high level of hybrid mortality while lineages are not as divergent as lineages W3 and W2 for example which resulted in almost no hybrid mortality.

influence the outcome of paternal leakage, if it exists: for example, paternal alleles dominance at nuO loci would impose selective pressure to keep the paternal organellar genome instead of the maternal one, because it would be the fittest with the hybrid nuclear background at nuO loci. Little is known about that but this would help to understand the speciation dynamic associated with the CNIs: in which case hybridization would result in hybrid breakdown and in which case it wouldn't because of paternal leakage? Is the presence of "fit" hybrids purely random (i.e. only depending on the already established dominance relationship between parental nuclear alleles) or is there some selective processes either driven by the nucleus or the organelles, to keep the fit combination and maintain CN co-adaptation? Are these dominance relationships conserved within species or labile?

Nuclear dominance architecture might be observed between lineages of *S. nutans*. W3 lineage is relatively divergent from the other three lineages at plastid and ptNu genes (nuclear genes with gene products targeted to the plastid). When looking at randomly selected nuclear loci in chapter 4, level of genetic divergence seems to be only slightly different between all four lineages, with W3 being equally genetically divergent from lineages E1, W1 and W2 and W1/W2 the least divergent from one another. Yet W3 seems to work better in diallele crosses with lower hybrid mortality and chlorotic seedlings, either used as the mother or as the father. It works even better than when W1 and W2 are crossed

together. When looking at results from chapter 2, hybrids resulting from crosses involving W3 often survived because they inherited W3 plastid genome (either through maternal or paternal inheritance). In conclusion, W3 seem then to be more compatible with the other lineages than any other one, yet being relatively highly genetically divergent from them. This would mean that either (i) W3 accumulated mutations that do not disrupt essential organellar pathways or (ii) that the nuclear alleles of W3 are dominant at one or more loci, with expression of its alleles rather than the alleles of the other lineages in inter-lineages hybrids, maintaining plastid-nuclear co-adaptation and (iii) that its plastid genome is “favored” in inter-lineages hybrids when paternal leakage can occur.

#### 1.1.4. Paternal leakage of the plastid genome: selection or not? That is the question.

Following up on paternal leakage, the above statements relate to the more general question of distinguishing between maintenance of the paternal plastid genome in hybrids through selection for the most compatible plastid or because of intrinsic characteristic of the plastid genome. Theoretically, several nonexclusive factors could influence the outcome of paternal leakage in inter-lineages hybrids. (i) When the elimination of the other plastid type occurs: if the elimination occurs directly in the male germinative cell, then potential for paternal leakage is near zero. (ii) The characteristics of the two plastid types themselves: do they have similar competitive ability or not? (iii) the size of the two gametes: if the male gamete is smaller than the female one, then the plastid genome in the sperm cells will be lost through drift during vegetative sorting out. (iv) The nuclear dominance architecture between the lineages at ptNu loci: how many of the maternal and paternal alleles are dominant at ptNu loci? (v) If selection for the most compatible plastid is acting, the functional impact of the incompatibility: if it disrupts an essential function, the strength of selection to recover it might be strong while if it impacts “unessential” pathways, it might not be strong enough to favor the paternal plastid genome. (vi) The number of reproductive barriers already present between hybridizing lineages: if other reproductive barriers are acting, then restoring plastid-nuclear co-adaptation might not be sufficient to save the hybrid.

With *Silene nutans*, most of the hybrids resulting from crosses with lineage E1 died any way, suggesting that, regardless of the presence of paternal leakage of a less incompatible plastid genome, genetic divergence and reproductive barriers might be too strong to allow for hybrid rescue and observation of paternal leakage. Also, presence of paternal plastid genomes in hybrids might not be possible in the first place if paternal plastid genome was not kept in the male germinative cells, suggesting constitutive pollen transmission of the plastid genome in this species. *S. nutans* is thought to be anisogamous so we could have expected loss of the paternal plastid genome through drift. Yet, we reported some cases of heteroplasmic individuals (i.e. the variegated ones), with segregation of maternal and paternal plastid genomes in different sectors of the plant, and homoplasmic individuals for the paternal plastid

genome (chapter 2). Variegation might be the result of (i) vegetative sorting of the two plastid types after fertilization and/or (ii) it could also represent equal competitive ability of the two plastid types, resulting in the maintenance of the two plastid types. Almost all variegated hybrids resulted from crosses with lineage W2 either as the mother or the father. This might suggest that its plastid genome have good competitive ability compared to the other western lineage's plastid genome. If selection was acting alone, it would have led to its elimination as its plastid genome is highly incompatible with the hybrid nuclear background. Yet, in crosses with W2 we also observed homoplasmic individuals for the other plastid genome, suggesting that intrinsic characteristic of the plastid itself might not be enough for paternal transmission. So, selection for and/or selective degradation of one of the two plastid types must have played a role somehow, favoring the plastid genome which is the less incompatible with the nuclear genome, which in some case would be the paternal plastid genome. Finally, as mentioned above, hybrids produced by W3 survived because they inherited its plastid genome, through paternal leakage or not, while it is quite genetically divergent from the other lineages. Here, the main factors influencing the outcome of paternal leakage seems to be a balance between intrinsic capacity of the plastid itself, strength of selection for the fittest plastid and the genetic architecture of the PNIs. The strength of selection might depend on the nature of the incompatibility (i.e. a few impacting essential organellar function or numerous incompatibilities acting on multiple complexes) and the nuclear dominance hierarchy at ptNu loci.

## *1.2. Methodologies perspectives to answer the above questions*

### *1.2.1. Which hybrids to work with?*

In the following sections, I propose methodologies that rely on analyzing the inter-lineages hybrids so the first thing to do is to choose which hybrids to work with. As explained before, the diallele cross with the four lineages of *S. nutans* resulted in various hybrid phenotypes and levels of hybrid seedling mortality depending on the direction of the cross. These variability in terms of hybrids survival and fitness could be very useful. I would suggest to analyze hybrids from all pairs, when possible. Analyzing hybrid resulting from crosses with lineage E1 could help us to identify strongly lethal PNIs and/or numerous ones, giving the level of mortality in crosses with E1 and the level of genetic divergence of E1 with the western lineages (Figure 2). Analyses of the hybrid with W2 as one of the parents could help further understand the dynamic of paternal leakage. Hybrids resulting from crosses with lineage W3 could also be very helpful to understand the dynamics of nuclear allele expression on the cytonuclear co-adaptation, as even though this lineage exhibit high level of genetic diversity, percentage of hybrid mortality are the lowest when it is involved (Figure 2). Finally crosses with W1 are also relatively lethal, also with W2 while levels of genetic divergence between the two is low (Figure

2). So, we can expect few strong incompatibilities playing a role here and analyses of the hybrids from this cross might allow us to identify them. More generally, working with hybrids from all pairs could also help us tackle the question of convergent evolution of PNIs as we could have an idea of the plastid complexes disrupted within each pair of lineages.

#### 1.2.2. Sugar hybrids rescuing

The first thing that we could do next is to rescue the hybrids of the reciprocal crosses between the four lineages. To save the hybrids, we would have to sow hybrid seeds on laboratory media supplemented with sugar (Kühn *et al.*, 2015). Because of the presence of sugar, hybrids could easily pass the step from heterotrophy to autotrophy, which seems to be the critical step in their development giving the observed hybrids phenotypes (i.e. white/yellowish which suggest default in photosynthetic complex) (Yao & Cohen, 2000; Massouh *et al.*, 2016; Liebers *et al.*, 2017). Thanks to that we could do several things. (1) Rescuing hybrids could first allow us to confirm that it is the photosynthetic pathways which is not working, rather the plastid gene expression machinery. (2) We could genotype these surviving hybrids and further identify which plastid genome they inherited: does the hybrids that would have died without sugar effectively inherited the maternal (incompatible?) plastid genome or the paternal one? (3) We could also assess which of the maternal and paternal nuclear alleles are expressed in these hybrids through analysis of hybrid nuclear gene expression and RNAseq sequencing of the whole nuclear set of genes targeted to the plastid. Having already the RNAseq data of pure individual, we could also compare expression ratio between hybrids and parents to assess whether disruption of co-adaptation between plastid and nuclear genes also has an impact or is impacted by altered gene expression. (4) We could further identify what plastid-nuclear gene pairs might generate PNIs. Using proteomic methods, we could assess which of the hybrid's proteins are effectively produced and which ones are functional or not (supposing that the dysfunctional one is the result of disruption of co-adaptation between plastid and nuclear genes). This proteomic approach could be enriched and complemented by bioinformatic modeling when 3D structures of complexes are known.

#### 1.2.3. Genetic mapping on backcrosses

Another way to map the incompatibility and the traits disrupted in hybrids would be to backcross the surviving hybrids with one of their parents. Backcrossing leads to increase genomic content of one of the two parents. Because of segregation, the phenotypes observed will be due to some segregating alleles that can be identify through quantitative trait loci (QTL) mapping. Yet, to find the nuclear loci involved in the incompatibility, we would need to generate many backcrossed individuals in order to have statistical power for the QTL to find them.

If hybrids rescued on sugar are fertile, we could work with all hybrids from all cross direction. If not, as not hybrids from all cross directions survived (e.g. when E1 is the mother, only some of the hybrids with W3 as the father survived and we do not have any surviving hybrids for the cross W2xE1 neither) we would have to work only with the “naturally” surviving hybrids. Among these surviving hybrids, most them produced viable pollens that could be used to do the backcross (data not shown). With them, we could backcross the hybrids that have inherited the paternal plastid genome as the paternal parent with their father. In this case, we would obtain hybrids with paternal plastid genomes with higher fraction of the nuclear genome being also paternal, so decreasing the proportion of the nuclear genome that would be maternal and mis-adapted, making it easier to identify the PNIs. Theoretically, we would need at least 200 individuals per backcross to have enough statistical power.

#### 1.2.4. Long-read sequencing and plastid genome assemblies

Finally, to find the missing genes in the nuclear or plastid genomes of *S. nutans* lineages, we could analyze the already available plastid genome assemblies of the four lineages. Supposing that gene order is conserved in the plastid genome of *S. nutans*, we could see whether at the expected position of these genes there are gaps or not in the assembly : if there are gaps, it would suggest that these genes are still in the plastid genomes but that we did not get them with our method ; if there is no gap, it might suggest that they were transferred to the nucleus. Quality of the available assemblies have not been investigated yet, so this would be the first thing to do. If quality is not good enough, we could re-sequence individuals following long-read data methodology. It could be a good option as DNA fragments to reconstruct are longer which facilitates the reconstruction of very divergent gene sequences and it could also improve assembly's quality. Long-read could also be a good option to get the highly divergent plastid genes. If we get them, we could reconstruct their sequences and assess whether or not the sequences in the plastid are still functional, conducting molecular data analysis.

## 2. The elevated rate of plastid genome evolution in *S. nutans*

Working on the plastid assemblies and long read data might also be useful for another question. Given the elevated rate of plastid genome evolution in all lineages of *S. nutans*, one might want to dig further to understand how such an acceleration can occur in the first place.

Giving the results in chapter 1, this acceleration might be the result of relaxed selective pressure on most of the plastid genes and some positive selection on the genes encoding the plastid ribosome. Demographic processes are supposed to influence the evolutionary dynamics of genomes in general (Olson & McCauley, 2000; Féart *et al.*, 2006; Rockenbach *et al.*, 2016). Reduction in population effective size ( $N_e$ ) would lower the impact of selection and increase the influence of genetic drift

(Rockenbach *et al.*, 2016; Havird *et al.*, 2017). *S. nutans* is thought to have experienced repeated bottlenecks, which are known to reduce population  $N_e$ , through post-glacial recolonization and results from chapter 4 indicated variable  $N_e$  during evolutionary history of the four lineages. These bottlenecks might have facilitated fixation of slightly deleterious mutations in *S. nutans* plastid genes and led to the observed pattern of increased genetic diversity. Similar evolutionary history and accelerated rate of evolution was observed in *Campanulastrum americanum* (Barnard-Kubow *et al.*, 2014, 2016). So, it is very likely that demographic processes leading to reduced  $N_e$  influence the evolutionary rates of plastid genome. But other angiosperm species exhibited similar, yet lower, increased in  $d_N/d_S$  without any indication of involvement of demographic processes (Sloan *et al.*, 2014b; Havird *et al.*, 2015). Although impacts of demographic processes in evolutionary rate of organellar genomes might be underestimated or understudied, these alone may not explain why some species exhibit evolutionary rate acceleration. In our case, they may have only accentuated it. So, what other processes could have led to relaxed selective constraints and accumulation of non-synonymous mutations in *S. nutans* plastid genome?

Initially, we thought that gynodioecy might be involved. Organellar genomes are supposed to be in strong linkage disequilibrium (Olson & Mccauley, 2000), so one could imagine “linked selection” between these two. This reproductive system is thought to be maintained either through selective sweep or via balancing selection (Ingvarsson & Taylor, 2002; Touzet & Delph, 2009). If selective sweep occurs on the mitochondrial genome, it could drag the plastid genome with it. Depending on how frequent these sweeps are, it could lead to the fixation of substantial number of plastid mutations. But regarding the results of chapter 3, balancing selection seems to be the main driver of gynodioecy maintenance in *S. nutans* and organellar genome are not in strong linkage-disequilibrium. So gynodioecy is unlikely involved in driving an accelerated rate of plastid genome evolution here.

General plastome instability could have led to this acceleration (Jansen *et al.*, 2007; Park *et al.*, 2017). Because of this instability, genomic rearrangement can occur and lead to relaxed constraint on the plastid genome (Jansen & Ruhlman, 2012). Genomic rearrangement can shuffle plastid gene order, which is normally strongly conserved (Sloan *et al.*, 2014a; Williams *et al.*, 2019) and lead to expansion/contraction of the inverted repeat present in the plastid genome (Shrestha *et al.*, 2019)). With the plastid genome assemblies available or the new ones constructed with long-read data, we could see if (i) the boundaries of the inverted repeats are similar to angiosperm species with regular rates of plastid genome evolution or not, as observed in other *Silene* species with accelerated rate of plastome evolution (Sloan *et al.*, 2014a) and (ii) also look at the gene order and see whether inversion/shuffling occurred, as witness for genomic rearrangement.

Local adaptation could also have played at least a partial role in the accelerated rate of plastome evolution in *S. nutans*, through fixation of adaptive plastid mutations. Plastid genome might have an

underestimated role in local adaptation (Levin, 2003; Budar & Roux, 2011). Lineages E1 and W1 represent edaphic ecotypes in Belgium with lineages E1 being adapted to calcicolous soil and lineage W1 to silicicolous ones (De Bilde, 1973; Van Rossum *et al.*, 1996). Local adaptation to different soil conditions arose after their divergence in allopatry. This differential local adaptation in Belgium could raise the question of the involvement of local adaptation in the observed genetic diversity of the plastid genome. Although this edaphic adaptation is only observed in lineages E1 and W1, one could wonder whether lineages are pre-adapted to specific ecological conditions or not. This potential pre-adaptation could have led to the fixation of some of the observed mutations in the plastid genes. Little is known about adaptation to ecological conditions in lineages W2 and W3, but lineage W3 is present at higher altitude than the other three, which could also lead to some degree of local adaptation. Regarding lineage W2, its post-glacial recolonization either stops in south of France or is not completed yet. If the former is true, then why W2 did not recolonize further in the north of France as lineages W1? W2 could also have become locally adapted either to soils conditions or other ecological constraint during migration to glacial refugia and/or subsequent recolonization, restricting its geographical range. It would be interesting to test whether lineages W2 and W3 are also locally adapted to different soil conditions. More generally, it would be interesting to assess (i) if the four lineages of *S. nutans* are locally adapted to different ecological conditions, potentially induced by the ecological characteristic of their distinct glacial refugia and (ii) what are these ecological conditions: do they all refer to soils characteristics or are other ecological factors involved? For the first question, we could collect data on climatic and environmental conditions in the ranges of the four lineages and see whether these environmental data differ from one lineage to the other. We could use PCA analysis or statistical ones to try to see whether environmental do have an effect on the distribution of the four lineages of *S. nutans*. If distinct environmental conditions are identified, common garden experiment with reciprocal transplant could help disentangle between plastic response to environmental variables and adaptively fixed genetic differences (Galloway & Fenster, 2001; Sambatti *et al.*, 2008; Sobel *et al.*, 2010).

### **3. What about the mito-nuclear coevolution in *Silene nutans*?**

In the current work we did not analyze at all the nuclear genes whose gene products are targeted to the mitochondria (mtNu). The mitochondria contain essential, if not vital, cellular function as it deals with cellular respiration and it produces the cellular energy needed for the cell to correctly function (Hill, 2016). Mitochondrial complexes are also encoded both by mitochondrial and nuclear genes, leading to strong co-evolution between mitochondrial and nuclear interacting genes (Sloan *et al.*, 2017). We could then wonder what evolutionary pattern would be observed in the nuclear genes

interacting with the mitochondrial ones: would we find shared polymorphism too? We could imagine that because of the evolutionary dynamics observed in the mitochondrial genome, that “kept” ancestral variation, the mtNu genes would also have kept this ancestral variation, representing “introgression islands” / “ancestral islands of introgression” where differentiation between lineages is close to 0, conversely to the “speciation islands” that represent highly differentiated region of the genome (Turner *et al.*, 2005; Feder *et al.*, 2012). Near absence of specific selective pressures were identified on the mitochondrial genes in chapter 3 suggesting either neutral evolution or relaxed selective pressure on these genes, except on *cox3*. We could then expect to identify at least some signature for positive selection on the interacting mtNu of complex IV, as a witness for their need to co-adapt with mitochondrial genes (Rand *et al.*, 2004; Sloan *et al.*, 2018). So, it would be interesting to study the evolutionary dynamics of these nuclear genes, and how contrasting the evolutionary pattern is with the ptNu genes and the rest of the nuclear genome. We could just follow the exact same strategy as the one followed in chapter 1: (i) extract the nuclear-encoded mitochondrial-targeted genes in the transcriptomic data that already are available, (ii) conduct molecular analysis to assess co-evolution between mitochondrial and nuclear interacting genes.

Besides, this species is gynodioecious. Not all populations contain female individuals (Dufaÿ, unpublished data), which suggests the fixation of one or more nuclear loci that restore the male reproductive function or that not all populations contain a CMS factor. Usually, nuclear restorer loci (*Rf*) are co-adapted with the CMS they restore, leading to lineage-specific co-adaptation between *Rf* and CMS factors (Hanson & Bentolila, 2004; McCauley & Bailey, 2009). Because of that, cryptic CMS can be revealed in inter-lineages hybrids as hybridization would bring together CMS from one lineage with the *Rf* of another one or with a hybrid nuclear background lacking the co-adapted *Rf* (e.g. in *Mimulus* (Case & Willis, 2008; Case *et al.*, 2016)). In our inter-lineage hybrids, some individuals were females (data not shown), suggesting that CMS are expressed in hybrids. This could mean that (i) not all lineages have the *Rf* in their nuclear genomes, which would lead to expression of CMS factors in hybrids or (ii) that lineage-specific CMS/*Rf* pairs are present in each lineage, with co-adaptation between them being disrupted in hybrids. The presence of shared polymorphism in *S. nutans* mitochondrial genes indicate that mitochondrial polymorphism, the genetic basis of CMS emergence, is shared between lineages. If mitochondrial polymorphism is common to lineages, then the *Rf*, if present, must also be common to all lineages. So, the second hypothesis should be the most likely. Nevertheless, we cannot exclude the possibility of the co-occurrence of old CMS shared between lineages and new lineage-specific CMS as observed in *Lobelia siphilitica* (Adhikari *et al.*, 2019).

Mitochondria is not involved in the RI through mito-nuclear incompatibilities because of gynodioecy and the subsequent dynamic of mitochondrial genome evolution, highlighting the importance of the reproductive system in the speciation process.

## II. Evolutionary processes of the organellar genomes

During this PhD, I had the chance to discover genome evolution in plants and especially the organellar ones. These two genomes are pretty intriguing in many ways: how did they evolve in the first place? Why in some species/lineages, do they exhibit accelerated rate of genome evolution, while they are supposed to be strongly conserved and under strong functional constraints as they are too important for the cells and the individual (Jansen & Ruhlman, 2012; Havird *et al.*, 2015)? Organellar genome evolution and their mode of transmission could greatly influence speciation. As demonstrated during this PhD, organellar genome fast evolution might influence the emergence of CNIs as a reproductive barrier between lineages and biparental inheritance might rescue some of the hybrids suffering from these incompatibilities, potentially slowing down the speciation process. In the following sections, I tried to sum up the mechanisms and open questions regarding organellar genome evolution and their involvement in speciation. Summary of these questions can be found in box 1.

### 1. Inheritance pattern of organellar genomes

#### 1.1. Uniparental inheritance = is it still the rule?

Around 20% of angiosperm species should have the potential for biparental transmission (Zhang *et al.*, 2003; Azhagiri & Maliga, 2007; Nagata, 2010; McCauley, 2013; Breton & Stewart, 2015) and it is described in nearly all clades of green plant species (Ramsey & Mandel, 2019). In addition to *S. nutans*, paternal leakage was also observed in *C. americanum* (Barnard-Kubow *et al.*, 2017), some pea accessions (Bogdanova & Kosterin, 2006), Geraniaceae (Ruhlman & Jansen, 2018) and *Passiflora* species (Shrestha *et al.*, 2021), *Arabidopsis thaliana* (Azhagiri & Maliga, 2007), *Silene vulgaris* (McCauley *et al.*, 2005, 2007; Pearl *et al.*, 2008) and others. In Dipsacales, Hu *et al.* (2008) observed active proliferation of plastids in male generative cells with active degradation at the same time, during pollen development which suggests that these species have the potential for paternal transmission but do not necessarily keep the paternal plastid (Hu *et al.*, 2008). If biparental inheritance and paternal leakage is widespread, then why maternal transmission is still view as the rule? Underestimation of biparental inheritance might come from the fact that (i) level of heteroplasmy due to this is generally low and vary in tissues and during plant development (Johnston, 2018), (ii) because this phenomenon is not investigated carefully enough or (iii) due to methodological limitations (Xu, 2005; Camus *et al.*, 2022).

### 1.2. Benefits of biparental transmission

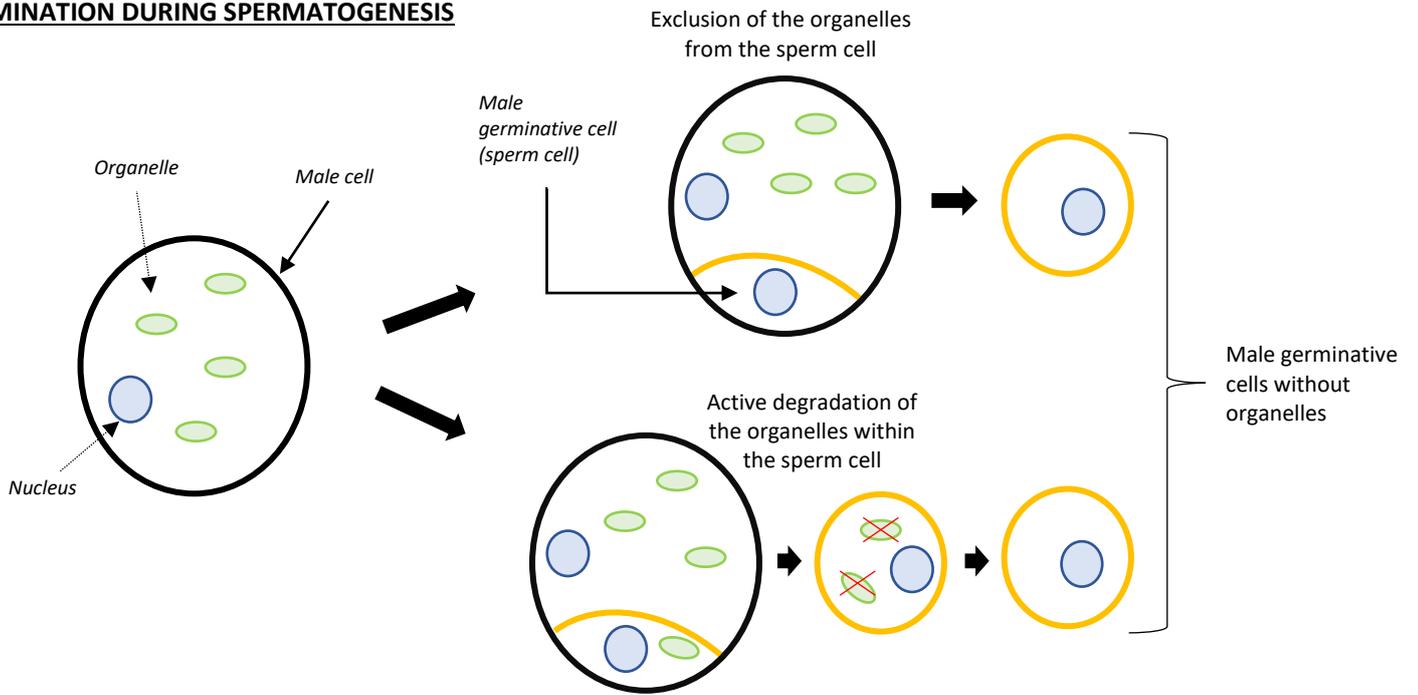
Biparental inheritance seems to have evolved repeatedly and independently with paternal leakage being lost and restored over evolutionary timeframes (Greiner *et al.*, 2014). Why is this loss/restoration dynamic observed for organellar inheritance pattern? Uniparental inheritance is thought to be the evolutionary most advantageous way of transmission. Briefly, biparental inheritance can lead to the spread of selfish cytoplasmic element and create intra-cytoplasmic conflict that can be costly for the organisms (Xu, 2005; Camus *et al.*, 2022) and it can breakdown cytonuclear co-adaptation (see (Greiner *et al.*, 2014 for more details). However, uniparental inheritance can also have deleterious consequences making this mode of transmission potentially unstable: (i) it can lead to mutation accumulation through Muller's ratchet (Greiner *et al.*, 2014), (ii) if the organellar genome structure is highly variable, the variation will ultimately be transmitted to progenies (Camus *et al.*, 2022), (iii) it can sometimes lead to a breakdown of the cytonuclear coevolution (Greiner *et al.*, 2014). Fall-back to biparental inheritance could allow some level of sexual organellar recombination and purge of the deleterious mutations (Greiner *et al.*, 2014). It could also lead to a breakdown of the linkage disequilibrium between organelles, so that the beneficial mutations are decoupled from the deleterious ones (Camus *et al.*, 2022). In case of hybridization, biparental inheritance can be advantageous as it increases the probability to have an organelle better matching the hybrid nuclear background (Greiner *et al.*, 2014; Barnard-Kubow *et al.*, 2017). For hybridizing populations then, biparental inheritance should be beneficial and stable if the fitness cost of incompatibilities is higher than the cost associated with stable heteroplasmy (Camus *et al.*, 2022). Finally, some studies highlighted the advantages of having organellar DNA in the pollen grains. For example, presence of plastid DNA in pollen grains could be beneficial as plastidial glycolysis could represent a great source of metabolic energy needed for pollen development and pollen tube growth, critical steps during plant sexual reproduction (Selinski & Scheibe, 2014). This would at least allow keeping the paternal organellar genome in the male generative cells. The balance between the cost associated with uniparental or biparental inheritance might then determine how organellar genomes will be transmitted.

### 1.3. Keeping only one organellar type

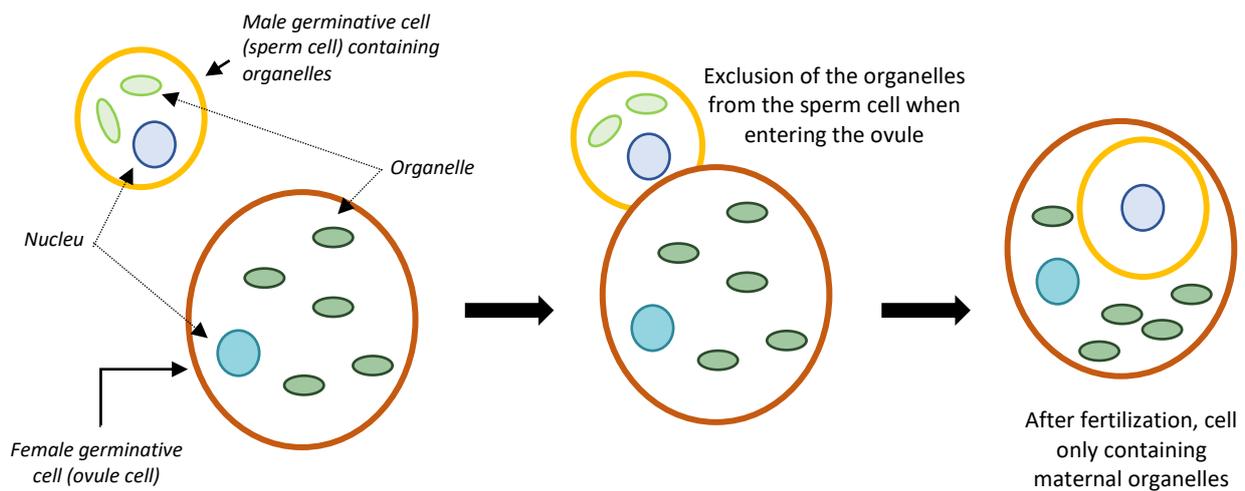
In case of biparental inheritance, both organellar type will not necessarily be maintained. Mechanisms of exclusion of one organellar type are not well understood but several explanations have been put forward, that we already discussed a bit in the above sections. Elimination of the paternal organellar genome is possible at several steps of sexual reproduction: (i) during gametogenesis, (ii) during

fertilization and (iii) post-fertilization, in the zygote (Birky, 2001; Barr *et al.*, 2005; Xu, 2005; Nagata, 2010) (Figure 3). Mechanisms involved at these different steps could work in combination or as back-up mechanisms if one of them fails (Xu, 2005). Elimination of one organellar type to reach homoplasmy is not well known (Shrestha *et al.*, 2021), but could be under nuclear control (Chiu *et al.*, 1988; Kmiec *et al.*, 2006; Thyssen *et al.*, 2012; McCauley, 2013; Greiner *et al.*, 2014; Radzvilavicius *et al.*, 2017; Broz *et al.*, 2022). Sato & Sato (2013) described an 'active degradation model' which can act before or after fertilization, and lead to the selective degradation of one of the two organellar types through various mechanisms, for example ubiquitin mediated degradation in male germinative cells in mammals but these have been mainly described outside the plant kingdom (Sato & Sato, 2013). Selection against one of the two types can also occur during the germline maturation, with fuel performance of the two types, their regulation dynamics and, in case of mitochondria, their ROS handling, influencing the force of selection for one of them (Latorre-pellicer *et al.*, 2019). Once kept in the male generative cells, organellar types can be eliminated either during or after entering the oocyte. Again, the active degradation model could play a role here, selectively preventing the paternal organellar genome from entering the oocyte (Sato & Sato, 2013). In inter-lineage hybrids, the presence of paternal organellar genomes could be the result of a breakdown of the potentially lineage-specific mechanisms normally excluding paternal genome during fertilization or later during zygote development (Barr *et al.*, 2005; McCauley, 2013; Breton & Stewart, 2015; Ramsey & Mandel, 2019). After fertilization, the sorting-out of the organellar types will be determined by either (i) genetic drift and random vegetative sorting, (ii) selection on particular organellar genomes regarding their composition in deleterious mutations or (iii) their competitive abilities (Greiner *et al.*, 2014). In case of anisogamous species, following the 'simple dilution model' of (Sato & Sato, 2013), one type would be lost just because one of the two gamete type is smaller than the other one and thus contains less organelles. Because organellar genomes experience a drastic bottleneck when entering the germline, and because the male gamete is usually much smaller than the female one, the number of organellar genomes in the pollen grain are supposed to be much lower than in the ovule (Barr *et al.*, 2005; Xu, 2005; Greiner *et al.*, 2014; Ramsey & Mandel, 2019; Parakatselaki & Ladoukakis, 2021). This results in biased input ratio of organelles of the two parents and higher probability of losing the minor organellar type (i.e. the paternal one) through vegetative segregation. So without selection acting, loss of the minor organellar type through drift would lead to homoplasmic individuals for the maternal organellar genomes (Breton & Stewart, 2015; Shrestha *et al.*, 2021). Competitive ability of the two transmitted types might also be of primary importance in the elimination of one of them. Factors determining the outcome of this competition may be diverse. For example organellar types could have different intrinsic rate of multiplication, with one type replicating faster and having a replicative advantage over the slower one that will persist until complete sort-out of the two types (Chiu *et al.*, 1988). Replicative advantage might also depend on the

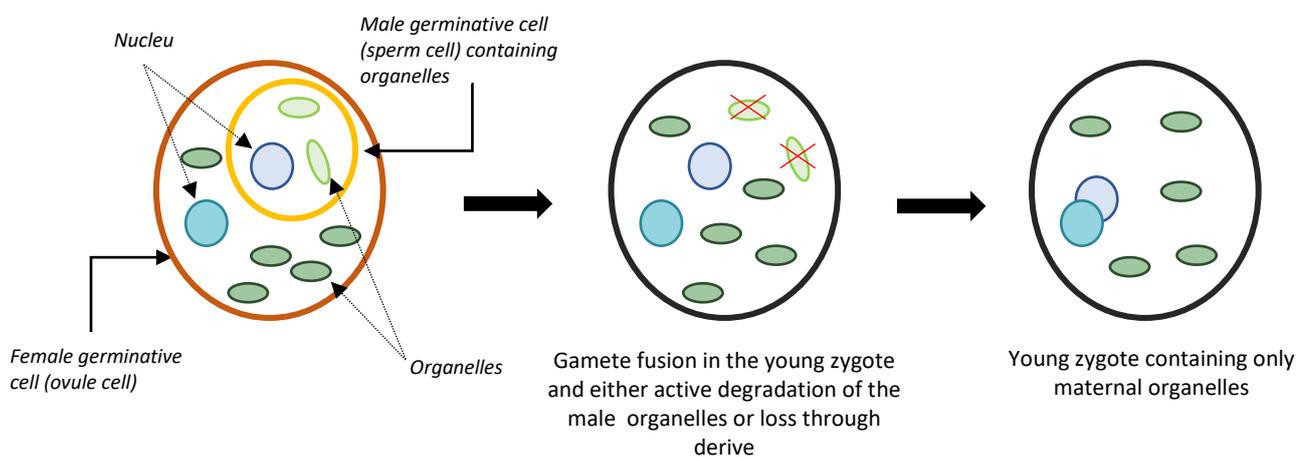
**ELIMINATION DURING SPERMATOGENESIS**



**ELIMINATION DURING FERTILIZATION**



**ELIMINATION DURING ZYGOTE DEVELOPMENT**



**Fig.3 Schematic representation of the three potential steps where the paternal organelles can be eliminated.** Inspired from Greiner et al, 2014 BioEssays and Sato & Sato, 2013 Biochimica & Biophysica acta

organellar genome size, with smaller genome generally replicating faster (Shrestha *et al.*, 2021; Camus *et al.*, 2022). Regarding the plastid, competitive ability is also thought to partially depend on the lipid composition of the plastid membrane, as it determines plastid stability and division rate (Sobanski *et al.*, 2019). Active degradation can also occur as for example selective autophagy of the paternal mitochondrial genome after fertilization in *C. elegans* (Sato & Sato, 2013). Another form of selective forces that would influence the exclusion of one organellar type, but only in inter-lineages hybrids, might be their level of compatibility with the hybrid nuclear background.

Though both organellar genome should experience at some point one of the above mechanisms to reach organellar homoplasmy, one major difference exist between mitochondrial and plastid : mitochondrial types can fuse and recombine but the two plastids will not often fuse and recombine (Sears, 1980; Birky, 2001). Mitochondria can recombine through the presence of numerous repeated sequences that are absent from the plastid genome, which explains the differential rate of intra-genomic recombination in organelles (Gualberto & Newton, 2017). This could mean that the selective pressure to eliminate one of the plastid types could be greater than the one observed on the mitochondrial types. Why can mitochondria fuse and plastid can't might rely on mechanic properties of organelles themselves.

#### 1.4. Maternal exclusion : sometimes the father is the best

For some species, sometimes the paternal organellar genome is kept. Keeping it or not might depend on how many defective mutations they contain (Hu *et al.*, 2008). Paternal organellar genomes are generally excluded because they are heavily damaged by reactive oxygen species (ROS) produced during spermatogenesis and the long sperm swim (Sato & Sato, 2013). Because of this high mutational load, it is more beneficial to keep the maternal organellar genome (Greiner *et al.*, 2014). Yet, if the male gametes experience less severe bottlenecks or less oxidative damage, the cost associated with their transmission would be reduced, allowing paternal inheritance of the organellar genome (Greiner *et al.*, 2014). Selection for one of the two organellar types and exclusion of the maternal genome, might also play at least a partial role somehow (Birky, 2001; Sato & Sato, 2013; Ramsey *et al.*, 2019; Sobanski *et al.*, 2019). (Wolff *et al.*, 2013) reported case of paternal organellar elements experiencing positive selection and replacing the maternal ones in *Drosophila simulans*. In plants, (Aksyonova *et al.*, 2005) reported the selective replication of the less incompatible plastid genome with the hybrid nuclear background in some barley-wheat hybrids. In the inter-lineages crosses, as we observed in *Silene nutans*, selection against the maladaptive cytotype might result in the loss of the maternal organellar genome (Greiner *et al.*, 2014). Still looking at inter-lineage hybrids, one could also imagine that the maternal lineage could have less competitive organellar genome that could lose the

competition and be excluded from the zygote. Plastid genome are supposed to transmit themselves better in crosses in which the nuclear genome of the progeny is closer to its native nuclear background (Chiu *et al.*, 1988), suggesting that the competitive ability might also depend on the hybrid nuclear background and its co-adaptation with it. Once organellar of the two parents are transmitted, exclusion of one of the parental organellar genome seems to be driven by a balance between selection for the fittest type and intrinsic capacity of the two organellar types themselves.

#### 1.5. Why do some species group have the potential and other don't?

Evidence are growing about species having the potential for biparental inheritance and paternal leakage, yet not all plant species are concerned. The first thing that could lead to this observation is the fact that it is easier to observe biparental inheritance and paternal leakage in inter-lineages hybrids. Hybridizing species might be more prone to transmit their paternal organellar genome because hybridization breaks the mechanisms preventing its transmission (conditioning on the initial presence of organelles in male reproductive cells) (Xu, 2005; Barnard-Kubow *et al.*, 2017; Shrestha *et al.*, 2021). Biparental inheritance could then be viewed as more common in those hybridizing species. Regarding the plastid, the presence of plastid heteroplasmy as a result of plastid biparental inheritance might also be more common in hybridizing species because of differential size of their plastome (Shrestha *et al.*, 2021). Supposing that the plastid genome is not systematically eliminated of the germinative cell (as observed in Dipsacales species), then when crossing individuals of distant lineages or different species, one might expect to observe differences in plastome size, so differences in replication speeds between the two species. The species with the smallest plastome will ultimately transmit its plastid genome to hybrid, regardless to its maternal or paternal origin, because it replicates faster. This would not be observed in intra-species hybrids as within species, the size of the plastome must be similar between individuals.

The mating system of the species might also greatly influence the inheritance pattern of the organellar genome (Xu, 2005). Gynodioecious species, with a cytonuclear sex determination system based on cytoplasmic male sterility factors in the mitochondrial genome, might favor paternal leakage of the mitochondrial genome (McCauley, 2013; Ramsey & Mandel, 2019). When organellar genomes are in tight linkage disequilibrium because of co-transmission, as described in *Silene vulgaris* (Olson & Mccauley, 2000), paternal leakage of the mitochondrial genome could drive the plastid genome with it, ending up in paternal leakage of both organellar genomes in this species.

General organellar genome instability, such as mechanisms mentioned above, might also increase the potential for biparental inheritance. Evidences points out to a correlation between plastome structural rearrangement (a hallmark of genome instability) and biparental inheritance in angiosperms (Jansen

& Ruhlman, 2012). Furthermore, several species with unstable organellar genomes and fast-evolving ones also display paternal leakage if not biparental inheritance of the organellar genomes (e.g. *Silene nutans*, *Campanulastrum americanum* (Barnard-Kubow *et al.*, 2014, 2017), some *Passiflora* species (Shrestha *et al.*, 2019, 2021). Imagine one unstable organellar genome that might be very labile and fast evolving as the one described in several *Silene* species, it could also disrupt either (i) its interaction with the nuclear genome, that it thought to partly control the inheritance pattern of the organelles or (ii) disrupt the mechanisms directly coming from the organelles itself. Further questions can emerge here such as, if there is a link between potential for biparental inheritance and organellar genome instability, which of the two evolved first? Is the general instability of the organellar genome a consequence or a cause of biparental inheritance? Given the fact that not all species with biparental inheritance exhibit unstable organellar genome disorder (e.g. *Silene vulgaris*), it is likely that genome instability might rather be a consequence of biparental inheritance than a cause for it. Still, organellar genome instability is not evolutionary stable as it increases evolutionary rates and with it, potentiality for unfit organelle production. So, speculatively, one could imagine that biparental inheritance could shift from uniparental inheritance to deal with this unstable organellar genome and allow the probability to gain a more evolutionary stable one.

## 2. Evolution of the organellar genomes in plants

The following sections discuss the rate of evolution of mitochondrial and plastid genomes. In order not to overload the manuscript, I sometimes give examples for the plastid or the mitochondrial genome and sometimes for both. I have tried to make this clear in each line of argument.

### 2.1. Organellar genome acceleration of evolutionary rate

Acceleration of both organellar genome rates of evolution have been observed repeatedly, in several genera and species (Jansen *et al.*, 2007; Magee *et al.*, 2010; Sloan *et al.*, 2012b; Weng *et al.*, 2012; Barnard-Kubow *et al.*, 2014; Sloan *et al.*, 2014a,b; Rockenbach *et al.*, 2016; Havird *et al.*, 2017; Park *et al.*, 2017; Shrestha *et al.*, 2019; Williams *et al.*, 2019; Abdel *et al.*, 2022). Though it seems to have occurred across independently evolving lineages (Sloan *et al.*, 2012b), common genomic signatures can be observed in these species: elevated  $d_N/d_S$ , i.e. relaxed selective constraint and presence of genome structural rearrangement, suggesting that common evolutionary mechanisms are acting (Sloan *et al.*, 2014a).

Organellar genome instability seems to be the mainly accepted driver of this acceleration. For example, correlation between changes in plastid gene order/loss of introns and lineage-specific increase in rate

of evolution have been identified in several angiosperm species (Jansen *et al.*, 2007). Still regarding the plastid genome, in several species with acceleration of its evolutionary rate, signatures of genome structural instability are observed such as contraction/expansion of the inverted repeat, introns losses, inversions and large indels along with elevated rates of amino acid substitutions (Weng *et al.*, 2012; Barnard-Kubow *et al.*, 2014; Sloan *et al.*, 2014a; Shrestha *et al.*, 2019). In *Silene* species with such fast-evolving plastome Sloan *et al.* identified inversions mediated by recombination between pair of small and imperfect repeats, which were absent from the “slowly” evolving *Silene* species (Sloan *et al.*, 2014a). Regarding the mitochondrial genome, in *Silene conica* and *Silene noctiflora*, the accelerated rate of substitution in the mitochondrial genome is associated with major expansion of the mitochondrial genome size (Sloan *et al.*, 2012a). But how might organellar genome instability be generated in the first place? The mitochondrial genome is supposed to be very labile in its structure, which might induce general instability of this genome (Wu *et al.*, 2020; Chevigny *et al.*, 2022). For the plastid genome, some species in the Geraniaceae family seems to exhibited fast-evolving plastid genome, triggered by DNA repair defect (Guisinger *et al.*, 2008). The organellar DNA reparation/replication/recombination machinery could be less effective (Jansen *et al.*, 2007; Drouin *et al.*, 2008; Guisinger *et al.*, 2008; Smith, 2016), which would result in sequence and structural changes consistent with relaxed purifying selection and elevated rate of amino-acid substitution (Jansen & Ruhlman, 2012; Sloan *et al.*, 2014b). The DNA repair machinery is partially encoded by nuclear genes (the RRR genes – repair/recombination/replication) regulating both organellar genomes (Sloan *et al.*, 2014a; Havird *et al.*, 2017). Mutations or loss of these nuclear genes would have significant effects of plant organelle genome evolution (Maréchal & Brisson, 2010; Shrestha *et al.*, 2019). A defective nuclear control of organellar gene expression could alter their expression and might result in increased rate of nucleotide substitution (Guisinger *et al.*, 2008). In some *Silene* species, the acceleration of both the mitochondrial and plastid genome evolutionary rates might be the result of pseudogenization/loss of these genes leading to instability of the organellar genome in these species (Sloan *et al.*, 2014a; Havird *et al.*, 2017). Analysis of the nuclear RRR genes in Geraniaceae species also points out in this direction leading to evolutionary rate acceleration in some plastid genes (Zhang *et al.*, 2015).

Other processes could influence the rate of evolution of the organellar genomes. The presence of genome mutator or stochastic mutagenic events (external or not) could lead to general or localized hypermutation (Jansen *et al.*, 2007; Magee *et al.*, 2010; Johnston, 2018). In the plastid genome for example, sites of gene-loss and genome rearrangement coincide with fast-evolving regions more often than expected by chance, consistent with the presence of genomic regions more or less prone to experience genetic changes (Jansen *et al.*, 2007; Magee *et al.*, 2010). Mutation in plastome mutator nuclear-encoded loci have been identified as resulting in increased rates of indels and substitution in *Oenothera* (Jansen *et al.*, 2007). Regarding the mitochondrial genome, changes in mtDNA stability

could be due to mutagenic environmental factors, such as increased ROS production, increased UV exposure, changes in physiology (Gaut *et al.*, 2011) that would increase mtDNA mutational load and rate of evolution. External stress could also alter mitochondria fusion/fission events (Johnston, 2018). Fusion between mitochondria leading to extensive recombination events (Gualberto & Newton, 2017; Chevigny *et al.*, 2022) could also generate extensive mitochondrial structural rearrangement (Shrestha *et al.*, 2019; Gonçalves *et al.*, 2020).

Finally, species-specific characteristic may also play a role in pace of evolution of both organellar genomes. For example, as mentioned above, biparental inheritance could be involved and drive destabilization of the organellar genomes, either through increased fusion and recombination of mitochondria types or increased competition between plastid types. Species can also use different pool of tRNA to translate the organellar genes (Drouin *et al.*, 2008). We can speculate that some tRNA pools could lead to less errors than other, generating differences in organellar genome evolution in different species. The metabolic rate, size of the body of the organisms, methylation processes or other species characteristic could also influence the rate of organellar genome evolution (Jansen *et al.*, 2007; Gaut *et al.*, 2011). For example, annual plant species are also supposed to have higher evolutionary rate than perennial plants (Gaut *et al.*, 2011). The number of observed nucleotide changes could also be due to the speciation rate: species rich group tends to have higher evolutionary rates because selection and/or genetic drift influence during speciation drives fixation of mutation (Gaut *et al.*, 2011). Another question relies on why these genomic changes will be more prone to be maintained. Decrease in efficacy of purifying selection could further play a role in maintaining these genomic rearrangement or sequence changes. Relaxed selective constraints can be the result of demographic events resulting in reduction of  $N_e$  (Drouin *et al.*, 2008; Sloan *et al.*, 2014a). Reduced  $N_e$  would increase influence of genetic drift, facilitating the maintenance of any genomic changes (Lynch *et al.*, 2006). Maintenance of these changes could also be the result of a “mutational masking”. Because there are many copies of the organellar genomes per cell, the deleterious changes can persist longer as they’re masked, subsequently being more common and at higher frequency than expected (Camus *et al.*, 2022). Finally, these structural and sequence changes could also be maintained in organellar genomes because there might be some adaptive advantages in shuffling gene order (Johnston, 2018). Yet, as mentioned earlier, signatures of relaxed selective pressure were identified in fast-evolving species, suggesting that local adaptation might not be the main player here.

Why does organellar genome in specific lineages or on specific genes sometimes exhibit acceleration of evolutionary rate is still largely mysterious. If genome instability is the main factor responsible for it, then why is instability maintained in the first place? Why in some species and not in others?

## 2.2. What factors drove gene-specific acceleration?

In the plastid genome, this acceleration mostly concern non-photosynthetic genes like the ones encoding the plastid ribosome, the RNA polymerase, *clpP1*, *accD*, *ycf1* and *ycf2* (Forsythe *et al.*, 2021). In the mitochondrial genome, the *atp9* gene is also often observed as fast evolving (Dan Sloan personal observation). Regarding the plastid genes, they are under the control of a different RNA polymerase than the photosynthetic plastid genes. Interaction with the transcriptional activity might affect the rate of molecular evolution (Guisinger *et al.*, 2008; Sloan *et al.*, 2014a). For both mitochondria and plastid genes, their genome localization could also influence its rate of molecular evolution: for example, the plastid genes localized in the inverted repeat tend to have a faster rate of evolution (Guisinger *et al.*, 2008). Levels of expression of the organellar genes correlated with their cellular function and could also explain their pace of evolutionary rate: the highly expressed genes, which also are the most functionally important, often evolve at lower rates (Guisinger *et al.*, 2008; Gaut *et al.*, 2011). The highly expressed genes need to be translated accurately as they are functionally essential, so selection for robustness against mistranslation is strong on these genes, leading to lower rates of evolution (Gaut *et al.*, 2011). Genes can also be more or less “labile”, and undergo illegitimate recombination like the *rpoA* plastid gene in the genus *Passiflora* (Shrestha *et al.*, 2019).

Several other mechanisms acting at the gene level might be responsible for gene-specific accelerated rate of evolution: the extent of methylation and the position of the gene in the protein network with the upstream gene being generally more constrained than the downstream ones (Gaut *et al.*, 2011). This gene-specific acceleration could also be the result of co-evolution with the nuclear genes in the organellar protein complexes, as observed for example in the plastid-encoded genes of the plastid ribosome in Geraniaceae species, where positive selection was detected on these genes as a result of plastid-nuclear co-evolution within the plastid ribosome (Weng *et al.*, 2016). Organellar genes can also be sometimes involved in adaptive response to environments (Budar & Roux, 2011; Budar & Fujii, 2012) so part of the mutations observed could be the results of adaptive evolution and contribute to a syndrome of gene-specific acceleration. Rate of transfer to the nucleus could play a role: if transfers are common, then we might observe remnant organellar genes that do not function in the organelles anymore but are not degraded yet (Shrestha *et al.*, 2019).

Focusing now only on the plastid genome, why the ribosomal and *clpP1* gene are almost always concerned by this acceleration? Even though they are not photosynthetic, some of them are still extremely important for the cell (Zoschke & Bock, 2018): ribosomal proteins deal with translation of the photosynthetic genes (Tiller *et al.*, 2012; Bieri *et al.*, 2017) and the *clpP1* gene is part of the CLP complex and an ATP-dependent proteolytic SU of caseinolytic peptidase, involved in protein metabolism within the plastid (Abdel *et al.*, 2022). Missense mutation in *rps4*, a ribosomal gene, leads to severe default in plants development (Tang *et al.*, 2018) further showing the importance of these

genes for cell and plant function. Most of the components of the translation machinery of the plastid proteins are encoded in the nucleus (Zoschke & Bock, 2018). Regarding the plastid ribosome, 50% of the small ribosomal proteins are encoded in the plastid and only 25% for the large one (Zoschke & Bock, 2018). Loss of ribosomal plastid genes is thought to be frequently associated with functional transfer to the nucleus (Shrestha *et al.*, 2019). Because of these transfers, and some transfers of the genes encoding RNA polymerase proteins to the nucleus, the plastid proteins could be mainly produced by the cytosolic ribosome rather than the plastid one, lowering the evolutionary constraint on the plastid ribosome (Zoschke & Bock, 2018). Yet, we did identify some positive selection in the ribosomal plastid genes in *S. nutans*, suggesting that their acceleration is not the result of relaxed selective constraint. Why then do these ribosomal gene exhibit such high level of amino acid substitutions? Do they have any influence on local adaptation that would explain the presence of positive selection on these genes in *S. nutans*? Is it an evolutionary answer to elevated rate of substitution in the nuclear genes? Regarding *clpP1*, some hypothesis have been put forward its elevated evolutionary rate such as its localization in a hypermutational hotspot (Magee *et al.*, 2010) or the presence of a mutagenic retroprocessing (Williams *et al.*, 2019) or a decrease of functional constraint on this gene because another protein took the precedence in its cellular function (Williams *et al.*, 2019). Yet, some positive selection was also identified on this gene in several species (Barnard-Kubow *et al.*, 2014; Sloan *et al.*, 2014a; Rockenbach *et al.*, 2016). If positive selection is identified it does not fit with relaxed selective constraint and fixation of mutation through genetic drift. Is the *clpP1* involved in local adaptation? Does it evolve in response to the numerous nuclear-encoded other subunits of the complex? These remain open questions.

### 2.3. Acceleration = speciation?

Emergence of CNIs might also be the result of general instability of the organellar genomes. Ability of CNIs to represent a reproductive barrier would depend on their outcome, which might be hard to predict as it involves several factors such as (i) the number of organellar and nuclear genes involved in the incompatibilities, which might dependent on the split time of the hybridizing species (Presgraves, 2010; Baack *et al.*, 2015; Schluter & Rieseberg, 2022) ; (ii) the functional impact of the CNIs : if it involved essential organellar pathways or not (this could also depend on the number of CNIs as the more there are, the more chance you have to get one in essential function) ; (iii) the presence of paternal leakage, as a mean of hybrids rescue from CNIs and (iv) the dominance architecture at the interacting nuclear loci : which of the maternal and paternal alleles would be expressed. Accelerated rate of evolution of organellar genomes could have implication for speciation as if the organellar genomes are unstable and fast-evolving, this could accelerate the CN co-adaptation in allopatric

lineages and increase their implication as reproductive barriers in case of secondary contact. So instable and fast-evolving organellar genome could, at the same time, accelerate the rate of speciation. CNIs are supposed to be among the first post-zygotic barriers to emerge during speciation (Levin, 2003; Fishman & Willis, 2006; Greiner *et al.*, 2011). It would be interesting to further investigate this idea and see whether the plant species exhibiting similar patterns of organellar genome diversity (i.e. fast-evolving organellar genome, structural rearrangement...) also contains more cryptic species that would be expected. RI can be easily detected when sampling and sequencing individuals in hybrid zones or where geographical ranges of divergent lineages overlap. RI is the most evident in experimental crosses when species are not in close geographical contact and have the possibility to hybridize in nature (Nosil *et al.*, 2017). So, it is likely that RI and cryptic species quantity is underestimated in nature. In case of speciation with gene flow, it would represent a fitness disadvantage to allow new mis-adapted nuclear alleles to enter a population, as it could disrupt co-evolution with the fast-evolving organellar genomes. So, this CN co-evolution, accelerated because of elevated rate of organellar genome evolution, could represent a selective force against introgression of plastid targeted nuclear loci. This would lead to high level of differentiation in these nuclear genes compared to the rest of the genome and make these nuclear loci kind of "speciation genes". This is true only if no paternal leakage is occurring, as if there are substantial levels of paternal leakage, selection could occur not against introgression of the mis-adapted nuclear alleles but favoring introgression of the most fitted organelle.

### BOX 1 - Open questions regarding organellar genome evolution (III)

Findings the cause of organellar genome instability and biparental inheritance in some species might greatly help understand the evolutionary dynamics at stakes in the organellar genomes and further understand its implication for genome evolutionary mechanisms.

#### Organellar genome evolutionary rate acceleration

- What factors mainly drive the acceleration of the organellar genome?
- Why is it repeatedly observed in a subset of plastid genes, which are essential to plastid complexes function?
- Does it speed up the speciation process through accelerated organello-nuclear coadaptation?
- Why does only a subset of species exhibit it?
- Any link with paternal leakage of the organellar genomes? Do the species exhibiting this phenomenon also exhibit paternal leakage?

#### Paternal leakage and biparental inheritance of the organellar genomes

- Any link with accelerated rate of evolution of the organellar genome? Is the accelerated rate of evolution of the organelles partially resulting from paternal leakage or biparental inheritance of the organelles?
- Why are only some species concerned? Are they the same species as the one exhibiting accelerated rate of organellar genome evolution?
- How is the paternal genome eliminated/excluded in isogamous species if not through loss of the minor organellar type? To what extent is selection acting? Are there “active” degradation mechanisms as the one observed outside the plant kingdom?
- Is biparental inheritance the ancestral state or uniparental transmission?
- To what extent can it slowdown the speciation process?

### III. References

- Abdel SE, Lisa G, Haleakala ML, Pal TH, Sloan DB. 2022.** Rapid sequence evolution is associated with genetic incompatibilities in the plastid Clp complex. *Plant Molecular Biology* **108**: 227–287.
- Adhikari B, Caruso CM, Case AL. 2019.** Beyond balancing selection: frequent mitochondrial recombination contributes to high-female frequencies in gynodioecious *Lobelia siphilitica* (Campanulaceae). *New Phytologist* **224**: 1381–1393.
- Aksyonova E, Sinyavskaya M, Danilenko N, Pershina L, Nakamura C, Davydenko O. 2005.** Heteroplasmy and paternally oriented shift of the organellar DNA composition in barley-wheat hybrids during backcrosses with wheat parents. *Genome* **48**: 761–769.
- Altschup SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990.** Basic Local Alignment Search Tool. *Journal of Molecular Biology* **215**: 403–410.
- Azhagiri AK, Maliga P. 2007.** Exceptional paternal inheritance of plastids in *Arabidopsis* suggests that low-frequency leakage of plastids via pollen may be universal in plants. *The Plant Journal* **52**: 817–823.
- Baack E, Melo MC, Rieseberg LH, Ortiz-Barrientos D. 2015.** The origins of reproductive isolation in plants. *New Phytologist* **207**: 968–984.
- Barnard-Kubow KB, McCoy MA, Galloway LF. 2017.** Biparental chloroplast inheritance leads to rescue from cytonuclear incompatibility. *New Phytologist* **213**: 1466–1476.
- Barnard-Kubow KB, Sloan DB, Galloway LF. 2014.** Correlation between sequence divergence and polymorphism reveals similar evolutionary mechanisms acting across multiple timescales in a rapidly evolving plastid genome. *BMC Evolutionary Biology* **14**: 1–10.
- Barnard-Kubow KB, So N, Galloway LF. 2016.** Cytonuclear incompatibility contributes to the early stages of speciation. *Evolution* **70**: 2752–2766.
- Barr CM, Barr CM, Neiman M, Taylor DR. 2005.** Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. *New Phytologist* **168**: 39–50.
- Bieri P, Leibundgut M, Saurer M, Boehringer D, Ban N. 2017.** The complete structure of the chloroplast 70S ribosome in complex with translation factor pY. *The EMBO Journal* **36**: 475–486.
- De Bilde J. 1973.** Etude génécologique du *Silene nutans* L. en Belgique : Populations du *Silene nutans* L. sur substrats siliceux et calcaires. *Revue Générale de Botanique* **80**: 161–176.
- Birky CW. 2001.** The inheritance of genes in mitochondria and chloroplasts: Laws, Mechanisms and Models. *Annual Review of Genetics* **35**: 125–148.
- Bogdanova VS, Kosterin OE. 2006.** A case of anomalous chloroplast inheritance in crosses of garden pea involving an accession of wild subspecies. *Doklady Biological Sciences* **406**: 44–46.
- Breton S, Stewart DT. 2015.** Atypical mitochondrial inheritance patterns in eukaryotes. *Genome* **58**: 423–431.

**Broz AK, Keene A, Gyorfy MF, Hodous M, Johnston IG, Sloan DB. 2022.** Sorting of plant mitochondrial and plastid heteroplasmy is extremely rapid and depends on MSH1 activity. *bioRxiv*.

**Budar F, Fujii S. 2012.** Cytonuclear Adaptation in Plants. *Advances in Botanical Research* **63**: 99–126.

**Budar F, Roux F. 2011.** The role of organelle genomes in plant adaptation: time to get to work! *Plant signaling & behavior* **6**: 635–639.

**Camus MF, Alexander-Lawrie B, Sharbrough J, Hurst GDD. 2022.** Inheritance through the cytoplasm. *Heredity* **129**: 31–43.

**Case AL, Finseth FR, Barr CM, Fishman L. 2016.** Selfish evolution of cytonuclear hybrid incompatibility in *Mimulus*. *Proceedings of the Royal Society B: Biological Sciences* **283**: 20161493.

**Case AL, Willis JH. 2008.** Hybrid male sterility in *Mimulus* (Phrymaceae) is associated with a geographically restricted mitochondrial rearrangement. *Evolution* **62**: 1026–1039.

**Chevigny N, Weber-Lotfi F, Le Blevenec A, Nadiras C, Fertet A, Bichara M, Erhardt M, Dietrich A, Raynaud C, Gualberto JM. 2022.** RADA-dependent branch migration has a predominant role in plant mitochondria and its defect leads to mtDNA instability and cell cycle arrest. *PLoS Genetics* **18**: e1010202.

**Chiu W-L, Stubbe W, Sears BB. 1988.** Plastid inheritance in *Oenothera*: organelle genome modifies the extent of biparental plastid transmission. *Current Genetics* **13**: 181–189.

**Drouin G, Daoud H, Xia J. 2008.** Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. *Molecular Phylogenetics and Evolution* **49**: 827–831.

**Feder JL, Egan SP, Nosil P. 2012.** The genomics of speciation-with- gene-flow. *Trends in Genetics* **28**: 342–350.

**Fénart S, Touzet P, Arnaud JF, Cuguen J. 2006.** Emergence of gynodioecy in wild beet (*Beta vulgaris* ssp . *maritima* L.): a genealogical approach using chloroplastic nucleotide sequences. *Proceedings of the Royal Society B: Biological Sciences* **273**: 1391–1398.

**Fishman L, Willis JH. 2006.** Cytonuclear Incompatibility Causes Anther Sterility in *Mimulus* Hybrids. *Evolution* **60**: 1372.

**Forsythe E, Williams AM, Sloan D. 2021.** Genome-wide signatures of plastid-nuclear coevolution point to repeated perturbations of plastid proteostasis systems across angiosperms. *Plant Cell* **33**: 980–997.

**Galloway LF, Fenster CB. 2001.** Nuclear and Cytoplasmic Contributions To Intraspecific Divergence in an Annual Legume. *Evolution* **55**: 488–497.

**Garraud C, Brachi B, Dufay M, Touzet P, Shykoff JA. 2011.** Genetic determination of male sterility in gynodioecious *Silene nutans*. *Heredity* **106**: 757–764.

**Gaut B, Yang L, Takuno S, Eguiarte LE, Gaut B, Yang L, Takuno S, Eguiarte LE. 2011.** The Patterns and Causes of Variation in Plant Nucleotide Substitution Rates. *Annual Review of Ecology, Evolution, and Systematics* **42**: 245–266.

**Gertz EM, Yu Y, Agarwala R, Schäffer AA, Altschul SF. 2006.** Composition-based statistics and translated nucleotide searches : Improving the TBLASTN module of BLAST. *BMC Biology* **14**: 1–14.

**Gonçalves DJP, Jansen RK, Ruhlman TA, Mandel JR. 2020.** Under the rug : Abandoning persistent misconceptions that obfuscate organelle evolution. *Molecular Phylogenetics and Evolution* **151**: 106903.

**Greiner S, Rauwolf U, Meurer J, Herrmann RG. 2011.** The role of plastids in plant speciation. *Molecular Ecology* **20**: 671–691.

**Greiner S, Sobanski J, Bock R. 2014.** Why are most organelle genomes transmitted maternally? *BioEssays* **37**: 80–94.

**Gualberto JM, Newton KJ. 2017.** Plant Mitochondrial Genomes: Dynamics and Mechanisms of Mutation. *Annual Review of Plant Biology* **68**: 225–252.

**Guisinger MM, Kuehl J V, Boore JL, Jansen RK. 2008.** Genome-wide analyses of Geraniaceae plastid DNA reveal unprecedented patterns of increased nucleotide substitutions. *PNAS* **105**: 18424–18429.

**Hanson MR, Bentolila S. 2004.** Interactions of mitochondrial and nuclear genes that affect male gametophyte development. *Plant Cell* **16**: S154–S169.

**Havird JC, Trapp P, Miller CM, Bazos I, Sloan DB. 2017.** Causes and Consequences of Rapidly Evolving mtDNA in a plant lineage. *Genome Biology and Evolution* **9**: 323–336.

**Havird JC, Whitehill NS, Snow CD, Sloan DB. 2015.** Conservative and compensatory evolution in oxidative phosphorylation complexes of angiosperms with highly divergent rates of mitochondrial genome evolution. *Evolution* **69**: 3069–3081.

**Hill GE. 2016.** Mitonuclear coevolution as the genesis of speciation and the mitochondrial DNA barcode gap. *Ecology and Evolution* **6**: 5831–5842.

**Hu Y, Zhang Q, Rao G, Sodmergen. 2008.** Occurrence of Plastids in the Sperm Cells of Caprifoliaceae : Biparental Plastid Inheritance in Angiosperms is Unilaterally Derived from Maternal Inheritance. *Plant Cell Physiol.* **49**: 958–968.

**Ingvarsson K, Taylor DR. 2002.** Genealogical evidence for epidemics of selfish genes. *PNAS* **99**: 11265–11269.

**Jansen RK, Cai Z, Raubeson LA, Daniell H, Claude W, Leebens-mack J, Mu KF, Guisinger-bellian M, Haberle RC, Hansen AK, et al. 2007.** Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. **104**: 19369–19374.

**Jansen RK, Ruhlman TA. 2012.** Plastid genomes of seed plants. *Genomics of chloroplasts and mitochondria* **35**: 103–126.

**Johnston IG. 2018.** Tension and Resolution : Dynamic , Evolving Populations of Organelle Genomes within Plant Cells. *Molecular Plant* **12**: 764–783.

**Kmiec B, Woloszynska M, Janska H. 2006.** Heteroplasmy as a common state of mitochondrial genetic

information in plants and animals. *Current Genetics* **50**: 149–159.

**Kühn K, Obata T, Feher K, Bock R, Fernie AR, Meyer EH. 2015.** Complete Mitochondrial Complex I Deficiency Induces an Up-Regulation of Respiratory Fluxes That Is Abolished by Traces of Functional Complex I. *Plant Physiology* **168**: 1537–1549.

**Lahiani E, Dufay M, Castric V, Le Cadre S, Charlesworth D, Van Rossum F, Touzet P. 2013.** Disentangling the effects of mating systems and mutation rates on cytoplasmic diversity in gynodioecious *Silene nutans* and dioecious *Silene otites*. *Heredity* **111**: 157–164.

**Latorre-pellicer A, Lechuga-vieco AV, Johnston IG, Hamalainen RH, Pellico J, Justo-Mendez R, Fernandez-Toro JM, Claveria C, Guaras A, Sierra R, et al. 2019.** Regulation of Mother-to-Offspring Transmission of mtDNA Heteroplasmy Short Article Regulation of Mother-to-Offspring Transmission of mtDNA Heteroplasmy. *Cell Metabolism* **30**: 1120–1130.

**Levin DA. 2003.** The cytoplasmic factor in plant speciation. *Systematic Botany* **28**: 5–11.

**Liebers M, Grübler B, Chevalier F, Lerbs-mache S, Merendino L, Blanvillain R, Pfannschmidt T. 2017.** Regulatory Shifts in Plastid Transcription Play a Key Role in Morphological Conversions of Plastids during Plant Development. *Frontiers in Plant Science* **8**: 1–8.

**Lynch M, Koskella B, Schaack S. 2006.** Mutation Pressure and the Evolution of Organelle Genomic Architecture. *Science* **311**: 1727–1730.

**Magee AM, Aspinall S, Rice DW, Cusack BP, Se M, Milbourne D, Barth S, Perry AS, Palmer JD, Gray JC, et al. 2010.** Localized hypermutation and associated gene losses in legume chloroplast genomes. *Genome Research* **20**: 1700–1710.

**Maréchal A, Brisson N. 2010.** Recombination and the maintenance of plant organelle genome stability. *New Phytologist* **186**: 299–317.

**Martin H. 2016.** Processus de spéciation et impact des systèmes de reproduction dans le genre *Silene*.

**Martin H, Touzet P, Dufay M, Godé C, Schmitt E, Lahiani E, Delph LF, Van Rossum F. 2017.** Lineages of *Silene nutans* developed rapid, strong, asymmetric postzygotic reproductive isolation in allopatry. *Evolution* **71**: 1519–1531.

**Martin H, Touzet P, Rossum F Van, Delalande D, Arnaud J. 2016.** Phylogeographic pattern of range expansion provides evidence for cryptic species lineages in *Silene nutans* in Western Europe. *Heredity* **116**: 286–294.

**Massouh A, Schubert J, Yaneva-Roder L, Ulbricht-Jones ES, Zupok A, Johnson MTJ, Wright SI, Pellizzer T, Sobanski J, Bock R, et al. 2016.** Spontaneous chloroplast mutants mostly occur by replication slippage and show a biased pattern in the plastome of *Oenothera*. *Plant Cell* **28**: 911–929.

**McCauley DE. 2013.** Paternal leakage, heteroplasmy, and the evolution of plant mitochondrial genomes. *New Phytologist* **200**: 966–977.

**McCauley DE, Bailey MF. 2009.** Recent advances in the study of gynodioecy : the interface of theory

and empiricism. *Annals of Botany* **104**: 611–620.

**McCauley DE, Bailey MF, Sherman NA, Darnell MZ. 2005.** Evidence for paternal transmission and heteroplasmy in the mitochondrial genome of *Silene vulgaris*, a gynodioecious plant. *Heredity* **95**: 50–58.

**McCauley DE, Sundby AK, Bailey MF, Welch ME. 2007.** Inheritance of chloroplast DNA is not strictly maternal in *Silene vulgaris* (Caryophyllaceae): evidence from experimental crosses and natural populations. *American Journal of Botany* **94**: 1333–1337.

**Nagata N. 2010.** Mechanisms for independent cytoplasmic inheritance of mitochondria and plastids in angiosperms. *Journal of Plant Research* **123**: 193–199.

**Nosil P, Feder JL, Flaxman SM, Gompert Z. 2017.** Tipping points in the dynamics of speciation. *Nature Ecology & Evolution* **1**: 0001.

**Olson MS, McCauley DE. 2000.** Linkage disequilibrium and phylogenetic congruence between chloroplast and mitochondrial haplotypes in *Silene vulgaris*. *Proceedings of the Royal Society of London B: Biological Sciences* **267**: 1801–1808.

**Parakatselaki M, Ladoukakis ED. 2021.** mtDNA Heteroplasmy : Origin, Detection, Significance, and Evolutionary Consequences. *Life* **11**: 633.

**Park S, Ruhlman TA, Weng ML, Hajrah NH, Sabir JSM, Jansen RK. 2017.** Contrasting patterns of nucleotide substitution rates provide insight into dynamic evolution of plastid and mitochondrial genomes of *Geranium*. *Genome Biology and Evolution* **9**: 1766–1780.

**Pearl SA, Welch ME, McCauley DE. 2008.** Mitochondrial Heteroplasmy and Paternal Leakage in Natural Populations of *Silene vulgaris*, a Gynodioecious Plant. *Molecular Biology and Evolution* **26**: 537–545.

**Postel Z, Poux C, Gallina S, Varré J-S, Godé C, Schmitt E, Meyer E, Van Rossum F, Touzet P. 2022.** Reproductive isolation among lineages of *Silene nutans* (Caryophyllaceae): A potential involvement of plastid-nuclear incompatibilities. *Molecular Phylogenetics and Evolution* **169**: 107436.

**Presgraves DC. 2010.** The molecular evolutionary basis of species formation. *Nature Reviews Genetics* **11**: 175–180.

**Radzvilavicius AL, Lane N, Pomiankowski A. 2017.** Sexual conflict explains the extraordinary diversity of mechanisms regulating mitochondrial inheritance. *BMC Biology* **15**: 94.

**Ramsey AJ, Mandel JR. 2019.** When one genome is not enough: organellar heteroplasmy in plants. *Annual Plant Reviews* **2**: 1–40.

**Ramsey AJ, McCauley DE, Mandel JR. 2019.** Heteroplasmy and Patterns of Cytonuclear Linkage Disequilibrium in Wild Carrot. *Integrative and Comparative Biology* **59**: 1005–1015.

**Rand DM, Haney RA, Fry AJ. 2004.** Cytonuclear coevolution: The genomics of cooperation. *Trends in Ecology and Evolution* **19**: 645–653.

**Rockenbach K, Havird JC, Monroe JG, Triant DA, Taylor DR, Sloan DB. 2016.** Positive Selection in

Rapidly Evolving Plastid – Nuclear Enzyme Complexes. *Genetics* **204**: 1507–1522.

**Van Rossum F, De Bilde J, Lefèbvre C. 1996.** Barriers to hybridization in calcicolous and silicicolous populations of *Silene nutans* from Belgium. *Royal Botanical Society of Belgium* **129**: 13–18.

**Van Rossum F, Martin H, Le Cadre S, Brachi B, Christenhusz MJM, Touzet P. 2018.** Phylogeography of a widely distributed species reveals a cryptic assemblage of distinct genetic lineages needing separate conservation strategies. *Perspectives in Plant Ecology, Evolution and Systematics* **35**: 44–51.

**Ruhlman TA, Jansen RK. 2018.** Aberration or Analogy ? The Atypical Plastomes of Geraniaceae. *Plastid Genome Evolution* **85**: 223–262.

**Sambatti JBM, Ortiz-Barrientos D, Baack EJ, Rieseberg LH. 2008.** Ecological selection maintains cytonuclear incompatibilities in hybridizing sunflowers. *Ecology Letters* **11**: 1082–1091.

**Sato M, Sato K. 2013.** Maternal inheritance of mitochondrial DNA by diverse mechanisms to eliminate paternal mitochondrial DNA. *Biochimica et Biophysica Acta* **1833**: 1979–1984.

**Schluter D, Rieseberg LH. 2022.** Three problems in the genetics of speciation by selection. *PNAS* **119**: e2122153119.

**Sears BB. 1980.** Elimination of plastids during spermatogenesis and fertilization in the plant kingdom. *Plasmid* **4**: 233–255.

**Selinski J, Scheibe R. 2014.** Pollen tube growth : where does the energy come from ? *Plant signaling & behavior* **9**: e977200.

**Shrestha B, Gilbert LE, Ruhlman TA. 2021.** Clade-Specific Plastid Inheritance Patterns Including Frequent Biparental Inheritance in *Passiflora* Interspecific Crosses. *International Journal of Molecular Sciences* **22**: 2278.

**Shrestha B, Weng M, Theriot EC, Gilbert LE, Ruhlman TA, Krosnick SE, Jansen RK. 2019.** Highly accelerated rates of genomic rearrangements and nucleotide substitutions in plastid genomes of *Passiflora* subgenus *Decaloba*. *Molecular Phylogenetics and Evolution* **138**: 53–64.

**Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, Mccauley DE, Palmer JD, Taylor DR. 2012a.** Rapid Evolution of Enormous, Multichromosomal Genomes in Flowering Plant Mitochondria with Exceptionally High Mutation Rates. *PLoS Biology* **10**: e1001241.

**Sloan DB, Alverson AJ, Wu M, Palmer JD, Taylor DR. 2012b.** Recent acceleration of plastid sequence and structural evolution coincides with extreme mitochondrial divergence in the angiosperm genus *Silene*. *Genome Biology and Evolution* **4**: 294–306.

**Sloan DB, Havird JC, Sharbrough J. 2017.** The on-again, off-again relationship between mitochondrial genomes and species boundaries. *Molecular Ecology* **26**: 2212–2236.

**Sloan DB, Triant DA, Forrester NJ, Bergner LM, Wu M, Taylor DR. 2014a.** A recurring syndrome of accelerated plastid genome evolution in the angiosperm tribe *Sileneae* (Caryophyllaceae). *Molecular Phylogenetics and Evolution* **72**: 82–89.

**Sloan DB, Triant DA, Wu M, Taylor DR. 2014b.** Cytonuclear interactions and relaxed selection accelerate sequence evolution in organelle ribosomes. *Molecular Biology and Evolution* **31**: 673–682.

**Sloan DB, Warren JM, Williams AM, Wu Z, Abdel-Ghany SE, Chicco AJ, Havird JC. 2018.** Cytonuclear integration and co-evolution. *Nature Reviews Genetics* **19**: 635–648.

**Smith DR. 2016.** The mutational hazard hypothesis of organelle genome evolution: 10 years on. *Molecular Ecology* **25**: 3769–3775.

**Sobanski J, Giavalisco P, Fischer A, Kreiner JM, Walther D, Schöttler MA, Pellizzer T, Golczyk H, Obata T, Bock R, et al. 2019.** Chloroplast competition is controlled by lipid biosynthesis in evening primroses. *PNAS* **116**: 5665–5674.

**Sobel JM, Chen GF, Watt LR, Schemske DW. 2010.** The biology of speciation. *Evolution* **64**: 295–315.

**Tang X, Wang Y, Zhang Y, Huang S, Liu Z, Fei D, Feng H. 2018.** A missense mutation of plastid RPS4 is associated with chlorophyll deficiency in Chinese cabbage (*Brassica campestris ssp. pekinensis*). *BMC Plant Biology* **18**: 1–11.

**Thyssen G, Svab Z, Maliga P, Brunswick N. 2012.** Exceptional inheritance of plastids via pollen in *Nicotiana sylvestris* with no detectable paternal mitochondrial DNA in the progeny. *The Plant Journal* **72**: 84–88.

**Tiller N, Weingartner M, Thiele W, Maximova E, Schöttler MA, Bock R. 2012.** The plastid-specific ribosomal proteins of *Arabidopsis thaliana* can be divided into non-essential proteins and genuine ribosomal proteins. *Plant Journal* **69**: 302–316.

**Touzet P, Delph LF. 2009.** The Effect of Breeding System on Polymorphism in Mitochondrial Genes of *Silene*. *Genetics* **644**: 631–644.

**Turner TL, Hahn MW, Nuzhdin S V. 2005.** Genomic Islands of Speciation in *Anopheles gambiae*. *PLoS Biology* **3**: e285.

**Weng ML, Ruhlman TA, Gibby M, Jansen RK. 2012.** Phylogeny, rate variation, and genome size evolution of *Pelargonium* (Geraniaceae). *Molecular Phylogenetics and Evolution* **64**: 654–670.

**Weng ML, Ruhlman TA, Jansen RK. 2016.** Plastid–nuclear interaction and accelerated coevolution in plastid ribosomal genes in Geraniaceae. *Genome Biology and Evolution* **8**: 1824–1838.

**Williams AM, Friso G, J. van Wijk K, Sloan DB. 2019.** Extreme variation in rates of evolution in the plastid Clp protease complex. *Plant Journal* **98**: 243–259.

**Wolff JN, Nafisinia M, Sutovsky P, Ballard JWO. 2013.** Paternal transmission of mitochondrial DNA as an integral part of mitochondrial inheritance in metapopulations of *Drosophila simulans*. *Heredity* **110**: 57–62.

**Wu Z, Waneka G, Broz AK, King CR, Sloan DB. 2020.** MSH1 is required for maintenance of the low mutation rates in plant mitochondrial and plastid genomes. *PNAS* **117**: 16448–16455.

**Xu J. 2005.** The inheritance of organelle genes and genomes : patterns and mechanisms. *Genome* **48**:

951–958.

**Yao JL, Cohen D. 2000.** Multiple gene control of plastome-genome incompatibility and plastid DNA inheritance in interspecific hybrids of *Zantedeschia*. *Theoretical and Applied Genetics* **101**: 400–406.

**Zhang Q, Liu Y, Sodmergen. 2003.** Examination of the Cytoplasmic DNA in Male Reproductive Cells to Determine the Potential for Cytoplasmic Inheritance in 295 Angiosperm Species. *Plant and Cell Physiology* **44**: 941–951.

**Zhang J, Ruhlman TA, Sabir J, Blazier JC, Jansen RK. 2015.** Coordinated rates of evolution between interacting plastid and nuclear genes in Geraniaceae. *Plant Cell* **27**: 563–573.

**Zoschke R, Bock R. 2018.** Chloroplast translation: Structural and functional organization, operational control, and regulation. *Plant Cell* **30**: 745–770.



## **Abstract (English)**

Speciation is the process by which the emergence of reproductive barriers isolate populations from one another and ultimately lead to the formation of new species. How these reproductive barriers emerge is a core question when thinking of speciation. Organellar genomes might be involved in the speciation process, through cytonuclear incompatibilities. Their mode of transmission might also influence the pace of reproductive isolation evolution. In my PhD, I worked on how organellar genomes influence the evolution of reproductive isolation between isolated lineages of *S. nutans* and which evo-demographic scenario shaped their evolution. Using plastid genomic and nuclear transcriptomic data we tried, in the first chapter, to identify candidates for plastid-nuclear incompatibilities involved in RI between lineages of *S. nutans*. We further dug into one plastid candidate complex, the plastid ribosome. Because RI seems to be incomplete between lineages of *S. nutans* as some inter-lineage hybrids survived, we tested for paternal leakage of the plastid genome. We genotyped the surviving hybrids for plastid SNPs and analyzed whether they inherited the paternal or maternal plastid genomes. By allowing the transmission of the less incompatible plastid genome, paternal leakage rescued some of the inter-lineage hybrids. The mitochondrial genome could also be involved in the RI, through mito-nuclear incompatibilities. Because of their co-transmission, organellar genomes are supposed to be in tight linkage-disequilibrium, so exhibiting similar evolutionary patterns. Using genomic data for both organellar genomes for individuals of the four lineages we compared their evolutionary patterns. They were different with mitochondrial genes exhibiting many shared polymorphisms while plastid genomes many fixed substitutions between lineages. Recombination-like events were also identified in the mitochondrial genes. Lastly, we reconstructed the evo-demographic histories of the four lineages of *S. nutans*, using RNAseq data and ABC methods. Allopatric speciation was identified between the four lineages, with split times consistent with the glacial maxima.

## **Résumé (Français)**

Via l'émergence de barrières à la reproduction qui isolent les populations les unes des autres, la spéciation est le processus qui conduit à la formation de nouvelles espèces. Par ailleurs, les génomes organellaires peuvent être impliqués dans ce processus, par le biais d'incompatibilités cytonucléaires. Leur mode de transmission peut également influencer l'évolution de l'isolement reproducteur (IR) entre populations. Dans ce travail de thèse, j'ai travaillé sur l'influence des génomes organellaires sur l'évolution de l'isolement reproducteur entre quatre lignées de *Silene nutans* et ai tenté de reconstruire le scénario évo-démographique qui a façonné leur évolution. Dans un premier temps, via l'utilisation de données génomiques et transcriptomiques, nous avons tenté d'identifier des candidats d'incompatibilités chloro-nucléaires impliquées dans l'IR entre ces lignées. Nous avons ensuite approfondi l'analyse d'un complexe candidat: le ribosome chloroplastique. Par ailleurs, l'IR semble être incomplet entre ces lignées puisque certains hybrides ont survécu. Nous avons donc testé une transmission paternelle du génome chloroplastique chez cette espèce, qui pourrait avoir sauvé certains de ces hybrides. Nous avons génotypé les hybrides survivants pour six SNP chloroplastiques et déterminé s'ils avaient hérité du génome chloroplastique paternel ou maternel. En permettant la transmission d'un génome chloroplastique moins incompatible, la fuite paternelle semble bien avoir sauvé certains de ces hybrides. Les génomes mitochondriaux pourraient également être impliqués dans l'IR, par le biais d'incompatibilités mito-nucléaires. Du fait de leur co-transmission, les génomes organellaires sont supposés être en déséquilibre de liaison étroit, présentant ainsi des schémas évolutifs similaires. Nous les avons comparés en utilisant les données génomiques des deux génomes organellaires, pour des individus des quatre lignées. Ces schémas évolutifs se sont révélés particulièrement contrastés, les gènes mitochondriaux présentant du polymorphisme partagé à l'inverse des gènes chloroplastiques contenant des substitutions fixées différemment entre lignées. Des événements de type recombinaison ont également été identifiés dans les gènes mitochondriaux. Enfin, nous avons reconstruit l'histoire évo-démographique des quatre lignées de *S. nutans*, en utilisant les données RNAseq et des méthodes ABC. Un scénario de spéciation allopatrique a été identifiée entre les quatre lignées, avec des temps de séparation cohérent avec les maximums glaciaires.