Zooming in on the lettuce mitochondrial genome, Study of induced and natural mitochondrial diversity within the Lactuca spp.

Arnaud Fertet

Université de Strasbourg

UNIVERSITÉ DE STRASBOURG

École Doctorale des Sciences de la Vie et de la Santé STRASBOURG

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTÉ

Institut de biologie moléculaire des plantes - UPR2357 - CNRS

THÈSE DE DOCTORAT

présentée par :

Arnaud FERTET

soutenu le : **17 septembre 2020**

Pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/Spécialité : Science du vivant/Aspect moléculaire et cellulaire de la biologie

Zooming in on the lettuce mitochondrial genome, Study of induced and natural mitochondrial diversity within the *Lactuca spp*.

Thèse dirigée par :		
Dr. GUALBERTO José Manuel	CNRS - Université de Strasbourg	Directeur de thèse
Rapporteurs :		
Dr. BUDAR Françoise	INRAE - IJPB Versailles	Rapportrice externe
Pr. SCHRANZ Eric	Université de Wageningen	Rapportrice externe
Autre membre du jury :		
Dr. GIÉGÉ Philippe	CNRS - Université de Strasbourg	Examinateur interne

REMERCIEMENTS/ACKNOWLDGMENTS

Merci ! / Thank you!

Thank you, first of all, to the doctors Françoise Budar, Philippe Giégé and professor Eric Schranz for having accepted to evaluate and discuss my thesis work.

Merci, ensuite, à mes deux encadrants José Gualberto et Émilie Guilloteau-Fonteny pour m'avoir offert la chance de prendre part à ce projet ainsi que pour la bienveillance dont ils ont fait preuve envers-moi tout au long de celui-ci. Merci pour vos conseils, votre aide et votre patience sans quoi tout cela n'aurait pas été possible. Merci aussi pour vos longues heures passées sur ce manuscrit.

Thank you to the doctor Gert-Jan de Boer for his time and wise advice. Thank you also for attending our meetings, there is no doubt that without you this thesis would look like completely different. Le Stade de La Meinau is still waiting for you.

Thank you to all the Enza-colleagues involved in this project. Thank you to Bas v.H. for the training in bioinformatics (that was definitely fun). Thank you also to Anke, Anaelle, Bas t.R., Camille, Elodie Fausto, Giuditta, Ilja, Lorena, Mathias, Mariann, Nadia, Richard and Willeke.

Merci également aux collègues d'Enza qui m'ont fait me sentir comme chez moi à Allonnes et Enkhuizen. Merci à Chantal, Elodie, Fausto, Judit, Nicolas et Samy. Merci à Emilie (encore) pour l'accueil que tu m'as réservé à chaque fois ainsi que pour les pizzas et les courgettes du jardin. Merci particulièrement à Camille (+ la colloc' de la petite bilange + Aux Brèves de Comptoir), Mathias et William pour Amsterdam et la vie à Saumur.

Merci à tous les membres de la team-312, passés et présents, que j'ai eu la chance de côtoyer. Merci à Déborah, Frédérique, José et Nicolas C.. Merci à Frédérique pour les relectures *in extremis* et inestimables. Merci à l'ouragan (de catégorie 4) José de toujours veiller au rangement du laboratoire. Merci à Déborah pour tout ce que tu es. Merci également pour les manips dont tu m'as soulagé et ce fameux gel qui te suivra encore longtemps. Merci aussi à tous les stagiaires qui ont eu la chance de me côtoyer. Merci à Denisa, Rebecca et Xavier.

Merci à tous les IBMPeers impliqués dans ce projet. Merci à Laurence D. pour le coup d'oeil sur les tRNAs de ma salade. Merci à Malek et Sandrine pour les nombreux séquençages de ce projet ainsi que pour votre accueil toujours chaleureux. Merci à David, Stéphanie et Valérie pour tous les conseils en

bio-informatique que vous m'avez fournis. Merci à la plateforme de production végétale et notamment à Sebastien responsable en chef des laitues sous la serre.

Merci à toutes les personnes de l'IBMP pour ce que vous faites de ce lieu où je me suis senti si bien pendant les 3 dernières années. Merci à Laurence H. de m'avoir fait intégrer les coureurs de l'IBMP. Merci aux autres coureurs plus ou moins réguliers que sont Bart, Kayo, Mathieu, Nicolas P., Thomas P., et Richard pour les kilomètres ensemble dans la bonne humeur. Merci également aux copains et copines rencontrés ici. Grazie à Marlène ma partenaire de rédaction pour tous ces thés partagés ensemble ainsi que ton inestimable aide sur la fin. Ευχαριστώ à Magdalène pour avoir fortement embelli mes trois années passées à l'IBMP. Merci également aux autres Docnald Trump que sont Chloé, Guillaume, Laetitia, Laura, Lucie et Valentin pour les bons moments partagés (à défaut d'avoir make IBMP great again).

Merci à mon amour de toujours le Racing Club de Strasbourg. Merci au président Marc Keller d'avoir remis notre ville sur la carte d'Europe du football. Merci à Dimitri Liénard, illustre terrifortain, pour ce coup franc qui me donne toujours 2 ans après autant de frissons. Merci à tous les supporters de ce club si particulier, de faire de notre maison l'un des endroits où je me sens le mieux. Merci à mes compères de supporter mon attitude et mes blagues à la Meinau. Merci à Benjamin de m'avoir suivi dans ce délire qui nous aura mené assez loin. Merci aussi à mes autres colocataires de la tribune Nord Honneur Haute Rang A que sont Axel et Victor. Merci à Nathan ensuite pour nos déplacements et weekends improvisés à suivre le Racing avec plus ou moins de réussite et pour l'hébergement toujours 5 étoiles lors de mes passages à Paris. Merci également à tous les copains présents à Lille pour ce moment magique et cette coupe qui pesait lourd dans le sac au retour.

Merci aux copains qui sont loin. Merci aux copains parisiens. Merci à Andrea pour nos concerts rock improvisés à chanter avec plus ou moins de yaourt et pour l'hébergement toujours 5 étoiles lors de mes passages à Paris (cette phrase me dit quelque chose …). Merci à Étienne qui restera mon seul et unique colocataire strasbourgeois. Merci au copain lyonnais, Florian qui était là pour les années les plus compliquées en Licence. Merci aux copains bruxellois, John et Juliette en tête pour les bons moments passés en Master. Merci à Paul qui vit loin sur son île de Mayotte.

Merci à ma famille, strasbourgeoise d'abord pour ces innombrables heures passées ensemble dans la bonne humeur. Merci à Emma d'accepter de débattre tous les jours de banane lors de notre pause-déjeuner. Merci à Paula pour ta joie de vivre et ta bonne humeur permanente qui sont un rayon de soleil pour nous tous. Merci à Robin d'avoir été un porte-bonheur efficace à la Meinau. Merci à Vincent

de nous supporter alors qu'aucun de nous n'est capable de chanter juste. Merci à Marine qui m'a connu et soutenu avant tous les autres. Merci également à tous les autres avec qui j'ai plaisir à trinquer. Merci à Mickael et Miriam d'avoir fait du Blue Moon l'endroit où nous avons plaisir à trinquer, justement.

Merci à mes parents pour votre soutien et vos encouragements tout au long de mes études. Merci Dodo de m'avoir poussé à continuer alors que j'étais prêt à tout lâcher lors de mes débuts à la fac. Merci Fertouille d'être toujours aussi disponible et pour cette passion imbriquable et dévorante que tu m'as transmis. Merci à vous deux d'être la chance qui force ma réussite.

Merci enfin à Adeline de m'avoir pris sous ses ailes et couvé lors des moments les plus difficiles. Merci pour tes attentions petites comme grandes. Merci pour tes petits plats et grands gâteaux qui ont (presque) toujours été bien accueillis par mon palais et mon estomac. Merci pour tout ce que tu fais/es pour moi, cette thèse c'est aussi un peu la tienne.

« *L’important, c’est les trois points !* »

Table of contents

CHAPTER 1: Introduction

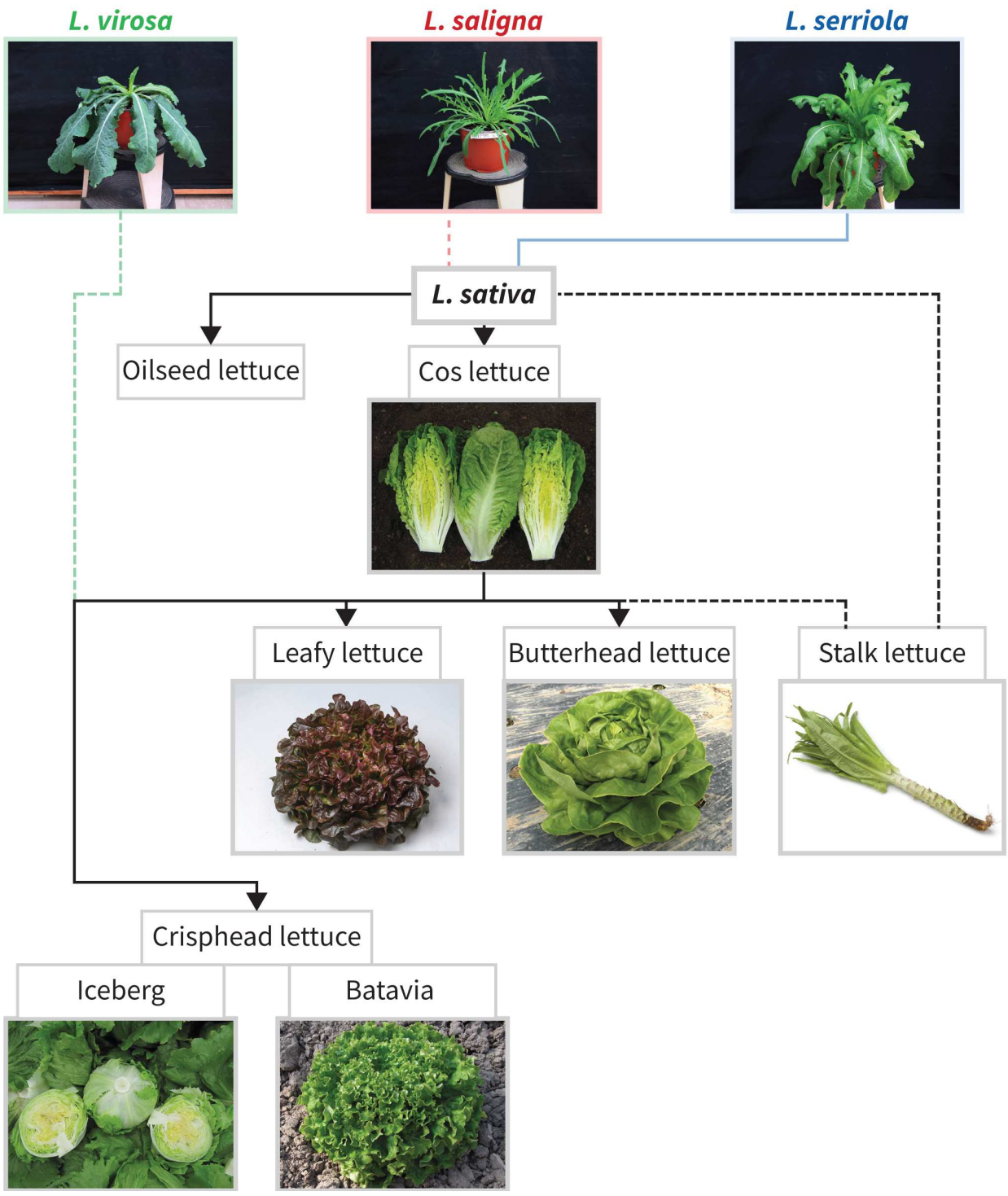


Figure I.1. Relationships among the *Lactuca* genus and descendants of the different typology groups. Within the *Lactuca* genus, it is admitted that there are four main species: three wild ones (*L. saligna*, *L. serriola* and *L. virosa*) and one domesticated (*L. sativa*). When there is documented evidence, relationships between species/cultivar are in full lines, when relationships are hypothetical they are represented in dotted lines. *L. serriola* is certainly one of the direct ancestors of *L. sativa*. However, there is evidence for the introgression of genes from other wild species during domestication. It is not clear if the stalk lettuce derives from the cos type or even earlier. Latin lettuce is not shown because of the lack of evidence regarding its origins. Adapted from de Vries, 1997.

1. Introduction to the *Lactuca* genus

Lactuca spp. are annual diploid (2n=18) plant members of the *Asteraceae* family. The name of the genus "*Lactuca*" derives from the Latin word "*lac*" (milk) due to the milky latex exuding from the plants when touched. Within the *Lactuca* genus which contains about 100 species, four species are well-described and characterized: the cultivated one, *Lactuca sativa*, and three wild species, *Lactuca saligna*, *Lactuca serriola* and *Lactuca virosa,* which can be crossed with *Lactuca sativa* (Ryder, 1999). All four are herbaceous hermaphroditic plants with a life cycle divided into two stages: the vegetative stage (from seed to plant) and the reproductive stage (from plant to seed).

The geographic origin of wild lettuce was presumed to be Egypt (Lindqvist, 1960b), but a more recent publication suggests that wild lettuces originated from South-West Asia, in the region between Egypt and Iran. Indeed, a higher number of wild lettuces can be found in Mesopotamia (de Vries, 1997). *L. saligna* and *L. serriola* are widely spread to every inhabited continent (Lebeda et al., 2004) and they both show prickles along their leaves. *L. saligna* with its almost linear leaves (Lindqvist, 1960b) is also called "least lettuce", while *L. serriola* is named "prickly lettuce", because of its harder prickles in its lobed leaves and stem. *L. virosa*, the "great lettuce", does not share the same geographic origin, and that is why it cannot be found in Asia (Lebeda et al., 2004). This last species shows many variations (leaves can be lobed or not, some with prickles others not) but all have broad leaves (Lindqvist, 1960b).

The domestication of *L. sativa* (**Figure I.1**) is not clearly understood. Indeed, even if *L. serriola* and *L. sativa* can be crossed and hybrids are self-fertile (Thompson et al., 1941; Thompson, 1943) and *L. serriola* is considered as one of the direct ancestor (de Vries, 1990; de Vries and van Raamsdonk, 1994), a polyphyletic origin of *L. sativa* was suggested (Kesseli et al., 1991). Consequently, the implication of *L. saligna* cannot be completely excluded.

Speaking of *L. sativa* involves a specific vocabulary that needs to be defined. *L. sativa* is divided into several morphological types/typologies/groups of cultivars according to their color, leaves shape, size, texture, taste and heading shape. Within each type, breeders developed varieties which can be seen as accessions.

The number of morphological types within *L. sativa* has been subject to debate (Davis et al., 1997; Thicoïpé, 1997), but since the work of Ryder the number of seven typologies resulting from

domestication seems to be accepted (**Figure I.1**) (Ryder, 1999; Křístková et al., 2008). The domestication of lettuce can be seen as a journey that starts with the **"cos lettuce"** (*L. sativa* var. *longifolia*, also called romaine lettuce) between the Euphrates and the Tigris (de Vries, 1997). The first mention of this type goes back to *ca*. 2500 B.C. in Egypt, where lettuce played an important role and was considered an aphrodisiac. Indeed, in Egyptian mythology the fertility and procreation god Min is often represented with erected lettuces, plants from which he draws his own erection (Harlan, 1986). Cos lettuces are tall plants with upright head and grouped leaves. However, it seems that cos lettuce was not the only type cultivated in Egypt, because there is also an indication of the use of **"oilseed lettuces"** (de Vries, 1994), which are not bred for eating but for the oil contained in their large seeds. The next step of this journey takes place on the shores of the Mediterranean sea under the rule of the Roman empire, where **"leaf lettuce"** (var. *acephala* Alef.) derived from cos lettuce appeared in the 1st century B.C.. Leafy types show an open rosette with a wide range of colors (green to red). But domestication not only happened to the west of Mesopotamia but also to the east, in China. Indeed, the **"stalk lettuce"** (var. *angustana*) crop appeared there in the 1st century A.D. (de Vries, 1997), where its swollen stalk is consumed. More recently (CA. 1500-1600), the cabbage lettuce types were derived from cos lettuce (Lindqvist, 1960a), leading to two new types: **"butterhead lettuce"** (var. *capitata* L. *nidus tenerrima*) in Europe and **"crisphead lettuce"** (var. *capitata* L. *salinas*) in North America. The latter can be subdivided into the two typologies "Iceberg" and "Batavia". Both butterheads and crispheads are eaten raw, but butterheads have soft and tender leaves while crispheads have thick and crispy ones. Crispheads also display a robust root system that derives from *L. virosa*. According to Lebeda, the seventh lettuce typology is "**Latin lettuce**", which is a minor type and shares similarities to both romaine and butterhead. All these morphological types have been selected over time, depending on climatic and physiologic factors, as well as from pathogen resistances.

2. The domesticated lettuce: *Lactuca sativa*

The domestication process of *L. sativa* led it to be very different from the wild species. The differences are morphological, like the formation of the head lettuce at the vegetative stage, and a less bitter taste through the reduction of latex production.

2.1. *L. sativa* physiology

Lettuce grows at temperate to cold temperatures, with an optimal growth temperature at 18-20 °C and seed dormancy induced above 28-30 °C. The production of anthocyanins and white latex (depending on cultivars) is enhanced by unfavorable growth conditions, like low temperature (less than 3 °C for germination and less than 7 °C for growth), negatively influencing the taste of the lettuce product. To avoid it lettuces can be grown under different conditions varying from high tech growing systems (hydroponic, heated greenhouses, vertical farming...) to simple growing systems (open-fields, greenhouse tunnel and heated greenhouse). The commercial varieties are adapted to one of these.

The life-cycle of lettuce is a short one, where it self-pollinates and the reproduction is ensured by seeds. The vegetative stage of the life cycle is dominated by the rosette, which is the consumed part of the plant, and ends with the bolting of the stem. Then the stem forms a corymb ramified flower stalk at the beginning of the reproductive stage. The lettuce inflorescence is only opened for 1 to 2 hours in the morning (depending on the climatic conditions and the season), a strong constraint to realize crosses. Moreover, lettuce is hermaphroditic, meaning that all flowers display both male and female reproductive organs. In optimal climatic conditions, when the flower open, the style elongates and the pollen is spread over the stigma. Then the flower closes a few minutes afterward and the ovule is fertilized. The fertilized ovule maturates for two weeks to become the seed. Both stages of its life cycle are useful for marketing: the vegetative stage is used for entire plant marketing and the reproductive stage is used for seed marketing.

2.2. The major typologies

Lettuce typologies differ in color, leaves shape, size, texture, taste and suitability to make a heading shape, meaning it grows into a round-head shape with the leaves growing from the base center. On the market, four typologies stand above the rest (Davis et al., 1997).

2.2.1. Butterhead

This type forms smaller heads than those produced by crisphead types. Because of its smaller size butterheads are lighter (~0.5 kg). Leaves are folded, broad rather than long and crinkled in the interior. They have a light to dark green outside and a creamy yellow inside. The texture is mostly smooth and oily and the taste ranges from bland to relatively sweet (Davis et al., 1997).

2.2.2. Cos lettuce

Cos (or romaine) lettuces have long and loaf-shaped heads, weighing about 0.7 kg. Leaves are longer than broad and spatulated. Outer leaves are normally dark green and inner leaves yellowish. The texture is characterized by its coarseness and crispness. Compared to crisphead types the taste ranges from sweet to strong. Different sub-typologies can be distinguished within Cos lettuces, depending on size and opened/headed top: from mini cos to midi cos, to little gem to romaine (Davis et al., 1997).

2.2.3. Iceberg and Batavia, the crisphead types

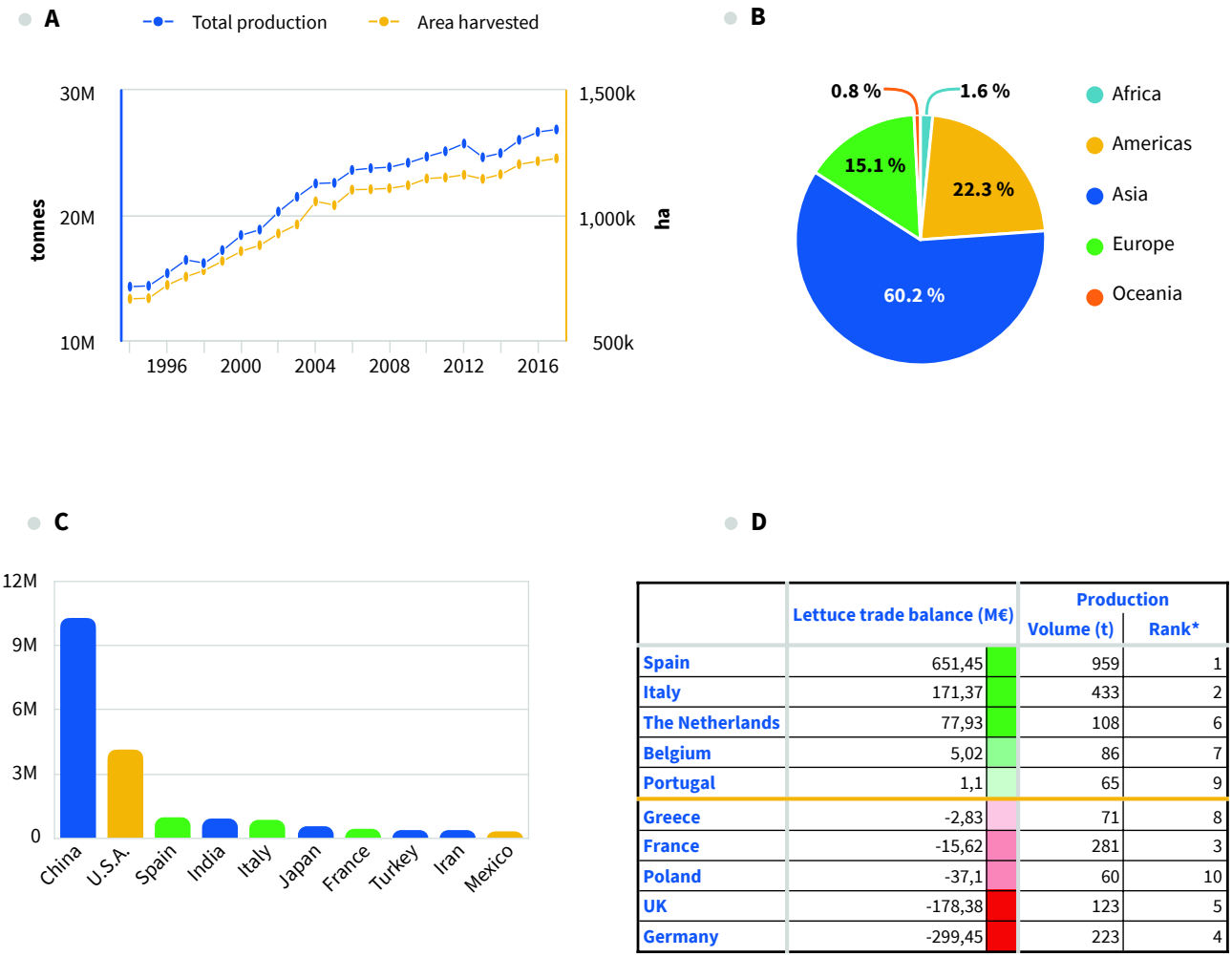
The crispheads are characterized by a crispy texture and a large and firm head. The plant weight is about 1 kg. The leaves are broader than long and dark green (bright or dull dark green), while the inner foliage is either white or creamy yellow. Batavia colors can vary from red, blond and dark green, whereas the iceberg type is more often green and the taste mild (Davis et al., 1997).

2.2.4. Leafy type

Leafy type lettuces show a wide range of sizes, shapes and colors, but the common point is the formation of a leaf rosette that may be long, broad, rounded, spatulated or lobed, with frilled or smooth margins. As butterheads the leafy lettuce weighs up to 0.5 kg. The leaf colors vary from type to type: dark or light green, with or without red color. Leaf lettuce has fewer bleached leaves than crisphead types, which is why they have a stronger taste and higher content in vitamins and minerals. Texture ranges from crisp to soft depending on the type (Davis et al., 1997).

2.3. The salad market: a worldwide context

In a worldwide context, clients can have very different expectations, which the processing and production industries must anticipate. But upstream from these industries breeding companies must create new varieties that offer new characteristics and that are compatible with processing and production industries. Because of its susceptibility to climatic conditions (temperature, photoperiod) the lettuce crop is mainly found in temperate areas, of both the Northern and Southern hemispheres. Moreover, lettuce's fragility in post-harvest stages limits commercial exchanges, which is why lettuce is mostly consumed in the region of production. But, for iceberg lettuce, which is more tolerant to oxidation and so more suitable to transports, exportation between continents is possible (Thicoïpé, 1997).



	Lettuce trade balance (M€)	Production Volume (t)	Production Rank*
Spain	651,45	959	1
Italy	171,37	433	2
The Netherlands	77,93	108	6
Belgium	5,02	86	7
Portugal	1,1	65	9
Greece	-2,83	71	8
France	-15,62	281	3
Poland	-37,1	60	10
UK	-178,38	123	5
Germany	-299,45	223	4

Figure I.2: Production of lettuce and chicory between 1994 and 2017. Data from FAOSTAT (www.fao.org) and UN Comtrade (www.comtrade.un.org). (**A**) Worldwide total production and yield quantities of lettuce and chicory. Between 1994 and 2017 the production of lettuce and chicory increased as well as the difference between production and surface harvested, suggesting that lettuce and chicory cultures have a better yield. (**B**) Mean repartition of the lettuce and chicory production by continent, between 1994 and 2017. The production in this period was led by Asia and North America, followed by Europe. (**C**) Mean production per year for the top 10 producers of lettuce and chicory, between 1994 and 2017. The world production is largely dominated by China, followed by the USA. Together, they produce half of the world's yearly production. (**D**) Lettuce production and exchanges in the European Union. The shades of green represent the strength of the excess of the lettuce balance trade while the shades of red stand for the strength of its deficit. * rank within the EU.

2.3.1. A worldwide market

The domestication process led lettuce to be consumed in all continents, due to agronomic and taste appeals. This worldwide consumption was accompanied by increased production. Indeed, according to the Food and Agriculture Organization (FAO), lettuce production and the related chicory (*Cichorium intybus* var. *foliosum*, *C. endivia* var. *crispa* and *C. endivia* var. *latifolia*) increased yearly from 14 Million metric tons in 1994 to 26 Million metric tons in 2017 (**Figure I.2.A**). The production is concentrated in Asia and America (**Figure I.2.B**) with, obviously, China and the USA as principal producers (**Figure I.2.C**). As an example of the economic value of lettuce the US production was worth 1.77 billion dollars in 2017, as reported by the U.S. Department of Agriculture (USDA). Despite its fragility, lettuce is exported. The USA is one of the main exporting countries with 300 000 metric tons per year, according to the U.S. Department of Commerce (data for 2018), most of it to Canada (85 %) but also to longer-range destinations like Honk-Kong (5 %) or the United Kingdom (UK). But, according to the U.S. Department of Commerce, the USA mainly ships iceberg crisphead types.

2.3.2. The European market

The European Union (EU) is both a very important production area and a major market for consumption. Indeed, according to UN Comtrade, the European market accounts for 62 % of the worldwide exports, while 68 % worldwide imports are made by EU members, meaning that 91 % of EU imports come from an EU member. This 91 % amounts to 1.37 billion euros per year. In 2018, the total EU lettuce production was about 2.5 million metric tons and the main producers were Spain, Italy and France, as reported by Eurostat (**Figure I.2.D**). However, it is possible to distinguish two groups of countries within the EU (**Figure I.2.D**): the countries with a positive “lettuce balance of trade” (Spain, Italy, the Netherlands, Belgium and Portugal) and the ones with a negative balance (Germany, UK, Poland, France and Greece). The lettuce production is concentrated in the Mediterranean areas, which is why Spain and Italy are the principal producers, with 959 000 and 433 000 metric tons in 2018, respectively. Moreover, Spain is the major exporter worldwide and bulks 28 % of the worldwide lettuce exports, mainly iceberg type to the UK (40 %) and to Germany (30 %). Italy produces mainly Iceberg for its exports to Germany (40 %), Austria (20 %) and the UK (12 %). France is the third lettuce producer within the EU in 2018 and the seventh world producer (**Figure I..2.C**). However, France exports yearly 121 million euros of lettuce while importing for 137 million euros, and is therefore considered as an importing country, according to UN Comtrade. French clients are Germany (60 % of the exports) and the UK (mostly iceberg type from Brittany).

Germany is the fourth producer, but also the first importer, mainly Iceberg types from Spain, Butterhead and Batavia from France and Belgium. Despite its large lettuce production, the German "lettuce balance of trade" is very unprofitable (**Figure I.2.D**). The Netherlands is the fifth producer, exporting 80 % of their lettuce production. Most of it goes, again, to Germany (65 %) and to the UK (8 %).

2.4. Current academic research hotspots

Due to the important economic role of *L. sativa*, studies on lettuce mainly focused on that domesticated species. Those studies include stress tolerance; as to cadmium (Bautista et al., 2013; Dawuda et al., 2019; Wang et al., 2019) or to salt (Xu and Mou, 2015; Di Mola et al., 2017; Hniličková et al., 2019); or resistance to bacterial pathogens (Hayes et al., 2014; Nicolas et al., 2018, 2019). But the most investigated field in lettuce research is the downy mildew caused by the *oomycota* pathogen *Bremia lactucae* (Giesbers et al., 2018; Missio et al., 2018; Dhar et al., 2020; Fletcher et al., 2019; Meisrimler et al., 2019; Pelgrom et al., 2019). Indeed, according to the International Bremia Evaluation Board (IBEB), *B. lactucae* is the major threat to lettuce production. The IBEB is a joint initiative of lettuce breeding companies from the USA, France and the Netherlands, the University of California Davis (UC Davis), the Dutch inspection service (Naktuinbouw) and the French National Seed Station (GEVES). Its mission is to identify new races of that pathogen that could be major threats to the American and European lettuce industries. Moreover, *B. lactucae* was ranked 13th oomycete pathogen in plants, based on its scientific and economic importance (Kamoun et al., 2015).

From the genomic point of view, the chloroplast genome (or cpDNA) was the first of the genomes of *L. sativa* to be assembled assembled and released (Kanamoto et al., 2006). The nuclear genome of is about 2.38 Gb and composed of 9 pairs of chromosomes (Michaelson et al., 1991; Koopman and De Jong, 1996), and its sequence was released 11 years after the chloroplastic one (Reyes-Chin-Wo et al., 2017). Its annotation was made possible thanks to the previous building of its transcriptome (Truco et al., 2013). Just recently, the mitochondrial genomes (or mtDNA) of *L. sativa* (363 kb), *L. saligna* (368 kb) and *L. serriola* (363 kb) were published (Kozik et al., 2019).

3. The mitochondria

Mitochondria are organelles of about 0.2–1.5 µm present in the cytosol of almost every eukaryotic cell (Roger et al., 1998; Karnkowska et al., 2016). Their main cellular function is the production of adenosine triphosphate (ATP) by oxidative phosphorylation, used by the cell as a source of energy, but mitochondria are also involved in signaling, cellular differentiation and cell death. In plants and more largely in eukaryotes, maternal inheritance is the predominant pattern of mitochondria transmission (Birky, 2001). However, in several plant species inheritance can be paternal, like in bananas (Fauré et al., 1994), cucumber (Havey et al., 2004) and the green alga *Chlamydomonas* (Aoyama et al., 2006), or biparental like in *Geranium* (Guo and Hu, 1995) and the Chinese pine (Guo et al., 2005).

3.1. A brief history of mitochondria discovery

The first observation of what could be mitochondria goes back to the 1840s, before they were identified by Richard Altmann in 1890. Altmann named them “bioblast”, an elementary organism that lives within the cell and carries vital function. Indeed, he discovered that his bioblasts display the same staining as bacteria. The name “mitochondria” was coined by Carl Benda in 1898 and comes from the Greek “*mitos*” (thread) and “*chondros*” (granule). Until the development of high-resolution electron microscopy, mitochondria were studied by staining with the redox-dye Janus Green B, thanks to the works of Leonor Michaelis in 1900. Even if mitochondria were known in animals since 1890, they were first described in plants in 1904 by Meves in *Nymphaea alba* (white nenuphar). Mitochondria isolation started in 1934 with the works of Bensley and Hoerr (Bensley and Hoerr, 1934), improved in 1946 by Albert Claude (Claude, 1946) and in 1948 by Hogeboom, Schneider, and Palade with the use of an isotonic sucrose medium.

Purified mitochondria set the start of biochemical approaches to elucidate its functions. In 1946, the exclusive mitochondrial localization of cytochrome oxidase activity was demonstrated (Hogeboom et al., 1946). The aerobic oxidation of citric acid cycle metabolites and of fatty acids accompanying the formation of ATP from inorganic phosphate and adenosine diphosphate (ADP) was elucidated by Eugene Kennedy and Albert Lehninger (Kennedy and Lehninger, 1949). In the same year, the coupling between the oxidation of nicotinamide adenine dinucleotide (NADH) and the phosphorylation of ADP was shown (Friedkin and Lehninger, 1949). The control of respiration by the availability of inorganic phosphate and ADP was demonstrated by several laboratories on

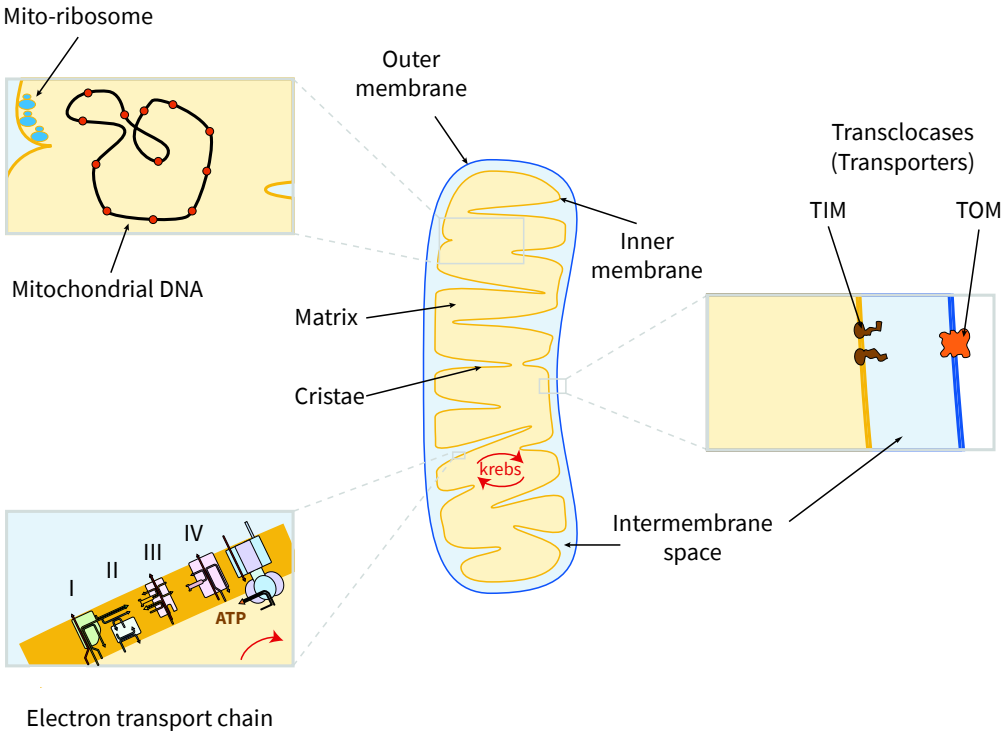


Figure I.3. Schematic view of the mitochondrial internal structure. The mitochondria is surrounded by two lipid bilayers: the Outer membrane (OM) and the Inner membrane (IM), delimiting the intermembrane space and the matrix. The import of proteins from the cell cytosol into the mitochondrial matrix is assured by the TOM (translocase of the outer membrane) and TIM (translocase of the inner membrane) translocases. The folds of the IM in the matrix are called cristae and are separated from the intermembrane space by junctions (not on the scheme). The cristae are the seat of the electron transport chain, which is why the pH is acid there, because of the proton pumps of complexes **I**, **II** and **IV** of the electron transport chain. The mitochondrial genome “swims” in the matrix but is probably attached to the inner surface of the IM and it is associated with proteins involved in its structuring and expression, together forming the nucleoids.

the period 1951-1952 (Rabinovitz et al., 1951; Lardy and Wellman, 1952). The discovery of the mitochondrial localization of the respiratory chain and ATP production led Philip Siekevitz to nickname it “the powerhouse of the cell” in 1957 (Ernster and Schatz, 1981).

3.2. Origin and evolution: the endosymbiotic theory

The endosymbiotic theory was first proposed by Mereschkowsky in 1905 to explain the origin of chloroplasts. However, he firmly rejected the possibility that mitochondria might have evolved from an endosymbiosis event. Mitochondria were linked 81 years later to that endosymbiotic theory by Lynn Margulis (born Sagan, Sagan, 1986). The theory proposed that a free-living organism, or proto-mitochondria, survived endocytosis and was incorporated into the cytoplasm. Then the ability of the proto-mitochondria to perform respiration within the cytoplasm gave a considerable evolutionary advantage to the host cell. Nowadays it is commonly accepted that mitochondria have originated through endosymbiosis and believed that the proto-mitochondrial ancestor was an aerobic *Alphaproteobacteria* related to the obligatory parasites of the *Rickettsiale* branch (Andersson et al., 1998). However, the identification of the closest relative to the proto-mitochondrial ancestor remains controversial (Thrash et al., 2011), as evidenced by a recent study suggesting that mitochondria evolved from a lineage that branched off before the divergence of *Alphaproteobacteria* (Martijn et al., 2018).

3.3. The internal structure of the mitochondria

Because of its endosymbiotic origin mitochondria displays two membranes: the outer mitochondrial membrane (OMM) and the inner mitochondrial membrane (IMM) (**Figure I.3**). The OMM is porous and allows exchanges between the cytosol and the inner membrane. To do so it contains three membrane protein families that form channels as part of larger protein complexes. Two of these three complexes, the TOM (translocase of the outer membrane) pore and the SAM (sorting and assembly machinery) insertase allow the translocation and insertion of almost all proteins targeted to the mitochondria. The third complex is the mitochondrial porin (or VDAC, for voltage-dependent anion channel) that allows ions and small uncharged molecules to go through the membrane. The IMM works slightly differently, it is a tight diffusion barrier to ions and molecules that cannot cross this membrane without help from specific transporters. A membrane potential of about 180 mV results from its selectivity for ions. The inner membrane is the home of oxidative phosphorylation.

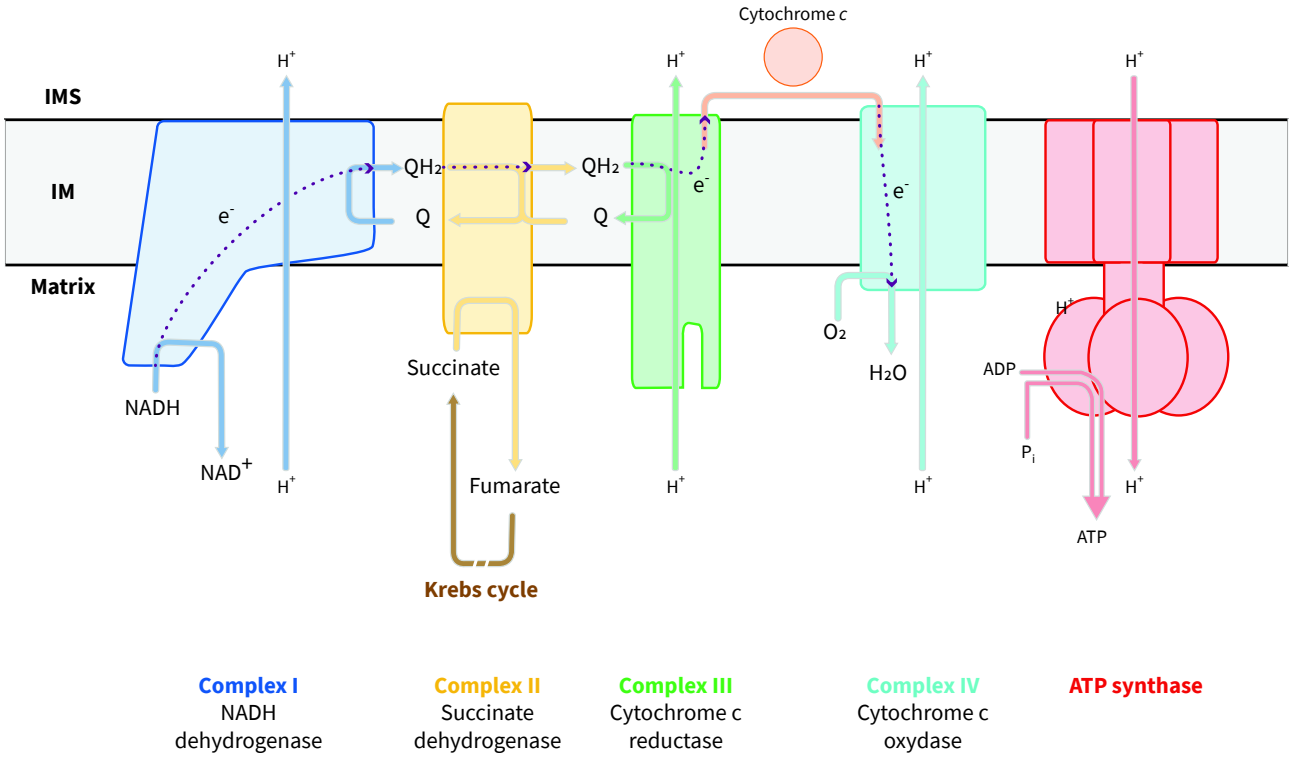


Figure I.4. Schematic view of the mitochondrial electron transport chain (mtETC). The proton pumps from complexes **I**, **III** and **IV** set a proton gradient through the inner membrane. The ATP synthase uses this gradient to produce ATP from ADP and inorganic phosphate. The journey of the electron through the mtETC is represented by purple dotted lines. IMS: Intermembrane space; IM: Inner membrane; Q: ubiquinone; Qh2: ubiquinol. Adapted from Sousa et al., 2018.

These two membranes define three compartments, each one with its distinct role: the intermembrane space between the OMM and the part of the IMM that is known as the inner boundary membrane, the mitochondrial matrix surrounded by the IMM and the cristae which are invaginations of the IMM into the matrix. The intermembrane space is a place where molecules are passing that is why carriers can be found as well as the ATP/ADP ones. The mitochondrial matrix has a pH of about 8 (Llopis et al., 1998) that drives the transmembrane electrochemical gradient needed for ATP synthesis. The matrix contains the citric acid cycle, or Krebs cycle, among other enzymes, as well as the mtDNA and ribosomes, and is the site of mtDNA replication, transcription and of protein translation. Finally, the cristae contain the complexes of the mitochondrial electron transport chain (mtETC), the full ATP synthase and a large amount of cytochrome *c*.

3.4. The electron transport chain

In the plant cell, energy production occurs in two organelles: the chloroplast and the mitochondria. Both rely on an electron transport chain which sets a proton gradient across the membrane used for the phosphorylation of ADP into ATP (Sousa et al., 2018). In chloroplasts, the energy production process is photosynthesis, which involves the conversion of water and CO_2 into complex organic molecules. It can be split on two reactions: the one that takes place in the light and the one that takes place in the dark (Johnson, 2016). The light-dependent reaction occurs in the thylakoid membranes and using the energy of light splits the water into oxygen, and the electrons freed up from H_2O are transferred to ATP and NADPH. The dark reaction occurs in the chloroplast stroma, where the energy released from ATP and NADPH is used to reduce CO_2 to carbohydrates, by the enzymes of the Calvin cycle (Johnson, 2016).

In mitochondria, the energy is produced during oxidative phosphorylation which relies on the oxidation of NADH through the mtETC located in the cristae of the IMM (Sousa et al., 2018). To do so, the chemical energy stored in the products of the citric acid cycle is converted into an electrochemical gradient over the inner membrane by a series of proton pumps which are three of the four complexes of the mtETC (**Figure I.4**). The **complex I** (or NADH ubiquinone oxidoreductase) is one of the three proton pumps and its role is oxidizing NADH into NAD^+, which is used by the Krebs cycle, a proton that is pumped out of the matrix into the intermembrane space and an electron that is used by **complex II** to reduce ubiquinone to ubiquinol. **Complex II,** or succinate dehydrogenase, is the only one of the four mtETC complexes that does not pump out protons. Besides its role in the mtETC **complex II** is also part of the Krebs cycle, where it catalyzes the oxidation of succinate to

fumarate. The ubiquinol is then used by the **complex III** to transfer electrons to cytochrome *c*. **Complex III** or cytochrome c reductase is the second proton pump of the mtETC. Finally, the electrons obtained from the reduction of cytochrome *c* are transferred by **complex IV**, or cytochrome c oxidase, to a molecule of oxygen that is the terminal electron acceptor and form a molecule of water. **Complex IV** is the third and last proton pump of the mtETC but also the less efficient. Indeed, **complexes I** and III pump out 4 protons per pair of electrons crossing the complexes, while **complex IV** pumps out 2 protons. The proton gradient thus established drives the membrane-embedded rotor of the ATP synthase (F_0 domain) and ATP is produced by conformational changes in the nucleotide-binding pockets of the F_1 part of the synthase (Sousa et al., 2018).

3.5. Organization of the mitochondrial population within the cell

The organization of the mitochondrial population, or chondriome, within the cytosol depends on the organisms. The chondriome can be seen as a dynamic structure where mitochondria are highly mobile and can merge or divide by fusion and fission. In mammals the chondriome is organized in extended reticular webs (Koning et al., 1993). The number of mitochondria in yeast chondriome varies according to the type of the media, depending on the energetic pathway required. Indeed, when respiration is needed like during growth on ethanol yeast cells carry around 30 small mitochondria, while on fermentation conditions on a glucose-rich medium they display 2 to 3 branched filamentous mitochondria (Visser et al., 1995; Raja et al., 2013). In mammals the number of mitochondria depends on the cell type studied (Veltri et al., 1990), but was estimated at 100 in human ovarian cancer cell line A2780 (Satoh and Kuroiwa, 1991). The plant chondriome is composed of a discrete and pleomorphic population where mitochondria display spherical to tubular forms. A mesophyll cell of *Arabidopsis thaliana* carries between 200 and 300 mitochondria per cell (Logan, 2006), but this number can be higher in other cell types, as in protoplasts prepared from rosette leaves where it is comprised between 300 and 400 (Preuten et al., 2010), or in other species, as it has been evaluated to be between 500 and 600 in mesophyll protoplasts of *Nicotiana tabaccum* (Sheahan et al., 2004).

4. The mitochondrial genomes

4.1. The mtDNA within mitochondria

The mtDNA, located in the matrix, is not free. Indeed, it is compacted and associated with other proteins as the genome of *Escherichia coli* that is 10^4-fold compacted (de Vries, 2010). This was shown in human mitochondria where the genome is located in a structure called nucleoid (Satoh and Kuroiwa, 1991) distributed all over the reticular web of mitochondria (Alam et al., 2003) and fold to ~100 nm (Kukat et al., 2011). The copy number of the mtDNA within the nucleoid seems to depend on the cell type. Indeed, it was shown that in an epithelial ovarian carcinoma cell line (A2780) nucleoids contain a single mtDNA copy (Kukat et al., 2011), unlike HeLa and 143B cell lines containing 6 and 8 mtDNA copies per nucleoid, respectively (Legros et al., 2004). Moreover, this publication shows that the number of nucleoids per cell is greater than 450 in both cell types, and that in animals, or at least in humans, mtDNA copies outnumber mitochondria. Nucleoid packaging requires several associated proteins organized into two groups: the core nucleoid factors and the peripheral factors. The core nucleoid factors are DNA binding proteins involved in transcription and/or replication, such as mitochondrial transcription factor A (TFAM) that is involved in transcription (Fisher and Clayton) by inducing a DNA U-turn which promotes transcription initiation (Ngo et al., 2011; Rubio-Cosials et al., 2011). Moreover, TFAM is also considered to be involved in the packaging of the mtDNA (Kaufman et al., 2007) and it seems to be the central factor in the structuring of the human mitochondrial genome (Kukat et al., 2015). The human core nucleoid also includes factors necessary for genome replication and transcription, as single-stranded binding protein mtSSB, the DNA polymerase POLγ, its accessory subunit POLG2, the replicative DNA helicase TWINKLE, the mitochondrial DNA-directed RNA polymerase POLRMT and transcription factor TFBM2. Peripheral factors are not directly associated with the mtDNA but through protein-protein interactions and seem to play a role in modulating ATP production by regulating the organization and metabolism of the mtDNA, such as protein M19 (Sumitani et al., 2009), or in stabilizing core nucleoid factors, such as prohibitin (PHB) that stabilizes TFAM (Kasashima et al., 2008). The membrane-bound ATPase ATAD3 (ATPase family AAA domain-containing protein 3) belongs to both groups: it is a DNA binding protein that can be associated with nucleoids, but it also has additional functions, such as mitochondrial-endoplasmic reticulum interactions (He et al., 2007).

Category	Name	AGI	Copies per mtDNA(*)	Human homologue
DNA Polymerase	POL1A	At1g50840	5	POLγ(**)
	POL1B	At3g20540	2	
Helicase/primase	TWINKLE	At1g30680	192 (32)	TWINKLE
Topoisomerase I	TOPI	At4g31210	61	-
Topoisomerase II	GYRA	At3g10690	981 (491)	-
	GYRB	At5g04130	676 (338)	
Recombinase	RECA2	At2g19490	462	-
	RECA3	At3g10140	?	
DNA-binding	ODB1	At1g71310	673 (61)	RAD52
	SWIB5	At1g31760	141	(RAD51?)
ssDNA-binding	OSB1	At1g47720	4	mtSSB
	OSB3	At5g44785	171	
	OSB4	At1g31010	180	
	SSB1	At4g11060	493 (123)	
	SSB2	At3g18580	657 (164)	
	WHY2	At1g71260	974 (41)	
MutS-like	MSH1	At3g24320	2(1)	-
DNA ligase	LIG1	At1g08130	42	LIGIII*
HMG-box protein	-	-		TFAM
Dimethyl transferase-like TF	-	-		TFB2M
RNA polymerase	RPOTm	At1g68990	5	POLRMT
	RPOTmp	At5g15700	31	

Table I.1. Plant mitochondrial nucleoid factors. * Data from Fuchs et al., 2020. For proteins known or assumed to be present in homo- or hetero-oligomers the inferred number of complexes is shown in bracket. ** Functional homologues, not evolutionary linked.

The composition of plant mitochondrial nucleoid is not yet fully resolved, but a set of factors involved in the expression, maintenance, and segregation of the organelle genomes have been associated with this complex (Chevigny et al., 2020) (**Table I.1**). Among the factors identified in plants there is single-stranded DNA (ssDNA) binding proteins SSB1 and SSB2 (Elo et al., 2003; Edmondson et al., 2005), orthologs to human mtSSB. The plant mitochondrial DNA polymerases are POL1A and POL1B that are functionally homologous to the human POLγ, but not evolutionary related to it (Parent et al., 2011), but to bacterial DNA polymerase I. Mitochondrial transcription is mediated by the phage-type DNA-directed RNA polymerases RPOTm and RPOTmp (Hedtke et al., 1997; Yadav et al., 2019). The replicative DNA helicase in plants is the ortholog of human TWINKLE (Diray-Arce et al., 2013). The DNA tangles and supercoiling are managed by the type II topoisomerase Gyrase, and by the type I topoisomerase TOP1 (Wall et al., 2004; Cho et al., 2004). The RecA-like recombinases RECA2 and RECA3 (Shedge et al., 2007) are also probably part of the mitochondrial nucleoid, as well as the MutS-like homolog MSH1 (Xu et al., 2011) and the plant-specific ssDNA-binding proteins of the OSB, WHIRLY and ODB families (Zaegel et al., 2006; Cappadocia et al., 2010; Janicka et al., 2012; Chevigny et al., 2020). But all do not have the same stoichiometry (**Table I.1**). Indeed, quantification of mitochondrial proteins absolute copy numbers showed that the ssDNA binding factors and the gyrase are present at high-copy numbers, while the transcriptional and replicative enzymes are present at a much lower level (Fuchs et al., 2020).

In a review, Gualberto and Kühn (Gualberto and Kühn, 2014) listed the factors that they considered as members of the mitochondrial nucleoid of *A. thaliana*. But none of the so important human TFAM or ATAD3 seems to have a homolog in plants. Furthermore, the DNA composition of the plant nucleoid remains unclear. Indeed, the plant mtDNA is considered as constituted by a group of subgenomic molecules, instead of a single circular genome as in animals, raising the question of the DNA composition of the nucleoid. Moreover, it has been shown that vegetative plant tissue contains less than one mitochondrial genome per mitochondria, and that during division not all mitochondria inherit a copy of the mtDNA (Preuten et al., 2010; Arimura et al., 2004).

The regulation of plant mtDNA copy numbers might differ according to the type of cell and developmental state. In *A. thaliana* it varies between organs and is not constant during the development of cotyledons and leaves (Preuten et al., 2010). In *Oryza sativa* (rice), a much higher mtDNA copy number was found in egg cells as compared to leaf mesophyll protoplasts (Takanashi et al., 2010). However, such a variation was not found in Arabidopsis and tobacco, where equivalent

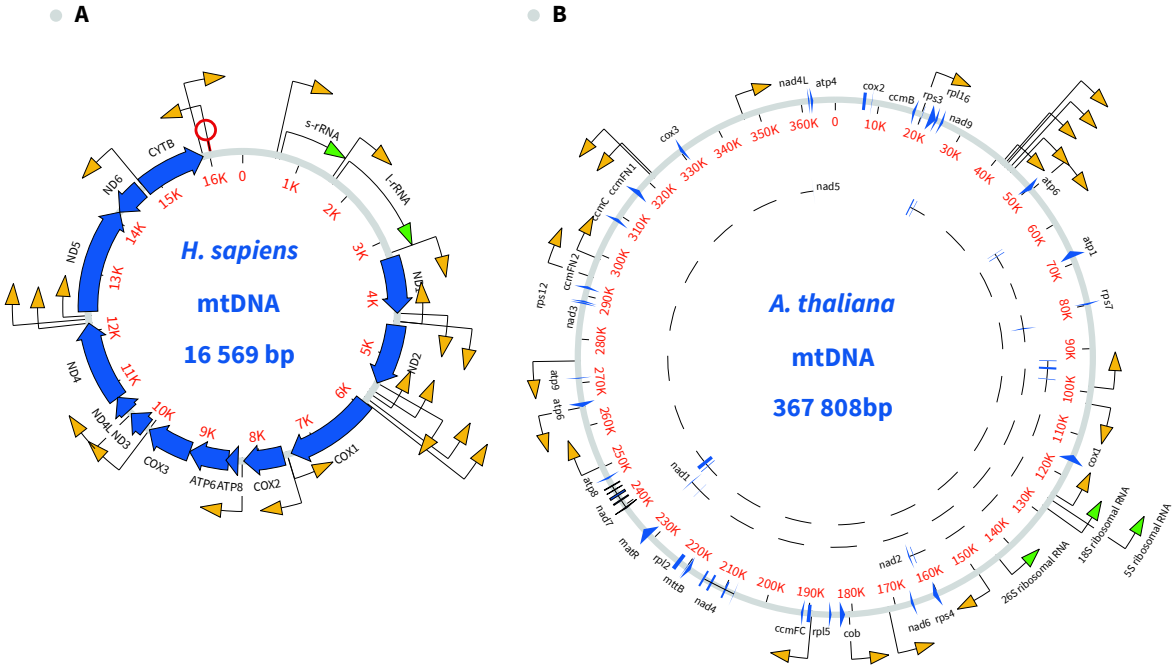


Figure I.5. Gene content in mitochondrial genomes. (**A**) In *Homo sapiens* (human) the 13 protein-coding genes (■▶), the 2 rRNA genes (➤) and the 22 tRNA genes (➤) carried by the mtDNA represent the major part of the 16.5 kb mtDNA. The red structure is the D-loop of the human mtDNA which is the start of its replication (Anderson et al., 1981). That is the opposite of the large mtDNA of *Arabidopsis thaliana* (**B**) and its 367 kb where the 32 protein-coding genes (▶), 22 tRNA genes (➤) and 3 rRNA genes (➤) represent only 26 % of the genome. The trans-spliced transcripts are show in dashed lines (Unseld et al., 1997; Sloan et al., 2018).

mtDNA copy numbers were determined in egg cells and in mesophyll cells (Wang et al., 2010). The authors of that study also found that egg cells of *Pelargonium zonal* own a very large content of mtDNA, implying that the mtDNA can be selectively amplified in the generative cells of some species (Wang et al., 2010). In contrast, on the other side of the plant life cycle, a 50-fold-mtDNA reduction was observed during pollen development (Sodmergen et al., 2002). This dramatic decrease would be because of selective degradation and could be a mean for rapid access to phosphate in case of nutrient-deficient conditions (Takami et al., 2018). Several nucleases have been identified that could be involved in that process (Tang and Sakamoto, 2011).

4.2. Size and content

Over the course of evolution, the endosymbiotic relationship between the host cell and the proto-mitochondria shaped the bacterial genome which would become the mtDNA. This process led the mtDNA to show a high size diversity across kingdoms and species. Indeed, even if the size of the mtDNA is quite conserved within mammals; 16 kb in humans (**Figure I.5.A**) (Anderson et al., 1981) and rats (Rat Genome Sequencing Project Consortium, 2004) or 17 kb in rabbits (Gissi et al., 1998); it varies substantially in size in plants; from 200 kb in rapeseed (Handa, 2003), 368 kb in Arabidopsis (**Figure I.5.B**) (Unseld et al., 1997; Sloan et al., 2018) and up to the several millions base pairs (11.3 Mb) in the striped corn catchfly, *Silene conica* (Sloan et al., 2012); or in fungi mtDNA; from 18.8 kb in *Hanseniaspora uvarum* (Pramateftaki et al., 2006) to 105 kb in *Nakaseomyces bacillisporus* (Bouchier et al., 2009) via 85 kb in *Saccharomyces cerevisiae* (Foury et al., 1998). All the listed genomes (excepted *Silene* sp. ones) have in common to be smaller than bacterial genomes. Indeed, an *Alphaproteobacteria* such as *Mesorhizobium loti* has a 7 Mb genome (Kaneko, 2000) which is far from the 16 kb of the human mtDNA. Even if a drastic genome size reduction can be observed in parasites, such as *Rickettsia prowazekii* and *Mycoplasma genitalium* and their 1.1 Mb and 580 kb genomes, respectively (Andersson et al., 1998; Su and Baseman, 1990), or the eukaryotic parasite *Encephalitozoon cuniculi* and its 2.9 Mb nuclear genome (Katinka et al., 2001), the mechanisms of genome reduction are different (Timmis et al., 2004). Whereas genome reduction in a parasite is a specialization to the nutrient-rich intracellular medium of his host and the loss of genes that are no longer needed, the reduction in size of the mtDNA during evolution was, in addition to the loss of redundant genes, by export of genetic materials to the nucleus, and the proteins that are still required for mitochondrial function are synthesized in the cytosol and imported into mitochondria.

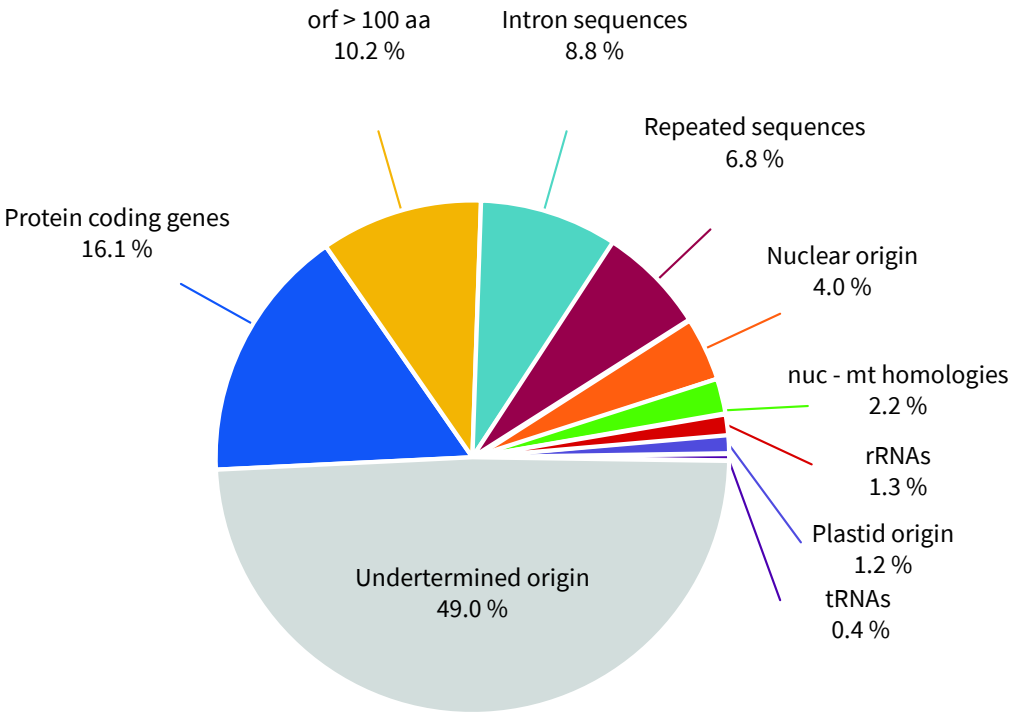


Figure I.6. Arabidopsis mtDNA composition. The origin of almost half of the sequences of the *A. thaliana* mtDNA could not be determined. Repeated sequences can be separated between 2.9 % of large repeated sequences (>500 bp) involved in frequent reciprocal recombination, responsible for the dynamic structure of the mtDNA, and 3.9 % of smaller repeated sequences. Homologies between nuclear (nuc) and mitochondrial (mt) genomes (■) are probably the result of the transfer of sequences from the nucleus into the mtDNA. Adapted from Marienfeld et al., 1999.

That is why, concerning the mtDNA, size doesn't always matter. Indeed, the mtDNA carries a variable number of protein-coding genes that is not correlated to its size. However, these genes are always involved in the mitochondrial electron transfer chain or in the mitochondrial gene expression system. For instance (**Figure I.5**), the large mtDNA of *A. thaliana* carries a total of 57 genes (32 protein-coding genes, 22 tRNA genes and 3 rRNA genes: 5S, 18S and 26S) (Unseld et al., 1997) which is only 20 genes more than the human mtDNA gene content (37 genes: 13 protein-coding genes, 22 tRNA genes and 2 rRNA genes: 12S and 16S) (Anderson et al., 1981) for a genome that is 23-fold larger. In fact, the plant mtDNA is mainly composed of non-coding sequences, like in Arabidopsis (**Figure I.6**) where the 57 mitochondrial genes and their introns cover only 26.6 % of the genome. In potato mitochondria, the transcribed regions corresponding to mature transcripts account to only 25 % of the genome (Varré et al., 2019). Among the 73 % non-coding sequences of the *A. thaliana* mtDNA 6.8 % are repeated regions, 4 % are of nuclear origin and 1.2 % were imported from the chloroplast. Finally, more than half of the mtDNA of *A. thaliana* is composed of sequences from unknown origin (Marienfeld et al., 1999) that might have been inherited through horizontal transfer (Bergthorsson et al., 2003). For larger mtDNA genomes, such as the ones of *S. conica* which do not encode for more factors, this proportion is even higher. Flower plant mtDNA-encoded protein genes can be divided into two groups: the 24 core genes that are present in all plant mitochondrial genomes (*atp1*, *atp4*, *atp6*, *atp8*, *atp9*, *ccmB*, *ccmC*, *ccmFc*, *ccmFn*, *cob*, *cox1*, *cox2*, *cox3*, *matR*, *mttB*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7*, and *nad9*) and ± 17 variable protein genes, which mostly code for subunits of the mitoribosome. The three rRNA genes 5S, 18S and 26S are also present in all plant mtDNA genomes, but the number of tRNA genes are variable, with several tRNA species that are imported from the cytosol and others that are expressed from chloroplast sequences inserted in the mtDNA (Marechal-Drouard et al., 1990; Skippington et al., 2015).

The plant mitochondrial genomes are characterized by their high proportion of repeated sequences. Small repeated sequences (< 35 bp) can also be found in yeast or animal mtDNA and can be involved in large scale deletion linked to diseases in humans (Schon et al., 1989; Phadnis et al., 2005). In plants, however, repeated sequences can go up to several tens of kilobase-pairs, and are not conserved in sequence or number (Wynn and Christensen, 2019). That is why they have been classified into three categories; large repeated sequences (≥ 500bp), intermediate-size repeats (50 - 500 bp) and microhomologies (< 50 bp). These sequences can be involved in homologous recombination events and consequently have a major impact on the structure of the mtDNA.

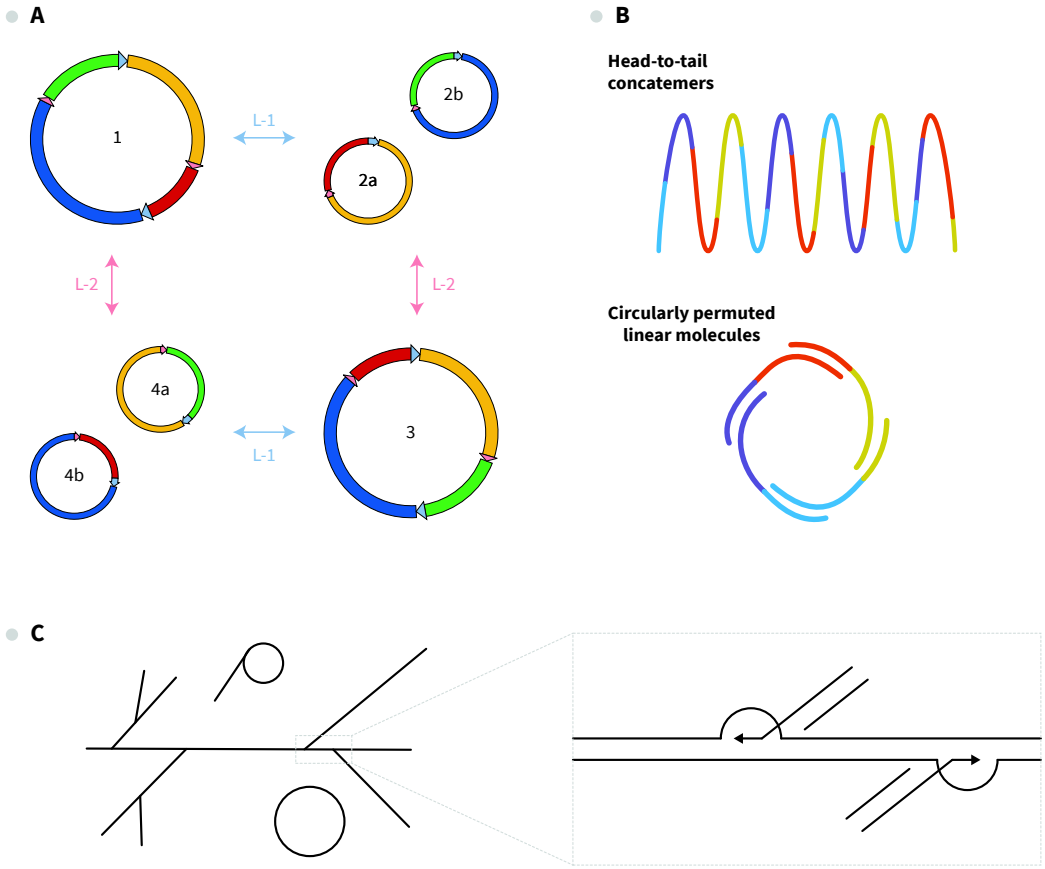


Figure I.7. Plant mtDNA structure. (**A**) The mtDNA of *A. thaliana* ecotype Col-0 carries two large repeated regions (L-1 and L-2) that are involved in frequent homologous recombination (in blue when L-1 involved or in pink when it is L-2). These recombination events lead to four different possible configurations of the mtDNA, two of them (2 and 4) subdivided in subgenomes (a and b). The different mtDNA molecules are represented in circular form, for convenience. (**B**) The plant mtDNA can exist in alternative linear structures that can produce circular genome maps through repeated regions. (**C**) Electron microscopy observations of mtDNA led to consider it as constituted by a mixture of branched, linear or circular molecules. Branched molecules might be the consequence of homologous recombination events. Adapted from Sloan et al., 2013 and Gualberto et al., 2014.

4.3. Structure of the plant mtDNA

The large repeats are involved in frequent and reciprocal homologous recombination events that shape the mtDNA, redistribute the sequences and generate sub-genomic mtDNA molecules (**Figure I.7.A**). The plasticity of the mtDNA by recombination makes it difficult to model the structure of the mtDNA, which is usually represented as a single circular structure called "master circle" because it is usually possible to map the genome into a circle. However, such genome-size circular structures have only been observed in very few species, in relatively old publications (Negruk et al., 1986; Oda et al., 1992) that were not reproduced (Oldenburg and Bendich, 1998). In fact, pulse-field gel electrophoresis (PFGE) and microscopy studies have shown that the plant mtDNA consists of a mixture of head-to-tail concatemers, circularly permuted linear molecules (**Figure I.7.B**) or branched molecules (**Figure I.7.C**) (Sloan, 2013). Indeed, electron microscopy showed that most large molecules observed are linear or rosette-like, while most of the circular molecules are small (Backert et al., 1996, 1997b), and that an important proportion of the mtDNA was found to be single-stranded (Backert et al., 1997a). Structural analysis using moving pictures and PFGE go in the same direction, suggesting that the plant mtDNA mainly exits as large and complex branched molecules or simple linear ones derived from the complex forms (Bendich, 1996). Recently, a study about the lettuce mtDNA confirmed these results and has shown that the lettuce mtDNA mostly consists of branched molecules, but that circular, linear, and branched circular structures can be observed as well (Kozik et al., 2019). It is generally accepted that the mitochondrial genome of plants consists of a collection of subgenomes produced by recombination between direct repeated sequences within one or several chromosomes. Another model could be of an mtDNA constituted by circularly permuted linear molecules containing overlapping ends (**Figure I.7.B**), allowing *in silico* mapping of these molecules one over the other into a circle, even if such a structure does not exist *in vivo* (Sloan, 2013). Interestingly, the concatemeric mitochondrial genome inserted in the nuclear genome of *A. thaliana* (Stupar et al., 2001) suggests a model where the genome is represented by concatemeric linear molecules (**Figure I.7.B**). Most of the models described suppose linear molecules, of which the linear ends need to be protected. This has been shown in the S-type cytoplasmic male sterility (**CMS-S**) genotype of maize. This particular cytoplasm carries two supplementary linear molecules S-1 and S-2 of 6 397 bp and 5 452 bp, respectively, of which the end of both are protected by attached proteins (Paillard et al., 1985; Levings and Sederoff, 1983; Kemble and Thompson, 1982). But those proteins have not been identified.

In contrast to the recombination involving large repeats, the ectopic recombination between intermediate-size repeats and microhomologies is infrequent, usually asymmetric, and not reversible (Woloszynska and Trojanowski, 2009; Janicka et al., 2012). These rare recombination events contribute to the heteroplasmic state of the mtDNA and are responsible for the different configurations of the mtDNA, called sublimons, that can co-exist with the main genome in the same individual, often less than one copy per cell (Small et al., 1989; Woloszynska and Trojanowski, 2009). Consequently, the stoichiometry of the different sequences that constitute the plant mtDNA might not be uniform, some being more represented than others in a cell (Preuten et al., 2010). But other studies showed, on the contrary, remarkable stability of the genome in the first stages of seed development (Paszkiewicz et al., 2017). During mitochondrial division the mtDNA is not equally transmitted to the daughter mitochondria, and the mitochondria which inherited DNA might not get a complete mtDNA genome (Arimura et al., 2004; Takanashi et al., 2010). Therefore, in a cell, mitochondria outnumber the copies of mtDNA, and some mitochondria do not contain any DNA (Preuten et al., 2010). Taken together, all these data led to build a global model where plant mtDNA is divided into several subgenomic molecules, obtained through recombination events, spread along the chondriome of the cell.

The increase in available data from assembled plant mitochondrial genomes tends to confirm their multipartite structure. This is true for large genomes composed of multiple chromosomes, like the one of cucumber and its 3 circular chromosomes (1.6 Mb, 84 kb, and 45 kb) (Alverson et al., 2011), and the ones of *Silene noctiflora* and *Silene conica* which are segmented in 58 and 128 circular chromosomes (from 44 kb up to 192 kb), respectively (Sloan et al., 2012). Quantitative data accompanying assembled genomes do not permit to conclude about the stoichiometry of the different chromosomes. Indeed, in cucumber the coverage of the 3 circular chromosomes is different (1.6 Mb: ~90X, 84 kb: ~65X, and 45 kb: ~41X) (Alverson et al., 2011), while the mitochondrial genome of *Psilotum nudum* shows 2 circular chromosomes (364 kb and 264 kb) with the same coverage (Wang et al., 2019). Furthermore, despite the possibilities of interconversion between sub-genomes by reciprocal recombination involving repeated sequences, it seems that only a particular configuration is transmitted to the plant progeny (Gualberto and Newton, 2017). Some of these chromosomes do not carry genes, with the consequence that they are not under functional pressure for maintenance. That is why they can be rapidly gained or lost (Wu et al., 2015). This versatility contributes to the intraspecific diversity of the plant mtDNA.

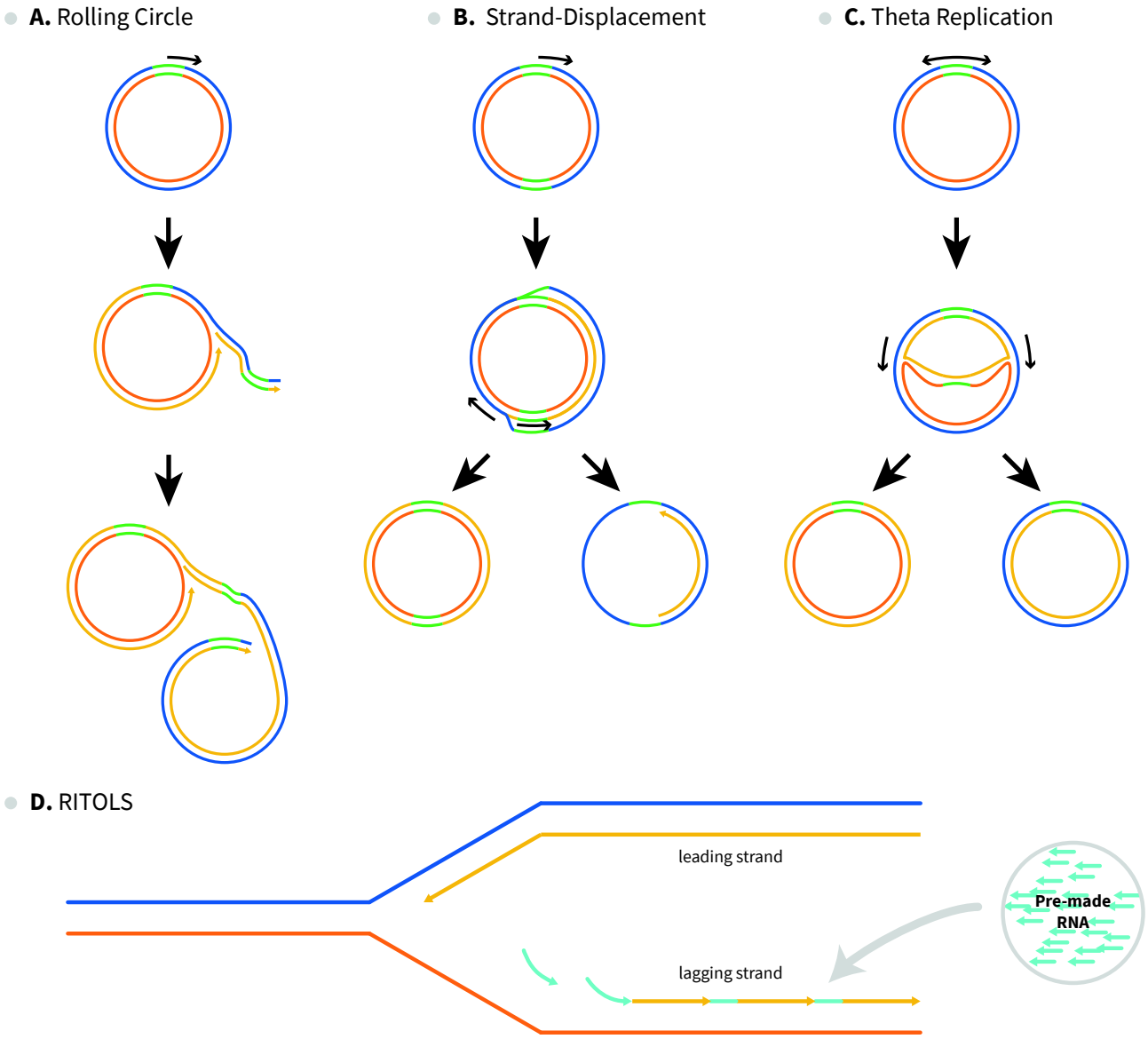


Figure I.8. Models for the replication of the animal mtDNA. Blue and orange represent the two original strands, in yellow are the strands newly synthesized and green show the replication initiation region . (**A**) Rolling circle replication. First, a nick is induced in one strand of DNA, then replication starts unidirectionally along the circular molecule while the nicked strand moves. Once the replication initiation site is reached, the displaced strand may be nicked again and ligated to form a new single-stranded circular molecule, or the synthesis proceeds and create a concatemeric molecule that is later converted into multiple single-stranded circular copies and later turned into a double-stranded DNA by the DNA polymerase. (**B**) Strand-displacement replication model. It involves an RNA primer that displaces one strand. Once two-thirds of the genome are synthesized, replication starts at the second origin of replication. When replication reaches the initiation site, the parent strand is displaced as a single-stranded circular molecule that is later converted into double-stranded DNA by the DNA polymerase. (**C**) Theta-like replication. Replication is initiated bi-directionally forming two replication forks leading to two double-strand circular copies of the genome. (**D**) The RITOLS replication model involves free pre-synthesized RNA molecules (in cyan) that hybridize to the lagging strand, starting from the 3' end. Then gaps are filled and the RNA primers removed. Adapted from Ciesielski et al., 2016 and Morley et al., 2019.

Of course, multipartite genomes do not only involve large chromosomes. Indeed, plant mitochondrial genomes can also show small replicative extrachromosomal replicons, circular or linear. This is the case of the 1.3 kb mp1 plasmid of *Chenopodium album* (Dörfel et al., 1991), able to replicate by rolling circle (Backert et al., 1996). However, these plasmids do not contribute to mitochondrial function and apparently their presence or absence does not affect plant fitness (Gualberto and Newton, 2017). Indeed, few open reading frames (ORFs) have been found in large plasmids, but they are mostly involved in the selfish replication of the plasmid that carries them. These plasmids evolve faster in sequence than the main mtDNA, and might be a key to explain the heterogeneity in substitution rates observed among mitochondrial genes in angiosperm (Smith et al., 2014): the periodic insertion of mitochondrial genes in plasmids would change their rates of sequence evolution (Warren et al., 2016).

4.4. Replication of the plant mitochondrial genome

Such a mixture of linear and/or circular subgenomes raises the question about their replication. This aspect of plant mtDNA maintenance is still not understood. In animals, several models have been proposed (**Figure I.8**) for different organisms (Ciesielski et al., 2016). These models include rolling circle in *Caenorhabditis elegans* (Lewis et al., 2015), theta replication in *Drosophila melanogaster* (Jõers and Jacobs, 2013), strand-displacement or D-loop replication (Doda et al., 1981) and RITOLS (ribonucleotide incorporation throughout the lagging strand) (Pohjoismäki et al., 2010) in vertebrates. Theta replication seems to be shared by all the invertebrates except for nematodes that adopted a rolling circle mechanism for the replication of their mtDNA. In vertebrates, D-loops and R-loops are commonly accepted mechanisms of mtDNA replication initiation. Animals use a minimal replisome for the replication of their mitochondrial genome, composed of the DNA helicase TWINKLE and of the DNA polymerase POLγ (Korhonen et al., 2004), both components of the nucleoid.

It was proposed that the plant mitochondrial genome replicates by mechanisms of recombination-dependent replication, similar to those found in bacteriophage T4 (Oldenburg and Bendich, 1996; Backert and Börner, 2000). Such a mechanism has been proposed for the replication of the yeast *Candida parapsilosis* mtDNA (**Figure I.9**) (Gerhold et al., 2010, 2014). Nonetheless, it has been shown that mitochondrial plasmids that can be found in certain plant species replicate through R-loop dependent rolling circle mechanism (Backert, 2002). Moreover, a recent study showed that specific cutting in mtDNA sequences with transcription activator-like effector nuclease

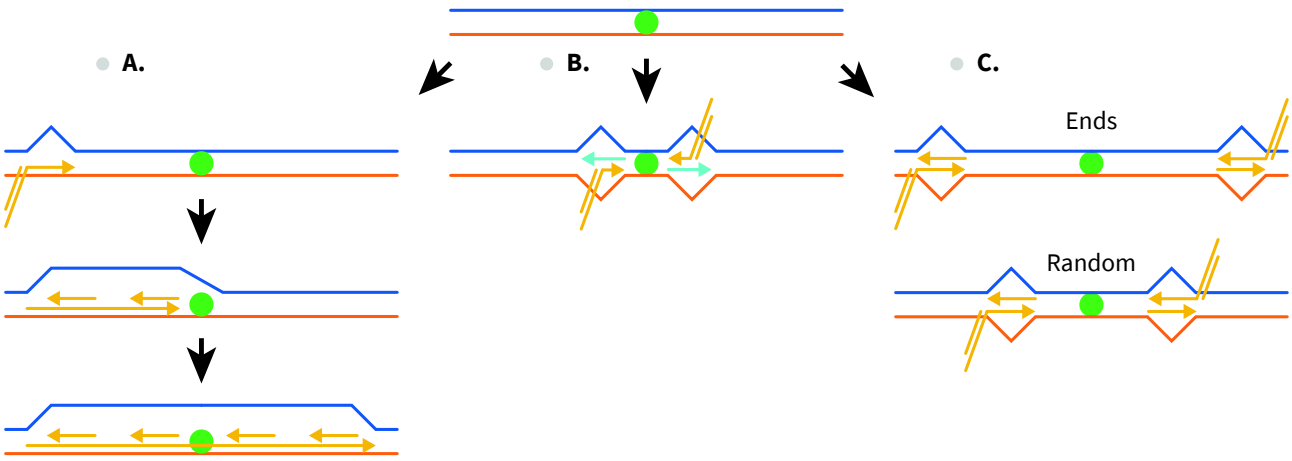


Figure I.9. Replication models for the mtDNA of the yeast *Candida parapsilosis*. The Green circle indicates the switching point of the two major transcription units. (**A**) This model relies on the unidirectional replication of the mtDNA that is initiated from the left telomeric repeat (TR) and exposed ssDNA of the upper strand. (**B**) In this model, the recombination-dependent replication is facilitated by transcription (RNA represented in cyan). Strand invasion initiates unidirectional and/or bidirectional replication. (**C**) The strand invasions may provide primers for replication initiation at random locations across the mtDNA, or from the TR toward the respective opposite ends. As soon as the replisome reaches the coding regions, replication proceeds in a strand-coupled mode. Adapted from Gerhold et al., 2014.

(TALEN) led to the deletion of the targeted sequences, but also of the surrounding regions (Kazama et al., 2019), as expected if the targeted sequences are carried by a subgenome no longer able to replicate after a double-strand break. The autonomous replication of circular subgenomes generated by recombination has been shown in mutants of recombination factors, like of helicase RECG1 (Wallet et al., 2015). Even if replication mechanisms are not fully understood, the set of mitochondrial proteins identified as potentially involved in replication supports the idea that plant mtDNA replication might use alternative mechanisms, from recombination-dependent replication to replication by a replisome similar to the one of bacteriophages (Moriyama and Sato, 2014; Morley et al., 2019; Brieba, 2019), requiring primers that could be synthesized by the TWINKLE replicative helicase that has primase activity (Diray-Arce et al., 2013; Peralta-Castro et al., 2017).

4.5. Plant mtDNA organizational diversity

Plant mtDNA evolves slowly in sequence. Indeed, the synonymous mutation rate of the mtDNA is different from the one of the nuclear genome. In animals, the mtDNA evolves faster than the nuclear genomes (Brown et al., 1979), which is expected because of the genotoxic environment of the mitochondrial matrix caused by the reactive oxygen species (ROS) produced through oxidative phosphorylation. In most plants, however, the opposite is observed, the plant mtDNA showing a lower mutation rate than the nuclear genome (Wolfe et al., 1987). This can be explained by the action of high fidelity repair mechanisms in the plant mitochondria that shape the mtDNA (Davila et al., 2011), such as homologous recombination (HR) that can copy-correct mutations by gene conversion (Gualberto and Newton, 2017). Moreover, the mutation rate is not equally distributed over the mtDNA. Indeed, the coding sequences show fewer substitution rates than the non-coding ones (Christensen, 2013). This can be explained by an evolutionary pressure that counter-selects mutated genes, while this evolutionary pressure does not apply to the non-coding sequences.

The evolutionary pressure on coding sequences pushes the HR repair mechanisms to be very active. On the other hand, their high activity forces the plant mtDNA to evolve rapidly in structure. That is why comparative studies show that the relative position of the genes within the mtDNA varies between closely related species and even within the same species. This has been shown in studies comparing different accessions of *A. thaliana* (Arrieta-Montiel et al., 2009; Davila et al., 2011), beet (Darracq et al., 2011) and maize (Allen et al., 2007), which highlighted the shuffling of sequences by recombination (deletions, duplications, insertions and inversions). This shuffling

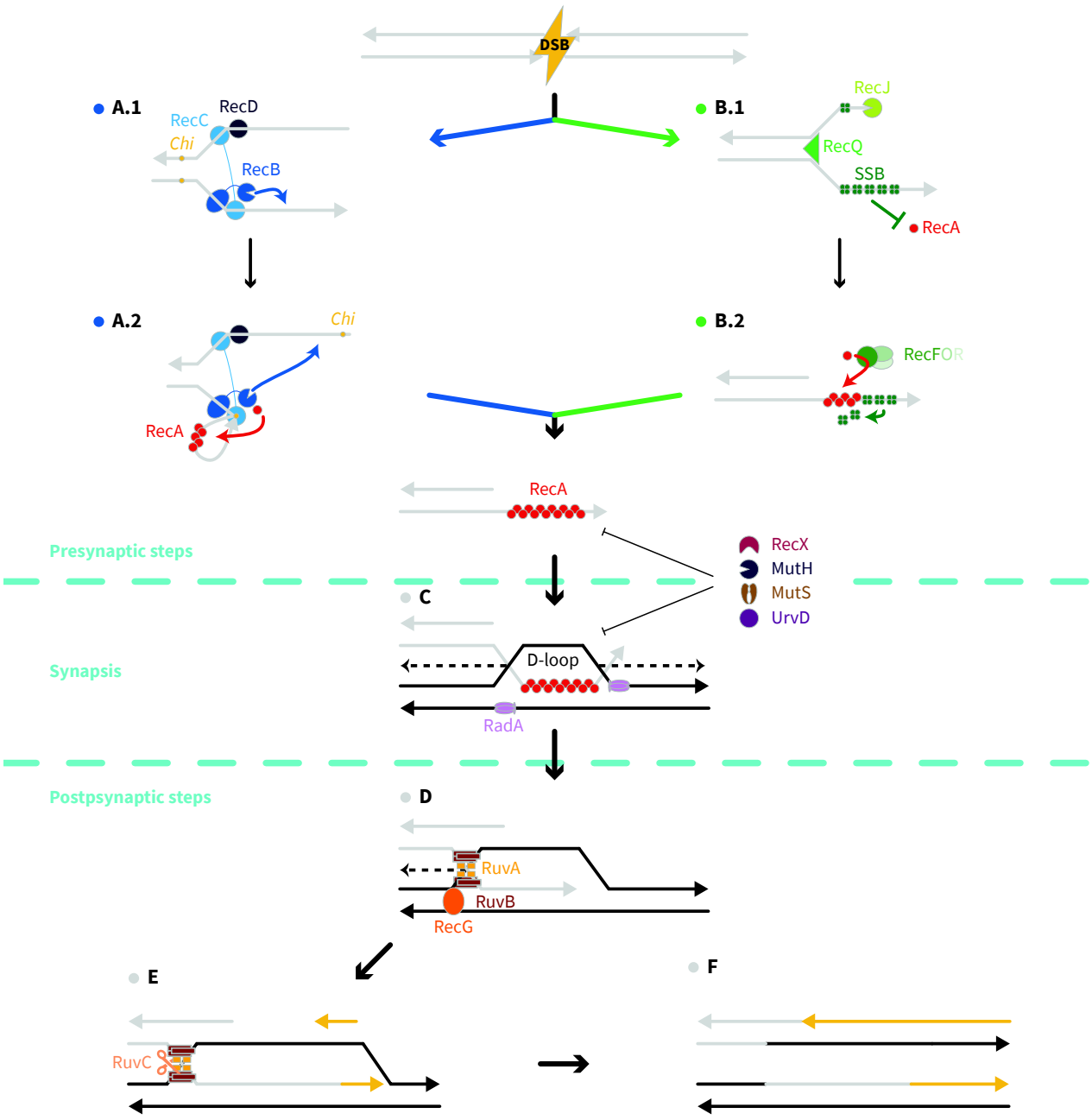


Figure I.10. Major steps of bacterial homologous recombination. In bacteria, the mechanism of repair of a DSB by HR starts with the resection of the DNA. (**A**) The recBCD pathway (blue) involves RecBCD that is composed of three factors and has exonuclease and helicase activities. The nuclease activity of RecB digests the DNA until a crossover hotspot instigator sequence (Chi). The joint action of these three factors led to the loading of RecA on ssDNA. (**B**) The loading of RecA on ssDNA can be also mediated by the RecFOR pathway. This pathway involves the 3'-5' helicase RecQ and the 5'-3' nuclease RecJ, that degrade and unwind the DNA, which is then coated with SSB proteins. The RecFOR pathway also involves the RecFOR complex, which manages the substitution of the SSBs by RecA. In both cases (A & B), the RecA-ssDNA nucleofilament scans the dsDNA to find a homologous sequence. (**C**) When a homologous sequence is found, the nucleofilament invades the dsDNA to form a D-loop. (**C** & **D**) The extension of the homology region is promoted by branch migration, mediated early by RadA and later by RecG and RuvAB, forming a four-strand structure named Holliday junction (HJ). (**E**) The HJ is resolved by the RuvC nuclease recruited by RuvAB and (**F**) the missing DNA is synthesized by a DNA polymerase. Courtesy of Nicolas Chevigny and adapted from Chevigny et al., 2020.

plays an important role in mitochondrial genome evolution by changing the genes synteny, but it can also lead to the creation and expression of chimeric genes. Indeed, mtDNA sequences shuffling can be neutral when affecting non-coding sequences, but when the shuffling is associated with the loss of an essential gene it leads to a severe defect in plant growth (Marienfeld and Newton, 1994; Newton et al., 1996). However, the mitochondrial rearrangements can open new ORFs leading, for some of them, to a gain of function for the mitochondria. Among the traits that can arise from these new ORFs, a prominent one for its economic importance is cytoplasmic male sterility (CMS).

4.6. Mechanisms of plant mtDNA dynamics

In plants mitochondria, HR is the main repair mechanism (Gualberto and Newton, 2017). HR is a process that repairs DNA double-strand breaks (DSBs) and single-strand gaps (SSGs). Through these repair mechanisms HR has a major impact on the dynamics of the genomes. The HR process can be dissected in three phases: presynaptic, synaptic and postsynaptic. During the presynaptic stage, the DSB is recognized and following resection of the DNA extremities the presynaptic filament is formed, which is constituted by the recombinase RecA (or Rad51 in the nucleus) bound to ssDNA. Then in the synaptic stage this filament invades a homologous dsDNA, to form a D-loop. Finally, during the postsynaptic step, the synapse is processed and the homologous DNA molecules released. HR is widely distributed through evolution, from prokaryotes and viruses to animals and higher plants. In fact, the faithful HR-dependent repair mechanisms are considered the main reason for the high conservation of mtDNA sequences and diversified mtDNA organization in plants (Palmer and Herbon, 1988). Due to the endosymbiotic origin of mitochondria its HR machinery is partially derived form the one of its prokaryotic ancestor. It is therefore necessary to understand the prokaryotic mechanisms of HR to elucidate the HR mechanism of mitochondria.

4.6.1. Homologous recombination in prokaryotes

In *Escherichia coli*, the main repair mechanism of a DSB starts with the resection of 5' phosphate DNA ends by RecBCD (**Figure I.10**). RecBCD is an enzymatic complex composed of three factors: RecB, RecC and RecD. RecB is a 3'-5' helicase with a nuclease activity, RecC is the recognition domain of the complex and RecD is a 5'-3' helicase. RecB and RecD unwind DNA in a are ATP-dependent way. The nuclease activity of RecB digests the DNA until a specific site called "crossover hotspot instigator" or *chi* (Taylor et al., 2014) of eight nucleotides long (5'-GCTGGTGG-3') (**Figure I.10.A**). This site is recognized by RecC and induces a change in the conformation of the RecBCD complex that stops the RecD helicase activity. On its side, the RecB activity is maintained and

induces a DNA loop, upstream of the *chi* site (Cockram et al., 2015). Then RecC digests the 3' end downstream *chi* and cleaves the 5' extremity. Then RecBCD begins loading RecA proteins onto the ssDNA of the newly generated 3' end (**Figure I.10.A**). RecA is a recombinase and its association with the ssDNA is called the presynaptic complex (Galletto et al., 2006).

However, if for some reason this main repair pathway is non-functional there's an alternative pathway that can be mobilized. This mechanism is the RecF pathway (Morimatsu and Kowalczykowski, 2003), specialized in the resolution of SSGs and that recruits the RecQ 3'-5' helicase and the RecJ 5'-3' nuclease (**Figure I.10.B**). RecQ unwinds the DNA from the DSB and allows RecJ to be loaded on one strand that is then digested. The single-stranded 3' end thus produced is then protected by the binding of SSB proteins. However, the bound SSB inhibits the loading of RecA (**Figure I.10.B**). The complex RecFOR makes that loading possible (Umezu et al., 1993; Umezu and Kolodner, 1994). To do so, RecF recognizes the ssDNA coated with SSBs that are removed by the RecOR complex (Sandler and Clark, 1994). Then, RecFOR mediates the loading of RecA onto the ssDNA to form the presynaptic complex (Galletto et al., 2006) (**Figure I.10.B**).

In both cases, the newly established presynaptic complex then scans for a homologous sequence in double-strand DNA (dsDNA). To do so, the RecA nucleoprotein filament randomly probes the dsDNA (Forget and Kowalczykowski, 2012) until homology recognition that can occur at as few as 6 nucleotides (Ragunathan et al., 2012). Once a homologous sequence is found, the RecA nucleoprotein filament invades the dsDNA to form a triple-stranded structure called D-loop (**Figure I.10.C**). The formation of a D-loop, that marks the end of the synaptic steps, is negatively controlled by UvrD (Veaute et al., 2005) or by RecX (Stohl et al., 2003), which inhibit the formation of the RecA nucleoprotein filament.

Finally, the postsynaptic step aims at the resolution of the D-loop through branch-migration. First, branch-migration extends the region of homology to prevent recombination between hemi-homologous regions. Then, two branch-migration pathways can be involved to maturate the D-loops. The first one involves the protein RadA that is an ATP-dependent helicase paralog to RecA (not to be confused with RadA from *Archaea*). To do so, RadA hexamers are recruited by RecA and loaded on both sides of the D-loop on the acceptor dsDNA (Marie et al., 2017). Then, the helicase activity of RadA unwinds the acceptor dsDNA and extends the D-loop on both sides (**Figure I.10.C**). The second pathway involves RecG (Whitby et al., 1994). RecG is also an ATP-dependent helicase, but moving on dsDNA. RecG, mobilized by SSB, recruits the second strand of

Molecular function	*E. coli*	*A. thaliana*	Localization
DNA resection	RecBCD	?	-
ssDNA protection	SSB	SSB1	Mt + Cp
		SSB2	Mt
Recombinase loading	RecFOR	ODB1	Mt
		WHY2	Mt
		OSB1	Mt
		OSB3	Mt + Cp
		OSB4	Mt
Recombination regulation	RecX	RECX	Mt + Cp
	UvrD	-	-
	MutH	MSH1	Mt + Cp
	MutS		
Formation of D-loop	RecA	RECA2	Mt + Cp
		RECA3	Mt
Branch migration	RuvAB	-	-
	RecG	RECG1	Mt + Cp
	RadA	RADA	Mt + Cp
Resolution of Holliday junctions	RuvC	?	-
	Yqgf?	YQGF?	?
DNA synthesis	PolI	POL1B	Mt + Cp
DNA ligation	LigA	LIG1	Mt + Cp

Table I.2. Comparison between bacterial and plant mitochondrial factors involved in homologous recombination. Mt: Mitochondria; Cp: Chloroplast. Adapted from Gualberto and Newton, 2017.

the invading DNA thus converting the triple-stranded D-loop into a four-stranded structure called Holliday junction (HJ) (**Figure I.10.D**) (Mawer and Leach, 2014). In the end, the HJ is resolved by the RuvABC complex (West, 1997). In this complex, a tetramer of RuvA is located to the cross between the four strands and recruits a hexamer of RuvB. RuvB has helicase activity and moves the HJ until a dimer of the RuvC endonuclease is recruited (**Figure I.10.D**). Finally, RuvC cleaves the HJ (**Figure I.10.E**) and the two DNA molecules are released (**Figure I.10.F**). Depending on how RuC cuts the HJs there is gene conversion without cross-over or with cross-over.

4.6.2. Specificities of homologous recombination in plant mitochondria

The factors involved in plant mitochondrial homologous recombination are largely inherited from the bacterial ancestors of mitochondria or of the chloroplast (Gualberto and Newton, 2017). The study and the elucidation of the bacterial mechanisms (**Figure I.10**) help to identify orthologous factors in plants or to assign a putative function to others (**Table I.2**). Thus, plant mtDNA HR requires many DNA binding proteins and several factors homologous to the bacterial recombinase RecA. Several steps remain unclear, especially the early and late stages involving the resection of DSB extremities and HJ resolution, for which the responsible factors have not yet been identified. But all factors characterized so far are encoded by the nucleus. Many are targeted to both mitochondria and chloroplasts or are part of small families comprising chloroplastic and mitochondrial isoforms. Like in bacteria (**Figure I.10**), plant mitochondrial repair of DSBs by HR should also start with the resection of DNA extremities. But the high proportion of ssDNA found in the plant mtDNA suggests that the repair of SSGs is also predominant. Afterward, the ssDNA ends (in DSB repair) of ssDNA gaps (in SSG repair) are protected from nucleases by ssDNA-binding proteins. Such action could be done by SSB1 and/or SSB2, homologs of bacterial SSB. The possible differential functions of SSB1 and SSB2 are not known, but both are essential to the recombination process (Edmondson et al., 2005). This same function, as well as the promotion/inhibition of HR could also be the roles of the OSB (organellar single-stranded binding) proteins or of the WHY proteins from the whirly family (Zaegel et al., 2006; Cappadocia et al., 2010). In *A. thaliana*, the members of these protein families can be targeted to mitochondria (OSB1, OSB4, WHY2), to chloroplasts (OSB2, WHY1, WHY3) or to both organelles (OSB3) and have the particularity of having a higher affinity to dsDNA than the SSB proteins (García-Medel et al., 2019). The loss of OSB1 and OSB4 leads to increase ectopic recombination events involving intermediate-size repeats, and consequently to genomic rearrangements (Zaegel et al., 2006). On the other hand, loss of WHY proteins leads to an increase in recombination involving microhomologies, both in mitochondria

(*why2* mutants) and in the chloroplast (double mutant *why1 why3*), and inhibits HR repair of DSBs induced by ciprofloxacin treatment (Marechal et al., 2009; Cappadocia et al., 2010). However, it has been shown that OSB and WHY factors do not interact with the POL1B organellar polymerase, the DNA polymerase involved in repair mechanisms (García-Medel et al., 2019). The authors of this publication have proposed alternative roles for the SSB, OSB and WHY proteins. In DSB, the SSBs (SSB1 and SSB2) would promote the association of POL1B with the ends of the breaks, favoring repair by microhomology-mediated end-joining, whereas OSB1 and WHY2 inhibit end-joining mechanisms by blocking the access of DNA 3'-OH ends to SSBs and to POL1B (García-Medel et al., 2019).

Then, the recombinase is loaded on the ssDNA to form the presynaptic filament, which requires the displacement of the ssDNA-binding proteins by recombination mediators. Such mediator function could be the role of several factors. The organellar DNA binding protein ODB1 is homologous to yeast Rad52 and could play that role. Indeed, ODB1 promotes the hybridization of complementary ssDNA sequences *in vitro*, supporting the mediator role of ODB1 in recombination (Janicka et al., 2012). The role of recombination mediator could also be performed by WHY2 (Cappadocia et al., 2010). ODB1 and WHY2 both co-purified with mtDNA and are considered as important components of the nucleoid. They could act in the same way or promote the recruitments of different factors in different recombination pathways (Gualberto and Kühn, 2014). As for WHY2, the loss or inactivation of ODB1 inhibits DSB repair mechanisms and leads to an increase of recombination involving micro-homologies (Cappadocia et al., 2010; Janicka et al., 2012).

In *A. thaliana* there are three recombinases homologous to bacterial RecA, two of which are targeted to mitochondria (**Table I.2**): RECA2 is targeted to both mitochondria and chloroplast and RECA3 is only imported by the mitochondria. RECA2 and RECA3 might be involved in different partially redundant HR pathways (Miller-Messmer et al., 2012). Both can partially complement the bacterial mutant *recA*- for survival after genotoxic treatment with hydroxyurea or mitomycin. However, the survival of the *recA*- mutant after UV stress was only complemented by RECA2. *A. thaliana* mutants for RECA2 or RECA3 do not share the same phenotype: the *recA2* mutant is lethal early in seedling development whereas the *recA3* mutant has little effect on plant development and only increases sensitivity to genotoxic stresses; but both mutants show an increase in ectopic recombination involving repeated sequences (Miller-Messmer et al., 2012). These results suggest that RECA2 may be the main mitochondrial recombinase whose loss leads to excessive and lethal

rearrangements of the mtDNA, whereas RECA3 rather participates in mtDNA repair, mainly through pathways that do not involve crossovers, such as synthesis-dependent strand annealing (Miller-Messmer et al., 2012). In agreement with this assumption the *recA3* mutation is highly synergistic with the mutation of other factors that modulate mitochondrial HR, such as MSH1, RECG1 and ODB1 (Arrieta-Montiel et al., 2009; Davila et al., 2011; Wallet et al., 2015 and unpublished results of the laboratory). RECA3 has the distinction of not possessing the important C-terminal acid extension found in all other RecA proteins, which is necessary in *E. coli* for the interaction with factors modulating its activity, such as SSB and RECX (Eggler et al., 2003; Drees et al., 2004). Thus, mitochondrial factors regulating RECA2 could be dodged by RECA3 that would so escape from their regulation. Among these regulatory factors, RECX, the mitochondrial ortholog of bacterial RecX, is found in all land plants and was recently shown in *Physcomitrella paten* to interact with mitochondrial RecA. Moreover its overexpression mimics the RecA mutation (Odahara and Sekine, 2018).

As shown in **Figure I.10**, during branch-migration the newly formed D-loop extends through the action of helicases. In bacteria (as described in **5.1.1**) there are two branch-migration pathways described as partially redundant: the RuvAB pathways and the RecG/RadA one. In plants however, the RuvAB pathway is absent from organelles, since there are no orthologs of RuvA and RuvB encoded by the plant genomes. But orthologs of RecG and of RadA are encoded by the nuclear genomes of all land plants (Chevigny et al., 2020). Both are described as dually targeted to the chloroplast and to mitochondria (Odahara et al., 2015; Wallet et al., 2015; Chevigny et al., 2019). The RECG1 protein, the homolog of RecG, can partially complement the bacterial mutant *recG*- in survival to UV treatment, suppression of pathological re-initiation of replication and to inhibition of replication by RNA primers (Wallet et al., 2015). These results suggest that RECG1 performs the same function in plant organelles as RecG in bacteria (Wallet et al., 2015). However, the *recG1* mutation does not significantly affect the phenotype of the plant, but does increase ectopic recombination involving small repeated sequences of the mtDNA (Odahara et al., 2015; Wallet et al., 2015). In bacteria, survival after genotoxic treatment is little affected when a single branch-migration pathway is mutated (RuvAB or RecG/RadA), but the survival decreases drastically when both pathways are affected (Beam et al., 2002). This last result put together with the lack of the RuvAB pathway in plant suggests that another branch-migration pathway should exist in plants. In a recent review, RADA is suggested to be involved in this pathway (Chevigny et al., 2020). The RADA protein is the ortholog of bacterial RadA and its loss has severe effects on mtDNA stability and plant

development (Chevigny et al., 2019). The authors also suggest that RECA3 could assume the branch-migration function based on the activities of bacterial RecA (Chevigny et al., 2020). Indeed, branch-migration is one of the RecA activities and it allows strand exchange reactions to occur *in vitro*, in the presence of RecA only. The deletion from *E. coli* RecA of its C-terminal acid extension (17 amino acids) affects every one of its functions by making them more robust (Cox, 2007). Indeed, for instance, this mutant no longer requires free Mg^{2+} ion for its optimal strand exchange activity and binds faster to DNA duplex (Lusetti et al., 2003a, 2003b). Thus, the loss of the C-terminal acid extension of RECA3 might have increased its branch-migration activity, allowing RECA3 able to partially complement the lack of RECG1 or RADA in plant mitochondria (Chevigny et al., 2020).

Finally, the resolution of HJ is mediated by RuvC in *E. coli*. But in plant organelles the absence of the RuvAB pathway results in the absence of the RuvC resolvase. However, RuvC is not universally present in bacteria, and RecU or YqgF might be used as alternative resolvases (Aravind et al., 2000; Rocha et al., 2005). A homolog of YqgF is found in plants and could potentially be involved in the resolution of HJ in mitochondria (Chevigny et al., 2020). But It is also possible that mitochondrial HJs are resolved not by a specific nuclease, but by dissolution due to the combined action of branch-migration factors and topoisomerase I. Moreover, RECG1 could resolve HJs by moving them to replication forks, as it was proposed in bacteria (Wardrope et al., 2009).

4.7. Mitochondrial genome inheritance

Like mitochondria, the mtDNA is transmitted maternally in most plants, (Birky, 2001). This uniparental mode of inheritance might be key to the plant offspring near-homoplasmy, meaning that in a non-mutagenic and non-recombinant environment nearly all mitochondrial genomes in an individual are identical (Birky, 2001). Nonetheless, this fragile balance can be disrupted by small-scale mutations, unwanted recombination events or paternal mtDNA leakage, leading to a heteroplasmic mitochondrial population. But due to the polyploidy of the mtDNA within a cell (Sheahan et al., 2004; Logan, 2006; Preuten et al., 2010) the sublimons or new mtDNA variants (mitotypes) remain confidential because the phenotype is determined by the predominant mitotype (Kmiec et al., 2006). However, the predominant mitotype can change very fast, in just two generations, by the so-called process of substoichiometric shifting (SSS) (Woloszynska, 2010). The mechanisms behind SSS remains unclear, but two possibilities are considered: an increase in homologous recombination and/or the preferential amplification of one of the alternative mtDNA variants that already co-existed with the predominant mitotype (Woloszynska, 2010). SSS does not

revert the mtDNA population to a homoplasmic state. Indeed, the population remains heteroplasmic, but with a new predominant mtDNA. An example in common bean showed that the process can be reversed. In this species the CMS-associated ORF *PVS*, present in a fertile line, became predominant after SSS, inducing male sterility. In a second generation the mtDNA reverted to male fertility (Janska et al., 1998; Kmiec et al., 2006). SSS might not be the only way to an mtDNA variant to become dominant. Another possibility could implicate cytoplasm sorting, which generation after generation promotes the segregation of a single predominant mitotype (Mishra and Chan, 2014; Stewart and Larsson, 2014). This has been shown in *A. thaliana*, where a *recG1* mutant showed an increase in ectopic recombination leading to heteroplasmic mtDNA with several co-existing mtDNA configurations that were sorted after backcross with WT pollen, leading to new stable configurations of the mtDNA (Wallet et al., 2015).

4.8. Nucleo-cytoplasmic interactions

The communication between the nucleus and the cytoplasmic organelles, mitochondria and chloroplasts, is required to cellular homeostasis. Communication from the nucleus to the organelles is known as anterograde signaling while the reverse communication is known as retrograde signaling. Both are used by most, if not all, eukaryotes but with little conservation of the signals and factors involved, indicating the specificity of the nucleus-cytoplasm relationship (Woodson and Chory, 2008).

4.8.1. Anterograde signaling

Mitochondrial biogenesis and function require more than a thousand proteins, but the mtDNA only encodes a (very) small portion of them (~3,5 %). That is why proper mitochondrial activity needs the coordinated expression of nuclear and mitochondrial genes. Anterograde signaling manages this coordination and adapts it to environmental changes that shift the mitochondrial functions (Woodson and Chory, 2008). The anterograde regulation can occur at three different levels: transcriptional, post-transcriptional and post-translation.

All known factors of the mitochondria transcriptional machinery are nuclear-encoded (Ikeda and Gray, 1999). For instance, plant mitochondria gene expression depends on RNA polymerases that are encoded by the nucleus. However, the transcriptional activity does not correlate with protein levels in plant mitochondria (Giegé et al., 2005), which is ultimately controlled by post-transcriptional and translational processes (Woodson and Chory, 2008).

The regulation of organelle gene expression mostly occurs through post-transcriptional regulation controlled by nuclear-encoded proteins (Leon et al., 1998). These proteins are numerous and dedicated to a reduced set of mitochondrial genes and act as adaptors by binding to specific transcripts and recruiting RNA processing enzymes or the translation machinery (Woodson and Chory, 2008). The majority of these factors are pentatricopeptide repeat (PPR) proteins. PPR proteins show a repeated 35 degenerate 35-amino-acid motif (Small and Peeters, 2000; Lurin et al., 2004) and are involved in RNA editing (Kotera et al., 2005), transcript processing (splicing) (Fisk et al., 1999), transcript stabilization (Wang et al., 2006) and transcript translation in both chloroplast and mitochondria (Schmitz-Linneweber et al., 2005).

The mechanisms of post-translational regulation act through the rapid degradation of unassembled proteins and the regulation of the assembly of multi-subunit complexes. These processes are conserved in plants (Lu et al., 1996) and yeast (Rep and Grivell, 1996) mitochondria. Moreover, it has been shown that the copy number of plant mitochondrial ribosomes and respiratory complex decrease during sucrose starvation, due to limiting nuclear-encoded proteins (Giegé et al., 2005). Consequently, levels of mitochondrial-encoded proteins must be adjusted to those of nuclear-encoded ones by post-translational mechanisms.

Anterograde regulation, at these three different levels, is highly important in plant mitochondria to suppress the potentially deleterious effects of ORFs resulting from mtDNA recombination, such as in CMS. Indeed, the CMS phenotype can be restored *via* nuclear-encoded restorer of fertility genes (*Rf*) that are widely present in plants (Fujii et al., 2011). Moreover, anterograde regulation of the interaction between the nuclear-encoded *Rf* and the CMS-associated mitochondrial *orf* hides the CMS phenotype in a natural population, such as in *Mimulus guttatus* (Fishman and Willis, 2006; Case and Willis, 2008) or as among *A. thaliana* ecotypes (Gobron et al., 2013; Roux et al., 2016).

4.8.2. Retrograde signaling

Retrograde regulation is used by the organelles to communicate their functional and developmental state to the nucleus and, consequently, modulates anterograde signaling. This modulation informs the nucleus of cellular stimuli which can then respond to, by regulating transcription factors associated to the response, or even by triggering apoptosis (Woodson and Chory, 2008). The most studied protein-induced retrograde response is the one modulated by AOX1 (for alternative oxidase 1) that provides a degree of signaling homeostasis to the organelle

(Vanlerberghe, 2013). By contrast to the anterograde signaling acting through proteins, retrograde signaling acts through metabolic signals of different nature (Ng et al., 2014).

During respiration, reactive oxygen species (ROS) are generated in the mtETC. Normal ROS production is very sensitive to mitochondrial well-being and can be increased upon an over-reduction of the mtETC (Maxwell et al., 1999) and during stresses like cold (Prasad et al., 1994), salt (Hernandez et al., 1993) and phosphate starvation (Parsons et al., 1999). The increase of cellular ROS levels can affect mitochondrial functions by damaging proteins of the mtETC or of the Krebs cycle and by lipoxidation of the mitochondrial membranes (Sweetlove et al., 2002). Oxidative damage of the mtDNA also results in the expression of aberrant proteins that, if assembled in mitochondrial complexes, further amplify the production of ROS. Although the mechanisms behind the transmission of ROS-driven signals remain unclear, a scenario of local detection mechanisms inside and outside mitochondria is supported by transcriptomic data (Gadjev et al., 2006). In such a scenario the peroxidation of mitochondrial lipids could provoke specific signaling effects (Tang et al., 2002). Another possibility could be through damaged mitochondrial proteins (Sweetlove et al., 2002), but it would require the export of these peptides across the mitochondrial membranes and their perception by specific receptors that have not been evidenced so far (Ng et al., 2014). However, ROS can leave mitochondria via permeability transition pores upon mitochondrial dysfunction (Maxwell et al., 1999), and then participate in the first steps of a signaling cascade (Ng et al., 2014).

The respiration, generating ROS as a side product, aims at the production of ATP, and so energy could be an initiator of mitochondrial retrograde signaling. There is indeed evidence in *A. thaliana* suggesting that energy control of metabolism might be a component of retrograde signaling (Meyer et al., 2009). Thus, the growth deficiency observed when ATP synthase is disturbed is not due to a lack of available energy, but to a whole cellular metabolic readjustment (Geisler et al., 2012).

The mtDNA mainly encodes for proteins of the mtETC, and their production and insertion into the membrane may trigger a variety of signals that can be relayed to activate nuclear gene expression (Law et al., 2012). There is evidence for mitochondrial proteins to have a signaling activity besides their biochemical role, such as shown on mutants for respiratory complex I subunits that have an increased level of TIM17:23, higher import rates of nuclear-encoded proteins and increased mitochondrial transcription activity (Wang et al., 2012). The relationship between

complex I and TIM17:23 can be explained by the dual location in both complexes of the preprotein transporters B14.7 and Tim23-2 (Wang et al., 2012).

It has also been proposed that dual-targeted proteins (nucleus-mitochondria) could act as retrograde signals (Duchêne and Giegé, 2012). Indeeed, over 250 genes encode for dual-targeted proteins in *A. thaliana* (Carrie and Whelan, 2013), arising from an early acquisition in land plants (Xu et al., 2013). Such an example has been highlighted with the PPR protein PNM1, which is able to regulate its own gene expression in the nucleus and could thus adjust gene expression between mitochondria and the nucleus (Hammani et al., 2011).

Another possible retrograde signal is calcium. Studies in yeast and animal mitochondria on uptake and release of calcium ions (Ca^{2+}) have shown that an increase in mitochondrial Ca^{2+} concentration leads to a positive effect on ATP synthesis, while Ca^{2+} import into the matrix is energetically demanding and can disturb the membrane potential (Duchen, 2000) or open the permeability transition pore, leading to reduced mitochondrial membrane potential and even cell death (Hajnóczky et al., 2006). In plants, Ca^{2+} has been linked with the promotion of mitochondrial protein import (Kuhn et al., 2009), the induction of mitochondrial swelling, cytochrome *c* release (Virolainen, 2002) and to fluctuations in the mitochondrial transmembrane potential (Schwarzländer et al., 2012). But at present very little is still known about its potential role as a retrograde signal in plants.

5. Cytoplasmic male sterility

5.1. Pollen development

5.1.1. Flower structure and development

Flowers are composed of different organs organized in four whorls. There are two vegetative organs: the sepals in the outermost whorl and the petals in the second whorl; and two reproductive organs: the stamens in the third whorl that are the male organs and produce pollen and the carpel, in the fourth and last whorl, which is the female organ that contains the ovules. This organization is a general organization and flowers across evolution do not share these four organs. Flower development is directed by several homeotic transcription factors that play a major role (Theißen et al., 2016). Historically, the ABC model was the first model proposed to explain that role

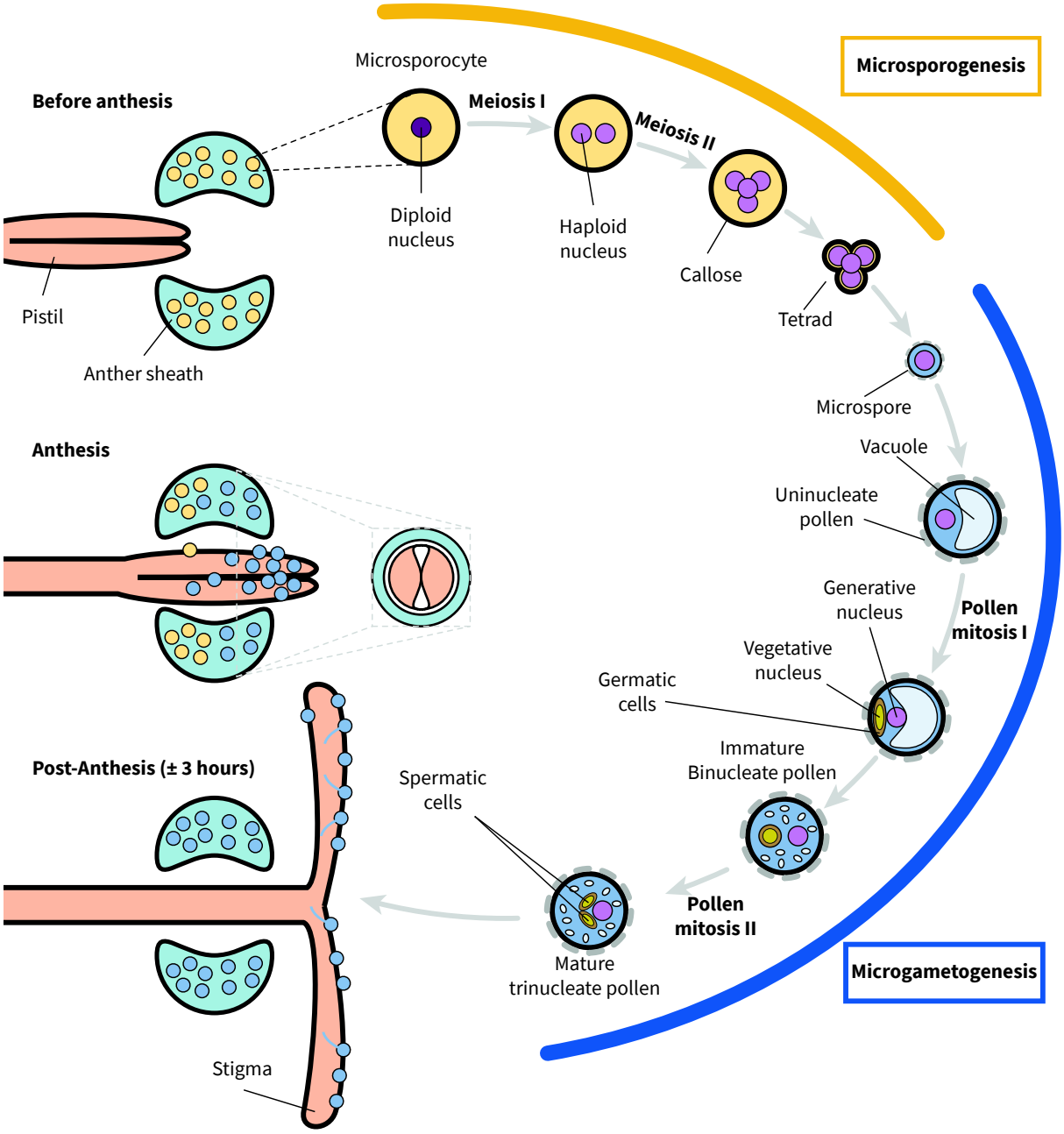


Figure I.11. Lettuce male gametophyte development. The germline development starts with the microsporogenesis where the diploid pollen mother cells undergo meiotic divisions (meiosis I and II) leading to four haploid spores grouped into tetrads. Microsporogenesis is then followed by microgametogenesis. During microgametogenesis the tetrad releases the four haploid microspores that undergo an asymmetric division, called Pollen Mitosis I, to produce a bicellular pollen grain with a small germ cell containing the vegetative nucleus engulfed within the cytoplasm of a large vegetative cell. The germ cell undergoes a further mitotic division at pollen mitosis II, to produce twin sperm cells in the mature pollen.

(Coen and Meyerowitz, 1991). According to this model the authors categorized the genes in three families “A”, “B” and “C”. Genes “A” are involved in sepals and petals development. Genes “B” interact with “A” in petals development and with “C” in stamen development. Finally, genes “C” are also involved in carpel development. In 2001 this model was completed to become the floral quartet model (Theißen and Saedler, 2001). This model keeps the ABC genes and adds two more classes of genes: D and E. The genes of class “E” are involved in the development of all floral organs including the ovules while the “D” genes are specific to the ovule. The proteins encoded by these genes are MIKC-type MADS-domain protein and act as tetramers, in quartet complexes (Theißen et al., 2016). Mutants for these genes, called homeotic mutations, have a strong floral phenotype that can go from the absence of male organs to their replacement by other flower parts (Yang et al., 2003).

5.1.2. Pollen development

The process of pollen development occurs in the anthers, located at the tip of the stamen, and is composed of two distinct sequential phases, microsporogenesis and microgametogenesis (Borg et al., 2009) (**Figure I.11**). The microsporogenesis starts with the diploid pollen mother cells undergoing meiotic division and ends when distinct unicellular haploid microspores are released from the tetrads by the activity of a mixture of enzymes secreted by the tapetum, the inner nutritive layer of the anther (Scott, 2004). During microgametogenesis microspores enlarge and undergo an asymmetric cell division known as pollen mitosis I (PMI) (Borg et al., 2009). The two cells resulting from PMI are the large vegetative cell and the small germinal cell within the cytoplasm of the first one. In some species, pollen development stops there and leads to mature binucleated pollen. However, in other species, such as lettuce and *A. thaliana*, the germinal cell undergoes a second round of mitosis called pollen mitosis II (PMII) to produce twin sperm cells, leading to trinucleated mature pollen (Borg et al., 2009) (**Figure I.11**).

Pollen development is a fragile balance requiring the coordinated expression of a high number of genes. Indeed, in *A. thaliana* this number represents half of the genes expressed in microspores and in binuclear pollen (Honys and Twell, 2004). Moreover, pollen development relies on the activity of the diploid sporophytic tissues and especially of the cells from the tapetum, which provide essential metabolic precursors (Ariizumi and Toriyama, 2011). Mutants of genes involved in any of these different steps show a deficiency in pollen viability or production (Sanders et al., 1999).

5.2. male-sterile phenotypes

In plants, the male-sterile phenotype is the inability to release functional pollen without affecting female fertility (Levings, 1993). This phenotype has been documented for the first time in 1763 by Joseph Gottlieb Kölreuter (Mayr, 1986). Nowadays, male-sterile phenotypes have been reported in more than 610 plant species (Chen and Liu, 2014). It would be reductive to limit the male-sterile phenotype to an absence of pollen production, because the whole process of male organ development can be affected. Indeed, three types of male sterility can be distinguished: homeotic sterility which results from the absence of male organs during flower development; sporophytic male-sterility which occurs on the diploid maternal tissues during microsporogenesis and leads to the absence of pollen; and gametophytic male-sterility which occurs, by contrast, during microgametogenesis on the haploid pollen (Guo and Liu, 2012; Chen and Liu, 2014). Male sterility can result from genetic and environmental causes, and can be induced by the application of chemical hybridizing agents (McRae, 1985). In the case of genetically-transmitted sterility a distinction is made, based on their mode of inheritance. Cytoplasmic male sterility (CMS) is maternally inherited and induced by a protein encoded by the mitochondrial genome. Genic male sterility (GMS) is governed by genes carried by the nuclear genome. It can be environment-sensitive and it is then called EGMS (Delourme and Budar, 1999).

Due to its nuclear origin GMS follows a Mendelian pattern of inheritance. The majority of the reported GMS mutants are in dicots and most of them arose spontaneously (Kaul, 1988). Among these naturally occurring GMS systems the large majority are governed by monogenic recessive alleles (Kaul, 1988; Delourme and Budar, 1999). Over the years, an increasing number of induced GMS systems became available. They have been obtained through physical (γ-ray, UV...) and chemical (ethyl methanesulfonate (EMS), colchicine...) mutagenesis, but most are also governed by monogenic recessive alleles (Kaul, 1988). The large majority of GMS mutants affect pollen development during early or late meiosis, and only a few acts before meiosis (Kaul, 1988). This means that the majority of GMS lead mostly to sporophytic or gametophytic sterility and only in rare cases to homeotic.

The EGMS is a GMS system that is influenced by environmental factors such as temperature, photoperiod or humidity (Virmani and Ilyas-Ahmed, 2001). According to Kaul, EGMS represents only a small proportion of the GMS systems and in this small proportion 44 % are influenced by temperature and 12 % by the photoperiod (Kaul, 1988). The 44 % remaining are modulated by

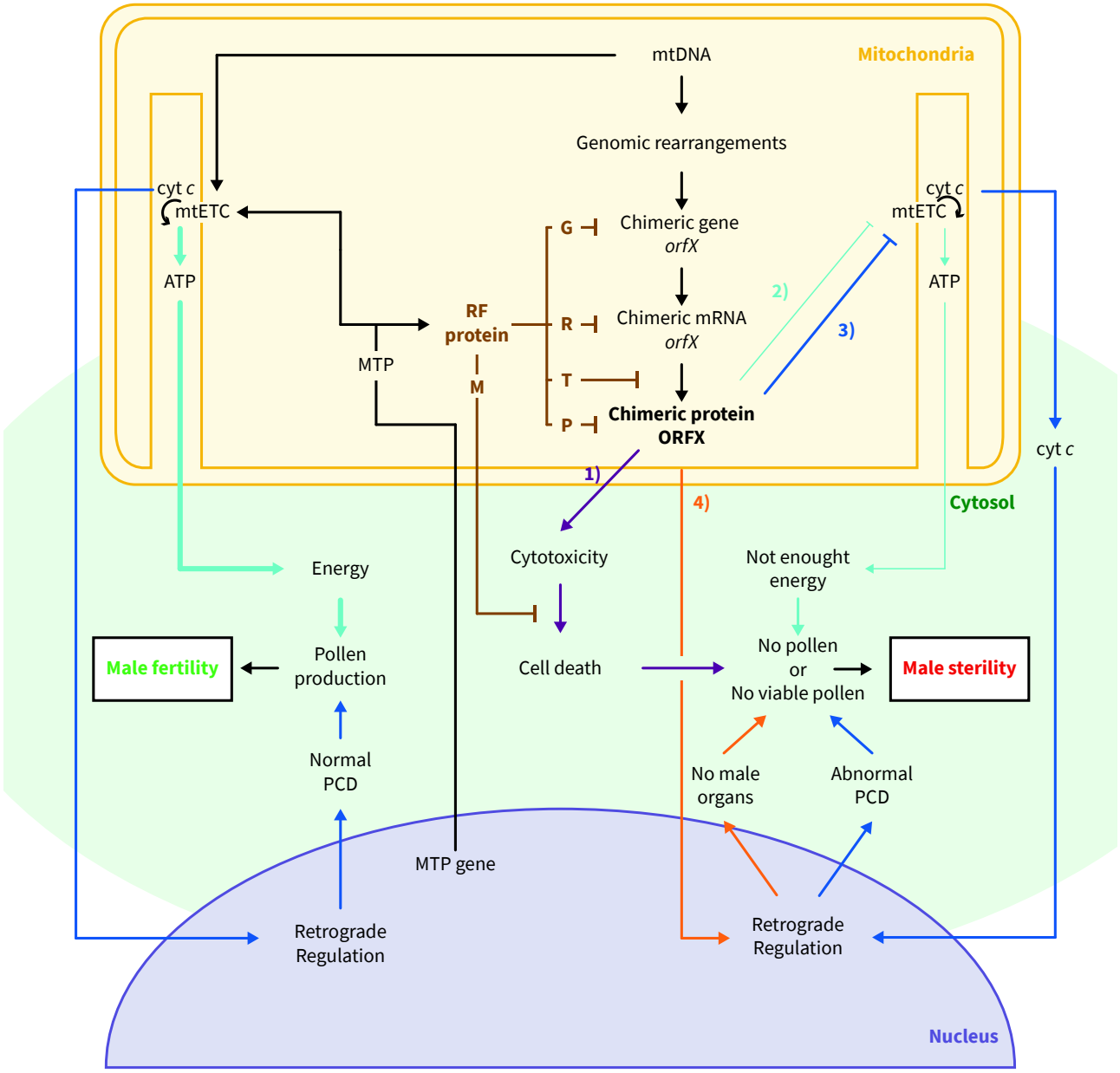


Figure I.12. A general model for CMS and fertility restoration. The mitochondrial-targeted proteins (MTP) are encoded by the nuclear genome and are involved in the mitochondrial electron transport chain (mtETC), mitochondrial genome recombination (not shown) and restoration of fertility (**RF**). Fertility (on the left) requires a large production of energy by mitochondria (**Turquoise**) and normal programmed cell death (PCD) in certain tissues mediated by the release of cytochrome *c* (*cyt c*) into the cytosol (**Blue**). Fertility can be lost through four models (starting from **ORFX** on the right): **1)** The cytotoxicity model which directly induces the death of the cell; **2)** The energy deficiency model by which the cell fails to reach the energy level required for male reproductive development; **3)** The aberrant programmed cell death model; **4)** The retrograde regulation model, by which a chimeric protein influences the expression of nuclear homeotic genes. However, plants developed strategies to counteract it, through nuclear restorer of fertility genes (**RF**) acting at the genomic level (**G**), at the transcript level (**R**), at the translation (**T**), at the post-translational level (**P**), or at the metabolic levels (**M**). Adapted from Chen and Liu, 2014.

diverse environmental stresses such as drought or salinity, or by a deficiency in certain micronutrients. The existence of EGMS underlines the complex epigenetic control of the reproductive development stages and sterility behaviors in response to environmental signals (Kim and Zhang, 2018). Indeed, in the rice photoperiod-thermo-sensitive EGMS line PA64S, high methylation was observed mainly in noncoding regions or promoters (Chen et al., 2014).

As the GMS and EGMS systems, the CMS phenotype is widely spread in natural populations, in particular in gynodioecious species in which female and hermaphrodite plants co-exist within the population (Dufaÿ et al., 2007). CMS has been reported in more than 150 vegetable species. It is a maternally inherited trait because the mitochondrial genome (mtDNA) is responsible for the expression of male sterility and mitochondria is virtually excluded from the pollen sperm cells in most plant species. Up to now no natural CMS system has been related to a mutation in the chloroplast genome, although it was reported that it is possible to artificially engineer CMS via the chloroplast genome in tobacco (Ruiz and Daniell, 2005). The CMS gene is usually a chimeric gene generated by recombination of the mtDNA. To identify a CMS-associated mitochondrial locus the finding of recombinant mitochondrial genes is only the first step, since there is a propensity for plant mitochondrial genomes to undergo a high amount of recombination. However, the number of natural CMS might be underestimated because of anterograde control by the nuclear genome through *Rf* genes, which suppress CMS and restore viable pollen production (Fishman and Willis, 2006; Case and Willis, 2008; Gobron et al., 2013; Roux et al., 2016). The large majority of CMS genes are associated with mitochondrial rearrangements. However, not all the mitochondrial genes are involved in CMS (Chen and Liu, 2014). Indeed, the genes involved in CMS are mostly genes coding for subunits of the mtETC (especially *cox1*, *atp6* and *atp8*) and the encoded proteins are predominantly transmembrane proteins (Chen and Liu, 2014). As described above for male sterility in general, the different types of CMS can be also grouped according to the tissues that are affected: CMS leading to an absence of male organs during flower development is called homeotic; CMS affecting microsporogenesis is called sporophytic; and CMS affected in microgametogenesis is called gametophytic (Guo and Liu, 2012; Chen and Liu, 2014).

5.3. The mechanisms of cytoplasmic male sterility

5.3.1. The cytotoxicity model

According to the cytotoxicity model (**Figure I.12**) the CMS protein directly kills the cells. Evidence for this model relies on the death of transgenic cells expressing proteins responsible for

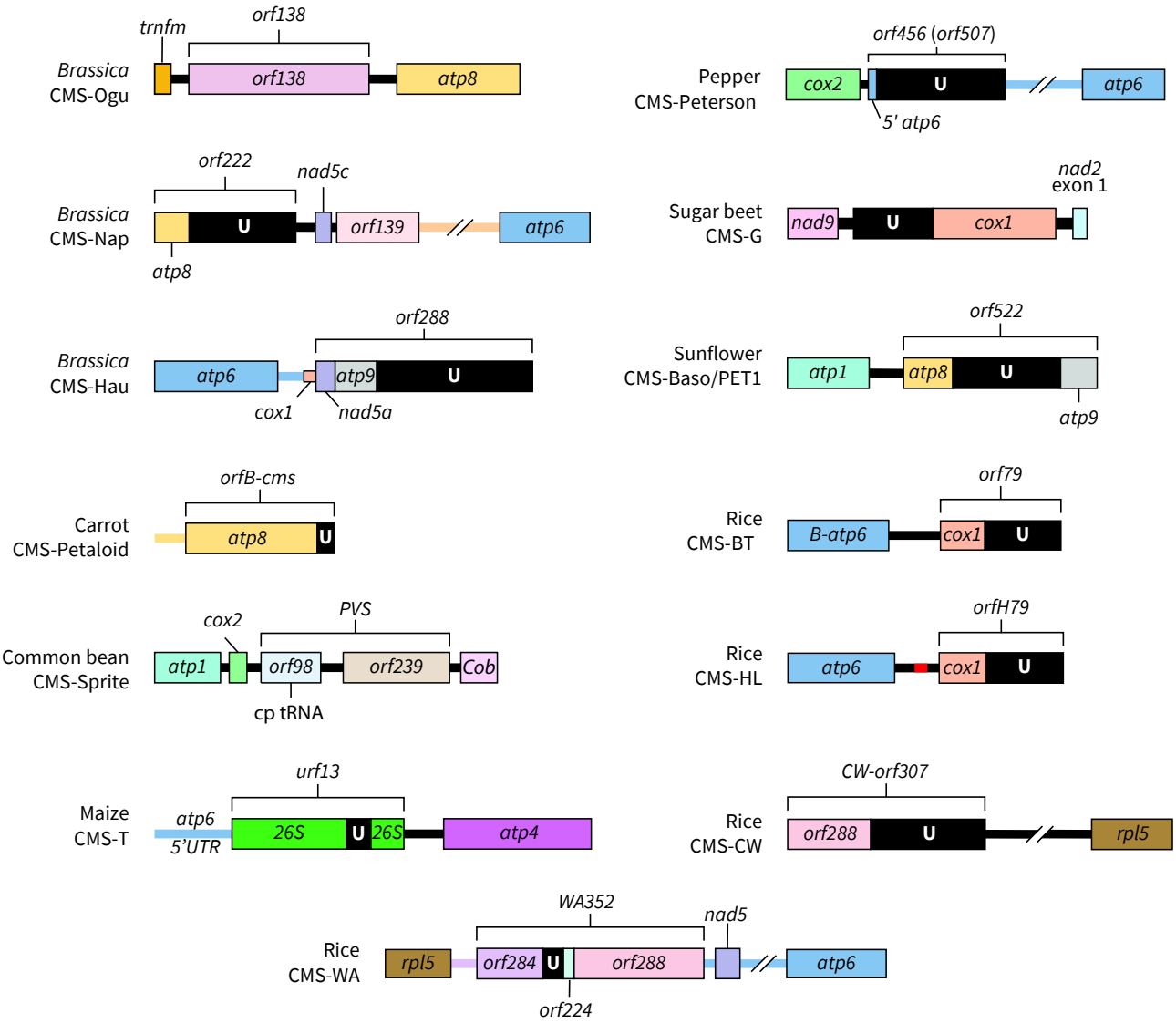


Figure I.13. Examples of CMS-associated chimeric genes. *Orf*s are listed by the convention of the number of codons, with a historical exception (*urf13* encodes a 13-kD protein and *orf522* encodes 522 nucleotides rather than 522 codons). Coding sequences are represented by boxes and flanking regions by horizontal lines. Black boxes (**U**) represent DNA sequences of unknown origin; *cp*: chloroplast; UTR: Untranslated region. Adapted from Chen and Liu, 2014.

the type of CMS that studied, but not on direct biochemical evidence (Chen and Liu, 2014). Historically, the first discovered CMS protein was URF13, which induces sporophytic sterility in the maize **CMS-T** system. It is a 13 kDa membrane protein resulting from hypothetical recombination events involving *atp6, atp4* and the *26S* rRNA gene (**Figure I.13**) (Dewey et al., 1986, 1987). URF13 has been expressed in bacteria (Dewey et al., 1987; Korth et al., 1991) or in eukaryotic cells (Korth and Levings, 1993) with the same conclusive toxicity. Another historical dicot CMS system is the one from sunflower **CMS-PET1**, which is caused by the expression of the chimeric ORF *orf522* originated from a hypothetical recombination event involving *atp8* and *atp9* (**Figure I.13**). *orf522* codes for a toxic membrane protein acting in the tapetum during the development of pollen, action that can be counteracted by the restorer protein encoded by the *Rf1* gene (Köhler et al., 1991; Laver et al., 1991; Horn et al., 2003). The toxic effect of ORF522 on *E. coli* was also shown (Nakai et al., 1995). Similar toxicity on bacterial growth has also been demonstrated for ORF138 from radish **CMS-Ogu** (Duroc et al., 2005), for ORF288 from *B. juncea* **CMS-Hau** (Jing et al., 2012) and for ORF79 and ORFH79, which share the N-terminus of COX1, from rice **CMS-BT** and its variant **CMS-HL**, respectively (Wang et al., 2006). These cytotoxic proteins share a transmembrane domain and a hydrophobic region and they lead to mitochondrial dysfunction in particular cells with a crucial role in pollen production, such as the ones of the tapetum (Levings, 1993). However, the reason why these cells are particularly sensitive remains unknown.

5.3.2. The energy deficiency model

According to the energy deficiency model (**Figure I.12**) the CMS protein causes mitochondrial dysfunction and a deficit of energy to the cell, which consequently fails to reach full male reproductive development (Chen and Liu, 2014). This model is based on the fact that the mitochondrial inner membrane is the home of the mtETC and that the majority of the CMS proteins characterized up to now have transmembrane domains. The mtETC relies on a proton gradient that needs an intact inner membrane to be correctly set. Moreover, most CMS proteins share a portion of their sequence with genes involved in the respiration pathway and they may interact with other subunits from mtETC complexes and therefore reduce their efficiency (Chen and Liu, 2014). Evidence for this model of how CMS works was found in sugar beet **CMS-G,** where genes coding for mutant proteins have been identified, one is a NAD9 protein (a core subunit of complex I) with a C-terminal extension and the other is a COX1 protein (a subunit of complex IV) with a large N-terminal extension, whose gene is co-transcribed with *NAD9*, while in the fertile line *COX1* is located downstream *COX2* (**Figure I.13**) (Ducos et al., 2001; Meyer et al., 2018). **CMS-G** mutants show a

reduction in cytochrome *c* oxidase that correlates with a lower activity of complex IV, providing a molecular link between mtETC activity and CMS (Ducos et al., 2001). More recent publications on pepper **CMS-Peterson** and rice **CMS-HL** go in the same direction. In pepper **CMS-Peterson** there are two analogs of the CMS gene, *orf456* and *orf507*, co-transcribed with *cox2* and correlating with a reduced cytochrome *c* oxidase activity (Ji et al., 2013). The rice **CMS-HL** encodes a small protein, ORFH79, with a N-terminus similar to COX1, that interacts with the nuclear-encoded subunit P61 of complex III, leading to a decrease in ATP concentration and an increase in ROS production (Wang et al., 2013).

5.3.3. The aberrant programmed cell death model

Programmed cell death (PCD) is the process of controlled and organized destruction of cells, which is in part controlled by mitochondrial-derived signals (Reape and McCabe, 2010). A major mitochondrial triggering signal of PCD is the release of cytochrome *c* into the cytosol (Liu et al., 1996). During pollen development PCD is required in the degeneration of the tapetum cells, and a defect in PCD compromises microsporogenesis and leads to male sterility (**Figure I.12**) (Kawanabe et al., 2006). In sunflower **CMS-PET1** ORF522 causes premature PCD of the tapetal cells, through a release of cytochrome *c* from mitochondria (Balk and Leaver, 2001). Moreover, ORF522 has been shown to be involved in the reduction of F_1F_0-ATP synthase and cytochrome *c* oxidase activities (Balk and Leaver, 2001; Sabar et al., 2003). In rice, premature PCD has been shown in the **CMS-WA** and **CMS-HL** systems. The **CMS-WA** system correlates with the expression of the WA352 protein, which accumulates preferentially in the tapetal cells and interacts with the nuclear-encoded protein COX11 (Luo et al., 2013). COX11 is highly conserved in eukaryotes and has been proposed to play a role in hydrogen peroxide degradation. It acts as a negative regulator of PCD, besides of being an assembly factor for cytochrome *c* oxidase (Banting and Glerum, 2006; Veniamin et al., 2011). The interaction between WA352 and COX11 leads to the inhibition of COX11 function in peroxide metabolism, and to premature PCD in tapetal cells (Luo et al., 2013). In the **CMS-HL** rice system premature PDC has also been observed, in anthers, and linked with excess production of ROS (Li et al., 2004).

5.3.4. The retrograde regulation model

It is commonly accepted that the mitochondrial genotype influences nuclear gene expression through retrograde signaling (Fujii and Toriyama, 2008). The retrograde regulation model proposes that the mitochondrial chimeric proteins responsible for CMS regulate genes

involved in male fertility (**Figure I.12**) (Chen and Liu, 2014). Such a regulation has been shown in carrot **CMS-Petaloid** (Linke et al., 2003), which is a homeotic CMS system where the expression of two homeotic genes is affected, leading to the stamens to be replaced by petals (Linke et al., 2003). A different retrograde regulation has been identified in rice **CMS-CW,** in which the chimeric gene *orf307* was identified as CMS gene candidate (Fujii et al., 2010). This CMS system involves a nuclear gene *Rf17* encoding the mitochondrial protein RMS (for Retrograde-regulated Male Sterility) (Fujii et al., 2007). Through a mechanism that remains unclear the CMS-CW cytoplasm up-regulates the allele *rf17,* but has no effect on the *Rf17* allele encoding RMS. This suggests that the increase in *rf17* expression maintains the gametophytic sterility (Fujii and Toriyama, 2009), while the reduced expression of RMS would restore fertility.

5.4. Mechanisms of fertility restoration

5.4.1. At the genomic level

CMS is a consequence of the dynamics of the mitochondrial genome, that by the reshuffling of sequences by recombination leads to the creation of chimeric ORFs that might be expressed. This same dynamic process can also lead to the spontaneous reversion to fertility (**Figure I.12**), by substoichiometric shifting of the relative copy numbers of the subgenomic molecules that contain the CMS genes (Chen and Liu, 2014). The first example of such a reversion has been found in common bean **CMS-Sprite,** caused by the *PVS* sequence located downstream of *atp1* (Chase and Ortega, 1992; Johns et al., 1992) and restored to fertility by the dominant nuclear gene *Fr*. In the presence of *Fr*, the *PVS*-associated sequence is reduced to a substoichiometric level, but still transmitted to the following generations that are male fertile, regardless of the presence of *Fr* (Mackenzie and Chase, 1990; Abad et al., 1995; Janska et al., 1998). Another example of control of CMS by substoichiometric shifting has been reported in millet **CMS-A1,** which is prone to spontaneous reversion to fertility. In this case, a subgenomic molecule containing the junction region between two copies of *cox1* (*cox1-1-2* region) was amplified 10-fold and triggered a recombination cascade, including an ectopic recombination through a microhomology of 7-bp, to produce a novel *cox1-3-2* unstable junction. This unstable junction is then stabilized by recombination to form a new copy of *cox1* followed by a deletion (Feng et al., 2009).

5.4.2. At the post-transcriptional level

Mitochondrial transcripts usually undergo post-transcriptional modifications such as editing, splicing and cleavage. Expression and sequencing analyzes showed that to be also the case

for CMS-associated transcripts, which are processed by different post-transcriptional mechanisms, including editing, polyadenylation, cleavage and degradation mediated by *Rf* gene products (Chen and Liu, 2014). RNA editing is the deamination of a specific cytidine (C) residue of the RNA sequence into uridine (U). It affects both plastidial and mitochondrial transcripts, but has a major impact in mitochondria, where it affects more than 500 positions in mitochondrial transcripts (Shikanai, 2006). RNA processing by cleavage can be exonucleolytic or endonucleolytic and may occur in the spacer sequences of polycistronic transcripts, in 5' and 3' UTRs, and in the coding regions.

In maize **CMS-T** the fertility is restored at the post-transcriptional level by *Rf1* which processes the *urf13-orf221* dicistronic transcript and causes the reduction in abundance of the cleaved *urf13* RNA fragment (Kennell and Pring, 1989). The fertility in the millet **CMS-A3** system is restored by the nuclear gene *Rf3.* The transcript associated, *orf107*, has four editing sites and among them two sites are differently edited. In sterile plants these sites are 81 % and 61 % edited, while in fertile restored plants these rates drop to 41 % and 10 %, respectively, and the highly edited transcript was shown to be rapidly degraded in the lines carrying *Rf3* (Pring et al., 1998; Tang et al., 1999). In *Brassica* **CMS-Pol**, the dicistronic CMS-associated transcript *orf224-atp6* is cleaved in the restored lines carrying *Rfp,* and consequently the *atp6* specific transcript increases in abundance (Li et al., 1998). The mechanism of fertility restoration is quite similar in *Brassica* **CMS-Nap**, where the CMS-associated transcript o*rf222-nad5c-orf139* is processed by *Rfn*, another allele of the *Rfp* locus (Brown, 1999).

5.4.3. At the translational or post-translational level

In some systems of fertility restoration the abundance and size of the CMS-associated mRNA are not affected, suggesting that restoration acts at the translational or post-translational levels (**Figure I.12**) (Chen and Liu, 2014). Among them the one of *Brassica* **CMS-Ogu** is the best documented (Bellaoui et al., 1991). The effect of the **CMS-Ogu**-associated protein ORF138 is restored by several *Rf,* including the PPR protein PPR-B, encoded by *Rf0* (Brown et al., 2003; Uyttewaal et al., 2008). In the anthers, the accumulation of the *orf138* mRNA is not affected, while expression of the ORF138 protein is suppressed by PPR-B that binds the mRNA and blocks its translation (Uyttewaal et al., 2008). Three others restorer systems have been reported that also revealed unchanged levels and processing of the CMS-associated mRNA, while the accumulation of the proteins is suppressed: the ones from common bean **CMS-Sprite** (Sarria et al., 1998), *Brassica* **CMS-Tour** (Landgren et al., 1996) and rice **CMS-WA** (Luo et al., 2013). In the **CMS-WA** system the

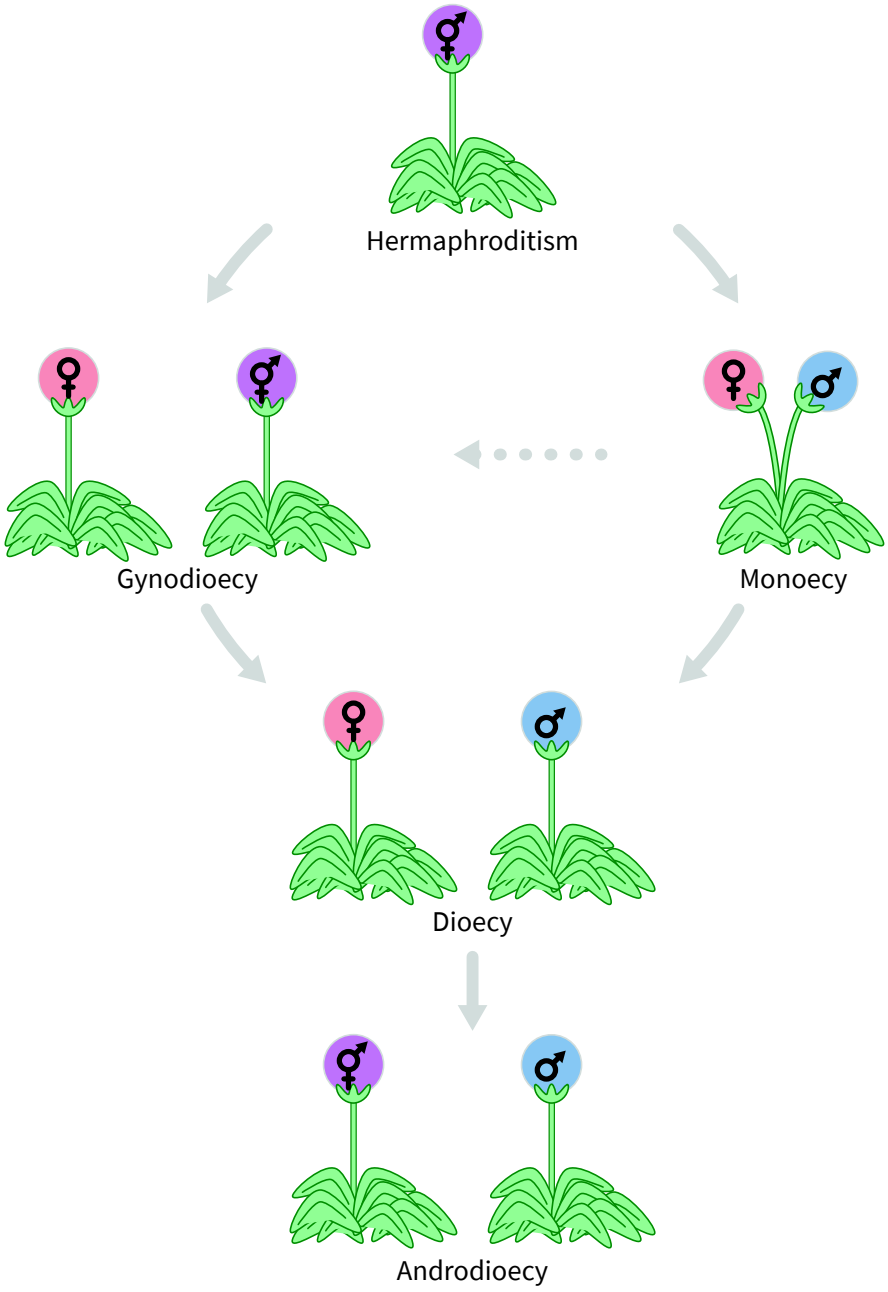


Figure I.14. Evolution of gender dimorphism in flowering plants. There are two main pathways from hermaphroditism to dioecism. The first pathway requires male sterility genes in a hermaphrodite population that allows female plants to appear, resulting in a gynodioecious population. Female fertility is then genetically reduced in hermaphrodites, which gradually become male plants, resulting in dioecy. The second pathway involves sterility mutations that produce unisexual flowers. This pathway is less known and characterized, but it appears that monoecy has evolved numerous times from hermaphroditism. The dashed line represents a hypothetical gynodioecious intermediate stage from monoecy. Finally, in rare cases like in *Datiscaceae*, female plants showed a conversion to hermaphrodite through mutation. Adapted from Barret, 2002.

CMS-associated membrane protein WA352 has an effect at the gametophytic stage. In this system fertility can be restored two ways: RF4 acts on the processing of mRNA *rpl5-WA352* while RF3 acts at the post-transcriptional level (Zhang et al., 1997; Luo et al., 2013).

5.4.4. At the metabolic level

When restoration of fertility acts at the metabolic level, the protein encoded by the *Rf* gene does not act directly on the CMS-responsible protein, but on its consequences. The maize **CMS-T** and its restorer protein RF2 might be an example of such a restoration system (Chen and Liu, 2014). The RF2 protein is an aldehyde dehydrogenase (Cui et al., 1996; Liu et al., 2001). These enzymes oxidize aldehydes to carboxylic acids thus preventing aldehyde damages to cells and tissues. However, in the presence of RF2 neither the *urf13-orf221* mRNA nor the URF13 protein changed, suggesting that RF2 may restore fertility by eliminating harmful molecules caused by the expression of *URF13* (Liu et al., 2001).

5.5. CMS and evolution

Cytoplasmic male sterility is the illustration of the complex relationship between the nucleus and mitochondria. Both are adapted to each other through retrograde and anterograde regulation mechanisms, leading the nucleus (via the *Rf* genes) to hide the deleterious effects of mitochondrial CMS-associated genes. However, the maintenance of these sterility genes in wild populations supports the idea that CMS may provide some selective advantage, a gain of function for the plant, at least under certain circumstances, especially because the nuclear genome does not take advantage of the female state (Lewis, 1941). Indeed, in a female context the nuclear genome is forced to split while in an autogamous context the nuclear genome is propagated.

The CMS phenotype can be seen as a way to force outbreeding, since CMS plants are no longer autogamous. In a natural population outbreeding is a way to acquire new genetic materials and generate a progeny possibly more adapted to the environment. Moreover, in certain species such as in radish the progeny is subject to inbreeding depression, which is the reduction in vitality and fertility of inbred offsprings, as compared to outbred offsprings.

CMS might be a key to sex evolution in higher plants (**Figure I.14**). Indeed, CMS is a way to switch from a hermaphrodite population, able to propagate by self-cross, to a gynodioecious population, in which both female and hermaphrodites are present (Budar et al., 2003; Dufaÿ et al.,

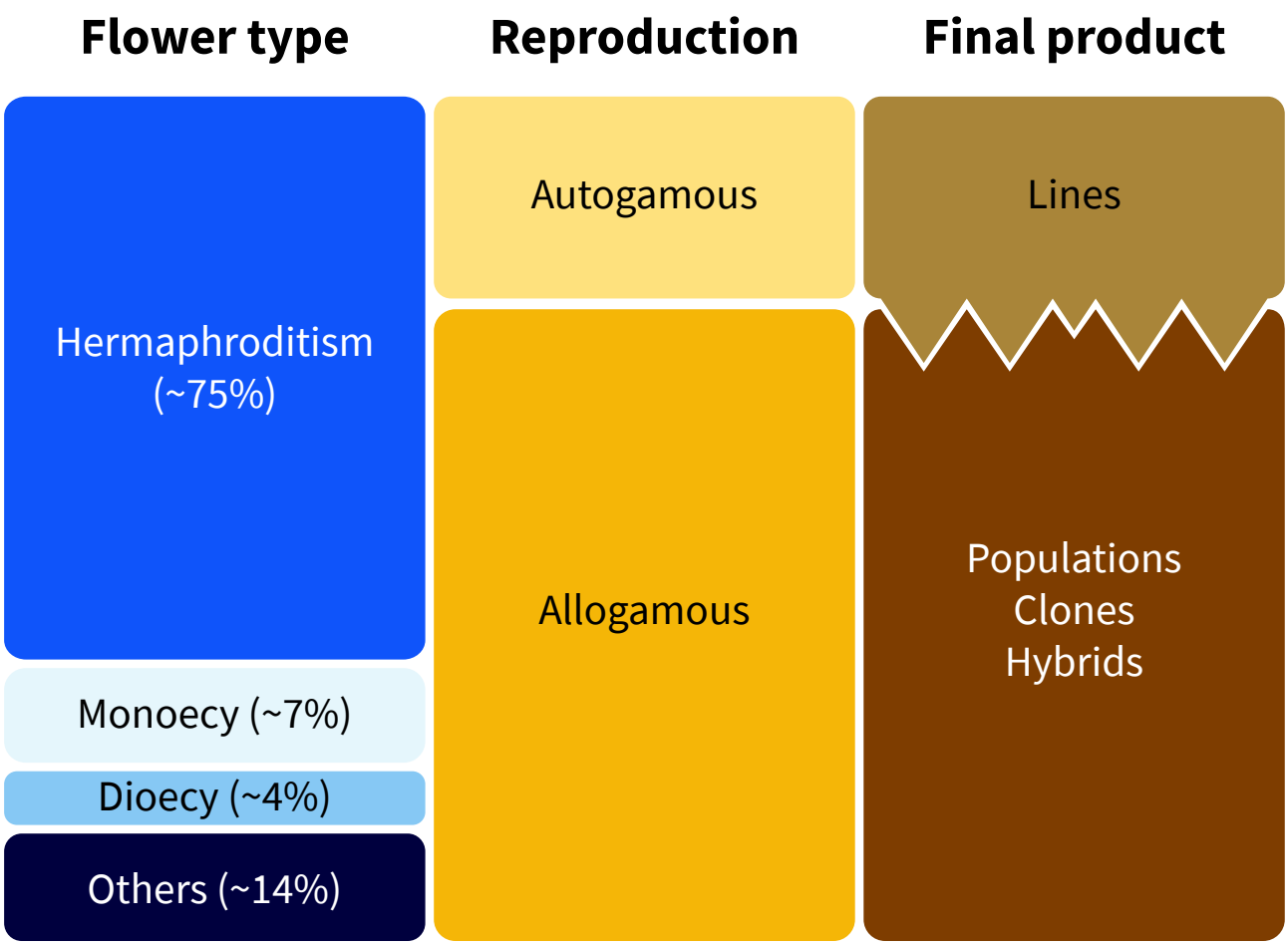


Figure I.15. General combination of flower types in crops related to their reproduction systems. Within vegetable crops, breeding systems are adjusted to plant specificities. Hermaphrodite plants, which represent 75 % of crops, can be autogamous or allogamous, while monoecious, dioecious and the other types reproduce only through allogamy. Allogamous plants are suitable for a hybrid selection system, while autogamous plants only in rare cases. The breeding of autogamous plants mainly goes through the lines selection system.

2007). It has also been shown that specific traits, like disease resistance, can drive a gynodioecious population towards dioecy (Miller and Bruns, 2016).

5.6. CMS in plant breeding

5.6.1. The CMS system

A complete CMS system involves three linked lines. The first line, or A-line, is the male-sterile carrying the sterilizing cytoplasm. The second line, or B-line, is the maintainer that has the exact same nuclear genetic content, but within a fertile cytoplasm context. This line is called maintainer because it is used as a father to produce seeds from the A-line. Due to the maternal inheritance of the cytoplasm the whole progeny of the cross between the A and B lines will carry the sterilizing cytoplasm, showing 100 % male sterility. The third line of the CMS system, or C-line, is the restorer line carrying the *Rf* gene(s). The C-line is highly important in a CMS system where the markable part of the plant results from a cross (fruits like sweet peppers and cereals like wheat). Indeed, the cross between the A and C lines leads to a male-fertile progeny able to self-pollinate and consequently allows the production of the desired product, in such case fruits or seeds.

5.6.2. The interest of CMS system in plant breeding

The extent of genetic variation in a vegetable population often relates to its breeding system. Plants are either self- or cross-pollinated, or a mixture of the two mechanisms may operate in a single plant or species. Self-pollination, or autogamy, is the transfer of pollen from the anthers to the stigma of the same flower or of another flower on the same plant, by a pollinating agent. Cross-pollination, allogamy, involves the transfer of pollen from an anther of one plant to the stigma of a flower on another plant. Plants with high levels of cross-pollination show greater genetic variation than those that are self-pollinated (prone to line production), as there is more opportunity for nuclear gene recombination and gene flow within the population.

Plants have evolved a wide range of outbreeding mechanisms that place restrictions on self-fertilization, with the overall effect of increasing cross-fertilization and hence genetic diversity. This maintains a high level of heterozygosity within a population and avoids the phenomenon called inbreeding depression, which relates to the accumulation of alleles that are deleterious at homozygous state (including self-incompatibility alleles). The main outbreeding mechanisms are spatial or temporal separation of sexes, sexual incompatibility and male or female sterility.

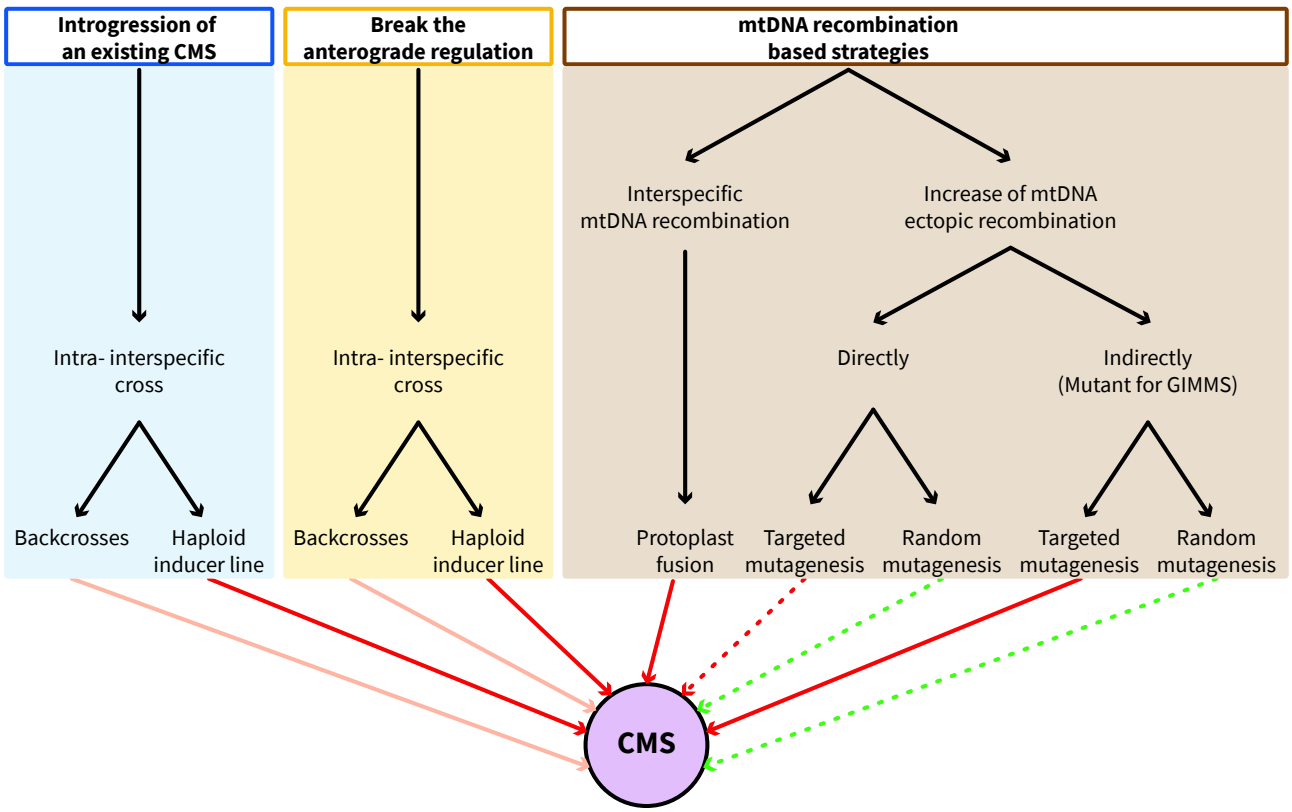


Figure I.16. Strategies to obtain CMS lines. Three main strategies can be used to obtain cytoplasmic male sterile lines (CMS). The first one (in light blue) relies on the cross with another CMS system and requires several backcrosses to obtain alloplasmic lines. However, if interactions with the chloroplast genome induces a severe development phenotype, protoplast fusion might be required to revert it (not shown). The second strategy (in light yellow) aims to break the relationship between the cytoplasm and the nucleus. To do so, two methods are possible: the cross with another cytoplasm, which also requires several backcrosses to obtain alloplasmic lines, or the use of a haploid inducer line. The third strategy (in light brown) relies on mtDNA recombination. One possible approach involves recombination between different mtDNA genomes that co-exist in the cell after interspecific protoplast fusion. Another possibility is to increase ectopic mtDNA recombination. Either directly, by using genetic tools to specifically induce cuts in the mtDNA (restriction enzymes, TALEN endonucleases) or by using genotoxic agents that unspecifically induce breaks in the mtDNA. Or indirectly, by using mutants of genes involved in mtDNA maintenance and segregation (GIMMS) obtained through random or targeted mutagenesis. Red arrows represent strategies and methods that might be considered as GMO; Orange arrows represent strategies and methods that might take a long time due to multiple backcrosses; Dotted lines stand for methods that increase mtDNA recombination but no resulting CMS was observed yet.

A line breeding system is suitable for crops that are not suffering from inbred depression and which are preferably autogamous, while a hybrid breeding system is used for crops where a heterosis effect is needed and for plants preferably allogamous (**Figure I.15**). Allogamy enhances diversity conservation and avoids inbreeding. It can be associated with populations as well as with hybrids. An allogamous breeding system requires some characteristics: a pollinating agent and a limitation of selfing. The pollen can be transported by the wind (anemophily) or by an insect vector (entomophily), which needs to have its active periods synchronized with the flower receptivity. There are several ways to limit selfing, such as monoecie (ex: squash, melon, maize…), flowers strictly male or female and both on the same plant; dioecie (ex: spinach, asparagus…), flowers also strictly male or female but on separate plants; dichogamy (ex: oregano…), flowers male and female on the same plant but flowering at different times; self-incompatibility (ex: cabbages…) and male sterility. Among the male-sterility systems, CMS and its uniform transmission of the male-sterile phenotype to progeny is a leading choice. Indeed, because of the mitochondrial origin of the male-sterile phenotype, all the progeny is male sterile if the paternal line does not carry any restorer fertility gene(s). Moreover, the uniform transmission of the male-sterile phenotype allows the commercial protection of the varieties which are not able to produce seeds in the absence of maintainer or restorer fertile line(s).

However, it is important to distinguish species where the fruit (pepper, tomato) or the seed (rice, maize, sunflower) are the marketable part of the plant from the species where it is the vegetative part (carrot, lettuce). Indeed, for the first ones the profitable part comes from the cross, that is why a restorer line with traceable *Rf* is needed, which is not the case when the marketable part doesn't come from a cross.

5.7. Obtention of CMS lines

Since its identification the CMS phenotype has been introduced in several species, following different strategies.

One strategy is to directly introduce a sterilizing cytoplasm in a new line, by crossing (**Figure I.16**). The sterilizing cytoplasm could origin from another crossable species or from another ecotype/line from the same species. The cross of the two parental lines leads to a hybrid with the desired CMS-inducing cytoplasm and a nuclear genome inherited from both. The unwanted nuclear-half can be removed by backcross. This strategy has been used for **CMS-Ogu**, which

originated from radish but is widely spread in rapeseed. To do so, radish **CMS-Ogu** has been crossed with rapeseed, followed by several backcrosses to obtain alloplasmic lines containing the **CMS-Ogu** cytoplasm and the rapeseed nuclear genome. These alloplasmic lines showed a deleterious phenotype (male-sterile but chlorotic plants), due to negative interaction between the nuclear and cytoplasm genomes, which was counteracted by protoplast fusion with rapeseed (Pelletier et al., 1983). The cybrids then obtained conserved the CMS trait, but dissociated from the cytoplasmic genome responsible for the severe phenotype (radish chloroplast). But the unwanted nuclear-half can be also removed by spontaneous androgenesis. This strategy has been used to introduce a sterilizing cytoplasm in tobacco through selection of spontaneous haploids (Horlow et al., 1993). In this case, selected plants only carried the paternal chromosome set. Chromosome doubling was performed using three methods: pollination, in vitro culture or acenaphthen treatment. Finally, spontaneous androgenesis can be be simulated by using a haploid inducer line (HIL). A HIL leads to the entire paternal chromosome set to be removed in the zygote. The zygote will do then a copy and paste of the remaining haploid genome leading to a homozygous diploid genome through a process that is called *in vivo* gynogenesis (Gilles et al., 2017)."

A second strategy is to break the anterograde regulation (**Figure I.16**). Indeed, CMS-associated genes are probably widely present in natural populations, but their effect is suppressed by anterograde control from the nucleus. To (re)activate CMS, the relationship between the nucleus and the mtDNA needs to be perturbed. This does not require an existing CMS. And so, the idea is to cross two distinct ecotypes or accessions (same species or not) with the aim of putting a cytoplasm in another nuclear context (or *vice versa*) after several backcrosses. Such a work has been done in *Mimulus guttatus* (Fishman and Willis, 2006; Case and Willis, 2008) as well as in *A. thaliana* (Gobron et al., 2013; Roux et al., 2016) and in sunflower, discovered by Serieys and Vincourt 1987 and characterized by the work the Pr. Renate Horn (Horn, 2002; Horn et al., 2003). As for the introduction of an existing CMS, such a strategy could be facilitated by using a haploid inducer line which does not require backcrosses (Wijnker et al., 2012). The haploid inducer line is used as a mother, leading to the loss of the maternal nuclear genotype while keeping only the paternal nuclear genotype, as demonstrated in *A. thaliana* (Flood et al., 2020) (**Figure I.16**). This process allows a quick introduction of the parental nuclear genome within another cytoplasmic context. Moreover, one of the cybrids then produced showed cytoplasmic male sterility in a specific cytoplasm/nucleus combination (Flood et al., 2020).

These two strategies rely on the relationship between the nuclear and cytoplasm genomes, but approaches based on the induction of cytoplasmic diversity through mitochondrial recombination remain largely unexploited, despite good promises. So, the idea behind this strategy is to induce the creation of new mitochondrial orfs that could be the origin of new CMS lines. To do so, there are two possibilities. The first involves recombination between sequences from different mitochondrial genomes put together by protoplast fusion. This was successfully attempted on protoplast fusions between sunflower CMS-Pet1 and *Cichorium intybus* and *C. endivia* (Dubreucq et al., 1999; Cappelle et al., 2007). Such a strategy has been also applied in *Nicotiana sylvestris* where CMSI and CMSII have been obtained through protoplast regeneration, and show large deletions in their mtDNA (Li et al., 1988; Chetrit et al., 1992).

The second strategy to induce the creation of new CMS-associated orfs is to increase mtDNA ectopic recombination. Indeed, mutants for nuclear genes involved in mitochondrial genome maintenance and segregation showed a shift in the stoichiometry of their mtDNA in *A. thaliana*. These mutants are *msh1*, coding for a homolog of bacterial MutS (Abdelnoor et al., 2003), *osb1* (Zaegel et al., 2006), *radA* (Chevigny et al., 2019), *recA2* and *recA3* (Miller-Messmer et al., 2012), *recG1* (Wallet et al., 2015) and *why2* (Cappadocia et al., 2010). Furthermore, a study showed that suppression of the expression of *MSH1* by RNA interference (RNAi) leads to mtDNA rearrangements and CMS in tobacco and tomato (Sandhu et al., 2007) (**Figure I.16**).

6. Lettuce breeding

6.1. Challenges of lettuce breeding

Over the past 60 years, lettuce breeding was mostly oriented to the development of varieties adapted to climate and growth conditions allowing lettuce production all over the year (Doré and Varoquaux, 2006). Nowadays, breeding is still oriented to the adaptation to growth conditions, with the development of new cropping systems (hydroponics, vertical farming …). But modern lettuce breeding is also oriented to new characteristics, such as conservation in the field (giving to the producer a certain latitude to harvest), leaf shape (wilting is an important trait for mechanical harvest and for the consumers), resistance to transport (post-harvest challenges), diversification, which led to new varieties or typologies to fit market trends, and resistance to phytopathogens. Among the lettuce pathogens, *Bremia lactucae* that causes the downy mildew is the major threat to

lettuce production (the International Bremia Evaluation Board). All types of lettuce are potential hosts to *B. lactucae,* which develops on the dorsal face of the leaves. The transmission of the pathogen from plant to plant is very fast, and once infected plants are unsuitable for sale, which explains the severe damages to the crops (Doré and Varoquaux, 2006). Furthermore, *B. lactucae* is known to overtake resistance rapidly, by mutations leading to new strains. That is why breeders looked for new resistance genes carried by wild lettuce species, but that remains difficult to introduce in cultivars through interspecific crosses. But downy mildew is not the only lettuce disease. Indeed, the fusarium wilt, caused by the fungi *Fusarium oxysporum f.* sp. *lactucae*, the corky root, caused by the bacterial pathogen *Rhizomonas suberifaciens,* and viral infections, such as by Lettuce Mosaic Virus (LMV) or Tomato Spotted Wilt Virus (TSWV), also concern the breeders (Davis et al., 1997; Thicoïpé, 1997; Doré and Varoquaux, 2006).

6.2. Interest in mitochondrial genome diversity

In breeding, nuclear diversity remains currently the most exploited sink of biodiversity. But the exploitation of mitochondrial genetic diversity might present new opportunities. Among the traits that can result from this diversity CMS would be the most valuable in lettuce breeding. Indeed, lettuce CMS could be a breakthrough technology in lettuce breeding.

Lettuce is not an easy plant to cross, because of its autogamous reproductive system and the short daily period of opening of the inflorescences. This is doable by emasculating the flowers before they open, but a laborious work on a large scale. A CMS lettuce system would render emasculation unnecessary, because plants would not produce viable pollen anymore. Such a tool would lead to easy crosses in lettuce and would allow the wide introgression of genes present in wild species, which necessitates many backcrosses to segregate away the non-desired gene pool of the wild parent. Moreover, these genes are mostly clustered in the lettuce nuclear genome and, consequently, require a very large number of crosses to be properly introgressed.

CMS would be key to the setting of a lettuce hybrid system. Hybrid systems are highly desired by breeders to profit from the heterosis effect. However, such a system is already possible in lettuce and it has been patented (US20140298546A1). The hybrid system possibilities in *L. sativa* rely on a panel of 7 GMS genes, all recessive (*ms1*, *ms2*, *ms3*, *ms4*, *ms5* and *ms6*) with the exception of *Ms7* which is dominant (Lindqvist, 1960b; Ryder, 1999). But even if hybrids resulting from this GMS system do benefit from heterosis effect, or combine characters or genes at the same locus, male

fertile progenies are still segregating with male-sterile progenies. Consequently, the protection of the varieties is only possible with a CMS system.

7. Our Project

This study was closely coordinated with an industrial partner, the breeding company Enza Zaden, through a Convention Industrielle de Formation par la Recherche (CIFRE) project.

7.1. Few words about Enza Zaden

Enza Zaden is an independent family company specialized in vegetable seeds breeding. The company was founded in 1938 by Jacob Mazereeuw as “De Enkhuizer Zaadwinkel” (The seed shop of Enkhuizen), now known by the abbreviation Enza Zaden. In the beginning Enza Zaden was a seed marketing company, but when Jacob’s son Piet Mazereeuw took his place in 1959 he started plant breeding, to better attune the products to the growers and market needs. Tomato Extase was his first creation and it was a great success. In 1985 the company developed internationally, in Europe, India, China and Brazil, for example. The headquarter is still located in Enkhuizen, in the Netherlands. The current managing director of Enza Zaden is Jaap Mazereeuw, Jacob’s grandson.

Nowadays there are 43 subsidiaries including 3 joint-ventures in 24 countries present on all continents. The company employed 1800 people in 2017. The turnover of Enza Zaden grew from 47 M€ in 2000 to 290 M€ in 2016. The company develops varieties of vegetables of more than 30 international and local species; the seeds are then produced and sold all around the world. In total, Enza Zaden has developed about 1200 varieties of vegetables, from sweet pepper, tomato, cucumber to lettuce. Enza Zaden’s significant investment in research and innovation allows the company to place on the market 100 new varieties of vegetables every year. Indeed, the long-term vision and innovative strength are the essential factors that distinguish Enza Zaden from competitors.

Enza Zaden is also present in France where there is a marketing center, distributing seeds across the country, and two research centers. One research center is located in Allonnes (Maine-et-Loire department), nearby Saumur, and focuses on leafy vegetables (lettuce, chicory). The other is in Châteaurenard (Bouches-du-Rhône department) and is specialized in fruity vegetables (melon, pumpkin, squash).

This PhD project was carried out mainly in collaboration with the research station of Allonnes, and co-supervised by Emilie Guilloteau-Fonteny, a senior researcher at the center. Enza Zaden's main activity in this center is associated with the breeding of lettuce and endive, as well as research programs in radish and lettuce. Fifty people work at this site.

7.2. Geographical repartition of the work

Our project took place at three different locations: two within the walls of Enza Zaden, in the research station of Allonnes in France and in the research & development headquarter of the company in Enkhuizen in the Netherlands; and at the IBMP (CNRS) in Strasbourg (Bas-Rhin).

The work relative to the mutant lines has been done within Enza Zaden. This work included mutagenesis by EMS treatment, identification of interesting mutants by TILLING (targeting induced local lesions in genomes), and their genotyping and phenotyping during the duration of the project. Their growth was insured during summer cycles in Allonnes, on greenhouse tunnels, and during winter cycles in Enkhuizen, on heated greenhouses.

The molecular biology and NGS work (DNA extractions, library preparations and bioinformatics) was done at the IBMP in Strasbourg, in close collaboration with the teams of Enza Zaden.

7.3. Aims of my thesis

The mitochondrial genome diversity within a species or a genus can be studied at two different levels. The first level is to explore the natural mtDNA diversity present within the species, or the genus. In lettuce, a recent study determined the sequence of the mitochondrial genomes of *L. saligna*, *L. serriola* and *L. sativa* (Kozik et al., 2019). However, this study was limited to the genome of a single accession per species and so was not able to explore the full diversity of the *lactuca* mtDNAs. The second level is the study of induced mtDNA diversity. For this purpose, previous studies showed that it is possible to induce mitochondrial genome diversity by using mutants of genes involved in the surveillance of mtDNA recombination. Most of these studies took were conducted in the model plant *A. thaliana,* but such results have also been seen in tobacco and tomato. Consequently, mutants for *msh1* (Sandhu et al., 2007), *osb1* (Zaegel et al., 2006), *osb4* (unpublished data), *radA* (Chevigny et al., 2019), *recA3* (Miller-Messmer et al., 2012) and *recG1* (Wallet et al., 2015) are a leading choice to induce mtDNA diversity in new species.

During my Ph.D. project, my aim was to explore the natural mtDNA diversity within *Lactuca* and to study the mtDNA diversity induced by mutants for genes involved in the surveillance of mtDNA recombination. When this project started, the sequence and organization of the *L. sativa* mtDNA were not available. To have a reference to study mtDNA diversity, both natural and induced, we sequenced and assembled *de novo* the mtDNA of *L. sativa*. This reference mtDNA genome was then used as a tool to study the natural mtDNA diversity present within the *Lactuca* genus through the sequencing and the *de novo* assembly of the mtDNAs from two *L. sativa*, one *L. saligna,* three *L. serriola,* and three *L. virosa* accessions. The mtDNAs were then compared between themselves to explore the natural mtDNA diversity present within the *Lactuca* genus.

The mtDNA of L. sativa was also used to study the diversity induced through two strategies: A direct strategy based on the treatment with a genotoxic agent, ciprofloxacin (CIP), targeting the organellar DNA gyrase and inducing DSBs; and an indirect strategy based on thirteen mutants for genes involved in the surveillance of mtDNA recombination, generated by Enza Zaden by EMS treatment and selected by TILLING. This treatment has been made on an old commercial variety of butterhead lettuce (var. *capitata* L. *nidus tenerrima*) called “Wendel” which is usually used in research by the company for its short cycle (< 6 months). Enza Zaden provided *msh1*, *osb1*, *osb2*, *radA*, *recA3,* and *recG1* mutants. Because these genes had not been studied in lettuce we needed to first confirm their gene structures and mitochondrial targeting, then study their impact on the stability of the mtDNA and the phenotype of the plants. Among these mutants we found that the lettuce *radA-2* knockout mutant is highly affected in its development, and not able to shed seeds, a much severer effect than in *A. thaliana*. We tested if the phenotype correlated with effects on mtDNA transcription.

Among the two mutant lines available for *OSB2*, one presented the particularity of not being able to segregate at the homozygous stage and showed a high proportion of male-sterile plants. We have therefore tried to identify the reasons for the absence of homozygous plants. To do so, we considered the hypothesis of a dominant effect of the mutant allele encoding an aberrant protein, and tested it by translational fusion to GFP, RT-qPCR, RNA sequencing and the heterologous expression of the lettuce mutant allele in *A. thaliana*. Finally, we investigated the origin of the phenotype of male sterility through bulked segregant RNA sequencing (BSR-Seq).

CHAPTER 2: Determination of the mitochondrial genome of *Lactuca sativa*

1. Challenges of plant mitochondrial genome assembly

The sequencing of the mtDNA can be a delicate process in some species. Indeed, even if plants mtDNA are taken as large genomes, they cannot compete in size with nuclear genomes. This is a relative issue in *Arabidopsis thaliana* and its nuclear genome of ± 157 Mb (The Genome Initiative, 2000; Bennett et al., 2003). However, it is a different situation for plants with much bigger nuclear genomes, such as lettuce (2.9 Gb) (Reyes-Chin-Wo et al., 2017). Indeed, based on the *A. thaliana* mtDNA size and cellular content (mtDNA size of 368 kb; 1 genome per 3 mitochondria; around 200 and 300 mitochondria per cell), a plant cell contains about 24.5 Mb of mitochondrial genome sequences, which in lettuce should be about 118 times less than nuclear sequences. This means that from a total genomic sequencing there will be 118 nuclear reads for a single mtDNA read. And so, sequencing an mtDNA of about the size of *A. thaliana* with an Illumina 2 x 150 nucleotides paired-end strategy will roughly require 285 000 reads to reach a 1x coverage. Moreover, the mitochondrial and nuclear genomes are not the only ones in the plant cell, there is also the one from the chloroplast (cpDNA). To apply the same calculation to it, we need first to agree on the number of chloroplasts per cell and the number of cpDNA per chloroplast, and on the literature different values can be found. Here I will consider a mean cell content of about 39 chloroplasts that contain each 58 cpDNA copies, of a size of 152 kb, leading to a cell content of about 343.8 Mb of cpDNA per cell (Rowan et al., 2009). Based on this data, for every mitochondrial read there would be 14 chloroplastic ones. Finally, it would require 3 990 000 reads to reach an mtDNA 1X coverage from a total DNA extraction of lettuce. There are two ways to work around this problem: choose a higher-throughput sequencing strategy (Illumina HiSeq for instance) or enrich the mtDNA concentration by extracting it from purified mitochondria.

But once the sequencing done, the challenge is not finished. Plant mitochondria genomes are known to have highly dynamic and multipartite structures because of large repeated sequences (≥ 500 bp) involved in frequent and reciprocal recombination, which makes their assembly tricky. Indeed, these large repeated regions cannot be crossed by a conventional short-read sequencing strategy. Consequently, these sequences remain contigs during genome assembly that can be linked to four other contigs (2 per side) leading to four possible configurations. There are two

possibilities to overcome such assembly issue. The first relies on determining the stoichiometry of each possible configuration. The second relies on the use of a combination of both short and long-read sequencing strategies.

2. The unexpected scoop: The *L. sativa* mtDNA (Kozik et al., 2019)

When we initiated the project in 2017 the mtDNA was the only genomic sequence missing from the three *L. sativa* genomes. However, a *L. sativa* mtDNA sequence was subsequently published (Kozik et al., 2019) (GenBank accession number **MK642355)**. This sequence and the one we assembled during our work described in this chapter inform us on the mtDNA diversity in *L. sativa*.

In this study, the authors built the mtDNA of *L. sativa* var. *capitata* L. *salinas* (Crisphead, Iceberg, acc. NC_042756). They assembled this genome into a single circular chromosome of 363 324 bp. To do so, they sequenced total DNA by three sequencing strategies: two Illumina HiSeq paired-end short-reads strategies (short insert size), Illumina HiSeq mate-pair (long insert size) and PacBio as long reads strategy. Thanks to those strategies they could identify *in silico* three different isoforms. However, they were unable to detect *in vitro* molecules of the expected length of ~363 kb, but only smaller subgenomes (mean size 113 kb). Based on coverage data they identified four large repeated regions, named M (30.579 kb), R (10.430 kb), N (4.115 kb) and T (3.552 kb). *L. sativa* mtDNA carries the 24 core mitochondrial genes commonly found in higher plants plus 7 additional ones (*rpl5*, *rpl10*, *rpl16*, *rps3*, *rps4*, *rps12*, and *rps13*).

3. Aim of this chapter

This chapter is a tale about the sequencing and assembly of the mtDNA of a *L. sativa* var. *capitata* L. *nidus tenerrima* (butterhead) from an old commercial variety from Enza Zaden named "Wendel". The sequencing and assembly of this genome was the cornerstone of our project. Indeed, its sequence was the start of a phylogenetic analysis about the natural diversity of the *Lactuca spp.* mtDNAs described in **CHAPTER 3**. It also allowed us to analyze the induced mtDNA diversity described in **CHAPTERS 4 & 5**. Consequently, the aim of this chapter was to assemble a Wendel mtDNA sequence as robust as possible.

Mito prep	mtDNA extraction			*Qubit 4 Fluorometer*	DNA extracted (ng)
	Starting material (g of seeds)	Method	Volume (µL)	Concentration (ng/µL)	
#1	0.2	CTAB	20	2	40
#2	12	Phenol:Chloroform	50	92	4600

Table II.1. Quantification of extracted mtDNA from two different mitochondrial purifications (Mito prep).

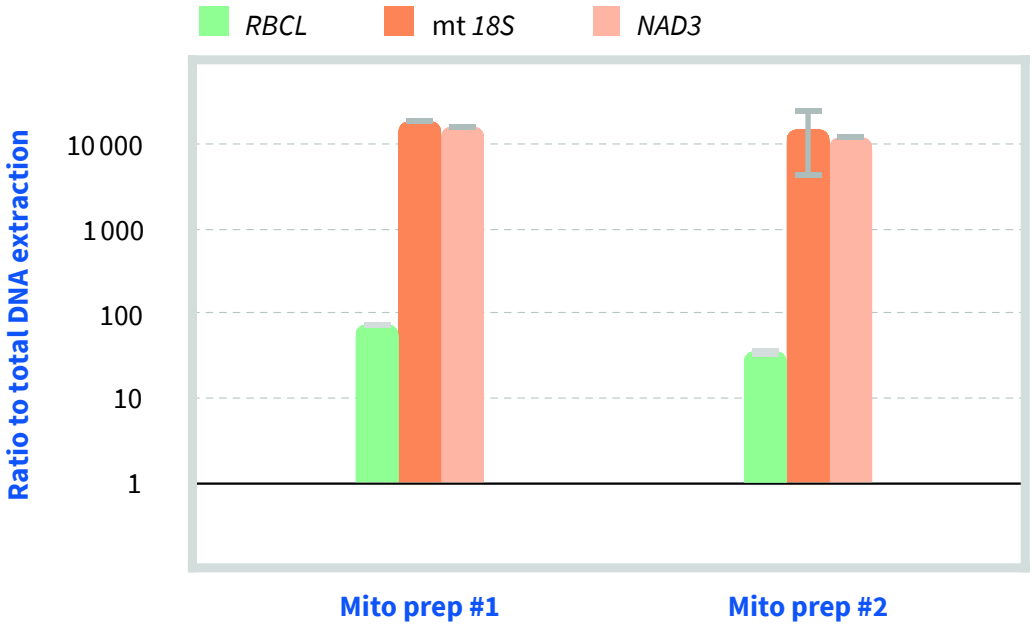


Figure II.1. Relative enrichment of the organellar genomes on two mtDNA extraction experiments. DNA extracted from two independent mitochondrial purification (Mito prep) were analyzed for the enrichment on mtDNA and cpDNA, by qPCR. The cpDNA-encoded *rbcL* gene and the mtDNA-encoded 18S rRNA and *nad3* genes were tested. Results were normalized to the nuclear *18S* gene sequence and are represented in a log10 scale. The chloroplastic gene level is shown in green and the mitochondrial gene levels are shown in oranges. Error bars are the SD values from three technical replicates.

1. Assembly of the *L. sativa* var. *capitata* L. *nidus tenerrima* mitochondrial genome

To build the mtDNA of our butterhead lettuce variety, we chose to extract the mitochondrial DNA from purified mitochondria and to sequence it by using a combination of short and long-reads sequencing strategies, Illumina MiSeq paired-end and PacBio Single-molecule Real-Time (SMRT), respectively.

1.1. Optimization of mitochondrial DNA extraction to *L. sativa*

To reduce sequencing costs and make the *L. sativa* mtDNA sequencing feasible we started by purifying mitochondria. To do so, we optimized to *L. sativa* the protocol of mitochondria purification routinely used in house for *A. thaliana*. We extracted lettuce mitochondria starting from etiolated seedlings, grown in dark for 7 days. We started from different amounts of seedlings, considering the requirements of the sequencing strategy used downstream. Indeed, the Illumina MiSeq paired-end sequencing requires as little as 1 ng of DNA, while PacBio SMRT needs 1 µg of DNA for a maximal insert size of 10 kb. In the end we managed to extract 40 ng and 4.6 µg of DNA from mitochondria purified from ~200 (± 0.2 g of seeds) and ~12 000 (± 12 g of seeds) seedlings, respectively (**Table II.1**). The relative levels of nuclear, plastidial and mitochondrial DNA were next determined by qPCR, to evaluate the mtDNA enrichment in these samples (**Figure II.1**). In both samples the mtDNA was enriched more than 10000 times (~17 000 and ~13 000, respectively) as compared to total leaf DNA, while the plastidial DNA was much less enriched (~75 and ~25, respectively). This confirmed that we could obtain fairly pure lettuce mitochondria.

Both samples were then sequenced by Illumina MiSeq paired-end (2 x 150 bases) and PacBio SMRT, respectively. The library preparation and Illumina sequencing were managed by the IBMP platform *"Gene expression analysis"* and produced 2 078 368 reads with a mean average insert size of about 834 bp. The PacBio SMRT sequencing was done by Keygene (Wageningen, The Netherlands) and led to 3 245 610 reads with a mean size of 3 061 pb.

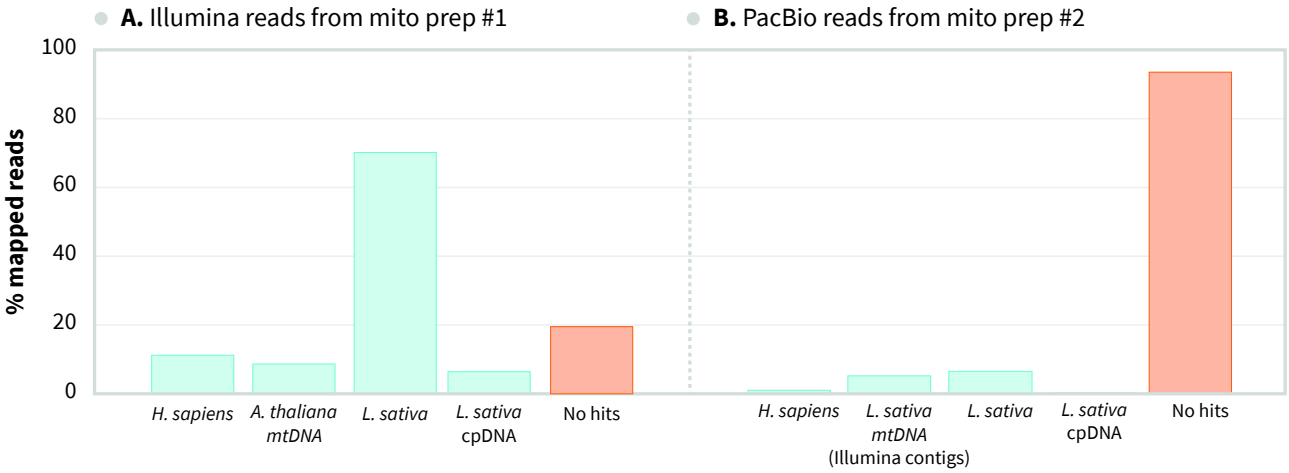


Figure II.2. Graphical output from FastQ Screen after mapping reads from mito preps against several reference genomes. Mapped reads are shown in pale turquoise on the corresponding genome, and unmapped ones are shown in orange in the column “no hits”.

# reads	2 078 368
# contigs	48
# contigs (>= 2 500 bp)	48
# contigs (>= 5 000 bp)	23
# contigs (>= 10 000 bp)	19
# contigs (>= 25 000 bp)	13
# contigs (>= 50 000 bp)	2
Largest contig (bp)	76 886
Total lenght (bp)	757 189
GC %	42.28
mtDNA contigs	21

Table II.2. Assembly statistics of *L. sativa* mtDNA contigs with Novoplasty.

	Size(pb)	Mean Coverage
Contig #01	76'888	1030
Contig #02	3'690	1284
Contig #03	11'986	742
Contig #04	11'031	737
Contig #05	35'835	654
Contig #06	41'967	731
Contig #07	9'881	673
Contig #08	12'776	801
Contig #09	9'478	629
Contig #10	45'897	781
Contig #11	12'376	765
Contig #12	9'542	635
Contig #13	10'723	709
Contig #14	9'118	589
Contig #15	30'730	1267
Contig #16	38'414	750
Contig #17	39'736	719
Contig #18	17'928	677
Contig #19	42'952	675
Contig #20	31'028	869
Contig #21	31'426	889

Table II.3. Mitochondrial contigs built from Illumina reads

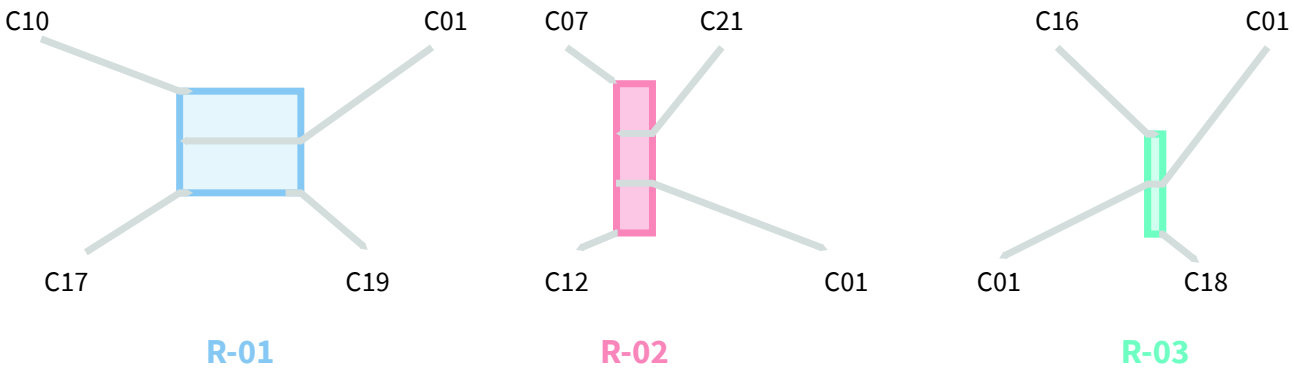


Figure II.3. Identification of large repeated sequences. The overlapping regions allowing the connection between four contigs were identified by BLAST alignment. Doing so, the region R-01 (light blue) is a node between contigs C10/C17-C01/C19, the region R-02 (light red) is a node between contigs C07/C12-C21/C01 and R-03 (cyan) is a node between contigs C16/C01-C01/C18.

# reads	3 245 610
# mitochondrial reads	277 292
# contigs	4
# contigs (>= 10 000 bp)	4
# contigs (>= 25 000 bp)	3
# contigs (>= 50 000 bp)	3
# contigs (>= 100 000 bp)	1
Largest contig (bp)	177 077
Total lenght (bp)	351 476
GC %	45.56

Table II.4. Assembly statistics of *L. sativa* mtDNA contigs with Canu

1.2. Contigs building

Despite a strong enrichment in mtDNA we started the analysis of the sequencing data by evaluating the possible contaminations (**Figure II.2**). To do so, we mapped the reads from both sequencing strategies to several genomes with *FastQ screen* (Wingett and Andrews, 2018). Illumina data showed a high content of nuclear DNA (~70 % of mapped reads) and PacBio sequencing lead to few *L. sativa* sequences (< 10 %). That was due to an important bacterial DNA contamination, probably because of growing the seedlings on wetted blotting paper, in non-sterile conditions.

1.2.1. Contigs building from Illumina paired-end reads

Contigs from Illumina MiSeq paired-end (2 x 150 bases) were built with the help of the IBMP platform "*Bioinformatic and bioimage*". To do so, and because of the potential high contamination in nuclear DNA, the organellar *de novo* assembler *Novoplasty* was used (**Table II.2**) (Dierckxsens et al., 2016). This led to 48 contigs among which 21 were identified as mitochondrial by *BLAST* (**Table II.3**) (Altschul et al., 1990). Moreover, we identified 3 repeated regions based on overlapping regions between 4 contigs: R01 (34 696 bp), R02 (10 430 bp) and R03 (3 552 bp) (**Figure II.3**).

1.2.2. Contigs building from PacBio reads

The high contamination of PacBio data with unidentified sequences makes a complete *de novo* assembly starting from PacBio reads impossible. To use the data we filtered the PacBio reads with the 12 Illumina contigs previously built. By doing so we reduced the number of reads to 277 292 which were then assembled into 4 contigs with the long-reads assembler *Canu* (**Table II.4**) (Koren et al., 2017). Three of those 4 contigs were mitochondrial DNA and the other was a bacterial sequence. The three mitochondrial contigs are long ones with sizes of 177 077 bp (PB01), 90 502 bp (PB02), and 73 475 bp (PB03), respectively. Interestingly, the beginning of the PB01 sequence maps to its end for 5 225 bp. Consequently, we circularized it into a circular chromosome (PBc01) of 171 852 bp. Furthermore, we found that the extremities of contigs PB02 and PB03 mapped to sequences internal to PBc01. But these contigs did not correspond to circular molecules.

1.3. mtDNA assembly

At the end of the contig-building step we had two sets of mitochondrial contigs: 21 contigs from Illumina reads and 3 from PacBio data, among which one corresponds to a circular chromosome. The second copy of the repeated regions R01 and R02 identified within Illumina contigs was not present in PacBio contigs. These two sequences are present in PBc01, and

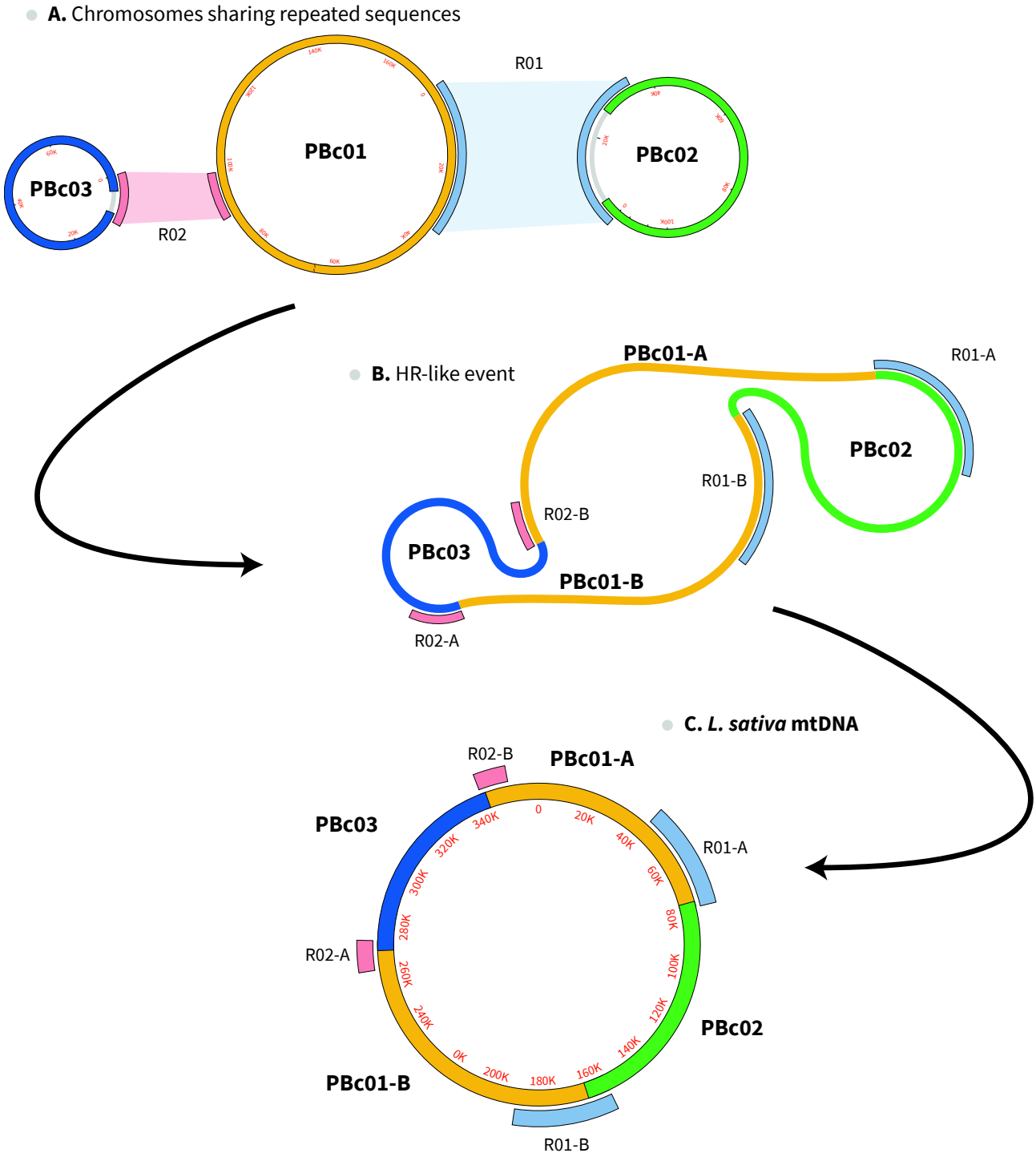


Figure II.4. mtDNA assembly through homologous recombination-like events. (**A**) Subgenomic chromosomes that share repeated sequences. PBc01 (171 852 bp) shares repeat R01 (34'696 bp) with PBc02 (112 968 bp) and repeat R02 (10 430 bp) with PBc03 (78 504 bp). (**B**) The circular chromosomes were then assembled through recombination-like events involving repeats R01 and R02 into (**C**) a single circular chromosome (363 324 bp).

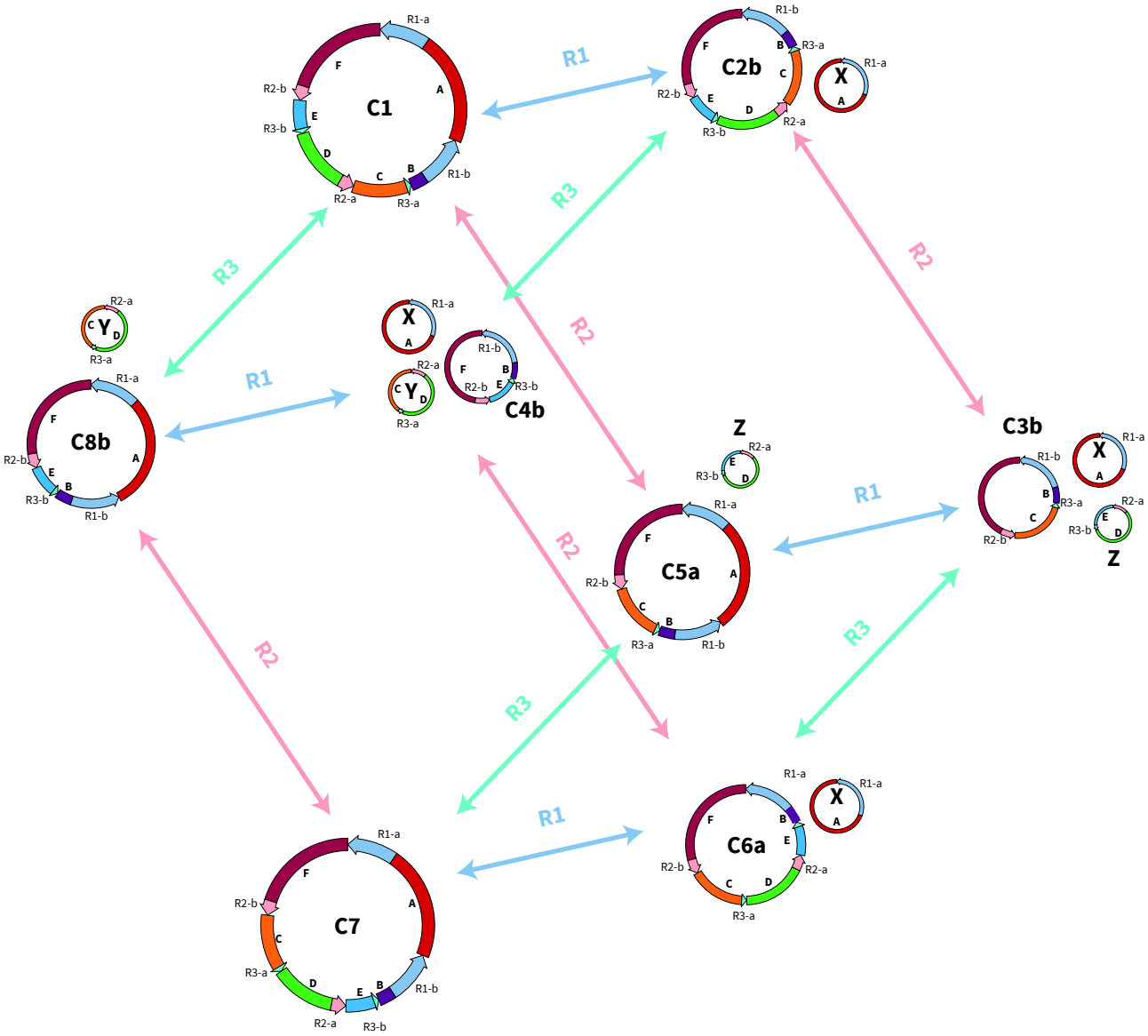


Figure II.5. Alternative configurations of the *L. sativa* mtDNA. Eight hypothetical alternative conformations (C1 to C8) are possible for the *L. sativa* mtDNA as the result of homologous recombination between one of the three large direct repeat pairs, and assuming that subgenomes are circular for convenience of representation. Recombination between direct repeats causes genomic fission or fusion events (brown arrows). The repeats involved in a particular recombination event are labeled above each arrow. The six single-copy genomic regions are shown in unique colors, which were used in all conformations. Subgenomic molecules present in several conformations were named according to a different path (X, Y, and Z).

Repeated sequence	Gene	Start	Stop	Strand	Remarks
R01-A	*18S*	3 797	5 743	+	
	5S	5 885	5 998	+	
	nad5b	6 217	6 238	+	
	orf163	6 306	6 797	+	
	26S	13 376	16 981	-	
	nad5c	26 886	27 035	+	
		27 956	28 350	+	
	orf52a	28 436	28 591	-	
	nad1a	36 438	36 824	+	
	orf15	50 046	50 090	+	
	cox2	50 161	51 873	+	
		51 873	52 322	+	
	nad6	54 574	55 374	+	
	rps12	70 453	70 830	-	
	nad3	70 892	71 248	-	
	nad1c	72 429	72 487	+	
	matR	73 156	75 123	+	
	nad1c	75 892	76 150	+	
	ccmFc	78 522	79 289	+	
		80 235	80 813	+	
	cob	82 829	84 004	-	
	orf43	84 368	84 757	-	
	nad4	102 192	102 652	+	
		104 100	105 126	+	
	mttB	109 094	109 975	+	No evidence of AUG created by editing
R01-B	*18S*	116 765	118 711	+	
	5S	118 853	118 966	+	
	nad5b	119 185	119 206	+	
	orf163	119 274	119 765	+	
	26S	126 344	129 949	-	
	nad5c	139 854	140 003	+	
		140 924	141 318	+	
	orf52a	141 404	141 559	-	
	rps3	147 863	147 936	+	Editing site 162570 (96%) creates internal stop codon.
		149 647	151 249	+	
	rpl16	151 140	151 655	+	
R03-A	*atp1*	161 181	163 127	+	
	nad2b	174 008	174 168	+	
		176 387	176 959	+	
		178 430	178 617	+	

Repeated sequence	Gene	Start	Stop	Strand	Remarks
	nad7	191 219	191 361	+	
		192 269	192 337	+	
		193 798	194 263	+	
		195 322	195 565	+	
		197 364	197 626	+	
	ccmFn	197 914	199 632	+	
	atp6	200 317	201 036	-	
	cox1	201 115	201 211	-	
R02-A	*rpl10*	207 527	208 015	+	
	ccmB	208 327	208 947	-	
	nad2a	212 297	212 449	+	
		213 658	214 049	+	
	atp8	220 937	221 416	+	
	cox3	221 988	222 785	+	
	sdh4	222 713	223 127	+	
	atp9	235 137	235 397	+	
	nad5A	235 723	235 928	+	
		236 805	238 004	+	
	rpl14	238 835	238 972	-	Pseudo-gene
	coxI	244 351	245 952	-	Initiation codon created by editing
R03-B	*atp1*	257 830	259 362	+	
	nad9	279 026	279 598	-	
R02-B	*rpl10*	286 031	286 519	+	
	ccmB	286 831	287 451	-	
	rps4	289 858	290 853	+	
	ccmC	297 593	298 345	-	
	orf278	308 436	309 272	-	
	nad4L	313 908	314 180	+	
	atp4	314 362	314 937	+	
	orf47	339 605	339 748	-	
	orf52b	340 461	340 619	-	
	rps13	346 593	346 943	+	
	nad1b	347 781	347 861	-	
		348 654	348 845	-	
	rpl5	354 550	355 113	+	
	rps14	355 115	355 415	+	

Table II.5. List of transcribed protein-encoding genes and open reading frames (orf) carried by the mtDNA of *L. sativa*. The trans-spliced genes are shown in the same highlighting color.

correspond to the sites were the beginnings and ends of contigs PB02 and PB03 mapped to PBc01. We assumed that during contig building the central parts of these repeated regions were not assembled within PB02 and PB03. Consequently, we extended manually PB02 and PB03 with R01 and R02, respectively. This resulted in circular molecules PBc02 and PBc03, of 112 968 bp and 78 504 bp, respectively, which as PBc01 can be considered as circular chromosomes that probably physically exist in lettuce mitochondria.

The three circular chromosomes PBc01, PBc02and PBc03 share several large repeated regions. PBc01 and PBc02 share repeated region R01 and PBc01 and PBc02 share repeated regions R02 and R03. Plant mitochondrial genomes are conventionally represented as a single and circular chromosome, despite the lack of evidence for such a model. Similarly, by assuming HR mechanisms involving repeats R01 and R02 we could configure the mtDNA of *L. sativa* var. *capitata* L. *nidus tenerrima* into a single circle of 363 324 bp (**Figure II.4**).

The presence of three large repeated sequences (R01, R02 and R03) can lead to 8 alternative configurations of the mtDNA (**Figure II.5**). Among these 8 alternative configurations only 2 are single circular chromosomes (C1 is our model). Interestingly, the chromosome PBc02 is present in four of those 8 configurations.

1.4. mtDNA gene content

1.4.1. Genes annotation

The analysis of the mtDNA of *L. sativa* var. *capitata* L. *nidus tenerrima* reveals the exact same protein genes content as *L. sativa* var. *capitata* L. *salinas*, the 24 core genes (*atp1*, *atp4*, *atp6*, *atp8*, *atp9*, *ccmB*, *ccmC*, *ccmFc*, *ccmFn*, *cob*, *cox1*, *cox2*, *cox3*, *matR*, *mttB*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7* and *nad9*) and 7 additional ones (*rpl5*, *rpl10*, *rpl16*, *rps3*, *rps4*, *rps12*, and *rps13*) that we identified with *GeSeq* (**Table II.5** & **Figure II.6**) (Tillich et al., 2017).

Among these, there is an *rps14*-related sequence located downstream *rpl5* that is described as a pseudogene in *Solanum tuberosum* and *A. thaliana* (Brandt et al., 1993; Quiñones et al., 1996; Scotti et al., 2004). The functional copy of *rps14* is in the nucleus where it was transferred roughly 80 million years ago (Ong and Palmer, 2006).

The mapping of public lettuce RNAseq data (*SRR080725*, *SRR085107*) on our *L. sativa* mtDNA allowed us to identify 7 expressed *orf*s (*orf15*, *orf43, orf47, orf52a, orf52b, orf163* & *orf278*), to confirm

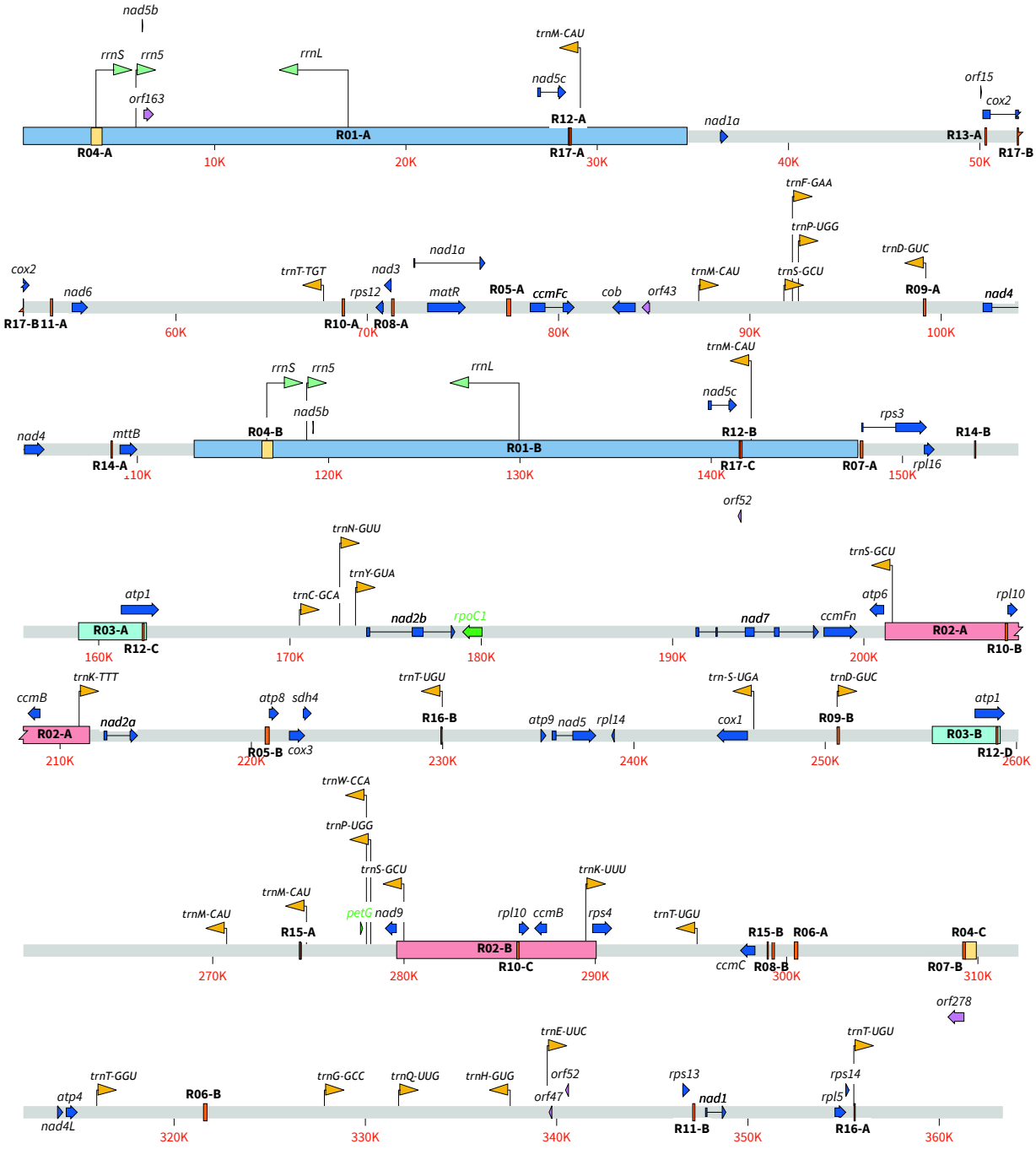


Figure II.6. Annotated mtDNA of *L. sativa*. The mtDNA is shown represented as a single linear molecule. The protein-coding genes are represented by dark blue arrows and the ORFs in light purple. The rRNA genes and the tRNA genes are shown in pale green and dark yellow, respectively. The large repeated sequences (> 500 bp) are highlighted in light blue, light red, cyan, and light yellow. The other repeated sequences larger than 50 bp are shown in orange.

tRNA Gene	Start	Stop	Strand	Sequence	R
trnC(GCA)	170496	170566	+	GGCTAGGTAACATAATGGAAATGTATCGGACTGCAAATCCTGGAATGACGGTTCGACCCCGTCCTTGGCCT	
trnD(GUC)	99119	99192	-	GGGATTGTAGTTCAATCGGTCAGAGCACCGCCCTGTCAAGGCGGAAGTTGCGGGTTCGAGCCCCGTCAGTCCCG	6
trnE(UUC)	339508	339579	+	TATCCCTTACGGGAATCGAACCCGTGTCTTCGACATGAAAAGACGATGTCCTAACCACTGGACGAAAGGGAC	
trnF(GAA)	92224	92297	+	GTTCAGGTAGCTCAGCTGGTTAGAGCAAAGGACTGAAAATCCTTGTGTCAGTGGTTCGAATCCACTTCTAAGCG	
trnG(GCC)	327857	327929	+	GCGGAAATAGCTTAATTGGTAGAGCATAGCCTTGCCAAGGCTGAGGTTGAGGGTTCAAGTCCCTCCTTCCGCT	
trnH(GUG)	337497	337570	-	GCGGATGTAGCCAAGTGGATCAAGGCAGTGGATTGTGAATCCACCACGCGCGGGTTCAATCCCCGTCGTTCGCC	
trnI(CAU)	270655	270728	-	GGGCTTATAGTTTAATTGGTTGAAACGTACCGCTCATAACGGTGATATTGTAGGTTCGAGCCCTACTAAGCCTA	
trnK(UUU)	210996	211068	+	GGGTGTATAGCTCAGTTGGTAGAGCATTGGGCTTTTAACCTAATGGTCGCAGGTTCAAGTCCTGCTATACCCA	2
trnM(CAU)-elongator	87325	87397	+	ACCTACTTGACTCAGCGGTTAGAGTATCGCTTTCATACGGCGAGAGTCATTGGTTCAAATCCAATAGTAGGTA	
trnM(CAU)-initiator	29047	29119	-	AGCGGGGTGGAGGAATGGTCAACTCATCAGGCTCATGATCTGAAGACTACAGGTTCGAATCCTGTCCCCGCCT	1
trnN(GUU)	172598	172669	+	TCCTCAGTAGCTCAGTGGTAGAGCGGTCGGCTGTTAACTGACTGGTCGTAGGTTCAAATCCTACTTGGGGAG	
trnP(UGG)	278185	278259	-	AGGGATGTAGCGCAGCTTGGTAGCGCCTTTTTTTTTGGGTAAAAAATATCACGGGTTCCAATCCTGTCATCCCTA	
trnP(UGG)	92537	92611	+	CGAGGTGTAGCGCAGTCTGGTCAGCGCATCTGTTTTGGGTACAGAGGGCCATAGGTTCGAATCCTGTCACCTTGA	
trnQ(UUG)	331736	331807	+	TGGAGTATAGCCAAGTGGTAAGGCATCGGTTTTTGGTACCGGCATGCAAAGGTTCGAATCCTTTTACTCCAG	
trnS(GCU)	91777	91864	+	GGAGGTATGGCTGAGTGGCTTAAGGCATTGGTTTGCTAAATCGACATACAAGAAGATTGTATCATGGGTTCAAATCCCATTTCCTCCG	
trnS(GCU)	201399	201490	-	GGGGGGATAGCACAAGGGCTTAAGGCATTGGTTTGCTAAATCGACATACAAGAAGATTGCATCATGCATGGGTTCGAATCCCATTTCTTCCA	2
trnS(UGA)	246179	246265	-	GGATGGATGTCTGAGCGGTTGAAAGAGTCGGTCTTGAAAACCGAAGTATTGATAGGAATACCGGGGGTTCGAATCCCTCTCCATCCG	
trnT(GGU)	315958	316031	+	ATTTCAAGCTCAGTGGTAGAGCGGTCGGCTGGTAACTGACTAGTCGGGCGTAGGTTCGAATCCTACTTGAAAAG	
trnT(UGU)	295240	295320	-	ACCTCGGTAGCTCAGCGGTAGAGCGGCCTCTTGTATAGTTCAACAGGTAGTCGGACGTGGGTTCAAATCCCGCTCGAGGGA	
trnT(UGU)	67637	67715	-	GTCACGATTAGAAGAAATGGTAGTTCGCTGGGATTGTCTTCATTTCCCAGAGGTGCTGGTTCGACTCCAGCTCGTGACA	
trnT(UGU)	355535	355613	+	GTCGCGATCAGAATGAATGGTAGTTCGCTGGGCCTGTCTTTTGCTCCCAGAGGTGCTGGTTCGAATCCAGTTCGCGACA	
trnT(UGU)	229901	229979	-	GTCGCGATCCGAATGAATGGTAGTTCGCTGGGCCTGTCTTTTGCTCCCAGAGGTGCTGGTTCGAATCCAGTTCGTGACA	
trnW(CCA)	277961	278034	-	GCGTTCTTAGTTCAGTTCGGTAGAACGTGGGTCTCCAAAACCTGATGTCGTAGGTTCAAATCCTACAGAGCGTG	
trnY(GUA)	173425	173507	+	GGGAGAGTGGCCGAGTGGTCAAAAGCGACAGACTGTAAATCTGTTGAAGTTTTTCTACGTAGGTTCGAATCCTGCCTCTCCCA	

Table II.6. List of tRNA genes carried by the mtDNA of *L. sativa*. The tRNA genes with a chloroplastic origins are highlighted in green. The repeated regions (R) bearing the tRNA gene are shown in the last column according to their color in Figure II.6.

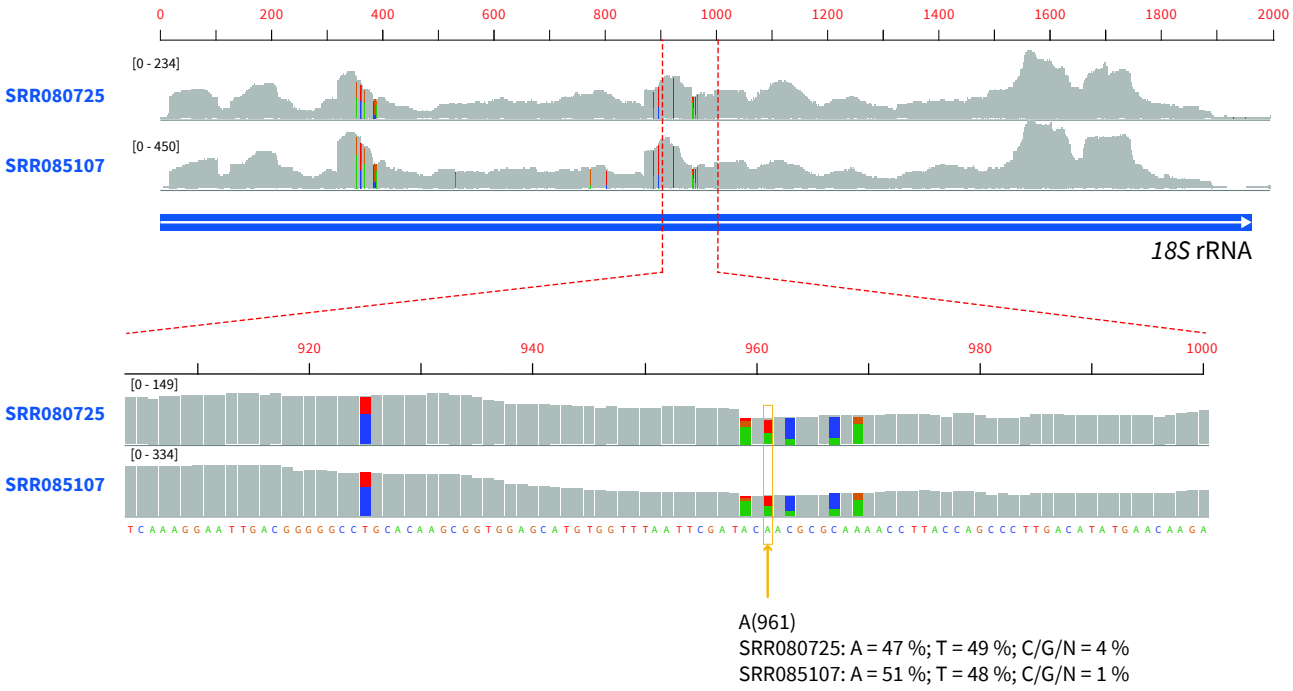


Figure II.7. Evidence for methylation of base A961 of rRNA *18S*. The coverage from RNA-sequencing of the 18S rRNA (1 947 bases) is represented by grey bars. Color bars show the proportion of the different nucleotides if a SNP is found. Among the 6 SNP focused bellow, the position A961 shows an A-to-T mismatch in both RNA-seq libraries that corresponds to the misincorporation by reverse transcriptase of a T in front of a methylated A. The other SNPs shown can be explained by the mapping of reads corresponding to the plastidial 16S rRNA. The genomic sequence is displayed at the bottom.

the structure of the 31 protein genes previously identified and to identify co-transcribed orfs. (**Table II.5**). Moreover, the data was used to identify 738 RNA-editing sites. Among these sites, the one at position 245 933 (A**C**G) creates the initiation codon (A**U**G) of *cox1,* in 89 % of the reads. The sequence of *rpl16* overlaps with the 3'-end of *rps3,* like in all higher plant mtDNAs already sequenced (Newton et al., 1996). It was therefore believed that translation of *rpl16* is initiated internal to the *rps3* sequence, but in a different frame (Takemura et al., 1992). However, it has been recently reported in *S. tuberosum* that an editing site that transforms an U**C**A serine codon into an U**U**A leucine codon in *rps3* creates an internal **U**AG stop codon in the overlapping *rpl16* reading frame, and that translation of *rpl16* should initiate downstream of *rps3* (Varré et al., 2019). We could identify the exact same event in *L. sativa* (editing site at position 151 176), which confirms the strong conservation of this editing site and its essential importance for the expression of *rpl16*.

The mtDNA of *L. sativa* carries a limited set of 29 tRNA genes, corresponding to 16 amino acid isoacceptors, among which seven have been transferred from the chloroplast (**Table II.6**). Consequently, tRNAs for alanine, arginine, isoleucine, leucine and valine need to be imported from the cytoplasm. Interestingly, genes for tRNAs decoding threonine are found in the genome, while in most plants the corresponding tRNAs are imported from the cytosol. Like other higher plant mitogenomes, that of lettuce also codes for ribosomal RNAs *5S*, *18S* and *26S*. The three rRNAs genes are in repeated sequence R01, and consequently the genome carries two copies of each. A recent publication about the mtDNA of *S. tuberosum* suggests methylation of the mitochondrial 18S rRNA (Varré et al., 2019). This hypothesis relies on RNAseq data analysis of the *18S,* showing that in 55 % of reads the base 960 of the *18S* rRNA reads as a T, while an A is found in the genomic sequence. Such an event is diagnostic for a mis-incorporation during cDNA synthesis in front of an m^1A nucleotide (Cognat et al., 2017). We observed the same event in *L. sativa*, where the A at position 961 of the 18S genomic sequence (or 4 757 of the mtDNA) corresponds to a T in 48.5 % of the reads from RNAseq data (**Figure II.7**). The mapping of reads on the mitochondrial *18S* reveals a certain number of other SNPs, but with a frequency lower than 40 %. We attributed the reads carrying these SNP by alignment to the chloroplastic *16S* rRNA gene. Because the methylation of the 18S rRNA affects a nucleotide in close vicinity of the tRNA anticodon at position P of the ribosome, the authors speculated that this methylation might be important for proper tRNA positioning and efficient translation. Our results show that methylation of the 18S rRNA at this position is conserved in higher plants, and are in line with the hypothesis of the important role of this post-transcription

# Repeat	Lenght	Copy number	Annotation	Start	Stop	Comment
R-01	34696	2	R01-A	1	34696	Direct repeats
			R01-B	112969	147664	
R-02	10430	2	R02-A	201113	211542	Direct repeats
			R02-B	279617	290046	
R-03	3552	2	R03-A	158948	162499	Direct repeats
			R03-B	255597	259148	
R-04	576	3	R04-A	3541	4116	A & B Direct repeats C inverted
			R04-B	116509	117084	
			R04-C	309342	309917	
R-05	215	2	R05-A	77284	77498	Direct repeats
			R05-B	220720	220934	
R-06	186	2	R06-A	300404	300589	Direct repeats
			R06-B	321538	321723	
R-07	137	2	R07-A	147794	147930	Direct repeats
			R07-B	309205	309341	
R-08	131	2	R08-A	71271	71401	Direct repeats
			R08-B	299235	299365	
R-09	123	2	R09-A	99073	99195	Direct repeats
			R09-B	250618	250740	
R-10	119	3	R10-A	68690	68808	B & C Direct repeats A inverted
			R10-B	207392	207510	
			R10-C	285896	286014	
R-11	117	2	R11-A	53439	53555	Inverted repeats
			R11-B	347111	347227	
R-12	97	4	R12-A	28549	28645	Direct repeats
			R12-B	141517	141613	
			R12-C	162282	162378	
			R12-D	258931	259027	
R-13	88	2	R13-A	50267	50354	Direct repeats
			R13-B	274546	274633	
R-14	68	2	R14-A	108627	108694	Direct repeats
			R14-B	153764	153831	
R-15	65	2	R15-A	274530	274594	Direct repeats
			R15-B	298983	299047	
R-16	64	2	R16-A	355545	355608	Direct repeats
			R16-B	229906	229969	
R-17	62	3	R17-A	28491	28552	A & C Direct repeats B inverted
			R17-B	51960	52021	
			R17-C	141459	141520	

Table II.7. List of the repeated sequences in the mtDNA of *L. sativa*. The largest repeated sequence are shown according to their annotation in Figure II.6.

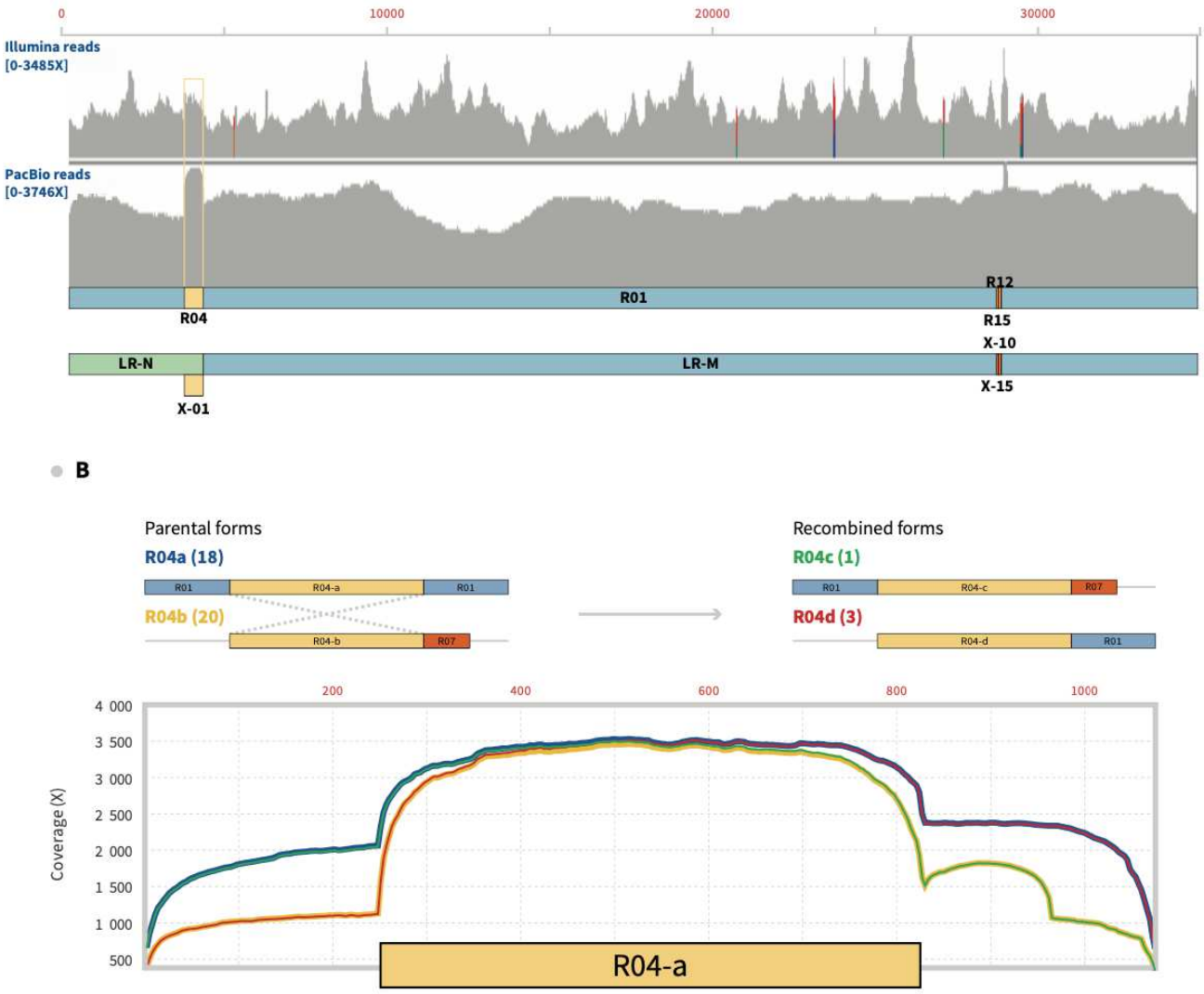


Figure II.8. Coverage of repeated sequences R01 and R04. (**A**) The coverage of the repeated sequence R01 from Illumina paired-end and PacBio SMRT reads are represented by grey barres. The structures of repeat R01 from butterhead lettuce and LR-M/LR-N from crisphead lettuce are shown at the bottom. The yellow box highlights the repeat R04/X-01, which is present in three copies in the genome. (**B**) The coverages from PacBio data of the parental forms are represented in blue (R04a) and yellow (R04b) and the coverages of the recombined forms from crossover are in green (R04c) and red (R04d). The number in brackets refers to the number of PacBio reads that spanned the different environments of the repeat.

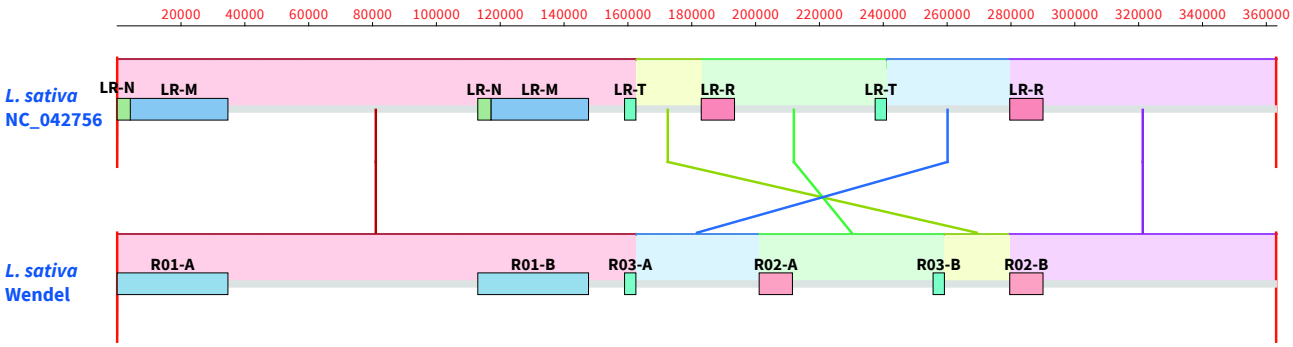


Figure II.9. *L. sativa* mitochondrial genome rearrangements. Alignment made with progressive MAUVE between *L. sativa* mtDNA NC_042756 (from Kozik et al., 2019) and the mtDNA of *L. sativa* Wendel. Regions of significant synteny are shown as colored blocks in the mauve alignment. Colored lines connect syntenic blocks of sequence between the strains. Sequences are represented with the large repeated sequences-structure.

modification for efficient mitochondrial translation. It remains unclear why just about 50 % of the sequences are methylated, both in lettuce as in potato mitochondria.

1.4.2. Identification of genome repeated sequences

The mtDNA of higher plants are usually rich in repeated sequences that are responsible for the rapid evolution of the genome organization by recombination (Gualberto and Newton, 2017). In the mtDNA of *L. sativa* var. *capitata* L. *nidus tenerrima* we identified 60 repeated sequences with *Rousfinder* (Wynn and Christensen, 2019). Among those 17 are longer than 50 bp (**Table II.7**). The annotation of the repeats in *L. sativa* var. *capitata* L. *salinas* mtDNA differs by how repeat R01 is represented (Kozik et al., 2019). Indeed, repeat R01 is present in our genome (var. *capitata* L. *nidus tenerrima*) as a single repeated sequence, while it is annotated as two distinct repeated regions LR-M (large repeat M, 30 580 bp) and LR-N (4 116 bp) in var. *capitata* L. *salinas*, based on coverage. Our coverage data do not allow us to follow that path and split repeat R01 in our assembly (**Figure II.8.A**). In the published lettuce mtDNA, at the junction between LR-M and LR-N there is a third repeated region of 576 bp named X01-a (R04 in our genome) (**Figure II.8.B**). We could detect a limited number of reads spanning the parental forms of this repeat. We detect fewer reads spanning the recombined crossover forms, meaning that we constructed the major forms of the genome, but the minor forms do exist at a substoichiometric level. Moreover, this translocation does not affect the assembled sequence of R01, because the two occurrences of repeats LR-M and LR-N are consecutive.

2. Butterhead mtDNA versus the Crisphead iceberg mtDNA

We compared our sequence of the mtDNA of Butterhead (*L. sativa* var. *capitata* L. *nidus tenerrima*), Wendel, to the recently published sequence of the mtDNA of crisphead iceberg (*L. sativa* var. *capitata* L. *salinas*; accession number of the variety NC_042756, accession of the mtDNA **MK642355**) (Kozik et al., 2019). To do so we aligned the sequences with *progressiveMauve* (Darling et al., 2010) (**Figure II.9**). The assembled genomes show different configurations involving repeated sequences R02 and R03 (or LR-R and LR-T, respectively). Such a difference is expected with respect to the dynamic structure of the mitochondrial genome. These two alternative configurations are predicted by our models based on hypothetical recombination events involving these large repeated sequences (**Figure II.5**). Moreover, Kozik et al., 2019 described two different isoforms to

their reference (NC_042756), and out of them the configuration that we built is considered. Despite this conformational divergence, the sequences are rigorously identical. The alignment of the synteny blocks did not allow us to identify any single nucleotide polymorphism (SNP) between the two Lettuce accessions.

DISCUSSION AND PERSPECTIVES

In the work described in this chapter our goal was to build the mtDNA of *L. sativa* var. *capitata* L. *nidus tenerrima* (butterhead). To do so, we successfully optimized the mitochondria purification and mtDNA extraction protocols to lettuce. Then, we sequenced the mtDNA through two different strategies, short and long-reads. The combination of both data sets was used to build the butterhead mtDNA genome, of a size of 363 324 bp. Finally, we identified our genome as an isoform of the recently published genome of *L. sativa* var. *capitata* L. *salinas,* which is confirmed by the total absence of SNPs and indels.

To build our genome we merged three circular chromosomes based on HR-like mechanisms involving repeated sequence R01 and R02. We chose these repeats which were the largest. However, if such a choice had been made with R01 and R03 we would have ended with the same conformation as the mtDNA of *L. sativa* var. *capitata* L. *salinas* (Kozik et al., 2019). There is here no bias because the dynamics of the mtDNA allows such alternative configurations, but just from our sequence data it is impossible to infer which are the conformations that truly exist in *L. sativa* mitochondria, or what are their proportions.

We determined eight possible alternative configurations based on HR-like mechanisms involving the three large repeated sequences R01, R02, and R03. But we only have data supporting three of those eight configurations: C1 is the model that we built, C3 regroups our three circular chromosomes, and C7 is the model published for iceberg lettuce (accession MK642355, Kozik et al., 2019) (**Figure II.5**). An unresolved issue is the length of the R01 repeated region (34 696), which can be spanned by very long reads only. PacBio data should provide very large reads; up to 30 kb or bigger. However, there were two problems with our PacBio run that did not allow us to obtain such data. The first concerned the high rate of non-plant contaminants found in our data. Contigs built with these sequences led to bacterial genomes. These contaminants could come from the buffer used (which was not autoclaved) or from the seedlings (non-sterile *in vitro* growth). The second

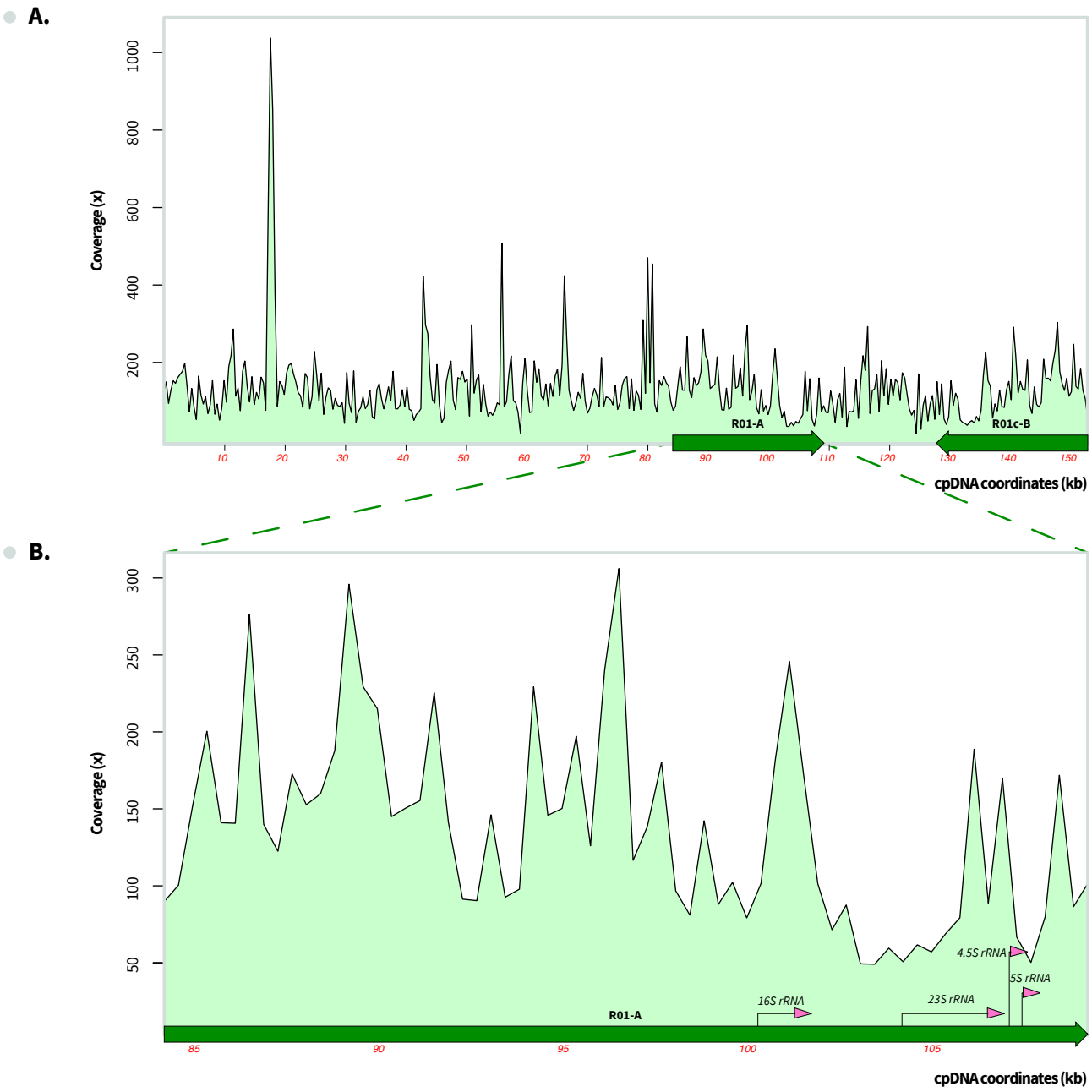


Figure II.10. Coverage of the *L. sativa* cpDNA by the Illumina data. (**A**) The Illumina paired-end reads were aligned on the cpDNA to determine its coverage. (**B**) The region on the repeated sequence R01 from the coordinates 102 000 to 107 000 shows a lower coverage than the rest of the genome, despite being duplicated.

problem concerned the size of mitochondrial reads, which were relatively small. Indeed, we only got 325 mitochondrial reads bigger than 10 kb (and 10 bigger than 20 kb), while bacterial reads went up to 100 kb. So theoretically it could have been possible to get mitochondrial reads that long. So, the question concerning the origin of bacterial contamination needs to be raised. For further studies on the mtDNA of *L. sativa* we highly recommend the use of sterile buffers and the extraction of mitochondria from seedlings grown in sterile conditions. It would be also important to avoid possible harmful steps that might break the mtDNA (extended grinding, for instance). We can however not exclude that the small size of mitochondrial reads directly reflects a fragmented physical structure of the *L. sativa* mtDNA. Indeed, data obtained by pulse-field gel electrophoresis (PFGE) and by analysis of DNA molecules using DNA Movies led the group of Arnold Bendich to propose that the organellar DNA in plants becomes highly damaged/degraded when cells develop into green leaf cells (Oldenburg and Bendich, 2015). This could explain the inability to amplify and sequence large mtDNA reads. But this hypothesis was contested by other research groups, regarding the fate of the cpDNA during leaf development (Golczyk et al., 2014). Among the 17 repeated sequences larger than 50 bp that we identified the structure of repeated sequence R01 remains unclear. In the assembly of the iceberg mtDNA (Kozik et al., 2019) the authors justified the separation of this repeated region in two (LR-M and LR-N) based on their coverage data that showed a break in that region. However, such a break in the coverage of a large repeated sequence has also been observed in *A. thaliana* cpDNA. Like most plastid genomes of higher plants the cpDNA of *A. thaliana* contains a pair of large repeated sequences (26 205 bp). The coverage across the cpDNA drops dramatically around position 100 kb, before increasing again around position 110 kb (Zampini et al., 2015). These positions correspond to the location of the rRNA operon, containing genes of rRNA *16S*, *23S* and *4.5*S (from 101 012 bp to 107 701 bp). We tested if the same happens for lettuce cpDNA sequence data. For that we mapped the Illumina reads corresponding to the cpDNA and plotted the coverage. We found that indeed the mapping of the Illumina reads on the cpDNA of *L. sativa* shows the same break of coverage at the same position, corresponding to the rRNA operon (**Figure II.10**). In Lettuce mtDNA the rRNA genes (for 18S, 26S and 5S rRNAs) are all three carried by the R01 repeated sequence. Interestingly, repeat R04/X-01 which is at the boundary between repeated sequences LR-M and LR-N is part of the *18S* rRNA gene. So, it is possible that such a drop in coverage observed in *L. sativa* mtDNA reads leads an assembler program to identify LR-M and LR-N as individual repeated regions, even if they are contiguous sequences. Furthermore, the PacBio reads spanning repeated sequences R04 allowed us to detect the four possible configurations that

result from reciprocal HR involving this pair of repeated sequences (**Figure II.8.A**). Such data confirmed that repeated sequence R01 can be split by the recombination involving R04. However, we only got very few reads that spanned forms corresponding to crossover products R04c (1 read) and R04d (3 reads), as compared to parental forms R04a (18 reads) and R04b (20 reads). Consequently, even if the small number of reads does not allow us to precisely determine the relative levels of each reciprocal recombination product, the numbers suggest that the split of R01 by recombination involving R04 corresponds to a minor configuration of the lettuce mtDNA.

To summarize this chapter, we built the mtDNA of a butterhead lettuce, which has a size of 363 324 bp. This size is consistent with the published mtDNA of a Crisphead iceberg lettuce which corresponds to an alternative configuration of the same genome. However, to study mtDNA evolution and diversity within lettuce the sequence of the mtDNA of additional accessions and species is necessary.

CHAPTER 3: Mitochondrial genome natural diversity within the genus Lactuca

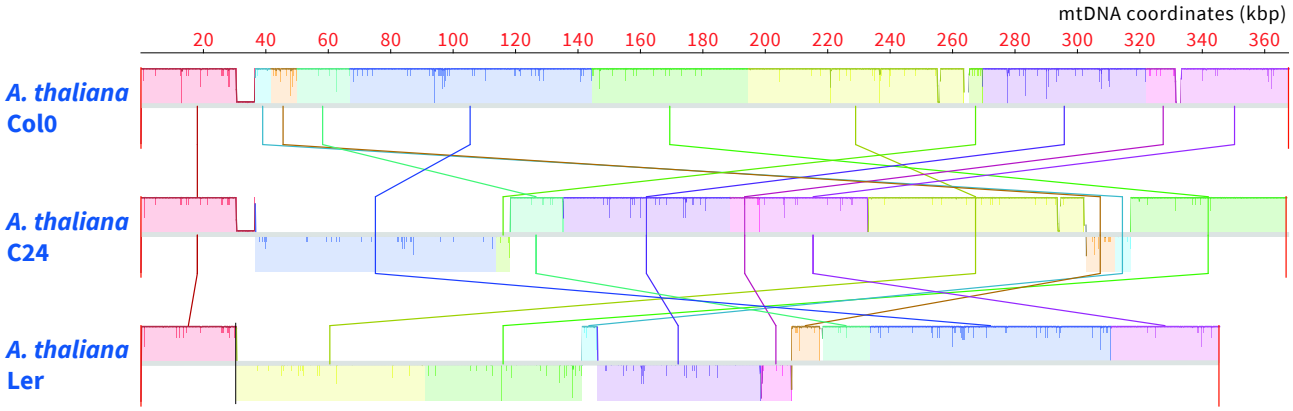


Figure III.1. Mitochondrial genome rearrangements in *A. thaliana* between ecotypes Columbia-0, C24, and Landsberg erecta. Alignment made with progressive MAUVE. Regions of significant synteny between the accessions are shown as colored blocks in the mauve alignment. *A. thaliana* accession Columbia-0: BK010421.1, C24: JF729200.1, Landsberg erecta: JF729202.1.

INTRODUCTION

1. Plant mitochondrial genome natural diversity

The content of plant mitogenomes varies little interspecies, with a set of 24 core protein genes coding for subunits of the mtETC and ±17 variable protein genes that mostly code for subunits of the mitoribosome. And mtDNA differences are further reduced when comparing evolutionary close species. Indeed, there are few differences in gene content within a genus, as in *Silene* spp. where few differences in genes content can be observed, despite massive changes in genome size (Sloan et al., 2012). And there are even fewer intraspecific differences, like in *A. thaliana*, where different accessions have the same content in protein and tRNA genes (Davila et al., 2011; Sloan et al., 2018). In fact, the diversity of plant mitogenomes is mainly found in their structure, such as in *A. thaliana* where the different sequences of the mtDNA are differently organized in different accessions, as in Col-0, C24 and Landsberg erecta (**Figure III.1**) (Davila et al., 2011).

2. Natural diversity within the *Lactuca* genus

The *Lactuca* genus contains about 100 species and among them *Lactuca sativa*, *Lactuca saligna*, *Lactuca serriola* and *Lactuca virosa*. Due to the economic importance of *L. sativa*, the phylogenetic relationship within the *Lactuca* genus has been studied. To do so, two sets of markers were used: a set of ribosomal internal transcribed spacer (ITS), ITS-1 and ITS-2, the *5.8* rRNA gene sequence as nuclear marker and the sequences of the *ndhF* and *trnL-F* genes as chloroplast markers (Koopman et al., 1998; Wei et al., 2017). These studies showed that the *Lactuca* spp. can be subdivided into several clades among which the *crop group* containing *L. sativa*, *L. saligna*, *L. serriola* and *L. virosa* (Wei et al., 2017). This subdivision goes further and the species can be distinguished, at the very exception of *L. sativa* and *L. serriola* that cannot be separated from each other (Koopman et al., 2001). Indeed, *L. serriola* is certainly one of the direct ancestors of *L. sativa* and consequently its closest wild relative (de Vries, 1990; de Vries and van Raamsdonk, 1994). However, the relative position of *L. virosa* and *L. saligna* in the cluster *L. serriola-L. sativa* remains unclear (Koopman et al., 2001). Moreover, the species and accessions tend to clusterize according to their geographical origins (Wei et al., 2017).

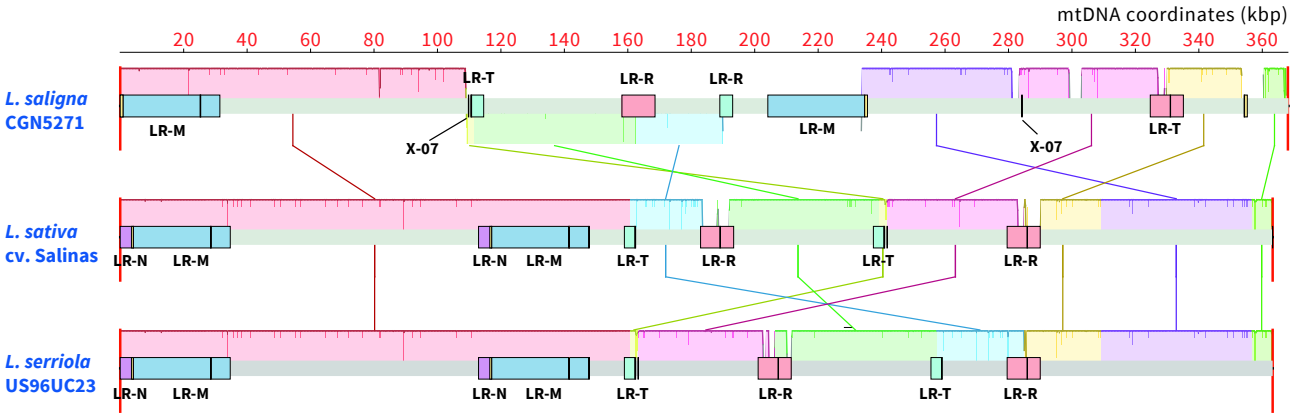


Figure III.2. Rearrangements between *Lactuca spp.* mitochondrial genomes. Alignment made with progressive MAUVE. Regions of significant synteny between the mtDNA from different species are shown as colored blocks in the mauve alignment. Accession numbers in GenBank: *L. saligna*, MK759657; *L. sativa*, MK642355; *L. serriola*, MK820672.

3. The unexpected scoop II: The mtDNA of *L. saligna* and of *L. serriola* (Kozik et al., 2019)

The release of the mtDNA sequence of *L. sativa* was accompanied by the ones of *L. saligna* and *L. serriola* (Kozik et al., 2019). Both mtDNA were assembled into a circular chromosome of 368 269 bp for *L. saligna* (GenBank accession number **MK759657**) and 363 328 bp for *L. serriola* (GenBank accession number **MK820672**). The genome of this accession of *L. serriola* (US96UC23) is only 4 bases longer than the mtDNA of *L. sativa,* due to a CTCG insertion, and there are only 8 SNPs between *L. sativa* and *L. serriola* **MK820672.** All three genomes share the same gene content (31 protein-coding genes, 3 rRNA and the same 25 tRNA genes described in *L. sativa* **MK642355**). However, *L. saligna* carries two supplementary sequences corresponding to *rpo* genes coding for a phage-type RNA polymerase and *dpo* coding for a type B2 superfamily DNA polymerase enclosed between inverted repeats of 1 218 bp (Kozik et al., 2019). These sequences resemble mitochondrial linear plasmids found and described in other plants which can be physically integrated into the mitochondrial genome (Warren et al., 2016). The three mtDNA sequences show cpDNA insertions, like most higher plant mitogenomes, but *L. sativa* and *L. serriola* display the same insertion between *nad2b* and *nad7. L. saligna* has also a chloroplastic insertion at this position, but a different one. The difference in structure between *L. sativa* **MK642355** and *L. serriola* **MK820672** can be explained by recombination events involving the large repeated sequences LR-R and LR-T (**Figure III.2**). Indeed, the configuration of *L. serriola* **MK820672** mtDNA corresponds to configuration C1 in **Figure II.5** and so to the mtDNA of Wendel. The differences between these mitogenomes and the one of *L. saligna* **MK759657** highlights a recombination event involving an intermediate-size repeated sequence, X-07 of 122 bp (Kozik et al., 2019). The sequence of this repeat can be also found in the mitogenomes of *L. sativa* and of *L. serriola,* but only as single copy sequences, with 2 and 3 SNP as compared to X-07 of *L. saligna,* respectively.

4. Aim of this chapter

In this chapter I describe the sequencing and assembly of 9 supplementary wild accessions of *L. saligna*, *L. sativa*, *L. serriola* and *L. virosa* that we performed during my PhD work. These sequences were then compared interspecies but also intraspecies to get a better insight into the wild and natural diversity of the mtDNA within the *Lactuca* genus.

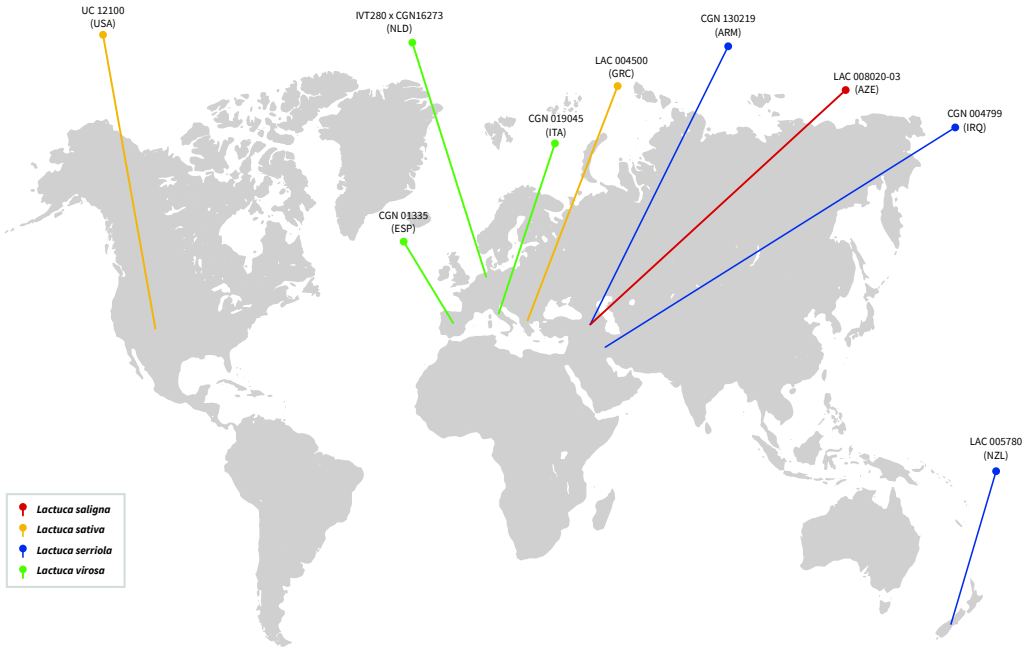


Figure III.3. Geographical origins of selected wild lettuce accessions. The 9 accessions selected for the work described in this chapter have been collected in wild environments. *L. saligna* (in red), *L. sativa* (in yellow), *L. virosa* (in green) and *L. serriola* (in blue). NDL: The Netherlands; ESP: Spain; ITA: Italy; GRC: Greece; ARM: Armenia; AZE: Azerbaijan; IRQ: Iraq; NZL: New-Zealand.

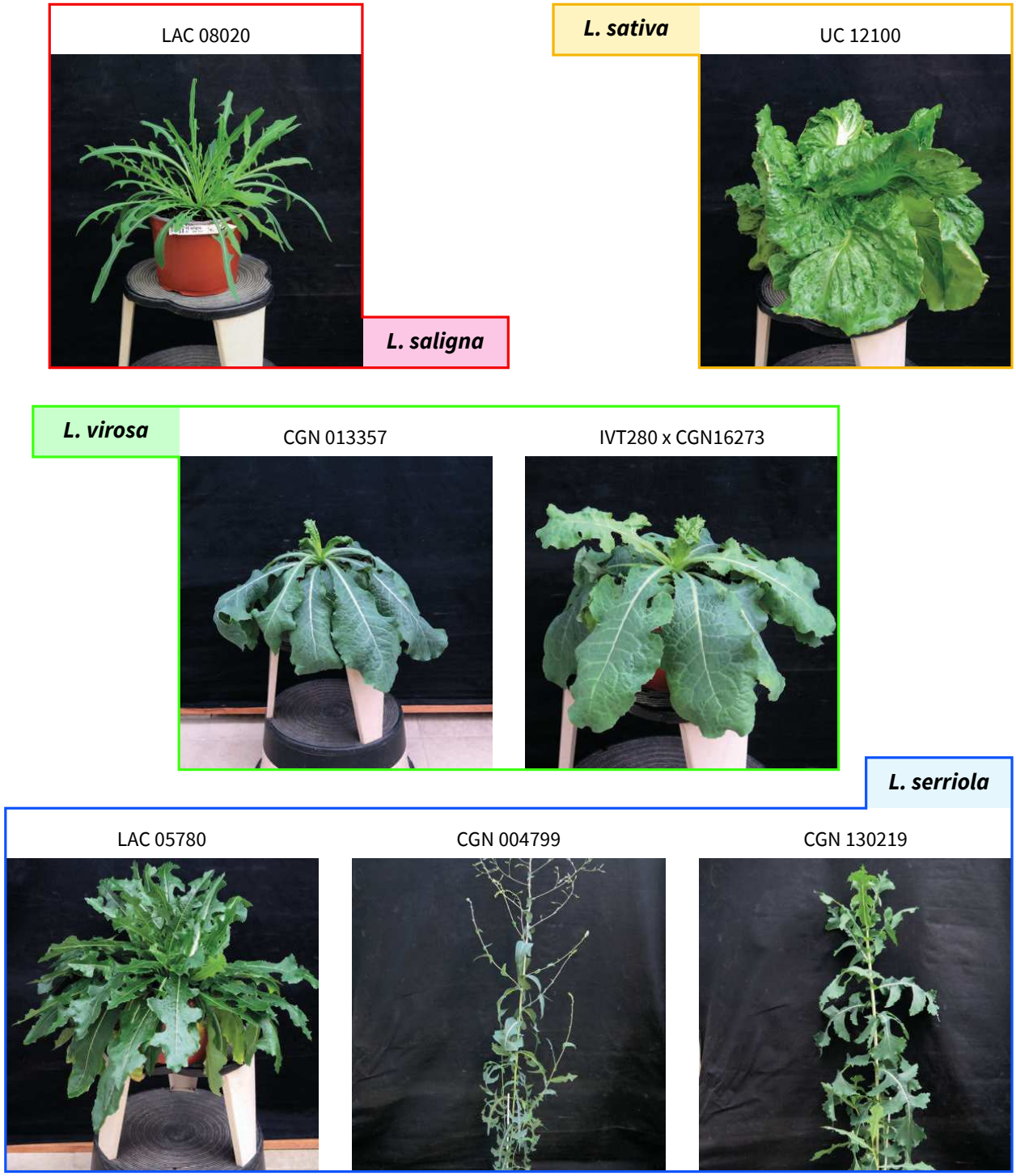


Figure III.4. Phenotypes of selected *Lactuca* spp. *L. saligna* (in red), *L. sativa* (in yellow), *L. virosa* (in green) and *L. serriola* (in blue) have distinguishable leaf phenotypes.

Species	Accession #	mtDNA extraction			Qubit 4 Fluorometer	DNA extracted (ng)
		Starting material (g of seeds)	Method	Volume (µL)	Concentration (ng/µL)	
L. saligna	LAC 008020	1	QIAamp	40	15	600
L. sativa	UC 12100	1	QIAamp	40	7	280
	LAC 004500	1	QIAamp	40	8	320
L. serriola	CGN 130219	1	QIAamp	40	26	1040
	CGN 004799	1	QIAamp	40	15	600
	LAC 005780	1	QIAamp	40	17	680
L. virosa	CGN 013357	1	QIAamp	40	10	400
	CGN 019045	1	QIAamp	40	3	120
	IVT280 x CGN 16273	1	QIAamp	40	13	520

Table III.1. Quantification of purified mtDNA from wild accessions of lettuce. *L. saligna* (in red), *L. sativa* (in yellow), *L. virosa* (in green) and *L. serriola* (in blue).

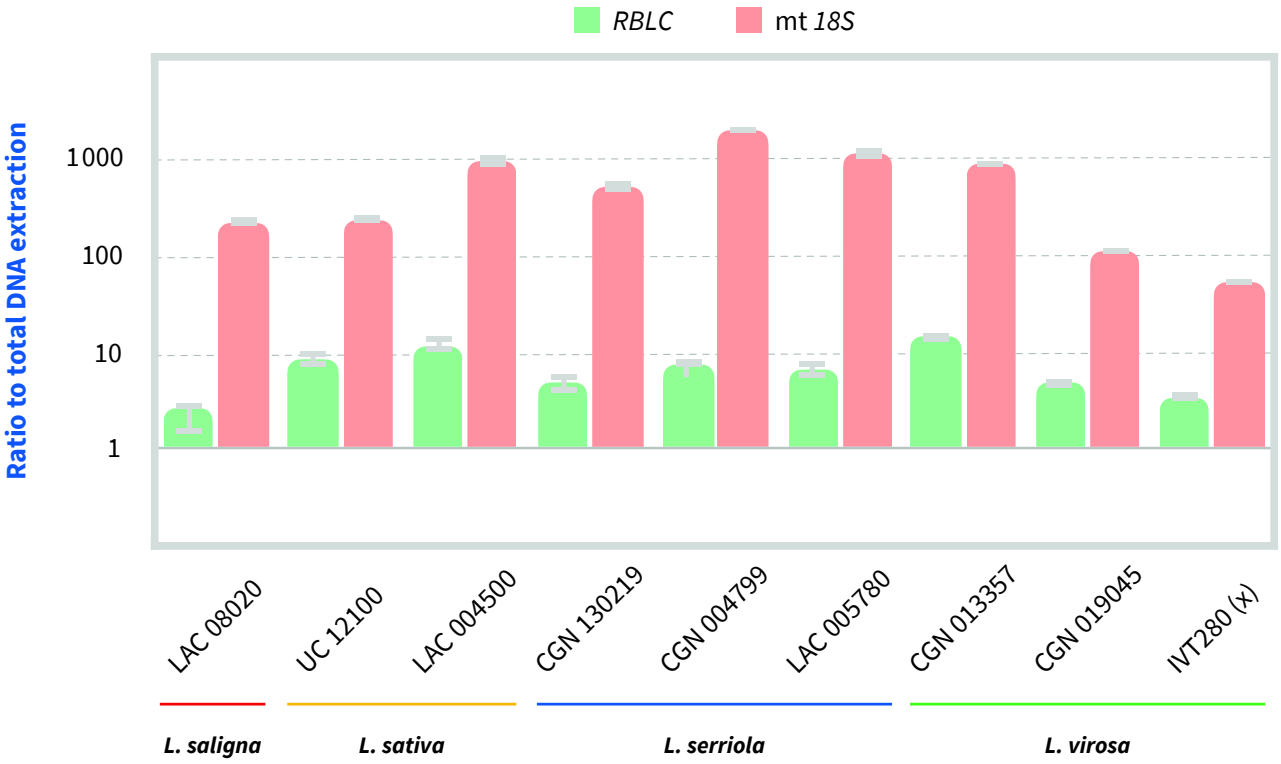


Figure III.5. Enrichment of the mtDNA extracted from *Lactuca spp.* Relative amplification of organellar genomes in DNA extracted from purified mitochondria, as compared to total plant DNA. Results are in log10 and normalized to the nuclear *18S* rRNA gene sequence. The chloroplast *RBCL* gene level is shown in green and the mitochondrial 18S rRNA gene level is shown in orange. Error bars are the SD values from three technical replicates.

1. Selection of accessions from wild lettuce

At the beginning of the project, in addition to *L. sativa* we planned to sequence and assemble the mtDNA from three different *L. saligna* accessions, three from *L. serriola* and three from *L. virosa* coming from four different continents **Figure III.3**). To do so we get access to the Genebank of Enza Zaden. The selection of the accessions was made by Mariann Börner and Anne Rood, among a shortlist of accessions that we proposed, and based on their geographical origins. The Genebank made the final selection according to three criteria: being able to ship 300 seeds of each accession, the seeds have a good germination rate and finally, the selected plants can rapidly produce seeds in case of need.

Among the selected accessions of *L. saligna* we found that accessions LAC004500 and UC12100 are phenotypically very close to *L. sativa*, and consequently were considered as *L. sativa* accessions, leaving LAC008020 as the only *L. saligna* accession that was retained (**Figure III.4**). In the end, we got seeds for the 9 accessions presented in this chapter.

2. Assembly of wild lettuces mitochondrial genomes

2.1. Mitochondrial DNA extraction and sequencing

The previously built mtDNA of *L. sativa* Wendel was a good reference to start the assembly of other *Lactuca spp*.. That is why we choose an Illumina MiSeq sequencing strategy. However, the etiolated seedlings from 300 seeds is hardly enough for one single mitochondria purification and mtDNA extraction, with no margin of error. Consequently, we started to produce seeds, and for each accession we had 3 plants in Strasbourg and 5 in Allonnes for seeds production. However, this required more time than planned, because the *L. virosa* accessions required vernalization to bolt.

Once we had a sufficient number of seeds we extracted mitochondria from etiolated seedlings, following the same protocol used for *L. sativa* Wendel. We started from a higher number of seeds than for *L. sativa,* because of the possible lower germination rate of wild lettuces seeds (**Table III.1**). We managed to get at least 100 ng of enriched mtDNA for each accession, ranging from

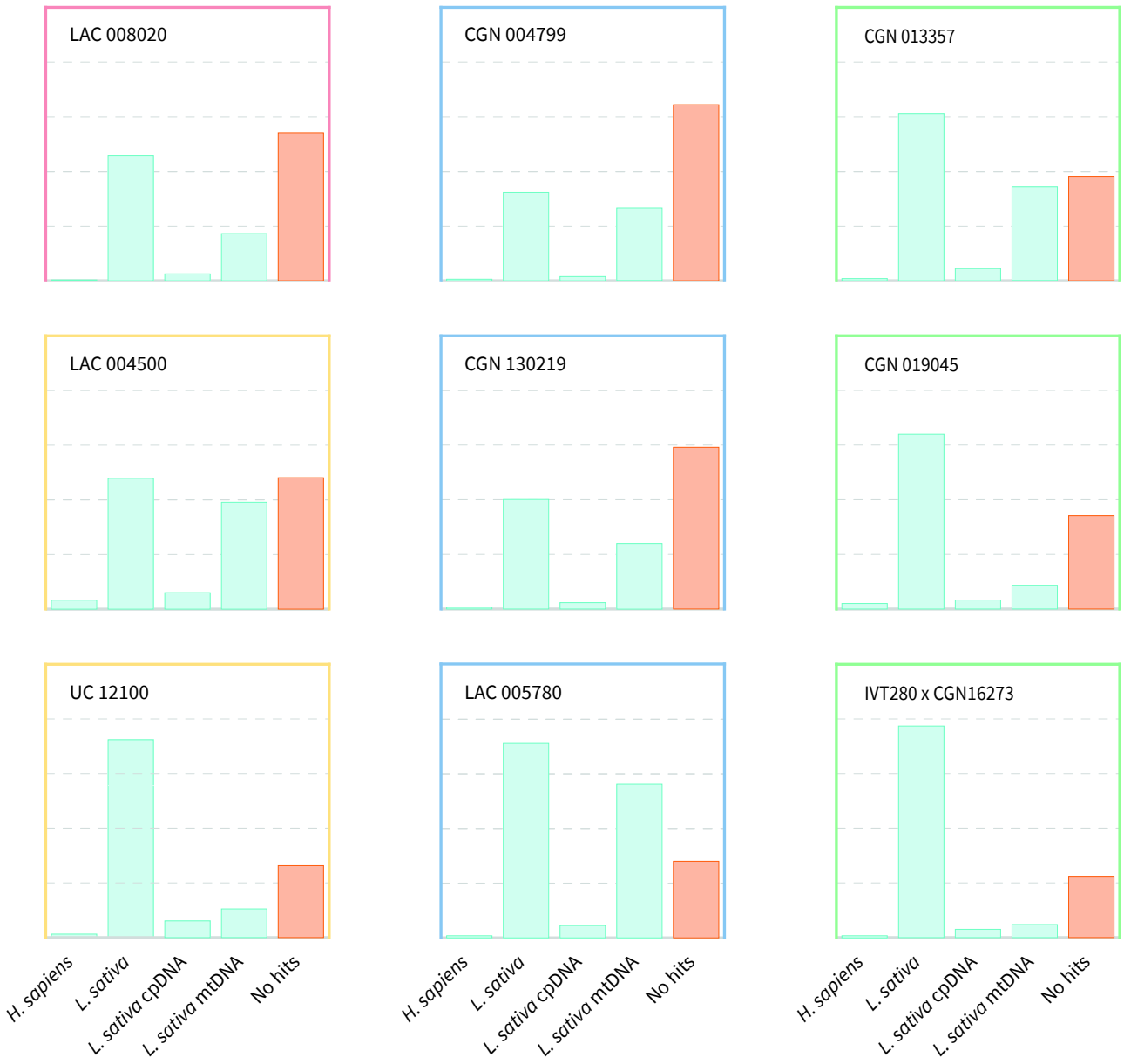


Figure III.6. Graphical output from FastQ Screen after mapping reads from DNA extracted from mito preps against several reference genomes. Mapped reads are shown in pale turquoise on the corresponding genome, and unmapped ones are shown in orange in the column “no hits”. *L. saligna* (in red), *L. sativa* (in yellow), *L. virosa* (in green) and *L. serriola* (in blue).

120 ng extracted from *L. virosa* CGN01045 to 1.04 µg extracted from *L. serriola* CGN130219. These DNA quantities are more than enough for Illumina MiSeq sequencing. Moreover, the determination of the relative levels of nuclear, chloroplastic and mitochondrial DNA by qPCR showed a sufficient mitochondrial DNA enrichment (**Figure III.5**). For all samples the mtDNA was enriched more than 100-fold (up to ~1800 fold), with the very exception of *L. virosa* IVT280xCGN16273, while the cpDNA was poorly enriched (~2.5 to ~20 fold). That was less mtDNA enrichment than what we could obtain for *L. sativa* Wendel, for which we had an mtDNA enrichment of more than 10 000-fold.

The eight samples were then sequenced through Illumina MiSeq paired-end (2 x 150 bp), which was managed by the IBMP platform *"Gene expression analysis"*. These sequencing produced ~340 000 up to ~500 000 reads per accession. However, and as shown by qPCR (**Figure III.5**), there was a high rate of nuclear contamination (from ~33 % of the reads from *L. serriola* CGN004799 to ~77 % of the reads from *L. virosa* IVT280xCGN16273), while the cpDNA reads only represented at maximum 6 % of the total reads (**Figure III.6**).

2.2. Genomic diversity within the selected accessions

To get an insight into the diversity within our population we followed two strategies. The first relied on the chloroplastic markers *ndhF* and *trnL-F* (Wei et al., 2017). For each accession (+ *L. sativa* Wendel) we assembled these two sequences and concatenated them. We aligned the concatenated sequences and draw a phylogenetic tree relying on 11 SNPs (**Figure III.7**). However, we were not able to build the *ndhF* sequence for accessions CGN004799 and CGN130219 of *L. serriola,* because of the very low coverage. Consequently, we used the *ndhF* of *L. serriola* LAC005780 in their concatenated sequences. On this tree, the accessions belonging to the same species clusterize, which confirmed our suspicion that accessions LAC004500 and UC12100 correspond to *L. sativa* accessions, since they clusterize with Wendel *L. sativa*. But these sequences are highly conserved within a species. However, besides the 11 SNPs that we found, the *L. saligna* accession LAC008020 shows a deletion of 17 bp in the spacer sequence between the second exon of *trnL* and *trnF*. But because it was our single *L. saligna* accession that was sequenced we were not able to confirm that result in other accessions of the same species.

The second strategy was based on the sequence of the two ITS, ITS-1 and ITS-2, present on both sides of the nuclear *5.8S rRNA* gene (Koopman et al., 1998). However, we could not assemble the corresponding wild lettuce sequences due to low coverage. Indeed, for some accessions we did

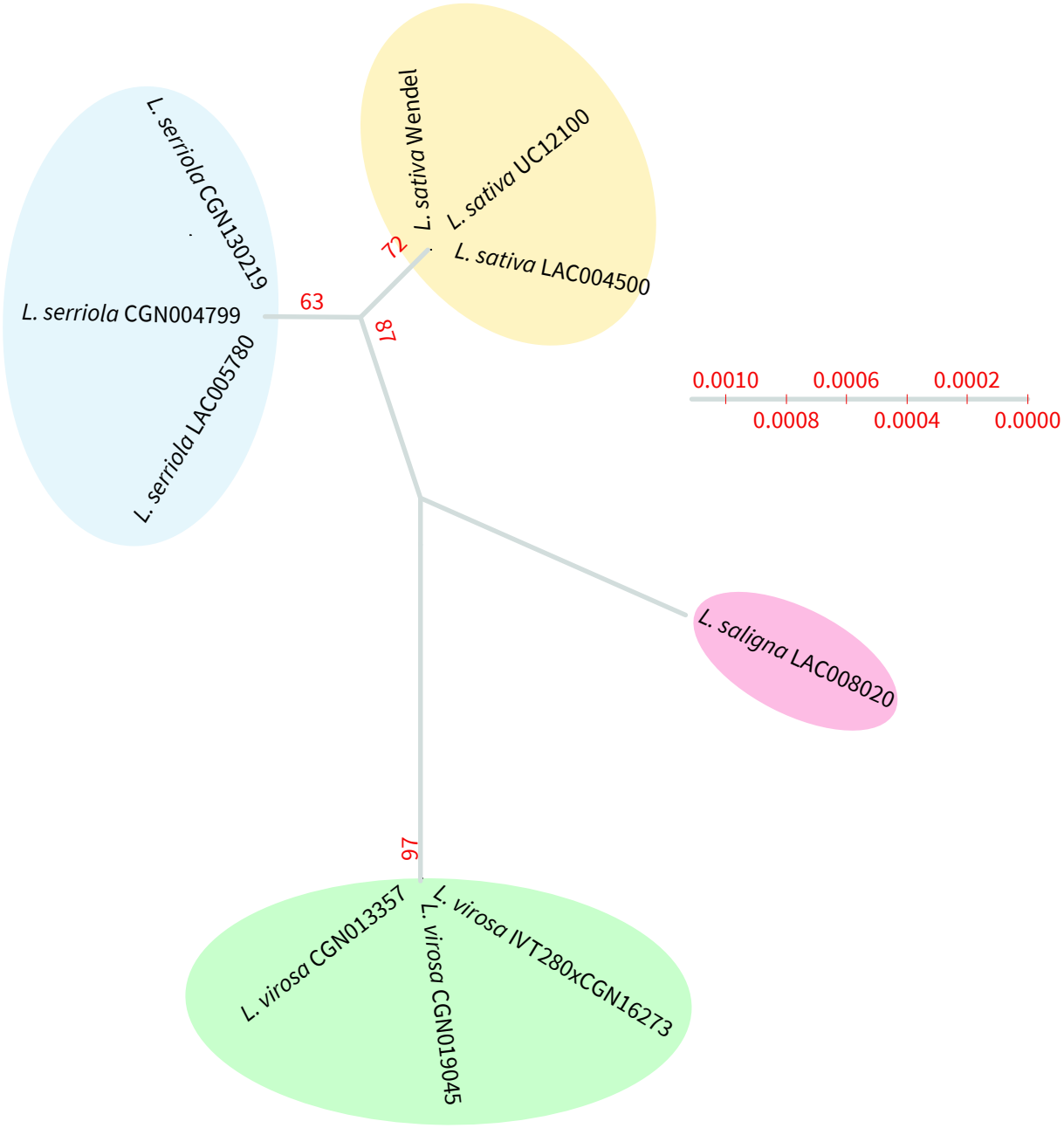


Figure III.7. Chloroplastic evolutionary relationship within *Lactuca spp*.. Neighbor-Joining tree based on the sequences of *ndhF* and *trnL-F*. Bootstrap values (1100 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. *L. saligna* (in red), *L. sativa* (in yellow), *L. virosa* (in green) and *L. serriola* (in blue).

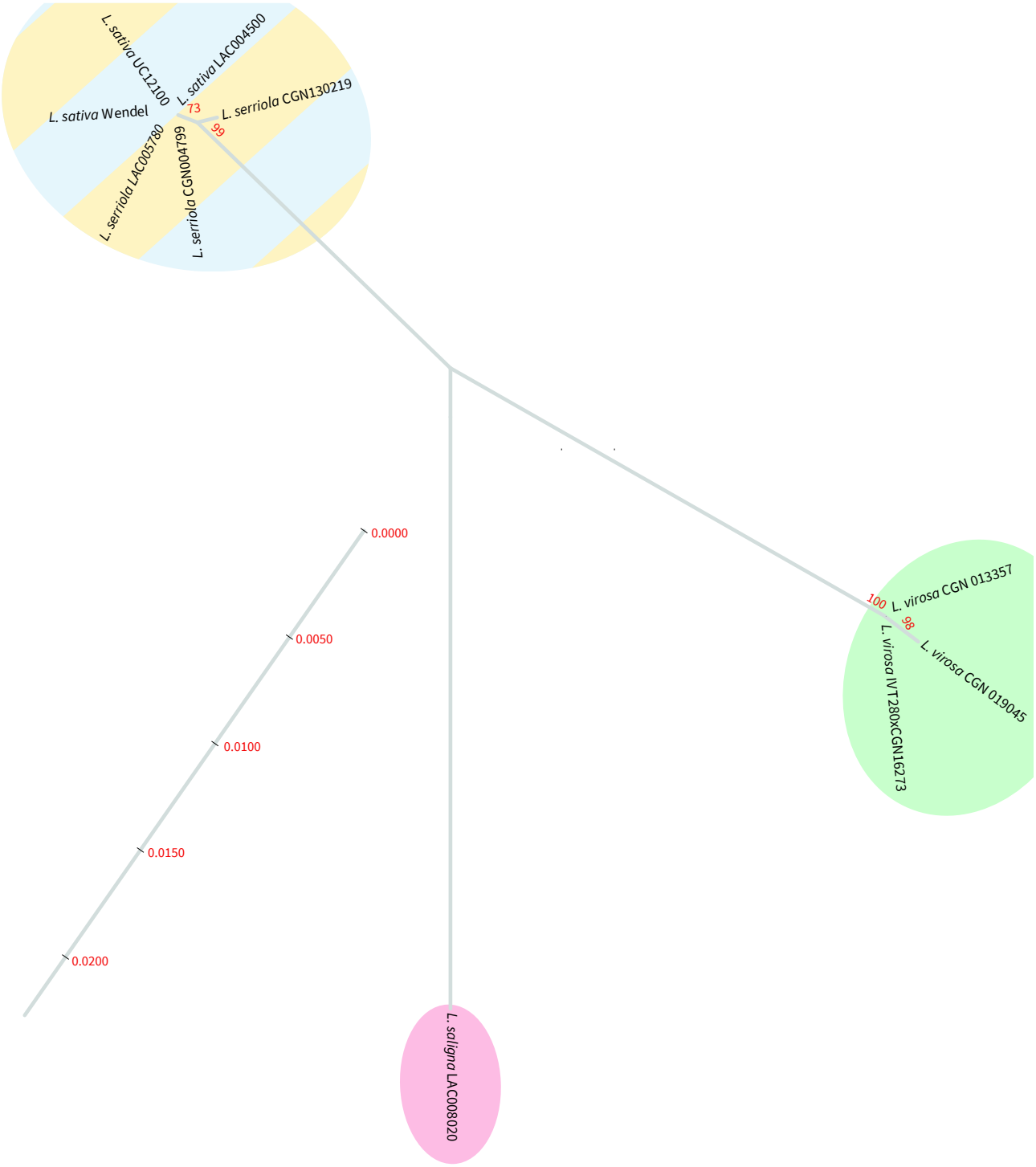


Figure III.8. Nuclear evolutionary relationship within *Lactuca spp.*. Neighbor-Joining tree based on the sequences of the ribosomal internal transcribed spacers (ITS), ITS-1 and ITS-2, and on the *5.8* rRNA gene. Bootstrap values (1100 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. *L. saligna* (in red), *L. sativa* (in yellow), *L. virosa* (in green) and *L. serriola* (in blue).

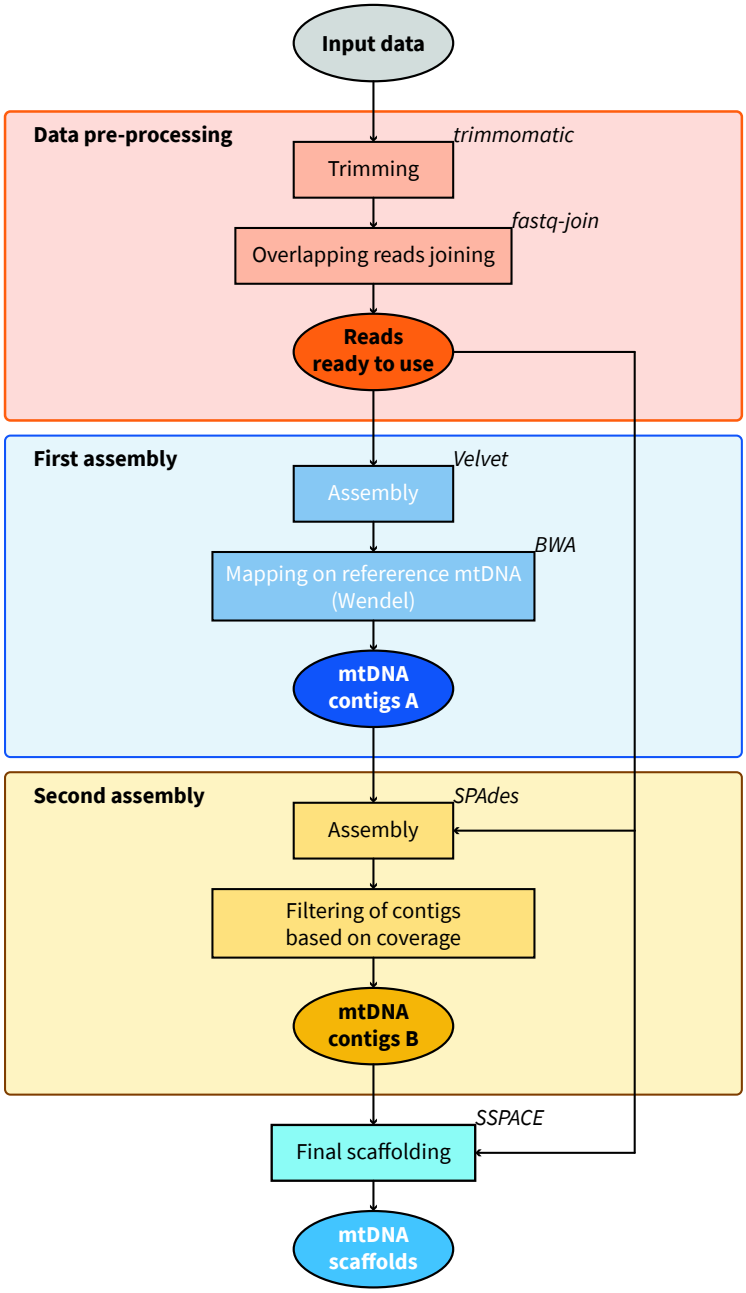


Figure III.9. The pipeline used for mtDNA scaffold assembly. The names of the different programs used are shown at the top-left of the boxes.

not attain 1X coverage of the nuclear genome. Consequently, we built the sequences based on the Wendel reference and the SNPs that could be detected on reads mapped using *IGV* (Robinson et al., 2011). The alignment of these sequences allowed us to draw a phylogenetic tree using the Neighbor- Joining method (**Figure III.8**). The nuclear markers confirmed that LAC004500 and UC12100 do not belong to *L. saligna,* as it was obtained through the chloroplastic markers. However, we were not able to identify any SNP between the *L. sativa* and *L. serriola* sequences, which consequently are not resolved in the tree of **Figure III.8.**

2.3. Mitochondrial genomes assembly

The genome assembly of the wild lettuce mtDNAs was not supported by a long reads strategy, as it was for Wendel. Consequentely we needed an assembly pipeline that could give us the longest contigs possible, starting from Illumina paired-end reads (2 x 150 bases).

To do so we built our pipeline based on a recent work which aimed to determine recombination sites in cybrids between *Nicotiana tabacum* and *Hyoscyamus niger,* by sequencing and assembly of their mtDNA starting from Illumina paired-end reads (2 x 125 bases) (Garcia et al., 2019). To do so they performed several *de novo* assemblies with different assemblers. The first assembly was done using Velvet (Zerbino and Birney, 2008) and the contigs produced were identified using BLAST against the parental mtDNA, to select the mitochondrial contigs. Then those contigs were used as trusty contigs in SPAdes, the second assembler (Bankevich et al., 2012). The contigs produced by SPAdes were then extended using SSAKE and scaffolded using SSPACE and Gapfiller (Warren et al., 2006; Boetzer et al., 2010; Boetzer and Pirovano, 2012). The scaffolds then obtained were finally assembled manually. We adapted this method to build our pipeline, with the advice and support of Bas van Hoewijk from the bioinformatics department of Enza Zaden (**Figure III.9**). There are few differences between the original and our adaptation of the pipeline. Indeed, we used the same assemblers (Velvet and SPAdes) but only a single scaffolder (SSPACE), and the contigs from velvet were mapped on the closest reference, which is the Wendel mtDNA, before being used in SPAdes.

At the end of the analysis we obtained a limited set of scaffolds for each accession and for Wendel (used to validate the pipeline), from 5 scaffolds for Wendel and *L. serriola* CGN130219 to 9 scaffolds for *L. serriola* CGN004799 (median = 6) (**Table III.2**). Interestingly we managed to get long contigs from this pipeline up to 167 kb, which could represent ~46 % of the mtDNA if referred to the

	L. sativa			*L. serriola*			*L. virosa*			*L. saligna*
	Wendel*	UC 12100	LAC 04500	CGN 130219	CGN 004799	LAC 05780	CGN 013357	CGN 019045	IVT280 x CGN 16273	LAC 08020
# reads	2 078 368	519 284	457 992	447 762	340 735	438 651	465 696	345 383	403 001	417 478
# scaffolds	5	6	7	5	9	6	6	5	8	8
# scaffolds (>= 25 000 bp)	3	3	5	5	5	3	2	3	4	3
# scaffolds (>= 50 000 bp)	2	3	2	3	2	2	1	2	1	3
Largest contig (bp)	166 885	166 885	113 223	88 886	112 883	112 883	163 898	94 333	163 770	114 696
Total lenght (bp)	316 040	404 785	312 225	317 802	313 268	311 365	298 891	246 413	334 855	321 585
GC %	45.29	43.21	45.29	45.23	45.26	45.29	45.13	41.02	43.80	45.15
Total lenght (bp)	363 324	363 324	363 324	396 235	367 647	367 647	373 019	373 027	403 737	368 269
Mean coverage (X)	736	309	922	497	415	1 361	666	119	79	379

Table III.2. Assembly statistics of wild lettuces assembly. *: Illumina data from Wendel has been used to validated the pipeline of mtDNA assembly.

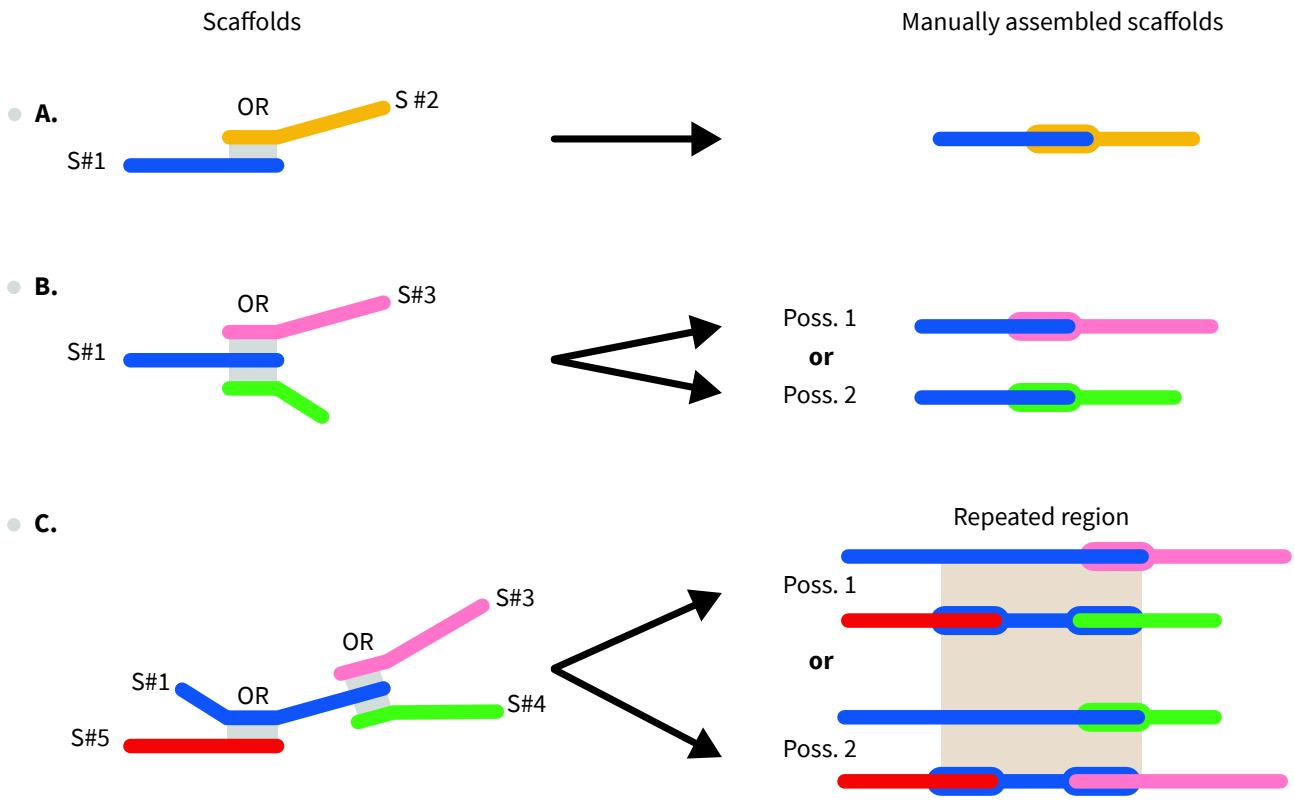


Figure III.10. Manual assembly of scaffolds. (**A**) Two overlapping scaffolds involve a single possibility of resolution. (**B**) Three overlapping scaffolds involve two possibilities of resolution but suggest the presence of a repeated sequence. (**C**) OR: Overlapping region. S: Scaffold. Poss.: Possibility.

mtDNA of Wendel. We got diversity in the total size of the sets of scaffolds, which can be explained by the following considerations. The first is that the assemblers do not take care of repeated sequences and built them only once in the set. That is why the set of scaffolds from Wendel has a total size of about 316 kb, while the Wendel mtDNA is 363 kb long (R01 = 34 kb, R02 = 10 kb, R03 = 4 kb). The second consideration is the persistence of certain scaffolds corresponding to cpDNA sequences to be rebuilt, even after mapping and filters. This appeared for *L. sativa* UC12100 which is the accession for which the mtDNA extraction gave the lower mtDNA/cpDNA ratio (**Figure III.5**). Consequently, the cpDNA contigs have coverages (80 and 113 X) that intersect with the range of coverages of the mtDNA contigs (38 to 127 X) in this accession. Moreover, it is possible that the assembler begins to build a contig with mitochondrial reads, and that after reaching a region of similarity between the two organellar genomes it continues by building chimeric contigs. However, we did not detect such events in our data set. Indeed, the cpDNA scaffold (109 519 bp) found in the set of scaffolds from *L. sativa* UC12100 corresponds to the cpDNA lacking the small single copy region and the second copy of the inverted repeat.

The next step was the manual assembly of the scaffolds to build the mtDNAs. To do so we started with the bigger scaffold for each set and mapped the other scaffolds on its extremities to assemble the overlapping scaffolds (**Figure III.10.A**). The procedure is then repeated with the extremity of the added scaffold. When two scaffolds map to the same extremity (**Figure III.10.B**) it suggests the presence of a repeated sequence. This could be confirmed by looking for third scaffolds of which an extremity mapped within the first one (**Figure III.10.C**). During the assembly of the ten different mtDNAs we used the Wendel mtDNA known sequence to identify the similar repeated sequences present in the wild lettuce mtDNAs. But in some cases we were not able to find a contig or a repeated sequence to continue the sequence. We suggested that this could be due to our pipeline, which could filter out contigs with very low coverage. To counteract that, on the advices of Axel Fischer of Dr. Stephan Greiner's laboratory we used contigs build with IDBA-UD, a metagenomic assembler (Peng et al., 2010) in addition to our set of scaffolds. Indeed, the strategy allowed us to close the circle of several genomes.

After manual assembly we got the 9 circular mtDNAs from the wild accessions and the one coming from the test on Wendel (named pip-Wendel for pipeline). First, we confirmed that the size of pip-Wendel is the same as the reference one, but the structure slightly differed. Indeed, pip-Wendel corresponds to configuration C7 of the *L. sativa* mtDNA (**Figure II.5**) and thus to the one

previously published by Kozik and collaborators (Kozik et al., 2019). This result validated that our pipeline works correctly to assemble lettuce mitochondrial genomes just from Illumina sequence data. The two other *L. sativa* mtDNAs, from accessions UC12100 and LAC04500, also have the same size of 363 324 bp as the mtDNA of Wendel (**Table III.2**). The mtDNA of LAC08020, our single *L. saligna* accession, is 368 269 bp-long, which corresponds to the size of the published one (Kozik et al., 2019). But we did get diversity in terms of mtDNA size within *L. serriola* and *L. virosa*. Indeed the two species display the same pattern: two genomes quite close in size (367 647 bp for *L. serriola* and 373 019-366 919 bp for *L. virosa*) and one genome ~30 kb bigger. This ~30 kb difference corresponds to the large repeated sequence R01 that we needed to consider repeated three times in the genome to build our circular models of the mtDNA of *L. serriola* CGN130219 and of *L. virosa* IVT280 x CGN16273. However, we know from the coverage data that such an event is very unlikely *in vivo,* and only results from the *in silico* assembly.

3. Analysis and comparison of the assembled mtDNA genomes

3.1. Gene content

We analyzed the different genomes for three different types of sequences: the genes, repeated sequences and the chloroplastic DNA insertion(s).

We determined the gene content of each accession by different methods. The first consisted in the comparison to the gene content of Wendel *L. sativa*. To do so, we used GMAP (Wu and Watanabe, 2005) to map the coding sequences (rRNA, tRNA and CDS) of Wendel on the other mtDNA genomes. We found that the nine accessions have the same gene content as Wendel. Later, we confirmed this result with *GeSeq* (Tillich et al., 2017).

In the publication of the *Lactuca spp.* mtDNAs, Kozik and collaborators reported the chloroplastic insertion of a part the second exon of *rpoC1* (1008 pb) and of a fragment of *ycf3* (115 bp) within the mtDNA of *L. sativa* and *L. serriola,* between the sequences of the mitochondrial genes *nad2b* and *nad7* (Kozik et al., 2019). The authors also reported a chloroplastic insertion between these two genes in *L. saligna,* but involving a fragment of *psaB* of 2253 bp instead of *rpoC1*. As we did for Wendel, we could observe a chloroplastic insertion between *nad2b* and *nad7* in all accessions. The two accessions of *L. sativa* and the three accessions of *L. virosa* show the same

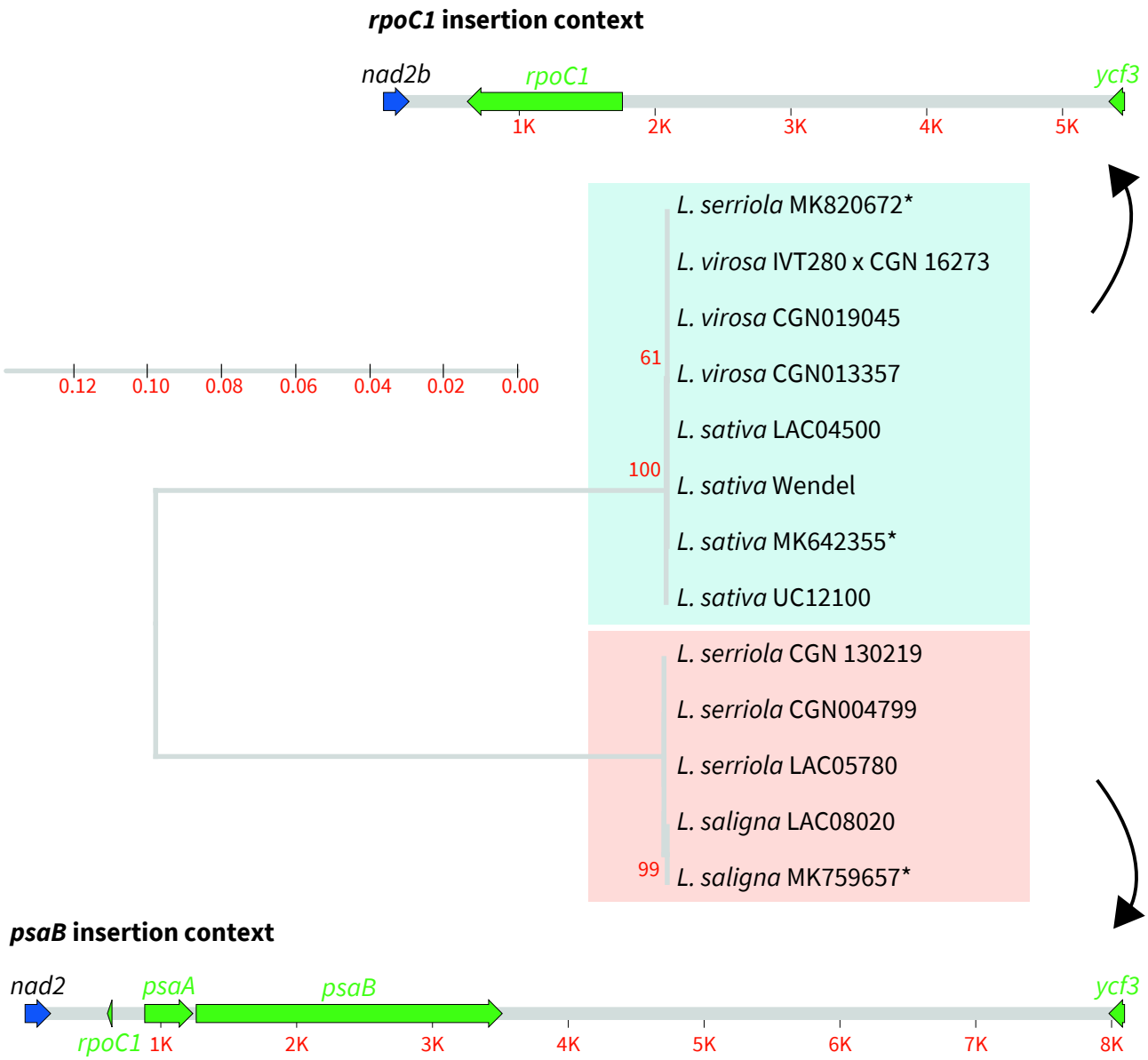


Figure III.11. Alignment of the regions containing the chloroplast DNA insertions. Neighbor-Joining tree based on the sequences of the chloroplastic insertions found in the mtDNA of lettuces. The region containing the *rpoC1* insertion is shown above, the one containing the *psaB* insertion is shown below. Bootstrap values (1100 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. *: accessions from Kozik et al., 2019.

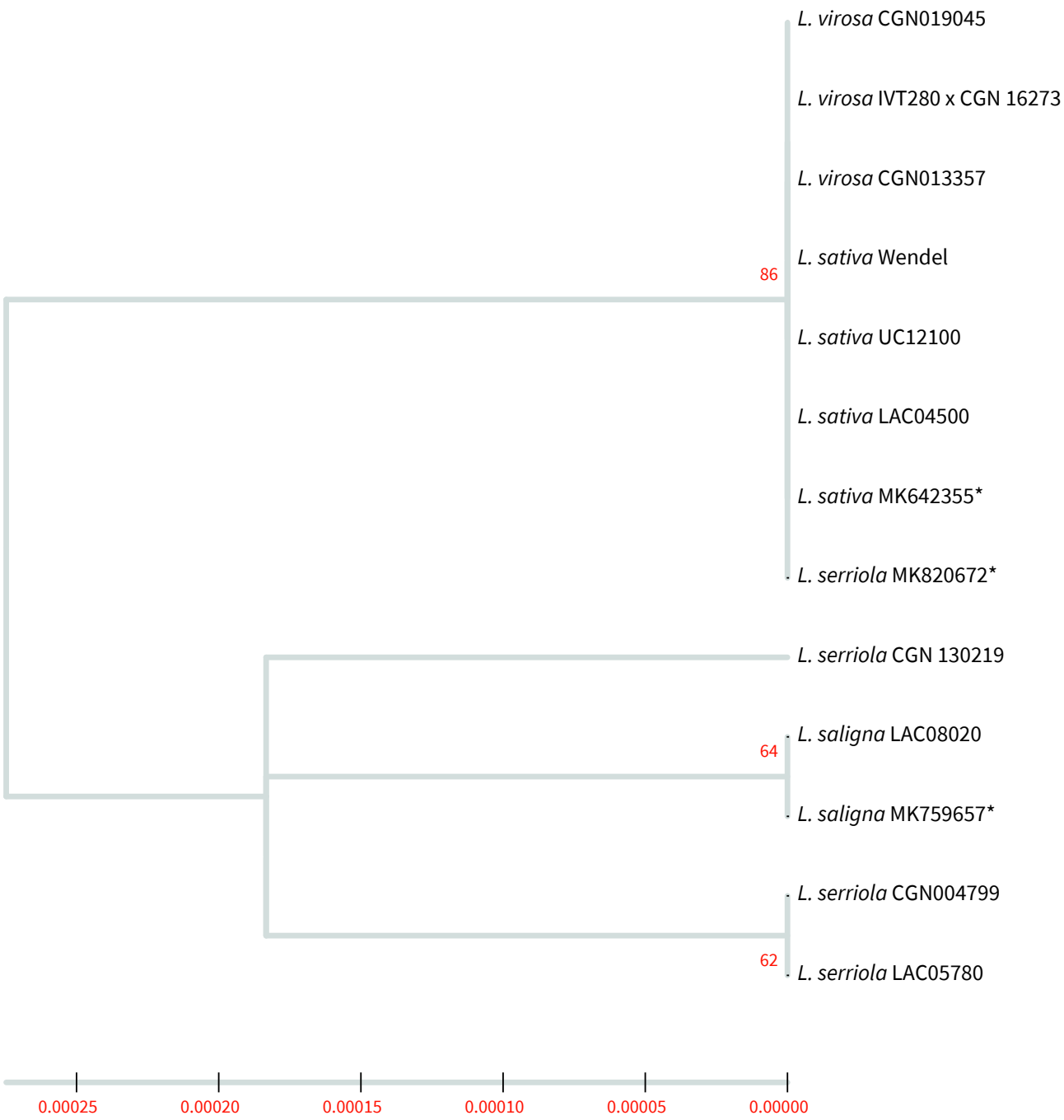


Figure III.12. Neighbor-joining phylogeny tree of mitochondrial concatenated sequences. Neighbor-Joining tree based on the catenated sequences of the *ccmFN*, *cob*, *ccmFN*, *cox2* (+ intronic sequence), *nad6* and *rps4* genes. Bootstrap values (1100 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. *: accessions from Kozik et al., 2019.

	L. sativa			*L. serriola*			*L. virosa*			*L. saligna*
	Wendel	UC 12100	LAC 04500	CGN 130219	CGN 004799	LAC 05780	CGN 013357	CGN 019045	IVT280 x CGN 16273	LAC 08020
Repeat_1	34 696 (2)	34 696 (2)	34 696 (2)	34 534 (2)	31 165 (2)	30 716 (2)	35 098 (2)	34 705 (2)	35 098 (2)	31 640 (2)
Repeat_2				34268 (3)					34 705 (3)	
Repeat_3	10430 (2)	10430 (2)	10430 (2)	10430 (2)	10430 (2)	10430 (2)	10430 (2)	10 430 (2)	10430 (2)	10430 (2)
Repeat_4						10137 (2)				
# >10 000	2	2	2	3	2	3	2	2	3	2
Repeat_5					8 524 (2)					
Repeat_6							4 116 (3)	4 116 (3)		
Repeat_7	3 552 (2)	3 552 (2)	3 552 (2)		3 905 (2)	3 905 (2)	3 552 (2)	3 447 (2)	3 552 (2)	4 048 (2)
Repeat_8					1 613 (2)					
# >1 000	3	3	3	3	5	4	4	4	4	3
Repeat_9	576 (3)	576 (3)	576 (3)	597 (2)	576 (3)	449 (3)				877 (3)
Repeat_10								393 (2)	266 (2)	
Repeat_11	215 (2)	215 (2)	215 (2)	215 (2)	215 (2)	215 (2)	215 (2)	215 (2)	215 (2)	215 (2)
Repeat_12	186 (2)	186 (2)	186 (2)	186 (2)	186 (2)	186 (2)	186 (2)	186 (2)	186 (2)	186 (2)
Repeat_13	137 (2)	137 (2)	137 (2)		137 (2)	137 (2)	137 (2)	137 (2)	131 (2)	131 (2)
Repeat_14	131 (2)	131 (2)	131 (2)	131 (2)	131 (2)	131 (2)	131 (2)	128 (2)	123 (2)	123 (2)
Repeat_15						127 (3)*				
Repeat_16	123 (2)	123 (2)	123 (2)	123 (2)	123 (2)	123 (3)	123 (2)	123 (2)		122 (2)
Repeat_17	119 (3)	119 (3)	119 (3)	119 (3)	119 (3)	119 (3)	119 (3)	119 (3)	119 (3)	119 (3)
Repeat_18	117 (2)	117 (2)	117 (2)	117 (2)	117 (2)	117 (2)	117 (2)	117 (2)	117 (2)	117 (2)
Repeat_19										112 (2)*
# >100	11	11	11	10	13	13	11	12	11	12

Table III.3. List of the repeated sequences of the different mtDNA built in this chapter.

insertion of *rpoC1,* while the three accessions of *L. serriola* and the accession of *L. saligna* show the same insertion of *psaB* (**Figure III.11**). These results are surprising with respect to the published *Lactuca spp.* mtDNAs and our phylogeny results from chloroplastic and nuclear markers. Indeed, the sequence of the chloroplastic insertions suggest that the mtDNA of *L. virosa* and of *L. sativa* are closer in evolution than in the two other species. In fact, these results are consistent with a result that we obtained by the alignment of concatenated sequences *ccmFC*, *cob*, *cox2* (+ intronic sequence), *nad6* and *rps4* (**Figure III.12**). These results and their possible causes are discussed further below.

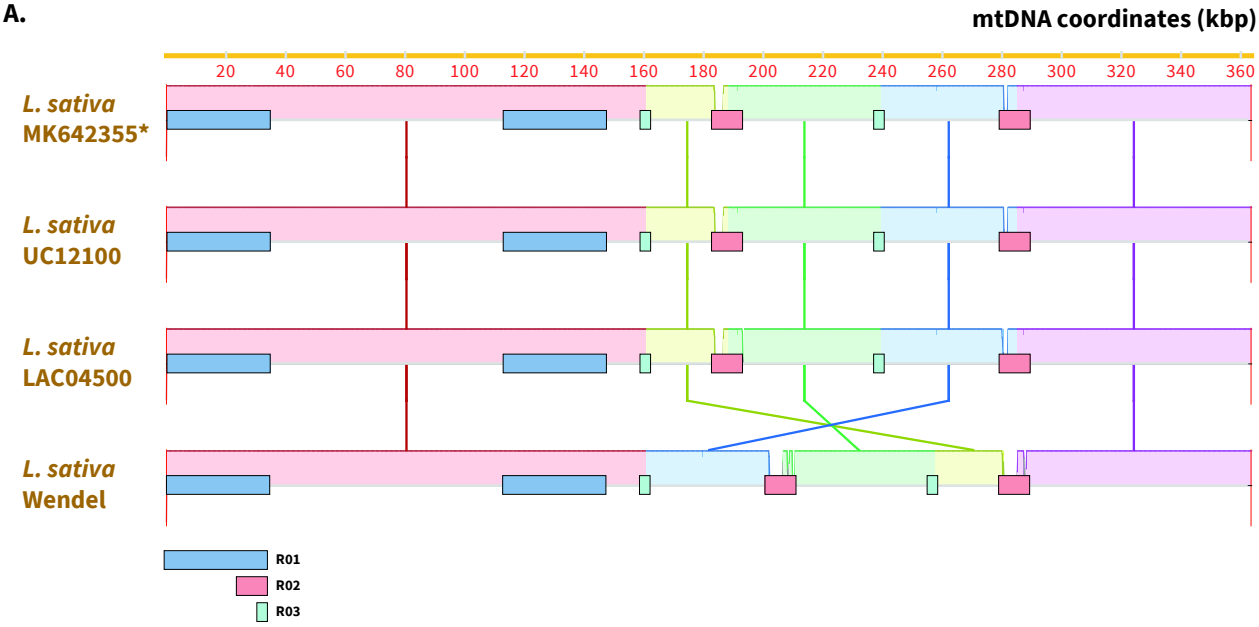
We also analyzed the repeated sequences carried by these genomes by using the python script *ROUSfinder* (**Table III.3**) (Wynn and Christensen, 2019) The genomes that we assembled contain 60 pairs of repeated sequence, in *L. sativa*, up to 68 in *L. saligna*. All these genomes bear a large repeated sequence (30 716 pb to 35 098 pb) homologous to repeat R01 from Wendel, containing the three rRNA genes. Two genomes, the ones from *L. serriola* CGN130219 and *L. virosa* IVT280 x CGN16273 contain a third smaller copy of their largest repeat, but as it was discussed above this probably results from assembly. The large pair of repeats R02 (10 430 bp) from Wendel is also found in all the mtDNAs that we built. It is extremely conserved, and diverges only in *L. saligna* LAC08020, where it has an insertion of 3 nucleotides. In addition, two accessions of *L. serriola* (CGN004799 and LAC05780) show a second repeated sequence about the size of R02, but not related. As well as repeat R03 from Wendel, which is also conserved in all accessions studied, at the one exception of *L. serriola* CGN 130219. Interestingly, this accession does not carry repeated sequences of sizes comprised between 1 000 bp and 10 000 bp. This could be because of mistakes introduced during hand assembly.

3.2. Genome structures

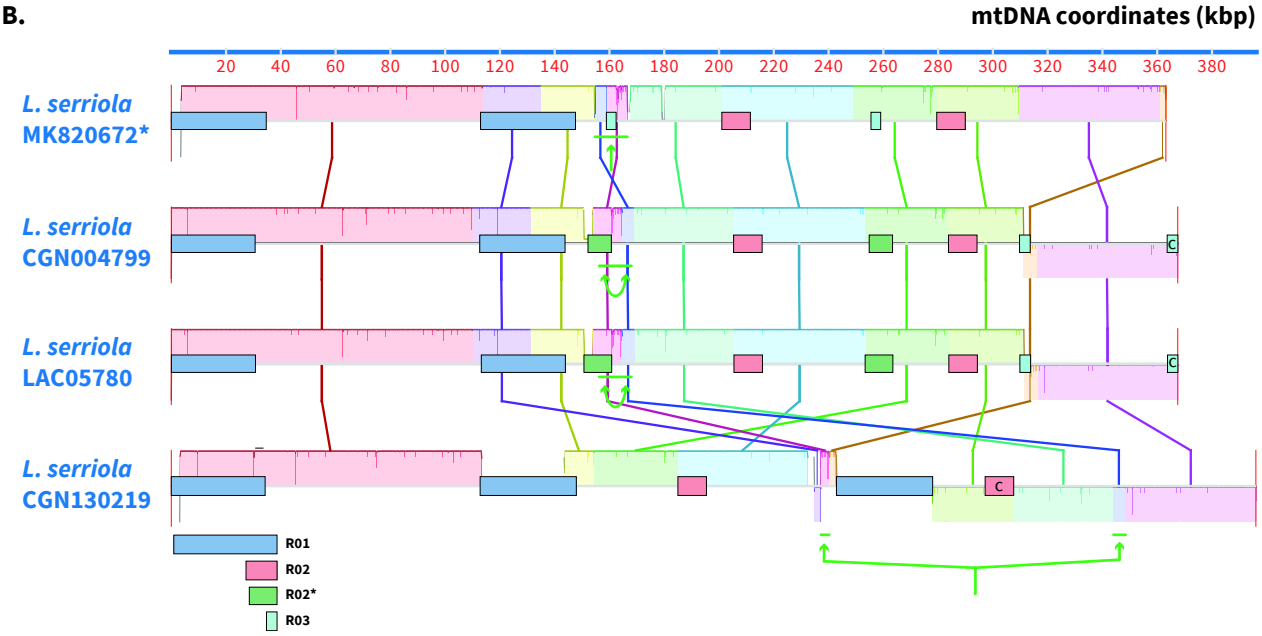
The frequent events of HR involving the large repeated sequences are responsible for the dynamic structure of the mitochondrial genome in plants. Consequently, the diversity of the mtDNA is mainly found in its structure. To evaluate the structural diversity present within the genomes that we built we aligned them per species with *progressive Mauve* (Darling et al., 2010). We used the genomes assembled by Kozik and collaborators as reference for our alignments.

The only structural diversity that we could observe among the *L. sativa* sequences involved the repeated sequences R03 and R02, through a mechanism of recombination (**Figure III.13.B**).

A.


mtDNA coordinates (kbp)
20 40 60 80 100 120 140 160 180 200 220 240 260 280 300 320 340 360
L. sativa MK642355*
L. sativa UC12100
L. sativa LAC04500
L. sativa Wendel
R01
R02
R03

B.


mtDNA coordinates (kbp)
20 40 60 80 100 120 140 160 180 200 220 240 260 280 300 320 340 360 380
L. serriola MK820672*
L. serriola CGN004799
L. serriola LAC05780
L. serriola CGN130219
R01
R02
R02*
R03

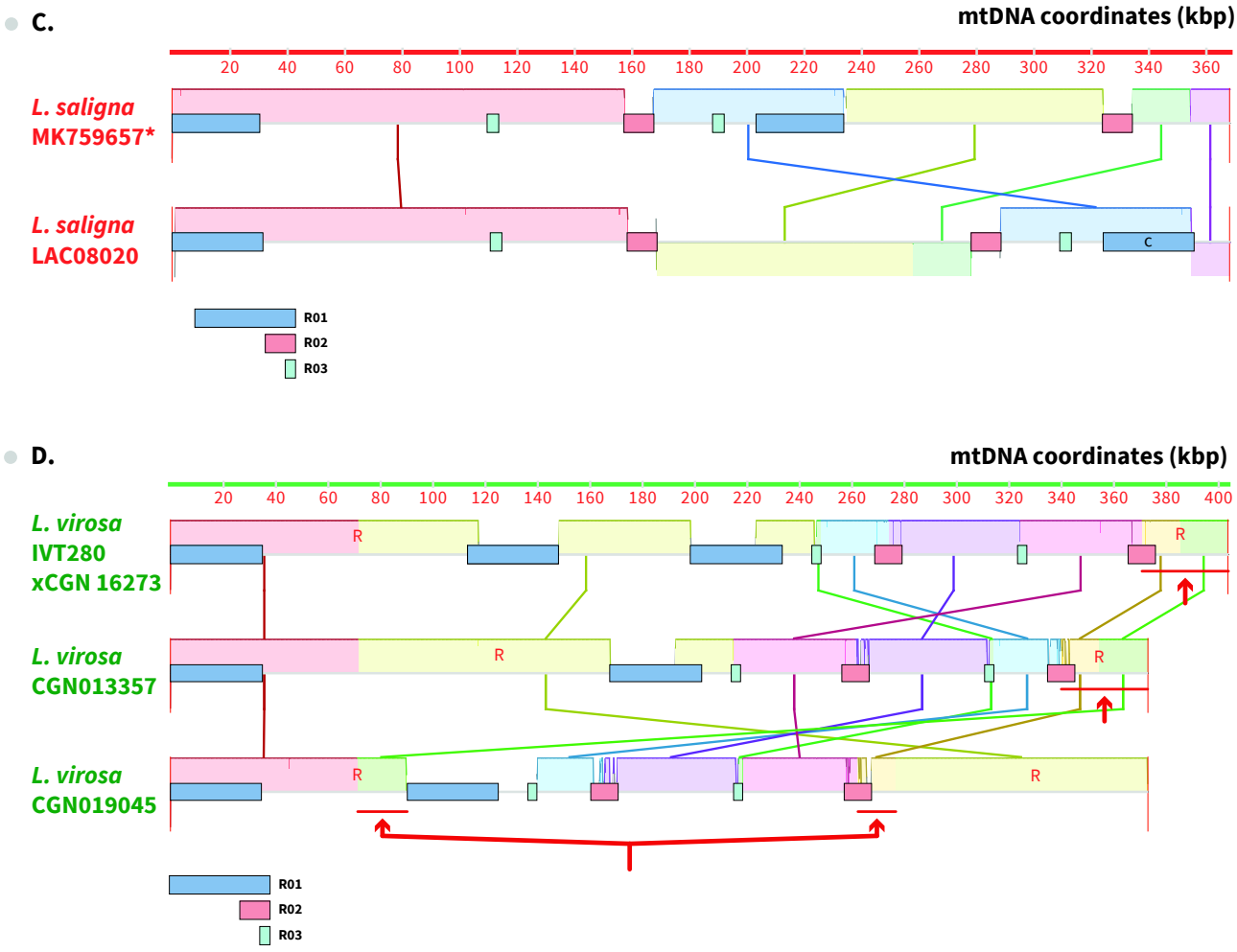


Figure III.13. Conservation of the mtDNA structure among the Lactuca spp. Alignment made with progressive MAUVE. Regions of significant synteny between the strains are shown as colored blocks in the mauve alignment. (**A**) *L. sativa*. (**B**) *L. serriola,* the recombination event of interested discussed in the main text is marked by green arrows, (**C**) *L. saligna* and (**D**) *L. virosa,* R*:* small repeat of 177 bp involved in the recombination event discussed in the main text*.* *: accessions from Kozik et al., 2019.

Indeed, the mitogenomes of *L. sativa* UC12100, LAC04500 and MK642355 have the same structure (which correspond to C7 in **Figure II. 5**) while Wendel corresponds to the C1 configuration. Such events involving large repeats reflect the dynamic structure of the mitochondria, and are not consequences of evolution. However, we found a SNP located in the *rpoC1* chloroplastic insertion, that is a T at base 257 991 in *L. sativa* UC12100, while a C is found at the corresponding position in the other *L. sativa* accessions and in the three *L. virosa* that we studied.

Within *L. serriola*, in accessions LAC05780 and CGN004799 the mtDNA has the same structure, which is close to MK820672 that we used as reference (**Figure III.13.B**). However, a region is inverted between the mtDNA of LAC05780-CGN004799 and MK820672 and split in CGN130219 (shown in green on **Figure III.13B**). But there is no repeated sequence or microhomology that could explain such event, that probably involved non-homologous ends joining. The other differences found between the mtDNA of these accessions and the one of the reference can be explained by events of recombination involving small and large repeats. Furthermore, at the extremities of their sequences LAC05780 and CGN004799 have an inverted repeat of 3905 bp, (indicated in turquoise in the figure). Most of this repeat is lost in MK820672, where only a fragment of 877 bp remains. Surprisingly, this fragment of 877 is carried three times by the large repeated sequences of CGN130219. However, the structure of CGN130219 is hard to analyze due to the third copy of repeat R01, which increases the number of possible alternative configurations of the genome that might exist.

The mtDNA of the *L. saligna* that we assembled, LAC08020, presents the particularity of having an inverted pair of the largest repeated sequence. However, this pair of repeats share with another portion of the genome a smaller and inverted repeated sequence of 877 bp, which can consequently isomerize the sequence between them by recombination, leading to a genome structure similar to the published one of reference MK759657 (**Figure III.13.C**).

Regarding the genomes of the *L. virosa* accessions that we studied, an event of recombination led a region to be split in CGN019045 as compared to the genomes of the other two accessions (shown in red on **Figure III.13.D**). We could link this event to a small repeated sequence of 177 bp located between the regions green and yellow and noted R in **Figure III.13.D**.

All the genomes, at one very exception (*L. virosa* CGN019045) present a large region of about 120 kb extremely conserved. This region is comprised between the two copies of the largest repeat

R01, except in *L. saligna* where the second copy of R01 has moved to another locus of the genome by recombination. This conserved region corresponds to subgenome X represented in **Figure II. 5**, and to our *L. sativa* circular chromosome PBc02.

DISCUSSION AND PERSPECTIVES

In this chapter we aimed to study the natural mtDNA diversity present within four species of *Lactuca spp*., which were *L. saligna*, *L. sativa*, *L. serriola* and *L. virosa*. To do so, we had to build the mtDNA of these species after evaluating the nuclear and chloroplastic diversities existing in this population to validate the genetic origin of the accessions that we studied.

The study of the nuclear and chloroplastic diversity within our accessions is in line with what is suggested by the theory about the domestication of *L. sativa*. Indeed, according to this theory *L. serriola* is one of the direct ancestors (de Vries, 1990; de Vries and van Raamsdonk, 1994). And according to these results*, L. saligna* and *L. virosa* are distant from the pair *L. serriola-L. sativa*. It is also accepted that *L. virosa* played a role in the domestication of *L. sativa,* which could explain the robust root system of the Crisphead lettuces. The mitochondrial sequences agree with that possibility. Indeed, *L. virosa* and *L. sativa* share the same chloroplastic insertion, while a different cpDNA sequence is found at the same locus in *L. serriola* and *L. saligna*. The possibility that the exact same event of horizontal transfer happened twice in evolution is unrealistic. Moreover, this result is corroborated by the phylogeny built from the sequences of *ccmC*, *cob*, *cox2*, *nad6* and *rps4*. These results suggest that Crisphead lettuce could derive from a cross where *L. virosa* was the mother. However, this hypothesis does not fit with the phylogeny result that we obtained based on the chloroplastic markers.

The nine genomes that we assembled result from a possible order of scaffolds inferred from overlapping sequences, but they are not validated by experimental data yet. Nevertheless we could already fetch interesting data from the genomes comparisons. Indeed, the chloroplastic insertion as well as the events of recombination identified in this chapter could be used as interesting markers in certain contexts. The chloroplastic insertion differs between *L. sativa-L. virosa* and *L.-saligna-L. serriola,* and could represent a mitochondrial marker for crosses between parents from each pool. Recombination events do not seem to be conserved and thus could be used as markers to cross a specific accession to another from a same specie. Furthermore, the ancestral events of

recombination leading to mtDNA diversity could provide clues for the project aiming to induce mtDNA diversity in *L. sativa*.

CHAPTER 4: Induction of mitochondrial genome diversity within L. sativa

There are several ways to increase or induce mtDNA recombination and, consequently, generate mtDNA diversity (**Figure I.16**). Some of these strategies produce lines that are considered as genetically modified organism (GMO), according to the European legislation. The production or use of GMOs is not in line with the policy of Enza Zaden. That is why in the work described in this chapter we focused on techniques that do not require GMOs. These were the indirect induction of mtDNA ectopic recombination using non-GMO mutants for genes involved in mtDNA maintenance and segregation (GIMMS) and the direct induction of mtDNA recombination using the genotoxic agent ciprofloxacin (CIP).

1. GIMMS and the indirect increase of mtDNA ectopic recombination

The GIMMS are nuclear genes coding for mitochondrial-targeted proteins, several of which are homologous to bacterial factors involved in bacterial DNA recombination. Consequently, mutants for these genes should have an impact on the recombination of the mtDNA. This has been shown for several genes, mostly in *A. thaliana*. For instance, in plants, the *MSH1* gene codes for a protein (targeted to both organelles) homologous to bacterial MutS, involved in mismatch repair. In *A. thaliana*, the *msh1* mutants have enhanced intramolecular ectopic recombination activity, detected as apparent amplification of novel mtDNA forms (Abdelnoor et al., 2003). Similar results were obtained in tobacco and tomato, where the suppression of *MSH1* expression by RNAi leads to increased mtDNA diversity by recombination (Sandhu et al., 2007). Similarly, mutants *osb1* and *osb4* deficient in the expression of the mitochondrial-targeted ssDNA-binding mediators of recombination OSB1 and OSB4 also show an increase in ectopic recombination in *A. thaliana* (Zaegel et al., 2006 and unpublished data). Such an increase in recombination has also been highlighted in mutants for DNA helicases RADA and RECG1, homologous to bacterial RecG and RadA, respectively, involved in alternative pathways of branch migration during HR. In *A. thaliana* both RADA and RECG1 are dually targeted to mitochondria and to chloroplasts (Wallet et al., 2015; Chevigny et al., 2020). Finally, *A. thaliana* mutants for homologs of bacterial recombinase RecA, RECA2 and RECA3, targeted to both organelles and to mitochondria respectively, show an increase in mtDNA recombination involving intermediate-size repeats (Shedge et al., 2007; Miller-Messmer et al., 2012). However, these two organellar recombinases have differential functions: *A. thaliana*

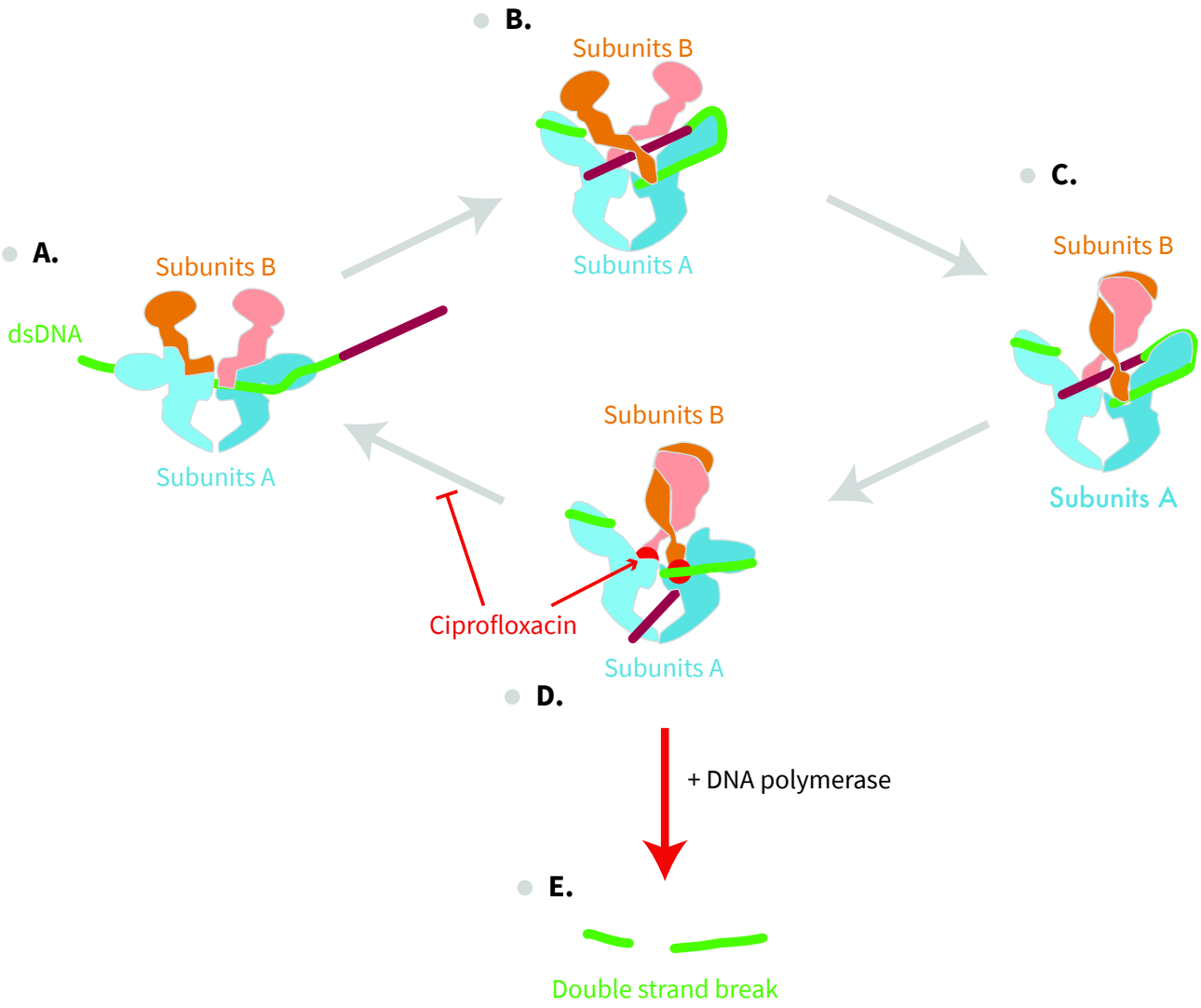


Figure IV.1. DNA gyrase activity and inhibition by ciprofloxacin. (**A**) Binding of ATP molecules causes the dimerization of the ATPase domains of the GyrB subunits. (**B**) Capture of DNA in a cavity between the GyrA and GyrB subunits. (**C**) On the next step, the enzyme cleaves the DNA, making a double-strand break. (**D**) The DNA is then transferred through the break. (**D**-**A**) Finally, the DNA-gyrase ligates the break and the DNA is released. These last actions are inhibited by ciprofloxacin which binds to the subunit GyrA of the gyrase, repressing the ligase activity. The blocked complex will be further removed by DNA polymerase during replication, leaving a double-strand break in the DNA.

recA2 mutants show a seedling lethal phenotype, while *recA3* plants are viable. However, mutants for the organellar DNA polymerase POL1B and for WHY2, a mitochondria-targeted ssDNA-binding mediator of recombination, do not show an increase in mtDNA recombination in normal grow conditions (Cappadocia et al., 2010; Parent et al., 2011). This is also true for the double mutant *pol1B* w*hy2*. But these mutants show increased sensitivity to mitochondrial genotoxic stresses. Such genotoxic stress can be mediated by the antibiotic CIP (Cappadocia et al., 2010; Parent et al., 2011).

2. Ciprofloxacin, an organelles genotoxic agent inducing mtDNA DSBs

CIP is a fluoroquinolone broad-spectrum antibiotic effective against a wide range of Gram-positive and Gram-negative bacteria, which acts by interfering with bacterial DNA gyrase, a type II DNA topoisomerase that is necessary for bacterial genome replication and transcription (**Figure IV.1**) (Gellert et al., 1976). Gyrase is a heterotetrameric enzyme composed of two GyrA and two GyrB subunits. Its bacterial role is to release DNA topology strain by creating a double-strand break (DSB), passing duplex DNA through the break, and then resealing the break. In bacteria, the subunit GyrA of gyrase is the target of CIP (Shen and Pernet, 1985). The binding to GyrA of a fluoroquinolone antibiotic represses the closing of the double-strand break induced by the gyrase, which remains bound to the DNA (**Figure IV.1**). In plants, genes coding for the two subunits of a DNA gyrase (GYRA and GYRB) are located in the nucleus and the proteins are targeted to both organelles. Like in bacteria, the plant organellar GYRA is sensitive to CIP, and CIP-treated plants are severely affected in their development (Evans-Roberts et al., 2016). The DSBs induced by CIP through the inhibition of DNA gyrase triggers repair by recombination and consequently, in plant mitochondria, to an increase in mitochondrial recombination.

3. Evaluation of the lettuce sterility

Beyond the induction of mitochondrial diversity, developing a cytoplasmic male sterility (CMS) system in *L. sativa* would provide tangible breeding applications. But to do so, we need tools to study sterility and strategies to determine its cytoplasmic origin.

The sterility needs to be studied at three different levels, according to the three different types of male sterility. Possible homeotic sterility must be checked by looking at the presence or

Strategies	Tools	In-field ease-of-use	Pollen defficiency tested	
			Wall	Cytoplasm
Enzymatic assay	FDA-test	+	x	
Tissue staining	Alexander staining	+++		x
Impedance flow cytometry	Ampha Z32®	-	x	x

Table IV.1. Tools used by Enza Zaden to determine pollen viability

Gene	Chromosome	Strand	Start codon position	Stop codon position
MSH1	2	-	95 589 978	95 580 708
OSB1	1	-	22 962 292	22 960 645
OSB2*	8	-	96 557 439	96 555 183
RADA	4	+	162 593 726	162 597 620
RECA3	9	-	34 496 942	34 494 765
RECG1	3	+	71 225 633	71 231 531

Table IV.2. Candidate genes involved in mtDNA maintenance and segregation and gene localization in the *L. sativa* nuclear genome. *: Because of the absence of *OSB4* in *L. sativa* we used mutant for *OSB2*

absence of male organs within the flowers. If the flowers do not show such a phenotype, the pollen production is then controlled, with respect to potential sporophytic male-sterility. And finally, if the flowers produce pollen, its viability must be checked. To do so three strategies are routinely used at Enza Zaden (**Table IV.1**). The first is the fluorescein diacetate test (FDA-test) used to determine membrane integrity. FDA is a substrate of esterases and the enzymatic reaction produces fluorescein, in viable cells with an integer membrane and functional enzymatic activity, and so viable pollen. The second strategy is Alexander staining (Alexander, 1969). This stain contains malachite green, acid fuchsin and orange G. Malachite green colors the membrane, and acid fuchsin stains viable cells with polysaccharides in the protoplasm. Consequently, by Alexander staining viable pollen shows red to deep red cytoplasm with a green callose wall, while the aborted pollen grains that only have walls are stained green. The last strategy to analyze pollen viability is impedance flow cytometry, using the *Ampha Z32* (Amphasys AG). It relies on the differential conductivity between viable pollen, with integer cytoplasm and membrane, and non-viable pollen with defects in membrane or cytoplasm.

To determine the cytoplasmic origin of a male-sterile phenotype, the transmission of the phenotype to the progeny needs to be determined. In the case of a CMS the phenotype is uniformly transmitted to the progeny through cross with a maintainer line, carrying the same nuclear genetic content as the male-sterile one.

4. Aim of this chapter

In this chapter the aim was to generate mtDNA diversity in the Wendel *L. sativa* Butterhead accession, by using mutants for GIMMS. These mutants have been generated by Enza Zaden, through ethyl methanesulfonate (EMS) mutagenesis. Among them we focused on one mutant that showed a severe development phenotype associated with mtDNA recombination. Also, we initiated a pilot study of the treatment of lettuce with CIP, to directly induce mitochondrial recombination through the repair of DSBs.

Figure IV.2. Phylogenetic tree of the OSB protein family. This tree supports the identification of the OSB1 and OSB2 from lettuce, which branch together with *A. thaliana* OSB1 and *A. thaliana* OSB2 and OSB4, respectively. The numbers attached to the species names are the number of PDF motifs in the protein, at the very exception of (*) *Arabidopsis thaliana 4* which carries only two PDF motifs. Bootstrap values are indicated.

1. Indirect induction of mitochondrial genome diversity

At the beginning of the project we established a shortlist of 8 candidate genes, based on the literature and previous work from the laboratory. These candidates are *MSH1*, *OSB1*, *OSB4*, *POL1B*, *RADA*, *RECA3*, *RECG1* and *WHY2*. However, we did not succeed in getting mutants for *POL1B* and *WHY2*. Consequently we will not discuss these genes any further.

1.1. Identification of the candidate genes in *L. sativa*

The genes *MSH1*, *OSB1*, *OSB4*, *RADA*, *RECA3*, and *RECG1* have been identified and mainly studied in *A. thaliana*. Therefore our first task was to identify the orthologs of *L. sativa*. We had to find the genomic sequences, determine the gene structures and confirm the mitochondrial targeting of the encoded proteins. We identified the genomic sequences by BLAST search on the nuclear genome of *L. sativa* (Altschul et al., 1990; Reyes-Chin-Wo et al., 2017). Such *BLAST* was done using the bacterial and plant homologs to the desired genes as query sequences. The identification of the *L. sativa* genes coding for MSH1, RADA and RECG1 was unambiguous (**Table IV.2**). However, the identification of the *OSB1*, *OSB4* and *RECA3* genes needed a more careful analysis.

The genes coding for the OSB family are plant specific and do not have bacterial homologs. Among plants, proteins of the OSB family share a C-terminal motif of about 50 amino acids called “PDF”, because these three residues are completely conserved in that motif, which is involved in ssDNA binding (Zaegel et al., 2006). The proteins in each species are named “OSB#” where “#” represents the number of PDF sites in the proteins. However, *A. thaliana* carries two proteins with 2 PDF motifs, OSB2 and OSB4. The difference between these two proteins lies in their targeting: OSB2 is targeted to the chloroplast while OSB4 is targeted to mitochondria (unpublished result from the laboratory). In *L. sativa*, there are only two genes coding proteins with PDF motifs: OSB1 and OSB2 with 1 and 2 PDF motifs, respectively (**Figure IV.2**). The sequences comparison indicates that *L. sativa OSB2* is the ortholog of *A. thaliana OSB4. OSB1* and *OSB2* are located in chromosomes 1 and 8, respectively.

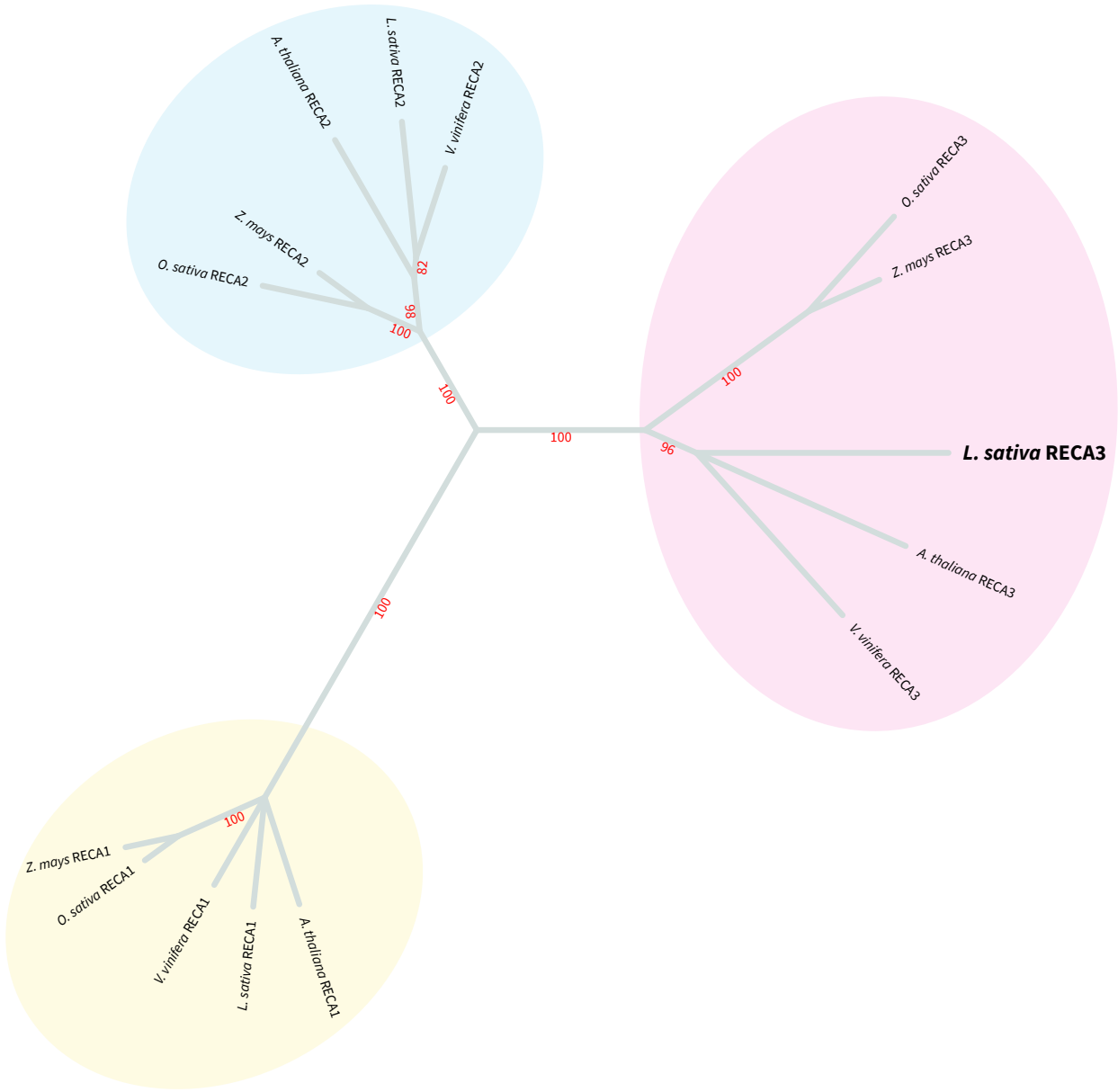


Figure IV.3. Phylogenetic tree of the higher plants RECA protein family. This tree supports the identification of RECA1, RECA2 and RECA3 from lettuce. The annotation of RECA2 and RECA3 was switched regarding what can be found in NCBI, to fit with the *A. thaliana* annotation. Bootstrap values are indicated.

In *A. thaliana*, there are three proteins homologous to bacterial RecA: RECA1, RECA2, and RECA3. RECA1 and RECA3 are targeted to chloroplast and mitochondria, respectively, whereas RECA2 is targeted to both organelles (Shedge et al., 2007). The two paralogs targeted to mitochondria, RECA2 and RECA3, are quite close in sequence, but are not redundant. The knockout *recA2* mutant is lethal at the seedling stage, while the *recA3* mutant is viable and not severely affected in its development (Miller-Messmer et al., 2012). In *L. sativa* we were able to find the three genes of the *RECA* family and to discriminate RECA3 from the two other encoded proteins, mainly because RECA3 lacks a C-terminal acidic sequence found in the other RECA proteins (**Figure IV.3**).

Once the genomic sequences identified, before further analysis we had to confirm the genes structures. For *MSH1*, *OSB1, OSB2* and *RADA* the proteins predicted from annotated genes models aligned correctly with the orthologs from other species (**Figure IV.5 to IV.10**) and were considered correct. For *OSB2* there exists an Expressed Sequence Tag (EST) sequence (accession DW087553) that confirmed the gene model. For *RECG1*, we used several available EST to correct the gene model. For *RECA3*, we had to be sure of the correct N-terminal part of the protein, since there are chloroplast and mitochondrial paralogs in the RECA protein family. We have therefore determined it by RT-PCR and found that the existing gene model was wrong. Indeed, the existing gene model lacks the first exon (102 bp) and the 5' of the second one (30 bp). This region is important because potentially carrying the organellar targeting sequence (OTS).

To confirm the mitochondrial targeting of these *L. sativa* proteins we evaluated it first by using the prediction software *Predotar* and *TargetP2.0* (Small et al., 2004; Almagro Armenteros et al., 2019) (**Table IV.3**). Both software agreed on the mitochondrial targeting of OSB2 and RADA, the chloroplast targeting of MSH1 and RECA3 and the undetermined targeting of RECG1. But *Predotar* was unable to determine the targeting of OSB1 while *TargetP2.0* identified it as targeted to mitochondria. However, these software are not reliable to predict dual targeting, because none of the 6 proteins was identified as dual-targeted, whereas three of them are, in *A. thaliana* (MSH1, RADA and RECG1) (**Table IV.3**). Consequently, confirmation of the subcellular localization required *in vivo* studies. To do so we cloned the N-terminal part of these proteins fused in frame with eGFP. The sequences were selected from the analysis of the alignments with the orthologs from other species, because organellar targeting sequences are poorly conserved in evolution. The genomic sequences coding for these N-terminal parts (MSH1: 193 aa, OSB1: 90 aa, OSB2: 126 aa, RADA: 196 aa, RECA3:136 aa) were amplified by PCR and BamHI and KpnI sites introduced by PCR were used to

Protein	Subcellular targeting			*A. thaliana* targeting
	Prediction		GFP-fusion	
	Predotar	*TargetP*		
MSH1	**Cp**	**Cp**	Undetermined	**Mt** + **Cp**
OSB1	None	**Mt**	**Mt** + **Cp**	**Mt**
OSB2	**Mt**	**Mt**	**Mt**	**Mt**
RADA	**Mt**	**Mt**	**Mt** + **Cp**	**Mt** + **Cp**
RECA3	**Cp**	**Cp**	**Mt**	**Mt**
RECG1	None	None	Non-tested	**Mt** + **Cp**

Table IV.3. Subcellular localization of *L. sativa* factors orthologous to the selected *A. thaliana* proteins involved in mtDNA maintenance and segregation. Cp: Chloroplast; Mt: Mitochondria.

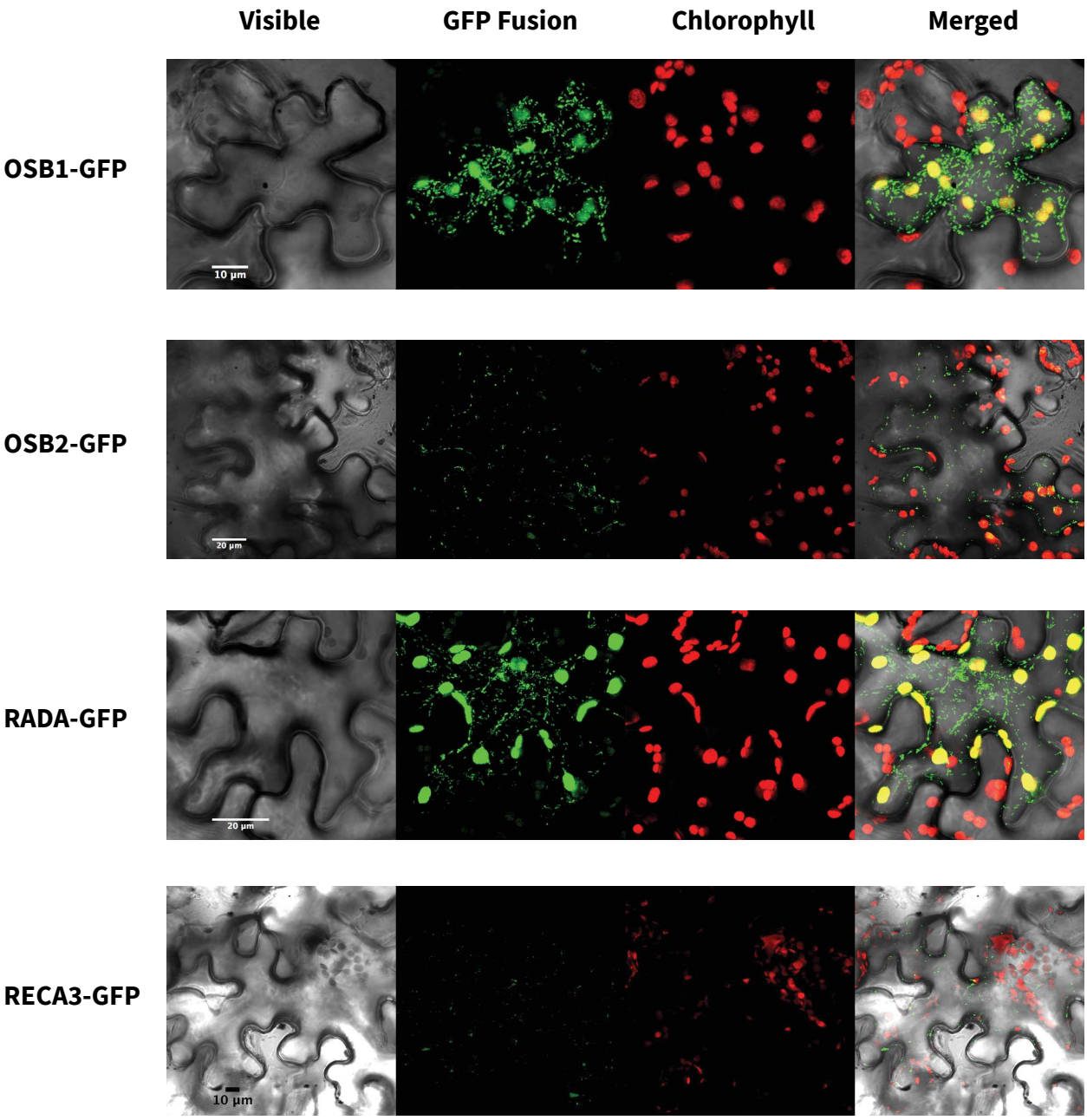


Figure IV.4. Subcellular localization of the selected genes involved in mtDNA maintenance and segregation in *L. sativa*. Transient expression of fusion proteins between the N-terminal part of different GIMMS (OSB1, OSB2, RADA and RECA3) and GFP, in *N. benthamiana* leaf epidermal cells transfected by biolistc. Chloroplasts are identified by the autofluorescence of chlorophyll.

Gene	Protein size	mutant	Mutation	Conservation of the position
MSH1	1065	*msh1-1*	D141>N	+++
		msh1-2	E155>K	+++
		msh1-3	Q157>*	
		msh1-4	R168>H	++++
		msh1-5	P583>S	+++
		msh1-6	P742>S	+++++
		msh1-7	E988>K	-
OSB1	286	*osb1-1*	Splicing deficiency between Q187 and D188	
		osb1-2	W204>*	
		osb1-3	W214>*	
OSB2	412	*osb2-1*	W140>*	
		osb2-2	W389>*	
RADA	602	*radA-1*	C128>Y	+++++
		radA-2	W139>*	
RECA3	433	*recA3-1*	A64>T	-
		recA3-2	L110>F	+++
		recA3-3	E190>K	+++++
RECG1	970	*recG1-1*	D83>N	+
		recG1-2	S99>F	+
		recG1-3	R375>*	

Table IV.4. *L. sativa* mutant lines obtained by EMS mutagenesis. The conservation of the amino acids have been evaluated according the alignment picture available in Figures 5 to 10.

clone in the corresponding sites of the pUCAP35S vector, in frame with eGFP and under the control of a double 35S promoter. In the case of RECA3, because the gene carries a BamHI site in its first exon the sequences was introduced in destination vector pB7FWG2 by Gateway® cloning, in frame with eGFP. The constructions were then used for transient expression following transfection of *Nicotiana benthamiana* leaves by biolistic. Doing so, we confirmed the strict mitochondrial targeting of OSB2 and RECA3 and the dual targeting of OSB1 and RADA (**Table IV.3** & **Figure IV.4**). While doing these experiments we also confirmed that WHY2 of lettuce is also only targeted into mitochondria, as in *A. thaliana*. However, we were not able to determine the targeting of MSH1, because we only observed week and diffuse fluorescence in the cytoplasm. We decided to postpone the determination of the subcellular localization of RECG1 due to a new orientation of the project, which is discussed below. But since there is only one *RECG1* gene, both in Arabidopsis and lettuce, we predict that the dual targeting of RECG1 observed in Arabidopsis is conserved in lettuce. To confirm the mitochondrial localizations we used the same biolistic approach to transfect *Nicotiana tabaccum* leaves of plants constitutively expressing a mitochondrial targeted red-fluorescent protein (Forner and Binder, 2007). Because of time and practical consideration this was just done for RADA and OSB2, as discussed in **CHAPTER 5 & 6**.

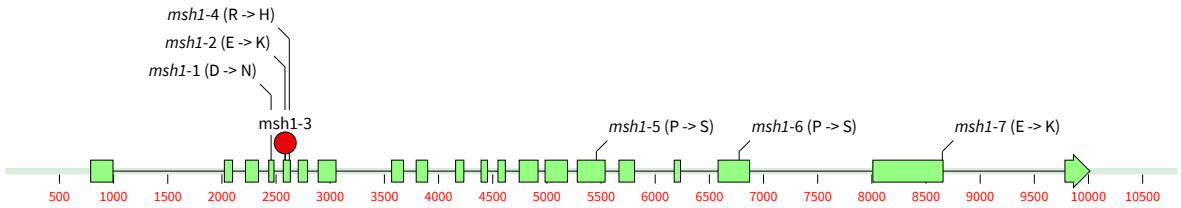
1.2. Study of the *L. sativa* mutant lines

1.2.1. Presentation of the mutant lines

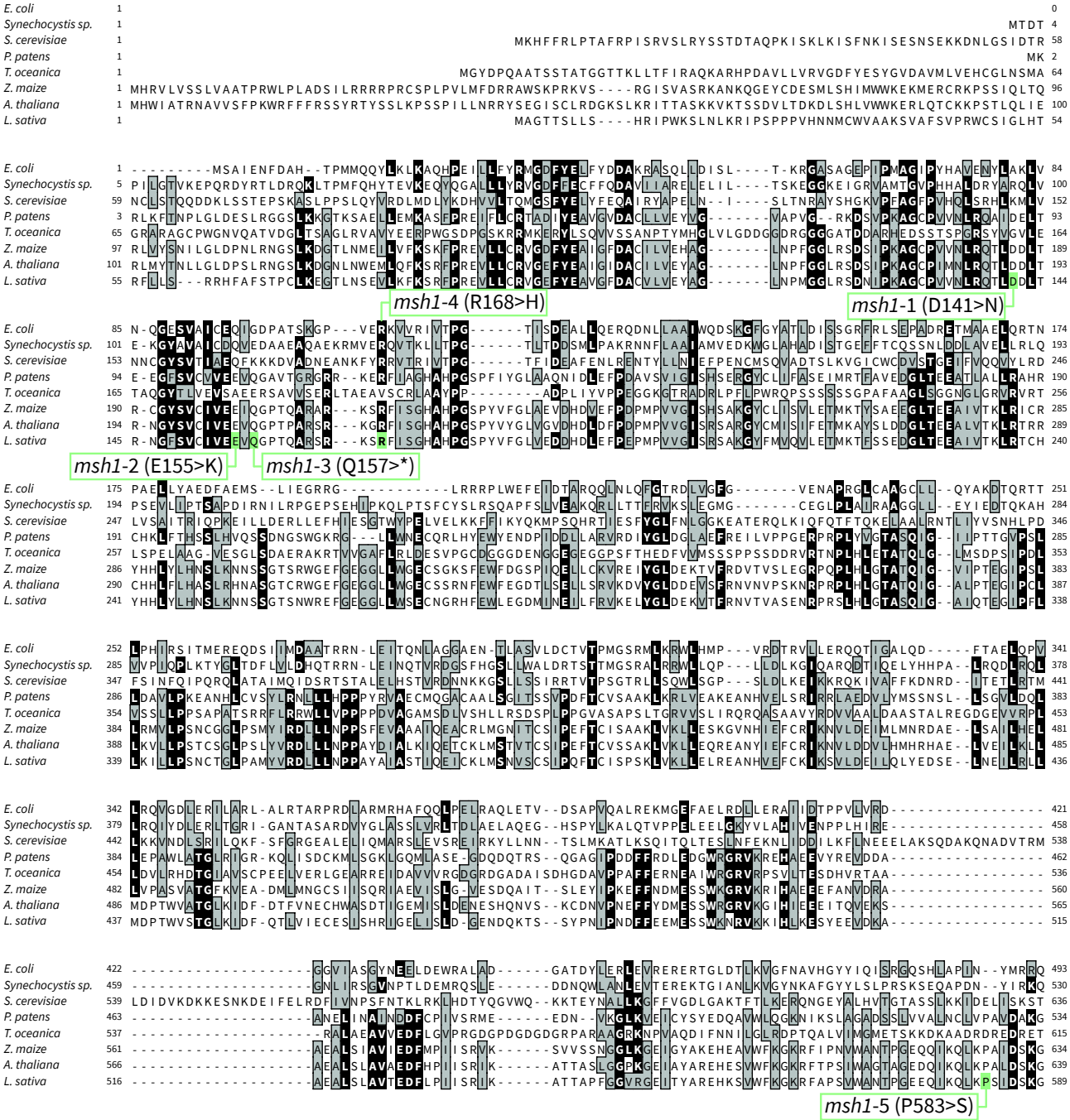
The *L. sativa* mutant lines for GIMMS have been generated by Enza Zaden by EMS treatment and selected by TILLING. The EMS treatment is a random mutagenesis strategy inducing G:C to A:T base substitutions. Consequently, it is possible to get from EMS treatment three kinds of mutation: Nonsense mutations through substitutions in arginine (**C**GA -> **T**GA), glutamine (**C**AA -> **T**AA and **C**AG -> **T**AG) and tryptophan (T**GG** -> T**A**G and TG**A**) codons, missense mutations, and splicing defects by mutations at important intron positions, like at intron borders (A**G** -> A**A** and **G**T -> **A**T).

The combination of EMS mutagenesis with high-throughput TILLING allowed the primal identification of a set of numerous mutant lines. These were pre-selected regarding the type of mutation (Nonsense > splicing deficiency > missense), the localization of the nonsenses and splicing deficiencies (First third of the protein > second > third) and the conservation of the amino acid for the missense mutations. Finally, the mutant lines have been certified through pooling. Resulting from this strategy we selected 20 mutant lines for 6 of the genes of interest (**Table IV.4**).

A.


msh1-4 (R -> H)
msh1-2 (E -> K)
msh1-1 (D -> N)
msh1-3
msh1-5 (P -> S)
msh1-6 (P -> S)
msh1-7 (E -> K)
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
5500
6000
6500
7000
7500
8000
8500
9000
9500
10000
10500

B.


E. coli
Synechocystis sp.
S. cerevisiae
P. patens
T. oceanica
Z. maize
A. thaliana
L. sativa
msh1-4 (R168>H)
msh1-1 (D141>N)
msh1-2 (E155>K)
msh1-3 (Q157>*)
msh1-5 (P583>S)

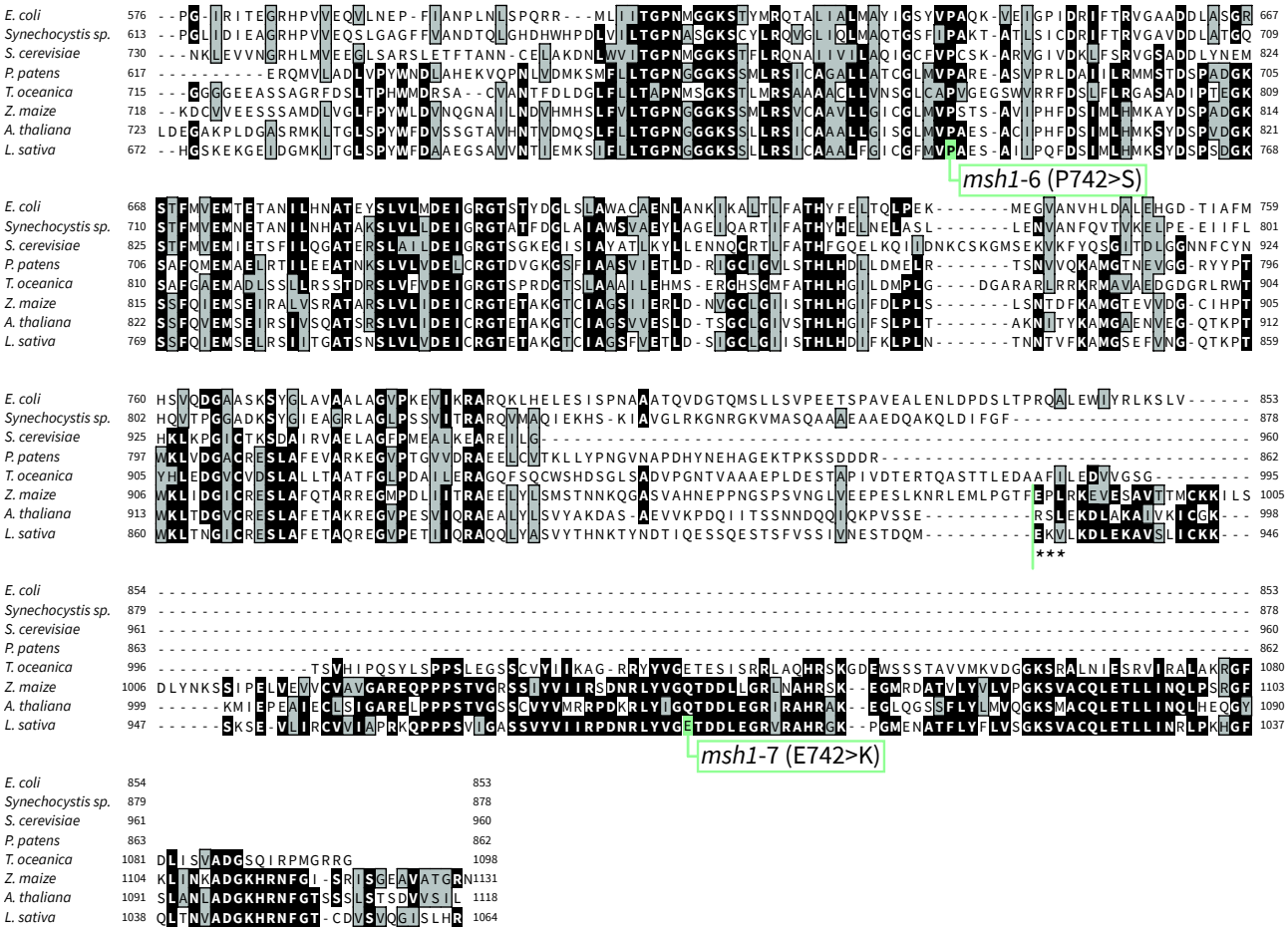


Figure IV.5. Selected *L.sativa* mutants for MSH1. (**A**) Structure of the *L. sativa MSH1* gene and localization of the selected mutations. Amino acid changes are represented by black lines and the nonsense mutation is represented by a red circle. (**B**) Sequence alignment (ClustalW) between representative MSH1 plant sequences and bacterial MutS. The localization of the selected mutations are indicated.

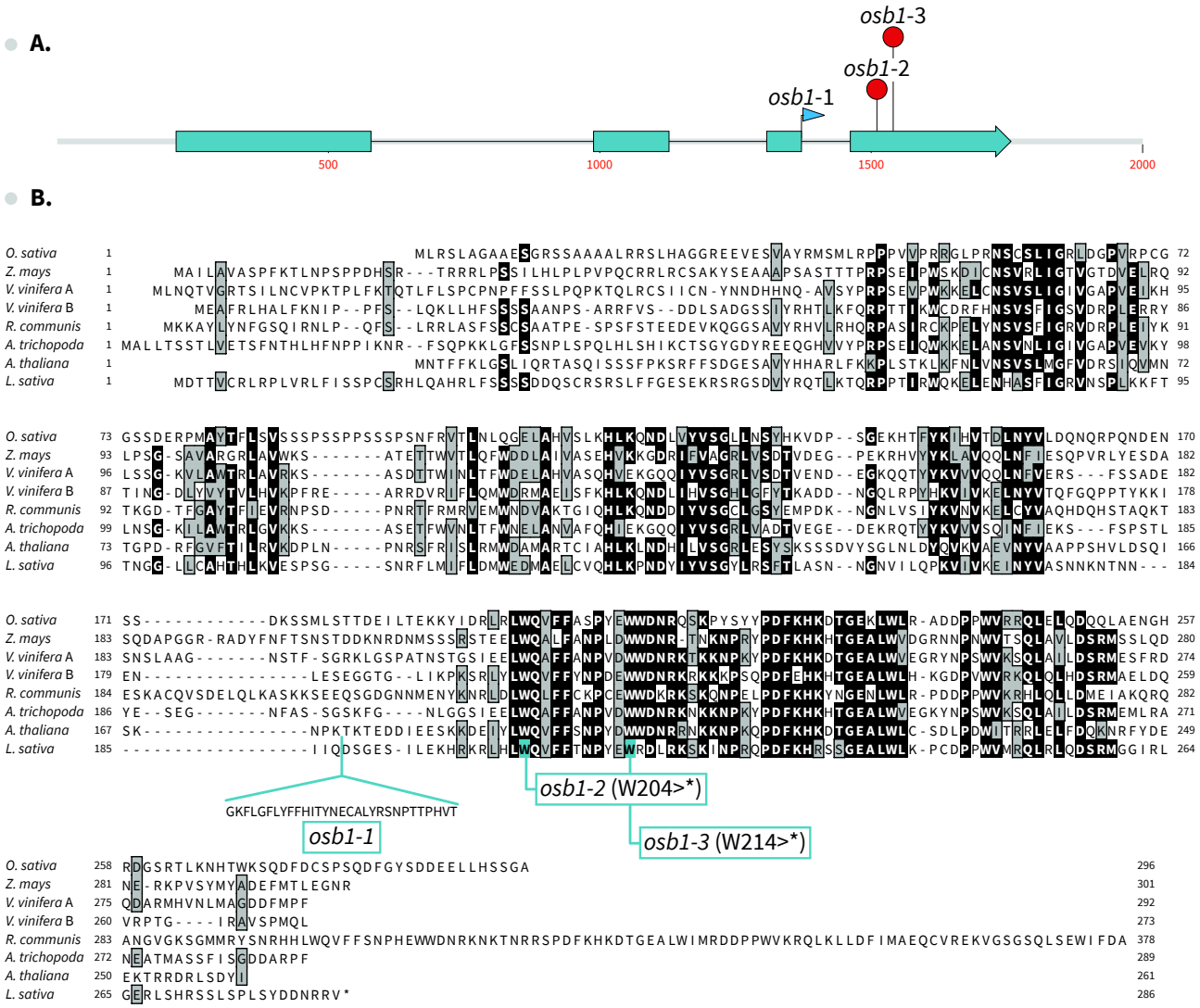


Figure IV.6. Selected *L.sativa* mutants for OSB1. (**A**) Structure of the *L. sativa OSB1* gene and localization of the selected mutations. Nonsense mutations are represented by red circles and the splicing deficiency is shown by a blue arrow. (**B**) Sequence alignment (ClustalW) between representative land plant OSB1 sequences. The localization of the selected mutations are indicated and the amino acids added by the splicing deficiency in *osb1-1* are shown.

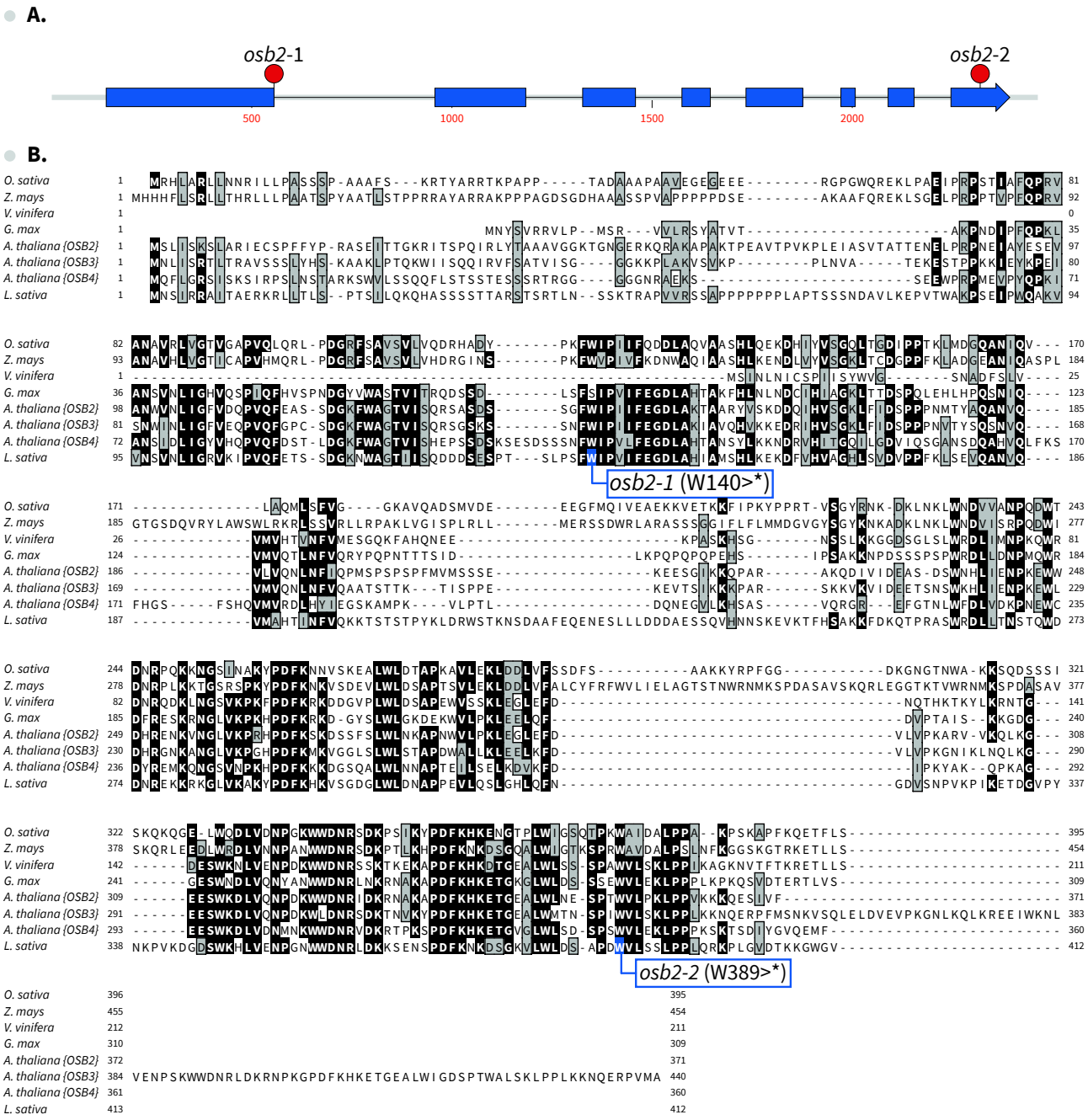


Figure IV.7. Selected *L.sativa* mutants for OSB2. (**A**) Structure of the *L. sativa OSB2* gene and localization of the selected mutations. Nonsense mutations are represented by red circles. (**B**) Sequence alignment (ClustalW) between representative land plant OSB2 sequences. The localization of the selected mutations are indicated.

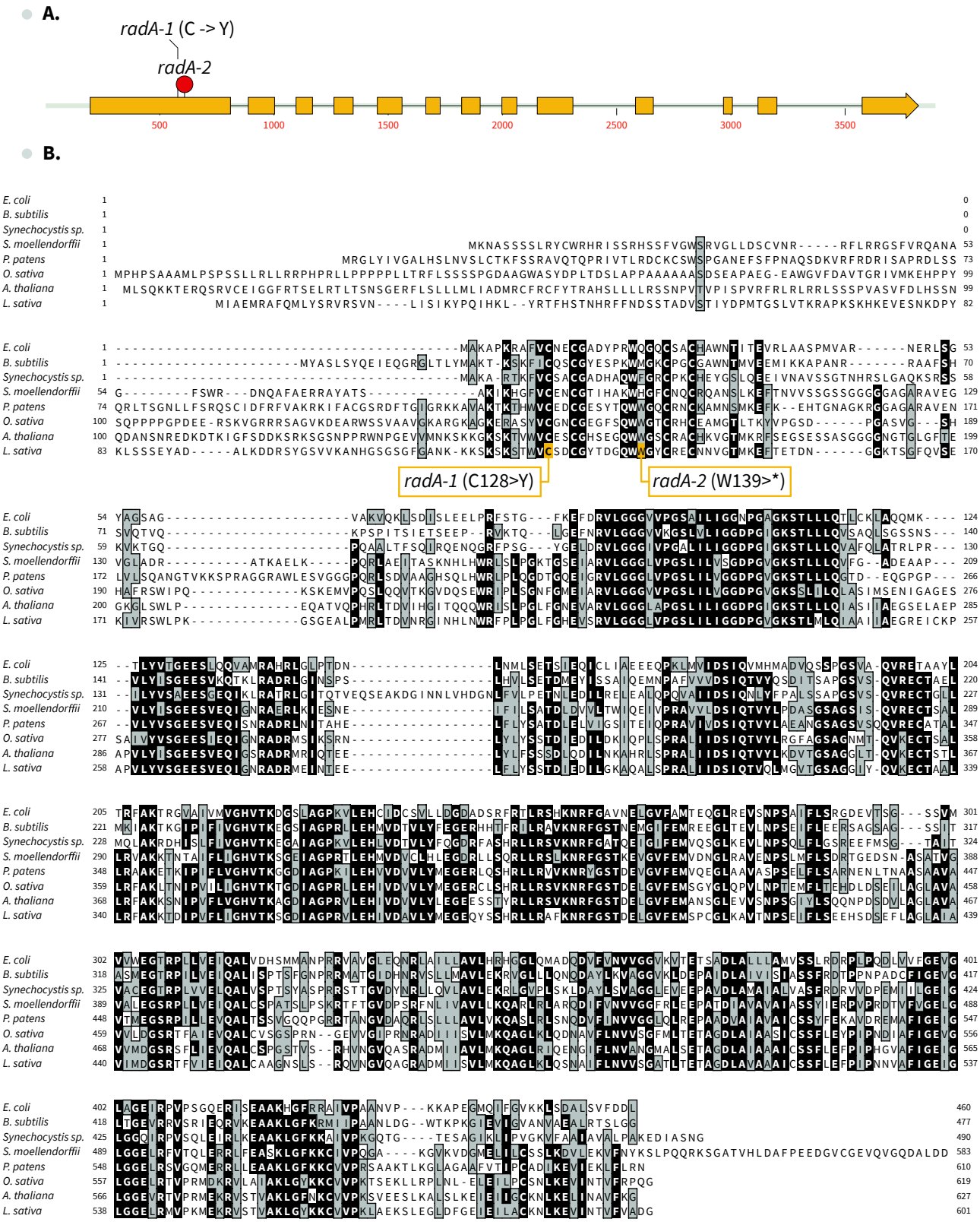


Figure IV.8. Selected *L.sativa* mutants for RADA. (**A**) *RADA* gene structure and localization of the selected mutations. The amino acid change is represented by a black line and the nonsense mutation is represented by a red circle. (**B**) Sequence alignment (ClustalW) between representative land plant RADA sequences and RadA from proteobacteria and cyanobacteria. The localization of the selected mutations are indicated.

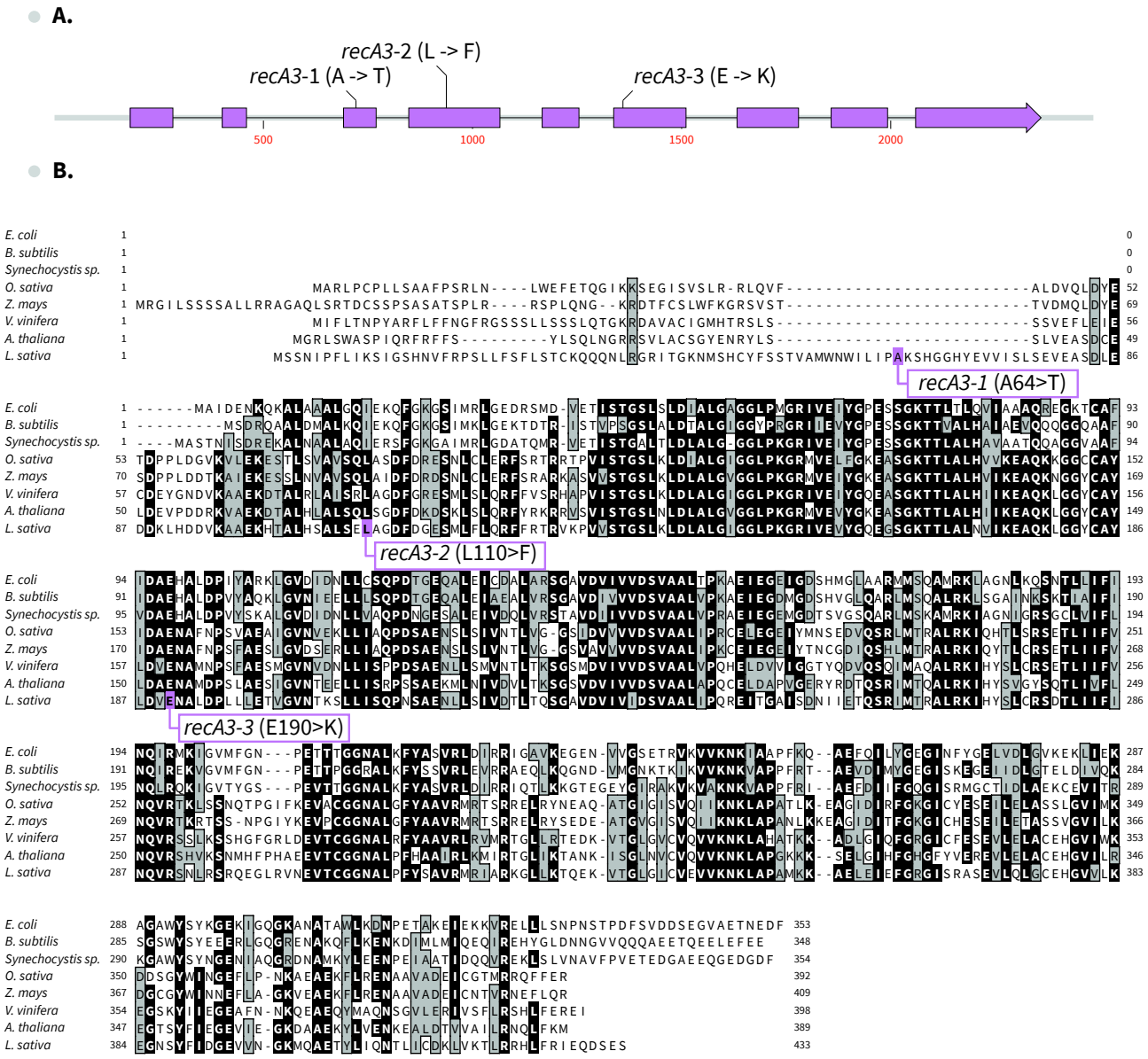
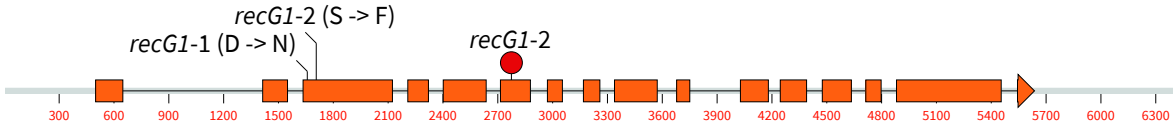
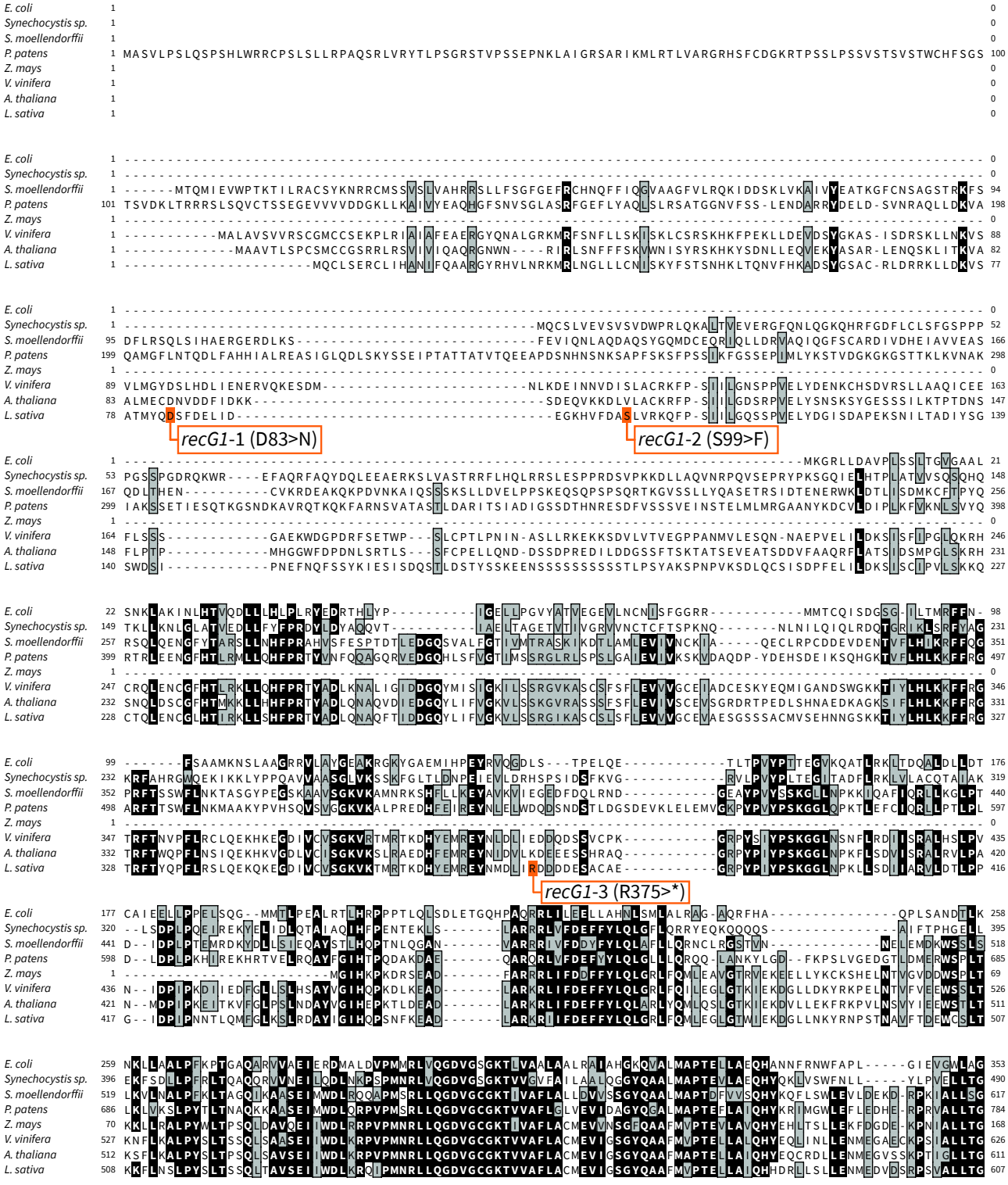


Figure IV.9. Selected *L.sativa* mutants for RECA3. (**A**) Structure of the *L. sativa RECA3* gene and localization of the selected mutations. Amino acid changes are represented by black lines. (**B**) Sequence alignment (ClustalW) between representative land plant RECA3 sequences and RecA from proteobacteria and cyanobacteria. The localization of the selected mutations are indicated.

A.


recG1-2 (S -> F)
recG1-1 (D -> N)
recG1-2
300
600
900
1200
1500
1800
2100
2400
2700
3000
3300
3600
3900
4200
4500
4800
5100
5400
5700
6000
6300

B.


E. coli
Synechocystis sp.
S. moellendorffii
P. patens
Z. mays
V. vinifera
A. thaliana
L. sativa
recG1-1 (D83>N)
recG1-2 (S99>F)
recG1-3 (R375>*)

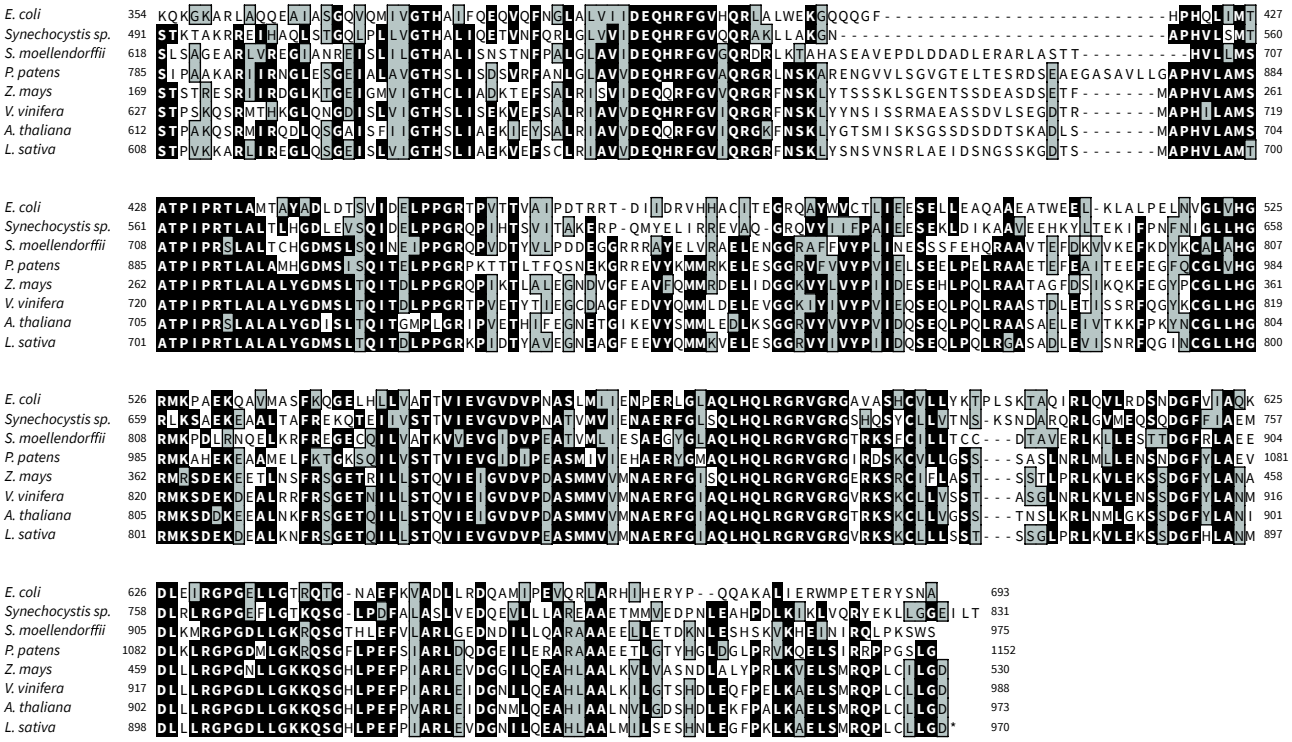


Figure IV.10. Selected *L.sativa* mutants for RECG1. (**A**) Structure of the *L. sativa RECG1* gene and localization of the selected mutations. Amino acid changes are represented by black lines. (**B**) Sequence alignment (ClustalW) between representative land plant RECG1 sequences and RecG from proteobacteria and cyanobacteria. The localization of the selected mutations are indicated.

Phenotype	Number of plants	Remarks
Total plants	1 446	-
Phenotyped plants	1 272	Dead plants because of the genotype or viral infection
Flowering plants	1 231	
=> Non-pollen producing plants	6	
=> Limited pollen production	58	
Pollen producing plants	1 225	
=> Not determined	129	Early selfing seeds production, or presence of both viable and non-viable pollen
=> Non-viable pollen	75	
Viable pollen	1 021	
=> Not determined	42	Plants not conserved
=> No selfing seeds production	116	Male sterility
=> Intermediate phenotype	67	Reduced fertility
Selfing seeds production	997	
=> No crossing seeds production	5	Female sterility
Crossing seeds production	178	Among the Male fertile and fertility reduced plants

Table IV.5. The global phenotyping data are reported in this table according to the different phenotypes tested.

For *MSH1* we obtained 7 mutant lines. These include a nonsense mutation for the glutamine at position 158 (Q158>STOP) in the *msh1*-3 line and six missense mutations (**Figure IV.5**). However, these mutants do not have the same potential. Indeed, the STOP codon in *msh1*-3 is in the first third of the protein sequence and consequently should have a strong impact on MSH1 activity, while the changed amino acids in the other *msh1* lines are not all conserved across evolution (**Figure IV.5.B)**. Thus, the proline that is changed in *msh1*-6 (P742>S) is highly conserved, from bacterial MUTS to plant MSH1, the arginine mutated in *msh1-5* (R168>H) is also shared by bacterial MUTS and the MSH1 of land plants, but not by Diatoms MSH1 (**Figure IV.5.B)**. On the other hand, the positions affected in *msh1*- 1 (D141>N), *msh1*-2 (E155>K) and *msh1*-5 (P583>S), are only shared by land plants, including the bryophyte *Physcomitrella patens,* while the position affected in mutant *msh1*-6 (E742>K) is only found in *L. sativa* and the oceanic diatom *Thalassiosira oceanica*.

We got three ***osb1*** mutant lines, including one (*osb1*-1) with a possible splicing deficiency in the third intron and two lines with nonsense mutations, *osb1-2* (W204>STOP) and *osb1*-3 (W214>STOP) (**Figure IV.6**). These nonsense mutations are interesting because they are located before the PDF motif, known to be involved in ssDNA binding. Consequently the truncated OSB1 in *osb1-2* and in *osb1-3* should not be able to bind ssDNA (**Figure IV.6.B**). However, the splicing deficiency in *osb1-1* might only have a minor effect on the protein activity, because the 90 nucleotides long intronic sequence does not induce a frameshift.

Two ***osb2*** mutants were found in the EMS treated population. These two mutants both result from nonsense mutation of tryptophan codons, at position 140 for *osb2*-1 (W140>STOP) and 389 for *osb2*-2 (W389>STOP) (**Figure IV.7**). The mutant protein encoded by *osb2*-2 is only truncated of few nucleotides (23), while the mutant protein in *osb2*-1 has lost its two PDF motifs (**Figure IV.7.B**).

We obtained two interesting mutants for ***RADA***, both affecting its N-terminal zinc finger domain, reported to be involved in DNA-binding (Cooper and Lovett, 2016). The mutant *radA- 1* is a missense mutation of cysteine-139 into a tyrosine (C128>Y), and mutant *radA*-2 should be a knockout mutant, because of a nonsense mutation (W139>STOP) (**Figure IV.8**).

We did not succeed to get any ***recA3*** nonsense mutant. However, we obtained three missense mutant lines for this gene, *recA3-1* (A64>T), *recA3-2* (L110>F) and *recA3-3* (E190>K) (**Figure IV.9**). The position of the mutation in *recA3-1* is within an exon only present in *L. sativa,* so

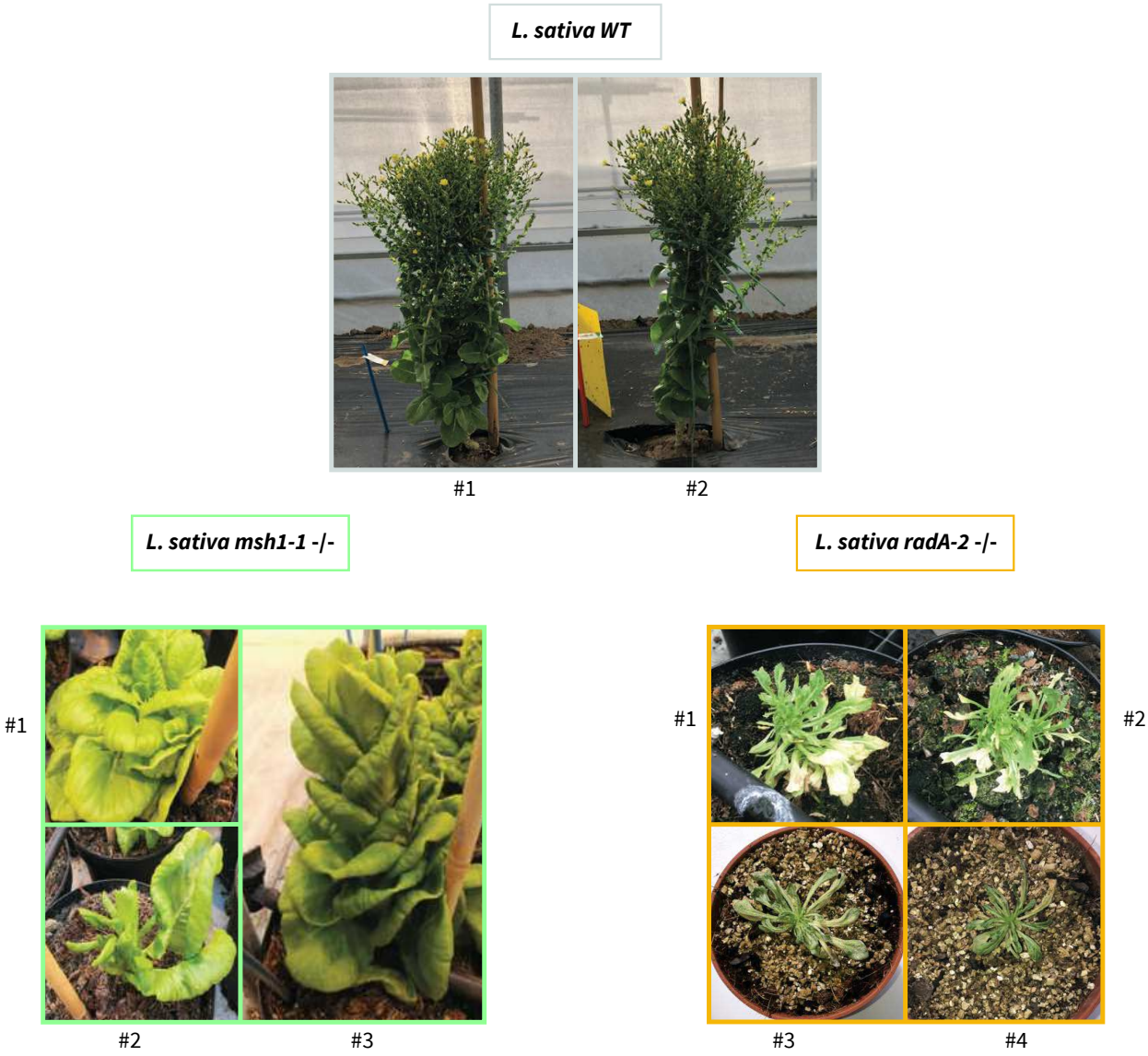


Figure IV.11. General appearance of the *L. sativa* mutant lines obtained by EMS mutagenesis. Only a few mutant lines showed a uniform phenotype. The *Msh1-1* homozygous mutant (-/-) shows a growth retarded phenotype and was not able to produce flowers, and the *radA-2* homozygous mutant (-/-) shows a severe phenotype which makes them unable to develop after the rosette stage.

we could not expect an important impact on the protein function. However, the positions of the mutations in *recA3*-2 and *recA3*-3 are conserved, across plants for leucine-110 of *recA3*-2 and across all empires for lysine 190 in *recA3*-3 (**Figure IV.9.B**). The latter is a position annotated as involved in the binding ATP, in the RecA of *E. coli*.

Finally, we obtained three ***recG1*** mutants, *recG1-1* (D83>N) affected at a position only shared by dicots, *recG1*-2 (S99>F) affected at a non-conserved position, and more interesting *recG1*- 3 (R375>STOP), coding a protein truncated of its main conserved regions (**Figure IV.10**).

For most of these mutants the homozygous (-/-) have been segregated. The (-/-) still have to be segregated for *msh1*-6, *radA*-1 and *recA3*-1. We were not able to segregate (-/-) mutants from *msh1*-7 and *recG1*-2, because of weaknesses of the single heterozygous (+/-) plants of both mutants that were not able to bolt and produce seeds. Moreover, the (-/-) mutant for *osb2*-1 was impossible to segregate despite extensive investigations (further discussed in **CHAPTER 6**).

At the beginning of our project, the strategy to reach a male sterility trait consisted in using these mutants to induce mtDNA diversity, but also to combine them in double or triple mutants. Consequently we adapted our crossing scheme to get these combinations. However, because the phenotypes observed in certain single mutant lines were highly promising (discussed below in **section 1.2.2**) we studied them in priority. Therefore we did not continue the combination strategy and during my PhD we only obtained two double (-/-) mutants, *osb1*-2 *radA2* and *osb1*-3 *radA-2*. We have also (-/-) (+/-) lines: *recA3-3* *osb1-3* and the reciprocal *osb1-3* *recA3-3*, *osb1-2* *osb2-1**, *radA-2* *osb2-1**, *osb1-2* *osb2-2* and the reciprocal *osb2-2* *osb1-2* (* *osb2-1* cannot be segregated at the (-/-) stage). Several double, triple and quadruple heterozygous mutants have also been produced.

1.2.2. Phenotyping of mutant lines

Two reasons complexified the phenotyping in our project. The first one was related to the EMS mutagenesis, which is a random mutagenesis strategy. Consequently, our mutants are carrying other mutations than the ones affecting GIMMS. To overtake this difficulty, unwanted mutations need to be cleaned up by outcrossing with the wild-type (WT). The second issue, inherent to the indirect strategy of mtDNA diversity induction, complexifies the phenotyping. Indeed, with such a strategy we promote mtDNA recombination, but we cannot control the regions of the mtDNA that are potentially affected. As shown in *A. thaliana*, mutants for the same GIMMS do not necessarily have the same impact on the mtDNA and consequently not the same phenotype. That is why after

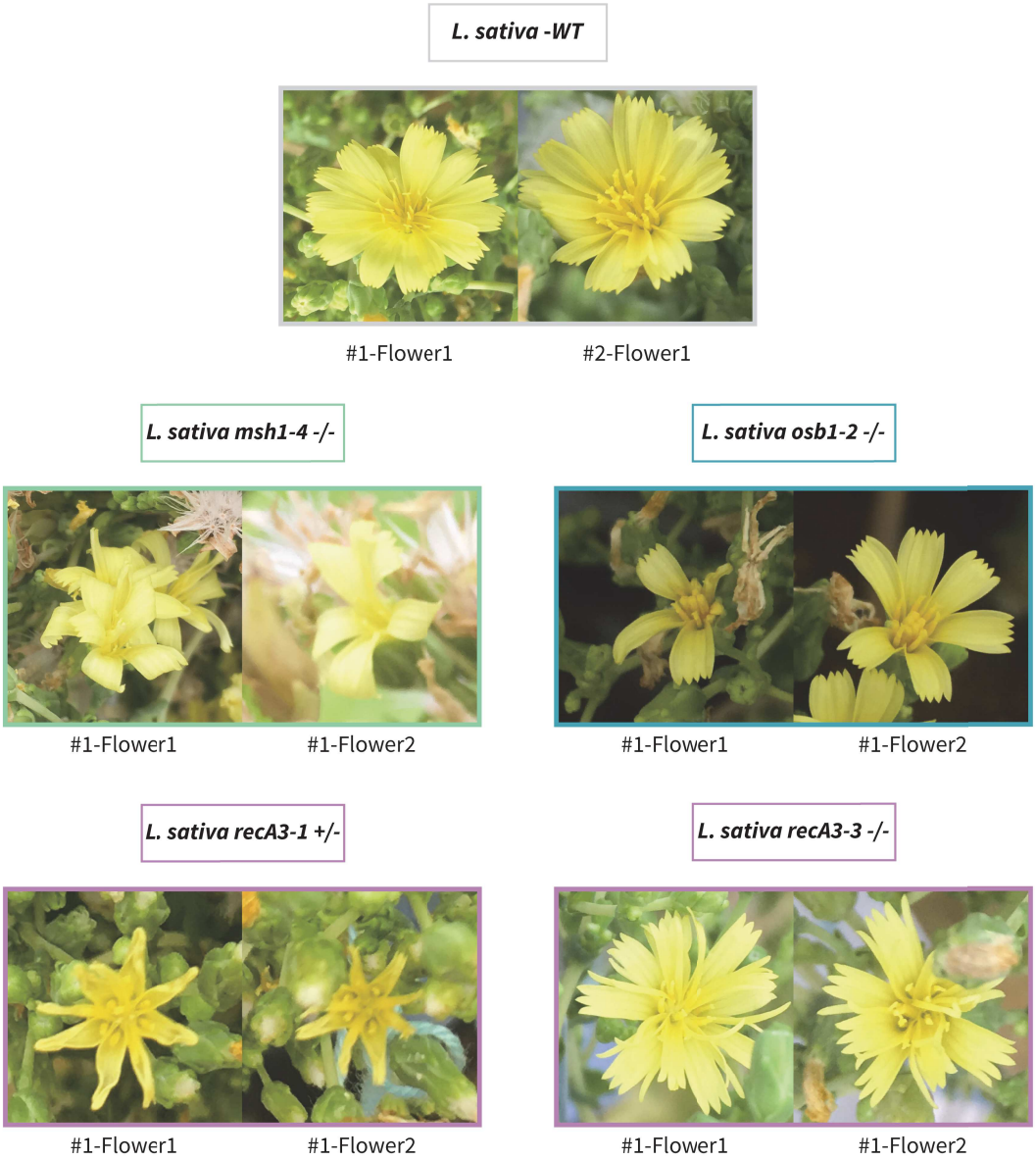


Figure IV.12. Phenotypes of flowers in *L. sativa* mutant lines obtained by EMS mutagenesis. In our mutant population, individual plants showed flower phenotypes. An *msh1-4* mutant shows distorted petals. The same phenotype in observed in a heterozygous *recA3-1* mutant. A few *osb1-2* mutants show a reduced number of open petals and a homozygous *recA3-3* mutant displays a mix of small and WT-like petals.

we had inspected the general appearance of the plant and the phenotype of the leaves, the phenotyping of the twenty mutant lines was focused on flowers and the male sterility trait, combined with female fertility. To do so, we phenotyped the flowers according to the three different types of male sterility. Possible homeotic sterility was checked by looking at the presence or absence of male organs within the flowers. Then we controlled pollen production, with respect to potential sporophytic male-sterility. And finally, we tested for gametophytic male-sterility, by Alexander staining for its ease-of-use. The candidate male-sterile phenotypes were next confirmed and validated by an absence of seeds by self-pollination, and the potential female-sterile phenotype was excluded by seeds production following outcrossing with a WT father or backcrossing with a male-fertile brother line.

The three years' duration of our project allowed us to phenotype plants during five different screens. On these screens, three took place during summer cycles in Allonnes, in greenhouse tunnels, and two during winter cycles in Enkhuizen, on heated greenhouses. The different growth conditions in these two environments could consequently affect the fertility of our mutant populations. The fifth screen will be discussed in **CHAPTER 5**. On the first four screens, our mutant population was composed of 1446 plants on four different generations, including single mutants (het and (-/-) lines), double mutants (HOM, (+/-) and HOM/het) triple mutants (HOM/het/het and het/het/het) and a quadruple heterozygous mutant (**Table IV.5**).

Among these mutant lines we were not able to phenotype 174 plants. Indeed, *Tomato spotted wilt virus* (*TSWV*) infection led to the death of 147 plants. Amidst the remaining 27 plants, 18 carried a (-/-) mutation: 1 *msh1-1* (on 4 (-/-) plant), 4 *osb1-2* (on 68 (-/-) plants), 8 *radA-2* (on 37 (-/-) plants), 2 *recA3-3* (on 32 (-/-) plants) and 2 *recG1-1* (on 11 (-/-) plants). The remaining 9 (+/-) include 4 *osb2-1* plants (on 601 (+/-) plants) (**Table IV.5**).

In the different mutant populations, we did not see a constant uniform phenotype shared by the whole population at the very exception of the *msh1*-1, *msh1*-5 and *radA-2* mutants (**Figure IV.11**). We phenotyped 3 (-/-) and 2 (+/-) *msh1*-1 mutant lines that were not able to properly bolt and develop flowers.

For *msh1*-5, we phenotyped 6 (-/-) and 4 (+/-) mutant lines. All besides one (+/-) were late flowering lines. Consequently, we were not able to include them in the program and be able to cross them. However, they produced a very limited number of seeds by self-cross, at the exception of the

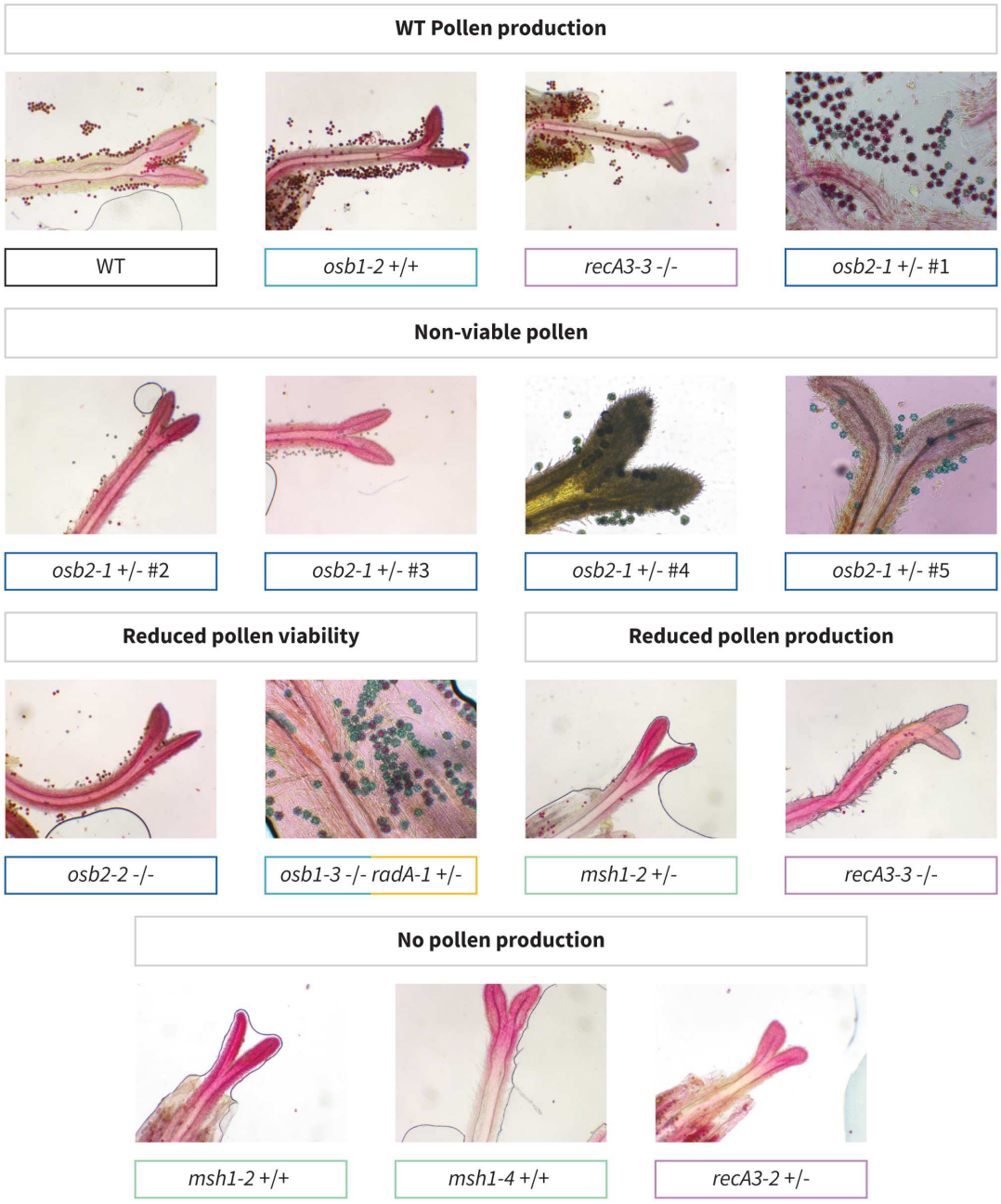


Figure IV.13. Pollen production and viability in *L. sativa* mutant lines obtained by EMS mutagenesis. Coloration with Alexander staining (Alexander, 1969). The Alexander staining colors viable cytoplasms in purple and the the pollen wall in green. Consequently, viable pollen is colored in purple while non-viable one remains blue. The majority of the mutants had a WT-like production of pollen. But within the different mutant populations, individual plants produced non-viable pollen, less viable pollen, less pollen or no pollen at all.

Mutant	Evaluated plants	Male-sterile plants	%	Generation	Remarks
msh1-1	0				Sterility cannot be evaluated
msh1-2	4	4	100	1/4	2 heterozygous & 2 homozygous
msh1-3	17	8	47,1	2/4	+ 4 reduced sterility
msh1-4	1	1	100	1/4	
msh1-5	3	3	100	1/4	Late flowering
*msh1-6**	4	4	100	2/4	Heterozygous
msh1-7	0				Sterility cannot be evaluated
osb1-1	6	0	0		Fertility reduction osberved in 4 heterozygous
osb1-2	11	4	36,4	1/4	Heterozygous all fertile
osb1-3	34	4	11,8	1/4	+ 2 heterozygous sterile
*osb2-1**	265	35	13,2	4/4	Heterozygous
osb2-2	20	7	35	2/4	Heterozygous all fertile
*radA-1**	2	0	0		Heterozygous
radA-2	74	3	4,1	2/4	Homozygous cannot be evaluated,data from the heterozygous
*recA3-1**	1	0	0		Heterozygous
recA3-2	8	1	12,5	1/4	
recA3-3	28	4	14,3	2/4	+ 5/11 heterozygous sterile
recG1-1	7	4	57,1	1/4	+ 1 reduced sterility
*recG1-2**	2	0	0		Heterozygous
recG1-3	10	4	40	1/4	+ 2 reduced sterility

Table IV.6. Male sterility proportion among *L. sativa* mutant populations.

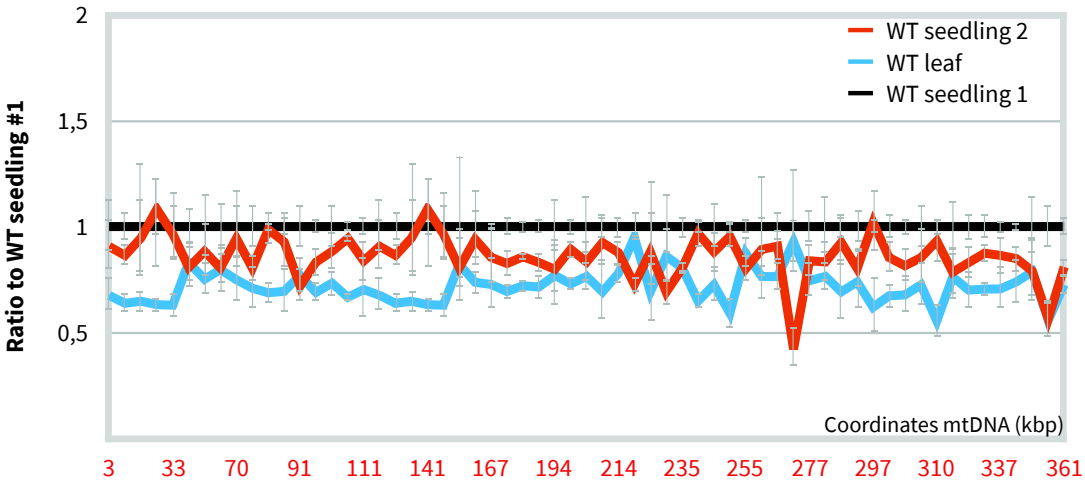


Figure IV.14. mtDNA stoichiometry in different parts of WT lettuce plants. The mtDNA from leaves and seedlings of WT plants were scanned for changes in the stoichiometry of different mtDNA regions. Sequences spaced 5-8 kb apart on the mtDNA were quantified by qPCR. Results were normalized against the nuclear 18S rRNA sequence and are the mean and SD values from three biological replicates. Coordinates correspond to the mtDNA of *L. sativa* var. *capitata* L. *nidus tenerrima,* determined in **CHAPTER 1**.

not late-flowering (+/-) line which produced a WT amount of seeds (>2000). Indeed, the other (+/-) produced between 50-160 seeds while half of the (-/-) were not able to produce any seeds and the three remaining 2, 3 and 3 seeds, respectively. This suggested that the WT-like (+/-) plant could have been wrongly genotyped and should be re-tested, which has not been done for practical reasons.

The study of the *radA*-2 lines involved 29 homozygous mutants segregating from a heterozygous, and all displayed severe phenotypes of retarded growth of both leaves and of distorted leaves (**Figure IV.11**). None of the 300 (+/-) plants carrying the *radA-2* mutation that were segregated did not show any phenotype. We were not able to propagate the *radA*-2 (-/-) because the mutants did not develop after the rosette stage, so it was not possible to phenotype their flowers. Consequently, we segregated them from a (+/-. The visible and molecular phenotypes of (-/-) *radA*- 2 is further discussed below (**CHAPTER 4**). Moreover, 10 additional mutant lines were not able to produce flowers, despite a WT-like leaf phenotype. This was observed for 1 out of 64 *osb1-2* plants, 1 out of 30 *osb2-2* plants, 1 (+/-) *osb2-1* out of 597 *and* 7 (+/-) *radA-2* out of 674 (**Table IV.5**). These could be because of additional EMS mutations affecting other genomic loci, but at present we cannot exclude an indirect effect because of organellar genomic changes, which were not tested.

We did not find homeotic sterility, but some mutants showed changes in their flower phenotypes (**Figure IV.12**). The flowers of the single homozygous *msh1*-4 mutant were distorted, with folded petals. The flowers' phenotype of three out of the sixty-three *osb1*-2 plants was different, because half of the petals remained closed. The flowers of two different *recA3* mutants were also affected. The petals of a (+/-) *recA3-1* mutant were curved-shaped, but this phenotype was not observed on the 12 other heterozygous plants and the two homozygous plants phenotyped. The second line affected was a *recA3*-3 mutant whose flowers displayed a higher number of small petals, but this phenotype was not shared by the 30 others (-/-) *recA3*-3 plants phenotyped. But overall the penetrance of these phenotypes was low, because all cases described above concerned a single plant, or a very small proportion of the population.

The phenotype of the mutant population was next focused on pollen production with respect to the sporophytic sterility. However, pollen production was not a binary phenotype. Indeed, we did find plants that do not produce pollen (the 4 *msh1-2* lines (2 (+/-) and 2 (-/-), the *msh1-4* mutant line and two of the *osb2-1* (+/-) mutants), but the main phenotype observed was a reduction in the quantity of pollen produced (**Table IV.5** & **Figure IV.13**). Such a phenotype was

observed in 58 plants, including 2 of the 6 *osb1-1,* 6 of the 46 *osb1-3*, 3 of the 29 *osb2-2*, 3 of the 30 *recA3-3*, and 26 of 596 *osb2-1* (+/-) mutants.

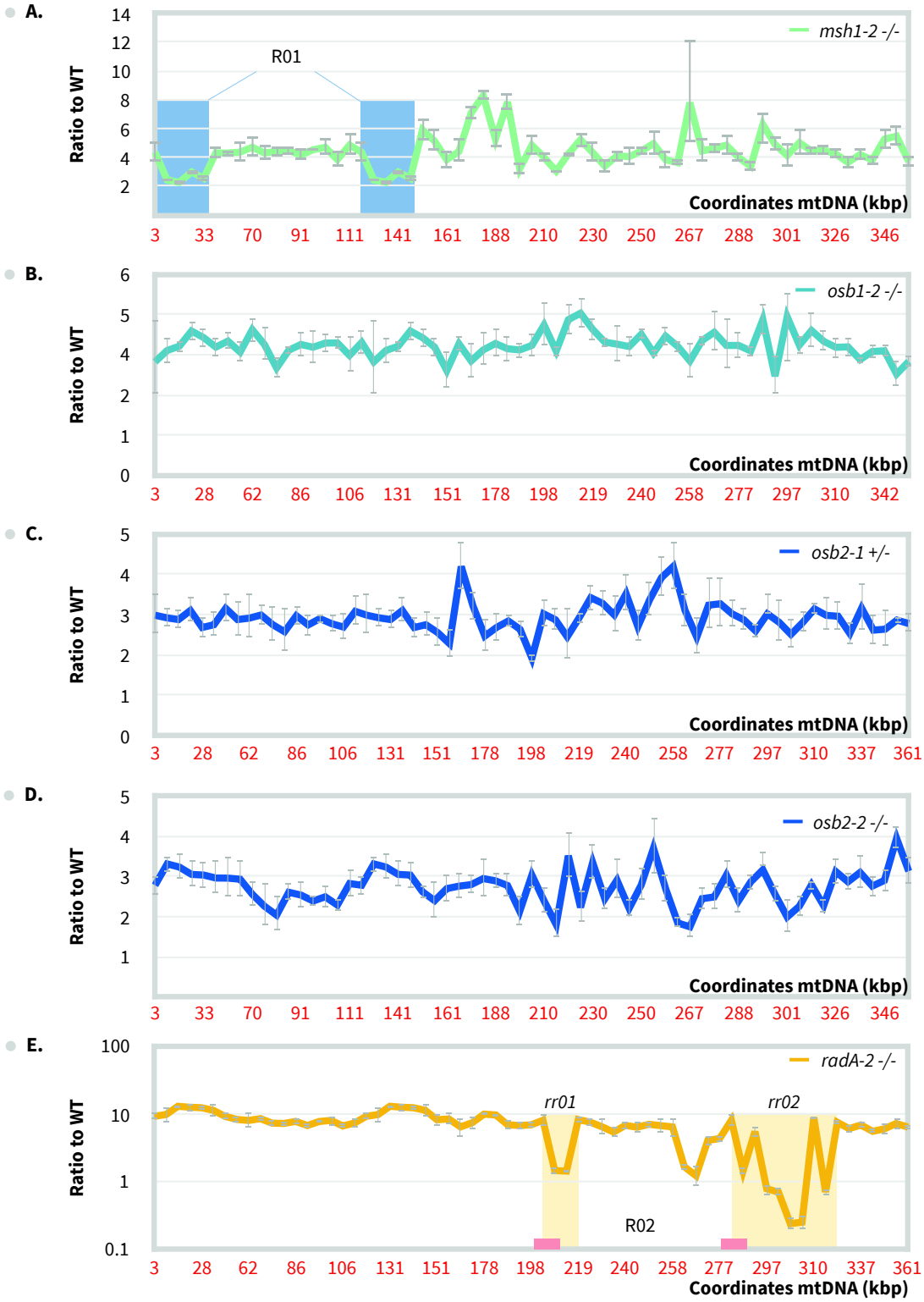
The viability of the pollen produced was next checked with respect to gametophytic sterility. To do so, the structure of the lettuce pollen grain oriented our strategy to determine its viability by Alexander staining. With this method, we identified 75 mutant lines with a defect in pollen viability (**Table IV.5** & **Figure IV.13**). These 75 lines included 7 of the 30 *msh1-3* lines (-/-), 3 of the 41 *osb1-3* lines (-/-), 1 of the 9 *recG1-1* lines (-/-), the 2 *osb2-2* mutant lines (-/-) with low pollen production and 67 *osb2-1* (+/-), among which 20 were lines with reduced pollen production.

The final item checked regarding male sterility phenotypes was the ability to produce seeds by selfing. In our flowering population (1231 plants), we found a total of 116 almost sterile plants (seeds produced < 40 seeds) and 67 plants with reduced fertility (seeds production comprised between 40 and 1000). In the single mutants, the higher proportions were found in *msh1* and *osb's* mutants (**Table IV.6**).

To determine whether the origin of the phenotype of male sterility was cytoplasmic, we aimed to cross the male-sterile plants with a potential maintainer line. However, we cannot predict the maintainer potential for a specific male sterile line before having the segregation results of the cross with this male sterile line. But we have sister plants segregating from the same mother that can mimic these maintainer lines through backcross. By doing this we found that the *msh1-2* mutants were not able to produce any seeds in following backcrosses, and so we considered them as also female sterile. For most of the other sterile plants, these backcrosses led to the loss of sterility, except for the *osb2-1* family. Consequently, because of time constraints the study of the transmission of the male-sterile phenotypes was postponed, for most of the mutant lines, and we decided to focus on this *osb2-1* family which is further discussed in **CHAPTER 5**.

1.2.3. Impact of the mutation on the mitochondrial genome

In *A. thaliana*, mutants of *MSH1*, *OSB1*, *OSB4*, *RADA*, *RECA3* or *RECG1* have been described as showing an increase in mtDNA ectopic recombination involving small repeated sequences. This because the ectopic recombination induced by these mutations led to sublimons or subgenomic molecules which replicate at a different rate than the other regions of the mtDNA. One way to evaluate these changes is to quantify the relative copy number of the different parts of the mtDNA, by qPCR. In *A. thaliana* this was done using a set of primer pairs spaced every ~5 kb (Wallet et al.,


A.
Ratio to WT
msh1-2 -/-
R01
Coordinates mtDNA (kbp)
B.
osb1-2 -/-
C.
osb2-1 +/-
D.
osb2-2 -/-
E.
radA-2 -/-
rr01
rr02
R02

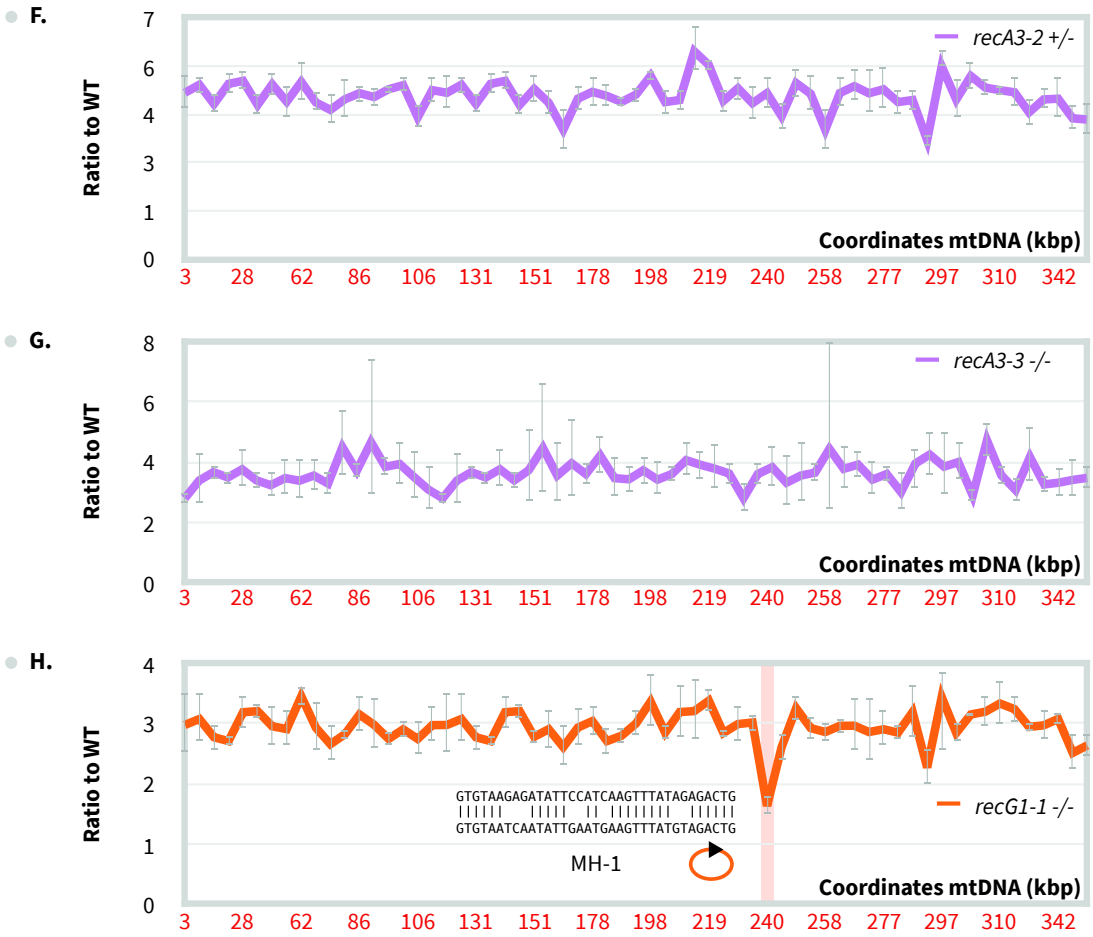


Figure IV.15. mtDNA stoichiometry in selected lettuce mutants. (**A-H**) The mtDNAs of different individual mutants were scanned for changes in their stoichiometry, as compared to the WT reference. Sequences spaced 5-8 kb apart on the mtDNA were quantified by qPCR as described in **Figure IV.14**. (**A**) Two regions with a reduced relative copy number in the mtDNA of the *msh1-2* mutant, corresponding to the large repeated sequence R01 are blue shaded. (**E**) The radA-2 mutant shows two regions (yellow shaded) with a strong reduction in relative copy numbers (**H**) The *recG1-1* mutant shows a small region (orange shaded) with a lower mtDNA stoichiometry. On both sides of the sequence, a direct microhomology can be found (the sequences are shown). Results are on a logarithmic scale. Coordinates correspond to the mtDNA of *L. sativa* var. *capitata* L. *nidus tenerrima* determined in **CHAPTER 1**.

2015; Le Ret et al., 2018; Chevigny et al., 2020). The mtDNA of *L. sativa* is 364 324 bp long, but due to repeated sequences only 313 875 bp are unique. Consequently, to scan by qPCR the mtDNA of *L. sativa,* for possible changes in the relative stoichiometry of the different mtDNA regions, we designed 60 pairs of primers, a pair of primers spaced every ~5.2 kb on the genome.

In a WT context and in a same tissue, the stoichiometry of the different parts of the mtDNA does not show significant changes, between two plants. However, this is not necessarily the case when different tissues are compared. Indeed, this method relies on the normalization to the nuclear genome. The nuclear content does not vary with respect to the cell tissue, except in plants undergoing endoreplication (*such as A. thaliana*), but the mtDNA does. Consequently, the mtDNA copy number can vary between different tissues. By qPCR, using DNA extracted from different tissues of WT plants, we could confirm that in *L. sativa* the relative stoichiometry of different mtDNA regions remains constant across the genome, in leaves and seedlings, although the genome copy number changes, with 1.14 less copies of the mtDNA in leaves than in seedlings, in relation to a haploid nuclear genome (**Figure IV. 14**).

We next evaluated the mtDNA stoichiometry in mutants *msh1-2* (-/-), *osb1-2* (-/-), *osb2-1* (+/-), *osb2-2* (-/-), *radA-2* (-/-), *recA3-1* (+/-), *recA3-2* (-/-) and *recG1-1* (-/-) (**Figure IV. 15**). For all mutant plants analyzed we found a higher mtDNA copy number (3 to 10-fold) as compared to the WT. Such a difference can be explained by the mtDNA content which varies according to the tissues and age. Indeed, the WT DNA used as reference were extracted from seedlings, while DNA of *msh1- 2*, *osb1-2*, *osb2-1*, *osb2-2*, *recA3-1*, *recA3-2* and *recG1-1* were from floral buds and DNA from *radA-2* was from a seedling. Despite the higher mtDNA copy numbers, *osb1-2*, *osb2-1*, *osb2-2, recA3-1* and *recA3-2* did not show any significant changes in the stoichiometry of the different regions of the mtDNA. For *recG1-1* we found a region with a lower stoichiometry, located between 235 and 245 kb on the genome. This region carries the two copies of an imperfectly repeated microhomology of 32-40 bp, in direct orientation (**Figure IV. 15.H**). This microhomology might recombine and lead to a small sublimon with a lower replication rate, explaining its lower stoichiometry. But this hypothesis was not tested and lacks experimental evidence.

More important changes in mtDNA sequences stoichiometry was observed in *msh1-2* mutant plants. Indeed, this mutant shows a surprising 2-fold reduction of the large repeated region R01, as compared to neighboring sequences (**Figure IV. 15.A**). Such a change could originate from a different mtDNA organization induced by the *msh1-2* mutation. In *A. thaliana*, mutants for MSH1

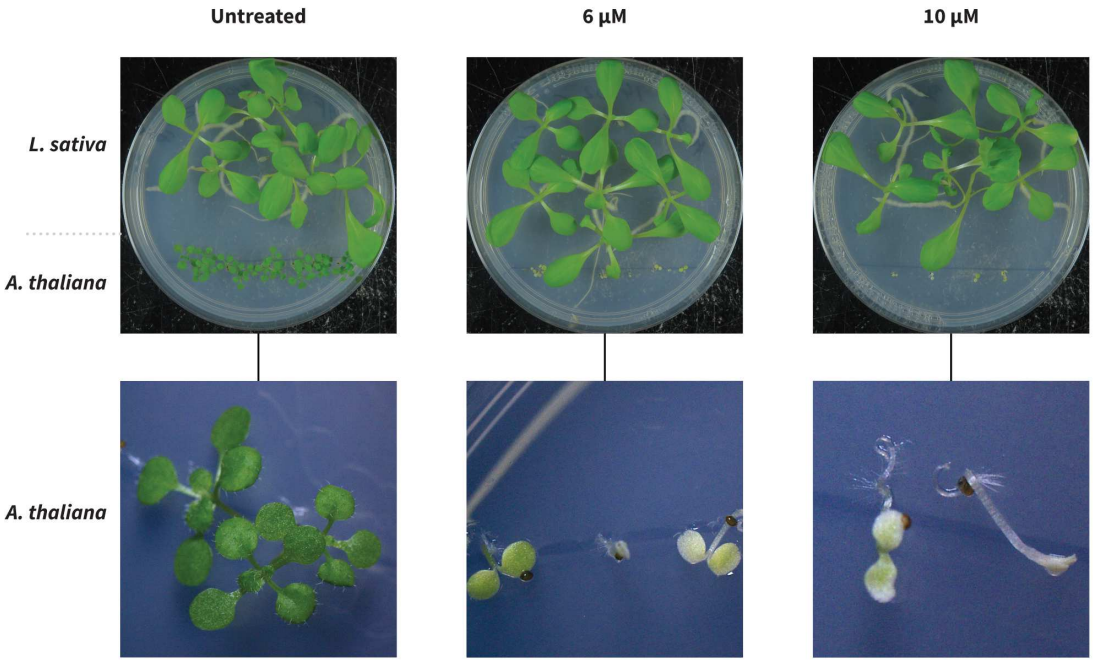


Figure IV.16. Growth of lettuce and Arabidopsis in the presence of ciprofloxacin. Plants of *L. sativa* and of *A. thaliana* (upper panels) were germinated and grown for 14 days in ciprofloxacin-containing agar plates. *A. thaliana* seedlings were observed with a x20 by stereomicroscope (bottom panels).

are known for unlocked mtDNA recombination. In the *L. sativa msh1-2* mutant unrestricted recombination might have led to a new configuration of the mtDNA, involving different ratios of the subgenomes containing R01. But as well as for the *recG1-1* mutant this hypothesis needs to be further investigated experimentally. That could be by long-read sequence analysis of the mtDNA, or by PFGE analysis of undigested mtDNA and of mtDNA digested with rare-cutter restriction enzymes. Because of priority decisions this could not be done during the relative short time of the PhD.

Finally, the biggest changes in mtDNA stoichiometry were observed in the *radA-2* mutant (**Figure IV. 15.E**). This mutant shows dramatic changes in its mtDNA stoichiometry, with important reduction of two large regions, *rr01* between 210 and 219 kb and *rr02* between 283 and 326 kb (rr for reduced region). These two regions are located immediately next to the two copies of the large repeated region R02, suggesting a possible involvement of this repeat in recombination events that led to the reduction in copy number of those regions. The range of these changes, up to ~100-fold reduction compared to the basal level of the mtDNA for *rr02,* indicates that the mitochondrial recombination induced by the *radA-2* mutation leads to their possible loss. Moreover, these regions contain protein-genes; *ccmB, nad2a, nad2b and rpl10* in *rr01,* and *atp4, ccmB, ccmC, nad2a, nad2b, nad4L and rpl10* in *rr02.* Their loss could explain the severe development phenotype of the *radA-2* plants. This hypothesis along with the investigation of the *radA-2* mitotype is further discussed in **CHAPTER 4**.

2. Pilot study: Direct induction of mitochondrial genome diversity with ciprofloxacin

The aim of this pilot study was to determine the suitable conditions to treat *L. sativa* with ciprofloxacin (CIP) to induce mtDNA recombination. Such a treatment could be useful to amplify the effects of mutations in GIMMS, by promoting repair of DSBs by recombination. But it could also be a way to directly induce mtDNA recombination in a WT context.

2.1. Strategies of treatment

In *A. thaliana*, CIP is used as an organellar genotoxic agent to evaluate the impact of a mutation on the organellar genome repair mechanisms. To do so, a range from 0.25 µM up to 0.75 µM of CIP is usually tested, on *A. thaliana* plants grown *in vitro* on MS medium (Cappadocia et

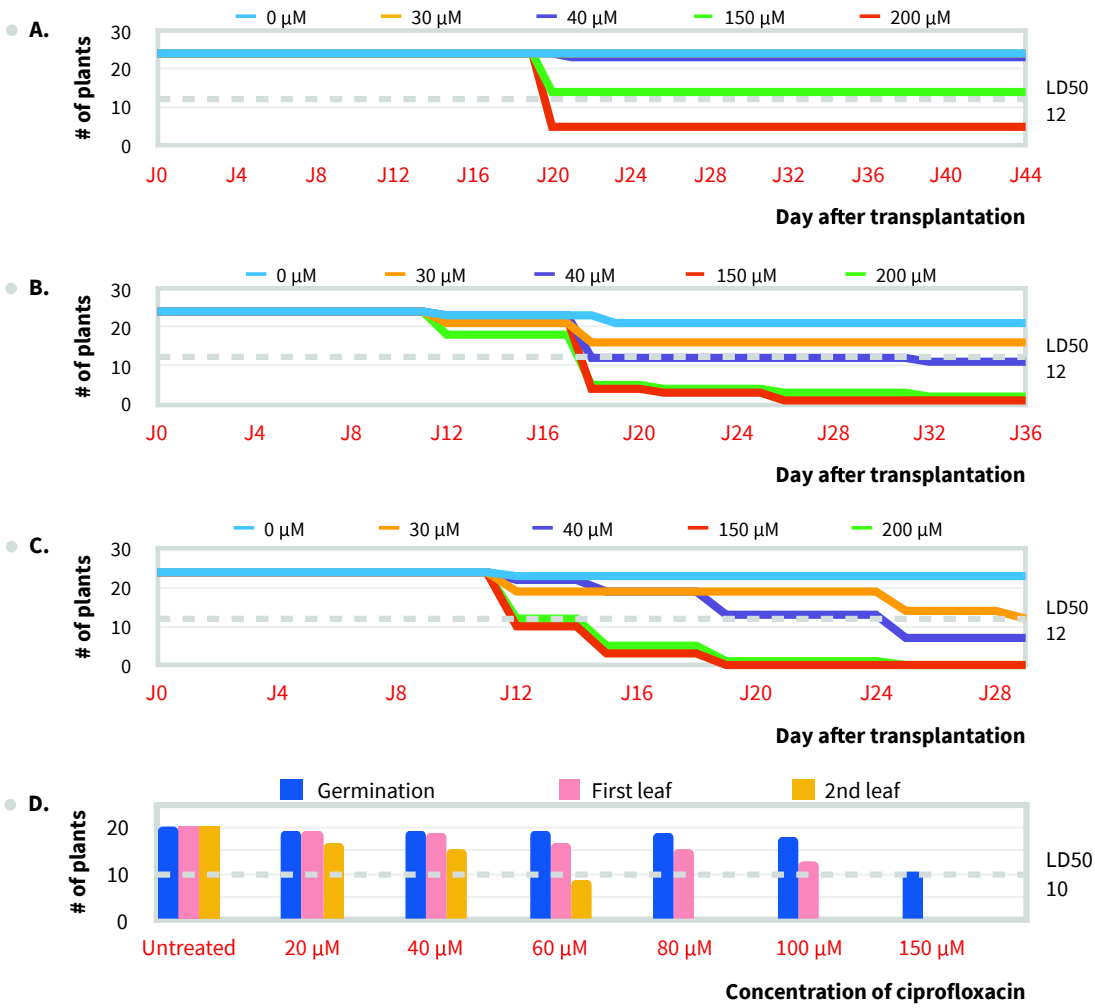


Figure IV.17. Lettuce survival to ciprofloxacin-induced genotoxic stress. (**A-C**) Lettuce plants were germinated and grown in ciprofloxacin-containing agar plates for 7 days (**A**), 14 days (**B**) or 21 days (**C**), and transplanted to soil. The number of dead plants was then followed every day (J0 to J28). (**D**) Lettuce plants were germinated and grown in ciprofloxacin-containing agar and the number of plant able to germinate and to develop first and second leaves were determined after 14 days.

al., 2010). But in all cases, WT and mutant *A. thaliana* plants died for CIP concentration higher than 1 µM.

In lettuce, we developed a treatment strategy that mimics an EMS treatment. Briefly, CIP is added to the seeds-germination medium and the seedlings grown on this media before being transplanted to CIP-free soil. We tested two germination protocols in the presence of CIP. According to the protocol, *L. sativa* seeds are germinated on blotting paper, wetted with ultrapure water containing the desired concentration of antibiotic. Seedlings are grown for 1 to 3 weeks and the medium is hydrated each week with ultrapure water containing CIP. The second protocol was inspired from the treatment of *A. thaliana* seedlings with CIP. It involves the germination and growth of seedlings on agar-plates containing CIP, before transplanting into soil. In both cases, we found that *L. sativa* had a much higher resistance to CIP than *A. thaliana* (**Figure IV.16**). Consequently, we used a wider range of CIP concentrations in our treatments, from 20 µM up to 2 mM.

2.2. Determination of the working concentration of ciprofloxacin in Lettuce

Using both protocols of CIP treatment we determined the dose of ciprofloxacin which kills half of the population (LD_{50}) (**Figure IV.17**). However, we evaluated the plant survival according to different strategies and criteria. For seedling germinated and grown on blotting paper, we evaluated the survival of plants after the transplantation on CIP-free soil, and considered as dead the plants that did not continue to produce leaves. To do so, we tested concentrations ranging between 20 µM and 200 µM CIP, and different treatment lengths from 1 up to 3 weeks. After that, 24 seedlings were transplanted into soil to study their survival. In these conditions the LD_{50} for CIP depended on the duration of the treatment (**Figure IV.17.A-B-C**). For one week treatment, the LD_{50} was between 150 and 200 µM. For treatments of two and three weeks the LD_{50} dropped to 30-40 µM CIPs. For seedling germinated and grown on solid media, we considered as dead the plants that did not produce a first true pair of leaves after 14 days (**Figure IV.17.D**). In this condition the LD_{50} was comprised between 40 and 60 µM CIP, which is in the same range as the one obtained with the other treatment protocol, for plant treated for more than one week with CIP.

In both cases, *L. sativa* confirmed its higher resistance to CIP than *A. thaliana*. However, the two treatment protocols showed a difference regarding germination. Indeed, germination was not possible beyond 200 µM CIP on solid medium, while seeds still germinated with a 2 mM concentration of CIP on blotting paper.

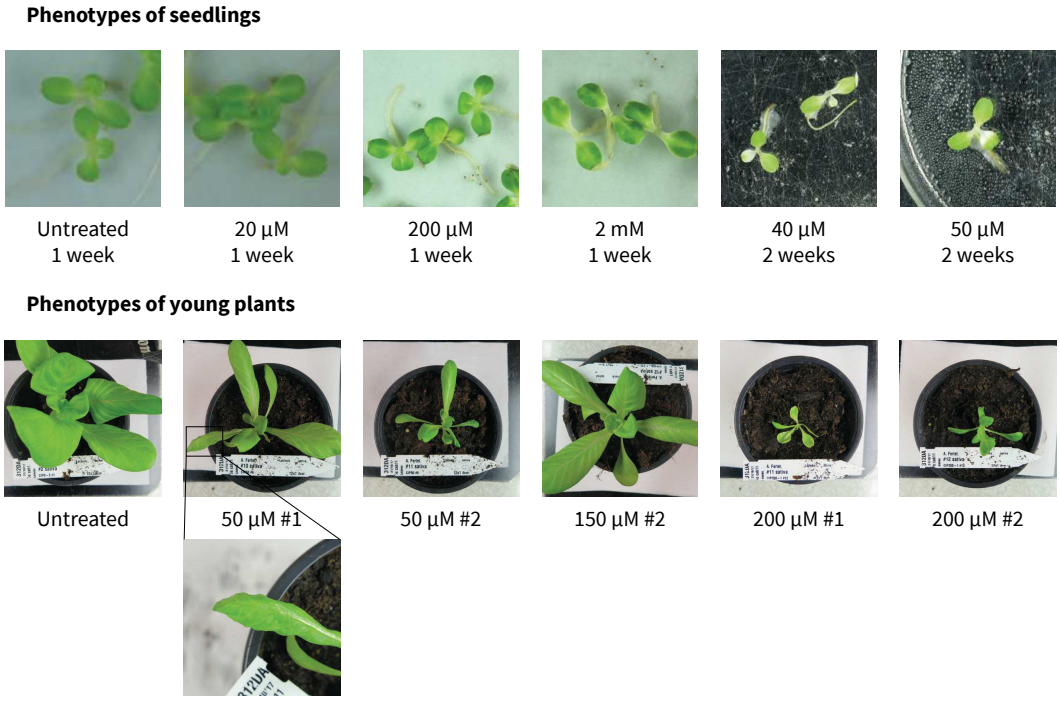


Figure IV.18. Lettuce phenotypes associated with ciprofloxacin treatments. The lettuce seedlings treated with ciprofloxacin showed a chlorosis halo around the cotyledons and a larger first leaf area (top panel). This phenotype was not conserved in young plants, but the young plants showed retarded growth and sectors of distorted leaves (middle panel). All the ciprofloxacin treated plants showed WT-like fertile pollen (bottom panel), with one exception which showed a reduction in pollen viability and in the production of seeds by selfing.

2.3. Phenotypes associated to CIP treatments

The phenotyping of CIP-treated plants followed the same strategy as the phenotyping of the mutants for GIMMS (**Section 1**). The general appearance of the seedlings and of the plant were first evaluated regarding the concentration and duration of the CIP treatment. After transfer to soil we phenotyped the surviving plants with respect to male fertility.

The CIP-treated seedlings showed a chlorosis halo around the cotyledon (**Figure 18.A**). The size of this halo increased with the CIP concentration and the duration of the treatment. The plants surviving the treatment showed different phenotypes, WT-like, reduced growth or distorted leaves (**Figure 18.B**). These distorted leaves looked like the leaves of *radA-2* mutants, severely affected in mtDNA maintenance. But in CIP-treated plants the phenotype was restrained to just a few leaves and leaf sectors.

Regarding fertility, all the plants developed normal flowers producing viable pollen according to Alexander staining, at the very exception of a single plant (**Figure 18.C**). This plant was treated for 4 weeks with 25 µM CIP and did not show any other development phenotype, else than a reduction in pollen viability (~⅔ of viable pollen) linked to a reduced production of seeds by selfing. Indeed, the flowers of a part of the plant produced less than 4 seeds by self-pollination, as compared to ~12 in other parts of the plant. We could restore seed production in this part of the plant by crossing it with WT pollen. However, all the progeny of this plant showed a WT-like fertility phenotype either from seeds obtained by selfing or from crosses.

2.4. Impact of CIP on the stability of *L. sativa* organellar genomes

CIP inhibits the organellar DNA gyrase, located in both chloroplasts and mitochondria. As described above, binding of CIP to the subunit A of gyrase represses the re-ligation of DSBs induced by the normal action of the enzyme. These DSBs can be the source of recombination in organellar genomes, because of the action of repair by recombination pathways. To determine the effects of CIP-treatment on the stability of the organellar genomes we evaluated the relative copy numbers of the mtDNA and cpDNA and the stoichiometry of the different parts of the two genomes, between CIP-treated and untreated *L. sativa* plants (**Figure 19**).

CIP treatment seemed to mainly affect the copy number of the mtDNA (**Figure 19.A&B**). Indeed, in treated seedlings the global mtDNA copy number was lower than in untreated plants,

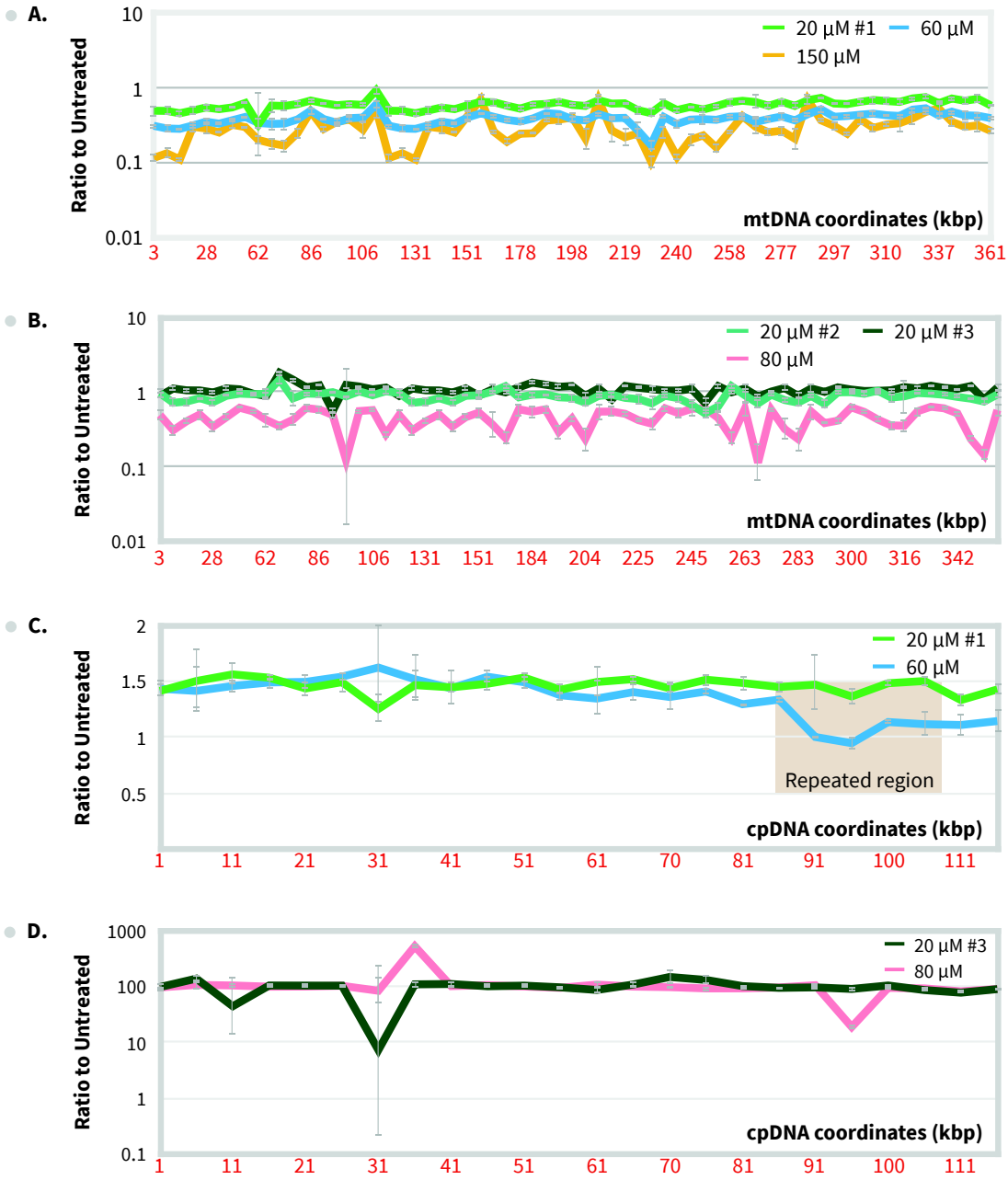


Figure IV.19. Impact of ciprofloxacin treatments on lettuce organellar genomes stoichiometry. (**A and B**) The mtDNA from ciprofloxacin treated seedlings (**A**) or leaves (**B**) were scanned for changes in the stoichiometry of their different regions as compared to the mtDNA of an untreated plant (UP). Sequences spaced 5-8 kb apart on the mtDNA were quantified by qPCR as described in **Figure IV.14**. Coordinates correspond to the mtDNA of *L. sativa* var. *capitata* L. *nidus tenerrima* determined in **CHAPTER 1**. (**C and D**) The same as in **A** and **B**, but for the cpDNA. Coordinates correspond to the published *L. sativa* cpDNA. Results are represented in a logarithmic scale and are the mean and SD of three technical replicates.

from 2-fold in a 20 µM treated plant up to 5-fold in a 150 µM treated plant (**Figure 19.A**). Moreover, at the higher CIP concentrations the mtDNA showed important variations in the stoichiometry between different regions, which could be because of recombination events induced by the CIP treatment. In the leaves of treated plants transferred to soil, for those treated with a low CIP concentration (20 µM) the mtDNA copy number recovered to normal levels, while those treated with higher CIP concentrations retained an affected mtDNA (**Figure 19.B**).

The copy number of the cpDNA was mildly affected by the CIP treatments (**Figure 19.C&D**). However, in the seedlings of plants treated with 60 µM CIP there was a significant reduction of the cpDNA repeated region and of the region comprised between the repeats (small single copy region, or SSC) (**Figure 19.C**). In Arabidopsis, results from our laboratory also showed that CIP treatment results in an important reduction of those regions of the cpDNA (Le Ret et al., 2018). It was speculated that it results from a differential effect of CIP on the progression of the divergently oriented replication forks, emanating from replication initiation sites that are located around the rRNAs genes, in the repeated region (Le Ret et al., 2018).

In leaves of treated plants transferred to soil, the 80 µM CIP treatment resulted in two regions with a variation close to 10-fold as compared to neighboring sequence, which could be because of recombination events affecting the cpDNA. But this remained to be confirmed (**Figure 19.D**). The very high cpDNA copy number indicated in the figure is because the results were compared to the cpDNA in seedlings, while leaves contain a higher number of chloroplasts per cell and consequently a much higher cpDNA copy number.

DISCUSSION AND PERSPECTIVES

In this chapter our goal was to induce mtDNA recombination in *L. sativa* using two different strategies. The first was to indirectly induce mtDNA recombination by using mutant for GIMMS, and the second was a direct strategy involving the organellar-specific genotoxic compound ciprofloxacin.

The pilot studies of direct induction of mtDNA recombination aimed to determine the conditions of CIP treatment more adapted to *L. sativa*. To do so we developed a protocol where seeds are treated with CIP and germinated on CIP-containing germination medium. We selected

two germination media appropriated to such treatment: blotting paper (wetted with ultrapure water containing the desired antibiotic concentration) and agar media (containing CIP). After 7 to 21 days the seedlings were transferred to CIP-free soil. In both germination conditions, *L. sativa* showed a much higher resistance to ciprofloxacin than *A. thaliana,* and the LD_{50} was close to ~50 µM for both germination conditions. These are preliminary results that need to be refined to closely determine the CIP LD_{50} for lettuce. Moreover, the lettuce resistance to CIP deserves to be further investigated. Two hypotheses could explain this resistance, either CIP is inefficiently absorbed and remains in the roots, where the concentration is higher than in the aerial tissues of the plant, or the antibiotic is more efficiently detoxified in lettuce than in other species. For such study, a comparative analysis of the mtDNA stability in roots and leaves from treated seedling could be conducted, as well as the determination of the CIP concentration in roots and leaves. Such results could lead to the development of new treatment protocols, such as to directly apply the CIP solutions to flowers or flower buds.

The treated plants showed development phenotypes and molecular phenotypes of reduced mtDNA copy number and changed mtDNA sequences stoichiometry, which can be linked to mtDNA recombination. However, this was not sufficiently investigated. The phenotypes induced by the CIP on plants were limited to a certain area (like a leaf) of the plant so it might be important to determine if the non-affected leaves show a same mtDNA profile as the affected ones. Moreover, we analyzed only a few CIP-treated plants, and it would be interesting to follow the mtDNA profile in the progeny of an affected plant.

We were not able in this preliminary study to produce a CMS plant, but among the treated plants one showed a reduction in fertility, which was however not transmitted to its progeny. This plant grew on a winter cycle, which for lettuce it is not the best period for pollen production and fertility. However, plants from the same batch did not show such reduction in fertility. We have not analyzed the mtDNA of this plant, and therefore do not know if the reduction in fertility correlated with changes in mtDNA stability. But overall, we found this strategy promising and applied in a wider scale might lead to good mutant candidates. To go further, we should treat a larger plant population with CIP and screen them for male sterility phenotypes.

Following the indirect strategy to induce mtDNA variability, using mutants for GIMMS, our shortlist of candidate genes was established according to previous works in *A. thaliana.* It comprised *MSH1*, *OSB1*, *OSB2*, *POL1B*, *RADA*, *RECA3*, *RECG1* and *WHY2*. We identified their orthologs in *L. sativa*

and determined the subcellular localization for most of the corresponding proteins. The localization remains to be determined for MSH1, POL1B and RECG1. Our attempt to determine the subcellular localization of MSH1 was not successful, the protein localized within the cytoplasm. That could be because our *MSH1*-gene model is incorrect. Indeed, the alignment of MSH1 sequences shows that the predicted *L. sativa* MSH1 lacks a 42 amino-acids N-terminal peptide found in *A. thaliana* and *Z. mays*. Consequently, the gene structure of *MSH1* needs to be refined. Nevertheless, because *MSH1*, *POL1B* and *RECG1* are single copy genes, the subcellular targeting of the corresponding proteins is most probably conserved among flowering plants.

Among the identified mutant lines we selected 20 interesting ones, for *MSH1*, *OSB1*, *OSB2*, *RADA*, *RECA3* and *RECG1*. Among these lines *radA-2* and *osb2-1* were most interesting for the apparent effects of the mutations on mtDNA recombination and male sterility, respectively, and are further discussed in the next chapters.

Our plan was to use these mutants as inducers of mtDNA recombination and consequently to maintain the mutation to increase the probability of a recombination event leading to CMS. However, such CMS plant was not obtained. One of the reasons for such a failure could be that, for a recombined form of mtDNA existing at low copy number to become the predominant mtDNA form involves a process of clonal amplification known as substoichiometric shifting (SSS) (Woloszynska, 2010), which is not well understood. In our plan, SSS should happen in a mutant background after few homozygous generations, by maintaining the mutation. But that was not observed. In previous work from the lab, an *A. thaliana recG1* mutant showed heteroplasmic mtDNA that did not induce a phenotype (Wallet et al., 2015). Several alternative configurations of the mtDNA could coexist, with none becoming predominant over the others. However, the reintroduction of the WT *RECG1* allele by backcross induced the segregation of the alternative mitotypes in the progeny. Such a strategy could be applied to our homozygous mutant lines.

One other hypothetical way to obtain CMS lines from these mutants could be the combination of mutations at the homozygous stage, to inactivate alternative recombination pathways that exist in plant organelles. We put aside this plan to focus on the single mutants. But we already got several double heterozygous mutants and so the double homozygous mutants are waiting to be segregated, phenotyped and mitotyped.

Regarding mtDNA recombination, because we decided to focus on *radA-2* and *osb2-1* we did not test all the mutant lines available, which consequently remain to be further analyzed.

Among the eight mutant lines for each we analyzed the mtDNA stoichiometry, some lines other than *radA-2* showed evidence for changes in mtDNA stability. One that should be interesting to study is *msh1-2,* which shows a 2-fold reduction of the large repeated region R01 (**Figure IV. 15.A**). Our hypothesis is that new mtDNA configuration induced by this mutation involved only one copy of R01. To confirm this hypothesis the solution could be to sequence through the whole repeated region. But the size of R01 (~35 kbp) involves the use of a long-read strategy, such as PacBio or MinION sequencing (Oxford Nanopore Technologies). Both strategies require a large amount of DNA as input, which cannot be obtained from purified mitochondria because this mutant does not produce seeds (male & female sterility). The *msh1-2* did not show other phenotype than sterility, and consequently it is possible to obtain a large amount of total DNA from it, starting back from the initial seed lot. And it is possible to enrich a sample in mtDNA starting from total DNA, by using the NEBNext® Microbiome DNA Enrichment Kit (New England Biolabs, Ipswich, Massachusetts, United States). This kit relies on the lower rate of CpG methylated cytosines in the bacterial genome (and also organellar genomes) than in the nuclear genome. It uses the methylated CpG-specific binding protein MBD2 fused to the Fc fragment of human IgG which binds readily to Protein A linked to magnetic beads. The magnetic field then pulls out the CpG-methylated (nuclear) DNA, leaving the non-CpG-methylated (organellar) DNA in the supernatant.

Another line that could be further analyzed is *recG1-1.* Its study seems more accessible, because of the small size of the mtDNA region that is reduced in copy number, and the ability to propagate the mutant by selfing. For this mutant, PCR and sequence of the molecule resulting from the hypothetical microhomology-mediated recombination could validate our hypothesis. In both cases, it would be interesting to determine the mtDNA configurations that exist in several individual plants of a same mutant, and in several different mutants of a same gene.

CHAPTER 5: Study of the *radA-2* mutant line

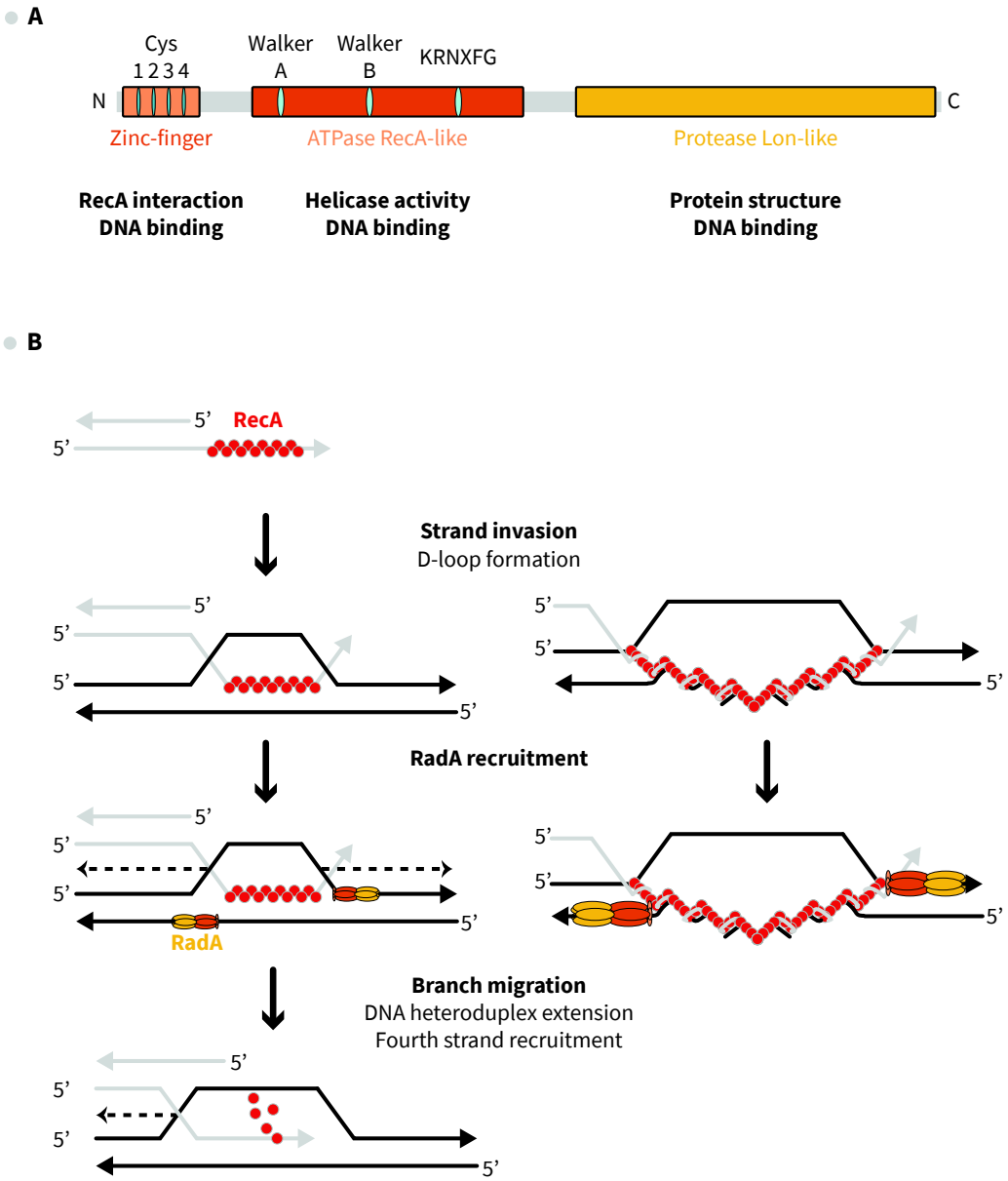


Figure V.1. Structure and function of bacterial RadA. (**A**) Schematic representation of the bacterial protein RadA. The three functional domains are shown with colored boxes and their function annotated. The 4 cysteines of the zinc-finger motive are shown by the turquoise boxes. The Walker A and B and the KRNXG motives are shown in cyan in the ATPase RecA-like domain. (**B**) Strand invasion is managed by the RecA-ssDNA nucleoprotein filament and forms a D-loop. RecA recruit RadA hexamers on each strand of the receiving DNA. RadA moves on ssDNA, opens the receiving molecule (black) and promotes the pairing of one of its strands with the donor DNA (grey). The region of homology enlarges with the translocation of RadA along the molecule and the fourth strand is recruited, to form a Holliday junction. Finally, RadA destabilizes RecA and guide the recombination mechanism for the resolution steps by recruiting RecG and RuvABC. Adapted from Cooper and Lovett, 2016 and Marie et al., 2017.

1. The bacterial RadA

The RadA protein was originally found in *E. coli* because of sensitivity to gamma radiation of the corresponding mutant (Diver et al., 1982). Even if its involvement in the HR mechanism was suggested since 1982 by Diver and collaborators, and confirmed twenty years later because of the synergy between the mutants affecting RadA, RuvA, and RecG (Beam et al., 2002), its precise role was only determined recently (Cooper and Lovett, 2016; Marie et al., 2017).

1.1. Structure of RadA

Bacterial RadA is a three-domain ATP-dependent helicase related in structure to the DnaB replicating helicase (Marie et al., 2017) (**Figure V.1.A**). The N-terminal part of the protein carries four highly conserved cysteines (**Figure IV.8**) forming a C4 zinc-finger structure that is required for the interaction of the protein with RecA. The substitution of any of these four cysteines abolishes the interaction between RadA and RecA (Cooper et al., 2015). The Internal region of the protein shows high homology to RecA and includes Walker A and B motifs characteristic of AAA+ ATPases. The C-terminus of RadA aligns with a region from the Lon protease of bacteria that is also found in RecA. Finally, in the region between the RecA and Lon homology there is a short motif, KNRXG, that is found in all bacterial and plant RadA sequences, but not in other archaeal or eukaryotic RecA-related proteins (Beam et al., 2002).

The motifs Walker A and KNRXG are involved in DNA binding and are essential to the functions of the protein (Cooper et al., 2015). The Lon protease-like domain has no protease activity, but it is required for the scaffolding of the protein structure (Inoue et al., 2017). Moreover, alone the Lon-like domains assemble themselves in a hexameric ring structure able to bind ssDNA. Consequently, the RadA Lon-like domain is required to the oligomerization of RadA into its hexameric active form.

1.2. Roles of RadA as branch migrating factor during HR

RadA is involved in the late steps of the bacterial HR and has been recently identified as a branch migration factor along with RecG and RuvAB (Cooper et al., 2015; Cooper and Lovett, 2016).

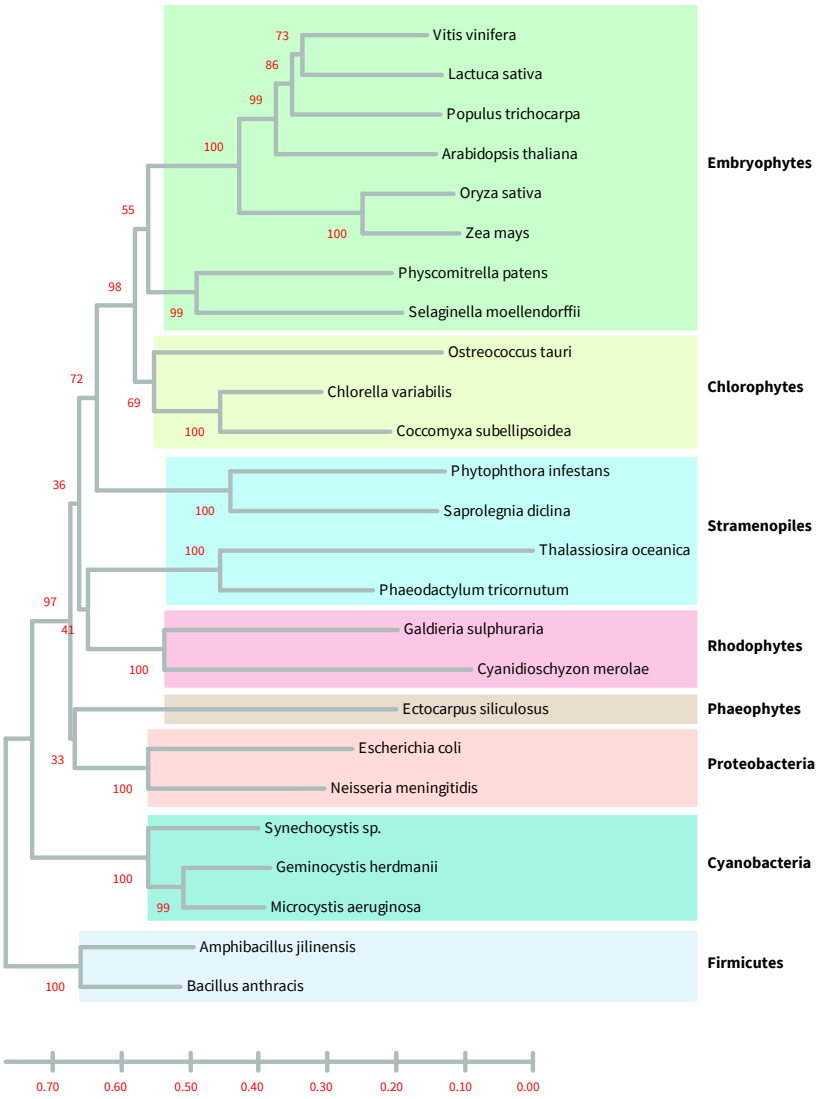


Figure V.2. Evolutionary relationships of RADA. (**A**) Genes coding for RadA-proteins can be found in all bacteria and all plants. Neighbor-Joining tree based on the sequences of *RADA*. Bootstrap values are indicated next to the branches (1100 replicates). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. Adapted from Chevigny et al., 2019.

The survival of the single *radA-* mutant is little affected following genotoxic stress (Beam et al., 2002). However, double mutants *radA- recG-* are severely affected in their ability to repair DNA, which can be surprisingly complemented by overexpression of RuvAB or by the mutation of RecF (Cooper et al., 2015). This result suggests that it is the accumulation of recombination intermediate products that is lethal for the bacteria on a r*adA- recG-* background. Such an accumulation is normally prevented by the three different branch migration factors that are partially redundant, but when two are affected the workload is too heavy for the remaining one, which leads to the accumulation of toxic recombination intermediate products.

During HR, the invasion of homologous dsDNA by the RecA nucleoprotein filament forms a D-loop, whose extremity needs to be extended to increase the region of homology and recruits the fourth strand of the donor DNA (**Figure V.1.B**). In a RecA context, the branch migration factors RecG and RuvB are unable to act, which leads to a dissociation of the D-loop heterocomplex (Whitby et al., 1993). It is the action of RadA interacting with RecA that allows extension of the heteroduplex and recruitment of the additional branch migration factors (Cooper and Lovett, 2016; Marie et al., 2017). Indeed, RadA can accelerate *in vitro* the strand exchange reaction initiated by RecA. The interaction between the two proteins is required for the RadA functions and the positioning of the RadA hexamers on both sides of the D-loop (Cooper and Lovett, 2016; Marie et al., 2017). The RadA ssDNA helicase activity promotes the separation of the two strands of the donor dsDNA and the hybridization of one of the strands with the invading ssDNA (Marie et al., 2017). The loading of RadA through its interaction with RecA on both sides of the D-loop allows bidirectional branch migration leading to a larger region of homology. Finally, RadA could destabilize RecA and consequently guide the HR mechanisms to the resolution steps (Cooper and Lovett, 2016).

2. The RADA protein in *A. thaliana*

Genes coding for RADA can be found in all plants (**Figure V.2**) and protein sequences show strong conservation with their bacterial homologs (Chevigny et al., 2019). Indeed, the sequence alignments showed that plant RADA carries all the important functional motifs described for the bacterial protein (Cooper and Lovett, 2016; Marie et al., 2017). However, RADA from plants show an extension at their N-terminus which could be organellar targeting peptides (**Figure IV.8**). Indeed, a

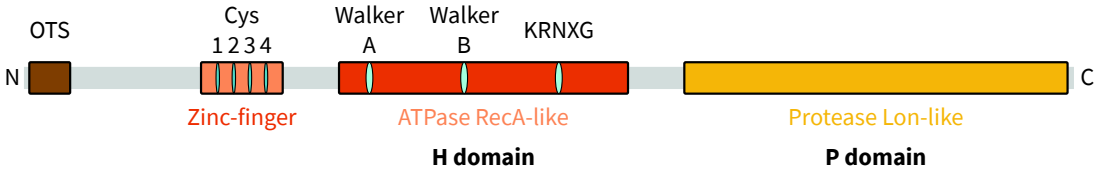


Figure V.3. Schematic representation of the plant protein RADA. The RADA modular structure is similar to the bacterial one. They are composed of the same three domains that are the zinc-finger, the helicase ATPase RecA-like domain (H domain) and the Lon protease-like domain (P domain). Plant RADA sequences show an N-terminal extension containing an organellar targeting sequence (OTS). Adapted from Chevigny et al., 2019.

dual organellar subcellular localization of plant RADA was determined by the translational fusion of the N-terminal part of the *A. thaliana* protein with eGFP (Chevigny et al., 2019).

Plant RADA has the same modular structure as bacterial RadA (Chevigny et al., 2019). This structure comprises the N-terminal C4 zinc-finger, the ATPase RecA-like domain containing the Walker A and B domains, the KNRXG motif and the Lon protease-like domain (**Figure V.3**). The high similarity between the sequences of bacterial RadA and plant RADA has been proved to be linked with similar functions, because the *A. thaliana* RADA can complement the *E. coli radA*- mutant for growth in presence of ciprofloxacin (Chevigny et al., 2019).

The expression and purification of recombinant RADA from *A. thaliana* showed that RADA is mainly purified in hexameric forms (Chevigny et al., 2019). The same authors performed *in vitro* studies and showed that RADA preferentially binds DNA structures containing ssDNA regions and can accelerate *in vitro* strand exchange reactions, as described for bacterial RadA. However, the plant protein seems to inhibit the initiation of new reactions, and consequently to limit the accumulation of final products (Chevigny et al., 2019).

A. thaliana radA mutant lines are severely affected in their development. They display phenotypes of retarded growth of both leaves and roots, and of distorted leaves. Electronic microscopy observation showed that mitochondria in mutant leaf cells are enlarged and less electron-dense, as compared to mitochondria from WT cells (Chevigny et al., 2019). Moreover, *radA* mutants are severely affected in their fertility and produce only few viable seeds. This sterile phenotype seems to originate from both male and female organs (Chevigny et al., 2019). The observation of embryos from the selfing of *radA* mutants demonstrated a diminution of embryo viability after fertilization.

Furthermore, the *A. thaliana radA* mutants showed a strong effect on mtDNA stability, due to an increase in ectopic recombination, with the severity of the developmental phenotypes correlating with the increased reshuffling of the mtDNA by recombination (Chevigny et al., 2019). But the mitochondrial transcriptome analysis did not permit to elucidate the causes of the severe phenotypes observed. Finally, RADA seems to have a more important role in mitochondria than in chloroplast because the cpDNA copy number and stoichiometry are not affected in the *radA* mutant (Chevigny et al., 2019).

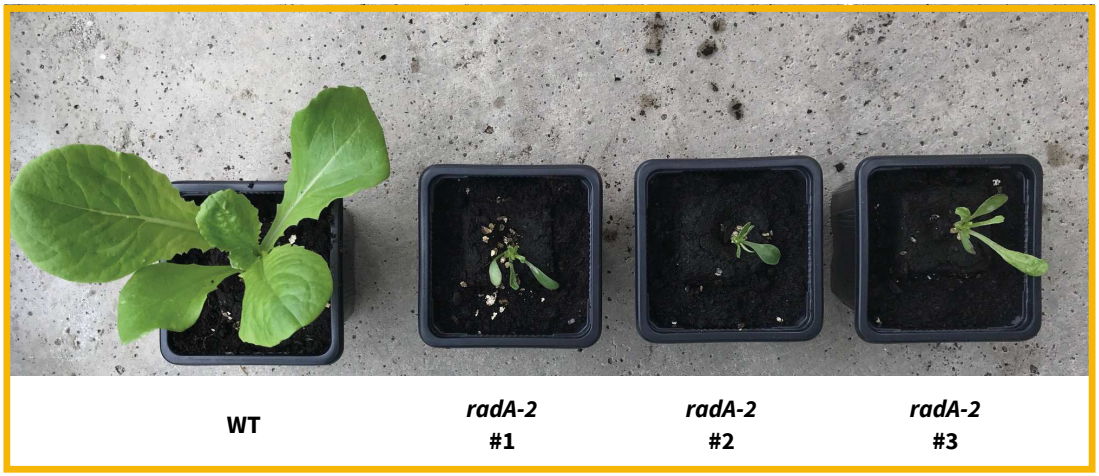


Figure V.4. The phenotype associated to the *radA-2* mutation. The homozygous *radA-2* mutants share the same severe development phenotype. The plants show important growth retardation (as compared to WT plants in the background) and distorted leaves. Moreover, the plants are not able to overtake that stage and bolt. Consequently, the effect of the *radA-2* mutation on plant fertility could not be evaluated.

3. The *L. sativa radA-2* mutant line

In *L. sativa* we identified the gene coding for the RADA ortholog and determined it to be dual targeted to both chloroplasts and mitochondria (see **CHAPTER 3**). We selected two mutants for RADA. The *radA-1* mutant is a missense mutation of cysteine-139 into tyrosine (C128>Y), which is cysteine C1 of the C4 zinc-finger. However, we did not investigate this line further and focused on mutant *radA-2* which codes a severely truncated protein because of a nonsense mutation at position 139 (W139>STOP). This mutant line showed severe developmental phenotypes (**Figure V.4**) and was unable to develop further. Consequently we obtained neither flowers nor seeds, unlike the *A. thaliana radA* mutant lines (Chevigny et al., 2019). Because of its inability to generate a progeny, the mutant needed to be maintained at the heterozygous state. The molecular analysis of a *L. sativa radA-2* mutant showed a strong effect on its mtDNA stability, where two large regions seem to be lost (**Figure IV. 15.E**).

4. Aim of this chapter

Besides the study of *A. thaliana* RADA, little is known about RADA in other plant species. Consequently, the study of the *L. sativa radA-2* mutant represents an opportunity to study this protein in another context. In this chapter we aimed to better investigate the relationship between the affected mtDNA and the severe phenotype displayed by *rada-2* mutants. To do so, we studied the phenotype associated to *radA-2* and we tried to confirm the loss of the two regions of the mutant mtDNA and its effect at the transcriptome level. Because RADA is dual targeted to mitochondria and chloroplast we also studied the impact of the mutation on the cpDNA.

RESULTS

1. Confirmation of the RADA mitochondrial targeting

In **CHAPTER 3** we identified *L. sativa* RADA as dual-targeted to mitochondria and chloroplast (**Figure IV.4**). The chloroplast targeting was unambiguous, because of the colocalization between the fluorescence of the N-RADA-GFP fusion protein and the autofluorescence of chlorophyll. Besides

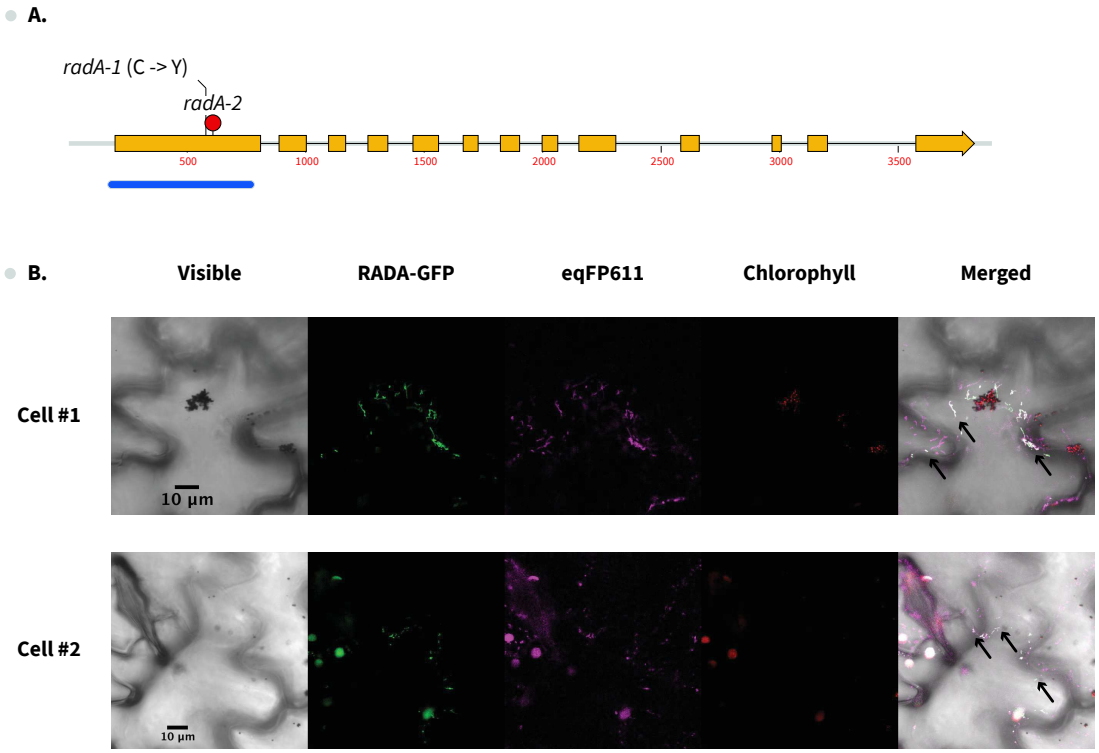


Figure V.5. Confirmation of the RADA mitochondrial subcellular localization. (**A**) *RADA* gene structure and localization of the selected mutations. The amino acid change is represented by a black line and the nonsense mutation is represented by a red circle. The blue bar below represents the sequence fused to eGFP for subcellular localization. (**B**) Transient expression of RADA N-terminus fused to GFP in leaf epidermal cells of *N. tabaccum* constitutively expressing a mitochondria-targeted eqFP red-fluorescent protein. Mitochondria are identified by the eqFP611 and the chloroplasts by the autofluorescence of chlorophyll. Examples of co-localization are

that, many fluorescent spots shaped like mitochondria were observed in the cytoplasm. However, this suspicion of mitochondrial targeting needed to be confirmed by co-localization with a mitochondria-specific marker.

To do so, we used biolistic transfection of the RADA-GFP construct into mesophyll cells of a *Nicotiana tabaccum* line constitutively expressing a mitochondria-targeted eqFP red-fluorescent protein (**Figure V.5**) (Forner and Binder, 2007). Unexpectedly, we found the red fluorescence of mitochondria in the reporter line to be weak, with a high background. Nevertheless, the visualization of the constructs by laser scanning confocal microscopy (Zeiss LSM 700) showed colocalization between the red fluorescence of mitochondrial targeted eqFP611 and the green fluorescence of N-RADA-GFP, in focused structures within the cytoplasm. This confirmed that the N-terminal part of *L. sativa* RADA contains a mitochondria targeting sequence.

2. Phenotype induced by the *radA-2* mutation

During the screens of the mutant lines the genotyping and growth of the plants was managed by Enza Zaden. The selection of the *radA-2* population included in this study took place on four generations (**Figure V.6**). The plants belonging to the *radA-2* family have been named R2#W-X-Y-Z where R2 stands for radA-2 and W, X, Y and Z for their parents in the previous generation. For instance, plant R2#6-4-1 (third generation) is the plant number 1 from the progeny of R2#6-4 (second generation), itself plant number 4 from the progeny of R2#6 (**Figure V.6**). In the first generation, two (+/-) plants were conserved to produce seeds by selfing. The first (-/-) mutants were segregated from the progeny of these two plants. We selected 6 (-/-) mutants from the progeny of R2#1 and 2 from the progeny of R2#6, and phenotyped them. All these eight (-/-) mutants showed the phenotype associated with *radA-2*. From the same segregation studies, the heterozygous *radA-2* plants did not show any particular phenotype. However, we were not able to segregate homozygous mutants from those heterozygous plants. The reason for that is not clear, because on the fourth generation we were able to segregate homozygous mutants from heterozygous plants of the third mutant generation. And once again, the homozygous plants showed the characteristic *radA-2* phenotypes. We were therefore confident that the phenotype observed was induced by the *radA-2* mutation.

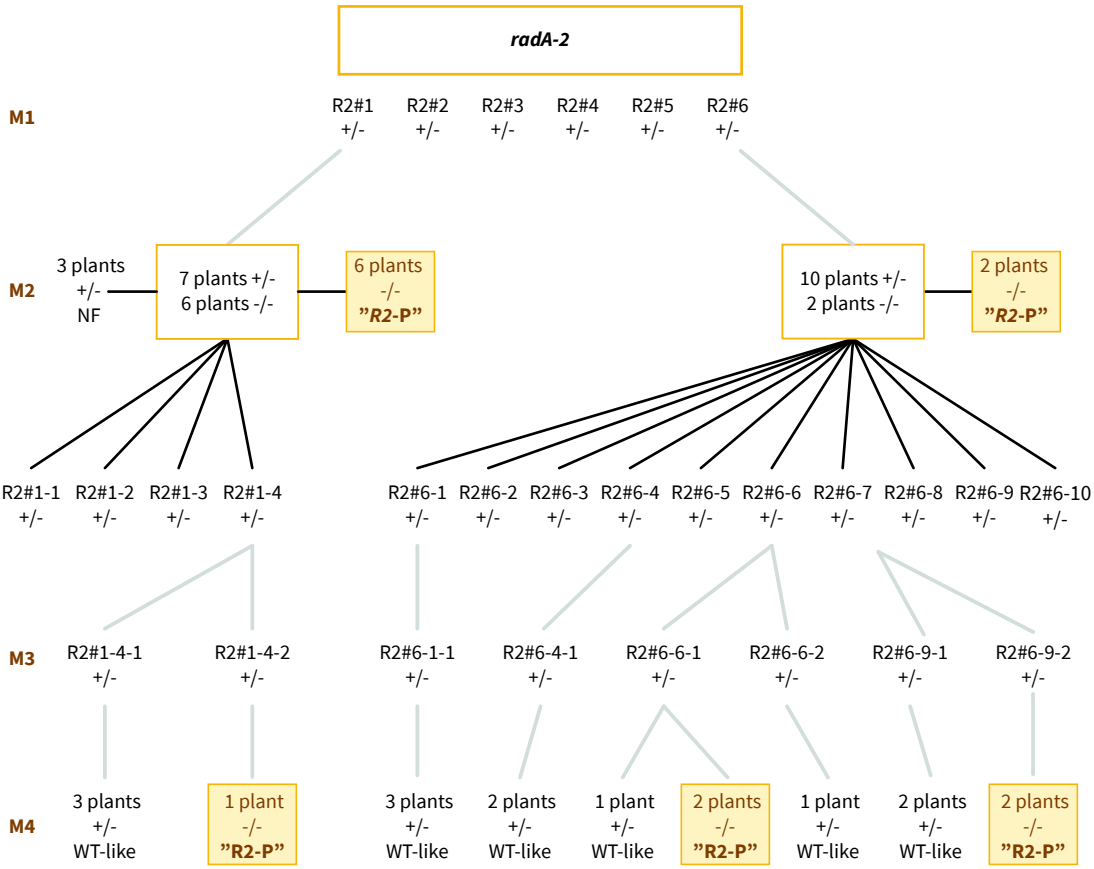


Figure V.6. Genealogy of the *radA-2* mutant line. This genealogy only takes in account the selfings (grey bars) within the mutant lines. Six *radA-2* heterozygous segregated from the EMS-treated plants by selfings. Among these five M1 mutants, two have been selected and their progeny studied: #1 and #6. In their M2 selfing-progeny, 13 and 12 plants have been selected, for #1 and #6 respectively. In both progenies, the homozygous mutants showed severe growth retardation and distorted leaves and were not able to develop further (the *radA-2* phenotype, indicated as R2-P). From the M2 progeny, five heterozygous plants were selected. In the different M3 progenies we did not segregate homozygous mutants. But in the selfing progeny of M4 there were homozygous mutants, with their characteristic phenotype.

The study of *radA-2* at the IBMP was done in plants from the seeds coming from the selfing of R2#6-4-1 (**Figure V.6**). During our study we were unable to setup the genotyping of the *radA-2* mutation by high-resolution melting analysis (HRM). Consequently, we relied on the observation that the severe developmental phenotype only affects the homozygous mutants. Moreover, from the progeny of R2#6-4-1 we phenotyped 90 seedlings, which were either WT-like or showed a typical *radA-2* phenotype (**Figure V.7.A**). Of these 90 plants, 15 showed the *radA-2* associated phenotype and 5 died at an early stage. That corresponds to a 22 % segregation, consistent with the expected Mendelian segregation of the mutation.

In **CHAPTER 3** we described the effect of the homozygous *radA-2* mutation on the development of the plant shoot. In *A. thaliana*, it was shown that the *radA* mutant is also severely affected in root development (Chevigny et al., 2019). We therefore also checked the root development, in the progeny of R2#6-4-1, and observed a strong reduction in roots length that could go up to the virtual absence of roots in certain *radA-2* seedling (**Figure V.7.B**). Moreover, the young plants retained very short roots, and also had a large reduction in lateral roots (**Figure V.7.A & V.8**).

The authors of the study of *A. thaliana* RADA showed that the severe phenotype can be partially relieved by growing the mutants under a short-day photoperiod (8 hours of light) (Chevigny et al., 2019). We tested the same strategy in *L. sativa,* to try to recover *radA-2* plants able to bolt and produce flowers. To do so, *radA-2* mutants were grown under a photoperiod of 12 hours of light. On this growth conditions the WT plants showed a retarded growth of both leaves and roots, as compared to the growth under a long day photoperiod (16h light), but we were unable to see any difference in the growth of *radA-2* seedlings (**Figure V.8**). The growth of *L. sativa radA-2* on a short-day photoperiod (8 hours of light) remains to be tested.

3. Impact of the *radA-2* mutation on the stability of the organellar genomes

By the experiments described in **CHAPTER 3** we got an insight into the impact of the *radA-2* mutation on the stability of the mtDNA, by the determination of the relative copy numbers of the different mtDNA regions (**Figure IV.15.A**). The mutant studied in **CHAPTER 3** belonged to the progeny of R2#1. This mutant showed dramatic changes in the stoichiometry of the mtDNA sequences. This was particularly important for two large regions carrying protein-coding genes that showed a reduction of about 10 and 100-fold, respectively.

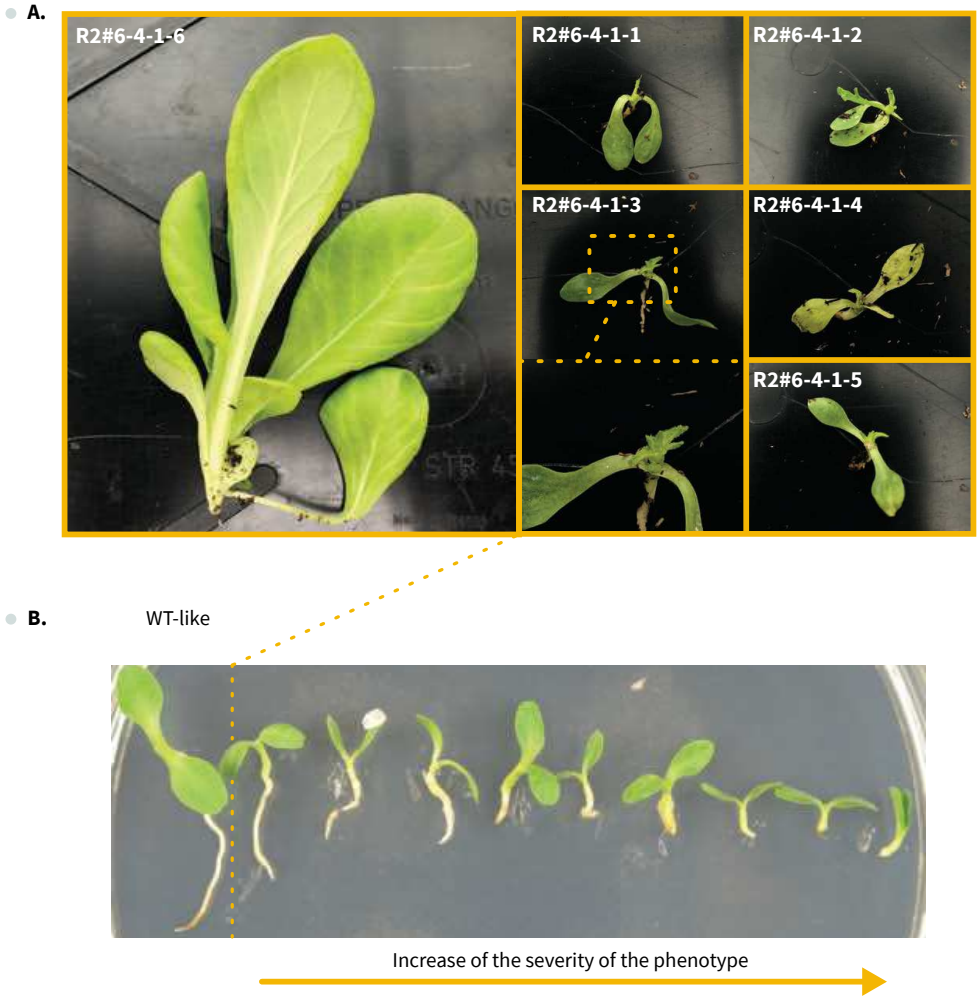


Figure V.7. Impact of the *radA-2* mutation on lettuce plant development. (**A**) The cotyledons of the *radA-2* mutants are not in size, but they are thicker. True leaves of the mutant are short and thin, with a rough and grainy appearance. (**B**) Besides its strong effect on the leaves' shape and length, the *radA-2* mutation induces a strong reduction in the development of the root.

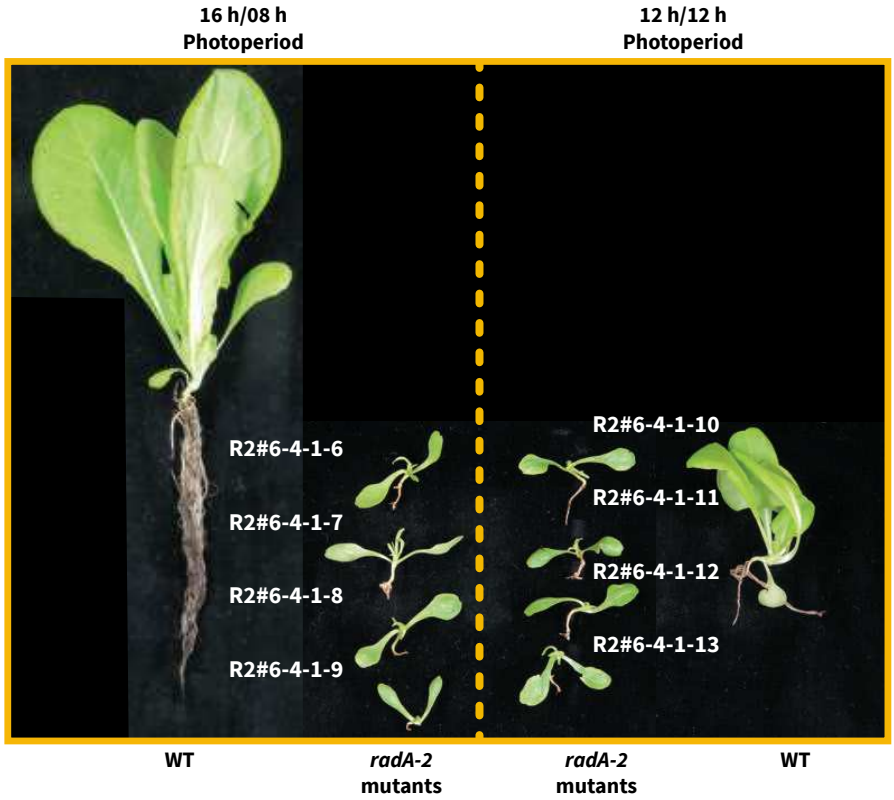


Figure V.8. Lettuce *radA-2* phenotype under different daylight conditions. Seeds from WT and a same *radA-2* heterozygous seed lot were grown under a 16 hour of light and 8 hours of dark photoperiod or under a 12 hours of light and 12 hours of dark photoperiod. The reduction in daylight time affects the development of WT plants, which display reduced leaves length and thickness. However, the *radA-2* homozygous mutants do not show this difference of phenotype.

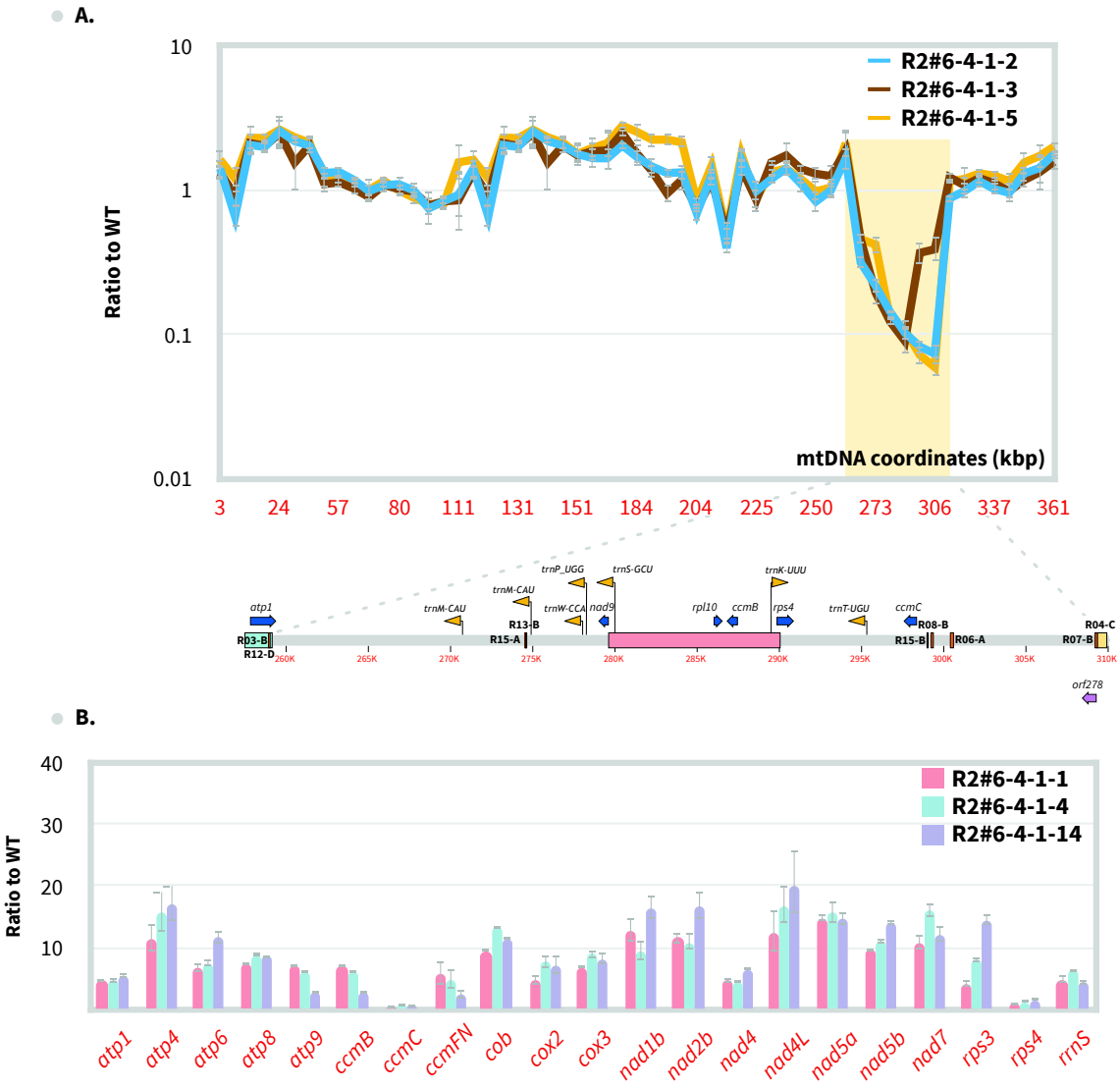


Figure V.9. Changes in mtDNA stoichiometry and gene expression in lettuce *radA-2*. (**A**) Changes in radA-2 mtDNA stoichiometry. Results are shown in logarithmic scale. The mtDNAs of seedlings from *radA-2* mutants were scanned for changes in their stoichiometry, as compared to WT. Sequences spaced 5-8 kb apart on the mtDNA were quantified by qPCR as described for WT stoichiometries in **Figure IV.14**. Results were normalized against the nuclear 18S rRNA gene sequence (**B**) Accumulation of mitochondrial transcripts in *radA-2* mutant seedlings. Mitochondrial transcripts were quantified by RT-qPCR and normalized against the housekeeping gene *TIP41*.

To confirm this result we also tested the mtDNA on additional severely affected *radA-2* plants belonging to the progeny of R2#6. More precisely, three plants coming from the selfing of R2#6-4-1 that had already been phenotyped (**Figure V.7.A**). These plants also presented dramatic changes in the stoichiometry of their mtDNA (**Figure V.9.A**). However, the mtDNA profile of these plants showed the loss of a large mtDNA region (abbreviated LMR-A2) between 257 kb and 310 kb. LMR-A2 corresponds to the region showing a reduction of about 100-fold in the mtDNA profile of R2#1-8 (**Figure IV.15.E**). We were not able to link this event with a recombination event involving a single pair of repeated sequences. But LMR-A2 is bordered by the large repeated sequence R03 on one side and by the large repeated sequence R04 and the small repeated region R07 on the other side. This suggests that the process that resulted in loss of the LMR-A2 probably involved multiple events of recombination, which we could not reconstruct.

LMR-A2 carries the single copies of mitochondrial genes *ccmC, nad9*, *rps4* and *trnW(CCA)*. Moreover, it also carries one copy of *atp1*, *ccmB*, *rpl10, trnK(UUU), trnS(GCU)*, all five carried by the large repeated sequence R02, and *trnM(CAU), trnP(UGG)* and *trnT(UGU)* which present other copies in the genome. The loss of a single copy mitochondrial gene should have a much deleterious effect on plant growth and consequently explain the phenotype associated to *radA-2*. To verify this hypothesis we determined the relative expression level of several mitochondrial protein-coding genes by qRT-PCR (**Figure V.9.B**). For three affected mutants obtained from the selfing of R2#6-4-1 and already phenotyped (**Figure V.7.A**) we confirmed the much-reduced accumulation of *ccmC* and *rps4* transcripts, by a factor of about 15 and 8-fold, respectively, compared to the other genes. The expression level of *nad9* and of the tRNA *trnW(CCA)* were not determined. But the result obtained for *ccmC* and *rps4* confirmed our hypothesis that the loss of LMR-A2 has as consequence the reduced expression of essential mitochondrial genes.

The protein RADA is dual-targeted to mitochondria and chloroplast, both in *A. thaliana* and in *L. sativa*. In *A. thaliana* the *radA* mutation had no apparent effect on the stability of the cpDNA (Chevigny et al., 2019). To study the possible effect of the *radA-2* mutation on the cpDNA of *L. sativa* we determined the relative copy number of the different cpDNA regions (**Figure V.10**). Two of the three tested mutants, R2#6-4-1-3 and R2#6-4-1-4, did not show any changes in their relative cpDNA stoichiometry. Surprisingly, the third mutant tested, R2#6-4-1-5, showed a region increased in copy number as compared to neighboring regions, along with a second region that was reduced. The causes for these changes in cpDNA stoichiometry still need to be determined.

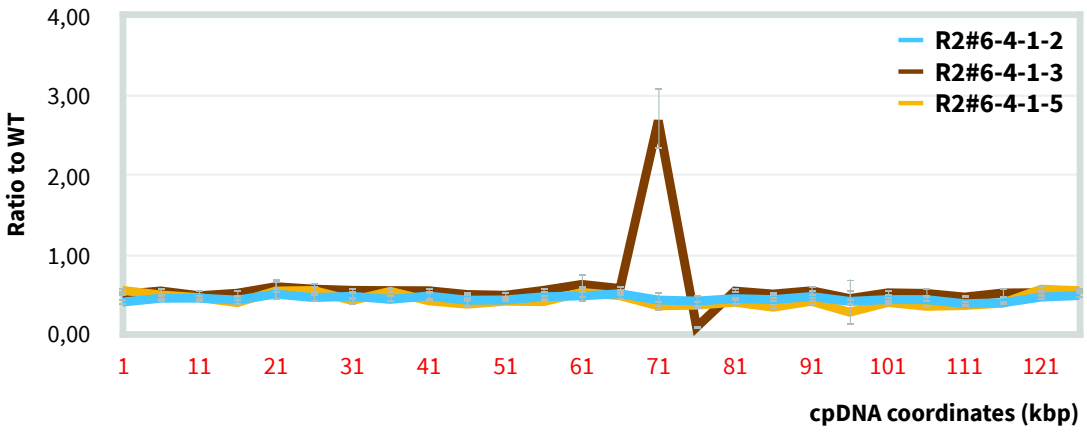


Figure V.10. Changes in lettuce *radA-2* mutant cpDNA stoichiometry. The cpDNA from the *radA-2* plants tested in **Figure V.9.A** and relative WT were scanned for changes in copy number and stoichiometry. Results are represented in logarithmic scale. Sequences spaced 5-8 kb apart on the cpDNA were quantified by qPCR as described for mtDNA stoichiometries in **Figure IV.14**. Coordinates correspond to the published *L. sativa* cpDNA.

In this chapter we aimed to investigate the causes of the *radA-2* phenotype. To do so we linked the phenotype to the *radA-2* mutation by looking at the progeny of two different mutants. Then we linked the *radA-2* mutation with the loss of a region of the mtDNA which was confirmed by a strong reduction in the expression of genes carried by it. Finally, we showed evidence that, despite of the chloroplast targeting of RADA, the stability of the cpDNA is not apparently involved in the *radA- 2* associated phenotype.

Most of the work of this chapter relies on the fact that the "*radA-2* phenotype" is only developed by the homozygous mutants. Consequently, and because we were not able to efficiently genotype this mutation in the laboratory, we based the selection of plants on the phenotype. However, the plants that were further analyzed for the stability of their organellar genomes (R2#6-4-1-2, R2#6-4-1-3, R2#6-4-1-4 and R2#6-4-1-5) should be properly genotyped. One solution is to externalize it within Enza Zaden, where the markers and the genotyping of this mutation work in routine.

We initiated a strategy to try to alleviate the severe "*radA-2* phenotype" based on previous work on the *A. thaliana radA* mutants (Chevigny et al., 2019). Because of time constraints we were unable to conduct the last screen, under short-day photoperiod (8 hours of light). The possibility of obtaining flowers from *radA-2* would allow us to phenotype the fertility of the mutant. In *A. thaliana* the *radA* mutant does produce viable-pollen, which apparently is unable to bind to the stigma, and consequently the mutant shows a reduction in fertility (Chevigny et al., 2019). It would be interesting to test if the same also happens in lettuce. We would also be able to outcross the *radA-2* mutation and re-introduce the WT allele of *RADA*. Such outcross could lead to the stabilization of new forms of the mtDNA. Consequently, these alternative mitotypes could be sorted in the progeny and potentially segregate mitochondria-associated phenotypes.

We identified LMR-A2 as a region of the mtDNA highly reduced in all individual mutants tested. This suggests that the same events of mtDNA recombination and/or segregation are triggered by the loss of RADA. But we could not determine the possible events leading to this loss. A possibility is that the mtDNA region LMR-A2 corresponds to a subgenome whose replication or segregation depends from RADA functions. This might be better understood by the sequencing of

the *radA-2* mtDNA. Moreover, the presence in LMR-A2 of the large repeated sequences R02 (10 430 bp) and R03 (3 552 bp) requires long reads sequencing strategies such as PacBio or MinION, to sequence across them. But that requires mtDNA enrichment that should be tricky on such affected mutant. Furthermore, following the mitochondrial purification protocol described in **CHAPTER 1** we started with etiolated seedlings, and we do not know if *radA-2* mutants can grow under such stress. The project could benefit from the NEBNext® Microbiome DNA Enrichment Kit (New England Biolabs, Ipswich, Massachusetts, United States) described previously, to enrich mtDNA from a total DNA sample.

Among the single copy genes carried by the LMR-A2 regions we could show the reduced accumulation of the *ccmC* and *rps4* transcripts. The same still needs to be done for *nad9* and *trnW(CCA)*. Indeed, nad9 is involved in the electron transport chains which produces ATP and is consequently a good candidate for such a phenotype. For studying *nad9* expression the strategy is the same as for *ccmC* and *rps4,* by qRT-PCR. There is also a good nad9-specific antibody available that should allow us to test *nad9* expression at the protein level (Lamattina et al., 1993). However, the analysis of tRNAs expression is difficult, because of their small size and base modifications that inhibit reverse transcriptase. So the study of *trnW(CCA)* expression requires northern blot analysis.

By doing so, we would be able to clarify the expression of the single copy genes present in LMR-A2 and therefore obtain stronger evidence for the correlation between mitochondrial dysfunction and the "*radA-2* phenotype". These results are complementary to the work on the *A. thaliana radA* mutant, where mitochondrial transcriptomic analysis did not reveal any significant changes in the steady state of mitochondrial transcripts that could explain the severe *radA* growth phenotype (Chevigny et al., 2019).

In *A. thaliana*, the instability of the mtDNA induced by the *radA* mutation has been linked to a defect in cell cycle progression (Chevigny et al., 2019). Besides its strong impact on the growth of the plant the *A. thaliana radA* mutant induces a leaf-phenotype of enlarged leaf epidermal cells and fewer stomata, visible under scanning electron microscopy (Chevigny et al., 2019). The DNA ploidy levels of these leaves determined by flow cytometry show that the cells of the *radA* mutant displayed significantly different nuclei contents than the WT, suggesting a blockage in cell cycle progression (Chevigny et al., 2019). The authors also showed the strong induction of two genes in *radA* mutants: *SMR5* encoding an inhibitor of cell cycle and *ANAC085,* involved in cell cycle inhibition in response to DNA damage. Such results lead to think that a retrograde response is involved.

Among the retrograde response effector reactive oxygen species (ROS) were shown to increase in the *radA* mutant, which could explain the induction of genes controlling cell cycle progression (Chevigny et al., 2019). We did not initiate such a study, but our work could benefit from further insights into the mitochondria-nucleus cross-talk triggered by mitochondrial dysfunction in *L. sativa*.

This study focused on the *radA-2* mutant, but it was not our only *radA* mutant. Indeed, the *radA-1* mutant that is a missense mutation of cysteine-139 into tyrosine (C128>Y) is also available, but not yet segregated to the homozygous stage. The affected cysteine is one of the four constituting the zinc-finger domain and should potentially affect DNA binding activity and/or the interaction with RECA proteins, leading to a knockdown mutation rather than to a knockout such as the *radA-2* mutation. Consequently, it would be interesting to segregate homozygous *radA-1* plants and study their development phenotypes and mitochondrial genome stability.

CHAPTER 6: Study of a male sterile line

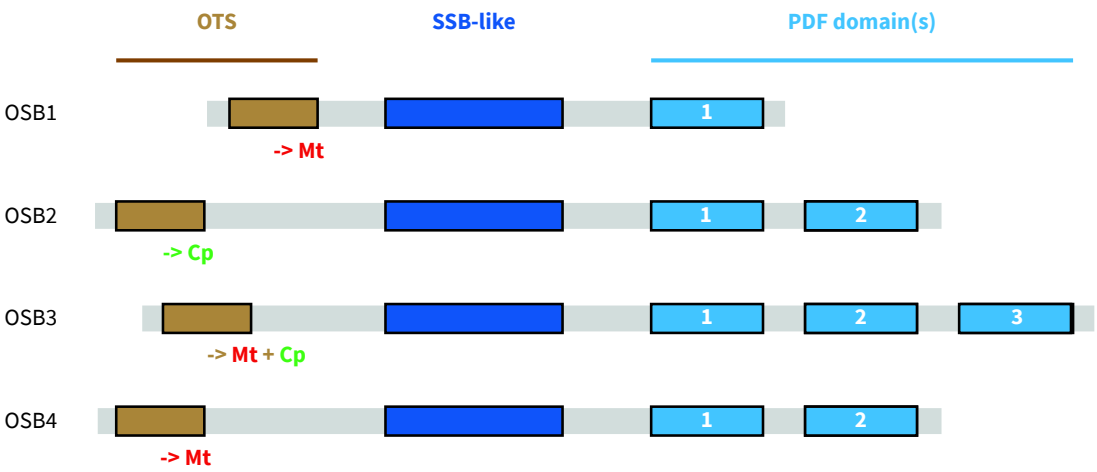


Figure VI.1. Schematic representation of the *A. thaliana* OSB proteins. The four *A. thaliana* OSBs proteins share an OB-fold ssDNA binding-like domain (SSB-like) followed by up to three highly conserved PDF domains. The OSBs are differentially targeted according to their organellar targeting sequence (OTS). Their experimentally determined subcellular localizations are specified below each OTS. Cp: Chloroplast; Mt: Mitochondria.

1. The plant organellar ssDNA-binding proteins (OSBs)

Proteins belonging to the OSB family are specific to plants. These proteins show a high affinity to ssDNA and are composed of two main domains: a poorly conserved oligonucleotide/oligosaccharide-binding (OB) fold–like domain and an OSB-specific domain named PDF (Zaegel et al., 2006). The OB fold–like domain is shared with bacterial SSB and is assumed to be involved in the oligomerization of OSBs, as it does in bacteria (Raghunathan et al., 1997, 2000). The PDF domain contributes to the interaction with ssDNA, and up to four can be found in OSB proteins (Zaegel et al., 2006). The nuclear genome of *A. thaliana* encodes for four OSB proteins which are targeted to mitochondria (OSB1 and OSB4), plastids (OSB2), or to both (OSB3). These proteins carry different numbers of PDF domains: one in OSB1, two in OSB2 and OSB4 and three in OSB3 (**Figure VI.1**). Mutants for *osb1* show phenotypes associated with mtDNA instability, induced by the increase of ectopic homologous recombination (Zaegel et al., 2006). Such results have been also observed in *osb4* mutants in the laboratory but have not been published yet.

In *L. sativa*, we identified only two genes coding for OSBs proteins: OSB1 and OSB2 carrying one and two PDF domains, respectively. In **CHAPTER 3** we determined that OSB1 is targeted to both organelles, while OSB2 is apparently only targeted into mitochondria.

2. The *L. sativa osb2-1* mutant line

In **CHAPTER 3** we described how we selected two mutants for OSB2, *osb2-1* and *osb2-2*, which are both nonsense mutations at positions 140 (W140>STOP) and 389 (W389>STOP), respectively. We were only able to segregate homozygous plants for *osb2-2*. These *osb2-2* homozygous plants as well as heterozygous *osb2-1* plants did not show any development phenotype. Moreover, for both mutants we were not able to detect mtDNA instability. However, a phenotype of male sterility was reproducibly found in the progeny of *osb2-1*, but its origin remained to be determined.

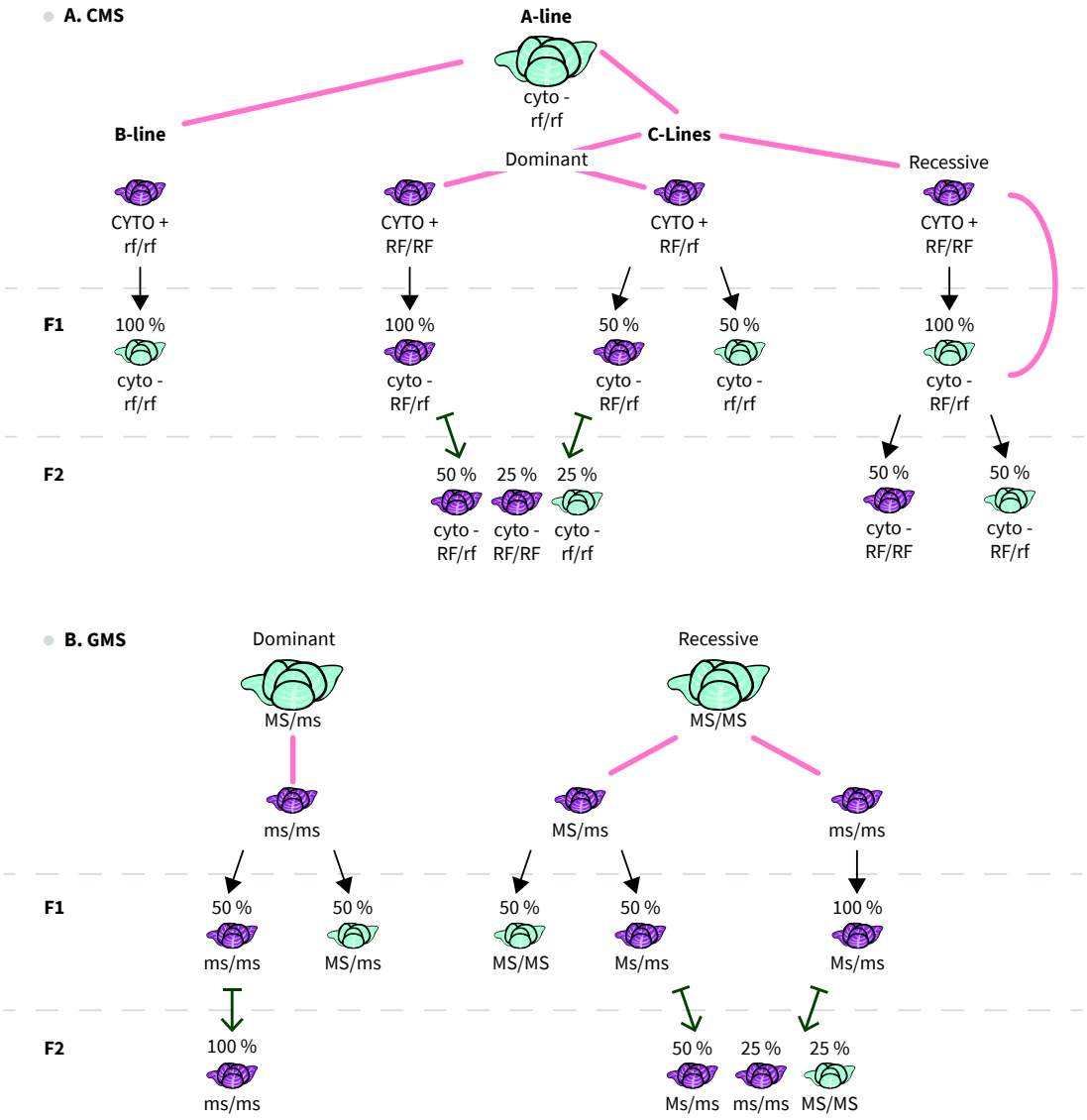


Figure VI.2. Segregations of the male-sterile phenotype. (**A**) Segregations of the male-sterile phenotype in a CMS system. The sterile cytoplasm is called cyto - while the fertile is named CYTO +. The restorer of fertility allele is annotated RF, unlinked to its dominance/recessivity, and the non-restorer allele rf. (**B**) Segregations of the male-sterile phenotype in a GMS system. The sterilizing allele is annotated MS, unlinked to its dominance/recessivity, and the non-sterilizing allele ms. The pink lines (-) link the two parents of a cross. The progeny of the cross is linked to one of the parents by a black arrow. The green arrows (I->) represent the progeny resulting from selfing. The blue and green colors of the plants represent the fertility and sterility of pollen as tested by Alexander staining, respectively.

3. Genetic rules of male sterility

A genetic phenotype of male sterility can have two origins. The phenotype can be the consequence of an allele of a nuclear gene. In such a system, named genic male sterility (GMS), the allele can induce sterility either at the homozygous or at the heterozygous stage, qualified as recessive-GMS or dominant-GMS, respectively. The other possible origin for genetic male sterility lies in the cytoplasm, and more precisely in the mitochondria. Such a system is named CMS and is usually induced by the expression of a chimeric gene generated by the recombination of the mtDNA.

Because of their different origins, CMS and GMS can be discriminated with respect to the segregation of the male-sterile phenotype (**Figure VI.2**). In the CMS system, the male-sterile line (A-line) can be crossed with two different lines (**Figure VI.2.A**): with the maintainer (B-line) which has the exact same nuclear genetic content but within a fertile cytoplasm context or with a restorer line (C-line) carrying a fertility-restorer gene(s). The cross between the A and B-lines leads to a whole male-sterile progeny due to the maternal transmission of the mtDNA responsible for the phenotype, while the cross between the A and C lines depends on the restorer allele. A dominant restorer of fertility allele introduced in the sterile-cytoplasm context leads to a whole male-fertile progeny, while a recessive restorer needs two generations to reach the homozygous stage and then a male-fertile progeny. The cross of a recessive-GMS line with a heterozygous shows a segregation of the male-sterile phenotype (½) while the selfing of the male-fertile progeny leads to new segregation of the male-sterile phenotype (¼) (**Figure VI.2.B**). The cross of the same recessive-GMS line with a homozygous leads to a whole male-fertile progeny, but the selfing of this progeny shows segregation of the male-sterile phenotype (¼). In the case of a dominant-GMS system, the homozygous stage does not exist, because the heterozygous stage is male sterile and can not be selfed. So the cross of such a line with a fertile line leads to a progeny half male-fertile, half male-sterile.

However, the systems can be indistinguishable from each other at some generations. Indeed, a segregation in F2 of the male-sterile phenotype at ¼ could be due to a CMS system in presence of a dominant gene restorer of fertility or to a recessive GMS. The solution to identify the origin of the male sterility is to realize a topcross; meaning that the male-sterile line is crossed with several male-fertile plants and the segregation of the phenotype within these progenies will help to

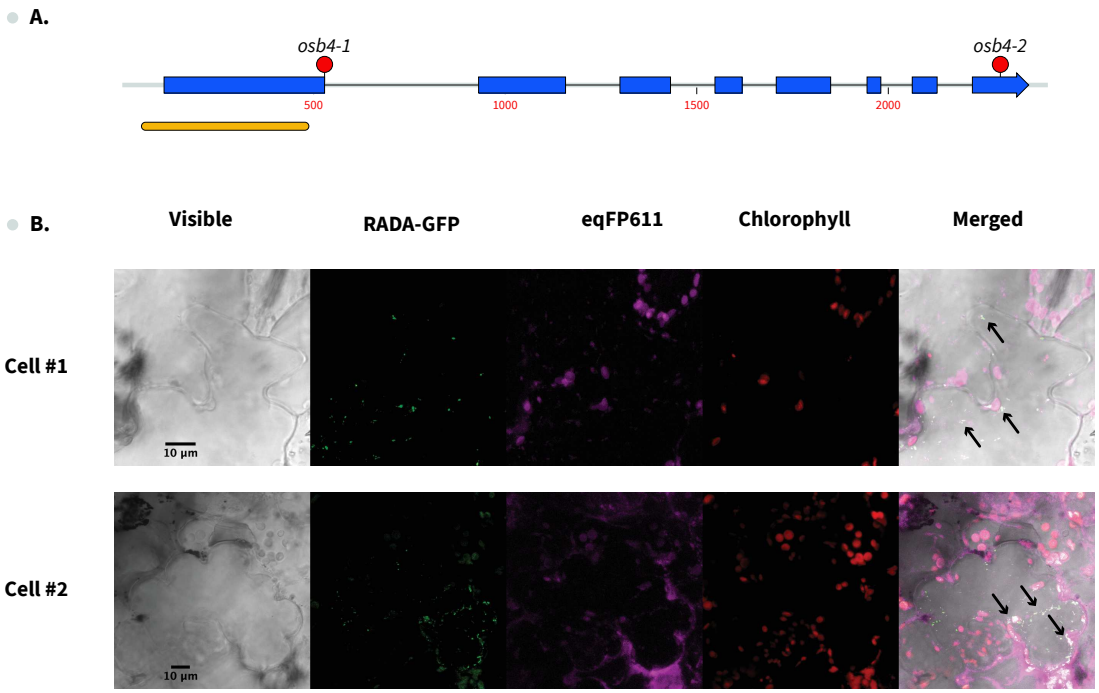


Figure VI.3. Confirmation of the OSB2 mitochondrial subcellular localization. (**A**) Lettuce *OSB2* gene structure and localization of the selected mutations. The amino acid change is represented by a black line and the nonsense mutation is represented by a red circle. The yellow bar below represents the sequence fused to eGFP that was tested for subcellular localization. (**B**) Transient expression of OSB2 N-terminus fused to GFP in leaf epidermal cells of *N. tabaccum* constitutively expressing a mitochondria-targeted eqFP red-fluorescent protein. Mitochondria are identified by the eqFP611 and the Chloroplasts by the fluorescence of chlorophyll. Examples of co-localization are indicated by the arrows.

elucidate the trait of male-sterility: a CMS will be approved when, within a progeny, 100 % male sterile plants are obtained.

4. Aim of this chapter

In this chapter, we describe our study of the *osb2-1* mutant line. To do so, we oriented our work in two directions. The first concerned the *osb2-1* mutation itself, and the impossibility to segregate a homozygous *osb2-1* mutation. The second was the identification of the origin of the male-sterility phenotype observed in the progeny of *osb2-1*.

RESULTS

1. Study of the *osb2-1* heterozygous mutant

1.1. Confirmation of the mitochondrial targeting of OSB2

Previously in **CHAPTER 3** we described that we found *L. sativa* OSB2 as targeted to mitochondria only. For further study of this protein and the possible effect of the mutant on the mtDNA we needed to confirm this result, by colocalization with a mitochondrial marker.

To do so, as described in **CHAPTER 2** for RADA, we transformed by biolistic mesophyll cells of the *Nicotiana tabaccum* line constitutively expressing a mitochondria-targeted eqFP red-fluorescent protein (**Figure VI.3**) (Forner and Binder, 2007). As for RADA, the observation of the transfected tobacco leaves showed a colocalization between the red fluorescence of eqFP611 and the green fluorescence of the OSB2-GFP fusion protein, confirming our previous observation and thus the strict mitochondrial targeting of OSB2.

1.2. Identification of the causes of the impossibility to get homozygous mutants

A particularity of the *osb2-1* mutant is the impossibility to segregate it at the homozygous stage from the heterozygous, by crosses or by selfing. We suggested the hypothesis that the *osb2-1* homozygous mutant was embryo lethal. In such a case the embryo-lethality would lead to aborted seeds.

A

Plant	Pistils					TOTAL	Seeds produced					Seeds	
	Fleur #1	Fleur #2	Fleur #3	Fleur #4	Fleur #5		Fleur #1	Fleur #2	Fleur #3	Fleur #4	Fleur #5	Viable	Aborted (%)
***osb2-1* +/- #1**	17	14	16	16	19	82	13	10	10	14	17	64	12.33
***osb2-1* +/- #2**	16	17	17			50	15	13	15			43	7.53
***osb2-1* +/- #3**	16	16	18	13		63	14	9	11	12		46	15.60
***osb2-1* +/- #4**	18	18	15	15	17	83	12	15	6	10	11	54	21.17
TOTAL						278						207	**14.64**
WT #1	18	13	13	14	17	75	15	13	10	14	16	68	4.90
WT #2	18	16	15	19	16	84	18	15	14	15	16	78	3.70
WT #3	20	14	14	14	15	77	18	14	14	13	15	74	1.99
WT #4	17	17	14	20	18	86	17	16	13	15	16	77	5.52
WT #5	20	13	15	16	14	78	17	13	15	15	13	73	3.31
WT #6	14	18	17	19		68	14	18	17	19		68	0.00
SOUS-TOTAL						468						438	**3.31**

B

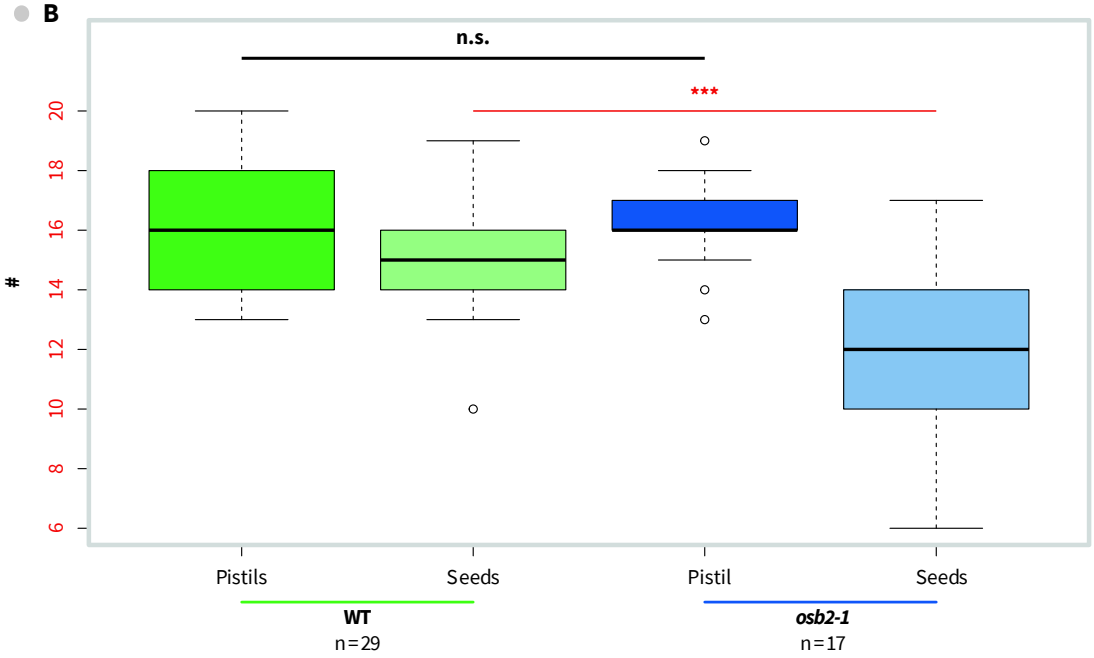


Figure VI.4. *osb2-1* +/- seeds production by selfing. (**A**) The pistils and seeds produced by 29 flowers coming from 5 WT plants and by 17 flowers from 3 different *osb2-1* plants were counted. (**B**) The *Student's t-test* on different populations shows no significant differences between WT and mutant in pistil production, but a significant difference in seeds production.

To confirm this hypothesis we determined the number of pistils of WT and heterozygous *osb2-1* male-fertile flowers. The number of pistils determines the theoretical maximum number of seeds that is possible to obtain from the flower. To ensure an optimal seed production by selfing we fertilized the flowers by hand, with the pollen from another flower of the same plant. Then, at seeds maturation, we counted the pistils of 29 flowers from 5 WT plants and 17 flowers from 4 *osb2-1* heterozygous plants (**Figure VI.4**). We found no significant difference between the WT and the mutant, regarding the number of pistils displayed by the flowers. We realized a t-test two-sample assuming equal variance between the two sets of data. For this we compared seeds quantities obtained with pistils quantities for both genotypes: (t-test = 6.666, k = 47 p = 2,62E-08); t > 2.012 t critical two-tail and p = 2.62E-08 < 0.05, so there is a significant difference between the two data sets. Then we compared seeds quantities obtained with the theoretical number of seeds expected (75% in heterozygous genotype, 100% in WT genotype): (t-test = 2.550, k = 47, p = 0,01); t > 2.012 t critical two-tail and p = 0.014 > 0.05, so there is no significant difference between the two data sets. Consequently, the quantity of seeds produced in this population is consistent with the expected reduction in seeds production (25 %) in a embryo lethal homozygous mutant.

1.3. Characterization of the *osb2-1* transcript

The mutant *osb2-1* is a SNP (G to A) at position 420 starting from the initiation codon. This is a particular position because it corresponds to the donor (5') splicing site of the first intron. Consequently, we presumed that this mutation induces a defect in the splicing of the first intron.

Preliminary *in silico* work showed that splicing deficiency would introduce inframe STOP-codons, and one at the very same site of the SNP. Indeed, the mutation created a STOP codon (UG**A**) in the frame of the first exon. To validate it *in planta* we prepared a construction corresponding to the *osb2-1* genomic sequence and realized the translational fusion of the genomic sequence with GFP, under the control of a ubiquitin (*UBQ10* gene) promoter sequence. This construct was used to transfect *N. benthamiana* leaves by agroinfiltration. In parallel we also tested a control construction, of GFP alone under a constitutive 35S promoter. As expected from our *in silico* model, we were not able to detect any fluorescence with this construction, while the control showed very good expression of GFP in the cytoplasm, showing that our transfection protocol worked correctly. This result supported the hypothesis that the mutation induces a premature stop-codon induced by a complete deficiency in the splicing of the first intron. Moreover, there was no evidence that a

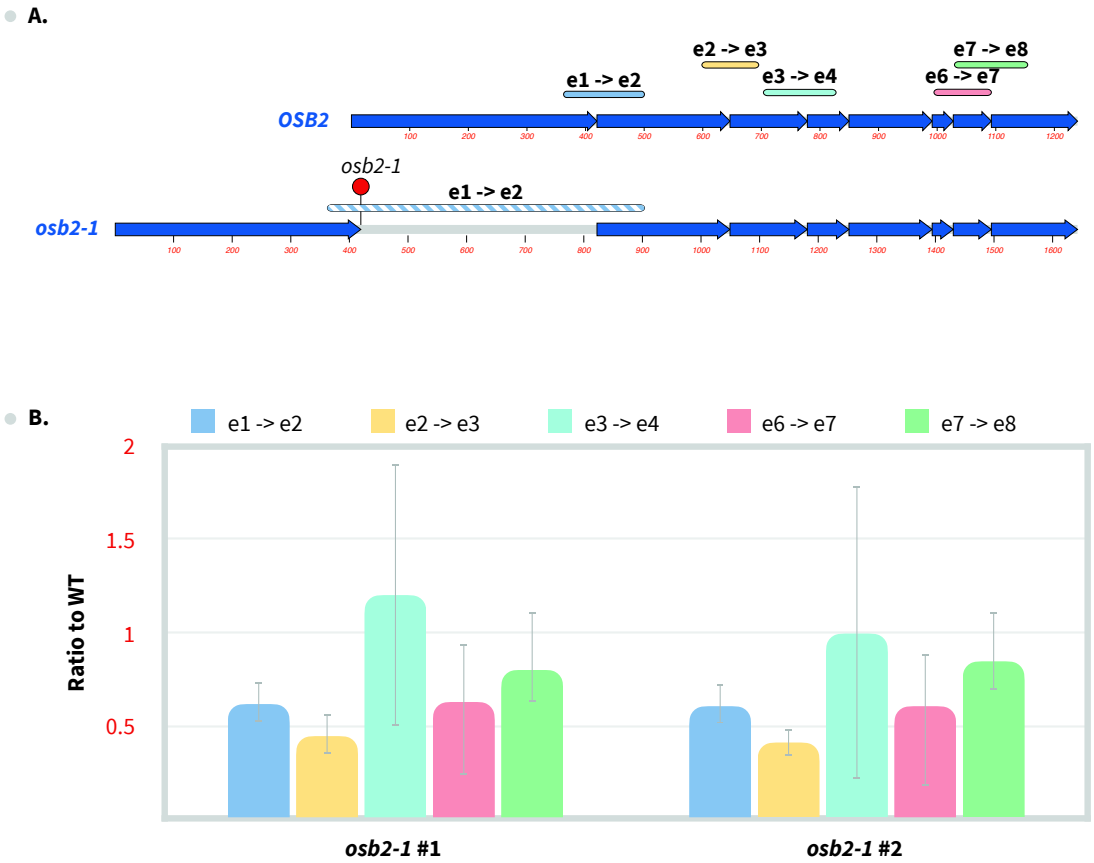


Figure VI.5. Accumulation of lettuce *OSB2* transcripts. (**A**) Structure of the *OSB2* transcript and of the hypothetical *osb2-1* cDNA retaining the intron 1 sequence. The colored bar represents the different exon-borders tested by qRT-PCR amplification in **B**. In *osb2-1* cDNA, the amplification of the border between exons 1 and 2 is represented by a dashed line because it should not be possible to efficiently amplify it in the qPCR run. (**B**) Quantifications obtained for two *osb1-2* heterozygous plants. Total RNA was from sterile flowers, and results were normalized to the nuclear *TIP41* transcript that is constitutively expressed. The level of transcripts showing amplifiable exon borders is represented. Error bars are the SD values from three technical replicates.

truncated protein could be synthesized, by re-initiation of transcription downstream from the mutation.

The splicing defect should result in a longer *OSB2* mRNA, retaining the first exon. To confirm it we tried to amplify such *osb2-1* transcript by RT-PCR. Because we are in a heterozygous system we were always able to detect the mRNA corresponding to the WT allele. But using several different pairs of primers we found no evidence for the presence of larger mature mRNA transcripts retaining intron 1 and with spliced intron 2. We also tried a qRT-PCR approach (**Figure.VI.5**). We designed pairs of primers to specifically amplify through the exon borders, which meant that if there is no splicing the amplicon would be too big to be efficiently detected (**Figure.VI.5.A**). We tested two independent *osb2-1* mutants, for five of the splicing sites of the gene. They both showed the same profiles, with a global reduction of the *OSB2* transcripts. This reduction is higher for the first two splicing sites, with a ratio to the WT reduced to 50 %. This could mean that in these samples half of the transcripts are not spliced, as suggested by our hypothesis. However, the global reduction also suggested that the whole *osb2-1* transcript is degraded. That could be because of nonsense-mediated decay (NMD) of the transcript containing an internal stop codon.

To further support this hypothesis we sequenced the RNA from *osb2-1* flowers. The library preparation and sequencing were managed by BGI GENOMICS (Shenzhen, Guangdong, China) by paired-end (2 x 100 bases) DNA nanoballs sequencing (DNBseq). We obtained 70 000 000 reads, and among them 610 mapped to the genomic sequence of *OSB2* (**Figure.VI.6**). The coverage of the gene sequences revealed no evidence for a transcript retaining the non-spliced intron 1. These results suggested that indeed the *osb2-1* transcript is degraded. However, the other exons show a coverage similar to the first exon, at the very exception of the third exon, suggesting that the transcriptions of these exons are not affected. Therefore, we cannot exclude a re-initiation of transcription somewhere between the first intron and the third exon. The third exon presents a higher coverage than the others as well as the first part of the third intron, but this result need to be statistically confirmed with the support of other RNAseq runs. Interestingly, we observed that the mutation site is covered by 10 reads (**Figure.VI.6.B**). Among these 10 reads 6 carry the WT-allele and 4 the *osb2-1* associated SNP. However, all 4 reads carrying the SNP are second reads, meaning they are at the 3' end of the insert. This suggests that the mutant transcript does not go further than base 434 of the gene.

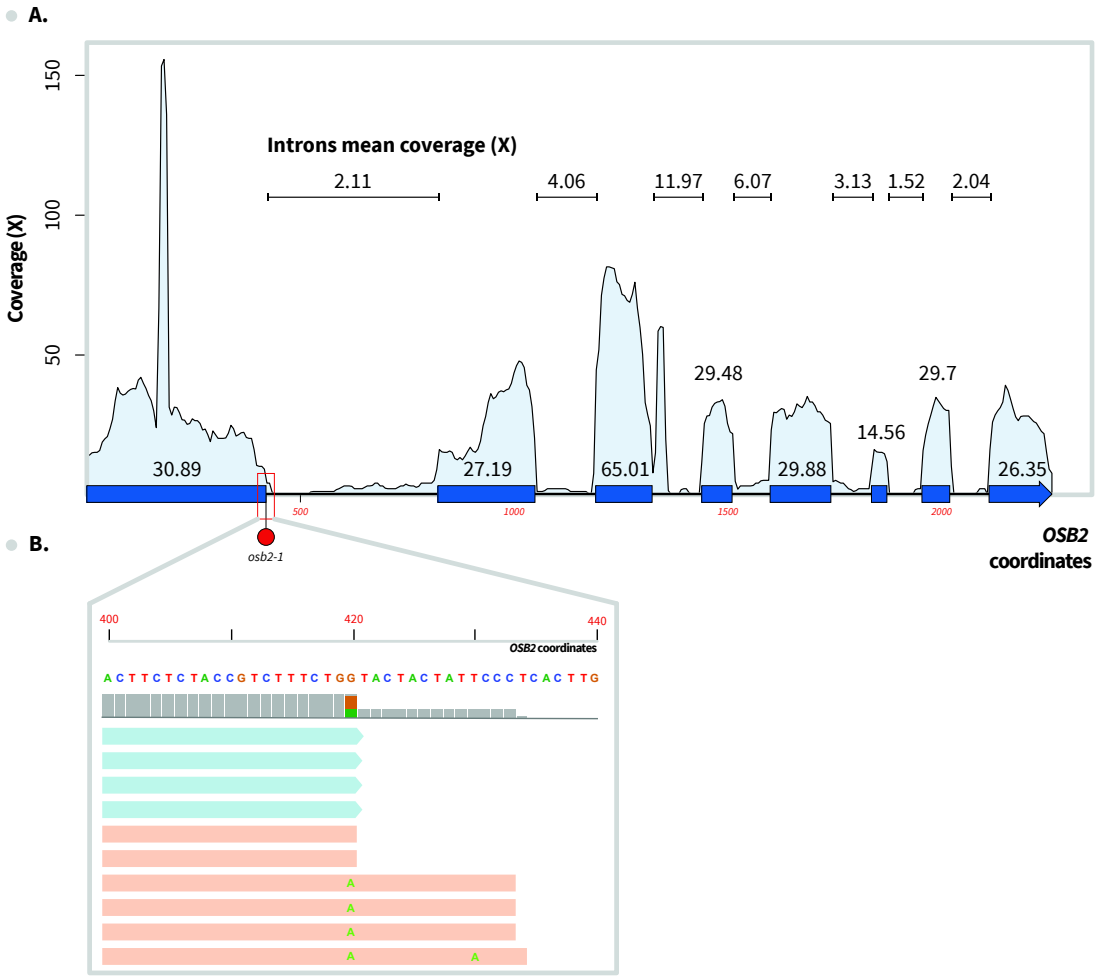


Figure VI.6. RNAseq *OSB2* coverage. (**A**) Coverage of the *OSB2* genomic sequence by reads from a RNAseq run . The gene structure of *OSB2* is represented on the x-axis. The means coverage of the different exons and introns is specified. (**B**) Zoom on the *osb2-1* mutation site. The reads are represented by a colored box: the first read (5' to 3') of a pair is colored in pale blue and the second read (3' to 5') of a pair is colored in orange. The SNPs are represented by the letter on the reads and by a different color on the coverage. The position of the SNP (position 420) is covered by 10 reads, 4 of which carried the "A" SNP of *osb1-2*.

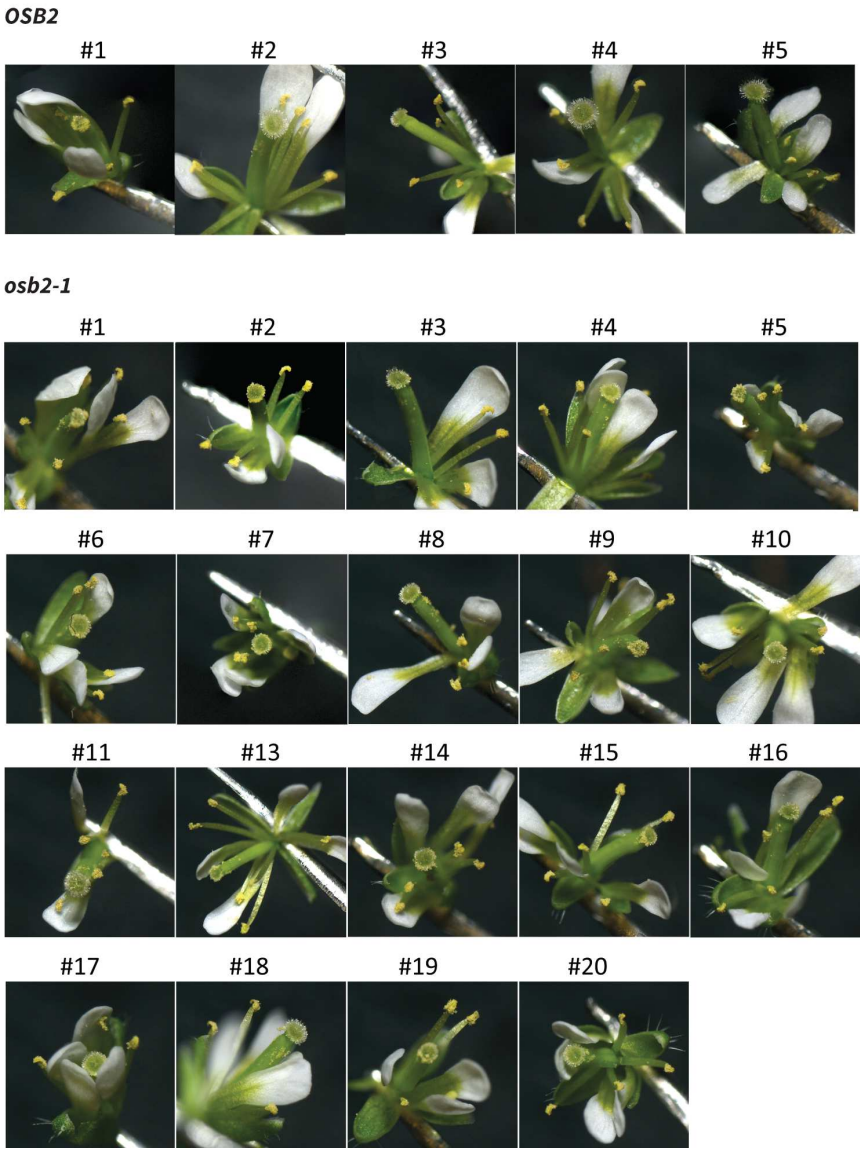


Figure VI.7. Phenotype of *A. thaliana* flowers transformed with *L. sativa OSB2* genes. The flowers of *A. thaliana* transformed either with *L. sativa OSB2* or with *osb2-1* show WT-like phenotype for the production of pollen.

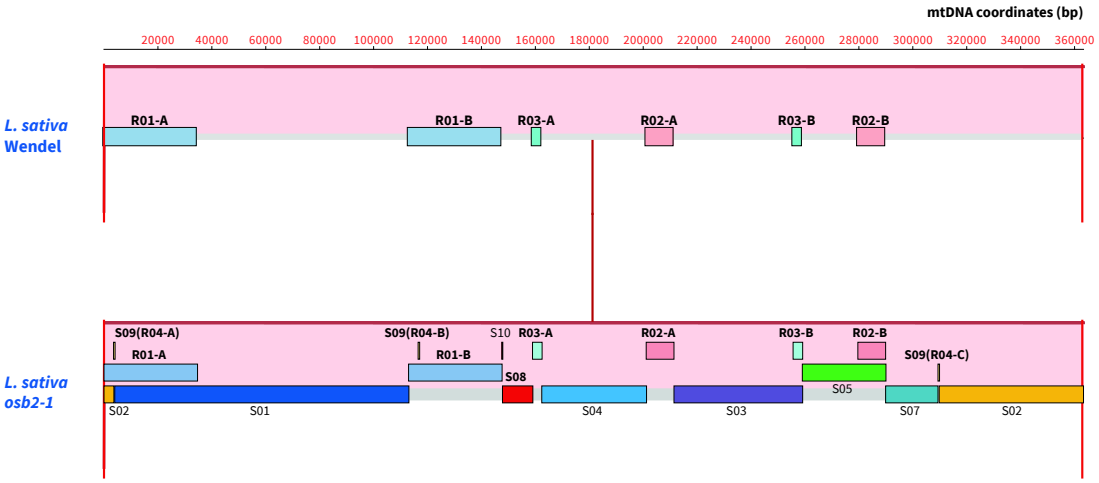


Figure VI.8. Structure of the *osb2-1* male-sterile mtDNA. Alignment made with progressive MAUVE. The *osb2-1* male-sterile mtDNA has the same structure as the WT (Wendel) mtDNA. The genomes are represented with the large repeated regions as well as the Illumina scaffolds (S-#) used to build the *L. sativa osb2-1* mtDNA.

1.4. Putative role of o*sb2-1 i*n the male-sterile phenotype

The main reason for us to focus on the *osb2-1* mutant was that one family carrying this mutation conserved its phenotype of male-sterility after backcrosses. Because *osb2-1* does not segregate at the homozygous stage we proposed that it could have a dominant effect on the WT allele. That could be because of the synthesis of a truncated toxic protein. To investigate that possible effect we studied the effect of *osb2-1* on two different systems

1.4.1. Expression of the *L. sativa osb2-1* allele in *A. thaliana*

To study the possible dominant effect of *L. sativa osb2-1* we introduced it in *A. thaliana*. To do so, we ordered the genomic sequences of *OSB2* and *osb2-1* (Genscript) and cloned them under their own promoting sequence (on destination Gateway vector pK7m24GW) or under the *UBQ10* gene promoter (on destination Gateway vector pUBC-GFP-Dest). These constructions were then used for the stable Agro-transformation of *A. thaliana* by floral-dip.

This part of our project is still ongoing. Indeed, we were able so far to select the first generation of mutants and we phenotyped them for male fertility. The *A. thaliana* plants transformed with either the WT *OSB2* gene or with the *osb2-1 allele* do not display any phenotype or an apparent male fertility defect. These transformants show a WT-like pollen production as well as normal seeds production by selfing (**Figure.VI.7**). This preliminary result suggests that *osb2-1* does not have a dominant effect on *A. thaliana* fertility.

1.4.2. Determination of possible changes in *osb2-1* mtDNA

In **CHAPTER 3**, we tried to determine the effect of the *osb2-1* heterozygous mutant on mtDNA stability. However, we were not able to see such an effect. There are two hypotheses to explain it, either the method we utilized does not have the resolution to see possible minor changes on the mtDNA, or the heterozygous mutation has no effect on the mtDNA. And then it would be difficult to link the male sterility phenotype to the *osb2-1* mutation.

To resolve the first issue (low resolution of the qPCR approach) we sequenced the mtDNA of *osb2-1* sterile heterozygous mutant through an Illumina MiSeq paired-end strategy (2 x 150 bases), managed by the IBMP platform *"Gene expression analysis"*. To do so, we purified mitochondria and mtDNA from flowers of heterozygous *osb2-1* plants, using the same protocol as for seedlings. The contigs were built according to the pipeline of analysis developed in **CHAPTER 2**. We obtained a set of 9 contigs that we manually assembled into a single circular chromosome of 363 324 bp. We

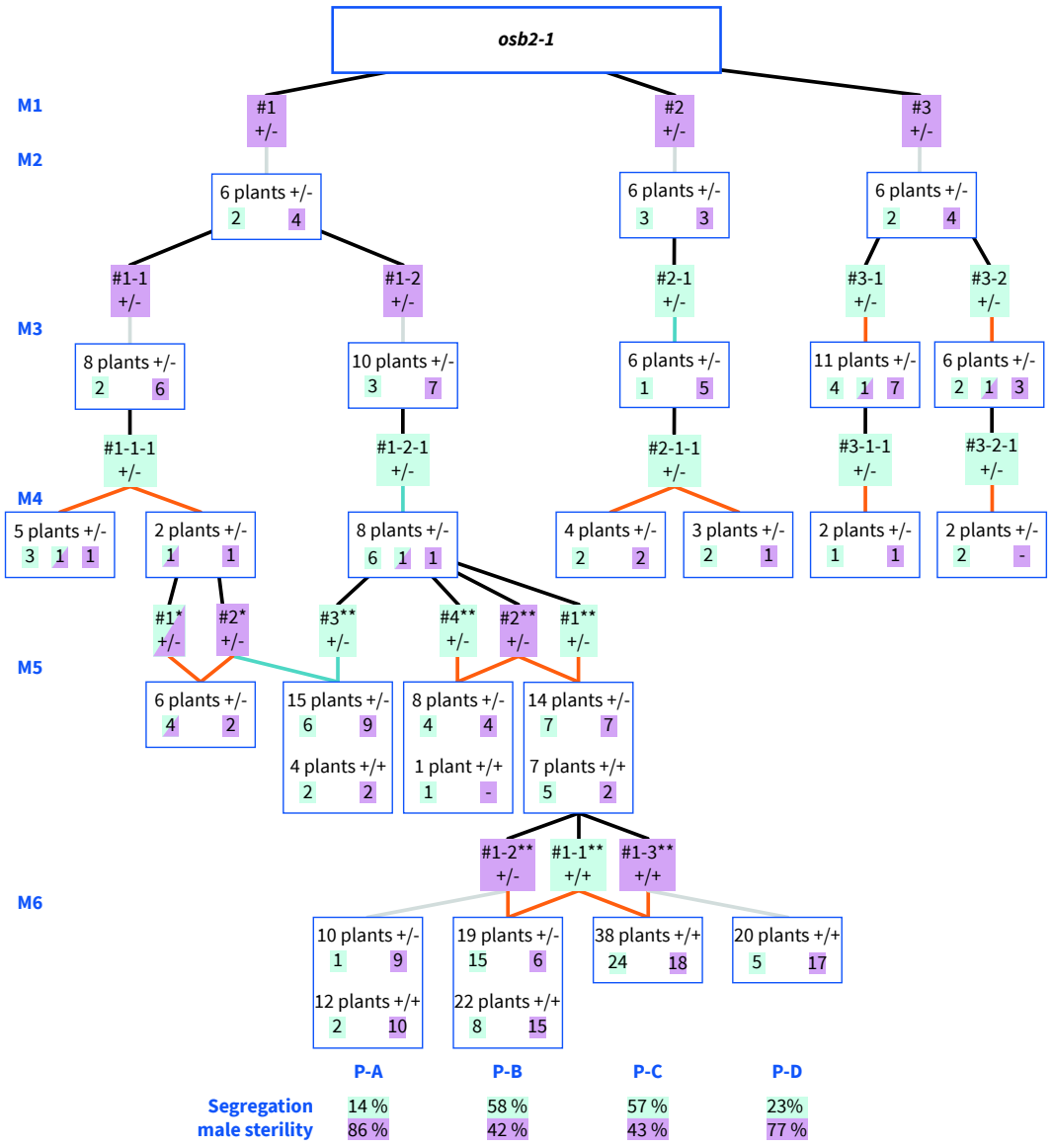


Figure VI.9. Segregation of the male-sterile phenotype within the progeny of *osb2-1*. The different plants from the *osb2-1* family used as parents are represented according to their *OSB2* genotype (Homozygous WT: +/+; heterozygous: +/-) and male-fertility phenotype (Sterile: pale turquoise; fertile: pale purple, half purple and turquoise: reduced male-fertility) for the six generations studied (M1 to M6). The crosses are represented by different colors regarding the type of cross: selfing is shown in grey, crosses with a sister plant are shown in orange and the crosses with a cousin plant are shown in turquoise. The black bars link a plant to the progeny where it comes from. For each cross, the segregation of the phenotype is shown in the blue box. For the ease of the reader, the plant names have been simplified starting from M4 (*: 1-1-1; **: 1-2-1). The segregation of the male-sterile phenotype has been evaluated in M6 and is shown under the different progenies.

aligned this sequence to the Wendel mtDNA (**Figure VI.8**). Besides conventional alternative configurations involving large repeated sequences, the two genomes are identical. We did not detect either any important change in mtDNA coverture. This result confirms what we obtained in CHAPTER 2 about the absence of evidence about an impact of the *osb2-1* heterozygous mutant on the mtDNA.

In short, the RT-PCR and RNAseq data showed no evidence for the accumulation of an aberrant *osb2-1* transcript retaining intron 1, that could code for a truncated *OSB2* protein, toxic for mitochondrial function. We could not exclude the possible existence of very minor transcripts resulting from re-initiation of transcription downstream exon 1. But the absence of visible fluorescence by expression of the gene-GFP fusion did not support such hypothesis. Neither the expression of the mutant gene in *A. thaliana.* And no evidence was found that heterozygous *osb2-1* has a negative effect on mtDNA stability. Consequently, we considered that the *osb2-1* heterozygous mutation is not likely at the origin of the sterility phenotype. As described below, we found indeed that it is different mutation, that could be segregated from *osb2-1.*

2. Identification of the origin of the male-sterile phenotype.

1.1. Genetic approach

During our project, we could study the *osb2-1* family over six generations (**Figure VI.9**). The plants belonging to the *osb2-1* family have been named #W-X-Y where W, X and Y and stand for their parents in the previous generation. For instance, plant#3-2-1 (third generation) is the plant number 1 from the progeny of #3-2 (second generation), itself plant number 2 from the progeny of #3 (**Figure VI.9**). For the ease of the reader, the plant names have been simplified starting from M4 (*: 1-1-1; **: 1-2-1). The roots of this family are three male fertile *osb2-1* heterozygous plants. The study of the male-sterile phenotype started in the selfing-progenies of these three plants, where the phenotypes segregated for the first time. Because we selected only the plants carrying *osb2-1*, we were not able to see the possible segregation of the phenotype from the *osb2-1* mutation. However, the molecular study that we conducted on *osb2-1* did not give any evidence linking the *osb2-1* mutation to the male-sterile phenotype. Consequently, we started the selection and the phenotyping of segregant WT plants for *OSB2* from the fifth generation on. Interestingly, we could

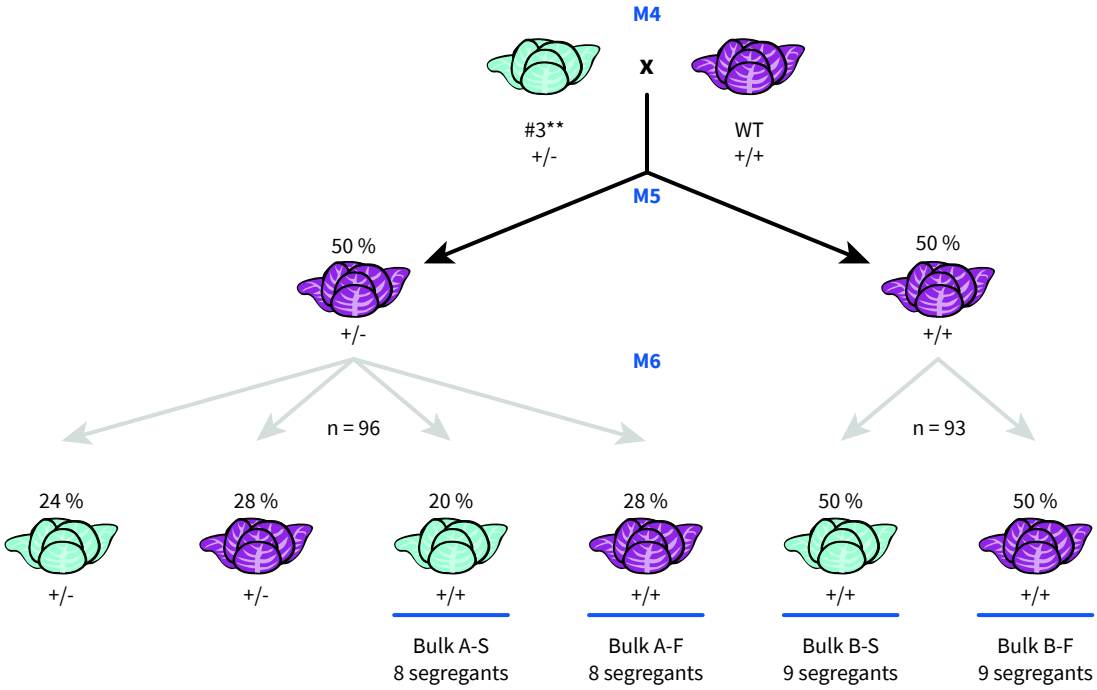


Figure VI.10 Bulked segregant RNA-seq strategy. A particular progeny has been studied to determine the origin of the phenotype of male-sterility, by bulked segregant RNA-seq (BSR-seq). The plants are represented according to their *OSB2* genotype (WT: +/+; heterozygous: +/-) and male-fertility phenotype (Sterile: pale turquoise; fertile: pale purple, half purple and turquoise: reduced male-fertility). The generations are as in **Figure VI.9**. The experimental segregation of the male-sterile phenotype within the progenies is shown. The bulks used for BSR-seq are underlined by a blue bar. (** = 1-2-1-)

observe the dissociation of *osb2-1* from the male-sterile phenotype. Therefore, the mutation leading to the male-sterile phenotype needed to be identified.

To do so, we studied four progenies in the sixth generation (**Figure VI.9**). The progenies A (P-A) and D (P-D) come from the selfing of a heterozygous *osb2-1* mutant (1-2**) or from a sister plant WT for *OSB2* (1-3**), respectively. The last two progenies, P-B and P-C result from the cross of these two plants, 1-2** and 1-3, and a sister male-sterile plant WT for *OSB2* (1-1**). The four progenies showed a segregation of the male-sterile phenotype (P-A = 14 %; P-B = 58 %; P-C = 57 %; P-D = 23 %) which could correspond to the segregation of a recessive GMS allele (½ male sterile in back crosses vs ¼ male sterile in selfings) (**Figure VI.2**). In such a case, 1-2** and 1-3** would be heterozygous for the male-sterile allele (MS) and 1-1** homozygous. The theoretical progenies should segregate the male-sterile phenotype at 25 % for P-A and P-D and 50 % for P-B and P-C, which is consistent with our experimental data. We confirmed this hypothesis by realizing a chi-square test on the four progenies (P-A: $\chi2$ (1, N = 22) = 1.52, p = 0.05; P-B: $\chi2$ (1, N = 44) = 0.90, p = 0.05; P-C: $\chi2$ (1, N = 42) = 0.86, p = 0.05; P-D: $\chi2$ (1, N = 23) = 0.06, p = 0.05). The chi-square test confirmed our hypothesis about the recessive monogenic origin of the male-sterile phenotype, but we could not exclude here, with such a segregation, that a CMS system with a dominant monogenic restorer fertility gene would be responsible for our trait of interest.

2.1. Transcriptomic approach: bulked segregant analysis

We proposed that the male-sterile phenotype that we observed in the *osb2-1* family is due to a recessive GMS. To identify the nuclear gene involved we realized a bulked segregant analysis (BSA). However, the large nuclear genome of lettuce (2.9 Gb) requires finding an alternative analysis. Consequently, we based our BSA on RNA sequencing. Such a strategy is called bulked segregant RNA-seq (BSR-seq). To do so, we based our BSR-seq on a branch of the family that is not represented in **Figure VI.9**. This branch starts with the cross of the male-sterile #3**, from M4, with a WT plant father (**Figure VI.10**). This cross leads to a progeny 100 % fertile in M5, with half of the plants carrying *osb2-1* at the heterozygous level, because of the genotype of the mother. The selfing-progeny of two M5-plants, one heterozygous and one WT for *osb2-1* was selected. Then, for both progenies, the plants were individually evaluated and flowers from different plants (but belonging to the same progeny) were bulked according to their phenotype male fertile or male sterile. The bulks from the progeny of the heterozygous *osb2-1* mother were composed of 8 flowers each (A-S and A-F) and the bulks for the *OSB2* WT mother were composed of 9 flowers each (B-S and B-F). Each flower came

from individual plants, meaning that the RNA extracted from these bulks was composed of the RNA pooled from 8 or 9 different segregants. The sampling was made in-field and thus the samples needed to be protected from RNA degradation with *RNALater* (Sigma-Aldrich). The sequencing of these four RNA bulks was managed by BGI GENOMICS through paired-end (2 x 100 bases) DNBseq.

We obtained many high-quality reads for each bulk (**Table VI.2**). We analyzed the data by mapping the reads on the *L. sativa* genomes (nuclear, cpDNA and mtDNA) and extracted the SNPs. Then we applied different filters to the SNP lists (**Table VI.2**). The SNP responsible for male sterility should be present in both sterile and fertile bulks. Because we were looking for a recessive allele, the fertile can be homozygous WT for this SNP, but also heterozygous with a mean frequency of 33.33 % (50% of total male fertile plants will segregate ⅓ WT against ⅔ heterozygous, so 50% of ⅔ leads to 33.33%), while the sterile bulk must be 100% homozygous mutant. Consequently, we set the frequency of the associated SNP in the fertile bulks lower than 51 % and superior to 99 % in the sterile bulks. After filtering we retained 48 SNPs for bulk A and 44 for bulk B. But because the mutagenesis was mediated by EMS, we filtered out all the non-C -> T or G -> A mutations. This led us to 15 and 14 remaining SNPs, in bulks A and B respectively, but only one was common to both. Moreover, a chi-square test showed that there were no significant differences between the frequencies observed and the ones expected (**Table VI.2**).

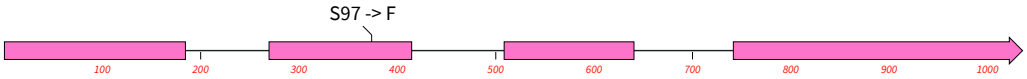
This SNP is located on chromosome 3, at position 374 after the ATG of the *caffeoyl CoA O-methyltransferase1* (*CCoAOMT1*) gene. This mutation is in the second exon of the gene and induces the substitution of the serine at the position 97 to a phenylalanine (S97>F) (**Figure VI.11**). This protein is extremely conserved in plants and shares few motifs with its functional homologs in fungi or the one that can be found in the arthropod *Folsomia candida* (**Figure VI.11**). The serine substituted in our GMS is conserved by all CCoAOMTs and seems to have a role in the *S-adenosyl-L-methionine* binding (Pakusch et al., 1989). This protein is described as the key methylation enzyme for caffeoyl-CoA during lignin biosynthesis in *A. thaliana* (Do et al., 2007). CCoAOMT1 has also be found to be expressed in the tapetum of young stamens, and linked to the pollen grain composition in spermidine phenylpropanoid conjugates (Fellenberg et al., 2008, 2012). Its possible role in GMS is discussed below.

To confirm that this SNP is responsible for the male-sterile phenotype, we initiated the genotyping of other mutants from the *osb2-1* family. The ideal would have been to genotype the SNP by HRM. However, we were not able to set the optimal condition to do it. Therefore, we did

Bulk	A-Sterile	A-Fertile	B-Sterile	B-Fertile
Expected frequency	100	33.33	100	33.33
Reads pairs	36 873 924	36 831 885	37 167 419	37 239 199
Total SNPs	157 267	165 215	142 672	153 429
Common (A or F)	102 524		92 254	
Freq in fertile < 51	22 011		18 374	
Freq in sterile > 99	48		44	
EMS mutation (C -> T or G -> A)	15		14	
Common to both bulks	**1**			
Genome	Nuclear			
Chromosome	3			
Postition (base)	134 804 317			
Gene	*CCOAOMT1*			
Mutation type	Missense mutation S97 ->F			
Observed frequency	100	36.46	100	26.53
Frequency chi-square test	$\chi 2$ (1) = 1.39, p = 0.05		$\chi 2$ (1) = 0.29, p = 0.05	

Table VI.2. Bulked segregant RNA-seq results

A.



B.

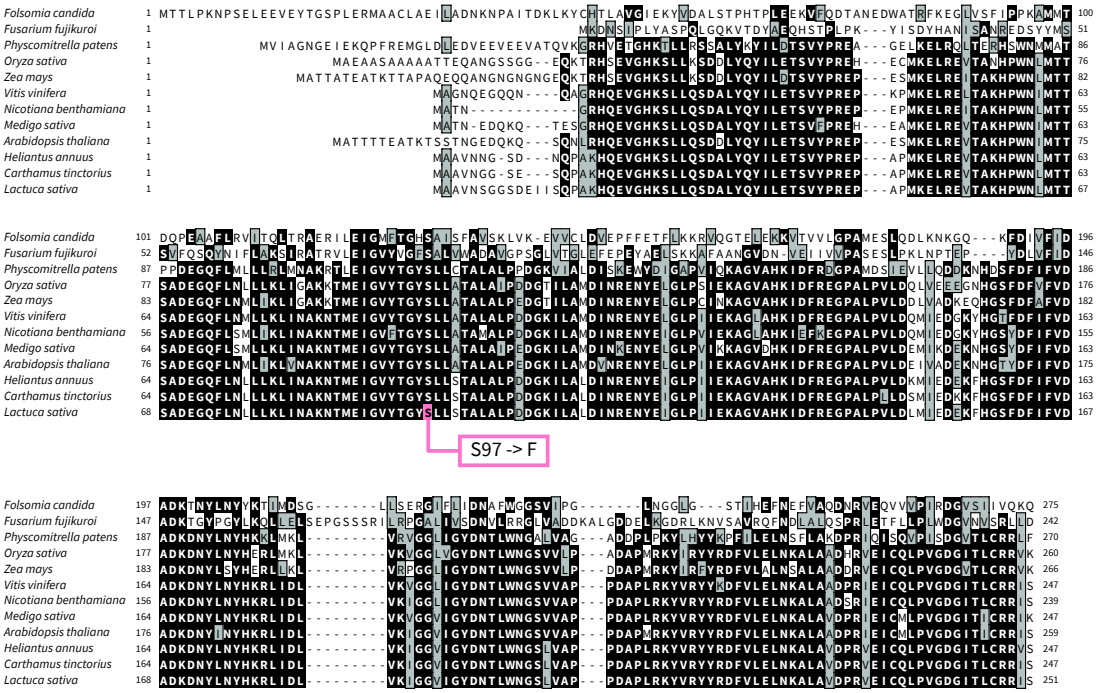


Figure VI.11. Alignment of *CCOAOMT1* protein sequences. (**A**) *CCOAOMT1* gene structure and localization of the SNP detected by BSR-seq (S97>F). (**B**) Sequence alignment (ClustalW) between representative *CCOAOMT1* sequences. The localization of the amino acid change presumed to induce the male-sterile phenotype is shown.

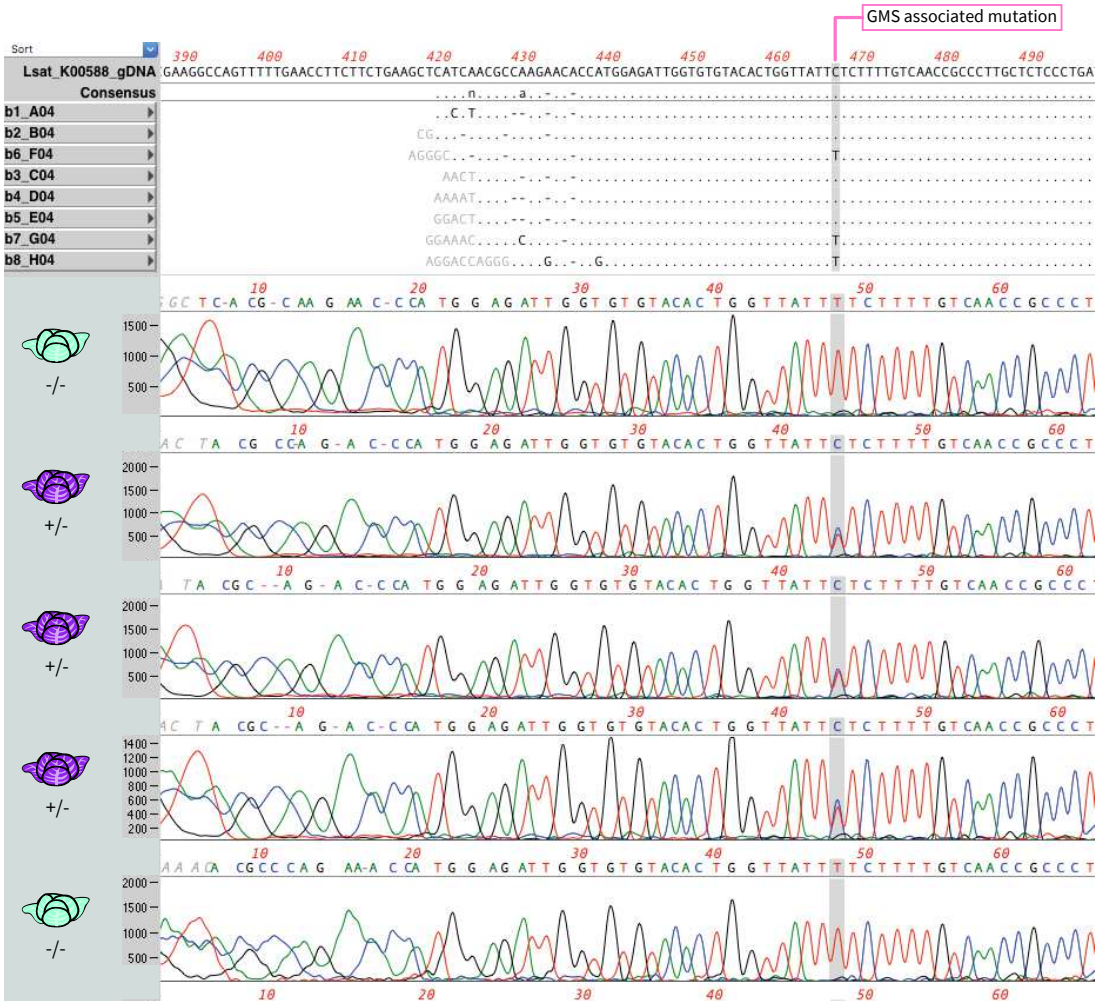


Figure VI.12. Genotyping of *CCoAOMT1* mutants. Snapshot from MacVector showing examples of sequence profiles on +/- and -/- plants (the phenotype of the related plant is shown on the left with the same color code as previously described). The localization of the mutation is indicated by the pink bar and highlighted in grey on the sequences.

genotyping by Sanger sequencing of PCR amplicons (**Figure VI.12**). Up to now, we genotyped 82 plants already phenotyped for male-sterility. Among these 82 plants, we phenotyped 38 sterile plants and 44 fertile ones. At one very exception (one false homozygous mutant), the genotype of homozygous mutation correlates with the sterility phenotype. Furthermore, we tested 43 fertile plants resulting from the cross between a male-sterile plant with a sister male fertile plant from the same seed lot, and as expected, all the 43 plants were heterozygous for the mutation.

DISCUSSION AND PERSPECTIVES

In the work described in this chapter we aimed to study the *osb2-1* mutant and to determine the origin of a male-sterile phenotype associated to the *osb2-1* line.

Our investigation on the *osb2-1* mutant showed that it produced fewer seeds than the WT, as expected for the embryo lethality of the (-/-) mutant. We also initiated the observation of the embryonic development, after selfing of an *osb2-1* (+/-) plant, on the technical advice of Anke Püschel from the cell biology department of Enza Zaden. It consisted of the bleaching of unmatured achenes at different times post-fecundation. But we did not investigate it further for lack of time.

We formulated the hypothesis that the mutation affecting the donor (5') splicing site of the first intron induces an aberrant transcript retaining the first exon. But the characterization of the *osb2-1* transcript by qRT-PCR and RNA-seq did not permit us to confirm that hypothesis and suggested instead that the transcript could be degraded by nonsense-mediated mRNA decay (NMD) (Chang and Kan, 1979; Losson and Lacroute, 1979). The level of expression of the other exons suggests that a possible reinitiation of transcription might be located upstream to the stop codon. To confirm this hypothesis as well as the aberrant peaks of coverage observed in the first exon and in the third intron, additional RNAseq runs should be performed for a statistical analysis. Because the role of *osb2-1* with respect to the male-sterile phenotype was already discarded, we did not investigate it further.

To explain the sterility phenotype associated with the *osb2-1* family we considered that the mutant allele could have a dominant effect. But if a dominant effect there is, it does not affect mtDNA stability. Furthermore, the *in vivo* studies by expression of the mutant *osb2-1* allele in *A. thaliana* suggested that by itself it is not detrimental to plant growth, pollen production or seeds

production by selfing. In *A. thaliana*, homozygous mutants for *OSBs* are not lethal, which is also true for the double mutants. Therefore, we could expect that the lack of *OSBs* would not be lethal neither, in L. sativa. Moreover, mutants for GIMMS that are severely affected like *recA2* in *A. thaliana* or *radA- 2* in *L. sativa* are not embryo lethal (Miller-Messmer et al., 2012). It was therefore surprising to find that the loss of OSB2 could be embryo lethal in *L. sativa.* In particular because *OSB1* is also targeted to mitochondria, and could theoretically complement the loss of OSB2. We should therefore conclude that OSB1 and OSB2 have differential functions in mitochondria, or that they are differentially expressed during development. It is also possible that the reason why an *OSB* gene is essential in *L. sativa* and not in *A. thaliana* is intrinsic to the mtDNA structure itself. Indeed, the *radA- 2* mutation of *L. sativa* showed us that it reproducibly induces the loss of a large mtDNA region containing essential genes, while in *A. thaliana* the loss of *radA* has no such dramatic effects, although it also severely affects mtDNA stability. It would be therefore interesting to introgress the *radA-2* and *osb2-1* mutations in the other *Lactuca* species, that have different organizations of the mtDNA, and test whether the mutations might be viable in a different mtDNA context.

We demonstrated that the *osb2-1* mutation cannot be the origin of the male-sterile phenotype. This was confirmed by a careful study of *osb2-1* and phenotypes segregation. This study oriented our research to a monogenic recessive GMS system that we identified by BSR-seq. The SNP identified which is associated to the phenotype induces an amino acid change at a highly conserved position of the protein CCoAOMT1. In *A. thaliana* this protein is mostly associated with lignin biosynthesis (Do et al., 2007). It is mainly expressed in the stems, but it has also been localized in the tapetum and the vascular bundles of the stigma and stamen connective tissue (Fellenberg et al., 2012). Moreover, the pollen of the mutant shows a reduction in N^1,N^{10}-bis-(5-hydroxyferuloyl)-N^5-sinapoyl-spermidine, a hydroxycinnamic acid amide (HCAA) that is a component of the pollen coat (Fellenberg et al., 2012). These compounds are essential for pollen viability, since in *A. thaliana* the *ms1* mutation that induces a gametophytic male-sterility with non-viable pollen displays an aberrant content in HCAA in its coat (Grienenberger et al., 2009). *CCoAOMT1* has also been shown to be downregulated in male-sterile citrus where the pollen interruption of pollen development has been links to a CoA deficiency. Finally, *CCoAOMT1* has been shown to be down-regulated in CMS anthers/buds while compared to the restorer line, in pepper (Wu et al., 2013). The literature therefore strongly supports our assumption that the mutant for *CCoAOMT1* is responsible for the male-sterile phenotype associated to the *osb2-1 line* progeny. However, at present we cannot be

sure that this gene is not one of the six other recessive GMS alleles already identified in *L. sativa,* because of the confidentiality surrounding them.

We used in this project specific and accurate procedures of phenotyping, in combination with targeted plant to plant crossing schemes which allowed us to obtain such a result. This was possible by the expertise and dedication of the scientist at the Allonnes station of Enza Zaden.

Conclusion & perspectives

My Ph.D. project brought together the breeding company Enza Zaden represented by Mme Emilie Guilloteau-Fonteny and the IBMP group "Maintenance and segregation of the mitochondrial genome" led by Dr. José Manuel Gualberto. The aim of this collaboration was to study mitochondrial diversity in lettuce.

The prerequisite for such a study was to build a reference genome (discussed in CHAPTER 2). That is why we started this project by the assembly of the mtDNA of *Lactuca sativa*, the domesticated lettuce. This part of the project beneficiated from the support of Enza Zaden by access to a PacBio platform. We found that the mitochondrial genome of *L. sativa* can be carried by three circular chromosomes that can theoretically be assembled into a single circular chromosome of 363 324 bp. The deposit of this sequence in Genbank would require a minimal work of annotation, including the annotation of the protein sequences taking in consideration the RNA-DNA changes by RNA editing. For that Pr. Jean-Stephane Varré (University of Lille) kindly provided us his script used for the annotation of the potato mtDNA.

The first aspect of the mtDNA diversity that we studied was the natural diversity existing in wild lettuce species (discussed in CHAPTER 3). To do so, we built the mtDNA of 9 wild accessions: 1 from *L. saligna;* 2 from *L. sativa*, 3 from *L. serriola* and 3 from *L. virosa*. All these genomes displayed the same gene content and a conserved set of repeated sequences at the very exception of the one from *L. serriola*. Moreover, we found chloroplastic insertions resulting from horizontal transfer that suggest that Lettuce Crisphead result from a cross involving an *L. virosa* mother. But this hypothesis is not supported by the usual chloroplastic markers supporting a more ancient lineage between the two species. Finally, the genomes that we built result from a final step of assembly by hand, of contigs sharing overlapping sequences, and could benefit from experimental confirmations by PCR of the scaffold borders. Such work is necessary before publication of our sequences, and for confirmation of our observations regarding the structural diversity of the mtDNA in *Lactuca spp.*.

The second aspect that I developed during my PhD project consisted in the induction of diversity within the mtDNA of *L. sativa.* The plan was to increase the probability of recombination events potentially leading to CMS (discussed in CHAPTER 4). To induce it, we tested two strategies.

The first one was a pilot experiment, to test the increase in mtDNA recombination resulting from the repair of DNA breaks induced by ciprofloxacin, an organellar genotoxic compound known to directly induce mtDNA diversity. We found that *L. sativa* is more resistant to this compound than *A. thaliana.* Moreover, our preliminary work showed that indeed ciprofloxacin has an impact on the diversity of the mtDNA, but at high concentrations that are detrimental to plant development. We could detect a plant with reduced pollen viability correlating with a reduced production of seeds. However, such defect was not observed in its progeny. This part of the project was managed at the IBMP, at a low scale, for proof of concept. The next step would be to upscale it to a higher population of treated plants, in the hope of inducing a CMS system from a ciprofloxacin-treated population.

The second strategy was an indirect one and relied on mutants for genes involved in mitochondrial genome maintenance and segregation (also discussed in CHAPTER 4). To do so we selected 8 genes and got 13 interesting mutants for 6 of them, namely genes *MSH1*, *OSB1*, *OSB2*, *RADA*, *RECA3,* and *RECG1*. We did find male sterility and mtDNA induced diversity within our population, but unlinked: associated to the *osb2-1* mutant line we found male-sterile plants, but no rearrangement of the mtDNA by recombination could be correlated to the *osb2-1* mutation; associated to *radA-2* we found important mtDNA instability by recombination, but the mutant was not able to produce flowers and it was therefore impossible to phenotype the plants for male sterility. Other mutants were completely sterile, such as *msh1-2* that was not able to produce seeds, even if outcrossed with pollen from a male-fertile plant. Consequently, we decided to focus on mutants *osb2-1* and *radA-2*. By doing so we might have missed other interesting mutants, but the seeds of those mutants (monogene homozygous, double homozygous mutants and heterozygous double and triple mutants) are available, if Enza Zaden decides to continue the project.

We focused the study in the *radA-2* mutant, that apparently is a knockout mutant because of a stop codon in the third part of the coding sequence (discussed in CHAPTER 5). This mutant shows a severe phenotype at the homozygous stage that we could link to a strong reduction in the stoichiometry of a region of the mtDNA comprised between the repeated sequences R03 and R04. We confirmed that two single copy genes comprised in this region, *ccmC* and *rps4,* are much reduced in expression, as evaluated by the reduced accumulation of their transcripts. Two other genes (*nad9* and *trnW(CCA)* comprised in that mtDNA regions remain to be tested. But it is already possible to conclude that it is the reduced expression of essential mitochondrial genes encoded by the region deleted by recombination that is responsible for the sever growth defects of the *radA-2* mutants.

The aim of the study of the *osb2-1* mutant line was to determine the determinant of the male-sterile phenotype segregating in that line. We also wanted to understand why the homozygous mutant was apparently not viable (discussed in CHAPTER 6). The *osb2-1* mutation affects the donor (5') splicing site of the first intron of the *OSB2* gene. We confirmed that the *osb2-1* heterozygous mutant produces significantly fewer seeds than the WT, which suggested an embryo lethality of the homozygous mutation. A study of the embryonic development within the selfing progeny of an *osb2-1* heterozygous plant should confirm it. A hypothesis for such the lethality could be the expression of an aberrant transcript coding for a toxic truncated protein. But RT-qPCR and RNAseq data excluded the accumulation of the transcripts retaining the first intron sequence. In fact, the RNAseq data rather suggests that the *osb2-1* transcript is degraded, which could lead to transcription reinitiating downstream of the first exon. Confirmation of any hypothesis, including of alternative spliced transcripts will require additional RNAseq runs as replicates.

With respect to the male-sterile phenotype, we determined that it was associated to a recessive GMS caused by a SNP mutation within the *CCoAOMT1* gene, which according to the literature can been linked to the failure to produce viable pollen, because the enzyme is required for the biosynthesis of hydroxycinnamic acid spermidine conjugates of the pollen coat. We confirmed the link between the mutation in this gene and the male-sterile phenotype, for some plants *a posteriori* (first phenotyped, then genotyped). But it will be important to confirm it within an *osb2-1* progeny and validate it as a marker for the screening of the trait *a priori* (first genotype, then phenotype), for possible breeding purposes.

Materials and methods

MATERIALS

1. Bacterial strains

1.1. Agrobacterium tumefaciens GV3101 pMP90

This nopaline-type phytopathogenic strain of *A. tumefaciens* derives from the C58C1 strain and had been used for the stable transformation of *Arabidopsis thaliana*. This bacterial strain is resistant to rifampicin and carry the disarmed Ti plasmid pMP90 giving resistance to gentamicin. This plasmid carries the virulence genes needed to transfer the T-DNA region from a binary vector into the plant cell.

1.2. Escherichia coli TOP10

This *E. coli* strain has been used for cloning and plasmid amplification. It allows stable replication of high-copy number plasmids because the strain is deficient for recA1, avoiding non-specific recombination of cloned DNA. In addition, this strain allows alpha-complementation and white-blue colony selection of recombinant clones on media containing IPTG and X-gal.

Genotype: F⁻*mcrA* Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74*recA1 ara*D139 Δ(*ara leu*) 7697 *galU galK rpsL* (StrR) *end*A1 *nup*G

2. Plant lines

2.1. Lactuca spp.

2.1.1. *Lactuca sativa* Enza Zaden old commercial variety “Wendel”

Domesticated lettuce, *L. sativa* was the main subject of this work to study the impact of MSH1, OSB1, OSB2, RADA, RECA3 and RECG1 on its mitochondrial genome (mtDNA). Seeds from the old commercial variety “Wendel” were provided by Enza Zaden. Plants were cultivated in the greenhouses of the IBMP under a day/night cycle of 16 hours of light at 21 °C and 8 hours in the dark at 18 °C. Plants grown *in vitro were also* under the same long day light cycle, but powered by neon lights (OSRAM, biolux T5).

Species	Geographical group	GB no.	Lot no.	Variety name	Supplier	Country of origin
L. saligna	Caucase	LAC 008020-03	2016NL	LAC 08020	Collected	AZE
L. sativa	Med. Basin	LAC 004500-02/03	2009NL/2008NL	LAC 04500	Collected	GRC
L. sativa	North America	LAC 011481-02	2016NL	UC 12100	UCAL	USA
L. serriola	Oceania	LAC 005780-03	2010NL	LAC 05780	Collected	NZL
L. serriola	Central Asia	LAC 005418-02	2007NL	CGN 004799	CGN	IRQ
L. serriola	Caucase	LAC 009042-03	2015NL	CGN 130219	CGN	ARM
L. virosa	Europe	LAC 006941-02	2010NL	IVT280 x CGN16273		NLD
L. virosa	Med. Basin	LAC 004015-05	2007NL	CGN 013357	CGN	ESP
L. virosa	Med. Basin	LAC 004030-04	2007NL	CGN 019045	CGN	ITA

Table MM.1. Origins of the selected wild accessions of lettuce.

2.1.2. *Lactuca sativa* EMS mutant lines

A total of 20 mutant lines (**Table IV.4**) for the six genes of interest (*MSH1*, *OSB1*, *OSB2*, *RADA*, *RECA3* and *RECG1*) were obtained and provided by Enza Zaden. Mutant lines have been generated by ethyl methanesulfonate (EMS) mutagenesis on Wendel and identified by TILLING.

2.1.3. Wild lettuce species

A collection of 9 accessions (**Table MM.1**) belonging to 3 different wild lettuce species (*Lactuca saligna*, *Lactuca serriola* and *Lactuca virosa*) originating from diverse geographical origins (Caucase, Central Asia, the Mediterranean basin, Europe, North America and Oceania) were used to study the natural diversity of the mtDNA within the *Lactuca sp*.. The seeds of these 9 accessions were provided by Enza Zaden.

2.2. Nicotiana spp.

2.2.1. Nicotiana benthamiana

N. benthamiana plants were used for subcellular localization of eGFP fusion proteins by biolistic transfection or agroinfiltration of mesophyll cells. Plants were provided by the plant-growth platform of the IBMP.

2.2.2. Nicotiana tabaccum

A mutant of *N. tabaccum* constitutively expressing a mitochondria-targeted eqFP red-fluorescent protein (Forner and Binder, 2007) was used for confirmation of the mitochondrial targeting of eGFP fusion proteins by biolistic transfection of mesophyll cells (a kind gift of Dr. Joachim Forner).

2.3. Arabidopsis thaliana

A. thaliana ecotype Columbia (Col-0) was used to study the possible dominant effect of the *L. sativa osb2-1* mutation, by transformation with *A. tumefaciens*. The Arabidopsis *OSB2* (At1g31010) mutant by T-DNA insertion *osb2-1* (SALK_018261) derives from the Col-0 accession. This mutant line was provided by the European Arabidopsis Stock Center (NASC).

Name	Concentration	Purpose
Ampicilin	100 µg/ml	Bacterial selection
Carbenicilin	25 µg/mL	*A. thaliana* selection
Cefotaxime	25 µg/mL	*A. thaliana* selection
Ciprofloxacin	1-2000 µM	Genotoxic stress on plants
Gentamicin	20 µg/ml	Bacterial selection
Hygromycin	15 µg/mL	*A. thaliana* selection
Kanamycin	50 µg/mL	Bacterial selection
Rifampicin	30 µg/mL	Bacterial selection
Streptomycin	50 µg/ml	Bacterial selection

Table MM.2. Antibiotics used in this project.

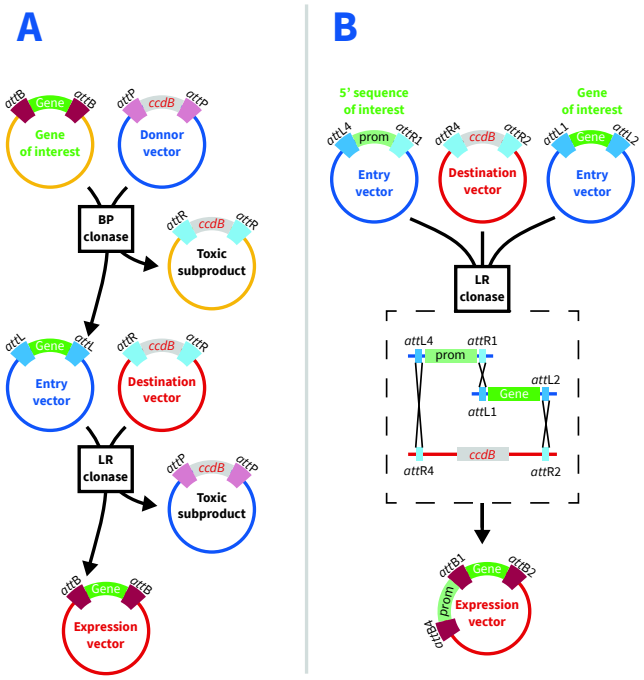


Figure MM.1. Gateway cloning principles. (**A**) Principles of single-site Gateway cloning. Recombination events between *attB* sites (burgundy, ■) and *attP* sites (light burgundy, ■) are mediated by the BP clonase. This recombination leads to a vector carrying the toxic *ccdB* gene and the entry vector carrying the gene of interest between *attL* sites. Recombination events between *attL* sites (blue, ■) and *attR* sites in a destination vector (light blue, ■) are mediated by the LR clonase. This recombination leads to a vector carrying the toxic gene and the expression vector carrying the gene of interest. (**B**) Principles of a multiSite Gateway cloning. Recombination events between *attL1* sites (blue, ■) and *attR1* sites (light blue, ■); *attL2* sites (blue, ■) and *attR2* sites (light blue, ■); *attL4* sites (blue, ■) and *attR4* sites (light blue, ■) are mediated by the LR clonase. This recombination leads to the expression vector carrying the gene of interest with the 5' sequence of interest.

3. Growth media

3.1. LB medium

Lysogeny broth medium was used for bacterial growth (tryptone 1 % (p/v); yeast extract 0.5 % (p/v); NaCl 0.5 % (p/v); pH 7.4). This medium can be complemented with antibiotics for bacterial selection (**Table MM.2**) and solidified by adding 1.5 % (p/v) agar.

3.2. MS media

Murashige and Skoog media (MS255) were used for *in vitro* growth of plants. It contains micro and macro elements M0255 (DUCHEFA BIOCHEMIE B.V., Haarlem, The Netherlands), sucrose 1 % (p/v) and pH adjusted to 5.8 and can be complemented with antibiotics for transgene selection. It can be used as liquid medium but it was mainly used as a solid medium by adding 1 % agar (p/v) (**Table MM.2**).

4. Plasmid vectors

4.1. Gateway vectors

This technique, based on site-specific recombination properties of *lambda phage*, requires Gateway *att* recombination sites and a recombination enzyme mix (BP and LR clonases) (**Figure MM.1**). The recombination between sites *attB* and *attP* will generate an entry vector with the desired sequence. Then, this sequence can be transferred to multiple destination vectors by a new recombination reaction between sites *attL* and *attR*. All Gateway plasmids carry a chloramphenicol resistance gene and the *ccdB* gene, encoding for a toxic protein, for negative selection.

4.1.1. pDONR207

The plasmid pDONR207 is an entry vector carrying the *attP1* and *attP2* recombination sites and a gentamicin resistance gene.

4.1.2. pB7FWG2

The plasmid pB7FWG2 is a destination binary vector (**Figure MM.1.A**) carrying the *attR1* and *attR2* recombination sites, the 35S promoter sequence of the *Cauliflower* mosaic virus (CaMV) and the *eGFP* gene. These last two items, from either side of the cloning site, allow to C-terminal fuse a

sequence of interest to *eGFP,* to determine its subcellular localization after biolistic transfection or agro-infiltration. Bacterial transformants can be selected thanks to its streptomycin resistance gene.

4.1.3. pK7m24GW

The plasmid pK7m24GW is a MultiSite destination binary vector (**Figure MM.1.B**) carrying the recombination sites *attR2* and *attR4*, and the 35S terminator. This configuration allows to test in this vector multiple promoting sequences for a gene. The vector is carrying two resistance genes for transformant selection purpose: the streptomycin resistance gene allows to select bacterial transformants and the kanamycin one allows to select transformed plants.

4.1.4. pUBC-Dest

This binary vector derives from the pB7WGR2 vector (Grefen et al., 2010) and is used to clone a gene under the control of a ubiquitin (*UBQ10* gene) promoter sequence and fused to the *GFP* sequence at his C-terminal part. The gene sequence is inserted through the *attR1* and *attR2* recombination sites. Like pK7m24GW, pB7WGR2 is carrying two resistance genes for selection: the streptomycin resistance gene for the selection of bacterial transformants and the *bar* gene conferring resistance to the herbicide Basta (containing glufosinate for the selection of transformed plants).

4.1.5. pGEM-T Easy

The pGEM-T Easy vector (Promega®, Madison, Wisconsin, USA) is marketed linearized for cloning of PCR products. Its 3'-terminal thymidine-overhangs allows the insertion of the PCR product containing adenosine overhangs, added by classic *Taq* polymerase. This vector carries an ampicillin resistance gene and the promoter sequences of the phagic RNA polymerases T7 and SP2 flanking the cloning region, in the β-galactosidase gene, allowing identification of recombinants by blue/white screening on plates containing X-gal and IPTG.

4.2. pUC57

The plasmid pUC57 is the vector containing the sequences ordered at Genscript (Piscataway, New-Jersey, USA) and carrying an ampicillin resistance gene.

4.3. pUCAP35S

This plasmid was built within the institute based on pUCAP and is used for cloning genes fused by the C-terminal part to *eGFP,* and under the control of a double 35S promoter. The constructs are transfected into plant cells by biolistic transformation, to determine the subcellular localization of the expressed fusion proteins. Transformants can be selected thanks to the ampicillin resistance gene carried by pUCAP35S.

4.4. pCK-*COX4*-GFP3

The plasmid pCK-*COX4*-GFP3 was used as positive control for subcellular localization. It codes GFP fused at its N-terminus to the presequence of the yeast COX4 protein, for targeting into the mitochondria (Menand et al., 1998). As in pUCAP35S, the construction is under the control of two copies of the CaMV 35S promoter, allowing overexpression of the protein and easy visualization by fluorescent microscopy. Bacteria carrying pCK-*COX4*-GFP3 can be selected thanks to the ampicillin resistance gene.

5. Synthetic gene

Four synthetic sequences have been built by Genscript and cloned into pUC57:

- *OSB2* promoter sequence (1000 bp upstream the start codon of OSB2) flanked by Gateway recombination sites *attL4* and *attR1*.
- *OSB2* WT complete sequence flanked by Gateway recombination sites *attL1* and *attL2*.
- *osb2-1* mutant complete sequence flanked by Gateway recombination sites *attL1* and *attL2*.
- *osb2-1* mutant sequence without STOP codon flanked by Gateway recombination sites *attL1* and *attL2*.

6. Oligonucleotide

The sequences of all primers used into this study for PCR, qPCR or HRM are detailed in the annexes.

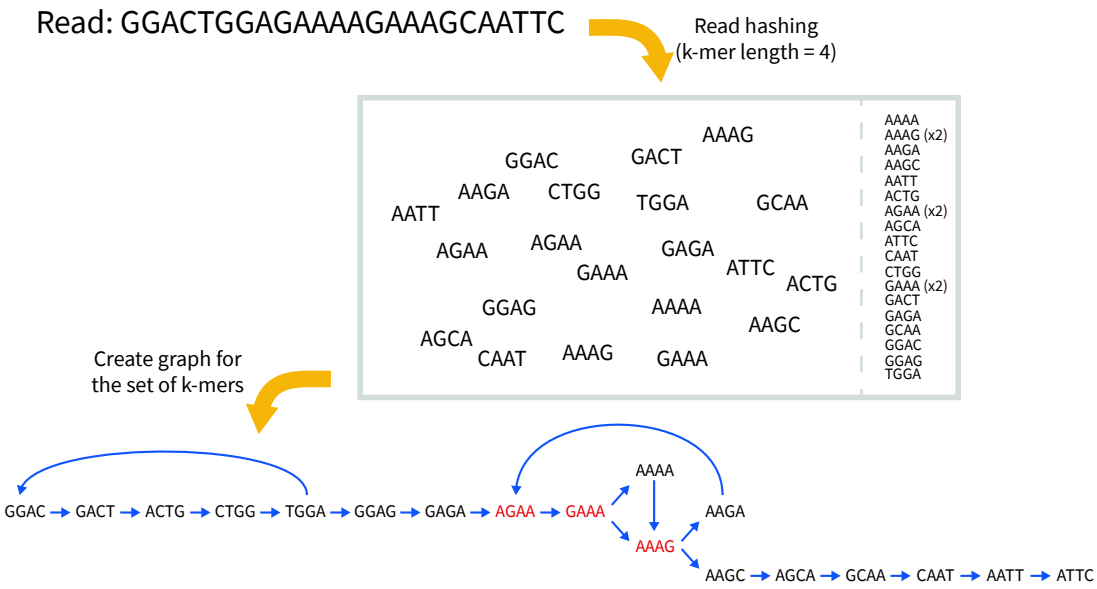


Figure MM.2. Construction of a de Bruijn graph for 24 nucleotides-long reads. To build a de Bruijn graph, reads are first hashed according to the k-mer setting(s) and put in the hash table. K-mers are then sorted and repeated ones are flagged. Finally the de Bruijn graph is built according to the hash table and all the pieces that overlap for k-1 (3 in this example) are linked. The final path needs to go through every piece only once, at the very exception of repeated ones (red, AGAA, GAAA and AAAG) where the path must go through them twice.

7. Informatic tools

7.1. Subcellular localization prediction tools

Prediction of the subcellular localization of proteins has been made through *Predotar* (urgi.versailles.inra.fr/predotar/) and *TargetP2.0* (www.cbs.dtu.dk/services/TargetP/) (Small et al., 2004; Almagro Armenteros et al., 2019).

7.2. Databases

NCBI tools and databases have been used for bibliographic purposes (PubMed; www.ncbi.nlm.nih.gov/pmc) and research of similarity between biological sequences (BLAST; www.blast.ncbi.nlm.nih.gov/Blast.cgi).

For researches specific to the *Lactuca sativa* genome (V8) the plant genomic resource Phytozome (www.phytozome.jgi.doe.gov) has been used.

The RNAseq data used for mitochondrial annotation are available on the database of the European Nucleotide Archive (www.ebi.ac.uk/ena) under the accession number SRR080725 and SRR085107.

7.3. Command line tools

7.3.1. Trimmomatic (v0.39)

Trimmomatic (Bolger et al., 2014) performs a variety of useful trimming tasks for both Illumina paired-end and single ended data. In this work it has been used on paired-end reads to remove Illumina adapters and low quality nucleotides. On paired-end reads, Trimmomatic gives as output files the paired reads in two separated files (R1 and R2) but also the unpaired reads in two separated files (R1U for forward unpaired and R2U for reverse unpaired).

7.3.2. Fastq-join (v1.3.1)

Fastq-join joins two paired reads that overlap.

7.3.3. Velvet (v1.2.10)

Velvet (Zerbino and Birney, 2008) is a *de novo* sequence assembler for short reads using the de Bruijn graph. To build the de Bruijn graph (**Figure MM.2**), velvet requires the setting of the k-mer

length (k). The k-mers are a subdivision of a read of length k. The de Bruijn graph is a directed graph representing all overlaps of the k-mers for k-1. To do the contigs assembly Velvet works in two steps: first Velveth for hashing the reads, and then the command Velvetg builds the de Bruijn graph, runs simplification and error corrections and finally extracts the contigs.

7.3.4. Burrows-Wheeler Aligner (BWA) (v0.7.17-r1188)

BWA (Li and Durbin, 2009) is a software package for the mapping of reads on a large reference genome. Its algorithm BWA-MEM, specialized in the mapping of reads from 70 bp to 1 Mbp, was used in this study to perform alignments. The algorithm works by seeding alignments with maximal exact matches (MEMs) and then extending seeds.

7.3.5. Hisat2 (v2.1.0)

Hisat2 (Kim et al., 2015) is a fast and sensitive aligner software that can be used on low configuration. It was used for the mapping of reads from transcriptomic data because its parameters can be easily adjusted to increase the number of mismatches.

7.3.6. SPAdes (v3.13.1)

SPAdes (Bankevich et al., 2012) is, as well as Velvet, a de Bruijn graph-based assembler. The difference between both assemblers is that SPAdes allows multiple k-mer settings and build as many de Bruijn graph as k-mers set and extracts contigs based on the information of all graphs.

7.3.7. Samtools (v1.9)

Samtools is a set of utilities to manipulate alignments in the Sequence Alignment Map (SAM) format, convert it into the Binary Alignment Map (BAM) format and even extract aligned sequences in FASTA or FASTQ format.

7.3.8. SSPACE (v1.12)

SSPACE (Boetzer et al., 2010) is a scaffolder able to extend pre-assembled contigs using one or more set of reads. This scaffold uses Bowtie to map the reads on the ends of the contigs and then extend them.

7.3.9. IDBA-UD (v1.1.3)

IDBA-UD (Peng et al., 2010) is, as well as Velvet and SPAdes, a de Bruijn graph-based assembler. Its particularity lies in the fact that it can assemble data from short reads sequencing with highly uneven sequencing depth. That’s why it is mostly used in metagenomic.

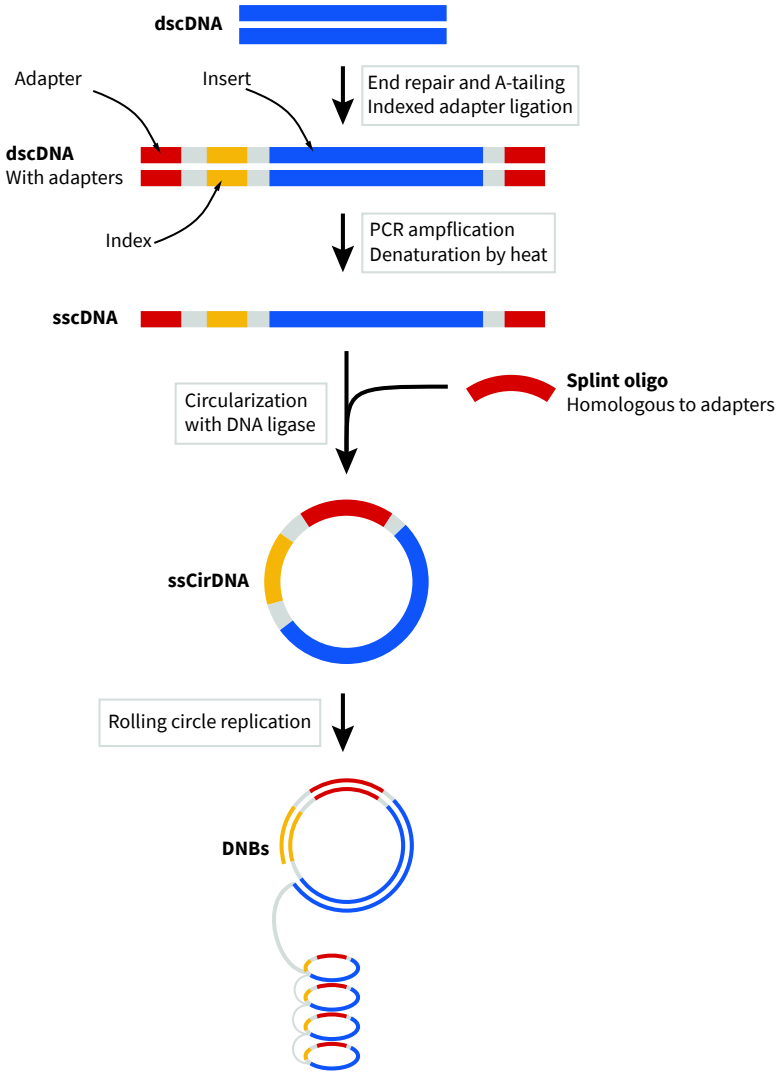


Figure MM.3 Library construction for DNB sequencing. Fragmented double-stranded cDNAq (dscDNA, blue, ▮) are repaired and A-tailed to be linked with adapter sequences (red, ▮) and index sequence (yellow, ▮). The dscDNA-adapters-index complex is then purified, amplified by PCR using primers specific to the adapters sequences and heat-denatured into a single-stranded cDNA (sscDNA). A splint oligo, homologous to adapter sequences, is next used to circularize the sscDNA into a single-stranded circular DNA (ssCirDNA). Finally, this ssCirDNA is used as a base for rolling-circle replication leading to the DNA nanoballs (DNB) that will be loaded on the BGISEQ-500 platform.

7.3.10. NOVOplasty (v2.4)

NOVOplasty (Dierckxsens et al., 2016) is an assembler for short circular genomes. To do so, it requires a close genome as a reference and a seed sequence. Then reads are hashed according to the k-mer parameter and the seed sequence is used to search for a region that matches the start of a read. Once this match is found, extension follows the same process. The positions of the paired reads are next verified by aligning each paired read to the assembled sequence. Finally, a consensus sequence is extracted.

7.3.11. Canu (v1.4)

Canu is an assembler specialized in long reads sequences (PacBio Single Molecule, Real-Time (SMRT) or Oxford Nanopore Technology (ONT) operating in three steps: First correction, to improve the accuracy of bases and so the reads quality. Then the reads are trimmed to the portion that appears to be high-quality sequence, and adapters are removed. Finally, trimmed reads are ordered into contigs.

7.3.12. Freebayes (v1.3.1-dirty)

Freebayes (Garrison and Gabor, 2012) is a genetic variant detector based on Bayesian statistics. It was used in this study to detect single nucleotide polymorphisms (SNP) from RNA sequencing reads mapped to mitochondrial genomes.

7.3.13. VarScan (v2.3)

VarScan (Koboldt et al., 2012)is a mutation caller for resequencing data used to detect SNPs from RNA-seq reads mapped from pools of samples.

7.3.14. Data.table (v.1.12.6)

Data.table is an R package specialized into the manipulation of data stored in table files and was used for the processing of data outputted from VarScan.

7.4. Visualization tools

Comparisons, sequence alignments, *in silico* of cloning have been made on *MacVector* (MacVecor INC., Apex, North Carolina, USA).

Alignments between lettuce mitochondrial genomes have been made with *MAUVE* (Darling et al., 2010)

Intra and interspecific phylogenetic trees regarding lettuce mitochondrial, plastid or protein sequences as well a protein has been made with *MEGAX* (Kumar et al., 2018; Stecher et al., 2020).

Aligned reads from both DNA and RNA sequencing had been visualized on Integrative Genomics Viewer or IGV (Robinson *et al.*, 2011)(Robinson et al., 2011).

METHODS

1. Nucleic acids methods

Quantification

1.1.1. Spectrophotometric quantification

Nucleic acids are quantified by measuring the absorbance at 260 nm (A260) through the spectrophotometer *NanoDrop 2000* (Thermo Fisher Scientific, Waltham, Massachusetts, USA). Ratios A260/A280 and A260/A230 quantify the contamination by protein, phenol or other contaminants that absorb strongly at or near 280 nm or 230 nm. A nucleic acid solution is accepted as pure if the ratio A260/A280, for DNA ~1.8; for RNA ~2.0; and the ratio A260/A230 is in the range of 2.0-2.2.

1.1.2. Fluorometric quantitation

For sequencing purposes, most of the platforms require a strict quantification by *Qubit 4 Fluorometer* (Thermo Fisher Scientific). This technique is more sensitive than UV absorbance thanks to the use of fluorometer dyes specific to DNA or RNA that emit only when bound to the target molecules.

1.2. Agarose gel electrophoresis

Agarose gel electrophoresis allows to separate nucleic acids through a matrix made of agarose according to their size or structure. The gel is made of 0.7 % to 1.7 % agarose (p/v) in Tris-acetate-EDTA buffer (TAE 0.5 X: 20 mM Tris-acetate pH 8.0; 0.5 mM EDTA). Nucleic acid samples are loaded after adding one-sixth of the volume of loading buffer (loading buffer 6X: 10 mM Tris-HCl pH

7.6; bromophenol blue 0.03 % (p/v); glycerol 60 %; 60 mM EDTA) and revealed, after 30 min of migration at 100 V, thanks to the ethidium bromide (EtBr) present in the gel (0.5 µg/mL EtBr). EtBr is an intercalating agent of nucleic acid observable under UV light. Finally, the size of nucleic acids is compared to a size marker, *MassRuler DNA Ladder* (Thermo Fisher Scientific) that contains 20 fragments of known sizes (80 bp to 10000 bp).

2. RNA methods

2.1. Preservation of samples until extraction

To transport samples for RNA extraction from the company to the laboratory, a solution to block RNases from degrading the RNA is needed. To do so, the *RNALater* (Sigma-Aldrich, St. Louis, Missouri, USA) storage reagent was used. Briefly, all the non-green parts (petals, carpels and stamens) of 10 lettuce flowers are completely immersed in 1 ml of *RNAlater* in a 2.2 ml tube, before completing the tube volume with *RNALater*.

Back in the lab, before extracting the RNA it is necessary to completely remove the *RNALater* solution. To do so, the bottom of the tube is pierced with a needle and the tube centrifuged 1 min at 3220 g. Finally, the samples are transferred to a new tub, frozen in liquid nitrogen and stored at -80 °C until use.

2.2. Plant total RNA extraction

RNA are extracted starting from samples frozen in liquid nitrogen (reproductive system, leaves or seedlings), then grinded with metal beads (3 mm) using the *TissueLyserII* (Qiagen, Venlo, The Netherlands), for 1 minute at 25 Hz.

2.2.1. TRIZOL total RNA isolation

The grinded tissue is resuspended in 800 µL of *Trizol* (Life Technologies, Carlsbad, California, USA) and mixed until defrosting. Then 200 µL of chloroform are added and the mix is incubated 5 min on ice before being centrifuged for 8 min at 16000 g. Afterwards the aqueous phase, containing the RNA, is collected and the RNA precipitated with 0.8 volumes of Isopropyl alcohol and 300 mM sodium acetate. After centrifugation at 16000 g for 5 min the pellet of RNA is washed with 70% ethanol and resuspended in 200 µL of nuclease free water. Subsequently, 100 µL of phenol are

added and, after 5 min incubation with agitation at room temperature, 100 µL of chloroform are added. After 8 min centrifugation at 16000 g the aqueous phase is collected and the phenol and chloroform step is repeated. Finally, there is a last extraction with 200 µL of chloroform and the RNA contained in the aqueous phase is precipitated with 3 volumes of absolute ethanol and 300 mM sodium acetate. After centrifugation at 4 °C for 30 min the RNA pellet is of washed with 70% ethanol and resuspended in 50 µL of nuclease free water. The RNA integrity and quantity are tested by agarose gel electrophoresis and spectrophotometry (*Nanodrop 2000*, Thermo Fisher Scientific).

2.2.2. RNA isolation for Illumina RNA sequencing.

For sequencing purposes, total RNA was extracted using the *NucleoSpin® RNA Plant isolation kit* (Macherey-Nagel, Düren, GERMANY). 100 mg of previously grinded plant tissue were resuspended in a lysis buffer containing guanidinium thiocyanate and complemented with β-mercaptoethanol. Then, the lysate is clarified by centrifugation through a filter to reduce its viscosity. Next, 70% ethanol is added to adjust the RNA binding condition and the sample is loaded on a NucleoSpin® RNA Plant Column. After centrifugation 1 min at 11000 g the silica membrane containing the bound nucleic acids is washed with a buffer containing ethanol and guanidinium thiocyanate. Afterwards, the DNA is digested on the column with RNase-free DNase (mg of rDNase in 90 µL of rDNase buffer). Finally, after washing with two different buffers and drying of the silica membrane, the RNA are eluted twice with 60 µL of RNase-free water.

2.3. DNase RQI treatment

For further applications, residual DNA contamination needs to be digested. To do so, 5 µg of RNA solution are incubated in 100 µL of nuclease free water containing *1X DNase RQ1 buffer* (400 mM Tris-HCl pH8 ; 100 mM $MgSO_4$; 10 mM $CaCl_2$) and 10 U of *RNase-free DNase RQ1* (Promega, Madison, Wisconsin, USA), for 45 min at 37 °C. Afterwards the volume is adjusted to 300 µL with water and the RNA purified by phenol:chloroform extraction (1:1 v/v) followed by ethanol precipitation, as described above.

2.4. Determination of the RNA Integrity Number (RIN)

For NGS purpose, the integrity of extracted RNA must be determined. To do so, RNA samples after DNA I treatment are tested by microfluidic electrophoresis on the *2100 Bioanalyzer* (Agilent, Santa-Clara, California, USA). The *2100 Bioanalyzer* is an automated tool for electrophoresis of biomolecules transferred to a chip format. Samples are driven, due to the electrode array, through

2 µg of RNA (DNase I treatment) 1 µL of 50 µM oligo-dT 1 µL of 200 ng/µL random primers 1 µL of 10 mM dNTP qsp 13 µL H2O
5 minutes incubation at 65°C
5 minutes incubation at 0°C
4 µL of 5X *Supercript IV* (INVITROGEN) 1 µL of 100 mM DTT 1 µL of 40 U/µL *RNaseOUT* (INVITROGEN) 1 µL of 200 U/µL *Supercript IV* (INVITROGEN)
5 minutes incubation at 25°C
60 minutes incubation at 50°C
15 minutes incubation at 70°C

Table MM.3. Retrotranscription protocol.

the microcapillaries on the chip where the electrophoresis occurs. Nucleic acids are detected by an optical detector thanks to an intercalating fluorescent dye. Finally, the size of nucleic acids is compared to a ladder size marker. The RIN, from 1 to 10 (1 being completely degraded RNA), of each sample is then calculated and given by the *2100 Expert Software* (Agilent) based on the ratio between the rRNA 26S and 18S peaks and the flatness of the region between the 5S and the 18S, called fast region. A RIN superior to 7 is required for NGS purposes.

2.5. RNA Sequencing

The different steps of RNA sequencing (cDNA synthesis, library construction and DNB sequencing) were performed by BGI GENOMICS (Shenzhen, Guangdong, CHINA). To do so, BGI GENOMICS requires a minimum of 200 ng of RNA diluted not more than 20 ng/µL, with a RIN≥6.0, ratios 28S/18S≥1.0, OD260/280≥1.8 and OD260/220≥1.8 to qualify a sample.

For library preparation (**Figure MM.3**), poly(A) mRNA are enriched with oligo(dT)-attached magnetic beads and fragmented into small pieces using divalent cations at high temperature. First cDNA strands are synthesized using reverse transcriptase and random primers. Second strands are synthesized using DNA Polymerase I and RNase H. The double stranded cDNAs (dscDNA) are repaired by phosphorylation of the 5′ ends and 3′ adenylation, allowing the dscDNA to be ligated to adaptors. These products are then purified and enriched by PCR amplification. PCR products are heat denatured and the single-stranded DNA (ssDNA) circularized by splint oligo and DNA ligase. Circularized ssDNA (ssCirDNA) are then used for the generation of DNA nanoballs (DNBs) by rolling circle replication. Finally, DNBs are used for 100 bp paired-reads sequencing on the BGISEQ-500 platform. Afterwards, reads are pre-processed by BGI to discard the adaptors and low-quality reads.

2.6. Reverse-transcription

Complementary DNA (cDNA) to RNA are synthesized using the enzyme reverse transcriptase, an RNA-dependent DNA polymerase, according to the protocol described in the **Table MM.3**. The cDNA can be stored or used as matrix for RT-PCR or RT-qPCR experiments.

Number of cycles	Step	*Phire Hot Start II*		*Phusion High Fidelity*	
		Temperature	Time	Temperature	Time
1	Initial denaturation	Hot Start		Hot Start	
		98°C	30 - 90 s	98°C	30 s
28 - 35	Denaturation	98°C	5 s	98°C	10 s
	Annealing	52 - 72 °C	5 s	52 - 72 °C	10 s
	Extending	72 °C	10 - 15 s/kb	72 °C	15 - 30 s/kb
1	Final extending	72°C	2 min	72°C	2 min
		4°C	Hold	4°C	Hold

Table MM.4. PCR conditions

3. DNA methods

3.1. Plant total DNA extraction

DNA is extracted from seedlings or leaves frozen in liquid nitrogen, then grinded with metal beads in the *TissueLyserII* (Qiagen), for 1 minute at 30 Hz. The grinded tissue is resuspended in 400 µL of cetyltrimethylammonium bromide (CTAB) buffer (CTAB buffer: 2 % cetyltrimethylammonium bromide (p/v); 1.4 M NaCl; 20 mM EDTA; 100 mM Tris-HCl pH 8.0) preheated at 65 °C and supplemented of 0.2 % β-mercaptoethanol (v/v) and 50 mg/mL of RNase A. After incubation for 1 hour at 65 °C, 400 µL of chloroform are added and the mix is centrifuged for 8 min at 13200 g. Finally, the aqueous phase containing the DNA is collected and the DNA precipitated with 2.5 volumes of absolute ethanol. The DNA quality and quantity can be tested by agarose gel electrophoresis and spectrophotometry (*Nanodrop 2000*, Thermo Fisher Scientific).

3.2. Polymerase chain reaction (PCR)

The PCR allows to amplify *in vitro* a specific sequence of DNA. To do so, a heat-resistant DNA polymerase will exponentially synthesize a new DNA strand complementary to the DNA template, primed by specific oligonucleotides flanking the DNA target sequence. The reaction starts with a denaturation step at high temperature (this step can be extended for PCR starting from bacterial cells, to break the bacterial cell wall and liberate the DNA template). In a second step of annealing, the oligonucleotide primers anneal to each of the single-stranded DNA templates at a temperature corresponding to their melting temperature (Tm). Finally, in a third step of elongation the DNA polymerase synthesizes the complementary strands. These three steps (denaturation, annealing and elongation) form a PCR cycle, which is repeated 28 to 35 times to exponentially amplify the target DNA to millions of copies.

In this work, two heat-resistant DNA polymerases have been used (specific conditions in **Table MM.4**). The *Phusion High Fidelity DNA Polymerase* (Thermo Fisher Scientific) was used for cloning techniques, because of its high fidelity, and the *Phire Hot Start II DNA Polymerase* (Thermo Fisher Scientific) was used for other purposes because of its high processivity and stability, and when a high fidelity was not needed.

Number of cycles	Step	Temperature	Time
1	Initial denaturation	95°C	7 min
40	Denaturation	95°C	10 s
	Annealing	58°C	15 s
	Extending	72 °C	20 s
1	Melting curve	95°C	5 s
		55°C	1 min
		95°C	0,11 °C/s
		40°C	30 s

Table MM.5. qPCR conditions

Number of cycles	Step	Temperature	Time
1	Initial denaturation	95°C	2 min
45	Denaturation	95°C	10 s
	Annealing & extending	72 °C	20 s
1	HRM melting curve	95°C	30 s
		60°C	1 min
		65°C	10 s
		95°C	0,02 °C/s
		40°C	30 s

Table MM.6 High resolution melt condition.

3.3. Quantitative polymerase chain reaction (qPCR)

The principle of qPCR is the same as for PCR, but this technique allows the quantification of a DNA sequence within a sample by measuring the release of fluorescence at each PCR cycle. The number of cycles required to exceed the fluorescence detection threshold (Cp: crossing point) is proportional to the number of molecules initially present in the sample.

In this work, qPCR reactions (conditions in **Table MM.5**) have been made in 384 well microplates in the *LightCycler® 480 II* (ROCHE, Basel, Switzerland). Each well contains (reaction volume 6 µL) 0.5 volume of *SYBR Green I Master* (ROCHE), 1.5 mM of primers and 2 ng of DNA. Each reaction is carried out in triplicate, and according to the following conditions: a preheating phase of 10 min at 95 °C, required to activate the polymerase, then 40 cycles of 15 sec at 95 °C, 15 sec at 58 °C and 15 sec at 72 °C. At the end of these amplification cycles, a melting curve is produced by heating the products from 55 °C to 95 °C, thereby verifying the unique nature of the amplicons and the specificity of the primers. Results are normalized against endogenous reference genes. The PCR efficiency for each pair of primers can be calculated using a DNA dilution curve, using the Roche device program. This efficiency is then used to establish the quantity of DNA relative to the reference.

3.4. High resolution melt (HRM) genotyping

HRM genotyping allows to detect single nucleotide polymorphisms (SNP) which is not possible by conventional genotyping by PCR. To do so, samples were run on a LightCycler® 480 system (384-well format) for PCR followed, by the HRM procedure (conditions in **Table MM.6**). Each well contains 2 ng of template DNA, 0.01 µL of each primer at 100 µM, 5 µL of the Precision Melt Supermix for High Resolution Melt Analysis (BIORAD, Hercules, California, USA; 2x real-time PCR mix containing dNTPs, iTaq™ DNA polymerase, $MgCl_2$, EvaGreen dye, stabilizers) and water to 10 µL. Melt temperatures (Tm) are examined by using the Tm Calling software module of the ROCHE LightCycler® 480 package and the genotype determine by using the HRM genotyping software module based on the profile of a known sample.

3.5. Cloning methods

3.5.1. Purification of PCR products

A PCR product needs to be purified after its amplification, before being inserted into a vector. To do so, the product is purified with the kit *Nucleospin Gel and PCR Clean-up* (Macherey-Nagel), allowing two possible purification protocols. If there is more than one PCR product then the desired product needs to be isolated from gel following electrophoresis. The DNA band is cut under UV light and the agarose piece is melted in a buffer containing sodium perchlorate and loaded onto a silica-based column. If there is a single PCR product, the PCR mix is directly loaded onto the column. The DNA is fixed on the silica membrane and the chaotropic agents present in the loading buffer are washed with an ethanol and salt solution and the DNA eluted with a low ionic strength buffer (5 mM Tris-HCl pH 8.5; 0.1 mM EDTA).

3.5.2. Restriction enzymes

A restriction enzyme is an enzyme of bacterial origin specialized in the recognition and cleavage of double stranded DNA. Most restriction enzymes used are of type II, and recognize and cut specific palindromic sequences. The cleavage can leave DNA with sticky ends (5' or 3') or with blunt ends.

In this work, the restriction enzymes were provided by either Promega or by Thermo Fischer Scientific, and the reaction conducted according to the supplier instructions. After digestion, in order to inactivate the enzyme and remove the reaction buffer the volume is adjusted to 300 µL with H_2O and the DNA is then purified by phenol:chloroform extraction followed by ethanol precipitation.

3.5.3. DNA Ligation

3.5.3.1. In a linearized vector

The ligation of a DNA fragment (PCR products and restriction fragments) into a linearized vector is mediated by the enzyme. The reaction in performed (Promega) in a 10 µL volume containing 100 ng of the plasmid vector, the DNA fragment (equimolar ratio to the vector of 1-5 fold), the rapid ligation buffer (Promega, 30 mM Tris-HCl pH 7.8 ; 10 mM $MgCl_2$; 10 mM DTT; 1 mM ATP; 5 % PEG-3000 (p/v) and 3 U T4 DNA ligase, at 4 °C overnight. Then the ligation mix can be directly used to transform competent bacteria.

3.5.3.2. 3′ adenylation and ligation in the pGEM-T Easy vector

For cloning purposes the PCR fragments were amplified with the Phusion High Fidelity DNA polymerase that generates products with blunt ends. To insert these fragment into the pGEM-T Easy vector (Promega) an adenosine needs to be added to each 3'OH end. This A-tailing is mediated by a classic *Taq* DNA polymerase on a purified PCR product, in a 10 µL volume containing 500 ng of purified DNA, 0.2 mM dATP, 1 U of *Taq* DNA polymerase and its buffer (10 mM Tris-HCl pH 8.3; 1.5 mM $MgCl_2$; 50 mM KCl) during 30 min at 72 °C. After incubation the DNA fragment can be ligated into the pGEM-T Easy vector. The ligation into this vector is in 10 µL volume containing 50 ng of the vector, ⅕ (100 ng) of the A-tailing reaction, the rapid ligation buffer (Promega, 30 mM Tris-HCl pH 7.8; 10 mM $MgCl_2$; 10 mM DTT; 1 mM ATP; 5 % PEG-3000 (p/v) and incubated at 4 °C overnight. Then the ligation mix can be directly used to transform competent bacteria.

3.5.4. Gateway cloning

The Gateway cloning system relies on the recombination between att sites on the DNA of interest with those on a vector, as explained in **Figure MM.1**.

In a first step, attB sequences flanking the DNA of interest need to recombine with attP sequences on the donor vector. This reaction is mediated by the BP clonase II mix (Invitrogen/Thermo Fischer Scientific) in a 5 µL volume containing 20 fmol of vector, 10 fmol of the DNA of interest, 1 µL of BP clonase II reaction mix and incubated overnight at 25 °C. The reaction is stopped by adding 2 µL of 20 mg/ml proteinase K (Thermo Fischer Scientific) and incubation for 10 min at 37 °C. Finally, 2-5 µL of the reaction mix are used to transform *E. coli* TOP10 cells, a strain that is sensitive to the product of the *ccdB* gene encoded by the non-recombinant donor plasmid.

In a second step, attL sequences on the entry vector from the first cloning step need to recombine with attR sequences on the destination vector. This reaction is mediated by the LR clonase II mix (Invitrogen/Thermo Fischer Scientific), in a 5 µL volume containing 20 fmol of each vector, 1 µL of LR clonase II reaction mix and incubated overnight at 25 °C. Then, the reaction is stopped by adding 2 µL of 20 mg/ml proteinase K (Thermo Fischer Scientific) and incubation for 10 min at 37 °C. Finally, 2 µL of this reaction are used to transform *E. coli* TOP10 cells sensitive to ccdB encoded by the non-recombinant destination vector.

3.6. Bacterial transformation by heat shock

Heat shock transformation relies on micropores on the bacterial membrane induced by a quick heat shock, allowing DNA to pass thought thanks to a calcium rich environment. Afterwards, bacteria are put back on ice to recover. For this purpose, 50 µL of competent bacteria stored at -80 °C (*E. coli* or *A. tumefaciens*) are thawed on ice, then mixed with 0.1-100 ng of DNA (volume of added DNA must not exceed 5 µL) and left for 30 min on ice. The mix is then put into a heated water bath for 45 sec at 42 °C (5 min at 37 °C for *A. tumefaciens*) and put back on ice for 2 min. Finally, 500 µL of LB media is added to the cells which are incubated 1 hour at 37 °C (28 °C for *A. tumefaciens*), allowing resistance genes to be expressed, before being plated on Pétri dish containing LB agar media supplemented with the appropriate antibiotic(s).

3.7. Plasmid DNA extraction

Plasmid DNA was purified using the kits *Nucleospin Plasmid QuickPure or Nucleospin Plasmid EasyPure* (Macherey-Nagel). A 2 ml volume of LB media with appropriate antibiotic is seeded with selected bacteria and incubated overnight at 37 °C, for *E. coli,* and 48 h at 28 °C, for *A. tumefaciens*. Bacteria are next sedimented by centrifugation and lysed in a buffer containing SDS and sodium hydroxide. Cellular debris and genomic DNA are removed by centrifugation and the supernatant loaded into a silica membrane column. This membrane is then washed with an ethanol and salt solution and the bound DNA eluted with a low ionic strength buffer (5 mM Tris-HCl pH 8.5; 0.1 mM EDTA).

When a high quantity of plasmid DNA is required (for biolistic experiments, for instance) the plasmid from 50-100 mL bacterial cell cultures is purified using the *NucleoBond Xtra Midi* kit (Macherey-Nagel), that is based on the binding of DNA into an anionic exchange column and elution with a high salt buffer, followed by ethanol precipitation.

3.8. Sequencing

3.8.1. Sanger sequencing

Sanger sequencing reactions were performed by Mr. A. Alioua and Ms. S. Koechler from the IBMP platform “gene expression analysis”. This method relies on the chain-termination sequencing of a DNA by the addition of dideoxynucleotides. Nucleic acids are then separated by capillary electrophoresis on the ABI PRISM 3100 Genetic Analyzer (Applied Biosystem, Foster City, California).

3.8.2. Illumina Miseq sequencing

Library preparation and sequencing reactions were performed by Mr. A. Alioua and Ms. S. Koechler from the IBMP platform “gene expression analysis”. Libraries were prepared with the Nextera XT DNA Library Prep kit (Illumina, San Diego, California, USA). First, 1 µg of mtDNA is tagmented, meaning that it is fragmented and tagged with adapter sequences in a single step. Tagmented DNA is then amplified by PCR using primers complementary to the adapter sequences, and cleaned from reactional media. Afterwards, the size of the libraries is checked on the 2100 Bioanalyzer (Agilent). Finally, libraries are sequenced with paired-end 2 x 150 reads on the MiSeq System (Illumina).

3.8.3. PacBio sequencing

Access to the Keygene (Wageningen, The Netherlands) PacBio platform was provided and financed by Enza Zaden. To do so, 1 mg of mtDNA is fragmented (possible sizes range from 250 bp to greater than 20,000 bp) and the DNA ends repaired. Then, adapters are ligated to the DNA fragments to get a circular molecule closed by two hairpins called SMRTbell™ Template (Pacific biosciences, Menlo Park, California, USA). Circular molecules are next purified and the sizes of the libraries checked on the 2100 Bioanalyzer (Agilent). Next, primers are annealed to the hairpin structures and DNA polymerase bound to these primers. Finally, SMRTbell™ Template are loaded into the PacBio Sequel Systems (Pacific biosciences) for single molecule, real-time (SMRT) sequencing.

4. Bioinformatic methods

4.1. Assembly of *L. sativa* mtDNA contigs from Illumina reads

L. sativa Illumina reads were analyzed by Ms. S. Graindorge from the IBMP platform “Bioimage and Bioinformatics”. Assembly of these reads was made with NOVOplasty (k-mer 22) and 48 contigs were obtained, among them 22 were confirmed as mtDNA contigs by BLAST.

4.2. Assembly of *L. sativa* mtDNA contigs from PacBio reads

Raw PacBio data delivered by Keygene were mapped on the 21 mtDNA contigs from Illumina with BWA-mem to remove contaminations (Keygene tried to build contigs but ended up only with

bacterial contigs). Mapped reads were then assembled using Canu to get 4 contigs among them 3 were identified as mitochondrial by BLAST (177 kb, 91 kb and 73 kb) and 1 of bacterial origin (10 kb).

4.3. Assembly of wild lettuce mtDNA from Illumina reads

Illumina reads from wild lettuce species were trimmed with Trimmomatic to remove adapter sequences and low quality ends and joined with Fastq-join when paired reads mapped to each others. Then those reads were assembled, first with Velvet. “Velvet contigs” were mapped on the *L. sativa* mtDNA and mapped “Velvet contigs” were used as “trusted contigs” for a second round of assembly with SPAdes. “SPAdes” contigs were identified by BLAST and filtered according to their coverage. Filtered “SPAdes contigs” were extended by SSPACE into long scaffolds (up 170 kb). Finally, contigs were assembled and gaps filled using contigs from a metagenomic assembler IDBA-UD.

4.4. Determination of RNA editing sites on *L. sativa* mtDNA

L. sativa RNA reads (accessions SRR080725 and SRR085107) were mapped on *L. sativa* mtDNA using Hisat2 with reduced mismatch penalties (--mp parameters to “2,0”), allowing the mapping of more reads. SAM format files were then converted to BAM format with Samtools. BAM files were sorted with Samtools for visualization on IGV. SNPs calling was made with Freebayes, using default parameters. SNPs were filtered to remove non C to U (or G to A) ones. Finally, only SNPs in coding regions were conserved.

4.5. Identification of SNPs associated with a qualitative phenotype

DNA nanoballs sequencing data were preprocessed by BGI to remove low quality reads and adapters. Preprocessed reads were then mapped on the *L. sativa* nuclear (V8), plastidial (NC007578) and mitochondrial genomes. SAM format files were converted to BAM format with Samtools. SNPs calling was made with VarScan with default parameters and tables manipulated on R with Data.table.

5. Plants methods

5.1. *In vitro* methods

5.1.1. Seeds sterilization

To remove bacterial and fungi contamination associated to the seeds surface, seeds need to be sterilized prior to growth. To do so, lettuce seeds are washed 30 sec with 70 % ethanol followed by 10 min wash with 1 % NaOH. Finally, seeds are washed in sterile water and dried on blotting paper.

5.1.2. *In vitro* growth on solid medium

Depending on the upcoming experiments, seeds were grown on agarose plates (either MS255 or agarose only media) in growth chambers, either in the dark at 23 ° C (for purification of mitochondria from etiolated seedlings) or under a photoperiod of 16 hours of light at 21 °C and 8 hours of dark at 18 °C.

5.1.3. *In vitro* growth in liquid medium

For growth in liquid medium, sterilized seeds were first grown on agarose plates (either MS255 or agarose only media) in a growth chamber under a day/night photoperiod of 16 hours light at 21 °C and 8 hours in the dark at 18 °C for seven days. Afterwards, seedlings were transferred to liquid MS255 medium by immerging the roots.

5.1.4. *In vitro* growth on blotting paper

Seeds were disposed on T-300 blotting paper (Ø 80 mm; All Paper b.v. Didam, the Netherlands) in Pétri dish and wet with 5 mL of water (or water containing genotoxic agent). The petri dish was next sealed with parafilm to keep humidity. Then the boxes are put in a growth chamber under a day/night cycle of 16 hours of light at 21 °C and 8 hours in the dark at 18 °C up to 21 days.

5.2. Stable transformation of *A. thaliana* by floral-dip

To optimize the number of transformants, the first floral stem is cut a week before transformation, to favor the growth of numerous flower stems. Plants are ready to be transformed when siliques start to form. The transformation is mediated by a *A. tumefaciens* strain carrying the construction of interest on a binary plasmid. A 25 ml preculture of this strain is grown, at 28 °C

overnight in LB media containing appropriate antibiotics, and then used to inseminate a 250 ml culture in LB media. When the 250 ml culture reaches OD_{600}= 0.8, bacteria are sedimented by centrifugation 20 min at 4000 g and resuspended in 5 % saccharose and 0.02 % Silwet L-77 to OD_{600}= 0.8. The aerial parts of the plant can now be plunged into the bacterial suspension for 30 sec. Afterwards, plants are put in the dark under a dome to maintain humidity for 2 days. Finally, plants are grown at normal conditions until seeds are obtained.

Transformants are next selected on agarose plates (MS255 media, Duchefa) containing 1% sucrose and the appropriate antibiotics. To do so, 150 µL of sterilized seeds are suspended in 8 ml of 0.1 agarose (p/v) and plated on 15 cm plates the solid media. The seeds are then stratified in the dark 48 h at 4 °C before normal growth on a growth chamber. Finally, seedlings resistant to the selection antibiotic are transplanted into soil for the production of seeds.

5.3. Transfection of tobacco leaves for subcellular localization of fluorescent proteins

Transfection of *Nicotiana benthamiana* or *Nicotiana tabaccum* was made on 4 weeks old leaves.

5.3.1. By agroinfiltration

This transfection is mediated by a *A. tumefaciens* strain carrying the construction of interest cloned in a binary vector. A 2 ml preculture of this strain is grown, at 28 °C overnight in LB media containing appropriate antibiotics, and then used to inseminate a 5 ml culture in the same media. When the 5 ml culture reaches OD_{600}= 1, bacteria are sedimented by centrifugation for 10 min at 4000 g and resuspended in 10 mM $MgCl_2$ and 100 µM acetosyringone to OD_{600}= 1. A tiny lesion is made on the underside of the *N. benthamiana* leaf and the bacterial suspension is infiltrated into the leaf using a 1 ml syringe without a needle. Plants are put back to growth for 72 h, before observation of the expression of fluorescent proteins at the confocal microscope.

5.3.2. By biolistics

Biolistic transfection is mediated by the PDS-1000/He Biolistic Particle Transfection System (BIORAD). Gold or tungsten particles (0.6 µm) are coated with DNA. For that, 20 µL of a 30 mg/mL particles suspension in 10 % glycerol is added of 7 µg of plasmid DNA, 30 µL of 2.5 M $CaCl_2$, 16 µL of 0.1 M spermidine and vortexed vigorously for 15 min. The particles are then washed four times with 200 µL of 100 % Ethanol, and finally resuspended in 35 µL of 100 % Ethanol.

For the biolistic transfection, the *N. benthamiana* or *N. tabaccum* leaves are placed on an agar plate, abaxial leaf face up, put in the chamber of the PDS-1000/He (BIORAD) and transfected with 10 µL of particles suspension previously spread on a macrocarrier, using 1350 psi rupture disks. Finally, the leaves are put in the dark overnight before observation at the confocal microscope.

5.3.3. Observation of transfected cells

Transfected cells are observed under an inverted laser scanning confocal microscope (LSM700, Carl Zeiss, Oberkochen, Germany). To do so, a piece of leaf is placed on a microscope slide, abaxial face up, a droplet of water is added and the leaf piece between slide and cover slip is sealed with tape, and ready to be observed.

5.4. Crosses between two *L. sativa* plants

To cross *L. sativa* plants two dimensions must be considered. The first one is the spatial dimension. Indeed, flowers of *L. sativa* are hermaphrodite, having both male and female organs. That is why to cross two plants, the flowers of the female plant need to be emasculated, by removing its anthers. However, emasculation must be done before the stigma grows through the anthers, and that is where a temporal dimension comes in. Indeed, lettuce flowers open only once, and for a few hours, and to avoid self-pollination the emasculation has to be done before the opening of the flower. In practice, we emasculate 5 flowers per cross. Afterwards, widely opened flowers of the male plant are harvested and used to pollinate emasculated (or male sterile) flowers (1 to 1) by brushing.

5.5. Evaluation of pollen viability by Alexander staining

The viability of the pollen of *L. sativa* plants was evaluated by Alexander staining (Alexander, 1969). The stain solution is composed of 10 % Ethanol (v/v), 25 % of glycerol (v/v), 4 % glacial acetic acid, 4 µM fuchsin acid, 1 µM malachite green and 500 nM orange G in water. In practice, all the sexual organs of a widely opened flower are removed and placed on a microscope slide. The sexual organs are next covered with 150 µl of Alexander staining solution diluted to the tenth and incubated for 2 hours before observation under optical microscope. The pollen cell wall is colored blue-green (due to malachite green), while the polysaccharides and cellulose present in the viable protoplasm are colored red-pink (due to orange G and fuchsin acid). This makes it easy to distinguish between viable pollen with a red-pink protoplasm from non-viable one that is stained only green.

5.6. Ciprofloxacin (CIP) genotoxic stress

To test the resistance of *L. sativa* to CIP, two conditions were determined. Sterilized seeds were distributed either on agarose plates (MS255 or agarose only media) containing CIP (from 250 nM to 2 mM) for 1 week or on blotting paper humidified with water containing CIP (from 2 µM to 2 mM) for 1 to 3 weeks. The surviving plants were then transplanted into soil until they produce flowers, to evaluate their fertility.

6. Plant mitochondria methods

6.1. Mitochondrial purification from etiolated seedlings

Every step leading to intact and purified mitochondria was done at 4 °C in a cold room, to avoid their degradation while minimizing waiting times. Seven days old etiolated seedlings are ground with a mortar and pestle in grinding buffer (0.6 M sucrose, 50 mM tetrasodium pyrophosphate, 4 mM EDTA, 20 mM KH_2PO_4, 2 % PVP-40, 2 % BSA, pH adjusted to 7.5, 40 mM ascorbate, 10 mM cysteine, 2 mM DTT; 10 ml for 100 seedlings). The suspension is filtered through two layers of nylon cloth (96 µm) and the solid material is ground a second time in 10 ml grinding buffer and filtered. Next, the filtrate undergoes differential centrifugation. The first step aims to remove cell debris by two consecutive centrifugations at 3500 g for 5 min. The supernatant is transferred to a new tube and centrifuged at 6000 g for 10 min. Finally, the supernatant is transferred into a new tube and mitochondria is sedimented by centrifugation at 17000 g for 20 min. The pellet is then resuspended into 400 µL of washing buffer (0,3 M sucrose, 10 mM MOPS, 0,1% BSA, pH adjusted to 7.5). Mitochondria are then purified on a discontinuous percoll gradient by centrifugation at 25000 g for 25 min, with break off. The percoll gradient is composed of 3 layers of different percoll solutions, 50%, 25% and 18% (v/v) in washing buffer. Mitochondria are carefully recovered from the 50-25% interphase. To remove the remaining percoll, mitochondria are diluted in 30 ml of 1 X washing buffer and centrifuged 15 min at 17000 g. The washing step is repeated twice. Finally, the mitochondrial pellet is transferred into a 1.5 ml Eppendorf tube until use.

6.2. DNase I treatment of mitochondria

After mitochondria purification, contaminant nuclear or plastid DNA bound to the mitochondria surface is removed by DNase treatment. For that mitochondria are resuspended into

100 µL of 1 X washing buffer with 50 µg DNase I and 0.55 µM Mg^{2+} and incubated 20 min at 25 °C. During the incubation, the mtDNA is protected by the intact mitochondrial membrane, while external DNA is digested. After incubation, 1 ml of washing medium containing EDTA and EGTA (both at 10 mM) is added and mitochondria collected by centrifugation at 18000 g for 15 min. Finally, the mitochondria pellet can be used directly or flash frozen on liquid nitrogen and stored at -80 °C for further applications.

6.3. Mitochondrial DNA extraction

DNA extraction from purified mitochondria was performed using the *QIAamp® DNA Micro* kit (Qiagen) following the protocol "Isolation of Genomic DNA from tissues". Mitochondria pellet is first resuspended in a lysis buffer containing SDS and lysed with proteinase K (Qiagen) overnight at 56 °C. After this incubation, the lysate is transferred to a silica column provided by the kit and DNA is bound to it. The column is then washed with an ethanol and salt solution and the DNA eluted with distilled water.

Bibliography

Abad, A.R., Mehrtens, B.J., and Mackenzie, S.A. (1995). Specific expression in reproductive tissues and fate of a mitochondrial sterility-associated protein in cytoplasmic male-sterile bean. Plant Cell **7**: 271–285.

Abdelnoor, R.V., Yule, R., Elo, A., Christensen, A.C., Meyer-Gauen, G., and Mackenzie, S.A. (2003). Substoichiometric shifting in the plant mitochondrial genome is influenced by a gene homologous to MutS. Proc. Natl. Acad. Sci. USA **100**: 5968–5973.

Alam, T.I., Kanki, T., Muta, T., Ukaji, K., Abe, Y., Nakayama, H., Takio, K., Hamasaki, N., and Kang, D. (2003). Human mitochondrial DNA is packaged with TFAM. Nucleic Acids Res. 31: 1640–1645.

Alexander, M.P. (1969). Differential staining of aborted and nonaborted pollen. Stain Technol. **44**: 117–122.

Allen, J.O., Fauron, C.M., Minx., P., Roark, L., Oddiraju, S., Lin, G.N., Meyer, L., Sun, H., Kim, K., Wang, C., Du, F., Xu, D., Gibson, M., Cifrese, J., Clifton, S.W., and Newton, K.J. (2007). Comparisons among two fertile and three male-sterile mitochondrial genomes of maize. Genetics 177: 1173–1192.

Almagro Armenteros, J.J., Salvatore, M., Emanuelsson, O., Winther, O., von Heijne, G., Elofsson, A., and Nielsen, H. (2019). Detecting sequence signals in targeting peptides using deep learning. Life Sci. Alliance 2: e201900429.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic Local Alignment Search Tool. J. Mol. Biol. 215: 403–410.

Alverson, A.J., Rice, D.W., Dickinson, S., Barry, K., and Palmer, J.D. (2011). Origins and recombination of the bacterial-sized multichromosomal mitochondrial genome of cucumber. Plant Cell **23**: 2499–2513.

Anderson, S., Bankier A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F, Schreier P.H., Smith, A.J., Staden, R., and Young, I.G. (1981). Sequence and organization of the human mitochondrial genome. Nature 290: 457–465.

Andersson, S.G.E., Zomorodipour, A., Andersson, J.O., Sicheritz-Pontén, T., Alsmark, U.C.M., Podowski, R.M., Näslund, A.K., Eriksson, A.-S., Winkler, H.H., and Kurland, C.G. (1998). The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. Nature 396: 133–140.

Aoyama, H., Hagiwara, Y., Misumi, O., Kuroiwa, T., and Nakamura, S. (2006). Complete elimination of maternal mitochondrial DNA during meiosis resulting in the paternal inheritance of the mitochondrial genome in *Chlamydomonas* species. Protoplasma **228**: 231–242.

Aravind, L., Makarova, K.S., and Koonin, E.V. (2000). SURVEY AND SUMMARY: Holliday junction resolvases and related nucleases: identification of new families, phyletic distribution and evolutionary trajectories. Nucleic Acids Res. **28**: 3417–3432.

Ariizumi, T. and Toriyama, K. (2011). Genetic regulation of sporopollenin synthesis and pollen exine development. Annu. Rev. Plant Biol. **62**: 437–460.

Arimura, S.-i., Yamamoto, J., Aida, G.P., Nakazono, M., and Tsutsumi, N. (2004). Frequent fusion and fission of plant mitochondria with unequal nucleoid distribution. Proc. Natl. Acad. Sci. USA **101**: 7805–7808.

Arrieta-Montiel, M.P., Shedge, V., Davila, J., Christensen, A.C., and Mackenzie, S.A. (2009). Diversity of the arabidopsis mitochondrial genome occurs via nuclear-controlled recombination activity. Genetics **183**: 1261–1268.

Backert, S. (2002). R-loop-dependent rolling-circle replication and a new model for DNA concatemer resolution by mitochondrial plasmid mp1. EMBO J. **21**: 3128–3136.

Backert, S. and Börner, T. (2000). Phage T4-like intermediates of DNA replication and recombination in the mitochondria of the higher plant *Chenopodium album* (L.). Curr. Genet. **37**:

304–314.
Backert, S., Dörfel, P., Lurz, R., and Börner, T. (1996). Rolling-circle replication of mitochondrial DNA in the higher plant *Chenopodium album* (L.). Mol. Cell. Biol. **16**: 6285–6294.
Backert, S., Lurz, R., Oyarzabal, O.A., and Borner, T. (1997a). High content, size and distribution of single-stranded DNA in the mitochondria of *Chenopodium album* (L.). Plant Mol. Biol. **33**: 1037–1050.
Backert, S., Lynn Nielsen, B., and Börner, T. (1997b). The mystery of the rings: structure and replication of mitochondrial genomes from higher plants. Trends Plant Sci. **2**: 477–483.
Balk, J. and Leaver, C.J. (2001). The PET1-CMS mitochondrial mutation in sunflower is associated with premature programmed cell death and cytochrome c release. Plant Cell **13**: 1803–1818.
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., and Pevzner, P.A. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19: 455–477.
Banting, G.S. and Glerum, D.M. (2006). Mutational Analysis of the *Saccharomyces cerevisiae* Cytochrome *c* Oxidase Assembly Protein Cox11p. Eukaryot. Cell **5**: 568–578.
Bautista, O.V., Fischer, G., and Cárdenas, J.F. (2013). Cadmium and chromium effects on seed germination and root elongation in lettuce, spinach and Swiss chard. Agron. Colomb. **31**: 48–57.
Beam, C.E., Saveson, C.J., and Lovett, S.T. (2002). Role for radA/sms in recombination intermediate processing in *Escherichia coli.* J. Bacteriol **184**: 6836–6844.
Bellaoui, M., Grelon, M., Pelletier, G., and Budar, F. (1991). The restorer Rfo gene acts post-translationally on the stability of the ORF138 Ogura CMS-associated protein in reproductive tissues of rapeseed cybrids. Plant Mol. Biol. **40**: 893–902.
Bendich, A.J. (1996). Structural analysis of mitochondrial dna molecules from fungi and plants using moving pictures and pulsed-field gel electrophoresis. J. Mol. Biol. **255**: 564–588.
Bennett, M.D., Leitch, I.J., Price, H.J., and Johnston, J.S. (2003). Comparisons with *Caenorhabditis* (approximately 100 Mb) and Drosophila (approximately 175 Mb) using flow cytometry show genome size in Arabidopsis to be approximately 157 Mb and thus approximately 25% larger than the Arabidopsis genome initiative estimate of approximately 125 Mb. Ann. Bot. **91**: 547–557.
Bensley, R.R. and Hoerr, N.L. (1934). Studies on cell structure by the freezing-drying method VI. The preparation and properties of mitochondria. Anat. Rec. **60**: 449–455.
Bergthorsson, U., Adams, K.L., Thomason, B., and Palmer, J.D. (2003). Widespread horizontal transfer of mitochondrial genes in flowering plants. Nature **424**: 197–201.
Birky, C.W. (2001). The inheritance of genes in mitochondria and chloroplasts: Laws, mechanisms, and models. Annu. Rev. Genet. **35**: 125–148.
Boetzer, M., Henkel, C.V., Jansen, H.J., Butler, D., and Pirovano, W. (2010). Scaffolding pre-assembled contigs using SSPACE. Bioinformatics **27**: 578–579.
Boetzer, M. and Pirovano, W. (2012). Toward almost closed genomes with GapFiller. Genome Biol. **13**: R56.
Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinforma. Oxf. Engl. **30**: 2114–2120.
Borg, M., Brownfield, L., and Twell, D. (2009). Male gametophyte development: a molecular perspective. J. Exp. Bot. **60**: 1465–1478.
Bouchier, C., Ma, L., Créno, S., Dujon, B., and Fairhead, C. (2009). Complete mitochondrial genome sequences of three *Nakaseomyces* species reveal invasion by palindromic GC clusters and considerable size expansion. FEMS Yeast Res. **9**: 1283–1292.
Brandt, P., Unseld, M., Eckert-Ossenkopp, U., and Brennicke, A. (1993). An rpsl4 pseudogene is

transcribed and edited in Arabidopsis mitochondria. Curr. Genet. **24**: 330–336.
Brieba (2019). Structure–function analysis reveals the singularity of plant mitochondrial dna replication components: A mosaic and redundant system. Plants **8**: 533.
Brown, G.G. (1999). Unique aspects of cytoplasmic male sterility and fertility restoration in *Brassica napus*. J. Hered. **90**: 351–356.
Brown, G.G., Formanová, N., Jin, H., Wargachuk, R., Dendy, C., Patil, P., Laforest, M., Zhang, J., Cheung, W.Y., and Landry, B.S. (2003). The radish *Rfo* restorer gene of Ogura cytoplasmic male sterility encodes a protein with multiple pentatricopeptide repeats: *Radish* Rfo *restorer gene of Ogura cytoplasmic male sterility*. Plant J. 35: 262–272.
Brown, W.M., George, M., and Wilson, A.C. (1979). Rapid evolution of animal mitochondrial DNA. Proc. Natl. Acad. Sci. USA **76**: 1967–1971.
Budar, F., Touzet, P., and Paepe, R.D. (2003). The Nucleo-Mitochondrial Conflict in Cytoplasmic Male Sterilities Revisited. Genetica **117**: 3–16.
Cappadocia, L., Maréchal, A., Parent, J.-S., Lepage, É., Sygusch, J., and Brisson, N. (2010). Crystal structures of DNA-Whirly complexes and their role in *Arabidopsis* organelle genome repair. Plant Cell **22**: 1849–1867.
Cappelle, C., Mörchen, M., Hilbert, J.-L., and Rambaud, C. (2007). Regeneration and molecular characterization of a male sterile interspecific somatic hybrid between Cichorium intybus and C. endivia. Plant Sci. **172**: 596–603.
Carrie, C. and Whelan, J. (2013). Widespread dual targeting of proteins in land plants: When, where, how and why. Plant Signal. Behav. **8**: e25034.
Case, A.L. and Willis, J.H. (2008). Hybrid male sterility in *Mimulus* (*Phrymaceae*) is associated with a geographically restricted mitochondrial rearrangement. Evolution **62**: 1026–1039.
Chang, J.C. and Kan, Y.W. (1979). Beta 0 thalassemia, a nonsense mutation in man. Proc. Natl. Acad. Sci. USA **76**: 2886–2889.
Chase, C.D. and Ortega, V.M. (1992). Organization of ATPA coding and 3' flanking sequences associated with cytoplasmic male sterility in *Phaseolus vulgaris* L. Curr. Genet. **22**: 147–153.
Chen, L. and Liu, Y.-G. (2014). Male sterility and fertility restoration in crops. Annu. Rev. Plant Biol. **65**: 579–606.
Chen, X., Hu, J., Zhang, H., and Ding, Y. (2014). DNA methylation changes in photoperiod-thermo-sensitive male sterile rice PA64S under two different conditions. Gene **537**: 143–148.
Chetrit, P., Rios, R., De Paepe, R., Vitart, V., Gutierres, S., and Vedel, F. (1992). Cytoplasmic male sterility is associated with large deletions in the mitochondrial DNA of two *Nicotiana sylvestris* protoclones. Curr. Genet. **21**: 131–137.
Chevigny, N., Nadiras, C., Raynaud, C., Le Ret, M., Bichara, M., Erhardt, M., Dietrich, A., and Gualberto, J.M. (2019). RADA is the main branch migration factor in plant mitochondrial recombination and its defect leads to mtDNA instability and cell cycle arrest. bioRxiv.
Chevigny, N., Schatz-Daas, D., Lotfi, F., and Gualberto, J.M. (2020). DNA repair and the stability of the plant mitochondrial genome. Int. J. Mol. Sci. **21**: 328.
Cho, H.S., Lee, S.S., Kim, K.D., Hwang, I., Lim, J.-S., Park, Y.-I., and Pai, H.-S. (2004). DNA gyrase is involved in chloroplast nucleoid partitioning. Plant Cell **16**: 2665–2682.
Christensen, A.C. (2013). Plant mitochondrial genome evolution can be explained by dna repair mechanisms. Genome Biol. Evol. **5**: 1079–1086.
Ciesielski, G.L., Oliveira, M.T., and Kaguni, L.S. (2016). Animal mitochondrial DNA replication. In The Enzymes (Elsevier), pp. 255–292.
Claude, B.A. (1946). Fractionation of mammalian liver cells by differential centrifugation. J. Exp. Med. **84**: 61–89.
Cockram, C.A., Filatenkova, M., Danos, V., El Karoui, M., and Leach, D.R.F. (2015). Quantitative

genomic analysis of RecA protein binding during DNA double-strand break repair reveals RecBCD action in vivo. Proc. Natl. Acad. Sci. USA **112**: E4735–E4742.

Coen, E.S. and Meyerowitz, E.M. (1991). The war of the whorls: genetic interactions controlling flower development. Nature **353**: 31–37.

Cognat, V., Morelle, G., Megel, C., Lalande, S., Molinier, J., Vincent, T., Small, I., Duchêne, A.-M., and Maréchal-Drouard, L. (2017). The nuclear and organellar tRNA-derived RNA fragment population in Arabidopsis thaliana is highly dynamic. Nucleic Acids Res. **45**: 3460–3472.

Cooper, D.L., Boyle, D.C., and Lovett, S.T. (2015). Genetic analysis of Escherichia coli RadA: functional motifs and genetic interactions. Mol. Microbiol. **95**: 769–779.

Cooper, D.L. and Lovett, S.T. (2016). Recombinational branch migration by the RadA/Sms paralog of RecA in *Escherichia coli*. eLife **5**: e10807.

Cox, M.M. (2007). Regulation of Bacterial RecA Protein Function. Crit. Rev. Biochem. Mol. Biol. **42**: 41–63.

Cui, X., Wise, R.P., and Schnable, P.S. (1996). The *rf2* nuclear restorer gene of male-sterile T-cytoplasm maize. Science **272**: 1334–1336.

Darling, A.E., Mau, B., and Perna, N.T. (2010). progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS ONE **5**: e11147.

Darracq, A., Varré, J.S., Maréchal-Drouard, L., Courseaux, A., Castric, V., Saumitou-Laprade, P., Oztas, S., Lenoble, P., Vacherie, B., Barbe, V., and Touzet, P. (2011). Structural and content diversity of mitochondrial genome in beet: A comparative genomic analysis. Genome Biol. Evol. 3: 723–736.

Davila, J.I., Arrieta-Montiel, M.P., Wamboldt, Y., Cao, J., Hagmann, J., Shedge, V., Xu, Y.-Z., Weigel, D., and Mackenzie, S.A. (2011). Double-strand break repair processes drive evolution of the mitochondrial genome in Arabidopsis. BMC Biol. 9: 64.

Davis, R.M., Subbarao, K.V., Raid, R.N., and Kurtz, A.K. (1997). Compendium of lettuce diseases (APS Press).

Dawuda, M.M., Liao, W., Hu, L., Yu, J., Xie, J., Calderón-Urrea, A., Jin, X., and Wu, Y. (2019). Root tolerance and biochemical response of Chinese lettuce (*Lactuca sativa* L.) genotypes to cadmium stress. PeerJ **7**: e7530.

Delourme, R. and Budar, F. (1999). Male sterility. In Biology of Brassica coenospecies (Elsevier), pp. 185–216.

Dewey, R., Levings, C.S., and Timothy, D.H. (1986). Novel recombinations in the maize mitochondrial genome produce a unique transcriptional unit in the texas male-sterile cytoplasm. Cell **44**: 439–449.

Dewey, R.E., Timothy, D.H., and Levings, C.S. (1987). A mitochondrial protein associated with cytoplasmic male sterility in the T cytoplasm of maize. Proc. Natl. Acad. Sci. USA **84**: 5374–5378.

Dhar, N., Mamo, B.E., Subbarao, K.V., Koike, S.T., Fox, A., Anchieta, A., and Klosterman, S.J. (2020). Measurements of aerial spore load by qpcr facilitates lettuce downy mildew risk advisement. Plant Dis. **104**: 82–93.

Di Mola, I., Rouphael, Y., Colla, G., Fagnano, M., Paradiso, R., and Mori, M. (2017). Morphophysiological traits and nitrate content of greenhouse lettuce as affected by irrigation with saline water. HortScience **52**: 1716–1721.

Dierckxsens, N., Mardulyn, P., and Smits, G. (2016). NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data. Nucleic Acids Res. **45**: e18.

Diray-Arce, J., Liu, B., Cupp, J.D., Hunt, T., and Nielsen, B.L. (2013). The Arabidopsis At1g30680 gene encodes a homologue to the phage T7 gp4 protein that has both DNA primase and DNA helicase activities. BMC Plant Biol. **13**: 36.

Doda, J.N., Wright, C.T., and Clayton, D.A. (1981). Elongation of displacement-loop strands in

human and mouse mitochondrial DNA is arrested near specific template sequences. Proc. Natl. Acad. Sci. USA **78**: 6116–6120.

Doré, C. and Varoquaux, F. (2006). Histoire et amélioration de cinquante plantes cultivées 1ere édition. (Éditions Quae: Versailles).

Dörfel, P., Weihe, A., Dolferus, R., and Börner, T. (1991). DNA sequence of a mitochondrial plasmid fromChenopodium album. Plant Mol. Biol. **17**: 155–156.

Drees, J.C., Lusetti, S.L., and Cox, M.M. (2004). Inhibition of RecA protein by the *Escherichia coli* RecX protein: modulation by the RecA c terminus and filament functional state. J. Biol. Chem. **279**: 52991–52997.

Dubreucq, A., Berthe, B., Asset, J.-F., Boulidard, L., Budar, F., Vasseur, J., and Rambaud, C. (1999). Analyses of mitochondrial DNA structure and expression in three cytoplasmic male-sterile chicories originating from somatic hybridisation between fertile chicory and CMS sunflower protoplasts: Theor. Appl. Genet. **99**: 1094–1105.

Duchen, M.R. (2000). Mitochondria and calcium: from cell signalling to cell death. J. Physiol. **529 Pt 1**: 57–68.

Duchêne, A.-M. and Giegé, P. (2012). Dual localized mitochondrial and nuclear proteins as gene expression regulators in plants? Front. Plant Sci. **3**: 221.

Ducos, E., Touzet, P., and Boutry, M. (2001). The male sterile G cytoplasm of wild beet displays modified mitochondrial respiratory complexes. Plant J. **26**: 171–180.

Dufaÿ, M., Touzet, P., Maurice, S., and Cuguen, J. (2007). Modelling the maintenance of male-fertile cytoplasm in a gynodioecious population. Heredity **99**: 349–356.

Duroc, Y., Gaillard, C., Hiard, S., Defrance, M.-C., Pelletier, G., and Budar, F. (2005). Biochemical and functional characterization of ORF138, a mitochondrial protein responsible for Ogura cytoplasmic male sterility in Brassiceae. Biochimie **87**: 1089–1100.

Edmondson, A.C., Song, D., Alvarez, L.A., Wall, M.K., Almond, D., McClellan, D.A., Maxwell, A., and Nielsen, B.L. (2005). Characterization of a mitochondrially targeted single-stranded DNA-binding protein in Arabidopsis thaliana. Mol. Genet. Genomics **273**: 115–122.

Eggler, A.L., Lusetti, S.L., and Cox, M.M. (2003). The C terminus of the *Escherichia coli* RecA protein modulates the DNA binding competition with single-stranded DNA-binding protein. J. Biol. Chem. **278**: 16389–16396.

Elo, A., Lyznik, A., Gonzalez, D.O., Kachman, S.D., and Mackenzie, S.A. (2003). Nuclear 33 DNA and RNA metabolism are clustered in the Arabidopsis genome. Plant Cell **15**: 1619–1631.

Ernster, L. and Schatz, G. (1981). Mitochondria: a historical review. J. Cell Biol. **91**: 227s–255s.

Evans-Roberts, K.M., Mitchenall, L.A., Wall, M.K., Leroux, J., Mylne, J.S., and Maxwell, A. (2016). DNA gyrase is the target for the quinolone drug ciprofloxacin in *Arabidopsis thaliana*. J. Biol. Chem. **291**: 3136–3144.

Fauré, S., Noyer, J.-L., Carreel, F., Horry, J.-P., Bakry, F., and Lanaud, C. (1994). Maternal inheritance of chloroplast genome and paternal inheritance of mitochondrial genome in bananas (*Musa acuminata*). Curr. Genet. **25**: 265–269.

Fellenberg, C., Milkowski, C., Hause, B., Lange, P.-R., Böttcher, C., Schmidt, J., and Vogt, T. (2008). Tapetum-specific location of a cation-dependent O-methyltransferase in *Arabidopsis thaliana*. Plant J. **56**: 132–145.

Fellenberg, C., van Ohlen, M., Handrick, V., and Vogt, T. (2012). The role of CCoAOMT1 and COMT1 in Arabidopsis anthers. Planta **236**: 51–61.

Feng, X., Kaur, A.P., Mackenzie, S.A., and Dweikat, I.M. (2009). Substoichiometric shifting in the fertility reversion of cytoplasmic male sterile pearl millet. Theor. Appl. Genet. **118**: 1361–1370.

Fisher, R.P. and Clayton, D.A. A transcription factor required for promoter recognition by Human mitochondrial RNA polymerase. J. Biol. Chem. **260**: 11330–11338.

Fishman, L. and Willis, J.H. (2006). A cytonuclear incompatibility causes anther sterility in Mimulus hybrids. Evolution **60**: 1372–1381.

Fisk, D.G., Walker, M.B., and Barkan, A. (1999). Molecular cloning of the maize gene crp1 reveals similarity between regulators of mitochondrial and chloroplast gene expression. EMBO J. **18**: 2621–2630.

Fletcher, K., Gil, J., Bertier, L.D., Kenefick, A., Wood, K.J., Zhang, L., Reyes-Chin-Wo, S., Cavanaugh, K., Tsuchida, C., Wong, J., and Michelmore, R. (2019). Genomic signatures of heterokaryosis in the oomycete pathogen *Bremia lactucae*. Nat. Commun. 10: 2645.

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., Becker, F.F.M., Schnabel, S.K., Willems, L.A.J., Ligterink, W., van Arkel, J., Mumm, R., Gualberto, J.M., Savage, L., Kramer, D.M., Keurentjes, J.J.B., van Eeuwijk, F., Koornneef, M., Harbinson, J., Aarts, M.G.M., and Wijnker, E. (2020). Reciprocal cybrids reveal how organellar genomes affect plant phenotypes. Nat. Plants 6: 13–21.

Forget, A.L. and Kowalczykowski, S.C. (2012). Single-molecule imaging of DNA pairing by RecA reveals a three-dimensional homology search. Nature **482**: 423–427.

Forner, J. and Binder, S. (2007). The red fluorescent protein eqFP611: application in subcellular localization studies in higher plants. BMC Plant Biol. **7**: 28.

Foury, F., Roganti, T., Lecrenier, N., and Purnelle, B. (1998). The complete sequence of the mitochondrial genome of *Saccharomyces cerevisiae*. FEBS Lett. **440**: 325–331.

Friedkin, M. and Lehninger, A.L. (1949). Oxidation-coupled incorporation of inorganic radiophosphate into phospholipide and nucleic acid in a cell-free system. J. Biol. Chem. **177**: 775–788.

Fuchs, P., Rugen, N., Carrie, C., Elsässer, M., Finkemeier, I., Giese, J., Hildebrandt, T.M., Kühn, K., Maurino, V.G., Ruberti, C., Schallenberg-Rüdinger, M., Steinbeck, J., Braun, H.P., Eubel, H., Meyer, E.H., Müller-Schüssele, S.J., and Schwarzländer, M. (2020). Single organelle function and organization as estimated from Arabidopsis mitochondrial proteomics. Plant J. 101: 420–441.

Fujii, S., Bond, C.S., and Small, I.D. (2011). Selection patterns on restorer-like genes reveal a conflict between nuclear and mitochondrial genomes throughout angiosperm evolution. Proc. Natl. Acad. Sci. USA **108**: 1723–1728.

Fujii, S., Kazama, T., Yamada, M., and Toriyama, K. (2010). Discovery of global genomic re-organization based on comparison of two newly sequenced rice mitochondrial genomes with cytoplasmic male sterility-related genes. BMC Genomics **11**: 209.

Fujii, S., Komatsu, S., and Toriyama, K. (2007). Retrograde regulation of nuclear gene expression in CW-CMS of rice. Plant Mol. Biol. **63**: 405–417.

Fujii, S. and Toriyama, K. (2008). Genome Barriers between Nuclei and Mitochondria Exemplified by Cytoplasmic Male Sterility. Plant Cell Physiol. **49**: 1484–1494.

Fujii, S. and Toriyama, K. (2009). Suppressed expression of retrograde-regulated male sterility restores pollen fertility in cytoplasmic male sterile rice plants. Proc. Natl. Acad. Sci. USA **106**: 9513–9518.

Gadjev, I., Vanderauwera, S., Gechev, T.S., Laloi, C., Minkov, I.N., Shulaev, V., Apel, K., Inzé, D., Mittler, R., and Van Breusegem, F. (2006). Transcriptomic footprints disclose specificity of reactive oxygen species signaling in Arabidopsis. Plant Physiol. 141: 436–445.

Galletto, R., Amitani, I., Baskin, R.J., and Kowalczykowski, S.C. (2006). Direct observation of individual RecA filaments assembling on single DNA molecules. Nature **443**: 875–878.

Garcia, L.E., Zubko, M.K., Zubko, E.I., and Sanchez-Puerta, M.V. (2019). Elucidating genomic patterns and recombination events in plant cybrid mitochondria. Plant Mol. Biol. **100**: 433–450.

García-Medel, P.L., Baruch-Torres, N., Peralta-Castro, A., Trasviña-Arenas, C.H., Torres-Larios, A., and Brieba, L.G. (2019). Plant organellar DNA polymerases repair double-stranded breaks by microhomology-mediated end-joining. Nucleic Acids Res. **47**: 3028–3044.

Garrison, E. and Gabor, M. (2012). Haplotype-based variant detection from short-read sequencing. ArXiv Prepr.: arXiv:1207.3907 [q-bio.GN].

Geisler, D.A., Päpke, C., Obata, T., Nunes-Nesi, A., Matthes, A., Schneitz, K., Maximova, E., Araújo, W.L., Fernie, A.R., and Persson, S. (2012). Downregulation of the δ-subunit reduces mitochondrial atp synthase levels, alters respiration, and restricts growth and gametophyte development in Arabidopsis. Plant Cell 24: 2792–2811.

Gellert, M., Mizuuchi, K., O'Dea, M.H., and Nash, H.A. (1976). DNA gyrase: an enzyme that introduces superhelical turns into DNA. Proc. Natl. Acad. Sci. USA **73**: 3872–3876.

Gerhold, J.M., Aun, A., Sedman, T., Jõers, P., and Sedman, J. (2010). Strand invasion structures in the inverted repeat of *Candida albicans* mitochondrial DNA reveal a role for homologous recombination in replication. Mol. Cell **39**: 851–861.

Gerhold, J.M., Sedman, T., Visacka, K., Slezakova, J., Tomaska, L., Nosek, J., and Sedman, J. (2014). Replication intermediates of the linear mitochondrial DNA of *Candida parapsilosis* suggest a common recombination based mechanism for yeast mitochondria. J. Biol. Chem. **289**: 22659–22670.

Giege, P. and Brennicke, A. (1999). RNA editing in Arabidopsis mitochondria effects 441 C to U changes in ORFs. Proc. Natl. Acad. Sci. USA **96**: 15324–15329.

Giegé, P., Sweetlove, L.J., Cognat, V., and Leaver, C.J. (2005). Coordination of nuclear and mitochondrial genome expression during mitochondrial biogenesis in Arabidopsis. Plant Cell **17**: 1497–1512.

Giesbers, A.K.J., Boer, E. den, Braspenning, D.N.J., Bouten, T.P.H., Specken, J.W., van Kaauwen, M.P.W., Visser, R.G.F., Niks, R.E., and Jeuken, M.J.W. (2018). Bidirectional backcrosses between wild and cultivated lettuce identify loci involved in nonhost resistance to downy mildew. Theor. Appl. Genet. 131: 1761–1776.

Gilles, L.M., Martinant, J.-P., Rogowsky, P.M., and Widiez, T. (2017). Haploid induction in plants. Curr. Biol. **27**: R1095–R1097.

Gissi, C., Gullberg, A., and Arnason, U. (1998). The complete mitochondrial DNA sequence of the Rabbit, *Oryctolagus cuniculus*. Genomics **50**: 161–169.

Gobron, N., Waszczak, C., Simon, M., Hiard, S., Boivin, S., Charif, D., Ducamp, A., Wenes, E., and Budar, F. (2013). A cryptic cytoplasmic male sterility unveils a possible gynodioecious past for *Arabidopsis thaliana*. PLoS ONE **8**: e62450.

Golczyk, H., Greiner, S., Wanner, G., Weihe, A., Bock, R., Börner, T., and Herrmann, R.G. (2014). Chloroplast DNA in mature and senescing leaves: A reappraisal. Plant Cell **26**: 847–854.

Grienenberger, E., Besseau, S., Geoffroy, P., Debayle, D., Heintz, D., Lapierre, C., Pollet, B., Heitz, T., and Legrand, M. (2009). A BAHD acyltransferase is expressed in the tapetum of Arabidopsis anthers and is involved in the synthesis of hydroxycinnamoyl spermidines. Plant J. **58**: 246–259.

Gualberto, J.M. and Kühn, K. (2014). DNA-binding proteins in plant mitochondria: Implications for transcription. Mitochondrion **19**: 323–328.

Gualberto, J.M. and Newton, K.J. (2017). Plant mitochondrial genomes: Dynamics and mechanisms of mutation. Annu. Rev. Plant Biol. **68**: 225–252.

Guo, F., Hu, S.-Y., Yuan, Z., Zee, S.-Y., and Han, Y. (2005). Paternal cytoplasmic transmission in Chinese pine (Pinus tabulaeformis). Protoplasma **225**: 5–14.

Guo, F.L. and Hu, S.Y. (1995). Cytological evidence of biparental inheritance of plastids and mitochondria inPelargonium. Protoplasma **186**: 201–207.

Guo, J.-X. and Liu, Y.-G. (2012). Molecular control of male reproductive development and pollen fertility in rice. J. Integr. Plant Biol. **54**: 967–978.

Hajnóczky, G., Csordás, G., Das, S., Garcia-Perez, C., Saotome, M., Sinha Roy, S., and Yi, M. (2006). Mitochondrial calcium signalling and cell death: Approaches for assessing the role of mitochondrial Ca2+ uptake in apoptosis. Cell Calcium **40**: 553–560.

Hammani, K., Gobert, A., Hleibieh, K., Choulier, L., Small, I., and Giegé, P. (2011). An *Arabidopsis* Dual-Localized Pentatricopeptide Repeat Protein Interacts with Nuclear Proteins Involved in Gene Expression Regulation. Plant Cell **23**: 730–740.

Handa, H. (2003). The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (*Brassica napus* L.): comparative analysis of the mitochondrial genomes of rapeseed and Arabidopsis thaliana. Nucleic Acids Res. **31**: 5907–5916.

Harlan, J.R. (1986). Lettuce and the Sycomore: sex and romance in ancient Egypt. Econ. Bot. **40**: 4–15.

Havey, M.J., Park, Y.H., and Bartoszewski, G. (2004). The Psm locus controls paternal sorting of the Cucumber mitochondrial genome. J. Hered. **95**: 492–497.

Hayes, R.J., Trent, M.A., Truco, M.J., Antonise, R., Michelmore, R.W., and Bull, C.T. (2014). The inheritance of resistance to bacterial leaf spot of lettuce caused by Xanthomonas campestris pv. vitians in three lettuce cultivars. Hortic. Res. **1**: 14066.

He, J., Mao, C.C., Reyes, A., Sembongi, H., Di Re, M., Granycome, C., Clippingdale, A.B., Fearnley, I.M., Harbour, M., Robinson, AJ., Reichelt, S., Spelbrink, J.N., Walker, J.E., Holt, IJ. (2007). The AAA+ protein ATAD3 has displacement loop binding properties and is involved in mitochondrial nucleoid organization. J. Cell Biol. 176: 141–146.

Hedtke, B., Börner, T., and Weihe, A. (1997). Mitochondrial and chloroplast phage-type RNA Polymerases in Arabidopsis. Science **277**: 809–811.

Hernandez, J.A., Corpas, F.J., Gomez, M., Rio, L.A., and Sevilla, F. (1993). Salt-induced oxidative stress mediated by activated oxygen species in pea leaf mitochondria. Physiol. Plant. **89**: 103–110.

Hniličková, H., Hnilička, F., Orsák, M., and Hejnák, V. (2019). Effect of salt stress on growth, electrolyte leakage, Na^+ and K^+ content in selected plant species. Plant Soil Environ. **65**: 90–96.

Hogeboom, G.H., Claude, A.C., and Hotchkiss, R.D. (1946). The distribution of cytochrome oxidase and succinoxidase in the cytoplasms of the mammalian liver cell. J. Biol. Chem. **165**: 615–629.

Honys, D. and Twell, D. (2004). Transcriptome analysis of haploid male gametophyte development in Arabidopsis. Genome Biol. **5**: R85.

Horlow, C., Defrance, M.C., Pollien, J.M., Goujaud, J., Delon, R., and Pelletier, G. (1993). Transfer of cytoplasmic male sterility by spontaneous androgenesis in tobacco (*Nicotiana tabacum* L.). Euphytica **66**: 45–53.

Horn, R. (2002). Molecular diversity of male sterility inducing and male-fertile cytoplasms in the genus *Helianthus*. Theor. Appl. Genet. **104**: 562–570.

Horn, R., Kusterer, B., Lazarescu, E., Prüfe, M., and Friedt, W. (2003). Molecular mapping of the *Rf1* gene restoring pollen fertility in PET1-based F1 hybrids in sunflower (*Helianthus annuus* L.). Theor. Appl. Genet. **106**: 599–606.

Ikeda, T.M. and Gray, M.W. (1999). Characterization of a DNA-binding protein implicated in transcription in Wheat mitochondria. Mol. Cell. Biol. **19**: 8113–8122.

Inoue, M., Fukui, K., Fujii, Y., Nakagawa, N., Yano, T., Kuramitsu, S., and Masui, R. (2017). The Lon protease-like domain in the bacterial RecA paralog RadA is required for DNA binding and repair. J. Biol. Chem. **292**: 9801–9814.

Janicka, S., Kühn, K., Le Ret, M., Bonnard, G., Imbault, P., Augustyniak, H., and Gualberto, J.M.

(2012). A RAD52-like single-stranded DNA binding protein affects mitochondrial DNA repair by recombination. Plant J. **72**: 423–435.

Janska, H., Sarria, R., Woloszynska, M., Arrieta-Montiel, M., and Mackenzie, S.A. (1998). Stoichiometric shifts in the common bean mitochondrial genome leading to male sterility and spontaneous reversion to fertility. Plant Cell **10**: 1163–1180.

Ji, J., Huang, W., Yin, C., and Gong, Z. (2013). Mitochondrial cytochrome c oxidase and f1fo-atpase dysfunction in peppers (Capsicum annuum L.) with cytoplasmic male sterility and its association with *orf507* and *Ψatp6*-2 Genes. Int. J. Mol. Sci. **14**: 1050–1068.

Jing, B., Heng, S., Tong, D., Wan, Z., Fu, T., Tu, J., Ma, C., Yi, B., Wen, J., and Shen, J. (2012). A male sterility-associated cytotoxic protein ORF288 in *Brassica juncea* causes aborted pollen development. J. Exp. Bot. **63**: 1285–1295.

Jõers, P. and Jacobs, H.T. (2013). Analysis Of Replication Intermediates Indicates That *Drosophila melanogaster* mitochondrial DNA replicates by a strand-coupled theta mechanism. PLoS ONE **8**: e53249.

Johns, C., Lu, M., Lyznik, A., and Mackenzie, S. (1992). A mitochondrial DNA sequence is associated with abnormal pollen development in cytoplasmic male sterile bean plants. Plant Cell **4**: 435–449.

Johnson, M.P. (2016). Photosynthesis. Essays Biochem. **60**: 255–273.

Kamoun, S., Furzer, O., Jones, J.D., Judelson, H.S., Ali, G.S., Dalio, R.J., Roy, S.G., Schena, L., Zambounis, A., Panabières, F., Cahill, D., Ruocco, M., Figueiredo, A., Chen, X.R., Hulvey, J., Stam, R., Lamour, K., Gijzen, M., Tyler, B.M., Grünwald, N.J., Mukhtar, M.S., Tomé, D.F., Tör, M., Van Den, Ackerveken, G., McDowell, J., Daayf, F., Fry, W.E., Lindqvist-Kreuze, H., Meijer, H.J., Petre, B., Ristaino, J., Yoshida, K., Birch, PR., and Govers, F. (2015). The Top 10 oomycete pathogens in molecular plant pathology: Top 10 oomycete plant pathogens. Mol. Plant Pathol. **16**: 413–434.

Kanamoto, H., Yamashita, A., Asao, H., Okumura, S., Takase, H., Hattori, M., Yokota, A., and Tomizawa, K.-I. (2006). Efficient and stable transformation of *Lactuca sativa* L. cv. *Cisco* (lettuce) Plastids. Transgenic Res. **15**: 205–217.

Kaneko, T. (2000). complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. DNA Res. **7**: 331–338.

Karnkowska, A., Vacek, V., Zubáčová, Z., Treitli, S.C., Petrželková, R., Eme, L., Novák, L., Žárský, V., Barlow, L.D., Herman, E.K., Soukal, P., Hroudová, M., Doležal, P., Stairs, C.W., Roger, A.J., Eliáš, M., Dacks, JB., Vlček, Č., and Hampl, V. (2016). A Eukaryote without a mitochondrial organelle. Curr. Biol. 26: 1274–1284.

Kasashima, K., Sumitani, M., Satoh, M., and Endo, H. (2008). Human prohibitin 1 maintains the organization and stability of the mitochondrial nucleoids. Exp. Cell Res. **314**: 988–996.

Katinka, M.D., Duprat, S., Cornillot, E., Méténier, G., Thomarat, F., Prensier, G., Barbe, V., Peyretaillade, E., Brottier, P., Wincker, P., Delbac, F., El, Alaoui, H., Peyret, P., Saurin, W., Gouy, M., Weissenbach, J., and Vivarès, C.P. (2001). Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. Nature 414: 450–453.

Kaufman, B.A., Durisic, N., Mativetsky, J.M., Costantino, S., Hancock, M.A., Grutter, P., and Shoubridge, E.A. (2007). The mitochondrial transcription factor TFAM coordinates the assembly of multiple DNA molecules into nucleoid-like structures. Mol. Biol. Cell **18**: 3225–3236.

Kaul, M.L.H. (1988). Male sterility in higher plants (Springer Berlin Heidelberg).

Kawanabe, T., Ariizumi, T., Kawai-Yamada, M., Uchimiya, H., and Toriyama, K. (2006). Abolition of the tapetum suicide program ruins microsporogenesis. Plant Cell Physiol. **47**: 784–787.

Kazama, T., Okuno, M., Watari, Y., Yanase, S., Koizuka, C., Tsuruta, Y., Sugaya, H., Toyoda, A., Itoh, T., Tsutsumi, N., Toriyama, K., Koizuka, N., and Arimura, S.I. (2019). Curing cytoplasmic

male sterility via TALEN-mediated mitochondrial genome editing. Nat. Plants 5: 722–730.
Kemble, R.J. and Thompson, R.D. (1982). S1 and S2, the linear mitochondial DNAs present in a male sterile line of maize, possess terminally attached proteins. Nucleic Acids Res. **10**: 8181–8190.
Kennedy, E.P. and Lehninger, A.L. (1949). Oxidation of fatty acids and tricarboxylic acid cycle intermediates by isolated rat liver mitochondria. J. Biol. Chem. **179**: 957–972.
Kennell, J.C. and Pring, D.R. (1989). Initiation and processing of atp6, T-urfl3 and ORF221 transcripts from mitochondria of T-cytoplasm maize. Mol. Gen. Genet. **216**: 16–24.
Kesseli, R., Ochoa, O., and Michelmore, R. (1991). Variation at RFLP loci in *Lactuca* spp. and origin of cultivated lettuce (*L. sativa*). Genome **34**: 430–436.
Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods **12**: 357–360.
Kim, Y.-J. and Zhang, D. (2018). Molecular control of male fertility for crop hybrid breeding. Trends Plant Sci. **23**: 53–65.
Kmiec, B., Woloszynska, M., and Janska, H. (2006). Heteroplasmy as a common state of mitochondrial genetic information in plants and animals. Curr. Genet. **50**: 149–159.
Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., McLellan, M.D., Lin, L., Miller, C.A., Mardis, E.R., Ding, L., and Wilson, R.K. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 22: 568–576.
Köhler, R., Horn, R., Lössl, A., and Zetsche, K. (1991). Cytoplasmic male sterility in sunflower is correlated with the co-transcription of a new open reading frame with the atpA gene. Mol. Gen. Genet. MGG **227**: 369–376.
Koning, A.J., Lum, P.Y., Williams, J.M., and Wright, R. (1993). DiOC6 staining reveals organelle structure and dynamics in living yeast cells. Cell Motil. Cytoskeleton **25**: 111–128.
Koopman, W.J.M. and De Jong, J.H. (1996). A numerical analysis of karyotypes and DNA amounts in lettuce cultivars and species (*Lactuca* subsect. *Lactuca*, Compositae). Acta Bot. Neerlandica **45**: 211–222.
Koopman, W.J.M., Guetta, E., van de Wiel, C.C.M., Vosman, B., and van den Berg, R.G. (1998). Phylogenetic relationships among *Lactuca* (Asteraceae) species and related genera based on ITS-1 DNA sequences. Am. J. Bot. **85**: 1517–1530.
Koopman, W.J.M., Zevenbergen, M.J., and Van den Berg, R.G. (2001). Species relationships in *Lactuca* s.l. (*Lactuceae*, *Asteraceae*) inferred from AFLP fingerprints. Am. J. Bot. **88**: 1881–1887.
Korhonen, J.A., Pham, X.H., Pellegrini, M., and Falkenberg, M. (2004). Reconstitution of a minimal mtDNA replisome in vitro. EMBO J. **23**: 2423–2429.
Korth, K.L., Kaspi, C.I., Siedow, J.N., and Levings, C.S. (1991). URF13, a maize mitochondrial pore-forming protein, is oligomeric and has a mixed orientation in *Escherichia coli* plasma membranes. Proc. Natl. Acad. Sci. USA **88**: 10865–10869.
Korth, K.L. and Levings, C.S. (1993). Baculovirus expression of the maize mitochondrial protein URF13 confers insecticidal activity in cell cultures and larvae. Proc. Natl. Acad. Sci. USA **90**: 3388–3392.
Kotera, E., Tasaka, M., and Shikanai, T. (2005). A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. Nature **433**: 326–330.
Kozik, A., Rowan, B.A., Lavelle, D., Berke, L., Schranz, M.E., Michelmore, R.W., and Christensen, A.C. (2019). The alternative reality of plant mitochondrial DNA: One ring does not rule them all. PLoS Genet. **15**: e1008373.
Křístková, E., Doležalová, I., Lebeda, A., Vinter, V., and Novotná, A. (2008). Description of morphological characters of lettuce (*Lactuca sativa* L.) genetic resources. Hortic. Sci. **35**: 113–129.

Kuhn, S., Bussemer, J., Chigri, F., and Vothknecht, U.C. (2009). Calcium depletion and calmodulin inhibition affect the import of nuclear-encoded proteins into plant mitochondria. Plant J. **58**: 694–705.

Kukat, C., Davies, K.M., Wurm, C.A., Spåhr, H., Bonekamp, N.A., Kühl, I., Joos, F., Polosa, PL., Park, C.B., Posse, V., Falkenberg, M., Jakobs, S., Kühlbrandt, W., and Larsson, N.G (2015). Cross-strand binding of TFAM to a single mtDNA molecule forms the mitochondrial nucleoid. Proc. Natl. Acad. Sci. USA 112: 11288–11293.

Kukat, C., Wurm, C.A., Spahr, H., Falkenberg, M., Larsson, N.-G., and Jakobs, S. (2011). Super-resolution microscopy reveals that mammalian mitochondrial nucleoids have a uniform size and frequently contain a single copy of mtDNA. Proc. Natl. Acad. Sci. USA **108**: 13534–13539.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Mol. Biol. Evol. **35**: 1547–1549.

Lamattina, L., Gonzalez, D., Gualberto, J., and Grienenberger, J.-M. (1993). Higher plant mitochondria encode an homologue of the nuclear-encoded 30-kDa subunit of bovine mitochondrial complex I. Eur. J. Biochem. **217**: 831–838.

Landgren, M., Zetterstrand, M., Sundberg, E., and Glimelius, K. (1996). Alloplasmic male-sterile *Brassica* lines containing *B. tournefortii* mitochondria express an ORF 3' of the atp6 gene and a 32 kDa protein. Plant Mol. Biol. **32**: 879–890.

Lardy, H.A. and Wellman, H. (1952). Oxidative phosphorylations; role of inorganic phosphate and acceptor systems in control of metabolic rates. J. Biol. Chem. **195**: 215–224.

Laver, H.K., Reynolds, S.J., Moneger, F., and Leaver, C.J. (1991). Mitochondrial genome organization and expression associated with cytoplasmic male sterility in sunflower (*Helianthus annuus*). Plant J. **1**: 185–193.

Law, S.R., Narsai, R., Taylor, N.L., Delannoy, E., Carrie, C., Giraud, E., Millar, A.H., Small, I., and Whelan, J. (2012). Nucleotide and RNA Metabolism Prime Translational Initiation in the Earliest Events of Mitochondrial Biogenesis during Arabidopsis Germination. Plant Physiol. **158**: 1610–1627.

Le Ret, M., Belcher, S., Graindorge, S., Wallet, C., Koechler, S., Erhardt, M., Williams-Carrier, R., Barkan, A., and Gualberto, J.M. (2018). An organellar thymidine kinase is required for the efficient replication of the maize plastidial genome. Plant Physiol.: pp.00976.2018.

Lebeda, A., Dolezalová, I., Feráková, V., and Astley, D. (2004). Geographical Distribution of Wild Lactuca Species (Asteraceae, Lactuceae). Bot. Rev. **70**: 328–356.

Legros, F., Malka, F., Frachon, P., Lombès, A., and Rojo, M. (2004). Organization and dynamics of human mitochondrial DNA. J. Cell Sci. **117**: 2653–2662.

Leon, P., Arroyo, A., and Mackenzie, S. (1998). Nuclear control of plastid and mitochondrial development in higher plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. **49**: 453–480.

Levings, C.S. (1993). Thoughts on cytoplasmic male sterility in CMS-T Maize. Plant Cell **5**: 1285–1290.

Levings, C.S. and Sederoff, R.R. (1983). Nucleotide sequence of the S-2 mitochondrial DNA from the S cytoplasm of maize. Proc. Natl. Acad. Sci. USA U.S.A. **80**: 4055–4059.

Lewis, D. (1941). Male sterility in natural populations of hermaphrodite plants. the equilibrium between females and hermaphrodites to be expected with different types of inheritance. New Phytol. **40**: 56–63.

Lewis, S.C., Joers, P., Willcox, S., Griffith, J.D., Jacobs, H.T., and Hyman, B.C. (2015). A rolling circle replication mechanism produces multimeric lariats of mitochondrial dna in *Caenorhabditis elegans*. PLOS Genet. **11**: e1004985.

Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinforma. Oxf. Engl. **25**: 1754–1760.

Li, S., Wan, C., Kong, J., Zhang, Z., Li, Y., and Zhu, Y. (2004). Programmed cell death during microgenesis in a Honglian CMS line of rice is correlated with oxidative stress in mitochondria. Funct. Plant Biol. **31**: 369–376.

Li, X.Q., Chetrit, P., Mathieu, C., Vedel, F., De Paepe, R., Remy, R., and Ambard-Bretteville, F. (1988). Regeneration of cytoplasmic male sterile protoclones of Nicotiana sylvestris with mitochondrial variations. Curr. Genet. **13**: 261–266.

Li, X.-Q., Jean, M., Landry, B.S., and Brown, G.G. (1998). Restorer genes for different forms of Brassica cytoplasmic male sterility map to a single nuclear locus that modifies transcripts of several mitochondrial genes. Proc. Natl. Acad. Sci. USA **95**: 10032–10037.

Lindqvist, K. (1960a). Inheritance studies in lettuce. Hereditas **46**: 387–470.

Lindqvist, K. (1960b). On the origin of cultivated lettuce. Hereditas **46**: 319–350.

Linke, B., Nothnagel, T., and Börner, T. (2003). Flower development in carrot CMS plants: mitochondria affect the expression of MADS box genes homologous to *GLOBOSA* and *DEFICIENS*. Plant J. **34**: 27–37.

Liu, F., Cui, X., Horner, H.T., Weiner, H., and Schnable, P.S. (2001). Mitochondrial aldehyde dehydrogenase activity is required for male fertility in maize. Plant Cell **13**: 1063–1078.

Liu, X., Kim, C.N., Yang, J., Jemmerson, R., and Wang, X. (1996). Induction of apoptotic program in cell-free extracts: Requirement for datp and cytochrome c. Cell **86**: 147–157.

Llopis, J., McCaffery, J.M., Miyawaki, A., Farquhar, M.G., and Tsien, R.Y. (1998). Measurement of cytosolic, mitochondrial, and Golgi pH in single living cells with green fluorescent proteins. Proc. Natl. Acad. Sci. USA **95**: 6803–6808.

Logan, D.C. (2006). The mitochondrial compartment. J. Exp. Bot. **57**: 1225–1243.

Losson, R. and Lacroute, F. (1979). Interference of nonsense mutations with eukaryotic messenger RNA stability. Proc. Natl. Acad. Sci. USA **76**: 5134–5137.

Lu, B., Wilson, R.K., Phreaner, C.G., Mulligan, R.M., and Hanson, M.R. (1996). Protein polymorphism generated by differential RNA editing of a plant mitochondrial rps12 gene. Mol. Cell. Biol. **16**: 1543–1549.

Luo, D., Xu, H., Liu, Z., Guo, J., Li, H., Chen, L., Fang, C., Zhang, Q., Bai, M., Yao, N., Wu, H., Wu, H., Ji, C., Zheng, H., Chen, Y., Ye, S., Li, X., Zhao, X., Li, R., and Liu, Y.G. (2013). A detrimental mitochondrial-nuclear interaction causes cytoplasmic male sterility in rice. Nat. Genet. 45: 573–577.

Lurin, C., Andrés, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyère, C., Caboche, M., Debast, C., Gualberto, J., Hoffmann, B., Lecharny, A., Le Ret, M., Martin-Magniette, M.L., Mireau, H., Peeters, N., Renou, JP., Szurek, B., Taconnat, L., and Small, I. (2004). Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. Plant Cell 16: 2089–2103.

Lusetti, S.L., Shaw, J.J., and Cox, M.M. (2003a). Magnesium ion-dependent activation of the RecA protein involves the C-Terminus. J. Biol. Chem. **278**: 16381–16388.

Lusetti, S.L., Wood, E.A., Fleming, C.D., Modica, M.J., Korth, J., Abbott, L., Dwyer, D.W., Roca, A.I., Inman, R.B., and Cox, M.M. (2003b). C-terminal deletions of the *Escherichia coli* RecA Protein: Characterization of in vivo and in vitro effects. J. Biol. Chem. 278: 16372–16380.

Mackenzie, S.A. and Chase, C.D. (1990). Fertility restoration is associated with loss of a portion of the mitochondrial genome in cytoplasmic male-sterile common bean. Plant Cell **2**: 905–912.

Marechal, A., Parent, J.-S., Veronneau-Lafortune, F., Joyeux, A., Lang, B.F., and Brisson, N. (2009). Whirly proteins maintain plastid genome stability in Arabidopsis. Proc. Natl. Acad. Sci. USA **106**: 14693–14698.

Marechal-Drouard, L., Guillemaut, P., Cosset, A., Arbogast, M., Weber, F., Weil, J.-H., and Dietrich, A. (1990). Transfer RNAs of potato (*Solanum tuberosum*) mitochondria have different

genetic origins. Nucleic Acids Res. **18**: 3689–3696.
Marie, L., Rapisarda, C., Morales, V., Bergé, M., Perry, T., Soulet, A.-L., Gruget, C., Remaut, H., Fronzes, R., and Polard, P. (2017). Bacterial RadA is a DnaB-type helicase interacting with RecA to promote bidirectional D-loop extension. Nat. Commun. 8: 15638.
Marienfeld, J., Unseld, M., and Brennicke, A. (1999). The mitochondrial genome of Arabidopsis is composed of both native and immigrant information. Trends Plant Sci. **4**: 495–502.
Marienfeld, R. and Newton, K.J. (1994). The maize ncs2 abnormal growth mutant has a chimeric *nad4-nad7* mitochondrial gene and is associated with reduced Complex I gunction. Genetics **138**: 855–863.
Martijn, J., Vosseberg, J., Guy, L., Offre, P., and Ettema, T.J.G. (2018). Deep mitochondrial origin outside the sampled alphaproteobacteria. Nature **557**: 101–105.
Mawer, J.S.P. and Leach, D.R.F. (2014). Branch migration prevents DNA loss during double-strand break repair. PLoS Genet. **10**: e1004485.
Maxwell, D.P., Wang, Y., and McIntosh, L. (1999). The alternative oxidase lowers mitochondrial reactive oxygen production in plant cells. Proc. Natl. Acad. Sci. USA **96**: 8271–8276.
Mayr, E. (1986). Joseph Gottlieb Kolreuter's contributions to biology. Osiris **2**: 135–176.
McRae, D.H. (1985). Advances in chemical hybridization. In Plant Breeding Reviews, J. Janick, ed (John Wiley & Sons, Inc.: Hoboken, NJ, USA), pp. 169–191.
Meisrimler, C., Pelgrom, A.J.E., Oud, B., Out, S., and Van den Ackerveken, G. (2019). Multiple downy mildew effectors target the stress-related NAC transcription factor Ls NAC 069 in lettuce. Plant J. **99**: 1098–1115.
Menand, B., Maréchal-Drouard, L., Sakamoto, W., Dietrich, A., and Wintz, H. (1998). A single gene of chloroplast origin codes for mitochondrial and chloroplastic methionyl-tRNA synthetase in Arabidopsis thaliana. Proc. Natl. Acad. Sci. USA **95**: 11014–11019.
Meyer, E.H., Tomaz, T., Carroll, A.J., Estavillo, G., Delannoy, E., Tanz, S.K., Small, I.D., Pogson, B.J., and Millar, A.H. (2009). Remodeled respiration in *ndufs4* with low phosphorylation efficiency suppresses arabidopsis germination and growth and alters control of metabolism at night. Plant Physiol. **151**: 603–619.
Michaelson, M.J., Price, H.J., Ellison, J.R., and Johnston, J.S. (1991). Comparison of plant DNA contents determined by Feulgen microspectrophotometry and laser flow cytometry. Am. J. Bot. **78**: 183–188.
Miller, I. and Bruns, E. (2016). The effect of disease on the evolution of females and the genetic basis of sex in populations with cytoplasmic male sterility. Proc. R. Soc. B Biol. Sci. **283**: 20153035.
Miller-Messmer, M., Kühn, K., Bichara, M., Le Ret, M., Imbault, P., and Gualberto, J.M. (2012). RecA-dependent dna repair results in increased heteroplasmy of the Arabidopsis mitochondrial genome. Plant Physiol. **159**: 211–226.
Mishra, P. and Chan, D.C. (2014). Mitochondrial dynamics and inheritance during cell division, development and disease. Nat. Rev. Mol. Cell Biol. **15**: 634–646.
Missio, J.C., Rivera, A., Figàs, M.R., Casanova, C., Camí, B., Soler, S., and Simó, J. (2018). A comparison of landraces vs. modern varieties of lettuce in organic farming during the winter in the mediterranean area: An approach considering the viewpoints of breeders, consumers, and farmers. Front. Plant Sci. **9**: 1491.
Morimatsu, K. and Kowalczykowski, S.C. (2003). RecFOR proteins load RecA protein onto gapped DNA to accelerate DNA strand exchange: A universal step of recombinational repair. Mol. Cell **11**: 1337–1347.
Moriyama, T. and Sato, N. (2014). Enzymes involved in organellar DNA replication in photosynthetic eukaryotes. Front. Plant Sci. **5**: 480.

Morley, S.A., Peralta-Castro, A., Brieba, L.G., Miller, J., Ong, K.L., Ridge, P.G., Oliphant, A., Aldous, S., and Nielsen, B.L. (2019). *Arabidopsis thaliana* organelles mimic the T7 phage DNA replisome with specific interactions between Twinkle protein and DNA polymerases Pol1A and Pol1B. BMC Plant Biol. **19**: 241.

Nakai, S., Noda, D., Kondo, M., and Terachi, T. (1995). High-level expression of a mitochondrial *orf522* gene from the male-sterile Sunflower is lethal to *E. coli*. Jpn. J. Breed. **45**: 233–236.

Negruk, V.I., Eisner, G.I., Redichkina, T.D., Dumanskaya, N.N., Cherny, D.I., Alexandrov, A.A., Shemyakin, M.F., and Butenko, R.G. (1986). Diversity of *Vicia faba* circular mtDNA in whole plants and suspension cultures. Theor. Appl. Genet. 72: 541–547.

Newton, K.J., Mariano, J.M., Gibson, C.M., Kuzmin, E., and Gabay-Laughnan, S. (1996). Involvement of S2 episomal sequences in the generation of NCS4 deletion mutation in maize mitochondria. Dev. Genet. **19**: 277–286.

Ng, S., De Clercq, I., Van Aken, O., Law, S.R., Ivanova, A., Willems, P., Giraud, E., Van Breusegem, F., and Whelan, J. (2014). Anterograde and retrograde regulation of nuclear genes encoding mitochondrial proteins during growth, development, and Stress. Mol. Plant **7**: 1075–1093.

Ngo, H.B., Kaiser, J.T., and Chan, D.C. (2011). The mitochondrial transcription and packaging factor Tfam imposes a U-turn on mitochondrial DNA. Nat. Struct. Mol. Biol. **18**: 1290–1296.

Nicolas, O., Charles, M.T., Jenni, S., Toussaint, V., and Beaulieu, C. (2018). Relationships between *Xanthomonas campestris* pv. *vitians* population sizes, stomatal density and lettuce resistance to bacterial leaf spot. Can. J. Plant Pathol. **40**: 399–407.

Nicolas, O., Charles, M.T., Jenni, S., Toussaint, V., Parent, S.-É., and Beaulieu, C. (2019). The Ionomics of Lettuce Infected by *Xanthomonas campestris* pv. *vitians*. Front. Plant Sci. **10**: 351.

Oda, K., Kohchi, T., and Ohyama, K. (1992). Mitochondrial DNA of *Marchantia polymorpha* as a single circular form with no incorporation of foreign DNA. Biosci. Biotechnol. Biochem. **56**: 132–135.

Odahara, M., Masuda, Y., Sato, M., Wakazaki, M., Harada, C., Toyooka, K., and Sekine, Y. (2015). RECG maintains plastid and mitochondrial genome stability by suppressing extensive recombination between short dispersed repeats. PLOS Genet. **11**: e1005080.

Odahara, M. and Sekine, Y. (2018). RECX Interacts with mitochondrial RECA to maintain mitochondrial genome stability. Plant Physiol. **177**: 300–310.

Oldenburg, D.J. and Bendich, A.J. (2015). DNA maintenance in plastids and mitochondria of plants. Front. Plant Sci. **6**: 883.

Oldenburg, D.J. and Bendich, A.J. (1996). Size and structure of replicating mitochondrial dna in cultured Tobacco cell. Plant Cell **8**: 447–461.

Oldenburg, D.J. and Bendich, A.J. (1998). The structure of mitochondrial DNA from the liverwort, Marchantia polymorpha. J. Mol. Biol. **276**: 745–758.

Ong, H.C. and Palmer, J.D. (2006). Pervasive survival of expressed mitochondrial *rps14* pseudogenes in grasses and their relatives for 80 million years following three functional transfers to the nucleus. BMC Evol. Biol. **6**: 55.

Paillard, M., Sederoff, R.R., and Levings, C.S. (1985). Nucleotide sequence of the S-1 mitochondrial DNA from the S cytoplasm of maize. EMBO J. **4**: 1125–1128.

Pakusch, A.-E., Kneusel, R.E., and Matern, U. (1989). S-adenosyl-l-methionine:trans-caffeoyl-coenzyme A 3-O-methyltransferase from elicitor-treated parsley cell suspension cultures. Arch. Biochem. Biophys. **271**: 488–494.

Palmer, J.D. and Herbon, L.A. (1988). Plant mitochondrial DNA evolved rapidly in structure, but slowly in sequence. J. Mol. Evol. **28**: 87–97.

Parent, J.-S., Lepage, E., and Brisson, N. (2011). Divergent Roles for the Two PolI-Like Organelle

DNA Polymerases of Arabidopsis. Plant Physiol. **156**: 254–262.

Parsons, H.L., Yip, J.Y.H., and Vanlerberghe, G.C. (1999). Increased respiratory restriction during phosphate-limited growth in transgenic Tobacco cells lacking alternative oxidase. Plant Physiol. **121**: 1309–1320.

Paszkiewicz, G., Gualberto, J.M., Benamar, A., Macherel, D., and Logan, D.C. (2017). Arabidopsis seed mitochondria are bioenergetically active immediately upon imbibition and specialize via biogenesis in preparation for autotrophic Growth. Plant Cell **29**: 109–128.

Pelgrom, A.J.E., Eikelhof, J., Elberse, J., Meisrimler, C.-N., Raedts, R., Klein, J., and Van den Ackerveken, G. (2019). Recognition of lettuce downy mildew effector BLR38 in *Lactuca serriola* LS102 requires two unlinked loci: Two loci mediate downy mildew effector recognition. Mol. Plant Pathol. **20**: 240–253.

Pelletier, G., Primard, C., Vedel, F., Chetrit, P., Remy, R., Rousselle, and Renard, M. (1983). Intergeneric cytoplasmic hybridization in cruciferae by protoplast fusion. Mol. Gen. Genet. MGG **191**: 244–250.

Peng, Y., Leung, H.C.M., Yiu, S.M., and Chin, F.Y.L. (2010). IDBA – A practical iterative de Bruijn graph de novo assembler. In Research in Computational Molecular Biology, B. Berger, ed (Springer Berlin Heidelberg: Berlin, Heidelberg), pp. 426–440.

Peralta-Castro, A., Baruch-Torres, N., and Brieba, L.G. (2017). Plant organellar DNA primase-helicase synthesizes RNA primers for organellar DNA polymerases using a unique recognition sequence. Nucleic Acids Res. **45**: 10764–10774.

Pohjoismäki, J.L., Holmes, J.B., Wood, S.R., Yang, M.Y., Yasukawa, T., Reyes, A., Bailey, L.J., Cluett, T.J., Goffart, S., Willcox, S., Rigby, R.E., Jackson, A.P., Spelbrink, J.N., Griffith, J.D., Crouch, R.J., Jacobs, H.T., and Holt, I.J. (2010). Mammalian mitochondrial DNA replication intermediates are essentially duplex but contain extensive tracts of RNA/DNA hybrid. J. Mol. Biol. 397: 1144–1155.

Pramateftaki, P.V., Kouvelis, V.N., Lanaridis, P., and Typas, M.A. (2006). The mitochondrial genome of the wine yeast *Hanseniaspora uvarum*: a unique genome organization among yeast/fungal counterparts. FEMS Yeast Res. **6**: 77–90.

Prasad, T.K., Anderson, M.D., and Stewart, C.R. (1994). Acclimation, Hydrogen Peroxide, and Abscisic Acid Protect Mitochondria against Irreversible Chilling Injury in Maize Seedlings. Plant Physiol. **105**: 619–627.

Preuten, T., Cincu, E., Fuchs, J., Zoschke, R., Liere, K., and Börner, T. (2010). Fewer genes than organelles: extremely low and variable gene copy numbers in mitochondria of somatic plant cells: Gene copy numbers in mitochondria. Plant J. **64**: 948–959.

Pring, D.R., Chen, W., Tang, H.V., Howad, W., and Kempken, F. (1998). Interaction of mitochondrial RNA editing and nucleolytic processing in the restoration of male fertility in sorghum. Curr. Genet. **33**: 429–436.

Quiñones, V., Zanlungo, S., Moenne, A., Gómez, I., Holuigue, L., Litvak, S., and Jordana, X. (1996). The *rpl5-rps14-cob* gene arrangement in *Solanum tuberosum*: *rps14* is a transcribed and unedited pseudogene. Plant Mol. Biol. **31**: 937–943.

Rabinovitz, M., Stulberg, M.P., and Boyer, P.D. (1951). The control of pyruvate oxidation in a cell-free rat heart preparation by phosphate acceptors. Science **114**: 641–642.

Raghunathan, S., Kozlov, A.G., Lohman, T.M., and Waksman, G. (2000). Structure of the DNA binding domain of *E. coli* SSB bound to ssDNA. Nat. Struct. Mol. Biol. **7**: 648–652.

Raghunathan, S., Ricard, C.S., Lohman, T.M., and Waksman, G. (1997). Crystal structure of the homo-tetrameric DNA binding domain of *Escherichia coli* single-stranded DNA-binding protein determined by multiwavelength x-ray diffraction on the selenomethionyl protein at 2.9-A resolution. Proc. Natl. Acad. Sci. USA **94**: 6652–6657.

Ragunathan, K., Liu, C., and Ha, T. (2012). RecA filament sliding on DNA facilitates homology search. eLife **1**: e00067.

Raja, M.R., Waterman, S.R., Qiu, J., Bleher, R., Williamson, P.R., and O'Halloran, T.V. (2013). A copper hyperaccumulation phenotype correlates with pathogenesis in *Cryptococcus neoformans*. Metallomics **5**: 363–371.

Rat Genome Sequencing Project Consortium (2004). Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature **428**: 493–521.

Reape, T.J. and McCabe, P.F. (2010). Apoptotic-like regulation of programmed cell death in plants. Apoptosis **15**: 249–256.

Rep, M. and Grivell, L.A. (1996). The role of protein degradation in mitochondrial function and biogenesis. Curr. Genet. **30**: 367–380.

Reyes-Chin-Wo, S., Wang, Z., Yang, X., Kozik, A., Arikit, S., Song, C., Xia, L., Froenicke, L., Lavelle, D.O., Truco, M.J., Xia, R., Zhu, S., Xu, C., Xu, H., Xu, X., Cox, K., Korf, I., Meyers, B.C., Michelmore, R.W. (2017). Genome assembly with in vitro proximity ligation data and whole-genome triplication in lettuce. Nat. Commun. 8: 14953.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. 29: 24–26.

Rocha, E.P.C., Cornet, E., and Michel, B. (2005). Comparative and evolutionary analysis of the bacterial homologous recombination systems. PLoS Genet. **1**: e15.

Roger, A.J., Svard, S.G., Tovar, J., Clark, C.G., Smith, M.W., Gillin, F.D., and Sogin, M.L. (1998). A mitochondrial-like chaperonin 60 gene in Giardia lamblia: Evidence that diplomonads once harbored an endosymbiont related to the progenitor of mitochondria. Proc. Natl. Acad. Sci. USA **95**: 229–234.

Roux, F., Mary-Huard, T., Barillot, E., Wenes, E., Botran, L., Durand, S., Villoutreix, R., Martin-Magniette, M.-L., Camilleri, C., and Budar, F. (2016). Cytonuclear interactions affect adaptive traits of the annual plant *Arabidopsis thaliana* in the field. Proc. Natl. Acad. Sci. USA 113: 3687–3692.

Rowan, B.A., Oldenburg, D.J., and Bendich, A.J. (2009). A multiple-method approach reveals a declining amount of chloroplast DNA during development in Arabidopsis. BMC Plant Biol. **9**: 3.

Rubio-Cosials, A., Sydow, J.F., Jiménez-Menéndez, N., Fernández-Millán, P., Montoya, J., Jacobs, H.T., Coll, M., Bernadó, P., and Solà, M. (2011). Human mitochondrial transcription factor A induces a U-turn structure in the light strand promoter. Nat. Struct. Mol. Biol. 18: 1281–1289.

Ruiz, O.N. and Daniell, H. (2005). Engineering cytoplasmic male sterility via the chloroplast genome by expression of {beta}-ketothiolase. Plant Physiol. **138**: 1232–1246.

Ryder, E.J. (1999). Lettuce, endive and chicory (CABI Publishin).

Sabar, M., Gagliardi, D., Balk, J., and Leaver, C.J. (2003). ORFB is a subunit of F_1F_O-ATP synthase: insight into the basis of cytoplasmic male sterility in sunflower. EMBO Rep. **4**: 381–386.

Sagan, L. (1986). On the Origin of Mitosing Cells. J. Theor. Biol. **14**: 225–274.

Sanders, P.M., Bui, A.Q., Weterings, K., McIntire, K.N., Hsu, Y.-C., Lee, P.Y., Truong, M.T., Beals, T.P., and Goldberg, R.B. (1999). Anther developmental defects in *Arabidopsis thaliana* male-sterile mutants. Sex. Plant Reprod. 11: 297–322.

Sandhu, A.P.S., Abdelnoor, R.V., and Mackenzie, S.A. (2007). Transgenic induction of mitochondrial rearrangements for cytoplasmic male sterility in crop plants. Proc. Natl. Acad. Sci. USA **104**: 1766–1770.

Sandler, S.J. and Clark, A.J. (1994). RecOR suppression of recF mutant phenotypes in Escherichia coli K-12. J. Bacteriol. **176**: 3661–3672.

Sarria, R., Lyznik, A., Vallejos, C.E., and Mackenzie, S.A. (1998). A cytoplasmic male sterility–

associated mitochondrial peptide in common bean is post-translationally regulated. Plant Cell **10**: 1217–1228.

Satoh, M. and Kuroiwa, T. (1991). Organization of multiple nucleoids and DNA molecules in mitochondria of a human cell. Exp. Cell Res. **196**: 137–140.

Schmitz-Linneweber, C., Williams-Carrier, R., and Barkan, A. (2005). RNA immunoprecipitation and microarray analysis show a chloroplast pentatricopeptide repeat protein to be associated with the 5′ region of mRNAs whose translation it activates. Plant Cell **17**: 2791–2804.

Schwarzländer, M., Logan, D.C., Johnston, I.G., Jones, N.S., Meyer, A.J., Fricker, M.D., and Sweetlove, L.J. (2012). Pulsing of membrane potential in individual mitochondria: A stress-induced mechanism to regulate respiratory bioenergetics in *Arabidopsis*. Plant Cell **24**: 1188–1201.

Scott, R.J. (2004). Stamen Structure and Function. Plant Cell **16**: S46–S60.

Scotti, N., Maréchal-Drouard, L., and Cardi, T. (2004). The rpl5 - rps14 mitochondrial region: a hot spot for DNA rearrangements in *Solanum spp*. somatic hybrids. Curr. Genet. **45**: 378–382.

Sheahan, M.B., Rose, R.J., and McCurdy, D.W. (2004). Organelle inheritance in plant cell division: the actin cytoskeleton is required for unbiased inheritance of chloroplasts, mitochondria and endoplasmic reticulum in dividing protoplasts: *Organelle inheritance requires actin*. Plant J. **37**: 379–390.

Shedge, V., Arrieta-Montiel, M., Christensen, A.C., and Mackenzie, S.A. (2007). Plant mitochondrial recombination surveillance requires unusual *RecA* and *MutS* homologs. Plant Cell **19**: 1251–1264.

Shen, L.L. and Pernet, A.G. (1985). Mechanism of inhibition of DNA gyrase by analogues of nalidixic acid: the target of the drugs is DNA. Proc. Natl. Acad. Sci. USA **82**: 307–311.

Shikanai, T. (2006). RNA editing in plant organelles: machinery, physiological function and evolution. Cell. Mol. Life Sci. **63**: 698–708.

Skippington, E., Barkman, T.J., Rice, D.W., and Palmer, J.D. (2015). Miniaturized mitogenome of the parasitic plant *Viscum scurruloideum* is extremely divergent and dynamic and has lost all *nad* genes. Proc. Natl. Acad. Sci. USA **112**: E3515–E3524.

Sloan, D.B. (2013). One ring to rule them all? Genome sequencing provides new insights into the 'master circle' model of plant mitochondrial DNA structure. New Phytol. **200**: 978–985.

Sloan, D.B., Alverson, A.J., Chuckalovcak, J.P., Wu, M., McCauley, D.E., Palmer, J.D., and Taylor, D.R. (2012). Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. PLoS Biol. **10**: e1001241.

Sloan, D.B., Wu, Z., and Sharbrough, J. (2018). correction of persistent errors in arabidopsis reference mitochondrial genomes. Plant Cell **30**: 525–527.

Small, I., Peeters, N., Legeai, F., and Lurin, C. (2004). Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. Proteomics **4**: 1581–1590.

Small, I., Suffolk, R., and Leaver, C.J. (1989). Evolution of plant mitochondrial genomes via substoichiometric intermediates. Cell **58**: 69–76.

Small, I.D. and Peeters, N. (2000). The PPR motif – a TPR-related motif prevalent in plant organellar proteins. Trends Biochem. Sci. **25**: 45–47.

Smith, D.R., Jackson, C.J., and Reyes-Prieto, A. (2014). Nucleotide substitution analyses of the glaucophyte Cyanophora suggest an ancestrally lower mutation rate in plastid vs mitochondrial DNA for the Archaeplastida. Mol. Phylogenet. Evol. **79**: 380–384.

Sodmergen, Zhang, Q., Zhang, Y., Sakamoto, W., and Kuroiwa, T. (2002). Reduction in amounts of mitochondrial DNA in the sperm cells as a mechanism for maternal inheritance in *Hordeum vulgare*. Planta **216**: 235–244.

Sousa, J.S., D'Imprima, E., and Vonck, J. (2018). Mitochondrial respiratory chain complexes. In

membrane protein complexes: structure and function, J.R. Harris and E.J. Boekema, eds (Springer Singapore: Singapore), pp. 167–227.

Stecher, G., Tamura, K., and Kumar, S. (2020). Molecular Evolutionary Genetics Analysis (MEGA) for macOS. Mol. Biol. Evol.: msz312.

Stewart, J.B. and Larsson, N.-G. (2014). Keeping mtDNA in Shape between Generations. PLoS Genet. **10**: e1004670.

Stohl, E.A., Brockman, J.P., Burkle, K.L., Morimatsu, K., Kowalczykowski, S.C., and Seifert, H.S. (2003). *Escherichia coli* RecX inhibits RecA recombinase and coprotease activities in vitro and in vivo. J. Biol. Chem. **278**: 2278–2285.

Stupar, R.M., Lilly, J.W., Town, C.D., Cheng, Z., Kaul, S., Buell, C.R., and Jiang, J. (2001). Complex mtDNA constitutes an approximate 620-kb insertion on Arabidopsis thaliana chromosome 2: Implication of potential sequencing errors caused by large-unit repeats. Proc. Natl. Acad. Sci. USA **98**: 5099–5103.

Su, C.J. and Baseman, J.B. (1990). Genome size of Mycoplasma genitalium. J. Bacteriol. **172**: 4705–4707.

Sumitani, M., Kasashima, K., Ohta, E., Kang, D., and Endo, H. (2009). Association of a novel mitochondrial protein M19 with mitochondrial nucleoids. J. Biochem. (Tokyo) **146**: 725–732.

Sweetlove, L.J., Heazlewood, J.L., Herald, V., Holtzapffel, R., Day, D.A., Leaver, C.J., and Millar, A.H. (2002). The impact of oxidative stress on Arabidopsis mitochondria. Plant J. 32: 891–904.

Takami, T., Ohnishi, N., Kurita, Y., Iwamura, S., Ohnishi, M., Kusaba, M., Mimura, T., and Sakamoto, W. (2018). Organelle DNA degradation contributes to the efficient use of phosphate in seed plants. Nat. Plants **4**: 1044–1055.

Takanashi, H., Ohnishi, T., Mogi, M., Okamoto, T., Arimura, S., and Tsutsumi, N. (2010). Studies of mitochondrial morphology and DNA amount in the rice egg cell. Curr. Genet. **56**: 33–41.

Takemura, M., Oda, K., Yamato, K., Ohta, E., Nakamura, Y., Nozato, N., Akashi, K., and Ohyama, K. (1992). Gene clusters for ribosomal proteins in the mitochondrial genome of a liverwort, *Marchantia polymorpha*. Nucleic Acids Res. **20**: 3199–3205.

Tang, D.G., La, E., Kern, J., and Kehrer, J.P. (2002). Fatty acid oxidation and signaling in apoptosis. Biol. Chem. **383**: 425–442.

Tang, H.V., Chen, W., and Pring, D.R. (1999). Mitochondrial *orf107* transcription, editing, and nucleolytic cleavage conferred by the gene *Rf3* are expressed in sorghum pollen. Sex. Plant Reprod. **12**: 53–59.

Tang, L.Y. and Sakamoto, W. (2011). Tissue-specific organelle DNA degradation mediated by DPD1 exonuclease. Plant Signal. Behav. **6**: 1391–1393.

Taylor, A.F., Amundsen, S.K., Guttman, M., Lee, K.K., Luo, J., Ranish, J., and Smith, G.R. (2014). Control of RecBCD enzyme activity by DNA binding- and Chi hotspot-dependent conformational changes. J. Mol. Biol. **426**: 3479–3499.

The Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature **408**: 796–815.

Theißen, G., Melzer, R., and Rümpler, F. (2016). MADS-domain transcription factors and the floral quartet model of flower development: linking plant development and evolution. Development **143**: 3259–3271.

Theißen, G. and Saedler, H. (2001). Floral quartets. Nature **409**: 469–471.

Thicoïpé, J.-P. (1997). Laitue CTIFL. (CTIFL).

Thompson, R.C. (1943). Futher studies on interspecific genetic relationships in Lactuca. J. Agric. Res. **66**: 41–48.

Thompson, R.C., Whitaker, T.W., and Kosar, W.F. (1941). Interspecific genetic relationships in

Lactuca. J. Agric. Res. **63**: 91–107.
Thrash, J.C., Boyd, A., Huggett, M.J., Grote, J., Carini, P., Yoder, R.J., Robbertse, B., Spatafora, J.W., Rappé, M.S., and Giovannoni, S.J. (2011). Phylogenomic evidence for a common ancestor of mitochondria and the SAR11 clade. Sci. Rep. 1: 13–13.
Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E.S., Fischer, A., Bock, R., and Greiner, S. (2017). GeSeq – versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 45: W6–W11.
Timmis, J.N., Ayliffe, M.A., Huang, C.Y., and Martin, W. (2004). Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat. Rev. Genet. **5**: 123–135.
Truco, M.J., Ashrafi, H., Kozik, A., van Leeuwen, H., Bowers, J., Wo, S.R.C., Stoffel, K., Xu, H., Hill, T., Van Deynze, A., and Michelmore, R.W. (2013). An ultra-high-density, transcript-based, genetic map of lettuce. G3 3: 617–631.
Umezu, K., Chi, N.W., and Kolodner, R.D. (1993). Biochemical interaction of the *Escherichia coli* RecF, RecO, and RecR proteins with RecA protein and single-stranded DNA binding protein. Proc. Natl. Acad. Sci. USA **90**: 3875–3879.
Umezu, K. and Kolodner, R.D. (1994). Protein interactions in genetic recombination in *Escherichia coli*. J. Biol. Chem. **269**: 30005–30013.
Unseld, M., Marienfeld, J.R., Brandt, P., and Brennicke, A. (1997). The mitochondrial genome of Arabidopsis thaliana contains 57 genes in 366,924 nucleotides. Nat. Genet. **15**: 57–61.
Uyttewaal, M., Arnal, N., Quadrado, M., Martin-Canadell, A., Vrielynck, N., Hiard, S., Gherbi, H., Bendahmane, A., Budar, F., and Mireau, H. (2008). Characterization of *raphanus sativus* pentatricopeptide repeat proteins encoded by the fertility restorer locus for Ogura cytoplasmic male sterility. Plant Cell 20: 3331–3345.
Vanlerberghe, G. (2013). Alternative oxidase: A mitochondrial respiratory pathway to maintain metabolic and signaling homeostasis during abiotic and biotic stress in plants. Int. J. Mol. Sci. **14**: 6805–6847.
Varré, D'Agostino, Touzet, Gallina, Tamburino, Cantarella, Ubrig, Cardi, Drouard, Gualberto, and Scotti (2019). complete sequence, multichromosomal architecture and transcriptome analysis of the *Solanum tuberosum* mitochondrial genome. Int. J. Mol. Sci. **20**: 4788.
Veaute, X., Delmas, S., Selva, M., Jeusset, J., Le Cam, E., Matic, I., Fabre, F., and Petit, M.-A. (2005). UvrD helicase, unlike Rep helicase, dismantles RecA nucleoprotein filaments in Escherichia coli. EMBO J. **24**: 180–189.
Veltri, K.L., Espiritu, M., and Singh, G. (1990). Distinct genomic copy number in mitochondria of different mammalian organs. J. Cell. Physiol. **143**: 160–164.
Veniamin, S., Sawatzky, L.G., Banting, G.S., and Glerum, D.M. (2011). Characterization of the peroxide sensitivity of COX-deficient yeast strains reveals unexpected relationships between COX assembly proteins. Free Radic. Biol. Med. **51**: 1589–1600.
Virmani, S.S. and Ilyas-Ahmed, M. (2001). Environment-sensitive genic male sterility (EGMS) in crops. In Advances in Agronomy (Elsevier), pp. 139–195.
Virolainen, E. (2002). Ca2+-induced high amplitude swelling and cytochrome c release from Wheat (*Triticum aestivum* L.) mitochondria under anoxic stress. Ann. Bot. **90**: 509–516.
Visser, W., van Spronsen, E.A., Nanninga, N., Pronk, J.T., Kuenen, J.G., and van Dijken, J.P. (1995). Effects of growth conditions on mitochondrial morphology in *Saccharomyces cerevisiae*. Antonie Van Leeuwenhoek **67**: 243–253.
de Vries, F. (1994). On genetically modified plants.
de Vries, I.M. (1990). Crossing experiments of lettuce cultivars and species (*Lactuca* sect. *Lactuca*, *Compositae*). Plant Syst. Evol. **171**: 233–248.
de Vries, I.M. (1997). Origin and domestication of *Lactuca sativa* L. Genet. Resour. Crop Evol. **44**:

165–174.
de Vries, I.M. and van Raamsdonk, L.W.D. (1994). Numerical morphological analysis of Lettuce cultivars and species (*Lactuca* sect. *Lactuca*, *Asteraceae*). Plant Syst. Evol. **193**: 125–141.
de Vries, R. (2010). DNA condensation in bacteria: Interplay between macromolecular crowding and nucleoid proteins. Biochimie **92**: 1715–1721.
Wall, M.K., Mitchenall, L.A., and Maxwell, A. (2004). *Arabidopsis thaliana* DNA gyrase is targeted to chloroplasts and mitochondria. Proc. Natl. Acad. Sci. USA **101**: 7821–7826.
Wallet, C., Le Ret, M., Bergdoll, M., Bichara, M., Dietrich, A., and Gualberto, J.M. (2015). The RECG1 DNA translocase is a key factor in recombination surveillance, repair, and segregation of the mitochondrial DNA in Arabidopsis. Plant Cell **27**: 2907–2925.
Wang, D.-Y., Zhang, Q., Liu, Y., Lin, Z.-F., Zhang, S.-X., Sun, M.-X., and Sodmergen (2010). The levels of male gametic mitochondrial DNA are highly regulated in angiosperms with regard to mitochondrial inheritance. Plant Cell **22**: 2402–2416.
Wang, H., Dittmer, T.A., and Richards, E.J. (2013). *Arabidopsis* CROWDED NUCLEI (CRWN) proteins are required for nuclear size control and heterochromatin organization. BMC Plant Biol. **13**: 200.
Wang, Y., Carrie, C., Giraud, E., Elhafez, D., Narsai, R., Duncan, O., Whelan, J., and Murcha, M.W. (2012). Dual location of the mitochondrial preprotein transporters B14.7 and Tim23-2 in complex I and the TIM17:23 complex in *Arabidopsis* links mitochondrial activity and biogenesis. Plant Cell **24**: 2675–2695.
Wang, Y., Yang, R., Zheng, J., Shen, Z., and Xu, X. (2019). Exogenous foliar application of fulvic acid alleviate cadmium toxicity in lettuce (*Lactuca sativa* L.). Ecotoxicol. Environ. Saf. **167**: 10–19.
Wang, Z., Zou, Y., Li, X., Zhang, Q., Chen, L., Wu, H., Su, D., Chen, Y., Guo, J., Luo, D., Long, Y., Zhong, Y., Liu, Y.G. (2006). Cytoplasmic male sterility of Rice with Boro II cytoplasm is caused by a cytotoxic peptide and is restored by two related PPR Motif Genes via Distinct Modes of mRNA Silencing. Plant Cell **18**: 676–687.
Wardrope, L., Okely, E., and Leach, D. (2009). Resolution of joint molecules by RuvABC and RecG following cleavage of the *Escherichia coli* chromosome by EcoKI. PLoS ONE **4**: e6542.
Warren, J.M., Simmons, M.P., Wu, Z., and Sloan, D.B. (2016). Linear plasmids and the rate of sequence evolution in plant mitochondrial genomes. Genome Biol. Evol. **8**: 364–374.
Warren, R.L., Sutton, G.G., Jones, S.J.M., and Holt, R.A. (2006). Assembling millions of short DNA sequences using SSAKE. Bioinformatics **23**: 500–501.
Wei, Z., Zhu, S.-X., Van den Berg, R.G., Bakker, F.T., and Schranz, M.E. (2017). Phylogenetic relationships within *Lactuca* L. (*Asteraceae*), including African species, based on chloroplast DNA sequence comparisons. Genet. Resour. Crop Evol. **64**: 55–71.
West, S.C. (1997). Processing of recombination intermediates by the RuvABC proteins. Annu. Rev. Genet. **31**: 213–244.
Whitby, M.C., Vincent, S.D., and Lloyd, R.G. (1994). Branch migration of Holliday junctions: identification of RecG protein as a junction specific DNA helicase. EMBO J. **13**: 5220–5228.
Wijnker, E., van Dun, K., de Snoo, C.B., Lelivelt, C.L.C., Keurentjes, J.J.B., Naharudin, N.S., Ravi, M., Chan, S.W.L., de Jong, H., and Dirks, R. (2012). Reverse breeding in *Arabidopsis thaliana* generates homozygous parental lines from a heterozygous plant. Nat. Genet. 44: 467–470.
Wingett, S. and Andrews, S. (2018). FastQ Screen: A tool for multi-genome mapping and quality control. F1000Research **7**: 1338.
Wolfe, K.H., Li, W.H., and Sharp, P.M. (1987). Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc. Natl. Acad. Sci. USA **84**: 9054–9058.
Woloszynska, M. (2010). Heteroplasmy and stoichiometric complexity of plant mitochondrial

genomes--though this be madness, yet there's method in't. J. Exp. Bot. **61**: 657–671.
Woloszynska, M. and Trojanowski, D. (2009). Counting mtDNA molecules in *Phaseolus vulgaris*: sublimons are constantly produced by recombination via short repeats and undergo rigorous selection during substoichiometric shifting. Plant Mol. Biol. **70**: 511–521.
Woodson, J.D. and Chory, J. (2008). Coordination of gene expression between organellar and nuclear genomes. Nat. Rev. Genet. **9**: 383–395.
Wu, Z., Cuthbert, J.M., Taylor, D.R., and Sloan, D.B. (2015). The massive mitochondrial genome of the angiosperm *Silene noctiflora* is evolving by gain or loss of entire chromosomes. Proc. Natl. Acad. Sci. USA **112**: 10185–10191.
Wynn, E.L. and Christensen, A.C. (2019). Repeats of unusual size in plant mitochondrial genomes: identification, incidence and evolution. G3 **9**: 549–559.
Xu, C. and Mou, B. (2015). Evaluation of lettuce genotypes for salinity tolerance. HortScience **50**: 1441–1446.
Xu, L., Carrie, C., Law, S.R., Murcha, M.W., and Whelan, J. (2013). Acquisition, conservation, and loss of dual-targeted proteins in land plants. Plant Physiol. **161**: 644–662.
Xu, Y.-Z., Arrieta-Montiel, M.P., Virdi, K.S., de Paula, W.B.M., Widhalm, J.R., Basset, G.J., Davila, J.I., Elthon, T.E., Elowsky, C.G., Sato, S.J., Clemente, T.E., and Mackenzie, S.A. (2011). MutS HOMOLOG1 Is a nucleoid protein that alters mitochondrial and plastid properties and plant response to high light. Plant Cell 23: 3428–3441.
Yadav, A.K., Sahoo, P.K., Goswami, H.N., and Jain, D. (2019). Transcriptional tidelity of mitochondrial RNA polymerase RpoTm from Arabidopsis thaliana. J. Mol. Biol. **431**: 4767–4783.
Yang, Y., Fanning, L., and Jack, T. (2003). The K domain mediates heterodimerization of the Arabidopsis floral organ identity proteins, APETALA3 and PISTILLATA. Plant J. **33**: 47–59.
Zaegel, V., Guermann, B., Le Ret, M., Andrés, C., Meyer, D., Erhardt, M., Canaday, J., Gualberto, J.M., and Imbault, P. (2006). The plant-specific ssDNA binding protein OSB1 is involved in the stoichiometric transmission of mitochondrial DNA in *Arabidopsis*. Plant Cell **18**: 3548–3563.
Zampini, É., Lepage, É., Tremblay-Belzile, S., Truche, S., and Brisson, N. (2015). Organelle DNA rearrangement mapping reveals U-turn-like inversions as a major source of genomic instability in *Arabidopsis* and humans. Genome Res. **25**: 645–654.
Zerbino, D.R. and Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. **18**: 821–829.
Zhang, G., Lu, Y., Bharaj, T.S., Virmani, S.S., and Huang, N. (1997). Mapping of the Rf-3 nuclear fertility-restoring gene for WA cytoplasmic male sterility in rice using RAPD and RFLP markers. Theor. Appl. Genet. **94**: 27–33.

Annexes

Annexe.1 List of Genbank accessions used in this work

Genbank n°	Description	Organism
AP007232.1	cpDNA	*Lactuca sativa* var. *capitata* L. *salinas*
JF729202.1	mtDNA	*Arabidopsis thaliana* ecotype Landsberg erecta
JF729200.1	mtDNA	*Arabidopsis thaliana* ecoytype C24
BK010421.1	mtDNA	*Arabidopsis thaliana* ecoytype Columbia-0
NC_012920.1	mtDNA	*Homo sapiens*
MK759657.1	mtDNA	*Lactuca saligna* accession CGN5271
MK642355.1	mtDNA	*Lactuca sativa* var. *capitata* L. *salinas*
MK820672.1	mtDNA	*Lactuca serriola* accession US96UC23
KT249801.1	ITS1, 5.8 rRNA, ITS2	*Lactuca sativa*
O49499	CCoAOMT1	*Arabidopsis thaliana*
BAG71891	CCoAOMT1	*Carthamus tinctorius*
OXA61824	CCoAOMT1	*Folsomia candida*
KLP16479.1	CCoAOMT1	*Fusarium fujikuroi*
XP_021998984	CCoAOMT1	*Heliantus annuus*
XP_023737725	CCoAOMT1	*Lactuca sativa*
AAC28973	CCoAOMT1	*Medigo sativa*
NP_001312329	CCoAOMT1	*Nicotiana benthamiana*
XP_015643224	CCoAOMT1	*Oryza sativa*
XP_024362766	CCoAOMT1	*Physcomitrella patens*
RVW55374	CCoAOMT1	*Vitis vinifera*
AAP37884	CCoAOMT1	*Zea mays*
NP_189075.2	MSH1	*Arabidopsis thaliana*
PLY73757.1	MSH1	*Lactuca sativa*
XP_024362883.1	MSH1	*Physcomitrella patens*
NP_011988.1	MSH1	*Saccharomyces cerevisiae*
EJK62659.1	MSH1	*Thalassiosira oceanica*
NP_001105898.1	MSH1	*Zea mays*
WP_052909840.1	MutS	*Escherichia coli*
WP_010872450.1	MutS	*Synechocystis sp.*
XP_006838874.1	OSB1	*Amborella trichopoda*
NP_175203.2	OSB1	*Arabidopsis thaliana*
XP_023728533.1	OSB1	*Lactuca sativa*
XP_015628693.1	OSB1	*Oryza sativa*
XP_002510278.1	OSB1	*Ricinus communis*
XP_010651930.1	OSB1	*Vitis vinifera A*
CBI19580.3	OSB1	*Vitis vinifera B*
ACG38623.1	OSB1	*Zea mays*
NP_568639.1	OSB2	*Arabidopsis thaliana*
XP_006370188.2	OSB2	*Populus trichocarpa*
CBI17867.3	OSB2	*Vitis vinifera*
PWZ10610.1	OSB2	*Zea mays*
XP_023761157.1	OSB3	*Arabidopsis thaliana*
XP_015622390.1	OSB3	*Oryza sativa*
NP_564370.1	OSB4	*Arabidopsis thaliana*

Genbank n°	Description	Organism
WP_017473696	RadA	*Amphibacillus jilinensis*
NP_842650	RadA	*Bacillus anthracis*
WP_001458566	RadA	*Escherichia coli*
WP_017296205	RadA	*Geminocystis herdmanii*
WP_002742386	RadA	*Microcystis aeruginosa*
WP_002258526	RadA	*Neisseria meningitidis*
WP_009633429	RadA	*Synechocystis sp*
NP_001056828	RADA	*Oryza sativa*
XP_003084463	RADA	*Ostreococcus tauri*
EFN56488	RADA	*Chlorella variabilis*
NP_199845	RADA	*Arabidopsis thaliana*
XP_005643141	RADA	*Coccomyxa subellipsoidea*
XP_005536634	RADA	*Cyanidioschyzon merolae*
CBJ25917	RADA	*Ectocarpus siliculosus*
XP_005709405	RADA	*Galdieria sulphuraria*
XP_023756175.1	RADA	*Lactuca sativa*
XP_002178713	RADA	*Phaeodactylum tricornutum*
XP_001757578	RADA	*Physcomitrella patens*
XP_002904225	RADA	*Phytophthora infestans*
EEE84551	RADA	*Populus trichocarpa*
EQC30001	RADA	*Saprolegnia diclina*
XP_002976563	RADA	*Selaginella moellendorffii*
EJK58798	RADA	*Thalassiosira oceanica*
XP_002277638	RADA	*Vitis vinifera*
NP_001170708	RADA	*Zea mays*
WP_095252583.1	RecA	*Bacillus subtilis*
ABJ02129.1	RecA	*Escherichia coli*
NP_565198.1	RECA1	*Arabidopsis thaliana*
XP_023736817.1	RECA1	*Lactuca sativa*
XP_015627776.1	RECA1	*Oryza sativa*
XP_002271727.1	RECA1	*Vitis vinifera*
PWZ58964.1	RECA1	*Zea mays*
NP_179539.2	RECA2	*Arabidopsis thaliana*
XP_023771242.1	RECA2	*Lactuca sativa*
XP_015612195.1	RECA2	*Oryza sativa*
XP_002279464.2	RECA2	*Vitis vinifera*
NP_001170211.1	RECA2	*Zea mays*
NP_187625.2	RECA3	*Arabidopsis thaliana*
EEE55988.1	RECA3	*Oryza sativa*
WP_010874337.1	RECA3	*Synechocystis sp.*
XP_002278976.2	RECA3	*Vitis vinifera*
NP_001358502.1	RECA3	*Zea mays*
AKK14591.2	RecG	*Escherichia coli*
WP_010873434.1	RecG	*Synechocystis sp.*
NP_178253.3	RECG1	*Arabidopsis thaliana*
XP_023764076.1	RECG1	*Lactuca sativa*
XP_024356452.1	RECG1	*Physcomitrella patens*
EFJ11466.1	RECG1	*Selaginella moellendorffii*
CBI26906.3	RECG1	*Vitis vinifera*
XP_008646149.1	RECG1	*Zea mays*

Annexe.2 List of primers used in this work

Subcellular localization by GFP-fusion		
Name	Sequence	Function
OSB1-F	TAGGGTACCAAGTCCACAACCTCAGCACAC	Restriction clonning of OSB1
OSB1-R	GATGGATCCATAGGGGAATTGACTCTGCCAATG	Restriction clonning of OSB1
OSB2-F	TAGGGTACCCACACACACACACTCTCC	Restriction clonning of OSB2
OSB2-R	GATGGATCCAATTGAGAAATGATAGTGCCGG	Restriction clonning of OSB2
RADA-F	TAGGGTACCAAACCCTAAACTTCTGTCCC	Restriction clonning of RADA
RADA-R	GATGGATCCGAGATCCCTCTGTTAACATCCG	Restriction clonning of RADA
WHY2-F	TAGGGTACCGTCCTCACACGGTACACAATG	Restriction clonning of WHY2
WHY2-R	GATGGATCCGGAAGAACAGGGGCAACAGAA	Restriction clonning of WHY2
RECA3-GW-F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGAGTTCAAACATTCCTTTC	Gateway cloning of RECA3
RECA3-GW-R	GGGGACCACTTTGTACAAGAAAGCTGGGTATTCCTCTTGCCCATAAACCTCAAC	Gateway cloning of RECA3
Quantification of mtDNA enrichment		
Name	Sequence	Function
N18S-1F	ACTCCGCTGGCACCTTATGAG	qPCR quantification of the nuclear genome
N18S-1R	GTGGTGCCCTTCCGTCAATTC	qPCR quantification of the nuclear genome
mt18S-1F	CGAGTGCGCGATCATGACAAG	qPCR quantification of the mtDNA
mt18S-1R	CGCCCAGTCATTCCGAAGAAC	qPCR quantification of the mtDNA
NAD3-1F	TCGGTGTTCCTTTTCCATTTGC	qPCR quantification of the mtDNA
NAD3-1R	AAAACGACTTCTGGCATCACCG	qPCR quantification of the mtDNA
rbcL-12-F	CAGTTCGGTGGAGGAACTTTAG	qPCR quantification of the cpDNA
rbcL-12-R	GCAAGATCGCGTCCCTCATTAC	qPCR quantification of the cpDNA
Genotyping of *CCoAOMT1*		
Name	Sequence	Function
CCoAOMT1-F	CATGACCTAATTTCCAGAGGTAGAG	Genotyping of CCoAOMT1
CCoAOMT1-R	AGGTCTAGAACAGGAAGTGCAGGAC	Genotyping of CCoAOMT1
CCoAOMT1-S	GCCAGTTTTTGAACCTTCTTC	Sanger sequencing for genotyping of CCoAOMT1
Primers specific to *OSB2*		
Name	Sequence	Function
OSB2-F	ATACGTCGAGCTATCACTGC	RT-PCR of the intron #1
OSB2w-R	TTCTTAGTGCTCCACCTATCC	RT-PCR of the intron #1
OSB2m1-RTPrimer-R	TCAAGTGAGGGAATAGTAG	Reverse primer for Retro-transcription
OSB2m2-RTPrimer-R	ACGCGAGCCAATATCAATATC	Reverse primer for Retro-transcription
OSB2w-RTPrimer-R	CTAAACCCCCCATCCCTTTTTCG	Reverse primer for Retro-transcription
OSB2-p-MUT	CTCTACCGTCTTTCTGATACTAC	Reverse primer for Retro-transcription
OSB2-RT-e1_e2-F	CACTATCATTTCTCAAGATG	qRT-PCR for the quantification of the splicing of the intron #1 of *OSB2 transcript*
OSB2-RT-e1_e2-R	CCAGCTACATGTACAAAATC	qRT-PCR for the quantification of the splicing of the intron #1 of *OSB2 transcript*
OSB2-RT-e2_e3-F	CTACTTCTACACCATACAAGC	qRT-PCR for the quantification of the splicing of the intron #2 of *OSB2 transcript*
OSB2-RT-e2_e3-R	GCATCATCATCAAGTAGCAGAC	qRT-PCR for the quantification of the splicing of the intron #2 of *OSB2 transcript*
OSB2-RT-e3_e4-F	CCCAAGTCCATAATAACTC	qRT-PCR for the quantification of the splicing of the intron #3 of *OSB2 transcript*
OSB2-RT-e3_e4-R	CGGTTATCATCCCATTGTGTG	qRT-PCR for the quantification of the splicing of the intron #3 of *OSB2 transcript*
OSB2-RT-e6_e7-F	CTGATGGTGTTCCTTACAAC	qRT-PCR for the quantification of the splicing of the intron #6 of *OSB2 transcript*
OSB2-RT-e6_e7-R	CTTATCTAGTCTATTGTCCCACC	qRT-PCR for the quantification of the splicing of the intron #6 of *OSB2 transcript*
OSB2-RT-e7_e8-F	GGTGATTCTTGGAAGCATTTG	qRT-PCR for the quantification of the splicing of the intron #7 of *OSB2 transcript*
OSB2-RT-e7_e8-R	CTATCTAACCACAACACTTTCC	qRT-PCR for the quantification of the splicing of the intron #7 of *OSB2 transcript*
OSB2_HRM-F	CTGAATCTCCTACTTCTCTACCG	HRM-genotyping
OSB2_HRM-R	GCCAATATCAATATCAAGTGAGGG	HRM-genotyping
Primers specific to *RECA3*		
Name	Sequence	
RECA3-rf2_M61	ATGAGTTCAAACATTCCTTTC	RT-PCR of the possible imitation codon
RECA3-rf2_M130	ATGAAAAGTTCAGAATCTCAG	RT-PCR of the possible imitation codon
RECA3-rf3_M73	ATGTCTTCCGACCGAGTCTAC	RT-PCR of the possible imitation codon
RECA3-rf3_M116	ATGTTTTCTAAGCTATATTTG	RT-PCR of the possible imitation codon
RECA3-rf3_MEnza	ATGAGCCATTGTTATTTTTCA	RT-PCR of the possible imitation codon
RECA3-Intron-R	GCTTCTTATATGGTATTAGC	RT-PCR of the intron #1
RECA3-Intron-F	CCAGCATTGTTTATATGTGG	RT-PCR of the intron #1

RECA3-Exon-R	GATGTTAAAGCTGCAGAAAAGC	RT-PCR of the exon #1
RECA3-Exon-F	CTTTGGTAATCCTCCAATTCC	RT-PCR of the exon #1
RECA3-Exon3-R	GCTTGATCTTGCTCTTGGAATTGG	RT-PCR from the exons #1 to #6
RECA3-Exon6-F	CGAAGTGCCCGTGTCATTATTCG	RT-PCR from the exons #1 to #6
RECA3-RTPrimer-R	AGCTTATTCTTTACCACCTCC	Reverse primer for Retro-transcription
Determination of the stoichiometry of the mtDNA of *L. sativa*		
Name	Sequence	Function
mtDNA-1-F	TACACCCTCATAATCACATAAC	Quantification of the mtDNA region around the coordinate 3 kb by qPCR
mtDNA-1-R	TCTAACTAACCTCTTCGATTTC	Quantification of the mtDNA region around the coordinate 3 kb by qPCR
mtDNA-2-F	AGGTTAGGTCTCTTTTATTCTG	Quantification of the mtDNA region around the coordinate 361 kb by qPCR
mtDNA-2-R	CATCTAAAACCAGACTCTCAG	Quantification of the mtDNA region around the coordinate 361 kb by qPCR
mtDNA-4-F	AGAGCAAGTCCCACCTACCA	Quantification of the mtDNA region around the coordinate 351 kb by qPCR
mtDNA-4-R	TTCCACGCCTAGCATCGTAC	Quantification of the mtDNA region around the coordinate 351 kb by qPCR
mtDNA-5-F	AACTAGTCAGAGCAACTAACTG	Quantification of the mtDNA region around the coordinate 346 kb by qPCR
mtDNA-5-R	TGAGCTTTCTCACTATACCTC	Quantification of the mtDNA region around the coordinate 346 kb by qPCR
mtDNA-6-F	GTCTTGAGGTAATTGTAGAAAG	Quantification of the mtDNA region around the coordinate 342 kb by qPCR
mtDNA-6-R	CTCTTCTTGGATGAAGACTATC	Quantification of the mtDNA region around the coordinate 342 kb by qPCR
mtDNA-7-F	CTAATACCTTCTCGACACTATC	Quantification of the mtDNA region around the coordinate 337 kb by qPCR
mtDNA-7-R	GTTATATGACCAAGCAAAGAC	Quantification of the mtDNA region around the coordinate 337 kb by qPCR
mtDNA-8-F	CACGACTCTAATAGAAAGAAAG	Quantification of the mtDNA region around the coordinate 332 kb by qPCR
mtDNA-8-R	AACGAGTCTAAGGTTACAGAAG	Quantification of the mtDNA region around the coordinate 332 kb by qPCR
mtDNA-9-F	AAGTCAGCCTATAAGATGAAAC	Quantification of the mtDNA region around the coordinate 326 kb by qPCR
mtDNA-9-R	GGACCTCTATCAAGTCATTATC	Quantification of the mtDNA region around the coordinate 326 kb by qPCR
mtDNA-10-F	CTATATATGGAATAGCGGTAAG	Quantification of the mtDNA region around the coordinate 301 kb by qPCR
mtDNA-10-R	ATCTTTCTCAATATCCTCTACG	Quantification of the mtDNA region around the coordinate 301 kb by qPCR
mtDNA-11-F	TATTCCTGCTACTATTAGATCC	Quantification of the mtDNA region around the coordinate 306 kb by qPCR
mtDNA-11-R	AAGAATCTGCTCTTCTCTTATC	Quantification of the mtDNA region around the coordinate 306 kb by qPCR
mtDNA-12-F	ACTATTTCCATCTCTTATGACC	Quantification of the mtDNA region around the coordinate 111 kb by qPCR
mtDNA-12-R	CATATCTATCTCAATCTTGTGG	Quantification of the mtDNA region around the coordinate 111 kb by qPCR
mtDNA-13-F	GATAAGAATCAGTTCCACTTTC	Quantification of the mtDNA region around the coordinate 106 kb by qPCR
mtDNA-13-R	CTAAGAGAGATGATAACTGCAC	Quantification of the mtDNA region around the coordinate 106 kb by qPCR
mtDNA-14-F	AGGTAGAATATTTATGTCCCAG	Quantification of the mtDNA region around the coordinate 100 kb by qPCR
mtDNA-14-R	GATAAGTAGTCGATTAGGAGAAC	Quantification of the mtDNA region around the coordinate 100 kb by qPCR
mtDNA-15-F	GGAATTCTATACTATAAAGGGG	Quantification of the mtDNA region around the coordinate 96 kb by qPCR
mtDNA-15-R	CTAGGAATTGAAGAGAGAACAG	Quantification of the mtDNA region around the coordinate 96 kb by qPCR
mtDNA-16-F	GCTTACCTTTAACCTAACTCTC	Quantification of the mtDNA region around the coordinate 91 kb by qPCR
mtDNA-16-R	GTACTCAAGAACACAAGAAAAG	Quantification of the mtDNA region around the coordinate 91 kb by qPCR
mtDNA-17-F	GTCTATCTCTCTGGTCAAGTG	Quantification of the mtDNA region around the coordinate 86 kb by qPCR
mtDNA-17-R	CACGACTGATAGAAAGAGTAAG	Quantification of the mtDNA region around the coordinate 86 kb by qPCR
mtDNA-18-F	GTCCATGTAAATAATCGAGAC	Quantification of the mtDNA region around the coordinate 80 kb by qPCR
mtDNA-18-R	ACACAAAGAAGATACAGTTCAC	Quantification of the mtDNA region around the coordinate 80 kb by qPCR
mtDNA-19-F	CCAGTAAGTAAGCAGGTAATAG	Quantification of the mtDNA region around the coordinate 219 kb by qPCR
mtDNA-19-R	GTATGTGTCAAAACCGTTATAG	Quantification of the mtDNA region around the coordinate 219 kb by qPCR
mtDNA-20-F	ACATAGAGGGTCTCTACAAGTC	Quantification of the mtDNA region around the coordinate 214 kb by qPCR
mtDNA-20-R	TATCGGATCGAGTATTACTTAG	Quantification of the mtDNA region around the coordinate 214 kb by qPCR
mtDNA-21-F	TCGTAGTCAAATCTGATGATAC	Quantification of the mtDNA region around the coordinate 210 kb by qPCR
mtDNA-21-R	ATAAGGAGTAGGTGAGTGAGTC	Quantification of the mtDNA region around the coordinate 210 kb by qPCR
mtDNA-22-F	TAGACTATCTCATTAAACGTCG	Quantification of the mtDNA region around the coordinate 204 kb by qPCR
mtDNA-22-R	CTTCCTTCCTCTACTAGTTCTG	Quantification of the mtDNA region around the coordinate 204 kb by qPCR
mtDNA-23-F	TTACCTAAGTTTCCAGATACG	Quantification of the mtDNA region around the coordinate 277 kb by qPCR
mtDNA-23-R	TACCTCTGTCACTCTATCTTTG	Quantification of the mtDNA region around the coordinate 277 kb by qPCR
mtDNA-24-F	AGCTGAATTAAGAAGATAGGC	Quantification of the mtDNA region around the coordinate 273 kb by qPCR
mtDNA-24-R	TAATTGACTAAACTCTGTCAGG	Quantification of the mtDNA region around the coordinate 273 kb by qPCR
mtDNA-25-F	AACCTACTTTCTCGGAATATC	Quantification of the mtDNA region around the coordinate 267 kb by qPCR
mtDNA-25-R	GTTTCCATGATAGTCTAGATCC	Quantification of the mtDNA region around the coordinate 267 kb by qPCR
mtDNA-26-F	CATTTAGAGACTCAATGGAAG	Quantification of the mtDNA region around the coordinate 263 kb by qPCR
mtDNA-26-R	GAGATCGACTAGAAGAGAAAAC	Quantification of the mtDNA region around the coordinate 263 kb by qPCR
mtDNA-27-F	TCTGTGTTAGAGGACTTAGATG	Quantification of the mtDNA region around the coordinate 161 kb by qPCR

mtDNA-27-R	GATTTCAGTCTTGTGTTAAGTG	Quantification of the mtDNA region around the coordinate 161 kb by qPCR
mtDNA-28-F	TGAGATTAAGTTAGTGAGATGC	Quantification of the mtDNA region around the coordinate 156 kb by qPCR
mtDNA-28-R	CTTTATCCCATCTAGGTATAGG	Quantification of the mtDNA region around the coordinate 156 kb by qPCR
mtDNA-29-F	ACCGACTCAATTAAGGTATTAG	Quantification of the mtDNA region around the coordinate 151 kb by qPCR
mtDNA-29-R	AGATCATACTGAGAAACAGAAG	Quantification of the mtDNA region around the coordinate 151 kb by qPCR
mtDNA-30-F	AGCTAGCTCATCTCTCAGAC	Quantification of the mtDNA region around the coordinate 33 kb by qPCR
mtDNA-30-R	GAGAGGTAGAGAGATTTTAAGC	Quantification of the mtDNA region around the coordinate 33 kb by qPCR
mtDNA-31-F	TAATGACTTTCTAGTCAGATCG	Quantification of the mtDNA region around the coordinate 28 kb by qPCR
mtDNA-31-R	AATGTGTACGAGATACCATAAG	Quantification of the mtDNA region around the coordinate 28 kb by qPCR
mtDNA-32-F	CTGAAATATCTCACTACGAAAC	Quantification of the mtDNA region around the coordinate 24 kb by qPCR
mtDNA-32-R	GTCGGTCTAGAAAGCTAGTAAC	Quantification of the mtDNA region around the coordinate 24 kb by qPCR
mtDNA-33-F	AAACTAGTCTATCTGGGATGTC	Quantification of the mtDNA region around the coordinate 18 kb by qPCR
mtDNA-33-R	GAGAAGTTATTCAATCTGCTG	Quantification of the mtDNA region around the coordinate 18 kb by qPCR
mtDNA-36-F	TTAGATAGGTAAGAACAAGTGC	Quantification of the mtDNA region around the coordinate 310 kb by qPCR
mtDNA-36-R	TTTGAGAAAGTATAGGAGAAGC	Quantification of the mtDNA region around the coordinate 310 kb by qPCR
mtDNA-37-F	TGTCTAAACTGGAATCTGTG	Quantification of the mtDNA region around the coordinate 316 kb by qPCR
mtDNA-37-R	ATAGTTGGTATTCCTTCCTTAG	Quantification of the mtDNA region around the coordinate 316 kb by qPCR
mtDNA-38-F	TAGAAAGATTGCTACTAACGC	Quantification of the mtDNA region around the coordinate 300 kb by qPCR
mtDNA-38-R	CTTACCCAGTTCTTTCTATTTC	Quantification of the mtDNA region around the coordinate 300 kb by qPCR
mtDNA-39-F	ATTAGGGTATGGGATAGAGTTC	Quantification of the mtDNA region around the coordinate 297 kb by qPCR
mtDNA-39-R	GTTGAGTTACAGTACAGACTCC	Quantification of the mtDNA region around the coordinate 297 kb by qPCR
mtDNA-40-F	AGACTAACCTTCGTTTAGAAAG	Quantification of the mtDNA region around the coordinate 291 kb by qPCR
mtDNA-40-R	AAGGGTAGAGATAGGAAGTAAG	Quantification of the mtDNA region around the coordinate 291 kb by qPCR
mtDNA-41-F	CGACACCATGATAAAATACTAC	Quantification of the mtDNA region around the coordinate 198 kb by qPCR
mtDNA-41-R	GTTCCGTCATATTCCTAATAAC	Quantification of the mtDNA region around the coordinate 198 kb by qPCR
mtDNA-42-F	AATAAGGTCACCCATAACAC	Quantification of the mtDNA region around the coordinate 194 kb by qPCR
mtDNA-42-R	TACTAGTTGTTACCTTACGCTG	Quantification of the mtDNA region around the coordinate 194 kb by qPCR
mtDNA-43-F	GAACTCATAAACAAGATGGAG	Quantification of the mtDNA region around the coordinate 188 kb by qPCR
mtDNA-43-R	ATAAGTATCCCTTAGTCTGAGC	Quantification of the mtDNA region around the coordinate 188 kb by qPCR
mtDNA-44-F	CATAAATAGGACTCCAGTAAGC	Quantification of the mtDNA region around the coordinate 184 kb by qPCR
mtDNA-44-R	AGGTTTCAATAAGTGTCTGTAG	Quantification of the mtDNA region around the coordinate 184 kb by qPCR
mtDNA-45-F	GATAAAGAGTAAGGATTGAAGC	Quantification of the mtDNA region around the coordinate 178 kb by qPCR
mtDNA-45-R	AAGTGCCTACTACTACACCAC	Quantification of the mtDNA region around the coordinate 178 kb by qPCR
mtDNA-47-F	CAACAAGAGTTCAAACCTATTC	Quantification of the mtDNA region around the coordinate 167 kb by qPCR
mtDNA-47-R	CTTTGATAAAGAAGAAGAGCAG	Quantification of the mtDNA region around the coordinate 167 kb by qPCR
mtDNA-48-F	CTGGTATTGCTGCTATAAGTC	Quantification of the mtDNA region around the coordinate 255 kb by qPCR
mtDNA-48-R	GAGATTTCTCACTATTCGTATG	Quantification of the mtDNA region around the coordinate 255 kb by qPCR
mtDNA-49-F	TCTAAGATCTATTATCCATCCC	Quantification of the mtDNA region around the coordinate 250 kb by qPCR
mtDNA-49-R	CACAACACTAGAACCAAGTTAG	Quantification of the mtDNA region around the coordinate 250 kb by qPCR
mtDNA-50-F	CATTATGTACTTTCTATGGGAG	Quantification of the mtDNA region around the coordinate 245 kb by qPCR
mtDNA-50-R	GACCTAAAGTTTCAGGATATG	Quantification of the mtDNA region around the coordinate 245 kb by qPCR
mtDNA-51-F	GAGAAGTAGAAGTGAGATAAAGC	Quantification of the mtDNA region around the coordinate 240 kb by qPCR
mtDNA-51-R	GTGTAAGAGATATTCCATCAAG	Quantification of the mtDNA region around the coordinate 240 kb by qPCR
mtDNA-52-F	GAACTGTTAAGTACAGACGATG	Quantification of the mtDNA region around the coordinate 235 kb by qPCR
mtDNA-52-R	CTTTCTGTTCCTAGTTAAAGTG	Quantification of the mtDNA region around the coordinate 235 kb by qPCR
mtDNA-53-F	TGTAGAAAGGTATGGTGTAGTC	Quantification of the mtDNA region around the coordinate 230 kb by qPCR
mtDNA-53-R	CTTCTAGATGCAATATCTTCAC	Quantification of the mtDNA region around the coordinate 230 kb by qPCR
mtDNA-54-F	GTTCTAATAGTACAAGCTCGC	Quantification of the mtDNA region around the coordinate 225 kb by qPCR
mtDNA-54-R	ATATAGTTGTCAAAGACACTGG	Quantification of the mtDNA region around the coordinate 225 kb by qPCR
mtDNA-55-F	ATACTAGGTAGTTACCGAATGC	Quantification of the mtDNA region around the coordinate 77 kb by qPCR
mtDNA-55-R	GGTAGTTAAGAGCATCAATTTC	Quantification of the mtDNA region around the coordinate 77 kb by qPCR
mtDNA-56-F	GATACGGAGACTACTCTACAGG	Quantification of the mtDNA region around the coordinate 70 kb by qPCR
mtDNA-56-R	TTCTCGTTATAGAATGTAGTGC	Quantification of the mtDNA region around the coordinate 70 kb by qPCR
mtDNA-58-F	GCTATTAGAGGAGAGTTTTGAC	Quantification of the mtDNA region around the coordinate 62 kb by qPCR
mtDNA-58-R	CTACTTTCTGTCAGCTATAACC	Quantification of the mtDNA region around the coordinate 62 kb by qPCR
mtDNA-59-F	ATAGATAGATAGTCAAAAGGGG	Quantification of the mtDNA region around the coordinate 57 kb by qPCR
mtDNA-59-R	GTTTATCGGATATCAGACTTTC	Quantification of the mtDNA region around the coordinate 57 kb by qPCR
mtDNA-60-F	TAGTTAGTACCGACCGTAAGAC	Quantification of the mtDNA region around the coordinate 52 kb by qPCR
mtDNA-60-R	GAATACATAGGGTAGGTAGGAC	Quantification of the mtDNA region around the coordinate 52 kb by qPCR

Determination of the stoichiometry of the cpDNA of *L. sativa*		
Name	Sequence	Function
cpDNA-1-F	GAATATGCAACAGCAATCCAAG	Quantification of the cpDNA region around the coordinate 1 kb
cpDNA-1-R	GGTGGTCCTTATGAACTAATTG	Quantification of the cpDNA region around the coordinate 1 kb
cpDNA-2-F	CGAGCCAAGAGCACCTTCATTC	Quantification of the cpDNA region around the coordinate 6 kb
cpDNA-2-R	ACGATGTGGTAGAAAGCAACG	Quantification of the cpDNA region around the coordinate 6 kb
cpDNA-3-F	CACCTGTCAATTCCATGAATGC	Quantification of the cpDNA region around the coordinate 11 kb
cpDNA-3-R	TACTATTCAATTCTCGACGACG	Quantification of the cpDNA region around the coordinate 11 kb
cpDNA-4-F	ACGGGTAGGAGAAATGGAGGTTTG	Quantification of the cpDNA region around the coordinate 16 kb
cpDNA-4-R	CCTGGCGCGCTCTAATATGATC	Quantification of the cpDNA region around the coordinate 16 kb
cpDNA-5-F	CGGCCATTCTCGACGGTAATTC	Quantification of the cpDNA region around the coordinate 21 kb
cpDNA-5-R	GGTATTAGTTGGTTCTCGCCTTC	Quantification of the cpDNA region around the coordinate 21 kb
cpDNA-6-F	CCGCTTCCGTTATTGCTGCTG	Quantification of the cpDNA region around the coordinate 26 kb
cpDNA-6-R	GTCTCGCGATACCTTCTACAGC	Quantification of the cpDNA region around the coordinate 26 kb
cpDNA-7-F	GGCGCTTCTGTCTATGTGTAG	Quantification of the cpDNA region around the coordinate 31 kb
cpDNA-7-R	GTTCATGGAATTCCTAAACG	Quantification of the cpDNA region around the coordinate 31 kb
cpDNA-8-F	CGCGCCGAAGAGAAGAACTG	Quantification of the cpDNA region around the coordinate 36 kb
cpDNA-8-R	GACGGCCTCTTTACCATTCTG	Quantification of the cpDNA region around the coordinate 36 kb
cpDNA-9-F	CATGCCACTCAGCCAAAGGAAG	Quantification of the cpDNA region around the coordinate 41 kb
cpDNA-9-R	CATGCTGATGCTCACGATTTCG	Quantification of the cpDNA region around the coordinate 41 kb
cpDNA-10-F	CATCCCTCGCCGCAATAGTATC	Quantification of the cpDNA region around the coordinate 46 kb
cpDNA-10-R	CGGGAGCTCGCCAATTAGTTAAC	Quantification of the cpDNA region around the coordinate 46 kb
cpDNA-11-F	CGGACTCGAACCGTAGACTTTC	Quantification of the cpDNA region around the coordinate 51 kb
cpDNA-11-R	CAGTTAATGAAAGAGCCCAATGC	Quantification of the cpDNA region around the coordinate 51 kb
cpDNA-12-F	CAGTTCGGTGGAGGAACTTTAG	Quantification of the cpDNA region around the coordinate 56 kb
cpDNA-12-R	GCAAGATCGCGTCCCTCATTAC	Quantification of the cpDNA region around the coordinate 56 kb
cpDNA-13-F	GGTGGAATACTAGGCAATCCGAAC	Quantification of the cpDNA region around the coordinate 61 kb
cpDNA-13-R	CCGAGTATTCCTTAATCATTTCCTC	Quantification of the cpDNA region around the coordinate 61 kb
cpDNA-14-F	GATTGCCGTTCCAGTTAGTGCTTC	Quantification of the cpDNA region around the coordinate 66 kb
cpDNA-14-R	GTTCGGTAGAACGTGGGTCTCC	Quantification of the cpDNA region around the coordinate 66 kb
cpDNA-15-F	GATCCCATTGAAGCGGCTGATC	Quantification of the cpDNA region around the coordinate 71 kb
cpDNA-15-R	CTCCTGGCGGATGGGTAATACC	Quantification of the cpDNA region around the coordinate 71 kb
cpDNA-16-F	CGATGGTCGGCAAGTATGATGG	Quantification of the cpDNA region around the coordinate 76 kb
cpDNA-16-R	CCCAGAACCACACCTGTAACC	Quantification of the cpDNA region around the coordinate 76 kb
cpDNA-17-F	CCTGAACCCATACGTGTTTCTGC	Quantification of the cpDNA region around the coordinate 81 kb
cpDNA-17-R	CGACGAGCAATGACACGAAATG	Quantification of the cpDNA region around the coordinate 81 kb
cpDNA-18-F	CGGTTCTGTCGGTGCTTGAAAC	Quantification of the cpDNA region around the coordinate 85 kb
cpDNA-18-R	CGGCAGCAAGTGATTGAGTTCAG	Quantification of the cpDNA region around the coordinate 85 kb
cpDNA-19-F	TGAACAACCGGGAGCAATTTAC	Quantification of the cpDNA region around the coordinate 91 kb
cpDNA-19-R	GCAAGGAATATCCGTCTTTCTG	Quantification of the cpDNA region around the coordinate 91 kb
cpDNA-20-F	ACGGATGCTCCTATTACACTCG	Quantification of the cpDNA region around the coordinate 95 kb
cpDNA-20-R	ATTACGGGTAGTTCCTACAAAG	Quantification of the cpDNA region around the coordinate 95 kb
cpDNA-21-F	AAGAAGCAATGACGGTATCTGG	Quantification of the cpDNA region around the coordinate 101 kb
cpDNA-21-R	AAGCCACCTACAGACGCTTTAC	Quantification of the cpDNA region around the coordinate 101 kb
cpDNA-22-F	TAGCGAAAGCGAGTCTTCATAG	Quantification of the cpDNA region around the coordinate 105 kb
cpDNA-22-R	CTCCACTTAGTTTCACCCAAGC	Quantification of the cpDNA region around the coordinate 105 kb
cpDNA-23-F	GAGGCATAAGTCGAGCTACAAG	Quantification of the cpDNA region around the coordinate 111 kb
cpDNA-23-R	GGTTACCTGATGCAATGGAAGG	Quantification of the cpDNA region around the coordinate 111 kb
cpDNA-24-F	CCTCTCCGTGGGTATCTGGTAAC	Quantification of the cpDNA region around the coordinate 116 kb
cpDNA-24-R	CAATCGTATCCTGTGGCTTTGG	Quantification of the cpDNA region around the coordinate 116 kb
cpDNA-25-F	ACTGCTATAGATGGTCCAATACTG	Quantification of the cpDNA region around the coordinate 121 kb
cpDNA-25-R	CGCAGGACCTTTGGGAATTCTC	Quantification of the cpDNA region around the coordinate 121 kb
cpDNA-26-F	GGATGTTCCGTTTCCATTCC	Quantification of the cpDNA region around the coordinate 126 kb
cpDNA-26-R	CCTTGGCACAGATCTAAGCTAAGAC	Quantification of the cpDNA region around the coordinate 126 kb
Mitochondriale transcript quantification		
Name	Sequence	Function
atp1-F	CCCCAGCTTGAATCTCGTTC	Quantification of the atp1 transcript by RT-qPCR

atp1-R	GGTCGAGTGGTCTCAGTTG	Quantification of the atp1 transcript by RT-qPCR
atp4-F	GGCATTTGGAAGTGTTGCTG	Quantification of the atp4 transcript by RT-qPCR
atp4-R	GTGCGAAAAGACAGTGCAAG	Quantification of the atp4 transcript by RT-qPCR
atp6-F	GTGGGTTCGCTTGGACTATG	Quantification of the atp6 transcript by RT-qPCR
atp6-R	CCAGACCGGTTAATGCAAGAAC	Quantification of the atp6 transcript by RT-qPCR
atp8-F	TCCTCGTGAGCCACTGATTTC	Quantification of the atp8 transcript by RT-qPCR
atp8-R	TAAGGCCGTCGACTTATTGGG	Quantification of the atp8 transcript by RT-qPCR
atp9-F	GCTTCGGTTAGAGCAAAGCC	Quantification of the atp9 transcript by RT-qPCR
atp9-R	TCCTCGATTCATTCCGTGGC	Quantification of the atp9 transcript by RT-qPCR
ccmB-F	CGGAATGGATCGGTTAAACA	Quantification of the ccmB transcript by RT-qPCR
ccmB-R	CCGAACGAGAATGAATACCAC	Quantification of the ccmB transcript by RT-qPCR
ccmC-F	AAGTCTTCGGGTAGCACCAC	Quantification of the ccmC transcript by RT-qPCR
ccmC-R	TGTACCGGTTCCGAAAGAGC	Quantification of the ccmC transcript by RT-qPCR
ccmFN-F	TGGCTGTTGGTCACGACTAC	Quantification of the ccmFN transcript by RT-qPCR
ccmFN-R	GGCAATCGTTGAGTGAACTG	Quantification of the ccmFN transcript by RT-qPCR
cob-F	GAACACATTATGAGAGATGTTGAAGG	Quantification of the cob transcript by RT-qPCR
cob-R	AACATACTTGCCCCATTAGCA	Quantification of the cob transcript by RT-qPCR
cox2-F	ATGCCATAAAGCGCGAACC	Quantification of the cox2 transcript by RT-qPCR
cox2-R	GACGCAGCAACACCTATGATG	Quantification of the cox2 transcript by RT-qPCR
cox3-F	GCCCAAACTGAGAAGTGTTG	Quantification of the cox3 transcript by RT-qPCR
cox3-R	GGCCTATTTCGGGTTCACTC	Quantification of the cox3 transcript by RT-qPCR
nad4-F	ATGCTACCTCCAATTCCCTG	Quantification of the nad4 transcript by RT-qPCR
nad4-R	ATACCCATGTTTCCTGAAGC	Quantification of the nad4 transcript by RT-qPCR
nad4L-F	CTACAGCAATAGTCCCTC	Quantification of the nad4L transcript by RT-qPCR
nad4L-R	GCTTCATTGGTTCCAACGGTG	Quantification of the nad4L transcript by RT-qPCR
nad5a-F	CCACCTACGGCTTTGATTGT	Quantification of the nad5a transcript by RT-qPCR
nad5a-R	GTTGCCGCAAGGAATGAC	Quantification of the nad5a transcript by RT-qPCR
nad5b-F	CAATTTTTGGGCCAATTCC	Quantification of the nad5b transcript by RT-qPCR
nad5b-R	TGGTTGGAGCAGCAAACTC	Quantification of the nad5b transcript by RT-qPCR
nad7-F	GGTCGTATGAAACTGGCATGC	Quantification of the nad7 transcript by RT-qPCR
nad7-R	TGGATGTGGGAGCATCAACTC	Quantification of the nad7 transcript by RT-qPCR
rps3-F	CTCGACCAGCGAGAAAAAGT	Quantification of the rps3 transcript by RT-qPCR
rps3-R	*GGTAGCACCGAAGAAAGGAAA*	Quantification of the rps3 transcript by RT-qPCR
rps4-F	CAGCGAAAGAACTGCCTAAG	Quantification of the rps4 transcript by RT-qPCR
rps4-R	GACCCGTTTTTGCAACAGTTG	Quantification of the rps4 transcript by RT-qPCR
rrnS-F	CACACTGGGACTGAGACACG	Quantification of the rrnS transcript by RT-qPCR
rrnS-R	GCCCATTGTCCAAGATTCC	Quantification of the rrnS transcript by RT-qPCR
TUA-3F	CTTCTTAGTGTTCAATGCTGTTGG	Reference gene for RT-qPCR
TUA-3R	GAAGGGTAGATAGTGAAACCGAGC	Reference gene for RT-qPCR
TIP41-F	TTTGTATGGAGATGAATTGGCTGATA	Reference gene for RT-qPCR
TIP41-R	CGTAAGAGAAGAAACCAACAGCTAGG	Reference gene for RT-qPCR
ELFα-F	CGGGTCAGATTGGAAACGGGTAT	Reference gene for RT-qPCR
ELFα-R	CTTCTTCGCAGCAGCCTTGGTGA	Reference gene for RT-qPCR

Université de Strasbourg

Arnaud FERTET

École Doctorale des Sciences de la Vie et de la Santé STRASBOURG

Zooming in on the lettuce mitochondrial genome:

Study of induced and natural mitochondrial diversity within the *Lactuca spp*.

Résumé

En sélection végétale, la diversité nucléaire est le réservoir de biodiversité encore majoritairement exploité. Cependant, la diversité génétique mitochondriale représente de nouvelles opportunités telles que la stérilité mâle cytoplasmique qui semble le plus intéressant caractère chez la laitue, dans une optique de sélection. Pour cette raison, au cours de ma thèse j'ai séquencé et assemblé le génome mitochondrial (ADNmt) de la laitue, *Lactuca sativa,* que j'ai utilisé comme référence pour mes études subséquentes. Puis j'ai étudié la diversité mitochondriale naturelle, présente au sein d'accessions sauvages de quatre espèces du genre *Lactuca*, ou la diversité mitochondriale induite. Celle-ci générée en utilisant soit un composé génotoxique soit des mutants pour des gènes nucléaires codant des protéines impliquées dans la maintenance et la ségrégation de l'ADNmt. Parmi les mutants sélectionnés nous avons décidé de nous focaliser sur deux d'entre eux : un mutant de la protéine RADA présentant un phénotype sévère, associé à une importante instabilité de l'ADNmt par recombinaison, et un mutant de la protéine OSB2 que nous pensions lié à un phénotype de stérilité mâle. Il s'est avéré que nous avions sélectionné de façon indirecte dans cette famille de mutants *osb2* un nouveau gène de stérilité que nous avons par la suite identifié comme étant *CCoAOMT1,* codant pour une caffeoyl coenzyme A méthyltransférase.

Mots clés : Laitues, ADNmt, diversité mitochondriale, stérilité mâle cytoplasmique, GMS, recombinaison, OSB2 RADA.

Abstract

In breeding, nuclear diversity remains the biodiversity reservoir most exploited. However, mitochondrial genetic diversity might present new opportunities in lettuce breeding, and cytoplasmic male sterility seems to be the most interesting trait, with a view to breeding. For this reason, during my thesis I sequenced and assembled the mitochondrial genome (mtDNA) of lettuce, *Lactuca sativa*, which I used as a reference for my subsequent studies. Then I studied the natural mitochondrial diversity, present within wild accessions of four species of the genus *Lactuca*, or the induced mitochondrial diversity, generated using either a genotoxic compound or mutants for nuclear genes encoding proteins involved in the maintenance and segregation of mtDNA. Among the mutants selected we decided to focus on two of them: a mutant of the RADA protein, with a severe phenotype associated to important instability of the mtDNA by recombination and a mutant of the protein OSB2, which we thought was linked to a male sterility phenotype. It turned out that we had indirectly selected within this *osb2* mutant family a new sterility gene, which we subsequently identified as CCoAOMT1, coding for a caffeoyl coenzyme A methyltransferase.

Key words: Lettuce, mtDNA, mitochondrial diversity, cytoplasmic male sterility, GMS, recombination, OSB2, RADA