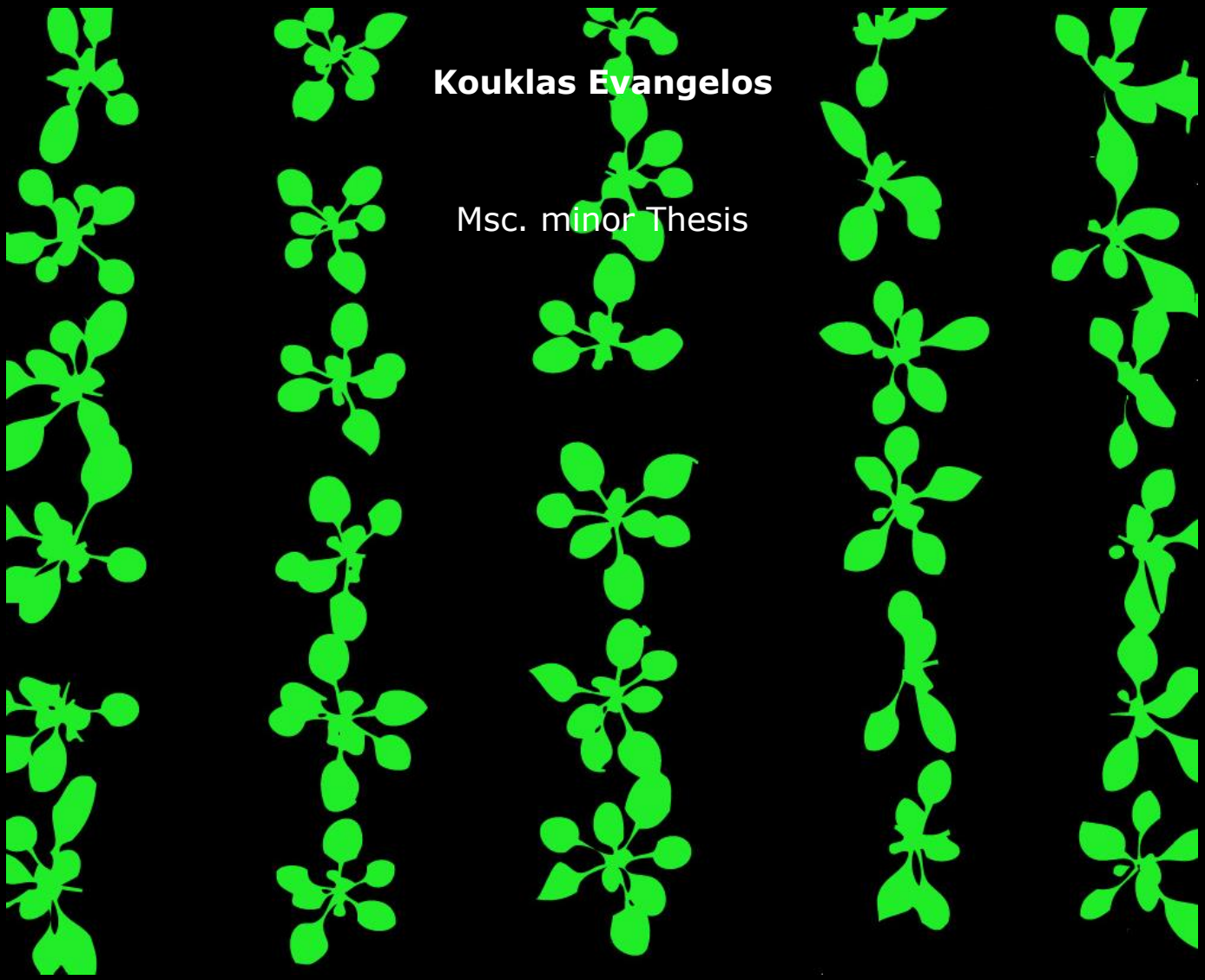
Construction of a cytoplasmic swap library in *Arabidopsis thaliana*

Kouklas Evangelos

Msc. minor Thesis



WAGENINGEN UNIVERSITY
WAGENINGEN UR

Construction of a cytoplasmic swap library in Arabidopsis thaliana

Master of Science minor Thesis (27 ECTS)

Kouklas Evangelos
Registration Number: 860226471130

Supervisor: Padraic Flood

Examiner a: Dr. ing. Joost Keurentjes

Examiner b: prof. Dr. Hans de Jong

Laboratory of Genetics

November 2012 – April 2013

Wageningen University

Table of Contents

1. Introduction

1.1 Plant cell components carrying genetic information

1.2 Evolution of mitochondria and plastids

1.3 Cytoplasmic and nuclear crosstalk

1.4 Crosstalk examples

1.5 CENH3 mutation: new breeding opportunities

2. Aim of thesis

3. Results

3.1 Creation of the cybrid haploid population

3.2 Marker-based molecular characterization of polymorphic patterns

3.2.1. Polymorphic pattern of wild-types

3.2.2. Polymorphic pattern of inducer lines (mothers)

3.3 Physiological experiments on cybrids

4. Discussion

5. References

6. Materials and Methods

7. Acknowledgements

8. Annex I

1. Introduction

Arabidopsis thaliana (common name: mouse-ear cress) is a small flowering annual plant species and a member of the Brassicaceae family. The plant's size, life cycle, low input growth requirements, together with other developmental and physiological properties made it an ideal model organism for biological studies for over 50 years. Two decades ago, scientists discovered that the genetic organization of the plant allows rapid and repetitive screening of genomic libraries with minimal effort, in addition to the easy induction and screening of mutants (Meyerowitz 1989). This discovery, combined with the release of its complete genomic sequence in 2000 (AGI 2000) provided a unique opportunity to establish a universal research source tool for higher organisms.

1.1 Plant cell components carrying genetic information

A generalized version of a plant cell consists out of: a cell wall, a nucleus, a vacuole (often contains ergastic substances like crystals), a plasma membrane, a cytoskeleton (microtubules, microfilaments), an ER (Endoplasmic reticulum), a Golgi apparatus, ribosomes, mitochondria, chloroplasts and a cytosol (Figure 1.1).

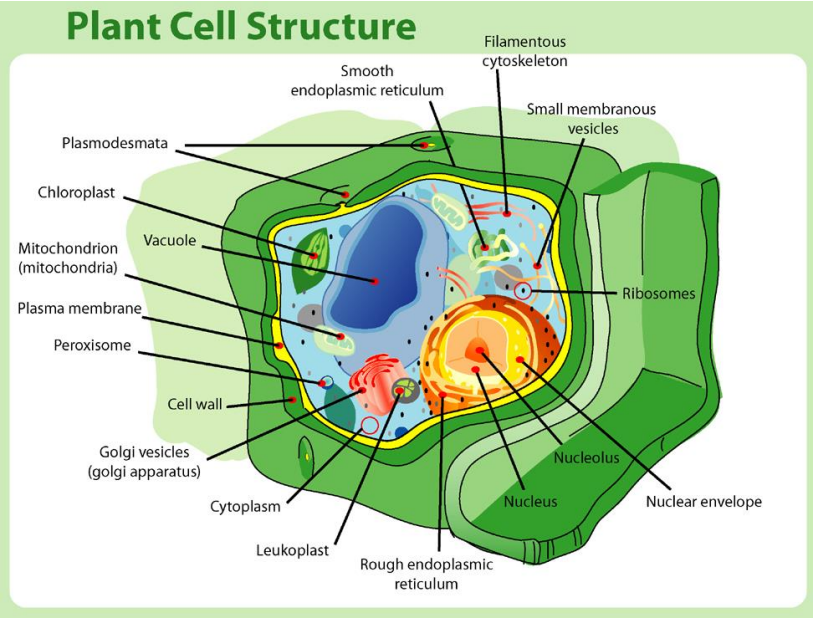


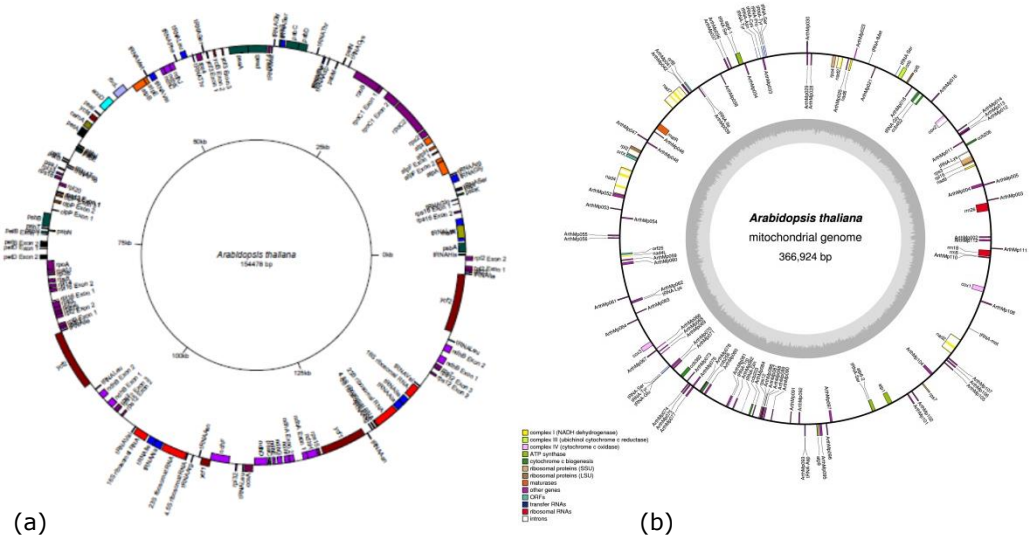
Figure 1.1: Plant cell components (http://en.wikipedia.org).

In eukaryotes, DNA is found in several distinct cellular compartments: the nucleus, mitochondria, and, in the case of algae and plants, plastids (Leucoplasts, chloroplasts and chromoplasts) (Kleine, Maier et al. 2009).

The chloroplasts are found only in plants and algae and perform one of the basic physiological processes, namely photosynthesis. For this purpose plants and photosynthetic algae use light to split water, forming O_2 and reducing CO_2 to an organic form, thus converting some of the absorbed light into chemical energy (Flood, Harbinson et al. 2011), a procedure that also provides the electrons for

the creation of ATP. Besides photosynthesis, chloroplasts carry out other essential plant functions, such as the synthesis of amino acids, fatty acids and lipids, plant hormones, nucleotides, vitamins and secondary metabolites (Leister 2003). Chloroplasts have two membranes, the inner and outer membrane. The interior of the plastids is called stroma, is connected by the inner membrane and contains enzymes that convert CO_2 into carbohydrates, mainly starch (Salisbury F. 1992). It also contains the DNA, RNA and proteins such as ribosomes of the chloroplast. Moreover, the stroma contains the thylakoids, the energy-producing systems of the plastid and of the plant. The chloroplast genome is a circular DNA molecule composed of 154,478 bp encoding a total of 87 potential protein-coding genes (Figure 2a) (Sato S 1999). Moreover, the chloroplast genome in Arabidopsis shows sequence variability amongst different genotypes worldwide.

Mitochondria, on the other hand, are found in all eukaryotes and play a very important role in energy handling, protein synthesis and respiration. This process consists essentially of the Krebs cycle, the electron-transport system and oxidative phosphorylation (Salisbury F. 1992). A mitochondrion contains outer and inner membranes composed of phospholipid bilayers and proteins, with different properties. Because of this organization, there are five distinct parts to a mitochondrion: the outer mitochondrial membrane, the intermembrane space (the space between the outer and inner membranes), the inner mitochondrial membrane, the cristae space (formed by infoldings of the inner membrane), and the matrix (space within the inner membrane). In mitochondria, ATP formation from ADP and Pi is indirectly driven by the strong thermodynamic tendency of O_2 to become reduced, a process called oxidative phosphorylation (Salisbury F. 1992). Moreover, mitochondria use the proton gradient across the membrane to synthesize ATP (Lewin 2007). Due to the presence of mitochondria in all cells and the differences in copy number per cell, they are not uniform or static entities within the cell; they can vary in composition and function between different cells, organs, and species over development and in response to different stresses (Carrie, Murcha et al. 2012). The mitochondrial genome was the first genomic component of *Arabidopsis thaliana* and the first mitochondrial genome in flowering plants that was sequenced. The established size of the genome was 366.924 nucleotides that encode 57 identified genes with a duplication rate of 7% (Figure 1.2b) (Unseld M 1997).



(a) (b)

Figure 1.2: (a) Complete chloroplast genome (constructed using GenomeVx and TAIR database), (b) complete mitochondrial genome (adapted from Unseld et. al, 1997).

1.2 Evolution of mitochondria and plastids

One of the key topics that need elucidation is the origin and evolution of the organelles. The most convincing evolutionary scenario is called serial endosymbiosis theory (SET) (Taylor 1974). There, it is stated that mitochondria in cells probably arose from an endosymbiosis between the bacterial ancestors of mitochondria together with a host cell. Another similar endosymbiosis event that occurred in a later evolutionary step is responsible for the introduction of the chloroplasts in plant cells and algae. Phylogenetic reconstructions based on mitochondrially encoded proteins have placed the mitochondrial ancestor in the a-proteobacterial subdivision (Andersson, Karlberg et al. 2003). The similarities between bacterial respiration and mitochondrial functions as well as the similarities between cyanobacterial photosynthesis and chloroplast functions in plants taken together with molecular phylogenies have provided convincing evidence of the bacterial origins of these modern cellular organelles (Gray 1992). The reasons for the events that actually created eukaryotic organisms are not known yet. There are a lot of different scenarios, like the host's mitochondrial ATP production, as well as the hydrogen hypothesis and the oxygen-respiring hypothesis (Allen 2003). The modern mitochondrial genome is able to encode only a handful of proteins and most of the proteins present in mitochondria are encoded from nuclear genes. In Arabidopsis, for example, the mitochondrial genes that are encoded cover only 10% of the mitochondrial genome, with the majority of the genome (62%) containing no obvious feature of information (Unseld M 1997). Such facts lead to several assumptions; one opinion supports that the loss of sequences from the ancestral genomes and the expansion of its proteome with novel contributions from the eukaryotic nuclear genome are incremental processes that are presumably spread over millions of years (Andersson, Karlberg et al. 2003).

Chloroplasts were first identified as organelles with unique features in the plant cell over a century ago (Mereschkowsky 1905; Martin and Kowallik 1999). In that report, Mereschkowsky mentions two very important hypotheses that indeed were proven years ago. First of all, chloroplasts started actually as symbiotic organisms in plant cells, and secondly that chloroplasts descend from cyanobacteria. Only cyanobacteria and chloroplasts have two photosystems and split water to obtain electrons as a source of reducing power (Raven 2003). Still, the actual ancestor of chloroplasts is not known due to the physiological and biochemical differences, but it seems like plastids arose from a strongly supported monophyletic group near the base of the cyanobacterial clade (Turner 1999). Like in the case of mitochondrial genome, the chloroplast genome encodes only a small fraction of the organelle's proteins: for plastids, the number ranges from 15 to 209 proteins (Keeling and Palmer 2008). In comparison to chloroplast genome the complete sequencing of cyanobacterial species has shown that their genome codes for at least 1,500 proteins, considerably larger than plastid genomes (Raven 2003). These findings about the chloroplast genome show that from an evolutionary point of view, the transition of a cyanobacterium into a plastid involved not only vertical transmission of organellar genomes, but also many evolutionary innovations, such as the light-harvesting antenna complex of higher plants (Martin 2002).

1.3 Cytoplasmic and nuclear crosstalk

The previous section highlighted the significant difference between the genome size and the proteins encoded by both chloroplasts and mitochondria compared with their ancestors. Consequently, scientists tried to find out what happened to the lost organelle genome. The main hypothesis was that parts of the organellar genome were transferred to the nucleus. The first organism that was discovered to have mitochondrial genes in the nucleus was budding yeast (*Saccharomyces cerevisiae*) (Thorsness 1990). For plants, this hypothesis was proven, at the time when fully sequenced genomes started to appear. First of all, a region about 620 kb of mitochondrial DNA (mtDNA) was discovered on chromosome 2 of Arabidopsis (Stupar, Lilly et al. 2001), and another big region of about 33 kb of cpDNA was discovered on chromosome 10L of rice (Yuan 2002). It seems that over evolutionary time, the DNA loss was caused by the redistribution of genetic material between nucleus, mitochondria and plastids via intercompartmental DNA transfer (Kleine, Maier et al. 2009). Since this part of the hypothesis was established quite firmly, the next question was referred to what happened to the nuclear genome after the endosymbiosis events. Phylogenetic analyses of the Arabidopsis nuclear genome combined with cyanobacterial sequences revealed that around 18% of the nuclear genome, a total of 4,500 cyanobacterial genes were identified in the five nuclear chromosomes (Martin 2002). Similarly, it was

reported that 45% of the Arabidopsis mtDNA gave rise to more than 500 insertions (NUMTs) of various sizes in the nucleus, representing a 0,17% of the nuclear genome (Richly 2004). As it seems, during evolution, organelles export their genes to the nucleus, but reimport the products with the help of transit peptides and protein-import machinery, so that proteins are retained in organelles, but most of the genes are not (Martin 1998). There are several reasons for this event in evolution. One of them is a phenomenon called Muller's ratchet (Muller 1964). There, the answer in the question of "what happens to an organism when it is deprived of sex" is elaborated. Apparently, the organism cannot recombine out the deleterious mutations that are inevitably going to accumulate in its genome, piling up. In the end, this would have irreversibly degenerative results for the organelle genome. So, when one of the genes is successfully transferred to the nucleus, it moves from a predominantly asexual to a predominantly sexual genome, restoring recombination, and freeing the gene from the fate of mutational meltdown (Martin 1998). In the end, the reduction of the organelle genome gave them a selective advantage over the unreduced ones, because they were replicating faster, so their inheritance was more certain. The mechanism underlying the organellar gene transfer in the nucleus has been experimentally demonstrated for mitochondria in yeast (Thorsness 1990) and for chloroplasts in tobacco (Huang 2003; Stegemann 2003). In yeast, mtDNA is inserted into the nuclear genome by nonhomologous end-joining (NHEJ) during the repair of double-stranded breaks (DSBs) in haploid mitotic cells. (Figure 1.3) (Ricchetti 1999; Yu 1999). Repair of DSBs by NHEJ requires little or no sequence homology (0 to 4 bp, 'micro-identities') between the termini, enabling the noncomplementary ends of DSBs and orgDNA to be 'pasted' to one another (Ricchetti 1999; Yu 1999; van Gent 2001). This mechanism shaped chloroplasts and mitochondria into highly specialized and efficient machinery of vital importance for eukaryotic organisms.

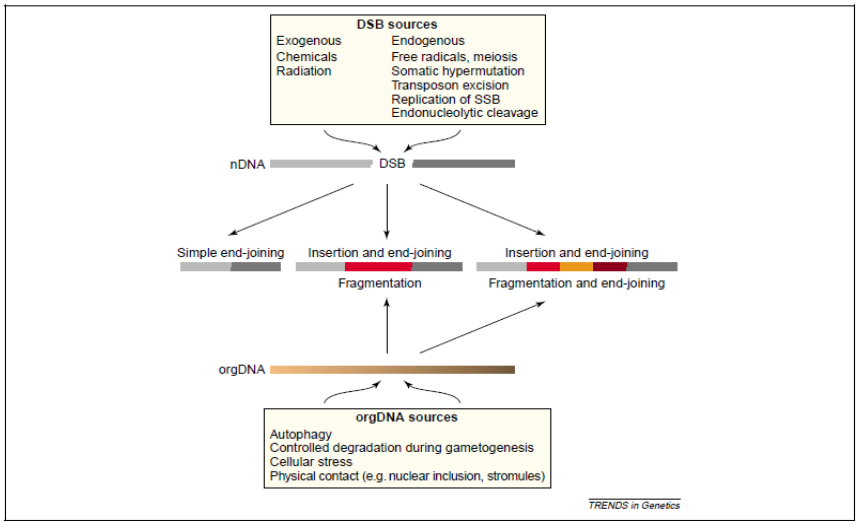


Figure 1.3: Mechanism of integration for organellar DNA into nucleic chromosomes (adapted from Leister, D. 2005).

1.4 Crosstalk examples

In chloroplasts, the photosynthetic apparatus carries out processes like light capture, electron transfer, ATP synthesis and CO_2 assimilation. These photosynthetic protein complexes are all genetic chimeras, because they contain subunits encoded and made in the chloroplast itself that are nevertheless assembled into close proximity and intimate functional contact with subunits that are encoded on chromosomes in the nucleus and imported, instead, from the cytosol (Allen, de Paula et al. 2011). In the protein interaction network between chloroplasts and nucleus, 19,368 interaction pairs of 3,565 proteins were predicted (Yu, Li et al. 2008). The vast majority of the chloroplast proteins are nuclear-encoded and require N-terminal presequences, termed 'chloroplast transit peptides' (cTPs), to target them back to the chloroplast (Leister 2003). There are several scenarios for the existence of such protein interaction networks within the cell. One of them suggests that safe and efficient production of ATP might require tight regulation of the expression of mitochondrial and chloroplast genes for functions directly involved in energy transduction, and thus favour rapid, onsite regulation within the organelle (Kleine, Maier et al. 2009). This theory is called Co-location for Redox Regulation (CoRR)((a)Allen 2003; (b)Allen 2003). CoRR states that the benefit of retaining rapid and unconditional redox control of expression of the genes for the key components is so great that, on its own, it repays the cost of maintaining a separate genetic system in close proximity to the electron transport chain (Allen, de Paula et al. 2011). On the other hand, more leisurely modes of regulation, such as acclimation responses that operate on a long timescale, involve both regulation of organelle gene expression and modification of the expression of nuclear gene products (Pfannschmidt 2003).

Probably the best example to observe the coordination of nuclear and chloroplast gene expression is RubisCO (ribulose 1,5-bisphosphate carboxylase/oxygenase), one of the most important protein complexes. It has eight large (RbcL) subunits encoded in the chloroplast and eight small (RbcS) nuclear-encoded small subunits (Gatenby 1990). Other counter-examples from chloroplasts are the nuclear-encoded but hydrophobic subunits of the chloroplast light-harvesting complexes, LHC II and LHC I (Allen 2003). The light-harvesting chlorophyll a/b binding (LHC) proteins are a group of nuclear-encoded thylakoid proteins, one of which (the major LHCII protein, or Lhcb1) is perhaps the most abundant membrane protein in nature (Jansson 1999).

The previous examples highlight the fact that although various genes from both mitochondria and chloroplasts migrated to the nucleus, their products are still essential for basic functions of the organelles.

1.5 CENH3 mutation: new breeding opportunities

So far, it has been established that the function of plant organelles needs additional genetic information from the nucleus, creating a genetic mosaic in which interactions are mediated between the organelles and the nucleus. The genetic contribution of the nucleus to the organellar biochemical pathways, as well as the signal transduction pathways that include transit peptides and protein-import machinery in plants have been in the research spotlight during the last decade. On the contrary, the actual genetic contribution of chloroplasts and mitochondria in physiological traits of the plant is only partially researched. Recently, the discovery of the CENH3 function in plants created a possibility to generate haploid plants that inherit chromosomes from only one parent and helped to investigate aspects of this topic.

In meiosis, a single round of DNA replication is followed by two rounds of nuclear division, meiosis I and meiosis II; meiosis I is unique and involves the segregation of homologous chromosomes (homologs), but meiosis II resembles mitosis because it results in the segregation of sister chromatids (Ma 2006). Both meiosis I and II are further divided into four phases: prophase, metaphase, anaphase, and telophase (Figure 1.4). In meiosis I, sister centromeres face in the same direction (mono-orientation), that allows sister chromatids to move to the same side of the spindle in anaphase I, when homologous chromosomes are segregated apart (Figure 1.4) (Copenhaver, Ravi et al. 2011). Centromeres are the chromosomal loci that mediate faithful inheritance of the genome during cell division (Ravi and Chan 2010) by providing a platform for the assembly of kinetochores, the protein complexes on eukaryotic chromosomes that attach to spindle microtubules (Ravi, Kwong et al. 2010).

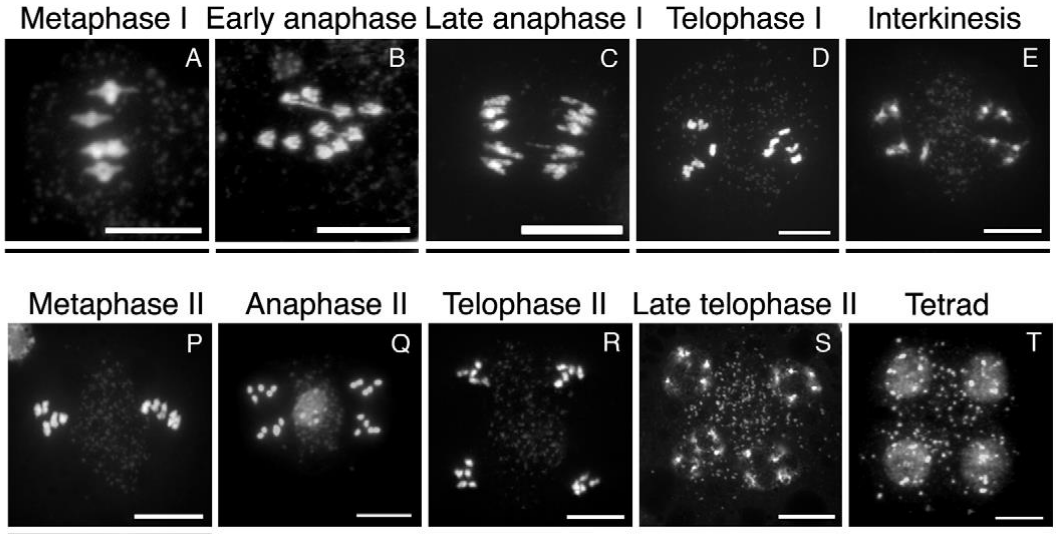


Figure 1.4: Male meiotic chromosome spreads from wild type *Arabidopsis* plant. Scale bar ~1 µm, figure adapted and modified from Figure 1 of Ravi M., 2011.

The centromeres, together with the molecular mechanism that involves the presence of different histones and post-translational modification events are most likely responsible for chromosome inheritance and segregation. In their absence, somatic tissues and organs would not develop and differentiate (Sullivan 2001). Although their DNA sequences are not conserved, animal and fungal centromeres are characterized by special histone H3 variants, or centromeric H3s (CenH3s), which are thought to identify centromeres (Talbert 2002). These histone H3 variants specify epigenetically the centromeres and replace conventional H3 in centromeric nucleosomes (Henikoff and Dalal 2005). The centromeric regions of *Arabidopsis thaliana* have a core consisting of large tandem arrays of 180-bp repeats (Martinez-Zapater 1986; Simoens 1988; Maluszynska 1991). In *Arabidopsis thaliana*, HTR12 gene encodes an H3-like variant histone that is found specifically at the cytologically visible centromeres and colocalizes with the 180-bp repeats (Talbert 2002). The *cenh3-1* mutant is an embryo-lethal null mutant; its discovery helped to replace native CENH3 with altered versions of the protein. The embryo-lethal phenotype of *cenh3-1* can be rescued by a transgene called GFP–tailswap, expressing a GFP-tagged chimeric protein containing the H3 N-terminal tail (using the H3.3 variant, encoded by At1g13370) and the CENH3 C-terminus, giving rise to sterile plants due to random meiotic chromosome segregation (Copenhaver, Ravi et al. 2011). When *cenh3* null mutants expressing altered CENH3 proteins are crossed to wild type, chromosomes from the mutant are eliminated, producing haploid progeny (Ravi and Chan 2010). From these haploid plants, sometimes doubled haploid branches emerge spontaneously, giving rise to plants that carry a uniparental genotype.

2. Aim of thesis

The inability to move makes plants susceptible in environmental changes, one of the most important evolutionary forces in nature, together with drift and recombination frequency. Over the wide geographical distribution of *Arabidopsis thaliana* a vast diversity of genotypes have been observed and documented. These populations show clear variations in development and physiology. The main research topic of this study is the role of exonuclear genome in natural variation of seven selected Arabidopsis accessions. For this reason, a cytoplasmic swap library of the seven accessions is created in order to observe the extent of the variation in physiological traits due to organellar genome. As mentioned before, previous studies have revealed a new way to create haploid plants (Ravi and Chan 2010). In this way, a library is created, where the offspring have the cytoplasm that belongs to the mother with the paternal nuclear genome. In the end, the library will contain 49 different combinations, 7 of which will be a wild-type-like combination of maternal and paternal DNA and the rest will have mixed nuclear and organelle genome. In this way, differences in photosynthesis and other physiological traits between the wild type parents and the offspring may be observed. Most likely, these differences are caused by organellar differentiation. The seven genotypes used in this study are the following:

- Shahdara (Shah) from Tajikistan,
- Columbia (Col-0) from Germany,
- Landsberg (Ler-0) from Poland,
- C24 (C24) from Portugal,
- WS4 (WS4) from Belarus (possibly),
- Burren (Bur-0) from Ireland,
- Ely (Ely) from England.

There are several reasons for the choice of these genotypes. First of all, two of them (Shah and C24) belong to different glacial refugia, namely Iberia and Central Asia and therefore fit in genetically diverged clusters (Schmid KJ 2006). This means that for more than 18,000 years the genotypes were genetically isolated from the other European genotypes and there was no exchange of genetic information. Secondly, Col-0, Ler-0 and C24 serve as reference genotypes and their genome has been fully sequenced. Apart from that, there are also individual characteristics for some of the genotypes that are interesting, as shown in several studies reviewed in (Koornneef, Alonso-Blanco et al. 2004). Shahdara, for example, causes male sterility when crossed to Monterosso (Mr-0) (P. Flood, personal communication). Ely, on the other hand, shows altered photosynthetic performance due to a single nucleotide substitution in the psbA gene (El-Lithy 2005). Moreover, experiments have shown that Burren shows a high photosynthetic ability.

The plan of action consists out of two distinct, but interconnected parts. The first one would be the creation of cytoplasmic swap offspring by crossing wt as male parents and plants carrying the partial *cenh3* mutation as females (from now on called inducer lines), as shown in figure 2.1. There, one can observe that 7 of the total 49 offspring will resemble wild-types, because the same genotype will be used as a mother and a father. These plants act as control samples. In that way, it would be possible to introduce the nuclear chromosomes of the wild-types to another cytoplasmic background. The second part would be the use of mitochondrial markers for cytoplasmic identification of the parents and also of the offspring population that will be generated.

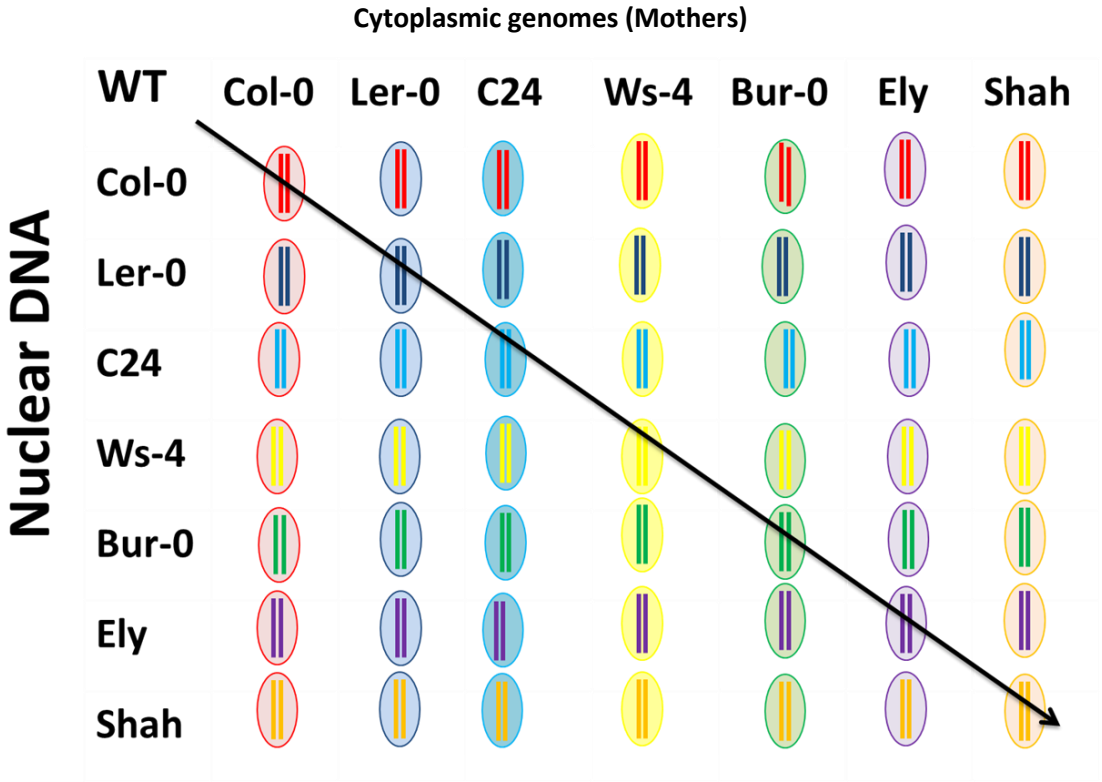


Figure 2.1: Crosses generated throughout the minor thesis project. The left column represents the male parents and the first row represents the female parents that carry the partial cenh3 mutation. 7 out of the 49 offspring would be wt-like.

3. Results

3.1 Creation of the cyto-swap haploid population

Firstly the mothers carrying the *cenh3* mutation were created. In order to do that, the plants with the chimeric CENH3 protein served as fathers and were crossed to wild-type plants. Then, the part of the offspring population that are heterozygotes are selfed and so the mutated protein is inserted into a new cytoplasmic background. Since the *cenh3* mutation is recessive, the segregants of the selfed heterozygotes served as the female inducer lines. The first step for the creation of the haploid population was to make sure that the plants used as mothers were carrying the *cenh3* mutation. For that reason, DNA was extracted from the leaves of the potential mothers and wild types and a PCR was run with the cenh3 primers (Annex I). Then, a digestion with EcoRV was performed to distinguish the wild-types from the mutants. As seen in figure 3.1, the 225bp fragment of the inducer lines is uncut due to the creation of the chimeric CENH3 protein and the wild types show a smaller fragment of 200bp.

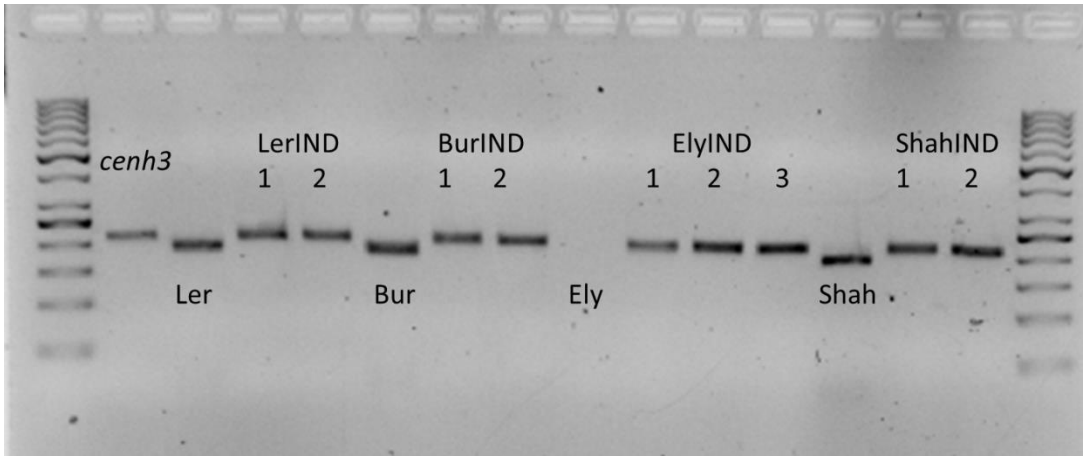


Figure 3.1: 2.5% (w/v) agarose gel containing the potential inducer lines and wild types. The 25bp difference between the wt and the inducer lines is visible in 3 out of 4 cases. 50bp GR was used as a ladder.

After establishing the inducer lines that would be used as mothers, all plants were ready for crossing. For that purpose, the female flowers were emasculated and a different string colour was used to distinguish the father in each cross. In general, 4-6 flowers were pollinated per wt that resulted in 4-5 silique per cross. Some of the crosses did not work and resulted in very small silique that contained no seeds. From literature it is known that each Arabidopsis fruit contains around 50 seeds (Meyerowitz 1989) and due to the mutation of CENH3 almost 50% of seeds per fruit were not viable. The fact that the crosses were performed between different genotypes in which the mother always contained the *cenh3* mutation that causes abortion of the nuclear chromosomes created a diverse offspring population. The population contained selfed (inducers), heterozygous, aneuploid, and haploid plants (Figure 3.2). A final estimation of

the seeds generated shows that in the end, around 85% of them were not haploid. In order to understand the offspring's level of ploidy, all the seeds were sown on rockwool blocks and were phenotypically distinguished by eye 2-3 weeks after germination. The haploid plants have certain characteristics that, in some cases, are easy to identify.

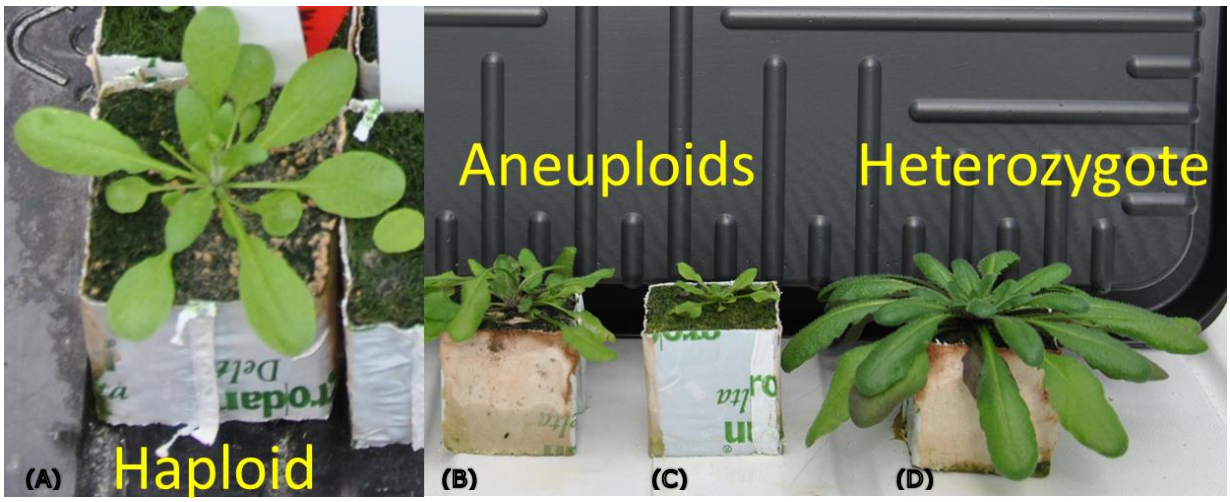


Figure 3.2: Overview of the population generated. (A) shows a haploid plant, (B) and (C) aneuploid plants, (D) heterozygote plant.

First of all, they have a smaller rosette that consists of narrower leaves. They resemble a smaller version of a wild-type plant (Figure 3.2). On the other hand, the aneuploid plants look similar to the plants under abiotic stress, and the heterozygotes are considerably bigger. After the first screen, leaf samples were sent to Rijk Zwaan Breeding B.V. for genotyping analysis. A variety of polymorphic markers was used per chromosome (Figure 3.3). The genotyping analysis is based on the polymorphism between Col-0 and Ler-0 ((A) for Col-0, (B) for Ler-0). For the selected markers these two genotypes are 100% polymorphic. The rest of the wt genotypes that were used show different patterns of polymorphisms for these alleles that are easy to distinguish. In order for a cross to be successful, the plant material has to follow the wt pattern, since the goal is to have the paternal nuclear genome in the maternal cytoplasmic background.

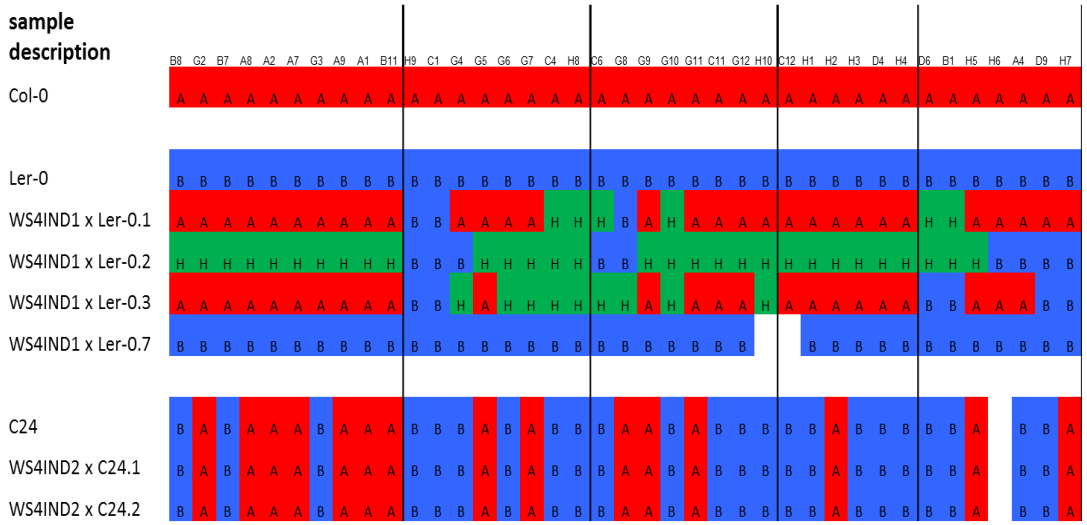


Figure 3.3: Genotyping output example from Rijk Zwaan. On the first row the sample description is observed. The first line shows the polymorphic markers used per chromosome, starting with chromosome 1. In red (A) the phenotypic pattern of Col-0 is shown and in blue (B) the phenotypic pattern of Ler-0. In green (H) a heterozygous allele is described.

As depicted in Figure 3.3, three of the crosses between WS4IND1 and Ler-0 did not work out, because a different pattern is shown than the expected Ler-0 pattern. On the other hand, the crosses between WS4IND2 and C24 show exactly the same pattern as wild-type C24, so this means that they contain the correct nuclear genome. In general, some genotypes were pollinated very easily and others acted very well as pollinators. Likewise, Ler-0 was a bad pollen donor and Ely was also pollinated repeatedly to produce haploid progenies.

A small part of the haploid plants is presented in Figure 3.4. The differences described previously are obvious also in latter developmental stages of the plant. In general, a smaller rosette with reduced leaf size is observed and the plants become very tall. This happens because the haploid plants are extremely infertile and their inflorescence keeps growing. In order to assure the cytoplasmic exchange and the creation of doubled haploid plants, one branch per plant was hand pollinated with the wild-type once again. The branch was labelled with a string so that it could be identified.

Figure 3.4: A small part of the total population of haploids that was created. The plants were individually labelled and organized based on the wt parent.

In the end, a rough estimate of the total plants generated during this minor thesis project reaches at least 700 plants. Out of this number, 47 out of 49 haploids were created, as shown in Table 3.1.

Table 3.1: Overview of the haploid cyto-swaps that were created throughout the minor thesis project. The first column shows the nuclear genotypes and the first row the cytoplasmic genomes. In green are the cybrids that were created, in white the ones that are still missing.

Nuclear genotypes (Fathers)	Cytoplasmic genomes (Mothers)						
	Col-0	**Shah**	**Ler-0**	**Ely**	**WS4**	**Bur-0**	**C24**
Col-0	C-C	C-S	C-L	C-E	C-W	C-B	C-C24
Shah	S-C	S-S	S-L	S-E	S-W	S-B	S-C24
Ler-0	L-C	L-S	L-L	L-E	L-W	L-B	L-C24
Ely	E-C	E-S	E-L	E-E	E-W	E-B	E-C24
WS4	W-C	W-S	W-L	W-E	W-W	W-B	W-C24
Bur-0	B-C	B-S	B-L	B-E	B-W	B-B	B-C24
C24	C24-C	C24-S	C24-L	C24-E	C24-W	C24-B	C24-C24

Some of the cyto-swaps, namely the plants that contained Col-0 cytoplasm were created before the start of the minor thesis by the project supervisor P. Flood. Also, a lot of the crosses did not work immediately. Actually, most of them were done repeatedly for the desired outcome. For the plants that are still missing there is a possibility of nuclear-cytoplasmic incompatibility between the genotypes. Such a scenario occurs in many plant species (reviewed in (Woodson and Chory 2008), like rice (Sato S 1999), pea (Carrie, Murcha et al. 2012), Brassicaceae (Koornneef, Alonso-Blanco et al. 2004) and is often associated with chimeric mitochondrial open reading frames (Flood P.J.).

3.2 Marker-based molecular characterization of polymorphic patterns

The detection of various polymorphisms is the source of exploring natural variation. Various molecular techniques are used to investigate the genetic diversity within and among populations, for example simple sequence repeat (SSR) that can also predict single nuclear polymorphisms (SNPs), random amplified polymorphic DNA (RAPDs), amplified fragment length polymorphism (AFLP) markers.

The second part of the minor thesis project was the molecular characterization of polymorphic marker-based patterns for the parents and the offspring generated. Since the project is focused on the cytoplasmic differences between the seven different genotypes, the markers in use should be either chloroplastic or mitochondrial. Based also on the fact that these two organelles are the only that carry exonuclear genome, three different approaches can be designed; one can use chloroplastic, mitochondrial or a combination of the above mentioned markers. For this project it was decided that the best strategy would be the use of mitochondrial markers only, because they were thought to be inexpensive, faster and easier to handle. Previous research has shown that mitochondrial polymorphisms can be established easily by PCR amplification. In some cases, additional restriction fragment length polymorphism (RFLP)-based techniques, will find the genomic rearrangements. On the other hand, the chloroplast markers used in previous research were found to be more time consuming since the PCR amplified fragments would need also sequencing to identify the genomic diversifications (Moison, Roux et al. 2010).

The first goal for this part was to create a polymorphic pattern for the wild types. Research conducted by (Allen 2003), (Allen 2003) and (Moison, Roux et al. 2010) was able to identify polymorphisms and to distinguish some of the genotypes that were used in this project. Moreover, plastid polymorphisms that were used in phylogeographic studies of *Arabidopsis thaliana* populations from the Iberian Peninsula (Pico, Mendez-Vigo et al. 2008) and China (Yin, Kang et al. 2010) were taken into consideration. In the end, 12 mitochondrial and 1 chloroplast markers were used to identify the seven different genotypes that acted as fathers and mothers. Most of them were presence/absence markers, but there were some multifragment markers as well, such as the ones designed for the *ccmC* region.

3.2.1 Polymorphic pattern of wild-types

First of all, the polymorphic wild-type pattern was explored. Figure 3.5 shows some examples of the mitochondrial markers that were used in this study. All of the markers are absence/presence markers. The first three show the presence of an open reading frame in the mitochondrial genome of the plants. Markers nad5-rrn26 rec I and recII show the presence of two recombination events that took place in the nad5-rrn26 region of the mitochondial genome (Allen 2003).

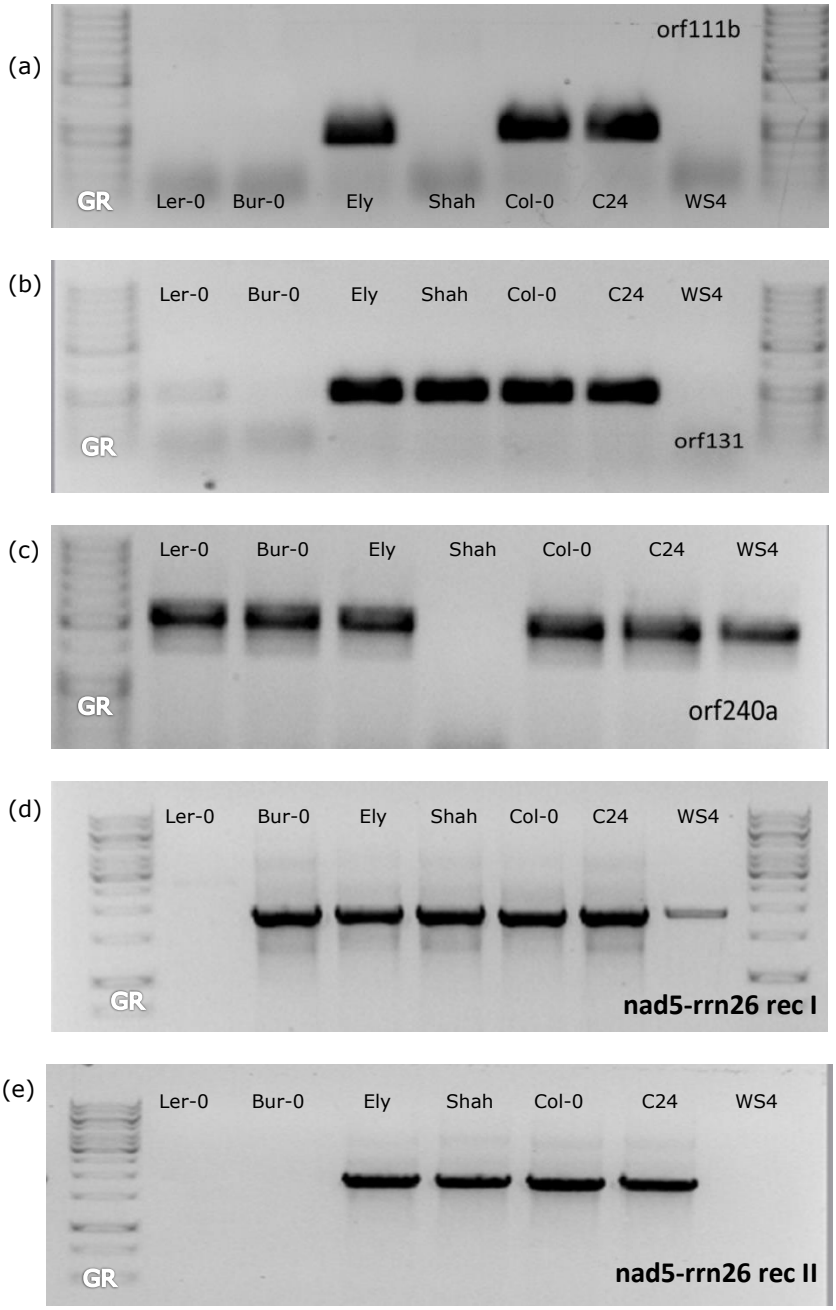


Figure 3.5: Mitochondrial polymorphic pattern for the seven wild-types. Marker examples: (a) orf111b, 253bp PCR product, (b) orf131, 239bp PCR product, (c) orf240a, 562bp PCR product, (d) nad5-rrn26 recI marker, 1,8kb PCR product, (e) nad5-rrn26 recII marker, 2kb PCR product. (a) and (b) PCR products are loaded on 1% agarose gel with 2µl of 50bp GR as ladder. For (c), (d) and (e) the ladder that was used was 1kb GR.

From this set of markers one can distinguish at least two of the genotypes. At first, Shahdara is easy distinguished because of the pattern shown in Figure 3.5 (a), (b) and (c). Moreover, Landsberg shows a polymorphic band only in orf240a and no band in the rest. These five markers are not enough to separate Burren from WS4 and Ely from Columbia and C24.

The separation of Burren with WS4 and Columbia with C24 is achieved with the use of the multifragmental marker of the *ccmC* region (Figure 3.6). With the ccmC marker, four distinct groups of polymorphic bands appear: in the first group, Ely and Columbia have fragments of the same size, one 400bp and one 260bp. In the second group, Landsberg, Shahdara and C24 have two fragments of 400 bp and 250bp length. And then, there is Burren that has a 400bp and a 240bp fragment, whereas WS4 has a 400bp and a 230bp fragment.

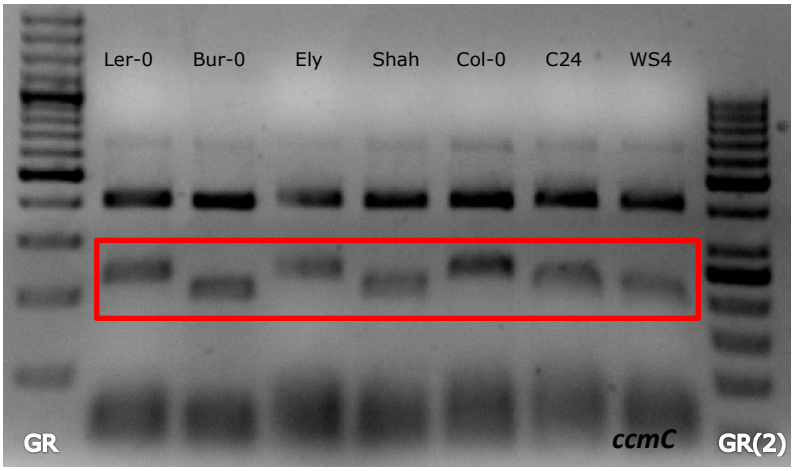


Figure 3.6: Multifragmental marker of the mt ccmC region. The red box shows the differences between the seven different genotypes. With this marker Burren (400bp and 240bp) and Columbia (400bp and 260bp) are separated from WS4 (400bp and 230bp) and C24 (400bp and 250bp), respectively. For that purpose, a 2% agarose gel was used, together with 2 different ladders: on the left 100bp+ GR and on the right 50bp GR.

The *ccmC* polymorphic differences of the genotypes are very small, so ideally the post-digested PCR products should be sequenced to identify SNPs and small in/dels, based on previous research (Moison, Roux et al. 2010). In the end, the mitochondrial markers were able to distinguish six out of seven genotypes. Ely showed an identical polymorphic profile with Col-0. From the literature it is known that Ely has a point mutation in the cp *psbA* gene (El-Lithy 2005). The primer set that was designed specifically for that area and around a restriction site distinguished Ely and Col-0 (Annex I).

3.2.2 Polymorphic pattern of inducer lines (mothers)

After creating the wild-type polymorphic profile, the same technique was applied on the inducer lines that acted as mothers in this project. The clear pattern that occurred on the wild-types changed completely (Figure 3.7).

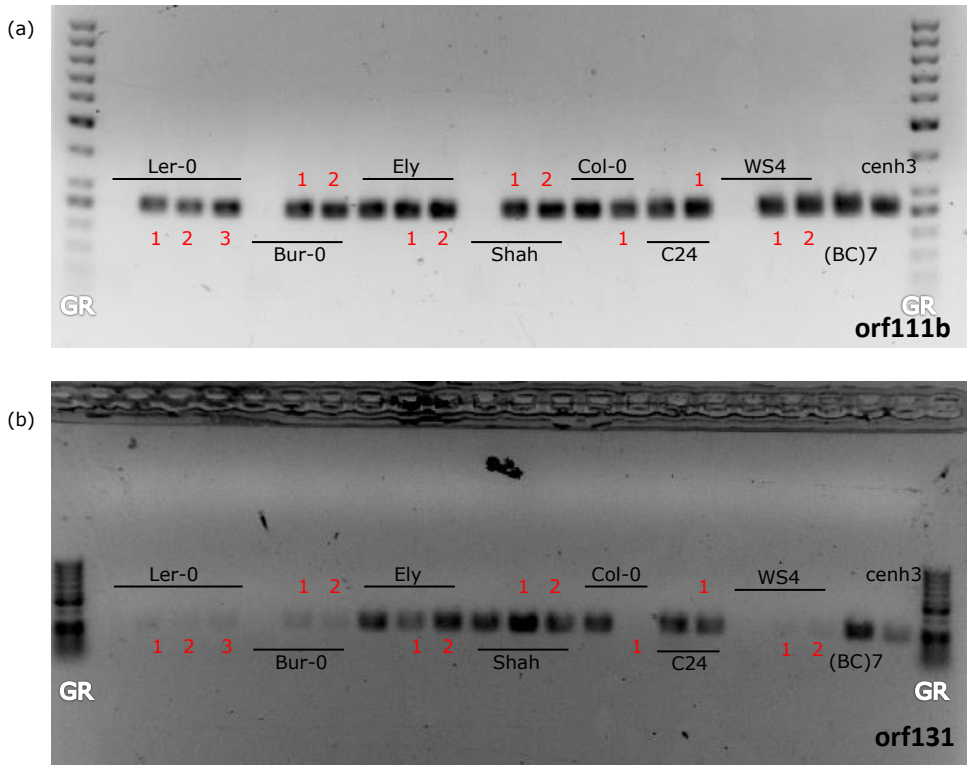


Figure 3.7: Mt markers orf111b (a) and orf 131 (b) on wt and inducer lines. The first line in each genotype is the wt. Highlighted with red numbers are the inducer lines. (BC)7 is a 7th generation backcross between Ely and Ler-0 that acts as a negative control together with *cenh3*. The PCR products are loaded on 1% agarose gel with 50bp GR as ladder.

As observed in figure 3.7, although some genotypes show no polymorphic band in the wild-type, the inducer lines show the polymorphic band that is expected, so this means that the fragments are not aspecific bands generated by the nuclear genome. At first this outcome seemed like cross-contamination of the samples, so DNA was re-extracted from wild-types and inducer lines. The same experimental conditions were applied for the second round of PCR reactions and the result was identical. The polymorphic pattern change was present only in the cases where Ler-0, Bur-0, Shah and WS4 did not have an allele present. When these genotypes showed a polymorphic band but Col-0 and C24 did not, the inducer lines were following the wild-type pattern (Annex I). The results became more confusing after a PCR with the multifragmental *ccmC* marker (Figure 3.8). In figure 3.6 two different molecular patterns were discovered; firstly, Col-0 and C24 both have a big (400bp) fragment and a smaller one (260bp (Col-0), 250bp (C24)). All the other genotypes have also two fragments, but the size of the smaller ones is easy to distinguish. In this case, the inducer lines of Bur, Shah and WS4 showed a diverse pattern.

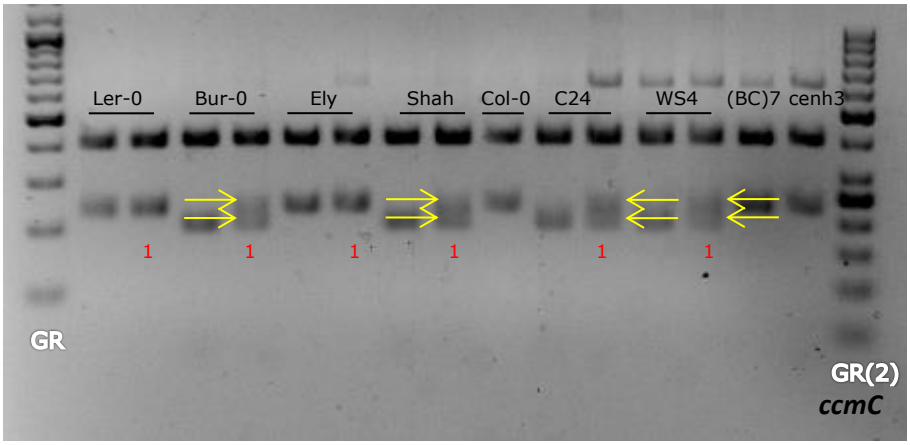


Figure 3.8: Results of *ccmC* marker. Every genotype has two samples: one wt and one inducer line (in red). The yellow arrows highlight the double bands that appear in the genotypes which show a smaller fragment than Ler-0, Ely and Col-0. The PCR products are loaded on a 2,5% agarose gel with two ladders, on the left GR 100bp+ and on the right GR 50bp.

The polymorphic pattern of the ccmC marker was puzzling. The inducer lines of the genotypes with the 260-250bp fragment follow the wild-type pattern, whereas the ones with the 240-230bp pattern appear to have a double band. The bigger band is 260-250 bp, similar to Col-0. After ruling out the chance of cross-contamination, one of the available explanations for this phenomenon would be the occurrence of heteroplasmy.

Heteroplasmy is defined as a state in which more than one mitochondrial genotype occurs in an organism (Gatenby 1990). It is known from literature that the mitochondria and chloroplasts follow a uniparental mode of inheritance, but their molecular and cellular mechanisms show that their genome partitioning does not comply with the Mendelian rules (van Gent 2001). Research has shown that the main reasons for this difference in genome partitioning, present in plants and animals, can be reduced recombination, small-scale mutations (for example replication errors, defective and inefficient repair or reaction of mtDNA with reactive oxygen metabolites) and paternal leakage (Gatenby 1990). It has been shown that the nuclear genotype strongly affects the recovery of plants bearing heritable changes in the mitochondrial genome (Herrmann, Maier et al. 2003). The ratio of different types of mtDNAs in a heteroplasmic population may be variable, but usually one mitotype is dominant and the alternative one(s) are present in a very low proportion (Gatenby 1990). Several cases in the plant kingdom have been reported, linking this phenomenon with important traits like cytoplasmic male sterility (CMS) (Noutsos 2005) and the chloroplast mutator mutant in Arabidopsis (CHM) (Jansson 1999; Herrmann, Maier et al. 2003). An interesting common feature of animal and plant heteroplasmy is that it is controlled by nuclear genes, which is a complex phenomenon involving many genes (Gatenby 1990).

The question raised about the putative event of heteroplasmy was whether this happens due to the *cenh3* mutation of the inducer lines or it is a general phenomenon that is applied in *Arabidopsis thaliana*. Previous knowledge shows that in case of chloroplasts there is no paternal transmission (Elsayed 2011). This result combined with the fact that there is lack of information about the mitochondrial inheritance pattern triggered the assumption of two distinct inheritance mechanisms for chloroplasts and mitochondria.
In order to check this hypothesis the same mitochondrial markers were used on F1 crosses between wild-type plants as a negative control. Two main reasons supported this choice: first of all, none of the parents carried the *cenh3* mutation; furthermore, previous research has shown that the core histone H3 is targeted to the mitochondria in *Brassica oleracea* (Ricchetti 1999).

Figure 3.9 (a), (b): Results of the orf111b marker on F1 wt crosses and wt plant material. The first genotype in every legend is the maternal and the second is the paternal. In this case, it seems that no paternal inheritance of mitochondria is observed for the F1 crosses. The only strange result is the polymorphic band that is observed for the cross between Ler and Col (red circle). A possible reason for that could be nuclear hybridization. The PCR products are loaded on 1% agarose gel with 50bp GR as ladder.

In figure 3.9 (a) and (b) it is observed that no mitochondrial genome is transferred from the fathers to the offspring. The best examples with orf111b are the crosses with Ely as the father where no polymorphic band is shown. The only cross that shows a different result than the one expected is Ler x Col.

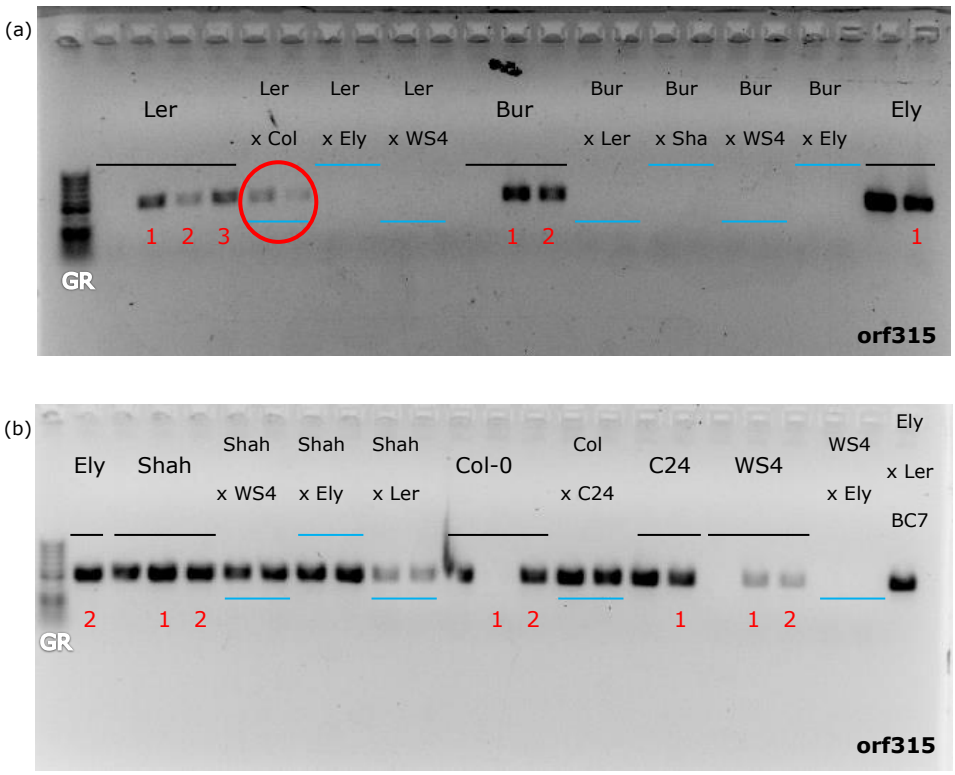


Figure 3.10 (a), (b): Results of the orf315 marker on wild-types, inducer lines and F1 wt crosses. As in 3.9, the first genotype in the F1 crosses is the maternal and the second is the paternal. In this case, it seems that no paternal inheritance of mitochondria is observed for the F1 crosses except the cross Ler x Col (red circle). The main difference remains between the wt that show no polymorphic bands and their inducer lines. The PCR products were loaded in a 1% agarose gel, the ladder was 50bp GR.

The result from orf315 (Figure 3.10) shows the same pattern as in orf111b (Figure 3.9). Again, no paternal mitochondrial leakage is observed, but the difference between orf315 and orf111b is that in 3.10 the polymorphic bands have levels of intensity. This is clear, for example, in the comparison between Shah x Ler and Shah x Ely.

3.3 Physiological experiments on cyto-swaps

As mentioned in 3.1, a small part of the population, namely the plants that contained Col-0 cytoplasm were created before the start of the minor thesis by the project supervisor. For these physiological experiments a high throughput in situ phenotyping platform was used that was built from members of Wageningen UR. This phenotyping system can measure the photosynthetic efficiency, projected leaf area, chlorophyll and anthocyanin content of 1440 Arabidopsis plants multiple times per day. The measurements of photosynthetic (photosystem II) efficiency are measured via chlorophyll fluorescence and projected leaf area is measured as a proxy for plant size (Flood P.J.).

The experiment started on the 25th of January in the growth chamber B6 in collaboration with Roxanne van Rooijen. The growth conditions were: 10 hours of light and 14 of dark, the temperature during the day was 20^{o}C and during the night 18^{o}C. At the beginning of the experiment and till the 18th of February the light intensity was 100 μmol/m^2/sec and next the plants were put under high light stress of 550 μmol/m^2/sec till the end of the experiment on the 22nd of February. The plants were measured for photosynthetic efficiency four times per day (at 09:00, 11:30, 14:30 and 16:30) and projected leaf area by measuring pixel count (Near InfraRed) five times per day (01:00, 04:00, 10:00, 13:00 and 16:00).

An overview of the results is given in figures 3.11 and 3.12. 3.11 depicts the photosynthetic efficiency of the thirteen different genotypes. The difference that is immediately obvious is the reduced photosynthesis of Ely due to the psbA mutation. Because of that, the plants show a difficulty in recovering from the high light stress. Ely plants need more time to adjust in new light conditions and from the graph it is safe to assume that this effect is caused by the chloroplast genes. This assumption is based on the fact that Ely with Columbia cytoplasm has an increased photosynthetic efficiency but also a totally different reaction in the high light stress. The cyto-swaps need less time to adjust to the new light intensity conditions and some of them, like Ely with Col-0 cytoplasm plants resembled the wild-type Col-0.

Figure 3.12 shows five different genotypes, three wild-types and two cyto-swaps. As expected, the impaired photosynthetic efficiency of Ely has an effect on the plant size as well. During the first part of the experiment with a constant low light intensity the difference is not that dramatic, but after the application of high light this becomes clearer due to the difficult adjustment of the chloroplasts to the new conditions. Again, it seems that the insertion of Columbia cytoplasm is able to partially rescue the reduced size of the plants. The pattern of the data points in the graph is caused by the periodic movement of the leaves throughout the day.

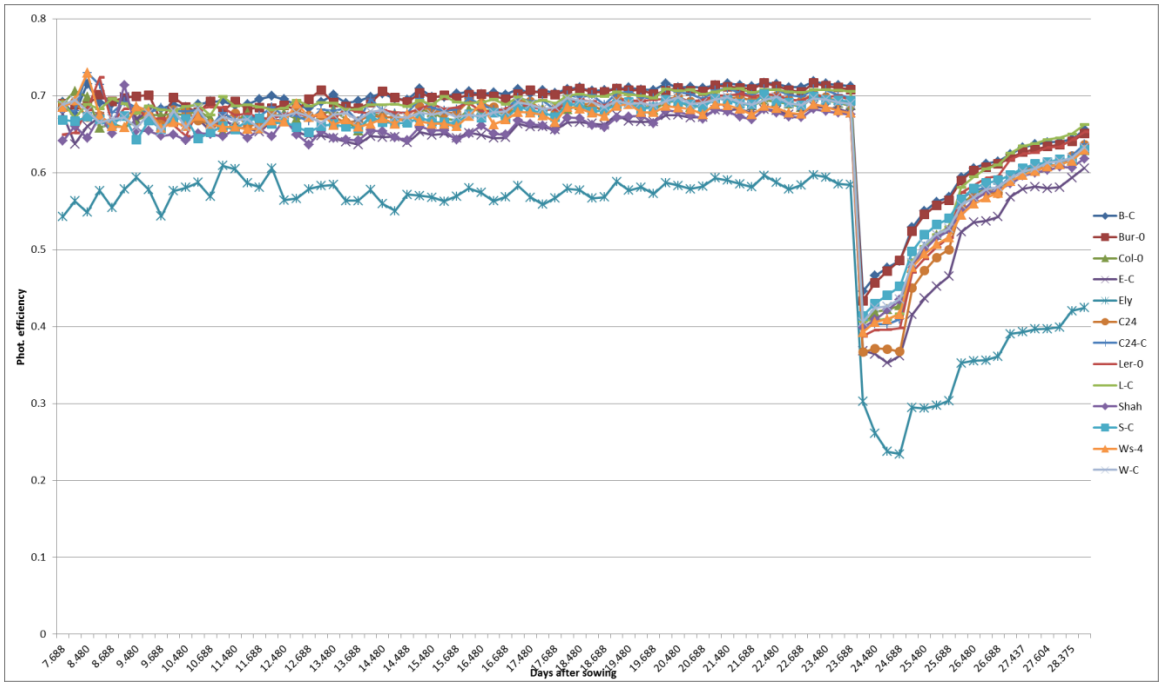


Figure 3.11: General overview of the photosynthetic efficiency from 13 genotypes: 7 wt and 6 cyto-swaps. The first observation of this graph is the immediate drop in photosynthetic efficiency after the application of 550 μmol/m2/sec of light. Ely clearly shows reduced photosynthetic efficiency due to the mutation in the psbA gene. The Columbia cytoplasm is able to partially rescue this effect, as observed in the graph. Y axis is the photosynthetic efficiency and in x axis the days after sowing are shown.

At first, five out of the six genotypes of the cyto-swap population generated show very small differences (1-2%) in photosynthetic efficiency. The same rates appear before and after the application of high light. These differences are not statistically significant (p-values ranging from 0,08 to 0,69) in all five genotypes. The same pattern appears for the dataset of projected leaf area. The increase in rosette size is slightly bigger (5-12%), but still these results are not statistically significant (two-tailed heteroscedastic t-test, p-values ranging from 0,26-0,76). The results generated from the Ely plants and the Ely plants with Columbia cytoplasm show a different aspect of the same story. The cyto-swaps demonstrate a 16% increase in photosynthesis before the high light stress and 42% after the application of high light. This is translated into an 87% and 78% increase in rosette size before and after high light stress, respectively. These results are statistically significant (two-tailed heteroscedastic t-test, p-value for plant size: 0.000600056 and for photosynthesis: 9.45531E-08). This set of experiments has shown that there is a strong correlation between photosynthetic efficiency and plant size. Moreover, after high light stress the cyto-swap plants grow even faster because they adapt easier to the new conditions.

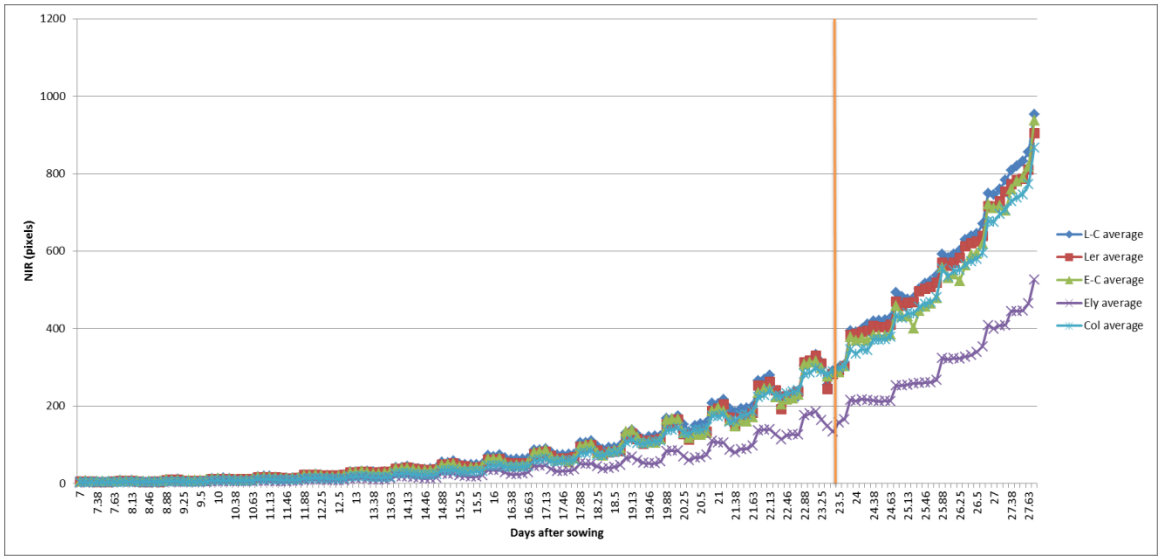


Figure 3.12: Near InfraRed measurements of five genotypes: two cyto-swaps and three wt. The orange line shows the time point when the light intensity was changed. The consequence of reduced photosynthesis of Ely is reflected in the smaller size of the plants. The Columbia cytoplasm again is able to restore the plant size, as observed in the graph. Y axis is NIR (size) in pixels and x-axis is days after sowing.

For a small part of the population generated (Ler-0, Ler-0 with Col cytoplasm, Ely, Ely with Col cytoplasm and Col-0) the chlorophyll A/B ratio was measured to observe if the ratio is changed as a result of lower PSII efficiency in different genotypes. The two-tailed heteroscedastic t-test that was performed on the chlorophyll results showed that they are not statistically significant (p value of Ler-C vs. Ler-0 = 0.693869486 and p value of Ely-C vs. Ely = 0.09392762). This experiment indicates that the reason for the previous results is not just the production of a certain pigment. This means that there are genetic reasons for this increase that may involve crosstalk between the nucleus and the organellar genome.

4. Discussion

The increasing need for food worldwide is a driving force for a big part of the plant scientific community. Therefore, the discovery of the CENH3 function in plants by Ravi and Chan 4 years ago was of great importance in the plant breeding world. This helped the development of new breeding strategies, like reverse breeding. This technique allows any desired heterozygote to be selected from a large population and be propagated indefinitely as an F_1 (Wijnker, van Dun et al. 2012).

Initially, the creation of the cyto-swap population, though labour-intensive and time-consuming, was extremely successful. The combination of numerous inhibiting factors reduced the likelihood to find a haploid plant from the seeds generated, but in the end 7% of the seeds generated were haploid with the correct nuclear SNP profile based on the Rijk-Zwaan data. During the procedure it was discovered that crosses between certain genotypes work immediately but others are very demanding. For example, C24 was a successful nuclear donor in all genotypes and on the other hand Ler-0 showed complications when crossed to WS4 and Ely. It seems though that there is no sign of genotype incompatibility and the population will be completed shortly. The next step would be a series of physiological experiments to observe whether the cytoplasmic change will affect basic plant functions, like alterations in seed germination or flowering time.

The physiological experiments are closely correlated with the molecular identification of the cyto-swap population. The first attempt of the cytoplasmic characterization involved only mitochondrial markers, as shown in 3.2. These experiments were not sufficient to distinguish all seven genotypes. For example, Ely and Col-0 shared the same polymorphic bands in every marker. Therefore, additional experiments are needed, like using chloroplast markers as well or partial sequencing of mitochondrial regions or genomes. This will give a more detailed view of the interspecific differences of different cytoplasmic genotypes in *Arabidopsis thaliana*.

Such sequence data should provide the necessary information to test whether or not heteroplasmy is truly present in the inducer lines that acted as mothers. There are several hints that point out to the involvement of the *cenh3* mutation in biparental inheritance of the mitochondria. The most convincing fact is the negative control of F_1 crosses between wild types that was presented in 3.2. There, it was shown that in the majority of the cases there is no paternal inheritance of mitochondria, so this means that they are maternally inherited. Of course the cross between Ler-0 and Col-0 needs to be further investigated, but there was no change of pattern with the use of different mitochondrial markers. So, the fact that the Ler-0 X Col-0 cross showed a heterplasmic genotype indicates a Col-0 specific problem which could be marker-related but could also

mean higher levels of paternal leakage in Col-0. The presence of a polymorphic band in both orf111b and orf315 supports the scenario of paternal leakage; however it is certain that the experiments with the mitochondrial markers need to be replicated. The next step would be to generate more crosses that will have Col-0 as the father. This will show the behaviour of a genotype that gave a polymorphic band for every marker. So, these crosses will help to establish if there is a constant inheritance pattern of chloroplasts and mitochondria when crossing different genotypes of Arabidopsis.

The small part of the cyto-swap population that was generated was used for a pilot phenotype experiment presented in 3.3. Primarily, for the traits of interest (photosynthetic efficiency, projected leaf area) the observed variations were of minor importance, but in general it was shown that different combinations of organellar and nuclear genome show different levels of response. The only significant difference was detected in the comparison between Ely and Ely with Col cytoplasm. This experiment revealed a 16% and 42% increase in photosynthesis before and after the high light stress that made the rosettes 87% and 78% bigger. First of all, this means that the Col cytoplasm is able to restore the impaired photosynthesis of Ely that is caused by the psbA mutation in the chloroplasts. Also, the crosstalk between the nuclear and organellar genome in the case of *Arabidopsis thaliana* is a plastic procedure and probably is affected by intraspecific differences. This could also suggest that the interspecific variation is probably larger than the intraspecific variation observed in these experiments. Another interesting finding is the way that the Ely cyto-swaps adapted to environmental changes, in this case high light stress. The cytoplasmic swap allowed the photosystem II to respond faster and more efficiently to the change of light intensity and therefore to imitate the response of Col-0. This means that the plants are more flexible to grow under various conditions that are not optimal. Consequently, the experiment has shown a cytoplasm-targeted approach on the increase of biomass production. The results bring out the considerable impact of photosynthesis that promotes growth of projected leaf area, but further research is needed to establish the actual correlation between photosynthesis and increase in biomass.

In the end, although the period of the minor thesis project was short, it was able to produce results that will need further investigation. The cyto-swap population that was generated will be the centre of research that will probably reveal the influence of organellar genome in basic physiological traits of the plants, something that was poorly studied in the past. Also, due to the important difference in projected leaf area there could be a prospect of studying the increase in biomass by modifying photosynthesis. In other words, with the biological tools that are available today it may be possible to allow plants to approach the theoretical yield maxima by genetically modifying the only function

that was thought to be very challenging to help breeding due to its complexity, namely photosynthesis.

The different experimental approaches of the minor thesis project point out interesting indications. First of all, the five month period of the project was sufficient to generate almost the complete cyto-swap population, Secondly; there are other aspects of the *cenh3* mutation, one of the mutations that will probably benefit plant breeding. One of them is the probability of creating a different inheritance system for mitochondria for the offspring of intraspecific crosses. It seems though that mitochondrial markers are not ideal for pattern detection. Lastly, strong indications point to the fact that organellar genome plays an important role in crucial plant traits, but different combinations of organellar and nuclear genome show different levels of response.

5. References

AGI, A. G. I. (2000). "Analysis of the genome sequence of the flowering plant Arabidopsis thaliana." Nature **408**: 796-815.

Allen, J. F. (2003). "The function of genomes in bioenergetic organelles." Philosophical Transactions of the Royal Society B: Biological Sciences **358**(1429): 19-38.

Allen, J. F. (2003). "Why Chloroplasts and Mitochondria Contain Genomes." Comparative and Functional Genomics **4**(1): 31-36.

Allen, J. F., W. B. M. de Paula, et al. (2011). "A structural phylogenetic map for chloroplast photosynthesis." Trends in Plant Science **16**(12): 645-655.

Andersson, G. E., O. Karlberg, et al. (2003). "On the origin of mitochondria: a genomics perspective." Philosophical Transactions of the Royal Society B: Biological Sciences **358**(1429): 165-179.

Carrie, C., M. W. Murcha, et al. (2012). "How do plants make mitochondria?" Planta **237**(2): 429-439.

Copenhaver, G. P., M. Ravi, et al. (2011). "Meiosis-Specific Loading of the Centromere-Specific Histone CENH3 in Arabidopsis thaliana." PLoS Genetics **7**(6): e1002121.

El-Lithy, M. E. (2005). "Altered photosynthetic performance of a natural Arabidopsis accession is associated with atrazine resistance." Journal of Experimental Botany **56**(416): 1625-1634.

Elsayed, A. (2011). Analysis of paternal plastid inheritance in Arabidopsis thaliana and its impact on biosafety of gene-modified plants. Faculty of Biology. Munich, LMU Munich. **PhD:** 80.

Flood P.J., H. J., Snel J., van der Schoor R., Jalink H., Aarts M.G.M., High-throughput phenotyping for genetic variation in Arabidopsis photosynthesis. Wageningen UR.

Flood, P. J., J. Harbinson, et al. (2011). "Natural genetic variation in plant photosynthesis." Trends in Plant Science **16**(6): 327-335.

Gatenby, A. A., Ellis, R.J. (1990). "Chaperones - assembly of ribulose bisphosphate carboxylase-oxygenase." Annu. Rev. Cell Biol. **6**: 125-149.

Gray, M. W. (1992). "The endosymbiont hypothesis revisited." Int. Rev. Cytol. **141**: 233-357.

Henikoff, S. and Y. Dalal (2005). "Centromeric chromatin: what makes it unique?" Current Opinion in Genetics & Development **15**(2): 177-184.

Herrmann, R. G., R. M. Maier, et al. (2003). "Eukaryotic genome evolution: rearrangement and coevolution of compartmentalized genetic information." Philosophical Transactions of the Royal Society B: Biological Sciences **358**(1429): 87-97.

Huang, C. Y., Ayliffe, M.A., and Timmis, J.N. (2003). "Direct measurement of the transfer rate of chloroplast DNA into the nucleus." Nature **422**: 72-76.

Jansson, S. (1999). "A guide to the Lhc genes and their relatives in Arabidopsis." Trends in Plant Science **4**: 1360-1385.

Keeling, P. J. and J. D. Palmer (2008). "Horizontal gene transfer in eukaryotic evolution." Nature Reviews Genetics **9**(8): 605-618.

Kleine, T., U. G. Maier, et al. (2009). "DNA Transfer from Organelles to the Nucleus: The Idiosyncratic Genetics of Endosymbiosis." Annual Review of Plant Biology **60**(1): 115-138.

Koornneef, M., C. Alonso-Blanco, et al. (2004). "Naturally Occurring Genetic Variation Inarabidopsis Thaliana." Annual Review of Plant Biology **55**(1): 141-172.

Leister, D. (2003). "Chloroplast research in the genomic age." Trends in Genetics **19**: 47-56.

Lewin, B., Ed. (2007). Cells. Sudburry, MA, Jones and Bartlett.

Ma, H. (2006). "A Molecular Portrait of Arabidopsis Meiosis." The Arabidopsis Book **53**(1): 1.

Maluszynska, J., Heslop-Harrison, J.S. (1991). "Localization of tandemly repeated DNA sequences in Arabidopsis thaliana." The Plant Journal **1**(2): 159-166.

Martin, W. (2002). "From the Cover: Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus." Proceedings of the National Academy of Sciences **99**(19): 12246-12251.

Martin, W., Herrmann, R.G. (1998). "Gene Transfer from Organelles to the Nucleus: How Much, What Happens, and Why?" Plant Physiol. **118**: 9-17.

Martin, W. and K. Kowallik (1999). "Annotated English translation of Mereschkowsky's 1905 paper 'Über Natur und Ursprung der Chromatophoren imPflanzenreiche'." European Journal of Phycology **34**(3): 287-295.

Martinez-Zapater, J. M., Estelle, M.A. and Somerville C.R. (1986). "Highly repeated DNA sequence in Arabidopsis thaliana " Molecular Genetics and Genomics **204**: 417-423.

Mereschkowsky, C. (1905). "Über Natur und Ursprung der Chromatophoren im Pflanzenreiche." Biol Centralbl **25**: 593-604.

Meyerowitz, E. M. (1989). "Arabidopsis, a useful weed." Cell **56**: 269-269.

Moison, M., F. Roux, et al. (2010). "Cytoplasmic phylogeny and evidence of cyto-nuclear co-adaptation in Arabidopsis thaliana." The Plant Journal **63**(5): 728-738.

Muller, H. J. (1964). "The relation of recombination to mutational advance." Mut. Res. **1**: 2-9.

Noutsos, C., Richly, E. and Leister, D. (2005). "Generation and evolutionary fate of insertions of organelle DNA in the nuclear genomes of flowering plants." Genome Research **15**(5): 616-628.

Pfannschmidt, T. (2003). "Chloroplast redox signals: how photosynthesis controls its own genes." Trends in Plant Science **8**: 33-41.

Pico, F. X., B. Mendez-Vigo, et al. (2008). "Natural Genetic Variation of Arabidopsis thaliana Is Geographically Structured in the Iberian Peninsula." Genetics **180**(2): 1009-1021.

Raven, J. A., Allen, J.F. (2003). "Genomics and chloroplast evolution: what did cyanobacteria do for plants?" Genome Biology **4**(3).

Ravi, M. and S. W. L. Chan (2010). "Haploid plants produced by centromere-mediated genome elimination." Nature **464**(7288): 615-618.

Ravi, M., P. N. Kwong, et al. (2010). "The Rapidly Evolving Centromere-Specific Histone Has Stringent Functional Requirements in Arabidopsis thaliana." Genetics **186**(2): 461-471.

Ricchetti, M., Fairhead, C., Dujon, B. (1999). "Mitochondrial DNA repairs double-strand breaks in yeast chromosomes " Nature **402**: 96-100.

Richly, E. a. L., D. (2004). "NUMTs in Sequenced Eukaryotic Genomes " Molecular Biology and Evolution **21**: 1081-1084.

Salisbury F., R. C. (1992). Plant Physiology. Belmont, California, Wadsworth.

Sato S, N. Y., Kaneko T, Asamizu E, Tabata S. (1999). "Complete structure of the chloroplast genome of Arabidopsis thaliana." DNA Res. **6**(5): 283-290.

Schmid KJ, T. O., Meyer R, Schmuths H, Hoffmann MH, Altmann T. (2006). "Evidence for a large-scale population structure of Arabidopsis thaliana from genome-wide single nucleotide polymorphism markers." Theor. Appl. Genet.(112): 1104–1114.

Simoens, C. R., Gielen, J., Van Montagu, M. and Inze, D. (1988). "Characterization of highly repetitive sequences of Arabidopsis thaliana." Nucleic Acids Res. **16**: 6753-6766.

Stegemann, S. (2003). "High-frequency gene transfer from the chloroplast genome to the nucleus." Proceedings of the National Academy of Sciences **100**(15): 8828-8833.

Stupar, R. M., J. W. Lilly, et al. (2001). "Complex mtDNA constitutes an approximate 620-kb insertion on Arabidopsis thaliana chromosome 2: Implication of potential sequencing errors caused by large-unit repeats." Proceedings of the National Academy of Sciences **98**(9): 5099-5103.

Sullivan, B. A., Blower, M.D. and Karpen G.H. (2001). "Determining centromere identity: Cyclical stories and forking paths " Nature Reviews Genetics **2**: 584-596.

Talbert, P. B. (2002). "Centromeric Localization and Adaptive Evolution of an Arabidopsis Histone H3 Variant." The Plant Cell Online **14**(5): 1053-1066.

Taylor, F. (1974). " Implications and extensions of the Serial Endosymbiosis Theory of the origin of eukaryotes. ." Taxon **23**: 229-258.

Thorsness, P. E., Fox, T.D. (1990). "Escape of DNA from mitochondria to the nucleus in Saccharomyces cerevisiae." Nature **346**: 376-379.

Turner, S., Pryer, K. M., Miao, V. P. W. & Palmer, J. D. (1999). "Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis." J Eukaryot Microbiol. **46**: 327-338.

Unseld M, M. J., Brandt P, Brennicke A (1997). "The mitochondrial genome of Arabidopsis thaliana contains 57 genes in 366,924 nucleotides." Nat. Genet. **15**: 57-61.

van Gent, D. C., Hoeijmakers, J.H.J. and Kanaar, R. (2001). "Chromosomal stability and the DNA double-stranded break connection." Nature Reviews Genetics **2**: 196-206.

Wijnker, E., K. van Dun, et al. (2012). "Reverse breeding in Arabidopsis thaliana generates homozygous parental lines from a heterozygous plant." Nat Genet **44**(4): 467-470.

Woodson, J. D. and J. Chory (2008). "Coordination of gene expression between organellar and nuclear genomes." Nature Reviews Genetics **9**(5): 383-395.

Yin, P., J. Kang, et al. (2010). "The origin of populations of Arabidopsis thaliana in China, based on the chloroplast DNA sequences." BMC Plant Biology **10**(1): 22.

Yu, Q.-B., G. Li, et al. (2008). "Construction of a chloroplast protein interaction network and functional mining of photosynthetic proteins in Arabidopsis thaliana." Cell Research **18**(10): 1007-1019.

Yu, X., Abram, G. (1999). "Patching Broken Chromosomes with Extranuclear Cellular DNA " Molecular Cell **4**: 873-881.

Yuan, Q., Hill, J., Hsiao, J., Moffat, K., Ouyang, S., Cheng, Z., Jiang, J., and Buell, C.R. (2002). "Genome sequencing of a 239-kb region of rice chromosome 10L reveals a high frequency of gene duplication and a large chloroplast DNA insertion." Molecular Genetics and Genomics **267**(6): 713-720.

6. Materials and Methods

Plant growth conditions

The seeds were put for stratification on filter paper together with milli-Q water for two days at 4°C. Then, the seeds were transferred to Radix Serre, at 6.11, 6.15 and 3.2 to grow on rockwool blocks. The plants were grown in the greenhouse under natural and artificial light that would ensure at least 16 hours of light per day at 21°C.

Molecular techniques

PCR

For this experiment 2,5 µl per reaction of the 5X Green GoTaq® Reaction Buffer and 0,25µl of GoTaq® polymerase (Promega) was used. As a template 0,5 µl of gDNA was used, along with 1 µl of forward and reverse primer (10 µM), 15,25 µl of water and 0,5 µl of dNTPs. The total volume of the reaction is 20 µl. The PCR conditions were varying, depending on the primers. Further information is provided in Annex I. A generalization of the protocol would be: 95^{o}C for 10 sec, then 35 cycles of: 95^{o}C for 30 sec, annealing temperature 54^{o}C for 30 sec, extension temperature 72^{o}C (30 seconds/1000bp). The next step was the final extension, 5 minutes at 72^{o}C. Then the samples were left at 4^{o}C.

DNA gel electrophoresis

DNA samples were mixed with 6x loading dye by Fermentas (10 mM Tris-HCl (pH 7.6), 0.03% bromophenol blue, 0.03% xylene cyanol FF, 60% glycerol, and 60 mM EDTA) and loaded onto aragose TAE gels containing ethidiumbromide. As a ladder various versions of GeneRuler (Fermentas) were used. Electrophoresis was performed at 100V and DNA bands were visualized using the Gel Doc™ XR+ System (Bio Rad). The agarose percentage of the gels was from 1 till 2,5%. The voltage for the gels was 120V.

Genotyping

The plants were genotyped by Rijk Zwaan Breeding B.V, with the help of Bastiaan de Snoo, molecular biology analyst. 6-10 markers per chromosome were used.

CTAB DNA extraction from plant material

The CTAB DNA extraction buffer consists of 75 ml water, 10 ml 1M Tris (pH 7,5), 14 ml 5M NaCl and 2 ml 0,5M EDTA. The total volume of the buffer is 100 ml. Just prior to use 1g of CTAB (Sigma M-7635) is added and fully dissolved. First, 300 µl of buffer is added to each sample and the tubes are inverted a few times. Then, the samples are put at 65^{o}C for at least 30 minutes. After that, 200 µl of chloroform is added; the tubes are shaken thoroughly and centrifuged at 3500 rpm for 15 min. The transparent phase of the mix is collected and 200 µl of isopropanol is added. Again, the samples are centrifuged for 30 minutes at 3500

rpm in the cold room. The DNA is gathered as a pellet in the bottom of the tube and the supernatant is discarded. Then, two washes with 200 µl of 70% EtOH are performed and the samples are centrifuged once more for 15 minutes in the same conditions. After that, the ethanol is discarded and the pellet is air-dried for at least 30 minutes. In the end, 200 µl of Tris is added. The samples are stored in the fridge (4^{o}C).

Concentration measuring

The concentration of samples was measured with a Nanodrop device according to the manufacturer's instructions. When this was not enough, a gel electrophoresis was used; containing 3 µl of gDNA mixed with 2 µl of 6x loading dye (Fermentas) and 3 µl of 1kb GeneRuler for comparison.

Chlorophyll A/B measuring

Extraction

Two leaves stored in -80^{o}C were put in a glass bottle and then 3 ml of DMF was added. The sample were left at -20^{o}C for at least one week for complete extraction.

Absorbance measurements

The measurements were performed with the spectrophotometer Varian Cary 4000 (Varian, Inc.) and the software CaryUV-Advanced Reads. For the green samples 1,25 ml DMF was put into a UV cuvette and then 1,25 ml of the extract was added. A new map was created and the already saved methods were copied into that file. Then the settings were checked and a sample sequence was created. The instrument was calibrated by using 'ZERO' and with a pure DMF sample in position 1 of the cell chamber. The absorbance was measured in four different wavelengths, 750, 663.8, 646.8 and 480 nm. If the absorbance at 750 nm is bigger than 0.015 it means that the extract is turbid and the rest of the measurements will give false values, so the solution to that is to centrifuge at maximum speed at 4^{o}C and the samples are re-measured.

Calculation

The Wellburn equations are used to determine concentrations of chlorophyll a (Chl_a) and b (Chl_b), as well as total carotenoids (C_{x+c}) in µg/ml (Huang 2003):

$$Chl_a = 12\ A_{663.8} - 3.11\ A_{646.8}$$

$$Chl_b = 20.78\ A_{646.8} - 4.88\ A_{663.8}$$

$$C_{x+c} = (1000\ A_{480} - 1.12\ Chl_a - 34.07\ Chl_b)/245$$

7. Acknowledgements

"Everything existing in the universe is the fruit of chance and necessity."

— Democritus

First of all, I would like to thank my parents, my brother and my aunt for their support, both financial and psychological. I feel really privileged and lucky that I was able to live my dream of studying abroad at a very high level. None of this would have happened without their help.

Secondly, I would like to express my gratitude to my professor, Dr. Mark Aarts and my supervisor, Padraic Flood, for their help and guidance throughout my minor thesis. They have been very kind and helpful whenever I needed them. I would like to thank them also for letting me work with them on a very interesting project.

Thirdly, I would like to thank all the members of the group and the rest of the people from the Laboratory of Genetics and Rijk Zwaan: Maarten, Ross, Charles, Ana Carolina, Frank, Nihal, Roxanne, Corrie, Ping Ping, Erik, Bas, Laurens, Jose, Mohammed, Hans. All of you were very helpful and I experienced a very nice atmosphere within the group that I really enjoyed.

8. Annex I

Table 1: List of mitochondrial primers used in this study.

Code	Gene	Sequence	F/Rev	Length (bp)	Tm (°C)	GC (%)	Product size (bp)
PF00028	orf131F	CTCTTACGGTCATCTATAT	For	19	47.61	36.8	239
PF00029	orf131R	GATCAACCATTTGGTGA	Rev	17	48.7	41.18	239
PF00030	cox2F	GAATAGCTATCCCGTCGTACT	For	21	56.5	47.62	PRC Product: 846bp. EcoRI dig: 1: 551, 296 bp. 2: 296 , 277 , 273 bp
PF00031	cox2R	TGTCTACGAAGCTCCGCTCCC	Rev	21	63.91	61.9	PRC Product: 846bp. EcoRI dig: 1: 551, 296 bp. 2: 296 , 277 , 273 bp
PF00032	orf240aF	ATATCTATCAACGCCTGA	For	18	49.47	38.89	562
PF00033	orf240aR	ATTTGGAGAGTACACGGAAAC	Rev	21	55.91	42.86	562
PF00034	ccmCF	GTTTGTTTTCGCTCTGCTTGT	For	21	58.2	42.86	PCR product: 631bp. ClaI dig: 1: 400, 200bp. 2: ca 400, 230bp
PF00035	ccmCR	AAGGAAAGTGGCTCAAAACGC	Rev	21	59.93	47.62	PCR product: 631bp. ClaI dig: 1: 400, 200bp. 2: ca 400, 230bp
PF00036	atp9-cox3F	GCTGCTATCGGTATTGGAAAC	For	21	57.2	47.62	1.1 kb
PF00037	atp9-cox3R	GTATTAAACCTTCCCCTCCAC	Rev	21	55.64	47.62	1.1kb
PF00038	orf262F	CGTTTCCCGTTTCAATCTTAT	For	21	54.59	38.1	628

PF00039	orf262R	GCGAGCTGTTCCGAAGTTGC	Rev	20	62.5	60	628
PF00040	orf111bF	GATATCCCTGGCACAACAGA	For	20	56.99	50	253
PF00041	orf111bR	TTTCTCGATTAAATGCCTG	Rev	19	50.8	36.84	253
PF00046	orf 120F	CTTTTATTCCTAGGTGGCTT	For	20	52.5	40	312
PF00047	orf 120R	GTGGACTTCCCAAGACTTTCA	Rev	21	57.7	47.6	312
PF00048	orf107dF	ATTTTGGTTTCACCACG	For	17	50.2	41.2	284
PF00049	orf107dR	GTCAAAATAGATCCTGGTGTG	Rev	21	54.2	42.8	284
PF00050	orf315F	AAATGCCTCTTACGGTCAT	For	19	54.1	42.1	493
PF00051	orf315R	CCCTGTTGGTACCTCC	Rev	16	52.2	62.5	493
PF00052	psbA2F	CTATGCATGGTTCCTTGGTAACTTC	For	25	54	44	350
PF00053	psbA2R	CGTTCATGCATAACTTCCATACCA	Rev	24	54	41.6	350

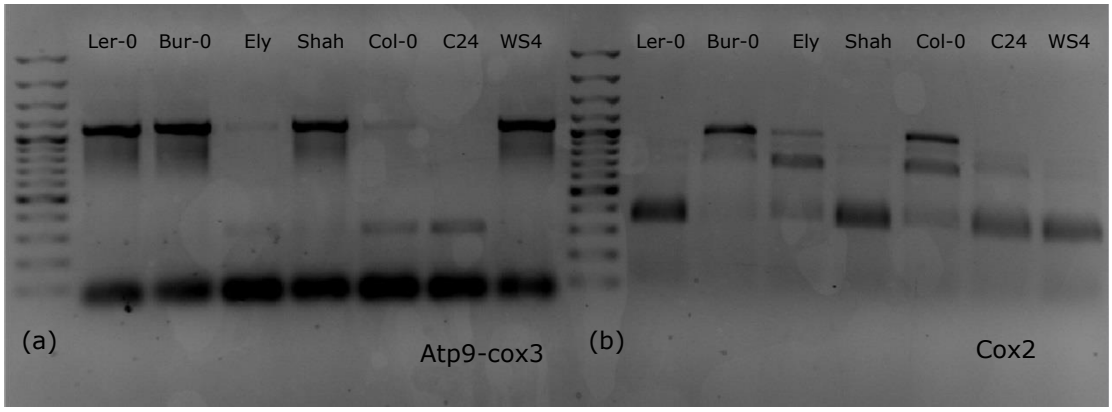


Figure 8.1: Mitochondrial polymorphic pattern for the seven wild-types: (a) atp9-cox3 , (b) cox2. PCR products are loaded on 1% agarose gel with 2µl of 50bp GR as ladder.

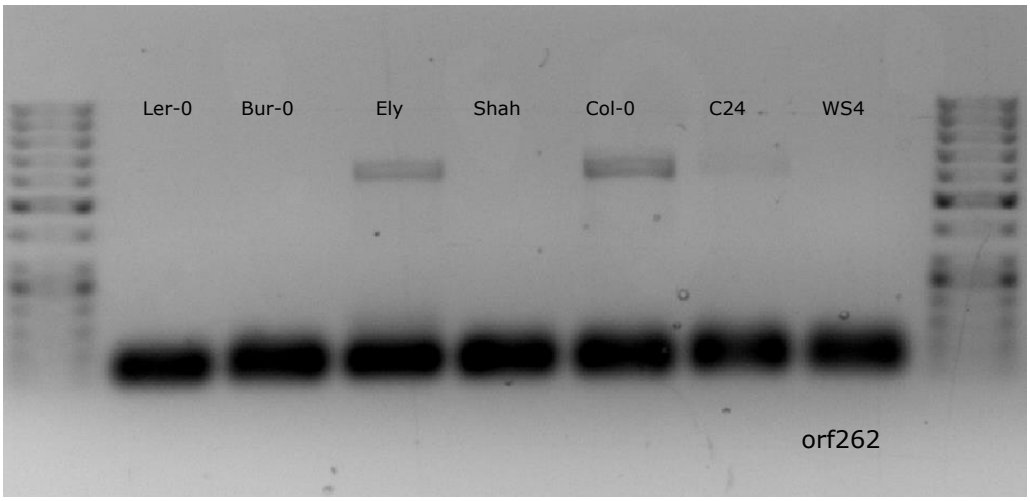
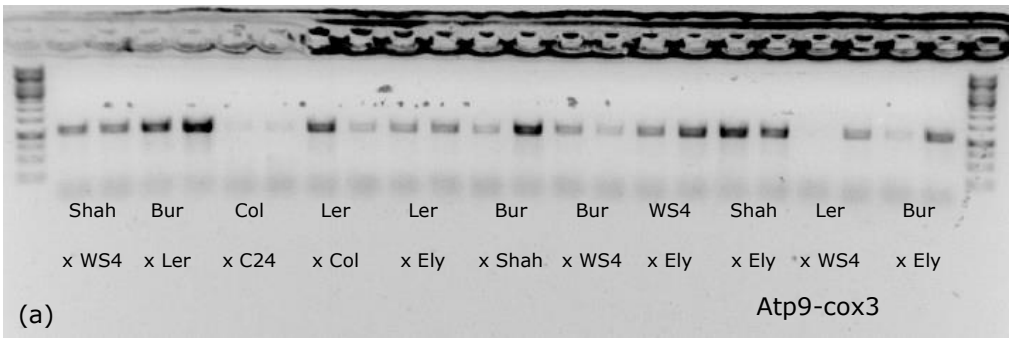


Figure 8.2: Mitochondrial polymorphic pattern for the seven wild-types with the orf262 marker. PCR products are loaded on 1% agarose gel with 2µl of 100bp+ GR as ladder



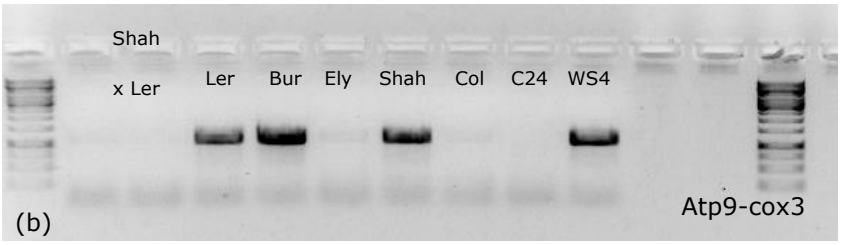


Figure 8.3 (a), (b): Results of the atp9-cox3 marker on F1 wt crosses and wt plant material. The first genotype in every legend is the maternal and the second is the paternal. PCR products are loaded on 1% agarose gel with 2µl of 50bp GR as ladder.

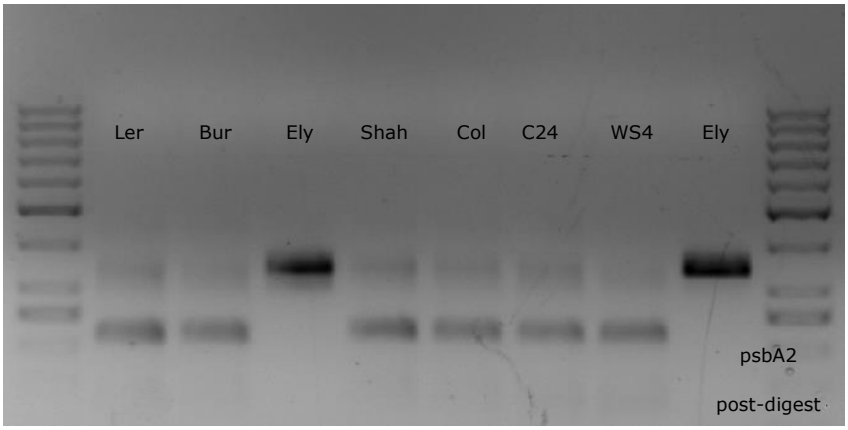


Figure 8.4: Polymorphic pattern, post digest, for the seven wild-types with the chloroplast marker designed for the psbA gene. PCR products are loaded on 1% agarose gel with 2μl of 50bp GR as ladder.

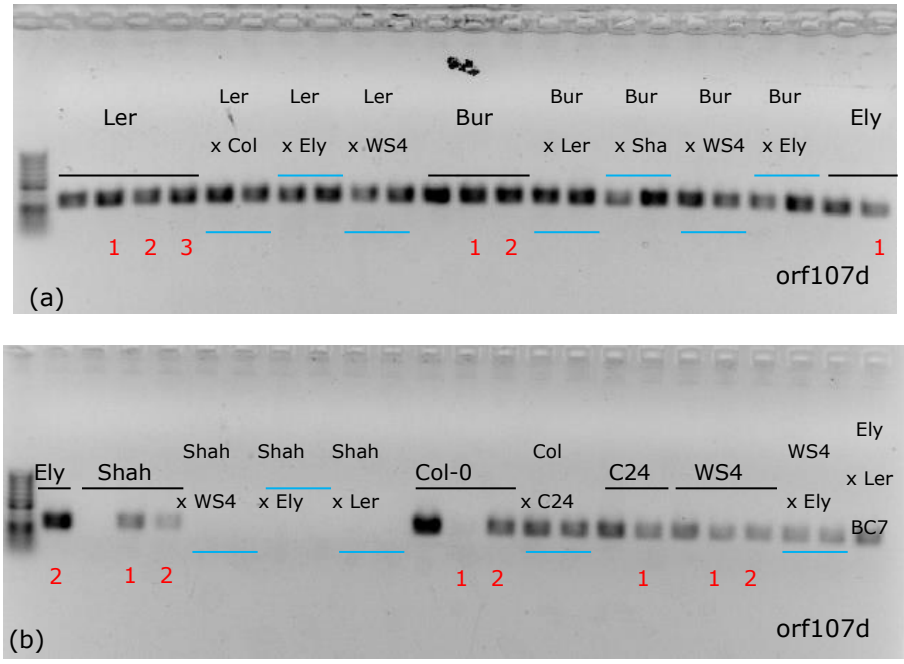


Figure 8.5 (a), (b): Results of the orf107d marker on wild-types, inducer lines and F1 wt crosses. The first genotype in the F1 crosses is the maternal and the second is the paternal. The main difference exists between some of the wt that show no polymorphic bands and their inducer lines that do. The PCR products were loaded in a 1% agarose gel, the ladder was 50bp GR.

Table 8.2: Clean data of the measurements for chlorophyll A and B for Col, Ely, Ely with Col cytopl., Ler and Ler with Col cytopl.

Sample descr.	Chlorophyll A	Chlorophyll B	A/B	t-test
blank	0.00144	-0.0005856	-2.45902	
blank	-0.0002312	-0.0014672	0.157579	
col1	2.4098038	0.5699156	4.228352	
col2	4.7684991	1.2064762	3.952419	
col3	2.5074575	0.6147042	4.079129	
col4	2.3357221	0.5701838	4.096437	
col5	4.8240427	1.2080874	3.993124	
col6	3.4423317	0.913263	3.769267	
col7	2.7451678	0.7379724	3.719879	
col8	4.3426803	1.2836274	3.383132	
col9	3.9168884	1.0526216	3.721079	
col10	3.8206474	0.9961348	3.835472	
col11	1.0928016	0.2862456	3.817706	
col12	2.2468666	0.5682636	3.953916	
Average	**2.7467227**	**0.714674486**	**3.160605**	
ec1	4.8611548	1.2884248	3.772944	
ec2	4.4714282	1.1748708	3.805889	
ec3	5.2919291	1.4424658	3.668669	
ec1	2.309112	0.5597336	4.125377	
ec2	2.3069042	0.5797564	3.979092	
ec3	4.3165353	1.0893494	3.962489	
ec1	3.3826473	0.8602542	3.932149	
ec2	6.4426702	1.735398	3.712503	
ec3	3.4735473	0.9403798	3.693771	
ec1	5.6306421	1.4940206	3.768785	
ec2	2.4087258	0.5274204	4.566994	
ec3	4.1344442	1.1245644	3.676485	
Average	**4.085811708**	**1.068053183**	**3.888762**	0.093928
ely1	3.0769018	0.8489596	3.624321	
ely2	1.8268425	0.4719094	3.871172	
ely2	2.4064615	0.6479698	3.713848	
ely1	1.9972676	0.5312088	3.759854	
ely2	3.0589974	0.8179908	3.739648	
ely2	1.5818073	0.4216798	3.751205	
Average	**2.324713017**	**0.623286367**	**3.743341**	
lc1	6.4034899	1.6289706	3.931004	
lc2	4.4167853	1.1461574	3.853559	
lc1	6.3695144	1.6827936	3.785084	
lc2	6.8597643	1.7353674	3.952918	
lc3	4.4003039	1.1282706	3.900043	
lc1	7.562357	1.9459196	3.886264	
lc2	7.4591966	1.980846	3.765662	
lc1	5.3697516	1.4156968	3.79301	
lc2	7.1409954	1.9330492	3.694161	
lc3	5.5791069	1.4821926	3.76409	
Average	**6.15612653**	**1.60792638**	**3.832579**	0.693869
ler1	7.6159806	1.9483116	3.909016	
ler2	6.1780943	1.6417898	3.763024	
ler1	5.9813147	1.6117474	3.711075	
ler2	6.2556368	1.5370736	4.069836	
ler1	4.1476219	1.1178394	3.710392	
ler1	4.8446904	1.2568704	3.854566	
ler2	6.3812522	1.7564716	3.632995	
Average	**5.914941557**	**1.552871971**	**3.807272**	