the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2010) **63**, 728–738 doi: 10.1111/j.1365-313X.2010.04275.x

Cytoplasmic phylogeny and evidence of cyto-nuclear co-adaptation in *Arabidopsis thaliana*

Michaël Moison[1], Fabrice Roux[2], Martine Quadrado[1], Romain Duval[1], Muriel Ekovich[1], Duc-Hoa Lê[1], Marie Verzaux[1] and Françoise Budar[1,*]

[1]*INRA, Institut Jean-Pierre Bourgin, UMR 1318, 78026 Versailles Cedex, France, and*

[2]*Laboratoire de Génétique et Evolution des Populations Végétales, UMR CNRS 8016, Bâtiment SN2 Université des Sciences et Technologies de Lille1, 59655 Villeneuve d'Ascq Cedex, France*

Received 15 April 2010; revised 28 May 2010; accepted 2 June 2010; published online 29 June 2010.
*For correspondence (fax +33 1 30 83 33 19; e-mail Francoise.Budar@versailles.inra.fr).

SUMMARY

In recent years *Arabidopsis thaliana* has become a model species for genomic variability and adaptation studies. Although impressive quantities of data have been gathered on the nuclear genomic diversity of this species, little has been published regarding its cytoplasmic diversity. We analyzed the diversity of plastid (pt) and mitochondrial (mt) genomes among 95 accessions, covering most Arabidopsis geographic origins. Four intergenic regions of the pt genome were sequenced, and a total of 68 polymorphisms and 65 pt haplotypes were identified. Several strategies were developed to identify mt polymorphisms among a subset of 14 accessions. Fifteen polymorphisms were further developed as PCR-based markers and used to analyze the whole set of 95 accessions. Using statistical parsimony, we built pt and mt phylogenetic networks of haplotype groups. To root the pt network, the pt intergenic regions of two related Arabidopsis species, *Arabidopsis lyrata* and *Arabidopsis arenosa*, were also sequenced. The mt and pt phylogenies are highly congruent and could be combined into a single cytoplasmic phylogeny. To estimate whether co-adaptation between nuclear and cytoplasmic genomes exists in *A. thaliana*, we tested the germination capacity in challenging conditions of 27 pairs of reciprocal F_2 families. We found that the cytoplasm donor had a significant effect on the germination capacity of some F_2 families.

Keywords: nucleo-cytoplasmic co-adaptation, natural variability, cytoplasmic phylogeny, *Arabidopsis thaliana*, mitochondrial polymorphism, plastid polymorphism.

INTRODUCTION

Organellar genomes co-evolved with the nucleus, potentially leading to co-adaptation of these genetic compartments in eukaryotic cells. In animals, examples of compensatory co-adaptation between mitochondrial (mt)- and nuclear-encoded respiratory complex subunits are known from interspecies studies (Rand *et al.*, 2004). Dowling *et al.* (2008) recently emphasized that the importance of mitochondrial–nuclear interactions in evolutionary issues has probably been widely underestimated. Plant alloplasmic lines (i.e. lines with cytoplasm from an alien species introduced into a given species) often have impaired phenotypes, such as defective chloroplast differentiation, sterility, or homeotic abnormalities in flower development, and this has been interpreted as breakdown of nucleo-cytoplasmic co-adaptation (Budar *et al.*, 2003; Herrmann *et al.*, 2003; Pelletier and Budar, 2007). Few of these situations have been investigated at the molecular level. Schmitz-Linneweber *et al.* (2005) showed that a factor necessary for the correct editing of the *atpA* transcript from the *Nicotiana tabacum* plastid (pt) is absent from the *Atropa belladona* nuclear genome, and identified this editing default as responsible for the albino phenotype of *A. belladona* plants with *N. tabacum* pt. Nucleo-cytoplasmic interactions may be important in adaptation by creating the best-fitted cyto-nuclear associations and participate in speciation by creating genetic barriers between individuals with divergent cytoplasms (Levin, 2003).

Cases of intraspecific cyto-nuclear co-adaptation are remarkably scarce in the literature. A genetic barrier resulting from nucleo-mt co-adaptation has been described within

a crustacean species (Rawson and Burton, 2002, 2006; Burton *et al.*, 2006; Harrison and Burton, 2006; Ellison and Burton, 2010). A case of breakdown of nucleo-pt co-adaptation was recently reported in *Pisum sativum*, where the pt of the subspecies *elatius* was shown to be incompatible with nuclear alleles of domesticated peas, resulting in sterile F_1 hybrids (Bogdanova *et al.*, 2009).

We initiated a project in *Arabidopsis thaliana* to study intraspecific nucleo-cytoplasmic co-adaptation. *Arabidopsis thaliana* may be an ideal species for such studies because its low out-crossing rate (Platt *et al.*, 2010) probably facilitates the linkage of nuclear alleles with the cytotype they are best fitted to. Data on cytoplasmic (ct) diversity in *A. thaliana* are very limited. Ullrich *et al.* (1997) studied mt RFLP polymorphisms among 23 *A. thaliana* genotypes. More recently, variants in the mt genome were shown to influence the maturation of mRNA ends from adjacent genes (Forner *et al.*, 2005, 2008). Chloroplast polymorphisms were used to elucidate the maternal origin of *Arabidopsis suecica*, an allotetraploid relative of *A. thaliana* (Jakobsson *et al.*, 2007), and to confirm rare paternal inheritance of plastids (Azhagiri and Maliga, 2007). Plastid polymorphisms were also used in phylogeographic studies of *A. thaliana* populations from the Iberian Peninsula (Pico *et al.*, 2008) and China (Yin *et al.*, 2010). We first evaluated and described pt and mt diversity in 95 accessions representing a good coverage of the nuclear diversity of Arabidopsis.

Mitochondria were shown to be involved in seed stress tolerance upon ageing and during germination (Macherel *et al.*, 2007), and Arabidopsis mutants with impaired mt respiration were reported to germinate less efficiently (Meyer *et al.*, 2009). In addition, pt play a central role in seed storage lipid synthesis, and impairment of this role may lead to germination deficiency (Baud *et al.*, 2007). Thus, we hypothesized that impaired mt and/or pt functions, due to breakdown of nucleo-cytoplasmic co-adaptation, might affect the capacity of seeds to germinate under challenging conditions, so we studied *in vitro* germination of reciprocal F_2 families without sugar and vitamins. Our results strongly suggest co-adaptation between the nuclear and cytoplasmic genomes within this species.

RESULTS

Analysis of plastid DNA diversity in *A. thaliana*

We sequenced four intergenic pt DNA regions in 95 accessions (Table S1 in Supporting Information). These included the 48 accessions of the core collection designed to maximize nuclear diversity (McKhann *et al.*, 2004). In addition to commonly described types of polymorphism for the pt genome, we identified an inversion in a pentanucleotide sequence in the *rbcL–accD* intergenic region. This polymorphic site is embedded in the middle of a 36-bp palindrome sequence (Figure 1). Table 1 lists the type and distribution of all pt polymorphisms detected. Altogether, the observed polymorphisms defined 65 pt haplotypes. The pt haplotype of each accession analyzed in this study, and the polymorphisms in each sequenced region, are provided as supporting information (Tables S1–S5).

Analysis of polymorphisms in the *A. thaliana* mitochondrial genome

Plant mitochondrial genomes differ mainly in organization and accumulate fewer substitutions than other genomes (Knoop, 2004; Kubo and Newton, 2008). Mitochondrial intergenic regions are generally very large, with a high proportion of non-coding sequences, and the gene order is

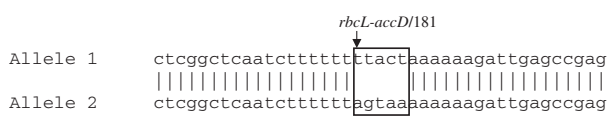

Figure 1. Inversion in the *rbcL–accD* intergenic region.
The palindromic sequence surrounding the *rbcL–accD*/181 inversion polymorphic site. The pentanucleotide stretch inverted between the two allelic forms is boxed. The coordinates of this sequence on the pt reference sequence are 56445–56485.

Table 1 Polymorphisms in plastid DNA

Plastid genome region	*MatK–trnK*	*ndhC–trnV*	*rbcL–accD*	*ndhF–rpl32*	All
Coordinates[a]	3378–4340	50289–51242	56283–57204	112547–113489	
Length of informative Sequence (bp)[b]	850	840	810	830	3330
SNPs[c]	5	18 (3)[c]	6	10	39 (3)[c]
Indels	2	7	1	4	14
Inversion	0	0	1	0	1
Microsatellites	2	3	3	6	14
All	9	28	11	20	68
Singletons	5	8	0	6	19

[a]The coordinates of the analyzed regions refer to the sequence of Col-0 plastid DNA (accession no. NC_000932; Sato *et al.*, 1999).
[b]Length with sequence reads for all accessions.
[c]Number of tri-allelic positions is indicated in parentheses. SNP, single nucleotide polymorphism.

variable between and within species (Palmer and Herbon, 1988; Unseld *et al.*, 1997; Kubo and Newton, 2008). We therefore developed a strategy to detect mt polymorphisms in a subsample of 14 accessions. These included the eight accessions of the smallest core collection defined by McKhann *et al.* (2004), the reference accessions for the pt and mt sequences (Col-0 and C24, respectively), and accessions currently used in the lab (L*er*-1, Mr-0, Ri-0, Ws) (Table S1). We took advantage of previously reported mt DNA variation (Ullrich *et al.*, 1997; Forner *et al.*, 2005, 2008) and added new polymorphisms by searching these accessions for variants in non-coding regions and in the occurrence of non-conserved open reading frames (orfs) (Appendix S1). The mt polymorphisms identified are shown in Table 2.

We then developed PCR-based strategies to genotype the polymorphic positions in the entire set of 95 accessions (Tables 2 and S6). In the *ccmC* region, sequencing of the 14 accession subsample revealed a peculiar situation. Three accessions (Col-0, L*er*-1, and Ita-0) had very similar, but non-identical, sequences that differed from that of C24 in multiple positions, whereas the sequence of the 10 remaining accessions was identical to that of C24. We concluded that two forms of the analyzed *ccmC* region exist, and developed a PCR-RFLP to score the entire set of accessions (Appendix S1 and Table S6). Accessions with the C24 form were assumed to have identical *ccmC* sequences (allele 1) whereas those with the other form might contain sequence variants. Therefore, we sequenced the *ccmC* fragment only in accessions with the second form. This identified six different alleles for the *ccmC* region (Table S7).

The results of mt genotyping of the 95 accessions are shown in Table S8. Unambiguous clear-cut results were obtained for all accessions with all markers based on PCR-RFLP, as well as amplifications to detect the rearrangement upstream of *cox3* (Forner *et al.*, 2005) and the presence/absence of *orf107d*, *orf120*, and *orf240a*. On the other hand, PCR designed to genotype other mt rearrangements gave weak amplification for several accessions, suggesting that the target sequence could be in substochiometric amounts in these samples (denoted SS in Table S8). The amplification of substochiometric molecules probably depends on the quantity and quality of the genomic DNA substrate. Thus, we only drew conclusions when weak amplification or failure to amplify was observed in at least two biological replicates.

Phylogeny of the Arabidopsis cytoplasm

We used statistical parsimony to construct a network of the phylogenetic relationships among pt haplotypes. Although rare paternal inheritance of plastids has already been evidenced in *A. thaliana* (Azhagiri and Maliga, 2007), for the construction of the network we assumed that the *A. thaliana* pt genome is maternally inherited (Martinez *et al.*, 1997; Nagata *et al.*, 1999) and evolves without genetic recombination, and regarded the four pt genome regions as a single locus. Microsatellite polymorphisms were not considered due to their high variation (Provan *et al.*, 2001), which would have blurred the resulting phylogeny. Thus, we generated 37 groups of pt haplotypes. Each group was named with a capital letter code (Table S1) and the network is shown in Figure 2.

The pt genome network showed a star-shaped topology and was reticulated at only one place. The [AK, AF, AG, AI/AJ] ring results from contradictions between single nucleotide polymorphism (SNP) *ndhF–rpl32*/672 and SNP *ndhC–trnV*/858. A recombination event between *ndhF–rpl32* and *ndhc–trnV* intergenic sequences was confirmed by the

Table 2 Polymorphisms in mitochondrial DNA

Mitochondrial genome region	Coordinates[a]	Type of polymorphism	Reference	Strategy for genotyping of 95 accessions
nad5–rrn26	11944–13795	Recombination event I	Ullrich *et al.* (1997)	Direct PCR
nad5–rrn26	16606–18591	Recombination event II	Ullrich *et al.* (1997)	Direct PCR
orf131	16965–17203	Presence/absence	This work	Direct PCR
orf315	16959–17451	Presence/absence	This work	Direct PCR
nad9-rpl16	24147–25309	indel 6 pb	This work	PCR-RFLP
cox2	39529–40374	SNP	This work	PCR-RFLP
atp8-orf107c	130231–131125	SNP	This work	PCR-RFLP
orf240a	203703–204264	Presence/absence	This work	Direct PCR
orf120	207558–207869	Presence/absence	This work	Direct PCR
orf107d	209506–209789	Presence/absence	This work	Direct PCR
cox3/atp9	217884/278985	Rearrangement	Forner *et al.* (2005)	Direct PCR
ccmC	240858–241488	Multiple SNPs	Forner *et al.* (2008), this work	PCR-RFLP + sequencing
orf262	279543–280170	Presence/absence	This work	Direct PCR
atp6-2	297337–298283	Presence/absence	Forner *et al.* (2008)	Direct PCR
orf111b	301020–301272	Presence/absence	This work	Direct PCR

[a]Coordinates refer to the sequence of reference (NC_001284; Unseld *et al.*, 1997).

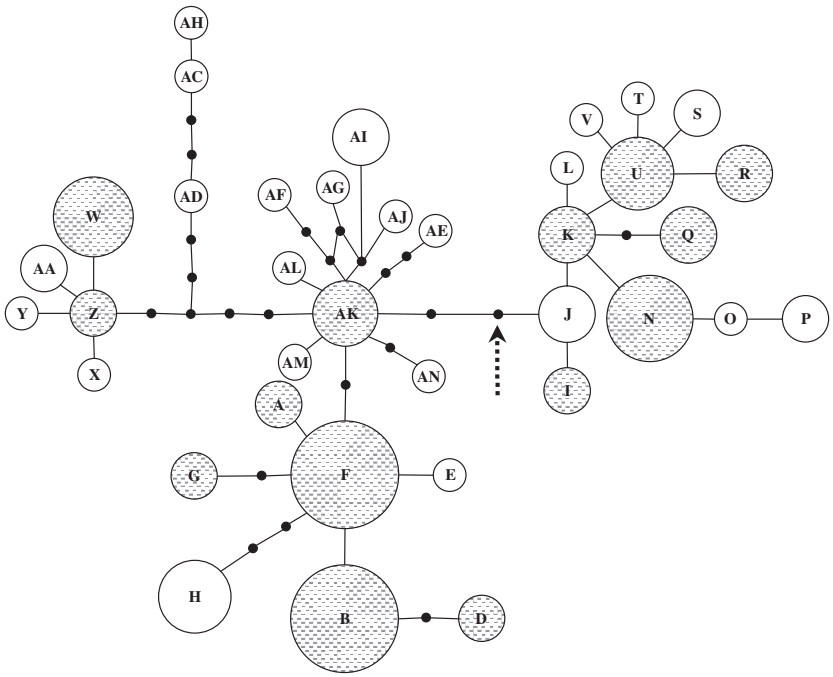

Figure 2. Phylogeny of the *Arabidopsis thaliana* plastid genome.
Each group is represented by a circle whose area is proportional to the number of accessions belonging to the group. Each group is labeled with a capital letter code. Gray filled circles correspond to groups clustering several haplotypes that differ only by microsatellite length. Each segment between circles represents a polymorphism; segment length has no particular significance. Black dots represent hypothetical intermediate haplotypes that were not observed in this study. The dotted arrow indicates the probable root of the network.

four-gamete test (Hudson and Kaplan, 1985). In addition, we observed homoplasy at position *rbcL–accD*/181 whose inversion was shared by the Y and AI groups. This inversion was also shared by *Arabidopsis arenosa*, but not *Arabidopsis lyrata*. We assume that a recombination mechanism led to the inversion at the *rbcL–accD*/181 site, similar to the flip–flop recombination observed at the large inverted repeat common in plant pt genomes. Unlike this large inverted repeat, however, the *rbcL–accD*/181 palindrome sequence would not isomerize because its relatively short length (36 bp) probably limits the frequency of recombination events. We interpreted the homoplasy observed for this polymorphism, both at the intraspecific and interspecific levels, as an indication that recombination at this site is more likely to occur several times on the time-scale of intraspecific evolution than a point mutation or an indel.

To root our *A. thaliana* pt phylogeny we sequenced the same four intergenic pt regions from two related species, *A. lyrata* and *A. arenosa*. Their sequences are very similar, but they differ from *A. thaliana* at 97 and 96 positions, respectively. We compared the *A. lyrata* and *A. arenosa* sequences with those of *A. thaliana* accessions at the sites considered for the phylogeny. At these positions, *A. thaliana*'s relatives are identical, except for the inversion at the *rbcL–accD*/181 site. They share more polymorphic positions with the J group than any other: 49 and 48 alleles out of 54 polymorphisms in *A. lyrata* and *A. arenosa*, respectively. The J group therefore appears to be closest to the common ancestral pt genome of *A. thaliana*. We placed the root on the branch between J and AK groups because the two relatives share a T at position *ndhC–trnV*/322 with the AK group.

Mitochondrial polymorphisms were then used to establish the Arabidopsis mt phylogenetic network. Rearrangements showing substochiometric shifting could lead to misinterpretations, and several cases remained undetermined after three biological replicates, leading to missing data for these particular polymorphic regions. Substochiometric shifting of mt DNA molecules is probably a frequent event, as are variations in pt microsatellite length, because such changes are observed during a plant's lifetime in a related species (Bellaoui *et al.*, 1998). In addition, a recent study suggested that the activity of *MSH1*, a nuclear gene involved in mt DNA recombination, could produce ecotype-specific stochiometric changes (Arrieta- Montiel *et al.*, 2009), and therefore these changes may not only reflect mt phylogeny. Thus, only SNPs, short indels, and genome rearrangements that did not show substochiometric shifting were taken into account. We considered the loss of *orf240a*, *orf120*, and *orf107d* as a single event because these deletions were always seen concomitantly and the genes are adjacent. We assumed that the multiple polymorphisms that distinguish the two forms of *ccmC* distinguished by a *Cla*I site resulted from a single recombination event. The single SNPs that differentiated between *ccmC* alleles 2–6 probably resulted from point mutations that occurred independently after this recombination event.

Mitochondrial haplotypes were classed into 11 groups named with lower case letters (Table S8). As with the pt genome phylogeny, and for the same reasons, we considered all mt polymorphic positions as a single locus and

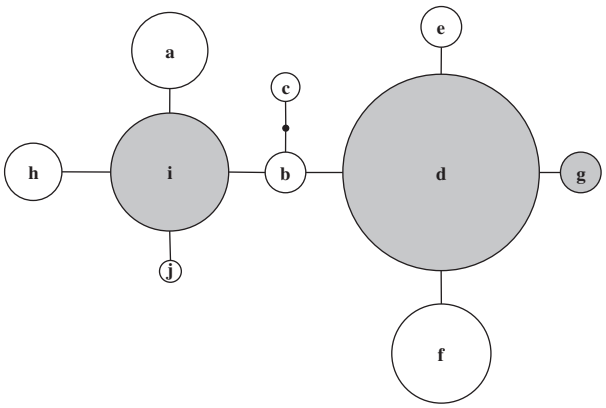

Figure 3. Phylogeny of the *Arabidopsis thaliana* mitochondrial (mt) genome. Legend is similar to Figure 2. Each group of mt haplotypes is labeled with a lower case letter. Gray filled circles correspond to groups clustering several haplotypes that differ by a rearrangement or substoichiometric shifting.

constructed a phylogenetic network using statistical parsimony (Figure 3). Its topology is star-shaped, without any reticulation.

Most mt haplotype groups coincided with one or several related groups of pt haplotypes. However, a few accessions belonging to the same pt haplotype group did not share the same mt haplotype. Firstly, Ita-0 was the only accession in pt group F which carried the c mt haplotype. Secondly, *ccmC* SNPs distinguished three mt haplotypes (h, i, and j) among accessions belonging to pt group B. Nevertheless, the pt and mt networks are remarkably congruent. This congruence is supported by a significant relationship between the pt and mt genetic distance matrices computed using the simple matching coefficient (Mantel test, Pearson $r = 0.51$, $P < 0.001$). Thus both cytoplasmic genomes appear to form a single evolutionary unit. Information from both genomes was combined to infer the ct phylogenetic network presented in Figure 4.

The sample used in our study comprised the 48 accessions defined by McKhann *et al.* (2004) to maximize the representation of nuclear diversity. The distribution of accessions from this core collection among the cytotypes indicates that this core collection also covers the cytoplasmic diversity of the considered sample. *Arabidopsis thaliana* nuclear genetic diversity is structured (Nordborg *et al.*, 2005; Ostrowski *et al.*, 2006; Beck *et al.*, 2008) despite clear evidence of recombination. We examined whether cytoplasmic and nuclear diversity patterns are correlated, using 68 accessions shared between our sample and the sample of Ostrowski *et al.* (2006). Figure 5 shows that in *A. thaliana* there is no obvious relationship between the structure of nuclear polymorphisms and the pattern of cytoplasm diversity. In addition, we found no relationship (Mantel test, Pearson $r = 0.04$, $P = 0.191$) between the nuclear and cytoplasmic distance matrices of the 86 accessions shared between our study and that of McKhann *et al.* (2004). This finding suggests that cyto-nuclear genetic associations due to the structure of genetic diversity are probably limited in *A. thaliana*, and any future data showing a strong association between a pt or mt polymorphism and a nuclear gene might be a sign of co-adaptation.

Is there nucleo-cytoplasmic co-adaptation in *A. thaliana*?

The above results were encouraging for using *A. thaliana* to study the genetic basis of nucleo-cytoplasmic co-adaptation.

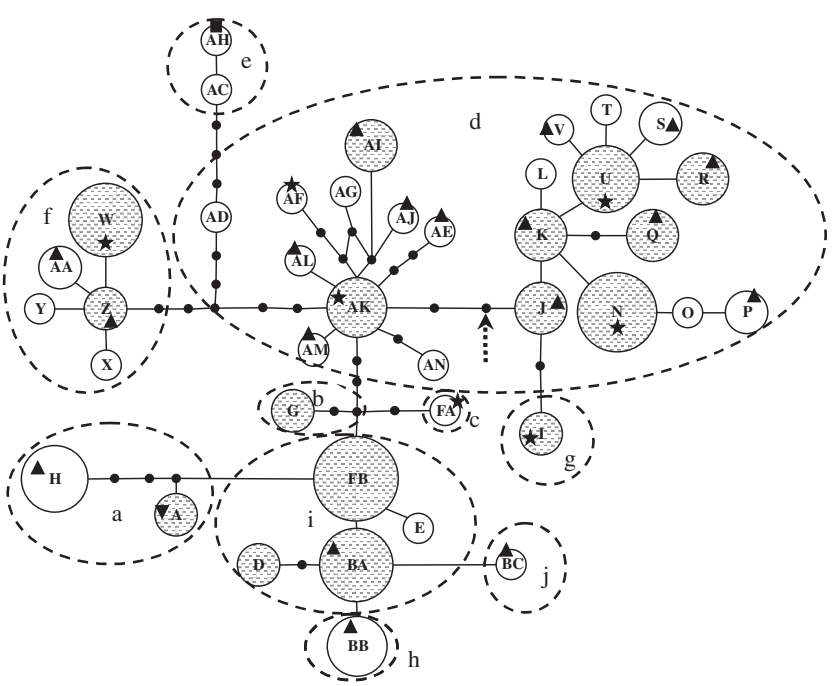

Figure 4. Phylogeny of *Arabidopsis thaliana* cytoplasm.
Legend is similar to Figure 2. Each group of cytotypes is labeled with a capital letter code that corresponds to its plastid (pt) haplotype group (except for the F pt group, divided into FA and FB, and for the B pt group, divided into BA, BB, and BC). Gray filled circles correspond to groups clustering several cytotypes. Dotted lines frame cytotypes clustered in the same mitochondrial (mt) group. Stars indicate groups containing at least one accession from the smallest core collection ($n = 8$) from McKhann *et al.* (2004). Triangles indicate groups containing at least one accession from the largest nested core collection ($n = 48$) from McKhann *et al.* (2004). The upside-down triangle indicates the group containing Col-0 (reference for the pt genome). The square indicates the group with C24 (reference for the mt genome).

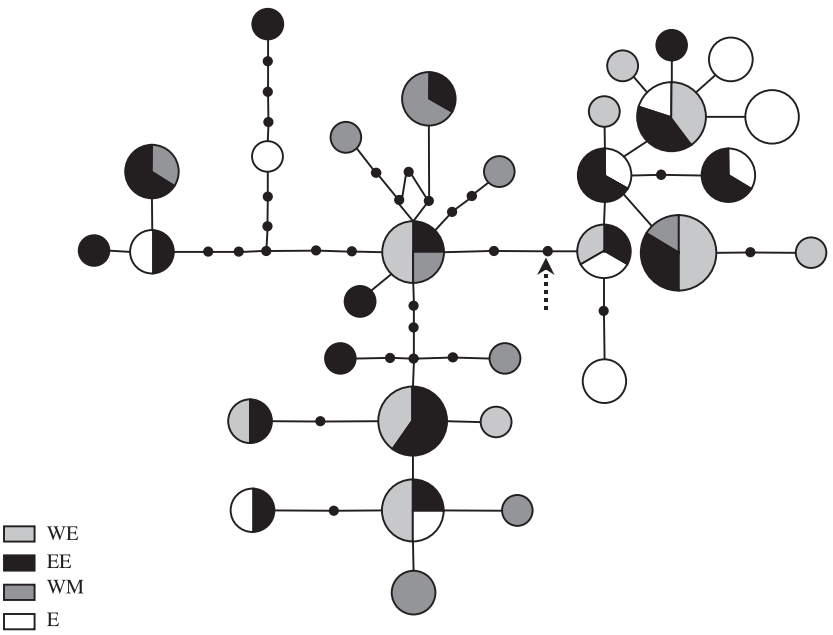

Figure 5. Distribution of nuclear diversity groups in the *Arabidopsis thaliana* cytoplasmic (ct) network.
The ct genome network of the 68 accessions in common between this study and that of Ostrowski *et al.* (2006) is presented. The legend is similar to Figure 4, except that the nuclear diversity groups defined in Ostrowski *et al.* (2006) are indicated by colors: black, EE; dark grey, WM; light grey, WE; white, E.

Nevertheless, before developing the genetic resources necessary for such a study, we tested whether nucleo-cytoplasmic interactions could be observed to influence a simple trait. If co-adaptation exists between some nuclear genes and the cytoplasmic genomes, combining nuclear alleles from one accession with the cytoplasm from another accession could sometimes lead to sub-optimal genotypes. Germination percentage has been reported to respond to differences in maternal cytoplasm in *A. thaliana* reciprocal F_1s (Corey *et al.,* 1976). We therefore selected germination as a character that could eventually be sensitive to misfit nucleo-cytoplasmic associations. However, the maternal nuclear genotype can also induce differences in reciprocal F_1s (Alonso-Blanco *et al.,* 1999). It was particularly important to design the experiment so that at least certain effects of nucleo-cytoplasmic interactions could be unambiguously detected. For this purpose, we used reciprocal F_2 families, which were obtained from selfing reciprocal F_1 hybrids from crosses of different accessions. Reciprocal F_2 families therefore have maternal parents with identical hybrid nuclear genotypes leading to the same combination of nuclear genotypes, but they differ in the cytoplasms of the two grandmothers (Figure 6).

Our first experiment measured *in vitro* germination in challenging conditions (no sugar, no vitamins) in 27 pairs of reciprocal F_2 families resulting from crosses between the eight accessions of the smallest core collection defined by McKhann *et al.* (2004) (out of the 28 theoretically possible pairs; Table S9). This core collection was designed to maximize nuclear diversity, and we found that it also provides good coverage of cytoplasmic diversity (stars in Figure 4). Seventeen pairs of F_2 families sharing the same grandparents gave similar germination percentages (data not shown). However, 10 pairs of reciprocal families showed significantly different germination percentages in our challenging experimental conditions. The germination test was repeated so that at least three biological replicates (i.e. from three genetically identical F_1 individuals) were performed for each of these 20 F_2 families (Table S10). Under our experimental conditions, the germination capacity was not affected by the time of seed dry storage (Table 3). The grandparental accessions had a significant effect on the germination capacity of the F_2 families (Table 3). This was an expected result, as incompatible allelic combinations from different accessions could lead to impaired seed development or germination (Bikard *et al.*, 2009). More interestingly, the cytoplasm donor had a significant effect within the pair of grandparents for *in vitro* germination under our challenging conditions (Table 3). Fisher's least-significance-difference test indicated that in four pairs of F_2 families (from grandparents: Jea/Blh-1, Ct-1/Cvi-0, Ct-1/Oy-0, and Bur-0/Oy-0) the germination percentage in these conditions differed depending on the grandmother's cytoplasm (Figure 7). The maximum difference in germination percentage was observed when Bur-0 and Oy-0 were used as the grandparents, the lowest germination percentage being associated with the cytoplasm from Oy-0.

Figure 8 shows the genetic distances (simple matching coefficient; Dillmann *et al.*, 1997) between nuclear and cytoplasmic genomes of the grandparental pairs used in this study. F_2 families with different *in vitro* germination capacities under our challenging conditions (black circles) appear to derive from accessions with rather distant nuclear genomes (nuclear distance > 0.5). Their cytoplasms are also

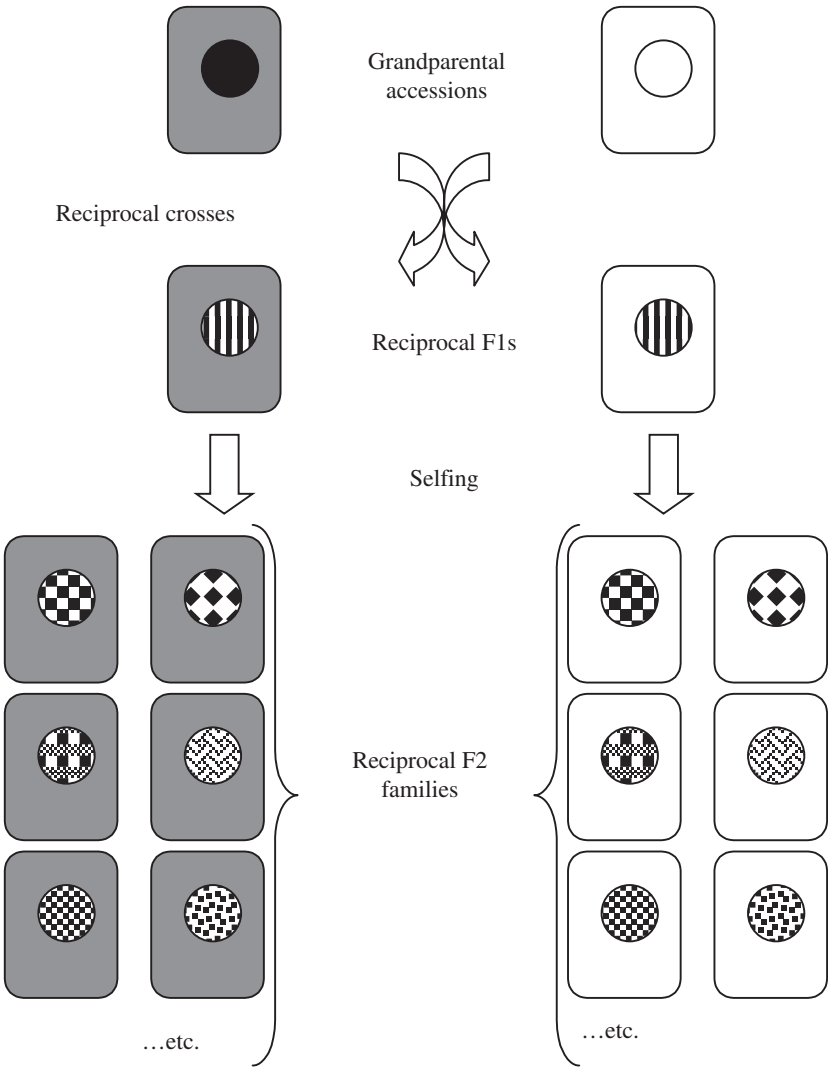

Figure 6. Genotypes of reciprocal F_2 families.
In this scheme, the circles stand for nuclear genotypes and the rectangles for the cytotype. Reciprocal crosses between grandparental accessions produce F_1s that are identical in their nuclear genome (heterozygous, as symbolized by stripped circles) but have different cytoplasms. F_2 families, resulting from selfing of F_1s, are collections of nuclear genotypes where the nuclear alleles of the grandparent accessions are distributed according to Mendel laws. Reciprocal F_2 families therefore contain the same collection of nuclear genotypes (symbolized by circles with different black and white motives) in the cytoplasm of their respective grandmothers.

rather distant (cytoplasmic distance > 0.3), but the strains Bur-0 and Oy-0, which could not be differentiated using any of our ct markers, produced the F_2 families that differed the most in their germination capacities. These two accessions could differ in an essential mt or pt gene (not detected in our ct diversity analysis), which would have selected for compensatory modifications in one or several nuclear genes.

Table 3 Statistical analysis of the germination of 20 F_2 families

Effect	d.f.	MS	*F*	*P*
Grandparents	9	0.167	11.11	***
Cytoplasm donor	10	0.054	3.57	**
Storage of seeds	1	0.029	1.95	n.s.
error	50	0.015		

The effects of the accessions used as grandparents and the cytoplasm donor (grandmother) for each pair of grandparents were tested. The length of seed storage was also tested.
d.f., degrees of freedom; MS, mean square; *F*, *F*-ratio (mean square of the considered effect/mean square error).
*0.05 > *P* > 0.01; **0.01 > *P* > 0.001; ***$P < 0.001$; n.s., not significant.

DISCUSSION

Unlike pt diversity, mt polymorphism has been rarely analyzed or used in angiosperms (Desplanque *et al.*, 2000; Mower *et al.*, 2007). The obvious reason for this is that plant mt genomes have low substitution rates (although different lineages may have variable substitution rates; Mower *et al.*, 2007). Nevertheless, our study shows that, despite its limited extent, mt polymorphism provides valuable information. Mitochondrial polymorphisms distinguished cytotypes that belonged to the same pt group. They also allowed more accurate placement of groups G and A in the phylogeny (compare Figures 2 and 4). However, it should be noted that we might have underestimated the diversity of mt groups d,

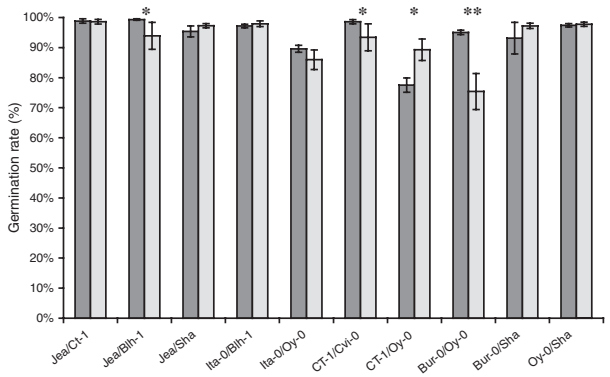

Figure 7. Cytoplasmic effect on germination efficiency in *Arabidopsis thaliana.*

The mean and standard error of the mean (SEM) of the germination percentage of F_2s plotted in pairs of grandparents. The first accession of the pair is the grandmother of the F_2 family plotted in dark gray. Planned comparisons were made for reciprocal F_2 families for each pair: *0.05 > *P* > 0.01, **0.01 > *P* > 0.001, no asterisk means 'not significant'.

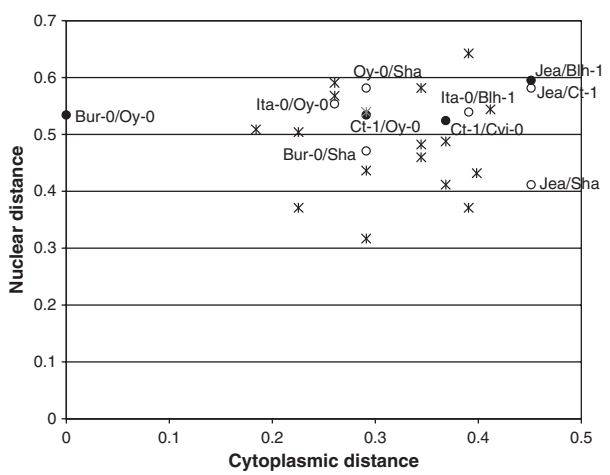

Figure 8. Genetic distances between accession pairs used as grandparents of F_2 families.

The pairs of accessions used as grandparents to produce the 54 F_2 families used in this work plotted according to their nuclear (*y*-axis) and cytoplasmic (*x*-axis) genetic distances. The grandparent pairs of F_2 families analyzed in triplicate are indicated with circles and their names are given. The grandparents whose F_2 family pairs showed significantly different germination capacities are in black.

e, f, and g by assuming that all accessions with the C24 type of *ccmC* have the same sequence in this region.

Very recently, Yin *et al.* (2010) analyzed pt diversity among 28 of the same *A. thaliana* accessions as in this study, and our results for these accessions are in agreement with theirs (Figure 8 in Yin *et al.*, 2010), except for the position of Cvi-0. Cvi-0 (AF ct group) carries the d mt haplotype, therefore its position in our network is more likely to be correct.

The congruence of pt and mt networks and the strong linkage disequilibrium between cytoplasmic genomes are expected for maternally inherited compartments, and have been observed in other species (Desplanque *et al.*, 2000; Olson and McCauley, 2000; Besnard *et al.*, 2002). Nevertheless, the reticulation of the pt network could be the result of rare recombination events following biparental transmission of pt. Paternal inheritance of pt was reported in *A. thaliana* in laboratory *in vitro* experiments under very strong selection to recover cell lineages with paternal chloroplasts (Azhagiri and Maliga, 2007). Such rare transmission of the pt genome via pollen under natural conditions could have led to the observed recombination between *ndhF–rpl32*/672 and *ndhC–trnV*/858 polymorphisms. Homoplasy could also result from frequently changing sites in the genome, as seems to be the case for the *rbcL–accD*/181 inversion.

The structure and geographic distribution of *A. thaliana* nuclear diversity were interpreted as signatures of historical demographic events such as repopulation of northern areas in post-glacial periods (Sharbel *et al.*, 2000) or dispersal by human population migrations (Beck *et al.*, 2008; Francois *et al.*, 2008). Nevertheless, we found no relationship between the patterns of ct and nuclear polymorphisms among 86 accessions. In addition, the ct diversity seems not to be structured by the geographical origin of accessions. For instance: the five accessions from Central Asia belong to cytotypes FB, I, and S; the four accessions from the American continent belong to cytotypes BA, FB, and W; the eight accessions from Japan belong to cytotypes AJ, BB, H, and X; the eight accessions from Scandinavia belong to cytotypes AK, FA, FB, G, W, and Z (Table S1, Figure 4).

On the one hand we expect the low outcrossing rate of *A. thaliana* to maintain associations between cytotypes and nuclear alleles during long periods of time, so that cyto-nuclear co-adaptation is facilitated, and therefore more likely to be detected than in outcrossing species. On the other hand, *A. thaliana*'s outcrossing rate is probably sufficient to introduce genetic recombination (Bomblies *et al.*, 2010; Platt *et al.*, 2010), which in turn shortens the length of the genome regions that are in linkage disequilibrium. The length of linkage disequilibrium was estimated at around 10 kb, on average, for *A. thaliana* (Kim *et al.*, 2007). This is in agreement with the lack of a relationship between nuclear and cytoplasmic distances in our study: it is unlikely that any of the 10 sequenced loci used in the analysis of nuclear diversity was tightly linked to genes selected for adaptation to cytotypes. So, associations between nuclear alleles and cytotypes, if found, might be signatures of co-adaptation rather than of historical demographic events.

The aim of this present work was to test for nucleo-cytoplasm co-adaptation within *A. thaliana*. *In vitro* germination capacity under challenging conditions appeared to be an appropriate, easy-to-score trait for this test, and a simple

and reasonably rapid way to explore a large number of genetic combinations was to compare the germination percentage of reciprocal F_2 families. We found a cytoplasm donor effect on germination percentage, which suggests that some nuclear genotypes (as a combination of grandparental alleles) could cope better with our challenging experimental germination conditions when associated with one cytoplasm than with the other. By studying only germination in challenging conditions, we are probably underestimating the number of sub-optimal or deleterious nucleo-cytoplasmic combinations.

Our results are an encouraging step towards the investigation of this aspect of plant biology. *Arabidopsis thaliana* provides a particularly favorable situation for identifying the factors involved in the concerted evolution of cytoplasmic and nuclear genomes and the impact of these factors on adaptation. This will imply observation of reasonably easily scored phenotypes impacted by a breakdown of nucleo-cytoplasmic co-adaptation, and identification of the genetic variations underlying the observed effects by fine genetic mapping, using dedicated genetic resources. We expect to identify genes that evolved in response to nucleo-cytoplasmic interactions and plan to study their role in fitness-related traits.

EXPERIMENTAL PROCEDURES

Plant material

All the *A. thaliana* genotypes used in this study were obtained from the Versailles Arabidopsis Resource Centre (http://dbsgap.versailles.inra.fr/vnat). Accessions used for the analysis of cytoplasmic diversity are listed in Table S1. The F_2 families used for the germination capacity measurements are listed in Table S9. They were produced and stored (4°C, 15–30% relative humidity) at the Versailles Arabidopsis Resource Centre. All F_2 families, except those involving Ita-0 as one of the grandparents, were produced in the same experiment, during autumn 2006. All F_2 families involving Ita-0 as one of the grandparents were produced in another single experiment, during summer 2008. Thus, pairs of reciprocal F_2 families were always produced in the same experiment.

The *A. lyrata* strain used was obtained from the Versailles Arabidopsis Resource Centre (accession 292AV). The *A. arenosa* ($2n = 2x = 16$) strain used was kindly provided by Dr Martin Lysak (Masaryk University, Brno, Czech Republic), and originates from Strecno, Slovakia.

Genomic DNA was extracted from plantlets. The entire set of *A. thaliana* accessions used in the analysis of cytoplasmic diversity was extracted in a 96-well plate following the procedure described in Loudet *et al.* (2002). When necessary, genomic DNA from individual accessions was extracted following the same procedure as in Bouchez *et al.* (1996).

Sequencing of pt intergenic regions

The intergenic regions considered in this study were chosen as follows. A first analysis was carried out on the smallest core-collection of accessions defined by McKhann *et al.* (2004) (eight genotypes; Table S1) and Col-0 by amplifying and sequencing eight intergenic fragments of the pt genome (data not shown). We then chose four intergenic regions exhibiting the most polymorphisms among this small sample of genotypes and carried out the analysis with these fragments on the whole sample of 95 accessions. Their positions on the pt genome are given in Table 1.

After PCR amplification from total genomic DNA, fragments were directly sequenced in both directions using the amplification primers. The primers used are listed in Table S11.

Sequences were aligned using GENALYS (Takahashi *et al.*, 2003). Whenever a polymorphism appeared in a unique accession (singleton), the sequencing was repeated on an independent amplification product before validation. The sequences of the four pt intergenic regions of *A. lyrata* and *A. arenosa* were deposited in the EMBL database (FN669599, FN669600, FN669601, FN669602, FN669603, FN669604, FN669605, FN669606).

Analysis of mt polymorphisms

The detection of mt polymorphisms and the development of PCR-based markers for genotyping are described in Appendix S1. New mt sequences were deposited in the EMBL database. Mitochondrial polymorphisms are described in Table S6. All primers used are described in Table S11.

For PCR-RFLP, amplification reaction was directly followed by digestion with restriction enzymes (Fermentas, http://www.fermentas.com/) in the conditions (reaction buffer and temperature) advocated by the manufacturer. Digestions were then analyzed by agarose gel electrophoresis (see also Table S6).

Construction of phylogenetic networks

Networks were drawn using TCS v.1.21 software (Clement *et al.*, 2000). The pt network was obtained with a connection limit of five steps. It was then manually modified to take homoplasy at the *rbcL–accD*/181 position into account (see Results). The position of the root was manually added. Nevertheless, when we added to the analysis the haplotypes of *A. lyrata* and *A. arenosa* at polymorphic positions in *A. thaliana*, the out-groups were connected to the network at the position inferred manually.

The minimum number of intragenic and intergenic recombination events was calculated using the four-gamete test (Hudson and Kaplan, 1985) using the DNASP program v.5 (Librado and Rozas, 2009).

In vitro germination capacity measurements and statistical analysis

For each reciprocal F_2 family, between 200 and 300 seeds were surface sterilized as described by Loudet *et al.* (2005) and sown on 7% agar (w/v). They were kept at 4°C in the dark for 3 days to optimize germination and transferred to a growth chamber with 50% atmospheric humidity, under a 16-h light/8-h dark cycle at 21°C during the day and 18°C at night. The number of germinated seeds was counted 4 days after transfer, as the number of seeds with an emerging root. Preliminary experiments showed that no new seeds germinated after 4 days under these conditions.

Biological replicates were performed using at least three different seed lots corresponding to three F_2 family samples, each originating from an independent F_1 plant. The results were analyzed using the following general linear model (GLM):

$$\text{germination capacity} \sim \text{mean} + \text{grandparents} + \text{cytoplasm donor (grandparents)} + \text{storage} + \text{error}$$

In this model, 'grandparents' accounts for the differences among crosses between different *A. thaliana* accessions and 'cytoplasm donor' nested in 'grandparents' measures the cytoplasmic effect

within each cross. Since the time of seed dry storage can affect the germination (Alonso-Blanco *et al.*, 2003), 'storage' is a co-variate that represents the number of months between seed harvesting and sowing (Table S10). All factors were treated as fixed effects.

Data on germination percentage were arcsin transformed to satisfy the normality and equal variance assumptions of linear regression. Model fitting was conducted using Proc GLM in SAS 9.1 (SAS Institute Inc., http://www.sas.com/), and the multiple Fisher's least-significance-difference statement was used for subsequent pairwise comparisons to examine the significance of cytoplasmic effect within each cross.

Genetic distance analyses

Genetic distances among the 86 accessions shared between this study and McKhann *et al.* (2004) were computed using the simple matching coefficient (Dillmann *et al.*, 1997). Nuclear genetic distance was calculated using 189 informative SNPs based on DNA sequence data from 10 fragments evenly distributed in the genome (McKhann *et al.*, 2004). Chloroplastic and mitochondrial genetic distances were calculated using 54 pt and 10 mt informative polymorphisms found in this study. Cytoplasmic genetic distance was calculated using the combination of 64 informative pt (n = 54) and mt (n = 10) polymorphisms. The relationship between pairs of genetic distance matrices was estimated by the Pearson correlation coefficient. Significance was assessed by a simple Mantel test (10 000 permutations). All analyses were performed under the R environment.

ACKNOWLEDGEMENTS

The excellent technical assistance of S. Hiard is gratefully acknowledged. We thank C. Waszczak for the design of some primer pairs and A. Bérard for her help in the sequencing work. We are grateful to J. Cuguen and B. Dubreucq for their advice on cytoplasm polymorphism analysis and germination capacity measurements, respectively. We are also very grateful to O. Loudet and C. Camilleri who shared their great knowledge on *A. thaliana* genetic diversity, and to H. Mireau, P. Touzet, and G. Pelletier for helpful discussions and critical reading of the manuscript. We thank Dr Martin Lysak who kindly shared *A. arenosa* material. An anonymous referee's advice for improvement of the manuscript is gratefully acknowledged. This work was partially supported by the Département de Génétique et d'Amélioration des Plantes of INRA through the 'Projet Innovant B2008-1'.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Table S1. List of accessions used.
Table S2. Polymorphisms in the *MatK–trnK* fragment.
Table S3. Polymorphisms in the *ndhC–trnV* fragment.
Table S4. Polymorphisms in the *rbcL–accD* fragment.
Table S5. Polymorphisms in the *ndhF–rpl32* fragment.
Table S6. Detection and description of mitochondrial polymorphisms.
Table S7. Polymorphisms in the *ccmC* upstream region.
Table S8. Mitochondrial haplotypes of accessions.
Table S9. List of F_2 families used.
Table S10. Results of *in vitro* germinations.
Table S11. List of primers used.
Appendix S1. Detection of mitochondrial polymorphisms.

Please note: As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

REFERENCES

Alonso-Blanco, C., Bentsink, L., Hanhart, C.J., Blankestijn-de Vries, H. and Koornneef, M. (2003) Analysis of natural allelic variation at seed dormancy loci of *Arabidopsis thaliana*. *Genetics*, **164**, 711–729.

Alonso-Blanco, C., Blankestijn-de Vries, H., Hanhart, C.J. and Koornneef, M. (1999) Natural allelic variation at seed size loci in relation to other life history traits of *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA*, **96**, 4710–4717.

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.

Azhagiri, A.K. and Maliga, P. (2007) Exceptional paternal inheritance of plastids in Arabidopsis suggests that low-frequency leakage of plastids via pollen may be universal in plants. *Plant J.* **52**, 817–823.

Baud, S., Wuilleme, S., Dubreucq, B., de Almeida, A., Vuagnat, C., Lepiniec, L., Miquel, M. and Rochat, C. (2007) Function of plastidial pyruvate kinases in seeds of Arabidopsis thaliana. *Plant J.* **52**, 405–419.

Beck, J.B., Schmuths, H. and Schaal, B.A. (2008) Native range genetic variation in Arabidopsis thaliana is strongly geographically structured and reflects Pleistocene glacial dynamics. *Mol. Ecol.* **17**, 902–915.

Bellaoui, M., Martin-Canadell, A., Pelletier, G. and Budar, F. (1998) Low-copy-number molecules are produced by recombination, actively maintained and can be amplified in the mitochondrial genome of Brassicaceae: relationship to reversion of the male sterile phenotype in some cybrids. *Mol. Gen. Genet.* **257**, 177–185.

Besnard, G., Khadari, B., Baradat, P. and Berville, A. (2002) Combination of chloroplast and mitochondrial DNA polymorphisms to study cytoplasm genetic differentiation in the olive complex (*Olea europaea* L.). *Theor. Appl. Genet.* **105**, 139–144.

Bikard, D., Patel, D., Le Mette, C., Giorgi, V., Camilleri, C., Bennett, M.J. and Loudet, O. (2009) Divergent evolution of duplicate genes leads to genetic incompatibilities within *A. thaliana*. *Science*, **323**, 623–626.

Bogdanova, V.S., Galieva, E.R. and Kosterin, O.E. (2009) Genetic analysis of nuclear-cytoplasmic incompatibility in pea associated with cytoplasm of an accession of wild subspecies Pisum sativum subsp. elatius (Bieb.) Schmahl. *Theor. Appl. Genet.* **118**, 801–809.

Bomblies, K., Yant, L., Laitinen, R.A., Kim, S.T., Hollister, J.D., Warthmann, N., Fitz, J. and Weigel, D. (2010) Local-scale patterns of genetic variability, outcrossing, and spatial structure in natural stands of Arabidopsis thaliana. *PLoS Genet.* **6**, e1000890.

Bouchez, D., Vittorioso, P., Courtila, B. and Camilleri, C. (1996) Kanamycin rescue: a simple technique for the recovery of T-DNA flanking sequences. *Plant Mol. Biol.* **14**, 115–123.

Budar, F., Touzet, P. and De Paepe, R. (2003) The nucleo-mitochondrial conflict in cytoplasmic male sterilities revisited. *Genetica*, **117**, 3–16.

Burton, R.S., Ellison, C.K. and Harrison, J.S. (2006) The sorry state of F2 hybrids: consequences of rapid mitochondrial DNA evolution in allopatric populations. *Am. Nat.* **168**(Suppl. 6), S14–S24.

Clement, M., Posada, D. and Crandall, K.A. (2000) TCS: a computer program to estimate gene genealogies. *Mol. Ecol.* **9**, 1657–1659.

Corey, L.A., Matzinger, D.F. and Cockerham, C.C. (1976) Maternal and reciprocal effects on seedling characters in *Arabidopsis thaliana* (L.) Heynh. *Genetics*, **82**, 677–683.

Desplanque, B., Viard, F., Bernard, J., Forcioli, D., Saumitou-Laprade, P., Cuguen, J. and Van Dijk, H. (2000) The linkage disequilibrium between chloroplast DNA and mitochondrial DNA haplotypes in Beta vulgaris ssp. maritima (L.): the usefulness of both genomes for population genetic studies. *Mol. Ecol.* **9**, 141–154.

Dillmann, C., Bar-Hen, A., Guerin, D., Charcosset, A. and Murigneux, A. (1997) Comparison of RFLP and morphological distances between maize Zea-mays L. inbred lines. Consequences for germplasm protection purposes. *Theor. Appl. Genet.* **95**, 92–102.

Dowling, D.K., Friberg, U. and Lindell, J. (2008) Evolutionary implications of non-neutral mitochondrial genetic variation. *Trends Ecol. Evol.* **23**, 546–554.

Ellison, C.K. and Burton, R.S. (2010) Cytonuclear conflict in interpopulation hybrids: the role of RNA polymerase in mtDNA transcription and replication. *J. Evol. Biol.* **23**, 528–538.

Forner, J., Weber, B., Wietholter, C., Meyer, R.C. and Binder, S. (2005) Distant sequences determine 5′ end formation of cox3 transcripts in Arabidopsis thaliana ecotype C24. *Nucleic Acids Res.* **33**, 4673–4682.

Forner, J., Holzle, A., Jonietz, C., Thuss, S., Schwarzlander, M., Weber, B., Meyer, R.C. and Binder, S. (2008) Mitochondrial mRNA polymorphisms in different Arabidopsis accessions. *Plant Physiol.* **148** 1106–1116.

Francois, O., Blum, M.G., Jakobsson, M. and Rosenberg, N.A. (2008) Demographic history of european populations of Arabidopsis thaliana. *PLoS Genet.* **4**, e1000075.

Harrison, J.S. and Burton, R.S. (2006) Tracing hybrid incompatibilities to single amino acid substitutions. *Mol. Biol. Evol.* **23**, 559–564.

Herrmann, R.G., Maier, R.M. and Schmitz-Linneweber, C. (2003) Eukaryotic genome evolution: rearrangement and coevolution of compartmentalized genetic information. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **358**, 87–97.

Hudson, R.R. and Kaplan, N.L. (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics*, **111**, 147–164.

Jakobsson, M., Sall, T., Lind-Hallden, C. and Hallden, C. (2007) The evolutionary history of the common chloroplast genome of Arabidopsis thaliana and A. suecica. *J. Evol. Biol.* **20**, 104–121.

Kim, S., Plagnol, V., Hu, T.T., Toomajian, C., Clark, R.M., Ossowski, S., Ecker, J.R., Weigel, D. and Nordborg, M. (2007) Recombination and linkage disequilibrium in Arabidopsis thaliana. *Nat. Genet.* **39**, 1151–1155.

Knoop, V. (2004) The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. *Curr. Genet.* **46**, 123–139.

Kubo, T. and Newton, K.J. (2008) Angiosperm mitochondrial genomes and mutations. *Mitochondrion*, **8**, 5–14.

Levin, D.A. (2003) The cytoplasmic factor in plant speciation. *Syst. Bot.* **28**, 5–11.

Librado, P. and Rozas, J. (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

Loudet, O., Chaillou, S., Camilleri, C., Bouchez, D. and Daniel-Vedele, F. (2002) Bay-0 x Shahdara recombinant inbred line population: a powerful tool for the genetic dissection of complex traits in Arabidopsis. *Theor. Appl. Genet.* **104**, 1173–1184.

Loudet, O., Gaudon, V., Trubuil, A. and Daniel-Vedele, F. (2005) Quantitative trait loci controlling root growth and architecture in Arabidopsis thaliana confirmed by heterogeneous inbred family. *Theor. Appl. Genet.* **110**, 742–753.

Macherel, D., Benamar, A., Avelange- Macherel, M.H. and Tolleter, D. (2007) Function and stress tolerance of seed mitochondria. *Physiol. Plant.* **129**, 233–241.

Martinez, P., Lopez, C., Roldan, M., Sabater, B. and Martin, M. (1997) Plastid DNA of five ecotypes of Arabidopsis thaliana: sequence of ndhG gene and maternal inheritance. *Plant Sci.* **123**, 113–122.

McKhann, H.I., Camilleri, C., Berard, A., Bataillon, T., David, J.L., Reboud, X., Le Corre, V., Caloustian, C., Gut, I.G. and Brunel, D. (2004) Nested core collections maximizing genetic diversity in Arabidopsis thaliana. *Plant J.* **38**, 193–202.

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.

Mower, J.P., Touzet, P., Gummow, J.S., Delph, L.F. and Palmer, J.D. (2007) Extensive variation in synonymous substitution rates in mitochondrial genes of seed plants. *BMC Evol. Biol.* **7**, 135.

Nagata, N., Saito, C., Sakai, A., Kuroiwa, H. and Kuroiwa, T. (1999) The selective increase or decrease of organellar DNA in generative cells just after pollen mitosis one controls cytoplasmic inheritance. *Planta*, **209**, 53–65.

Nordborg, M., Hu, T.T., Ishino, Y. *et al.* (2005) The pattern of polymorphism in Arabidopsis thaliana. *PLoS Biol.* **3**, e196.

Olson, M.S. and McCauley, D.E. (2000) Linkage disequilibrium and phylogenetic congruence between chloroplast and mitochondrial haplotypes in Silene vulgaris. *Proc. Biol. Sci.* **267**, 1801–1808.

Ostrowski, M.F., David, J., Santoni, S. *et al.* (2006) Evidence for a large-scale population structure among accessions of Arabidopsis thaliana: possible causes and consequences for the distribution of linkage disequilibrium. *Mol. Ecol.* **15**, 1507–1517.

Palmer, J.D. and Herbon, L.A. (1988) Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. *J. Mol. Evol.* **28**, 87–97.

Pelletier, G. and Budar, F. (2007) The molecular biology of cytoplasmically inherited male sterility and prospects for its engineering. *Curr. Opin. Biotechnol.* **18**, 121–125.

Pico, F.X., Mendez-Vigo, B., Martinez- Zapater, J.M. and Alonso-Blanco, C. (2008) Natural genetic variation of Arabidopsis thaliana is geographically structured in the Iberian peninsula. *Genetics*, **180**, 1009–1021.

Platt, A., Horton, M., Huang, Y.S. *et al.* (2010) The scale of population structure in Arabidopsis thaliana. *PLoS Genet.* **6**, e1000843.

Provan, J., Powell, W. and Hollingsworth, P.M. (2001) Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends Ecol. Evol.* **16**, 142–147.

Rand, D.M., Haney, R.A. and Fry, A.J. (2004) Cytonuclear coevolution: the genomics of cooperation. *Trends Ecol. Evol.* **19**, 645–653.

Rawson, P.D. and Burton, R.S. (2002) Functional coadaptation between cytochrome c and cytochrome c oxidase within allopatric populations of a marine copepod. *Proc. Natl Acad. Sci. USA*, **99**, 12955–12958.

Rawson, P.D. and Burton, R.S. (2006) Molecular evolution at the cytochrome oxidase subunit 2 gene among divergent populations of the intertidal copepod, Tigriopus californicus. *J. Mol. Evol.* **62**, 753–764.

Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E. and Tabata, S. (1999) Complete structure of the chloroplast genome of Arabidopsis thaliana. *DNA Res.* **6**, 283–290.

Schmitz-Linneweber, C., Kushnir, S., Babiychuk, E., Poltnigg, P., Herrmann, R.G. and Maier, R.M. (2005) Pigment deficiency in nightshade/tobacco cybrids is caused by the failure to edit the plastid ATPase alpha-subunit mRNA. *Plant Cell*, **17**, 1815–1828.

Sharbel, T.F., Haubold, B. and Mitchell-Olds, T. (2000) Genetic isolation by distance in Arabidopsis thaliana: biogeography and postglacial colonization of Europe. *Mol. Ecol.* **9**, 2109–2118.

Takahashi, M., Matsuda, F., Margetic, N. and Lathrop, M. (2003) Automated identification of single nucleotide polymorphisms from sequencing data. *J. Bioinform. Comput. Biol.* **1**, 253–265.

Ullrich, H., Lättig, K., Brennicke, A. and Knoop, V. (1997) Mitochondrial DNA variations and nuclear RFLPs reflect different genetic similarities among 23 *Arabidopsis thaliana* ecotypes. *Plant Mol. Biol.* **33**, 37–45.

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.

Yin, P., Kang, J., He, F., Qu, L.J. and Gu, H. (2010) The origin of populations of Arabidopsis thaliana in China, based on the chloroplast DNA sequences. *BMC Plant Biol.* **10**, 22.

Accession numbers of new sequences: FN669554, FN669555, FN669556, FN669557, FN669558, FN669559, FN669560, FN669561, FN669562, FN669563, FN669564, FN669565, FN669566, FN669567, FN669568, FN669569, FN669570, FN669571, FN669572, FN669573, FN669574, FN669575, FN669576, FN669577, FN669578, FN669579, FN669580, FN669581, FN669582, FN669583, FN669584, FN669585, FN669586, FN669587, FN669588, FN669589, FN669590, FN669591, FN669592, FN669593, FN669594, FN669595, FN669596, FN669597, FN669598, FN669599, FN669600, FN669601, FN669602, FN669603, FN669604, FN669605, FN669606, FN669607.

