Mol Genet Genomics (2002) 267: 329–337
DOI 10.1007/s00438-002-0663-8

ORIGINAL PAPER

K. Loridon · P. Saumitou-Laprade

Detection of gene-anchored amplification polymorphism (GAAP) in the vicinity of plant mitochondrial genes

Received: 7 November 2001 / Accepted: 18 February 2002 / Published online: 24 April 2002

Abstract A simple, semi-automatable method was established for assessing polymorphism in plant mitochondrial genome. A set of 41 mitochondrial markers based on the published *Arabidopsis thaliana* sequence was developed in Brassicaceae using a gene-anchored amplification polymorphism (GAAP) strategy. PCR primers were selected based on conserved coding regions of mitochondrial genes and used to amplify the corresponding 5′ and/or 3′ non-coding flanking regions in order to maximise sequence variability between haplotypes. The variations in fragment size were analysed on a LiCor DNA sequencer, but the methodology is compatible with various sequencing systems using denaturing polyacrylamide gels. One advantage of the method is that GAAP products can be directly sequenced (without any cloning steps) through labelled M13 consensus sequences. Mitochondrial GAAP loci gave clear and simple patterns (one or two bands) that were easy to score and highly reproducible. Nearly all mitochondrial loci examined in *A. thaliana* were conserved within the Brassicaceae family, and half of the primers generated products when DNA from a distant species, *Beta vulgaris* (Chenopodiaceae), was used as template. The GAAP markers revealed low levels of polymorphism within species but exhibited a high level of polymorphism among genera and families. Our results showed some discrepancies with respect to the published mtDNA sequence of *A. thaliana*.

Communicated by R. G. Herrmann

K. Loridon[1] · P. Saumitou-Laprade (✉)
Laboratoire de Génétique et Evolution des Populations Végétales,
CNRS UPRESA 8016, FR CNRS 1818, Bât. SN2,
Université des Sciences et Technologies de Lille,
59655 Villeneuve d'Ascq Cedex, France
E-mail: saumitou@univ-lille1.fr
Fax: +33-3-20436979

Present address:
[1] Institut de la Recherche Agronomique INRA,
Unité de Recherche en Génétique et Amélioration
des Plantes de Dijon, Domaine d'Epoisses,
Bretenières, 21110 Genlis, France

Keywords *Arabidopsis thaliana* (Brassicaceae) · Plant mitochondrial DNA · Universal primers · Polymorphism · Biodiversity

Introduction

There is an increasing interest in the use of DNA-based techniques to assess genetic diversity in natural populations and cultivated plants and to study phylogenetic relationships among species. Surveys of cytoplasmic variability provide specific information because chloroplast (cp) and mitochondrial (mt) genomes are uniparentally (mainly maternally) inherited in plants. The analysis of complete chloroplast sequences, which are now available for several species (e.g. *Marchantia polymorpha*, *Nicotiana tabacum*, *Oryza sativa* and *Arabidopsis thaliana*), shows that most chloroplast genomes are highly conserved in terms of size, structure, and gene content and order (Ohyama et al. 1986; Palmer 1987; Hiratsuka et al. 1989). This synteny has allowed the design of universal primers derived from the most conserved genes, which can be used to amplify polymorphic intergenic sequences (Taberlet et al. 1991; Demesure et al. 1995). In contrast, plant mitochondrial genomes are characterised by a very high degree of variability in size and structure (Wolfe et al. 1987; Levings and Brown 1989; Palmer 1990): intramolecular recombination events have been shown to occur via repeated sequences located in the so-called master circle. Recombination between inverted repeats generates an isomeric form of the master circle characterised by a "flip-flop" of the sequences located between the two copies of the repeat, whereas recombination between a pair of direct repeats converts the master circle into two subgenomic circles. Such tripartite structures were reported in Brassicaceae, first in *Brassica campestris* (Palmer and Shields 1984) and then in *B. nigra* and *Raphanus sativa* (Palmer and Herbon 1986). The presence of a larger number of direct repeats leads to complex multipartite structures, as in maize mtDNA (Fauron et al. 1995). Consequently, the

order of mitochondrial genes is rarely conserved among species, effectively precluding the amplification of intergenic regions with universal primers, except in the reported case of the *rps14-cob* gene complex (Demesure et al. 1995). Despite its structural diversity, plant mtDNA appears to be conserved in sequence, with a low rate of nucleotide substitution within the coding sequences relative to chloroplast and nuclear DNAs (Wolfe et al. 1987; Laroche et al. 1997). Therefore the chloroplast genome has been preferred to the mitochondrial genome as a source of markers for phylogenetic studies (Olmstead and Palmer 1994), and, more recently, for intra-population genetic studies (Dumolin-Lapegue et al. 1997; Forcioli et al. 1998). Nevertheless, the combined use of both cytoplasmic genomes is of marked interest, as it allows one to measure the degree of linkage desequilibrium between cpDNA and mtDNA haplotypes, and thus to detect paternal leakage of genomes assumed to be maternally inherited, or evaluate the role played by recurrent mutations in organellar genomes (Dumolin-Lapegue et al. 1998; Desplanque et al. 2000). Taking advantage of the availability of the complete mtDNA sequence from *Arabidopsis thaliana* (Unseld et al. 1997), we defined a strategy for assessing polymorphism in the vicinity of mitochondrial genes, which is automatable for high throughput. We describe here a study of 41 mitochondrial loci, using a gene-anchored amplification polymorphism (GAAP) approach. We report the first results of validation of the method, its application to other species of the same or unrelated families, and polymorphisms detected with these new molecular tools in several crucifers and a distant species belonging to another family.

Materials and methods

Plant materials and DNA extraction

Several accessions of the crucifers *Arabidopsis thaliana*, *A. lyrata* ssp. *petraea*, *A. halleri* and *Brassica oleracea* (cabbage), and one individual of *Beta vulgaris* (sugarbeet; a member of the family Chenopodiaceae) were studied. Their names and origins are given in Table 1. *A. thaliana* Na-16 seeds and total DNAs from *A. lyrata* ssp. *petraea* K24 and M24 were provided by Dr. Outi Savolainen (University of Oulu, Finland). Total DNAs from sugarbeet and cabbage were obtained from Dr. Pierre Devaux (Florimond-Desprez, Cappelle-en-Pévèle, France) and Dr. Ulf Lagercrantz (University of Uppsala, Sweden), respectively. Crucifers were grown from seeds under standard greenhouse conditions (23°C, 16 h light/8 h dark cycle). Total DNA was extracted from fresh leaves (2 g) following the Dellaporta protocol, modified as described by Saumitou-Laprade et al. (1999). To remove RNA, crude DNA preparations were treated with RNase.

Principle of the method, and choice of mitochondrial genes and restriction enzymes to define GAAP markers

The gene-anchored amplification polymorphism (GAAP) strategy is presented in Fig. 1. The mitochondrial genome of *A. thaliana* (accession C24) has been completely sequenced (Unseld et al. 1997). It contains 57 identified genes, 85 ORFs, and two large repeated sequences of 6.5 kb and 4.2 kb that are active in recombination. We examined the conservation of these mitochondrial genes in various plant species by performing BLAST searches in GenBank/EMBL/DDBJ, and designed gene-specific primers based on the most conserved parts of the 5′ and 3′ coding regions. The 5′ and 3′ flanking regions of these genes were surveyed for the presence of 4-base restriction sites in the *A. thaliana* mtDNA sequence. The best candidate enzyme had one recognition site in the vicinity of genes but none between the gene border and the gene-specific primer. The distance between a gene-specific primer and its related restriction site should be smaller than 600 bp in *A. thaliana* C24, in order to obtain fragments in the optimal size range for separation on denaturing polyacrylamide gels.

Table 1 Source of plant material

Species	Accession/origin	Abbreviation
Arabidopsis thaliana	C24 (Acc. N906)	C24
	Col-4 (Acc. N933)	Col-4
	Ler-0 (Acc. NW20)	Ler-0
	Naantali (Finland)	Na-16
A. lyrata ssp. *petraea*	Karhumäki (Russia)	K24
	Mjallöm (Sweden)	M24
	Cp 1-1 (Czech Republic)	Cp1-1
A. halleri	Tch6 D1-1 (Czech Republic)	C1
	Ch9 Auby VB (France)	C3
	Courcelles 830-1 (France)	Cou
Brassica oleracea	cv. Borecole (CGN11130)	BOR
	Doubled haploid (22.2.69)	DH
Beta vulgaris	Haploid individual FD 017804-1-4 (France)	FD1-4

Adapters and primers

Unlabelled oligonucleotides were synthesised by Genset SA (Evry, France), and primers labelled either with IRD700 or IRD800 for use in the Li-Cor automated sequencer were obtained from MWG Biotech (Ebersberg, Germany). The following double-stranded adapters were ligated to total genomic DNA cut with *Hpa*II or *Mse*I (the sequence of the top strand is shown in upper case and that of the bottom strand in lower case letters): 5′-CGCAGCACGCTCAGGACTCAT-3′ (*Hpa*II adapter)/3′-gtcgtgcgagtcctgagta-5′ (consensus primer) and 5′-TACAGCACGCTCAGGACTCAT-3′ (*Mse*I adapter)/3′-gtcgtgcgagtcctgagta-5′ (consensus primer).

These double-stranded adapters differ by two enzyme-specific nucleotides. The consensus primer (5′-ATGAGTCCTGAGCGTGCTG-3′) was also used to achieve non-specific amplifications. The mitochondrial gene-specific primers used are listed in Table 2. Each primer contains a forward or reverse consensus M13 tail, followed by a gene-specific sequence. The annealing temperature of the gene-specific part of the primers varies between 53°C and 60°C, depending on the locus (Table 2). The annealing temperature of the forward and reverse consensus M13 sequence is 56°C.

Restriction digestion and adapter ligation

Genomic DNA (0.5 µg) was incubated for 3 h at 37°C in 40 µl of a solution containing 4 U of *Hpa*II or *Mse*I (New England Biolabs, Ozyme, St Quentin en Yvelines, France), 1 U of T4 DNA ligase (Roche Diagnostics, Meylan, France), 1 µM *Hpa*II or *Mse*I adapter, 1 µM consensus-primer, 10 mM TRIS-acetate pH 7.5, 10 mM magnesium acetate, 50 mM potassium acetate, 5 mM DTT and 0.2 mM ATP. Then the reaction mixture was diluted 5-fold with 10 mM TRIS-HCl pH 8.3/0.1 mM EDTA pH 8.0 and stored at –20°C. Restriction and adapter ligation were carried out simultaneously to increase the efficiency of the reaction.

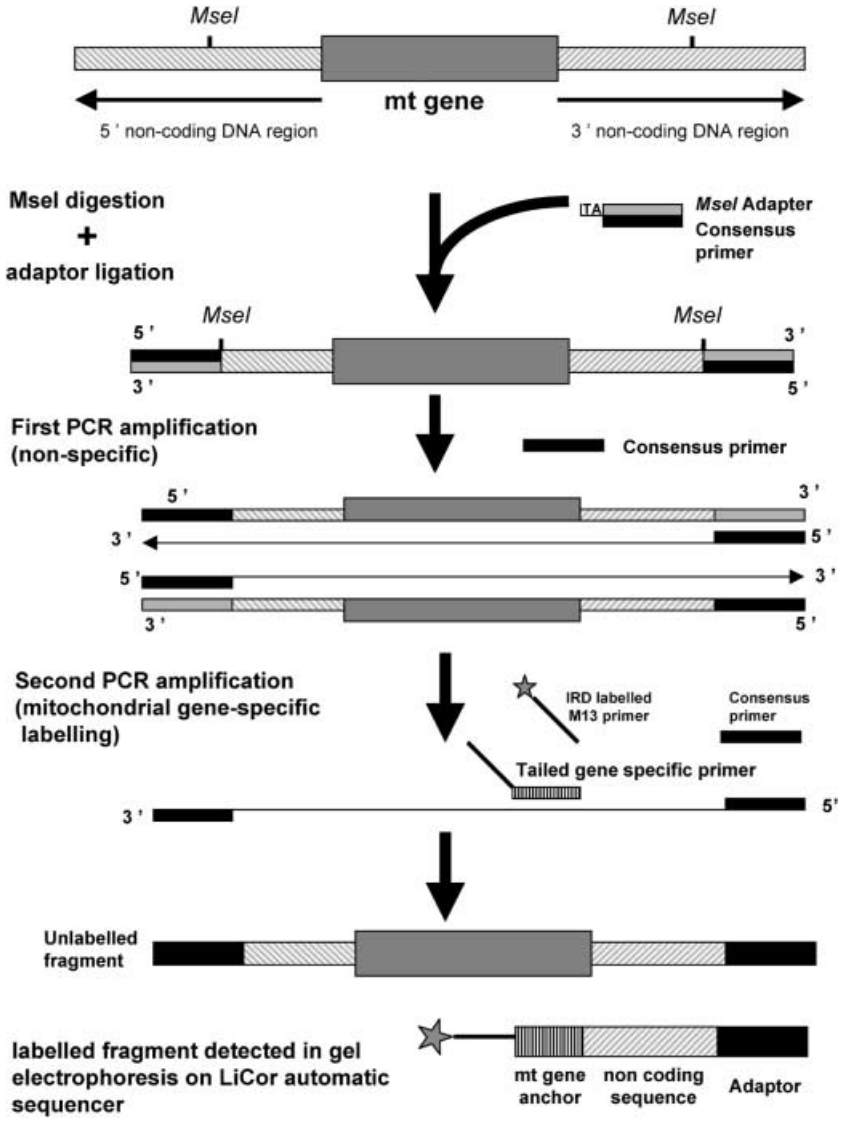

Fig. 1 Schematic representation of the gene-anchored amplification polymorphism (GAAP) method. Total genomic DNA is digested with a restriction enzyme that recognises a 4-base sequence (*Mse*I or *HpaII*) and double-stranded adapters are ligated to the ends of the restriction fragments to create target sites for consensus primer annealing and subsequent non-specific preamplification. A second PCR is then carried out with the same consensus primer and a long gene-specific primer anchored in the 5′ or 3′ coding region of a given mitochondrial gene (e.g. 5′ region) and containing a floating consensus M13 tail, allowing the incorporation of a complementary labelled M13 primer during the PCR cycles. In this procedure, the same labelled reverse or forward consensus M13 primers are used for all mitochondrial loci

Non-specific preamplification

The reaction mixtures (20 µl) contained 5 µl of the restriction-ligation mix, 10 mM TRIS-HCl pH 8.3, 1.5 mM $MgCl_2$, 50 mM KCl, 0.8 µM consensus primer, 0.25 mM of all four dNTPs and 0.5 U of AmpliTaq DNA polymerase (PE Biosystems, Courtaboeuf, France). After an initial denaturation step at 94°C for 5 min, PCR was carried out for 20 cycles of 30 s at 94°C, 1 min at 56°C and 1 min at 72°C, followed by a final incubation for 7 min at 72°C in a GeneAmp PCR System 9700 (PE Biosystems, Courtaboeuf, France). A 1-µl aliquot of this mixture was loaded onto a 1% agarose gel to check the amount and size of the PCR products; in most cases, a smear of low-molecular-weight material (200–500 bp) was observed. The reaction mixture was diluted 5-fold with 10 mM TRIS-HCl pH 8.3, 0.1 mM EDTA pH 8.0 and stored at –20°C.

Gene-specific PCR amplification

Each PCR was performed with a 5-µl sample of the diluted non-specific preamplification mixture and 0.5 U of AmpliTaq DNA polymerase (PE Biosystems, Courtaboeuf, France) in a total volume of 20 µl, containing 10 mM Tris-HCl pH 8.3, 1.5 mM $MgCl_2$, 50 mM KCl, 0.2 µM consensus primer, 0.1 µM gene-specific primer, 0.1 µM IRD 700- or IRD 800-labelled reverse or forward consensus M13 primer and 0.25 mM of all four dNTPs. The cycle parameters used for the amplification reaction were as follows: 1 cycle of 5 min at 94°C, 5 cycles of 30 s at 94°C, 45 s at 53–60°C (depending on the locus) and 1 min at 72°C, then 30 cycles of 30 s at 94°C, 45 s at 56°C, and 1 min at 72°C, followed by a final incubation for 7 min at 72°C.

Gel analysis

Amplification products were mixed with half a volume (10 µl) of formamide dye (92% formamide, 10 mM EDTA, 2 mg/ml Congo Red). The resulting mixtures were then denatured 2 min at 94°C, quickly cooled on ice, loaded on 7% denaturing polyacrylamide gels and analysed on a Li-Cor automated DNA sequencer 4200 (Li-Cor-ScienceTec, Les Ulis, France). PCR fragment sizes were assessed with Li-Cor's Base ImagIR software (version 4.1) by comparison with an appropriate labelled molecular-weight marker (Li-Cor-ScienceTec).

Sequencing of PCR products

For each sample, two to four identical gene-specific PCRs were performed as previously described (except that no labelled consensus M13 tail was added and the Taq Gold polymerase was used)

Table 2 Mitochondrial GAAP loci, primer sequences and annealing temperatures

Locus	Enzyme	Gene-specific primer (5′→3′)	Annealing temperatures	
			Cruciferae	*Beta vulgaris*[b]
3′F-ATp1	*Hpa*II	**GGATAACAATTTCACACAGGT**TCTGTGATCGAATGCCACT	53°C	53°C
3′F-CCb203	*Hpa*II	**GGATAACAATTTCACACAGG**GTGGCGCGAAGTACTCTTGT	56°C	53°C
3′F-CCb206	*Hpa*II	**CACGACGTTGTAAAACGACT**TTTTCCAATTTCGGTCTCG	56°C	53°C
3′F-CCb452-e2	*Hpa*II	**GGATAACAATTTCACACAGG**CGAGAAAGTTCCGTGGAGTT	56°C	nd
3′F-cox3	*Hpa*II	**CACGACGTTGTAAAACGACTG**CATGGTACTGGCATTTTG	56°C	53°C
3′F-nad1-e1	*Hpa*II	**GGATAACAATTTCACACAGGTG**CCATATCTTCGCTAGGTG	56°C	56°C
3′F-nad1-e3	*Hpa*II	**GGATAACAATTTCACACAGGTC**CGTTTGATCTCCCAGAAG	56°C	56°C
3′F-nad5-e2	*Hpa*II	**GGATAACAATTTCACACAGG**CCATTCCTATGGCCATTCCT	56°C	nd
3′F-rps4	*Hpa*II	**GGATAACAATTTCACACAGGT**TCCTCTTCGGAGCAGAAAC	56°C	56°C
3′F-rps7	*Hpa*II	**CACGACGTTGTAAAACGACC**GCATTTCAGATGGTGGTAA	56°C	nd
3′F-rrn26	*Hpa*II	**GGATAACAATTTCACACAGG**GCGGTCAGTCTCGATACCTC	56°C	nd
5′F-CCb256	*Hpa*II	**CACGACGTTGTAAAACGACC**GAGAATTTCCACCTTGTTGA	56°C	53°C
5′F-cox2-e1	*Hpa*II	**CACGACGTTGTAAAACGACATC**CTAATTGCCATGGTTCC	56°C	56°C
5′F-cox3	*Hpa*II	**GGATAACAATTTCACACAGG**CTCCGAGTGAACCCGAAATA	56°C	nd
5′F-nad5-e1	*Hpa*II	**CACGACGTTGTAAAACGACG**AATGAAACGCACGTAGTGG	56°C	nd
5′F-nad5-e4	*Hpa*II	**CACGACGTTGTAAAACGACG**AAAGGCTCGTTGGAATTGA	56°C	53°C
5′F-rpl5	*Hpa*II	**GGATAACAATTTCACACAGG**CACGATCCAGGAACTTCCATA	56°C	nd
5′F-rps4	*Hpa*II	**CACGACGTTGTAAAACGACG**CGGCGTTGTATTATTGTCA	53°C	nd
5′F-rps7	*Hpa*II	**GGATAACAATTTCACACAGGTA**CATCGCGTTCAGTTCGAG	53°C	nd
5′F-rps12/nad3	*Hpa*II	**CACGACGTTGTAAAACGACTT**GGAAGCAAATGGAAAAGG	53°C	nd
5′F-rrn5/rrn18	*Hpa*II	**CACGACGTTGTAAAACGACAG**CTAGCGTTCCTTCTGAGC	56°C	56°C
3′F-ATp9	*Mse*I	**CACGACGTTGTAAAACGACGG**AGCTGCTATCGGTATTGG	60°C	60°C
3′F-CCb382	*Mse*I	**GGATAACAATTTCACACAGGA**GCGAGAAGCCACTCGACTA	60°C	60°C
3′F-cob	*Mse*I	**CACGACGTTGTAAAACGACGT**TGGATCGGATGTCAACCT	56°C	nd
3′F-cox1	*Mse*I	**GGATAACAATTTCACACAGG**GTCCTTGGGCTCTTGAACTG	60°C	nd
3′F-nad2-e2	*Mse*I	**GGATAACAATTTCACACAGGTC**TGAATTTTCCACGGAAGC	56°C	56°C
3′F-nad4-e4	*Mse*I	**CACGACGTTGTAAAACGACGG**TGTTCACCCCAAAGTGTT	56°C	nd
3′F-nad5-e5	*Mse*I	**GGATAACAATTTCACACAGG**CCTTTTCTCGTATGTGGGACTC	60°C	nd
3′F-nad6	*Mse*I	**GGATAACAATTTCACACAGGT**TCCGACGAAATGCTATTGA	56°C	56°C
3′F-nad7-e5	*Mse*I	**CACGACGTTGTAAAACGACA**AACATCACATGCCAGCAGA	56°C	56°C
3′F-nad9	*Mse*I	**GGATAACAATTTCACACAGGA**CCCATTGAGATGACCCAAG	60°C	60°C
3′F-rpl16	*Mse*I	**GGATAACAATTTCACACAGGG**CGGCTCATAAACCATGTTC	56°C	53°C
3′F-rrn5/rrn18	*Mse*I	**GGATAACAATTTCACACAGGA**GATTGTTCGGGAGACATGG	60°C	60°C
5′F-ATp1	*Mse*I	**CACGACGTTGTAAAACGACA**CTCGACCGATCTCATCCAC	60°C	60°C
5′F-CCb203	*Mse*I	**CACGACGTTGTAAAACGACC**ATTACGAACGACACGCTTC	60°C	nd
5′F-cox1	*Mse*I	**CACGACGTTGTAAAACGACTG**TGGTTTGTGGAGAACAGC	56°C	nd
5′F-nad2-e3	*Mse*I	**CACGACGTTGTAAAACGACTC**GAAGTGGGTAGCTCCAGT	56°C	nd
5′F-nad6	*Mse*I	**CACGACGTTGTAAAACGACG**GGCTCGACAAAACAGAAAG	56°C	56°C
5′F-nad7-e1	*Mse*I	**GGATAACAATTTCACACAGGA**GCAGCAGGATGTTGAGGTC	60°C	53°C
5′F-nad9	*Mse*I	**CACGACGTTGTAAAACGACT**TCCCATGTTCCGATCTTTC	56°C	nd
5′F-rpl2-e1	*Mse*I	**GGATAACAATTTCACACAGG**AAAACCGTAATACGCCCTGA	56°C	nd

[a]M13 consensus sequences (reverse for the 3′ regions and forward for the 5′ regions of the genes) are shown in *bold*
[b]nd, not determined (no fragment detected)

and pooled to obtain enough template for direct sequencing. The pooled reaction mixtures were purified with the QIAquick PCR Purification Kit (Qiagen S.A., Courtaboeuf, France), according to the supplier's instructions. Half of the eluted DNA (15 µl) was used as a template for cycle sequencing with ABI Prism (PE Biosystems, Courtaboeuf, France), following the supplier's protocol and using an IRD-labelled reverse or forward consensus M13 primer. Sequencing fragments were separated on a 6% denaturing polyacrylamide gel (41 cm in length) and analysed on a Li-Cor automated DNA sequencer 4200.

Results and discussion

GAAP markers

We chose 41 mitochondrial gene-specific loci for the study (21 with *Hpa*II, 20 with *Mse*I), which are listed in Table 2. The mitochondrial GAAP markers defined gave simple and clearly legible patterns consisting of one or two amplified fragments, as illustrated in Fig. 2. All mitochondrial gene-specific loci defined with *Hpa*II (21 loci) and *Mse*I (20 loci) were tested in 13 accessions belonging to the species *A. thaliana*, *A. lyrata* ssp. *petraea*, *A. halleri*, *Brassica oleracea* (these four species belong to the Brassicaceae family), and *Beta vulgaris* (Chenopodiaceae). Gene-specific amplifications were achieved two to four times and produced highly reproducible fragments. Of the 533 locus/accession combinations, 487 gave a single fragment and 19 produced two fragments, while 27 (20 of which involved *B. vulgaris* and seven *B. oleracea*) did not give any product. PCR product sizes are listed in Table 3.

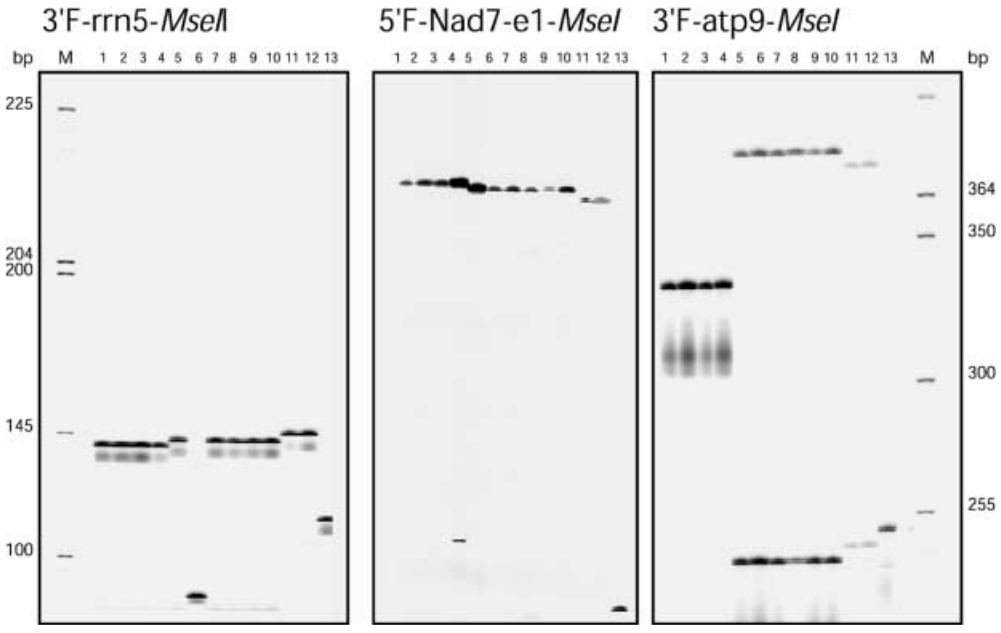

Fig. 2 Mitochondrial variations detected at three GAAP loci in different accessions of the four species of Brassicaceae and *Beta vulgaris* (Chenopodiaceae). The names and origins of accessions are listed in Table 1. Lanes 1–4, *A. thaliana* C24, Col-4, Ler-0 and Na-16; lanes 5–7, *A. lyrata* ssp. *petraea* K24, M24 and Cp 1-1; lanes 8–10, *A. halleri* C1, C3 and Cou; lanes 10 and 11, *Brassica oleracea* BOR and DH; lane 13, *Beta vulgaris* FDH1–4. M, molecular weight marker 50–700 bp (LiCor)

Validation of GAAP markers: size and sequence analysis of GAAP products obtained from *A. thaliana* C24

The size of PCR fragments obtained in *A. thaliana* were estimated and compared with the expected values (Table 3) calculated from the published sequence of mtDNA from *A. thaliana* C24 (Unseld et al. 1997). The sizes were found to be in good agreement with the published values for 35 of the 41 mitochondrial loci examined. These observations were confirmed at the nucleotide level for 12 loci, by direct sequencing of PCR products obtained from *A. thaliana* C24, Col-4 and Ler-0. Perfect identity in size and sequence was found at all of these loci (data not shown). When two PCR products were expected, two fragments were observed. At 5′F-nad6, the presence of two GAAP products is due to the duplication of the 5′ coding region of *nad6* in *A. thaliana* mtDNA. The long PCR fragment (235 bp) corresponds to the 5′ flanking region of *nad6*. The small fragment (139 bp) is anchored in a duplication of the 5′ coding region of the genuine *nad6* gene present as an inverted repeat in *orf313*. The size of this second PCR fragment is in perfect agreement with the position of the nearest *Mse*I site.

Some mitochondrial loci (6/41) produced unexpected results. Their corresponding products in C24, Col-4 and Ler-0 were sequenced two to four times. In five cases [5′F-nad5-e1 (*HpaII*), 5′F-cox3 (*HpaII*), 3′F-nad1-e3 (*HpaII*), 5′F-nad2-e3 (*Mse*I) and 5′F-ccb382 (*Mse*I)], small indels (1–2 bp) were noticed when the published C24 sequence was compared with the sequences we obtained for this ecotype, suggesting that a few errors had crept into the original mtDNA sequence. At 5′F-ccb382 (*Mse*I), the PCR product was much longer than expected (583 bp vs. 457 bp). The sequence data show that the sequence at the expected *Mse*I restriction site, ...GTTAAG..., is actually ...GTAAG..., and the endonuclease recognised the following TTAA tetramer, located 126 bp upstream.

Amplification products obtained from all *A. thaliana* accessions at 3′F-nad5-e2 were smaller than expected (206 bp vs. 247 bp). All these PCR products were sequenced (including that from C24) and appeared to be identical. The 3′ coding region of nad5-e2 and its flanking region were retrieved in the first 136 bp but the homology with the next part of the published C24 mtDNA sequence was completely lost. The non-homologous part of the sequence was found to match the reported sequence of chromosome 2 in *A. thaliana* Col-4 (Lin et al. 1999). As a matter of fact, a large portion (75%) of the mitochondrial genome has been retrieved in the centromere of chromosome 2 (to differentiate this insertion in nuclear DNA from the genuine mtDNA, we will call it cenX2mtDNA). The homology between the published Col-4 cenX2mtDNA and C24 mtDNA sequences is very high (99%), suggesting that this gene transfer from organelle to nucleus occurred recently and that sequence polymorphisms are due to divergence between the two *A. thaliana* accessions. The sequences obtained at 3′F-nad5-e2 for C24, Col-4, Ler-0 are identical and match cenX2mtDNA rather than mtDNA; they display a 50 bp insertion absent in the published mtDNA. This 50 bp insertion contains a *Hpa*II restriction site responsible for the 206 bp GAAP fragment instead of the 247 bp expected for the mitochondrial DNA. In order to determine why only a single GAAP product is generated and why this product corresponds to the nuclear copy, we checked whether the two different forms of the 3′F-nad5-e2 region (a nuclear single-copy form and a mitochondrial multicopy form) occur in *A. thaliana*. We performed a Southern analysis, using as a probe the 3′F-nad5-e2 region labelled with the PCR-DIG Labelling Mix (Roche). Total DNA from

Table 3 Size of PCR amplification products for 41 mitochondrial GAAP loci in five different species

Locus	Enzyme	Size (bp)[a]	Species[b]												
			At C24	At Col-4	At Ler-0	At Na-16	Ap K24	Ap M24	Ap Cp 1-1	Ah C1	Ah C3	Ah Cou	Bo BOR	Bo DH	Bv FD1-4
3′F-atp1	*Hpa*II	272	272	272	272	272	272	272	272	272	272	272	272	287
3′F-ccb203	*Hpa*II	161	161	161	161	161	161	161	161	161	161	161	161	377
3′F-ccb206	*Hpa*II	132	132	132	132	132	132	132	132	132	132	132	132	94
3′F-ccb452-e2	*Hpa*II	354	354	354	354	354	354	354	354	354	354	350	350	–
3′F-cox3	*Hpa*II	165	165	165	165	165	165	165	165	165	165	166	166	511
3′F-nad1-e1	*Hpa*II	293	293	293	293	293	293	293	293	293	293	293	293	264
3′F-nad1-e3	*Hpa*II	148	146	146	146	146	146	146	146	146	146	146	146	146
3′F-nad5-e2	*Hpa*II	247	206	206	206	206	206	206	206	206	206	206	206	–
3′F-rps4	*Hpa*II	117	117	117	117	117	117	117	117	117	117	315	315	171
3′F-rps7	*Hpa*II	159	159	159	159	159	159	159	159	159	159	159	159	–
3′F-rrn26	*Hpa*II	566	567	567	567	567	567	567	567	567	567	567	–	–	–
5′F-ccb256	*Hpa*II	509	509	509	509	509	551	533	557	551	551	551	547	547	215
5′F-cox2-e1	*Hpa*II	205	205	205	205	205	205	205	205	205	205	205	212	212	172
5′F-cox3	*Hpa*II	161	162	162	162	162	162	162	162	162	162	162	618	618	–
5′F-nad5-e1	*Hpa*II	238	237	237	237	237	232, 285	232	232	232	232, 285	232, 285	232	232	–
5′F-nad5-e4	*Hpa*II	237	237	237	237	237	237	237	237	237	237	237	237	237	237
5′F-rpl5	*Hpa*II	486	486	486	486	486	486	486	486	486	486	486	442	442	–
5′F-rps4	*Hpa*II	344	344	344	344	344	344	344	344	344	344	344	344	344	–
5′F-rps7	*Hpa*II	394	394	394	394	394	394	394	394	394	394	394	394	394	–
5′F-rps12/nad3	*Hpa*II	316	316	316	316	316	367	316	316	316	316	316	–	–	–
5′F-rrn5/rrn18	*Hpa*II	119	118	119	119	119	119	119	119	119	119	119	119	119	119
3′F-atp9	*Mse*I	334	334	334	334	334	241, 380	241, 380	241, 380	241, 380	241, 380	241, 380	245, 375	245, 375	249
3′F-ccb382	*Mse*I	457	583	583	583	583	552	552	552	552	552	583	–	388	224
3′F-cob	*Mse*I	366	366	366	366	366	366	366	366	366	366	366	525	525	–
3′F-cox1	*Mse*I	229	229	229	229	229	229	229	229	229	229	229	–	–	–
3′F-nad2-e2	*Mse*I	143	143	143	143	143	143	143	143	143	143	143	143, 235	143	143
3′F-nad4-e4	*Mse*I	156	156	156	156	156	156	156	156	156	156	156	156	156	–
3′F-nad5-e5	*Mse*I	342	342	342	342	342	342	342	342	342	342	342	446	446	–
3′F-nad6	*Mse*I	252	252	252	252	252	680	680	680	680	680	680	780	780	183
3′F-nad7-e5	*Mse*I	197	197	197	197	197	188	188	188	188	188	188	195	195	227
3′F-nad9	*Mse*I	282	282	282	282	282	270	270	270	270	270	270	270	270	241
3′F-rpl16	*Mse*I	99	99	99	99	99	99	99	99	99	99	99	99	99	235
3′F-rrn5	*Mse*I	143	143	143	143	143	145	86	145	145	145	145	147	147	116
5′F-atp1	*Mse*I	150	150	150	150	150	154	154	154	154	154	154	340	340	158
5′F-ccb203	*Mse*I	188	188	188	188	188	188	188	188	188	188	188	264	264	–
5′F-cox1	*Mse*I	552	552	552	552	552	552	552	552	552	552	552	250	250	–
5′F-nad2-e3	*Mse*I	279	278	>1000	>1000	>1000	278	278	278	278	278	278	274	274	–
5′F-nad6	*Mse*I	139, 235	139, 235	139, 235	139, 235	139, 235	235	235	235	235	235	235	235, 387	235, 387	160, 545
5′F-nad7-e1	*Mse*I	372	372	372	372	372	370	370	370	370	370	370	367	367	227
5′F-nad9	*Mse*I	258	258	258	258	258	258	258	258	258	258	258	258	258	–
5′F-rpl2-e1	*Mse*I	305	305	301	301	301	301	301	301	301	301	301	298	298	–

[a]Size expected based on the published sequence for *A. thaliana* C24
[b]At, *Arabidopsis thaliana*; Ap, *A. lyrata* ssp. *petraea*; Bo, *Brassica oleracea*; Bv, *Beta vulgaris*. For further details, see Table 1

C24 and Col-4 individuals was digested with *Hha*I, electrophoresed on a 1% agarose gel and transferred onto a nylon membrane (Biodyne A from Pall). Under these conditions, the 3′F-nad5-e2 probe is expected to hybridise with a 537 bp nuclear DNA fragment and a 487 bp mtDNA. After hybridisation only one signal was detected at 537 bp for both individuals (data not shown). This result demonstrates that both mitochondrial DNA and nuclear DNA contain the 50 bp insertion. Therefore the unexpected sizes of the GAAP fragments observed at the six loci are due to errors in the published mtDNA sequence.

Application of mitochondrial GAAP markers across species

All the mitochondrial loci examined in *A. thaliana* C24 gave the expected PCR products when used on the four *A. thaliana* accessions included in the experiment. We also asked whether these GAAP markers could be used in other species belonging to the same genus, or even in other genera and families. The PCR fragments obtained from *A. thaliana*, *A. lyrata* ssp. *petraea*, *A. halleri*, *B. oleracea* and *B. vulgaris* are listed in Table 3. All mitochondrial loci (41/41) were successfully amplified from *A. lyrata* ssp. *petraea* and *A. halleri*. Most amplifications gave a single fragment. The 3′F-atp9 (*Mse*I) locus is an exception: two bands of 241 bp and 380 bp were systematically observed in all accessions of *A. lyrata* ssp. *petraea* and *A. halleri*. A fragment of the 3′ coding region of the *atp9* gene has probably been duplicated in the mitochondrial genomes of both species, but not in *A. thaliana*, where only one product of the expected size was found. In *B. oleracea*, 38/41 mitochondrial loci yielded PCR fragments (one in most cases). Interestingly, two bands were found at 5′F-nad6 (*Mse*I), as in *A. thaliana*. All the crucifers studied here share a 235 bp fragment corresponding to a PCR product anchored in the genuine *nad6* gene. As explained above, the second band observed in *A. thaliana* is due to a chimeric copy of *nad6*. Similarly, duplication of part of this gene may have occurred in *B. oleracea* mtDNA, leading to the 387 bp product, whereas no extra band was detected in *A. lyrata* ssp. *petraea* or *A. halleri*. In *B. vulgaris* (Chenopodiaceae), amplification products were found for 21/41 GAAP markers, one of which [5′F-nad6 (*Mse*I)] gave two bands. Our failure to detect PCR fragments for the other loci could be explained if (1) amplified fragments were larger (>1 kb) than the optimal size range detectable on polyacrylamide gels or if (2) no products were amplified due to mismatches between sugarbeet mitochondrial gene sequences and the *A. thaliana*-specific primers (null alleles). Thus, mitochondrial GAAP markers appear to be reliably applicable within the same genus (*A. thaliana* vs. *A. lyrata* ssp. *petraea* and *A. halleri*) and family (*A. thaliana* vs. *B. oleracea*) and conserved to a certain extent between families (Brassicaceae vs. Chenopodiaceae).

Levels of polymorphism detected by GAAP markers

The results presented in Table 3 were analysed to assess the degree of polymorphism detectable with mitochondrial GAAP markers at various taxonomic levels (family, genus, species). First, the sizes of PCR products were compared in Chenopodiaceae and Brassicaceae, revealing a high level of polymorphism: 18/21 loci yield fragments of different size. Only three mitochondrial loci gave rise to the same PCR fragment in all species studied: 5′F-nad5-e4 (*Hpa*II), 3′F-nad1-e3 (*Hpa*II), and 5′F-rrn5 (*Hpa*II). However, in the case of 3′F-nad2-e2 (*Mse*I), an extra band was observed in *B. oleracea*. Secondly, we focused on the 38 mitochondrial loci that yielded PCR products in all crucifers; the three loci 3′F-rrn26 (*Hpa*II), 5′F-rps12/nad3 (*Hpa*II) and 3′F-cox1 (*Mse*I) were not taken into account because no fragments were obtained from *B. oleracea*. Of these 38 loci, 14 were monomorphic in all species and 10 revealed polymorphism between *B. oleracea* and all other crucifers. Thirdly, results were compared in the three species belonging to the genus *Arabidopsis*. Of 41 mitochondrial loci, 27 displayed a fragment of the same size in all species, while 7 were polymorphic between *A. thaliana* on the one hand and *A. lyrata* ssp. *petraea*/*A. halleri* on the other. GAAP products were found to be very similar in *A. lyrata* ssp. *petraea* and *A. halleri*. For example, the same allele was found in both species for 5′F-nad6 (*Mse*I), whereas two PCR products were observed in *A. thaliana*. Conversely, 3′F-atp9(*Mse*I) yields two fragments of identical size in *A. lyrata* ssp. *petraea* and *A. halleri*, but only one in *A. thaliana*. When no intraspecific polymorphism was detected, it was impossible to distinguish *A. halleri* from *A. lyrata* ssp. *petraea* on the basis of the size of GAAP products. The genetic similarity between these species is in agreement with phylogenetic data based on ITS sequence variation in Brassicaceae (Koch et al. 1999), which shows that *A. halleri* and *A. lyrata* ssp. *petraea* are closer than any of the others to *A. thaliana*. Fourthly, very few cases of intraspecific polymorphism were observed: two in *A. thaliana* [5′F-rpl2-e1 (*Mse*I) and 5′F-nad2-e3 (*Mse*I)], four in *A. lyrata* ssp. *petraea* [5′F-nad5-e1 (*Hpa*II), 5′F-ccb256 (*Hpa*II), 5′F-rps12/nad3 (*Hpa*II), and 3′F-rrn5 (*Mse*I)], two in *A. halleri* [5′F-nad5-e1 (*Hpa*II) and 3′F-ccb382 (*Mse*I)] and one in *B. oleracea* [3′F-nad2-e2 (*Mse*I)]. In *A. thaliana*, the intraspecific polymorphism noticed at 5′F-rpl2-e1 results from a 4-bp insertion (TTAC) in C24 as compared to Col-4, Ler-0 (and very probably Na-16). At 5′F-nad2-e3 (*Mse*I), the intraspecific polymorphism observed, with a small fragment (278 bp) for C24 and a long one (over 1 kb) for the other three ecotypes can be explained by a single A/C mutation at the third base of the *Mse*I site found in C24 and hence the use of next site located 743 bp upstream for the other three ecotypes.

The level of polymorphism detected with the 41 mitochondrial GAAP markers is high between families, intermediate between genera, low between species

belonging to the same genus and very low within the same species. This is in agreement with the extremely low levels of mitochondrial DNA polymorphism that have been reported within eight *Brassica* species based on RFLP markers (Palmer 1988) and within intron sequences in *Quercus robur* (Demesure et al. 1995).

Conclusions

We have developed a set of 41 mitochondrial markers in Brassicaceae using a gene-anchored amplification polymorphism (GAAP) strategy. The first steps of the method share some similarities with the AFLP procedure (Vos et al. 1995), except that genomic DNA is cleaved with only one endonuclease (a frequent cutter), and the consensus primer designed for preamplification does not overlap the restriction site. In AFLP, the number of genomic fragments obtained after preamplification is reduced by adding selective nucleotides at the 3′ ends of the primers used in the subsequent PCR. The GAAP approach is different in that the second amplification is specific for a given mitochondrial gene, which implies the availability of sequence information for at least in one reference species. A similar strategy for revealing sequence-specific amplification polymorphism (SSAP) has been designed which is based on the use of an anchor in the LTR part of retrotransposons in barley and pea (Waugh et al. 1997; Ellis et al. 1998). The GAAP method has several advantages. The template consists of total genomic DNA, avoiding the laborious and time-consuming preparation of mitochondrial DNA. The consensus primer designed for preamplification is compatible with any endonuclease, as it does not overlap the restriction site sequence. The gene-specific primers are not labelled (labelling is carried by a consensus M13 primer homologous to their 5′ M13 tail), so they can be applied in different labs using various sequencing systems. The separation of PCR products on denaturing polyacrylamide gels allows visualisation of small to large allelic indels (1 bp to a few hundred bp). One main interest of the method is that GAAP products can be directly sequenced (without any cloning step) using the labelled M13 consensus sequences. Mitochondrial GAAP loci give clear and simple patterns (one or two bands) that are easy to score and highly reproducible. Nearly all mitochondrial loci examined in *A. thaliana* were conserved in Brassicaceae and half of them could be applied in a distant species, *B. vulgaris* (Chenopodiaceae). These markers tend to be very polymorphic between families and genera. On the other hand, the level of polymorphism is rather low within and among the three *Arabidopsis* species. These tendencies have to be modulated depending on the markers. It would be interesting to assess the intraspecific level of polymorphism of GAAP markers in the *Arabidopsis* genus on a larger set of accessions, in order to examine to what extent they could be used for genotyping various accessions/cultivars in a given species. Besides, sequencing of GAAP products would certainly give some clues as to how mitochondrial genomes evolved in Brassicaceae.

Acknowledgements The authors would like to thank Monika Mörchen for technical expertise in molecular work, Pascal Touzet for helpful comments, and Robert Dron for taking care of the plant material. We are grateful to Dr. Outi Savolainen and members of the Motors consortium for providing biological material. This work, funded by the EU Biotechnology Program (under Contract MOTORs Bio CT98-0299), was carried out in compliance with the current laws governing genetic experimentation in France.

References

Demesure B, Sodzi N, Petit RJ (1995) A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. Mol Ecol 4:129–131

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

Dumolin-Lapegue S, Pemonge MH, Petit RJ (1997) An enlarged set of consensus primers for the study of organelle DNA in plants. Mol Ecol 6:393–397

Dumolin-Lapegue S, Pemonge MH, Petit RJ (1998) Association between chloroplast and mitochondrial lineages in oaks. Mol Biol Evol 15:1321–1331

Ellis TH, Poyser SJ, Knox MR, Vershinin AV, Ambrose MJ (1998) Polymorphism of insertion sites of *Ty1-copia* class retrotransposons and its use for linkage and diversity analysis in pea. Mol Gen Genet 260:9–19

Fauron C, Casper M, Gao Y, Moore B (1995) The maize mitochondrial genome: dynamic, yet functional. Trends Genet 11:228–235

Forcioli D, Saumitou-Laprade P, Vernet P, Cuguen J (1998) Distribution of chloroplast DNA diversity within and among populations gynodioecious species, *Beta vulgaris* ssp. *maritima* (Chenopodiaceae). Mol Ecol 7:1193–1204

Hiratsuka J, Shimada H, Whittier R, Ishibashi T, Sakamoto M, Mori M, Kondo C, Honji Y, Sun C-R, Meng B-Y, Li Y-Q, Kanno A, Nishizawa Y, Hirai A, Shinozaki K, Sugiura M (1989) The complete sequence of the rice (*Oriza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. Mol Gen Genet 217:185–194

Koch M, Bishop J, Mitchell-Olds T (1999) Molecular systematics and evolution of *Arabidopsis* and *Arabis*. Plant Biol 1:529–537

Laroche J, Li P, Maggia L, Bousquet J (1997) Molecular evolution of angiosperm mitochondrial introns and exons. Proc Natl Acad Sci USA 94:5722–5727

Levings CS, Brown GG (1989) Molecular biology of plant mitochondria. Cell 56:171–179

Lin X, et al (1999) Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. Nature 402:731–732

Ohyama K, Fukuzawa H, Kohchi T, Shirai H, Sano T, Sano S, Umesono K, Shiki Y, Takeuchi M, Chang Z, Aota S-I, Inokuchi H, Ozeki H (1986) Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. Nature 322:572–574

Olmstead RG, Palmer JD (1994) Chloroplast DNA systematics: a review of methods and data analysis. Am J Bot 8:1205–1224

Palmer JD (1987) Chloroplast DNA evolution and biosystematics uses of chloroplast DNA variation. Amer Nat 130:S7–S29

Palmer JD (1988) Intraspecific variation and multicircularity in *Brassica* mitochondrial DNAs. Genetics 118:341–351

Palmer JD (1990) Contrasting modes and tempos of genome evolution in land plant organelles. Trends Genet 6:115–120

Palmer JD, Herbon LA (1986) Tricircular mitochondrial genomes of *Brassica* and *Raphanus*: reversal of repeat configurations by inversion. Nucleic Acids Res 14:9755–9765

Palmer JD, Shields CR (1984) Tripartite structure of the *Brassica campestris* mitochondrial genome. Nature 307:437–440

Saumitou-Laprade P, Piquot Y, Raspé O, Bernard J, Vrieling K (1999) Plant DNA fingerprinting and profiling. In: Epplen JT, Lubjuhn T (eds) DNA profiling and DNA fingerprinting. Birkhäuser, Basel, pp 17–38

Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol Biol 17:1105–1109

Unseld M, Marienfeld JR, Brandt P, Brennicke A (1997) The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. Nature Genetics 15:57–61

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414

Waugh R, McLean K, Flavell AJ, Pearce SR, Kumar A, Thomas BB, Powell W (1997) Genetic distribution of *Bare-1*-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). Mol Gen Genet 253: 687–694

Wolfe KH, Li WH, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci USA 88:9054–9058