Physiologia Plantarum
An International Journal for Plant Biology

Physiologia Plantarum 137: 592–608. 2009

Carrot alternative oxidase gene *AOX2a* demonstrates allelic and genotypic polymorphisms in intron 3

Hélia Guerra Cardoso[a], Maria Doroteia Campos[a], Ana Rita Costa[a], Maria Catarina Campos[a], Thomas Nothnagel[b] and Birgit Arnholdt-Schmitt[a,*]

[a]EU Marie Curie Chair, ICAAM, University of Évora, Apartado 94, 7002-554 Évora, Portugal
[b]Julius Kühn-Institut, Federal Research Centre for Cultivated Plants, Institute for Breeding Research on Horticultural and Fruit Crops, Erwin-Baur-Str. 27, D-06484 Quedlinburg, Germany

Correspondence
*Corresponding author,
e-mail: eu_chair@uevora.pt

Received 9 June 2009; revised 24 September 2009

doi:10.1111/j.1399-3054.2009.01299.x

Single nucleotide polymorphisms (SNPs) and insertion–deletions (InDels) are becoming important genetic markers for major crop species. In this study, we focus on variations at genomic level of the *Daucus carota* L. *AOX2a* gene. The use of gene-specific primers designed in exon regions on the boundaries of introns permitted to recognize intron length polymorphism (ILP) in intron 3 *AOX2a* by simple polymerase chain reaction (PCR) assays. The length of intron 3 can vary in individual carrot plants. Thus, allelic variation can be used as a tool to discriminate between single plant genotypes. Using this approach, individual plants from cv. Rotin and from diverse breeding lines and cultivars were identified that showed genetic variability by *AOX2a* ILPs. Repetitive patterns of intron length variation have been observed which allows grouping of genotypes. Polymorphic and identical PCR fragments revealed underlying high levels of sequence polymorphism. Variability was due to InDel events and intron single nucleotide polymorphisms (ISNPs), with a repetitive deletion in intron 3 affecting a putative pre-miRNA site. The results suggest that high *AOX2a* gene diversity in *D. carota* can be explored for the development of functional markers related to agronomic traits.

Introduction

Sequence polymorphisms commonly occur in both protein-coding and non-coding regions. These variations can affect gene structures and expression patterns, and may contribute to phenotypic adaptations. Detection and analysis of those variations can help us to understand the molecular basis of various biological phenomena in plants. Because the entire plant kingdom cannot be covered under sequencing projects, molecular markers and their correlation to phenotypes provide us with requisite landmarks for the elucidation of important genetic variation.

Functional markers (FMs) for stress tolerance can be used in breeding programs to select genotypes with high yield stabilities under various conditions. Thus, a good marker should show a strong correlation with favorable adaptive plant behavior. The efficient reprogramming of target cells for yield determination is currently considered to be the most important step toward defining abiotic stress tolerance (Arnholdt-Schmitt 2005). The alternative oxidase (AOX) genes were previously suggested to play a crucial role in cell reprogramming under stress (Clifton et al. 2006) and proposed by Arnholdt-Schmitt et al. (2006) as a FM for genetic variation in cell reprogramming and yield stability.

Abbreviations – AS, alternative splicing; EPIC, exon-primed-intron-crossing; FM, functional marker; InDel, insertion and deletion; ISNP, intron single nucleotide polymorphism; ORF, Open reading frame; ROS, reactive oxygen species; SNP, single nucleotide polymorphism.

AOX is an inner mitochondrial membrane protein that functions as terminal oxidase in the alternative (cyanide resistant) pathway of respiration where it generates water from ubiquinol (Umbach et al. 2002). AOX serves to relieve oxidative stress originating from environmental stresses by limiting mitochondrial reactive oxygen species (ROS) formation and preventing specific components of the respiration chain from over-reduction (Popov et al. 1997, Purvis 1997) and canalizing ROS signals (Amirsadeghi et al. 2007). In higher plants, AOX is codified by a small multigene family with at least five genes (Clifton et al. 2006) belonging to two subfamilies: *AOX1*-type and *AOX2*-type genes.

Carrot is among the top-ten most economically important vegetable crops in the world, in terms of both area of production and market value (Simon et al. 2008), and is the most studied species for cell reprogramming, which recently obtained further novel importance from abiotic stress research (Costa et al. 2009a, Nishiwaki et al. 2000, Kikushi et al. 2006).

The availability of complete *AOX2* gene sequences at the cDNA level in *Daucus carota* L. (Campos et al. 2009) allows to investigate *AOX2a* gene variability at the genomic level. Single nucleotide polymorphism (SNP), insertion and deletion (InDel) events are the major driving forces that have shaped genomes (Zhang and Gerstein 2003) and are highly abundant and distributed throughout the genomes in various species including plants (Batley et al. 2003a, Drenkard et al. 2000, Garg et al. 1999, Nasu et al. 2002). This kind of changes mainly occurs in the non-protein-coding parts of a gene. The reason suggested is to reflect the strict functional requirements of the coding regions, being a clue that evolution has worked differently on protein-coding compared with intron sequences (Wang et al. 2005). SNPs and InDels can contribute directly to a phenotype (Thornsberry et al. 2001) or they can be associated with a phenotype shown by linkage disequilibrium (Daly et al. 2001). SNPs and InDels are becoming important genetic markers for major crop species for genetic research (e.g. construction of genetic maps, mapping of genes or quantitative trait loci) and breeding (e.g. marker-assisted selection) (Bi et al. 2006).

Polymorphisms located at the intronic sequences can be functionally critical in view of its potential to influence the binding of transcription factors (Xie et al. 2005), the process of alternative splicing (AS) (Baek et al. 2008, Ner-Gaon et al. 2007, Noh et al. 2006), the coding of intronic regulatory elements, such as micro- or small nucleolar-RNAs (Li et al. 2007) as well as nonsense-mediated mRNA decay (Jaillon et al. 2008). The organization of DNA into the highly condensed structure of eukaryotic chromosomes plays an important role in the regulation of gene expression, for DNA synthesis, recombination, and repair by modulating accessibility of DNA, and can also rely on motifs located in introns, such as scaffold or matrix attachment regions (S/MARs) (Arnholdt-Schmitt 2004, Shaposhnikov et al. 2007).

SNPs are the most abundant sequence variations encountered in most genomes (Cho et al. 1999; Griffin and Smith 2000). In plants, SNPs in gene regions are at least as abundant as in humans with the preliminary estimates ranging from 1 SNP per 60 bp in outbreeding maize (Ching et al. 2002) to ca. 1 SNP per 300 bp for inbreeding rice and *Arabidopsis thaliana* (Schmid et al. 2003, Yu et al. 2005). Some plant species also display a large number of insertions and deletions (Bhattramakki et al. 2002, Rafalski 2002). In addition, its abundance, ubiquity and interspersed nature of SNPs make them attractive tools for plant marker-assisted breeding and map-based gene cloning (Batley et al. 2003b, Bi et al. 2006, Chao et al. 2009, Gupta et al. 2001, Hayashi et al. 2004, Rafalski 2002, Rostoks et al. 2005, Sharma and Chauhan 2008). Several authors report the linkage between intron single nucleotide polymorphisms (ISNPs) at specific genes and phenotype changes, mostly related to human diseases (Hiratani et al. 2005, Jou et al. 2009).

InDels have been successfully exploited as genetic markers. There are several reports about its application in population genetics surveys (Bardini et al. 2004, Breviario et al. 2008, Côrte-Real et al. 1994, Daguin and Borsa 1999, Lessa 1992) and gene mapping (Wydner et al. 1994). Shimada et al. (2009) reported the use of intron length polymorphisms (ILPs) to distinguish gentian cultivars for plant protection and Song et al. (2008) could distinguish the geographical origin of *Clianthus* populations based on a short InDel (7 bp) in intron 2 of the gene *LEAFY*. Several authors report the application of ILPs as markers for agronomic traits, such as biotic stress tolerance (Slabaugh et al. 2003) or sensitivity to vernalization (Szücs et al. 2006). In addition, a recently genome-wide investigation was performed in *Oryza sativa* and highlighted ILPs as a potential source of molecular markers (Wang et al. 2005).

Recent knowledge on the regulatory functionality of intronic regions, and the knowledge of the *DcAOX2a* gene sequence made it reasonable to investigate the variability in intronic regions among genotypes in view of future application in carrot plant breeding. In this report, the variability of three intronic regions of *DcAOX2a* gene was analyzed among different carrot genotypes. ILPs were detected by varying polymerase chain reaction (PCR)-fragment patterns in intron 3. ISNPs were found as an additional source of polymorphism between alleles and genotypes. Further, *DcAOX2a* alleles were mapped to a linkage group in a segregating F2 population.

Materials and methods

Plant material

Genomic *DcAOX2a* gene isolation and characterization was performed from individual plants of the open pollinated cv. Rotin (*D. carota* L.). Seeds were germinated and grown on MS (Murashige and Skoog 1962) solid medium under sterile and controlled-climate conditions ($25 \pm 1^{\circ}$C at 16 h photoperiod: 95–100 µmol m^{-2} s^{-1}, Philips, Amsterdam, the Netherlands). Eight weeks after germination, an individual plant (whole plant sample) was selected for DNA extraction and *DcAOX2a* gene isolation.

To study sequence variation between diverse genomes of individual plants, seeds of cv. Rotin were germinated in a greenhouse in pots with 12 plants under controlled conditions. Twelve 10-week-old plants were harvested and DNA was extracted from the secondary phloem of the tap roots (named R1 to R12). In addition, DNA from leaves samples of 41 breeding lines of various origins from seed companies was included into the study.

For mapping the *DcAOX2a* gene, the genetically well-described carrot F_2 population DM19 was chosen. The population was obtained from a cross between the carrot mutants *yellow leaf* (*YEL*) (Nothnagel and Straka 2003) and *compressed lamina* (*COLA*) (Nothnagel et al. 2005). The F_2 plants ($n = 161$) were cultivated in 12 cm plastic pots and evaluated under glasshouse conditions. Focus of this double mutant assay was the investigation of the inheritance of both morphological traits and their localization in the carrot genome. The 161 F_2 plants of the mapping population DM19 segregated into four distinct phenotype classes, exactly in a Mendelian fashion as expected for two independent genes (9:3:3:1; $\chi^2 = 5.54$, $P = 0.14$). Genomic DNA used for molecular evaluation was isolated from leaves of the parental and individual F_2 plants.

DNA extraction

DNA extraction was performed using DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. DNA quantification was made in comparison to standard DNA concentrations of lambda DNA after electrophoresis in 1% agarose gel (Invitrogen, Karlsruhe, Germany) and visualization by ethidium bromide (EtBr, 2 ng/ml) using the Gene Flash Bio Imaging system (Syngene, Cambridge, UK).

Isolation of genomic *DcAOX2a* sequences

To isolate the complete genomic sequence of *DcAOX2a*, PCR was conducted using 10 ng of DNA as template in a reaction volume of 50 µl and specific primers designed according to the cDNA sequence already available (Campos et al. 2009). A forward primer located at the beginning of exon 1 DcAOX2a_30Fw and a reverse primer located at the 3′UTR region DcAOX2a_3UTRev were used (see Table 1). PCR was carried out for 35 cycles in the 2720 Thermocycler (Applied Biosystems, Foster City, CA). Each cycle consisted in 10 s at 98°C for denaturation, 20 s at 52°C for primers annealing and 2 min at 72°C for DNA synthesis. An initial step at 98°C for 30 s and a final step at 72°C for 10 min were included. For the PCR, a mix of 0.4 U of Phusion™ High-Fidelity DNA Polymerase (Finnzymes, Espoo, Finland) were used with 1× manufacturer-supplied Phusion HF buffer containing 1.5 m*M* $MgCl_2$, 0.2 m*M* of each dNTPs (Fermentas, Ontario, Canada) and 0.2 µ*M* of each primer were added. PCR products were analyzed by electrophoresis in 1.4% (w/v) agarose gel (Invitrogen, Karlsruhe, Germany) as previously described under DNA extraction section.

The amplicons were gel-purified with GFX PCR DNA and Gel Band Purification Kit (GE Healthcare, Little Chalfont, UK), and separately cloned into a bacterial plasmid pGem®-T Easy vector (Promega, Madison, WI), transformed into bacterial strain JM109 and selected according to the manufacturer's instructions (Promega, Madison, WI). Plasmid DNA was extracted according to the alkaline lyses protocol (Birnboim and Doly 1979) and analyzed with the restriction enzyme *Eco*RI (Fermentas, Ontario, Canada). Three clones per genotype showing

Table 1. List of the primers used in the study, correspondent sequence, location and aim of it use.

Primers	Sequence	Location	Aim	Origin
DcAOX2a_int1Fw	5′-CATCTGTTAGCCAAGTCTGTGA-3′	Exon 1	Intron 1 study	STABVIDA, Oeiras, Portugal
DcAOX2a_int1Rev	5′-CCGTCGGAATCCTTAGTAGTTT-3′	Exon 2	Intron 1 study	
DcAOX2a_int2Fw	5′-TAAGCACCATGTACCAAAGAC-3′	Exon 2	Intron 2 study	
DcAOX2a_int2Rev	5′-TCAGCACGGATCACAGTTAT-3′	Exon 3	Intron 2 study	
DcAOX2aFw	5′-TGCTGCATCTGAGGTCTCTCC-3′	Exon 3	Intron 3 study/gene mapping	Eurofins MWG Operon, Ebersberg, Germany
DcAOX2a_3UTRev	5′-TTCAGAGATATATAGCTATGTGG-3′	3′-UTR	Intron 3 study/gene isolation	
DcAOX2a_30Fw	5′-ATGAATCATCTGTTAGCCAAGTCTG-3	ORF	Gene isolation	
DcAOX2a_1303Rev	5′-CCAATCAATTCTACACAAACAACC-3′	3′-UTR	Gene mapping	

the fragment of interest were completely sequenced (Macrogen Company: www.macrogen.com) in the directions of sense and antisense strands using the primers T7 and SP6 (Promega, Madison, WI).

Sequence homology was searched in the NCBI data bases (National Center for Biotechnology Information, Bethesda, MD) using BLAST algorithm (Karlin and Altschul 1993) (http://www.ncbi.nlm.nih.gov/) (BLASTn and BLASTp). Intron location was made using the software Spidey, which is publicly available at (http://www.ncbi.nlm.nih.gov/IEB/Research/Ostell/Spidey/).

Search of ILPs in *DcAOX2a* introns by exon-primed-intron-crossing-polymerase chain reaction (EPIC-PCR)

Intron regions of *DcAOX2a* were amplified by specific primers designed in the exon boundaries for each of the three introns (see Table 1). To amplify intron 1, the forward primer DcAOX2a_int1Fw and the reverese primer DcAOX2a_int1Rev were used (amplicon with an expected size of 2511 bp), for intron 2 the primer pair DcAOX2a_int2Fw and DcAOX2a_int2Rev (amplicon with expected size of 1497 bp) was applied and for intron 3 primer pair DcAOX2aFw and DcAOX2a_3UTRev (amplicon with expected size of 1741 bp) succeeded amplification. PCR conditions were the same as described above using a temperature of 58°C instead of 52°C for primer annealing.

PCR products were analyzed as described under Isolation of genomic *DcAOX2a* sequences section.

Characterization of intron 3 sequence variability

Verification of fragment heterozygosity

To examine the nature of heterozygous fragments, 1:1 artificial mixtures of PCR fragments characteristic of both alleles were incubated at 94°C during 2 min for denaturation followed by an incubation at room temperature for 2 h for renaturation (Cavagnaro, personal communication). As a control, artificial mixtures were incubated on ice and run in parallel with the treated fragments in 1.4% (w/v) agarose gel (Invitrogen, Karlsruhe, Germany) electrophoresis. The band pattern was analyzed according to the procedure described under DNA extraction section.

Sequence analysis

Amplicons corresponding to the expected intron 3 fragments were cloned, selected and sequenced according to the procedure already described under Isolation of genomic *DcAOX2a* sequences section. Sequence homology was searched in the NCBI data bases using BLAST algorithm (Karlin and Altschul 1993) (http://www.ncbi.nlm.nih.gov/) (BLASTn). For sequence polymorphism analysis, the softwares MegAlign, SeqMan and EditSeq (LASERGENE 7, GATC Biotech, Konstanz) were used.

Pre-miRNAs–sequence and structure prediction

Predicting hairpins from scaffold sequences by means of Srnaloop. During miRNA maturation period, all full-length transcripts of miRNA gene form a hairpin structure, also called stem-loop structure. Such secondary structure is folded via intra-molecular base pairing and has been an important criterion for computational identification of miRNAs (Berezikov et al. 2005, Grad et al. 2003, Lai et al. 2003, Li et al. 2006). In this research, the program Srnaloop, developed by Grad et al. (2003), was used to identify putative hairpins from the intronic sequences. The optimized parameters for Srnaloop program were described previously (Li et al. 2006). Predicted candidate hairpins were further screened based on 'sequence and structure features filter'.

Sequence and structure features filter. Many genomic fragments other than authentic pre-miRNAs could be predicted to fold into hairpin structures. Therefore, we need more precise classification criteria specific to miRNAs to distinguish authentic candidates from false positives. miRNAs as well as pre-miRNAs have specific features on their sequences and structures, which comprise sequence and structural features filter, which is described previously (Li et al. 2006). Such strategy was also applied in viral miRNA identification (Li et al. 2007).

Features' values of plant miRNAs were used as reference values. A candidate whose guaninis and cytosines (GC) content, core mfe, hairpin mfe and ch_ratio values were within the individual reference range values was a positive candidate.

Secondary structure. Prediction of secondary structure of pre-miRNA was run on the web-based software MFOLD 3.4, which is publicly available at (http://frontend.bioinfo.rpi.edu/applications/mfold/cgi-bin/rna-form1.cgi) (Mathews et al. 1999, Zuker 2003).

Prediction of miRNAs and targets of miRNAs. For screening the candidates of potential miRNAs, the validated pre-miRNAs were run in the software miRBase::Sequences, which is publicly available at (http://microrna.sanger.ac.uk/sequences/search.shtml). BLASTx from web-site (http://www.ncbi.nlm.nih.gov/BLAST/) was used for the analysis of potential target genes (Mathews et al. 1999, Zuker 2003).

Gene mapping and map construction

For mapping, PCRs were conducted using 10 ng of DNA as template and the specific forward primer DcAOX2aFw combined with the reverse primer DcAOX2a_1303Rev (see Table 1). PCR was carried out in a Thermocycler (DNA Engine PTC-200 Bio-Rad, Hercules, CA) for 35 cycles each one consisting of 30s at 94°C, 30s at 58°C and 2 min at 72°C. An initial step of 94°C during 5 min and a final step at 72°C during 10 min were also performed.

For the PCR, a mix of 0.5 U of FastStart-Taq-DNA-Polymerase (Roche, Mannheim, Germany) were used with 1× manufacturer-supplied buffer with 2.2 m*M* $MgCl_2$, 0.35 m*M* of each dNTPs and 0.5 µ*M* of each primer were added. The band pattern analysis was performed in 1.5% (w/v) agarose gel LE (Biozym, Landgraaf, The Netherlands) as previously described under DNA extraction section.

An integrated linkage map using the JoinMap 3.0 software with a conservative logarithm of the odds (LOD) score of 6.0 (Van Ooijen and Voorrips 2001) was designed. Map distances (cM) between ordered loci were calculated using the recombination fraction and Kosambi mapping function (Kosambi 1944).

Results

The genomic structure of *DcAOX2a* and polymorphisms in intron 3

The complete cDNA sequence of the *DcAOX2a* gene had been published by Campos et al. (2009). A structure of four exons interrupted by three introns was indicated. Here, we report the complete genomic *DcAOX2a* sequence and show that the total genomic gene size can vary depending on the variable length of intron 3. Fig. 1 displays the result of EPIC-PCR analyses in 12 individual carrot plants of cv. Rotin for the three introns. It can be seen that EPIC-PCR for intron 1 and intron 2 show stable fragment lengths of around 2.5 and 1.5 kb in all plants. The length of the intron 1 and intron 2 sequences is 2063 and 957 bp, respectively. For intron 3, three plants of cv. Rotin display length variability. Three PCR fragments of different sizes (near 1.4, 1.7 and 1.8 kb) have been identified in all three individuals. However, the upper band was proved to be an artifact (see below). The two PCR fragments that included intron 3 contained also partial exon sequences of exon 3 (419 bp) and the complete exon 4 (57 bp). The sequence of intron 3 varied between 941 (in allele *S*, accession number GQ248713) and 1226 bp (in allele *L*, accession number GQ248714), whereas the included exon sequences were stable in length. The stability in length of intron 1 and intron 2 between individual plants could be confirmed in additional studies in 41 breeding lines of various origins (data not shown). The same length polymorphism was identified in these lines for intron 3. Both bands could be separated in breeding lines, which indicate the allelic nature of the two bands (data not shown).

The PCR-fragment patterns related to homozygocity (single bands corresponding to allele *S* or *L*) were found more frequently in this study than patterns related to heterozygocity (three band pattern), with a frequency rate of 75% (n = 9) to 25% (n = 3) in cv. Rotin and 90% (n = 37) to 10% (n = 4) in the breeding lines. The allele *L* was observed more often in cv. Rotin presenting a frequency rate of 75% (n = 9) and in the breeding lines of 51% (n = 21). The allele *S* with a frequency rate of 39% (n = 16) in the breeding lines was not present in the genotypes of cv. Rotin.

In Figs 2 and 3, it is shown that the length variation of intron 3 between both alleles of cv. Rotin is due to an InDel of 286 bp with the start in position 665 bp. Analysis in breeding lines highlighted the conserved nature of this InDel site, which appears to be independent from plants from diverse origins (data not shown). In addition, one gap of a single nucleotide, which was linked to the large InDel event, and five ISNPs were identified (Table 2). In positions 273, 976, 978, 1127, 1176 and 1205, the smaller allele (*S*) contains a C, C, T, A, C, C and G, whether it demonstrates T, T, C, a deletion, G, T and A in the larger intron (allele *L*). Further five ISNPs were observed that characterized individual genotypes. Four of them occurred in the larger fragments that have been repetitive between two genotypes of different origins. In position 351, T is substituted by G (genotypes J4 and V1), in position 455 it is a G/T and in positions 1127 and 1136 G/C and T/C substitutions were identified. A unique substitution of T by C was found in the smaller allele in position 978 of genotype V3 (Table 2).

Only one SNP has been identified in an exon region in the partial sequence of 419 bp of exon 3 of allele *S* (genotype V3), a SNP was identified that was responsible for a putative synonymous translation (data not shown). No polymorphisms were identified in exon 4.

The complete gene sequence of *DcAOX2a*, including exons and introns, consisted of 5263 or 4977 bp, depending on the size of intron 3 (Table 3). By comparing total gene sizes from all complete *AOX* genes available, *DcAOX2a* together with two *AOX2* genes of *Vitis vinifera* presents a group of larger *AOX* gene sequences. This is due to larger sizes of all three introns in these species (Table 3). Several of the other species present very small *AOX* introns. The smallest size for intron 3 was found in *AOX1a* from *A. thaliana* with 80 bp. Although exons in *AOX* gene sequences

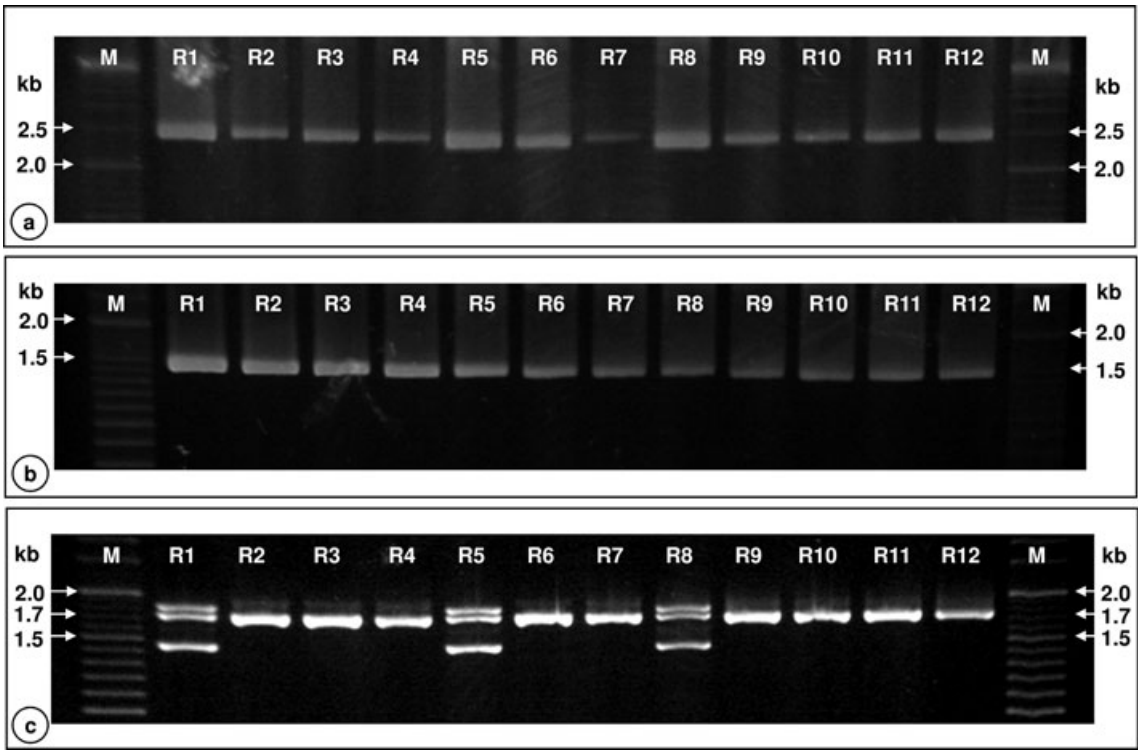


Fig. 1. Photograph of agarose gels showing the results of the introns analysis in 12 genotypes of *D. carota* cv. Rotin corresponding to individual plants. (a) intron 1, (b) intron 2 and (c) intron 3.

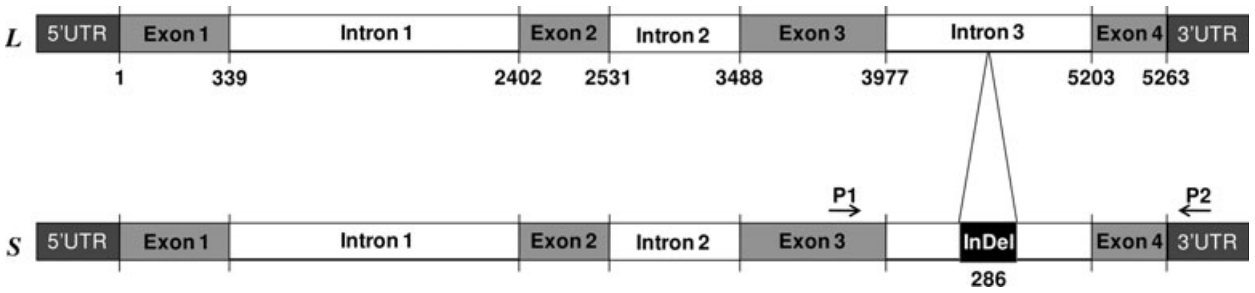


Fig. 2. Scheme of the genomic sequence of *DcAOX2a*. Analysis of intron 3 was performed using a primer located in exon 3 (P1: DcAOX2aFw) and a primer located in the 3′UTR (P2: DcAOX2a_3UTRev).

present a characteristic size pattern (Campos et al. 2009), comparing the intron sizes of *AOX2a* genes from several species demonstrate a high level of variation (Table 3). The introns did not reveal any kind of common length pattern between species.

In all intron sequences, the typical characteristic of the nuclear exon–intron junctures 5′-GU/AG-3′ was observed. The 5′ splice site is 5′-GU and at the 3′ splice site is AG-3′ at the extreme 5′ and 3′ ends of the intron (Saisho et al. 1997). The intronic region is more U rich than the flanking exons (37 vs 26% U on average). Similar results were previously reported in *A. thaliana* showing 41 vs 26% U on average (Deutsch and Long, 1999, Ko et al. 1998). According to Ko et al. (1998), U richness might be important for intron recognition.

Analysis of heterozygosity

D. carota is known to show diploidy. Thus, it was expected to observe a maximum of two alleles related to the studied gene. According to the present results, the single band patterns, showing a single band of 1.4 kb (correspondent to allele *S*) or 1.7 kb (correspondent to allele *L*) can correspond to the homozygous genotype for this gene (*SS* or *LL*). The heterozygous should present

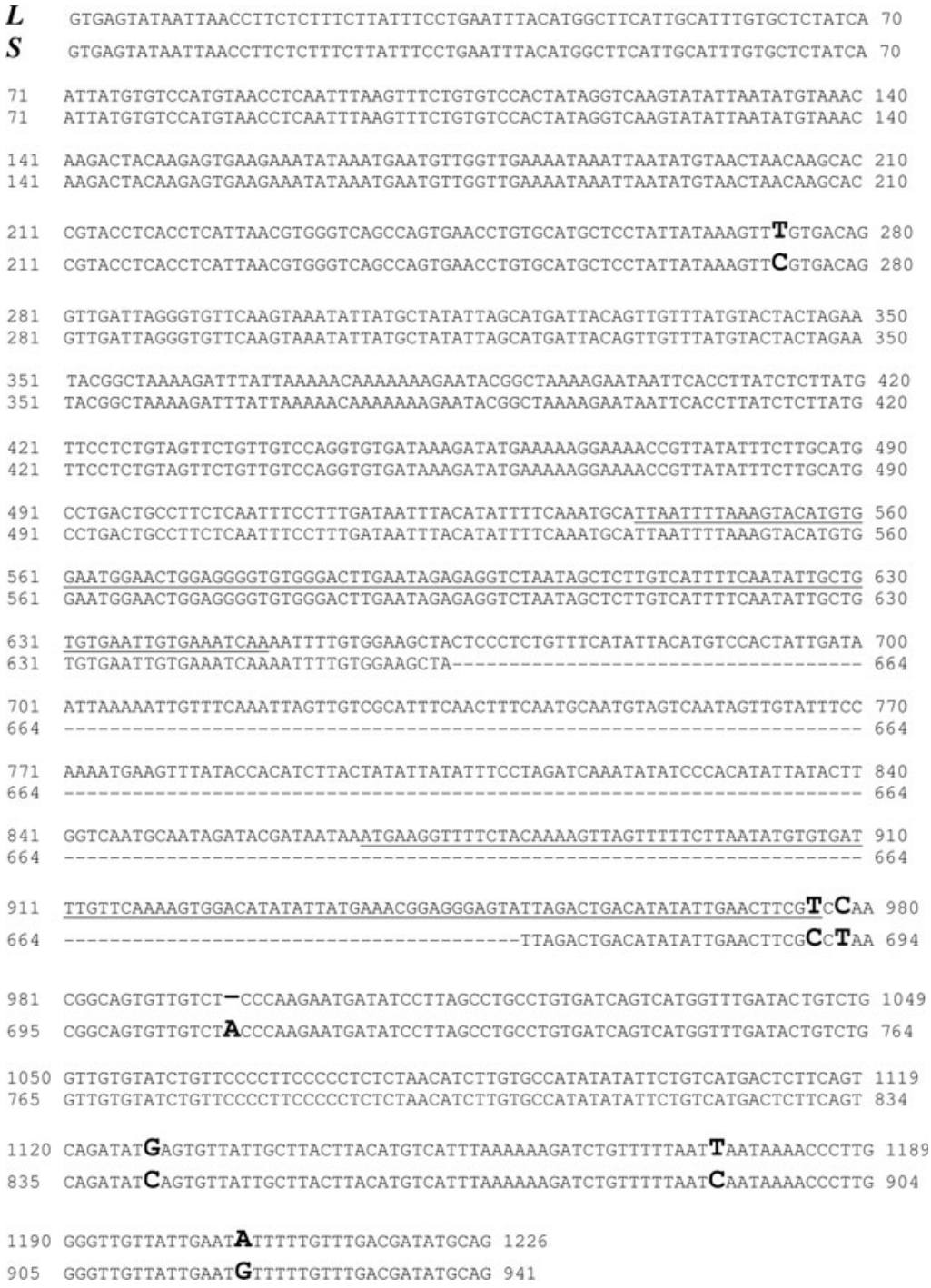


Fig. 3. PCR-fragment sequences of intron 3 corresponding to the fragment *L* and *S* identified in *D. carota* cv. Rotin. The different nucleotides (SNPs) identified between both sequences are in big letters. The positions of predicted pre-miRNAs are underlined.

both fragments (corresponding to *SL*). However, in addition to *S*- and *L*-associated fragments, the developed assay generates a third, seemingly larger PCR band that is specific to the heterozygote state (Fig. 4). This slower moving band was confirmed to be a heteroduplex resulting from one *S*-associated DNA strand pairing with one *L*-associated DNA strand. Although the strands from the different alleles are homologous in a large part of their sequence, the presence of the InDel is the *L*-associated strand results in conformational change

Table 2. Polymorphic nucleotide positions in intron 3 of D. Carota *AOX2a*. Alignment gaps are indicated by dashes. The nucleotide position is reoffered according the sequence *L*. Rt corresponds to a plant from cv. Rotin, all the others correspond to plants from breeding lines.

		Nucleotide position									
Fragment	Genotype	273	351	455	976	978	995	1127	1136	1176	1205
	Rt	T	T	G	T	C	–	G	T	T	A
	J1	T	T	G	T	C	–	G	T	T	A
	J3	T	T	G	T	C	–	G	T	T	A
	J4	T	G	T	T	C	–	C	C	T	A
L	3J2	T	T	G	T	C	–	G	T	T	A
	3J3	T	T	G	T	C	–	G	T	T	A
	3J4	T	T	G	T	C	–	G	T	T	A
	3J5	T	T	G	T	C	–	G	T	T	A
	V1	T	G	T	T	C	–	C	C	T	A
	V3	T	T	G	T	C	–	G	T	T	A
	Rt	C	T	G	C	T	A	C	T	C	G
	J2	C	T	G	C	T	A	C	T	C	G
	J3	C	T	G	C	T	A	C	T	C	G
	3J1	C	T	G	C	T	A	C	T	C	G
S	3J3	C	T	G	C	T	A	C	T	C	G
	3J4	C	T	G	C	T	A	C	T	C	G
	3J5	C	T	G	C	T	A	C	T	C	G
	V3	C	T	G	C	C	A	C	T	C	G
	V4	C	T	G	C	T	A	C	T	C	G
	V5	C	T	G	C	T	A	C	T	C	G

in the heteroduplex. This change results in retarded migration through an agarose gel, giving the appearance of a larger DNA fragment. Artificial 1:1 mixtures of the *S*- and *L*-associated fragments resulted in the generation of the heteroduplex band if the mixture is first heat-denaturated and then allowed to reanneal. Unheated samples, in contrast, showed no heteroduplex band formation (Fig. 4). In addition, cloned heteroduplex bands yielded plasmid inserts of only two sizes: 1456 bp (identical in size to the *S*-associated insert) and 1741 bp (identical in size to the *L*-associated insert). As expected, the DNA sequence of the smaller and larger heteroduplex inserts matched exactly those of the *S*- and *L*-associated inserts, respectively. In no case did the heteroduplex yield an insert size larger than the *L*-associated insert. The presence of this band in a PCR sample serves as an added confirmation of heterozygosity.

Identification of a pre-miRNA site in a polymorphic region of intron 3

Prediction of pre-miRNA sequences in intron 3 revealed two locations coding for precursors of miRNAs (Table 4). The sequence located in positions 541–648 bp (108 bp) was named pre-miRNA I, and the sequence located between 867 and 976 bp, with 110 bp, was named pre-miRNA II. Pre-miRNA II, whose secondary structure is presented in Table 4, is located in a region where the sequence *S* presents the deletion. This means that this sequence does not present a predicted pre-miRNA, and consequently does not codify the putative correspondent miRNA.

The analysis performed at the software miRBase using the complete sequence of predicted pre-miRNA II allowed to select miR1436 and miR1439 of *O. sativa* as more homologous (see alignments at Table 5). The BLAST analysis performed using the *D. carota* sequence homologous to those sequences revealed 100% homology in both cases (nts 7–21 for osa-miR1436; 2–18 for osa-miR1439) with the *D. carota C-EF1* gene, which codifies for the embryonic element-binding factor 1 (accession number AB188289.1). This result allows to hypothesize that this will be a target gene. The region of homology corresponds to a region in the 3′UTR (555–569 for osa-miR1436; 553–569 for osa-miR1439) in both cases.

Gene mapping: integration of *DcAOX2a* in a linkage group

The *DcAOX2a* primers amplified three bands with 1600, 1900 and 2000 bp for the maternal parent and one band with 1600 bp for the paternal parent of the DM19 population. Consequently, a homozygous state was postulated for the paternal parent and a heterozygous

Table 3. Comparison of intron size and complete gene size at the genomic level of the *AOX* genes.

Specie	NCBI accession	Gene	Intron size (bp)				Gene size exon/intron
			Intron 1	Intron 2	Intron 3	Intron 4	
A. thaliana	D89875	*AOX1a*	82	300	80		1 527
		AOX1b	83	82	86		1 149
	AB003175	*AOX1c*	95	138	84		1 307
	NC_003070	*AOX1d*	99	180	97		1 333
G. max	AF083880	*AOX1*	333	1040	149		2 488
O. sativa	AB004813	*AOX1a*	159	932	102		2 192
		AIX1b	111	108			1 227
	AB074004	*AOX1c*	126	90	90		1 344
S. tuberosum	DQ270421	*AOX1a*	75	164			1 310
V. unguiculata	DQ100440	*AOX1*	157	346	89		1 543
V. vinifera	CAAP01000118	*AOX1a*	93	106	83	125	1 245
	AM466432	*AOX1a*	109	106	83	125	1 261
	AM472072	*AOX1b*	80	109	94		1 252
A. thaliana	AB003176	*AOX2*	87	88	221	244	1 702
C. sativus	AY258276	*AOX2*	306	318	115		1 780
D. carota	GQ248714	*AOX2a*	2063	957	1226		5 263
	GQ248713				941		4 977
	Not published	*AOX2b*	822	91	85		1 958
G. max	AY303971	*AOX2a*	301	782	718		2 783
		AOX2b	244	510	318		2 053
V. unguiculata	EF187463	*AOX2a*	525	428	736		2 679
	DQ100439	*AOX2b*	168	226	194		1 566
V. vinifera	CAAP01005007	*AOX2*	2776	6885	1372		11 996
	AM459831	*AOX2*	2735	1857	1383		6 938

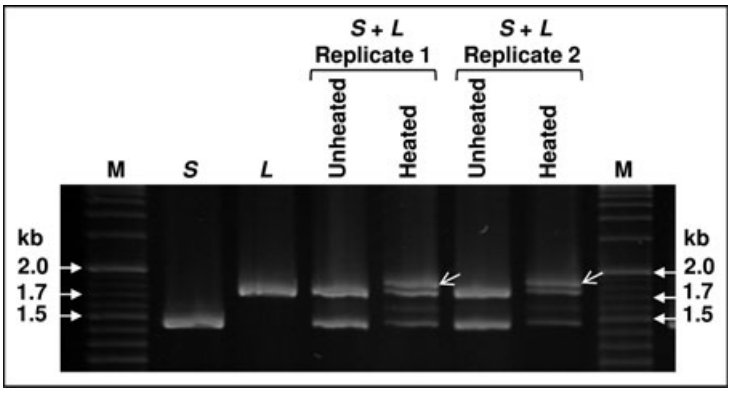


Fig. 4. Agarose gel showing results of the reannealing experiment (see section on Materials and methods). Lanes 2 and 3 show PCR products generated from genomic DNA using the described PCR assay. Predicted genotypes are *S* and *L* as shown by the presence of a single, allele-specific band in each of these lanes. Arrow shows the location of the heterozygote-specific PCR fragment. Artificial 1:1 mixtures of *S* and *L* PCR products result in little or any of the heterozygote-specific product if the mixtures are kept on ice (lanes 4 and 6). If the same samples are subjected to a single round of denaturation and reannealing, the heterozygote-specific band is visible (5 and 7), suggesting that the larger heterogygote-specific band is actually a heteroduplex resulting from one strand of the *S*-specific product and one strand of the *L*-specific product.

state for the maternal parent (data not shown). Only the two bands correspondent to alleles *L* and *S* were considered for mapping, the band correspondent to the heteroduplex was excluded from the analysis.

Following individual analysis of the 161 F_2 plants revealed segregation for three *DcAOX2a* genotype classes *LL:LS:SS* in an ratio of 37:54:33, respectively. Chi-square test was used to test the goodness-of-fit to expected ratio for a monogenic inheritance. A good fit to a co-dominant inheritance ($\chi^{2(1:2:1)} = 2.3$, $P = 0.31$) was observed. Thirty-seven plants were identified that did not show any of the two bands.

The *DcAOX2a* marker data were integrated into the existing DM19 data matrix generated by different

Table 4. Structure of the predicted pre-miRNAs identified in intron 3 of *AOX2a* in *D. carota* cv. Rotin.

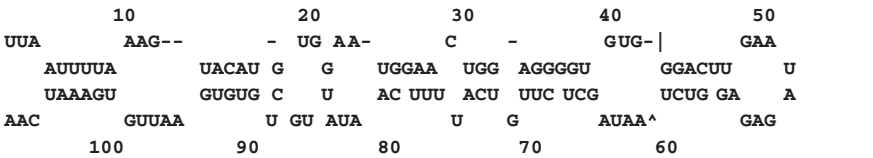
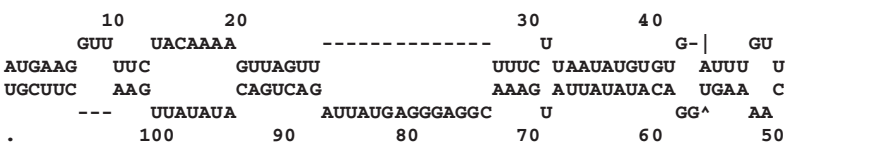
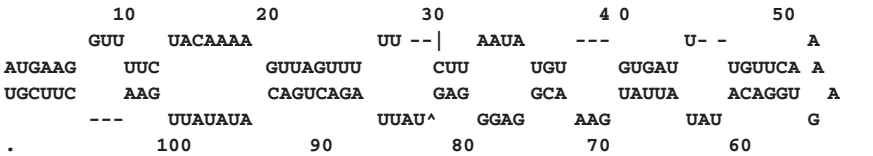
		Structure of the predicted pre-miRNA
pre-miRNA I	Length: 108 bp Location: 541–648 bp MFE: –23.70 kcal mol^{-1}	
pre-miRNA II	Length: 110 bp Location: 867–976 bp MFE: –19.60 kcal mol^{-1}	
	Length: 110 bp Location: 867–976 bp MFE: –19.40 kcal mol^{-1}	

Table 5. Result of the alignment obtained in the software miRBase::Sequences using the predicted pre-miRNA II sequence.

			Alignment	
osa-miR1436 score: 78 evalue: 13	Carrot	63	AUAUUAUGAAACGGAGGGAGU	83
	osa-miR1436	1	ACAUUAUGGGACGGAGGGAGU	21
osa-miR1439 score: 78 evalue: 13	Carrot	66	UUAUGAAACGGAGGGAGUAUU	86
	osa-miR1439	1	UUUUGGAACGGAGUGAGUAUU	21

PCR-based techniques such as RAPD, AFLP, SSR (Niemann 2001), *DcMaster* transposons (Grzebelus et al. 2007) and STS. The preliminary map contains more than 276 markers distributed over 14 linkage groups for a total length of 1171 cM. The average distance between markers was estimated to be 4.2 cM (unpublished data, publication in progress).

The *DcAOX2a* gene could be mapped to the DM19 population on group 7 (Fig. 5) linked together with 20 AFLP, 8 RAPD and 2 retrotransposon marker.

Discussion

The genomic organization of plant *AOX* genes is characterized by conserved intron positions that commonly consist of four exons interrupted by three introns (Considine et al. 2002, Li et al. 2008, Polidoros et al. 2005, Velasco et al. 2007). Intron loss and gain (Considine et al. 2002, Ito et al. 1997, Saisho et al. 2001) as well as variable intron length were reported (see Table 3). More information about intron loss or gain can be consulted in Polidoros et al. (2009). Intron loss is considered by several authors as the main driving force for intron evolution (Blake 1978, Roy 2003). *DcAOX2a* revealed the most common structure for *AOX* genes. The exons show a well-conserved size in relation to other plant species (Campos et al. 2009). As previously reported for other genes of eukaryotic model organisms (Deutsch and Long 1999), the size of *AOX* introns varies greatly within genes and from one organism to another. In comparison to the other *AOX* genes of *D. carota*, in *DcAOX2a* larger intron sizes and consequently a larger primary transcript size were detected. The bigger size of *DcAOX2a* gene in comparison to *DcAOX2b* gene

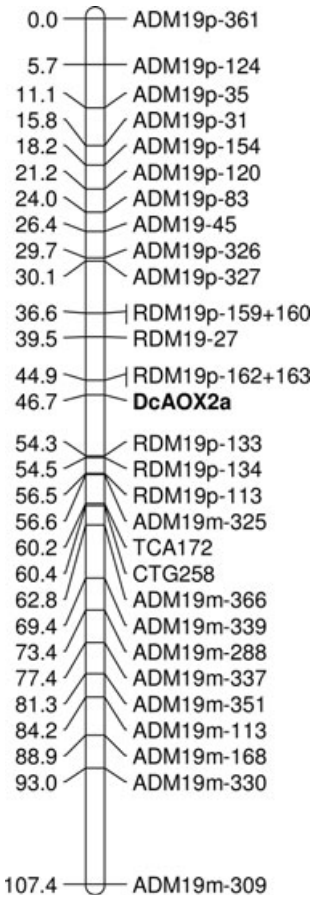


Fig. 5. Linkage group 7 of *D. carota* (mapped to the DM19 population) showing the three *DcAOX2a* markers, linked together with 22 AFLP, 9 RAPD and 2 retrotransposon markers previously identified.

can be one reason why less transcript abundance of *DcAOX2b* was detected in different organs and tissues and during growth regulation (Campos et al. 2009). In animals, highly expressed genes typically have smaller primary transcripts with fewer and smaller introns. In plants, several highly expressed genes were found to have more and longer introns and a larger primary transcript than genes expressed at low level (Castillo-Davis et al. 2002, Vinogradov 2004). Higher expressed genes tend to be less compact than lower expressed genes (Ren et al. 2006). Costa et al. (2009b) identified higher numbers of expressed sequence tags (ESTs) for the largest *AOX* gene, *AOX2*, in comparison to the two smaller *AOX1* genes discovered in grapevine. Two grapevine cultivars could be discriminated by ILP of *AOX2*. Variability of the length of *AOX2* was importantly influenced by the insertion of a large retroelement of >5000 bp in intron 2.

For most of the last century, research in genetics tended to focus on the stability of gene sequences from generation to generation and the stability of genomes through evolutionary time. More recently, rapid accumulation of data on genomic changes has shifted the emphasis from the stability of genomes to their potential plasticity. In addition to retroelements, mutations through spontaneous InDel events and nucleotide substitutions are thought to be the major driving forces in genome evolution (Gregory 2004, Zhang and Gerstein 2003). Introns as non-protein-coding intragene sequences with highly polymorphic rates are considered as the major source of genomic differences among species (Gibbs 2003). Evidences showing the important regulatory role of plant introns in the control of gene expression (Fiume et al. 2004, Gianì et al. 2003, Rose 2002) allowed lately that introns gained new credit in the scientific community (Rodríguez-Trelles et al. 2006, Roy and Gilbert 2006).

Natural variability in introns of *AOX* genes was also reported in *AOX1a* of *Triticum aestivum* (Naydenov et al. 2005) and *AOX1b* of *Hypericum perforatum* (Ferreira et al. 2009). Both studies revealed InDels consisting of few nucleotides that were dispersed in different introns. In wheat, the authors referred small deletions of 6–17 bp and in *H. perforatum* of 1–35 bp. In addition, a study performed in intron 3 of *AOX2* of *Olea europaea* (Macedo et al. 2009) revealed also ISNPs and an InDel of a single nucleotide, which appears not to be random. In all cases, the InDels are conserved between several sequences/genotypes. A genome-wide investigation of ILP made in rice revealed that the length difference between allelic introns in candidate ILP marker loci appeared to follow an exponential distribution with a mean value of 11.42 bp. Most (72.6%) had a size <5 bp, 23.5% fell between 5 and 50 bp and very few (3.9%) were larger than 50 bp (Wang et al. 2005). The interesting aspect related to *D. carota* is the length of the InDel with 286 bp and a non-random distribution, in a region of the intron, which was conserved among various genotypes. This disparity at position 665 bp of intron 3 is difficult to attribute to an insertion or a deletion event. In general, deletions are over three times more common than insertions (Blumenstiel et al. 2002, Zhang and Gerstein 2003), and small deletions occur frequently in higher plants (Bennetzen et al. 2005, Gu et al. 2006), which suggest a deletion rather than an insertion event for the InDel.

The utility of ILPs as FMs in plants for agronomic traits was reported (Slabaugh et al. 2003, Szücs et al. 2006). Several authors using ILPs as molecular markers referred a correlation between the number of amplified bands and the ploidy level of the taxon (Breviario et al. 2007, Gianì et al. 2003). Analysis of intron 3 of *AOX2a* in carrot revealed a three band pattern in the diploid

plant, which contradicts those observations. However, evaluation of the nature of the heterozygote bands identified the highest band as a heteroduplex artifact. Similar results were already reported by Bradeen and Simon (1998) when studying a PCR-based marker linked to the carrot Y_2 locus.

Compared with mitochondrial or chloroplast-located molecular markers (Grivet et al. 1996), the ILPs of intron 3 in *DcAOX2a* gene are based on Mendelian inheritance of nuclear DNA sequences, which permitted the mapping of the identified fragments in the linkage group 7 of a F_2 population (DM19) of *D. carota*. Recent reports have shown ILP to be a convenient and reliable molecular marker with high interspecies transferability. It can be exploited for the construction of genetic maps because it directly reflects within-gene variation (Wang et al. 2005).

Sequencing of intron regions permits to identify variations such as the ISNPs that are not detectable by electrophoresis. In intron 3 of *DcAOX2a*, the frequency of ISNPs was higher as the SNP frequency in the partial open reading frame (ORF) (partial exon 3 and complete exon 4). Genetic variability in intron sequences is known to be higher than in exon sequences. Wang et al. (2005) discovered that the number of ISNPs in rice was three to six times higher than for exons. Also Ferreira et al. (2009) identified a high frequency of ISNPs in comparison to the conserved nature of exon sequences in a partial *AOX* gene of *H. perforatum*.

The involvement of introns in the regulation of gene expression can be due to intronic sites for important regulatory elements, such as miRNAs. miRNAs inhibit translation of target genes by binding to their mRNAs. Recently discovered, the miRNAs have emerged as important players in plant stress responses, playing vital roles in plant resistance to abiotic as well as biotic stresses (Chiou et al. 2006). The control of plant development has already been related to miRNA, and there is evidence for the regulation of root growth (Wang et al. 2004), leaf development (Mallory et al. 2004), flower development and fertility (Achard et al. 2004). The identification of a putative pre-miRNA sequence in a region with an InDel allowed us to predict that the *S*-homozygous lack a pre-miRNA sequence and consequently will be unable to codify the corresponding miRNA. In future experiments, differences between *S*-homozygous and *L*-homozygous will be validated for differential expression of miRNA and a link to plant phenotypes.

Many miRNA families are evolutionarily conserved across all major lineages of plants, including bryophytes, lycopods, ferns, and monocots and dicots (Axtell and Bartel 2005). This conservation makes it possible to identify homologous miRNAs in other species (Axtell and Bartel 2005, Zhang et al. 2005). Sunkar et al. (2008) presented the osa-miRNA1436 as conserved in rice and *Aegilops*. A BLAST analysis using the correspondent miRNA sequences revealed high homology with *C-EF8* (accession number AB188296.1), a *D. carota* gene coding for the embryonic element-binding factor 8. Interestingly, homology of both miRNA sequences from rice and *Aegilops* with carrot, and the homology with the predicted miRNA sequences from carrot are all related to genes involved in embryogenesis. Somatic embryogenesis is currently studied in carrot as an important example of stress-induced cell reprogramming and differential expression of *AOX* during the induction phase (Costa et al. 2009a, Frederico et al. 2009).

Introns can also influence splicing decisions and induce either exon skipping or intron retention (Aoufouchi et al. 1996, Valentine 1998). Single base changes that affect splicing can have dramatic effects on gene function and consequently on the phenotype. ISNPs were already related to AS (Kawase et al. 2007). A strong correlation was found for those close to the intron–exon boundaries of the splicing events (Hull et al. 2007). This makes the ISNP identified in carrot at position–22 from the intron–exon boundary of high interest for further studies.

The results obtained for within-gene variability of *DcAOX2a* clearly contradicts any concept of co-evolution of introns and protein-coding sequences. Exon and intron sequences of *DcAOX2a* underwent different rates of evolutionary changes. Protein-coding sequences experienced a higher level of constraint, and thus, a lower rate of change. It may be suggested that this was a consequence of the key function of AOX in the energy bypasses of the respiratory chain. In contrast, *DcAOX2a* intron evolution has been more relaxed. The variability discovered in intron 3 of the *DcAOX2a* gene can be an important source of polymorphisms, which will permit to access gene diversity in *D. carota*. A later association of these polymorphisms with important plant traits would contribute to bridge the gap between genomics and phenomics.

Acknowledgements – This work was supported by the European Commission through a Marie Curie Chair for Birgit Arnholdt-Schmitt and by Fundação para a Ciência e a Tecnologia for scholarships to M.D.C. (SFRH/BI/15991/2006) and H.C. (SFRH/BPD/27016/2006). Collaboration with Germany was supported by a bilateral co-operation contract GRICES/FCT and DAAD. The authors would like to thank Vilmorin & Cie, France, for financial support and providing us carrot breeding material. The authors would like to thank Pablo Cavagnaro (University

of Wisconsin, USA) for important technical discussions related to artificial heteroduplex analysis, and to Raymond Li (National Yang-Ming University, Taiwan) for the in silico analysis of pre-miRNA in intron 3 sequences of *DcAOX2a* gene.

References

Achard P, Herr A, Baulcombe DC, Harberd NP (2004) Modulation of floral development by a gibberelin-regulated microRNA. Development 131: 3357–3365

Amirsadeghi S, Robson CA, Vanlerberghe GC (2007) The role of the mitochondrion in plant responses to biotic stress. Physiol Plant 129: 253–266

Aoufouchi S, Yélamos J, Milstein C (1996) Nonsense mutations inhibit RNA splicing in a cell-free system: recognition of mutant codon is independent of protein synthesis. Cell 85: 415–422

Arnholdt-Schmitt B (2004) Stress-induced cell reprogramming. A role for global genome regulation? Plant Physiol 136: 2579–2586

Arnholdt-Schmitt B (2005) Functional markers and a 'systemic strategy': convergency between plant breeding, plant nutrition and molecular biology. Plant Physiol Biochem 43: 817–820

Arnholdt-Schmitt B, Costa JH, Fernandes de Melo D (2006) AOX–A functional marker for efficient cell reprogramming under stress? Trends Plant Sci 11: 281–287

Axtell MJ, Bartel DP (2005) Antiquity of microRNAs and their targets in land plants. Plant Cell 17: 1658–1673

Baek J-M, Han P, Iandolino A, Cook DR (2008) Characterization and comparison of intron structure and alternative splicing between *Medicago truncatula*, *Populus trichocarpa*, *Arabidopsis* and rice. Plant Mol Biol 67: 499–510

Bardini M, Lee D, Donini P, Mariani A, Gianì S, Toschi M, Lowe C, Breviario D (2004) Tubulin-based polymorphism (TBP): a new tool, based on functionally relevant sequences, to assess genetic diversity in plant species. Genome 47: 281–291

Batley J, Barker G, O'Sullivan H, Edwards KJ, Edwards D (2003a) Mining for single nucleotide polymorphisms and insertions/deletions in maize expressed sequence tag data. Plant Physiol 132: 84–91

Batley J, Mogg R, Edwards D, O'Sullivan H, Edwards KJ (2003b) A high-throughput SNuPE assay for genotyping SNPs in the flanking regions of *Zea mays* sequence tagged simple sequence repeats. Mol Breed 11: 111–120

Bennetzen JL, Ma J, Devos KM (2005) Mechanisms of recent genome size variation in flowering plants. Ann Bot 95: 127–132

Berezikov E, Guryev V, van de Belt J, Wienholds E, Plasterk RH, Cuppen E (2005) Phylogenetic shadowing and computational identification of human microRNA genes. Cell 120: 21–24

Bhattramakki D, Dolan M, Hanafey M, Wineland R, Vaske D, Register JC III, Tingey SV, Rafalski A (2002) Insertion-deletion polymorphisms in 3′ regions of maize genes occur frequently and can be used as highly informative genetic markers. Plant Mol Biol 48: 539–547

Bi IV, McMullen MD, Sanchez-Villeda H, Schroedor S, Gardiner J, Polacco M, Soderlund C, Wing R, Fang Z, Coe EH Jr (2006) Single nucleotide polymorphisms and insertion-deletions for genetic markers and anchoring the maize fingerprint contig physical map. Genomics Mol Gen Biotech 46: 12–21

Birnboim HC, Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res 24: 1513–1523

Blake C (1978) Do genes-in-pieces imply proteins-in-pieces. Nature 273: 267–268

Blumenstiel JP, Hartl DL, Lozovsky EL (2002) Patterns of insertion and deletion in contrasting chromatin domains. Mol Biol Evol 19: 2211–2225

Bradeen JM, Simon PW (1998) Conversion of an AFLP fragment linked to the carrot Y_2 locus to a simple, codominant, PCR-based marker form. Theor Appl Genet 97: 960–967

Breviario D, Baird W, Sangoi S, Hilu K, Blumetti P, Gianì S (2007) High polymorphism and resolution in targeted fingerprinting with combined β-tubulin introns. Mol Breed 20: 249–259

Breviario D, Gianì S, Ponzoni E, Mastromauro F, Morello L (2008) Plant tubulin intronics. Cell Biol Int 32: 571–573

Campos MD, Cardoso HG, Linke B, Costa JH, Fernandes de Melo D, Justo L, Frederico AM, Arnholdt-Schmitt B (2009) Differential expression and co-regulation of carrot *AOX* genes (*Daucus carota*). Physiol Plant 137: 578–591

Castillo-Davis CI, Bedford TBC, Hartl DL (2002) Selection for short introns in highly expressed genes. Nat Genet 31: 415–418

Chao S, Zhang W, Akhunov E, Sherman J, Ma Y, Luo M-C, Dubcovsky J (2009) Analysis of gene-derived SNP marker polymorphism in US wheat (*Tritricum aestivum* L.) cultivars. Mol Breed 23: 23–33

Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski AJ (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genet 3: 19–33

Chiou TJ, Aung K, Lin SI, Wu CC, Chiang SF, Su CL (2006) Regulation of phosphate homeostasis by MicroRNA in *Arabidopsis*. Plant Cell 18: 412–421

Cho RJ, Mindrinos M, Richards DR, Sapolsky RJ, Anderson M, Drenkard E, Dewdney J, Reuber TL, Stammers M, Federspiel N, Theologis A, Yang WH, Hubbell E, Au M, Chung EY, Lashkari D, Lemieux B, Dean D, Lipshutz RJ, Ausubel FM, Davis RW, Oefner PJ (1999) Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. Nat Genet 23: 203–207

Clifton R, Millar AH, Whelan J (2006) Alternative oxidases in *Arabidopsis*: a comparative analysis of differential expression in the gene family provides new insights into function of non-phosphorylating bypasses. Biochim Biophys Acta 1757: 730–741

Considine M, Holtzapffel R, Day D, Whelan J, Millar A (2002) Molecular distinction between alternative oxidase from monocots and dicots. Plant Physiol 129: 949–953

Côrte-Real HBSM, Dixon DR, Holland PWH (1994) Intron-targeted PCR: a new approach to survey neutral DNA polymorphism in bivalve populations. Mar Biol 120: 407–413

Costa JH, Cardoso HC, Campos MD, Zavattieri A, Frederico AM, Fernandes de Melo D, Arnholdt-Schmitt B (2009a) *D. carota* L.–an old model for cell reprogramming gains new importance through a novel expansion pattern of *AOX* genes. Plant Physiol Biochem 47: 753–759

Costa JH, Fernandes de Melo D, Gouveia Z, Cardoso HG, Peixe A, Arnholdt-Schmitt B (2009b) The alternative oxidase family of *Vitis vinifera* reveals an attractive model for genomic design. Physiol Plant 137: 553–565

Dagiun C, Borsa P (1999) Genetic characterization of *Mytilus galloprovincialis* Lmk. In North West Africa using nuclear DNA markers. J Exp Mar Biol Ecol 235: 55–65

Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES (2001) High-resolution haplotype structure in the human genome. Nat Genet 29: 229–232

Deutsch M, Long M (1999) Intron–exon structures of eukaryotic model organisms. Nucleic Acids Res 27: 3219–3228

Drenkard E, Richter BG, Rozen S, Stutius LM, Angell NA, Mindrinos M, Cho RJ, Oefner PJ, Davis RW, Ausubel FM (2000) A simple procedure for the analysis of single nucleotide polymorphisms facilitates map-based cloning in *Arabidopsis*. Plant Physiol 124: 1483–1492

Ferreira A, Cardoso HG, Macedo ES, Breviario D, Arnholdt-Schmitt B (2009) Intron polymorphism pattern in *AOX1b* of wild St John's wort (*Hypericum perforatum*) allows discrimination between individual plants. Physiol Plant 137: 520–531

Fiume E, Christou P, Gianì S, Breviario D (2004) Introns are key elements of rice tubulin expression. Planta 218: 693–703

Frederico AM, Zavattieri MA, Campos MD, Cardoso HG, McDonald AE, Arnholdt-Schmitt B (2009) The gymnosperm *Pinus pinea* contains both *AOX* gene subfamilies, *AOX1* and *AOX2*. Physiol Plant 137: 566–577

Garg K, Green P, Nickerson DA (1999) Identification of candidate coding region single nucleotide polymorphisms in 165 human genes using assembled expressed sequence tags. Genome Res 9: 1087–1092

Gianì S, Morello L, Bardini M, Breviario D (2003) Tubulin intron sequences: multi-functional tools. Cell Biol Int 27: 203–205

Gibbs WW (2003) The unseen genome: gems among the junk. Sci Am 289: 46–53

Grad Y, Aach J, Hayes GD, Reinhart BJ, Church GM, Ruvkun G, Kim J (2003) Computational and experimental identification of *C. elegans* microRNAs. Mol Cell 11: 1253–1263

Gregory TR (2004) Insertion-deletion bases and the evolution of genome size. Gene 423: 15–34

Griffin TJ, Smith LM (2000) Genetic identification by mass spectrometric analysis of single-nucleotide polymorphisms: ternary encoding of genotypes. Anal Chem 72: 3298–3302

Grivet L, d'Hont A, Roques D, Feldmann P, Lanaud CE, Glaszmann JC (1996) RFLP mapping in cultivated sugarcane (*Saccharum* spp.): genome organization in a highly polyploid and aneuploid interespecific hybrid. Genetics 142: 987–1000

Grzebelus D, Jagosz B, Simon PW (2007) The DcMaster Transposon Display maps polymorphic insertion sites in the carrot (*Daucus carota* L.) genome 1. Gene 390: 67–74

Gu Y, Salse J, Coleman-Derr D, Dupin A, Crossman C, Lazo G, Huo N, Belcram H, Ravel C, Charmet G, Charles M, Anderson O, Chalhoub B (2006) Types and rates of sequence evolution at the high-molecular-weight glutenin locus in hexaploid wheat and its ancestral genomes. Genetics 174: 1493–1504

Gupta PK, Roy JK, Prasad M (2001) Single nucleotide polymorphisms: a new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. Curr Sci 80: 524–535

Hayashi K, Hashimoto N, Daigen M, Ashikawa I (2004) Development of PCR-based SNP markers for rice blast resistance genes at the Piz locus. Theor Appl Genet 108: 1212–1220

Hiratani H, Bowden DW, Ikegami S, Shirasawa S, Shimizu A, Iwatani Y, Akamizu T (2005) Multiple SNPs in intron 7 of Thyrotropin Receptor are associated with graves' disease. J Clin Endocrin Metab 90: 2898–2903

Hull J, Campino S, Rowlands K, Chan M-S, Copley RR, Taylor MS, Rockett K, Elvidge G, Keating B, Knight J, Kwiatkowski D (2007) Identification of common genetic variation that modulates alternative splicing. PLoS Genet 3: 1009–1018

Ito Y, Saisho D, Nakazono M, Tsutsumi N, Hirai A (1997) Transcript levels of tandem-arranged alternative oxidase genes in rice are increased by low temperature. Gene 203: 121–129
Jaillon O, Bouhouche K, Gout JF, Aury JM, Noel B, Saudemont B, Nowacki M, Serrano V, Porcel BM, Segurens B, Mouël A, Lepère G, Schächter V, Bétermier M, Cohen J, Wincker P, Sperling L, Duret L, Meyer E (2008) Translational control of intron splicing in eukaryotes. Nature 451: 359–362
Jou YS, Lo YL, Hsiao CF, Chang GC, Tsai YH, Su WC, Chen YM, Huang MS, Chen HL, Chen CJ, Yang PC, Hsiung CA (2009) Association of an *EGFR* intron 1 SNP with never-smoking female lung adenocarcinoma patients. Lung Cancer 64: 251–256
Karlin S, Altschul SF (1993) Applications and statistics for multiple high-scoring segments in molecular sequences. Proc Natl Acad Sci USA 90: 5873–7
Kawase T, Akatsuka Y, Torikai H, Morishima S, Oka A, Tsujimura A, Miyazaki M, Tsujimura K, Miyamura K, Ogawa S, Inoko H, Morishima Y, Kodera Y, Kuzushima K, Takahashi T (2007) Alternative splicing due to an intronic SNP in HMSD generates a novel minor histocompatibility antigen. Blood 110: 1055–1063
Kikushi A, Sanuki N, Higashi K, Koshiba T, Kamada H (2006) Abscisic acid and stress treatment are essential for the acquisition of embryogenic competence by carrot somatic cells. Planta 223: 637–645
Ko CH, Brendel V, Taylor RD, Walbot V (1998) U-richness is a defining feature of plant introns and may function as an intron recognition signal in maize. Plant Mol Biol 36: 573–583
Kosambi DD (1944) The estimation of map distances from recombination value. Ann Eugen 12: 172–175
Lai EC, Tomancak P, Williams RW, Rubin GM (2003) Computational identification of Drosphila micro RNA genes. Genome Biol 4: R42–42.20
Lessa EP (1992) Rapid surveying of DNA sequence variation in natural populations. Mol Biol Evol 9: 323–330
Li SC, Pan CY, Lin WC (2006) Bioinformatic discovery of microRNA precursors from human ESTs and introns. BMC Genomics 7: 164
Li SC, Shiau CK, Lin WC (2007) Vir-Mir db: prediction of viral microRNA candidate hairpins. Nucleic Acids Res 36: D184–D189
Li F, Zhang Y, Wang M, Zhang Y, Wu X, Guo X (2008) Molecular cloning and expression characteristics of alternative oxidase gene of cotton (*Gossypium hirsutum*). Mol Biol Rep 35: 97–105
Macedo ES, Cardoso HG, Hernández A, Peixe AA, Polidoros A, Ferreira A, Cordeiro A, Arnholdt-Schmitt B (2009) Physiologic responses and gene diversity indicate olive alternative oxidase as a potential source for markers involved in efficient adventitious root induction. Physiol Plant 137: 532–552
Mallory AC, Reinhart BJ, Jones-Rhoades MW, Tang G, Zamore PD, Barton MK, Bartel DP (2004) MicroRNA control of PHABULOSA in leaf development: importance of pairing to the microRNA 5′ region. EMBO J 23: 3356–3364
Mathews DH, Sabina J, Zuker M, Turner DH (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J Mol Biol 288: 911–940
Murashige T, Skoog F (1962) A revised medium for rapid growth and bio assays with tobacco tissues cultures. Physiol Plant 15: 473–497
Nasu S, Suzuki J, Ohta R, Hasegawa K, Yui R, Kitazawa N, Monna L, Minobe Y (2002) Search for and analysis of single mucleotide polymorphisms (SNPs) in rice (*Oryza sativa*, *Oryza rufipogon*) and establishment of SNP markers. DNA Res 9: 163–171
Naydenov N, Takumi S, Sugie A, Ogihara Y, Atanassov A, Nakamura C (2005) Structural diversity of the wheat nuclear gene *WAOX1a* encoding mitochondrial alternative oxidase, a single unique enzyme in the cyanide-resistant alternative pathway. Biotech Biotech Equip 19: 48–56
Ner-Gaon H, Leviatan N, Rubin E, Fluhr R (2007) Comparative cross-species alternative splicing in plants. Plant Physiol 144: 1632–1641
Niemann M (2001) Entwicklung von Mikrosatelliten-Markern bei der Möhre (*Daucus carota* L.) und die Markierung eines *Alternaria*-Resistenzgens. Univ Shaker, Aachen, Germany, pp. 101
Nishiwaki M, Fujino K, Koda Y, Masuda K, Kikuta Y (2000) Somatic embryogenesis induced by the simple application of abscisic acid to carrot (*Daucus carota* L.) seedlings in culture. Planta 211: 756–759
Noh SJ, Lee K, Paik H, Hur CG (2006) TISA: tissue-specific alternative splicing in human and mouse genes. DNA Res 13: 229–243
Nothnagel T, Straka P (2003) Inheritance and mapping of a yellow leaf mutant of carrot (*Daucus carota*). Plant Breed 122: 339–342
Nothnagel T, Ahne R, Straka P (2005) Morphology, inheritance and mapping of a compressed lamina mutant of carrot. Plant Breed 124: 481–486
Polidoros AN, Mylona PV, Arnholdt-Schmitt B (2009) Aox gene structure, transcript variation and expression in plants. Physiol Plant 137: 342–353
Polidoros NA, Mylona PV, Pasentsis K, Scandalios JG, Tsaftaris AS (2005) The maize alternative oxidase 1a (*Aox1a*) gene is regulated by signals related to oxidase stress. Redox Rep 10: 71–78
Popov VN, Simonian RA, Skulachev VP, Starkov AA (1997) Inhibition of the alternative oxidase stimulates

H_2O_2 production in plant mitochondria. FEBS Lett 415: 87–90
Purvis AC (1997) Role of the alterantive oxidase in limiting superoxide production by plant mitochondria. Physiol Plant 100: 164–170
Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 5: 94–100
Ren X-Y, Vorst O, Fiers MWEJ, Stiekema WJ, Nap J-P (2006) In plants, highly expressed genes are the least compact. Trends Genet 22: 528–532
Rodríguez-Trelles F, Tarrío R, Ayala FJ (2006) Origins and evolution of spliceosomal introns. Annu Rev Genet 40: 47–76
Rose AB (2002) Requirements for intron-mediated enhancement of gene expression in *Arabidopsis*. RNA 8: 1444–1453
Rostoks N, Mudie S, Cardle L, Russel J, Ramsay L, Booth A, Svensson JT, Wanamaker SI, Walia H, Rodriguez EM, Hedley PE, Liu H, Morris J, Close TJ, Marshall DF, Waugh R (2005) Genome-wide SNP discovery and linkage analysis in barley based on genes responsive to abiotic stress. Mol Gen Genomics 274: 515–527
Roy SW (2003) Recent evidence for the exon theory of genes. Genetica 188: 251–266
Roy SW, Gilbert W (2006) The evolution of spliceosomal introns: patterns, puzzles and progress. Nat Rev Genet 7: 211–221
Saisho D, Nambara E, Naito S, Tsutsumi N, Hirai A, Nakazono M (1997) Characterization of the gene family for alternative oxidase from *Arabidopsis thaliana*. Plant Mol Biol 35: 585–596
Saisho D, Nakazono M, Lee K-H, Tsutsumi N, Akita S, Hirai A (2001) The gene for alternative oxidase-2 (*AOX2*) from *Arabidopsis thaliana* consists of five exons unlike other *AOX* genes and is transcribed at an early stage during germination. Genes Genet Syst 76: 89–97
Schmid KJ, Sorensen TR, Stracke R, Torjek O, Altmann T, Mitchell-Olds T, Weisshaar B (2003) Large-scale identification and analysis of genome-wide single-nucleotide polymorphisms for mapping in *Arabidopsis thaliana*. Genome Res 13: 1250–1257
Shaposhnikov AS, Akopov SB, Chernov IP, Thomsen PD, Joergensen C, Collins AR, Frengen E, Nikolaev LG (2007) A map of nuclear matrix attachment regions within the breast cancer loss-of-heterozygosity region on human chromosome 16q22.1. Genomics 89: 354–361
Sharma A, Chauhan RS (2008) Identification of candidate gene-based markers (SNPs and SSRs) in the zinc and iron transporter sequences of maize (*Zea mays* L.). Curr Sci 95: 1051–1059
Shimada N, Nakatsuka T, Nakano Y, Kakizaki Y, Abe Y, Hikage T, Nishihara M (2009) Identification of gentian cultivars using SCAR markers based on intron-length polymorphisms of flavonoid biosynthetic genes. Sci Hortic 119: 292–296
Simon PW, Freeman RE, Vieira JV, Boiteux LS, Briard M, Nothnagel T, Michalik B, Kwon Y (2008) Carrot. In: Prohens J, Nuez F (eds) Handbook of Plant Breeding, Vol. 2. Springer, New York, pp 327–357
Slabaugh MB, Yu J-K, Tang S, Heesacker A, Hu X, Lu G, Bidney D, Han F, Knapp S (2003) Haplotyping and mapping a large cluster of downy mildew resistance gene candidates in sunflower using multilocus intron fragment length polymorphisms. Plant Biotech J 1: 167–185
Song J, Murdoch J, Gardiner SE, Young A, Jameson PE, Clements J (2008) Molecular markers and a sequence deletion in intron 2 of the putative partial homologue of LEAFY reveal geographical structure to genetic diversity in the acutely threatened legume genus *Clianthus*. Biol Conserv 141: 2041–2053
Sunkar R, Zhou X, Zheng Y, Zhang W, Zhu JK (2008) Identification of novel and candidate miRNAs in rice by high throughput sequencing. BMC Plant Biol 8: 1–17
Szücs P, Skinner JS, Karsai Ï, Cuesta-Marcos A, Haggard KG, Corey AE, Chen THH, Hayes PM (2006) Validation of the VRN-H2/VRN-H1 epistatic model in barley reveals that intron length variation in VRN-H1 may account for a continuum of vernalization sensitivity. Mol Gen Genomics 277: 249–261
Thornsberry J, Goodman MM, Doebley JF, Kresovich S, Nielsen D, Buckler E (2001) *Dwarf8* polymorphisms associate with variation in flowering time. Nat Genet 28: 286–289
Umbach AL, Gonzàlez-Meler MA, Sweet CR, Siedow JN (2002) Activation of the plant mitochondrial alternative oxidase: insights from site-directed mutagenesis. Biochim Biophys Acta 1554: 118–128
Valentine CR (1998) The association of nonsense codons with exon skipping. Mutat Res 411: 87–117
Van Ooijen JW, Voorrips RE (2001) JoinMap® 3.0 Software for the Calculation of Genetic Linkage Maps. Plant Research International, Wageningen
Velasco R, Zharkikh A, Troggio M, Cartwright DA, Cestaro A, Pruss D, Pindo M, Fitzgerald LM, Vezzulli S, Reid J, Malacarne G, Iliev D, Coppola G, Wardell B, Micheletti D, Macalma T, Facci M, Mitchell JT, Perazzolli M, Eldredge G, Gatto P, Oyzerski R, Moretto M, Gutin N, Stefanini M, Chen Y, Segala C, Davenport C, Dematte L, Mraz A, Battilana J, Stormo K, Costa F, Tao Q, Si-Ammour A, Harkins T, Lackey A, Perbost C, Taillon B, Stella A, Solovyev V, Fawcett JA, Sterck L, Vandepoele K, Grando SM, Toppo S, Moser C, Lanchbury J, Bogden R, Skolnick M, Sgaramella V, Bhatnagar SK, Fontana P, Gutin A, Van de Peer Y, Salamini F, Viola R (2007) A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. PLoS ONE 2: e1326

Vinogradov AE (2004) Compactness of human housekeeping genes: selection for economy or genomic design? Trends Genet 20: 248–253

Wang XJ, Reyes JL, Chua NH, Gaasterland T (2004) Prediction and identification of *Arabidopsis thaliana* microRNAs and their mRNA targets. Genome Biol 5: R65–R65.15

Wang X, Zhao X, Zhu J, Wu W (2005) Genome-wide investigation of intron length polymorphisms and their potential as molecular markers in rice (*Oryza sativa* L.). DNA Res 12: 417–427

Wydner KS, Sechler JL, Boyd CD, Passmore HC (1994) Use of an intron length polymorphism to localize the tropoelastin gene to chromosome 5 in a region of linkage conservation with human chromosome 7. Genomics 23: 125–131

Xie X, Lu J, Kullbokas EJ, Golub T, Mootha V, Lindblad-Toh K, Lander ES, Kellis M (2005) Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. Nature 434: 338–345

Yu J, Wang J, Lin W, Li S, Li H, Zhou J, Ni P, Dong W, Hu S, Zeng C, Zhang J, Zhang Y, Li R, Xu Z, Li S, Li X, Zheng H, Cong L, Lin L, Yin J, Geng J, Li G, Shi J, Liu J, Lv H, Li J, Wang J, Deng Y, Ran L, Shi X, Wang X, Wu Q, Li C, Ren X, Wang J, Wang X, Li D, Liu D, Zhang X, Ji Z, Zhao W, Sun Y, Zhang Z, Bao J, Han Y, Dong L, Ji J, Chen P, Wu S, Liu J, Xiao Y, Bu D, Tan J, Yang L, Ye C, Zhang J, Xu J, Zhou Y, Yu Y, Zhang B, Zhuang S, Wei H, Liu B, Lei M, Yu H, Li Y, Xu H, Wei S, He X, Fang L, Zhang Z, Zhang Y, Huang X, Su Z, Tong W, Li J, Tong Z, Li S, Ye J, Wang L, Fang L, Lei T, Chen C, Chen H, Xu Z, Li H, Huang H, Zhang F, Xu H, Li N, Zhao C, Li S, Dong L, Huang Y, Li L, Xi Y, Qi Q, Li W, Zhang B, Hu W, Zhang Y, Tian X, Jiao Y, Liang X, Jin J, Gao L, Zheng W, Hao B, Liu S, Wang W, Yuan L, Cao M, McDermott J, Samudrala R, Wang J, Wong GK, Yang H (2005) The genomes of *Oryza sativa*: a history of duplications. PLoS Biol 3: e38

Zhang Z, Gerstein M (2003) Patterns of nucleotide substitution, insertion and deletion in the human genome inferred from pseudogenes. Nucleic Acids Res 31: 5338–5348

Zhang BH, Pan XP, Wang QL, Cobb GP, Anderson TA (2005) Identification and characterization of new plant microRNAs using EST analysis. Cell Res 15: 336–360

Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31: 3406–3415

Edited by V. Hurry