Physiologia Plantarum
An International Journal for Plant Biology

Physiologia Plantarum 137: 520–531. 2009

Intron polymorphism pattern in *AOX1b* of wild St John's wort (*Hypericum perforatum*) allows discrimination between individual plants

Alexandre Oliveira Ferreira[a], Hélia Guerra Cardoso[a], Elisete Santos Macedo[a], Diego Breviario[b] and Birgit Arnholdt-Schmitt[a,*]

[a]EU Marie Curie Chair, ICAAM, University of Évora, Apartado 94, 7002-554 Évora, Portugal
[b]Istituto di Biologia e Biotecnologia Agraria-CNR, Via Bassini 15, 20133 Milano, Italy

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

Received 16 June 2009;
revised 7 September 2009

doi:10.1111/j.1399-3054.2009.01291.x

The present paper deals with the analysis of natural polymorphism in a selected alternative oxidase (*AOX*) gene of the medicinal plant, St John's wort. Four partial *AOX* gene sequences were isolated from the genomic DNA of a wild plant of *Hypericum perforatum* L. Three genes belong to the subfamily *AOX1* (*HpAOX1a, b* and *c*) and one to the subfamily *AOX2* (*HpAOX2*). The partial sequence of *HpAOX1b* showed polymerase chain reaction (PCR) fragment size variation as a result of variable lengths in two introns. PCR performed by Exon Primed Intron Crossing (EPIC)-PCR displayed the same two-band pattern in six plants from a collection. Both fragments showed identical sequences for all exons. However, each of the two introns showed an insertion/deletion (InDel) in identical positions for all plants that counted for the difference in the two fragment sizes. The InDel in intron 1 influenced the predictability of a pre-microRNA site. The almost identical PCR fragment pattern was characterized by a high variability in the sequences. The InDels in both introns were linked to repetitive intron single nucleotide polymorphisms (ISNP)s. The polymorphic pattern obtained by InDels and ISNPs from both fragments together was appropriate to discriminate between all individual plants. We suggest that *AOX* sequence polymorphism in *H. perforatum* can be used for studies on gene diversity and biodiversity. Further, we conclude that *AOX* sequence polymorphism of individual plants should be considered in biological studies on AOX activity to exclude the influence of genetic diversity. The identified polymorphic fragments are available to be explored in future experiments as a potential source for functional marker development related to the characterization of origins/accessions and agronomic traits such as plant growth, development and yield stability.

Introduction

The nucleotide sequence of spliceosomal introns is not neutrally evolving, as originally thought. Evidences for the existence of selective constraints that reduce evolutionary divergence have been reported. The mutational rate that affects intron evolution is higher than that affecting exon sequences (Wang et al. 2005).

Abbreviations – AOX, alternative oxidase; EPIC, exon primed intron crossing; ILP, intron length polymorphism; InDel, insertion/deletion; ISNP, intron single nucleotide polymorphism; NCBI; National Center for Biotechnology Information; PCR, polymerase chain reaction; S/MAR, scaffold or matrix attachment region; SNP, single nucleotide polymorphism; ORF, Open Reading Frame.

Thus, evolution has worked differently on two types of sequences of the same gene (Bardini et al. 2004). Intron loss and gain are considered by several authors as the main driving force for intron evolution (Blake 1978, Knowles and McLysaght 2006, Roy 2003). The more relaxed evolution of introns has been attributed to the functional role that they can play with regard to gene regulation (Fiume et al. 2004, Gianì et al. 2003, Rose 2002).

How introns exactly affect gene expression is a matter of intense investigation that is changing our current view of gene definition (Encode Project Consortium 2007). Polymorphisms in intron sequences are expected to play a critical role in gene regulation because of their potential to influence the binding of transcription factors (Xie et al. 2005), the process of alternative splicing (Baek et al. 2008, Ner-Gaon et al. 2007, Noh et al. 2006), the maintenance of those structural features that are required to support intron-located promoter activities and intron-mediated transcription of micro or small nucleolar RNAs (Li et al. 2007, Louro et al. 2007, Nakaya et al. 2007) as well as nonsense-mediated mRNA decay (Jaillon et al. 2008). Introns can affect either the level or the site of gene expression through intron-mediated enhancement of gene expression and intron-dependent spatial expression, respectively. It is often observed that introns that are most proximal to the 5′ end of a gene are the ones that exert a more pronounced effect on gene expression (Breviario et al. 2008, Rose et al. 2008). Chromosomal structures and chromatin modeling supporting transcription and recombination activities can also rely on motifs located in introns, such as scaffold or matrix attachment regions (S/MARs) (Shaposhnikov et al. 2007). In silico genome-wide screening done in *Arabidopsis thaliana* has indicated that most of the intragenic S/MARs are present within introns, especially at the most proximal 5′ end of the genes (Rudd et al. 2004). Intronic S/MARs were related to lower rates of differential expression (Tetko et al. 2006).

Variations between DNA sequences are mostly due to mutations related with insertion/deletions (InDels) and single nucleotide polymorphisms (SNPs) (Nasu et al. 2002). They can contribute directly to characterize a genotype and can be associated with a phenotype through linkage disequilibrium analyses (Risch and Merikangas 1996). Because of their abundance and amenability to high-throughput screening, InDels and SNPs are the markers of choice for genotyping and mapping (Bi et al. 2006). In medicinal research, several reports show a linkage between intron single nucleotide polymorphisms (ISNPs) and phenotype changes related to human diseases (e.g. Hiratani et al. 2005, Jou et al. 2009).

While maintaining conserved positions throughout the genome, introns can vary in length. Cardoso et al. (2009) report allelic and genotypic differences in intron length for *Daucus carota* L. Further, (alternative oxidase) *AOX* genes from different species show intron length variability (Cardoso et al. 2009). Intron length and numbers can be crucial for plant gene expression as can be seen by the analysis of massively parallel sequencing signatures and microarray experiments (Ren et al. 2006, see also Costa et al. 2009). Investigating intron length polymorphisms (ILPs) is a fast, reliable and convenient method to assess genetic variability in search of associated agronomic traits. ILP was, for example, used as functional marker for biotic stress tolerance (Slabaugh et al. 2003) and sensitivity to vernalization (Szücs et al. 2006). Recently, a genome-wide investigation has been done in *Oryza sativa* L. to screen for ILPs as a potential source of molecular markers (Wang et al. 2005). Ten different rice accessions were characterized identifying 173 co-dominant ILP markers. Seventy-one percent of these markers exhibited subspecies specificity. Moreover, a high fraction of the rice ILP markers was shown to be transferable to other plants (Wang et al. 2005). ILPs were successfully applied to assess genetic relationship among varieties using two introns of the beta-tubulin gene family (Breviario et al. 2007). In addition, ILP-based markers based on the first or third intron of diverse flavonoid biosynthetic genes could be used to distinguish gentian cultivars for plant protection (Shimada et al. 2009). A short InDel of 7 bp in intron 2 of a homologue of the floral meristem identity gene *LEAFY* revealed a link to the geographical origin of *Clianthus* populations (Song et al. 2008). Intron size variation was studied also in other organisms. For example, ILPs were applied to characterize *Mytilus* spp. populations and hybridization events between *Mytilus edulis* and *Mytilus galloprovincialis* (Daguin et al. 2001) and were recommended for taxonomic studies in the fish species *Serrasalmus* (Hubert et al. 2006).

AOX is an inner mitochondrial membrane protein codified in plants by a small multigene family with at least five genes (Clifton et al. 2006) belonging to two subfamilies: *AOX1*-type and *AOX2*-type. The genomic organization of plant *AOX* genes is characterized by conserved intron positions. Commonly, four exons are interrupted by three introns (Considine et al. 2002, Li et al. 2008, Polidoros et al. 2005). Several exceptions from this rule were referred to intron loss and gains (Considine et al. 2002, Ito et al. 1997, Saisho et al. 2001). Studying the occurrence of intron length and sequence polymorphism in members of *AOX*, an enzyme involved

in stress acclimation and adaptation (McDonald and Vanlerberghe 2006, Plaxton and Podestá 2006, Umbach et al. 2005), can be relevant for a better understanding of adaptation at evolutionary scale (Frederico et al. 2009, McDonald 2008) to study biodiversity and ecology, and for promoting breeding programs on stress tolerance (Arnholdt-Schmitt et al. 2006).

In this paper, we report about repetitive and cross-linked intron length and sequence polymorphism in *HpAOX1b*, a gene of the *AOX* multigene family of *H. perforatum*. We document that the identified polymorphic pattern is appropriate to discriminate individual plants of *H. perforatum*.

Materials and methods

Plant material

Seeds and shoots from wild *H. perforatum* plants were collected from various accessions of Portugal. For the present study, plant material from six accessions located in the region of Alentejo [Mora (M), Aljustrel (A), Beja (B) and Évora (E)], and in the region of Estremadura [Torres Vedras (T) and Alenquer (Q)] were included.

After disinfection by washing in a 70% (v/v) ethanol solution for 2 min and with a calcium hypocloride solution (10% activated chlorine, w/v) containing Tween 20 (0.1% v/v) for 20 min, and finally rinsing three times with sterilized distilled water, the seeds were aseptically inoculated in pots containing basal medium of Murashige and Skoog (1962), supplemented with 2% sucrose and solidified with 0.7% (w/v) agar (Invitrogen Life technologies, UK). The pH of the medium was adjusted to 5.75 prior to autoclaving (121°C, 1 kg cm^{-2} for 15 min). Cultures were kept at 25°C under a 16 h photoperiod, and 34 µmol $m^{-2}s^{-1}$ of light intensity provided by day-light Philips fluorescent lamps. Plantlets of each genotype were propagated by culture of micro-shoots in fresh culture medium with the same basal composition. The cultures were maintained under the same physical conditions. Subcultivation was repeated monthly under the same conditions.

Plants of accession T and Q had been propagated by rooted shoot cuttings under greenhouse conditions. T can be considered as an outgroup because plants are phenotypically different in comparison to all other plants of the collection; the growth type of the shoots during spring and summer is more compact and the leaves are bigger.

DNA extraction

Young leaves from a selected plant of each accession were collected for DNA extraction using the DNEasy plant mini kit (Qiagen, Hilden, Germany). Genomic DNA integrity was analyzed by electrophoresis in 1% agarose gel (Invitrogen Life technologies, UK) after staining in ethidium bromide (Merck, Germany) (0.2 µg ml^{-1}). DNA visualization was made with the Gene Flash Bio Imaging system (Syngene, Cambridge, UK) and the quantification was performed in comparison of defined amounts of lambda (λ) DNA (Fermentas, EU) as a standard with the help of the software GeneTools (SynGene, Cambridge, UK). For PCR, a DNA work solution of 10 ng $µl^{-1}$ DNA was prepared of each sample.

Alternative oxidase gene isolation

PCR fragments were amplified in a 2720 thermocycler (Applied Biosystems, Singapore) by using two degenerated primer pairs, P1/P2 previously reported by Saisho et al. (1997) that amplifies a site in the conserved region of exon 3; and 41AOX1Fw (5′-TG GAVRTGGAHYTGHTTYAG-3′)/44AOXRev (5′-GTCAT BARRTGCATBCKYTC-3′), designed to amplify intronic *AOX1* gene sequences also. PCR was carried out with TaqDNA polymerase (Fermentas, EU) using 10 ng of DNA of accession M as template. PCR was conducted with an initial step of 5 min at 94°C followed by 35 cycles, each consisting of 1 min at 94°C for DNA denaturation, 2 min at 50°C for primers' annealing and 2 min at 72°C for DNA synthesis. Final extension was performed by 5 min at 72°C. PCR fragments were analyzed in 1.4% agarose (Invitrogen Life technologies, UK) gel according to the procedure described under the section DNA extraction.

Fragments generated by PCR were purified using the GFX PCR DNA and Gel Band Purification Kit (GE Healthcare, Little Chalfont, England), cloned into a pGEM®-T Easy System I vector (Promega, Madison, WI, USA) and used for transformation of *E. coli* competent cells JM109 (Promega, Madison, WI, USA). Plasmidic DNA of putative recombinant clones was extracted according to the alkaline lyses protocol (Birnboim and Doly, 1979) and analyzed by the restriction enzymes *Eco*RI, *Hpy*F3I and *Alu*I (Fermentas, Ontario, Canada). Selected plasmid DNA was completely sequenced (Macrogen company: http://www.macrogen.com) in the directions of sense and antisense strands using specific primers of the cloning vector (T7 and SP6, Promega, Madison, WI, USA).

Research on sequence homologies was performed in the National Center for Biotechnology Information (NCBI) database using BLAST algorithm (Karlin and Altschul 1993) (http://www.ncbi.nlm.nih.gov) (BLASTx and BLASTn). Intron location was made using the software Spidey, which is publicly available at

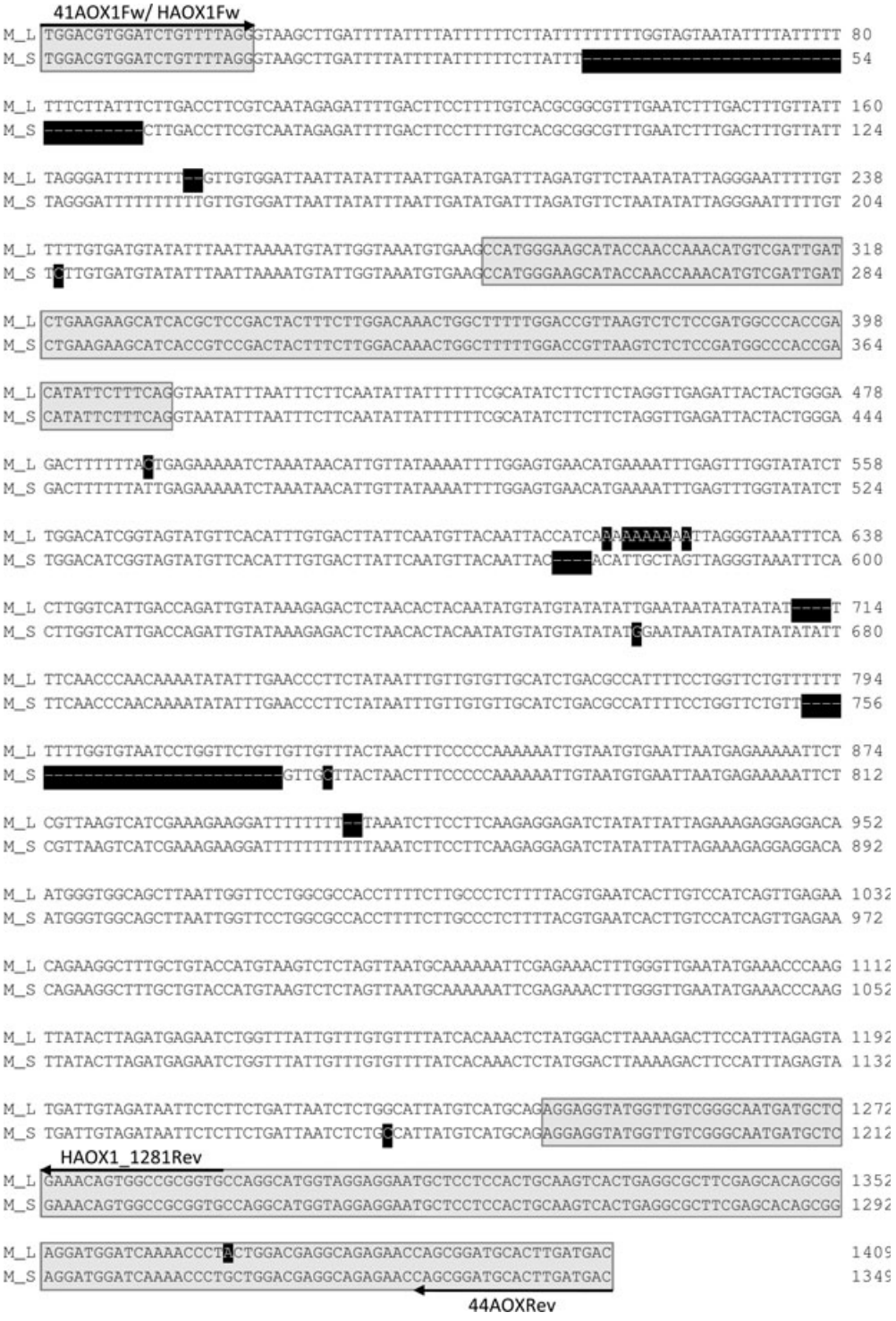

Fig. 1. Alignment of the sequences correspondent to the two PCR fragments of *HpAOX1b*. The fragments were isolated by the primer pair 41AOX1Fw/44AOXRev from the single plant of accession M. Nucleotide differences are highlighted on a black background. The deletions are marked by a minus sign. Gray boxes mark the partial sequence of exon 1, complete exon 2 and partial exon 3. The locations of primers used for gene isolation and EPIC-PCR analysis are indicated by arrows.

http://www.ncbi.nlm.nih.gov/IEB/Research/Ostell/Spidey/. For deduction of peptide sequences, the software EditSeq (Lasergene 7, DNASTAR, Madison, WI, USA) was used. The alignments among the peptide sequences of *H. perforatum* and homologous sequences from the NCBI databases were performed using the ClustalW method from the BioEdit software (Hall 1999).

Analysis of intron polymorphisms by EPIC-PCR

In order to analyze intron polymorphisms, specific primers were designed in the exons flanking both intron 1 and intron 2 regions. The forward primer was located in exon 1 (HAOX1Fw: 5′-TGGACGTGGATCTGTTTTAGG-3′, MWG, Germany) and the reverse primer in exon 3 (HAOX1_1281Rev: 5′-CACCGCGGCCACTGTTTC-3′, MWG, Germany). EPIC-PCRs were carried out using the Taq DNA polymerase (Fermentas, Ontario, Canada) and 10 ng of DNA from a single plant of each accession as template. An initial step of 5 min at 94°C was followed by 35 cycles, each consisting of 1 min at 94°C, 2 min at 53°C and 90 s at 72°C. Final extension was performed by 5 min at 72°C. PCR fragments were analyzed as described in the section on Alternative oxidase gene isolation.

DNA cloning and sequencing were performed as described in the section Alternative oxidase gene isolation. Selection of putative recombinant clones was made by the restriction enzyme *Eco*RI (Fermentas, Ontario, Canada).

Computational prediction and validation of miRNA percursors in intron sequences

Putative miRNA precursors were searched in the two intron sequences of the *HpAOX1b* gene from six accessions of *H. perforatum* (in total 17 sequences of intron 1 and 17 sequences of intron 2, corresponding to 9 sequences of small fragments and 8 sequences of large fragments) by using the software miR-*abela*, which is publicly available at http://www.mirz.unibas.ch/cgi/pred_miRNA_genes.cgi. For validation of potential pre-miRNAs the software MiPred was applied, which is publicly available at http://www.bioinf.seu.edu.cn/miRNA/ (Jiang et al. 2008).

Prediction of the secondary structure of pre-miRNA was run on the web-based software MFOLD 3.1, which is available at http://mfold.bioinfo.rpi.edu/cgi-bin/rnaform1.cgi (Mathews et al. 1999, Zuker 2003).

Results

Four partial *AOX* gene sequences were isolated from the genomic DNA of an individual *H. perforatum* plant of accession M. Three gene sequences could be classified as members of the subfamily *AOX1* (*HpAOX1a*: EU330414, *HpAOX1b*: EU330415 and *HpAOX1c*: EU330416) and one as belonging to the subfamily *AOX2* (*HpAOX2*: EU330413). The partial sequence of *HpAOX1b* contained two partial exons and two complete introns with a complete exon sequence of 129 bp included. Campos et al. (2009) have pointed out that exon 2 was present in various species and gene subfamilies of 129 bp. In addition, the terminal partial exon sequence in *HpAOX1b* shows homology to exon 3 from *A. thaliana* (D89875). Consequently, the complete exon was named exon 2 and the two introns classified as intron 1 and intron 2.

Two PCR fragments were isolated from *HpAOX1b* with primer pair 41AOX1Fw/44AOXRev from genomic DNA with a size of 1408 bp (M_L, GQ266688) and 1349 bp (M_S, GQ266687) (Fig. 1). Sequencing and alignment revealed that the fragments contained two complete introns with a size of 227 bp in intron 1 and 805 bp in intron 2 of the small fragment (M_S), and 260 bp in intron 1 and 831 bp in intron 2 of the large fragment (M_L) (Fig. 1). The observed fragment length variability was a result of InDel events in both introns.

Following this observation, *HpAOX1b* was selected to initiate gene diversity studies from a small number of six selected plants from a wild collection established for future experiments in a field. The selected plants were tetraploid (data not shown). Fig. 2 shows the result of EPIC-PCR analysis with specific primers that covered both intron regions. All six plants revealed the two-band pattern. The larger fragment consisted of around 1300 bp and the smaller fragment of around 1225 bp.

All isolated large and small PCR fragments from the six plants included two partial exon sequences, exon 1 (22 bp) and exon 3 (148 bp), and the complete exon 2 sequence (129 bp). Together, this corresponds to a small partial Open Reading Frame (ORF) of 299 bp. All isolated sequences also included the two complete intron sequences, intron 1 and intron 2.

The exon regions in *HpAOX1b* demonstrated strong sequence conservation. No variability was observed in

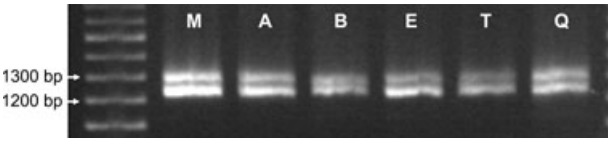

Fig. 2. Agarose gel showing the results of the EPIC-PCR for six individual *H. perforatum* plants from six accessions in Southern Portugal. The letters indicate each accession: (M) Mora, (A) Aljustrel, (B) Beja, (Q) Alenquer, (T) Torres Vedras and (E) Évora.

exon regions between large and small fragments from individual plants or in any large or small fragment of the six plants (not shown). In all sequences, the typical characteristics of the nuclear exon–intron junctures 5′-GU/AG-3′ (Saisho et al. 1997) were observed. The 5′ splice site is 5′-GU and the 3′ splice site is AG-3′ at the extreme ends of the introns. The intronic regions presented higher U contents than the exons. Intron 1 had 50.5–53.5% U and intron 2 had around 38%. The partial ORF consisted of ca. 22% U. The higher U content in introns is in agreement with the observation of other authors. U richness is suggested to play an important role for intron recognition (Deutsch and Long 1999, Ko et al. 1998).

Fig. 3 gives a summary of all polymorphisms identified in the two introns. All different sequences that were identified from the six plants are included. The upper part in Fig. 3 shows the sequences of all large fragments, whereas the sequences from the small fragments are grouped in the lower part. It can be seen that the ILPs were a result of InDels that occurred simultaneously in both introns in positions conserved among all plant genotypes. In all cases, an InDel of 35–36 bp in intron 1 (positions 55–90) was linked to an InDel of 26–27 bp in intron 2 (positions 817–844).

Besides the repetitive length polymorphism through the large InDels, a high number of repetitive or unique ISNPs were observed. The identified ISNPs can be grouped as follows:

1. single nucleotides present in all small fragments, but not in the large fragments,
2. single nucleotides characteristic for all large fragments, but not found in the short fragments,
3. single nucleotides observed only in small fragments, but not in all of them,
4. groups of ISNPs in the large fragment and
5. plant-specific polymorphisms
   (A) insertion of a 10× repetition of the dinucleotide AT in the small fragment (plant/accession B) and
   (B) single nucleotides typically observed in one or the other fragment recombined in the large fragment (T1 of plant/accession T).

Conserved ISNPs that were characteristic of smaller or larger fragments [points (1) and (2) respectively] can be found in Fig. 3 in eight nucleotide positions (not considering T1, as it marks a special case described below): 175 (T/-), 625 (G/A), 700 (G/T), 725 (A/T), 728 (T/-), 925 (T/C), 1220 (T/G) and 1256 (C/G). Repetitive ISNPs only in smaller fragments [point (3)] can be found in positions 99, 174, 244, 727, 729, 730, 931 and 932. This type of ISNP was observed across various locations, including both the regions, Alentejo and Estremadura. Polymorphic groups of sequences [point (4)] can be found between positions 491 and 674 bp in intron 2. The pattern is characterized through the InDel CATC, followed by an ISNP through replacing C by A and substituting the sequence TTGCT by AAAAA. This sequence pattern characterizes sequences from individual plants of the locations M, B and E in the Alentejo region.

The most pronounced plant-specific polymorphism [point (5)] was the 10-fold repetition of the dinucleotide

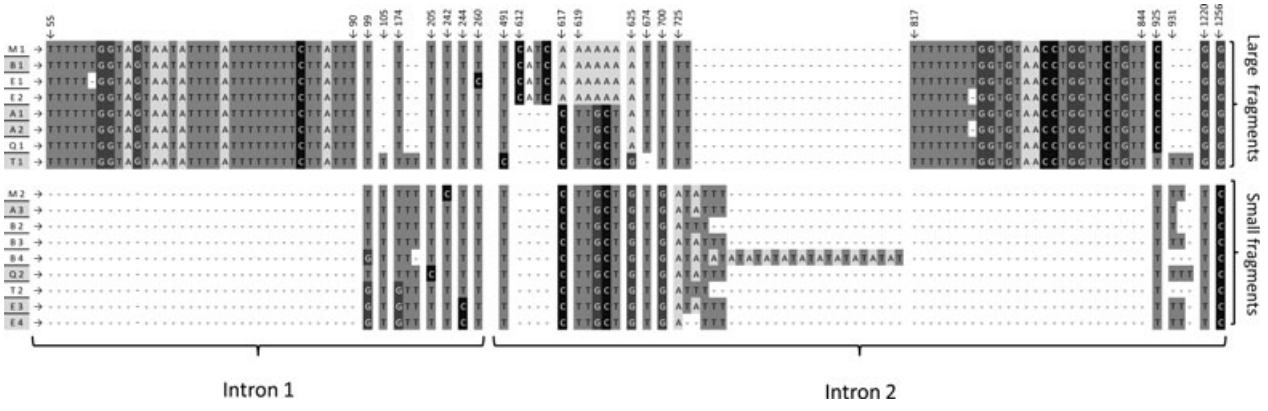

Fig. 3. Polymorphic sequences in intron 1 and intron 2 of *HpAOX1b* sequences identified for six individual *H. perforatum* plants provided from six accessions in Southern Portugal. The letters indicate each accession: (M) Mora, (A) Aljustrel, (B) Beja, (Q) Alenquer, (T) Torres Vedras and (E) Évora (see left side). The number added following the letter is referring to different sequences, e.g. E4 means that four different sequences were identified in the plant from Évora. Large and small fragments (see right side) are referring to the PCR products obtained by EPIC-PCR (see Fig. 2). Positions of the nucleotides in the sequence alignment are given on top of the figure and are related to the alignment of the complete sequence fragments (including exons and introns).

AT inserted in position 731 in only one of the four discovered sequences of plant B (B4). This insertion marks plant B easily as unique.

The plant of accession T demonstrated a small fragment (T2) that integrates well in the typical sequence pattern of the smaller fragments. However, the large fragment of the same plant characterized by the sequence T1 shows a unique pattern of polymorphism among the six plants. Basically, it contains the typical sequence pattern of larger fragments, but also contains nucleotide substitutions, which are characteristic of the smaller fragment, such as in positions 105 (T/-), 175 (T/-), 176 (T/-), 625 (G/A), 931 (T/-) and 932 (T/-). In addition, this sequence shows two unique ISNPs in positions 491 (C/T) and 674 (-/T).

Considering InDels and ISNPs in both fragments of the partial sequence of *HpAOX1b*, all plants could be discriminated from each other. The maximal number of different sequences that was discovered in individual plants has been four. This can be seen by the sequence pattern in Fig. 3 of two large and two small fragments of the individual plant from accession E (E1, E2 and E3, E4) and one large and three small fragments of the plant from accession B (B1 and B2, B3, B4). The observation is in agreement with the tetraploidy status determined for all plants.

Intron 1 and intron 2 sequences were searched for predictable pre-miRNA sites. In all intron 2 sequences from the six plants, a pre-miRNA could be predicted with a probability around 75%. This was independent on any insertion, deletion or substitutions of single nucleotides (data not shown). However, the result was different for intron 1 (Table 1). In all small fragments, pre-miRNA sites could be predicted in intron 1 with a probability between 61.9 and 67.8%. In contrary, prediction of a pre-miRNA in large fragments failed in all but one sequence. Prediction of a pre-miRNA in intron 1 in the large fragments was possible only in the sequence E1 with a probability of 68.9%. E1 shows in the inserted region of intron 1 two SNPs (60: -/T; 260: C/T), which marks this sequence as unique (Fig. 3).

Discussion

Novel isolation of four *AOX* genes from *H. perforatum* confirms the presence of both *AOX* gene families, *AOX1* and *AOX2*, in this plant species. Three genes were identified from the *AOX1* subfamily and one gene could be classified as belonging to *AOX2*. In this paper, gene diversity of the partial sequence of *HpAOX1b* is described in six plants from diverse accessions.

The partial sequence of *HpAOX1b* includes two complete introns. Introns are generally short in plants (Ko et al. 1998, Lorkovic et al. 2000, Reddy 2001) with less than 200 bp (Luehrsen and Walbot 1994). In contrast to the conserved size of *AOX* exon sequences (Campos et al. 2009), intron sizes of *AOX* genes vary highly within the *AOX* gene family of a species and seem to characterize the genes of a species (Cardoso et al. 2009). From the *AOX* genes available at NCBI databases, the smallest intron size was detected for *AOX1a* (75 bp) in *Solanum tuberosum* (DQ270421) and the larger intron was found for *Vitis vinifera AOX2* (6885 bp) (CAAP01005007) (see Costa et al. 2009). In *HpAOX1b*, intron 1 has a size of 226–263 bp and is significantly smaller than intron 2 with 801–831 bp.

Both *HpAOX1b* introns demonstrated high sequence diversity. This was a result of InDels and ISNPs. The variability seen for the introns is in high contrast to the strong sequence conservation of the included ORF of 299 bp. This observation is in line with the current view of a more relaxed evolution of intron sequences in comparison to exons and suggests strong evolutionary constraints to maintain the functionality of the AOX protein-coding sequence (Wang et al. 2005). AOX plays an important role in alternative respiration in mitochondria and is in plants involved in all types of stress reactions (McDonald and Vanlerberghe 2006, Plaxton and Podestá 2006), including morphogenic responses, such as adaptive growth and development (Campos et al. 2009, Fiorani et al. 2005, Frederico et al. 2009, Ho et al. 2007). On the other side, it may be suggested that the higher diversity in *AOX* intron sequences can point to the role of introns in providing more genetic flexibility to AOX regulation.

Intron loss and gain were considered as the main driving force in gene evolution (Coulombe-Huntington and Majewski 2007, Knowles and McLysaght 2006, Roy 2003). The most pronounced polymorphisms in both *HpAOX1b* intron sequences were the large InDels. Strikingly, the two InDels in intron 1 and intron 2 were conserved from their positions and related to the link to each other. In addition, smaller events were also repetitive in the genomes and linked to the InDels, such as single nucleotide substitutions or groups of nucleotide substitutions.

Intron 1 presented an InDel of 35–36 bp in comparison to the InDel of 26–27 bp in intron 2. Thus, in view of the smaller size of intron 1, the relative effect of the InDel on the structure and functionality of intron 1 can be expected to be much stronger than that of intron 2. The insertion in intron 1 influenced the predictability of a pre-miRNA site. In all small fragments, the site was identified (Table 1). However, in the majority of the larger intron 1 sequences, this was not possible although the identified pre-miRNA site is not located

Table 1. Computational prediction of intronic miRNA precursors in *HpAOX1b* of plants from six accessions in Portugal: Mora (M), Aljustrel (A), Beja (B), Évora (E), Torres Vedras (T) and Alenquer (Q). The number following the letter of each accession corresponds to the sequences identified (see Fig. 3). bp, length of the pre-miRNA sequence in bp; MFE, minimal free energy in kcal mol^{-1}; Prob, probability to be a real pre-miRNA sequence in percentage.

	Genotypes	Pre-miRNA sequences in intron 1 and secondary structure	bp	MFE	Prob
Small fragments	*A3* *B2, B3, B4* *T2*	UUAUAUUUAAUUGAUAUGAUUUAGAUGUUCUAAUAUAUUAGGGAAUU UUUGUUUUUGUGAUGUAUAUUUAAUUAAAAUGUAUUGGUAAAUGUGA 10 20 30 40 | A UG G A GUUCUA G UUU UUAUAUUUA U AUAU AUUU GAU AUAUAUUA GGAA \ AGUGUAAAU G UAUG UAAA UUA UAUGUAGU UUUU U ^ - GU - A AUUUA- G UGU 90 80 70 60 50	94	-16.60	61.9
	E3, E4	UUAUAUUUAAUUGAUAUGAUUUAGAUGUUCUAAUAUAUUAGGGAAUU UUUGUUUUCGUGAUGUAUAUUUAAUUAAAAUGUAUUGGUAAAUGUGA 10 20 30 40 | A UG G A GUUCUA G UUU UUAUAUUUA U AUAU AUUU GAU AUAUAUUA GGAA \ AGUGUAAAU G UAUG UAAA UUA UAUGUAGU CUUU U ^ - GU - A AUUUA- G UGU 90 80 70 60 50	94	-18.2	62.3
	M2	UUAUAUUUAAUUGAUAUGAUUUAGAUGUUCUAAUAUAUUAGGGAAUU UUUGUU**C**UUGUGAUGUAUAUUUAAUUAAAAUGUAUUGGUAAAUGUGA 10 20 30 40 | A UG G A GUUCUA - U UUAUAUUUA U AUAU AUUU GAU AUAUAUUA GGGAAU U AGUGUAAAU G UAUG UAAA UUA UAUGUAGU UUCUUG U ^ - GU - A AUUUA- G U 90 80 70 60 50	94	-17.6	64.9
	Q2	UUAUAUUUAAUUGAUA**C**GAUUUAGAUGUUCUAAUAUAUUAGGGAAUU UUUGUUUUUGUGAUGUAUAUUUAAUUAAAAUGUAUUGGUAAAUGUGA 10 20 30 40 | A UG A A UUCUA G UUU UUAUAUUUA U AUACG UUUAG UG AUAUAUUA GGAA \ AGUGUAAAU G UAUGU AAAUU AU UAUGUAGU UUUU U ^ - GU A A UUA-- G UGU 90 80 70 60 50	94	-18.7	67.8
Large fragments	*E1*	UUAUAUUUAAUUGAUAUGAUUUAGAUGUUCUAAUAUAUUAGGGAAUU UUUGUUUUUGUGAUGUAUAUUUAACUAAAAUGUAUUGGUAAAUGUGA 10 20 30 40 | A UG A A UUCUA G UUU UUAUAUUUA U AUAUG UUUAG UG AUAUAUUA GGAA \ AGUGUAAAU G UAUGU AAAUC AU UAUGUAGU UUUU U ^ - GU A A UUA-- G UGU 90 80 70 60 50	94	-18.3	68.9

within the insertion. Pre-miRNAs are determined by their hairpin-forming structure. The method for prediction of a pre-miRNA also takes into account the neighbor sequences given by the intron. Prediction of the site was possible in the presence of two additional ISNPs in the inserted region, as shown in the sequence E1 (Fig. 3, Table 1). Thus, insertion of the sequence (InDel) alone was not sufficient to exclude prediction of the

site. Prediction of a pre-miRNA site in intron 2 was confirmed in all small and large fragments from all plants. Pre-miRNA sites are the sites for microRNA synthesis and thus important for the regulation of target genes. Whether the prediction of pre-miRNA in *HpAOX1b* and the suggestion that polymorphisms can interfere with miRNA synthesis in intron 1 have practical impact need to be validated by 'wet' experimentation. Validation would give support to the current view that 5′ located introns should be more important for gene regulation than introns more directed to the 3′end of the gene (Breviario et al. 2008, Rose et al. 2008, Rudd et al. 2004, Tetko et al. 2006).

The *H. perforatum* plants were all tetraploid. This is the predominant ploidy found for this plant in most parts of the world (Barcaccia et al. 2006, Mártonfi et al. 1996, Pank et al. 2003). The maximal number of different *HpAOX1b* sequences found in this study from one plant was in fact four (Fig. 3: E1–E4, B1–B4). Cross-fertilization is possible in this species (Pank et al. 2003), although the predominant mode of propagation occurs via apomixis (Matzk et al. 2001, Noack 1939). Consequently, in natural populations, a high number of identical genotypes can be found in one accession. Besides, clonal diversity through clearly distinct genotypes and recombined genotypes can be found (Arnholdt-Schmitt 2000, Barcaccia et al. 2006, Pank et al. 2003). Plant B revealed four different partial *HpAOX1b* sequences (Fig. 3: B1–B4). One of these demonstrated a unique sequence by a 10-fold repetition of AT in intron 2. This sequence marks the plant. Plant T shows a unique pattern in the *HpAOX1b* gene. This was a result of a recombination of single nucleotides seen only in the large fragment (T1) caused by intron insertions. Here, single nucleotides that were typically found in the large fragment were recombined with single nucleotides otherwise only observed in introns from small fragments (Fig. 3).

It can be concluded that the polymorphic pattern observed in two introns of the partial *HpAOX1b* gene sequence was sufficient to distinguish six individual plants from *H. perforatum*. Similar results are reported also from other species, such as carrot (Cardoso et al. 2009) and grapevine (Costa et al. 2009). Thus, screening of *AOX* gene sequences for intron polymorphisms to study gene diversity and biodiversity seems to be reasonable. It can be supposed that the identified polymorphisms can have influence on *AOX* gene regulation that might affect phenotypes. Studies are in progress to validate these assumptions. Gene diversity in *AOX* genes among individual plants should be excluded in biological studies that aim to understand the role of AOX activity.

Acknowledgements – This work was supported by the European Commission through providing a Marie Curie Chair to Birgit Arnholdt-Schmitt and by FCT providing scholarships to Alexandre Oliveira Ferreira (SFRH/BI/15993/2006), Elisete Macedo (SFRH/ BD/22061/2005) and Hélia Cardoso (SFRH/BPD/27016/2006). We would like to thank Hélio Costa (UFC, Fortaleza, Brazil) for the design of the degenerated primers used for *HpAOX* genes' isolation. We are grateful to Timothy Sharbel, IPK-Gatersleben, Germany, and Otto Schrader, JKI-Quedlinburg, Germany, for their collaboration and also for determining the ploidy level of all *H. perforatum* plants. We would also like to appreciate the support of Manuel Maria Tavares de Sousa (INRB/L-INIA, Elvas, Portugal) for the establishment of the field collection in Elvas.

References

Arnholdt-Schmitt B (2000) RAPD analysis: a method to investigate aspects of the reproduction biology of *Hypericum perforatum* L. Theor Appl Genet 100: 906–911

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

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

Barcaccia G, Arzenton F, Sharbel TF, Varotto S, Parrini P, Lucchin M (2006) Genetic diversity and reproductive biology in ecotypes of the facultative apomict *Hypericum perforatum* L. Heredity 96: 322–334

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

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 7: 1513–1523

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

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 intronicas. 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

Cardoso HG, Campos MD, Costa AR, Campos MC, Nothnagel T, Arnholdt-Schmitt B (2009) Carrot alternative oxidase gene *AOX2a* demonstrates allelic and genotypic polymorphisms in intron 3. Physiol Plant 137: 592–608

Clifton R, Millar AH, Whelan J (2006) Alternative oxidase 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

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

Coulombe-Huntington J, Majewski J (2007) Characterization of intron loss events in mammals. Genome Research 17: 23–32

Daguin C, Bonhomme F, Borsa P (2001) The zone of sympatry and hybridization of *Mytilus edulis* and *M. galloprovincialis,* as described by intron length polymorphism at locus mac-1. Heredity 86: 342–354

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

Encode Project Consortium (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447: 799–816

Fiorani F, Umbach AL, Siedow JN (2005) The alternative oxidase of plant mitochondria is involved in the acclimation of shoot growth at low temperature. A study of *Arabidopsis AOX1a* transgenic plants. Plant Physiol 139: 1795–1805

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

Frederico AM, Campos MD, Cardoso HG, Imani J, Arnholdt-Schmitt B (2009) Alternative oxidase involvement in *Daucus carota* somatic embryogenesis. Physiol Plant 137: 498–508

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

Hall TA (1999) BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT. Nucleic Acids Symp Ser 41: 95–98

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 Endocrinol Metab 90: 2898–2903

Ho LH, Giraud E, Lister R, Thirkettle-Watts D, Low J, Clifton R, Howell KA, Carrie C, Donald T, Whelan J (2007) Characterization of the regulatory and expression context of an alternative oxidase gene provides insights into cyanide-insensitive respiration during growth and development. Plant Physiol 143: 1519–1533

Hubert N, Duponchelle F, Nuñez J, Rivera R, Renno J-F (2006) Evidence of reproductive isolation among closely related sympatric species of *Serrasalmus* (*Ostariophysii*, Characidae) from the upper Madeira River, Amazon, Bolivia. J Fish Biol 69: 31–51

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 J-F, Aury J-M, Noel B, Saudemont B, Nowacki M, Serrano V, Porcel BM, Ségurens B, Le Mouël AL, Lepère G, Schächter, Bétermier M, Cohen J, Wincker P, Sperling L, Duret L, Meyer E (2008) Translational control of intron splicing in eukaryotes. Nature 451: 359–362

Jiang J, Gusev Y, Aderca I, Mettler TA, Nagorney DM, Brackett DJ, Roberts LR, Schmittgen TD (2008) Micro-RNA expression in hepatocellular carcinoma is associated with hepatitis infection, cirrhosis and patient's survival. Clin Cancer Res 14: 419–427

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: 5873–5877

Knowles DG, McLysaght A (2006) High rate of recent intron gain and loss in simultaneously duplicated *Arabidopsis* genes. Mol Biol Evol 23: 1548–1557

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

Li SC, Tang P, Lin W-C (2007) Intronic microRNA: discovery and biological implications. DNA Cell Biol 26: 195–207

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

Lorkovic ZJ, Kirk DAW, Lambermon MH, Filipowicz W (2000) Pre-mRNA splicing in higher plants. Trends Plant Sci 5: 160–167

Louro R, Nakaia HTI, Amaral PP, Festa F, Sogayar MC, Verjosvki-Almeida S, Reis EMR (2007) Androgen responsive intronic non-coding RNAs. BMC Biol 5: 1–14

Luehrsen KR, Walbot V (1994) Addition of A- and U-rich sequence increases the splicing efficiency of a deleted from a maize intron. Plant Mol Biol 24: 449–463

Mártonfi P, Brutovská R, Cellárová E, Repcák M (1996) Apomixis and hybridity in *Hypericum perforatum*. Folia Geobot Phytotax 31: 389–396

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

Matzk F, Meister A, Brutovská R, Schubert I (2001) Reconstruction of reproductive diversity in *Hypericum perforatum* L. opens novel strategies to manage apomixis. Plant J 26: 275–282

McDonald AE (2008) Alternative oxidase: an inter-kingdom perspective on the function and regulation of this broadly distributed 'cyanide-resistant' terminal oxidase. Funct Plant Biol 35: 535–552

McDonald AE, Vanlerberghe GC (2006) The organization and control of plant mitochondrial metabolism. In: Plaxton WC, McManus MT (eds) Control of Primary Metabolism in Plants. Annual Plant Reviews, Vol. 22. Blackwell, Oxford, pp 290–324

Murashige T, Skoog F (1962) A revised medium for rapid growth and bio assays with tobacco tissues cultures. Physiol Plant 15: 473–497

Nakaya HI, Amaral PP, Louro R, Lopes A, Fachel AA, Moreira YB, El-Jundi TA, da Silva AM, Reis EM, Verjovski-Almeida S (2007) Genome mapping and expression analyses of human intronic noncoding RNAs specific patterns and enrichment in genes related to regulation of transcription. Genome Biol 8: R43.1–R43.25

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

Ner-Gaon H, Leviatan N, Rubin E, Fluhr R. (2007) Comparative cross-species alternative splicing in plants. Plant Physiol 144: 1632–1641

Noack KL (1939) Über *Hypericum*–Kreuzungen VI: Fortpflanzungsverhältnisse und bastarde von *Hypericum perforatum* L. Z Indukt Abstamm Vererbungsl 76: 569–601

Noh S-J, Lee K, Paik H, Hur C-G (2006) TISA: tissue-specific alternative splicing in human and mouse genes. DNA Res 10: 1093

Pank F, Matzk F, Kästner U, Blüthner WD, Garcia EF, Meister A, Ryschka U, Schumann G (2003) Reproductive diversity and strategies for breeding in St. John's wort (*Hypericum perforatum* L.). Euphytica 134: 77–84

Plaxton WC, Podestá FE (2006) The functional organisation and control of plant respiration. Crit Rev Plant Sci 25: 159–198

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

Reddy ASN (2001) Nuclear pre-mRNA splicing in plants. Crit Rev Plant Sci 20: 523–571

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

Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273: 1516–1517

Rose AB (2002) Requirements for intron-mediated enhancement of gene expression in *Arabidopsis*. RNA 8: 1444–1453

Rose AB, Elfersi T, Parra G, Korf I (2008) Promoter-proximal introns in *Arabidopsis thaliana* are enriched in dispersed signals that elevate gene expression. Plant Cell 20: 543–551

Roy SW (2003) Recent evidence for the exon theory of genes. Genetica 188: 251–266

Rudd S, Frisch M, Grote K. Meyers BC, Mayer K, Werner T (2004) Genome-wide *in silico* mapping of scaffold/matrix attachment regions in *Arabidopsis* suggests correlation of intragenic scaffold/matrix attachment regions with gene expression. Plant Physiol 135: 715–722

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

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

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 16*q*22.1. Genomics 89: 354–361
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
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
Szücs P, Skinner JS, Karsai I, 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 Genet Genomics 277: 249–261
Tetko IV, Haberer G, Rudd S, Meyers B, Mewes HW, Mayer KF (2006) Spatiotemporal expression control correlates with intragenic scaffold matrix attachment regions (S/MARs) in *Arabidopsis thaliana*. PLoS Comput Biol 2: 0136–0145
Umbach AL, Fiorani F, Siedow JN (2005) Characterization of transformed arabidopsis with altered alternative oxidase levels and analysis of effects on reactive oxygen species in tissue. Plant Physiol 139: 1806–1820
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
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
Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31: 3406–3415

Edited by V. Hurry