

# Ascorbate peroxidase-related (APx-R) is not a duplicable gene

Christophe Dunand,<sup>1,2,\*</sup> Catherine Mathé,<sup>1,2</sup> Fernanda Lazzarotto,<sup>3</sup> Rogério Margis<sup>4,5</sup> and Marcia Margis-Pinheiro<sup>3</sup>

<sup>1</sup>Université de Toulouse; Laboratoire de Recherche en Sciences Végétales; <sup>2</sup>CNRS; Castanet-Tolosan, France; <sup>3</sup>Departamento de Genética; <sup>4</sup>Centro de Biotecnologia; <sup>5</sup>Departamento de Biofísica; Universidade Federal do Rio Grande do Sul; Porto Alegre, Brazil

**Keywords:** single-copy gene, gene duplication, ascorbate peroxidase, selection pressure

Phylogenetic, genomic and functional analyses have allowed the identification of a new class of putative heme peroxidases, so called APx-R (Apx-Related). This new class, mainly present in the green lineage (including green algae and land plants), can also be detected in other unicellular chloroplastic organisms. Except for recent *Apx-R* extra-copies lost after chromosomal or segmental duplications. In a similar way, most *Apx-R* co-expressed genes in *Arabidopsis* genome do not have conserved extra-copies after chromosomal duplications and are predicted to be localized in organelles, as are the *Apx-R*. The member of this gene network can be considered a unique gene, well conserved through the evolution due to a strong negative selection pressure and a low evolution rate.

Ascorbate peroxidases (APx) belong to the class I peroxidase. They have been detected in all chloroplastic containing organisms in which they form a small multigenic family in green lineage.<sup>1</sup> They were subjected to some species specific duplications which produced punctual variation in a number of isoforms from 3 to 10. These duplications are probably associated with subfunctionalization. Indeed three major subclasses are defined based on their cellular localizations: cytoplasmic, peroxisomal and chloroplastic/mitochondrial.<sup>2</sup> Recently, an additional group of sequences closely related to APx has been characterized and named ascorbate peroxidase-related (APx-R). Noteworthy, this new class does not seem to be subjected to functional duplication.<sup>3</sup>

Exhaustive datamining of multiple sequence resources have been performed with available genome and EST libraries to confirm the previous observations. No functional *APx-R* gene duplication has been detected. Duplicated *APx-R* are only observed in polyploid organisms: *Triticum aestivum*, an allohexaploid, possesses 3 *APx-R*,<sup>4</sup> and *Brassica napus*, an allotetraploid, contains at least two independent expressed *APx-R* with no evidence of conservation of all expected paralogs sequences. *Glycine max*, an ancient polyploid (paleopolyploid, tetraploid)<sup>5</sup> possesses a single *APx-R* sequence and one pseudogene while most of *APx* genes have been detected in duplicated forms. Exhaustive data mining shows that *APx-R* genes are present in green algae (Chlorophyceae such as *Chlamydomonas reinhardtii* and Charophyceae such as *Klebsormidium flaccidum*) and streptophytes, even though two marginal presences have been detected in chloroplastic diatoms (Table 1). APx-R sequence can be considered as good functional molecular marker because APx-R phylogenetic tree

and taxonomic tree are congruent (Fig. 1). More genomic data are needed to determine if all APx-R sequences share the same ancestral sequence or if APx-R from diatoms resulted from a convergent evolution.

The search performed in EST libraries demonstrated that *APx-R* are poorly or not expressed in all analyzed organisms with an expression average of 0.003%. Expression analysis in *Arabidopsis thaliana* with Genevestigator<sup>6</sup> confirmed the low level of expression.

In addition to the absence of conserved duplication, high level of sequence conservation is detected (minimum of 50% identity between green algae and streptophyte, and 40% between chloroplastic diatoms and streptophytes). High variability intron positions and number is observed in diatoms and green algae (Fig. 1). However, intron positions and number are highly conserved in higher plants. Only low conservation of the gene structure is observed in the 5'end of the sequences which coincides with the variability of the coding sequence.

Detailed analysis of *Arabidopsis thaliana* *APx-R* co-expression network demonstrated that among the 42 genes listed, 31 encode proteins that are predicted to be localized in organelles, in most cases chloroplasts. These proteins display a great variety of biological functions, but a considerable number of them are implicated in chloroplasts protection against photooxidative damage, which suggests that APx-R could play a role in this protective mechanism as well. Interestingly, more than half of those genes are present as single-copy or as low-copy number in *Arabidopsis thaliana* (24 among the 42 genes, Table 2), but also in *Oryza sativa*, *Populus trichocarpa* and *Vitis vinifera* genomes. This data confirms that plant proteins predicted to be targeted to organelles are more likely single-copy

\*Correspondence to: Christophe Dunand; Email: christophe.dunand@lrsv.ups-tlse.fr  
Submitted: 08/23/11; Accepted: 09/15/11  
DOI: 10.4161/psb.6.12.18098

**Table 1.** Ascorbate peroxidase-related (APx-R)-encoding genes identified in different plant species

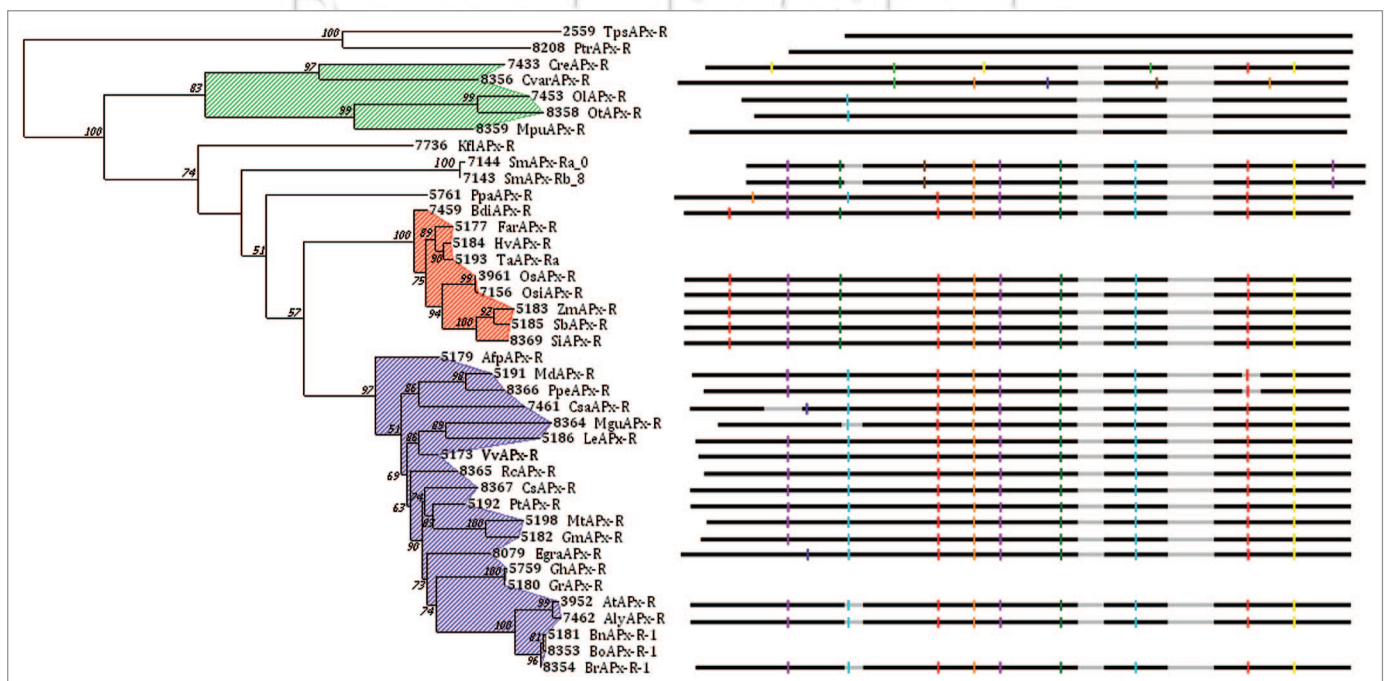
Name	Taxonomic group	Organism	Sequence Status	Expression (EST count)	Intron number
PtAPx-R	Bacillariophyta (diatoms)	<i>Phaeodactylum tricorutum</i>	complete	4/133887	0
TpsAPx-R	Bacillariophyta (diatoms)	<i>Thalassiosira pseudonana</i>	complete	0/61913	0
CreAPx-R	Chlorophyta (green algae)	<i>Chlamydomonas reinhardtii</i>	complete	18/204076	6
CvarAPx-R	Chlorophyta (green algae)	<i>Chlorella variabilis</i>	complete	0/413	5
MpuAPx-R	Chlorophyta (green algae)	<i>Micromonas pusilla</i>	complete	no	0
OIAPx-R	Chlorophyta (green algae)	<i>Ostreococcus lucimarinus</i>	complete	0/17592	1
OtAPx-R	Chlorophyta (green algae)	<i>Ostreococcus tauri</i>	complete	no	1
VcaAPx-R	Chlorophyta (green algae)	<i>Volvox carteri</i>	partial	/132038	
KfIAPx-R	Other Streptophyta	<i>Klebsormidium flaccidum</i>	complete	*	na
AcvAPx-R	Cryptogam	<i>Adiantum capillus-veneris</i>	partial	1/30540	na
MpAPx-R	Cryptogam	<i>Marchantia polymorpha</i>	partial	1/33692	na
PpaAPx-R	Cryptogam	<i>Physcomitrella patens</i>	complete	7/362131	10
SmAPx-Ra_0	Cryptogam	<i>Selaginella moellendorffii</i>	complete	4/93811	10
SmAPx-Rb_8	Cryptogam	<i>Selaginella moellendorffii</i>	complete	0/93811	10
PgAPx-R	Gymnospermae	<i>Picea glauca</i> (white spruce)	partial	3/313110	na
PsiAPx-R	Gymnospermae	<i>Picea sitchensis</i> (Sitka spruce)	partial	1/186637	na
AmaAPx-R	Eudicotyledons	<i>Antirrhinum majus</i> (snapdragon)	partial	1/25310	na
AfpAPx-R	Eudicotyledons	<i>Aquilegia formosa</i> x <i>Aquilegia pubescens</i>	complete	4/85039	na
AlyAPx-R	Eudicotyledons	<i>Arabidopsis lyrata</i>	complete	no	9
AtAPx-R	Eudicotyledons	<i>Arabidopsis thaliana</i>	complete	17/1529700	9
BnAPx-R-1	Eudicotyledons	<i>Brassica napus</i> (oilseed rape)	complete	3/643937	na
BnAPx-R-2	Eudicotyledons	<i>Brassica napus</i> (oilseed rape)	partial	1/643937	na
BoAPx-R-1	Eudicotyledons	<i>Brassica oleracea</i> (Cauliflower)	complete	5/179150	na
BrAPx-R-1	Eudicotyledons	<i>Brassica rapa</i>	complete	0/194305	9
CclAPx-R	Eudicotyledons	<i>Citrus clementina</i>	complete	0/118365	9
CsAPx-R	Eudicotyledons	<i>Citrus sinensis</i>	complete	1/213830	9
CsaAPx-R	Eudicotyledons	<i>Cucumis sativus</i>	partial	0/8128	9
EgraAPx-R	Eudicotyledons	<i>Eucalyptus grandis</i>	complete	0/1910	9
EeAPx-R	Eudicotyledons	<i>Euphorbia esula</i>	partial	1/47543	na
GmAPx-R	Eudicotyledons	<i>Glycine max</i> (soybean)	complete	13/1461624	9
GmAPx-R[P]	Eudicotyledons	<i>Glycine max</i> (soybean)	pseudogene	no	nd
GhAPx-R	Eudicotyledons	<i>Gossypium hirsutum</i> (cotton)	complete	8/273779	na
GrAPx-R	Eudicotyledons	<i>Gossypium raimondii</i>	complete	3/63577	na
HarAPx-R	Eudicotyledons	<i>Helianthus argophyllus</i>	partial	1/35720	na
LjAPx-R	Eudicotyledons	<i>Lotus japonicus</i>	partial	10/242432	na
LeAPx-R	Eudicotyledons	<i>Lycopersicon esculentum</i> (Tomato)	complete	6/298289	9
MdAPx-R	Eudicotyledons	<i>Malus domestica</i> (apple tree)	complete	2/324565	9
MeAPx-R	Eudicotyledons	<i>Manihot esculenta</i> (cassava)	partial	1/80681	9
MtAPx-R	Eudicotyledons	<i>Medicago truncatula</i> (barrel medic)	complete	4/269238	9
MguAPx-R	Eudicotyledons	<i>Mimulus guttatus</i>	complete	20/261907	8
NtAPx-R	Eudicotyledons	<i>Nicotiana tabacum</i>	partial	1/332667	na
PtAPx-R	Eudicotyledons	<i>Populus trichocarpa</i> (poplar)	complete	1/89943	9
PpeAPx-R	Eudicotyledons	<i>Prunus persica</i> (peach)	complete	0/79584	9

Exhaustive data mining was performed with all available resources (JGI, NCBI, Phytozome...). When available, EST count and intron number were determined and included in the 5th and 6th columns. no: no EST was found; nd: gene structure cannot be determined; na: no genomic sequence available; \*sequence kindly provided by R.Timme.

**Table 1.** Ascorbate peroxidase-related (APx-R)-encoding genes identified in different plant species (continued)

RcAPx-R	Eudicotyledons	<i>Ricinus communis</i>	complete	1/62582	9
StAPx-R	Eudicotyledons	<i>Solanum tuberosum</i> (Potato)	partial	6/249614	na
ToAPx-R	Eudicotyledons	<i>Taraxacum officinale</i> (dandelion)	partial	2/41296	na
VvAPx-R	Eudicotyledons	<i>Vitis vinifera</i> (Grape)	complete	7/362674	9
AGcAPx-R	Monocotyledons	<i>Agrostis capillaris</i>	partial	1/7743	na
AsAPx-R	Monocotyledons	<i>Avena sativa</i> (Oat)	partial	1/25344	na
BdiAPx-R	Monocotyledons	<i>Brachypodium distachyon</i>	complete	23/128092	10
FarAPx-R	Monocotyledons	<i>Festuca arundinacea</i>	complete	4/63758	na
HvAPx-R	Monocotyledons	<i>Hordeum vulgare</i> (barley)	complete	19/525781	na
OmAPx-R	Monocotyledons	<i>Oryza minuta</i>	partial	1/5760	na
OsiAPx-R	Monocotyledons	<i>Oryza sativa</i> (indica)	complete	?/203447	10
OsAPx-R	Monocotyledons	<i>Oryza sativa</i> (japonica)	complete	?/987318	10
ShyAPx-R	Monocotyledons	<i>Saccharum hybrid cultivar</i> (sugarcane)	partial	3/282809	na
SiAPx-R	Monocotyledons	<i>Setaria italica</i>	complete	0/2741	10
SbAPx-R	Monocotyledons	<i>Sorghum bicolor</i>	complete	6/209828	10
TaAPx-Ra	Monocotyledons	<i>Triticum aestivum</i> (bread wheat)	complete	4/1071453	na
TaAPx-Rb	Monocotyledons	<i>Triticum aestivum</i> (bread wheat)	partial	2/1071453	na
TaAPx-Rd	Monocotyledons	<i>Triticum aestivum</i> (bread wheat)	partial	2/1071453	na
ZmAPx-R	Monocotyledons	<i>Zea mays</i>	complete	16/2019105	10

Exhaustive data mining was performed with all available resources (JGI, NCBI, Phytozome...). When available, EST count and intron number were determined and included in the 5th and 6th columns. no: no EST was found; nd: gene structure cannot be determined; na: no genomic sequence available; \*sequence kindly provided by R.Timme.



**Figure 1.** Phylogenetic analyses of APx-R were conducted using complete protein sequences from 39 organisms (Gene accession numbers and name are given as PeroxiBase, peroxibase.toulouse.inra.fr/). Alignment was performed with MAFFT<sup>10</sup> and gap columns were discarded. Phylogenetic analysis was conducted with PhyML, using JTT substitution matrix (4), and among-site rate variation was modeled using a Gamma distribution plus a percent of invariant sites and observed frequencies, as advised by ProtTest (JTT + I + G + F). Branch support was calculated by approximate Likelihood Ratio Test (aLRT), and displayed when higher than 50%. Intron positions relatively to the protein alignment are shown, as obtained with CIWOG<sup>11</sup> and were included in front of each branch. The hatched areas in green, red and purple stand for green algae, monocots and dicots respectively.

**Table 2.** Co-The list of APx- R co-expressed genes was obtained through the network generated with ATTED-II ver. 6.0 (*atted.jp/*)

Gene	Annotation	Subcellular Localization			Single-copy gene*	Low-copy gene**
		Target P	Psort	TAIR		
At1g05385	<i>LOW PSII ACCUMULATION 19 (LPA19)</i>	Chlo	Chlo	chloroplast, chloroplast thylakoid lumen	Yes	-
At1g08550	<i>NON-PHOTOCHEMICAL QUENCHING 1 (NPQ1); ARABIDOPSIS VIOLAXANTHIN DE-EPOXIDASE 1 (AVDE1)</i>	Other	Cyto	chloroplast photosystem II, chloroplast thylakoid lumen	No	Yes
At1g10830	<i>15-CIS-ZETA-CAROTENE ISOMERASE (Z-ISO)</i>	Chlo	Chlo	chloroplast	Yes	-
At1g27385	Unknown protein	Chlo	Chlo	chloroplast	No	Yes
At1g33290	Sporulation protein-related	Chlo	Chlo	n/d	No	No
At1g33330	Peptide chain release factor	Mito	Chlo	chloroplast	Yes	-
At1g54520	Unknown protein	Chlo	Chlo	chloroplast	Yes	-
At1g64430	Unknown protein	Chlo	Chlo	n/d	No	Yes
At1g67840	<i>CHLOROPLAST SENSOR KINASE (CSK)</i>	Chlo	Chlo	chloroplast, chloroplast stroma	No	Yes
At1g76730	5-formyltetrahydrofolate cyclo-ligase family protein	Chlo	Chlo	chloroplast	No	Yes
At1g78140	Methyltransferase-related protein	Mito	Chlo	chloroplast, plastoglobule	No	No
At1g78995	Unknown protein	Chlo	Chlo	n/d	No	Yes
At2g01620	<i>MATERNAL EFFECT EMBRYO ARREST 11 (MEE11)</i>	Other	Chlo	n/d	No	No
At2g03390	uvrB/uvrC motif-containing protein	Chlo	Chlo	chloroplast	No	No
At2g20860	<i>LIPOIC ACID SYNTHASE 1 (LIP1)</i>	Mito	Chlo	mitochondrial matrix, mitochondrion	No	No
At2g30170	Unknown protein	Chlo	Chlo	chloroplast	No	No
At2g37920	<i>EMBRYO DEFECTIVE 1513 (emb1513)</i>	Chlo	Chlo	n/d	No	Yes
At2g38270	<i>CAX-INTERACTING PROTEIN 2 (CXIP2); GLUTAREDOXIN (ATGRX2)</i>	Chlo	Chlo	chloroplast, chloroplast stroma	Yes	-
At3g10970	Haloacid dehalogenase-like hydrolase family protein	Chlo	Chlo	chloroplast	Yes	-
At3g48560	<i>CHLORSULFURON/IMIDAZOLINONE RESISTANT 1 (CSR1); ACETOLACTATE SYNTHASE (ALS); ACETOHYDROXY ACID SYNTHASE (AHAS); TRIAZOLOPYRIMIDINE RESISTANT 5 (TZP5); IMIDAZOLE RESISTANT 1 (IMR1)</i>	Chlo	Chlo	chloroplast	No	No
At3g53920	<i>RNA POLYMERASE SIGMA-SUBUNIT C (SIGC); SIGMA FACTOR 3 (SIG3)</i>	Chlo	Chlo	chloroplast	No	No
At3g55630	<i>A. THALIANA DHFS-FPGS HOMOLOG D (ATDFD)</i>	Other	Cyto	cytosol	No	No
At4g02260	<i>RELA-SPOT HOMOLOG 1 (RSH1); RELA-SPOT HOMOLOG 1 (AT-RSH1); RELA/SPOT HOMOLOG 1 (ATRSH1)</i>	Chlo	Plast	chloroplast	No	No
At4g10000	Electron carrier protein; disulfide oxidoreductase	Chlo	Chlo	chloroplast	Yes	-
At4g25650	<i>ACD1-LIKE (ACD1-LIKE); PROTOCHLOROPHYLLIDE-DEPENDENT TRANSLOCON COMPONENT 52 KDA (PTC52)</i>	Chlo	Plast	chloroplast, chloroplast envelope	No	No
At4g27600	<i>NECESSARY FOR THE ACHIEVEMENT OF RUBISCO ACCUMULATION 5 (NARA5)</i>	Chlo	Chlo	chloroplast	Yes	-
At4g30310	Ribitol kinase protein	Other	Chlo	chloroplast	No	No
At4g32320	<i>ASCORBATE PEROXIDASE-RELATED (APX-R)</i>	Chlo	Chlo	cytosol	Yes	-

The putative subcellular localization was predicted through TargetP ver. 1.1 ([www.cbs.dtu.dk/services/TargetP/](http://www.cbs.dtu.dk/services/TargetP/)) and Psort ver. 3.0 ([www.psort.org/psortb/](http://www.psort.org/psortb/)) and from TAIR databases ([www.arabidopsis.org/](http://www.arabidopsis.org/)). The number of copies of each gene was estimated from the data published by Duarte et al., 2010, which listed single and low copy genes in *Oryza sativa*, *Vitis vinifera*, *Populus trichocarpa* and *Arabidopsis thaliana* genomes. \*Single-copy genes in *Oryza sativa*, *Vitis vinifera*, *Populus trichocarpa* and *Arabidopsis thaliana* genomes, according to Duarte et al. 2010.<sup>7</sup> \*\*Genes present as one or two copies in at least one of the analyzed genomes. \*\*\*Not present in *Oryza sativa*.

**Table 2.** Co-The list of APx- R co-expressed genes was obtained through the network generated with ATTED-II ver. 6.0 (*atted.jp/*) (continued)

Gene	Annotation	Subcellular Localization			Single-copy gene*	Low-copy gene**
		Target P	Psort	TAIR		
At5g02250	EMBRYO DEFECTIVE 2730 (EMB2730); RIBONUCLEOTIDE REDUCTASE 1 (RNRI); ARABIDOPSIS THALIANA MITOCHONDRIAL RNASE II (ATMTRNASEII)	Chlo	Chlo	chloroplast, mitochondrion	Yes	-
At5g03900	Unknown protein	Chlo	Plast	chloroplast envelope	Yes	-
At5g04360	PULLULANASE 1 (ATPU1); LIMIT DEXTRINASE (ATLDA); PULLULANASE 1 (PU1)	Chlo	Chlo	chloroplast	No	Yes
At5g06340	ARABIDOPSIS THALIANA NUDIX HYDROLASE HOMOLOG 27 (ATNUDX27)	Chlo	Chlo	chloroplast	No	No
At5g08340	Riboflavin biosynthesis protein-related	Other	Chlo	cellular_component unknown	No	No
At5g08410	FERREDOXIN/THIOREDOXIN REDUCTASE SUBUNIT A2 (FTRA2)	Chlo	Chlo	chloroplast	No	Yes
At5g13720	Unknown protein	Chlo	Plast	chloroplast, chloroplast inner membrane, chloroplast envelope	No	No
At5g18140	DNAJ heat shock N-terminal domain-containing protein	Chlo	Nuclear	n/d	No	No
At5g19540	Unknown protein	Chlo	Chlo	chloroplast	No	Yes
At5g26820	MULTIPLE ANTIBIOTIC RESISTANCE 1 (MAR1); IRON REGULATED 3 (IREG3)	Chlo	Plast	chloroplast, chloroplast envelope	No	Yes
At5g38510	Rhomboid family protein	Chlo	Nuclear	integral to membrane	Yes***	-
At5g57040	Lactoylglutathione lyase family protein	Chlo	Chlo	chloroplast	Yes	-
At5g65685	Soluble glycogen synthase-related protein	Chlo	Chlo	chloroplast	No	No

The putative subcellular localization was predicted through TargetP ver. 1.1 ([www.cbs.dtu.dk/services/TargetP/](http://www.cbs.dtu.dk/services/TargetP/)) and Psort ver. 3.0 ([www.psort.org/psortb/](http://www.psort.org/psortb/)) and from TAIR databases ([www.arabidopsis.org/](http://www.arabidopsis.org/)). The number of copies of each gene was estimated from the data published by Duarte et al., 2010, which listed single and low copy genes in *Oryza sativa*, *Vitis vinifera*, *Populus trichocarpa* and *Arabidopsis thaliana* genomes. \*Single-copy genes in *Oryza sativa*, *Vitis vinifera*, *Populus trichocarpa* and *Arabidopsis thaliana* genomes, according to Duarte et al. 2010.<sup>7</sup> \*\*\*Genes present as one or two copies in at least one of the analyzed genomes. \*\*Not present in *Oryza sativa*.

than expected by chance.<sup>7</sup> This could happen because these proteins, when present in the organelles, interact with proteins that are encoded by the organellar genome. In this case, the level of nuclear genome encoded proteins has to be very well controlled inside the cell, so the interaction network will not be disturbed. Looking specifically to the network genes that are single-copy in the specified genomes, we noticed that the majority of the extra-copies of these genes were lost after chromosomal duplications, in a situation very similar to *APx-R* gene. Thus, it is possible to infer that a great number of single and low-copy genes in this co-expression network could reflect a dose-dependent system, where a raise in copy numbers of such genes would not be favorable to the network. In **Figure 2**, *LPA19* (At1g05385), peptide release factor (*At1g33330*) and 15-cis-zeta-carotene isomerase (*At1g10830*) genes were used

as examples. The chromosomal segments that contain these genes in *Arabidopsis* were duplicated during the evolution and genomic analyses showed that the extra copies were lost during this process (red dashed lines).

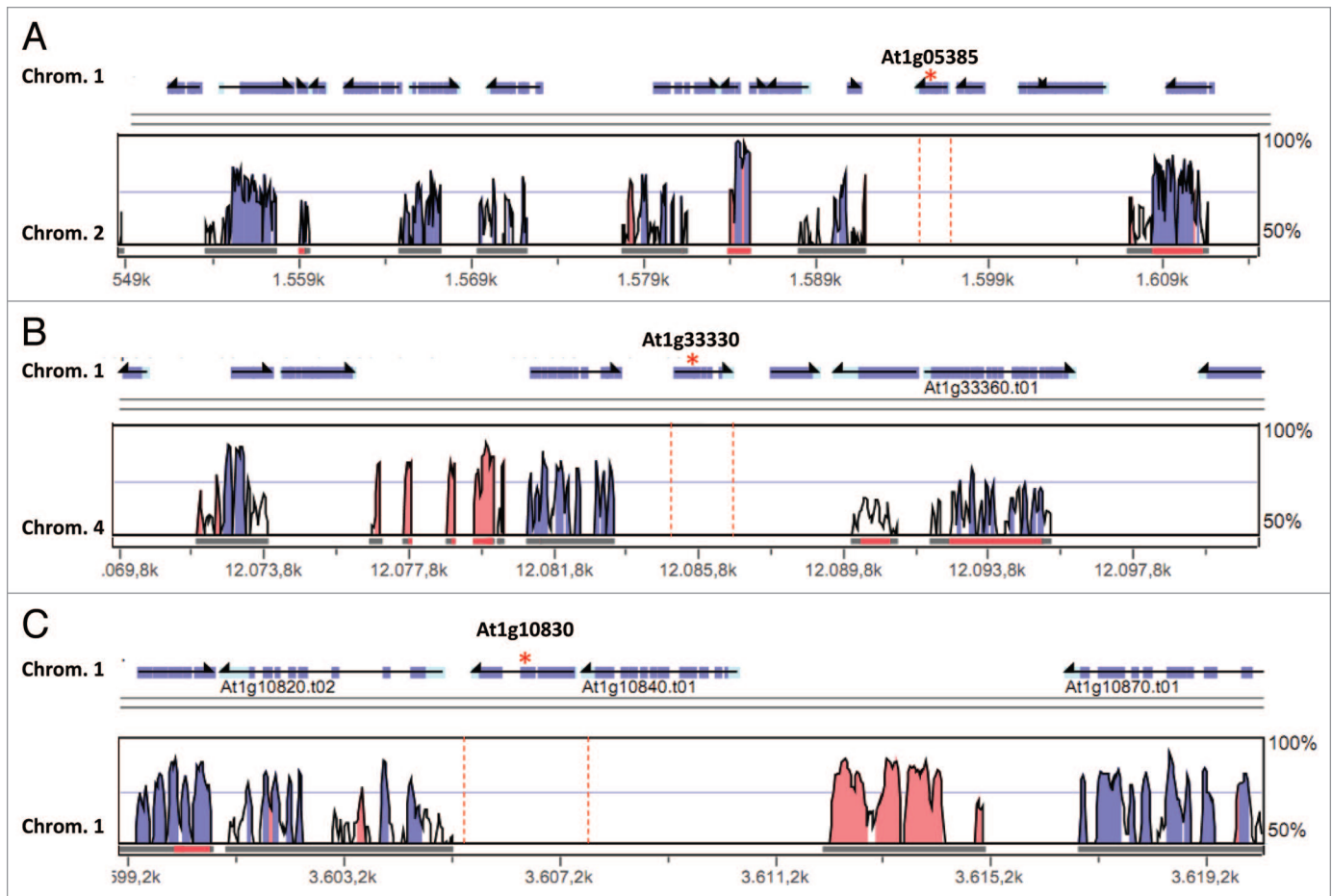
The hypothesis of conserved unique genes has already been proposed in reference 8. However further analyses are mandatory to precisely evaluate the extension of the proposal of a complex network of unique gene, taking into consideration that many other neighbor genes were also deleted from these genomic regions. The conservation of this unique gene network indicates that they are under a strong negative selection pressure and subjected to low evolution rate.

#### Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

#### References

1. Passardi F, Bakalovic N, Teixeira FK, Margis-Pinheiro M, Penel C, Dunand C. Prokaryotic origins of the non-animal peroxidase superfamily and organelle-mediated transmission to eukaryotes. *Genomics* 2007; 89:567-79; PMID:17355904; DOI:10.1016/j.ygeno.2007.01.006.
2. Teixeira FK, Menezes-Benavente L, Galvao VC, Margis R, Margis-Pinheiro M. Rice ascorbate peroxidase gene family encodes functionally diverse isoforms localized in different subcellular compartments. *Planta* 2006; 224:300-14; PMID:16397796; DOI:10.1007/s00425-005-0214-8.
3. Lazzarotto F, Teixeira FK, Rosa SB, Dunand C, Fernandes C, Fontenele AD, et al. Ascorbate peroxidase-related (APx-R) is a new heme-containing protein functionally associated with ascorbate peroxidase but evolutionarily divergent. *New Phytol* 2011; 191:234-50; PMID:21352234; DOI:10.1111/j.1469-8137.2011.03659.x.



**Figure 2.** Comparative analyses between the genomic regions containing *APx-R* co-expressed genes *At1g05385* (A), *At1g33330* (B) and *At1g10830* (C) and the DNA segments generated after these regions were duplicated. Peaks correspond to conserved DNA sequences (exons, in most cases) in the *Arabidopsis* duplicated genomic regions. Horizontal arrows above the graphics represent the annotated genes in *Arabidopsis* genome (AGI codes are shown). The red dot lines indicate the absence of the *APx-R* co-expressed gene in duplicated region. The VISTA program ([www-gsd.lbl.gov/vista/](http://www-gsd.lbl.gov/vista/)) was used in these analyses.

- Kerby K, Kuspira J. The phylogeny of the polyploid wheats *Triticum aestivum* (bread wheat) and *Triticum turgidum* (macaroni wheat). *Genome* 1987; 29:722-37; DOI:10.1139/g87-124.
- Schmutz J, Cannon SB, Schlueter J, Ma JX, Mitros T, Nelson W, et al. Genome sequence of the palaeopolyploid soybean. *Nature* 2010; 463:178-83; PMID:20075913; DOI:10.1038/nature08670.
- Hruz T, Laule O, Szabo G, Wessendorp F, Bleuler S, Oertle L, et al. Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. *Adv Bioinformatics* 2008; 2008:420747; PMID:19956698.
- Duarte JM, Wall PK, Edger PP, Landherr LL, Ma H, Pires JC, et al. Identification of shared single copy nuclear genes in *Arabidopsis*, *Populus*, *Vitis* and *Oryza* and their phylogenetic utility across various taxonomic levels. *BMC Evol Biol* 2010; 10:61; PMID:20181251; DOI:10.1186/1471-2148-10-61.
- Armisen D, Lecharny A, Aubourg S. Unique genes in plants: specificities and conserved features throughout evolution. *BMC Evol Biol* 2008; 8:280; PMID:18847470; DOI:10.1186/1471-2148-8-280.
- Koua D, Cerutti L, Falquet L, Sigrist CJA, Theiler G, Hulo N, et al. PeroxiBase: a database with new tools for peroxidase family classification. *Nucleic Acids Res* 2009; 37:261-6; PMID:18948296; DOI:10.1093/nar/gkn680.
- Katoh K, Toh H. Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform* 2008; 9:286-98; PMID:18372315; DOI:10.1093/bib/bbn013.
- Wilkerson MD, Ru YB, Brendel VP. Common introns within orthologous genes: software and application to plants. *Brief Bioinform* 2009; 10:631-44; PMID:19933210; DOI:10.1093/bib/bbp051.