Genome-wide analysis of plant glutaredoxin systems

Nicolas Rouhier*, Jérémie Couturier and Jean-Pierre Jacquot

Unité Mixte de Recherches 1136 Interaction arbres microorganismes, INRA, Université Henri Poincaré, IFR 110, Génomique Ecologie et Ecophysiologie Fonctionnelles, Faculté des Sciences, BP 239, F-54506 Vandoeuvre-lès-Nancy Cedex, France

Received 16 March 2006; Accepted 27 March 2006

Abstract

The recent release of the first tree genome (Populus trichocarpa) has allowed a comparison to be made of the multigenic glutaredoxin (Grx) and glutathione reductase (GR) families of this tree with those of other sequenced organisms and especially of the two other fully sequenced plant species, Arabidopsis thaliana and Oryza sativa. Grxs are small proteins involved in disulphide bridge or protein–glutathione adduct reduction, and they are maintained in a reduced form using glutathione and an NADPH-dependent GR. While the P. trichocarpa and O. sativa genomes are nearly five times larger than that of A. thaliana, they contain ~45 000 and 37 500 genes compared with the 25 500 genes of A. thaliana. On the one hand, the GR gene composition varies little between species and the gene structures are relatively conserved. On the other hand, the Grx gene family can be divided into three subgroups and the gene content is larger in P. trichocarpa (36 genes) compared with A. thaliana and O. sativa (31 and 27 genes, respectively). This could be partly explained by the occurrence of more duplication events, and this is especially true for one of the three identified Grx subgroups (subgroup III). The expression of most of these genes was confirmed by analysing expressed sequence tags present in various databases. In addition, the expression of Grx of subgroups I and II was examined by RT–PCR in various poplar organs. A complete classification based essentially on gene structure and sequence identity is proposed.

Key words: Genome, glutaredoxin, glutathione reductase, oxidative stress, poplar.

Introduction

When submitted to adverse environmental conditions (biotic or abiotic stresses), plants very often react by generating oxidative bursts. To avoid biological damage, the concentration of the oxidizing species must be kept under control. One of the most documented functions of glutaredoxins (Grxs) in plants is their involvement in the oxidative stress response. They are implicated in many different ways, for example by directly reducing peroxides or dehydroascorbate (DHA), by reducing peroxideroxins (Prx), and also by protecting thiol groups on other enzymes via glutathionylation/deglutathionylation mechanisms (for a review, see Rouhier et al., 2004). Grxs need to be reduced in order to function, the reducing system being composed of an NADPH-dependent pyridine nucleotide oxidoreductase called glutathione reductase (GR) and the small tripeptide, glutathione. Populus has become a model organism in many laboratories to study the physiology, ecology, and genetics of forest trees for the following reasons. It has a relatively small nuclear genome, ~550 Mbp, which is quite similar in size to that of rice and about four times larger than that of Arabidopsis. All species are diploid and can be easily crossed. In addition, poplars grow quickly, they can be multiplied by clonal propagation and methods for transformation and regeneration are available. Finally, extensive genetic maps, including the identification of quantitative trait loci (QTLs) are available. With the release of version 1.1 of the Populus trichocarpa genome (http://genome.jgi-psf.org/Poptr1/Poptr1.home.html) by the DOE JGI (US Department Of Energy, Joint Genome Institute), restricted to researchers involved in the annotation process, a process of manual curation and correction started. The initial task was to collect all the sequences of interest in the Joint Genome Institute Populus Genome Sequence Database,
and to correct genes with false annotations. Additional searches, essentially using the BLASTN search and annotated Arabidopsis thaliana sequences, were performed on the entire genome to ensure that no gene of interest was missing.

In this study, the Grx system, one of the two major systems [together with the thioredoxin (Trx) system] involved in dithiol–disulphide exchanges was investigated. As GR is the key enzyme required for the reduction of Grx since it reduces glutathione using NADPH, and glutathione is the Grx reductant, a section describing the organization of the genes coding for those proteins is also included in this study. The Grx and GR gene distribution and organization in poplar were compared with the two other known plant genomes. Only two review articles have been devoted to the description of Grx in higher plants and more precisely in A. thaliana, but there is no information about the gene distribution in other higher plants (Lemaire, 2004; Rouhier et al., 2004). Additional information comes from other photosynthetic organisms such as Chlamydomonas reinhardtii or Synechocystis sp, which possess a smaller set of Grx genes than A. thaliana Lemaire, 2004). This review will thus mostly emphasize the genome-wide distribution of GR and Grx in plants including poplar, rice and Arabidopsis Genome Initiative, 2000; International Rice Genome Sequencing Project, 2005) and focus on recent data showing the involvement of Grx in the oxidative stress response. Indeed, one of the challenges of the future is to understand why so many genes, with putative identical or similar functions, are present. One way to answer the question of gene redundancy is to study their spatial and temporal expression, the localization of their products, and their specificity towards target enzymes.

Materials and methods

RT–PCR

Total RNA isolation was performed with the RNeasy Plant Mini kit (Qiagen) from ~100 mg of frozen tissue from various poplar organs [roots, young and mature leaves, stems, petioles, fruits, and flowers (more precisely stamens)]. To remove contaminating DNA, the samples were treated with DNase I (Qiagen). Except for the stem, for which 300 ng was used, 1 µg of total RNA was converted to cDNA using reverse transcriptase (Qiagen). PCR was performed at an annealing temperature of 54 °C, and 35 cycles were used except for Grx C1,2 for which 40 cycles were used. The Ptrc Trx h1 gene, which is expressed in all the tissues tested, was amplified simultaneously (35 cycles) and used as a control. The primers used for these RT–PCR experiments were originally designed for cloning the full-length open reading frame of poplar Grx sequences into prokaryotic expression vectors, and thus they contain Ncol or BamHI restriction sites (underlined). In order to differentiate between Ptc Grx C1,1 and 2, a common reverse primer was used (Grx C1 rev) but with a different forward primer (either Grx C1,1 for or Grx C1,2 for).

The primer sequences are as follows: Grx C1,1 for, 5’ CCCCCCATGGCTATGAAACAGCGAAGGAG 3’; Grx C2 rev, 5’ CCCCCCATGGCTATGAAACAGCGAAGGAG 3’; Grx C2 rev, 5’ CCCCCCATGGCTATGAAACAGCGAAGGAG 3’; Grx C3 for, 5’ CCCCCGATCCCTAACGCTGAGA 3’; Grx C3 rev, 5’ CCCCCGATCCCTAACGCTGAGA 3’; Grx C4 for, 5’ CCCCCGATCCCTAACGCTGAGA 3’; Grx C4 rev, 5’ CCCCCGATCCCTAACGCTGAGA 3’; Grx C5 for, 5’ CCCCCGATCCCTAACGCTGAGA 3’; Grx C5 rev, 5’ CCCCCGATCCCTAACGCTGAGA 3’; Grx S12 for, 5’ CCCCCCATGGCTCTTGGTGTTGCCAGTGAAG 3’; Grx S12 rev, 5’ CCCCCCATGGCTCTTGGTGTTGCCAGTGAAG 3’; Grx S14 for, 5’ CCCCCCATGGCTCTTGGTGTTGCCAGTGAAG 3’; Grx S14 rev, 5’ CCCCCCATGGCTCTTGGTGTTGCCAGTGAAG 3’; Grx S15 for, 5’ CCCCCCATGGCTCTTGGTGTTGCCAGTGAAG 3’; Grx S15 rev, 5’ CCCCCCATGGCTCTTGGTGTTGCCAGTGAAG 3’; Grx S16 for, 5’ CCCCCCATGGCTCTTGGTGTTGCCAGTGAAG 3’; Grx S16 rev, 5’ CCCCCCATGGCTCTTGGTGTTGCCAGTGAAG 3’; Trx h1 for, 5’ GGGGCGAATCATGCTCAGTTGC 3’; Trx h1 rev, 5’ GGGGCGAATCATGCTCAGTTGC 3’.

The glutaredoxin family

Identified Grx functions in plant cells

As explained above, the high number of genes coding for Grx suggests that these proteins should have many important biological functions. Figure 1 summarizes established roles for plant Grxs and putative functions based on available data in other organisms. A recent finding, not linked to oxidative stress, is that some Grxs of subgroup II (AtGrx S14 and AtGrx S16) interact with Cyt2+ transporters and probably regulate their activity (Cheng and Hirschi, 2003). Indirectly, this suggests that Grxs could be involved in cell signalling by regulating the Cyt2+ concentration. Plant Grxs could play additional roles, uncharacterized so far, in signalling, such as the regulation of phosphatases or transcription factors, two well-known reactions in other organisms. For example, a protein tyrosine phosphatase 1B from soybean was shown to be regulated via cysteine glutathiolation (Dixon et al., 2003). Another well documented indirect
role for plant Grx is the possibility to reduce specifically one Prx subgroup. Prxs are haem-free thiol-dependent peroxidases which reduce various peroxides using conserved catalytic cysteines. Plant Prxs are divided into five subgroups, depending essentially on the catalytic mechanism used and the position of these conserved catalytic cysteines (Dietz, 2003; Rouhier and Jacquot, 2005a).

Among the five subgroups, Grxs are only able to reduce type II Prxs, whereas Prxs of other groups are reduced by Trxs but not by Grxs (Rouhier et al., 2001; Bréhelin et al., 2003; Finkemeier et al., 2005; N Rouhier, unpublished results). In poplar, as well as in A. thaliana and O. sativa, Prxs II are present in the cytosol (Prx IIB–D), in the chloroplasts (Prx IIE), and in mitochondria (Prx IIF). Although the proteins are somewhat different, all the poplar type II Prxs are probably able to use Grx as a reductant (Rouhier et al., 2002; N Rouhier, unpublished results), which is apparently not the case in A. thaliana, for example where Prx IIE was not reduced by the Grx tested (Bréhelin et al., 2003). This observation is confirmed by the fact that Prxs IIB, IIE, and IIF were trapped on a monocysteinic Grx column (Rouhier et al., 2005b). As at least one Grx is putatively present in each compartment (see below), this in vitro specificity could be of physiological significance. Another function demonstrated for a Grx from O. sativa (OsGrx C2, Os04g42930) is its ability to reduce hydrogen peroxide directly (Lee et al., 2002). Nevertheless, its catalytic efficiency is rather low. In addition, this activity is probably not a general characteristic, since all poplar Grxs tested so far do not possess such an activity (Rouhier et al., 2003; N Rouhier, unpublished results). Recently, a cis-element, which responds to a methylviologen-generated oxidative stress, was identified in the promoter of the Grx from O. sativa, mentioned above, confirming that some Grx isoforms should play an important role in the antioxidative network (Tsukamoto et al., 2005). It was described earlier that one way to control the reactive oxygen species levels in cells is to oxidize methionine residues randomly on proteins. The methionine sulfoxide residues generated need to be converted back to methionine to restore the integrity of the damaged proteins, and this is performed by an enzyme called methionine sulfoxide reductase (Stadtman et al., 1996). This enzyme has a special catalytic mechanism which generally requires Trx as a regenerating system. In plants, some Grxs can replace Trx for this function (N Rouhier et al., unpublished results).

Among other functions identified for plant Grxs is the possibility of reducing a specific poplar Trx h isoform, called Ptrc Trxh4 (Gelhaye et al., 2003). This interaction between Grx and Trx systems in plants is probably not the only example, since a chloroplastic Trx f was shown to be regulated by glutathiolation (Michelet et al., 2005). Grxs are thus the prime candidates to catalyse both the glutathiolation and deglutathiolation processes.
Many other putative functions for plant Grxs arise from a proteomic study in which 94 putative target proteins were identified (Rouhier et al., 2005b). As most of these proteins were also identified as putative Trxs partners, the physiological significance of this observation needs to be evaluated with caution. Nevertheless, it could be that most of these enzymes and especially those belonging to the Calvin or the Krebs cycle can be regulated by glutathiolation. It is already known that poplar Grx C4 is able to catalyse deglutathiolation in in vitro tests, and that several plant proteins, including type II Prx and aldolase, can be glutathiolated in vitro or in vivo (Ito et al., 2003; Rouhier et al., 2006). It is likely that in plants submitted to abiotic or biotic stresses, these phenomena occur on a large scale especially in the organelles where the majority of the re-active oxygen species are produced, and it will thus be of considerable interest to know which among the Grx isoforms are involved in those post-translational modifications.

The original discovery of Grx in *Escherichia coli* was related to its involvement in deoxyribonucleotide synthesis as a ‘feeding system’ for the enzyme ribonucleotide reductase (Fernandes and Holmgren, 2004). This reaction is obviously also essential for DNA synthesis in plant cells and so it is clear that plant Grxs should be involved. This area of research remains largely unexplored, however, except that a plant ribonucleotide reductase has been identified in green algae (Feller et al., 1980), and that plant Grx was shown to be able to reduce bacterial ribonucleotide reductase (Rouhier et al., 2003).

In other organisms such as mammals, yeast or *E. coli*, Grxs have also been shown to participate in the haem or Fe–S assembly machinery, but no clear role has been defined (Rodriguez-Manzaneque et al., 2002; Mulhenhoff et al., 2003; Achebach et al., 2004; Wingert et al., 2005) and this has not been addressed in plants yet. Likewise, a poplar Grx which contains an iron–sulphur centre has recently been identified (Feng et al., 2005, 2006; N Rouhier et al., unpublished results). Other iron–sulphur-containing Grxs such as human mitochondrial Grx2 have also been identified (Lillig et al., 2005). As plant organelles possess their own electron transfer chains with iron–sulphur-containing proteins, it is likely that some plant Grx isoforms will be involved in these processes. Likewise, the maturation of cytochrome c and haem attachment require the presence of ‘thioredoxin-like’ molecules (Meyer et al., 2005).

**Comparison of Grxs in higher plant genomes**

Based on sequence alignments, active site sequences and construction of unrooted phylogenetic trees, three Grx subgroups can be distinguished. These subgroups are similar to those already described for *A. thaliana* Grx (Rouhier et al., 2004).

Subgroup I contains ‘classical’ Grxs with CPYC, CGYC, CPFC, and CSY[C/S] active sites. This group comprises five different classes of Grx (Grx C1–C4 and S12) which differ in their active site sequences. The nomenclature used (C or S) is based on the presence of a cysteine or a serine in the fourth position of the active site (CxxC or CxxS). The situation is different for the three plants analysed (Table 1; Fig. 2). For example, there is no CGYC-containing Grx (Grx C1) in *O. sativa* but two in *P. trichocarpa*. On the other hand, *O. sativa* contains two CPFC isoforms (Grx C2) whereas only one is present in *P. trichocarpa* and none in *A. thaliana*. Finally, instead of two CPYC Grxs as in *P. trichocarpa* or *A. thaliana*, the rice genome possesses only one CPYC isoform, but it also displays one CPYS isoform which is not present in the two other genomes. Using a BLAST search, this CPYS sequence was only found in other plants belonging to the Poaceae or in some fungi (data not shown and Table 2).

This could constitute a specificity of Grxs from cereal as already observed for the Trx h gene family (Gelhaye et al., 2005). Finally, two isoforms with WCSY or WCSS active sites are present in *A. thaliana*, but only one, with a WCSS active site, is present in *P. trichocarpa* and *O. sativa*. These proteins display an N-terminal transit peptide, which is predicted to convey the protein into plastids (Table 1).

The proteins of subgroup II possess CGFS active sites, but they differ in the number of repeated modules (one in Grx S14, S15 and S16, and three in Grx S17) and thus in their size, ranging from 170 to 492 amino acids. In addition to the Grx domain, Grx S16 possesses an additional module in the N-terminal part, which is related to a Trx domain, as it presents a WCDAS sequence, close to some atypical Trx active sites. Subgroup II is thus made up of four different classes, as also shown in Fig. 2. These Grxs were previously identified as a PICOT HD domain (for protein kinase C-interacting cousin of thioredoxin homology domain) (Isakov et al., 2000), but some characteristics other than the active site sequences and especially a conserved motif [KG][GE][LF][I/V][GG][C/S], specific to Grx and present in the C-terminus part of these proteins, allow them to be identified as Grxs. At least one gene from the three species analysed is present in each class, but the situation is also different among plants (Table 1; Fig. 2). There are two Grx S15 in *O. sativa* but only one in *P. trichocarpa* and *A. thaliana*. In contrast, there are two Grx S17 in *P. trichocarpa* but one in *A. thaliana* and *O. sativa*. Again, the presence of two closely related genes in some species probably arises from duplication events. This observation is reinforced by the very high identity between these sequences (92% between *PbrGrx S17,1*, and 2, and 88% between *OsGrx S15,1* and 2).

In *A. thaliana*, all the proteins of subgroup III possess active sites of the CC[M/L][C/S] and the CCLG active site (At1g03850) (Rouhier et al., 2004). A situation is almost similar in *P. trichocarpa*; only one sequence is divergent, with a CYMS active site. In *O. sativa*, the active site sequences vary greatly compared...
Table 1. Characteristics of *Populus trichocarpa* glutathione reductases and subgroup I and II glutaredoxins, and comparison with *Oryza sativa* and *Arabidopsis thaliana* genomes

Only the two most defined subgroups are included since subgroup III with CxC/S active sites is too large and variable between these organisms. The nomenclature Ptrc was used for *P. trichocarpa* and GrxC1–C4 or Grx S12 and Grx S14–S17 was shortened from the Grx CxxC1 or Grx CxxS14 initially defined for *A. thaliana* (Rouhier et al., 2004), according to the nature of the fourth amino acid of the active site, either a C or an S. The predicted localizations of each protein are indicated. They have been obtained with three different software packages and are generally identical [Psort (http://psort.nibb.ac.jp/form.html), TargetP (http://www.cbs.dtu.dk/services/TargetP/), and Predotar (http://urgi.infobiogen.fr/predotar/)] (C, cytosol; P, plastid; ER, endoplasmic reticulum; M, mitochondria). In addition, the sequences of the redox centre, the size of the proteins, the molecular masses of the full-length proteins, the theoretical isoelectric points and the number of ESTs present in the GenBank database are also presented. The size, molecular mass, and pl of each protein were obtained using an Expasy program called ProtParam tool (http://www.expasy.ch/tools/protparam.html). For some genes of *A. thaliana* or *O. sativa*, the active site is indicated in parentheses when it is different from typical ones. The closest homologues in *A. thaliana* and *O. sativa* were found respectively in the MIPS (http://mips.gsf.de/proj/thal/db/index.html) and TIGR (http://www.tigr.org/planProjects.shtml) databases.

<table>
<thead>
<tr>
<th>Group</th>
<th>Names</th>
<th>Gene models</th>
<th>Predicted localization</th>
<th>Redox centres</th>
<th>Sizes/molecular mass (aa/Da)</th>
<th>pl</th>
<th>No. of ESTs</th>
<th>Homologues in <em>A. thaliana</em></th>
<th>Homologues in <em>O. sativa</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>GR</td>
<td>PtcGR1,1</td>
<td>grail3.0008003701</td>
<td>C</td>
<td>CVIRGC</td>
<td>499/54249</td>
<td>5.95</td>
<td>7</td>
<td>At3g24170</td>
<td>Os02g56850</td>
</tr>
<tr>
<td></td>
<td>PtcGR1,2</td>
<td>estExt_Genewise1_v1.C_LG_III0933</td>
<td>P</td>
<td>CVLRGC</td>
<td>498/53804</td>
<td>6.09</td>
<td>4</td>
<td>At3g54660</td>
<td>Os10g28000, Os03g06740</td>
</tr>
<tr>
<td></td>
<td>PtcGR2</td>
<td>eugene3.00150408</td>
<td>ER/C</td>
<td>CPFC</td>
<td>561/60583</td>
<td>7.69</td>
<td>2</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Grx Cxx/S</td>
<td>PtcGrxC1,1</td>
<td>estExt_fgenesh4_pg.C_LG_XV0532</td>
<td>Poptr1_1</td>
<td>C</td>
<td>CGYC</td>
<td>126/13421</td>
<td>8.46</td>
<td>18</td>
<td>At5g63030</td>
</tr>
<tr>
<td></td>
<td>PtcGrxC1,2</td>
<td>eugene3.00120763</td>
<td>Poptr1_1</td>
<td>C</td>
<td>123/13210</td>
<td>8.45</td>
<td>1</td>
<td>At5g40370 (CPYC)</td>
<td>Os02g40500, Os04g42930</td>
</tr>
<tr>
<td></td>
<td>PtcGrxC2</td>
<td>gw1.2484.1</td>
<td>Poptr1_1</td>
<td>ER/C</td>
<td>CPFC</td>
<td>109/11445</td>
<td>6.81</td>
<td>71</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>PtcGrxC3</td>
<td>estExt_fgenesh4_pg.C_LG_VII1147</td>
<td>Poptr1_1</td>
<td>ER</td>
<td>CPYC</td>
<td>134/14838</td>
<td>6.59</td>
<td>12</td>
<td>At1g77370</td>
</tr>
<tr>
<td></td>
<td>PtcGrxC4</td>
<td>estExt_Genewise1_v1.C_LG_XVIII0284</td>
<td>Poptr1_1</td>
<td>ER</td>
<td>CPYC</td>
<td>136/14788</td>
<td>5.77</td>
<td>58</td>
<td>At5g20500</td>
</tr>
<tr>
<td></td>
<td>PtcGrxS12</td>
<td>fgenesh4_pg.C_LG_II002543</td>
<td>Poptr1_1</td>
<td>P</td>
<td>CSYS</td>
<td>184/19911</td>
<td>9.56</td>
<td>5</td>
<td>At4g28730 (CSYC), At2g20270</td>
</tr>
<tr>
<td>Grx CGFS</td>
<td>PtcGrxS14</td>
<td>gw1.XIV.2949.1</td>
<td>Poptr1_1</td>
<td>P</td>
<td>CGFS</td>
<td>170/18681</td>
<td>9.02</td>
<td>29</td>
<td>At3g54900</td>
</tr>
<tr>
<td></td>
<td>PtcGrxS15</td>
<td>estExt_Genewise1_v1.C_LG_III2036</td>
<td>Poptr1_1</td>
<td>M</td>
<td>CGFS</td>
<td>172/19159</td>
<td>7.77</td>
<td>13</td>
<td>At3g15660</td>
</tr>
<tr>
<td></td>
<td>PtcGrxS16</td>
<td>estExt_fgenesh4_pg.C_LG_XVII087</td>
<td>Poptr1_1</td>
<td>P</td>
<td>CGFS</td>
<td>296/32760</td>
<td>7.65</td>
<td>6</td>
<td>At2g38270</td>
</tr>
<tr>
<td></td>
<td>PtcGrxS17,1</td>
<td>grail3.0124002401</td>
<td>Poptr1_1</td>
<td>C</td>
<td>CGFS (X3)</td>
<td>492/53484</td>
<td>5.02</td>
<td>3</td>
<td>At4g04950</td>
</tr>
<tr>
<td></td>
<td>PtcGrxS17,2</td>
<td>eugene3.00040524</td>
<td>Poptr1_1</td>
<td>C</td>
<td>492/53632</td>
<td>5.28</td>
<td>14</td>
<td>No</td>
<td></td>
</tr>
</tbody>
</table>
with *A. thaliana* or *P. trichocarpa*. Indeed, some atypical active sites, differing in the second or fourth position or both, such as CFMC or CPMC, CGMC, CGMS, CCMA, CCLI, and CYMA, are found in *O. sativa* [respective accession numbers Os01g70990, Os12g35340, Os11g43520, Os05g05730, Os01g13950, Os01g47760, and Os01g09830 of The Institute of Genome Research (TIGR)]. Again, these sequences are not restricted to *O. sativa*, since a BLAST search against plant ESTs indicates that similar active site sequences are mostly present in Poaceae such as *Hordeum vulgare*, *Triticum aestivum* or *Zea mays* (data not shown).

In addition to cytosolic isoforms, prediction softwares indicate that N-terminal extensions may target Grx to
As already proposed for Arabidopsis thaliana (Herald & et al., 2002). In the three genomes, all genes of subgroup III contain at least one Grx gene, suggesting that Grx is an essential protein (Table 2 and data not shown). This further confirms that these Grxs constitute independent classes. Interestingly, the Grx S14 gene is composed of three exons in P. trichocarpa, two in O. sativa, and one in A. thaliana. For unknown reasons, this gene is the only one which differs between the three organisms. When present, the N-terminal extension, which is supposed to direct the proteins into the organelles, is included in the first exon (Ptrc Grx C3, C4, S12, S14 and S16), except for Ptrc Grx S15, for which this extension is also present in the second exon.

### Comparison with other sequenced organisms
Among plants, the number of Grx genes is larger in P. trichocarpa (36), compared with A. thaliana (31) and O. sativa (27) (Table 2). This difference mainly arises from a larger subgroup III. In other photosynthetic organisms (C. reinhardtii and Synechocystis sp), the number of genes is lower (six and three, respectively), although C. reinhardtii possesses the four different CGFS-containing Grxs. This strongly suggests that some functions, specific to higher plants, are regulated via specific Grxs. On the other hand, all other genomes analysed in this study, including viruses, contain at least one Grx gene, suggesting that Grx is an essential protein (Table 2 and data not shown). Surprisingly, there are only three Grx genes in mammals and up to five or six genes in some bacteria, yeast, or fungi.

In addition to the plant-specific subgroup III, one way to explain the difference between plant and non-plant genomes or within plant genomes is to look for gene duplications. Interestingly, the Grx C1,2 and 2, Grx C2, and Grx S17,1 or 2) in the cytosol (Table 1). Most of the proteins of subgroup III are extremely small, as they possess only 99–102 amino acids. This suggests that they should be essentially located in the cytosol, unless they can be imported into organelles without an N-terminal extension. Indeed, one of these Grxs (At3g15660), devoid of an N-terminal extension, was found in a mitochondrial proteomic study devoted to the identification of proteins able to bind divalent cations (Herald et al., 2003).

### Gene structure
Among the three sequenced plant genomes, the Grx gene organization is almost conserved (Fig. 3A and data not shown). As already proposed for A. thaliana Trxs, one of the criteria to classify genes into families is to consider the position of introns in the genomic sequences (Meyer et al., 2002). In the three genomes, all genes of subgroup III are encoded by a unique exon, indicating that they belong to a unique class (not shown in Fig. 3). All the genes of subgroup I are made up of four exons and three introns except the closely related genes Grx C5 or S12 (with CSYC/S active sites), which are composed of five exons (Fig. 3A and data not shown). In subgroup II, the situation is a bit more complex; the Grx S15, S16, and S17 genes contain six exons, two exons, and three exons, respectively, whatever the organism considered (Fig. 3A and data not shown). This further confirms that these Grxs constitute independent classes. Interestingly, the Grx S14 gene is composed of three exons in P. trichocarpa, two in O. sativa, and one in A. thaliana. For unknown reasons, this gene is the only one which differs between the three organisms. When present, the N-terminal extension, which is supposed to direct the proteins into the organelles, is included in the first exon (Ptrc Grx C3, C4, S12, S14 and S16), except for Ptrc Grx S15, for which this extension is also present in the second exon.
was performed using the double-affine Smith–Waterman algorithm (Tuskan et al., 2006 in preparation). This analysis indicates that 13 Grx genes, which belong essentially to subgroup III (nine genes), were duplicated in *P. trichocarpa*. Nevertheless, the four other genes, which encode Grx C1,1 or 2 and Grx S17,1 or 2, were only duplicated in poplar but not in *A. thaliana* or *O. sativa*. However, it seems that Grx S15 was duplicated only in rice.

**Grx expression in** *P. trichocarpa*, *O. sativa*, and *A. thaliana*

The expression of Grxs of subgroups I and II was first analysed by identifying among the 375 000 poplar ESTs available in the GenBank database those representing Grxs. All these Grxs are expressed, but two isoforms, Grx C2 and Grx C4, are highly represented with 71 and 58 ESTs, respectively (Table 1). Other genes are expressed at lower levels, from one EST for Grx C1,2 to 29 ESTs for Grx S14 (Table 1).

These data were completed by following *in planta* Grx transcript levels in various plant organs using reverse transcription–polymerase chain reaction (RT–PCR) (Fig. 4). A saturating number of cycles (generally 35 cycles) was used to make sure that the gene products could be detected, preventing quantification and comparison of expression levels between plant organs. Of course, it cannot be excluded that some genes, which were not detected under the experimental conditions used, can be expressed in other growth conditions or can be temporally regulated. This could be addressed, for example, by generating reporter lines using promoter sequences. All the Grxs tested are expressed in mature leaves, flowers, stems, petioles, and fruits, except for Grx S17 (1 or 2) which is only expressed at detectable levels in flowers and petioles, and Grx C4 which is not present in flowers, taking into account the sensitivity of the detection technique employed. On the other hand, only a few genes are expressed in roots (Grx C4 and Grx S12) and in young leaves (Grx C1,2, and Grx S14 and S15). Finally, as a sequence specific enough to discriminate between the two isoforms of Grx S17,1 or 2 could not be found in the cDNA and as ESTs representing these two genes were found, the signal obtained is probably the sum of the expression of the two genes. While all the genes of subgroups I and II are expressed in at least one tissue, the situation is not so clear in subgroup III. Indeed, for some genes, either ESTs could not be found in the GenBank
of 16 exons, whereas three genes were found in *O. sativa*. This difference probably arises from a duplication event in these two species. In poplar, the gene annotated as *At4g15690*, *At1g28480*, and *At3g62950* is also modified by various treatments such as salt, salicylic acid, or ozone applications. They also respond to the application of many chemicals, especially acrylate ester or cycloheximide. Finally, programmed cell death or biotic infections by *Pseudomonas syringae* or *Botrytis cinerea* are also a source of major modifications. In addition, the data described by Mittler and colleagues indicate that the expression of four other genes (*At1g06830*, *At4g15660*, *At4g15700*, and *At5g18600*) is modified and almost exclusively repressed following heat, drought, salt, or high light treatments (Mittler et al., 2004). The products of all those genes, which belong to subgroup III, have not been characterized yet. All other genes and especially the most characterized Grxs of subgroups I and II seem to be poorly regulated at the transcript level in stress situations. This behaviour is slightly surprising as ‘classical’ Grxs were shown to be involved in stress responses in many ways.

**The glutathione reductase family**

GRs are flavoproteins in which electrons and protons are transferred from NADPH to a FAD prosthetic group and finally to a reactive disulphide bridge. The size of these proteins ranges from 497 to 565 amino acids, depending on the presence of an N-terminal extension, which should direct the protein into different subcellular compartments (Fig. 5; Table 1). This N-terminal extension is predicted to be a signal for protein import into plastids. Two classes of proteins can thus be distinguished: GR1 are the shorter cytosolic enzymes, and GR2 are the elongated organellar proteins. The size of the mature proteins is thus ~500 amino acids. As shown in the amino acid sequence alignment presented in Fig. 5, the two cysteines of the GR redox centre are separated by four amino acid residues in the highly conserved motif GGTCV[I/L]RGCVPKK[I/L]LVY. The amino acid sequence identity of GR1 of the three model plants ranges from 71% to 91%, whereas the identity of GR2 oscillates between 64% and 77%. When the two subgroups are compared, identities are lower, comprised between 43% and 51%. Other characteristics are presented in Table 1.

The GR gene structure is well conserved among higher plants, since they are composed either of 10 exons (GR2) or of 16 exons (GR1) (Fig. 3B). Both the exon sizes and the intron positions of homologous genes are well conserved within a single species (for example between GR1.1 and 1.2 from *P. trichocarpa*) but also between plant species (GR1 from *P. trichocarpa*, *A. thaliana*, and *O. sativa*, or GR2 from *P. trichocarpa*, *A. thaliana*, and *O. sativa*). More surprising is the presence of only two genes coding for GR in *Arabidopsis*, belonging to the two categories described above, whereas three genes were found in *O. sativa* and *P. trichocarpa*. This difference probably arises from a duplication event in these two species. In poplar, the gene probably duplicated is composed of 16 exons, whereas that of *O. sativa* is the 10 exon variant. Examination of the various databases [Munich Information Center for Protein Sequences (MIPS), TIGR or Genbank at NCBI] confirmed that all these isoforms are expressed.
A maximum of three GR genes were identified, and only one of them contains a transit sequence. Thus, if it is confirmed experimentally that Grxs are present in nearly all plant compartments, the corresponding reducing system should be expressed in these compartments. One possible explanation is that the transit sequence of GR contains signals for a bipartite localization (chloroplast and mitochondria) as demonstrated in pea (Chew et al., 2003a). This feature is not unusual as other enzymes involved in the ascorbate/glutathione cycle possess similar properties (Chew et al., 2003b). Another possibility is that one sequence devoid of transit peptide could be destined for the organelles. This needs to be proven experimentally, by green fluorescent protein (GFP) fusion experiments for example, but proteomics could also provide an answer to this puzzling question.

Concluding remarks

This study highlights the great diversity of genes coding for Grxs in higher plants by providing evidence that a model tree, poplar, is even more complex than the known...
Acknowledgements

The authors would like to thank Francis Martin, Jerry Tuskan, Uffe Hellsten, and Stephen DiFazio for helpful discussions and help in duplication event analysis.

References


