Evolution and functional differentiation of recently diverged phytochelatin synthase genes from *Arundo donax* L.

Mingai Li1,†, Luca Stragliati2,†, Erika Bellini3, Ada Ricci2,*, Alessandro Saba4,*, Luigi Sanità di Toppi3,*, and Claudio Varotto1,*,

1 Department of Biodiversity and Molecular Ecology, Research and Innovation Centre, Fondazione Edmund Mach, via Mach 1, 38010, San Michele all’Adige (TN), Italy
2 Dipartimento di Scienze Chimiche, della Vita e della Sostenibilità Ambientale, Università degli studi di Parma, Parco Area delle Scienze 11/A, 43125, Parma, Italy
3 Dipartimento di Biologia, Università di Pisa, via Luca Ghini 13, 56126, Pisa, Italy
4 Dipartimento di Patologia Chirurgica, Medica, Molecolare e dell’Area Critica, Università di Pisa, via Roma 67, 56126, Pisa, Italy

† These authors contributed equally to this work.

* Correspondence: luigi.sanita@unipi.it or claudio.varotto@fmach.it

Received 21 August 2018; Editorial decision 24 May 2019; Accepted 24 May 2019

Editor: Hendrik Küpper, Biology Center of the Czech Academy of Sciences, Czech Republic

Abstract

Phytochelatin synthases (PCSs) play pivotal roles in the detoxification of heavy metals and metalloids in plants; however, little information on the evolution of recently duplicated PCS genes in plant species is available. Here we characterize the evolution and functional differentiation of three PCS genes from the giant reed (*Arundo donax* L.), a biomass/bioenergy crop with remarkable resistance to cadmium and other heavy metals. Phylogenetic reconstruction with PCS genes from fully sequenced monocotyledonous genomes indicated that the three *A. donax* PCSs, namely AdPCS1-3, form a monophyletic clade. The AdPCS1-3 genes were expressed at low levels in many *A. donax* organs and displayed different levels of cadmium-responsive expression in roots. Overexpression of AdPCS1-3 in *Arabidopsis thaliana* and yeast reproduced the phenotype of functional PCS genes. Mass spectrometry analyses confirmed that AdPCS1-3 are all functional enzymes, but with significant differences in the amount of the phytochelatins synthesized. Moreover, heterogeneous evolutionary rates characterized the AdPCS1-3 genes, indicative of relaxed natural selection. These results highlight the elevated functional differentiation of *A. donax* PCS genes from both a transcriptional and an enzymatic point of view, providing evidence of the high evolvability of PCS genes and of plant responsiveness to heavy metal stress.

Keywords: Cadmium, divergence, gene duplication, giant reed, phytochelatin synthase, phytochelatins, subfunctionalization.

Introduction

Most transition elements, including the metalloid arsenic, are often collectively defined as ‘heavy metals’ (HMs), due to their high density (Clemens *et al.*, 2002). Some of these elements (e.g. Cu, Mn, Ni, Zn, Fe) are essential nutrients for the majority of organisms, while others, such as Cd, Hg, Pb, and As, lack any known biological role and are toxic even at low concentrations (Tchounwou *et al.*, 2012). HM pollution represents a threat to the environment as well to human populations (Jarup, 2003;
Thompson and Bannigan, 2008). Usually, HMs occur naturally in the environment as trace components of the earth’s crust (Fraústo da Silva and Williams 2001). Their presence as pollutants is therefore most often due to anthropogenic activities, such as mining, motorized transport, and industry (Nagajyoti et al., 2010; Tchounwou et al., 2012). The mechanisms underlying HM toxicity are still not completely understood. What is known is that excessive amounts of essential metals and traces of non-essential metals cause at least two major responses: (i) displacement of the correct cellular cofactors, for example, Cd is able to displace fundamental bivalent ions such as Zn$^{2+}$, Fe$^{2+}$, Cu$^{2+}$, and Ca$^{2+}$; and (ii) a cascade of aberrant reactions, with protein thiol groups binding the metal ions and producing reactive oxygen species (Rea et al., 2004). In the case of plants, toxic HM ions present in the rhizosphere are taken up into cells together with essential metal ions by relatively non-specific plasma membrane transporters (Luo et al., 2016). Among the HMs, possibly the most studied is Cd, given its high toxicity and widespread occurrence in soils worldwide (Tchounwou et al., 2012; Mahar et al., 2016). Uptake of Cd$^{2+}$ from the rhizosphere is the result of ‘hitchhiking’ of this toxic ion on the plasma membrane transporters necessary for the uptake of essential metal ions, especially Mn and possibly Fe in rice (Redjala et al., 2009; Sasaki et al., 2012; Uraguchi and Fujiwara, 2013). Once in the cytoplasm, Cd has to be detoxified as soon as possible to prevent cellular damage. The major mechanism of intracellular Cd detoxification is based on phytochelatins (PCs), a family of cysteine-rich oligopeptides synthesized from glutathione (GSH) by the enzyme phytochelatin synthase (PCS), a γ-glutamylcysteine dipeptidyl transpeptidase (EC 2.3.2.15) (Grill et al., 1989; Vatamaniuk et al., 2004) belonging to clan CA of the papain–like cysteine proteases (Vivares et al., 2005; Romanyuk et al., 2006; Rea, 2012). Elevated HM concentrations in the cytoplasm and the presence of reduced GSH activate the PCS enzyme, readily starting the biosynthesis of PCs. PCs chelate the HM and the HM–PC complex is then transported into the vacuoles, where the HM can be detoxified (Song et al., 2014). PC biosynthesis carried out by PCs represent the main mechanism for plants to detoxify the HMs present in their rhizosphere. PCs and PCS-like genes have been found in a wide range of organisms, from prokaryotes, such as cyanobacteria (Bhargava et al., 2005; Chaurasia et al., 2008), to eukaryotes, such as yeasts (Grill et al., 1985; Ha et al. 1999; Shine et al., 2013), plants, and animals (Clemens et al., 2001; Ray and Williams, 2011; Polak et al., 2014). PCs are evolutionarily conserved in all land plants as well as in charophytes (Fontanini et al., 2018), their sister group, and recent evidence indicates that PC biosynthesis is a plesiomorphic character for plants (Degola et al., 2014; Petraglia et al., 2014). Most of our knowledge of PCS genes derives from Arabidopsis thaliana, whose genome encodes two different PCS genes, AtPCS1 (Vatamaniuk et al., 1999) and AtPCS2 (Cazalé and Clemens, 2001; Kühnlenz et al., 2014). AtPCS1 is the major PCS isoform, and loss-of-function mutations in the AtPCS1 gene render plants extremely sensitive to Cd stress. AtPCS2 is also active, but plays a minor role in Cd detoxification, as the phenotype of AtPCS2 mutants becomes visible only after AtPCS1 has been knocked out (Lee and Kang, 2005; Kühnlenz et al., 2014). More recently, two PCS genes from rice, OsPCS1 and OsPCS2, have been characterized in detail (Li et al., 2007; Das et al., 2017; Hayashi et al., 2017; Uraguchi et al., 2017; Yamazaki et al., 2018). The two isoforms have different specificities for Cd and As, and contribute differentially to detoxification of these HMs (Hayashi et al., 2017; Yamazaki et al., 2018).

Since the discovery of the PCS enzymes, the potential of exploiting them for bioengineering plant detoxification of HMs has attracted much attention. Many attempts to increase resistance to HMs in a number of plant species by overexpression of PCS genes from different sources resulted in a variety of experimental outcomes, ranging from decreased to enhanced resistance to HMs, with various types of relationship (positive, negative, or no relationship) to the accumulation of different HMs. As of 2015, Cd resistance was obtained in roughly one-quarter of the tested PCS-overexpressing lines, while in another quarter of the cases PCS overexpression caused hypersensitivity to HMs (Lee and Hwang, 2015). Early attempts to increase HM detoxification in Arabidopsis by overexpression of AtPCS1, for instance, unexpectedly resulted in higher sensitivity to Cd (Lee, 2003). Apparently, the overexpression of AtPCS1 might have caused as a side effect depletion of the intracellular pool of GSH, which, besides PCs, can play a relevant role in HM chelation (Jozefczak et al., 2012). Indeed, overexpression of both AtPCS and γ-glutamylcysteine synthetase, the enzyme catalyzing the first committed step of GSH biosynthesis, can overcome the problem, leading to enhanced HM resistance (Guo et al., 2008). By contrast, overexpression in Arabidopsis of PCS genes from other plant species directly resulted in higher Cd resistance (Guo et al., 2008; Liu et al., 2012; Fan et al., 2018), indicating that significant functional differences may exist among PCS enzymes. Despite these difficulties, the capacity to modulate HM detoxification through careful adjustments of PC biosynthesis is of extreme relevance for phytoremediation, a relatively new branch of science that uses plants and their associated microbes to reduce the concentration of HMs in soils and freshwaters (Krämer, 2005).

Phytoremediation is now probably one of the most promising technologies at our disposal for the process of decontamination of water and soil from those harmful elements (Oyuela Leguizamo et al., 2017). Good candidate plant species for this technology should ideally be those with fast growth and remarkable biomass yield, as well as high ability to accumulate and inactive HMs (Salt et al., 1998; Suresh and Ravishankar, 2004; Peuke and Rennenberg, 2005). Several hyperaccumulators (i.e. plant species that are able to accumulate HMs in their aerial parts at concentrations two to three orders of magnitude higher than normal plants; van der Ent et al., 2013) are highly tolerant to specific HMs, but they usually grow relatively slowly (Souza et al., 2013), thus limiting their overall effectiveness in phytoremediation. Recently, the giant reed (Arundo donax L.), a perennial rhizomatous grass of the Poaceae family, has been proposed as a potential phytoremediation species, as it is able to accumulate and tolerate high concentrations of HMs such as Ni, Cd, and As without showing major stress symptoms (Papazoglou et al., 2005, 2007; Papazoglou, 2007; Sabeen et al., 2013). The potential use of A. donax for phytoremediation has been confirmed.

Downloaded from https://academic.oup.com/jxb/article-abstract/70/19/5391/5506725 by Fondazione Miani user on 27 February 2020
for a variety of HM concentrations and conditions in soil (Guo and Miao, 2010; Mirza et al., 2011; Alshaal et al., 2013; Barbosa et al., 2015; Fernando et al., 2016; Atma et al., 2017), and especially for wastewaters and polluted aquatic environments (Mirza et al., 2010; Sagehashi et al., 2011; Kausar et al., 2012; Sabeen et al., 2013; Ellhawat et al., 2014; Richveisová et al., 2014). Moreover, A. donax is a vigorous perennial plant that is capable of the high biomass yields necessary to rapidly remove HMs from soils. Although A. donax represents a promising non-crop plant for phytoremediation, to date no molecular characterization of the genes involved in HM detoxification has been carried out in this species. In this work we studied three recently diverged PCS genes of A. donax, and characterized them both in the species of provenance and in transgenic model organisms, in order to dissect their role in the molecular and physiological bases of resistance to Cd. Furthermore, we investigated whether functional diversification could be detected among the gene copies, to gain novel insights into the evolutionary trajectories and retention of recently duplicated PCS genes in plants.

Materials and methods

Plant materials, growth conditions, and stress treatment

Cohorts of Anundo donax L. cuttings (Sesto Fiorentino, Florence, Italy: 43°49′01.8″N, 11°11′37.0″E) and Ambrosia italicana L. Heynh. Col-0 wild-type and transgenic plants were used in this study. The procedures for stress treatments in A. donax were the same as those described previously (Fu et al., 2016; Li et al., 2017). Three biological replicates were used for all the treatments at every sampling time point. Sterilized seeds from AdPCS1-3 transgenic plants were sown on half-strength Murashige & Skoog solid medium supplemented with 1% sucrose, stratified at 4°C for 3 days, and transferred to a growth chamber. After 3.5 days of growth, the young seedlings were transferred to square petri dishes containing the same medium as above (control), or supplemented with 150 µM CdSO₄ (for HM treatment), and grown vertically for a further 10 days. Ten plants for Col-0 and two independent lines from each construct were grown in one plate, and at least 80 plants from each line and Col-0 were analyzed. All plants were grown in long-day conditions (16 h light/8 h dark) in a growth chamber at 23°C with light intensity 100–120 µmol m⁻² s⁻¹ and 40% relative humidity.

Sequence homology searches for novel PCS genes in A. donax

PCS homologs were mined from the A. donax transcriptome (Sablok et al., 2014) using the BLAST+ suite (Camacho et al., 2009). Oryza sativa PCS2 homolog (LOC_Os06g01260) was used as query to perform a BLASTn search, due to its close evolutionary distance from the Ambrosia italicana PCS sequence: Cys56, His162, Asp180). Three complete coding DNA sequences (CDSs) were identified and classified as putative A. donax PCS1, PCS2, and PCS3 (henceforth called AdPCS1, AdPCS2, and AdPCS3).

Genomic DNA extraction and cloning

Genomic DNAs from Anundo accessions/species growing in a growth chamber were isolated using the hexadecyltrimethylammonium bromide method (Doyle and Doyle, 1987). The last intron of AdPCS1, AdPCS2, and AdPCS3 was amplified using Phusion High Fidelity DNA Polymerase (Thermo Scientific) with the primers listed in Supplementary Table S1 at JXB online. After A-tailing with Taq polymerase (Sigma), amplicons were cloned into the pGEM-T vector and at least three clones were bidirectionally sequenced for each amplicon.

Total RNA extraction and real-time PCR analyses

Total RNA extraction and cDNA synthesis were performed as previously reported (Fu et al., 2016). Semi-quantitative RT–PCR was carried out for different organs using the A. donax glyceraldehyde 3-phosphate dehydrogenase (GAPDH) gene as a reference (Li et al., 2017) and AtAct2 for transgenic analysis. For stress treatments, qRT–PCR was conducted with Platinum® SYBR® Green qPCR SuperMix-UDG (Invitrogen) using a putative A. donax actin as a reference gene (Fu et al., 2016) in a Bio-Rad C1000 Thermal Cycler detection system. All reactions for qRT–PCR analyses were performed in triplicate and the 2⁻ΔΔCT method was used to calculate fold changes. Primers used for PCR are listed in Supplementary Table S1.

Phylogenetic reconstruction

A BLASTp search was conducted using AtPCS1 as query against the proteome of all Poaceae and Brassicaceae species present in the Phytozome v.12 database (Goodstein et al., 2012). The resulting homologs were manually checked for the presence of the canonical catalytic triad (Cys56, His162, Asp180) as well as for N- and C-terminal domain integrity. The Phragmites australis PCS sequence was additionally retrieved from the nr GenBank database (Zhao et al., 2014). Sequences were then aligned using the MAFFT program (algorithm E-INS-i, dedicated to proteins with two or more conserved domains; Katoh and Standley, 2013) and processed using GBlocks (Talavera and Castresana, 2007) with the following parameters: minimum number of sequences for a conserved position: 13; minimum number of sequences for a flanking position: 13; maximum number of contiguous non-conserved positions: 8; minimum length of a block: 5; allowed gap positions: with; use similarity matrices: yes. Maximum likelihood (ML) phylogenetic reconstruction was performed with the PhyML 3.0 web server (Guindon and Gascuel, 2003; Guindon et al., 2010). The best-fitting evolutionary model for Poaceae ML reconstruction was evaluated using the Smart Model Selection method (Ronquist et al., 2012), using two independent runs, each with one cold and three heated chains over 5 000 000 generations. The program was allowed to average over the first 10 amino acid rate matrix models by specifying the setting: amodelpr=mixed. Trees were sampled every 1000 generations and posterior probabilities of splits were obtained from the 50% majority rule consensus of the sampled trees, discarding the first 25% as burn-in. Poaceae ML and BI trees were combined using the program TreeGraph 2 (Stöver and Müller, 2010). The phylogenetic reconstruction of the last intron of AdPCS1-3 from different Anundo species/accessions was carried out by ML as described above.

Analyses of molecular evolution

Different programs of the HyPhy package (Kosakovsky Pond et al., 2005) were used for molecular evolution analyses of the codon–aligned AdPCS CDS (502 codons) using the HKY+5 model. GARD (Kosakovsky Pond et al., 2006) was used to infer putative recombination sites with no site-to-site rate variation and two rate classes. The BUSTED program (Murrell et al., 2015) was used for testing gene-wide episodic diversifying selection in all branches of the AdPCS1-3 phylogeny inferred with the neighbor-joining algorithm. In addition, a total of three branches were formally tested for diversifying selection using the aBSREL program (Smith et al., 2015). Significance was assessed using the likelihood ratio test at a threshold of P<0.05, after correcting for multiple testing. Analysis of selective pressures acting on the single alignment sites was carried out with the FUBAR program (Murrell et al., 2013) using a posterior
probability of 0.9 as cutoff. Relaxation/intensification of selection on specific branches of the tree was assessed with the RELAX program (Wertheim et al., 2015).

Molecular dating
Patristic distances (number of substitutions per site) among sequences of the last intron of AdPCS1-3 were calculated with T-REX software (Boc et al., 2012). Representative sequences for each gene were used to estimate divergence time (T) using the formula \( T = K/2r \), where \( r \) is the absolute rate of substitutions per site per year and \( K \) is the estimated numbers of substitutions per site between homologous sequences. The value of \( r \) used (9E-09) corresponds to the rate for the third intron of the Adl gene (Zhu and Ge, 2005), corrected for lineage-specific evolution of Poaceae (Wang et al., 2015).

Yeast complementation assay
Full-length AdPCS1, AdPCS2, and AdPCS3 CDSs were amplified with primers listed in Supplementary Table S1, cloned into pENTR/D-Topo, recombined into the pYES-DEST52 vector (Invitrogen™), and transformed into Cd-sensitive Saccharomyces cerevisiae strain 2452 (ana-3-52 his-3-200, ΔZRCDC3, mating type α) using the lithium acetate method (Gietz and Schiestl, 2007). The culture for each transformant and spotting on YPGL containing supplemented medium was used to determine the growth rate of substitutions per site per year and \( T \) is the external calibration curves and certified glycine-13C2,15N-labelled GSH (Sigma-Aldrich, Saint Louis, MO, USA) and glycine-13C2,15N-labelled PC2 (AnaSpec Inc., Fremont, CA, USA) as internal standards. The calibration curves were built by plotting the peak area ratio (analyte/internal standard) against the concentration ratio (analyte/internal standard). For accurate quantification of thiols-peptides using the present HPLC-ESI-MS-MS method, each sample was run at dilutions of 10^−6, 10^−7, and 10^−8.

Growth kinetic analyses of yeast transformed with AdPCS1-3
Single colonies from each transformant were grown overnight in YSD-U medium and each aliquot with OD600=0.5 was inoculated for further growth in YPGAL medium. During the following 48 h culture at 30 °C, 1 ml aliquots were taken at fixed time intervals (0, 4, 8, 12, 16, 20, 24, 36, and 48 h) and their OD600 measured. All experiments were done in triplicate.

Expression of recombinant protein and PCS assay
The coding sequences of full-length AdPCS1, AdPCS2, and AdPCS3 proteins were amplified from cDNA using primers listed in Supplementary Table S1 and cloned into pET28b in-frame with an N-terminal 6xHis-tag. The expression plasmids were transformed into Escherichia coli Rosetta (DE3) cells, which were induced with 0.5 mM isopropyl β-D-thiogalactopyranoside and cultured overnight at room temperature. The cells were collected by centrifugation and the soluble fraction of recombinant protein was purified as previously described (Pilat et al., 2014), quantified with the Quant-iT Protein Assay Kit (Thermo Fisher Scientific), and verified on 10% SDS-PAGE. The recombinant enzyme activity assay was performed in 100 µl reaction buffer (200 mM HEPES–NaOH pH 8.0, 10 mM β-mercaptoethanol, 12.5 mM GSH, and 100 µM CdSO4) containing recombinant protein (500 ng ml⁻¹) at 35 °C for 60 min, then stopped by adding 25 µl 10% (v/v) trifluoroacetic acid and immediately analyzed for production of PCs with high-performance liquid chromatography–electrospray ionization tandem mass spectrometry (HPLC–ESI-MS-MS; Bellini et al., 2019).

Mass spectrometry analyses of production of PCs by AdPCS1-3 in yeast and in vitro
Single colonies from each yeast transformant line were precultured overnight in YSD-U selective medium. A 500 µl aliquot (OD600=0.5) of each line was inoculated into YPGAL containing 100 µM CdSO4 and 1 mM GSH, and cultured for 24 h. Afterwards, 2 ml of each culture (OD600=2) was pelleted, washed once with sterile water, and resuspended with 200 µl 5% (v/v) sulfosalicylic acid. Cells were ground using glass beads with a tissue lyser, the recovered supernatant was rapidly frozen in liquid nitrogen, and a 10 µl aliquot was used for mass spectrometry (MS) analysis as previously described (Petragna et al., 2014). For recombinant AdPCS1-3 enzymes, thiols-peptides were characterized and quantified by HPLC–ESI-MS-MS using an Agilent 1290 Infinity UHPLC (Santa Clara, CA, USA) with a thermostated autosampler, a binary pump, and a column oven, coupled to an AB Sciex API 4000 triple quadrupole mass spectrometer (Concord, ON, Canada) equipped with a Turbo-V ion spray source (Concord, ON, Canada). Chromatographic separation was performed by using a reverse-phase Waters (Milford, MA, USA) X-Select® HSS T3 2.5 µm XP-C18, 100×3 mm HPLC column, protected by a guard cartridge. Separation was achieved by using a gradient solvent system (solvent A, acetonitrile with 0.1% v/v formic acid; solvent B, water with 0.1% v/v FA) as follows (Bellini et al., 2019): solvent A was set at 2% for 5 min, raised with a linear gradient to 44% in 4.5 min, and then raised with a linear gradient to 95% in 1 min. Solvent A was maintained at 95% for 1 min before column re-equilibration (2.5 min). The flow rate and column oven temperature were set to 300 µl min⁻¹ and 30 °C, respectively. The identification and quantification of thiols-peptides (GSH and PCn) was performed by tandem mass spectrometry (MS/MS) using certified standards (GSH, PC2-4; AnaSpec Inc., Fremont, CA, USA) to build external calibration curves and certified glycine-13C2,15N-labelled GSH (Sigma-Aldrich, Saint Louis, MO, USA) and glycine-13C2,15N-labelled PC2 (AnaSpec Inc., Fremont, CA, USA) as internal standards. The calibration curves were built by plotting the peak area ratio (analyte/internal standard) against the concentration ratio (analyte/internal standard). For accurate quantification of thiols-peptides using the present HPLC–ESI-MS-MS method, each sample was run at dilutions of 10^−6, 10^−7, and 10^−8.

Results
The A. donax genome encodes at least three different PCS gene copies
All three putative AdPCS full-length CDSs isolated contained a single open reading frame, encoding a protein of 502 residues and a predicted molecular weight of ~55 kDa. Once translated, the three proteins showed an average identity with A. thaliana PCS1 (NCBI accession number NP_199220) of 55% and an average similarity of 70%. This level of divergence was higher than those usually observed among alleles of single genetic loci in plants (e.g. Brown et al., 2004; Tatarinova et al., 2016; Zhao et al., 2017). Phylogenetic analyses of the last intron of
AdPCS1-3 isolated from four different *Arundo* species clearly showed that divergence among sequences predated the evolutionary split among species and ruled out recombination (Supplementary Fig. S1). It is thus almost certain that the isolated CDSs are not alleles of a single gene, but different genes encoded in the *A. donax* genome. In line with previous reports (Matsumoto et al., 2004), the N-terminal of the AdPCS1-3 proteins was more conserved than the C-terminal; this was also the case for the rice paralogs OsPCS1 (Os05g0415200) and OsPCS2 (Os06g0102300) (Fig. 1). AdPCS1-3 proteins are significantly less divergent than other duplicated PCSs from previous studies (Supplementary Table S2), displaying only 25–29 substitutions. All the canonical features of PCSs were present, namely the catalytic triad Cys56, His162, and Asp180, and the lengths of both N- and C-terminal domains were comparable to those of previously validated PCSs (Fig. 1). In addition, in the Brassicaceae, we detected a PCS duplication present in the genomes of all species that have been fully sequenced to date (Supplementary Fig. S2), suggesting that retention of PCS duplicates for long evolutionary time spans is a relatively common phenomenon in angiosperms.

Two different statistical methods suggested JTT+G+I (Jones et al., 1992) as the best-fitting evolutionary model for the alignment used for the phylogenetic reconstruction of monocots and *Amborella trichopoda* PCS proteins shown in Fig. 2. Accession numbers for all sequences used are reported in Supplementary Table S3. Using *A. trichopoda* PCS as an outgroup to polarize the tree, *P. australis* PCS was basal to the Panicoideae clade. With the exception of *Zea mays*, which has only one PCS gene (which could be the result of misannotation), all other Panicoideae species had at least two PCS genes, always present in two distinct clades, clearly indicating an early duplication event predating the Panicoideae radiation (Fig. 2). A further species-specific duplication was evident in *Panicum virgatum*. The *A. donax* sequences characterized in this study grouped together at the base of only one of the two Panicoideae clades, indicating that the three gene copies originated after the split of *A. donax* from the other members of the PACMAD clade. The two PCS copies present in *O. sativa* (Yamazaki et al., 2018) clustered together, indicating a lineage-specific duplication independent from the major duplication event in the Panicoideae.

---

Fig. 1. Multiple sequence alignment of *A. donax* and *O. sativa* PCS proteins. Dashes indicate gaps; dots indicate residues identical to those in the sequence of AdPCS1. The amino acids of the catalytic triad (Cys56, His162, and Asp180 in *A. thaliana*, corresponding to positions 57, 163, and 181 in the AdPCS1-3 proteins) are highlighted. (This figure is available in colour at JXB online.)
Expression of the AdPCS 1-3 genes is responsive to HM stress

AdPCS1-3 transcript expression levels were analyzed in both physiological and stressed conditions. In normal conditions, AdPCS1-3 were expressed in all organs analyzed, as indicated by semi-quantitative RT-PCR analyses (Fig. 3). In response to treatment with a concentration of 500 μM CdSO₄, the three transcripts reacted independently and in a tissue-specific manner. In the shoot, expression of the three genes did not change in response to CdSO₄, with the exception of a very modest, although statistically significant, decrease of AdPCS1 expression in the first 90 min of stress application (Fig. 4). By contrast, in root, all three genes showed significant increases of expression at various time points, with a general trend towards up-regulation, especially after 6 h after stress onset. AdPCS1 was the most reactive gene in root, with its expression level reaching 3-fold higher than in the absence of stress. Notably, AdPCS1 transcriptional up-regulation was most prominent in root between 11 and 24 h after stress application, whereas in shoot the gene was either unresponsive or mildly down-regulated by the presence of Cd²⁺ at these time points (Fig. 4). AdPCS3 responded so little to CdSO₄ treatment in absolute terms (maximal fold change 1.5) that its expression can be considered practically not inducible by HM.
Expression of AdPCS1-3 was confirmed by semi-quantitative RT–PCR in two A. thaliana transgenic lines per construct representing independent transformation events (Supplementary Fig. S3). No differences in growth could be detected in the absence of Cd among A. thaliana Col-0 untransformed controls and transgenic lines transformed with AdPCS1-3 CDSs. Independent lines overexpressing either the AdPCS2 or AdPCS3 CDS exhibited a significant reduction in growth (measured as the fresh weight of the aerial part) compared with Col-0 when treated with 150 µM CdSO₄ in the growth medium (Fig. 5A, B). In addition, chlorosis of transgenic plants overexpressing AdPCS2 or AdPCS3 was clearly visible (Fig. 5B). By contrast, the transgenic plants overexpressing AdPCS1 closely resembled the Col-0 ecotype and did not display any discernible phenotype in terms of fresh weight or pigmentation.

Fig. 3. Expression pattern of AdPCS1-3 in different organs/tissues of A. donax as demonstrated by semi-quantitative RT–PCR. Expression levels of AdPCS1-3 transcripts were measured in physiological, unstressed conditions. The AdGAPDH gene was used as a normalization reference, using 26 PCR cycles, while 33 cycles were used for AdPCS1-3.

Fig. 4. Responsivity of AdPCS1, AdPCS2, and AdPCS3 to Cd stress. Time-course variation by real-time PCR of AdPCS1-3 expression in A. donax roots (A) and shoots (B) in response to HM stress resulting from treatment of the root system with 500 µM CdSO₄. The Y-axis reports variations in transcript level compared with the non-stressed condition (0 h, indicated by the continuous line at Y=1). Bars indicate SD (n=3 biological replicates); asterisks indicate statistically significant differences from untreated controls (P<0.05, t-test corrected with false discovery rate).
Expression of AdPCS1-3 in yeast confers Cd resistance

AdPCS1-3 CDSs subcloned in the pYES-DEST52 vector were transformed into the S. cerevisiae YK44 highly Cd sensitive strain, previously used for the same purpose by Zhao et al. (2014). We observed that the growth of AdPCS1 transformants was delayed compared with AdPCS2-3 transformants. To further investigate the basis of this difference, a growth analysis was conducted on YPGAL media without any HM stress (Supplementary Fig. S4). The results clearly showed that the AdPCS1 transformant had a longer lag phase, resulting in slower initial growth. This difference was rescued after 24 h, with the growth kinetics of the AdPCS1 transformant becoming normal, indicating that the difference resulted from decreased translational efficiency in the

Fig. 5. Phenotype of A. thaliana plants overexpressing AdPCS1-3 in the presence of Cd. (A) Phenotype of two independent transgenic lines per construct and the control untransformed line (Col-0) grown on vertical plates for 10 days without (top) or with (bottom) the addition of 150 μM CdSO₄. (B) Fresh weight of the plants shown in A. Bars indicate the SD of n=3 biological replicates; asterisks indicate statistically significant differences from the untreated controls (P<0.05, t-test corrected with false discovery rate). (This figure is available in colour at JXB online.)

Fig. 6. Growth of yeast cells expressing AdPCS1-3 CDSs. The four yeast strains, transformed with AdPCS1-3 CDSs or the empty vector, were plated in serial dilutions (dilution factors are indicated above the image) in either the absence or presence of 100 μM CdSO₄. Four independent replicates were performed; the picture shows a representative example of one replicate.
yeast heterologous system. Comparison of OD-normalized cultures showed that AdPCS1-3 transformants acquired equivalent levels of Cd resistance in the presence of 100 μM CdSO₄ (Fig. 6). This result indicates that the three different CDSs code for functional PCSs and that A. donax possesses at least three different genes for PCS enzymes.

AdPCS1-3 produce phytochelatins in yeast and in vitro

To prove that Cd resistance was due to PC synthesis, MS analyses were carried out. As expected, all three PCS enzymes produced detectable amounts of PCs in yeast (Supplementary Fig. S5). There was significant heterogeneity in the mean quantities of the major PC polymerization forms (PC2 to PC5) produced by the strains overexpressing different AdPCS genes (one-way ANOVA, $F_{2,6}=5.149, P=0.042$). We attempted, without success, to normalize the amounts of enzyme in the AdPCS1-3 transformants by western blotting using an antibody raised against Arabidopsis PCS1 (data not shown). We thus could not rule out the possibility that the cause of the differences in total PCs among yeast transformants could lie in differences in the translational efficiency of the AdPCS transgenes in the yeast heterologous system (see above).

Recombinant AdPCS1-3 proteins purified from E. coli (Supplementary Fig. S6) were further assayed in vitro to compare their enzymatic activity in the absence of confounding effects due to in vivo differences in expression. The specific activities of AdPCS1-3 enzymes for the major PC polymerization forms (PC2 to PC4, either as gross total or individually) were significantly heterogeneous (one-way ANOVA; for total PCs: $F_{2,12}=98.35, P=3.62E-05$; for PC2: $F_{2,12}=49.2, P=1.65E-06$; for PC3: $F_{2,12}=46.6, P=2.34E-03$; for PC4: $F_{2,12}=98.39, P=3.61E-05$). Significant differences were detected among the amounts of both total PCs and single polymerization forms produced in all pairwise comparisons among enzymes (Tukey’s post-hoc test; total PCs: $P<0.00024$; PC2: $P<0.00304$; PC3: $P<0.00234$; PC4: $P<0.01109$ for all comparisons; Fig. 7). The total PCs produced ranged (mean ± SD) from 10.43±0.29 to 15.47±0.73 nmol mg protein⁻¹ min⁻¹. The rank in terms of total production was AdPCS1>AdPCS2>AdPCS3 (Fig. 7). Only minor differences in the spectrum of PCs produced (as referred to PC1 for each enzyme) were detected.

AdPCS1-3 genes evolve at different evolutionary rates

To shed additional light on the functional differences observed among the AdPCS1-3 genes and reconstruct their evolutionary history, we carried out a series of analyses on the patterns of molecular evolution of their sequences. The three genes did not show significant evidence of either gene-wide (likelihood ratio test, $P=0.859$) or branch-specific (likelihood ratio test, $P=1.00$ for AdPCS1, 1.00 for AdPCS2, 0.15 for AdPCS3) episodic diversifying selection.

Therefore, there is no evidence that any sites have experienced diversifying selection along the test branch(es). A putative break-point, indicative of either rate variation or topological incongruence due to recombination, was identified (position 296 of the alignment; corrected Akaike Information Criterion score of the best fitting GARD model: 5362.96; AICc score of null model: 5365.01). The Kishino–Hasegawa topological incongruence test resulted negative (Shimodaira–Hasegawa test applied to the left partition, $P=0.997$; applied to the right partition, $P=1.00$), suggesting that the different domains of AdPCS genes underwent evolutionary rate variation rather than recombination. Two additional findings support evolutionary rate heterogeneity among AdPCS sequences. First, episodic positive/diversifying selection was detected at four sites, all of them located at the C-terminus of the proteins (Supplementary Table S4), while the 29 sites under episodic negative/purifying selection were distributed throughout the alignment. Second, the sequences of AdPCS1 and AdPCS2 displayed statistically significant relaxation ($K=0.10, P=0.030$, likelihood ratio=4.72) compared with AdPCS3 (Fig. 8). Given such evolutionary rate variation, CDS could not be reliably used for estimation of duplication time. Theoretically more reliable estimations of divergence times from nucleotide divergence of the AdPCS1-3 last introns varied depending on the pairs of genes used (Supplementary Tables S5 and S6), providing an approximate time range for AdPCS1-3 duplications. These and the results described above (i.e. phylogeny, transcriptional responsiveness to Cd) allowed reconstruction of the plausible events leading to evolution of the three homologous PCS genes characterized in this study. According to the evolutionary model shown in Fig. 9A, we propose that two rounds of polyploidization happened in rapid succession, in a time range between ~8.5±0.1 and ~11.3±0.1 million years ago. The first duplication generated PCS ancestral copies A and B. While PCS B evolved by mutation/drift into present-day AdPCS3, the second duplication of PCS A gave rise to AdPCS1 and AdPCS2. During the whole process, functional divergence among the resulting copies involved the changes in gene transcription, enzyme activity, and evolutionary rates that characterize the present-day AdPCS1-3 genes (Fig. 9B).
Fig. 8. Relaxation of selective pressure acting on AdPCS1 and AdPCS2. The distribution of $\omega$ (omega ratio of non-synonymous to synonymous substitution rates) across alignment sites is shown by the black bars for the null model in which the same $\omega$ distribution is assumed for the AdPCS3 (reference) and AdPCS1-2 (test) branches. The grey bars represent the $\omega$ distribution for the alternative model in which different $\omega$ distributions are assumed for the reference and test branches. The arrows indicate the direction of $\omega$ classes variation from the null model to the alternative model distribution. The shift observed towards $\omega$ values closer to neutrality ($\omega=1$, vertical dotted line) indicates relaxation of the selective pressure acting on AdPCS1 and AdPCS2 compared with AdPCS3.

Fig. 9. Evolutionary model and summary of functional diversification among AdPCS1-3. (A) Scheme of the two rounds of duplication giving rise to AdPCS1-3. (B) Functional features of AdPCS1-3 with respect to transcriptional up-regulation by Cd, total amount of PCs produced, and the presence of reduced growth and a chlorotic phenotype when overexpressed in A. thaliana, and relaxation of the evolutionary constraints of purifying selection acting on the single genes. – indicates (near) absence and + indicates presence, with more + indicating more pronounced features. (This figure is available in colour at JXB online.)
Discussion

Given the reported high resistance of A. donax to Cd and other HMs (e.g. Papazoglou et al., 2007; Mirza et al., 2011) and its relevance as a biomass/bioenergy crop (Angelini et al., 2009), this species has been repeatedly suggested as a strong candidate for phytoremediation (Fernando et al., 2016). To date, however, the molecular bases for the high resistance of A. donax to HMs has not been functionally investigated in detail. The fundamental role of the PCS genes in HM detoxification is well established in a number of higher plants (Clemens, 2006; Yadav, 2010); thus, the characterization of the transcriptional and enzymatic activity of the three putative PCS genes carried out in this study constitute an important step towards the dissection of the mechanisms underlying the resistance of A. donax to HMs. The results obtained demonstrate that all the AdPCS isoforms characterized are fully functional, as they are transcribed in various A. donax organs (Fig. 3), can confer enhanced resistance to Cd in yeast (Fig. 6), and they all synthesize PCs (Fig. 7, Supplementary Fig. S5). However, they also display enhanced resistance to Cd in yeast (Fig. 6), and they all synthesize PCs (Fig. 7, Supplementary Fig. S5). However, they also display marked functional differences in the amount and, to a lower extent, levels of polymerization of PCs they produce, as well as in their transcriptional regulation upon exposure to HM stress (Figs 4 and 7). In addition, PCS paralogs previously characterized in other species (Lotus japonicus, O. sativa, A. thaliana, and Morus notabilis) were found to display functional diversification in terms of the amount of PCs produced and the specificity of metal-mediated activation, as well as differential regulation of transcription in response to HMs (Lee and Kang, 2005; Blum et al., 2007; Ramos et al., 2008; Moudouma et al., 2011; Kühnlenz et al., 2014; Fan et al., 2018; Yamazaki et al., 2018). In particular, the recent work carried out in O. sativa var. japonica clearly demonstrates that OsPCS1 and OsPCS2 have specialized roles in Cd and As resistance. OsPCS2 is the major isozyme for biosynthesis of PCs and is the most responsive to Cd (Hayashi et al., 2017; Yamazaki et al., 2018), while OsPCS1 plays a major role in allowing the accumulation of As in rice grains but is also important in resistance to both As and Cd (Li et al., 2007; Hayashi et al., 2017; Uraguchi et al., 2017). This high level of functional divergence parallels the medium to high sequence divergence of PCS paralogs from these species (52–54% identity between LjPCS1 and LjPCS2/LjPCS3; 72% identity between OsPCS1 and OsPCS2; 84% identity between AtPCS1 and AtPCS2) compared with that among AdPCS1-3 (94–95%), indicating that they originated from more ancient duplication events than AdPCS1-3. For instance, two PCS paralogs were found in the genomes of all Brassicaceae analyzed in our study (Supplementary Fig. S1), and functional AtPCS1 and AtPCS2 orthologs have been characterized in Arabidopsis thaliana and Noccaea caerulescens (formerly named Thlaspi caerulescens; Meyer et al., 2011). The AtPCS1 and AtPCS2 paralogs thus originated from a duplication preceding the divergence between the Arabidopsis and Noccaea genera, which has been estimated to have taken place in the Oligocene about 27 million years ago (Huang et al., 2016). Unfortunately, the lack of a reference genome makes it difficult to infer the timing and mechanism of the AdPCS1-3 duplication. The best estimates we could obtain place both duplication events at roughly 9–12 million years ago, corroborating the notion that the AdPCS1-3 duplications are more recent than those of previously characterized PCS genes. The increased relaxation of selective pressure, closer phylogenetic relationship of AdPCS1 and AdPCS2 compared with AdPCS3, and the overall low sequence divergence among AdPCS1-3 proteins together suggest that two rounds of duplication took place in close succession: the first duplication gave rise to two ancestral PCS copies (PCS A and B in Fig. 9), followed relatively soon after by a second duplication responsible for the evolution of AdPCS1 and AdPCS2 from the ancestral PCS A copy. Another example of such an evolutionary pattern has been reported for LjPCS1-3, but in this case the second duplications happened much later than the first (Ramos et al., 2007). Given the transcriptional responsiveness of PCS genes to HMs in other species (e.g. Moudouma et al., 2012; Yamazaki et al., 2018), the ancestral PCS existing before both duplications may have been up-regulated by Cd stress, like the present-day AdPCS1 and AdPCS2 genes; thus, we propose that the PCS B ancestral copy underwent subfunctionalization of regulatory promoter elements after divergence. This scenario is corroborated by the fact that stress-responsive genes tend to be over-represented among paralogous pairs with different expression levels and reduced conservation of cis-acting regulatory elements (Hoffmann and Palmgren, 2016). Interestingly, transcriptional differences, with one copy dominant over the other, seem to be a relatively common feature of duplicated PCS genes across different plant families and originating from independent duplication events (e.g. Lee and Kang, 2005; Fan et al., 2018; Yamazaki et al., 2018), suggesting that transcriptional divergence may be a relevant mechanism for long-term retention of multiple PCSs. In addition to non-coding regulatory regions, subfunctionalization is known to be a common outcome for protein-coding sequences of retained duplicates (Hoffmann and Palmgren, 2016). Even in the absence of pseudogenization, duplicated genes can undergo functional decay for millions of years simply due to genetic drift (Panchy et al., 2016). The presence of several sites under purifying selection and some under intensified/positive selection (Supplementary Table S4), however, indicates that this is not the case for the AdPCS1-3 genes, further suggesting that loss of any of the three gene copies, including AdPCS1, would have a detrimental effect on plant fitness (Panchy et al., 2016). It is noteworthy that the relatively large functional variation among AdPCS isoforms required a limited number of amino acid substitutions (<30) to take place (Supplementary Table S2). The fact that the majority of the substitutions took place in the C-terminal domain of the protein suggests that regulation rather than core catalytic activity is affected, in agreement with previous reports demonstrating that this fast-evolving domain is responsible for metal sensing and enzyme stability (Cobbett and Goldsbrough, 2002; Ruotolo et al., 2004).

Taken together, these results strongly suggest that AdPCS1, AdPCS2, and AdPCS3 most likely contribute to Cd detoxification in A. donax. The presence of multiple PCS copies seems to be advantageous as on the one hand redundancy provides higher overall levels of PC biosynthesis, while on the other hand functional specialization provides increased flexibility in HM resistance (Panchy et al., 2016). According to the
ubiquitous and conserved level of constitutive expression of the three isozymes in all A. donax organs/tissues but the lower PC biosynthetic activity of AdPCS3, it seems that AdPCS1 and AdPCS2 play a major role in basal detoxification of Cd and possibly other HMs. Transcriptional inducibility by Cd further suggests that AdPCS1 and AdPCS2 provide the most relevant contribution to HM detoxification in roots at high metal concentrations (Fig. 4). A. donax accumulates the majority of absorbed Cd in the roots and rhizomes (Sagehashi et al., 2011; Yu et al., 2018), implying that below-ground organs act as the main centers of bioaccumulation (Bonanno, 2012). AdPCS1 and AdPCS2 thus likely contribute to avoidance of metal toxicity by preventing Cd sequestration in non-photosynthetic tissues. While the reduced growth observed for AdPCS2- and AdPCS3-overexpressing plants in Arabidopsis is in line with former reports (Lee, 2003), the reason why overexpression of AdPCS1 does not cause any visible phenotype under the conditions tested is currently not clear. Although the transcriptional levels of the AdPCS1 transgene were similar to those of AdPCS2 and AdPCS3, the amount of AdPCS1 protein expression in this heterologous system may be limited by translational efficiency, analogous to what was detected in yeast. Further studies will be needed to investigate this point.

More generally, the AdPCS1-3 triplication is, to the best of our knowledge, by far the most recent set of paralogs functionally characterized in plants until now. The results obtained therefore provide novel insights into the mechanisms of functional diversification that, ultimately, are responsible for long-term retention of PCS duplicates in plants. In fact, they demonstrate that functional diversification happens relatively early after duplication and entails a small number of amino acid changes in the C-terminal domain of PCS, suggesting a mechanism underlying the flexibility of species with multiple PCSs to adapt to HM stress.

**Supplementary data**

Supplementary data are available at JXB online.

Table S1. Primers used in this study.

Table S2. Amino acid divergence among PCS duplicates from different species.

Table S3. List of species and PCS accession numbers used for phylogenetic reconstruction.

Table S4. Codons of AdPCS1-3 under episodic positive/ divergising selection or episodic negative/purifying selection as detected by FUBAR, software.

Table S5. Summary of sequences of the AdPCS1-3 last intron from different Arundo species/accessions.

Table S6. Estimated divergence times among AdPCS genes.

Fig. S1. Maximum likelihood phylogenetic tree of the AdPCS1-3 last intron from four Arundo species.

Fig. S2. Bayesian inference phylogenetic tree of PCS proteins present in Brassicaceae with fully sequenced genomes.

Fig. S3. Semi-quantitative RT–PCR of AdPCS1-3 transcription levels in two A. thaliana transgenic lines.

Fig. S4. Liquid growth kinetics of yeast strains overexpressing AdPCS1-3 CDS in non-stressed conditions.

Fig. S5. Total production of PCs (PC2–5) in yeast lines overexpressing AdPCS1, AdPCS2, or AdPCS3 CDS.

Fig. S6. SDS-PAGE of recombinant AdPCS1-3 enzymes purified from E. coli.

Fig. S7. Summary and sequence alignment of the AdPCS1-3 last intron from different Arundo species.

**Acknowledgements**

This work has been funded by the Autonomous Province of Trento through core funding of the Ecogenomics group of Fondazione E. Mach. LS was the recipient of a PhD fellowship from the University of Parma. This work was partly supported by MIUR–PRIN 2015 funds (prot. 20158HTL58, PI Prof. Luigi Sanità di Toppi) and by the Evolutionary Biology and Ecology PhD course of the University of Parma. The authors thank Enrico Barbaro (Fondazione E. Mach) for technical support in sampling of A. donax material, Andrea Faccini (Centro Interdipartimentale Misure ‘G. Casnati’, University of Parma) for assistance in mass spectrometry analyses, Stefania Pilati for help in recombinant protein purification, Massimo Pindo for technical support in Sanger sequencing, and Dietrich H. Nies for the kind gift of the YK44 yeast strain.

**Conflict of interest**

The authors declare no conflict of interest.

**References**


Shine AM, Shakya VPS, Idrunn A. 2015. Phytochelatin synthase is required for tolerating metal toxicity in a basidioomyete yeast and is a conserved factor involved in metal homeostasis in fungi. Fungal Biology and Biotechnology 2, 3.
Song WY, Mendoza-Cózatl DG, Lee Y, Schroeder JL, Ahn SN, Lee HS, Wicker T, Martinoia E. 2014. Phytochelatin-metal(loid) transport into...
vacuoles shows different substrate preferences in barley and Arabidopsis.


