



HAL
open science

Global analysis of non-animal peroxidases provides insights into the evolutionary basis of this gene family in green lineage

Duchesse Lacour, Qiang Li, Philippe Ranocha, Yves Martinez, Christophe Dunand

► **To cite this version:**

Duchesse Lacour, Qiang Li, Philippe Ranocha, Yves Martinez, Christophe Dunand. Global analysis of non-animal peroxidases provides insights into the evolutionary basis of this gene family in green lineage. *Journal of Experimental Botany*, 2020. <hal-03372634>

HAL Id: hal-03372634

<https://hal.science/hal-03372634v1>

Submitted on 11 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

Global analysis of non-animal peroxidases provides insights into the evolutionary basis of this gene family in green lineage

Duchesse Lacour Mbadinga Mbadinga^{1,§}, Qiang Li^{1,2,§}, Philippe Ranocha¹, Yves Martinez³, Christophe Dunand^{1,*}

¹ Laboratoire de Recherche en Sciences Végétales, Université de Toulouse, CNRS, UPS, Toulouse, France.

² Citrus Research Institute, Southwest University/Chinese Academy of Agricultural Sciences, Beibei, Chongqing 400700, PR China.

³ Fédération de Recherche 3450, Plateforme Imagerie, Pôle de Biotechnologie Végétale, Castanet-Tolosan 31326, France.

[§] Authors have contributed equally to this work.

***Corresponding author: Christophe Dunand**

E-mail: dunand@lrsv.ups-tlse.fr

Tel: +33 (0) 5 34 32 38 58

21 **Abstract**

22 The non-animal peroxidases belong to a superfamily of oxidoreductases that reduce the
23 hydrogen peroxide and oxidize numerous substrates. Since their initial characterization in
24 1992, several advances have provided an understanding into the origin and evolutionary
25 history of this family of proteins. Here, we report for the first time an exhaustive evolutionary
26 analysis of non-animal peroxidases using integrated *in silico* and biochemical strategies.
27 Thanks to the availability of numerous genomic sequences from many species belonging to
28 different kingdoms together with expert and exhaustive annotation of peroxidase sequences
29 centralized in a dedicated database, we have deepened our understanding of the evolutionary
30 process underlying non-animal peroxidases through phylogenetic reconstructions. We
31 analysed the distribution of all non-animal peroxidases in more than 200 eukaryotic
32 organisms *in silico*. First, we show that the presence or absence of non-animal peroxidases
33 can be correlated with the presence or absence of certain organelles or with specific
34 biological processes. Examining a wide range of organisms, we confirmed that ascorbate
35 peroxidases (APx) and cytochromes c peroxidases (CcP) were detected respectively in
36 chloroplast and mitochondria containing organisms. Plants, which contain both organelles,
37 are an exception and contain only APxs without CcP. Class II peroxidases (CII Prx) only
38 found in fungi related to wood decay and plant degradation and Class III peroxidases (CIII
39 Prx) only detected in both Streptophytes algae and land plants, have been subjected to large
40 family expansion in the latter group. Moreover, we demonstrate that biochemical activities
41 (APx, CcP and CIII Prx) assayed in protein extracts obtained from 30 different eukaryotic
42 organisms strongly support the distribution of the sequences resulting from our *in silico*
43 analysis. The biochemical results confirmed both the presence and classification of non-
44 animal peroxidase encoding sequences.

45

46 **Key words:** Class I peroxidase, Class III peroxidase, non-animal peroxidase, ancestral genes,
47 *Spirogyra sp.*

48

49 **Introduction**

50 The increase of oxygen level, 2.3 billion years ago at the beginning of the Proterozoic,
51 forced the ancestral organisms to adapt to oxidative stress (Becker *et al.*, 2004). The reactive
52 oxygen species (ROS) produced are necessary to cells, but toxic, at high levels, and are
53 consequently managed by an ROS gene network (Mittler *et al.*, 2004). Peroxidases are
54 components of this network that regulate ROS homeostasis through oxidation-reduction
55 reactions using hydrogen peroxide (H_2O_2) as an electron acceptor along with different
56 oxidized substrates as electron donors. The term “peroxidase” encompasses two major
57 protein families that differ based on in the presence or absence of a haem, namely "haem
58 peroxidase" and "non-haem peroxidase". The haem peroxidases are the most diverse group in
59 terms of taxonomic distribution, and the non-animal peroxidase family belongs to this group
60 (Passardi *et al.*, 2007a). This family was first described by Welinder in 1992 based on
61 structural homology (Welinder, 1992) and comprises three peroxidase classes: Class I (CI
62 Prx), Class II (CII Prx) and Class III (CIII Prx) (Fig. 1A). This large superfamily is regrouped
63 under a unique PFAM (PF00141), which described the conserved peroxidase domain.
64 However, this grouping does not discriminate between the members of these three classes and
65 can produce mis-annotation which needs to be corrected by manual annotation.

66 The CI Prxs are widely distributed in prokaryotes and eukaryotes and were originally
67 described as intracellular peroxidases (lacking signal peptide). They are composed of an
68 unglycosylated haem, without calcium ions neither disulphide bridges. This class is the most
69 widespread of the peroxidases and is subdivided into four subclasses: Catalase Peroxidases
70 (CPs), Cytochrome c Peroxidases (CcPs), Ascorbate Peroxidases (APxs) and Hybrid
71 Ascorbate Peroxidase-Cytochrome c Peroxidases (APx-CcPs). CPs are predominantly found
72 in bacteria and only occasionally in other organisms due to horizontal gene transfer (Passardi
73 *et al.*, 2007b). All CPs are characterized by the presence of two independent peroxidase
74 domains. CcPs are present in organisms with mitochondria but are absent from plants. CcPs
75 catalyse the oxidation of ferricytochrome c in the presence of H_2O_2 (Finzel *et al.*, 1984) in
76 order to remove toxic oxygen generated during the mitochondrial respiratory process (Kwon
77 *et al.*, 2003). APx are specific to chloroplast-containing organisms with the notable exception
78 of cyanobacteria (Miyake *et al.*, 1991). They catalyse the oxidation of ascorbate in the
79 presence of H_2O_2 for photoprotection and regulation of photooxidative stress (Maruta *et al.*,
80 2010). The last CI Prx subclass corresponds to proteins with common characteristics to both
81 APx and CcP, and has been named hybrid Ascorbate Peroxidase-Cytochrome c Peroxidase

82 (APx-CcP). A predominant example of this class is found in *Leishmana major* for which both
83 APx and CcP activities have been detected (Adak and Datta, 2005).

84 The CII Prxs are described as secreted fungal lignin peroxidases. They belong to large
85 multigenic families and can oxidize molecules with high redox potential such as lignin. CII
86 Prxs include lignin peroxidases (LiP), manganese peroxidases (MnP), and versatile
87 peroxidases (VPs), so far only described in fungi involved in the degradation of wood
88 (Welinder et al. 1992). More recently, six other classes have been identified in Ascomycetes
89 (Asco class II type A, B and C) and in Basidiomycetes (Basidio class II type A, B and C)
90 (Mathé *et al.*, 2019). These new classes have been primarily found in plant pathogenic and
91 plant degrading fungi.

92 The CIII Prxs also known as plant secreted peroxidases have been described as
93 potential multifunctional proteins, which can participate in auxin metabolism, cell wall
94 elongation, stiffening and protection against pathogens (Hiraga *et al.*, 2001; Passardi *et al.*,
95 2004). However, the precise *in vivo* role of any given single peroxidase is still marginal,
96 primarily because of the wide range of peroxidase substrates (lignin subunits, lipid
97 membranes, and some amino acid side chains (Lazzarotto *et al.*, 2011)) and the probable
98 functional redundancy of this large multigenic family that is highly conserved in terms of
99 amino acid sequence. CIII Prxs have been detected in both green algae that predate land
100 colonization and in land plants, but they are absent from Chlorophyte algae (Mathé *et al.*,
101 2010; Passardi *et al.*, 2007a). This family was subjected to numerous genes, and segmental or
102 genome-wide duplications leading to a large increase in copy numbers from the probable first
103 non-aquatic organisms (Bryophyte). For example, late diverging plants such as *Arabidopsis*
104 *thaliana* (Tognolli *et al.*, 2002), and *Oryza sativa* (Passardi *et al.*, 2004) contain high copy
105 numbers generated by gene duplications (73 and 138, respectively). These explosive
106 duplications make the genes ideal markers for the mapping of evolutionary lineages in plants.

107 Unlike CI Prxs, CII and CIII Prxs have more complex structures marked for several
108 glycosylations, four to five disulfide bridges, two calcium binding sites, and one signal
109 peptide (Martinez, 2002). The genes encoding these two classes containing several introns at
110 conserved positions and were subjected to numerous recent duplication events. Gene
111 duplications, together with intronic sequences, play important roles in organism evolution.
112 Intronic sequences are required for various processes such as alternative splicing, translation
113 efficiency and contain regulatory sequences, encoding other genes or non-coding RNAs
114 (Fawcett *et al.*, 2012). The sequences of new paralogs may diverge in regulatory regions,

115 changing its expression patterns or in coding regions may result in acquisition of new
116 functions leading to either subfunctionalization or neofunctionalisation, (Xu *et al.*, 2012). If
117 the duplicated sequences have no additional function, the accumulation of mutations will
118 rapidly lead to pseudogenisation.

119 To reveal the emergence and analyse the evolution of the CIII Prxs in green plant
120 lineage, we annotated the CI Prxs (APx, CcP and CP) and CIII Prxs families from publically
121 available genomic and expression data. Particular attention was paid to cover most of the
122 green lineage from the early to late divergent plants. In parallel, we confirmed the presence
123 and absence of the three classes of proteins based on APx, CcP and CIII Prxs enzymatic
124 activities. Finally, we focused on the evolution of the CIII Prx copy number in the green
125 lineage, with a particular interest to the CIII Prxs from *Spirogyra sp.*

126

127 **Results and Discussion**

128 **The 3D structure is strikingly conserved between CI, CII and CIII Prxs**

129 The three subclasses CPs, APxs, and CcPs which belong to the CI Prxs have very
130 similar tertiary structures, including a conserved number of α -helix domains and overall
131 organization (Fig. 1B). The CPs which contain two independent peroxidase domains,
132 consequently as many α -helices. In addition, residues proximal to the haem such as arginine,
133 as well as distal residues such as histidine, are conserved at the same relative positions among
134 the 3 CI Prxs subclasses (Fig. 1C). These structural similarities in term of common haem,
135 iron, key residues and structure conservations, strongly suggest that the CPs, APxs, and CcPs
136 originated from a common ancestral sequence. In addition, the comparison between the
137 tertiary structures of the three classes part of the non-animal peroxidases (CI, CII and CIII
138 Prxs) revealed similarly shared common structural elements including haem, iron, and key
139 residues and structural conservations, strongly suggesting that the three Prx classes could
140 have originated from the same ancestral sequence (Fig. 1) (Lazzarotto *et al.*, 2015; Passardi *et*
141 *al.*, 2007a).

142 Although CII Prxs are not the primary focus of this study, it was necessary to gain
143 insight into the evolution of this class in order to obtain a more global appreciation of the
144 evolution of all non-animal peroxidases. The high duplication rate of CII Prxs, as well as and
145 the absence of CII Prxs in organisms like Chytrids at the base of fungi and even in some
146 species of basidiomycetes, hinder the clear determination of the evolutionary basis behind

147 this class of peroxidases (Mathé *et al.*, 2019). However, the functional similarity and tertiary
148 structure conservation of CII Prxs with CI Prxs suggest that the CII Prx sequences could have
149 derived from CI Prxs. Therefore, it is likely that the ancestor of the CII Prxs had duplicated
150 and emerged independently during the separation of Basidiomycetes and Ascomycetes to
151 give rise to these different subclasses of CII Prxs (Mathé *et al.*, 2019).

152

153 **CI, CII and CIII Prxs have specific taxonomic distribution**

154 Several taxonomic distributions of CI, CII and CIII Prxs have already been described
155 (Lazzarotto *et al.*, 2015; Passardi *et al.*, 2007a), but more recent advances in available
156 comprehensive genomic resources (Mukherjee *et al.*, 2018), allowed us to perform exhaustive
157 data mining and expert sequence annotation in order to define the evolutionary basis giving
158 rise to the actual taxonomic distribution of the non-animal Prxs in the different kingdoms.

159 To do this, we performed data mining for peroxidase sequences and organelle gain or
160 loss events from different eukaryotic taxonomic groups that yielded a fine distribution of the
161 different classes (Fig. 2, Table 1 and Supplementary Table 1). We employed two
162 complementary strategies to perform exhaustive and expert data mining. First, NCBI, JGI and
163 1KP databases were queried using the PFAM annotation (PF00141). As this PFAM
164 encompasses the entire set of haem containing peroxidases (animal and non-animal
165 peroxidases), it allowed us to collect most of the non-animal peroxidase sequences, but did
166 not allow to discriminate between the different classes and subclasses. To complete this
167 approach, we used a homology-based strategy to determine the class/subclass memberships
168 and to detect new sequences not already annotated. All the sequences detected were available
169 from the RedoxiBase (Savelli *et al.*, 2019).

170 Thanks to the exhaustive collection and analysis of sequences, we have obtained a
171 fine distribution of non-animal peroxidases (Fig. 2 and Table 1). Our results confirm and
172 strongly support the concept that APx sequences are only detected in the eukaryotic groups
173 that having acquired and maintained chloroplasts. In the other way around, all the
174 chloroplastic containing organisms analysed (based on available genome sequences)
175 possessed at least one to ten APx sequence (Table 1). Several duplications occurred in
176 Streptophytes but without significant variations of the copy number between streptophyte
177 algae and land plants. A more divergent sequence, APx-related (APx-R) proteins have only
178 been detected in the green lineage (Chlorophyceae and Streptophyte) and not in the other

179 chloroplastic organisms. Only two diatoms (*Phaeodactylum tricornutum* and *Thalassiosira*
180 *pseudonana*) and one early divergent fungus (*Gonapodya prolifera*) possessed one APx-R
181 encoding sequences, which may have resulted from lateral gene transfers from a green
182 lineage organism (Table 1). Furthermore, the APx-R may have emerged more recently in the
183 green lineage from an existing APX (neofunctionalisation) and still presented a high identity
184 level. The sequence has been probably lost in Charales (missing sequence or short DNA
185 fragment in *Nitella hyaline* and *Chara braunii* respectively). In addition, CcPs are present in
186 all eukaryotic organisms but absent from Streptophytes and organisms that have undergone a
187 mitochondrial loss event such as Microsporida (obligate intracellular parasite fungi) or which
188 have developed parasitic style of life (*Apicomplexa* animal parasites, *Pyromyces* commensal
189 organisms of the herbivore gut) (Fig. 2 and Table 1). Interestingly, we can notice that all the
190 organisms having acquired a chloroplast, have preserved their mitochondria. On the other
191 hand, a loss of mitochondria is never associated with a gain of chloroplast. The majority of
192 the detected CP sequences belong to the prokaryotes (bacteria) while the rest was distributed
193 punctually in several eukaryotic genomes due to probable lateral gene transfer (Passardi *et*
194 *al.*, 2007a) (Table 1). We found that all CI Prxs are distributed throughout the eukaryotic
195 kingdom confirming the hypothesis of a common ancestral sequence for the different
196 members of the CI Prx subclasses. The APx-CcP hybrid sequences were detected in several
197 kingdoms without taxonomic relations (Table 1). The APx-CcP have conserved key residues
198 from the Class I but are clustered together within the kingdoms meaning that they diverged
199 independently from an ancestral APx or CcP sequence.

200 Next, with the use of a larger dataset, we confirmed and supported, the hypothesis that
201 APxs are specific to chloroplast containing organisms, and that CcPs are specific to
202 mitochondrial containing organisms with the exception of all embryophytes. In the Excavata,
203 two phyla perfectly illustrated these gain or lost events. On the one hand, *Giardia intestinalis*,
204 *Spironucleus barkhanus* and *Spironucleus salmonicida* belonging to the Diplomonadida
205 phylum have lost their mitochondria and lack the CcP sequence. On the other hand, *Euglena*
206 *mutabilis*, *E. gracilis* and *Eutreptiella gymnastica* part of Euglenida phylum, have gained
207 chloroplasts and contained both APx and CcP sequences.

208 A marginal presence of APx sequences was found in non chloroplast-containing
209 organisms, likely due to lateral gene transfer (LGT) (Table 1). This phenomenon is most often
210 observed in living water organisms (proximity with algae containing APx) and for symbiotic
211 organisms. *Hydra viridis*, belonging to animals contained a non-animal Prx with both

212 characteristic of APx and traits of CIII Prx. Based on the homology with APx sequences
213 found in green algae, a LGT from *Chlorella* can be postulated, and that some subsequent
214 evolutionary convergence or divergence occurred, leading to the homology with CIII Prx. We
215 also note the marginal presence of CI Prxs in organisms belonging to animal kingdoms, even
216 though these proteins are known to be generally absent from the animal kingdom. For
217 example, the presence of APx and CcP encoding sequences in the order of Choanoflagellates
218 demonstrates this.

219

220 In addition, we observed a lack of CcP for several parasitic organisms containing
221 mitochondria. Among them, is Apicomplexa, a unicellular animal parasite containing a
222 mitochondria and an organelle referred to as the apicoplast (a relic of chloroplasts obtained
223 by secondary endosymbiosis of red algae, (Lim and McFadden, 2010). More than 10
224 Apicomplexa genomes are available from NCBI and JGI but no Class I has been detected in
225 these organisms. A chloroplast-containing Apicomplexa, *Chromera* (small EST library) was
226 analysed and an APx like protein was found but no CcP.

227

228 Regarding the CII and CIII Prxs, the taxonomic distribution is more precise and has
229 recently been described (Mathé *et al.*, 2019). However, we found that it was necessary to
230 include CII and CIII Prxs in our analysis of non-animal peroxidase evolution, because it
231 allowed us to draw a more integrative evolutionary scenario.

232 Briefly, the set of CII Prxs sequences was only detected in Ascomycetes and
233 Basidiomycetes that interact with living or dead plants. As mitochondrial containing
234 organisms, Ascomycetes and Basidiomycetes contained CcP in two copies in most cases. This
235 is likely not due to recent gene duplication, as the two copies are phylogenetically distant. We
236 observed that most of the fungi containing CII Prxs lack the second CcP copy, which suggest
237 that CII Prx could have originated from one CcP copy. Microsporidia, obligate intracellular
238 parasite fungi, and *Pyromyces* (Neocallimastigomycota order), a commensal organism of the
239 herbivore gut, have both lost mitochondria during evolution and did not contain CcP
240 sequence (absence confirmed from more than 10 available genomes).

241 Similarly to the CII Prxs evolutive scenario, CIII Prxs sequences were only found in
242 Streptophytes and are absent from Chlorophyceae. Such as CII Prxs, CIII Prxs belong to
243 multigenic families whose size has largely increased along the green lineage (from 0 in

244 *Mesostigma viride* or 1 in *Chlorokybus atmophyticus* to 191 *E. grandis*, Fig. 2 and Table 1).
245 The CIII Prxs copy number found in Streptophyte algae is low in comparison to land plants.
246 The late diverging plants contain more CIII Prxs copies than early diverging plants due to
247 explosive and recent duplications (tandem (TDs), segmental (SDs) and whole genome
248 duplications (WGDs), Li *et al.*, 2015), which make the size of CIII family quite variable (Fig.
249 2) including a significant number of pseudogene (as also found for the CII Prxs). Since
250 Streptophytes are the only mitochondria containing organisms lacking CcP, and because CI
251 and CIII Prxs present high level of similarity and several conserved key residues, we raise a
252 hypothesis that the ancestral CIII Prx was derived from the ancestral CcP sequence.

253 Other non-animal peroxidases, not belonging to any of the well described CI, CII and
254 CIII Prxs, are non-specifically distributed. Therefore, they can be considered as the result of
255 an evolutive divergence or convergence from an ancestral CI Prxs.

256

257 **In land plants, the CIII Prx activities are greater than those of other plants**

258 APxs, CcP and CIII Prx specific activities have been tested in different species
259 including early to late divergent plants, several fungi, and chloroplast-containing protists
260 (Fig. 3).

261 Significant variations of APx activities were observed between species likely due to
262 the organ variability or different substrate affinity. However, the different activities are in the
263 same order of magnitude between the different organisms from the green lineage tested (Fig.
264 3A), which is in agreement with the conserved number of APx sequences detected with the *in*
265 *silico* analysis (Fig. 2). As expected, no APx activity was detected in non chloroplast-
266 containing organisms. CcP activity was detected in all mitochondria-containing organisms
267 excepted in the eight organisms belonging to the Streptophytes (Fig. 2B). This was in
268 agreement with the presence or absence of CcP encoding sequences (Fig. 2).

269 CIII Prxs activity was only detected in the Streptophytes (Fig. 3C) which is consistent
270 with our *in silico* analysis. The differences between early divergent plants (*K. flaccidum*, *S.*
271 *pratensis* and *C. orbicularis*) and late divergent plants were significant ($P = 0.01$), and the
272 range of values was very high (a ratio of 1000:1 between *C. orbicularis* and *E. grandis*).
273 Between the terrestrial streptophytes tested, we observed that the activities were also
274 significantly different possibly due to certain variability in plant ages or stages development.
275 This is in agreement with the finding that the activities of different-aged *A. thaliana* were

276 significantly different (data not shown (Cosio and Dunand, 2010)). Surprisingly, low activity
277 was detected from *K. flaccidum*, *S. pratensis* and *C. orbicularis* protein extracts despite the
278 presence of CIII Prx sequences in these organisms (Fig. 2, 3C and Supplementary Fig. S1).
279 This low activity could be due to the poor cell wall protein extractability and/or to the lower
280 affinity for *in vitro* substrate, to the reduced sequence similarity when compared with *A.*
281 *thaliana* sequences (36 to 42% of identity for the sequences of *K. flaccidum*, *S. pratensis* and
282 *C. orbicularis* when compared with the closest *A. thaliana* orthologous sequence) or to the
283 low expression level of the CIII Prx in Streptophyte algae. In Zygnematales, the 12 species
284 analysed contained similar number of APx sequences, but 5 did not present CIII Prx
285 suggesting a lower expression of CIII Prxs than APx.

286 In the green lineage, it appeared that the activities of CIII Prxs increased in the land
287 plant vs the green algae while no significant difference of APx between plants or CcP activity
288 has been detected (Fig. 3). The explosion of CIII Prx activities can be correlated with the
289 explosion of the gene copy number. Although no direct evidence has been demonstrated yet,
290 this explosion could be related to the variation of environmental oxygen concentration, to the
291 organs and cell wall complexification, to the climatic changes, to the colonization of new
292 biotopes, to the constant appearance of new pathogens and, most recently, to human impact
293 on cultivated plants (Passardi *et al.*, 2004).

294

295 **The features of CIII Prx in early divergent plants**

296 In order to confirm the increase in gene number from early to late divergent plants,
297 *Spirogyra* sp. genome and *S. pratensis* RNA-seq data were analysed (Delaux *et al.*, 2015; Van
298 de Poel *et al.*, 2016). *Spirogyra* belongs to the Zygnemophyceae, a close relative of land
299 plants. At least seven independent CIII Prx and five independent APx encoding sequences
300 were detected from *Spirogyra* sp and *S. pratensis*. In parallel, a search of *K. flaccidum* and
301 *Closterium peracerosum* available sequences allowed for the identification of 3 CIII Prx from
302 the two species and seven and six APx, respectively. The CIII Prxs from the two *Spirogyra* sp
303 encode for proteins with regular predicted 3D structure containing 12 alpha-helices
304 (Supplementary Fig. S2). Although the sequences were not shorter, no predicted signal
305 peptide or C-terminal extension can be detected with various prediction programs (SignalP
306 4.1 (Petersen *et al.*, 2011), PrediSi (Menne *et al.*, 2000), Signal-3L (Shen and Chou, 2007)
307 and SOSUIsignal (Hirokawa *et al.*, 1998)) for two CIII Prxs of the seven (considering that

308 two sequences are partial). The canonical CIII Prx sequences have eight conserved cysteine
309 (Cys) residues (C1-C8) that are necessary for disulfide bond formation (Fig. 1B). CIII Prx
310 genes lacking some Cys residues can be found all along the green lineage with a frequency
311 lower than 5 % (41 CIII Prx with missing cysteine for 959 CIII Prx analysed). No genes
312 lacking more than two disulfide bonds have been identified in this study. Despite the absence
313 of data concerning the influence of the lack of disulfide bonds on the protein stability or
314 activity, we could assume that the thermo stability could be affected due to missing disulfide
315 bonds. In the seven CIII Prxs from *Spirogyra sp*, the cysteines C2 and C3 surrounding the
316 distal histidine are missing while they are present in the CIII Prx from *K. flaccidum* and *C.*
317 *peracerosum*. The transient expression in tobacco leaves of the native SsPrx03 and SsPrx03
318 with the two missing cysteines have been performed in *N. benthamiana*. No fluorescence was
319 detected in any cell compartments of *N. benthamiana* leaves transformed with SsPrx03 fused
320 to the TagRFP. However, the cell wall localisation was detected when leaves were
321 transformed with sequence containing the two cysteines C2 and C3 (SsPrx03m1, Fig. 4A).
322 However, similar signal was detected in the leaves transformed with CIII Prx from *A.*
323 *thaliana* (native AtPrx36 and AtPrx36 mutated for the cysteines C2 and C3 (AtPrx36m3)).

324 In parallel, total CIII Prx activity was measured in the infected areas of the leaves.
325 First, the lack of the cysteines C2 and C3 did not significantly affect the CIII Prx activity
326 detected in the leaves of the AtPrx36m3 vs AtPrx36. The addition of the cysteines C2 and C3
327 in the *Spirogyra sp* sequence did not increase the initial activity (SsPrx03 vs SsPrx03m1) and
328 was not significantly different from the activity detected with the empty vector in both cases
329 (Fig. 4B). The low activity detected with SsPrx03, lacking the cysteines C2 and C3, could
330 support the very low activity detected from *Spirogyra sp* fresh material (Fig. 3).

331

332 **Conclusions**

333 Based on previous studies (Lazzarotto *et al.*, 2015; Passardi *et al.*, 2007a), we have
334 performed exhaustive data mining and expert sequence annotation. We have greatly increased
335 the number of sequences belonging to underrepresented kingdoms. We conclude that the
336 three CI Prx subclasses (APx, CcP and CP) have originated from the same ancestral CI Prx
337 sequence initially present in the Neomura (Fig. 5). This sequence is no longer present in all of
338 the current prokaryotes and only the bifunctional CP containing two fused peroxidase
339 domains has been marginally found. The CP are also detected in some eukaryotes with a

340 chaotic distribution due to LGT. The punctual distribution of CP in most of the kingdoms and
341 the lack of other non animal Prx in prokaryotes strongly supports the existence of a mono
342 functional ancestral CI Prx. The presence of APx in most of photosynthetic organisms (uni,
343 multi and pluricellular organisms) suggests that sequence gains followed independent
344 primary endosymbiosis. The evolutive scenario leading to the modern CII and CIII Prxs
345 appeared to be similar. They probably both emerged independently from an ancient CcP in
346 the ascomycetes and basidiomycetes for the CII Prx and in the streptophyte algae for the
347 CIII Prxs. The emergence of the CII and CIII Prxs was accompanied by the acquisition of a
348 secretion signal and several cysteine residues allowing disulfide bridge formation, which was
349 coherent with the extracellular activity of these proteins. These common innovations could
350 strongly support a similar evolutionary of CII and CIII Prxs in relation with the plant cell wall
351 loosening and/or stiffening. In the perspective of the better understanding of CIII Prxs
352 evolution and functions, the use of new steptophyte algae containing few CIII Prxs copies
353 will be of major interest.

354

355 **Materials and Methods**

356 **Cultivated and collected organisms**

357 The protist (*Euglena mutabilis* (Euglenida), *Chlorarachnion reptans* (Cercozoa),
358 *Synura petersenii* (Other Stramenopiles), *Moneuplotes crassus* (Alveolates), *Tetrahymena*
359 *thermophila* (Ciliophora, ciliates), *Emiliana huxleyi* (Isochrysidales)) and the algae
360 (*Chlamydomonas reinhardtii* (Chlorophyceae), *Klebsormidium flaccidum*, *Coleochaete*.
361 *Orbicularis*, *Spirogyra pratensis*) were provided by the Culture Collection of Algae and
362 Protozoa (CCAP, <https://www.ccap.ac.uk/>) and cultivated *in vitro* with culture media
363 compositions detailed in supplementary Table 2. Illustrations of cultivated cultures are visible
364 in supplementary Fig. S3 *Saccharomyces cerevisiae* (*Saccharomycetes*) is baker yeast and
365 protein extraction was made directly without *in vitro* culture. The two diatoms (*Moneuplotes*
366 *crassus* and *Phaeodactylum tricornutum* (Bacillariophyta) were provided by the LISBP
367 laboratory and their growth conditions are also describe in supplementary Table 2.

368 The Oomycete (*Aphanomyces euteiches*) and the Glomeromycete (*Glomus irregularis*)
369 were available from our laboratory and have been cultivated as indicated in supplementary
370 Table 2. The land plants *Marchantia polymorpha*, *Physcomitrella patens*, *Selaginella*

371 *moellendorffii*, *Gingko biloba*, *Brachypodium distachyon*, *Arabidopsis thaliana*, *Eucalyptus*
372 *grandis* were available directly from the growth room facility.

373 **Total protein extraction and enzyme assays of APx, CcP and CIII Prxs**

374 For multicellular organisms, approximately 100 mg fresh material was ground into
375 fine powder with a mortar and pestle under liquid nitrogen. For cell culture, to extract
376 proteins 200 µl extraction buffer containing potassium phosphate buffer (pH 7.0, 50 mM/L),
377 EDTA (5 mM/L), PVPP (16 g/L) was used for extraction. The homogenate was centrifuged at
378 10, 000 g for 10 min at 4 °C and the supernatant was used for enzyme assays and total protein
379 assay.

380 The APx activity was determined based on the oxidation of ascorbic acid (ASA).
381 Oxidation was determined based on measurements of the decrease in absorbance at 290 nm
382 between 30 sec to 5 min. one ml reaction mixture contained potassium phosphate buffer (pH
383 7.0, 50 mM/L), ASA (0.5 mM/L), H₂O₂ (0.1mM/L) and 20 µl of protein extract already
384 containing ASA, 2 mM/L. The reaction was initiated with the addition of the of protein
385 extract. A correction was made for the low, non-enzymatic oxidation of ascorbate by H₂O₂
386 (Nakano *et al.*, 1981). The APx unit is defined as the decrease in 1 min of 1 µg total proteins
387 under the above assay conditions.

388 The CcP activity was measured by monitoring at 550 nm the oxidation of Cytochrome
389 c (from the bovine heart, Sigma-Aldrich) previously reduced. The reduction was carried out
390 by mixing a solution of 0.22 mM Cytochrome c with a 0.5 mM final concentration DTT
391 solution for 15 min at room temperature. After this time, the color of the solution should
392 change from red to pale pink-red for a good reduction. Nevertheless, the reduction must be
393 estimated by spectrometry by measuring the ratio of OD 555 nm / OD 565 nm of an aliquot
394 of the solution diluted 20-fold in the assay buffer. For a good reduction, the ratio should be
395 between 10 and 20. The assay conditions for the CcP activity are a few microliters of protein,
396 50 mM sodium phosphate buffer pH = 7, 200 µM H₂O₂ and 20 µM Cytochrome c in a plastic
397 tub. After adding the proteins or H₂O₂, the OD is measured between 10 and 30 seconds.

398 The CIII Prx activity was assayed in a 1 ml reaction mixture consisting of potassium
399 phosphate buffer (50 mM/L, pH 6.0), 0.125% guaiacol (v/v) and 20 µL of protein extract. The
400 reaction was initiated by adding 125 µL H₂O₂ (11 mM/L), and the oxidation of guaiacol was
401 determined based on the increase in A470. One CIII Prx unit is defined as the increase

402 between 1 min and 2 min of 1 µg total proteins under the above assay conditions (Jia *et al.*,
403 2013).

404 The concentration of total proteins, used to determine the enzyme activity was
405 detected using Coomassie Protein Assay (Bradford).

406

407 **Identification and annotation of CI Prxs and CIII Prx genes in green lineage**

408 The presence or numbers of CI Prxs and CIII Prxs were analyzed from various
409 organisms belonging to the green lineage including early and late divergent plants. First an
410 exhaustive data mining was performed from available annotated repositories (Phytozome v12
411 (<http://www.phytozome.net/>), NCBI (Benson *et al.*, 1990)) using PFAM annotation
412 (PF00141) and keywords searches. All the collected peroxidase sequences were included in
413 the RedoxiBase (Fawal *et al.*, 2013) and classified based on homology, phylogeny and
414 homemade Prosite profile. In parallel, non-annotated repositories (1 kP and EST libraries
415 from NCBI) were also analyzed thank to homology based strategy.

416

417 **Subcellular localization of CIII Prx protein SsPrx03**

418 The full-length ORF frame of SsPrx01 (without stop codon) was amplified by PCR
419 from synthetic sequence using 5'-CCGGAATTCATGTCTCTTCTTCCCC-3' and 5'-
420 CCGCCCGGGAAGATACAAGCAATAC-3'. The PCR product was digested with *EcoRI* and
421 *SmaI*, and cloned into TagRFP-AS-N vector (Evrogen). The SsPrx03 and SsPrx03m1-
422 TagRFP C-terminal fusion was then subcloned (LR reaction, Gateway Technology,
423 Invitrogen) into pEAQ-HT-DEST1 vector (Sainsbury *et al.*, 2009). This construct was
424 confirmed by restriction analysis and sequencing, and transferred into *Agrobacterium*
425 *tumefaciens* (strain GV3101). The YFP fusion protein pm-yb CD3-1006 (Nelson *et al.*, 2007)
426 was used as a plasma membrane control. Both constructs were co-infiltrated into *N.*
427 *benthamiana* 30-day old leaves at an optical density of 0.5 at 600 nm. 36 hours after
428 infiltration leaves were detached and used for image acquisition on a Leica TCS SP2
429 Confocal Laser Scanning Microscope with the following excitation wavelengths: 561 nm for
430 TagRFP and 514 nm for YFP.

431 **SsPrx03-TagRFP, SsPrx03m-TagRFP activity measurement**

432 *N. benthamiana* leaves overexpressing SsPrx03-TagRFP, SsPrx03m-TagRFP, and
433 AtPrx36-TagRFP were dissected, weighed (40-60 mg), and frozen in liquid nitrogen. Leaves
434 not transformed or transformed with the pEAQ empty vector were used as negative controls.
435 Tissues were grinded twice for 3 min at 30 Hz using a Retsch MM400 ball mill, with a quick
436 spin between the two cycles. Soluble proteins were extracted in the following extraction
437 buffer (50 ml for 100 mg leaves): 20 mM HEPES, pH 7.0, 1 mM EGTA, 10 mM vitamin C,
438 and PVP PolyclarAT (100 mg·g⁻¹ fresh weight). The extract was centrifuged twice 10 min at
439 10,000 g (4°C) to remove insoluble material. The protein content was determined using the
440 Bradford reagent (Serva) with BSA (Sigma-Aldrich) as a standard (Bradford, 1976). CIII Prx
441 activity kinetic was measured at 25°C by following the oxidation of 8 mM guaiacol (Fluka) at
442 470 nm in the presence of 2 mM H₂O₂ (Carlo Erba) in a phosphate buffer (200 mM, pH 6.0).
443 Results were expressed in nanokatal / mg proteins ± SD (n ≥ 4).

444

445 **Acknowledgments**

446 The authors are thankful to the Paul Sabatier Toulouse 3 University and to the Centre
447 National de la Recherche Scientifique (CNRS) for granting their work. QL was supported by
448 PhD grants from the China Scholarship Council (CSC). The authors especially thank the
449 Delwiche Lab for the unpublished sequence data available for our analysis, Endymion D.
450 Cooper for the spirogyra DNA and Aurélie Le Ru (UMR5546 CNRS/UPS, Castanet-Tolosan,
451 France) for their valuable help in performing and imaging the transient expression
452 experiments.

453

454

455 **Figure legends**

456 **Fig. 1. Non-animal peroxidases: 3D structures and catalytic domain are conserved.** (A)
457 Class composition and taxonomic distribution of the 3 non-animal peroxidase classes. Among
458 the CI Prx, CP are mainly found in bacteria but also marginally in other organisms due to
459 LGT, APX are only detected in chloroplastic organisms, CcP in all organism with
460 mitochondria with the exception of plants and finally the hybrid APx-CcP found in most of
461 the eukaryotes (plants excepted). The CII Prxs and the CIII Prxs are only detected in fungi
462 and viridiplantae respectively (B) The available tertiary structure of 3 CI Prxs (CP: 5L05
463 from *Burkholderia pseudomallei*, APX: 1APx from *Pisum sativum* and CcP: 2CYP from
464 *Saccharomyces cerevisiae*), 1 CIII Prx (1HCH from *Armoracia rusticana*) and 1 CII Prx
465 (3Q3U from *Trametes cervina*) were visualized and compared using Swiss-PdbViewer. The
466 3D models show strong conservation in the number of α -helixes (blue) and their overall
467 spatial organisation. (C) Moreover, the key residues necessary for haem binding and electron
468 transfer are also highly conserved amongst the 3 main Prx classes.

469

470 **Fig. 2. Evolution of CI and CIII Prxs isoform numbers in green lineage.** The number of
471 genes encoding CI Prxs (CP, APx and CcP) and CIII Prx was analysed for 21 species from
472 viridiplantae and Chlorophyceae. All the sequences detected and included in the table are
473 available from the RedoxiBase (www.peroxibase.toulouse.inra.fr). Other: other non-animal
474 Prx no considered as belonging to CI or CIII Prxs but containing the key residues
475 characteristic of the PF00141, ns: no sequence found. ^a only very short sequence encoding for
476 an APx-R has been found. Marginal unexpected detection of CcP and CP sequences due to
477 putative sample contamination (^b, hits only with chlorophyceae and fungi), and Lateral Gene
478 Transfer (^c, LGT)

479

480 **Fig. 3. APx, CcP and CIII Prx activities were correlated with the presence of the**
481 **corresponding encoding proteins.** Total protein was extracted from 19 species belonging to
482 green plants, fungi and protists. (A) APx activity was determined by the oxidation of ascorbic
483 acid based on the decrease of the absorbance at 290 nm. (B) CcP activity was measured by
484 monitoring the oxidation of Cytochrome c previously reduced at 550 nm. (C) CIII Prx
485 activity was assayed at 470 nm to follow the oxidation of guaiacol. (D) Table
486 summarizing, the presence of mitochondria (Mito), CcP encoding sequences (CcP seq),
487 chloroplast (Chloro), APx encoding sequences (APx seq) and CIII Prx encoding sequences
488 (CIII Prx seq)) for the 19 studied species. \surd : organite or sequences have been detected, no:
489 organite or sequence has not been detected, ^a : Presence of CI, CII and CIII Prxs in *C.reptans*
490 is based on the analysis of related species *Bigelowiella natans*, ^b : presence of CI, CII and
491 CIII Prxs in *Moneuplotes (Euplotes) crassus* is based on the analysis of two closely related
492 species *E.octocarinatus* (<http://ciliates.ihb.ac.cn/database/home/#eo>) and
493 *E.vannus* (<http://evan.ciliate.org>). The enzymatic activities are expressed in nanokatal per mg of total
494 proteins. Activities are the mean of 3 independent assays \pm SD.

495 **Fig. 4. The lack of cysteine C2 and C3 could affect the cell wall localisation of Spirogyra**
496 **CIII Prx.** (A) Alignment of SsPrx03, SsPrx03m1 and AtPrx36 partial region surrounding the
497 conserved distal histidine (Hd) and containing the cysteine C2 and C3 (missing in Spirogyra
498 sequence). (B) Pieces of *Nicotiana benthamiana* leaves transformed with constructs

499 expressing SsPrx03-TagRFP, SsPrx03m1-TagRFP and AtPrx36-TagRFP were cut 48 h post-
500 agroinfiltration. The fluorescence signal was observed with a Leica TCS SP2 Confocal Laser
501 Scanning Microscope (B) Proteins were extracted from *N. benthamiana* leaves transformed
502 with constructs expressing SsPrx03-TagRFP, SsPrx03m1-TagRFP and AtPrx36-TagRFP and
503 assayed for peroxidase activity using guaiacol/hydrogen peroxide as substrate. Enzyme
504 activity values (expressed as nkatal/mg protein) are the mean of at least three replicates \pm SD.
505 P-value of one-way anova followed by Tukey's significance test, (***) $P < 0.0001$, (**) $P <$
506 0.001 , (ns) non significantly different to leaves transformed with empty vector.

507 **Fig. 5. An evolutionary model to explain the current distribution of the different non-**
508 **animal peroxidase classes.** *: Ancestral CI Prx sequence probably existing in Neomura and
509 lost in the Archebacteria, Proteobacteria and Cyanobacteria; **: chaotic and marginal
510 distribution of CP resulting from LGT; ***: several independent events of CcP lost.

511

512

513

514 References

515 **Adak S, Datta A.** 2005. Leishmania major encodes an unusual peroxidase that is a close homologue
516 of plant ascorbate peroxidase: a novel role of the transmembrane domain. *Biochemical Journal* **390**,
517 465-474.

518 **Becker K, Tilley L, Vennerstrom JL, Roberts D, Rogerson S, Ginsburg H.** 2004. Oxidative stress in
519 malaria parasite-infected erythrocytes: host-parasite interactions. *Int J Parasitol* **34**, 163-189.

520 **Benson D, Boguski M, Lipman D, Ostell J.** 1990. The National Center for Biotechnology Information.
521 *Genomics* **6**, 389-391.

522 **Cosio C, Dunand C.** 2010. Transcriptome analysis of various flower and silique development stages
523 indicates a set of class III peroxidase genes potentially involved in pod shattering in *Arabidopsis*
524 *thaliana*. *BMC Genomics* **11**, 528.

525 **Delaux PM, Radhakrishnan GV, Jayaraman D, Cheema J, Malbreil M, Volkening JD, Sekimoto H,**
526 **Nishiyama T, Melkonian M, Pokorny L, Rothfels CJ, Sederoff HW, Stevenson DW, Surek B, Zhang Y,**
527 **Sussman MR, Dunand C, Morris RJ, Roux C, Wong GK, Oldroyd GE, Ané JM.** 2015. Algal ancestor of
528 land plants was preadapted for symbiosis. *Proc Natl Acad Sci U S A* **112**, 13390-13395.

529 **Fawal N, Li Q, Savelli B, Brette M, Passaia G, Fabre M, Mathé C, Dunand C.** 2013. PeroxiBase: a
530 database for large-scale evolutionary analysis of peroxidases. *Nucleic Acids Res* **41**, D441-444.

531 **Fawcett JA, Rouzé P, Van de Peer Y.** 2012. Higher intron loss rate in *Arabidopsis thaliana* than *A.*
532 *lyrata* is consistent with stronger selection for a smaller genome. *Mol Biol Evol* **29**, 849-859.

533 **Finzel BC, Poulos TL, Kraut J.** 1984. Crystal structure of yeast cytochrome c peroxidase refined at 1.7-
534 Å resolution. *J Biol Chem* **259**, 13027-13036.

535 **Hiraga S, Sasaki K, Ito H, Ohashi Y, Matsui H.** 2001. A large family of class III plant peroxidases. *Plant*
536 *Cell Physiol* **42**, 462-468.

537 **Hirokawa T, Boon-Chieng S, Mitaku S.** 1998. SOSUI: classification and secondary structure prediction
538 system for membrane proteins. *Bioinformatics* **14**, 378-379.

539 **Jia L, Xu W, Li W, Ye N, Liu R, Shi L, Bin Rahman AN, Fan M, Zhang J.** 2013. Class III peroxidases are
540 activated in proanthocyanidin-deficient *Arabidopsis thaliana* seeds. *Ann Bot* **111**, 839-847.

541 **Kwon M, Chong S, Han S, Kim K.** 2003. Oxidative stresses elevate the expression of cytochrome c
542 peroxidase in *Saccharomyces cerevisiae*. *Biochim Biophys Acta* **1623**, 1-5.

543 **Lazzarotto F, Teixeira FK, Rosa SB, Dunand C, Fernandes CL, Fontenele AeV, Silveira JA, Verli H,**
544 **Margis R, Margis-Pinheiro M.** 2011. Ascorbate peroxidase-related (APx-R) is a new heme-containing

545 protein functionally associated with ascorbate peroxidase but evolutionarily divergent. *New Phytol*
546 **191**, 234-250.

547 **Lazzarotto F, Turchetto-Zolet A, Margis-Pinheiro M.** 2015. Revisiting the Non-Animal Peroxidase
548 Superfamily. *Trends in Plant Science* **20**, 807-813.

549 **Li Q, Yu H, Cao PB, Fawal N, Mathé C, Azar S, Wang H, Myburg AA, Grima-Pettenati J, Marque C,**
550 **Teulière C, Dunand C.** 2015. Explosive tandem and segmental duplications of multigenic families in
551 *Eucalyptus grandis*. *Genome Biol Evol* **7**, 1068-1081.

552 **Lim L, McFadden G.** 2010. The evolution, metabolism and functions of the apicoplast. *Philosophical*
553 *Transactions of the Royal Society B-Biological Sciences* **365**, 749-763.

554 **Martinez AT.** 2002. Molecular biology and structure function of lignin-degrading heme peroxidase.
555 *Enz Microb Technol* **30**, 425-444.

556 **Maruta T, Tanouchi A, Tamoi M, Yabuta Y, Yoshimura K, Ishikawa T, Shigeoka S.** 2010. Arabidopsis
557 Chloroplastic Ascorbate Peroxidase Isoenzymes Play a Dual Role in Photoprotection and Gene
558 Regulation under Photooxidative Stress. *Plant and Cell Physiology* **51**, 190-200.

559 **Mathé C, Barre A, Jourda C, Dunand C.** 2010. Evolution and expression of class III peroxidases. *Arch*
560 *Biochem Biophys* **500**, 58-65.

561 **Mathé C, Fawal N, Roux C, Dunand C.** 2019. *In silico* definition of new ligninolytic peroxidase sub-
562 classes in fungi and putative relation to fungal life style. *Scientific Reports* **in press**.

563 **Menne K, Hermjakob H, Apweiler R.** 2000. A comparison of signal sequence prediction methods
564 using a test set of signal peptides. *Bioinformatics* **16**, 741-742.

565 **Mittler R, Vanderauwera S, Gollery M, Van Breusegem F.** 2004. Reactive oxygen gene network of
566 plants. *Trends Plant Sci* **9**, 490-498.

567 **Miyake C, Michihata F, Asada K.** 1991. Scavenging of hydrogen peroxide in prokaryotic and
568 eukaryotic algae: acquisition of ascorbate peroxidase during the evolution of cyanobacteria. *Plant*
569 *Cell Physiol* **32**, 33-43.

570 **Mukherjee S, Seshadri R, Varghese N, Eloie-Fadrosch E, Meier-Kolthoff J, Goker M, Coates R,**
571 **Hadjithomas M, Pavlopoulos G, Paez-Espino D, Yoshikuni Y, Visel A, Whitman W, Garrity G, Eisen J,**
572 **Hugenholtz P, Pati A, Ivanova N, Woyke T, Klenk H, Kyrpides N.** 2018. 1,003 reference genomes of
573 bacterial and archaeal isolates expand coverage of the tree of life (vol 35, pg 676, 2017). *Nature*
574 *Biotechnology* **36**, 368-368.

575 **Nakano Y, Nakano K, Hara I.** 1981. Induction of leucocyte migration inhibitory factor (LIF) by
576 stimulation with free hapten and water-insoluble epoxy resin. *Clin Exp Immunol* **45**, 419-426.

577 **Nelson BK, Cai X, Nebenführ A.** 2007. A multicolored set of in vivo organelle markers for co-
578 localization studies in Arabidopsis and other plants. *Plant J* **51**, 1126-1136.

579 **Passardi F, Bakalovic N, Teixeira FK, Margis-Pinheiro M, Penel C, Dunand C.** 2007a. Prokaryotic
580 origins of the non-animal peroxidase superfamily and organelle-mediated transmission to
581 eukaryotes. *Genomics* **89**, 567-579.

582 **Passardi F, Favet J, Zamocky M, Jakopitsch C, Penel C, Obinger C, Dunand C.** 2007b. Phylogenetic
583 distribution of catalase-peroxidase: are there patches of order in chaos? *Gene* **397(1-2):101-13**.

584 **Passardi F, Longet D, Penel C, Dunand C.** 2004. The class III peroxidase multigenic family in rice and
585 its evolution in land plants. *Phytochemistry* **65**, 1879-1893.

586 **Petersen TN, Brunak S, von Heijne G, Nielsen H.** 2011. SignalP 4.0: discriminating signal peptides
587 from transmembrane regions. *Nat Methods* **8**, 785-786.

588 **Sainsbury F, Thuenemann EC, Lomonossoff GP.** 2009. pEAQ: versatile expression vectors for easy
589 and quick transient expression of heterologous proteins in plants. *Plant Biotechnol J* **7**, 682-693.

590 **Savelli B, Li Q, Webber M, Jemmat AM, Robitaille A, Zamocky M, Mathé C, Dunand C.** 2019.
591 RedoxiBase: A database for ROS homeostasis regulated proteins. *Redox Biol* **26**, 101247.

592 **Shen HB, Chou KC.** 2007. Signal-3L: A 3-layer approach for predicting signal peptides. *Biochem*
593 *Biophys Res Commun* **363**, 297-303.

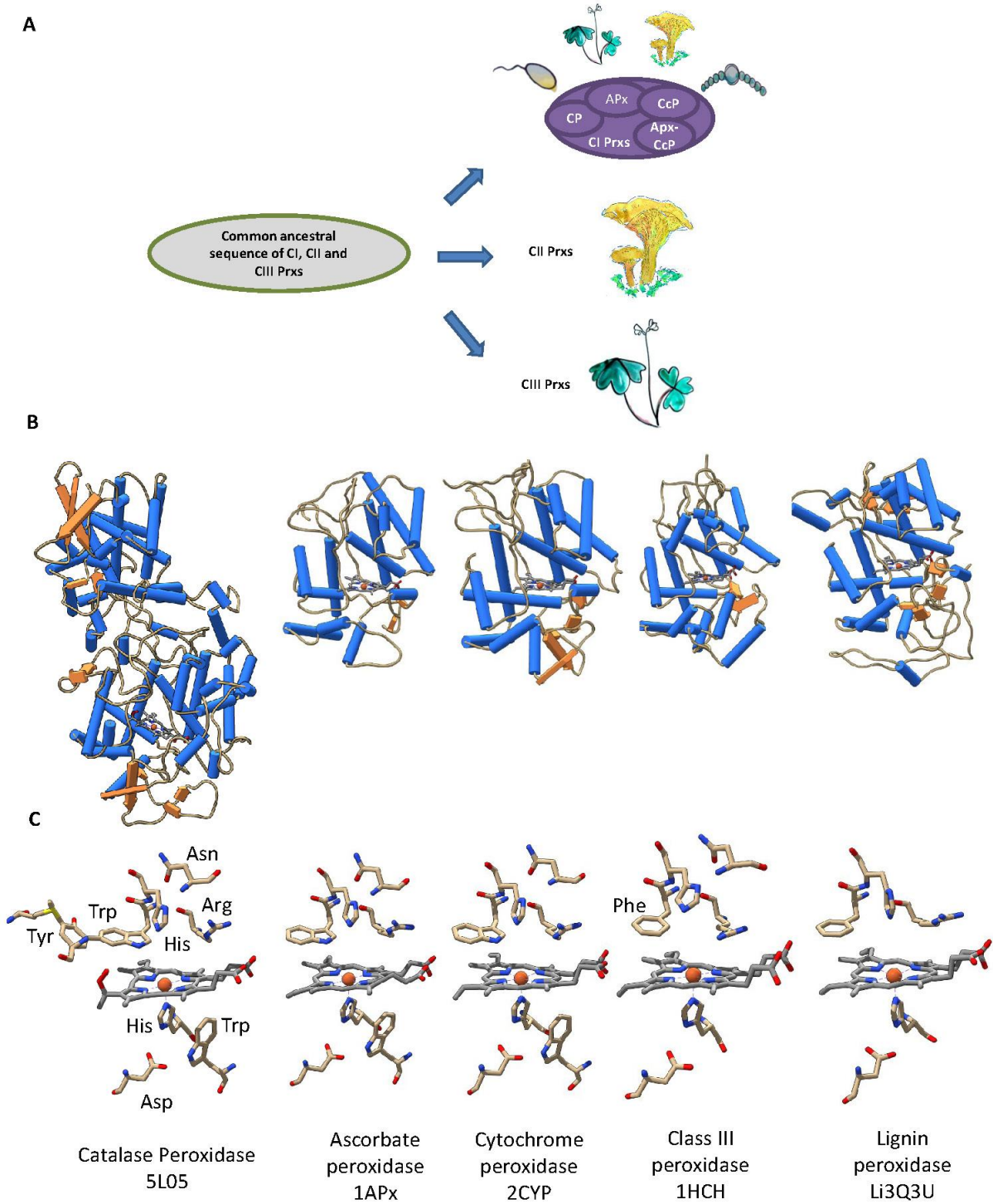
594 **Tognolli M, Penel C, Greppin H, Simon P.** 2002. Analysis and expression of the class III peroxidase
595 large gene family in Arabidopsis thaliana. *Gene* **288**, 129-138.

596 **Van de Poel B, Cooper E, Van Der Straeten D, Chang C, Delwiche C.** 2016. Transcriptome Profiling of
597 the Green Alga *Spirogyra pratensis* (Charophyta) Suggests an Ancestral Role for Ethylene in Cell Wall
598 Metabolism, Photosynthesis, and Abiotic Stress Responses(1[OPEN]). *Plant Physiology* **172**, 533-545.
599 **Welinder KG.** 1992. Plant peroxidases: structure-function relationships. In: Penel C, Gaspar T, Greppin
600 H, eds. *Plant Peroxidases*: University of Geneva, Switzerland, 1-24.
601 **Xu G, Guo C, Shan H, Kong H.** 2012. Divergence of duplicate genes in exon-intron structure. *Proc Natl*
602 *Acad Sci U S A* **109**, 1187-1192.

603

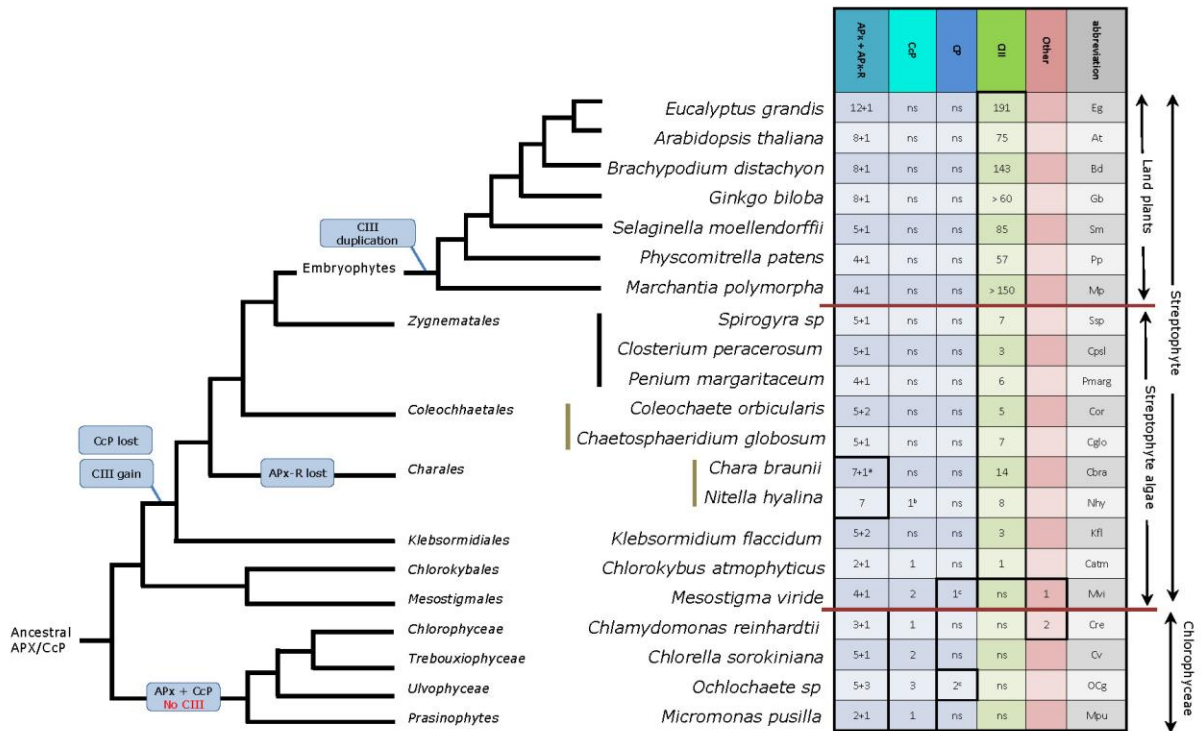
604

Figure 1 3D structure and catalytic domain of non-animal peroxidase classes are highly conserved.

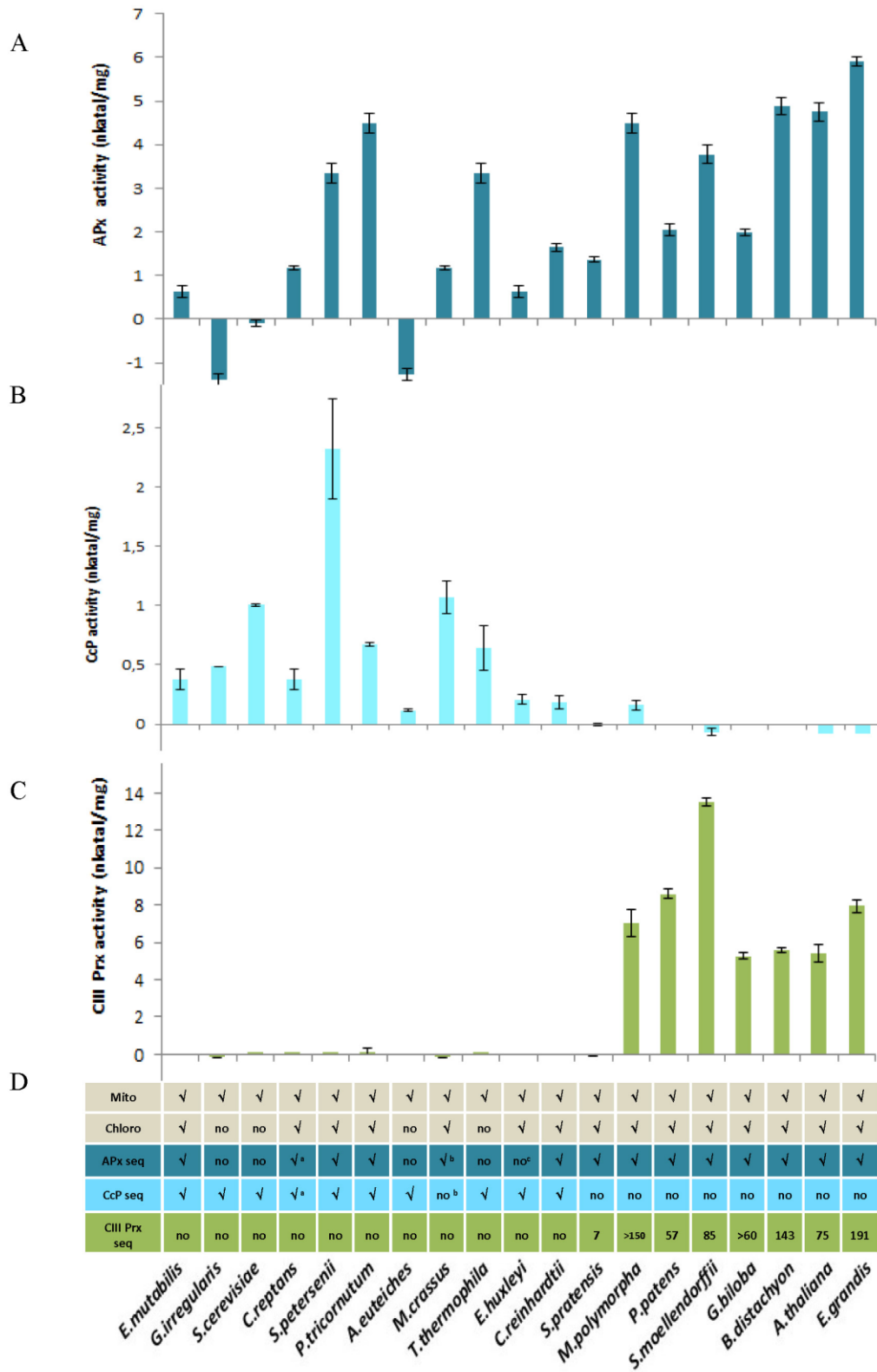


605
606

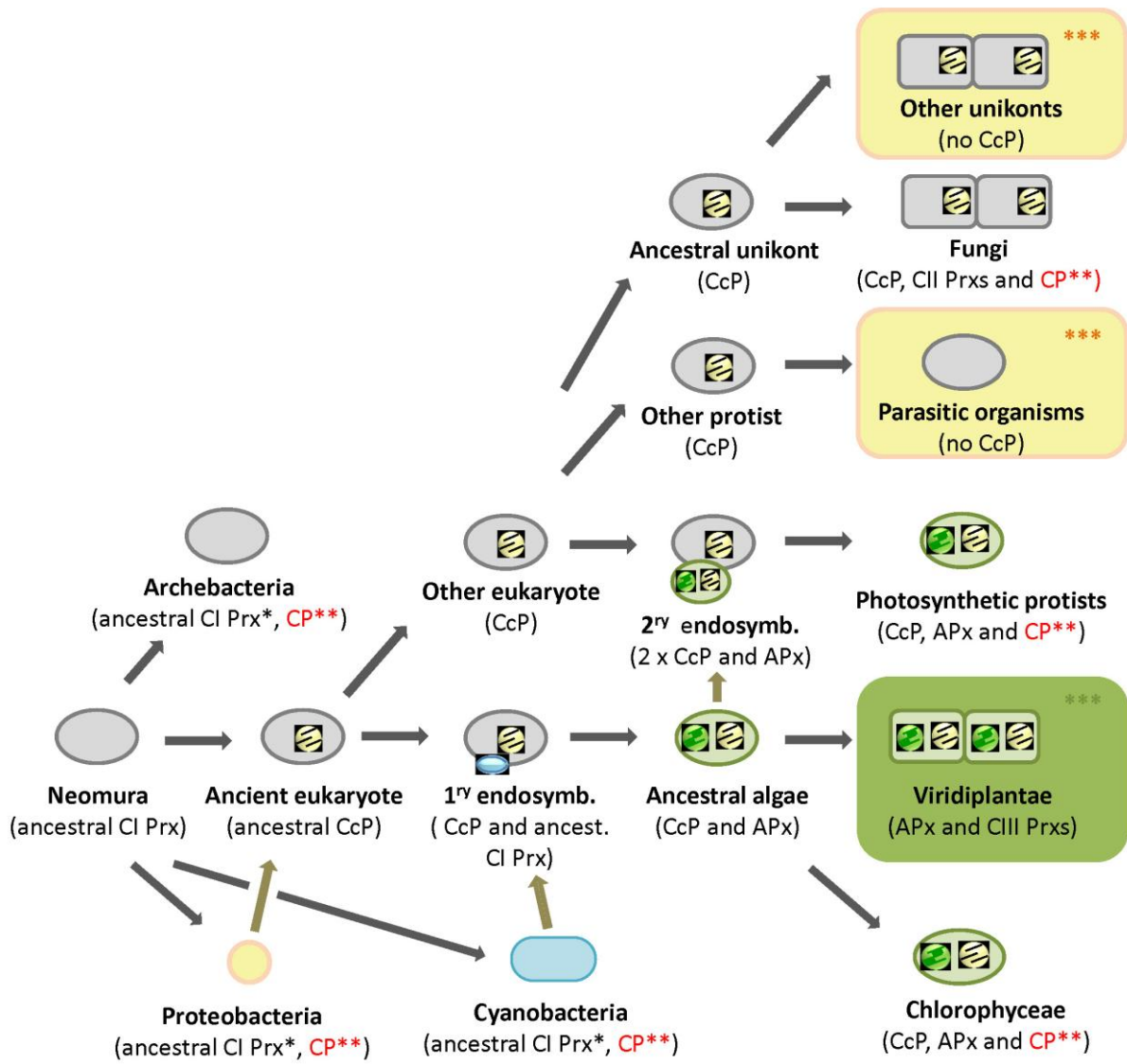
Fig. 2. CI and CIII Prxs isoform numbers evolution in green lineages.



607
608



609
610



613