The FEBS Journal

FEBS PRESS science publishing by scientists

Phylogenetic reconciliation provides new insights into the evolutionary diversification of the glutamine synthetase gene family in seed plants

Elena Aledo, Rafael Antonio Cañas, Francisco R. Cantón and Juan Carlos Aledo

Department of Molecular Biology and Biochemistry, Integrative Molecular Biology Lab, Universidad de Málaga, Spain

Keywords
dynamics of gene family; glutamine synthetase; nitrogen metabolism; ORTHGS; plant evolution; R package

Correspondence
J. C. Aledo, Department of Molecular Biology and Biochemistry, Integrative Molecular Biology Lab, Universidad de Málaga, Campus Universitario de Teatinos, Málaga 29071, Spain
Tel: +34952137129
E-mail: caledo@uma.es

(Received 28 February 2025, revised 19 July 2025, accepted 19 September 2025)

doi:10.1111/febs.70276

Glutamine synthetase (GS) catalyzes the incorporation of ammonium into glutamate, a crucial reaction in nitrogen metabolism. Despite advances in plant genomics, the evolutionary relationships among GS isoforms in seed plants remain incompletely understood. In this study, we selected 155 GS genes from 45 phylogenetically well-characterized seed plant species. Our analyses consistently support the existence of three distinct evolutionary lineages of GS genes in seed plants: *GS2* (chloroplastic), and *GS1a* and *GS1b* (cytosolic). Using a Bayesian molecular clock dating approach, we estimate that *GS2* diverged approximately 560 million years ago, whereas *GS1a* and *GS1b* began to diverge around 70 million years later. Additionally, we developed ORTHGS, a software tool that implements tree reconciliation and enables analyses of orthology and paralogy among GS enzymes. Phylogenetic reconciliation offers new insights into the evolutionary dynamics of the GS gene family in plants, revealing the existence of new paralogs of *GS1a* and *GS2*. Most of the genomes analyzed contain multiple paralogs of *GS1b*, whereas *GS1a* and *GS2* appear as singleton genes. Although these single-copy genes have traditionally been considered orthologs, we present evidence challenging this assumption. Thus, our findings suggest that both *GS1a* and *GS2* have undergone multiple duplication events throughout evolutionary history, similar to *GS1b*. However, unlike *GS1b*, only a single paralog of *GS1a* (or *GS2*) is retained per genome. Overall, our study reshapes the understanding of GS gene family evolution in seed plants by uncovering hidden duplication events in GS1a and GS2, highlighting dynamic evolutionary patterns previously overlooked in this gene family.

Introduction

Glutamine synthetase (GS; EC6.3.1.2) catalyzes the incorporation of ammonium to glutamate to form glutamine in an ATP-dependent fashion. This amide ligase, present in virtually all living organisms, seems to be vital for life. Consequently, there is a broad consensus that the gene encoding GS is extremely old [1,2], most likely already present in the Last Universal Common Ancestor [3]. Consequently, GS-encoding

Abbreviations
Abi, *Abies firma*; Ang, angiosperm; Ap, *Abies pinsapo*; Ara, *Araucaria cunninghamii*; Ath, *Arabidopsis thaliana*; Atr, *Amborella trichopoda*; Bdi, *Brachypodium distachyon*; C, Carboniferous; Cb, Cambrian; Cr, Cretaceous; Cyc, *Cycas revoluta*; D, Devonian; GS, glutamine synthetase; Gym, gymnosperm; J, Jurassic; MAD, minimal ancestor deviation; MCMC, Markov Chain Monte Carlo; Mg, *Magnolia grandiflora*; ML, *Maximum likelihood*; MP, midpoint; MV, minimum variance; Nag, *Nageia nagi*; Ne, Neoproterozoic; O, Ordovician; OG, outgroup; P, Permian; Pa, *Picea abies*; pI, isoelectric point; Pin, *Pinus armandii*; Pl, Paleogene; Pod, *Podocarpus macrophyllus*; Pp, *Pinus pinaster*; Pr, Precrambian; Psm, *Picea smithiana*; Pta, *Pinus taeda*; Sci, *Sciadopitys verticilloata*; T, Triassic; Tax, *Taxus chinensis*; Tba, *Taxus baccata*.

genes form a large and intricate multigenic family comprising three distinct groups, characterized by their primary and quaternary structures. Type I glutamine synthetases (GSI) are dodecameric enzymes containing protomers of around 55 kDa, present in the three domains of life. Type II (GSII) was first described as a group characteristic of Eukarya, forming octamers with a subunit of around 36 kDa. Later on, GSII has been described as decamers composed of two pentameric rings [4], found in Eukarya, as well as in some Bacteria and Archaea [3]. Finally, type III glutamine synthetases (GSIII), although initially described as hexamers, are currently recognized as dodecamers formed by subunits of around 75 kDa [5]. This type of GS has been found in species representing all three domains of life (Archaea, Bacteria, and Eukarya). Within Eukarya, GSIII has been described in Chlorophyta, Fungi, and Heterokonta, but not in Embryophyta nor Metazoa. Embryophyta, also known as land plants, exhibits GS forms grouped into both *GSI* and *GSII* families. With a few notable exceptions, most plant genomes examined contain only one *GSI* gene. In contrast, *GSII* genes have diversified significantly within this group of organisms, and most plant species express numerous isoenzymes classified as GSII. Additionally, since GSI homologs in plants appear to lack glutamine synthetase activity [6,7], the enzymes that have traditionally been the primary focus of research belong to the GSII type.

In plants, GS plays a crucial role in nitrogen nutrition, being a key component of nitrogen use efficiency and plant yield [8–10]. Moreover, GS has been implicated in all aspects of plant development, from germination [11,12] to senescence [13–15]. Thus, understanding the mechanisms regulating GS activity is of utmost importance. Not surprisingly, this enzyme has been the focus of interest for decades. Despite all this interest and the enormous effort devoted to the study of this enzyme, the progress in our understanding has been rather limited, mainly because the wide diversity of functions and complex regulation of this enzyme rests on a gene family with a long and intricate evolutionary history behind it [16]. In accordance with this diversity of roles, different paralogs of *GS* have arisen through gene duplication, followed by processes of neofunctionalization and subfunctionalization. Although some works have been made to dissect the subfunctionalization experienced by *GS* paralogs in some plant species [17–20], the results obtained in one species are hardly transferable to other species, as it has been repeatedly stated [9,21]. The observation that the results obtained in different plants are not directly comparable has led to the importance of studying GS in both model and crop species [9]. Regardless of the advantage of broadening the spectrum of species being studied, the diversity of *GS* gene endowments makes it difficult to transfer conclusions obtained from studies in model species to other species, and this situation could be improved if data obtained in experimental work on a limited set of species could be interpreted and discussed within the framework of a well-established network of orthology/paralogy relationships.

Indeed, the evolution of GS has been the subject of numerous studies dating back to the 1990s [1,2,22–25]. This foundational research was developed before the accumulation of genomic data. Nevertheless, the molecular evolution of this enzyme has been revisited multiple times [26–28], extending up to the present day [3,16]. Altogether, these studies represent a valuable contribution. However, none of them has systematically addressed the establishment of orthology and paralogy relationships among members of the *GS* family. Within spermatophytes, the *GS* family has been traditionally considered to be formed by the cytosol-localized GS1 and the plastid-localized GS2 [29–31]. Recently, it has been proposed to distinguish two groups within GS1, designated as GS1a and GS1b based on their sequence and functional similarity to the gymnosperm GS1a and GS1b [16]. In the current study, we provide further evidence supporting that these three *GS* groups (*GS1a*, *GS1b*, and *GS2*) represent *bona fide* evolutionary lineages. More interestingly, we present findings that corroborate our hypothesis that the three GS-encoding genes are ancient paralogs already present in the last common ancestor of seed plants. Furthermore, in the current study, we addressed the origins of the three *GS* groups throughout evolutionary time, concluding that *GS2* was the first lineage to emerge and has persisted to the present day. The *GS2* branch may have preceded the divergence between *GS1a* and *GS1b* by approximately 70 million years. Our results also suggest that while *GS1a* and *GS2* have been subjected to a tight selection, *GS1b* has been heavily diversified. Finally, we have also analyzed the orthology/paralogy relationship of 155 GS sequences from 45 seed plant species. In the case of *GS2* and *GS1a*, although in general they are single-copy genes (singletons), they have undergone different processes of duplication and subsequent gene loss, leaving only one gene copy. This has led to the existence of paralogs and non-orthologous genes for these two families in nearby species. The results of such analyses can be explored using an R package accompanying this paper, written *ad hoc* and available at https://cran.r-project.org/web/packages/orthGS/index.html.

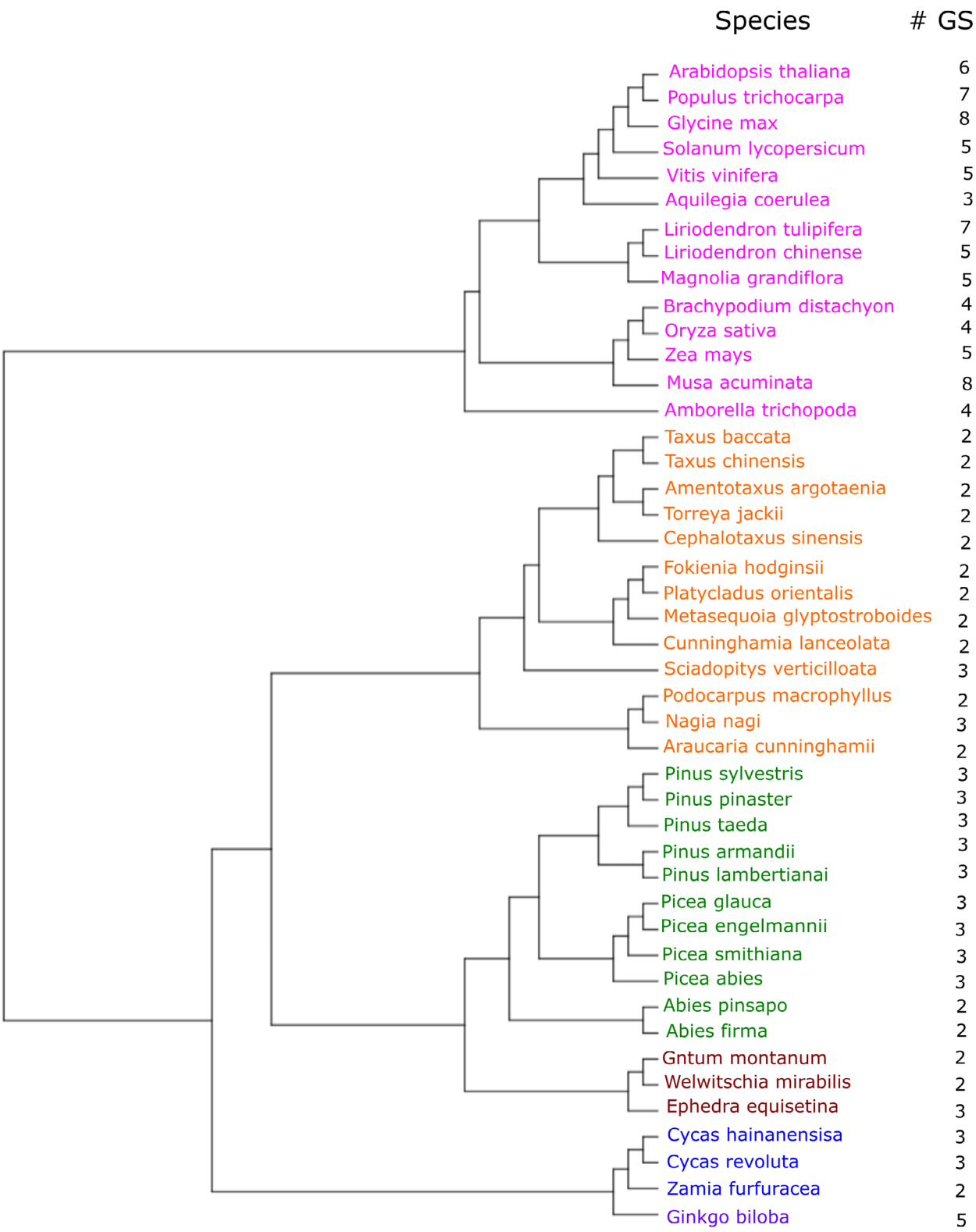

Fig. 1. Tree of selected species. 45 species of Spermatophyta were selected based on their well-established phylogenetic relationships [32]. 14 species of angiosperms (magenta) account for 76 glutamine synthetase (GS) sequences. On the other hand, within gymnosperms (79 GS sequences), we distinguish Conifers II (orange), Pinacea (green), Gnetales (brown), Cycadales (blue), and Ginkgo (purple).

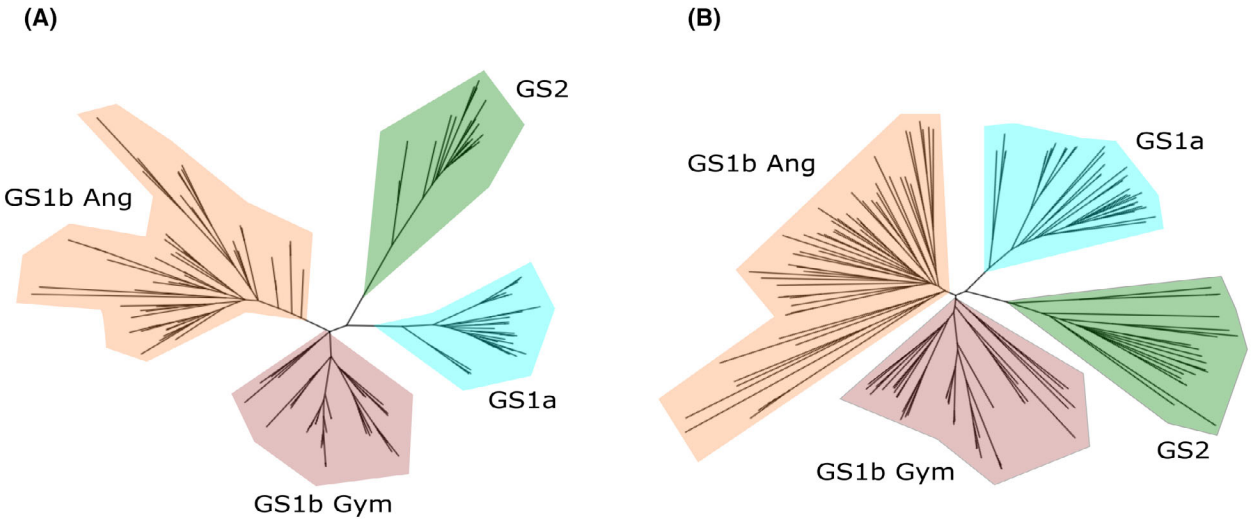

Fig. 2. Glutamine synthetases from seed plants cluster into four apparent groups. Unrooted phylogenetic trees were inferred using 155 glutamine synthetase protein sequences from 45 Spermatophyta species (see Table S1 for further details). (A) Tree obtained using a maximum likelihood analysis. (B) Phylogeny was reconstructed using an alignment-free approach. The tetrapeptide frequencies in each sequence (4-gram) were used to compute a distance matrix and its corresponding tree. Ang, angiosperms; Gym, gymnosperms. An R script that reproduces this figure can be accessed at https://github.com/jcaledo/Phylogeny-of-plant-GS/blob/main/Scripts/Figure2_Seed_Plants_4_Clusters.R.

Results

Seed plants GS proteins cluster into three main phylogenetic lineages

A total of 155 GSII protein sequences from 45 different species of Spermatophyta were collected as described in the Materials and methods section. These sequences were selected because of the well-established phylogeny of the hosting species (Fig. 1). The taxonomic groups represented included 24 Pinopsida (13 Conifer II and 11 Pinaceae), 3 Gnetales, 3 Cycadales, 1 Ginkgoaceae, and 14 Angiospermae (4 primitive and 10 modern angiosperms). The number of GS isoforms found in each species is indicated in Table S1 [17,21,33–38]. Overall, 76 GS sequences from angiosperms and another 79 from gymnosperms were subjected to phylogenetic reconstruction. Figure 2 shows the unrooted tree obtained using either a maximum likelihood approach (Fig. 2A) or an alignment-free method (Fig. 2B).

Although the detailed tree topology varies depending on the approach used to construct the tree, it is notable that both the alignment-based and alignment-free approaches consistently cluster all 155 GS proteins into four distinct groups. Following previous work, these well-defined groups were designated as GS2, GS1a, GS1b-Ang (angiosperms), and GS1b-Gym (gymnosperms). The so-called GS2 group contained all the proteins clustering together with the chloroplast glutamine synthetase from the angiosperm *Arabidopsis thaliana*. This group included enzymes from three species belonging to gymnosperm lineages, such as Ginkgosipda and Cycadopsida (see Table S1 for details). On the other hand, the labeled GS1a group comprised all the proteins that clustered together with the well-characterized cytosolic enzyme GS1a from *Pinus sylvestris* [39,40]. This last group consisted mainly of GS proteins from gymnosperm plants, but primitive angiosperms also contributed to the GS1a group (Table S1). Analogously, the GS1b-Gym group consisted of all the sequences from gymnosperms that clustered with the *P. sylvestris* GS1b enzyme [40,41], while the GS1b-Ang cluster was constituted by the GS enzymes from angiosperms that grouped with the five cytosolic GS1 isoforms from *A. thaliana* [21]. Although Fig. 2 seems to suggest four different groups, as we shall elaborate next, rooting these protein GS trees underlines the existence of three evolutionary *GS* lineages because GS1b from angiosperms and GS1b from gymnosperms most properly conform to a single evolutionary lineage, sister to GS1a (Fig. 3).

GS2 was the first modern lineage to emerge in Spermatophyta

Early phylogenetic studies conducted with a limited set of plant GS sequences indicated that *GS2* likely emerged before *GS1a* and *GS1b* began to diverge from

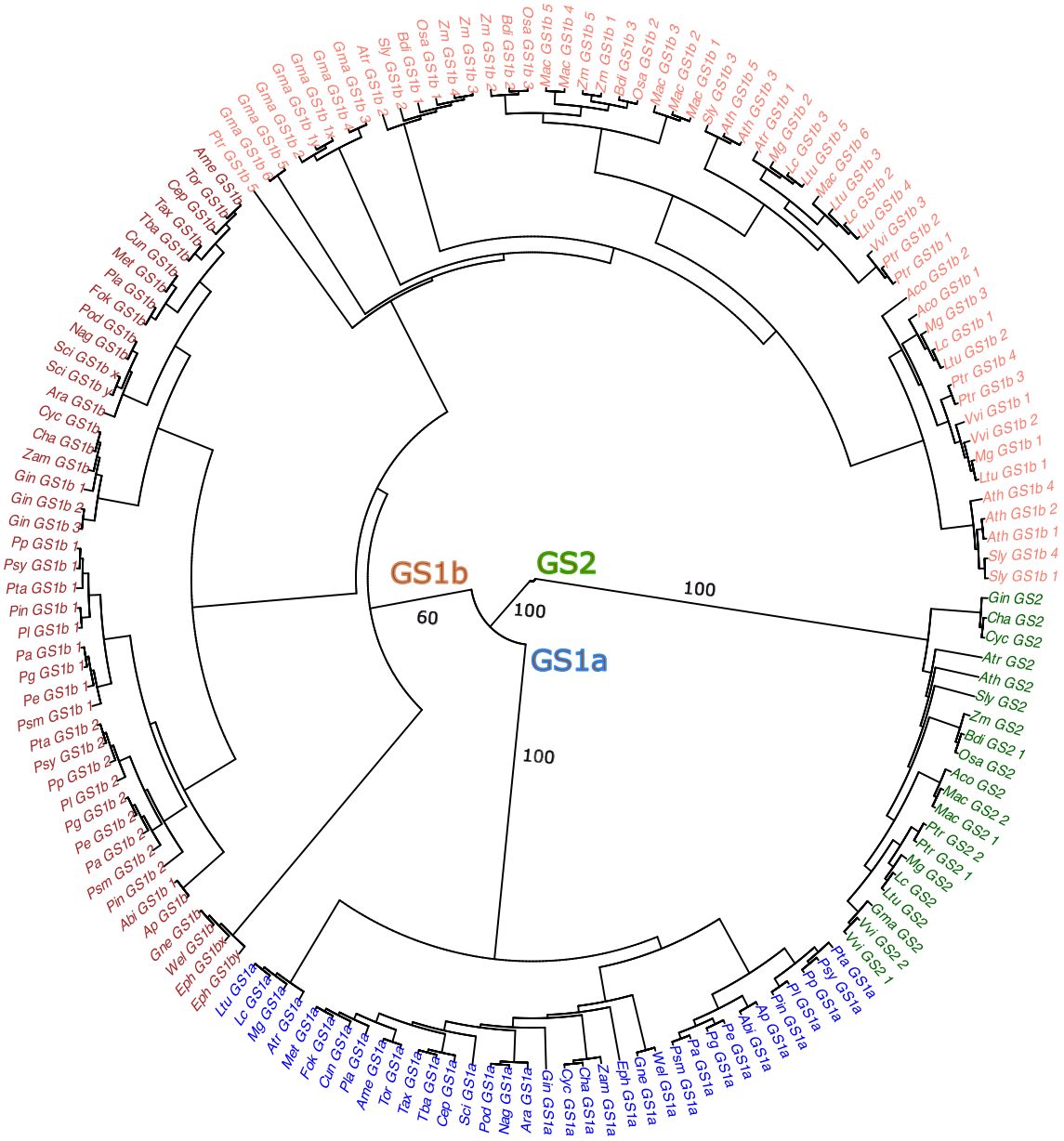

Fig. 3. Rooted tree of the glutamine synthetase protein family. The unrooted tree obtained by maximum likelihood was rooted using the method of minimal ancestor deviation. The numbers next to the branches corresponding to the main lineages indicate the percent bootstrap support. An R script that reproduces this figure can be accessed at https://github.com/jcaledo/Phylogeny-of-plant-GS/blob/main/Scripts/Figure3_Rooted_GS_Protein_Tree.R.

each other [23,41]. However, gene expression patterns, conservation of specific Cys residues, and nucleotide sequence-based phylogeny of *GS* may suggest a closer phylogenetic relationship between *GS1a* and *GS2* genes with respect to the *GS1b* group [16,42]. To elucidate the branching order of these three *GS* lineages, we have employed various approaches to control factors that might affect tree topology and rooting, with the goal of achieving robust and reliable conclusions. Among these factors, we have closely examined sampling bias and the rooting procedure as two critical elements. We began by randomly sampling our set of seed plant species. From the 45 species used in this study, we extracted 100 random samples, each consisting of six species (see Table S2 for the raw data). While the sampling procedure was random, we

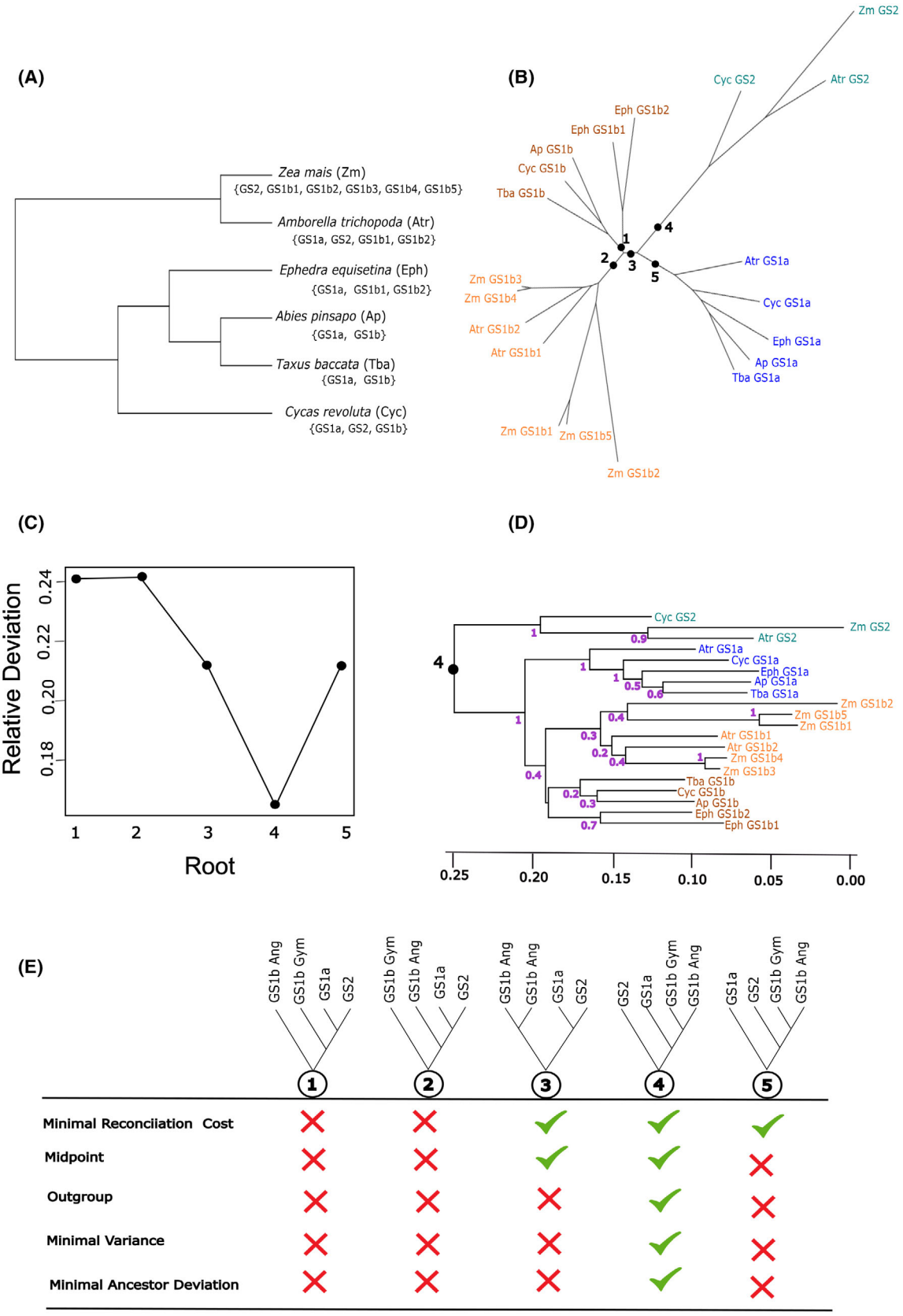

Fig. 4. Rooting the glutamine synthetase protein family using sets of representative species. (A) Organism tree for a set of species representative of the main seed plant groups (primitive and modern angiosperms, Gnetales, Pinales, Conifers II, Ginkgoales-Cycadales), whose phylogenetic relationships are well established. (B) Unrooted glutamine synthetase (GS) protein family tree of the isoform found in the representative species shown in (A). Each of the four GS groups detected in Fig. 2 has been colored differentially (GS2: green, GS1a: blue, GS1b Gym: brown, and GS1b Ang: orange). The branches in which *a priori* the root could be located have been marked with a filled circle and numbered from 1 to 5. (C) Using the minimal ancestor deviation algorithm, the relative deviations at each of the five potential places for the root were computed and plotted. (D) The rooted GS tree yielding the minimal relative deviation is shown. Bootstrap support values are shown in purple. The horizontal axis gives the estimated evolutionary distances (substitutions per site). (E) One hundred species sets of representative organisms (one species from each taxon) were phylogenetically analyzed. The unrooted trees were rooted using different approaches, including minimal cost for species tree-gene tree reconciliation, midpoint, outgroup (isoform of *Chlamydomonas reinhardtii* as outgroup), minimal variance, and minimal ancestor deviation. The results of such analyses (Table S2) are summarized in this figure. A green tick indicates that at least one of the 100 samples has been successfully rooted at the specified node using the chosen rooting method. Ang, angiosperms; Gym, gymnosperms. This figure can be reproduced using the script available at https://github.com/jcaledo/Phylogeny-of-plant-GS/blob/main/Scripts/Figure4_Rooted_Representative_Trees.R.

ensured the inclusion of one representative from each of the following taxa: (a) primitive angiosperm, (b) modern angiosperms, (c) ginkgo-cycad, (d) Gnetales, (e) Pinacea, and (f) Conifers II. In Fig. 4A–D, we have taken one of the 100 samples to illustrate the procedure followed with all the random samples. Thus, the species tree is shown in Fig. 4A. All the GS forms present in these six species (20 sequences in the current example) were retrieved and used to build an unrooted tree (Fig. 4B). The determination of ancestor–descendant relationships in an unrooted tree is accomplished by inferring a root node, which can initially be positioned on any branch of the unrooted tree. In our unrooted tree (Fig. 4B), potential locations for the root have been indicated with filled circles and numbered from 1 to 5. For instance, if we placed the root at the point 4, then it would be concluded that the *GS2* lineage was the first to diverge from the common line. Conversely, if the root were located at point 5, then *GS1a* would be considered the first to emerge as a distinct entity up to the present day, and so on. Therefore, an accurate placement of the root node is crucial for our purposes. Figure 4D shows the same tree as Fig. 4B once it has been rooted. In this case, the position of the root node was inferred using a recently introduced outgroup-free method called minimal ancestor deviation (MAD) [43]. Under a strict molecular clock assumption, that is, when the tree exhibits ultrametricity, the root can be easily found at the midpoint of the path between any two tips. However, strict ultrametricity is rarely met in practice, and the midpoint deviates from the actual position of the ancestor node. The MAD method allows us to quantify the deviations of the midpoint criterion for all potential root positions of the unrooted tree, and the tree root is chosen as the position that minimizes these deviations. Figure 4C shows the relative deviations at the five potential root positions within our unrooted tree. As can be seen, the root at position 4 exhibits the minimum relative deviation, while placing the root either at position 1 or 2 would introduce maximum deviations with respect to a molecular clock.

In addition to the MAD method, four other rooting procedures were applied to the 100 trees obtained from the random samples described above. Among these alternative strategies of rooting a tree, we considered the popular outgroup (OG) and midpoint (MP) methods, as well as two more recently introduced algorithms based on the branch length distribution, such as the MAD and minimum variance (MV) methods. The latter minimizes the variance of root to tip distances [44]. The fifth method used was based on minimizing the cost of tree reconciliation (see Materials and methods for details). As summarized in Fig. 4E, in all the 100 samples, the OG, MAD, and MV methods failed to place the root at any branch other than the labeled as number 4. On the other hand, while the MP method placed the root at branch 4 in two thirds of the rooted trees, some samples were rooted at branch number 3. Finally, the minimal reconciliation cost ruled out the root at positions 1 and 2, but it could not discriminate between trees with roots at positions 3, 4, or 5.

In extant organisms, GS1 is present in the cytosol while GS2 is mainly found in chloroplasts. Thus, the presence of a transit peptide is common in GS2 preproteins that are synthesized in the cytosol and must be usually targeted to the chloroplast. The transit peptide is a sequence at the N-terminus of the protein that generally directs it to the chloroplast, where it is cleaved off once the protein has reached its destination. Since the transit peptide often contains basic residues, its presence contributes to a significant increase in both the molecular size and the isoelectric

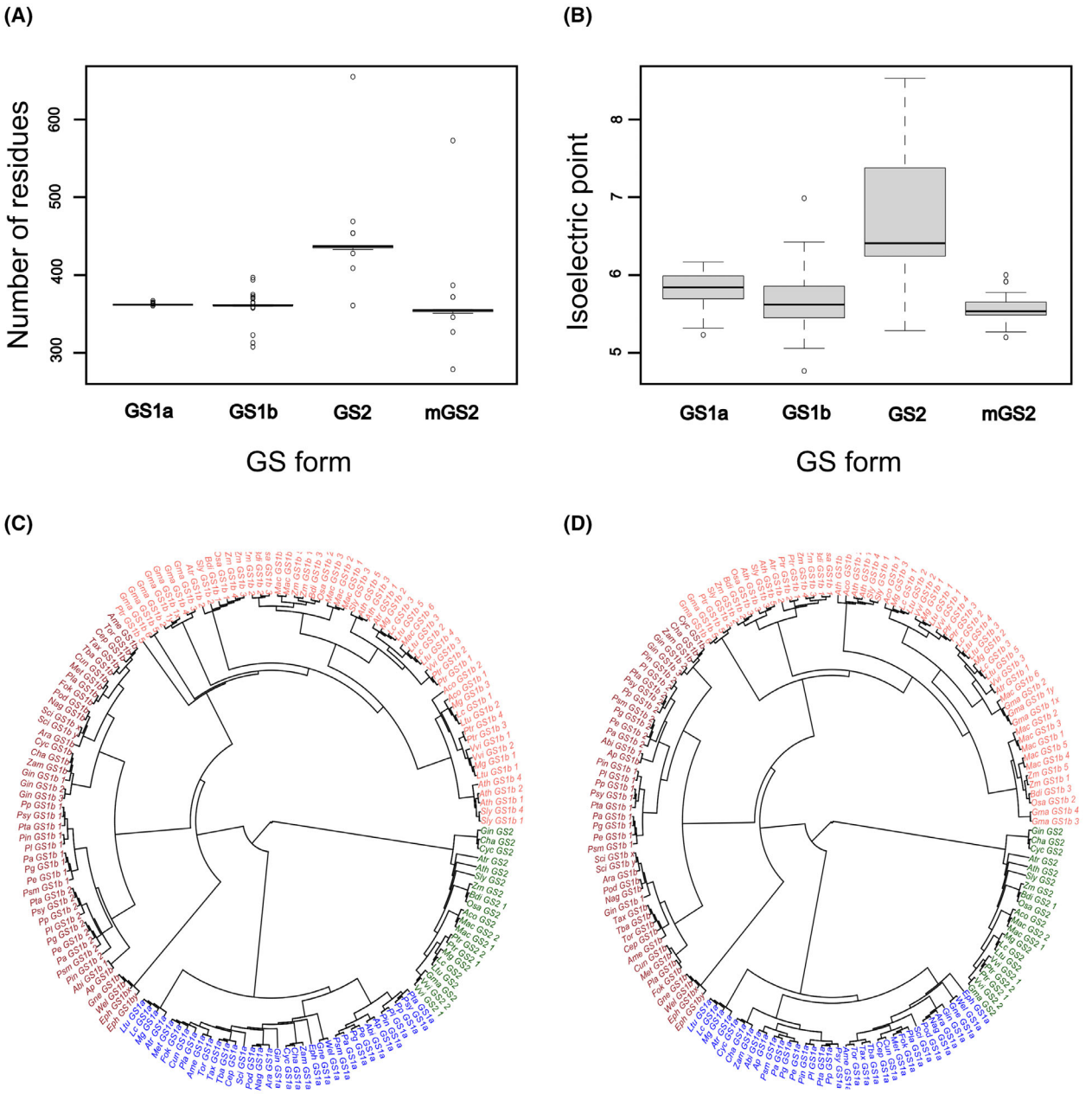

Fig. 5. Glutamine synthetase (GS) 2 is the modern lineage that appears first within the GS protein family of seed plants, regardless of the transit peptide. The size (A) and isoelectric point (pI) (B) distributions of the GS isoforms are shown as boxplots. Phylogenetic trees inferred using the preprotein GS2 (C), or the mature GS2 protein, mGS2 (D). Despite the differences between GS2 and mGS2 in size (*P*-value = 2.1×10^{-5}) and isoelectric point (*P*-value = 1.6×10^{-5}), both forms produce comparable phylogenies. Each lineage has been colored differently: GS2 (green), GS1a (blue), GS1b Gymnosperm (brown), and GS1b Angiosperm (orange). This figure can be reproduced using the script available at https://github.com/jcaledo/Phylogeny-of-plant-GS/blob/main/Scripts/Figure5_MatureGS2.R.

point of the GS2 preprotein (Fig. 5A,B, respectively). At this point, we wondered whether this transit peptide might contribute to confounding the distances between GS proteins from different lineages. In other words, we sought to rule out the possibility that *GS2* incorrectly appeared in the first place as a lineage branch. For this purpose, the phylogenetic relationships between the 155 GS proteins were inferred before and after removing the transit peptide from all the GS2 forms included in our analyses (Fig. 5C,D,

respectively). As expected, we observed moderate changes in topology, reflected by a normalized Robinson–Foulds distance of 0.32. Nevertheless, the conclusion of the existence of three major *GS* lineages, with *GS2* emerging as the first to establish itself as distinct, remained well supported.

Dating GS evolution in seed plants

Hitherto, we have shown evidence that supports the following sequence of evolutionary events: first, an ancestral form of *GS* split into the *GS2* lineage and an ancestral *GS1* form, which subsequently gave rise to the *GS1a* and *GS1b* lineages. After establishing the temporal sequence of gene duplications and diversification that led to the three main *GS* lineages in spermatophytes, we aimed to date these two splitting events. For this purpose, we implemented a Bayesian molecular clock dating methodology using 25 GS proteins from six representative species. As can be observed in Fig. 6, the emergence of *GS2* around 560 million years ago marks a very early event in the history of life. Our results also indicate that it took an additional 70 Myr for *GS1a* to split from *GS1b*.

Unraveling the hidden GS1a paralogs from conifers

Genes or proteins that diverge through speciation are known as orthologs, while those that diverge through gene duplication events are referred to as paralogs [45]. Conifers possess two distinct families of cytosolic GS isoforms: GS1a, primarily expressed in photosynthetic tissues and encoded by a single locus, and GS1b, predominantly found in non-photosynthetic tissues and encoded by multiple loci. Given that *GS1a* is a single-copy gene, one might assume that the sequences of this isoform found in different species are orthologs. However, using GS data from six species within the order Pinales and applying tree reconciliation techniques to align species and protein trees (see Materials and methods for details), we made an unexpected discovery. The *GS1a* isoforms in the genus *Pinus* are paralogs of the *GS1a* isoforms in the genera *Picea* and *Abies*. In contrast, all *GS1a* isoforms from *Picea* and *Abies* are orthologs to each other. The inferred sequence of evolutionary events leading to such reconciliation is shown in Fig. 7. As observed, ancestral duplicated forms of both GS1a and GS1b were present in the last common ancestor of all the studied Pinaceae species. GS1a subsequently underwent three independent loss events, one in each of the lineages leading to pines, spruces, and firs. This resulted in the GS1a isoforms in *Pinus* and *Picea* being paralogs, while those in *Picea* and *Abies* remained orthologs. In contrast, GS1b underwent only a single loss event, which occurred in the lineage leading to fir trees. The species-protein tree reconciliation, which extends to other conifers such as Cupressales and Araucariales, is shown in Fig. 8. Surprisingly, the data indicate that GS1a of Cupressales and Araucariales are orthologs to each other but analogous to GS1a of Pinales (Figs 8 and 9). Additionally, we have provided a resource (https://r-pkg.org/pkg/orthGS) that allows interested readers to infer tree reconciliations and establish orthology/paralogy relationships for any subset of the 155 GS isoforms included in this study. Although ORTHGS was originally developed to distinguish between orthologs and paralogs within the GS family, the current version extends this functionality to other gene families. To illustrate this, we have provided a vignette titled 'Orthology Beyond GS', which demonstrates how to use this new functionality.

Discussion

The *GS* gene superfamily is one of the oldest among living organisms, and the proteins they encode have an essential function in incorporating inorganic nitrogen into organic molecules [3]. Until recently, it was considered that in plants, *GS* could be subdivided into two gene subfamilies simply based on their subcellular localization (cytosolic GS1 and plastidial GS2) [46]. This was due to the lack of gene sequence information in large groups of plants, as most of the information gathered was from agronomically important plants, almost all of which are monocots and eudicot angiosperms. However, new genomic data have revealed that the evolutionary history of *GS* genes in plants is much more complex [16]. Focusing on seed plants, at least three major lineages of *GS* genes have been identified: *GS1a*, *GS1b*, and *GS2*. Both *GS1a* and *GS1b* genes encode proteins with cytosolic localization; however, *GS1b* is present in all seed plants, whereas *GS1a* is found only in gymnosperms, the ANA grade, and magnoliids, having been lost in more recently evolved angiosperms. On the other hand, the *GS2* product is a plastidial protein, with this gene being present in all seed plants except for the Gnetophyta/Coniferophyta group.

In this work, we have attempted to delve deeper into the phylogeny of *GS* in seed plants using methodological approaches distinct from those used in previous studies. This includes reconciliation techniques to align species with phylogenetic trees, the use of different tree

rooting methods, and molecular clock techniques. Our data robustly confirm our previous proposal of the existence of these three lineages of *GS* in seed plants, originating from a common ancestral *GS* gene (Figs 2–6). In addition, they shed light on the question previously raised about two possible scenarios for the emergence of *GS2*. These results suggest that the three gene lineages must have been present in the common

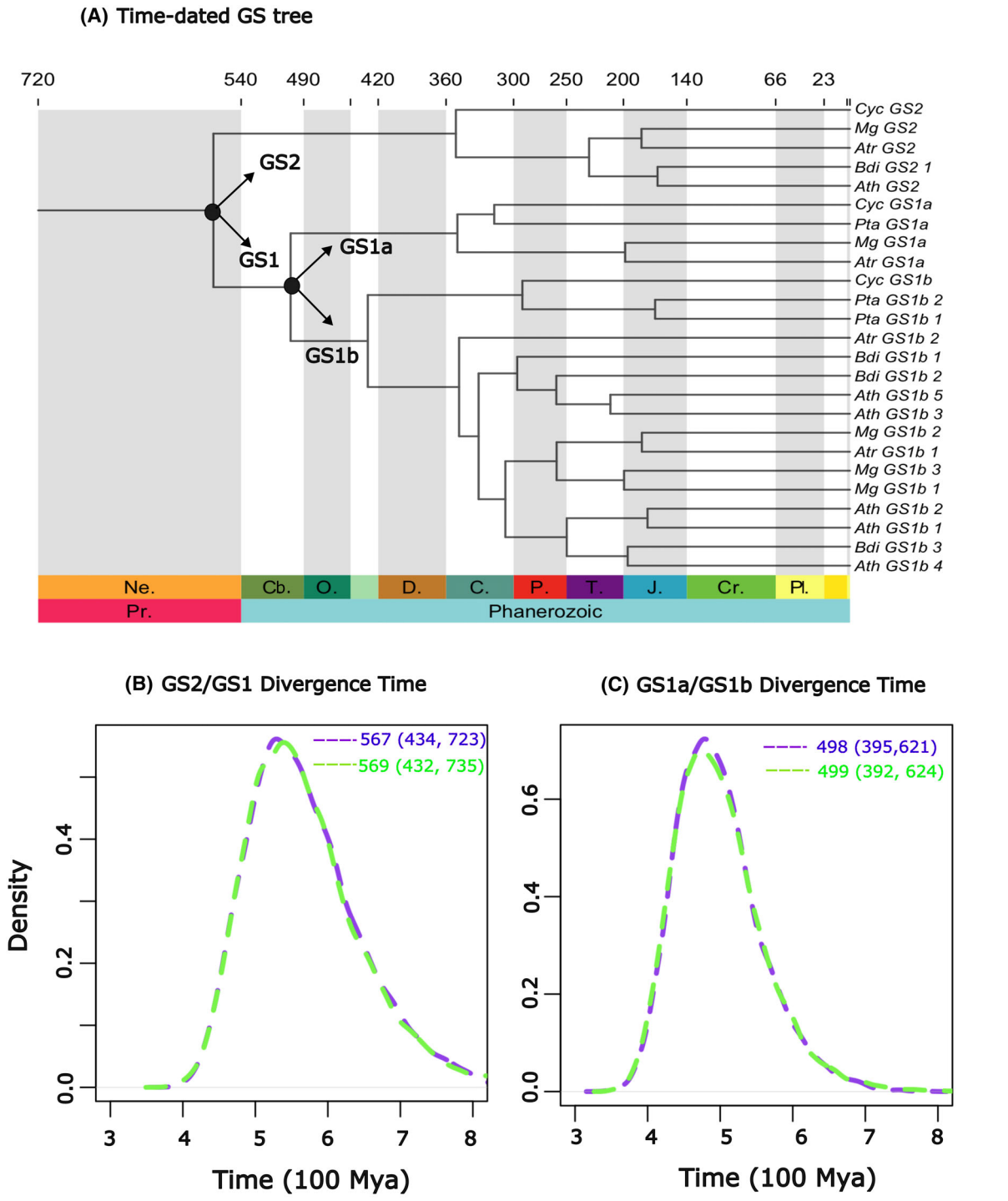

Fig. 6. Molecular clock dating of glutamine synthetase (GS) splits under relaxed models. (A) Data-time tree of seed plant GS proteins (Cyc, *Cycas revoluta*; Mg, *Magnolia grandiflora*; Atr, *Amborella trichopoda*; Bdi, *Brachypodium distachyon*; Ath, *Arabidopsis thaliana*; Pta, *Pinus taeda*). The horizontal axis at the top of the figure shows the time in million years. The colored rectangles at the bottom of the image represent the geological periods within the eras of the Phanerozoic Eon and part of the Precambrian. Each color corresponds to a specific period: Pr., Precambrian (dark pink); Ne., Neoproterozoic (orange); Cb., Cambrian (olive green); O., Ordovician (green); D., Devonian (brown); C., Carboniferous (teal); P., Permian (red); T., Triassic (purple); J., Jurassic (blue); Cr., Cretaceous (lime green); Pl., Paleogene (yellow). The light blue horizontal bar labeled 'Phanerozoic' indicates the Phanerozoic Eon, which includes all periods from the Cambrian to the present. (B) Posterior distribution of divergence times for GS2-GS1. (C) Posterior distribution of divergence times for GS1a-GS1b. The purple and green dashed lines represent the distributions of divergence times sampled after analyses where the root age was bounded at 800 and 1000 million years, respectively. The mean and the 95% CI (Highest Posterior Density Credible Interval) are displayed in the top-right corner of the box enclosing each panel. This figure can be reproduced using the script available at https://github.com/jcaledo/Phylogeny-of-plant-GS/blob/main/Scripts/Figure6_Dating_GS_Evolution.R.

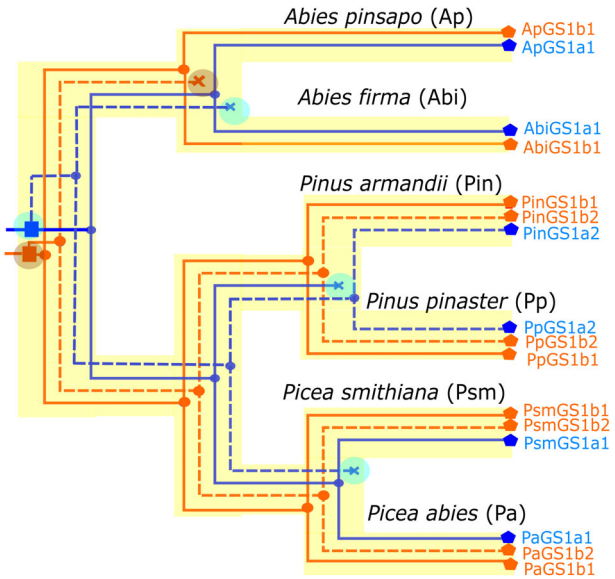

Fig. 7. Species and protein trees reconciliation focused on the order Pinales. Duplications (squares) and losses (crosses) have been emphasized with colored circles: turquoise for GS1a and brown for GS1b. Solid and dashed lines differentiate paralogs originated from a duplication event. Small blue (GS1a) and orange (GS1b) circles indicate speciation events. This figure can be reproduced using the script available at https://github.com/jcaledo/Phylogeny-of-plant-GS/blob/main/Scripts/Figure7_Tree_Reconciliation.R.

ancestor of seed plants, with *GS2* being the first to diverge. This suggests that the *GS2* gene was lost in the common ancestor of gnetales and conifers, as previously suggested [41]. Following this idea, it is necessary to consider which gene family presents characteristics most similar to the ancestral *GS* gene. Despite being the only group present in all seed plants, *GS1b* can be ruled out since in older plant groups such as ferns there are no sequences related to this lineage, although there are sequences related to GS1a and GS2 [16]. Additionally, GS1b usually forms small gene subfamilies where members have diversified their functions, unlike GS1a and GS2 which maintain similar functions and are usually single-copy genes. On the other hand, it is difficult to decide between *GS2* and *GS1a*. Although *GS2* was the first lineage to diverge, both *GS1a* and *GS2* are expressed in photosynthetic tissues and perform similar metabolic functions. Additionally, GS2 protein usually has a plastidial localization, which is not found in the proteins encoded by the phylogenetically related *GS* genes of ferns, the closest relatives of seed plants [16]. This implies that it is necessary to delve deeper into the phylogeny of ferns and more primitive plant groups to elucidate this issue. However, the availability of genome sequences for these plant groups remains limited, with a focus on only a few taxa. Additionally, many intermediate groups that preceded seed plants are found only in the fossil record, which hampers the resolution of this question.

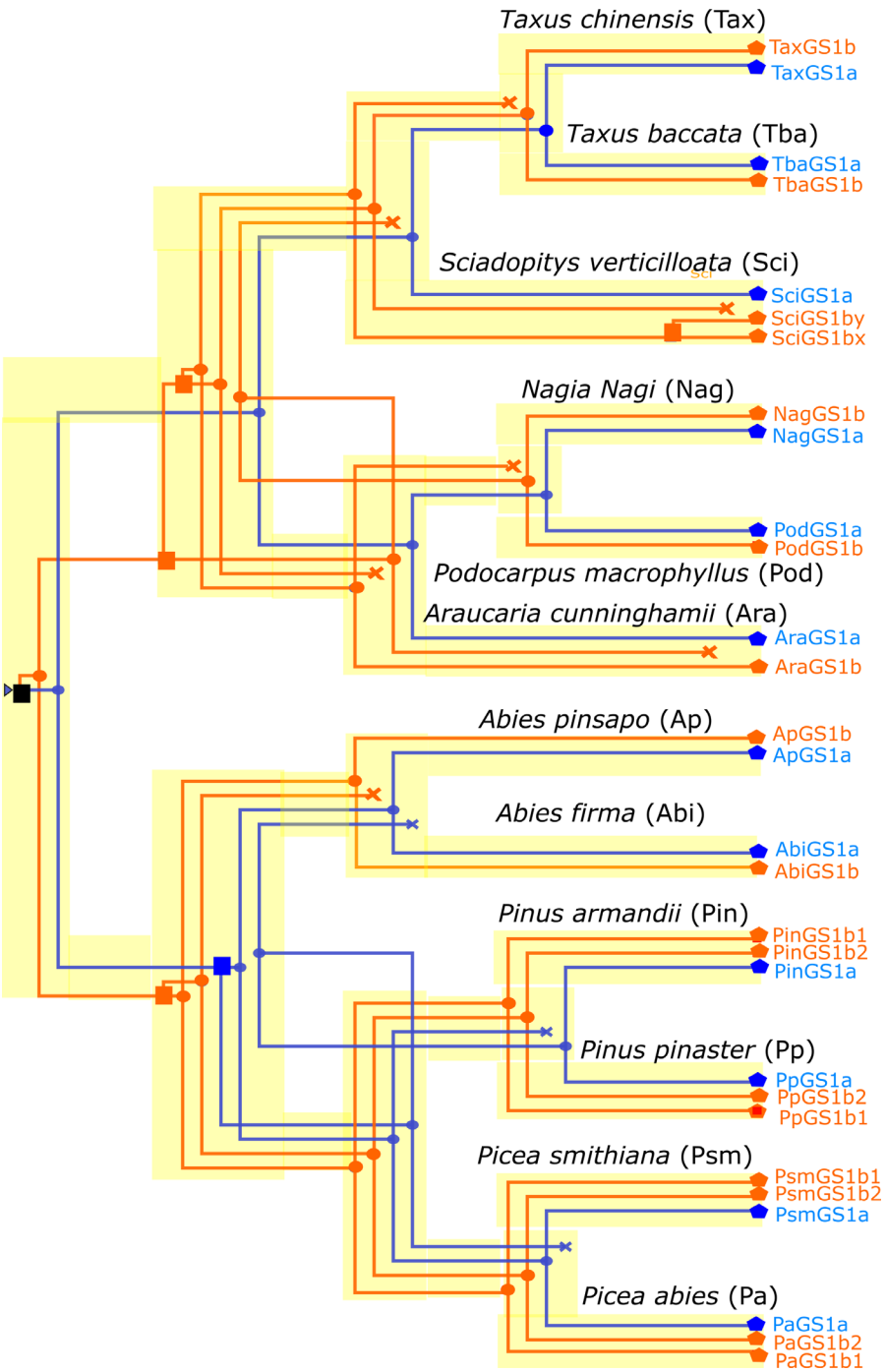

Fig. 8. Reconciliation of species and protein trees in conifer species. The analyzed conifers consisted of a set of 12 species, including representatives of Conifer II (Tax, *Taxus chinensis*; Tba, *Taxus baccata*; Sci, *Sciadopitys verticilloata*; Nag, *Nageia nagi*; Pod, *Podocarpus macrophyllus*; Ara, *Araucaria cunninghamii*) as well as representatives of Pinaceae (Ap, *Abies pinsapo*; Abi, *Abies firma*; Pin, *Pinus armandii*; Pp, *Pinus pinaster*; Psm, *Picea smithiana*; Pa, *Picea abies*). Circles, squares, and crosses indicate speciation, duplication, and losses, respectively. The glutamine synthetase (GS) gene phylogeny was built using the function *mltree* from the R package ORTHGS. The tree reconciliation was performed with RANGER-DTL and plotted with THIRDKIND.

Our results also raise additional considerations regarding the relationship between the functions of the *GS1a* and *GS2* and their evolutionary events. The proteins produced by both types of genes are expressed in photosynthetic tissues and seem to be linked to photorespiration and nitrate photoassimilation [16]. This

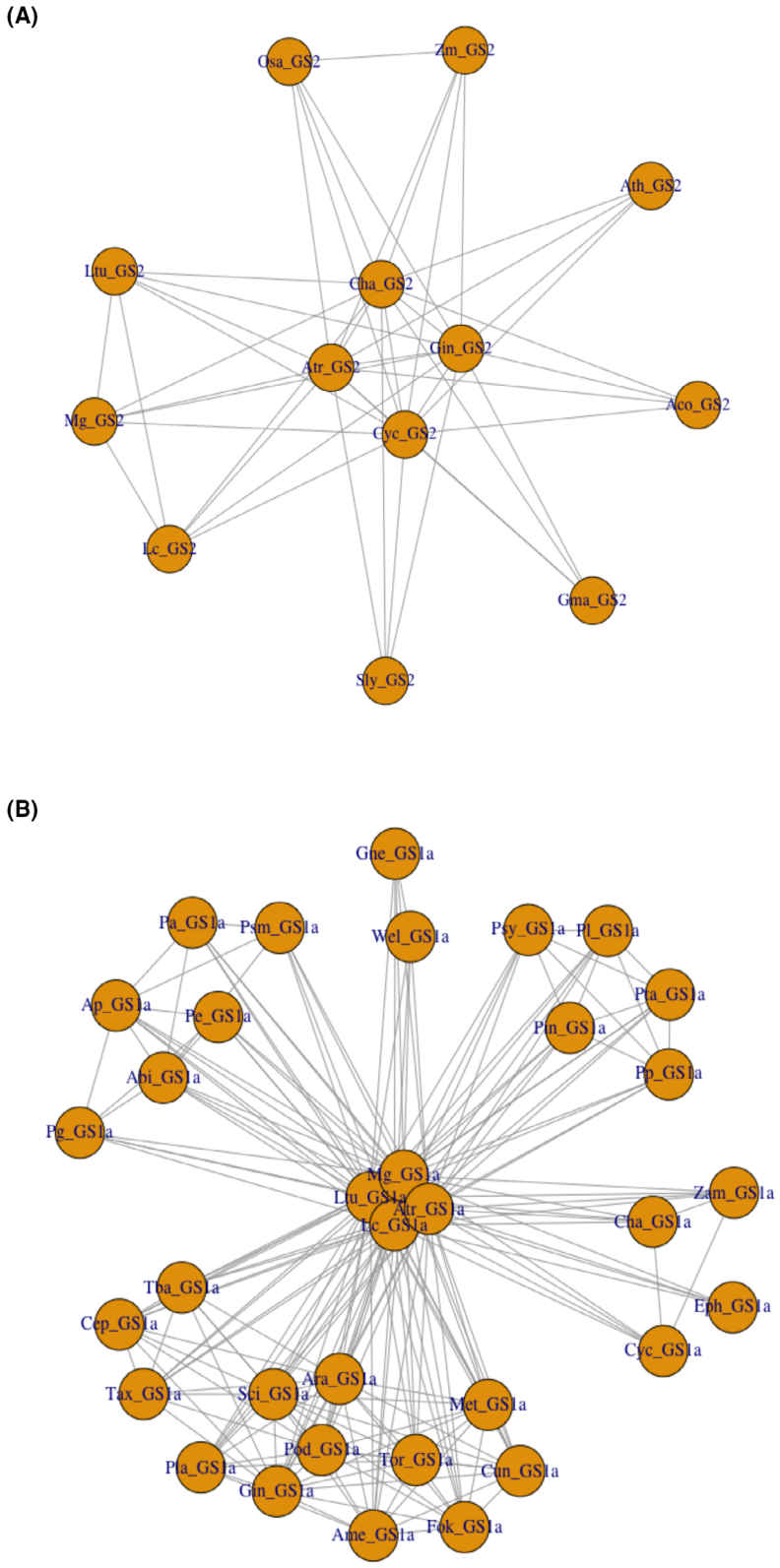

Fig. 9. Orthology networks for glutamine synthetase. Proteins from different species are connected by lines when they are orthologs. Unconnected nodes represent paralogs. (A) Lineage GS2. (B) Lineage GS1a.

implies a certain degree of equivalence in their functions in plant metabolism, so there are groups of plants that can live without *GS1a* (modern angiosperms) and others without *GS2* (gnetales/conifers). However, this fact, along with the observation that a single copy of these gene families is generally found in plant genomes, except in cases of recent genetic duplications, suggests a purifying selective pressure. It is possible that an excess of GS activity (higher gene dosage) could be detrimental in the tissues and conditions where *GS1a* and *GS2* genes are expressed. In this regard, it has been suggested that high levels of GS2 enzyme activity in *Arabidopsis thaliana* decrease ammonium tolerance [47].

To gather evidence supporting the presence of evolutionary forces limiting the dosage of *GS1a* and *GS2*, we resorted to analyses of orthology/paralogy. To this end, we have developed an R package, ORTHGS, which implements tree reconciliation. This software enabled us to identify ortholog and paralog genes within selected clades. In this way, we revealed that both *GS1a* and *GS2* genes have undergone duplications throughout evolution. However, unlike what occurs with *GS1b*, in all species, one of the two paralogs is eventually lost, meaning that not all *GS1a* genes are orthologs to each other, and the same can be sustained for *GS2* (Fig. 9). Together, these results support our hypothesis of selection against an increase in GS1a and GS2 dosage, which may be related to their roles in photosynthesis and photorespiration.

Since ortholog genes are typically the ones that mirror the evolutionary relationships between species, accurately predicting orthology is essential in evolutionary biology. Furthermore, as argued by Koonin, making a clear distinction between orthologs and paralogs is critical for the construction of a robust evolutionary classification of genes. In addition, this distinction between orthologs and paralogs also plays a key role in ensuring the reliable functional annotation of newly sequenced genomes [48]. However, single-copy genes are frequently employed as molecular markers for phylogenetic analysis in seed plants [49]. This approach rests on the critical assumption that these genes are orthologs across different species. However, that may not always be always the case. Much of the confusion arises from the misconception that two paralog genes must be present within the same genome when nothing in the definition imposes such a restriction [45]. Of course, at the time when paralogs originated by duplication, both copies were present in the same genome, but it does not imply that these paralogs should be restricted to the same genome as evolution progresses. Over time, gene families expand (gene duplication) and contract (gene loss) in such a way that we can end up finding paralogs in different species. Sometimes these paralogs are easy to recognize when multiple homologs are in the same genome, as it is the case, for instance, of *GS1b1* and *GS1b2* from *Pinus pinaster* [50]. However, other times paralogs are easy to confuse with orthologs. For instance, *GS1a* from *P. pinaster* and *Picea abies*, thought to be orthologs, have been unraveled as paralogs by means of tree reconciliation (Fig. 7). Regarding *GS2*, a similar situation occurs in angiosperms, with all genes being orthologs of those in gymnosperms and *Amborella trichopoda*, but paralogs among different groups of angiosperms (Fig. 9). This implies gene duplication processes followed by the elimination of one of the resulting copies. Taken together, these results support the existence of purifying selection of the pseudogenization and/or deletion type, probably related to genetic dosage and not to genetic sequence [51]. Both *GS1a* and *GS2* are highly expressed genes, so it is possible that additional copies of these genes increase GS activity to harmful levels or may produce gene interference effects. This could explain why in most groups of seed plants one of the lineages has been eliminated, such as *GS2* in gnetales and conifers and *GS1a* in modern angiosperms. On the other hand, *GS1b* remains in all seed plants and even undergoes gene expansion processes based on neofunctionalization of new genes, for example, *GS1b1* and *GS1b2* of *P. pinaster* [50].

Genome sequencing is changing our perception of the evolutionary history of gene families, which has been highly influenced by the initial biochemical data resulting from the characterization of gene products. This has been the case for the *GS* family in plants. Thanks to recent advances, we now understand that there are three main lineages in seed plants: *GS1a*, *GS1b*, and *GS2*. This has raised different questions that we have attempted to answer in the present work.

Our results indicate that the three lineages were already present in the common ancestor of seed plants. We have also established a timeline for the divergence of the three lineages currently present in plants. Thus, the first lineage to diverge was *GS2*, while it took an additional 70 Myr for *GS1a* to split from *GS1b*.

Finally, the results support a selection pressure on the *GS1a* and *GS2* lineages to maintain them as single-copy genes. This pressure may be related to their shared metabolic functions and a potential gene dosage effect.

Materials and methods

Taxon sampling

A total number of 45 Spermatophyta species were chosen following two criteria. First, their phylogenetic

relationships must be well established and widely accepted in the literature [32]. Secondly, the four Gymnosperm classes (Coniferophyta, Cycadophyta, Ginkgophyta, and Gnetophyta), as well as primitive and modern Angiosperms, should be represented in the species set. Details regarding the selected species, their taxonomy, and phylogeny are provided as Table S1.

Glutamine synthetase sequences

Nucleotide sequences of genes encoding plant glutamine synthetase were retrieved from various public databases or assembled from transcriptomic NGS data available in the SRA database at the National Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov) and the European Nucleotide Archive (ENA, https://www.ebi.ac.uk/ena), as described elsewhere [16]. All nucleotide and protein sequences used in the current work can be obtained after installing the R package ORTHGS and typing *AngGym* into the console.

Phylogenetic inference and tree rooting procedures

Phylogenetic relationships between GS protein sequences were inferred using two different approaches. One of these approaches used maximum likelihood (ML) analysis, which relies on multiple sequence alignment. The alternative approach consisted of an alignment-free method based on word statistics, which avoids the risks derived from incorrect sequence alignments [52]. Both approaches are implemented in R packages (ORTHGS and *EnvNJ*) available at CRAN. Further details for reproducing these analyses can be obtained at (https://github.com/jcaledo/Phylogeny-of-plant-GS/blob/main/Scripts/Figure2_Seed_Plants_4_Clusters.R).

Several methods for root inference have been described in the literature, including the popular outgroup (OG) and midpoint (MP) methods [53]. More recently, new methods of rooting a tree, based on their branch length distribution, have been introduced with considerable success [43,44]. Thus, in addition to the well-known midpoint and outgroup methods, we have implemented two additional methods aimed at accurately inferring the root position. The so-called minimal ancestor deviation (MAD) method is based on the principle of minimizing the deviation from the molecular clock assumption. Thus, the core idea is to identify the root position that minimizes the deviation in branch lengths from the expected values under a molecular clock [43]. On the other hand, the minimum variance (MV) method minimizes the variance of root to tip distances [44]. A fifth method for inferring the root node involves using tree reconciliation. Methods based on tree reconciliation extract rooting information from the discrepancies between gene trees and species trees. They consider gene duplications, transfers, and losses to find the root that minimizes these evolutionary events. All five different alternatives to locate the root node were implemented using the software indicated below. For the MP method, we used the function *midpoint* from the *phangorn* R package [54]. The OG and MV rooting procedures were developed using PYTHON software available at https://github.com/uym2/MinVar-Rooting, which implement the algorithm previously described in [44]. The sequences GLN2.1 and GLN2.2 from *Chlamydomonas reinhardtii* were used as outgroup. Regarding the method MAD, it was implemented by the R function *madRoot* from the package ORTHGS accompanying this paper. We must acknowledge that the algorithm used in *madRoot* is a slight modification of the code provided by Tria *et al.* [43], available at https://www.mikrobio.uni-kiel.de/de/ag-dagan/ressourcen. Finally, the rooting approach based on minimal reconciliation cost was implemented using the software RANGER-DTL, as previously described by [55].

Molecular clock dating analyses

Molecular clock dating was performed using a Bayesian approach implemented into MCMCTREE [56], a computer program that forms part of the PAML 4 package [57]. A detailed description of the protocol followed can be found at https://github.com/jcaledo/Phylogeny-of-plant-GS/tree/main/MCMCtree_GS_tutorial. In brief, an alignment in phylip format of 25 GS plant proteins from six representative species (see the figure in the tutorial cited above) and the corresponding calibrated tree were used as input data. The tree calibration was done using published fossil dating data [58]. Using these files as input, the maximum likelihood estimation of the branch lengths, the gradient vector, and the Hessian matrix were estimated in CODEML, using the LG + G amino acid substitution model [59], with five gamma rate categories. Divergence times were estimated under the independent rates model. The prior for the overall substitution rate was specified by gamma distributions with mean values of 0.005, 0.025, 0.05, and 0.4 substitutions/site/100 million years to assess the robustness of the resulting posterior distribution of divergence times. Once the robustness of the results with respect to the mean value of the gamma distribution was ensured, we fixed the mean at 0.05 substitutions/site/100 million years, which is a reasonable value for the rate of amino acid substitutions reported for Spermatophyta [60]. Subsequently, the MCMC analyses were repeated using different bounds for the root age, ranging from 500 to 1000 million years. For each of these conditions, two independent runs were performed, each consisting of 401 000 iterations, discarding the first 1000 generations as burn-in and sampling every 20 iterations, which resulted in a total of 20 000 samples post burn-in. To confirm convergence to the posterior time distribution and ensure the quality of posterior MCMC

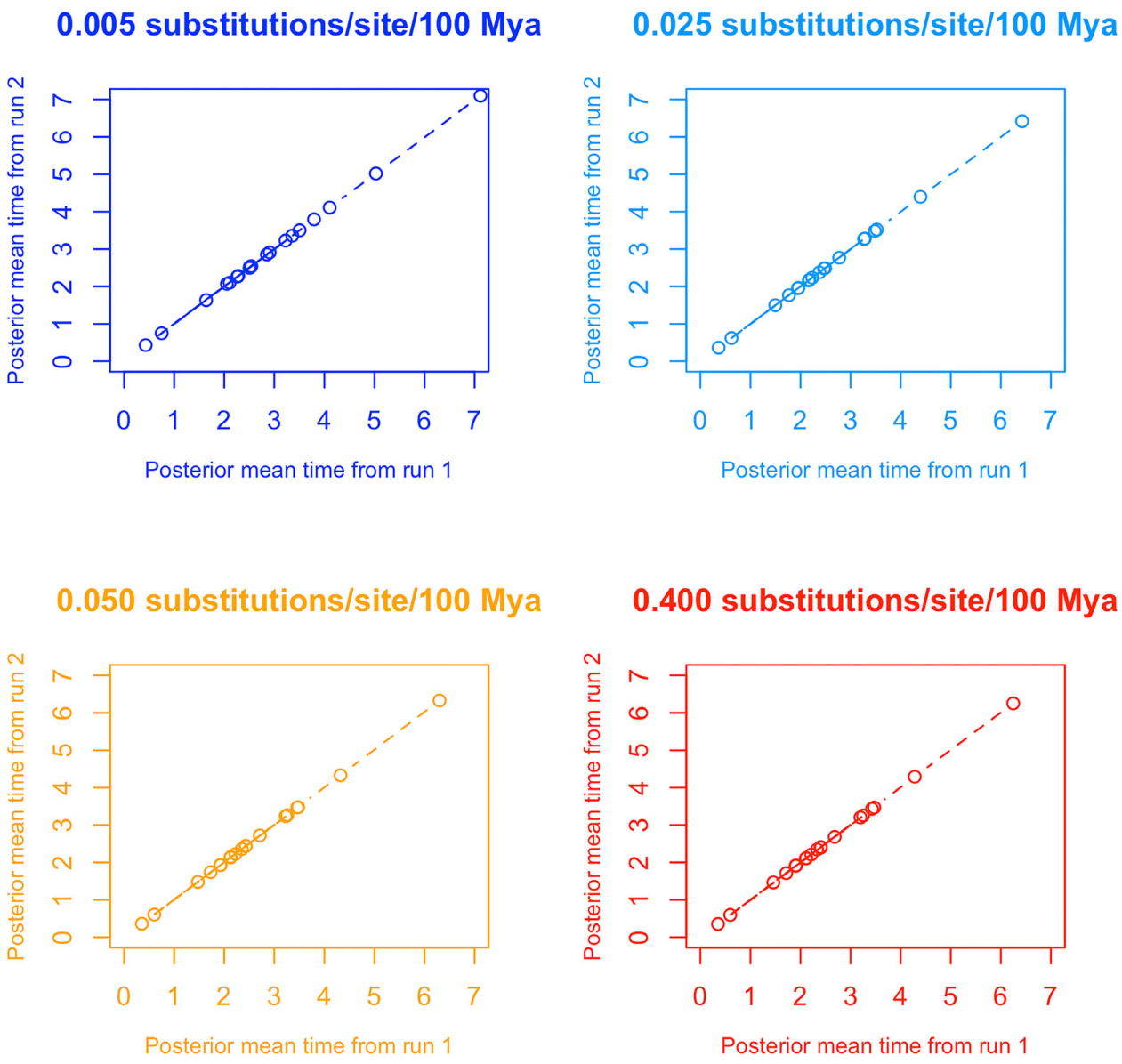

Fig. 10. Markov Chain Monte Carlo (MCMC) convergence plots. Using the fossil-calibrated tree of representative Spermatophyta species shown in Fig. 6A, two MCMC posterior sampling runs (starting with different seeds) were conducted for each parameter set. The parameter sets differed in the RootAge bound indicated in the figure. After the MCMC posterior sampling was finished, the mean times for the 24 internal nodes were computed and plotted (run 1 vs. run 2, for each parameter set).

samples, we conducted independent MCMC runs and examined convergence plots as diagnostic tests (Fig. 10).

Tree reconciliation and orthology networking

To infer the evolutionary histories of the GS gene family in the context of complex events like gene duplications, horizontal transfer, and losses, we resorted to a probabilistic framework based on parsimony implemented in the software RANGER-DTL previously described [55]. This software uses the species and the gene/protein trees as input, and it returns a file with annotation of the nodes of the gene/protein tree with evolutionary events (speciation, duplication, transfer, loss), along with a mapping onto the species tree. To facilitate the interpretation of the evolutionary history of the GS gene family, the mapping provided by RANGER-DTL was plotted using the software THIRDKIND [61]. To this end, it was necessary to convert the reconciliation file to an xml file beforehand, which was facilitated by the program RECPHYLOXML [62]. Further details and a script outlining the workflow

described above are available at https://github.com/jcaledo/Phylogeny-of-plant-GS/blob/main/Scripts/Figure7_Tree_Reconciliation.R. To construct orthology network graphs, we used the R packages ORTHGS and *igraph*. Detailed instructions are available in the accompanying vignette, accessible by entering *browseVignettes('orthGS')* in the RSTUDIO terminal. Additionally, a script to reproduce the results for the *GS1a* and *GS2* orthology networks (Fig. 9) is provided at https://github.com/jcaledo/Phylogeny-of-plant-GS/blob/main/Scripts/Figure9_GS1a_GS2_orth_networks.R.

Acknowledgements

We wish to express our special thanks to Dr Simon Penel from the Université de Lyon (Laboratoire de Biométrie et Biologie Évolutive), who generously provided us with a version of the THIRDKIND software specifically tailored to summarize our results on tree reconciliation and orthology networking. This work was funded by Ministerio de Ciencia e Innovación, grant number PID2021-122641NB-C21, MICINN, FEDER, UE. Funding for open access charge: Universidad de Málaga/CBUA.

Conflict of interest

The authors declare no conflict of interest.

Author contributions

EA and JCA wrote the ORTHGS software and performed the analyses. FRC, RAC, and JCA collected sequences. EA, RAC, and JCA wrote the article. FRC edited the manuscript. JCA planned and designed the research. All the authors discussed the results and reached a consensus on the conclusions.

Peer review

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/febs.70276.

Data availability statement

All data related to this manuscript can be found within this paper and its Supporting Information (Tables S1 and S2), as well as in a GitHub repository linked to this paper at https://github.com/jcaledo/Phylogeny-of-plant-GS. The code developed to address the analyses reported in this work can be downloaded from the CRAN repository at https://cran.r-project.org/web/packages/orthGS/index.html.

References

1 Kumada Y, Benson DR, Hillemann D, Hosted TJ, Rochefort DA, Thompson CJ, Wohlleben W & Tateno Y (1993) Evolution of the glutamine synthetase gene, one of the oldest existing and functioning genes. *Proc Natl Acad Sci U S A* **90**, 3009–3013.
2 Tateno Y (1994) Evolution of glutamine synthetase genes is in accordance with the neutral theory of molecular evolution. *Jpn J Genet* **69**, 489–502.
3 de Carvalho FG, Turchetto-Zolet AC & Pereira Passaglia LM (2022) Glutamine synthetase evolutionary history revisited: tracing back beyond the last universal common ancestor. *Evolution (N Y)* **76**, 605–622.
4 Unno H, Uchida T, Sugawara H, Kurisu G, Sugiyama T, Yamaya T, Sakakibara H, Hase T & Kusunoki M (2006) Atomic structure of plant glutamine synthetase: a key enzyme for plant productivity. *J Biol Chem* **281**, 29287–29296.
5 Van Rooyen JM, Abratt VR, Belrhali H & Sewell T (2011) Crystal structure of type III glutamine synthetase: surprising reversal of the inter-ring interface. *Structure* **19**, 471–483.
6 Doskočilová A, Plíhal O, Volc J, Chumová J, Kourová H, Halada P, Petrovská B & Binarová P (2011) A nodulin/glutamine synthetase-like fusion protein is implicated in the regulation of root morphogenesis and in signalling triggered by flagellin. *Planta* **234**, 459–476.
7 Silva LS, Seabra AR, Leitão JN & Carvalho HG (2015) Possible role of glutamine synthetase of the prokaryotic type (GSI-like) in nitrogen signaling in *Medicago truncatula*. *Plant Sci* **240**, 98–108.
8 Bernard SM & Habash DZ (2009) The importance of cytosolic glutamine synthetase in nitrogen assimilation and recycling. *New Phytol* **182**, 608–620.
9 Thomsen HC, Eriksson D, Møller IS & Schjoerring JK (2014) Cytosolic glutamine synthetase: a target for improvement of crop nitrogen use efficiency? *Trends Plant Sci* **19**, 656–663.
10 Fortunato S, Nigro D, Lasorella C, Marcotuli I, Gadaleta A & de Pinto MC (2023) The role of glutamine synthetase (GS) and glutamate synthase (GOGAT) in the improvement of nitrogen use efficiency in cereals. *Biomolecules* **13**, 1771.
11 Guan M, Møller IS & Schjoerring JK (2015) Two cytosolic glutamine synthetase isoforms play specific roles for seed germination and seed yield structure in *Arabidopsis*. *J Exp Bot* **66**, 203–212.
12 Moison M, Marmagne A, Dinant S, Soulay F, Azzopardi M, Lothier J, Citerne S, Morin H, Legay N, Chardon F *et al.* (2018) Three cytosolic glutamine synthetase isoforms localized in different-order veins act together for N remobilization and seed filling in *Arabidopsis*. *J Exp Bot* **69**, 4379–4393.

13 Kawakami N & Watanabe A (1988) Senescence-specific increase in cytosolic glutamine synthetase and its mRNA in radish cotyledons. *Plant Physiol* **88**, 1430–1434.
14 Kamachi K, Yamaya T, Mae T & Ojima K (1991) A role for glutamine synthetase in the remobilization of leaf nitrogen during natural senescence in rice leaves. *Plant Physiol* **96**, 411–417.
15 May O (2019) Industrial enzymes: overview of industrial enzyme applications. In Industrial Enzyme Applications (Andreas Vogel OM, ed.), pp. 3–24. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
16 Valderrama-Martín JM, Ortigosa F, Ávila C, Cánovas FM, Hirel B, Cantón FR & Cañas RA (2022) A revised view on the evolution of glutamine synthetase isoenzymes in plants. *Plant J* **110**, 946–960.
17 Martin A, Lee J, Kichey T, Gerentes D, Zivy M, Tatout C, Dubois F, Balliau T, Valot B, Davanture M *et al*. (2006) Two cytosolic glutamine synthetase isoforms of maize are specifically involved in the control of grain production. *Plant Cell* **18**, 3252–3274.
18 Cai H, Zhou Y, Xiao J, Li X, Zhang Q & Lian X (2009) Overexpressed glutamine synthetase gene modifies nitrogen metabolism and abiotic stress responses in rice. *Plant Cell Rep* **28**, 527–537.
19 Konishi N, Ishiyama K, Beier MP, Inoue E, Kanno K, Yamaya T, Takahashi H & Kojima S (2017) Contributions of two cytosolic glutamine synthetase isozymes to ammonium assimilation in *Arabidopsis* roots. *J Exp Bot* **68**, 613–625.
20 Kusano M, Fukushima A, Tabuchi-Kobayashi M, Funayama K, Kojima S, Maruyama K, Yamamoto YY, Nishizawa T, Kobayashi M, Wakazaki M *et al*. (2020) Cytosolic glutamine synthetase1;1 modulates metabolism and chloroplast development in ROOTs1 [open]. *Plant Physiol* **182**, 1894–1909.
21 Moreira E, Coimbra S & Melo P (2022) Glutamine synthetase: an unlikely case of functional redundancy in *Arabidopsis thaliana*. *Plant Biol* **24**, 713–720.
22 Doyle JJ (1991) Evolution of higher-plant glutamine synthetase genes: tissue specificity as a criterion for predicting orthology. *Mol Biol Evol* **8**, 366–377.
23 Pesole G, Bozzetti MP, Lanave C, Preparata G & Saccone C (1991) Glutamine synthetase gene evolution: a good molecular clock. *Proc Natl Acad Sci U S A* **88**, 522–526.
24 Brown JR, Masuchi Y, Robb FT & Doolittlel WF (1994) Evolutionary relationships of bacterial and archaeal glutamine synthetase genes. *J Mol Evol* **38**, 566–576.
25 Biesiadka J & Legocki AB (1997) Evolution of the glutamine synthetase gene in plants. *Plant Sci* **128**, 51–58.
26 Robertson DL & Tartar A (2006) Evolution of glutamine synthetase in heterokonts: evidence for endosymbiotic gene transfer and the early evolution of photosynthesis. *Mol Biol Evol* **23**, 1048–1055.
27 Ghoshroy S, Binder M, Tartar A & Robertson DL (2010) Molecular evolution of glutamine synthetase II: phylogenetic evidence of a non-endosymbiotic gene transfer event early in plant evolution. *BMC Evol Biol* **10**, 198.
28 Ghoshroy S & Robertson DL (2012) Molecular evolution of glutamine synthetase II and III in the chromalveolates. *J Phycol* **48**, 768–783.
29 McNally SF, Hirel B, Gadal P, Mann AF & Stewart GR (1983) Glutamine synthetases of higher plants: evidence for a specific isoform content related to their possible physiological role and their compartmentation within the leaf. *Plant Physiol* **72**, 22–25.
30 Hirel B, Mcnally SF, Gadal P, Sumar N & Stewart GR (1984) Cytosolic glutamine synthetase in higher plants. A comparative immunological study. *Eur J Biochem* **138**, 63–66.
31 Tingey SV, Tsai FY, Edwards JW, Walker EL & Coruzzi GM (1988) Chloroplast and cytosolic glutamine synthetase are encoded by homologous nuclear genes which are differentially expressed in vivo. *J Biol Chem* **263**, 9651–9657.
32 Ran J-H, Shen T-T, Wang M-M & Wang X-Q (2018) Phylogenomics resolves the deep phylogey of seed plants and indicates partial convergent or homoplastic evolution between Gnetales and angiosperms. *Proc R Soc B Biol Sci* **285**, 20181012.
33 Swarbreck SM, Defoin-Platel M, Hindle M, Saqi M & Habash DZ (2011) New perspectives on glutamine synthetase in grassses. *J Exp Bot* **62**, 1511–1522.
34 Zhao XQ & Shi WM (2006) Expression analysis of the glutamine synthetase and glutamate synthase gene families in young rice (*Oryza sativa*) seedlings. *Plant Sci* **170**, 748–754.
35 Paczek V, Dubois F, Sangwan R, Morot-Gaudry JF, Roubelakis-Angelakis KA & Hirel B (2002) Cellular and subcellular localisation of glutamine synthetase and glutamate dehydrogenase in grapes gives new insights on the regulation of carbon and nitrogen metabolism. *Planta* **216**, 245–254.
36 Martins M, Sousa B, Lopes J, Soares C, Machado J, Carvalho S, Fidalgo F & Teixeira J (2020) Diclofenac shifts the role of root glutamine synthetase and glutamate dehydrogenase for maintaining nitrogen assimilation and proline production at the expense of shoot carbon reserves in *Solanum lycopersicum L. Environ Sci Pollut Res Int* **27**, 29130–29142.
37 Morey KJ, Ortega JL & Sengupta-Gopalan C (2002) Cytosolic glutamine synthetase in soybean is encoded by a multigene family, and the members are regulated in an organ-specific and developmental manner. *Plant Physiol* **128**, 182–193.

38 Castro-Rodríguez V, García-Gutiérrez A, Canales J, Avila C, Kirby EG & Cánovas FM (2011) The glutamine synthetase gene family in *Populus*. *Plant Biol* **24**, 713–720.
39 Cantón FR, Suárez MF, Josè-Estanyol M & Cánovas FM (1999) Expression analysis of a cytosolic glutamine synthetase gene in cotyledons of Scots pine seedlings: developmental, light regulation and spatial distribution of specific transcripts. *Plant Mol Biol* **40**, 623–634.
40 De la Torre F, García-Gutiérrez A, Crespillo R, Cantón FR, Ávila C & Cánovas FM (2002) Functional expression of two pine glutamine synthetase genes in bacteria reveals that they encode cytosolic holoenzymes with different molecular and catalytic properties. *Plant Cell Physiol* **43**, 802–809.
41 Ávila Sáez C, Muñoz-Chapuli R, Plomion C, Frigerio JM & Cánovas FM (2000) Two genes encoding distinct cytosolic glutamine synthetases are closely linked in the pine genome. *FEBS Lett* **477**, 237–243.
42 Valderrama-Martín JM, Ortigosa F, Cantón FR, Ávila C, Cañas RA & Cánovas FM (2024) Emerging insights into nitrogen assimilation in gymnosperms. *Trees* **38**, 273–286.
43 Tria FDK, Landan G & Dagan T (2017) Phylogenetic rooting using minimal ancestor deviation. *Nat Ecol Evol* **1**, 193.
44 Mai U, Sayyari E & Mirarab S (2017) Minimum variance rooting of phylogenetic trees and implications for species tree reconstruction. *PLoS One* **12**, e0182238.
45 Fitch WM (1970) Distinguishing homologous from analogous proteins. *Syst Zool* **19**, 99–113.
46 Mondal R, Kumar A & Chattopadhyay SK (2021) Structural property, molecular regulation, and functional diversity of glutamine synthetase in higher plants: a data-mining bioinformatics approach. *Plant J* **108**, 1565–1584.
47 Hachiya T, Inaba J, Wakazaki M, Sato M, Toyooka K, Miyagi A, Kawai-Yamada M, Sugiura D, Nakagawa T, Kiba T *et al*. (2021) Excessive ammonium assimilation by plastidic glutamine synthetase causes ammonium toxicity in *Arabidopsis thaliana*. *Nat Commun* **12**, 4944.
48 Jordan IK, Kondrashov FA, Adzhubei IA, Wolf YI, Koonin EV, Kondrashov AS & Sunyaev S (2005) A universal trend of amino acid gain and loss in protein evolution. *Nature* **433**, 633–638.
49 Li Z, De La Torre AR, Sterck L, Cánovas FM, Avila C, Merino I, Cabezas JA, Cervera MT, Ingvarsson PK & Van De Peer Y (2017) Single-copy genes as molecularmarkers for phylogenomic studies in seed plants. *Genome Biol Evol* **9**, 1130–1147.
50 Valderrama-Martín JM, Ortigosa F, Aledo JC, Ávila C, Cánovas FM & Cañas RA (2023) Pine has two glutamine synthetase paralogs, GS1b.1 and GS1b.2, exhibiting distinct biochemical properties. *Plant J* **113**, 1330–1347.
51 Panchy N, Lehti-Shiu M & Shiu SH (2016) Evolution of gene duplication in plants. *Plant Physiol* **171**, 2294–2316.
52 Aledo JC (2022) Phylogenies from unaligned proteomes using sequence environments of amino acid residues. *Sci Rep* **12**, 7497.
53 Kinene T, Wainaina J, Maina S & Boykin LM (2016) Rooting trees, methods for. In Encyclopedia of Evolutionary Biology (Kliman RM, ed.), pp. 489–493. Academic Press, Oxford.
54 Schliep KP (2011) Phangorn: phylogenetic analysis in R. *Bioinformatics* **27**, 592–593.
55 Bansal MS, Kellis M, Kordi M & Kundu S (2018) RANGER-DTL 2.0: rigorous reconstruction of gene-family evolution by duplication, transfer and loss. *Bioinformatics* **34**, 3214–3216.
56 dos Reis M (2022) Dating microbial evolution with MCMCtree. *Methods Mol Biol* **2569**, 3–22.
57 Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**, 1586–1591.
58 Clarke JT, Warnock RCM & Donoghue PCJ (2011) Establishing a time-scale for plant evolution. *New Phytol* **192**, 266–301.
59 Le SQ & Gascuel O (2008) An improved general amino acid replacement matrix. *Mol Biol Evol* **25**, 1307–1320.
60 De La Torre AR, Li Z, Van De Peer Y & Ingvarsson PK (2017) Contrasting rates of molecular evolution and patterns of selection among gymnosperms and flowering plants. *Mol Biol Evol* **34**, 1363–1377.
61 Penel S, Menet H, Tricou T, Daubin V & Tannier E (2022) Thirdkind: displaying phylogenetic encounters beyond 2-level reconciliation. *Bioinformatics* **38**, 2350–2352.
62 Duchemin W, Gence G, Chifolleau AMA, Arvestad L, Bansal MS, Berry V, Boussau B, Chevenet F, Comte N, Davín AA *et al*. (2018) RecPhyloXML: a format for reconciled gene trees. *Bioinformatics* **34**, 3646–3652.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. Distribution of GS isoforms across species.

Table S2. Data from rooting analyses of 100 random samples of representative species.