Structure and evolution of Alanine/Serine Decarboxylases and the engineering of theanine production

Hao Wang[a,1], Biying Zhu[b,1], Siming Qiao[b], Chunxia Dong[b], Xiaochun Wan[b], Weimin Gong[a,*], Zhaoliang Zhang[b,*]

[a]Department of Life Sciences and Medicine, University of Science and Technology of China, Hefei, China.

[b]State Key Laboratory of Tea Plant Biology and Utilization, Anhui Agricultural University, Hefei, Anhui 230036, China.

[1]These authors contributed equally to this work.

* To whom correspondence should be addressed. Email: zhlzhang@ahau.edu.cn or wgong@ustc.edu.cn.

Author contributions: W.G, Z.Z and X.W designed research; H.W, B.Z, S.Q and C.D performed research; H.W and B.Z analyzed data; H.W, B.Z, W.G, and Z.Z wrote the paper.

Competing interest statement: The authors declare no competing financial interests.

Classification: Biochemistry (major), Plant Sciences (minor)

Keywords: alanine decarboxylase, serine decarboxylase, substrate specificity, theanine, *Camellia sinensis*

Abstract

Ethylamine (EA), the precursor of theanine biosynthesis, is synthesized from alanine decarboxylation by Alanine Decarboxylase (AlaDC) in tea plants. AlaDC evolves from Serine Decarboxylase (SerDC) through neofunctionalization and has lower catalytic activity. However, lacking structure information hinders the understanding of the evolution of substrate specificity and catalytic activity. In this study, we solved the X-ray crystal structures of AlaDC from *Camellia sinensis* (CsAlaDC) and SerDC from *Arabidopsis thaliana* (AtSerDC). Tyr^{341} of AtSerDC or the corresponding Tyr^{336} of CsAlaDC is essential for their enzymatic activity. Tyr^{111} of AtSerDC and the corresponding Phe^{106} of CsAlaDC determine their substrate specificity. Both CsAlaDC and AtSerDC have a distinctive zinc finger that contributes to the activity and has not been identified in any other Group II PLP-dependent amino acid decarboxylases. Based on the structure comparison and evolution analysis AlaDC or SerDC led us to perform mutation screen of CsAlaDC. The results indicated that the mutation of L110F or P114A in the CsAlaDC dimerization interface significantly improved the catalytic activity by 110% and 59%, respectively. Combining double mutant $CsAlaDC^{L110F/P114A}$ with theanine synthetase increased 672% of theanine production in an *in vitro* system. This study provides the structural basis for the substrate selectivity and catalytic activity of CsAlaDC and AtSerDC and sheds light on a new direction for the efficient industrial synthesis of theanine.

Introduction

Tea is one of the most popular natural non-alcoholic beverages consumed worldwide. About 2 billion cups of tea are consumed worldwide daily (1). The popularity of tea is determined by its favorable flavor and numerous health benefits (2-4). The flavor and health-beneficial effects of tea are conferred by the abundant secondary metabolites, including catechins, caffeine, theanine, volatiles, etc (5). Theanine (γ-glutamylethylamide) is a unique non-protein amino acid and the most abundant free amino acid in tea plants (*Camellia sinensis*). It accounts for more than 50% of the total free amino acids and approximately 1-2% of the dry weight of the new shoots of tea plants (6). Theanine is the secondary metabolite conferring the umami taste of tea infusion and also balances the astringency and bitterness of tea infusion caused by catechins and caffeine (7). It has also many health-promoting functions, such as reducing stress and anxiety, improving mood and cognition, protecting the nervous system, etc (6, 8-13). Therefore, theanine content is highly correlated with green tea quality (14).

Theanine is synthesized from EA and glutamate (Glu) by theanine synthetase (15, 16). Importantly, the large amount of theanine biosynthesis is determined by the high availability of EA in tea plants (17). Therefore, the evolution of EA biosynthesis in tea plants largely determined the quality formation of tea. EA is synthesized from alanine decarboxylation by alanine decarboxylase (CsAlaDC) (18) (Figure 1A). Indeed, CsAlaDC expression level and catalytic activity largely determine the theanine accumulation level in tea plants (19). As a novel gene, *CsAlaDC* had not been reported

62 in other plant species before it was identified in tea plants. Previous studies indicated
63 that *AlaDC* originated from the serine decarboxylase gene (*SerDC*) by gene duplication
64 in tea plants (20). However, CsAlaDC and CsSerDC specifically catalyze alanine and
65 serine decarboxylation, respectively, despite they share highly conserved amino acid
66 sequences (20) (Figure 1A, 1C). In addition, CsAlaDC exhibited a much lower
67 enzymatic activity compared with CsSerDC. However, the structural basis and key
68 amino acids underlying the evolution of the substrate specificity and enzymatic activity
69 of CsAlaDC are unknown.

70 SerDC belongs to the Group II pyridoxal-5-phosphate (PLP)-dependent
71 decarboxylase superfamily (21). These Group II amino acid decarboxylases produce
72 many bioactive secondary metabolites and signaling molecules (22-24). In plants,
73 SerDC catalyzes the biosynthesis of ethanolamine from serine, which was first
74 functionally characterized in *Arabidopsis thaliana* (25) (Figure 1A). Ethanolamine is
75 an important metabolite in the synthesis of phosphatidylethanolamine and
76 phosphatidylcholine, two main phospholipids involved in maintaining eukaryotic
77 membrane structures in eukaryotic cell membranes (26-28). Furthermore, the study has
78 shown that SerDC is essential in the embryogenesis of *Arabidopsis* (29). SerDC is of
79 vital importance for plant growth and development, but the mechanism of substrate
80 recognition and catalytic activity of SerDC also remains unclear.

81 In this work, we attempt to understand the mechanism of functional divergence of
82 plant AlaDC and SerDC by analyzing their structural characteristics. Here, we obtained
83 the X-ray crystal structure of CsAlaDC and AtSerDC. According to the crystal

84 structure, we found a distinctive zinc finger structure located at both CsAlaDC and
85 AtSerDC, which has not been identified in any other Group II PLP-dependent amino
86 acid decarboxylases that have been previously characterized. By comparing the
87 substrate binding pockets, we identified Phe^{106} of CsAlaDC and Tyr^{111} of AtSerDC as
88 the crucial sites for substrate specificity. By conducting mutation screening based on
89 the protein structures, we identified the amino acids repressing the catalytic activity and
90 discovered that $CsAlaDC^{L110F/P114A}$ exhibited a 2.3-fold increase in catalytic activity
91 compared to that of CsAlaDC and enhanced the engineering of theanine production *in*
92 *vitro*.
93
94 **Results**
95 **Enzymatic properties of CsAlaDC, AtSerDC, and CsSerDC**
96 CsAlaDC originates from CsSerDC by gene duplication and neofunctionalization in tea
97 plants, but they catalyze different metabolic processes (Figure 1A). A phylogenetic tree
98 was constructed to further reveal the evolutionary relationship between CsAlaDC and
99 its homologs in *Algal*, *Pteridohyta*, *Bryophyta*, *Monocots* and *Dicots* (Figure 1B).
100 Phylogenetic analysis indicated that CsAlaDC is homologous with SerDCs, and its
101 closer relationship was Dicots plants. As expected, CsSerDC was most closed to
102 AtSerDC, which implies that they shared similar functions. However, CsAlaDC is
103 relatively distant from CsSerDC. Next, we performed a multiple sequence alignment of
104 the amino acid sequences of the 6 SerDCs and CsAlaDC (Figure 1C). The results
105 showed that the amino acid sequences of CsAlaDC and SerDCs were highly conserved,

106 but CsAlaDC has amino acid mutations in conserved motifs compared with other
107 SerDCs.

108 Next, to verify the substrate specificity and enzyme activity, we conducted enzyme
109 activity detection and enzyme kinetics analysis for CsAlaDC, CsSerDC and AtSerDC.
110 The 5′ truncated *CsAlaDC* and *AtSerDC* were constructed into the pET22b and pET28a
111 expression vectors to generate recombinant plasmids pET22b-*CsAlaDC* and pET28a-
112 *AtSerDC*, correspondingly. Additionally, the full-length protein of CsSerDC was
113 inserted into the pET28a expression vector to generate recombinant plasmid pET28a-
114 *CsSerDC*. Subsequently, the recombinant proteins were expressed in *Escherichia coli*
115 and purified via nickel affinity chromatography. The purified proteins were examined
116 through SDS-PAGE analysis (Figure 2A). To investigate the substrate specificity of
117 CsAlaDC, AtSerDC, and CsSerDC, enzyme activity assays were conducted using Ala
118 and Ser as substrates, followed by product identification via UPLC analysis. The results
119 showed that CsAlaDC can effectively participate in the decarboxylation of alanine to
120 generate EA, whereas its catalytic efficacy towards serine is relatively inferior.
121 Conversely, AtSerDC and CsSerDC selectively catalyzed serine decarboxylation to
122 yield ethanolamine but did not exhibit activity towards alanine (Figure 2B). These
123 findings confirmed that CsAlaDC and SerDCs do not have promiscuous decarboxylase
124 activity on Ala and Ser.

125 The kinetic properties of CsAlaDC, AtSerDC, and CsSerDC were determined
126 through the use of corresponding substrates (Figure 2C), and the results are presented
127 in Table 1. The kinetic parameter Km of CsAlaDC was determined to be 1.215 mmol/L,

which is similar to that of AtSerDC and CsSerDC at 1.522 mmol/L and 2.364 mmol/L, respectively. However, AtSerDC has a Vmax of 7.053 μmol/L/s, which is 4-fold that of CsAlaDC's Vmax of 1.709 μmol/L/s; while CsSerDC has a Vmax of 9.031 μmol/L/s, which is 5.3-fold that of CsAlaDC. These findings indicated that the catalytic efficiency of CsAlaDC is considerably lower than that of both CsSerDC and AtSerDC. Due to the similarity in catalytic activity between CsSerDC and AtSerDC, we chose the more representative AtSerDC for further analysis.

The overall structures of CsAlaDC and AtSerDC

To enhance our comprehension of CsAlaDC and AtSerDC, we conducted structural analyses of these two proteins. Through optimization of crystallization conditions, we successfully determined the crystal structures of CsAlaDC, as well as the CsAlaDC-EA, and AtSerDC at resolutions of 2.50 Å, 2.60 Å and 2.85 Å, respectively (Supplementary table 1).

The overall architecture of CsAlaDC and AtSerDC comprises homodimeric structures with two subunits exhibiting an asymmetrical arrangement (Figure 3A and 3D). The monomer of CsAlaDC and AtSerDC is divided into three distinct structural domains: an N-terminal domain (N-terminal -104 aa in CsAlaDC, N-terminal -109 aa in AtSerDC), a large domain (105-364 aa in CsAlaDC, 110-369 aa in AtSerDC), and a small C-terminal domain (365 aa-C-terminus in CsAlaDC, 370-C-terminus as in AtSerDC), colored light pink, khaki, and sky blue, respectively. Compared to the full-length protein, the N-terminal structures of CsAlaDC and AtSerDC are truncated by 60

150 and 65 amino acid residues, respectively. The truncated protein N-terminus contains a
151 long α-helix, which is connected to the large domain through a long loop. The large
152 domain contains a seven-stranded mixed β-sheet surrounded by eight α-helices, while
153 the C-terminal domain composes of a three-stranded antiparallel β-sheet and three α-
154 helices. The catalytic site of the enzyme is positioned within a superficial crevice at the
155 junction of two subunits forming a dimeric arrangement. The cofactor-binding process
156 involves amino acid residues derived from both subunits. One monomer accommodates
157 both PLP and substrate, while the other monomer also contributes to cofactor binding
158 and enzymatic function (Figure 3A and 3D).

159 Notably, our investigation has revealed the presence of a distinctive zinc finger
160 structure (as depicted in Figure 3A and 3D) located at the C-terminus of CsAlaDC and
161 AtSerDC. This structure is composed of a ring structure spanning 17 amino acid
162 residues, wherein coordination of Zn^{2+} is facilitated by three Cys residues and one His
163 residue. Importantly, this particular configuration is exclusive to the two proteins under
164 examination and has not been identified in any other Group II PLP-dependent amino
165 acid decarboxylases that have been previously characterized.

166 The crystal structures of CsAlaDC-EA complex and AtSerDC were obtained and
167 further analyzed. In the former complex, PLP bound to Lys^{309} via a Schiff base linkage
168 (internal aldimine form), and the pyridine moiety of PLP is positioned between the
169 imidazole ring of His^{196} and the methyl group of Ala^{279} in a parallel orientation to the
170 imidazole ring. The N1 of PLP establishes a salt bridge interaction with Asp^{277}, while
171 the O3 forms hydrogen bonds with Thr^{247} and Lys^{309}. Additionally, the phosphate group

172 of PLP participates in hydrogen bonding interactions with His^{308}, Gly^{169}, Thr^{170}, Lys^{309},
173 and Ser^{347}*("*" denotes the amino acids on adjacent subunits.) (Figure 3C). Similarly,
174 in the latter structure, the pyridine ring of PLP was sandwiched between the imidazole
175 ring of His^{201} and the methyl group of Ala^{284}. The N1 of PLP formed a salt bridge with
176 Asp^{282}, while the O3 formed hydrogen bonds with Thr^{252}. The phosphate group of PLP
177 established hydrogen bonding interactions with Lys^{314}, His^{313}, Gly^{174}, Thr^{175}, and
178 Ser^{352}* (Figure 3E).
179
180 **Factors affecting CsAlaDC and AtSerDC activity**
181 The Group II PLP-dependent amino acid decarboxylases are dimeric enzymes with
182 each monomer containing an active site located within a shallow cavity (30). A flexible
183 loop of amino acids originating from one monomer extends into the active site of the
184 other monomer and plays a crucial role in catalysis. The catalytic loop in CsAlaDC,
185 consisting of amino acid residues 328-341, is well-structured and exhibits clear electron
186 density. However, the corresponding loop in AtSerDC is disordered (Figure 3F). The
187 aforementioned loop harbors a conserved Tyr residue that is believed to function as a
188 proton donor of the carbanion in the catalytic process (31). In particular, CsAlaDC
189 Tyr^{336} and AtSerDC Tyr^{341} correspond to that Tyr residue. Upon superimposing the
190 active site residues of CsAlaDC apo and CsAlaDC-EA complex, it was observed that
191 the hydroxyl group of Tyr^{336}* in the CsAlaDC-EA complex exhibits a substantial 60-
192 degree deviation along the Cα-Cβ bound (Figure 3G). To assess the function of the Tyr,
193 we substituted the corresponding Tyr residues in CsAlaDC and AtSerDC with Phe. The

194 resulting CsAlaDC^{Y336F} and AtSerDC^{Y341F} mutants were then exposed to L-alanine and
195 L-serine, respectively. We measured the production of EA or ethanolamine in the
196 reaction mixtures. Our findings indicate that these mutants catalyze abortive
197 decarboxylation, as opposed to wild-type CsAlaDC and AtSerDC, as evidenced by the
198 absence of detectable EA and only a small amount of ethanolamine observed in the
199 reaction mixture (Figure 3H). This result suggested this Tyr is required for the catalytic
200 activity of CsAlaDC and AtSerDC.

201 CsAlaDC and AtSerDC both possess a C-terminal zinc finger motif whose removal
202 led to protein insoluble when expressed in *E. coli*. Therefore we used dithiothreitol
203 (DTT) to block the zinc finger structure to evaluate the role of this zinc finger structure
204 in the catalytic activity. The results showed that 5 mM L-DTT reduced the relative
205 activity of CsAlaDC and AtSerDC to 22.0% and 35.2%, respectively, compared to the
206 control (Figure 3I). These results suggested that this characteristic zinc finger motif is
207 probably critical for the protein folding of CsAlaDC and AtSerDC.

208

209 **Identification of key amino acids for the substrate specificity**

210 Through the superposition of the amino acid residues present in the substrate binding
211 pocket of CsAlaDC and AtSerDC, it was observed that except for Phe at position 106
212 in CsAlaDC (Figure 4A), the residues were identical. In AtSerDC, this position is
213 occupied by Tyr at 111, suggesting that this specific amino acid residue is related to
214 substrate specificity.

215 To obtain more insights into the role of specific amino acid residue in substrate

216 specificity, we utilized amino acid sequences of AtSerDC to identify all potential serine
217 decarboxylases in *Embryophyta* and constructed phylogenetic trees (Supplementary
218 figure 1). We further identified conserved motifs in these SerDC homologs in
219 *Embryophyta* (Figure 4B). Importantly, the Phe^{106} of CsAlaDC or Tyr^{111} of AtSerDC
220 locates in a conserved motif (Figure 4B). In this conserved motif, the first two amino
221 acids Y and P are 100% conserved, but the third amino acid is variable and can be Y, T,
222 F, V, A, I, or other amino acids (Figure 4B and C); interestingly, the Phe^{106} of CsAlaDC
223 and Tyr^{111} of AtSerDC locates at this variable amino acid position. For these first three
224 amino acids, the YPY was identified in 83.7% of the SerDC homologs, while the YPF
225 motif was found in 2.3% of these SerDC homologs (Figure 4C). These analyses
226 provided more insights into the role of Phe^{106} of CsAlaDC and Tyr^{111} of AtSerDC in the
227 substrate specificity.

228 To verify the role of Phe^{106} of CsAlaDC or Tyr^{111} of AtSerDC in substrate
229 specificity, we mutated Phe^{106} of CsAlaDC into Tyr ($CsAlaDC^{F106Y}$) and Tyr^{111} of
230 AtSerDC into Phe ($AtSerDC^{Y111F}$) and tested enzymatic activity *in vitro*. The results
231 showed that $CsAlaDC^{F106Y}$ totally lost the alanine decarboxylase activity, and
232 $AtSerDC^{Y111F}$ gained alanine decarboxylase activity (Figure 4E). We also performed
233 other mutations of Tyr^{111} of AtSerDC, including $AtSerDC^{Y111A}$, $AtSerDC^{Y111I}$,
234 $AtSerDC^{Y111L}$, $AtSerDC^{Y111V}$ and $AtSerDC^{Y111W}$. Within these mutations, only
235 $AtSerDC^{Y111I}$ exhibited a low level of alanine decarboxylase activity (Figure 4E). These
236 results verified that the Phe^{106} of CsAlaDC is required for the alanine decarboxylase

activity from serine decarboxylase activity.

On the other side, both CsAlaDC and CsAlaDC^{F106Y} exhibit a very low level of serine decarboxylase activity (Figure 4F). Moreover, compared with AtSerDC, the serine decarboxylase activity of AtSerDC^{Y111F} was about 30% of AtSerDC; AtSerDC^{Y111I} retained lower than 5% of serine decarboxylase activity of AtSerDC; while other mutations, including AtSerDC^{Y111A}, AtSerDC^{Y111L}, AtSerDC^{Y111V} and AtSerDC^{Y111W} abolished the serine decarboxylase activity (Figure 4F). These results suggested the Tyr^{111} of AtSerDC is also important for the substrate specificity of AtSerDC.

To further verify that Phe^{106} of CsAlaDC and Tyr111 of AtSerDC were key amino acid residues determining its substrate recognition *in planta*, *Nicotiana benthamiana* transient expression system was carried out. To this end, *A. tumefaciens* strain GV3101 (pSoup-p19), carrying recombinant plasmid, was infiltrated into leaves of 5-week-old *N. benthamiana* plants, and the pCAMBIA1305 empty vector was used as the control (EV). Relative mRNA levels were detected (Supplementary figure 2), indicating that they had been successfully overexpressed in tobacco. Here, we found that a high level of EA was produced in the CsAlaDC-expressing tobacco leaves; while no EA product was detected in tobacco leaves infiltrated with the mutant CsAlaDC^{F106Y}. In addition, we did not detect EA products in the AtSerDC-expressing tobacco leaves, while a high level of EA was detected in tobacco leaves infiltrated with mutant AtSerDC^{Y111F}. As anticipated, EA was not detected in tobacco leaves infiltrated with EV (Figure 4G and H). These results further verified the critical role of Phe^{106} in the substrate specificity

259 of CsAlaDC.

260

261 **Key amino acids for the evolution of CsAlaDC enzymatic activity**

262 CsAlaDC and AtSerDC have a high sequence similarity of 74.5% and a nearly identical
263 structure with an RMSD (root mean square deviation) of only 0.77Å for monomer
264 structures. However, the two enzymes catalyze different amino acid decarboxylation
265 reactions, and AtSerDC exhibits significantly higher activity than CsAlaDC (Figure 2B
266 and C). To elucidate this observation, we proposed a prediction test: EA, generated via
267 Ala decarboxylation by CsAlaDC, can be toxic and harmful to plants if accumulated
268 excessively. Thus, during the evolution of plant serine decarboxylase into alanine
269 decarboxylase, the enzyme has evolved not just to alter the substrate preference but also
270 to reduce catalytic activity to control EA production within a suitable range. Based on
271 this hypothesis, we suggest that mutating specific amino acids in CsAlaDC to those
272 corresponding amino acids in SerDCs could enhance its activity, and the results could
273 provide insights into the evolution of CsAlaDC enzymatic activity.

274 Dimerization is essential to AlaDC/SerDC activities because the active site is
275 composed of residues from two monomers. To identify the amino acids repressing the
276 enzymatic activity of CsAlaDC during evolution from SerDC, we analyzed the crystal
277 structures and the amino acids at the dimer interface between CsAlaDC and AtSerDC.
278 This analysis revealed that the amino acid residues at positions 66, 97, 110, 114, 116,
279 117, 122, 315 and 345 are different in CsAlaDC and SerDCs (Figure 1C). Therefore,
280 we mutated these amino acids of CsAlaDC into those in the corresponding positions of

281 AtSerDC and CsSerDC, and performed enzyme activity assays (Figure 5A). The results
282 demonstrated that CsAlaDC^{L110F} and CsAlaDC^{P114A} exhibited significantly enhanced
283 enzyme activity compared to the wild-type CsAlaDC, with a 2.1-fold and 1.59-fold
284 increase, respectively. Furthermore, the catalytic activity of the CsAlaDC$^{L110F/P114A}$
285 double mutant exhibits a remarkable 2.3-fold enhancement compared to that of the
286 wild-type protein (Figure 5B). These findings suggested a critical role of Leu^{110} and
287 Pro^{114} of CsAlaDC in the evolution of enzymatic activity and provided a basis to
288 improve CsAlaDC activity. It is possible that these amino acid residues could
289 potentially augment the hydrophobic nature of the protein dimer interface.
290

291 ***In Vitro* synthesis of theanine**

292 The biosynthetic pathway of theanine in tea plants comprises two sequential enzymatic
293 steps: alanine decarboxylase mediates alanine decarboxylation to yield EA, which is
294 subsequently utilized by theanine synthetase to effectuate the condensation reaction
295 with glutamate to synthesize theanine. By conducting mutation screening, we
296 discovered a CsAlaDC mutant protein (L110F/P114A) that exhibited a 2.3-fold increase
297 in catalytic activity compared to the wild-type protein. Subsequently, we employed an
298 *in vitro* theanine synthesis system utilizing CsAlaDC and either glutamate synthetase
299 (PtGS [P*seudomonas syringae* pv. syringae]) or gamma-glutamate methylamine ligase
300 (GMAS [*Methylovorus mays*]) which have the activity to synthesize theanine from EA
301 and Glu (32, 33).

302 The results illustrated the successful synthesis of theanine using CsAlaDC in

303 conjunction with two theanine synthetases, employing Ala and glutamate as substrates
304 (Figure 5C). The theanine content generated by the combination of CsAlaDC^{L110F} and
305 GMAS, as well as CsAlaDC$^{L110F/P114A}$ and GMAS, are 4.57-fold and 6.72-fold higher
306 than the content produced in the combination of wild-type CsAlaDC and GMAS,
307 respectively. Similarly, when combined with the PsGS, comparable outcomes are
308 observed as well. The theanine content resulting from the combination of CsAlaDC^{L110F}
309 and PsGS, as well as CsAlaDC$^{L110F/P114A}$ and PsGS, exhibit enhancements of 1.62-fold
310 and 4.33-fold compared to the wild-type protein combination, respectively. (Figure 5C).
311 Hence, the utilization of CsAlaDC$^{L110F/P114A}$ could effectively enhance the theanine
312 production yield and thus, holds potential for large-scale engineering production of
313 theanine.
314

315 **Discussion**

316 **Comparison of CsAlaDC and AtSerDC with other decarboxylases**

317 The Dali search (34) revealed that AtSerDC and CsAlaDC exhibit structural similarity
318 to several other decarboxylases, including 7CIG (MetDC [*Streptomyces sp*. 590]),
319 7ERV (HisDC1 [*Photobacterium phosphoreum*]), 6KHO (TrpDC [*Oryza sativa*
320 *Japonica* Group]), 4E1O (HisDC2 [*Homo sapiens*]), 6JY1 (AspDC
321 [*Methanocaldococcus jannaschii*]), and 5GP4 (GluDC [*Levilactobacillus brevis*]). The
322 monomers of these amino acid decarboxylases all consist of three characteristic
323 domains (Figure 3 and Supplementary figure 4). The C-terminal domain, also known
324 as the small domain, is composed of three antiparallel β-sheets and three α-helices,

325 facing toward the opposite monomers. The large domain contains seven stranded β-
326 sheets surrounded by eight α-helices, exhibiting characteristics typical of Type I
327 enzymes. This domain includes a conserved PLP-binding active center and amino acid
328 residues that are conserved across species. The N-terminal domain consists of elongated
329 α-helices and is connected to the large domain via a loop. The helix of one subunit is
330 antiparallel to the corresponding helix in the neighboring subunit, forming a clamp.
331 Therefore, the N-terminal domain may not be an independent folding unit and is likely
332 only stable in dimers.

333 The monomeric structure of CsAlaDC and AtSerDC exhibits numerous structural
334 features common to other Group II PLP-dependent amino acid decarboxylases, and the
335 amino acid residues that bind to PLP cofactors are conserved across multiple enzymes
336 (Supplementary figure 3 and 4). Out of the eight proteins analyzed, a total of thirteen
337 amino acid residues were found to be conserved across all sequences (Glu^{143}, Glu^{171},
338 Lys^{201}, Gly^{245}, Thr^{247}, Asp^{253}, His^{275}, Asp^{277}, Ala^{279}, Pro^{286}, Ser^{302}, Lys^{309}, and Tyr^{336} in
339 CsAlaDC). Some of them are situated within the active center and play a role in
340 stabilizing PLP or facilitating catalytic reactions. Conversely, other residues reside
341 outside the active center and their function remains unclear. (Supplementary figure 3).
342 PLP is bound to conserved lysine residues in the active center through the Schiff base.
343 The carboxylic group of conserved Asp residues stabilizes protonated pyridine nitrogen
344 of PLP by electrostatic action, providing the latter with strong electrophilicity to
345 stabilize carbanion intermediates during enzymatic catalysis. Another conserved His
346 residue is situated laterally to the PLP, while on the opposite side, there is a conserved

347 Ala residue. The oxygen atom at position 3 of PLP is stabilized by hydrogen bonding
348 with the Thr residue. The phosphate moiety of PLP is additionally secured to the protein
349 through an elongated system of hydrogen bonding. The α5 helix located within the large
350 domain exhibits evolutionary conservation in Group II enzymes, facilitating interaction
351 with the PLP phosphate moiety and conferring appropriate stabilization thereof.

352 The superimposition of the crystal structures of CsAlaDC and AtSerDC with those
353 of related enzymes suggests that residue Phe106 in CsAlaDC (corresponding to Tyr in
354 AtSerDC and Glu in MetDC) may play a role in determining substrate specificity. This
355 hypothesis was tested via mutagenesis experiments, which revealed that substitution of
356 Phe^{106} for Tyr rendered CsAlaDC inactive, while substitution of Tyr^{111} for Phe in
357 AtSerDC enabled alanine decarboxylase activity. Likewise, the substitution of Glu^{64} for
358 Ala in MetDC reduced enzyme activity by approximately 50-fold compared to the wild-
359 type (35), supporting this conclusion.

360 Group II amino acid decarboxylase is characterized by the existence of highly
361 flexible catalytic rings, which are of great significance for the catalytic mechanism of
362 decarboxylase (31). The catalytic rings are located at the dimer interface and extend to
363 the active sites of other monomers in a closed conformation. In the structure of
364 CsAlaDC, the catalytic ring is orderly and allocated, while in the structure of AtSerDC,
365 it is disordered and lacks electron density (Figure 3F). Prior research has established
366 that the conserved amino acid residue Tyr within the ring plays a crucial role in catalysis
367 by donating protons to carbanions of quinone intermediates that arise following
368 decarboxylation (36, 37). Our experiments demonstrate that the substitution of the

369 corresponding Tyr with Phe in the loop renders the protein inactive. Furthermore,
370 alterations in the structure of CsAlaDC when combined with EA provide further
371 evidence supporting this conclusion. Specifically, the structure of CsAlaDC-EA
372 complex displayed a significant conformational change in Tyr^{336}* located in the
373 catalytic site (as depicted in Figure 3G). Comparing the structure of CsAlaDC apo, the
374 CsAlaDC-EA complex exhibited a notable 60-degree shift in the p-hydroxyl group of
375 Tyr^{336}* along the Cα-Cβ bond.

376 The monomeric configuration of CsAlaDC and AtSerDC is akin to that of other
377 Group II amino acid decarboxylases, except for a conspicuous dissimilarity. Unlike
378 other Group II amino acid decarboxylases, the C-terminal of CsAlaDC and AtSerDC
379 harbors an evident zinc finger structure. We propose that this structure could potentially
380 influence enzyme stability. To test this hypothesis, we truncated the zinc finger structure
381 from both proteins and expressed them. Our findings indicate that, following the
382 excision of the zinc finger structure, CsAlaDC became insoluble, while AtSerDC
383 displayed a similar trait to a certain extent.

384

385 **Evolution of AlaDC and SerDC**

386 The sequence identity between CsAlaDC and AtSerDC is 74.5%, indicating a
387 significant degree of homology. Moreover, the two proteins exhibit a high level of
388 monomer structural similarity. Notably, they differ in substrate specificity. Using the
389 crystal structure of CsAlaDC and AtSerDC as a basis, we identified pivotal amino acid
390 residues that govern the substrate specificity of serine or alanine decarboxylase. We

391 observed that sequences proximal to these sites exhibit a high degree of conservation
392 in both proteins and possess characteristic motifs (Figure 4B). Moreover, we analyzed
393 serine decarboxylase-like proteins in *Embryophyta* that are highly homologous to
394 CsAlaDC and constructed phylogenetic trees to gain insight into their evolutionary
395 relationships. Based on our experimental findings and evolutionary evidence, we have
396 identified a conserved YPX motif at the substrate binding pocket. For most plants, the
397 residue X in this motif is predominantly Tyr, which corresponds to the serine
398 decarboxylase protein. However, other residues such as Thr, Phe, Val, Ala, and Ile can
399 also occupy this position (Figure 4B and C). Interestingly, serine decarboxylase-like
400 proteins containing YPF motifs, which have the potential to function as alanine
401 decarboxylases, were found to be distributed throughout the phylogenetic tree,
402 including *Asteroideae*, *Ericales*, *Chenopodiaceae*, *Poaceae*, and *Ranunculales*
403 (Supplementary figure 1B). However, these proteins were absent in some more recent
404 species. This observation implies that the emergence of alanine decarboxylase in these
405 particular species could be attributed to convergent evolution.

406 Moreover, we have observed the enrichment of serine decarboxylase-like proteins
407 that possess a YPT motif within the *Fabales*. Prior research has demonstrated that
408 XP_004496485-1, which possesses a YPT motif, displays catalytic efficacy toward the
409 processes of decarboxylation and oxidative deamination of Phe, Met, Leu, and Trp (38).
410 Given that the YPT motif is highly conserved and widely distributed in *Fabales*, serine
411 decarboxylase-like proteins bearing the YPT motif may have developed a unique
412 substrate specificity in *Fabales*, beyond their conventional decarboxylation functions,

413 as exemplified by XP_004496485-1 protein mentioned above. We speculate that these
414 proteins may be capable of catalyzing other reactions, potentially involving non-protein
415 amino acids or other substances as substrates.
416
417 **Applied to improve the synthesis of theanine**
418 Theanine is an important indicator of green tea quality. Therefore, improving the
419 synthesis of theanine in tea plants is the focus of research. In this study, through crystal
420 structure analysis and mutation verification, we found that the catalytic activity of
421 $CsAlaDC^{L110F/P114A}$ is 2.3 times higher than that of the wild-type protein, resulting in a
422 more abundant synthesis of theanine *in vitro*. This gives us great inspiration to improve
423 the synthesis of EA in tea plants by gene editing, thus increasing the content of theanine
424 in tea plants (Figure 5).
425 Theanine is highly demanded, by the market, due to its health effects and medicinal
426 value, and as a constituent in food, cosmetics and in other fields. To meet the market
427 demand, a variety of methods have been used to acquire theanine, with the main
428 methods including direct extraction, chemical synthesis, biotransformation (microbial
429 fermentation) and plant cell culture. Based on this study, we can use mutation
430 technology to modify engineering bacteria and improve the possibility of synthesis of
431 theanine in bacteria, yeast, model plants, etc, promoting its use for industrial theanine

432 production.

433

434 **Conclusions**

435 In summary, we determined the crystal structures of the CsAlaDC-EA complex and
436 AtSerDC. We employed structural biology to reveal the substrate selection mechanism
437 of these two decarboxylases. By structural analysis and enzyme activity experiments,
438 we found that Tyr^{111} of AtSerDC and Phe^{106} of CsAlaDC as crucial sites for substrate
439 specificity. Next, through mutation screening, we identified a CsAlaDC mutant protein
440 (L110F/P114A) with catalytic activity 2.3 times higher than that of the wild-type
441 protein. In an *in vitro* L-theanine biosynthesis system with added PsGS or GMAS, we
442 demonstrated that the L-theanine yield of the $CsAlaDC^{L110F/P114A}$ group was 4.3 and 6.7
443 times higher than that of the wild-type control group, respectively. In addition, by
444 utilizing the substrate recognition key motif of CsAlaDC and AtSerDC, evolutionary
445 analysis was conducted on protein sequences exhibiting high homologous to CsAlaDC
446 in *Embryophyta*, which resulted in the discovery of 13 potential alanine decarboxylases.
447 Furthermore, we discovered serine protease-like proteins containing special motifs that
448 have evolved in *Fabale*, which may possess unique substrate specificity and catalytic
449 functions. On the other hand, our study has effectively enhanced the efficiency of the
450 initial stage of theanine synthesis and is anticipated to offer a promising avenue for
451 developing a novel approach to producing L-theanine, thereby facilitating the progress
452 of the tea industry.

453

454 **Materials and Methods**

455 **Plants materials.**

456 Tobacco (*Nicotiana benthamiana*) plants were grown in a controlled chamber, under a
457 16-h light and 8-h dark photoperiod at 25 °C. Leaves of 5-week-old tobacco plants were
458 used for transient transformation, mediated by *Agrobacterium tumefaciens* strain
459 GV3101.

460

461 **Gene cloning and protein expression**

462 The coding sequences of AtSerDC and CsAlaDC, amino acid residues 66-482 and 61-
463 478, respectively, were amplified using PCR and then ligated into the *Nde* I and *Xho* I
464 restriction sites of the pET-28a and pET-22b vector, respectively, providing the
465 recombinant vector pET-28a-*AtSerDC* and pET-22b-*CsAlaDC*. The cDNAs encoding
466 CsSerDC were amplified using PCR and then ligated into the *Nde* I and *Xho* I restriction
467 sites of the pET-28a vector. Gene-specific primers were listed in Supplementary table
468 2. The recombinant plasmid was transformed into *E. coli* BL21 (DE3) competent cells.
469 Positive transformants were grown in a 5 mL LB medium containing 30 μg/mL
470 kanamycin or 50 μg/mL ampicillin at 37 °C overnight and then subcultured into an 800
471 mL LB medium containing the corresponding antibiotic. Protein expression was
472 induced by the addition of 0.2 mM isopropyl-β-d-thiogalactoside (IPTG) for 20 h at
473 16 °C when the optical density (OD) at 600 nm reached 0.6-0.8, harvested by

474 centrifugation at 4 °C and 4,000 rpm for 30 min.

475

476 **Protein production and crystallization**

477 The cell pellet was suspended in 30 mL of lysis buffer (20 mM HEPES, pH 7.5, 200
478 mM NaCl, 0.1 mM PLP), disrupted by High-Pressure Homogenizer, and then
479 centrifuged at 16,000 rpm for 30 min at 4 ℃ to remove the cell debris. The supernatants
480 were purified with a Ni-Agarose resin column followed by size-exclusion
481 chromatography. Before crystallization, purified proteins were concentrated at 10
482 mg/mL. Crystallization conditions were screened by the sitting-drop vapor diffusion
483 method using the reservoir solutions supplied in commercially available screening kits
484 (Crystal Screen, Crystal Screen 2, PEGRx 1, 2, and SaltRx 1, 2). A droplet made by
485 mixing 1.0 μL of purified AtSerDC or CsAlaDC (10 mg/mL) with an equal volume of
486 a reservoir solution was equilibrated against 100 μL of the reservoir solution at 16℃.
487 The crystal of AtSerDC was obtained using buffer pH 8.0 containing 20% (w/v)
488 PEG400 as a precipitate and 0.2 M $CaCl_2$. The crystal of the CsAlaDC-EA complex
489 was obtained at pH 7.5 containing 2.6 M sodium acetate. The crystal of CsAlaDC was
490 obtained at pH 6.0 containing 3.5 M sodium formate.

491

492 **Data collection and processing**

493 Crystals were grown by sitting-drop vapor diffusion method at 16 °C. The volume of
494 the reservoir solution was 100 μL and the drop volume was 2 μL, containing 1 μL of
495 protein sample and 1 μL of reservoir solution. The reservoir solution of AtSerDC

496 contained 0.1 M Tris-HCl (pH 8.0), 0.2 M $CaCl_2$, and 20% PEG 400. The reservoir
497 solution of CsAlaDC contained 0.1 M HEPES (pH 7.5), 2.6 M sodium acetate or 0.1 M
498 Bis-Tris (pH 6.0), and 3.5 M sodium formate. Crystals grew in 3-5 days using a protein
499 concentration of 10 mg/mL. Diffraction data were collected at Shanghai Synchrotron
500 Radiation Facility (China). The collected data sets were indexed, integrated, and scaled
501 using the HKL3000 software package. The structure of AtSerDC was solved by
502 molecular replacement using the structure of HisDC (PDB code: 7ERV
503 [*Photobacterium phosphoreum*]) as the model, and utilizing AtSerDC as a molecular
504 displacement model for both CsAlaDC and the CsAlaDC-EA complex. The AtSerDC,
505 CsAlaDC, and CsAlaDC-EA complex exhibit resolutions of 2.85 Å, 2.50 Å, and 2.60
506 Å, respectively. The statistics for data collection and processing are summarized in
507 Supplementary table 1.
508
509 **Enzyme activity assays**
510 Decarboxylase activity was measured by detecting products (EA or ethanolamine) in
511 Waters Acquit ultraperformance liquid chromatography (UPLC) system (20, 39). The
512 100 μL reaction mixture, containing 20 mM substrate (Ala or Ser), 100 mM potassium
513 phosphate, 0.1 mM PLP, and 0.025 mM purified enzyme, was prepared and incubated
514 at standard conditions (45 °C and pH 8.0 for CsAlaDC, 40 °C and pH 8.0 for AtSerDC
515 for 30 min). Then, the reaction was stopped with 20 μL of 10% trichloroacetic acid.
516 The detection methodology for theanine production remains consistent with the
517 aforementioned approach. The 100 μL reaction mixture, containing 20 mM Ala, 45 mM

518 Glu, 50 mM HEPES (pH 7.5), 0.1 mM PLP, 30 mM $MgCl_2$, 10 mM ATP, 0.03 mM
519 PsGS/GMAS, and 0.025 mM CsAlaDC/CsAlaDC ^{L110F}/CsAlaDC $^{L110F/P114A}$, was
520 prepared and incubated at standard conditions for 1 h. Subsequently, the reaction was
521 terminated via immersion of the reaction vessel in a metal bath at 96 °C for 3 min.
522 The product was derivatized with 6-aminoquinolyl-N-hydroxy-succinimidyl
523 carbamate (AQC) and subjected to analysis by UPLC. All enzymatic assays were
524 performed in triplicate.
525
526 **Site-directed mutagenesis**
527 Site-directed mutagenesis experiment was conducted using a PCR method from the
528 wild-type construct pET-28a-*AtSerDC* and pET-22b-*CsAlaDC*, respectively. *Dpn* I
529 endonucleases were used to digest the parental DNA template. The reaction mixture
530 was used to transform *E. coli* DH5α competent cells and the plasmids from positive
531 strains were extracted to *E. coli* BL21 (DE3) for protein expression, purification, and
532 analyzing the enzymatic activity.
533
534 ***In vivo* enzyme activity assay in *N. benthamiana***
535 The amplified PCR products were fused to the plant expression vector,
536 pCAMBIA1305. Linearization was conducted by restriction digest with *Spe* I and
537 *Bam*H I. The recombinant colonies were selected for PCR validation on the appropriate
538 antibiotics plate. After validation, the plasmids were electroporated into *Agrobacterium*
539 *tumefaciens* strain GV3101(pSoup-p19). Empty pCAMBIA1305 vector, with an intron

540 containing the GFP gene, was treated as the control.

541 *Agrobacterium* transient expression assays were performed on 5-week-old *N.*
542 *benthamiana* plants. *Agrobacterium tumefaciens* strain GV3101 (pSoup-p19), carrying
543 the above-described vectors, were cultured in Luria Bertani (LB) medium, containing
544 appropriate antibiotics, at 28 °C. When the absorbance of bacteria colonies reached
545 OD_{600} = 0.6-0.8, bacterial cells were collected and resuspended in MMA solution (10
546 mM $MgCl_2$, 10 mM 2-(N-morpholino) ethane sulfonic acid (MES), pH 5.6). After the
547 OD_{600} of the resuspended bacterial solution was adjusted to approx. 1.0, acetosyringone
548 (AS) was added with a final concentration of 200 μM, and this solution was then
549 incubated, at room temperature, for at least 3 h in darkness. Next, cell suspensions were
550 infiltrated into *N. benthamiana* leaves with a needle-free syringe. These *N.*
551 *benthamiana* leaves were then collected 3 days post infiltration, frozen in liquid
552 nitrogen and stored at -80 °C. Internal EA in *N. benthamiana* leaves was extracted as
553 previously described (22), and then the solvent was subjected to gas chromatography-
554 mass spectrometry GS-MS system.

555

556 **Transcript level analysis in *N. benthamiana***

557 Total RNA was isolated from samples using the RNAprep Pure Plant Kit (Tiangen,
558 Beijing, China), according to the manufacturer's protocol. The cDNAs were
559 synthesized using TransScript One-Step gDNA Removal and cDNA Synthesis
560 SuperMix Kit (TransGen Biotech, Beijing, China). The qRT-PCR assays were
561 performed, as previously described (40). Primers used for qRT-PCR assays were listed

562 in Supplementary Table 3 and the qRT-PCR was run on a Bio-Rad CFX96TM RT PCR
563 detection system and CFX Manager Software. Each reaction reagent (20 μL) contained
564 0.4 μL forward and reverse primers (10 μM), 2 μL cDNA (200±5 ng/μL), 10 μL SYBR
565 Green Supermix (Vazyme, Nanjing, China) and 6.2 μL double-distilled water. Reaction
566 was performed by a two-step method: 95 ℃ for 5 min; 40 cycles of 95 ℃ for 10 s; and
567 60 ℃ for 30 s. The glyceraldehyde-3-phosphate dehydrogenase (GAPDH) gene was
568 used for internal normalization in each RT-PCR, and the $2^{-\Delta\Delta CT}$ method (41) was used
569 to calculate the relative gene expression. All samples were performed with three
570 replicates.
571
572 **Multiple sequence alignment**
573 MUSCLE was used to generate the protein multiple sequence alignment of amino acid
574 decarboxylases with default settings (42). ESPript 3.x was used to display the multiple
575 sequence alignment (43).
576
577 **Phylogenetic analysis**
578 The CsAlaDC was used as a query for searching its homologs in *Embryophyta* using
579 BLASTP (44) (version 2.13.0). To fetch homologs of CsAlaDC in distant species, we
580 performed a two-round blast. We first search for the best hit in the given order under
581 *Embryophyta*. Then we blast each species under a provided order using the order best
582 hit as a query and as many as 10 hits were kept. The sequences with e values <= 0.001
583 were removed. We further dropped sequences with lengths less than 200 or larger than

800 amino acids. Then multiple sequence alignment was performed via clustalo (version 1.2.4) with default settings on these filtered sequences (45). The NJ algorithm implemented in clustalw (version 2.1) was adopted to build the phylogenetic tree (46). The R package ggtree (version 3.2.1) and universalmotif (version 1.12.4) were used for tree and motif visualization (47).

Acknowledgments

We thank the team of beamline BL18U1 in the Shanghai Synchrotron Radiation Facility for diffraction data collection.

Funding

This study is supported by grants from the Ministry of Science and Technology of China (2019YFA0904100 to W.G, 2022YFF1003103 to X.W) and the Natural Science Foundation of China (T2221005 to W.G, 32072624 to Z.Z).

Data availability

The structures of AtSerDC, CsAlaDC-EA complex and CsAlaDC have been deposited in the Protein Data Bank (PDB) under accession numbers 8JG7, 8JIK and 8JIJ respectively.

References

1. Drew, L. (2019) The growth of tea. *Nature* **566**, S2-S4.
2. Wheeler, D. S., Wheeler, W. J. (2004) The medicinal chemistry of tea. *Drug Dev. Res.* **61**,

45-65.

3. Xia, E. H. *et al*. (2017) The tea tree genome provides insights into tea flavor and independent evolution of caffeine biosynthesis. *Mol. Plant* **10**, 866-877.
4. Yang, C. S., Hong, J. G. (2013) Prevention of Chronic Diseases by Tea: Possible Mechanisms and Human Relevanc. *Annu. Rev. Nut.* **33**, 161-181.
5. Xia, E. *et al*. (2020) The reference genome of tea plant and resequencing of 81 diverse accessions provide insights into its genome evolution and adaptation. *Mol. Plant* **13**, 1013-1026.
6. Juneja, L. R., Chu, D. C., Okubo, T., Nagato, Y., Yokogoshi, H. (1999) L-theanine - a unique amino acid of green tea and its relaxation effect in humans. *Trends Food Sci. Technol.* **10**, 425-425.
7. Lin, J. Z. *et al*. (2022) Effect of the presence of stem on quality of oolong tea. *Foods* **11**, 3439.
8. Haskell, C. F., Kennedy, D. O., Milne, A. L., Wesnes, K. A., Scholey, A. B. (2008) The effects of L-theanine, caffeine and their combination on cognition and mood. *Biol. Psychol.* **77**, 113-122.
9. Kakuda, T. *et al*. (2000) Protective effect of gamma-glutamylethylamide (theanine) on ischemic delayed neuronal death in gerbils. *Neuro. Sci. Lett.* **289**, 189-192.
10. Liu, Q., Duan, H., Luan, J., Yagasaki, K., Zhang, G. (2009) Effects of theanine on growth of human lung cancer and leukemia cells as well as migration and invasion of human lung cancer cells. *Cytotechnology* **59**, 211-217.
11. Lu, K. *et al*. (2004) The acute effects of L-theanine in comparison with alprazolam on anticipatory anxiety in humans. *Hum. Psychopharmacol.* **19**, 457-465.
12. Takagi, Y. *et al*. (2010) Combined administration of (L)-cystine and (L)-theanine enhances immune functions and protects against influenza virus infection in aged mice. *J. Vet. Med. Sci.* **72**, 157-165.
13. Zheng, G., Sayama, K., Okubo, T., Juneja, L. R., Oguni, I. (2004)Anti-obesity effects of three major components of green tea, catechins, caffeine and theanine, in mice. *In Vivo* **18**, 55-62.

14. Lin, S. *et al*. (2023) Theanine metabolism and transport in tea plants (*Camellia sinensis* L.): advances and perspectives. *Crit. Rev. Biotechnol*. **43**, 327-341.
15. Deng, W., Ogita, S., Ashihara, H. (2008) Biosynthesis of theanine (γ-ethylamino-l-glutamic acid) in seedlings of *Camellia sinensis*. *Phytochem. Lett.* **1**, 115-119.
16. Deng, W.-W., Ogita, S., Ashihara, H. (2009) Ethylamine content and theanine biosynthesis in different organs of camellia sinensis seedlings. *Z. Naturforsch. C* **64**, 906-906.
17. Cheng, S. *et al*. (2019) Differential accumulation of specialized metabolite L-theanine in green and albino-induced yellow tea (*Camellia sinensis*) leaves. *Food Chem.* 276, 93-100.
18. Takeo, T. (1974) L-alanine as a precursor of ethylamine in *Camellia sinensis*. *Phytochem.* **13**, 1401-1406.
19. Zhu, B. *et al*. (2021) CsAlaDC and CsTSI work coordinately to determine theanine biosynthesis in tea plants (*Camellia sinensis* L.) and confer high levels of theanine accumulation in a non-tea plant. *Plant Biotechnol. J.* **19**, 2395-2397.
20. Bai, P. *et al*. (2021) Biochemical characterization of specific Alanine Decarboxylase (ADC) and its ancestral enzyme Serine Decarboxylase (SDC) in tea plants (*Camellia sinensis*). *BMC Biotechnol.* **21**,17.
21. Han, S. W., Shin, J. S. (2022) Aromatic L-amino acid decarboxylases: mechanistic features and microbial applications. *Appl. Microbiol. Biotechnol.* **106**, 4445-4458.
22. Berger, M., Gray, J. A., Roth, B. L. (2009) The expanded biology of serotonin. *Annu. Rev. Med.* **60**, 355-366.
23. Khan, M. Z., Nawaz, W. (2016) The emerging roles of human trace amines and human trace amine-associated receptors (hTAARs) in central nervous system. *Biomed. Pharmacother*. **83**, 439-449.
24. Ng, J., Papandreou, A., Heales, S. J., Kurian, M. A. (2015) Monoamine neurotransmitter disorders--clinical advances and future perspectives. *Nat. Rev. Neurol.* **11**, 567-584.
25. Rontein, D. *et al*. (2001) Plants synthesize ethanolamine by direct decarboxylation of serine using a pyridoxal phosphate enzyme. *J. Biol. Chem.* **276**, 35523-35529.
26. Gibellini, F., Smith, T. K. (2010) The Kennedy pathway-De novo synthesis of phosphatidylethanolamine and phosphatidylcholine. *IUBMB Life* **62**, 414-428.

27. Mudd, S. H., Datko, A. H. (1989) Synthesis of ethanolamine and its regulation in *Lemna paucicostata*. *Plant Physiol.* **91**, 587-597.

28. Zinser, E. *et al*. (1991) Phospholipid synthesis and lipid composition of subcellular membranes in the unicellular eukaryote *Saccharomyces cerevisiae*. *J. Bacteriol.* **173**, 2026-2034.

29. Yunus, I. S., Liu, Y.-c., Nakamura, Y. (2016) The importance of SERINE DECARBOXYLASE1 (SDC1) and ethanolamine biosynthesis during embryogenesis of *Arabidopsis thaliana*. *Plant J.* **88**, 559-569.

30. Paiardini, A., Giardina, G., Rossignoli, G., Voltattorni, C. B., Bertoldi, M. (2017) New insights emerging from recent investigations on human group II pyridoxal 5'-phosphate decarboxylases. *Curr. Med. Chem.* **24**, 226-244.

31. Torrens-Spence, M. P. *et al*. (2020) Structural basis for divergent and convergent evolution of catalytic machineries in plant aromatic amino acid decarboxylase proteins. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 10806-10817.

32. Hagihara, R., Ohno, S., Hayashi, M., Tabata, K., Endo, H. (2021) Production of L-theanine by *Escherichia coli* in the absence of supplemental ethylamine. *Appl. Environ. Microbiol.* **87**, e00031-21.

33. Yamamoto, S., Wakayama, M. & Tachiki, T. (2007) Characterization of theanine-forming enzyme from *Methylovorus mays* no. 9 in respect to utilization of theanine production. *Biosci. Biotechnol. Biochem.* **71**, 545-552.

34. Holm, L. (2022) Dali server: structural unification of protein families. *Nucleic. Acids Res.* **50**, W210-W215.

35. Okawa, A. *et al*. (2021) Structural basis for substrate specificity of L-methionine decarboxylase. *Protein Sci.* **30**, 663-677.

36. Fenalti, G. *et al*. (2007) GABA production by glutamic acid decarboxylase is regulated by a dynamic catalytic loop. *Nat. Struct. Mol. Biol.* **14**, 280-286.

37. Gayathri, S. C., Manoj, N. (2019) Structural insights into the mechanism of internal aldimine formation and catalytic loop dynamics in an archaeal Group II decarboxylase. *J. Struct. Biol.* **208**, 137-151.

38. Torrens-Spence, M. P., von Guggenberg, R., Lazear, M., Ding, H., Li, J. (2014) Diverse functional evolution of serine decarboxylases: identification of two novel acetaldehyde synthases that uses hydrophobic amino acids as substrates. *BMC Plant Biol.* **14**, 247.
39. Qu, L., Gu, S., Zhang, J., Zhao, C., Deng, X. (2021) Determination of 18 amino acids in three different kinds of milk powder by ultra performance liquid chromatography coupled with pre-column derivatization. *Chin. J. Chromatogr.* **39**, 472-477.
40. Madeira, F. *et al*. (2022) Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.* **50**, W276-W279.
41. Dong C. *et al*. (2020) Theanine transporters identified in tea plants (*Camellia sinensis* L.). *Plant J.* **101**, 57-70.
42. Schmittgen T. D. *et al*. (2000) Quantitative reverse transcription-polymerase chain reaction to study mRNA decay: comparison of endpoint and real-time methods. *Anal. Biochem.* **285**, 194–204.
43. Robert X., Gouet P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320-W324.
44. Camacho, C. *et al*. (2009) BLAST+: architecture and applications. *BMC Bioinf.* **10**, 421.
45. Sievers, F., Higgins, D. G. (2018) Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci.* **27**, 135-145.
46. Larkin, M. A. *et al*. (2007) Clustal W and clustal X version 2.0. *Bioinformatics.* **23**, 2947-2948.
47. Yu, G., Lam, T. T.-Y., Zhu, H., Guan, Y. (2018) Two methods for mapping and visualizing associated data on phylogeny using ggtree. *Mol. Biol. Evol.* **35**, 3041-3043.

Figures and table

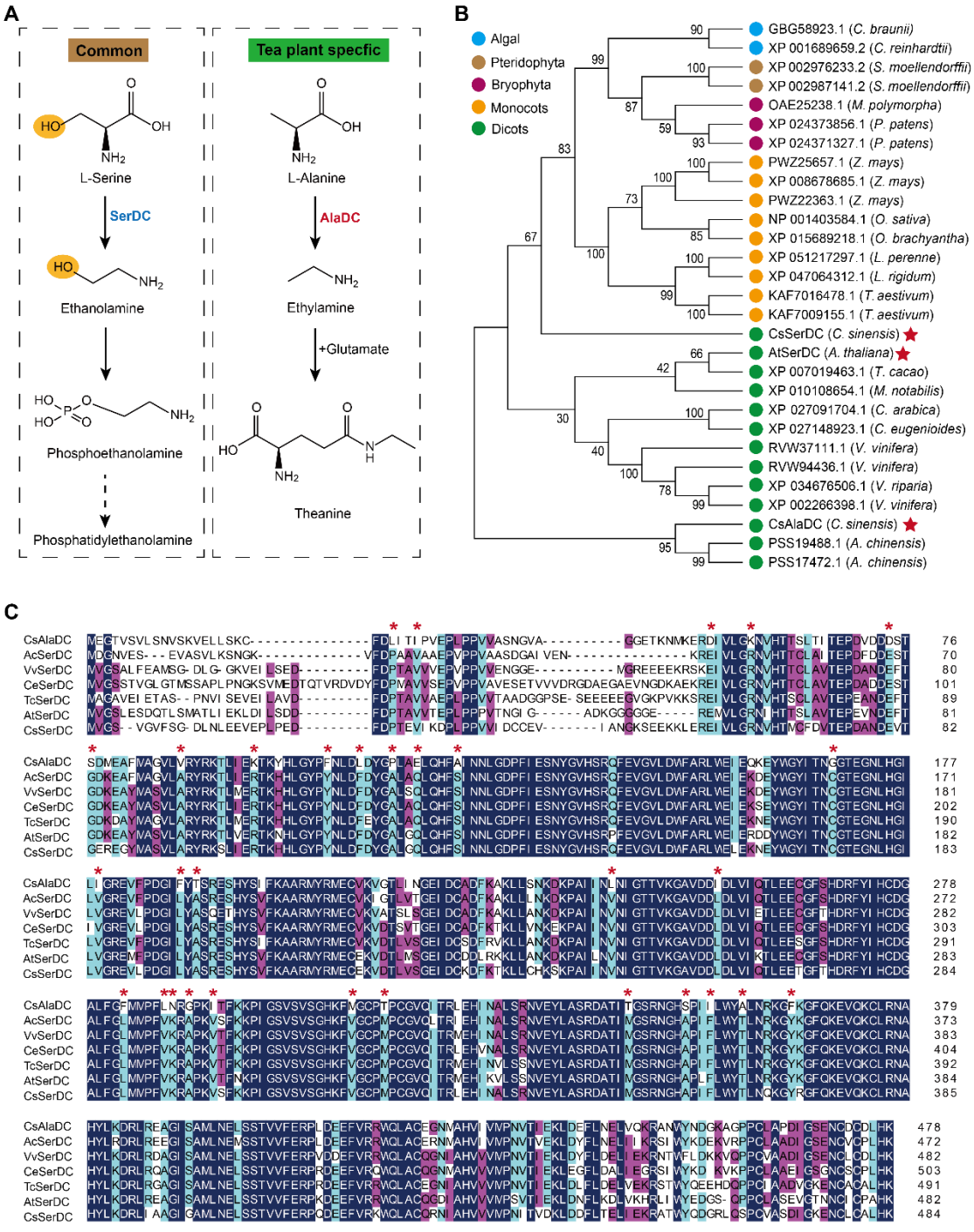


Figure 1. Metabolic pathways and phylogeny analysis of SerDC and AlaDC in plants. (**A**) The decarboxylation of serine and alanine in plants. SerDC, serine decarboxylase; AlaDC, alanine decarboxylase. (**B**) Phylogenetic analysis of the SerDC homologous proteins in Algal, Pteridohyta, Bryophyta, Dicots and Monocots. The

729 phylogenetic tree was performed using the neighbor-joining method in MAGA 7.0
730 software and bootstrap values based on 1000 replicates were displayed at branch points.
731 These proteins were represented by their GenBank accession number. (**C**) Multiple
732 alignment of the amino acid sequences of AlaDC and SerDCs. Primary (100%),
733 secondary (80%), and tertiary (60%) conserved percent of similar amino acid residues
734 were shaded in deep blue, light blue and cheer red, respectively. The multiple alignment
735 was constructed by MAGA 7.0 software by ClustalW method, and GeneDOC software
736 was used to visualize multiple sequence alignments. "*" indicated amino acid residues
737 mutated only in CsAlaDC.
738
739
740

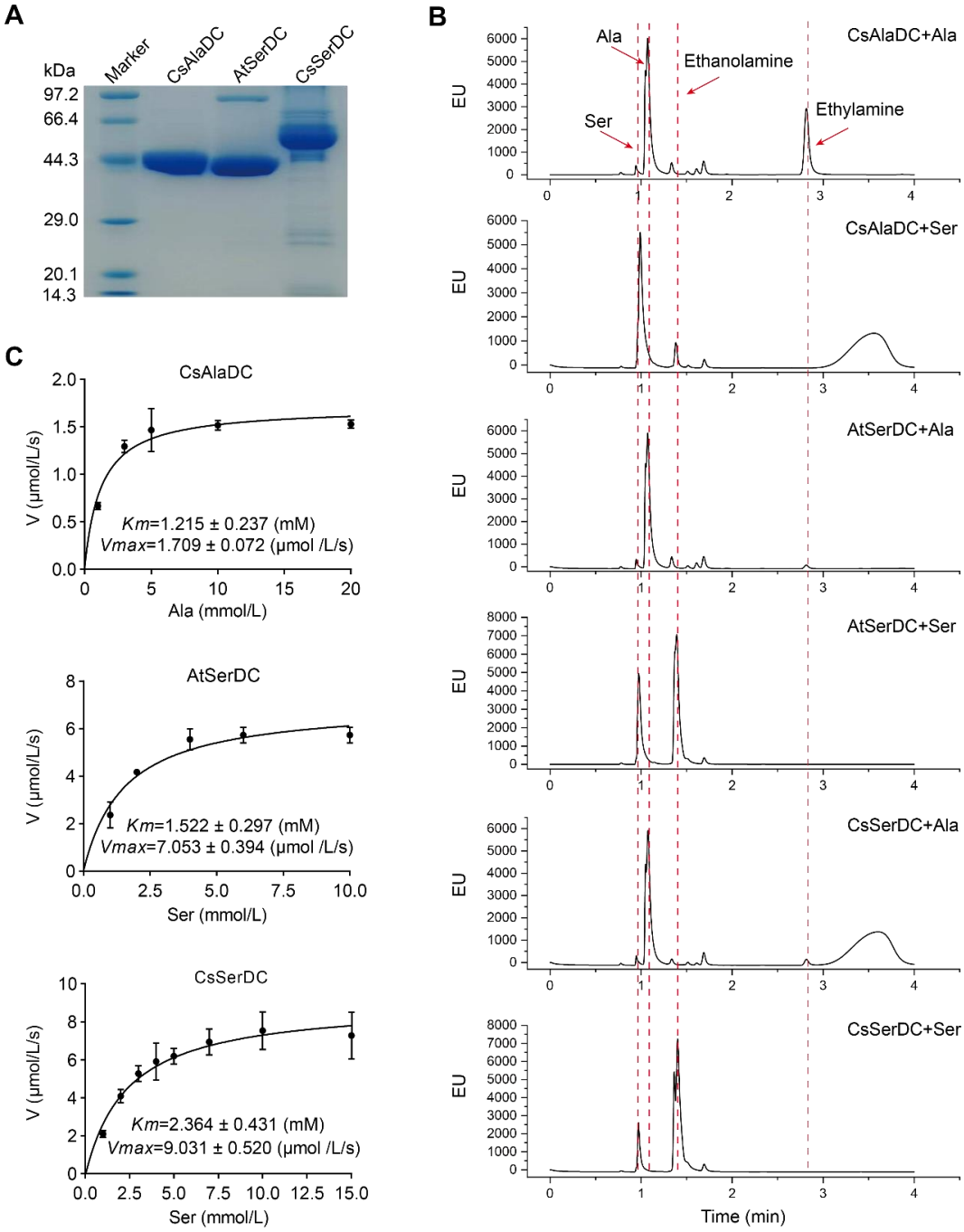


Figure 2. Purification and characterization of CsAlaDC, AtSerDC, and CsSerDC. (**A**) Identification of CsAlaDC, AtSerDC, and CsSerDC using SDS-PAGE. (**B**) Detection of enzyme activities of CsAlaDC, AtSerDC, and CsSerDC by UPLC. (**C**) Reaction rates of substrates with different concentrations catalyzed by CsAlaDC, AtSerDC, and CsSerDC.

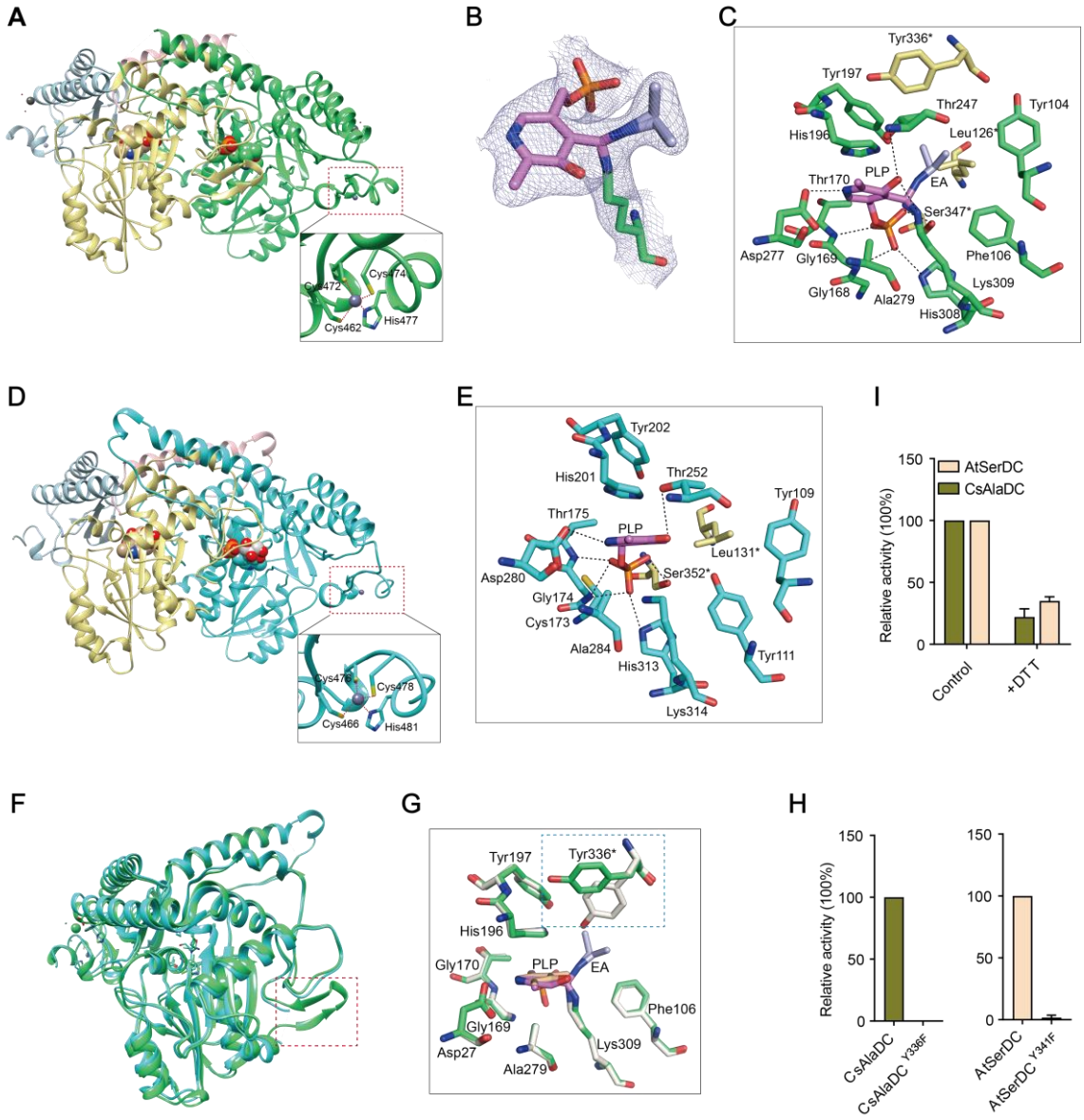


Figure 3. Crystal structures of CsAlaDC and AtSerDC. (**A**) Dimer structure of CsAlaDC. The color display of the N-terminal domain, large domain, and C-terminal domains of chain A is shown in light pink, khaki and sky blue, respectively. Chain B is shown in spring green. The PLP molecule is shown as a sphere model. The zinc finger structure at the C-terminus of CsAlaDC is indicated by the red box. The gray spheres represent zinc ions, while the red dotted line depicts the coordination bonds formed by zinc ions with cysteine and histidine. (**B**) The 2Fo-Fc electron density maps of K309-

757 PLP-EA (contoured at 1σ level). The PLP is shown in violet, the K309 is shown in
758 spring green, and the EA is shown in lightblue. (**C**) Active center of the CsAlaDC-EA
759 complex, with hydrogen bonds denoted by black dotted lines. "*" denotes the amino
760 acids on adjacent subunits. (**D**) Dimer structure of AtSerDC. The color display of the
761 N-terminal domain, large domain, and C-terminal domains of chain A is shown in light
762 pink, khaki and sky blue, respectively. Chain B is shown in cyan. The PLP molecule is
763 shown as a sphere model. The zinc finger structure at the C-terminus of AtSerDC is
764 indicated by the red box. The gray spheres represent zinc ions, while the red dotted line
765 depicts the coordination bonds formed by zinc ions with cysteine and histidine. (**E**)
766 Active center of the AtSerDC, with hydrogen bonds denoted by black dotted lines. "*"
767 denotes the amino acids on adjacent subunits. (**F**) The monomer of CsAlaDC and
768 AtSerDC are superimposed. CsAlaDC is depicted in spring green, while AtSerDC is
769 shown in cyan. The conserved amino acid catalytic ring is indicated by the red box (**G**)
770 Amino acid residues of the active center in CsAlaDC apo and CsAlaDC-EA complex
771 are superimposed. CsAlaDC apo is shown in floral white, while CsAlaDC-EA complex
772 is shown in spring green. (**H**) The relative activity of wild-type CsAlaDC and its Y336F
773 mutant (left), as well as wild-type AtSerDC and its Y341F mutant (right) is shown. (**I**)
774 The effects of DTT on the activity of CsAlaDC and AtSerDC are depicted. Three
775 independent experiments were conducted.
776

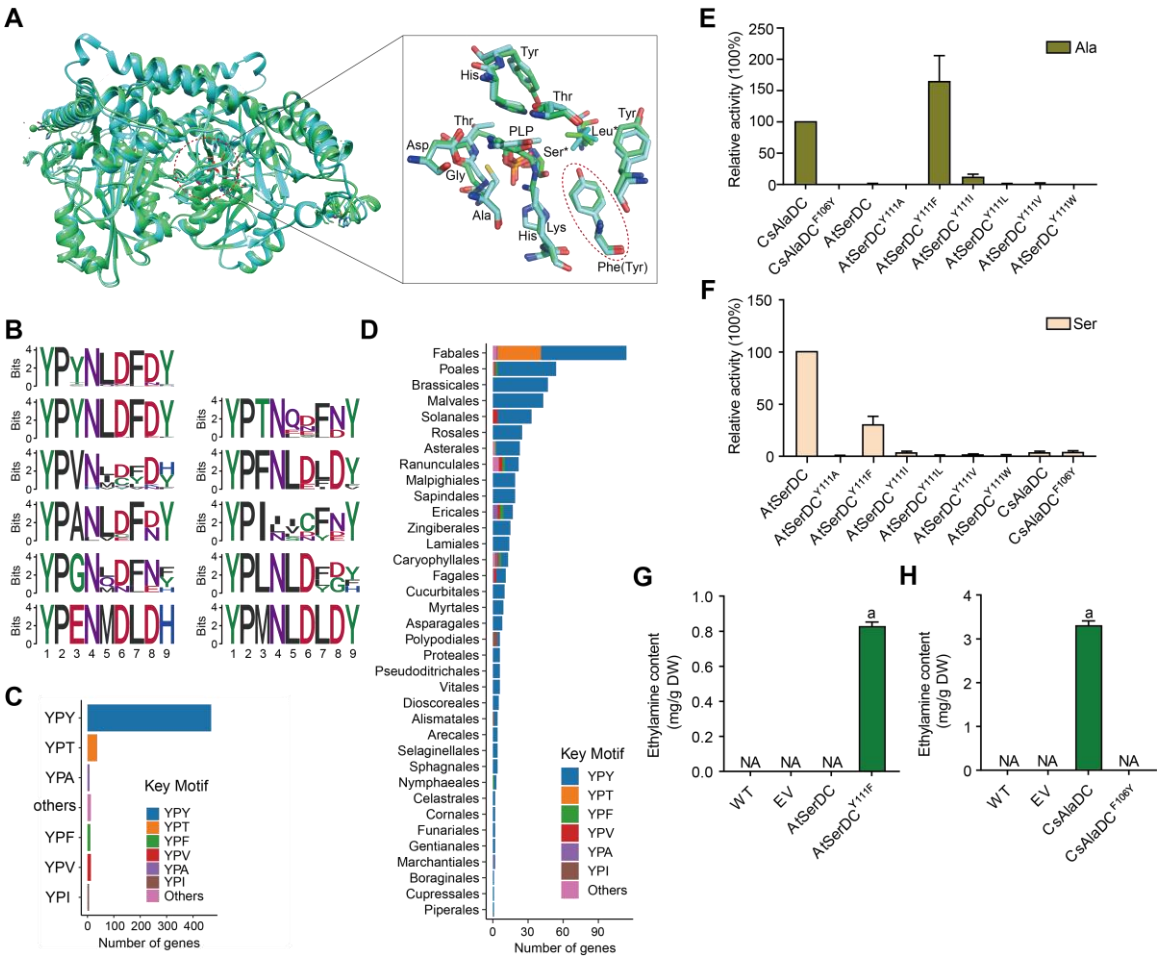


Figure 4. Key amino acid residues for substrate recognition. (**A**) Superposition of substrate binding pocket amino acid residues in CsAlaDC and AtSerDC. The amino acid residues of CsAlaDC are shown in spring green, the amino acid residues of AtSerDC are shown in cyan, with the substrate specificity-related amino acid residue highlighted in a red ellipse. (**B**) Active-site-lining residues from *Embryophyta* AlaDC were identified. The height of the residue label displays the relative amino acid frequency. (**C**) Histogram showing the distribution of the number of key motifs. (**D**) Histogram showing the number of key motifs in different plant orders. (**E**) Relative enzyme activities of wild-type CsAlaDC and mutant protein CsAlaDC^{F106Y} against Ala substrate (columns 1 and 2), and enzyme activities of wild-type AtSerDC and various

788 AtSerDC mutant proteins against Ala substrate (columns 3-9) are presented. The
789 percentage graph shows the relative activity of each protein compared to wild-type
790 CsAlaDC activity (taken as a 100% benchmark). (**F**) Relative enzyme activities of wild-
791 type AtSerDC and AtSerDC mutant proteins (columns 1-7) against Ser substrates, and
792 enzyme activities of wild-type CsAlaDC and mutant protein CsAlaDC^{F106Y} against Ser
793 substrates(columns 8, 9) were measured. The percentage graph shows the relative
794 activity of each protein compared to the wild-type AtSerDC (taken as a 100%
795 benchmark). Three independent experiments were conducted. (**G**) The EA contents of
796 AtSerDC and its mutant AtSerDC^{Y111F} in *N. benthamiana*. (**H**) The EA contents of
797 CsAlaDC and its mutant CsAlaDC^{F106Y} in *N. benthamiana*. The significance of the
798 difference (P<0.05) was labeled with different letters according to Duncan's multiple
799 range test.

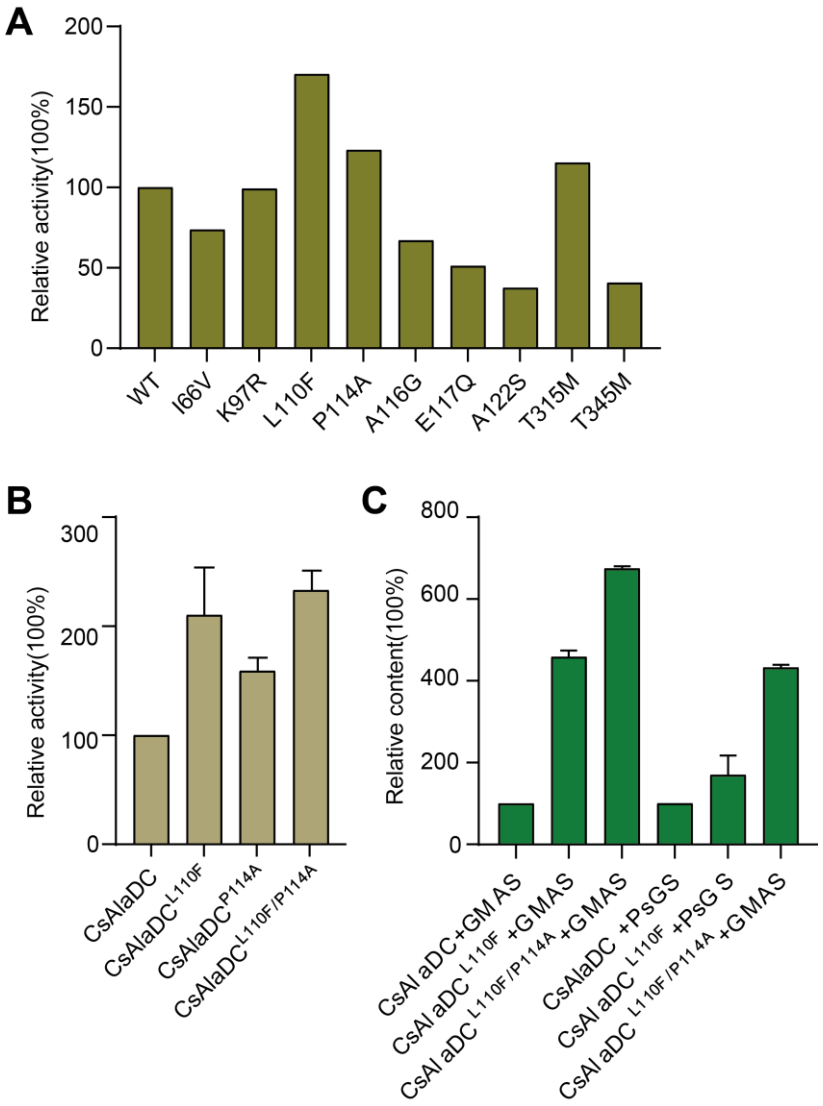


Figure 5. Mutations enhance CsAlaDC enzyme activity and theanine synthesis *in vitro*. (**A**) Relative enzyme activities of CsAlaDC mutant proteins against Ala substrate. (**B**) Relative enzyme activities of CsAlaDC^{L110F}, CsAlaDC^{P114A}, and CsAlaDC$^{L110F/P114A}$ against Ala substrate. (**C**) Histogram showing the relative content of theanine resulting from different combinations of alanine decarboxylase and theanine synthetase. Three independent experiments were conducted.

808 **Table 1.** Kinetic parameters of CsAlaDC, AtSerDC and CsSerDC

Enzyme	Substrate	*Km* (mmol/L)	*Vmax* (μmol/L/s)
CsAlaDC	Alanine	1.215 (0.802, 1.766)	1.709 (1.568, 1.865)
AtSerDC	Serine	1.522 (1.010, 2.234)	7.053 (6.311, 7.945)
CsSerDC	Serine	2.364 (1.634, 3.351)	9.031 (8.087, 10.160)

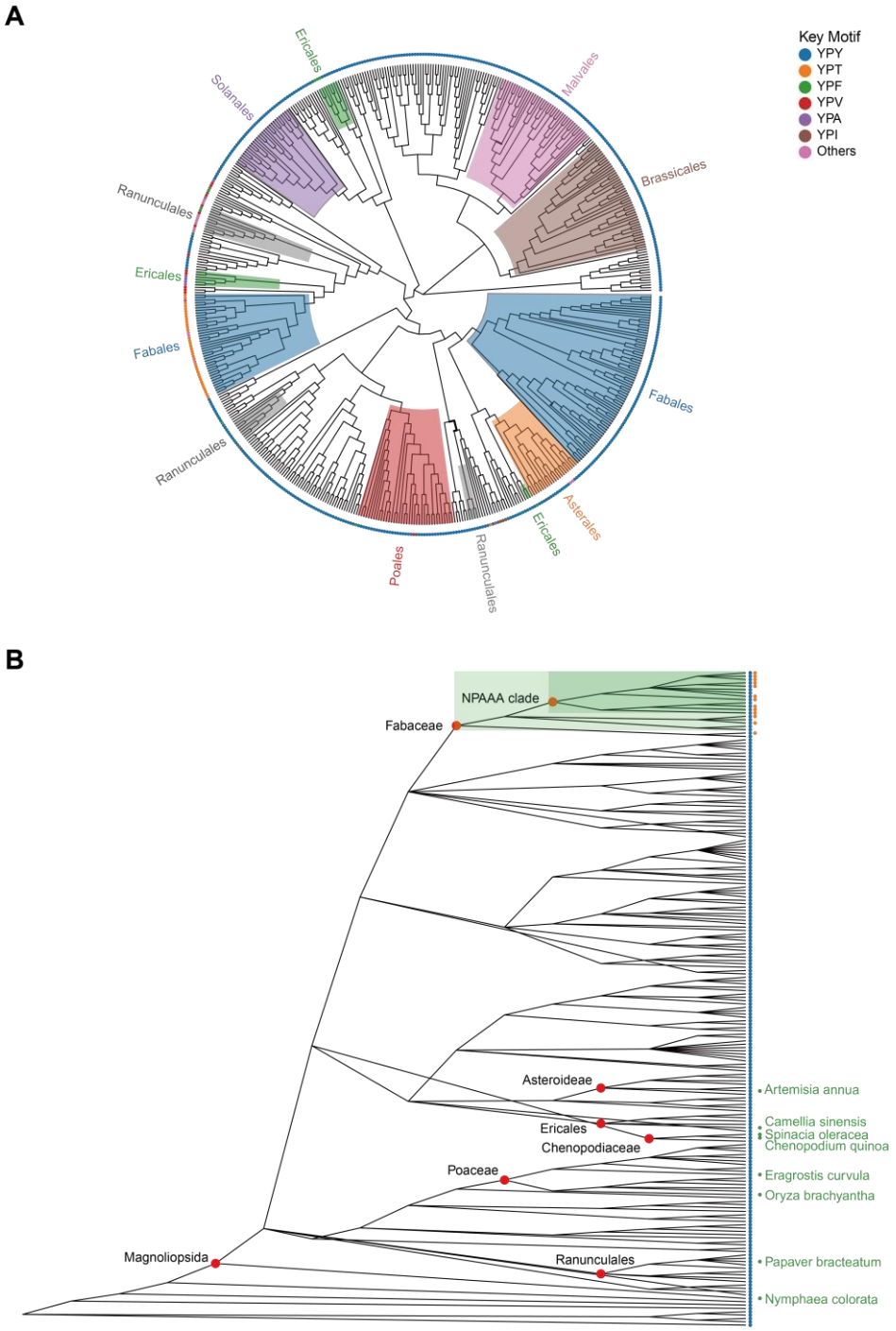
809

810 The values enclosed in parentheses denote the 95% confidence intervals for the
811 corresponding parameters.

812

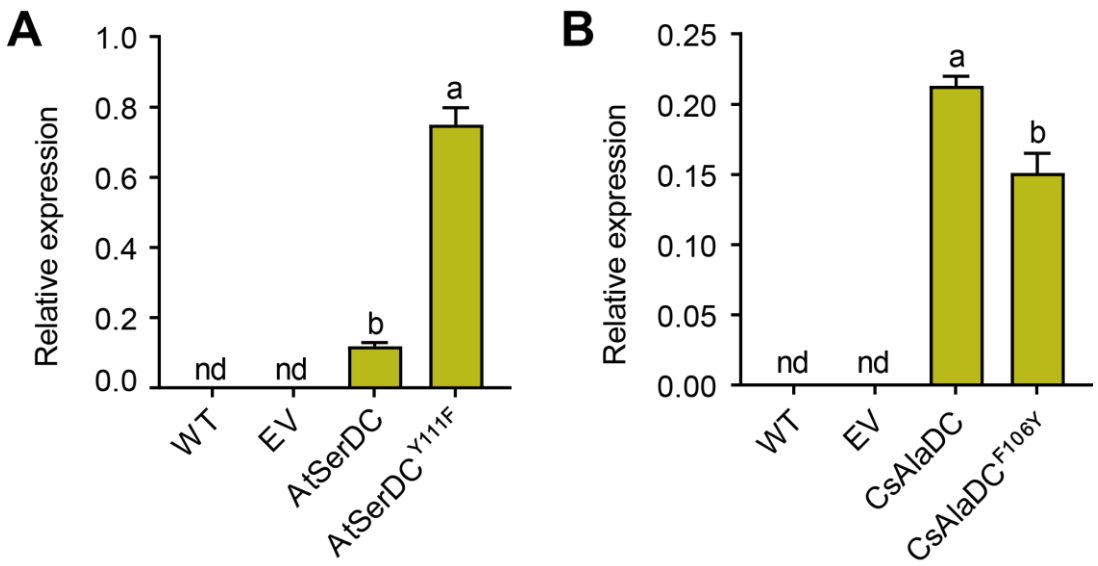
813

Supplementary figures

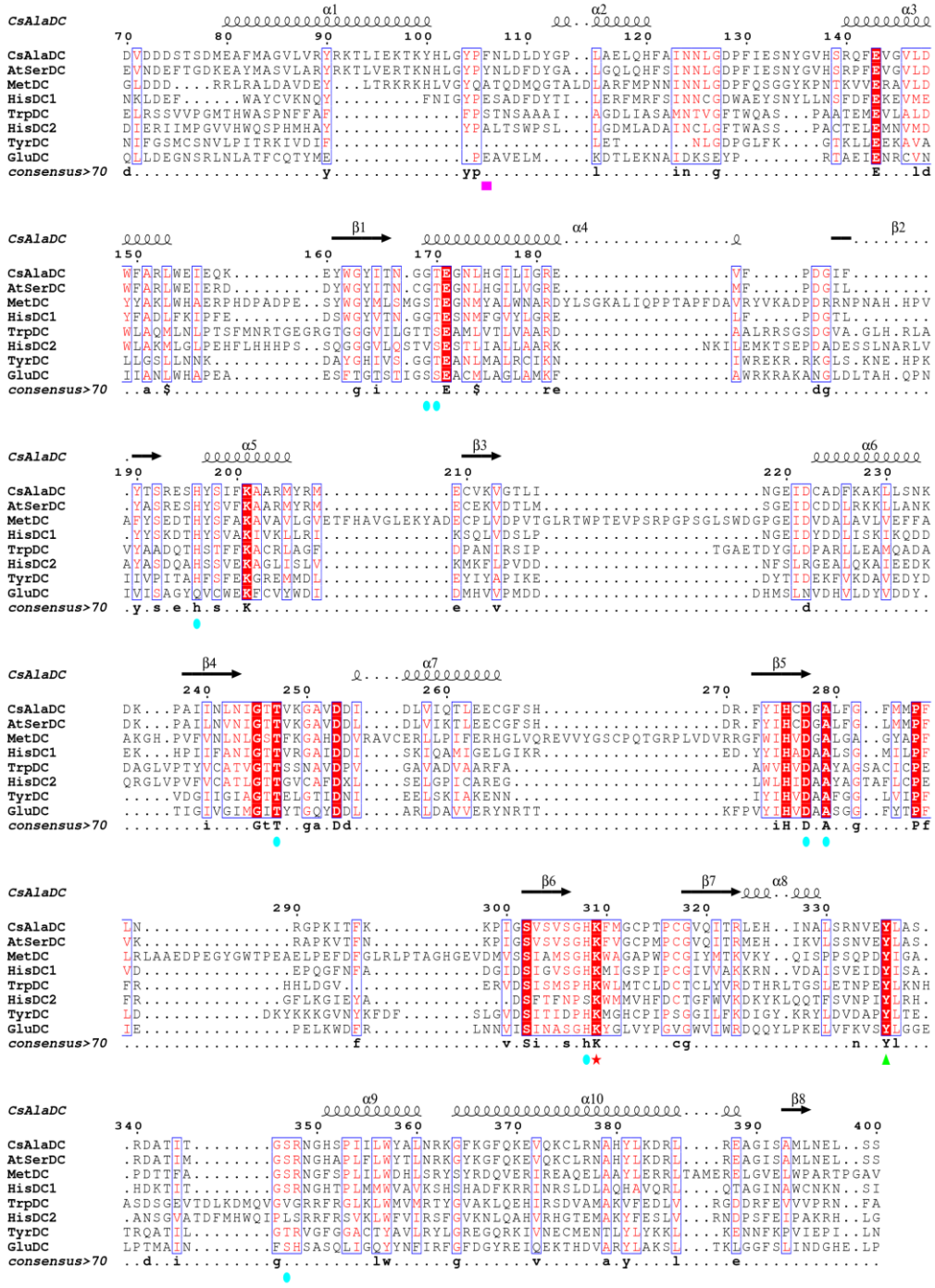


Supplementary figure 1. Evolutionary analysis of CsAlaDC in *Embryophyta*. (**A**) The presented diagram depicts an evolutionary tree of CsAlaDC, which is devoid of a root and solely portrays the topological structure of the tree without including distance

information. The color of the inner ring corresponds to various orders, while the outer ring's leaf nodes are colored based on the motif types that the sequence exhibits. (**B**) Diversity of serine decarboxylase-like proteins in *Embryophyta* (196 species). Colored scatter spots on the right side of leaf nodes correspond to the respective motifs shown in Figure A.



Supplementary figure 2. The relative mRNA levels of AtSerDC and its mutant Y111F (*A*), CsAlaDC and its mutant F106Y (*B*) in *N. benthamiana* leaves were measured by two primers. WT, wild type of *N. benthamiana*; EV, empty vector control; *NbGAPDH* was used as an internal control. Data represent mean ± SD (n=3). The significance of the difference (P<0.05) was labeled with different letters according to Duncan's multiple range test.

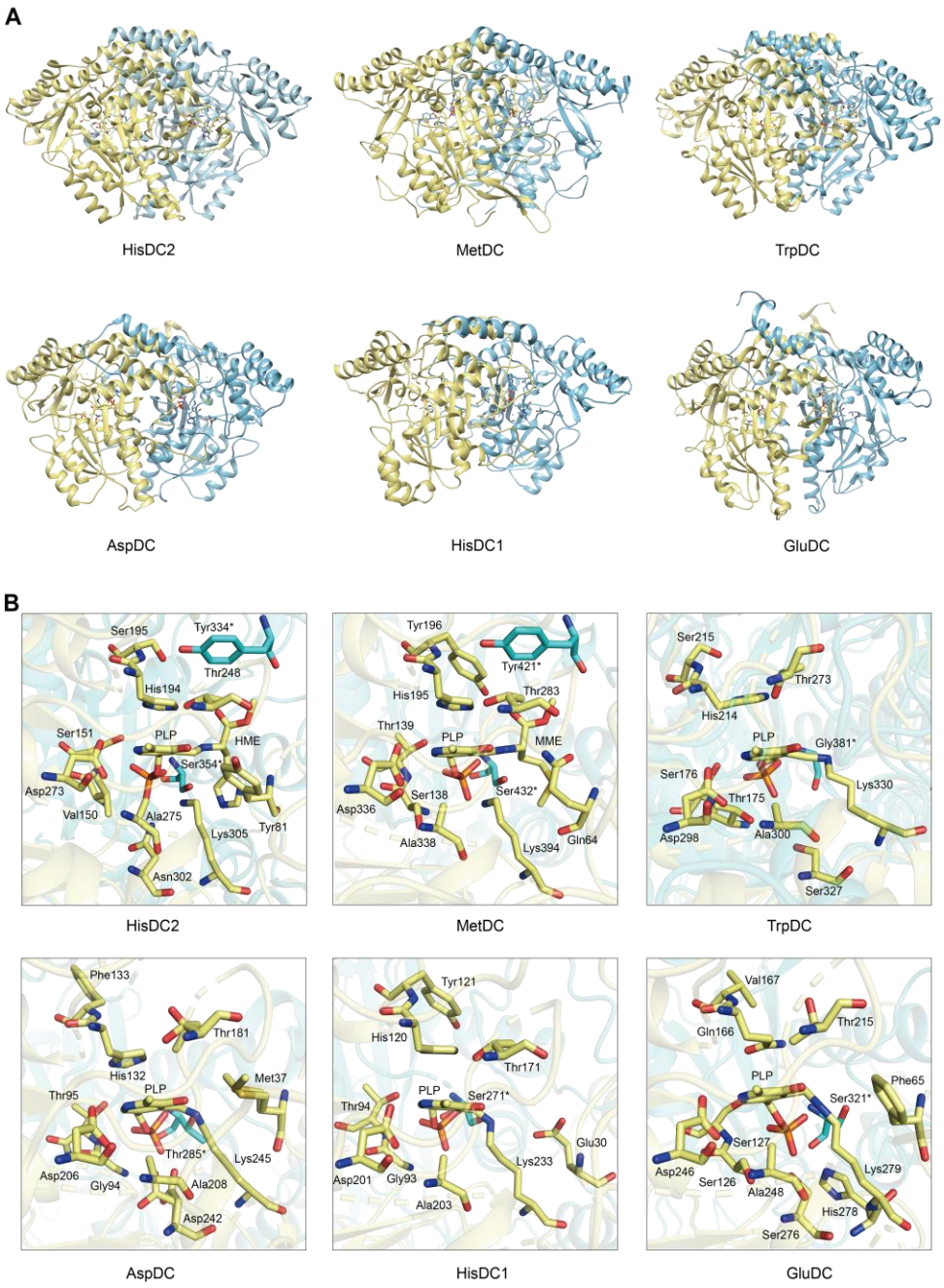


834

835 **Supplementary figure 3.** Multiple sequence alignment of CsAlaDC, AtSerDC, MetDC,

836 HisDC1, TrpDC, HisDC2, TyrDC, and GluDC were generated using MUSCLE and

837 visualized with ESPript 3.x. Conserved amino acid residues in all eight proteins are
838 highlighted with red backgrounds. The magenta box marks key amino acid residues
839 involved in substrate recognition for CsAlaDC and AtSerDC. The Lys residue
840 covalently bound to the PLP cofactor is denoted by a red star, while the green triangle
841 indicates the Tyr residue associated with enzymatic activity. Amino acid residues
842 involved in the CsAlaDC substrate binding pocket are marked with blue circles.
843



Supplementary figure 4. Structures of HisDC2, MetDC, TrpDC, AspDC, HisDC1 and GluDC. (**A**) The Overall Structures of HisDC2, MetDC, TryDC, AspDC, HisDC1 and GluDC. Chain A is shown in khaki, chain B is shown in cyan. (**B**) Amino acid residues in the substrate binding pocket of HisDC2, MetDC, TryDC, AspDC, HisDC1 and GluDC. The amino acid residues in chain A are shown in khaki, and the amino acid residues in chain B are shown in cyan.

851 **Supplementary tables**

852

853 **Supplementary table 1**. Data collection and refinement statistics

	CsAlaDC	CsAlaDC-EA complex	AtSerDC
Data Collection			
Space Group	*P3121*	*P3121*	*C2*
Cell Dimensions			
a, b, c (Å)	163.5,163.5, 50.2	164.0,164.0,50.2	115.4,149.5,139.1
α, β, γ (°)	α=β=90, γ=120	α=β=90, γ=120	α=γ=90, β=109.2
Wavelength (Å)	0.979	0.979	0.979
Resolution (Å)	50-2.50	50-2.60	50-2.85
Completeness (%)	99.3 (99.9)	99.6 (100.0)	99.6 (98.1)
Redundancy	5.4 (5.1)	13.0 (12.5)	5.9 (5.0)
$<I>/<\sigma I>$	9.3 (1.9)	18.9 (4.6)	10.6 (2)
R_{merge}	0.128 (0.603)	0.156 (0.562)	0.173 (0.551)
Refinement			
R_{work}/ R_{free}	0.173/0.222	0.172/0.223	0.214/0.262
No. reflections	25175	22813	49543
B-facator (Å^2)			
Protein	43.3	48.3	57.6
PLP	29.5	31.5	52.8
Zn^{2+}	58.7	43.8	71.6
Ca^{2+}	-	55.5	-
Ethylamine	-	27.5	-
Glycerol	-	-	51.7
Water	38.5	40.9	-
R.m.s. deviations			
Bond lengths (Å)	0.008	0.008	0.007
Bond angles (°)	1.531	1.601	1.410
No. of atoms			
Protein	3292	3265	3250
PLP	15	15	60
Zn^{2+}	2	1	4
Ca^{2+}	-	1	-
Ethylamine	-	5	-
Glycerol	-	-	24
Water	115	60	-
Ramachandran statistics (%)			
Favoured	94.01	93.58	91.52
Allowed	5.74	5.68	7.61
Outliers	0.25	0.74	0.87

854 R_{merge}=∑hkl∑i|Ii(hkl)-<I(hkl)>|/∑hkl∑i Ii(hkl), where<I(hkl)>is the main value of I(hkl).

855 R_{work}=∑||Fobs|-|Fcalc||/∑|Fobs|, where Fobs and Fcalc are observed and calculated structure

factors.

The free R factor was calculated using 5% of reflections ommitted from the refinement.

Highest resolution shell is shown in parenthesis

Supplementary table 2. Primers used for gene cloning.

Name	Forward primers (5'-3')	Reverse primers (5'-3')
CsAlaDC	GGAATTCCATATGACAACAAGCCTCACCATCACGG	GATCTCGAGTCATTTATGAAGATCACAATCACAATTCTCACTTC
AtSerDC	GGAATTCCATATGACCACCAGCCTGGCCG	GATCTCGAGTTATTTATGGGCCGGACAAATGCAATTATTG
CsSerDC	ATGGTGGGAAGTGTTGGGGT	TCACTTGTGTAGTGCACAAAGACAGT

Supplementary table 3. Primers used for real-time PCR.

Name	Forward primers (5'-3')	Reverse primers (5'-3')
CsAlaDC	CACTGTGATGGGGCTCTGTT	TGTTATCTGGACACCGCACG
AtSerDC	TCACTCGCTGTAACGGAACC	AACTGACCAAGCGCACCATA
NbGAPDH	AGCTCAAGGGAATTCTCGATG	AACCTTAACCATGTCATCTCCC