Tropane alkaloids and terpenes synthase genes of *Datura stramonium* (Solanaceae)

Sabina Velázquez-Márquez [1] , **Iván M De-la-Cruz** [1] , **Rosalinda Tapia-López** [1] , **Juan Núñez-Farfán** [Corresp. 1]

[1] Laboratorio de Genética Ecológica y Evolución, Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, Ciudad de México, D F, MEXICO

Corresponding Author: Juan Núñez-Farfán
Email address: farfan@unam.mx

Background. Plants have evolved physical-chemical defense to prevent/diminish damage by their enemies. Chemical defense involves the synthesis' pathways of specialized toxic, repellent, or anti-nutritive metabolites to herbivores. Molecular evolutionary studies have revealed the origin of new genes, acquisition and functional diversification along time in different plant lineages. Methods. Using bioinformatic tools we analyze gene divergence of tropane alkaloids (TAs) and terpene synthases (TPSs) in *Datura stramonium* and other species of Solanaceae; compared gene and amino acids sequence of TAs and TPSs on genomes, cDNA and proteins sequences of Viridiplantae. We analyzed two recently assembled genomes of *D. stramonium* (Ticumán and Teotihucán), transcriptomes of *Datura metel* and genomes of other Solanaceae. Hence, we analyzed variation of TAs and TPSs to infer genes involved in plant defense and plant responses before stress. We analyzed protein modeling and molecular docking to predict interactions between H6H and ligand; we translated the sequences (Teo19488, Tic8550 and Tic8549) obtained from the two genomes of *D. stramonium* by using Swiss-Model and Ramachandran plot and MolProbity structure validation of Teo19488 protein model. Results. For TAs, we detected an expansion event in the tropinone reductase II (TRII) and the ratio synonymous/non-synonymous substitutions indicate positive selection. In contrast, a contraction event and negative selection was detected in tropinone reductase I (TRI). In Hyoscyamine 6 b-hydroxylase (H6H), enzyme involved in the production of tropane alkaloids atropine and scopolamine, the synonymous/non-synonymous substitution ratio in its dominion indicates positive selection. For terpenes (TPS), we found 18 DsTPS in *D. stramomiun* and 7 in *D. metel*; evolutionary analyses detected positive selection in TPS10.1 and TPS10.2 of *D. stramonium* and *D. metel*. Comparison of copies of TPSs in *D. stramonium* detected variation among them in the binding site. Duplication events and differentiation of TAs and TPSs of *D. stramonium*, as compared to other Solanaceae, suggest their possible involvement on adaptive evolution of defense to herbivores. Protein modeling and docking

show that the three protein structures obtained of DsH6H from Teo19488, Tic-8550 and Tic8549 maintain the same interactions and the union site of 2OG-FeII_Oxy with the Hy-o ligand as in 6TTM of *D. metel.* Conclusion. Our results indicate differences in the number of gene copies involved in the synthesis of tropane alkaloids, between the genomes of *D. stramonium* from two Mexican populations. More copies of genes related to the synthesis of tropane alkaloids (TRI, TRII, H6H, PMT-10) are found in *D. stramonium*as compared to Viridiplantae. Likewise, for terpene synthases (TPS), TPS-10 is duplicated in *D. stramonium* and *D. metel*. Further studies should be directed to experimentally assess gain (overexpression) or loss (silencing) of function of duplicated genes.

Tropane alkaloids and terpenes synthase genes of *Datura stramonium* (Solanaceae)

3 Sabina Velázquez-Márquez[1], Iván. M. De-la-Cruz[1], Rosalinda Tapia-López[1] and Juan Núñez-

4 Farfán[1]*

6 [1]Laboratorio de Genética Ecológica y Evolución, Departamento de Ecología Evolutiva, Instituto

7 de Ecología, Universidad Nacional Autónoma de México, Ciudad de México, D F, MEXICO

9 *Corresponding Author:

10 Juan Núñez-Farfán[1]

14 Email address: farfan@unam.mx

16   ABSTRACT

18   Background. Plants have evolved physical and chemical defense to prevent/diminish damage by

19   their natural enemies. Chemical defense involves the synthesis' pathways of specialized toxic,

20   repellent, or anti-nutritive metabolites to herbivores. Molecular evolutionary studies have

21   revealed the origin of new genes, acquisition and functional diversification along time in

22   different plant lineages. Here, using bioinformatic tools we analyze gene divergence of tropane

23   alkaloids (TAs) and terpene synthases (TPSs) in *Datura stramonium* and other species of

24   Solanaceae.

26   Methods. We compared the gene and amino acids sequence of TAs and TPSs by analyzing the

27   genomes, cDNA and proteins sequences of species Viridiplantae. We analyzed two genomes of

28   *Datura stramonium* (localities of Ticumán and Teotihucán), recently assembled in our lab,

29   transcriptomes of *Datura metel* and the genomes of other species of Solanaceae. Hence, we

30   analyzed the variation of TAs and TPSs to infer the genes involved in plant defense and plant

31   responses before stress. In addition, we analyzed protein modeling and molecular docking to

32   predict interactions between H6H and ligand; we translated the sequences (Teo19488, Tic8550

33   and Tic8549) obtained from the two genomes of *D. stramonium* by using Swiss-Model and

34   Ramachandran plot and MolProbity structure validation of Teo19488 protein model.

36   Results. In the TAs' we detected an expansion event in the tropinone reductase II (TRII) and the

37   ratio synonymous/non-synonymous substitutions indicate positive selection. In contrast, a

38   contraction event and negative selection was detected in the tropinone reductase I (TRI). In

39   Hyoscyamine 6 β-hydroxylase (H6H), the enzyme involved in the production of the tropane

40   alkaloids atropine and scopolamine, the synonymous/non-synonymous substitutions ratio in its

41   dominion indicates positive selection. In the case of terpenes (TPS), responsible of inducible

42   chemical defense in plants, we found 18 DsTPS of *D. stramomiun* and 7 in *D. metel*;

43   evolutionary analyses detected positive selection in TPS10.1 and TPS10.2 of *D. stramonium* and

44   *D. metel*. Further, comparison of copies of TPSs in *D. stramonium* detected variation among

45   them in the binding site. Duplication events and differentiation of TAs and TPSs of *D.*

46   *stramonium*, as compared to other solanaceous species, suggest their possible involvement on

47   adaptive evolution of chemical defense against herbivores. Protein modeling and docking show

48   that the three protein structures obtained of DsH6H from Teo19488, Tic-8550 and Tic8549

49   maintain the same interactions and the union site of 2OG-FeII_Oxy with the Hy-o ligand as in

50   6TTM of *Datura metel*. Further studies should be directed to experimentally assess gain

51   (overexpression) or loss (silencing) of function of duplicated genes.

52

53   Conclusion. Our results indicate differences in the number of gene copies involved in the

54   synthesis of tropane alkaloids, between the genomes of *Datura stramonium* from the two

55   Mexican populations (Ticumán and Teotihuacán). Furthermore, more copies of genes related to

56   the synthesis of tropane alkaloids (TRI, TRII, H6H, PMT-10) are found in *D. stramonium* as

57   compared to Viridiplantae. Likewise, for terpene synthases (TPS), TPS-10 is duplicated in *D.*

58   *stramonium* and *D. metel*. These results point future experimental studies of gain/loss of

59   duplicated genes.

60

61

62   INTRODUCTION
63
64   Plants are exposed to manifold environmental factors, biotic and abiotic, that affect their lifetime

65   reproductive success. To cope with the different types of environmental stress, plants have

66   evolved different mechanisms, ranging from morpho-anatomical, physiological, biochemical,

67   molecular and epigenetic modifications, among others (Lamalakshmi et al., 2017). An ubiquitous

68   defensive mechanism that protects plants from physical and biotic stresses is represented by the

69   synthesis of specialized metabolites (SM; the so-called "secondary metabolites"). Thousands of

70   chemical compounds found in plants promote protection against plants' natural enemies (i.e.,

71   pathogens, viruses and herbivores) (Mithöffer & Boland, 2012). More than 40 000 and 12 000

72   terpenoids and alkaloids, respectively, have been described in plants (Zhou & Pichersky, 2020).

73   Tropane alkaloids (TAs) and Terpene synthases (TPSs) are important specialized

74   molecules of plants that help to protect them from herbivores and pathogens (Kessler & Baldwin,

75   2002). Tropane alkaloids are distinctive, but not exclusive, of Solanaceae (Wink, 2003;

76   Mithöffer & Boland, 2012; Pigatto et al., 2015) and in *Datura stramonium,* implicated in plant

77   defense to herbivores (Shonle & Bergelson, 2000; Castillo et al., 2014; Miranda-Pérez et al.,

78   2016). Terpene synthases are enzymatic genes involved in the synthesis of volatile organic

79   compounds (VOCs) in many organisms (Picazo-Aragonés et al., 2020). Alkaloids and VOCs

80   negatively affect herbivores. The concentration of VOCs may even rise before herbivores start to

81   feed (Heil et al., 2008). Recent molecular studies in plants indicate modifications in the

82   expression of SM as defense mechanisms (Mansfield, 2000; Van Etten et al., 2001). Such

83   modifications may result from the composition of the herbivore community, their abundance and

84   dominance (Poelman & Keizer, 2016). In this study we present a bioinformatic analysis of the

85   changes in genes involved in the synthesis of tropane alkaloids and terpene synthases in

86   Solanaceae, with particular attention to *Datura stramonium*. Gene modifications may include

87   changes in the regulatory sequence or in the dominion, in the length of genes (including

88   duplication), and post-transduction modifications (Defoort et al., 2019). Here, we analyze events

89   of duplication in Solanaceae, with reference to whether duplicates are involved in modifications

90   of biosynthetic pathways of TAs and TPSs.

91   Generally, copies of genes can have different fates. If different copies are retained, they

92   may undergo sub-functionalization, dividing the original function. Alternatively, one copy may

93   undergo neofunctionalization, acquiring a new function. However, the most frequent outcome is

94   pseudogenization, whereby one copy becomes non-functional (pseudogene) and tends to

95   disappear due to accumulation of deleterious mutations. A pseudogene shows strong similarity to

96   the parental copy, which encodes a particular protein, but has one or more alterations such as

97   premature stop codons, mutations that cause phase changes, deletions and/or insertions that

98   prevent a protein from being functional (Ohno, 1970). It has been reported that in some species

99   the retention of pseudogenes for a long time only accumulates neutral mutations. However,

100 several questions arise regarding the survival mode of these inactivated genes (Ohno, 1970;

101 Lynch & Conery, 2003; Conant & Wolfe, 2008; Wang et al., 2012). Some retained functions

102 maintain developmental homeostasis, for instance, the ability to rapid respond to a wide variety

103 of environmental cues (Wang et al., 2012). One first defense response of plants involves the

104 activation of signaling mechanisms (De-La-Cruz et al., 2020), and therefore protein turnover. It

105 is estimated that more than 80% of the proteins are degraded through the proteosome pathway

106 (Maltsev, 1999). In recent years it has been reported that the responses of specialized metabolites

107 are grouped into several multigenic families. Therefore, we aim to elucidate the evolutionary

108 process of two gene families, TAs (Tropane alkaloids) and TPSs (terpenes synthases)

109 (Karunanithi et al., 2019*; De-La-Cruz et al., 2021). Studies conducted in Solanaceae have found

110 rapid responses to biotic and abiotic stimuli, in a matter of minutes, of genes involved in the

111 synthesis of specialized metabolites (atropine, scopolamine, VOCs) (Brille et al., 2019). To date,

112 several groups have reported different TAs and TPSs and have found similar results (Bharat &

113 Sharma, 2015).

114   In this study we analyze the evolutionary history of genes involved in the synthesis of

115 tropane alkaloids and terpene synthase in the Solanaceae family and specifically in draft

116 genomes of two plants of *Datura stramonium* from populations of México (Ticumán and

117 Teotihuacán). Overall, our bioinformatic analyses infer duplication of the genes of the main

118 biosynthetic pathways leading to specialized metabolites.

119

120 MATERIALS AND METHODS

121

122 **Selection and bioinformatic analyses: Identification of TAs and TPs, genomics, transcripts**
123 **and protein sequences**

124

125 We obtained the sequences of nucleotide, transcripts and proteins of TAs and TPSs. Nine

126 genomes (Table 1) of Solanaceae species were sourced for protein coding genes and CDS

127 (Coding Sequence) genomes from the Sol Genomics Network (see links in Table 1): *Nicotiana*

128 *tabacum* (Edwards et al. 2017), *Nicotiana sylvestris* (Sierro et al. 2013), *Nicotiana attenuata* (Xu

129 et al. 2017), *Nicotiana tomentosiformis* (Sierro et al. 2013), *Solanum pimpinellifolium* (Razali et

130 al. 2018), *Solanum lycopersicum* (The Tomato Genome Consortium, 2012), *Solanum pennellii*

131 (Bolger et al. 2014 b), *Solanum tuberosum* (Xu et al. 2011), *Capsicum annuum*, CM334 v1.55

132 (Kim et al. 2014). In addition, data of *Datura stramonium* were extracted from protein whereas

133 TAs and TPSs' sequence were obtained from the genome of *D. stramonium* from the Mexican

134 populations of Ticumán, Morelos (18°44'28.19"N, 99° 7'44.26"W), and Teotihuacán, Estado de

135 México (19°41'16.58"N, 98°50'4.14"W) (https://github.com/icruz1989/Datura-stramonium-

136 genome-project) (De-la-Cruz et al., 2021) and transcriptomes (https://medplantrnaseq.org/).

137   We downloaded the whole genome of *D. stramonium* to extract the TPSs genes and gene

138 prediction was performed using AUGUSTUS (Stanke et al., 2006). All sequences were uploaded

139 and mapped using as reference species of Solanaceae. A total of 366 000 genes were extracted,

140 and gene redundancy was reduced with CDHIT (Chen et al., 2016). The selected TPSs genes

141 were blasted against the NCBI, UNIPROT databases and the nine genomes of Solanaceae, using

142 a threshold value of $1E^{-5}$. All TPSs were downloaded from Viridiplantae followed by a protein-

143 protein blast; hits above 80% of similarity and 70% of cover were saved for further analyses.

144 Meta-alignments of protein sequences were performed using T-coffee (Wallace et al., 2006),

145 MUSCLE v3.8 (Edgar, 2004) and MAFF (https://mafft.cbrc.jp/alignment/server/) v6. (Rowicki

146 et al., 2019).

147

148 **Identification of orthogroups and inference of phylogenetic tree of TPa and TPSs.**

149 Construction of orthogroups, gene families, was performed according to De-la-Cruz et al. (2021)

150 using Orthofinder v2.3.3 (Emms & Kelly 2015, 2019). The phylogenetic tree inference was

151 accomplished using the program BEAST with the pair-wise deletion option, and reliability of the

152 obtained phylogenetic tree was tested by bootstrapping with 1000,000 MKKM chains. The

153 model of substitution selected was JTT+G+F, with a branch support of 1 (Drummond et al.,

154 2002). The trees were edited at the interactive tree of life (https://itol.embl.de) (Letunic & Bork,

155 2019).

156 We downloaded all orthologous sequences of TAs and TPSs of *D. stramonium*, *D. metel* and

157 nine species of Solanaceae (see Table 1) to construct the orthogroups. Downloaded sequences of

158 proteins and ORFs were extracted using TransDecoder and corroborated with GenScan. TPS

159 genes were translated with GenScan and checked for the right direction and correct CDS; we

160 corroborated this with the NCBI. The database used for BLAST was Uniprot; all available data

161 for the analyzed plant species were downloaded. These amounts 847,544 sequences of TAs and

162 TPSs. BLAST was carried out with a filtering E-value $< 1e^{-5}$ and only hits above 85 % were

163 selected (Data S1, S2). Orthogroups were inferred using Orthfinder v 2.3.3. DIAMONS (Emms

164 & Kelly 2015, 2019) while OrthoMCL was used to assess orthologs among species (Chen et al.,

165 2006).

166

Protein modeling

167

168 Molecular modelling and protein-ligand docking was performed with GLIDE software

169 (Schrödinger Release 2021-1: Glide, Schrödinger, LLC, New York, NY, 2021). Simulations of

170 interactions for each protein was done in a quadrat of 6LU/ processed with the same coordinates

171 of the crystalized ligand as described in Gutierrez et al. (2020).

172 The SWISS-MODEL (http://www.expasy.org/swissmod/SWISS-MODEL.html)

173 (Waterhouse et al., 2018) was employed to model the tertiary structure of proteins, from the

174 previously determined structure by the dominion DIOX_N (PF14226) (Hagel et al., 2010) as a

175 blueprint of modelling. Models were visualized and analyzed through Visual Molecular

176 Dynamics (Waterhouse et al., 2018; Bienert et al., 2017)

177

Molecular Docking

178

179 The *in-silico* docking (AutodockVina) of the PDB files (produced by the SWISS-Model) of

180 Tic8550 (Ticumán) and Teo19488 (Teotihuacán) was performed in a simulation of molecular

181 coupling with the native ligand of H6H, which is Hyoscyamine (Hy-o). For docking, the protein

182 was prepared with AutodockTools, assigning the parameters of energies and force fields of

183 AutoDock4.0. Then, we obtained the results of the thermodynamically favored conformation for

184 the Hy-o ligand using the algorithm VINA. The selected conformation had affinity coupling

185 parameters of -7.3 and -7.5 kcl/mol for Teo19488 and Tic8550, respectively. As positive control,

186   we performed the re-docking of the molecule of reference 6TTM (*Datura metel*), eliminating the

187   ligand of the crystallized structure; the adopted conformation is similar to that obtained by the

188   method of crystallization, and the coupling affinity for the reference model was -7.9 kcal/mol.

189   The produced models are theoretically correct, and it is predicted that these have an

190   oxidoreductase and use Hy-o as substrate.

191

192   *Tropane alkaloids*

193   Analyses detected variation in the number of TAs gene copies in *D. stramonium*: DsTRI,

194   DsTRII, DsPMT and DsH6H. Compared to Viridiplantae *D. stramonium* has one or more

195   copies of these genes suggesting a genome expansion of TAs.

196   We detected eight copies of tropinone reductase I (DsTRI) and four copies of DsTRII. TRI

197   and TRII define two groups in Solanaceae (Fig. 1A). The gene DsTRII have had duplications

198   events in *D. stramonium* (Fig. 1B). The first duplication occurs in Rosidae-Asteridae,

199   followed by three duplications at the base of Solanaceae and one more in *D. stramonium* (Fig.

200   1B). The phylogeny and structure 32 of protein sequences (TRI) from different plant species

201   show variability in the ADH_Short_C2 TRI dominion (Fig. 2); in *D. stramonium*

202   (Tic23_dati33027) there a duplication of this dominion but shorter (Fig. 2).

203   In Hyoscyamine- 6-ß-hydroxylase (*DsH6H*) in the Ticumán genome we detected two

204   copies of the gene, Tic8550 and Tic8549 (Fig. 3A). Tic 8550 has a tandem duplication of the

205   DIOX-N dominion of 79 amino acids (Fig. 3) whereas Tic8549 has the dominion 2-

206   oxaglutarate (65 amino acids) (Fig. 3A). In contrast, in the Teotihuacán genome, *DsH6H* (i.e.,

207   Teo19488) has only one copy (Fig. 3). The alignment of 29 sequences of H6H of different

208   plant species indicate the variation in the duplicated DIOX-N dominion in Tic8550 (Fig. 3B).

209   In N-methyl putremescine transferase (DsPMT) we found variation in the number of gene

210   copies in comparison to other Solanaceae (Fig. 4). In Tucumán's genome there are three

211   copies while two copies were detected in Teotihuacán's genome (Fig. 4, Suppl. Fig. 1).

212   *Terpene synthases*

213 Eighteen TPSs genes were found in the genomes of *D. stramonium* (Fig. 5; Table S1) and seven

214 in *Datura metel* (Fig. S2). The 18 *DsTPS* found in *D. stramonium* are distributed in four

215 subfamilies, identified for other angiosperms: Nine in TPSa, 3 in TPSb, 2 in TPSc, 2 in TPSg

216 and 2 TPS- unknown (Fig. S3, Table S1 and S2) (Huang, 2013, 2017; Falara et al., 2011).

217   The domains of these TPSs, directly involved in the biosynthesis of terpenoids, show

218 expansion events and positive selection in TPS10 (Fig. 6; Table S3). Changes in the dominions

219 are: Terpene synthase, N-terminal domain (IPR001906), Terpenoid cyclases/protein prenyl

220 transferase alpha-alpha toroid (IPR008930), Terpene synthase, metal-binding domain

221 (IPR005630), Terpene cyclase-like 1, C-terminal (IPR034741). Analyses indicate that

222 Solanaceae have the gene TPS10.1 whereas only *D. stramonium* and *D. metel* possess TPS10.2

223 (Fig. 7 and Table S2 and S3,); this possibly is a duplicate of TPS10.1_like.

224

225 **Protein modeling and molecular docking**

226

227 Superimposition of the reported protein structure of H6H from *D. metel* (PDB ID 6TTM) with

228 the corresponding model obtained from *DsH6H* sequences for the two genomes of *D.*

229 *stramonium* from two Mexican populations (Teo19488 and Tic8550), indicate similarities (Fig.

230 7A). The Teo19488 predicted structure match very well with the DmH6H structure (Fig. Suppl.

231 2); similarly, Tic8550 matches well except for some residues that were not modeled (Fig. 7B,

232 Fig. Suppl 3). This fragment of Tic8550 have a big terminal protein sequence

233 (NQTMMLKWLLFLKLVYKQYLYKLYSTIGEKAEKDVINHGVPEKIMVEAMEFTKSFL

234 HCLLRKKKSLSQKEASIKQSSML) of 79 amino acids was not modeled (Fig. 7B, cf. Fig.

235 3B). Further, a NCBI protein-protein blast of this sequence aligns well to the region: 33-72 of

236 DmH6H fragment; we analyzed this fragment separately as s Tic8550_Fragment (Fig.86). The

237 Tic8549   sequence   corresponds   to   a   65   amino   acids   sequence

238 (MLPIIPRPKSTLGAGGHYDGNIITFLQQDCLACNNSLLRMTNGLLLNLSYCFCGLSGT

239 HSKGYEQ) that was modeled as hyoscyamine 6 beta-hydroxylase like fragment (Fig. 8). This

240 fragment (Tic8549) aligns to DmH6H in one section of the binding pocket, where His-217 and

241 Asp-219 are coordinated with Ni^{2+} (Fig. 8C). The presence of these amino acids in Tic8549

242   sequence is interesting since Ni^{2+} ion, a surrogate of the natively present Fe^{2+} ion, is coordinated

243   by the side chains of His-217 (strand βII), Asp-219 (loop βII/βIII) and His-274 (strand βVII)

244   that forms a metal binding **His-X-Asp…His** motif, highly conserved in the oxoglutarate

245   dependent oxygenases (ODD) family (Kluza *et al.*, 2020).

246   Protein models of H6H for *D. metel* and *D. stramonium* (Teo19488 and Tic8550) predict

247   more interaction between residues in Tic8550 than on *D. metel* and Teo19488 (Fig. 9). It has

248   recently been reported that the binding pocket (BP) of DmH6H is mainly formed by

249   hydrophobic amino acids. It has been demonstrated that in the crystalized H6H the phenyl ring

250   of Hyo is bounded in an aromatic cage formed by Phe-103, Tyr-295, Tyr-319, Phe-322 and

251   Tyr-326, of which the most prominent bound is with Tyr-326 which forms CH-π hydrogen

252   bonds with the phenyl ring of Hyo in an edge-to-face bidentate manner (Kluza *et al.,* 2020)

253   (Fig. 9A, D). Our results shown that the predicted model for Teo19488 have the same amino

254   acids forming the aromatic cage composed by Phe-88, Tyr-255, Tyr-279, Phe-282 and Tyr-286

255   and the predicted pose of Hyo obtained by the docking *in silico* shows that the phenyl ring is

256   inside of this aromatic cage (Fig. 9B). Nevertheless, the main interaction with Tyr-286 was not

257   detected (Fig. 9E). This could be because the cutoff distance was set 4 Å and the predicted pose

258   of this residue is beyond this cutoff distance. The predicted structure of Tic8550 has different

259   amino acids arrangement inside the binding pocket. This predicted structure does not contain

260   the aromatic cage, instead it is predicted that the phenyl ring of Hyo inside of a cavity formed

261   by the non-polar amino acids Leu-308, Leu-92, Ile-179, Met-181, Met-292 and the polar

262   uncharged amino acids Asn-206 and Asn-289 (Fig. 9C). Also, it is predicted that the phenyl

263   ring has several non-polar interactions with these side chain amino acids (Fig. 9F). The Asn-

264   206 forms a conventional hydrogen bond with the hydroxyl group from Hyo which may

265   stabilize the structure in the BP.

266

267

268

269   DISCUSSION
270

271   Gene duplicates product of molecular evolution are the raw material for evolutionary innovation

272   (Defoort et al., 2019). Recent findings have uncovered the interactions and contrasting functions

273   of genes that may help plants to confront environmental stresses in nature and serve as important

274   agronomic characteristics. Hence, to trace the origin, molecular mechanisms, evolutionary fate

275   and function of gene duplicates is a main goal (Panchy et al., 2016; Soltis & Soltis 2016; Van de

276   Peer et al., 2017). The phylogenetic analyses of genes involved in the synthesis of tropane

277   alkaloids (*DsPMT, DsTRI, DsTrII* and *DsH6H*) here presented, indicate that two genes (DsTrII

278   and DsH6H) have expanded in *D. stramonium*, presenting five and two copies, respectively.

279   Previous studies have shown that tropinone reductases are involved in the bifurcation of the

280   synthesis pathways in Solanaceae (Dräger, et al., 1992; Nakajima 1993; Kanehisa & Sato, 2019).

281   These genes are responsive to environmental stimuli and activate the signaling pathways (Tian et

282   al., 2016; Campos et al., 2014). The presence of these genes in the Solanaceae, as revealed by

283   phylogenetic analyses, suggests that these duplicates in *D. stramonium* may contribute to

284   increase the production of tropane alkaloids. The absence of TRII in *Zea mays* and *A. thaliana*

285   suggest that these species never had this gene or lost it during their evolution. However, the gene

286   is also absent in species of other clades (e.gr., *Brassica rapa, Glycine max, Vitis vinifera,* etc.;

287   Fig. 1). Another possibility is that TRII is not related to the synthesis of tropane alkaloids in

288   these species (Miller, 1999).

289   The gene DsPMT of *D. stramonium* found in the population of Ticumán, México, had an

290   extra domain of spermine-synthase in comparison with its homologous gene found in the same

291   species but from the population of Teotihuacán, México (De-la-Cruz et al., 2021). We named

292   this gene as *Ds*PMT7568ti (Fig. 4), Spermine gene is a potent plant defense activator, with

293   protective effects of broad-spectrum (Seifi et al., 2019). Reports indicate that overexpression of

294   spermidine synthase enhances tolerance to multiple environmental stresses, including the attack

295   by herbivores and pathogens (Kasukabe et al., 2004). In addition, PMT is a key enzyme in the

296   catalysis of N-methylputrescine from putrescine and S-adenosyl-L- methionine and triggers the

297   production of hygrine and other different tropane alkaloids (Kanehisa & Sato, 2019).

298   In contrast to other plants, in Solanaceae (*Atropa belladona, Hyosciamus niger* and

299   *Datura stramonium*), TRII is duplicated and under positive selection. In this line, we hypothesize

300   that TRII may have acquired a different functional importance after its duplication in *D.*

301   *stramonium* and hence, it has retained multiple copies of the gene. Yet, this hypothesis warrants

302 further analysis. The phylogeny indicates an independent history, for instance, from

303 Brassicaceae.

304 Perhaps TRs in Brassicales have other functions, different to the production of alkaloids,

305 but using the same pathways. For instance, they may have conferred plasticity to deal with

306 environmental stresses. Plant species with small genome size may have lost considerable

307 fractions of their genome. Thus, enzymes with similar function may have been replaced by TRII.

308 Of the four subfamilies of TPSs found in the *D. stramonium*'s genome, three have been reported

309 in other angiosperms (TPSa, TPSb and TPSg). TPS10 is member of the subfamily *a* and its role

310 in defense against pathogens and herbivores has been dilucidated in *Zea mays* (Huang et al.,

311 2017; Tholl, 2006). We detected positive selection on TPS; positive selection on TPS10.1 and

312 TPS10.2 was detected in *D. stramonium* and *D. metel*. We speculate that the presence of two

313 copies of this gene can contribute to plant defense against plants' natural enemies (Köllner et al.,

314 2009, 2020).

315 In this line, duplicated genes involved in the specialized metabolism and plant defense

316 often show differential restrictions (constraints, trade-offs, limitations) either indicating positive

317 selection or differential selection (Cotton et al., 2006). Apparently, TRII displays genic

318 redundancy since it possesses different copies and polymorphic sites. However, this may not be

319 the case since this process is only present in genes of primary metabolism or development

320 (Cotton et al., 2006; Maltsev et al., 2005) which are constitutive and display low variation. The

321 gene H6H is involved in the last step in the production of scopolamine and atropine in one of the

322 pathways for the production of tropane alkaloids (Kanehisa & Sato 2019). Here, we detected one

323 duplicated in tandem of this gene in *D. stramonium* (Ticumán has two domains of DIOX_N

324 (PF14226), although each copy belongs to two different gene families (OG0028637 and

325 OG0043057; cf. Fig. 3). Similarly, we found 18 TPSs in *D. stramonium* that likely are involved

326 in plant defense against herbivores like in other plants species (e.gr., Bharat & Sharma 2015).

327 Lineages in the Solanaceae have high values of the ratio Ka/Ks (Table 2). Thus, positive

328 selection may have acted on TPS10.1 and TPS10.2 of *D. stramonium*, not recently and the

329 selection effect on allelic diversity may be masked by the accumulation of neutral mutation.

330 On the other hand, results of protein simulation and molecular docking suggest that the

331 structure and interactions of this protein (H6H) in Tic8550 might have a different structure to

332   that of the known H6H; we speculate if this may be related to enhancing the efficiency of

333   alkaloids' synthesis.

334   It is interesting to note that the Tic8549 possess only the fragment 2OG-FeII_Oxy,

335   PF03171) whereas Tic8550 possess an extra dominion (DIOX_N) which aligns structurally to a

336   highly conserved region of H6H proteins (Fig. S2 and Fig. S3) (and associated to a coordination

337   of binding metal Ni^{+2} (Fig. S5). Further, it is interesting noting that a particular gene Tic8549 is

338   associated to this fragment DIOX_N (PF14226).

339   Finally, the number of copies and differentiation of TAs and TPSs of *D. stramonium*, as

340   compared to other solanaceous species, suggest their involvement on adaptive evolution of

341   chemical defense against herbivores. Further studies should be directed to experimentally assess

342   gain (overexpression) or loss (silencing) of function of duplicated genes.

343

344   CONCLUSION

345

346   Our results indicate differences in the number of gene copies involved in the synthesis of tropane

347   alkaloids, between the genomes of *Datura stramonium* from the two Mexican populations

348   (Ticumán and Teotihuacán). Furthermore, more copies of genes related to the synthesis of

349   tropane alkaloids (TRI, TRII, H6H, PMT-10) are found in *D. stramonium* as compared to

350   Viridiplantae. Likewise, for terpene synthases (TPS), TPS-10 is duplicated in *D. stramonium* and

351   *D. metel*. These results point future experimental studies of gain/loss of duplicated genes.

352

353   Links to data on Figshare (Data S3)

354   ACKNOWLEDGEMENTS

355   We are grateful to the three reviewers whose comments and suggestions improved this

356   contribution. We thank to Tonatiuh Campos García his advice in protein modelling. We

357   appreciate the strong support from CONACyT and DGAPA UNAM. To the members of the

358   *Laboratory of Ecological Genetics and Evolution* for encouragement and help.

359

360   REFERENCES

361

362 Bailey TL, Williams N, Misleh C, Li WW. 2006. MEME: discovering and analyzing DNA and

363 protein sequence motifs. *Nucleic Acids Research* **34(2)**:W369-W373 DOI: 10.1093/nar/gkl198.

364

365 Bienert, S, Waterhouse A, de Beer TAP, Tauriello G, Studer G, Bordoli L, Schwede T. 2017.

366 The SWISS-MODEL Repository - new features and functionality *Nucleic Acids Res*earch

367 **45(D1)**: D313-D319 DOI:10.1093/nar/gkw1132.

368

369 Birgit D. 2006. Tropinone reductases, enzymes at the branch point of tropane alkaloid

370 metabolism. *Phytochemistry* **67(4):**327–337 DOI: 10.1016/j.phytochem.2005.12.001.

371 Bolger A, Scossa F, Bolger M E, Lanz C, Maumus F, Tohge T, Quesneville H, Saleh Alseekh S,

372 Sørensen I, Lichtenstein G, Fich E A, Conte M, Keller H, Schneeberger K, Schwacke R, Ofner I,

373 Vrebalov J, Xu Y, Osorio S, Alves Aflitos S, Schijlen E, Jiménez-Goméz J M, Ryngajllo M,

374 Kimura S, Kumar R, Koenig D, Headland L R, Maloof J N, Sinha N, van Ham R C H, Klein

375 Lankhorst R M, Mao L, Vogel A, Arsova B, Panstruga R, Fei Z, Rose J K C, Zamir D, Carrari F,

376 Giovannoni J J, Weigel D, Usadel B, Fernie A R. 2014 The genome of the stress-tolerant wild

377 tomato species *Solanum pennellii*. *Nature Genetics* **46**:1034-1038 DOI: 10.1038/ng.3046.

378

379 Brille F, Loreto F, Baccelli I. 2019. Exploiting plant Volatile Organic Compounds (VOCs) in

380 agriculture to improve sustainable defense strategies and productivity of crops. *Frontiers in*

381 *Plant Science* **10**: 264 DOI: 10.3389/fpls.2019.00264.

382

383 Campos ML, Kang JH, Howe GA. 2014. Jasmonate-triggered plant immunity. *Journal of*

384 *Chemical Ecology*, **40(7)**:657-675 DOI 10.1007/s10886-014-0468-3.

385

386 Castillo G, Cruz LL, Tapia-López R, Olmedo-Vicente E, Carmona D, Anaya-Lang AL, Fornoni

387 J, Andraca-Gómez G, Valverde PL, Núñez-Farfán J. 2014. Selection mosaic exerted by specialist

388 and generalist herbivores on chemical and physical defense of *Datura stramonium*. *PLoS ONE*

389 **9(7):** e102478 DOI: 10.1371/journal.pone.0102478.

390

391   Chen F, Mackey AJ, Stoeckert CJ, Roos D. 2006. OrthoMCL-DB: querying a comprehensive

392   multi-species collection of ortholog groups. *Nucleic Acids Research* **34**:363-368.

393   https://doi.org/10.1093/nar/gkj123

394

395   Chen Q, Wan Y, Lei Y, Zobel J, Verspoor K. 2016. Evaluation of CD-HIT for constructing non-

396   redundant databases IEEE *International Conference on Bioinformatics and Biomedicine (BIBM)*

397   pp. 703-706 DOI: 10.1109/BIBM.2016.7822604.

398

399   Conan GC, Wolfe KH. 2008. Turning a hobby into a job: How duplicated genes find new

400   functions. *Nature Reviews Genetics* **9**:938-950 DOI: 10.1038/nrg2482.

401

402   Cotton JA, Page RDM. 2006. Multiple mechanisms promote the retained expression of gene

403   duplicates in the tetraploid frog *Xenopus laevis*. *PLoS Genetics* 2(4):e56 DOI:

404   10.1371/journal.pgen.0020056.

405

406   Defoort J, Van de Peer Y, Carretero-Paule L. 2019. The Evolution of gene duplicates in

407   angiosperms and the impact of protein–protein interactions and the mechanism of duplication.

408   *Genome Biology and Evolution* **11(4)**:2292-2305 DOI: 10.1093/gbe/evz156.

409

410   De-la-Cruz I M, Hallab, A, Olivares-Pinto U, Tapia-López R, Velázquez-Márquez S, Piñero D,

411   Oyama K, Usadel B, Núñez-Farfán, J. 2021. Genomic signatures of the evolution of defence

412   against its natural enemies in the poisonous and medicinal plant *Datura stramonium*

413   (Solanaceae). *Scientific Reports* **11(1)**:1-19. https://doi.org/10.1038/s41598-020-79194-1

414

415   De-la-Cruz I.M., Velázquez-Márquez S., Núñez-Farfán J. 2020. What do we know about the

416   genetic basis of plant defensive responses to herbivores? A minireview. In: Núñez-Farfán J.,

417   Valverde P. (eds) Evolutionary Ecology of Plant-Herbivore Interaction. *Springe*r

418   DOI:10.1007/978-3-030-46012-9_16

419

420   Dräger B, Portsteffen A, Schaal A, McCabe P, Peerless A. 1992. Levels of tropinone- reductase

421   activities influence the spectrum of tropane esters found in transformed root cultures of *Datura*

422   *stramonium (L.). Planta* **188(4)**:81-586 DOI: 10.1007/BF00197052.

423

424   Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W. 2002. Estimating mutation parameters,

425   population history and genealogy simultaneously from temporally spaced sequence data.

426   *Genetics* 161:1307-132.

427

428   Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high

429   throughput. *Nucleic Acids Research* **32(5)**:1792-7 DOI: 10.1093/nar/gkh340.

430

431   Edwards KD, Fernandez-Pozo N, Drake-Stowe K, Humphry M, Evans AD, Bombarely A, Allen

432   F, Hurts R, White B, Kernodle SP, Bromley J, Sanchez-Tamburrino J, Lewis R, Mueller L. 2017.

433   A reference genome for *Nicotiana tabacum* enables map-based cloning of homologous loci

434   implicated in nitrogen utilization efficiency. *BMC Genomics* **8**:448 DOI: 10.1186/s12864-017-

435   3791-6.

436

437   Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome

438   comparisons dramatically improves orthogroup inference accuracy *Genome Biology* 16,157 DOI:

439   10.1186/s13059-015-0721-2.

440

441   Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative

442   genomics. *Genome Biology* **20**:238 DOI: 10.1186/s13059-019-1832-y.

443

444   Falara V, Akhtar TA, Thuong TH, Spyropoulou EA, Bleeker PM, Schauvinhold I, Matsuba Y,

445   Bonini ME, Schilmiller AL, Last RL, Schuurink RC, Pichersky E. 2011. The Tomato Terpene

446   Synthase Gene Family. *Plant Physiology* 1**57**:770–789 DOI: 10.1104/pp.111.179648.

447

448   Hagel JM & Facchini PJ. 2010. Dioxygenases catalyze the O-demethylation steps of morphine

449   biosynthesis in opium poppy. *Nature Chemical Biology* **6**:273 DOI:10.1038/nchembio.317.

450

451 Heil M, Lion U, Boland W. 2008. Defense-Inducing Volatiles: In search of the active motif.

452 *Journal of Chemical Ecology* **34(5)**:601–604 DOI: 10.1007/s10886-008-9464-9.

453

454 Huang AC, Kautsar SA, Hong YJ, Medema MH, Bond AD, Tantillo DJ and Osbourn A. 2017.

455 Unearthing a sesterterpene biosynthetic repertoire in the Brassicaceae through genome mining

456 reveals convergent evolution. *Proceedings of the National Academy of Sciences USA* .

457 **114**:E6005–E6014 DOI: 10.1073/pnas.1705567114

458

459 Huang X, Xiao Y, Köllner T G, Zhang W, Wu J, Wu J, Zhang Y. 2013. Identification and

460 characterization of (E)-β-caryophyllene synthase and α/β-pinene synthase potentially involved in

461 constitutive and herbivore-induced terpene formation in cotton. *Plant Physiology and*

462 *Biochemistry* **73:**302-308 DOI: 10.1016/j.plaphy.2013.10.017.

463 Kanehisa M, Sato Y. 2019. KEGG Mapper for inferring cellular functions from protein

464 sequences. *Protein Science* **29**:28-35 DOI: 10.1002/pro.3711.

465

466 Kasukabe Y, He L, Nada K, Misawa S, Ihara I, Tachibana S. 2004. Overexpression of

467 spermidine synthase enhances tolerance to multiple environmental stresses and up-regulates the

468 expression of various stress-regulated genes in transgenic *Arabidopsis thaliana*. *Plant & Cell*

469 *Physiology* **45**:712-22 DOI: 10.1093/pcp/pch083.

470

471 Kessler A., Baldwin I. 2002. Plant responses to insect herbivory: the emerging molecular

472 analysis. *Annual Review of Plant Biology* **53**: 99-328 DOI: 10.1126/science.291.5511.2141.

473

474 Kluza A, Wojdyla Z, Mrugala B, Kurpiewska K, Porebski PJ, Niedzialkowska E, Borowski T.

475 2020. Regioselectivity of hyoscyamine 6β-hydroxylase-catalysed hydroxylation as revealed by

476 high-resolution structural information and QM/MM calculations. *Dalton Transactions* **49**:4454-

477 4469 DOI: 10.1039/d0dt00302f.

478

479 Köllner TG, Degenhardt J & Gershenzon J. 2020. The product specificities of maize terpene

480 synthases TPS4 and TPS10 are determined both by active site amino acids and residues adjacent

481 to the active site. *Plants* **9:**552 DOI:10.3390/plants9050552.

482

483   Köllner TG, Gershenzon J & Degenhardt J. 2009. Molecular and biochemical evolution of maize

484   terpene synthase 10, an enzyme of indirect defense. *Phytochemistry* **70**:1139 1145 DOI:

485   10.1016/j.phytochem.2009.06.011.

486

487   Kurananithi PS & Zerbe P. 2019. Terpene Synthases as Metabolic Gatekeepers in the Evolution

488   of Plant Terpenoid Chemical Diversity. *Frontiers in Plant Science* **10**:1166 DOI:

489   10.3389/fpls.2019.01166

490

491   Lamalakshmi E, Kumar S, Basanta T, Susheel S. Beemrote A, Chingakham P, Chongtham SK,

492   Chongtham HS, Yumlembam R, Haribhushan A, Prakash N, Wani S.H .2017. Adaptation

493   strategies and defence mechanisms of plants during environmental stress. In: Ghorbanpour M,

494   Varma A (eds), *Medicinal Plants and Environmental Challenges.* pp. 359-413 *Springer* DOI:

495   org/10.1007/978-3-319-68717-9_20.

496

497   Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F,

498   Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007. Clustal W and

499   Clustal X version 2.0. *Bioinformatics* **23**:2947–2948 DOI: 10.1093/bioinformatics/btm404.

500   Letunic I & Bork P. 2019. Interactive Tree of Life (iTOL) v4: recent updates and new

501   developments. *Nucleic Acids Research* **47**:256-259 DOI: 10.1093/nar/gkz239.

502

503   Lynch M & Conery JS. 2003. The origins of genome complexity. *Science* **5649(302)**:1401-1409

504   DOI: 10.1126/science.1089370.

505

506   Maltsev N, Glass EM, Ovchinnikova G, Gu Z. 2005. Molecular mechanisms involved in

507   robustness of yeast central metabolism against null mutation. *Journal of Biochemistry*

508   **34(5)**:601–604. DOI: 10.1093/jb/mvi017.

509

510   Miller JD, Arteca RN, Pell EJ. 1999. Senescence-associated gene expression during ozone-

511   induced leaf senescence in Arabidopsis. *Plant Physiology* **120**:1015–1024 DOI:

512   10.1104/pp.120.4.1015.

513

514   Miranda-Pérez A, Castillo G, Hernández-Cumplido J, Valverde PL, Borbolla M, Cruz LL,

515   Tapia-López R, Fornoni J, Flores-Ortiz CM, Núñez-Farfán J. 2016. Natural selection drives

516   chemical resistance of *Datura stramonium*. *PeerJ* 4: e1898 DOI: 10.7717/peerj.1898.

517

518   Mithöfer A, Boland W. 2012. Plant defense against herbivores: chemical aspects. *Annual Review*

519   *of Plant Biology* **63**:431-50 DOI: 10.1146/annurev-arplant-042110-103854.

520

521   Mondragon-Palomino M, Meyers BC, Michelmore W, Gaut BS. 2002. Patterns of positive

522   selection in the complete NBS-LRR gene family of *Arabidopsis thaliana*. *Genome Research*

523   **12(9)**:1305–1315 DOI: 10.1101/gr.159402.

524

525   Nakajima K, Hashimoto T, Yamada Y. 1993. Two tropinone reductases with different stereo

526   specificities are short-chain dehydrogenases evolved from a common ancestor. *Proceedings of*

527   *the National Academic of Science USA* **90**:9591-9595 DOI: 10.1073/pnas.90.20.9591.

528

529   Ohno S. 1970. Evolution by gene duplication. Springer-Verlag Press.

530

531   Panchy N, Lehti-Shiu M, Shiu SH. 2016. Evolution of gene duplication in plants. *Plant*

532   *Physiology* **171(4)**:2294–2316 DOI: 10.1104/pp.16.0052.

533

534   Picazo-Aragonés J, Terrab A, Balao F. 2020. Plant volatile organic compounds evolution

535   transcriptional regulation, epigenetics and polyploidy. *International Journal of Molecular*

536   *Sciences* 218986 DOI:10.3390/ijms21238956

537

538   Pigatto AS, Blanco CC, Mentz LA, Soares GL. 2015. Tropane alkaloids and calystegines as

539   chemotaxonomic markers in the Solanaceae. *Anais de Academia Brasiliera de Ciencias*

540   **87(4)**:2139 49.

541

542   Razali R, S. Bougouffa, Morton M, Lightfoot DJ, Alam I, Essack M, Arold S, Kamau A,

543   Schmöckel S, Pailles Y, Shaid M, Michel C. Al-Babili S, Shwen Y, Tester M, Bajic, Negrap S.

544   2018. The genome sequence of the wild Tomato *Solanum pimpinellifolium* provides insights into

545   salinity tolerance. *Frontiers in Plant Science* **9**:1402 DOI: 10.3389/fpls.2018.01402.

546

547   Seifi HS & Shelp BJ. 2019. Spermine differentially refines plant defense responses against biotic

548   and abiotic stresses. *Frontiers in Plant Science* **10**:117 DOI: 10.3389/fpls.2019.00117.

549

550   Shidore T, Triplett LR. 2017. Toxin-antitoxin systems: implications for plant disease. *Annual*

551   *Review of Phytopathology* **55(1)**:161-179 DOI: 10.1146/annurev-phyto-080516-035559.

552

553   Shonle I, Bergelson J. 2000. Evolutionary ecology of the tropane alkaloids of *Datura*

554   *stramonium L.* (Solanaceae). *Evolution* **54(3)**:778–788 DOI: 10.1111/j.0014-3820.

555   2000.tb00079. x.

556

557   Sierro, N. Battey JN, Quadi S, Boved L, Goeofert S, Bakaher N, Ivanov N.2013. Reference

558   genomes and transcriptomes of *Nicotiana sylvestris* and *Nicotiana tomentosiformis*. *Genome*

559   *Biology***14**, R60 DOI: 10.1186/gb-2013-14-6-r60.

560

561   Simonsen M, Mailund T, Pedersen CNS. 2008. Rapid neighbor-joining. *Lectures Notes*

562   *Computational Sciences* **5251**: 113-122 DOI: 10.1007/978-3-540-87361-7_10.

563

564   Singh B, Sharma RA. 2015. Plant terpenes: defense responses, phylogenetic analysis, regulation

565   and clinical applications. *Biotechnology* **5(2)**:129–151 DOI: 10.1007/s13205-014-0220-2.

566

567   Soltis PS, Soltis DE. 2016. Ancient WGD events as drivers of key innovations in angiosperms.

568   *Current Opinion in Plant Biology* **30**:159–165 DOI: 10.1016/j.pbi.2016.03.015.

569

570   Stanke M, Keller O, Gunduz I, Hayes A, Waak S. Morgenstern B. 2006. AUGUSTUS: ab initio

571   prediction of alternative transcripts. *Nucleic Acids Research* **34(2):**W435:439 DOI:

572   10.1093/nar/gkl200.

573

574   The Tomato Genome Consortium. 2012. The tomato genome sequence provides insights into

575   fleshy fruit evolution. *Nature* **485**:635–641 (2012). https://doi.org/10.1038/nature11119

576   Schrödinger Release 2021-1**:** Desmond Molecular Dynamics System, D. E. Shaw Research, New

577   York, NY, 2021. Maestro-Desmond Interoperability Tools, Schrödinger, New York, NY, 2021.

578   Seifi H S, Zarei A, Hsiang T, Shelp BJ. 2019. Spermine is a potent plant defense activator

579   against gray mold disease on *Solanum lycopersicum*, *Phaseolus vulgaris*, and *Arabidopsis*

580   *thaliana*. *Phytopathology.* **109(8)**:1367-1377 DOI: 10.1094/PHYTO-12-18-0470-R.

581

582   Tholl, D. 2006. Terpene synthases and the regulation, diversity and biological roles of terpene

583   metabolism. *Current Opinion in Plant Biology* **9(3)**:297-304 DOI: 10.1016/j.pbi.2006.03.014.

584   Van de Peer Y, Mizrachi E, Marchal K. 2017. The evolutionary significance of polyploidy.

585   *Nature Review Genetics* **18(7)**:411–424 DOI: 10.1038/nrg.2017.26.

586

587   Wallace I.M, Sullivan O, Higgins D G, Notredame C. 2006. M-Coffee: combining multiple

588   sequence alignment methods with T-Coffee *Nucleic Acids Research* **34(6)**:1692-1699 DOI:

589   10.1093/nar/gkl091.

590

591   Waterhouse A, Bertoni MB, Studer S, Tauriello G, Gumienny R, Heer FT, de Beer TA, Rempfer

592   C, Bordoli L, Lepore R, Schwede T. 2018. SWISS-MODEL: homology modelling of protein

593   structures and complexes. *Nucleic Acids Research* 46(W1): W296-W303.

594

595   Wink M. 2003. Evolution of secondary metabolites from an ecological and molecular

596   phylogenetic perspective. *Phytochemistry* **64**:3–19 DOI: 10.1016/s0031-9422(03)00300-5

597

598   Xu S, Brockmöller T, Navarro-Quezada A, Kuhl H, Gase K, Ling Z, Zhou W, Kreitzer C, Stanke

599   M, Tang H, Lyons E, Pandey P, Pandey SP, Timmermann B, Gaquerel E, Baldwin IT. 2017.

600   Wild tobacco genomes reveal the evolution of nicotine biosynthesis. *Proceedings of the National*

601   *Academy of Sciences USA* 114:6133LP-LP6138 DOI: 10.1073/pnas.1700073114.

602

603 Yang Z, Nielsen R. 2000. Estimating synonymous and nonsynonymous substitution rates under

604 realistic evolutionary models. *Molecular Biology and Evolution* **17(1)**:32–43 DOI:

605 10.1093/oxfordjournals.molbev.a026236.

606 Zhou F., Pichersky E. 2020. The complete functional characterization of the terpene synthase

607 family in tomato. *New Phytologist* **226(5):**1341-1360 DOI: 10.1111/nph.16431.

608

609

Figure 1

Figure 1. Phylogenetic analysis of TRI and TRII protein sequences

Phylogenetic analysis of TRI and TRII protein sequences. TR-I and TR-II define Group I and II, respectively, in Solanaceae (A). Substitution rates (ω) of *TR-II* (B). Three duplication events in the genome of *Datura stramonium* are indicated (diamonds).
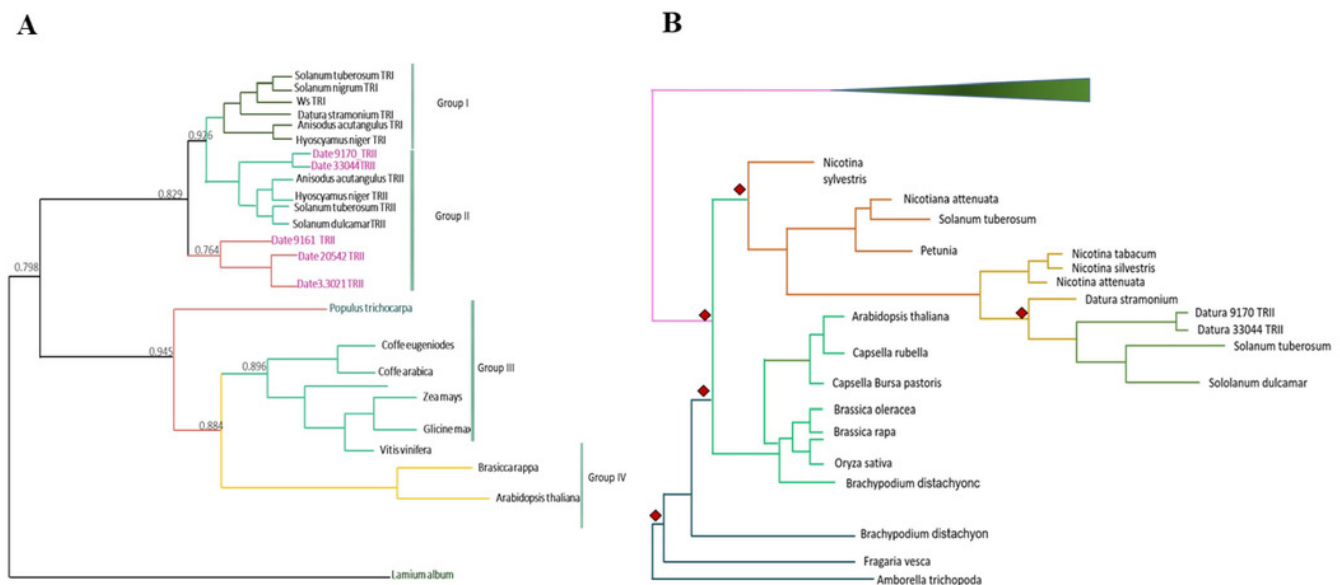
Figure 2

Figure 2. Structure of domains of TRI

Phylogenetic analysis and structure of domains of TRI gene in different species. Six copies are distributed in different clades. Tic23dati33027 has an extra domain adh_short_C.
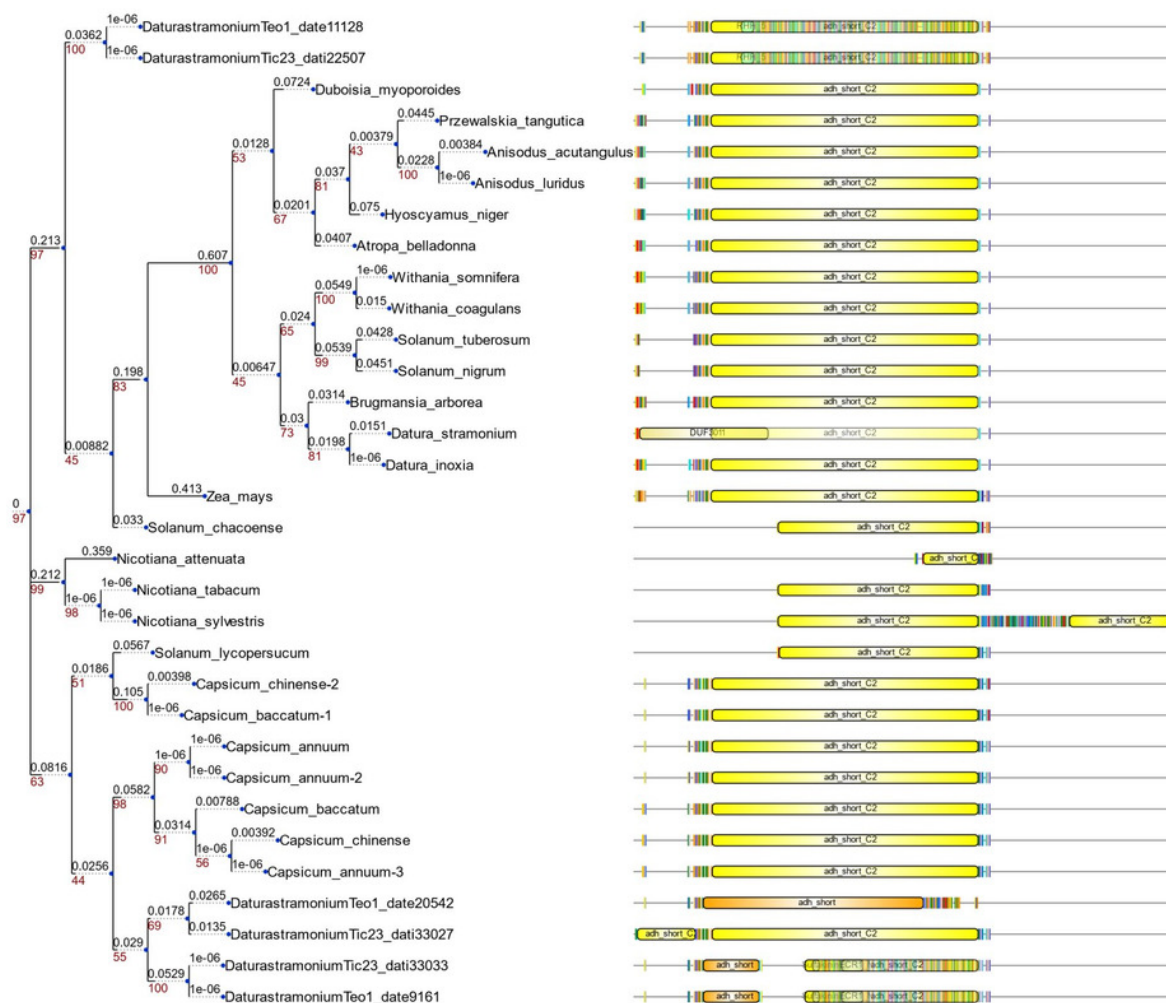
Figure 3

Phylogeny of gene H6H

Phylogeny of gene H6H in 19 species. In A we observe two copies of the gene are present in the two sequenced genomes of *D. stramonium* (Tic and Teo). In Tic8550 there is a tandem duplication of dominion DIOX N (PF14226) vs. Teo 19488; two copies of the gene are present in both genomes (segmental duplication) distributed in different clades. B. Alignment of sequences of protein H6 with the conserved zones highlighted. In the terminal carboxyl, the DIOX-N dominion is duplicated in TIC8550.
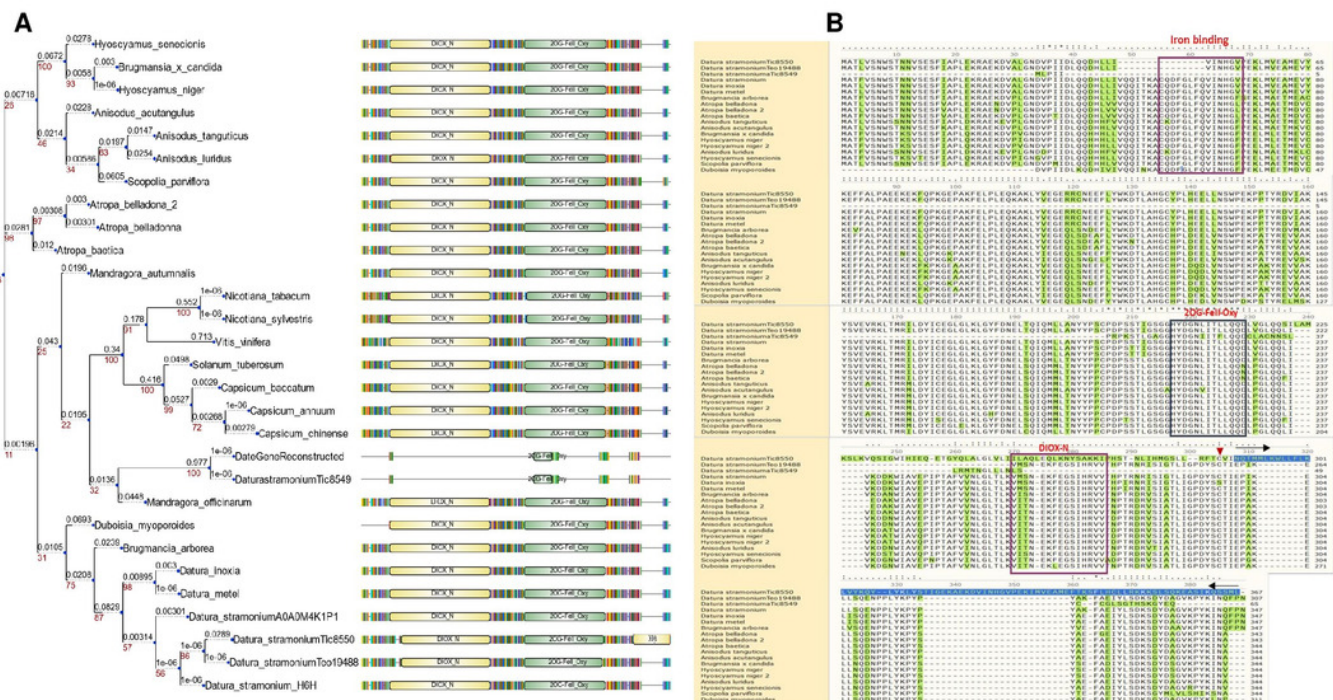
Figure 4

Phylogeny of pmt

Phylogeny of Putrescine N-methyltransferase (*pmt*). date = *Datura stramonium* Teotihuacán, dati = *Datura stramonium*Ticumán. The genomic analyses indicate an expansion of *pmt*, two copies in date and five in dati. The latter has an additional dominion of spermine synthase.
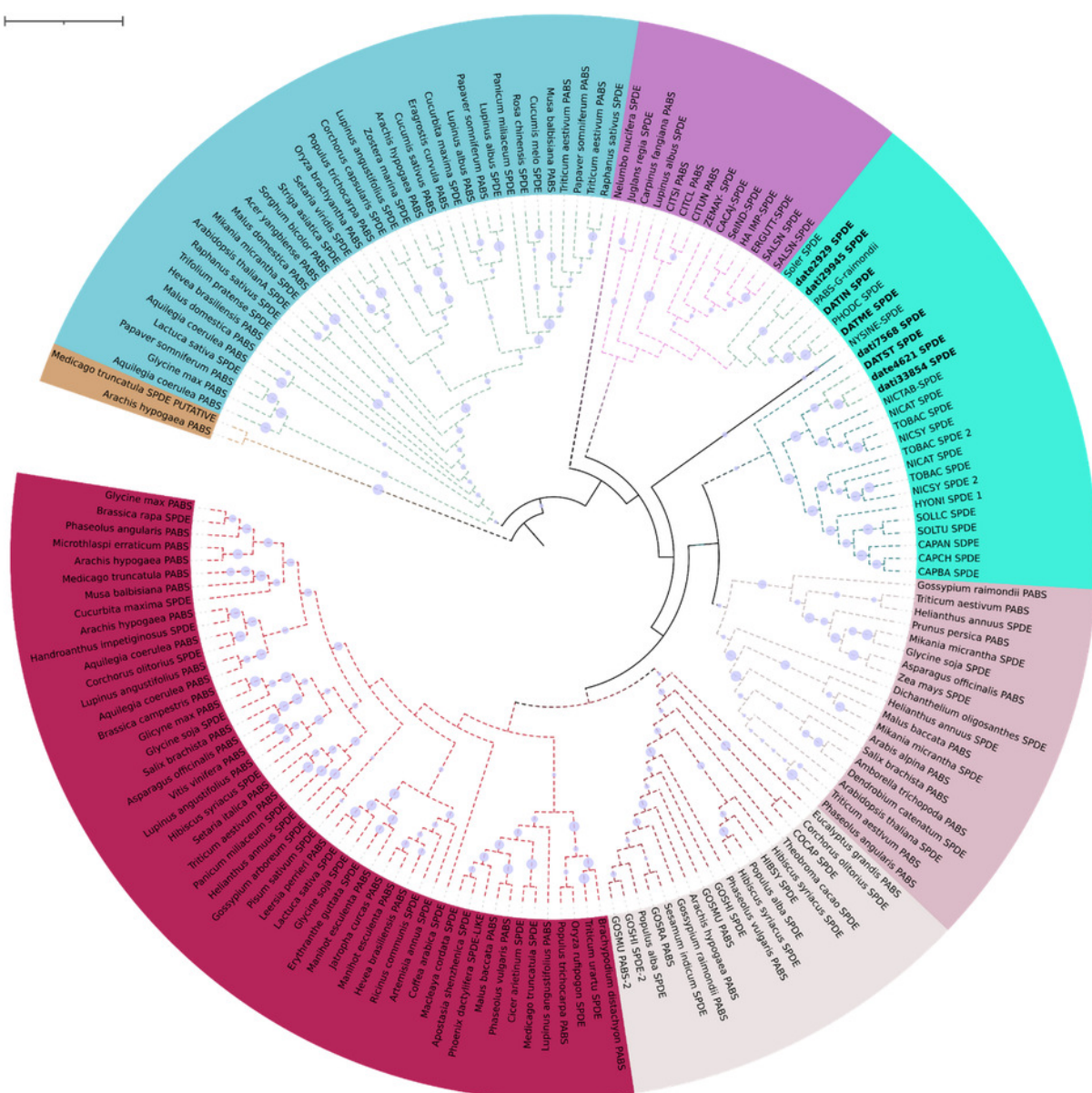
Figure 5

Phylogeny of TPS

Phylogeny of TPSs. 18 TPSs were found in *D. stramonium* (white letters) and Solanaceae and other angiosperms. Two copies of TPS-10 and TPS-13 in the subfamily TPSa. Tree obtained by Bayesian inference, with JTT+G+F evolutionary model. Most branches have a bootstrap support of 1.
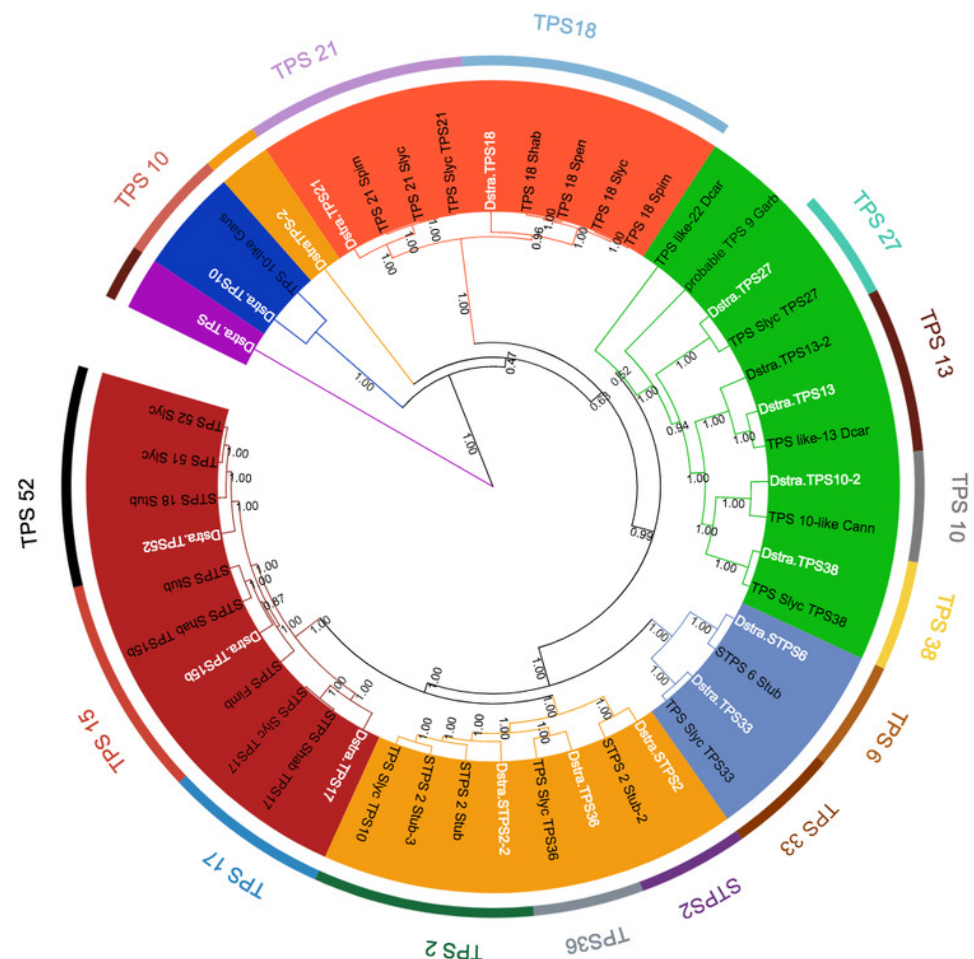
Figure 6

Phylogeny of TPS-10

Phylogeny of TPS-10 (green) of *Datura stramonium*, *D. metel* and other Solanaceae. In these species the TPS-10 is duplicated. Tree obtained by Bayesian inference, with JTT+G+F evolutionary model. Most branches have a support above 80%.
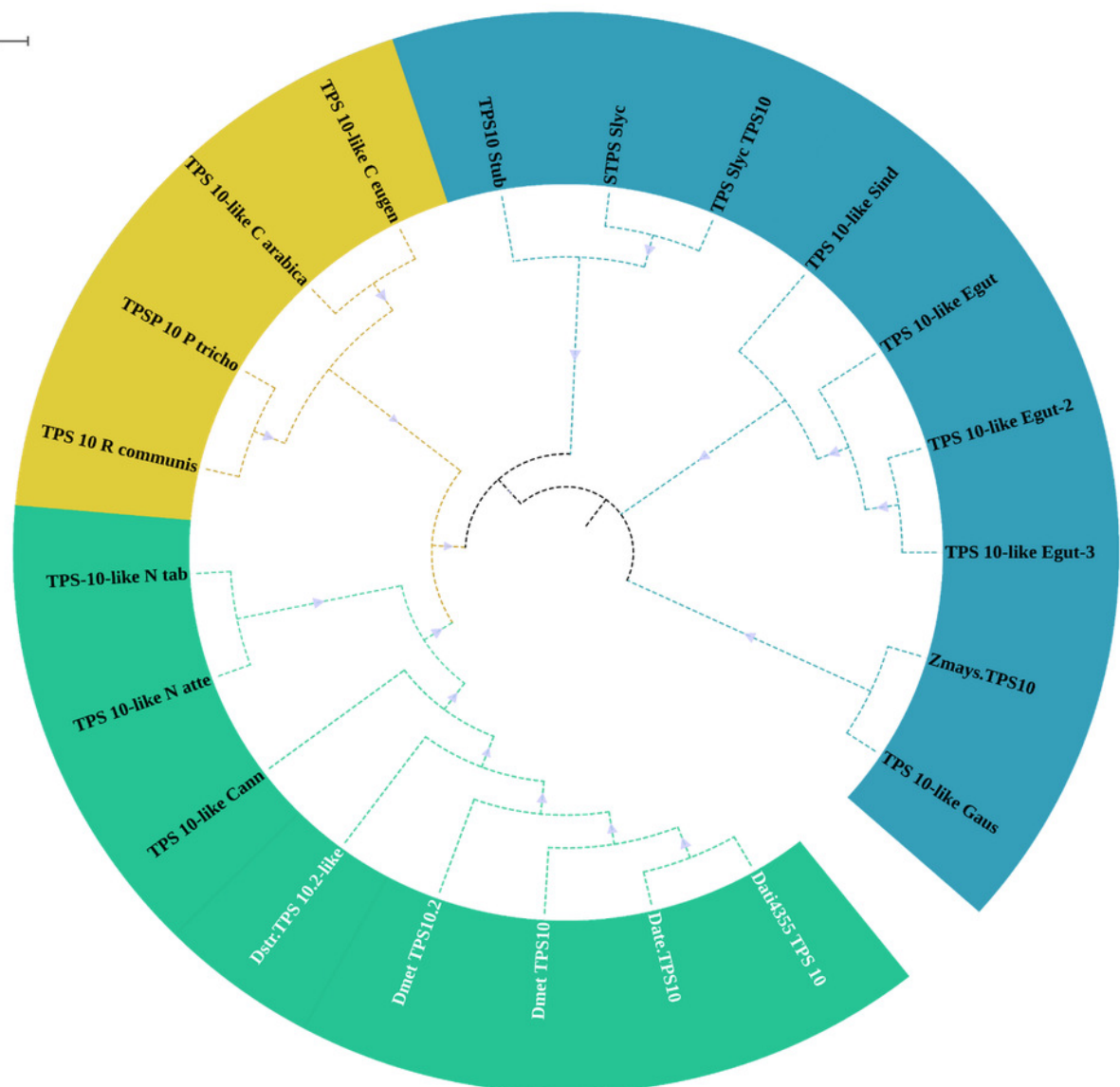
Figure 7

Structure of 6TTM in *Datura metel*

(A) Structure of 6TTM in *Datura metel* (yellow). The ligand (blue) inside the BP is surrounded by amino acids (gray). 8B) Structure modelled for *Datura stramonium*Teo_19488 (cyan) and (C) *D. stramonium* Tic_8550. The interactions between the predicted residues of the BP and the ligand Hyo are illustrated in D, E and F.
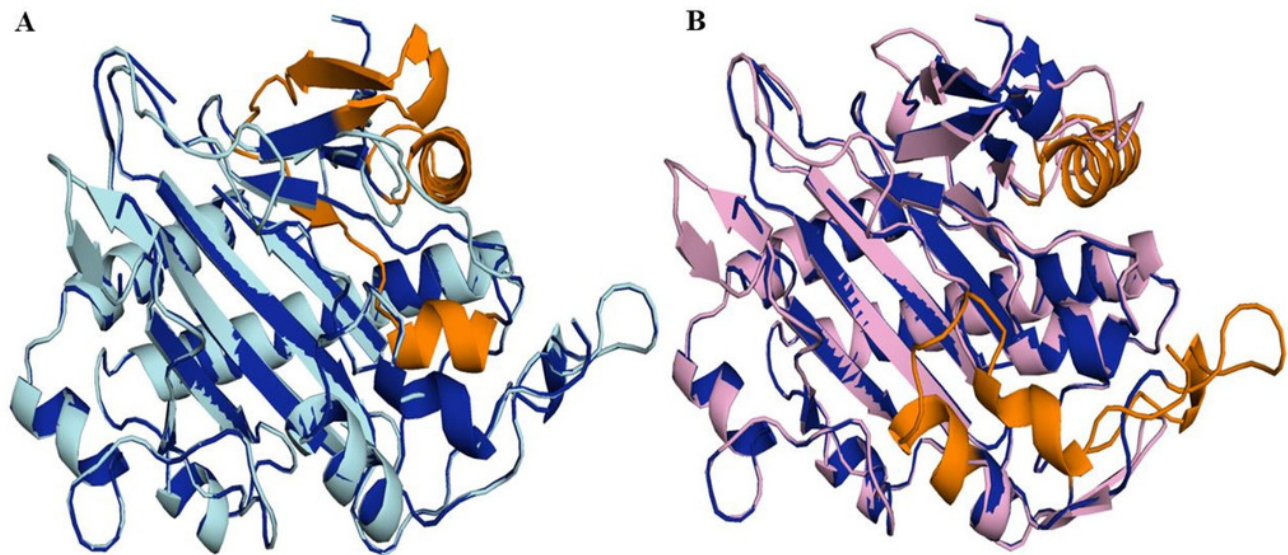
Figure 8

Models of H6H

Superimposition of reported structure of DmH6H (PDB ID 6TTM) (blue) and models obtained for the sequence Ticu8549 (yellow) (A) and a fragment of the sequence Ticu8550 (green)(B). Both sequences possess structural similarity with other sequences of the same protein. The fragment Ticu8549 (yellow) aligns in the region where residues of the active site of H6H are located (gray) and involved in the coordination of the structure with metals (C). This site is highly conserved in the enzyme families ODD. (D) surface of fragments Ticu8549 (Yellow) and Tic8550_Fragment (green) superimposed on H6H structure. In red are the residues that compose the active (binding) site.

Figure 9

Structure of 6TTM

(A) Structure of 6TTM in *Datura metel* (yellow). The ligand (blue) inside the BP is surrounded by amino acids (gray). B) Structure modelled for *Datura stramonium*Teo_19488 (cyan) and (C) *D. stramonium* Tic_8550. The interactions between the predicted residues of the BP and the ligand Hyo are illustrated in D, E and F.
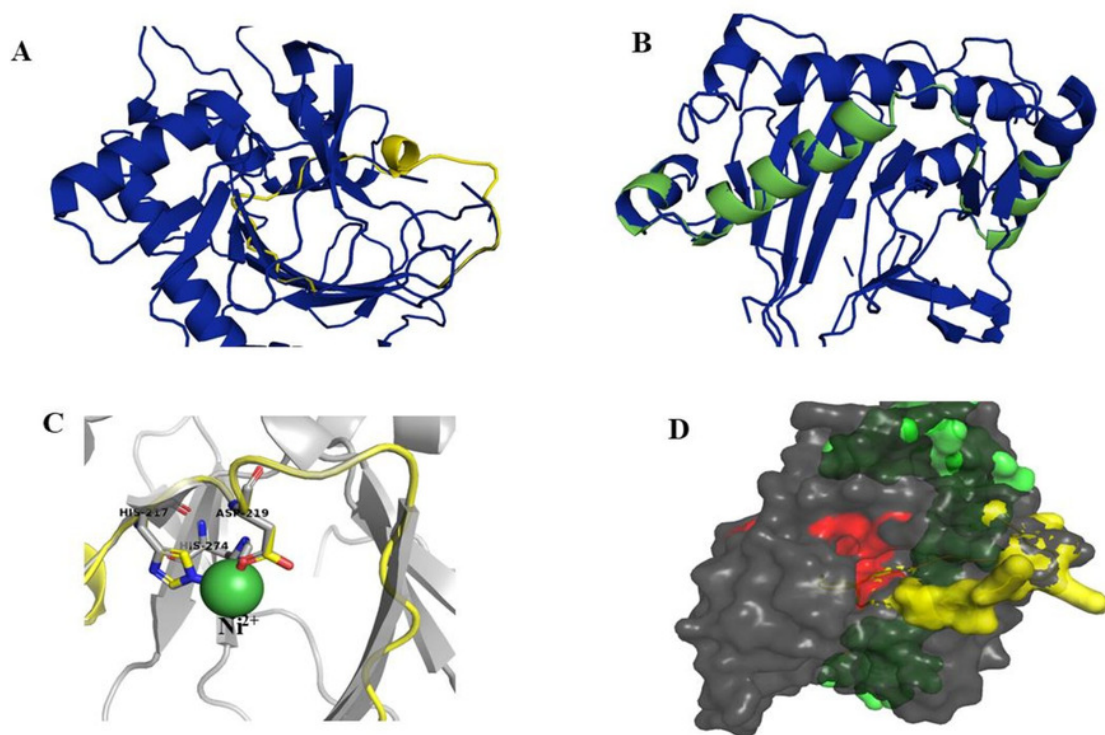
Table 1

Genomes of reference employed in this study.

Table 1. Genome of Solanaceae employed in this study

Specie	Link	Link
Nicotiana tabacum	https://solgenomics.net/organism/Nicotiana_tabacum/genome	https://solgenomics.net/tools/blast/
Nicotiana sylvestris	https://solgenomics.net/organism/Nicotiana_sylvestris/genome	https://solgenomics.net/tools/blast/
Nicotiana attenuata	https://solgenomics.net/organism/Nicotiana_attenuata/genome	https://solgenomics.net/tools/blast/
Nicotiana tomentosiformis	https://solgenomics.net/organism/Nicotiana_tomentosiformis/genome	https://solgenomics.net/tools/blast/
Solanum pimpinellifolium	https://solgenomics.net/organism/Solanum_pimpinellifolium/genome	https://solgenomics.net/tools/blast/
Solanum lycopersicum	https://solgenomics.net/organism/Solanum_lycopersicum/genome	https://solgenomics.net/tools/blast/
Solanum pennellii	https://solgenomics.net/organism/Solanum_pinnelii/genome	https://solgenomics.net/tools/blast/
Solanum tuberosum	https://solgenomics.net/organism/Solanum_tuberosum/genome	https://solgenomics.net/tools/blast/
Capsicum annuum	https://solgenomics.net/organism/Capsicum_annuum/genome	https://solgenomics.net/tools/blast/

Table 2

Selection test for tropane alkaloids (TAs) and terpene syntahses (TPSs) genes of *Datura stramonium*.

Genes	Chi-Square	Proportion of sites selected	ω
PMT	5.43432	0.1389	2.5778
TRI	0	NA	
TRII	31.5335	0.09994	10.6356
H6H	15.66146	0.15136	6.469
TPS10_1	22.75614	0.07163	1.704
TPs10_2	16.16614	0.014589	2.295
TPS14	4.67532	0.098812	3.7809
TPS21	1.43981	NA	
GGPS1	3.89631	0.0899	2.5988