DNA Research, 2022, 29(5), 1–8
https://doi.org/10.1093/dnares/dsac031
Advance Access Publication Date: 18 August 2022
Resource Article: Genomes Explored

OXFORD

Resource Article: Genomes Explored

Chromosomal-level genome of velvet bean (*Mucuna pruriens*) provides resources for L-DOPA synthetic research and development

Shijie Hao[1,2†], Qijin Ge[1†], Yunchang Shao[3,4†], Benqin Tang[3], Guangyi Fan[1,4], Canyu Qiu[4], Xue Wu[3], Liangwei Li[1], Xiaochuan Liu[1], Chengcheng Shi[1]*, and Simon Ming-Yuen Lee[3]*

[1]BGI-Qingdao, BGI-Shenzhen, Qingdao 266555, China, [2]College of Life Sciences, University of Chinese Academy of Sciences, Beijing 101408, China, [3]State Key Laboratory of Quality Research in Chinese Medicine, Institute of Chinese Medical Sciences, University of Macau, Macao 999078, China, and [4]BGI-Shenzhen, Shenzhen 518083, China

*To whom correspondence should be addressed. Tel: +86 15986611742 (C.S), Tel: +853 8822 4695 (S.M.-Y.L.). Email: shichengcheng@genomics.cn (C.S.); simonlee@um.edu.mo (S.M.-Y.L.)

†These authors contributed equally to this work.

Received 6 June 2022; Editorial decision 15 August 2022; Accepted 17 August 2022

Abstract

Mucuna pruriens, commonly called velvet bean, is the main natural source of levodopa (L-DOPA), which has been marketed as a psychoactive drug for the clinical management of Parkinson's disease and dopamine-responsive dystonia. Although velvet bean is a very important plant species for food and pharmaceutical manufacturing, the lack of genetic and genomic information about this species severely hinders further molecular research thereon and biotechnological development. Here, we reported the first velvet bean genome, with a size of 500.49 Mb and 11 chromosomes encoding 28,010 proteins. Genomic comparison among legume species indicated that velvet bean speciated ~29 Ma from soybean clade, without specific genome duplication. Importantly, we identified 21 polyphenol oxidase coding genes that catalyse L-tyrosine to L-DOPA in velvet bean, and two subfamilies showing tandem expansion on Chr3 and Chr7 after speciation. Interestingly, disease-resistant and anti-pathogen gene families were found contracted in velvet bean, which might be related to the expansion of polyphenol oxidase. Our study generated a high-quality genomic reference for velvet bean, an economically important agricultural and medicinal plant, and the newly reported L-DOPA biosynthetic genes could provide indispensable information for the biotechnological and sustainable development of an environment-friendly L-DOPA biosynthesis processing method.

Key words: medicinal plant, *Mucuna pruriens*, chromosome-level assembly, gene family evolution, polyphenol oxidase genes

1. Introduction

Velvet bean (*Mucuna pruriens*, $2n = 22$) is a tropical Fabaceae plant mostly distributed in Asia, Africa and America.[1] It is a versatile potential source of nutrition and cover crop. Velvet bean seeds contain 20.2–29.3% crude protein,[2] which is comparable with the legumes *Cajanus cajan*[3] and *Cicer arietinum*.[4] As a legume cover crop, velvet bean has the characteristic of nitrogen fixation, and its fast-growing could improve biomass and soil quality for farming.[5,6]

Despite its obvious advantages as a crop, velvet bean is best known as an important herbal medicine; it is a renowned and indispensable natural source of levodopa (L-DOPA),[2,7,8] which is widely used for the treatment of mental disorders. Velvet bean extracts can relieve Parkinson's disease (PD).[9–11] Moreover, other unknown compounds in velvet bean could improve motor, olfactory, mitochondrial and synaptic impairments in PD models.[12] Natural source of L-DOPA from velvet bean seed powder exhibit comparable efficacy, but surprisingly a faster onset, than standard synthetic L-DOPA/carbidopa in PD treatment.[13] Additionally, velvet bean is effective for treating male infertility,[14,15] neutralizing snake venom[16] and improving sleep quality.[17]

L-DOPA is the oxidation product of L-tyrosine and is eventually decarboxylated into dopamine.[8,18] Identifying the relevant enzymes in this process and their coding sequences could provide insights aiding the industrial development and production of L-DOPA, and agricultural improvement of velvet bean. However, in contrast to the widespread application and daily use of velvet bean, genetic and molecular information about this species is scarce. No clear genetic information or large-scale molecular research on velvet bean was available before 2017. The first transcriptomic study of two accessions of velvet bean identified several differential expressed transcripts, such as MYB, MADS, WRKY and bHLH, related to secondary metabolite biosynthesis.[19] That report, however, did not clearly address the pathway underlying L-DOPA's metabolism. In this study, we reported a chromosome-level genome for velvet bean, investigated its genomic evolutionary history and identified genes involved in L-DOPA synthesis, with the goal of paving the way for further in-depth research aiding velvet bean breeding.

2. Materials and methods

2.1. Sample collection and library preparation

The velvet bean (*M. pruriens* var. utilis) plants were collected from Napo County, Baise, Guangxi Province, China, in November 2017 and identified by Dr Bao-jing Li, Yunnan University of Chinese Medicine, Kunming, China. A voucher specimen (No. 2017110601A) was deposited in the State Key Laboratory of Quality Research in Chinese Medicine, Institute of Chinese Medical Sciences, University of Macau, Macao, China. Genomic DNA of velvet bean leaf was extracted with QIAGEN Genomic kit following manufacturer's protocol. After fragmentation, we concentrated different sizes of DNA fragments (350, 2,000 and 5,000 bp) with MGIEasy DNA purification kit and checked the library quality with Agilent 2100 Bioanalyzer. After sequencing adapter ligation, the libraries were amplified and sequenced with BGISEQ500 platform. We prepared the Hi-C libraries following a Hi-C protocol.[20] Briefly, the fresh velvet bean leaves were shredded and cross-linked with 2% formaldehyde followed by digesting DNA with a restriction enzyme and biotin-labelling the ends of fragments. Then, fragmented DNA was ligated and sheared, and the biotin-labelled fragments were enriched with streptavidin beads and used to build sequencing library with MGI CoolMPS sequencing kit following manufacturer's protocol too.

2.2. Genome assembly

All the reads were sequenced on the BGISEQ-500 platform. The SOAP*denovo* v2.04[21] (parameters: -K 63 -R -k 41 -F) was used to assemble the scaffolds, followed by two rounds of gap closing, conducted by using GapCloser v1.12[22] (default settings). For Hi-C assembly, data were first processed by HiC-Pro v2.8.0[23] (default settings), and Juicer v1.5[24] (default settings) was then used to map reads onto the genome to validate available pair reads. The 3D-DNA pipeline v180922[25] (default settings) was used to construct the chromosome-level genome.

2.3. Genome annotation

Repeat sequences were detected with homology-based and *de novo* approaches, respectively. The RepeatMasker v4.0.7 program (http://www.repeatmasker.org/, 10 July 2020, date last accessed, parameters: -nolow -no_is -norna -engine ncbi -parallel 1) was used to search for repetitive sequences based on the Repbase v21.01 library. Tandem repeat elements were detected using the TRF program v4.04[26] (parameters: 2 7 7 80 10 50 2000 -d -h). The RepeatModeler v1.03 (http://www.repeatmasker.org/RepeatModeler/, 10 July 2020, date last accessed, parameters: -engine ncbi -database mydb -pa) and LTR_finder v1.04[27] (default settings) were used to build a *de novo* repeat library spanning the whole-genome sequence. The long terminal repeat (LTR) pairs (5′ and 3′ terminal sequences) were extracted and aligned with MUSCLE v3.8.31[28] (default settings), and sequence distances (K) were then calculated by the distmat in the EMBOSS v6.6.0 package (http://www.bioinformatics.nl/cgi-bin/emboss/help/distmat, 1 February 2021, date last accessed, parameter: -nucmethod 2). The insertion time of LTR elements was estimated according to the formula $T = K/2r$, where r refers to the evolution ratio, with 7e−9 per site per year.[29]

Gene models were predicted with *de novo* predictions, homologue (*Glycine max*, *Phaseolus vulgaris*, *Medicago truncatula* and *Vigna unguiculata*) and RNA-seq data. *De novo* predictions were made using GlimmerHMM v3.0.4[30] (parameters: -d arabidopsis -f -g) and Augustus v3.2.1 software[31] (parameters: –species=arabidopsis –uniqueGeneId=true –noInFrameStop=true –gff3=on –strand=both). Gene sets from homologous species were first mapped to velvet bean genome sequences using BLAT v0.36[32] (parameters: -q=prot -t=dnax -noHead); then, coding phases were determined using GeneWise v2.4.1[33] (default settings). RNA-seq data were mapped to the genome with HISAT2 v2.1.0[34] (default settings), and gene models were then predicted by Braker v2.0[35] (default settings). Finally, we used GLEAN v1.0.1[36] (default settings) to integrate the three kinds of evidences. Finally, BUSCO v3.0.2[37] (parameters: -c 10 -e 0.001 -l embryophyta_odb9 -sp arabidopsis) was used to evaluate the genome assembly and gene models.

2.4. Genomic comparative analysis

The coding sequences from seven species (*M. pruriens*, *Arabidopsis thaliana*, *M. truncatula*, *G. max*, *P. vulgaris*, *V. unguiculata* and *Oryza sativa*) were aligned with the all-vs-all BLASP method embedded in blastall package v2.2.26 (parameters: -p blastp -m 8 -e 1e-5 -F F). Then, gene families were clustered by OrthoMCL v2.0.9[38] (parameter: –mode 3). Single-copy genes were extracted, and a taxonomy tree was then constructed using PhyML v3.0[39] (default settings) and TreeBeST v 1.9.2[40] (default settings) based on the global alignment result using MUSCLE v3.8.31 (default settings) with *O. sativa* as outgroup. Divergence time was estimated by MCMCTree v4.4[41] (parameters: clock=3, model=0, BDparas=1 1 0). Gene families' expansion and contraction were calculated by cafe tool v4.2.1[42] (-f Mucunapruriens -n Mucunapruriens='M.pruriens'). Whole-genome synteny blocks were detected using the Python version of MCScan package v0.7.4 [https://github.com/tanghaibao/jcvi/wiki/MCscan (Python-version), 17 March 2021, date last accessed, parameter: jcvi.formats.gff bed –type=mRNA, -m

jcvi.graphics.dotplot –nostdpf –genomenames]. The pairwise blocks were extracted, and then 4-fold degenerate synonymous site (4DTv) values were calculated and counted to estimate the whole-genome duplication (WGD) events.

2.5. Study of L-DOPA synthetic genes

All annotated gene models were first searched against the Kyoto Encyclopedia of Genes and Genomes (KEGG) database, producing a unique gene annotation for each gene. All genes annotated as polyphenol oxidase were recognized as candidate L-DOPA synthetic gene members (KEGG pathway ko00350).[43] Genes trees were constructed by FastTree v 2.1.3[44] based on the global alignment of MUSCLE v3.8.31 (default settings) and visualized by ggtree package v2.4.1.[45] The RNA-seq data from different tissues (leaf, pod, bean, stem and root) were mapped to the gene models using bowtie v2.[46,47] (default settings), and the fragments per kilobase of exon model per million mapped fragments values were then calculated for each gene and each sample to represent the expression level. The distance value between all gene pairs was calculated using distmat program (parameter: -nucmethod 2) in EMBOSS package v3.0.2. For the six velvet bean cultivars that grow in different zones and with different seed colour traits, a library of 350 bp was constructed for each line and sequenced on the BGISEQ500 platform. We also used SOAPdenovo v2.04 (parameters are the same as described in Section 2.2) for draft assembly. The gene prediction and annotation were performed with the same pipeline as described above.

2.6. Identification of R-genes

Domain models containing resistance-related sites or domains were acquired from the Pfam database[48] (Leucine Rich Repeat: PF00560/PF07723/PF07725/PF12799; NB-ARC: PF00931; NB-LRR: PF12061; Pkinase: PF00069; TIR: PF01582; BED zinc finger: PF02892). Then hmmsearch v3.1b2[49] (default settings) was used to search gene sets with R-gene models.

3. Results and discussion

3.1. Genome assembly and annotation

A total of 50.11 Gb (~100×) whole-genome data were produced for assembly (Supplementary Table S1). We assembled a genome with a total length of 500.49 Mb. Using 73 Gb Hi-C data, we anchored ~470 Mb scaffolds (~93.83%) to 11 chromosomes, with a final scaffold N50 of 48.43 Mb, and contig N50 of 92.52 kb, respectively (Table 1, Fig. 1A and B). A public RNA-seq data[19] was mapped to the genome with a high alignment ratio of 95.57%. We predicted 28,010 coding genes in velvet bean genome, of which 94.92% have been functionally annotated with Non-Redundant (NR), KEGG, UniProt and Gene Ontology (GO) database. BUSCO evaluation results revealed that about 93.5% and 96.4% of complete embryophyte BUSCO genes could be covered, respectively by our genome and genes models, for high completeness of the velvet bean genome.

Table 1. Summary of velvet bean genome assembly

	Value
Total size of scaffold (Mb)	500.49
Total size of contig (Mb)	494.05
Anchored scaffold (Mb)	470.15 (93.83%)
Scaffold N50 (Mb)	48.43
Contig N50 (kb)	92.53
GC content (%)	32.21
Number of genes	28,010
Repeat contents (%)	39.75
BUSCO evaluation of genome (%)	93.5
BUSCO evaluation of genes (%)	96.4
All functional annotated (%)	94.92
NR (%)	94.61
Swissprot (%)	78.11
KEGG (%)	72.55
TrEMBL (%)	94.63
Interpro (%)	91.09
GO (%)	68.25

3.2. Repetitive elements in velvet bean genome

We detected about 39.75% (~199 Mb) repetitive sequences in the velvet bean genome (Supplementary Table S2), which is larger than that of *Lotus japonicus* (~31.85%) and smaller than that of *Arachis duranensis* (~44.62%), *V. unguiculata* (~46.37), *G. max* (~47.57%) and *P. vulgaris* (~50.50%) and almost equal to *M. truncatula* (~39.67%) (Supplementary Table S2). The percentages of long terminal repeats (LTR) in entire transposable element (TE) sequences were also different among species, with the LTR percentage of the velvet bean being ~74.82% (Supplementary Table S2). In velvet bean, *A. duranensis* and *C. arietinum* genomes, we found a distinct common peak in the insert time distribution of the LTR, suggesting that one recent LTR burst event occurred around 3 Ma. Meanwhile, no peaks were detected in the *G. max*, *P. vulgaris*, *L. japonicus* or *M. truncatula*, suggesting successive LTR bursting events (Supplementary Fig. S1).

3.3. Comparative genomic analysis

Genes from seven species, including *O. sativa*, *A. thaliana* and four legumes (*G. max*, *P. vulgaris*, *M. truncatula* and *V. unguiculata*), together with the velvet bean genome, were aligned and clustered into 24,940 families, and velvet bean occupied 17,098 families. Using 1,213 conserved single copy genes of seven species, phylogenic tree and their speciation time were estimated. The results revealed that velvet bean is a sister clade of soybean (*G. max*), common bean (*P. vulgaris*) and cowpea (*V. unguiculata*), and diverged from those three species ~29 Ma (Fig. 2A).

Expansion and contraction of gene families could be important events during the species' evolution. We found that velvet bean genome contained 838 expanded families but more contracted families compared with the other three Phaseoleae legumes (Supplementary Fig. S2). The expanded families contained 3,073 genes, and KEGG enrichment analysis suggested that they are significantly related with 28 pathways, including ABC transporters (ko02010, *P*-value: 7.42e−20), pentose and glucuronate interconversions (ko00040, *P*-value: 6.00e−19), isoquinoline alkaloid biosynthesis (ko00950, *P*-value: 3.79e−16) and tyrosine metabolism (ko00350, *P*-value: 8.08eE−14). Interestingly, L-DOPA is involved in the tyrosine metabolism pathway. Besides, the contracted families contained 3,163 genes in velvet bean which significantly enriched in 26 pathways, including ribosome biogenesis in eukaryotes (ko03008, *P*-value: 1.03e−15), linoleic acid metabolism (ko00591, *P*-value: 5.69e−15) and plant–pathogen interaction (ko04626, *P*-value: 8.65e−15). Both expanded and contracted genes were distributed in several common pathways; however, the gene number in the environmental adaption pathway of contracted families is much more than that of expanded

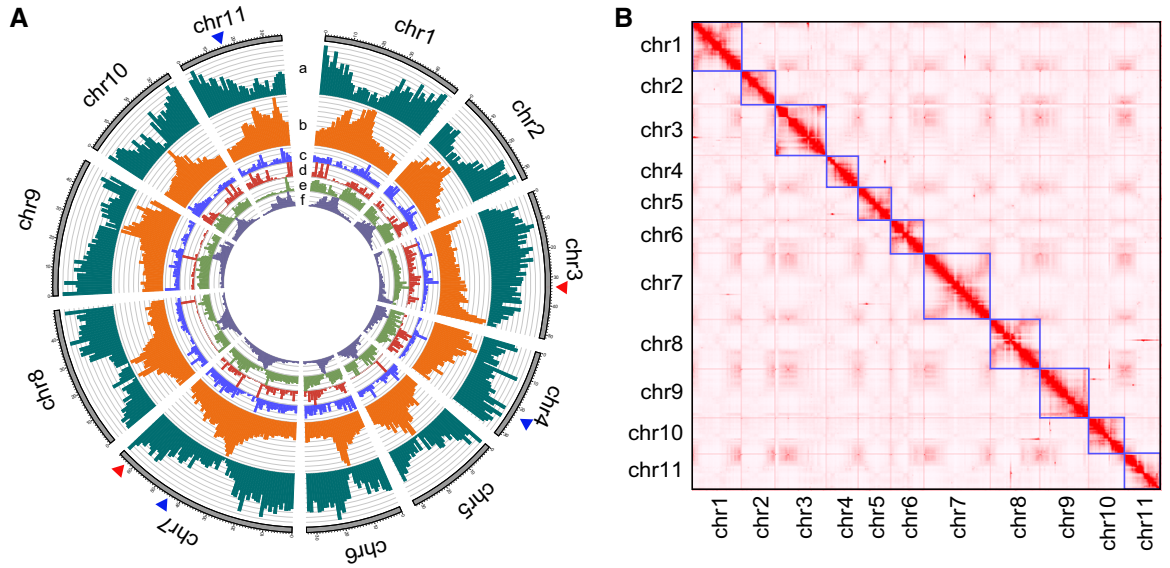

Figure 1. Genomic overview. (A) Genomic overview of velvet bean (*Mucuna pruriens*). (a) Gene model density. (b) Repeat sequence density. (c–f) Density of LINE, SINE, DNA transposon and LTR elements. The values were calculated within a 1-Mb window. Red triangles denote the positions of tandem duplicated polyphenol oxidase genes and the blue triangles are the remaining genes. (B) Heatmap of the Hi-C assembly (A color version of this figure appears in the online version of this article).

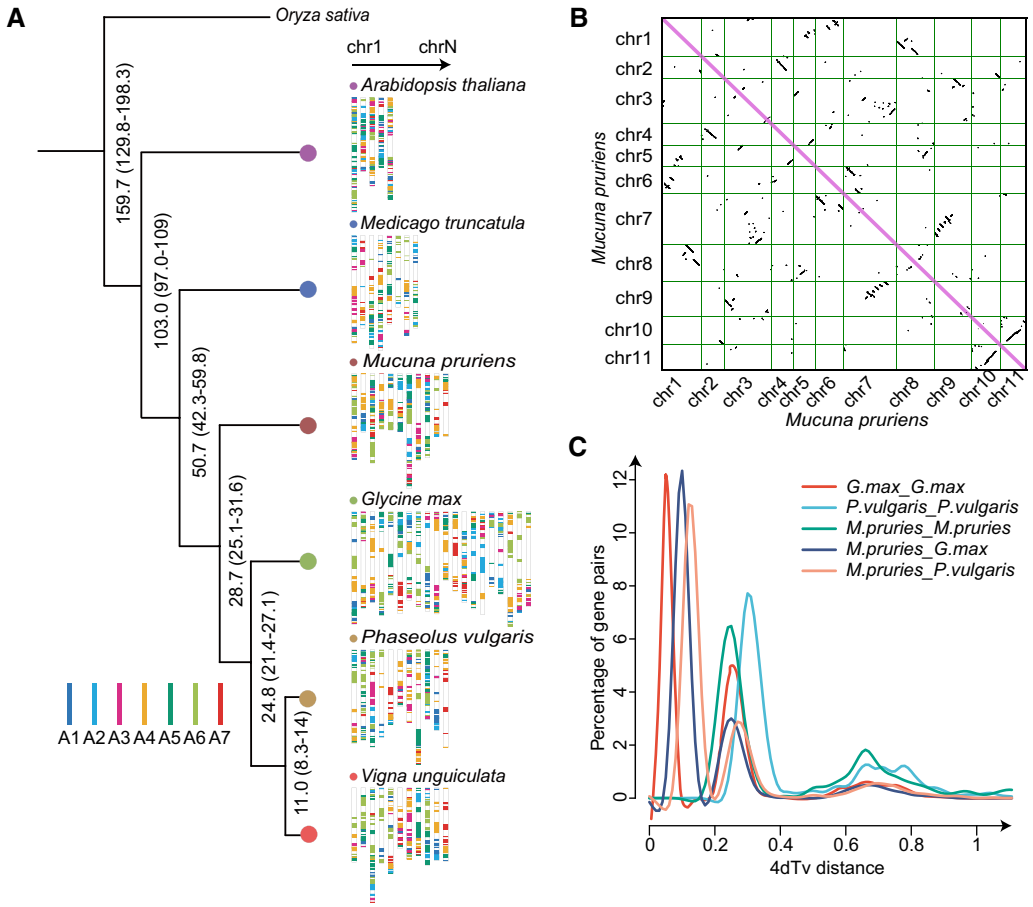

Figure 2. Comparative genomic analysis and WGD among legume genomes. (A) Estimated phylogenetic divergence time and chromosome evolution model of legume species. (B) Whole-genome synteny within velvet bean chromosomes. (C) 4DTv distributions of velvet bean, common bean, soybean, velvet bean vs. common bean and velvet bean vs. soybean.

gene families. In addition, there are 237 families unique to velvet bean genome. Enrichment analysis of these specific gene families showed the genes are associated with 13 pathways, including protein processing in the endoplasmic reticulum (ko04141, *P*-value 1.11e−6), selenocompound metabolism (ko00450, *P*-value 2.74e−5) and diterpenoid biosynthesis (ko00904, *P*-value 3.08e−5) (Supplementary Fig. S3).

3.4. WGD events and chromosome synteny

WGD plays an important role in the genome evolution and chromosome evolvement of plants. Most eudicots shared an ancient whole-genome triplication (gamma WGT) event that occurred ∼125 Ma; for leguminous plants, a Fabaceae-specific duplication event occurred ∼55 Ma in addition to the gamma WGT.[50,51] Within velvet bean genome, at least 12 obvious self-syntenic segments were found (e.g. chr1, chr5, chr6 and chr8; Fig. 2B). Synteny between velvet bean and common bean, as well as velvet bean and cowpea, indicated obvious collinear correspondences (Supplementary Fig. S4). To date, no other sequenced legume has been shown to experience a recent species-specific WGD event, except soybean.[50] We calculated the 4DTv distribution to make inferences regarding the WGD events. The 4DTv distribution of velvet bean showed similarities with cowpea and common bean and experienced an ancestral hexaploidization event and Fabaceae shared WGD; there was no species-specific WGD (Fig. 2C). This is consistent with the number of chromosomes, which is almost twice as high for soybean as velvet bean, cowpea and common bean, resulting from *G. max* experiencing an extra WGD event. The gamma WGT event changed the chromosome number of ancestral dicotyledons from 7 to 21.[51] We reconstructed the chromosome evolutionary process of velvet bean by detecting collinearity between the legumes and grape, which is regarded as a representative species of ancestral eudicot karyotype.[52] Based on the components of the karyotype in velvet bean, we assumed that at least four fissions and fusions occurred in each chromosome on average. For example, chr2 is derived from ancestral chromosomes A1, A2, A4 and A5 (Fig. 2A).

3.5. Expansion and evolution of L-DOPA synthetic genes

Tyrosinase (TYR, EC: 1.14.18.1), tyrosine 3-monooxygenase (TH, EC: 1.14.16.2) and polyphenol oxidase (EC: 1.10.3.1, also called catechol oxidase) are involved in the process of transforming L-tyrosine into L-DOPA.[53,54] We identified 21 polyphenol oxidase genes in velvet bean, 7 in cowpea, 21 in soybean, 4 in barrelclover (*M. truncatula*) and 3 in common bean, respectively (Fig. 3A and Supplementary Table S3). We also looked into the number of polyphenol oxidase genes in the assemblies of six velvet bean cultivars grown in different regions with different seed colour traits (Supplementary Table S4). By assembling the whole-genome sequencing data (Section 2.5), we obtained six draft assemblies ranging from 405 to 500 Mb in size, and the scaffold N50 values ranged from 45.16 to 58.09 kb. There are 12–14 polyphenol oxidase genes identified in the six lines. The reason why the gene number is smaller than the chromosome version is that the sequences of the draft assemblies were relatively fragmented, and some genes were not annotated (Fig. 3C and D). These results demonstrate that the polyphenol

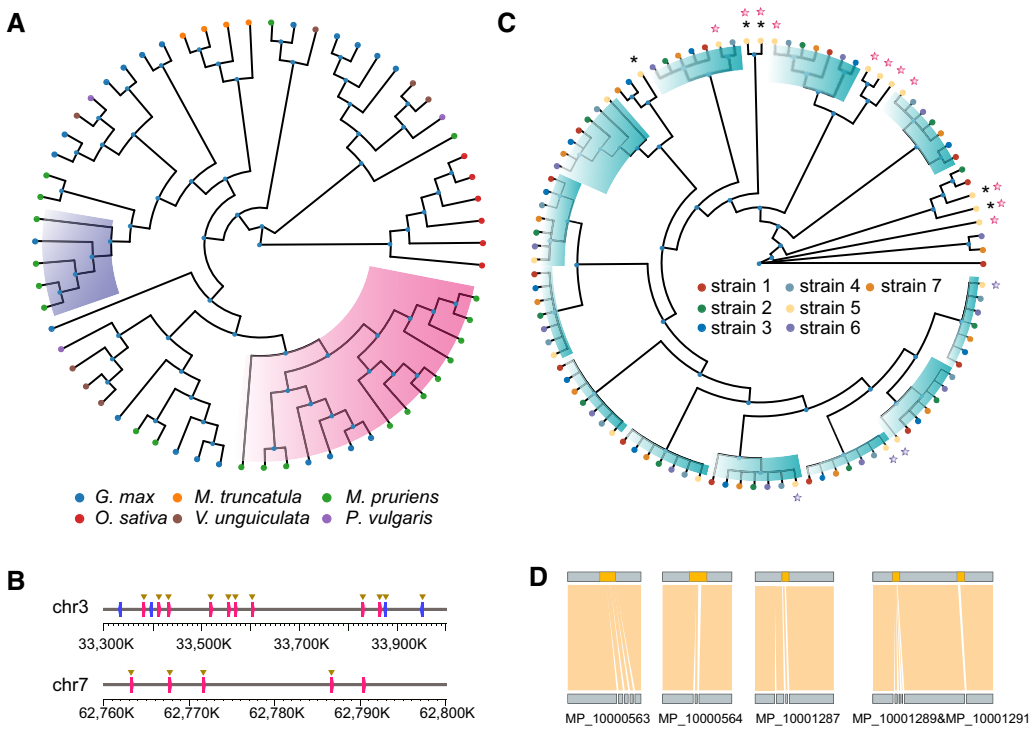

Figure 3. Expansion of L-DOPA polyphenol oxidase genes in velvet bean genome. (A) Gene phylogenetic tree of five legume genomes. (B) Tandem duplicated locations of 18 polyphenol oxidase genes on chr3 and chr7. The genes represented by triangles are clustered into one clade in Fig. 1A (eight on chr3 with pink background and four on chr7 with lavender background in A). (C) Gene phylogenetic tree of seven velvet beans; the common genes shared by seven cultivars are marked by the green background. The pink and lavender symbols correspond to the tandem polyphenol oxidase genes in A. (D) Validation of five polyphenol oxidase genes marked with black '*' in C in other velvet bean cultivars. Strain5 represents the chromosomal-level assembly, and genes denoted by '*' are complete in Strain5, and present in non-chromosomal-level strains with fragmented sequences (A color version of this figure appears in the online version of this article).

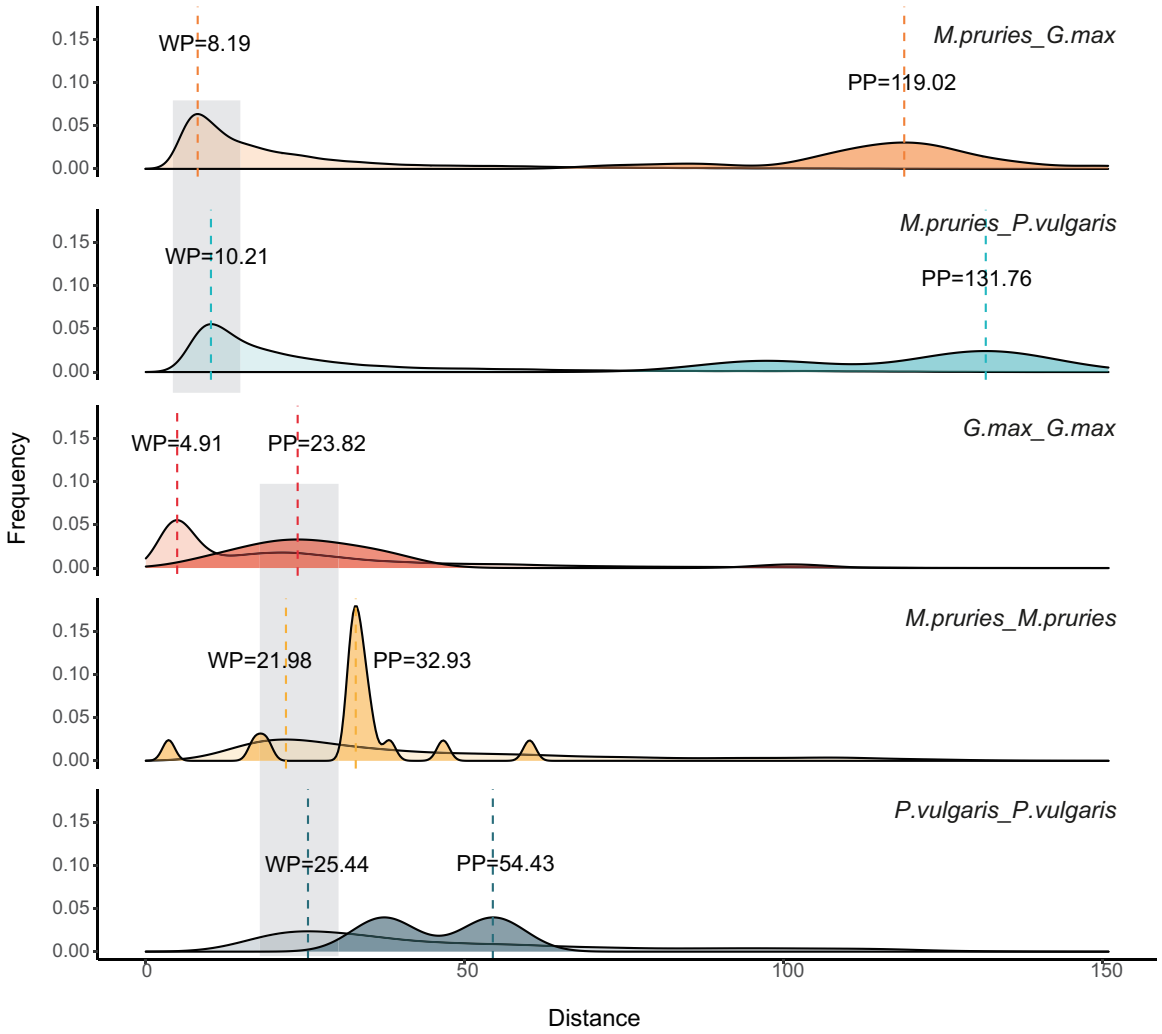

Figure 4. The distance distribution of whole-genome gene pairs and L-DOPA polyphenol oxidase gene pairs in legumes. Pairs in velvet bean, common bean and soybean genome as well as velvet bean vs. common bean and velvet bean vs. soybean are shown. Two grey background areas spanning top2 and bottom3 figures mark the divergence between velvet bean and soybean/common bean, and Fabaceae shared WGD events, respectively. Dotted lines indicate the peak values of each line. WP: whole-genome gene pairs, transparent colour. PP: polyphenol oxidase gene pairs, regular colour.

oxidase gene is indeed expanded in the velvet bean genome. In addition, as supported by RNA-seq data, we identified all annotated polyphenol oxidase genes expressed in velvet bean tissues (Supplementary Fig. S5), suggesting that they play a role in the progress of L-DOPA synthesis in velvet bean. Our findings are in line with a recent study reporting that L-DOPA synthesis in velvet bean is catalysed by polyphenol oxidase (EC: 1.10.3.1), where no tyrosinase gene or tyrosine 3-monooxygenase (EC: 1.14.16.2) gene was found in the transcriptome data.[18]

The polyphenol oxidase gene tree of five legumes showed two notably expanded subfamilies in velvet bean (Fig. 3A). These genes are tandemly distributed on chr3 and chr7 (Figs 1A and 3B), indicating the tandem expansion of polyphenol oxidase genes in the velvet bean genome. To investigate the polyphenol oxidase genes' insertion time, we compare the polyphenol oxidase gene distance (PP values) and whole-genome orthologous/paralogous distance (WP values), which indicates speciation and WGD events (Fig. 4). We found that PP values between two species were much larger than the WP value, indicating that the polyphenol oxidase genes have already existed in their ancestral genome. At the same time, the PP value within the species is similar to or even slightly larger than the Fabaceae WGD distance value (Fig. 4, e.g. *M.pruries_M.pruries*: PP 32.93 > WP 21.98), so we speculate that the L-DOPA synthase is likely to expand with the Fabaceae WGD event, but during the process of speciation, these genes are retained in the velvet bean genome, while other legumes such as common bean may have lost them.

The accumulation of L-DOPA in the seed of velvet bean likely involves the regulation of upstream and downstream genes. We also investigated the gene members of tyrosine decarboxylase (EC: 4.1.1.25), an enzyme that can degrade L-DOPA. Interestingly, we found that velvet bean genome not only contained more polyphenol oxidase genes (EC: 1.10.3.1) but also fewer tyrosine decarboxylase genes than other legumes (13 in velvet bean and 20–63 in soybean/cowpea/common bean/barrelclover, Supplementary Table S3). We

believe that our results provide a foundation for future genetic studies of velvet bean, for example concerning the gene activity of polyphenol oxidase in different life stages and conditions, and the synthesis regulation and transport system of L-DOPA.

3.6. Contraction of R-genes

Resistance genes (R-genes) play a vital role in plant defence against biotic stresses.[55] We identified R-genes in velvet bean and six other species including soybean, *Arabidopsis*, *Medicago*, common bean, cowpea and rice (*O. sativa*), based on the conserved domains of the R-genes. Compared with the other six species, we found a notable contraction of R-genes in velvet bean, with 433 R-genes in velvet bean (the smallest number among legumes), 486 in *Arabidopsis* (the smallest number among non-legume species) and 649 in common bean (Supplementary Table S5). Nucleotide-binding leucine-rich repeat domains (NBS-LRR) genes, which sense pathogen invasion of plants,[56] showed less gene number than other species, with approximately one-quarter of that in soybean. Interestingly, we found that 45.56% (190/417) of R-genes of velvet bean were located in ancient chromosome blocks, which was much more than the proportions of 30.28% (295/974), 24.46% (147/601), 21.17% (151/713) and 14.50% (176/1214) for soybean, common bean, cowpea and *Medicago*, respectively. This suggests that the R-genes in other legumes probably expanded much more than those in velvet bean after speciation. Given that velvet bean shows strong adaptation to various environments and the accumulation of L-DOPA (an allelochemical substance that protects plants from competitor and invader),[7,57–59] we speculated that a compensation mechanism might account for the reduced number of R-genes in velvet bean genome; that is, the high levels of L-DOPA that prevented an invasion by pathogens of velvet bean might be a factor in its small number of NBS-LRR genes.

4. Conclusions

Mucuna pruriens is not only an excellent additive for the food industry but also an important natural material for extracting L-DOPA for the treatment of PD. This study is the first to report the chromosome-level assembly of *M. pruriens* along with the genomic evolutionary analysis and a description of essential enzyme coding genes for L-DOPA biosynthesis. These genomic results will aid further genetic study and development of *M. pruriens* and L-DOPA biosynthesis.

Acknowledgements

Research at University of Macau was funded by the Shenzhen-Hong Kong-Macao Science and Technology Innovation Project (Category C) (Ref No: EF038/ICMS-LMY/2021/SZSTIC), Shenzhen Science and Technology Innovation Committee and the Science and Technology Development Fund, Macau SAR (File no. 0016/2019/AKP) and University of Macau (MYRG2022-00263-ICMS and CPG2022-00023-ICMS). We thank the China National GeneBank for supporting the study.

Accession number

PRJNA862264

Conflict of interest

The authors declare no conflict of interest.

Data availability

All the data that support this project including whole-genome sequencing data and genome resources have been deposited at NCBI with the Project ID PRJNA862264 and CNGB Sequence Archive (CNSA, https://db.cngb.org/cnsa/) with the Project ID CNP0002743.

Supplementary data

Supplementary data are available at *DNARES* online.

References

1. Janardhanan, K., Gurumoorthi, P. and Pugalenthi, M. 2003, Nutritional potential of five accessions of a South Indian tribal pulse, *Mucuna pruriens* var utilis, *Trop. Subtrop. Agroecosyst.*, **1**, 141–52.
2. Vadivel, V. and Janardhanan, K. 2000, Nutritional and anti-nutritional composition of velvet bean an under-utilized food legume in South India, *Int. J. Food Sci. Nutr.*, **51**, 279–87.
3. Saxena, K.B., Kumar, R.V. and Sultana, R. 2010, Quality nutrition through pigeonpea—a review, *Health*, **2**, 1335–44.
4. Wallace, T.C., Murray, R. and Zelman, K.M. 2016, The nutritional value and health benefits of chickpeas and hummus, *Nutrients*, **8**, 766.
5. Ceballos, A.I.O., Rivera, J.R.A., Arce, M.M.O. and Valdivia, C.P. Velvet bean (*Mucuna pruriens* var. utilis) a cover crop as bioherbicide to preserve the environmental services of soil. In: Alvarez-Fernandez R. ed. *Herbicides—Environmental Impact Studies and Management Approaches*. London: IntechOpen, 2012.
6. Blanchart, E., Villenave, C., Viallatoux, A., et al. 2006, Long-term effect of a legume cover crop (*Mucuna pruriens* var. utilis) on the communities of soil macrofauna and nematofauna, under maize cultivation, in southern Benin, *Eur. J. Soil Biol.*, **42**, S136–44.
7. Soares, A.R., Marchiosi, R., Siqueira-Soares Rde, C., Barbosa de Lima, R., Dantas dos Santos, W. and Ferrarese-Filho, O. 2014, The role of L-DOPA in plants, *Plant Signal. Behav.*, **9**, e28275.
8. Lampariello, L.R., Cortelazzo, A., Guerranti, R., Sticozzi, C. and Valacchi, G. 2012, The magic velvet bean of *Mucuna pruriens*, *J. Tradit. Complement. Med.*, **2**, 331–9.
9. Cilia, R., Laguna, J., Cassani, E., et al. 2018, Daily intake of *Mucuna pruriens* in advanced Parkinson's disease: a 16-week, noninferiority, randomized, crossover, pilot study, *Parkinsonism Relat. Disord.*, **49**, 60–6.
10. Cassani, E., Cilia, R., Laguna, J., et al. 2016, *Mucuna pruriens* for Parkinson's disease: low-cost preparation method, laboratory measures and pharmacokinetics profile, *J. Neurol. Sci.*, **365**, 175–80.
11. Srivastav, S., Fatima, M. and Mondal, A.C. 2017, Important medicinal herbs in Parkinson's disease pharmacotherapy, *Biomed. Pharmacother.*, **92**, 856–63.
12. Poddighe, S., De Rose, F., Marotta, R., et al. 2014, *Mucuna pruriens* (velvet bean) rescues motor, olfactory, mitochondrial and synaptic impairment in PINK1B9 *Drosophila melanogaster* genetic model of Parkinson's disease, *PLoS One*, **9**, e110802.
13. Katzenschlager, R., Evans, A., Manson, A., et al. 2004, *Mucuna pruriens* in Parkinson's disease: a double blind clinical and pharmacological study, *J. Neurol. Neurosurg. Psychiatry*, **75**, 1672–7.
14. Shukla, K.K., Mahdi, A.A., Ahmad, M.K., Shankhwar, S.N., Rajender, S. and Jaiswar, S.P. 2009, *Mucuna pruriens* improves male fertility by its action on the hypothalamus-pituitary-gonadal axis, *Fertil. Steril.*, **92**, 1934–40.
15. Shukla, K.K., Mahdi, A.A., Ahmad, M.K., Jaiswar, S.P., Shankwar, S.N. and Tiwari, S.C. 2010, *Mucuna pruriens* reduces stress and improves the quality of semen in infertile men, *Evid. Based Complement. Alternat. Med.*, **7**, 137–44.
16. Kumar, A., Gupta, C., Nair, D.T. and Salunke, D.M. 2016, MP-4 contributes to snake venom neutralization by *Mucuna pruriens* seeds through an indirect antibody-mediated mechanism, *J. Biol. Chem.*, **291**, 11373–84.

17. McCarthy, C.G., Alleman, R.J., Bell, Z.W. and Bloomer, R.J. 2012, A dietary supplement containing *Chlorophytum borivilianum* and velvet bean improves sleep quality in men and women, *Integr. Med. Insights*, **7**, 7–14.
18. Singh, S.K., Dhawan, S.S., Lal, R.K., Shanker, K. and Singh, M. 2018, Biochemical characterization and spatio-temporal analysis of the putative L-DOPA pathway in *Mucuna pruriens*, *Planta*, **248**, 1277–87.
19. Sathyanarayana, N., Pittala, R.K., Tripathi, P.K., et al. 2017, Transcriptomic resources for the medicinal legume *Mucuna pruriens*: *de novo* transcriptome assembly, annotation, identification and validation of EST-SSR markers, *BMC Genomics*, **18**, 409.
20. Sandoval-Velasco, M., Rodríguez, J.A., Perez Estrada, C., et al. 2020, Hi-C chromosome conformation capture sequencing of avian genomes using the BGISEQ-500 platform, *GigaScience*, **9**, giaa087.
21. Luo, R., Liu, B., Xie, Y., et al. 2012, SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler, *Gigascience*, **1**, 18.
22. Bannar-Martin, K.H., Kremer, C.T., Ernest, S.K.M., et al. 2018, Integrating community assembly and biodiversity to better understand ecosystem function: the Community Assembly and the Functioning of Ecosystems (CAFE) approach, *Ecol. Lett.*, **21**, 167–80.
23. Servant, N., Varoquaux, N., Lajoie, B.R., et al. 2015, HiC-Pro: an optimized and flexible pipeline for Hi-C data processing, *Genome Biol.*, **16**, 259.
24. Durand, N.C., Shamim, M.S., Machol, I., et al. 2016, Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments, *Cell Syst.*, **3**, 95–8.
25. Dudchenko, O., Batra, S.S., Omer, A.D., et al. 2017, *De novo* assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds, *Science*, **356**, 92–5.
26. Benson, G. 1999, Tandem repeats finder: a program to analyze DNA sequences, *Nucleic Acids Res.*, **27**, 573–80.
27. Xu, Z. and Wang, H. 2007, LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons, *Nucleic Acids Res.*, **35**, W265–268.
28. Edgar, R.C. 2004, MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Res.*, **32**, 1792–7.
29. Hu, T.T., Pattyn, P., Bakker, E.G., et al. 2011, The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change, *Nat. Genet.*, **43**, 476–81.
30. Majoros, W.H., Pertea, M. and Salzberg, S.L. 2004, TigrScan and GlimmerHMM: two open source *ab initio* eukaryotic gene-finders, *Bioinformatics*, **20**, 2878–9.
31. Stanke, M. and Morgenstern, B. 2005, AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints, *Nucleic Acids Res.*, **33**, W465–467.
32. Kent, W.J. 2002, BLAT—the BLAST-like alignment tool, *Genome Res.*, **12**, 656–64.
33. Birney, E., Clamp, M. and Durbin, R. 2004, GeneWise and Genomewise, *Genome Res.*, **14**, 988–95.
34. Kim, D., Langmead, B. and Salzberg, S.L. 2015, HISAT: a fast spliced aligner with low memory requirements, *Nat. Methods*, **12**, 357–60.
35. Hoff, K.J., Lange, S., Lomsadze, A., Borodovsky, M. and Stanke, M. 2015, BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS, *Bioinformatics*, **32**, 767–9.
36. Elsik, C.G., Mackey, A.J., Reese, J.T., Milshina, N.V., Roos, D.S. and Weinstock, G.M. 2007, Creating a honey bee consensus gene set, *Genome Biol.*, **8**, R13.
37. Simao, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. and Zdobnov, E.M. 2015, BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs, *Bioinformatics*, **31**, 3210–2.
38. Li, L., Stoeckert, C.J., Jr. and Roos, D.S. 2003, OrthoMCL: identification of ortholog groups for eukaryotic genomes, *Genome Res.*, **13**, 2178–89.
39. Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W. and Gascuel, O. 2010, New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0, *Syst. Biol.*, **59**, 307–21.
40. Postuma, R.B., Anang, J., Pelletier, A., et al. 2017, Caffeine as symptomatic treatment for Parkinson disease (Cafe-PD): a randomized trial, *Neurology*, **89**, 1795–803.
41. Yang, Z. 1997, PAML: a program package for phylogenetic analysis by maximum likelihood, *Comput. Appl. Biosci.*, **13**, 555–6.
42. De Bie, T., Cristianini, N., Demuth, J.P. and Hahn, M.W. 2006, CAFE: a computational tool for the study of gene family evolution, *Bioinformatics*, **22**, 1269–71.
43. Saranya, G., Jiby, M.V., Jayakumar, K.S., Padmesh Pillai, P. and Jayabaskaran, C. 2020, L-DOPA synthesis in *Mucuna pruriens* (L.) DC. is regulated by polyphenol oxidase and not CYP 450/tyrosine hydroxylase: an analysis of metabolic pathway using biochemical and molecular markers, *Phytochemistry*, **178**, 112467.
44. Price, M.N., Dehal, P.S. and Arkin, A.P. 2010, FastTree 2—approximately maximum-likelihood trees for large alignments, *PLoS One*, **5**, e9490.
45. Yu, G., Lam, T.T., Zhu, H. and Guan, Y. 2018, Two methods for mapping and visualizing associated data on phylogeny using Ggtree, *Mol. Biol. Evol.*, **35**, 3041–3.
46. Langdon, W.B. 2015, Performance of genetic programming optimised Bowtie2 on genome comparison and analytic testing (GCAT) benchmarks, *BioData Min.*, **8**, 1.
47. Langmead, B. and Salzberg, S.L. 2012, Fast gapped-read alignment with Bowtie 2, *Nat. Methods.*, **9**, 357–9.
48. Finn, R.D., Bateman, A., Clements, J., et al. 2014, Pfam: the protein families database, *Nucleic Acids Res.*, **42**, D222–230.
49. Johnson, L.S., Eddy, S.R. and Portugaly, E. 2010, Hidden Markov model speed heuristic and iterative HMM search procedure, *BMC Bioinformatics*, **11**, 431.
50. Van de Peer, Y., Mizrachi, E. and Marchal, K. 2017, The evolutionary significance of polyploidy, *Nat. Rev. Genet.*, **18**, 411–24.
51. Salse, J. 2012, *In silico* archeogenomics unveils modern plant genome organisation, regulation and evolution, *Curr. Opin. Plant Biol.*, **15**, 122–30.
52. Jaillon, O., Aury, J.-M., Noel, B., et al. 2007, The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla, *Nature*, **449**, 463–7.
53. Duckworth, H.W. and Coleman, J.E. 1970, Physicochemical and kinetic properties of mushroom tyrosinase, *J. Biol. Chem.*, **245**, 1613–25.
54. Ali, S. and Rizvi, N. 2014, Microbiological transformation of L-tyrosine to L-dopa from methanol pretreated biomass of a novel *Coriolus versicolor* under submerged culture, *Appl. Biochem. Biotechnol.*, **172**, 2041–54.
55. Zhang, R., Zheng, F., Wei, S., et al. 2019, Evolution of disease defense genes and their regulators in plants, *IJMS*, **20**, 335.
56. DeYoung, B.J. and Innes, R.W. 2006, Plant NBS-LRR proteins in pathogen sensing and host defense, *Nat. Immunol.*, **7**, 1243–9.
57. Guidotti, B.B., Gomes, B.R., Siqueira-Soares Rde, C., Soares, A.R. and Ferrarese-Filho, O. 2013, The effects of dopamine on root growth and enzyme activity in soybean seedlings, *Plant Signal. Behav.*, **8**, e25477.
58. de Cassia Siqueira-Soares, R., Soares, A.R., Parizotto, A.V., de Lourdes Lucio Ferrarese, M. and Ferrarese-Filho, O. 2013, Root growth and enzymes related to the lignification of maize seedlings exposed to the allelochemical L-DOPA, *ScientificWorldJournal*, **2013**, 134237.
59. Rehr, S.S., Janzen, D.H. and Feeny, P.P. 1973, L-DOPA in legume seeds: a chemical barrier to insect attack, *Science*, **181**, 81–2.