The Chromosome-level Genome of *Dracaena cochinchinensis* Provides Insights into its Biological Features and the Mechanism of Dragon's Blood Formation

Yanhong Xu^{1,6}, Kaijian Zhang^{4,6}, Zhonglian Zhang^{3,6}, Yang Liu^{1,6}, Feifei Lv², Peiwen Sun¹, Shixi Gao¹, Qiuling Wang¹, Cuicui Yu¹, Jiemei Jiang¹, Chuangjun Li⁵, Meifang Song³, Zhihui Gao¹, Chun Sui¹, Haitao Li³, Yue Jin¹, Xinwei Guo¹, Jianhe Wei^{1,2}

¹ Key Laboratory of Bioactive Substances and Resources Utilization of Chinese Herbal Medicine, Ministry of Education & National Engineering Laboratory for Breeding of Endangered Medicinal Materials, Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100193, China

² Hainan Branch of the Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College, Haikou 570311, China

³ Yunnan Key Laboratory of Southern Medicine Utilization, Yunnan Branch of the Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College, Jinghong 666100, China

⁴ Novogene Bioinformatics Institute, 100083 Beijing, China

⁵ Institute of Materia Medica, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100050, China

⁶ These authors contributed equally to this article.

*Correspondence: Jian-He Wei, wjianh@263.net

SHORT SUMMARY

This study reports a chromosome-level genome assembly for *D. cochinchinensis*, the first genome of *Dracaena* Vand. ex L. It provides valuable genetic resources and creates a large scope for studying *Dracaena*. We found the significant expansion of genes associated with its longevity and slow growth, and genes in flavonoids

biosynthesis were first completely identified. Moreover, transcription factors bHLH and MYB as the core regulator of flavonoids biosynthesis, and ROS as the specific signal molecule during the process of injury-induced dragon's blood formation were also identified. These results not only deciphered the mystery of its longevity, but also elucidated the molecular mechanism of dragon's blood formation preliminarily, which will facilitate the resource protection and sustainable utilization of *Dracaena*.

ABSTRACT

Dracaena, a remarkably long-lived and slowly maturing species, is famous all over the world for the production of dragon's blood, a precious traditional medicine used by different cultures since ancient times. However, lacking a high-quality genome, the molecular mechanisms underlying these traits are largely unknown and that greatly restricts the protection and regeneration of the rare and endangered plant resources. Here, we sequenced and assembled a chromosome-level genome of the *Dracaena cochinchinensis*, the first to be sequenced of *Dracaena* Vand. ex L. The *D. cochinchinensis* genome covering 1.21 Gb with a scaffold N50 of 50.06 Mb, and encodes 31,619 predicted protein-coding genes. We found *D. cochinchinensis* has undergone two whole genome duplications (WGDs) and two long terminal repeats (LTRs) insertion burst events. The expansion of *cis*-zeatin O-glucosyltransferase (*cZOGT*) and small auxin up-regulated RNA (*SAUR*) genes is account for its longevity and slow growth. In flavonoids biosynthesis pathway, transcription factors bHLH and MYB were predicted as the core regulators, and ROS as the specific signal molecule during the process of injury-induced dragon's blood formation. Our study not only provides high-quality genomic knowledge of *D. cochinchinensis*, but also deciphered the mystery of its longevity, and preliminarily elucidated the molecular mechanism of dragon's blood formation, which will facilitate the resource protection and sustainable utilization of *Dracaena*.

Key words: genome, *Dracaena cochinchinensis*, dragon blood, whole genome duplication, longevity, flavonoids biosynthesis, ROS.

58

59 INTRODUCTION

60 Dragon's blood, the red resin produced by tree species in the genus *Dracaena*
 61 (Asparagaceae), is a precious traditional Chinese medicine, and a folk medicine used
 62 by many nationalities in the world. It has several therapeutic functions of activating
 63 blood circulation and dissipating blood stasis, relieving inflammation and pain,
 64 astringency and hemostasis, etc (Gupta *et al.*, 2008; Fan *et al.*, 2014; Sun *et al.*, 2019).
 65 Besides being a famous medicine, it is also used as the colorant of art works by many
 66 cultures (Zheng *et al.*, 2012). It is called “dragon’s blood” because the resin or sap
 67 excreted by a part of the genus *Dracaena* plants is deep red and looks like the blood
 68 of a legendary dragon when encountered external injury or microbial invasion (Gupta
 69 *et al.*, 2008), and therefore, *Dracaena* has been famous all over the world for its red
 70 resin, and belongs to the cultural heritage of humanity.

71 *Dracaena* species are monocotyledonous evergreen plants mainly distributed in
 72 tropical and subtropical regions of Asia and Africa. They are also a remarkably
 73 long-lived and slowly-growing species that known as "plant longevity star" (Zheng *et*
 74 *al.*, 2012; Tomlinson & Huggett, 2012; Maděra *et al.*, 2018, 2020; Scheiger *et al.*,
 75 2020). The age of a fabulous dragon blood tree was overestimated for several
 76 thousands of years (Humboldt, 1814; Christ *et al.*, 1886; Schenck, 1907). Thereafter,
 77 through the observation of a few largest trees, it was concluded that their ages are
 78 about several hundred years, and the flowering time varies from ten to several decades
 79 (Pütter, 1925; Symon, 1974; Mägdefrau, 1975; Zheng *et al.*, 2012; Maděra *et al.*,
 80 2018). According to the research of Maděra (2018), the juvenile stages of *D.*
 81 *cinnabari* from seedling to the first flowering will take about 200-250 years, even
 82 more (Maděra *et al.*, 2018), demonstrating it is an extremely slowly growth and
 83 slowly maturing species. Due to the difficulties in the authentic study of natural
 84 population aging, several methodologies of age estimation were developed and
 85 improved (Adolt & Pavliš, 2004; Adolt *et al.*, 2012; Jura-Morawiec, 2019; Lengálová
 86 *et al.*, 2020). Anyway, they all showed that the dragon tree is a long-lived species.

87 The origin and identity of dragon's blood have been confused, and there are
 88 many source plants of dragon's blood, such as *Croton*, *Dracaena*, *Daemonorops* and
 89 *Pterocarpus* (Gupta *et al.*, 2008; Ding *et al.*, 2020). In nature, *Dracaena* grows slowly,
 90 and only trees over 30-years-old have the possibility to produce dragon's blood
 91 (Edwards *et al.*, 1997; Wang *et al.*, 2013; Ding *et al.*, 2020). For a long time, due to
 92 the continuous over-exploitation, its natural resources are threatened with depletion.
 93 In 1998, many of the *Dracaena* species were list in the International Union for
 94 Conservation of Nature Red List (IUCN red list, 2017.2). Of them, *D. cochinchinensis*
 95 and *D. Cambodian* species in China were rated as vulnerable (Hubáľková *et al.*, 2015;
 96 Kamel *et al.*, 2018). Soon afterwards, the two species had been listed in the List of
 97 National Key Protected Wild Plants of China, which are the second grade of national
 98 key endangered protection and prohibited to felling. For the endangered status
 99 coupled with limited genetic resources, it is critical to consider adequate measures for
 100 conservation management and the exploitation of dragon's blood.

101 In China, *D. cochinchinensis* is the official original species produced dragon's
 102 blood (Zheng *et al.*, 2009; Luo *et al.*, 2011), which was first discovered by Xitao Cai
 103 in 1972 in Yunnan province and has been used to be the substitute of the traditionally
 104 imported dragon's blood, called Long-Xue-Jie (Chinese dragon's blood). According
 105 to literature research, the dragon's blood has been recorded and applied in China for at
 106 least 1500 years, and the authentic source of Chinese ancient *Dracaena* should be
 107 imported from Africa (Cai & Xu, 1979). With the rising of shipping, the genuine
 108 *Dracaena* has changed from the resin of *Dracaena* in agave family to the red resin of
 109 the fruit of *rattan* in Southeast Asia until the discovery of *D. cochinchinensis* which is
 110 exclusively distributed in China (Cai & Xu, 1979; Zheng *et al.*, 2006; Wang *et al.*,
 111 2011).

112 Flavonoids are the main components of dragon's blood. Flavonoid oligomers,
 113 composing of one dihydrochalcone unit condensed with one or more chalcone,
 114 flavane or homoisoflavane units, constituted the major identified components of
 115 dragon's blood from the genus *Dracaena* (Pang *et al.*, 2016). The oligomeric
 116 flavonoids account for over 50% of the resin from *Dracaena* spp. by weight (Hao et

117 al., 2015b). Up to now, hundreds of flavonoids have been isolated from the resin
118 wood of *D. cochinchinensis* (for review Gupta *et al.*, 2008; Fan *et al.*, 2014, Sun *et al.*,
119 2019). Additionally, terpenoids and steroidal saponins are also effective components
120 of dragon's blood (Fan *et al.*, 2014; Teng *et al.*, 2015; Maděra *et al.*, 2020; Vanickova
121 *et al.*, 2020), which were reported to be cytotoxic and could be used as an
122 anticarcinogen (Darias *et al.*, 1989; Mimaki *et al.*, 1999; González *et al.*, 2003;
123 Hernández *et al.*, 2004; Ding *et al.*, 2020; Maděra *et al.*, 2020).

124 Though there are many studies on the chemical constituents and pharmacological
125 effects of dragon's blood (for review Gupta *et al.*, 2008; Fan *et al.*, 2014; Sun *et al.*,
126 2019; Maděra *et al.*, 2020), few about its formation mechanism and the molecular
127 basis about its biology, even many knowledge gaps in the research of *Dracaena*. In
128 this study, we first report the genome draft of *D. cochinchinensis* at chromosome level,
129 as well as the transcriptome and metabolome analysis were conducted, which is
130 undoubtedly an important step to understand the genetic and molecular basis of its
131 fundamental biology and conservation management.

132 **RESULT**

133 **Genome Sequencing, Assembly and Annotation**

134 The genome survey using a k-mer analysis ($k = 17$) revealed that the genome size of
135 *D. cochinchinensis* was approximately 1.257 Gb, with the heterozygosity of 0.52% and
136 the repeat sequence ratio of 67.8% (Fig. S1; Table S1). Based on the estimated
137 genome size, using a combination of SMRT sequencing technology from PacBio,
138 10xGenomics and short-read sequencing from Illumina platforms, we totally obtained
139 407.06 Gb clean data, which represented x 323.75 coverage (Table S2). The initial
140 assembly is 1.21 Gb, containing 1898 scaffolds and 3042 contigs, with a scaffold N50
141 size of 2.07 Mb and contig N50 size of 1.1 Mb (Table S3). Assessment of the
142 completeness of the genome assembly with CEGMA indicated 96.77% coverage of
143 the conserved core eukaryotic genes, and Benchmarking Universal Single-Copy
144 (BUSCO) (Simao *et al.*, 2015) analyses showed that the genome was 93.6% complete
145 (**Table S4**), suggesting that the genome assembly was relatively complete and of high

quality. Next, ~137.26 Gb high-quality Hi-C data were used to further improved the genome assembly, and 90.26% of the genome were anchored onto 20 chromosomes (2n=40) (**Fig. 1b**; **Fig. S2** ; **Table S5**). The final assembly consists of 1835 scaffolds, which spans 1.21 Gb, with a scaffold N50 size of 50.05 Mb (**Table S6**). The *D. cochinchinensis* chromosomes showed a significant level of synteny with its related species *A. officinalis* (**Fig. S3**).

Based on *de novo* and homology-based predictions and transcriptome data, a total of 31,619 protein-coding genes were predicted, and 31,402 genes (99.3%) were assigned with functions (**Table S7**). In addition, by comparing with the known ncRNA libraries, the ncRNA information of the *D. cochinchinensis* genome of was obtained (**Table S8**), including 1570 transfer RNAs (tRNAs), 1278 ribosomal RNAs (rRNAs), 1576 microRNAs (miRNAs) and 722 small nuclear RNAs (snRNAs).

158 **Genomic Evolution of *D. cochinchinensis***

D. cochinchinensis belongs to the Asparagales clade of monocots. It is close to *A. officinalis*, *D. officinalis* and *A. shenzhenica* that have completed genome sequencing. We clustered the annotated genes into gene families among them plus other additional monocotyledon and dicotyledon species, including *E. guineensis*, *O. sativa*, *A. comosus*, *Arabidopsis*, *Populus*, *A. sinensis*, *H. brasiliensis*, *C. grandis*, *C. papaya*, *E. grandis*, *V. vinifera*, *A. Trichopoda*. Family clustering yielded 31,324 gene families (groups), including 252 single-copy families. CDS-based evolutionary trees were constructed using RAXML software. *D. cochinchinensis* and *A. officinalis*, *D. officinalis* and *A. shenzhenica* are gathered in one branch, respectively, which indicated that *Asparaginaceae* and *Orchidaceae* had an independent diverge long ago (138.4 Mya), while diverge time between *D. cochinchinensis* and *A. officinalis* is 87.7 Mya (**Fig. 2a**).

To investigate the genome expansion in *D. cochinchinensis*, we analyzed whole-genome duplication (WGD) events. 4DTv values was estimated on the basis of the paralogous gene pairs in collinear regions detected in *D. cochinchinensis* and three other representative plant species, *A. officinalis*, *D. officinalis* and *A. shenzhenica*.

The distribution of 4DTv distances in *D. cochinchinensis* showed two peaks at the values of approximately 0.12 and 0.45 (**Fig. 2b**). The first peak shared by *D. cochinchinensis* and *A. officinalis* may correspond to the Asparagales- β event previously identified in the *A. officinalis* genome (Harkess *et al.*, 2017; Li *et al.*, 2020). The second peak at approximately 0.12 indicate that *D. cochinchinensis* underwent another WGD event in dependently after diverging from *A. officinalis* (**Fig. 2b, Fig. S4**). The 29,280 duplicated genes were classified into five different categories: 6864 whole-genome duplicates (WGD duplicates, 23.1%), 1506 tandem duplicates (TD, 5.1%), 1196 proximal duplicates (PD, 4.1%), 7301 transposed duplicates (TRD, 24.9%), and 19613 dispersed duplicates (DSD, 66.7%). We further investigate the duplications of genes involved in flavonoids biosynthesis (i.e., *PAL*, *4CL*, *C4H*, *CHI*, *CHS*, *DFR*, *F3H*, *FLS*, *FAR*, *PPO*, *OMT*), and found some genes were differently duplication, and there were more genes retained by WGD, suggesting the recent WGD event was important to the evolution of flavonoid biosynthesis in *D. cochinchinensis* (**Table S9**).

60.15% (729.11 Mb) of the *D. cochinchinensis* genome comprises repetitive sequences (**Table S10**), similar to *D. officinale* (63.33%) (Yan *et al.*, 2015) and *E. guineensis* (57%) (Singh *et al.*, 2013), but higher than *A. shenzhenica* (42.05%) (Zhang *et al.*, 2017) and *A. comosus* (43.7%) (D'Hont *et al.*, 2012). As in many other sequenced plant genomes (Kim *et al.*, 2014; Teh *et al.*, 2017; Zhang *et al.*, 2017), long terminal repeats retrotransposons (LTR-RTs) were the dominate, accounts for 51.06% of the assembly, with the *Gypsy* (501,175,881bp, 41% of the genome sequence) and *Copia* (72,176,608 bp, 5.9% of the genome sequence). Furthermore, analyzing the insertion time of intact LTR, it is found there are two insertion burst events in *D. cochinchinensis* genome occurred recently and before, about 0.3-0.4 million years and 2 million years, respectively (Fig. 2c), dominated by both autonomous and non-autonomous LTRs (Fig. 2d). In order to investigate the evolution of transposable elements (TEs) in *D. cochinchinensis*, phylogenetic trees of domains in reverse transcriptase genes were constructed for both *Copia* and *Gypsy* superfamily. We found the majority of *Gypsy* was species-specific in *D. cochinchinensis*, while the

205 *Copia* superfamily displayed a different pattern with consisting of elements from all
206 three species (**Fig. S5**).

207 **Comparative Genomic Shows the Adaption of *D. cochinchinensis***

208 Copy numbers of homologous gene families vary greatly among different
209 species, which is caused by the differences in the rates of gene gain and loss. Among
210 the 31,318 gene families inferred to be present in the most recent common ancestor of
211 the 17 species that used for gene family analysis, 81 gene families were found to have
212 undergone expansion in *D. cochinchinensis* relative to the other 16 species (**Fig. 2a**).
213 A total of 730 and 1088 genes in the expanded families were annotated to Gene
214 Ontology (GO) terms and Kyoto Encyclopedia of Genes Genomes (KEGG) pathways,
215 respectively. GO analysis revealed that these expanded orthogroups were related to
216 several metabolism process (**Table S11, Figs, S6-7**), such as “ATP metabolic
217 process”, “defense response” based on the biological process category, and
218 “binding”, “catalytic activity”, “oxidoreductase activity”, “peroxidase activity” based
219 on molecular function category. The enriched KEGG categories were the categories
220 of “plant-pathogen interaction”, “phenylpropanoid biosynthesis”, “zeatin
221 biosynthesis”, “sesquiterpene and triterpenoid biosynthesis”, “flavone and flavonol
222 biosynthesis”, and “photosynthesis”, etc (**Table S12, Fig. S8**). We speculate that these
223 expanded genes are closely related in some way to the function requirements of
224 environmental adaptability of *D. cochinchinensis* and the formation of dragon’s
225 blood.

226 It is noted that 28 genes were found to be under positive selection, which
227 significantly enriched in KEGG terms related to base excision repair (BER),
228 nucleotide excision repair (NER), mismatch repair (MMR) and DNA replication
229 (**Table S13**). They include DNA ligase 1 and DNA excision repair protein which are
230 DNA-repair pathways that play crucial roles in maintaining genomic integrity
231 particularly in response to exposure to environmental genotoxins including
232 ultraviolet and reactive oxygen produced from the harsh rays of the sun (Schuch *et al.*,
233 2017; Nisa *et al.*, 2019). The response of cells to DNA damage will restore the DNA

234 structure to its original function, or just enable the cell to survive the damage.

235 We further identified unique and shared gene families among the four *Asparagus*
 236 species, and found 8056 were shared as families with *A. officinalis*, *D. officinalis* and
 237 *A. shenzhenica*, and 1367 gene families, containing 2954 genes appeared to be unique
 238 (**Fig. 2c; Tables S14-15**). Functional annotation revealed that two classes of genes,
 239 small auxin up-regulated RNA (*SAUR*) genes and *cis*-zeatin O-glucosyltransferase
 240 (*cZOGT*) in zeatin biosynthesis, may be closely related to its slow growth and
 241 longevity. There are totally 40 *SAURs* identified in *D. cochinchinensis*, while 20 in *D.*
 242 *officinalis*, 14 in *A. officinalis*, 18 in *A. shenzhenica*, and 23 *cZOGTs* is unique to *D.*
 243 *cochinchinensis*, which was not identified in the three closely related species (**Fig.**
 244 **2f-g**). *SAUR* is an auxin-responsive protein that acts as the plant's toolbox for
 245 adaptation of growth and development (Ren & Gray, 2015; Stortenbeker & Beme,
 246 2019). Studies in rice confirmed *SAUR* inhibit growth via negatively regulates auxin
 247 synthesis and transport, in addition, *SAUR* significantly increased anthocyanin
 248 content and the ABA level (Kant *et al.*, 2009). *ZOGT* are known led to growth
 249 retardation and senescence delay (Rodo *et al.*, 2008; Kudo *et al.*, 2012). Therefore,
 250 the expansion of *SAURs* and *cZOGTs* is predicted to at least partly account for its
 251 slow growth and long life, even its adaptation to the arid environment of karst
 252 landform.

253 **Metabolic Profiling and Differential Metabolite Pathways in the Process of** 254 **Dragon's Blood Formation**

255 Dragon's blood does not appear directly after the wound, the process of its
 256 formation takes time. From the slices, we observed the color of resin gradually
 257 deepened during the formation of dragon's blood, with filling the vascular bundle and
 258 surrounding parenchymal cells after 90 d (**Fig. 3a**). Eventually, the resin might be in
 259 the form of tear-like drops or chips which coats the place of wound (**Fig.1a**),
 260 protecting the plant from the spread of pathogens. Thus, it is the product of dragon
 261 tree defense mechanism (Wang, 2007; Jura-Morawiec & Tulik, 2016; Maděra *et al.*,
 262 2020).

263 To elucidate the changes in metabolite composition during the formation of
264 dragon's blood, we focused on identifying metabolites with significant difference by
265 filtering the picked peak. Ultimately, a total of 148 differential metabolites were
266 identified or tentatively identified (**Table S16; Fig. 3c**). As shown in Fig. 3c, content
267 of most compounds changed significantly after injury. Among them, 85 increased,
268 including flavonoids, phenols, polyphenols, phenylpropanoids and phenolic acids, and
269 63 compounds, mainly organic acids, amines and steroid were decreased. Consistently,
270 the total normalized peak areas showed the same results (**Fig. S9**).

271 Biological processes associated with the formation of dragon's blood were
272 investigated using transcriptome data. The up-regulated differentially expressed genes
273 (DEGs) are mainly enriched in "metabolic pathway", "biosynthesis of secondary
274 metabolites", showing that different time points correspond to different metabolic
275 processes (**Fig. 4**). For example, the first metabolic pathways to change in 2 h are
276 plant-pathogen interaction, plant hormone signal transduction as well as starch and
277 sucrose metabolism. Then, biosynthesis of amino acid and citrate cycle were
278 significantly enriched, and flavonoid biosynthesis appeared at 3 d and lasted until 17
279 M, which was the most active metabolic pathway, and terpenoid backbone
280 biosynthesis at 15 d, sesquiterpenoid and triterpenoid biosynthesis at 30 d,
281 phenylpropanoid biosynthesis at 6 M. The differentially expressed metabolic
282 pathways combined with the metabolic profiling showed in Fig. 3c, we infer that the
283 synthesis of dragon's blood involves a series of very complex signal transduction,
284 gene expression regulation and metabolism processes.

285 **Genes in Flavonoids Biosynthesis in *D. cochinchinensis***

286 Numerous studies have shown that flavonoids are the main components of dragon's
287 blood (Gupta *et al.*, 2008; Fan *et al.*, 2014). It occurs via phenylpropane biosynthesis
288 pathway (**Fig. 5**). From our genome assembly, 11 gene families related to dragon's
289 blood formation were identified (**Table S17**), including some common enzymes in the
290 flavonoid synthesis pathway: PAL, C4H, 4CL, CHS, CHI, F3H, FLS, DFR, and
291 downstream synthetic genes producing specific components of dragon's blood: such

292 as OMT (O-methyltransferase) to form loureirin A and PPO (polyphenol oxidase) to
 293 form loureirin B, which is the index components of the standard of dragon's blood in
 294 China. Gene expression results showed that four of the five *CHS* genes and two *CHI*
 295 genes showed significantly upregulated after injury and lasted until 17 M. Similarly,
 296 two of the three *F3H* genes are not expressed in healthy tissues, and the highest
 297 expression level is at 3 d and 5 d after injury. Some of *FLS* and *PPO* members are
 298 also almost not expressed in health stem, but expressed after injury (**Fig. 5**). We
 299 speculated that these genes may be specifically involved in the formation of dragon's
 300 blood, and their expression is significantly induced by injury.

301 Further, based on co-expression model, two kinds of transcription factors bHLH
 302 and MYB, expressed in the wounded stems, were predicted as the core regulators in
 303 the synthesis of flavonoids, which are likely target to *PAL*, *CHS* and *FLS* and
 304 participate in the regulation of flavonoid synthesis pathway (**Fig. S10; Table S18**).
 305 Analyzing the 2000 bp upstream of the gene coding region, we found many *cis*-acting
 306 elements specifically recognized by bHLH and MYB (**Table S19**), which further
 307 demonstrated these enzyme genes are likely regulated by the two groups of
 308 transcriptional factors. Moreover, from the expression heat map of bHLH and MYB
 309 (**Figs. S11-12**), we can see that they are significantly induced after injury 2 h. We
 310 speculate that bHLH and MYB not only regulates a key enzyme gene *PAL*, *CHS* and
 311 *FLS*, but also regulates the whole flavonoids synthetic pathway.

312 **Specific Regulation of ROS Signaling during Dragon's Blood Formation**

313 Focused on analysis genes involved in different signaling pathways showed that genes
 314 related to ROS production and scavenging is substantially up-regulated (**Fig. S13**).
 315 Oxidative burst is a common early event in plant defense responses. ROS will
 316 accumulate rapidly when plants are injured outside. ROS generation depends on
 317 NADPH oxidase and apoplastic peroxidases (Bolwell & Wojtaszek, 1997), and
 318 peroxisomes are probably considered the major sites of intracellular H₂O₂
 319 accumulation as a consequence of their oxidative metabolism (del Río *et al.*, 2006).
 320 From the transcriptome data, we found that several enzymes, such as respiratory burst
 321 oxidase homolog (RBOH), amine oxidases (AOC), NAD(P)H oxidases (NOX),

322 peroxidase (PRX), which contribute to ROS generation were significantly upregulated
323 at 2 h after injury (**Fig. S13**). Meanwhile, ROS scavenging enzymes, such as
324 superoxide dismutase (SOD), ascorbate peroxidase (APX), glutathione peroxidase
325 (GPX) and catalase (CAT), are also up-regulated subsequently (**Fig. S13**). Obviously,
326 the degree of oxidative stress is determined by the level of ROS, and the balance
327 between ROS and antioxidants is essential to maintain balanced redox state.

328 H_2O_2 is the most stable type of ROS within plants. To preliminarily validate the
329 H_2O_2 response during the dragon's blood, the change of H_2O_2 content within stem of
330 *D. cochinchinensis* over time after mechanical damage is detected. It was showed
331 H_2O_2 concentration obviously improved 1 h after mechanical damage, and continued
332 to increase until the maximum value at 6 h. After 8 h, it began decreased slightly until
333 the minimum value was reached at 5 d (**Fig. S14**). In contrast, other signals, such as
334 ethylene, nitric oxide and brassinosteroids, have no obvious changing rule (**Figs.**
335 **S15-17**). We infer ROS as the first sensor and the core regulator in the process of
336 dragon's blood formation (**Fig. 6**), which is consistent with the conclusion that ROS
337 represent a point at which various signaling pathways come together (Camejo *et al.*,
338 2016; Sewelam *et al.*, 2016),

339 **DISCUSSION**

340 Dragon's blood is a precious traditional medicine with multiple therapeutic uses
341 (Masaoud & Schmidt, 1995; Gupta *et al.*, 2008; Yang & Yao, 2011; Xin *et al.*, 2013;
342 Fan *et al.*, 2014; Ding *et al.*, 2020; Maděra *et al.*, 2020). *D. cochinchinensis* is one of
343 the most important genuine resources of Chinese dragon's blood as medicine. Here,
344 we constructed its high-quality reference genome at chromosome level combined with
345 PacBio sequencing and Hi-C data. As the first genome report of *Dracaena*, the
346 genome sequence reported herein is expected to contribute to clarify its biology as
347 well as the protection and regeneration of the precious plant resources.

348 *Dracaena* has been famous all over the world for its longevity (Gupta *et al.*,
349 2008; Zheng *et al.*, 2012; Tomlinson & Huggett, 2012; Maděra *et al.*, 2018). However,
350 the molecular mechanism underlying its unique biology traits has been unclear. Our

analysis revealed *D. cochinchinensis* experienced two genome duplication events, one shared with *A. officinalis* and another independently after their divergence (**Fig. 2b**). Additionally, two LTR insertion burst occurred recently (**Fig. 2c**). Comparative genomic analysis revealed the expanded and unique genes *SAUR* and *cZOGT* are likely the main cause of its longevity and slow growth (**Fig. 2f-g**). Genes in the DNA repair pathway were positively selected which contributes to genome maintenance and environmental adaptability (**Table S13**).

The dragon's blood, red resin of *Dracaena*, is a defensive substance produced by *Dracaena* when it is damaged or infected by pathogenic microorganisms (Wang, 2007; Jura-Morawiec & Tulik, 2016; Ding *et al.*, 2020; Maděra *et al.*, 2020). As the main component of dragon's blood, in addition to pharmacological effects, flavonoids also have many benefits to *Dracaena* itself during the growth development. For example, they are involved in plant defense pathogens and microbes, in absorption of free radicals and ultraviolet light (Neill & Gould, 2003; Crozier *et al.*, 2006; Agati *et al.*, 2009, 2011; Agati & Tattini, 2010; Landi *et al.*, 2015). Our current study focused on analyzing differentially expressed pathways and differential metabolites in the process of dragon's blood formation, and biosynthesis and regulatory genes in flavonoids biosynthesis in *D. cochinchinensis*. We observed secondary metabolic-related pathways, including flavonoid, terpenoid backbone and phenylpropanoid biosynthesis, were upregulated in 0-6 M after injury (Fig. 4). Correspondingly, metabolome analysis showed the contents of flavonoids compounds are significantly increased, as well as phenylpropanoids, phenols, and phenolic acids. Whereas, the contents of coumarins, organic acids and amines compounds are significantly decreased (**Fig. 3c; Fig. S9; Table S16**). Enrichment in upregulated genes related to biosynthesis of secondary metabolites includes *PAL*, *CHS*, methyltransferase, etc, which are involved in the flavonoid biosynthesis pathway. Analyzing the dynamic expression profiles of genes in flavonoid biosynthesis pathway during the dragon's blood formation, showed that most of genes induced expression, even some, such as *CHS*, *CHI*, *F3H*, *FLS* and *PPO*, almost not expressed in health stem, but rapidly and significantly induced after injury, which are considered specifically involved in the formation of dragon's blood

(**Fig. 5**). The expression pattern is essentially consistent with the results we observed differential metabolites and metabolic pathways (**Fig. 3c, Fig. 4**). Further, transcription factors bHLH and MYB are predicted as the core regulators involved in flavonoid biosynthesis pathway (**Figs. S10-12**), which is consistent with previously studies that MYB, bHLH and WD40 TFs are involved in flavonoid metabolism by regulating many structural genes (Hichri *et al.*, 2011; Schaart *et al.*, 2013). The corresponding relationship between transcription factor and its target was predicted which will provide important clues for studying the function of these genes and their regulation mechanism. Further studies are definitely required to comprehensively elucidate the process of the formation of the red resin in *D. cochinchinensis*, and hence, enable controlled and reasonable dragon's blood extraction for commercial purposes.

To reveal the regulatory network of dragon's blood formation induced by injury, we analyzed the signal pathways systematically and considered ROS is the crucial signal stimuli. It is found genes in ROS generation and scavenging are all upregulated after injury (**Fig. S13**), the content of H₂O₂ showed regular dynamic changes at different time points after injury (**Fig. S14**), and genes related to the main place of ROS production are expanded (**Table S12, S20**). As we know, *D. cochinchinensis* grows in limestone area where is full of sunlight and strong UV-B, which can cause the degradation of CP43 and CP47 and the decrease of photosynthetic capacity, leads to ROS production and cause DNA damage and protein denaturation (for review Takshak & Agrawal, 2019). KEGG pathway analysis indicated that many of the positively selected genes are related to that, such as base excision repair (BER), nucleotide excision repair (NER), mismatch repair (MMR) and DNA replication (**Table S13**), which implies the *D. cochinchinensis* has a great adaptation to the environment in the evolution to maintain genomic integrity. In addition, genes family in phenylpropanoid biosynthesis and photosynthesis pathway is expanded (**Table S12, S20**). More and more evidences showed flavonoids have multiple photoprotective effects (Winkle-Shirley, 2002; Agati & Tattini, 2010) and functions as scavenger of ROS (Yamasaki *et al.*, 1997; Agati *et al.*, 2009), as well as moderator of auxin

transport (Peer & Murphy, 2007). Taken together, we believe there are some potential associations among “ROS signal--flavonoids biosynthesis--environmental adaptability” in *D. cochinchinensis* (Fig. 6). This should be the results of adaptation to the environment in the process of evolution. We speculate that their longevity depends on their strong self-healing ability and adaptability to the environment.

CONCLUSION

This study reports a chromosome-level genome assembly for *D. cochinchinensis*, the first genome for *Dracaena* Vand. ex L. It provides valuable genetic resources and creates a large scope for studying *Dracaena*. We revealed the genetic basis of its longevity and slow growth at the genomic level, and preliminarily deciphered the molecular mechanism of injury induced flavonoids biosynthesis to form dragon's blood through transcriptome and metabolome method analysis. These findings are important for an improved understanding of *Dracaena* adaptation, resource protection and utilization and further studies on serving human medicine and health.

METHODS

Library Construction and Sequencing

Combined with four types of libraries constructed: (1) the short-read sequencing library with 350-bp insertions from Illumina HiSeq X Ten instrument platforms; (2) The linked read sequencing libraries were constructed on 10X Genomics GemCode platform and sequencing from Illumina platforms; (3) the SMRT Cell libraries with an insert size of 20 Kb for PacBio sequencing; and (4) the chromatin interaction mapping (Hi-C) library with 350-bp insertions for Illumina sequencing.

Genome Assembly and Assessment of the Assembly Quality

De novo assembly of the PacBio single-molecule long reads from SMRT Sequencing was performed using FALCON (<https://github.com/PacificBiosciences/FALCON/>) (Chin *et al.*, 2013). BWA-MEM (Li *et al.*, 2013) was used to align the 10X Genomics data to the assembly using default settings. Scaffolding was performed by fragScaff (version 1.1) with the barcoded sequencing reads. These contigs were then used to form super-scaffolds. BWA (v0.7.8) (Li & Durbin, 2009) software was used to align

the Hi-C clean data to the preceding assembly. According to the linkage information and restriction enzyme site, the string graph formulation was used to construct the scaffold graph with LACHESIS (Burton *et al.*, 2013).

Genome Annotation

Genome repeat sequences were annotated de novo by using RepeatMasker (<http://www.repeatmasker.org>). The protein-coding genes were predicted based on the repeat-masked genome, conserved protein homologues and assembled transcripts. miRNA and snRNA genes were predicted by INFERNAL software using the Rfam database (Supplementary method).

Repetitive Sequences Analysis

TEs identification was performed using a combination of homology-based and ab initio-based approaches. For the homology-based method, RepeatMasker (v3.3.0) was applied to identify the TEs with a known repeat library (Repbase 15.02). For the ab initio-based approach, TEs prediction was implemented by RepeatScout (Price *et al.*, 2005), RepeatModeler (v1.0.5; [http:// www.repeatmasker.org/RepeatModeler/](http://www.repeatmasker.org/RepeatModeler/)), Piler, and LTR_FINDER (Xu *et al.*, 2007) to build the de novo repeat libraries. RepeatMasker (v3.3.0) (Chen, 2004.) was run on the *D. cochinchinensis* genome assembly using the de novo libraries as reference libraries. Tandem Repeats Finder (TRF) (Benson, 1999) was used to identify the tandem repeats with integrated repeat library.

We further investigated the dynamics of long terminal repeats retrotransposons (LTR-RTs) in the *D. cochinchinensis* genome. *De novo* detection of LTR-RTs was performed using LTRharvest (-motif tgca -motifmis 1) and LTRfinder (-minlenltr 100 -maxlenltr 5000 -mindistltr 1000 -maxdistltr 20000). The primer binding site (PBS) motifs were annotated using LTRdigest, and only elements containing PBS motifs were retained for further domain annotation. Domains of retrotransposons were identified by searching against HMM profiles collected by Gypsy Database (<http://gydb.org/>), and internal features of LTR-RTs were annotated by LTRdigest. Intact LTR-RTs that contained domains (gag domain, protease domain, reverse transcriptase domain and integrase domain) were clustered, and those belonged to

families with less than 5 members were discarded. Nucleotide divergence rate (λ) was calculated from the MUSLCE alignment of the 5' and 3' LTR sequences of LTR-RTs by EMBOSS program distmat. The insertion date (T) was computed for each LTR-RT ($T = K/2r$, K: genetic distance, $K = -0.75 \ln(1 - 4\lambda/3)$), which was set substitution rate (r) of 1.3×10^{-8} substitutions per site per year.

Gene Family Construction and Phylogenetic Analysis

Gene families were generated by OrthoMCL (<http://orthomcl.org/orthomcl/>) (Li *et al.*, 2003) from 17 species, including *Dracaena cochinchinensis*, *Asaragus officinalis*, *Dendrobium officinale*, *Apostasia shenzhenica*, *Elaeis guineensis*, *Oryza sativa* (phytozomev10), *Ananas comosus*, *Musa acuminata*, *Arabidopsis thaliana* (phytozomev10), *Populus trichocarpa*, *Aquilaria sinensis*, *Hevea brasiliensis*, *Citrus grandis*, *Carica papaya*, *Eucalyptus grandis*, *Vitis vinifera* (phytozomev10), *Amborella trichopoda* with the parameter of 1.5 inflation index.

Protein sequences from 252 shared single-copy gene were utilized to construct the phylogenetic tree. Protein sequences of these orthologs were aligned by MUSCLE (Edgar, 2004), and the phylogenetic tree was constructed by the ML (maximum likelihood) TREE algorithm in RAxML software (Stamatakis, 2006). Then mcmctree program of PAML (<http://abacus.gene.ucl.ac.uk/software/paml.html>) (Yang, 2007) was applied to estimate divergence time among 17 species. Five calibration points were selected and applied in the present study: *V. vinifera* and *O. sativa* time (130~200 Mya) (Iorizzo, *et al.*, 2016), *O. sativa* and *A. shenzhenica* time (137~156 Mya) (Zhang *et al.*, 2017), *V. vinifera* and *C. papaya* time (114~134 Mya) (Ming *et al.*, 2008), *P. trichocarpa* and *A. thaliana* time (107~109 Mya) (Tuskan *et al.*, 2006), *H. brasiliensis* and *P. trichocarpa* time (65~86 Mya) (Tang *et al.*, 2016; Tuskan *et al.*, 2006). Gene gain and loss in gene families were determined using CAFÉ (v2.2) (Computational Analysis of gene Family Evolution). The program uses a birth and death process to model gene gain and loss over a phylogeny.

Expansion and Contraction of Gene Families

Gene gain and loss in gene families were determined using CAFÉ (v2.2)

(Computational Analysis of gene Family Evolution). The program uses a birth and death process to model gene gain and loss over a phylogeny. Enrichment of KEGG pathway for expanded gene families of *D. cochinchinensis* was summarized and visualized using KOBAS software.

RNA Sequencing and Analysis

Five tissues (root, stem, leaves, flower, fruit) were harvested from *D. cochinchinensis*, and 33 cut injury-treated stems at different time points (0h, 2h, 6h, 12h, 24h, 2d, 3d, 4d, 5d, 10d, 15d, 30d, 6M, 17M). Three biological replicates for each tissue were collected. Raw reads from different samples were filtered to obtain high-quality clean reads following standard process. For gene expression analysis, clean reads were mapped to *D. cochinchinensis* genome using Tophat (v2.0.8) (Kim *et al.*, 2013). Then the readcounts of each gene was normalized for downstream analysis. A gene is defined as expressed if its RPKM value is ≥ 1 in at least one of all the transcriptomes.

Differential expression analysis between a treatment sample and the control sample was performed using the DESeq R package (v1.18.0). The resulting P-values were adjusted using the Benjamini and Hochberg's approach for controlling the false discovery rate. Genes with an adjusted P-value < 0.05 found by DESeq were assigned as differentially expressed. KEGG pathway enrichment analysis of the DEGs was performed using KOBAS software (Mao *et al.*, 2005).

Metabolic Profiling Acquisition Based On UHPLC-Q-Orbitrap-MS

The freeze-dried samples were dissolved in 50% methanol-water (v/v) solutions at the concentration of $5 \text{ g} \cdot \text{mL}^{-1}$, and the solutions were filtered, centrifuged at $13,200 \times g$ for 10 min, and 10 μL of the resulting supernatant was directly injected into the LC-MS system.

Untargeted metabolomics was accomplished using a U3000 UHPLC equipped with aQ-Exactive™ Q-Orbitrap MS through a HESI source (Thermo Fisher Scientific, Waltham, MA, USA). A Waters Acquity UPLC HSS T₃ column ($2.1 \times 100 \text{ mm}$, $1.8 \mu\text{m}$) was used for chromatographic separation with a column temperature of 35°C . The flow rate was set as $0.4 \text{ mL} \cdot \text{min}^{-1}$. The autosampler temperature was maintained

528 at 4°C.

529 The mobile A and B phases were 0.1% formic acid–water (v/v) and acetonitrile,
530 respectively. The following elution gradient was used: 10% (B) in 0–3 min, 10–22%
531 (B) in 3–5 min, 22–27% (B) in 5–7 min, 27–45% (B) in 7–25 min, 45–57% (B) in 25–35
532 min, 57–100% (B) in 35–36 min, 100% (B) in 36–37 min, 100–10% (B) in 37–37.5 min,
533 and 10% (B) in 37.5–40 min.

534 The HESI source parameters were as follows: the scan mode was full-ms/dd-ms
535 mode and the scanning range was m/z 150–1,500 at a resolution of 70,000; the spray
536 voltage was 3.0 kV in the positive mode and 2.8 kV in negative mode; the capillary
537 temperature and the probe heater temperatures were 320°C and 350°C, respectively;
538 the sheath gas (N₂) was 35 arbitrary units (arb) and the auxiliary gas was 10 arb (99.9%
539 pure N₂). The S-lens level value was 50 V. The normalized collision energy was
540 20/40/60 V. Data acquisition was conducted using Xcalibur 4.1 software (Thermo
541 Fisher Scientific).

542 **FOUND**

543 This work was supported by the program of CAMS Initiative for Innovative Medicine
544 (No. 2021-1-I2M-032); the National Natural Science Foundation of China (82173925,
545 81573525); the National Key Research and Development Program of China (Grant
546 NO. 2018YFC1706401).

547 **Author Contributions**

548 YX and JW conceived and designed the project, and participated in all the related
549 analyses. YX and KZ performed the data analyses and drafted the manuscript. KZ
550 performed genome assembly and annotated the genome. ZZ conducted metabolome
551 analysis. YL conducted H₂O₂ analysis. YX, ZZ and YL contributed to the sample
552 preparation used for genome and RNA sequencing. All authors contributed to the
553 writing of the paper.

554 **Data Availability Statement**

555 The raw data of genome and transcriptome sequencing reported in this paper will be
556 deposited in the Genome Sequence Archive (Wang et al., 2017) in BIG Data Center,

557 Beijing Institute of Genomics (BIG), Chinese Academy of Sciences under the
558 accession number (PRJCA007701). It will be publicly accessible at
559 <http://bigd.big.ac.cn/gsa>) when paper is accepted.

560 ACKNOWLEDGMENTS

561 The authors declare no competing financial interests

562 REFERENCES

- 563 **Adolt, R., Habrova, H., Madera, P.** (2012). Crown age estimation of a
564 monocotyledonous tree species *Dracaena cinnabari* using logistic regression.
565 *Trees* DOI 10.1007/s00468-012-0704-9.
- 566 **Adolt, R. and Pavlis, J.** (2004). Age structure and growth of *Dracaena cinnabari*
567 populations on Socotra. *Trees* **18**, 43–53.
- 568 **Agati, G., Biricolti, S., Guidi, L., Ferrini, F., Fini, A., Tattini, M.** (2011). The
569 biosynthesis of flavonoids is enhanced similarly by UV radiation and root zone
570 salinity in *L. vulgare* leaves. *J Plant Physiol.* **168**, 204–212.
- 571 **Agati, G., Stefano, G., Biricolti, S., Tattini, M.** (2009) Mesophyll distribution of
572 'antioxidant' flavonoid glycosides in *Ligustrum vulgare* leaves under contrasting
573 sunlight irradiance. *Annals of Botany* **104**, 853–861.
- 574 **Agati, G. and Tattini, M.** (2010). Multiple functional roles of flavonoids in
575 photoprotection. *New Phytol.* **186**, 786–793.
- 576 **Benson, G.** (1999). Tandem repeats finder: a program to analyze DNA sequences.
577 *Nucleic Acids Res.* **27**: 573–580.
- 578 **Bolwell, G.P. and Wojtaszek, P.** (1997). Mechanisms for the generation of reactive
579 oxygen species in plant defence-a broad perspective. *Physiol. Mol. Plant Pathol.*
580 **51**, 347–366.
- 581 **Burton, J.N., Adey, A., Patwardhan, R.P., Qiu, R., Kitzman, J.O., Shendure, Jay.**
582 (2013). Chromosome-scale scaffolding of de novo genome assemblies based on
583 chromatin interactions. *Nature Biotechnol.* **31**, 1119-25.
- 584 **Cai, X.T. and Xu Z.F.** (1979). A Study on the resource of Chinese dragon's blood.
585 *Acta Botanica Yunnanica* **1(2)**, 1-10.

- 586 **Camejo, D., Guzman-Cedeno, A., Moreno, A.** (2016). Reactive oxygen species,
587 essential molecules, during plant-pathogen interactions. *Plant Physiol. Biochem.*
588 **103**, 10-23.
- 589 **Chin, C.S., Alexander, D.H., Marks, P., Klammer, A.A., Drake, J., Heiner, C.,**
590 **Clum, A. et al.** (2013). Nonhybrid, finished microbial genome assemblies from
591 long-read SMRT sequencing data. *Nature Methods* **10**, 563.
- 592 **Christ, H.** (1886). *Eine Frühlingsfahrt Zu Den Canarischen Inseln*; H. Georg's Verlag:
593 Basel, Switzerland.
- 594 **Crozier, A., Clifford, M.N., Ashihara, H.** (2006). *Plant Secondary Metabolites*. UK:
595 Blackwell Publishers.
- 596 **D'hont, A., Denoeud, F., Aury, J.M., Baurens, F.C., Carreel, F., Garsmeur, O.,**
597 **Noel, B. et al.** (2012). The banana (*Musa acuminata*) genome and the evolution
598 of monocotyledonous plants. *Nature* **488 (7410)**, 213-217.
- 599 **Darias,V., Bravo, L., Rabanal, R., Sanchez Mateo, C., Gonzalez Luis, R.M.,**
600 **Hernandez Perez, A.M.** (1989). New contribution to the ethnopharmacological
601 study of the Canary Islands. *J Ethnopharmacol.* **25**: 77–92.
- 602 **del Río, L.A., Sandalio, L.M., Corpas, F.J., Palma, J.M., Barroso, J.B.** (2006).
603 Reactive oxygen species and reactive nitrogen species in peroxisomes.
604 Production, scavenging, and role in cell signalling. *Plant Physiol.* **141**,330e335.
- 605 **Ding, X.P., Zhu, J.H., Wang, H., Chen, H.Q., Mei, W.L.** (2020). Dragon's Blood
606 from *Dracaena cambodiana* in China: Applied History and Induction
607 Techniques toward Formation Mechanism. *Forests* **11**, 372.
- 608 **Edgar, R.C.** (2004). MUSCLE: multiple sequence alignment with high accuracy and
609 high throughput. *Nucleic Acids Res.* **32**, 1792-1797.
- 610 **Edwards, H.G.M., Farwell, D.W., Quye, A.** (1997). 'Dragon's
611 blood'-characterization of an ancient resin using Fourier transform Raman
612 spectroscopy. *J. Raman Spectrosc.* **28**, 243–249.
- 613 **Fan, J.Y., Tao, Y., Chui-Mei S.T. □ Lin, Z., Wan-Ling, P., Ya-Zhou,**
614 **Z., Zhong-Zhen, Z., et al.** (2014). A systematic review of the botanical,

615 phytochemical and pharmacological profile of *Dracaena cochinchinensis*, a plant
616 source of the ethnomedicine “Dragon’s Blood”. *Molecules* **19**, 10650–10669.

617 **González, A.G., Hernández, J.C., León, F., Padron, J.I., Estevez, F., Quintana, J.,**
618 **Bermejo, J.** (2003). Steroidal saponins from the bark of *Dracaena draco* and their
619 cytotoxic activities. *J. Nat. Prod.* **66**, 793– 798.

620 **Gupta, D., Bleakley, B., Gupta, R.K.** (2008). Dragon's blood: botany, chemistry and
621 therapeutic uses. *J. Ethnopharmacol.* **115**, 361–80.

622 **Hao, H.Z., He, A.D., Wang, D.C., Yin, Z., Zhou, Y.J., Liu, G., Liang, M.L. et al.**
623 (2015a). Antiplatelet activity of loureirin A by attenuating Akt phosphorylation:
624 In vitro studies. *European J Pharmacol.* **746**, 63–69.

625 **Hao, Q., Saito, Y., Matsuo, Y., Li, H.Z., Tanaka, T.** (2015b). Chalcane-stilbene
626 conjugates and oligomeric flavonoids from Chinese dragon's blood produced
627 from *Dracaena cochinchinensis*. *Phytochemistry* **119**, 76–82.

628 **Harkess, A., Zhou, J.S., Xu, C.Y., Bowers, J.E., der Hulst, R.V., Ayyampalayam,**
629 **S., Mercati, F. et al.** (2017). The asparagus genome sheds light on the origin and
630 evolution of a young Y chromosome. *Nature Commu.* **8(1)**, 1279.

631 **Hernández, J.C., León, F., Quintana, J., Estevez, F., Bermejo, J.** (2004). Icogenin,
632 a new cytotoxic steroidal saponin isolated from *Dracaena draco*. *Med. Chem.* **12**,
633 4423–4429.

634 **Hichri, I., Barrieu, F., Jochen Bogs, Kappel, C., Delrot, S., Lauvergeat, V.** (2011).
635 Recent advances in the transcriptional regulation of the flavonoid biosynthetic
636 pathway. *J Exp. Bot.* **62**, 2465–2483.

637 **Humboldt, A.V.** (1814) *Relation Historique Du Voyage Aux Régions Équinoxiales*
638 *Du Nouveau Continent*.

639 **Hubáľková, I., Maděra, P., Volařík, D.** (2015) Growth dynamics of *Dracaena*
640 *cinnabari* under controlled conditions as the most effective way to protect
641 endangered species. *Saudi J Biol. Sci.* **24(7)**, 1445-1452.

642 IUCN red list of threatened species version. [www document] URL
643 <http://www.iucnredlist.org>. [accessed on 25 October 2017].

644 **Jura-Morawiec, J. and Tulik, M.** (2016) Dragon’s blood secretion and its ecological

significance. Chemoecol. **26**, 101–105.

Kamel, M., Ghazaly, U.M., Callmander, M.W. (2018). Conservation status of the endangered Nubian dragon tree *Dracaena ombet* in Gebel Elba national park. Egypt. Oryx **49**, 704–709.

Kant, S., Bi, Y.M., Zhu, T., Rothstein, S.J. (2009). *SAUR39*, a small auxin-up RNA gene, act as a negative regulator of auxin synthesis and transport in rice. Plant Physiol. **151**, 691–701.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. **14**, R36.

Kim, S., Park, M., Yeom, S.I., Kim, Y.M., Lee, J.M., Lee, H.A., Seo, E. et al. (2014). Genome sequence of the hot pepper provides insights into the evolution of pungency in Capsicum species. Nat. Genet. **46**(3), 270–279.

Kudo, T., Makita, N., Kojima, M., Tokunaga, H., Sakakibara, H. (2012). Cytokinin Activity of cis-Zeatin and Phenotypic Alterations Induced by Overexpression of Putative cis-Zeatin-O-glucosyltransferase in Rice. Plant Physiol. **160**, 319–331.

Landi, M., Tattini, M., Gould, K.S. (2015). Multiple functional roles of anthocyanins in plant-environment interactions. Environ. Exp. Bot. **119**, 4–17.

Lengálová, K., Kalivodová, H., Habrová, H., Maděra, P., Tesfamariam, B., Šenfeldr, M. (2020) First Age-Estimation Model for *Dracaena ombet* and *Dracaena draco* subsp. *Caboverdeana*. Forests **11**(264), doi:10.3390/f11030264.

Li, H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. [arXiv:1303.3997](https://arxiv.org/abs/1303.3997) [q-bio.GN]

Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25**, 1754–1760 .

Li, S.F., Wang, J., Dong, R., Zhu, H.W., Lan, L.N., Zhang, Y.L. et al. (2020) Chromosome-level genome assembly, annotation and evolutionary analysis of the ornamental plant *Asparagus setaceus*. Hortic. Res. **7**, 48.

Luo, Y., Wang, H., Zhao, Y.X., Zeng, Y.B., Shen, H.Y., Dai, H.F., Mei, W.L.

(2011). Cytotoxic and antibacterial flavonoids from dragon's blood of *Dracaena cambodiana*. *Planta Medica* **77**(18), 2053-2056.

Maděra, P., Forrest, A., Hanáček, P., Vahalík, P., Gebauer, R., Plichta, R., Jupa R. et al. (2020). What We Know and What We Do Not Know About Dragon Trees? *Forests* **11**, 236.

Maděra, P., Habrová, H., Šenfeldr, M., Kholová, I., Lvončík, S., Ehrenbergerová, L., Roth, M. et al. (2018). Growth dynamics of endemic *Dracaena cinnabari* Balf. f. of Socotra Island suggest essential elements for a conservation strategy. *Biologia* doi: 10.2478/s11756-018-0152-0.

Mägdefrau, K. (1975) Das alter der drachenbäume auf Tenerife. *Flora* 164, 347–357.

Mao, X., Cai, T., Olyarchuk, J.G., Wei, L. (2005). Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics* **21**, 3787-93.

Masaoud, M. and Schmidt, J.J.P. (1995). Sterols and triterpenoids from *Dracaena cinnabari*. *Phytochemistry* **38**(3), 795-796.

Mimaki, Y., Kuroda, M., Ide Atsushi Kameyama, A., Yokosuka, A., Sashida, Y. (1999). Steroidal saponins from the aerial parts of *Dracaena draco* and their cytostatic activity on HL-60 cells. *Phytochemistry* **50**, 805–813.

Neill, S.O. and Gould, K.S. (2003). Anthocyanins in leaves: Light attenuators or antioxidants. *Funct. Plant Biol.* **30**, 865–873.

Nisa, M., Ying, H.Y., Benhamed, M., Raynaud, C. (2019). The plant DNA damage response: signaling pathways leading to growth inhibition and putative role in response to stress conditions. *Front. plant sci.* **10**, 653.

Pang, D.R., Su, X.Q., Zhu, Z.X., Sun, J., Li, Y.T., Song, Y.L. et al. (2016). Flavonoid dimers from the total phenolic extract of Chinese dragon's blood, the red resin of *Dracaena cochinchinensis*. *Fitoterapia* **115**, 135–141.

Peer, W.A. and Murphy, A.S. (2007). Flavonoids and auxin transport: modulators or regulators? *Trends Plant Sci.* **12**, 556-563.

Price, A.L., Jones, N.C., Pevzner, P.A. (2005). De novo identification of repeat families in large genomes. *Bioinformatics* **21**, 351–358.

- 705 **Pütter, A.** (1925) Altersbestimmung an Drachenbäumen von Tenerife. Sitz. Heidelb.
706 Akad. Wissenschaften. Math.Nat. Kl. **12**, 12–18.
- 707 **Ren, H. and Gray, W.M.** (2015). SAUR proteins as effectors of hormonal and
708 environmental signals in plant growth. Mol. Plant **8**, 1153–1164.
- 709 **Rodo, A.P., Brugiére, N., Vankova, R., Malbeck, J., Olson, J.M., Haines, S.C.,**
710 **Martin, R.C. et al.** (2006). Biochemical characteristics and ligand-binding
711 properties of Arabidopsis cytokinin receptor AHK3 compared to CRE1/AHK4 as
712 revealed by a direct binding assay. J. Exp. Bot. **57**, 4051–4058.
- 713 **Rodo, A.P., Brugiére, N., Vankova, R., Malbeck, J., Olson, J.M., Haines, S.C.,**
714 **Martin, R.C. et al.** (2008). Over-expression of a zeatin O-glucosylation gene in
715 maize leads to growth retardation and tasselseed formation. J Exp. Bot. **59(10)**,
716 2673–2686.
- 717 **Schaart, J.G., Dubos, C., Fuente, I.R.D.L., van Houwelingen, A.M.M.L., de Vos,**
718 **R.C.H., Jonker, H.H., Xu, W.J., et al.** (2013). Identification and characterization
719 of MYB-bHLH-WD40 regulatory complexes controlling proanthocyanidin
720 biosynthesis in strawberry (*Fragaria x ananassa*) fruits. *New Phytol.* **197**,
721 454–467.
- 722 **Schuch, A.P., Moreno, N.C., Schuch, N.J., Menck, C.F.M., Garcia, C.C.M.** (2017)
723 Sunlight damage to cellular DNA: Focus on oxidatively generated lesions. Free
724 Radical Bio.Med. **107**, 110–124.
- 725 **Schenck, H.** (1907) Beitrage Zur Kenntnis Der Vegetation Der Canarischen Inseln;
726 Wissenschaftliche Ergebnisse der Deutschen Tiefsee-Expedition auf dem
727 Dampfer “Valdivia”; Fischer: Jena, Germany, **2(1)**, 225–406.
- 728 **Sewelam, N., Kazan, K., Schenk, P.M.** (2016). Global Plant Stress Signaling:
729 Reactive Oxygen Species at the Cross-Road. Front Plant Sci. **7**, 187.
- 730 **Simao, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M.**
731 (2015). BUSCO: assessing genome assembly and annotation completeness with
732 single-copy orthologs. Bioinformatics **31**, 3210–3212.
- 733 **Stamatakis, A.** (2006) RAXML-VI-HPC: maximum likelihood-based phylogenetic
734 analyses with thousands of taxa and mixed models. Bioinformatics **22**,

2688-2690.

Stortenbeker, M. and Bemer, M. (2019) The SAUR gene family: the plant's toolbox for adaptation of growth and development. *J Exp. Bot.* **70** (1), 17–27.

Sun, J., Liu, J.N., Fan, B., Chen, X.N., Pang, D.R., Zheng, J., Zhang, Q. et al. (2019). Phenolic constituents, pharmacological activities, quality control, and metabolism of *Dracaena* species: A review. *J. Ethnopharmacol.* **224**, 112138.

Symon, D.E. (1974) The growth of *Dracaena draco*—Dragon's blood tree. *J. Arnold Arbor.* **55**, 51–58.

Takshak, S. and Agrawal, S.B. (2019). Defense potential of secondary metabolites in medicinal plants under UV-B stress. *J Photoch. Photobio. B.* **193**, 51–88.

Tang, C.R., Yang, M., Fang, Y.J., Luo, Y.F., Gao, S.H., Xiao, X.H., An, Z.W. et al. (2016). The rubber tree genome reveals new insights into rubber production and species adaptation. *Nature plants* **2**(6), 1-10.

Teh, B.T., Lim, K., Yong, C.H., Ng, C.C.Y., Rao, S.R., Rajasegaran, V., Lim, W.K. et al. (2017). The draft *genome* of tropical fruit durian (*Durio zibethinus*). *Nat. Genet.* **49**(11), 1633-1641.

Teng, Z., Zhang, M., Meng, S., Dai, R., Meng, W., Deng, Y., Huang, L.J.B.C. (2015). A comparative study on volatile metabolites profile of *Dracaena cochinchinensis* (Lour.) S.C.Chen xylem with and without resin using GC-MS. *Biomed. Chromatogr.* **29**(11), 1744-1749.

Tomlinson, P.B. and Huggett, B.A. (2012). Cell longevity and sustained primary growth in palm stems. *Am. J. Bot.* **99**(12), 1891-1902.

Tuskan, G.A., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N. et al. (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**(5793), 1596-1604.

Vaničková, L., Pompeiano, A., Maděra, P., Massad, T.J., Vahalik, P. (2020). Terpenoid profiles of resin in the genus *Dracaena* are species specific. *Phytochemistry* **170**, 112197.

Wang, X.H. (2007). Flavones formed from xylem stem of *Dracaena cochinchinensis* by co-culture of fungus strain and bacteria strain. *Nat. Pro. Res. Dev.* **19**, 11–15.

- 765 **Wang, X.H., Gong, M., Tang, L., Zheng, S., Lou, J.D., Ou, L.** (2013).
766 Gomes-Laranjo J; Zhang C. Cloning, bioinformatics and the enzyme activity
767 analyses of a phenylalanine ammonia-lyase gene involved in dragon's blood
768 biosynthesis in *Dracaena cambodiana*. *Mol. Biol. Rep.* **40**, 97–107.
- 769 **Wang, X.H., Zhang, C., Yang, L.-L., Gomes-Laranjo, J.** (2011). Production of
770 dragon's blood in *Dracaena cochinchinensis* plants by inoculation of *Fusarium*
771 *proliferatum*. *Plant Sci.* **180**(2), 0-299.
- 772 **Wang, Y., Song, F., Zhu, J., Zhang, S., Yang, Y., Chen, T. et al.** (2017). GSA:
773 Genome Sequence Archive . *Genom, Proteom Bioinf.* **15**, 14–18.
- 774 **Winkel-Shirley, B.** (2002). Biosynthesis of flavonoids and effects of stress. *Curr.*
775 *Opin. Plant Bio.* **5**, 218–223.
- 776 **Xin, N., Yang, F.J., Li, Y., Li, Y.J., Dai, R.J., Meng, W.W., Chen, Y. et al.** (2013).
777 Dragon's blood dropping pills have protective effects on focal cerebral ischemia
778 rats model. *Phytomedicine* **21**(1), 68-74.
- 779 **Xu, Z. and Wang, H.** (2007). LTR_FINDER: an efficient tool for the prediction of
780 fulllength LTR retrotransposons. *Nucleic Acids Res* **35**, 265–268.
- 781 **Yan, L., Wang, X., Liu, H., Tian, Y., Lian, J.M., Yang, R.J., Hao, S.M., et al.**
782 (2015). The genome of *Dendrobium officinale* illuminates the biology of the
783 important traditional Chinese orchid herb. *Mol. Plant.* **8**: 922-934.
- 784 **Yang, X. and Yao, L.** (2011). Pharmacological Action and Clinical Experience of
785 Dragon's Blood. *Heilongjiang Medicine Journal* **24**(2), 265-266.
- 786 **Yang, Z.** (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Bio*
787 *Evol.* **24**, 1586-1591.
- 788 **Yu, X.J., Zheng, H.K., Wang, J., Wang, W., Su, B.** (2006). Detecting
789 lineage-specific adaptive evolution of brain-expressed genes in human using
790 rhesus macaque as outgroup. *Genomics* **88**, 745-751.
- 791 **Zhang, G.Q., Liu, K.W., Li, Z., Lohaus, R., Hsiao, Y.Y., Niu, S.C., Wang, J.Y. et**
792 **al.** (2017). The *Apostasia* genome and the evolution of orchids. *Nature* **549**,
793 379-383.
- 794 **Zhang, L.J., Li, X.X., Ma, B., Gao, Q., Du, H.L., Han, Y.H., Li, Y. et al.** (2017).

795 The artary Buckwheat Genome Provides Insights into Rutin Biosynthesis and
796 Abiotic Stress Tolerance. *Mol. Plant* **10**, 1224–1237.

797 **Zheng, D.J., Xie, L S., Wang, Y., Zhang, Z.L., Zhang, W.** (2009). Research
798 Advances in Dragon’s Blood Plants in China. *Chin. Wild Plant Res.* **28**, 15–20.

799 **Zheng, D.J., Xie, L.S., Zhu, J.H., Zhang, Z.L.** (2012). Low genetic diversity and
800 local adaptive divergence of *Dracaena cambodiana* (Liliaceae) populations
801 associated with historical population bottlenecks and natural selection: an
802 endangered long-lived tree endemic to Hainan Island, China. *Plant Biol.* **14(5)**,
803 828-838.

804 **Zheng, Q.A., Zhang, Y.J., Yang, C.R.** (2006). A new meta-homoisoflavane from the
805 fresh stems of *dracaena cochinchinensis*. *J Asian Nat. Prod. Res.* **8(6)**, 571-577.

806

807 **Figure Legends**

808 **Fig. 1 Morphology and genome features of *D. cochinchinensis*.** (a) morphology of
809 *D. cochinchinensis* and dragon’s blood. (b) Genome features and syntenic information.
810 The circus map, from outside to inside, shows GC density, repeat density and gene
811 density, respectively.

812 **Fig. 2 Comparative Genomic Analysis.** (a) Phylogenetic tree of 17 plant species.
813 Divergence time (Mya) estimations are indicated by the blue numbers beside the
814 branch nodes. The number of gene-family contraction and expansion is indicated by
815 green and red numbers respectively. (b) Distribution of 4DTv for paralogous genes
816 among *D. cochinchinensis*, *A. officinalis*, *D. officinalis* and *A. shenzhenica*. (c)
817 Estimation of LTR activity showed there are two insertion burst events in *D.*
818 *cochinchinensis* genome. (mya). (d) In this recent burst of activity, the majority of
819 elements responsible were both autonomous (3359 copies of Ty1-copia and 16897
820 copies of Ty3-gypsy) and non-autonomous LTRs (5861 copies), with peaks of activity
821 <6 mya. (e) Venn diagram shows shared and unique orthologues among the four
822 asparagus species, including *D. cochinchinensis*, *A. officinalis*, *D. officinalis* and *A.*
823 *shenzhenica*, and genes unique to *D. cochinchinensis*. (f-g) Phylogeny of the SAURs
824 and ZOGTs family in *D. cochinchinensis*.

Fig. 3 Histological and Metabolic Profilings during dragon's blood formation. (a) Histological images during the formation of dragon's blood in *D. cochinchinensis*. (The yellow and red substances in the picture are dragon's blood resin. VB = Vascular Bundle, PC: parenchymal cell.) (b) Typical total ion chromatograms of *D. cochinchinensis*. (a: positive TIC before dragon's blood formation; b: positive TIC after dragon's blood formation; c: negative TIC before dragon's blood formation; d: negative TIC after dragon's blood formation.) (c) Heat map of identified metabolites of *D. cochinchinensis* before and after the formation of dragon's blood (BFD 1~6 represents six duplications of the analytical materials before formation of dragon's blood; AFD1~6 represents six duplications of the analytical materials after formation of dragon's blood).

Fig. 4 Differentially expressed metabolic pathways during dragon's blood formation. Transcriptome analysis identifies differentially expressed pathways during dragon's blood formation. The up-regulation pathway was above the oblique line, and the down-regulation pathway was below the oblique line. The circle size represents the number of genes, and the coordinate value represents the average up-regulation or down-regulation times.

Fig. 5 Flavones biosynthesis in *D. cochinchinensis*. The flavones biosynthesis pathway and expression profiles (reads per kilobase per million reads mapped; RPKM) of the genes involved in rubber biosynthesis. PAL, phenylalanine ammonia-lyase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumarate CoA ligase; CHS, chalcone synthase; CHI, chalcone isomerase; FNS, flavone synthase; F3H, flavanone 3-hydroxylase; DFR, dihydroflavonol 4-reductase; FLS, flavonol synthase; LAR, leucoanthocyanidin reductase; PPO, polyphenol oxidase; OMT: O-methyltransferase.

Fig. 6 A propose model of ROS as a specific regulating signal in the adaptation of *D. cochinchinensis*. External wound leads to ROS burst and excessive ROS causes DNA damage. Fortunately, most of the positive selection genes in the genome are enriched in DNA-repair pathways, such as base excision repair (BER), nucleotide excision repair (NER), mismatch repair (MMR), which play crucial roles in maintaining genomic integrity. Meanwhile, ROS as a signal molecule activates

855 transcription factors, thus activating the flavonoid synthesis pathway to synthesize a
 856 large amount of flavonoids. Flavonoid is the main component of dragon's blood, as a
 857 defensive substance to resist further damage. In turn, Flavonoids together with
 858 enzymes in ROS scavenging system scavenge the excess ROS accumulated in vivo to
 859 maintain homeostasis.

860

(a) (b)

D. cochinchinensis

wounded stem with resin

resin extract

(a)

bioRxiv preprint doi: <https://doi.org/10.1101/2022.02.08.479636>; this version posted February 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(b)

(c)

(d)

(e)

(f)

(g)

(a)

(b)

(c)