

Supporting Information for

Short communication

Whole-genome sequencing and analysis of the Chinese herbal plant *Gelsemium elegans*

Yisong Liu^{a,b,c,†}, Qi Tang^{b,†}, Pi Cheng^{b,†}, Mingfei Zhu^d, Hui Zhang^d, Jiazhe Liu^d, Mengting Zuo^c, Chongyin Huang^c, Changqiao Wu^b, Zhiliang Sun^{a,b,c,*}, Zhaoying Liu^{a,b,c,*}

^a*Hunan Engineering Technology Research Center of Veterinary Drugs, Hunan Agricultural University, Changsha 410128, China*

^b*Hunan Key Laboratory of Traditional Chinese Veterinary Medicine, Hunan Agricultural University, Changsha 410128, China*

^c*College of Veterinary Medicine, Hunan Agricultural University, Changsha 410128, China*

^d*Nextomics Biosciences Institute, Wuhan 430000, China*

Received 15 April 2019; received in revised form 27 June 2019; accepted 26 July 2019

*Corresponding authors. Tel./fax: +86 731 84635054.

E-mail addresses: sunzhiliang1965@aliyun.com (Zhiliang Sun), liu_zhaoying@hunau.edu.cn (Zhaoying Liu).

†These authors made equal contributions to this work.

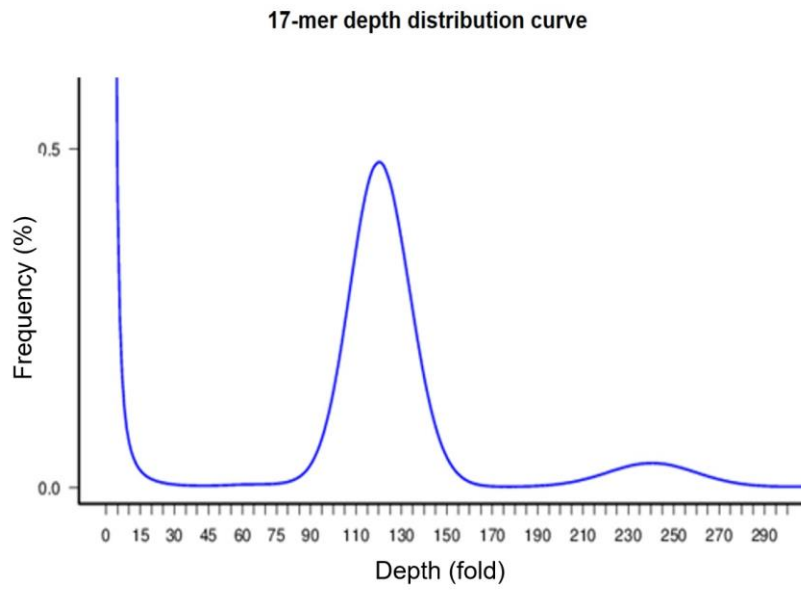


Figure S1 Frequency distribution of the 17-mer graph analysis used to estimate the size of the *Gelsemium elegans* genome

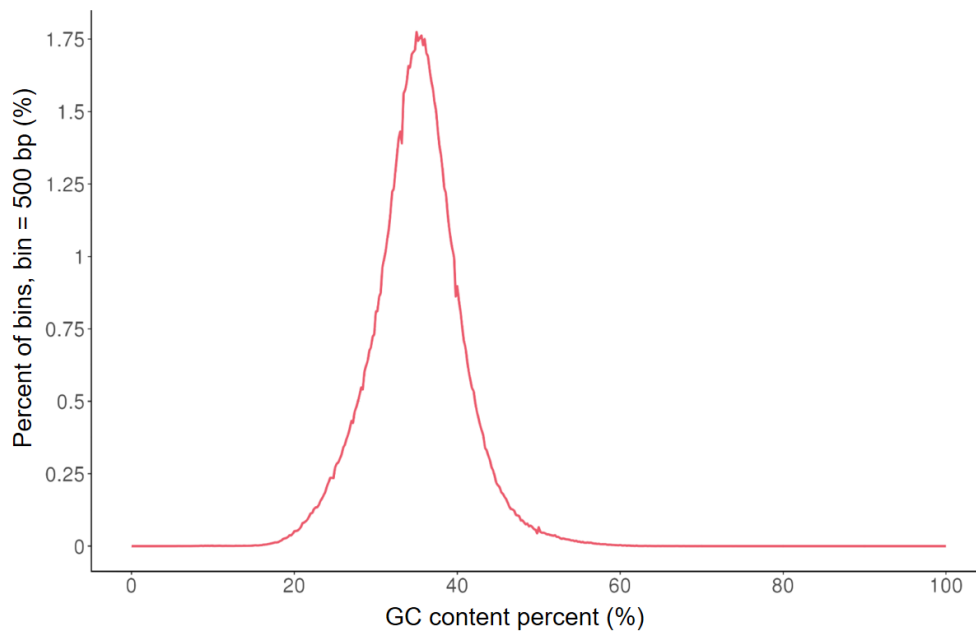


Figure S2 GC content distribution of the *Gelsemium elegans* genome. The GC content was established using 500 bp sliding windows.

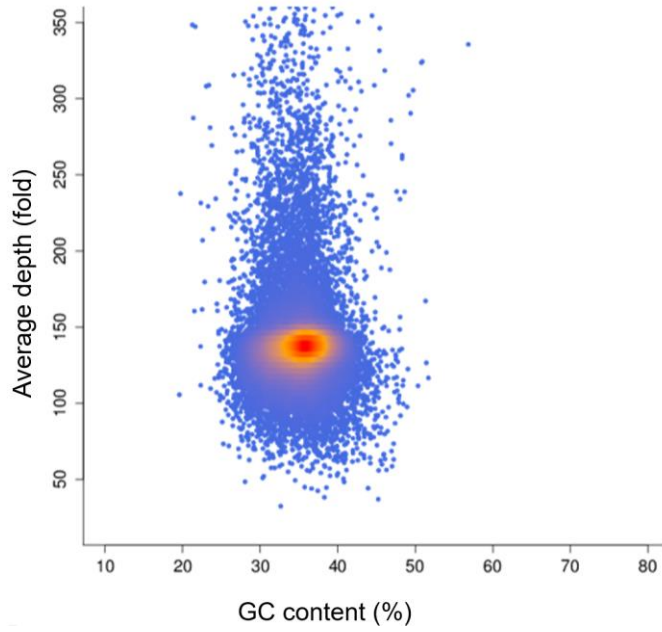


Figure S3 The GC depth distribution of the *Gelsemium elegans* genome.

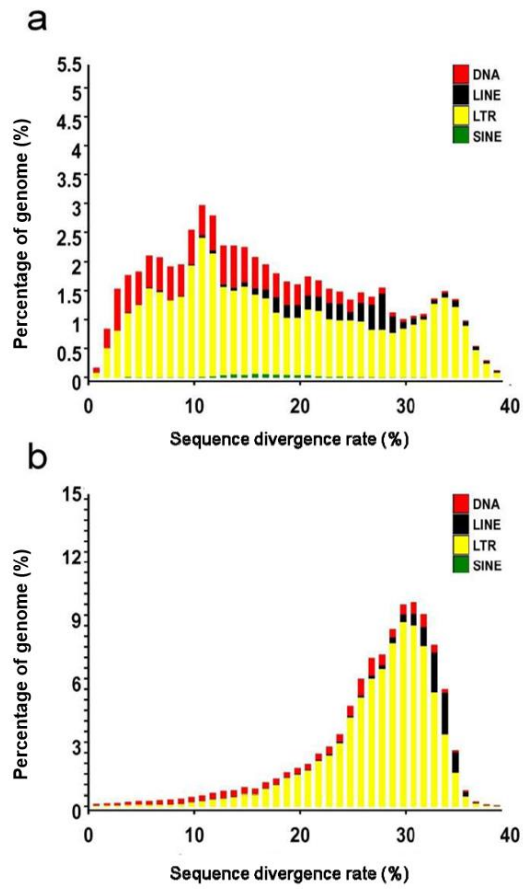


Figure S4 RepeatMasker analysis of the *Gelsemium elegans* genome assembly. (a) *De novo*; (b) Rebase.

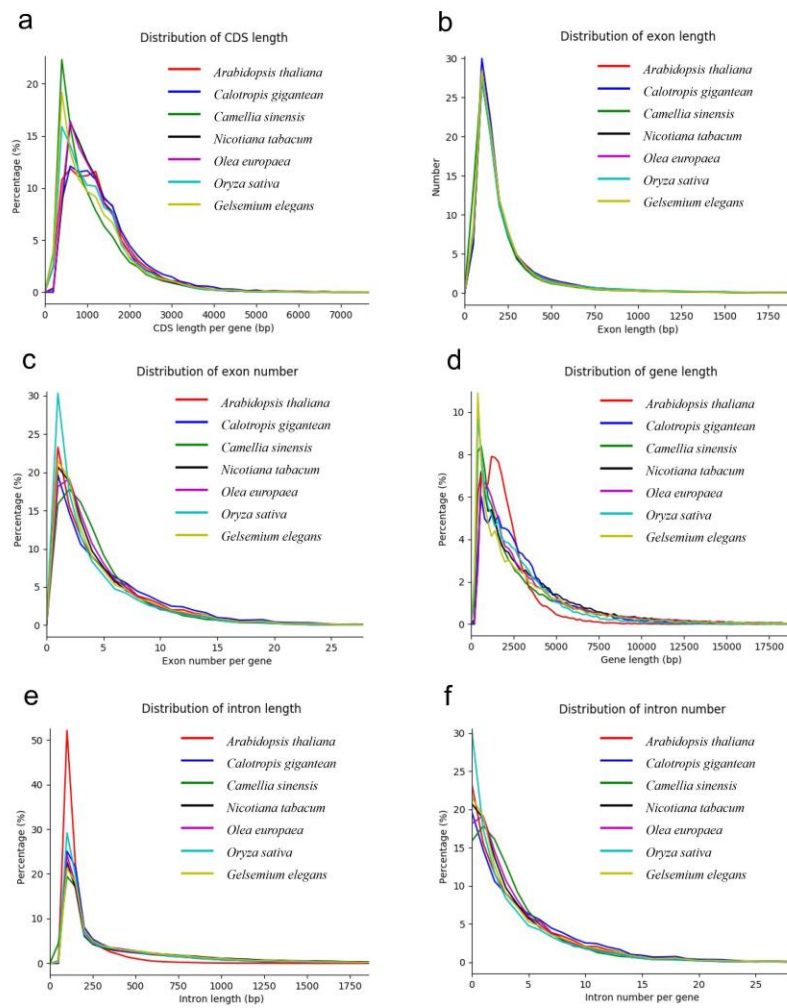


Figure S5 Comparison of gene structure characteristics in *Gelsemium elegans* to that in other plants. (a) CDS length; (b) exon length; (c) exon number; (d) gene length; (e) intron length; (f) intron number.

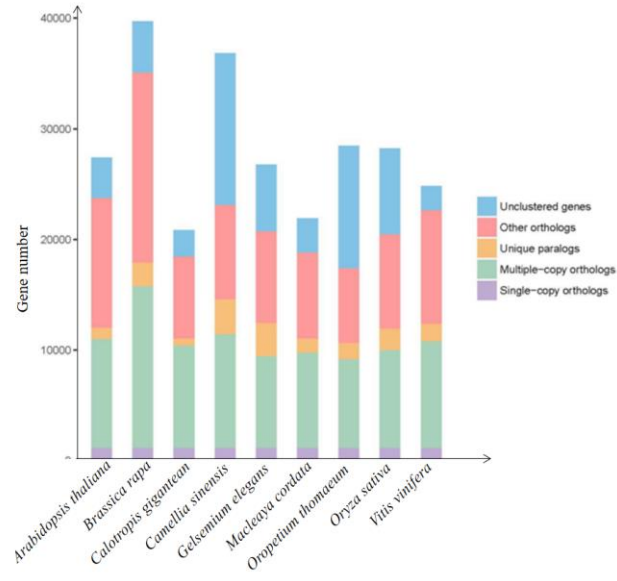


Figure S6 Gene orthology determined by comparing genomes using the OrthoMCL software.

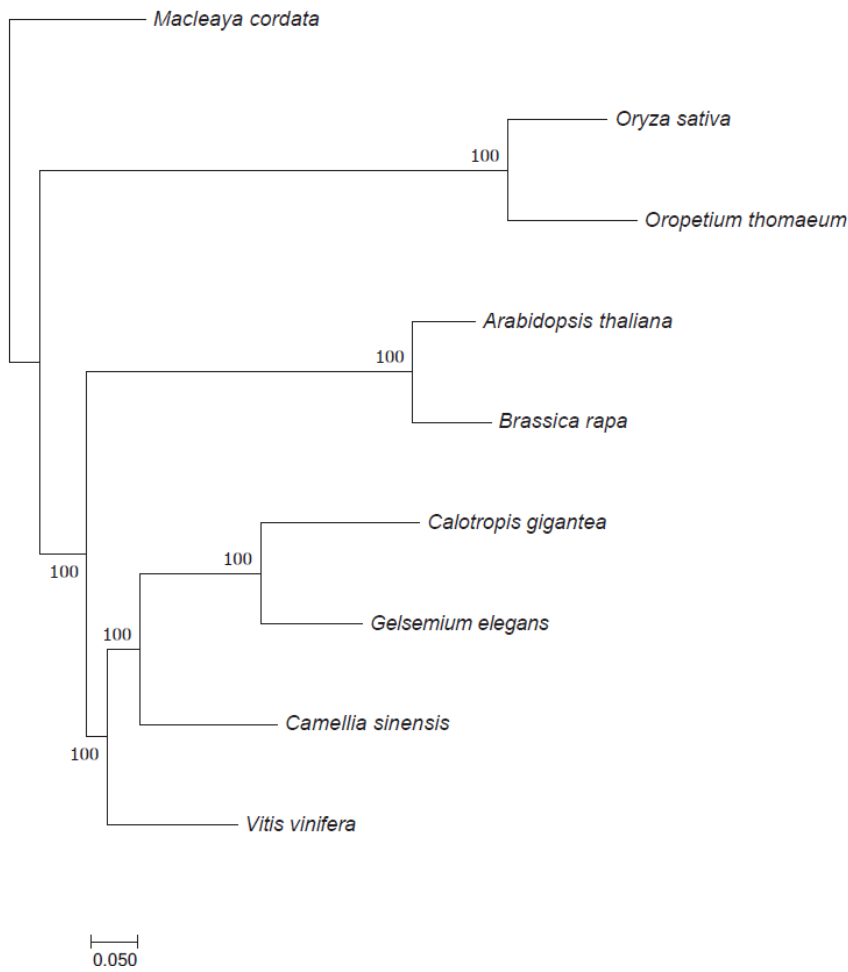


Figure S7 The maximum-likelihood phylogeny of *Gelsemium elegans* and 8 other plants. The bootstrap value is shown at each node.

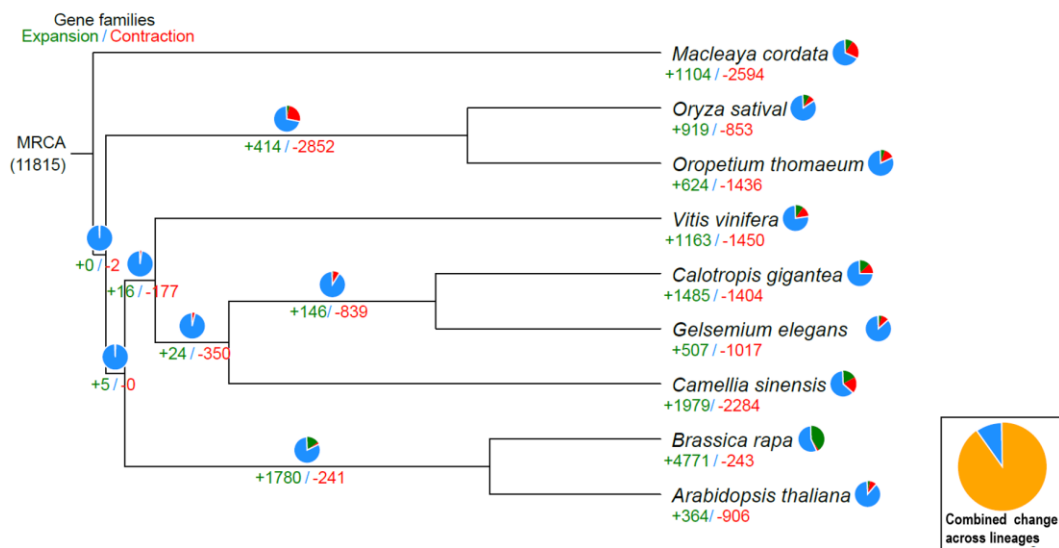


Figure S8 Gene family expansions and contractions in *Gelsemium elegans* and 8 other plants.

Table S1 Sequencing statistics from the nanopore sequencing.

Total pass reads bases (Gb)	Total pass reads number	Pass reads mean length (kb)	Pass reads max length (kb)	Pass reads N50 length (kb)	Pass reads > 10kb ratio	Pass reads > 40kb ratio
56.11	3,663,275	14.59	153.61	36.22	51.49%	2.96%

Table S2 Summary of the transcriptome and their mapping rate on the genome assembly.

Sample	Tissue	Total reads number	Mapped reads number	Mapping rate (%)
R1704065	Leaf	23,785,097	21,901,917	92.08
R1704066	Leaf	22,455,658	20,668,723	92.04
R1704067	Leaf	25,131,247	23,132,168	92.04
R1704068	Root	20,748,256	19,109,832	92.10
R1704069	Root	31,713,585	29,130,646	91.85
R1704070	Root	28,468,029	26,068,705	91.57

Table S3 Summary of clean reads generated from transcriptome sequencing by Illumina HiseqX Ten.

Sample	Clean reads	Clean bases	Q30 rate (%)
GEflower_a	67,922,518	10,168,126,580	0.94
GEflower_b	70,570,098	10,546,666,299	0.96
GEflower_c	89,613,594	13,392,480,334	0.96
GEleaf_a	81,439,082	12,189,647,061	0.97
GEleaf_b	75,611,228	11,314,844,846	0.97
GEleaf_c	85,305,224	12,764,365,164	0.97
GEroot_a	69,298,662	10,368,060,758	0.97
GEroot_b	82,342,200	12,322,538,800	0.95
GEroot_c	71,294,892	10,674,266,631	0.95
GEstem_a	73,210,680	10,948,432,092	0.96
GEstem_b	72,638,140	10,858,155,675	0.96
GEstem_c	69,291,652	10,359,264,597	0.96

Table S4 Estimation of genome size based on 17-mer statistics.

<i>k</i> -mer value	<i>k</i> -mer number	<i>k</i> -mer depth	Genome size (bp)	Heterozygosity rate (%)
17	40,563,763,104	120	338,031,359	0.38

Table S5 Statistics of initial assembly results.

Statistic type	Contig length (bp)	Contig number
N50	10,143,347	12
N60	7,085,978	15
N70	5,466,148	21
N80	1,321,283	34
N90	236,259	100
Longest	18,101,142	1
Total	331,825,569	738
Length \geq 1kb	331,825,569	738
Length \geq 2kb	331,825,569	738
Length \geq 5kb	331,825,569	738

Table S6 Summary of the genome final assembly after polish.

Statistic type	Contig length (bp)	Contig number
N50	10,238,528	12
N60	7,156,190	15
N70	5,523,542	21
N80	1,337,470	34
N90	239,091	100
Longest	18,270,620	1
Total	335,134,780	738
Length \geq 1kb	335,134,780	738
Length \geq 2kb	335,134,780	738
Length \geq 5kb	335,134,780	738

Table S8 The contig cluster of 8 pseudo-chromosomes length.

Pseudo-chromosome	Size	Scaffold number
Chr01	52,329,522	63
Chr02	45,875,187	124
Chr03	44,416,115	69
Chr04	40,473,783	6
Chr05	38,673,583	55
Chr06	38,231,740	72
Chr07	36,223,536	91
Chr08	36,075,847	65
Total	332,299,313	545

Table S9 Summary of BUSCO analysis results according to assembly.

Type	Number	Compare to the total BUSCO groups (%)
Complete BUSCOs (C)	1338	92.9
Complete and single-copy BUSCOs (S)	1308	90.8
Complete and duplicated BUSCOs (D)	30	2.1
Fragmented BUSCOs (F)	29	2.0
Missing BUSCOs (M)	73	5.1
Total BUSCO groups searched	1440	100

Table S10 Summary of the SSR search results.

Repeat type	Unit size (repeat number)	Number
Mono-nucleotide	1 (≥ 10)	134,047
Di-nucleotide	2 (≥ 6)	29,668
Tri-nucleotide	3 (≥ 5)	9,336
Tetra-nucleotide	4 (≥ 5)	1,557
Penta-nucleotide	5 (≥ 5)	409
Hexa-nucleotide	7 (≥ 5)	444

Table S11 The repeat of the *G. elegans* genome assembly.

Type	Repeat size (bp)	Percentage of genome (%)
<i>Trf</i>	32,316,704	9.64
RepeatProteinMask	46,644,858	13.92
RepeatMasker (Rebase)	43,991,917	13.13
RepeatMasker (Mips-REdat)	41,664,552	12.43
<i>Denovo</i> (RepeatModeler) <i>novo</i> (RepeatModeler)	89,166,197	26.61
Total	144,642,909	43.16

Table S12 Repeat annotation of the *Gelsemium elegans* genome assembly.

Type*	Rebase TEs		Mips-REdat TEs		TE proteins		RepeatModeler		Combined TEs	
	Length (Mb)	% in genome	Length (Mb)	% in genome	Length (Mb)	% in genome	Length (Mb)	% in genome	Length (Mb)	% in genome
DNA	3,771,418	1.13	1,074,439	0.32	3,848,392	1.15	10,689,208	3.19	12,050,722	3.6
LINE	3,594,794	1.07	1,519,190	0.45	6,342,959	1.89	4,580,193	1.37	7,176,479	2.14
LTR	36,260,317	10.82	38,907,712	11.61	36,289,011	10.83	40,751,378	12.16	80,110,096	23.9
SINE	5,000	0	11,617	0	0	0	670,701	0.2	618,147	0.18
Other	344,384	0.11	151,594	0.05	164,496	0.05	0	0	26,856,213	8.02
Unknown	16,004	0	0	0	0	0	33,125,998	9.88	17,831,252	5.32
Total	43,991,917	13.13	41664552	12.43	46,644,858	13.92	89,166,197	26.61	144,642,909	43.16

*DNA, DNA transposon; LINE, long interspersed nuclear element; TEs, transposable elements; SINE, short interspersed nuclear element; LTR, long terminal repeat.

Table S13 Summary of non-protein-coding gene annotations in the *Gelsemium elegans* genome assembly.

Type	Copy number	Average length (bp)	Total length (bp)	Percentage (%) of genome
rRNA	279	1079.33	301,132	0.065048
18S	29	1899.62	55,089	0.011900
28S	39	5680.03	221,521	0.047851
5.8S	20	153.60	3,072	0.000664
5S	191	112.30	21,450	0.004633
snRNA	1257	106.79	134237	0.028997
CD-box	1150	103.59	119,129	0.025733
HACA-box	28	131.00	3,668	0.000792
Splicing	79	144.81	11,440	0.002471
miRNA	208	122.62	25,505	0.005509
tRNA	531	74.88	39,761	0.008589

Table S14 Gene annotation statistics of the *Gelsemium elegans* genome assembly.

Gene set		Total number of CDS	Average transcript length (bp)	Average CDS length (bp)	Average exon number pergene	Average exon length (bp)	Average intron length (bp)
<i>De novo</i>	Augustus	30,257	2,323.9	962.68	4.61	208.74	376.88
	Gene ID	26,980	5,890.56	967.15	5.12	188.95	1,195.44
	Snap	40,763	6,687.38	912.33	7.1	128.46	946.39
Homology	<i>A. thaliana</i>	24,855	2,840.18	967.74	3.79	255.27	670.9
	<i>C. gigantea</i>	21,287	3,350.2	932.06	3.64	256.01	915.74
	<i>C. sinensis</i>	22,049	3,573.91	1,067.08	4.4	242.67	737.9
	<i>N. tabacum</i>	22,963	2,368.12	885.53	3.32	266.66	638.84
	<i>O. europaea</i>	22,903	3,004.06	1,029.97	4.06	253.6	644.88
	<i>O. sativaL</i>	22,907	2,309.66	877.79	3.17	277.28	661.18
Final set	EVM	26,768	3,961.71	1,088.11	4.98	218.63	722.54

Table S15 Comparative gene statistics.

Specie	Totalnumero fgene	Average transcript length (bp)	Average CDS length (bp)	Average exon number per gene	Average exon length (bp)	Average intron length (bp)
<i>G.elegans</i>	26,768	3,961.71	1,088.11	4.98	218.63	722.54
<i>A. thaliana</i>	27,416	1,875.54	1,220.7	5.14	237.3	158.02
<i>C. gigantea</i>	20,831	3,126.12	1,345.81	5.84	230.49	367.92
<i>C. sinensis</i>	36,841	3,285.39	988.26	4.69	210.89	623.17
<i>N.tabacum</i>	60,288	3,718.43	1,163.23	4.74	245.58	683.82
<i>O.europaea</i>	39,579	3,656.34	1,168.54	4.78	244.33	657.7
<i>O.sativaL</i>	28,225	2,454.91	1,086.74	4.27	254.55	418.5

Table S16 Functional annotation of predicted genes of *Gelsemium elegans*.

	Database	Number	Annotated percent (%)
Annotation	Swissprot	16,817	62.83
	TrEMBL	22,545	84.22
	KEGG	7,919	29.58
	GO	12,832	47.94
	COG	12,552	46.89
	InterProScan	18,280	68.29
Total	Annotated	22,636	84.56
	Gene	26,768	–

–Not applicable

Table S17 Summary of BUSCO analysis results according to prediction.

Type	Number	Annotated percent (%)
Complete BUSCOs (C)	1375	95.5
Complete and single-copy BUSCOs (S)	1349	93.7
Complete and duplicated BUSCOs (D)	26	4.8
Fragmented BUSCOs (F)	16	1.1
Missing BUSCOs (M)	49	3.4
Total BUSCO groups searched	1440	–

–Not applicable

Table S18 Summary statistics of gene families in 9 plant species.

Specie	Gene number	Gene family	Unclustered gene number	Unclustered gene family	Unique family number	Average gene
<i>Arabidopsis thaliana</i>	27,416	23,776	3,640	15,745	347	1.51
<i>Brassica rapa</i>	39,717	35,083	4,634	16,051	602	2.19
<i>Calotropis gigantea</i>	20,831	18,474	2,357	12,313	216	1.50
<i>Camellia sinensis</i>	36,841	23,127	13,714	12,543	953	1.84
<i>Gelsemium elegans</i>	26,768	20,755	6,013	13,792	903	1.50
<i>Macleaya cordata</i>	21,907	18,845	3,062	12,066	311	1.56
<i>Oropetium thomaeum</i>	28,446	17,425	11,021	12,783	621	1.36
<i>Oryza sativa</i> L	28,225	20,499	7,726	13,120	633	1.56
<i>Vitis vinifera</i>	24,821	22,644	2,177	13,156	390	1.72

Table S19 Candidate positively selected genes in the *Gelsemium elegans* lineage

Group	<i>P</i> value	Positive site number	Gene	Swissprot function
ortholog05642	0.000629	1	evm.model.Contig5.1222	Phosphatidylinositol/phosphatidylcholine transfer protein SFH12
ortholog00678	0.037136	1	evm.model.Contig9.685	Cullin-associated NEDD8-dissociated protein 1
ortholog09581	0.036572	3	evm.model.Contig22.161	Checkpoint protein hus1 homolog
ortholog07073	1.86×10^{-7}	11	evm.model.Contig9.218	Zinc finger CCCH domain-containing protein 1
ortholog15048	0.044781	4	evm.model.Contig1.330	Uncharacterized protein C1683.06c
ortholog06103	0.018971	1	evm.model.Contig2.1905	WD repeat-containing protein 3
ortholog19149	0.000184	2	evm.model.Contig9.557	Probable Xaa-Pro aminopeptidase 3
ortholog05212	0.01273	1	evm.model.Contig11.399	Protein CLP1 homolog
ortholog05899	0.018062	2	evm.model.Contig22.659	E3 ubiquitin-protein ligase MARCH1
ortholog20263	0.037057	1	evm.model.Contig3.1279	Condensin-2 complex subunit G2
ortholog02837	0.029401	1	evm.model.Contig20.646	RNA exonuclease 4
ortholog06318	0.005725	1	evm.model.Contig3.468	Uncharacterized tRNA/rRNA methyltransferaseYsgA
ortholog13296	0.024789	2	evm.model.Contig15.297	PsbP domain-containing protein 6, chloroplastic
ortholog15845	0.009238	1	evm.model.Contig19.432	Protein CHROMATIN REMODELING 25
ortholog00407	0.023281	1	evm.model.Contig22.289	Methyltransferase-like protein 22
ortholog03122	0.028539	46	evm.model.Contig3.2082	Pentatricopeptide repeat-containing protein At5g14080
ortholog10186	0.03142	1	evm.model.Contig7.1120	Protein MIZU-KUSSEI 1
ortholog08097	0.001847	2	evm.model.Contig2.1862	Pentatricopeptide repeat-containing protein At5g03800
ortholog01318	0.027228	2	evm.model.Contig6.871	Uncharacterized protein At2g39910
ortholog06837	9.44×10^{-5}	4	evm.model.Contig3.2020	COP9 signalosome complex subunit 3
ortholog12455	0.000311	2	evm.model.Contig6.1339	Phosphomannomutase/phosphoglucomutase

ortholog17353	0.009301	2	evm.model.Contig4.615	Replication factor C subunit 1
ortholog11825	0.034555	1	evm.model.Contig19.2	AT-rich interactive domain-containing protein 3
ortholog00415	0.043436	1	evm.model.Contig5.74	DEAD-box ATP-dependent RNA helicase 24
ortholog10208	0.047319	1	evm.model.Contig4.421	DNA polymerase epsilon catalytic subunit A
ortholog03255	1.31×10^{-6}	14	evm.model.Contig20.188	Protein CWC15 homolog A
ortholog06371	0.046945	2	evm.model.Contig8.491	Protein THYLAKOID FORMATION1, chloroplastic
ortholog20051	9.04×10^{-7}	7	evm.model.Contig1.81	Probable prefoldin subunit 3
ortholog16201	0.006934	2	evm.model.Contig6.1545	Gamma-soluble NSF attachment protein
ortholog05551	0.033047	1	evm.model.Contig7.1244	Pentatricopeptide repeat-containing protein At1g19525
ortholog09635	0.045576	1	evm.model.Contig1.1299	Probable acyl-activating enzyme 18, peroxisomal
ortholog09255	0.000123	2	evm.model.Contig6.1532	Peptide chain release factor PrfB1, chloroplastic
ortholog00196	0.034474	1	evm.model.Contig3.1139	ATP-dependent Clp protease proteolytic subunit 3, chloroplastic
ortholog03004	0.034315	1	evm.model.Contig14.548	Putative DEAD-box ATP-dependent RNA helicase 29
ortholog02844	0.024832	1	evm.model.Contig4.556	Ribonuclease inhibitor
ortholog00001	0.00854	1	evm.model.Contig28.10	Vacuolar protein-sorting-associated protein 11 homolog
ortholog17130	0.015842	5	evm.model.Contig4.438	Beta-1,6-galactosyltransferase GALT29A
ortholog13337	0.014621	1	evm.model.Contig3.2051	Protease HtpX homolog 2
ortholog13201	2.07×10^{-8}	16	evm.model.Contig3.2119	Histone-lysine <i>N</i> -methyltransferase, H3 lysine-9 specific SUVH4
ortholog01743	0.002436	3	evm.model.Contig2.764	APO protein 2, chloroplastic
ortholog08204	8.93×10^{-8}	14	evm.model.Contig1.481	Protein CDC73 homolog
ortholog04915	0.000112	6	evm.model.Contig10.606	Probable dolichyl pyrophosphate Glc1Man9GlcNAc2 alpha-1,3-glucosyltransferase
ortholog15219	4.04×10^{-5}	3	evm.model.Contig8.668	Membrane metalloprotease ARASP, chloroplastic
ortholog01622	0.000561	7	evm.model.Contig20.686	Protein DJ-1 homolog B
ortholog02404	0.002514	2	evm.model.Contig10.869	CRS2-associated factor 1, mitochondrial
ortholog05992	1.02×10^{-8}	7	evm.model.Contig19.400	Armadillo repeat-containing protein 6

ortholog04006	0.030848	1	evm.model.Contig11.815	Masparidin
ortholog08100	2.76×10 ⁻⁵	11	evm.model.Contig7.327	Uncharacterized protein C57A10.07
ortholog00483	0.000683	2	evm.model.Contig10.270	50S ribosomal protein L1, chloroplastic
ortholog07104	0.010222	1	evm.model.Contig7.780	Protein CLT2, chloroplastic
ortholog00621	0.012093	2	evm.model.Contig11.172	WD repeat-containing protein WRAP73
ortholog05207	0.046867	1	evm.model.Contig7.1252	Uncharacterized sugar kinase slr0537
ortholog08484	0.015047	1	evm.model.Contig14.275	Probable dolichyl-diphosphooligosaccharide-protein glycosyltransferase subunit 3B
ortholog00141	0.000148	2	evm.model.Contig11.732	Chaperone protein dnaJ 13
ortholog07284	0.010601	7	evm.model.Contig4.646	Putative pentatricopeptide repeat-containing protein At1g02420
ortholog06462	0.009994	1	evm.model.Contig7.855	Chaperone protein dnaJ
ortholog01779	0.021626	1	evm.model.Contig6.654	Hydroxyacylglutathione hydrolase cytoplasmic
ortholog03952	0.000704	1	evm.model.Contig5.404	Methylcrotonoyl-CoA carboxylase subunit alpha, mitochondrial
ortholog04243	0.01528	8	evm.model.Contig2.1229	rRNA methyltransferase 1, mitochondrial
ortholog08493	0.0459	3	evm.model.Contig6.1588	Oxygen-dependent coproporphyrinogen-III oxidase, chloroplastic
ortholog07701	0	11	evm.model.Contig11.1048	Probable LRR receptor-like serine/threonine-protein kinase At3g47570
ortholog16622	0.037958	1	evm.model.Contig12.930	Pentatricopeptide repeat-containing protein At3g18020
ortholog05078	0.028468	1	evm.model.Contig2.1159	Probable acyl-[acyl-carrier-protein]-UDP- <i>N</i> -acetylglucosamine- <i>O</i> -acyltransferase, mitochondrial
ortholog01700	0.022949	2	evm.model.Contig5.356	Chloroplast stem-loop binding protein of 41 kDa, chloroplastic
ortholog04031	0.011329	2	evm.model.Contig6.1175	Ribonuclease H2 subunit A
ortholog08157	0.029378	1	evm.model.Contig10.977	Dynamin-like protein ARC5
ortholog12998	0.02605	1	evm.model.Contig6.178	Cell division control protein 14
ortholog09954	0.025488	2	evm.model.Contig6.235	Probable protein phosphatase 2C 76
ortholog03091	0.016347	1	evm.model.Contig7.468	Uncharacterized aarF domain-containing protein kinase At1g71810,

ortholog13718	0.031591	2	evm.model.Contig15.176	chloroplastic Probable arabinosyltransferase ARAD1
ortholog02624	0.002228	6	evm.model.Contig10.837	Fasciclin-like arabinogalactan protein 8
ortholog18122	0.004715	7	evm.model.Contig8.315	DDB1- and CUL4-associated factor 13
ortholog16265	0.027547	1	evm.model.Contig15.355	DNA-repair protein XRCC1
ortholog18609	0.006866	2	evm.model.Contig8.853	Fasciclin-like arabinogalactan protein 7
ortholog18594	0.014358	1	evm.model.Contig2.257	Probable plastid-lipid-associated protein 13, chloroplastic
ortholog08194	7.07×10^{-5}	3	evm.model.Contig7.1228	Probable methyltransferase PMT28
ortholog01210	0.046787	1	evm.model.Contig5.783	Uncharacterized CRM domain-containing protein At3g25440, chloroplastic

Table S20 The genes involved in indole alkaloids biosynthesis.

Pathway	Gene	Enzyme No.	Number	Seq No.	
MEP Pathway	1-Deoxy-D-xylulose-5-phosphate synthase	[EC:2.2.1.7]	4	Contig1.321, Contig2.892, Contig3.463, Contig21.325	
	1-Deoxy-D-xylulose-5-phosphate reductoisomerase	[EC:1.1.1.267]	1	Contig12.343	
	2-C-Methyl-D-erythritol 4-phosphate cytidyltransferase	[EC:2.7.7.60]	1	Contig9.677	
	4-Diphosphocytidyl-2-C-methyl-D-erythritol kinase	[EC:2.7.1.148]	1	Contig4.740	
	2-C-Methyl-D-erythritol 2,4-cyclodiphosphate synthase	[EC:4.6.1.12]	1	Contig21.242	
	(E)-4-Hydroxy-3-methylbut-2-enyl-diphosphate synthase (ferredoxin)	[EC:1.17.7.1]	1	Contig5.1404	
	4-Hydroxy-3-methylbut-2-enyl diphosphate reductase	[EC:1.17.1.2]	1	Contig19.347	
	MVA Pathway	Acetyl-CoA acetyltransferase	[EC:2.3.1.9]	2	Contig23.94, Contig32.7
		Hydroxymethylglutaryl-CoA synthase	[EC:2.3.3.10]	1	Contig11.883
		3-Hydroxy-3-methylglutaryl-coenzyme A reductase	[EC:1.1.1.34]	3	Contig3.680, Contig7.1553, Contig19.3
Mevalonate kinase		[EC:2.7.1.36]	1	Contig9.284	
5-Phosphomevalonate kinase		[EC:2.7.4.2]	1	Contig11.139	
Diphosphomevalonatedecarboxylase		[EC:4.1.1.33]	1	Contig6.1194	
Isopentenyl-diphosphate delta-isomerase		[EC:5.3.3.2]	2	Contig6.1499, Contig6.1500	
Monoterpenoid biosynthesis		Geranylgeranyldiphosphatesynthase	[EC:2.5.1.1]	5	Contig20.363, Contig3.1039, Contig3.522, Contig7.1184, Contig19.86
		Geraniol synthase	[EC:3.1.7.11]	1	Contig8.533
		Geraniol 8-hydroxylase(CYP76B6)	[EC:1.14.14.83]	4	Contig3.1451, Contig3.1452, Contig11.148, Contig11.809
	8-Hydroxygeraniol dehydrogenase	[EC:1.1.1.324]	9	Contig1.1602, Contig3.473, Contig3.495, Contig3.496, Contig3.502, Contig3.503, Contig3.514, Contig3.515, Contig10.255	
	7-Deoxyloganate 7-hydroxylase	[EC:1.14.14.85]	0	Not Found	
	Secologaninsynthase	[EC:1.14.19.62]	4	Contig1.875, Contig10.302, Contig20.787, Contig20.789	
	S-Adenosyl-L-methionine-dependent methyltransferase	[EC:2.1.1.50]	2	Contig5.182, Contig7.1229	
	Indole alkaloid biosynthesis	L-Tryptophan decarboxylase	[EC:4.1.1.105]	1	Contig8.637
		Strictosidinesynthase	[EC:4.3.3.2]	17	Contig2.1043,

			Contig3.748, Contig3.749, Contig4.259, Contig4.262, Contig4.611, Contig8.636, Contig8.825, Contig12.325, Contig12.326, Contig12.327, Contig12.756, Contig12.757, Contig15.402, Contig15.403, Contig16.448, Contig16.449
Strictosidine- <i>O</i> -beta-D-glucosidase	[EC:3.2.1.105]	5	Contig3.622, Contig9.61, Contig9.70, Contig9.75, Contig9.82
Geissoschizinedehydrogenase	[EC:1.3.1.36]	0	Not Found
Sarpagan bridge enzyme		0	Not Found
Polyneuridine-aldehyde esterase	[EC:3.1.1.78]	15	Contig1.1627, Contig1.1628, Contig3.488, Contig3.491, Contig3.500, Contig3.539, Contig3.540, Contig3.541, Contig3.542, Contig4.1007, Contig4.1008, Contig4.1010, Contig9.60, Contig11.1021, Contig26.32
Vinorine synthase	[EC:2.3.1.160]	20	Contig1.1102, Contig1.926, Contig1.927, Contig1.928, Contig1.930, Contig5.1734, Contig5.1734, Contig6.1902, Contig6.805, Contig9.136, Contig9.137, Contig9.139, Contig10.686, Contig11.1060, Contig12.263, Contig12.93, Contig12.96, Contig12.97,

Raucaffricine- β -D-glucosidase	[EC:3.2.1.21]	9	Contig13.599, Contig15.392 Contig2.805, Contig2.806, Contig2.809, Contig9.66, Contig9.67, Contig9.69, Contig9.86, Contig26.100, Contig26.99
---------------------------------------	---------------	---	---

Total

113

Table S21 Versions and main parameters of the software used in this study.

Software	Version	Parameter
Canu	1.3	length_cutoff, 11kb; length_cutoff_pr, 11.5 kb
WTDBG	1.2.8	Default
PBJelly	PBSuite_14.9.9	Default
BWA	0.7.12-r1039	Default
Pilon	1.20	--changes --vcf --diploid --fix bases --threads 32 --mindepth 10
HISAT	2.0.0-beta	hisat2 -N 1
RepeatMasker	open-4.0.6	RepeatMaskernolow -q -no_is -gff -norna -parallel 1 -engine wublast -species Viridiplantae
RepeatModeler	1.0.4	RepeatModeler -pa 30 -database zhongrenkeyuan.contigs.2.fasta -engine wublast
tRNAscan-SE	1.3.1	tRNAscan-SE -i -o -m
RNAmmer	1.2	rnammer -S euk -m lsu,ssu,tsu -gff
GeneWise	wise2-4-1	genewise -genesf -gff -sum
Augustus	3.0	augustus-UTR=off --gff3=on --genemodel=complete --strand=both --min_intron_len=15
GeneID	1.4	geneid -3 -P arabidopsis.param.Aug_4_2004
GlimmerHMM	3.0.4	default
SNAP	2013/2/16	default
MCMCTREE	PAML v4.9e	Clock=2, RootAge≤1.73, model=7, BDparas=110, kappa_gamma = 62, alpha_gamma = 11, rgene_gamma = 23.18, sigma2_gamma = 14.5
Codeml	PAML v4.9e	Model A: model = 2, NSsites = 2, fix_omega = 0; Model A1: model = 2, NSsites = 2, fix_omega = 1, omega = 1
CAFE	3.0	load -p 0.05 -r 10,000 , lambda -s

Table S22 Genome summary of *Gelsemium elegans* and closely related species.

Specie	Assembly Genome size (Mb)	Contig N50 (kb)	Scaffold N50 (Mb)	Complete BUSCOs (%)	Protein-coding genes
<i>Gelsemium elegans</i>	335.13	10,230	40.47	92.9	26,768
<i>Calotropis gigantea</i>	157.28	48.58	0.81	89.8	18,197
<i>Camellia sinensis</i>	3,140	67.07	1.39	98.0	33,932
<i>Macleaya cordata</i>	378	25	0.308	90.4	22,328