

# Exploring evolution characteristic between cultivated tea and its wild relatives using complete chloroplast genomes

**CURRENT STATUS:** UNDER REVIEW

BMC Evolutionary Biology  BMC Series

Jiao Peng  
Central South University of Forestry and Technology  
ORCID: <https://orcid.org/0000-0002-0495-8510>

Yunlin Zhao  
Central South University of Forestry and Technology

Meng Dong  
Hunan City University

Shiquan Liu  
Hunan City University

Zhiyuan Hu  
Hunan City University

Xiaofen Zhong  
Hunan City University

Zhenggang Xu

 [rssq198677@163.com](mailto:rssq198677@163.com) Corresponding Author  
ORCID: <https://orcid.org/0000-0002-7401-5163>

## DOI:

10.21203/rs.3.rs-15441/v1

## SUBJECT AREAS

*Evolutionary Developmental Biology*    *Evolutionary Biology*

## KEYWORDS

*chloroplast genome, cultivated tea, evolution, ycf1, Camellia*

## Abstract

**Background:** The cultivated tea is one of the most important economic and ecological crops worldwide. The cultivated tea and wild tea suffered from long-term targeted selection of traits and overexploitation of habitats by human beings, which may change the genetic structure. Chloroplast is an organelle with a conserved cyclic structure, and can help us better understand the evolutionary relationship of *Camellia* plants.

**Results:** The study conducted comparative analysis and evolution analysis between cultivated tea and wild tea, and detected the evolution characteristic in cultivated tea. Chloroplast genome sizes of cultivated tea were slightly different, ranged from 157,025 bp to 157,085 bp. These cultivated species were more conservative than wild species, in terms of the genome length, genes number, genes arrangement and GC contents. However, the IRs length of cultivated species was about 20 bp shorter than that of *C. sinensis* var. *sinensis*. We also found that the nucleotide diversity of 14 sequences in cultivated tea was higher than that of wild tea. These results prove the evidence on the variation of chloroplast genomes of cultivated tea. Detail analysis on the chloroplast genome variation and evolution of cultivated tea showed that 67 SNPs and 46 indels were found and 16 protein coding genes had nucleotide substitutions. The most common variation gene was *ycf1*. It has the largest number of nucleotide substitution. At the same time, in *ycf1*, five amino acid sites were exhibiting site-specific selection, and a 9 bp sequence insertion was found in the *C. sinensis* cultivar. Anhua. The phylogenetic tree constructed by *ycf1* sequence shows that two cultivated tea are not completely clustered, and the evolutionary relationship between *C. sinensis* var. *sinensis* and *C. sinensis* cultivar. Longjing is closer than that of *C. sinensis* cultivar. Anhua.

**Conclusions:** The cultivated species are more conservative than wild species in terms of architecture and linear sequence order. The variation of chloroplast genome of cultivated tea is mainly manifested in the nucleotide polymorphism of some sequences. The *ycf1* gene plays an important role in the adaptive evolution of cultivated tea. These results provide evidence regarding the influence of human activities on tea.

## Full Text

Due to technical limitations, full-text HTML conversion of this manuscript could not be completed.

However, the manuscript can be downloaded and accessed as a PDF.

## Tables

**Table 1 Summary of research species information**

Species	GenBank ID	Subgenus	Types	Sample location	Location	References
<i>Camellia sinensis</i> var. <i>sinensis</i>	KJ806281	Thea	wild	Yunnan Academy of Agricultural Science	Yunnan, China	[52]
<i>Camellia sinensis</i> cultivar Anhua	MH042531	Thea	cultivar	Hunan City University	Hunan, China	[2]
<i>Camellia sinensis</i> cultivar Longjing	KF562708	Thea	cultivar	Huajiachi campus of Zhejiang University	Zhejiang, China	[3]
<i>Camellia sinensis</i> var. <i>assamica</i>	MH394410	Thea	wild	Kunming Institute of Botany, Kunming	Yunnan, China	[53]
<i>Camellia sinensis</i> var. <i>pubilimba</i>	KJ806280	Thea	wild	Yunnan Academy of Agricultural Science	Yunnan, China	[52]
<i>Camellia grandibracteata</i>	NC024659	Thea	wild	Yunnan Academy of Agricultural Science	Yunnan, China	[52]
<i>Camellia taliensis</i>	NC022264	Thea	wild	Kunming Institute of Botany	Yunnan, China	[54]
<i>Camellia impressinervis</i>	NC022461	Thea	wild	Kunming Institute of Botany	Yunnan, China	[54]
<i>Camellia pubicosta</i>	NC024662	Thea	wild	International Camellia Species Garden	Zhejiang, China	[52]
<i>Camellia azalea</i>	NC035574	Camellia	wild	Yangchun County	Guangdong, China	[5]
<i>Camellia pitardii</i>	NC022462	Camellia	wild	Kunming Institute of Botany	Yunnan, China	[54]
<i>Camellia reticulata</i>	NC024663	Camellia	wild	Kunming Institute of Botany	Yunnan, China	[52]
<i>Camellia crapnelliana</i>	NC024541	Camellia	wild	Kunming Botanical Garden	Yunnan, China	[6]
<i>Camellia cuspidata</i>	NC022459	Metacamellia	wild	Kunming Institute of Botany	Yunnan, China	[54]
<i>Camellia petelotii</i>	NC024661	Protocamella	wild	International Camellia Species Garden	Zhejiang, China	[52]
<i>Camellia yunnanensis</i>	NC022463	Protocamella	wild	Kunming Institute of Botany	Yunnan, China	[54]

**Table 2. Chloroplast genomic features of sixteen *Camellia* species, with the *C.sinensis* var. *sinensis* as the outgroup**

Species	CSVS	CSCA	CSCL	CSA	CSP	CGR	CTA	CIM	CPU	CAZ	CPI	CRE	CCR	CCU	CPE	CYU
Length (bp)	157117	157025	157085	157028	157086	157127	156974	156892	157076	157039	156585	156971	156997	156618	157121	156592
Genes	115	115	115	113	115	115	115	115	115	115	115	115	114	115	115	115
CDS genes*	81	81	81	79	81	81	81	81	81	81	81	81	81	81	81	81
tRNA genes	30	30	30	30	30	30	30	30	30	30	30	30	29	30	30	30
Introns	18	18	18	22	18	18	21	21	18	18	21	18	20	21	18	21
CDS region	80542	80620	80650	79093	80622	80656	79577	79655	80665	80629	79619	76224	79649	79643	80650	79655
Introns region	15192	15196	15198	17902	15210	15205	16947	16897	15198	15195	16937	15182	16239	16917	15196	16935
IGS region	49535	49361	49389	48200	49405	49418	48591	48481	49365	49367	48171	53717	49321	48199	49427	48143
tRNA region	2802	2802	2802	2789	2802	2802	2813	2813	2802	2802	2812	2802	2742	2813	2802	2813
rRNA region	9046	9046	9046	9044	9047	9046	9046	9046	9046	9046	9046	9046	9046	9046	9046	9046
Genome GC	37.3	37.3	37.29	37.3	37.32	37.29	37.32	37.33	37.3	37.3	37.34	37.31	37.3	37.31	37.29	37.33
CDS GC	37.58	37.57	37.56	37.47	37.58	37.56	37.57	37.54	37.57	37.56	37.56	37.54	37.54	37.56	37.56	37.54
Introns GC	36.41	36.38	36.38	37.91	36.42	36.39	37.25	37.28	36.42	36.41	37.22	36.4	37.54	37.25	36.41	37.25
IGS GC	32.93	32.94	32.94	32.48	32.97	32.94	32.68	32.72	32.93	32.95	32.71	33.39	32.64	32.63	32.92	32.68
tRNA GC	52.86	52.86	52.86	52.99	52.86	52.86	52.86	52.9	52.89	52.86	52.92	52.86	52.88	52.9	52.86	52.9
rRNA GC	55.39	55.41	55.41	55.4	55.41	55.41	55.38	55.41	55.42	55.39	55.36	55.34	55.41	55.38	55.39	55.41
Gene losses				<i>orf42</i> , <i>ycf1</i> , <i>ycf15</i>			<i>orf42</i> , <i>ycf1</i>	<i>orf42</i> , <i>ycf1</i>			<i>orf42</i> , <i>ycf1</i>		<i>orf42</i> , <i>trnG</i>	<i>ycf1</i> , <i>orf42</i>		<i>orf42</i> , <i>ycf1</i>
Intron losses	<i>rps12</i>	<i>rps12</i>	<i>rps12</i>		<i>rps12</i>	<i>rps12</i>			<i>rps12</i>	<i>rps12</i>		<i>rps12</i>			<i>rps12</i>	

Abbreviations: *Camellia sinensis* var. *sinensis*, CSVS; *Camellia sinensis* cultivar Anhua, CSCA; *Camellia sinensis* cultivar Longjing, CSCL; *Camellia sinensis* var. *assamica*, CSA; *Camellia sinensis* var. *pubilimba*, CSP; *Camellia grandibracteata*, CGR; *Camellia taliensis*, CTA; *Camellia impressinervis*, CIM; *Camellia pubicosta*, CPU; *Camellia azalea*, CAZ; *Camellia pitardii*, CPI; *Camellia reticulata*, CRE; *Camellia crapnelliana*, CCR; *Camellia cuspidate*, CCU; *Camellia petelotii*, CPE; *Camellia yunnanensis*, CYU.

## Figures

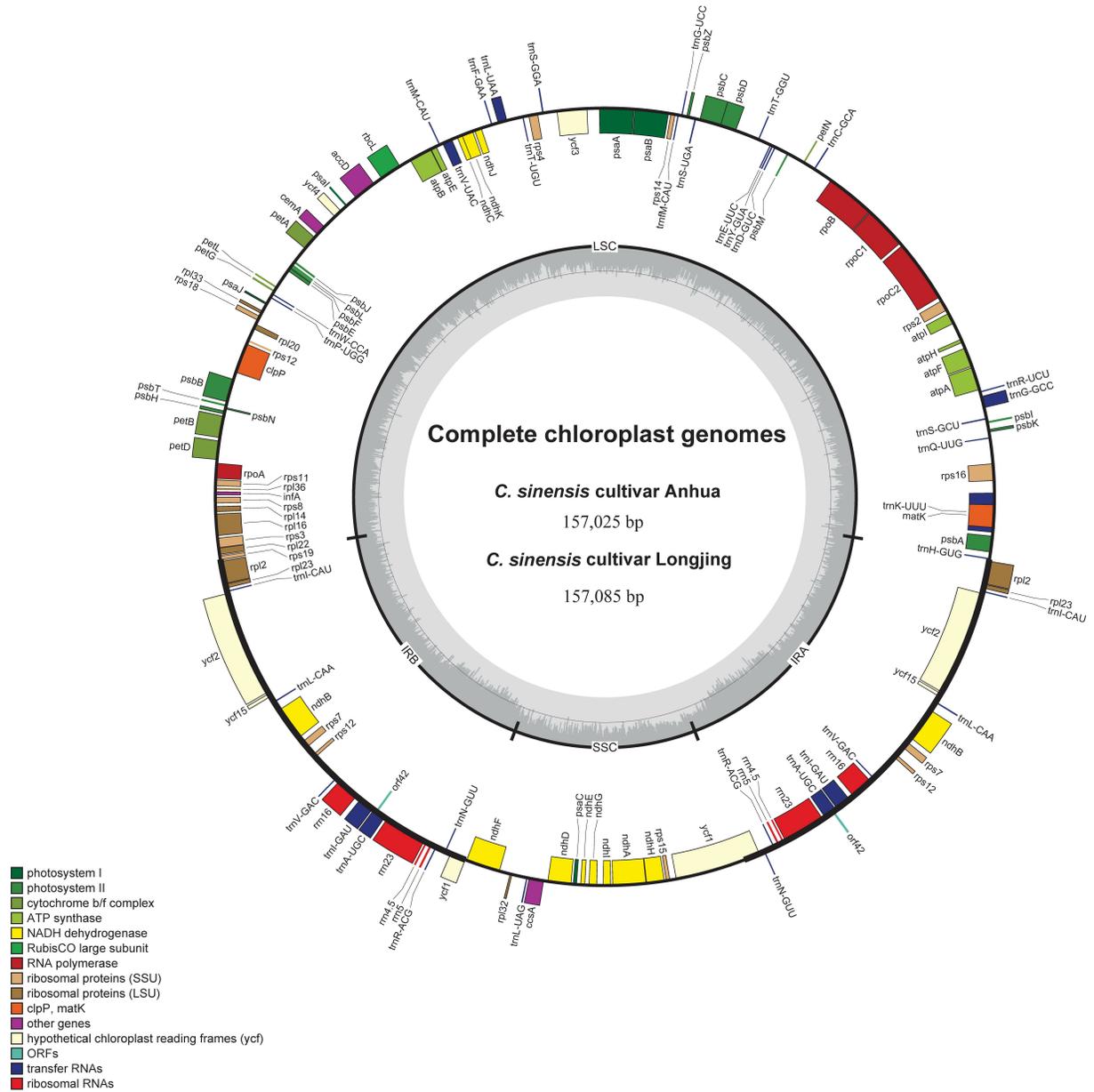


Figure 1

Gene map of the complete chloroplast genome of cultivated tea. Genes lying inside of the circle are transcribed clockwise, and those outside are transcribed counterclockwise.

Different color of blocks represents different functional groups. The darker gray color of the inner circle corresponds to the GC contents, and the lighter gray color corresponds to the AT contents.

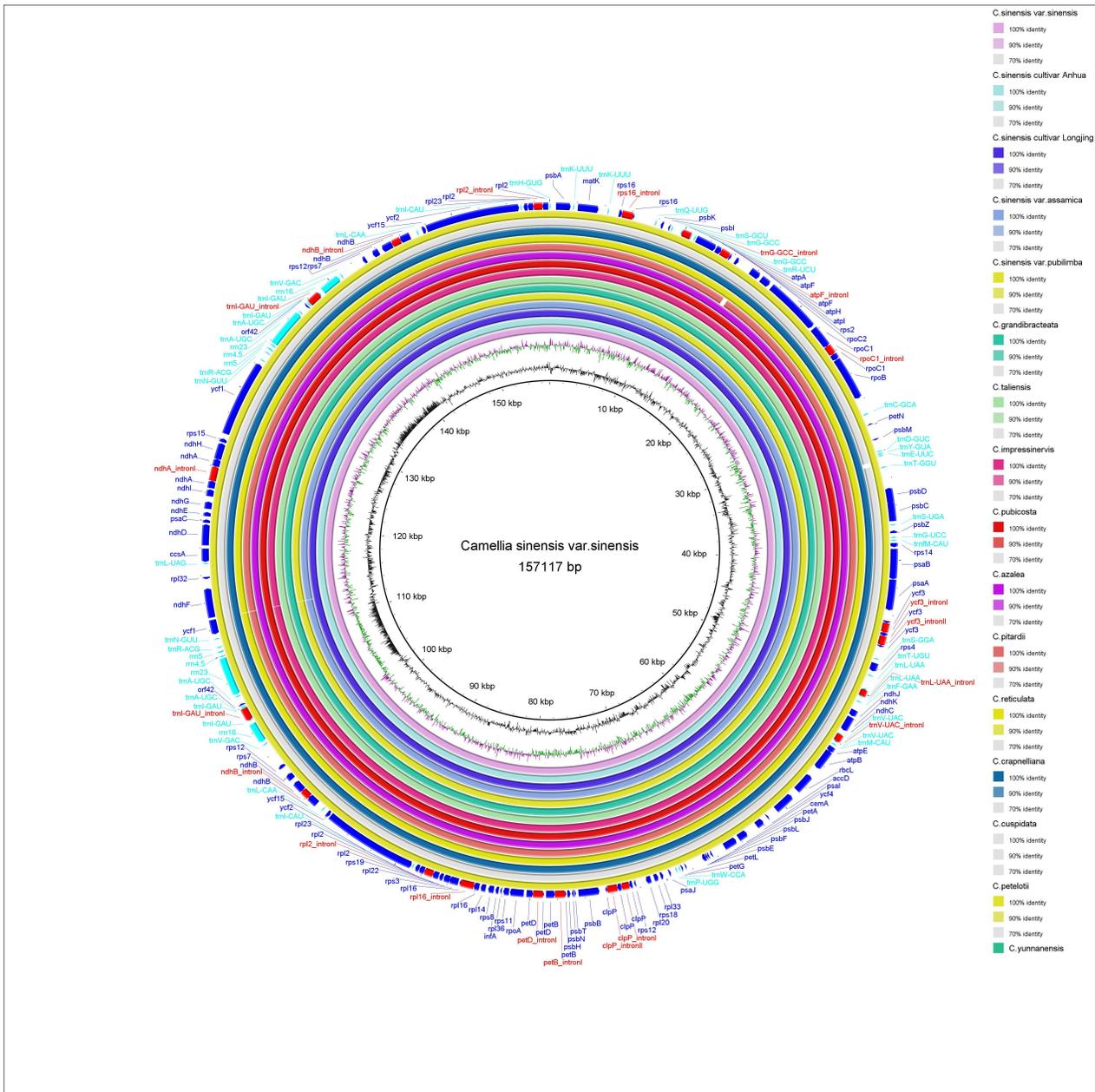


Figure 2

The sequence identity of sixteen *Camellia* species. The inner circle is the reference genome.

Next circles represent the sequence identity between *C. sinensis* var. *sinensis* and fifteen other species. The outermost circle corresponds to the protein-coding genes and intergenic spacer regions. Genes with clockwise arrows represent reverse strands, while genes with the counterclockwise arrow represent forward strands.

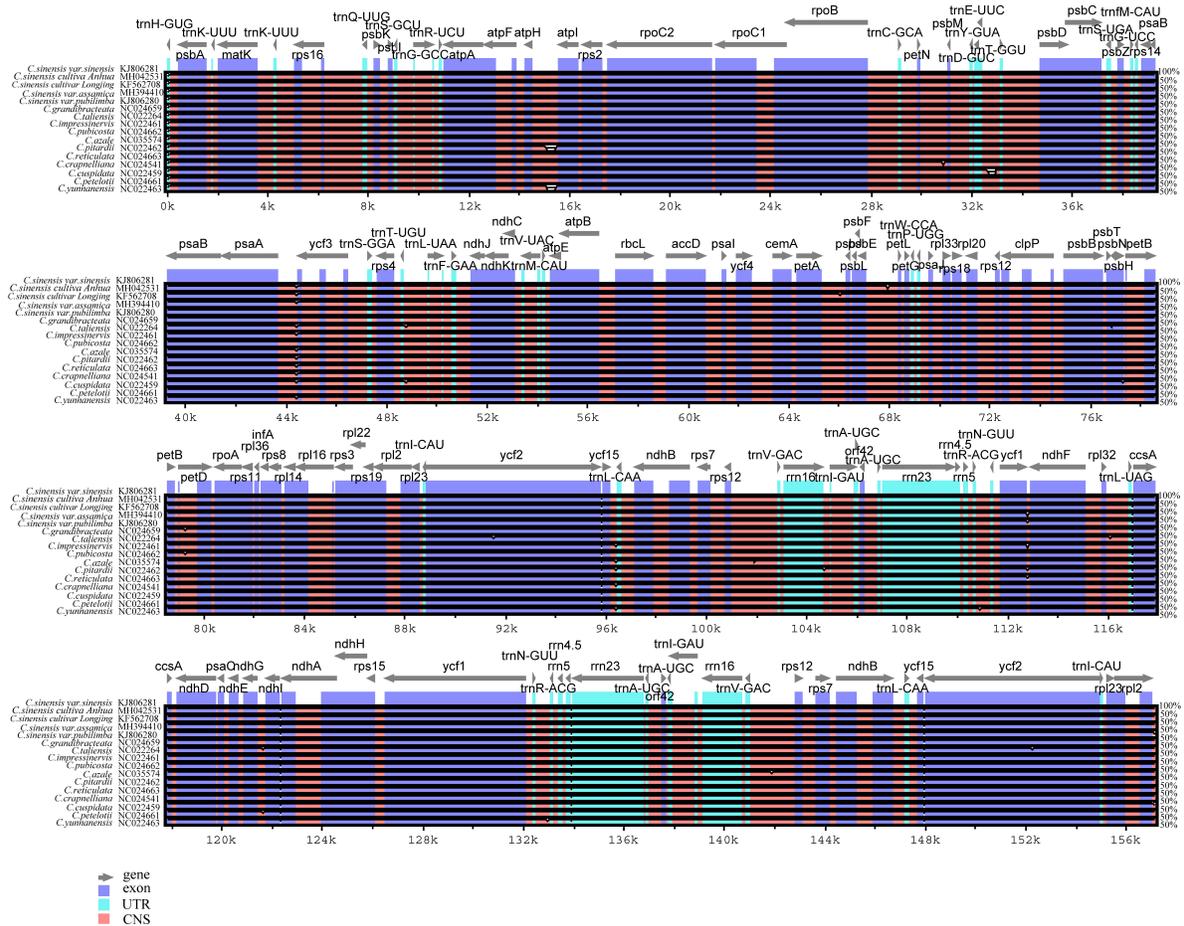
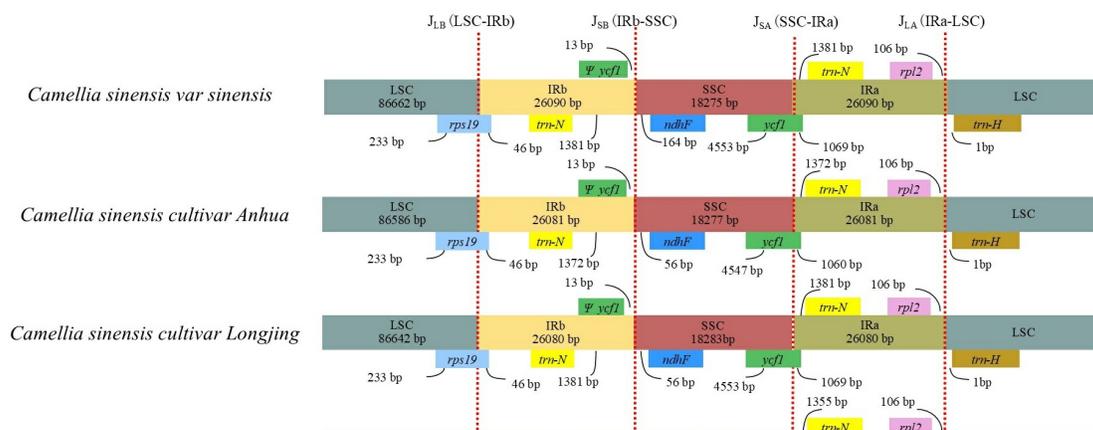


Figure 3

Alignment visualization of the sixteen *Camellia* chloroplast genome sequences using *C. sinensis* var. *sinensis* as reference. Vertical scale indicates the percentage of identity, ranging from 50% to 100%. Arrows indicate the annotated genes and their transcriptional direction. The different colored boxes correspond to exons, tRNA or rRNA, and noncoding sequences (CNSs).



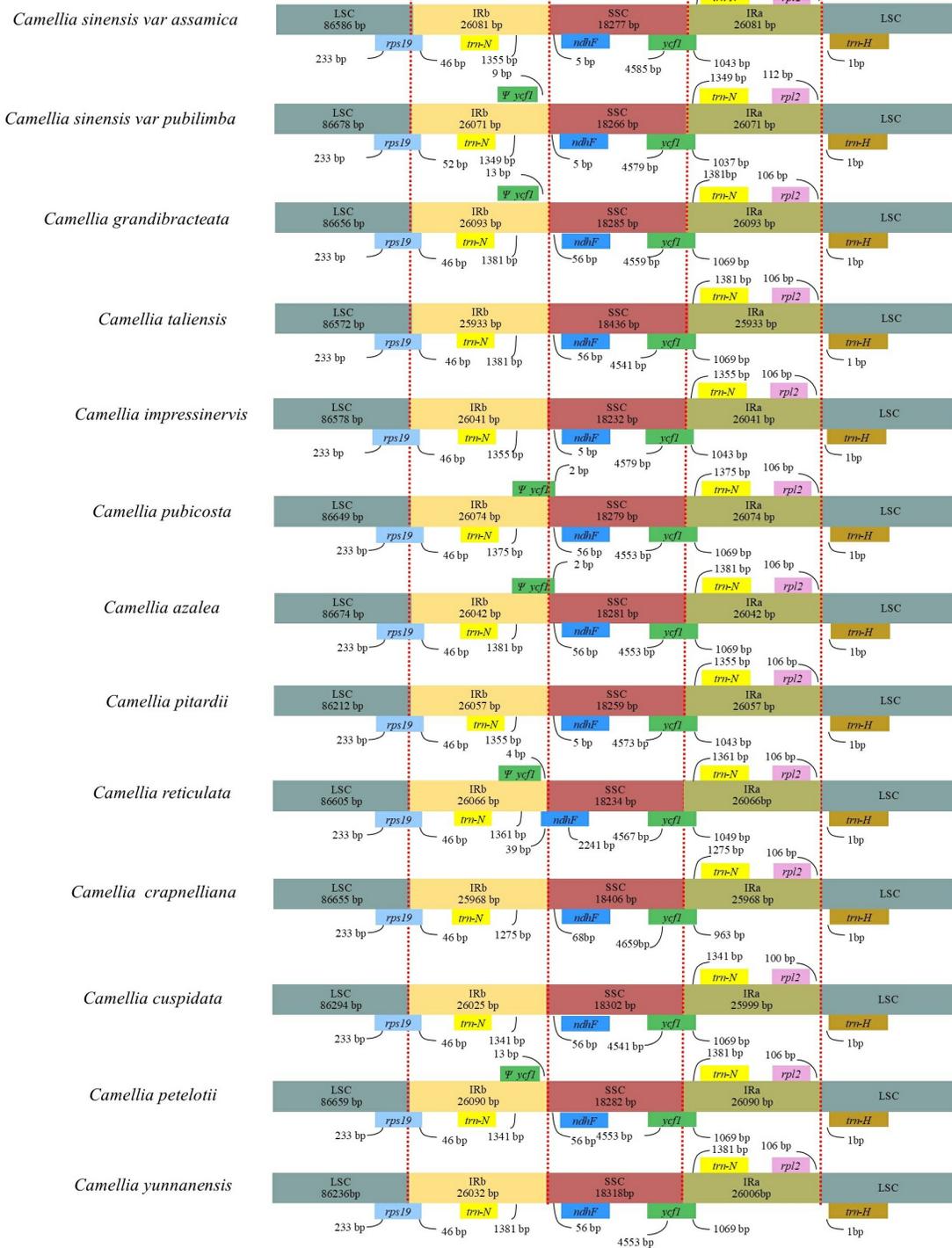


Figure 4

Comparison of IRs boundary regions among the 16 *Camellia* chloroplast genomes, using *C. sinensis* var. *sinensis* as the reference. Boxes above or below the line are forward strands and reverse strands, respectively.

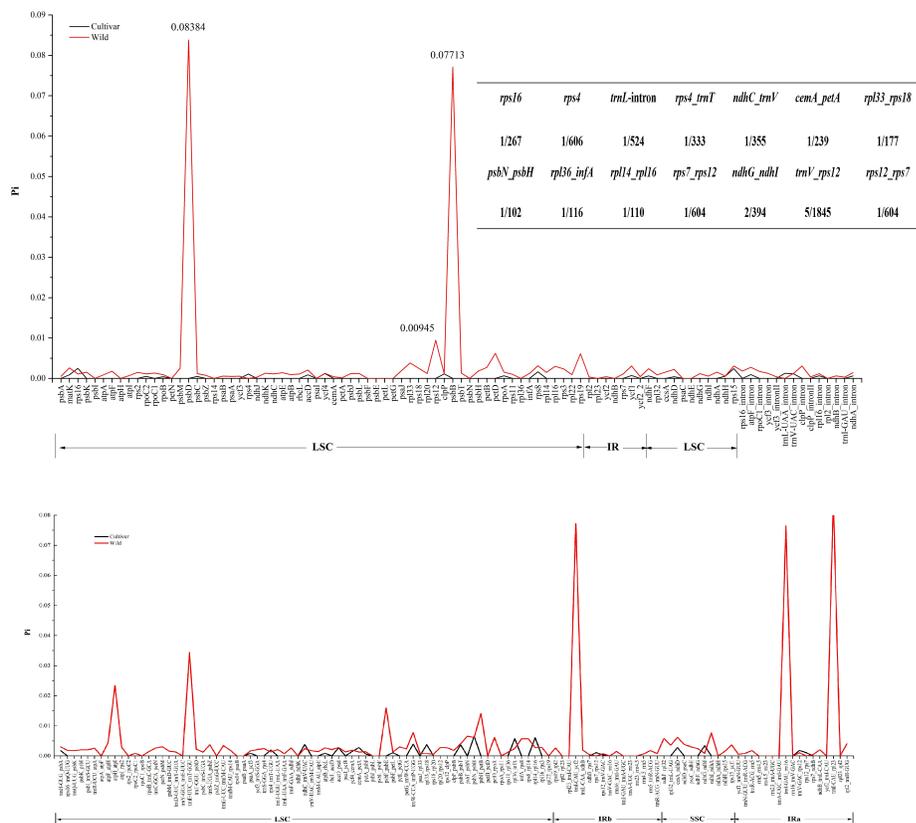


Figure 5

Comparative analysis of nucleotide variability (Pi) values between the cultivated tea and wild tea cp genome sequences. X-axis: position of the midpoint of a window, Y-axis: nucleotide diversity of each window.

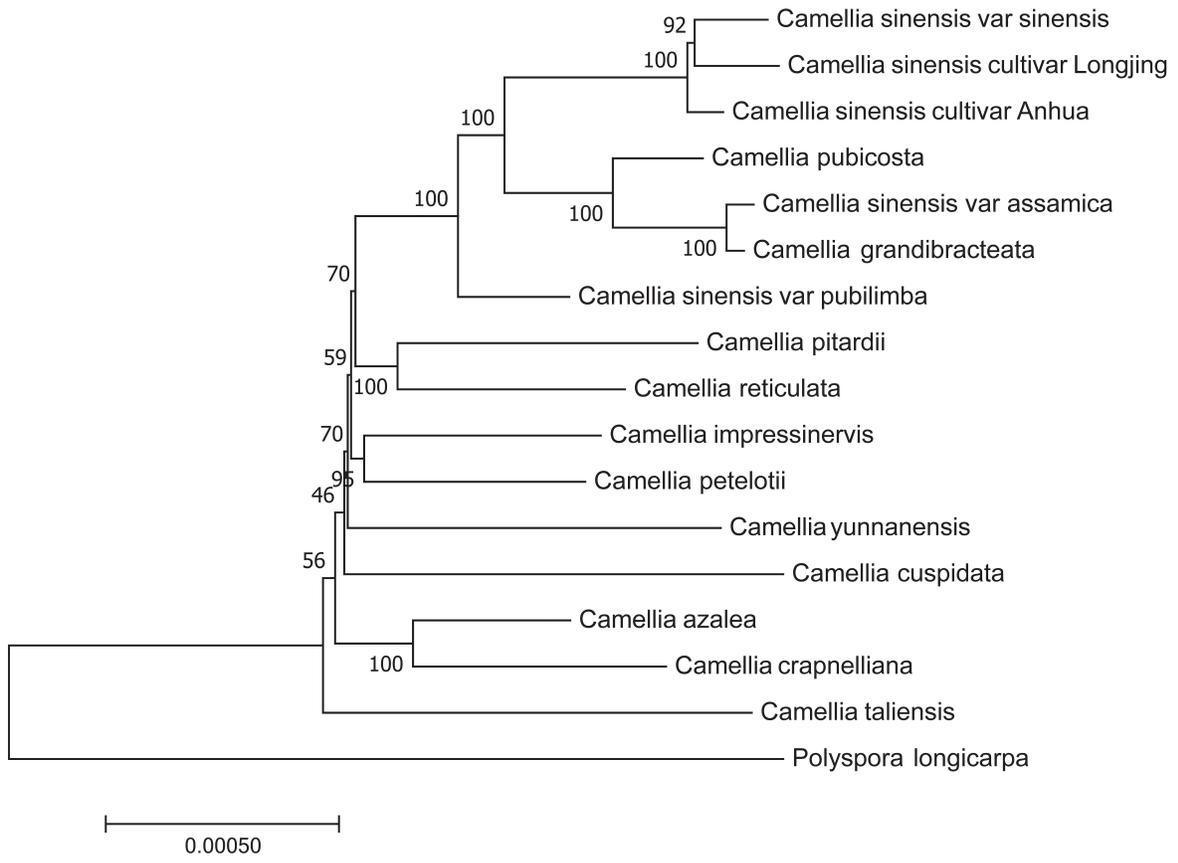


Figure 6

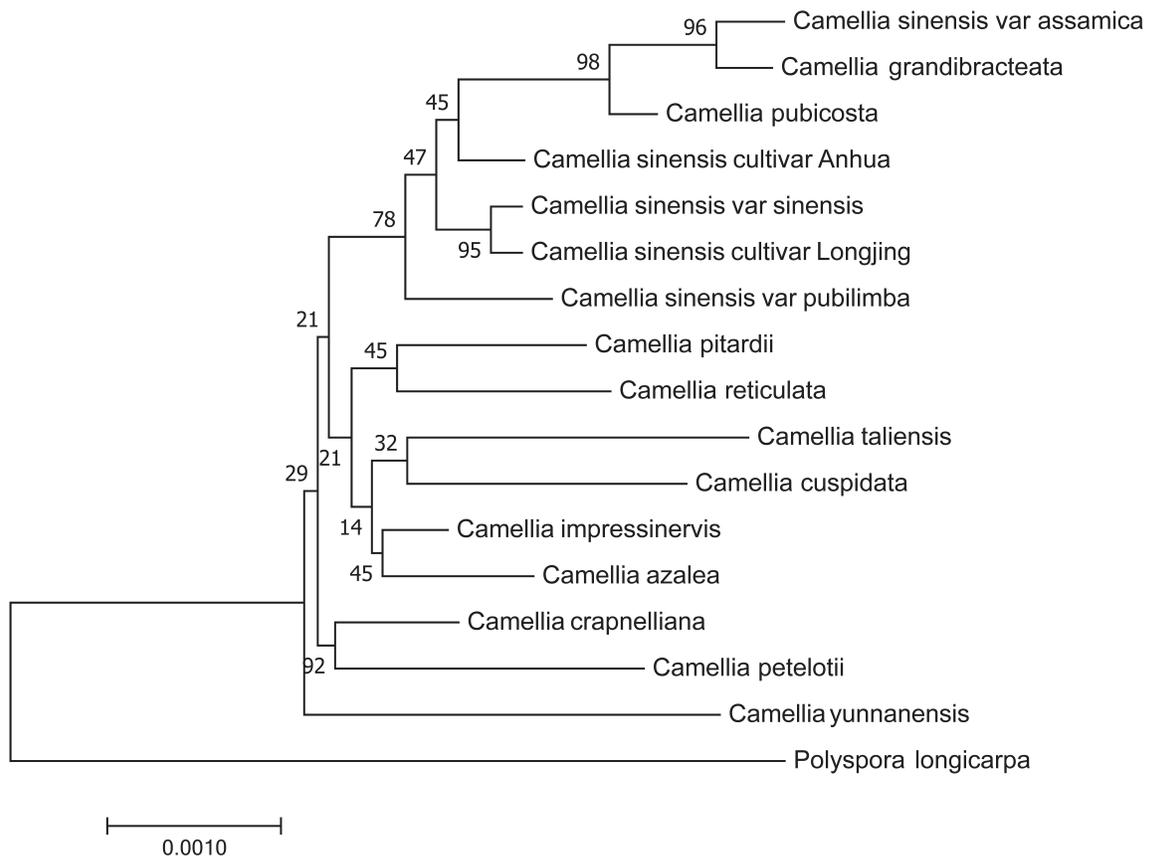


Figure 7

Phylogenetic trees of *Camellia* plants based on based on *ycf1* gene. *Polyspora longicarpa* was selected as the outgroup.

