scientific reports

Check for updates

OPEN

Comparative analysis of chloroplast genomes indicated different origin for Indian tea (*Camellia assamica* cv TV1) as compared to Chinese tea

Hukam C. Rawal[1], Sangeeta Borchetia[2], Biswajit Bera[3], S. Soundararajan[3], R. Victor J. Ilango[4], Anoop Kumar Barooah[2], Tilak Raj Sharma[1], Nagendra Kumar Singh[1] & Tapan Kumar Mondal[1]✉

Based upon the morphological characteristics, tea is classified botanically into 2 main types i.e. Assam and China, which are morphologically very distinct. Further, they are so easily pollinated among themselves, that a third category, Cambod type is also described. Although the general consensus of origin of tea is India, Burma and China adjoining area, yet specific origin of China and Assam type tea are not yet clear. Thus, we made an attempt to understand the origin of Indian tea through the comparative analysis of different chloroplast (cp) genomes under the Camellia genus by performing evolutionary study and comparing simple sequence repeats (SSRs) and codon usage distribution patterns among them. The Cp genome based phylogenetic analysis indicated that Indian Tea, TV1 formed a different group from that of China tea, indicating that TV1 might have undergone different domestications and hence owe different origins. The simple sequence repeats (SSRs) analysis and codon usage distribution patterns also supported the clustering order in the cp genome based phylogenetic tree.

Tea is natural morning drink consumed by majority of the world population. Tea is a woody, perennial and highly cross-pollinated crop, so the genus is very dynamic as indicated with the recent discovery of several new species[1]. At present, more than 350 species are available in this genus Camellia[2]. However among them, mainly two species i.e. *Camellia sinensis* (L.) O. Kuntz and *C. assamica* produce tea that we drink in various forms such as black tea, green tea oroolong tea[3]. However, in some areas of China, tea is also being made from a different species i.e. *C. taliensis*. Based upon the morphological characters, they are divided broadly into three types such as China tea, Assam tea and a hybrid between them called Cambod tea. Due to their high out-crossing nature, they breed freely among themselves which produce plant type that is an intermediate between the two extreme forms i.e. Assam type big leaf and China type small leaf. As a consequence, the taxonomic classification of tea, based on morphology, is still confusing and species name is also a misnomer. For example, though China tea and Assam tea are considered to be separate species yet there is no restriction of gene flow. In fact desirable traits such as anthocyanin pigmentation or special quality characters of Darjeeling tea might have introduced from two wild species such as *C. Irrawadiensis* and *C. taliensis*[4]. Besides taxonomy, origin of tea is also debatable and needs to be clarified. Although Indo-Burma region near Irrawaddy river is considered to be the centre of origin[5], yet it is not clear whether Assam and China type tea have same or different domestication origin.

Chloroplasts (cp) are essential organelles that are primarily responsible for photosynthesis and hence found from green algae to higher plants[6]. They are maternally inherited and do not participate in genetic recombination and hence are highly conserved[7]. The cp genome is circular with double stranded DNA molecule with a length of about 120–220 kb; that codes for 100–200 genes including protein-coding genes, rRNA and tRNA genes[8]. It also has a highly conserved quadripartite structure that includes a large single copy (LSC) region and a small single copy (SSC) region separated by two copies of inverted repeat regions (IR-A and IR-B). Due to this high conservation of chloroplast genomes compared to nuclear and mitochondrial genomes[9], they are widely used to differentiate closely related taxa, particularly some taxa below the species level with unclear taxonomic

[1]ICAR-National Institute for Plant Biotechnology, LBS Centre, IARI, New Delhi 110012, India. [2]Tocklai Tea Research Institute, Tea Research Association, Jorhat, Assam 785008, India. [3]Tea Board, Ministry of Commerce and Industry, Govt. of India, 14, B.T.M. Sarani, Kolkata 700 001, India. [4]UPASI Tea Research Foundation, Tea Research Institute, Coimbatore, Tamil Nadu, India. ✉email: mondaltk@yahoo.com

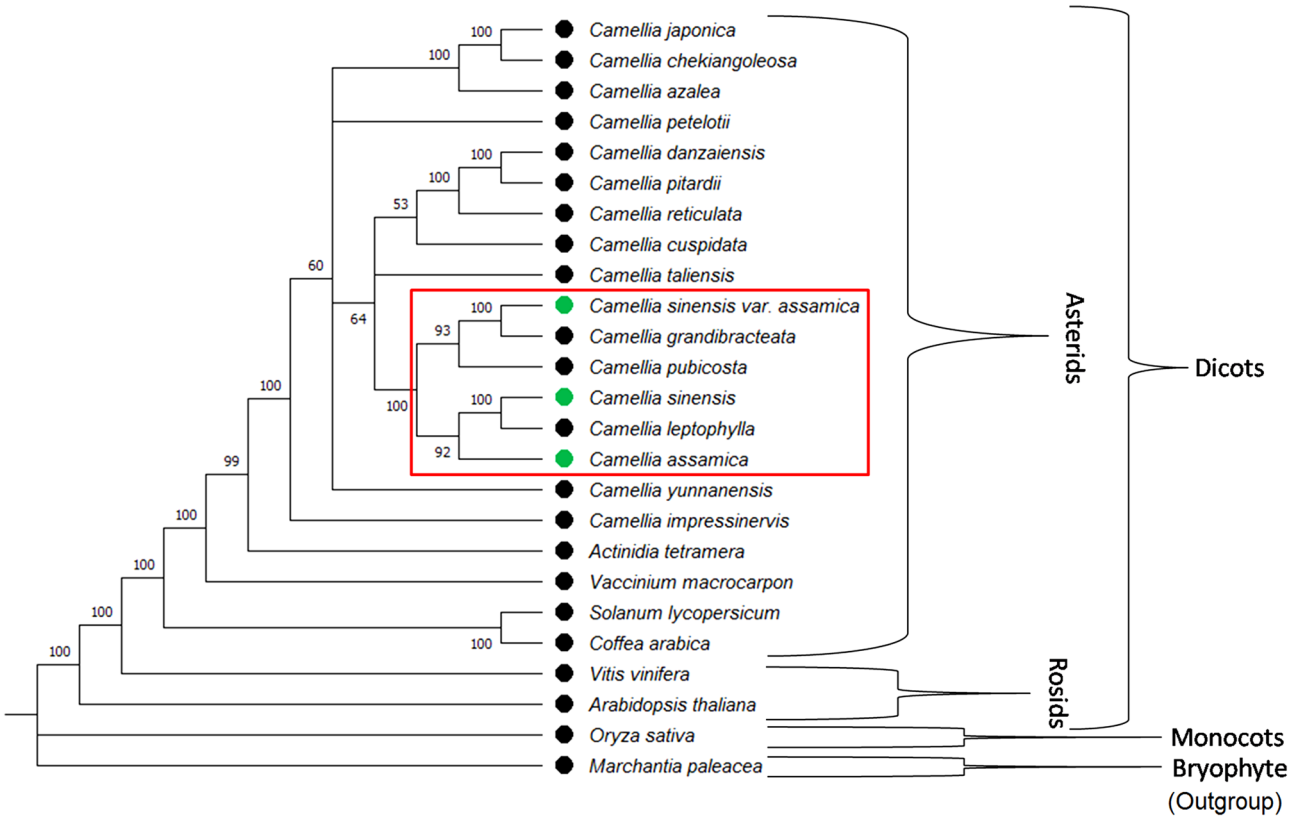

Figure 1. Chloroplast (cp) genome sequence based ML tree showing the phylogenetic relation between China tea, China Assam tea and Indian Assam tea (green bullet) to confirm their different origins.

relationships. Complete chloroplast genome thus can be used for deciphering phylogenetic relationships between closely related taxa to improve the understanding of the evolution of plant species. In the present study, we have used well annotated cp genome sequences of 17 species from Camellia genus including China type (*C. sinensis*), Assam type (*C. sinensis* var. assamica) and typical Indian Tea (Indian Assam type, TV1) that we decoded recently to understand the domestication, evolution of Indian tea and for exploiting DNA barcodes to geographically specific tea variety.

Materials and methods

Data collection. To perform the chloroplast based phylogenetic analysis, we have chosen previously published and well-annotated cp genome sequences of 15 Camellia species (*C. sinensis*, *C. grandibracteata*, *C. leptophylla*, *C. petelotii*, *C. pubicosta*, *C. reticulata*, *C. azalea*, *C. japonica*, *C. chekiangoleosa*, *C. cuspidata*, *C. danzaiensis*, *C. impressinervis*, *C. pitardii*, *C. yunnanensis* and *C. taliensis*)[10] and 2 recently published cp genome sequences including *C. assamica*[11] and *C. sinensis* var. *assamica*[12]. All these 17 Camellia species belongs to the Theaceae family under Asterids group. Moreover, in the analysis, while *Oryza sativa* and *Arabidopsis thaliana* were included as model monocot and dicot species, respectively, a bryophyte *Marchantia paleacea* was included to serve as the out-group. We have also included 5 other dicot species comprising *Vitis vinifera* (grapes), *Solanum lycopersicum* (tomato) and *Coffea arabica* (coffee), all three are believed to be evolutionary close to tea plant, 2 members from Asterids group but from family other than Theaceae i.e. *Actinidia tetramera* (crab-apple kiwi) and *Vaccinium macrocarpon* (cranberry), representing the family Actinidiaceae and Ericaceae, respectively. The cp genome sequences were downloaded for these 25 species from the NCBI Organelle Genome Resources database (Table S1).

Phylogenetic analysis. We have used the method described in our previous study[11]. The cp genome sequences of all selected 25 species were aligned by MAFFT v7.402[13] at default parameters and the best-fit model for the downstream phylogenetic analysis was determined by ModelTest (ModelTest-NG v0.1.3)[14]. Finally, the Maximum likelihood (ML) tree was generated with RAxMLv8.2.12[15] by using the best-fit substitution model (GTRGAMMAX model) with 1000 bootstrap replicates. *M. paleacea*, a bryophyte served as the out-group in this analysis.

Comparative analysis of seven Camellia species. Seven cp genomes of Camellia genus were further shortlisted based upon the results of phylogenetic analysis and compared for their cp genomic features including length, LSC and SSC and IR regions number of protein-coding genes, tRNA and rRNA genes. Further simple sequence repeats (SSRs) were identified in these cp genomes by MicroSAtellite identification tool (MISA)[16] with

Repeats	*C. assamica*	*C. leptophylla*	*C. sinensis*	*C. pubicosta*	*C. grandibracteata*	*C. sinensis* var. *assamica*	*C. taliensis*
A/T	152	152	152	151	154	155	153
C/G	5	5	5	5	5	5	5
AG/CT	16	16	16	16	16	16	16
AT/AT	25	25	25	25	25	25	25
AAG/CTT	1	1	1	1	1	1	1
AAT/ATT	0	0	0	0	0	0	1
AAAG/CTTT	3	3	2	4	3	3	3
AAAT/ATTT	2	3	3	3	3	3	1
ACAG/CTGT	1	1	1	1	1	1	1
AGAT/ATCT	3	3	3	3	3	3	3
AGGG/CCCT	2	2	2	0	0	0	2
AAAAAG/CTTTTT	2	2	2	2	2	2	0

Table 1. Frequency of classified SSR repeat types in 7 cp genomes.

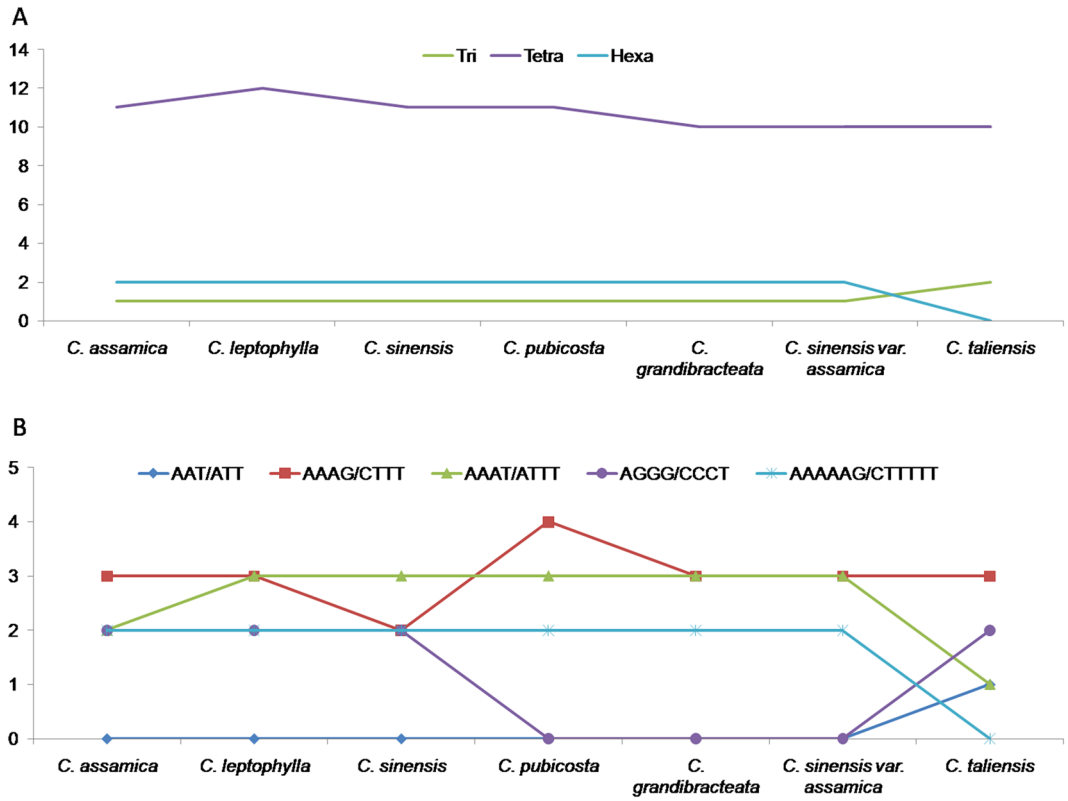

Figure 2. Distribution pattern of tri, tetra and hexa type SSR number (**A**) and SSR motifs (**B**) showing related variation between first three (*C. assamica*, *C. leptophylla* and *C. sinensis*), next three (*C. pubicosta, C. grandibracteata* and *C. sinensis* var. *assamica*) or last species (*C. taliensis*) under analysis.

a threshold of 8, 4, 4, 3, 3 and 3 for minimum number of nucleotide repeats in mono-, di-, tri-, tetra-, penta- and hexa-nucleotide repeats, respectively[17,18].

For a specified codon, the Relative Synonymous Codon Usage (RSCU) value is the ratio of its actual usage frequency to expected frequency when it is used without bias. These RSCU values were calculated for seven Camellia species using ACUA[19] for 60 codons (excluding 3 stop and 1 start codon from the total 64 codons).

Results and discussion

Phylogenetic analysis based on cp genomes. The ML tree based on alignment of cp genome sequences using GTRGAMMAX model at 1000 bootstrap replicates supported the branching-off of the bryophyte *M. paleacea*, an outgroup in this analysis, from rest of the plant species and serving as the root for the tree. *O. sativa*,

Amino Acid	Codon	*C. assamica*	*C. sinensis*	*C. sinensis* var. *assamica*	*C. taliensis*	Remarks
A	GCA	1.11	1.14	1.14	1.15	#,CSA-CSS
A	GCC	0.66	0.65	0.64	0.65	#,CSS-CST
A	GCG	0.39	0.39	0.38	0.38	Ind-CSS,CSA-CST
A	GCT	1.83	1.83	1.84	1.83	Ind-CSS,Ind-CST,CSS-CST
C	TGC	0.49	0.47	0.48	0.47	#,CSS-CST
C	TGT	1.51	1.53	1.53	1.53	#,CSA-CSS-CST
D	GAC	0.38	0.37	0.37	0.36	#,CSA-CSS
D	GAT	1.62	1.63	1.63	1.64	#,CSA-CSS
E	GAA	1.49	1.51	1.51	1.51	#,CSA-CSS-CST
E	GAG	0.51	0.49	0.49	0.49	#,CSA-CSS-CST
F	TTC	0.73	0.71	0.72	0.72	#,CSA-CST
F	TTT	1.27	1.29	1.28	1.28	#,CSA-CST
G	GGA	1.63	1.64	1.64	1.65	#,CSA-CSS
G	GGC	0.42	0.42	0.41	0.41	Ind-CSS,CSA-CST
G	GGG	0.68	0.68	0.67	0.67	Ind-CSS,CSA-CST
G	GGT	1.28	1.27	1.28	1.27	Ind-CSA,CSS-CST
H	CAC	0.44	0.43	0.43	0.43	#,CSA-CSS-CST
H	CAT	1.56	1.58	1.57	1.57	#,CSA-CST
I	ATA	0.93	0.96	0.96	0.96	#,CSA-CSS-CST
I	ATC	0.61	0.60	0.60	0.60	#,CSA-CSS-CST
I	ATT	1.46	1.44	1.44	1.44	#,CSA-CSS-CST
K	AAA	1.46	1.49	1.48	1.49	#,CSS-CST
K	AAG	0.54	0.51	0.52	0.51	#,CSS-CST
L	CTA	0.81	0.80	0.81	0.81	Ind-CSA,Ind-CST,CSA-CST
L	CTC	0.46	0.43	0.45	0.43	#,CSS-CST
L	CTG	0.39	0.39	0.39	0.39	*
L	CTT	1.25	1.25	1.26	1.27	Ind-CSS
L	TTA	1.85	1.91	1.89	1.90	#
L	TTG	1.24	1.22	1.20	1.20	#,CSA-CST
N	AAC	0.47	0.46	0.46	0.46	#,CSA-CSS-CST
N	AAT	1.53	1.54	1.54	1.54	#,CSA-CSS-CST
P	CCA	1.18	1.19	1.18	1.19	Ind-CSA,CSS-CST
P	CCC	0.70	0.70	0.70	0.70	*
P	CCG	0.50	0.50	0.51	0.51	Ind-CSS,CSA-CST
P	CCT	1.63	1.60	1.61	1.61	#,CSA-CST
Q	CAA	1.51	1.53	1.53	1.53	#,CSA-CSS-CST
Q	CAG	0.49	0.48	0.47	0.47	#,CSA-CST
R	AGA	1.78	1.86	1.86	1.86	#,CSA-CSS-CST
R	AGG	0.63	0.62	0.62	0.61	#,CSA-CSS
R	CGA	1.45	1.44	1.44	1.43	#,CSA-CSS
R	CGC	0.34	0.33	0.33	0.32	#,CSA-CSS
R	CGG	0.43	0.42	0.42	0.42	#,CSA-CSS-CST
R	CGT	1.37	1.33	1.32	1.35	#
S	AGC	0.35	0.34	0.34	0.34	#,CSA-CSS-CST
S	AGT	1.25	1.22	1.23	1.23	#,CSA-CST
S	TCA	1.14	1.19	1.19	1.19	#,CSA-CSS-CST
S	TCC	0.96	0.94	0.94	0.95	#,CSA-CSS
S	TCG	0.52	0.51	0.52	0.51	Ind-CSA,CSS-CST
S	TCT	1.79	1.80	1.78	1.79	Ind-CST
T	ACA	1.21	1.23	1.24	1.22	#
T	ACC	0.73	0.73	0.73	0.73	*
T	ACG	0.42	0.41	0.42	0.42	Ind-CSA,Ind-CST,CSA-CST
T	ACT	1.64	1.63	1.61	1.64	Ind-CST
V	GTA	1.49	1.50	1.49	1.50	Ind-CSA,CSS-CST
V	GTC	0.47	0.46	0.48	0.47	Ind-CST
V	GTG	0.56	0.56	0.55	0.55	Ind-CSS,CSA-CST
V	GTT	1.48	1.48	1.48	1.48	*
W	TGG	1.00	1.00	1.00	1.00	*
Y	TAC	0.38	0.40	0.39	0.39	#,CSA-CST
Y	TAT	1.62	1.61	1.61	1.61	#,CSA-CSS-CST

Table 2. The codon usage frequency distribution in terms of RSCU values among four cp genomes of tea with colour coded for blue–yellow–red scale for maximum to minimum values.

the only monocot included in the study, is the next to branched off from rest of the species (dicots) in the tree (Fig. 1). *V. vinifera* and *A. thaliana* (the two Rosids member) are the next species to get separated from rest of the species, all of which belongs to Asterids group. Among Asterids, *C. Arabica* and *S. lycopersicum* are the first one to get separated from the remaining species under analysis, all of which represents the order Ericales. These species under Ericales includes *A. tetramera*, *V. macrocarpon* and all Camellia species, representing three families including Actinidiaceae, Ericaceae and Theaceae, respectively.

Among Camellia genus or Theaceae family, all three domesticated tea *C. assamica* (Indian Assam tea), *C. sinensis* and *C. sinensis* var. *assamica* (China Assam tea) are present in a single major cluster adjacent to the single clade of *C. taliensis*, a well known close wild relative of domesticated tea[20]. Here within this cluster, we observed that Indian Assam tea and China Assam tea belongs to different clades and must have different origins as predicted previously[21,22]. Interestingly, *C. sinensis* was found evolutionary close to Indian Assam tea and seems to be branched-off from Indian Assam tea along with *C. leptophylla* in a similar manner and time as the China Assam tea and *C. grandibracteata* gets branched-off from *C. pubicosta*. This presence of *C. leptophylla* as sister clade with *C. sinensis* and existence of *C. pubicosta* as sister clade to *C. sinensis* var. *assamica* and *C. grandibracteata* are consistent with the previous study[10]. Moreover, both China type tea (*C. sinensis* and China Assam tea) seems to be evolved after Indian Assam tea. Similar phylogenetic relations were found among these seven Camellia species (*C. assamica*, *C. leptophylla*, *C. sinensis*, *C. pubicosta, C. grandibracteata*, *C. sinensis* var. *assamica*, and *C. taliensis*) when inferred with other different and fast phylogenetic tree construction methods in MEGAX[23] including Unweighted Pair Group Method with Arithmetic mean (UPGMA) (Fig. S1), Neighbour-joining (NJ) and Minimum Evolution method with strong bootstrap supports.

Comparative study of seven Camellia species. A comparative study of cp genome features of these seven species revealed that there is hardly any differences or patterns to notice in terms of cp genome size, LSC, SSC or IR region, number of genes, tRNAs and rRNAs (Table S2) to correlate with their phylogenetic positions in the tree. Even list of all genes found almost similar and conserved and has no clues about their relations (Table S3). Interestingly, distribution and frequency of SSR repeat types follow the pattern that these seven genomes have in the phylogenetic tree (Tables S4 and S5). Distribution of SSR types shows that *C. taliensis*, a wild relative to domesticated tea, has a different pattern with 2 tri-mers as compared to 1 in other six cp genomes and no hexa-mers as compared to 2 in other six (Table 1, S4 and Fig. 2A). No AAT/ATT motif was found in these cp genomes, except 1 in *C. taliensis* and two AAAAAG/CTTTTT motifs were there in these six genomes but none in *C. taliensis* (Table 1, Fig. 2B). More importantly, AGGG/CCCT motif has frequency of two in each of *C. assamica*, *C. leptophylla* and *C. sinensis* (which are present in one cluster in the tree) but totally missing in *C. pubicosta, C. grandibracteata* and *C. sinensis* var. *assamica* (which altogether represents a different cluster in the tree).

The codon usage frequency distribution among seven Camellia species showed that Indian Assam type tea has a different pattern than the rest of the six Chinese cp genomes (Table S6). Among these, when we further compared this codon usage frequency of Indian tea with two other domesticated tea (*C. sinensis* and *C. sinensis* var. *assamica*) and one of their close wild relative *C. taliensis*, we observed that out of 60 codons, only 5 codons have similar RSCU values in all 4 cp genomes (Table 2). Among rest of the 55 codons, the distribution pattern of 39 codons (marked with # sign) in Indian tea was found different from other three Chinese tea and only 7 were similar with *C. sinensis* and 6 were similar with each of *C. sinensis* var. *assamica* and *C. taliensis*. Hence, out of total 60 codons, 65% codons have different patterns in Indian tea, 11% shared with *C. sinensis*, 10% shared with each of *C. sinensis* var. *assamica* and *C. taliensis* and around 8% shared common pattern between all four. Apart from these 8% shared codons in all 4 cp genomes, the three Chinese genomes shared a very good number of similar codon usage patterns with 23 shared between *C. sinensis* and *C. sinensis* var. *assamica*, 30 between *C. sinensis* var. *assamica* and *C. taliensis*, and 25 between *C. sinensis* and *C. taliensis*. Overall, these results indicated that while the Indian cp genome has different codon usage patterns, the three Chinese cp genomes have quite similar patterns. Hence there must be different codon usage selection between cp genomes of Indian and Chinese tea with possible chances of different domestications and origins of Indian and Chinese tea.

Conclusions

In the present study, no deviation was observed from the known taxonomic classification from sub-class or order level to family or genus level by following this cp genome based phylogenetic relationship study. Thus one can rely on cp based evolutionary study and the observed phylogenetic relations among species under consideration. This analysis supported the possible different domestication origins of Indian Assam tea and China Assam tea with the existence of Indian Assam tea prior to *C. sinensis* and China Assam tea which is well supported with SSR and codon usage frequency distribution.

Received: 7 October 2020; Accepted: 9 December 2020
Published online: 08 January 2021

References

1. Lee, S. L. & Yang, T. U. A. *Camellia chinmeii*, a new species of *Camellia* sect. Para *Camellia* in Taiwan. *Taiwania* **64**(3), 321–325 (2019).
2. Mukhopadhyay, M., Mondal, T. K. & Chand, P. K. Biotechnological advances in tea (*Camellia sinensis* [L.] O. Kuntze): A review. *Plant Cell Rep.* **35**, 255–287. https://doi.org/10.1007/s00299-015-1884-8 (2016).
3. Mondal, T. K., Bhattacharya, A., Laxmikumaran, M. & Ahuja, P. S. Recent advance in tea biotechnology. *Plant Cell Tissue Organ Cult.* **75**, 795–856 (2004).
4. Wood, D. J. & Barua, D. N. Species hybrids of tea. *Nature* **181**, 1674–1675. https://doi.org/10.1038/1811674a0 (1958).

5. Mondal, T.K. Tea. In *Breeding Plantation Tree crops Tropical Species* (Eds. M. Prydarsini & S.M. Jain) (ISBN: 978-0-387-71199-7) 545–587 (Springer USA, New York, 2009).
6. Xiong, A. S. *et al.* Gene duplication, transfer, and evolution in the chloroplast genome. *Biotechnol. Adv.* **27**(4), 340–347. https://doi.org/10.1016/j.biotechadv.2009.01.012 (2009).
7. Palmer, J. D., Jansen, R. K., Michaels, H. J., Chase, M. W. & Manhart, J. R. Chloroplast DNA variation and plant phylogeny. *Ann. Missouri Botl. Gard.* **75**, 1180–1206 (1988).
8. Rogalski, M., do N Vieira, L., Fraga, H. P. & Guerra, M. P. Plastid genomics in horticultural species: Importance and applications for plant population genetics, evolution, and biotechnology. *Front. Plant Sci.* **6**, 586. https://doi.org/10.3389/fpls.2015.00586 (2015).
9. Bi, Y. *et al.* Chloroplast genomic resources for phylogeny and DNA barcoding: A case study on *Fritillaria*. *Sci. Rep.* **8**, 1184. https://doi.org/10.1038/s41598-018-19591-9 (2018).
10. Huang, H., Shi, C., Liu, Y., Mao, S. Y. & Gao, L. Z. Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: Genome structure and phylogenetic relationships. *BMC Evol. Biol.* **14**, 151. https://doi.org/10.1186/1471-2148-14-151 (2014).
11. Rawal, H. C., Kumar, P. M., Bera, B., Singh, N. K. & Mondal, T. K. Decoding and analysis of organelle genomes of Indian tea (*Camellia assamica*) for phylogenetic confirmation. *Genomics* **112**(1), 659–668. https://doi.org/10.1016/j.ygeno.2019.04.018 (2020).
12. Zhang, F., Li, W., Gao, C. W., Zhang, D. & Gao, L. Z. Deciphering tea tree chloroplast and mitochondrial genomes of *Camellia sinensis* var. *assamica*. *Sci. Data* **6**(1), 209. https://doi.org/10.1038/s41597-019-0201-8 (2019).
13. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780. https://doi.org/10.1093/molbev/mst010 (2013).
14. Posada, D. & Crandall, K. A. MODELTEST: Testing the model of DNA substitution. *Bioinformatics* **14**, 817–818. https://doi.org/10.1093/bioinformatics/14.9.817 (1998).
15. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033 (2014).
16. Thiel, T., Michalek, W., Varshney, R. K. & Graner, A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* **106**, 411–422. https://doi.org/10.1007/s00122-002-1031-0 (2003).
17. Chen, J. *et al.* The complete chloroplast genome sequence of the relict woody plant *Metasequoia glyptostroboides* Hu et Cheng. *Front. Plant Sci.* **6**, 447. https://doi.org/10.3389/fpls.2015.00447 (2015).
18. Tian, N., Han, L., Chen, C. & Wang, Z. The complete chloroplast genome sequence of *Epipremnum aureum* and its comparative analysis among eight Araceae species. *PLoS ONE* **13**(3), e0192956. https://doi.org/10.1371/journal.pone.0192956 (2018).
19. Vetrivel, U., Arunkumar, V. & Dorairaj, S. ACUA: A software tool for automated codon usage analysis. *Bioinformation* **2**(2), 62–63. https://doi.org/10.6026/97320630002062 (2007).
20. Li, M. M., Meegahakumbura, M. K., Yan, L. J., Liu, J. & Gao, L. M. Genetic involvement of *Camellia taliensis* in the domestication of *Camellia sinensis* var. *assamica* (Assamica Tea) revealed by nuclear microsatellite markers. *Plant Divers. Res.* **37**, 29–37. https://doi.org/10.7677/ynzwyj201514048 (2015).
21. Meegahakumbura, M. K. *et al.* Indications for three independent domestication events for the tea plant (*Camellia sinensis* (L.) O. Kuntze) and new insights into the origin of tea germplasm in China and India revealed by nuclear microsatellites. *PLoS ONE* **11**(5), e0155369 (2016).
22. Wambulwa, M. C. *et al.* Nuclear microsatellites reveal the genetic architecture and breeding history of tea germplasm of East Africa. *Tree Genet. Genomes* **12**, 11. https://doi.org/10.1007/s11295-015-0963-x (2016).
23. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549. https://doi.org/10.1093/molbev/msy096 (2018).

Acknowledgements

The authors are grateful to National Tea Research Foundation, Tea Board, Kolkata, India for financial assistance and Director, ICAR-National Institute for Plant Biotechnology, New Delhi for providing the facility.

Author contributions

T.K.M. conceived the idea, H.C.R. executed all the analysis, T.R.S., S.B., B.B., S.S., R.V.J.I., and A.K.B. gave their inputs while writing the paper. N.K.S. guided the work. H.C.R., T.K.M. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-020-80431-w.

Correspondence and requests for materials should be addressed to T.K.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.