scientific reports

Check for updates

OPEN

Genome-wide identification and development of miniature inverted-repeat transposable elements and intron length polymorphic markers in tea plant (*Camellia sinensis*)

Megha Rohilla[1], Abhishek Mazumder[1], Dipnarayan Saha[2], Tarun Pal[1], Shbana Begam[1] & Tapan Kumar Mondal[1]✉

Marker-assisted breeding and tagging of important quantitative trait loci for beneficial traits are two important strategies for the genetic improvement of plants. However, the scarcity of diverse and informative genetic markers covering the entire tea genome limits our ability to achieve such goals. In the present study, we used a comparative genomic approach to mine the tea genomes of *Camellia sinensis* var. *assamica* (CSA) and *C. sinensis* var. *sinensis* (CSS) to identify the markers to differentiate tea genotypes. In our study, 43 and 60 *Camellia sinensis* miniature inverted-repeat transposable element (CsMITE) families were identified in these two sequenced tea genomes, with 23,170 and 37,958 putative CsMITE sequences, respectively. In addition, we identified 4912 non-redundant, *Camellia sinensis* intron length polymorphic (CsILP) markers, 85.8% of which were shared by both the CSS and CSA genomes. To validate, a subset of randomly chosen 10 CsMITE markers and 15 CsILP markers were tested and found to be polymorphic among the 36 highly diverse tea genotypes. These genome-wide markers, which were identified for the first time in tea plants, will be a valuable resource for genetic diversity analysis as well as marker-assisted breeding of tea genotypes for quality improvement.

Tea is an important plantation crop in India and is widely consumed as a non-alcoholic beverage around the world. As tea is a perennial, woody, cross-pollinated plant[1], the conventional breeding program is extremely slow. Being a recalcitrant plant (i.e., difficult to regenerate in vitro), the transgenic or genome-editing approach for genetic improvement of tea is difficult[2]. Modern tea cultivars still rely primarily on hybridization as a method of genetic improvement. There are three botanical subgroups of tea plants (i.e., Assam, China, and Cambod type) based on morphological parameters, but due to their high outcrossing nature, they can all interbreed freely. The existing tea population today is mostly genetic admixtures of these three types[3]. Therefore, estimating the purity of tea genotypes using molecular markers is an important criterion for precious tea breeding.

Tea breeding is restricted to clonal selection of superior bush from the existing natural population. A systematic breeding technique for tea genetic improvement is not obscure. It is noteworthy to mention that a few draft genomes of tea, including the Assam and China types, have been reported[4,5], providing insights into the tea genome's organization and genetic information. The development of molecular markers using the draft genomes of these two cultivars is one of the useful strategies for tagging the important QTLs and marker-assisted breeding for agronomically important traits. The development of a large number of diverse and informative genetic markers to cover the entire tea genome is thus necessary to accomplish such goals. Several DNA markers in tea plant, such as randomly amplified polymorphic DNA (RAPD), inter simple sequence repeats (ISSR), amplified fragment length polymorphism (AFLP), and simple sequence repeats (SSR), were reported primarily from the

[1]ICAR-National Institute for Plant Biotechnology, LBS Centre, Pusa, New Delhi 110012, India. [2]Division of Crop Improvement, ICAR-Central Research Institute for Jute and Allied Fibres, Barrackpore, Kolkata, West Bengal 700121, India. ✉email: mondaltk@yahoo.com