

(05-5-02)

Brassica-Arabidopsis Genome Browser: Overview of Brassica Genome based on Comparative Genome Analysis with Arabidopsis

Tae-Jin Yang¹, Jungsun Kim¹, Ki-Byung Lim¹, Soo-Jin Kwon¹, Jin-A Kim¹, Mina Jin¹, Jee Young Park¹, Beom-Soon Choi¹, Hyo-Jin Lee¹, Myung-Ho Lim¹, Ho-Il Kim¹, Seok-Hyoung Kim¹, Yong Pyo Lim², Seung-Wook Lee³, Tae-Suk Park³, Jin-Han Hong³,
and Beom-Seok Park^{1*}

1. Brassica Genomics Team, National Institute of Agricultural Biotechnology (NIAB), 224 Suinro Gwonseon-gu, Suwon, Gyeonggi-do, 441-707, Republic of Korea 2. Chungnam National University, Daejeon, , 305-764, Republic of Korea 3. MacroGen Co. Gasan-dong, Seoul 153-023.

The complete genome sequence provides unlimited information in the sequenced organism as well as in the related taxa. Korea Brassica Genome Project (KBGP) have shotgun sequenced 70 BAC clones of which 50% are for chromosome #1 sequencing and the others are for the study of genome evolution and flowering related genes. Comparative genome analyses of the sequenced BAC clones revealed overall collinearity with 82% sequence similarity with their homeologous partner regions on *Arabidopsis* genome. Comparison of all the Brassica BAC clones corresponding to an *Arabidopsis* chromosome segment represents that *Brassica* genome is undertaken triplication and subsequent gene losses after divergence of *Arabidopsis* and *Brassica*. We have obtained 91,179 BAC end sequences (BES) from 46848 BAC clones originated three BAC libraries (*Hind*III, *Bam*HI, and *Sau*3AI). All BES were used for comparative genome analysis with the *Arabidopsis*. A total of 45232 (45%) BES hits significantly (E-6) on a spot of *Arabidopsis* pseudo chromosomes. And a total of 4317 BAC clones (9.5%) landed on *Arabidopsis* chromosomes by directional matches of both ends (8634 BES), within 30-500 kb interval on *Arabidopsis* chromosome. These 4317 clones are grouped into 54 pseudo contigs (13, 7, 12, 12, and 10 for *Arabidopsis* chromosome 1, 2, 3, 4, and 5, respectively, the biggest contig is 7.3 Mb on chromosome 5) based on overlapping of the hit sites on *Arabidopsis* chromosomes. The 54 pseudo contigs span a total of 72 Mb of *Arabidopsis* euchromatin sequence without internal gaps. Each contig may represent the counterpart chromosomes of the triplicated Brassica genome. About 35 Mb of *Arabidopsis* sequences remained gaps uncovered. More than 95% of the gaps are from heterochromatin blocks (70% is from pericentromeric, 5% is from telomeric, and 20% is from heterochromatin knob). The remaining 5% uncovered region which is from about 3 Mb euchromatin regions is composed of 40 small gaps (less than 200 kb). The 40 small gaps may represent the chromosomal rearrangements between *Arabidopsis* and *Brassica*. Single *Brassica* BAC clone spans an average of 147 kb (\pm 74 kb) *Arabidopsis* sequence (Fig. 3). We have selected a total of 629 BACs that are on the minimum tiling path of 86 Mb *Arabidopsis* genome. The sequence and chromosomal locations of the 629 minimum tiled Brassica BACs will show the comparative overview of *Arabidopsis* and *Brassica* genome and will be provided for the Brassica genome sequencing. All the comparative genome analysis of the BAC and BES are available from our *Arabidopsis*-Brassica Genome Browser (www.brassica-rapa.org) showing the positions of Brassica BAC clones on the counterpart *Arabidopsis* chromosomes. The genome browser will give great opportunity to the enlargement of understanding on evolution of polyploidized Brassica genome.

*Corresponding author : Beom-Seok Park, Tel : 031-299-1677, E-mail : pbeom@rda.go.kr; tjyang@rda.go.kr)200