

Comparative genomics of *Vibrio campbellii* strains and core species of the *Vibrio* Harveyi clade

Huei-Mien Ke^{1,2}, Anuphap Prachumwat^{3,4}, Chun-Ping Yu², Yi-Ting Yang^{5,6}, Sutitcha Promsri^{3,7}, Kuan-Fu Liu⁸, Chu-Fang Lo^{5,6,9}, Mei-Yeh Jade Lu², Mei-Chin Lai^{1,10,11}, Isheng J Tsai^{1,2,*}, and Wen-Hsiung Li^{1,2,12,*}

1 Ph.D. Program in Microbial Genomics, National Chung Hsing University and Academia Sinica, Taiwan.

2 Biodiversity Research Center, Academia Sinica, Taipei, Taiwan.

3 Centex Shrimp, Faculty of Science, Mahidol University, Bangkok, Thailand.

4 Shrimp-Virus Interaction Laboratory, Animal Biotechnology Research Unit, National Center for Genetic Engineering and Biotechnology (BIOTEC), National Science and Technology Development Agency (NSTDA), Pathum Thani, Thailand.

5 Institute of Bioinformatics and Biosignal Transduction, College of Bioscience and Biotechnology, National Cheng Kung University, Tainan, Taiwan.

6 Department of Life Science, National Taiwan University, Taipei, Taiwan.

7 Department of Biotechnology, Faculty of Science, Mahidol University, Bangkok, Thailand.

8 Tungkang Biotechnology Research Center, Fisheries Research Institute, Council of Agriculture, Pingtung, Taiwan.

9 Center of Bioscience and Biotechnology, National Cheng Kung University, Tainan, Taiwan.

10 Department of Life Sciences, National Chung Hsing University, Taichung, Taiwan.

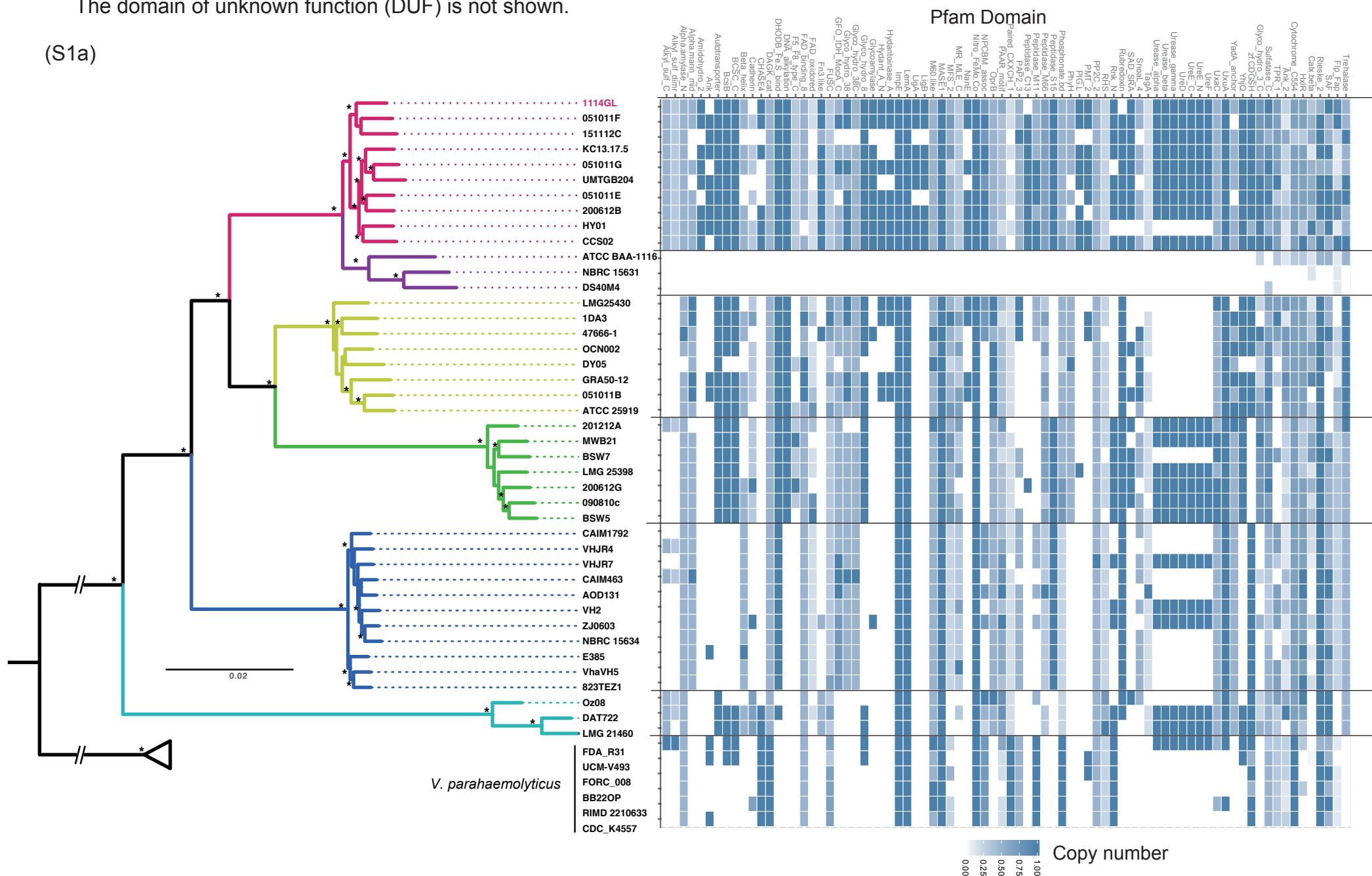
11 Agricultural Biotechnology Center, National Chung Hsing University, Taichung, Taiwan.

12 Department of Ecology and Evolution, University of Chicago, Chicago, US.

*Author for Correspondence: Isheng J Tsai, Biodiversity Research Center, Academia Sinica, Taipei, Taiwan, ijtsai@gate.sinica.edu.tw, and Wen-Hsiung Li, Biodiversity Research Center, Academia Sinica, Taipei, Taiwan, whli@uchicago.edu.

Figure S1. Protein domain enrichment across strains. Protein domains expanded (Wilcoxon rank sum test, $p < 0.05$; 4-fold copy number) either in Group 1 (S1a) or Group 2 (S1b) were colored from light to dark blue and the number is rescaled between 0 to 1 (for each domain, copy number from each of 48 strains is divided by the maximum copy number). The names of protein domains are labeled in the horizontal axis and the names of strains are labeled in the vertical axis. The domain of unknown function (DUF) is not shown.

(S1a)



(S1b)

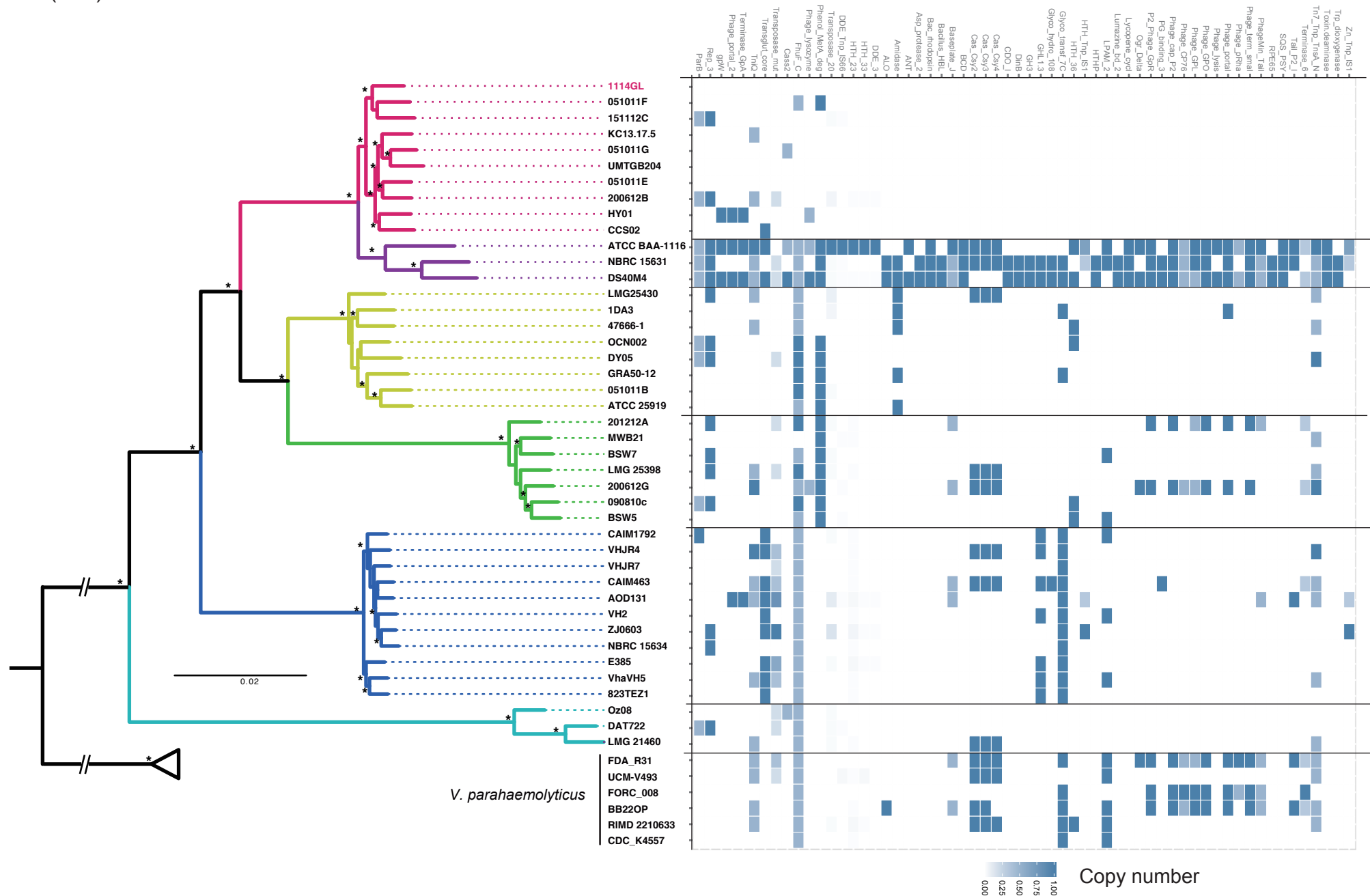


Figure S2a. The carbohydrate fermentation test of 1114GL showed acid production when using D-galactose as the sole carbon source.

Tube 1: basal medium without D-galactose and without 1114GL;

Tubes 2-4: basal medium without D-galactose but inoculated with 1114GL;

Tube 5: basal medium with D-galactose but without 1114GL;

and Tubes 6-8: basal medium with D-galactose but inoculated with 1114GL.

The color was observed after 10 hours.

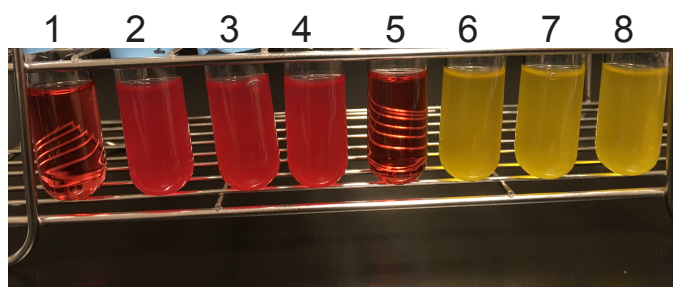


Figure S2b. The growth curves of 1114GL on basal medium and on basal medium with D-galactose. Three replicates were conducted for both conditions.

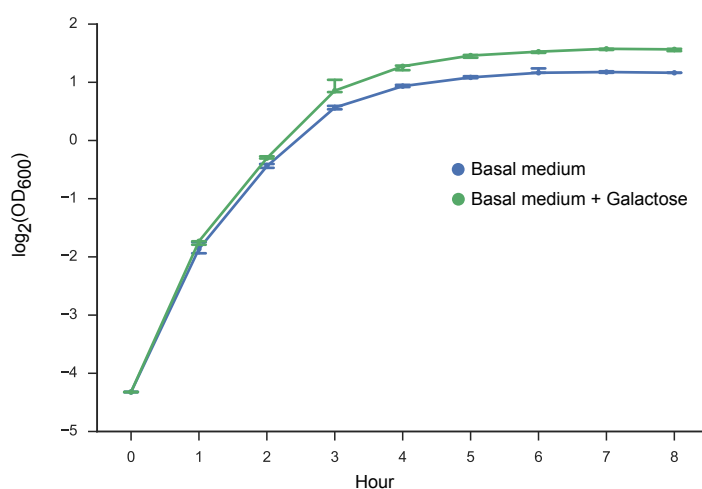


Figure S3. The number of the proteins related to galactose fermentation in *V. campbellii*.

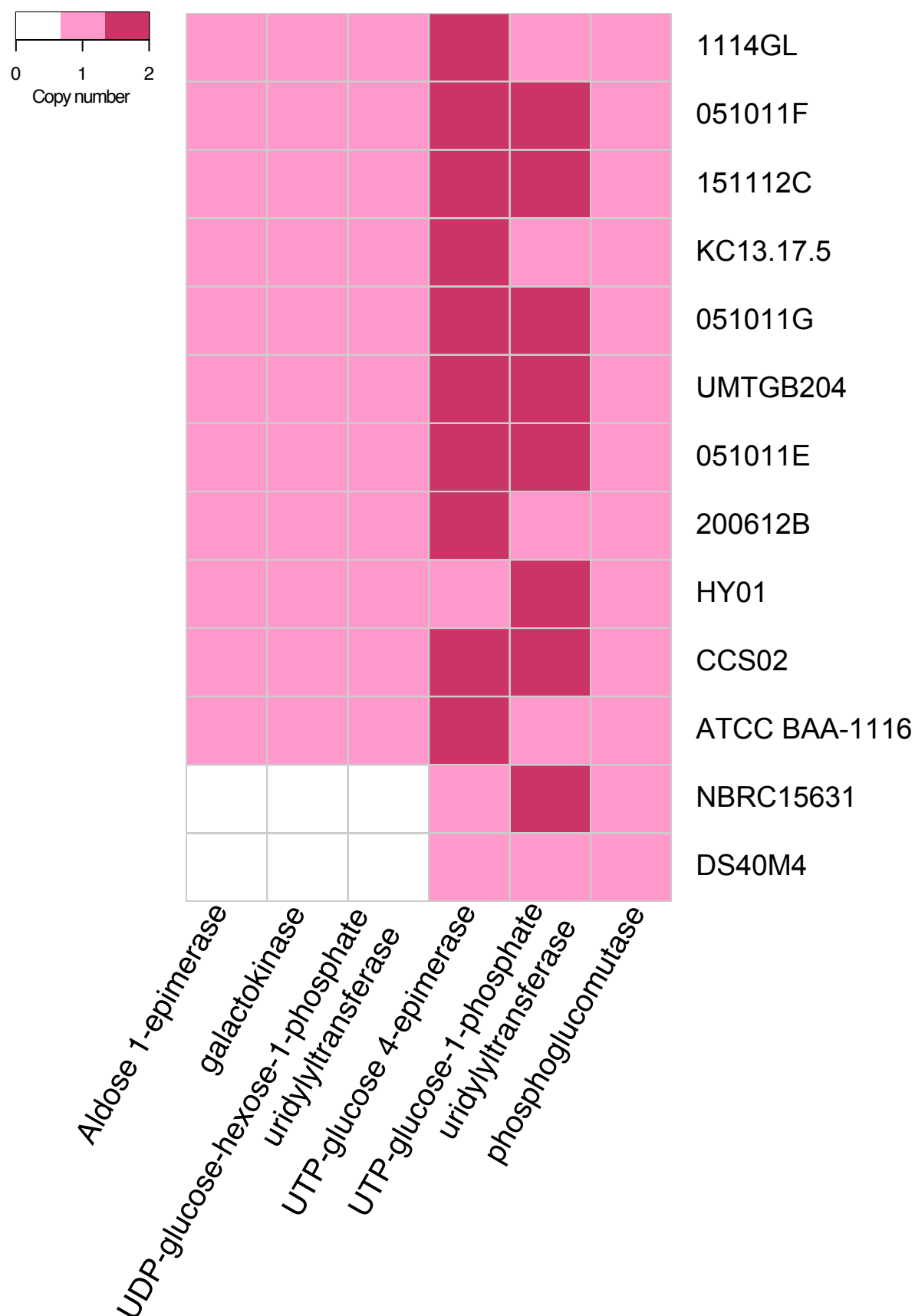


Figure S4. The sequence alignment of Chromosome I (a) and Chromosome II (b) between *V. campbellii* 1114GL and ATCC BAA-1116.

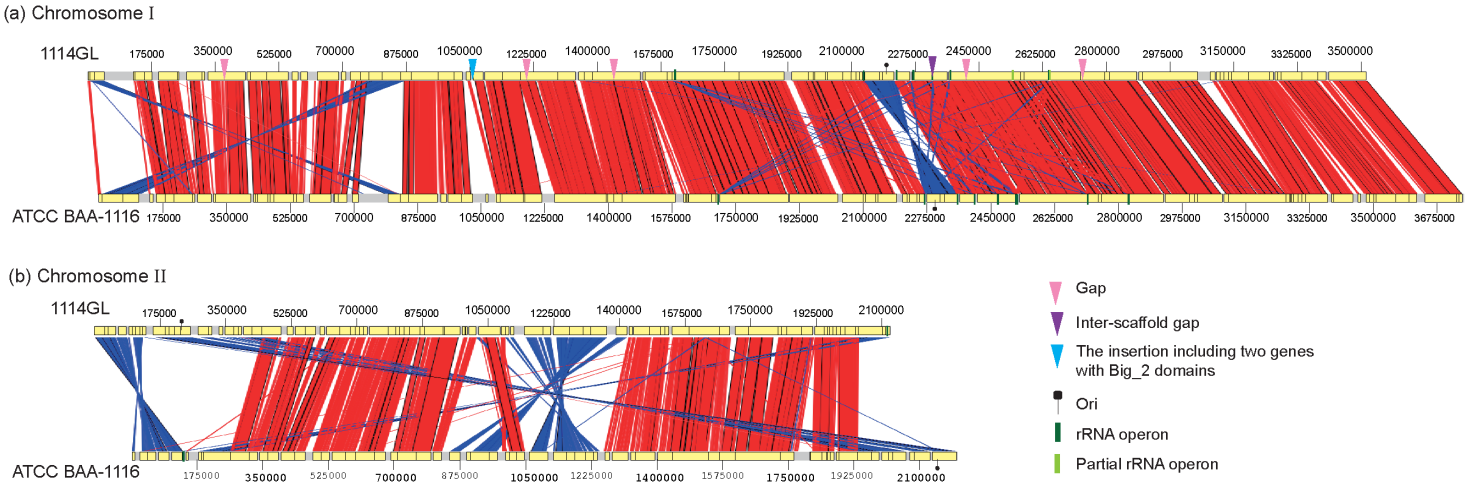


Figure S5. Correlation between transposase gene number and the size of syntenic break.

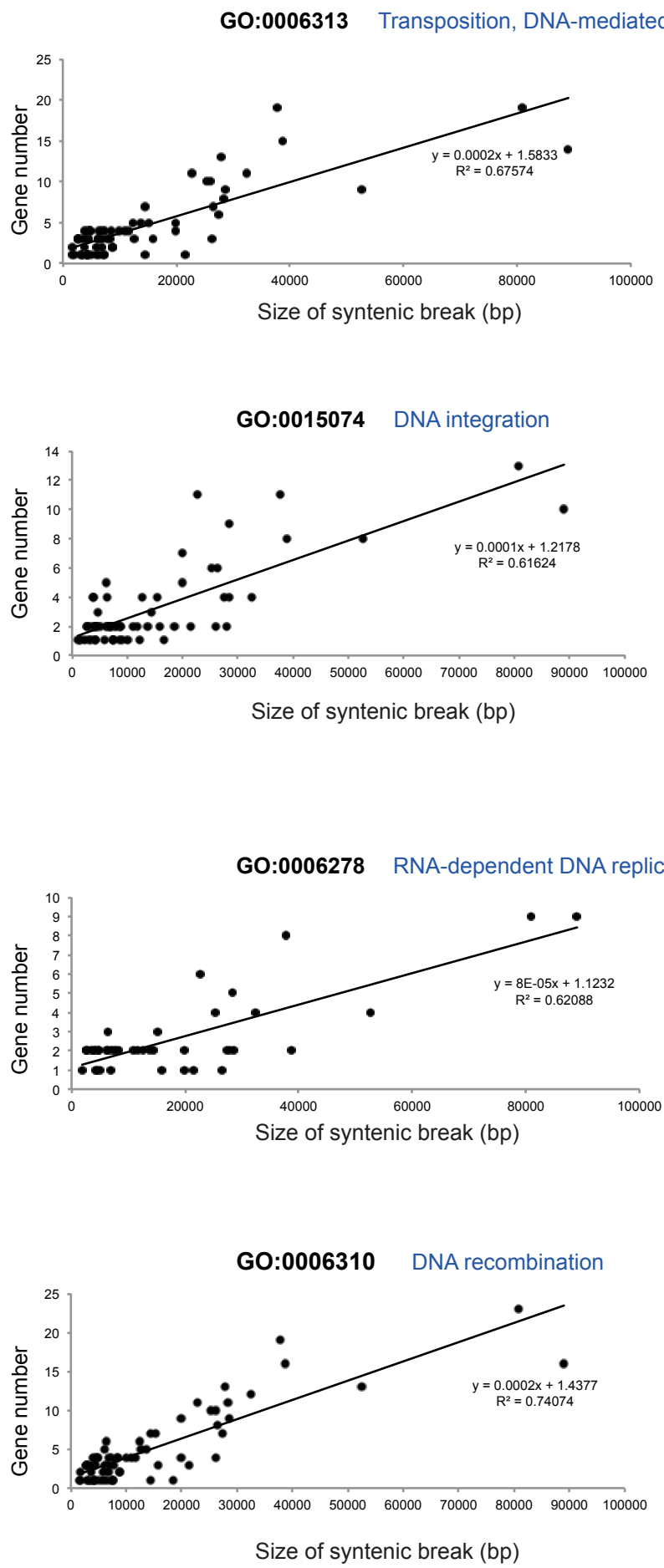


Figure S6. Phylogenetic tree based on the transposase protein sequences from the synteny break. The color of a lineage is labeled according to its position in the synteny break.

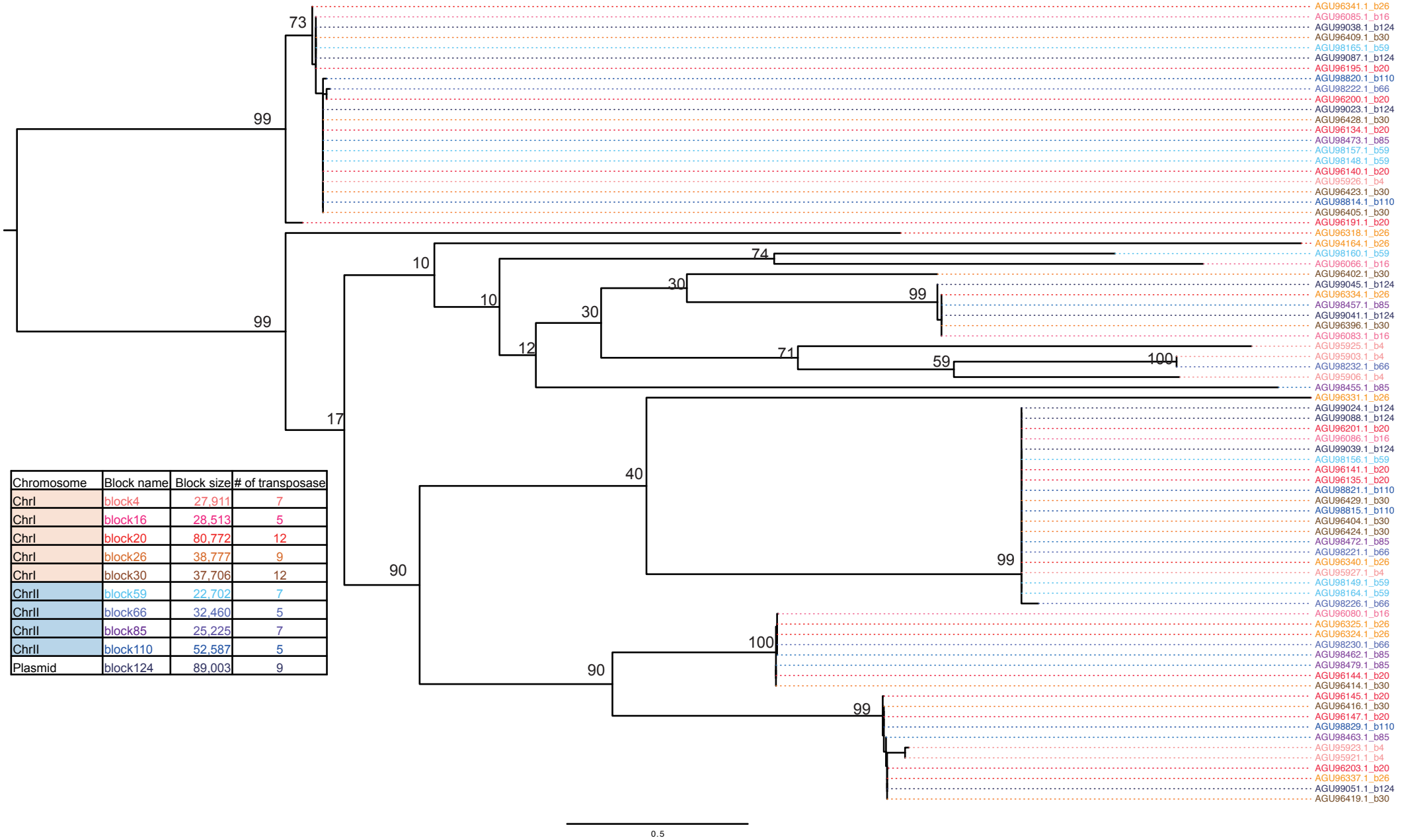


Figure S7a. Diagram of *V. campbellii* 1114GL superintegron gene cassette array. 39 *attC* sites were predicted across the coding sequence from Vca1114GL_00040 to Vca1114GL_00146 (box arrow).

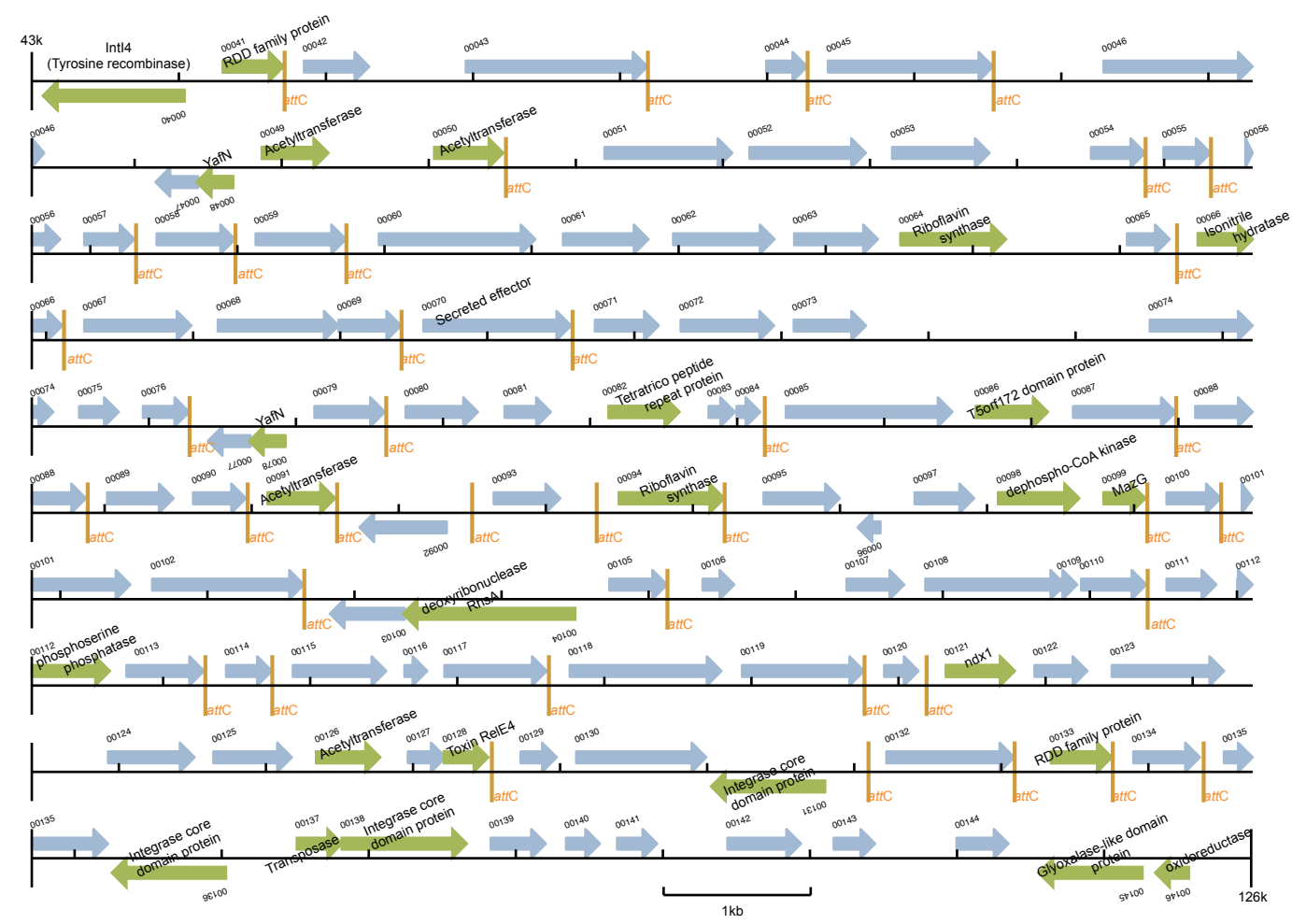


Figure S7b. The alignment of 39 predicted *attC* site containing partially palindromic sequences (labeled with R', R'', L', and L'').

