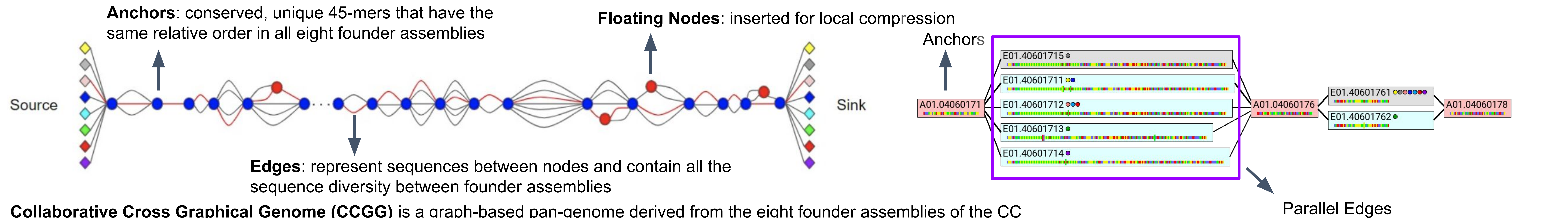




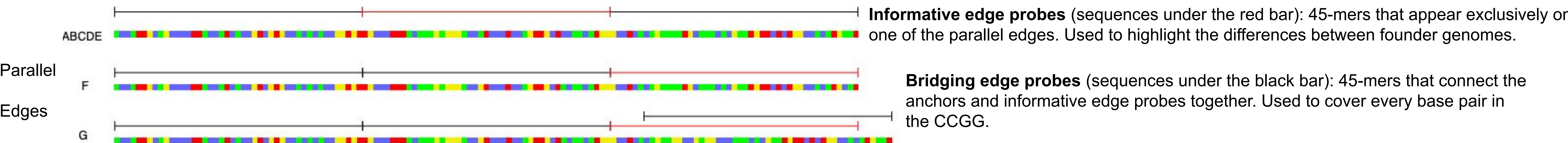
An Online Viewer and Database for the Collaborative Cross Graphical Genome

Ziwei Chen¹, Hang Su^{1,2}, Maya Najarian¹, Fernando Pardo Manuel de Villena³ and Leonard McMillan¹

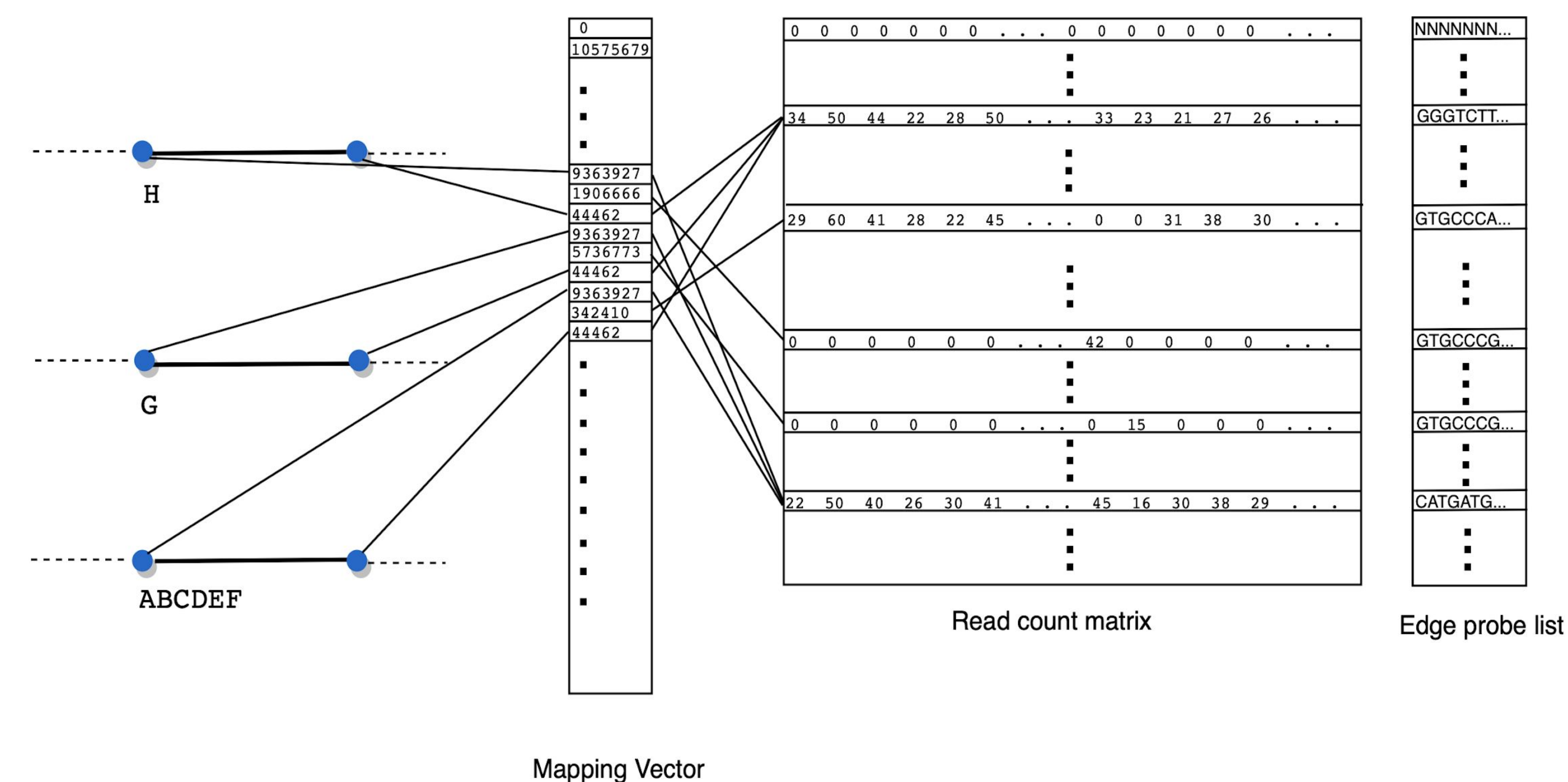
¹Department of Computer Science, University of North Carolina, Chapel Hill, NC 27514, USA; ²Curriculum of Bioinformatics and Computational Biology, University of North Carolina, Chapel Hill, NC 27514, USA; ³Department of Genetics, University of North Carolina, Chapel Hill, NC 27514, USA



Edge Probes: 45-mers from the CCGG to help characterize the CC samples and founder assemblies



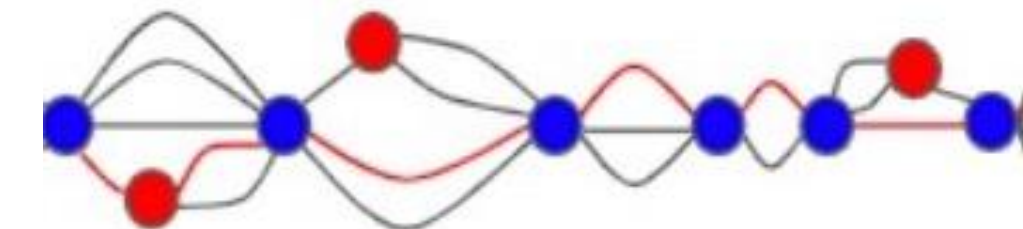
All edge probes derived from the CCGG are queried in the sequencing data of 96 CC samples and 8 newly sequenced CC founders using the msBWT for read counts. The result was organized using a mapping vector to reduce the redundancy. Edge probes with the same sequence content are always mapped to the same row in the read count matrix and edge probe list.



The total number of probes selected from the CCGG: 289,683,815
The total number of entries in the read count matrices: 148,183,468
The total number of 45-mers from reference genome ~ 58,750,000

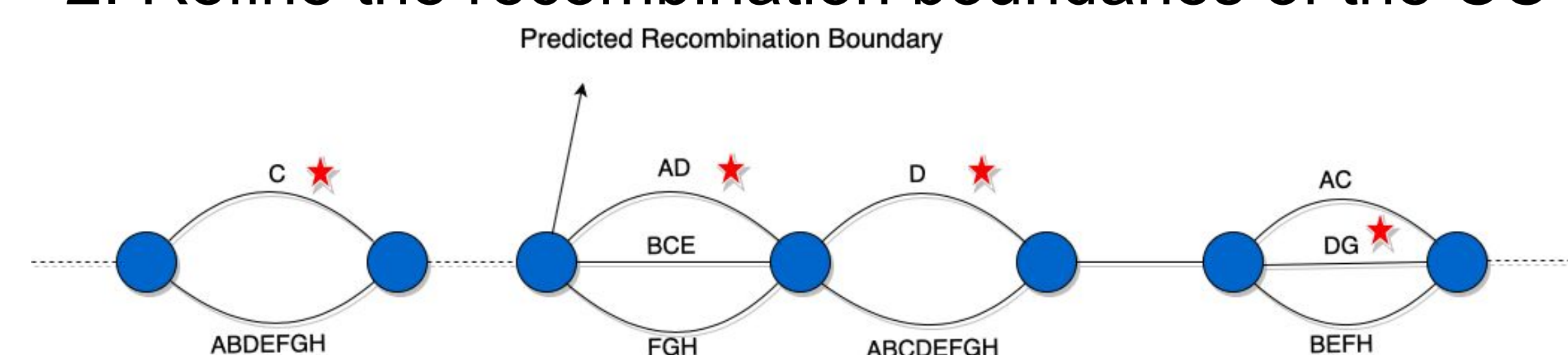
Applications

1. Compress the CCGG



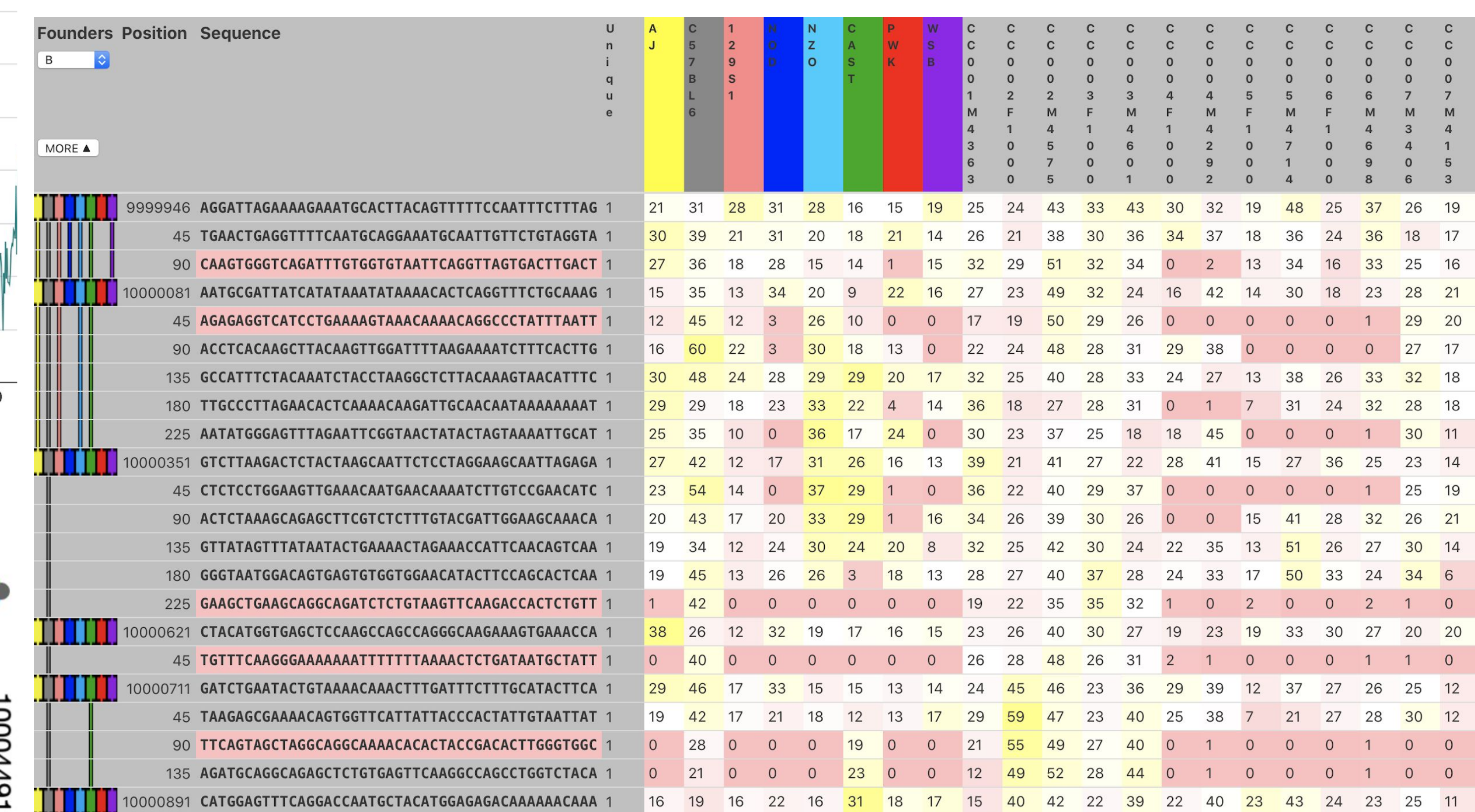
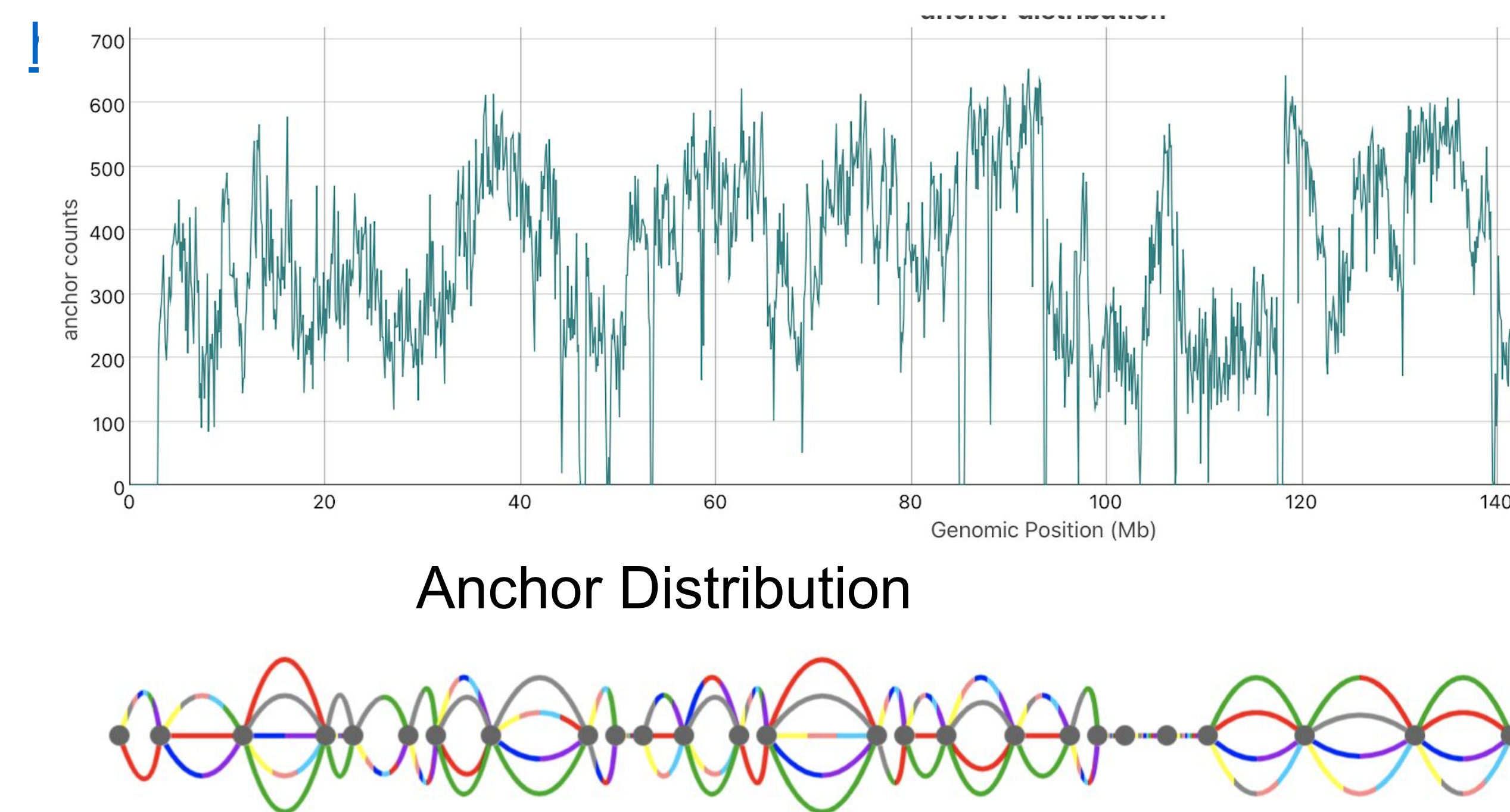
The mapping vectors reveal the local sharing of 45-mers
3,198,957 floating nodes are inserted into the CCGG for compression

2. Refine the recombination boundaries of the CC

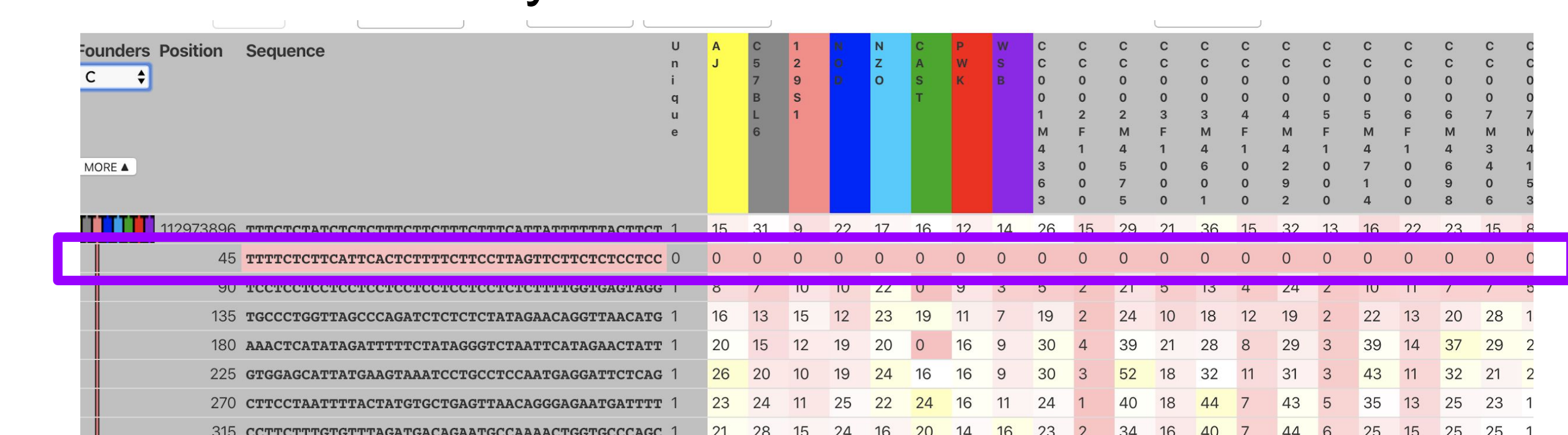


There are 7,583 recombination in 75 CC strains across all chromosomes
For 87% of them, we are able to refine the recombination boundary.

The online CCGG viewer



3. Locate assembly errors



1,040,276 unsupported edge probes are located