

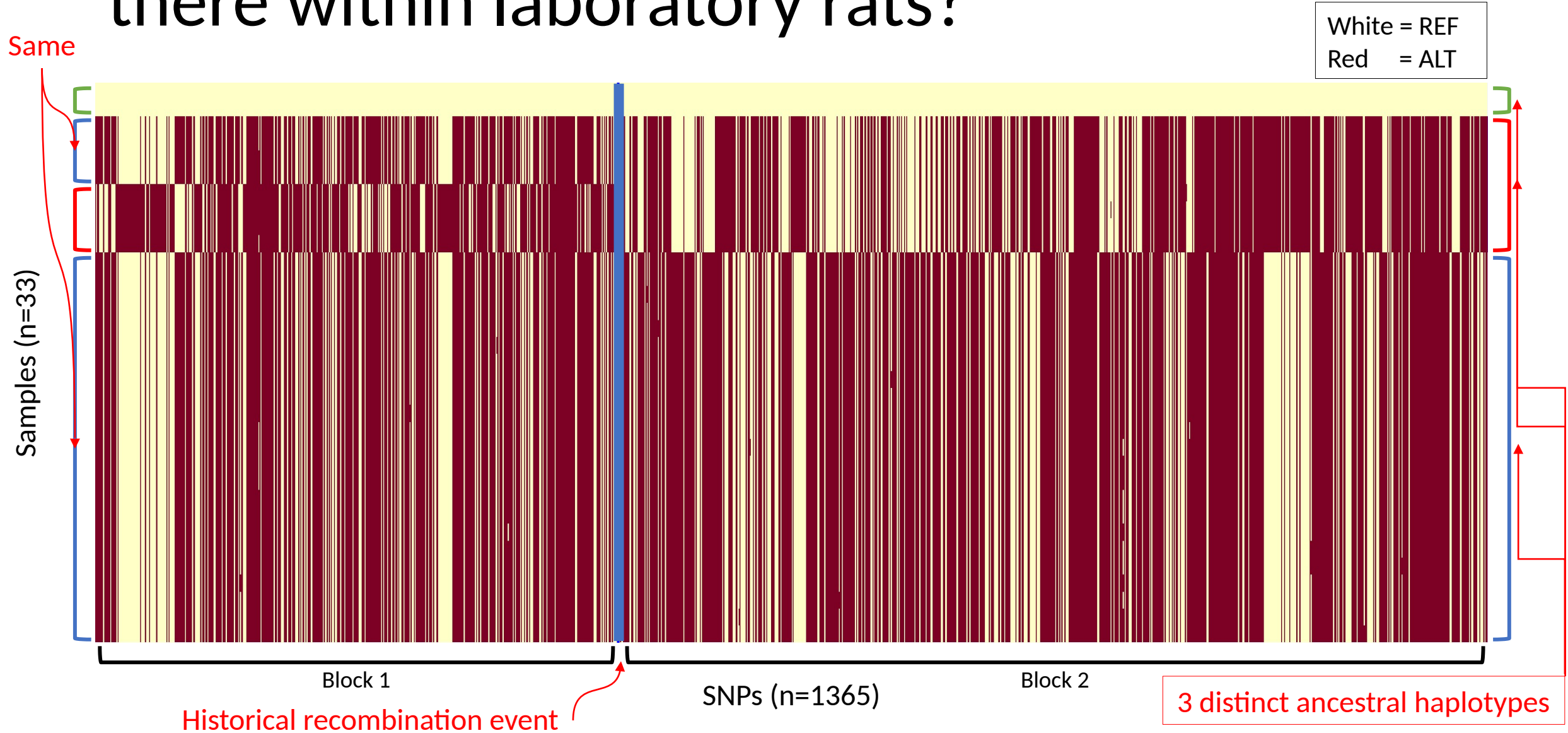
Haplotype-based analyses of phylogeny and regional genome diversity in laboratory rats

Yanchao (Han) Pan¹, Hao Chen², Jun Z. Li¹

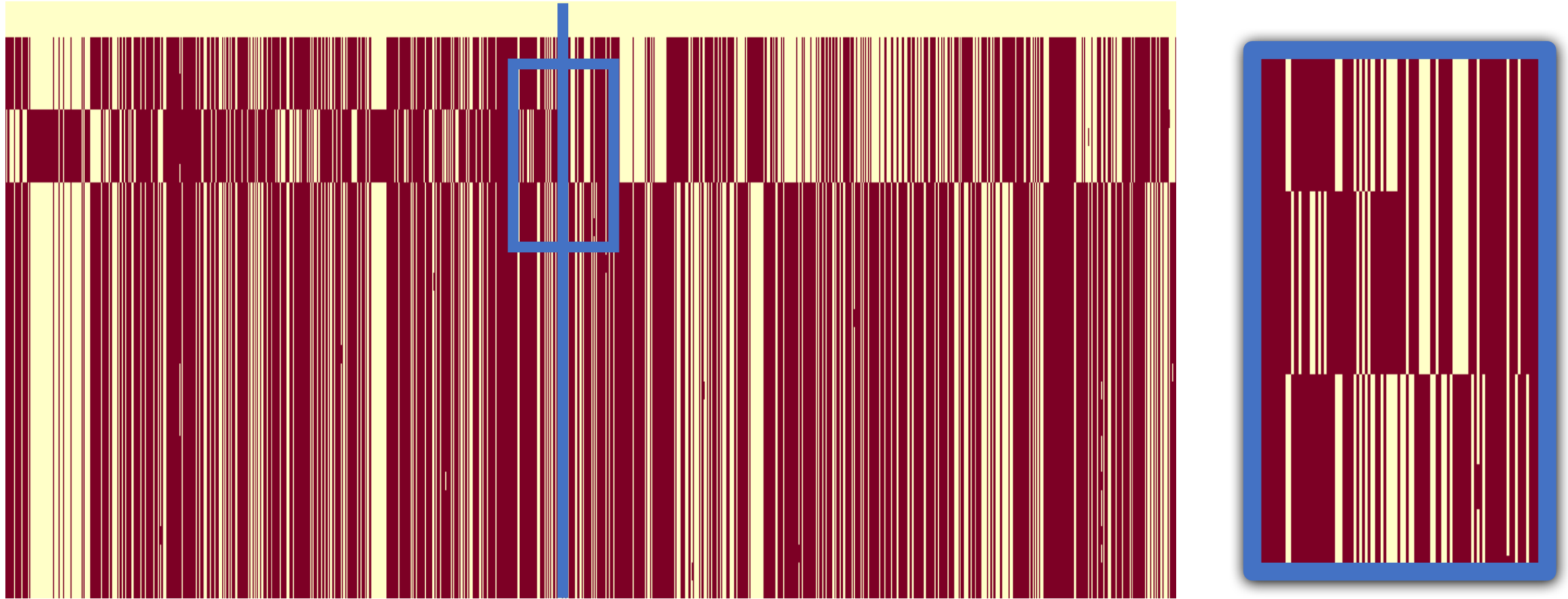
¹ Department of Human Genetics, University of Michigan, Ann Arbor, MI, USA.

² Department of Pharmacology, Addiction Science, and Toxicology, University of Tennessee Health Science Center, Memphis, TN, USA.

How many distinct ancestral haplotypes are there within laboratory rats?

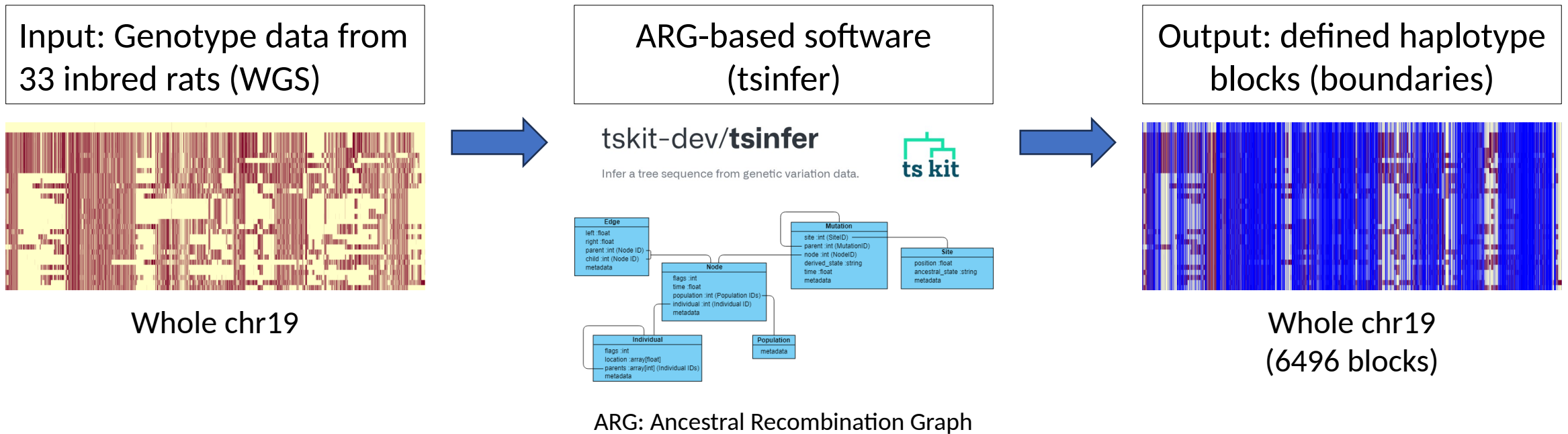


Step 1: divide genome into haplotype blocks

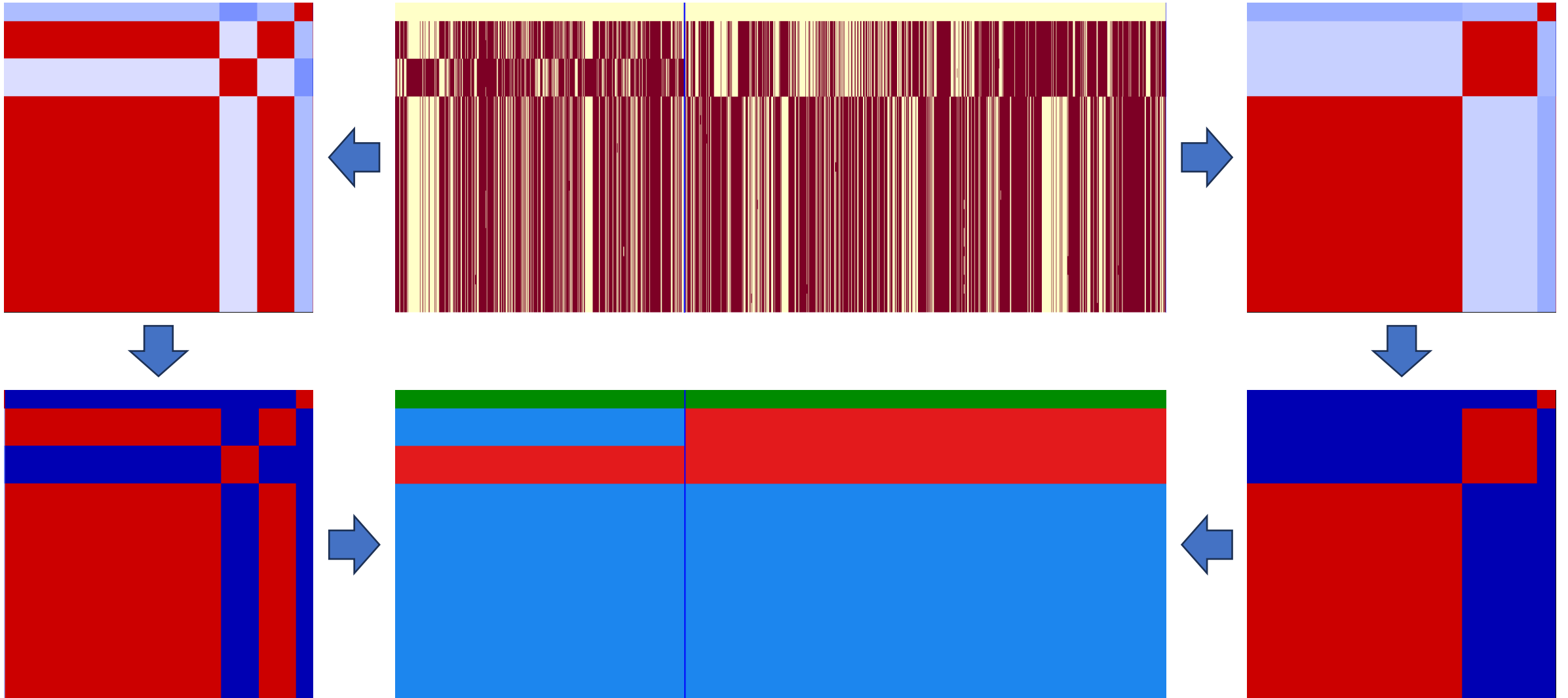


Find these boundaries at whole chromosomal scale

Step 1: divide genome into haplotype blocks



Step 2: define distinct ancestral haplotypes within each haplotype blocks



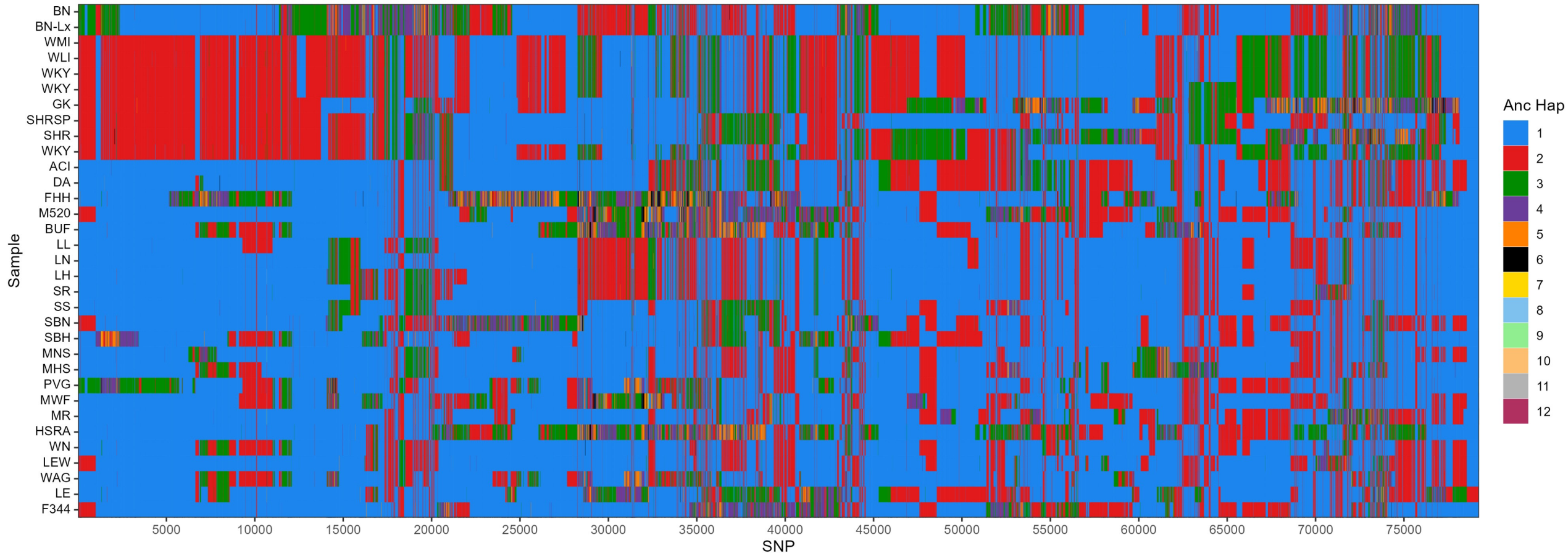
Step 2: define distinct ancestral haplotypes within each haplotype blocks

ah_chr19_s300__ts_vm0.95

Number of samples = 33 | Number of blocks = 6496 | Number of sites = 79250 | Variante sites per Kb = 1.38 | Jump % = 27.98 | Flicker % = 10.3

Block length: Median = 5, Mean = 12 | Block length physical: Median = 2214.5, Mean = 8091 | Number of AH per block: Median = 3, Mean = 3.45, Weighted Mean = 3.85, Max = 12

chr19



Number of blocks: 6496

Step 3: horizontal merge over-segmented blocks

- Treat every 2 adjacent blocks as a potential merging candidate
- For each merging candidate we calculate the cost of merging

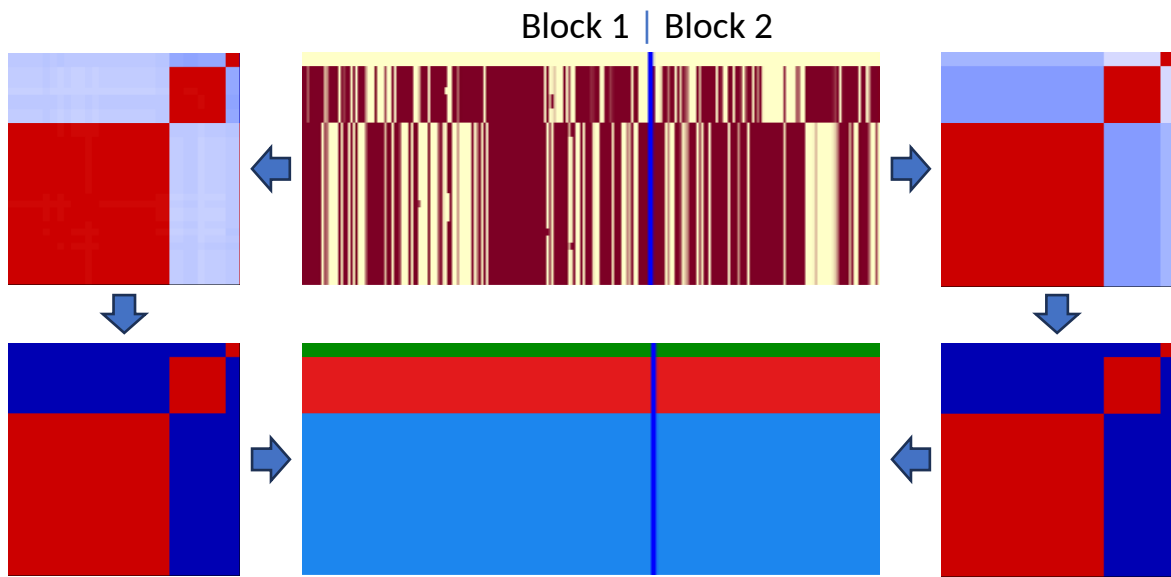
$$\text{Cost} = D_1 * L_1 + D_2 * L_2$$

D_1 : discordance between block 1 and the merged block

L_1 : length of block 1

D_2 : discordance between block 2 and the merged block

L_2 : length of block 2



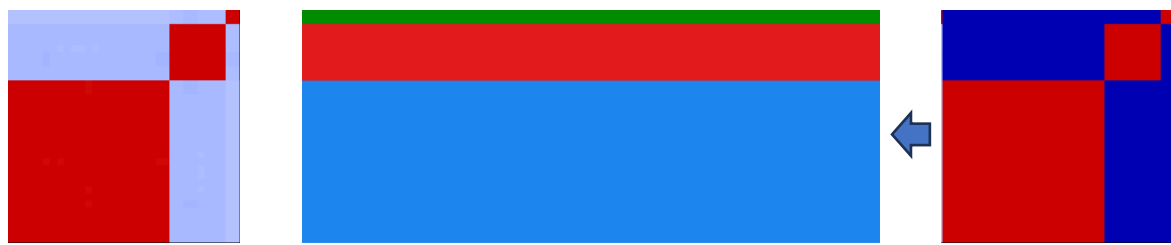
$$D_1 = 0$$

$$L_1 = 163$$

$$D_2 = 0$$

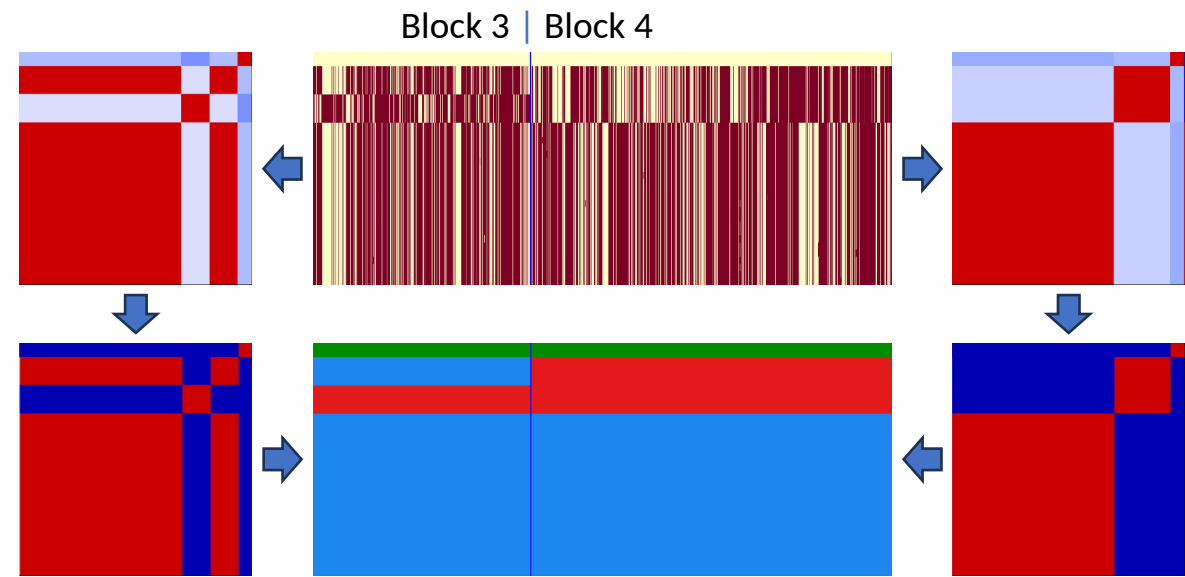
$$L_2 = 108$$

Block (1+2)



$$\text{Cost} = D_1 * L_1 + D_2 * L_2 = 0 * 163 + 0 * 108 = 0$$

Merge



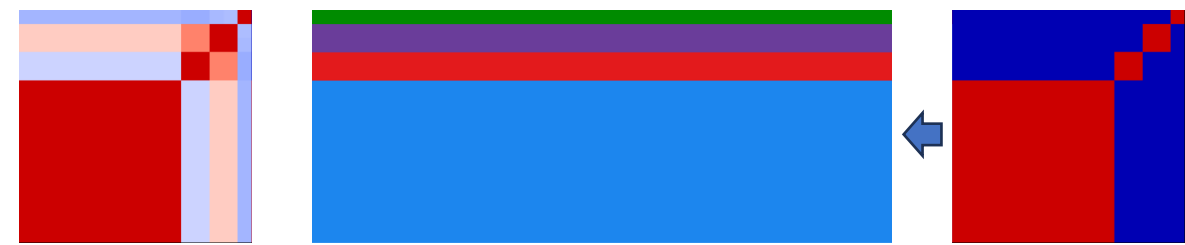
$$D_1 = 4$$

$$L_1 = 512$$

$$D_2 = 4$$

$$L_2 = 853$$

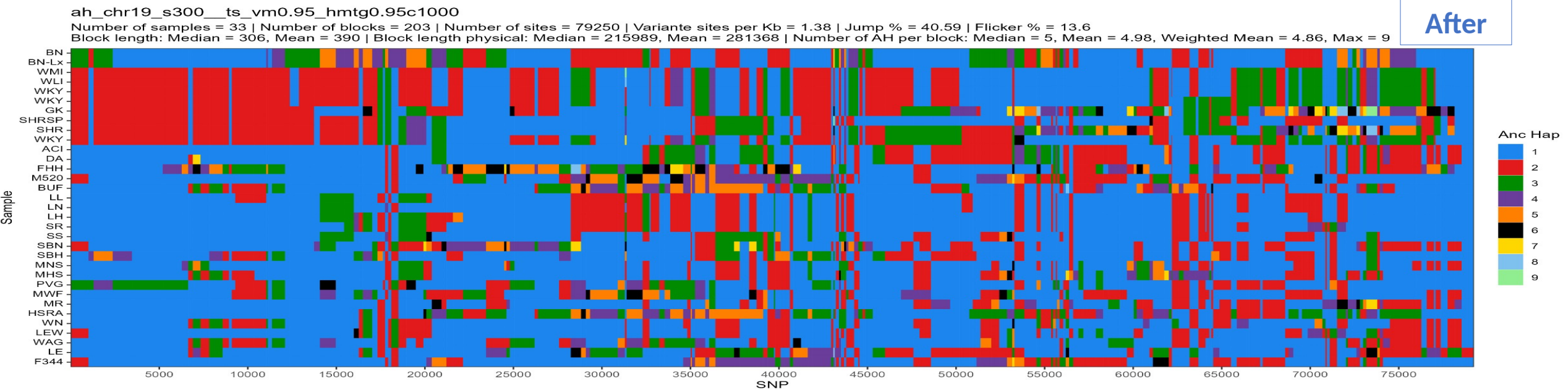
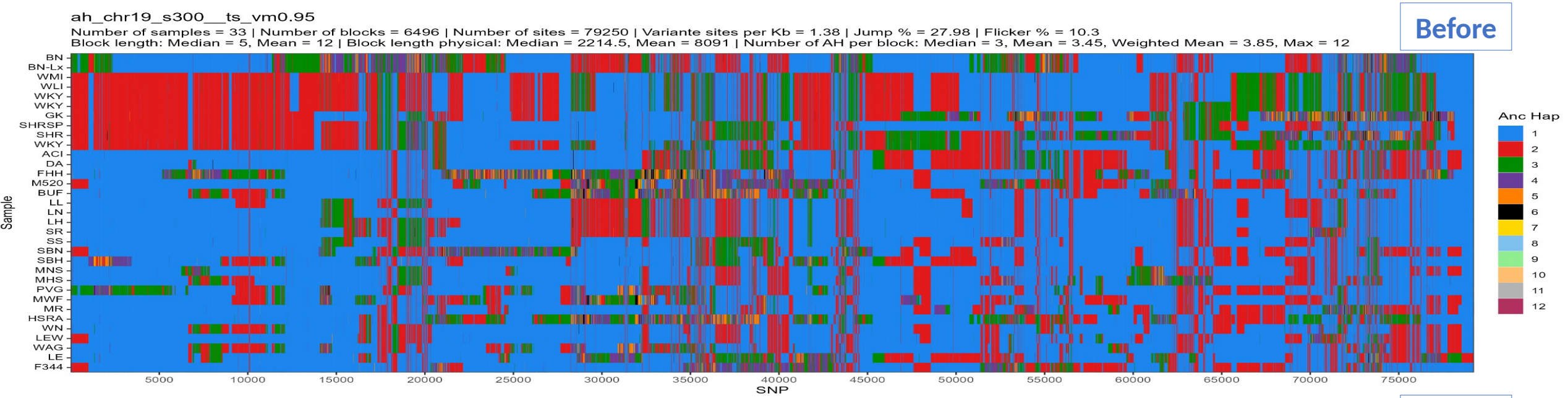
Block (3+4)



$$\text{Cost} = D_1 * L_1 + D_2 * L_2 = 4 * 512 + 4 * 853 = 5460$$

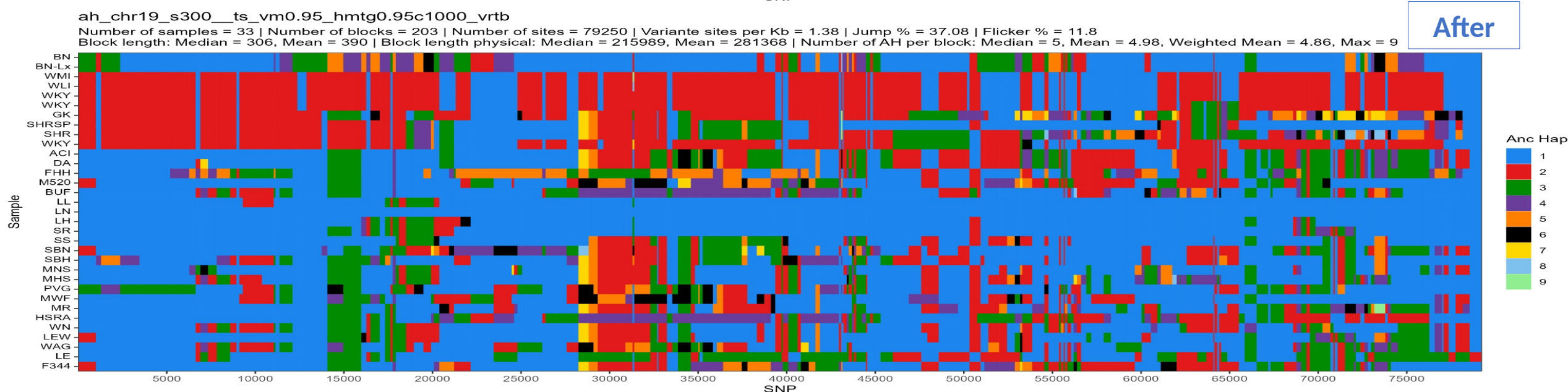
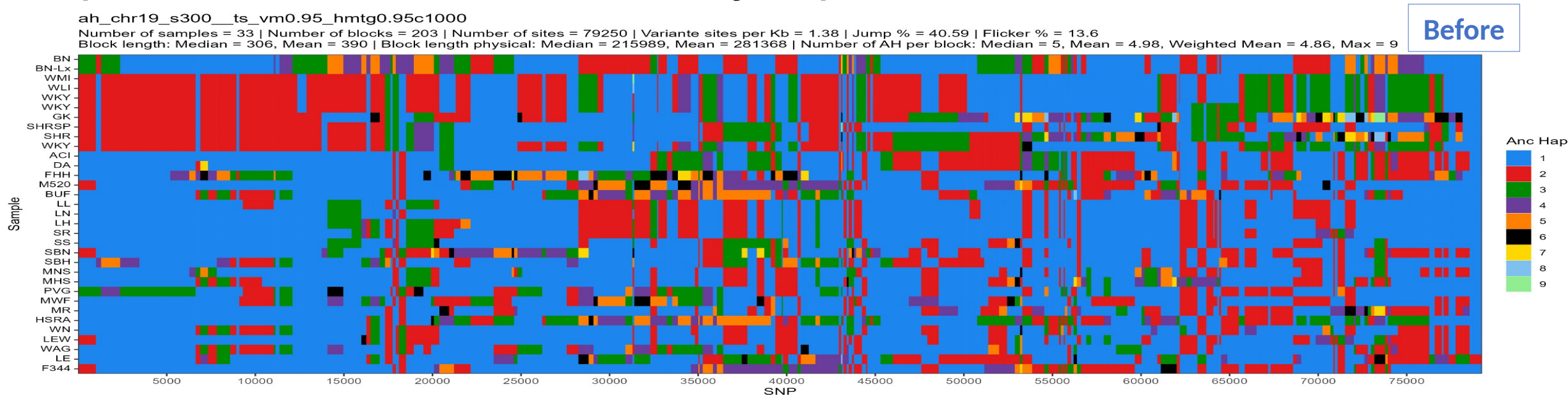
No merge

Step 3: horizontal merge over-segmented blocks



Number of blocks: 6496 -> 203

Step 4: vertical rename to reduce jumps and flickers



Jump %: 41.59 -> 37.08

Flicker %: 13.6 -> 11.8

Summary

- Extensive sharing of ancestral genome materials when the modern-day laboratory rat strains were created
 - Most genomic regions contains only 2-6 distinct ancestral haplotypes
 - The 5 most common ancestral haplotypes make up more than 95% of the overall genomes
- Insights into the evolution of rat strains and inform regional-specific diversity to support haplotype-based genetic mapping/association studies.
- The methodologies we developed do **not** require a reference founder panel and should have wider adaptability.

Acknowledgment

Jun Li lab @ University of Michigan

- Jun Li

Dissertation committee

- Jun Li
- David Burke
- Jacob Mueller
- Jeffrey Kidd
- Xiang Zhou

International Rat Omics Consortium (IROC)

- Hao Chen
- Abraham Palmer
- Robert Williams
- Melinda (Mindy) Dwinell
- Anne Kwitek
- Peter A Doris
- Terence Murphy
- Fergal Martin
- Tristan de Jong