Supplement to Chapter 3 from the dissertation:

Anderson, T (2020) Genetics and breeding of early blight and bacterial spot resistant tomatoes. Cornell University, Ithaca NY

Fig. S1 Silhouette scores and the number of clusters corresponding as a function of *d* (distance merging threshold) for three genomic window sizes. The data are from 780 tomato sequences beginning at SL4.0 Chromosome 9 position 62,512,575. When the genetic distance between two samples falls under a distance threshold (*d*) during hierarchical clustering, these samples are clustered together. The average pairwise distance among samples increases with the size of the genomic window, as a larger window captures more genotypic information. As a result, the optimal value of *d* will also increase. A value of *d* that is too low will result in too many clusters, while a *d* that is too large will give too few clusters. Thus, the user sets a range of *d* values for the algorithm to investigate that balances classification accuracy and sensitivity

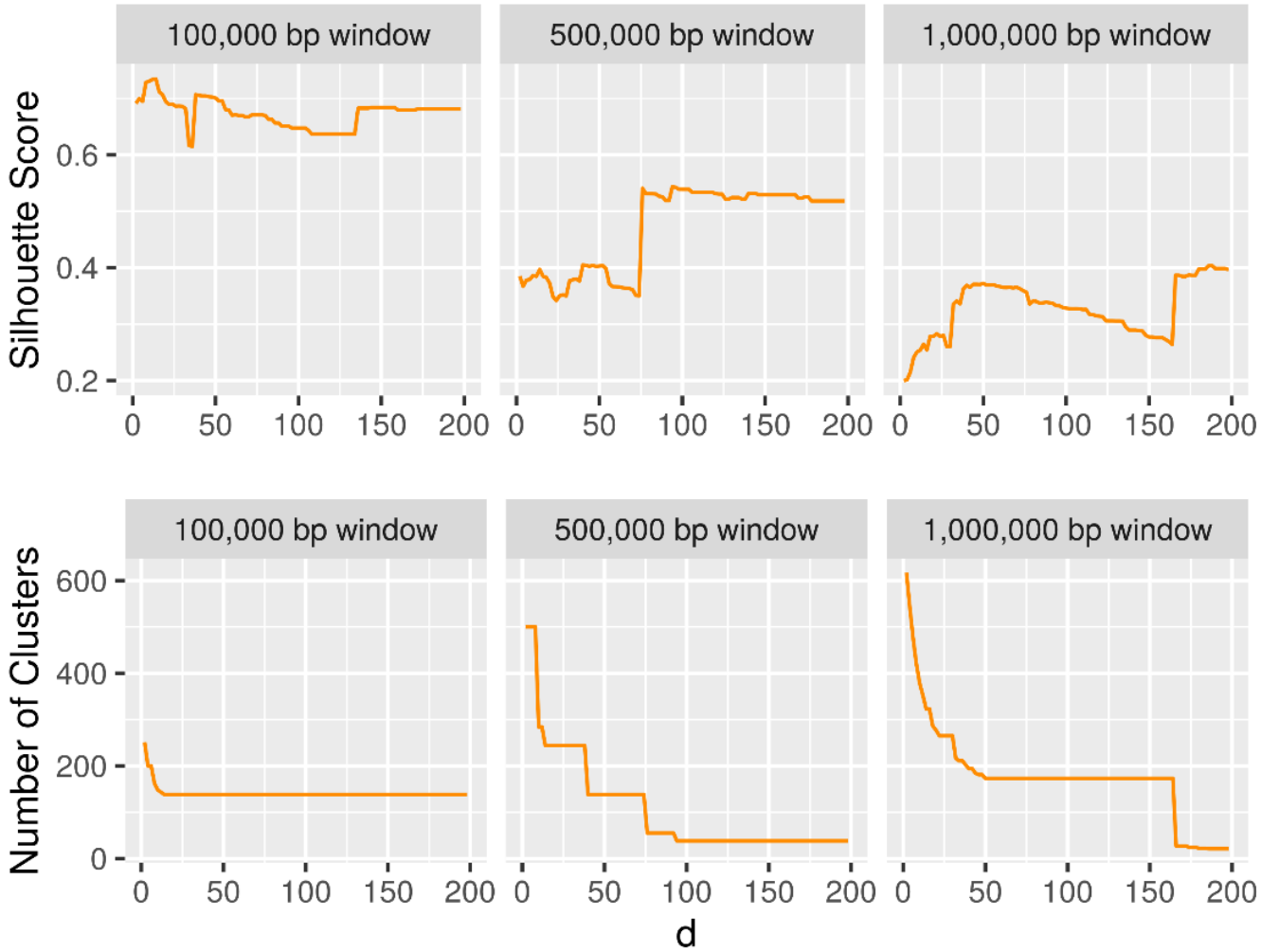
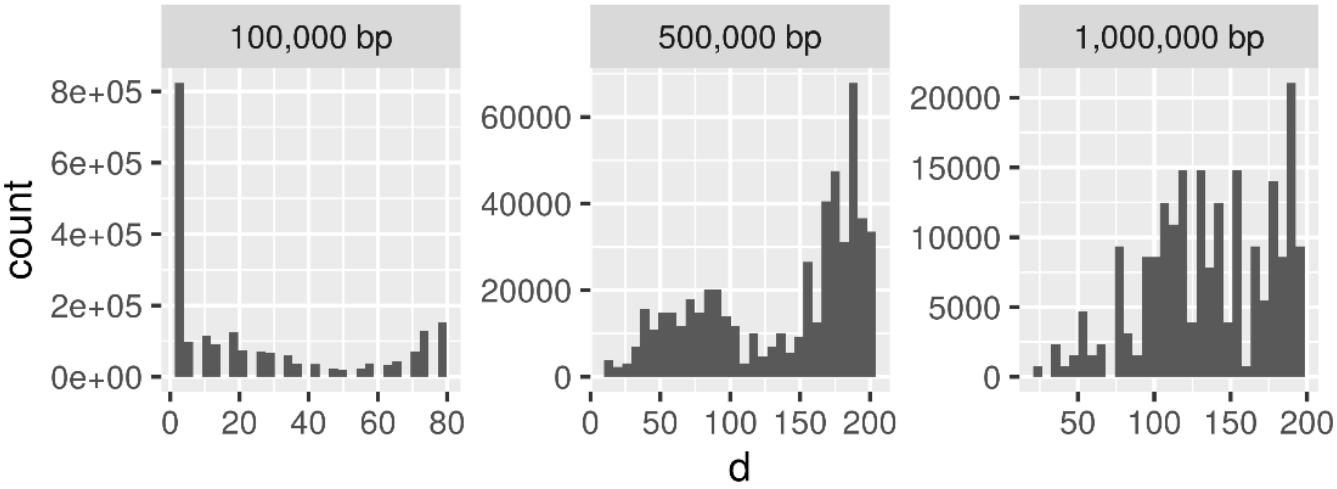


Fig. S2 A. Distribution of clustering distance thresholds (*d*) as determined from the hierarchical classification algorithm for all genomic windows on chromosome 9. The analysis included 780 individual sequences for tomatoes and wild relatives **B.** Genome-wide distribution of algorithmically determined *d* thresholds for a sliding window size of 250 Kb and step size of 100 Kb for 780 genome sequences

A



B

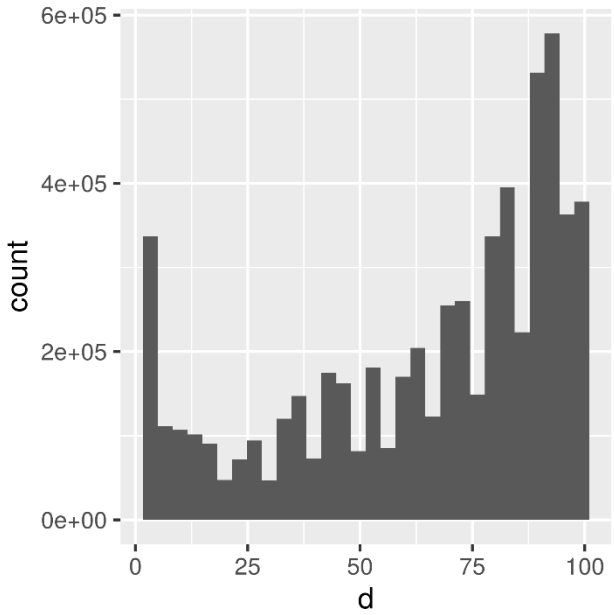


Fig. S3 The haplotype clustering methodology was evaluated by attempting to detect introgressions in a well-known breeding line, CU17NBL. Tomato CU17NBL is known to have the following introgressions from *Solanum pennellii* LA0716: a 1.5 Mb introgression on chromosome 2, a 58.4 Mb upper and a 1.4 Mb lower introgression on chromosome 3, a 260 Kb introgression on chromosome 7, a 3.5 Mb introgression on chromosome 8, and a 1.9 Mb introgression on chromosome 10. The locations of known introgressions are shown at top. Haplotype detection was repeated with four window sizes, shown below. Haplotypes with homology to LA0716 are colored, bar widths indicate the size of the detected introgression

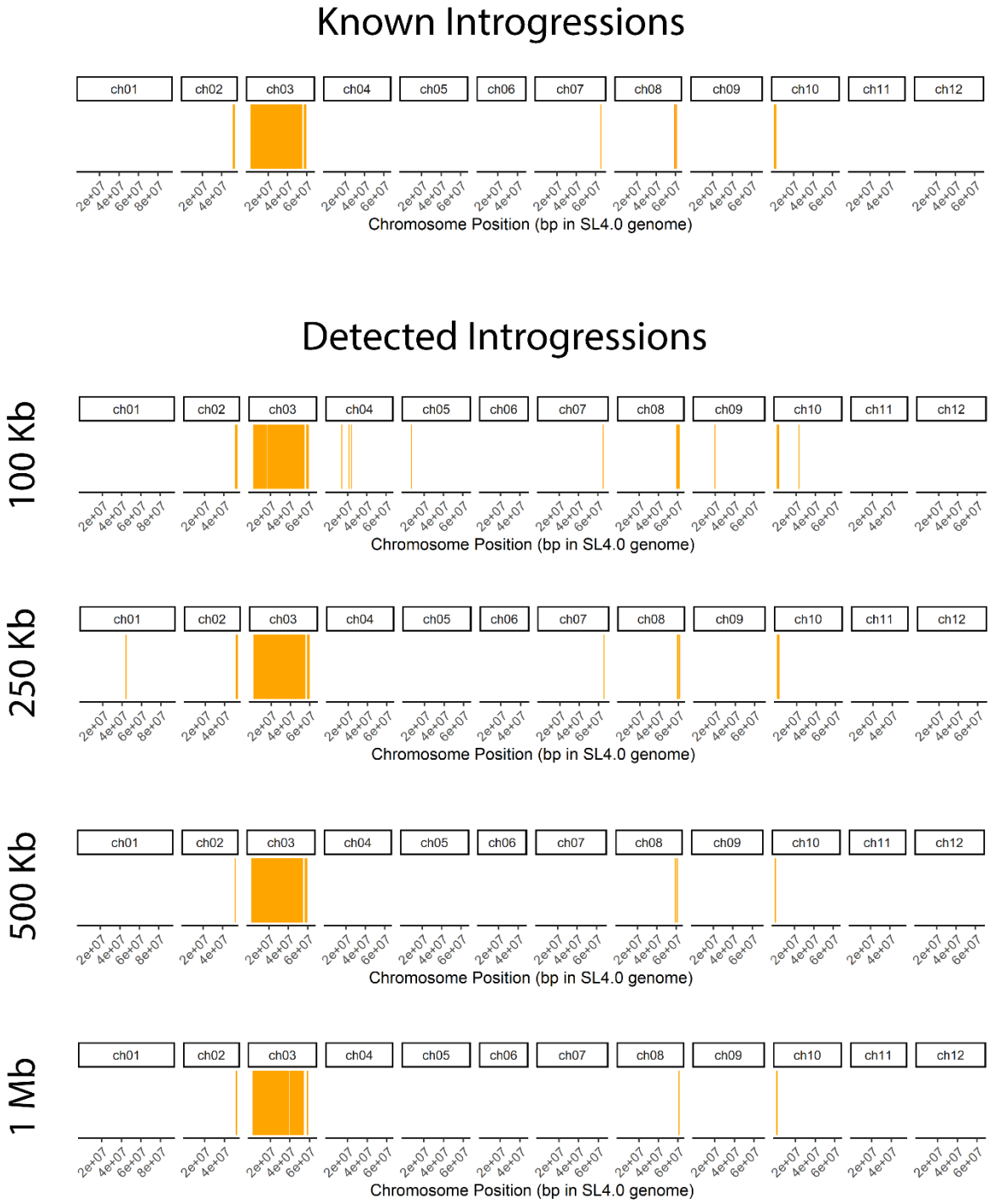


Fig. S4 Evidence for the chromosome 8 and 12 haplotype homology with *Devon Surprise* among early blight resistant breeding lines and other accessions

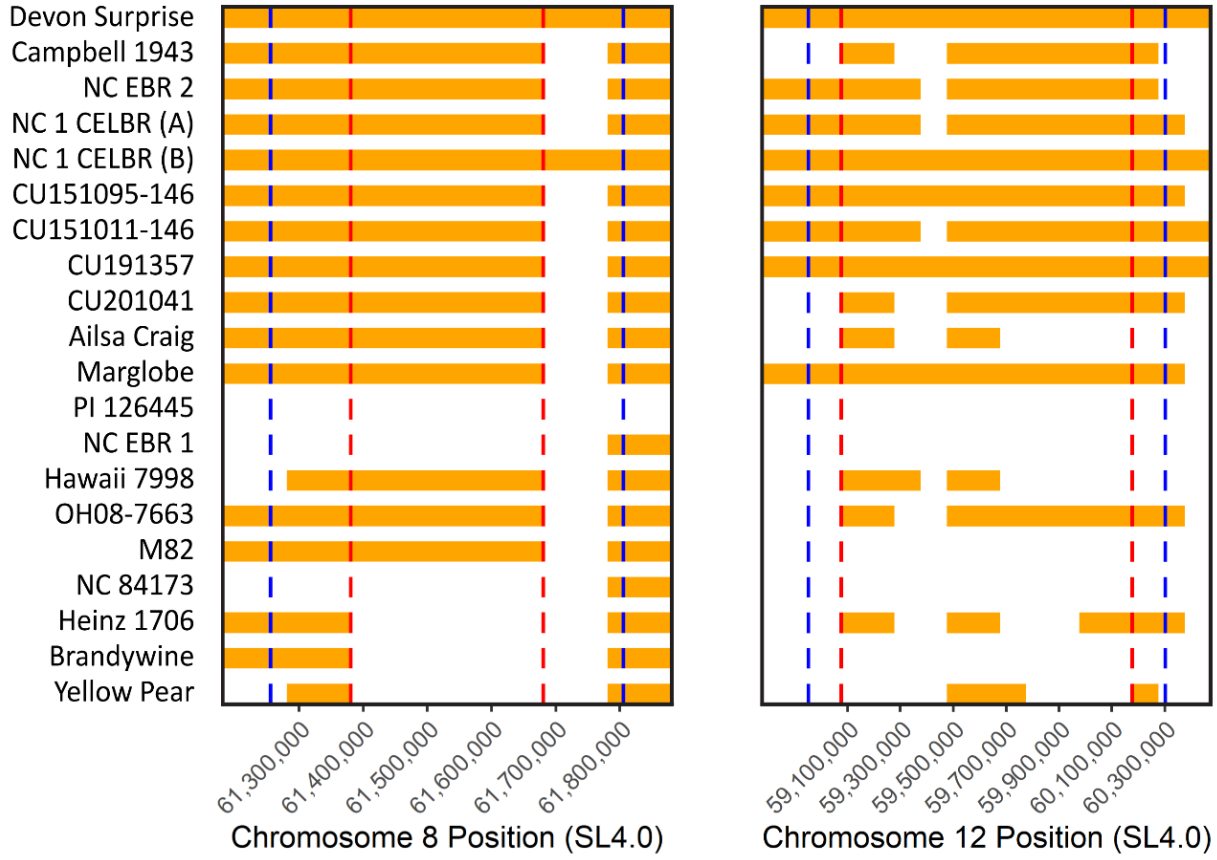


Fig. S5 A. Genetic windows that are shared between *Devon Surprise* and *Ailsa Craig* according to our data. Haplotypes with homology to *Devon Surprise* are colored orange. **B.** Coalescent species tree hierarchical clustering of the whole genome sequences for a relevant subset (to aid visualization only) of sequences, indicating strong support for the similarity between *Devon Surprise* and the two independent sequences of *Ailsa Craig*

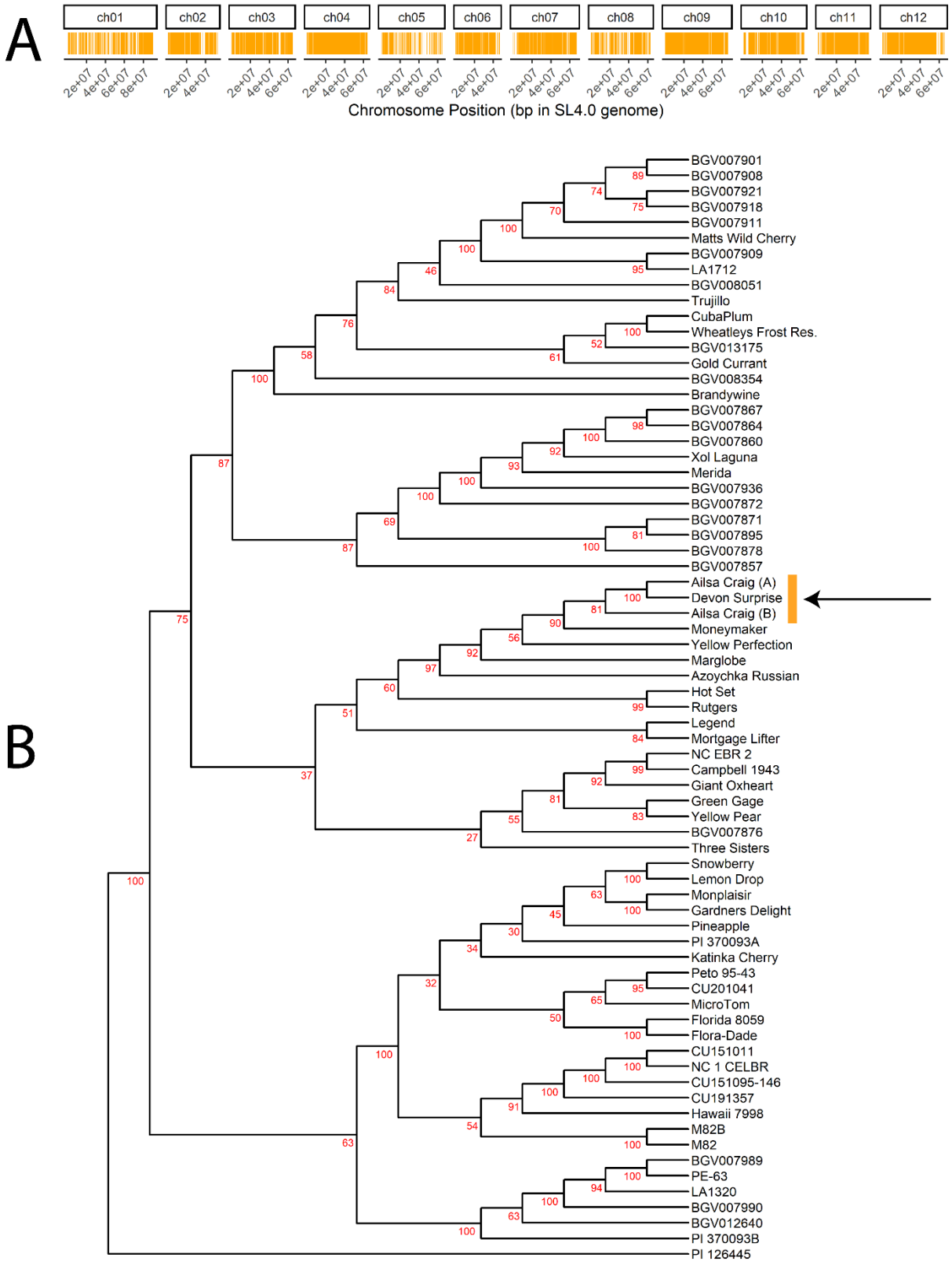


Fig. S6 Pairwise ancestry painting of the *EB-9* interval for the contrasts: **A.** *Devon Surprise vs.* NC 84173 **B.** *Devon Surprise vs.* Heinz 1706 and **C.** *Devon Surprise vs. Yellow Pear*

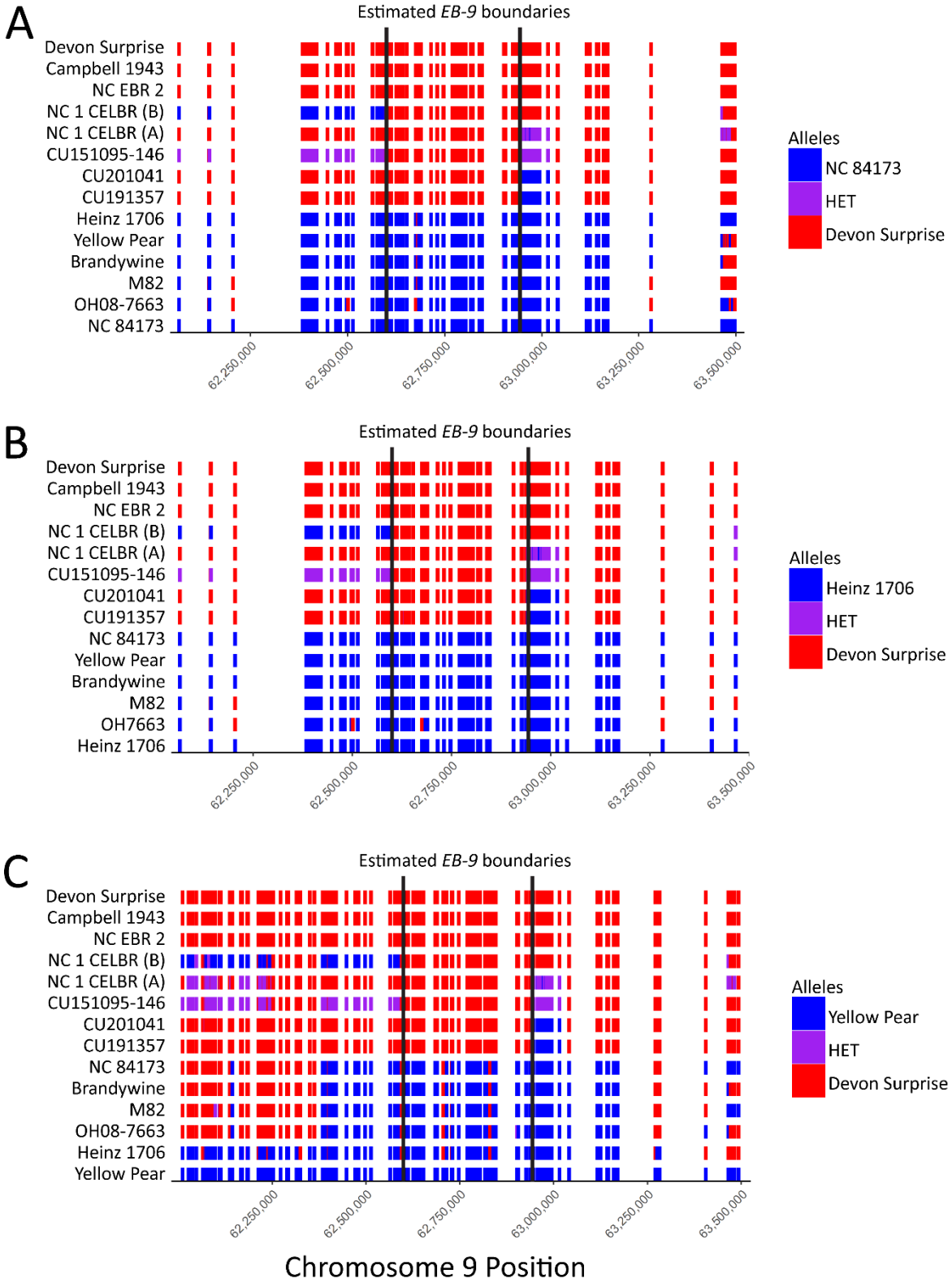


Fig. S7 Accessions predicted to have *EB-9* resistance based on similarity to *Devon Surprise*. Seven haplotype windows were tested, ranging in size from the red bars (smallest, 300 Kb) to the x-axis limits (largest, 750 Kb). High confidence accessions were grouped with *Devon Surprise* in 6-7 of these windows, while medium confidence accessions were included in 3-5 windows, and low-confidence accessions in 1-2 windows. Fine-scale sub-haplotype evidence is shown for a 250 Kb sliding window analysis with a 25 Kb step size. Windows clustering with *Devon Surprise* are colored. The most predictive marker from prior QTL mapping work is shown as a black bar (solcap_snp_sl_29188). NC 1 CELBR, *Gardner's Delight*, and PI 370093, were represented twice in the dataset

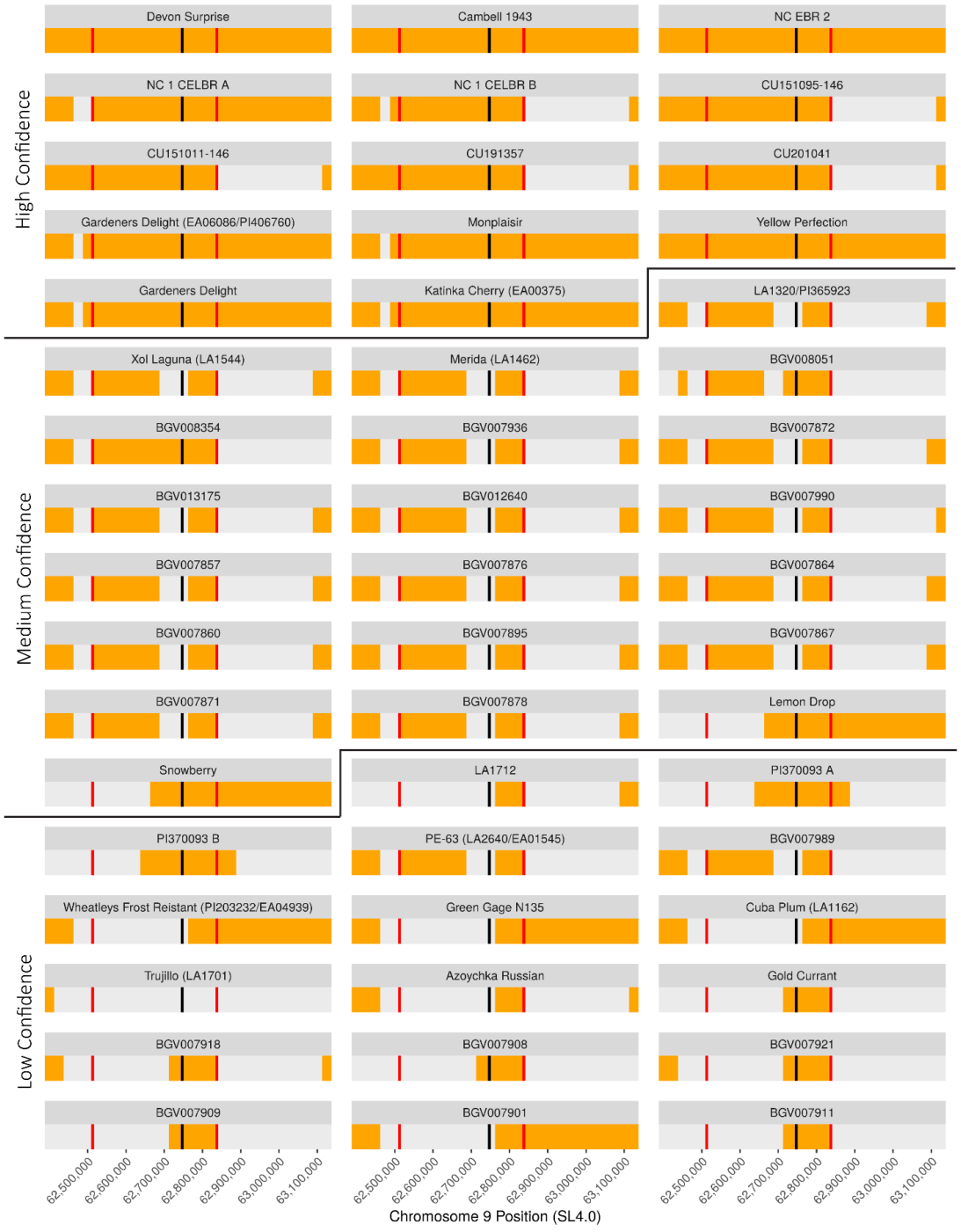


Fig. S8 Pairwise ancestry painting for the *EB*-5 interval contrasts: **A.** *Hawaii 7998 vs.* CU151095-146 **B.** *Hawaii 7998 vs.* Heinz 1706 **C.** *Hawaii 7998 vs. Yellow Pear*

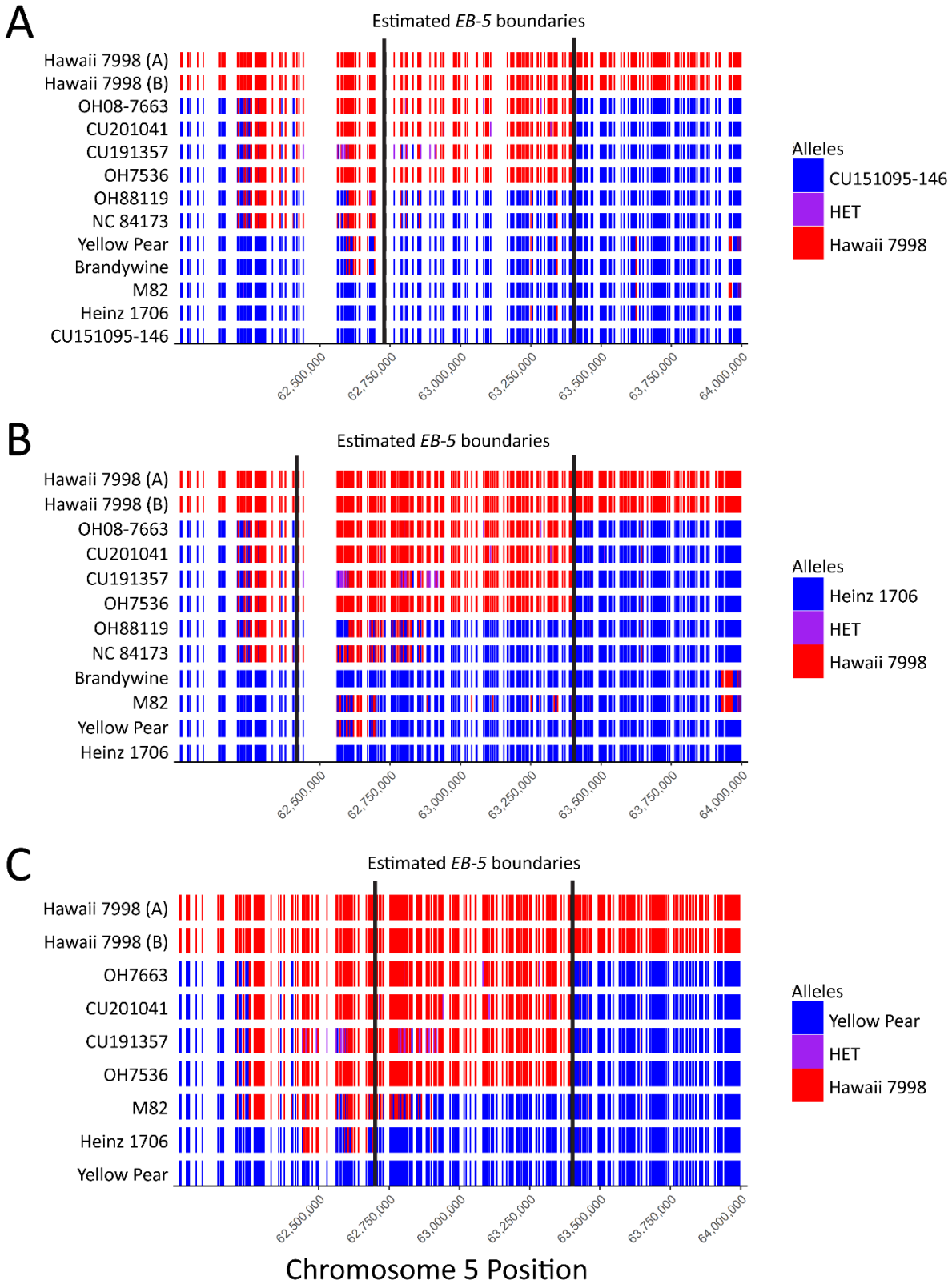


Fig. S9 Accessions predicted to have *EB-5* resistance based on haplotype similarity to *Hawaii 7998*. Seven haplotype windows were tested, ranging in size from the red bars (smallest, 375 Kb) to the x-axis limits (largest, 750 Kb). The black bar is the most predictive marker for *EB-5* from prior QTL mapping (solcap_snp_sl_231). High confidence accessions were grouped with *Hawaii 7998* in 6-7 of these windows, while medium confidence accessions were included in 3-5 windows, and low-confidence accessions in 1-2 windows. Fine-scale sub-haplotype evidence is shown for a 25 Kb sliding window analysis with a 25 Kb step size. Windows clustering with *Hawaii 7998* are colored. There are two sequences for *Hawaii 7998* in our dataset, the first supplied by Dr. David Francis (OSU)

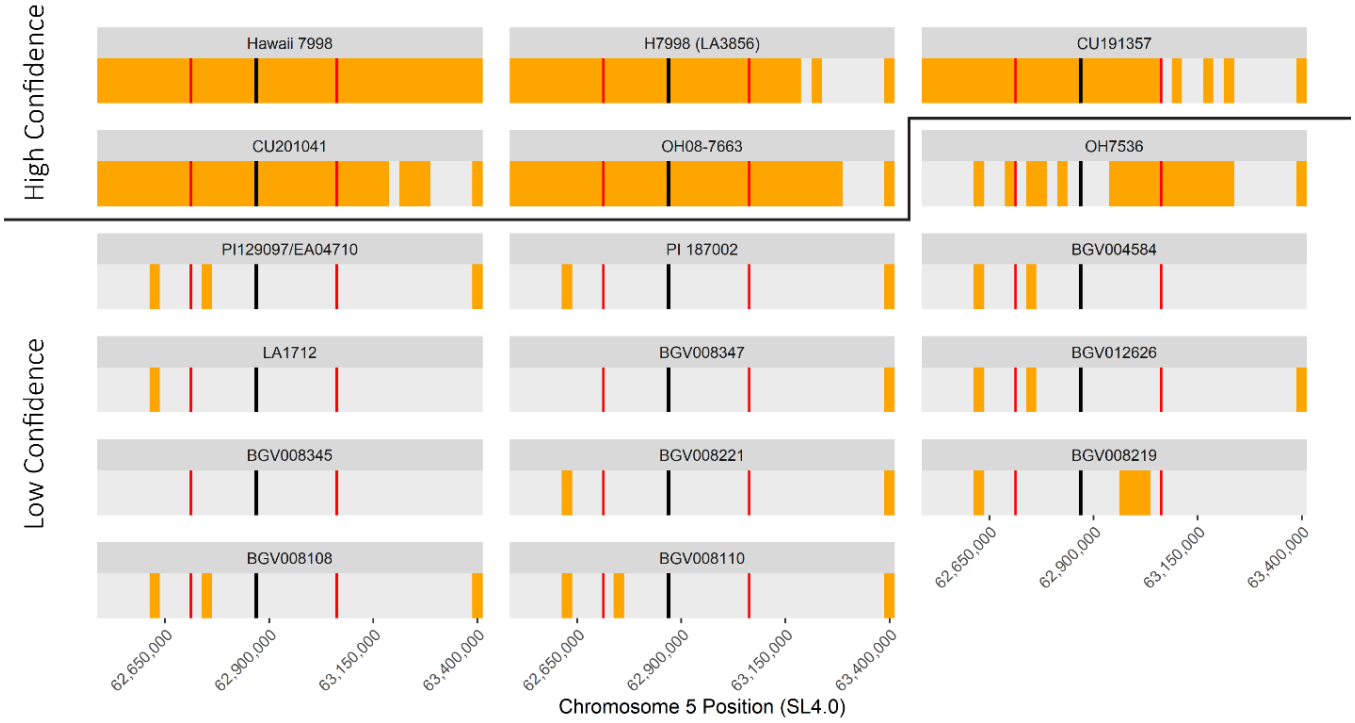


Fig. S10 Accessions with homology (orange color) for the chromosome 11 centromeric region with *Hawaii 7998* by our clustering analysis. Seven haplotype windows were tested, ranging in size from the red bars (smallest, 32.5 Mb) to the x-axis limits (largest, 38.1 Mb). High confidence accessions were grouped with *Hawaii 7998* in 6-7 of these windows, while medium confidence accessions were included in 3-5 windows, and low-confidence accessions in 1-2 windows. Fine-scale (sub-)haplotype evidence is shown for a 250 Kb sliding window analysis with a 100 Kb step size. There are two sequences for *Hawaii 7998* in our dataset, the first supplied by Dr. David Francis (OSU)

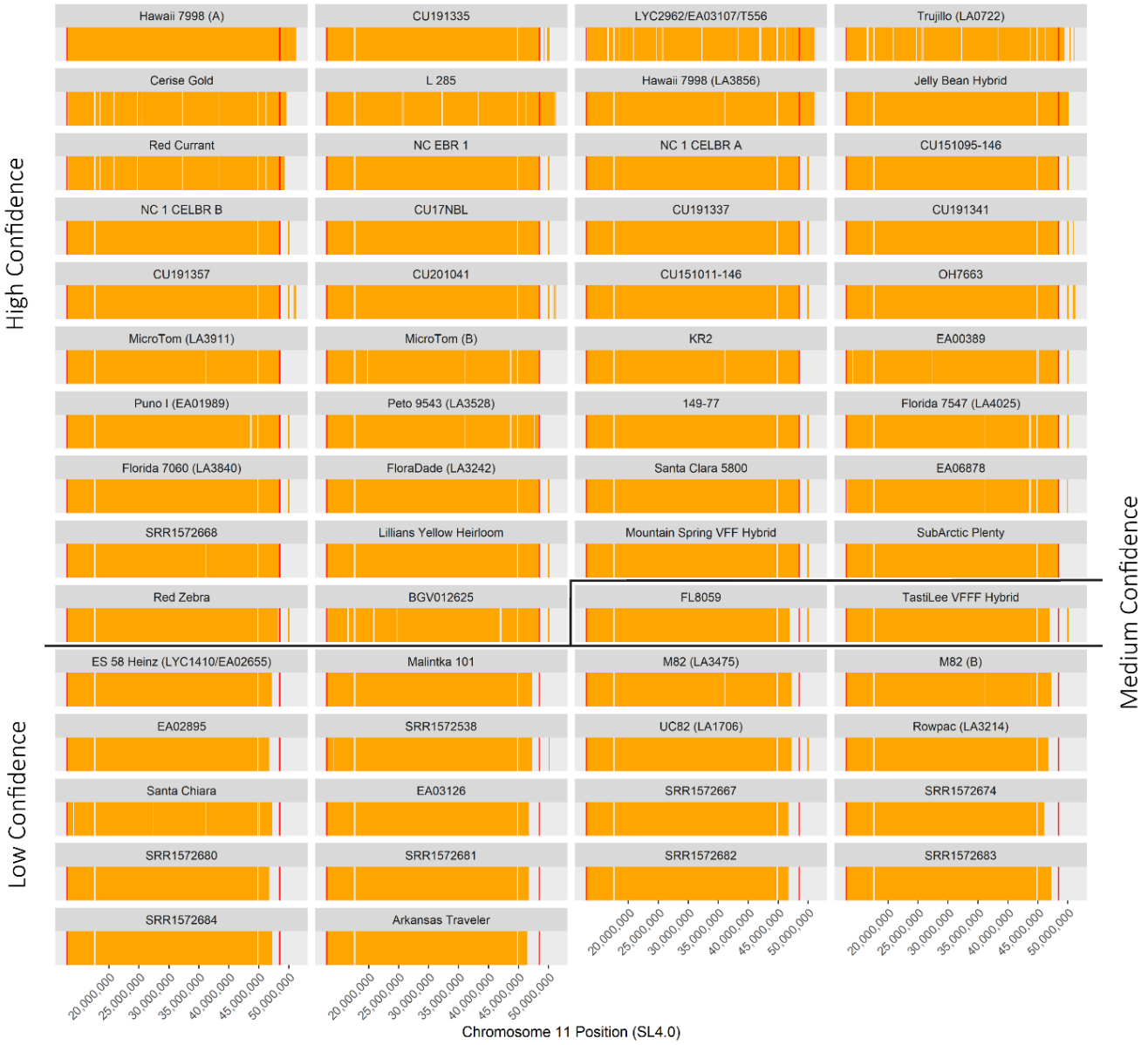


Fig. S11 Evidence for chromosome 11 homology (orange color) between *Hawaii 7998* and *S. pimpinellifolium* accessions in the dataset using a 250 Kb window size and 100 Kb step size