Supplementary Figure 11: Representative roots of resequenced carrot accessions. Samples include eastern (C1, C2, C5) and western (C9, C12, C13) cultivated (*D. carota* subsp. *sativus*) carrot phenotypes, and examples of wild (*D. carota* subsp. *carota*) carrots. Details for each sample are reported in Supplementary Table 16.

C1

C2

C5

C9

C12

C13

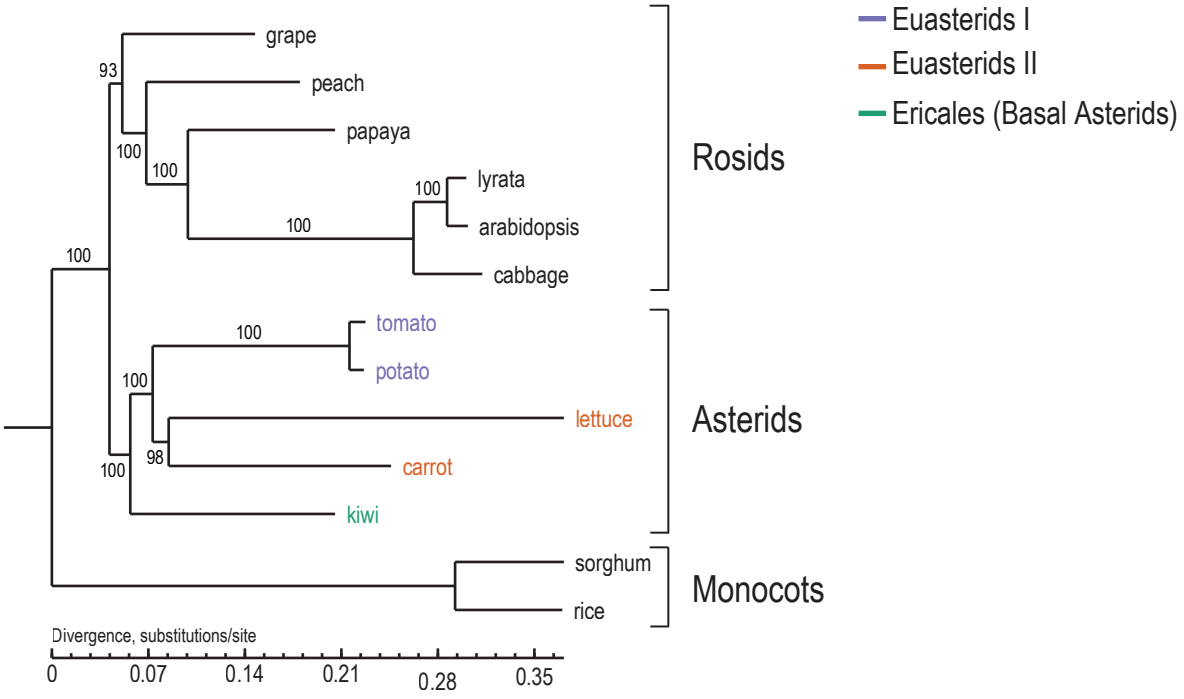
W3

W5

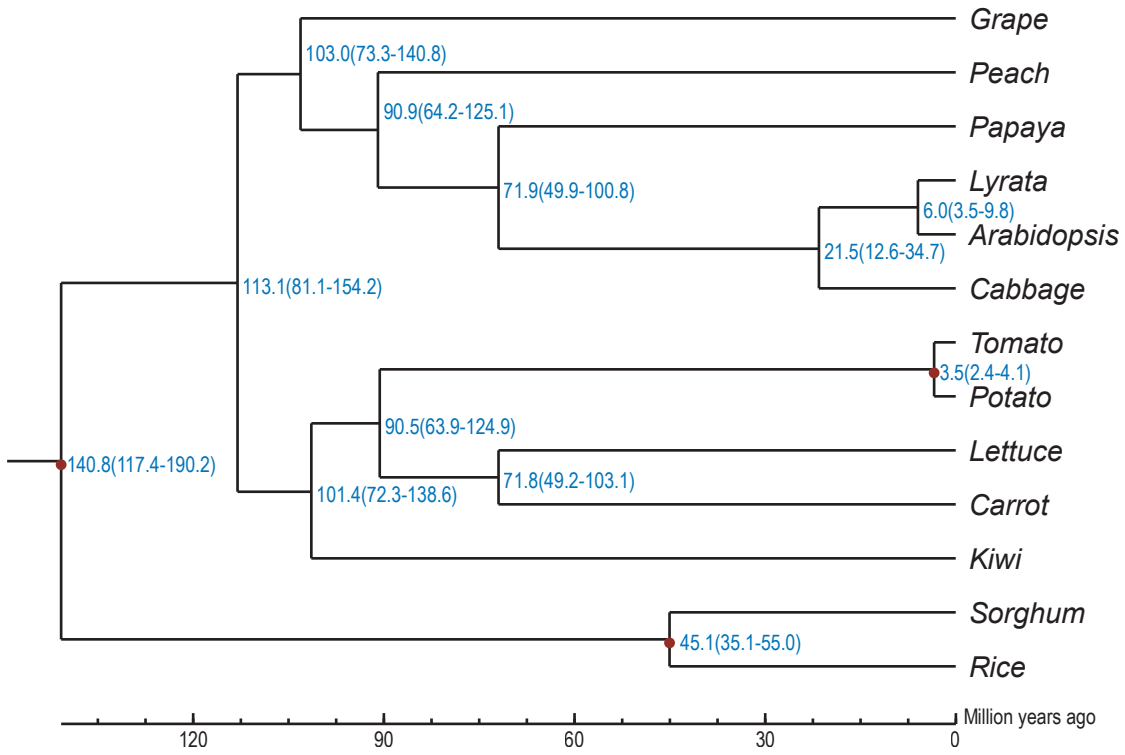
W7

Supplementary Figure 12: Comparative gene analysis. A: Maximum likelihood tree constructed with 312 single copy orthologous genes. Bootstrap values are shown at nodes. The scale is amino acid substitutions per site. B: Time divergence estimation of 13 dicot and monocot plants. The red dots represent divergence time estimates used to calibrate the analysis.

A

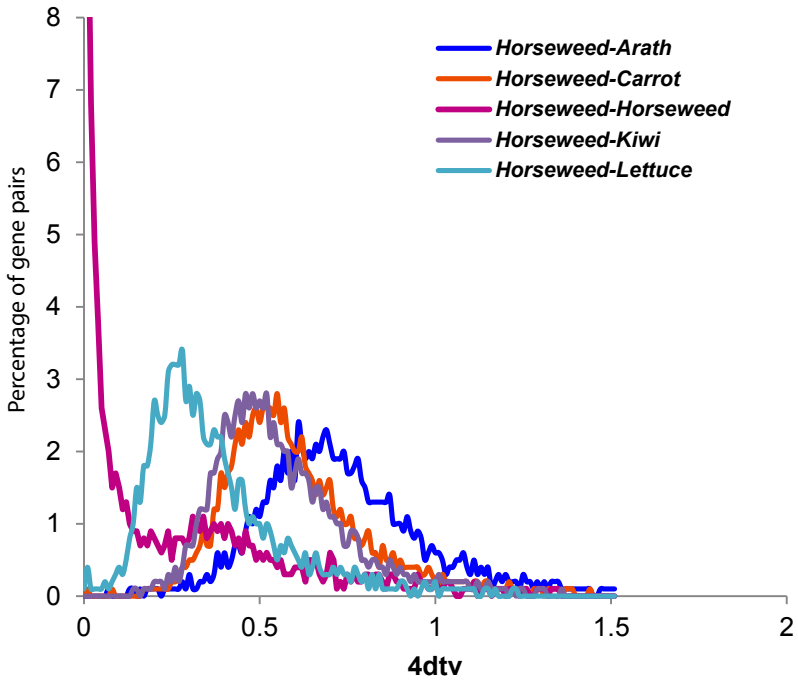


B

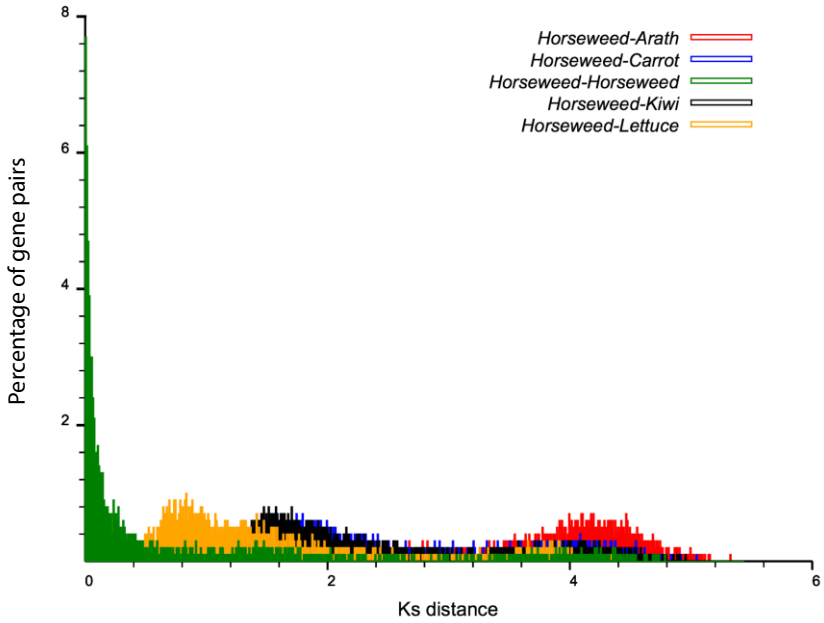


Supplementary Figure 13: Age distribution of 4DTv and Ks analyses. 4DTv analysis (Panel A) and Ks analysis (Panel B) are presented for genes from the Horseweed (*Conyza canadensis*), carrot, *A. thaliana*, kiwi and lettuce genomes. X-axis indicates 4DTv values and Ks distance, respectively; Y-axis indicates percentage of ortholog/paralog gene pairs.

A

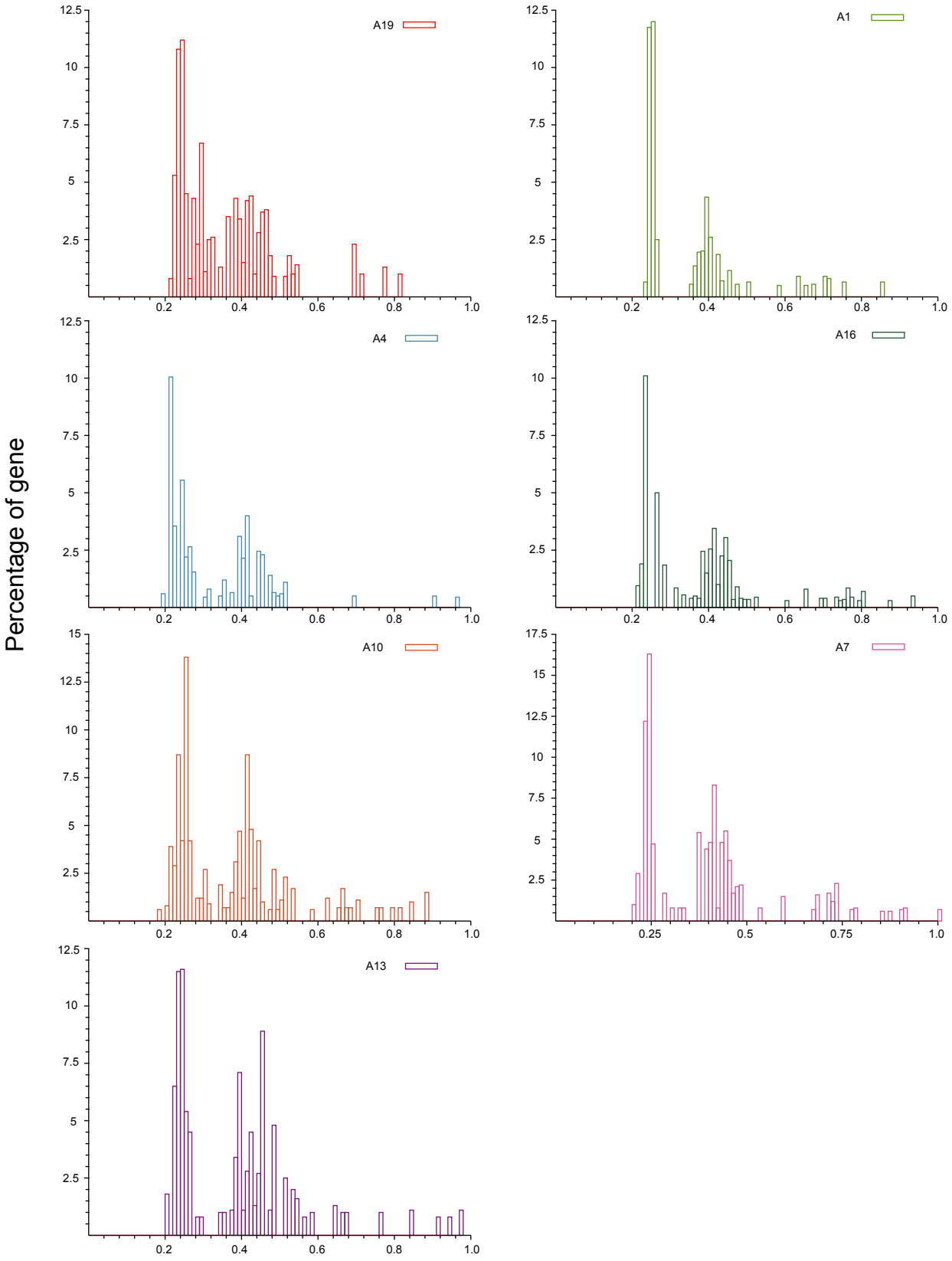


B



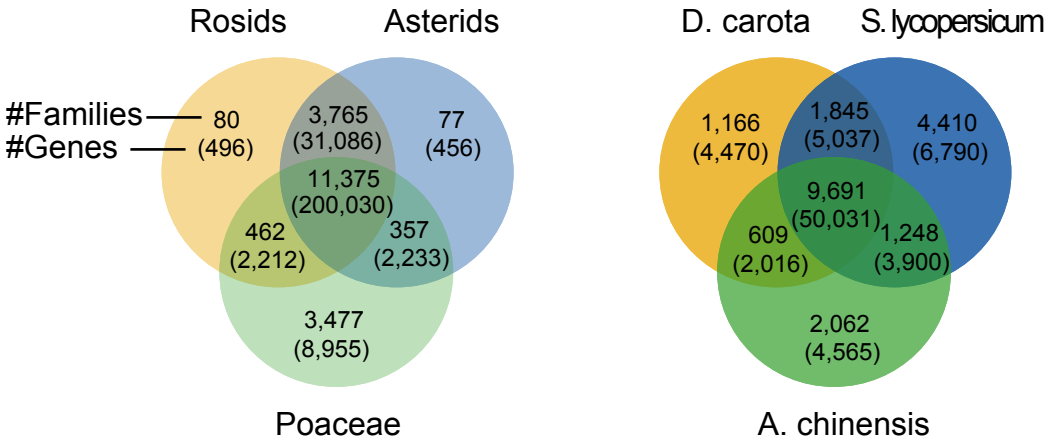
Nature Genetics: doi:10.1038/ng.3565

Supplementary Figure 14: Age distribution of 4DTv for carrot gene paralogs descending from the seven ancestral core eudicot chromosomes. A1 to A19 are the ancestral protochromosomes.

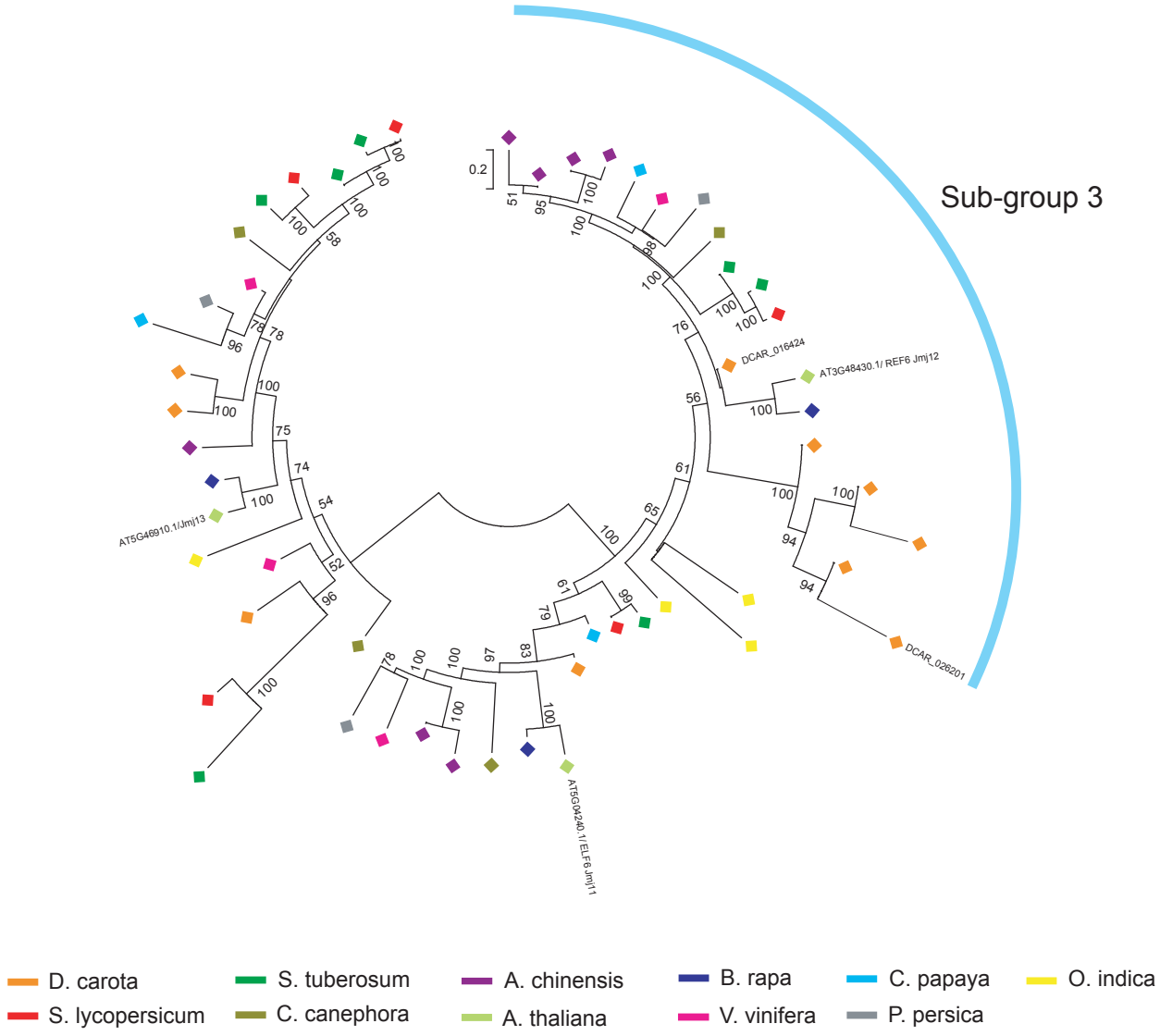


Nature Genetics: doi:10.1038/ng.3565

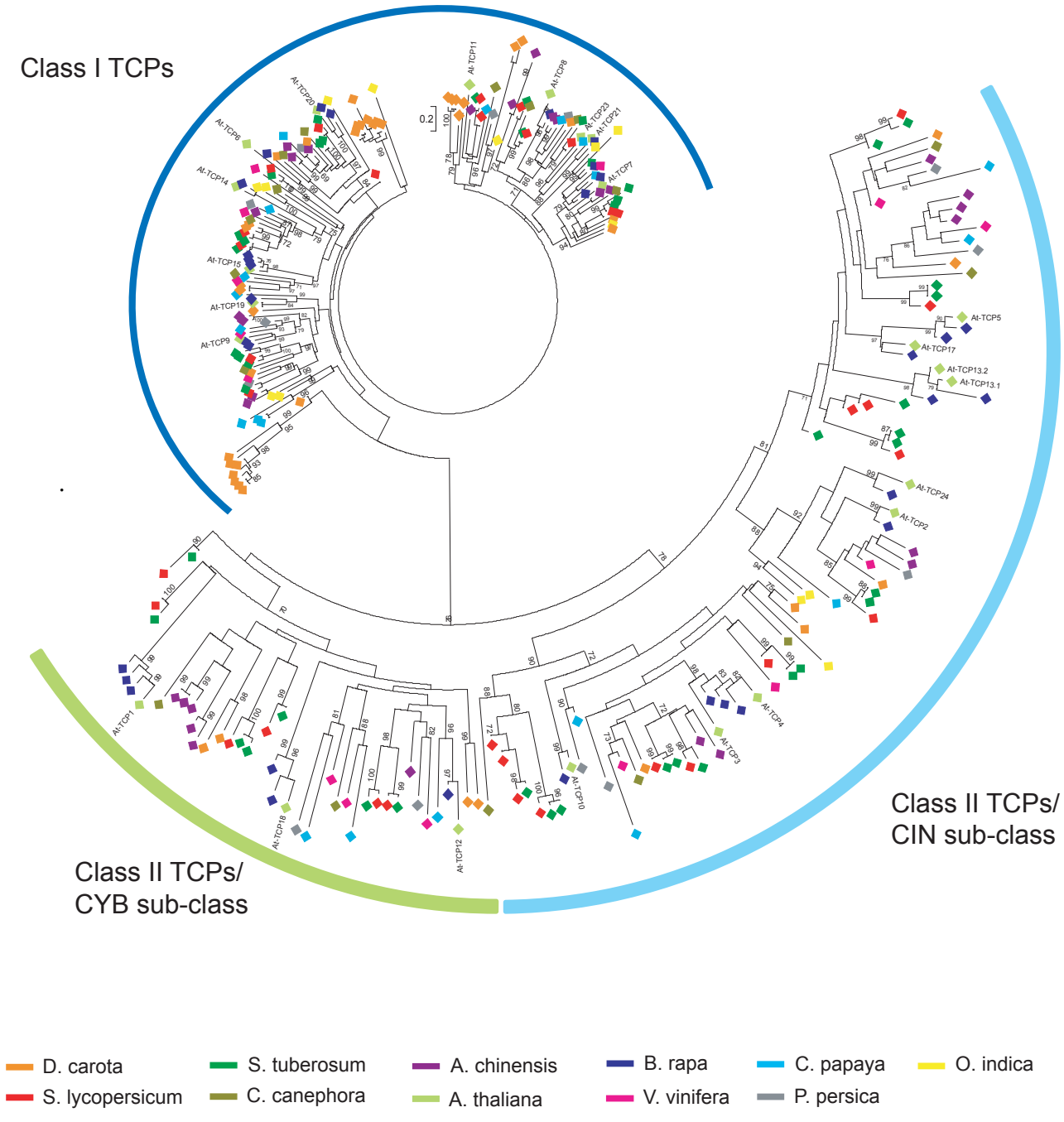
Supplementary Figure 15: Gene family cluster analysis. The Poaceae group includes orthologous genes from *O. sativa* and *S. bicolor* genomes, representatives of the Monocot clade. The Rosids group includes orthologous genes from *A. thaliana*, *A. lyrata*, *B. rapa*, *C. papaya*, *P. persica* and *V. vinifera* genomes. The Asterids group includes orthologous genes from *D. carota*, *S. lycopersicum*, *S. tuberosum* and *A. chinensis* genomes.



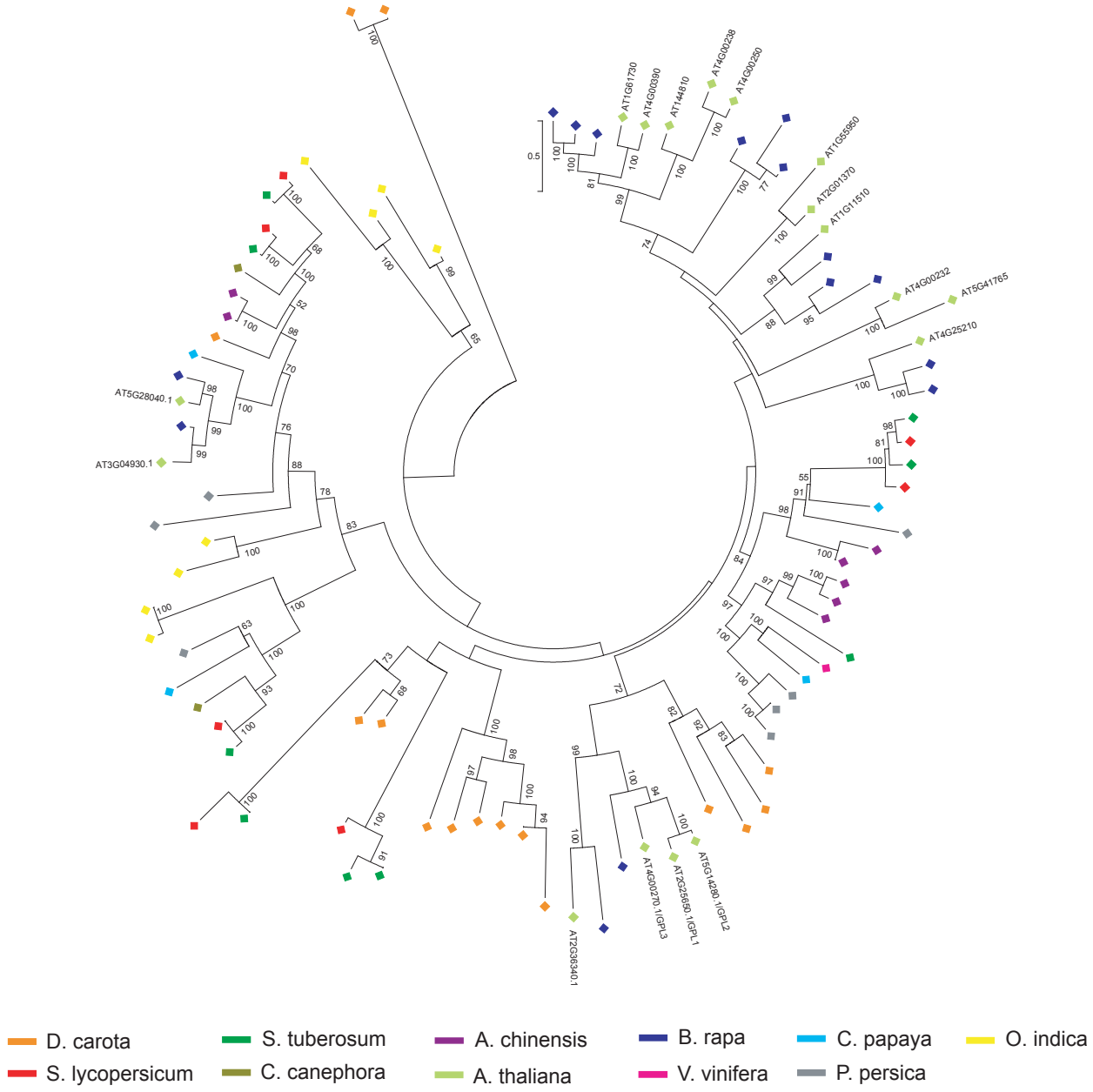
Supplementary Figure 16: Phylogenetic analysis of the JMJD2 subfamily of JmjC transcription factors. The phylogenetic tree was constructed using the Neighbor-Joining method and a bootstrap test was performed with 1,000 iterations. Bootstrap values over 50% are shown. The scale bar (0.2) shows the number of amino acid substitutions per site. Previously characterized JmjC genes from A. thaliana are labeled with AT prefix[109]. JMJD2 sub-group 3 was expanded in carrot and it includes the functionally characterized Arabidopsis *REF6*. Carrot DCAR_016424 and DCAR_026201 were retained from the eudicot gamma whole genome triplication.



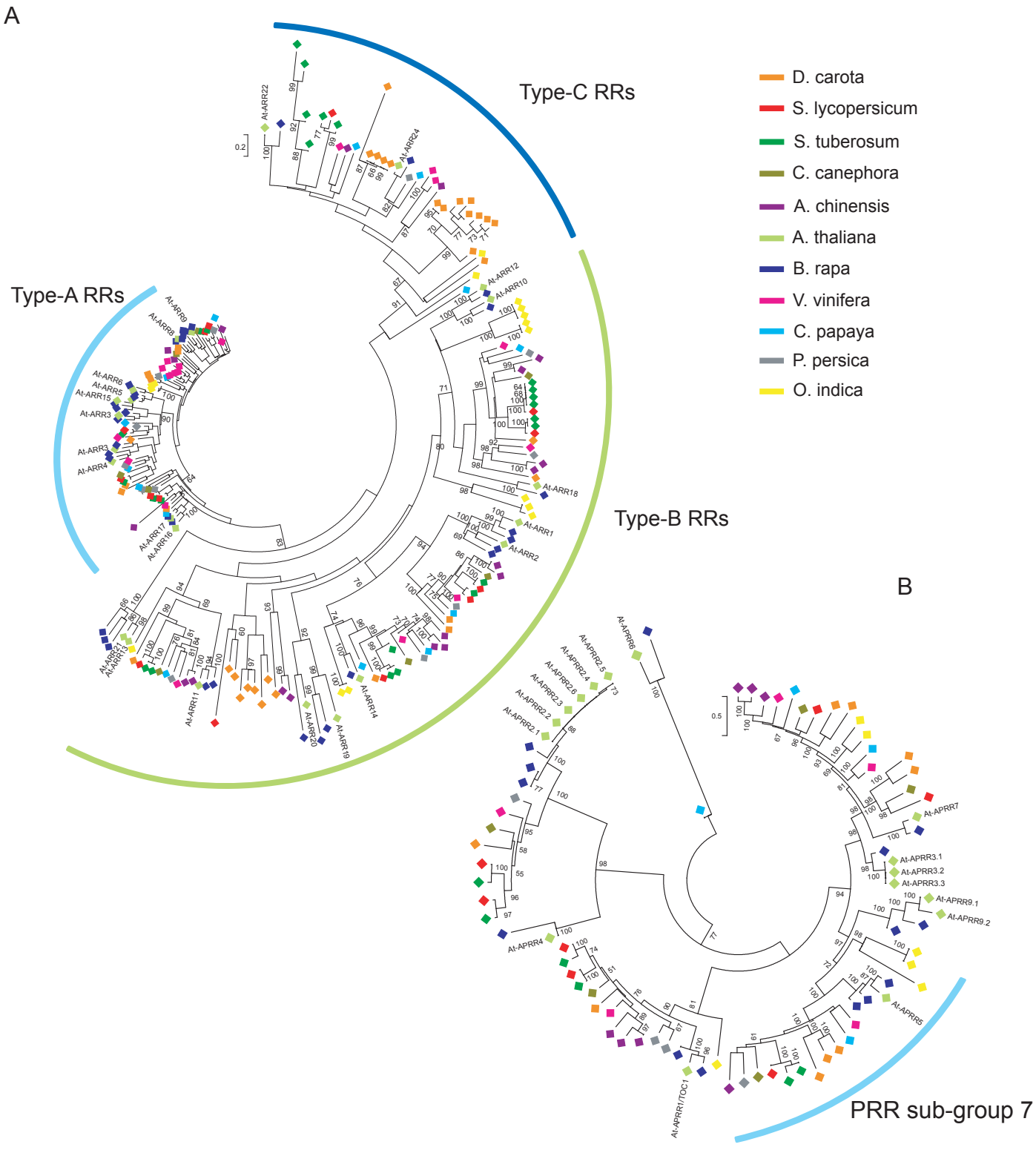
Supplementary Figure 17: Phylogenetic analysis of the TCP transcription factors. The phylogenetic tree was constructed using the Neighbor-Joining method and a bootstrap test was performed with 1,000 iterations. Bootstrap values over 50% are shown. The scale bar (0.2) shows the number of amino acid substitutions per site. Previously characterized Arabidopsis TCPs are labeled with AT prefix[125]. TCP sub-group 11 (highlighted with thick black branches) has expanded in carrot and it includes a functionally characterized Arabidopsis *At-TCP11*.



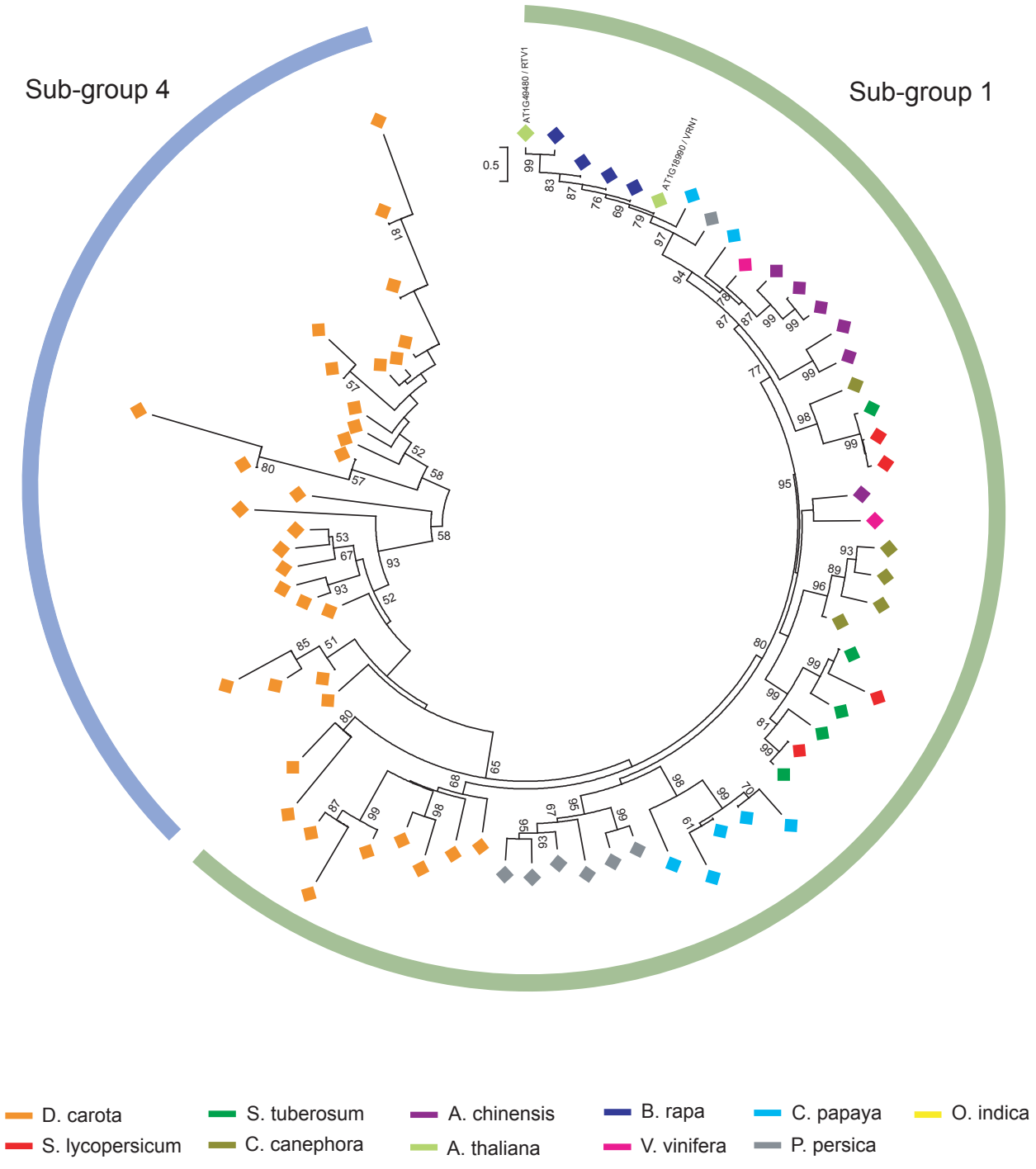
Supplementary Figure 18: Phylogenetic analysis of the GeBP transcription factors. The phylogenetic tree was constructed using the Neighbor-Joining method and a bootstrap test was performed with 1,000 iterations. Bootstrap values over 50% are shown. The scale bar (0.5) shows the number of amino acid substitutions per site. Annotated Arabidopsis GeBP genes from the plant transcription factor database are labeled with AT prefix. GeBP sub-group 1 has expanded in carrot and it includes four genes that are homologous to the functionally characterized Arabidopsis *GPL1-2-3* (ref. 126).



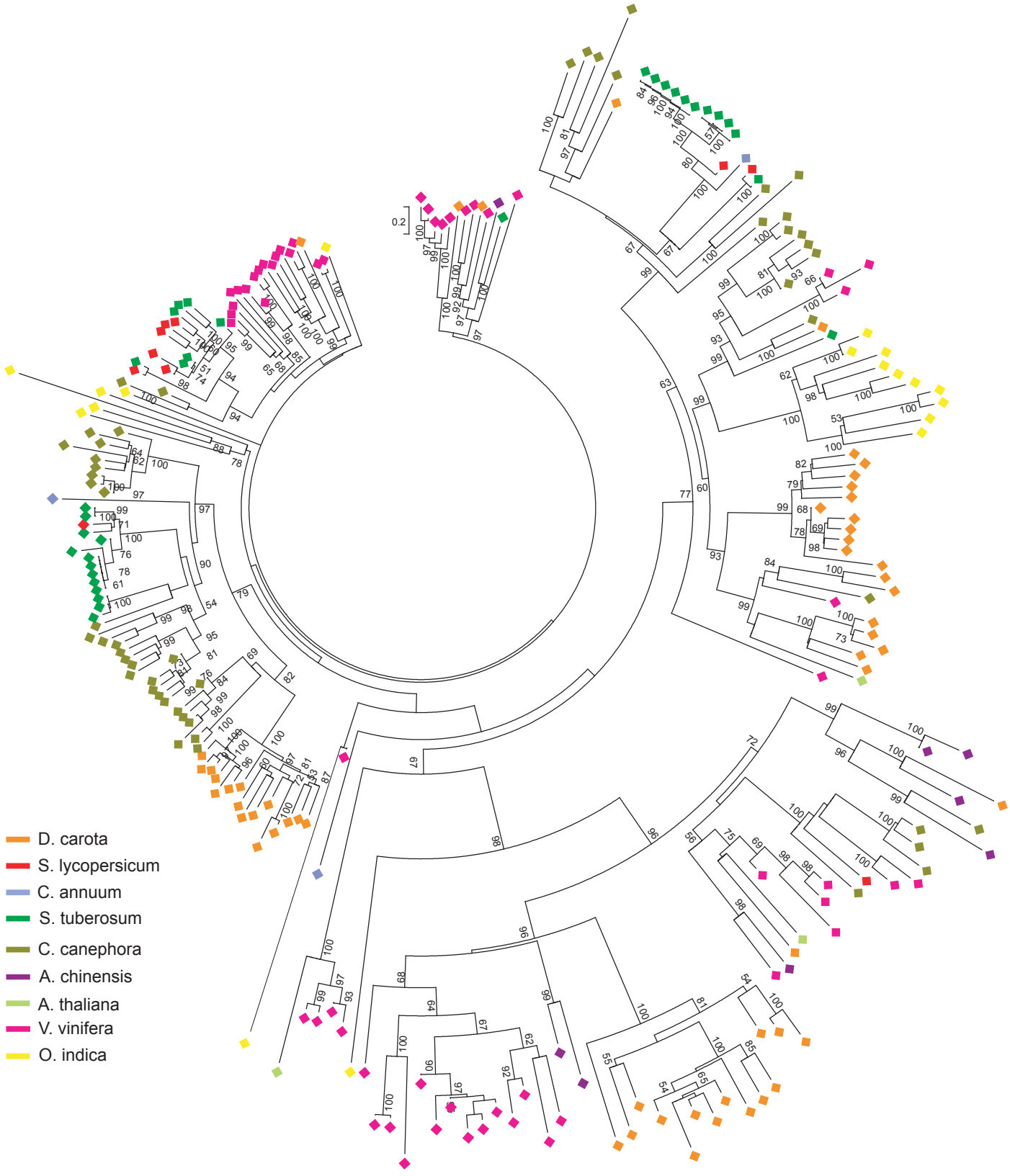
Supplementary Figure 19: Phylogenetic analysis of the response regulator genes (RR). The phylogenetic tree was constructed using the Neighbor-Joining method and a bootstrap test was performed with 1,000 iterations. Bootstrap values over 50% are shown. The scale bars (0.2 and 0.5) show the number of amino acid substitutions per site. A: phylogenetic tree of type A-B-C regulators. B: phylogenetic tree of PRR regulators. Previously characterized RR genes from *A. thaliana* are labeled with AT prefix[129].



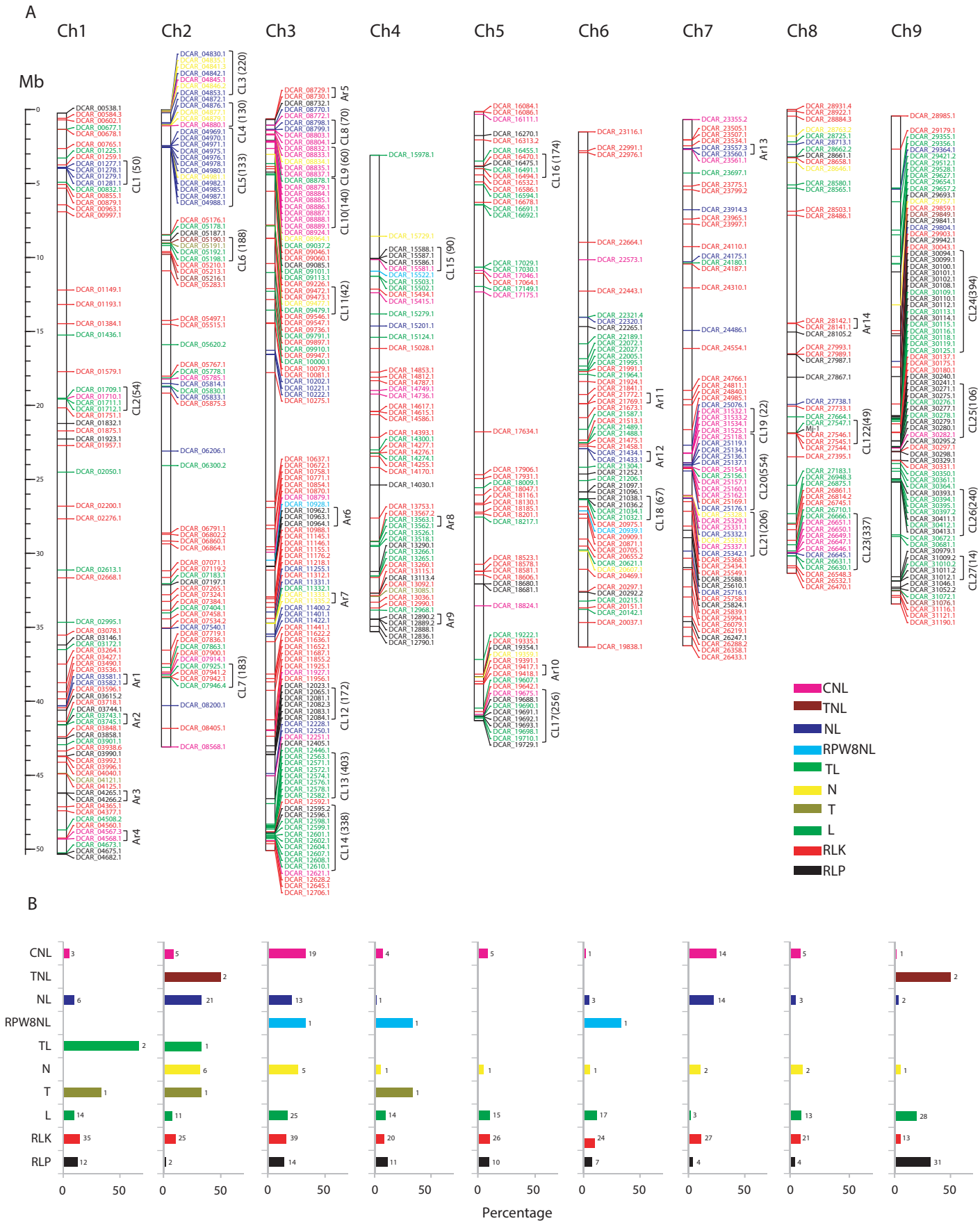
Supplementary Figure 20: Phylogenetic analysis of the REM sub-groups 1 and 4 of the B3-domain transcription factors. The phylogenetic tree was constructed using the Neighbor-Joining method and a bootstrap test was performed with 1,000 iterations. Bootstrap values over 50% are shown. The scale bar (0.5) shows the number of amino acid substitutions per site. Previously characterized REM genes from *A. thaliana* are labeled with AT prefix[132]. Sub-group 1 was expanded in carrot and it includes the functionally characterized Arabidopsis *VRN1*. Sub-group 4 is carrot specific and it is the sub-group most closely related to sub-group1.



Supplementary Figure 21: Phylogenetic analysis of the CNL R gene class. The phylogenetic tree was constructed using the Neighbor-Joining method and a bootstrap test was performed with 1,000 iterations. Bootstrap values over 50% are shown. The scale bar (0.2) shows the number of amino acid substitutions per site.

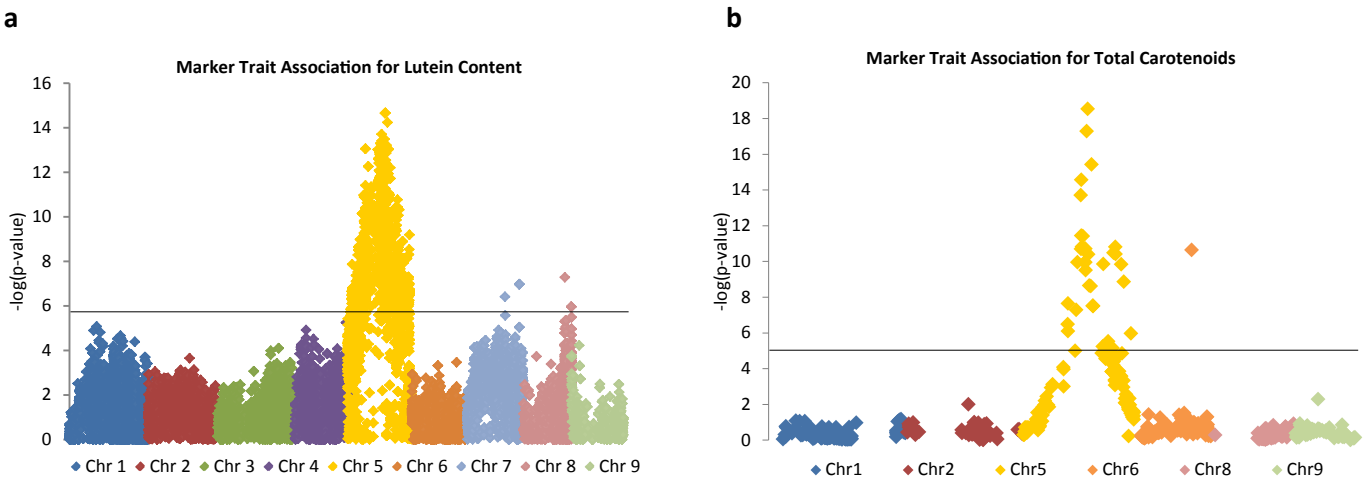


Supplementary Figure 22: Carrot R genes. A: Distribution of candidate R genes in the nine carrot chromosomes. Brackets on the right side of each chromosome indicate the set of R genes organized in clusters (CL) or arrays (Ar), and the length in kb of the genomic region spanning each cluster (in parentheses). B: Summary distribution of each R gene class along the nine carrot chromosomes. Bars represent the percentage of genes for each family located on each chromosome. Numbers above the bars indicate the number of genes.

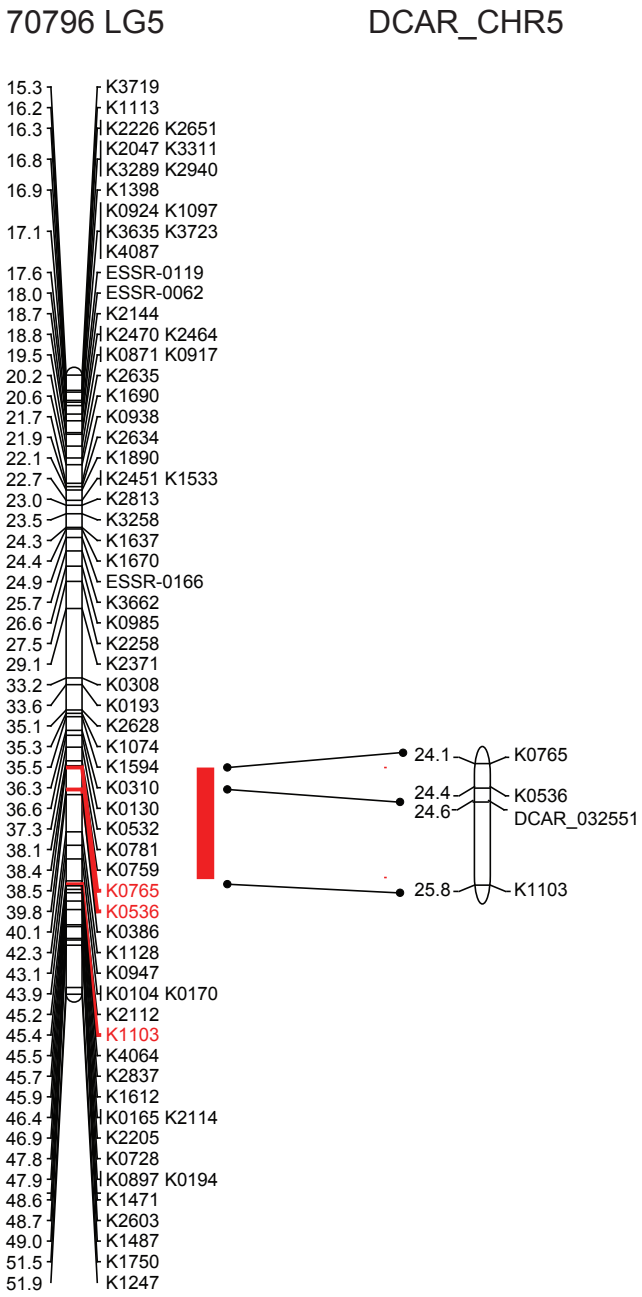


Nature Genetics: doi:10.1038/ng.3565

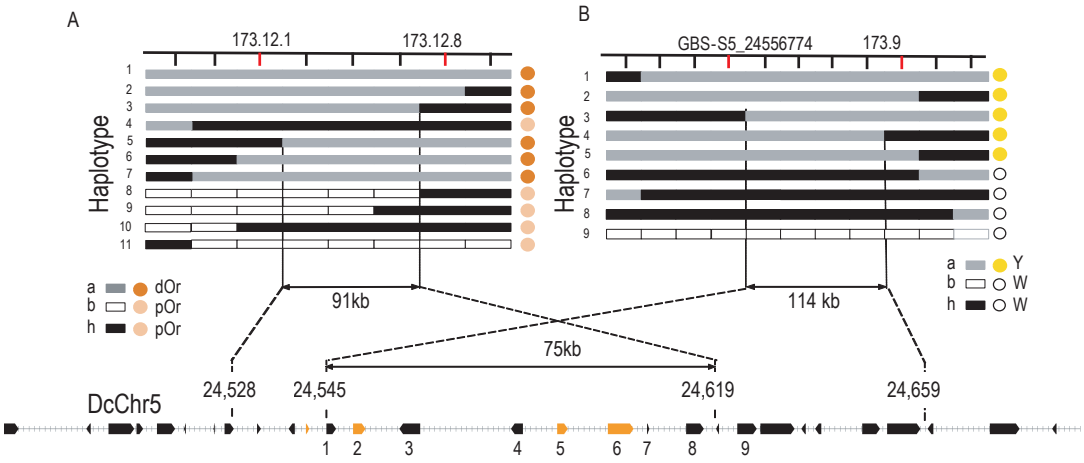
Supplementary Figure 23: Manhattan plots for marker-trait associations using GLM analysis. a) Lutein content in population 97837. b) Total carotenoids in population 70796. The gray dotted line indicates significance cut-off after using a Bonferroni correction.



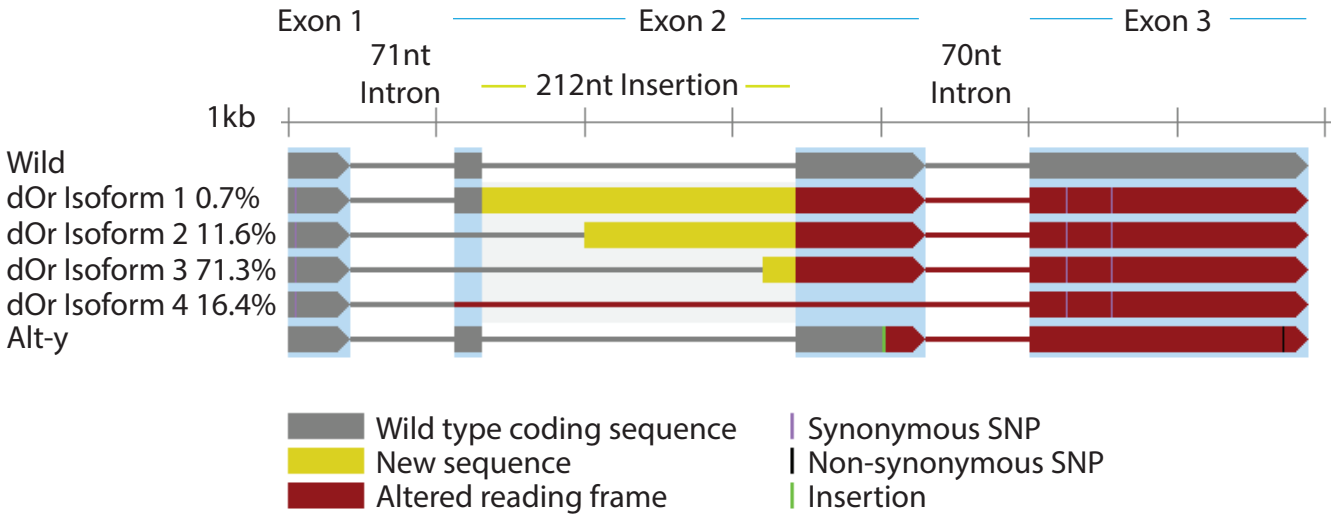
Supplementary Figure 24: : Population 70796 carotenoid QTL mapping results. Map on the left corresponds to LG5. Markers highlighted in red indicate mapping positions (cM) flanking the QTL detected for total carotenoid,content. Map on the right indicates the corresponding physical position in carrot chromosome 5 with markers flanking the QTL interval. DCAR_032551 corresponds to the candidate gene controlling the *Y* locus. To optimize the visualization of the QTL region, the start (from position 0 to 15.3 cM) and the end (from position 51 to 63 cM) portions of LG5 are not shown.



Supplementary Figure 25: Fine mapping of the carrot *Y* locus. A: Haplotype blocks associated with dark orange (dOr) and pale orange (pOr) root phenotypes. B: Haplotype blocks associated with yellow (Y) and White (W) root phenotypes. Overlapping haplotype blocks associated with dOr and Y phenotypes and spanning 75 kb were identified across the two populations as the most significant location harboring the gene controlling the *Y* locus.



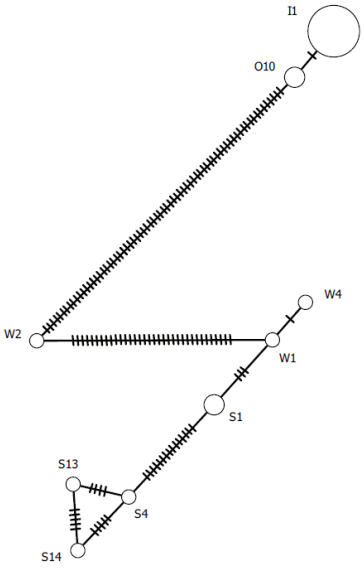
Supplementary Figure 26: Schematic representation of the polymorphisms detected in DCAR_032551. “Wild” indicates the wild type allele without the insertions. dOrF1 to dOrF4 indicate all the isoforms identified in the *Y* mutant which includes a 212 nt insertion in the second exon. The relative percent for each isoform is reported. Alt-y represents the *Y* allele identified in two resequenced samples, C1 and I2, that contains a 1nt insertion in the second exon.



Supplementary Figure 27: Haplotype network analysis. Panel B includes SNP data from all *D. carota* wild and cultivated accessions in the haplotype block associated with the *Y* locus. Panels A and C include SNPs from the regions covering 8 kb upstream and 8 kb downstream from the start and the end of the haplotype block, respectively.

A

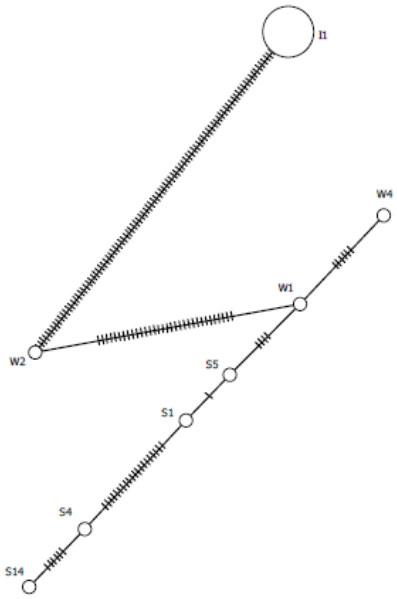
24,550,000 - 24,622,997
(72,997nt)
Y haplotype + ~8kb upstream



Number of segregating sites: 105
Number of parsimony -informative sites: 104
I1 includes I3, I4, O2, O3, O4, O5, O7, O8, O9, O12, O13 and O14; S1 includes S5; O10 includes O11

B

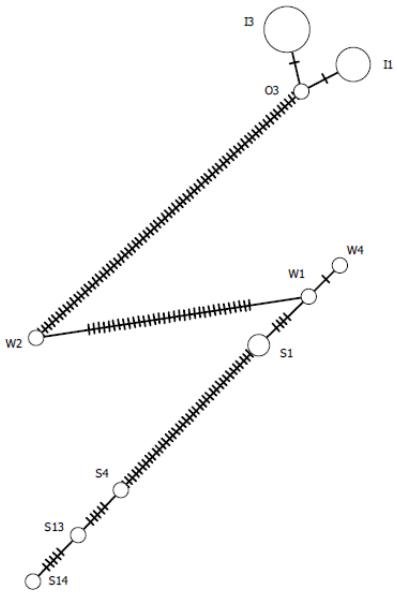
24,557,852 - 24,622,997
(65,145nt)
Y haplotype



Number of segregating sites:140
Number of parsimony -informative sites: 137
I1 includes I3, I4, O2, O3, O4, O5, O7, O8, O9, O10, O11, O12, O13 and O14

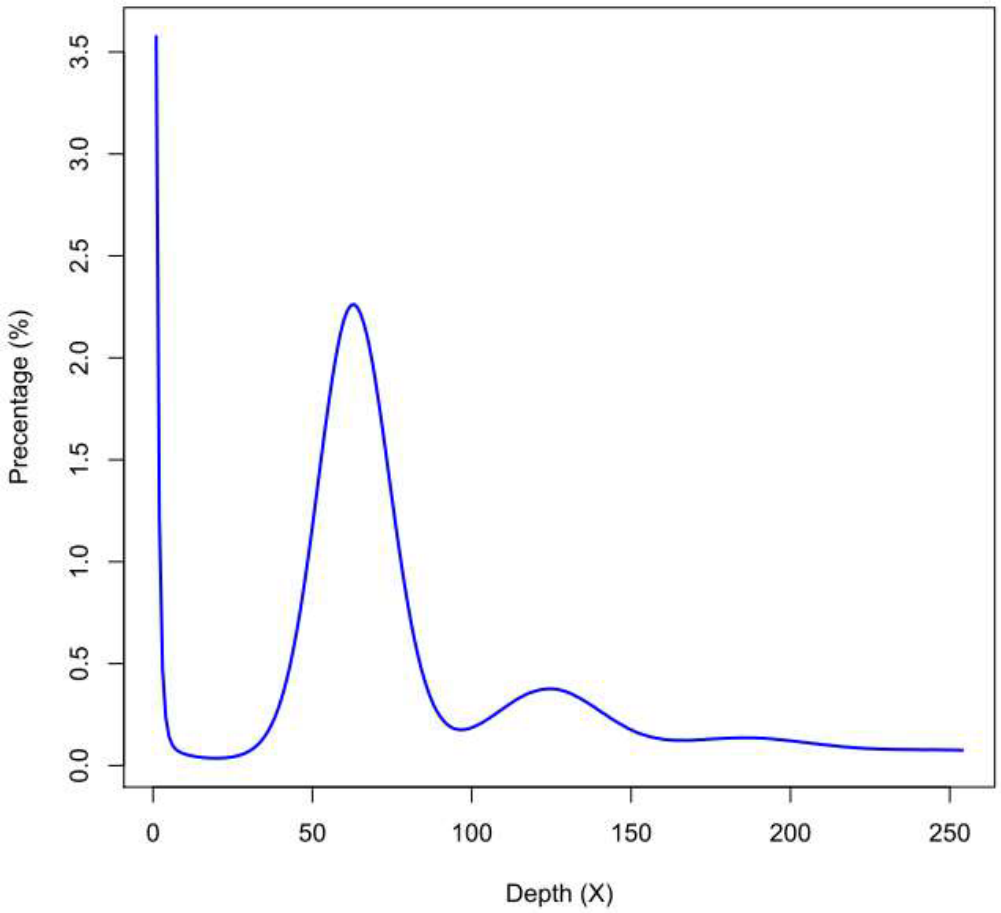
C

24,557,852 - 24,631,000
(73,148nt)
Y haplotype + ~8 kb downstream

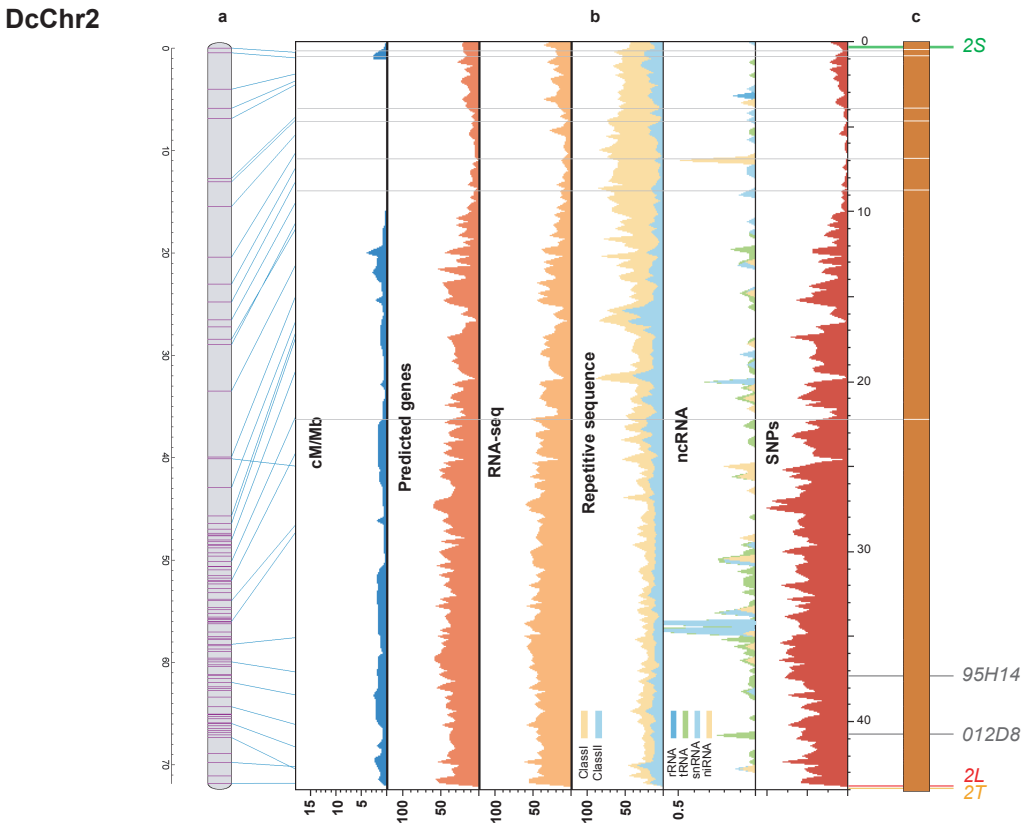


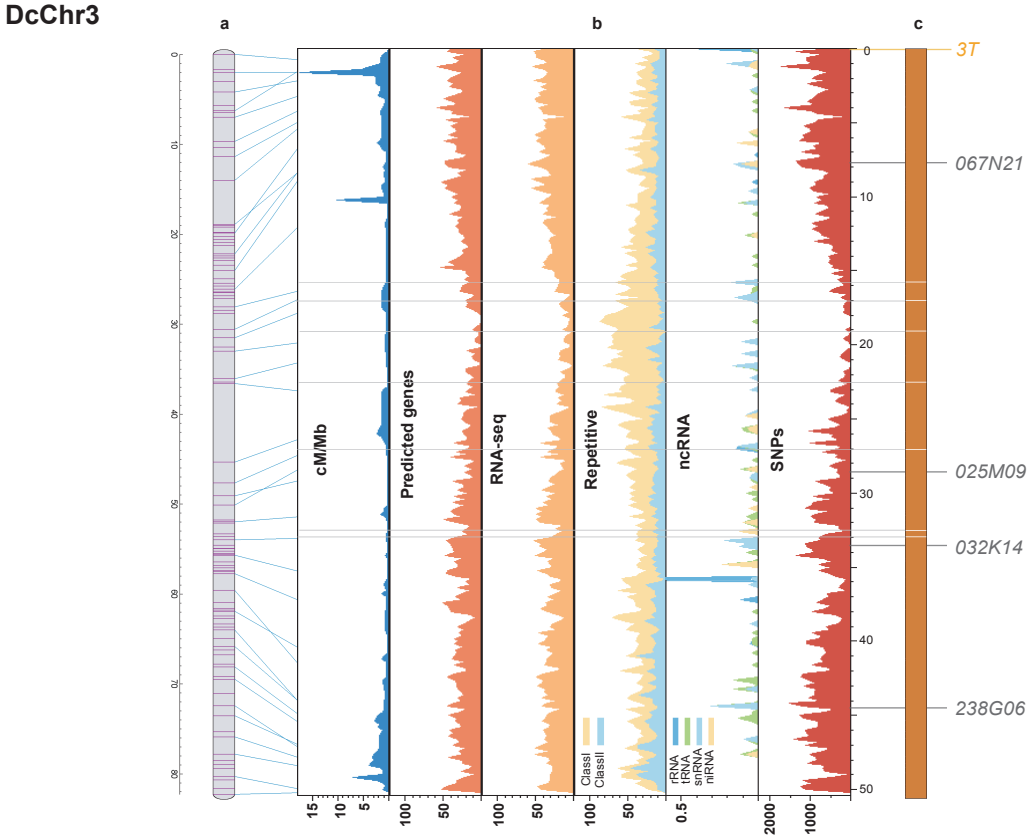
Number of segregating sites: 131
Number of parsimony -informative sites: 130
I3 includes O5, O7, O8, O10, O11, O12, O13 and O14; I1 includes I4, O2, O4 and O9; S1 includes S5

Supplementary Figure 28: 17-mer estimated carrot genome size. The x-axis is depth (X). The y-axis is the proportion of sequences which represents the frequency at that depth, divided by the total frequency of all the depths.

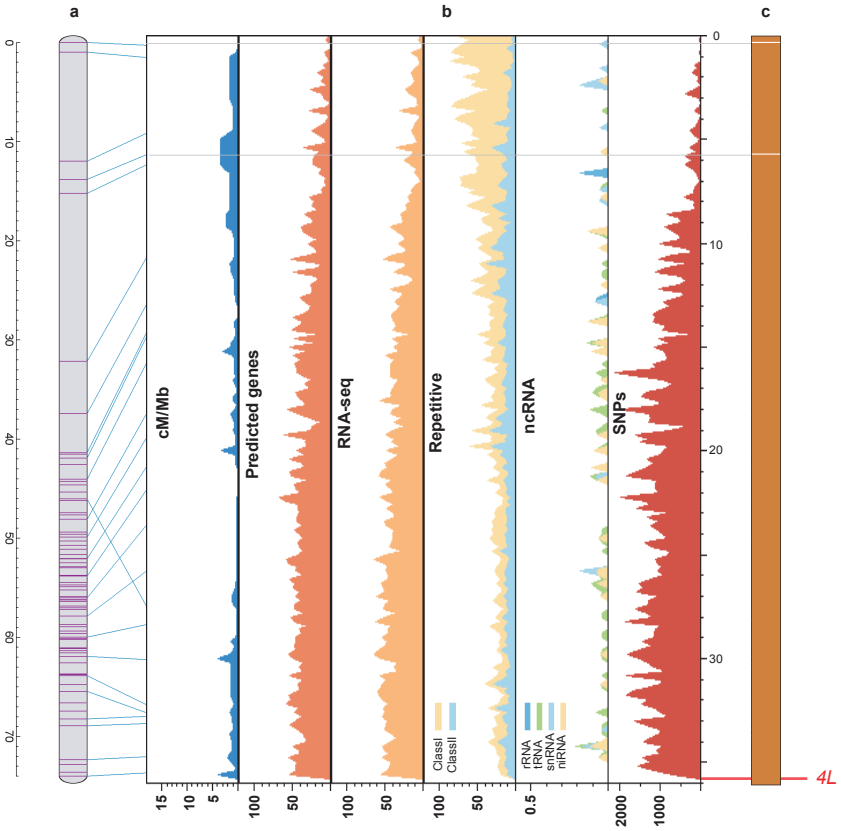


Supplementary Figure 29. Multi-dimensional topography of carrot chromosomes. Chromosome 1 is illustrated in Fig. 1. **a**) Carrot integrated linkage map. Vertical bar to the left indicates the genetic distance in cM. Lines to the right connect a subset of markers to the assembled pseudomolecule. **b**) From the left to the right: linkage map distance (cM/Mb), predicted genes (% nucleotide per 200 kb), RNA-Seq from 20 different sequencing libraries (% nucleotide per 200 kb), class I and class II repetitive sequences (% nucleotide per 200 kb), non-coding RNA (% nucleotide per 200 kb), SNPs detected comparing resequencing data from 35 different genotypes (number of SNPs per 100 kb). Horizontal gray lines represent gaps in the pseudomolecule. **c**) DNA pseudomolecules. Gaps between superscaffolds are indicated by gray horizontal lines. Location of BAC probes hybridized to pachytene chromosomes are identified by horizontal green and red lines and labeled on the right. Horizontal orange lines indicate location of the telomeric repetitive sequence (T). Dark gray lines at the right indicate the location of BAC clone end sequences previously used to anchor the genetic map to carrot chromosomes[10].

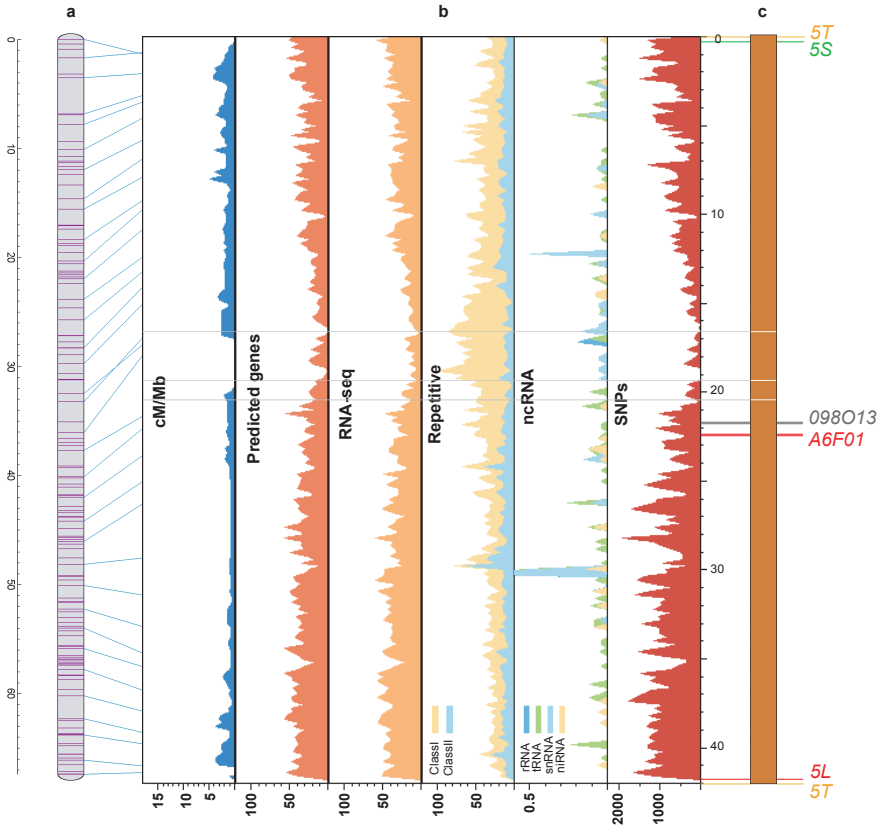

DcChr2
a
b
c
cM/Mb
Predicted genes
RNA-seq
Repetitive sequence
ncRNA
SNPs
ClassI
ClassII
rRNA
tRNA
snRNA
niRNA
2S
95H14
012D8
2L
2T


DcChr3
a
b
c
cM/Mb
Predicted genes
RNA-seq
Repetitive
ncRNA
SNPs
ClassI
ClassII
rRNA
tRNA
snRNA
niRNA
3T
067N21
025M09
032K14
238G06

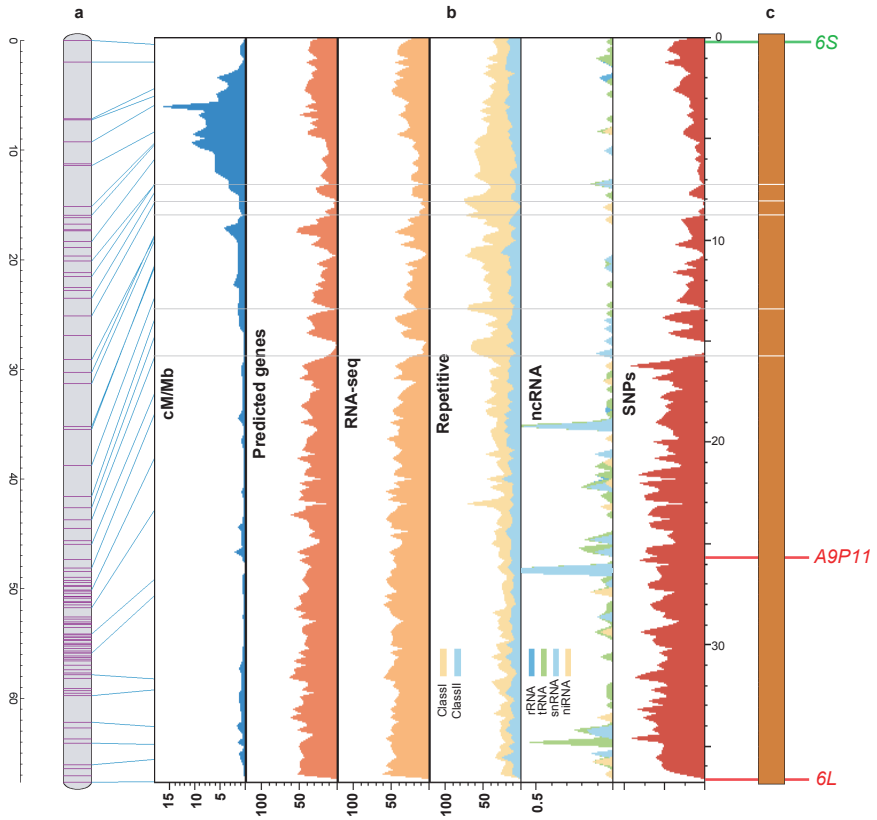
DcChr4



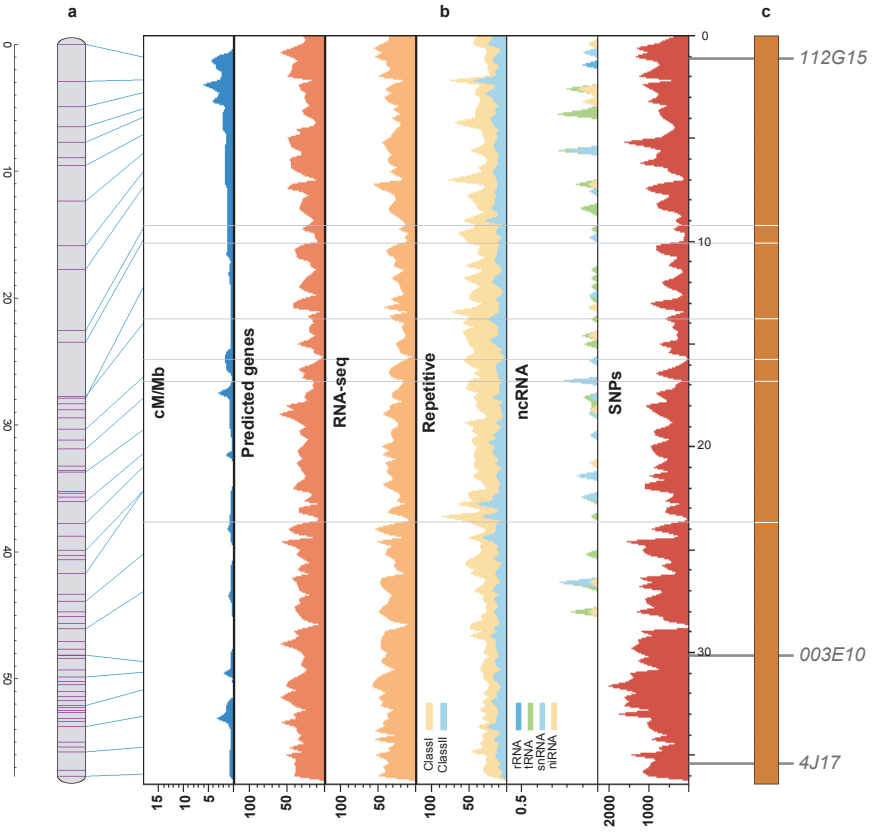
DcChr5



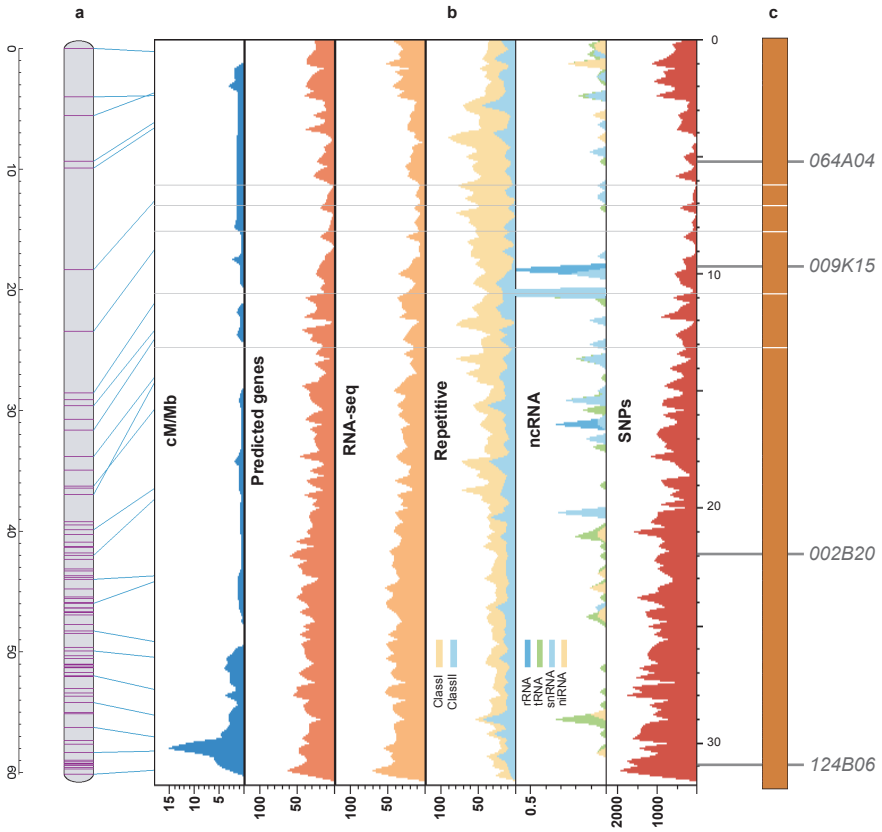
DcChr6



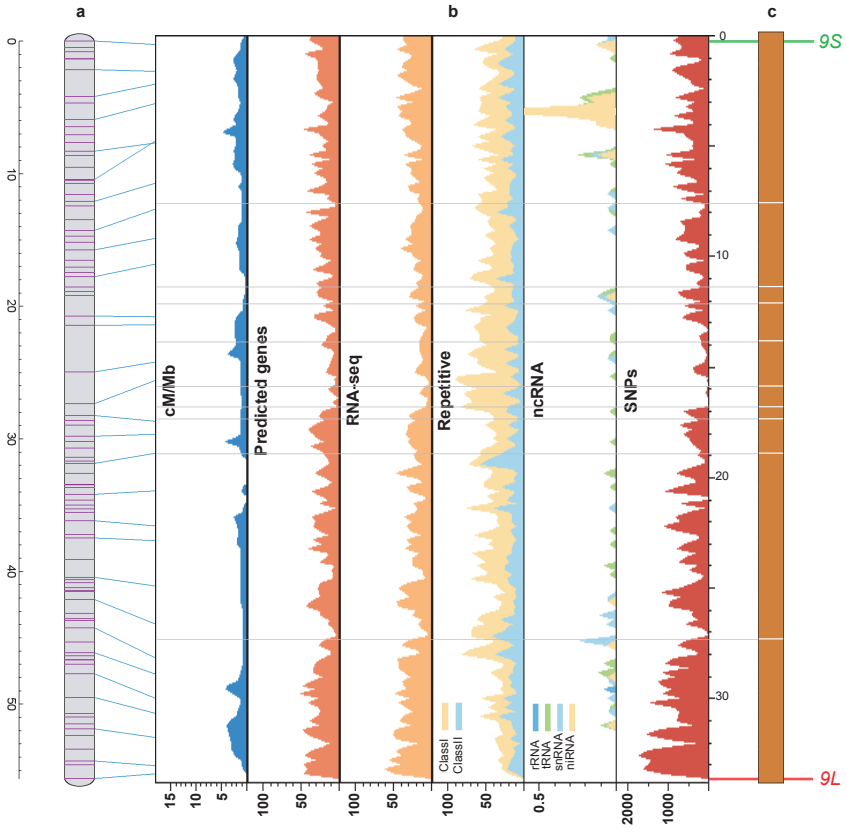
DcChr7



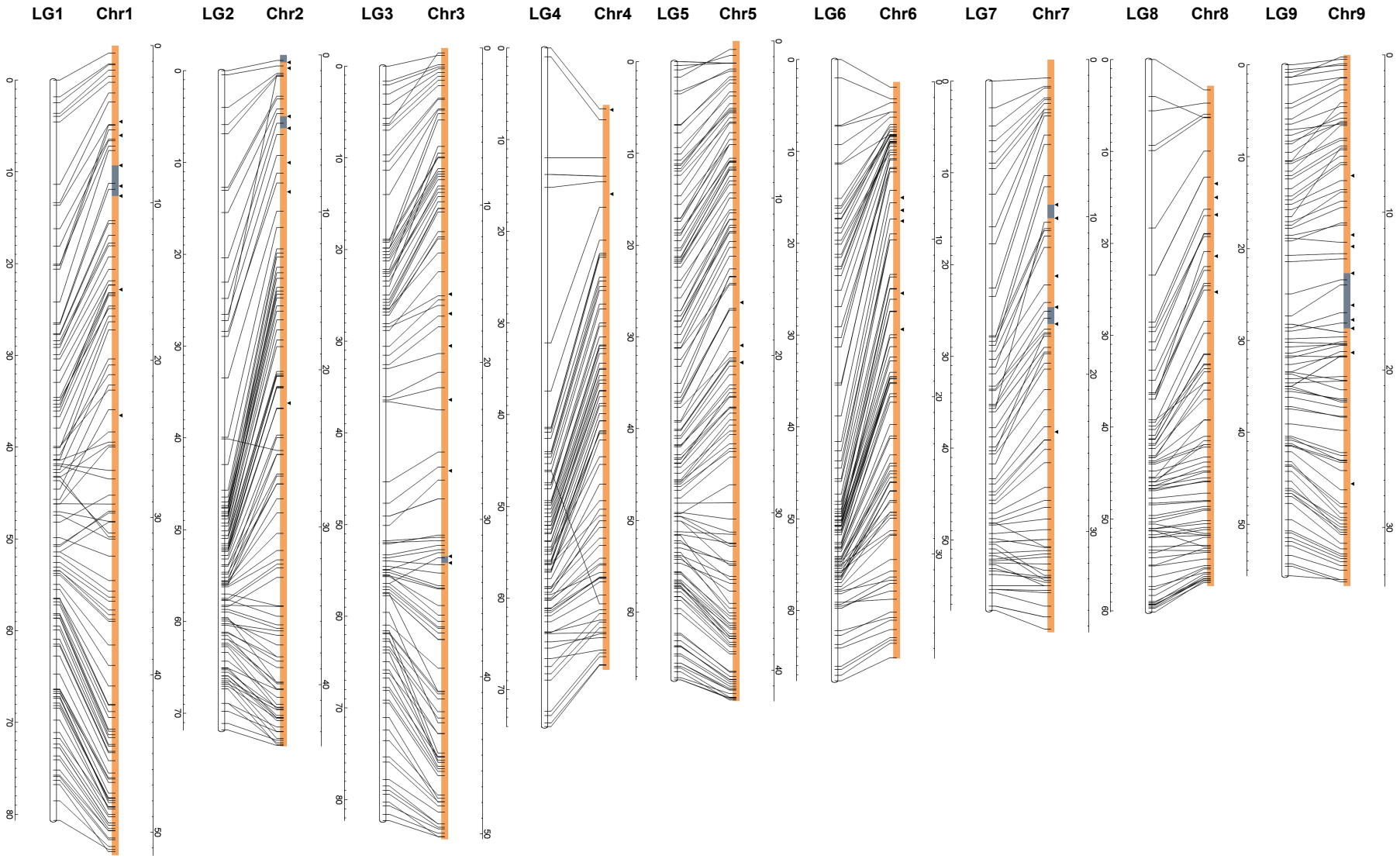
DcChr8



DcChr9



Supplementary Figure 30: Anchoring the carrot genome assembly to the carrot reference genetic map. The carrot chromosome pseudomolecules (right) are shown in orange if superscaffold is oriented, blue if ambiguous orientation. Connections between superscaffolds are marked by triangles. Superscaffolds were anchored to the linkage groups (left) of the *D. carota* genetic bin map with 918 SNP markers.

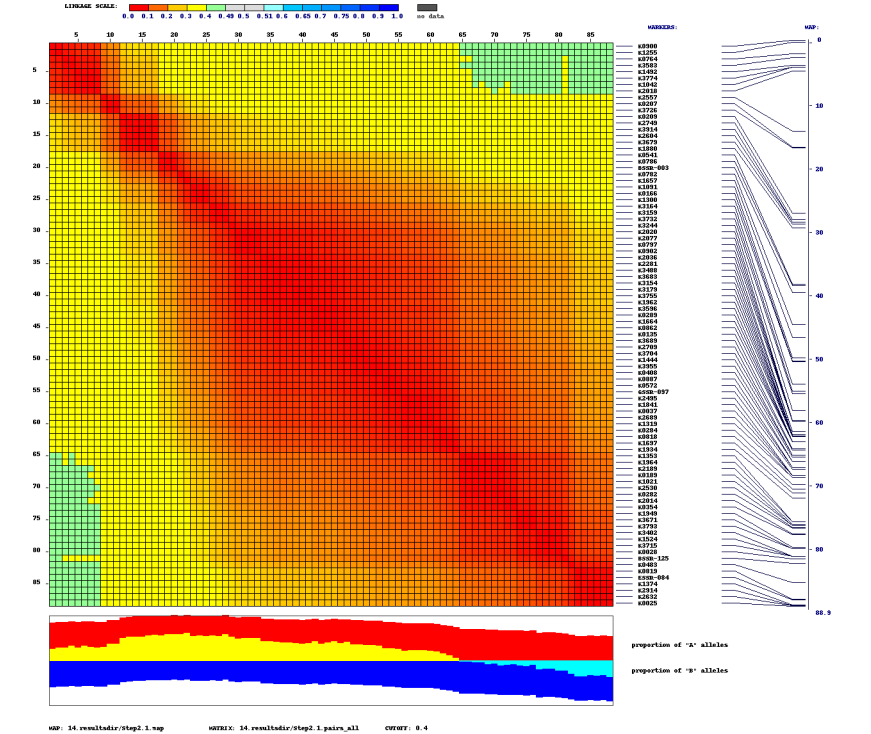


Nature Genetics: doi:10.1038/ng.3565

Supplementary Figure 31: Distribution of haplotypes along the nine linkage groups from the mapping population 85036 using Checkmatrix. The 84 genotypes are arranged along the x-axis and the loci displayed in linear order along the y-axis. Red indicates parent A alleles; blue indicates parent B 85alleles; yellow indicates heterozygous loci; gray indicates missing data. The first column at the right lists the allele identifier. The second column indicates the number of inconsistent scores. The third column indicates the genetic distance. The proportion of each haplotype and the number of crossovers are below each genotype.

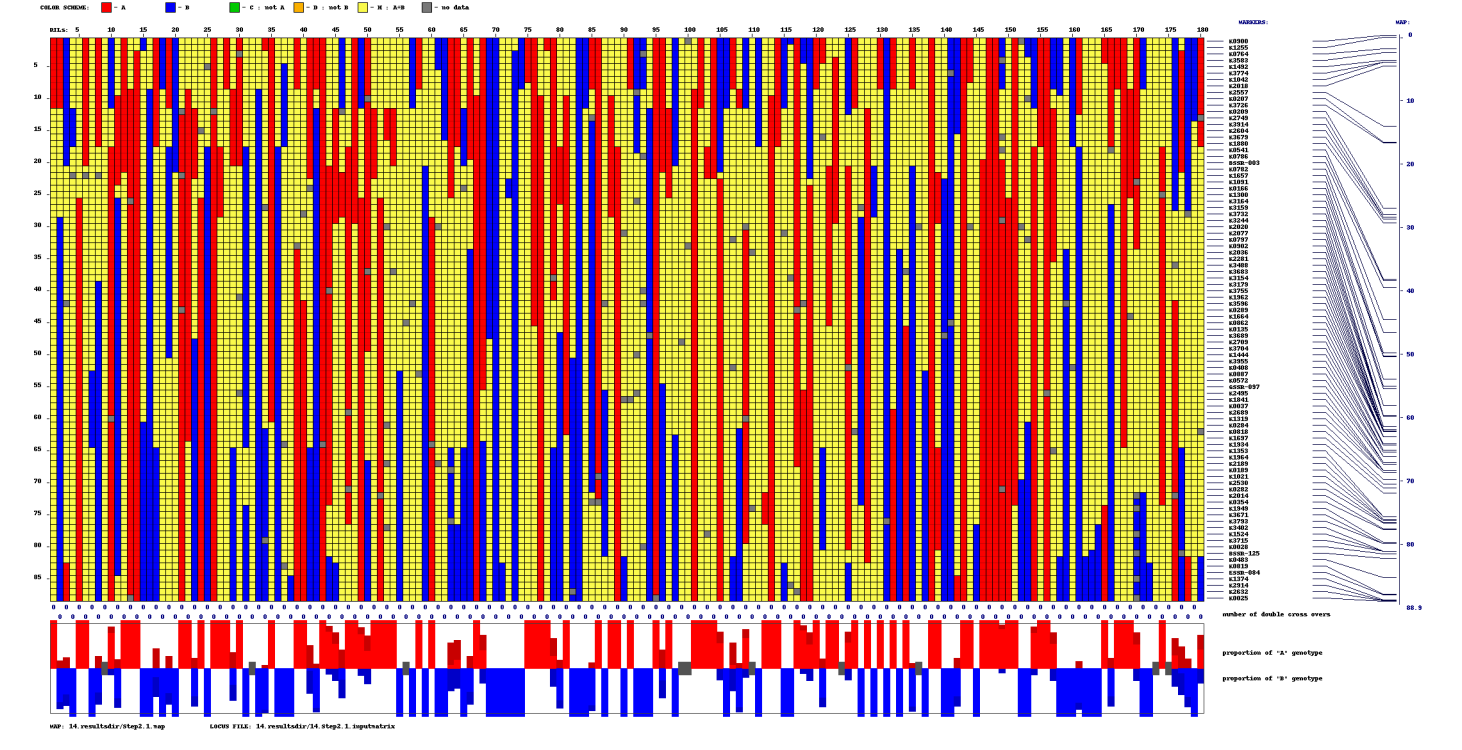
Nature Genetics: doi:10.1038/ng.3565

Supplementary figure 26: Chromosome DCARv2_Chr1 - Heatmap



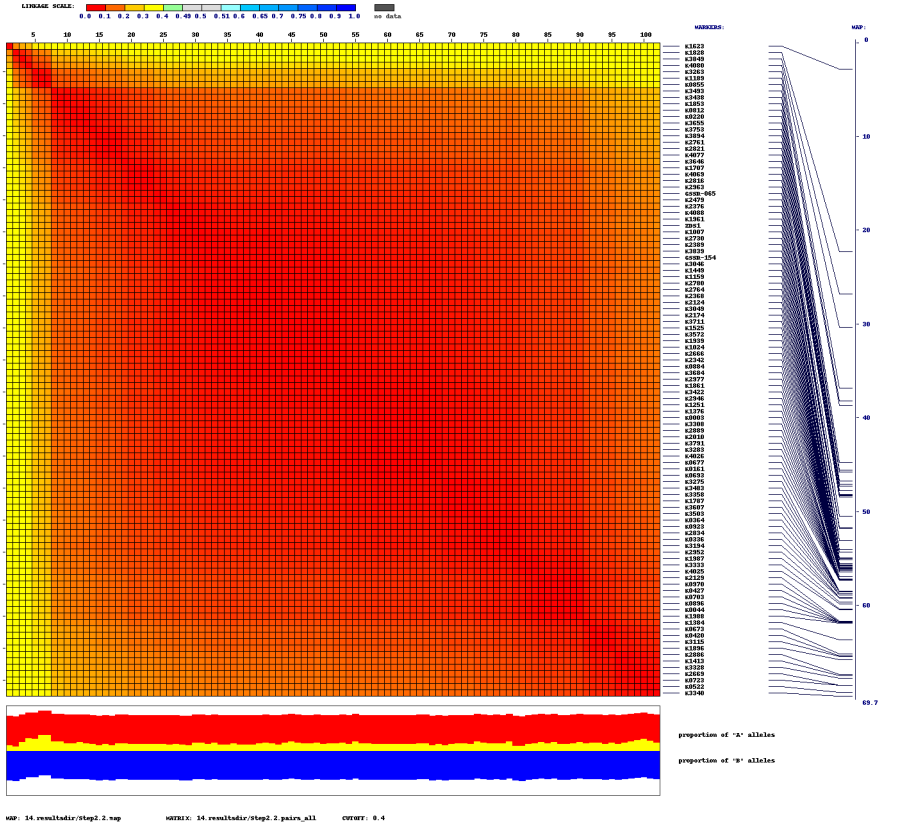
Nature Genetics: doi:10.1038/ng.3565

Supplementary figure 26: Chromosome DCARv2_Chr1 - Recombination matrix

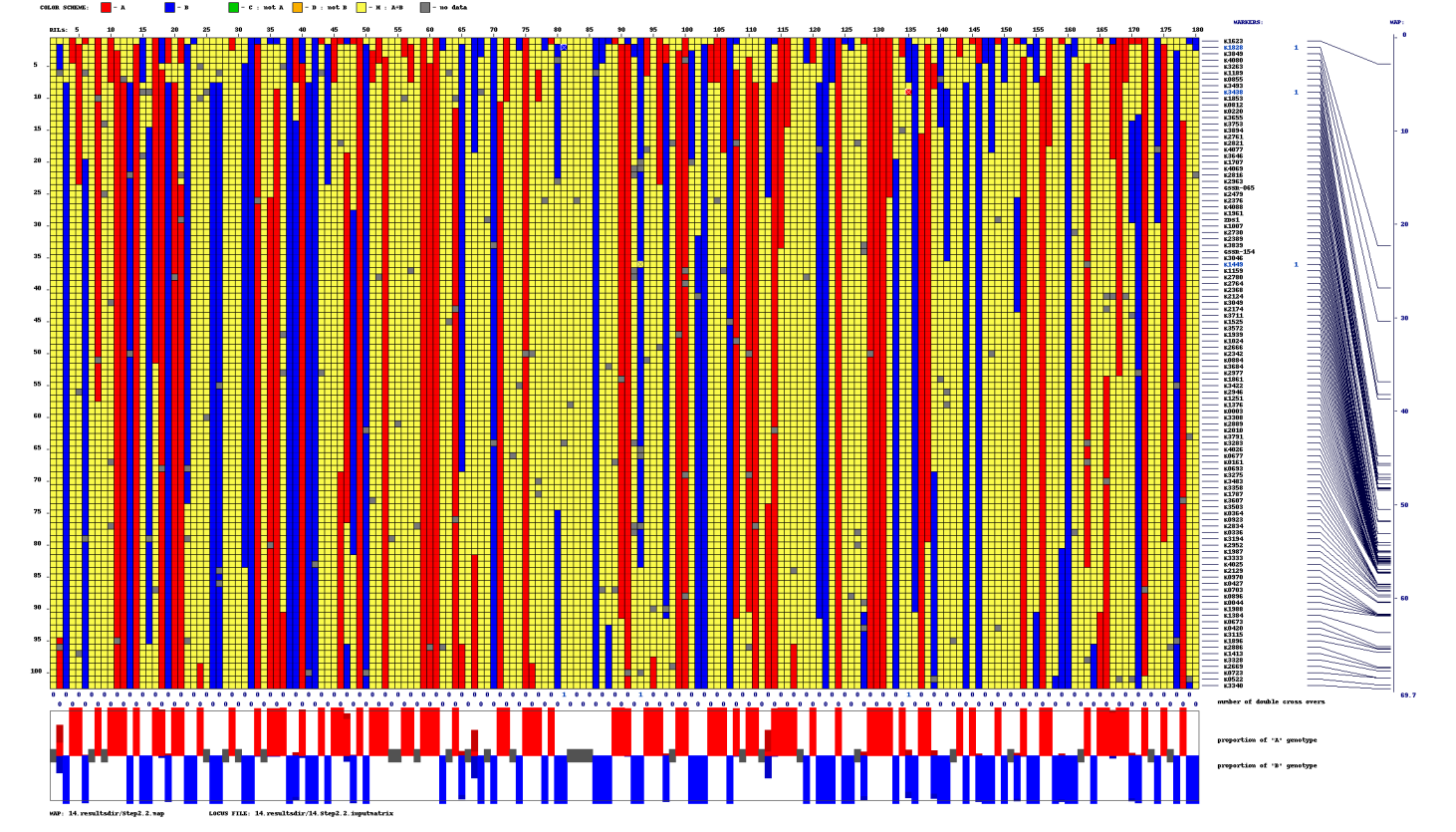


Nature Genetics: doi:10.1038/ng.3565

Supplementary figure 26: Chromosome DCARv2_Chr2 - Heatmap

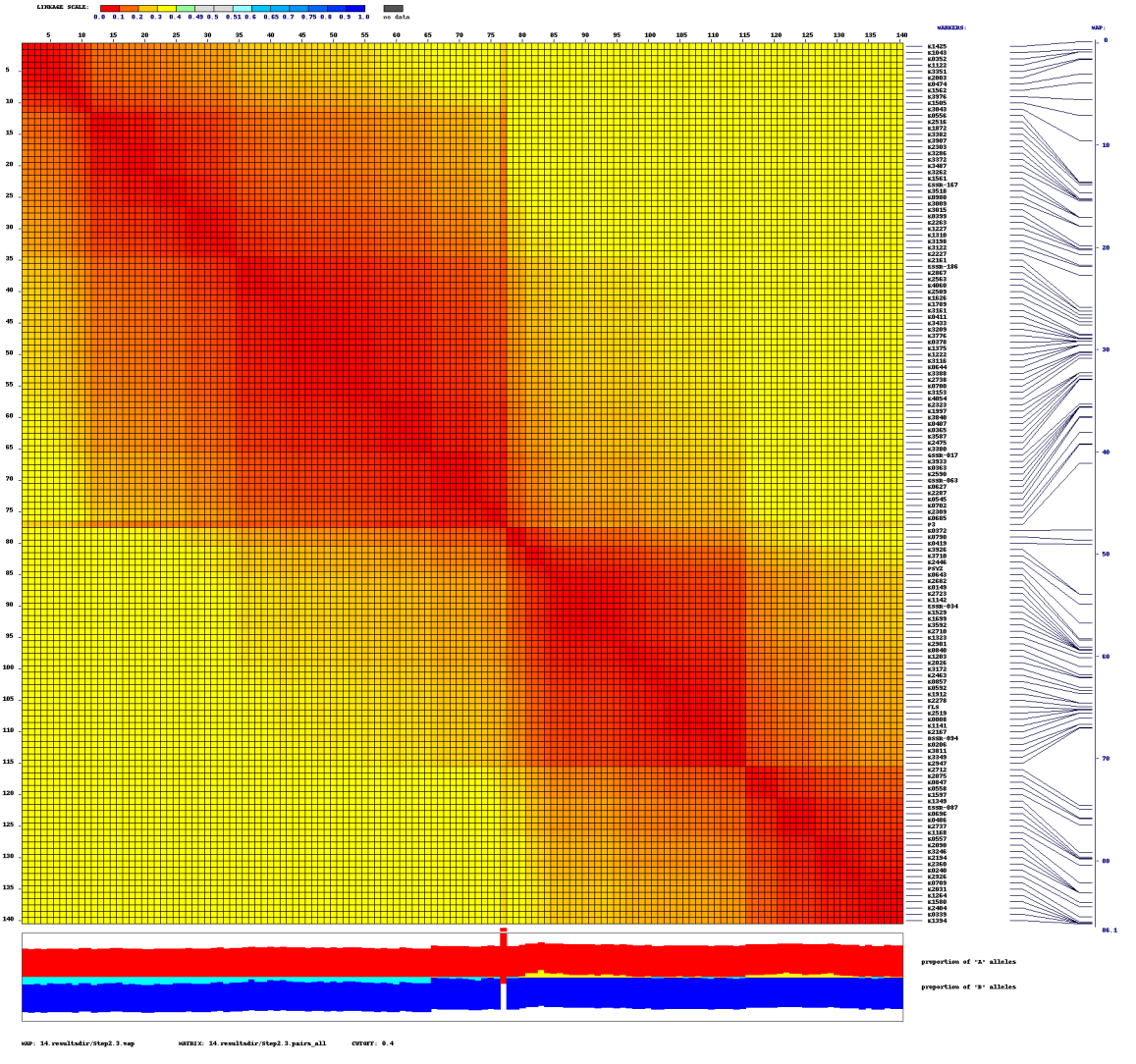


Supplementary figure 26: Chromosome DCARv2_Chr2 - Recombination matrix

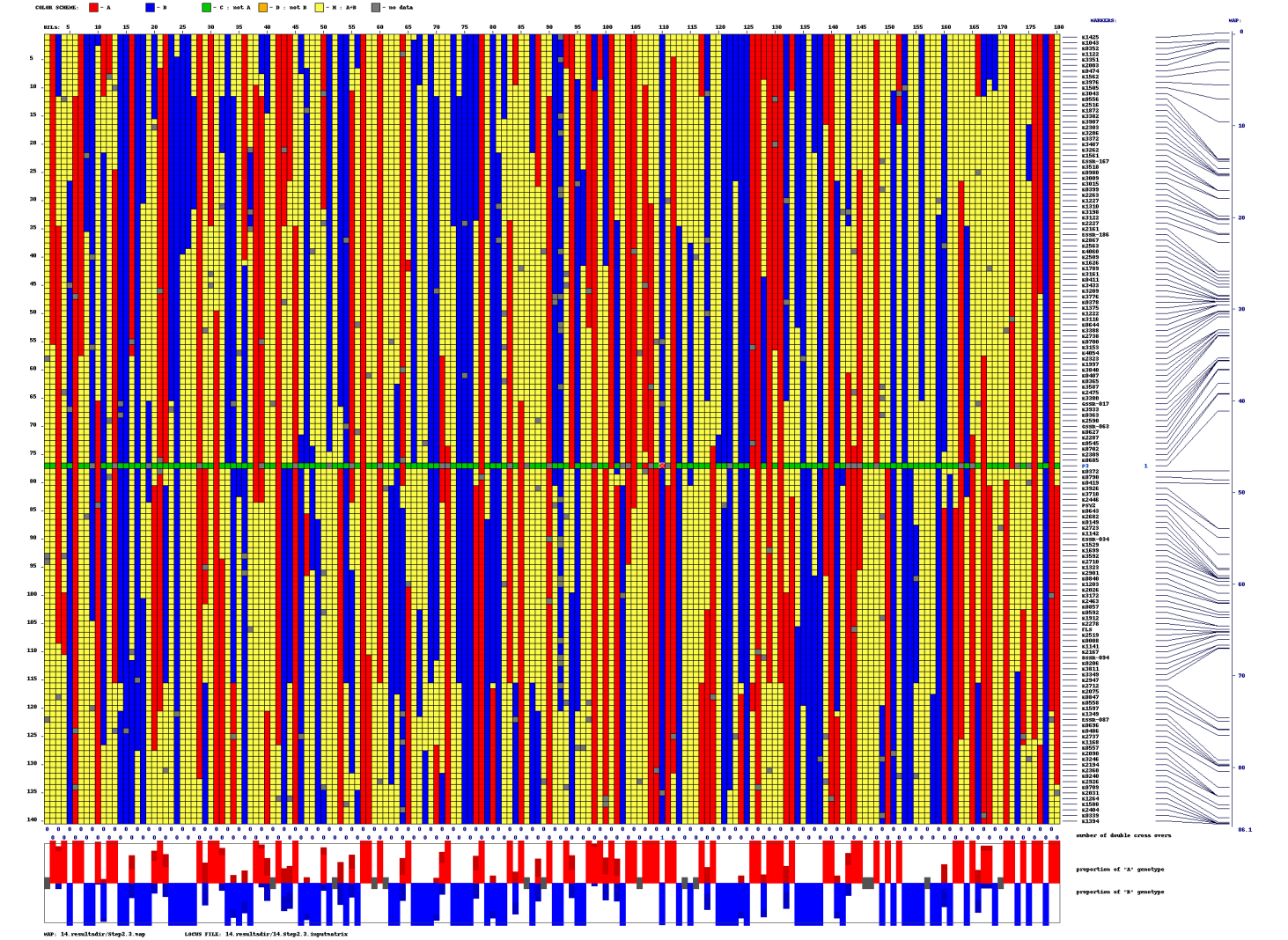


Nature Genetics: doi:10.1038/ng.3565

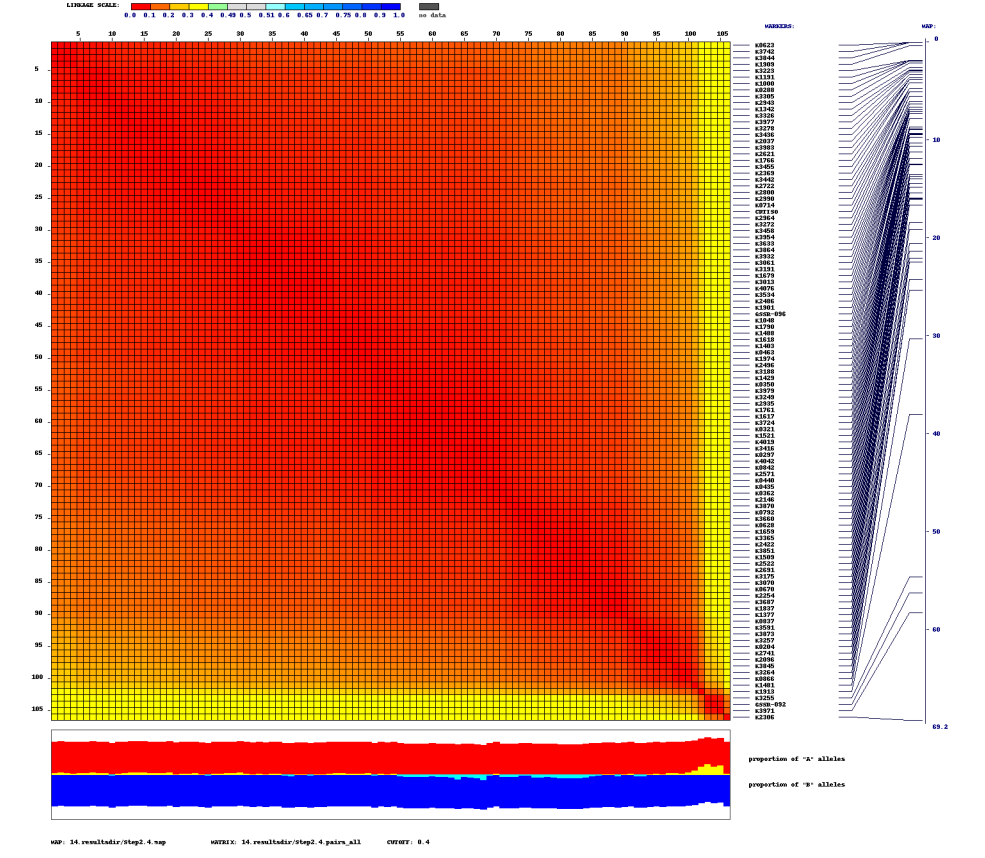
Supplementary figure 26: Chromosome DCARv2_Chr3 - Heatmap



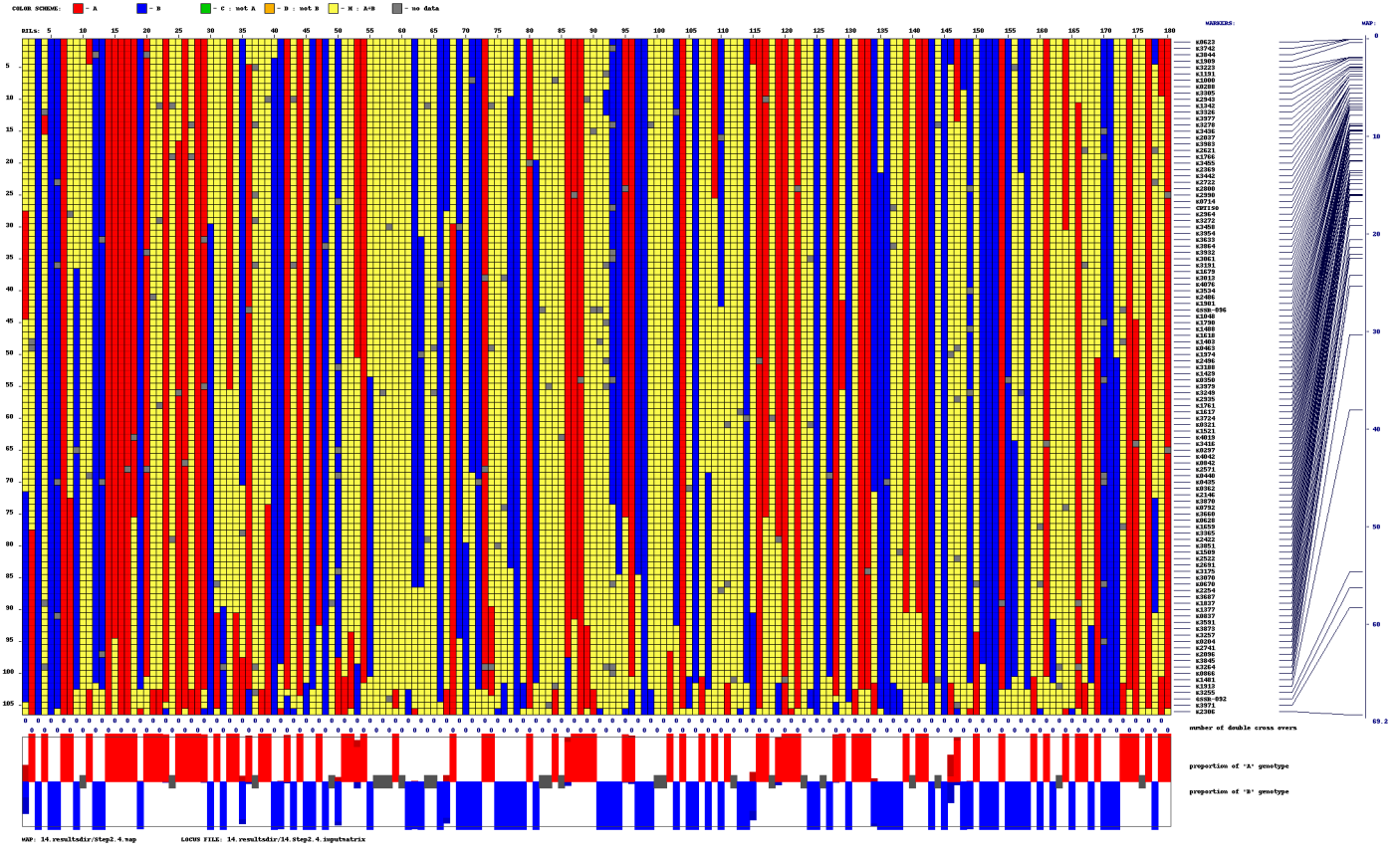
Supplementary figure 26: Chromosome DCARv2_Chr3 - Recombination matrix



Supplementary figure 26: Chromosome DCARv2_Chr4 - Heatmap

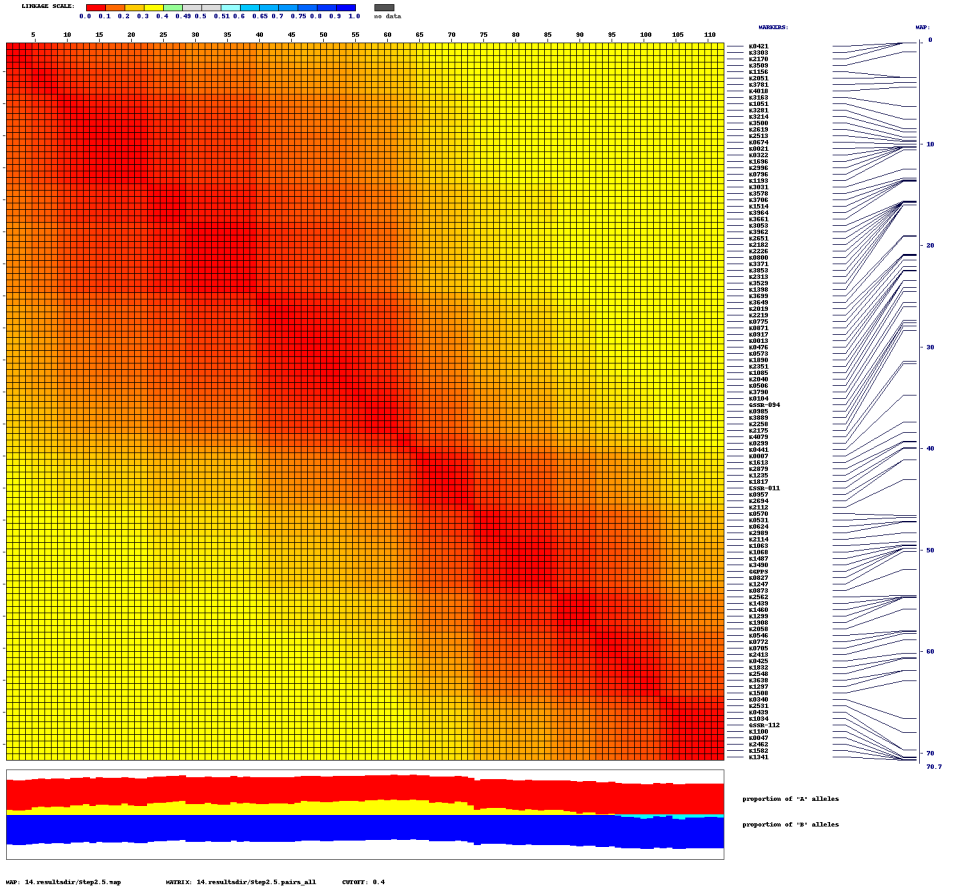


Supplementary figure 26: Chromosome DCARv2_Chr4 - Recombination matrix

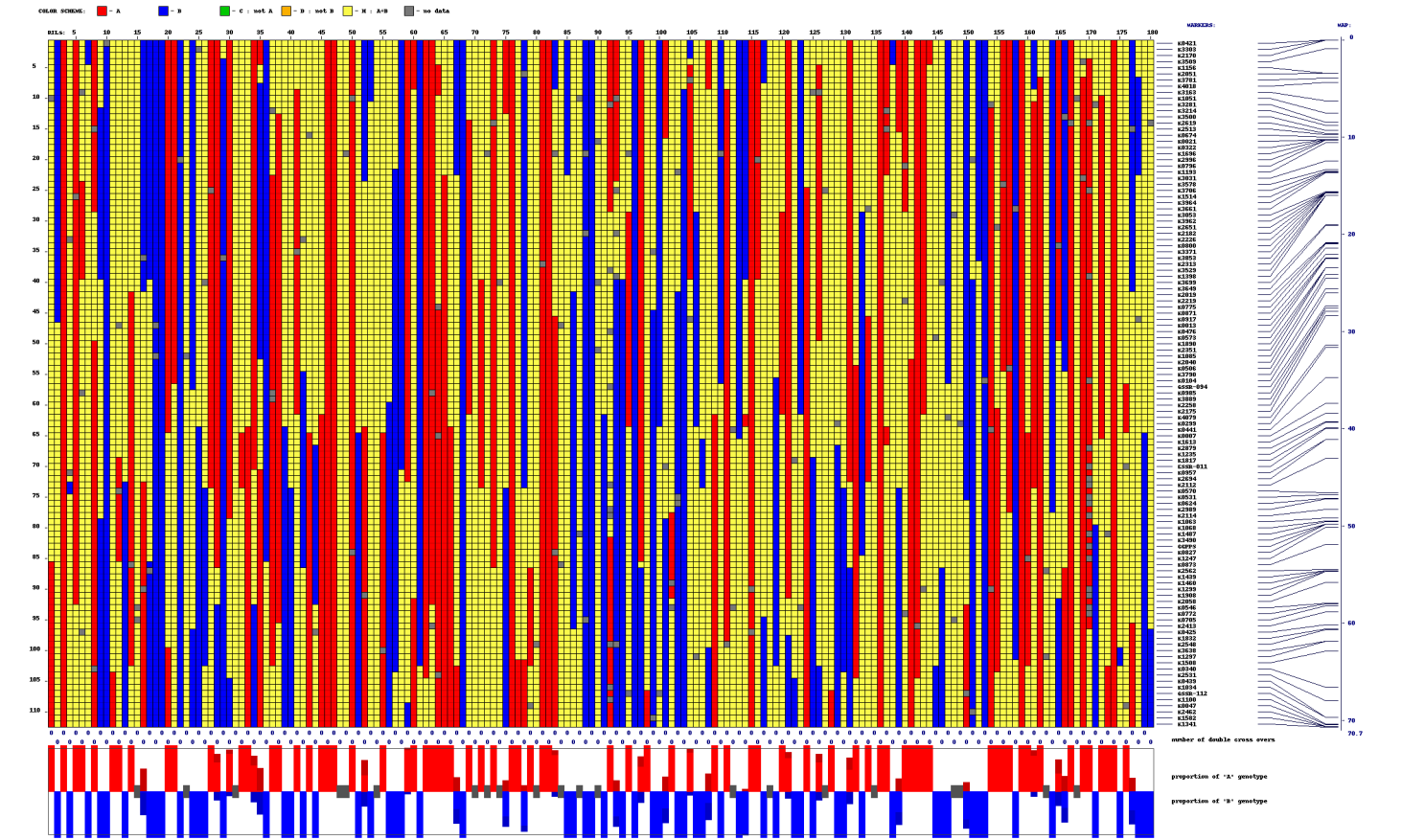


Nature Genetics: doi:10.1038/ng.3565

Supplementary figure 26: Chromosome DCARv2_Chr5 - Heatmap

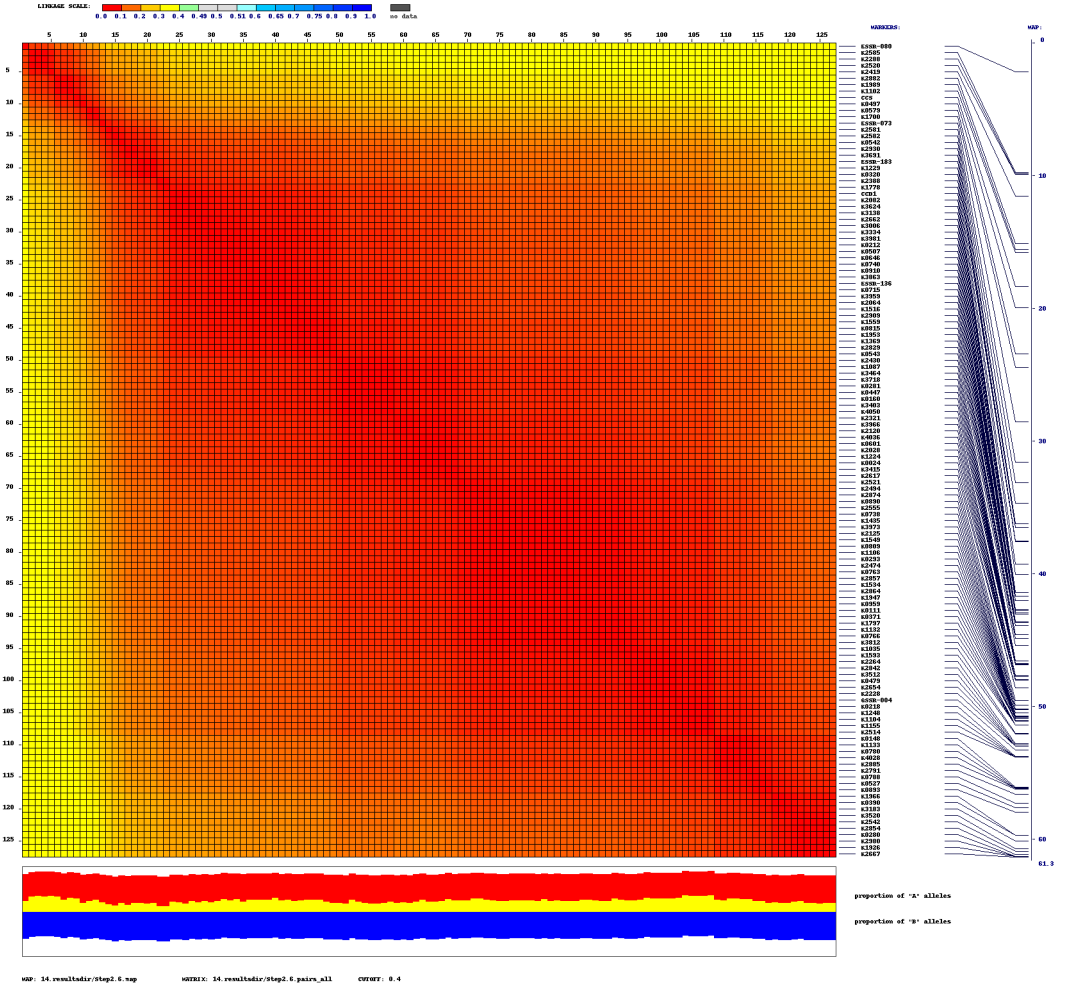


Supplementary figure 26: Chromosome DCARv2_Chr5 - Recombination matrix

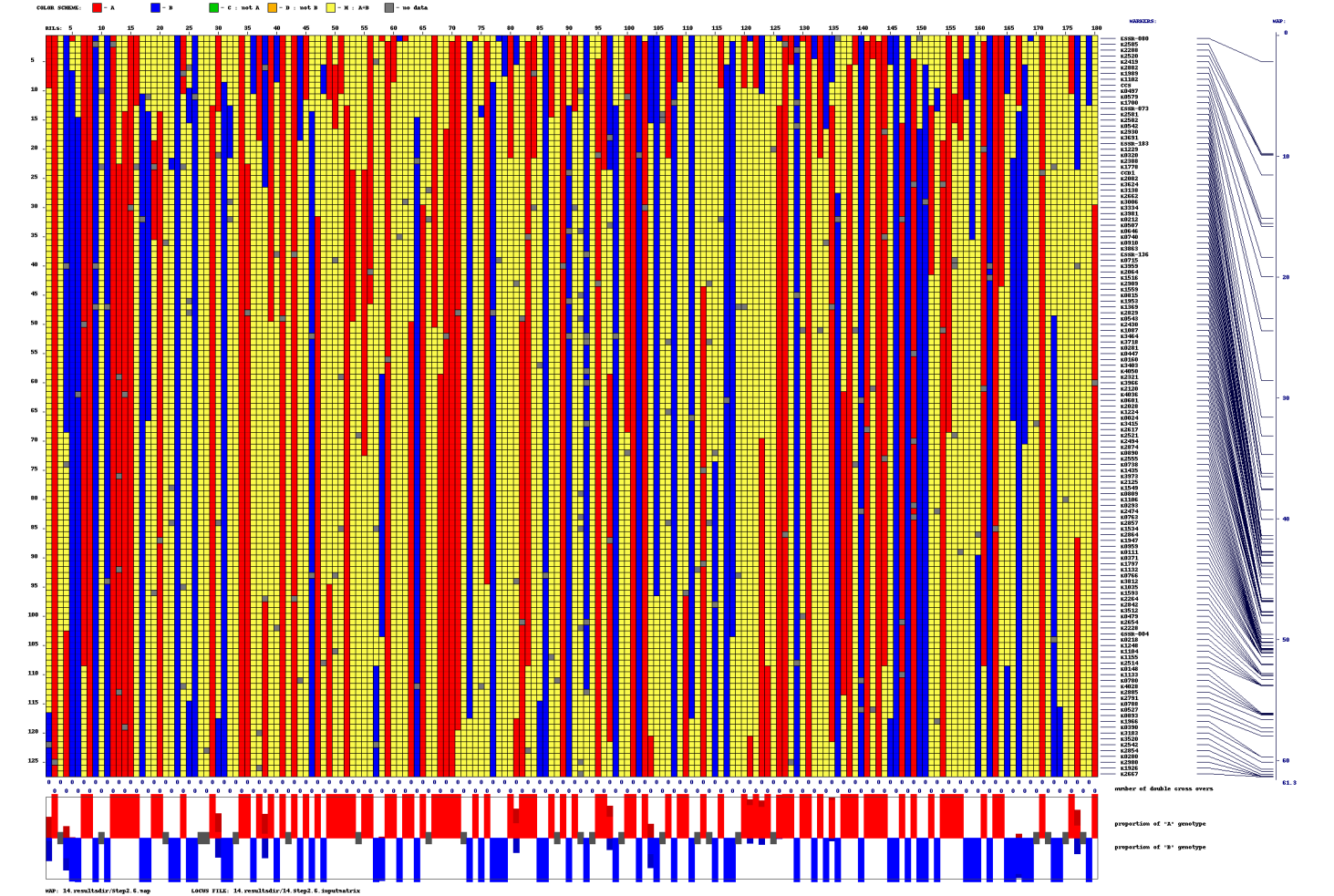


Nature Genetics: doi:10.1038/ng.3565

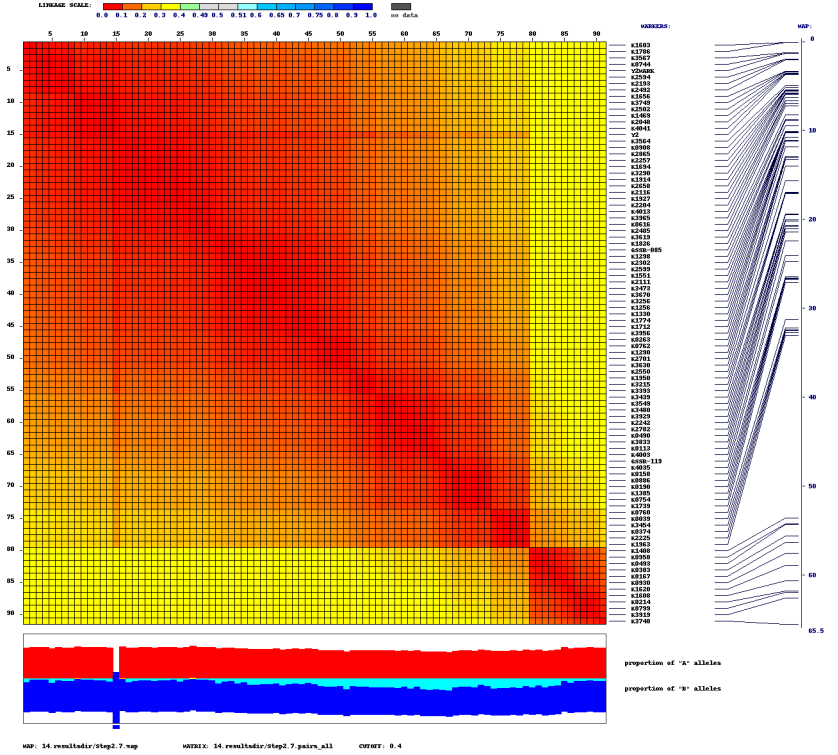
Supplementary figure 26: Chromosome DCARv2_Chr6 - Heatmap



Supplementary figure 26: Chromosome DCARv2_Chr6 - Recombination matrix

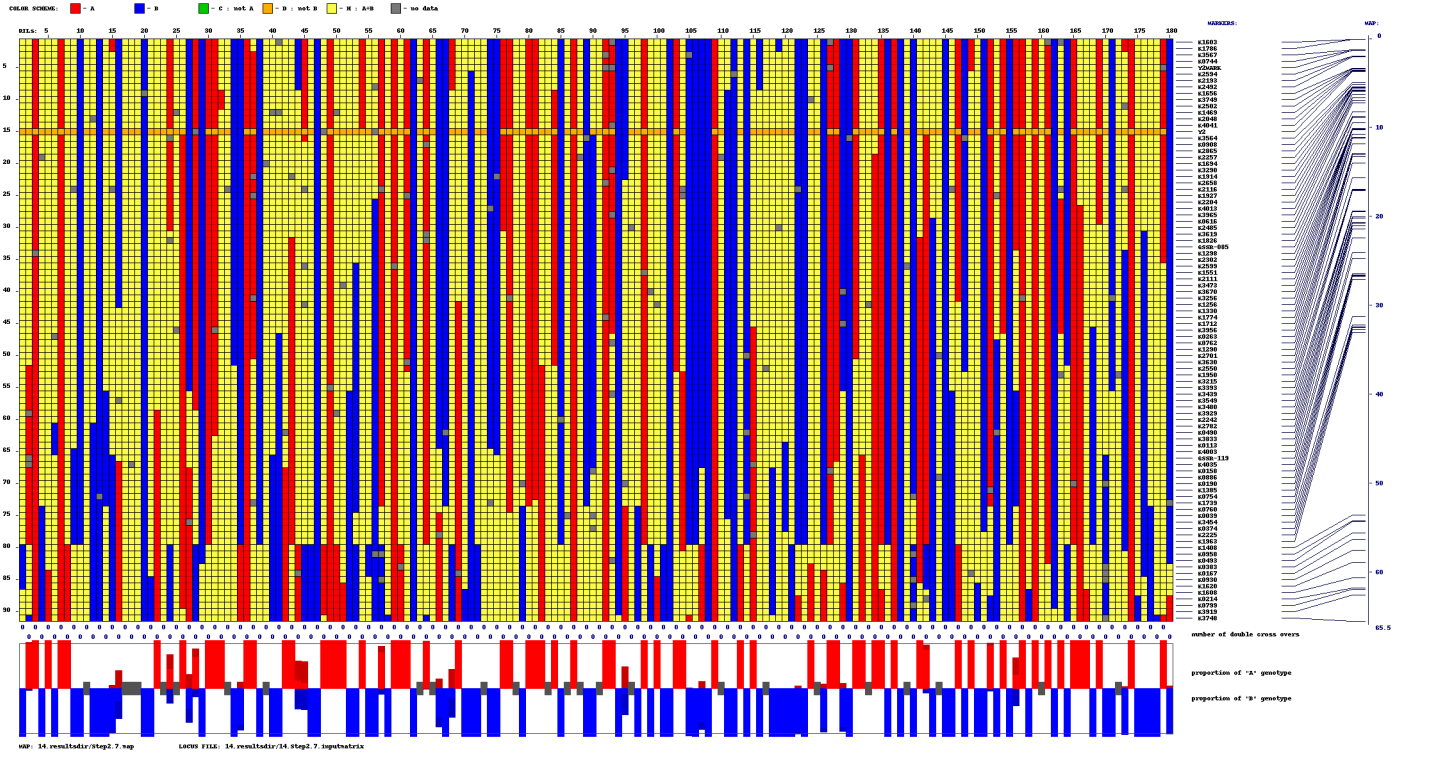


Supplementary figure 26: Chromosome DCARv2_Chr7 - Heatmap



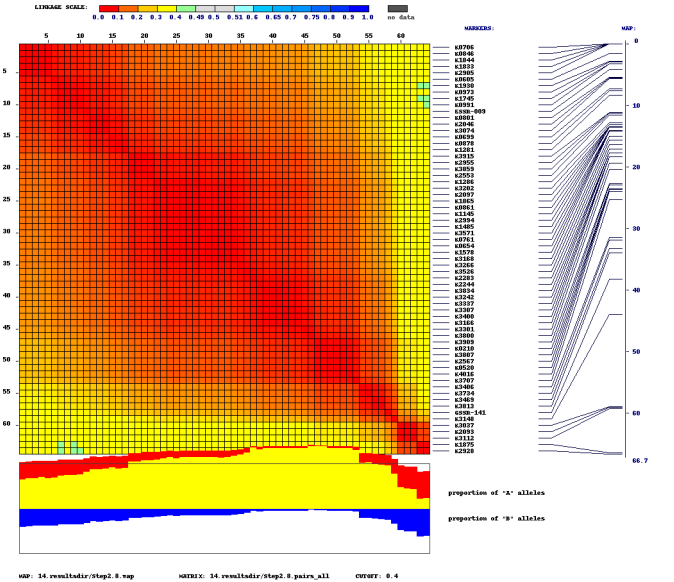
Nature Genetics: doi:10.1038/ng.3565

Supplementary figure 26: Chromosome DCARv2_Chr7 - Recombination matrix



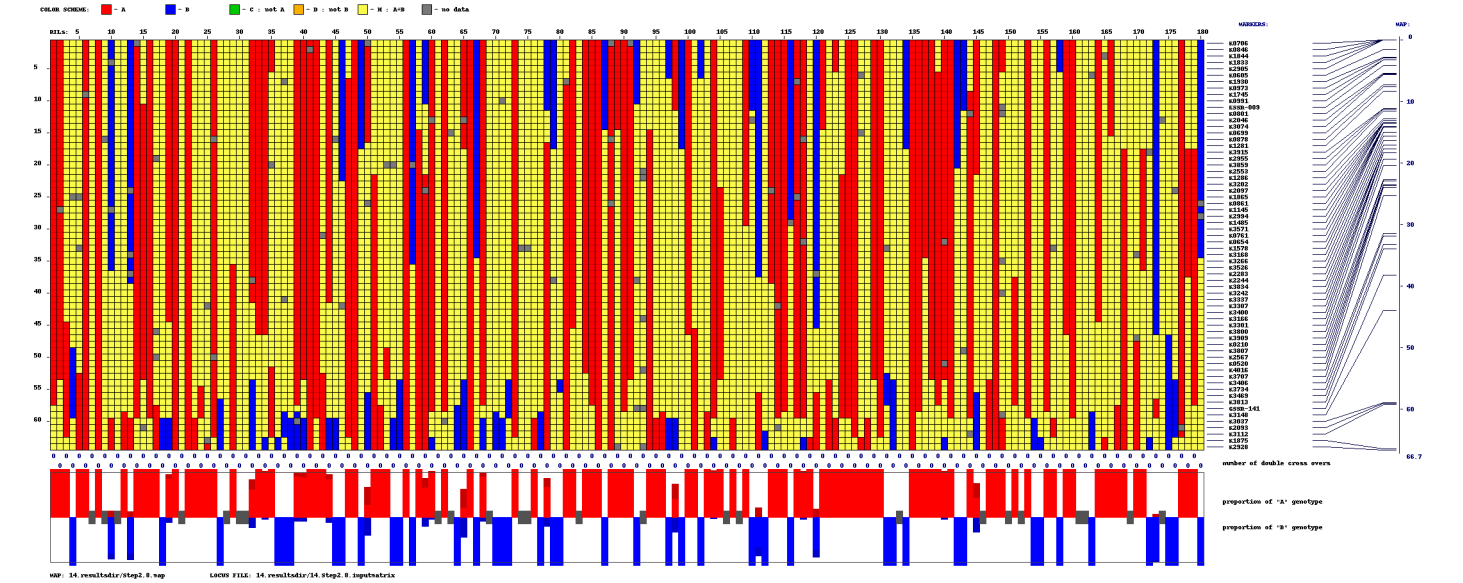
Nature Genetics: doi:10.1038/ng.3565

Supplementary figure 26: Chromosome DCARv2_Chr8 - Heatmap

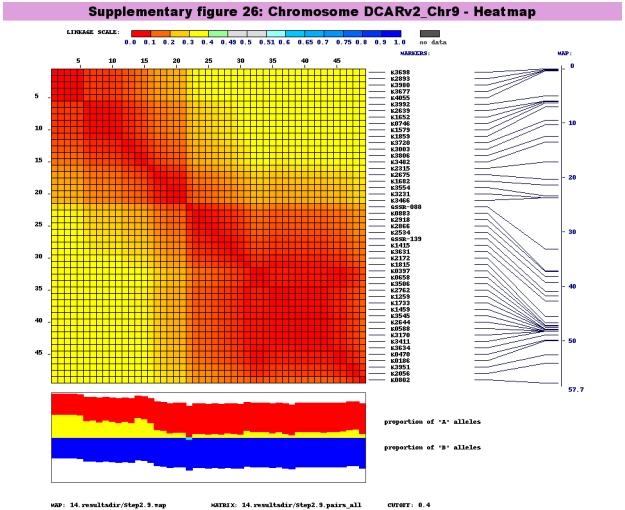


Nature Genetics: doi:10.1038/ng.3565

Supplementary figure 26: Chromosome DCARv2_Chr8 - Recombination matrix

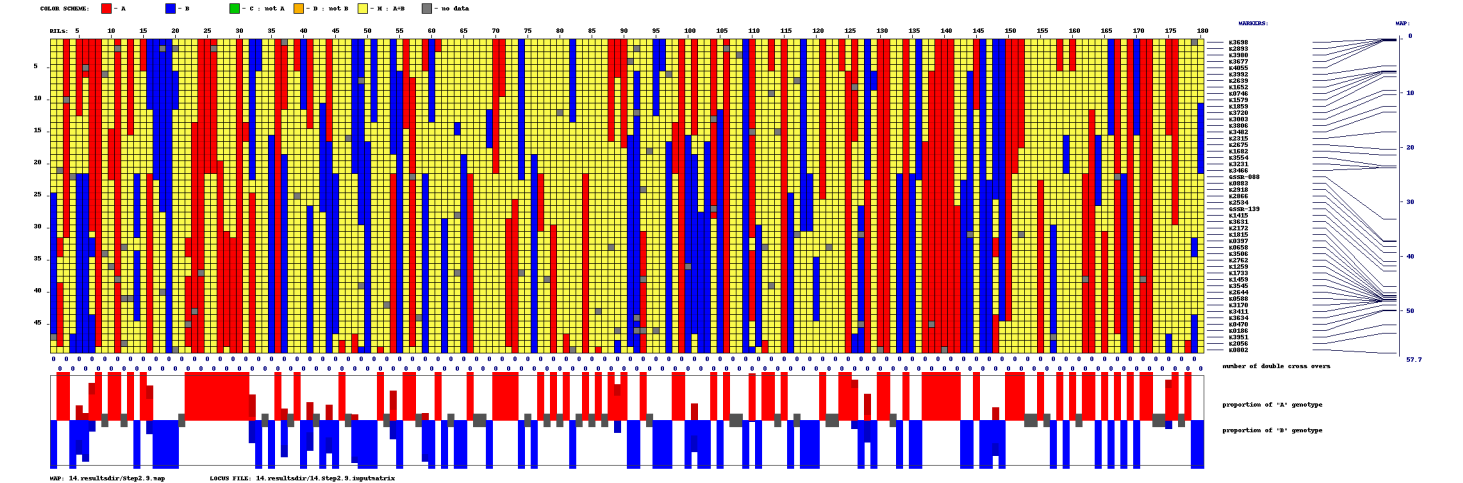


Nature Genetics: doi:10.1038/ng.3565



Nature Genetics: doi:10.1038/ng.3565

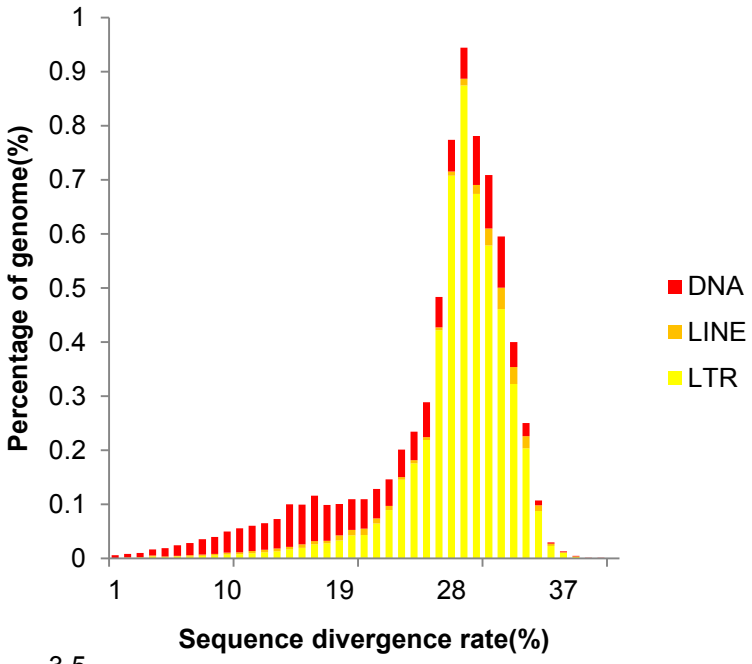
Supplementary figure 26: Chromosome DCARv2_Chr9 - Recombination matrix



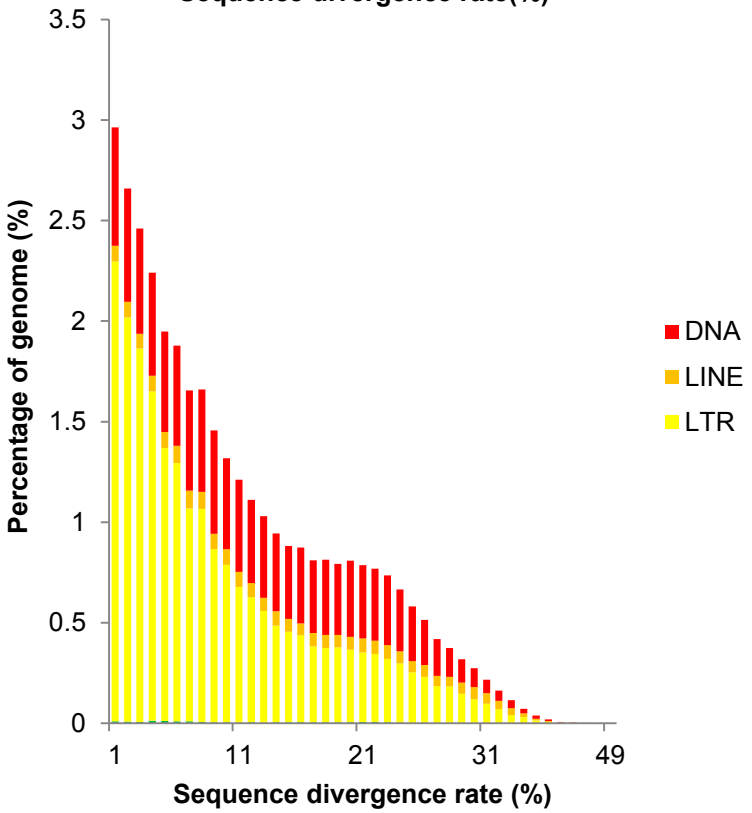
Nature Genetics: doi:10.1038/ng.3565

Supplementary Figure 32: Distribution of divergence rates of mobile elements (ME) in carrot DH1. The divergence rate was calculated between the ME elements identified in the genome by homology compared to the consensus sequence in the Repbase (panel A) and in the predicted TE library (Panel B). SINE elements are not included due their small representation among MEs.

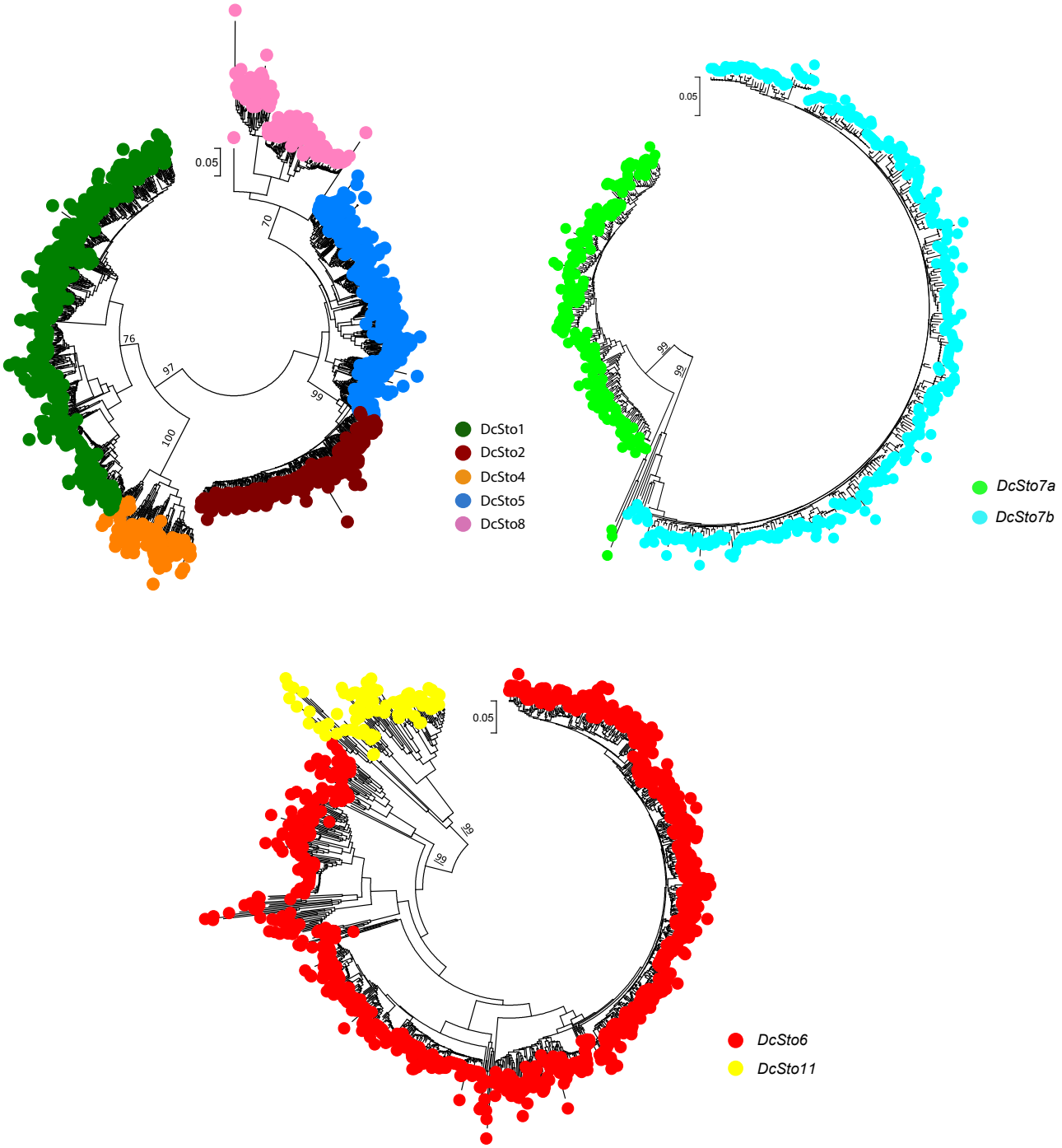
A



B



Supplementary Figure 33: Phylogenetic tree of DcSto families in carrot DH1 based on the genetic distances calculated with the neighbor-joining method (NJ). To evaluate their phylogenetic relationships and the robustness of our classification, DcSto families with shared similarity from the analysis with Circoletto were analyzed together.



Supplementary Figure 34: Inter- and intra-specific similarity among families of carrot (*DcSto*), potato (*StSto*), pepper (*CaSto*) and tomato (*SlSto*) Stowaway-like elements. Each family is represented by a consensus sequence. Colored ribbons indicate regions of similarity based on blastn results. Colors represent similarity levels (blue<green<orange<red). The diagram was drawn with Circoletto[42].