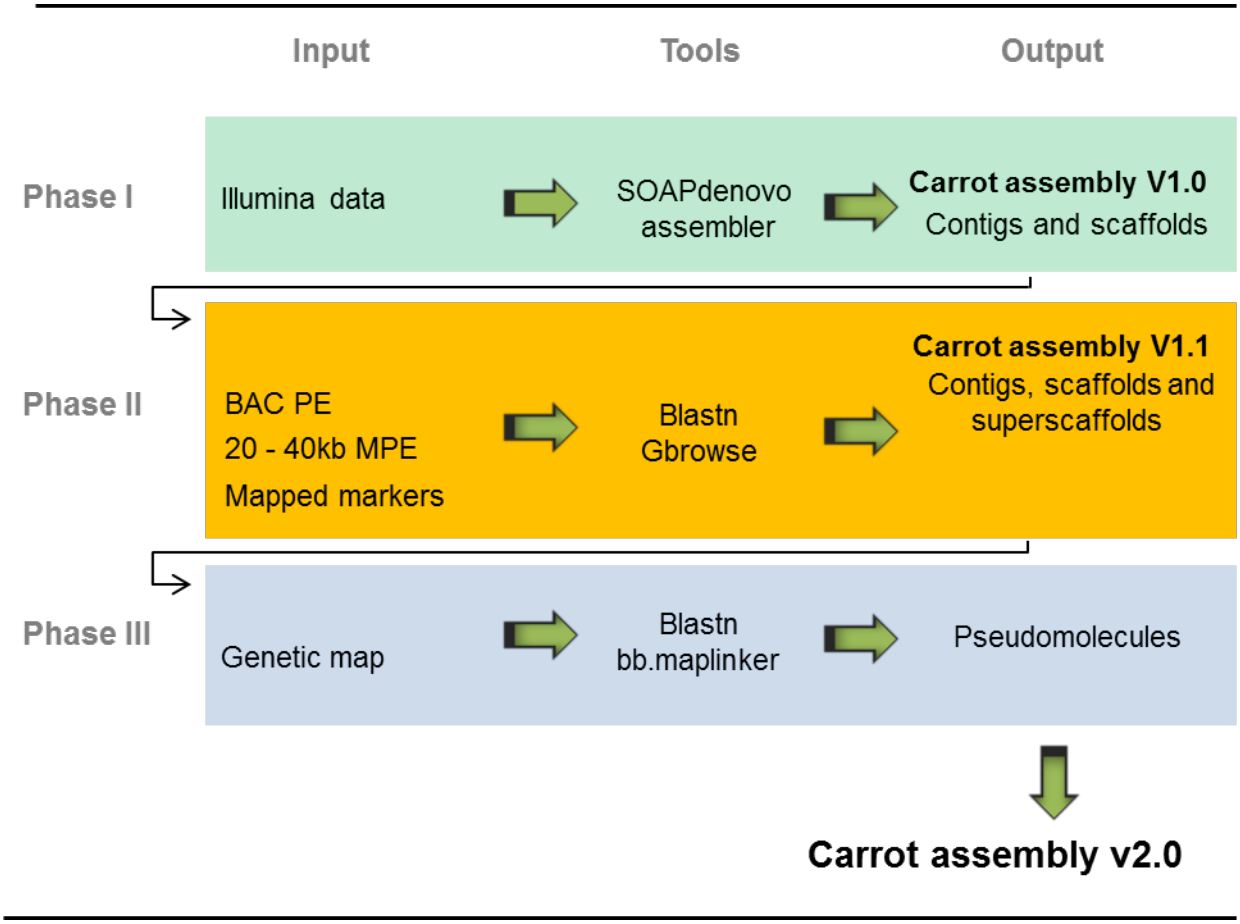
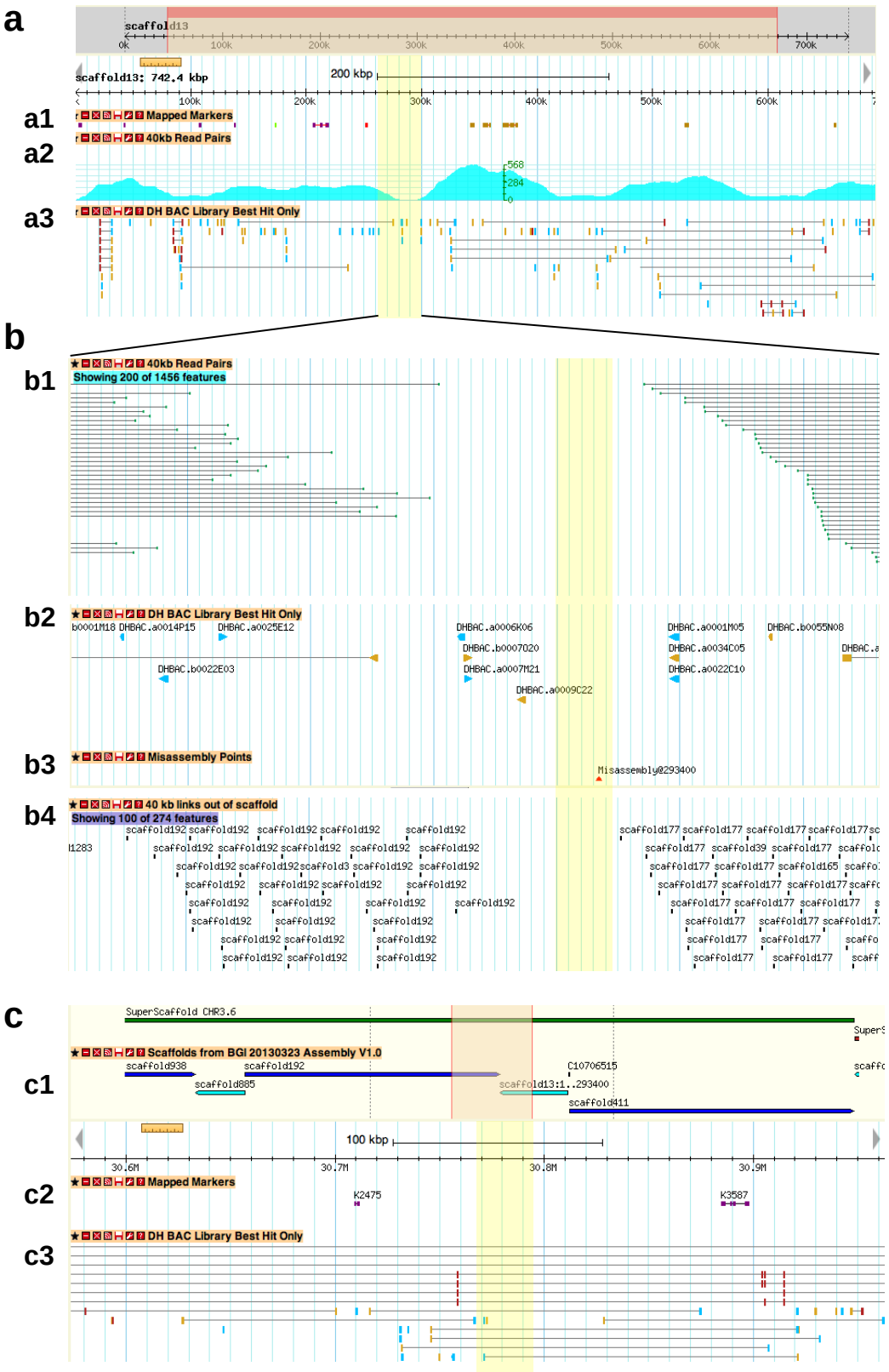
Supplementary Figure 1: Scheme of the carrot genome assembly pipeline. In Phase I, quality filtered Illumina data from eight insert libraries (Supplementary Table 49) were assembled using SOAPdenovo (http://soap.genomics.org.cn) producing the carrot assembly v1.0, which included contigs and scaffolds. In Phase II, unambiguously aligned sequences from mapped molecular markers, BAC end sequences and 20 and 40 kb Illumina MPE were visualized in Gbrowse to manually inspect and correct chimeric regions and construct superscaffolds. This process produced the carrot assembly v1.1 which included contigs, scaffolds and superscaffolds. In Phase III, the integrated linkage map was used to anchor superscaffolds and construct the nine carrot pseudomolecules (chromosomes). The final assembly named carrot assembly v2.0 includes pseudomolecules and the remaining unanchored contigs, scaffolds and superscaffolds.

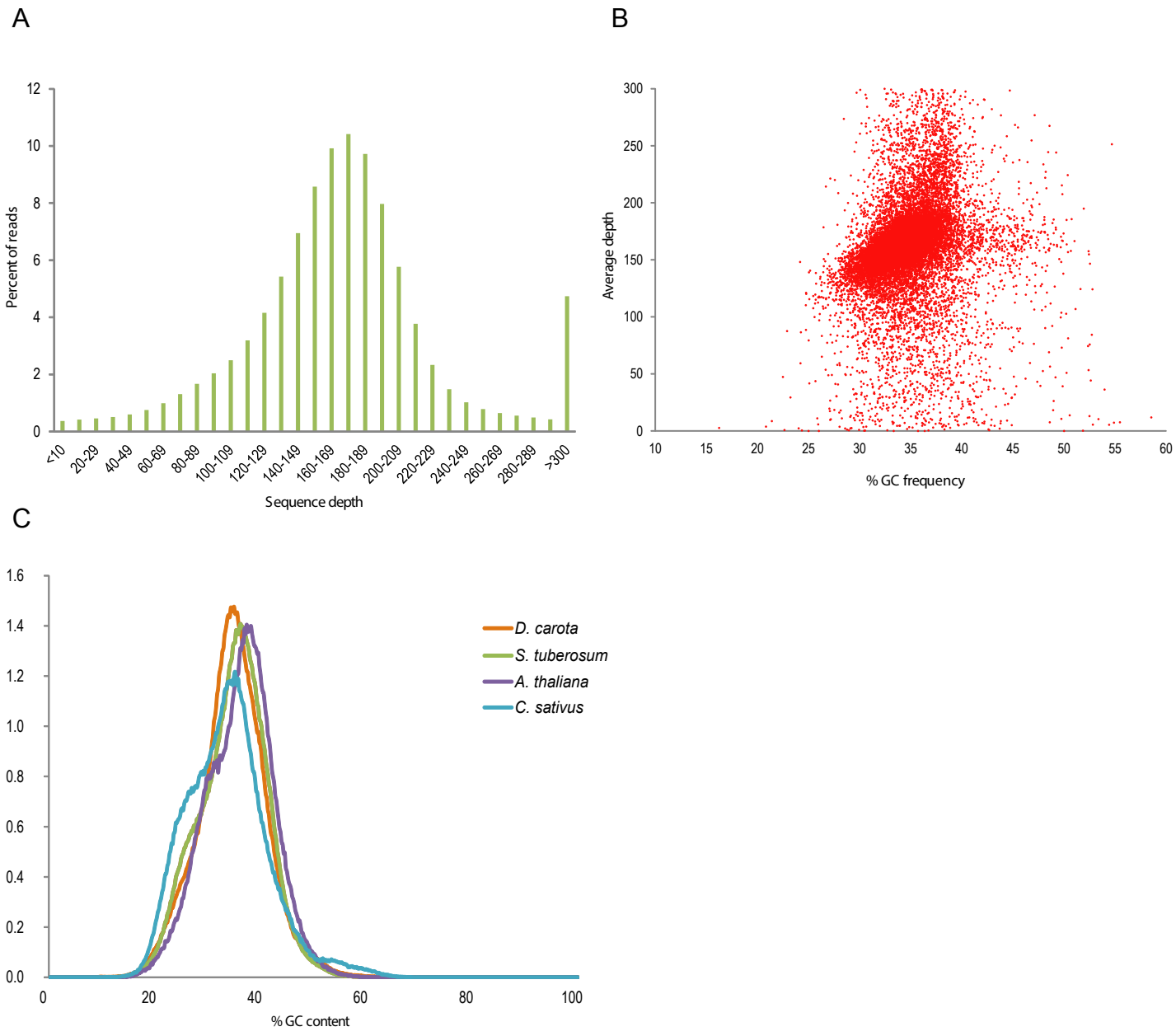


Supplementary Figure 2: **Example of a chimeric scaffold in the carrot assembly v1.0**. The yellow vertical highlight identifies the chimeric region on scaffold13 of carrot assembly v1.0 (Panel A and B) and its correction in the carrot assembly v2.1 (Panel C). Panel A) Gbrowse window of scaffold13 (carrot assembly v1.0). ***a1***: indicates the mapping location of markers unambiguously aligned to scaffold13; markers in purple mapped to LG3(CHR3) and markers in brown mapped to LG5(CHR5); ***a2-a3***: 40kb paired end reads (PE) and DH BAC end sequences that unambiguously aligned to scaffold13; Panel B) Gbrowse enlarged window of scaffold13 covering the chimeric region. ***b1-b2***; enlarged window of the region where neither the 40kb nor the BAC end sequences spanned a contiguous connection on scaffold13; ***b3***: misassembly point were the scaffold13 has been split; ***b4***: 40kb PE reads that unambiguously aligned to one side of the misassembled region on scaffold13. The PE reads on the left side unambiguously link to scaffold192, the PE reads on the right unambiguously link to scaffold 177. Panel C) Gbrowse window of carrot assembly v2.0, CH3, superscaffold6 (CH3.6). ***c1***: order of scaffolds in superscaffold6 spanning the corrected chimeric assembly. ***c2***: mapping location of markers unambiguously aligned to superscaffold CH3.6; markers on both side of the corrected chimeric assembly mapped to LG3(CHR3) at 30.4 cM; ***c3***: BAC end sequences that unambiguously aligned to superscaffold CH3.6. The BAC end sequences span the region connecting the misassembly point on scaffold13 and its connection to scaffold192 at an expected average distance of 150 kb.

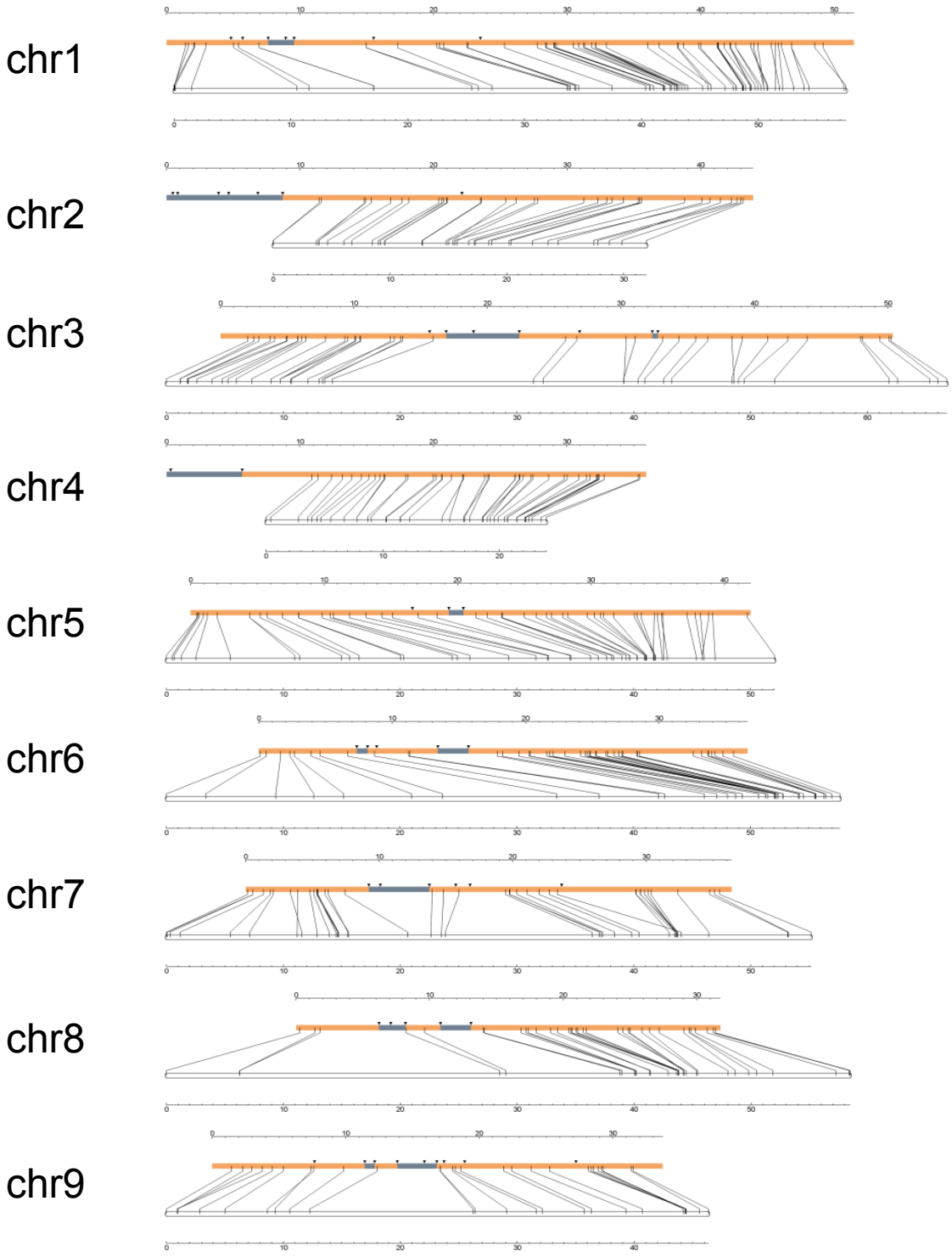


Nature Genetics: doi:10.1038/ng.3565

Supplementary Figure 3: Relationships between sequence depth and GC content of the carrot genome. A: Read depth distribution on the carrot genome assembly. B: Relationship between percent of GC content and sequencing depth. The x-axis represents the percent GC content; the y-axis represents the average sequence depth. A 10 kb non-overlapping sliding window was used to calculate the GC content and the average depth. C: Comparisons of percent GC content across four species, including carrot (Daucus carota), potato (Solanum tuberosum), Arabidopsis thaliana and cucumber (Cucumis sativus).

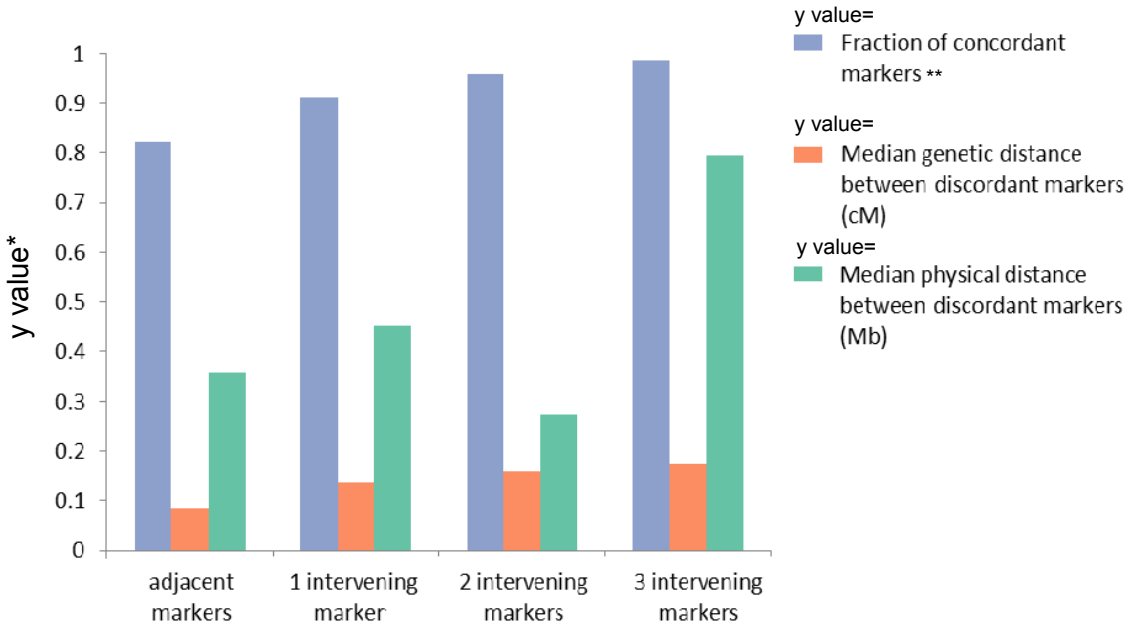


Supplementary Figure 4: Comparison of the genetic map of population 85036 to the physical map of the DH1 anchored genome. The upper bars indicate pseudomolecules, and the lower bars represent linkage groups, corresponding to the nine chromosomes. The orange blocks and gray lines indicate scaffolds matched to the genetic map. Gray blocks indicate scaffolds that did not match with any markers. Triangles indicate break points between scaffolds.



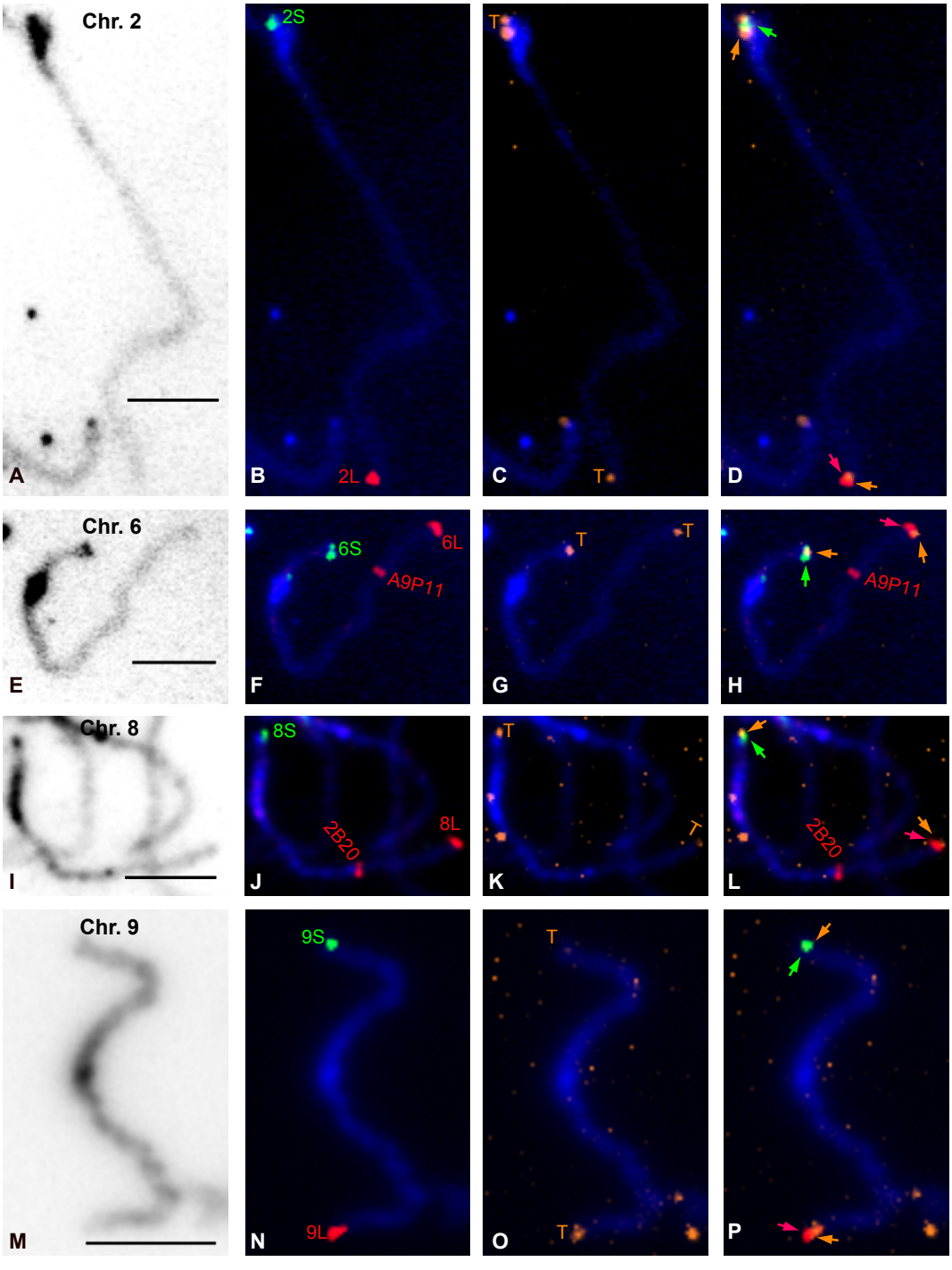
Nature Genetics: doi:10.1038/ng.3565

Supplemementary Figure 5: Evaluation of alignment consistency between the 85036 linkage map and the order of sequences in the nine carrot pseudomolecules. The analysis was carried out by comparing adjacent pairs of markers, and pairs with one, two or three intervening mapped markers. Blue bars represent the fraction of markers that matched the order of sequences in the nine pseudomolecules. Orange bars represent the median genetic distance between discordant markers. Green bars represent the physical median distance between discordant markers. In pairwise comparisons of adjacent, those with one intervening marker, those with two, and those with three intervening markers, the percent of concordant markers ranged from 82 to 98%, and the median genetic distance between discordant markers ranged from 0.08 to 0.18 cM suggesting that inconsistent alignments likely reflect technical accuracy and low map density.



*y axis values are indicated on the legend on right side of the figure.

**calculated as %/100.



Supplementary Figure 6: FISH-based confirmation of the consistency and coverage of the carrot genome assembly at the telomeric regions of the DH1 chromosomes. Chr. 2 (A-D), 6 (E-H), 8 (I-L) and 9 (M-P). DAPI stained pachytene Chr. 2 (A), Chr. 6 (E), Chr. 8 (I) and 9 (M) are converted to black and white images. (B-F-J-N): FISH signals derived from BAC clones that unambiguously aligned to sequences close to the start ('S', green signals) and the ends ('L', red signals) of the pseudomolecules of Chr. 2 (B), 6 (F), 8 (J) and 9 (N). BACs A9P11 and 2B20 (red signals) are specific for Chr. 6 and 8, respectively. (C-G-K-O) Same chromosomes probed with a telomeric oligo-based probe ('T', orange signals). (D-H-L-P) Merged images demonstrating the co-localization of the 'S' (green signals/arrows) and 'L' (red signals/arrows) BACs with the telomeric probe (orange signals/arrows). Bar scale=5 μm.

Nature Genetics: doi:10.1038/ng.3565

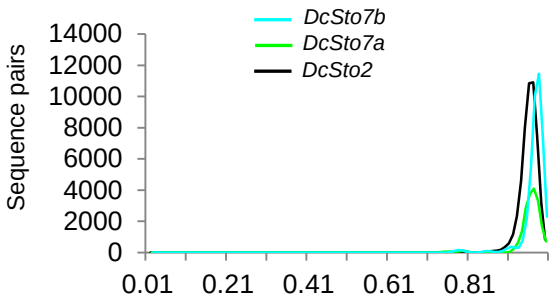
Supplementary Figure 7. MITE family analysis.

Panel a1-b1-c1: Histograms of unimodal (a1) bimodal (b1) and multimodal (c1) distributions of pairwise distances calculated using the Kimura 2-parameter mode among members of some MITE families. Sharp modal distributions suggest a rapid amplification burst rather than a gradual amplification.

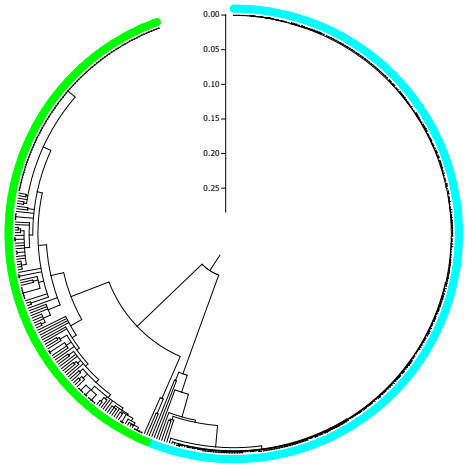
Panel a2-b2-c2: Timetree analysis by Maximum Likelihood method for some MITE families. The length of the branch indicates relative divergence time among elements represented in each phylogenetic tree.
Colors of *DcSto* families' histograms represented in the a1, b1, c1 panels correspond to family colors represented in the divergence time tree (Except for *DcSto*2 and *Dcsto*12 which were too divergent to be aligned with *DcSto*7a, *DcSto*7b, *DcSto*1 and *DcSto*5).

Nature Genetics: doi:10.1038/ng.3565

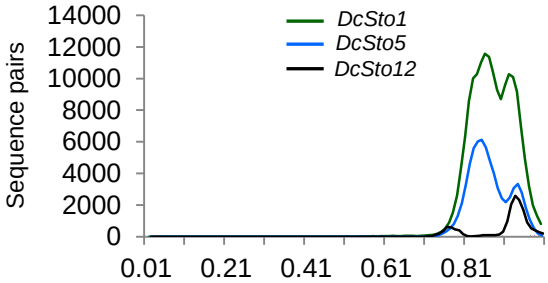
a1



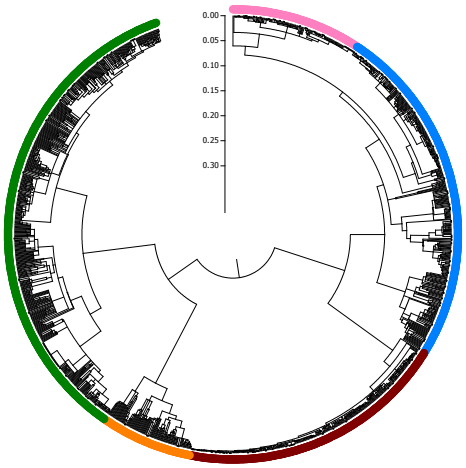
a2



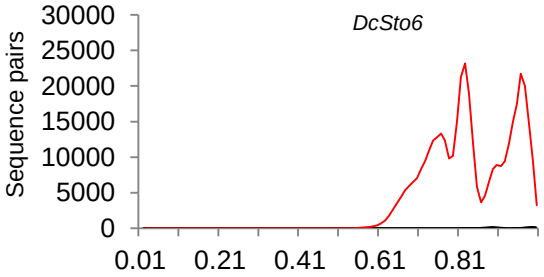
b1



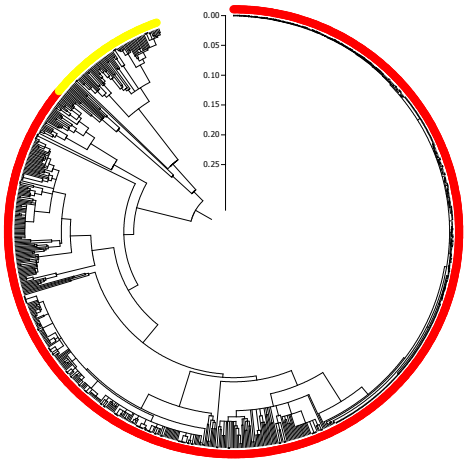
b2



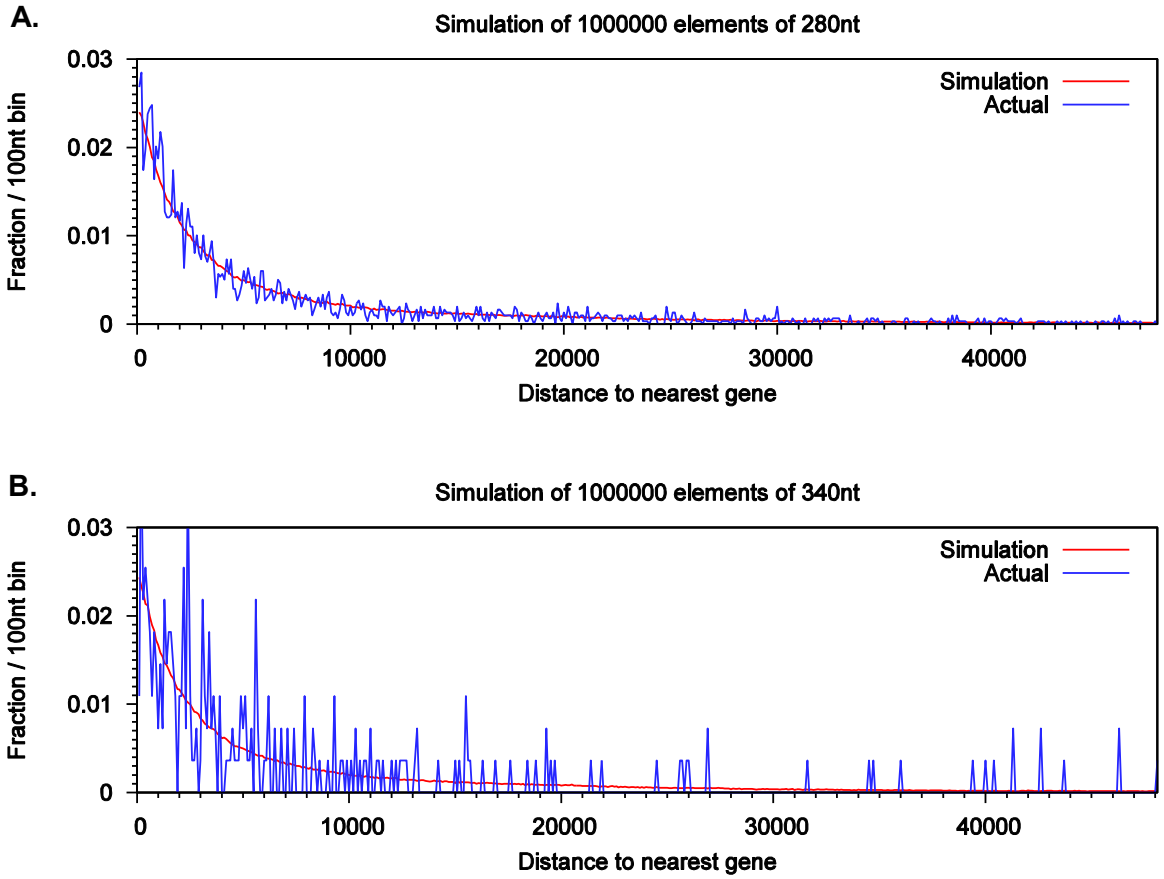
c1



c2



Supplementary Figure 8: Distribution of distance from MITE or *Krak* insert site to nearest predicted gene in carrot DH1. MITE *DcSto* (Panel A, blue line) and Krak (Panel B, blue line) elements, and simulated datasets (red lines). The simulated datasets represent the distribution of 1M random insertion sites of 280 or 340 nt, which in turn represent the average sizes of *DcSto* and *Krak* elements, respectively. The maximum interval plotted represents the point where the simulation reaches 95% of the elements that do not overlap genes.



Supplementary Figure 9: Evolution of tandem repeats in the carrot genome and *Daucus* genus. Panel I: Estimated genome proportion of Cent-Dc-like and CL80 repetitive sequences across six *Daucus* species included in this study. Panel II: Figures on the left illustrate the similarity and structure of the A-B-C-D Cent-Dc monomers in K11, DH1 and *D. syrticus* (Dsyr) organized in a higher order repeat structure (HOR). *D. aureus* (Daur) harbors a single 40 nt monomer most similar to monomer A. The panel on the right illustrates the sequences of the alignments of the 40 nt monomers. Arrows above each base indicates the private polymorphic sites that each individual A-B-C-D monomer has accumulated. Panel III: Phylogenetic analysis of the single ABCD monomers extracted from DH1 BAC end sequences. Panel IV: Schematic illustration of the hypothetical evolution of carrot Cent-Dc sequence.