Iorizzo, Massimo, Douglas A. Senalik, Shelby L. Ellison, Dariusz Grzebelus, Pablo F. Cavagnaro, Charlotte Allender, Johanne Brunet, David M. Spooner, Allen Van Deynze, and Philipp W. Simon. 2013. Genetic structure and domestication of carrot (*Daucus carota* subsp. *sativus*) (Apiaceae). *American Journal of Botany* 100(5): 939-950.

**Appendix S2.** Parameters used for analysis programs.

**A.** Parameters used for mapping Illumina read to the 59,493 carrot EST (2) consensus sequences with Mosaik:

|  |  |
| --- | --- |
| hash size (-hs) | 15 |
| number of mismatches allowed (-mm) | 4 |
| alignment mode (-m) | all |
| maximum hash positions per seed (-mhp) | 100 |
| alignment candidate threshold (-act) | 20 |

**B.** Parameters used for preliminary filtering with a custom Perl program:

|  |  |
| --- | --- |
| gigaBayes quality | ≥ 0.99 |
| distance from either end of sequences  | ≥ 20 nt |
| read coverage for most abundant allele | ≥ 15 (10)\* |
| read coverage for second most abundant allele | ≥ 15 (10) |
| amount of sequence on either side of the SNP in the output file | 100 nt |
| Isolation range | 100 nt |
| amount of sequence on either side of the SNP that cannot contain any other SNPs on one side | 30 nt (22 nt) |
| maximum number of other nearby SNPs within range of isolation range on one side | 10 (20) |
| amount of sequence on either side of the SNP that cannot contain any other SNPs on other side | 25 nt (3 nt) |
| maximum number of other nearby SNPs within range of isolation range on other side | 3 (20) |

**C.** Parameters used for megablast alignment of EST contig against whole genome assembly:

|  |  |
| --- | --- |
| low complexity filtering | off |
| e-value | ≤1 |
| amount of sequence flanking SNP on each side used for alignment (and blast hit must include SNP location) | 50 nt |

\*Parameters in parentheses were applied to develop the list of SNPs included in the high priority list, see table S6 (Priority 1 and 2).