A contiguous *de novo* genome assembly of sugar beet EL10 (*Beta vulgaris* L.).

Short title: EL10 sugar beet genome.

J. Mitchell (Mitch) McGrath[1*], Andrew Funk[2], Paul Galewski[2], Shujun Ou[2], Belinda Townsend[3], Karen Davenport[4], Hajnalka Daligault[4], Shannon Johnson[4], Joyce Lee[5], Alex Hastie[5], Aude Darracq[6], Glenda Willems[6], Steve Barnes[6], Ivan Liachko[7], Shawn Sullivan[7], Sergey Koren[8], Adam Phillippy[8], Jie Wang[9], Tiffany Liu[9], Jane Pulman[9], Kevin Childs[9], Anastasia Yocum[10], Damian Fermin[10], Effie Mutasa-Göttgens[11], Piergiorgio Stevanato[12], Kazunori Taguchi[13], Kevin Dorn[14*]

[1*]USDA-ARS, 1066 Bogue St., Michigan State Univ., East Lansing, MI USA 48824 (co-corresponding author: mitchmcg@msu.edu, ORCID ID 0000-0002-7365-9901)

[2]Plant Breeding, Genetics, and Biotechnology Program, Michigan State Univ., East Lansing, MI USA 48824.

[3]Department of Plant Sciences, Rothamsted Research, West Common, Harpenden, Hertfordshire AL5 2JQ, UK

[4]Los Alamos Nat'l Lab, Biosecurity and Public Health, Los Alamos, NM USA 87545

[5]BioNano Genomics, 9640 Towne Centre Drive, San Diego, CA USA 92121

[6]SESVANDERHAVE N.V., Industriepark Soldatenplein Zone 2 Nr 15, 3300 Tienen, Belgium

[7]Phase Genomics, 4000 Mason Road, Suite 225, Seattle, WA USA 98195

[8]Genome Informatics Section, Computational and Statistical Genomics Branch, National Human Genome Research Institute, Bethesda, Maryland, USA

[9]Center for Genomics-Enabled Plant Science, Plant Biology Department, Michigan State Univ., East Lansing, MI USA 48824

[10]A2IDEA, 674 S. Wagner Rd., Ann Arbor, MI USA 48103

[11]University of Hertfordshire, Hatfield, Hertfordshire AL10 9AB UK

[12]DAFNAE, University of Padova, Viale Università 16, 35020 Legnaro (PD), Italy

[13]Hokkaido Agricultural Research Center, National Agriculture and Food Research Organization, Shinsei Memuro, Hokkaido, 082-0081 Japan

[*14]USDA-ARS, 1701 Centre Ave, Crops Research Laboratory, Fort Collins, CO 80526. (co-corresponding author: kevin.dorn@usda.gov)

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Mitch McGrath (mitchmcg@msu.edu).

Key Words:

Genome assembly, genome annotation, *Beta vulgaris*, beet, genome size, read depth mapping, synteny, LTR, repetitive elements, comparative genomics

Abstract

A contiguous assembly of the inbred 'EL10' sugar beet (*Beta vulgaris* ssp. *vulgaris*) genome was constructed using PacBio long read sequencing, BioNano optical mapping, Hi-C scaffolding, and Illumina short read error correction. The EL10.1 assembly was 540 Mb, of which 96.7% was contained in nine chromosome-sized pseudomolecules with lengths from 52 to 65 Mb, and 31 contigs with a median size of 282 kb that remained unassembled. Gene annotation incorporating RNAseq data and curated sequences via the MAKER annotation pipeline generated 24,255 gene models. Results indicated that the EL10.1 genome assembly is a contiguous genome assembly highly congruent with the published sugar beet reference genome. Gross duplicate gene analyses of EL10.1 revealed little large-scale intra-genome duplication. Reduced gene copy number for well-annotated gene families relative to other core eudicots was observed, especially for transcription factors. Variation in genome size in *B. vulgaris* was investigated by flow cytometry among 50 individuals drawn from EL10 progeny and three unrelated germplasm accessions, producing estimates from 633 to 875 Mb/1C. Read depth mapping with short-read whole genome sequences from other sugar beet germplasm suggested that relatively few regions of the sugar beet genome appeared associated with high-copy number variation.

Introduction

Humans have used beet (*Beta vulgaris* spp. *vulgaris* L.) as early as the late Mesolithic, initially as leafy pot herb and for medicinal uses (Biancardi et al. 2012). It was not until the Middle Ages that the enlarged taproot was widely used as a vegetable. The origin of the enlarged taproot is not clear, but by the 18th century beets were widely used as fodder and fueled the prelude to the Industrial Revolution in Europe. Sugar beet was selected from lower sucrose fodder beets (6-8% sucrose fresh weight) from the late 1700's, with the first true sugar beet commercial varieties available by 1860 (Galon and Zallen 1998). Since then, improvements in sucrose content and processing quality have been continuous, resulting in an industry average in the US and Europe approaching 19% sucrose fresh weight (~75% dry weight). Breeding methods for sugar beet are applicable to the *B. vulgaris* vegetable crop types (table beet/ beet root and leafy chard) and fodder/ biofuel/ industrial chemical feedstock crop types (McGrath and Panella 2019, McGrath and Townsend 2015). Public sector sugar beet breeding today focuses generally on crop protection traits (Panella et al. 2008, 2015). The EL10.1 genome summarized here was recently interrogated for resistance gene signatures (Funk et al. 2018) and crop-type attributes (Galewski and McGrath 2020). An alternate assembly, EL10.2, is available but not as well characterized as EL10.1.

Beta vulgaris is a basal eudicot in the family Amaranthaceae (Caryophyllales) (Yang et al. 2015). Wild forms are found around European and Mediterranean coastlines and collectively classified as subspecies *maritima* (Biancardi et al. 2012, Andrello et al. 2016, 2017). There are no known barriers to cross-fertilization among beet crop and wild types, and the genomes of crop wild relatives are beginning to be described in detail (del Rio et al. 2019). Most *Beta vulgaris* types, and all characterized *maritima* types, are diploid. Chromosomes are morphologically similar at mitotic metaphase, and highly repetitive DNA sequences comprise ~60% or more of the beet genome (Flavell et al. 1974, Dohm 2014). Each chromosome shows different patterns of repeat-sequence distribution (Schmidt and Heslop-Harrison 1998, Paesold et al. 2012) supporting the notion that sugar beet genomes are true diploids (Halldén et al. 1998, Dohm et al. 2014). An ancient genome triplication appears to be shared with the basal asterid and rosid eudicot clades (Dohm et al. 2012). A uniform linkage group nomenclature was derived from Schondelmaier and Jung's (1997) linkage group assignments and made more portable with SSR markers (McGrath

et al. 2007). Extensive marker technologies remain proprietary within the commercial sugar beet breeding sector who supply hybrid seed to growers worldwide.

We seek to fill knowledge gaps in understanding of sugar beet traits by completing a genome framework for beet and then building crop genetic traits into the framework, focusing on crop quality and preservation traits. Creating highly contiguous genome assemblies is challenging, especially in plants due to the generally high-repetitive nature of portions of their genomes. Genome annotation is perhaps more challenging as expressed gene functions are generally predicted from relatively few physiologically-verified protein functions derived from unrelated plant taxa on the basis of nucleotide and amino acid sequence similarity. Improved approaches are becoming available and more commonly used (Jung et al. 2019). Many of these approaches were used in create the EL10 genome assemblies described here, including long-read length technologies which can span many (but not all) longer low-complexity repeat regions, optical mapping which can create larger scaffolds from long-read contig assemblies, and Hi-C which can link together scaffolds across the genome into chromosome-sized scaffolds. Highly contiguous assemblies exhibit the full organization of hereditary material, thus little uncertainty of position and distribution of genetic markers, for instance, allows closer focus on any region of the genome.

Scaffolds of the EL10 assemblies show high concordance with genetic maps and the RefBeet genome sequence (Dohm et al. 2014), which is an excellent but fragmented genome sequence assembled using first- and second-generation sequencing technologies. The current work from less fragmented assemblies used here provide a more comprehensive picture of genome size variability of sugar beet which surprisingly varies extensively each generation, and global changes in repeat sequence depth and coverage within and between sugar beet inbreds and breeding populations. Genome fluidity generates mutations, and assessing whether these recombinational mutation events are useful for sugar beet improvement, or simply a hinderance, can be investigated. Further, gene number in beet appears to be uniformly diminished relative to other eudicots, at least for gene classifiers that are shared among representative angiosperm genomes. Contiguous genome assemblies will allow routine inter-cultivar comparisons between accessions that vary for important traits, and thus help deduce casual from associated genomic features influencing a trait of interest, general performance, or otherwise suggest that an association is merely an historical coincidence of shared parentage and breeding.

Results

A five-generation inbred genome of the sugar beet 'C869' (PI 628755) was released as a genetic stock ‘EL10’ (PI 689015). C869 is the common seed parent for East Lansing recombinant inbred populations (McGrath et al. 2005). Five plants from one inbred family showing no gross phenotypic differences and no polymorphism among 24 selected unlinked SSR markers (McGrath et al. 2007) were chosen for nuclei isolation, long read sequencing, and assembly. The resulting assembly, using only one of the five plants, was scaffolded via opto-physical mapping, and the two assemblies described here share this common backbone. Hi-C scaffolding was largely able to reduce the number of scaffolds to the haploid chromosome number in beet (n=9), with assembly EL10.2 slightly improved in contiguity over assembly EL10.1. Insights described below pertain to EL10.1 since the annotation of EL10.2 is on-going.

Sequencing and Assembly: High molecular weight DNA was isolated from intact gel-embedded nuclei of true leaves from young seedlings and pooled from the five inbred plants for long-read sequencing using standard protocols for BAC library construction (Amplicon Express, Pullman, WA). Eighty-six PacBio SMRT cells yielded 79.3-fold coverage (58,655 Mb) of the *circa* 750 Mb *Beta vulgaris* genome size (see below). The Falcon Assembler (version 0.2.2) was used to assemble long reads (Table 1), initialized with reads exceeding 40 kb in length. The Falcon assembly resulted in 938 primary contigs, 70.9% with a length greater than 100,000 nucleotides and a total length of 562.76 Mb (Table 2). Both assembly versions (EL10.1 and EL10.2) relied on this intermediate assembly (e.g. SBJ_80X_BN, Table 2). G+C content was similar between EL10 and RefBeet contigs (35.8% vs. 36.1%, respectively).

Scaffolding the Falcon assembly with a BioNano two-enzyme (*Bsp*QI and *Bss*SI) sequential hybrid optical (physical) map resulted in substantial improvement. The *Bsp*QI optical map was generated from 141,462 molecules with an average length of 285 kb and labelled to an average density of 11.8 sites kb^{-1}, and the *Bss*SI optical map was generated from 270,071 molecules, also with an average length of 285 kb, labelled to a density of 7.7 sites kb^{-1}. Optical maps were aligned to PacBio Falcon contigs and the resulting *Bsp*QI and *Bss*SI map lengths were 628 Mb and 590 MB with N50 contig sizes of 1.99 Mb and 1.21 Mb, respectively. After merging PacBio,

*Bsp*QI, and *Bss*SI contigs, the final hybrid genome map consisted of 86 scaffolds with a total length of 566.8 Mb, and an N50 of 12.5 MB (Table 2).

Hi-C Proximity Guided Assembly, using selfed progeny of the individual that was optically mapped (see below), was applied to the merged PacBio/BioNano assembly (e.g. SBJ_80X_BN). This assembly was polished and gap-filled using a combination of approaches (PBJelly, Arrow, and Pilon; following Bickhart et al. 2017). The resulting 540.5 Mb assembly consisted of 9 chromosome-sized scaffolds, numbered via Butterfass chromosome nomenclature (Butterfass 1964), and 31 unscaffolded contigs. These comprise the genome assembly version described here, EL10.1. The 9 chromosome-sized scaffolds (designated Chromosomes below) were relatively similar in size (mean = 57.8 Mb, std. dev. = 3.9 Mb) (Table 2). Chromosomally-unscaffolded contigs (n=31, hereafter referred to as Scaffolds) represented 3.9% of the final EL10.1. A second assembly (EL10.2) was created using a second set of Hi-C reads and the default Dovetail Genomics HiRise assembly pipeline, starting from the PacBio-BioNano assembly SBJ_80X_BN (Table 2).

The EL10.1 genome assembly was not corrected for the differences described below, however EL10.1 has been used in at least two publications (Funk et al. 2018, Galewski and McGrath 2020), and therefore is important to document. Differences between assemblies have not been fully annotated and further improvement of the EL10 assembly is likely, thus, the EL10.2 assembly is being reported here simply as a publicly available resource for community inspection, assessment, and basis for refinement. Initial assessment suggested the EL10.2 assembly appeared to have resolved the major assembly-associated inversions on Chromosomes 7 and 9 (see below), as well as placed the 31 Scaffolds into the larger whole genome chromosome context (Figure 2), where many unplaced Scaffolds of EL10.1 appeared to be better placed within the context of Chromosome 5 in EL10.2.

Assessment: No complete chloroplast or mitochondrial genomes were incorporated into the EL10.1 assembly, although fragments of both plastid genomes were detected in the EL10.1 assembly. The position of RefBeet 1.2 scaffolds were determined for EL10.1 Chromosomes (Figure 1). Contigs > 5 kb in length were largely colinear between the two assemblies. Two small inverted-orientation contigs were evident on Chromosome 7, as were small inverted (e.g. Chromosome 6) and misplaced segments (e.g. Chromosomes 3 and 7). RefBeet 1.2 was

anchored with genetic markers (Dohm et al. 2014), and 345 of these with 100% match identity across 75 nt or greater were placed in concordant order on the EL10.1 assembly. In addition, 3,279 proprietary SNP markers from the SESVanderhave (Tienen, Belgium) molecular marker genetic map were placed to the EL10.1 assembly. Most marker orders were highly concordant. However, a third of the mapped markers were inverted on Chromosome 9, and a complex rearrangement involving 40% of markers was evident on Chromosome 7 (mapped inversions; Table 3). Genetic markers also added 9 Scaffolds to five Chromosomes (mapped integrations; Table 4). Genetic markers used to orient the cytogenetic map (Paesold et al. 2012) also aligned with the EL10.1 assembly. Chromosomes 1 and 3 were cytogenetically congruent with their North-South orientation, and the rest were reversed relative to the orientations given in that publication. Scaffold 5 was located to the South end of cytogenetic Chromosome 5 (Table 5), consistent with SESVanderhave marker data (Table 4).

Annotation: Of note, the EL10.1 assembly contained the entire first linkage group described in beet (Keller 1936), the *R-Y-B* linkage group on Chromosome 2. Each of these genes has been recently cloned (*R*, for the red alkaloid betalain synthesis by a novel cytochrome P450; Hatlestad et al. 2012), *Y*, a Myb transcription factor required for production of red color (Hatlestad et al. 2014), and *B* for the bolting gene which determines annual or biennial life habit; Pin et al. 2010, 2012). Both the direction and the distance agree with published genetic map intervals, and the EL10.1 assembly indicates that the bolting gene is physically located proximal towards the centromere and the color genes are more distal (Table 6).

Results from the MAKER annotation pipeline (Holt and Yandell 2011) conservatively predicted 24,255 proteins, numerically 88.5% of the 27,421 predicted in RefBeet (Dohm et al. 2014). For functional annotation, three sources were used, in the priority: 1) UniProt, 2) Pfam-A, and 3) Uniref90. If no functional annotation was found in these three highly curated sets, predicted proteins were assigned to the class of ‘hypothetical' proteins. Gene model completeness was checked using BUSCO (Table 7) (Simão et al. 2015). A higher proportion of missing BUSCO's was seen in EL10.1 than either RefBeet 1.1 or Arabidopsis. Overall, protein coding gene predictions covered a relatively small proportion of the assembled EL10.1 genome (39,161,207 nt; 7.2%). GC content of predicted coding genes was marginally higher than that of the whole genome (41.1% vs. 35.8%, respectively). Predicted proteins were named using the underlined characters in the key: <u>EL10</u> / annotation version <u>A</u> / <u>c</u>hromosome or <u>s</u>caffold number

/ genomic in origin / a sequential number / and appended with .1 to signify that only one isoform was considered at this level of analysis (e.g. EL10Ac7g16740.1).

The number of MAKER annotations ascribed across Chromosomes of EL10.1 was relatively consistent (mean = 2,559, stdev = 173.8), but highly variable between Scaffolds (mean = 44.0, stdev = 47.6) (excluding Scaffolds 23, 29, 30, and 31 for which no gene models were predicted) (Table 8). A total of 3,940 gene models had no functional annotation among curated comparative databases (and thus were designated hypothetical), and these were also evenly distributed among Chromosomes but not necessarily Scaffolds (Table 8). Fewer than 55% of gene models were considered unique in the sense their curated-database annotations only occurred once in the list of gene models (Table 8), and thus, at this level of analysis, more than 45% of predicted genes could be members of gene families.

Self-synteny of MAKER gene models with the EL10.1 genome sequence was explored using the CoGe SynMap platform (Lyons et al. 2008). Few internal syntenies were detected. Mean copy number of the 2,327 discovered tandem gene models was 2.82 (stdev = 1.96), and 65.8% of these tandem duplications were two copies. For syntenic regions with at least 5 matches in a span of 20 gene models (encompassing 1,858 genes in 268 synteny blocks), average Kn/Ks values were all less than 1, suggesting stabilizing selection for genes in these blocks. For individual gene pairs, only five gene pairs had Kn/Ks values >1 (suggesting diversifying selection) but only two of these pairs had interpretable annotations. EL10Ac6g14284.1 & EL10Ac9g20883.1 were predicted as Clathrin heavy chain 2 genes (i.e., vesicle trafficking) and EL10Ac1g01568.1 & EL10Ac5g12109.1 were predicted SET Domain Protein genes (i.e., chromatin structure modulation).

A comparative gene annotation perspective was gained using the MapMan4 ontology of plant proteomes (Schwacke et al. 2019). EL10.1 MAKER gene models were placed in 99.6% of 4,145 ontologies assigned to one of 28 'bins' (infrequently allowing for assignment to more than one bin), organized in a hierarchal, conceptual, plant-specific context (e.g., Photosynthesis, Cell cycle, Hormones, etc.). Where possible, each bin resolves to a gene from a high-quality genome assembly in the Mercator4 web implementation of MapMan4. Specific comparisons for each of the 4,127 EL10.1 occupied terminal, termed 'leaf', bins were made with five other angiosperms (e.g., *Arabidopsis thaliana*, *Oryza sativa*, *Brachypodium distachyon*, *Solanum lycopersicum*, and

Manihot esculenta). Most EL10.1 predicted proteins in the found set were placed in one (or more) MapMan4 leaf bins (Table 9). Since the MapMan4 ontology is hierarchal, the number of genes in each leaf bin was averaged for all five angiosperms, and compared to EL10.1. Surprisingly, the number of genes in the EL10.1 gene set was 69% that of the average of five angiosperms (Table 9).

Enrichment analysis can shed light on biological processes that may have assumed greater or lesser importance in the evolutionary success of a lineage. Given the general reduced gene copy number in EL10.1, genes whose copy number equaled or exceeded the mean of five angiosperms were tentatively considered as enriched, and those that were substantially lower than the overall mean of EL10.1 were considered as reduced. EL10.1 appeared particularly depauperate in at least two top-level ontologies; Cellular respiration (Bincode 2) and Phytohormones (Bincode 11) (Table 9). Equal or over-represented ontologies included DNA Damage Response (Bincode 14) and Coenzyme metabolism (Bincode 7) (Table 9).

Proteome content of the five averaged angiosperms relative to EL10.1 was gauged for missing members, which could suggest regions in EL10.1 that were not assembled, genes that were not annotated, or perhaps reflect biological divergence or biochemical alternatives that beet followed during its evolution. Not detected in EL10.1 were 154 genes that were present in at least one copy in each of the five angiosperms. Missing annotations were assignable across all 28 top-level bins, with the exception of Bincode 8 (Polyamine metabolism) (Table 10). In this set, mean copy number was low (1.6 genes per leaf bin) in the five taxa, and failure to assemble or annotate a low-copy number genes in EL10.1 was possible. However, in 12 cases, each of the five other plants had small gene families (mean copy number = 3.7 genes per family) but no EL10.1 homologue was annotated, which seemed less probable that all would have been missed during assembly and annotation, thus their functions in beet may have been dispensable, their genes diverged, or their functions assumed by other genes.

Remaining annotations were surveyed for potential biological interest, but not exhaustively evaluated (Table 11). Under-represented genes in 'Cell wall' (Bincode 21) included those involved with hemicellulose, lignin, cutin, and suberin metabolism, as might be expected from selection for a mechanically-sliced root crop for sucrose extraction (e.g., less knife wear during processing, which is a trait that has not necessarily been under conscious selection).

Phytohormone representation was low across all second-level categories, especially salicylic acid (Bincode 11.8). External stimuli response (Bincode 26) was rich in drought response but poor in biotic stress response genes. Multi-process regulation (e.g., integration of development with response-to-environment) was over-represented by the TOR signaling pathway (Bincode 27.2) and under-represented in the SnRK1 metabolic regulator system (Bincode 27.3). RNA biosynthesis (Bincode 15) was generally over-represented, however Bincode 15.8 (transcriptional repression) was greatly under-represented. Overall, 138 leaf bins were similar or over-represented and 447 were under-represented in EL10.1.

Transcription factor genes (Bincode 15.7) were under-represented overall. On average, there were ~10 fewer genes in EL10.1 than the average of five other angiosperms. Transcription factor genes with a >50 gene deficiency between the angiosperm average and EL10.1 included MADS box, NAC , MYB, and bHLH transcription factors (Table 12). Most of the transcription factor classes showing larger deficiency in copy number were members of larger gene families. Few transcription factor classes were equally- or over-represented, and most of these were from gene families characterized by lower copy number (Table 12). However, the FAR1 transcription factor class was abundant in EL10.1, and highly variable in the group of five other angiosperms (Table 12). It is likely that each of these differences in transcription factor copy number has potential to impact plant phenotype, development, and/ or response to the environment.

Genome size: Reported genome sizes (714 - 758 Mb; Arumuganathan and Earle 1991, derived from estimates for one plant each of table and sugar beet, respectively) and assembled genome sizes of sugar beet (~540.5 - 566.6 Mb, Table 2) may be explained by failure to assemble repetitive sequence arrays completely. To better assess genome size as a gauge of the completeness of assemblies in *Beta vulgaris*, an additional 50 independent cytometrically-determined nuclear DNA content estimates were obtained from four unrelated germplasm accessions; two traditional out-crossing progenies and two from progeny of deeply inbred accessions of EL10 and an inbred table beet derived from germplasm 'W357B'. Nuclear DNA content estimates of these materials ranged from 633 Mb to 875 Mb, as estimated from at least four biological replicates from each accession (at least 20 from inbreds) with four technical replicates performed per biological replicate (Table 13). Overall, genome size between crop types was not statistically different (sugar beet, n = 120, mean = 729.0 Mb/1C, std. dev. = 51.2; table beet, n=80, mean = 742.3, std. dev. = 52.8; p = 0.079). Average genome size differences of

each sugar beet accession were significantly different from one another ($p < 0.001$, means and dispersion values are presented in Table 13), and only the difference between sugar beet '5B sugar breeding population' and Inbred Table beet was not significantly different than the other two sugar beet accessions. Inbreds showed a statistically-significant smaller average genome size (Table 13: inbreds, mean = 728.5 Mb/1C, out crossed, mean = 764.9 Mb/1C, $p = 0.0002$), and at least 2-fold higher variation than out-crossers (Table 13). The average cytometrically-determined genome size of all tested accessions was 734.3 Mb (stdev = 50.3 Mb). The smallest cytologically-estimated DNA content (633 Mb), coincidentally present in the progeny of EL10, closely approximated EL10's optical map length of 628 Mb, and curiously, the average genome size of EL10's progeny was 88 Mb larger than the assembled EL10.1 genome. Thus, average genome size appeared to increase over a single generation of selfing, and to an extent that reflected the size range observed within the species. This also implied that genome size also decreased at some point during the generation of these materials.

Repetitive element content estimation: Plant genomes are characterized by high repetitive sequence content, found either as tandem arrays or as multiple copies distributed throughout the genome (Bennetzen and Wang 2014). More than 180,000 named repetitive elements (as deduced by RepeatMasker) were placed on the EL10.1 assembly (Table 14). DNA class transposable elements were the most frequent (58.1%), which appears to be at odds with RefBeet (Dohm et al. 2014), and LTR elements the next most frequent class (36.0%) of annotated transposable elements (Table 7A). Numbers and types of LTR elements were estimated similarly using RepeatMasker and LTR_Retriever (Ou and Jiang 2018). However, distribution of the filtered high-confidence intact LTR_Retriever-predicted Gypsy and Copia elements (Table 14B) showed Copia elements generally more frequent towards the ends of Chromosomes and Gypsy elements biased towards centromeric regions (Figure 4).

Repeats associated with centromeric histone variants have been characterized in beets (Kowar et al. 2016), and these consist of the Gypsy element Beetle7 as well the pBV class of major satellites (Table 14C). High-similarity Beetle7 sequences (90% identity over 1,000 nt or better) were located on all Chromosomes and eight of the Scaffolds. The 35S and 5S ribosomal RNA genes are also tandemly arrayed in beets (Paesold et al. 2012). The 35S arrays in EL10.1 were localized to Chromosome 2, as expected, and also to Scaffolds 7 and 19. The 5S array localized to Chromosome 4, also as expected, and to Scaffold 11. Only one canonical plant telomere array

$(TTTAGGG)_n$ greater than three tandem copies was found in the EL10.1 assembly, near the end of Scaffold 5. However, terminal repeat arrays defined by the major satellite class pAV (Dechyeva and Schmidt 2006) were found near the ends of most Chromosomes, except at one end each of Chromosomes 1, 5, 7, and 9 (Table 14D). pAV arrays were seen on each of these except Chromosome 1, where the South terminus appeared absent. Evidence suggests Chromosome 5 South is Scaffold 5, Chromosome 9 may have a pericentric inversion or an assembly artifact that misplaced Chromosome 9 South, and complex inversions in Chromosome 7 may have failed to accurately assemble the North terminal repeat region (these appeared to have been resolved in the EL10.2 assembly). Notably, interstitial pAV arrays were evident in both Chromosomes 5 and 7 (Table 14D).

Tandem repeats (unit length 500 nt or less assessed with Tandem Repeat Finder) were evenly spread across the EL10.1 assembly (Table 15), with an average of 630.4 repeats Mb^{-1} (stdev = 19.3) across Chromosomes, and similar for Scaffolds but with 25-fold higher variation (Mean 661.0 repeats Mb^{-1}, stdev = 460.8). Shorter repeats were more frequent, and the most frequent size class was 21 nt (23,163 instances). Size classes of tandem repeats may reflect the predominant repeat unit size for centromeric sequence in a species (Melters et al. 2013), and for EL10.1, the most frequent repeat size above 100 nt was 160 nt (781 copies), followed by 170 nt (382 copies) (Figure 5). Relatively high numbers of repeats (67-134 copies) in the 314-325 nt repeat unit size range were evident, as might be consistent with a heterodimeric model of centromere repeats (Melters et al. 2013).

Assembly continuity was accessed using the LTR Assembly Index (LAI) (Ou and Jiang 2018). After adjusting for the amplification time of LTR-RTs, the whole-genome LAI of the EL10 assembly was estimated to be 13.3, which is considered reference quality and improves upon the RefBeet assembly (LAI = 6.7) (Figure 6). Thus, the EL10.1 sugar beet genome assembly appeared to be largely complete with respect to repetitive element landmarks and assembled in a largely congruent fashion with respect to genetic markers.

Read count mapping: Read depth variation provided a means to compare accessions using readily available and deeper coverage short reads. Low variation in read depth suggests relatively even distribution of coverage across assembly coordinates, while higher variation suggests regions of low sequence complexity that may not have assembled in a consistent fashion,

perhaps contributing to differences in genome size between cytometry and assembly estimates. Five independent Illumina-derived read sets were read mapped to the EL10.1 genome assembly, one from EL10 and one each from four other sugar beet germplasms (including two EL10 relatives and two unrelated germplasms). Overall, more than 99.6% of EL10's cleaned reads mapped to the EL10.1 assembly, with relatively even coverage (e.g. ~ 36 reads per assembled nucleotide), but Scaffold coverage was slightly less and the standard deviation was 22-fold higher. Similar results were evident in the other four germplasms (Table 16). There appeared no 'degree-of-relatedness' discrimination between disparate germplasm at this level of analysis, as EL10 relatives showed as much overall difference in read-depth variation as individuals drawn from unrelated populations (Figure 7).

High read-depth locations were localized using a conservative, computationally facile, and relatively crude sequence-independent approach. High read-depth locations were defined as a region of 5 kb with average per-base read mapping depth above 500 in one or more of the five tested germplasms (indexed from the lower nucleotide position of the EL10.1 assembly, Figure 8). This binning approach is conservative in the sense that most highly repetitive elements are shorter than the 5 kb window size used, but provided a computational advantage for an initial assessment whether changes in genome size could crudely be restricted to specific genomic bins, or were otherwise more or less independently distributed across the genome. The difference between C869_25 (i.e., the base genotype for EL10 and C869_UK) and each other accession flagged 47 such bins along Chromosome 1. Each flagged bin in each of the five germplasms occurred predominantly in the same places on Chromosome 1 (Figure 8). Most of these bins were occupied by Gypsy or Copia LTRs, however Bin 44,615,000 was occupied by chloroplast sequence (Sequence ID: KR230391.1) and Bins 8,100,000, 22,360,000, and 22,365,000 were occupied with mitochondrial sequences (Sequence ID: FP885845.1). It is not unusual to find plastid sequences within plant genomes (Pichersky et al. 1991), and plastid sequence read-depths are likely subject to external influences (e.g. plant growth and DNA isolation methods). The large differences in the remaining read-depth estimates at specific sites suggests that copy number changed since a last common ancestor. In the case of C869_UK and EL10, two generations of selfing had elapsed, and here differences were localized to 8 of the 47 bins on Chromosome 1 (Figure 8). These sites have the potential to contribute to intra-specific genome size variation. Further evaluation of such sites across the genome in a more precise sequence-

specific fashion (e.g., not binned) may help deduce special features related to their lability and whether changes in genome size at this level of resolution have phenotypic effects.

Broader synteny: Caryophyllales members spinach (*Spinacea oleracea*), grain amaranth (*Amaranthus hypochondriacus*), and quinoa (*Chenopodium quinoa*) have annotated genome assemblies that were used to compare with EL10.1 (Yang et al. 2016, Lightfoot et al. 2017, Jarvis et al. 2017, respectively; note that quinoa and amaranth are each amphidiploid). Chromosome 4 synteny appeared maintained in chromosome-sized blocks among Caryophyllales, as well as *Vitis vinefera* to a lesser extent, but not *Arabidopsis thaliana*, as outgroup representatives of the Rosids (Figure 9). Chromosome 1 synteny also appeared relatively conserved in chromosome-sized blocks among the Caryophyllales, with the exception of the spinach assembly version used here, which will likely improve in the future. Elements of Chromosomes 2, 6, and 9 were found in extended blocks in quinoa and amaranth, but also not spinach. Extended synteny for Chromosomes 5 and 8 were evident in quinoa but were not as extended in amaranth, while extended blocks for Chromosomes 3 and 7 were present in amaranth but not as well maintained in quinoa. Genome evolution within the Caryophyllales produced significant genomic variation in chromosome number, number of syntenic regions, and size of syntenic regions relative to beet (Table 17).

Discussion

A high-quality *de novo* assembly of the sugar beet genome was created. The EL10.1 assembly contains most of the 'EL10' genome organized into nine linkage groups plus 31 extra unplaced scaffolds. Most scaffolds contain predicted genes, and many were able to be placed in context of the larger chromosome-sized assemblies using genetic markers. Ends of chromosomes were captured to some degree, however additional work will be required to finish the EL10 genome assembly to exacting standards. Efforts to this end are underway and it should be noted that the EL10.2 assembly appears to resolve at least the major assembly-induced inversions evident of Chromosomes 7 and 9. Genome assembly is fraught with uncertainty. In most cases, there is no *a priori* information to gauge the completeness and correctness of an assembly. In this case, the fortuitous availability of a published sugar beet assembly (Dohm et al. 2014) allows for comparisons, however EL10 provides an independent perspective on the organization of the beet

genome. The major difference between the two studies is that the sequencing technologies have improved to provide longer range scaffolding resulting in a substantially improved contiguity of sugar beet genome assembly. Such improvements also presumably better reflect copy number variations.

Long reads alone are currently insufficient for a high-quality assembly of a plant genome of moderate to large genome size. Beet might be considered a moderately-sized plant genome. The addition of an opto-physical map to long reads alone provided a ~10-fold reduction in the number of contigs, as well as set an upper bound for the size of the sequenced EL10 genome assembly (628 Mb). However, this also was insufficient to achieve chromosome-level assembly. Further addition of Hi-C data, where intact nuclei are cross-linked *in vivo* and where the native genome organization is presumably preserved, provided the means to map chromosome level associations. The sequential application of at least four independent technologies (i.e., short- and long-read sequencing, physical / optical maps, Hi-C chromatin conformation capture, and genetic maps) seems to have overcome many limitations spanning low-complexity regions of a genome over previous technologies in creating contiguous *de novo* genome assemblies of moderate to large plant genomes.

A reduction in gene copy number in beet (relative to annotated protein genes generated for comparative purposes, e.g., MapMan4) was observed. No clear evidence of gene copy number amplification was seen among the EL10.1 predicted protein set. Clear reductions were observed for transcription factors in particular, also observed by Dohm et al. (2014). Exceptions to the transcription factor reduction observed in EL10.1 included the FAR1 class of transcription factors, which may be anciently-derived from Mutator-like transposons and coopted in Arabidopsis for red-light perception and signaling (Hosoda et al. 2002, Mason et al. 2005). The role for this class of sequences remains unknown in beets, and copy number variation was high for FAR1 between the five other angiosperms considered. Discounting the occasional exception to lower overall gene copy number in beets, it may be suggestive of a basal gene copy number in dicots where beet numbers (or Caryophyllids in general) approximate a baseline condition, and other dicots have increased their copy numbers, as opposed to beets missing copies.

Genome size estimates of the cultivated beets examined here were quite variable, ranging from 633.0 to 875.5 MB per haploid genome. Genome size estimates of 21 wild *Beta vulgaris* spp.

maritima genotypes from Portugal ranged from 660.1 to 753.1 MB (Castro et al. 2013), thus variability in genome size is known to occur in the species. The range of estimates was 2.6 times higher in the cultivated beets relative to the wild types. This was also observed relative to the breeding system of the cultigens, where the range in genome size among the out-crossers was 2.7 times lower than that of the inbreds (e.g., Table 13). Inbreeding *per se* may have effects on genome size in maize, including substantial loss of chromatin (Roessler et al. 2019). High variability in genome size between generations may be expected if copy-number variations were generated at each generation, as suggested by changes in genome size and read depth variation at specific loci. When during the life cycle such changes occur is not known, but presumably during a phase associated with DNA replication, and it is likely that transposable elements are at least partially involved (Whitney et al. 2010).

Variation in read-depth coverage may be useful for tracking genome size changes (Pucker 2019). Areas of high variation are intriguing from a chromosome biology and evolution perspective, as well as their potential effect on phenotype and on the origin of novel variation. It is no surprise that many plant genomes are large because of their highly repetitive nature, and many classes of repetitive elements are known to vary across kingdoms, often with little in common other than size, the fact they are repetitive, and characteristic footprints (target site duplications, terminal repeats, etc.) (Bennetzen and Wang 2014). Speciation seems to favor whole-scale sequence replacement of repeat elements while retaining their size, however inter-specific amplified repeats seem to be present at low copy number in related genomes (Schmidt et al. 1991). Exactly how, and in particular when and what effects the efficiency, distribution, and specificity of divergent repeat amplification, is not as easy to investigate. Genome size reduction occurs rarely in plants, and improved read depth mapping approaches may be helpful in identifying additional examples and underlying mechanisms.

Two related, and two unrelated, germplasms contributed short-reads to the read depth differences observed, and are not necessarily representative breeding populations but rather have been crafted for genetic analyses. Beet is naturally a wind-pollinated out-crosser, which means that genetic diversity is partitioned within populations rather than between populations, as for inbreds. Each of the germplasms examined here, with the exception of C869_25, is highly inbred, using one of three different breeding methods. Both C869_UK and EL10 were derived from C869_25 through single seed descent, for three and five generations, respectively. RefBeet

(aka KWS2320) was derived as a doubled haploid, and NK-388mm-O is a seed parent for hybrids inbred through conventional sib-mating (Taguchi 2014, Taguchi et al. 2019). The method used to generate the inbred seems not to relate to generation of read depth differences. Relatives of C869_25 showed as much difference in copy number as did the unrelated germplasm. However, each germplasm had a set of events specific to their own lineage, and others that were shared among two, three, or all germplasms. For instance, NK-388mm-O was enriched in depth at EL10.1 Chromosome 1 positions 53,310,000 Mb to 53,325,000 Mb, KWS2320 depauperate at positions 40,540,000 to 44,615,000, and C869_25 over represented from 22,735,000 to 22,7750,000. Responsible sequences underlying these regions have not yet been investigated, except where wide differences in chloroplast content, and less so mitochondria, were particularly rich in NK-388mm-O. We recognize the speculative nature of these interpretations, but they do generate testable hypotheses in a difficult to access arena of chromosome biology.

Exploration of synteny between species is accessible from a contiguous well-annotated genome sequence. For EL10.1, annotations were conservatively estimated from well curated plant gene resources, which likely improved confidence in assessing similarity between well known plant genes. Following the syntenic organization of such genes across phylogenetic groups showed that closely related species retained higher levels of synteny than more distantly related species, as expected. Also expected, was that recombination and schism of synteny blocks increased with increasing phylogenetic distance. Perhaps unexpected was differential synteny conservation by individual chromosomes. However, relatively few plant genomes are available that are highly contiguous, and this caveat limits interpreting results (for instance, the spinach genome is still under construction).

Methods

Plant material: USDA-ARS germplasm release EL10 (PI 689015) was derived by single seed descent from C869 (PI 628755) by self-pollinating over six generations. C869 is a biennial sugar beet conditioned by the self-fertile (S^fS^f) allele and is segregating for nuclear male sterility (*Aa*), with resistance to several diseases (Lewellen 2004). The initial selfing occurred from one self-fertile C869 CMS plant (EL-A013483) in 2002. Seed was field grown at the Bean and Beet Farm

(Saginaw, MI) in 2005, roots were harvested, potted into fiber pots (5 L, Stock # ITMLFNP08090RBRD040TW, BFG Supply, Burton, OH), vernalized for 16 weeks, and grown in the greenhouse until flowering. Flowers were inspected for visible pollen, and when present, a #16 white grocery bag (Duro Bag, Novolex, Hartsville, SC) was placed over the bolting stem to effect self-pollination. Seed harvested from a single plant (EL-A018880) was considered the S1 generation, and subsequent generations were derived by single seed descent using field grown mother roots and selfing with the same methods. The S2 generation (EL-A022144) was obtained in 2007, and the S3 (EL-A025943) in 2010. Nine individuals of this population were genotyped with 69 SESVanderhave proprietary SNP markers evenly spaced across the beet linkage map, and a single homozygous individual (#17) of this population was sequenced for a preliminary assembly (named C869_UK here, McGrath et al. 2013). A sibling of this line (EL-A026195) with good field performance in the 2011 Michigan field (Saginaw Valley Research and Extension Center, SVREC, Richville, MI) was selfed in the same manner to yield the S4, while S5 (EL-A13-03870) and S6 generations were produced solely under greenhouse conditions in 2013 and 2015, respectively. Sixteen S6 individuals were genotyped with 24 SSR markers (McGrath et al. 2007), and six individuals (EL-A15-01096, EL-A15-01098, EL-A15-01099, EL-A15-01101, EL-A15-01102, and EL-A15-01103) were chosen as sequencing candidates based on marker homozygosity and similar growth habit and appearance, and pooled for long-read sequencing. One of these (EL-A15-01101) provided the sole tissue source for Illumina sequencing and nuclear DNA content estimation, and seed was named and released as EL10. Seed of EL10 was increased and deposited in the National Plant Germplasm System repository as a genetic stock (PI 689015).

Additional taxa were used, depending on the availability of materials. For the assessment of genome size, cytometric estimates were obtained from progeny of EL-A15-01101 whose genome was assembled here, advanced progeny of table beet W357B (a self-fertile parental line graciously provide by Dr. Irwin Goldman) which was inbred by single seed descent for five generations (accession EL-A1400766), an East Lansing open-pollinated self-sterile sugar beet breeding population (termed "5B"), and an open-pollinated USDA-ARS release used for a disease nursery check entry (F1042, PI 674103).

Genome sequencing, assembly, and finishing: High molecular weight DNA for PacBio sequencing isolated nuclei using the HMW preparation protocols suitable for BAC library

construction by Amplicon Express (Pullman, WA). PacBio RSII sequencing was performed at the Los Alamos National Laboratory (Los Alamos, NM), in 86 single-molecule, real-time cells using P6-C4 chemistry. PacBio reads greater than 6 kb were assembled with the Falcon Assembler (version 0.2.2), resulting in 938 primary contigs. Optical mapping was performed using the BioNano Irys sequential hybrid protocol with enzymes *Bss*SI and *Bsp*QI. For the EL10.1 assembly, scaffolding was accomplished using Proximity Guided Assembly (PGA) and Hi-C reads by Phase Genomics (Seattle, WA). Resulting scaffolds were polished and gap-filled using PBJelly, Arrow, and Pilon, following Bickhart et al. (2017). Briefly, PBJelly from PBSuite v15.8.24 was run using the Protocol.xml (https://gembox.cbcb.umd.edu/shared/Protocol.xml) with default parameters and minimum gap size set to 3 as: Jelly.py setup Protocol.xml --minGap=3, Jelly.py mapping Protocol.xml, Jelly.py support Protocol.xml, Jelly.py extraction Protocol.xml, Jelly.py assembly Protocol.xml, and finally Jelly.py output Protocol.xml. Pilon v1.13 was run using --fix local bases and the is pipeline at: https://github.com/skoren/PilonGrid. Arrow v2.0.0 was run using the pipeline available at: https://github.com/skoren/ArrowGrid. And, Pilon v1.21 was run using --fix indels using the pipeline at: https://github.com/skoren/PilonGrid. For the EL10.2 assembly, 462 million Hi-C read pairs were input into the SBJ_80X_BN assembly (Table 2) and assembled via Chicago and Dovetail Hi-C technologies using the HiRise algorithm as described (Meyer and Kircher 2010, Putman et al. 2015). Bioinformatic manipulations during sequential assembly steps were performed by the respective organizations, and the 'best' assembly was then used as input for the next assembly step. Assembly metrics were assessed using assemblathon_stats.pl with default parameters (github.com/KorfLab/Assemblathon) (Earl et al. 2011).

Whole-genome alignment: Whole-genome alignment of the EL10.1 assembly (as reference) and the RefBeet-1.2 assembly (as query) was conducted using modules from MUMmer v.4.0.0beta2. Initial alignments were created with the nucmer module, with options --mum --minmatch 30 (uses only anchor matches that are unique in both the reference and the query, and sets the minimum length of a single exact match to 30 bp). The resulting delta alignment was filtered using the delta-filter module with options -1 -i 70 -l 5000 (to use only 1-to-1 alignments, with a minimum 70% sequence identity, and minimum alignment length of 5,000 bp). Summary reports were created using dnadiff, and plots were created from the filtered delta file using

mummerplot with options --png --fat -r (with output image as png, and using layout sequences using fattest alignment only).

Annotation: The EL10.1 assembly was annotated using the MAKER pipeline (Holt and Yandell 2011). The EL10.2 assembly has not been annotated. A custom repeat library for EL10 was created and used for repeat masking (Campbell et al. 2014). Protein and transcript evidence were used to aid gene model prediction. Protein evidence was obtained from the following species or databases: *Arabidopsis thaliana* proteins from Araport11 (Cheng et al. 2017), *Solanum lycopersicum* proteins from IPTG 2.4 (Fernandez-Pozo et al. 2015), *Populus trichocarpa* proteins from Phytozome genome v3.0 (Tuskan et al. 2006), and curated plant proteins from UniProt release 2017_03 (The UniProt Consortium 2017). Transcript evidence was derived from 25 RNA-seq read sets (BioProject PRJNA450098, Illumina 2500, 150 bp paired-end) using StringTie v1.3.3b (Pertea et al. 2015) and TransDecoder v5.0.1 (Haas & Papanicolaou et al., manuscript in prep. http://transdecoder.github.io).

Gene prediction programs AUGUSTUS (Stanke and Waack 2003) and SNAP (Korf 2004) were trained using the transcript sequences generated by StringTie (above), and both AUGUSTUS and SNAP were used to predicted gene models within the MAKER pipeline (Holt and Yandell 2011). When AUGUSTUS and SNAP predicted genes at the same locus, MAKER chose the gene model that was the most concordant with the transcript and protein evidence, and that model was retained at that locus. HMMER v 3.1 (Finn et al. 2011) was used to determine the presence of Pfam-A protein domains in the initial predicted protein sequences. Gene models supported either by protein or transcript evidence or by the presence of a Pfam domain were collected as high-quality gene models for the final genome annotation. Both transcript and protein sequences were searched against the SwissProt and UniRef databases using BLAST (Altschul et al. 1990). HMMER v3.1 (Finn et al. 2011) identified PfamA domains within predicted protein sequences. Signal peptide and transmembrane domains were predicted using SignalP v4.1 (Petersen et al. 2011) and TMHMM v2.0 (Krogh et al. 2001), respectively. Searches and predicted results were parsed and combined in the final functional annotation.

The online sequence functional classification and annotation tool Mercator4 ver. 2.0 (Schwacke et al. 2019) was supplied with the EL10.1 MAKER predicted protein fasta file using default settings. Four gene models were excluded from analysis due to their short length (<5 amino

acids) (e.g. EL10Ac2g04429.1, EL10Ac8g20093.1, EL10Ac1g00658.1, EL10Ac7g16947.1). Comparisons were made with Mercator4-supplied representatives of the Tracheophyta (i.e. *Oryza sativa, Brachypodium distachyon, Arabidopsis thaliana, Solanum lycopersicum, and Manihot esculenta*).

LTR annotation: *De novo* identification of intact LTR retrotransposons were performed using LTR_Retriever v1.6 with default parameters (Ou and Jiang 2018). The insertion time of each intact LTR-RT is estimated by LTR_retriever based on $T = K/2\mu$ where K is the divergence between an LTR pair and μ is the mutation rate of 1.3×10^{-8} per bp per year. Whole-genome LTR sequence annotations were achieved using the non-redundant LTR library generated by LTR_Retriever and RepeatMasker v4.0.0 (www.repeatmasker.org).

LAI estimation: The assembly continuity of repeat space is accessed using the LTR Assembly Index (LAI) deployed in the LTR_retriever package (v1.6) (Ou and Jiang 2018). LAI was calculated based on either 3 Mb sliding windows or the whole assembly using raw_LAI = (Intact LTR-RT length * 100)/Total LTR-RT length. For the sliding window estimation, a step of 300 Kb was used (-step 300000 -window 3000000). The estimation of LAI was adjusted using the mean identity of LTR sequences in the genome based on all-versus-all BLAST.

Tandem Repeats: Tandem Repeats Finder Program Version 4.09 was used to characterized tandemly dupicated sequences. using the default Alignment Parameters (e.g. match = 2, mismatch = 7, indels = 7, PM=80, PI=10, Minimum alignment score = 50, Maximum period size = 500) (Benson 1999).

Self-synteny: CoGe SynMap (Lyons et al. 2008) was used, inputting *Beta vulgaris* (vEL10_1.0, id37197) and EL10.1 MAKER annotation gff files. Coding sequences were compared using LAST (Kielbasa et al. 2011) and DAGChainer (Haas et al. 2004) (with input settings Maximum distance between two matches = 20 genes, Minimum number of aligned genes = 5). Kn/Ks ratios (Yang 2007) were calculated using default parameters on CoGe (genomevolution.org/wiki/index.php/SynMap).

Genome size variation: Four *Beta vulgaris* populations were evaluated for nuclear DNA content as described (Arumaguthan and Earle 1991). Briefly, young and healthy true leaf tissues from greenhouse grown seedlings were placed in between moist paper towels in zip-lock bags and shipped to the Flow Cytometry Lab at Benaroya Research Institute at Virginia Mason

(Seattle, WA) for next day delivery. 50 mg of leaf tissue from each sample was finely chopped using a razor edge to release intact nuclei for flow cytometric analysis. Chicken erythrocyte nuclei (2.50 pg/2C) were used as an internal standard. A value of 978 Mb per pg was used for genome size conversion (Doležel et al. 2003). Statistical analyses were performed with JMP Pro version 14 (SAS, Cary, NC).

Read count mapping: Reads from five Illumina paired-end sequencing datasets were trimmed and subsampled to produce sets of 25 GB for normalized mapping to the EL10.1 assembly. These were the single sequenced EL10 plant, a single plant two generations less inbred than EL10 (i.e., C869_UK), a pool of 25 individual from the parental population from which EL10 was derived (C869_25), the doubled haploid from which RefBeet was generated (KWS2320), and a single plant of a Japanese O-type breeding line (NK-388mm-O) (each accessible at NCBI BioProject PRJNA563463). Four samples of KWS2320 genomic reads (SRR869628, SRR869631, SRR869632, and SRR869633) were obtained from the NCBI SRA and pooled prior to filtering. FASTQ reads from the 5 mapping samples were filtered for a minimum FASTQ quality of 6 and minimum length of 80 bp after trimming. The reads that passed the filter were randomly subsampled to obtain 25 GB of reads per sample. Each pool of 25 GB was independently mapped to the EL10 assembly using BBMap v. 36.67 (Bushnell 2014). Read mapping was done with default parameters and kmer length = 13 with the addition of 'local=t 'to allow soft-clipping the ends of alignments and 'ambiguous=random' to randomly assign reads with multiple best matches among all best sites, to facilitate mapping of repetitive sequences evenly across the genome. For plotting read depth, 5 kb bins were created across each chromosome and the read coverage per base pair was calculated for each bin. The ‘basecov’ and ‘covstats’ outputs of BBMap were used to determine read depths and their standard deviations.

Multispecies Synteny: The analysis of synteny was accomplished by plotting collinear blocks relative to beet chromosomes. Collinear blocks were defined using the program MCScanX using default recommendations (Wang et al. 2012). Protein sets for *A. thaliana*, *V. vinifera*, *S. oleraceae*, and *A. hypocondriacus* were downloaded from phytozome (https://phytozome.jgi.-doe.gov/pz/portal.html) with their corresponding gff files. Quinoa data were downloaded from chenopodiumdb (*www.cbrc.kaust.edu.sa/chenopodiumdb/*) and the *B. vulgaris* proteins and gff files were developed for this report.

Accession Numbers: Sequence data from this article can be found in the EMBL/GenBank data libraries. The EL10 sugar beet whole genome project has been deposited in NCBI under the accession PCNB00000000. EL10.1 is version PCNB01000000. Associated NCBI database pointers are BioSample SAMN07736104, BioProject PRJNA413079; Assembly GCA_002917755.1, and WGS Project PCNB01. All raw reads used in EL10 genome assemblies are deposited in the short-read archive (SRA): Illumina reads SRR6305245; PacBio Reads SRR6301225; and Hi-C Library reads SRR10011257 (Phase Genomics) and SRR12507442 & SRR12507443 (Dovetail Genomics). BioNano Maps are located at SAMN08939661 (*Bsp*Q1) and SAMN08939667 (*Bss*S1). Read mapping accessions are deposited under BioProject PRJNA563463, and BioSamples SAMN12674955 (C869_UK), SAMN12674956 (C869_25), SAMN12674957 (NK-388mm-O). The EL10 genome assemblies and annotations can be viewed and downloaded via the CoGe Genome Browser available at genomevolution.org/coge/, both EL10.1 (Genome ID = 54615) and EL10.2 (Genome ID = 57232), and Phytozome only for EL10.1 (phytozome-next.jgi.doe.gov/info/Bvulgaris_EL10_1_0).

Genome browsing and file resources including transcript assemblies are available at sugarbeets.msu.edu. Transcript assemblies were constructed from root development and leaf RNA-seq reads derived from C869 (the EL10 progenitor) from 3 to 10 weeks post emergence (Trebbi and McGrath 2009) [3-week-old root (SRR10039097), 4-week-old root (SRR10039086), 5-week-old root (SRR10039081), 6-week-old root (SRR10039080), 7-week-old root (SRR10039079), 10-week-old root (SRR10039098), and mature leaf (SRR10037935)]. Also included were RNA-seq sets of 96 hr germinated seedlings from other germplasm germinated under aqueous stress conditions (McGrath et al. 2000), including 150 mM NaCl, 0.3% hydrogen peroxide, and biologically extreme temperatures (10 and 41 ^{o}C) (SRR10039075, SRR10039076, SRR10039077, SRR10039078, SRR10039082, SRR10039083, SRR10039084, SRR10039085, SRR10039087, SRR10039088, SRR10039089, SRR10039090, SRR10039091, SRR10039092, SRR10039093, SRR10039094, SRR10039095, and SRR10039096). The transcript assemblies are located at http://sugarbeets.msu.edu/data/EL10.1/.

Supplemental Data files: To be selected from the contributed, at the discretion of the editors as appropriate.

Acknowledgements: We thank Safa Alzohairy for isolating RNA from germinating sugar beet seedlings under various germination regimes. We also thank K. Arumuganathan, Director of the Flow Cytometry Lab of the Benaroya Research Institute at Virginia Mason, Seattle, WA for contracting the genome size estimates, and the staff at Dovetail Genomics (Santa Cruz, CA) for their assistance with the EL10.2 assembly. JMM: Funding provided by USDA-ARS CRIS 3635-21000-011-00D and the Beet Sugar Development Foundation, Denver, CO, USA. SK and AP were supported by the Intramural Research Program of the National Human Genome Research Institute, National Institutes of Health. Authors assert no conflicts of interest.

Author Contributions:

JMM, BT, EM-G, KD: Conceived and organized the work and wrote the manuscript; AF, PG, SO: Characterized EL10.1 assembly sequence organization; KD, HD, SJ: Created PacBio resources; JL, AH: Created BioNano resources; IL, SS, SK, AP: Conducted Hi-C assembly and finishing of EL10.1; AD, GW, SB, PS, KT: Applied proprietary genetic markers and materials to assess integrity of the EL10 assemblies; JW, TL, JP, KC: Provided MAKER gene annotations for EL10.1; AY, DF: Created RNA-seq transcriptome assemblies used in the EL10.1 annotation.

Author emails:

Mitch McGrath <mitchmcg@msu.edu>

Andrew Funk <afunk1@carthage.edu>

Paul Galewski <paulgalewski@gmail.com>

Shujun Ou <oushujun@iastate.edu>

Belinda Townsend <townsendbj810@gmail.com>

Karen Walston Davenport <kwdavenport@lanl.gov>

Hajnalka Daligault <hajkis@lanl.gov>

Shannon Lyn Johnson <shannonj@lanl.gov>

Joyce Lee <jlee@bionanogenomics.com>

Alex Hastie <ahastie@bionanogenomics.com>

Aude Darracq <Aude.Darracq@sesvanderhave.com>

Glenda Willems <Glenda.Willems@sesvanderhave.com>

Steve Barnes <barnes.sr@gmail.com>

Ivan Liachko <ivan@phasegenomics.com>

Shawn Sullivan <shawn@phasegenomics.com>

Sergey Koren <sergey.koren@nih.gov>

Adam Phillippy <adam.phillippy@nih.gov>

Jie Wang <jwang7@lbl.gov>

Tiffany Liu <liutiff1@msu.edu>

Jane Pulman <jane.pulman@googlemail.com>

Kevin Childs <kchilds@msu.edu>

Anastasia Yocum <anastasia.yocum@a2idea.com>

Damian Fermin <damian.fermin@a2idea.com>

Effie Mutasa-Gottgens <effie@ebi.ac.uk>

Piergiorgio Stevanato <stevanato@unipd.it>

Kazunori Taguchi <ktaguchi@affrc.go.jp>

Kevin Dorn <Kevin.Dorn@usda.gov>

References:

Abe, J., Guan, G-P., and Shimamoto, Y. (1993) Linkage maps for nine isozyme and four marker loci in sugarbeet (*Beta vulgaris* L.). Euphytica 66:117-126.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Journal of Molecular Biology 215:403-410

Andrello, M., K. Henry, P. Devaux, B. Desprez, and S. Manel. 2016. Taxonomic, spatial and adaptive genetic variation of *Beta* section *Beta*. Theor. Appl. Genet. 129:257–271.

Andrello, M., K. Henry, P. Devaux, D. Verdelet, D., B. Desprez, and S. Manel. 2017. Insights into the genetic relationships among plants of *Beta* section *Beta* using SNP markers. Theor. Appl. Genet. 130:1857-1866.

Arumuganathan, K., and E.D. Earle. 1991. Nuclear DNA content of some important plant species. Plant Mol. Biol. Rep. 9:208-218.

Bennetzen JL, Wang H (2014) The contributions of transposable elements to the structure, function, and evolution of plant genomes. Annual Review of Plant Biology 65:505–530.

Benson, G. (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acid Research 27:573-580.

Biancardi, E., L.W. Panella, and R.T. Lewellen. 2012. *Beta maritima*: The origin of beets. Springer, New York.

Bickhart DM, Rosen BD, Koren S, Sayre BL, Hastie AR, Chan S, Lee J, Lam ET, Liachko I, Sullivan ST, Burton JN, Huson HJ, Nystrom JC, Kelley CM, Hutchison JL, Zhou Y, Sun J, Crisà A, Ponce de León FA, Schwartz JC, Hammond JA, Waldbieser GC, Schroeder SG, Liu GE, Dunham MJ, Shendure J, Sonstegard TS, Phillippy AM, Van Tassell CP, Smith TPL (2017) Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. Nature Genetics 49: 643-650.

Bushnell, B. (2014). BBMap: A fast, accurate, splice-aware aligner. - Report Number: LBNL-7065E.

Butterfass, T. 1964. Die chloroplastenzahlen in verschiedenartigen zel- len trisomer zuckerruben (*Beta vulgaris* L.). Z. Bot. 52:46–77.

Campbell, M.S., Law, M., Holt, C., Stein, J.C., Moghe, G.D., Hufnagel, D.E., Lei, J., Achawanantakun, R., Jiao, D., Lawrence, C.J., et al. (2014). MAKER-P: A tool kit for the rapid creation, management, and quality control of plant genome annotations. Plant Physiol. 164: 513–524.

Castro S, Romeiras MM, Castro M, Duarte MC, Loureiro J. (2013) Hidden diversity in wild *Beta* taxa from Portugal: Insights from genome size and ploidy level estimations using flow cytometry. Plant Science 207: 72–78.

Cheng, C.-Y., Krishnakumar, V., Chan, A.P., Thibaud-Nissen, F., Schobel, S., and Town, C.D. (2017). Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. Plant J. 89: 789–804.

Dechyeva D, Schmidt T. (2007) Molecular organization of terminal repetitive DNA in *Beta* species. Chromosome Res. 214:881–97.

del Rio AR, Minoche AE, Zwickl NF, Friedrich A, Liedtke S, Schmidt T, Himmelbauer H, Dohm JC (2019) Genomes of the wild beets *Beta patula* and *Beta vulgaris* ssp. *maritima*. The Plant Journal: DOI:10.1111/tpj.14413.

Dohm JC, Lange C, Holtgräwe D, Rosleff Sörensen T, Borchardt D, Schulz B, Lehrach H, Weisshaar B, Himmelbauer H. (2012) Palaeohexaploid ancestry for Caryophyllales inferred from extensive gene-based physical and genetic mapping of the sugar beet genome (*Beta vulgaris*) Plant J. 70:528-540.

Dohm, J.C., C. Lange, R. Reinhardt, and H. Himmelbauer. 2009. Haplotype divergence in *Beta vulgaris* and microsynteny with sequenced plant genomes. Plant Journal 57:14-26.

Dohm, J.C., Minoche, A.E., Holtgräwe, D., Capella-Gutiérrez, S., Zakrzewski, F., Tafer, H., Rupp, O., Sörensen, T.R., Stracke, R., Reinhardt, R., A. Goesmann, B. Schulz, P.F. Stadler, T. Schmidt, T. Gabaldón, H. Lehrach, B. Weisshaar, and H. Himmelbauer (2014). The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). Nature 505, 546–549.

Doležel, J., Bartoš, J., Voglmayr, H. and Greilhuber, J. (2003), Letter to the editor. Cytometry, 51A: 127–128. doi:10.1002/cyto.a.10013

Earl D, Bradnam K, St John J, Darling A, Lin D, Fass J, Yu HO, Buffalo V, Zerbino DR, Diekhans M, Nguyen N, Ariyaratne PN, Sung WK, Ning Z, Haimel M, Simpson JT, Fonseca

NA, Birol İ, Docking TR, Ho IY, Rokhsar DS, Chikhi R, Lavenier D, Chapuis G, Naquin D, Maillet N, Schatz MC, Kelley DR, Phillippy AM, Koren S, Yang SP, Wu W, Chou WC, Srivastava A, Shaw TI, Ruby JG, Skewes-Cox P, Betegon M, Dimon MT, Solovyev V, Seledtsov I, Kosarev P, Vorobyev D, Ramirez-Gonzalez R, Leggett R, MacLean D, Xia F, Luo R, Li Z, Xie Y, Liu B, Gnerre S, MacCallum I, Przybylski D, Ribeiro FJ, Yin S, Sharpe T, Hall G, Kersey PJ, Durbin R, Jackman SD, Chapman JA, Huang X, DeRisi JL, Caccamo M, Li Y, Jaffe DB, Green RE, Haussler D, Korf I, Paten B. (2011) Assemblathon 1: a competitive assessment of de novo short read assembly methods. Genome Res. 21:2224-2241. doi: 10.1101/gr.126599.111.

Fernandez-Pozo, N., Menda, N., Edwards, J.D., Saha, S., Tecle, I.Y., Strickler, S.R., Bombarely, A., Fisher-York, T., Pujar, A., Foerster, H., et al. (2015). The Sol Genomics Network (SGN)—from genotype to phenotype to breeding. Nucleic Acids Res. 43, D1036–D1041.

Finn, R.D., Clements, J., and Eddy, S.R. (2011). HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 39, W29–W37.

Flavell R.B., M.D. Bennet, and J.B. Smith. 1974. Genome size and the proportion of repeated nucleotide sequence DNA in plants. Biochem. Genet. 12:257-269.

Funk, A., Galewski, P., McGrath, J. M. (2018) Nucleotide-binding resistance gene signatures in sugar beet, insights from a new reference genome. The Plant Journal 95:659-671

Galon, J., and D.T. Zallen. 1998. The role of the Vilmorin Company in the promotion and diffusion of the experimental science of heredity in France, 1840–1920. J. Hist. Biol. 31:241–262.

Galewski, P., McGrath, J.M. 2020. Genetic diversity among cultivated beets (Beta vulgaris) assessed via population-based whole genome sequences. BMC Genomics 21:189. https://doi.org/10.1186/s12864-020-6451-1

Goldman, I.L., and D. Austin. 2000. Linkage among the *R*, *Y* and *BI* loci in table beet. Theor. Appl. Genet. 100:337–343.

Haas BJ, Delcher AL, Wortman JR, Salzberg SL (2004) DAGchainer: a tool for mining segmental genome duplications and synteny. Bioinformatics 20: 3643–3646

Halldén, C., D. Ahrén, A. Hjerdin, T. Säll, and N.O. Nilsson. 1998. No conserved homoeologous regions found in the sugar beet genome. J. Sugar Beet Res. 35:1-13.

Hatlestad, G.J., N.A. Akhavan, R.M. Sunnadeniya, L. Elam, S. Cargyle, A. Hembd, A. Gonzalez, J.M. McGrath, and A.M. Lloyd. 2014. The beet *Y* locus is a co-opted anthocyanin MYB that regulates betalain pathway structural genes. Nature Genetics 47:92–96.

Hatlestad, G.J., R.M. Sunnadeniya, N.A. Akhavan, A. Gonzalez, I.L. Goldman, J.M. McGrath, and A.M. Lloyd. 2012. The beet *R* locus encodes a new cytochrome P450 required for red betalain production. Nature Genetics 44:816–820.

Holt, C., and Yandell, M. (2011). MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics 12, 491.

Hosoda K, Imamura A, Katoh E, Hatta T, Tachiki M, Yamada H, Mizuno T and Yamazaki T (2002) Molecular structure of the GARP family of plant Myb-related DNA binding motifs of the Arabidopsis response regulators. The Plant Cell 14:2015–2029

Jarvis DE, Ho YS, Lightfoot DJ, Schmöckel SM, Li B, Borm TJA, et al. 2017. The genome of *Chenopodium quinoa*. Nature 542: 307-312

Jung H, Winefield C, Bombarely A, Prentis P, Waterhouse P. (2019) Tools and strategies for long-read sequencing and *de novo* assembly of plant genomes. Trends Plant Sci. 24:700-724.

Keller, W. 1936. Inheritance of some major color types in beets. J. Agric. Res. 52:27-38.

Kielbasa, S.M., Wan, R., Sato, K., Horton, P., and Frith, M.C. (2011). Adaptive seeds tame genomic sequence comparison. Genome Research 21:487–493.

Korf, I. (2004). Gene finding in novel genomes. BMC Bioinformatics 5, 59.

Kowar T, Zakrzewski F, Macas J, Kobliz□ková A, Viehoever P, Weisshaar B, Schmidt T (2016) Repeat composition of CenH3-chromatin and H3K9me2-marked heterochromatin in sugar beet (*Beta vulgaris*)BMC Plant Biology (2016) 16:120

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L.L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes11 Edited by F. Cohen. J. Mol. Biol. 305, 567–580.

Lewellen, R.T. (2004) Registration of C869 and C869CMS. Crop Sci. 44:357.

Lightfoot DJ, Jarvis DE, Ramaraj T, Lee R, Jellen EN and Maughan PJ (2017) Single-molecule sequencing and Hi-C-based proximity-guided assembly of amaranth (*Amaranthus hypochondriacus*) chromosomes provide insights into genome evolution. BMC Biology (2017) 15:74

Lyons, E., Pedersen, B., Kane, J., Alam, M., Ming, R., Tang, H., Wang, X., Bowers, J., Paterson, A., and Lisch, D. (2008) Finding and comparing syntenic regions among Arabidopsis and the outgroups papaya, poplar and grape: CoGe with rosids. Plant Phys. 148:1772–1781.

Mason MG, Mathews DE, Argyros DA Maxwell BB, Kieber JJ, Alonso JM, Ecker JR, and Schaller GE. (2005) Multiple Type-B response regulators mediate cytokinin signal transduction in *Arabidopsis*. The Plant Cell 17:3007–3018

McGrath JM, Drou N, Waite D, Swarbreck D, Mutasa-Gottgens E, Barnes S, Townsend B. (2013) The 'C869' Sugar Beet Genome: A Draft Assembly. https://pag.confex.com/pag/xxi/webprogram/Paper5768.html. Accessed 8/29/2020.

McGrath, J. M., Koppin, T. K. and Duckert, T. M. (2005) Breeding for genetics: Development of Recombinant Inbred Lines (RILs) for gene discovery and deployment. Proceedings of the American Society of Sugar Beet Technologists: 124-132. (www.bsdf-assbt.org/wp-content/uploads/2017/04/PASSBTAgp124to132BreedingforGeneticsDevelopmentofRecombinantInbredLinesforGeneDiscoveryandDeployment.pdf; accessed June 16, 2019)

McGrath, J. M., Trebbi, D., Fenwick, A. Panella, L., Schulz, B., Laurent, V., Barnes, S. and Murray, S. C. (2007) An open-source first-generation molecular genetic map from a sugarbeet x table beet cross and its extension to physical mapping. Crop Science 47:S-27-S-44.

McGrath, J.M., and B.J. Townsend. (2015) Sugar beet, energy beet, and industrial beet. p. 81-99. In: V.M.V. Cruz and D.A. Dierig (eds.), Handbook of Plant Breeding. Volume 9. Industrial Crops: Breeding for Bioenergy and Bioproducts. Springer, New York.

McGrath, J.M., and Panella, L. (2019) Sugar Beet Breeding. Plant Breeding Reviews 42:167-218.

McGrath, J.M., Derrico, C.A., Morales, M., Copeland, L.O., Christenson, D.R. (2000) Germination of sugar beet (*Beta vulgaris* L.) seed submerged in hydrogen peroxide and water

as a means to discriminate cultivar and seedlot vigor. Seed Science and Technology. 28:607-620.

Melters DP, Bradnam KR, Young HA, Telis N, May MR, Ruby JG, Sebra R, Peluso P, Eid J, Rank D, Garcia JF, DeRisi JL, Smith T, Tobias C, Ross-Ibarra J, Korf I, and Chan SWL (2013) Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. Genome Biology 2013 14:R10.

Meyer M, Kircher M. 2010. Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. Cold Spring Harbor Protocol doi:10.1101/pdb.prot5448.

Ou S, and Jiang N. 2018. LTR_retriever: a highly accurate and sensitive program for identification of long terminal-repeat retrotransposons. Plant Physiology 176:1410–1422.

Paesold, S., D. Borchart, T. Schmidt, D. Dechyeva. 2012. A sugar beet (*Beta vulgaris* L.) reference FISH karyotype for chromosome and chromosome-arm identification, integration of genetic linkage groups and analysis of major repeat family distribution. Plant J. 72:600-611.

Panella, L., L.G. Campbell, I.A. Eujayl, R.T. Lewellen, and J.M. McGrath. 2015. USDA-ARS sugar beet releases and breeding over the past 20 years. J. Sugar Beet Research 52:22-67.

Panella, L., R.T. Lewellen, and L.E. Hanson. 2008. Breeding for multiple disease resistance in sugar beet: Registration of FC220 and FC221. J. Plant Registrations 2:146–155.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.-C., Mendell, J.T., and Salzberg, S.L. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. 33, 290–295.

Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods 8, 785–786.

Pichersky, E., Logsdon, J. M., McGrath, J. M. and Stasys, R. A. 1991. Fragments of plastid DNA in the nuclear genome of tomato: prevalence, chromosomal location and possible mechanism of integration. Molecular and General Genetics 225:453-458.

Pin, P.A., R. Benlloch, D. Bonnet, E. Wremerth-Weich, T. Kraft, J. Gielen, and O. Nilsson. 2010. An antagonistic pair of FT homologs mediates the control of flowering time in sugar beet. Science 330:1397-1400.

Pin, P.A., W. Zhang, S.H. Vogt, N. Dally, B. Büttner, G. Schulze-Buxloh, N.S. Jelly, T.Y.P. Chia, E.S. Mutasa-Göttgens, J.C. Dohm, H. Himmelbauer, B. Weisshaar, J. Kraus, J.J.L. Gielen, M. Lommel, G. Weyens, B. Wahl, A. Schechert, O. Nilsson, C. Jung, T. Kraft, and A.E. Müller. 2012. The role of a pseudo-response regulator gene in life cycle adaptation and domestication of beet. Current Biology 22:1095-1101.

Pucker B (2019) Mapping-based genome size estimation. bioRxiv preprint first posted online Apr. 13, 2019; doi: http://dx.doi.org/10.1101/607390.

Putnam NH, O'Connell B, Stites JC, Rice BJ, Blanchette M, Calef R, Troll CJ, Fields A, Hartley PD, Sugnet CW, Haussler D, Rokhsar DS, Green RE. (2016) Chromosome-scale shotgun assembly using an *in vitro* method for long-range linkage. Genome Res. 26: 342-350.

Roessler K, Muyle A, Diez CM, Gaut GRJ, Bousious A, Stitzer MC, Seymour DK, Doebley JF, Liu Q, Gaut BS (2019) The genome-wide dynamics of purging during selfing in maize. Nature Plants 5: 980-990.

Schmidt T, Heslop-Harrison JS (1998) Genomes, genes and junk: the large-scale organization of plant chromosomes. Trends Plant Sci 3: 195-199.

Schmidt, T., Jung, C., Metzlaff, M. (1991) Distribution and evolution of two satellite DNAs in the genus *Beta*. Theor. Appl. Genetics 82:793-799.

Schondelmaier, J., and C. Jung. 1997. Chromosomal assignment of the nine linkage groups of sugar beet (*Beta vulgaris* L.) using primary trisomics. Theor. Appl. Genet. 95:590-596.

Schwacke R, Ponce-Soto GY, Krause K, Bolger AM, Arsova B, Hallab A, Gruden K, Stitt M, Bolger ME, Usadel B (2019). MapMan4: A refined protein classification and annotation framework applicable to multi-omics data analysis. Mol. Plant. 12: 879–892.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, and Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics, published online June 9, 2015, doi: 10.1093/bioinformatics/btv351

Stanke, M., and Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics 19, ii215-ii225.

Taguchi K, Kuroda Y, Okazaki K, Yamasaki M. (2019) Genetic and phenotypic assessment of sugar beet (*Beta vulgaris* L. subsp. *vulgaris*) elite inbred lines selected in Japan during the past 50 years. Breeding Science 69:255-265.

Taguchi, K. (2014) Genetics and breeding studies on Aphanomyces root rot resistance of sugar beet, covers from the discovery of genetic resources to development of new varieties. Breed. Res. 16:186–191

The UniProt Consortium (2017). UniProt: the universal protein knowledgebase. Nucleic Acids Res. 45, D158–D169.

Trebbi D, McGrath JM (2009) Functional differentiation of the sugar beet root system as indicator of developmental phase change. Physiologia Plantarum 135: 84–97.

Tuskan, G.A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, A., et al. (2006). The Genome of Black Cottonwood, *Populus trichocarpa* (Torr. & Gray). Science 313, 1596–1604.

Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H, Kissinger JC, Paterson AH. (2012) MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res, 40(7): e49.

Whitney KD, Baack EJ, Hamrick JL, Godt MJW, Barringer BC, Bennett MD, Eckert CG, Goodwillie C, Kalisz S, Leitch IJ, Ross-Ibarra J. (2010) A role for nonadaptive processes in plant genome size evolution? Evolution 64: 2097–2109.

Yang W-D, Tan H-W, Zhu W-M (2016) SpinachDB: A well-characterized genomic database for gene family classification and SNP information of spinach. PLoS ONE 11: e0152706. doi:10.1371/journal.pone.0152706.

Yang Y, Moore MJ, Brockington SF, Soltis DE, Wong GK, Carpenter EJ, Zhang Y, Chen L, Yan Z, Xie Y, Sage RF, Covshoff S, Hibberd JM, Nelson MN, Smith SA. (2015) Dissecting molecular evolution in the highly diverse plant clade Caryophyllales using transcriptome sequencing. *Mol Biol Evol* 32:2001-2014.

Yang, Z. (2007) PAML 4: Phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution 24:1586-1591.

Table legends:

Table 1: Sequence inputs and metrics used in construction of EL10.1.

Table 2: Assembly metrics for EL10.1 and sequence assembly iterations.

Table 3: Inversions in the EL10.1 genome assembly assessed using genetic markers.

Table 4: Co-locations of Scaffolds and Chromosomes deduced by genetically mapped markers.

Table 5: Orientation of EL10.1 Chromosomes relative to the cytogenetic map of Paesold et al. (2012).

Table 6: The Y-R-B linkage group in the EL10.1 genome assembly.

Table 7: Gene models detected via Benchmarking Universal Single-Copy Orthologs (BUSCO).

Table 8: Distribution of MAKER annotations across the EL10.1 genome assembly.

Table 9: Comparison of MapMan4 first-order functional classifications for EL10.1 and four other Tracheophytes [*Oryza sativa (Os) Brachypodium distachyon (Bd), Arabidopsis thaliana (At), Solanum lycopersicum (Sl), and Manihot esculenta (Me*)].

Table 10: Comparison of MapMan4 leaf bins where no gene was predicted by MAKER in EL10.1 and five Tracheophytes [*Oryza sativa (Os) Brachypodium distachyon (Bd), Arabidopsis thaliana (At), Solanum lycopersicum (Sl), and Manihot esculenta (Me*)].

Table 11: Comparison of over- and under-represented MapMan4 second-order functional classifications for EL10.1 and four other Tracheophytes [*Oryza sativa (Os) Brachypodium distachyon (Bd), Arabidopsis thaliana (At), Solanum lycopersicum (Sl), and Manihot esculenta (Me*)].

Table 12: Comparison of Transcription factor classes for EL10.1 and four other Tracheophytes [*Oryza sativa (Os) Brachypodium distachyon (Bd), Arabidopsis thaliana (At), Solanum lycopersicum (Sl), and Manihot esculenta (Me*)].

Table 13: Beet genome size estimates obtained by flow cytometry.

Table 14: Frequency of Transposable Element (TE) classes in the EL10.1 genome assembly. A. RepeatMasker-derived annotations. B. Complete LTR retrotransposons. C. Interstitial repeat classes (from Kowar et al. 2016), D. Terminal repeat locations (from Dechyeva and Schmidt

2006) integrated with cytogenetic orientation (parentheses indicate reversed orientation relative to Paesold et al. 2012).

Table 15: Characteristics of tandem repeats in the EL10.1 genome assembly.

Table 16: Read count mapping of short reads from EL10 and four other germplasms to the EL10.1 genome assembly.

Table 17: Proportion and metrics of synteny (co-linear blocks of MAKER beet gene predictions) shared among five species.

Figure Legends:

Figure 1: Chromosome alignment of the EL10 assembly (x-axis) versus RefBeet-1.2 assembly (y-axis) by EL10.1 Chromosome. Alignments less than 5 kb in length were removed before plotting. Alignments with matching orientation are shown in red, inversions are show in blue. Unassembled RefBeet regions are indicated by gaps.

Figure 2: Comparison of contiguity between EL10.1 and EL10.2 genome assemblies. Alignments with matching orientation are shown in red, inversions are show in blue.

Figure 3: Self-synteny of EL10.1 Chromosomes against the EL10.1 predicted protein set.

Figure 4: Distribution of LTR Copia and Gypsy retrotransposon elements across the EL10.1 Chromosomes.

Figure 5: Copy number per consensus tandem repeat length in the EL10.1 genome assembly.

Figure 6: LTR Assembly Index (LAI) of the RefBeet assembly (A) and EL10 assembly (B) of the sugar beet genome. X-axes denote pseudochromosomes of the two assemblies. Each dot represents regional LAI in a 3 Mb window. Red-dotted lines indicate the LAI cutoff of the reference genome quality (LAI = 10). Blue-dotted lines indicate the mean LAI.

Figure 7: Read count mapping of short reads from EL10 and four other germplasms to the EL10.1 genome assembly and the standard deviation of reads mapped to each 5 kb window across the entire EL10.1 genome assembly.

Figure 8: Distribution of high-copy number variant differences (>2000 copies per 5 kb window) between open pollinated population C869_25 and four inbred sugar beets across Chromosome 1 of the EL10.1 genome assembly.

Figure 9: Visualization of syntenic blocks among Caryophyllales genomes relative to *B. vulgaris* EL10.1 Chromosomes compared with two representative Rosid species, color coded by EL10.1 Chromosome.

Figure 1: Chromosome alignment of the EL10 assembly (x-axis) versus RefBeet-1.2 assembly (y-axis) by EL10.1 Chromosome. Alignments less than 5 kb in length were removed before plotting. Alignments with matching orientation are shown in red, inversions are show in blue. Unassembled RefBeet regions are indicated by gaps.

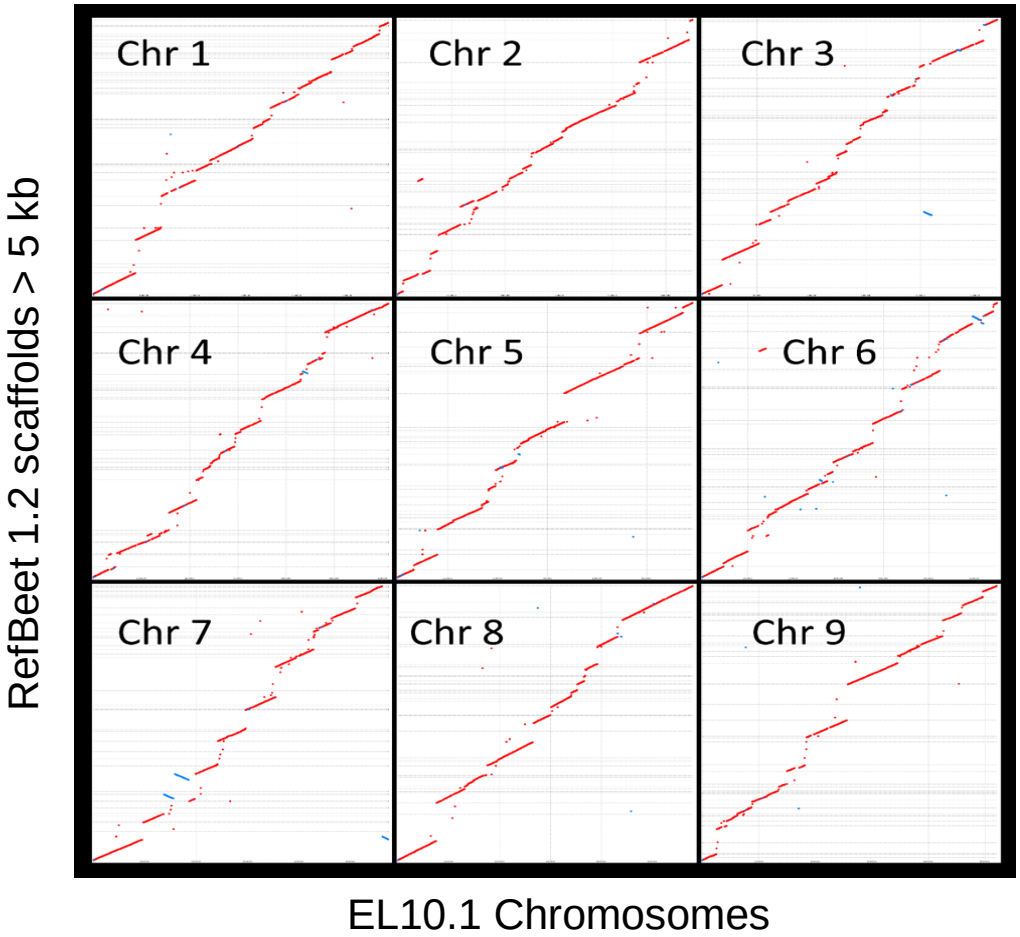

Figure 2: Comparison of contiguity between EL10.1 and EL10.2 genome assemblies. Alignments with matching orientation are shown in red, inversions are show in blue.

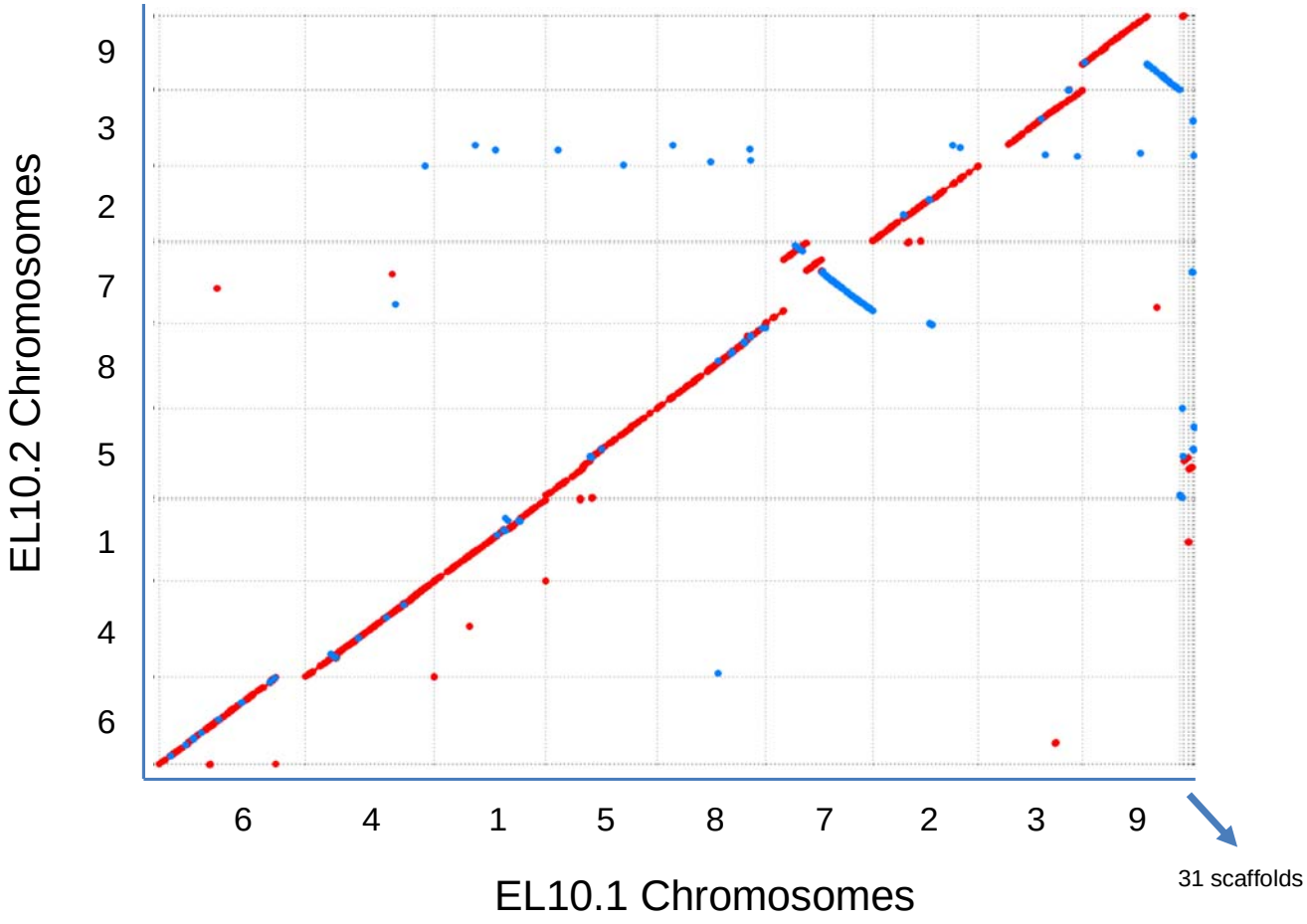

Figure 3: Self-synteny of EL10.1 Chromosomes against the EL10.1 predicted protein set.

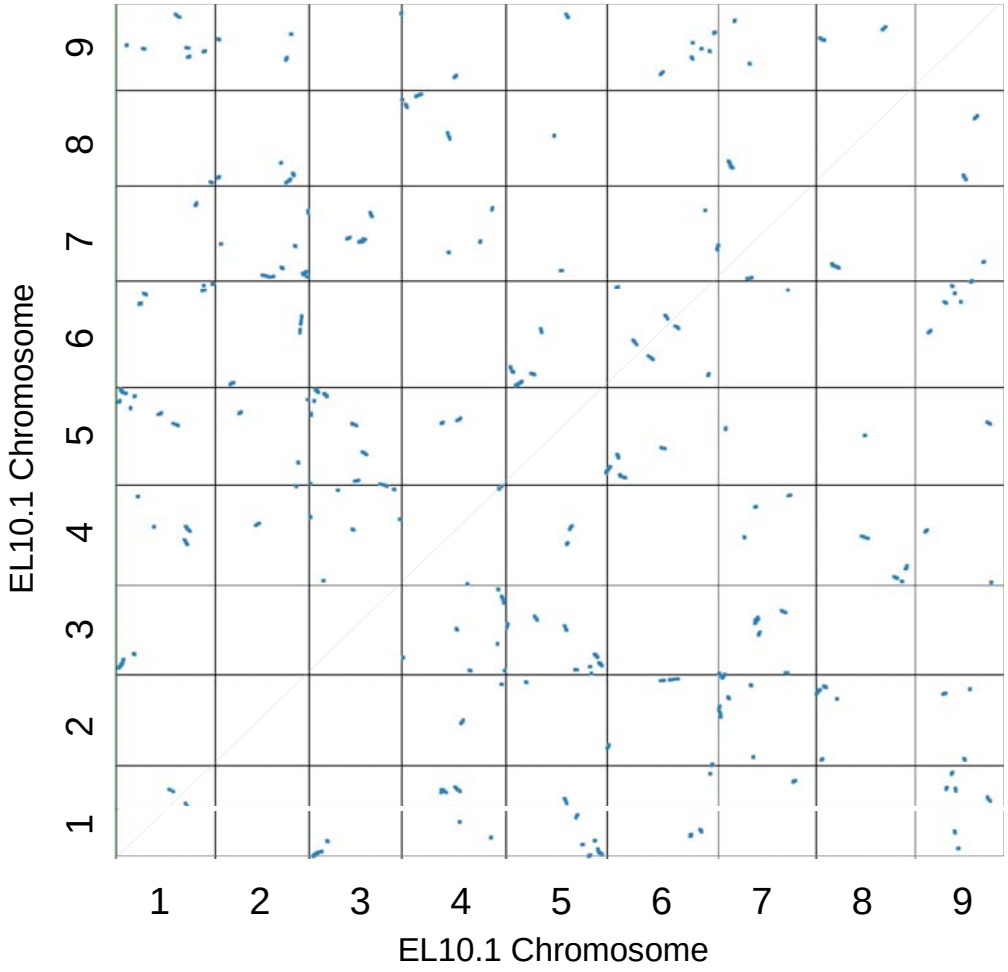

Figure 4: Distribution of LTR Copia and Gypsy retrotransposon elements across the EL10.1 Chromosomes.

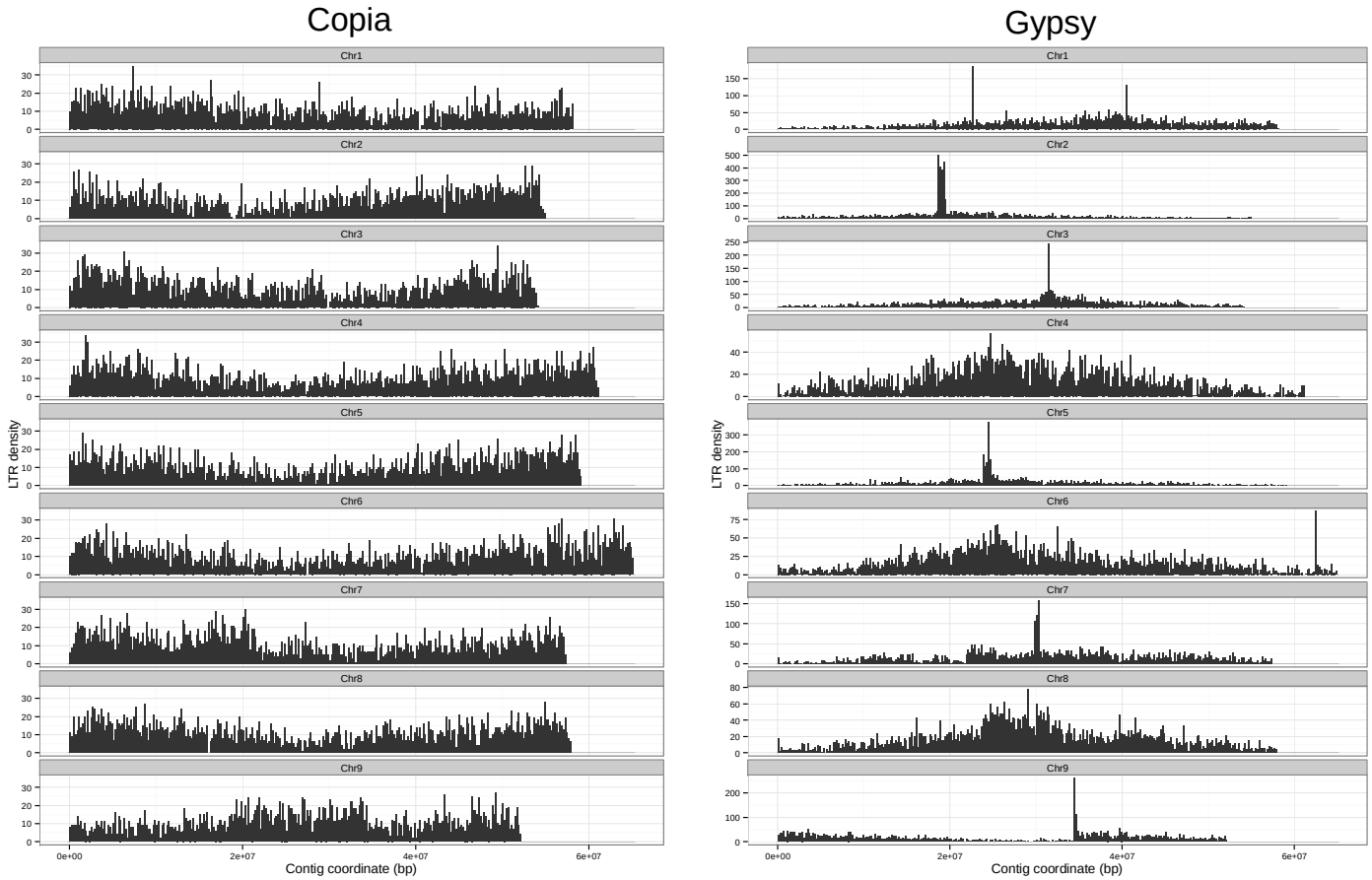

Figure 5: Copy number per consensus tandem repeat length in the EL10.1 genome assembly.

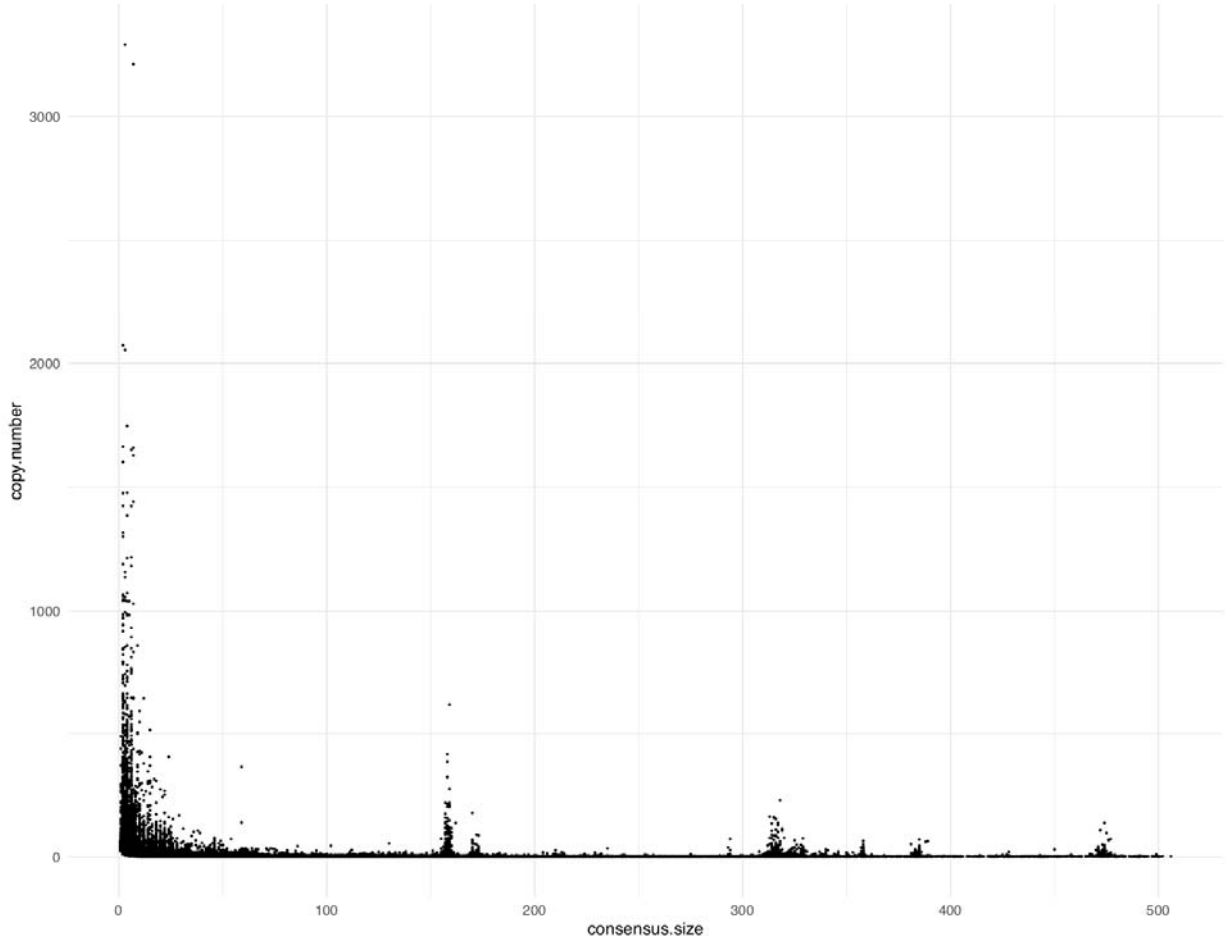

Figure 6: LTR Assembly Index (LAI) of the RefBeet assembly (A) and EL10 assembly (B) of the sugar beet genome. X-axes denote pseudochromosomes of the two assemblies. Each dot represents regional LAI in a 3 Mb window. Red-dotted lines indicate the LAI cutoff of the reference genome quality (LAI = 10). Blue-dotted lines indicate the mean LAI.

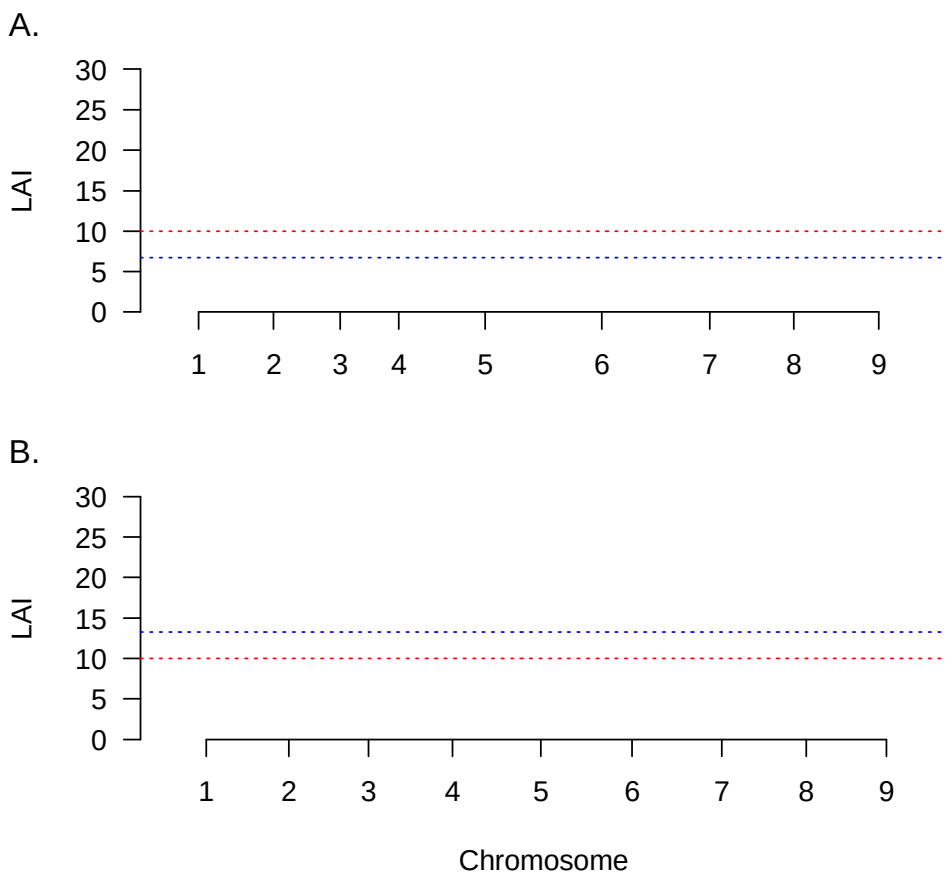

Figure 7: Read count mapping of short reads from EL10 and four other germplasms to the EL10.1 genome assembly and the standard deviation of reads mapped to each 5 kb window across the entire EL10.1 genome assembly.

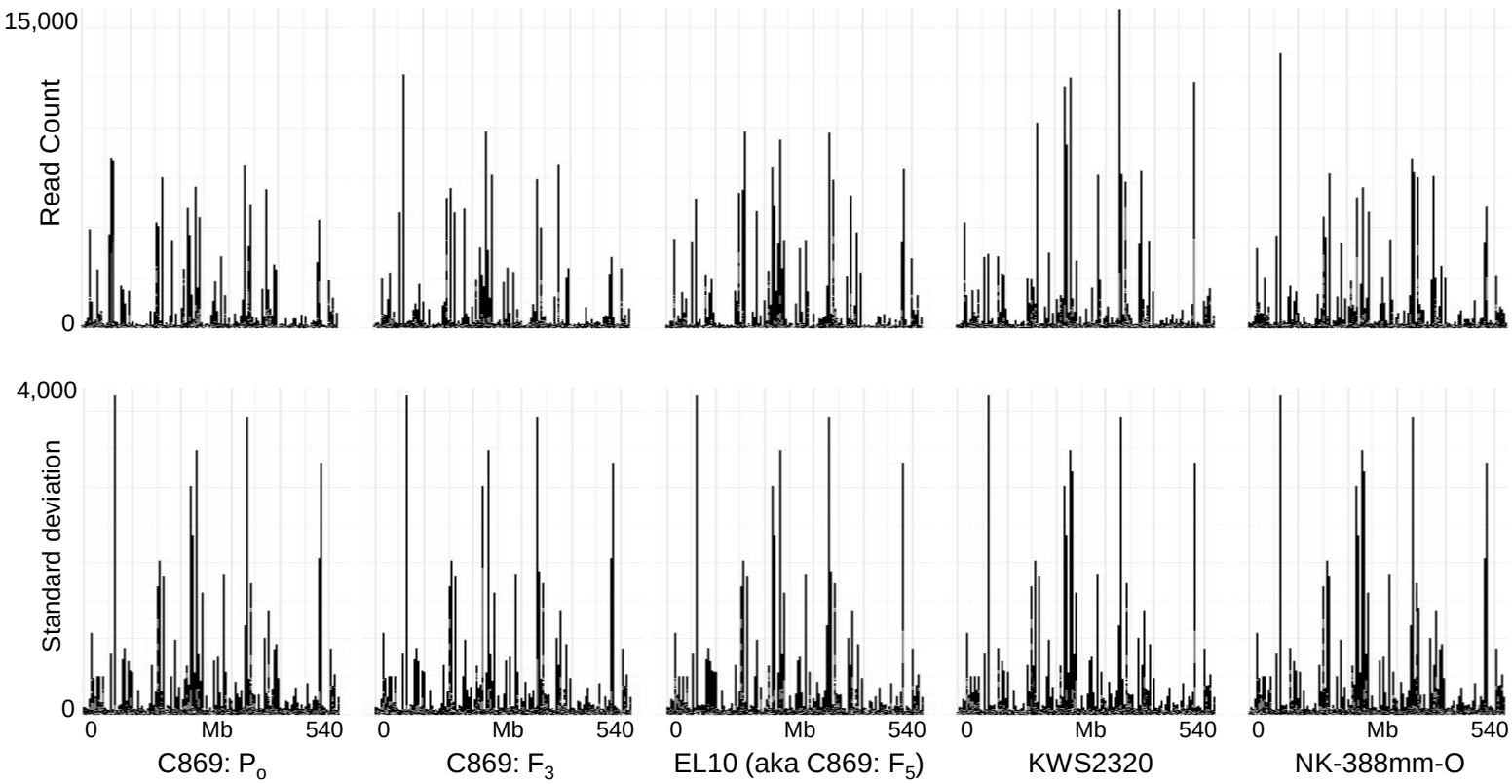

Figure 8: Distribution of high-copy number variant differences (>2000 copies per 5 kb window) between open pollinated population C869_25 and four inbred sugar beets across Chromosome 1 of the EL10.1 genome assembly.

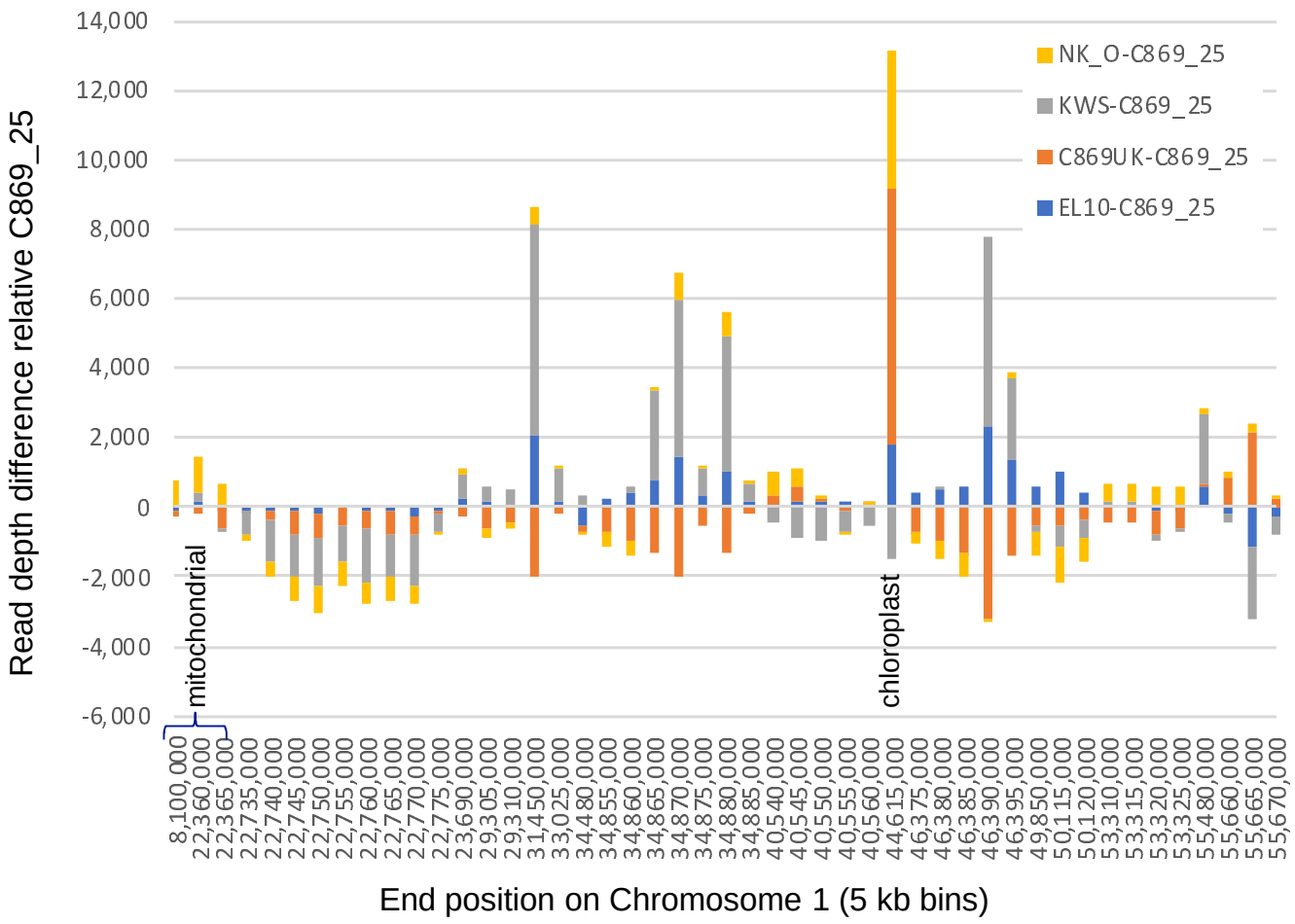

Figure 9: Visualization of syntenic blocks among Caryophyllales genomes relative to *B. vulgaris* EL10.1 Chromosomes compared with two representative Rosid species, color coded by EL10.1 Chromosome.

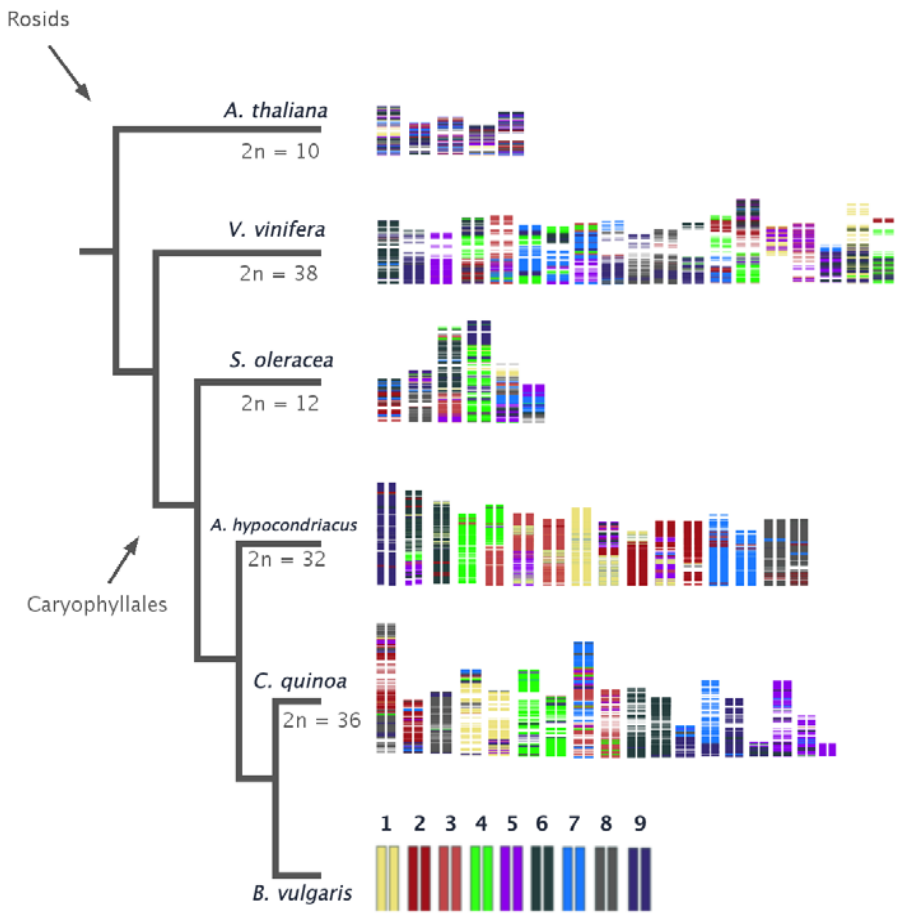

Table 1: Sequence inputs and metrics used in construction of EL10.1

Technology	Library		Coverage[1]
		PacBio passed reads	
PacBio long reads	RS II, P6-C4 chemistry	6,540,795	79.3
	Mean length = 9,096 nt (std.dev = 6,528)		
	> 40 kb initial mapping and pre-assembly	5,176	0.38
		BioNano passed labels	
Optical physical map	BioNano Genomics *Bss* SI - *Bsp* PQ1 Hybrid Scaffold	121 Gb	161.3
	Bsp PQ1 (7.6 labels/100 kb)	40 Gb	
	Bss SI (10 labels/100 kb)	81 Gb	
		Illumina passed reads	
Paired-End short reads	HiSeq 2500, TruSeq Libraries, 125bp PE	447,211,041	149.0
Cross-linked *in vivo*	Phase Genomics Hi-C library, HiSeq 2500, TruSeq Libraries (EL10.1)	355,892,798	118.6
	Dovetail Genomics Hi-C library, HiSeq 10X, TruSeq Libraries (EL10.2)	927,545,984	183.3

[1] Using genome size of 758 Mb

Table 2: Assembly metrics for EL10.1 and sequence assembly iterations.

Assembly by input and method	Name	# Contigs	% Scaffolded	Total size	N50	NG50[1]	% >100 kb	# Scaffolds	Total size	N50	NG50[1]	%N	Coverage %[1]
				(x 1,000 nt)					(x 1,000 nt)				
RefBeet 1.2 (Dohm et al. 2014)	RefBeet	60,051	93.7	517,882	43.8	nd	1.0	40,508	566,571	2,013	nd	8.60	nd
EL10.1 PacBio	SBJ_80X	938	na	562,760	1,394	1,228	70.9	938	562,760	1,394	1,228	0.00	89.6
EL10.1 PacBio BioNano	SBJ_80X_BN	2,983	99.2	533,042	1,340	1,093	21.5	86	566,848	12,513	10,655	5.90	90.3
EL10.1 PacBio BioNano Hi-C	EL10.1	364	96.2	540,479	2,701	2,335	96.7	40	540,537	57,939	57,353	0.01	86.1
	Chromosome_1	47	100	58,076	2,421	nd	100.0	1	58,086	na	nd	0.02	9.2
	Chromosome_2	30	100	54,968	2,834	nd	96.7	1	54,972	na	nd	0.01	9.2
	Chromosome_3	22	100	54,096	3,728	nd	100.0	1	54,100	na	nd	0.01	8.6
	Chromosome_4	47	100	61,154	2,396	nd	97.9	1	61,163	na	nd	0.01	9.7
	Chromosome_5	30	100	59,218	3,579	nd	93.3	1	59,225	na	nd	0.01	9.4
	Chromosome_6	52	100	65,091	2,381	nd	98.1	1	65,097	na	nd	0.01	10.4
	Chromosome_7	40	100	57,345	2,831	nd	95.0	1	57,354	na	nd	0.02	9.1
	Chromosome_8	37	100	57,932	2,335	nd	97.3	1	57,939	na	nd	0.01	9.2
	Chromosome_9	28	100	52,176	2,382	nd	100.0	1	52,180	na	nd	0.01	8.3
	Unplaced Scaffolds	31	0	20,421	1,679	nd	87.1	31	20,421	na	nd	0.00	3.3
EL10.2 PacBio BioNano Hi-C	EL10.2	3,098	99.9	533,041	1,283	nd	21.1	18	567,031	61,792	nd	5.99	90.3
	Chromosome_1	505	100	58,689	1,093	nd	16.0	1	64,154	na	nd	8.52	10.2
	Chromosome_2	253	100	53,946	1,566	nd	21.3	1	56,769	na	nd	4.97	9.0
	Chromosome_3	181	100	54,788	1,907	nd	28.7	1	57,123	na	nd	4.00	9.1
	Chromosome_4	406	100	61,919	1,225	nd	20.0	1	66,143	na	nd	6.39	10.5
	Chromosome_5	276	100	64,991	1,635	nd	21.4	1	67,720	na	nd	4.03	10.8
	Chromosome_6	423	100	68,152	1,004	nd	23.6	1	72,250	na	nd	5.67	11.5
	Chromosome_7	349	100	57,001	1,144	nd	22.6	1	60,906	na	nd	6.41	9.7
	Chromosome_8	245	100	59,411	1,155	nd	31.8	1	61,792	na	nd	3.85	9.8
	Chromosome_9	308	100	51,533	1,527	nd	18.8	1	55,602	na	nd	7.32	8.9
	Unplaced Scaffolds	152	97.4	2,613	252	nd	0.0	9	4,573	na	nd	42.86	0.7

[1] Based on 628 Mb Physical Map

Table 3: Inversions in the EL10.1 genome assembly assessed using genetic markers.

EL10.1	Position (Mb)	SES Marker Position
Chromosome_7	0.2	7
Chromosome_7	9.5	45
Chromosome_7	10.3	30
Chromosome_7	14.1	28
Chromosome_7	15.8	18
Chromosome_7	19.7	27
Chromosome_7	19.9	18
Chromosome_7	21.8	3
Chromosome_7	24.8	32
Chromosome_7	57.3	44
Chromosome_9	3.1	67
Chromosome_9	25.2	50
Chromosome_9	25.8	8
Chromosome_9	34.5	50
Chromosome_9	41.3	71
Chromosome_9	50.6	74

Table 4: Co-locations of Scaffolds and Chromosomes deduced by genetically mapped markers.

EL10.1 Chromosome	EL10.1 Scaffold	Orientation	Bin Position (genomic coordinates)
Chromosome_1	Scaffold_07	reverse	58,086,001 - end
Chromosome_2	Scaffold_19	unknown	30,001,550 - 30,051,550
Chromosome_3	Scaffold_03	reverse	27,470,050 - 27,610,050
Chromosome_5	Scaffold_05	reverse	end - 216,554
Chromosome_5	Scaffold_04	forward	19,422,050 - 19,462,050
Chromosome_5	Scaffold_01	forward	24,502,050 - 24,632,050
Chromosome_5	Scaffold_14	reverse	45,192,050 - 45,212,050
Chromosome_6	Scaffold_02	forward	14,450,050 - 14,610,050
Chromosome_6	Scaffold_08	forward	61,492,050 - 61,552,050

Table 5: Orientation of EL10.1 Chromosomes relative to the cytogenetic map of Paesold et al. (2012).

EL10.1 chromosome	EL10.1 position	EL10.1 orientation	Cytogenetic orientation[1]	Orientation match	Marker Name[1]	Forward primer[1]
Chromosome_1	225,015	North	1-North	yes	KWS_m2937	GGCCAAACATAGCCAGCTTA
Chromosome_1	218,834	North	1-North	yes	KWS_m3793	GAGAACGGGAGTGGAATGAAC
Chromosome_1	15,717,829	South	1-South	yes	KWS_m3888	TCTTTGTTGGAATTTCTCAGG
Chromosome_2	54,934,037	South	2-North	no	KWS_m2759	TTCCAGTCTCGTCTCTTTCACA
Chromosome_2	54,930,551	South	2-North	no	KWS_m4860	CCTTAGAGCACCCACAAATGA
Chromosome_2	2,092,719	North	2-South	no	KWS_m3192	TGAGAGAGGGAAACCTCCAAT
Chromosome_3	341,073	North	3-North	yes	KWS_m4507	CTTCTCCTGACCCAGATACCC
Chromosome_3	309,368	North	3-North	yes	KWS_m4995	GGGGTGTTGATGTTGCTGTAT
Chromosome_3	53,402,665	South	3-South	yes	KWS_m2641	GAGAAAGACCAAAAAGATGCAGA
Chromosome_4	60,902,981	South	4-North	no	KWS_m4363	CGCTGGACGTGAGAGTTAGAG
Chromosome_4	203,419	North	4-South	no	KWS_m5057	GGTATTGATGGGGTGAAGGTT
Chromosome_5	59,033,483	South	5-North	no	KWS_m4394	AGTGCCCTCACAACTCCATC
Chromosome_5	59,027,813	South	5-North	no	KWS_m4890	ACTCAACAAAGGGGCATCAC
Scaffold_5	1,244,762	North	5-South	no	KWS_m3442	TTCCTCTTCTCCCAACAACCT
Scaffold_5	1,167,474	North	5-South	no	KWS_m4060	TGAATCTTCCCCAGACCATC
Chromosome_6	65,032,325	South	6-North	no	KWS_m4895	CGGTGGAGCGAGTTTTAGAG
Chromosome_6	551,047	North	6-South	no	KWS_m4682	GGTGACATCCAACTCCGCTAC
Chromosome_7	21,678,063	South	7-North	no	KWS_m4047	ACACAACCGCATTCTCTTCC
Chromosome_7	81,844	North	7-South	no	KWS_m4448	TGAGAGCTGGAACAAACAAGA
Chromosome_8	56,578,861	South	8-North	no	KWS_m2221	CCATAGTGGTGGTGCTTTTCA
Chromosome_8	866,773	North	8-South	no	KWS_m3801	CGGAGAGCAGAGCATTACTTC
Chromosome_9	50,453,056	South	9-North	no	KWS_m4595	TGTTGCGATTCCTGTGCAT
Chromosome_9	34,138,985	uncertain	9-South	no	KWS_m3315	TGGCCTTGACATACTTCCAAC

[1] From Paesold et al. (2012) Supplemental Table, based on RefBeet.

Table 6: The Y-R-B linkage group in the EL10.1 genome assembly.

Gene	ID	MAKER Inferred Annotation	EL10.1_start..stop (strand)	cM[1]	Mb	Mb/cM
Y	EL10Ac2g04466	MYB114	49,675,759..49,679,064 (plus)			
				7.4	2.4	0.32
R	EL10Ac2g04268	Geraniol 8-hydroxylase	47,304,905..47,309,543 (minus)			
				17.3	14.7	0.85
B	EL10Ac2g03535	Response regulator-like PRR73	32,610,511..32,619,244 (minus)			

[1] From Abe (1993), Goldman and Austin (2000)

Table 7: Gene models detected via Benchmarking Universal Single-Copy Orthologs (BUSCO).

BUSCO	EL10.1	RefBeet 1.1	TAIR 10
Complete	1,251	1,302	1,414
Complete and single-copy	1,223	1,268	1,401
Complete and duplicated	28	34	13
Fragmented	36	37	7
Missing	153	101	19
Total groups searched	1,440	1,440	1,440
% Missing	10.6	7.0	1.3

Table 8: Distribution of MAKER annotations across the EL10.1 genome assembly.

Location	# genes	# hypothetical	%	Unique	%
Total	24,255	3,940	16.24	13,220	54.50
Chromosome 1	2,393	354	14.79	1,375	57.46
Chromosome 2	2,525	383	15.17	1,379	54.61
Chromosome 3	2,574	372	14.45	1,458	56.64
Chromosome 4	2,903	463	15.95	1,410	48.57
Chromosome 5	2,687	422	15.71	1,452	54.04
Chromosome 6	2,681	446	16.64	1,478	55.13
Chromosome 7	2,487	385	15.48	1,380	55.49
Chromosome 8	2,391	353	14.76	1,316	55.04
Chromosome 9	2,387	338	14.16	1,320	55.30
Scaffold 1	117	54	46.15	55	47.01
Scaffold 2	141	60	42.55	65	46.10
Scaffold 3	87	39	44.83	48	55.17
Scaffold 4	94	38	40.43	46	48.94
Scaffold 5	162	22	13.58	121	74.69
Scaffold 6	79	28	35.44	36	45.57
Scaffold 7	119	50	42.02	54	45.38
Scaffold 8	91	15	16.48	66	72.53
Scaffold 9	35	15	42.86	16	45.71
Scaffold 10	25	10	40.00	15	60.00
Scaffold 11	19	12	63.16	3	15.79
Scaffold 12	46	10	21.74	33	71.74
Scaffold 13	21	10	47.62	11	52.38
Scaffold 14	30	6	20.00	15	50.00
Scaffold 15	52	10	19.23	20	38.46
Scaffold 16	5	3	60.00	2	40.00
Scaffold 17	3	1	33.33	1	33.33
Scaffold 18	15	2	13.33	9	60.00
Scaffold 19	30	17	56.67	5	16.67
Scaffold 20	16	9	56.25	7	43.75
Scaffold 21	5	3	60.00	0	0.00
Scaffold 22	10	4	40.00	6	60.00
Scaffold 24	2	1	50.00	2	100.00
Scaffold 25	1	0	0.00	1	100.00
Scaffold 26	12	3	25.00	9	75.00
Scaffold 27	4	2	50.00	0	0.00
Scaffold 28	6	0	0.00	6	100.00
Chromosome mean	2,558.7	390.7		1,396.4	
Chromosome stdev	173.8	43.7		58.2	
Scaffold mean	43.8	15.1		23.3	
Scaffold stdev	47.6	17.5		28.8	

Table 9: Comparison of MapMan4 first-order functional classifications for EL10.1 and four other Tracheophytes

				Mean Gene Count per Leaf Bin									
Bincode	Top level bin	# MapMan4 leaf bins	# EL10.1 leaf bins	*At*	*Os*	*Bd*	*Sl*	*Me*	Angiosperm Mean	std dev	EL10 mean	std dev	EL10 mean / Angiosperm mean
1	Photosynthesis	226	223	1.29	1.71	1.78	1.54	1.50	**1.57**	1.26	**1.22**	0.94	78.2
2	Cellular respiration	136	135	1.81	1.73	1.47	1.73	1.86	**1.72**	1.17	**0.93**	0.84	54.1
3	Carbohydrate metabolism	92	92	2.53	2.51	2.30	2.64	3.05	**2.61**	2.14	**1.83**	1.65	70.0
4	Amino acid metabolism	135	134	1.77	1.84	1.70	1.81	2.04	**1.83**	1.27	**1.46**	1.03	79.9
5	Lipid metabolism	173	173	2.57	2.73	2.55	2.81	2.88	**2.71**	3.02	**1.95**	2.10	71.9
6	Nucleotide metabolism	53	53	1.94	1.81	1.77	1.83	2.02	**1.88**	1.20	**1.47**	0.93	78.5
7	Coenzyme metabolism	158	158	1.40	1.39	1.41	1.43	1.53	**1.43**	0.89	**1.27**	0.85	88.8
8	Polyamine metabolism	12	12	2.08	2.00	1.50	2.08	2.67	**2.07**	1.78	**1.58**	1.38	76.6
9	Secondary metabolism	93	90	2.48	1.88	1.63	2.00	2.34	**2.07**	3.76	**1.32**	1.98	64.0
10	Redox homeostasis	47	47	2.64	2.64	2.45	2.91	3.11	**2.75**	2.56	**2.13**	2.23	77.4
11	Phytohormones	140	140	4.19	3.72	3.65	4.26	5.42	**4.25**	5.02	**2.36**	2.67	55.6
12	Chromatin organisation	113	113	2.76	2.73	2.96	3.16	2.93	**2.91**	3.23	**2.15**	2.23	74.0
13	Cell cycle	258	258	1.74	1.59	1.66	1.67	1.85	**1.70**	1.64	**1.34**	1.03	78.6
14	DNA damage response	67	67	1.25	1.16	1.31	1.24	1.24	**1.24**	0.69	**1.16**	0.86	93.8
15	RNA biosynthesis	295	295	7.84	7.81	7.89	8.70	9.26	**8.30**	21.32	**5.28**	13.15	63.6
16	RNA processing	328	327	1.52	1.50	1.54	1.58	1.65	**1.56**	1.10	**1.26**	0.93	81.0
17	Protein biosynthesis	328	328	1.92	1.98	1.84	1.97	2.00	**1.94**	1.24	**1.45**	0.93	74.9
18	Protein modification	299	299	4.98	5.68	5.13	4.90	6.11	**5.36**	10.08	**3.71**	5.90	69.2
19	Protein degradation	187	187	5.58	5.81	5.87	5.83	6.26	**5.87**	20.35	**3.98**	13.21	67.9
20	Cytoskeleton	107	107	2.87	2.35	2.39	2.63	3.14	**2.67**	3.09	**1.90**	1.85	70.9
21	Cell wall	126	126	4.64	4.24	3.90	4.29	5.10	**4.43**	7.22	**2.71**	4.24	61.2
22	Vesicle trafficking	212	212	2.60	2.37	2.27	2.54	3.00	**2.55**	3.14	**1.91**	2.09	74.8
23	Protein translocation	135	135	1.48	1.44	1.53	1.57	1.83	**1.57**	1.04	**1.19**	0.72	75.9
24	Solute transport	174	173	6.58	7.14	6.73	7.33	7.95	**7.15**	10.81	**5.38**	7.95	75.3
25	Nutrient uptake	52	50	3.18	2.54	2.66	2.68	3.52	**2.92**	2.99	**1.96**	2.38	67.2
26	External stimuli response	111	111	3.25	2.28	2.18	2.82	3.63	**2.83**	6.57	**1.77**	2.94	62.7
27	Multi-process regulation	38	38	3.66	3.55	3.45	3.82	4.26	**3.75**	4.22	**2.39**	1.88	63.9
50	Enzyme classification	50	44	26.64	42.70	33.25	41.52	39.84	**36.79**	76.72	**25.66**	50.56	69.7
	total	4145	4127 (99.6%)										
	mean			3.43	3.60	3.41	3.69	4.04	3.69	12.19	2.54	7.87	69.0
	stdev			10.03	13.31	11.38	12.57	13.33	12.43		7.87		

Table 10: Comparison of MapMan4 leaf bins where no gene was predicted by MAKER in EL10.1 and five Tracheophytes.

Bincode	Description	*At*	*Os*	*Bd*	*Sl*	*Me*	Angiosperm Mean
1.1.1.1.5	Photosynthesis.photophosphorylation.photosystem II.LHC-II complex.LHCq component	1	1	1	2	1	1.2
1.1.1.2.1.5.2	Photosynthesis.photophosphorylation.photosystem II.PS-II complex.reaction center complex.cytochrome b559 heterodimer.beta component PsbF	1	1	3	1	1	1.4
1.1.1.2.1.6	Photosynthesis.photophosphorylation.photosystem II.PS-II complex.reaction center complex.component PsbI	1	1	2	3	2	1.8
1.1.1.2.4	Photosynthesis.photophosphorylation.photosystem II.PS-II complex.component PsbJ	1	1	5	1	2	2.0
1.1.1.2.5	Photosynthesis.photophosphorylation.photosystem II.PS-II complex.component PsbK	1	5	1	1	1	1.8
1.1.1.2.6	Photosynthesis.photophosphorylation.photosystem II.PS-II complex.component PsbL	1	1	6	1	1	2.0
1.1.1.2.7	Photosynthesis.photophosphorylation.photosystem II.PS-II complex.component PsbM	1	1	2	2	1	1.4
1.1.1.2.8	Photosynthesis.photophosphorylation.photosystem II.PS-II complex.component PsbN	1	1	3	1	1	1.4
1.1.1.2.10	Photosynthesis.photophosphorylation.photosystem II.PS-II complex.component PsbTc	1	1	4	1	1	1.6
1.1.1.3.9	Photosynthesis.photophosphorylation.photosystem II.assembly and maintenance.Psb27 protein	1	1	1	1	1	1.0
1.1.1.6.1.2	Photosynthesis.photophosphorylation.photosystem II.LHC-related protein groups.one-helix LHC-related protein group.OHP2 protein	1	1	1	2	2	1.4
1.1.2.5	Photosynthesis.photophosphorylation.cytochrome b6/f complex.component PetG/V	1	1	6	1	1	2.0
1.1.2.6	Photosynthesis.photophosphorylation.cytochrome b6/f complex.component PetL/VI	1	1	3	2	1	1.6
1.1.2.9.2.1	Photosynthesis.photophosphorylation.cytochrome b6/f complex.assembly.CCS cytochrome f/c6 maturation system (system II).CcsA component	1	2	2	6	2	2.6
1.1.2.9.3	Photosynthesis.photophosphorylation.cytochrome b6/f complex.assembly.HCF153 factor	1	1	1	1	1	1.0
1.1.4.1.1	Photosynthesis.photophosphorylation.photosystem I.LHC-I complex.LHCa1-type component	1	1	1	2	2	1.4
1.1.4.2.9	Photosynthesis.photophosphorylation.photosystem I.PS-I complex.component PsaI	1	1	5	2	1	2.0
1.1.4.2.10	Photosynthesis.photophosphorylation.photosystem I.PS-I complex.component PsaJ	1	1	3	2	1	1.6
1.1.5.2.1	Photosynthesis.photophosphorylation.linear electron flow.ferredoxin-NADP reductase (FNR) activity.ferredoxin-NADP oxidoreductase	2	2	2	2	3	2.2
1.1.8.1.1.7	Photosynthesis.photophosphorylation.chlororespiration.NADH dehydrogenase-like (NDH) complex.membrane subcomplex M.NdhG component	1	3	5	4	1	2.8
1.1.8.1.2.1	Photosynthesis.photophosphorylation.chlororespiration.NADH dehydrogenase-like (NDH) complex.subcomplex A.NdhH component	1	3	1	1	1	1.4
1.1.8.1.2.2	Photosynthesis.photophosphorylation.chlororespiration.NADH dehydrogenase-like (NDH) complex.subcomplex A.NdhI component	1	2	2	4	1	2.0
1.1.8.1.4.1	Photosynthesis.photophosphorylation.chlororespiration.NADH dehydrogenase-like (NDH) complex.lumen subcomplex L.PnsL1 component	1	1	1	1	1	1.0
1.1.9.1.4	Photosynthesis.photophosphorylation.ATP synthase complex.membrane CF0 subcomplex.subunit c	1	3	1	1	2	1.6
2.4.1.4.1.5	Cellular respiration.oxidative phosphorylation.NADH dehydrogenase complex.non-core components.alpha subcomplex.NDUFA9 component	1	1	1	1	2	1.2
2.4.1.4.2.3	Cellular respiration.oxidative phosphorylation.NADH dehydrogenase complex.non-core components.beta subcomplex.NDUFB7 component	1	1	1	1	2	1.2
2.4.3.6	Cellular respiration.oxidative phosphorylation.cytochrome c reductase complex.QCR7 component	2	3	2	1	2	2.0
3.1.2.5	Carbohydrate metabolism.sucrose metabolism.synthesis.cytosolic phosphoglucomutase	2	1	1	1	2	1.4
3.2.2.1.3	Carbohydrate metabolism.starch metabolism.degradation.phosphorylation.ESV1 dikinase regulator	1	1	1	1	2	1.2
3.2.2.4.2	Carbohydrate metabolism.starch metabolism.degradation.sugar translocation.glucose transporter	1	1	1	1	2	1.2
3.8.9.2.1	Carbohydrate metabolism.nucleotide sugar biosynthesis.UDP-N-acetylglucosamine synthesis.salvage biosynthesis.N-acetylglucosamine kinase	1	2	2	2	3	2.0
4.1.1.1.1.2	Amino acid metabolism.biosynthesis.glutamate family.glutamate-derived amino acids.ornithine.N-acetylglutamate kinase	1	2	2	1	2	1.6
4.1.2.2.3.3	Amino acid metabolism.biosynthesis.aspartate family.aspartate-derived amino acids.lysine.LL-diaminopimelate aminotransferase	1	1	2	1	1	1.2
5.7.3.1	Lipid metabolism.lipid degradation.fatty acid degradation.peroxisomal long-chain acyl-CoA synthetase	2	2	2	1	3	2.0
5.8.1.3	Lipid metabolism.lipid transport.plastidial lipid import.TGD5 lipid trafficking cofactor	1	1	1	1	1	1.0
6.2.4.3	Nucleotide metabolism.pyrimidines.salvage pathway.ribokinase	1	1	1	1	1	1.0
7.2.1.2	Coenzyme metabolism.thiamine pyrophosphate synthesis.hydroxymethylpyrimidine diphosphate synthesis.bifunctional hydroxymethylpyrimidine kinase a	1	1	1	1	1	1.0
7.12.1.2	Coenzyme metabolism.tetrapyrrol biosynthesis.5-aminolevulinic acid formation.glutamyl-tRNA reductase	3	1	2	2	2	2.0
9.1.2.3	Secondary metabolism.terpenoids.methylerythritol phosphate pathway.DXR 1-deoxy-D-xylulose 5-phosphate reductase	1	1	1	1	2	1.2
9.1.3.4.5	Secondary metabolism.terpenoids.terpenoid synthesis.carotenoid metabolism.LCY-e lycopene epsilon cyclase	1	1	1	1	1	1.0
9.2.2.8.1	Secondary metabolism.phenolics.flavonoid synthesis and modification.isoflavonoids.isoflavone synthase	1	1	1	1	4	1.6
9.2.3.2	Secondary metabolism.phenolics.regulation of key enzymes.KFB-CHS proteolytic chalcone synthase regulator	1	2	1	1	1	1.2
10.6.2	Redox homeostasis.cytosol/mitochondrion/nucleus redox homeostasis.O-type thioredoxin	2	1	1	1	1	1.2
11.3.1.3	Phytohormones.brassinosteroid.synthesis.steroid 5-alpha-reductase (DET2)	1	1	1	1	1	1.0
11.5.2.2	Phytohormones.ethylene.perception and signal transduction.protein kinase (CTR1)	1	2	2	3	2	2.0
11.8.2.1	Phytohormones.salicylic acid.perception and signal transduction.NPR3/4 receptor protein	2	2	2	2	2	2.0
11.10.1.4.1	Phytohormones.signalling peptides.NCRP (non-cysteine-rich-peptide) category.CIF family.CIF precursor polypeptide	2	2	3	1	1	1.8
11.10.1.5.1	Phytohormones.signalling peptides.NCRP (non-cysteine-rich-peptide) category.IDL family.IDA/IDL precursor polypeptide	6	1	1	3	9	4.0
11.10.1.6.1	Phytohormones.signalling peptides.NCRP (non-cysteine-rich-peptide) category.DVL/ROT family.DVL/RTFL precursor polypeptide	24	7	15	3	22	14.2
11.10.2.3.3	Phytohormones.signalling peptides.CRP (cysteine-rich-peptide) category.EPF/EPFL family.TMM peptide receptor	2	1	2	3	5	2.6
12.4.1.1.1	Chromatin organisation.chromatin remodeling complexes.ATPase core components.Snf2-like group.Alc chromatin remodeling factor	1	1	1	1	1	1.0
12.4.1.4.2	Chromatin organisation.chromatin remodeling complexes.ATPase core components.SSO1653-like group.Mot1 chromatin remodeling factor	1	2	1	1	1	1.2
12.4.1.5.3	Chromatin organisation.chromatin remodeling complexes.ATPase core components.Rad5/16-like group.Ris1 chromatin remodeling factor	5	3	4	1	2	3.0
13.1.1.5	Cell cycle.regulation.cyclins.CYCH-type cyclin	1	1	1	2	1	1.2
13.1.1.9	Cell cycle.regulation.cyclins.SDS-type cyclin	1	1	1	1	1	1.0
13.2.2.3.1.1	Cell cycle.interphase.DNA replication.elongation.DNA polymerase alpha complex.POLA1 catalytic component	1	1	1	1	1	1.0
13.2.2.3.4.2.4	Cell cycle.interphase.DNA replication.elongation.DNA-tracking platform.PCNA sliding clamp loader complex.RFC4 component	1	1	1	1	1	1.0
13.3.2	Cell cycle.mitosis and meiosis.TPX2 prospindle assembly factor	1	1	1	3	1	1.4
13.3.3.1.3	Cell cycle.mitosis and meiosis.chromosome segregation.centromere assembly and maintenance.KNL2/Mis18 CENH3 recruitment factor	1	2	2	1	1	1.4
13.3.5.5.1	Cell cycle.mitosis and meiosis.sister chromatid separation.spindle assembly checkpoint machinery.BUB1 checkpoint protein	1	1	1	1	1	1.0
13.3.6.2.1.1	Cell cycle.mitosis and meiosis.meiotic recombination.meiotic double strand break initiation.meiotic topoisomerase-VI complex.component a (SPO11)	2	2	1	1	2	1.6
13.3.6.2.5	Cell cycle.mitosis and meiosis.meiotic recombination.meiotic double strand break initiation.accessory protein (DFO)	1	2	1	1	1	1.2
13.3.6.5.1.7	Cell cycle.mitosis and meiosis.meiotic recombination.meiotic crossover.class I interference-sensitive crossover pathway.double Holliday junction resolvir	1	1	1	1	1	1.0
13.3.6.5.2.2.2	Cell cycle.mitosis and meiosis.meiotic recombination.meiotic crossover.class II interference-insensitive crossover pathway.MUS81-independent pathway	1	1	1	1	1	1.0
13.3.6.5.4.3	Cell cycle.mitosis and meiosis.meiotic recombination.meiotic crossover.FANCM-MHF DNA remodeling complex.MHF2 component	1	1	1	1	1	1.0
13.4.1.5	Cell cycle.cytokinesis.preprophase microtubule organization.SABRE microtubule orientation-stabilizing factor	1	1	2	2	2	1.6
13.4.4.1.1	Cell cycle.cytokinesis.phragmoplast disassembly.NACK-PQR signalling pathway.NACK microtubule-destabilizing kinesin	2	1	3	1	1	1.6
13.5.2.2.2	Cell cycle.organelle machineries.organelle fission.plastid division.MinD FtsZ assembly factor	1	1	1	1	1	1.0
14.5.2.4	DNA damage response.DNA repair mechanisms.base excision repair (BER).bifunctional DNA glycosylase/lyase (ROS1)	2	3	4	2	2	2.6
14.5.2.10.2	DNA damage response.DNA repair mechanisms.base excision repair (BER).apurinic/apyrimidinic (AP) endonuclease activities.APE2 AP-endonuclease	1	1	1	2	1	1.2
14.5.5.1.1	DNA damage response.DNA repair mechanisms.nonhomologous end-joining repair (NHEJ).Ku70-Ku80 helicase complex.KU70 component	1	1	1	1	2	1.2
15.3.2.7.2	RNA biosynthesis.RNA polymerase II-dependent transcription.pre-initiation complex.TATA box-binding protein (TBP) regulation.TBP-associated factor (M	1	2	1	1	1	1.2
15.3.6.2.3	RNA biosynthesis.RNA polymerase II-dependent transcription.MEDIATOR transcription co-activator complex.middle module.MED9 component	2	1	1	2	2	1.6
15.3.7.4	RNA biosynthesis.RNA polymerase II-dependent transcription.ELONGATOR transcription elongation complex.ELP4 component	1	1	1	1	1	1.0
15.4.5.2	RNA biosynthesis.RNA polymerase III-dependent transcription.TFIIIe transcription factor complex.RPC34 component	1	1	2	1	1	1.2
16.4.2.1	RNA processing.RNA splicing.U1/U2/U4/U5-associated Sm accessory ribonucleoprotein complex.Sm-B component	2	1	1	2	2	1.6

16.4.5.2.8	RNA processing.RNA splicing.spliceosome-associated non-snRNP MOS4-associated complex (MAC).associated components.MOS2 component	2	1	1	1	3	1.6
16.4.7.1.1	RNA processing.RNA splicing.spliceosome assembly/disassembly.RNA helicase activities.Sub2 RNA helicase	1	2	3	2	3	2.2
16.7.9.1	RNA processing.RNA modification.mRNA demethylation.ALKBH10 N6-methyladenosine demethylase	2	1	1	2	2	1.6
16.8.1.1.4	RNA processing.RNA decay.exosome complex.EXO9 core complex.RRP43 component	1	1	1	2	1	1.2
16.10.1.1.1.5	RNA processing.organelle machineries.RNA splicing.plastidial RNA splicing.group-II intron splicing.RH3 basal splicing factor	1	1	1	1	1	1.0
16.10.1.1.1.6	RNA processing.organelle machineries.RNA splicing.plastidial RNA splicing.group-II intron splicing.mTERF4 splicing factor	1	1	1	1	1	1.0
16.10.1.2.1.6	RNA processing.organelle machineries.RNA splicing.mitochondrial RNA splicing.group-II intron splicing.OTP43 splicing factor	1	1	1	1	1	1.0
16.10.1.2.2	RNA processing.organelle machineries.RNA splicing.mitochondrial RNA splicing.RUG3 splicing factor	1	1	1	1	1	1.0
16.10.3.3.4	RNA processing.organelle machineries.RNA editing.mitochondrial RNA editing.MEF7 RNA editing factor	1	1	1	1	1	1.0
16.10.3.3.7	RNA processing.organelle machineries.RNA editing.mitochondrial RNA editing.MEF10 RNA editing factor	1	1	1	1	1	1.0
16.10.3.3.8	RNA processing.organelle machineries.RNA editing.mitochondrial RNA editing.MEF11 RNA editing factor	1	1	1	1	1	1.0
16.10.3.4.6	RNA processing.organelle machineries.RNA editing.plastidial RNA editing.CRR21 RNA editing factor	1	1	1	1	1	1.0
16.10.3.4.12	RNA processing.organelle machineries.RNA editing.plastidial RNA editing.OTP85 RNA editing factor	1	1	1	1	1	1.0
17.1.1.1.38	Protein biosynthesis.cytosolic ribosome.large subunit (LSU).LSU proteome component.RPL36a component	2	4	2	2	2	2.4
17.1.1.1.48	Protein biosynthesis.cytosolic ribosome.large subunit (LSU).LSU proteome component.RPP3 component	2	2	1	1	1	1.4
17.1.1.2.12	Protein biosynthesis.cytosolic ribosome.large subunit (LSU).LSU processome component.RSA4 assembly factor (NLE)	1	2	1	1	2	1.4
17.1.2.2.21	Protein biosynthesis.cytosolic ribosome.small subunit (SSU).SSU processome.SWA3/RH36 assembly factor	1	1	1	1	1	1.0
17.1.2.2.22.3	Protein biosynthesis.cytosolic ribosome.small subunit (SSU).SSU processome.pre-40S subunit nuclear export.Rio2 kinase	1	1	1	1	1	1.0
17.6.1.2.7	Protein biosynthesis.organelle translation machineries.mitochondrial ribosome.small subunit proteome.mtRPS9 component	1	1	1	1	1	1.0
17.6.1.2.15	Protein biosynthesis.organelle translation machineries.mitochondrial ribosome.small subunit proteome.mtRPS17 component	2	1	1	1	1	1.2
17.6.2.1.14	Protein biosynthesis.organelle translation machineries.plastidial ribosome.large subunit proteome.psRPL16 component	1	1	1	1	1	1.0
17.6.2.1.18	Protein biosynthesis.organelle translation machineries.plastidial ribosome.large subunit proteome.psRPL20 component	1	3	1	4	1	2.0
17.6.2.1.20	Protein biosynthesis.organelle translation machineries.plastidial ribosome.large subunit proteome.psRPL22 component	1	4	3	1	1	2.0
17.6.2.1.21	Protein biosynthesis.organelle translation machineries.plastidial ribosome.large subunit proteome.psRPL23 component	2	7	1	3	2	3.0
17.6.2.1.27	Protein biosynthesis.organelle translation machineries.plastidial ribosome.large subunit proteome.psRPL32 component	1	1	4	3	2	2.2
17.6.2.1.31	Protein biosynthesis.organelle translation machineries.plastidial ribosome.large subunit proteome.psRPL36 component	1	1	1	1	1	1.0
17.6.2.2.3	Protein biosynthesis.organelle translation machineries.plastidial ribosome.small subunit proteome.psRPS3 component	1	5	3	1	1	2.2
17.6.2.2.8	Protein biosynthesis.organelle translation machineries.plastidial ribosome.small subunit proteome.psRPS8 component	1	4	3	1	1	2.0
17.6.2.2.15	Protein biosynthesis.organelle translation machineries.plastidial ribosome.small subunit proteome.psRPS15 component	1	4	3	3	1	2.4
17.6.2.2.16	Protein biosynthesis.organelle translation machineries.plastidial ribosome.small subunit proteome.psRPS16 component	1	1	2	1	1	1.2
17.6.2.2.19	Protein biosynthesis.organelle translation machineries.plastidial ribosome.small subunit proteome.psRPS19 component	1	1	3	1	1	1.4
18.1.6.2.1	Protein modification.N-linked glycosylation.complex N-glycan maturation.class-II glucosidase II complex.subunit alpha	1	1	1	1	1	1.0
18.1.6.3	Protein modification.N-linked glycosylation.complex N-glycan maturation.class-I alpha-mannosidase I	2	1	1	2	2	1.6
18.2.2.3	Protein modification.O-linked glycosylation.serine/threonine O-linked glycosylation.OFT1 O-fucosyltransferase	1	1	1	1	1	1.0
18.7.4.1.6	Protein modification.lipidation.Glycophosphatidylinositol (GPI)-anchor addition.GPI pre-assembly.PIG-N phosphoethanolamine transferase-I	1	1	1	1	1	1.0
18.8.1.40	Protein modification.phosphorylation.TKL kinase superfamily.RLCK-X kinase	4	3	3	2	4	3.2
18.8.5.1.1	Protein modification.phosphorylation.CAMK kinase superfamily.SNF1-related SnRK1 kinase complex.alpha-type catalytic subunit	3	4	3	2	2	2.8
18.8.9	Protein modification.phosphorylation.BUB kinase	1	1	1	1	1	1.0
19.4.1.5.4.2.2.5	Protein degradation.peptide tagging.Ubiquitin (UBQ)-anchor addition (ubiquitylation).UBQ-ligase E3 activities.Cullin-based ubiquitylation complexes.CUL3-	2	1	1	1	1	1.2
19.4.1.5.4.3.1	Protein degradation.peptide tagging.Ubiquitin (UBQ)-anchor addition (ubiquitylation).UBQ-ligase E3 activities.Cullin-based ubiquitylation complexes.CUL4-	1	1	1	2	2	1.4
19.4.6.3	Protein degradation.peptide tagging.Ubiquitin-fold-modifier (UFM)-anchor addition.UFM conjugation E2 protein	1	1	1	1	1	1.0
19.5.2.1.3	Protein degradation.peptidase families.serine-type peptidase activities.subtilisin-type protease families.SBT3 protease	15	2	2	1	2	4.4
19.5.2.5.5	Protein degradation.peptidase families.serine-type peptidase activities.chloroplast Clp-type protease complex.ClpD chaperone component	1	1	2	1	1	1.2
20.1.3.7	Cytoskeleton.microtubular network.Kinesin microtubule-based motor protein activities.Kinesin-8 motor protein	2	2	2	2	4	2.4
20.2.2.7.2	Cytoskeleton.microfilament network.actin polymerisation.actin capping protein heterodimer.beta component	1	1	1	1	2	1.2
21.3.5.1.1	Cell wall.pectin.modification and degradation.polygalacturonase activities.QRT2 polygalacturonase	3	3	2	2	3	2.6
21.3.5.1.3	Cell wall.pectin.modification and degradation.polygalacturonase activities.PGX1 polygalacturonase	2	1	1	1	2	1.4
21.6.1.1	Cell wall.lignin.monolignol synthesis.hydroxycinnamoyl-CoA:quinate/shikimate O-hydroxycinnamoyltransferase (HCT)	1	2	2	1	2	1.6
21.6.1.7	Cell wall.lignin.monolignol synthesis.caffeic acid O-methyltransferase (COMT)	1	1	1	1	2	1.2
21.6.2.1	Cell wall.lignin.monolignol conjugation and polymerization.p-coumaroyl-CoA:monolignol transferase (PMT)	1	3	1	1	1	1.4
21.9.3.3	Cell wall.cutin and suberin.biosynthesis regulation.CFL regulator	2	1	1	1	2	1.4
22.1.3.3	Vesicle trafficking.clathrin coated vesicle (CCV) machinery.AP-2 cargo adaptor complex.AP2M medium mu subunit	1	1	1	2	1	1.2
22.7.7.1.1	Vesicle trafficking.target membrane tethering.TRAPP (Trafficking-Protein-Particle) complexes.TRAPP-I/II/III complex-shared components.BET5 componer	1	1	1	1	1	1.0
23.1.3.5	Protein translocation.chloroplast.inner envelope TIC translocation system.Tic40 component	1	1	1	1	1	1.0
23.1.4.3	Protein translocation.chloroplast.inner envelope Sec2 post-import insertion system.SecE2 component	1	1	1	1	1	1.0
23.1.5.1	Protein translocation.chloroplast.thylakoid membrane Sec1 translocation system.SecA1 component	1	1	1	1	1	1.0
23.2.1.1	Protein translocation.mitochondrion.outer mitochondrion membrane TOM translocation system.Tom5 component	1	1	2	1	1	1.2
23.2.1.3	Protein translocation.mitochondrion.outer mitochondrion membrane TOM translocation system.Tom7 component	2	1	2	2	2	1.8
23.2.1.5	Protein translocation.mitochondrion.outer mitochondrion membrane TOM translocation system.Tom20 component	4	1	1	2	4	2.4
23.2.1.7	Protein translocation.mitochondrion.outer mitochondrion membrane TOM translocation system.OM64 component	1	1	1	1	1	1.0
23.3.2.3.1	Protein translocation.endoplasmic reticulum.GET post-translational insertion system.GET4-GET5 scaffold subcomplex.GET4 GET3-recruitment componer	1	1	1	2	1	1.2
23.5.1.1.6.3	Protein translocation.nucleus.nucleocytoplasmic transport.nuclear pore complex (NPC).central subcomplex.NUP58 nucleoporin	1	2	2	1	1	1.4
24.2.1.7	Solute transport.carrier-mediated transport.DMT superfamily.TPPT-type solute transporter	4	5	4	3	2	3.6
24.2.9.2.1	Solute transport.carrier-mediated transport.CDF superfamily.CDF family.iron/zinc cation transporter (Fe/Zn-CDF-type)	1	1	1	2	1	1.2
25.5.2.3	Nutrient uptake.copper uptake.reduction-based uptake.CCH copper chaperone	2	2	2	2	1	1.8
26.1.2.2.3	External stimuli response.light.UV-A/blue light.phototropin-mediated photoperception.PKS phototropin signalling factor	4	3	3	3	6	3.8
26.2.1.5	External stimuli response.gravity.sensing and signalling.SCR transcription factor	1	2	1	1	2	1.4
26.3.2.5.3	External stimuli response.temperature.Hsp (heat-shock-responsive protein) families.sHsp (small heat-shock-responsive protein) families.class-C-III prote	1	1	1	1	1	1.0
26.3.2.5.10	External stimuli response.temperature.Hsp (heat-shock-responsive protein) families.sHsp (small heat-shock-responsive protein) families.class-PX proteir	1	1	1	1	1	1.0
26.5.1.1.1	External stimuli response.salinity.SOS (Salt Overly Sensitive) signalling pathway.SOS3-SOS2 signalling.SOS3 calcium sensor component	1	2	1	3	3	2.0
26.6.2.2.1.2	External stimuli response.biotic stress.pathogen effector.ETI (effector-triggered immunity) network.RIN4-RPM1 immune signalling.RIPK RIN4-protein kina	2	1	1	3	5	2.4
26.6.4.4	External stimuli response.biotic stress.systemic acquired resistance (SAR).FMO1 pipecolate N-hydroxylase	1	1	1	1	2	1.2
26.6.5.2	External stimuli response.biotic stress.tobamovirus multiplication.TOM2A replication host factor	1	1	1	1	1	1.0
26.6.6.1.6.1	External stimuli response.biotic stress.symbiont-associated response.symbiosis signalling pathway.NSP1-NSP2 nodulation initiation complex.NSP1 comp	1	1	1	2	2	1.4
26.6.6.1.6.2	External stimuli response.biotic stress.symbiont-associated response.symbiosis signalling pathway.NSP1-NSP2 nodulation initiation complex.NSP2 comp	1	1	1	1	2	1.2
27.1.4.4	Multi-process regulation.circadian clock.morning element regulation.TZP repression factor	1	1	1	1	2	1.2
27.1.6	Multi-process regulation.circadian clock.TIC circadian clock regulator	2	1	1	1	2	1.4
27.3.1.1	Multi-process regulation.SnRK1 metabolic regulator system.SnRK1 kinase complex.alpha catalytic subunit	3	4	3	2	2	2.8

Table 11: Comparison of over- and under-represented MapMan4 second-order functional classifications for EL10.1 and four other Tracheophytes

			gene count									EL10 mean / Angiosperm mean (%)	
								Angiosperm					
Bincode	Top level bin	Second level bin	*At*	*Os*	*Bd*	*Sl*	*Me*	Mean	std dev	EL10 mean	std dev	'under-represented'	'over-represented'
2.2	Cellular respiration	pyruvate oxidation	1.8	2.4	2.2	2.0	1.8	**2.04**	0.68	**1.20**	0.45	58.8	-
3.3	Carbohydrate metabolism	trehalose metabolism	5.0	4.7	4.0	3.3	4.7	**4.33**	4.08	**2.33**	2.31	53.8	-
3.4	Carbohydrate metabolism	raffinose family oligosaccharide biosynthesis	3.3	1.0	1.3	2.7	4.0	**2.47**	2.47	**2.33**	0.58	-	94.6
3.8	Carbohydrate metabolism	nucleotide sugar biosynthesis	2.7	2.3	2.1	3.3	2.8	**2.65**	1.97	**1.46**	0.93	55.0	-
3.9	Carbohydrate metabolism	fermentation	2.0	3.3	2.3	2.7	4.3	**2.93**	1.58	**1.67**	0.58	56.8	-
5.9	Lipid metabolism	lipid bodies-associated activities	6.7	4.7	5.0	4.5	5.2	**5.20**	3.56	**3.00**	1.79	57.7	-
7.10	Coenzyme metabolism	FMN/FAD biosynthesis	1.1	1.3	1.5	1.3	1.2	**1.27**	0.56	**1.18**	0.40	-	92.9
7.1	Coenzyme metabolism	molybdenum cofactor synthesis	1.0	1.1	1.0	1.0	1.0	**1.03**	0.17	**1.00**	0.00	-	97.2
7.11	Coenzyme metabolism	iron-sulfur cluster assembly machineries	1.5	1.4	1.5	1.5	1.4	**1.43**	0.78	**1.35**	0.98	-	95.0
7.13	Coenzyme metabolism	phylloquinone synthesis	1.3	0.9	1.4	1.3	1.3	**1.20**	0.61	**1.13**	0.35	-	93.8
7.14	Coenzyme metabolism	lipoic acid synthesis	2.0	1.5	2.5	2.0	1.5	**1.90**	1.10	**1.00**	0.00	52.6	-
7.3	Coenzyme metabolism	S-adenosyl methionine (SAM) cycle	3.0	3.0	2.0	3.5	4.0	**3.10**	1.66	**1.50**	0.71	48.4	-
7.5	Coenzyme metabolism	tetrahydrofolate synthesis	1.8	1.4	1.4	1.6	1.9	**1.61**	1.03	**1.86**	1.61	-	115.0
7.6	Coenzyme metabolism	biotin synthesis	1.3	1.3	1.3	1.0	1.0	**1.15**	0.37	**1.25**	0.50	-	108.7
9.1	Secondary metabolism	terpenoids	2.9	2.8	2.2	3.1	3.4	**2.87**	5.17	**1.68**	2.41	58.6	-
9.3	Secondary metabolism	nitrogen-containing secondary compounds	2.3	0.7	0.9	0.6	0.7	**1.01**	1.66	**0.59**	1.18	58.2	-
10.4	Redox homeostasis	hydrogen peroxide removal	4.9	4.7	4.1	4.9	4.7	**4.66**	2.81	**4.29**	3.55	-	92.0
10.5	Redox homeostasis	chloroplast redox homeostasis	2.2	2.0	1.9	2.5	2.5	**2.22**	1.62	**1.27**	0.90	57.4	-
11.1	Phytohormones	signalling peptides	8.4	6.2	6.4	6.4	10.8	**7.65**	8.33	**3.23**	4.08	42.3	-
11.5	Phytohormones	ethylene	3.8	4.1	3.7	5.1	4.3	**4.20**	2.73	**2.44**	1.59	58.2	-
11.6	Phytohormones	gibberellin	2.6	3.2	3.2	3.2	5.5	**3.55**	2.98	**1.69**	1.03	47.6	-
11.6	Phytohormones	jasmonic acid	2.0	1.9	1.9	2.8	2.7	**2.27**	1.68	**1.14**	0.77	50.3	-
11.8	Phytohormones	salicylic acid	1.7	0.7	0.7	0.9	1.4	**1.09**	0.92	**0.29**	0.49	26.3	-
14.1	DNA damage response	DNA damage sensing and signalling	1.0	1.0	1.0	1.3	1.8	**1.20**	0.70	**1.25**	0.50	-	104.2
14.4	DNA damage response	DNA repair polymerase activities	1.0	0.7	1.0	1.0	1.0	**0.94**	0.24	**1.00**	0.58	-	106.1
14.5	DNA damage response	DNA repair mechanisms	1.2	1.2	1.3	1.2	1.2	**1.23**	0.58	**1.16**	0.93	-	94.2
15.2	RNA biosynthesis	RNA polymerase I-dependent transcription	1.4	1.4	1.1	1.6	1.6	**1.43**	0.61	**1.71**	0.95	-	120.0
15.4	RNA biosynthesis	RNA polymerase III-dependent transcription	1.9	1.6	1.9	1.2	1.2	**1.56**	0.88	**1.50**	1.40	-	96.3
15.5	RNA biosynthesis	siRNA biogenesis	1.9	1.8	2.2	3.1	1.8	**2.16**	2.03	**2.00**	1.63	-	92.6
15.6	RNA biosynthesis	rRNA biogenesis	1.3	1.3	1.3	1.0	1.3	**1.20**	0.41	**1.25**	0.50	-	104.2
15.8	RNA biosynthesis	transcriptional repression	4.5	4.8	5.0	10.5	8.5	**6.65**	4.17	**2.75**	1.26	41.4	-
16.5	RNA processing	ribonuclease activities	1.9	2.3	2.2	2.4	2.2	**2.22**	1.52	**2.00**	1.32	-	90.0
16.6	RNA processing	RNA editing	1.0	1.0	1.5	1.0	1.0	**1.10**	0.32	**2.00**	0.94	-	181.8
17.2	Protein biosynthesis	aminoacyl-tRNA synthetase activities	2.2	2.5	2.3	2.0	2.2	**2.25**	1.10	**2.12**	1.13	-	94.3
18.1	Protein modification	N-linked glycosylation	1.4	1.3	1.3	1.3	1.6	**1.39**	0.90	**1.33**	0.77	-	95.9
18.14	Protein modification	peptide maturation	1.9	2.0	2.0	1.7	2.2	**1.93**	1.45	**1.75**	1.48	-	90.5
18.4	Protein modification	disulfide bond formation	1.3	1.1	1.2	1.2	1.2	**1.22**	0.42	**1.11**	0.33	-	90.9
18.5	Protein modification	ADP-ribosylation	2.5	2.0	2.5	3.5	2.5	**2.60**	1.26	**4.00**	1.41	-	153.8
18.9	Protein modification	tyrosine sulfation (only 1 entry)	1.0	1.0	1.0	1.0	1.0	**1.00**	0.00	**1.00**	NA	-	100.0
20.3	Cytoskeleton	actin and tubulin folding	1.1	1.2	1.2	1.2	1.5	**1.21**	0.41	**1.10**	0.31	-	90.9
21.2	Cell wall	hemicellulose	3.2	4.5	3.9	3.0	3.7	**3.65**	3.32	**2.04**	1.86	55.8	-
21.4	Cell wall	cell wall proteins	7.8	7.5	7.1	6.1	8.2	**7.33**	8.57	**4.28**	4.11	58.3	-
21.6	Cell wall	lignin	2.8	3.3	3.2	3.5	3.6	**3.29**	3.15	**1.77**	1.92	53.7	-
21.9	Cell wall	cutin and suberin	3.3	2.2	2.3	2.7	3.2	**2.71**	2.26	**1.57**	1.34	57.7	-
26.4	External stimuli response	drought	1.0	1.0	1.0	1.0	1.5	**1.10**	0.32	**1.00**	0.00	-	90.9
26.6	External stimuli response	biotic stress	3.8	1.6	1.6	2.7	3.6	**2.66**	8.73	**1.09**	1.16	41.1	-
27.2	Multi-process regulation	TOR signalling pathway	2.3	1.5	1.5	1.8	2.3	**1.85**	1.04	**1.75**	0.50	-	94.6
27.3	Multi-process regulation	SnRK1 metabolic regulator system	4.8	7.3	6.7	4.8	6.2	**5.97**	7.69	**2.50**	2.51	41.9	-
50.5	Enzyme classification	EC_5 isomerases	3.3	3.0	2.8	3.8	3.3	**3.20**	3.04	**3.50**	4.12	-	109.4
50.6	Enzyme classification	EC_6 ligases	1.8	2.4	2.0	1.6	1.6	**1.88**	2.98	**1.80**	2.49	-	95.7

Table 12: Comparison of Transcription factor classes for EL10.1 and four other Tracheophytes

		gene count								
							Angiosperm			
Bincode	Description	*At*	*Os*	*Bd*	*Sl*	*Me*	**mean**	stdev	**EL10**	**Difference**
15.7.49	FAR1 transcription factor	18	89	142	44	36	**77.8**	48.8	**106**	28.25
15.7.3.7	HB (Homeobox) superfamily.HOX-like transcription factor	3	2	2	4	4	**3.0**	1.2	**6**	3.00
15.7.27	ELF3 transcription factor	2	2	1	4	2	**2.3**	1.3	**3**	0.75
15.7.4.3	bZIP superfamily.bZIP19/23/24 transcription factor	3	3	4	2	2	**2.8**	1.0	**3**	0.25
15.7.3.10	HB (Homeobox) superfamily.SAWADEE transcription factor	2	3	2	4	3	**3.0**	0.8	**3**	0.00
15.7.45	HRT transcription factor	3	1	1	1	1	**1.0**	0.0	**1**	0.00
15.7.5.4	B3 superfamily.LAV-VAL transcription factor	3	2	2	4	4	**3.0**	1.2	**3**	0.00
15.7.51	SAP transcription factor	1	0	0	2	2	**1.0**	1.2	**1**	0.00
15.7.53	DPB3 transcription factor	2	2	2	2	2	**2.0**	0.0	**2**	0.00
15.7.13	HSF (heat shock) transcription factor	24	25	24	27	32	**27.0**	3.6	**16**	-11.00
15.7.46	PLATZ transcription factor	12	16	15	25	21	**19.3**	4.6	**8**	-11.25
15.7.5.1	B3 superfamily.ARF transcription factor	23	24	24	25	29	**25.5**	2.4	**14**	-11.50
15.7.3.13	HB (Homeobox) superfamily.zf-HD transcription factor	17	14	21	37	22	**23.5**	9.7	**11**	-12.50
15.7.7.3	AP2/ERF superfamily.AP2-type transcription factor	18	22	24	24	30	**25.0**	3.5	**12**	-13.00
15.7.16	C3H zinc finger transcription factor	57	52	49	76	68	**61.3**	12.9	**48**	-13.25
15.7.19	TCP transcription factor	24	20	21	38	35	**28.5**	9.3	**14**	-14.50
15.7.1.5	C2C2 superfamily.DOF transcription factor	36	30	29	36	44	**34.8**	6.9	**20**	-14.75
15.7.1.3	C2C2 superfamily.GATA transcription factor	30	25	27	33	35	**30.0**	4.8	**15**	-15.00
15.7.3.3	HB (Homeobox) superfamily.HD-ZIP IV transcription factor	18	12	17	49	13	**22.8**	17.6	**6**	-16.75
15.7.39	OFP transcription factor	18	31	32	25	23	**27.8**	4.4	**11**	-16.75
15.7.24	AS2/LOB transcription factor	43	36	28	52	56	**43.0**	13.2	**26**	-17.00
15.7.7.2	AP2/ERF superfamily.DREB-type transcription factor	53	35	37	45	54	**42.8**	8.7	**25**	-17.75
15.7.3.1	HB (Homeobox) superfamily.HD-ZIP I/II transcription factor	27	27	24	36	38	**31.3**	6.8	**13**	-18.25
15.7.2.2	MYB superfamily.MYB-related transcription factor	81	72	64	81	85	**75.5**	9.4	**55**	-20.50
15.7.2.3	MYB superfamily.G2-like GARP transcription factor	36	37	38	38	51	**41.0**	6.7	**18**	-23.00
15.7.7.1	AP2/ERF superfamily.ERF-type transcription factor	63	60	52	87	97	**74.0**	21.4	**38**	-36.00
15.7.12	GRAS transcription factor	34	57	63	53	78	**62.8**	11.0	**26**	-36.75
15.7.4.1	bZIP superfamily.bZIP transcription factor	76	96	87	75	81	**84.8**	9.0	**46**	-38.75
15.7.22	WRKY transcription factor	72	100	88	82	101	**92.8**	9.3	**45**	-47.75
15.7.15	C2H2 zinc finger transcription factor	106	119	107	108	158	**123.0**	24.0	**75**	-48.00
15.7.14	MADS box transcription factor	109	75	79	146	82	**95.5**	33.8	**43**	-52.50
15.7.17	NAC transcription factor	113	136	135	102	111	**121.0**	17.1	**61**	-60.00
15.7.2.1	MYB superfamily.MYB transcription factor	135	118	121	139	176	**138.5**	26.7	**71**	-67.50
15.7.33	bHLH transcription factor	172	178	162	175	207	**180.5**	19.0	**113**	-67.50
	mean of 91 classes	21.5	21.4	21.4	23.8	25.6	23.1	5.2	13.7	-9.4
	stdev of 91 classes	32.0	33.1	33.0	33.9	38.1	33.7	7.8	21.3	15.6

Table 13: Beet genome size estimates obtained by flow cytometry.

				Genome size (Mb.1C)				
Sample	Individual	Type	Replicates	Mean	Std Dev	Median	Range	CV
"5B" sugar breeding population	1	out crossed	4	749.92	5.35	750.83	12.38	0.71
"5B" sugar breeding population	2	out crossed	4	760.80	8.23	763.36	18.82	1.08
"5B" sugar breeding population	3	out crossed	4	727.35	13.40	729.47	28.81	1.84
"5B" sugar breeding population	4	out crossed	4	750.63	7.87	752.97	17.54	1.05
"5B" sugar breeding population	**combined**	**out crossed**	**16**	**747.17**	**15.07**	**749.91**	**56.83**	**2.02**
Sugar beet F1042	1	out crossed	4	745.39	7.85	747.40	17.79	1.05
Sugar beet F1042	2	out crossed	4	794.33	11.71	799.64	24.46	1.47
Sugar beet F1042	3	out crossed	4	790.83	5.98	788.76	13.24	0.76
Sugar beet F1042	4	out crossed	4	800.38	21.15	799.68	50.00	2.64
Sugar beet F1042	**combined**	**out crossed**	**16**	**782.73**	**25.39**	**788.76**	**91.59**	**3.24**
EL10 sugar beet selfed progeny	1	inbred	4	648.38	29.18	655.63	64.59	4.50
EL10 sugar beet selfed progeny	2	inbred	4	683.63	9.26	683.94	18.91	1.35
EL10 sugar beet selfed progeny	3	inbred	4	689.56	4.82	690.80	11.27	0.70
EL10 sugar beet selfed progeny	4	inbred	4	696.15	6.04	697.35	13.55	0.87
EL10 sugar beet selfed progeny	5	inbred	4	700.22	1.62	699.83	3.78	0.23
EL10 sugar beet selfed progeny	6	inbred	4	721.03	14.05	725.69	31.33	1.95
EL10 sugar beet selfed progeny	7	inbred	4	738.38	6.67	737.64	14.64	0.90
EL10 sugar beet selfed progeny	8	inbred	4	739.62	57.81	748.17	135.17	7.82
EL10 sugar beet selfed progeny	9	inbred	4	760.37	18.00	763.75	41.94	2.37
EL10 sugar beet selfed progeny	10	inbred	4	811.07	20.86	812.56	50.83	2.57
EL10 sugar beet selfed progeny	11	inbred	4	758.57	10.29	760.44	23.13	1.36
EL10 sugar beet selfed progeny	12	inbred	4	761.82	4.04	761.24	9.39	0.53
EL10 sugar beet selfed progeny	13	inbred	4	742.38	10.34	745.67	23.53	1.39
EL10 sugar beet selfed progeny	14	inbred	4	751.87	10.14	753.90	20.80	1.35
EL10 sugar beet selfed progeny	15	inbred	4	755.48	5.36	755.22	10.66	0.71
EL10 sugar beet selfed progeny	16	inbred	4	749.47	13.08	748.63	31.76	1.75
EL10 sugar beet selfed progeny	17	inbred	4	633.04	10.47	634.16	25.46	1.65
EL10 sugar beet selfed progeny	18	inbred	4	640.26	5.33	640.76	12.19	0.83
EL10 sugar beet selfed progeny	19	inbred	4	741.55	5.39	742.21	12.08	0.73
EL10 sugar beet selfed progeny	20	inbred	4	738.27	3.12	739.22	7.14	0.42
EL10 sugar beet selfed progeny	21	inbred	4	633.67	40.08	640.45	95.55	6.33
EL10 sugar beet selfed progeny	22	inbred	4	656.84	31.56	665.09	72.46	4.81
EL10 sugar beet selfed progeny	**combined**	**inbred**	**88**	**715.98**	**51.71**	**732.89**	**255.88**	**7.22**
Sugar beet	**combined**	**combined**	**120**	**729.04**	**51.21**	**741.00**	**255.88**	**7.02**
W357B table beet selfed progeny	1	inbred	4	875.51	14.94	877.06	35.37	1.71
W357B table beet selfed progeny	2	inbred	4	801.81	19.32	804.73	46.61	2.41
W357B table beet selfed progeny	3	inbred	4	679.18	22.38	677.74	51.52	3.29
W357B table beet selfed progeny	4	inbred	4	696.88	12.58	698.39	29.43	1.80
W357B table beet selfed progeny	5	inbred	4	731.51	10.82	732.63	25.64	1.48
W357B table beet selfed progeny	6	inbred	4	738.44	9.66	737.94	20.08	1.31
W357B table beet selfed progeny	7	inbred	4	748.06	17.36	747.85	42.52	2.32
W357B table beet selfed progeny	8	inbred	4	756.51	19.60	759.35	47.33	2.59
W357B table beet selfed progeny	9	inbred	4	761.98	13.85	766.23	30.31	1.82
W357B table beet selfed progeny	10	inbred	4	780.58	13.66	785.08	30.69	1.75
W357B table beet selfed progeny	11	inbred	4	700.88	22.90	708.29	51.95	3.27
W357B table beet selfed progeny	12	inbred	4	783.55	26.24	776.21	59.04	3.35
W357B table beet selfed progeny	13	inbred	4	779.73	24.70	774.50	57.85	3.17
W357B table beet selfed progeny	14	inbred	4	767.52	19.22	764.14	42.75	2.50
W357B table beet selfed progeny	15	inbred	4	725.14	4.06	725.07	7.55	0.56
W357B table beet selfed progeny	16	inbred	4	758.03	14.64	753.73	33.60	1.93
W357B table beet selfed progeny	17	inbred	4	748.50	5.19	748.19	12.53	0.69
W357B table beet selfed progeny	18	inbred	4	679.69	9.91	677.85	23.71	1.46
W357B table beet selfed progeny	19	inbred	4	669.48	8.51	672.18	19.08	1.27
W357B table beet selfed progeny	20	inbred	4	662.36	10.75	662.23	22.59	1.62
Table beet	**combined**	**inbred**	**80**	**742.27**	**52.84**	**746.38**	**240.45**	**7.12**
Inbred	**combined**	**inbred**	**168**	**728.50**	**53.73**	**737.55**	**312.53**	**7.38**
Out crossed	**combined**	**out crossed**	**32**	**764.95**	**27.35**	**756.15**	**115.25**	**3.58**
Grand Total	**combined**	**combined**	**200**	**734.33**	**52.14**	**743.68**	**312.53**	**7.10**

Table 14: Frequency of Transposable Element (TE) classes in the EL10.1 genom

A: RepeatMasker TE's

Repeat Class	TE Type	Number in EL10.1
DNA	TcMar-Stowaway	48,575
DNA	En-Spm	23,710
DNA	hAT-Tip100	17,754
DNA	MULE-MuDR	6,473
DNA	PIF-Harbinger	3,878
DNA	TcMar-Mogwai	358
DNA	MuDR	87
DNA	Maverick	46
LINE	L1	2,380
LINE	RTE-BovB	1,830
LINE	L2	968
LINE	DRE	5
LTR	Gypsy	34,588
LTR	Copia	17,342
LTR	undefined	1,657
LTR	Caulimovirus	924
LTR	ERV1	78
LTR	Ngaro	78
LTR	Pao	35
MITEs	-	4,481
RC	Helitron	3,869
rRNA	-	51
Satellite	-	5,587
Simple	-	2,227
SINE	tRNA	2,971
SINE	7SL	679
	sum	180,631
	mean	6,947
	stdev	12,050

B: Complete LTRs (LTR_Retriever)

Repeat Class	TE Type	Number in EL10.1
LTR	unknown	999
LTR	Gypsy	1,030
LTR	Copia	574

C: Major interstitial satellite sequences

Repeat Class	Name	Number in EL10.1
Gypsy CenH3	Beetle7	1,937
CenH3	pBV_I	4,029
CenH3	pBV_II	3,863
CenH3	pBV_III	4,387
CenH3	pBV_IV	3,854
CenH3	pBV_V	5,206
CenH3	pBV_VI	5,090
H3K9me2	pEV1	10,463
5S rDNA	pXV1_5S	107
35S rDNA	pZR1_18S	199

D: Major terminal satellite sequences (pAV34)

Chromsome orientation	Position in EL10.1	Number in EL10.1
Chr1 - N	34,591..49,546	38
Chr1	not found	-
Chr2-(S)	1,010..7,486	21
Chr2-(N)	40,597,947..54,946,504	22
Chr3 -N	16,918..24,744	16
Chr3-S	54,068,612..54,100,447	83
Chr4-(S)	6..15,705	42
Chr4-(N)	39,426,832..61,140,574	95
Chr5-(S) Scaffold 5	1,596,750..1,657,270	152
Chr5	6,732,312..6,741,739	19
Chr5	41,593,251..41,598,882	13
Chr5-(N)	59,198,215..59,224,585	71
Chr6-(S)	4..4,595	11
Chr6-(N)	6,5072,913..65,073,571	3
Chr7-(S)	1,002..15,434	45
Chr7	8,866,608..8,879,321	24
Chr7	21,826,741..21,836,557	26
Chr8-(S)	5..16,336	35
Chr8-(N)	43,367,644..57,938,902	101
Chr9	25,507,827..26,781,523	3
Chr9-(N)	52,020,904..52,159,048	256

Table 15: Characteristics of tandem repeats in the EL10.1 genome assembly.

Target	Length	Number of Tandem Repeats	Tandem Repeats/ Mb	Mean Copy Number	stdev	Median Copy Number	Maximum Copy Number	Distinct Types	Mean Percent Match	Mean Indels
Chromosome 1	58,086,001	36,834	634.1	7.8	35.7	2.8	3290.3	816	86.1	5.2
Chromosome 2	54,971,872	37,041	673.8	7.5	27.3	2.9	2073.0	784	85.9	5.4
Chromosome 3	54,100,447	33,183	613.4	7.8	24.1	2.8	1424.0	781	86.3	5.0
Chromosome 4	61,163,185	38,116	623.2	7.9	26.4	2.9	1441.0	843	86.2	5.2
Chromosome 5	59,224,585	37,876	639.5	7.8	23.3	2.8	1477.2	845	86.1	5.3
Chromosome 6	65,096,967	39,448	606.0	8.2	28.3	2.9	1747.0	898	86.3	5.1
Chromosome 7	57,353,724	36,137	630.1	7.5	23.3	2.8	1217.3	794	86.3	5.1
Chromosome 8	57,938,902	36,560	631.0	7.8	22.2	2.9	1136.3	840	86.2	5.2
Chromosome 9	52,180,088	32,462	622.1	8.1	26.3	2.8	1663.5	815	86.1	5.2
Chromosome mean			630.4							
Chromosome std dev			19.3							
Scaffold 1	2,519,862	1,618	642.1	7.9	18.4	2.8	373.8	238	86.1	5.1
Scaffold 2	2,420,327	1,529	631.7	8.9	21.0	2.8	383.0	261	86.4	5.1
Scaffold 3	1,921,869	1,050	546.3	6.2	12.6	2.8	223.0	169	86.4	4.9
Scaffold 4	1,802,165	1,173	650.9	8.9	30.8	2.9	661.5	201	85.9	5.4
Scaffold 5	1,679,391	799	475.8	8.3	58.2	2.9	1628.4	159	86.5	4.8
Scaffold 6	1,639,599	1,061	647.1	9.1	33.2	2.8	703.0	185	86.1	5.1
Scaffold 7	1,327,247	705	531.2	7.2	16.4	3.0	269.0	142	86.9	4.9
Scaffold 8	1,116,489	686	614.4	7.0	10.1	2.9	82.7	159	86.6	4.7
Scaffold 9	912,454	420	460.3	9.0	59.2	2.6	1190.0	97	87.0	4.7
Scaffold 10	511,739	290	566.7	10.2	23.3	2.9	230.2	98	86.7	5.2
Scaffold 11	457,683	546	1193.0	4.1	9.8	2.9	196.0	66	84.9	8.0
Scaffold 12	413,183	174	421.1	9.3	35.3	2.6	391.3	57	87.5	4.7
Scaffold 13	378,582	244	644.5	9.6	31.9	2.8	384.8	77	86.5	5.0
Scaffold 14	366,083	279	762.1	5.5	9.5	2.7	98.5	74	86.0	5.2
Scaffold 15	344,704	109	316.2	8.2	14.9	2.7	88.8	49	89.0	4.1
Scaffold 16	282,902	812	2870.3	3.3	3.6	3.0	70.9	47	84.5	9.8
Scaffold 17	281,078	61	217.0	4.5	4.9	2.4	22.2	27	90.4	2.7
Scaffold 18	278,623	108	387.6	7.2	12.2	2.6	89.0	48	88.2	4.0
Scaffold 19	267,222	181	677.3	6.0	9.8	2.7	53.0	54	88.8	4.1
Scaffold 20	264,226	131	495.8	6.9	11.5	3.1	93.0	55	85.5	6.0
Scaffold 21	232,983	107	459.3	20.1	39.1	5.4	219.3	63	81.6	5.4
Scaffold 22	195,802	96	490.3	6.3	10.3	2.7	61.5	40	88.7	4.6
Scaffold 23	155,017	24	154.8	9.1	16.5	2.6	76.2	15	93.4	2.3
Scaffold 24	136,997	260	1897.9	3.5	3.9	2.9	38.0	31	85.1	10.0
Scaffold 25	135,273	39	288.3	6.0	8.5	3.0	46.0	29	84.4	6.5
Scaffold 26	131,654	116	881.1	5.3	7.1	3.0	48.0	51	85.0	5.0
Scaffold 27	130,158	55	422.6	11.0	15.9	3.1	57.0	38	86.1	6.9
Scaffold 28	91,000	39	428.6	4.0	3.8	2.5	22.0	24	87.7	4.0
Scaffold 29	14,771	6	406.2	6.7	10.0	2.9	27.0	5	92.7	2.0
Scaffold 30	11,875	11	926.3	12.5	11.7	9.8	41.1	11	82.6	3.9
overall mean			661.0							
overall stdev			460.8							

Table 16: Read count mapping of short reads from EL10 and four other germplasms to the EL10.1 genome assembly.

		EL10 self mapping			C869_25			C869_UK			KWS2320			NK-388mm-O		
Target	Length	Mapped (%)	Median Read depth	Std Dev Read Depth	Mapped (%)	Median Read depth	Std Dev Read Depth	Mapped (%)	Median Read depth	Std Dev Read Depth	Mapped (%)	Median Read depth	Std Dev Read Depth	Mapped (%)	Median Read depth	Std Dev Read Depth
Chromosome 1	58,086,001	99.67	36	245.4	99.91	34	198.8	98.42	34	210.5	98.78	37	320.5	98.78	39	229.0
Chromosome 2	54,971,872	99.75	37	155.7	99.96	34	140.9	98.46	34	182.8	98.60	37	166.4	99.05	40	153.1
Chromosome 3	54,100,447	99.73	36	35.8	99.93	33	38.0	98.55	34	52.3	97.83	34	50.4	98.87	39	65.9
Chromosome 4	61,163,185	99.71	36	97.5	99.81	33	84.8	98.21	33	102.0	97.83	35	133.1	98.78	39	93.0
Chromosome 5	59,224,585	99.74	36	239.2	99.84	33	186.9	98.88	35	241.0	98.71	36	183.8	99.09	39	204.6
Chromosome 6	65,096,967	99.72	36	183.7	99.86	33	215.6	97.83	32	272.0	97.61	34	150.1	98.44	38	288.9
Chromosome 7	57,353,724	99.60	36	293.3	99.89	33	228.2	98.94	35	183.6	98.56	36	380.7	99.13	40	264.4
Chromosome 8	57,938,902	99.69	36	102.7	99.86	33	81.6	98.68	34	109.3	98.83	37	114.6	98.97	39	93.0
Chromosome 9	52,180,088	99.74	36	194.4	99.93	33	152.6	98.09	33	144.5	97.62	34	299.0	99.03	39	179.9
Scaffold 1	2,519,862	99.76	36	587.3	99.88	30	539.8	99.42	35	904.0	99.66	44	416.5	99.82	39	573.4
Scaffold 2	2,420,327	99.81	36	879.5	99.99	31	605.9	95.15	28	889.2	96.20	29	798.4	96.99	34	699.4
Scaffold 3	1,921,869	99.90	36	658.1	99.99	31	516.9	99.45	33	1027.7	98.52	43	385.5	98.30	39	757.8
Scaffold 4	1,802,165	99.85	37	530.0	99.98	34	345.9	99.39	37	627.1	99.38	41	298.9	98.55	38	404.7
Scaffold 5	1,679,391	99.92	36	223.1	99.98	34	233.8	99.25	36	269.7	99.46	35	219.0	99.69	40	228.8
Scaffold 6	1,639,599	99.67	36	624.3	99.98	31	379.1	99.54	35	238.5	96.97	31	1082.3	97.33	38	506.3
Scaffold 7	1,327,247	99.81	35	858.4	100.00	32	518.6	98.79	30	308.7	98.29	37	1731.4	99.28	38	902.8
Scaffold 8	1,116,489	99.87	36	17.7	99.89	33	19.2	98.71	34	23.2	98.36	30	23.4	98.49	37	26.8
Scaffold 9	912,454	99.84	37	1475.4	99.96	37	1117.3	99.53	40	1200.9	99.51	42	1329.4	99.96	41	1099.2
Scaffold 10	511,739	98.55	36	3172.2	98.75	35	1884.7	99.26	37	1124.1	98.16	41	5187.9	98.84	42	2344.3
Scaffold 11	457,683	99.94	53	1459.9	99.99	55	890.9	99.47	58	921.4	99.84	83	2464.1	99.94	65	1404.3
Scaffold 12	413,183	99.86	41	5184.2	99.99	34	3978.5	99.56	42	4847.5	99.36	55	4276.1	98.56	41	3898.0
Scaffold 13	378,582	99.43	36	11.4	99.98	32	13.6	99.49	33	23.5	97.26	31	21.2	99.51	39	17.5
Scaffold 14	366,083	99.45	36	13.1	98.78	31	20.7	98.10	32	26.9	95.62	29	32.9	97.33	35	28.3
Scaffold 15	344,704	97.07	51	3099.3	96.99	44	2641.9	96.90	51	3887.8	95.92	44	3244.1	94.96	47	2837.0
Scaffold 16	282,902	99.96	547	1740.7	99.95	545	2253.9	99.96	567	3376.0	99.95	388	1478.6	99.93	475	3093.4
Scaffold 17	281,078	99.91	299	783.1	99.71	313	619.3	99.57	387	1308.8	99.87	314	406.7	99.68	245	828.8
Scaffold 18	278,623	99.18	36	2774.8	99.06	32	2779.0	97.86	29	2788.5	96.55	38	3372.3	98.50	37	2827.8
Scaffold 19	267,222	99.83	39	979.7	99.55	34	685.4	99.15	30	543.4	99.03	47	1808.8	99.22	38	1082.9
Scaffold 20	264,226	98.58	45	1760.8	98.63	37	1908.8	97.84	43	2357.0	97.40	54	820.0	97.28	42	1073.7
Scaffold 21	232,983	99.97	220	152.7	99.96	181	159.1	99.86	122	238.5	99.94	289	235.6	99.36	145	164.9
Scaffold 22	195,802	99.93	41	2483.3	99.98	35	2416.5	97.40	31	1980.0	98.64	50	1809.9	97.96	38	1903.3
Scaffold 23	155,017	99.76	835	2449.1	99.86	606	3300.6	99.92	612	3582.5	99.83	426	1417.7	99.46	540	3106.8
Scaffold 24	136,997	100.00	960	3837.2	100.00	922	4081.7	100.00	833	5049.4	100.00	527	3487.4	100.00	799	3957.9
Scaffold 25	135,273	95.76	34	15.8	96.13	29	19.0	95.04	26	76.2	95.10	43	25.4	95.73	37	38.2
Scaffold 26	131,654	98.13	35	377.4	97.92	31	414.8	97.44	29	484.3	97.55	38	569.4	97.69	35	481.0
Scaffold 27	130,158	96.49	34	21.4	96.92	24	19.6	93.89	13	37.5	93.34	11	48.1	92.24	16	39.8
Scaffold 28	91,000	99.99	75	4410.0	99.99	46	3162.4	99.44	45	3291.8	99.81	50	2210.2	99.99	38	2187.7
Scaffold 29	14,771	99.90	42	14.0	99.89	43	13.6	99.70	32	18.1	99.95	46	31.9	99.80	44	20.9
Scaffold 30	11,875	98.35	333	2809.1	97.78	268	1864.4	97.06	146	1188.4	98.11	463	5209.8	97.54	224	2927.4
Chrs Mean		99.71			99.89			98.45			98.26			98.90		
Chrs Stdev		0.05			0.05			0.36			0.52			0.22		
Scaffold Mean		99.30			99.33			98.54			98.25			98.41		
Scaffold Stdev		1.11			1.10			1.59			1.74			1.75		

Table 17: Proportion and metrics of synteny (co-linear blocks of MAKER beet gene predictions) shared among five species.

Species	*C. quinoa*	*A. hypocondriacus*	*S. oleracea*	*V. vinefera*	*A. thaliana*
Common name	quinoa	amaranth	spinach	grape	Arabidopsis
Chromosome number	2n=4x=36	2n=4x=32	2n=2x=12	2n=2x=38	2n=2x=10
Number of synteny blocks	854	599	410	547	734
Number of genes in blocks	25,832	14,519	8,437	10,711	9,228
Mean number of genes per block	30.2	24.2	20.6	19.6	12.6
Stdev	51.4	31.2	27.6	22.4	8.7
Range	5 - 490	5 - 245	6 - 261	6 - 223	6 - 74