GENOMICS OF *BETA VULGARIS* CROP TYPES: INSIGHTS INTO TAP ROOT DEVELOPMENT AND STORAGE CHARACTERISTICS

By

Paul John Galewski

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Plant Breeding, Genetics and Biotechnology - Crop and Soil Sciences - Doctor of Philosophy

2020

ABSTRACT

GENOMICS OF *BETA VULGARIS* CROP TYPES: INSIGHTS INTO TAP ROOT DEVELOPMENT AND STORAGE CHARACTERISTICS

By

Paul John Galewski

Cultivated *Beta vulgaris* L. (beet) is a species complex composed of several distinct crop types developed for specific end uses. The crop types include sugar beet, fodder beet, table beet and leaf beet/chard. The evolution of each crop type appears to have resulted from interactions between selection, drift, gene flow, recombination, and the sorting of ancestral variation. Beets are generally heterozygous and contain self-incompatibility mechanisms. Therefore, reproducing and maintaining the genetic constitution of a single individual for genetic and phenotypic analysis is a challenge. Beet populations are the fundamental unit of improvement and contain the evolutionary and adaptive potential of the species. This research used several approaches which explore the utility of pooled population genomic sequencing to survey the organization and distribution of genetic diversity within cultivated *B. vulgaris* lineages, and give context and clarity to the genetics underlying important agronomic characters.

Whole genome sequence data was produced for important varieties and germplasm releases which represent the *B. vulgaris* crop type lineages. Using population genetic and statistical methods, relationships were determined between populations. Lineage-specific variation, or variation unique to specific crop types, was uncovered and used to quantify the level of support for these groups as discrete units. Allele frequency was able to differentiate between crop types using Principle Components Analysis (PCA), suggesting positive selection for end use was a major driver of crop type divergence. PCA carried out on a chromosome-by-

chromosome basis showed the relative contributions of specific chromosomes to crop type diversification. Gene diversity (e.g., expected heterozygosity) and F_{ST} proved powerful indicators of selection along the chromosome at nucleotide resolution. In total, 12.13% of loci within the genome were differentiated with respect to crop type. Interestingly, this corresponds to levels of divergence observed in studies of incipient speciation. Differentiated regions, indicated by F_{ST} outliers, contained 472 genes, or 1.6% of the 24,255 genes predicted in the reference genome assembly.

The content and organization of diversity in beet genomes reflects a complex history related to *B. vulgaris* crop type diversification. With the exception of chard, much of the species' historical selection has focused on the improvement of root characters (e.g., root enlargement, biomass, dry matter content, and sucrose concentration). As a result, major differences in root morphology and physiology can be observed between these lineages. Measures of root development and physiology between crop types were compared, and interestingly, much of the phenotypic variation partitioned between crop types corresponds to candidate genes identified from analyses of genome-wide variation using F_{ST} and *2pq*. Admixture and introgression appear to have shared specific variation involved in the reduction of lateral roots (e.g., Root primordium defective 1), root enlargement (e.g., Brevis radix-like 4, putative NAC domain-containing protein 94, cytokinin dehydrogenase 3), and biomass accumulation (e.g., 6-phosphofructo-2-kinase). High relationship coefficients and high correlations in allele frequency for this variation were observed, indicating the genetic variation influencing these characters may have been derived from a single origin. Integrating selection, drift, and admixture into a putative demographic history of beet provides evidence for the role of specific genes in the development of beet crop types and the expression of novel phenotypic characters.

TABLE OF CONTENTS

LIST OF TABLES vii

LIST OF FIGURES viii

KEY TO ABBREVIATIONS x

INTRODUCTION 1

LITERATURE CITED 7

CHAPTER 1 GENETIC DIVERSITY AMONG OF CULTIVATED BEETS (*BETA VULGARIS)* ASSESSED VIA POPULATION-BASED WHOLE GENOME SERQUENCES 11

INTRODUCTION 12

MATERIALS AND METHODS 16

Beta vulgaris populations and sequencing 16

Data processing and variant detection 16

AMOVA 17

Crop type relationships 17

Population size history 18

Lineage-specific variation 18

RESULTS 19

DISCUSSION 35

LITERATURE CITED 40

CHAPTER 2 QUANTIFYING *BETA VULGARIS* GENOME DIFFERENTIATION WITH RESPECT TO CROP TYPE USING WHOLE GENOME POOLED SEQUENCING 45

INTRODUCTION 46

MATERIALS AND METHODS 50

Beet populations and sequencing 50

Data processing and variant detection 50

2pq – Gene diversity/expected heterozygosity of bi-allelic sites 51

F_{ST} *– differentiation* 51

Lineage-specific variation 53

Genes/ F_{ST} Outliers 54

Visualization of genome differentiation 54

Visualization of crop type differentiation 55

Gene plots (allele frequency) 55

RESULTS 56

Genetic variation within cultivated B. vulgaris 56

Lineage-specific variation (individuals) 56

Gene diversity/ expected heterozygosity 57

Crop type differentiation (F_{ST}) 60

Differentiation of B. vulgaris crop types....61
Lineage-specific variation (crop type)....66
F_{ST} outliers and associated genes....66
Crop type genes (sugar beet)....69
Crop type genes (table beet)....70
Crop type genes (fodder beet)....70
Crop type genes (chard)....71
Selective sweeps....71
DISCUSSION....77
APPENDIX....86
LITERATURE CITED....105

CHAPTER 3ADMIXTURE AND INTROGRESSION IN THE DIVERSIFICATION OF *BETA VULGARIS* CROP TYPES....113
INTRODUCTION....115
MATERIALS AND METHODS....118
Admixture, introgression and the origin of important variation....118
Comparisons and evaluation standing genetic diversity....118
RESULTS....120
Variation in B. vulgaris genomes and the history of crop type lineages....120
Evolutionary history of root types involves admixture and introgression....124
Standing genetic diversity in beets....127
DISCUSSION....131
LITERATURE CITED....137

CONCLUSIONS....140

LITERATURE CITED....145

LIST OF TABLES

Table 1-1: List of materials for sequencing 21

Table 1-2: SNP and INDEL variation in cultivated *B. vulgaris* 22

Table 1-3: Analysis of molecular variance (AMOVA) 23

Table 1-4: Accumulation of lineage-specific variation along chromosomes 24

Table 1-5: Pairwise relationship matrix 28

Table 1-6: Historical time line highlighting evidence of beet utilization 34

Table 2-1: Results of Wilson Cox test 61

Table 2-2: Differentiated regions (F_{ST}) crop type 62

Table 2-3: Diverged SNP loci with respect to crop type and chromosome.....63

Table 2-4: Differentiated regions (F_{ST}) by chromosomes 65

Table 2-5: Significant genes based on F_{ST} outliers 68

Table 2-S1 Genes with significant F_{ST} values ($F_{ST} > 0.6$).....99

Table 3-1: Comparison of genome-wide variation.....126

Table 3-2: Comparisons of local candidate gene variation.....129

LIST OF FIGURES

Figure 1-1 Images of select *B. vulgaris* populations representing differences between important varieties and crop types. ……20

Figure 1-2: Gene diversity/expected heterozygosity (*2pq*) of *B. vulgaris* lineages. …… 25

Figure 1-3: Lineage relationships inferred by hierarchical clustering of pairwise relationship coefficients. 29

Figure 1-4: PCA plot showing the separation of crop types using genome-wide allele frequency data 30

Figure 1-5: PCA plot showing the separation of crop types using allele frequency data on a chromosome by chromosome basis 31

Figure 1-6: Inferred historical N_e of *B. vulgaris* crop types using the program SMC++ 33

Figure 2-1: Distribution of lineage-specific variation across chromosomes of cultivated beet 58

Figure 2-2: Topology of crop type variation across the genome 59

Figure 2-3: Topology of crop type variation along Chromosome 3 64

Figure 2-4: Allele frequency data for R locus (EL10Ac2g04268) 76

Figure 2-S1: Topology of crop type variation along Chromosome 1 87

Figure 2-S2: Topology of crop type variation along Chromosome 2 88

Figure 2-S3: Topology of crop type variation along Chromosome 4 89

Figure 2-S4: Topology of crop type variation along Chromosome 5 90

Figure 2-S5: Topology of crop type variation along Chromosome 6 91

Figure 2-S6: Topology of crop type variation along Chromosome 7 92

Figure 2-S7: Topology of crop type variation along Chromosome 8 93

Figure 2-S8: Topology of crop type variation along Chromosome 9 94

Figure 2-S9: Allele frequency data for Root Primordium Defective 1, *RPD1,* (EL10Ac4g09126) 95

Figure 2-S10: Allele frequency data for NAM/NAC (EL10Ac2g02976) 96

Figure 2-S11: Allele frequency data for Cytokinin dehydrogenase 1 (EL10Ac8g19202) 97

Figure 2-S13: Allele frequency data for the Y locus (EL10Ac2g04466) 98

Figure 3-1: Classification of standing genetic variation within *B. vulgaris* lineage genomes 122

KEY TO ABBREVIATIONS

LSV – Lineage-specific variation

N_e – Effective population size

NGS – Next generation sequencing

PCA – Principal components analysis

SNP – Single nucleotide polymorphism

WGS – Whole genome sequencing

IBS – Identity by state

AMOVA – Analysis of molecular variation

SNP – Single nucleotide polymorphism

Indel – Insertion/Deletion

ILS – Incomplete lineage sorting

AI – Admixture and introgression

LSE – Lineage-specific evolution

LSV – Lineage-specific variation

WAG – Weeks after germination

FPKM – Fragments Per Kilobase of transcript per Million mapped reads

CMS – Cytoplasmic male sterility

CWR – Crop wild relatives

INTRODUCTION

Beta vulgaris L. is a species within the order Caryophyllales, family Amaranthaceae. The species is composed of wild *B. vulgaris* ssp. *Maritima* and several crop types that fill distinct production niches. Sugar beet, fodder beet, table beet, and chard are produced as a sugar crop, feed crop, root vegetable, and leaf vegetable, respectively. The crop type lineages contain important phenotypic variations, which are the major determinants of end use and production.

Sugar beet is one of two economically viable sugar crops, the other being sugar cane (*Saccharum officinarum* L.). Together these crops satisfy the global demand for sucrose. Sugar beet represents a significant crop to the US and to the state of Michigan. Sugar beet accounts for 50% of US sugar production and 25% of global sugar production. Historically an old-world crop, sugar beet represents an important temperate source for sucrose. Considerable time and energy have been put into the adaptation of the crop to the major growing regions of the US. These regions include the Upper Midwest (e.g Michigan, Minnesota, and North Dakota), Great Plains (Colorado, Montana, Nebraska, and Wyoming) and the Far West (California, Idaho, Oregon, and Washington) (ERS 2019). Sugar beet differs from other crop types, mainly in root characteristics such as sucrose content and yield. Sucrose concentration can exceed 18% in modern hybrids. Sugar beet is also largely adapted to regional growing environments and management practices determined by sugar yield per hectare.

The other crop types represent important but minor crops based on total acres in cultivation. Table beet is a biennial root vegetable prized for sweet flesh and nutritional value (Goldman and Navazio 2002). Breeding practices of the crop are similar to that of sugar beet (Goldman and

Navazio 2008) and the history of breeding of table beets in the US has been well documented (Goldman 1996). Fodder beet, also referred to as forage beet, mangle, or mangle-wurzel is used as animal feed. Fodder beet is less frequently utilized in the US than abroad owing to the prevalence of other feed crops. Fodder beet expresses an expanded root similar to sugar beet but contains more diversity in terms of shape and composition (e.g., dry matter content, sucrose concentration) (Henry 2010). Chard represents lineages selected for leaf quality and likely represents the first cultivated beet types (Biancardi et al. 2012). It is plausible that chard was selected from sea beet more than once. All beet types are ultimately derived from *B. vulgaris* spp. *maritima* (Winner 1993), and to date, how the genomes of these ancestral populations reflect genomes of cultivated lineages is unknown aside from a reduction of genetic diversity in sugar beet gene pools (Bosemark 1979).

Potential for beet improvement include traits related to sugar and dry matter concentration, root and leaf quality for human consumption and feed, yield, and biomass accumulation. Other end use niches for beet production are possible (e.g., energy beet, industrial chemical stocks) and will likely follow similar breeding methods as a consequence of the genetics of the species regardless of phenotypes being measured and selected for (McGrath and Panella 2018). Irrespective of crop type, breeders of *B. vulgaris* report similar breeding practices and recognize similar genetic resources (e.g., gene pools) for improvement. Relative to the other crop types, sugar beet has seen greater investments in genetics and genomics research because of its economic importance, but for the most part insights gained regarding the genetics and breeding of beet appear highly transferable irrespective of crop type.

B. vulgaris L. is diploid species with nine chromosomes (2n=2x=18). Wild-type populations are generally outcrossing, self-incompatible, and wind pollinated. The high heterozygosity has large implications on diversity, breeding, and adaptation of beet to diverse regions/environments. Few barriers to hybridization exist and thus important agronomic characters developed within a crop type lineage are likely transferable to others through hybridization, introgression, and backcross strategies. Cytoplasmic male sterility (CMS) systems have been transferred to table beet through such strategies for hybrid seed production (Goldman and Navazio 2002).

The gene pools for beet improvement include the crop types, diverse populations of *B. vulgaris* spp. *maritima,* and several related species such as *Patellifolia procumbens* and *P. webbiana.* Research in the US has been focused on local adaptation and identification of resistance to devastating pathogens. This is mirrored by the plethora of historical seed releases of improved germplasm for sugarbeet and the systematic incorporation of genetic diversity into public breeding programs (Panella et al. 2015). *B. vulgaris* spp. *maritima* has been used extensively as a source for resistance to Cercospora (Munerati et al. 1913). Activities of national programs have focused on widening the genetic base of sugar beet as it is reported that early improvement focused solely on sucrose concentration and extraction (Pannella and Lewellen 2007). As a result, the genetic base of sugar beet is suggested to be less diverse than other outcrossing crops (Boesmark 1979). *P. procumbens* and *P. webbiana* have been used as a source for variation to improve cultivated beet types. Nematode resistance was introduced to sugar beet by hybridization with *P. procumbens* (Savitsky 1975). Further experiments have identified a source of resistance in the $Hs1^{pro\text{-}1}$ gene (Jung and Wricke 1987). Although successfully introgressed, the source of this resistance is rarely used owing to high yield penalties in environments with low

disease pressure. Gaskill (1954) reported swiss chard as a bridging species for hybridization and introgression between sugar beet and interspecific species, *P. procumbens* and *P. webbiana*. The fact that hybridization was variable between crop types hints at genome divergence between crop types.

New technologies have offered ways to measure the genomic diversity of crops and their genetic resources (e.g., related species). The availability of genome sequence has provided useful measures of diversity and the content and organization of variation contained within genomes of a species. Genomes representing these important lineages provide an opportunity to detect the heritable genome variation underlying important phenotypes with agronomic potential and give context and clarity to the subspecific diversity of beet. Reference genome sequences EL10 (Funk et al. 2018), RefBeet (Dohm et al. 2014) along with *in situ* hybridization of chromosomes (Paesold et al. 2012) have offered a perspective of unique features and evolutionary history of understanding of the Amaranthaceae and order Caryophyllales.

Roots are important plant organs that exhibit a large array of morphological and functional diversity. This diversity functions in the stabilization, adaptation, and interaction with the rhizosphere. In a handful of crops roots are the economic tissues of interest (e.g., beet, sweet potato, turnip, carrots, parsnips, radish). Beet is predominantly thought of as a root crop, with the exception of leaf beet/chard, which is used for leaves and lacks the enlarged root character. This subspecific diversity results from hundreds to thousands of years of selective breeding. The ability to generate sequence from phenotypically distinguishable lineages provides an opportunity to quantify the genomic diversity and divergence with respect to the mechanisms governing root expansion and differences in physiological traits. The enlarged root may serve

several purposes, one includes a switch to biennial habit whereby the first year is vegetative growth and second year is reproductive growth that relies on energy “sucrose” stored in the first year (Cooke and Scott 1993). This switch is thought to occur through the role of pseudo response regulators and has been implicated as a switch in the life cycle of beet and likely a key domestication trait in beet due to associated changes in carbohydrate metabolism (Pin et al. 2012).

Selection for sucrose occurs by measuring sucrose yield per hectare. Gains in sucrose per hectare have been largely accomplished by first increasing sucrose content within the roots and secondly by root yield (growth and development). Evidence for negative linkage between yield (biomass) and sucrose concentration may limit the efficiency of selection in beet (Boesmark 2006). The E and Z types represent lineages with yield and sucrose, respectively, as the primary trait under selection and may represent important subspecific diversity that underlies this negative linkage. Understanding sucrose accumulation in beet requires an understanding of root development and physiology of the root. Both traits are highly influenced by environment, and thus, crop management strategies (e.g., seeding rates and nitrogen application) must also be considered for improvement (McGrath and Panella 2018).

Root enlargement occurs, in part, by the formation of supernumerary cambia (Artschwager 1926). From these secondary cambia, cell growth occurs first by division and then by cell expansion. As cell type differentiation terminates in the formation of tissues of specialized function (e.g., vasculature), new rings continue to form, repeating this process. Developing roots experience a morphophysiological change at around five weeks in development, and correlated shifts in gene expression and morphology can be observed (Trebbi and McGrath 2009). This

correlates well with a formation of rings and the accumulation of sucrose. Ring density was found to be correlated with sucrose concentration but negatively correlated with yield (Milford 1973, 1976). Parenchyma cells close to phloem are thought to be higher in sucrose. The sucrose gradient hypothesis (Wyse 1979) suggests sucrose diffuses into the cytosol of parenchyma cells neighboring the phloem using a series of invertases, which establish a gradient for passive diffusion. Trafficking into the vacuole is thought to occur by similar mechanisms or potentially through ATP-dependent vesicle trafficking (Getz 2000). Colocalization of sucrose synthase with locations of tissues and cells involved in energy dependent processes such as cell wall biosynthesis and sucrose accumulation in the vacuole suggest a role for this enzyme in maintaining sink strength (Fugate et al., 2019). Molecular genetic explanations for this important process remain unknown. Furthermore, little is known about the differences in genome variation between beet crop types that contribute to phenotypic differences observed in important traits. Understanding the relationships between these lineages is critical for identification of the genetic basis of important agronomic adaptations. An understanding of how the variation is distributed between important lineages and populations will be useful for identifying additional sources of variation for important traits and breeding varieties that impact local adaptation, productivity, and sustainability of the crop.

LITERATURE CITED

LITERATURE CITED

Artschwager, E., 1926 Anatomy of the vegetative organs of the sugar beet. J. Agric. Res. 143–176.

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

Bosemark, N. O., 1979 Genetic poverty of the sugar beet in Europe. In: Proceedings of the Conference Broadening Genetic Base of Crops. Pudoc, Wageningen, the Netherlands, pp. 29–35.

Bosemark, N. O., 2006 Genetics and Breeding, pp. 50–88 in *Sugar Beet*, Blackwell Publishing Ltd, Oxford, UK.

Cooke, D. A., and R. K. Scott, 1993 *The Sugar Beet Crop*. Chapman and Hall Publishers, London.

Dohm, J. C., A. E. Minoche, D. Holtgräwe, S. Capella-Gutiérrez, F. Zakrzewski et al., 2014 The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). Nature 505: 546–549.

ERS (Economic Research Service) United States Department of Agriculture. 2019 Table 14-U.S. sugarbeet crops: area planted, acres harvested, yield per acre, and production, by State and region https://www.ers.usda.gov/data-products/sugar-and-sweeteners-yearbook-tables/sugar-and-sweeteners-yearbook-tables/#World%20Production,%20Supply,%20and%20Distribution Last Updated 2019.

Fugate, K. K., J. D. Eide, D. N. Martins, M. A. Grusak, E. L. Deckard *et al.*, 2019 Colocalization of sucrose synthase expression and sucrose storage in the sugarbeet taproot indicates a potential role for sucrose catabolism in sucrose accumulation. J. Plant Physiol. 240: 153016.

Gaskill, J. O., 1954 Viable hybrids from matings of chard with *Beta procumbens* and *B. webbiana*. Am. Soc. Sugar Beet Technol. 8: 148–152.

Getz, H. P., 1991 Sucrose transport in tonoplast vesicles of red beet roots is linked to ATP hydrolysis. Planta 185: 261–268.

Goldman, I. L., 1996 A list of germplasm releases from the University of Wisconsin table beet breeding program, 1964-1992. HortScience 31: 880–881.

Goldman, I. L., and J. P. Navazio, 2008 Table Beet, pp. 219–238 in *Vegetables I. Handbook of Plant Breeding*, edited by J. Prohens and F. Nuez. Springer New York, New York, NY.

Goldman, I. L., and J. P. Navazio, 2002 History and breeding of table beet in the United States. Plant Breed. Rev. 22: 357–388.

Henry, K., 2010 Fodder Beet, pp. 221–243 in *Root and Tuber Crops*, edited by J. E. Bradshaw. Springer New York, New York, NY.

Jung, C., and G. Wricke, 1987 Selection of diploid nematode-resistant sugar beet from monosomic addition lines. Plant Breed. 98: 205–214.

McGrath, J. M., and L. Panella, 2018 Sugar beet breeding. Plant Breed. Rev. 42: 167–218.

Milford, G. F. J., 1973 The growth and development of the taproot of sugar beet. Ann. Appl. Biol. 75: 427–438.

Milford, G. F. J., 1976 Sugar concentration in sugar beet: varietal differences and the effects of soil type and planting density on the size of the root cells. Ann. Appl. Biol. 83: 251–257.

Munerati, O., Mezzadroli, G. and Zapparoli, T.V., 1913 Osservazioni sulla *Beta maritima* L., nel triennio 1910–1912. *Le Stazioni Sperimentali Agrarie Italiane*, *46*(6): 347-371.

Paesold, S., D. Borchardt, T. Schmidt, and 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 Sugarbeet Releases and Breeding Over the Past 20 Years. J. Sugarbeet Res. 52: 40-85.

Panella, L., and R. T. Lewellen, 2007 Broadening the genetic base of sugar beet: Introgression from wild relatives. Euphytica 154: 383–400.

Pin, P. A., W. Zhang, S. H. Vogt, N. Dally, B. Büttner *et al.*, 2012 The role of a pseudo-response regulator gene in life cycle adaptation and domestication of beet. Curr. Biol. 22: 1095–1101.

Trebbi, D., and J. M. McGrath, 2009 Functional differentiation of the sugar beet root system as indicator of developmental phase change. Physiol. Plant. 135: 84–97.

Wyse, R., 1979 Parameters controlling sucrose content and yield of sugarbeet roots. J. Sugarbeet Res. 20: 368–384.

Winner, C., 1993 History of the crop., pp. 1–22 in *The Sugar Beet Crop.*, edited by D. Cooke and S. RK. Chapmann & Hall, London, UK.

Savitsky, H., 1975 Hybridization between *Beta vulgaris* and *B. procumbens* and transmission of nematode (*Heterodera schachtii*) resistance to sugar beet. Can. J. Genet. Cytol. 197–209.

CHAPTER 1

GENETIC DIVERSITY AMOUNG CULTIVATED BEETS (*BETA VULGARIS)* ASSESSED VIA POPULATION-BASED WHOLE GENOME SEQUENCES

INTRODUCTION

Beta vulgaris L. (beet) is an economically important plant species consisting of several distinct cultivated lineages. These lineages, or "crop types," include sugar beet, table beet, fodder beet, and chard. The crop types have been adapted for specific end uses and thus exhibit pronounced phenotypic differences. Crop type lineages breed true, indicating a genetic basis for these phenotypes. Cultivated beets likely originated from wild progenitors of *B. vulgaris* spp. *maritima*, also called "sea beet" (Biancardi et al. 2012). It is widely accepted that beet populations were first consumed for leaves. The earliest evidence for lineages with expanded roots occurs in Egypt around 3500 BC. The root types and the origin of the enlarged root is thought to have occurred in the Near East (Iraq, Iran, and Turkey) and spread west (Europe) (Zossimovich 1940). Interestingly, beet production for roots as an end use was first described along trade routes across Europe. Historically, Venice represented a major European market of the Silk Road and facilitated the distribution of eastern goods across Europe (Kuzmina 2008). Table beet has been proposed to have been developed within Persian and Assyrian gardens (Goldman and Navazio 2002). Whether this specifically corresponds to the origin of the expanded root character or a restricted table beet phenotype remains unknown. In fact, early written accounts regarding the use of root vegetables often confused beet with turnip (*Brassica rapa*).

Hybridization between diverged beet lineages has long been recognized as a source of genetic variability available for the selection of new crop types and improving adaptation (Schukowsky 1950 cited in Winner et al. 1993, Cooke and Scott 1993). In 1747, Margraff was the first to

recognize the potential for sucrose extraction from beet. Achard, a student of Margraff, was the first to describe specific fodder lineages that contained increased quantities of sucrose and the potential for an economically viable source of sucrose for commoditization (Winner 1993). In 1787, Abbe de Commerell suggested red mangle (fodder) resulted from a red table beet/chard hybrid and that the progenitors of sugar beet arose from hybridizations between fodder and chard lineages (Fischer 1989, Ford-Lloyd 1995). Louise de Vilmorin (1816-1860), a French plant breeder, first detailed the concept of progeny selection in sugar beet, a method of evaluating the genetic merit of lineages based on progeny performance (Gayon and Zallen 1998). Vilmorin used differences in specific gravity to select beet populations. This approach led to increases in sucrose concentration from ~4% in fodder beet to ~18% in current US hybrids (reviewed in McGrath and Fugate 2012).

B. vulgaris is a diploid organism (2n = 18) with a predicted genome size of 758 Mb (Arumuganathan and Earle 1991). Chromosomes at metaphase exhibit similar morphology (Paesold et al. 2012). The first complete reference genome for *B. vulgaris* (e.g., Refbeet) provided a new perspective regarding the content of the genome (e.g., annotated gene models, repeated sequences, and pseudomolecules) (Dohm et al. 2014). This research confirmed whole genome duplications and generated a broader view of genome evolution in the Eudicots, Caryophyllales, and *Beta*. The EL10.1 reference genome (Funk et al., 2018) represents a contiguous chromosome scale assembly resulting from a combination of PacBio, BioNano optical mapping and Hi-C. Together, EL10.1 and Refbeet provide new opportunities for studying the content and organization of the beet genome. Resequencing of important beet populations has

the potential to characterize the landscape of variation and inform recent demographic history of beet, including the development of crop types and other important lineages.

Population genetic inference leveraging whole genome sequencing (WGS) data have proven powerful tools for understanding evolution from a population perspective (Stortz 2005, Lynch 2009, Casillas and Barbadilla 2017). Knowledge of the quantity and distribution of genetic variation within a species is critical for the conservation and preservation of genetic resources in order to harness the evolutionary potential required for the success of future beet cultivation. Recent research has revealed the complexity of relationships within *B. vulgaris* crop types (Andrello et al. 2017). Studies have shown sugar beet is genetically distinct and exhibits reduced diversity compared to *B. vulgaris* spp. *maritima*. Geography and environment are major factors in the distribution of genetic variation within sugar beet populations in the US (McGrath et al. 1999). Furthermore, spatial and environmental factors were evident in the complex distribution of genetic variation in wide taxonomic groups of *Beta* (Andrello et al. 2016), which include the wild progenitors of cultivated beet.

Here we present a hierarchical approach to characterize the genetic diversity of cultivated *B. vulgaris* using pooled sequencing of populations representing the crop type lineages. These populations contain a wide range of phenotypic variation including leaf and root traits, distinct physiological/biochemical variation in sucrose accumulation, water content, and the accumulation and distribution of pigments (e.g., betaxanthin and betacyanin). These phenotypic traits, along with disease resistance traits, represent the major economic drivers of beet production. Developmental genetic programs involved in cell division, tissue patterning, and

organogenesis likely underlie the differences in root and leaf quality traits observed between crop types. Improvement for these traits as well as local adaptation and disease resistance occurs at the level of the population. Pooled sequencing provides a means to characterize the diversity of beet populations and generate nucleotide variation, which has utility in marker-based approaches for a diverse community of breeders and researchers interested in *B. vulgaris*. Pooled sequencing works in synergy with both the reproductive biology of the crop as well as the means by which phenotypic data is collected (e.g., populations' mean phenotypes) and beets are improved through selection. Knowledge regarding the genetic control of important traits, currently unknown, will help prioritize existing variation and access novel genetic variation in order to address the most pressing problems related to crop production and sustainability.

MATERIALS AND METHODS

Beta vulgaris populations and sequencing

Twenty-three beet populations were sequenced to 80X coverage relative to the predicted 758 Mb *B. vulgaris* genome using a pooled sequencing approach. The populations selected are representative of the four recognized crop types and capture the range of phenotypic diversity found within cultivated beet (Table 1). Populations were grown in the greenhouse and leaf material was harvested from 25 individuals per population. Leaf material, one young expanding leaf of similar size from each individual within a population, was combined, homogenized, and DNA was extracted using the Macherey-Nagel NucleoSpin Plant II Genomic DNA extraction kit (Bethlehem, PA). NGS libraries were constructed using TruSeq bar-code adapters from one microgram of DNA from each population and sequenced as paired end reads of 150 bp on the Illumina Hi-Seq 2500. The resulting reads were assessed for quality using FastQC (Andrews 2010), library bar-code adapters were removed, and reads were trimmed according to a quality threshold using TRIMMOMATIC (Bolger et al. 2014) invoking the following options (ILLUMINACLIP:adapters.fa-:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36). These filtered reads were used for downstream analysis.

Data processing and variant detection

Variants for each population were called by aligning the filtered reads to the EL10.1 reference genome assembly (Funk et al. 2018) using bowtie2 v2.2.3 (options -q --phred33-quals -k 2 -x) (Langmead and Salzberg 2012). The resulting alignment files were sorted and merged using

SAMtools version 0.1.19 (Li et al. 2009). SNP variants were called for each population using BCFtools (Li 2011), filtered for mapping quality (MAPQ >20) and read depth (n > 15), and then combined using VCFtools (Danecek et al. 2011). The combined data was again filtered to obtain biallelic sites across all populations. Indels were evaluated using the Genome Analysis Toolkit (GATK) haplotype caller (McKenna et al. 2010). The 'm*pileup*' subroutine in SAMtools was then used to quantify the alignment files and extract allele counts. Allele frequency was estimated within individual populations for SNP loci identified as biallelic across all populations. Population parameters were then estimated using allele frequencies within each population such that ($p + q = 1$), where p was designated as the allele state of the EL10.1 reference genome and q, the alternate, detected in each sequenced population. Expected heterozygosity ($2pq$), also termed gene diversity (Nei 1987), was used to compare diversity contained within each population.

AMOVA

Analysis of molecular variance (AMOVA) was used to assess the distribution of genetic variation within the species (Excoffier et al. 1992). AMOVA was performed using the *ade4* package in R (Thioulouse et al. 1997) following the approach for pooled sequence data outlined in Gompert et al. (2010).

Crop type relationships

Biallelic SNPs were used to calculate pairwise relationship coefficients between populations using an identity by state (IBS) approach within the Kinship Inference for Association Genetic Studies (KING) package (Manichaikul et al. 2010). Neighbor joining trees were generated in

order to extract bootstrap support for clusters using the *ape* package (Analyses of Phylogenetics and Evolution) in R (Paradis and Schliep 2004).

Population size history

Composite likelihood methods were used to estimate historical population sizes and infer demographic history from genome sequences of populations using the program SMC++ (Terhorst et al. 2016).

Lineage-specific variation

Lineage-specific variation (LSV), defined as homozygous private variation (e.g., apomorphy), was extracted from the merged VCF file containing variants for all populations. Variants that were fixed within a particular population or assemblage of populations (lineage), and not detected within any other lineage, were considered LSV. Variant files representing LSV were produced for each lineage in a hierarchical fashion (e.g., species, crop type and individual populations). LSV was then evaluated with respect to lineage as well as its distribution along chromosomes.

RESULTS

Twenty-five individuals from each of the 23 *B. vulgaris* populations were chosen to represent the cultivated *B. vulgaris* crop types (Table 1-1 and Figure 1-1). Leaf tissue was pooled, DNA extracted and sequenced using the Illumina 2500 in paired end format. On average, 61.84 ± 12.22 GB of sequence data was produced per population, with an average depth of 81.5X. After processing for quality, reads were aligned to the EL10.1 reference genome. Approximately 20% of bases were discarded owing to trimming of low-quality base calls and adapter sequences. Biallelic SNP and lineage-specific variants were used to estimate the quantity and organization of genome-wide variation within these *B. vulgaris* populations and groups (e.g., species, crop types, and populations). On average 90.74% of the filtered reads aligned to the EL10.1 reference genome. A total of 14,598,354 variants were detected across all populations, and 12,411,164 (85.0%) of these were classified as a SNP, and of these 10,215,761 (82.3%) were biallelic. Thus, most variants appeared to be biallelic, as only 2,718,205 (18.6%) variants were characterized as multiallelic. After filtering for read depth ($n \geq 15$), 8,461,457 biallelic SNPs remained for computational analysis. Insertions and deletions (indels) accounted for 2,187,190 (14.9%) of the variants detected (Table 1-2).

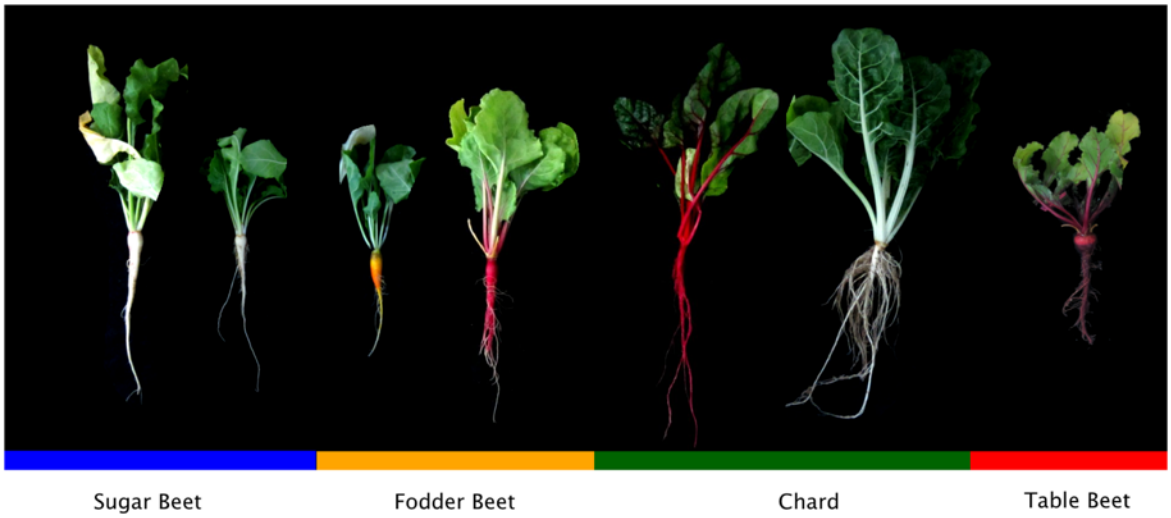

Figure 1-1: **Images of select *B. vulgaris* populations representing differences between important varieties and crop types.**

Table 1-1: **List of materials for sequencing.**

Crop Type	Entry	Name	Pop ID	PI # / Source	Total Reads	Gb	Coverage (X)	Year Released	Description
Sugar Beet	1	EL10	EL10	requested	-	-	-	2018	Reference genome assembly
	2	C869	C869	628754	549262696	68.7	90.6	2002	Parent population of EL10
	3	EL50/2	EL50	598073	487259716	60.9	80.4	1994	Cercospora Resistance
	4	EL51	EL51	598074	456623952	57.1	75.3	2000	Rhizoctonia Resistance
	5	East Lansing Breeding Population	GP10	-	492970286	61.6	81.3	Pending	OP Recurrent Selection Population
	6	SR102	SR102	675153	462483116	57.8	76.3	2016	Smooth Root/Low Tare
	7	East Lansing Breeding Population	GP9	-	847319042	105.9	139.7	Pending	OP Recurrent Selection Population
	8	SP6322	SP7322	615525	549262696	68.7	90.6	1973	Adaptation to Eastern US
	9	SR98/2	SR98/2	655951	482270894	60.3	79.5	2011	Rhizoctonia Resistance
	10	L19	L19	590690	767383878	76.7	101.2	1978	High Sucrose (>20%)
Table Beet									
	11	Bulls Blood Table Beet	BBTB	Chriseeds	519832300	65.0	85.7	1700	Historic ornamental and vegetable variety
	12	Crosby Egyptian Table Beet	Crosby	Chriseeds	466455846	58.3	76.9	1869	US variety with Egyptian background
	13	Ruby Queen Table Beet	RQ	Chriseeds	500356022	62.5	82.5	1950	Current production
	14	Touch Stone Gold Table Beet	TG	Chriseeds	396335036	49.5	65.4	Unknown	Golden Root
	15	Albino Table Beet	WT	Chriseeds	503139454	62.9	83.0	Unknown	White root
	16	Detroit Dark Red Table Beet	DDTB	Chriseeds	473659992	59.2	78.1	1892	US variety
	17	Wisconsin Breeding Line	W357B	Univ. WI	538981844	53.9	71.1	1982	Self-fertile O-type
Chard									
	18	Fordhook Giant	FGSC	Chriseeds	484646866	60.6	79.9	1934	Green chard
	19	Vulcan Swiss Chard	Vulcan	Chriseeds	547992902	68.5	90.4	Unknown	Red chard
	20	Lucellus Chard	LUC	Chriseeds	617051314	61.7	81.4	Pre-1700s	Historic green chard variety
	21	Rhubarb Swiss Chard	RHU	Chriseeds	538577146	53.9	71.1	1857	Red chard
Fodder Beet									
	22	Mammoth Red Fodder	MAM	Burpee	400297680	40.0	52.8	1800	Heirloom fodder beet variety
	23	Wintergold Fodder	WGF	Local stock	545378784	54.5	71.9	Unknown	Winter beet with gold skin pigment

[1] OP = open pollinated

Table 1-2**: SNP and INDEL variation in cultivated *B. vulgaris.***

	POP ID	Entry	Variation Detected			Lineage-specific Variation			Gene diversity
Populations			Total variants	SNP variants	Indel variants	Total variants	SNP variants	Indel variants	2pq
	EL10	1	34,870	30,686	4,184	1,149	689	460	0.027
	C869	2	635,471	588,096	47,375	9,514	8,290	1,224	0.194
	EL50	3	828,626	767,954	60,672	30,712	27,667	3,045	0.159
	EL51	4	830,003	768,406	61,597	17,464	15,547	1,917	0.195
	GP10	5	754,888	698,729	56,159	9,051	7,999	1,052	0.230
	GP9	6	649,330	599,372	49,958	6,094	5,366	728	0.253
	L19	7	809,158	748,133	61,025	19,938	17,854	2,084	0.187
	SP7322	8	840,925	778,082	62,843	15,528	13,942	1,586	0.213
	SR102	9	757,464	701,432	56,032	8,765	7,846	919	0.232
	SR98	10	795,193	736,344	58,849	16,241	14,612	1,629	0.202
	BBTB	11	953,871	884,972	68,899	88,129	79,236	8,893	0.087
	Crosby	12	872,503	809,544	62,959	21,882	19,436	2,446	0.198
	DDRT	13	852,400	791,076	61,324	24,180	21,592	2,588	0.185
	RQ	14	884,050	818,829	65,221	31,786	28,714	3,072	0.154
	TGSC	15	786,306	730,401	55,905	37,213	33,887	3,326	0.103
	W357B	16	878,640	815,237	63,403	81,786	74,941	6,845	0.043
	WT	17	867,720	804,159	63,561	30,371	27,613	2,758	0.159
	MAM	18	723,004	669,180	53,824	11,969	10,716	1,253	0.221
	WGF	19	879,000	813,515	65,485	25,210	22,850	2,360	0.202
	FGSC	20	1,033,473	958,024	75,449	31,764	28,455	3,309	0.241
	LUC	21	1,133,038	1,047,169	85,869	35,097	31,341	3,756	0.240
	RHU	22	965,749	894,064	71,685	29,089	26,138	2,951	0.195
	Vulcan	23	1,012,869	939,067	73,802	37,056	33,650	3,406	0.190
Crop Type									
	Sugar (Entries 1-10)		2,295,573	2,101,855	193,718	3,659	3,317	342	0.207 ± 0.002
	Table (Entries 11-17)		2,155,105	1,981,659	173,446	1,937	1,379	558	0.147 ± 0.044
	Fodder (Entries 18-19)		1,200,301	1,107,357	92,944	848	643	205	0.221 ± 0.013
	Leaf (Entries 20-23)		2,129,588	1,957,348	172,240	4,217	3,359	858	0.216 ± 0.027
B. vulgaris (cultivated)									
	B. vulgaris (GATK)		4,180,197	3,809,937	370,260	n/a	n/a	n/a	0.178 ± 0.060
	B. vulgaris (SamTools)		14,598,354	12,411,164	2,187,190	n/a	n/a	n/a	0.182 ± 0.040

AMOVA was performed in order to quantify the distribution of variation within and among cultivated *B. vulgaris* crop types. The results showed no strong population subdivision with respect to crop type. The variation shared among crop types (99.37%), far exceeded the variation apportioned between crop type lineages (0.40%). The variation detected between populations within a crop type was also low (0.23%) (Table 1-3). This result suggested a small proportion of the total variation is unique to any given population. This was confirmed by the low quantity of lineage-specific variation (LSV) detected, evaluated in a hierarchical fashion. Lineages were defined as individual populations, crop types, and species (Table 1-2). In total, 600,239 variants (4.0%) were unique and fixed within a single population. The accumulation of variation for specific chromosomes and populations was informative (Table 1-4). Individual populations of sugar beet contained a large quantity of LSV on Chromosome 6 relative to other sugar beet chromosomes and indicated that either divergent selection or drift has occurred on this sugar beet chromosome. The population Bulls Blood contained the greatest amount of LSV detected, 8,893 indels and 79,236 SNP variants. Table beet populations contained the most LSV which suggested they are the most divergent of the crop types (Table 1-4).

Table 1-3: **Analysis of molecular variance (AMOVA).**

Variance components	Sigma	%
Between Crop Type	0.005	0.40
Between Populations Within Crop Type	0.003	0.23
Populations (Species)	1.266	99.37
Total variation	1.274	100

Table 1-4: **Accumulation of lineage-specific variation along chromosomes.**

Pop ID	Entry	Chr 1	Chr 2	Chr 3	Chr 4	Chr 5	Chr 6	Chr 7	Chr 8	Chr 9	**mean**
EL10	1	91	170	103	114	96	229	147	95	104	138
C86925	2	680	562	1,547	933	2,365	1,101	482	1,316	528	1,057
EL50	3	1,482	1,496	5,328	2,414	5,141	4,722	3,356	4,244	2,529	3,412
EL51	4	978	2,436	1,852	1,830	2,019	3,361	1,825	1,772	1,391	1,940
GP10	5	398	787	964	642	776	2,376	1,331	1,116	661	1,006
GP9	6	491	521	864	1,023	892	1,839	821	1,028	510	888
L19	7	568	1,248	993	4,438	845	5,175	3,374	1,918	1,379	2,215
SP7322	8	467	1,190	1,696	2,026	1,475	4,125	1,906	1,601	1,042	1,725
SR102	9	406	683	1,081	1,115	1,000	1,458	1,021	1,368	633	974
SR98	10	419	1,356	1,364	2,056	3,158	3,757	1,423	1,691	1,017	1,805
BBTB	11	17,632	10,425	8,148	9,559	12,067	9,383	4,597	6,131	10,187	9,792
Crosby	12	2,210	1,172	2,772	2,584	2,511	3,857	2,470	2,548	1,758	2,431
DDRT	13	2,175	1,314	2,874	3,007	1,776	4,559	4,431	2,195	1,849	2,687
RQ	14	3,186	3,402	3,680	2,937	4,053	5,349	3,356	3,691	2,132	3,532
TGSC	15	3,014	8,486	3,732	3,625	2,971	4,290	3,988	3,716	3,391	4,135
W357B	16	7,806	4,186	7,661	6,766	16,835	2,011	8,723	5,947	2,102	6,893
WT	17	3,347	1,577	3,508	4,084	2,777	4,790	3,203	4,876	2,209	3,375
MAM	18	698	1,014	885	1,628	1,758	2,820	1,044	1,030	1,092	1,330
WGF	19	1,014	2,074	4,929	2,468	4,923	4,288	2,041	1,886	1,587	2,801
FGSC	20	2,883	3,738	2,480	4,665	3,768	4,286	4,181	3,224	2,539	3,529
LUC	21	2,615	3,570	3,269	3,376	4,834	7,489	4,063	3,118	2,763	3,900
RHU	22	2,631	2,996	2,249	3,421	2,649	5,019	2,872	3,880	3,372	3,232
Vulcan	23	3,662	3,977	3,694	4,243	3,343	5,800	3,841	5,054	3,442	4,117
mean		2,558	2,538	2,855	2,998	3,566	4,003	2,804	2,758	2,096	

Within the crop types, 10,661 variants were crop type specific and were not found within any other crop type. Of these, 8,098 were characterized as SNPs and 1,963 as indel. The number of SNP LSV detected within sugar beet, table beet, fodder beet, and chard crop types were as follows: 3,317, 1,379, 643, and 3,359, respectively. Indel LSV detected for the crop types were 342, 558, 205, and 858 (Table 1-2b). Interestingly, chard contained the most LSV of the crop types yet showed high diversity (*2pq*), suggesting some unique variation supports the divergence of this lineage. Diversity contained within the species, crop type, and individual populations was estimated using expected heterozygosity (*2pq*) (Table 1-2 and Figure 1-2). Expected *heterozygosity (2pq)* varied from 0.027 in our inbred reference EL10 to 0.253 in the recurrent selection population GP9. Within the crop types, the mean expected heterozygosity for sugar beet was 0.207, table beet = 0.148, fodder beet = 0.221, and chard = 0.216.

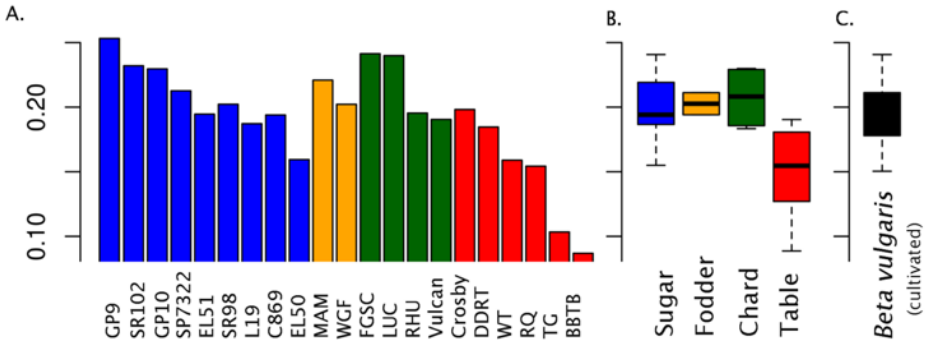

Figure 1-2: **Gene diversity/expected heterozygosity (*2pq*) of *B. vulgaris* lineages.** (A) Populations, (B) Crop types, and (C) Species

The expected heterozygosity (*2pq*) for populations such as EL10 and W357B was low. This was expected owing to inbreeding via the presence of self-fertility alleles. These populations were excluded from further analysis because of the lack of variation. Interestingly, the population Bulls Blood lacks variation relative to other beet populations, it remains unknown if selection, sib mating, or self-fertility underlie this result. The variation in diversity estimates as measured by expected heterozygosity (*2pq*) in these populations suggests the level of diversity is highly dependent on the breeding system, selection history, and sample size (N).

The variation detected was used to cluster populations in two ways: (1) a hierarchical clustering based on relationship coefficients estimated using the quantity of shared variation between populations, and (2) a principal components analysis using allele frequency in each population, estimated using an IBS (Identity by State) approach. The resulting dendrogram and heatmap showed that the table beet crop type was the only group to have strong evidence (e.g., high relationship coefficients and bootstrap values) supporting it as a unique group harboring significant variation (Table 1-5). Likewise, the green (LUC and FGSC) and red (RHU and Vulcan) chard populations showed evidence for two distinct groups (Figure 1-3). Sugar beet lineages with known pedigree relationships and high probability for shared variation (e.g., SR98/2 and EL51) also had strong evidence, which supports the delineation of population structure on the basis of shared variation. Additionally, the clade composed of SP7322, SR102, GP10, and GP9 resolved in a similar fashion of population delineation on the basis of shared variation.

PCA used genome-wide allele frequency estimates for individual populations. PC1 explained 75.6% of the variance in allele frequency and separated the table beet crop type from other crop types. PC2 explained 15.25% of the variance (Figure 1-4). Sugar and table appear the most divergent and were able to be separated along both dimensions. Chard and fodder crop types were distinguishable but appeared less divergent. Allele frequency estimates analyzed on a chromosome-by-chromosome basis demonstrated that specific chromosomes cluster the populations by crop type (Figure 1-5). Chromosomes 3, 8, and 9 appear to be important for the divergence between sugar beet and other crop types. All chromosomes were able to separate table beet with the exception of Chromosomes 7 and 9.

Table 1-5: **Pairwise relationship matrix.**

Table 5- Relationship Coefficients

	BBTB	C86925	Crosby	DDRT	EL50	EL51	FGSC	GP10	GP9	L19	LUC	MAM	RHU	RQ	SP7322	SR102	SR98	TGSC	Vulcan	W357B	WGF	WT
BBTB	0.5/**218557**	0.07	0.10	0.09	0.05	0.08	0.05	0.08	0.07	0.06	0.06	0.08	0.06	0.09	0.07	0.07	0.08	0.07	0.06	0.05	0.08	0.08
C86925	**52750**	0.5/**511754**	0.13	0.12	0.12	0.16	0.12	0.19	0.17	0.15	0.13	0.17	0.12	0.12	0.17	0.19	0.17	0.07	0.12	0.02	0.14	0.10
Crosby	**70695**	**130759**	0.5/**505245**	0.19	0.11	0.13	0.12	0.14	0.13	0.12	0.13	0.16	0.12	0.16	0.14	0.14	0.13	0.11	0.12	0.03	0.15	0.17
DDRT	**59441**	**113827**	**188971**	0.5/**470498**	0.10	0.11	0.11	0.13	0.11	0.10	0.12	0.14	0.10	0.16	0.12	0.12	0.12	0.11	0.10	0.04	0.13	0.17
EL50	**34786**	**115514**	**104394**	**86483**	0.5/**423153**	0.13	0.10	0.17	0.16	0.12	0.10	0.13	0.11	0.10	0.17	0.17	0.14	0.08	0.11	0.03	0.12	0.09
EL51	**58216**	**164300**	**129653**	**106058**	**125234**	0.5/**527330**	0.12	0.22	0.18	0.16	0.12	0.16	0.12	0.11	0.18	0.20	0.22	0.07	0.12	0.02	0.14	0.10
FGSC	**48756**	**151732**	**146481**	**127319**	**112853**	**146910**	0.5/**702758**	0.13	0.13	0.12	0.26	0.14	0.14	0.10	0.13	0.14	0.12	0.07	0.14	0.02	0.13	0.09
GP10	**59830**	**204521**	**152642**	**130954**	**168981**	**237545**	**169436**	0.5/**571955**	0.22	0.19	0.14	0.19	0.14	0.12	0.21	0.24	0.23	0.08	0.13	0.02	0.16	0.11
GP9	**51352**	**180426**	**134556**	**108977**	**154461**	**192621**	**162204**	**246429**	0.5/**558778**	0.16	0.13	0.16	0.13	0.10	0.19	0.22	0.18	0.07	0.12	0.02	0.14	0.09
L19	**50472**	**156454**	**130465**	**108893**	**123389**	**175104**	**156569**	**211054**	**180008**	0.5/**566702**	0.13	0.17	0.12	0.10	0.18	0.19	0.17	0.07	0.12	0.02	0.14	0.09
LUC	**55846**	**165652**	**164455**	**146440**	**125134**	**161585**	**381976**	**189218**	**177229**	**180158**	0.5/**784655**	0.15	0.16	0.10	0.14	0.14	0.12	0.07	0.15	0.02	0.14	0.10
MAM	**61623**	**176572**	**160351**	**137020**	**124312**	**170896**	**174972**	**205630**	**178356**	**181943**	**199127**	0.5/**526512**	0.14	0.12	0.17	0.19	0.17	0.08	0.13	0.02	0.18	0.12
RHU	**52499**	**136685**	**128676**	**108450**	**112381**	**137265**	**184945**	**158959**	**151926**	**141290**	**214992**	**157764**	0.5/**593875**	0.09	0.13	0.14	0.12	0.06	0.26	0.02	0.13	0.10
RQ	**57154**	**107465**	**148436**	**140456**	**85037**	**101362**	**110558**	**120391**	**94413**	**98677**	**123557**	**117199**	**91943**	0.5/**420894**	0.11	0.11	0.11	0.13	0.09	0.04	0.12	0.15
SP7322	**53847**	**185002**	**151158**	**126094**	**171708**	**199097**	**172329**	**251045**	**219342**	**205770**	**192226**	**193456**	**158604**	**117055**	0.5/**599548**	0.22	0.18	0.08	0.12	0.02	0.16	0.11
SR102	**54973**	**211401**	**150303**	**127972**	**177872**	**223012**	**177448**	**284159**	**253960**	**216634**	**191365**	**210374**	**161968**	**116369**	**266349**	0.5/**596710**	0.20	0.08	0.13	0.02	0.15	0.11
SR98	**55697**	**179546**	**136061**	**119493**	**133219**	**230872**	**145391**	**247637**	**194809**	**183163**	**161550**	**180241**	**138750**	**106581**	**201371**	**228776**	0.5/**523580**	0.08	0.12	0.02	0.15	0.10
TGSC	**32552**	**52159**	**78982**	**78294**	**49332**	**52930**	**61870**	**62273**	**58319**	**53403**	**69764**	**61460**	**50803**	**83935**	**63587**	**63884**	**56711**	0.5/**222986**	0.06	0.06	0.08	0.09
Vulcan	**48283**	**129522**	**125065**	**107622**	**107822**	**127374**	**174260**	**149596**	**139797**	**132969**	**200013**	**144474**	**308198**	**87807**	**146542**	**152448**	**129704**	**48370**	0.5/**577065**	0.02	0.12	0.09
W357B	**14009**	**11506**	**18890**	**19363**	**13181**	**12743**	**13185**	**13356**	**11162**	**13191**	**14970**	**13110**	**13110**	**19781**	**13957**	**14014**	**12822**	**17771**	**13120**	0.5/**75094**	0.02	0.04
WGF	**59949**	**142340**	**161910**	**134396**	**118967**	**146166**	**156673**	**175510**	**158721**	**156317**	**182120**	**196282**	**146812**	**110507**	**178625**	**174029**	**154415**	**59683**	**135022**	**14953**	0.5/**539956**	0.12
WT	**53539**	**94388**	**153744**	**151484**	**74886**	**91696**	**105962**	**106858**	**86166**	**92859**	**119696**	**112372**	**96396**	**123629**	**109331**	**107723**	**96302**	**59246**	**90946**	**17906**	**116081**	0.5/**415564**

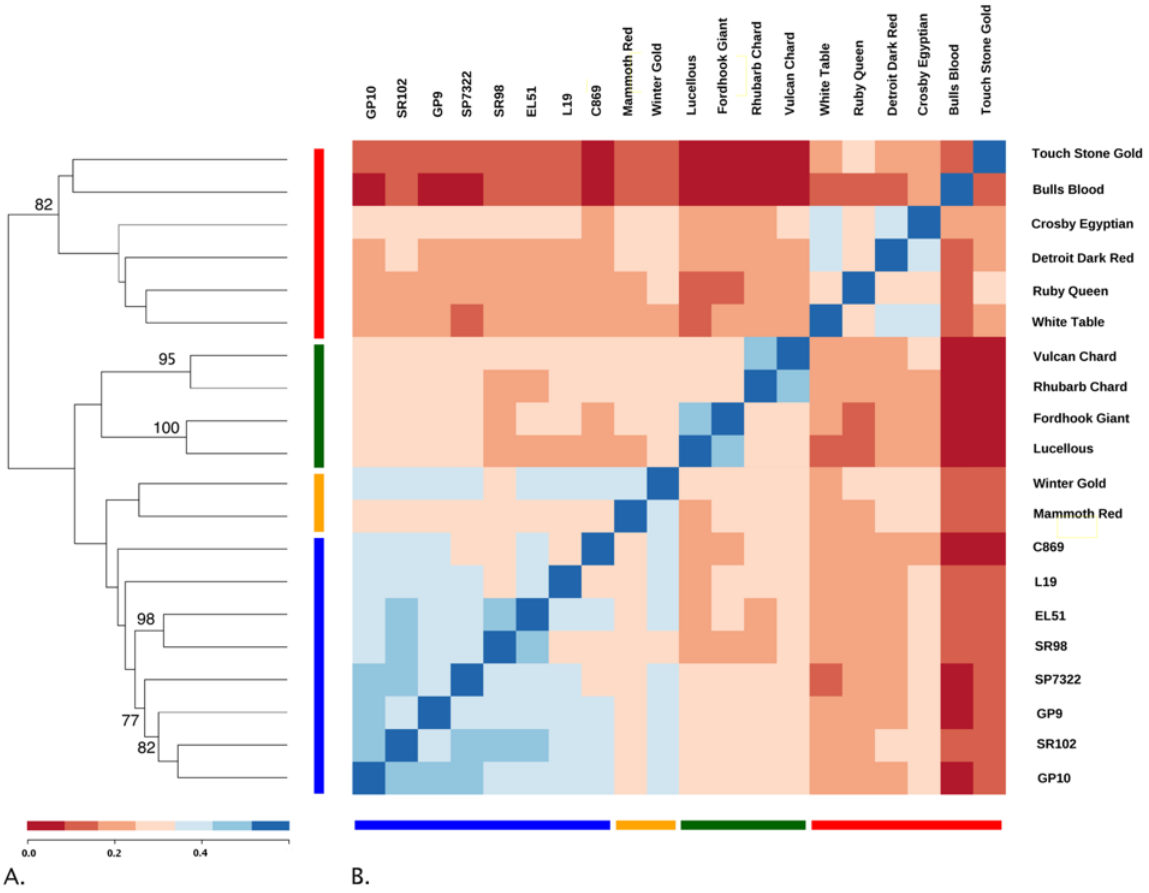

Figure 1-3: **Lineage relationships inferred by hierarchical clustering of pairwise relationship coefficients.** (A) Dendrogram reflects support for clusters and (B) heatmap shows relationship coefficient values for all comparisons.

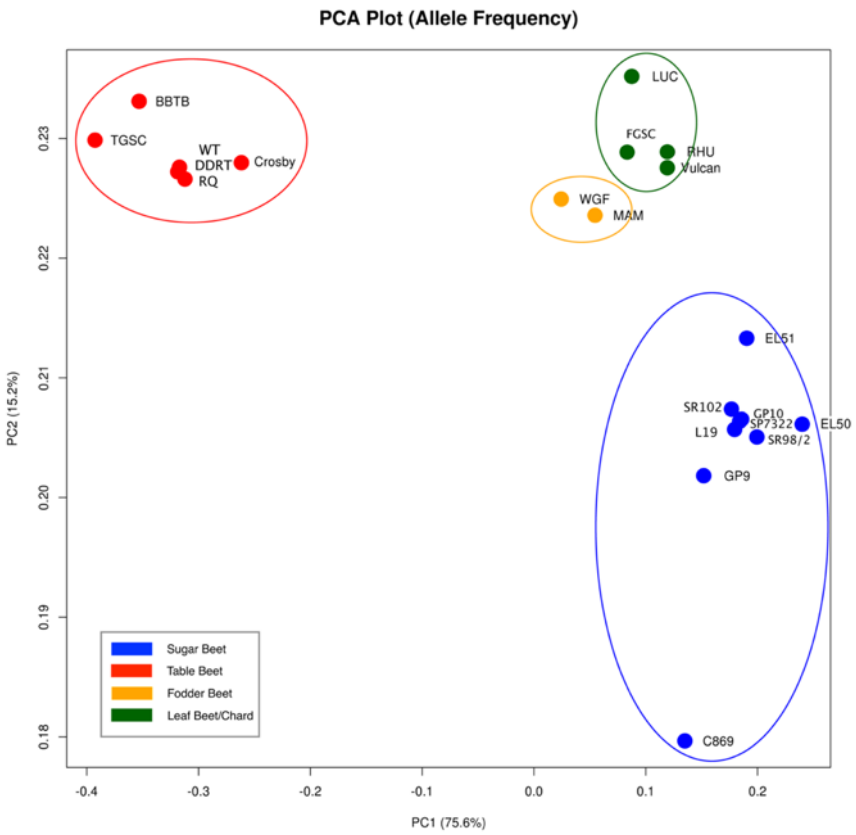

Figure 1-4: **PCA plot showing the separation of crop types using genome-wide allele frequency data.**

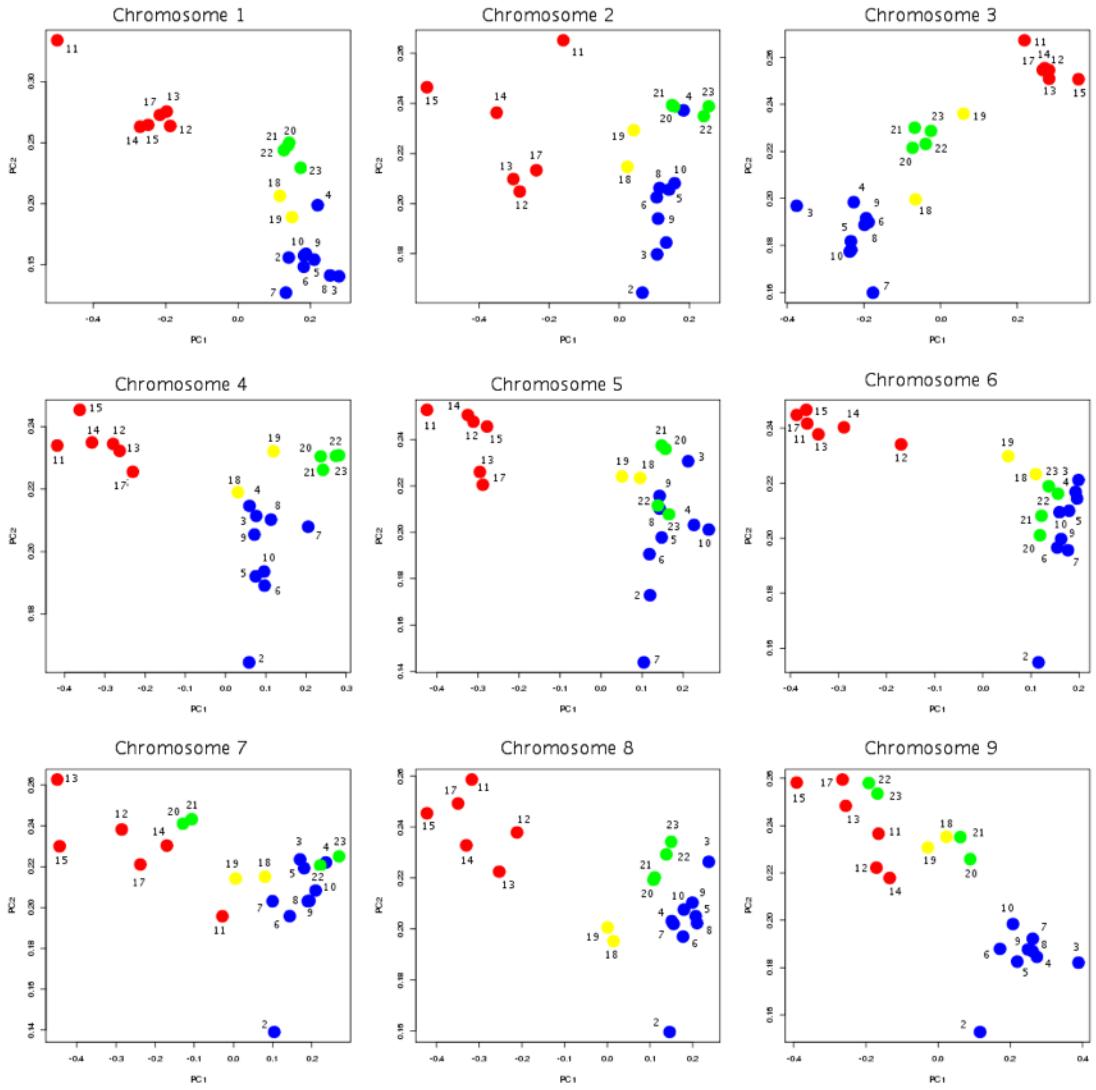

Figure 1-5: **PCA plot showing the separation of crop types using allele frequency data on a chromosome by chromosome basis.**

Finally, using our population genomic data we tested a composite likelihood method to estimate historical effective population size (Ne) and infer demographic histories for crop type lineages. Table beet appears to have a distinct history in terms of historical population size trends as well as demographic splits when compared with the other three lineages. Trends in historical Ne for fodder and sugar groups were quite similar to each other, and no early divergence was detected between them. The chard group appeared to share early demographic history with the fodder/sugar group but showed a different trend later, suggesting it diverged early with respect to the other crop types (Figure 1-6). The demographic history of *B. vulgaris* crop type correlates well to the historical evidence (e.g., records of antiquity, archeological evidence, and scientific literature) detailing the development of distinct crop type lineages (Table 1-6).

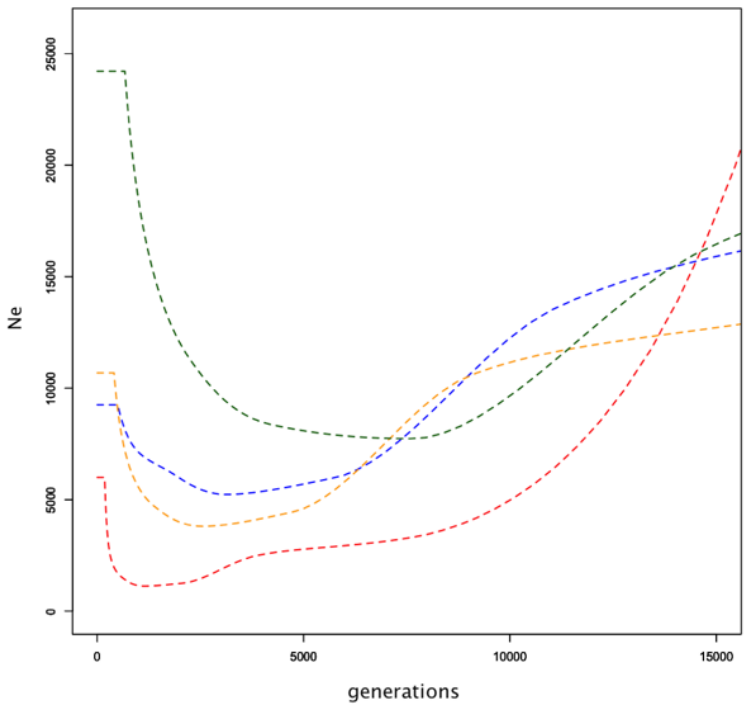

Figure 1-6: **Inferred historical N_e of *B. vulgaris* crop types using the program SMC++.**

Table 1-6: **Historical time line highlighting evidence of beet utilization.**

Date	Source	Description
before 8500 BCE	[1,3,4]	*B. vulgaris* gathered as potherb in Eroupe
8500 BCE	[1,2,3]	The domestication of leaf beet in eastern Turkey
3500 BCE	[1,2]	Leaf and root types present in Egypt
1200 BCE	[1,2]	Leaf beet present in Syria
1000 BCE	[1,2,3]	Leaf beet present in Greece
600 BCE	[1,2]	Leaf beet present in China
460 BCE	[1,4]	Black beet mentioned (perhaps a reference to table beet)
250 BCE	[1,2]	Table beet cultivation spreads
50 BCE	[1,2]	*Beta* cultivation spreads in Roman Empire
1,000 – 1300 CE	[1,2]	Beet described as a garden vegetable, with many types.
1500 CE	[1,2]	Fodder beet spreads across Europe
1747 CE	[1,2,3,4]	Margraff demonstrates sucrose can be extracted from beet
1800 CE	[1,2]	Achard identifies fodder lineages with potential use as a sugar crop
1816–1850 CE	[1,2,4]	Vilmorin develops progeny selection to increase sugar content using differences in specific gravity

[1] Biancardi et al. 2012

[2] Zossimovich 1940

[3] Cook and Scott 1993

[4] Schukowsky 1950

DISCUSSION

The populations sampled here represent significant divergent lineages used in the production of beet. All have undergone significant breeding effort, which has served to capture and fix genetic variation resulting in predictable phenotypes characteristic of each individual within a population or crop type. The organization and distribution of genetic variation within and among populations reflects the historical selection and evolutionary pressures experienced as these crop types, populations, and varieties were developed. Pooled sequencing allowed us to make the cogent genomic comparisons that informs the history of beet development, from ancestral gene pools and domestication to the development of varieties and germplasm within modern breeding programs. Using population genomic data, we were able to support *B. vulgaris* as a species complex, uncover genomic variation associated with development of beet crop types, and gain fundamental insight into the natural history of beet.

Two biological groups could be identified with high confidence using these data, a table beet group and a group encompassing chard, fodder beet, and sugar beet. Previous research, which used genetic markers to cluster crop types, reported similar findings (Mangin et al. 2015, Andrello et al. 2016). The strong evidence for a unique table beet group hints at both genetic drift, resulting from reproductive isolation, as well as positive selection for end use. In general, selection and drift act to change allele frequency within a population (Hedrick 2005), but the effects are relative to the effective population size (Ne) of the populations under selection. Effective population size is an important consideration because it relates to the standing genetic diversity within populations (Crow and Denniston 1988, Waples 1990). The patterns of variation

resulting from drift and selection are distinct. For example, table beet populations had low diversity (*2pq*) relative to other crop types, and the ability to separate table beet populations using allele frequency is suggestive of selection. Relationship coefficients, on the other hand, highlight the differences in the quantity of shared variation within and between crop types, suggesting table beet may have been less connected to other crop type populations. Allele frequency showed signals of differentiation distributed across all chromosomes for table beet, likely reflecting both selection and drift. The low quantity of shared variation between crop types did not support long term phylogeographic explanations for the differentiation observed. Long periods of geographic isolation can produce barriers to reproduction, further reinforcing isolation and divergence of populations (Palumbi 1994). This appears not to be the case in cultivated beet, as experimental hybrids between crop types show few barriers to hybridization and produce viable progeny, which does not suggest a large degree of chromosomal variation between the groups. The creation of segregating populations from crosses between sugar and table beet crop types support this observation (McGrath et al. 2007, Laurent et al. 2007).

The lesser degree of separation between chard, fodder, and sugar crop types may be the result of increased connectivity (e.g., historical gene flow) between these lineages versus table beet. High gene flow exerts a homogenizing effect on the diversity contained within populations and increases the quantity of shared variation. This may explain a lack of clear delineation of these crop types using genome-wide markers. Fodder and sugar crop types separated using allele frequency but not shared variation. This was not unexpected given the known history between these lineages. The development of fodder lineages that accumulate sucrose have occurred in recent history (~200 years), giving rise to the progenitor of sugar beet, the ‘White Silesian’

(Fischer 1989, Winner 1993). This was reflected in the low quantity of indel LSV detected within both crop types. Interestingly, phenotypic divergence between species is attributed more to indel variation than to SNP variation owing to their greater consequences on gene expression and gene regulation (Chen et al. 2009). This phenomenon may be visible in population divergence as well as speciation. The high quantity of shared variation between sugar and fodder crop types relative to comparisons between other crop types suggests a close relationship and shared demographic history that includes selection. The high quantity of shared variation between the sugar beet, fodder beet, and chard crop types versus table beet highlights the variable extent and timing of gene flow between lineages.

Chard, being was the first crop type developed from diverse ancestral *B. vulgaris* spp. *maritima* populations (Biancardi et al. 2012, Winner 1993) is supported by the high level of diversity (*2pq*), a high quantity of LSV, and an interesting demographic history. The clear delineation of two distinct chard groups suggests different demographic histories. Although the chards share similar leaf morphology, color, and root morphology of these groups is different in that the roots of the red chard group were enlarged and had fewer 'sprangles' (adventitious roots branching from the tap root) with respect to the green chards but not to the extent as in the root types (e.g. sugar, fodder, and table). This may reflect introgressions between the red chard and a root type, potentially fodder or table beet, and potentially an unintended consequence of breeding for color, but this was not obvious at the whole genome level or even at the level of chromosomes.

The enlarged tap root character appears to have been first developed in table beet lineages (Biancardi et al. 2012), but the expanded root character is shared across crop type lineages. This suggests two plausible hypotheses: (1) the root character in fodder beet reflects the introgression

of this character from a table beet to a chard background, or (2) an ancestral population gave rise to the root character that diverged into fodder and table lineages. Historically, it appears admixture, hybridization, and introgression were fundamental to the development of beet lineages and populations. Schukowsky (1950) suggested that the broad adaptation of beet to novel growing environments may be due to variation accumulated in geographically diverse ancestral populations and shared via admixture and gene flow between lineages. Adaptive trait variation from wild relatives is becoming increasingly important in light of changing conditions across the growing regions of many crop species (Takuno et al. 2015). Distinguishing between sorting ancestral variation and introgression events remains a challenge but could yield important insight into beet crop type development, and other cultivated species as well.

The beet crop types have appeared to have diverged by selection. The variance in allele frequency of bi-allelic SNPs between populations was able to separate the crop type groups. This suggests that the allele frequency data contains a signal related to selection. Sugar and table beet appear to be the most diverged, which is consistent with large breeding efforts for each of these crop types. Allele frequency data on a per chromosome basis demonstrated that crop types are variable with respect to specific chromosomes. Ostensibly the presence of variation located on specific chromosomes is under positive selection for end use, leading to an accumulation of lineage-specific differences including those linked to defining phenotypic characters. Many quantitative trait loci studies support the fact that specific regions along chromosomes contain the variation that ultimately influences phenotype (Doerge 2002). Population divergence in the presence of gene flow produces distinct patterns of variation with respect to selection (Martin et al. 2013). Cryptic relationships within other species complexes have been explained by the

islands-of-differentiation model (Waples 1998, Bickford et al. 2007). Islands of differentiation may be common in species with high gene flow because selection increases the frequency of beneficial alleles and gene flow acts to return neutral variation to equilibrium frequencies. Allele frequency estimates for specific chromosomes as well as the distribution of lineage-specific variation for crop type on specific chromosomes suggests a small degree of total genome differentiation, which appears to be localized to specific chromosomes and likely localized chromosome regions. Interestingly, small amounts of variation can have profound effects on phenotypic variation (Doebley and Stec 1993, Meyer and Purugganan 2013).

Given the support for crop type relationships it appears the divergence of beet crop types occurred in the presence of high gene flow. Admixture and introgression events may have served to share genetic variation across cultivated beet populations and crop type lineages, which in turn, created challenges for the clear delineation of subpopulations. This is confounded by the fact that, as lineages evolve, a lesser quantity of variation with greater agricultural importance contributes to our notion of economic and agronomic value. Resolving the degree to which historical admixture and introgression has contributed to the development of beet crop type will require more in-depth analysis of the variation at nucleotide level within local chromosome regions.

LITERATURE CITED

LITERATURE CITED

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, B. Desprez *et al.*, 2017 Insights into the genetic relationships among plants of *Beta* section *Beta* using SNP markers. Theor. Appl. Genet. 130: 1857–1866.

Andrews, S., 2010 FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

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

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

Bickford, D., D. J. Lohman, N. S. Sodhi, P. K. L. Ng, R. Meier *et al.*, 2007 Cryptic species as a window on diversity and conservation. Trends Ecol. Evol. 22: 148-155

Bolger, A. M., M. Lohse, and B. Usadel, 2014 Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114-2120.

Casillas, S., and A. Barbadilla, 2017 Molecular population genetics. Genetics 205: 1003–1035.

Chen, J. Q., Y. Wu, H. Yang, J. Bergelson, M. Kreitman *et al.*, 2009 Variation in the ratio of nucleotide substitution and indel rates across genomes in mammals and bacteria. Mol. Biol. Evol. 26: 1523–1531.

Cooke, D. A., and R. K. Scott, 1993 *The Sugar Beet Crop*. Chapman and Hall Publishers, London.

Crow, J. F., and C. Denniston, 2006 Inbreeding and variance effective population numbers. Evolution 42: 482–495.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks *et al.*, 2011 The variant call format and VCFtools. Bioinformatics 27: 2156-2158.

Doebley, J., and A. Stec, 1993 Inheritance of the morphological differences between maize and teosinte: Comparison of results for two F2 populations. Genetics 134: 559–570.

Doerge, R. W., 2002 Mapping and analysis of quantitative trait loci in experimental populations. Nat. Rev. Genet. 3: 43–52.

Dohm, J. C., A. E. Minoche, D. Holtgräwe, S. Capella-Gutiérrez, F. Zakrzewski *et al.*, 2014 The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). Nature 505: 546–549.

Excoffier, L., P. E. Smouse, and J. M. Quattro, 1992 Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. Genetics 131: 479–491.

Fischer, H. E., 1989 Origin of the "Weisse Schlesische Rübe" (white Silesian beet) and resynthesis of sugar beet. Euphytica 41: 75–80.

Ford Lloyd, B. V., 1995 Sugarbeet, and other cultivated beets. In: Evolution of crop plants (J. Smartt & N. W. Simmonds, Eds.). Longman Scientific & Technical, Essex, U.K.

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

Gayon, 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. Journal of the History of Biology 31: 241-262.

Goldman, I. L., and J. P. Navazio, 2002 History and breeding of table beet in the United States. Plant Breed. Rev. 22: 357–388.

Gompert, Z., M. L. Forister, J. A. Fordyce, C. C. Nice, R. J. Williamson *et al.*, 2010 Bayesian analysis of molecular variance in pyrosequences quantifies population genetic structure across the genome of *Lycaeides* butterflies. Molecular Ecology 19: 2455–2473.

Hedrick, P., 2005 *Genetics of Populations*. Jones and Bartlett Publishers, Sudbury, Massachusetts.

Kuzmina, E. E., 2008 *The Prehistory of the Silk Road*. University of Pennsylvania Press, Philadelphia, Pennsylvania.

Langmead, B., and S. L. Salzberg, 2012 Fast gapped-read alignment with Bowtie 2. Nat. Methods 9: 357-359.

Laurent, V., Devaux, P., Thiel, T., Viard, F., Mielordt, S., Touzet, P., and Quillet, M., 2007 Comparative effectiveness of sugar beet microsatellite markers isolated from genomic libraries and GenBank ESTs to map the sugar beet genome. Theor. Appl. Genet. 115: 793-805.

Li, H., 2011 A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27: 2987-2993.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan *et al.*, 2009 The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078-2079.

Lynch, M., 2009 Estimation of allele frequencies from high-coverage genome-sequencing projects. Genetics 182: 295–301.

Mangin, B., F. Sandron, K. Henry, B. Devaux, G. Willems *et al.*, 2015 Breeding patterns and cultivated beets origins by genetic diversity and linkage disequilibrium analyses. Theor. Appl. Genet. 128: 2255-2271.

Manichaikul, A., J. C. Mychaleckyj, S. S. Rich, K. Daly, M. Sale *et al.*, 2010 Robust relationship inference in genome-wide association studies. Bioinformatics 26: 2867-2873.

Martin, S. H., K. K. Dasmahapatra, N. J. Nadeau, C. Salazar, J. R. Walters *et al.*, 2013 Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. Genome Res. 23: 1817-1828.

McGrath, J. M., C. A. Derrico, and Y. Yu, 1999 Genetic diversity in selected, historical US sugarbeet germplasm and *Beta vulgaris* ssp. *maritima*. Theor. Appl. Genet. 98: 968–976.

McGrath, J. M., D. Trebbi, A. Fenwick, L. Panella, B. Schulz *et al.*, 2007 An open-source first-generation molecular genetic map from a sugarbeet × table beet cross and its extension to physical mapping. Crop Sci. 47: S27-S44.

McGrath, J.M., and K.K. Fugate, 2012 *Analysis of Sucrose from Sugar Beet. In: Preedy VR, editor. Dietary Sugars: Chemistry, Analysis, Function and Effects. Food and Nutritional Components in Focus No. 3.* (V. R. Preedy, Ed.). Royal Society of Chemistry Publishing, Cambridge, UK.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis *et al.*, 2010 The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20: 1297–1303.

Meyer, R. S., and M. D. Purugganan, 2013 Evolution of crop species: Genetics of domestication and diversification. Nat. Rev. Genet. 14: 840-852.

Nei, M., 1987 *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Paesold, S., D. Borchardt, T. Schmidt, and 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.

Palumbi, S. R., 2003 Genetic divergence, reproductive isolation, and marine speciation. Annu. Rev. Ecol. Syst. 25: 547-572.

Paradis, E., and Schliep K., 2018 ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35: 526-528.

Schukowsky, P.M. (1950). *The Cultivated Plants and their Relatives* (in Russian). Moscow.

Storz, J. F., 2005 Using genome scans of DNA polymorphism to infer adaptive population divergence. Mol. Ecol. 14: 671–88.

Takuno, S., P. Ralph, K. Swart, R. J. Elshire, J. C. Glaubitz *et al.*, 2015 Independent molecular basis of convergent highland adaptation in maize. Genetics 200: 1297-1312.

Terhorst, J., Kamm, J.A., Song, Y.S., 2016 Robust and scalable inference of population history from hundreds of unphased whole genomes. Nat. Genet. 49: 303–309.

Thioulouse, J., D. Chessel, S. Dolédec, and J. M. Olivier, 1997 ADE-4: A multivariate analysis and graphical display software. Stat. Comput. 7: 75-83.

Waples, R. S., 1990 Conservation genetics of Pacific salmon. II. Effective population size and the rate of loss of genetic variability. J. Hered. 81: 267–276.

Waples, R. S., 1998 Separating the wheat from the chaff: Patterns of genetic differentiation in high gene flow species. J. Hered. 89: 438–450.

Winner, C., 1993 History of the crop. In: *The Sugar Beet Crop* (D.A. Cooke and R. K. Scott, eds.) pp 1-35. Chapman and Hall Publishers, London.

Zossimovich, V. P., 1940 *Wild species and origin of cultivated beets*. p. 17-44. Sveklovodstvo.

CHAPTER 2

QUANTIFYING *BETA VULGARIS* GENOME DIFFERENTIATION WITH RESPECT TO CROP TYPE USING WHOLE GENOME POOLED SEQUENCING

INTRODUCTION

The distribution and organization of genetic diversity within a species results from complex interactions between selection, drift, mutation, migration, recombination, and ancestral variation. Population divergence occurs by selection and drift and can result in heterogeneous genome differentiation (Nosil et al. 2009). Domestication and long-term selective breeding provide an interesting experimental system to study genome differentiation with respect to selection, drift, and the development of important lineages that contain phenotypic characters (Schreiber et al. 2018). The success of plant and animal breeding results, in large part, from our ability to partition heritable variation into lineages with predictable phenotypic outcomes. Selection and drift play a large role in this process, but the effectiveness of selection strategies is influenced by intrinsic factors of the species, including ploidy, reproductive biology, chromosome structure, and standing genome variation.

Root crops are important for food security because of storability and availability as a source of calories when other foodstuffs are not available. *Beta vulgaris* (beet) domestication is unique in that it resulted in the development of distinct crop types. *Brassica spp*. are similar to beet in that selection has produced significant morphotype diversity that fill distinct production niches based on end use. Significant divergence has been found between these groups (Bird et al. 2017). Evolution in *Brassica* differs from beet in that divergence has also been accomplished by changes in ploidy and subgenome dominance (Osborn 2004). *B. vulgaris* crop types include both root types and leaf types. Chard, also referred to as "leaf beet," is consumed as a leaf vegetable and exhibits enlarged leaves and petioles relative to the other beet crop types. The root types

include table beet, which is consumed as a fresh or processed market root vegetable, fodder beet, used for animal feed (Cooke and Scott 1993; Biancardi 2012), and sugar beet, produced for sucrose extraction. Sugar beet was developed recently compared to the other beet crop types (Dohm et al. 2014) and represents an important source of sucrose in temperate regions. Historically, sucrose was a scarce resource, and its production and commoditization was at the center of the global economy (McGrath and Panella 2018).

The domestication of root crops is less understood and differs significantly from grain crops, including common features of the "domestication syndrome" such as reduced seed shattering and synchronous flowering (Zohary and Hopf 2000). Given the importance of nongrain crops in agricultural production, the definition of domestication has recently been revised to include the modification of any plant feature of economic interest (Doebley et al. 2006). Research in sweet potato, yam, turnip, radish, carrot (Scotland et al. 2018; Akakpo et al. 2017; Bird et al. 2017; Kim et al. 2016; Macko-Podgórni et al. 2017; Ellison et al. 2018), and now beet provides an opportunity to compare similarities and differences of genetic mechanisms and pathways involved in root enlargement, expansion, and biomass accumulation. Roots are important plant organs as they provide stability to the aboveground tissues, facilitate nutrient and water uptake, store plant products, and interact with diverse communities of organisms in the rhizosphere. Molecular markers studies have shown selection in different grain crops have targeted orthologous genes such as *shattering1* (Lin et al. 2012). Understanding the loci under historical selection that influence important biology in one species may inform the potential for development of these characters in related species as well (Rendón-Anaya and Herrera-Estrella

2018). The idea of parallel evolution is not new; in fact, these ideas are similar to the law of homologous series proposed by Vavilov (1922).

The Caryophyllales represent a basal eudicot order containing few sequenced genomes. The order is characterized by herbaceous habit and odd ecology (Stevens 2001). Specific families and species include diverse examples of adaptation to extreme environments, such as ice plants (Aizoaceae), cactus (Cactaceae), and fly traps (Droseraceae). Important food crops in the Caryophyllales include beets (*Beta vulgaris*), quinoa (*Chenopodium quinoa)*, amaranth (*Amaranthus* spp.), spinach (*Spinacia oleracea* L.), and various cacti (*Opuntia* spp.). This order is unique in that the majority of plant species produce pigments that are characterized as betalains versus anthocyanins which are color compounds distributed across the majority of plant taxa. The genes coding for the enzymes which drive the biosynthesis of yellow and red pigments in beet, the R and Y locus have been cloned (Halsted et al. 2012 and Halsted et al. 2015). Historically, color has been a useful phenotypic marker because it is erasily scored and the YRB linkage group (Owen 1942) which includes a bolting (B) locus was the first linkage groups described in beet.

Beets are diploid (2n = 18), outcrossing, and generally self-incompatible. Breeding and improvement are accomplished at the level of the population, which contains the requisite diversity for selection. The quantity and distribution of diversity within the genomes of beet populations reflects the timing and intensity of historical selection, drift, and admixture. To date, the result and extent of selective sweeps, historical bottlenecks, and founder effects in the development of distinct crop types and adaptation to growing regions and conditions remains

unknown. Pooled sequencing of beet populations fits the breeding practices, reproductive biology of the species, and the methods for evaluating phenotypic diversity in the field. Often, important traits (e.g., yield, productivity, and disease resistance) are reported as population means. As a result of the high heterozygosity and diversity within populations, a single individual is not necessarily representative of the population from which it was derived. Additionally, the genetic constitution of an individual is hard to maintain because of self-incompatibility and tendency to outbreed. The maintenance and preservation of genetic resources for beet occurs *in vivo* (e.g., seed banks, collections), whereby a lineage is represented by a population of individual seeds. Pooled sequencing data better represents the diversity of a population and its derivatives because allele frequency can be estimated and the diversity reflects the evolutionary pressures a population has experienced. A pooled approach can inform the process of germplasm enhancement, breeding populations, and hybrid seed production. Population comparisons using measures such as F_{ST} that calculate the ratio of variances between two populations can quantify the level of divergence between two populations. Several studies have demonstrated the utility of population genetic inference using pooled data (Ferretti et al. 2013, Kofler et al. 2011). Additionally, genome-wide association and genomic prediction models have been carried out using pooled sequencing data (Gaj et al. 2012). In beets and species with similar genetics, pooled sequencing provides a means to survey the diversity within a species, characterize the genetic base, and inform the efficient utilization of genetic resources for breeding and improvement.

MATERIALS AND METHODS

Beet populations and sequencing

Twenty-five individuals from each of the 23 *B. vulgaris* populations were pooled and sequenced using a pooled sequencing approach. The populations selected represent the four recognized crop types and capture a wide range of phenotypic diversity found within cultivated beet (Chapter 1). Populations were grown in the greenhouse, and leaf material was harvested from 25 individuals per population. Leaf material was pooled and homogenized, and DNA was extracted using the Macherey-Nagel NucleoSpin Plant II Genomic DNA extraction kit (Bethlehem, PA). One microgram of DNA for each population was submitted to the MSU Genomics Core, where NGS libraries were constructed using TruSeq bar-code adapters. The sequencing reactions were carried out on the Illumina Hi-Seq 2500 in a 2 x 150 bp paired-end format with a target coverage of 80x relative to the predicted 758 Mb genome size of beet (Arumuganathan and Earle 1991). Post sequencing, read quality was assessed using FastQC (Andrews 2010). Library bar-code adapters were removed and reads were trimmed according to a quality threshold using TRIMMOMATIC (Bolger et al. 2014) invoking the following options (ILLUMINACLIP:adapters.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36). These filtered reads were used for downstream analysis.

Data processing and variant detection

The reference genome generated from sugar beet accession EL10 represents the most contiguous and complete *B. vulgaris* genome assembly to date (Funk et al. 2018). Variants for each population were called by aligning the filtered reads to the EL10.1 *B. vulgaris* reference genome

assembly using Bowtie2 v2.2.3 with the following parameters (bowtie2 -q --phred33-quals -k 2 -x) (Langmead and Salzberg 2012). The resulting alignment files were sorted and merged using SAMtools (Li et al. 2009). SNP (single nucleotide polymorphism) variants were called for each population using BCFtools (Li 2011), filtered for mapping quality (MAPQ >20) and read depth (n > 15) and combined using VCFtools (Danecek et al. 2011). The data was filtered to obtain biallelic SNP loci across all populations.

2pq – Gene diversity/expected heterozygosity of biallelic sites

The mpileup subroutine in SAMtools was used to quantify the alignment files and extract allele counts. Allele frequency was estimated allele counts for biallelic SNP sites determined at the species level. Population parameters were then estimated using the allele frequency within each population such that ($p + q = 1$). The variable p was designated as the allele state of the EL10.1 reference genome and q as the alternate state. Expected heterozygosity ($2pq$), also termed gene diversity (Nei 1987), ranges from 0 to 0.5 and was used as the means to compare diversity contained within the genomes for each crop type.

F_{ST} – *differentiation*

F_{ST} was used to calculate differentiation between a single crop and all other crop types. F_{ST} is defined as the ratio of variances between two populations (Wright 1951); subsequently it was used to determine population structure and divergence (Weir and Cockerham 1984). Weir and Hill (2002) define F_{ST} as the correlation between alleles drawn at random from two populations relative to the most common ancestral population. Genome scans using SNP data and population genetic inference is a powerful tool in order to identify causal variation (Nielsen et al. 2005).

The allele counts for each biallelic SNP loci were combined across populations representing a specific crop type and used to estimate allele frequency for the crop type. Allele frequency was used to determine the differentiation of each crop type relative to all other crop types by estimating F_{ST} for all loci (Eq. 1). F_{ST} was calculated at the locus level, within a 5000 bp and 50,000 bp window, with a step size of 100 and 1000 bp, respectively. Ultimately, a sliding window of 25 biallelic variant sites, 12 upstream and 12 downstream from a given locus, was used in order to obtain a uniform sample size for use in the equation to maintain statistical power. The distribution of F_{ST} across the *B. vulgaris* genome with respect to crop type differentiation was evaluated. The numerator of the equation represents the variance in allele frequency of a single crop type and the denominator, the total variance in allele frequency in all crop types. The result is the proportion of variance in allele frequency explained by a single crop type or the genetic differentiation of a single crop type relative to all other crop types. Values for F_{ST} range from 0 to 1 with values close to 0 indicating panmixia, high gene flow and little divergence (e.g. less population structure) and values close to 1 suggesting a high degree of divergence (e.g. high degree of population structure). A one-sided Wilcoxon test was performed using the function (wilcox.test) in R in order to determine the level of significance (p-value) of any biallelic SNP within the distribution. Both the empirical distribution of FST and traditional thresholds (Meirmans and Hendrik 2011) for interpreting F_{ST} were considered. F_{ST} values from 0 to 0.3 were deemed undifferentiated (e.g. weak population structure), 0.30 to 0.60 were considered differentiating (e.g. some population structure), 0.6 to 0.9 were considered differentiated (e.g. population structure), and >0.90 were considered highly differentiated (greatest degree of population structure). The degree of differentiation and significance of F_{ST} values are dependent on many factors including the choice of estimators, N size of populations, and comparisons performed. Specific factors related to the population and

species include the reproductive biology of the species, and complex interactions between selection, mutation, migration, and drift. A closer examination of the F_{ST} distribution allowed the identification of outliers by selecting sites on the upper tail of the distribution in order to reduce the number of genes for further investigation.

(Eq. 1) $F_{ST} = \frac{\sigma_S^2}{\sigma_T^2} = \frac{\sigma_S^2}{\bar{p}(1-\bar{p})}$

Equation 1: shows F_{ST} is defined as the ratio of variance in allele frequency of the subpopulation (s) relative to the total population (t), where p is the allele frequency of allele (p).

The span of significant F_{ST} values across large regions was considered important owing to potential linkage disequilibrium (LD), although LD was not directly measured. Significant regions were quantified by evaluating the size of the region that contained a signal of significant loci ($F_{ST} > 0.6$), allowing the signal to drop below the threshold across two consecutive loci before estimating its size (bp). Additionally, loci with significant F_{ST} were characterized as genic, exonic, intronic, or within 500 and 1000 bp flanking a gene. Differentiation was evaluated for crop types, chromosomes, and crop type by chromosome using F_{ST}.

Lineage-specific variation

LSV or homozygous private variation was extracted from the merged VCF file containing the variants for all populations. The characterization of variation as LSV required the variant to be fixed within a defined population or crop type and not detected within any other population or

crop type. VCF files representing LSV were produced for each population and crop type (Chapter 1).

Genes/F_{ST} Outliers

Genes in close proximity to differentiated loci (e.g., within 1000 bp) were evaluated for putative biological functions and potential involvement with important phenotypic variation. Gene coordinates were extracted from the annotation file (.gff) for the EL10.1 reference genome assembly (http://sugarbeets.msu.edu/data). Gene function was evaluated using the EL10.1 annotation file, InterPro scan output for predicted proteins, and the BLASTp output using predicted proteins against TAIR. Best hits from blast were used to query GO terms using Gene Ontology Consortium enrichment analysis tool (Ashburner et al. 2000, GO Consortium 2017) using Arabidopsis gene identifiers.

Visualization of genome differentiation

Python and bash were used to extract and filter the data in order to visualize population genomic variation with respect to gene density, repeat density, and useful cytogenetic landmarks. Gypsy and copia repeats were extracted from the output of LTR_Retriever (Ou et al. 2018). Gene density was calculated on the basis of positional information within the (.gff) file (Funk et al. 2018). Sequences representing the main satellites used in florescent *in situ* hybridization with *B. vulgaris* chromosomes (Paesold et al. 2012) were aligned to the EL10 reference genome using BLAST (blastall -p blastn -d ${genome} -i ${Var} -o ${Var}.out -e 0.001 -a 4 -m 8) (Altschul et al. 1990). The location of each sequence was plotted and used to link the *in silico* bioinformatic analysis with physical chromosome marks. Plotting these data allowed the visualization of

unique variation within individual populations. The function used for the placement of variation in a circular output was extracted from the source code Rcircos (Zhang et al. 2013). Otherwise general R plotting libraries (R Core Team 2013) were used.

Visualization of crop type differentiation

Genome-wide differentiation was plotted using averaged expected heterozygosity (*2pq*) for all crop type populations, and F_{ST} calculated on the basis of crop type. The raw values for *2pq* were not informative because of their high variability. Ultimately, a rolling average was calculated using 100 kb windows with a 20 kb step proved to be the most informative at the level of whole genome. LTR_Retriever was used to identify gypsy elements and density plots across the genome was used to determine putative centromere locations. The delineation of chromosome features and suspected gene function was evaluated to assess the accumulation of genetic variation and evolutionary potential of these regions (e.g., euchromatic, pericentric, centromeric). This procedure was done for the whole genome as well as on a chromosome by chromosome basis. Code is available for these plots (www.github.com/beetgenomeninja/).

Gene plots (allele frequency)

Gene coordinates were extracted from the (.gff) file and the allele frequency data for all populations were used to plot local allele frequency for the gene plus 1000 bp of sequence flanking the gene on each end. Plots include the predicted gene model, which allowed for a characterization of variation (e.g., gene body, start, stop, introns, exons, and promoters).

RESULTS

Genetic variation within cultivated B. vulgaris

To understand the degree of genome differentiation between *Beta vulgaris* crop type lineages, 25 individuals from each of the 23 *B. vulgaris* populations were pooled and sequenced in a 2 x 150 bp paired end format with a target coverage of 80x relative to the predicted 758 Mb size of the beet genome. On average, 61.84 ± 12.22 GB of sequence data was produced per population, with an average depth of 81.5X. After processing for quality, reads were aligned to EL10.1 reference genome. Biallelic SNP markers and lineage-specific variation (LSV) (Chapter 1) were used to estimate the quantity and organization of genome-wide variation within *B. vulgaris* populations and hierarchical groups (e.g., species, crop types, and populations). On average, 90.74% of the filtered reads aligned to the EL10.1 reference genome. Approximately 20% of bases were discarded as a result of trimming of low-quality base calls and adapter sequences. A total of 14,598,354 variants were detected across all populations, and 12,411,164 (85.0%) of these were classified as SNP variation, and of these SNPs, 10,215,761 (82.3%) were biallelic. After filtering for read depth ($n \geq 15$), 8,461,457 biallelic SNPs remained for computational analysis. Insertion and deletions (indels) accounted for 2,187,190 (14.9%) of the variants detected. Additionally, 2,718,205 (18.6%) variants were characterized as multiallelic.

Lineage-specific variation (individuals)

Lineage-specific variation was evaluated for individual populations. The unique variation with respect to individual populations and crop types reflects the evolutionary history of the species. (Chapter 1; Figure 2-1). Regions that lacked LSV suggest physical positions where variation is

shared between related populations and/or crop type lineages. The accumulation of LSV across the genome highlighted both regions of differentiation as well as the similarity between genomes of cultivated beet populations and crop types.

Gene diversity/expected heterozygosity

Regions devoid of sequence polymorphism across the genome with respect to crop type were inferred by the distribution of expected heterozygosity (*2pq*). This was done for each population using the allele frequencies of biallelic SNP markers (n = 8,461,457). A rolling average was performed on the expected heterozygosity estimates for each crop type using a window size of 100 kb with a step of 20 kb (Figure 2-2).

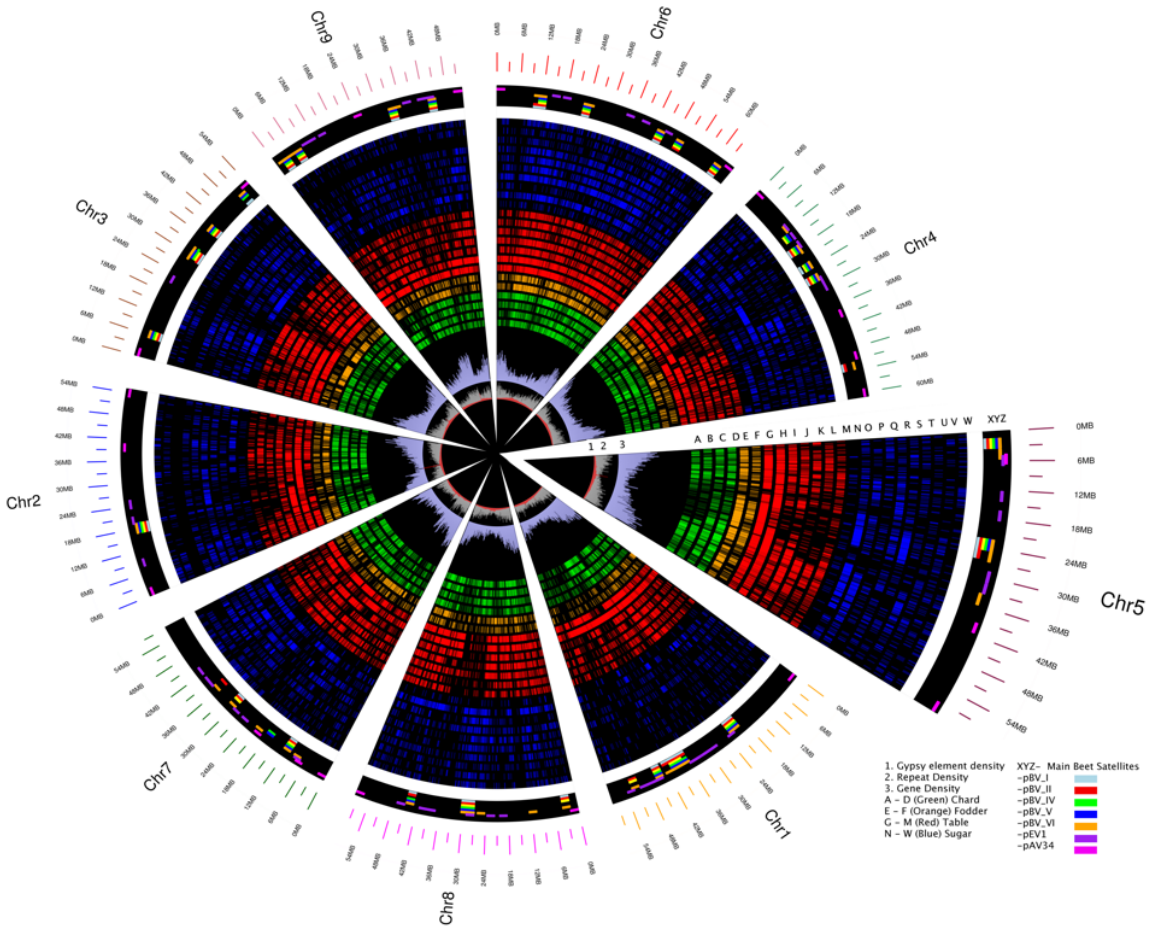

Figure 2-1: **Distribution of lineage-specific variation across chromosomes of cultivated beet.** Crop types are represented by colored bars, chard (green), fodder beet (orange), table beet (red) and sugar beet (blue). Individual populations by letters (Tracks A-W). Lineage specific variation is plotted with respect to (1) Gyspy element density, (2) repeat element density, (3) gene density and (xyz) major satellites used in cytogenetic studies of beet chromosomes (Paesold et al. 2012).

B. vulgaris genome

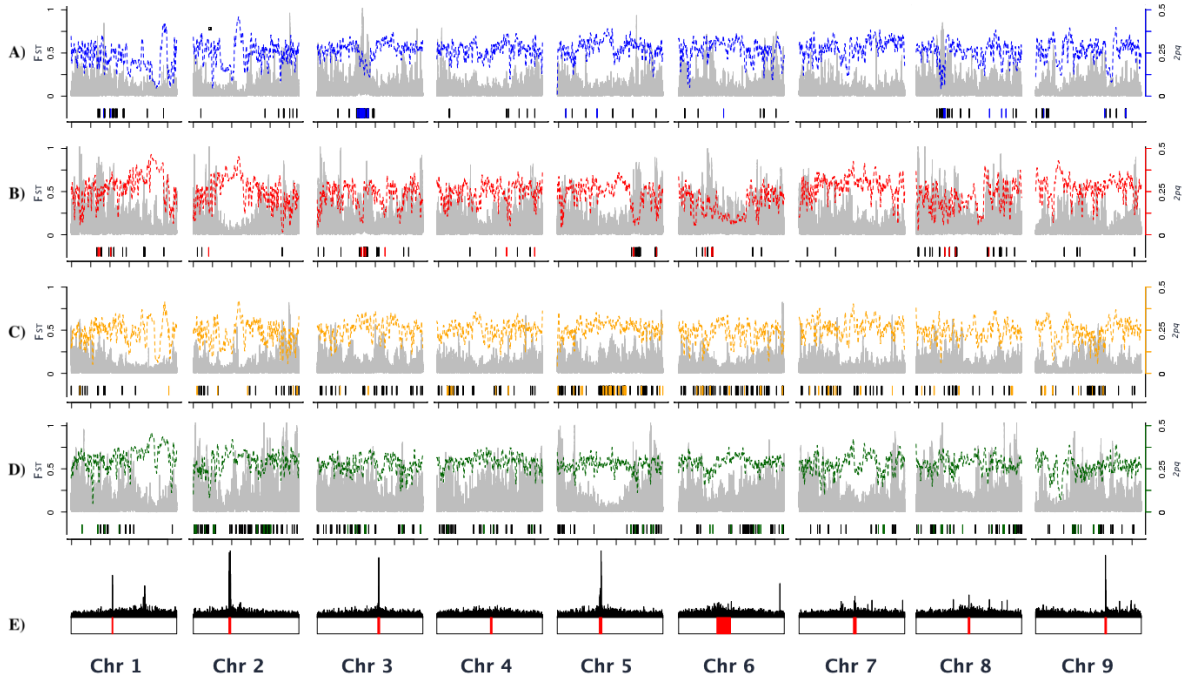

Figure 2-2: **Topology of crop type variation across the genome.** Expected heterozygosity and F_{ST} plotted across *B. vulgaris* chromosomes 1 through 9 (left to right). (A) Sugar beet, (B) table beet, (C) fodder beet, (D) chard/leaf beet. Dashed lines represent average *2pq* for all populations representing cultivated *B. vulgaris*. Gray background represents the statistic F_{ST}. Below each plot is the crop type specific variation; color = Indel, black = SNP. € Putative centromere (red) indicated is by gypsy element density along chromosome.

Crop type differentiation (F_{ST})

Allele frequency estimates were used to calculate F_{ST} and measure the degree of differentiation between *B. vulgaris* crop type genomes. The distribution of F_{ST} across all loci was skewed toward zero (Table 2-1), showing a small percent of the genome was differentiated ($F_{ST} > 0.6$) with respect to crop type. Percent differentiated was calculated as number of SNP loci ($F_{ST} > 0.6$) /Total number of biallelic SNP loci (n = 8,461,457). In total 12.13% (1,020,913 bp) of the genome was differentiated with an average of 3.03% per crop type (Figure 2-2 and Figure 2-3). Of these differentiated sites, 33.71% of were detected in genic regions. Within genic regions, differentiated sites were further divided into intron (27.38%) and exon (6.33%) regions. Furthermore, 13.25% of the differentiated loci were detected within 1000 bp flanking a gene (Table 2-2). The distribution of this differentiation across all nine *B. vulgaris* chromosomes is shown in Figure 2-2, Table 2-S1, and Table 2-S2. SNP loci with significant F_{ST} values ($F_{ST} > 0.6$) were distributed within 20,249 regions across the genome with a mean size of 1,402 bp per region. Regions of differentiation ($F_{ST} > 0.6$) for Chromosome 3 in sugar beet had a mean size of 2,650 bp and a large quantity of the differentiation was located between 20-28 Mb. This highlights the importance of this region in the development of sugar beet lineages and potential linkage disequilibrium resulting from historical selection (Figure 2-3). Regions of significance on other chromosomes with respect to crop type can be observed in Figures 2-S1 through Figure 2-S8.

Table 2-1: **Results of Wilson-Cox test.**

	F_{ST}	N SNPs	Percentile
Total SNP		8414286	1
Undifferentiated	x < 0.3	7832938	0.9309
Starting to be differentiated	x > 0.3, x < 0.6	550446	0.0654
Differentiated	x > 0.6, x < 0.9	29218	0.0035
Highly differentiated	x > 0.9	1684	0.0002

** P-values calculated from a one-sided Wilson-Cox Test of the F_{ST} distribution*

Differentiation of B. vulgaris crop types

Specific chromosomes were more or less differentiated with respect to crop type (Figure 2-2, Table 2-3). In sugar beet, 1.23% (103,903 bp) of loci were characterized as differentiated. Chromosomes 3, 6, and 8 accounted for 0.5%, 0.14%, and 0.22% of the total differentiation, respectively. In total, 5.18% (436,106 bp) of loci were characterized as differentiated in table beet and Chromosomes 1, 6, and 8 contained 0.73%, 0.84%, and 1.05% of the total differentiation, respectively. Only 0.56% (47006 bp) of loci were characterized as differentiated in fodder beet. This differentiation was distributed across the genome and no specific chromosomes appeared to explain the divergence of this crop type. In the chard crop type, 5.16% (433898 bp) of loci were characterized as differentiated. Chromosomes 2, 5, and 8 appear to be the most differentiated and contained 1.19%, 0.69%, and 0.75%, of the total differentiation respectively. Differentiated sites appeared restricted to specific regions along these chromosomes. Many independent datapoints (e.g., sites supported by independent reads) reflect both the quantity and magnitude of these signals. Further characterization of differentiated SNP loci as genic, exonic, intronic, or flanking sequence did not appear variable with respect to crop type or chromosome (Table 2-S1).

Table 2-2: **Differentiated regions (F_{ST}) crop type.**

Chromosome	Number (bp) F_{ST} > 0.6	Percent SNP Differentiated	Percent genic (SNP)	Percent exonic (SNP)	Percent SNP within 1000bp of gene	Percent SNP within 500bp of gene
Sugar	103,903	0.01	0.33	0.06	0.16	0.07
Table	436,106	0.05	0.31	0.06	0.13	0.06
Fodder	47,006	0.01	0.38	0.07	0.12	0.06
Chard	433,898	0.05	0.33	0.07	0.13	0.07
***B. vulgaris* Total**	**1,020,913**	**0.12**	**0.34**	**0.06**	**0.13**	**0.06**

Table 2-3: **Diverged SNP loci with respect to crop type and chromosome.**

Crop type	Chromosome	Number (bp) F_{ST} > 0.6	Percent SNP Differentiated	Percent genic (SNP)	Percent exonic (SNP)	Percent SNP within 1000bp of gene	Percent SNP within 500bp of gene
Sugar	Chr1	7881	0.09	0.27	0.04	0.14	0.08
	Chr2	7357	0.09	0.36	0.08	0.15	0.06
	Chr3	42004	0.50	0.24	0.05	0.10	0.05
	Chr4	2049	0.02	0.26	0.03	0.24	0.08
	Chr5	5094	0.06	0.34	0.04	0.19	0.07
	Chr6	11604	0.14	0.44	0.07	0.15	0.08
	Chr7	2639	0.03	0.35	0.09	0.12	0.06
	Chr8	18492	0.22	0.29	0.05	0.13	0.06
	Chr9	6783	0.08	0.46	0.06	0.17	0.08
	Mean	11545	0.14	0.33	0.06	0.16	0.07
Table	Chr1	61654	0.73	0.28	0.06	0.12	0.06
	Chr2	36564	0.43	0.34	0.09	0.14	0.08
	Chr3	52342	0.62	0.32	0.07	0.15	0.07
	Chr4	27374	0.33	0.31	0.05	0.12	0.05
	Chr5	53529	0.64	0.28	0.07	0.14	0.07
	Chr6	70558	0.84	0.26	0.04	0.09	0.04
	Chr7	25793	0.31	0.32	0.07	0.15	0.08
	Chr8	88582	1.05	0.30	0.05	0.12	0.05
	Chr9	19710	0.23	0.36	0.06	0.11	0.06
	Mean	48456	0.58	0.31	0.06	0.13	0.06
Fodder	Chr1	5929	0.07	0.51	0.09	0.09	0.04
	Chr2	5740	0.07	0.34	0.04	0.14	0.06
	Chr3	7209	0.09	0.51	0.07	0.12	0.06
	Chr4	2173	0.03	0.31	0.14	0.17	0.11
	Chr5	5574	0.07	0.33	0.04	0.09	0.05
	Chr6	6379	0.08	0.27	0.06	0.09	0.07
	Chr7	3737	0.04	0.32	0.07	0.12	0.04
	Chr8	6934	0.08	0.40	0.03	0.06	0.02
	Chr9	3331	0.04	0.42	0.06	0.16	0.06
	Mean	5223	0.06	0.38	0.07	0.12	0.06
Chard	Chr1	29700	0.35	0.34	0.08	0.14	0.07
	Chr2	100148	1.19	0.37	0.07	0.13	0.07
	Chr3	37364	0.44	0.33	0.06	0.11	0.07
	Chr4	53902	0.64	0.29	0.06	0.13	0.07
	Chr5	57733	0.69	0.33	0.08	0.14	0.07
	Chr6	32273	0.38	0.35	0.08	0.13	0.06
	Chr7	29716	0.35	0.27	0.07	0.13	0.06
	Chr8	63351	0.75	0.33	0.05	0.12	0.06
	Chr9	29711	0.35	0.32	0.07	0.14	0.08
	Mean	48211	0.57	0.33	0.07	0.13	0.07

Chromosome 3

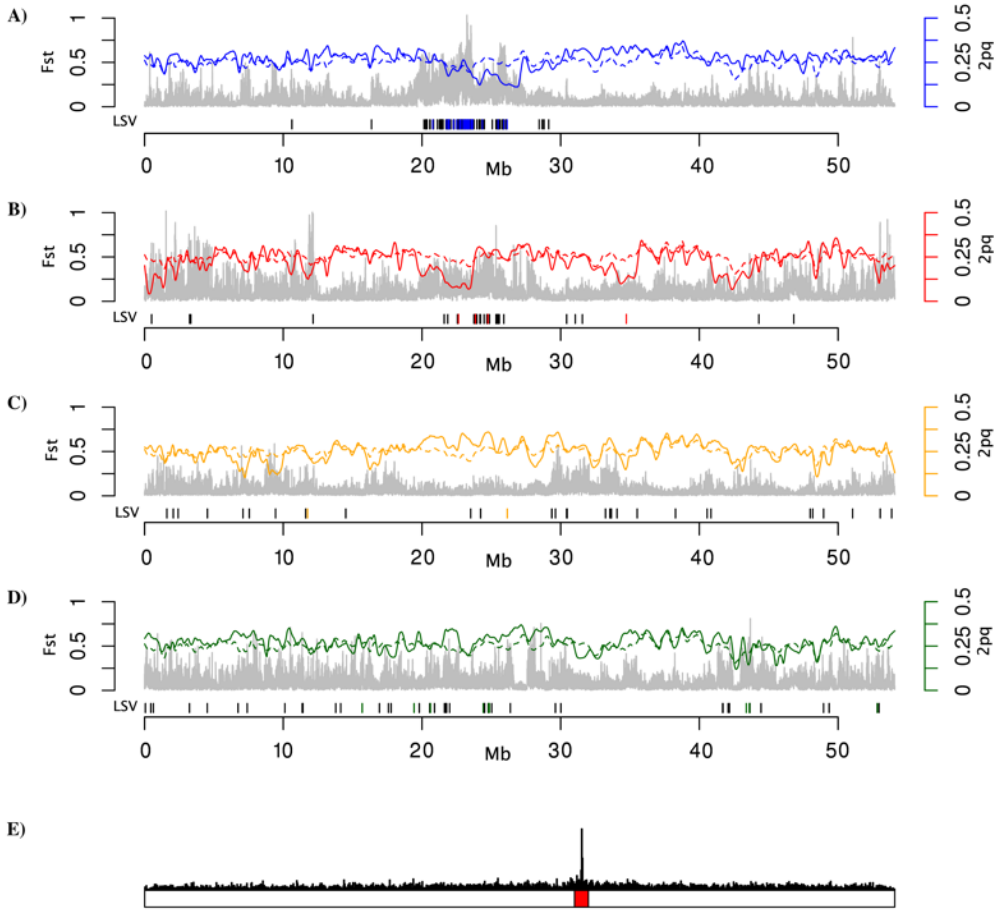

Figure 2-3: **Topology of crop type variation along Chromosome 3.** Expected heterozygosity and F_{ST} plotted across *B. vulgaris* chromosomes. (A) Sugar beet, (B) table beet, (C) fodder beet, (D) chard/leaf beet. Solid colored lines represent *2pq* for crop types. Dashed lines represent average *2pq* for all populations representing cultivated *B. vulgaris*. Gray background represents the F_{ST} statistic. Below each plot is the crop type specific variation; indels (color) and SNP (black). (E) Putative centromere indicated by gypsy element density along chromosome (red).

Table 2-4: **Differentiated regions (F_{ST}) by chromosomes.**

Chromosome	Number (bp) F_{ST} > 0.6	Percent SNP Differentiated	Percent genic (SNP)	Percent exonic (SNP)	Percent SNP within 1000bp of gene	Percent SNP within 500bp of gene
Chr1	26291	0.003	0.35	0.07	0.13	0.06
Chr2	37452	0.004	0.35	0.07	0.14	0.07
Chr3	34730	0.004	0.35	0.06	0.12	0.06
Chr4	21375	0.003	0.30	0.07	0.17	0.08
Chr5	30483	0.004	0.32	0.06	0.14	0.07
Chr6	30204	0.004	0.33	0.06	0.11	0.06
Chr7	15471	0.002	0.31	0.07	0.13	0.06
Chr8	44340	0.005	0.33	0.05	0.11	0.05
Chr9	14884	0.002	0.39	0.06	0.15	0.07

Lineage-specific variation (crop type)

Genome-wide SNP and indel variation was evaluated for lineage-specific variation (LSV). In total, 10,661 variants were detected as crop type specific (e.g., distribution restricted to a single crop type). Of these, 8,098 were SNPs and 1,963 indels. The number of SNP LSV detected within sugar beet, table beet, fodder beet, and chard were as follows: 3,317, 1,379, 643, and 3,359, respectively. Indel LSV detected for the crop types were 342, 558, 205, and 858, respectively. The significance of the quantity and distribution of lineage-specific variation within each crop type was described in more detail in Chapter 1. Interestingly, a high correlation (R^2 = 0.85) between crop type LSV and differentiated regions ($F_{ST} > 0.6$) was found (Figure 2-3 and Figure 2-S1 through Figure 2-S8). This high correlation suggests the accumulation of variation in specific chromosome regions was important for crop type diversification and divergence on the basis of end use.

F_{ST} outliers and associated genes

In total, 472 genes (1.6%) of the 24,255 genes predicted within the EL10.1 reference genome had a significant SNP ($F_{ST} > 0.6$) associated with them. The association was defined as a significant SNP located within the gene boundary or within 1000 bp of flanking sequence. Sixteen genes were discovered in sugar beet, 283 genes in table beet, 2 genes in fodder beet, and 171 genes in chard. Annotations for these genes provided an interesting perspective regarding the putative function of these genes and the processes they are involved with. Of the genes identified as F_{ST} outliers ($F_{ST} > 0.6$), 116 contained experimental evidence in Arabidopsis. One gene was characterized as an ortholog of ATCOL2 BBX3 CONSTANS-LIKE 2 B-box domain protein 3 (EL10Ac2g04397) and was evaluated with respect to bolting in beet (Chia et al. 2008). The most

significant genes for each crop type are reported (Table 2-5) and the complete list is present in Table 2-S1.

Table 2-5: **Significant genes based on F_{ST} outliers.**

Crop Type	Chr	Start	Stop	Gene ID	Max Fst	Mean Fst	N SNP	Annotation
Chard	Chr2	1124105	1130461	EL10Ac2g02466	0.98	0.59	209	Monogalactosyldiacylglycerol synthase, chloroplastic
Chard	Chr2	1132286	1139044	EL10Ac2g02467	0.94	0.48	197	Auxin-binding protein ABP
Chard	Chr2	36903129	36908364	EL10Ac2g03693	0.94	0.47	176	Protein AIG2
Chard	Chr2	48405004	48411761	EL10Ac2g04361	0.89	0.74	144	hypothetical protein
Chard	Chr2	48426379	48444840	EL10Ac2g04365	0.95	0.79	74	Structural maintenance of chromosomes protein 5
Chard	Chr2	48445630	48450656	EL10Ac2g04366	0.94	0.58	88	50S ribosomal protein L
Chard	Chr2	48456005	48458989	EL10Ac2g04368	0.90	0.44	61	ADP-ribosylation factor
Chard	Chr2	48460958	48467377	EL10Ac2g04369	0.90	0.69	86	F-box/WD-40 repeat-containing protein
Chard	Chr5	52292141	52294929	EL10Ac5g12586	0.90	0.52	77	hypothetical protein
Chard	Chr8	55179554	55187589	EL10Ac8g20440	0.89	0.50	228	hypothetical protein
Fodder	Chr2	6525742	6547542	EL10Ac2g02806	0.67	0.26	114	Probable tRNA N6-adenosine threonylcarbamoyltransferase, mitochondrial
Fodder	Chr2	6584270	6585540	EL10Ac2g02808	0.65	0.41	67	Two-component response regulator ARR9
Sugar	Chr1	17999804	18002243	EL10Ac1g01251	0.71	0.44	56	Probable trehalose-phosphate phosphatase D
Sugar	Chr1	18082596	18098518	EL10Ac1g01252	0.76	0.30	256	Endoplasmic reticulum-Golgi intermediate compartment protein 3
Sugar	Chr2	50160084	50163080	EL10Ac2g04512	0.87	0.62	92	Pentatricopeptide repeat-containing protein, mitochondrial
Sugar	Chr2	50164439	50167338	EL10Ac2g04513	0.87	0.67	83	cAMP-regulated phosphoprotein/endosulfine conserved region
Sugar	Chr3	23241971	23242579	EL10Ac3g06337	0.87	0.52	94	hypothetical protein
Sugar	Chr3	23266082	23284333	EL10Ac3g06338	0.86	0.50	218	hypothetical protein
Sugar	Chr3	23313137	23313525	EL10Ac3g06339	0.75	0.66	51	gag-polypeptide of LTR copia-type
Sugar	Chr3	23317099	23333823	EL10Ac3g06340	0.79	0.56	395	DUF2
Sugar	Chr3	23317814	23326286	EL10Ac3g06341	0.79	0.61	215	hypothetical protein
Sugar	Chr3	23419906	23432678	EL10Ac3g06342	0.77	0.46	269	DUF2
Sugar	Chr3	23494631	23513691	EL10Ac3g06343	0.76	0.43	296	hypothetical protein
Sugar	Chr3	23527425	23528852	EL10Ac3g06344	0.86	0.74	97	hypothetical protein
Sugar	Chr3	51060282	51063512	EL10Ac3g07284	0.74	0.41	101	Pentatricopeptide repeat-containing protein
Sugar	Chr4	2887833	2899041	EL10Ac4g07734	0.71	0.28	415	hypothetical protein
Sugar	Chr5	4400661	4403470	EL10Ac5g10742	0.63	0.40	89	Dof zinc finger protein DOF5
Sugar	Chr8	14505353	14510538	EL10Ac8g19192	0.84	0.37	148	Putative transcription factor bHLH04
Table	Chr1	4631423	4639952	EL10Ac1g00390	0.90	0.46	251	Protein of unknown function (DUF3522)
Table	Chr1	5742359	5753296	EL10Ac1g00472	0.85	0.53	251	Transcription factor DIVARICATA
Table	Chr2	8096936	8100260	EL10Ac2g02886	0.89	0.57	127	Cytokinin dehydrogenase 6
Table	Chr2	8163438	8169350	EL10Ac2g02888	0.91	0.75	259	hypothetical protein
Table	Chr3	11878635	11890018	EL10Ac3g05841	0.87	0.65	273	E3 ubiquitin protein ligase RIN2
Table	Chr3	53555895	53561372	EL10Ac3g07455	0.86	0.45	247	Werner Syndrome-like exonuclease
Table	Chr6	17874246	17879128	EL10Ac6g13977	0.88	0.66	176	Geranylgeranyl transferase type-2 subunit alpha
Table	Chr6	18352959	18367495	EL10Ac6g13989	0.87	0.53	363	Reverse transcriptase-like
Table	Chr6	18609565	18622088	EL10Ac6g13995	0.86	0.49	316	Protein NRT
Table	Chr8	1260672	1275448	EL10Ac8g18344	0.87	0.42	549	Cell division cycle protein 27 homolog B
Table	Chr8	46449727	46460636	EL10Ac8g20022	0.86	0.49	353	Serine/threonine-protein kinase PBS

Crop type genes (sugar beet)

Sugar beet genes identified in close proximity to loci with significant F_{ST} values were further investigated for function using gene annotations, experimental evidence in Arabidopsis, and GO terms. The GO categories these genes belong to include: negative regulation of protein dephosphorylation (GO:0035308), phloem or xylem histogenesis (GO:0010087), procambium histogenesis (GO:0010067), response to chitin (GO:0071323), retrograde endoplasmic reticulum to Golgi vesicle mediated transport (GO:2000156), and trehalose biosynthetic processes (GO:0005992). Chromosomes 3, 5, and 8 appear to contain the signal for divergence of sugar beet relative to the other crop types. Chromosome 3 showed a large extended signal of differentiation around 20 Mb to 25 Mb, with the most significant peak centered at 23 Mb (Figure 2-3). Several genes surrounding this region with significant F_{ST} values were annotated as 'domain of unknown function' and 'hypothetical protein'. Several of these predicted genes had no annotation, and two targets were identified as an LTR associated gag-polypeptide (EL10Ac3g06339) and a lncRNA (EL10Ac3g06344) (Table 2-5). The composition and function of this region may partially explain the unique biology and divergence of sugar beet relative to other crop types. Chromosome 8 of sugar beet contained loci with significant F_{ST} values, and the gene associated with this signal was identified as a Myc-type, basic helix-loop-helix (bHLH) domain protein (EL10Ac8g19192). Chromosome 5 also contained loci with significant F_{ST} values associated with a gene coding for a Dof zinc finger protein DOF5.6 (EL10Ac5g10742). Interestingly this gene appears to be a transcription factor involved with procambium histogenesis and differentiation of vascular tissues. Significant loci ($F_{ST} > 0.6$) within glutamate receptor 2.7 (EL10Ac5g12159) suggests genes involved in cellular carbohydrate metabolism may be under selection.

Crop type genes (table beet)

In table beet, 283 genes were associated with significant SNP loci ($F_{ST} > 0.6$), the most of all crop types. The quantity of significant genes and putative functions based on annotations, GO terms, and experimental evidence in Arabidopsis suggest major differences in physiology, metabolism, and development of table beet lineages relative to other crop types. These genes included MADS box genes, homeodomain transcription factors, auxin and cytokinin biosynthesis, hormone perception and signaling, oxidative stress response genes, and genes which code for disease resistance proteins. Sugar and aquaporin genes were also recovered, suggesting differences in physiology and metabolism related to water content and sugar. Other notable results included a large number of genes involved with DNA replication, mitosis, and meiosis. These included chromosome checkpoint regulators, sister chromatid cohesion proteins, mitotic spindle proteins, replication fork arrest, telomere maintenance, and resolution of holiday junctions. These genes are interesting because of their potential effects on gene flow and the transmission of genetic information across generations, as well as cell cycle progression and effects on morphology. The most significant genes for table beet are presented in Table 2-5 and the complete list available in Table 2-S1.

Crop type genes (fodder beet)

Only two genes were associated with significant SNP loci in fodder beet ($F_{ST} > 0.6$). These genes included a probable tRNA N6-adenosine threonylcarbamoyltransferase (EL10Ac2g02806) and a two-component response regulator, ARR9 (EL10Ac2g02808), involved in histidine kinase signaling. The GO terms associated with these proteins include cytokinin response, signal transduction, development, and circadian rhythm. The proximity of these two genes on

Chromosome 2 suggests only one may be important. The low number of genes supporting the divergence of fodder relative to other crop types may reflect the high heterozygosity within fodder populations, small number of representative fodder beet populations (N=2), or the low degree of divergence between sugar and fodder resulting from common ancestry (e.g. high relationship coefficients) (Chapter 1).

Crop type genes (chard)

In chard, 171 genes were identified in close proximity to significant SNP loci ($F_{ST} > 0.6$). Many of these genes were involved in root, shoot, and flower development as well as pathogen response. A notable quantity of genes detected (47.4%) were located on Chromosome 2, suggesting this chromosome was important for the differentiation of chard relative to the other crop types. The distribution of LSV (Figure 2-1) and quantity of shared variation suggest the four chard populations sampled likely represent two distinct subpopulations (Chapter 1). The reduced number of unique, or diverged samples for population genomic comparisons may have affected the ability of this approach to distinguish between divergence resulting from historical selection versus by chance, as a result of the low number of unique samples. The substructure within chard lineages showed two distinct groups but these differences were not accumulated on Chromosome 2. Since divergent subpopulations are less likely to share variation, the lack of divergence on Chromosome 2 between the two chard subpopulations further supports the role of undefined variation located on Chromosome 2 in conditioning economic phenotypes associated with chard (e.g. expanded leaves and petioles). Another observation was that the low number (N = 4) of chard samples used likely had a negative effect on the ability to resolve specific variation on

Chromosome 2 explaining the differentiation between chard and other crop types. These signals warrant further investigation using increased N sizes of the chard crop type.

Selective sweeps

F_{ST} can determine the apportionment of variation between populations. The statistic F_{ST} was useful in detecting historical selection which occurred within a single crop type lineage. The majority of variation was not differentiated with respect to crop type which suggests it is not under selection or it is distributed among crop types and populations as a result of a complex evolutionary history (e.g., common ancestry, admixture and introgression, and the random sorting of ancestral polymorphism). The utility of detecting significant variation using F_{ST} outliers was limited in all but the most obvious cases of selection for unique crop type variation detailed above. Low F_{ST} values could indicate myriad explanations for a lack of divergence but by examining genomic regions devoid of genetic polymorphism (*2pq)* with respect to crop type we found regions indicative of selective sweeps (e.g. low diversity [*2pq*] and low F_{ST} values) within and between crop types and populations. Shared historical selection was not entirely unexpected because of known common ancestry (Chapter 1) between specific lineages. These regions revealed several notable observations. 1) The expression and distribution of color phenotypes within and among crop type populations was complex and although F_{ST} was not significant at color loci, signals of selection (e.g. low [*2pq*]) were observed in table beet and in all beets that express color. 2) Fodder and sugar crop types share regions of low diversity shared between these crop types suggest historical selection for important phenotypes may have occurred within common ancestors of these lineages. 3) The root types (e.g., sugar beet, fodder beet, table beet) shared several regions of low genetic diversity relative to leaf types. This is

consistent with genetic variation with the potential to influence root enlargement and supports previously unknown events in the demographic history of these lineages.

The genes coding for the key enzymes involved in the biosynthesis of betalain pigments (e.g. betacyanin [rev/violet] and betaxanthin [orange/yellow]) have been cloned and functionally evaluated in beet (Halsted et al. 2012 and Halsted et al. 2015). This provided an opportunity to evaluate the utility of population genetic measures (e.g., allele frequency, *2pq,* and F_{ST}) to understand patterns of variation within the genome by looking closer at targets of historical selection such as the Y locus (EL10Ac2g04466.1) and the R locus (EL10Ac2g04268.1). The R locus, located at 49 Mb on Chromosome 2, showed low genetic diversity indicative of intense historical selection and specific patterns (e.g. fixation for alternate alleles) restricted to table beet lineages (Figure 2-4). A closer look at the Y locus, located at 47.3 Mb along Chromosome 2 (Figure 2-S12), codes for the yellow color, showed a high degree of fixation for the alternate 'non sugar beet' allele. The reduction of heterozygosity within the gene as well as in regions flanking the coding region is consistent with selection for populations that express color in the root. Furthermore, there were obvious patterns of variation present in the promoter sequence of the Y locus. The expression of color among beet crop types provides an interesting example of variation that appears to result from a selective sweep within a lineage (e.g., table beet) but provides little significance through F_{ST} as a result of this variation being shared among crop types and populations which express color.

Fodder and sugar beet crop types exhibited less divergence than the table beet or chard crop types. N size for fodder populations was limited but nonetheless close relationships between

sugar and fodder beet suggests common ancestry may be one explanation for the lack of divergence observed for these crop types. Within the genomes of these lineages, specific chromosome regions lacked significant F_{ST} and exhibited low diversity (*2pq)* relative to genome wide data. A region on chromosome 8 (13.5 Mb) was one such region (Figure 2-S7) and underlying this region was the transcription factor, radix-brevis like (EL10Ac8g19137). Experimental evidence in Arabidopsis suggests this gene regulates root and shoot growth by modulating auxin signaling and controls quantitative aspects of root growth in Arabidopsis (Mouchel et al. 2004). The distribution of this variation within sugar and fodder beet indicates the potential for a genetic mechanism controlling components of root shape and root elongation shared between sugar and fodder lineages. Chromosome 9 contained a large region (34.5 Mb–38 Mb) with similar characteristics (e.g., lacked significant F_{ST} and exhibited low diversity) in sugar beet . This region was indicative of a selective sweep but due may have a complex distribution between crop types and was not detected using our estimate of F_{ST}. On chromosome 9 (37 Mb), 6-phosphofructo-2-kinase (EL10Ac9g22391) was identified as a potential candidate due to another potential selective sweep and its putative role in cellular carbohydrate metabolism.

The root types of *B. vulgaris* shared three undifferentiated regions exhibiting low diversity (*2pq*) that correspond to major differences between genomes of root types (e.g., sugar, fodder, table) versus leaf types (e.g. chard). These regions included Chromosome 2 (26 Mb–27 Mb), Chromosome 4 (42 Mb–43 Mb), and Chromosome 8 (14 Mb–15 Mb) (Figure 2-S2, Figure 2-S3, Figure 2-S7). Several candidate genes were identified within these regions on the basis of gene diversity (*2pq*) and local allele frequencies which supported these candidates as potential targets of selection. These genes include Cytokinin dehydrogenase 3 (EL10Ac8g19202), NAM/NAC

(EL10Ac2g02976), RPD1 (EL10Ac4g09126), and Homeodomain transcription factor (EL10Ac4g09093) (Figure 2-S9, Figure 2-S10, and Figure 2-S11). Functional evidence in Arabidopsis agreed with their potential functions in beet and may explain the unique biology of beet roots (e.g., root enlargement and biomass accumulation).

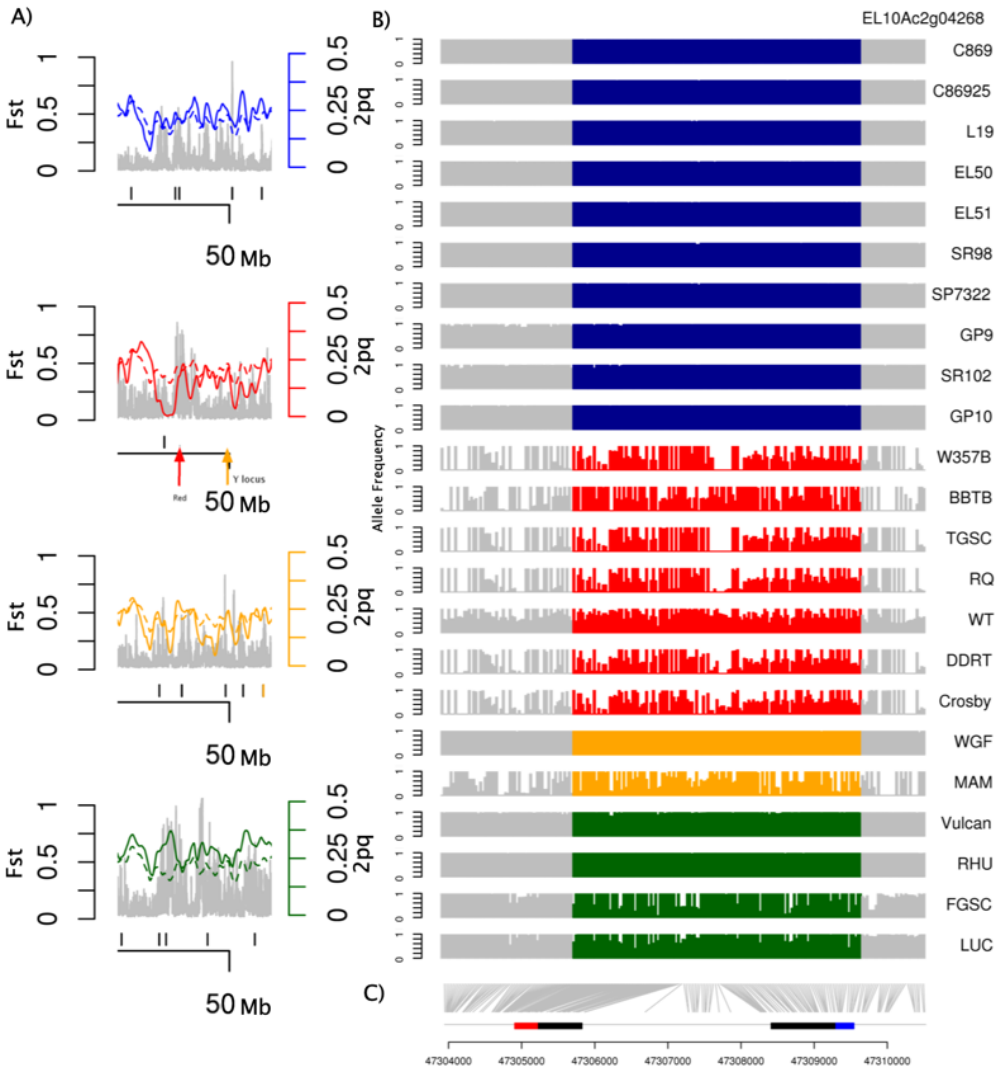

Figure 2-4: **Allele frequency data for R locus (EL10Ac2g04268)**. (A) F_{ST} and *2pq* plot of chromosome region containing gene of interest. (B) Allele frequency plots range from 0 to 1. Color indicates crop type (blue = sugar beet, red = table beet, orange = fodder beet, green = chard). Color also indicates the variation within gene boundaries; gray variation represents 1000 bp flanking the gene. (C) Physical position of each variant relative to the gene model. Blue and red color represent the start and stop sequence. Black represents the exons.

DISCUSSION

Genomic variation distributed within and among beet crop types correlates with the unique biology and important phenotypes contained within these lineages. Previously unknown features were identified within the genomes of diverse beet populations and showed the utility of estimating population genetic parameters (e.g., lineage-specific variation [LSV], diversity [*2pq*], and differentiation [F_{ST}]) for understanding phenotypic divergence of these linages. Genome differentiation in beet likely results from selection, drift, and mating closely related individuals. This process acts to sort and fix ancestral polymorphism within discrete lineages while increasing the frequency of beneficial alleles conferring desired phenotypes. The total genome differentiation detected in the cultivated species with respect to crop type was 12.13%. sugar 1.23%, table 5.18%, fodder 0.56%, chard 5.16%. These results are similar to what has been reported previously in incipient speciation literature (e.g., 5% ~ 10% of the genome) (Nosil et al. 2009). Estimating genome differentiation and substructure is subjective and influenced by the choice of estimators, thresholds for determining differentiation, and representative populations sampled. Our estimate of differentiation tested the degree of divergence between a single crop type relative to all other crop types using F_{ST}. In this way we detected important crop type variation and generated additional lines of inquiry based on empirical observations. This included the presence of selective sweeps, bottlenecks, and admixture across the genome. When selective sweeps were unique to a single crop type, F_{ST} was informative. In cases where selective sweeps appear shared between crop types, F_{ST} was limited and likely impacted by close relationships, common ancestry and introgression between lineages. This was highlighted in the low proportion of differentiated SNP loci across both sugar beet (1.23%) and fodder beet

(0.56%) genomes. Signals pertaining to these shared regions were present in the allele frequency data. The reduction of diversity of genomic regions, measured by (*2pq*), suggest these regions were important for the development and diversification of specific crop type lineages. Admixture and gene flow between populations negatively affects the ability to resolve population structure (differentiation) and suggests prior knowledge of the demographic history, historical selection and admixture would benefit these analyses by allowing more informed comparisons and better estimation of selective sweeps, population bottlenecks, and founder effects. Knowledge of these features is lacking in beet and this study provides a high-density dataset capable of discovering and characterizing these regions and the extent of these features within the genome. Negative correlations between traits as a result of population history and linkage disequilibrium within the genome can have unintended consequences on selection efficiency within a species (Slatkin 2008). In turn this can affect the rate of genetic gain in crop improvement. Negative linkages between yield and sucrose concentration in sugar beet have been reported and may be a limiting factor in increasing sucrose on a per hectare basis (Boesmark 2006).

To date, only a handful of genes have been functionally evaluated in beet. These include several genes related to bolting, BvBTC1 (Pin et al. 2010) and two CONSTANSE-LIKE genes (Dally et al. 2018). Since the populations represented within this research are biennial these genes were not investigated as a means to validate the approach used here. The betalain biosynthesis (color) genes (Hatlestad et al. 2012; Hatlestad et al., 2015) were more suited to validation and benchmarking the utility of the population genetic measures to describe the allelic variation and test the degree to which this variation explains the distribution of color within and among crop type lineages. Color in beet ranges from yellow to orange and violet to red. Yellow pigments

produced first and are converted to red. Red beets possess functional gene which codes for enzyme. The pathway originates from the tyrosine pathway (WISC pub). (BIOCHEMICAL MECHANISM) The red locus (EL10Ac2g04268), annotated as Geraniol 8-hydroxylase, was not significant using our F_{ST} estimator. However, due to the lack of diversity (*2pq*) in the region surrounding the gene, appeared highly selected within beet crop types, specifically within table beet. Much of this variation appeared to be consistent with historical breeding and color as a target trait for improvement. Additionally, the Y locus (EL10Ac2g04466) identified as a transcription factor MYB114 showed similar patterns of variation in all beet populations that expressed color. Fixation of specific variation unique to beet lineages which produce color pigments appeared in the upstream promoter region of the Y locus, suggesting transcription factor binding might be important for the up-regulation of this gene and the expression of color pigments. The expression of color within diverse tissue types suggests this pathway has a great deal of complexity in its regulation. The two table beets that exhibit intense color, BBTB and TGTB, lacked diversity relative to other table beets, suggesting additional genes are involved and intense selection may have been required to achieve such pronounced phenotypes.

The genes associated with significant F_{ST} values suggest a large degree of differentiation in physiology, morphology, and metabolism between crop types. The number of genes recovered for each crop type was influenced by the number of populations per crop type, relationships between crop types, and choice of F_{ST} estimator (Bhatia et al. 2013). The average size of a differentiated region was small (1,400 bp). This size suggests a high marker density may be needed in beet. Presumably, the size of differentiated regions can be used to infer time and intensity of selection as well as rates of recombination within the genome. This was evident

along Chromosome 3 of sugar beet, where an extensive region of differentiation appears to result from linkage. This potentially reflects both the time and intensity of selection in this region. To date, beet research has lacked high density marker data to resolve regions of agronomic importance. A recent study leveraged pooled data for a segregating population and identified casual variation associated with hypocotyl color of sugar beet (Ries et al. 2016). The combination of pooled data and WGS proved informative to this end. Segregating populations are quite useful in beet. RIL populations are one example of this owing to the linkage generated across few generations and limited recombination. QTL studies have resulted in the identification of large chromosome regions influencing important trait variation in beet (CITATIONS). Until recently, the size of these regions, lack of reference genome sequence and the identity of genes within these regions has made the selection of candidate genes for functional analysis difficult. The recent publication of several beet genomes has provided physical location and content of genes within the sugar beet genome. Together, molecular maps from QTL studies and physical maps have provided important insight into our understanding of trait heritability and trait performance across years and environments.

Common ancestry between root types was not evident in relationship coefficients and clustering based on genome-wide markers (Chapter 1). This suggests the evolution of the expanded root character results from either convergence or is shared via introgression. Regions with low diversity (*2pq*) were evident within root lineages, which indicate a selective sweep. The identity of the genes underlying these regions suggest potential functional roles in root enlargement. The regions on Chromosomes 2, 4, and 8 lacked diversity (*2pq*) in root types and appeared unselected in chard. Root morphology of chard is similar to the wild progenitor of beet, *B. vulgaris* spp.

maritima. The most probable candidates were identified on the basis of allele frequency and diversity (*2pq*) within these regions. On Chromosome 4 an ortholog of root primordium defective 1 (RPD1) was identified. Functional experiments using *rpd1* mutants showed RPD1 is part of a unique gene family in plants and required for adventitious/lateral root development (Konishi and Munetaka 2006). Interestingly, *rpd1* did not affect the development of root primordium or the initiation of cell division required for lateral root formation. Local allele frequency for this gene was consistent with expectations of a candidate gene having undergone a selective sweep for root enlargement. Chromosome 2 contained a gene coding for a no apical meristem NAC domain protein (NAM/NAC). These proteins are involved in hormone regulation and influence meristem function with large effects on the development of tissues and organs (Willemsen et al. 2008). Experimental evidence in Arabidopsis showed NAM/NAC proteins interact with scarecrow (SCR) and short root (SHR), two genes involved in root development and patterning of tissues within the root. Interactions between auxin and cytokinin, specifically antagonisms between them, have been demonstrated for proper root development and the maintenance of specific cell types (Chapman and Estelle 2010). On Chromosome 8, another region indicative of a sweep within root types was identified. A promising candidate was identified as cytokinin dehydrogenase 3. The role of cytokinin in root development is well recognized and has been postulated as being involved in the enlargement of beet roots (Smigocki and Owens 1988, 1989).

This research produced a list of genes underlying the differences in root development between crop types. Several candidates appear to be good targets for further functional validation and research into developmental genetic networks underlying root development, including several

related to hormone biosynthesis, perception, and signaling. The number of regions with low diversity corresponding to potential sweeps for root enlargement suggests genetic variation within multiple genes may be required for expression of this phenotype. Furthermore, the absence of an enlarged root within wild populations, suggests root enlargement occurring spontaneously through mutation is a low probability event. This might suggest variation in many genes is required for the expression of this trait or it is selected against in wild populations. This observation is of importance because root enlargement was likely paramount to the development of beet lineages that contain the agronomic potential to accumulate large quantities of sucrose but independent of physiological changes that are required to realize that potential.

The mechanism underlying sucrose accumulation is likely the same for all beet crop types (Goldman and Navazio 1996). Differences in the ability of beet varieties to accumulate sucrose has been proposed to result from relationships between water and dry matter (sucrose) within roots (Carter 1987 and Bergen 1967). Sucrose accumulation and water content are negatively correlated in most instances. Given the relationship between water and dry matter, selection for high sucrose (e.g., sugar beet) could have resulted from selection on water use or water use efficiency genes. The development of beet roots shows a transition between juvenile and adult stages (Trebbi and McGrath 2009), which corresponds to physiological changes (Milford 1973, Wyse 1979). Gene expression differences were also evident across this transition, suggesting different genetic pathways underlie these physiological changes in water content, sucrose content, and relative abundance of storage tissues (Trebbi and McGrath 2009).

Chromosomes 3, 5 and 8 appear to contain signal for sugar beet domestication. Understanding the basis for sugar accumulation has been a major focus of sugar beet research (e.g., genetics, local adaptation, management practices). The significant region on Chromosome 3 contained many hypothetical protein predictions, domains of unknown function as well as an LTR - gag polypeptide. This may indicate that transposon/repeat-based sequence evolution may have had a large effect on the unique biology of sugar beet. The silencing of transposable elements is demonstrated to have consequences on gene expression of neighboring genes and thus potentially major consequences on phenotype (Sigman and Slotkin 2015). The diversity of this region was also a surprise, and in reality, the region was identified as significant owing to the absence of variation within all other crop types. The nature of this region and close proximity to centromere could mean significantly lower recombination rates and may help explain the strong negative correlation between sucrose content and root yield. This correlation exists in sugar beet but is not present in wide hybrids (McGrath unpublished).

Previous research reported extensive linkage disequilibrium along Chromosome 3 (Adetunji et al. 2014). This was attributed to introgression and selection of the disease resistance loci Rz1, which codes for rhizomania resistance. The sugar beet populations sampled in this research represent germplasm developed before the widespread utilization of Rhizomania resistance and suggests this signal represents the differentiation and divergence between fodder and sugar lineages. Explicitly identifying the genetic basis of selection for sugar beet from fodder may aid in the understanding of the physiological differences observed between these lineages, specifically in regards to biomass and sucrose accumulation. Chromosome 8 (13 – 15 Mb) contained low diversity (2pq) and high divergence (F_{ST}) across multiple crop types. The location

of this region within the gene rich, euchromatic arm of Chromosome 8 and the quantity and distribution of signals within this region may reflect a high degree of recombination. This suggests this region may possess a greater ability to respond to selection and may have been significant to the development of beet crop types.

Mapping studies have identified several regions in close proximity to genomic locations we identified as likely targets for physiological differences in sugar beet lineages. A genome wide association study (Würschum et al. 2011) and a recent QTL study (Wang et al. 2019) identified significant regions related to sucrose accumulation on Chromosome 9. Direct comparisons of regions discovered between studies are challenging due to lack of published markers as well as differences between molecular maps and reference genomes used. This study identified 6-phosphofructo-2-kinase (EL10Ac9g22391), on Chromosome 9, as a potential candidate for the altered carbohydrate metabolism exhibited across beet crop types.

Purging genetic variation through selection appears important in the development of stable phenotypes within a lineage and may reflect the number of genes involved in producing a variety with a given trait. The fact that these traits appear to be under selection but were not significant in our analysis highlights the limitations of F_{ST} to detect important variation due to a complex evolutionary history of the species and the diversification of beet crop types. Even with these limitations hundreds of genes were recovered which were previously unknown in conditioning the underlying phenotypic differences between beet crop types. One advantage of F_{ST} was that phenotypic data was not required but can be utilized in order to gain perspective on the phenotypic divergence between populations and crop types. The complex relationships and

degree to which variation is shared across beet lineages may be approachable using pairwise F_{ST} for each population and may be one way to tease out significant variation that is shared. Aside from F_{ST} outliers and the most diverged regions, low F_{ST} values support a hypothesis of panmixia and greater probability for geneflow between populations at these loci which result in no divergence. Highly selected sites showing low F_{ST} values are good targets for investigating admixture and gene flow between populations and likely explain how genomic variation is shared between crop types and identify the important variation associated with phenotypes corresponding to these events.

APPENDIX

Chromosome 1

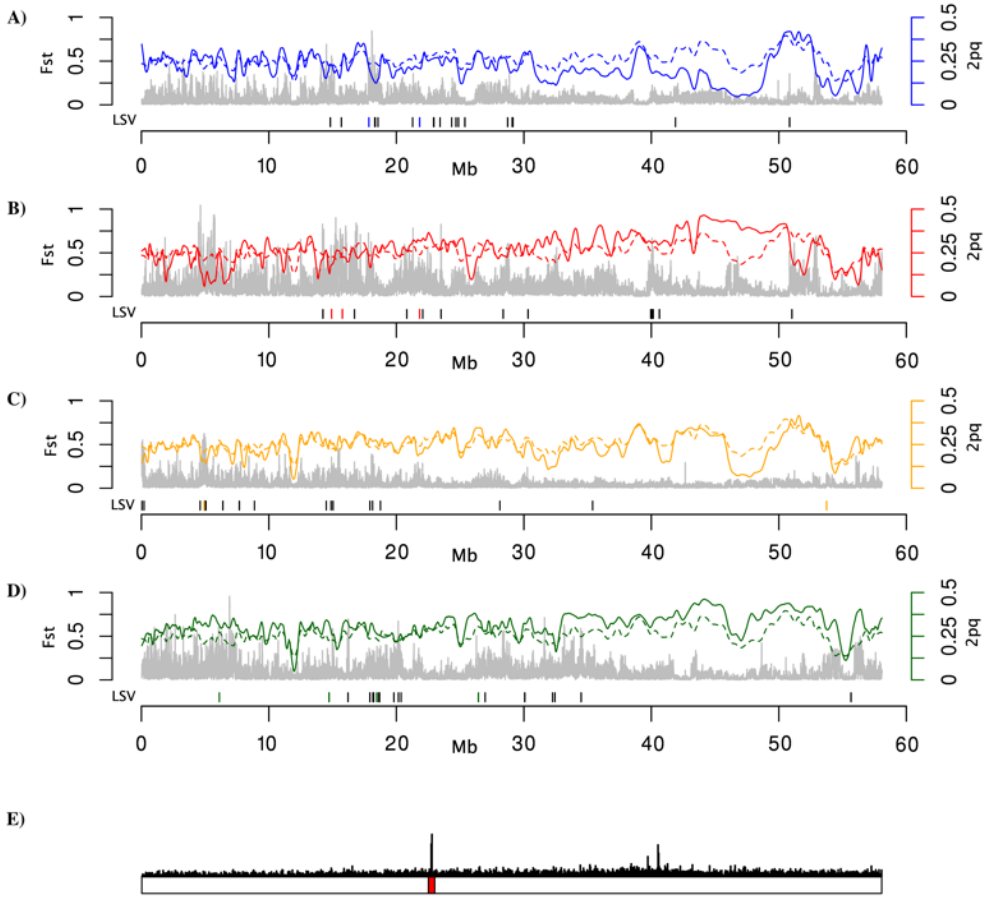

Figure 2-S1: **Topology of crop type variation along Chromosome 1.** Expected heterozygosity and F_{ST} plotted across *B. vulgaris* chromosomes. (A) Sugar beet, (B) table beet, (C) fodder beet, (D) chard/leaf beet. Solid colored lines represent *2pq* for crop types. Dashed lines represent average *2pq* for all populations representing cultivated *B. vulgaris*. Gray background represents the F_{ST} statistic. Below each plot is the crop type specific variation; indels (color) and SNP (black). (E) Putative centromere indicated by gypsy element density along chromosome (red).

Chromosome 2

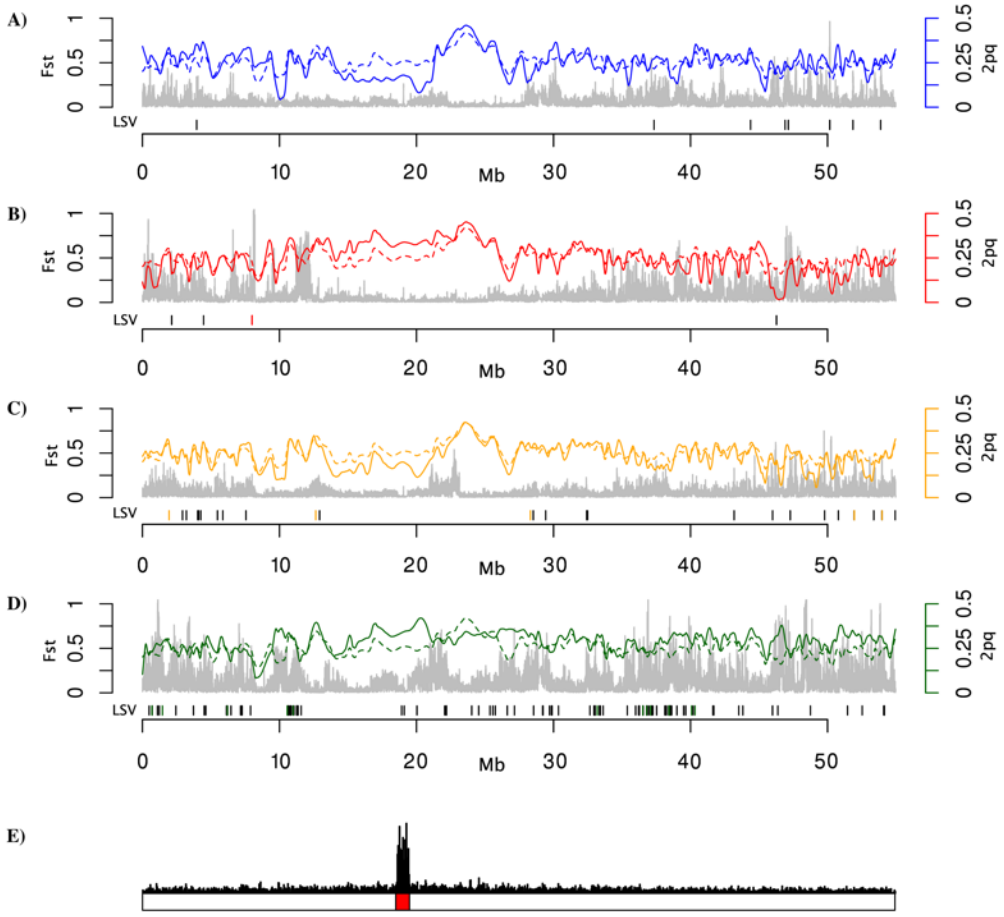

Figure 2-S2: **Topology of crop type variation along Chromosome 2.** Expected heterozygosity and F_{ST} plotted across *B. vulgaris* chromosomes. (A) Sugar beet, (B) table beet, (C) fodder beet, (D) chard/leaf beet. Solid colored lines represent *2pq* for crop types. Dashed lines represent average *2pq* for all populations representing cultivated *B. vulgaris*. Gray background represents the F_{ST} statistic. Below each plot is the crop type specific variation; indels (color) and SNP (black). (E) Putative centromere indicated by gypsy element density along chromosome (red).

Chromosome 4

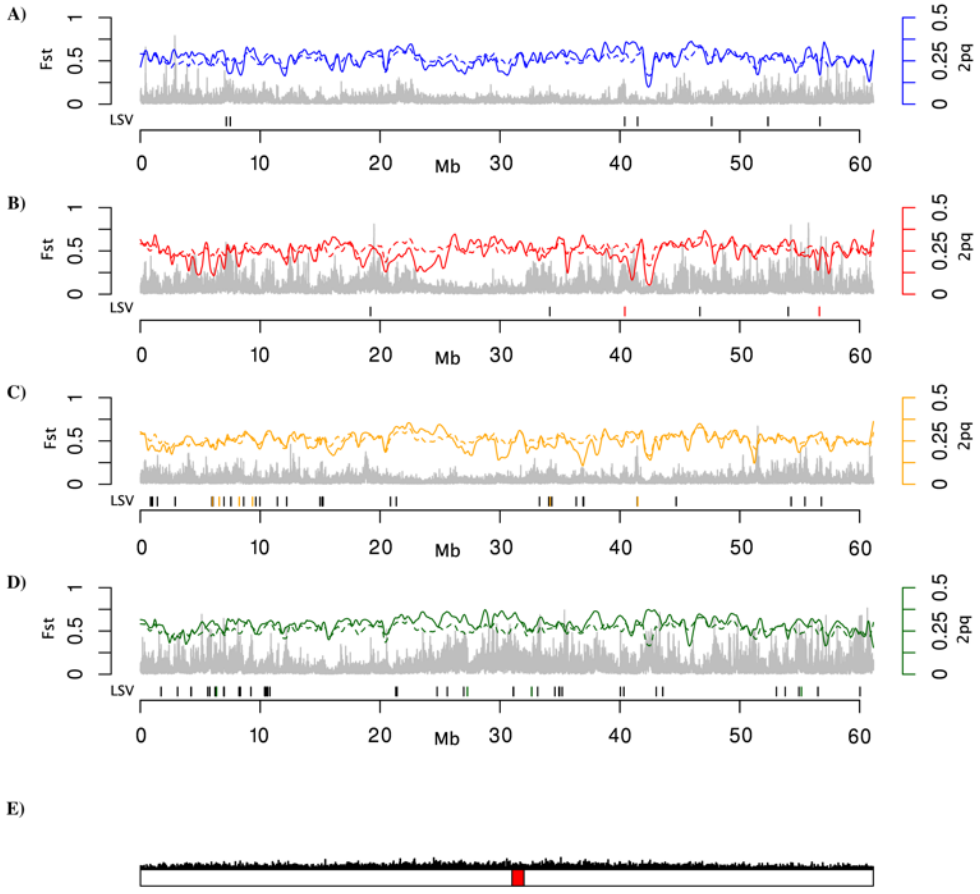

Figure 2-S3: Topology of crop type variation along Chromosome 4. Expected heterozygosity and F_{ST} plotted across *B. vulgaris* chromosomes. (A) Sugar beet, (B) table beet, (C) fodder beet, (D) chard/leaf beet. Solid colored lines represent *2pq* for crop types. Dashed lines represent average *2pq* for all populations representing cultivated *B. vulgaris*. Gray background represents the F_{ST} statistic. Below each plot is the crop type specific variation; indels (color) and SNP (black). (E) Putative centromere indicated by gypsy element density along chromosome (red).

Chromosome 5

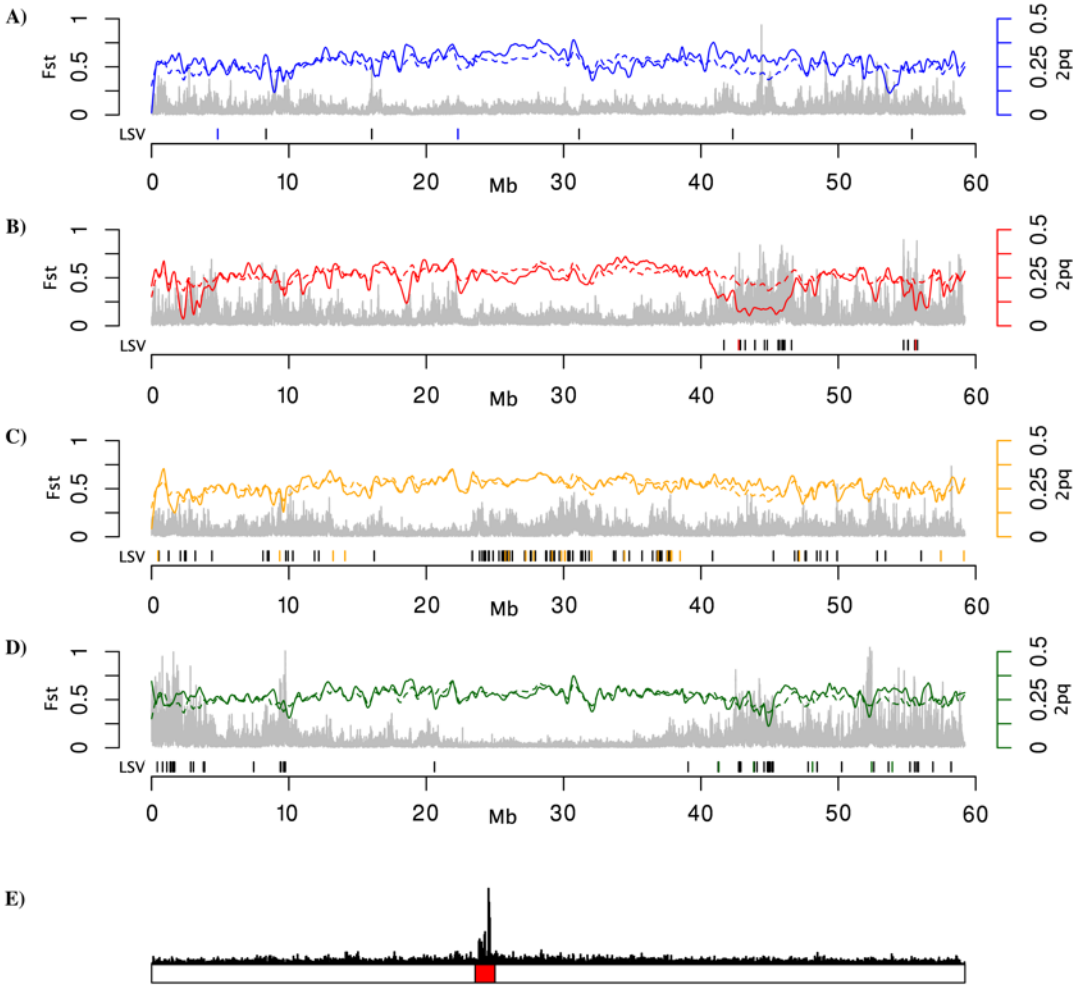

Figure 2-S4: **Topology of crop type variation along Chromosome 5.** Expected heterozygosity and F_{ST} plotted across *B. vulgaris* chromosomes. (A) Sugar beet, (B) table beet, (C) fodder beet, (D) chard/leaf beet. Solid colored lines represent *2pq* for crop types. Dashed lines represent average *2pq* for all populations representing cultivated *B. vulgaris*. Gray background represents the F_{ST} statistic. Below each plot is the crop type specific variation; indels (color) and SNP (black). (E) Putative centromere indicated by gypsy element density along chromosome (red).

Chromosome 6

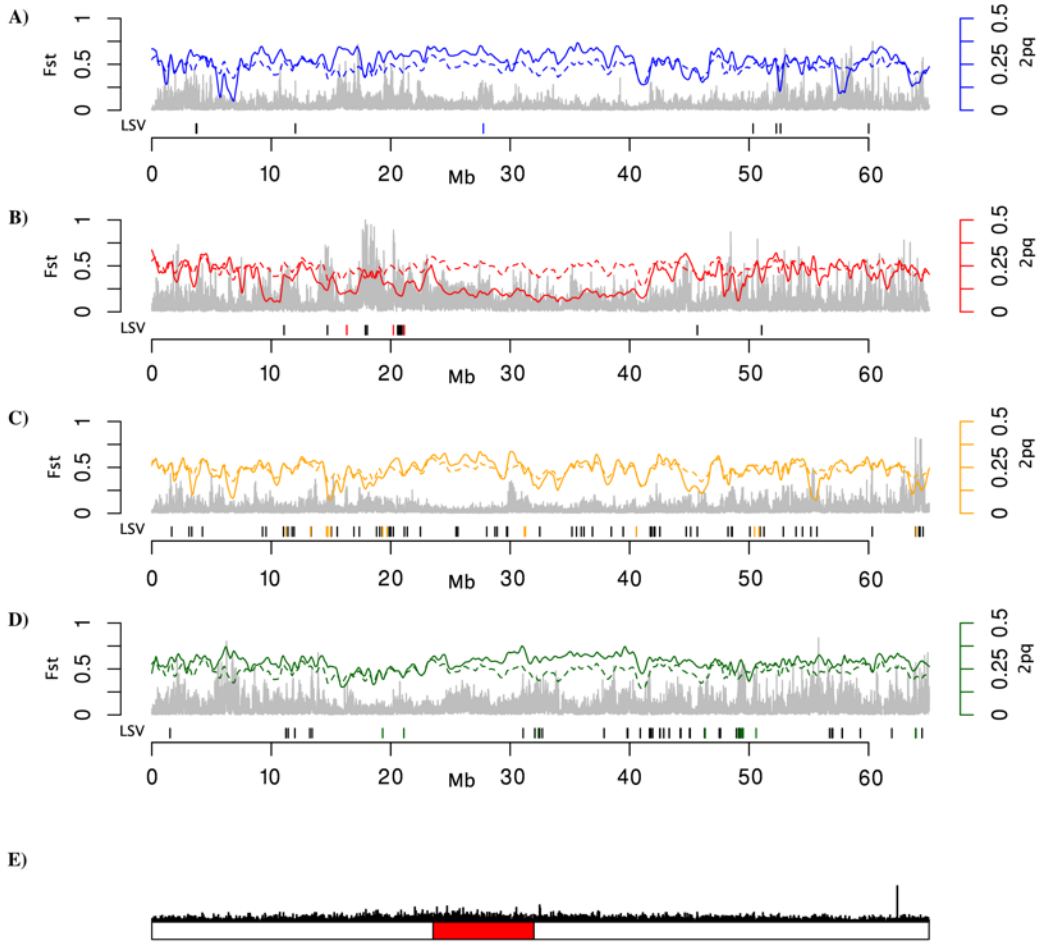

Figure 2-S5: **Topology of crop type variation along Chromosome 6.** Expected heterozygosity and F_{ST} plotted across *B. vulgaris* chromosomes. (A) Sugar beet, (B) table beet, (C) fodder beet, (D) chard/leaf beet. Solid colored lines represent *2pq* for crop types. Dashed lines represent average *2pq* for all populations representing cultivated *B. vulgaris*. Gray background represents the F_{ST} statistic. Below each plot is the crop type specific variation; indels (color) and SNP (black). (E) Putative centromere indicated by gypsy element density along chromosome (red).

Chromosome 7

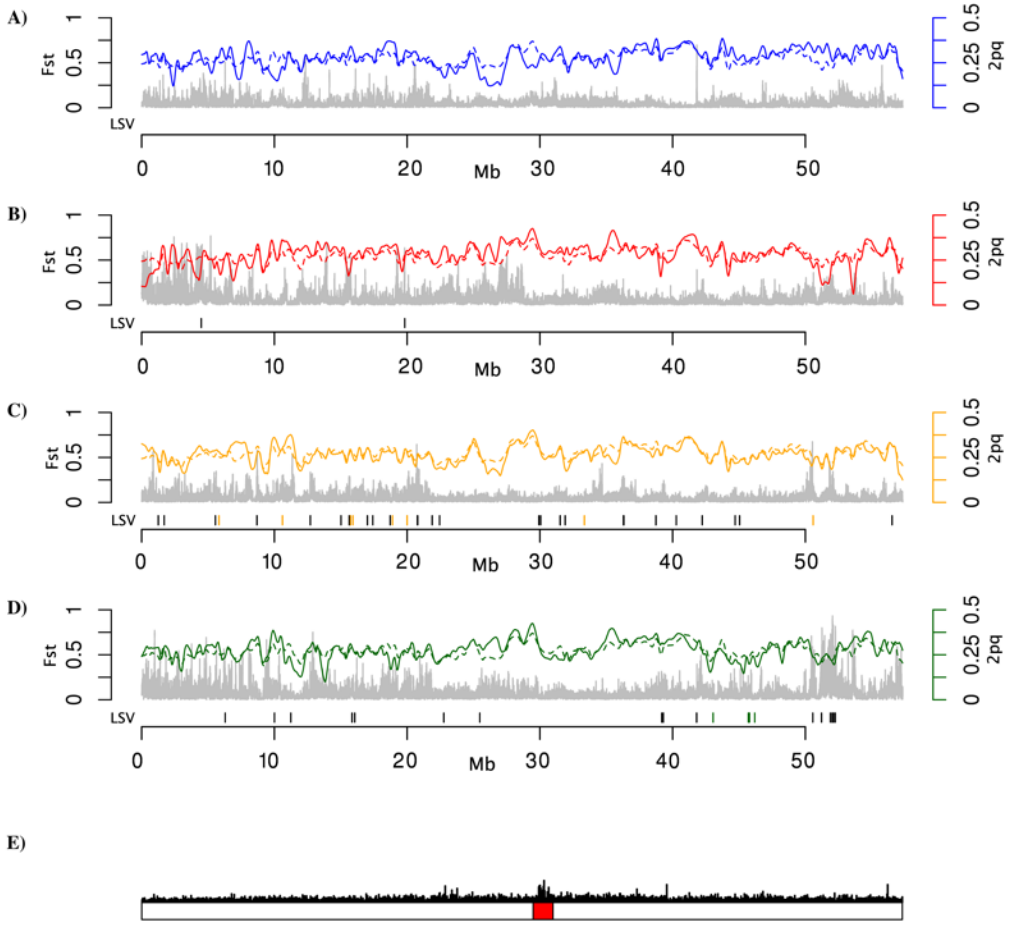

Figure 2-S6: **Topology of crop type variation along Chromosome 7.** Expected heterozygosity and F_{ST} plotted across *B. vulgaris* chromosomes. (A) Sugar beet, (B) table beet, (C) fodder beet, (D) chard/leaf beet. Solid colored lines represent *2pq* for crop types. Dashed lines represent average *2pq* for all populations representing cultivated *B. vulgaris*. Gray background represents the F_{ST} statistic. Below each plot is the crop type specific variation; indels (color) and SNP (black). (E) Putative centromere indicated by gypsy element density along chromosome (red).

Chromosome 8

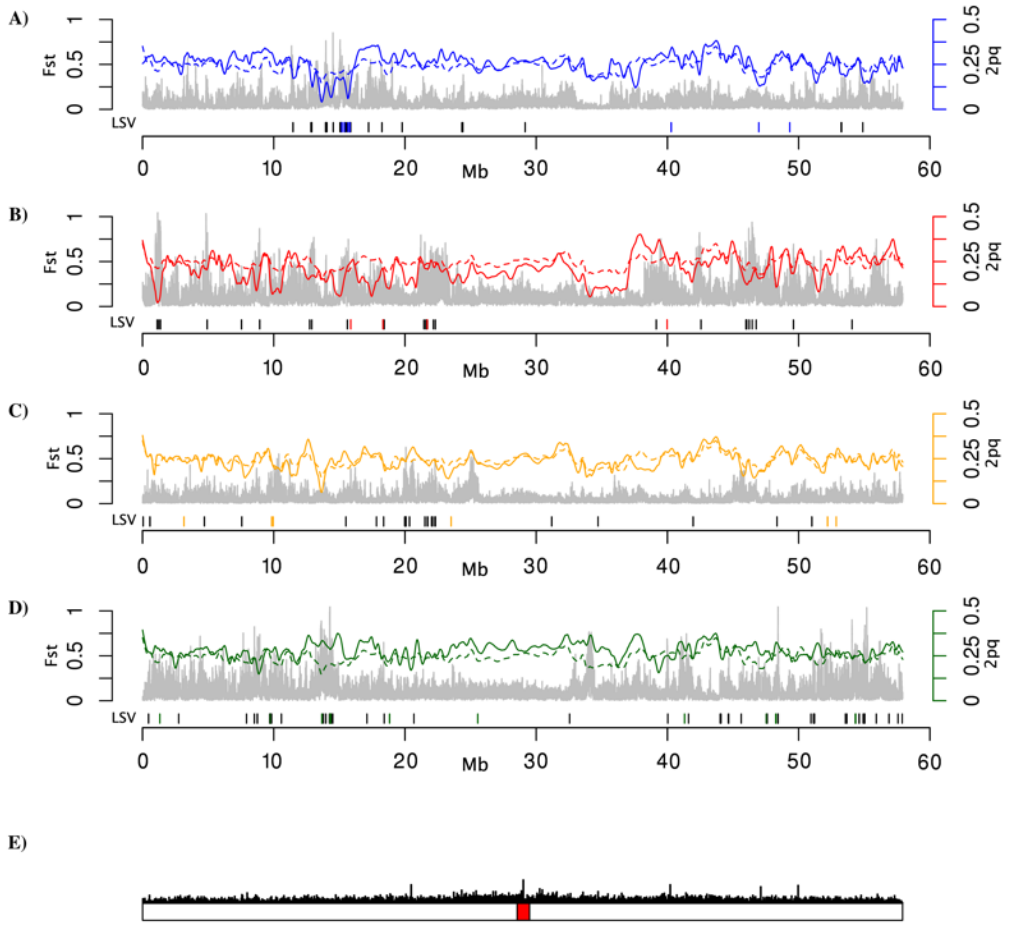

Figure 2-S7: **Topology of crop type variation along Chromosome 8.** Expected heterozygosity and F_{ST} plotted across *B. vulgaris* chromosomes. (A) Sugar beet, (B) table beet, (C) fodder beet, (D) chard/leaf beet. Solid colored lines represent *2pq* for crop types. Dashed lines represent average *2pq* for all populations representing cultivated *B. vulgaris*. Gray background represents the F_{ST} statistic. Below each plot is the crop type specific variation; indels (color) and SNP (black). (E) Putative centromere indicated by gypsy element density along chromosome (red).

Chromosome 9

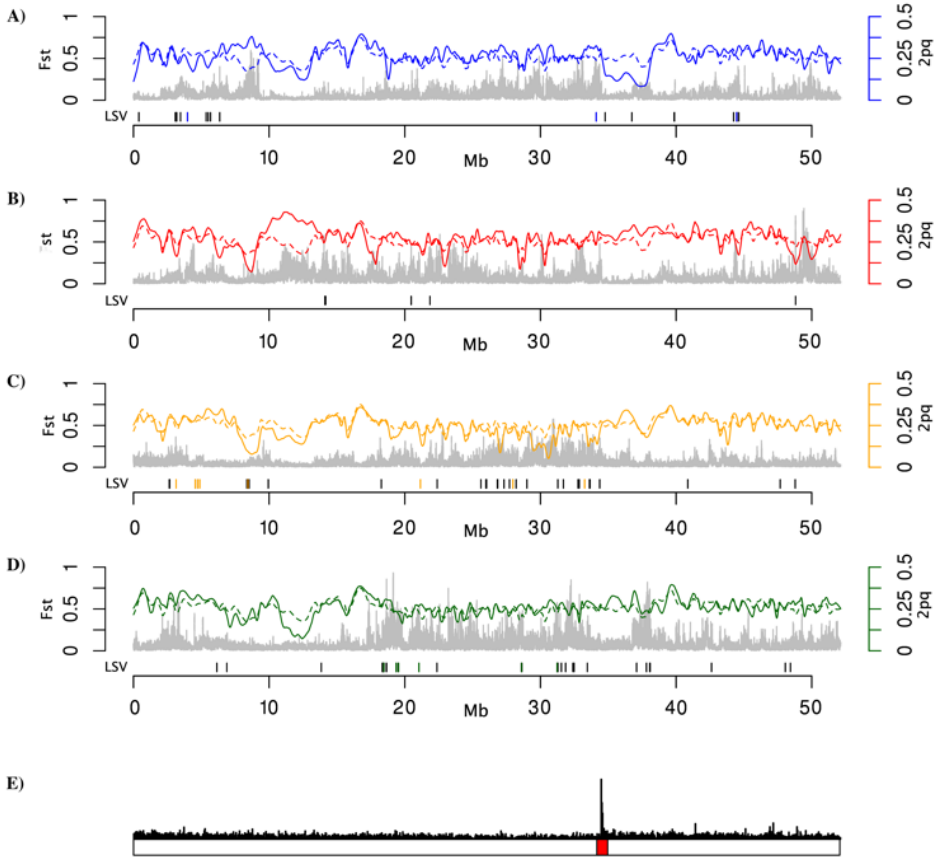

Figure 2-S8: **Topology of crop type variation along Chromosome 9.** Expected heterozygosity and F_{ST} plotted across *B. vulgaris* chromosomes. (A) Sugar beet, (B) table beet, (C) fodder beet, (D) chard/leaf beet. Solid colored lines represent *2pq* for crop types. Dashed lines represent average *2pq* for all populations representing cultivated *B. vulgaris*. Gray background represents the F_{ST} statistic. Below each plot is the crop type specific variation; indels (color) and SNP (black). (E) Putative centromere indicated by gypsy element density along chromosome (red).

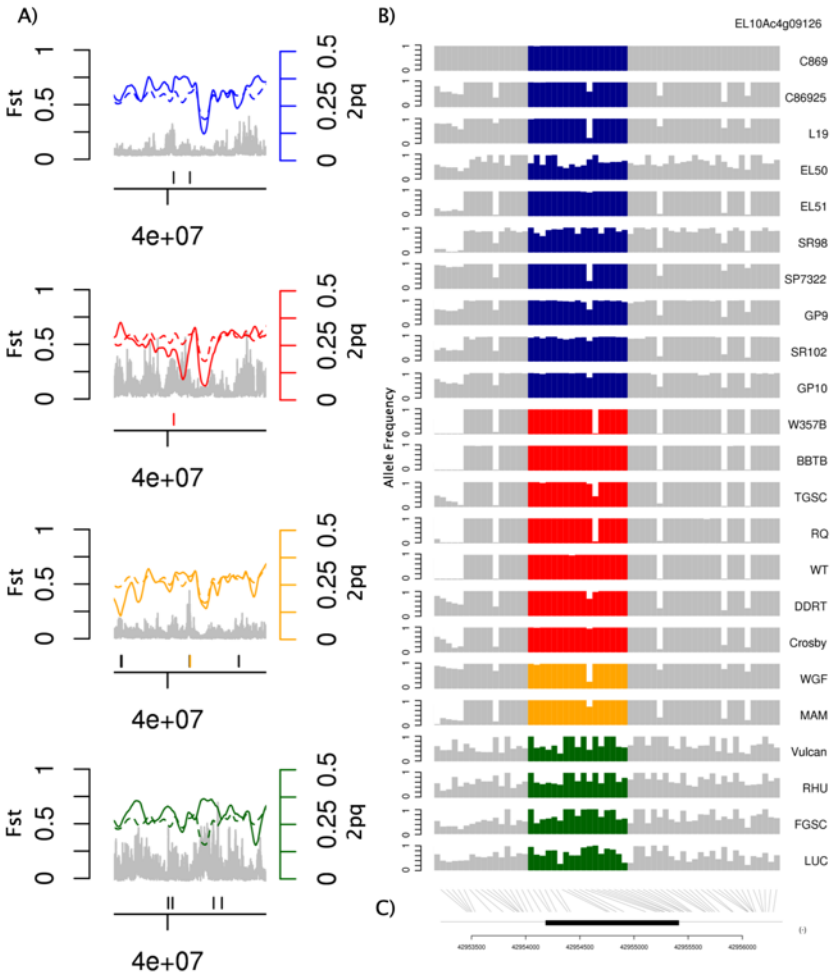

Figure 2-S9: **Allele frequency data for Root Primordium Defective 1, *RPD1*, (EL10Ac4g09126).** (A) F_{ST} and *2pq* plot of chromosome region containing gene of interest. (B) Allele frequency plots range from 0 to 1. Color indicates crop type (blue = sugar beet, red = table beet, orange = fodder beet, green = chard). Color also indicates the variation within gene boundaries; gray variation represents 1000 bp flanking the gene. (C) Physical position of each variant relative to the gene model. Blue and red color represent the start and stop sequence. Black represents the exons.

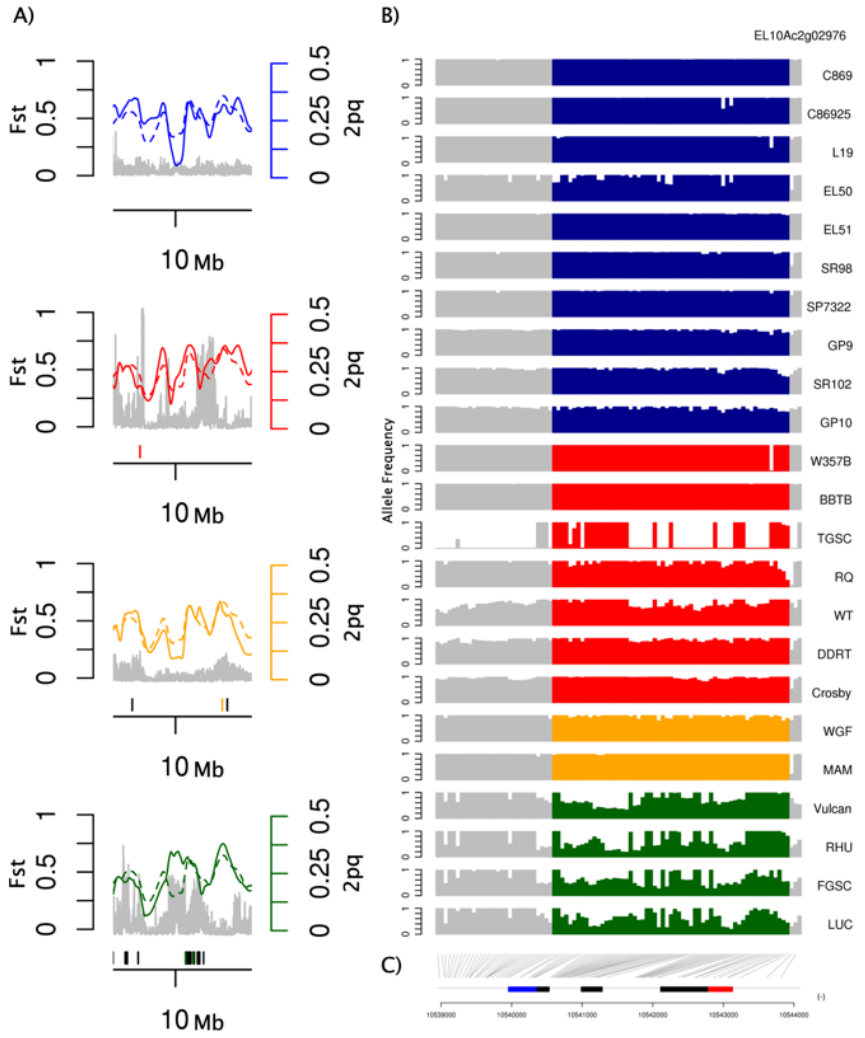

Figure 2-S10: **Allele frequency data for NAM/NAC (EL10Ac2g02976).** (A) F_{ST} and *2pq* plot of chromosome region containing gene of interest. (B) Allele frequency plots range from 0 to 1. Color indicates crop type (blue = sugar beet, red = table beet, orange = fodder beet, green = chard). Color also indicates the variation within gene boundaries; gray variation represents 1000 bp flanking the gene. (C) Physical position of each variant relative to the gene model. Blue and red color represent the start and stop sequence. Black represents the exons.

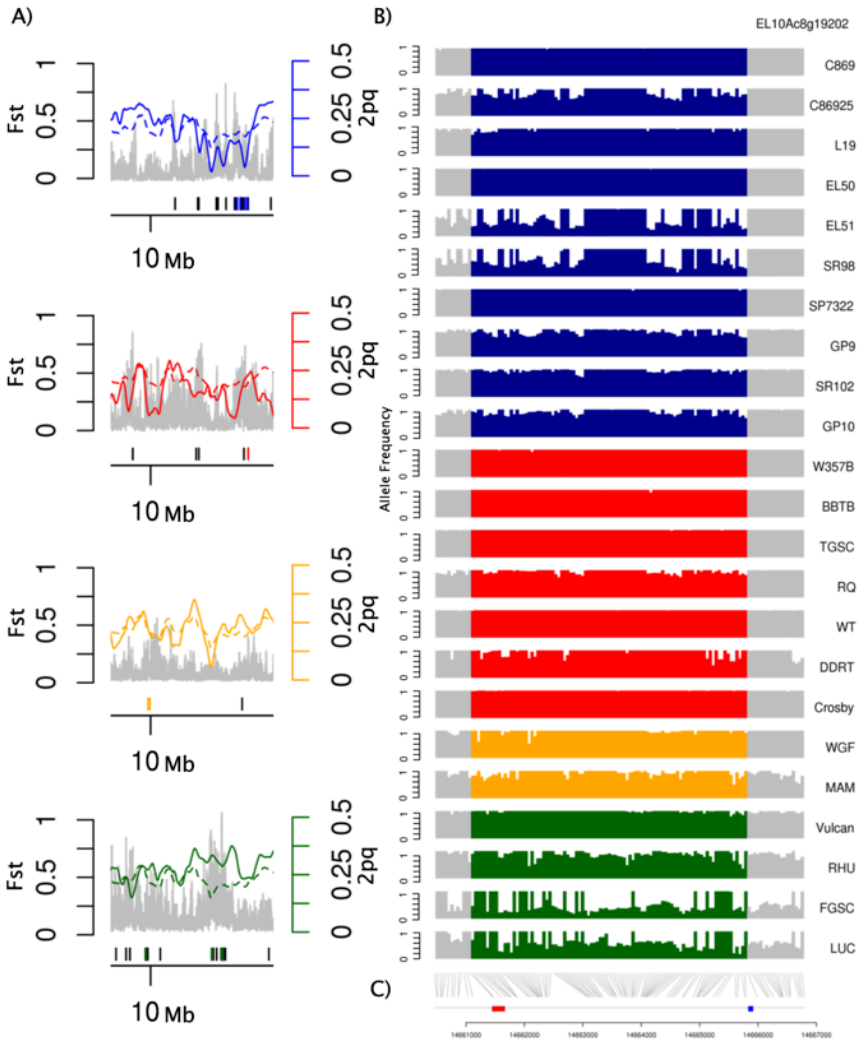

Figure 2-S11: **Allele frequency data for Cytokinin dehydrogenase 1 (EL10Ac2g02976).** (A) F_{ST} and *2pq* plot of chromosome region containing gene of interest. (B) Allele frequency plots range from 0 to 1. Color indicates crop type (blue = sugar beet, red = table beet, orange = fodder beet, green = chard). Color also indicates the variation within gene boundaries; gray variation represents 1000 bp flanking the gene. (C) Physical position of each variant relative to the gene model. Blue and red color represent the start and stop sequence. Black represents the exons.

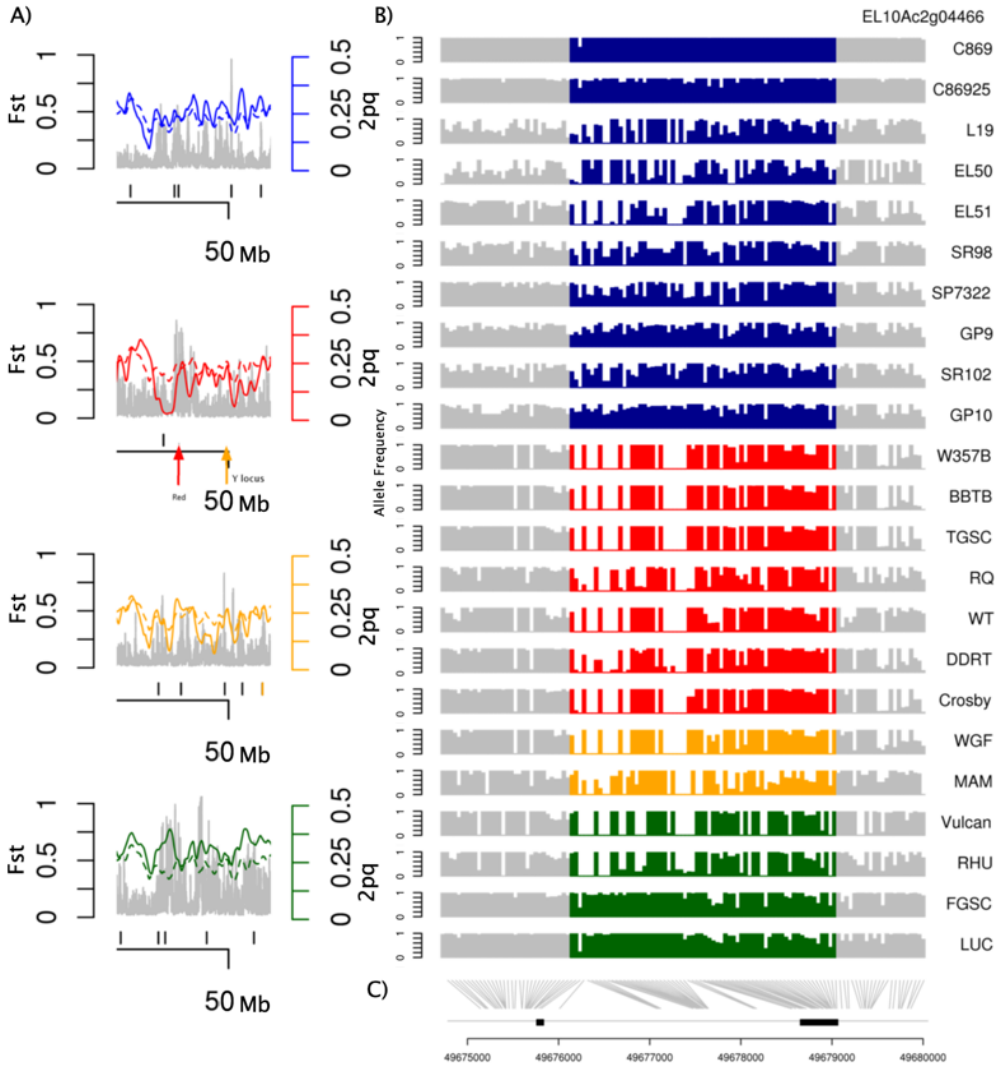

Figure 2-S12. **Allele frequency data for the Y locus (EL10Ac2g04466).** (A) F_{ST} and *2pq* plot of chromosome region containing gene of interest. (B) Allele frequency plots range from 0 to 1. Color indicates crop type (blue = sugar beet, red = table beet, orange = fodder beet, green = chard). Color also indicates the variation within gene boundaries; gray variation represents 1000 bp flanking the gene. (C) Physical position of each variant relative to the gene model. Blue and red color represent the start and stop sequence. Black represents the exons.

Table 2-S1 **Genes with significant F_{ST} values (F_{ST} > 0.6).**

Crop Type	Chr	Start	Stop	Length	Gene ID	Max Fst within gene	Mean Fst withiin gene	Number of varients	Annotation
Chard	Chr2	1103409	1107878	4469	EL10Ac2g02464	0.88	0.36	132	Sirohydrochlorin ferrochelatase
Chard	Chr2	1111507	1120352	8845	EL10Ac2g02465	0.81	0.55	172	hypothetical protein
Chard	Chr2	1124105	1130461	6356	EL10Ac2g02466	0.98	0.59	209	Monogalactosyldiacylglycerol synthase, chloroplastic
Chard	Chr2	1132286	1139044	6758	EL10Ac2g02467	0.94	0.48	197	Auxin-binding protein ABP
Chard	Chr2	1132548	1132757	209	EL10Ac2g02468	0.83	0.64	49	Auxin-binding protein ABP
Chard	Chr2	1172421	1185869	13448	EL10Ac2g02469	0.85	0.51	263	Auxin-binding protein ABP
Chard	Chr2	1179655	1181225	1570	EL10Ac2g02470	0.83	0.67	51	Auxin-binding protein ABP
Chard	Chr2	1209990	1219099	9109	EL10Ac2g02472	0.84	0.39	133	Auxin-binding protein ABP
Chard	Chr2	3334136	3341792	7656	EL10Ac2g02616	0.79	0.43	90	Protein NRT
Chard	Chr2	3344051	3345821	1770	EL10Ac2g02617	0.73	0.32	95	hypothetical protein
Chard	Chr2	3349132	3350821	1689	EL10Ac2g02618	0.72	0.47	56	hypothetical protein
Chard	Chr2	3352175	3357233	5058	EL10Ac2g02619	0.73	0.36	249	WD repeat-containing protein 6
Chard	Chr2	3366358	3367068	710	EL10Ac2g02620	0.72	0.29	38	Probable sugar phosphate/phosphate translocator
Chard	Chr2	3376356	3383824	7468	EL10Ac2g02621	0.76	0.25	151	Alpha-galactosidase
Chard	Chr2	3378935	3397281	18346	EL10Ac2g02622	0.82	0.48	282	hypothetical protein
Chard	Chr2	3407225	3416783	9558	EL10Ac2g02623	0.78	0.46	263	tRNA (guanine(26)-N(2))-dimethyltransferase
Chard	Chr2	3418367	3421554	3187	EL10Ac2g02624	0.73	0.52	190	40S ribosomal protein S26-2
Chard	Chr2	3428455	3432574	4119	EL10Ac2g02625	0.75	0.50	137	Superoxide dismutase [Mn], mitochondrial
Chard	Chr2	3435755	3445765	10010	EL10Ac2g02626	0.84	0.39	390	Uncharacterized membrane protein At
Chard	Chr2	36841601	36846652	5051	EL10Ac2g03686	0.83	0.42	198	Putative glutathione-specific gamma-glutamylcyclotransferase 2
Chard	Chr2	36853067	36861026	7959	EL10Ac2g03687	0.85	0.70	328	Proteasome subunit beta type-6
Chard	Chr2	36886439	36888938	2499	EL10Ac2g03688	0.71	0.65	10	Putative AC transposase
Chard	Chr2	36891688	36894110	2422	EL10Ac2g03689	0.77	0.56	128	F-box/kelch-repeat protein
Chard	Chr2	36894717	36896298	1581	EL10Ac2g03690	0.73	0.54	64	hypothetical protein
Chard	Chr2	36898871	36900529	1658	EL10Ac2g03691	0.88	0.51	79	F-box/kelch-repeat protein
Chard	Chr2	36903129	36908364	5235	EL10Ac2g03693	0.94	0.47	176	Protein AIG2
Chard	Chr2	36909913	36911650	1737	EL10Ac2g03694	0.83	0.49	80	GDSL esterase/lipase At
Chard	Chr2	39898033	39901860	3827	EL10Ac2g03828	0.77	0.32	234	Domain of unknown function (DUF35)
Chard	Chr2	39905220	39918200	12980	EL10Ac2g03829	0.71	0.53	367	Probable magnesium transporter NIPA9
Chard	Chr2	39925315	39929008	3693	EL10Ac2g03830	0.70	0.43	165	Cytokinin riboside 5'-monophosphate phosphoribohydrolase LOG8
Chard	Chr2	39960403	39965306	4903	EL10Ac2g03831	0.77	0.26	183	Protein of unknown function (DUF86)
Chard	Chr2	39977374	39980750	3376	EL10Ac2g03832	0.71	0.39	170	Cytochrome P450 7
Chard	Chr2	39980421	39981826	1405	EL10Ac2g03833	0.71	0.44	103	Cytochrome P450 7
Chard	Chr2	46177056	46182651	5595	EL10Ac2g04181	0.82	0.56	221	Cysteine--tRNA ligase
Chard	Chr2	46938751	46939986	1235	EL10Ac2g04234	0.86	0.55	82	Core-2/I-Branching enzyme
Chard	Chr2	46941574	46947249	5675	EL10Ac2g04235	0.87	0.49	82	RNA-dependent RNA polymerase 6
Chard	Chr2	48300766	48307491	6725	EL10Ac2g04350	0.87	0.70	186	Pentatricopeptide repeat-containing protein
Chard	Chr2	48306857	48312755	5898	EL10Ac2g04351	0.87	0.68	74	Putative disease resistance protein RGA3
Chard	Chr2	48316224	48318590	2366	EL10Ac2g04352	0.86	0.60	85	hypothetical protein
Chard	Chr2	48319662	48319902	240	EL10Ac2g04353	0.86	0.69	47	hypothetical protein
Chard	Chr2	48376778	48379857	3079	EL10Ac2g04357	0.88	0.70	71	Notchless protein homolog
Chard	Chr2	48380227	48383787	3560	EL10Ac2g04358	0.88	0.80	77	Putative disease resistance protein RGA4
Chard	Chr2	48387218	48392993	5775	EL10Ac2g04359	0.85	0.71	89	Ankyrin repeat, PH and SEC7 domain containing protein secG
Chard	Chr2	48397725	48402781	5056	EL10Ac2g04360	0.88	0.74	77	Uncharacterized protein family, UPF0
Chard	Chr2	48405004	48411761	6757	EL10Ac2g04361	0.89	0.74	144	hypothetical protein
Chard	Chr2	48405925	48407670	1745	EL10Ac2g04362	0.88	0.76	59	Pentatricopeptide repeat-containing protein
Chard	Chr2	48413276	48416707	3431	EL10Ac2g04363	0.87	0.69	79	Probable mitochondrial chaperone bcs
Chard	Chr2	48419937	48420828	891	EL10Ac2g04364	0.87	0.79	28	Structural maintenance of chromosomes protein 5
Chard	Chr2	48426379	48444840	18461	EL10Ac2g04365	0.95	0.79	74	Structural maintenance of chromosomes protein 5
Chard	Chr2	48445630	48450656	5026	EL10Ac2g04366	0.94	0.58	88	50S ribosomal protein L
Chard	Chr2	48451959	48455260	3301	EL10Ac2g04367	0.82	0.56	102	Domain of unknown function (DUF34)
Chard	Chr2	48456005	48458989	2984	EL10Ac2g04368	0.90	0.44	61	ADP-ribosylation factor
Chard	Chr2	48460958	48467377	6419	EL10Ac2g04369	0.90	0.69	86	F-box/WD-40 repeat-containing protein
Chard	Chr2	48469098	48471956	2858	EL10Ac2g04370	0.81	0.33	62	N-alpha-acetyltransferase
Chard	Chr2	48475512	48476089	577	EL10Ac2g04371	0.70	0.57	17	hypothetical protein
Chard	Chr2	48481821	48486219	4398	EL10Ac2g04372	0.77	0.47	66	CTL-like protein DDB_G0274487
Chard	Chr2	48493517	48494642	1125	EL10Ac2g04373	0.74	0.47	57	Protein PLANT CADMIUM RESISTANCE 2
Chard	Chr2	48496483	48499486	3003	EL10Ac2g04374	0.78	0.55	104	hypothetical protein
Chard	Chr2	48512304	48513945	1641	EL10Ac2g04375	0.73	0.54	63	Probable glutamine--fructose-6-phosphate aminotransferase [isomerizing]
Chard	Chr2	48521055	48528596	7541	EL10Ac2g04376	0.74	0.55	168	Glutamine--fructose-6-phosphate aminotransferase [isomerizing] 2
Chard	Chr2	48529203	48540948	11745	EL10Ac2g04377	0.71	0.36	311	Serine carboxypeptidase-like 40
Chard	Chr2	48556600	48562746	6146	EL10Ac2g04380	0.65	0.31	181	Agamous-like MADS-box protein AGL
Chard	Chr2	48569012	48574444	5432	EL10Ac2g04381	0.64	0.38	164	Methyltransferase-like protein
Chard	Chr2	48595215	48605207	9992	EL10Ac2g04383	0.68	0.39	232	Protein of unknown function (DUF760)
Chard	Chr2	48605702	48607379	1677	EL10Ac2g04384	0.72	0.37	131	Xylose isomerase
Chard	Chr2	48651714	48656323	4609	EL10Ac2g04388	0.63	0.44	152	Pheophytinase, chloroplastic
Chard	Chr2	48745996	48754847	8851	EL10Ac2g04393	0.69	0.45	42	F-box/FBD/LRR-repeat protein
Chard	Chr2	48768187	48777813	9626	EL10Ac2g04395	0.64	0.39	84	hypothetical protein
Chard	Chr2	48808223	48811623	3400	EL10Ac2g04397	0.69	0.43	146	Zinc finger protein CONSTANS-LIKE 2
Chard	Chr2	48819176	48824863	5687	EL10Ac2g04398	0.71	0.38	277	APO protein 4, mitochondrial
Chard	Chr2	48858258	48861577	3319	EL10Ac2g04401	0.77	0.49	82	Probable galacturonosyltransferase 9
Chard	Chr2	48865832	48867583	1751	EL10Ac2g04402	0.83	0.76	36	Basic leucine zipper 43
Chard	Chr2	48867317	48869250	1933	EL10Ac2g04403	0.83	0.59	57	Basic leucine zipper 43
Chard	Chr2	53186471	53199696	13225	EL10Ac2g04775	0.86	0.62	235	Probable leucine-rich repeat receptor-like protein kinase
Chard	Chr2	53200336	53203485	3149	EL10Ac2g04776	0.83	0.47	118	LIM domain-containing protein WLIM2b
Chard	Chr2	53817886	53825086	7200	EL10Ac2g04828	0.85	0.42	172	Peptidyl-prolyl cis-trans isomerase CYP20-1
Chard	Chr2	53828886	53834895	6009	EL10Ac2g04829	0.83	0.65	123	Phosphoinositide phospholipase C 6
Chard	Chr2	53835464	53836831	1367	EL10Ac2g04830	0.83	0.69	51	Probable aspartic protease
Chard	Chr2	53840847	53849061	8214	EL10Ac2g04831	0.88	0.68	180	Phosphoinositide phospholipase C 2
Chard	Chr2	53850555	53858744	8189	EL10Ac2g04832	0.82	0.64	267	NF-X
Chard	Chr4	60646595	60650030	3435	EL10Ac4g10352	0.81	0.55	151	GPI mannosyltransferase 2
Chard	Chr5	790952	792974	2022	EL10Ac5g10460	0.87	0.45	83	Photosystem II reaction center W protein, chloroplastic
Chard	Chr5	1119486	1123411	3925	EL10Ac5g10484	0.82	0.67	115	2-methyl-6-phytyl-1,4-hydroquinone methyltransferase, chloroplastic

Table 2-S1 (cont'd)

Chard	Chr5	1124574	1131860	7286	EL10Ac5g10485	0.79	0.62	222	Peptidyl-prolyl cis-trans isomerase CYP63
Chard	Chr5	1135895	1144954	9059	EL10Ac5g10486	0.79	0.25	408	hypothetical protein
Chard	Chr5	1143840	1144625	785	EL10Ac5g10487	0.79	0.58	89	SWIM zinc finger
Chard	Chr5	1580576	1588723	8147	EL10Ac5g10519	0.81	0.38	237	KIP
Chard	Chr5	1605834	1614295	8461	EL10Ac5g10520	0.76	0.31	281	Methylcrotonoyl-CoA carboxylase subunit alpha, mitochondrial
Chard	Chr5	1840594	1843276	2682	EL10Ac5g10537	0.82	0.54	93	Protein YLS3
Chard	Chr5	1863461	1873957	10496	EL10Ac5g10539	0.84	0.32	217	Non-specific lipid transfer protein GPI-anchored 2
Chard	Chr5	1882332	1885775	3443	EL10Ac5g10540	0.75	0.55	120	Putative acyl-activating enzyme
Chard	Chr5	1887154	1892438	5284	EL10Ac5g10541	0.79	0.65	171	tRNA (guanine-N(7)-)-methyltransferase non-catalytic subunit wdr4
Chard	Chr5	1913318	1916612	3294	EL10Ac5g10542	0.73	0.39	46	hypothetical protein
Chard	Chr5	9643119	9645705	2586	EL10Ac5g11039	0.74	0.47	105	Thioredoxin-like
Chard	Chr5	9656309	9657058	749	EL10Ac5g11040	0.71	0.53	142	hypothetical protein
Chard	Chr5	9658989	9669528	10539	EL10Ac5g11041	0.78	0.50	422	UDP-glycosyltransferase 86A
Chard	Chr5	9683276	9694496	11220	EL10Ac5g11042	0.79	0.43	244	11S globulin seed storage protein 2
Chard	Chr5	9713828	9723540	9712	EL10Ac5g11043	0.71	0.56	50	Domain of unknown function (DUF42)
Chard	Chr5	9736158	9739590	3432	EL10Ac5g11044	0.81	0.43	118	hypothetical protein
Chard	Chr5	52109963	52115621	5658	EL10Ac5g12574	0.75	0.39	204	ABC transporter G family member
Chard	Chr5	52171132	52171353	221	EL10Ac5g12575	0.75	0.59	43	hypothetical protein
Chard	Chr5	52196680	52201021	4341	EL10Ac5g12576	0.71	0.41	197	Outer envelope pore protein
Chard	Chr5	52202891	52206643	3752	EL10Ac5g12577	0.72	0.41	194	Uncharacterized protein C24B
Chard	Chr5	52212726	52224917	12191	EL10Ac5g12578	0.70	0.34	377	WAT
Chard	Chr5	52227723	52234818	7095	EL10Ac5g12579	0.80	0.36	159	Transglutaminase-like superfamily
Chard	Chr5	52239950	52248632	8682	EL10Ac5g12581	0.74	0.28	434	hypothetical protein
Chard	Chr5	52250000	52250833	833	EL10Ac5g12582	0.74	0.58	65	Glucuronoxylan 4-O-methyltransferase
Chard	Chr5	52252986	52259955	6969	EL10Ac5g12583	0.76	0.52	273	Superoxide dismutase [Fe] 2, chloroplastic
Chard	Chr5	52261046	52263035	1989	EL10Ac5g12584	0.80	0.52	157	Mediator of RNA polymerase II transcription subunit 22b
Chard	Chr5	52265278	52270380	5102	EL10Ac5g12585	0.79	0.53	150	F-box protein SKIP3
Chard	Chr5	52292141	52294929	2788	EL10Ac5g12586	0.90	0.52	77	hypothetical protein
Chard	Chr5	52295782	52316349	20567	EL10Ac5g12587	0.81	0.49	341	Nucleotide-diphospho-sugar transferase
Chard	Chr5	52346665	52347789	1124	EL10Ac5g12588	0.69	0.33	55	hypothetical protein
Chard	Chr5	52387597	52391277	3680	EL10Ac5g12589	0.74	0.38	146	Abscisic acid 8'-hydroxylase
Chard	Chr5	52417560	52423097	5537	EL10Ac5g12590	0.88	0.55	133	Long-chain-alcohol oxidase FAO4A
Chard	Chr5	52427354	52429979	2625	EL10Ac5g12591	0.84	0.60	66	Protein of unknown function (DUF)
Chard	Chr5	54628530	54633867	5337	EL10Ac5g12744	0.63	0.31	279	Replication factor C subunit 2
Chard	Chr5	54637771	54643286	5515	EL10Ac5g12745	0.69	0.33	230	E3 ubiquitin-protein ligase
Chard	Chr5	54645591	54664625	19034	EL10Ac5g12746	0.70	0.49	680	Protein of unknown function (DUF8)
Chard	Chr5	54646417	54646834	417	EL10Ac5g12747	0.69	0.50	64	Photosystem I P700 chlorophyll a apoprotein A2
Chard	Chr5	54725605	54727324	1719	EL10Ac5g12757	0.63	0.37	115	hypothetical protein
Chard	Chr5	54729805	54738375	8570	EL10Ac5g12758	0.65	0.20	207	Thaumatin-like protein
Chard	Chr5	54739421	54740358	937	EL10Ac5g12759	0.65	0.23	108	Ribosomal protein S3, mitochondrial
Chard	Chr5	54740382	54740870	488	EL10Ac5g12760	0.66	0.26	92	Cytochrome c oxidase subunit
Chard	Chr5	54754919	54755488	569	EL10Ac5g12761	0.63	0.39	24	Reverse transcriptase-like t EL10Ac5g12761 Reverse transcriptase-like
Chard	Chr5	54759051	54760790	1739	EL10Ac5g12762	0.64	0.44	77	C2 domain-containing protein
Chard	Chr5	54761304	54775050	13746	EL10Ac5g12763	0.65	0.44	186	Methyl-CpG-binding domain-containing protein
Chard	Chr5	54784102	54789074	4972	EL10Ac5g12764	0.71	0.39	220	Stress responsive A/B Barrel Domain
Chard	Chr5	54794574	54805457	10883	EL10Ac5g12765	0.82	0.27	388	Putative DEAD-box ATP-dependent RNA helicase 33
Chard	Chr6	6256016	6265065	9049	EL10Ac6g13521	0.81	0.65	56	Histidine kinase 3
Chard	Chr6	55811037	55814587	3550	EL10Ac6g15092	0.82	0.52	72	Succinate dehydrogenase [ubiquinone] iron-sulfur subunit 3, mitochondrial
Chard	Chr7	52022369	52022908	539	EL10Ac7g17979	0.82	0.50	80	Auxin-induced in root cultures protein
Chard	Chr7	52087288	52089912	2624	EL10Ac7g17980	0.78	0.42	91	Cytochrome b56
Chard	Chr8	1120040	1124678	4638	EL10Ac8g18334	0.65	0.47	262	hypothetical protein
Chard	Chr8	1126966	1131181	4215	EL10Ac8g18335	0.65	0.39	204	Pentatricopeptide repeat-containing protein
Chard	Chr8	1132697	1148684	15987	EL10Ac8g18336	0.68	0.33	309	Probable zinc protease PqqL
Chard	Chr8	1155190	1156788	1598	EL10Ac8g18337	0.68	0.46	85	Transcription factor RAX2 t EL10Ac8g18337 Transcription factor RAX2
Chard	Chr8	13583853	13591148	7295	EL10Ac8g19141	0.77	0.54	367	Mitotic checkpoint regulator, MAD2B-interacting
Chard	Chr8	13604511	13613921	9410	EL10Ac8g19142	0.77	0.38	260	Protein bem46
Chard	Chr8	13619876	13624798	4922	EL10Ac8g19143	0.85	0.48	164	PHD finger protein ALFIN-LIKE 5
Chard	Chr8	13633189	13641186	7997	EL10Ac8g19144	0.75	0.39	198	Uncharacterized membrane protein C776
Chard	Chr8	13653071	13654657	1586	EL10Ac8g19145	0.74	0.55	126	GDSL esterase/lipase
Chard	Chr8	13654696	13666106	11410	EL10Ac8g19146	0.83	0.44	324	GDSL esterase/lipase
Chard	Chr8	13686308	13720246	33938	EL10Ac8g19147	0.77	0.37	453	GDSL esterase/lipase At5g03980
Chard	Chr8	13747193	13759733	12540	EL10Ac8g19148	0.76	0.52	208	GDSL esterase/lipase
Chard	Chr8	13782384	13794865	12481	EL10Ac8g19149	0.75	0.45	367	GDSL esterase/lipase
Chard	Chr8	13798264	13798458	194	EL10Ac8g19150	0.71	0.54	36	Photosystem I P700 chlorophyll a apoprotein A
Chard	Chr8	13805532	13827999	22467	EL10Ac8g19151	0.76	0.36	251	Myosin
Chard	Chr8	34052253	34055074	2821	EL10Ac8g19655	0.77	0.61	48	U-box domain-containing protein 9
Chard	Chr8	34099593	34106209	6616	EL10Ac8g19656	0.76	0.38	247	Peroxisomal (S)-2-hydroxy-acid oxidase GLO
Chard	Chr8	34118661	34120224	1563	EL10Ac8g19657	0.79	0.44	114	Protein TIFY 5A
Chard	Chr8	34121890	34122224	334	EL10Ac8g19658	0.74	0.62	92	hypothetical protein
Chard	Chr8	34122288	34122595	307	EL10Ac8g19659	0.74	0.65	78	hypothetical protein
Chard	Chr8	34122638	34123935	1297	EL10Ac8g19660	0.74	0.67	106	DDE superfamily endonuclease
Chard	Chr8	34123938	34125028	1090	EL10Ac8g19661	0.76	0.61	90	hypothetical protein
Chard	Chr8	34158224	34195575	37351	EL10Ac8g19662	0.81	0.36	807	Protein of unknown function (DUF)
Chard	Chr8	51750009	51752074	2065	EL10Ac8g20254	0.75	0.39	53	Heavy-metal-associated domain
Chard	Chr8	51774038	51782626	8588	EL10Ac8g20255	0.81	0.34	196	Heavy-metal-associated domain
Chard	Chr8	54065232	54065480	248	EL10Ac8g20375	0.85	0.42	98	hypothetical protein
Chard	Chr8	55043396	55048539	5143	EL10Ac8g20430	0.84	0.47	162	Protein DEHYDRATION-INDUCED
Chard	Chr8	55062471	55064365	1894	EL10Ac8g20431	0.74	0.41	157	hypothetical protein
Chard	Chr8	55065399	55068301	2902	EL10Ac8g20432	0.74	0.51	199	hypothetical protein
Chard	Chr8	55072112	55074736	2624	EL10Ac8g20433	0.69	0.18	181	Chaperone protein DnaJ
Chard	Chr8	55148828	55150276	1448	EL10Ac8g20438	0.71	0.46	142	Pentatricopeptide repeat-containing protein
Chard	Chr8	55151246	55157051	5805	EL10Ac8g20439	0.71	0.36	294	Protein of unknown function (DUF679)
Chard	Chr8	55179554	55187589	8035	EL10Ac8g20440	0.89	0.50	228	hypothetical protein
Chard	Chr9	32200425	32210685	10260	EL10Ac9g22127	0.82	0.31	194	DNA polymerase V
Chard	Chr9	32214654	32217703	3049	EL10Ac9g22128	0.69	0.51	124	Transcription factor GTE7
Chard	Chr9	32231255	32233969	2714	EL10Ac9g22129	0.72	0.37	110	

Table 2-S1 (cont'd)

Chard	Chr9	32251785	32253197	1412	EL10Ac9g22130	0.73	0.39	67	PB
Fodder	Chr2	6525742	6547542	21800	EL10Ac2g02806	0.67	0.26	114	Probable tRNA N6-adenosine threonylcarbamoyltransferase, mitochondrial
Fodder	Chr2	6584270	6585540	1270	EL10Ac2g02808	0.65	0.41	67	Two-component response regulator ARR9
Sugar	Chr1	17999804	18002243	2439	EL10Ac1g01251	0.71	0.44	56	Probable trehalose-phosphate phosphatase D
Sugar	Chr1	18082596	18098518	15922	EL10Ac1g01252	0.76	0.30	256	Endoplasmic reticulum-Golgi intermediate compartment protein 3
Sugar	Chr2	50160084	50163080	2996	EL10Ac2g04512	0.87	0.62	92	Pentatricopeptide repeat-containing protein, mitochondrial
Sugar	Chr2	50164439	50167338	2899	EL10Ac2g04513	0.87	0.67	83	cAMP-regulated phosphoprotein/endosulfine conserved region
Sugar	Chr3	23241971	23242579	608	EL10Ac3g06337	0.87	0.52	94	hypothetical protein
Sugar	Chr3	23266082	23284333	18251	EL10Ac3g06338	0.86	0.50	218	hypothetical protein
Sugar	Chr3	23313137	23313525	388	EL10Ac3g06339	0.75	0.66	51	gag-polypeptide of LTR copia-type
Sugar	Chr3	23317099	23333823	16724	EL10Ac3g06340	0.79	0.56	395	DUF2
Sugar	Chr3	23317814	23326286	8472	EL10Ac3g06341	0.79	0.61	215	hypothetical protein
Sugar	Chr3	23419906	23432678	12772	EL10Ac3g06342	0.77	0.46	269	DUF2
Sugar	Chr3	23494631	23513691	19060	EL10Ac3g06343	0.76	0.43	296	hypothetical protein
Sugar	Chr3	23527425	23528852	1427	EL10Ac3g06344	0.86	0.74	97	hypothetical protein
Sugar	Chr3	51060282	51063512	3230	EL10Ac3g07284	0.74	0.41	101	Pentatricopeptide repeat-containing protein
Sugar	Chr4	2887833	2899041	11208	EL10Ac4g07734	0.71	0.28	415	hypothetical protein
Sugar	Chr5	4400661	4403470	2809	EL10Ac5g10742	0.63	0.40	89	Dof zinc finger protein DOF5
Sugar	Chr8	14505353	14510538	5185	EL10Ac8g19192	0.84	0.37	148	Putative transcription factor bHLH04
Table	Chr1	4631423	4639952	8529	EL10Ac1g00390	0.90	0.46	251	Protein of unknown function (DUF3522)
Table	Chr1	4639507	4645464	5957	EL10Ac1g00391	0.75	0.48	210	Calmodulin-binding receptor-like cytoplasmic kinase 2
Table	Chr1	4648990	4650936	1946	EL10Ac1g00392	0.82	0.52	103	Pentatricopeptide repeat-containing protein, mitochondrial
Table	Chr1	5660004	5665192	5188	EL10Ac1g00465	0.73	0.42	109	Oligopeptide transporter 2
Table	Chr1	5668843	5670684	1841	EL10Ac1g00466	0.85	0.48	60	Pentatricopeptide repeat-containing protein At
Table	Chr1	5687918	5691714	3796	EL10Ac1g00467	0.82	0.73	111	hypothetical protein
Table	Chr1	5697203	5698050	847	EL10Ac1g00468	0.82	0.76	37	Agamous-like MADS-box protein AGL
Table	Chr1	5712716	5714024	1308	EL10Ac1g00469	0.84	0.66	49	MADS-box transcription factor ANR
Table	Chr1	5724012	5725622	1610	EL10Ac1g00470	0.84	0.50	81	Putative GEM-like protein 8
Table	Chr1	5738322	5739939	1617	EL10Ac1g00471	0.72	0.46	74	GEM-like protein 4
Table	Chr1	5742359	5753296	10937	EL10Ac1g00472	0.85	0.53	251	Transcription factor DIVARICATA
Table	Chr1	14217184	14218739	1555	EL10Ac1g01074	0.70	0.33	114	NAD(P)H-quinone oxidoreductase subunit N
Table	Chr1	14249315	14252968	3653	EL10Ac1g01077	0.73	0.38	170	hypothetical protein
Table	Chr1	14255208	14266282	11074	EL10Ac1g01078	0.77	0.28	311	hypothetical protein
Table	Chr1	14273877	14280658	6781	EL10Ac1g01079	0.77	0.26	249	Glucose-6-phosphate isomerase
Table	Chr1	14285048	14289090	4042	EL10Ac1g01080	0.75	0.51	102	DnAJ-like protein slr0093
Table	Chr1	14289472	14304514	15042	EL10Ac1g01081	0.82	0.41	333	Protein TRANSPARENT TESTA
Table	Chr1	15245152	15246375	1223	EL10Ac1g01121	0.85	0.55	106	E3 ubiquitin-protein ligase ATL6
Table	Chr1	16878566	16908888	30322	EL10Ac1g01197	0.77	0.37	73	Putative pentatricopeptide repeat-containing protein
Table	Chr1	16908955	16918844	9889	EL10Ac1g01198	0.78	0.35	30	Putative pentatricopeptide repeat-containing protein
Table	Chr2	8096936	8100260	3324	EL10Ac2g02886	0.89	0.57	127	Cytokinin dehydrogenase 6
Table	Chr2	8121488	8126036	4548	EL10Ac2g02887	0.71	0.56	21	hypothetical protein
Table	Chr2	8163438	8169350	5912	EL10Ac2g02888	0.91	0.75	259	hypothetical protein
Table	Chr2	8198940	8208452	9512	EL10Ac2g02889	0.84	0.31	123	Putative calcium-transporting ATPase
Table	Chr2	11922362	11924071	1709	EL10Ac2g03009	0.75	0.64	27	Mannose/glucose-specific lectin
Table	Chr2	11928837	11929608	771	EL10Ac2g03010	0.74	0.65	19	SPX domain-containing protein 4
Table	Chr2	11965616	11967182	1566	EL10Ac2g03011	0.73	0.54	39	Mannose/glucose-specific lectin
Table	Chr2	11977163	11977816	653	EL10Ac2g03012	0.75	0.39	35	SPX domain-containing protein 4
Table	Chr2	11989752	11991075	1323	EL10Ac2g03013	0.78	0.33	33	hypothetical protein
Table	Chr2	11991152	11991660	508	EL10Ac2g03014	0.79	0.57	21	hypothetical protein
Table	Chr2	12008833	12012683	3850	EL10Ac2g03015	0.72	0.50	89	Transmembrane emp24 domain-containing protein p24delta7
Table	Chr2	12031539	12032693	1154	EL10Ac2g03016	0.70	0.57	12	Protein of unknown function (DUF3755)
Table	Chr2	12062142	12067141	4999	EL10Ac2g03017	0.77	0.60	27	Transposase-associated domain
Table	Chr2	12072832	12076756	3924	EL10Ac2g03018	0.82	0.71	112	Pectinesterase 3
Table	Chr2	12083661	12088812	5151	EL10Ac2g03019	0.80	0.47	78	Protein of unknown function (DUF)
Table	Chr2	12105780	12136191	30411	EL10Ac2g03020	0.82	0.50	594	CSC1-like protein HYP1
Table	Chr2	47016285	47019498	3213	EL10Ac2g04244	0.83	0.42	195	Putative methyltransferase NSUN6
Table	Chr2	47019972	47030041	10069	EL10Ac2g04245	0.75	0.38	359	B-box zinc finger
Table	Chr2	47069413	47069691	278	EL10Ac2g04247	0.64	0.52	37	hypothetical protein
Table	Chr2	47075889	47081920	6031	EL10Ac2g04248	0.63	0.40	248	Endo-1,31,4-beta-D-glucanase
Table	Chr2	47095856	47103859	8003	EL10Ac2g04249	0.65	0.38	232	Potassium transporter 2
Table	Chr2	47105103	47106513	1410	EL10Ac2g04250	0.63	0.43	115	hypothetical protein
Table	Chr2	47106285	47107613	1328	EL10Ac2g04251	0.62	0.50	105	hypothetical protein
Table	Chr2	47122003	47134377	12374	EL10Ac2g04255	0.68	0.31	245	Isoflavone 2'-hydroxylase
Table	Chr2	47142881	47152204	9323	EL10Ac2g04256	0.65	0.39	260	TLC ATP/ADP transporter
Table	Chr2	47152665	47156761	4096	EL10Ac2g04257	0.70	0.33	125	Phosphoglucan phosphatase LSF2, chloroplastic
Table	Chr2	47167548	47169317	1769	EL10Ac2g04258	0.70	0.33	93	Adenine/guanine permease AZG
Table	Chr2	47234430	47238546	4116	EL10Ac2g04263	0.68	0.39	194	Uroporphyrinogen decarboxylase, chloroplastic
Table	Chr2	47240625	47248240	7615	EL10Ac2g04264	0.74	0.29	258	Phosphorylated carbohydrates phosphatase
Table	Chr2	47251517	47254016	2499	EL10Ac2g04265	0.72	0.43	160	High mobility group B protein 7
Table	Chr2	47256282	47262559	6277	EL10Ac2g04266	0.76	0.31	198	WEB family protein
Table	Chr3	1549708	1555092	5384	EL10Ac3g05026	0.82	0.42	174	Probable polygalacturonase
Table	Chr3	2183863	2187291	3428	EL10Ac3g05089	0.81	0.35	161	UDP-glycosyltransferase 78D2
Table	Chr3	2189984	2198435	8451	EL10Ac3g05090	0.73	0.30	428	ADP-ribosylation factor
Table	Chr3	2199858	2219280	19422	EL10Ac3g05091	0.85	0.53	447	Probable GTP diphosphokinase RSH3, chloroplastic
Table	Chr3	3220257	3223390	3133	EL10Ac3g05180	0.71	0.54	84	Ribosome-binding factor PSRP
Table	Chr3	3224878	3226906	2028	EL10Ac3g05181	0.72	0.55	56	mTERF
Table	Chr3	3244544	3248941	4397	EL10Ac3g05183	0.75	0.51	86	StAR-related lipid transfer protein 7, mitochondrial
Table	Chr3	3257425	3257700	275	EL10Ac3g05184	0.69	0.43	51	hypothetical protein
Table	Chr3	3270931	3275379	4448	EL10Ac3g05186	0.69	0.29	110	Granule-bound starch synthase
Table	Chr3	3310665	3324644	13979	EL10Ac3g05189	0.74	0.17	323	Transcription factor IIIC subunit delta N-term
Table	Chr3	3338658	3345650	6992	EL10Ac3g05190	0.73	0.33	141	Polyadenylate-binding protein-interacting protein
Table	Chr3	3357865	3358395	530	EL10Ac3g05191	0.71	0.44	46	Domain of unknown function (DUF4228)
Table	Chr3	3382289	3382777	488	EL10Ac3g05193	0.70	0.53	23	Domain of unknown function (DUF4228)

Table 2-S1 (cont'd)

Table	Chr3	3390130	3393409	3279	EL10Ac3g05194	0.70	0.47	149	Jasmonate-induced protein homolog
Table	Chr3	3394334	3396469	2135	EL10Ac3g05195	0.70	0.42	105	Small heat shock protein, chloroplastic
Table	Chr3	3417710	3424710	7000	EL10Ac3g05196	0.69	0.37	122	Growth-regulating factor 8
Table	Chr3	3511037	3511456	419	EL10Ac3g05203	0.73	0.43	86	Auxin-induced protein
Table	Chr3	3602728	3605443	2715	EL10Ac3g05210	0.70	0.48	126	Ubiquitin-60S ribosomal protein L40
Table	Chr3	3607907	3618455	10548	EL10Ac3g05211	0.84	0.46	344	Nuclear pore complex protein NUP96
Table	Chr3	11863913	11868795	4882	EL10Ac3g05839	0.82	0.46	147	Vesicle-associated protein
Table	Chr3	11875997	11876509	512	EL10Ac3g05840	0.83	0.63	76	Transcriptional regulator TAC
Table	Chr3	11878635	11890018	11383	EL10Ac3g05841	0.87	0.65	273	E3 ubiquitin protein ligase RIN2
Table	Chr3	11897295	11905272	7977	EL10Ac3g05842	0.84	0.32	196	Proteasome subunit alpha type-5
Table	Chr3	11908720	11914910	6190	EL10Ac3g05843	0.84	0.62	283	Domain of unknown function (DUF4535)
Table	Chr3	11917980	11921127	3147	EL10Ac3g05844	0.76	0.46	118	Putative glycerol-3-phosphate transporter
Table	Chr3	11949749	11955487	5738	EL10Ac3g05845	0.70	0.52	103	Luc7-like protein 3
Table	Chr3	11956482	11957977	1495	EL10Ac3g05846	0.72	0.59	55	Probable aquaporin TIP5
Table	Chr3	11959217	11959960	743	EL10Ac3g05847	0.73	0.65	53	Zinc finger protein
Table	Chr3	11961195	11967400	6205	EL10Ac3g05848	0.76	0.63	159	Cell number regulator 6
Table	Chr3	11979209	11981985	2776	EL10Ac3g05849	0.74	0.61	104	Cytochrome c-type biogenesis protein CcmE
Table	Chr3	11985234	11992667	7433	EL10Ac3g05850	0.74	0.58	141	NO-associated protein
Table	Chr3	12001742	12005256	3514	EL10Ac3g05851	0.80	0.55	111	Aldo-keto reductase family 4 member C9
Table	Chr3	12004548	12026199	21651	EL10Ac3g05852	0.80	0.46	317	Aldo-keto reductase family 4 member C
Table	Chr3	12033774	12042539	8765	EL10Ac3g05853	0.73	0.36	186	Uncharacterized PKHD-type hydroxylase
Table	Chr3	12045378	12048604	3226	EL10Ac3g05854	0.82	0.26	160	Receptor-like protein
Table	Chr3	12058448	12064374	5926	EL10Ac3g05855	0.78	0.48	103	Acetyltransferase (GNAT) domain
Table	Chr3	12070552	12097386	26834	EL10Ac3g05856	0.81	0.53	451	Structural maintenance of chromosomes protein 6B
Table	Chr3	53025474	53031833	6359	EL10Ac3g07411	0.83	0.42	294	MACPF domain-containing protein
Table	Chr3	53039104	53041730	2626	EL10Ac3g07412	0.78	0.38	216	40S ribosomal protein S30
Table	Chr3	53044926	53045774	848	EL10Ac3g07413	0.70	0.31	82	Protein MIZU-KUSSEI
Table	Chr3	53185694	53193924	8230	EL10Ac3g07421	0.64	0.17	256	Kinesin-like protein KIN
Table	Chr3	53207600	53218800	11200	EL10Ac3g07424	0.63	0.26	317	Probable acyl-activating enzyme
Table	Chr3	53236037	53238577	2540	EL10Ac3g07426	0.66	0.38	182	Probable receptor protein kinase TMK
Table	Chr3	53243896	53245332	1436	EL10Ac3g07427	0.69	0.32	166	Crocetin glucosyltransferase, chloroplastic
Table	Chr3	53260302	53263780	3478	EL10Ac3g07428	0.66	0.43	220	hypothetical protein
Table	Chr3	53263426	53281155	17729	EL10Ac3g07429	0.76	0.38	527	Probable xyloglucan endotransglucosylase/hydrolase protein
Table	Chr3	53289542	53293052	3510	EL10Ac3g07430	0.76	0.34	155	Probable xyloglucan endotransglucosylase/hydrolase protein
Table	Chr3	53305490	53312975	7485	EL10Ac3g07432	0.70	0.34	251	Putative E3 ubiquitin-protein ligase RF298
Table	Chr3	53335779	53343052	7273	EL10Ac3g07435	0.70	0.44	302	Dihydroorotase, mitochondrial
Table	Chr3	53519734	53523248	3514	EL10Ac3g07453	0.61	0.36	126	Serine hydroxymethyltransferase 4
Table	Chr3	53524500	53553849	29349	EL10Ac3g07454	0.62	0.27	388	DNA repair protein RAD50
Table	Chr3	53555895	53561372	5477	EL10Ac3g07455	0.86	0.45	247	Werner Syndrome-like exonuclease
Table	Chr4	54281404	54285100	3696	EL10Ac4g09768	0.71	0.37	50	Ammonium transporter
Table	Chr4	54284425	54285918	1493	EL10Ac4g09769	0.70	0.33	65	Ammonium transporter
Table	Chr4	54322888	54334029	11141	EL10Ac4g09774	0.69	0.25	370	Domain of unknown function (DUF4409)
Table	Chr4	54488779	54494782	6003	EL10Ac4g09785	0.71	0.34	234	Uncharacterized protein At
Table	Chr4	54496402	54505069	8667	EL10Ac4g09786	0.69	0.45	173	Pentatricopeptide repeat-containing protein
Table	Chr4	54505409	54510232	4823	EL10Ac4g09787	0.70	0.45	152	SufE-like protein, chloroplastic
Table	Chr4	54519720	54541720	22000	EL10Ac4g09788	0.71	0.40	608	Protein PIR
Table	Chr4	54691398	54697063	5665	EL10Ac4g09803	0.64	0.33	156	Violaxanthin de-epoxidase, chloroplastic
Table	Chr4	54695620	54699187	3567	EL10Ac4g09804	0.64	0.32	113	Serine/threonine-protein kinase
Table	Chr4	54701581	54705468	3887	EL10Ac4g09805	0.63	0.43	171	Uncharacterized protein
Table	Chr4	54710230	54712251	2021	EL10Ac4g09806	0.65	0.36	101	Pentatricopeptide repeat-containing protein
Table	Chr4	54837482	54841273	3791	EL10Ac4g09818	0.63	0.48	105	Universal stress protein A-like protein
Table	Chr4	54841626	54845836	4210	EL10Ac4g09819	0.63	0.49	105	Calreticulin
Table	Chr4	54856906	54863178	6272	EL10Ac4g09820	0.63	0.45	96	RNA pseudouridine synthase
Table	Chr4	54864583	54873052	8469	EL10Ac4g09821	0.64	0.43	262	Peptide chain release factor PrfB2, chloroplastic
Table	Chr4	54939640	54940644	1004	EL10Ac4g09822	0.64	0.42	63	hypothetical protein
Table	Chr4	54957955	54958146	191	EL10Ac4g09823	0.62	0.57	19	hypothetical protein
Table	Chr4	54958255	54959033	778	EL10Ac4g09824	0.62	0.56	28	Putative pentatricopeptide repeat-containing protein
Table	Chr4	54959049	54959246	197	EL10Ac4g09825	0.62	0.57	21	Pentatricopeptide repeat-containing protein, mitochondrial
Table	Chr4	54959258	54960303	1045	EL10Ac4g09826	0.62	0.56	22	Putative pentatricopeptide repeat-containing protein
Table	Chr4	54966645	54968102	1457	EL10Ac4g09827	0.62	0.49	96	hypothetical protein
Table	Chr4	54967840	54970416	2576	EL10Ac4g09828	0.62	0.49	87	Pentatricopeptide repeat-containing protein
Table	Chr4	54972790	54979936	7146	EL10Ac4g09829	0.62	0.40	124	Zinc finger matrin-type protein 2
Table	Chr4	54982169	54989211	7042	EL10Ac4g09830	0.62	0.46	211	Probable protein disulfide-isomerase A6
Table	Chr4	54990346	54996981	6635	EL10Ac4g09831	0.61	0.60	2	RNA recognition motif
Table	Chr4	55026020	55027213	1193	EL10Ac4g09832	0.62	0.34	65	Protein of unknown function (DUF)
Table	Chr4	55037615	55046402	8787	EL10Ac4g09833	0.66	0.37	319	DUF76
Table	Chr4	55048669	55055577	6908	EL10Ac4g09834	0.69	0.39	195	ATP-dependent DNA helicase Q-like
Table	Chr4	55057225	55060730	3505	EL10Ac4g09835	0.69	0.47	133	Calmodulin binding protein-like
Table	Chr4	55060062	55061609	1547	EL10Ac4g09836	0.67	0.43	87	Pentatricopeptide repeat-containing protein
Table	Chr4	55063765	55068989	5224	EL10Ac4g09838	0.67	0.39	239	ABC transporter F family member 5
Table	Chr4	55073565	55078727	5162	EL10Ac4g09839	0.67	0.48	111	E3 ubiquitin-protein ligase
Table	Chr4	55078292	55087140	8848	EL10Ac4g09840	0.66	0.45	160	Protein DEHYDRATION-INDUCED
Table	Chr4	55092908	55097893	4985	EL10Ac4g09841	0.68	0.33	159	Syntaxin-4
Table	Chr4	55114257	55115266	1009	EL10Ac4g09843	0.66	0.49	56	GATA transcription factor
Table	Chr4	55123568	55124374	806	EL10Ac4g09844	0.67	0.45	72	Probable ribose-5-phosphate isomerase 2
Table	Chr4	55125054	55125633	579	EL10Ac4g09845	0.66	0.48	106	hypothetical protein
Table	Chr4	55125797	55126027	230	EL10Ac4g09846	0.66	0.49	112	hypothetical protein
Table	Chr4	55127227	55127626	399	EL10Ac4g09847	0.66	0.46	115	hypothetical protein
Table	Chr4	55134390	55138303	3913	EL10Ac4g09848	0.65	0.30	150	DnaJ homolog subfamily B member 6
Table	Chr4	55143097	55151177	8080	EL10Ac4g09849	0.62	0.59	2	hypothetical protein
Table	Chr4	55161680	55162321	641	EL10Ac4g09850	0.66	0.36	63	Putative pentatricopeptide repeat-containing protein
Table	Chr4	55162769	55169917	7148	EL10Ac4g09851	0.66	0.29	262	Calmodulin binding protein-like
Table	Chr4	55179272	55182991	3719	EL10Ac4g09853	0.64	0.45	87	hypothetical protein
Table	Chr4	55231840	55237187	5347	EL10Ac4g09855	0.68	0.43	129	Cell division control protein 48 homolog C
Table	Chr4	55240675	55247186	6511	EL10Ac4g09856	0.67	0.35	164	hypothetical protein
Table	Chr4	55254801	55256218	1417	EL10Ac4g09857	0.70	0.66	5	Domain of unknown function (DUF4283)
Table	Chr4	55268855	55273484	4629	EL10Ac4g09858	0.69	0.40	196	Calmodulin binding protein-like

Table 2-S1 (cont'd)

Table	Chr4	55280230	55281125	895	EL10Ac4g09859	0.64	0.45	22	hypothetical protein
Table	Chr4	55281159	55287576	6417	EL10Ac4g09860	0.70	0.40	119	Calmodulin binding protein-like
Table	Chr4	55305859	55326389	20530	EL10Ac4g09861	0.70	0.33	378	Exopolyphosphatase
Table	Chr4	55327174	55327545	371	EL10Ac4g09862	0.70	0.49	71	Domain of unknown function (DUF35
Table	Chr4	55350877	55357918	7041	EL10Ac4g09864	0.68	0.32	183	Kinesin-4
Table	Chr4	55359731	55363711	3980	EL10Ac4g09865	0.73	0.40	108	Probable protein phosphatase 2C 5
Table	Chr4	55395037	55399654	4617	EL10Ac4g09868	0.69	0.32	187	Single-stranded DNA-binding protein, mitochondrial
Table	Chr4	55410865	55420172	9307	EL10Ac4g09869	0.70	0.46	402	Eukaryotic translation initiation factor 3 subunit A
Table	Chr4	55465972	55467464	1492	EL10Ac4g09873	0.70	0.43	93	hypothetical protein
Table	Chr4	55471884	55477173	5289	EL10Ac4g09874	0.70	0.30	207	Probable protein phosphatase 2C 73
Table	Chr4	55518027	55530479	12452	EL10Ac4g09878	0.69	0.32	325	Phospholipase D
Table	Chr4	55545782	55548216	2434	EL10Ac4g09881	0.68	0.37	183	Tetratricopeptide repeat
Table	Chr4	55549277	55553010	3733	EL10Ac4g09882	0.67	0.50	171	60S ribosomal protein L
Table	Chr4	55696832	55701200	4368	EL10Ac4g09895	0.72	0.38	164	Bifunctional epoxide hydrolase 2
Table	Chr4	55701830	55704832	3002	EL10Ac4g09896	0.72	0.32	154	60S ribosomal protein L
Table	Chr4	55707754	55712972	5218	EL10Ac4g09897	0.74	0.35	160	Malignant T-cell-amplified sequence
Table	Chr4	55723621	55726727	3106	EL10Ac4g09898	0.83	0.54	83	Bidirectional sugar transporter SWEET
Table	Chr4	55734181	55740992	6811	EL10Ac4g09899	0.83	0.65	70	Putative splicing factor C222
Table	Chr4	55743036	55747227	4191	EL10Ac4g09900	0.78	0.31	135	Serine/threonine-protein phosphatase PP
Table	Chr5	42787256	42787570	314	EL10Ac5g12096	0.81	0.62	38	hypothetical protein
Table	Chr5	44329571	44332777	3206	EL10Ac5g12156	0.71	0.31	50	Protein FAR
Table	Chr5	44361572	44372492	10920	EL10Ac5g12157	0.73	0.46	265	Probable serine/threonine-protein kinase
Table	Chr5	44376295	44384701	8406	EL10Ac5g12158	0.74	0.42	299	Ent-kaurenoic acid oxidase 2
Table	Chr5	44391780	44406558	14778	EL10Ac5g12159	0.71	0.41	128	Glutamate receptor 2
Table	Chr5	44429172	44435821	6649	EL10Ac5g12160	0.62	0.42	58	Nuclear cap-binding protein subunit 2
Table	Chr5	44456897	44464307	7410	EL10Ac5g12161	0.69	0.41	148	Sister chromatid cohesion
Table	Chr5	44464218	44466276	2058	EL10Ac5g12162	0.69	0.49	67	Glycine cleavage system H protein, mitochondrial
Table	Chr5	44524019	44524354	335	EL10Ac5g12164	0.71	0.61	9	Zinc-finger homeodomain protein 9
Table	Chr5	46277751	46285517	7766	EL10Ac5g12239	0.81	0.38	208	Decapping nuclease DXO homolog, chloroplastic
Table	Chr6	1562468	1563154	686	EL10Ac6g13186	0.63	0.50	44	Trypsin inhibitor
Table	Chr6	1624032	1627180	3148	EL10Ac6g13193	0.68	0.29	173	Origin of replication complex subunit 6
Table	Chr6	1787997	1794102	6105	EL10Ac6g13204	0.73	0.36	291	Protein of unknown function (DUF)
Table	Chr6	1856434	1860539	4105	EL10Ac6g13207	0.63	0.42	221	E3 ubiquitin-protein ligase MARCH2
Table	Chr6	1864371	1873464	9093	EL10Ac6g13208	0.63	0.21	206	Probable apyrase 6
Table	Chr6	2038709	2040497	1788	EL10Ac6g13220	0.65	0.42	115	Myb family transcription factor APL
Table	Chr6	2055511	2056689	1178	EL10Ac6g13221	0.69	0.37	53	hypothetical protein
Table	Chr6	2076042	2078764	2722	EL10Ac6g13222	0.74	0.54	75	Cyclic dof factor
Table	Chr6	2084451	2118867	34416	EL10Ac6g13223	0.67	0.14	681	ABC transporter C family member 2
Table	Chr6	17691899	17698716	6817	EL10Ac6g13974	0.81	0.52	152	Alpha-mannosidase
Table	Chr6	17783384	17799591	16207	EL10Ac6g13975	0.74	0.51	46	Ubiquitin-like domain-containing CTD phosphatase
Table	Chr6	17874246	17879128	4882	EL10Ac6g13977	0.88	0.66	176	Geranylgeranyl transferase type-2 subunit alpha
Table	Chr6	18011611	18018194	6583	EL10Ac6g13978	0.72	0.58	21	Dynamin-2A
Table	Chr6	18022058	18024560	2502	EL10Ac6g13979	0.80	0.54	118	hypothetical protein
Table	Chr6	18061419	18062548	1129	EL10Ac6g13980	0.73	0.46	45	Probable transcriptional regulator SLK2
Table	Chr6	18115882	18120231	4349	EL10Ac6g13981	0.71	0.46	116	GTP cyclohydrolase
Table	Chr6	18130141	18159312	29171	EL10Ac6g13982	0.74	0.43	406	Cullin-associated NEDD8-dissociated protein
Table	Chr6	18229764	18255079	25315	EL10Ac6g13983	0.75	0.35	484	ATP-dependent Clp protease ATP-binding subunit ClpX
Table	Chr6	18267843	18278353	10510	EL10Ac6g13984	0.80	0.54	254	Putative Holliday junction resolvase
Table	Chr6	18306151	18317415	11264	EL10Ac6g13986	0.83	0.47	214	GPN-loop GTPase 3
Table	Chr6	18306950	18309773	2823	EL10Ac6g13987	0.81	0.51	82	Probable glutathione peroxidase 8
Table	Chr6	18349885	18350241	356	EL10Ac6g13988	0.82	0.50	23	20 kDa chaperonin, chloroplastic
Table	Chr6	18352959	18367495	14536	EL10Ac6g13989	0.87	0.53	363	Reverse transcriptase-like
Table	Chr6	18390664	18410663	19999	EL10Ac6g13990	0.79	0.50	271	Survival of motor neuron-related-splicing factor 30
Table	Chr6	18405942	18412091	6149	EL10Ac6g13991	0.79	0.52	111	hypothetical protein
Table	Chr6	18505763	18506914	1151	EL10Ac6g13992	0.74	0.34	142	Putative ribonuclease H protein
Table	Chr6	18568990	18576706	7716	EL10Ac6g13994	0.76	0.48	73	Armadillo repeat-containing kinesin-like protein 3
Table	Chr6	18609565	18622088	12523	EL10Ac6g13995	0.86	0.49	316	Protein NRT
Table	Chr6	18757141	18776854	19713	EL10Ac6g13996	0.76	0.37	267	hypothetical protein
Table	Chr6	18809816	18833641	23825	EL10Ac6g13997	0.79	0.56	627	Mitotic spindle checkpoint protein MAD
Table	Chr6	18852291	18853115	824	EL10Ac6g13998	0.72	0.40	99	hypothetical protein
Table	Chr6	18855299	18872492	17193	EL10Ac6g13999	0.77	0.51	436	Superkiller viralicidic activity 2-like 2
Table	Chr6	18886424	18899463	13039	EL10Ac6g14000	0.74	0.37	249	Auxin response factor
Table	Chr6	19094463	19107114	12651	EL10Ac6g14001	0.71	0.33	330	ATP-dependent RNA helicase SUV3L, mitochondrial
Table	Chr6	19185570	19200991	15421	EL10Ac6g14004	0.71	0.40	238	Dynamin-2A
Table	Chr6	19212456	19214392	1936	EL10Ac6g14005	0.71	0.45	83	60S ribosomal protein L6
Table	Chr6	19240934	19268163	27229	EL10Ac6g14006	0.73	0.39	603	Protein of unknown function, DUF482
Table	Chr6	19434055	19438027	3972	EL10Ac6g14009	0.76	0.32	119	Core-2/I-Branching enzyme
Table	Chr6	19484639	19485787	1148	EL10Ac6g14011	0.66	0.45	43	BTB/POZ domain-containing protein
Table	Chr6	19586938	19589540	2602	EL10Ac6g14016	0.62	0.57	9	GDSL esterase/lipase
Table	Chr6	19596509	19597256	747	EL10Ac6g14017	0.63	0.51	7	GDSL esterase/lipase
Table	Chr6	19786281	19831964	45683	EL10Ac6g14025	0.68	0.27	1225	hypothetical protein
Table	Chr6	19981357	20000759	19402	EL10Ac6g14031	0.66	0.29	203	Protein BASIC PENTACYSTEINE7
Table	Chr6	20004604	20024481	19877	EL10Ac6g14032	0.67	0.23	437	Ras-related protein RABD
Table	Chr6	20216640	20226594	9954	EL10Ac6g14035	0.85	0.42	285	Endoglucanase
Table	Chr6	20262555	20288651	26096	EL10Ac6g14036	0.81	0.57	412	U3 small nucleolar RNA-associated protein 2
Table	Chr7	5200878	5203819	2941	EL10Ac7g16236	0.82	0.47	68	hypothetical protein
Table	Chr8	1004278	1018088	13810	EL10Ac8g18327	0.83	0.38	455	Cadmium/zinc-transporting ATPase HMA2
Table	Chr8	1226877	1231934	5057	EL10Ac8g18341	0.79	0.46	209	Gamma-glutamyltranspeptidase 3
Table	Chr8	1243297	1248839	5542	EL10Ac8g18342	0.71	0.41	120	ABC transporter B family member 2
Table	Chr8	1251556	1258089	6533	EL10Ac8g18343	0.77	0.18	201	Domain of unknown function (DUF4283)
Table	Chr8	1260672	1275448	14776	EL10Ac8g18344	0.87	0.42	549	Cell division cycle protein 27 homolog B
Table	Chr8	1281593	1283487	1894	EL10Ac8g18345	0.73	0.24	159	F-box/FBD/LRR-repeat protein
Table	Chr8	1290271	1293637	3366	EL10Ac8g18346	0.72	0.44	105	F-box/FBD/LRR-repeat protein At
Table	Chr8	1299474	1302692	3218	EL10Ac8g18347	0.73	0.46	168	AP-4 complex subunit sigma
Table	Chr8	1304840	1308523	3683	EL10Ac8g18348	0.72	0.37	155	ATP-dependent 6-phosphofructokinase 3
Table	Chr8	1322600	1329995	7395	EL10Ac8g18349	0.73	0.34	335	Mitochondrial-processing peptidase subunit alpha
Table	Chr8	1350616	1360842	10226	EL10Ac8g18350	0.73	0.41	334	NAC domain-containing protein 8
Table	Chr8	1373867	1378952	5085	EL10Ac8g18351	0.74	0.47	239	Putative dual specificity protein phosphatase DSP8

Table 2-S1 (cont'd)

Table	Chr8	1381725	1383023	1298	EL10Ac8g18352	0.82	0.58	61	Leucine-rich repeat extensin-like protein 4
Table	Chr8	1393125	1398885	5760	EL10Ac8g18353	0.82	0.48	214	Branched-chain-amino-acid aminotransferase 2, chloroplastic
Table	Chr8	4770121	4775565	5444	EL10Ac8g18598	0.69	0.26	41	hypothetical protein
Table	Chr8	4780618	4787565	6947	EL10Ac8g18599	0.75	0.47	272	Acyl-protein thioesterase 2
Table	Chr8	4831731	4834439	2708	EL10Ac8g18604	0.70	0.19	100	Protein kinase PINOID 2
Table	Chr8	4849366	4852284	2918	EL10Ac8g18605	0.78	0.61	118	Early nodulin-93
Table	Chr8	4860895	4862399	1504	EL10Ac8g18606	0.77	0.60	79	Early nodulin-93
Table	Chr8	4872275	4874482	2207	EL10Ac8g18608	0.65	0.37	105	Pentatricopeptide repeat-containing protein
Table	Chr8	4937262	4937795	533	EL10Ac8g18615	0.68	0.45	85	Auxin-binding protein ABP
Table	Chr8	4938928	4944977	6049	EL10Ac8g18616	0.69	0.45	368	Nudix hydrolase
Table	Chr8	4946515	4948426	1911	EL10Ac8g18617	0.70	0.41	110	Probable amino-acid racemase
Table	Chr8	4950753	4957298	6545	EL10Ac8g18618	0.83	0.49	211	NADH-cytochrome b5 reductase-like protein
Table	Chr8	46239642	46243654	4012	EL10Ac8g20012	0.74	0.19	110	Protein of unknown function (DUF36)
Table	Chr8	46246441	46249899	3458	EL10Ac8g20013	0.74	0.39	160	hypothetical protein
Table	Chr8	46250421	46253491	3070	EL10Ac8g20014	0.72	0.41	174	Double-stranded RNA-binding protein
Table	Chr8	46273580	46277464	3884	EL10Ac8g20015	0.74	0.47	112	Beta-glucosidase 46
Table	Chr8	46278067	46294962	16895	EL10Ac8g20016	0.76	0.47	133	RNA pseudouridine synthase 6, chloroplastic
Table	Chr8	46330266	46343124	12858	EL10Ac8g20017	0.80	0.40	196	RNA pseudouridine synthase 6, chloroplastic
Table	Chr8	46393141	46396779	3638	EL10Ac8g20018	0.75	0.35	139	Cytochrome P450
Table	Chr8	46426935	46431415	4480	EL10Ac8g20019	0.72	0.38	187	Adenosine deaminase-like protein
Table	Chr8	46435507	46445329	9822	EL10Ac8g20020	0.75	0.41	314	Phospholipase A
Table	Chr8	46447499	46447876	377	EL10Ac8g20021	0.81	0.38	69	hypothetical protein
Table	Chr8	46449727	46460636	10909	EL10Ac8g20022	0.86	0.49	353	Serine/threonine-protein kinase PBS
Table	Chr8	46484240	46488043	3803	EL10Ac8g20023	0.85	0.58	148	Cardiolipin synthase, mitochondrial
Table	Chr8	46488622	46491084	2462	EL10Ac8g20024	0.80	0.48	101	Pentatricopeptide repeat-containing protein At
Table	Chr8	46494896	46502641	7745	EL10Ac8g20025	0.76	0.40	236	Tobamovirus multiplication protein
Table	Chr8	46508499	46514051	5552	EL10Ac8g20026	0.75	0.46	204	Derlin-2
Table	Chr8	46566616	46575252	8636	EL10Ac8g20027	0.70	0.45	328	Abnormal spindle-like microcephaly-associated protein homolog
Table	Chr8	46571138	46577944	6806	EL10Ac8g20028	0.74	0.38	279	
Table	Chr8	46583794	46585692	1898	EL10Ac8g20029	0.71	0.54	79	Zinc finger MYND domain-containing protein
Table	Chr8	46594477	46595019	542	EL10Ac8g20030	0.72	0.58	40	Reverse transcriptase-like
Table	Chr8	46615848	46621153	5305	EL10Ac8g20031	0.81	0.57	170	Polyadenylate-binding protein RBP45
Table	Chr9	49350195	49352478	2283	EL10Ac9g22862	0.84	0.48	131	7-deoxyloganetic acid glucosyltransferase
Table	Chr9	49353611	49354759	1148	EL10Ac9g22863	0.83	0.39	67	hypothetical protein
Table	Chr9	49356165	49360589	4424	EL10Ac9g22864	0.71	0.51	181	Aspartate-semialdehyde dehydrogenase
Table	Chr9	49367357	49381163	13806	EL10Ac9g22866	0.70	0.43	234	Protein of unknown function (DUF3755)
Table	Chr9	49383136	49388434	5298	EL10Ac9g22867	0.73	0.52	183	TIMELESS-interacting protein
Table	Chr9	49400185	49404950	4765	EL10Ac9g22868	0.73	0.37	302	Reticulon-like protein B8
Table	Chr9	49408723	49421908	13185	EL10Ac9g22869	0.72	0.41	526	Chitobiosyldiphosphodolichol beta-mannosyltransferase
Table	Chr9	49425612	49431678	6066	EL10Ac9g22870	0.85	0.37	233	Uncharacterized oxidoreductase At

LITERATURE CITED

LITERATURE CITED

Adetunji, I., G. Willems, H. Tschoep, A. Bürkholz, S. Barnes *et al.*, 2014 Genetic diversity and linkage disequilibrium analysis in elite sugar beet breeding lines and wild beet accessions. Theor. Appl. Genet. 127: 559–571.

Akakpo, R., N. Scarcelli, H. Chaïr, A. Dansi, G. Djedatin *et al.*, 2017 Molecular basis of African yam domestication: analyses of selection point to root development, starch biosynthesis, and photosynthesis related genes. BMC Genomics 18: 782.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, 1990 Basic local alignment search tool. J. Mol. Biol. 215: 403–410.

Andrews, S., 2010 FastQC - A quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

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

Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler *et al.*, 2000 Gene ontology: Tool for the unification of biology. Nat. Genet. 25: 25–29.

Bergen, P., 1967 Seasonal patterns of sucrose accumulation and weight increase in sugar beets. J. Sugarbeet Res. 14: 538–545.

Bhatia, G., N. Patterson, S. Sankararaman, and A. L. Price, 2013 Estimating and interpreting FST: The impact of rare variants. Genome Res. 23: 1514–1521.

Biancardi, E., J. McGrath, P. L, L. R, and P. Stevanato, 2010 Sugar Beet, in *Root and Tuber Crops. Handbook of plant breeding, vol 7.*, edited by J. E. Bradshaw. Springer, New York, NY.

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

Bird, K. A., H. An, E. Gazave, M. A. Gore, J. C. Pires *et al.*, 2017 Population structure and phylogenetic relationships in a diverse panel of *Brassica rapa* L. Front. Plant Sci. 8:321

Bolger, A. M., M. Lohse, and B. Usadel, 2014 Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114–2120.

Carbon, S., E. Douglass, N. Dunn, B. Good, N. L. Harris *et al.*, 2019 The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res. 47: 330–338.

Carter, J. N., 1987 Sucrose production as affected by root yield and sucrose concentration of sugarbeet. J. Am. Soc. Sugar Beet Technol. 24: 14–31.

Chapman, E. J., and M. Estelle, 2009 Cytokinin and auxin intersection in root meristems. Genome Biol. 10: 210.

Chia, T. Y. P., A. Müller, C. Jung, and E. S. Mutasa-Göttgens, 2008 Sugar beet contains a large CONSTANS-LIKE gene family including a CO homologue that is independent of the early-bolting (B) gene locus. J. Exp. Bot. 59: 2735–2748.

Cooke, D. A., and R. K. Scott, 1993 *The Sugar Beet Crop*. Chapman and Hall Publishers, London.

Dally, N., M. Eckel, A. Batschauer, N. Höft, and C. Jung, 2018 Two CONSTANS-LIKE genes jointly control flowering time in beet. Sci. Rep. 8: 16120.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks *et al.*, 2011 The variant call format and VCFtools. Bioinformatics 27: 2156–2158.

Doebley, J. F., B. S. Gaut, and B. D. Smith, 2006 The molecular genetics of crop domestication. Cell 127: 1309–1321.

Dohm, J. C., A. E. Minoche, D. Holtgräwe, S. Capella-Gutiérrez, F. Zakrzewski *et al.*, 2014 The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). Nature 505: 546–549.

Ellison, S. L., C. H. Luby, K. E. Corak, K. M. Coe, D. Senalik *et al.*, 2018 Carotenoid presence is associated with the *Or* gene in domesticated carrot. Genetics 210: 1497–1508.

Ferretti, L., S. E. Ramos-Onsins, and M. Pérez-Enciso, 2013 Population genomics from pool sequencing. Mol. Ecol. 22: 5561–5576.

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

Gaj, P., N. Maryan, E. E. Hennig, J. K. Ledwon, A. Paziewska *et al.*, 2012 Pooled sample-based GWAS: A cost-effective alternative for identifying colorectal and prostate cancer risk variants in the Polish population (K. M. Lau, Ed.). PLoS One 7: e35307.

Gene Ontology Consortium, 2004 The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res. 32: 258–261.

Goldman, I. L., and J. P. Navazio, 2002 History and breeding of table beet in the United States. Plant Breed. Rev. 22: 357–388.

Hatlestad, G. J., N. A. Akhavan, R. M. Sunnadeniya, L. Elam, S. Cargile *et al.*, 2015 The beet Y locus encodes an anthocyanin MYB-like protein that activates the betalain red pigment pathway. Nat. Genet. 47: 92–96.

Hatlestad, G. J., R. M. Sunnadeniya, N. a Akhavan, A. Gonzalez, I. L. Goldman *et al.*, 2012 The beet R locus encodes a new cytochrome P450 required for red betalain production. Nat. Genet. 44: 816–820.

Hufford, M. B., P. Lubinksy, T. Pyhäjärvi, M. T. Devengenzo, N. C. Ellstrand *et al.*, 2013 The genomic signature of crop-wild introgression in maize. PLoS Genet. 9: e1003477.

Kim, N., Y.-M. Jeong, S. Jeong, G.-B. Kim, S. Baek *et al.*, 2016 Identification of candidate domestication regions in the radish genome based on high-depth resequencing analysis of 17 genotypes. Theor. Appl. Genet. 129: 1797–1814.

Kofler, R., A. J. Betancourt, and C. Schlötterer, 2012 Sequencing of pooled DNA samples (Pool-Seq) uncovers complex dynamics of transposable element insertions in *Drosophila melanogaster*. PLoS Genet. 8: e1002487.

Konishi, M., and M. Sugiyama, 2006 A novel plant-specific family gene, root primordium defective 1, is required for the maintenance of active cell proliferation. Plant Physiol. 140: 591–602.

Kraft, T., B. Fridlund, A. Hjerdin, T. Säll, S. Tuvesson *et al.*, 1997 Estimating genetic variation in sugar beets and wild beets using pools of individuals. Genome 40: 527–533.

Langmead, B., and S. L. Salzberg, 2012 Fast gapped-read alignment with Bowtie 2. Nat. Methods 9: 357–359.

Li, H., 2011 A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27: 2987–2993.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan *et al.*, 2009 The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078–2079.

Lin, Z., X. Li, L. M. Shannon, C. T. Yeh, M. L. Wang *et al.*, 2012 Parallel domestication of the *Shattering1* genes in cereals. Nat. Genet. 44: 720–724.

Lynch, M., 2009 Estimation of allele frequencies from high-coverage genome-sequencing projects. Genetics 182: 295–301.

Lynch, M., D. Bost, S. Wilson, T. Maruki, and S. Harrison, 2014 Population-genetic inference from pooled-sequencing data. Genome Biol. Evol. 6: 1210–1218.

Ma, T., K. Wang, Q. Hu, Z. Xi, D. Wan *et al.*, 2017 Ancient polymorphisms and divergence hitchhiking contribute to genomic islands of divergence within a poplar species complex. Proc. Natl. Acad. Sci. U. S. A. 115: E236–E243.

Macko-Podgórni, A., G. Machaj, K. Stelmach, D. Senalik, E. Grzebelus *et al.*, 2017 Characterization of a genomic region under selection in cultivated carrot (*Daucus carota* subsp. *sativus*) reveals a candidate domestication gene. Front. Plant Sci. 8:12.

Mangin, B., F. Sandron, K. Henry, B. Devaux, G. Willems *et al.*, 2015 Breeding patterns and cultivated beets origins by genetic diversity and linkage disequilibrium analyses. Theor. Appl. Genet. 128: 2255–2271.

Martin, S. H., K. K. Dasmahapatra, N. J. Nadeau, C. Salazar, J. R. Walters *et al.*, 2013 Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. Genome Res. 23: 1817–1828.

Maynard Smith, J., and J. Haigh, 2008 The hitch-hiking effect of a favourable gene. Genet. Res. (Camb). 89: 391–403.

McGrath, J. M., and L. Panella, 2018 Sugar beet breeding. Plant Breed. Rev. 167–218.

McGrath, J. M., C. A. Derrico, and Y. Yu, 1999 Genetic diversity in selected, historical US sugarbeet germplasm and *Beta vulgaris* ssp. *maritima*. Theor. Appl. Genet. 98: 968–976.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis *et al.*, 2010 The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20: 1297–1303.

Meirmans, P. G., and P. W. Hedrick, 2011 Assessing population structure: FST and related measures. Mol. Ecol. Resour. 11: 5–18.

Meyer, R. S., and M. D. Purugganan, 2013 Evolution of crop species: Genetics of domestication and diversification. Nat. Rev. Genet. 14: 840–852.

Milford, G. F. J., 1973 The growth and development of the taproot of sugar beet. Ann. Appl. Biol. 75: 427–438.

Muñoz-Rodríguez, P., T. Carruthers, J. R. I. Wood, B. R. M. Williams, K. Weitemier *et al.*, 2018 Reconciling conflicting phylogenies in the origin of sweet potato and dispersal to polynesia. Curr. Biol. 28: 1246-1256.e12.

Nei, M., 1987 *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nielsen, R., S. Williamson, Y. Kim, M. J. Hubisz, A. G. Clark *et al.*, 2005 Genomic scans for selective sweeps using SNP data. Genome Res. 15: 1566–1575.

Noguero, M., R. M. Atif, S. Ochatt, and R. D. Thompson, 2013 The role of the DNA-binding One Zinc Finger (DOF) transcription factor family in plants. Plant Sci. 209: 32–45.

Nosil, P., D. J. Funk, and D. Ortiz-Barrientos, 2009 Divergent selection and heterogeneous genomic divergence. Mol. Ecol. 18: 375–402.

Osborn, T. C., 2004 The contribution of polyploidy to variation in *Brassica* species. Physiol. Plant. 121: 531–536.

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

Owen, F., and G. Ryser, 1942 Some Mendelian characters in *Beta vulgaris* and linkages observed in the Y-R-B group. J. Agric. Res. 65: 155–171.

Paesold, S., D. Borchardt, T. Schmidt, and 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.

Perna, N. T., G. Plunkett, V. Burland, B. Mau, J. D. Glasner *et al.*, 2001 Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature 409: 529–533.

Pin, P. A., R. Benlloch, D. Bonnet, E. Wremerth-weich, T. Kraft *et al.*, 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 *et al.*, 2012 The role of a pseudo-response regulator gene in life cycle adaptation and domestication of beet. Curr. Biol. 22: 1095–1101.

Reif, J. C., W. Liu, M. Gowda, H. P. Maurer, J. Möhring *et al.*, 2010 Genetic basis of agronomically important traits in sugar beet (*Beta vulgaris* L.) investigated with joint linkage association mapping. Theor. Appl. Genet. 121: 1489–1499.

Rendón-Anaya, M., and A. Herrera-Estrella, 2018 The advantage of parallel selection of domestication genes to accelerate crop improvement. Genome Biol. 19:147.

Ries, D., D. Holtgräwe, P. Viehöver, and B. Weisshaar, 2016 Rapid gene identification in sugar beet using deep sequencing of DNA from phenotypic pools selected from breeding panels. BMC Genomics 17: 236.

Savage, M., N. I. Vavilov, and D. Love, 1994 Origin and Geography of Cultivated Plants. Geogr. Rev. 84: 231.

Schneider, K., R. Schäfer-Pregl, D. C. Borchardt, and F. Salamini, 2002 Mapping QTLs for sucrose content, yield and quality in a sugar beet population fingerprinted by EST-related markers. Theor. Appl. Genet. 104: 1107–1113.

Schreiber, M., N. Stein, and M. Mascher, 2018 Genomic approaches for studying crop evolution. Genome Biol. 19: 140.

Sigman, M. J., and R. K. Slotkin, 2015 The first rule of plant transposable element silencing: Location, location, location. Plant Cell 28: 304–313.

Smigocki, A. C., and L. D. Owens, 1988 Cytokinin gene fused with a strong promoter enhances shoot organogenesis and zeatin levels in transformed plant cells. Proc. Natl. Acad. Sci. 85: 5131–5135.

Smigocki, A. C., and L. D. Owens, 1989 Cytokinin-to-auxin ratios and morphology of shoots and tissues transformed by a chimeric isopentenyl transferase gene. Plant Physiol. 91: 808–811.

Stevens, P. F., 2012 Angiosperm Phylogeny Website. Version 12, July 2012.

Storz, J. F., 2005 Using genome scans of DNA polymorphism to infer adaptive population divergence. Mol. Ecol. 14: 671–688.

Takuno, S., P. Ralph, K. Swarts, R. J. Elshire, J. C. Glaubitz *et al.*, 2015 Independent molecular basis of convergent highland adaptation in maize. Genetics 200: 1297–1312.

Tanaka, M., Y. Takahata, H. Nakayama, M. Nakatani, and M. Tahara, 2009 Altered carbohydrate metabolism in the storage roots of sweetpotato plants overexpressing the SRF1 gene, which encodes a Dof zinc finger transcription factor. Planta 230: 737–746.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Weber, W. E., D. C. Borchardt, and G. Koch, 1999 Combined linkage maps and QTLs in sugar beet (*Beta vulgaris* L.) from different populations. Plant Breed. 118: 193–204.

Weir, B. S., and C. C. Cockerham, 1984 Estimating F-statistics for the analysis of population structure. Evolution (N. Y). 38: 1358–1370.

Weir, B. S., and W. G. Hill, 2002 Estimating F-Statistics. Annu. Rev. Genet. 36: 721–750.

Willemsen, V., M. Bauch, T. Bennett, A. Campilho, H. Wolkenfelt *et al.*, 2008 The NAC domain transcription factors *FEZ* and *SOMBRERO* control the orientation of cell division plane in arabidopsis root stem cells. Dev. Cell 15: 913–922.

Wright, S., 1943 Isolation by distance. Genetics 28: 114–138.

Wright, S., 1990 Evolution in mendelian populations. Genetics 16: 97–159.

Wright, S., 1951 The genetical structure of populations. Ann. Eugen. 15: 323–354.

Würschum, T., H. P. Maurer, T. Kraft, G. Janssen, C. Nilsson *et al.*, 2011 Genome-wide association mapping of agronomic traits in sugar beet. Theor. Appl. Genet. 123: 1121–1131.

Wyse, R., 1979 Parameters controlling sucrose content and yield of sugarbeet roots. J. Sugarbeet Res. 20: 368–384.

Zhang, H., P. Meltzer, and S. Davis, 2013 RCircos: An R package for Circos 2D track plots. BMC Bioinformatics 14: 244.

Zohary, D., and M. Hopf, 2013 Domestication of plants in the old world: the origin and spread of cultivated plants in West Asia, Europe, and the Nile Valley. Choice Rev. Online.

CHAPTER 3

ADMIXTURE AND INTROGRESSION IN THE DIVERSIFICATION OF *BETA VULGARIS* CROP TYPES

INTRODUCTION

The organization and content of *Beta vulgaris* crop type genomes reflect the demographic history and complex interactions between populations and crop type lineages. The crop types are classified on the basis of end use and include sugar beet, fodder beet, table beet, and chard. Relationships determined between *B. vulgaris* populations demonstrated a varying degree of support for crop types as discrete units. Cryptic relationships between lineages likely result from a complex evolutionary history (Chapter 1). Total genome differentiation was measured using F_{ST} and the variance in allele frequency within and between crop types showed a small proportion of the genome (~12%) was diverged with respect to crop type (Chapter 2). This suggested a relatively small proportion of the total genome variation underlies the different economic phenotypes observed between crop types. It also appeared that selection is likely the major driver of this differentiation. In order to describe the natural history of cultivated beet, the demographic history of the crop types, degree of genome divergence with respect to the crop type, and the magnitude of variation that is shared between crop types must be addressed. Ultimately, such explanations require a description of the standing genetic diversity of the species, crop type lineages and populations in the context of divergence (e.g. selection and drift) and coalescence (e.g., mutation, migration and common ancestry). This chapter specifically addresses the potential for pooled population sequencing to survey the later, specifically the effects of mutation, migration and common ancestry on the standing genetic diversity of beets.

Evidence for selective sweeps shared between crop types, specifically, those restricted to root and leaf types, prompted further inquiry into how variation is distributed among crop types, and

the effect of migration (e.g. admixture and introgression) in the development of important crop type characters. Beet improvement has largely focused on the improvement of root characters and not unexpectedly, a large number of candidate genes discovered were identified as orthologs to genes characterized within Arabidopsis root development pathways. These candidates may prove useful for understanding the genetic mechanisms underlying the unique biology of beet and more generally, root development and morphology in non-model species. The phenotypic diversity present in beet provides an opportunity to compare and contrast the genomes of phenotypically distinct lineages in order to identify genomic variation associated with traits of economic importance (e.g., root enlargement and biomass accumulation). Root morphology of Chard is similar to that of the wild progenitors of all beet types, *B. vulgaris* spp *maritima*, which is characterized as spangled, containing many lateral roots, and exhibits significantly less root enlargement compared to beet lineages cultivated for roots. These differences are likely influenced by a large genetic component as they breed true across environments (e.g. population phenotypes are reproducible), which provides a suitable contrast for comparative genomic approaches.

Admixture and introgressive hybridization are important processes that influence the diversity contained within a species. Migration and gene flow have the potential to introduce adaptive trait variation to distinct populations, lineages, and species at several orders of magnitude greater than mutation alone (Grant and Grant 1994). This directly influences the evolutionary trajectory of populations and the species. For example, specific trait variation identified in humans (*Homo sapiens*) shows DNA sequence evolution likely occurred in related hominid species (e.g., Neanderthals and Denisovans) and has been introgressed into the human genome as a source of

adaptive trait variation related disease resistance and human survival in extreme climates (Gittelman et al. 2016; Jeong et al. 2014). Admixture plays an important role in adaptive trait variation with respect to predator prey interactions across diverse geographic regions in *Heliconius* butterfly species (Martin et al. 2013). In poplar (*Populus* species), the extent and timing of gene flow has influenced the standing genetic diversity within phenotypically distinct lineages (Ma et al. 2018). Adaptive trait variation with respect to altitude in maize may help expand the range in which the crop can be cultivated (Hufford et al. 2013). Aromatic traits in cultivated rice have been suggested to result from admixture (Choi et al. 2017, Civáň et al. 2019). In fact, the majority of species we rely on for food, fuel, and fiber likely inherited important variation from antecedents versus *de novo* generation across short time scales such as crop domestication.

Recent research has highlighted the genetic cost associated with domestication including the loss of genetic diversity (Moyers et al. 2018). Modern breeding programs are interested in identifying and incorporating novel sources of variation can increase the rate of genetic gain for polygenic traits (e.g., yield, local adaptation, disease resistance) (Burgarella et al. 2019). In soybean (*Glycine max*), a population bottleneck resulting from domestication has been characterized and currently efficient strategies have been devised to incorporate genetic variation within specific genomic regions to ameliorate effects of negative trait linkages (Wang et al. 2019). A complete picture of the evolutionary history of a species requires testing the degree of admixture and introgression. To date, a litany of approaches can be found in population genetics literature which seek to estimate admixture and introgression. These include genealogy-based approaches, discordant phylogenies (Martin et al. 2013), F statistics (Wright 1951), and D statistics (Durand

et al. 2011), which serve to estimate the presence of shared derived alleles (Green et al. 2010). In *Heliconius* butterflies, introgression between closely related species has led to demonstrable effects on the complexity of genome variation between these species (Edelman et al., 2019).

Given the reproductive biology of beet (e.g., outcrossing, wind pollinated, self-incompatible and few barriers to reproduction between crop types), admixture and introgression likely occurred throughout the development of beet crop types given these lineages were not reproductively isolated (e.g., geographic separation, breeding methods, or asynchronous flowering). By exploring the evolutionary history of *Beta vulgaris* crop types, the importance of admixture and introgression was evident at local regions within the genome. This further suggests these regions contain important candidates. Furthermore, the origin of important candidate gene variation was explored, along with the putative effects these genes may have on the development of crop type phenotypes.

MATIERALS AND METHODS

Admixture, introgression and the origin of important variation

Population genetic parameters were used to test the evolutionary history of specific genomic regions. Diversity and divergence within and between *B. vulgaris* crop types was measured using gene diversity (*2pq*) and F_{ST} following the procedure outlined in Chapter 2. Correlations in allele frequency between populations and lineages (AF100) and relationship coefficients between populations and lineages (Rel100) were investigated in 100 kb bins across the genome. A bin size of 100 kb was large enough to visualize the variation within genomic regions at nucleotide resolution and scan regions of several Mb in size. Correlations in allele frequency were carried out using the cor() function in R (R Core Team 2013). Relationships coefficients were determined pairwise between each population using the Kinship Inference for Association Genetic Studies (KING) package (Manichaikul et al. 2010) detailed further in Chapter 1. Mean and standard deviation were calculated for each parameter using the empirical distribution of each parameter across the genome. This allowed comparisons between parameter estimates for local regions, containing specific candidate genes, and genome-wide estimates. Leveraging the information from all four parameters (e.g *2pq*, F_{ST}, AF100, Rel100), the evolutionary history of specific regions was examined.

Comparisons and evaluation standing genetic diversity

Comparisons within and between crop types were made by estimating parameters for individual crop types (CT) and by grouping crop types (e.g., [CT x CT], [CT x CT x CT] and [CT x CT x CT x CT]). This provided a picture of how variation is shared between lineages and the

significance of specific regions. Variation across the genome as well as variation within important candidate genes were categorized according to support for evolutionary hypothesis. These categories include, lineage-specific evolution (LSE), admixture and introgression (AI), and incomplete lineage sorting (ILS). The criterion for placement of genes into these categories was as follows:

1) Lineage-specific evolution (LSE) was defined as sequence variation with high probability for having evolved within independent crop type lineages. These regions appear unique to a lineage, contain significant F_{ST} values, high relationship coefficients (Rel100) within a crop type, and high correlation in allele frequency (AF100) within a crop type.

2) Admixture and introgression (AI) was defined as sequence variation with a high probability for having evolved independently and shared through admixture and introgression events. AI was evaluated by sites with low gene diversity (*2pq*) shared across two or more crop types, low F_{ST} values indicating little divergence between crop types, a high correlation in allele frequency between crop types, and significant relationship coefficients between two or more crop types, suggesting the origin of this variation may be the same.

3) Incomplete lineage sorting (ILS) refers to the segregation of polymorphism within ancestral populations. ILS was estimated using difference between total sites/regions and sites/regions characterized as lineage-specific evolution, and, admixture and introgression. There is a challenge in determination of old AI events and ILS as well as efficient ILS and LSE. This approach likely overestimates this category but with sufficient data, or different statistical tests, loci may be accurately placed within the LSE or AI categories.

RESULTS

Genome wide sequence diversity was used to describe how genetic diversity is distributed within and among crop types lineages. A population genomic dataset was generated for 23 beet populations representing a sample of the cultivated lineages of the species *B. vulgaris*. The parameters *2pq*, F_{ST}, relationship coefficients (Rel100), and correlations in allele frequency (AF100) were estimated across the whole genome and used to compare crop types and groups of crop types. Whole genome data (e.g., mean and standard deviation) for these parameters were used to determine significance of variation within local genome regions relative to genome-wide averages. Local regions were chosen on the basis of candidate genes previously identified as targets of selection, with potential roles in conditioning important economic and agronomic variation observed between beet crop type lineages (Chapter 2). Genome sequence data of representative beet populations was used to probe the evolutionary history of beet crop type lineages and to further define the role of admixture and introgression (AI), incomplete lineage sorting (ILS) and lineage specific evolution (LSE) in the development of these lineages. The complex distribution of variation within and between crop types is relevant to the origin of important genetic and phenotypic variation.

Variation in B. vulgaris genomes and the history of crop type lineages

The genetic variation detected within crop type genomes was used to estimate population genetic parameters (e.g., divergence [F_{ST}], diversity [*2pq*], relationships coefficients, and correlations in allele frequency). Using the aforementioned parameters, total genome variation was categorized as lineage-specific evolution (LSE), admixture and introgression (AI), and incomplete lineage

sorting (ILS). LSE with respect to crop type accounted for 2.3% (197074 bp) of the total variation. Putative AI between crop types accounted for 4.8% (410819 bp) of the total genome variation with respect to crop type, and ILS represented the majority of variation within crop type genomes, representing 92.8% (7853564 bp) of the total variation (Figure 3-1).

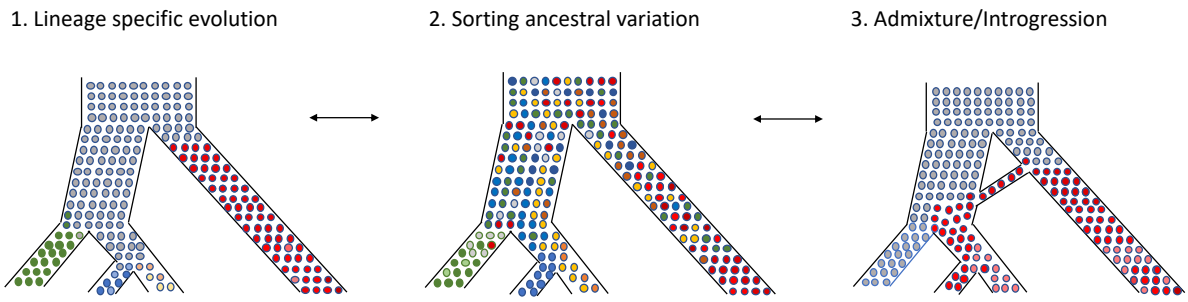

Figure 3-1. **Classification of standing genetic variation within *B. vulgaris* lineage genomes.** (1) Lineage-specific evolution (LSE), (2) Incomplete lineage sorting (ILS), and (3) Admixture and introgression (AI).

Common ancestry between crop type lineages was evident in the number of sites determined to be ILS as well as the mean values calculated for *2pq*, F_{ST}, allele frequency correlations (AF100) and relationship coefficients (Rel100) (Table 3-1). It is widely accepted that fodder and sugar crop types have a shared demographic history which was visible within comparisons of population genetic parameters measured. The number of shared sites with low diversity (*2pq*) was high. The level of divergence (F_{ST}) was the lowest between fodder and sugar crop types (F_{ST} = 0.31) relative to all other possible pairwise comparisons between crop types. This can be interpreted as a higher degree of connectivity or "gene flow" between specific crop types. Correlations in allele frequency estimates between crop type linages were the highest between sugar and fodder comparisons (R^2 = 0.57), suggesting a large degree of shared historical selection, which presumably occurred within a common ancestor. Mean relationship coefficients were the greatest between sugar and fodder lineages which indicates a larger quantity of shared variation between these lineages. Together, the parameters indicate signal related to the timing and extent of admixture between crop types is visible in this data. Fodder beet shared more variation with all the crop types suggesting fodder beet may be a less selected intermediate to other beet crop type lineages. Chard exhibited high diversity (*2pq*) contained within their genomes relative to other crop types. which indicates a greater likelihood of sharing variation by chance but this was not the case. Chard did not appear to share as much of this diversity with other crop types, rather this diversity appeared restricted within chard lineages. This suggests chard was historically isolated from other crop types. The data also supported table beet as the most diverged group with the lowest mean relationship coefficients observed between table and chard (0.072) and greatest level of divergence (F_{ST} = 0.39) observed between these two crop types.

Evolutionary history of root types involves admixture and introgression

The delineation of *B. vulgaris* crop types revealed relationships between and crop types and the degree to which genetic variation is shared between crop types. (Table 3-1). Two explanations for the degree of shared variation between crop types include 1) incomplete lineage sorting (ILS) and 2) admixture and introgression (AI) between populations whereby genetic variation is shared either by common ancestry or gene flow. The population genetic parameters estimated for all crop type linages showed that the root types (e.g., sugar beet, fodder beet, and table beet) shared more loci characterized as low diversity (*2pq*) than was expected given the distant relationships detected between these crop types. F_{ST} and correlations in allele frequency were used to highlight variation as same or different. This helped to characterize the evolutionary history of specific regions and classify the variation as LSE, AI or ILS. Discordance in clustering was observed between clusters constructed on the basis of local variation and those constructed on the basis of genome-wide variation. Differences between parameters estimated for genome-wide data and local regions is present in comparisons between Table 3-1 and Table 3-2 respectively. Local regions were chosen based the fact that they contain genes identified to be likely candidates with important functional roles in the evolutionary history of cultivated *B. vulgaris* (Chapter 2).

Patterns of gene diversity (*2pq*), divergence (F_{ST}) and what appeared to be shared selective sweeps restricted to the lineages which exhibit an enlarged root character (Chapter2). These patterns produced a list of candidates for further inquiry and include homeobox-leucine zipper protein ATHB-5 (EL10Ac4g09093), putative NAC domain-containing protein 94 (EL10Ac2g02976), cytokinin dehydrogenase 3 (EL10Ac8g19202), and ROOT PRIMORDIUM DEFECTIVE 1 (RPD1) (EL10Ac4g09126). The low diversity (*2pq*) of these regions, low F_{ST},

high correlations in allele frequency (100 Kb), and high relationship coefficients (100 Kb) observed between the root types (e.g., sugar, fodder, table) supports admixture and introgression in the evolutionary history of this variation and the enlarged root character. RPD1 and NAM/NAC (Table 3-2) contained the greatest signal for AI. The high relationship coefficients for these genes relative to genome-wide averages can be explained by a single origin for this variation.

Table 3-1 **Comparison of genome-wide variation.**

Group Comparison	Mean *2pq*	sd	Lower 95% CI (P < 0.05)	N (Bp) P < (0.05)	Mean F_{ST}	sd	Upper 95% CI (P = 0.05)	N (Bp)	Sig *2pq* & F_{ST}	(N) *2pq* - F_{ST}	Mean Relationship values (Rel100)	sd	R^2 Allele frequency (AF100)	sd
Sugar	0.259	0.096	0.102	304601	0.296	0.181	0.676	51734	9434	295167	0.197	0.046	0.712	0.101
Table	0.237	0.110	0.056	394153	0.351	0.196	0.774	57827	11869	382284	0.170	0.039	0.689	0.150
Fodder	0.246	0.093	0.093	248797	0.315	0.194	0.725	30559	4497	244300	0.330	0.041	0.824	0.080
Chard	0.277	0.087	0.133	375414	0.282	0.191	0.677	56954	7596	367818	0.246	0.045	0.741	0.110
Sugar Table	0.248	0.103	(0.102, 0.056)	37269	0.375	0.203	0.814	60533	2469	29673	0.076	0.040	0.449	0.172
Sugar Fodder	0.252	0.094	(0.102, 0.093)	117971	0.310	0.193	0.716	61107	1915	115502	0.128	0.053	0.575	0.148
Sugar Chard	0.268	0.091	(0.102, 0.133)	40258	0.343	0.199	0.767	65204	1348	38343	0.100	0.048	0.471	0.164
Table Fodder	0.257	0.099	(0.056, 0.093)	67984	0.365	0.205	0.808	65204	963	66636	0.091	0.046	0.504	0.173
Table Chard	0.257	0.099	(0.056, 0.133)	36205	0.390	0.214	0.854	61107	1776	35242	0.072	0.039	0.403	0.173
Fodder Chard	0.261	0.090	(0.093, 0.133)	49980	0.352	0.206	0.794	60533	647	48204	0.115	0.052	0.503	0.164
Sugar Table Fodder	0.247	0.100	(0.102, 0.056, 0.093)	32495	0.282	0.191	0.677	56954	1703	31848	0.136	0.034	0.578	0.113
Sugar Table Chard	0.252	0.094	(0.102, 0.056, 0.133)	1308	0.315	0.194	0.725	30559	88	1220	0.128	0.032	0.543	0.108
Sugar Fodder Chard	0.260	0.092	(0.102, 0.093, 0.133)	27638	0.351	0.196	0.774	57827	781	25935	0.160	0.039	0.602	0.106
Table Fodder Chard	0.253	0.097	(0.102, 0.093, 0.133)	18304	0.296	0.181	0.676	51734	267	18216	0.132	0.031	0.552	0.116
Sugar Table Fodder Chard	0.253	0.096	(0.102, 0.056, 0.093,0.133)	180	0.311	0.190	0.713	-	-	-	0.128	0.032	0.545	0.108

Standing genetic diversity in beets

Putative admixture events appear to have played a significant role in the development of beet crop types. Based on the functional annotations of genes with sequence variation classified as AI, the root types (e.g., sugar beet, fodder beet, and table beet) share variation which appears to condition lateral root formation, root expansion, and biomass accumulation. These traits are requisite to the development of an economically viable sugar crop. Additionally, a host of physiological changes (e.g., water content, dry matter content, and sucrose content) underlie the phenotypic differences between sugar beet and all other crop types. Similar to the analysis of root development genes described previously (e.g., RPD1, ATHB-5, and NAM/NAC), the same population genetic parameters used to compare averages of genome-wide variation with the variation residing within local regions. Local regions were chosen based on candidate genes with potential impact on important sugar beet characters. These genes include 6-phosphofructo-2-kinase (EL10Ac9g22391) and Brevis radix-like 4 (EL10Ac8g19137). Interestingly, these genes appeared to be important selection targets in sugar lineages but also appeared under selection in either chard and fodder, respectively. Functional annotations for these genes suggest putative involvement in sugar metabolism and root elongation. The variation in 6-phosphofructo-2-kinase (EL10Ac9g22391) exhibited low gene diversity (*2pq*) and low relationship coefficients between sugar and chard lineages relative to genome-wide averages. In addition to low gene diversity, a low correlation in allele frequencies between sugar and chard lineages within this region was observed. This suggests this gene is fixed for different alleles and indicates the selection history for these lineages was different and likely occurred independently within each lineage. A survey of standing genetic variation in Brevis radix-like 4 (EL10Ac8g19137) showed that a majority of sites with low diversity were shared, but some sites were unique to both sugar and fodder. No

significant divergence (F_{ST}) between sugar and fodder beets was observed and the average relationship coefficients suggest this variation results from ILS. Given the close relationships between sugar and fodder lineages, it is plausible that this variation is shared due to common ancestry and is identical by decent. The sequence variation within this gene, Brevis radix-like 4 (EL10Ac8g19137), likely results from drift and selection after the divergence of sugar and fodder lineages from a common ancestor.

Sugar beet specific genes, represented by genes classified as LSE, were confirmed by significant F_{ST} values when regions containing these genes were compared with all other crop types. The annotations associated with these genes were developmental and physiological in nature, which is consistent with phenotypic differences observed between sugar beet and the other crop types. A list of candidates that represent lineage-specific evolution with respect to sugar beet are detailed in Chapter 2. The annotations of these genes as well as experimental evidence in Arabidopsis point to divergence in root development and patterning of tissues (e.g., Dof zinc finger protein DOF5.6 [EL10Ac5g10742]), root physiology (e.g., probable trehalose-phosphate phosphatase D [EL10Ac1g01251], Glutamate receptor 2.7 [EL10Ac5g12159] and transcription factor bHLH041 [EL10Ac8g19192]). An extended region along Chromosome 3, likely represents a major determinant of sugar beet domestication. This region showed an interesting pattern of divergence and the region contained several hypothetical proteins, domains of unknown function and several functional elements including a gag-polypeptide of LTR copia-type (EL10Ac3g06339), and a lncRNA (EL10Ac3g06344) (Table 3-2).

Table 3-2 **Comparisons of local candidate gene variation.**

ROOT PRIMORDIUM DEFECTIVE 1 (RDP1) (EL10Ac4g09126)

Crop Type Comparison	Number of loci *2pq (p < 0.05)*	Number of loci F_{ST} *(p < 0.05)*	Mean Relationship (Rel100)	(R^2) Allele frequency (AF100)
Sugar	0	0	0.243	0.781
Table	0	0	0.172	0.935
Fodder	0	0	0.310	0.781
Chard	0	0	0.324	0.596
Sugar Table	0	0	0.140	0.766
Sugar Fodder	0	0	0.195	0.711
Sugar Chard	0	0	0.142	0.274
Table Fodder	35	0	0.116	0.703
Table Chard	0	0	0.079	0.170
Fodder Chard	0	0	0.133	0.249
Sugar Table Fodder	2	0	0.179	0.780
Sugar Table Chard	0	0	0.166	0.601
Sugar Fodder Chard	0	0	0.203	0.556
Table Fodder Chard	0	0	0.146	0.523
Sugar Table Fodder Chard	0	0	0.166	0.605

Transcription factor bHLH041 (EL10Ac8g19192)

Crop Type Comparison	Number of loci *2pq (p < 0.05)*	Number of loci F_{ST} *(p < 0.05)*	Mean Relationship (Rel100)	(R^2) Allele frequency (AF100)
Sugar	88	58	0.157	0.585
Table	0	0	0.110	0.907
Fodder	0	0	0.297	0.772
Chard	0	0	0.228	0.641
Sugar Table	60	0	0.027	0.065
Sugar Fodder	0	0	0.043	0.120
Sugar Chard	0	0	0.029	-0.005
Table Fodder	0	0	0.032	0.715
Table Chard	0	0	0.012	0.161
Fodder Chard	0	0	0.084	0.314
Sugar Table Fodder	0	0	0.080	0.386
Sugar Table Chard	0	0	0.072	0.290
Sugar Fodder Chard	0	0	0.101	0.310
Table Fodder Chard	0	0	0.078	0.527
Sugar Table Fodder Chard	0	0	0.069	0.305

Putative NAC domain-containing protein 94 (NAM/NAC) (EL10Ac2g02976)

Crop Type Comparison	Number of loci *2pq (p < 0.05)*	Number of loci F_{ST} *(p < 0.05)*	Mean Relationship (Rel100)	(R^2) Allele frequency (AF100)
Sugar	3	0	-	0.894
Table	0	0	0.165	0.486
Fodder	1	0	0.399	0.893
Chard	0	0	0.360	0.787
Sugar Table	0	0	0.125	0.536
Sugar Fodder	71	0	0.235	0.836
Sugar Chard	0	0	0.183	0.451
Table Fodder	0	0	0.171	0.552
Table Chard	0	0	0.139	0.332
Fodder Chard	0	0	0.253	0.466
Sugar Table Fodder	0	0	-	0.674
Sugar Table Chard	0	0	-	0.578
Sugar Fodder Chard	15	0	-	0.705
Table Fodder Chard	0	0	0.189	0.488
Sugar Table Fodder Chard	0	0	-	0.597

lncRNA (EL10Ac3g06344)

Crop Type Comparison	Number of loci *2pq (p < 0.05)*	Number of loci F_{ST} *(p < 0.05)*	Mean Relationship (Rel100)	(R^2) Allele frequency (AF100)
Sugar	1	96	-	0.552
Table	19	0	0.153	0.958
Fodder	0	0	0.314	0.866
Chard	0	0	0.278	0.944
Sugar Table	4	0	0.007	0.113
Sugar Fodder	0	0	0.014	0.173
Sugar Chard	0	0	0.022	0.095
Table Fodder	0	0	0.108	0.831
Table Chard	55	0	0.053	0.845
Fodder Chard	0	0	0.114	0.749
Sugar Table Fodder	0	0	-	0.419
Sugar Table Chard	0	0	-	0.426
Sugar Fodder Chard	0	0	-	0.393
Table Fodder Chard	0	0	0.125	0.872
Sugar Table Fodder Chard	0	0	-	0.443

Cytokinin dehydrogenase 3 (EL10Ac8g19202)

Crop Type Comparison	Number of loci *2pq (p < 0.05)*	Number of loci F_{ST} *(p < 0.05)*	Mean Relationship (Rel100)	(R^2) Allele frequency (AF100)
Sugar	0	0	-	0.559
Table	45	0	0.147	0.679
Fodder	3	0	0.321	0.784
Chard	0	0	0.211	0.652
Sugar Table	18	0	0.040	0.385
Sugar Fodder	0	0	0.062	0.250
Sugar Chard	0	0	0.048	0.079
Table Fodder	42	0	0.043	0.387
Table Chard	1	0	0.024	0.214
Fodder Chard	0	0	0.117	0.475
Sugar Table Fodder	14	0	-	0.459
Sugar Table Chard	0	0	-	0.381
Sugar Fodder Chard	0	0	-	0.361
Table Fodder Chard	0	0	0.096	0.452
Sugar Table Fodder Chard	0	0	-	0.378

Probable trehalose-phosphate phosphatase D (EL10Ac1g01251)

Crop Type Comparison	Number of loci *2pq (p < 0.05)*	Number of loci F_{ST} *(p < 0.05)*	Mean Relationship (Rel100)	(R^2) Allele frequency (AF100)
Sugar	0	7	-	0.472
Table	0	0	0.198	0.866
Fodder	0	0	0.318	0.824
Chard	0	0	0.268	0.782
Sugar Table	0	0	0.051	0.243
Sugar Fodder	0	0	0.071	0.309
Sugar Chard	0	0	0.045	0.327
Table Fodder	0	0	0.076	0.746
Table Chard	0	0	0.042	0.148
Fodder Chard	0	0	0.102	0.332
Sugar Table Fodder	0	0	-	0.443
Sugar Table Chard	0	0	-	0.384
Sugar Fodder Chard	0	0	-	0.418
Table Fodder Chard	0	0	0.126	0.537
Sugar Table Fodder Chard	0	0	-	0.400

Table 3-2 (cont'd)

Dof zinc finger protien DOF5.6 (EL10Ac5g10742)

Crop Type Comparison	Number of loci 2pq (p < 0.05)	Number of loci FST (p < 0.05)	Mean Relationship (Rel100)	(R2) Allele frequency (AF100)
Sugar	0	0	0.158	0.628
Table	2	0	0.135	0.799
Fodder	0	0	0.283	0.838
Chard	0	0	0.225	0.710
Sugar Table	0	0	0.047	0.226
Sugar Fodder	0	0	0.100	0.475
Sugar Chard	0	0	0.079	0.439
Table Fodder	0	0	0.065	0.575
Table Chard	0	0	0.036	0.214
Fodder Chard	0	0	0.086	0.535
Sugar Table Fodder	0	0	0.100	0.478
Sugar Table Chard	0	0	0.093	0.436
Sugar Fodder Chard	0	0	0.126	0.546
Table Fodder Chard	0	0	0.098	0.528
Sugar Table Fodder Chard	0	0	0.093	0.454

Glutamate receptor 2.7 (EL10Ac5g12159)

Crop Type Comparison	Number of loci *2pq (p < 0.05)*	Number of loci F_{ST} *(p < 0.05)*	Mean Relationship (Rel100)	(R^2) Allele frequency (AF100)
Sugar	0	26	0.184	0.612
Table	16	0	0.160	0.950
Fodder	0	0	0.272	0.736
Chard	26	0	0.192	0.937
Sugar Table	0	0	0.042	0.514
Sugar Fodder	0	0	0.093	0.435
Sugar Chard	0	0	0.040	0.212
Table Fodder	0	0	0.052	0.264
Table Chard	8	0	0.034	-0.134
Fodder Chard	0	0	0.070	0.649
Sugar Table Fodder	0	0	0.106	0.568
Sugar Table Chard	0	0	0.091	0.452
Sugar Fodder Chard	0	0	0.120	0.483
Table Fodder Chard	0	0	0.097	0.434
Sugar Table Fodder Chard	0	0	0.090	0.450

6-phosphofructo-2-kinase (EL10Ac9g22391)

Crop Type Comparison	Number of loci 2pq (p < 0.05)	Number of loci FST (p < 0.05)	Mean Relationship (Rel100)	(R2) Allele frequency (AF100)
Sugar	130	0	-	0.716
Table	0	0	0.150	0.497
Fodder	0	0	0.279	0.610
Chard	0	0	0.332	0.887
Sugar Table	0	0	0.017	0.385
Sugar Fodder	0	0	0.005	0.423
Sugar Chard	289	0	0.005	0.227
Table Fodder	0	0	0.097	0.404
Table Chard	0	0	0.132	0.499
Fodder Chard	0	0	0.099	0.402
Sugar Table Fodder	0	0	-	0.501
Sugar Table Chard	0	0	-	0.476
Sugar Fodder Chard	40	0	-	0.500
Table Fodder Chard	0	0	0.151	0.518
Sugar Table Fodder Chard	0	0	-	0.466

Homeobox-leucine zipper protein ATHB-5 (EL10Ac4g09093)

Crop Type Comparison	Number of loci *2pq (p < 0.05)*	Number of loci F_{ST} *(p < 0.05)*	Mean Relationship (Rel100)	(R^2) Allele frequency (AF100)
Sugar	0	0	0.170	0.898
Table	21	0	0.126	0.969
Fodder	11	0	0.297	0.941
Chard	0	0	0.236	0.802
Sugar Table	7	0	0.085	0.872
Sugar Fodder	0	0	0.140	0.895
Sugar Chard	0	0	0.129	0.696
Table Fodder	46	0	0.084	0.911
Table Chard	0	0	0.073	0.623
Fodder Chard	0	0	0.123	0.684
Sugar Table Fodder	0	0	0.122	0.898
Sugar Table Chard	0	0	0.121	0.816
Sugar Fodder Chard	0	0	0.156	0.811
Table Fodder Chard	0	0	0.116	0.793
Sugar Table Fodder Chard	0	0	0.122	0.824

Brevis radix-like 4 (EL10Ac8g19137)

Crop Type Comparison	Number of loci *2pq (p < 0.05)*	Number of loci F_{ST} *(p < 0.05)*	Mean Relationship (Rel100)	(R^2) Allele frequency (AF100)
Sugar	13	0	0.176	0.547
Table	0	0	0.127	0.500
Fodder	19	0	0.273	0.663
Chard	0	0	0.219	0.693
Sugar Table	0	0	0.067	0.270
Sugar Fodder	167	0	0.147	0.409
Sugar Chard	0	0	0.036	0.214
Table Fodder	0	0	0.077	0.264
Table Chard	0	0	0.051	0.323
Fodder Chard	0	0	0.033	0.330
Sugar Table Fodder	0	0	0.117	0.398
Sugar Table Chard	0	0	0.096	0.370
Sugar Fodder Chard	0	0	0.123	0.415
Table Fodder Chard	0	0	0.097	0.409
Sugar Table Fodder Chard	0	0	0.098	0.368

gag-polypeptide of LTR copia-type (EL10Ac3g06339)

Crop Type Comparison	Number of loci *2pq (p < 0.05)*	Number of loci F_{ST} *(p < 0.05)*	Mean Relationship (Rel100)	(R^2) Allele frequency (AF100)
Sugar	0	30	-	0.593
Table	5	0	0.127	0.969
Fodder	0	0	0.297	0.866
Chard	0	0	0.258	0.958
Sugar Table	0	0	0.008	0.087
Sugar Fodder	0	0	0.025	0.177
Sugar Chard	0	0	0.035	0.139
Table Fodder	0	0	0.038	0.835
Table Chard	0	0	0.048	0.923
Fodder Chard	0	0	0.145	0.803
Sugar Table Fodder	0	0	-	0.423
Sugar Table Chard	0	0	-	0.449
Sugar Fodder Chard	0	0	-	0.427
Table Fodder Chard	0	0	0.104	0.909
Sugar Table Fodder Chard	0	0	-	0.464

DISCUSSION

The high quantity of shared variation between crop types (Chapter 1), as well as the low degree of total genome divergence between *B. vulgaris* crop types (Chapter 2) can be explained by ILS (e.g. the segregation of ancestral variation) and by historical admixture and introgression events between crop types. Both genome-wide variation and variation within local regions, containing gene candidates of interest, were used to classify specific variation and test hypotheses of LSE, ILS and AI, and further explore the roles of specific genes which likely influenced phenotypic evolution across cultivated *B. vulgaris* lineages. Sugar beet represents the most economically important crop type and, to date, lacks molecular genetic explanations for the vast majority of important traits. We limited the scope of this discussion to genome variation that appeared important to the development of sugar beet for this reason. However, many characteristics that make sugar beet a successful crop appear shared among beet crop types, complicating simple explanations. Thus, understanding how genetic and phenotypic diversity is distributed among beet lineages provides the necessary information to group populations and lineages to compare crop type variation and in doing so provide the contrast to describe the unique nature of sugar beet.

Understanding the timing of crop type diversification and divergence is important because it reflects the potential for gene flow between crop type lineages, which serves to obfuscate evolutionary history, homogenize genome variation, and produce cryptic relationships. Historical accounts suggest chard was the first crop type selected (Ford-Lloyd et al. 1975), followed by table beet and fodder beet (Biancardi et al. 2012). Sugar beet was developed from fodder lineages in the last ~200 years (Fischer 1989) and was evident in the genetic data. The

development of distinct crop types appears to coincide with the accumulation of important variation across time. Understanding how variation is accumulated and retained within linages (e.g., lineage-specific evolution, sorting of ancestral variation, and, admixture and introgression) can help explain the origin of important variation, identify potential sources of novel genetic and capture phenotypic variation for traits critical to future productivity an sustainable production of the crop.

Low genetic diversity (*2pq*) within specific genomic regions was indicative of selective sweeps across the genome. In specific cases these regions were shared across all lineages exhibiting a trait. The enlarged root character represents on such trait and the regions identified contained genes with potential for influencing root enlargement. These genes include RPD1 (EL10Ac4g09126), homeobox-leucine zipper protein ATHB-5 (EL10Ac4g09093), NAM/NAC (EL10Ac2g02976), and cytokinin dehydrogenase 3 (EL10Ac8g19202). F_{ST} values for these genes suggested little divergence between root types. High relationship coefficients, discordance in genome-wide versus local trees, and strong correlations in allele frequency for the region surrounding these genes hint at a single origin for this variation. Selection for genetic variation within and around these genes may have occurred within a single lineage and was subsequently shared through admixture and introgression. Results did not indicate the direction or origin of important variation but do indicate regions in the genome where phased haplotype data would be useful. Orthologs of these genes have been functionally characterized in Arabidopsis and found to affect root growth and development. Additionally, these genes were recovered as differentially expressed in maize roots and shoots (Hwang et al. 2018), supporting the function of these candidates in root development. These results suggest a large degree of conservation of these

developmental genetic pathways between phylogenetically distant taxa. The mechanisms responsible for root enlargement in beet may not be unique to beet. In fact, enlargement and growth by successive cambia is reported as a pervasive character in the Caryophyllales (Carlquist 2010). Beet may have exploited this mechanism characteristic of the order for root enlargement. Using the variability that exists between the root type and leaf types as a comparison, uncovered several genes that may influence this character and may explain potential mechanisms of biomass accumulation in beet and more broadly, species within the order Caryophyllales.

Admixture and introgression accounted for a small proportion (4.8%) of the genome but appears to be an important feature in the evolution of beet and the development of important phenotypic variation such as an enlarged root. Another root development gene identified by a potential selective sweep observed between sugar and fodder beet was the protein coding gene brevis radix like 4 (EL10Ac8g19137). Given the close relationships of fodder and sugar lineages and the quantity of shared variation within this gene, this variation likely results from common ancestry and may explain some of the shared root morphology between sugar and fodder lineages. Signals for admixture were clear if the underlying variation was fixed, owing to the observation that the majority of variation was segregating between crop type genomes (92.8%). This suggests ILS is the major determinant of standing genetic variation between crop types. Our estimate of AI was likely biased toward important variation that was fixed as a result of selection and provided a clear signal. Pooled data leverages allele frequency versus sequence evolution that is common in haplotype-based approaches for the determination of admixture. Although biased, this approach detected some important events in the development of *B. vulgaris* crop

types. Without representative ancestral populations the distinction between old admixture and ILS will remain a challenge. The difference between efficient sorting versus lineage-specific also presents a challenge. Further sampling of beet populations, historical and current, as well as, haplotype level data will be needed to further classify genome variation to accurately characterize the evolutionary history of crop type genomes.

Both developmental and physiological traits were required for the development of sugar beet (e.g. lineages with the agronomic potential to accumulate large quantities of sucrose). Root enlargement appears underlie the agronomic potential for sucrose accumulation but is not mutually exclusive to the physiological changes associated with differences in carbohydrate metabolism and source sink relationships observed between crop types. A list of interesting candidate genes detected as diverged with respect to crop type (Chapter 2) could largely be categorized as developmental and physiological in nature. The identity of these genes implicates their role in pathways with the potential to alter physiological properties of the root. One gene of interest due to its role in cellular carbohydrate metabolism is 6-phosphofructo-2-kinase (EL10Ac9g22391). The region containing this gene appeared selected in both sugar and chard lineages due to the lack of genetic diversity (*2pq*) but the variation did not appear the same suggesting the region was fixed for different alleles as a result of divergent selection, which likely occurred independently within both lineages.

Lineage-specific evolution in beet accounted for 2.3% of the genome. The low degree of lineage-specific variation and divergence between independent lineages (crop types) is consistent with the time (4000–8000 years) since beets were derived from wild progenitors of *B. vulgaris* ssp

maritima. The development of novel crop types terminated with the development of sugar beet, which was largely accomplished though progeny selection (Gayon and Zallen 1998). In total, 16 genes were identified, which correspond to the selection of sugar beet, genetic bottlenecks, and the reduction of diversity at specific regions which explain the genetic and phenotypic divergence of sugar beet relative to other crop types. Sugar beet genomes represent cultivated *B. vulgaris* lineages optimized for these developmental and physiological traits, especially those related to sucrose accumulation. Selection for these traits and the reduced diversity as a result of genetic bottlenecks may have produced negative linkages between important traits such as those seen between yield and sucrose content (Boesmark 2006). Some studies suggest limitations on yield have been reached (CITE). If the genes and genomic regions influencing these characters, were known, experimental strategies could be devised to validate and potentially break these linkages. The following genes were confirmed to result of LSE (e.g., contain high divergence (F_{ST}) and unique variation) and may affect physiological features of sugar beet roots: Trehalose 6-phosphate (EL10Ac1g01251), transcription factor bHLH041 (EL10Ac8g19192) and a glutamate receptor (EL10Ac5g12159). Chromosome 3 showed a large degree of differentiation between sugar beet and all other crop types. We evaluated the most significant genes (e.g., lncRNA [EL10Ac3g06344], LTR associated gag-polypeptide [EL10Ac3g06339]) and confirmed they likely arose from LSE. How these genes function with respect to the unique phenotypic diversity of sugar lineages is of considerable interest.

In conclusion, much of the genetic variation available to plant breeders results from mutation across large evolutionary time scales. The potential for genetic variation and thus traits to be shared between diverged populations by admixture and migration is orders of magnitude greater

than mutation alone (Grant and Grant 1992). The variation contained within lineages and sub-populations represents the evolutionary potential of the species. Understanding how the standing genetic variation in modern populations is derived from variation segregating within ancestral populations is complex but an important feature of crop evolution and improvement (Stetter et al. 2018). Selection experiments are a means to uncover adaptive trait variation and to use these strategies to uncover the genetic mechanisms underlying adaptation in an agricultural setting has been proposed (Ross-Ibarra et al. 2007). Leveraging pooled data has many advantages, such as species with variable ploidy, species that are a challenge to isolate and maintain a single individual for sequencing and analysis, and species where populations are the evolutionary unit of improvement. Considering that the success of agriculture depends on adaptation to novel growing environments, understanding the diversity of a species through dissecting the evolutionary history of important lineages, targets of historical selection within the genome, and the mechanisms of polygenic adaptation will help integrate genomics into the decision-making process of crop improvement.

LITERATURE CITED

LITERATURE CITED

Burgarella, C., A. Barnaud, N. A. Kane, F. Jankowski, N. Scarcelli *et al.*, 2019 Adaptive introgression: An untapped evolutionary mechanism for crop adaptation. Front. Plant Sci. 10: 1–17.

Carlquist, S., 2010 Caryophyllales: a key group for understanding wood anatomy character states and their evolution oj_1095 342..393. Bot. J. Linn. Soc. 164: 342–393.

Choi, J. Y., A. E. Platts, D. Q. Fuller, Y. I. Hsing, R. A. Wing *et al.*, 2017 The rice paradox: Multiple origins but single domestication in Asian Rice. Mol. Biol. Evol. 34: 969–979.

Civáň, P., S. Ali, R. Batista-Navarro, K. Drosou, C. Ihejieto *et al.*, 2019 Origin of the aromatic group of cultivated rice (*Oryza sativa* L.) traced to the indian subcontinent. Genome Biol. Evol. 11: 832–843.

Durand, E. Y., N. Patterson, D. Reich, and M. Slatkin, 2011 Testing for ancient admixture between closely related populations. Mol. Biol. Evol. 28: 2239–2252.

Edelman, N.B., Frandsen, P.B., Miyagi, M., Clavijo, B., Davey, J., Dikow, R.B., García-Accinelli, G., Van Belleghem, S.M., Patterson, N., Neafsey, D.E., et al. (2019). Genomic architecture and introgression shape a butterfly radiation. Science 366, 594–599.

Ford-Lloyd, B. V., and J. T. Williams, 1975 A revision of *Beta* section *Vulgares* (Chenopodiaceae), with new light on the origin of cultivated beets. Bot. J. Linn. Soc. 71: 89–102.

Gittelman, R. M., J. G. Schraiber, B. Vernot, C. Mikacenic, M. M. Wurfel *et al.*, 2016 Archaic hominin admixture facilitated adaptation to out-of-africa environments. Curr. Biol. 26: 3375–3382.

Grant, P. R., and B. R. Grant, 2008 Hybridization of bird species. Science 256: 193–197.

Green, R. E., J. Krause, A. W. Briggs, T. Maricic, U. Stenzel *et al.*, 2010 A draft sequence of the neandertal genome. Science. 328: 710–722.

Hufford, M. B., P. Lubinksy, T. Pyhäjärvi, M. T. Devengenzo, N. C. Ellstrand *et al.*, 2013 The genomic signature of crop-wild introgression in maize. PloS Genet. 9: e1003477.

Hwang, S.-G., K.-H. Kim, B.-M. Lee, and J.-C. Moon, 2018 Transcriptome analysis for identifying possible gene regulations during maize root emergence and formation at the initial growth stage. Genes Genomics 40: 755–766.

Jeong, C., G. Alkorta-Aranburu, B. Basnyat, M. Neupane, D. B. Witonsky *et al.*, 2014 Admixture facilitates genetic adaptations to high altitude in Tibet. Nat. Commun. 5: 1–7.

Ma, T., K. Wang, Q. Hu, Z. Xi, D. Wan *et al.*, 2017 Ancient polymorphisms and divergence hitchhiking contribute to genomic islands of divergence within a poplar species complex. Proc. Natl. Acad. Sci. U. S. A. 115: E236–E243.

Maherali, H., 2017 The evolutionary ecology of roots. New Phytol. 215: 1295–1297.

Manichaikul, A., J. C. Mychaleckyj, S. S. Rich, K. Daly, M. Sale *et al.*, 2010 Robust relationship inference in genome-wide association studies. Bioinformatics 26: 2867-2873.

Martin, S. H., K. K. Dasmahapatra, N. J. Nadeau, C. Salazar, J. R. Walters *et al.*, 2013 Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. Genome Res. 23: 1817–1828.

Moyers, B. T., P. L. Morrell, and J. K. McKay, 2018 Genetic costs of domestication and improvement. J. Hered. 109: 103–116.

Ross-ibarra, J., P. L. Morrell, and B. S. Gaut, 2007 Ross-Ibarra, et al. PNAS May 15, 2007 vol. 104 suppl. 1 8641–8648.pdf. Proc. Natl. Acad. Sci. 104: 8641–8648.

Stetter, M. G., K. Thornton, and J. Ross-Ibarra, 2018 Genetic architecture and selective sweeps after polygenic adaptation to distant trait optima. PLOS Genet. 14: e1007794.

Wang, X., L. Chen, and J. Ma, 2019 Genomic introgression through interspecific hybridization counteracts genetic bottleneck during soybean domestication. Genome Biol. 20: 22.

Wright, S., 1951 The genetical structure of populations. Ann. Eugen. 15: 323–354.

CONCLUSIONS

Pooled sequencing offered an effective strategy for measuring genetic diversity in cultivated *B. vulgaris.* This research supports the idea that cultivated *B. vulgaris* lineages "crop types" represent a species complex (Fénart et al. 2008). The effectiveness of pooled population sequencing to inform the evolutionary history of beet crop types can be explained by how the genetic diversity is held in the sub populations that compose the species. This is influenced by the reproductive biology of the species and the effects phenotypic selection has on the variation contained within the genome. Pooled sequencing has the ability to measure the enrichment of beneficial alleles associated with selection for characters which define crop type end use. The high degree of diversity and outcrossing nature of beet produced clear signals related to the diversification of the species into distinct cultivated forms (e.g. crop types). The availability of a complete and contiguous genome sequence coupled with WGS of pooled populations was effective for the identification of important regions and underlying genes at nucleotide resolution. Pooled sequencing offers an effective means to estimate genetic diversity in beet and other outcrossing species where the genetic potential for important traits is contained within populations (e.g., crop wild relatives (CWR), in-situ populations, core collections, breeding programs). As a consequence of the species reproductive biology (e.g. self-incompatibility), the advancement of materials occurs as a population because it is a challenge to maintain a single individual or inbred line. The method could inform other species with variable ploidy and for species where a single individual is a challenge to isolate or study *in-situ* (e.g. bacteria and fungi). Population level data better represents the genomic diversity within populations and linages because it not only reflects the genetic variation of the generation measured but can also

estimate its future derivatives. Phenotypic diversity in beet is evaluated in the field as populations, often reported as plot averages. Measuring phenotypic diversity is important but limited by resource constraints. The number of individuals per pool is an important consideration. In beet, twenty-five individuals represent a total of fifty parental gametes and is roughly the number of individuals contained within a field plot aimed at screening functional diversity. This suggests pooled sequencing can provide a genomics perspective to field-based research and aid in beet improvement. This research attempts to address several fundamental questions. How well are the crop types supported from a genomics perspective? What variation in the genome explains crop type differentiation and what appear to be the major evolutionary forces behind this diversification? What factors explain complex distribution of genome variation and complex relationships observed between crop type lineages?

How well are the crop types supported from a genomics perspective?

Beet crop types represent important lineages which exhibit pronounced genetic and phenotypic divergence. Support these groups as significant biological units was observed on the basis of *de novo* clustering of pooled populations using both allele frequency estimates and quantity of shared variation (e.g. pairwise relationship coefficients). It appears that selection for end use qualities and genetic drift were major factors in the divergence between crop type lineages and explains the apportionment of genetic variation between crop types. This divergence was visible at the genome-wide level as well as at distinct chromosome locations. Common ancestry and, admixture and introgression likely maintained levels of genetic variation between crop types and suggests a complex demographic history between crop types. The majority of genetic variation detected in beet crop types were biallelic SNPs, but lineage specific variation, including indels

and structural variants may have had a greater role in crop diversification with table beet showing the greatest degree of differentiation. The majority of variation is held within the species, shared among crop type lineages, and only a small amount of the total variation was partitioned within individual crop types.

What variation in the genome explains crop type differentiation and what appear to be the major evolutionary forces behind this diversification?

Chapter 2 further explored the delineation of the species based on genome-wide data, specifically by measuring the degree of differentiation along chromosomes with respect to crop type. We found specific chromosomes had a greater ability to differentiate the crop types. Specific regions along chromosomes contained genes that were associated with these signals. An average of 3.03% of crop type genomes were diverged ($F_{ST} > 0.6$) and the total degree of divergence between crop types detected was 12.13%. The levels of divergence estimated in beet correspond to those found within incipient speciation literature. On average, between 5 and 10% of the genome were found to be differentiated for species involved in recent speciation events (Nosil et al. 2009). Differentiated regions with respect to crop type contained 472 genes, or 1.6% of the 24,255 genes predicted in the reference genome assembly. Respectively, sugar beet, table beet, fodder beet, and chard genomes contained 16, 283, 2, and 171 genes characterized as differentiated. Interestingly**,** SNP and indel LSV was concentrated in regions of significant F_{ST}, further supporting the importance of these regions to crop diversification. The annotations associated with genes determined to be diverged with respect to crop type suggest they may play functional roles in the morphological and physiological differences observed between crop types.

What factors explain complex distribution of genome variation and complex relationships observed between crop type lineages?

Relationships between crop types were determined in Chapter 1 and supported the crop types as discrete units, yet the majority of the genetic variation was detected to be shared between crop type lineages. Furthermore, the parameters F_{ST} and *2pq* were used to investigate variation in allele frequency within genomes of *B. vulgaris* crop types. These parameters, determined across set distances, were used to describe putative locations within the genome where divergence has occurred, highlighting specific genomic variation, which explain these relationships and may influence the phenotypic variation associated with end use. A relatively small proportion of the genome was diverged with respect to crop type, indicating a need to quantify the degree of shared variation in order to understand the evolutionary history of beet. The four parameters (*2pq*, F_{ST}, relationship coefficients and allele frequency correlations) were used to characterize the standing genome variation within crop type lineages. Furthermore, these parameters were used to test the evolutionary history of beet by characterizing genome variation as having resulted from admixture and introgression (AI), incomplete lineage sorting (ILS) or lineage specific evolution (LSE). Several regions within the genome appeared to be the result of selective sweeps which were shared between crop types. As an example, one such region was restricted to the root types and indicates potential genomic variation involved in conditioning the enlarged root phenotype. Candidate gene variation involved in root enlargement supported a hypothesis of admixture and introgression development of this character versus convergence. The genes were identified as ROOT PRIMORDIUM DEFECTIVE 1 (RPD1) (EL10Ac4g09126) and putative NAC domain-containing protein 94 (NAM/NAC) (EL10Ac2g02976). The high similarity of this variation suggests a single origin of the enlarged root character. Specific

instances of common ancestry and sorting of ancestral variation were also identified which helped explain the degree of divergence observed between specific crop types. Based on functional annotations, the gene Brevis radix-like 4 (EL10Ac8g19137) is suggested to control quantitative aspects of root growth, specifically root elongation. This variation appeared shared between fodder and sugar lineages. Due to the degree of common ancestry between these lineages, this variation likely represents identity by decent (IBD) and may be reflected in similar root phenotypes.

Understanding the evolutionary history of beet crop types through measuring heterogenous genome differentiation and the corresponding divergence of phenotypes may help to identify and recover a genetic basis for phenotypes of economic and agronomic interest. Genetic data for these groups as discrete biologically relevant units and allowed for the identification of specific variation with a high probability of conditioning important phenotypes. In fact, a handful of genes were identified which represent putative targets in the domestication of sugar beet. Shared genome variation among crop types was another feature that proved useful for understanding important traits due to the fact *B. vulgaris* crop type lineages appear to have a complex evolutionary history.

LITURATURE CITED

LITURATURE CITED

Fénart, S., J. F. Arnaud, I. De Cauwer, and J. Cuguen, 2008 Nuclear and cytoplasmic genetic diversity in weed beet and sugar beet accessions compared to wild relatives: New insights into the genetic relationships within the Beta vulgaris complex species. Theor. Appl. Genet.

Nosil, P., D. J. Funk, and D. Ortiz-Barrientos, 2009 Divergent selection and heterogeneous genomic divergence. Mol. Ecol.