2014-06-01

Characterization of the Granule-Bound Starch Synthase I Gene in Chenopodium

Douglass C. Brown
Brigham Young University - Provo

Characterization of the *Granule-Bound Starch Synthase I* Gene in *Chenopodium*

Douglass C. Brown

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Eric Nathan Jellen, Chair
Peter J. Maughan
Clinton Jeremy Whipple

Department of Plant and Wildlife Sciences

Brigham Young University

June 2014

ABSTRACT

Characterization of the *Granule-Bound Starch Synthase I* Gene in *Chenopodium*

Douglass C. Brown
Department of Plant and Wildlife Sciences, BYU
Master of Science

Chenopodium L. is a relatively under-studied genus that includes the cultivated seed crop quinoa (*Chenopodium quinoa* Willd.). Quinoa is an allotetraploid (2*n*=4*x*=36, AABB genomes) that is cultivated by subsistence farmers and commercial growers in the Andean regions of South America. Approximately 60% of a quinoa seed is starch, a glucose polymer that is an important carbohydrate energy source in the human diet. Seed starch is normally comprised of amylose and amylopectin in a 1:3 ratio, but starches with different amylose:amylopectin ratios have different properties and potential uses. The accumulation of the amylose fraction of starch is controlled by a single dominant gene in quinoa, *GBSSI*. We report the sequencing and characterization of the *GBSSI* gene in 18 accessions of *Chenopodium*, including Andean quinoa and the related Mesoamerican chenopod grain species, *C. berlandieri* subsp. *nuttalliae* Saff. Two distinct homoeologs (*GBSSIa* and *GBSSIb*) were identified in the tetraploid accessions, and 19 different alleles were identified, including three null mutants – one in an accession of quinoa and two in a *waxy* landrace of *C. berlandieri* subsp. *nuttalliae*, referred to as 'H02'. Expression analysis of the null mutants revealed that *GBSSIa* and *GBSSIb* were both strongly expressed late in quinoa and *C. berlandieri* subsp. *nuttalliae* seed development. Starch phenotyping indicated that 'H02' produced no amylose, likely due to its having two mutated copies of *GBSSI*. *GBSSI* sequences were used to analyze the phylogenetic relationships between quinoa and other members of the *Chenopodium* genus. This study and the discovery of *Chenopodium GBSSI* null-mutants will assist in the development of new *Chenopodium* crops with novel starches.

Keywords: *Chenopodium*, quinoa, huauzontle, amylose, *waxy*, *GBSSI*

ACKNOWLEDGEMENTS

I want to express my gratitude to all of those who helped me with this challenging undertaking, especially my graduate committee. Thank you Dr. Rick Jellen for always being willing to meet and work things through with me. Thank you Dr. Jeff Maughan for offering patient advice and counsel. Thank you Dr. Clint Whipple for your critical thinking and valuable insights.

I am grateful for Brigham Young University and the Plant and Wildlife Sciences Department in particular for providing me with both spiritual and financial support while I worked on my degree.

Lastly, I want to thank my wonderful wife Emily Paxton, my sister Rachel Brown, brother-in-law Jonathan Strange, and the rest of my friends and family for occasionally providing me with a much needed social and psychological respite these past two years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1: Characterization of the Granule-Bound Starch Synthase I Gene *in* Chenopodium

Introduction

The genus *Chenopodium* L. is a relatively under-studied group of flowering plants with a worldwide distribution. *Chenopodium*, while widely known for its weedy species, also includes a large number of wild as well as several domesticated species: *C. album* L. in Europe, *C. giganteum* D.Don in Asia, *C. berlandieri* subsp. *nuttalliae* Saff. in Mesoamerica, *C. pallidicaule* Aellen in the Andes, and, most notably, quinoa (*C. quinoa* Willd.) in South America. Quinoa is an allotetraploid South American psuedocereal (AABB, 2n= 4x= 36) (Maughan et al., 2004; Storchova et al., unpublished data, 2014; Walsh et al., unpublished data, 2014; Ward, 2000). Its relatives include several commercially important plant species, such as amaranth (*Amaranthus* L.), spinach (*Spinacea oleracea* L.) sugar beet (*Beta vulgaris* L.), and garden orache (*Atriplex hortensis* L.). Quinoa is grown in a variety of different environments, from high elevations in the Andean mountains to coastal regions nearer to sea level. Quinoa was an important crop to the ancient Incan empire and was regarded as the sacred "mother of grains" (Risi and Galwey, 1984). Because of its use in indigenous religious ceremonies, its cultivation was suppressed by Spanish conquerors, who encouraged the local populace to grow introduced crops like wheat (*Triticum aestivum* L.) and barley (*Hordeum vulgare* L.). Quinoa was preserved by poor and isolated farming communities that grew the crop for sustenance, but recently, the global demand for quinoa has increased dramatically (Table 1; FAOSTAT, 2014). Quinoa has been successfully marketed in South America, North America, Europe, Australia, India, Africa, and Japan. Quinoa possesses an impressive nutritional profile, a nearly ideal amino acid profile, as well as significant tolerance to many abiotic stresses, including saline soils, drought, and frost (Gonzales et al., 1989; Ruales and Nair, 1992). These traits make quinoa a potentially valuable alternative

food crop, especially to less-developed regions where malnutrition is high and access to arable land is limited.

Starch is the most common and important carbohydrate energy source in the human diet. It accounts for a significant portion of global caloric intake, making up as much as 80% of the daily calories consumed in some parts of the world (Keeling and Myers, 2010). Starch is the main component of quinoa seeds. Approximately 60% of the quinoa seed is starch, the remainder of the seed is comprised of protein, moisture, fiber, fat, and ash, which account for roughly 15, 9, 10, 5, and 3%, respectively (Wright et al., 2002). Seed starch is comprised of two polysaccharides: amylose and amylopectin. Amylose is mostly linear and consists of glucose molecules bound by α-1,4 linkages. Amylopectin is also comprised of α-1,4 linkages, but has frequent α-1,6 branches as well. Most plant starches are approximately 20 to 30% amylose, while quinoa starches have amylose concentrations ranging from 3 to 20% (Atwell et al., 1983; Lindeboom et al., 2005).

The amylose:amylopectin ratio can drastically effect the properties and uses of starch. High-amylose starches are a type of resistant starch, an important component of dietary fiber. Resistant starches have a positive effect on the functioning of the digestive tract, microbial flora, blood cholesterol level, and glycemic index, and can assist in the control of diabetes, similarly to soluble fiber (Fuentes-Zaragoza et al., 2010). Flours made from purely amylopectin (*waxy*) starches have more fiber than normal flours, presumably due to more glucose being available for partitioning into cellulose and hemicellulose (Morita et al., 2002). Both high-amylose and *waxy* wheat flours have characteristics that are beneficial to food processing, though high-amylose and *waxy* flours have inferior bread-making properties (Hung et al., 2006). *Waxy* rice (glutinous rice; *Oryza sativa L.* var. glutinosa) and other *waxy*/low-amylose grains possess desired traits that

influence both cooking quality and seed softness, and are highly valued in East Asian cuisine (Liu et al., 2009).

Of the suite of genes controlling amylose and amylopectin biosynthesis in plant seeds, *Granule-Bound Starch Synthase I* (*GBSSI*) is one of the most important, as it is the only gene involved in amylose production. Plants lacking a functional copy of *GBSSI* produce *waxy* seed starches that consist entirely of amylopectin. *GBSSI* mutants have been well characterized in several crop species, including rice, maize, barley, wheat, and amaranth (Fujita et al., 2001; Nakamura et al., 1995; Park et al., 2010; Patron et al., 2002; Sano, 1984; Tsai, 1974). *GBSSI* interacts pleiotropically with other starch biosynthesis genes and has been shown to respond to changes in the expression of other genes, specifically *SSIIIa* (Fujita et al., 2007).

Although there has been extensive genetic research published on protein composition and abiotic stress tolerance in quinoa (Balzotti et al., 2008; Morales, 2009; Sun et al., 2014; Watanabe et al.; 2003), relatively little research has been conducted on the genetics of starch content. Previous research on quinoa starch has focused mainly on starch quantity and composition (Atwell et al., 1983; Lindeboom et al., 2005a, 2005b; Qian and Kuhn, 1999; Watanabe et al., 2007; Wolf et al., 1950, Wright et al., 2002). There has been virtually no research published on the genetics of starch biosynthesis in quinoa or its sister taxa from Mesoamerica, *C. berlandieri* subsp. *nuttalliae* ('huauzontle'). A better understanding of the genetics controlling starch biosynthesis in *Chenopodium* and in quinoa in particular is needed. Such an understanding could be used to design breeding programs and genetic engineering strategies geared towards developing crops with novel starch traits, such as high or low amylose concentrations. In this study, we report the sequencing, characterization, and expression analysis of two homoeologs of the *GBSSI* gene (*GBSSIa* and *GBSSIb*) in quinoa and other members of the

Chenopodium genus. We report the identification of a *waxy* variety of huauzontle, as well as several different mutant *GBSSI* alleles. We also report the phylogenetic relationships between quinoa and other members of the *Chenopodium* genus based on *GBSSI* sequences, including a *GBSSI* sequence from the related species, garden orache, which we use as an outgroup for the analysis.

Materials and Methods

Plant Materials and DNA Extraction

Germplasm and plant material was acquired from the collections of Brigham Young University, International Potato Center and U.N. Food and Agriculture Organization (CIP-FAO), E. De la Cruz (Instituto Nacional de Investigaciones Nucleares), D. Bertero (Facultad de Agronomía, Universidad de Buenos Aires), and H. Storchova (Institute of Experimental Botany, AS CR). In total, we analyzed 18 different accessions of *Chenopodium*, representing 12 New World allotetraploids (AABB), five New World diploids (AA), and one Old World diploid (BB). The accessions were selected in order to represent the two progenitor diploid genomes [*C. pallidicaule* (AA), *C. standleyanum* Aellen (AA), *C. dessicatum* A. Nels. (AA), *C. hians* Standl. (AA), *C. neomexicanum* Standl. (AA), and *C. ficifolium* Sm. (BB)], the wild tetraploids (*C. hircinum* Schrad., *C. berlandieri* Moq., *C. berlandieri* var. *boscianum* Moq., and *C. berlandieri* var. *macrocalycium* Aellen), several cultivated varieties of quinoa (including three highland accessions and three lowland accessions); and a related vegetable species from Mexico, *C. berlandieri* subsp. *nuttalliae* (hereafter referred to with its common name ‘huauzontle’). We also included a single accession of the closely related European vegetable species, garden orache

(*Atriplex hortensis*). This accession was selected in order to provide an outgroup for phylogenetic analysis of Chenopodium. A list of the selected germplasm is supplied in Table 2.

Plants were grown in potting soil at 25º C under natural light in 6" pots at the Brigham Young University greenhouses in Provo, Utah. Young leaf tissue was harvested from plants and lyophilized, and genomic DNA was extracted as described by Todd and Vodkin (1996). DNA was quantified using a NanoDrop 1000 spectrophotometer (Thermo Scientific, Wilmington, Delaware).

Primer Design, PCR, and PCR Product Purification

Primers for the PCR amplification of internal fragments of the *Chenopodium GBSSI* gene were designed based on consensus *GBSSI* cDNA sequences from amaranth (*Amaranthus caudatus* L*., A. cruentus* L.*,* and *A. hypochondriacus* L.; Park et al., 2010) using Geneious Pro v. 6 (Biomatters Ltd., New Zealand) and *Primer3* (Rozen and Skaletsky, 1998). Primers specific to the 5' and 3' untranslated regions (UTRs) of the gene were designed from sequences obtained using Rapid amplification of cDNA ends (RACE) protocol (Yeku and Frohman, 2011) and the SMARTer® RACE cDNA Amplification Kit (Clontech, Mountain View, Califonia). Lists of the primers used and the gene region they amplified are supplied in Table 3 and Table 4.

PCR amplification was carried out in a 25 µL reaction using the designed primers, diluted to 0.2 µM, 12.5 µL *Taq* 2x Master Mix (New England Biolabs Inc., Ipswich, Massachussets), and 30 to 100 ng genomic DNA. The PCR amplification protocol consisted of an initial denaturation step of 30 s at 94º C, followed by 30 cycles of amplification consisting of a 30 s denaturation step at 94º C, 30 s for primer annealing (temperature dependent on primer pair; Table 3 and Table 4), and a 60 s of DNA extension at 68º C. A final extension of 5 min at 68º C followed. PCR products were electrophoresed on 1% agarose gels and were visualized using

ethidium bromide and UV light. Amplicons were purified using EXOSAP PCR product cleanup (New England Biolabs Inc., Ipswich, Massachussets).

Cloning and Sequencing

Since homoeologous copies of the *GBSSI* gene were expected for the tetraploid accessions in the panel, all PCR products were cloned prior to sequencing, according to the manufacture's recommended instructions using the pGEM-T Easy Vector System (Promega, Fitchburg, Wisconsin). Plasmid DNA was extracted from overnight cultures using the GenElute Plasmid Miniprep Kit (Sigma-Aldrich, St. Louis, Missouri). Colonies were sequenced at the Brigham Young University DNA Sequencing Center using Big Dye v3.1 chemistry and electrophoresed on Applied Biosystems 3730xl DNA Analyzer. Six or more colonies were sequenced to acquire full-length sequences from each subgenome in the tetraploids. Assembly and alignment of the sequences were performed using Geneious Pro v. 6 (Biomatters Ltd., New Zealand).

RNA Extraction and cDNA Synthesis

Seeds from the quinoa accession 'G205-95' from and two landraces of huauzontle, 'H04' and 'H02', were planted in 6" pots filled with potting soil and grown in the Brigham Young University greenhouses at 25 ºC under natural light. Starting ten days post-anthesis (DPA), seed samples from two biological replicates of 'G205-95' were collected every ten days until maturation at 70 DPA. Seed samples from two biological replicates of 'H04' and 'H02' were taken at 40, 50, and 60 DPA, and at 50, 60, and 70 DPA, respectively. Seed samples were placed in liquid nitrogen immediately after harvesting and then stored at -80º C. The seed tissue was ground to a fine powder under liquid nitrogen and total RNA was extracted using the RNeasy Mini Kit (Qiagen, Valencia, California). Extracted RNA was quantified using a NanoDrop 1000

spectrophotometer (Thermo Scientific, Wilmington, Delaware) and visualized on a 1% agarose gel with ethidium bromide. High-quality samples of total RNA, as determined by the 260/280 ratio and the presence of strong, intact 18S and 28S rRNA bands, were treated with gDNA Eraser (Clontech, Mountain View, California) to digest any residual genomic DNA contamination. The cDNA was then synthesized from the treated RNA samples using the PrimerScript RT Reagent Kit (Clontech, Mountain View, California).

Real-Time PCR

Subgenome-specific *GBSSI* expression levels were quantified using SYBR® Select Master Mix (Life Technologies, Waltham, Massachussets) on an AB 7300 Real-Time PCR System (Applied Biosystems, Waltham, Massachussets). Each 20 µL reaction consisted of 10 µL SYBR® Select Master Mix, 400 nM forward primer, 400 nM reverse primer, and 1 µL cDNA. Subgenome-specific primers were used to measure homoeolog-specific (*GBSSIa* or *GBSSIb*) expression (Table 3). *GAPDH* was used as an internal reference. Primers were designed to amplify a ~200 bp fragment from the 3' end of the gene. Primers were designed such that at least one member of the pair spanned an exon-intron junction in order to limit amplification of any potential genomic DNA contamination. Amplification products were sequenced to confirm their subgenome specificity. All reactions were performed in triplicate and a standard curve was used to calculate amplification efficiencies for each primer set. Relative gene expression values (RE) for each sample were then calculated using the equation:

$$RE = (1 + E_T)^{-Ct} \div (1 + E_R)^{-CtR}$$

In this equation, E_T is the amplification efficiency of the primer set amplifying the target, Ct is the Ct value of the target gene, E_R is the amplification efficiency of the primer set amplifying the reference, and CtR is the Ct value of the reference gene.

Starch Phenotyping

Two or three mature seeds from each accession were ground to powder and suspended in 100µL of an I_2/KI solution (5% (wt/v) I_2 and 10% (wt/v) KI; Hunt et al., 2010). 20µL of the crushed seed/iodine mixture was examined under 40X objective magnification on a Zeiss Axioplan 2 microscope (Oberkochen, Germany) and starch contents were evaluated qualitatively. Slides were photographed using a Zeiss AxioCam ERc 5s (Oberkochen, Germany). Waxy starches, which contain no amylose, stain reddish-brown, while normal starches with amylose stain blue to purple. All analyses were performed in duplicate.

Phylogenetic Analysis

Phylogenetic analysis was performed in Geneious Pro v. 6 (Biomatters Ltd., New Zealand). All Chenopodium *GBSSI* sequences were aligned and a phylogenetic tree was constructed using the MrBayes plugin (Huelsenbeck and Ronquist, 2001), which employed the HKY85 substitution model. Each Markov chain was started from a random tree and run for 900,000 generations with every 1000th tree sampled from the chain. All sample points prior to the 800,000th generation were discarded as burn in and the remaining trees combined to find the *a posteriori* probability estimate of phylogeny.

Results and Discussion

GBSSI *Gene in* Chenopodium

We used a genomic PCR-based approach to amplify, clone, and sequence six contiguous fragments of the *GBSSI* gene in 18 different accessions of *Chenopodium* (including 12 tetraploids and six diploids), as well as a single accession of garden orache. Cloning was necessary due to the likelihood of amplifying multiple PCR targets, especially in the polyploids.

These six primer sets amplified fragments ranging from 700 to 1,200 bp in length, and fragments overlapped by approximately 200 bp to allow for contiguous sequence assembly. Sequence assembly and alignment in Geneious v. 6 (Biomatters Ltd., New Zealand) produced a *Chenopodium* consensus sequence of 3,237 bp. Exon splice sites were identified using Spidey (Wheelan et al., 2001), and were based on alignments between *Chenopodium GBSSI* sequences and amaranth *GBSSI* cDNA sequences (Park et al., 2010). The *Chenopodium* consensus *GBSSI* gene consisted of 13 exons, which spanned from 236 to 563, 667 to 748, 847 to 946, 1,061 to 1,151, 1,219 to 1,283, 1,381 to 1,482, 1,567 to 1,677, 1,764 to 2,008, 2,089 to 2,266, 2,356 to 2,548, 2,644 to 2,731, 2,846 to 2,975, and 3,065 to 3,182. Exons ranged in size from 64 bp to 327 bp, and the 12 introns ranged in size from 68 bp to 115 bp (Figure 1). All introns in the *Chenopodium* consensus *GBSSI* gene followed the universal GT-AG rule (Breathnach and Chambon, 1981). The structure of the *GBSSI* gene in *Chenopodium* is similar to the structure of the *GBSSI* gene in other plants, including maize, rice, sweet potato, amaranth, potato, and millet (Fukunaga et al., 2002; Kimura et al., 2000; Klosgen et al., 1986; Okagaki, 1992; Park et al., 2009; van der Leij et al., 1991).

The consensus coding sequence (CDS) consisted of 1,818 bp and translated into a protein with 605 amino acid residues and an estimated weight of 66.9 kDa. Pfam detected a starch synthase catalytic domain and a glycosyl transferase group in the protein, with E-values of 4.9e-70 and 1.2e-16, respectively (Punta et al., 2012; Figure 3).

A transit peptide was predicted using the ChloroP program with a score of 0.505 (Emanuelsson et al., 1999). Gavel and von Heijne (1990) proposed that the cleavage site for chloroplast transit peptides is I/V X A/C ^ A. Salehuzzaman et al. (1993) suggested that this may change to I/V X A/C ^ G for dicots. Based on the chloroplast transit peptide cleavage site

sequence proposed by Gavel and von Heijne (1990) and Salehuzzaman et al. (1993), we predicted a transit peptide of 77 amino acids and a cleavage site of ITC^G at the N-terminal end of the protein (Figure 3). This suggested that the mature GBSSI protein would be 528 amino acids in length and have an estimated molecular weight of 58.5 kDa. The *Chenopodium* GBSSI consensus transit peptide would have an estimated weight of 8.4 kDa.

The *Chenopodium* GBSSI consensus protein sequence was submitted to a BLASTP search of the Reference Proteins Database using the default settings. The *Chenopodium* GBSSI consensus protein showed a high degree of similarity to cacao (*Theobroma cacao L.*), orange (*Citrus sinensis L.*), soybean (*Glycine max L.*), and tomato (*Solanum lycopersicum L.*) GBSSI proteins (e-values of 0). Aligning the *Chenopodium* GBSSI consensus protein sequence to GBSSI proteins from other plants revealed that GBSSI proteins are highly conserved across plant species. The full length *Chenopodium* GBSSI consensus protein shared 94.7% identity with garden orache GBSSI, 79.2% identity with amaranth, 65.9% identity with *Arabidopsis thaliana*, 68.2% with potato, and 57.9% to 85% identity with other plant species (Table 5).

The 77 amino acid transit peptide of the *Chenopodium* GBSSI consensus sequence is similar in size to GBSSI transit peptides in garden orache, amaranth, sweet potato, and potato (Kimura et al., 2000; Park et al., 2009; van der Leij et al., 1991), though it is not highly conserved. The *Chenopodium* GBSSI consensus transit peptide shared 81.8% identity with the transit peptide from garden orache GBSSI, 60.8% identity with the amaranth transit peptide, 34.6% with *Arabidopsis thaliana*, 32.5% identity with potato, and 4.1% to 74.7% identity with the GBSSI transit peptides from other plant species (Table 5).

The mature *Chenopodium* GBSSI consensus protein sequence was much more conserved across species than the transit peptide. The mature protein from *Chenopodium* shared 96.6%

identity with the mature GBSSI protein from garden orache, 82.2% identity with amaranth, 71% identity with *Arabidopsis thaliana*, 73.3% with potato, and 64.3% to 86.6% identity with other plant species (Table 5).

Aligning the *Chenopodium* GBSSI consensus protein sequence to the GBSSI protein sequences of other plants revealed three conserved protein domains (Figure 3). These domains were similar in structure across plant species, and were highly similar to the *Escherichia coli* glycogen synthase protein as well (Kumar et al., 1986; van der Leij et al., 1991; Park et al., 2009). The first conserved region spanned amino acids 93 to 97 in the *Chenopodium* protein and consisted of the KTGGL motif. This motif is believed to be the ADP-glucose binding site (Furukawa et al., 1990). As ADP-glucose is the precursor molecule for starch, this motif is essential to normal starch synthesis and GBSSI function. The second conserved region spanned amino acids 477 to 485 in the *Chenopodium* protein and consisted of the sequence PSRFEPCGL, which is completely conserved in dicots, rice, and the *E. coli* glycogen synthase protein, and is highly conserved in monocots other than rice. The third conserved region spanned amino acids 503 to 507 and consisted of the KTGGL "look-alike" motif, STGGL (Dry et al., 1992). This motif, as well as the C-terminal tail of the GBSSI protein, are specific to GBSSI and confer most of the GBSSI-specific properties (Edwards et al., 1999). The KTGGL "look-alike" motif is named for its resemblance to the KTGGL motif, and because of its structural similarities, it is likely involved in ADP/ADP-glucose binding as well.

We estimated the weight of the full-length *Chenopodium* GBSSI consensus protein to be 66.9 kDa. We estimated the weight of mature *Chenopodium* GBSSI consensus protein to be 58.5 kDa. In a previous study, Lindeboom et al. (2005) identified two quinoa GBSSI proteins by SDS-page analysis. One of these proteins weighed approximately 62 kDa, while the other weighed

approximately 56 kDa. Lindeboom et al. suggested these two GBSSI proteins could be different alleles or isoforms of *GBSSI*. The weights calculated by Lindeboom et al. (2005) are reasonably close to the weights we calculated for the complete and mature GBSSI protein (approximately 66.9 kDa and approximately 58.5 kDa, respectively). We therefore suggest that the two quinoa GBSSI proteins identified by Lindeboom et al. are not different alleles or isoforms, but are instead just the complete and mature versions of the quinoa GBSSI protein.

Sequence Variation in Chenopodium *Diploids and Homoeolog Designation*

The assembled *GBSSI* sequences from the diploid accessions of *Chenopodium* grouped into two distinct variants. One group of sequences was found in New World diploids and the other distinct variant in the Eurasian diploid *C. ficifolium*. We arbitrarily designated the New World variant as "A-genome" and the Eurasian variant as "B-genome". These two distinct variants shared 91.5% identity. Figure 2 shows aligned sequences of New World diploids and *C. ficifolium*.

Alignments of assembled *GBSSI* sequences from the tetraploid accessions of *Chenopodium* showed that there were two distinct gene copies per accession. One copy was most similar (>99% identity) to sequences from the A-genome diploids and the other copy was most similar (>99% identity) to the sequence from the B-genome diploid. This was consistent with the hypothesized allotetraploid nature of these accessions (Maughan et al., 2004; Storchova et al., unpublished data, 2014; Walsh et al., unpublished data, 2014; Ward, 2000). This pattern in the data indicate two likely hypotheses: 1) that the analyzed allotetraploids have a monophyletic origin; and 2) that, like *Gossypium hirsutum L.*, *Chenopodium* tetraploids arose from hybridization between Old and New World-origin diploids. This finding was previously suggested by Kolano et al. (2011) after *in situ* hybridization experiments with a repetitive

sequence, 18-24J, that hybridized abundantly to one subgenome (18 chromosomes) of the *C. berlandieri-C. quinoa* allotetraploid complex and to one subgenome in hexaploid Eurasian *C. album*, but not to diploids native to the New World. Those tetraploid sequences that were most similar to the gene sequences of the presumptive A-genome diploid group were designated as A-genome homoeologs (*GBSSIa)*. Those tetraploids sequences that were most similar to the gene sequences of the presumptive B-genome diploid ancestor were designated as B-genome homoeologs (*GBSSIb*). However, caution must be exercised since we did not sequence a comprehensive sample of all known or suspected A- and B-genome *Chenopodium* diploids. Collections of *Chenopodium* diploids are currently limited by several constraints, among them a lack of taxonomic consensus regarding species delineations; seed germination recalcitrance or inviability in existing collections; and the rare or episodic distribution and appearance of a number of species, for example the North American taxa *C. cycloides* A. Nels., *C. nevadense* Standl., and *C. pallescens* Standl. (Jellen et al., 2011).

DNA and protein sequences from the *GBSSI* alleles of the same subgenome were highly conserved for the *Chenopodium* genus. The genomic sequences for the A-genome homoeologs, *GBSSIa*, were over 99% identical, as were coding sequences and protein sequences. The genomic sequences for the B-genome homoeologs, *GBSSIb*, shared over 99% identity, as did the coding sequences and protein sequences. There were however, significant differences between the alleles of different subgenomes. The *GBSSIa* and *GBSSIb* genomic consensus sequences shared 91.9% identity. Most (73.9%) of these differences were in intron regions, which were 84.6% identical. Of the 176 single-nucleotide polymorphisms (SNPs) identified, 116 (65.9%) were found in introns. Of the 23 indels identified, 23 (100%) were found in introns. The coding sequences were 95.6% identical. Those sequence differences that were found in CDS were

mostly (52.3%) nonsynonymous, and GBSSIa and GBSSIb consensus protein sequences shared 97.7% identity.

Allelic Variation in Chenopodium GBSSI

We identified a total of 19 different *GBSSI* alleles in the analyzed accessions (Table 6). One *GBSSIa* allele was shared by five accessions of *Chenopodium* and one *GBSSIb* allele was shared by six accessions. Based on their prevalence in the study panel, these two alleles, referred to as *GBSSIa-1* and *GBSSIb-1*, were arbitrarily selected for use as references and models for the *Chenopodium GBSSIa* and *GBSSIb* genes.

Three putative mutant copies of *GBSSI* were identified in the surveyed accessions of *Chenopodium*. One mutant allele was found in the B-genome homoeolog of the lowland quinoa accession 'G205-95'. Sequence analysis of 'G205-95' revealed a G to A substitution in the third position of codon 129 in its copy of *GBSSIb* relative to *GBSSIb-1.* The substitution creates a premature termination codon (TGG to TGA; W129X), truncating the protein by 477 amino acids. We refer to this allele as *gbssib-t* (Table 6).

Sequence analysis of the huauzontle landraces 'H02' and 'H04' revealed another putative mutant copy of the *GBSSI* gene. The *GBSSIa* homoeolog in 'H02' and 'H04' had a normal CDS, normal intron splice sites, and a normal number of exons. In addition, the predicted protein from this allele was the same length as the GBSSIa-1 reference and contained all essential domains. However, a T to C substitution in the second position of codon 54 was observed in this allele of *GBSSIa* that changed the codon from ATA in *GBSSIa-1* to ACA (I54T). This mutation was within the predicted transit peptide. In general, the transit peptide is not conserved across plant species. However, this position was highly conserved, and in all dicot GBSSI proteins analyzed, there is either an isoleucine or a leucine present at this position in the transit peptide. We refer to this

allele as *gbssia-tp* (Table 6).

The third putative mutant copy of *GBSSI* was also found in the huauzontle landrace, 'H02'. The *GBSSIb* homoeolog in 'H02', which we refer to as *gbssib-Δ* (Table 6), had a deletion that spanned from bp 192 to 635 in the *GBSSIb-1* reference sequence. This 440 bp mutation deleted the last 182 bp of exon one, all 103 bp of intron one, all 81 bp of exon two, all 99 bp of intron two, and the first 24 bp of exon 3. Interestingly, the deletion is in-frame and is predicted to allow for translation of *gbssib-Δ* until the normal stop codon is reached. This mutant GBSSI protein would be 526 amino acid residues in length, as opposed to the 605 amino acids residues found in the GBSSIb-1 protein. The deleted portion of the mutant GBSSI protein included a portion of the transit peptide, the transit peptide cleavage site, as well as the important KTGGL motif.

Expression Analysis

Gene-specific primer pairs were used to measure the individual expression levels of *GBSSIa* and *GBSSIb* in seeds from the accessions bearing the putative mutant alleles, including the quinoa accession 'G205-95' and the huauzontle accessions 'H04' and 'H02'. Measurements for 'G205-95' were taken starting at 10 days post anthesis (DPA) and ending at maturation (Figure 5). Transcripts for both *GBSSI* homoeologs were detected at all developmental stages, but expression levels varied significantly at different sampling times. *GBSSIa* relative expression levels increased from 0.03 at 10 DPA, to 0.50 at 20 DPA, to 1.73 at 30 DPA, and to their peak of 3.17 at 40 DPA. Relative expression levels decreased thereafter, from 1.31 at 50 DPA, to 0.33 at 60 DPA, to 0.18 at 70 DPA. *GBSSIb* relative expression levels were measured at 0.04 at 10 DPA, but increased to 0.93 at 20 DPA, to 2.51 at 30 DPA, and to their peak of 5.70 at 40 DPA. Expression levels decreased to 4.42 at 50 DPA, to 1.04 at 60 DPA, and to 0.45 at 70 DPA.

Expression analysis of the quinoa accession ‘G205-95’ indicated that *GBSSI* expression levels in quinoa peak in the middle or later periods of development, that the *GBSSIa* and *GBSSIb* homoelogs are expressed at all developmental time periods, and that the two homoeologous genes followed the same general expression patterns.

Seed samples for the huauzontle accessions ‘H04’ and ‘H02’ were taken at 40, 50, and 60 DPA, and at 50, 60, and 70 DPA, respectively (Figure 6). For ‘H04’, *GBSSIa* and *GBSSIb* relative expression levels at 40 DPA were 4.14 and 4.81, respectively. *GBSSIa* and *GBSSIb* expression levels increased to 17.80 and 26.04, respectively, at 50 DPA, then decreased to 4.22 and 7.06, respectively, at 60 DPA. For ‘H02’, *GBSSIa* and *GBSSIb* relative expression levels peaked at 50 DPA and were 32.51 and 53.91, respectively. *GBSSIa* and *GBSSIb* expression levels then decreased to 30.44 and 44.72, respectively, at 60 DPA, and to 19.64 and 36.97, respectively, at 70 DPA. In the huauzontle accessions, as in the quinoa accession ‘G205-95’, *GBSSI* relative expression levels peaked in the middle or later periods of development, both homoeologs were expressed and followed the same general expression patterns.

GBSSI expression patterns have been well studied in other plants, especially cereals. These previous studies have shown that *GBSSI* is generally expressed late in seed development (Dry et al., 1992; Hirose and Terao; 2004, Park et al., 2011). We found that, as in previously studied plants, *GBSSI* was expressed strongly in the middle and later periods of seed development in quinoa and huauzontle. We also found that both the *GBSSIa* and *GBSSIb* homoeologs were expressed at all developmental time points and that they followed the same general expression patterns.

These expression patterns presented here are validated by a study published by Lindeboom et al. (2005). Lindeboom et al. that found that starch concentrations in some lines of

quinoa peak six weeks post anthesis. Because amylose concentrations are correlated with *GBSSI* activity, this suggests that *GBSSI* expression levels reached their maximum sometime around six weeks after anthesis.

Starch Phenotyping

We screened all 18 Chenopodium accessions for *waxy* mutants using 5% I_2/KI solution (Hunt et al., 2010). The majority of Chenopodium seeds contained starch that stained bluish-purple (Figure 4). This indicated the presence of amylose and suggested that, given the dominant nature of the *GBSSI* gene, these accessions possessed at least one functional copy of GBSSI (Kempton, 1919, Park et al., 2010). We identified one landrace of huauzontle, 'H02', whose seed starch stained reddish-brown (Figure 4), suggesting a *waxy* phenotype and no functional copies of the *GBSSI* gene.

Hypotheses Regarding Allelic Functionality

The quinoa accession 'G205-95' contained a truncated copy of the *GBSSIb* gene, *gbssib-t.* The protein translated from this allele was 477 amino acids shorter that the reference and was missing two of the three conserved GBSSI domains. It is therefore unlikely that this protein would be capable of functioning normally. The starch from quinoa accession 'G205-95' stained bluish-purple, in the presence of I_2/KI, which indicated that it did not have the *waxy* mutant phenotype. This was consistent with the dominant nature of the *GBSSI* gene (Kempton, 1919, Park et al., 2010). Since 'G205-95' possessed a functional copy of *GBSSIa* to mask its truncated *gbssib-t* allele, its phenotype was non-*waxy*.

Seed starch from the huauzontle landrace 'H02' stained reddish-brown in the presence of I_2/KI. This suggested that it contained no amylose and therefore no functional copies of *GBSSI*. Sequence analysis of 'H02' revealed a mutated copy of *GBSSIb* (*gbssib-Δ*). The *gbssib-Δ* allele

possessed a mutation that deleted the last 182 bp of exon one, all 103 bp of intron one, all 81 bp of exon two, all 99 bp of intron two, and the first 24 bp of exon 3. The mutant gbssib-Δ protein would be 526 amino acid residues in length, as opposed to the 605 amino acids residues found in the reference protein. The deleted portion of the mutant gbssib-Δ protein included a portion of the transit peptide, the tentative transit peptide cleavage site, and the important KTGGL motif.

The deleted portions of *gbssib-Δ* in 'H02' likely have serious functional repercussions. With a large section of the transit peptide, including the transit peptide cleavage site, missing from the translated protein, it is unlikely that the protein would be capable of locating its intended plastid. It is also unlikely that this protein would be capable of being cleaved correctly into the transit peptide and mature protein. In addition, this mutation extends past the transit peptide into the functional part of the *GBSSI* gene. The conserved motif KTGGL, believed to be involved in ADP-glucose binding, is deleted in *gbssib-Δ* (Furukawa et al., 1990). If the mutant *gbssib-Δ* in 'H02' lacked the ability to bind to ADP-glucose, the precursor of starch, then it is unlikely that it would be able to synthesize amylose.

Interestingly, although 'H02' contained no amylose, we found that it had a seemingly-functional copy of *GBSSIa.* This copy of *GBSSIa*, referred to as *gbssia-tp*, appeared normal, except that a T to C substitution in the second position of codon 54 changed the codon from ATA in *GBSSIa-1* to ACA (I54T).

We predict that the *waxy* phenotype of 'H02' is due to a lack of functional *GBSSI* expression in the seed. It is possible that through subfunctionalization, the B-genome homoeolog, *gbssib-Δ*, was expressed in the seed, while the seemingly-functional A-genome homoeolog, *gbssia-tp*, was expressed elsewhere in the plant. This would result in *waxy* seed starch, as no functional copy of *GBSSI* would be expressed in the seed. However, this hypothesis

was not supported by our expression analysis. The results of RT-PCR clearly indicate that both copies of the gene are expressed in the seed at all developmental time points.

Another hypothesis is that *gbssib-Δ* interacts in a dominant negative fashion to disable or silence the seemingly-functional A-genome homoeolog, *gbssia-tp*. To test this hypothesis, we crossed 'H02' (*waxy*, AAbb) with 'H04' (non-*waxy*, AABB), which contains a functional copy of *GBSSIb* in addition to the *gbssia-tp* allele. If *gbssib-Δ* is capable of silencing functional copies of *GBSSI*, then we would expect that the F1, which would harbor an introduced copy of *gbssib-Δ* from 'H02', as well as a normal B-genome copy of *GBSSI* from 'H04', would also be *waxy* (AABb). F1 plants are currently growing and their seed starch will be analyzed as soon as possible in order to confirm or refute this hypothesis.

A third hypothesis is that the seemingly-functional A-genome copy of *GBSSI*, *gbssia-tp*, in 'H02' is in fact, non-functional. The I54T mutation in *gbssia-tp* was located in the transit peptide. In general, the transit peptide is not conserved across plant species. However, in all dicot GBSSI proteins analyzed, there was either an isoleucine or a leucine present at this position in the transit peptide. The change at this conserved position from a non-polar, hydrophobic amino acid (isoleucine or leucine) to a polar, hydrophilic amino acid (threonine) may be sufficient to disrupt the normal function of the *gbssia-tp* transit peptide and effectively silence the gene. Given the established dominant/recessive nature of *GBSSI*, this hypothesis is the most likely. We are currently developing F1 and F2 plants from a cross between 'H02' (*waxy*, aabb) and 'H04' (non-*waxy*, aaBB) to test this hypothesis.

This hypothesis is validated by the characterization of *GBSSI* in a closely related landrace of huauzontle. Cepeda-Cornejo et al. (unpublished data, 2014) analyzed a non-*waxy* huauzontle landrace that was homozygous for the *gbssib-Δ* allele, and heterozygous for the *gbssia-tp* allele

and another *GBSSIa* allele that lacked the I54T substitution in the transit peptide. Essentially, this plant's *GBSSI* genotype was identical to that of the *waxy* 'H02' (aabb), except that it possessed a single *GBSSIa* allele without the substitution mutation in the transit peptide (Aabb). The non-*waxy* phenotype in 'H3' must therefore have been due to this *GBSSIa* allele without the substitution. These results help to verify the dysfunctionality of the *gbssia-tp* allele. Future work will focus on making crosses between 'H02', 'H04', and 'H3' in order to further validate this hypothesis.

There is a precedent for small, seemingly innocuous changes in *GBSSI* having a significant effect on amylose accumulation. Liu et al. (2009) describe a novel *GBSSI* gene, Wx^{hp}, in rice in which an A to G change results in an aspartate to glycine substitution at codon 165. Although this small change did not lead to decreased *GBSSI* activity *in vitro*, it did noticeably reduce the ability of *GBSSI* to bind to starch granules, and thereby reduced amylose content *in vivo*. The Wx^{hp} allele in rice therefore establishes a precedent for seemingly innocuous changes in *GBSSI* having a significant effect on amylose accumulation. Such may be the case for the *gbssia-tp* allele.

Phylogenetic Analysis

The taxonomy and evolutionary history of the quinoa genus *Chenopodium* is widely recognized as being problematic at all levels (Jellen et al., 2011). Notably, the species complex *C. album* consists of diploid, tetraploid, and hexaploid individuals (Bhargava et al., 2005). There are a number of established species that are fully capable of crossing and producing fertile hybrids with other chenopods (Bonifacio, 1995; Wilson, 1980). It has been established that quinoa is an allotetraploid, and strong evidence suggests that quinoa evolved from the North American tetraploid *C. berlandieri*, but further details on the phylogeny and evolutionary history

of *Chenopodium* are lacking (Kolano et al., 2011; Maughan et al., 2004; Maughan et al., 2006; Heiser and Nelson, 1974; Walters, 1988; Ward, 2000; Wilson, 1980; Wilson and Heiser, 1979).

Breeding resources for cultivated *Chenopodium* crops are somewhat limited. The wild *Chenopodium* germplasm represents an untapped genetic resource, so the *Chenopodium* plylogeny and quinoa ancestry are of great interested to those interested in cross-breeding quinoa and its relatives. Previous studies have implicated *C. incanum* S. Watts., *C. neomexicanum*, *C. fremontii* S. Watts., and *C. watsonii* A. Nels. as possible A-genome progenitors of quinoa and *C. berlandieri* (Aellen and Just, 1929; Jellen et al., 2011; Sederberg, 2008; Wilson, 1988; Wilson, 1980). The B-genome parent was likely an Old World *Chenopodium* diploid, such as *C. ficifolium* (Jellen et al., 2011).

In an attempt to elucidate the ancestral origins of quinoa, we constructed a phylogenetic tree from all *Chenopodium GBSSI* sequences using the default settings of the MrBayes plugin (Huelsenbeck and Ronquist, 2001) in Geneious v. 6 (Biomatters Ltd., New Zealand). A garden orache *GBSSI* sequence was used as the outgroup (Figure 7). Two main clades were observed (posterior probability values of 99.97 each). One clade corresponded to the A-genome homoeologs, while the other corresponded to the B-genome homoeologs. Each tetraploid had one allele in each clade. The putative A-genome diploids [BYU accession numbers 835 (*C. dessicatum*), 843 (*C. neomexicanum*), 921 (*C. standleyanum*), 1005 (*C. hians*), and 1302 (*C. pallidicaule*)] all clustered with the *GBSSIa* alleles from the tetraploids, but were scattered across the A-genome clade. The presumably B-genome diploid, *C. ficifolium*, was found at the base of the *GBSSIb* clade.

This analysis validates the hypothesized allopolyploid nature of quinoa and its close relationship to the *C. berlandieri* and *C. hircinum* taxa. Unfortunately, there was a lack of

support for the majority of the relationships predicted by this analysis. We were unable to infer any new information about specific diploid ancestors of the *berlandieri-quinoa* complex. It is possible that the *GBSSI* gene is too highly conserved in *Chenopodium* to be useful in elucidating a relatively recent event such as the evolution of quinoa. It is also possible that our sample size was too small to yield meaningful results. Future work will focus on sequencing the *GBSSI* gene in additional accessions of *Chenopodium*, including additional A-genome and B-genome diploids, in order to improve the phylogenetic analysis.

Conclusions

Amylose content in seed starch is one of the most important factors in determining seed quality, as amylose contents can affect the dietary fiber content, stability, viscosity, and texture of processed foods (Morita et al., 2002). Amylose contents are particularly important in East Asian cuisines, where sticky, low-amylose grains are often used and preferred. Understanding the genetic mechanisms that control amylose production and accumulation in *Chenopodium* and in the seed crop quinoa will be helpful in developing crops with novel traits and in designing breeding programs and genetic engineering strategies geared towards crop improvement. This is the first report in which the amylose-producing *GBSSI* gene has been sequenced and characterized in the genus *Chenopodium*. *GBSSI* expression levels in quinoa and huauzontle were also analyzed for the first time. Three mutant *GBSSI* alleles were identified, and a phylogenetic tree was constructed using *GBSSI* sequences.

This study provides useful information about the structure and function of the *Chenopodium GBSSI* gene. These results will provide the groundwork to identify and develop novel *GBSSI* mutants in quinoa, and will assist in the designing of breeding programs and

genetic engineering strategies geared towards crop improvement. Future work will focus on improving our *Chenopodium* phylogenetic analysis, developing crop lines with novel starches, further characterizing the mutant alleles, and further examining mutant phenotypes.

References

Aellen, P., and T. Just. 1929. Key and synopsis of the American species of the genus *Chenopodium L.* Am. Midl. Nat. 30:47-67.

Atwell, W.A., B.M. Patrick, L.A. Johnson, and R.W. Gloss. 1983. Characterization of quinoa starch. Cereal Chem. 60:9-11.

Balzotti, M.R.B., M. Stevens, P.J. Maughan, D.J. Fairbanks, J.N. Thornton, D.A. McClellan, E.N. Jellen, and C.E. Coleman. 2008. Expression and evolutionary relationships of the *Chenopodium quinoa* 11S seed storage protein gene. Int. L. Plant Sci. 169(2):281-291

Bonifacio, A. 1995. Interspecific and intergeneric hybridization in chenopod species. M.S. thesis, Brigham Young Univ.

Bhargava, A., T.S. Rana, S. Shukla, D. Ohri. 2005. Seed protein electrophoresis of some cultivated and wild species of *Chenopodium*. Biologia Plantarum. 49:505-511.

Breathnach, R. and P. Chambon. 1981. Organization and expression of eukaryotic split genes coding for proteins. Ann. Rev. Biochem. 50:349–383.

Dry, I., A. Smith, A. Edward, M. Bhattacharyya, P. Dunn, and C. Martin. 1992. Characterization of cDNAs encoding two isoforms of granule-bound starch synthase which show differential expression in developing storage organs of pea and potato. The Plant J. 2:193-202.

Edwards, A., A. Borthakur, S. Bornemann, J. Venail, K. Denyer, D. Waite, D. Fulton, A. Smith, and C. Martin. 1999. Specificity of starch synthase isoforms from potato. Eur. J. Biochem. 266:724–736.

Emanuelsson, O., H. Nielsen, and G. von Heijne. 1999. ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein Sci. 8:978-984.

FAOSTAT. 2014. FAOSTAT Database.

Fuentes-Zaragoza, E., M.J. Riquelme-Navarrete, E. Sánchez-Zapata, and J.A. Pérez-Álvarez. 2010. Resistant starch as functional ingredient: A review. Food Res. Int. 43:931-942.

Fujita, N., H. Hasegawa, and T. Taira. 2001. The isolation and characterization of a *waxy* mutant of diploid wheat (*Triticum monococcum L.*). Plant Sci. 160:595-602.

Fujita, N., M. Yoshida, T. Kondo, K. Saito, Y. Utsumi, T. Tokunaga, A. Nishi, H. Satoh, J.H. Park, J.L. Jane, A. Miyao, H. Hirochika, and Y. Nakamura. 2007. Characterization of *SSIIIa*-deficient mutants of rice: the function of *SSIIIa* and pleiotropic effects by *SSIIIa* deficiency in the rice endosperm. Plant Physiol. 144:2009-2023.

Fukunaga, K., M. Kawase, and K. Kato. 2002. Structural variation in the *Waxy* gene and differentiation in foxtail millet [*Setaria italica (L.) P. Beauv*.]: implications for multiple origins of the *waxy* phenotype. Mol. Genet. Genomics. 268:214-222.

Furukawa, K., M. Tagaya, M. Inouye, J. Preiss, and T. Fukui. 1990. Identification of lysine 15 at the active site in *Escherichia coli* glycogen synthase. Conservation of Lys-X-Gly-Gly sequence in the bacterial and mammalian enzymes. J. Biol. Chem. 265:2086-2090.

Gavel, Y. and G. von Heijne. 1990. A conserved cleavage-site motif in chloroplast transit peptides. FEBS Lett. 261:455–458.

Gonzales, J.A., A.Roldan, M. Gallardo, T. Escudero, and F.E. Prado. 1989. Quantitative determination of chemical compounds with nutritional value from Inca crops: *Chenopodium quinoa* (‘quinoa’). Plant Food Hum. Nutr. 39:331-337.

Hirose, T. and T. Terao. 2004. A comprehensive expression analysis of the starch synthase gene family in rice (*Oryza sativa L.*). Planta 220:9-16.

Heiser, C.B., and C.D. Nelson. 1974. On the origin of cultivated Chenopods (*Chenopodium*). Genetics. 78:503-505.

Huelsenbeck, J.P., and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogeny. Bioinformatics. 17:754-755.

Hung, P., T. Maeda, and N. Morita. 2006. *Waxy* and high-amylose wheat starches and flours - characteristics, functionality and application. Trends in Food Sci. and Technol. 17:448-456.

Hunt, H.V., K. Denyer, L.C. Packman, M.K. Jones, and C.J. Howe. 2010. Molecular basis of the waxy endosperm starch phenotype in broomcorn millet (*Panicum miliaceum L.*). Mol. Biol. Evol. 27(7):1478-1494.

Jellen, E.N., B.A. Kolano, M.C. Sederberg, A. Bonifacio, and P.J. Maughan. 2011. *Chenopodium*. In. Chittaranjan Kole, editor, Wild crop relatives: genomic and breeding resources: legume crops and forages. Springer Heidelberg Dordrecht London New York. 35-61

Jeon, J., N. Ryoo, T. Hahn, H. Walia, and Y. Nakamura. 2010. Starch biosynthesis in cereal endosperm. Plant Phisiol. Bioch. 48:383-392.

Keeling, P.L., and A.M. Myers. 2010. Biochemistry and genetics of starch synthesis. Ann. Rev. Food Sci. and Technol. 1:271-303.

Kempton, J. H. 1919. Inheritance of *waxy* endosperm in maize. USDA Bulletin 754. Washington D.C.

Kimura, T., O. Ideta, and A. Saito. 2000. Identification of the gene encoding granule-bound starch synthase I in sweet potato (*Ipomoea batatas (L.) Lam*.). Plant Biotech. J. 17:247-252.

Klosgen, R., A. Gierl, Z. Schwartz-Sommer, and H. Saedler. 1986. Molecular analysis of the *waxy* locus of maize. Mol. Gen. Genet. 203:237-244.

Kolano, B., B.W. Gardunia, M. Michalska, A. Bonifacio, D. Fairbanks, P.J. Maughan, C.E. Coleman, M.R. Stevens, E.N. Jellen, and J. Maluszynska. 2011. Chromosomal localization of two novel repetitive sequences isolated from the *Chenopodium quinoa* Willd. genome. Genome 54:710-717.

Lindeboom, N., P.R. Chang, K.C. Falk, and R.T. Tyler. 2005. Characteristics of starch from eight quinoa lines. Cereal Chem. 82:216-222.

Lindeboom, N., P.R. Chang, R.T. Tyler, and R.N. Chibbar. 2005. Granule-bound starch synthase I (GBSSI) in quinoa (*Chenopodium quinoa Willd*.) and its relationship to amylose content. Cereal Chem. 82:246-250.

Liu, L., X. Ma, S. Liu, C. Zhu, L. Jiang, Y. Wang, Y. Shen, Y. Ren, H. Dong, L. Chen, X. Liu, Z. Zhao, H. Zhai, and J. Wan. 2009. Identification and characterization of a novel *waxy* allele from a Yunnan rice landrace. Plant Mol. Biol. 71:609-626.

Maughan, P.J., A. Bonifacio, E.N. Jellen, M.R. Stevens, C.E. Coleman, and M. Ricks. 2004. A genetic linkage map of quinoa (*Chenopodium quinoa*) based on AFLP, RAPD, and SSR markers. Theor. Appl. Genet. 109:1188–1195.

Maughan, P.J., B.A. Kolano, J. Maluszynska, N.D. Coles, A. Bonifacio, J. Rojas, C.E. Coleman, M.R. Stevens, D.J. Fairbanks, S.E. Perkinson, E.N. Jellen. 2006. Molecular and cytological characterization of ribosomal RNA genes in *Chenopodium quinoa* and *Chenopodium berlandieri*. Genome. 49:825-839.

Morales, A.J. 2009. Physiological assessment of *Chenopodium quinoa* to salt stress. M.S. thesis, Brigham Young Univ.

Morita, N., T. Maeda, M. Miyazaki, M. Yamamori, H. Miura, and I. Ohtsuka. 2002. Dough and baking properties of high-amylose and *waxy* wheat flours. Cereal Chem. 79:491-495.

Nakamura, T., M. Yamamori, H. Hirano, S. Hidaka, and T. Nagamine. 1995. Production of *waxy* (amylose-free) wheats. Mol. Gen. Genet. 248:253-259.

Okagaki, R. 1992. Nucleotide sequence of a long cDNA from the rice *waxy* gene. Plant Mol. Biol. 19:513-516.

Park, Y., K. Nemoto, T. Nishikawa, K. Matsushima, M. Minami, and M. Kawase. 2009. Molecular cloning and characterization of granule bound starch synthase I cDNA from a grain amaranth (*Amaranthus cruentus L.*). Breeding. Sci. 59:351-360.

Park, Y., K. Nemoto, T. Nishikawa, K. Matsushima, M. Minami, and M. Kawase. 2010. *Waxy* strains of three amaranth grains raised by different mutations in the coding region. Mol. Breeding. 25:623-635.

Park, Y., K. Nemoto, T. Nishikawa, K. Matsushima, M. Minami, and M. Kawase. 2011. Genetic diversity and expression analysis of granule bound starch synthase I gene in the new world grain amaranth (*Amaranthus cruentus L.*). J. Cereal Sci. 53:298-305.

Patron, N.J., A.M. Smith, B.F. Fahy, C.M. Hylton, M.J. Naldrett, B.G. Rossnagel, and K. Denyer. 2002. The altered pattern of amylose accumulation in the endosperm of low-amylose barley cultivars is attributable to a single mutant allele of *granule-bound starch synthase I* with a deletion in the 5'-non-coding region. Plant Physiol. 130:190-198.

Punta, M., P.C. Coggill, R.Y. Eberhardt, J. Mistry, J. Tate, C. Boursnell, N. Prang, K. Forslund, G. Ceric, J. Clements, A. Heger, L. Holm, E.L.L. Sonnhammer, S.R. Eddy, A. Bateman, and R.D. Finn. 2012. The Pfam protein families database. Nucleic Acids Res. 40:290-301.

Qian J. and Kuhn M. (1999). Characterization of *Amaranthus cruentus* and *Chenopodium quinoa* starch. Starch-Starke. 51:116-120.

Risi, J.C., and N.W. Galwey. 1984. The *Chenopodium* grains of the Andes: Incan crops for modern agriculture. Adv. Appl. Biol. 10:145-216.

Rozen, S., and H.J. Skaletsky. 1998. Primer3 on the WWW for general users and biologist programmers. Methods Mol. Biol. 132:365.

Ruales, J., and B.M. Nair. 1992. Nutritional quality of the protein in quinoa (*Chenopodium quinoa* Willd.) seeds. Plant Food Hum. Nutr. 42:1-11.

Salehuzzaman, S., E. Jacobsen, and R. Visser. 1993. Isolation and characterization of a cDNA encoding granule-bound starch synthase in cassava (*Manihot esculenta Crantz*) and its antisense expression in potato. Plant Mol. Biol. 23:947–962.

Sano, Y. 1984. Differential regulation of *waxy* gene expression in rice endosperm. Theor. Appl. Genet. 64:467-473.

Sederberg, M.C. 2008. Physical mapping of ribosomal RNA genes in new world members of the genus *Chenopodium* using fluorescence in situ hybridization. M.S. thesis, Brigham Young Univ.

Sun, Y., F. Liu, M. Bendevis, S. Shabala, and S. Jacobsen. 2014. Sensitivity of two quinoa (*Chenopodium quinoa Willd*.) varieties to progressive drought stress. J. Agron. Crop Sci. 200:12-23.

Todd, J.J., and L.O. Vodkin. 1996. Duplications that suppress and deletions that restore expression from a chalcone synthase multigene family. Plant Cell. 8:687-699.

Tsai, C.Y. 1974. The function of *waxy* locus in starch synthesis in maize endosperm. Biochem. Genet. 11:83-96.

van der Leij, F., R. Visser, A. Ponstein, E. Jacobsen, and W. Feenstra. 1991. Sequence of the structural gene for granule-bound starch synthase of potato (*Solanum tuberosum L.*) and evidence for a single point deletion in the amf allele. Mol. Gen. Genet. 228:240-248.

Walters, T.W. 1988. Relationship between isozymic and morphologic variations in the diploids *Chenopodium fremontii, C. neomexicanum, C. palmeri*, and *C. watsonii*. Am. J. of Bot. 75:97-105.

Ward, S.M. 2000. Allotetraploid segregation for single-gene morphological characters in quinoa (*Chenopodium quinoa Willd*.) Euphytica. 116:11-16.

Watanabe, K., A. Ibuki, Y. Chen, Y. Kawamura, and T. Mitsunaga 2003. Composition of quinoa protein fractions. J. Jpn. Soc. Food Sci. Technol. 50:546-549.

Watanabe, K., N.L. Peng, H. Tang. And T. Mitsunaga 2007. Molecular structural characteristics of quinoa starch. Food Sci. and Technol. Res. 13:73-76.

Wheelen, S.J., D.M. Church, and J.M. Ostell. 2001. Spidey: a tool for mRNA-to-genomic alignments. Genome Res. 11:1952-1957.

Wilson, H.D. 1988. Allozyme variation and morphological relationships of *Chenopodium hircinum*. Syst. Bot. 13:215-228.

Wilson, H.D. 1980. Artificial hybridization among species of *Chenopodium* sect. *Chenopodium*. Syst. Bot. 5:253-263.

Wilson, H.W., and C.B. Heiser. 1979. The origin and evolutionary relationships of 'huauzontle' (*Chenopodium nuttaliae Safford*), domesticated chenopod of Mexico. Am. J. of Bot. 66:198-206.

Wolf, M.J., M.M. MacMasters, and G.E. Rist. 1950. Some characteristics of the starches of three South American seeds used for food. Cereal Chem. 27:219-222.

Wright, K.H., O.A. Pike, D.J. Fairbanks, and C.S. Huber. 2002. Composition of *Atriplex hortensis*, sweet and bitter *Chenopodium quinoa* seeds. J. Food Sci. 67:1383-1385.

Yeku, O., and M.A. Frohman. 2011. Radip amplification of cDNA ends (RACE). Methods Mol. Biol. 703:107-122.

Chapter 2: Literature review of starch biosynthesis and the GBSSI *gene in plants*

Introduction to Starch

Starch is the most common and important carbohydrate energy source in the human diet. It accounts for a significant portion of global caloric intake, making up as much as 80% of the daily calories consumed in some parts of the world (Keeling and Myers, 2010). In addition to being an essential food source for humans, starch is also important for feeding livestock and has a myriad of other industrial uses. Recently, starch has become one of the world's most important sources of renewable energy as a main contributor to biofuel production (Keeling and Myers, 2010).

Starch is a polysaccharide, made up of α 1-4-linked glucose molecules, that is produced by plants to store energy. There are two different types of plant starch: amylose and amylopectin. Amylose exists as linear chains and normally makes up ~20-30% of total starch content. Due to its consistent, repeating structure of long, straight, overlapping glucose units, amylose is insoluble in water. Unlike amylose, amylopectin forms loose molecular aggregates due to periodic α 1-6-linked branches that allow it to dissolve in water. Amylopectin makes up the remaining 70-80% of total starch content (Jeon et al., 2011; Smith, 2001). Because of the molecular, structural, and compositional differences between amylose and amylopectin, starches with different amylose:amylopectin concentration ratios will stain differently in the presence of iodine. Normal starches stain blue, low-amylose starches bluish-purple, and no-amylose starches reddish-brown in the presence of I_2/KI. (Hunt et al., 2010).

Localization of Starch

Starch is found throughout different areas of the plant in the form of either transitory starch or seed starch. Transitory starch is used as a short-term energy source and is synthesized

and metabolized on a regular basis. Seed starch is present in the seeds, where it is kept as an energy store for the embryo to use during development. Because seed/endosperm starch is not metabolized except by the embryo, it is more stable, more accessible, and is of greater economic value than transitory starch.

The endosperm forms as the result of one of the two pollen sperm nuclei fusing with the two polar nuclei in the embryo sac, forming the primary endosperm cell with its triploid nucleus. The primary endosperm cell then divides and develops through the syncytial phase, the cellular phase, and the maturation phase (Li and Berger, 2012). The syncytial phase begins when the endosperm undergoes mitoses without going through cell division, leading to the endosperm containing as many as several hundred nuclei in one cell. The syncytial phase ends with the nuclei separating into individual cells (Li and Berger, 2012). In the cellular phase, cells continue to divide and differentiate into four distinct cell types: transfer cells, aleurone cells, starchy endosperm cells, and embryo-surrounding region cells (Sabelli and Larkins, 2009). Each of these cell types serves a unique function in the seed. In the final phase, maturation, the endosperm stops cell differentiation and division, changes its metabolism, and begins to accumulate energy reserves, including proteins, lipids, and starch. Endosperm development concludes with programmed cell death of all the endosperm cells except for the aleurone and is induced by increased concentrations of the plant hormones ethylene and abscisic acid (Sabelli and Larkins, 2009).

Starchy endosperm cells make up the bulk of the endosperm and contain most of the storage molecules and nutrients that the embryo needs for early development. The transition from cell division to storage molecule production in the starchy endosperm cells is accomplished by a dramatic change in gene expression and is triggered by changes in the relative sucrose and

glucose concentrations. When glucose levels are high relative to sucrose levels, endosperm cells continue to divide. When the relative sucrose (from which starch is synthesized) levels are high, endosperm cells stop dividing and begin synthesizing storage compounds (Sabelli and Larkins, 2009).

Within the endosperm, starch synthesis and storage takes place within organelles called amyloplasts. Amyloplasts are a non-pigmented, specialized type of plastid categorized as luecoplasts (Yun and Kawagoe, 2009). Inside the amyloplasts, starch is stored as water insoluble granules, which efficiently package the starch (Smith, 2001). Amylopectin molecules are the major contributors to the granule structure, forming clustering arrangement that contribute to the semicrystalline structure of the starch granules. Adjacent amylopectin chains form double helices with each other and pack together into ordered crystalline lamellae. These crystalline layers alternate with amorphous layers where the amylopectin branch points occur, giving the granule the appearance of growth rings. Amylose is thought to accumulate mostly in the amorphous layers, and different concentrations of amylose and amylopectin can influence the granule structure.

Starch Biosynthesis

Generally, the biochemical pathway for starch biosynthesis relies on several different enzymes, including adenosine 5' diphosphate-glucose pyrophosphorylase (AGPase), granule-bound starch synthase (GBSS), soluble starch synthase (SS), starch branching enzymes (BE), and starch debranching enzymes (DBE) (Jeon et al., 2010; Figure 8). A myriad of genes contribute to starch synthesis, as most of these enzymes have several isoforms. There appears to be little functional redundancy, as most of the isoforms are expressed at different periods in

development and have distinct mutant phenotypes (Jeon et al., 2010; Hirose and Terao, 2004; Yan et al., 2009; Ding et al., 2009).

Starch biosynthesis begins with AGPase producing ADP-glucose from glucose-1-phosphate. This first step requires the use of ATP and is the rate limiting step for starch synthesis (Wang et al., 2007). The AGPase enzyme exists as a heterotetramer of two large subunits and two small subunits, and both cytosolic and plastidial versions of the enzyme exist (Jeon et al., 2010). ADP-glucose produced in the cytosol must be transported into the plastid before it can be synthesized into starch.

Inside the plastid, ADP-glucose is added to the non-reducing end of existing starch chains by either GBSS or SS (Tetlow 2011). GBSS synthesizes amylose, while the SS enzymes synthesize amylopectin molecules of differing lengths. There are two forms of the GBSS enzyme. GBSSII is found throughout the plant and is essential for the production of the amylose portion of transitory starch. GBSSI is found in developing seeds and is responsible for synthesizing amylose in seed starch (Jeon et al., 2010; Tetlow, 2011) *GBSSI* mutants have *waxy* starch that lacks amylose.

There are at least four SS classes, and each, with the exception of SSI, has its own combination of isoforms. Despite its lack of isoforms, *SSI* is expressed steadily through endosperm development and is the most active SS enzyme, accounting for up to 70% of SS activity in rice endosperm (Jeon et al., 2010). By analyzing SSI deficient mutants in rice, Fujita et al. (2006) found that SSI binds to and extends the shortest amylopectin chains. SSII has three isoforms in the cereal grains, but their functions are not fully understood. Of the two isoforms of SSIII, SSIIIb synthesizes transitory starch, while SSIIIa synthesizes starch in the endosperm (Hirose and Terao, 2004). Mutations in *SSIIIa* produce the *dull1* phenotype in maize and the *flo5*

phenotype in rice (Jeon et al., 2010). SSIIIa is the most active SS enzyme in the endosperm after SSI, and plants that have reduced SSIIIa activity have an increased concentration of amylopectin chains of medium length, suggesting that SSIIIa elongates medium-length chains into longer chains (Jeon et al., 2010). *SSIV* mutants in *Arabidopsis* showed decreased levels of leaf starch, a normal amylose to amylopectin ratio, and normal chain length distributions, suggesting that SSIV may have a function different than the typical amylopectin elongation of the other SS enzymes (Roldan et al., 2007). *Arabidopsis SSIV* mutants also showed severe growth defects, a decreased number of starch granules, and an increase in starch granule size, suggesting that SSIV may play a role in the formation and/or use of transitory starch as well as in starch granule formation. (Roldan et al., 2007).

BE introduce the 1,6 branch points that are characteristic of amylopectin. There are two classes of BE. The rice BEI mutant *sbe1* shows a decrease in the concentration of longer amylopectin chains and an increase in the concentration of intermediate and small sized chains (Satoh et al., 2003). Of the BEII isoforms, BEIIa is expressed throughout the plant, while BEIIb is expressed mostly in the endosperm (Tetlow, 2011). *BEIIb* mutants are also known as *amylose extender* (*ae*) mutants because they have starch with a high concentration of amylose (Sestili et al., 2010). This suggests that BEIIb has an especially important role in starch branching.

DBE systematically remove excess branch points to ensure proper amylopectin structure. There are at least three genes for isoamylase (ISA) type DBE, and a single gene for pullulanase (PUL) type DBE in rice (Jeon et al., 2010). Although it seems counter-intuitive, debranching is critical for normal starch biosynthesis. Mutations in DBE genes lead to the accumulation highly and randomly branched chains of glucans called phytoglycogen (Fujita et al., 2009).

Between AGPase, GBSS, SS, BE, and DBE, and their various isoforms, starch biosynthesis is an intricate and sensitive process. Recent evidence has shown that individual starch synthesis enzymes may complex with each other, making the process even more complicated. Protein-protein interactions between SS, BE, and DBE have all been reported (Lin et al., 2012; Tetlow et al., 2008; Tetlow et al., 2004).

Starch Metabolism and Mobilization

In seeds, starch metabolism begins when the embryo first begins to grow. The embryo releases gibberellin hormones that stimulate the living aleurone layer of the endosperm to synthesize and release various proteins and enzymes. These proteins and enzymes help to break down and mobilize the nutrients in the endosperm so that they can be used by the embryo (Sabelli and Larkins, 2009). Compared to starch biosynthesis, starch metabolism is relatively straightforward. Within the plastid, starch is broken down into glucose, maltose, and other sugars by DBE, α-amylase, and β-amylase. The sugars are then carried out of the plastids and into the cytosol by transporter proteins and diffusion, where they feed into the plant's metabolic cycles (Stitt and Zeeman, 2012).

Starch Research

Because starch feeds the world and has a suite of important industrial applications, a finer understanding of the genetic controls behind starch biosynthesis would be invaluable for multiple reasons. Enhanced starch biosynthesis has a large effect on the yield of crops that use starch as their main energy storage molecule (Jeon et al., 2010; Wang et al., 2007). A better understanding of starch biosynthesis could therefore lead to increased crop yields that consequently could help feed a world with a growing population. Starches with different molecular properties and amylose to amylopectin concentration ratios have different uses for cooking, eating, and

industrial processing (Jeon et al., 2010). Resistant starches, such as those with an increased amylose concentration, have a lower glycemic index and may improve digestive tract health and blood cholesterol levels (Fuentes-Zaragoza et al., 2010). Low-amylose, or *waxy*, starches possess desired traits that influence cooking quality and seed softness, especially in rice (Liu et al., 2009). If the genetics behind starch biosynthesis were better understood, plants having starches with specific properties could be developed so that starch could require less processing and could be used more widely and more efficiently. Starch biosynthesis is also related to carbon and resource partitioning, seed germination and viability, and other plant processes and functions. For these reasons, a more complete understanding of the controls behind starch synthesis would be scientifically and economically valuable, and while the biochemical pathways are fairly well understood and available for review, detailed information on specific genes in specific plants is lacking.

GBSSI

There are more than twenty genes governing starch biosynthesis, mobilization, and metabolism in plants, but one of the most important is *Granule-bound Starch Synthase I* (*GBSSI*). Of the two GBSS isoforms, GBSSI and GBSSII, GBSSI is more important for seed development. In general, both proteins can be found throughout the plant, but GBSSI is typically the only one of the two found in the endosperm, and it therefore plays a more significant role in synthesizing seed starch (Hirose and Terao, 2004; Vrinten and Nakamura, 2000; Young-Jun et al., 2011). *Waxy* mutants have a defective *GBSSI* gene, and although *waxy* seeds have a normal weight and amount of starch, *waxy* starch contains little or no amylose (Jeon et al., 2010) *Waxy* starches are important because they have unique properties, uses, and health benefits (Fuentes-Zaragoza et al., 2010). In particular, *waxy* and low-amylose starches are desired for their

influence on cooking quality and seed softness, especially in rice (Liu et al., 2009). In addition, *GBSSI* appears to influence and interact with other starch synthesis genes, specifically *SSIIIa* (Fujita et al., 2011; Jeon et al., 2010). Characterization of *GBSSI* could therefore facilitate the study of the interactions between starch synthesis genes and aid in the generation of novel high and low-amylose starches. Furthermore, analysis of *GBSSI* between species can help to answer evolutionary questions.

GBSSI *in* Arabidopsis

Arabidopsis thaliana is often used as a model organism in plant biology, but research and information on *Arabidopsis* starch is lacking. This is perhaps because so much starch research has already been conducted on the more relevant food crops, but is more likely because *Arabidopsis* seeds do not accumulate starch (Andriotis et al., 2010). *GBSSI* has never been specifically characterized in *Arabidopsis*, nor has a *waxy* mutant been characterized.

GBSSI *in Rice*

The *GBSSI* gene in rice is comprised of 13 exons and the genomic sequence spans roughly 4000 base pairs (Sano, 1991). Expression is limited to the seed and the pollen grain. *GBSSI* in rice has two main functional alleles, Wx^a, and Wx^b (Hirano et al., 1998). Wx^a is a normal, functional allele, but Wx^b, due to a mutation at the 5' splice site of the first intron, has reduced transcription which leads to lower amylose content (Jeon et al., 2010). Liu et al. (2009) describe a novel *GBSSI* gene, Wx^{hp}, in rice in which an A to G change results in an aspartate to glycine substitution at codon 165. Although this small change did not lead to decreased GBSSI activity *in vitro*, it did noticeably reduce the ability of GBSSI to bind to starch granules, and thereby reduced amylose content *in vivo*. The Wx^b and Wx^{hp} alleles establish a precedent for *GBSSI* genes to have reduced transcription without completely losing functionality.

Hanashiro et al. (2008) demonstrated that GBSSI in rice synthesizes extra-long unit chains (ELC) of amylopectin in addition to synthesizing amylose. Hanashiro et al. inserted a functional copy of *GBSSI* into *waxy* rice and isolated the ELC portion of the starch. The transgenic *waxy* rice starch was 7.5-8.4% ELC, whereas non-transgenic *waxy* rice starch had no ELC. Based on this information, Hanashiro et al. conclude that the GBSSI protein plays a role in ELC synthesis in addition to amylose synthesis. These results are consistent with studies that have shown that *waxy* wheat starches have no ELC and that increasing *GBSSI* expression can increase ELC levels (Jeon et al., 2010).

An expression analysis conducted by Hirose and Terao (2004) found that *GBSSI* in rice is expressed during the middle to late stage of seed development, between 5 and 20 days after fertilization, and that the expression of *GBSSI* is almost entirely specific to the developing rice seed. Another expression analysis by Ohdan et al. (2005) confirmed these results, but added that starting five days after fertilization, *GBSSI* transcript levels in developing rice seeds rise dramatically and remain elevated until the end of endosperm development. *GBSSI* transcript levels were reported to be nearly 300 times greater than the next most abundant starch synthase transcripts.

GBSSI *in Maize*

GBSSI, the *waxy* locus, has been extensively studied in maize. The wild-type maize genomic gene sequence is 3,718 base pairs in length, and is comprised of 14 exons and 13 introns (Klösgen et al., 1968). Over 50 *waxy* mutants have been identified in maize (Huang et al., 2010). Most mutants are the result of simple insertions and/or deletions, but the insertion of transposable elements also accounts for some of the mutations.

GBSSI activity in developing maize seeds increases during the time between 10 and 25 days after pollination and it reaches its peak at 25 days after pollination (Guo et al., 2006). *GBSSI* activity decreases from that point on, though amylose concentrations take until 45 days after pollination to obtain their highest levels. The proportion of amylose in the seed increases gradually during seed development, while the proportion of amylopectin decreases (Guo et al., 2006).

GBSSI *in* Amaranthus

The *GBSSI* gene was recently characterized in grain amaranth (genus *Amaranthus*), a psuedocereal crop that is closely related to *Chenopodium*. *GBSSI* in amaranth has 13 exons and 12 introns, and has an average length of 3,236 base pairs across the three analyzed species, *A. caudatus*, *A. cruentus*, and *A. hypochondriacus*. (Park et al., 2010). Park et al. (2010) also analyzed *GBSSI* mutants in the same three species of *Amaranthus* and found that in all cases, the mutation was caused by a simple, single-base substitution or insertion. These small changes led to nonsense or frameshift mutations that resulted in non-functional GBSSI proteins. *Waxy A. caudatus* had a T inserted in exon eight. *Waxy A. cruentus* had a T substituted for a G in exon ten, and *waxy A. hypochondriacus* had an A substituted for a G in exon six (Park et al., 2010). *GBSSI* sequence conservation between different species of *Amaranthus* was very high. *GBSSI* in *Amaranthus* is expressed throughout the plant during all developmental periods, but is more strongly expressed in the seed during later periods of seed development.

GBSSI *in Quinoa*

GBSSI has never been sequenced and characterized in *Chenopodium*, but Lindeboom et al. (2005) examined amylose concentration and the GBSSI protein and its activity in different lines of *Chenopodium quinoa*. Amylose concentrations in certain accessions of quinoa range

from 3% to 20%, but average around 7% (Lindeboom et al., 2005; Qian et al., 1999; Tang et al., 2001). This categorizes quinoa as having low-amylose and possibly *waxy* starches. This variability in amylose concentration also suggests that different lines of quinoa may differentially express *GBSSI*, and/or have copies of the GBSSI protein that are either less functional or non-functional. Lindeboom et al. identify a possible GBSSI isoform which, at 56 kDA, contrasts the normal 62 kDA GBSSI protein.

Apart from this study, there has been no evidence of additional GBSSI isoforms in quinoa. In fact, the low-amylose content of quinoa seeds points to fewer functional copies of *GBSSI*, not to more. It is possible that the 56kDA protein identified as an isoform of GBSSI is actually a non-functional, truncated version of the protein. Alternatively, the lighter protein could represent a mature protein, while the heavier, 62 kDa protein could represent an immature version, complete with transit peptide. In either case, the work of Lindeboom et al. (2005) paves the way for further investigation of the quinoa *GBSSI* gene.

References

Andriotis, V.M.E., M.J. Pike, B. Kular, S. Rawsthorne, A.M. Smith. 2010. Starch turnover in developing oilseed embryos. New Phytol. 187:791-804

Ding, X., B. Wang, Q. Gao, Q. Zhang, G. Yan, K. Duan, and J. Huang. 2009. Molecular diversity and differential expression of starch-synthesis genes in developing kernels of three maize inbreds. Plant Cell Rep. 28:1487-1495.

Fuentes-Zaragoza, E., M.J. Riquelme-Navarrete, E. Sánchez-Zapata, and J.A. Pérez-Álvarez. 2010. Resistant starch as functional ingredient: A review. Food Res. Int. 43:931-942.

Fujita, N., R. Satoh, A. Hayashi, M. Kodama, R. Itoh, S. Aihara, and Y. Nakamura. 2011. Starch biosynthesis in rice endosperm requires the presence of either *starch synthase I or IIIa*. J. Environ. Bot. 62:4819-4831.

Fujita, N., Y. Toyosawa, Y. Utsumi, T. Higuchi, I. Hanashiro, A. Ikegami, S. Akuzawa, M. Yoshida, A. Mori, K. Inomata, R. Itoh, A. Miyao, H. Hirochika, H. Satoh, and Y. Nakamura. 2009. Characterization of pullulanase (PUL)-deficient mutants of rice (*Oryza sativa L.*) and the function of PUL on starch biosynthesis in the developing rice endosperm. J. Exp. Bot. 60:1009-1023.

Fujita, N., M. Yoshida, N. Asakura, T. Ohdan, A. Miyao, H. Hirochika, and Y. Nakamura. 2006. Function and characterization of *starch synthase I* using mutants in rice. Plant Physiol. 140:1070.

Guo, S., J. Li, W. Qiao, and X. Zhang. 2006. Analysis of amylose accumulation during seed development in maize. Acta Genet. Sin. 33:1014-1019.

Hanashiro, I., K. Itoh, Y. Kuratomi, M. Yamazaki, T. Igarashi, J.I. Matsugasako, and Y. Takeda. 2008. Granule-bound starch synthase I is responsible for biosynthesis of extra-long unit chains of amylopectin in rice. Plant and Cell Physiol. 49:925-933.

Hirano, H.Y., M. Eiguchi, and Y. Sano. 1998. A single-base change altered the regulation of the *Waxy* gene at the posttranscriptional level during the domestication of rice. Mol. Biol. Evol. 15:978.

Hirose, T. and T. Terao. 2004. A comprehensive expression analysis of the starch synthase gene family in rice (*Oryza sativa L.*). Planta 220:9-16.

Huang, B., M. Tian, J. Zhang, and Y. Huang. 2010. *Waxy* locus and its mutant types in maize (*Zea mays L.*). Agr. Sci. China. 9:1-10.

Hunt, H.V., K. Denyer, L.C. Packman, M.K. Jones, and C.J. Howe. 2010. Molecular basis of the waxy endosperm starch phenotype in broomcorn millet (Panicum miliaceum L.). Mol. Biol. Evol. 27(7):1478-1494.

Jeon, J., N. Ryoo, T. Hahn, H. Walia, and Y. Nakamura. 2010. Starch biosynthesis in cereal endosperm. Plant Phisiol. Bioch. 48:383-392.

Keeling, P.L., and A.M. Myers. 2010. Biochemistry and genetics of starch synthesis. Ann. Rev. Food Sci. and Technol. 1:271-303.

Kloesgen, R.B., and A. Gierl. 1986. Molecular analysis of the *waxy* locus of *Zea mays*. Mol. and General Gen. 203:237-244.

Li, J., and F. Berger. 2012. Endosperm: food for humankind and fodder for scientific discoveries. New Phytol. 195:290-305.

Lin, Q., B. Huang, M. Zhang, X. Zhang, J. Rivenbark, R.L. Lappe, M.G. James, A.M. Myers, and T.A. Hennen-Bierwagen. 2012. Functional interactions between starch synthase III and isoamylase-type starch-debranching enzyme in maize endosperm. (Report). Plant Physiol. 158:679.

Lindeboom, N., P.R. Chang, R.T. Tyler, and R.N. Chibbar. 2005. Granule-bound starch synthase I (GBSSI) in quinoa (*Chenopodium quinoa Willd*.) and its relationship to amylose content. Cereal Chem. 82:246-250.

Liu, L., X. Ma, S. Liu, C. Zhu, L. Jiang, Y. Wang, Y. Shen, Y. Ren, H. Dong, L. Chen, X. Liu, Z. Zhao, H. Zhai, and J. Wan. 2009. Identification and characterization of a novel *waxy* allele from a Yunnan rice landrace. Plant Mol. Biol. 71:609-626.

Ohdan, T., P.B.J. Francisco, T. Sawada, T. Hirose, T. Terao, H. Satoh, and Y. Nakamura. 2005. Expression profiling of genes involved in starch synthesis in sink and source organs of rice. J. Exp. Bot. 56:3229-3244.

Park, Y., K. Nemoto, T. Nishikawa, K. Matsushima, M. Minami, and M. Kawase. 2010. *Waxy* strains of three amaranth grains raised by different mutations in the coding region. Mol. Breeding. 25:623-635.

Park, Y., K. Nemoto, T. Nishikawa, K. Matsushima, M. Minami, and M. Kawase. 2011. Genetic diversity and expression analysis of granule bound starch synthase I gene in the new world grain amaranth (*Amaranthus cruentus L*.). J. Cereal Sci. 53:298-305.

Qian J. and Kuhn M. (1999). Characterization of *Amaranthus cruentus* and *Chenopodium quinoa* starch. Starch-Starke. 51:116-120.

Roldan, I., F. Wattebled, M.M. Lucas, D. Delvalle, V. Planchot, S. Jimenez, R. Perez, S. Ball, C. D'Hulst, and A. Merida. 2007. The phenotype of soluble starch synthase IV defective mutants of *Arabidopsis thaliana* suggests a novel function of elongation enzymes in the control of starch granule formation. The Plant J. 49:492-504

Sabelli, P.A., and B.A. Larkins. 2009. The development of endosperm in grasses. (update on endosperm development in grasses) (report). Plant Physiol. 149:14.

Sano, Y. 1991. Molecular characterization of the *waxy* locus of rice *Oryza sativa*. Plant and Cell Physiol. 32:989-998.

Satoh, H. 2003. Starch-branching enzyme I-deficient mutation specifically affects the structure and properties of starch in rice endosperm. Plant Physiol. 133:1111-1121.

Sestili, F., M. Janni, A. Doherty, E. Botticella, R. D'Ovidio, S. Masci, H. Jones, and D. Lafiandra. 2010. Increasing the amylose content of durum wheat through silencing of the *SBEIIa* genes. BMC Plant Biol. 10:144.

Smith, A.M. 2001. The biosynthesis of starch granules. Biomacromol. 2:335.

Stitt, M., and S.C. Zeeman. 2012. Starch turnover: pathways, regulation and role in growth. Curr. Opin. Plant Biol. 15:282-292.

Tang, H., K. Watanabe, and T. Mitsunaga. 2002. Characterization of storage starches from quinoa, barley and adzuki seeds. Carbohydr. Polym. 49:13-22.

Tetlow, I.J. 2011. Starch biosynthesis in developing seeds. Seed Sci. Res. 21:5-32.

Tetlow, I.J., K.G. Beisel, S. Cameron, A. Makhmoudova, F. Liu, N.S. Bresolin, R. Wait, M.K. Morell, and M.J. Emes. 2008. Analysis of protein complexes in wheat amyloplasts reveals functional interactions among starch biosynthetic enzymes. Plant Physiol. 146:1878.

Tetlow, I.J., M.K. Morell, and M.J. Emes. 2004. Recent developments in understanding the regulation of starch metabolism in higher plants. J. Exp. Bot. 55:2131-2145.

Vrinten, P.L., and T. Nakamura. 2000. Wheat granule-bound starch synthase I and II are encoded by separate genes that are expressed in different tissues. Plant Physiol. 122:255-263.

Wang, Z., X. Chen, J. Wang, T. Liu, Y. Liu, L. Zhao, and G. Wang. 2007. Increasing maize seed weight by enhancing the cytoplasmic ADP-glucose pyrophosphorylase activity in transgenic maize plants. Plant Cell: Tiss. Organ Cul. 88:83-92.

Yan, H., X. Pan, H. Jiang, and G. Wu. 2009. Comparison of the starch synthesis genes between maize and rice: copies, chromosome location and expression divergence. Theor. and App. Gen. 119:815-825.

Yun, M., and Y. Kawagoe. 2009. Amyloplast division progresses simultaneously at multiple sites in the endosperm of rice. Plant and Cell Physiol. 50:1617-1626.

TABLES

Table 1. World quinoa production (thousand metric tons) (FOASTAT, 2014)

Country	1961	1970	1980	1990	2000	2010	2012
Peru	22.5	7.3	16.3	6.3	28.2	41.1	44.2
Bolivia	9.2	9.7	8.9	16.1	23.8	36.1	37.5
Ecuador	0.7	0.7	0.5	0.7	0.7	0.9	0.8
Total	32.4	17.7	25.7	23.1	52.7	78.1	82.5

Table 2. Selected germplasm used in this study

Species	Accession	Origin	Ploidy	Source
Chenopodium quinoa	0654	Peru	2*n*=4*x*=36	CIP-FAO
C. quinoa	Chucapaca	Bolivia	2*n*=4*x*=36	CIP-FAO
C. quinoa	Ollague	Chile	2*n*=4*x*=36	CIP-FAO
C. quinoa	G205-95	Peru	2*n*=4*x*=36	CIP-FAO
C. quinoa	NL6	Chile	2*n*=4*x*=36	CIP-FAO
C. quinoa	KU2	Chile	2*n*=4*x*=36	CIP-FAO
C. berlandieri	BYU 652	Utah, US	2*n*=4*x*=36	BYU Collection
C. b. var. *macrocalycium*	BYU 803	Maine, US	2*n*=4*x*=36	BYU Collection
C. b. var. *boscianum*	BYU 937	Texas, US	2*n*=4*x*=36	BYU Collection
C. b. nuttalliae	H04	Mexico	2*n*=4*x*=36	E. De la Cruz
C. b. nuttalliae	H02	Mexico	2*n*=4*x*=36	E. De la Cruz
C. hircinum	BYU 1101	Argentina	2*n*=4*x*=36	D. Bertero
C. pallidicaule	BYU 1302	Bolivia	2*n*=2*x*=18	BYU Collection
C. standleyanum	BYU 921	Illinois, US	2*n*=2*x*=18	BYU Collection
C. dessicatum	BYU 835	Nevada, US	2*n*=2*x*=18	BYU Collection
C. hians	BYU 1005	California, US	2*n*=2*x*=18	BYU Collection
C. neomexicanum	BYU 843	New Mexico, US	2*n*=2*x*=18	BYU Collection
C. ficifolium	BYU 943	Czech Republic	2*n*=2*x*=18	H. Storchova
Atriplex hortensis	BYU 1436	Utah	2*n*=2*x*=18	BYU Collection

Table 3. *Chenopodium* cloning, sequencing, and RT-PCR primers

Primer Set	Gene Amplified	Region Amplified	Tm	Primer Sequence
M20FA 2171RA	*GBSSIa*	~40 bp before start codon to ~1,830	49º C	TTGGTTGTTTGAGGTATTAGGACA TTACAAAAGCAAAAATCCTG
M80FB 1539RB	*GBSSIb*	~100 bp before start codon to ~1,280	49º C	TGGGGATATACAAACTCAAATCA CAACCAAGAAAACACCCCTA
646F 1718R	*GBSSIa/GBSSIb*	~650 to ~1,420	55º C	TTCCACACCTACAAGCGAGG CAGGCAAATGAAGACGCGAG
1454F 2223R	*GBSSIa/GBSSIb*	~1,190 to ~1,930	57º C	GGCATAGTGCTCTTCTCCCAGCC ACCAACTTCTGCTTGTAGGGCTTCC
2071F 2785R	*GBSSIa/GBSSIb*	~1,730 to 2,510	54º C	GGATGTGCAGGAATGGAATC GAAACGGCCCATATGGAATC
1820FA 3245RA	*GBSSIa*	~1,460 to ~20 bp after stop codon	49º C	GTAAAGTATCTTACCTACTAAATA CAGCCACATGAGTATTCCACA
1200FB 3415RB	*GBSSIb*	~1,280 to ~140 bp after stop codon	49º C	AGGGGTGTTTTCTTGGTTGT GCACTCATGACACATTAATAATCAAA
QA1278F QA1380R	*GBSSIa* (Quinoa CDS)	1,278 to 1,381	55º C	GCAGATTATTGTTCTTGGGACAGG ATTGAATTTGGTCACTCCTCTCG
QB1557F QB1750R	*GBSSIb* (Quinoa CDS)	1,557 to 1,751	55º C	CCGTTTCAGTGCAAATTGTGAC CAGCAACCCCTAGGCTCAG
HA868F HA1012R	*GBSSIa* (Huauzontle CDS)	868 to 1,013	55º C	GATGGGCACAACAAACCTGTAA TATCATTGAGCTCCACACCTCTCT
HB1008F HB1300R	*GBSSIb* (Huauzontle CDS)	1008 to 1,301 (*wt*) 771 to 1,064 (*mt*)	55º C	CGATGTAGTTCAGAGGACTGGA CAGTCCCAAGAACAATAATCTGCA
GAPDHF GAPDHR	*GAPDH*	-	55º C	GGTTACAGTCATTCAGACACCATCA AACAAAGGGAGCCAAGCAGTT

Table 4. *Atriplex hortensis* sequencing primers

Primer Set	Gene Amplified	Region Amplified	Tm	Primer Sequences
AtrM100F Atr1518R	*GBSSI*	~170 bp before start codon to 1,011	49º C	GGGACTACACATACAAACTGAGC TCACCATTCTTTCACGTCTTTT
646F 1718R	*GBSSI*	643 to 1,290	55º C	TTCCACACCTACAAGCGAGG CAGGCAAATGAAGACGCGAG
1454F 2223R	*GBSSI*	1,062 to 1,823	57º C	GGCATAGTGCTCTTCTCCCAGCC ACCAACTTCTGCTTGTAGGGCTTCC
2071F 2785R	*GBSSI*	1,625 to 2,410	54º C	GGATGTGCAGGAATGGAATC GAAACGGCCCATATGGAATC
Atr2604F Atr3425R	*GBSSI*	2,212 to ~170 bp after stop codon	49º C	CATGCAATGCGTTATGGAAC ACAGCACTCATGACACTTCAA

Table 5. *Chenopodium GBSSI* consensus sequence comparisons

Species	CDS Identity	Complete Protein Sequence Identity	Mature Protein Identity	Transit Peptide Identity	GC Content
Chenopodium Consensus	-	-	-	-	45.4%
Atriplex hortensis	93%	94.7%	96.6%	81.8%	43.8%
Beta vulgaris	85.1%	85%	86.6%	74.7%	40.4%
Amaranthus cruentus	82.3%	79.2%	82.2%	60.8%	44%
Nelumbo nucifera	71.6%	72.3%	77.3%	42.5%	43.8%
Theobroma cacao	70.7%	70.2%	75.6%	40.3%	46%
Gossypium hirsutum	70.6%	69.9%	75%	38.5%	45.1%
Nicotiana tabacum	70.3%	69.4%	75.2%	36.3%	44.9%
Ipomoea batatas	69.3%	68.5%	73.7%	36.4%	46.5%
Pisum sativum	69.4%	66.6%	72.9%	26.9%	41.3%
Solanum tuberosum	68.6%	68.2%	73.3%	32.5%	44.9%
Musa acuminata	68.2%	64.9%	72.5%	25%	46.3%
Arabidopsis thaliana	68.1%	65.9%	71%	34.6%	45.2%
Oryza sativa	63.3%	61.6%	68.8%	14.1%	66.1%
Hordeum vulgare	62.7%	60.1%	66.4%	4.1%	64.1%
Zea mays	61.9%	62%	68.5%	9.5%	66.2%
Triticum aestivum	61.1%	57.9%	64.3%	18.8%	64.3%
Escherichia coli	-	32.5%	-	-	-

Table 6. *Chenopodium* GBSSI alleles identified in this study

Allele	Accession/s	Mutation
GBSSIa-1	'Ollague', 'G205-95', 'NL6', 'KU2', 835	Reference Allele
GBSSIa-2	'0654', 'Chucapaca'	D393E
gbssia-tp	'H04', 'H02'	Transit Peptide Substitution, I54T, I325V, V456L
GBSSIa-3	937	I325V, V456L
GBSSIa-4	803	I325V, V456L, M555I
GBSSIa-5	652	I325V, V456L, N575D
GBSSIa-6	1101	I325V, V456L, L463S
GBSSIa-7	1302	V28I, R246P, V289F, I325V
GBSSIa-8	1005	K294Q
GBSSIa-9	921	K458R
GBSSIa-10	843	S274P
GBSSIb-1	'Ollague', '0654', 'NL6', 'KU2', 803, 1101	Reference Allele
GBSSIb-2	'Chucapaca'	R142M
gbssib-t	'G205-95'	Early Termination, W129X
GBSSIb-3	'H04'	T74P
gbssib-Δ	'H02'	Deletion, 64-142, A417E
GBSSIb-4	937	K55N
GBSSIb-5	652	K55N, T74P
GBSSIb-6	943	T531I

FIGURES

Figure 1. Gene structure of the consensus *Chenopodium GBSSI* gene

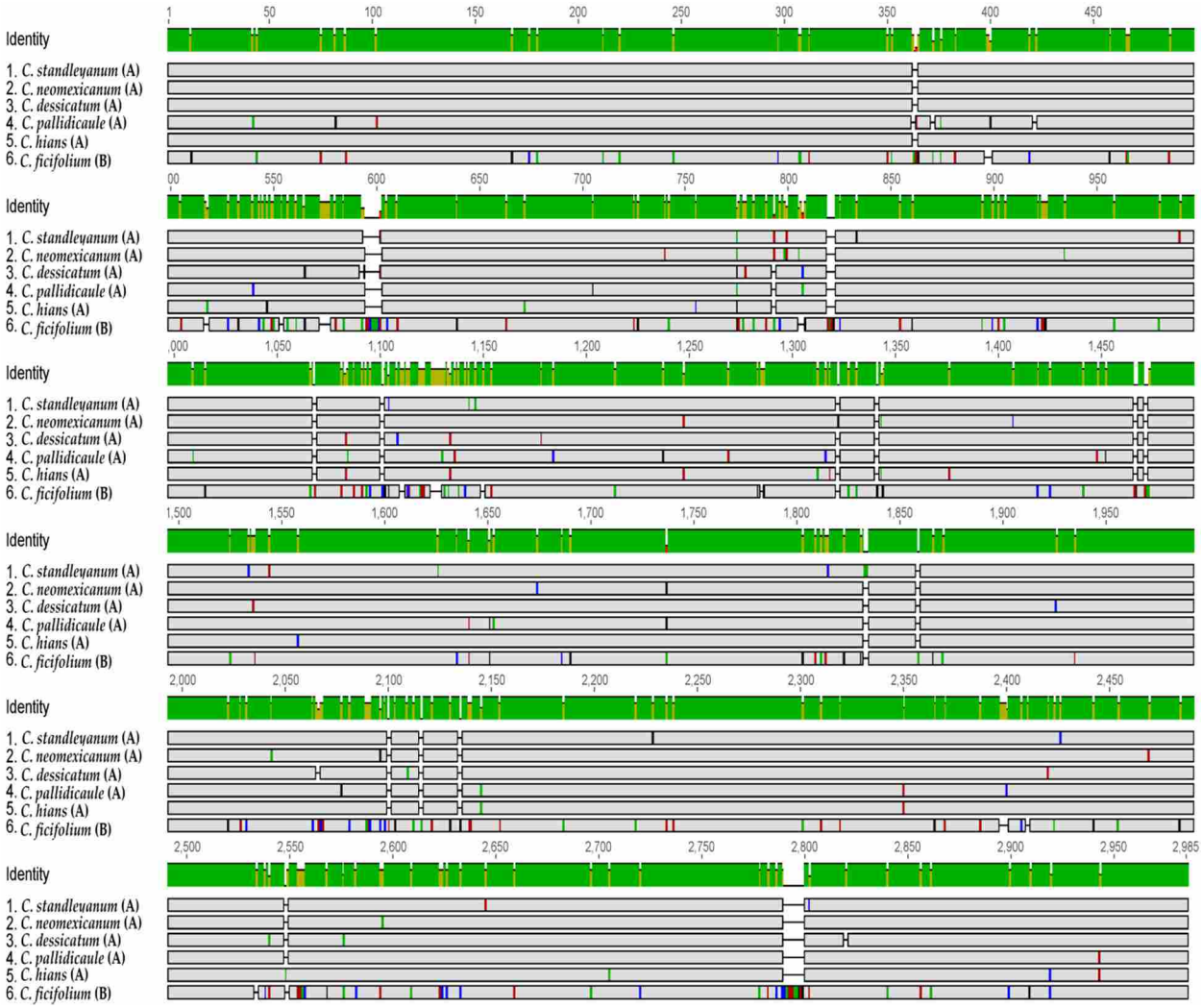

Figure 2. Aligned *GBSSI* sequences of New World *Chenopodium* diploids and *C. ficifolium*

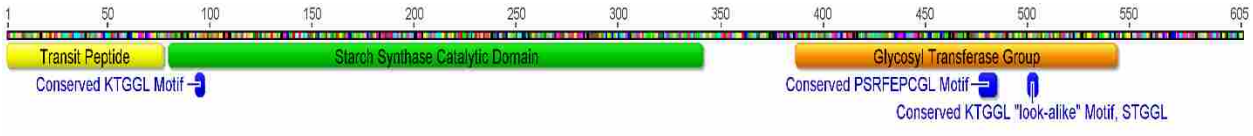

Figure 3: Generic structure of the *Chenopodium* GBSSI protein

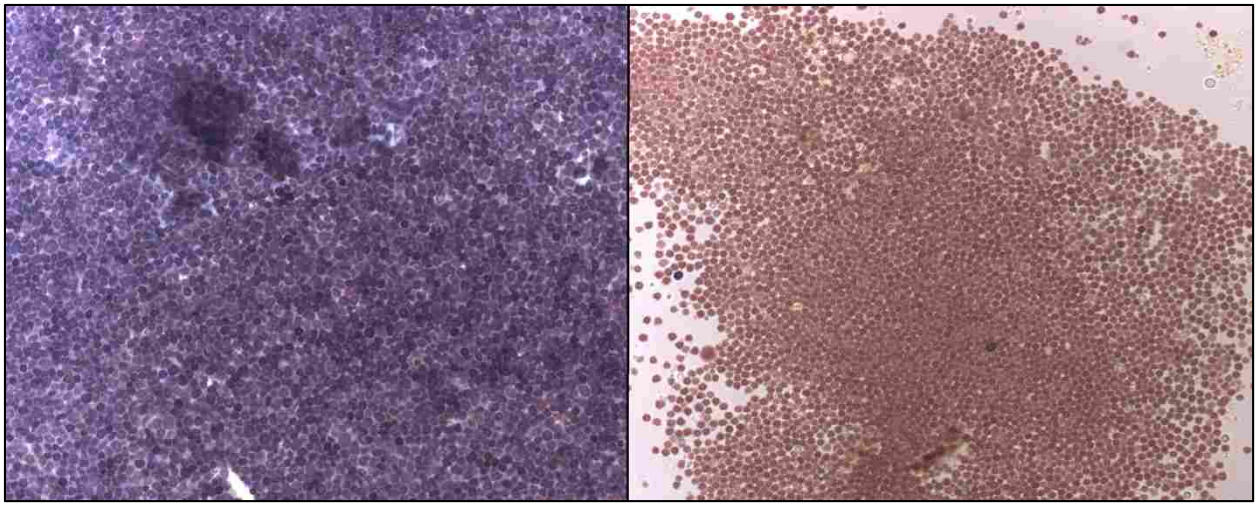

Figure 4. Comparison of non-*waxy* (‘H04’) and *waxy* (‘H02’) seed starch

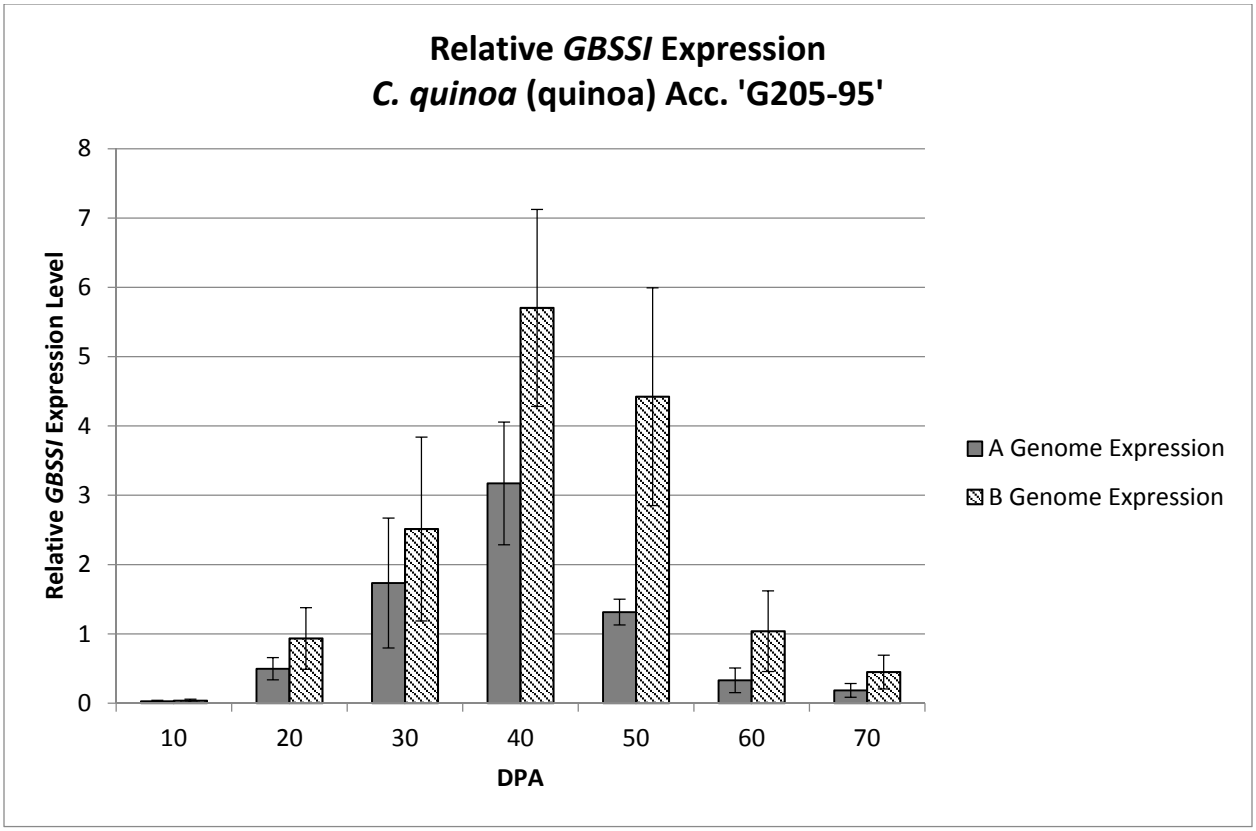

Figure 5. Expression profiles for *GBSSIa* and *GBSSIb* in the quinoa accession 'G205-95'

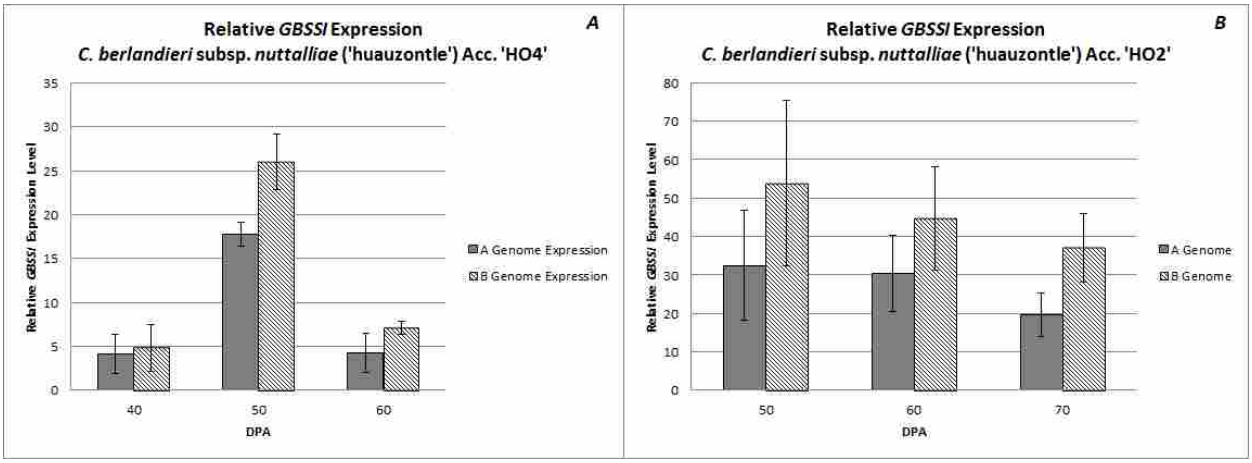

Figure 6. Expression profile for *GBSSIa* and *GBSSIb* in the huauzontle accessions 'H04' (A) and 'H02' (B)

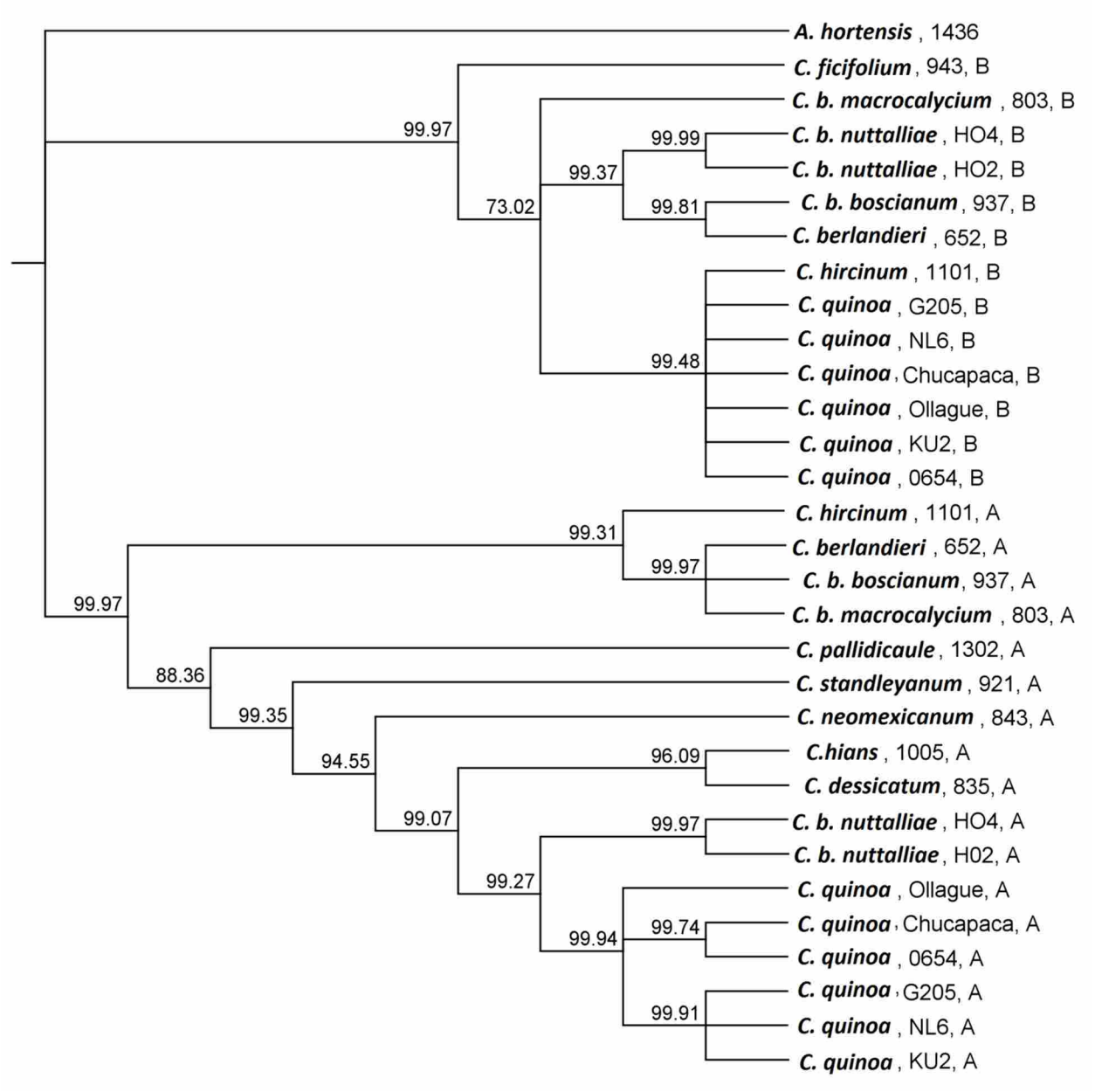

Figure 7. *Chenopodium GBSSI* phylogenetic tree

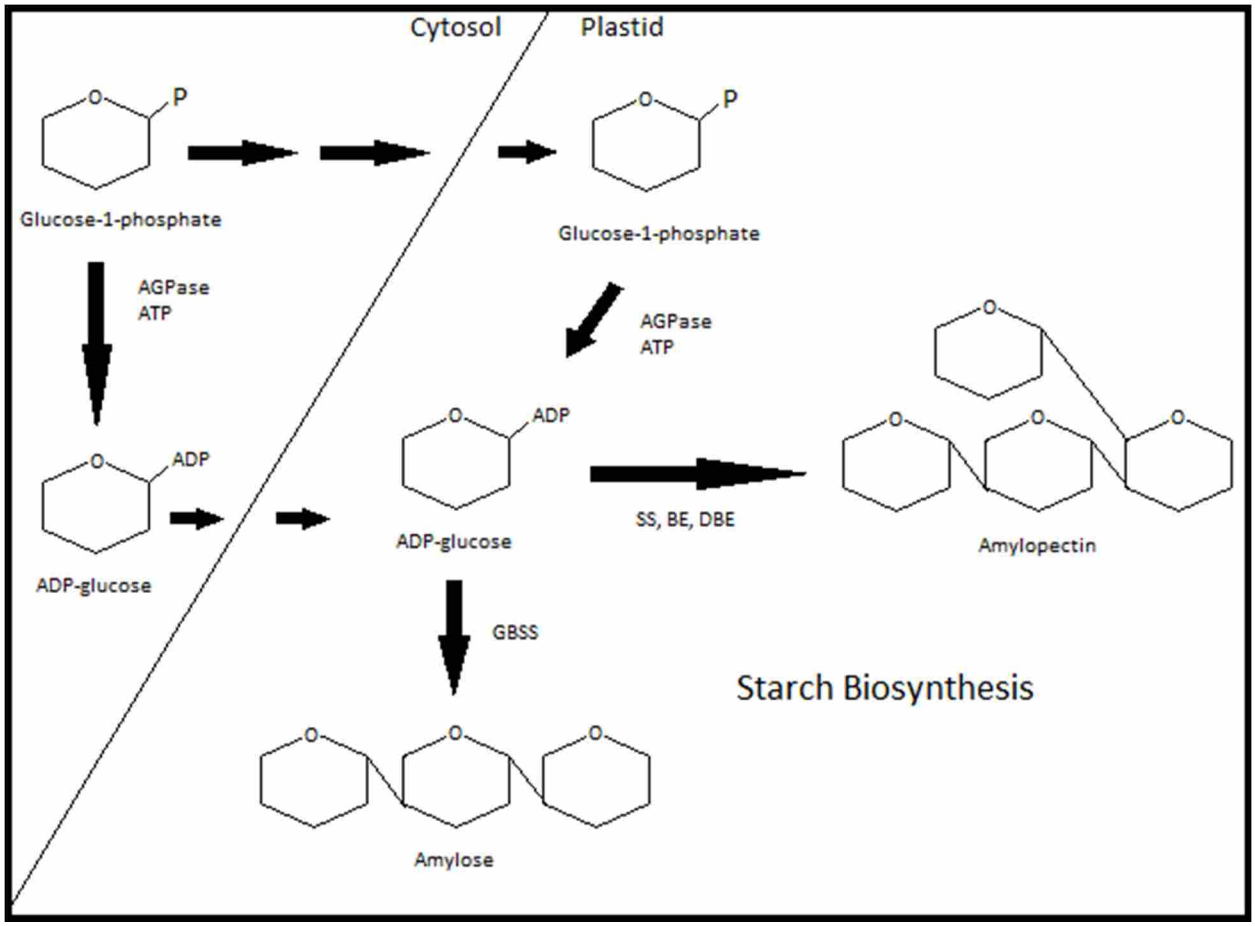

Figure 8. Starch biosynthesis