An Investigation of ßglux, a Glucosidase Co-Expressed with CsIf6 in Oat (Avena sativa) and Barley (Hordeum vulgare)

Michael Christopher Gines
Brigham Young University

Follow this and additional works at: https://scholarsarchive.byu.edu/etd

Part of the Plant Sciences Commons

BYU ScholarsArchive Citation
Gines, Michael Christopher, "An Investigation of ßglux, a Glucosidase Co-Expressed with CsIf6 in Oat (Avena sativa) and Barley (Hordeum vulgare)" (2016). Theses and Dissertations. 6595. https://scholarsarchive.byu.edu/etd/6595
An Investigation of βglux, a Glucosidase Co-Expressed with Cslf6 in Oat (Avena sativa) and Barley (Hordeum vulgare)

Michael Christopher Gines

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

Eric N. Jellen, Chair
Peter J. Maughan
Kathy Klos

Department of Plant and Wildlife Sciences
Brigham Young University

Copyright © 2016 Michael Christopher Gines
All Rights Reserved
ABSTRACT

An Investigation of βglux, a Glucosidase Co-Expressed with CslF6 in Oat (Avena sativa) and Barley (Hordeum vulgare)

Michael Christopher Gines
Department of Plant and Wildlife Sciences, BYU
Master of Science

Mixed Linkage Glucan (MLG, or (1,3;1,4)-β-D glucan) is a component of cell walls for major cereal crops and is significant to food and beverage industries. To better understand genetic factors affecting MLG content in oats, this study investigates the presence of glucosidases likely to participate in MLG production. A glucosidase showing co-expression with CslF6—the primary gene responsible for MLG synthesis—could indicate a hand in MLG production by association. Reference genes for expression analysis as well as glucosidase candidates were first selected using in silico methods. In both cases, barley was used as model species because it has abundant public bioinformatic resources for in silico data mining, and it generates large amounts of MLG, like oats. Actin, malate dehydrogenase, and elongation factor 2, were validated in oat and barley as top reference genes. They were then used to compare the expression activity of the top glucosidase candidate gene, βglux, with CslF6. βglux was found to have increased activity with CslF6 during caryopsis development. It is a strong candidate for future transgenic experiments regarding its effect on MLG production.

Keywords: reference gene, real-time PCR, CslF6, mixed linkage glucan, β (1,3;1,4) D glucan
ACKNOWLEDGMENTS

First, I would like to thank Eric Jellen, my advisor, for opening this door and for his continual guidance. I would also like to thank Peter Maughan and Kathy Klos for their contributions to not only my master’s degree, but to my growth as a scientist in other areas such as tutoring, lecturing, work experience through the USDA, and attending genetics conferences. I am grateful for the aid and partnership of fellow USDA co-workers, namely Thomas Baldwin, Robert Campbel, and Abdur Rashid. These three have helped share the workload for various tasks around the lab. I would like to express deep appreciation for Melissa Fogarty who has been a mentor, role model, and leader for students in the BYU Genetics and Biotechnology program. I also want to thank the BYU undergraduates Sarah Spendlove, Merylanne Dexter, and Michael Welch for their time and excellent lab work through this project. I want to thank BYU, General Mills, and the USDA-ARS whose valuable funds are necessary to advance such projects. And, finally, my wife and children who have been the most affected and most supportive in this phase of my education.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>TITLE PAGE</td>
<td>i</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>iii</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>viii</td>
</tr>
<tr>
<td>CHAPTER 1: (1,3;1,4) β-D-Glucan Literature Review Mixed Linkage Beta</td>
<td>1</td>
</tr>
<tr>
<td>GENETIC ORIGIN</td>
<td>2</td>
</tr>
<tr>
<td>LOCALIZATION OF PROTEIN AND PRODUCT</td>
<td>4</td>
</tr>
<tr>
<td>GLUCOSIDASE HYPOTHESIS</td>
<td>5</td>
</tr>
<tr>
<td>COMMERCIAL IMPORTANCE</td>
<td>6</td>
</tr>
<tr>
<td>CONCLUSION</td>
<td>7</td>
</tr>
<tr>
<td>LITERATURE CITED</td>
<td>8</td>
</tr>
<tr>
<td>CHAPTER 2: Identification and Validation of Superior Reference Genes</td>
<td>14</td>
</tr>
<tr>
<td>Genes for Barley Using In Silico Analysis</td>
<td>14</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>14</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>15</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 2-1: Validated primers of the 20 candidates used on barley cultivars. Genes which were also included in cited studies are underlined. Genes that demonstrated the most stability in this study are in bold. ................................................................. 28

Table 2-2: The RefFinder geomean stability ranking of candidates Hv.22848, Hv.9509, Hv.22901, Hv.2973, Hv.3271, Hv.20658, and Hv.23088. .................................................. 29

Table 3-1: The top ten candidates ranked from each statistical method approached. ................................................................................................................................. 52

Table 3-2: Primers used for the sequencing of cloned ßglux inserts. Position is described here as relative to the first nucleotide to the start codon. The * indicates sequences of the M13 primers included for the vector regions surrounding the cloned inserts. ........................................................................................................ 52

Table 3-3: Exon position and length as figured from the alignment of 3 HiFi contigs and 3 SolFi contigs to the CORE EST consensus contig. The positions are numbered by their relation to the first nucleotide of the start codon. ............................... 53

Table 3-4: Primers used for real-time PCR with their efficiency and r2 values. The * indicate primers used by Coon 2012. ......................................................................................... 53
LIST OF FIGURES

Figure 2-1: A box-and-whisker plot of the Cq data for each candidate across each Conlon tissue and biological replicate. The x-axis is labeled by the 20 UNIGENE clusters in the order they appeared in the in silico analysis rank. The bars represent minimum and maximum Cq values, the boxes indicate the first to third quartiles and the line in each box represents the median. Hv.3082 NADH dehydrogenase, Hv.674 Phosphoglycerate mutase, Hv.2613 Proteasome subunit beta type-7-B, Hv.22848 GAPDH, Hv.9509 Elongation factor EF-2, Hv.22901 Malate dehydrogenase, Hv.19033 Heat Shock Protein 70, Hv.735 Beta-tubulin 1, Hv.22789 Translationally-controlled tumor protein homolog (HTP), Hv.2973 3-ketoacyl-CoA thiolase like, Hv.3271 Superoxide dismutase [Cu-Zn], Hv.22823 Mitochondrial ATP synthase beta subunit, Hv.22783 Elongation factor 1-alpha, Hv.20658 S-adenosylmethionine synthetase 4, Hv.22798 Cytosolic heat shock protein 90, Hv.23243 Zinc finger A20 & AN1 domain-containing stress-associated, Hv.23088 Actin, Hv.3750 Peroxiredoxin-5, mitochondrial, Hv.3381 Translation initiation factor eIF-5A, Hv.12544 Proteasome subunit alpha type-5-A.......................................................................................................................... 30

Figure 2-2: A bar chart of the calculated stability values of the 20 candidate genes across 8 Conlon tissues. The y axis measures their geomean stability value calculated through RefFinder and the x-axis shows the candidates labeled by their UNIGENE cluster derived from the in silico rank. Hv.3082 NADH dehydrogenase, Hv.674 Phosphoglycerate mutase, Hv.2613 Proteasome subunit beta type-7-B, Hv.22848 GAPDH, Hv.9509 Elongation factor EF-2, Hv.22901

Figure 2-3: The box-and-whisker plot for the Cq data for candidates Hv.22848 (GAPDH), Hv.9509 (elongation factor 2), Hv.22901 (malate dehydrogenase), Hv.2973 (3-ketoacyl-CoA thiolase like protein), Hv.3271 (superoxide dismutase), Hv.20658 (S-adenosylmethionine synthetase 4), and Hv.23088 (actin) across Conlon, Golden Promise, Harrington, and HiFi. The bars represent minimum and maximum Cq values, the boxes indicate the first to third quartiles and the line in each box represents the median.

Figure 3-1: Glucosidase transcript levels in developing influorescence 1cm tissue. Arbitrary transcript units (Y-axis) for each accession (X-axis). Window cap at 100 units. MLOC_57200 is colored black and is the last accession.

Figure 3-2: Glucosidase transcript levels in developing influorescence 5mm tissue. Arbitrary transcript units (Y-axis) for each accession (X-axis). Window cap at 100 units. MLOC_57200 is colored black and is the last accession.
Figure 3-3: Glucosidase transcript levels in shoot tissue. Arbitrary transcript units (Y-axis) for each accession (X-axis). Window cap at 100 units. MLOC_57200 is colored black and is the last accession ................................................................. 55

Figure 3-4: Glucosidase transcript levels in root tissue. Arbitrary transcript units (Y-axis) for each accession (X-axis). Window cap at 100 units. MLOC_57200 is colored black and is the last accession ................................................................. 55

Figure 3-5: Glucosidase transcript levels in germinating embryo tissue. Arbitrary transcript units (Y-axis) for each accession (X-axis). Window cap at 100 units. MLOC_57200 is colored black and is the last accession ................................................................. 56

Figure 3-6: Glucosidase transcript levels in internode tissue. Arbitrary transcript units (Y-axis) for each accession (X-axis). Window cap at 100 units. MLOC_57200 is colored black and is the last accession ................................................................. 56

Figure 3-7: Glucosidase transcript levels in caryopsis tissue 5 days post anthesis (DPA). Arbitrary transcript units (Y-axis) for each accession (X-axis). Window cap at 100 units. MLOC_57200 is colored black and is the last accession ................................................................. 57

Figure 3-8: Glucosidase transcript levels in caryopsis tissue 15 days post anthesis (DPA). Arbitrary transcript units (Y-axis) for each accession (X-axis). Window cap at 100 units. MLOC_57200 is colored black and is the last accession ................................................................. 57

Figure 3-9: A bar chart representing the levels of CsLF6 and each of the top 5 glucosidase candidates from the in silico across all eight tissues. DPA=days post anthesis ........................................................................................................................................ 58
Figure 3-10: Expression of CslF6 and βglux. Expression levels are represented in a logarithmic scale relative to the target gene expression in the stem control. DPA = days post anthesis. DPG = days post germination. Points represent median value and bars represent the minimum and maximum values. Dotted lines were used to connect chronological data. Blue diamonds represent βglux. Green squares represent CslF6.
Cellulose, hemicellulose, and pectin form the three principle classes of polysaccharides in plant cell walls, the largest contributor being cellulose (Cosgrove, 2005). Cellulose is produced in coils of up to 36 chains of (1,4) β linked glucose molecules (Cutler and Somerville, 1997). These strands give the plant cell wall a rigid structure. The pectin class of polysaccharides fills the cell walls with a gelatinous material, increasing wall thickness and tensile strength (Domon et al. 2013, Fry 1989). Hemicelluloses include saccharides with equatorial 1-4 β linked backbones such as xyloglucans, xylans, mannans and glucomannans, and (1-3,1-4) β D glucan (Scheller and Ulskov 2010). The latter, also called mixed linkage glucan (MLG), is found in the grass family Poaceae (Scheller and Ulskov 2010) and in a few other species such as Equisetum arvense via convergent evolution (Simons 2013). Studies on MLG have found that it is prevalent in the plant endosperm but also in lower amounts in vegetative organs of grass plants (Vega-Sánchez et al. 2013). Its evolutionary advantage has been speculated to provide a flexible and strong tethering trait in cell walls during cell growth or as storage of carbon and energy for seed embryos (Burton and Fincher 2012, Meier and Reid 1982).

MLG, like cellulose, is an unbranched polysaccharide of glucose molecules where other hemicelluloses have branches or non-glucose components (Scheller and Ulskov 2010). The MLG polysaccharide chain is composed of repeating D-glucose molecules in identical orientation (beta linkage) with two alternating linkages. The glucose chain is formed from covalent bonds between either carbons 1 and 4 or 1 and 3 of adjoining monomers. The linkages are mostly repeating (1,4) linkages with interrupting (1,3) linkages. The number of (1,4) linkages between (1-3) linkages are not consistent in this polysaccharide. The ratios of (1,4) to (1,3) linkages differ...
between the plants. On average, ratios range between 2.2-2.6 (1,4) linkages for each (1,3) linkage (Burton and Fincher 2009).

It is not yet known how the pattern of (1,3) and (1,4) links are determined. If these linkages were random, then we would expect to see the occurrence of tandem (1,3) linkages as often as the frequency of one (1,3) linkage squared. So, we would expect 0.077-0.097% of all linkages between three sequential monomers to occur as two (1,3) links. But sequential (1,3) links seldom occur (Burton and Fincher 2009). The inclusion of the (1,3) linkages then is, though inconsistent, not completely random. The frequency of four or more sequential (1,4) linkages between (1,3) linkages is also below what random incorporation would show. Segments of three (cellotriosyl) or four (cellotetrasyl) beta glucan monomers in (1,4) linkage between (1,3) linkages make up 90% of the MLG polysaccharide (Burton and Fincher 2009).

The inconsistent ratio of MLG (1,3) linkages to (1,4) linkages across plant species, and even tissues, has been investigated. Buckridge et al (2004) proposed a process dependent upon the concentration of monomers present during synthesis as well as a glucosyltransferase. Jobing (2015) performed chimeric protein experiments of Cslf6 from several grass species in a transformed tobacco plant. His results showed a single Ile/Leu difference in transmembrane helix 4 correlated with high/low (1,3):(1,4) ratios, respectively.

**GENETIC ORIGIN**

Investigation of genetic factors responsible for MLG was conducted by QTL mapping in barley (Han et al 1995); this study identified a region on chromosome 2H that was highly correlated with grain MLG content. The genetic markers from this barley 2H QTL were later associated with a similar region on rice chromosome 7 containing six known Csl (cellulose synthase-like) genes (Burton et al 2006). QTL analysis determined the CslF6 gene correlates
with the highest transcript levels and MLG production in the grasses (Nemeth et al. 2010, Wilkinson et al. 2010).

Cellulose synthase-like proteins were designated as part of a superfamily also containing CesA proteins (Hazen, Scott-Craig, and Walton 2002), the latter being a glucosyltransferase family 2 protein responsible for cellulose production (Delmer 1999, Pear et al 1996, Campbell, Davies, Bulone, and Bernard 1997). As the structure of single cellulose and MLG chains are both unbranched lengths of the same monomers in the same orientation, only differing in occasional (1,3) linkages, it is not surprising to find that genes primarily responsible for MLG production are closely related to CesA genes. The superfamily was described as sharing an amino acid motif associated with glycosyltransferase capacity, DDDQXXRW (Saxena and Brown 1995). Currently, nine Csl families have been discovered in this superfamily and were labeled as CslA-H and CSLJ, along with the CesA family (Yin, Huang, and Xu 2009). Each Csl protein contains three to six transmembrane domains and is expected to localize to the Golgi apparatus (Richmond and Somerville 2000).

Though all ten of these families are expected to associate with the production of some polysaccharide, only a portion have been characterized as having a known function. Csl genes expressed in insects revealed the CslA to be responsible for synthesizing mannan (Liepman, Wilkerson, and Keegstra 2005). High expression of CslC was associated with xyloglucan deposition in nasturtium seeds (Cocuron et al 2007). One analysis suggests CslA and CslC originated through a gene duplication event and so CslA likely codes for a mannan synthase protein also (Yin, Huang, and Xu 2009). The CslF and CslH families have been shown to directly correlate with MLG production (Burton et al 2006, Doblin et al 2009).
Arabidopsis thaliana does not normally produce MLG but a transformation experiment with the organism using a single CslH or CslF gene (Burton et al. 2006, Doblin et al. 2009) was enough to produce the polysaccharide in the plant. This shows that MLG production can occur from a single gene, providing an explanation on how a several plants could acquire a Csl gene through separate evolutionary events. The low amounts of MLG produced in the transformed Arabidopsis, however, hints at other contributing factors in MLG synthesis for grass plants. When CslH and CslF genes were expressed jointly in Arabidopsis, the expression level was near normal (Doblin et al 2010), suggesting a dimer complex. Wheat MLG levels are also optimal when transcripts of both CSLF and CSLH are present (Pellny et al 2012).

Taketa et al (2012) further clarified the importance of CslF6 by testing null mutations in barley for MLG presence. A mutation in a conserved region of the HvCslF6 gene caused a betaglucanless (bgl) phenotype. In the same study, the native HvCslF6 was transiently expressed in N. benthamiana leaves, which then began producing MLG.

LOCALIZATION OF PROTEIN AND PRODUCT

Amino acid sequencing and analysis of CslF6 reveals integral membrane motifs (Hazen, Scott-Craig, and Walton 2002, Saxena and Brown 1995). For some time, MLG had been found only in the plant cell wall while CslF6 could be found only in the ER and Golgi vesicles (Philippe et al. 2006). This suggested that CslF6 was a necessary instigator for MLG synthesis but required additional steps for the polysaccharide to emerge as a complete product at the cell wall. One possibility was that CslF6 produced a type of saccharide which was later modified in the cell wall by an endotransglycosylase (Simmons et al. 2013). One study in maize showed the presence of MLG in Golgi bodies with CslF6 (Carpita and Maureen 2010). The identification was performed through immunolabeling of MLG and may not have been successful in previous
attempts because of some inaccessibility of the epitopes. A more recent study, however, placed CslF6 at the plasma membrane, diverging from other Csl gene locations (Wilson et al. 2015). Modifications such as acetyl esters were proposed as cellular mechanisms to increase solubility during hemicellulose export and may have impeded antibody labeling until the modifications were removed beyond the plasma membrane (Gille and Pauly 2012). Wilson et al. (2015) performed treatments which would have removed these ester groups, though, and found no change in labeling. MLG was only seen in the wall while other hemicelluloses were found in walls and adjacent Golgi bodies. The study found further immunolabeling of CslF6 proteins at the plasma membrane, while both CslF6 and CslH1 were found in the Golgi, Golgi vesicles, and the ER. The results concluded that CslF6 generated MLG at the cell wall.

GLUCOSIDASE HYPOTHESIS

The locations of MLG and CslF6 have been associated with the cell wall, but factors involving MLG length, quantity, and linkage ratio are still under investigation. A recent study (Jobling 2015) determined a single amino acid change in the CslF6 protein is responsible for significant changes in DP3:DP4 (cellotriosyl:cellotetrasyl) ratio across plant species. Transgenic tobacco plants expressing chimeric variants of CslF6 revealed a single isoleucine or leucine at the 757th amino acid resulted in high or low DP3/DP4 ratio respectively. This variation in the peptide was shown to occur in a transmembrane helix (TMH4). This implies that the transmembrane pore is mainly responsible for the difference in cellotriosyl and cellotetrasyl members.

The amount and length of MLG chains are dependent on other factors, perhaps including monomer availability. Gibeaut and Carpit (1993) showed MLG synthesis is dependent upon UDP-glucose monomers. To arrive at UDP-glucose, a lone glucose is phosphorylated by a
glucokinase into glucose-6-phosphate, which is then acted upon by phosphoglucomutase to become a glucose-1-phosphate. Glucose-1-phosphate is then reacted with UTP by UDP-glucose phosphorylase to produce UDP-glucose and the pyrophosphatase byproduct. A cytosolic UDP-glucose, therefore, depends on the presence of cytosolic glucose. The amount and length of MLG produced may depend on the quantity of glucose molecules locally available to meet the demand, along with the glucosidases that provide the glucose.

Genes with expression profiles similar to that of the CslF6 may likewise be contributors to MLG synthesis, leading to a better understanding of its creation. Hemicelluloses, such as xyloglucan, require glucosidases for their synthesis or structure arrangement (Farroki et al. 2006); consequently, those glucosidases having expression profiles similar to CslF6 could be potential candidates in helping to define MLG biosynthesis (Lee et al. 2004; Oliver 2000; Wolfe et al. 2005).

COMMERCIAL IMPORTANCE

According to the United States Department of Agriculture March report of “Grain: World Trades Market” the production tally in the last year reached 22,589 thousand metric tons (tmt) for oats and 140,720 tmt for barley (USDA 2015). This is compared to the world production of other grains: 989,661 tmt of maize; 724,759 tmt of wheat; 474,856 tmt of rice; 62,025 tmt of sorghum; and 14,487 tmt of rye. This puts barley as the fourth largest grain crop and oats as the sixth. However, MLG content for the grains of these crops, as determined by dry mass percentage, is 2.87-5.4% for oats; 2.36-5.4% for barley; 1-2% for rye; less than 0.6% for sorghum and wheat; and 0.12% maximum for maize (Jorgensen 1988; McClear 1985). These data show oats and barley are the greatest cereal producers of MLG.
Barley and oats are used for both human and livestock consumption, though mainly for the latter. MLG polysaccharides from the two grains act as dietary fibers. Such dietary fibers help clean out the digestive track and lower blood cholesterol and blood glucose levels (Brennan et al. 2005). Bread baked with MLG, oat cereals, and other nutrition experiments demonstrate the correlation between MLG presence and reduction of blood glucose levels and the occurrence of postprandial hypoglycemia (Caval et al. 2002, Poppit et al. 2007). Barley is also largely used in the production of alcoholic malt beverages. Malting quality, however, is reduced by the presence of proteins and MLG (Zhang et al. 2001). Therefore, as the presence of MLG can have negative or positive effects for targeted uses, it is beneficial to identify the associated genes which contribute to MLG levels in oat and barley grains.

Unlike cellulose, the MLG chain can be digested by humans because of its 1-3 glycosidic linkages (Brennan et al. 2005). Its cholesterol-reducing effect is hypothesized as being due to the stimulatory effect of high viscosity and high molecular weight MLG on bile secretion into the bowel—bile being a sink for LDL cholesterol removed from the blood (Lia et al. 1995).

CONCLUSION

The MLG content of a variety of grass grains have been measured across several cultivars. Such content holds positive or negative value depending on the desired end product. Though key genes are understood to be largely responsible for MLG presence, the degree and quality of that presence is not well explained. Understanding genetic components participating in the biosynthetic process of MLG could introduce new targets in crop breeding to achieve grain products with desirable qualities. To this end, this thesis endeavors to find appropriate candidates and methods of comparing genetic activity to CsIF6.
LITERATURE CITED


CHAPTER 2 : IDENTIFICATION AND VALIDATION OF SUPERIOR REFERENCE GENES FOR BARLEY USING IN SILICO ANALYSIS.

ABSTRACT

• **Premise of the study:** Reference genes are selected based on the assumption of temporal and spatial expression stability and on their widespread use in model species. They are often used in new target species without validation, presumed as stable. For barley, reference gene validation is lacking, but publicly-available bioinformatic resources are available to predict the expression stability of experimental candidates.

• **Methods:** EST profile viewer data from the UniGene library were used to estimate the expression stability of 655 barley genes. Twenty gene candidates predicted as most-stable were evaluated in the barley cultivar ‘Conlon’ by qRT-PCR across eight tissues. The five most-stable genes were then tested in the barley cultivars ‘Golden Promise’ and ‘Harrington,’ and (to test potential applicability to other *Poaceae* species) in the oat cultivar ‘HiFi.’

• **Results:** The traditional candidate actin (Hv.23088) and novel candidates elongation factor 2 (Hv.9509) and malate dehydrogenase (Hv.22901) demonstrated the most stable expression across barley and oat.

• **Discussion:** The predictive capacity of bioinformatics to identify suitable reference genes was demonstrated. Several novel reference genes were found to have similar stability to the top candidate, actin. These reference genes are recommended for barley under normal conditions and should be validated under experimental conditions if used.
INTRODUCTION

The advent of large-scale gene expression analysis by microarrays and, increasingly, by RNA-seq is changing the landscape of molecular biology research. These methods produce large datasets that are informative and reliable, but time-consuming to mine and costly to produce. For these reasons, classic techniques for gene expression analysis like northern blots and quantitative PCR are still effective and widely used for targeted, small-scale analysis. These methods can provide insights on gene regulation, protein interactions, novel gene discovery, and biosynthesis steps (Lee et al. 2004; Oliver 2000; Wolfe et al. 2005). However, they require normalization of target gene expression based on the constitutive expression of reference genes in the tissues being investigated.

Reference genes used for expression studies in barley (Hordeum vulgare L.) and other cereals have generally been chosen based on their use in studies of other species, often without further validation. Previous studies designed to identify or validate appropriate reference genes for barley have been conducted, but each was limited with respect to the cultivars, tissues, or genes studied. Many relied on traditional candidates. For instance, Hua et al. (2014) and Ferdous et al. (2015) identified the most-stable reference genes from a pool of traditionally-used genes including GADPH and actin, and Ovensa et al (2001) used caryopsis tissues to test a list of genes obtained from a separate study. More recently, the use of microarray-based expression to develop reference genes has been used as a more powerful approach to determine the most stable reference gene for a set of tissues. Janská et al. (2013) surveyed 13 genes in leaf and crown seedling tissue subjected to abiotic stress and Zmienko et al. (2015) surveyed 181 genes in senescing leaf tissues to identify the most stable genes within their respective tissue set. However, these genes might not be applicable to other tissues or conditions. Overall, these
approaches have one or more of these limitations: i) genes may be differentially expressed among species and in tissues within a species; ii) restriction of reference genes to a pool of previously-used reference genes may overlook superior reference candidates, and iii) microarray analysis is expensive and time consuming.

A more powerful approach may be to use current, publically-available bioinformatic resources to enable analysis of a far greater number of potential candidates. Such in silico mining techniques involve stability prediction of hundreds or thousands of genes followed by experimental validations of top candidates. In silico analysis has the distinct advantage of comparing multiple tissues, times points, and conditions across many experiments. Ovensa et al. (2012) used such a technique to identify reference genes of specific utility for normalizing Bmy1 expression at five developmental stages of barley caryopses. Lateral comparison of these publically-available data sets narrows the selection choice for finding adequate reference genes.

The objective of this study was to use in silico analysis to identify a set of highly reliable and stably expressed reference genes, and then validate them across genetically diverse barley cultivars and diverse plant tissues. The tissues selected included those important for the development of β-glucan fiber associated with cardiovascular health (United States Food and Drug Administration 2005) and malting quality (Jin et al. 2004). The three barley cultivars studied were selected based on their importance to genetic transformation studies and/or, commercial utility, and their genetic diversity. To test the potential for extending results to other members of the Poaceae, top reference gene candidates were also evaluated in an oat (Avena sativa) cultivar.
MATERIALS AND METHODS

Plant materials—Plant materials were obtained from Conlon, Golden Promise, and Harrington barley cultivars and from the HiFi cultivar of oat. The barley cultivars were chosen based on their importance for malting (Harrington and Golden Promise) and their use in transformation studies (Conlon and Golden Promise), and they are relatively genetically diverse. Eight tissues were chosen for each based on their relevance to a variety of important barley developmental stages: 12-day-old seedling roots and shoots; stems and tiller leaves at the six tiller stage; caryopsis at 0-1, 5-7, and 12-15 days post anthesis (0, 5, and 15 DPA); and seed tissue four days post-germination (4 DPG). For each cultivar and tissue type, thee biological replicates were assayed (each replicate sampled from a separate plant). All plants were grown in a greenhouse. Samples were harvested and immediately used for extraction or frozen in liquid nitrogen and stored at -80° C.

In silico analysis—Candidate reference genes were found using an in silico estimation of expression stability using a public database. A UniGene (http://www.ncbi.nlm.nih.gov/unigene) search was performed for barley EST clusters. Robust, well-represented candidates were selected by using the search term “Hordeum vulgare 100:20000[ESTC]” to exclude clusters with less than 100 ESTs, resulting in a pool of 689 candidate genes. Of these, 665 accessions had EST profiles for tissues assayed in this study (see below), and were used for estimating gene expression stability. For each cluster’s EST profile, coefficient of variation (CV) was derived from the Transcripts per Million (TPM) values across leaf, pericarp, pistil, root, seed, spike, and stem tissues. Anther, callus and meristem were excluded as they were underrepresented in their EST library submissions. The candidates were reduced to 430 after eliminating those with 0 listed in the TPM field for any of the seven tissues. The 20 clusters with the lowest CV were
selected as candidates for the empirical expression study, and ranked from 1 to 20 relative to their CV values.

**Primer design and validation**—Primers were validated to efficiently amplify barley and oat cDNA in real-time PCR. For application across both species, the first round of primer design used homologous regions for barley and oat sequences. A primer set was first validated for each candidate using barley cDNA. For genes used in oat experiments, those primer sets which did not perform optimally using oat cDNA were redesigned specifically for oat sequences. Barley sequences were acquired from NCBI (http://www.ncbi.nlm.nih.gov) accessions linked to UNIGENE clusters. Oat sequences were predicted by aligning CORE (Oliver et al. 2013) EST libraries of 22 different lines to the barley candidate gene NCBI accessions. Oat ESTs were aligned to barley sequences using Sequencher version 5.2.4 (Gene Codes Corporation, Ann Arbor, MI, USA) with 85% minimum match, minimum 20 bp overlap, and under the dirty data setting. All primer pairs were validated through end-point PCR and gel electrophoresis for specific amplification of single products. Satisfactory primer pairs were then validated for real-time PCR applications by performing a standard curve analysis. Primer pairs were accepted that produced products from 100-400 bp with amplification efficiencies between 90-110%, with $r^2$ above 0.98 (Table 1), and with single melt peaks (Appendix 1, 2). Some of the primer sets, including that for candidate 13, were observed to amplify genomic DNA through PCR as well. To ensure no genomic carry-over from RNA extraction interfered in real-time quantification, all RNA samples were run parallel with their cDNA counterparts using primer set 13 (Hv.22783; see Table 1).

**RNA extraction**—RNA extraction was performed using the GeneJET Plant Purification Minikit (Thermo Fisher Scientific, Waltham, MA, USA). Samples were then treated with DNase I and
DNase 5x buffer (New England Biolabs, Ipswich, MA, USA). The RNA was then isolated through LiCl precipitation. Samples with high starch levels (all 4 DPG seeds and leaf and shoot tissues for oats) clogged the GeneJet spin columns, and were therefore isolated through a Trizol (Thermo Fisher Scientific, Waltham, MA, USA) extraction method described by Li and Trick (2005) with further purification through LiCl precipitation. RNA samples were observed for quality using gel electrophoresis and OD readings. Samples showing defined 28S and 18S bands, and with values for OD 260/280 and 260/230 >1.7 were selected for use. The RNA samples were checked for genomic DNA contamination by running them in parallel with cDNA through real-time PCR. Sample cDNA was generated using the iScript kit (Bio-Rad, Hercules, CA, USA).

**Quantitative reverse transcriptase PCR**— Quantification of transcript levels for each sample was performed in triplicate through real-time PCR. All candidate reference gene expression measurements were performed using the SsoFast EvaGreen Supermix (Bio-Rad, Hercules, CA, USA) with the Sybrgreen fluorophore. The standard protocol recommended by the manufacturer was followed using 0.5 µL of each primer pair for each primer set, and 2.5 ng of template CDNA for each tissue. The thermal cycler program was 95° C activation for 30 seconds, 40 cycles of a melting step at 95° C for 5 seconds and an annealing/extension step of 60° C, followed by a melt curve step from 65-95° C in increments of 0.5° C every 5 seconds. A CFX thermal cycler and CFX Manager v. 3.21 (Bio-Rad, Hercules, CA, USA) were used to generate and analyze the data. Transcript abundance was measured from the cycle in which amplification of the template-primer product exhibited fluorescence above the threshold detection level for each target, manifesting as a Cq value.

**Statistical comparison of reference genes**— To assess the performance of each candidate gene empirically, the data retrieved from the real time experiments were uploaded to an online data
analysis tool, RefFinder (Fuliang et al. 2012, http://fulxie.0fees.us/?i=1). RefFinder returned the results from several statistical models used to calculate gene expression stability. Stability rank was assigned for each individual model, and then the placement for each reference gene under each model was used to derive a geomean value. The geomean values were used to assign a comprehensive rank for the reference genes. The statistical models used include BestKeeper, comparative delta Ct, geNorm, and Normfinder (Pfaffl et al. 2004, Silver et al. 2006, Vandesompele et al. 2002, and Andersen et al. 2004). The final geomean values were used to assign stability to candidate genes.

The empirical analysis of all 20 candidates was first performed on Conlon. Based on the results of this analysis, the top five novel candidates and two traditional reference genes (actin and GAPDH) across diverse cultivars were measured across Golden Promise, Harrington, and HiFi. The Cq values for these seven genes in eight tissues were applied to the RefFinder program for Conlon, Golden Promise, Harrington, and HiFi individually, then across all three barley cultivars, and finally across HiFi oat and all three barley cultivars together.

RESULTS

Twenty candidate reference genes from in silico analysis, validation in Conlon—The in silico method identified the 20 top candidates based on the lowest CV values. Seven were used in one or more of the six cited reference gene articles for barley (Table 1). A box-and-whisker plot representing the range of each candidate's expression from all tissues and biological replicates is shown in Fig. 1. Based on the RefFinder analysis, the five most stable candidates were Hv.3271, Hv.20658, Hv.9509, Hv.2973, and Hv.22901 (Fig. 2, Appendix 3) in descending order of stability.
A comparison was made between the predicted and measured stability of the 20 candidate genes. There was no significant correlation between the \textit{in silico} coefficient of variation rank and the rank based on expression analysis of the top 20 candidates in Conlon ($r^2 = -0.0511$, $p=0.8306$). Top candidates Hv.3271, Hv.20658, Hv.9509, Hv.2973, and Hv.22901 were ranked 11\textsuperscript{th}, 6\textsuperscript{th}, 5\textsuperscript{th}, 10\textsuperscript{th}, and 14\textsuperscript{th} \textit{in silico} respectively. The candidate that returned the lowest coefficient of variation, Hv.3082 (predicted as the most stable), placed 15\textsuperscript{th} in the comprehensive ranking (13\textsuperscript{th} in the Delta CT method, 16\textsuperscript{th} in the BestKeeper method, 13\textsuperscript{th} in the Normfinder method, and 14\textsuperscript{th} in the geNorm method). The candidate that returned the highest coefficient of variation (predicted as the least stable), Hv.12544, resulted as 12\textsuperscript{th} in the comprehensive ranking (7\textsuperscript{th} under the Delta CT method, 15\textsuperscript{th} for BestKeeper, 7\textsuperscript{th} for Normfinder, and 11\textsuperscript{th} for geNorm) (Appendix 3).

\textit{Top candidate reference genes’ validation in barley and oat} — The seven-gene box-and-whisker plots for the individual RefFinder analyses of Conlon, Golden Promise, Harrington, and HiFi are shown in Fig. 3. The RefFinder comprehensive ranks are shown in Table 2, along with the broad inclusion of combined barley and combined barley and oat data. Actin (Hv.23088), which ranked 7\textsuperscript{th} of 20 in the original empirical Conlon analysis, ranked 6\textsuperscript{th} of 7 in this analysis for Conlon, and was still behind the top candidates but ahead of GAPDH (Hv.22848). However, actin was ranked 3\textsuperscript{rd} for Golden Promise and Harrington, 1\textsuperscript{st} for HiFi, and 1\textsuperscript{st} for the combined barley and the combined barley plus oat analyses. GAPDH ranked 7\textsuperscript{th} in Conlon, 7\textsuperscript{th} in Golden Promise, 1\textsuperscript{st} in Harrington, 6\textsuperscript{th} in HiFi, 7\textsuperscript{th} for all barley cultivars, and 6\textsuperscript{th} across barley and oat. Candidate Hv.3271 which ranked 1\textsuperscript{st} in both analyses of Conlon, was less stable in the other tested cultivars.
DISCUSSION

There was a low correlation between the *in silico* 20-candidate preliminary rank and the measured expression stability in the 20-candidate Conlon validation. However, this approach identified several genes, notably actin, previously regarded as stable reference genes, and additional candidates with similar or better stability than other commonly-used reference genes – notably, GAPDH. This supports the idea that while initial screening for candidates is beneficial to identify a pool of potential reference genes, experimental validation is still required to ensure selection of the most stable reference genes. This lack of correlation between *in silico* and empirical rankings among the top 20 candidates may be due to the different cultivars used in this study from the pools of experiments used in the UNIGENE database.

The top candidates appear to have little difference in stability. In Conlon, actin (Hv.23088) and the top novel candidates have similarly low stability values and their rank placement is susceptible to rearrangement under the small change from the 20-candidate to the 7-candidate RefFinder analysis. Through the included cultivars, the top candidates swapped rank placement while GAPDH (Hv.22848) continued to rank as less stable with the exception of Harrington, which was generally more variable for gene stability (Fig. 3).

The more stable candidates from this study serve for future reference genes under normal plant growth conditions but need to be validated for any experiment involving treatments. The novel candidates have not been validated in previous barley studies to compare their performance, but actin and GAPDH have a history for reference. The two traditional candidates have been tested in previous experiments and the superior stability of one over the other depended on the treatments. Janská et al. (2013) recommended actin for cold conditions and GAPDH for drought. Ferdous et al. (2015) found actin to be more stable than GAPDH overall...
through five stresses, though it was shown to be less stable in nitrate and salt treatments. These novel candidates, then, should be reviewed for their stability if used under new conditions.

The use of public expression data was sufficient for this study’s purposes. The in silico analysis of publicly-available expression data in this study included seven traditional reference genes in the top 20 candidates while identifying novel candidates with high predicted stability. This coincides with the expectations that, though traditional reference genes have been validated as stable in other studies, there can be untested genes which perform better. Ideally, the most accurate information is probably best achieved with microarray or RNA-seq comparisons of individual tissue samples. If economic restrictions prevent the use of these methods, surveying expression data across large bioinformatics resources is a suitable approach, as seen here. Furthermore, in silico analysis is the most practical approach for those species like oat, which do not yet have large bioinformatics resources. Model species can be utilized to find candidate reference genes, as represented in this study where HiFi oat exhibits ranking of the candidate genes similar to that of the barley cultivars.

When Cq values for all barley cultivars combined, for oat (HiFi), and for barley plus oat combined were considered, candidate Hv.23088 (actin) emerged as the most stable candidate, followed by candidates Hv.9509 (elongation factor 2) and Hv.22901 (malate dehydrogenase). Similarly to previous studies, we found there is no ideal reference gene as the accuracy for such is refined by the tissue and genotype involved. For instance, actin was not shown to be the most stable for any single barley cultivar, but was the most stable overall. This supports the need to evaluate reference genes over multiple genotypes when considering a species in general. Additionally, the in silico analysis identified two novel and broadly applicable reference genes,
elongation factor 2, and malate dehydrogenase, validated as barley reference genes under normal conditions for use in future studies.

ACKNOWLEDGMENTS

Funding for this research was provided by the Agricultural Research Service, USDA, project 5366-21000-031-00D, and by General Mills, Inc. The USDA is an equal opportunity employer.
LITERATURE CITED


### Table 2-1: Validated primers of the 20 candidates used on barley cultivars. Genes which were also included in cited studies are underlined. Genes that demonstrated the most stability in this study are in bold.

<table>
<thead>
<tr>
<th>In silico rank</th>
<th>CV</th>
<th>UNIGENE Cluster</th>
<th>NCBI accession</th>
<th>Gene</th>
<th>Forward primer</th>
<th>Reverse primer</th>
<th>Size (bp)</th>
<th>Efficiency (%)</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.154</td>
<td>Hv.3082</td>
<td>AK362890</td>
<td>NADH dehydrogenase</td>
<td>AGATCGCAATTCAGCAGACC</td>
<td>CCAGTCTGCCCCCTTAATCA</td>
<td>112</td>
<td>107.8</td>
<td>.987</td>
</tr>
<tr>
<td>2</td>
<td>0.205</td>
<td>Hv.674</td>
<td>AK251074</td>
<td>Phosphoglycerate mutase</td>
<td>GCACAGTGTGGTGGTCTGTA</td>
<td>CTTGGGCTCTGTCTCTTAG</td>
<td>135</td>
<td>109.5</td>
<td>.981</td>
</tr>
<tr>
<td>3</td>
<td>0.205</td>
<td>Hv.2613</td>
<td>AK251152</td>
<td>Proteasome subunit beta type-7-B</td>
<td>AGTTGCTTGGAAAGGTGCC</td>
<td>CCAAGGGTCATGTCATGCTG</td>
<td>145</td>
<td>101.3</td>
<td>.993</td>
</tr>
<tr>
<td>4</td>
<td>0.221</td>
<td>Hv.22848</td>
<td>X60343</td>
<td>GAPDH</td>
<td>GAGGTCTGATGACACTGTG</td>
<td>AGATCAACTGACATCCAC</td>
<td>221</td>
<td>93.7</td>
<td>.982</td>
</tr>
<tr>
<td>5</td>
<td>0.231</td>
<td>Hv.9509</td>
<td>AK250157</td>
<td>Elongation factor EF-2</td>
<td>AACTGGCATGAAGGTGCC</td>
<td>CCAAGGGTCATGTCATGCTG</td>
<td>218</td>
<td>107.9</td>
<td>.988</td>
</tr>
<tr>
<td>6</td>
<td>0.240</td>
<td>Hv.22901</td>
<td>AK358926</td>
<td>Malate dehydrogenase</td>
<td>GCACCATGAGCTGAATGCAGT</td>
<td>CTTTCAGGGATAGATGAGC</td>
<td>218</td>
<td>102.4</td>
<td>.98</td>
</tr>
<tr>
<td>7</td>
<td>0.242</td>
<td>Hv.19033</td>
<td>AK248694</td>
<td>Heat Shock Protein 70</td>
<td>CTTTCCTCCAGCTGCCAAA</td>
<td>AGTCGCTGAACTGGAAT</td>
<td>228</td>
<td>110.5</td>
<td>.986</td>
</tr>
<tr>
<td>8</td>
<td>0.244</td>
<td>Hv.735</td>
<td>Y09741</td>
<td>Beta-tubulin 1</td>
<td>GCACCATCTCTCTCACCAG</td>
<td>GTTAGCAACATCCAGCTGTC</td>
<td>239</td>
<td>107.0</td>
<td>.981</td>
</tr>
<tr>
<td>9</td>
<td>0.254</td>
<td>Hv.22789</td>
<td>AF230786</td>
<td>Translationaly-controlled tumor protein homolog (HTP)</td>
<td>GTGCTCTGGGAAATGCAATG</td>
<td>CAAACTGACTACCTTGAC</td>
<td>135</td>
<td>109.7</td>
<td>.997</td>
</tr>
<tr>
<td>10</td>
<td>0.255</td>
<td>Hv.2973</td>
<td>AK251228</td>
<td>3-ketoacyl-CoA thiolase like</td>
<td>AAGTGAGTGATGTCCTGGA</td>
<td>GATCCACTCCACAGCAACG</td>
<td>111</td>
<td>109.6</td>
<td>.997</td>
</tr>
<tr>
<td>11</td>
<td>0.259</td>
<td>Hv.3271</td>
<td>AK252295</td>
<td>Superoxide dismutase [Cu-Zn]</td>
<td>GCACCATCTCTCCACCAG</td>
<td>GTTAGCAACATCCAGCTGTC</td>
<td>239</td>
<td>107.0</td>
<td>.981</td>
</tr>
<tr>
<td>12</td>
<td>0.261</td>
<td>Hv.22823</td>
<td>AK359072</td>
<td>Mitochondrial ATP synthase beta subunit</td>
<td>GCCACAGGACGCAAAATCT</td>
<td>CACCAAGCAACAGAGAAA</td>
<td>178</td>
<td>92.1</td>
<td>.984</td>
</tr>
<tr>
<td>13</td>
<td>0.264</td>
<td>Hv.22783</td>
<td>Z50789</td>
<td>Elongation factor 1-alpha</td>
<td>TTTGGTGCCATTGGAACGTG</td>
<td>CAAACCACGGCTGGATCC</td>
<td>207</td>
<td>108.9</td>
<td>.998</td>
</tr>
<tr>
<td>14</td>
<td>0.276</td>
<td>Hv.20658</td>
<td>AM039896</td>
<td>S-adenosylmethionine synthetase 4</td>
<td>CCAATTGAGCACGATCC</td>
<td>GCACCAATCTCATCGTTC</td>
<td>342</td>
<td>109.9</td>
<td>.986</td>
</tr>
<tr>
<td>15</td>
<td>0.287</td>
<td>Hv.22798</td>
<td>AY325266</td>
<td>Cytosolic heat shock protein 90</td>
<td>AGATCAACCGCTTCTTCG</td>
<td>CAAAGGTAGGCCAGTGA</td>
<td>359</td>
<td>103.6</td>
<td>.983</td>
</tr>
<tr>
<td>16</td>
<td>0.289</td>
<td>Hv.23243</td>
<td>AK250311</td>
<td>Zinc finger A20 &amp; AN1 domain-containing stress-associated</td>
<td>TGAGCAGTGGCTCAGAATGC</td>
<td>TGGACAGAATGCTGCAGCAT</td>
<td>371</td>
<td>100.0</td>
<td>.994</td>
</tr>
<tr>
<td>17</td>
<td>0.292</td>
<td>Hv.23088</td>
<td>AY145451</td>
<td>Actin</td>
<td>GAATGGTGACGACGTGGTCTG</td>
<td>TCCATGTCATCCCAGTGGT</td>
<td>205</td>
<td>110.0</td>
<td>.981</td>
</tr>
<tr>
<td>18</td>
<td>0.296</td>
<td>Hv.3750</td>
<td>AK371415</td>
<td>Peroxiredoxin-5, mitochondrial</td>
<td>GGTGCATCTCCTCGCGGTCT</td>
<td>CTCACGTGCTGGGAGATCTC</td>
<td>380</td>
<td>98.7</td>
<td>.983</td>
</tr>
<tr>
<td>19</td>
<td>0.299</td>
<td>Hv.3381</td>
<td>AK252614</td>
<td>Translation initiation factor eIF-5A</td>
<td>CACCACTTGAGTCACAGGC</td>
<td>CCAAGTGCTTCCAGCTTGG</td>
<td>146</td>
<td>108.0%</td>
<td>.995</td>
</tr>
<tr>
<td>20</td>
<td>0.304</td>
<td>Hv.12544</td>
<td>AK367709</td>
<td>Proteasome subunit alpha type-5-A</td>
<td>TCACGAGGACTGTAGTAC</td>
<td>CTCAACAGGCCAGAAACAC</td>
<td>146</td>
<td>104.6%</td>
<td>.99</td>
</tr>
</tbody>
</table>
Table 2-2: The RefFinder geomean stability ranking of candidates GAPDH, EF-2, Mal de, 3KACoA, SOD, S-AMS4, and Actin.

<table>
<thead>
<tr>
<th></th>
<th>More stable</th>
<th>Less stable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conlon</td>
<td>SOD</td>
<td>EF-2</td>
</tr>
<tr>
<td></td>
<td>S-AMS4</td>
<td>3KACoA</td>
</tr>
<tr>
<td></td>
<td>Mal de</td>
<td>Actin</td>
</tr>
<tr>
<td></td>
<td>GAPDH</td>
<td></td>
</tr>
<tr>
<td>Golden Promise</td>
<td>Mal de</td>
<td>EF-2</td>
</tr>
<tr>
<td></td>
<td>Actin</td>
<td>S-AMS4</td>
</tr>
<tr>
<td></td>
<td>3KACoA</td>
<td>SOD</td>
</tr>
<tr>
<td></td>
<td>GAPDH</td>
<td></td>
</tr>
<tr>
<td>Harrington</td>
<td>GAPDH</td>
<td>3KACoA</td>
</tr>
<tr>
<td></td>
<td>Actin</td>
<td>EF-2</td>
</tr>
<tr>
<td></td>
<td>S-AMS4</td>
<td>Mal de</td>
</tr>
<tr>
<td></td>
<td>SOD</td>
<td></td>
</tr>
<tr>
<td>HiFi</td>
<td>Actin</td>
<td>EF-2</td>
</tr>
<tr>
<td></td>
<td>Mal de</td>
<td>SOD</td>
</tr>
<tr>
<td></td>
<td>3KACoA</td>
<td>GAPDH</td>
</tr>
<tr>
<td></td>
<td>S-AMS4</td>
<td></td>
</tr>
<tr>
<td>All barley</td>
<td>Actin</td>
<td>EF-2</td>
</tr>
<tr>
<td></td>
<td>Mal de</td>
<td>S-AMS4</td>
</tr>
<tr>
<td></td>
<td>SOD</td>
<td>3KACoA</td>
</tr>
<tr>
<td></td>
<td>GAPDH</td>
<td></td>
</tr>
<tr>
<td>Barley plus oat</td>
<td>Actin</td>
<td>Mal de</td>
</tr>
<tr>
<td></td>
<td>EF-2</td>
<td>S-AMS4</td>
</tr>
<tr>
<td></td>
<td>SOD</td>
<td>GAPDH</td>
</tr>
<tr>
<td></td>
<td>3KACoA</td>
<td></td>
</tr>
</tbody>
</table>
Figure 2-1: A box-and-whisker plot of the Cq data for each candidate across each Conlon tissue and biological replicate. The x-axis is labeled by the 20 UNIGENE clusters in the order they appeared in the in silico analysis rank. The bars represent minimum and maximum Cq values, the boxes indicate the first to third quartiles and the line in each box represents the median. Hv.3082 NADH dehydrogenase, Hv.674 Phosphoglycerate mutase, Hv.2613 Proteasome subunit beta type-7-B, Hv.22848 GAPDH, Hv.9509 Elongation factor EF-2, Hv.22901 Malate dehydrogenase, Hv.19033 Heat Shock Protein 70, Hv.735 Beta-tubulin 1, Hv.22789 Translationally-controlled tumor protein homolog (HTP), Hv.2973 3-ketoacyl-CoA thiolase like, Hv.3271 Superoxide dismutase [Cu-Zn], Hv.22823 Mitochondrial ATP synthase beta subunit, Hv.22783 Elongation factor 1-alpha, Hv.20658 S-adenosylmethionine synthetase 4, Hv.22798 Cytosolic heat shock protein 90, Hv.23243 Zinc finger A20 & AN1 domain-containing stress-associated, Hv.23088 Actin, Hv.3750 Peroxiredoxin-5, mitochondrial, Hv.3381 Translation initiation factor eIF-5A, Hv.12544 Proteasome subunit alpha type-5-A.
Figure 2-2: A bar chart of the calculated stability values of the 20 candidate genes across 8 Conlon tissues. The y axis measures their geomean stability value calculated through RefFinder and the x-axis shows the candidates labeled by their UNIGENE cluster derived from the in silico rank. Hv.3082 NADH dehydrogenase, Hv.674 Phosphoglycerate mutase, Hv.2613 Proteasome subunit beta type-7-B, Hv.22848 GAPDH, Hv.9509 Elongation factor EF-2, Hv.22901 Malate dehydrogenase, Hv.19033 Heat Shock Protein 70, Hv.735 Beta-tubulin 1, Hv.22789 Translationally-controlled tumor protein homolog (HTP), Hv.2973 3-ketoacyl-CoA thiolase like, Hv.3271 Superoxide dismutase [Cu-Zn], Hv.22823 Mitochondrial ATP synthase beta subunit, Hv.22783 Elongation factor 1-alpha, Hv.20658 S-adenosylmethionine synthetase 4, Hv.22798 Cytosolic heat shock protein 90, Hv.23243 Zinc finger A20 & AN1 domain-containing stress-associated, Hv.23088 Actin, Hv.3750 Peroxiredoxin-5, mitochondrial, Hv.3381 Translation initiation factor eIF-5A, Hv.12544 Proteasome subunit alpha type-5-A.
Figure 2-3: The box-and-whisker plot for the Cq data for candidates Hv.22848 (GAPDH), Hv.9509 (elongation factor 2), Hv.22901 (malate dehydrogenase), Hv.2973 (3-ketoacyl-CoA thiolase like protein), Hv.3271 (superoxide dismutase), Hv.20658 (S-adenosylmethionine synthetase 4), and
APPENDIX 2-1

Melt curves for gene candidates 1-20 using Conlon for barley validation.
APPENDIX 2-2

Melt curves, sequences, efficiencies, and $r^2$ values for candidate primer standard curves for candidates Hv.22848 (GAPDH), Hv.9509 (elongation factor 2), Hv.22901 (malate dehydrogenase), Hv.2973 (3-ketoacyl-CoA thiolase like protein), Hv.3271 (superoxide dismutase), Hv.20658 (S-adenosylmethionine synthetase 4), and Hv.23088 (actin) using HiFi for oat validation. Oat primer sequences Hv.9509, Hv.22901, Hv.20658, and Hv.22823 are the same for oat as for barley.

<table>
<thead>
<tr>
<th>Candidate</th>
<th>Forward</th>
<th>Reverse</th>
<th>Efficiency</th>
<th>$r^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAPDH</td>
<td>GTCAACGACCCCTTCATCA</td>
<td>GCCTTGTCCTTTGTCATGA</td>
<td>107.7%</td>
<td>0.999</td>
</tr>
<tr>
<td>EF-2</td>
<td>AACTGGCATGAAGGCTGGTA</td>
<td>GGCAGGCATCAACTTCCTTC</td>
<td>94.0%</td>
<td>0.997</td>
</tr>
<tr>
<td>Mal de</td>
<td>GCACTGGGTGTGAATGTTGC</td>
<td>CTTCTCAGGGATAGATGGAGC</td>
<td>109.3%</td>
<td>0.987</td>
</tr>
<tr>
<td>3KACoA</td>
<td>GACTGGGGAAGCTAAAAACG</td>
<td>AAGAACCAGCCTCCAGCACCAT</td>
<td>93.3%</td>
<td>0.993</td>
</tr>
<tr>
<td>SOD</td>
<td>AAGTGAGTGATGTTGCTGGA</td>
<td>GATCCACTCCAACAGCAACG</td>
<td>104.4%</td>
<td>0.99</td>
</tr>
<tr>
<td>S-AMS4</td>
<td>CAACATTGAGCGAGCAGTCACC</td>
<td>GCAGCAATCTCATCGTGTGGT</td>
<td>98.7%</td>
<td>0.986</td>
</tr>
<tr>
<td>Actin</td>
<td>GAATGGTCAAGGCTGGTTTCG</td>
<td>TCCATGTCATCCAGTGGTCTT</td>
<td>107.3%</td>
<td>0.989</td>
</tr>
</tbody>
</table>
The stability placement of the 20 candidate genes from Conlon expression by 4 individual analytical programs and their comprehensive ranking by RefFinder.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Acc.</th>
<th>Candidate</th>
<th>Delta CT</th>
<th>Bestkeeper</th>
<th>Normfinder</th>
<th>Genorm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NADH de</td>
<td>11</td>
<td>1st</td>
<td>6th</td>
<td>1st</td>
<td>4th</td>
</tr>
<tr>
<td>2</td>
<td>PGM</td>
<td>6</td>
<td>2nd</td>
<td>10th</td>
<td>2nd</td>
<td>5th</td>
</tr>
<tr>
<td>3</td>
<td>PSMB7</td>
<td>5</td>
<td>3rd</td>
<td>9th</td>
<td>3rd</td>
<td>7th</td>
</tr>
<tr>
<td>4</td>
<td>GAPDH</td>
<td>10</td>
<td>5th</td>
<td>4th</td>
<td>5th</td>
<td>6th</td>
</tr>
<tr>
<td>5</td>
<td>EF-2</td>
<td>14</td>
<td>4th</td>
<td>5th</td>
<td>4th</td>
<td>8th</td>
</tr>
<tr>
<td>6</td>
<td>Mal de</td>
<td>20</td>
<td>8th</td>
<td>18th</td>
<td>8th</td>
<td>Tied 1st</td>
</tr>
<tr>
<td>7</td>
<td>HSP 70</td>
<td>17</td>
<td>6th</td>
<td>15th</td>
<td>6th</td>
<td>3rd</td>
</tr>
<tr>
<td>8</td>
<td>β tub</td>
<td>19</td>
<td>10th</td>
<td>2nd</td>
<td>10th</td>
<td>11th</td>
</tr>
<tr>
<td>9</td>
<td>HTP</td>
<td>1</td>
<td>11th</td>
<td>17th</td>
<td>12th</td>
<td>Tied 1st</td>
</tr>
<tr>
<td>10</td>
<td>3KCoA</td>
<td>9</td>
<td>14th</td>
<td>1st</td>
<td>14th</td>
<td>15th</td>
</tr>
<tr>
<td>11</td>
<td>SOD</td>
<td>16</td>
<td>7th</td>
<td>13th</td>
<td>7th</td>
<td>12th</td>
</tr>
<tr>
<td>12</td>
<td>ATP5B</td>
<td>13</td>
<td>12th</td>
<td>8th</td>
<td>11th</td>
<td>10th</td>
</tr>
<tr>
<td>13</td>
<td>EF-1α</td>
<td>7</td>
<td>9th</td>
<td>12th</td>
<td>9th</td>
<td>13th</td>
</tr>
<tr>
<td>14</td>
<td>S-AMS4</td>
<td>3</td>
<td>18th</td>
<td>3rd</td>
<td>18th</td>
<td>17th</td>
</tr>
<tr>
<td>15</td>
<td>HSP 90</td>
<td>2</td>
<td>13th</td>
<td>16th</td>
<td>13th</td>
<td>9th</td>
</tr>
<tr>
<td>16</td>
<td>ZF-stress as.</td>
<td>8</td>
<td>16th</td>
<td>7th</td>
<td>16th</td>
<td>16th</td>
</tr>
<tr>
<td>17</td>
<td>Actin</td>
<td>12</td>
<td>15th</td>
<td>11th</td>
<td>15th</td>
<td>14th</td>
</tr>
<tr>
<td>18</td>
<td>PRDX5</td>
<td>15</td>
<td>17th</td>
<td>14th</td>
<td>17th</td>
<td>18th</td>
</tr>
<tr>
<td>19</td>
<td>eIF-5A</td>
<td>4</td>
<td>19th</td>
<td>19th</td>
<td>19th</td>
<td>19th</td>
</tr>
<tr>
<td>20</td>
<td>PSAM5</td>
<td>18</td>
<td>20th</td>
<td>20th</td>
<td>20th</td>
<td>20th</td>
</tr>
</tbody>
</table>
CHAPTER 3: βGLUX IS A GLUCOSIDASE CO-EXPRESSED WITH CSLF6 THROUGH OAT AND BARLEY CARYOPSIS STAGES

ABSTRACT

The biosynthesis and pathway of hemicellulose (1,3;1,4)-β-D (mixed linkage glucan or MLG) has not yet been detailed. The production of MLG is directly linked to the expression of the gene CslF6. CslF6 is known to use UDP-glucose and its productivity may depend on the activity of a glucosidase supplying these monomers. A glucosidase showing co-expression with CslF6 may, therefore, play an important role in MLG quantity. A search for glucosidase candidates was performed in silico for oat using the public expression data of the model species barley. Seventy-three glucosidase accessions were analyzed to find the glucosidase candidate most likely to co-express with CslF6 in developing grains. The accession MLOC_58487, referred to here as βglux, was selected for expression analysis in HiFi oat using quantitative PCR. βglux activity was shown to generally coincide with that of CslF6 through caryopsis stage development. This glucosidase is recommended for further research to uncover any effects it may have on (1,3;1,4)-β-D glucan content.

INTRODUCTION

Mixed linkage glucan ((1,3;1,4)-β-D glucan or MLG) is a hemicellulose recognized for lowering serum cholesterol and increasing bile excretion (Lia et al. 1995). It is present in Poaceae crops and is highest in oat (Avena sativa) and barley (Hordeum vulgare) (Jorgensen 1988; McClear 1985). These two grass plants are among the most used crops in the world, others being corn, rice, wheat, sorghum, and rye (USDA 2015). Though sought for its dietary benefits, MLG is also detrimental to barley malting quality (Zhang et al. 2001). This hemicellulose is an
unbranched polysaccharide of D-glucose monomers with intermittent 1-3 and 1-4 linkages in β orientation (Burton and Fincher 2009). The presence of MLG has been directly linked to cellulose-synthase like (Csl) genes, principally CslF6 (Taketa 2012). However, the biosynthetic pathway for MLG is still unexplained, along with the factors that influence MLG quantity and quality.

Because MLG synthesis is dependent upon UDP-glucose (Gibeaut and Carpita 1993), it is likely that a glucosidase may be upregulated in tandem with CslF6 for its cytosolic glucose needs. By comparing CslF6 expression patterns with those of candidate glucosidases, a glucosidase may be found to be co-expressed, hinting at involvement in the same pathway (Lee et al. 2004; Oliver 2000; Wolfe et al. 2005). With the use of public bioinformatics databases, candidate genes can be screened for expression in silico. Because of the limitations in oat databases, barley is used here as a model because the two are closely related and barley, too, produces large quantities of MLG. The candidates co-expressing with CslF6 through in silico analysis can then be measured empirically. The candidate shown to successfully co-express with CslF6 should be further investigated for any influences it has on the MLG content in producing cells.

MATERIALS AND METHODS

Plant materials- ‘Hifi’ oat seeds from the USDA-ARS in Aberdeen, Idaho, were grown under normal greenhouse conditions. Tissues for this experiment were: roots and shoots of 10 day old seedlings, tiller leaves and stems of plants at a six-tiller stage, caryopsis seeds at 0, 5, 10, 15, 20, 25, and 30 days post anthesis (DPA), and seeds 4 days post germination (DPG)
**DNA extraction**-DNA was extracted from oat leaf tissue after being freeze-dried overnight and ground to powder. The powder was added to an extraction made from 100% ethanol, salts buffer, SDS, phenanthroline, and β-mercaptoethanol. The extraction was incubated at 68°C before adding KOAc. DNA was precipitated on ice before adding cold isopropanol, then physically removed with a hook and dissolved in TE buffer overnight. The DNA was then isolated using chloroform-phenol, chloroform-octanol (sevag), and isopropanol sequentially with centrifugation between each step. DNA was precipitated using 3M sodium acetate and cold 100% isopropanol at -70°C, then resuspended in TER buffer.

**RNA extraction**- Three RNA samples were extracted for each plant material. RNA extraction was performed using the GeneJET Plant Purification Minikit (Thermo Fisher Scientific, Waltham, MA, USA). Samples were then treated with DNase I in DNase 5x buffer (New England Biolabs, Ipswich, MA, USA). The RNA was then isolated through LiCl precipitation. Samples with high starch levels (all 4 DPG seeds and leaf and shoot tissues for oats), which clogged the GeneJet spin columns, were processed through a Trizol (Thermo Fisher Scientific, Waltham, MA, USA) extraction described by Li and Trick (2005) and isolated through LiCl precipitation. RNA samples were observed for quality using gel electrophoresis and OD measurements. Samples having intact 28s and 18s bands with OD 260/280 and 260/230 greater than 1.7 were selected for use. The RNA samples were checked for genomic DNA contamination by running them parallel with cDNA through real-time PCR. Sample cDNA was generated through the iScript kit (Bio-Rad, Hercules, CA, USA).
**In silico analysis of glucosidase candidates for CslF6 co-expression** - A search was performed for barley glucosidases using the Plant Genome and Systems Biology (PGSB, http://pgsb.helmholtz-muenchen.de/plant/index.jsp) online database. Those accessions with expression data were accessed through Expression Atlas (https://www.ebi.ac.uk/gxa/home). The expression data for the eight tissues (caryopsis, 15 days post anthesis, caryopsis 5 days post anthesis, germinating, embryo, developing inflorescence at 1 cm, developing inflorescence at 5 mm, internode, root, and shoot) were recorded for *in silico* co-expression analysis. CslF6 expression was also acquired through Expression Atlas.

**Statistical Analysis** - The expression data for each glucosidase was compared to that of CslF6 through several methods. Pearson’s correlation, distance rank average, and Euclidean distance were used to determine the glucosidases with the greatest conformity to CslF6 expression. The Pearson’s Product-Moment Correlation divided the covariance between each glucosidase accession data set and the CslF6 accession data set by the product of their respective standard deviations. A distance rank average was made by finding the distance between each glucosidase and CslF6 data point at each tissue, ranking each distance by the smallest value, and then averaging the eight ranks for each candidate to a consensus rank. The Euclidean distance between each glucosidase and CslF6 was found by treating their data sets as a point in eight dimensions. To illustrate the degree of difference across the expression levels, bar graphs were constructed to represent each candidate in each tissue.

**Sequencing of the glucosidase ßglux in oat** - Oat ESTs from the CORE libraries were assembled to the reference of the barley glucosidase MLOC_58487 from Ensembl Plants.
(http://plants.ensembl.org/index.html) using Sequencher (Sequencher 5.2.4) with 85% minimum match, minimum 20 bp overlap, and under the dirty data setting. A consensus sequence was derived from the aligned oat reads. Primers were then designed using Primer 3 (http://primer3.ut.ee) over multiple sections of the predicted coding sequence and tested on genomic DNA and cDNA. Primers which encompassed the coding sequence were used to amplify the entire nucleotide sequence.

To sequence ßglux in hexaploid Avena sativa, gene copies were isolated in bacterial clones prior to sequencing. ßglux amplicons were ligated into copies of a PUC vector, used in the transformation of One-Shot Competent Cells (Invitrogen, Carlsbad, CA) and then grown in colonies. Single colonies were screened by drug resistance to kanamycin and also through confirmation of the insert through PCR and electrophoresis. DNA was then used from positive colonies for sequencing.

Samples for sequencing ßglux were prepared using Big Dye protocol and submitted for standard Sanger dideoxy sequencing to the Brigham Young University DNA Sequencing Center. Resulting reads, labeled with their originating colony and primers used, were trimmed and assembled using Sequencher. The reads were trimmed and aligned with 80% minimum match, 20 bp minimum overlap, and under the dirty data setting. After forming a single contig, reads were removed when observed to have reoccurring disagreements with the consensus, then reassembled separately. When three contigs with very few disagreements were found, the HiFi and SolFi reads for each contig were separated and then realigned to a consensus sequence of
that contig. The six resulting contigs (1, 2, and 3 for HiFi and SolFi) were aligned to the previous CORE EST βglux contig to identify exon-intron junctions.

**Gene quantification and analysis**- Transcript quantification was performed through real-time PCR. The HiFi cDNA samples were run in triplicate using SsoFast EvaGreen Supermix (Bio-Rad, Hercules, CA, USA). Each cDNA sample was run with primer pairs for the reference genes, actin, elongation factor 2, and malate dehydrogenase, and with the target genes, CslF6 and βglux. Primers were validated for efficient real-time PCR amplification. For all real-time experiments, the SsoFast Evagreen Supermix (Bio-Rad) protocol was followed for 20 µL reactions, using 0.5 µL of each forward and reverse primer, and 2.5 ng of template cDNA for each tissue. The thermal cycler program went through an activation step of 95° C for 30 seconds, 40 cycles of a melting step at 95° C for 5 seconds and an annealing/extension step of 60° C, followed by a melt curve step from 65-95° C in increments of 0.5° C every 5 seconds. A CFX machine and manager program (Bio-Rad, Hercules, CA, USA) measured and interpreted fluorescent signals for quantitative data.

The expression of βglux and CslF6 were normalized through the ΔΔct method. Actin, elongation factor 2, and malate dehydrogenase were used to normalize expression and the stem tissue data was used as relative control.

**RESULTS**

**In silico analysis glucosidase candidates for CslF6 co-expression** - The PGSB search returned 139 glucosidase accessions. Expression data was found for 73 of them. Expression data was presented as internal comparison units (FPKM/TPM for transcripts) for each gene found in each
of eight tissues (Appendix 3-1). The accession for CslF6 was MLOC_57200. Table 3-1 shows the top ten glucosidase candidates ranked by each approach: correlation, the distance rank average, Euclidean distance, and overall placement. The Pearson’s Product-Moment Correlation showed 40 of 73 candidates with positive correlations to CslF6. The average distance rank Euclidean distance rank shows the placement of each accession by its distance to MLOC_57200 from each calculation. Overall, MLOC_58487 was the top final candidate.

Bar graphs illustrating the expression levels of glucosidase candidates and MLOC_57200 (HvCslF6) are shown in Figures 3-1 to 3-8. A visual cap was set at 100 FPKM/TPM units because the presences of excessively high transcript values distorted the visualization, and were much larger than MLOC_57200.

**Oat βglux sequence**- The oat βglux contig derived from CORE assembly encoded a hypothetical protein 616 aa in length—two amino acids longer than the MLOC_58487 coding sequence. Primers that amplified portions of βglux and were used in sequencing are shown in Table 3-3 with their nucleotide positions. Six contigs were assembled, belonging to three groups of ‘HiFi’ or ‘SolFi’ colonies. Reads from HiFi colonies 16, 24, 12, 34, and 22 were found in HF_contig_1; reads from colonies 50, 28, and 15 in HF_contig_2; and reads from colony 29 in HF_contig_3. Reads from SolFi colonies 26, 47, and 43 were found in SF_contig_1; reads from colonies 78, 56, 65, 48, and 55 in SF_contig_2; and reads from colonies 91, 32, 5, 95, and 71 in SF_contig_3. The sequences are available in Appendix 2. Once aligned, the distance from the first nucleotide of the start codon (at bp 100) to the last nucleotide of the stop codon (at bp 3,096)
was 2,997 bp in length. Upon their alignment to the oat CORE contig, the coding sequence is divided into eight exons listed in Table 3-3.

**Real-time PCR and quantification**- Primers used for qPCR are given in Table 3-4. The expression patterns for CslF6 and βglux are shown in figure 3-10. Throughout the caryopsis stages, βglux expression tends to be higher than CslF6 expression. The two gene activities tend to follow the same trend through caryopsis except for the 25 DPA time points where CslF6 decreases sharply before increasing again, but βglux appears to maintain steady expression before sharply increasing in expression.

**DISCUSSION**

Several methods were used to determine which glucosidase candidate has the expression pattern best matching that of HvCslF6 and each method offered different results. The Pearson’s Product Moment Correlation results included candidates, like the top ranked MLOC_57785, which express in tissues at magnitudes lower than CslF6—even not at all—and follow the same trend. In this case, correlation alone may not be so significant. Therefore, proximity to expression was also judged using distance rank and the Euclidean distance, each showing different top candidates. Averaging the rank each candidate placed for each method, MLOC_58487 was observed to be the most correlated and most proximate candidate. Figure 3-8 represents the transcript levels of the final top five candidates in comparison to MLOC_57200.

The most highly correlated candidates still had low coefficients of variation with CslF6, requiring a closer look at variations in tissues. It is possible that there is no glucosidase upregulated with CslF6 activity or that a glucosidase exists but had no expression data for the in
**silico** analysis. However, CslF6 may use UDP-glucose derived from free glucose produced by MLOC_58487 in 15 DPA caryopsis, germinating embryo, and such, but another glucosidase may be actively supplying the free sugars in the internode. Perhaps a biological process in the internode tissue already produces free sugars and MLOC_58487 is down regulated. Under this hypothesis, or another where MLOC_58487 is only sometimes participating in the MLG synthetic pathway, the tissues of interest need to be considered. For MLG synthesis in barley grain, the most important time point among tissues from the **in silico** investigation is the 15 days post anthesis for caryopsis tissue. Coles (1979) showed an increase of free sugars from 7 to 13 DPA, followed by a transient decline, with an increase again at 15 DPA until 31 DPA. From that study, MLG content was seen to be increasing sharply in the cultivars ‘Minerva’ and ‘Triumph’ until about 20-25 DPA, depending on the treatment, and then increased at a slower rate. Tsuchiya (2005) also found MLG production to increase in ‘Shikokuhadaka 97’ barley from 12 to 30 DPA, with MLG synthase enzyme activity peaking around 19 DPA. Considering the glucosidases from the current **in silico** approach, there were three represented in the 15 DPA barley caryopsis tissue that were expressed at similar levels as MLOC_57200, and which therefore might be providing the free sugars for MLG production. They were accessions MLOC_9865 (number 12), MLOC_68876 (number 67), and MLOC_58487 (number 27).

Of the three candidates with expression levels close to CslF6 in 15 DPA tissue, MLOC_58487 was the most likely candidate. MLOC_9865 showed the largest expression in any data point across all tissues at 2419 FPKM/TPM for 15 DPA. In 5 DPA caryopsis tissue, it was measured at 249 FPKM/TPM. In all other tissues, it expressed little or not at all with the
exception of 30 FPKM/TPM in germinating embryos. This accession likely represents a glucosidase that is highly expressed in grain development and germination but does not associate well with the level of MLOC_57200 expression. This candidate was annotated as a beta-mannosidase by UniProt (http://www.uniprot.org/), releasing mannose monomers rather than glucose monomers. MLOC_68876, expresses at levels close to MLOC_57200 in caryopses 15 DPA, and root. It expresses as lower in the internode, but higher in the tissues caryopsis 5 DPA, germinating embryo, shoot, and inflorescence. It is annotated by UniProt as an α-glucosidase, breaking starch and disaccharides to glucose monomers. MLOC_58487 was predicted earlier as the most highly co-expressed option for CslF6. The two had relatively similar expression patterns in all tissues except in the internode and caryopsis 5 DPA tissue, where MLOC_58487 was expressed in a noticeably smaller amount. In the internode, the expression was negligible but in the caryopsis 5 DPA tissue its presence was still significant. Of the three accessions having significant expression in the maturing caryopsis at 15 DPA, MLOC_58487 appeared to be the best candidate for co-expression of CslF6, possibly supplying glucose for MLG production.

MLOC_58487 is an uncharacterized protein in UniProt, but its sequence is classified as belonging to the glycoside hydrolase 3 family. This family consists of several subcategories, including β-glucosidase enzymes. In barley, 614 amino acids of βglux (positions 30-359 and 397-605) identify with the glycosyl hydrolase 3 family. A BLAST search through NCBI (http://www.ncbi.nlm.nih.gov/) for the translated sequence returned the specific hit for a periplasmic E. coli beta glucosidase described by Yang et al (1996), and a 3D structure by Varghese, Hrmova, and Fincher (1999). A beta glucosidase, enzyme classification 3.2.1.21
(http://www.chem.qmul.ac.uk/iubmb/enzyme/EC3/2/1/21.html), was expected by Shu and Rasmussen (2014) to break down cellulose into β-glucose monomers, which could subsequently become converted to GDP or UDP glucose for synthesizing other polysaccharides, including MLG. It is possible, therefore, that MLG quantity in cereal grains may depend in part on the activity of those enzymes breaking down cellulose during caryopsis development.

The contigs found for βglux from sequencing HiFi and SolFi DNA appear to represent hexaploid genomes. Reads sometimes had better similarity across cultivars rather than to reads of the same cultivar. This pairing suggests the homeologous contigs were found and are descended from the same ancestral genome. Having divided the contigs into three pairs, it appears each of the oat genomes are represented. There is a strong similarity between contigs 1 and 2 for both HiFi and SolFi. It may be that they belong to the same genome and the differences seen are due to heterologous alleles or sequencing errors. Concerning the possibility of missing one genome, however, if the A, C, and D genomes between the SolFi and HiFi are considered as the same class of genome, and if the likelihood of a competent cell being transformed by either genome PCR product is equal, then the trinomial probability of having each genome represented in at least one of the twenty-two colonies is above 99.9%. If one of the genomes is not represented, it may be because the gene-encompassing primers for the PCR product were not applicable to all genomes.

The expression data coincides with the expectation set up in chapter 3; βglux is active during caryopsis development and shows an expression pattern similar to that of CsIF6. There appears to be a steep reduction of CsIF6 expression at day 25, but it returns to a higher net gain
at the next time point. From 15 DPA, βglux continues to rise in activity. Also, as noticed in the investigation for a glucosidase in barley, βglux does not always co-express with CslF6. In shoot and leaf tissue, βglux activity was far lower than CslF6 activity. CslF6 appears very active at these time points, and, following the hypothesis set out in this project, would need another active glucosidase. It may be that, during these stages, another biochemical pathway upregulates a separate glucosidase which is able to satisfy the demand for glucose monomers.

Now that this co-expression is verified in HiFi oats, opportunities for future research open. There is still much about the glucosidase βglux to be known. As mentioned previously, co-expression is a lead for finding genes which participate in the same pathway (Lee et al. 2004; Oliver 2000; Wolfe et al. 2005). This co-expression, however, was observed in the available barley expression data and one oat variety. Therefore, it should be observed further in other oat varieties, especially those with known MLG profiles. Do those with lower MLG level tendencies also express βglux at low levels during caryopsis development? Is there a correlation between βglux activity and the length of βglux chains? In regards to oat genomes, Coon (2012) observed differences between oat genomes for CslF6 activity in several cultivars. Would βglux be more active in the same genome as that with the highest CslF6 activity? Sequence data in this report can help design experiments to answer that. Concerning transgenic experiments, would a distinct effect on MLG production be seen with knockout, knockdown, or upregulation of βglux genes? Future work is needed to characterize this new gene of interest and verify its effects, if any, on MLG production. If it is an important contributor, then it may be another targetable resource for better cereal grains.
CONCLUSION

Having compared expression levels from Expression Atlas through statistical and visual methods, the candidate accession MLOC_58487 was selected as being the most likely to represent a glucosidase gene whose expression is closely associated with HvCslF6. The gene will henceforth be referred to as βglux. The candidate will be utilized in further analysis by comparing expression of the oat homologues for these genes.
LITERATURE CITED


TABLES

Table 3-1: The top ten candidates ranked from each statistical method approached.

<table>
<thead>
<tr>
<th>Correlation</th>
<th>Distance Rank Average</th>
<th>Euclidean Distance</th>
<th>Final</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MLOC_57785</td>
<td>MLOC_20723</td>
<td>MLOC_10919</td>
</tr>
<tr>
<td>2</td>
<td>MLOC_63574</td>
<td>MLOC_58487</td>
<td>MLOC_20723</td>
</tr>
<tr>
<td>3</td>
<td>MLOC_11508</td>
<td>MLOC_54331</td>
<td>MLOC_15619</td>
</tr>
<tr>
<td>4</td>
<td>MLOC_58487</td>
<td>MLOC_66806</td>
<td>MLOC_8615</td>
</tr>
<tr>
<td>5</td>
<td>MLOC_11885</td>
<td>MLOC_10919</td>
<td>MLOC_58487</td>
</tr>
<tr>
<td>6</td>
<td>MLOC_9865</td>
<td>MLOC_11508</td>
<td>MLOC_54331</td>
</tr>
<tr>
<td>7</td>
<td>MLOC_16595</td>
<td>MLOC_5621</td>
<td>MLOC_1898</td>
</tr>
<tr>
<td>8</td>
<td>MLOC_66806</td>
<td>MLOC_8669</td>
<td>MLOC_11508</td>
</tr>
<tr>
<td>9</td>
<td>MLOC_37740</td>
<td>MLOC_61073</td>
<td>MLOC_61073</td>
</tr>
<tr>
<td>10</td>
<td>MLOC_15694</td>
<td>MLOC_51368</td>
<td>MLOC_66806</td>
</tr>
</tbody>
</table>

Table 3-2: Primers used for the sequencing of cloned ßglux inserts. Position is described here as relative to the first nucleotide to the start codon. The * indicates sequences of the M13 primers included for the vector regions surrounding the cloned inserts.

<table>
<thead>
<tr>
<th>Direction</th>
<th>Primer Sequence</th>
<th>Sense 5’ End</th>
<th>Sense 3’ End</th>
</tr>
</thead>
<tbody>
<tr>
<td>FWD*</td>
<td>TGTAACACGACGGCCAGT</td>
<td>Trimmed</td>
<td></td>
</tr>
<tr>
<td>FWD</td>
<td>CGTTTGTAGGGATCCGAGGT</td>
<td>730</td>
<td>749</td>
</tr>
<tr>
<td>REV</td>
<td>CCACGTCTCCATGCAAG</td>
<td>1592</td>
<td>1611</td>
</tr>
<tr>
<td>FWD</td>
<td>TCACGGAATTGATGATGCTGT</td>
<td>1592</td>
<td>1611</td>
</tr>
<tr>
<td>REV</td>
<td>CCCTTAGAATCCGCTCAACA</td>
<td>1609</td>
<td>1628</td>
</tr>
<tr>
<td>REV</td>
<td>ACAAGCATCTGGAGGCACAT</td>
<td>2594</td>
<td>2613</td>
</tr>
<tr>
<td>REV*</td>
<td>AGCGGATAACACATTTTCACACAGG</td>
<td>Trimmed</td>
<td></td>
</tr>
</tbody>
</table>
Table 3-3: Exon position and length as figured from the alignment of 3 HiFi contigs and 3 SolFi contigs to the CORE EST consensus contig. The positions are numbered by their relation to the first nucleotide of the start codon.

<table>
<thead>
<tr>
<th>Exon</th>
<th>Position</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1-377</td>
<td>377</td>
</tr>
<tr>
<td>2</td>
<td>453-543</td>
<td>91</td>
</tr>
<tr>
<td>3</td>
<td>640-788</td>
<td>147</td>
</tr>
<tr>
<td>4</td>
<td>871-1114</td>
<td>244</td>
</tr>
<tr>
<td>5</td>
<td>1247-1351</td>
<td>105</td>
</tr>
<tr>
<td>6</td>
<td>1426-1608</td>
<td>183</td>
</tr>
<tr>
<td>7</td>
<td>1722-1923</td>
<td>202</td>
</tr>
<tr>
<td>8</td>
<td>2501-2997</td>
<td>497</td>
</tr>
</tbody>
</table>

Table 3-4: Primers used for real-time PCR with their efficiency and r2 values. The * indicate primers used by Coon 2012.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Efficiency</th>
<th>r2</th>
</tr>
</thead>
<tbody>
<tr>
<td>ßglux Fwd</td>
<td>AGATGCTCTAGTTGCTGCGT</td>
<td>107.5%</td>
<td>0.980</td>
</tr>
<tr>
<td>ßglux Rev</td>
<td>AAACATTTTCAACCCCATATCCAAAAGG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Csff6 Fwd_16*</td>
<td>AGGACGTCGTCACAGGCTAC</td>
<td>102.8%</td>
<td>0.983</td>
</tr>
<tr>
<td>Csff6 Rev_16*</td>
<td>ACGATGCCAGGTAGACGTA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Actin Fwd</td>
<td>GAATGGTCAAGGCTGGTTTCG</td>
<td>107.3%</td>
<td>0.989</td>
</tr>
<tr>
<td>Actin Rev</td>
<td>TCCATGTCAATCCAGTTGCT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EF2 Fwd</td>
<td>AACTGGCATGAAGGCTCCGTA</td>
<td>94.0%</td>
<td>0.997</td>
</tr>
<tr>
<td>EF2 Rev</td>
<td>GGCAGGCATCAACTTCTCTTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Malate De Fwd</td>
<td>GCACTGGTGTAAGATGTTGC</td>
<td>109.3%</td>
<td>0.987</td>
</tr>
<tr>
<td>Malate De Rev</td>
<td>CTTCTCAGGGATAGATGGAGC</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
FIGURES

Figure 3-1: Glucosidase transcript levels in developing influence 1cm tissue. Arbitrary transcript units (Y-axis) for each accession (X-axis). Window cap at 100 units. MLOC_57200 is colored black and is the last accession.

Figure 3-2: Glucosidase transcript levels in developing influence 5mm tissue. Arbitrary transcript units (Y-axis) for each accession (X-axis). Window cap at 100 units. MLOC_57200 is colored black and is the last accession.
Figure 3-3: Glucosidase transcript levels in shoot tissue. Arbitrary transcript units (Y-axis) for each accession (X-axis). Window cap at 100 units. MLOC_57200 is colored black and is the last accession.

Figure 3-4: Glucosidase transcript levels in root tissue. Arbitrary transcript units (Y-axis) for each accession (X-axis). Window cap at 100 units. MLOC_57200 is colored black and is the last accession.
Figure 3-5: Glucosidase transcript levels in germinating embryo tissue. Arbitrary transcript units (Y-axis) for each accession (X-axis). Window cap at 100 units. MLOC_57200 is colored black and is the last accession.

Figure 3-6: Glucosidase transcript levels in internode tissue. Arbitrary transcript units (Y-axis) for each accession (X-axis). Window cap at 100 units. MLOC_57200 is colored black and is the last accession.
Figure 3-7 Glucosidase transcript levels in caryopsis tissue 5 days post anthesis (DPA). Arbitrary transcript units (Y-axis) for each accession (X-axis). Window cap at 100 units. MLOC_57200 is colored black and is the last accession.

Figure 3-8: Glucosidase transcript levels in caryopsis tissue 15 days post anthesis (DPA). Arbitrary transcript units (Y-axis) for each accession (X-axis). Window cap at 100 units. MLOC_57200 is colored black and is the last accession.
Figure 3-9: A bar chart representing the levels of CslF6 and each of the top 5 glucosidase candidates from the *in silico* across all eight tissues. DPA=days post anthesis.
Figure 3-10: Expression of CslF6 and βglux. Expression levels are represented in a logarithmic scale relative to the target gene expression in the stem control. DPA = days post anthesis. DPG = days post germination. Points represent median value and bars represent the minimum and maximum values. Dotted lines were used to connect chronological data. Blue diamonds represent βglux. Green squares represent CslF6.
The 73 glucosidase candidates with expression data from Expression Atlas for Barley. CslF6 is represented in bold as MLOC_57200.

<table>
<thead>
<tr>
<th>Accession</th>
<th>Developing</th>
<th>Developing</th>
<th>Shoot</th>
<th>Root</th>
<th>Germinating</th>
<th>Intermode</th>
<th>Caryopsis</th>
<th>Caryopsis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accession</td>
<td>1 cm</td>
<td>5 mm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MLOC_61980</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MLOC_18750</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>MLOC_56792</td>
<td>0</td>
<td>0</td>
<td>14</td>
<td>56</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MLOC_37976</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>MLOC_80849</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MLOC_10669</td>
<td>0</td>
<td>0</td>
<td>0.7</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MLOC_13831</td>
<td>0.7</td>
<td>0.8</td>
<td>0</td>
<td>0.9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MLOC_66576</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>7</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MLOC_57283</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>0.7</td>
<td>0</td>
<td>0</td>
<td>0.7</td>
</tr>
<tr>
<td>MLOC_62746</td>
<td>0</td>
<td>0</td>
<td>84</td>
<td>408</td>
<td>89</td>
<td>115</td>
<td>0</td>
<td>27</td>
</tr>
<tr>
<td>MLOC_79756</td>
<td>0</td>
<td>0</td>
<td>18</td>
<td>15</td>
<td>249</td>
<td>4</td>
<td>80</td>
<td>0</td>
</tr>
<tr>
<td>MLOC_9865</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>7</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MLOC_34767</td>
<td>0</td>
<td>0</td>
<td>22</td>
<td>28</td>
<td>21</td>
<td>126</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>MLOC_8615</td>
<td>0</td>
<td>0</td>
<td>42</td>
<td>38</td>
<td>23</td>
<td>68</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>MLOC_7890</td>
<td>0</td>
<td>0</td>
<td>60</td>
<td>61</td>
<td>25</td>
<td>0.6</td>
<td>2</td>
<td>19</td>
</tr>
<tr>
<td>MLOC_44074</td>
<td>0</td>
<td>0</td>
<td>27</td>
<td>29</td>
<td>19</td>
<td>15</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>MLOC_57785</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>9</td>
<td>2</td>
<td>1</td>
<td>13</td>
<td>8</td>
</tr>
<tr>
<td>MLOC_16506</td>
<td>10</td>
<td>0.7</td>
<td>17</td>
<td>0.9</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MLOC_66740</td>
<td>0</td>
<td>0</td>
<td>16</td>
<td>2</td>
<td>0.8</td>
<td>6</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>MLOC_24253</td>
<td>77</td>
<td>85</td>
<td>1</td>
<td>15</td>
<td>20</td>
<td>0</td>
<td>28</td>
<td>18</td>
</tr>
<tr>
<td>MLOC_11508</td>
<td>0</td>
<td>2</td>
<td>11</td>
<td>21</td>
<td>34</td>
<td>23</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>MLOC_5621</td>
<td>0</td>
<td>3</td>
<td>20</td>
<td>23</td>
<td>22</td>
<td>12</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>MLOC_75585</td>
<td>10</td>
<td>12</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>MLOC_4960</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>12</td>
<td>3</td>
<td>0</td>
<td>0.6</td>
</tr>
<tr>
<td>MLOC_37740</td>
<td>155</td>
<td>150</td>
<td>77</td>
<td>284</td>
<td>560</td>
<td>256</td>
<td>711</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>MLOC_69905</td>
<td>89</td>
<td>92</td>
<td>19</td>
<td>46</td>
<td>79</td>
<td>40</td>
<td>110</td>
</tr>
<tr>
<td>27</td>
<td>MLOC_58487</td>
<td>10</td>
<td>15</td>
<td>29</td>
<td>67</td>
<td>71</td>
<td>6</td>
<td>36</td>
</tr>
<tr>
<td>28</td>
<td>MLOC_66806</td>
<td>12</td>
<td>9</td>
<td>9</td>
<td>31</td>
<td>110</td>
<td>9</td>
<td>127</td>
</tr>
<tr>
<td>29</td>
<td>MLOC_10919</td>
<td>25</td>
<td>5</td>
<td>12</td>
<td>43</td>
<td>58</td>
<td>12</td>
<td>82</td>
</tr>
<tr>
<td>30</td>
<td>MLOC_1898</td>
<td>26</td>
<td>26</td>
<td>6</td>
<td>11</td>
<td>16</td>
<td>8</td>
<td>26</td>
</tr>
<tr>
<td>31</td>
<td>MLOC_54331</td>
<td>8</td>
<td>10</td>
<td>15</td>
<td>19</td>
<td>14</td>
<td>47</td>
<td>16</td>
</tr>
<tr>
<td>32</td>
<td>MLOC_61073</td>
<td>23</td>
<td>21</td>
<td>11</td>
<td>17</td>
<td>16</td>
<td>10</td>
<td>26</td>
</tr>
<tr>
<td>33</td>
<td>MLOC_37773</td>
<td>26</td>
<td>27</td>
<td>3</td>
<td>9</td>
<td>19</td>
<td>7</td>
<td>24</td>
</tr>
<tr>
<td>34</td>
<td>MLOC_76578</td>
<td>26</td>
<td>22</td>
<td>10</td>
<td>7</td>
<td>10</td>
<td>6</td>
<td>19</td>
</tr>
<tr>
<td>35</td>
<td>MLOC_8669</td>
<td>8</td>
<td>6</td>
<td>4</td>
<td>18</td>
<td>14</td>
<td>18</td>
<td>6</td>
</tr>
<tr>
<td>36</td>
<td>MLOC_20526</td>
<td>7</td>
<td>6</td>
<td>4</td>
<td>8</td>
<td>11</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>37</td>
<td>MLOC_67725</td>
<td>0.7</td>
<td>1</td>
<td>1</td>
<td>10</td>
<td>34</td>
<td>0.6</td>
<td>6</td>
</tr>
<tr>
<td>38</td>
<td>MLOC_44982</td>
<td>3</td>
<td>3</td>
<td>11</td>
<td>12</td>
<td>2</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>39</td>
<td>MLOC_60711</td>
<td>10</td>
<td>12</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>40</td>
<td>MLOC_15694</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>19</td>
</tr>
<tr>
<td>41</td>
<td>MLOC_11885</td>
<td>0.9</td>
<td>0.8</td>
<td>5</td>
<td>5</td>
<td>0.9</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>42</td>
<td>MLOC_61690</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>43</td>
<td>MLOC_79885</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>44</td>
<td>MLOC_65592</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.7</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>45</td>
<td>MLOC_56112</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.7</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>46</td>
<td>MLOC_16595</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>47</td>
<td>MLOC_4276</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>48</td>
<td>MLOC_77701</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>49</td>
<td>MLOC_76387</td>
<td>0</td>
<td>0.8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>50</td>
<td>MLOC_73077</td>
<td>0</td>
<td>0</td>
<td>59</td>
<td>3</td>
<td>62</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>51</td>
<td>MLOC_13908</td>
<td>0</td>
<td>0</td>
<td>37</td>
<td>36</td>
<td>15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>52</td>
<td>MLOC_67705</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>30</td>
<td>7</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>53</td>
<td>MLOC_72542</td>
<td>0</td>
<td>0</td>
<td>0.8</td>
<td>21</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>54</td>
<td>MLOC_81871</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>55</td>
<td>MLOC_39219</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0.8</td>
<td>0.9</td>
<td>0</td>
</tr>
<tr>
<td>56</td>
<td>MLOC_51439</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.7</td>
<td>0.9</td>
<td>0</td>
</tr>
<tr>
<td>57</td>
<td>MLOC_13058</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>0.8</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>58</td>
<td>MLOC_60415</td>
<td>0.7</td>
<td>2</td>
<td>0</td>
<td>14</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>59</td>
<td>MLOC_45963</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.8</td>
</tr>
<tr>
<td>60</td>
<td>MLOC_51849</td>
<td>0</td>
<td>0.8</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>MLOC_15992</td>
<td>0</td>
<td>0</td>
<td>44</td>
<td>20</td>
<td>3</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>---</td>
<td>------------</td>
<td>---</td>
<td>---</td>
<td>----</td>
<td>----</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>62</td>
<td>MLOC_75464</td>
<td>23</td>
<td>32</td>
<td>2</td>
<td>9</td>
<td>19</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>63</td>
<td>MLOC_14239</td>
<td>13</td>
<td>18</td>
<td>0.9</td>
<td>5</td>
<td>6</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>64</td>
<td>MLOC_3522</td>
<td>1</td>
<td>0.8</td>
<td>0.7</td>
<td>7</td>
<td>8</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>65</td>
<td>MLOC_63284</td>
<td>0.8</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>0.7</td>
</tr>
<tr>
<td>66</td>
<td>MLOC_822</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>67</td>
<td>MLOC_68876</td>
<td>66</td>
<td>100</td>
<td>52</td>
<td>51</td>
<td>100</td>
<td>35</td>
<td>324</td>
</tr>
<tr>
<td>68</td>
<td>MLOC_15619</td>
<td>26</td>
<td>37</td>
<td>20</td>
<td>22</td>
<td>42</td>
<td>17</td>
<td>135</td>
</tr>
<tr>
<td>69</td>
<td>MLOC_20723</td>
<td>19</td>
<td>24</td>
<td>27</td>
<td>39</td>
<td>32</td>
<td>16</td>
<td>126</td>
</tr>
<tr>
<td>70</td>
<td>MLOC_51368</td>
<td>13</td>
<td>10</td>
<td>4</td>
<td>9</td>
<td>7</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>71</td>
<td>MLOC_12571</td>
<td>3</td>
<td>0.6</td>
<td>5</td>
<td>47</td>
<td>6</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>72</td>
<td>MLOC_63574</td>
<td>1</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>8</td>
<td>4</td>
<td>31</td>
</tr>
<tr>
<td>73</td>
<td>MLOC_10974</td>
<td>3</td>
<td>3</td>
<td>7</td>
<td>2</td>
<td>1</td>
<td>17</td>
<td>1</td>
</tr>
</tbody>
</table>

HvCslF6  MLOC_57200  10  18  25  42  64  77  77  78
The six contigs acquired from βglux sequencing and assembly are shown here in FASTA format.

>HF_contig_1_@7/20/2016,_9_14_AM
----------------------ATGCTCGAGCGGCCGCCAGT-GTGATGG
ATA-TCTGC-AGAATTCAGGTCGCTCGTCTTCTTCTTACCCCATGGCGGCCG
GGCGCAAGGCGAGACGGCTGCGCTGTACAAAGGCAGCCTGCAGGGAGAAGGGCA
GGCGCGCGTGCGCGACCTGCTGGGCCGGATGACGCTGCGGGAGAAGGGCGG
GCCAGATGGCCCAGATCGAGCGCACCGTCGTGTCGCCCCGCGCCCTCACA
GAGCTCGCCCGCCGCAGCGCATCCTCTCTGCTCAACCCGCCCCGGCCCTACA
GCCGCGCTCCCCGGCGACGACTGCGGCCCAGTAGGTGACAGCAGCAGCGGCC
TCGCCCTCTCCTCTCGGCGCTGCGCTGGCCAATCTCTTGACGCCAGCGCC
GTCCACGGGACAAACACGTCTTCGGCGCCACCGCTTTCCTTTCTCAAACTGT
TGGGCTCGGAGGCTCCTCCAGGTATCCAATCCTGACGCCCCATATGAGCCTCA
TTGACCACGTCGACGCACTTGTACCGGTACCCGCTACTCGGACGGGAGCG
GAGCTTGTCGTTAAGATTGCGAAGGTGACGCCGCTGCTGCTCCGACGCCAC
CGGCATACACTGGGCTTGGCGCCCTGCTGCTGGCCATACCATTCCAA
AAGGCTTCGAGATTAACTGCCGGAATTATTTTGGGGTCTTGGGTGATGGA
G-TTGGCTCTGTCTGGCAGGTTATGCTCAAGGGGTAGGGTAGGATCCAGGG
GGGAGAGTGTTACAGAGGTCTACAGCGAGGCAGGACAGATGCTGCGCTCAG
TCACCAACATTGTACCGGCTTGGAGCAGCGACAGCGAGATCCACCTGAC
CATGTTACCGGCTTCTCTGCTTCTGCTGTAGGTAAATTCCACATGGTGAGAT
AGCTGTGACTTCTCAGACGCCAATCCTGTGCTGGTTATTGAAATGCTCTCTAT
GTGATTTACAGAGGAGATTGCTGCTTGCTGCTGCTGCTGCTGCTGCTGCTG
GGGAGATGTGTTACAGAGGTCTACAGCGAGGAGAGGACAGATGCTGCGCTC
TCACCAACATTGTACCGGCTTGGAGCAGCGACAGCGAGATCCACCTGAC
CATGTTACCGGCTTCTCTGCTTCTGCTTCTGCTGTAGGTAAATTCCACATGG
AGCTGTGACTTCTCAGACGCCAATCCTGTGCTGGTTATTGAAATGCTCTCTAT
GTGATTTACAGAGGAGATTGCTGCTTGCTGCTGCTGCTGCTGCTGCTGCTG
GGGAGATGTGTTACAGAGGTCTACAGCGAGGAGAGGACAGATGCTGCGCTC
TCACCAACATTGTACCGGCTTGGAGCAGCGACAGCGAGATCCACCTGAC
CATGTTACCGGCTTCTCTGCTTCTGCTTCTGCTGTAGGTAAATTCCACATGG
AGCTGTGACTTCTCAGACGCCAATCCTGTGCTGGTTATTGAAATGCTCTCTAT
GTGATTTACAGAGGAGATTGCTGCTTGCTGCTGCTGCTGCTGCTGCTGCTG
GGGAGATGTGTTACAGAGGTCTACAGCGAGGAGAGGACAGATGCTGCGCTC
TCACCAACATTGTACCGGCTTGGAGCAGCGACAGCGAGATCCACCTGAC
CATGTTACCGGCTTCTCTGCTTCTGCTTCTGCTGTAGGTAAATTCCACATGG
AGCTGTGACTTCTCAGACGCCAATCCTGTGCTGGTTATTGAAATGCTCTCTAT
GTGATTTACAGAGGAGATTGCTGCTTGCTGCTGCTGCTGCTGCTGCTGCTG
GGGAGATGTGTTACAGAGGTCTACAGCGAGGAGAGGACAGATGCTGCGCTC
TCACCAACATTGTACCGGCTTGGAGCAGCGACAGCGAGATCCACCTGAC
CATGTTACCGGCTTCTCTGCTTCTGCTTCTGCTGTAGGTAAATTCCACATGG
AGCTGTGACTTCTCAGACGCCAATCCTGTGCTGGTTATTGAAATGCTCTCTAT
GTGATTTACAGAGGAGATTGCTGCTTGCTGCTGCTGCTGCTGCTGCTGCTG
GGGAGATGTGTTACAGAGGTCTACAGCGAGGAGAGGACAGATGCTGCGCTC
TCACCAACATTGTACCGGCTTGGAGCAGCGACAGCGAGATCCACCTGAC
CATGTTACCGGCTTCTCTGCTTCTGCTTCTGCTGTAGGTAAATTCCACATGG
AGCTGTGACTTCTCAGACGCCAATCCTGTGCTGGTTATTGAAATGCTCTCTAT
GTGATTTACAGAGGAGATTGCTGCTTGCTGCTGCTGCTGCTGCTGCTGCTG
GGGAGATGTGTTACAGAGGTCTACAGCGAGGAGAGGACAGATGCTGCGCTC
TCACCAACATTGTACCGGCTTGGAGCAGCGACAGCGAGATCCACCTGAC
CATGTTACCGGCTTCTCTGCTTCTGCTTCTGCTGTAGGTAAATTCCACATGG
TTGTGAACATGTCATCATGTGTAAGCTCTTCTTTCAGGGTTTTGTGGTGT
CAGACTGGGAGGGTATTTGACAGGCTCTGCGGAGCCCTCGAGGGTCTGATTAT
TATGAAGTATGTCTTAAGATATATTTCTTCAGAGACAGGTTCTTTGCTCATT
TCTGATCGACCTACAGATTATGATACCTCACAGATTTGAGAAATTCTTGG
AAGATCTTTGTGTTCTTGGTAGACAGGAGACAGTACAAATACACGAGATT
GATGATGCTTGGAGCGGATTTCTCTAAGGTTAAGTCTATCTCGGAGGTT
TGAGCATCCAATTTTCAGATCCGTTCTCTGCTAGACATAGTCCGCTGTAAGG
TAGCATACTATTTTGTGCTATCTATATGTTTCTGCTTTATATATTACATCT
TAATTGTCTCAATATGCGCATATCATATAGAAATAATAGATATCAGACGCTAAATT-
TTCAACCTTTACAGAGATCGATGTCCTACT GGCCGTTGAGCCGCTGAAATGC
TTTGGTAACCTTCTCATAAAAATGGCAAGAATCAGAAAGAAACTGTTCCCGTGAGGCCGTTCGAAAGTC
TG GCCAAAAATGCAAAACAGAATACACTCGTTGCAGGGACACATGCTGTACAGAT
ATTGGGATACGTGCGGTGTTGAGCAGAAATAGCTTGGCAGGACATAGTCCGCTGTAAGG
AAAATAACTCTTTGTGTAAGCAAAAAACTCAACTTTTTTCTTTCACTATTCTT
TTGTTTTGAGGAGTACCTACATTACGGTGTTACTGAATAAAACTGTTCGTC
TCATCTAGAATTCTAGCTACCAAAGTAGTTAGGTTAACAATATGTCTCTTATT
TTAAACTGCTATAGTGCTACAGATGTTGAACTGTTGCTTCAAAGTTGAT
GTTTAAGATGTTTCAATGAGATCGTTCCAAGTATTGTGTTTGGTTGTTGCTGT
TTTAGACTGATGGTGATGGCATTGCTGGTTTGTACTGAAAAGCCAAAAAT
AAATATAAGATACAGATCTTTACAGATCTACTTACATTTGCAAAACAA
TGCAGAGATGCCAACAATGCAGACTAT-GTCACCAATTACTGAACATACACA
ATTTGGGTTAAAATAAAATTAAGCAACATTTTTTTGCTTTCTAGCCGAG
CATCAACCAGTTTTAATTGAGTTTATTTTTATTATAAAAACCTT
TCATCAGGTGAAGAGCAGTATAAAATCTGCAACTTTGCACAATA-TATGCTA
AGAAACAGATATTATTTATGATATCTAAGTCTACACCAAAGGTAACAGCA
TACTAGAGGCATACAAGAATCCGTCGGAGTGGAAAATCGAAGTTGTCAT
GAGCAATGCGCCAACAGGAGCTATCATTTGAAAACTGAGGATTTTCTTTATGC
AGTTGTTTTTGTTGCGAGTCTTCTTTATGCTAATGGTTTGGGAGATAGAA
CTGACCTTTGTATTTCCCTAATGGTTCCGAACCTGATTACTCGGGTTGCT

64
AGTAAAGTCCCTACACTAGTGATTGTCATATCTGGAAGGCCCTTGGTTAT
TGATCCACAAGTACTGGAAAAGATAGATGCTCTAGTTGCTGCGTGGCTAC
CTGGAAGTGAGGGCAACGGAATTACTGATTGCATCTTTGGGGACCATGAT
TTTTGTTGGGACCATGCTTAACGGTTTTAGGTTTTTCTGATCAATTC
TATAAAATATGGAGATGCAAAATATGACCCCTTATTTCTTTTTGGATATG
GGTTGAAATATGGTAAAGTGATAAGGTTCCACCATAACCTGGAATTTC
CAGCACT---CAYKKSGYMKWSWAGCGSYKGAAYCSAGC-TMSRYAY
SGMRGCTGYWTRCATAGCTKGATGCCAGCTGCTGATTACGCTTGGATGCAT
AGCT
>SF_contig_1_07/20/2016,_9_14_AM
-----------------------------------------------
--------CTGGAATTCAGGTCGCTCGTTCTTCTTCTTACCCCATKGCAGCCG
GCCCAAGGCGAGGACGTGCGCTGTCATACAGGACCGGCTGCGGCGAGGAGGA
GGCGCGCTGGCGCCGATCTGCTGGGCCGGATGACGCTGCGGGAGAAGGCGG
GCCAGATGGCCACGTGACGCACCGTCTGTCGCCCGCGCCCTACA
GAGCTCGGCCCGCCACGCGTCTCCTCAACGCCGCGCGAAGCGCCCCCGCGA
ACGCGCCCTCCCCGCGCGAAGCTGGGCGCCATGCTGACGACATGCAAGGCG
TCGCCCTCTCCCTCCCGCTTGCCGCTGGGATCCCCTTCTCAAGGCAAGGACGCC
GTCACCGGGCACAACAGTCTTTCGGCGCCACCGTCTCTCCCTCAACAGCT
TGGCCTCGGAGCCTCCAGGTACCCCAATCACCTGACCGCCATATGACGCTCA
TTGACACTGACGTCGACTTGACCTGACCGTACCGCGTTGACCTCCAGGACGCC
GAGCTTGTGCATAAGTGGCGAGGTGAAGGCGCTCAGGGCTCCGGCAGGCACCC
CGCATAACGACTGGGCTTGTGCGCTGCCTCGCCGTGAGTTAATTCTCA
AAGGCTTCCGGGTATTAACCTGGGCTGTACTTCTTGGGGTCTTGTATGTTG
GTTTGCTCTGCTGCTGGGGTTATGTCACAGGTGTTGAGGGATCC--------
-----------------------------------------------
-----------------------------------------------CAGCAGATCACCT
CATGGTTACCGGCTTTCCTGCTGGTGAAGTTAATTCCACTGTTGAGATCT
AGCTGTGACTTTCGTGCAGACGCCACATCCTGTGTATTTATTGAAATGGTCTAT
GTTGATTACCGGGAATGTGCTTGGTTTGCCCAAGCATTGTGAGCGGA

65
TCATCAGGTGAAGCAGTATAAAATCTGCACTTTGCACAATA-TATGCTA
AGAAACAGATATTTATGATATCCTAAGGTAACCAACAGGTACAAGCA
TACTAGAGGCATACAAGAAATCCGTCGGAGTGGAAACTGAAGTTGTGCAT
GAGCAATGCCCAACAGATGCTATCAATGAAACTGGAGGATTTTCTTATGC
AGTGGTGTTGGTTGGCCGAGGGTCTCTTTATGCTGAAATGGTTGGGAGATAGA
CTGACCTTTGTATCTCCCTTCAATGGGTTCGACCTGATTACTCGGGTTGCT
AGTAAAGTCCCTCACTAGTGTAGTTGCATATCTGCTATCTGGAAGGCCCTTGGTTAT
TGATCCACAAGTACTGGGAAGATAGATGCTCCTAGTTGCTGCGTGGCTAC
CTGGAAGTTGAGGCAACGGGAATTACTGATTGCTATCTTTGGGACCATGAT
TTTGGTGGCACATTGCGCTGAACTGGGTGTTTAGGTTTCTGTGATCAATTGCC
TATAAATTGAGGATGCAAAATATGACCCTTTTATTTTTCTGCTTATGATATG
GGTTGAAAATGTTTGAAGTAKATAAAGGGTTACCATACCATCTCTGAATTC
C--------------------------------------------------
--------------------------------------------------
----
>HF_contig_2_@7/20/2016, _9_13_AM
---------AGCTTGGTACCGAGCTCG-G-ATCCACTAGTAACGCGC-G
CGCGAAGGCAGAGGCACGTGGCTGTAACAAGGACGCCCTGGCGCCGGTGA
GCCGCGCTGCGGCGACCTGCTGGGCCGATGAGCGCTCGGGAGAAGGCGG
GCCAGATGGCCCAGATCGAGCGCACCGTCGTGTCGCCCCGCGCCCTACG
GAGCTCGGCCGCCGGACCGCTGCCCTCAACGCGCAGGCGAAGCGCCCGCCCGA
ACGCGCCTCCCCGGCCGACTGGGCCGCGATGCTGACGACATGCAGCGCC
TCGCCCTCTCCTCCGCGGCTCCGTGCCCTATCCCTCTACGCGAAGCCGACGCC
GTCCACGCGGACAAACAACGCTTCTTGGGCCCCACCCTGCTTCTTCCTACAACGT
TGGCCTCGGACCTCCAGTACCCAAACACTGACGCCCAACTATGACGCTCA
TTGACACGTGGAGCCACCTTGGTACCCGCTTGCAGCTCAGCTGAGGACGC
GAGCTTGTGCAGATGTGCGGAGGTAGGTACGGCGCCGCTCGAGGGTCCGCAC
CGCGCATACTGGGGCTTTCGCGGCTCCCTCGGCTGAGTTTAAATTCAA
AAGGCTTCCGGATTA----------------------------------
GGCAGATCACCCCT

CATGGTTACCGTTCTGCTTCTCGGTTAGGTAATTTCCACTTGGTATCT
AGCTGCTTCTCGGCTAACGCGGCCCACTCCTGTGTTTATGGAAATGGTCTAT
GTGATTCAACGGGAAATGTGCTTTGTGCTTGGCAAGCATTTCGATACGCGA
TGGTGGACATGAAAGGGACTTCAATAGGGGAAACACCATATGGCTCGCCAG
AAGATTTGGAAGAGGATCCACAGTTACCTTACCTGAGTTGCTAACTCAA
GGTTGGCAACCGGTACGTGCGCCTCTACTCTCTATGAATGGGAAATGGGAGCCATT
GCATGCCAGCGCCTATTGTGCTACAGATGTTTAAAGGGCAAGCTAGGCT
TCAAGGTTAGTGTGACATGTGCTATTTGTGATATCTGTGATCGAAACTCC
GGATCTGAAATGCTAGATGTGCTATTGCCGCTGAAAAAGGAAATAGTAGTCTTAGAG
TTGCAGACATGTCATCATGTGTAAGCTCTTCTCTCTCTCTGCTCTTTTGTGTGT
CAGATGGAAGGCTATTGACACGCCTCCTGAGGCTGCAGTCTGAGATTTAT
CGCTACTGCAATCGCACAATCGTTAATGCTGGAATGGATATGTATTAGT
TATTAAGTGATCTTTAAGTTTATTCTACAGCGACGCTTCTCTTCATTTT
GCTGATCAATCTACAGATATTAGTCTCTCASAGATTTGGAGAAATTCTTGG
AAGATCTTGGTGTCTTTGCTGAGACAGAGATACCAATATCAGGATT
GATGATGCTTGGAGCGGATTTCTAAGGGTTAAGTTTCAATNTCTGGAGGT
TGAGCATCCATTCTTGCTGGAGACATGACATGTCGCTGTGAAGG
TAGCATAACATTTGCTGTTTCTATATGTGTGTTCTGGCGTTAATATATTACCATCT
TAATTGTCCAATTTGCACTATGTCATACATATACGATTACGCTGATATTG
TTCAACTTTTGCAGAGCTCGGCTGCTGCTGACGCTGAGCTGTGAAGG
TTGGTACTCTGAAATGACAGGAAATCGGAGCTCCTTTCCAT
TGCCAAAAAATGCAAAACGATACTTTTGGACAGGACACATGCTGAGCAT
ATTGGATAATCGCTTGCGGTTGGGACAAATAGCTTTTGTGACTTGCAGGACAGTG
AAAGATAACTTTTGTAAGCAAAAAACTCGACTTTTTCTTCATATTACTTT
TATTTTGAGGAGGATCTCTGACATTAGCTTGGACTGAAATAACTGTGTGC
TCATCTGAAATCTGTGTTTACAGAGTTAGGGATAAACATATGTGCTCTATT
TTAACAGATCATGCTACAGATGTTGGAACATATTGTGCTCAAGGATTTGAT
GTCCACGGGCACAACAACGTCTTCGGGCACCACGTCTTCCCTCACAAACGT
TGGCCTCGGACCTCCCAGGTACCACAAACACCTGACGCCCATATGACGCTCA
TTGACCACTGACGGTGCTACGCGGTAGCTGCCTGGAGGTACCCAAACACCT
GAGCTTTGCAGTTAAGATTGGCGAGGATTGACGCCCTGCGAGGCTCAGGGCC
CGGCATACTGGGGCCTTTGCGCCCTGCCTCGCCGTGAGTTTAATTCGA
AAGGCTTCCGGATTAACCTGGCCTGCTCTATTTTGCGGTCTTGGGTTGAG
G-TTTTGCTCTGTCTGGGGGTATGTTATGTCCAGGTTTTGTAGGGATTCGG
GGGAGATGTTACGAGAGCTACAGCGAGGACCCAGAGATCGTGCGCTCGT
TCACCAACATTGTCACCGGCTTGGCAGGCCAGCCACCGGCAGATCCACCT
CATGTTACCCGGTCTTCTGGCTTCCTAGGTAGTAAATCCACTGTGGAGATCT
AGCTGTGACTTCTTCTGCAGACGCCACCTACATCTGGAAATGGTTCTAT
GTGATTTCAGGGGAGATTGCTTTGGCTTTGCTGCCAAGGCATTTCGTAGCG
TTGGTGCACTGACAAGGACTCAATGGAAAGGAACCATATGTCCTGCAGG
AAGATTTGGAAAGGATCCACATGAAAACCTACCCTGATTTGTATAACTCA
GGGGTGCAACCGTCACTGGCGGTTCTATCCTACTCTAAGTGGAATGGGGAG
GCATGCCAGCCGCTATTTGCTCACAGATGTTTTAAAGGGCAAGCTAGGCT
TCAGACGTGAGTTAAGGCTAGCTACGTGAGGGTTCTTGTGGGAGAGGATC
CGCTACTGCAATTCGACATCTACATCTGCTGCTCGGCTACAGATTATGGA
TATTTAAGTGTTACATGCTAGGCCTAGAGACGAAACATGAAATGAGGAT
TTGGCAACATGTCATCGATGTGAAGCTCTTCTCCCCAGGGTTTGTGGGT
CAGACTGGGAGGTTATTGACAGGCTCTGTGAGCCTGAGGCTTGCTGATT
CGCTACTGCAATTCGACATCTACATCTGCTGCTCGGCTACAGATTATGGA
TATTTAAGTGTTACATGCTAGGCCTAGAGACGAAACATGAAATGAGGAT
TTGGCAACATGTCATCGATGTGAAGCTCTTCTCCCCAGGGTTTGTGGGT
CAGACTGGGAGGTTATTGACAGGCTCTGTGAGCCTGAGGCTTGCTGATT
CGCTACTGCAATTCGACATCTACATCTGCTGCTCGGCTACAGATTATGGA
TATTTAAGTGTTACATGCTAGGCCTAGAGACGAAACATGAAATGAGGAT
TTGGCAACATGTCATCGATGTGAAGCTCTTCTCCCCAGGGTTTGTGGGT
TGGCCAAAATGCAAACGAATACTTGGTGCAGGCACACATGCTGACGAT
ATTGGATATCTCTTGGTAAGCAAAAAACTCGACTTTCTTTTACATATTCTTT
TTT--TTTGAGGAGATCTTACCATTAGCTTGGGTACTGAATAAACCTGTTGCT
TCATCTAGAATTTCTAGCTACAAAGTAGTTAGGGATAAACAATTGCTCCTATT
TTAAACTGCTATGATGCGATACAGGATATTGCTCTCCAAAGTTGAT
GTAAAAATGTGTCAATGGATTGTTCAAGTATTATGTTGTGGTTGTCTGTT
TTTAGACTGATGGTGAGGACTGCGCTGTGGTTCTTGGTTACTGAAAGCCTTT
AAAATATAACATACACTGGATAGTAACTGAGTCTACTATCAACAAAGAGG
TGCAATGCCCAACTGCGACTATCAGCAACATTACCTGAACATCTGAAAATAAA
ATGTTGGTGAIAAAATAATTAAGCAGCATCTTCTTTCTTTGTTCTTTAAGCCAGA
CATCACACTTTAATTTAGATTTTTTTTGGTTCTATTATATATATAAAGGA
TCATCAGGTAAGACGATATAAAATCTGTACACTCAACAAATATTTGCTA
AGAAACAGATATTATTGTATACCATCAAAGTCAAACAGGTACAGGCA
TACTGAGGAGGCTACAAGAATCCGTCGGATGGAACACTGAAGTTGTGCAT
GAGCAATGCCCAACAGATGCTATCATATTGAAACTGAGGATTTTCTTATTCG
AGTTGGTGTGTTGGCCAGGTCTTTCTTTATGCTGATGTTGGGAGATAGAA
CTGACCTTTAGTATTTCCCTTTAAATGTTTCACGATTAACCCGCTTTGCT
AGTAAAGTCCCTTACACTAGTGATTTGTGCATATCTGGAAGGCCCTTGGTTAT
TGATCAACAAAGTAATGGAAAAAGATAGATGCTCTTAGTGCTGCGTGCTAC
CTGGAAGTGGGACCCGGAAATTACGTAGTGGCATCTTTGGGGGACCATTGAT
TTTGGTTGGCAGGACTTGCGGCTCGCTTCTTACCCCATGGCGGCG

>HF_contig_3 @ 7/20/2016, 9:16 AM
-------------GCATGCT--CGAG--CG--CGCGCAGT-GTGATGG
ATA-TCTGC-AGAATCCNGTCGCTGCTTCTTCTTACCACATGGCGGCG
GCGCAAGGCGAGGACGTGCCGCTGTACAAGGACGCGTCGGCGCCGGTGGA
GGCGCGCTGCGCGACCTGCTGGGCCGGATGACGCTGCGGGAGAAGGCGG
GCCAGATGGCCCAGATCGAGCGCACCGTCGTGTCGCCCCGCGCCCTACG
GAGCTCGCCGCGCCAGCGCTCTCCTCAACGCCGGCCCGGAAGCGCCCCCCGCGA
ACGCGCCTCCCCGGCCGACTGGGCCCCGACATGGTCGACGACATGACGCGCC
TCGACCTCTCTCTCCCCCTGCGTCCCCATCTCTTACGCGACGAGGACGCC
GTCCACGACAGACCAACACGTCTTTGGCCGGACCGGACCGGACCGGACG
GGCGCATACACTGGGCTTTGCGCCCTGCCTCGCCGTGAGTTTAAATTCAA
AAGGCTTCCCGGATTTAACCTGGCGCGTGTCTATTTTTGGGTCTTGTAGTTGT
GTTTTGCTCTGTCTGGCCGGGTTATGTCCAGGTT------------------
--------------------------------------------------
-------ACATTGTCACCGGCCTTGCGAGGCGGAGGACGACATCACCCT
CATGGTTACCCGGCTTCTTGCAGCCGCTTGCGAGGACGACATCACCCTT
AGCTGTGACTTCTCGAGGACAGGCCCACACTTCCTGTTTATTGAAATGGTTCTAT
GTGATTTCACAGGGAAATGTGCTTTGCTTTGGCCAAGCAATTTCGTAGGGA
TGGTGGCAGAGCAGAGGACTCAATGAGGGGCAACCATATGCTCGCCAG
AAGATTTGGAAAGGATCCACATGAAACCCTACCTTGATTGTATTAACCTCA
GGGTGCGACGGCTATGCGGCTCTTACTCTAAGTGGAAATGGGGAGCATT
GCATGCCAGGCCTATTTGCTCACAGATGGTTTAAAGGGCAAGCTAGGCT
TCAAGGCTATGACATGCTTTTATTGATATCTTGATGCAAACTTCC
GGATCTGAATAGTGACTATGGCCTGTCAAAAGAAGAATAGTGACTATGAG
TTGCGAACATGTCATCATGCTCTTGAGCTTGTCTTCTTTGCCAGGCTTGGGG
CAGACTGGGAGGGTATTGACAGGCTCTGTGAGCCTCGAGGGTCGTATTTT
CGCTACTGCAATTGCAATAAGTTAATGCTGGAATGATGGATATTAGT
TATTAAGTGTATTAAATTATTTACTACAGCGACGGGTTCTCTTCATTTT
GCTGATCAATCTACTAGATTATGATTCCTCACAGATTTGAGAAATTCTTG
AAGATCTTTGTTCTTTGGGAGACAGGAGAGATACCACATATCAGGATT
GATGATGCTGGATGCGGATTCTAAATGGATATCTGGGAGGTT
TGAGCATTCCATTTTCAGATCCGTCTGCTGCTAGACATAGTCGGCTGTAAGG
TAGCATACATTTTGCAATCATCTATATGTATTGCTGCTAATTACATCT
TAATTGTCCAAATTCGTCATACAGTGATCATATACGATTACGCTAAATTG
TTCAACTTTTCGAGGACATCGGCTGGCTGGCCGCTGAGGGCGGCTTTTGAAAGTC
TTTGGTACCTTCTGGAAAAATGGCAAGAAATCAAGAAAGCTTCTTCCAT
TGCCAAAAATGTGAAAACGAATAACTTTGTTGCAGGGACACATGCTGACGAT
ATTGGATATCAGTGCGGTGGGTGGACAATAGCTTGGCATGGAGACAGTGG
AAAGATAACTCTTTGTAAGCAAAAACTCGACTTTTCTTTTCACATTACTTT
TCTCTAGATATTCTAGCTAAGCCTACATTACGTCTACATTCAAAACAG
TGCGAGATGCCAAATGCGGAGCTATCATCACGCAATTACTGAACTAAAAAT
ATGTTGGTGAATATTTAAATTAACGACAATTTCCCTTTTCTTTTTACAGCCAGA
CATCACCACCTTTTAAATTAGATTTTTTCTATTTATATATAAAAAAGGT
TCATCGTGAAGCGATATTTAAATACGACTCTCAACAAATTTGCTA
AGAAACAGATATTTGATATCTAATAACGTTCAACACACAGTACAGCA
TACGAGGAGGCAATACAAAGTAATCCTACGCAGGATTGCTTATGCTGGCAT
GAGCAGAACAAATGAGATCCTATACATTGGAAACTGGAGGATTTTCTTGCA
AGTTTGTTGGTTGGCAGGTTCCCTTTATGGGTATGGAAGTGTGAG
CTGCACCTGGTAATTTCTCTTTTAAATGTTCCGACCTGGTATCTGCTGGT
AGTAAAGTCCCTACACTAGTGTATCTATCTCGAAGGCTCTTGGTTAT
TGATCAACAAGTTGGAAGGATGAGATGCTCTGCTTAGTGGCTCTCCCTC
CTGGAAGGTCAGGGCAAGGATTACTGATTGCTATCTTTGGGACCATGAT
TTTGGTGTCAGGCTTCTGTAAGATGTTTGACTTCTGCTGGGCTATAC
TTATGATGTCAGGCAATGCGGAGCTCTGCTGAGGCTTCTTCTGGGAGGATAG
TATAAAATTGGGAGATGGCCAAATATGCAACGCTTTTCTTTTGGGATATG
GGTTGAAAAATGTTTAGAAGGTGATAAAGGTTCTCACCATAACCTCTGAAATT
----------CAGCACACTGGCGGCCGTTACTAGTGG-ATC-CGAGCTCG
G-T-ACCA-AGCTTGATGCATAGC--------------------------

>SF_contig_3_07/20/2016,_9_14_AM
GCTATGCATCAAGCTTGGTACCGAGCTCG-G-ATCCACTAGTAACGGCCG
CCAGTGTGCTGGAATTCAGGTCGCTGCTTCTTCTTCTTACCCCATGGCGGCG
GCGCAAGGCGAGGACGTGCCGCTGTACAAGGACGCGTCGGCGCCGGTGGA
GGCGCGCGTGCAGGCGACCTGCTGGGCCGGATGACGCTGCGGGAGAAGGCGG
GCCAGATGGCCCAGATCGAGCGCACCGTCGTGTCGCCCCGCGCCCTCACG
GAGCTCGCCGCAGCAGCCGCAGCGCTCTCAACAGCGCCCGCGGAAAGCGCCGCCCCCGCA
ACGCGCCTCCCCGCGGCGACTGGGGCGGCGATGTTCGACGACATGACGCGCC
TCGCCCTCTCTCTCCTCGCCGCTGGCTCCCATCTCTACTGCGCCCGAGGCACCC
GTCCACGGGCAACACACGTCTTCTGGGCGCAGCCACTCTTTCTACACACACGT
TGCCCTCGGAGCCTCCAGGTACCAATCACCTGACGAGCGCCTCAACGCTCA
TGAGCTCGTCGACGACTTGGACGTACCGTACGCTGCTGTGCCCTTCTGGAC
GATGTGCTTGTTGAGGTGAGGGTCGCTGCTCGAGGTCCGCGCCACGCGGCATACACTGGGCTTTGGCCTGCTGCCTGGCGAGTTTAAATTCAA
AAGGCTTCCCGGATTAACCTGGCGGCTGCTATTTTGGGCTCTTGGATTGTCT
G-TTGTCTCTGCTCTGGGCTTTATATGTACGCTTAGGGATCGCAGGTG
GGGAGATGTTACGAGAGCTACAGCGAGGACCAGAGATCGTGCGCTCGT
TCACCAACATTGTCACCGGGCTTGCGAGGCGCACCACRGCAGATCACCTC
CATGGTACCCGCTCTTCTGCTGTTAGGTAATTTCCACTGGTGAGATCT
AGCTGTGACCTTCTGCAAGGACCCCCACATCTGTGTTATTGAATGTTCTAT
GTGATTTTACGAGGAAATGTTGCTTTTGGTGAACCGATTTGCAGGAG
TGGTGCCACTGACAGGGACTCAATGAGGGGAACACCATATGCTGCGGCA
AAGATTTGGAAAGGATCCACATGAAACTCTACCCTGATTTGTAAACTCAA
GGGTTGGCAAGCGTGCTACCGCTCTACTCTAAATGGAATGGGGAGCCATT
GCATGCCAGCCGCTATTGCTTACGGATGTTTTAAAGGGCAAGCTAGGCT
TCAAGGTAGTAGAGCATGTGCTATTTGATTGATATCTTGATGCAAACTTC
GGGTTTGCAATAGTGACTATGGCGCTTCTTGGTCCTTGCTAAAGAATAGTGACTATGRG
TTGCGAACATGTTCATCATGTGTAAGCTCTTCTTTTCAGGGTTTTGTGGTGT
CAGACTGGGAGGTTATTTGACAGGCTCCTGAGCCTCAGGGGCTGTGATTAT
CGCTACTGCATTTGCAACATCAGTTTAATGCTGGAATGGATATGGTATTAGT
TATTAAGTAGTCCTTAAGTTATTATCATACAGCCACRGGTCTCTCTGCATT
GCTGATCAATCTCACAGATTATGATTCTCTCACAGATTTTGAGAAATTCTTGG
AGAGCATCCATTTTCAGATCCGTCTCTGCTAGACATAGGCTGTAAGG
TAGCATACATTTCAGATCATATCTATATGTTGCTGCCAATATATGCTACTT
TAATTGTCCAAAATTGCAATACATATGCTATACATTACGTGATCATATACGATTACGCTAAAATTG
TTCAACTTTTCCAGGAGCTCGGCTGCTGCGCCCTGGAGCCGCTGTCAGAAGTCT
TTTAGTACTCCTACTGGTAAGGCCAACACTGCACTTTTTTTCTCATTACTTT
TTT--TTGAGGAGTACTTACATTAGCTTGGTGACTGAATAAACCTGTGTC
TCATCTAGAATTCTAGCTACCAAGTAGATTAGGGATAATAAGCTAATATGCTCTATT
TTAAACTGCTATGACAGCAGTGGTGAACATATTTGCTCCAAAGTGTGAT
GTTAAAAATGGTTCAATGGAATTGTTCAAGTATTATGTTGTTGTGTCTGTT
TTTAGACTGATGGTGAAGGCACGTGCTGGTTGTGACTGAAAAAGCCAAAAT
AAAATAAATAACATACTGGAATAGTAACATTACGAGCTCATATTCAAAAACAAG
TGCAGAGATGCCAACAATGCGAGCTATCATCACCAATTACTGAACTAAAAA
ATGTTGGTGAAAAATAATTAAGCAGCATATTCTCATTTCATTTCTATTGAAGCCAGA
CATCACACTTTTTATTATTAACTTTTTTATTATTAAAAAGGTT
TCATCAGGTAGAAGCAGTATAAAAAACTCGACTCTCCTTCTGGAAGGGAATTTTTC
AGAAAACAGATATTATTGATATCTAAGTTCTACACCAACAGGTAACAAGCA
TACTGGAGGCAATACAAGATCCGGGAGTGGGAAAAGCGAAGTTGGTGTGACAT
GAGCAATGCGCAACAGATGCTACTATGAAAACGTGAGATTATTGTGTTATGC
AGTTGTTGTTGTCGGGAGGTTCTTATGCTGAATGGTTGGGAGATAGAA
CTGACCTTAGATTTCTTTTATTGTTGTCGACCTGATTACTCGGGTGTGCT

75
AGTAAAGTCCCTACACTAGTGATTGTCATATCTGGAAGGCCCTTGGTTAT
TGATCCACAAGTACTGGAAAAGATAGATGCTCTAGTTGCTGCGTGGCTAC
CTGGAAGTGGAGGGACGGAATTACTGATTGCATCTTTGGGACCATGAT
TTTGTGGCACATTGCCTGTAACGTTGTTTAGGGTTTCTGATCAATTGCC
TATAAATATTTGGAGATGCCAAATATGACCCCTTTATTTCTTTTGTGATATG
GGTTGAAAATGTTTAGAAGTGATAAACGGTCCCTCACCATAACCTCTGAAATTC
TGCAGATATCCCATCACACTGGCCGCCT-CGACATGCATCT-------
--------------------------------------------------
----