



Theses and Dissertations

---

2016-12-01

## An Investigation of $\beta$ glux, a Glucosidase Co-Expressed with Cslf6 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  $\beta$ glux, a Glucosidase Co-Expressed with Cslf6 in Oat (*Avena sativa*) and Barley (*Hordeum vulgare*)" (2016). *Theses and Dissertations*. 6595.  
<https://scholarsarchive.byu.edu/etd/6595>

This Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact [scholarsarchive@byu.edu](mailto:scholarsarchive@byu.edu), [ellen\\_amatangelo@byu.edu](mailto:ellen_amatangelo@byu.edu).

An Investigation of  $\beta$ gluc, 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 $\beta$ glux, 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)- $\beta$ -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,  $\beta$ glux, with CslF6.  $\beta$ glux 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,  $\beta$  (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 Campbell, 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

TITLE PAGE .....	i
ABSTRACT .....	ii
ACKNOWLEDGMENTS .....	iii
TABLE OF CONTENTS.....	iv
LIST OF TABLES .....	vii
LIST OF FIGURES .....	viii
CHAPTER 1 : (1,3;1,4) $\beta$ -D-GLUCAN LITERATURE REVIEW MIXED LINKAGE BETA (1,3;1,4) D-GLUCAN (MLG) AS A POLYSACCHARIDE .....	1
GENETIC ORIGIN .....	2
LOCALIZATION OF PROTEIN AND PRODUCT.....	4
GLUCOSIDASE HYPOTHESIS .....	5
COMMERCIAL IMPORTANCE .....	6
CONCLUSION.....	7
LITERATURE CITED.....	8
CHAPTER 2 : IDENTIFICATION AND VALIDATION OF SUPERIOR REFERENCE GENES FOR BARLEY USING IN SILICO ANALYSIS.....	14
ABSTRACT.....	14
INTRODUCTION .....	15

MATERIALS AND METHODS.....	17
RESULTS .....	20
DISCUSSION.....	22
LITERATURE CITED .....	25
TABLES .....	28
FIGURES.....	30
APPENDIX 2-1 .....	33
APPENDIX 2-2 .....	34
APPENDIX 2-3 .....	35
CHAPTER 3 : $\beta$ GLUX IS A GLUCOSIDASE CO-EXPRESSED WITH CSLF6 THROUGH OAT AND BARLEY CARYOPSIS STAGES.....	36
ABSTRACT.....	36
INTRODUCTION .....	36
MATERIALS AND METHODS.....	37
RESULTS .....	41
DISCUSSION.....	43
CONCLUSION.....	48
LITERATURE CITED .....	49
TABLES .....	52
FIGURES.....	54

APPENDIX 3-1 .....	60
APPENDIX 3-2 .....	63

## 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 $\beta$ gluc 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

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..... 31

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. .... 32

Figure 3-1: Glucosidase transcript levels in developing inflorescence 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..... 54

Figure 3-2: Glucosidase transcript levels in developing inflorescence 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..... 54

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 CsIF6 and  $\beta$ glux. 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  $\beta$ glux. Green squares represent CsIF6 ..... 59

## CHAPTER 1 : (1,3;1,4) $\beta$ -D-GLUCAN LITERATURE REVIEW MIXED LINKAGE BETA (1,3;1,4) D-GLUCAN (MLG) AS A POLYSACCHARIDE

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)  $\beta$  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  $\beta$  linked backbones such as xyloglucans, xylans, mannans and glucomannans, and (1-3,1-4)  $\beta$  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 (cellotetrasy) 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 *CsIF6* by testing null mutations in barley for MLG presence. A mutation in a conserved region of the *HvCsIF6* gene caused a betaglucanless (*bgl*) phenotype. In the same study, the native *HvCsIF6* 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:cellotetrasy) ratio across plant species. Transgenic tobacco plants expressing chimeric variants of CslF6 revealed a single isoleucine or leucine at the 757<sup>th</sup> 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 cellotetrasy 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 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 Cs1F6.

## LITERATURE CITED

- Brennan, C.S. and Cleary, L.J., 2005. The potential use of cereal (1→3, 1→4)-β-D-glucans as functional food ingredients. *Journal of cereal science*, 42(1), pp.1-13.
- Buckeridge, M.S., Rayon, C., Urbanowicz, B., Tiné, M.A.S. and Carpita, N.C., 2004. Mixed linkage (1→3),(1→4)-β-D-glucans of grasses. *Cereal Chemistry*, 81(1), pp.115-127.
- Burton, R.A. and Fincher, G.B., 2009. (1, 3; 1, 4)-β-d-glucans in cell walls of the Poaceae, lower plants, and fungi: a tale of two linkages. *Molecular Plant*, 2(5), pp.873-882.
- Burton, R. and Fincher, G., 2012. Current challenges in cell wall biology in the cereals and grasses. *Frontiers in plant science*, 3, p.130.
- Burton, R.A., Wilson, S.M., Hrmova, M., Harvey, A.J., Shirley, N.J., Medhurst, A., Stone, B.A., Newbigin, E.J., Bacic, A. and Fincher, G.B., 2006. Cellulose synthase-like CslF genes mediate the synthesis of cell wall (1, 3; 1, 4)-β-D-glucans. *Science*, 311(5769), pp.1940-1942.
- Campbell, J.A., Davies, G.J. and Bulone, V., 1998. A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. *Biochemical Journal*, 329(Pt 3), p.719.
- Carpita, N.C. and McCann, M.C., 2010. The maize mixed-linkage (1→3),(1→4)-β-d-glucan polysaccharide is synthesized at the Golgi membrane. *Plant physiology*, 153(3), pp.1362-1371.
- Cavallero, A., Empilli, S., Brighenti, F. and Stanca, A.M., 2002. High (1→3, 1→4)-β-glucan barley fractions in bread making and their effects on human glyceimic response. *Journal of Cereal Science*, 36(1), pp.59-66.

- Cocuron, J.C., Lerouxel, O., Drakakaki, G., Alonso, A.P., Liepman, A.H., Keegstra, K., Raikhel, N. and Wilkerson, C.G., 2007. A gene from the cellulose synthase-like C family encodes a  $\beta$ -1, 4 glucan synthase. *Proceedings of the National Academy of Sciences*, 104(20), pp.8550-8555.
- Cosgrove, D.J., 2005. Growth of the plant cell wall. *Nature reviews molecular cell biology*, 6(11), pp.850-861.
- Cutler, S. and Somerville, C., 1997. Cellulose synthesis: cloning in silico. *Current Biology*, 7(2), pp.R108-R111. Delmer, Deborah P. "Cellulose biosynthesis: exciting times for a difficult field of study." *Annual review of plant biology* 50, no. 1 (1999): 245-276.
- Doblin, M.S., Pettolino, F.A., Wilson, S.M., Campbell, R., Burton, R.A., Fincher, G.B., Newbigin, E. and Bacic, A., 2009. A barley cellulose synthase-like CSLH gene mediates (1, 3; 1, 4)- $\beta$ -D-glucan synthesis in transgenic Arabidopsis. *Proceedings of the National Academy of Sciences*, 106(14), pp.5996-6001. Domon, Jean-Marc, Laëtitia Baldwin, Sébastien Acket, Elodie Caudeville, Stéphanie Arnoult, Hélène Zub, Françoise Gillet, Isabelle Lejeune-Hénaut, Maryse Brancourt-Hulmel, Jérôme Pelloux, Catherine Rayon. "Cell wall compositional modifications of Miscanthus ecotypes in response to cold acclimation." *Phytochemistry* 85 (2013): 51-61.
- Farrokhi, N., Burton, R.A., Brownfield, L., Hrmova, M., Wilson, S.M., Bacic, A. and Fincher, G.B., 2006. Plant cell wall biosynthesis: genetic, biochemical and functional genomics approaches to the identification of key genes. *Plant biotechnology journal*, 4(2), pp.145-167.

- FRY, S.C., 1989. The structure and functions of xyloglucan. *Journal of Experimental Botany*, 40(1), pp.1-11.
- Gibeaut, D.M. and Carpita, N.C., 1993. Synthesis of (1→3),(1→4)-beta-D-glucan in the Golgi apparatus of maize coleoptiles. *Proceedings of the National Academy of Sciences*, 90(9), pp.3850-3854.
- Gille, S. and Pauly, M., 2007. O-acetylation of plant cell wall polysaccharides. *Current challenges in plant cell walls*, p.9.
- Han, F., Ullrich, S.E., Chirat, S., Menteur, S., Jestin, L., Sarrafi, A., Hayes, P.M., Jones, B.L., Blake, T.K., Wesenberg, D.M. and Kleinhofs, A., 1995. Mapping of β-glucan content and β-glucanase activity loci in barley grain and malt. *Theoretical and Applied Genetics*, 91(6-7), pp.921-927.
- Hazen, S.P., Scott-Craig, J.S. and Walton, J.D., 2002. Cellulose synthase-like genes of rice. *Plant physiology*, 128(2), pp.336-340.
- Jobling, S.A., 2015. Membrane pore architecture of the CslF6 protein controls (1-3, 1-4)-β-glucan structure. *Science advances*, 1(5), p.e1500069.
- Jørgensen, K.G. and Aastrup, S., 1988. Quantification of high molecular weight (1→3)(1→4)-β-d-glucan using Calcofluor complex formation and flow injection analysis. II. determination of total β-glucan content of barley and malt. *Carlsberg Research Communications*, 53(5), pp.287-296.
- Lee, Homin K., Amy K. Hsu, Jon Sajdak, Jie Qin, and Paul Pavlidis. "Co-expression analysis of human genes across many microarray data sets." *Genome research* 14, no. 6 (2004): 1085-1094.

- Lia, Angstrom, et al. "Oat beta-glucan increases bile acid excretion and a fiber-rich barley fraction increases cholesterol excretion in ileostomy subjects." *The American journal of clinical nutrition* 62.6 (1995): 1245-1251.
- Liepman, A.H., Wilkerson, C.G. and Keegstra, K., 2005. Expression of cellulose synthase-like (Csl) genes in insect cells reveals that CslA family members encode mannan synthases. *Proceedings of the National Academy of Sciences of the United States of America*, 102(6), pp.2221-2226.
- McClellan, B.V. and Glennie - Holmes, M., 1985. ENZYMIC QUANTIFICATION OF (1→3)(1→4)-β-D-GLUCAN IN BARLEY AND MALT. *Journal of the Institute of Brewing*, 91(5), pp.285-295.
- Meier, H. and Reid, J.S.G., 1982. Reserve polysaccharides other than starch in higher plants. In *Plant carbohydrates I* (pp. 418-471). Springer Berlin Heidelberg.
- Nemeth, C., Freeman, J., Jones, H.D., Sparks, C., Pellny, T.K., Wilkinson, M.D., Dunwell, J., Andersson, A.A., Åman, P., Guillon, F. and Saulnier, L., 2010. Down-regulation of the CSLF6 gene results in decreased (1, 3; 1, 4)-β-D-glucan in endosperm of wheat. *Plant physiology*, 152(3), pp.1209-1218.
- Oliver, S., 2000. Proteomics: guilt-by-association goes global. *Nature*, 403(6770), pp.601-603.
- Pear, J.R., Kawagoe, Y., Schreckengost, W.E., Delmer, D.P. and Stalker, D.M., 1996. Higher plants contain homologs of the bacterial celA genes encoding the catalytic subunit of cellulose synthase. *Proceedings of the National Academy of Sciences*, 93(22), pp.12637-12642.

- Pellny, T.K., Lovegrove, A., Freeman, J., Tosi, P., Love, C.G., Knox, J.P., Shewry, P.R. and Mitchell, R.A., 2012. Cell walls of developing wheat starchy endosperm: comparison of composition and RNA-Seq transcriptome. *Plant physiology*, 158(2), pp.612-627.
- Philippe, S., Saulnier, L. and Guillon, F., 2006. Arabinoxylan and (1→3),(1→4)-β-glucan deposition in cell walls during wheat endosperm development. *Planta*, 224(2), pp.449-461.
- Poppitt, S.D., Van Drunen, J.D., McGill, A.T., Mulvey, T.B. and Leahy, F.E., 2007. Supplementation of a high-carbohydrate breakfast with barley β-glucan improves postprandial glycaemic response for meals but not beverages. *Asia Pacific journal of clinical nutrition*, 16(1), pp.16-24.
- Richmond, T.A. and Somerville, C.R., 2000. The cellulose synthase superfamily. *Plant physiology*, 124(2), pp.495-498.
- Saxena, I.M. and Brown, R.M., 1995. Identification of a second cellulose synthase gene (acsAII) in *Acetobacter xylinum*. *Journal of bacteriology*, 177(18), pp.5276-5283.
- Simmons, T. J. (2013). Considerations in the search for mixed-linkage (1→3),(1→4)-β-d-glucan-active endotransglycosylases. *Plant signaling & behavior*, 8(4), e23835.
- Taketa, S., Yuo, T., Tonooka, T., Tsumuraya, Y., Inagaki, Y., Haruyama, N., Larroque, O. and Jobling, S.A., 2012. Functional characterization of barley betaglucanless mutants demonstrates a unique role for CslF6 in (1, 3; 1, 4)-β-D-glucan biosynthesis. *Journal of experimental botany*, 63(1), pp.381-392.
- United States Department of Agriculture (March 2015) Grain: World Market and Trade. Retrieved from <http://www.fas.usda.gov/data/grain-world-markets-and-trade>.



- Vega-Sanchez, M., Verhertbruggen, Y., Scheller, H.V. and Ronald, P., 2013. Abundance of mixed linkage glucan in mature tissues and secondary cell walls of grasses. *Plant signaling & behavior*, 8(2), p.e23143.
- Nemeth, C., Freeman, J., Jones, H.D., Sparks, C., Pellny, T.K., Wilkinson, M.D., Dunwell, J., Andersson, A.A., Åman, P., Guillon, F. and Saulnier, L., 2010. Down-regulation of the CSLF6 gene results in decreased (1, 3; 1, 4)- $\beta$ -D-glucan in endosperm of wheat. *Plant physiology*, 152(3), pp.1209-1218.
- Wilson, S.M., Ho, Y.Y., Lampugnani, E.R., Van de Meene, A.M., Bain, M.P., Bacic, A. and Doblin, M.S., 2015. Determining the subcellular location of synthesis and assembly of the cell wall polysaccharide (1, 3; 1, 4)- $\beta$ -d-glucan in grasses. *The Plant Cell*, 27(3), pp.754-771.
- Wolfe, C.J., Kohane, I.S. and Butte, A.J., 2005. Systematic survey reveals general applicability of "guilt-by-association" within gene coexpression networks. *BMC bioinformatics*, 6(1), p.1.
- Yin, Y., Huang, J. and Xu, Y., 2009. The cellulose synthase superfamily in fully sequenced plants and algae. *BMC plant biology*, 9(1), p.1.
- Zhang, G., Chen, J., Wang, J. and Ding, S., 2001. Cultivar and environmental effects on (1 $\rightarrow$  3, 1 $\rightarrow$  4)- $\beta$ -D-glucan and protein content in malting barley. *Journal of Cereal Science*, 34(3), pp.295-301.

## 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  $\beta$ -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, three 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  $\mu$ L 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 *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<sup>th</sup>, 6<sup>th</sup>, 5<sup>th</sup>, 10<sup>th</sup>, and 14<sup>th</sup> *in silico* respectively. The candidate that returned the lowest coefficient of variation, Hv.3082 (predicted as the most stable), placed 15<sup>th</sup> in the comprehensive ranking (13<sup>th</sup> in the Delta CT method, 16<sup>th</sup> in the BestKeeper method, 13<sup>th</sup> in the Normfinder method, and 14<sup>th</sup> in the geNorm method). The candidate that returned the highest coefficient of variation (predicted as the least stable), Hv.12544, resulted as 12<sup>th</sup> in the comprehensive ranking (7<sup>th</sup> under the Delta CT method, 15<sup>th</sup> for BestKeeper, 7<sup>th</sup> for Normfinder, and 11<sup>th</sup> for geNorm) (Appendix 3).

***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<sup>th</sup> of 20 in the original empirical Conlon analysis, ranked 6<sup>th</sup> 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<sup>rd</sup> for Golden Promise and Harrington, 1<sup>st</sup> for HiFi, and 1<sup>st</sup> for the combined barley and the combined barley plus oat analyses. GAPDH ranked 7<sup>th</sup> in Conlon, 7<sup>th</sup> in Golden Promise, 1<sup>st</sup> in Harrington, 6<sup>th</sup> in HiFi, 7<sup>th</sup> for all barley cultivars, and 6<sup>th</sup> across barley and oat. Candidate Hv.3271 which ranked 1<sup>st</sup> 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

- Andersen C.L., Jensen J.L., and Orntoft T.F. 2004. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer research* 64: 5245-5250.
- Ferdous, J., Li, Y., Reid, N., Langridge, P., Shi, B.J. and Tricker, P.J. 2015. Identification of reference genes for quantitative expression analysis of MicroRNAs and mRNAs in barley under various stress conditions. *PloS one* 10: e0118503.
- Hua, W., Zhu, J., Shang, Y., Wang, J., Jia, Q. and Yang, J. 2015. Identification of suitable reference genes for barley gene expression under abiotic stresses and hormonal treatments. *Plant Molecular Biology Reporter* 33: 1002-1012.
- Janská, A., Hodek, J., Svoboda, P., Zámečník, J., Prášil, I.T., Vlasáková, E., Milella, L. and Ovesná, J. 2013. The choice of reference gene set for assessing gene expression in barley (*Hordeum vulgare* L.) under low temperature and drought stress. *Molecular genetics and genomics* 288: 639-649
- Jin, Y.L., Speers, R.A., Paulson, A.T., and Stewart, R.J. 2004. Barley  $\beta$ -glucans and their degradation during malting and brewing. *Technical Quarterly-Master Brewers Association of the Americas* 41: 231-240.
- Jørgensen, K.G. and Aastrup, S. 1988. Quantification of high molecular weight  $(1 \rightarrow 3)(1 \rightarrow 4)$ - $\beta$ -d-glucan using Calcofluor complex formation and flow injection analysis. II. determination of total  $\beta$ -glucan content of barley and malt. *Carlsberg Research Communications* 53: 287-296.

- Lee, H.K., Hsu, A.K., Sajdak, J., Qin, J. and Pavlidis, P. 2004. Co-expression analysis of human genes across many microarray data sets. *Genome research*, 14: 1085-1094.
- Li, Z. and Trick, H.N. 2005. Rapid method for high-quality RNA isolation from seed endosperm containing high levels of starch. *Biotechniques*, 38: 872-876.
- Oliver, S. 2000. Proteomics: guilt-by-association goes global. *Nature* 403: 601-603.
- Oliver, R.E., Tinker, N.A., Lazo, G.R., Chao, S., Jellen, E.N., Carson, M.L., Rines, H.W., Obert, D.E., Lutz, J.D., Shackelford, I., Korol, A.B., Wight, C.P., Gardner, K.M., Hattori, J., Beattie, A.D., Bjornstad, A., Bonman, J.M., Jannink, J.L., Sorrells, M.E., Brown-Guedira, G.L., Mitchell Fetch, J.W., Harrison, S.A., Howarth, C.J., Ibrahim, A., Kolb, F.J., McMullen, M.S., Murphy, J.P., Ohm, H.W., Rossnagel, B.G., Yan, W., Miclaus, K.J., Hiller, J., Maughan, P.J., Redman Hulse, R.R., Anderson, J.M., Islamovic, E. and Jackson, E.W. 2013. SNP discovery and chromosome anchoring provide the first physically-anchored hexaploid oat map and reveal synteny with model species. *PLOS One* 8: e58068.
- Ovesna, J., Kučera, L., Vaculová, K., Štrymlová, K., Svobodova, I. and Milella, L. 2012. Validation of the  $\beta$ -amy1 transcription profiling assay and selection of reference genes suited for a RT-qPCR assay in developing barley caryopsis. *PloS one* 7: e41886.
- Paolacci, A.R., Tanzarella, O.A., Porceddu, E. and Ciaffi, M., 2009. Identification and validation of reference genes for quantitative RT-PCR normalization in wheat. *BMC molecular biology* 10: 1-27.

- Pfaffl, M.W., Tichopad, A., Prgomet, C. and Neuvians, T.P. 2004. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper–Excel-based tool using pair-wise correlations. *Biotechnology letters* 26: 509-515.
- Silver N, Best S, Jiang J, and Thein S.L. 2006. Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. *BMC molecular biology* 7: 1-9.
- United States Food and Drug Administration 2005. Food labeling: Soluble dietary fibre from certain foods and coronary heart disease. *Federal Register* 70: 76150–76162.
- Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A. and Speleman, F. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome biology*, 3: 1-12.
- Wolfe, C.J., Kohane, I.S. and Butte, A.J. 2005. Systematic survey reveals general applicability of "guilt-by-association" within gene co-expression networks. *BMC bioinformatics* 6: 1-10.
- Zmienko, A., Samelak-Czajka, A., Goralski, M., Sobieszczuk-Nowicka, E., Kozlowski, P. and Figlerowicz, M. 2015. Selection of reference genes for qPCR-and ddPCR-based analyses of gene expression in Senescing Barley leaves. *PloS one* 10: e0118226.

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.**

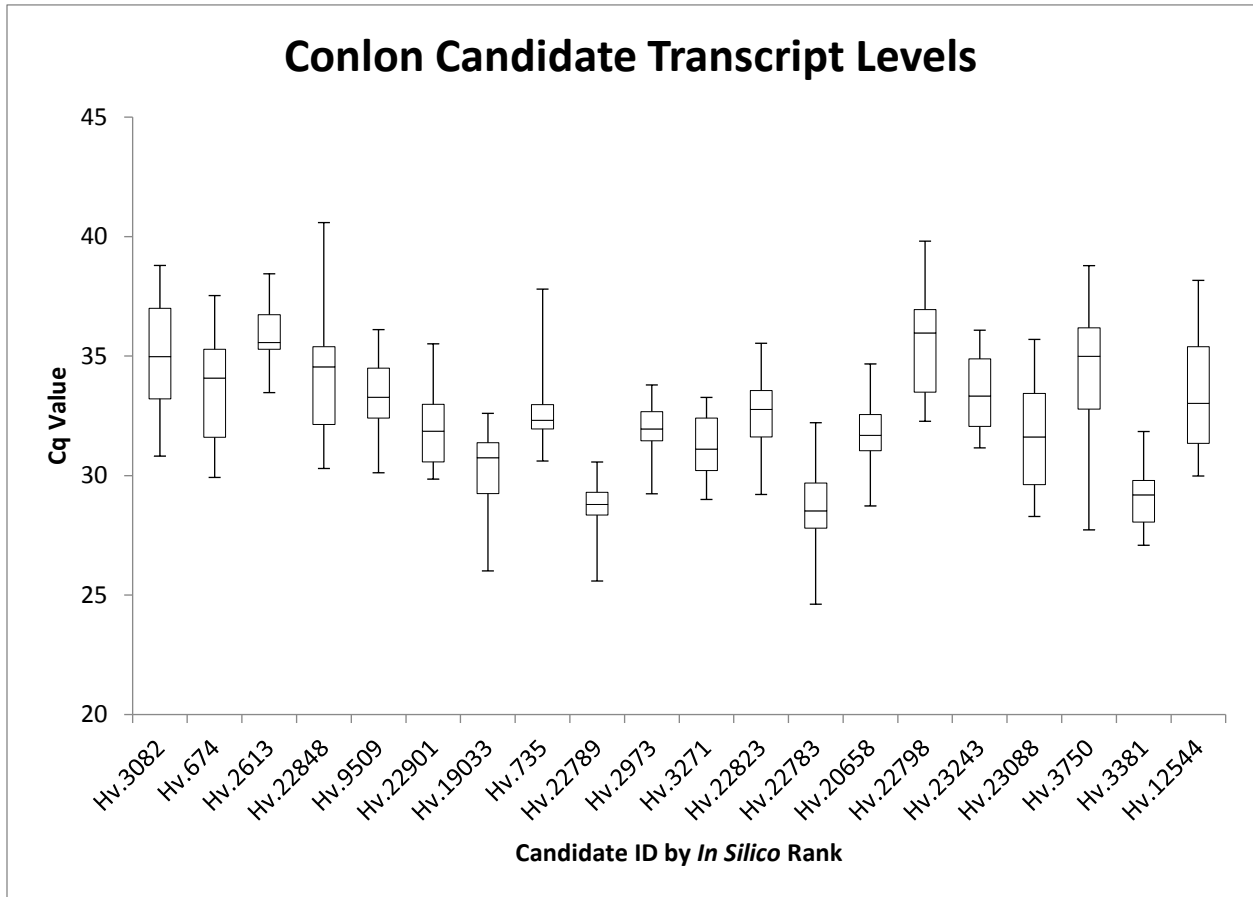
<i>In silico rank</i>	CV	UNIGENE Cluster	NCBI accession	<i>Gene</i>	<i>Forward primer</i>	<i>Reverse primer</i>	<i>Size (bp)</i>	<i>Efficiency (%)</i>	<i>R<sup>2</sup></i>
1	0.154	Hv.3082	AK362890	<u>NADH dehydrogenase</u>	AGATCGCATTTTCACGAACC	CCAGTCTGCCCTTTAATCA	112	107.8	.987
2	0.205	Hv.674	AK251074	Phosphoglycerate mutase	GCAGCAGTGTGGTTTCGTA	CCTGGGCTCTGCTCTTAG	135	109.5	.981
3	0.205	Hv.2613	AK251152	Proteasome subunit beta type-7-B	AGTTGCTTGTAAGGTGCC	CCAAGGTCATGTCAGTGCTG	145	101.3	.993
4	0.221	Hv.22848	X60343	<u>GAPDH</u>	GAGGGTCTGATGACCACTGT	AGATCAACAACAGACATCCAC	221	93.7	.982
5	0.231	Hv.9509	AK250157	<b>Elongation factor EF-2</b>	AACTGGCATGAAGGTCGGTA	GGCAGGCATCAACTTCCTTC	218	107.9	.988
6	0.240	Hv.22901	AK358926	<b>Malate dehydrogenase</b>	GCACTGGTGTGAATGTTGC	CTTCTCAGGGATAGATGGAGC	218	102.4	.98
7	0.242	Hv.19033	AK248694	<u>Heat Shock Protein 70</u>	CCTTTCTTCCACTGCCAAA	AGCTGCTGAACTCTGGGAAT	228	110.5	.986
8	0.244	Hv.735	Y09741	<u>Beta-tubulin 1</u>	CAGTTGGAGCGTGCAATGT	GCACCAGATTGCCAAAAGACA	175	100.4	.98
9	0.254	Hv.22789	AF230786	Translationally-controlled tumor protein homolog (HTP)	GTGCTCTGGGAAGTCGATG	CAACAATGTCAACCACCTTCAC	135	109.7	.997
10	0.255	Hv.2973	AK251228	3-ketoacyl-CoA thiolase like	AAGTGAGTGATGGTGCTGGA	GATCCACTCCAACAGCAACG	111	109.6	.997
11	0.259	Hv.3271	AK252295	Superoxide dismutase [Cu-Zn]	GCACCATCTTCTCACCCAG	GTTAGCAACACCATCCGCTC	239	107.0	.981
12	0.261	Hv.22823	AK359072	Mitochondrial ATP synthase beta subunit	GCCACAGAGCAGCAAATTCT	CACCAGCGAACACAGAGAAA	178	92.1	.984
13	0.264	Hv.22783	Z50789	Elongation factor 1-alpha	TTGGTGGCATTGGAACGTG	CAAACCCACGCTTGAGATCC	207	108.9	.998
14	0.276	Hv.20658	AM039896	S-adenosylmethionine synthetase 4	CAACATTGAGCAGCAGTCCC	GCAGCAATCTCATCGTTGGT	342	109.9	.986
15	0.287	Hv.22798	AY325266	<u>Cytosolic heat shock protein 90</u>	AGATCAACCAGTCTCTCTCG	CACAAGGTAGGCGGAGTAGA	359	103.6	.983
16	0.289	Hv.23243	AK250311	Zinc finger A20 & AN1 domain-containing stress-associated	TGAACATGTGCTCCAAGTGC	GTAGTCGAACTGGCAGTCATG	371	100.0	.994
17	0.292	Hv.23088	AY145451	<u>Actin</u>	GAATGGTCAAGGCTGGTTTCG	TCCATGTCATCCCAGTTGCT	205	110.0	.981
18	0.296	Hv.3750	AK371415	Peroxiredoxin-5, mitochondrial	GGTCATCTCTTCGGCGTC	CCTACAGTGCCTTGAGGATCTC	380	98.7	.983
19	0.299	Hv.3381	AK252614	Translation initiation factor eIF-5A	CACCACCTCGAGTCCAAGGC	CCATGCTTCCCAGTCTTGGA	146	108.0%	.995
20	0.304	Hv.12544	AK367709	Proteasome subunit alpha type-5-A	TCACGAGGACTGAGTACGAC	CTCAACAGCCAGGACAACAC	146	104.6%	.99



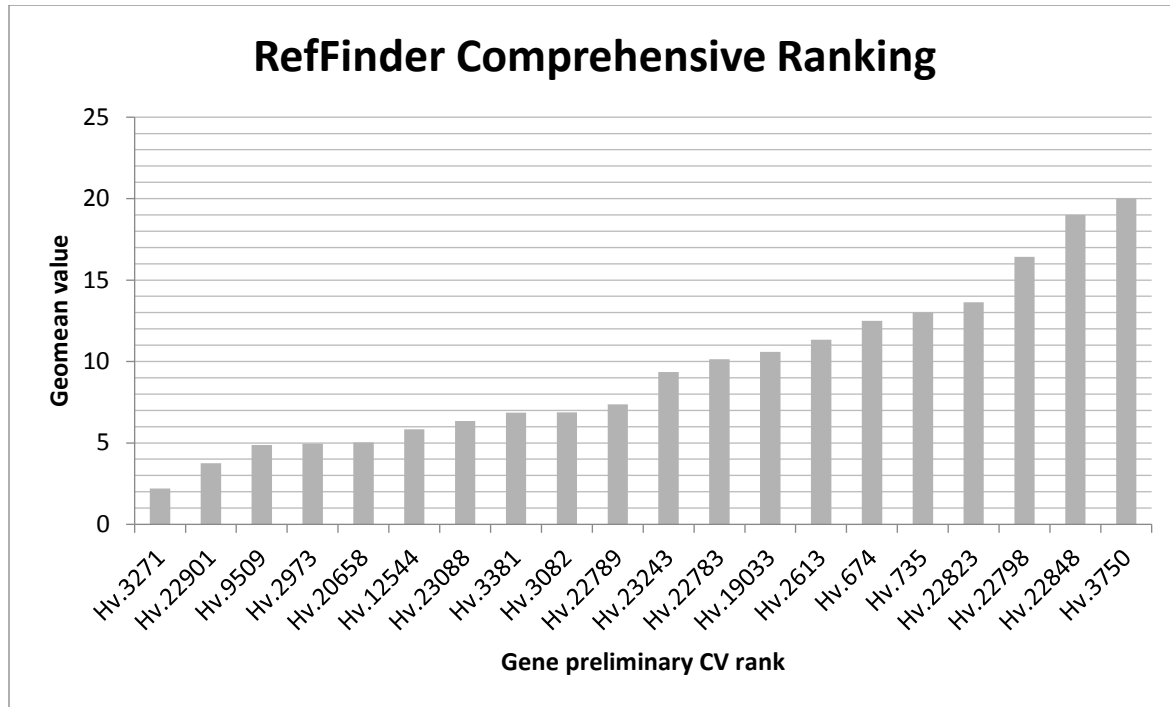
**Table 2-2: The RefFinder geomean stability ranking of candidates GAPDH, EF-2, Mal de, 3KACoA, SOD, S-AMS4, and Actin.**

	More stable						Less stable
Conlon	SOD	EF-2	S-AMS4	3KACoA	Mal de	Actin	GAPDH
Golden Promise	Mal de	EF-2	Actin	S-AMS4	3KACoA	SOD	GAPDH
Harrington	GAPDH	3KACoA	Actin	EF-2	S-AMS4	Mal de	SOD
HiFi	Actin	EF-2	Mal de	SOD	3KACoA	GAPDH	S-AMS4
All barley	Actin	EF-2	Mal de	S-AMS4	SOD	3KACoA	GAPDH
Barley plus oat	Actin	Mal de	EF-2	S-AMS4	SOD	GAPDH	3KACoA

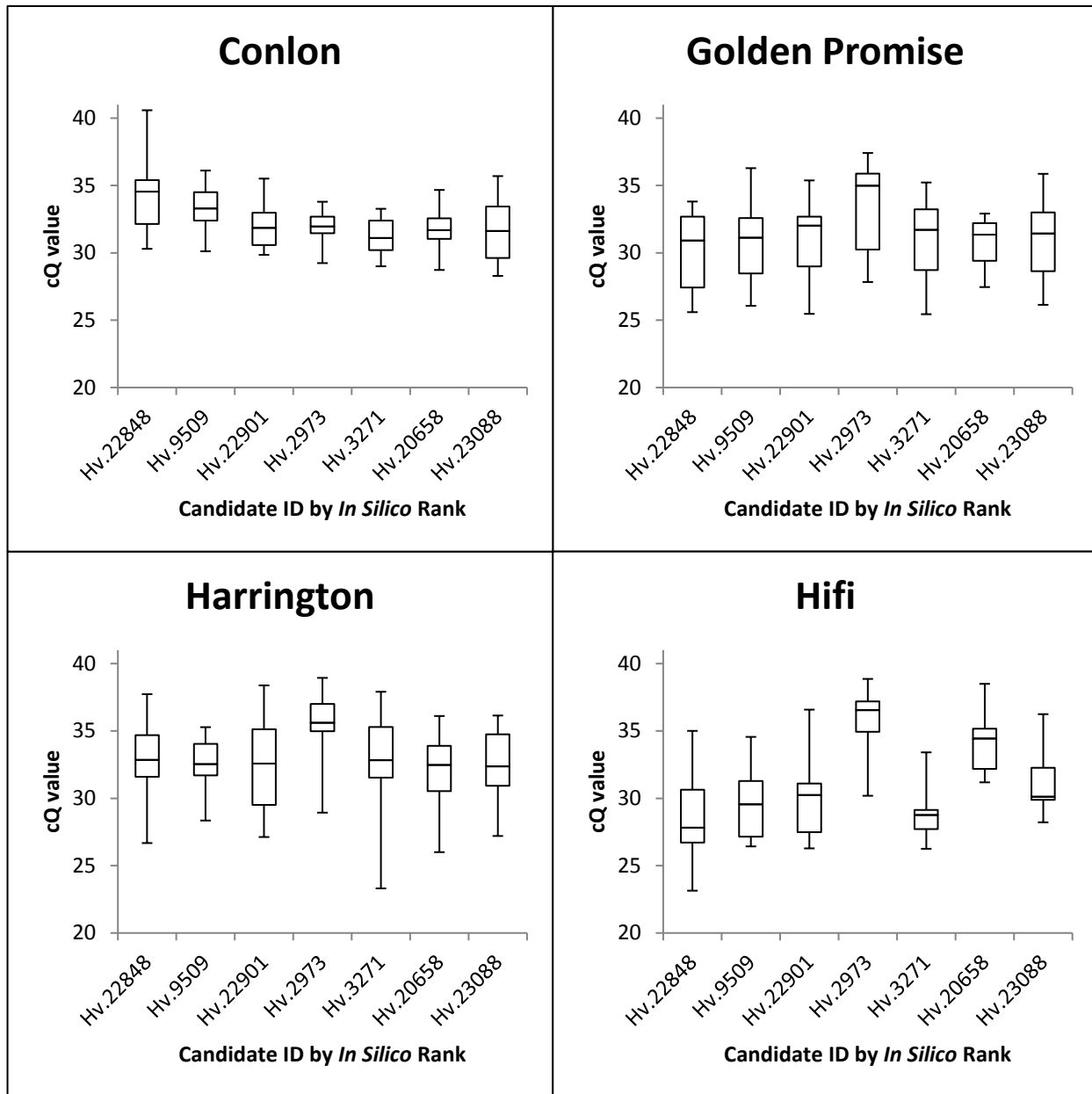
## FIGURES



**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.**



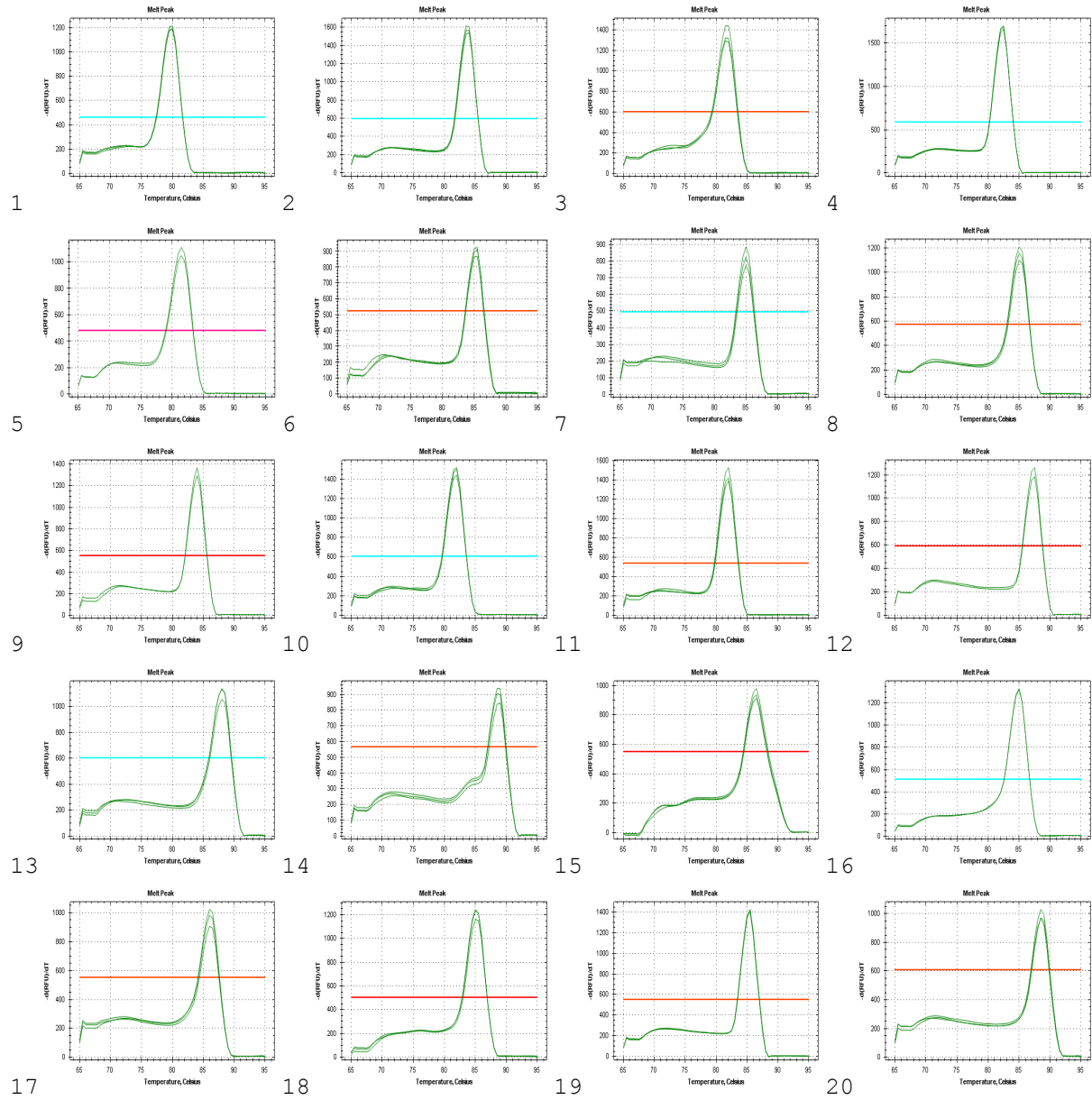
**Figure 2-2: A bar chart of the calculated stability values of the 20 candidate genes across 8 Conlon tissues. The y axis measures their geomean stability value calculated through RefFinder and the x-axis shows the candidates labeled by their UNIGENE cluster derived from the *in silico* rank. Hv.3082 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**

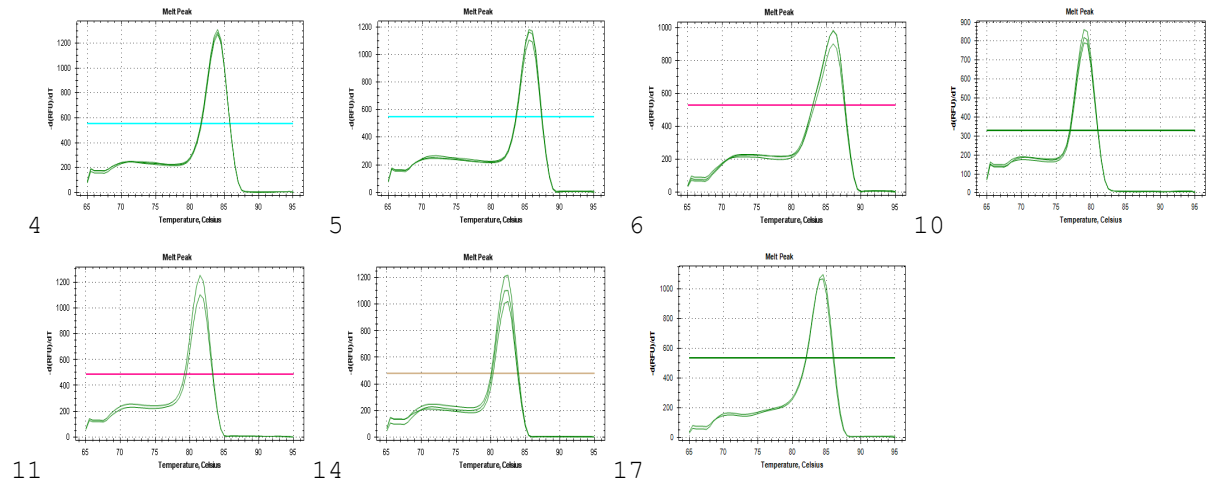
# APPENDIX 2-1

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.



<i>Candidate</i>	<i>Forward</i>	<i>Reverse</i>	<i>Efficiency</i>	$r^2$
GAPDH	GTCAACGACCCCTTCATCA	GCCTTGTCTTGTGTCAGTGAA	107.7%	0.999
EF-2	AACTGGCATGAAGGTCCGTA	GGCAGGCATCAACTTCCTTC	94.0%	0.997
Mal de	GCACTGGTGTGAATGTTGC	CTTCTCAGGGATAGATGGAGC	109.3%	0.987
3KACoA	GACTGGCGAAGCTTAAAACG	AAGAACAGCTCCAGCACCAT	93.3%	0.993
SOD	AAGTGAGTGATGGTGCTGGA	GATCCACTCCAACAGCAACG	104.4%	0.99
S-AMS4	CAACATTGAGCAGCAGTCCC	GCAGCAATCTCATCGTTGGT	98.7%	0.986
Actin	GAATGGTCAAGGCTGGTTTCG	TCCATGTCATCCCAGTTGCT	107.3%	0.989

## APPENDIX 2-3

The stability placement of the 20 candidate genes from Conlon expression by 4 individual analytical programs and their comprehensive ranking by RefFinder.

Rank	Acc.	Candidate	Delta CT	Bestkeeper	Normfinder	Genorm
1	NADH de	11	1st	6th	1st	4th
2	PGM	6	2nd	10th	2nd	5th
3	PSMB7	5	3rd	9th	3rd	7th
4	GAPDH	10	5th	4th	5th	6th
5	EF-2	14	4th	5th	4th	8th
6	Mal de	20	8th	18th	8th	Tied 1st
7	HSP 70	17	6th	15th	6th	3rd
8	$\beta$ tub	19	10th	2nd	10th	11th
9	HTP	1	11th	17th	12th	Tied 1st
10	3KCoA	9	14th	1st	14th	15th
11	SOD	16	7th	13th	7th	12th
12	ATP5B	13	12th	8th	11th	10th
13	EF-1 $\alpha$	7	9th	12th	9th	13th
14	S-AMS4	3	18th	3rd	18th	17th
15	HSP 90	2	13th	16th	13th	9th
16	ZF-stress as.	8	16th	7th	16th	16th
17	Actin	12	15th	11th	15th	14th
18	PRDX5	15	17th	14th	17th	18th
19	eIF-5A	4	19th	19th	19th	19th
20	PSAM5	18	20th	20th	20th	20th

## CHAPTER 3 : $\beta$ GLUX IS A GLUCOSIDASE CO-EXPRESSED WITH CSLF6 THROUGH OAT AND BARLEY CARYOPSIS STAGES

### ABSTRACT

The biosynthesis and pathway of hemicellulose (1,3;1,4)- $\beta$ -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  $\beta$ glux, was selected for expression analysis in HiFi oat using quantitative PCR.  $\beta$ glux 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)- $\beta$ -D glucan content.

### INTRODUCTION

Mixed linkage glucan ((1,3;1,4)- $\beta$ -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  $\beta$  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  $\beta$ -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  $\beta$ glux 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  $\beta$ glux in hexaploid *Avena sativa*, gene copies were isolated in bacterial clones prior to sequencing.  $\beta$ glux 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  $\beta$ glux 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  $\beta$ gluc 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  $\beta$ gluc. 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  $\mu$ L reactions, using 0.5  $\mu$ L 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  $\beta$ gluc and CslF6 were normalized through the  $\Delta\Delta$ Act 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  $\beta$ glux sequence-** The oat  $\beta$ glux 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  $\beta$ glux 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  $\beta$ gluc are shown in figure 3-10. Throughout the caryopsis stages,  $\beta$ gluc 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  $\beta$ gluc 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  $\alpha$ -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  $\beta$ -glucosidase enzymes. In barley, 614 amino acids of  $\beta$ glux (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  $\beta$ -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  $\beta$ glux 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;  $\beta$ glux is active during caryopsis development and shows an expression pattern similar to that of CslF6. There appears to be a steep reduction of CslF6 expression at day 25, but it returns to a higher net gain

at the next time point. From 15 DPA,  $\beta$ gluc continues to rise in activity. Also, as noticed in the investigation for a glucosidase in barley,  $\beta$ gluc does not always co-express with CslF6. In shoot and leaf tissue,  $\beta$ gluc 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  $\beta$ gluc 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  $\beta$ gluc at low levels during caryopsis development? Is there a correlation between  $\beta$ gluc activity and the length of  $\beta$ gluc chains? In regards to oat genomes, Coon (2012) observed differences between oat genomes for CslF6 activity in several cultivars. Would  $\beta$ gluc 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  $\beta$ gluc 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  $\beta$ glux. The candidate will be utilized in further analysis by comparing expression of the oat homologues for these genes.

## LITERATURE CITED

- Burton, R.A. and Fincher, G.B., 2009. (1, 3; 1, 4)- $\beta$ -d-glucans in cell walls of the Poaceae, lower plants, and fungi: a tale of two linkages. *Molecular Plant*, 2(5), pp.873-882. Coles, Graeme. "Relationship of mixed-link beta-glucan accumulation to accumulation of free sugars and other glucans in the developing barley endosperm." *Carlsberg Research Communications* 44, no. 6 (1979): 439-453.
- Coon, M.A., 2012. Characterization and Variable Expression of the CslF6 Homologs in Oat (*Avena* sp.).
- Gibeaut, D.M. and Carpita, N.C., 1993. Synthesis of (1 $\rightarrow$ 3),(1 $\rightarrow$ 4)-beta-D-glucan in the Golgi apparatus of maize coleoptiles. *Proceedings of the National Academy of Sciences*, 90(9), pp.3850-3854.
- Lee, H.K., Hsu, A.K., Sajdak, J., Qin, J. and Pavlidis, P., 2004. Coexpression analysis of human genes across many microarray data sets. *Genome research*, 14(6), pp.1085-1094.
- Lia, A., Hallmans, G., Sandberg, A.S., Sundberg, B., Aman, P. and Andersson, H., 1995. Oat beta-glucan increases bile acid excretion and a fiber-rich barley fraction increases cholesterol excretion in ileostomy subjects. *The American journal of clinical nutrition*, 62(6), pp.1245-1251.
- McClellan, B.V. and Glennie - Holmes, M., 1985. ENZYMIC QUANTIFICATION OF (1 $\rightarrow$ 3)(1 $\rightarrow$ 4)- $\beta$ -D-GLUCAN IN BARLEY AND MALT. *Journal of the Institute of Brewing*, 91(5), pp.285-295.

- Oliver, S., 2000. Proteomics: guilt-by-association goes global. *Nature*, 403(6770), pp.601-603.
- Shu, X. and Rasmussen, S.K., 2014. Quantification of amylose, amylopectin, and  $\beta$ -glucan in search for genes controlling the three major quality traits in barley by genome-wide association studies. *Frontiers in plant science*, 5, p.197.
- Taketa, S., Yuo, T., Tonooka, T., Tsumuraya, Y., Inagaki, Y., Haruyama, N., Larroque, O. and Jobling, S.A., 2012. Functional characterization of barley betaglucanless mutants demonstrates a unique role for CslF6 in (1, 3; 1, 4)- $\beta$ -D-glucan biosynthesis. *Journal of experimental botany*, 63(1), pp.381-392.
- Tsuchiya, K., Urahara, T., Konishi, T., Kotake, T., Tohno - oka, T., Komae, K., Kawada, N. and Tsumuraya, Y., 2005. Biosynthesis of (1 $\rightarrow$  3),(1 $\rightarrow$  4)- $\beta$ -glucan in developing endosperms of barley (*Hordeum vulgare*). *Physiologia Plantarum*, 125(2), pp.181-191.
- Varghese, J.N., Hrmova, M. and Fincher, G.B., 1999. Three-dimensional structure of a barley  $\beta$ -D-glucan exohydrolase, a family 3 glycosyl hydrolase. *Structure*, 7(2), pp.179-190.
- Varghese, J.N., Hrmova, M. and Fincher, G.B., 1999. Three-dimensional structure of a barley  $\beta$ -D-glucan exohydrolase, a family 3 glycosyl hydrolase. *Structure*, 7(2), pp.179-190.
- Wolfe, C.J., Kohane, I.S. and Butte, A.J., 2005. Systematic survey reveals general applicability of "guilt-by-association" within gene coexpression networks. *BMC bioinformatics*, 6(1), p.1.

Yang, M., Luoh, S.M., Goddard, A., Reilly, D., Henzel, W. and Bass, S., 1996. The bgIX gene located at 47.8 min on the Escherichia coli chromosome encodes a periplasmic  $\beta$ -glucosidase. *Microbiology*, 142(7), pp.1659-1665.

Zhang, G., Chen, J., Wang, J. and Ding, S., 2001. Cultivar and environmental effects on (1 $\rightarrow$  3, 1 $\rightarrow$  4)- $\beta$ -D-glucan and protein content in malting barley. *Journal of Cereal Science*, 34(3), pp.295-301.

TABLES

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

	Correlation	Distance Rank Average	Euclidean Distance	Final
1	MLOC_57785	MLOC_20723	MLOC_10919	<b>MLOC_58487</b>
2	MLOC_63574	<b>MLOC_58487</b>	MLOC_20723	MLOC_11508
3	MLOC_11508	MLOC_54331	MLOC_15619	MLOC_20723
4	<b>MLOC_58487</b>	MLOC_66806	MLOC_8615	MLOC_54331
5	MLOC_11885	MLOC_10919	<b>MLOC_58487</b>	MLOC_66806
6	MLOC_9865	MLOC_11508	MLOC_54331	MLOC_10919
7	MLOC_16595	MLOC_5621	MLOC_1898	MLOC_63574
8	MLOC_66806	MLOC_8669	MLOC_11508	MLOC_15619
9	MLOC_37740	MLOC_61073	MLOC_61073	MLOC_8615
10	MLOC_15694	MLOC_51368	MLOC_66806	MLOC_34767

**Table 3-2: Primers used for the sequencing of cloned *Bglux* 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.**

Direction	Primer Sequence	Sense 5' End	Sense 3' End
FWD*	TGTAACGACGGCCAGT	Trimmed	
FWD	CGTTTGTAGGGATCCGAGGT	730	749
REV	CCACTGTCTCCATGCCAAG	1592	1611
FWD	TCACGGATTGATGATGCTGT	1592	1611
REV	CCCTTAGAATCCGCTCAACA	1609	1628
REV	ACAAGCATACTGGAGGCCAT	2594	2613
REV*	AGCGGATAACAATTCACACAGG	Trimmed	



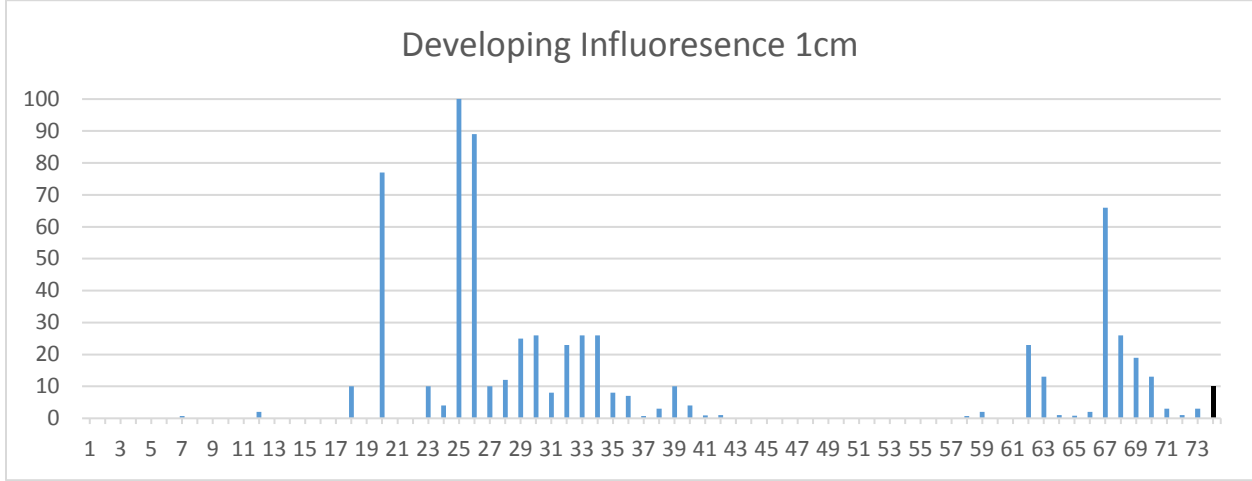
**Table 3-3: Exon position and length as figured from the alignment of 3 HiFi contigs and 3 SolFi contigs to the CORE EST consensus contig. The positions are numbered by their relation to the first nucleotide of the start codon.**

Exon	Position	Length
1	1-377	377
2	453-543	91
3	640-788	147
4	871-1114	244
5	1247-1351	105
6	1426-1608	183
7	1722-1923	202
8	2501-2997	497

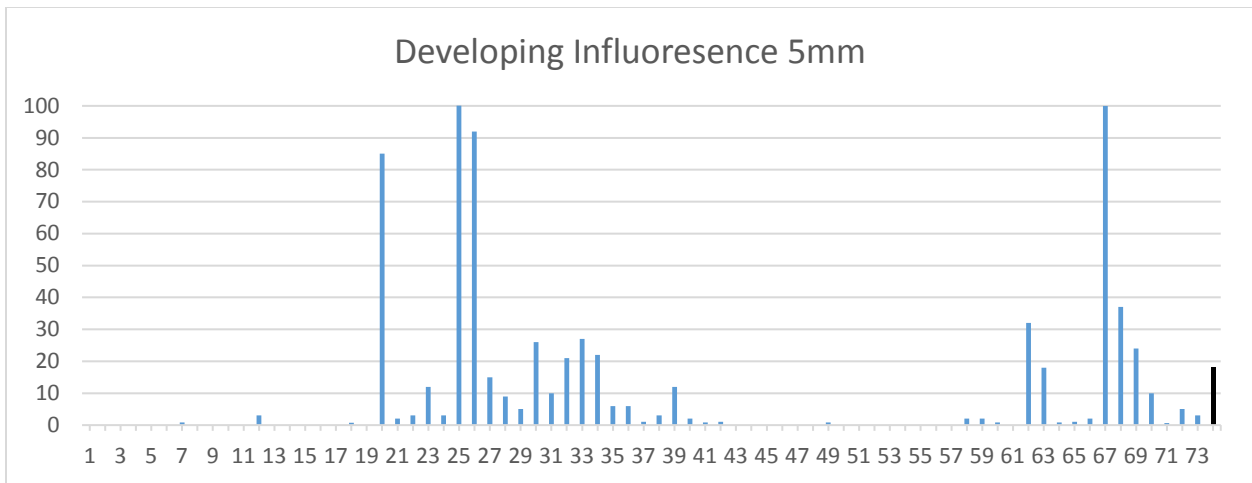
**Table 3-4: Primers used for real-time PCR with their efficiency and r<sup>2</sup> values. The \* indicate primers used by Coon 2012.**

Primer	Sequence	Efficiency	r <sup>2</sup>
βglux Fwd	AGATGCTCTAGTTGCTGCGT	107.5%	0.980
βglux Rev	AAACATTTTCAACCCATATCCAAAAGG		
CsfF6 Fwd_16*	AGGACGTCGTCACAGGCTAC	102.8%	0.983
CsfF6 Rev_16*	ACGATGCCCGAGGTAGACGTA		
Actin Fwd	GAATGGTCAAGGCTGGTTTCG	107.3%	0.989
Actin Rev	TCCATGTCATCCCAGTTGCT		
EF2 Fwd	AACTGGCATGAAGGTCCGTA	94.0%	0.997
EF2 Rev	GGCAGGCATCAACTTCCTTC		
Malate De Fwd	GCACTGGTGTGAATGTTGC	109.3%	0.987
Malate De Rev	CTTCTCAGGGATAGATGGAGC		

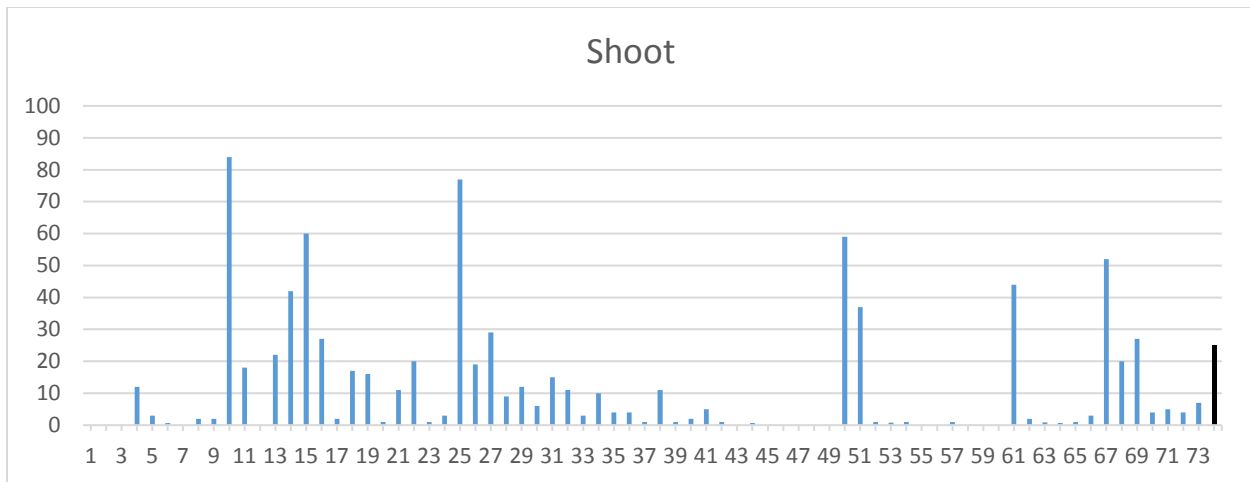
## FIGURES



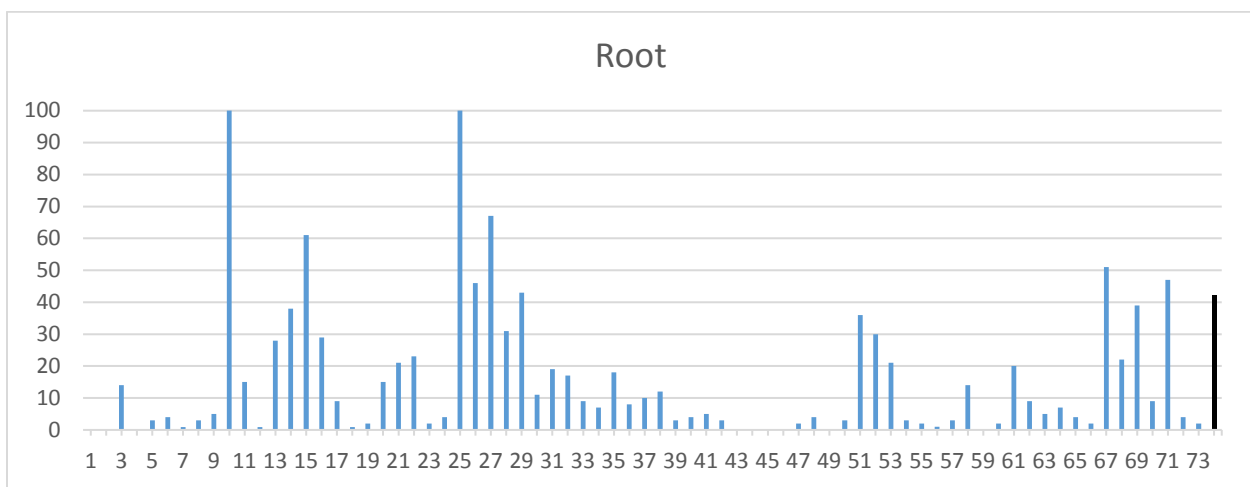
**Figure 3-1: Glucosidase transcript levels in developing inflourescence 1cm tissue. Arbitrary transcript units (Y-axis) for each accession (X-axis). Window cap at 100 units. MLOC\_57200 is colored black and is the last accession.**



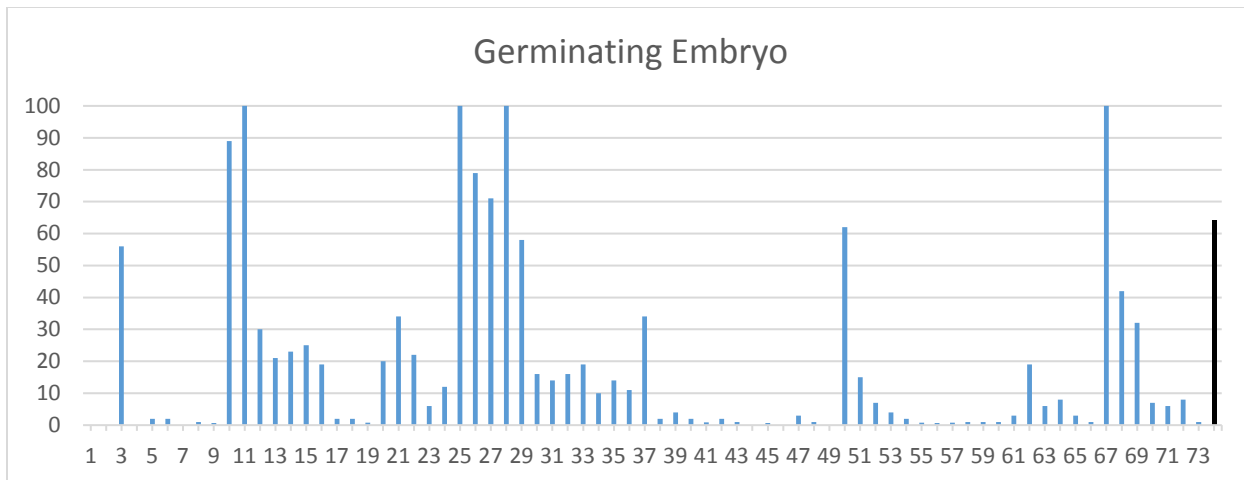
**Figure 3-2: Glucosidase transcript levels in developing inflourescence 5mm tissue. Arbitrary transcript units (Y-axis) for each accession (X-axis). Window cap at 100 units. MLOC\_57200 is colored black and is the last accession.**



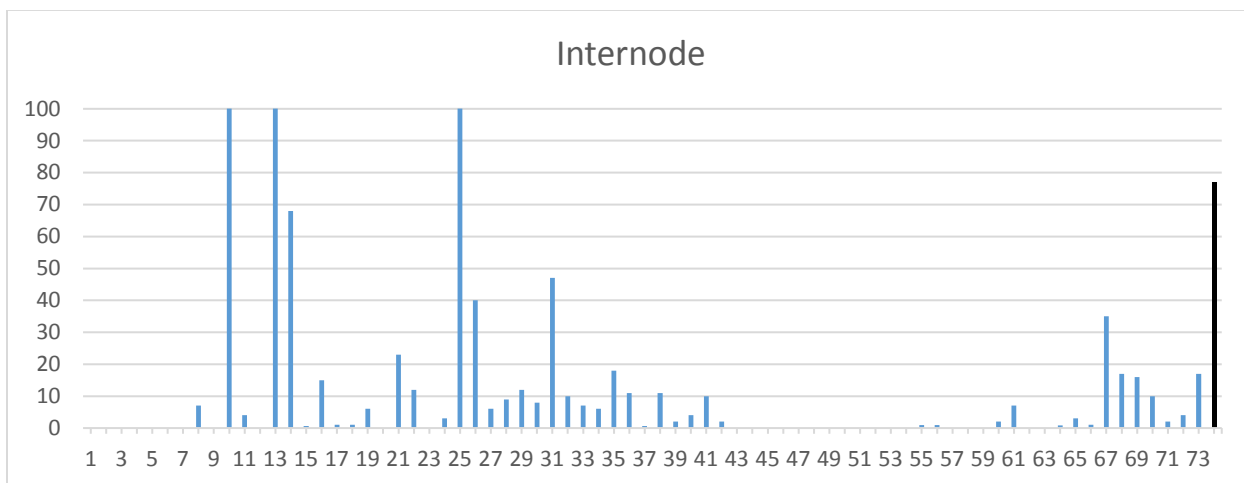
**Figure 3-3: Glucosidase transcript levels in shoot tissue. Arbitrary transcript units (Y-axis) for each accession (X-axis). Window cap at 100 units. MLOC\_57200 is colored black and is the last accession.**



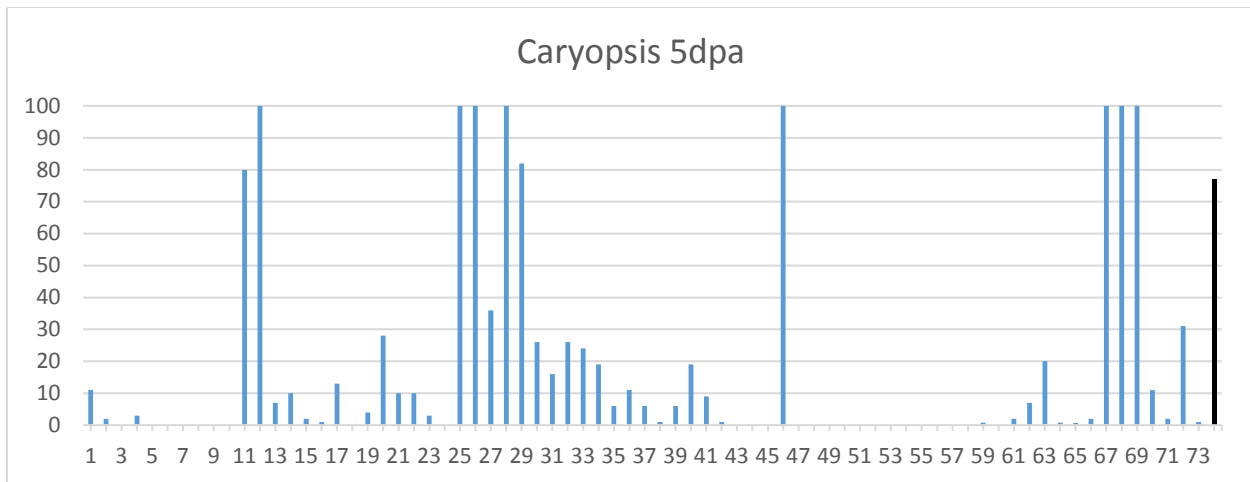
**Figure 3-4: Glucosidase transcript levels in root tissue. Arbitrary transcript units (Y-axis) for each accession (X-axis). Window cap at 100 units. MLOC\_57200 is colored black and is the last accession.**



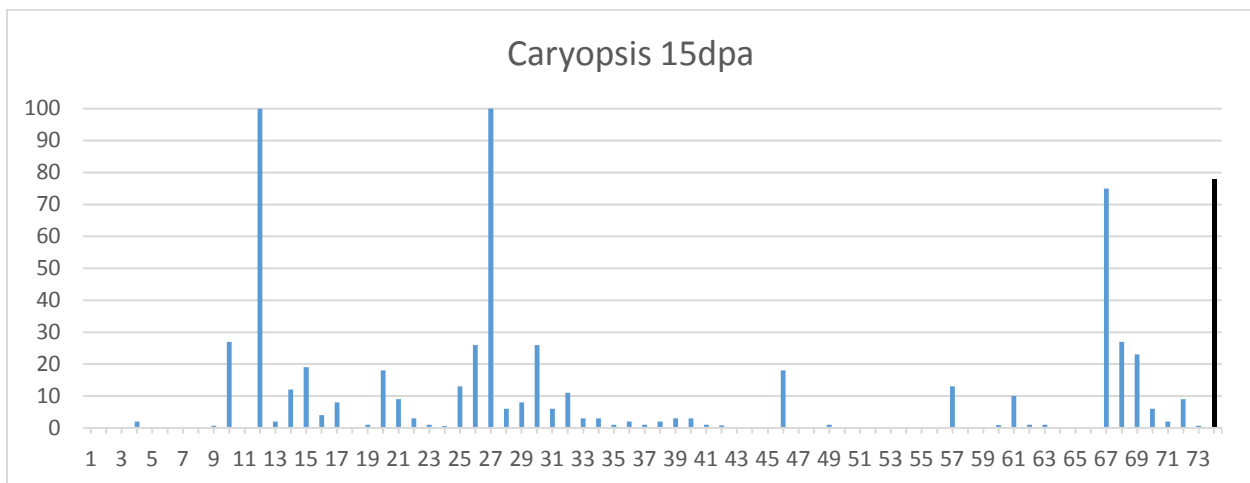
**Figure 3-5: Glucosidase transcript levels in germinating embryo tissue. Arbitrary transcript units (Y-axis) for each accession (X-axis). Window cap at 100 units. MLOC\_57200 is colored black and is the last accession.**



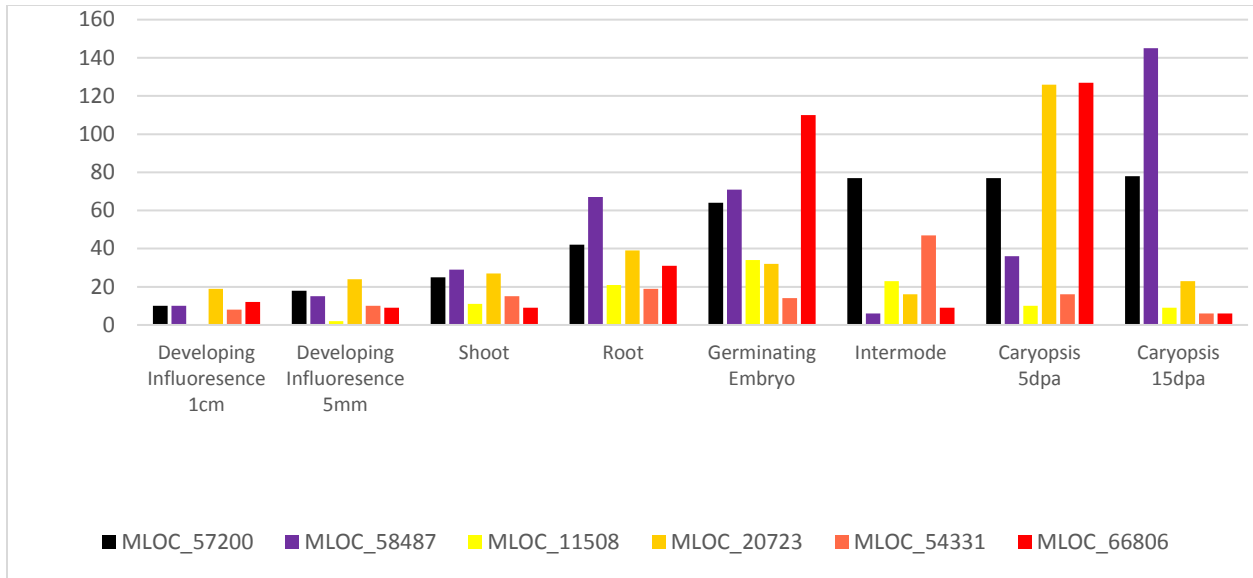
**Figure 3-6: Glucosidase transcript levels in internode tissue. Arbitrary transcript units (Y-axis) for each accession (X-axis). Window cap at 100 units. MLOC\_57200 is colored black and is the last accession.**



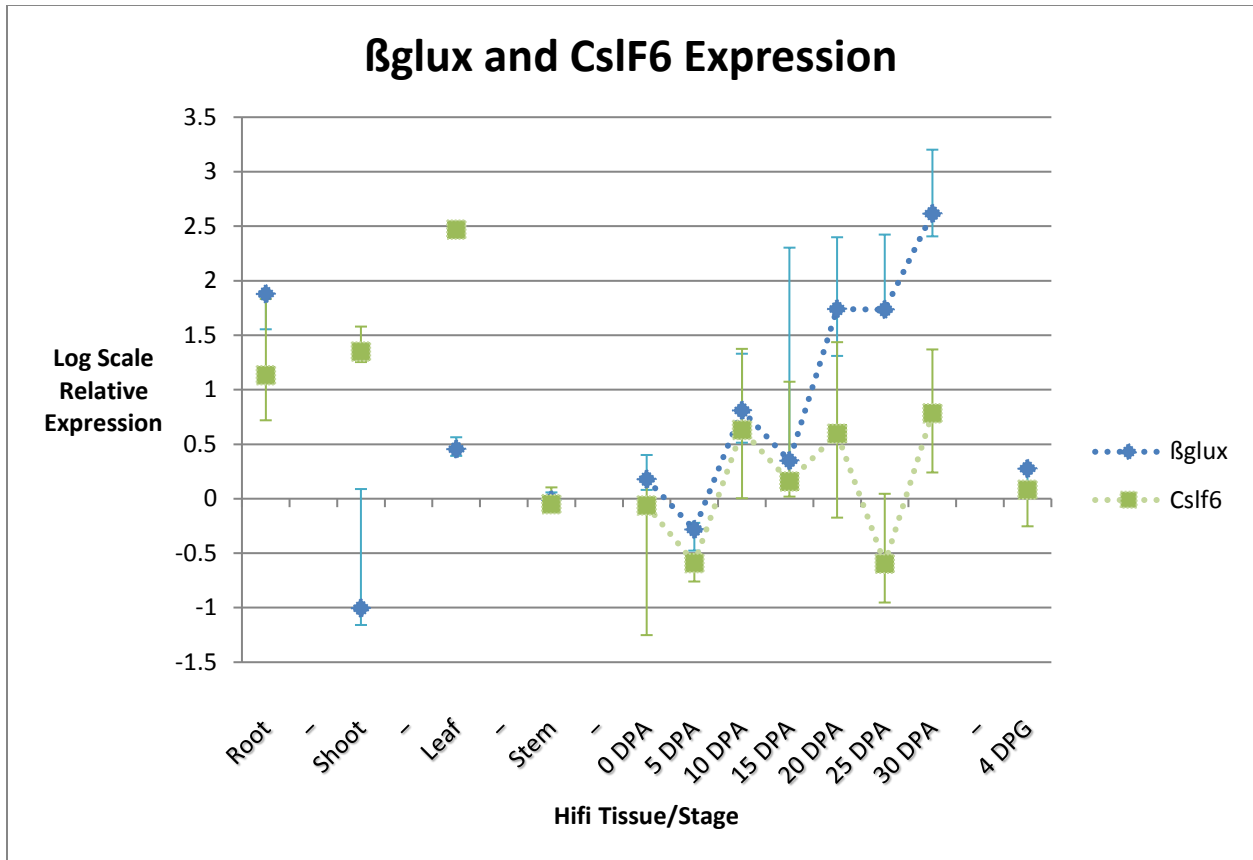
**Figure 3-7** Glucosidase transcript levels in caryopsis tissue 5 days post anthesis (DPA). Arbitrary transcript units (Y-axis) for each accession (X-axis). Window cap at 100 units. MLOC\_57200 is colored black and is the last accession.



**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 CsIF6 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 CsIF6.

APPENDIX 3-1

The 73 glucosidase candidates with expression data from Expression Atlas for Barley. CslF6 is represented in bold as MLOC\_57200.

	Accession	Developing Influorescence 1 cm	Developing Influorescence 5 mm	Shoot	Root	Germinating Embryo	Internode	Caryopsis 5 dpa	Caryopsis 15 dpa
1	MLOC_61980	0	0	0	0	0	0	11	0
2	MLOC_18750	0	0	0	0	0	0	2	0
3	MLOC_56792	0	0	0	14	56	0	0	0
4	MLOC_37976	0	0	12	0	0	0	3	2
5	MLOC_80849	0	0	3	3	2	0	0	0
6	MLOC_10669	0	0	0.7	4	2	0	0	0
7	MLOC_13831	0.7	0.8	0	0.9	0	0	0	0
8	MLOC_66576	0	0	2	3	1	7	0	0
9	MLOC_57283	0	0	2	5	0.7	0	0	0.7
10	MLOC_62746	0	0	84	408	89	115	0	27
11	MLOC_79756	0	0	18	15	249	4	80	0
12	MLOC_9865	2	3	0	0.9	30	0	249	2419
13	MLOC_34767	0	0	22	28	21	126	7	2
14	MLOC_8615	0	0	42	38	23	68	10	12
15	MLOC_7890	0	0	60	61	25	0.6	2	19
16	MLOC_44074	0	0	27	29	19	15	1	4
17	MLOC_57785	0	0	2	9	2	1	13	8
18	MLOC_16506	10	0.7	17	0.9	2	1	0	0
19	MLOC_66740	0	0	16	2	0.8	6	4	1
20	MLOC_24253	77	85	1	15	20	0	28	18
21	MLOC_11508	0	2	11	21	34	23	10	9
22	MLOC_5621	0	3	20	23	22	12	10	3
23	MLOC_75585	10	12	1	2	6	0	3	1
24	MLOC_4960	4	3	3	4	12	3	0	0.6
25	MLOC_37740	155	150	77	284	560	256	711	13



26	MLOC_69905	89	92	19	46	79	40	110	26
27	MLOC_58487	10	15	29	67	71	6	36	145
28	MLOC_66806	12	9	9	31	110	9	127	6
29	MLOC_10919	25	5	12	43	58	12	82	8
30	MLOC_1898	26	26	6	11	16	8	26	26
31	MLOC_54331	8	10	15	19	14	47	16	6
32	MLOC_61073	23	21	11	17	16	10	26	11
33	MLOC_37773	26	27	3	9	19	7	24	3
34	MLOC_76578	26	22	10	7	10	6	19	3
35	MLOC_8669	8	6	4	18	14	18	6	1
36	MLOC_20526	7	6	4	8	11	11	11	2
37	MLOC_67725	0.7	1	1	10	34	0.6	6	1
38	MLOC_44982	3	3	11	12	2	11	1	2
39	MLOC_60711	10	12	1	3	4	2	6	3
40	MLOC_15694	4	2	2	4	2	4	19	3
41	MLOC_11885	0.9	0.8	5	5	0.9	10	9	1
42	MLOC_61690	1	1	1	3	2	2	1	0.8
43	MLOC_78985	0	0	0	0	1	0	0	0
44	MLOC_65592	0	0	0.7	0	0	0	0	0
45	MLOC_56112	0	0	0	0	0.7	0	0	0
46	MLOC_16595	0	0	0	0	0	0	145	18
47	MLOC_4276	0	0	0	2	3	0	0	0
48	MLOC_77701	0	0	0	4	1	0	0	0
49	MLOC_76387	0	0.8	0	0	0	0	0	1
50	MLOC_73077	0	0	59	3	62	0	0	0
51	MLOC_13908	0	0	37	36	15	0	0	0
52	MLOC_67705	0	0	1	30	7	0	0	0
53	MLOC_72542	0	0	0.8	21	4	0	0	0
54	MLOC_81871	0	0	1	3	2	0	0	0
55	MLOC_39219	0	0	0	2	0.8	0.9	0	0
56	MLOC_51439	0	0	0	1	0.7	0.9	0	0
57	MLOC_13058	0	0	1	3	0.8	0	0	13
58	MLOC_60415	0.7	2	0	14	1	0	0	0
59	MLOC_45963	2	2	0	0	1	0	0.8	0
60	MLOC_51849	0	0.8	0	2	1	2	0	0.9

61	MLOC_15992	0	0	44	20	3	7	2	10
62	MLOC_75464	23	32	2	9	19	0	7	1
63	MLOC_14239	13	18	0.9	5	6	0	20	1
64	MLOC_3522	1	0.8	0.7	7	8	0.8	0.8	0
65	MLOC_63284	0.8	1	1	4	3	3	0.7	0
66	MLOC_822	2	2	3	2	1	1	2	0
67	MLOC_68876	66	100	52	51	100	35	324	75
68	MLOC_15619	26	37	20	22	42	17	135	27
69	MLOC_20723	19	24	27	39	32	16	126	23
70	MLOC_51368	13	10	4	9	7	10	11	6
71	MLOC_12571	3	0.6	5	47	6	2	2	2
72	MLOC_63574	1	5	4	4	8	4	31	9
73	MLOC_10974	3	3	7	2	1	17	1	0.7
<b>HvCsIF6</b>	<b>MLOC_57200</b>	<b>10</b>	<b>18</b>	<b>25</b>	<b>42</b>	<b>64</b>	<b>77</b>	<b>77</b>	<b>78</b>

## APPENDIX 3-2

The six contigs acquired from *β*gluc sequencing and assembly are shown here in FASTA format.

```
>HF_contig_1_@7/20/2016,_9_14_AM
-----ATGCTCGAGCGGCCGCCAGT-GTGATGG
ATA-TCTGC-AGAATTCAGGTCGCTCGCTTCTTCTTACCCCATGGCGGCG
GCGCAAGGCGAGGACGTGCCGCTGTACAAGGACGCGTCGGCGCCGGAGGA
GGCGCGCGTGC GCGACCTGCTGGGCCGGATGACGCTGCGGGAGAAGGCGG
GCCAGATGGCCCAGATCGAGCGCACCGTCGTGTCGCCCCGCGCCCTCACA
GAGTCGCCGCGGCAGCGTCCTCAACGCCGGCGGAAGCGCCCCCGCGA
ACGCGCCTCCCCGGCCGACTGGGCCCCGCATGGTCGACGACATGCAGCGCC
TCGCCCTCTCCTCCCGCCTCGCCGTCCCCATCCTCTACGGCACCGACGCC
GTCCACGGGCACAACAACGTCTTCGGCGCCACCGTCTTCCCTCACAACGT
TGGCCTCGGAGCCTCCAGGTACCCAATCACCTGACGCCATATGACGCTCA
TTGACCACGTCGACGCACTTGACCGTACGCCGTTGACTTCCAGGGACGCG
GAGCTTGTGCGTAAGATTGGCGAGGTGACGGCGCTCGAGGTCCGCGCCAC
CGGCATACACTGGGCCTTTGCGCCCTGCCTCGCCGTGAGTTTAAATTCAA
AAGGCTTCCGGATTAAC TGGCGGTGTCTATTTTGGGGTCCTTGATTGTGT
G-TTTGCTCTGTCTGGCGGGTTATGTCCAGGTTTGTAGGGATCCGAGGTG
GGGAGATGTTACGAGAGCTACAGCGAGGACCCAGAGATCGTGCGCTCGT
TCACCAACATTGTCACCGGCTTG CAGGGCCAGCCACCAGCAGATCACCT
CATGGTTACCCGTTCCCTTGCTTCGGTTAGGTAATTCCACTGGTGAGATCT
AGCTGTGACTTCTGCAGACGCCACATCCTGTTTATTGAAATGGTTCTAT
GTGATTTACAGGGAGAATGTGCTTGCTTGTGCCAAGCATTTTGTAGCGGA
TGGTGGCACTGACAAGGACTCAATGAGGGGAACACCATATGCTCGCCAG
AAGATTTGGAAAGGATCCACATGAAACCTTACCCTGATTGTATAACTCAA
GGGGTGGCAACGGTCATGGCGTCTACTCTAAGTGGAATGGGGAGCCATT
GCATGCCAGCCGCTATTTGCTTACGGATGTTTTAAAGGGCAAGCTAGGCT
TCAAGGTAGTAGAGCATGTGCTATTTGATTGATATCTTGATGCAAACCTC
TGATCTGAATAGTACTCTGGCCTGTCAAAAAAGAATAGGGACTATGGG
```

TTGTGAACATGTCATCATGTGTAAGCTCTTCTTTCAGGGTTTTGTGGTGT  
CAGACTGGGAGGGTATTGACAGGCTCTGCGAGCCTCGAGGGTCTGATTAT  
CGCTACTGCATTGCACAGTCAGTTAATGCTGGAATGGATATGGTATTATT  
TATGAAGTAGTCTTAAGATATATTCTTCAGAGACAGGTTCTCTTGCATTT  
TCTGATCGACCTACAGATTATGATACCTCACAGATTTGAGAAATTCTTGG  
AAGATCTTGTGTTCTTGGTGGAGACAGGGGAGATACCAATATCACGGATT  
GATGATGCTGTTGAGCGGATTCTAAGGGTTAAGTTCATATCTGGAGTGTT  
TGAGCATCCATTTTCAGATCCGTCTCTGCTAGACATAGTCGGCTGTAAGG  
TAGCATAcATTTTGCATATCTATATGTTTCTGCTGTTAATATTTACATCT  
TAATTGTCTCAAATTGCATATCAGTGATAATATACGATTATGCTAAATT-  
TTCAACTTTCAGGAGCATCGGCTACTGGCCCGTGAGGCCGTTTCGAAAGTC  
TTTGGTACTTCTGAAAAATGGCAAGAATCAGAAGGAAACGTTCCCTCCAT  
TGGCCAAAAATGCAAAACGAATACTCGTTGCAGGGACACATGCTGACGAT  
ATTGGATATCAGTGCGGTGGGTGGACAATAGCTTGGCATGGAGACAGTGG  
AAAGATAACTCTTGGTAAGCAAAAACCTCAACTTTTCTTTCACATTACTTT  
TTTGTGTTGAGGGAGTACTTACCATTAGCTTGGTACTGAATAAACTGTTGC  
TCATCTAGAATTCTAGCTACCAAGTAGTTAGGGATAACATATGTCCTATT  
TTAAACTGCTATGATGCATACAGATGTGGAAGTGTGCTCCAAAGTTGAT  
GTTTAAGATGTTCAATGGATCGTTCAAGTATTGTGTTGTGGTTGTCTGTT  
TTTAGACTGATGGGTGATGGCACTGCCTGGTTTGTACTGAAAGCCAAAAT  
AAAATATAAGATACACTGGATAGTAACTTACGAGTCTACATTCAAACAAC  
TGCAGAGATGCCACAATGCGAGCTAT-GTCACCAATTACTGAACTAACAA  
ATGTTGGTGAAAAATAAATTAAGCACATTTCCTTTTCGTTTTCAAGCCAGA  
CATCACCCTTTTAATTAGATTTATTTGTTCTATTTATATTAATAAAACTT  
TCATCAGGTAGAAGCAGTATAAAATCTGCACTTTGCACAATA-TATGCTA  
AGAAACAGATATTTATGATATCTAAAGTTCTACACCAACAGGTACAAGCA  
TACTAGAGGCCATACAAGAATCCGTCCGAGTGGAAACTGAAGTTGTGCAT  
GAGCAATGCCCAACAGATGCTATCATTGAAACTGGAGGATTTTCTTATGC  
AGTTGTTGTTGTTGGCGAGGTTCCCTTATGCTGAATGGTTGGGAGATAGAA  
CTGACCTTTGTATTCCCTTCAATGGTTCCGACCTGATTACTCGGGTTGCT

AGTAAAGTCCCTACACTAGTGATTGTCATATCTGGAAGGCCCTTGGTTAT  
TGATCCACAAGTACTGGAAAAGATAGATGCTCTAGTTGCTGCGTGGCTAC  
CTGGAAGTGAGGGCAACGGAATTACTGATTGCATCTTTGGGGACCATGAT  
TTTGTGGCACATTGCCTGTAACGTGGTTTAGGTTTTCTGATCAATTGCC  
TATAAATATTGGAGATGCCAAATATGACCCCTTATTTCTTTTGGATATG  
GGTTGAAAATGTTTAGAAGTGATAAAGGTTACCCATAACCTCCTGAATTC  
CAGCACT---CAYKKSGYMRKSWAGCGSYKGAATYCSAGC-TMSRYAY  
SGMRGCTYGWTRCATAGCTKGATCCGAGCTCGGTACCAAGCTTGATGCAT  
AGCT

>SF\_contig\_1\_@7/20/2016,\_9\_14\_AM

-----  
-----CTGGAATTCAGGTCGCTCGCTTCTTCTTACCCCATKGCGGCG  
GCGCAAGGCGAGGACGTGCCGCTGTACAAGGACGCGTCGGCGCCGGAGGA  
GGCGCGCGTGC GCGACCTGCTGGGCCGGATGACGCTGCGGGAGAAGGCGG  
GCCAGATGGCCCAGATCGAGCGCACCGTCGTGTCGCCCCGCGCCCTCACA  
GAGCTCGCCGCCGGCAGCGTCTCAACGCCGGCGGAAGCGCCCCCGCGA  
ACGCGCCTCCCCGGCCGACTGGGCCCGCATGGTCGACGACATGCAGCGCC  
TCGCCCTCTCCTCCCGCCTCGCCGTCCCCATCCTCTACGGCACCGACGCC  
GTCCACGGGCACAACAACGTCTTCGGCGCCACCGTCTTCCCTCACAACGT  
TGGCCTCGGAGCCTCCAGGTACCCAATCACCTGACGCCATATGACGCTCA  
TTGACCACGTGACGCACTTGACCGTACGCCGTTGACTTCCAGGGACGCG  
GAGCTTGTGCGTAAGATTGGCGAGGTGACGGCGCTCGAGGTCCGCGCCAC  
CGGCATACTGGGCCTTTGCGCCCTGCCTCGCCGTGAGTTTAAATTCAA  
AAGGCTTCCGGATTAAC TGGCGGTGTCTATTTTGGGGTCCTTGATTGTGT  
G-TTTGCTCTGTCTGGCGGGTTATGTCCAGGTTTGTAGGGATCC-----  
-----  
-----CAGCAGATCACCT  
CATGGTTACCCGTTCCCTTGCTTCGGTTAGGTAATTCCACTGGTGAGATCT  
AGCTGTGACTTCTGCAGACGCCACATCCTGTTTATTGAAATGGTTCTAT  
GTGATTTACAGGGAGAATGTGCTTGCTTGTGCCAAGCATTTTGTAGCGGA

TGGTGGCACTGACAAGGGACTCAATGAGGGGAACACCATATGCTCGCCAG  
AAGATTTGGAAAGGATCCACATGAAACCTTACCCTGATTGTATAACTCAA  
GGGGTGGCAACGGTCATGGCGTCCTACTCTAAGTGGAATGGGGAGCCATT  
GCATGCCAGCCGCTATTTGCTTACGGATGTTTTAAAGGGCAAGCTAGGCT  
TCAAGGTAGTAGAGCATGTGCTATTTGATTGATATCTTGATGCAAACCTC  
TGATCTGAATAGTGA CTCTGGCCTGTCAAAAAAGAATAGGGACTATGGG  
TTGTGAACATGTCATCATGTGTAAGCTCTTCTTTCAGGGTTTTGTGGTGT  
CAGACTGGGAGGGTATTGACAGGCTCTGCGAGCCTCGAGGGTCTGATTAT  
CGCTACTGCATTGCACAGTCAGTTAATGCTGGAATGGATATGGTATTATT  
TATGAAGTAGTCTTAAGATATATTCTTCAGAGACAGGTTCTCTTGCATTT  
TCTGATCGACCTACAGATTATGATACCTCACAGATTTGAGAAATTCTTGG  
AAGATCTTGTGTTCTTGGTGGAGACAGGGGAGATACCAATATCACGGATT  
GATGATGCTGTTGAGCGGATTCTAAGGGTTAAGTTCATATCTGGAGTGTT  
TGAGCATCCATTTTCAGATCCGTCTCTGCTAGACATAGTCGGCTGTAAGG  
TAGCATACATTTTGCATATCTATATGTTTCTGCTGTTAATATTTACATCT  
TAATTGTCTCAAATTGCATATCAGTGATAATATACGATTATGCTAAATT-  
TTCAACTTTCAGGAGCATCGGCTACTGGCCCGTGAGGCCGTTTCGAAAGTC  
TTTGGTACTTCTGAAAATGGCAAGAATCAGAAGGAAACGTTCTTCCAT  
TGGCCAAAATGCAAAACGAATACTCGTTGCAGGGACACATGCTGACGAT  
ATTGGATATCAGTGCGGTGGGTGGACAATAGCTTGGCATGGAGACAGTGG  
AAAGATAACTCTTGGTAAGCAAAA ACTCAACTTTTCTTTCACATTACTTT  
TTTGTGTTGAGGGAGTACTTACCATTAGCTTGGTACTGAATAAACTGTTGC  
TCATCTAGAATTCTAGCTACCAAGTAGTTAGGGATAACATATGTCCTATT  
TTAAACTGCTATGATGCATACAGATGTGGA ACTGTTGCTCCAAAGTTGAT  
GTTTAAGATGTTCAATGGATCGTTCAAGTATTGTGTTGTGGTTGTCTGTT  
TTTAGACTGATGGGTGATGGCACTGCCTGGTTTGTACTGAAAGCCAAAAT  
AAAATATAAGATACACTGGATAGTAACTTACGAGTCTACATTCAAACAAC  
TGCAGAGATGCCACAATGCGAGCTAT-GTCACCAATTACTGAACTAACAA  
ATGTTGGTAAAAATAAATTAAGCACATTTCTTTTCGTTTCAAGCCAGA  
CATCACC ACTTTTAATTAGATTTATTTGTTCTATTTATATTA AAAA ACTT

TCATCAGGTAGAAGCAGTATAAAAATCTGCACTTTGCACAATA-TATGCTA  
AGAAACAGATATTTATGATATCTAAAGTTCTACACCAACAGGTACAAGCA  
TACTAGAGGCCATACAAGAATCCGTCGGAGTGGAAGTGAAGTTGTGCAT  
GAGCAATGCCCAACAGATGCTATCATTGAAACTGGAGGATTTTCTTATGC  
AGTTGTTGTTGTTGGCGAGGTTTCCTTATGCTGAATGGTTGGGAGATAGAA  
CTGACCTTTGTATTCCCTTCAATGGTTCCGACCTGATTACTCGGGTTGCT  
AGTAAAGTCCCTACACTAGTGATTGTCATATCTGGAAGGCCCTTGTTAT  
TGATCCACAAGTACTGGAAAAGATAGATGCTCTAGTTGCTGCGTGGCTAC  
CTGGAAGTGAGGGCAACGGAATTACTGATTGCATCTTTGGGGACCATGAT  
TTTGTGGCACATTGCCTGTAACGTGGTTTAGGTTTTCTGATCAATTGCC  
TATAAATATTGGAGATGCCAAATATGACCCCTTATTTCTTTTGGATATG  
GGTTGAAAATGTTTAGAAGTKATAAAGGTTACCATAACCTCCTGAATTC  
C-----  
-----  
-----

>HF\_contig\_2\_@7/20/2016,\_9\_13\_AM

-----AGCTTGGTACCGAGCTCG-G-ATCCACTAGTAACGGC-G  
CNGGNGTGCTGGAATTCAGGTTCGCTCGCTTCTTCTTACCC-ATGGNNGCG  
GCGCAAGGCGAGGACGTGCCGCTGTACAAGGACGCGTCGGCGCCGGTGA  
GGCGCGCGTGCGCGACCTGCTGGGCCGGATGACGCTGCGGGAGAAGGCGG  
GCCAGATGGCCCAGATCGAGCGCACCGTCGTGTCGCCCCGCGCCCTCACG  
GAGCTCGCCGCCGCGCAGCGTCCTCAACGCCGGCGGAAGCGCCCCCGCGA  
ACGCGCCTCCCCGGCCGACTGGGCCCGCATGGTCGACGACATGCAGCGCC  
TCGCCCTCTCCTCCCGCCTCGCCGTCCCCATCCTCTACGGCACCGACGCC  
GTCCACGGGCACAACAACGTCTTCGGCGCCACCGTCTTCCCTCACAACGT  
TGGCCTCGGAGCCTCCAGGTACCCAAACACCTGACGCCATATGACGCTCA  
TTGACCACGTGACGCACTTGACCGTACGCCGTTGACTTCCAGGGACGCG  
GAGCTTGTGCGTAAGATTGGCGAGGTGACGGCGCTCGAGGTCCGCGCCAC  
CGGCATACACTGGGCCTTTGCGCCCTGCCTCGCCGTGAGTTTAAATTCAA  
AAGGCTTCCGGATTA-----

-----  
-----  
-----GGCAGATCACCCCT  
CATGGTTACCCGTTCCCTTGCTTCGGTTAGGTAATTCCACTGGTGAGATCT  
AGCTGTGACTTCTGCAGACGCCACATCCTGTTTATTGAAATGGTTCTAT  
GTGATTTACAGGGAGAATGTGCTTGCTTGTGCCAAGCATTTCGTAGCGGA  
TGGTGGCACTGACAAGGGACTCAATGAGGGGAACACCATATGCTCGCCAG  
AAGATTTGGAAAGGATCCACATGAAACCTTACCCTGATTGTATAACTCAA  
GGGGTGGCAACGGTCATGGCGTCTACTCTAAGTGAATGGGGAGCCATT  
GCATGCCAGCCGCTATTTGCTCACAGATGTTTTAAAGGGCAAGCTAGGCT  
TCAAGGTAGTAGATCATGTGCTATTTGATTGATATCTTGATGCAAACCTC  
GGATCTGAATAGTGACTATGGCCTGTCAAAGAAGAATAGTGACTATGAG  
TTGCGAACATGTCATCATGTGTAAGCTCTTCTTTCAGGGTTTTGTGGTGT  
CAGACTGGGAGGGTATTGACAGGCTCTGTGAGCCTCGAGGGTCTGATTAT  
CGCTACTGCATTGCACAATCAGTTAATGCTGGAATGGATATGGTATTAGT  
TATTAAGTAGTCTTAAGATTTATTCTACAGCGACGGGTCTCTTGCATTT  
GCTGATCAATCTACAGATTATGATTCCCTCASAGATTTGAGAAATTCTTGG  
AAGATCTTGTGTTCTTGGTGGAGACAGGAGAGATAACCAATATCACGGATT  
GATGATGCTGTTGAGCGGATTCTAAGGGTTAAGTTCATNTCTGGAGTGTT  
TGAGCATCCATTTTCAGATCCGTCTCTGCTAGACATAGTCGGCTGTAAGG  
TAGCATACATTTTGCATATCTATATGTTTCTGCCGCTAATATTTACATCT  
TAATTGTCCCAAATTGCATATCAGTGATCATATACGATTACGCTAAATTG  
TTCAACTTTCAGGAGCATCGGCTGCTGGCCCGTGAGGCCGTTGAAAGTC  
TTTGGTACTTCTGAAAAATGGCAAGAATCAGAAGGAAACGTTCCCTCCAT  
TGGCCAAAAATGCAAAACGAATACTTGTGTCAGGGACACATGCTGACGAT  
ATTGGATATCAGTGCGGTGGGTGGACAATAGCTTGGCATGGAGACAGTGG  
AAAGATAACTCTTGGTAAGCAAAAACCTCGACTTTTCTTTCACATTACTTT  
TTT--TTGAGGGAGTACTTACCATTAGCTTGGTACTGAATAAACTGTTGC  
TCATCTAGAATTCTAGCTACCAAGTAGTTAGGGATAACATATGTCCTATT  
TTAAACTGCTATGATGCATACAGATGTGGAACCTATTGCTCCAAAGTTGAT



GTTTAAAATGTTCAATGGATTGTTCAAGTATTATGTTGTGGTTGTCTGTT  
TTTAGACTGATGGGTGAAGGCACTGCCTGGTTTGTACTGAAAGCCAAAAT  
AAAATATAACATACTGGATAGTAACTTACGAGTCTACATTCAAACAAG  
TGCAGAGATGCCACAATGCGAGCTATCATCACCAATTACTGAACTAAAA  
ATGTTGGTGAAAAATAAATTAAGCACATTTCCTTTTTCATTTCAAGCCAGA  
CATCACCCTTTTAATTAGATTTATTTGTTCTATTTATATAAAAAAGGTT  
TCATCAGGTAGAAGCAGTATAAAATCTGCACTCTCAACAATATTATGCTA  
AGAAACAGATATTTATGATATCTAAAGTTCTACACCAACAGGTACAAGCA  
TACTGGAGGCCATAACAAGAATCCGTCGGAGTGGAAGTGAAGTTGTGCAT  
GAGCAATGCCCAACAGATGCTATCATTGAAACTGGAGGATTTTCTTATGC  
AGTTGTTGTTGTTGGCGAGGTTCCCTTATGCTGAATGGTTGGGAGATAGAA  
CTGACCTTAGTATTCCTTTAATGGTTCCGACCTGATTACTCGGGTTGCT  
AGTAAAGTCCCTACACTAGTGATTGTCATATCTGGAAGGCCCTTGGTTAT  
TGATCAACAAGTAATGGAAAAGATAGATGCTCTAGTTGCTGCGTGGCTAC  
CTGGAAGTGAGGGCACCGAATTACTGATTGCATCTTTGGGGACCATGAT  
TTTGTGGCACATTGCCTGTAACGTGGTTTAGGTCTTCTGATCAATTGCC  
TATAAATATTGGAGATGCCAAATATGACCCCTTATTTCCTTTTGGATATG  
GRTTGAAAATGTTTAGAAGTGATAAAGGTTACCATAACCTCCTGAATTC  
TGCAGATATCCATCACACTGGCGGCCGCTCGAGC-----  
-----  
-----

>SF\_contig\_2\_@7/20/2016,\_9\_08\_AM  
-----GCATCAAGCTTGGTACCGAGCTCG-G-ATCCACTAGTAACGGCCG  
CCAGTGTGCTGGAATTCAGGTCGCTCGCTTCTTCTTACCCCATGGCGGCG  
GCGCAAGGCGAGGACGTGCCGCTGTACAAGGACGCGTCGGCGCCGGTGA  
GGCGCGCGTGC GCGACCTGCTGGGCCGGATGACGCTGCGGGAGAAGGCGG  
GCCAGATGGCCCAGATCGAGCGCACCGTCGTGTCGCCCCGCGCCCTCACG  
GAGCTCGCCGCCGCGCAGCGTCTCAACGCCGGCGGAAGCGCCCCCGCGA  
ACGCGCCTCCCCGCGCGACTGGGCCCGCATGGTCGACGACATGCAGCGCC  
TCGCCCTCTCCTCCCGCCTCGCCGTCCCCATCCTCTACGGCACCGACGCC

GTCCACGGGCACAACAACGTCTTCGGCGCCACCGTCTTCCCTCACAACGT  
TGGCCTCGGAGCCTCCAGGTACCCAAACACCTGACGCCATATGACGCTCA  
TTGACCACGTGACGCACTTGACCGTACGCCGTTGACTTCCAGGGACGCG  
GAGCTTGTGCGTAAGATTGGCGAGGTGACGGCGCTCGAGGTCCGCGCCAC  
CGGCATACTGGGCCTTTGCGCCCTGCCTCGCCGTGAGTTTAAATTCAA  
AAGGCTTCCGGATTAAC TGGCGGTGTCTATTTTGGGGTCCTTGATTGTGT  
G-TTTGCTCTGTCTGGCGGGTTATGTCCAGGTTTGTAGGGATCCGAGGTG  
GGGGAGATGTTACGAGAGCTACAGCGAGGACCCAGAGATCGTGCGCTCGT  
TCACCAACATTGTCACCGGCTTG CAGGGCCAGCCACCGGCAGATCACCT  
CATGGTTACCCGTTCCCTTGCTTCGGTTAGGTAATTCCACTGGTGAGATCT  
AGCTGTGACTTCTGCAGACGCCACATCCTGTTTATTGAAATGGTTCTAT  
GTGATTTACAGGGAGAATGTGCTTGCTTGTGCCAAGCATTTTCGTAGCGGA  
TGGTGGCACTGACAAGGGACTCAATGAGGGGAACACCATATGCTCGCCAG  
AAGATTTGAAAGGATCCACATGAAACCTTACCCTGATTGTATAACTCAA  
GGGGTGGCAACGGTCATGGCGTCCTACTCTAAGTGAATGGGGAGCCATT  
GCATGCCAGCCGCTATTTGCTCACAGATGTTTTAAAGGGCAAGCTAGGCT  
TCAAGGTAGTAGATCATGTGCTATTTGATTGATATCTTGATGCAAACCTC  
GGATCTGAATAGTGACTATGGCCTGTCAAAGAAGAATAGTGACTATGAG  
TTGCGAACATGTCATCATGTGTAAGCTCTTCTTTCAGGGTTTTGTGGTGT  
CAGACTGGGAGGGTATTGACAGGCTCTGTGAGCCTCGAGGGTCTGATTAT  
CGCTACTGCATTGCACAATCAGTTAATGCTGGAATGGATATGGTATTAGT  
TATTAAGTAGTCTTAAGATTTATTCTACAGCGACGGGTCTCTTGCATTT  
GCTGATCAATCTACAGATTATGATTCCCTCAGAGATTTGAGAAATCTTGG  
AAGATCTTGTGTTCTTGGTGGAGACAGGAGAGATACCAATATCACGGATT  
GATGATGCTGTTGAGCGGATTCTAAGGGTTAAGTTCATATCTGGAGTGTT  
TGAGCATCCATTTTCAGATCCGTCTCTGCTAGACATAGTCGGCTGTAAGG  
TAGCATACATTTTGCATATCTATATGTTTCTGCCGCTAATATTTACATCT  
TAATTGTCCCAAATTCATATCAGTGATCATATACGATTACGCTAAATTG  
TTCAACTTTCAGGAGCATCGGCTGCTGGCCCGTGAGGCCGTTTCGAAAGTC  
TTTGGTACTTCTGAAAAATGGCAAGAATCAGAAGGAAACGTTCCCTTCCAT

TGGCCAAAATGCAAAACGAATACTTGTTCAGGGACACATGCTGACGAT  
ATTGGATATCAGTGCGGTGGGTGGACAATAGCTTGGCATGGAGACAGTGG  
AAAGATAACTCTTGGTAAGCAAAAACCTCGACTTTTCTTTCACATTACTTT  
TTT--TTGAGGGAGTACTTACCATTAGCTTGGTACTGAATAAACTGTTGC  
TCATCTAGAATTCTAGCTACCAAGTAGTTAGGGATAACATATGTCCTATT  
TTAAACTGCTATGATGCATACAGATGTGGAAC TATTGCTCCAAAGTTGAT  
GTTTAAAATGTTCAATGGATTGTTCAAGTATTATGTTGTGGTTGTCTGTT  
TTTAGACTGATGGGTGAAGGCACTGCCTGGTTTGTACTGAAAGCCAAAAT  
AAAATATAACATACACTGGATAGTAACTTACGAGTCTACATTCAAACAAG  
TGCAGAGATGCCACAATGCGAGCTATCATCACCAATTACTGAACTAAAA  
ATGTTGGTGAAAAATAAATTAAGCACATTTCTTTTCATTTCAAGCCAGA  
CATCACCCTTTTAATTAGATTTATTTGTTCTATTTATATAAAAAAGGTT  
TCATCAGGTAGAAGCAGTATAAAATCTGCACTCTCAACAATATTATGCTA  
AGAAACAGATATTTATGATATCTAAAGTTCTACACCAACAGGTACAAGCA  
TACTGGAGGCCATACAAGAATCCGTCGGAGTGGAAC TGAAGTTGTGCAT  
GAGCAATGCCCAACAGATGCTATCATTGAAACTGGAGGATTTTCTTATGC  
AGTTGTTGTTGTTGGCGAGGTTCTTATGCTGAATGGTTGGGAGATAGAA  
CTGACCTTAGTATTTCCCTTTAATGGTTCCGACCTGATTACTCGGGTTGCT  
AGTAAAGTCCCTACACTAGTGATTGTCATATCTGGAAGGCCCTTGGTTAT  
TGATCAACAAGTAATGGAAAAGATAGATGCTCTAGTTGCTGCGTGGCTAC  
CTGGAAGTGAGGGCACCGGAATTACTGATTGCATCTTTGGGGACCATGAT  
TTTGTGGCACATTGCCTGTAACGTGGTTTAGGTCTTCTGATCAATTGCC  
TATAAATATTGGAGATGCCAAATATGACCCCTTATTTCTTTTGGATATG  
GGTTGAAAATGTTTAGAAGTGATAAAGGTTACCCATAACCTCCTGAATTC  
TGCAGATATCCATCACACTGGCGGCCGCT-CGAGCATGCATCTAGA----

-----  
----

>HF\_contig\_3\_@7/20/2016,\_9\_16\_AM  
-----GCATGCT--CGAG--CG-G---CCGCCAGT-GTGATGG  
ATA-TCTGC-AGAATTCNNGTCGCTCGCTTCTTCTTACCCCATGGCGGCG

GCGCAAGGCGAGGACGTGCCGCTGTACAAGGACGCGTCGGCGCCGGTGGAGGCGCGGTGCGCGACCTGCTGGGCCGGATGACGCTGCGGGAGAAGGCGG  
GCCAGATGGCCCAGATCGAGCGCACCGTCGTGTCGCCCCGCGCCCTCACG  
GAGCTCGCCGCCGGCAGCGTCCTCAACGCCGGCGGAAGCGCCCCCGCGA  
ACGCGCCTCCCCGGCCGACTGGGCCCGCATGGTCGACGACATGCAGCGCC  
TCGCCCTCTCCTCCCGCCTCGCCGTCCCCATCCTCTACGGCACCGACGCC  
GTCCACGGGCACAACAACGTCTTCGGCGCCACCGTCTTCCCTCACAACGT  
TGGCCTCGGAGCCTCCAGGTACCCAATCACCTGACGCCATATGACGCTCA  
TTGACCACGTGACGCACTTGACCGTACGCCGTTGACTTCCAGGGACGCG  
GAGCTTGTGCGTAAGATTGGCGAGGTGACGGCGCTCGAGGTCCGCGCCAC  
CGGCATACACTGGGCCTTTGCGCCCTGCCTCGCCGTGAGTTTAAATTCAA  
AAGGCTTCCGGATTAAC TGGCGGTGTCTATTTTGGGGTCCTTGATTGTGT  
GTTTTGCTCTGTCTGGCGGGTTATGTCCAGGTT-----  
-----  
-----ACATTGTCACCGGCTTGCAGGGCCAGCCACCGGCAGATCACCT  
CATGGTTACCCGTTCCCTTGCTTCGGTTAGGTAATTCCACTGGTGAGATCT  
AGCTGTGACTTCTGCAGACGCCACATCCTGTTTATTGAAATGGTTCTAT  
GTGATTTACAGGGAGAATGTGCTTGCTTGTGCCAAGCATTTTCGTAGCGGA  
TGGTGGCACTGACAAGGACTCAATGAGGGGAACACCATATGCTCGCCAG  
AAGATTTGGAAAGGATCCACATGAAACCTTACCCTGATTGTATAACTCAA  
GGGTTGGCAACGGTCATGGCGTCCTACTCTAAGTGGAATGGGGAGCCATT  
GCATGCCAGCCGCTATTTGCTCACAGATGTTTTAAAGGGCAAGCTAGGCT  
TCAAGGTAGTAGATCATGTGCTATTTGATTGATATCTTGATGCAAACCTC  
GGATCTGAATAGTGA CTATGGCCTGTCAAAGAAGAATAGTGA CTATGAG  
TTGCGAACATGTCATCATGTGTAAGCTCTTCTTTCAGGGTTTTGTGGTGT  
CAGACTGGGAGGGTATTGACAGGCTCTGTGAGCCTCGAGGGTCTGATTAT  
CGCTACTGCATTGCACAATCAGTTAATGCTGGAATGGATATGGTATTAGT  
TATTAAGTAGTCTTAAGATTTATTCTACAGCGACGGGTTCTTTCGATTT  
GCTGATCAATCTACAGATTATGATTCCCTCACAGATTTGAGAAATTCTTGG  
AAGATCTTGTGTTCTTGGTGGAGACAGGAGAGATAACCAATATCACGGATT

GATGATGCTGTTGAGCGGATTCTAAGGGTTAAGTTCATATCTGGAGTGTT  
TGAGCATCCATTTTCAGATCCGTCTCTGCTAGACATAGTCGGCTGTAAGG  
TAGCATAACATTTTGCATATCTATATGTTTCTGCCGCTAATATTTACATCT  
TAATTGTCCCAAATTGCATATCAGTGATCATATACGATTACGCTAAATTG  
TTCAACTTTCAGGAGCATCGGCTGCTGGCCCGTGAGGCCGTTTCGAAAGTC  
TTTGGTACTTCTGAAAAATGGCAAGAATCAGAAGGAAACGTTTCCTTCCAT  
TGGCCAAAAATGCAAAACGAATACTTGTGTCAGGGACACATGCTGACGAT  
ATTGGATATCAGTGCGGTGGGTGGACAATAGCTTGGCATGGAGACAGTGG  
AAAGATAACTCTTGGTAAGCAAAAACTCGACTTTTCTTTCACATTACTTT  
TTT--TTGAGGGAGTACTTACCATTAGCTTGGTACTGAATAAACTGTTGC  
TCATCTAGAATTCTAGCTACCAAGTAGTTAGGGATAACATATGTCCTATT  
TTAAACTGCTATGATGCATACAGATGTGGAAC TATTGCTCCAAAGTTGAT  
GTTTAAAATGTTCAATGGATTGTTCAAGTATTATGTTGTGGTTGTCTGTT  
TTTAGACTGATGGGTGAAGGCACTGCCTGGTTTGTACTGAAAGCCAAAAT  
AAAATATAACATACACTGGATAGTAACTTACGAGTCTACATTCAAACAAG  
TGCAGAGATGCCACAATGCGAGCTATCATCACCAATTACTGAACTAAAAA  
ATGTTGGTGAAAAATAAATTAAGCACATTTCTTTTTCATTTCAAGCCAGA  
CATCACCCTTTTAATTAGATTTATTTGTTCTATTTATATAAAAAAGGTT  
TCATCAGGTAGAAGCAGTATAAAATCTGCACTCTCAACAATATTATGCTA  
AGAAACAGATATTTATGATATCTAAAGTTCTACACCAACAGGTACAAGCA  
TACTGGAGGCCATACAAGAATCCGTCCGAGTGGAAACTGAAGTTGTGCAT  
GAGCAATGCCCAACAGATGCTATCATTGAAACTGGAGGATTTTCTTATGC  
AGTTGTTGTTGTTGGCGAGGTTCCCTTATGCTGAATGGTTGGGAGATAGAA  
CTGACCTTAGTATTCCCTTTAATGGTTCCGACCTGATTACTCGGGTTGCT  
AGTAAAGTCCCTACACTAGTGATTGTCATATCTGGAAGGCCCTTGGTTAT  
TGATCAACAAGTAATGGAAAAGATAGATGCTCTAGTTGCTGCGTGGCTAC  
CTGGAAGTGAGGGCACCGGAATTACTGATTGCATCTTTGGGGACCATGAT  
TTTGTGGCACATTGCCGTGAACGTGGTTTAGGTCTTCTGATCAATTGCC  
TATAAATATTGGAGATGCCAAATATGACCCCTTATTTCTTTTGGATATG  
GGTTGAAAATGTTTAGAAGTGATAAAGGTTACACATAACCTCCTGAATTC

-----CAGCACACTGGCGGCCGTTACTAGTGG-ATC-CGAGCTCG  
G-T-ACCA-AGCTTGATGCATAGC-----  
----

>SF\_contig\_3\_@7/20/2016,\_9\_14\_AM

GCTATGCATCAAGCTTGGTACCGAGCTCG-G-ATCCACTAGTAACGGCCG  
CCAGTGTGCTGGAATTCAGGTCGCTCGCTTCTTCTTACCCCATGGCGGCG  
GCGCAAGGCGAGGACGTGCCGCTGTACAAGGACGCGTCGGCGCCGGTGA  
GGCGCGCGTGC GCGACCTGCTGGGCCGGATGACGCTGCGGGAGAAGGCGG  
GCCAGATGGCCCAGATCGAGCGCACCGTCGTGTCGCCCCGCGCCCTCACG  
GAGCTCGCCGCCGGCAGCGTCTCAACGCCGGCGGAAGCGCCCCCGCGA  
ACGCGCCTCCCCGGCCGACTGGGCCCGCATGGTCGACGACATGCAGCGCC  
TCGCCCTCTCCTCCCGCCTCGCCGTCCCCATCCTCTACGGCACCGACGCC  
GTCCACGGGCACAACAACGTCTTCGGCGCCACCGTCTTCCCTCACAACGT  
TGGCCTCGGAGCCTCCAGGTACCCAATCACCTGACGCCATATGACGCTCA  
TTGACCACGTGACGCACTTGACCGTACGCCGTTGACTTCCAGGGACGCG  
GAGCTTGTGCGTAAGATTGGCGAGGTGACGGCGCTCGAGGTCCGCGCCAC  
CGGCATACACTGGGCCTTTGCGCCCTGCCTCGCCGTGAGTTTAAATTCAA  
AAGGCTTCCGGATTAAC TGGCGGTGTCTATTTTGGGGTCCTTGATTGTGT  
G-TTTGCTCTGTCTGGCGGGTTATGTCCAGGTTTGTAGGGATCCGAGGTG  
GGGGAGATGTTACGAGAGCTACAGCGAGGACCCAGAGATCGTGC GCTCGT  
TCACCAACATTGTCACCGGCTTGCAGGGCCAGCCACCRGCAGATCACCT  
CATGGTTACCCGTTCCCTTGCTTCGGTTAGGTAATTCCACTGGTGAGATCT  
AGCTGTGACTTCTGCAGACGCCACATCCTGTTTATTGAAATGGTTCTAT  
GTGATTTACAGGGAGAATGTGCTTGCTTGTGCCAAGCATTTCGTAGCGGA  
TGGTGGCACTGACAAGGACTCAATGAGGGGAACACCATATGCTCGCCAG  
AAGATTTGGAAAGGATCCACATGAAACCTTACCCTGATTGTATAACTCAA  
GGGGTGGCAACGGTCATGGCGTCTACTCTAAGTGGAATGGGGAGCCATT  
GCATGCCAGCCGCTATTTGCTTACGGATGTTTTAAAGGGCAAGCTAGGCT  
TCAAGGTAGTAGAGCATGTGCTATTTGATTGATATCTTGATGCAAACCTC  
GGATCTGAATAGTACTATGGCCTGTCAAAAAAAGAATAGTACTATGRG

TTGCGAACATGTCATCATGTGTAAGCTCTTCTTTCAGGGTTTTGTGGTGT  
CAGACTGGGAGGGTATTGACAGGCTCTGTGAGCCTCGAGGGTCTGATTAT  
CGCTACTGCATTGCACAATCAGTTAATGCTGGAATGGATATGGTATTAGT  
TATTAAGTAGTCTTAAGATTTATTCTACAGCGACRGGTCTCTTGCATTT  
GCTGATCAATCTACAGATTATGATTCCCTCACAGATTTGAGAAATTCTTGG  
AAGATCTTGTGTTCTTGGTGGAGACAGGAGAGATACCAATATCACGGATT  
GATGATGCTGTTGAGCGGATTCTAAGGGTTAAGTTCATATCTGGAGTGTT  
TGAGCATCCATTTTCAGATCCGTCTCTGCTAGACATAGTCGGCTGTAAGG  
TAGCATAcATTTTGCATATCTATATGTTTCTGCCGCTAATATTTACATCT  
TAATTGTCCCAAATTGCATATCAGTGATCATATACGATTACGCTAAATTG  
TTCAACTTTCAGGAGCATCGGCTGCTGGCCCGTGAGGCCGTTCGAAAGTC  
TTTGGTACTTCTGAAAAATGGCAAGAATCAGAAGGAAACGTTCTTCCAT  
TGGCCAAAAATGCAAAACGAATACTTGTGTCAGGGACACATGCTGACGAT  
ATTGGATATCAGTGCCGGTGGGTGGACAATAGCTTGGCATGGAGACAGTGG  
AAAGATAACTCTTGGTAAGCAAAAACCTCGACTTTTCTTTCACATTACTTT  
TTT--TTGAGGGAGTACTTACCATTAGCTTGGTACTGAATAAACTGTTGC  
TCATCTAGAATTCTAGCTACCAAGTAGTTAGGGATAACATATGTCCTATT  
TTAAACTGCTATGATGCATACAGATGTGGAAC TATTGCTCCAAAGTTGAT  
GTTTAAAATGTTCAATGGATTGTTCAAGTATTATGTTGTGGTTGTCTGTT  
TTTAGACTGATGGGTGAAGGCACTGCCTGGTTTGTACTGAAAGCCAAAAT  
AAAATATAACATACACTGGATAGTAACTTACGAGTCTACATTCAAACAAG  
TGCAGAGATGCCACAATGCGAGCTATCATCACCAATTACTGAACTAAAAA  
ATGTTGGTGAAAAATAAATTAAGCACATTTCTTTTTCATTTCAAGCCAGA  
CATCACCCTTTTAATTAGATTTATTTGTTCTATTTATATAAAAAAGGTT  
TCATCAGGTAGAAGCAGTATAAAATCTGCACTCTCAACAATATTATGCTA  
AGAAACAGATATTTATGATATCTAAAGTTCTACACCAACAGGTACAAGCA  
TACTGGAGGCCATACAAGAATCCGTCCGAGTGGAAACTGAAGTTGTGCAT  
GAGCAATGCCCAACAGATGCTATCATTGAAACTGGAGGATTTTCTTATGC  
AGTTGTTGTTGTTGGCGAGGTTCTTATGCTGAATGGTTGGGAGATAGAA  
CTGACCTTAGTATTCCTTTAATGGTTCCGACCTGATTACTCGGGTTGCT

AGTAAAGTCCCTACACTAGTGATTGTCATATCTGGAAGGCCCTTGGTTAT  
TGATCCACAAGTACTGGAAAAGATAGATGCTCTAGTTGCTGCGTGGCTAC  
CTGGAAGTGAGGGCAMECGGAATTACTGATTGCATCTTTGGGGACCATGAT  
TTTGTGGCACATTGCCTGTAACGTGGTTTAGGTTTTCTGATCAATTGCC  
TATAAATATTGGAGATGCCAAATATGACCCCTTATTCCTTTTGGATATG  
GGTTGAAAATGTTTAGAAGTGATAAAGGTTACACCATAACCTCCTGAATTC  
TGCAGATATCCATCACACTGGCGGCCGCT-CGAGCATGCATCT-----  
-----  
-----