Identification and Manipulation of Starch Synthases from Marama Bean and Potato

Lionel Hartzenberg

Thesis presented in partial fulfilment of the requirements for the degree of Master of Science in Plant Biotechnology in the Faculty of Natural Sciences at Stellenbosch University

Supervisor: Dr James R Lloyd

Co-Supervisor: Prof. J. Kossmann

December 2016

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: 17-08-2016

Abstract

Starch is a major storage carbohydrate in plants and algae. It consist of amylose and amylopectin and provides energy during heterotrophic growth. The accumulation of starch differs between certain families. Viridiplantae (plants and green algae) utilize ADP-glucose for starch synthesis and store it within the chloroplast, whereas Rhodophyceae (red algae) utilize UDP-glucose and both synthesize and store starch in the cytoplasm. Starch is a commercial product of worldwide importance and research into improving current starch sources and examining alternatives may provide benefits to the starch industry. This study consisted of two parts. The first involved increasing the physical and genetic information available for the southern African legume *Tylosema esculentum*. This involved obtaining and analysing transcriptomic data in order to find genes relating to starch synthesis. The second part aimed to randomly mutagenize a *Solanum tuberosum* SSI gene for the utilization of UDP-glucose instead of ADP-glucose for starch synthesis.

Tylosema esculentum plants were grown for approximately one year before the leaf and tuber material was harvested. RNA extraction and cDNA synthesis was performed and the samples were sent for next-generation sequencing at the Agricultural Research Council (Pretoria). Reads were compiled and trimmed to produce three contigs related to starch metabolism. Using the Phytozome soybean database, we selected three genes (granule-bound starch synthase (GBSS), starch synthase 2 (SSII), starch synthase 4 (SS4)), and soybean-like actin, relating to starch synthesis within soybean. We matched them to open reading frames within the marama bean transcriptome and designed primers for their amplification from cDNA. All were successfully amplified.

We also obtained additional information on the physical properties of *T. esculentum* starch granules. Two methods were performed in order to accurately obtain data on the marama bean starch granule size distribution. Both showed that the starch granule size is predominantly between 15-35 µm, with no particles above 50 µm. A reducing end assay was also performed to determine the amylose and amylopectin average chain length of marama bean starch. The amylose chain was found to be 67.5% longer than that of harvested potato starch, and 31.8% longer than that of

commercial grade potato starch (Sigma-Aldrich). The amylopectin chain length was found to be 34.8% lower than that of harvested potato starch, but similar to that of commercial grade potato starch. The higher accumulation of long amylose chains could account for previous Rapid Visco Analyser (RVA) analyses performed that showed marama bean starch having unique viscosity properties, and may in turn lead to several other novel uses for marama bean.

In the second part of this study, we amplified the UGP*ase* (UDP-glucose pyrophosphorylase) gene from *Arabidopsis thaliana* to facilitate the production of UDP-glucose. The UGP*ase* gene was ligated into the pACYC 184 vector and transformed into G6MD2, an *E. coli* strain lacking the *glg* operon necessary for glycogen metabolism. The *S. tuberosum* SSI gene within pBluescript.SK was randomly mutagenized using XL1-red *E. coli* cells, which are deficient in 3 of the primary proof reading mechanisms. Of these mutagenized plasmids, 33 were pooled and analysed against the unaltered SSI gene as template. We obtained 192 mutations in total, 4 per kb, with 98 point mutations, 57 insertions and 37 deletions. The *SSI* library was transformed into G6MD2::pACUG and screened for activity via iodine vapour. 150 000 colonies were screened but no expressing colonies were found. In the future alternative mutation methods as well as larger colony screenings may yield better results.

Opsomming

Stysel is 'n groot stoor-koolhidraat in plante en alge. Dit bestaan uit amilose en amilopektien en verskaf energie tydens heterotrofiese groei. Die opeenhoping van stysel verskil tussen sekere families. Viridiplantae (plante en groen alge) gebruik ADP-glukose vir stysel sintese en stoor dit in die chloroplast, terwyl Rhodophyceae (rooi alge) UDP-glukose gebruik en albei sintetiseer en stoor stysel in die sitoplasma. Stysel is 'n kommersiële produk met wêreldwye belang en navorsing in die verbetering van die huidige styselbronne en die ondersoek van alternatiewes kan voordele inhou vir die styselbedryf. Hierdie studie het bestaan uit twee dele. Die eerste deel omvat die toename in die fisiese en genetiese inligting wat beskikbaar is vir die Suid-Afrikaanse peulplant *Tylosema esculentum*. Dit behels die verkryging en ontleding van transkriptomiese data om gene wat verband hou met stysel sintese op te spoor. Die tweede deel was gemik om ‘n *Solanum tuberosum* SSI gene lukraak te muteer vir die benutting van UDP-glukose in plaas van ADP-glukose vir stysel sintese.

Tylosema esculentum plante was gekweek vir ongeveer een jaar voordat die blaar en knol materiaal geoes. RNS ekstraksie en cDNA sintese was uitgevoer en die monsters is gestuur vir volgende-generasie volgorde-bepaling by die Landbounavorsingsraad (Pretoria). Reekse is saamgestel en afgewerk om drie samestellings te produseer wat verwant is aan stysel metabolisme. Deur gebruik te maak van die Phytozome sojaboon databasis, het ons drie gene gekies (granule-gebinde stysel sintase (GBSS), stysel sintase 2 (SSII), stysel sintase 4 (SS4)), asook sojaboon-agtige aktien, verwant aan stysel sintese binne sojabone. Ons het hulle ooreengestem met oop leesrame binne die marama boontjie transkriptoom en inleiers is ontwerp om hulle te amplifiseer vanaf kDNS. Almal was suksesvol geamplifiseer.

Ons het ook bykomende inligting verkry oor die fisiese eienskappe van *T. esculentum* styselgranules. Twee metodes is uitgevoer om die data op die marama boontjie styselgranule grootte-verspreiding akkuraat te bekom. Beide het getoon dat die styselgranule-grootte is hoofsaaklik tussen 15-35 µm, met geen deeltjies bo 50

µm. ‘n Reduserende endtoets is ook uitgevoer om die amilose en amilopektien gemiddelde kettinglengte van marama boontjie stysel te bepaal. Die amilose ketting is gevind om 67,5% langer as dié van geoeste aartappelstysel en 31,8% langer as dié van kommersiële graad aartappelstysel (Sigma-Aldrich) te wees. Die amilopektien kettinglengte is gevind om 34,8% laer as dié van geoeste aartappelstysel te wees, maar soortgelyk aan dié van kommersiële graad aartappelstysel. Die hoër opeenhoping van lang amilosekettings kan van rekenskap gee vir die vorige RVA ontledings wat uitgevoer was en aangedui het dat marama boontjie stysel unieke viskositeit-eienskappe het, en kan op sy beurt lei tot verskeie ander nuwe gebruike vir die marama boontjie.

In die tweede deel van hierdie studie, het ons die UGP*ase* (UDP-glukose pyrophosphorylase) geen van *Arabidopsis thaliana* geamplifiseer om die produksie van UDP-glukose te fasiliteer. Die UGP*ase* geen is afgesnoer in die pACYC 184 vektor en omskep in G6MD2, 'n *E. coli* stam sonder die glg operon wat nodig is vir glikogeen metabolisme. Die *S. tuberosum* SSI gene binne pBluescript.SK is lukraak muteer met behulp van XL1-rooi *E. coli* selle, wat gebrekkig is in 3 van die primêre proeflees meganismes. Van hierdie muteerde plasmiede, is 33 saamgevoeg en teen die onveranderde SSI gene as sjabloon ontleed. Ons het 192 mutasies in totaal verkry, 4 per kb, met 98 puntmutasies , 57 invoegings en 37 skrappings. Die SSI biblioteek is omskep in G6MD2::pACUG en gesif vir aktiwiteit via jodium dampe. 150 000 kolonies is ondersoek, maar geen uitdrukking kolonies is gevind nie. In die toekoms kan alternatiewe metodes van mutasie asook groter kolonie siftings beter resultate oplewer.

Acknowledgements

I would like to thank my supervisor, Dr James Lloyd, for the patience and guidance shown to me. Thank you for your support and your editing skills.

I would also like to thank Professor Jens Kossmann (co-supervisor) for the help and support shown to me. I shall never cease to be grateful.

To Dr Shaun Peters, Dr Paul Hills, Dr Christell van der Vyver, and Marnus Smith, your insight and input through the years transformed me into the scientist I am today. Thank you.

Special thanks to Prof Percy Chimwamurombe and Dr Emmanuel Nepolo at the University of Namibia for providing the marama bean seeds and the past information that allowed me to take on this project.

I would like to thank Dr Hano Maree at the Department of Genetics for the swift and in-depth transcriptomic analysis provided.

To the IPB staff and students, there are too many to name, but every one of you is responsible for me being here today. If not for the friendships, the discussions, the laughter and the tears shed together, I would not be where I am today. Thank you for everything.

I would like to thank the IPB, the NRF and Stellenbosch University for the financial assistance I received during my time here.

Lastly, to my family. Thank you to each and every one of you. The love and support you showed me while being far away from home was irreplaceable.

List of Abbreviations

3-PGA	Glycerate-3-phosphate
ADP	Adenosine diphosphate
ADP-glucose	Adenosine diphosphate glucose
AGP*ase*	Adenosine diphosphate glucose pyrophosphorylase
AMY	α-amylases
ATP	Adenosine triphosphate
BAM	β-amylases
BLAST	Basic Local Alignment Search Tool
Bp	Base pair
cDNA	Complementary deoxyribonucleic acid
cP	Viscosity
CTAB	Cetyl trimethylammonium bromide
DAPA	2, 6-Diaminopimelic acid
DBE	Debranching enzyme
DNS	Dinitrosalicylic acid
DPE	Disproportionating enzymes
EDTA	Ethylenediaminetetraacetic acid
G1P	Glucose-1-phosphate
G6P-DH	Glucose-6-phosphate dehydrogenase
GBSS	Granule-bound starch synthase
GWD	Glucan, water dikinase
ISA	Isoamylase
Km	Michaelis Menten constant
LB	Lysogeny broth
LDA	Limit dextrinases
LiCl	Lithium chloride
$MgCl_2$	Magnesium chloride
NaAC	Sodium acetate
NaCl	Sodium chloride
NAD:	Nicotinamide adenine dinucleotide (oxidised)
NADH:	Nicotinamide adenine dinucleotide (reduced)
NaOH	Sodium hydroxide

NDP-glucose	Nucleoside diphosphate glucose
NMR	Nuclear Magnetic Resonance
OD_{540}	Optical Density at 540 nanometres
PCR	Polymerase chain reaction
PGM	Phosphoglucomutase
PHS	α-glucan phosphorylases
Pi	Inorganic phosphate
PPi	Inorganic pyrophosphate
PVP	Polyvinylpyrrolidone
PWD	Phosphoglucan, water dikinase
RNA	Ribonucleic acid
RPM	Rotations per minute
RVA	Rapid Visco Analyser
SBE	Starch branching enzyme
SdFFF	Sedimentation field-flow fractionation
SEM	Scanning Electron Microscope
SS	Starch synthase
SSR	Single Sequence Repeats
TB	Terrific broth
TBE	Tris (hydroxymethyl) aminomethane borate EDTA
Tris-HCL	Tris (hydroxymethyl) aminomethane hydrochloric acid
UDP	Uridine diphosphate
UDP-glucose	Uridine diphosphate glucose
ε_{340}	Extinction coefficient at 340 nanometres

List of Figures

List of Tables

1. Introduction and Literature Review

Starch is a glycogen analogue produced by land plants and some algal species and functions as a storage polysaccharide (Viola et al. 2001). It is a major carbohydrate of the human diet contained within many staple foods such as potato, maize and rice. Starch and its derivatives are also of worldwide industrial importance, being utilized for a variety of products including papers, textiles and gelling agents (Keeling & Myers 2010). In 2012, worldwide sales of starch-based products were reported to be $51.2 billion and are predicted to rise to $77.4 billion within the next 4 years (prweb.com/releases/2013/7/prweb10923341.htm). This increase is partly due to the rise in the human population, which is currently over 7 billion and is expected to reach roughly 10 billion by the year 2050, creating a higher demand for starch and starch-based products (http://www.worldometers.info/world-population). Starch from different plant species is also a major source of renewable energy in the form of biofuels (Keeling & Myers 2010; Morell et al. 1995). In recent years, improvements in the fields of genomic and transcriptomic analysis has increased our understanding of the biology surrounding starch synthesis and its associated genes (Stitt & Zeeman 2012; Keeling & Myers 2010; Tiessen et al. 2002). This has motivated research into the improvement of starch yield within industrially important cereal crops. These include both the up and down-regulation of genes involved in starch synthesis and degradation as well as research into novel plant sources of starch (Holse et al. 2011; Cannon et al. 2009). In order to accomplish these feats, it is important to have a good understanding of the mechanisms of starch metabolism, the enzymes involved, their evolution and how current technology can aid us in expanding what we know presently.

1.1 Starch Metabolism

Starch is comprised of two polyglucans, amylose and amylopectin. Amylose consists of linear glucose polymers linked by α-1, 4-bonds, whereas amylopectin consists of α-1, 4-bonds and α-1, 6-branches (Fig. 1) (Dauvillée et al. 2009). There are four primary enzymes involved in the production of these two polymers: ADP-glucose pyrophosphorylase (AGP*ase*), starch synthases (SS), starch branching enzymes

(SBE) and starch debranching enzymes (DBE) (Deschamps et al. 2006). Secondary enzymes involved include α-glucan phosphorylases (PHS), disproportionating enzymes (DPE) and α-amylases and β-amylases (AMY & BAM) (Kötting et al. 2010). During the first step of starch biosynthesis, AGP*ase* catalyses the formation of ADP-glucose and pyrophosphate from glucose-1-phosphate and ATP. Starch synthase and branching enzymes utilize the ADP-glucose in tandem with debranching enzymes to form amylose and amylopectin, which make up the starch granule (Zeeman et al. 2010). Synthesized starch is stored in the chloroplasts of leaves, and in amyloplasts in storage organs such as potato tubers (Kötting et al. 2010; Zeeman et al. 2007). To further understand starch metabolism, it is necessary to analyse the effect that each enzyme has on starch synthesis and degradation at each step in the pathway.

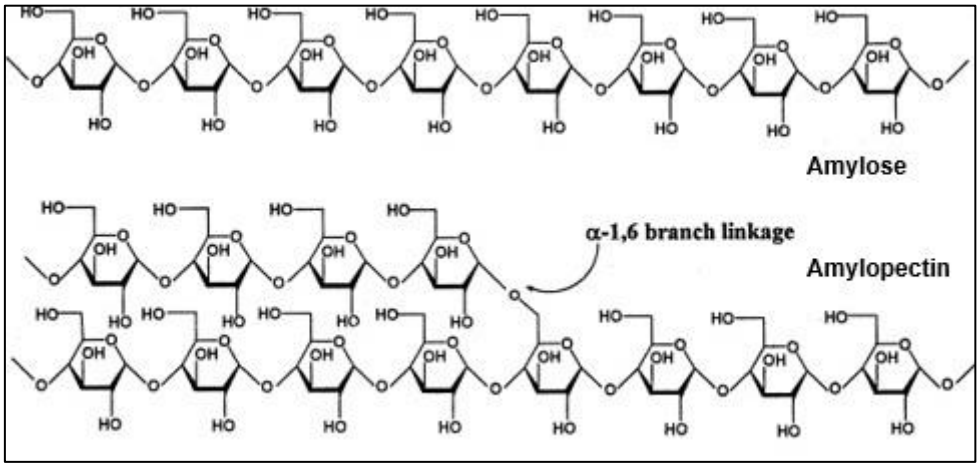

Figure 1: Structure of amylose and amylopectin. (Modified from http://biology.hi7.co/-9--carbohydrates---polysaccharides-56ce2a9717376.html)

1.2 Starch Synthesis

ADP-glucose pyrophosphorylase (AGP*ase*) is the first enzyme in the committed pathway of starch synthesis (Fig. 2) (Kötting et al. 2010; Zeeman et al. 2010). In plants it is a heterotetramer consisting of two large and two small subunits, the alteration of which can have varying effects on starch synthesis and seed formation. Hydroxylamine mutagenesis of a large subunit resulted in a higher affinity for

AGP*ase* subunit dimerization, a lower propensity for disassociation and higher enzymatic heat resistance (Greene & Hannah 1998), with a consequent increased sugar content and seed biomass in wheat (Smidansky et al. 2007). The introduction of an unregulated AGP*ase* gene from *E. col* in rice has also led to a similar increase in biomass (Sakulsingharoj et al. 2004). Conversely, a barley mutant lacking the cytosolic AGP*ase* small subunit led to reduced starch content, despite the presence of a functional plastidial AGP*ase* (Johnson et al. 2003).

AGP*ase* activity is regulated by three known mechanisms: The first is transcriptional regulation, with sugars increasing expression and nitrates and phosphates decreasing expression of the genes encoding different subunits. This allows for regulation of starch synthesis based on the nutritional status of the plant (Tiemann et al. 2002; Kossman et al. 1999). The second form of control is allosteric regulation, with glycerate-3-phosphate (3-PGA) acting as an activator and Pi as an inhibitor (Boehlein et al. 2010; Kötting et al. 2010). The balance between the rates of sucrose synthesis and photosynthesis alters the ratio of 3PGA/Pi within chloroplasts, which, in turn, alters the activation levels of AGP*ase*. The third is redox regulation. It has been shown that the detachment of potato tuber from the main plant leads to a loss of sucrose production in the plant, leading to AGP*ase* dimerization and increasing its sensitivity to 3PGA and Pi (Tiessen et al. 2002). The alternation of a regulatory mechanism of AGP*ase* results in varying degrees of starch synthesis alteration and coupled with further mutations within the synthesis pathway, may result in more predictable alterations in starch synthesis.

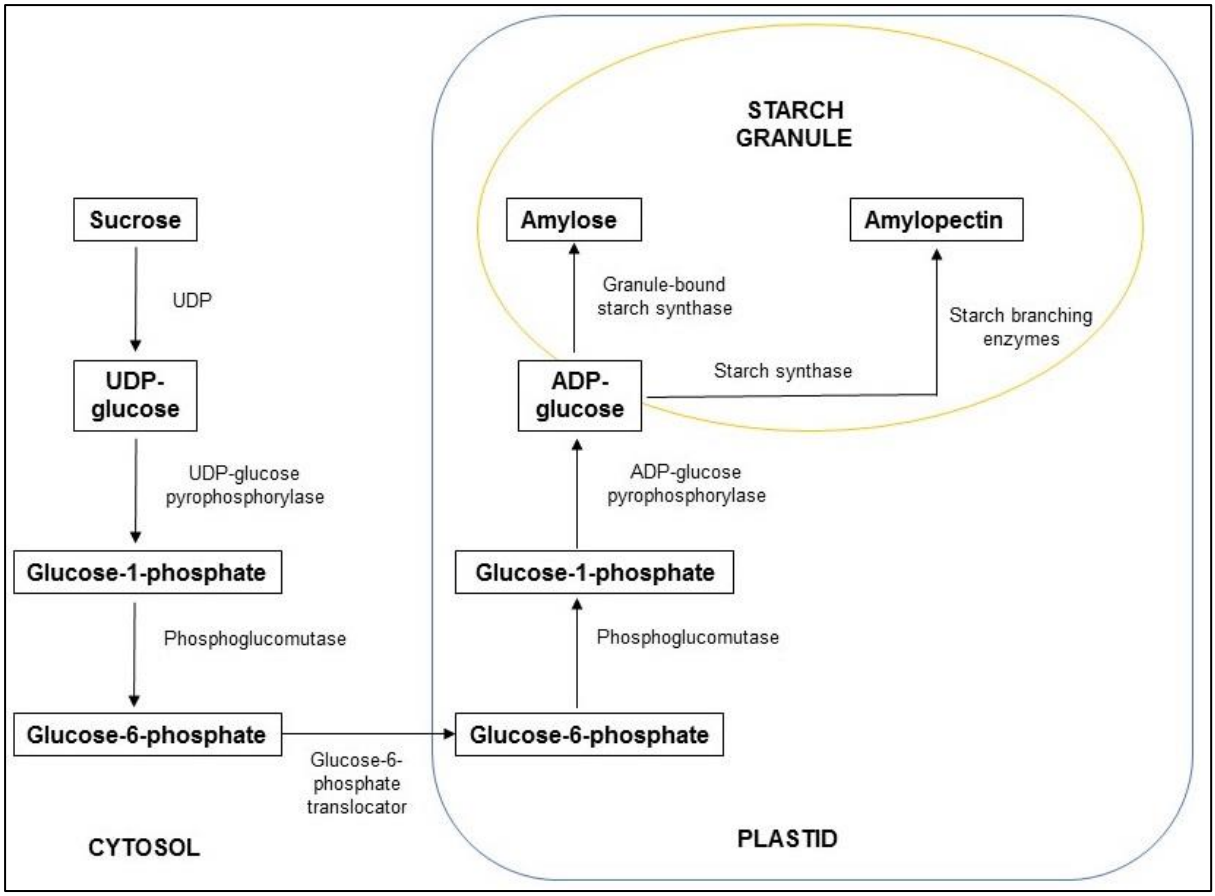

Figure 2: Simplified pathway of starch synthesis in heterotrophic cells via sucrose breakdown in the cytoplasm and transport to the plastid. (Modified from http://www.jic.ac.uk/STAFF/trevor-wang/images/full/starchpath2.jpg)

ADP-glucose acts as the substrate for starch synthases in most higher and many lower plant species, whereas certain lower plants utilize UDP-glucose (Viola et al. 2001). Granule-bound starch synthases (GBSS) are responsible for producing amylose, while starch synthases I–IV generate the chains incorporated into amylopectin. Mutations in genes encoding the SSI, SSII, and SSIII classes indicated that they elongate short, medium, and long chains respectively (Zeeman et al. 2010). Starch synthase IV has recently been found to play a significant role in starch granule formation, with *Arabidopsis SS4* mutants accumulating only one large starch granules per chloroplast. It was also found that *Arabidopsis SSIII SSIV* double mutants accumulated no starch during the day/night cycles, indicating that SSIII-SSIV activity are required for starch synthesis (Szydlowski et al. 2009; Roldán et al. 2007).

Granule-bound starch synthase is responsible for amylose synthesis, and generates long-chained glucans from the glucosyl residues of ADP-glucose (Fig. 2) (Grimaud et al. 2008). Amylose is synthesized in the starch granule within an amorphous ring structure, protecting it from the actions of branching enzymes (Zeeman et al. 2010; Tatge et al. 1999). GBSS can also act on the branch-points of amylopectin, creating longer side chains for denser starch granules and better carbon storage (Ral 2006).

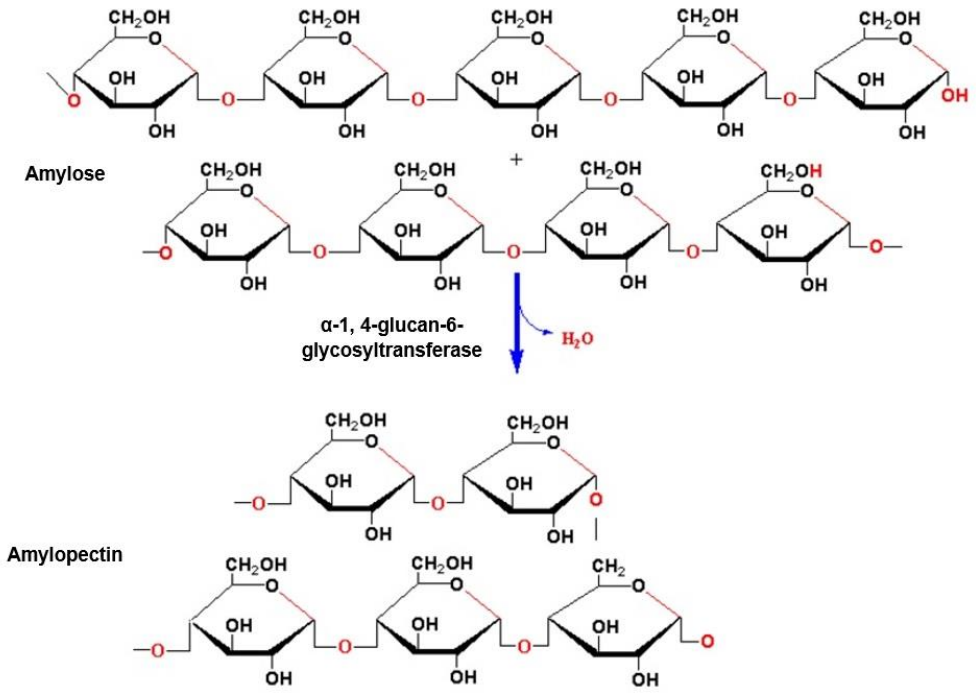

Figure 3: Amylopectin synthesis from amylose via α-1, 4-glucan-6-glycosyltransferase. (Modified from http://www.uky.edu/~dhild/biochem/11A/lect11A.html)

Amylopectin branching occurs concurrently with chain elongation, catalysed by branching enzymes (α-1, 4-glucan-6-glycosyltransferase). These enzymes cleave α-1, 4-glucan chains and transfer the chain to the C6 position of another glucan chain (Fig. 3) (Zeeman et al. 2010). *Solanum tuberosum* mutants lacking active starch branching enzymes have been shown to accumulate amylose at levels comparable to that of the highest commercially available starch (Schwall et al. 2000). The relative activities of starch synthase and branching enzymes vary between the organs of species, leading to differences in starch structure (Zeeman et al. 2010). Starch synthase I accounts for 60% of the soluble starch synthase activity in maize

endosperm, SSII accounts for 60% activity in pea embryos, and SSIII accounts for 80% activity in potato tubers (Zeeman et al. 2010). Evidence of complexes between starch synthases and branching enzymes has also been found in wheat endosperm. Using immunoprecipitation with antibodies specific for peptides associated with starch synthetic proteins, SSI, SSII, and SBE were identified as being present in phosphorylated complexes. Treatment of these complexes with alkaline phosphatase resulted in a separation into their monomeric proteins, indicating that protein complex formation is phosphorylation-dependant (Grimaud et al. 2008; Tetlow et al. 2008). Debranching enzymes such as isoamylases (ISA) and limit-dextrinases (LDA) also participate in glucose chain cleavage and aid in amylopectin synthesis. The number of debranching enzymes and associated isoforms differs between species. In *Arabidopsis* there are 4 debranching enzymes: ISA 1, ISA 2, ISA3, and LDA. Single, double and triple mutants of these genes produce starch granules with altered starch structures, but quadruple mutants produce no starch granules and instead accumulate branched-glucans which are degraded to malto-oligosaccharides by an α-amylase (Streb et al. 2008). More research into the interconnectivity between starch synthases, branching enzymes and debranching enzymes is necessary in order to fully understand the complexity of amylopectin synthesis.

1.3 Starch Degradation

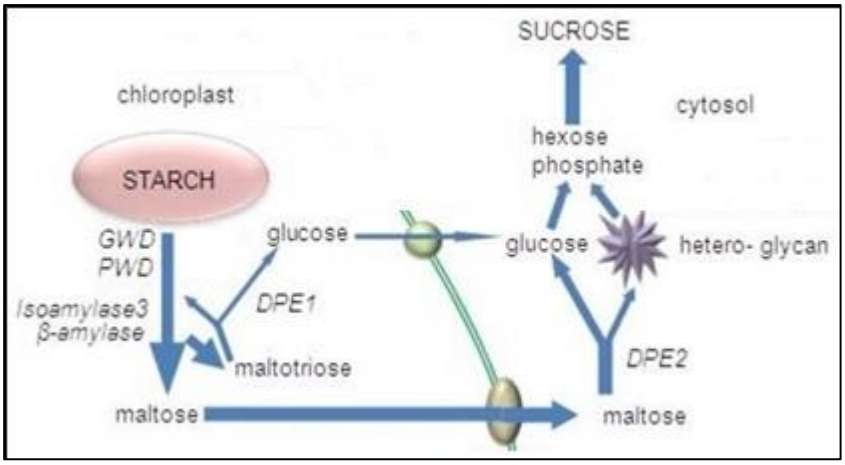

Figure 4: Diagram of the pathway of starch degradation in Arabidopsis leaves. Modified from (https://www.jic.ac.uk/staff/alison-smith/Control%20of%20starch_metabolism_in_leaves.html)

Typically, leaf starch degradation takes place during the absence of photosynthesis, when stored starch granules are broken down to sugars for metabolism. The granules are catabolised by hydrolases to glucose and maltose which are exported from the chloroplasts and metabolized in the cytosol (Fig 4) (Zeeman et al. 2010, 2007). β-amylases (BAMs) are responsible for hydrolyzing the α-1,4 bonds between glucans and form maltose, but cannot hydrolyze the α-1,6-branches of amylopectin (Crumpton-Taylor et al. 2012). That is accomplished by debranching enzymes such as isoamylase 3 (ISA3). The repression of a number of BAM and ISA genes associated with starch degradation has been performed in *Arabidopsis*, resulting in increased starch accumulation over several days in the *Arabidopsis* leaf (Zeeman et al. 2010). As starch granules are insoluble, starch degradation also involves the solubilisation of the granule surface through phosphorylation, which allows access to hydrolytic enzymes and subsequent hydrolysis of glucan chains (Zeeman et al. 2010). Phosphate addition is mediated by glucan, water dikinase (GWD) and phosphoglucan, water dikinase (PWD), which phosphorylate the C3-position and C6-positions of glucosyl residues in amylopectin (Fig. 4) (Kötting et al. 2010). Hansen et al (2009) showed via Nuclear Magnetic Resonance (NMR) spectroscopy that phosphate induces structural changes in starch, which allows hydrolytic enzymes access to the glucans in amylose in the centre of the semi-crystalline structure and

disrupts the amylopectin double helix structure. As GWD leads to the addition of phosphate groups to amylopectin which are essential for starch degradation (Edner et al. 2007), the down regulation of GWD has been shown to lead to an increase in starch accumulation in the leaves of transgenic clovers and rygrass (Zeeman et al. 2010). In rice, the disruption of the GWD gene OsGWD1 results in a 10-fold increase in leaf starch, while growth remains unaffected (Hirose et al. 2013). This is interesting as in similar experiments within *Arabidopsis* and the legume *Lotus japonicus* growth was affected (Caspar et al. 1991; Vriet et al. 2010). Due to its sequenced genome and 3-6 week growth cycle, genetic alterations in *Arabidopsis* have led to a better understanding of how to increase the availability of starch in crops species such as rice and maize (Grimaud et al. 2008; Koornneef & Meinke 2010; Fujita et al. 2011). However, experimentation involving model plants such as *Arabidopsis* cannot be applied to all crop species, as many come from different families and as such may have different pathways. One such family is Fabaceae, which contains the legume plants.

1.4 Legume Starch

Legumes are plants from the family Fabaceae and popular legumes include beans, peanuts and peas. A common trait among legumes is the presence of nitrogen-fixing bacteria within their root system (Deacon 2015). This is advantageous, as the legumes can be used to nourish soil with nitrogen before planting other crops. The three main plants used as models for legume research are currently soybean (*Glycine max*), barrel medic (*Medicago truncatula*) and birdsfoot trefoil (*Lotus japonicus*), all of which have had their genome sequenced (Cannon et al. 2009). These can be used as biological models for various crop species. Soybean in particular serves as a model for various developmental processes; however, some legumes are better suited for translational genomics than others, depending on the area of research. *Lotus* has been utilized to study plant-signalling (Oldroyd & Downie, 2008), *Medicago* for isoflavanoid pathways (Farag et al. 2008) and root architecture (Gonzalez-Rizzo et al. 2006), while soybean is a well-known model for both seed and root hair development and oil biosynthesis (Vodkin et al. 2008). The utilization of legumes for genomic comparison, and the results that followed, has led

researchers to explore other legumes not currently under study. One particular legume of interest is the marama bean (*Tylosema esculentum*) (Chingwaru et al. 2011; Holse et al. 2011; Cannon et al. 2009).

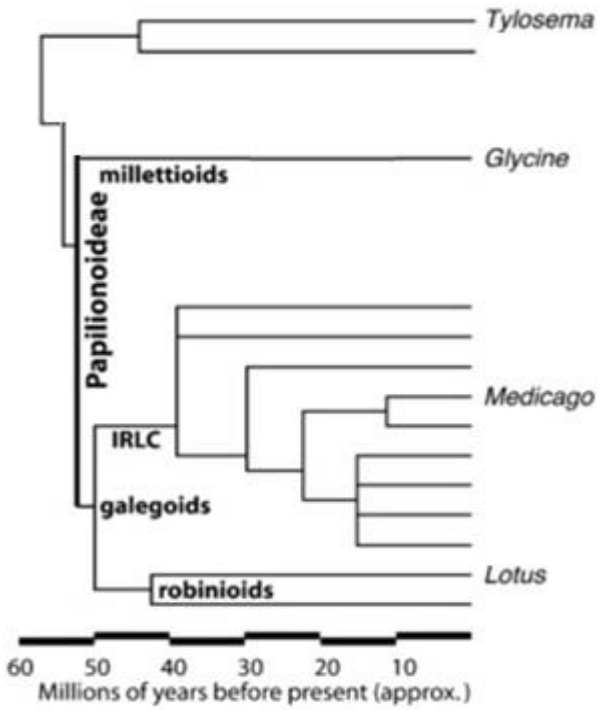

Figure 5: Taxonomic relationship of various legume genera. *Tylosema esculentum* shared a common ancestor with members of the Papilionnoideae and millettloids approximately 5 – 10 million years ago (Modified from Cannon et al. 2009).

Marama bean is a perennial plant that grows primarily in Sub-Saharan Africa, and is believed to be present in an early branch of the legume subfamily Papilionaceae (Fig. 5) (Cannon et al. 2009). Individual marama plants can grow several metres in size and produce large tubers of upwards of 100kg in weight (Chingwaru et al. 2011). The tuber is a staple food source for the indigenous people of the Kalahari, due to its high protein content (Mazimba et al. 2011), and it is highly adapted to this environment (Mitchell et al. 2005). Marama bean is considered to be an orphan crop; a plant which is grown agriculturally as a food source for humans and animals, but has not obtained enough interest to merit any genetic research (Armstead et al. 2009). However, recent findings have prompted interest in the marama bean's physiological properties. It has been shown to respond rapidly to increased

temperatures without risking severe water loss or increased photosynthesis, both of which can have detrimental effects on the plant (Mitchell et al. 2005). In 2013, Takundwa et al. developed single sequence repeats (SSRs) for the detection of polymorphisms in marama bean, and was successful in uncovering data that could be used for genetic variation analysis and breeding.

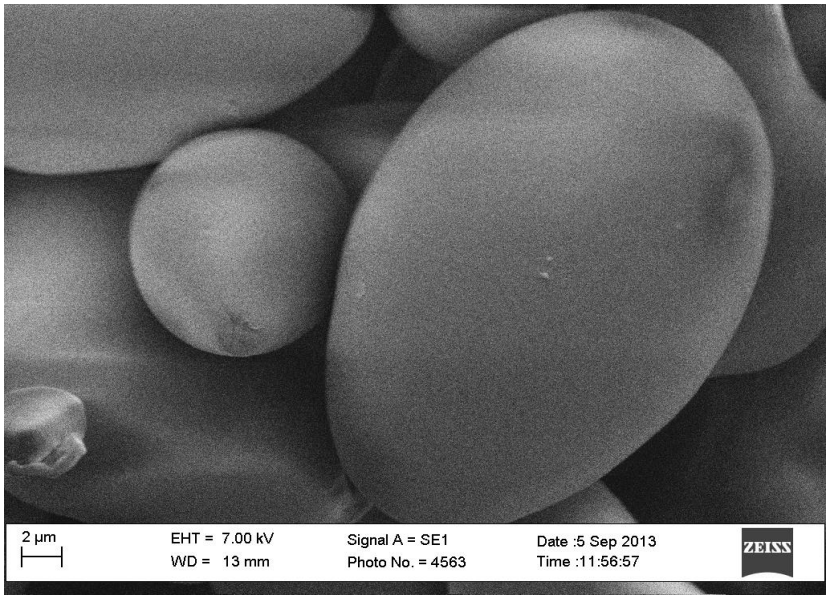

Figure 6: Scanning Electron Microscope (SEM) of marama bean starch granules. Obtained by Dr Emmanuel Nepolo, University of Namibia.

In 2013, a collaboration was established between the Institute of Plant Biotechnology (IPB) at Stellenbosch University and the University of Namibia between Dr James Lloyd and Prof Percy Chimwamurombe. The aim was to produce *T. esculentum* cDNA libraries to isolate starch biosynthetic genes. During this time, scanning electron microscope images of Marama bean starch granules were also obtained (Fig. 6). As the gelling properties of starch determine its industrial use, *Tylosema* starch granules and potato starch were subjected to Rapid Visco Analysis (RVA). This involves solubilizing starch in water at high temperatures and measuring their change in viscosity over time. The results show that *Tylosema* tuber starch had an exceptionally high gelling strength, compared to that of potato tuber starch (Fig. 7).

An alternative method for identifying marama bean starch genes is through analysing the transcriptome using RNA-Sequencing. Transcriptomics is defined as the study of the total RNA transcribed by the genome in a particular tissue (Sánchez-Pla et al. 2012). There are several factors that influence RNA transcription and accumulation, including environmental effects and developmental stages at harvest (Sánchez-Pla et al. 2012). From its emergence, next-generation sequencing methods utilized for transcriptome and genome assembly had several restrictions, such as long analysis times due to slow processing capabilities. Compared to the First-generation sequencing platforms which cost over $100 000 and assembled data over periods of years, modern technology such as Illumina 454 have decreased massively both price and analysis time, due to the development of improved extraction methods and data mining programs (Martin & Wang 2011; Severin et al. 2010; Grada & Weinbrecht 2013). Since then, many plant species with fully sequenced genomes have had their transcriptomes assembled, creating a more complete picture of their expression pathways, and allowing for gene identification and comparisons between various species (Palovaara et al. 2013; Severin et al. 2010).

An alternative method for understanding the pathways of higher plant starch synthesis is to compare the pathway to those of lower plants that differ in certain aspects. During the evolution of the plastidial starch storage, Rhodophyceae evolved starch synthesis pathways that do not require ADP-glucose as their substrate, but instead utilize UDP-glucose. Another key difference is that red algae do not have starch storage organs, and instead manufacture and store their starch in the cytoplasm of the

1.5 Starch in Red Algae

More than 1.5 billion years ago, an Archaeplastidial eukaryote entered into symbiotic relationship with a photosynthetic bacterium (Ball et al. 2005). The acquisition of plastids is believed to have occurred through this and several other similar events, leading to the evolution of the three major photosynthetic lineages: Chlorophyta (green algae and land plants), Rhodophyta (red algae), and Glaucophyta (unicellular algae) (Deschamps et al. 2008). Prior to endosymbiosis, the host organism utilized

UDP-glucose for cytosolic glycogen production, whereas the endosymbiont utilized ADP-glucose. Once endosymbiosis was complete, both substrates were used for starch production, each of which required enzymes from both the host and the cyanobiont. ADP-glucose produced by the cyanobiont was then directed to the host cytoplasm, enabling a metabolic link between the two organisms (Ball et al. 2009). When Rhodophyta and Glaucophyta branched away from the Archaeplastida, starch production remained in the cytoplasm alongside the use of UDP-glucose as a substrate, whereas in Chlorophyta, ADP-glucose utilization and starch production were maintained within the plastid (Ball et al. 2009). Although several lower plant species such as Glaucophyta produce cytosolic starch, it is specifically known as floridean starch when produced from organisms within the red algal family Florideophycideae (Fig. 8) (Ball et al. 2009; Viola et al. 2001).

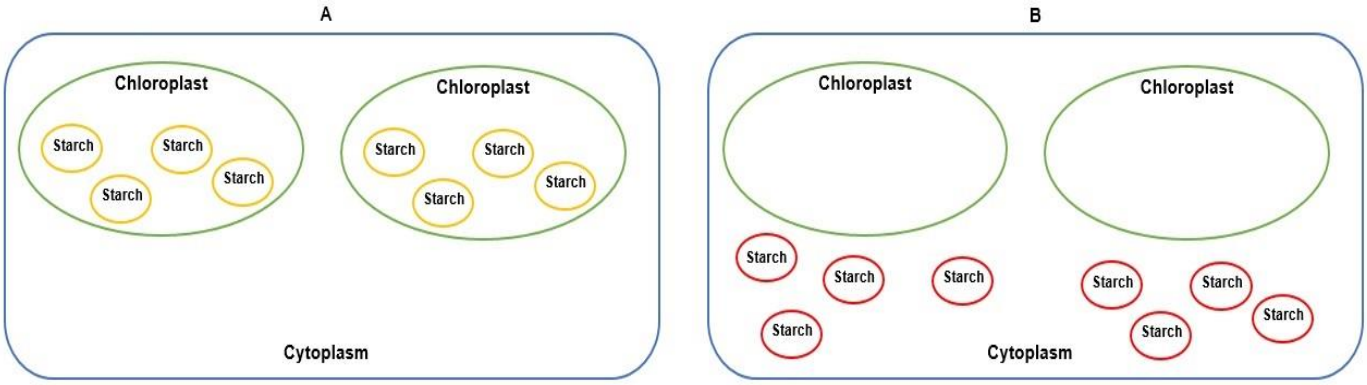

Figure 8: Simplified representation of starch granule partitioning. A) Starch granules within chloroplasts, typical of higher plant species. **B)** Starch granules within cytoplasm, common within red algal species and various lower plant species.

There are several theories as to why the compartmentalization of starch occurred during evolution. Higher temperatures have been shown to negatively affect floridean starch, showing lowered viscosity and gelatinization temperatures, but leave plastidial starch unaffected, offering an evolutionary advantage to survival at higher temperatures (Yu et al. 2002; Deschamps et al. 2008). These advantages are believed to be partly behind the successful evolution of Viridìplantae and the rise of the plant kingdom. Research has also been conducted comparing floridean starch to plastidial starch, and has shown to differ based on the species studied. The red

algae *Porphyridium purpureum* was found to contain amylose analogous to higher plant starch, while *Cyanidium caldarium* floridean starch resembles neither amylose nor amylopectin, but instead a glycogen-like polyglucans (Shimonaga et al. 2007). *Galdieria maxima*, a thermophillic red algae, was also found to produce floridean starch with a degree of branching higher than yeast glycogen, cyanobacterial phytoglycogen and other red algal floridean starch.

The study of red algae has led to increased understanding of plastidial evolution within the Chlorophyta and a broader understanding of starch synthesis evolution (Stiller & Hall 1997; Viola et al. 2001; Patron & Keeling 2005). Although these pathways evolved differently within Chlorophyta and Rhodophyta, ADP-glucose and UDP-glucose have similar structure and function, and the ability to utilize UDP-glucose as an alternative substrate in higher plants could lead to changes in substrate presence and starch levels. By utilizing random mutagenesis, we attempted to shift the preference of a starch synthase gene towards UDP-glucose utilization.

With next generation sequencing methods becoming cheaper and faster, more data about starch evolution and the mechanisms involved in starch synthesis can be obtained and analysed. This data can be used in conjunction with alternative starch sources in order to create a broader overview of the evolution of starch synthesis throughout history, and lead to more accurate genetic manipulation methods for increased starch production, the betterment of the starch industry.

Aims and Objectives

The first aim of this study involved acquiring physical and transcriptomic data for *Tylosema esculentum*. Our objectives were:

- Using flow cytometry and particle size analysis to determine marama bean starch granule size distribution.
- Determining marama bean amylose and amylopectin chain length using a reducing end assay.
- Growth and cultivation of marama bean plant.
- Leaf and tuber RNA extraction for transcriptomic analysis.
- The identification of marama bean starch synthases by mapping selected *Glycine max* starch synthases to marama bean transcriptome for comparison.
- Amplification of marama bean starch synthases from cDNA using designed primers

The second aim of this study was to establish a screening system for the utilization of UDP-glucose for starch synthesis, by randomly mutagenizing the *S. tuberosum* SSI gene. Our objectives were:

- Examination of conserved binding motifs within red algae species and *S. tuberosum.*
- Expression of the *A. thaliana* UDP-glucose pyrophosphorylase gene within the pACYC 184 plasmid, creating pACUG.
- Transformation of pACUG into the glycogen-deficient strain G6MD2.
- Random mutagenesis of *S. tuberosum SSI* gene within the pBluescript.SK plasmid.
- Transformation of the *pBluescript::SSI* library within G6MD2::pACUG.
- Colony screening using iodine vapour for glucan accumulation in G6MD2.

2. Transcriptomic analysis of *Tylosema esculentum* mRNA and identification of starch synthases

2.1 Introduction

Marama bean is native to the arid and semi-arid areas of South Africa, Botswana, and Namibia. Although little scientific interest about it has been raised in the past, it is interesting on a number of levels: Despite being a legume, it produces a tuber that can grow upwards of 100kgs in size, the plant is drought-tolerant, and indigenous people have been utilizing the plant and tuber for many years for food and medicine (Chingwaru et al. 2011). An interest in other orphan crops including marama bean has risen within the past two decades as they have the potential to help agriculture in developing countries. The Cassava plant stands as a strong example of the benefits of alternative starch sources, where previously it existed as an orphan crop, but has been developed into an international starch competitor (Sriroth et al. 2000). Cassava and marama also share similar points of interest; both are successfully utilized indigenously, both have high amounts of starch production, and both are cultivatable under adverse environmental conditions (Cock 1982; Sriroth et al. 2000; Mitchell et al. 2005; Holse et al. 2010).

In recent years, there have been several successful attempts at furthering the understanding of the marama plant, such as the protein composition of young tubers and the characterisation of associated compounds, such as phenolics and flavonoids (Amonsou et al. 2012; Shelembe et al. 2012). More recently, microsatellites have been successfully used to identify genetic variation between marama bean populations (Takundwa et al. 2010) as well as the construction of a marama bean cDNA library to study marama bean starch synthase genes (Swart 2013). In this study we aimed to expand on the information obtained by previous authors, while focusing mainly on two parts. The first is details regarding several aspects of starch structure, while the second would involve the transcriptomic analysis of marama bean RNA, and identification of genes involved in starch synthesis.

2.2 Methods and Materials

2.2.1 Starch particle size analysis

Starch granule sizes were determined using two methods. The first involved flow cytometry, based on the methodology developed by Cledat et al (2004). 5 µm, 15 µm and 25 µm polymer microspheres (Duke Scientific Corporation) with a refractive index of 1.54 were selected as measurement parameters, as it is almost identical to that of starch (Clédat et al. 2004). Once each size parameter was established, varying dilutions of starch were added to the flow cytometer to measure the refractive index of each particle. Once complete, the data was translated to show the corresponding particle size distribution of each dilution.

The second involved using a Saturn DigiSizer 5200 particle size analyser. 3 - 5 grams of starch was suspended in Milli Q water within the machine. Using a refractive index value of 1.54 (Clédat et al. 2004), and a 30 mW, 687 nm solid-state laser, starch granule size distribution was determined based upon the amount of light scattered by each particle.

2.2.2 Amylose and amylopectin separation

Amylose and amylopectin were separated by thymol precipitation as described by Tomlinson et al (1997). Briefly, 200 mg starch was dissolved in 12 ml of 90% (v/v) DMSO. 3 volumes of ethanol were added and vortexed until dissolved. The solution was centrifuged at 3000 x g for 5 minutes and the supernatant removed. Large particles of precipitate were ground if necessary. The ethanol wash step and centrifugation was repeated as before. The powder was dissolved in 40 ml of 1% (w/v) NaCl between 70^{o}C and 80^{o}C. Once dissolved, the solution was cooled to 30^{o}C. 80 milligrams of thymol was added with stirring, and left at 30^{o}C for 60 hours, after which it was centrifuged at 3000 x g for 5 minutes. The supernatant was removed for amylopectin isolation, while the precipitate contained the amylose fraction. To precipitate the amylopectin, 3 volumes of EtOH were added to the

supernatant and centrifuged at 3000 x g for 5 minutes. The supernatant was discarded and the wash step repeated twice. After each step the supernatant was removed and the precipitate ground to a powder which was washed twice with -20^{o}C cooled acetone and air-dried. For amylose, the powder was washed twice with thymol-saturated water and resuspended in 3 volumes EtOH before centrifugation at 3000 x g. The resulting powder was washed twice with -20^{o}C acetone and air-dried.

2.2.3 Examination of the branching frequency of starch.

A reducing end assay was performed to determine the frequency of branching in amylose and amylopectin molecules, based on the methodology developed by Regina et al. (2006). Five milligrams of amylose/amylopectin was suspended in 187.5 µl of water and 12.5 µl of 2 M NaOH, after which samples were boiled for 30 minutes with vortexing every 2 minutes to allow for complete dissolution. Once dissolved, samples were cooled, and 8 µl of glacial acetic acid was added, followed by 25 µl of 1 M (w/v) sodium acetate ($C_2H_3NaO_2$) and 250 µl water. For debranching of amylopectin, 2.5 U of isoamylase (*Pseudomonas sp.*, Megazyme) was added to starch samples and incubated at 37^{o}C for 2 hours. The isoamylase was deactivated by boiling for 10 minutes. Dinitrosalicylic acid (DNS) solution was prepared by dissolving 5 g of DNS in 100 ml of 2 M NaOH, followed by 60% (w/v) di-sodium tartrate solution to a final volume of 500 ml. 50 µl of starch solution was mixed with 150 µl of DNS solution and boiled at 100^{o}C for 10 minutes. The starch-DNS solution was cooled for 10 minutes at room temperature, before adding 180 µl of MilliQ water. The absorbance was measured at 540 nm, and a malto-triose solution was used to generate a calibration curve, with amounts ranging from 0-200 nmol.

2.2.4 Growth and cultivation

T. esculentum seeds were a gift from Prof Percy Chimwamurombe and Dr Emmanuel Nepolo (University of Namibia). Seeds were grown in a mixture of 80% sand and 20% soil for approximately 12 months between January 2014 and January 2015.

2.2.5 RNA extraction from *Tylosema esculentum* leaf and tuber

T. esculentum leaves and tubers were sliced into small discs and immediately frozen in liquid nitrogen, before being stored at -80°C. RNA extraction was performed according to a modified CTAB buffer RNA extraction protocol [CTAB buffer: 2% (w/v) cetyl trimethylammonium bromide; 2% (w/v) Polyvinylpyrrolidone (PVP); 100 mM TRIS-HCl, pH 8.0; 25 mM EDTA; 2 M NaCl]. β-mercaptoethanol was added to a final concentration of 3% (v/v) directly before use. Leaf and tuber samples were ground in liquid nitrogen. Approximately 300 mg of samples were added to 1.5 ml preheated CTAB buffer, and placed at 65°C for 30 minutes, vortexing for 10 seconds every 5 minutes. Centrifugation was performed at 13 000 x g for 10 minutes at room temperature before the supernatant was transferred to a fresh tube and one volume of chloroform/isoamylalcohol (24:1) (v/v) was added. The sample was vortexed for 30 seconds and centrifuged at 13 000 x g for ten minutes at 4°C. This step was repeated, after which the supernatant was collected and precipitated with 2 M LiCl at 4°C overnight. The RNA was centrifuged at 13 000 x g for 1 hour at 4°C, the supernatant discarded and the pellet washed with 70% (v/v) ethanol. The ethanol was removed before sterile Milli Q H_20 was used to resuspend the resulting RNA pellet. RNA samples were packaged in dry ice at -70°C and sent for transcriptomic analysis at the Agricultural Research Council (ARC) in Pretoria.

2.2.6 Transcriptomic analysis and primer design

Transcriptomic results were analysed by Dr Hano Maree (Stellenbosch University). A de novo sequence assembly was performed on the marama bean leaf and tuber samples, assembling putative full length transcripts from the 125bp reads. The ends of some reads were trimmed to lessen over-representation due to sequencing error. These full length sequences were pooled into one dataset for both leaf and tuber. Utilizing Phytozome 10.3 (http://phytozome.jgi.doe.gov/pz/portal.html) database, we selected 4 *Glycine max* starch synthases to be mapped against the pooled datasets, shown in Table 1.

Table 1: Glycine max starch synthase genes selected for comparison to *Tylosema esculentum* transcriptome. BlastX was utilized to compare *G. max* genes to *Arabidopsis thaliana* genes for closest sequential identification.

Gene ID	Blastx gene description
J01298.1	Soybean actin gene
Glyma.06G129400.1	Granule-bound starch synthase
Glyma.13G062700.1	Starch synthase II
Glyma.05G127800.1	Starch synthase IV

Once genes of similarity were found in the *T. esculentum* transcriptomic data, primers were designed to amplify the genes from *T. esculentum* cDNA (Table 2).

Table 2: Oligonucleotide primers for directed amplification of *Tylosema esculentum SSI* genes.

Directed gene	Forward Primer	Reverse Primer
Actin	5'-TTGGACTCTGGTGATGGTGT-3'	5'-TATATAAGAACAGAAACAGACCTGCTG-3'
GBSS	5'-TGGAGCAAAACTGGTGGACTTG-3'	5'-TGGAGCAATCTCCTCCCCTTCT-3'
SSII	5-'ATGGCGGCGCAGCTATGTACTTG-3'	5'-ATGTTCATGTCCTTCTGAACAAGCTGCC-3'
SSIV	5-'ATGGACGTGAAGAGGAAGATG-3'	5'-ATGTCCAACCAAGACCTGATT-3'

2.2.7 *T. esculentum* cDNA synthesis and gene amplification

T. esculentum cDNA synthesis was performed using the RevertAid First-Strand cDNA Synthesis Kit (Thermo-Scientific), according to manufacturer's protocol. *T. esculentum* starch synthase genes were amplified via PCR using GoTaq® (Promega) and Q5® High Fidelity DNA Polymerase (New England Biolabs), according to manufacturer's protocol. PCR products were separated in a Tris-borate-EDTA (TBE), 1% (w/v) agarose gel.

2.3 Results

2.3.1 *T. esculentum* starch particle size distribution

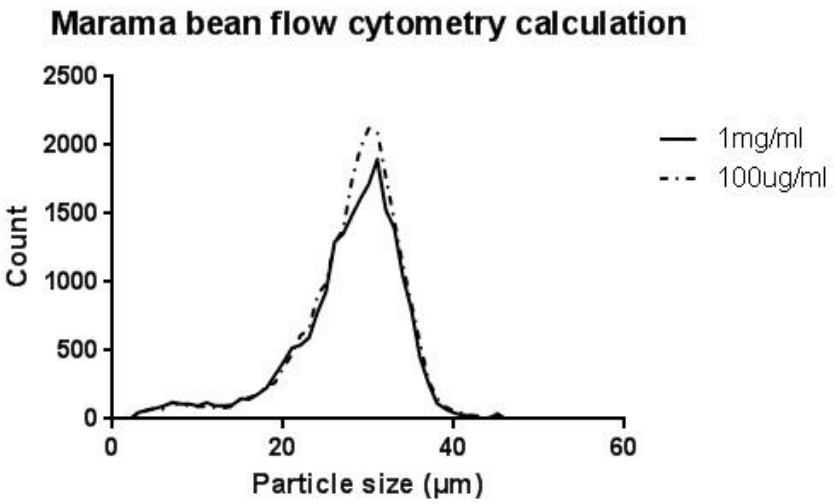

Figure 9: Flow Cytometry results of marama bean starch granule size. All samples were repeated in triplicate.

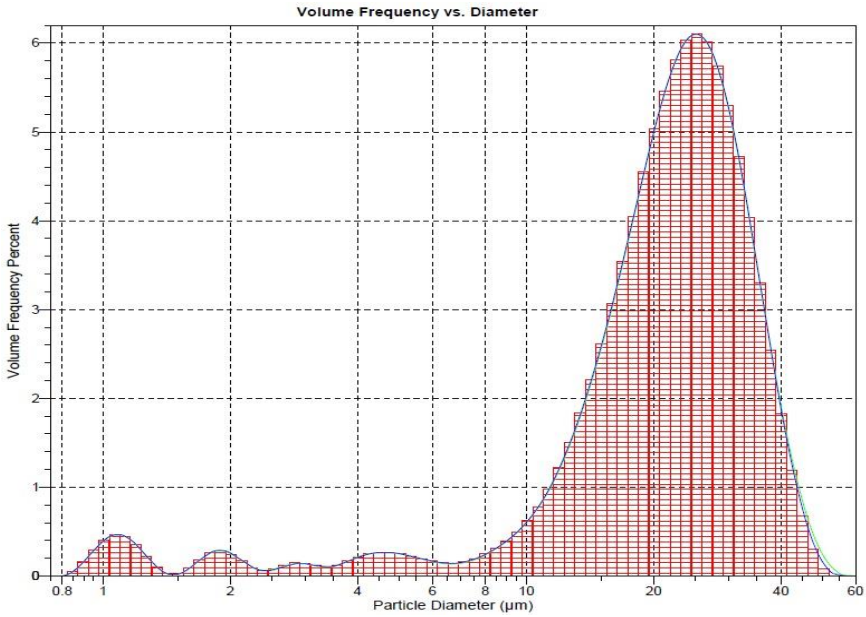

Figure 10: *T. esculentum* starch particle diameter by laser light scattering. Samples were repeated in triplicate.

Two methods were examined to measure the average size of starch granules from marama bean tubers. Initially, flow cytometry results were obtained (Fig. 9), but were restricted to particle sizes below 50 µm due to filter size restrictions and larger particles blocking liquid flow within the machine. In order to verify that these results had not excluded larger particles, a Saturn Digisizer particle size analyser was used to verify the readings and include any particles above 50 µm (Fig. 10). In both the Forward Light Scattering and Particle Size Analyser measurements, the starch granules showed a larger size distribution above 10 µm, with particle sizes ranging between 15 and 35 µm. No marama starch particles were measured at sizes above 50µm using either technique. All technical reps showed the same size distribution pattern. The results displayed by both methods verified that marama bean starch granules have a size distribution of between 15 µm and 35 µm.

2.3.2 Reducing end assay of *T. esculentum* starch

The reducing end assay allows the calculation of the average chain length within amylose or amylopectin by examining the number of reducing ends before and after debranching by isoamylase. The 3 samples selected for assay were marama tuber starch, potato starch extracted by standard starch extraction protocol, and commercially obtained potato starch (Sigma, S9765). The samples were separated into amylose and amylopectin by thymol precipitation before the average chain length was determined.

Table 3: Average glucose units per chain within amylose and amylopectin.

	Average glucose unit per chain		Average glucose unit per chain
Potato amylose	217 ± 11	Potato amylopectin	42 ± 3
Marama amylose	363 ± 4	Marama amylopectin	27 ± 3

As Table 3 shows, marama bean tuber starch accumulates longer average amylose chains than potato starch being approximately 70% longer. The average number of glucose moieties per chain of amylopectin for the marama bean starch is 35% lower than potato amylopectin, at 41 glucose molecules per chain. Amylopectin is the major component of starch, typically making up 70% or more of the total fraction of starch. Marama bean starch shows longer average amylose chains, but shorter amylopectin chains than that of potato. Amylose synthesis typically takes place within a semi-crystalline structure surrounded by amylopectin, and the resulting amylose/amylopectin ratio determines the starch physiological properties, such as melting temperature and viscosity (Zeeman et al. 2010; Tatge et al. 1999). This could be one possible explanation for the interesting viscosity properties of marama bean starch (Figure 7). Further study regarding the physical properties of marama plant amylose and amylopectin is required to provide evidence towards their contributions to these unique starch properties.

2.3.3 *Tylosema esculentum* leaf and tuber RNA extraction

The use of transcriptomics to identify genetic and evolutionary linkages between different species has already been demonstrated. Hefer et al. (2015) successfully used comparative transcriptomics to identify clusters of homologous genes during xylem formation between two different wood-forming species, *Populous trichocarpa* and *Eucalyptus grandis*. Comparing the xylem and leaf transcriptomes of both

species, there was evidence ranging from conserved expression of the genes involved in secondary cell wall biosynthesis, to species-specific expression of genes within both the xylem and leaf transcriptomes. With this in mind, a similar comparison could be made between marama bean (*T. esculentum*) and other legume crops. Total RNA was extracted from both leaf and tuber material using a modified CTAB method (Fig. 11).

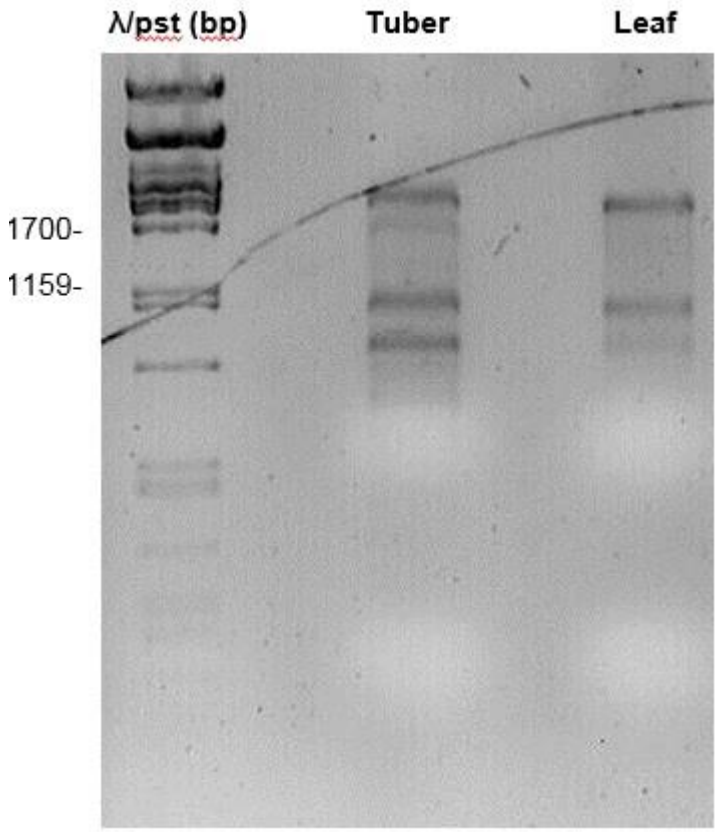

Figure 11: Marama tuber and leaf RNA. RNA was visualised by 1% (w/v) agarose gel. 28S and 18S RNA bands are visible in both leaf and tuber RNA.

Marama leaf sample concentrations were 500 ng/µl and tuber samples averaged 35-50 ng/µl. Tuber samples were eluted in smaller volumes and combined to form a total tuber extract concentration of 140 ng/µl, as concentrations above 50 ng/µl were necessary for transcriptomic analysis. Samples were sequenced using a commercial service and transcriptomic results analysed using the Illumina Basespace® analysis tools provided (https://basespace.illumina.com).

2.3.4 Transcriptomic analysis of *T. esculentum* RNA

In collaboration with Dr Hano Maree at the Department of Genetics, Stellenbosch, we analysed the transcriptomic data received.

Table 4: Transcriptomic results of *T. esculentum* mRNA.

	Leaf	Tuber
Total sequence bp reads	18884732	10110753
GC content	42%	44%
Phred Quality Score	36 (99.9% accuracy)	36 (99.9% accuracy)
Per Base Quality Score	34 (Good quality read. Total out of 40)	34 (Good quality read. Total out of 40)

Table 4 shows the results obtained for the marama leaf and tuber mRNA samples sequenced. All samples were run in triplicate using the Illumina HiSeq 2500 sequencing system (Illumina Inc. 2013). Nearly 30 million reads in total were obtained for the combined samples at 125 bp per read, with the leaf sample having nearly double the number of reads than the tuber, possibly due to the difficulty of obtaining high concentrations of RNA from marama tuber. The Q score value of 36 for both samples shows a 99.9% accuracy, translated as 1 sequence error per 1000 bps, and the GC content for both leaf and tuber samples were similar at 42% and 44% respectively, which has not been shown before. Slight deteriorated end occurred due to the end-to-end sequencing nature, so sample data were trimmed prior to contig sequence assembly.

2.3.5 Identification of starch biosynthetic genes within the *T. esculentum* transcriptome

Using Phytozome 10.3 (http://phytozome.jgi.doe.gov/pz/portal.html) and *Glycine max* as a reference genome, 4 starch synthase genes labelled starch synthase I-IV were

identified (Glyma.06G129400.1; Glyma.13G062700.1; Glyma.07G260500.1; Glyma.05G127800.1), and one granule-bound starch synthase gene (XM_003556383.2) and a soybean actin gene [J01298.1]. Each sequence was mapped to the marama transcriptome data to search for contigs and open reading frames. Four consensus sequences were found with marked open reading frames, and these sequenced were subjected to BlastX to identify the most likely gene relation within the NCBI database. The results obtained are shown in Table 5.

Table 5: BlastX results for *T. esculentum* consensus sequences.

	BP size	BlastX gene relation	Max score	Query cover	E.value	%Identity	Accession value
Concensus_1a	2023	*Glycine max* granule-bound starch synthase	1217	99%	0.0	100%	XP_003556431.1
Concensus_1b	3180	*Theobromo cacao* starch synthase 4 isoform	1532	99%	0.0	73%	KHN25377.1
Concensus_3	375	*Glycine soja* soluble starch synthase 3	238	99%	3e-70	87%	KHN33026.1
Concensus_5	1860	*Theobroma cacao* starch synthase 2	962	93%	0.0	80%	XP_007032555.1

Using the CLC Genomics Workbench (Quagen 2015), the open reading frames of 3 consensus sequences in Table 5 (1a, 1b and 5) were selected for primer design and gene amplification from *T. esculentum* cDNA. The figures 12 and 13 show successful PCR amplification of 3 genes on 1% (w/v) agarose gel.

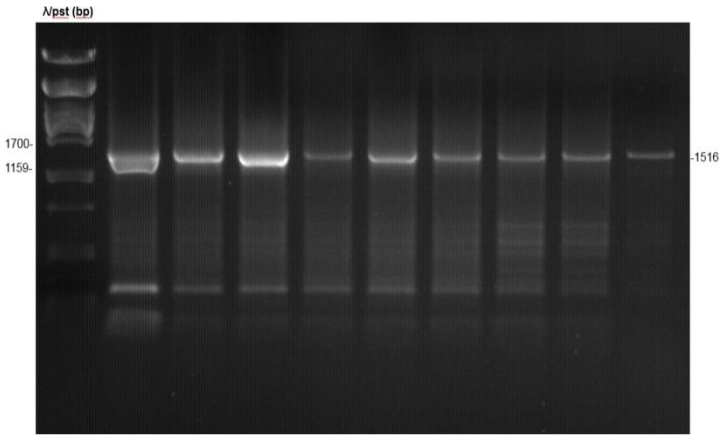

Figure 12: Gradient PCR amplification of Concensus_1a: *T. esculentum* granule-bound starch synthase. Temperature range from 56°C to 76°C.

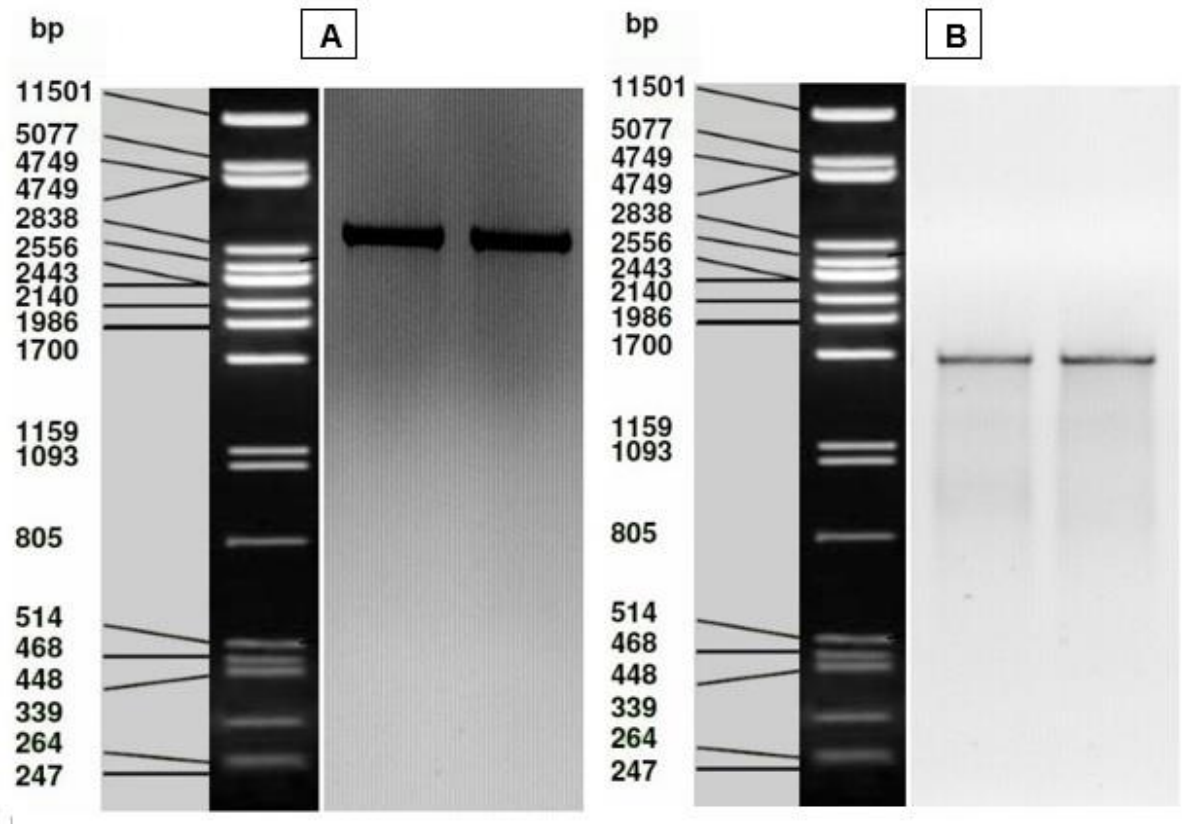

Figure 13: PCR amplification of *T. esculentum* genes. A) Concensus_1b: Starch synthase 4 gene, approximately 3.1 kb. B) Concensus_5: Starch synthase 2 gene, approximately 1.68 kb.

The amplified *T. esculentum* starch synthase genes were then ligated into the pBluescript.sk+. As marama bean starch synthases utilize ADP-glucose as a substrate for starch synthases, we aimed to transform these genes into G6MD2 *E. coli* cells containing an unregulated bacterial AGP*ase* gene (*GlgC16*) within pACYC 184 (Kossmann et al. 1999), similar to the G6MD2::pACUG screening we performed with *SSI,* and utilize iodine screening for expression tests. Unfortunately, complications with regards to the competency of the *G6MD2::pACAG* strain resulted in no further transformations of the *T. esculentum* starch synthase plasmids being transformed for screening.

2.4 Discussion

The marama bean is capable of growing in semi-arid to arid conditions in soil conditions where traditional crop species such as potato cannot grow. In order to help develop it as a crop, this study aimed to examine the physical properties of marama bean starch and provide transcriptomic data about genes involved in starch synthesis. A previous attempt to screen a *T. esculentum* cDNA library for active starch synthase clones yielded none (Swart 2013), possibly due to the cDNA library having a transcript profile of low starch synthase expression coupled with a low screening number of 100 000 colonies. This study aimed to provide a template for the amplification of specific *T. esculentum* starch synthase genes and acquire large-scale data of the marama bean transcriptome for future assembly and analysis.

Certain physical and structural properties, such as starch granule size and amylose and amylopectin ratios, determine the industrial uses of starch, and various methods exist at measuring these properties. Scanning Electron Microscopy (SEM) produces highly accurate particle pictures for determining starch granule size, but has the risk of observer bias, meaning measurements may not be accurate. Flow cytometry, however, determines the size of particles within an elution droplet by the amount of light scattered by the particle (Jones et al. 2004). This light scatter is compared to a standard curve, utilizing microspheres of similar size and refractive index to the particles being tested. This allows for a more accurate readout of size distribution. In 2004, Clédat et al. utilized sedimentation field-flow fractionation to measure the size of rice starch granules using microspheres of a similar size and refractive index to that of starch. With the same principle and two similar methods, an accurate model of marama starch granule size distribution was obtained. Both flow cytometry and particle size analysis displayed marama bean starch granule sizes concentrated between 15 µm and 35 µm, with no granule sizes above 50µm. In comparison, potato starch granules have a far larger particle distribution pattern, averaging between 5 and 100 µm in size (Ellis et al. 1998).

Other properties that determine starch properties are the length of amylose of amylopectin chains within the starch granules, and highly branched and unbranched

starch has been shown to change the properties of starch granules (Stadnichuk et al. 2007; Streb et al. 2008). We performed a reducing end assay to determine the amylose and amylopectin chain length, and found that the marama bean tuber starch has a significant difference in the amylose chain length, 67.5% higher than that of potato tuber starch extracted through the same method. The amylopectin chain length, however, was 34.8% lower than that of the extracted potato starch. The fine structures of amylose and amylopectin have been shown to greatly affect starch granule physiochemical properties (Zeeman et al. 1998; Tatge et al. 1999). Coupled with the narrow starch granule size distribution, this could also explain the RVA results shown previously (Fig. 7), which may lead to several unique uses for marama bean starch. Future plans include the assembly of the full transcriptome from the data acquired in this project for analysis in the field of comparative transcriptomics, which will hopefully lead to traits beneficial to the starch industry.

Three genes related to starch synthase sequences found within the soybean genome were successfully identified using primers designed based on the transcriptome data; a granule-bound starch synthase (GBSS), a starch synthase II (SSII), and a starch synthase IV (SSIV). All 3 were successfully amplified; however, demonstration that they encoded active proteins by expression in *E. coli* was unsuccessful during the frame of the project. The use of the *T. esculentum* transcriptome helping to design primers allowing for their amplification, however, demonstrates the benefit of having such data.

3. Mutation of *Solanum tuberosum* starch synthase gene and screening for UDP-glucose utilization

3.1 Introduction

As stated in the general introduction, the mechanisms involved in starch synthesis and degradation involve a large network of interconnectivity, and new information arises each year to further our understanding of these processes. This complexity most likely stems from the primary endosymbiotic event between Archaeplastida and the endosymbiont. All Viridìplantae can be directly traced back to this event and analysis of *Arabidopsis* has shown that 18% of its genome can be traced directly to the endosymbiont itself (Ball 2005). This same event also resulted in the utilization of ADP-glucose and the compartmentalization of starch granules within chloroplasts in the Viridìplantae lineage. The Rhodophyceae phylum branched away from Archaeplastida after the endosymbiotic event, but prior to the emergence of plastidial starch storage (Stiller & Hall 1997; Viola et al. 2001), presenting an alternative study model for starch synthesis in plants.

One difference in starch synthesis between Viridiplantae and Rhodophyceae is that the first enzyme in the committed pathway is AGP*ase* in the former, and UGP*ase* in the latter (Viola et al. 2001). UDP-glucose is still utilized by land plants for sucrose biosynthesis, and is coupled to the production of ADP-glucose for starch synthesis in certain seed endosperms (Kleczkowski et al. 2004). Starch synthases contain a catalytic domain at the C-terminus, while the N-terminus contains sequences that are thought to influence the specificity of the enzymes. Previous studies of higher plant starch synthases have shown them to contain two apparent ADP-glucose binding motifs within the catalytic domain. One, which is the main binding site, is present close to the N-terminal end of this domain, while a ”lookalike” sequence is present at the C-terminus (Edwards et al. 1999). Similar NDP-glucose binding motifs have been shown in certain red algal species (Edwards et al. 1999), suggesting that sequence specificity is not essential for NDP-glucose binding.

With this in mind, our project consisted of determining whether a higher plant starch synthase from *Solanum tuberosum* could be altered to utilize UDP-glucose for starch

synthesis, similar to that shown within red algal species. For this approach, we examined the starch synthase protein sequences of several red algae species and compared them to *S. tuberosum* starch synthase, to try and identify amino acids for site-directed mutagenesis that would lead to a change in substrate specificity. In addition, random mutagenesis of a *S. tuberosum* starch synthase gene was performed, and the mutated cDNA was screened for activity using UDP-glucose as a substrate in an *E. coli* expression system. A similar screening system has been used to identify higher plant starch synthases forming glucans in the presence of ADP-glucose within G6MD2 *E. coli* (Abel et al. 1996; Kossmann et al. 1999), and colonies producing glucans will stain blue in the presence of iodine vapour (Fig. 14)

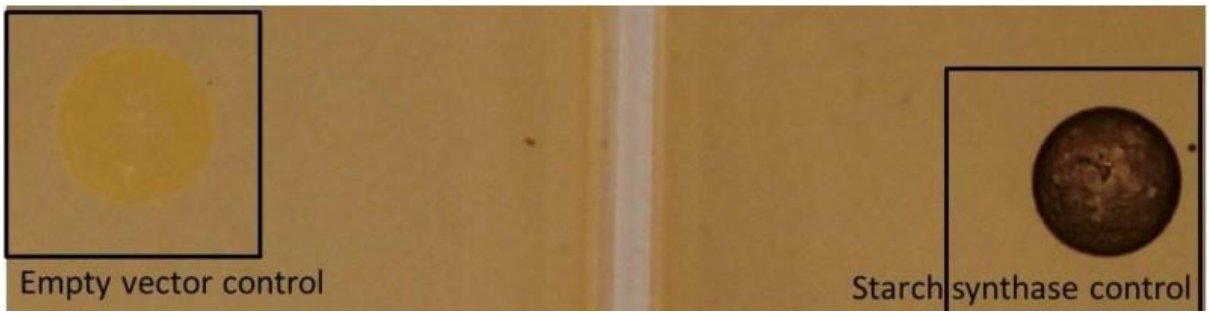

Figure 14: Iodine staining of glucan production. Colonies not producing glucans stain yellow (left), while colonies actively producing glucans stain blue (Swart 2013)

3.2 Methods and Materials

3.2.1 *E. coli* genotypes

The following *E. coli* genotypes were used in the study:

- **DH5α (Invitrogen)**: F- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoRnupG Φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(rK-mK+), λ–
- **XL1 Red (Agilent Technologies)**: F- endA1 gyrA96 (naIR) thi-1 hsdR17 supE44 relA1 lac mutD5 mutS mutT Tn10 (Tetr)a
- **G6MD2**: Hfr (PO101), *hisA323* (Stable), Δ(bioasd)29

3.2.2 Sequence Alignment

Sequences were compared using the CLC Genomics Workbench (Version 3.6.5.0). Sequences used were:

- *Chondrus crispus* starch synthase (XP_005718355.1)
- *Porphyridium purpureum* starch synthase (BAF49289.1)
- *Galdieria sulphuraria* starch synthase (XP_005707965.1)
- *Cyanidioschyzon merolae* starch synthase (XP_005537174.1)
- *Solanum tuberosum* soluble starch synthase I (P93568.1)

3.2.3 PCR amplification

PCR was performed using GoTaq® DNA Polymerase (Promega) according to manufacturer's instructions. Primer sequences for *S. tuberosum SSI* and *A. thaliana* UGP*ase* are detailed in Table 6. Genes were isolated using the Genejet Gel Extraction kit and PCR Purification kit, according to manufacturer instructions (Fermentas, USA).

Table 6: Oligonucleotide primers for amplification of *SSI* gene from *S. tuberosum* and UGP*ase* gene from *A. thaliana*.

Directed gene	Forward Primer	Reverse Primer
SSI	5'-AGTGGATCCCCCGGGCT-3'	5'-AATTGGGTACCGGGCCC-3'
UGP*ase*	5'-ACCACTACTCCCTTGGCCTA-3'	5'-CTTGCTTTGCTGTGCCACTT-3'

3.2.4 Preparation of competent cells

Chemically competent *E. coli* cells were prepared according to the method of Inoue et al. (1990).

3.2.5 Plasmid Isolation

Plasmid isolation was performed using the Genejet Plasmid miniprep kit according to manufacturer instructions (Fermentas, USA). Large-scale plasmid isolation was performed according to the maxiprep protocol of Sambrook and Russell (2001).

3.2.6 Transformation of *E. coli* cells

E. coli cells were thawed on ice for 5 minutes. One to two milligrams of plasmid DNA was added to the cells and incubated on ice for 15 minutes. The reaction was incubated at 42°C for 45 seconds before incubation on ice for 2 minutes. Three hundred and fifty microliters of LB media [1% (w/v) peptone, 1% (w/v) NaCl, 0.5% (w/v) yeast extract] was added to the cells and incubated at 37°C for 90 minutes with shaking at 200 rpm. The cells were plated onto solid LB agar [1% (w/v) peptone, 1% (w/v) NaCl, 0.5% (w/v) yeast extract, 1.5% (w/v) bacterial agar,] containing 1% (w/v) glucose, 100 µg/ml 2, 6-Diaminopimelic acid (DAPA) and appropriate antibiotics (35 µg/ml chloramphenicol; 100 µg/ml ampicillin).

3.2.7 Mutagenesis of *SSI* gene

E. coli XL1-Red was used to introduce random mutations according to manufacturer instructions (Aligent Technologies). Resulting colonies were picked at random from transformation plates using sterile toothpicks to inoculate thirty-three 3 ml LB broth [1% (w/v) peptone, 1% (w/v) NaCl, 0.5% (w/v) yeast extract] containing ampicillin (100µg/ml). Cultures were combined and a plasmid maxiprep was performed (Sambrook & Russell 2001).

3.2.8 Screening for accumulation of glucans in *E. coli*

E. coli (G6MD2) were grown overnight on LB agar with 1% (w/v) glucose, 100 µg/ml DAPA and appropriate antibiotic selection (35 µg/ml chloramphenicol; 100 µg/ml ampicillin). Colonies were stained with iodine crystal vapour.

3.2.9 Preparation of cells for enzymatic activity assay

E. coli colonies were inoculated overnight in 2 mls TB broth [1.2% (w/v) peptone, 2.4% (w/v) yeast extract, 0.4% (v/v) glycerol, 1.25% (w/v) K_2HPO_4, 0.23% (w/v) KH_2PO_4] with antibiotic selection (chloramphenicol, 34 µg/ml) and DAPA (100 µg/ml). The 2 ml culture was reinoculated into 50 ml of TB broth with antibiotics, and incubated for 3 hours at 37°C shaking. Cells were decanted into 50 ml tubes and centrifuged at 4000 x g for 20 minutes at 4°C. All following steps were performed on ice. Cells were resuspended in 20 ml ice-cold extraction buffer [50 mM Tris-HCL (pH 8), 1 mM $MgCl_2$, 100 mM EDTA, 1 mM β-mercaptoethanol], and sonicated on ice for 30 seconds at 12 V. This was repeated three times with 1 minute recovery on ice. The protein solution was then centrifuged at 4000 x g for 20 minutes at 4°C.

3.2.10 UGP*ase* activity assay

Ten microliters of crude protein extract was added to 200 µl of assay buffer [100 mM Tris-HCL, pH8, 2 mM MgCl2, 20 uM Glc-1,6-bisphosphate, 0.25 mM NAD+, 2 mM

UDP-glucose, 3 U/ml PGM, 2.5 U/ml G6P-DH]. The reaction was started with the addition of PPi (2 mM), and the activity determined by change in absorption (OD_{340}).

3.3 Results

3.3.1 Motif comparison of *S. tuberosum* and red algae starch synthases

Using the CLC Genomics Workbench, we compared the starch synthase sequences of 4 UDP-glucose-utilizing algal species (*Chondrus crispus, Galderia sulpheraria, Cyanidoschyzon merolae* and *Porphyridium purpureum*) as well as potato starch synthase (Fig. 15).

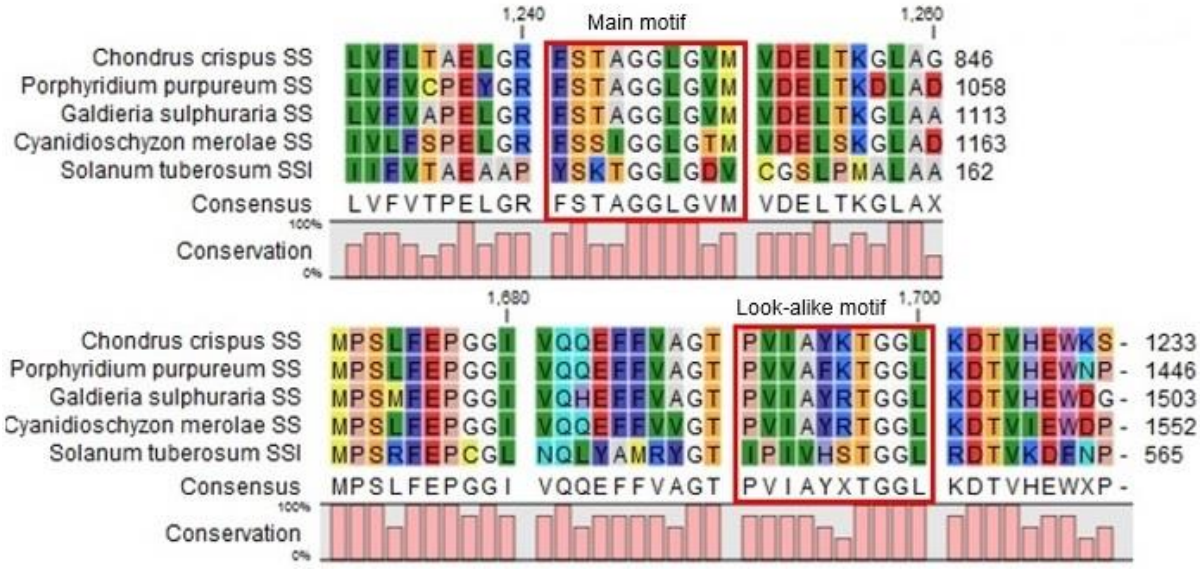

Figure 15: Primary amino acid sequence comparisons between sections of the catalytic domain of starch synthase isoforms from several algal species that accumulate floridean starch and *S. tuberosum.* Red bars indicate sections of alignment with percentage of sequence conservation. Consensus sequence indicates most likely amino-acid sequence resulting from alignment. Potential ADP-glucose binding motifs are enclosed in red.

The catalytic domain within starch synthases contains two potential sugar nucleotide binding motifs: the KTGGL binding motif and a secondary “look-alike” binding motif (Edwards et al. 1999). Sequences from the red algae differ in that within the main motif the lysine (K) is substituted with either serine (S) or threonine (T), while the threonine (T) is substituted by either alanine (A) or isoleucine (I). In addition, within the look-alike motif (which, unusually, is STGGL in the case of the potato starch synthase) the red algal starch synthases is K/RTGGL. There are several other amino acid substitutions in the red algal sequence compared with that from potato, meaning

that it is difficult to identify specific regions that might influence the specificity of the substrate used.

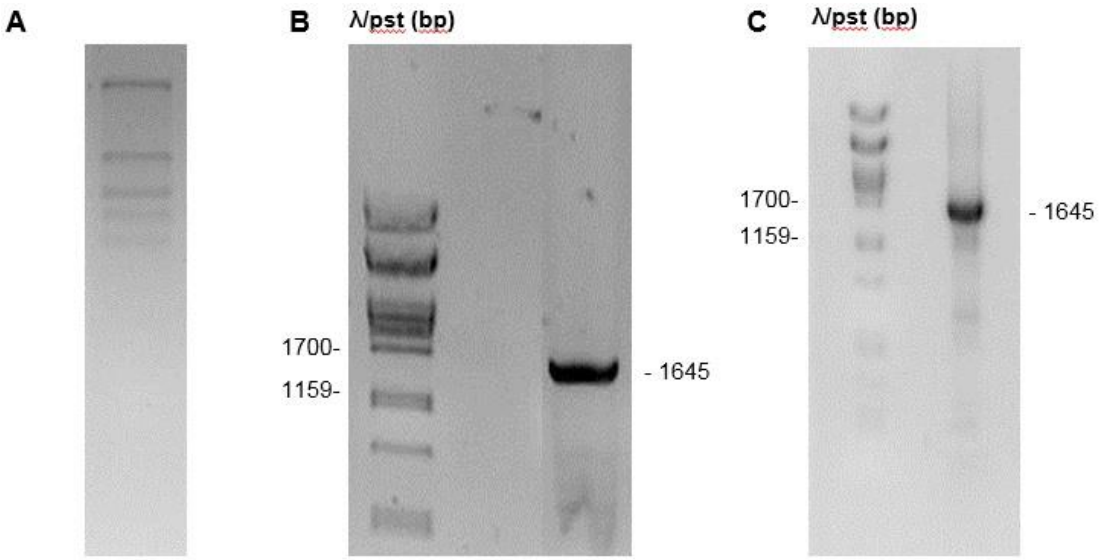

Figure 16: *A. thaliana* RNA extraction and UGP*ase* amplification. A) RNA extracted from *A. thaliana* leaf**. B)** Amplification of UGP*ase* amplicon from *A. thaliana* cDNA. **C)** UGP*ase* gene within pBluescript.SK+. T7 primer and UGP*ase* reverse primer were used to confirm orientation.

3.3.2 Production of construct for expression of UGP*ase* in *E. coli*

In order to develop an *E. coli* functional screen that identifies UDP-glucose utilizing starch synthases, a UDP-glucose pyrophosphorylase needed to be amplified to provide the substrate. Total RNA was extracted from *A. thaliana* leaf samples using a modified CTAB method and visualised by agarose gel electrophoresis (1% w/v) (Fig. 16A). RNA integrity was verified, and cDNA synthesis was performed to provide a template for PCR amplification. UGP*ase* flanking primers were designed using the NCBI Sequence [NM_111195.3] as a reference, resulting in a 1645 base pair amplicon (Fig 16B). This was purified and ligated into the EcoRV site of pBluescript.SK, and transformed into DH5-α *E. coli* cells. Colony PCR was performed to ensure that the UGP*ase* gene was present in the sense orientation with respect to the LAC promotor and terminator (Fig. 16 C). The active UGP*ase* gene, promotor and terminator were restricted from pBluescript.SK using EcoRI, blunted

and ligated into the EcoRV site of pACYC 184, making the vector pACUG. This vector was used as it is capable of coexisting in *E. coli* with other vectors containing the ColE1 origin of replication like pBluescript.

3.3.3 Enzymatic activity of UGP*ase* gene

The pACUG vector was transformed into *E. coli* cells and colonies grown prior to extraction of crude protein. UGP*ase* activity was measured by the conversion of NAD to NADH in both DH5-α and G6MD2 cells, as shown in Fig 17.

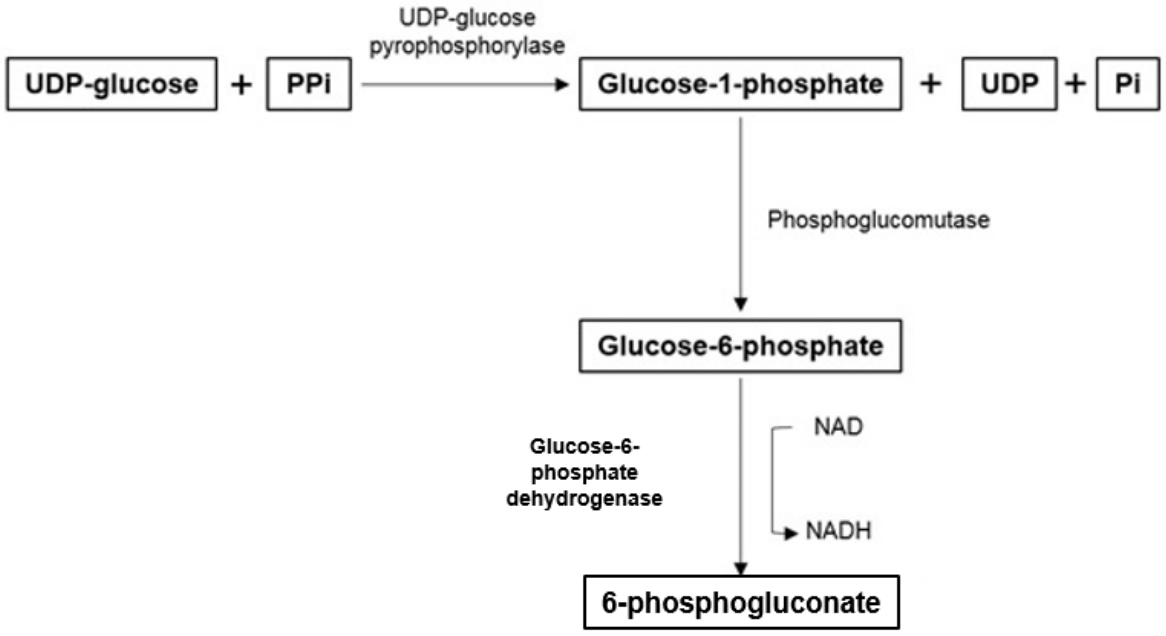

Figure 17: Chemical pathway of UDP-glucose conversation to 6-phosphogluconate. The reaction yields NADH as a by-product, measured at 340 nm.

The presence of active UGP*ase* would lead to the conversion of UDP-glucose to glucose-1-phosphate. Subsequent reactions would release NADH which absorb at a wavelength of 340 nm. As a positive control, commercial UDP-glucose pyrophosphorylase was added to sample buffer, while *E. coli* cells for each strain containing pBluescript with no insert were measured as negative controls. The reaction was started by the addition of pyrophosphate (2 mM final concentration).

Table 7: G1P production within DH5-α *E. coli* and G6MD2 cells.

	G1P production (µmol.ml^{-1}.mg protein^{-1})
DH5-a pBluescript	0.007
DH5-a pBluescript::UGP*ase*	0.090
G6MD2:pBluescript	0.010
G6MD2:pBluescript::UGP*ase*	0.042
G6MD2::pACUG	0.156
Pos. Control.	0.315
Neg. Control	0.004

Table 7 shows successful activity of the UGP*ase* gene in both DH5-α cells and G6MD2 cells through measureable G1P production whether present in pBluescript or pACYC 184.

3.3.4 Mutagenesis of *S. tuberosum SSI* gene

S. tuberosum SSI was randomly mutagenized to alter it so that the protein it produces utilizes UDP-glucose rather than ADP-glucose for starch synthesis. Random mutagenesis was performed using the mutator strain *E. coli* XL1-Red, and plasmid DNA was isolated from the pooled cultures (Sambrook and Russell 2001). To determine the rate and types of mutation, 33 samples of the mutated *SSI* library were sent for sequencing at the Macrogen (South Korea), using T7 Promoter (TAATACGACTCACTACTATAGGG) and T3 Promoter (AATTAACCCTCACTAAAGGG) primers. Sequence was compared using the SSI sequence (NM_001288145.1) as a template.

The results identified a total of 192 mutations; 98 point mutations (51%), 57 insertions (29.7%) and 37 (19.3%) deletions. The overall rate of mutation was

determined to be approximately 4 per kb, which corresponds to the reported mutation rate of this strain (Greener 1994; Greener & Callahan 1997).

3.3.5 Screening of mutated *S. tuberosum SSI* library

The screening of the *SSI* library was performed in the *E. coli* G6MD2 containing the UGP*ase* gene within the pACYC 184 vector (G6MD2::pACUG). This strain is unable to produce glycogen due to a deletion of the *glg* operon (Schwartz 1966). Cells were grown on a medium containing 1% (w/v) glucose to provide a carbon source. Colonies should produce linear glucans when starch synthase is expressed without starch branching enzymes, and iodine screening will lead to blue-staining colonies (Kossmann et al. 1999).

Over 150 000 clones of the *pBluescript:SSI* library were screened within *G6MD2::PACUG*, but all were unstained, meaning that no functional mutated *SSI* clone utilizing UGP*ase* for starch synthesis was identified. This method has previously been utilized to screen for functional starch synthases successfully (Kossmann et al. 1999), however, we were met with several difficulties that made large-scale and timely screening of the *SSI* library difficult, such as creating highly competent *G6MD2::PACUG* cells. Another potential problem was the 49% insertions and deletions obtained from the mutation process. These result in frame shifts in the amino acid sequence of the genes, which could result in the production of non-active proteins. In the future, a larger screening pool will improve the overall screening process. Repeat XL1-Red mutagenesis in order to lower the amount of insertions and deletions will also be beneficial, as there will be a lower chance of frame-shifts affecting gene expression.

3.4 Discussion

The pathways involved in starch synthesis form a complex network of regulations and modifications and their alteration can have a variety of results on starch structure and host plant physiology (Lorberth et al. 1998; Scheidig et al. 2002). In Chlorophyta, ADP-glucose conversion to G1P via AGP*ase* represents the first committed step in starch synthesis, and it has been demonstrated that altering AGP*ase* activity can cause various phenotypic changes in the mutant plant and alter starch granule formation (Smidansky et al. 2002; Smidansky et al. 2007; Hädrich et al. 2012). In Rhodophyceae the study of floridean starch, UDP-glucose and UGP*ase* has led to new findings and relations between the two separate pathways of starch synthesis (Viola et al. 2001; Shimonaga et al. 2007; Dauvillée et al. 2009), and this project aimed to test whether random mutagenesis could alter a higher plant starch synthase to utilize UDP-glucose instead of ADP-glucose. The starch synthase amino-acid sequences of four UDP-glucose-utilizing algae were found to contain similar binding motifs to the KTGGL binding motif and "look-alike" domain previously reported (Edwards et al. 1999), with some changes. As amino acids vary in charge, polarity and hydrophobicity, this could result in the different NDP-glucose specificity observed between higher plants such as potato and red algae. However, these results were not available to us during the beginning stages of the project, hence why random mutagenesis was the method utilized.

To develop a screen for the identification of UDP-glucose utilizing starch synthase genes, an *A. thaliana* UGP*ase* gene was amplified from cDNA and ligated into pBluescript.sk- plasmid. The UGP*ase* gene and the LAC promotor were both restricted from this plasmid and ligated into pACYC 184, before being transformed into the G6MD2 *E. coli* strain, which is unable to produce glycogen due to the deletion of the *Glg* operon. The UGP*ase* produced by the Arabidopsis gene was shown to be active within the *E. coli*, as previously reported (Kleczkowski et al. 2004). The *Solanum tuberosum* SSI gene was chosen to be mutated as it is the smallest of the higher plant starch synthases that is still active. Starch synthases contain a conserved C-terminal catalytic domain alongside N-terminal extensions which are variable in length between the different isoforms. These hydrophilic extensions are thought to alter the specificity of the starch synthases in terms of their

production of different chain lengths, but are not involved in substrate binding (Edwards et al. 1999). That means that the use of the smallest active starch synthase isoform in this screen would have the most chance of success as there would be a greater proportion of plasmids carrying genes without frame shift mutations than would be the case with a gene encoding larger isoform.

Using XL1-red *E. coli* we successfully mutated the SSI cDNA within pBluescript.SK. XL1-Red lacks 3 essential DNA repair pathways and as a result has a mutation rate approximately 5000 fold higher than wild-type *E. coli* (Greener 1994). Sequence analysis demonstrated a mutation rate of approximately 4 mutations per 1000 base pairs. Of these, 48% consisted of insertions and deletions, with the rest being point mutations. This presented a problem as these mutations could lead to frame shifts within our *SSI* library, which would likely lead to inactive protein (Wong et al. 2006). Despite this the plasmid was transformed into G6MD2::pACUG competent *E. coli* cells and screened using iodine vapour. G6MD2 colonies showing both the UGP*ase* gene and mutated starch synthase activity would synthesize a UDP-glucose utilizing polypeptide, accumulate linear glucans and stain blue.

Over 150 000 colonies were tested for starch synthase activity through UDP-glucose utilization, but none stained. This screening process has been confirmed to work in the past for identifying ADP-glucose utilizing enzymes (Kossmann et al. 1999). In the future a larger number of colonies will result in a more accurate representation of screening. Repeating the random mutagenesis of the *SSI* gene until the insertion and deletion number is smaller would also allow for a greater chance of gene activity. Other possibilities include the UGP*ase* activity being too low to cause any visual change due to iodine vapour. Another possibility could be differences in the ADP-glucose and UDP-glucose binding sites. Chlorophyta starch synthase and Rhodophyceae floridean starch synthase are part of the glycosyltransferase family (EC 2.4.1.21 2015; EC 2.4.1.183 2015) which are responsible for the conversion of NDP-glucose to NDP and the addition of glucose to the glucosyl residue of the amylose and amylopectin chain. As shown previously, red algal and higher plant starch synthases do not differ significantly in NDP-glucose binding domains in both binding motifs, but do differ in polarity, hydrophobicity, and charge. In the future, primer design for site-directed mutagenesis of the KTGGL binding motif could result

in a shift in chemical properties closer to those observed within red algae for UDP-glucose utilization.

4. Conclusion

As the rise in the human population leads to the increased consumption of foods and utilization of starch and starch-based products, previous methods of growth and cultivation may become inadequate if and when demand exceeds supply. As a result, alternative novel sources of starch need to be explored, as well as increased understanding of the mechanisms involved in starch synthesis for genetic modification. Although we were successful in causing random mutagenesis of the *Solanum tuberosum SSI* gene, our screening system was unable to find successful utilization of UDP-glucose for starch synthesis. In the future, larger colony screening numbers would be a better model for screening. Insertions and deletions in our random mutagenesis lead to frame-shifting within the gene, and could result in the expression of inactive protein. Reattempting mutagenesis for less insertions and deletions could lead to better chances of UDP-glucose utilization for starch synthesis, as well as site-directed mutagenesis of the ADP-glucose binding motifs.

Due to the rise of alternative starch sources such as cassava (Sriroth et al. 2000), the search for other sources requires the acquisition of data sets that can in the future, be used to accumulate new information, such as genetic linkage between species. The emergence of –omics technologies has led to an increased understanding of the intricacies surrounding the central dogma of molecular genetics. Post-transcriptional and post-translational modifications result in a wide variety of proteins and metabolites, leading to a very complex web of connections between each step within the central dogma (Franklin & Vondriska 2011). In order to gain a more accurate understanding of the pathways of molecular biology, it is necessary to understand the intricacies of each step (Sánchez-Pla et al. 2012; Keurentjes et al. 2008).

Utilizing next-generation sequencing technology, we successfully obtained transcriptomic data of *Tylosema esculentum* leaf and tuber, and using several *Glycine max* starch synthase reference genes, we located 3 open-reading frames within the marama bean transcriptome that showed starch synthase sequence comparison, all of which were successfully amplified from marama bean cDNA.

Future prospects include testing for activity of each starch synthase gene within an expression vector, as well as the assembly of the *Tylosema esculentum* transcriptome for future bioinformatics analysis.

References

Abel, G.J., Springer, F., Willmitzer, L., Kossmann, J. 1996. Cloning and functional analysis of a cDNA encoding a novel 139 kDa starch synthase from potato (*Solanum tuberosum* L.). *Plant Journal,* 10, pp.981–991.

Amonsou, E.O., Taylor, J.R.N., Beukes, M., Minnaar, A. 2012. Composition of marama bean protein. *Food Chemistry*, 130, pp.638–643.

Armstead, I. Huang, L., Ravagnani, A., Robson, P., Ougham, H. 2009. Bioinformatics in the orphan crops. *Briefings in Bioinformatics*, 10, pp.645–653.

Ball, S.G. 2005. Eukaryotic microalgae genomics. The essence of being a plant. *Plant physiology*, 137, pp.397–398.

Boehlein, S.K., Shaw, J.R., Hannah, L.C., Stewart, J.D. 2010. Probing allosteric binding sites of the maize endosperm ADP-glucose pyrophosphorylase. *Plant physiology*, 152, pp.85–95.

Cannon, S.B., May, G.D. & Jackson, S. 2009. Three sequenced legume genomes and many crop species: rich opportunities for translational genomics. *Plant physiology*, 151, pp.970–977.

Caspar, T., Tsan-Piao, L., Kakefuda, G., Benbow, L., Preiss, J., Somerville, C. 1991. Mutants of Arabidopsis with altered regulation of starch degradation. *Plant physiology*, 95, pp.1181–1188.

Chingwaru, W., Majinda, R.T., Yeboah, S.O., Jackson, J.C., Kapewangolo, P.T., Kandawa-Schulz, M., Cencic, A. 2011. *Tylosema esculentum* (Marama) Tuber and Bean Extracts Are Strong Antiviral Agents against Rotavirus Infection. *Evidence-based complementary and alternative medicine : eCAM*, p.284795.

Clédat, D., Bauut, S., Mikrini, R., Cardot, PJP. 2004. Rice starch granule characterization by flow cytometry scattering techniques hyphenated with sedimentation field–flow fractionationation. *J. Chromatography A*, 1049, pp.131–138.

Cock, J.H. 1982. Cassava: a basic energy source in the tropics. *Science*, 218, pp.755–762.

Crumpton-Taylor, M., Grandison, S., Png, K.M.Y., Bushby, A.J., Smith, A.M. 2012. Control of starch granule numbers in Arabidopsis chloroplasts. *Plant physiology*, 158, pp.905–16.

Dauvillée, D., Deschamps, P., Ral, J.P., Plancke, C., Putaux, J.L., Devassine, J., Durant-Terrasson, A., Devin, A., Ball, S. 2009. Genetic dissection of floridean starch synthesis in the cytosol of the model dinoflagellate *Crypthecodinium cohnii*. *Proceedings of the National Academy of Sciences of the United States of America*, 106, pp.21126–30.

Deacon, J., 2015. Nitrogen fixation. Available at: http://archive.bio.ed.ac.uk/jdeacon/microbes/nitrogen.htm [Accessed November 19, 2015].

Deschamps, P., Moreau, H., Worden, A.Z., Dauvillée, D., Ball, S. 2008. Early gene duplication within Chloroplastida and its correspondence with relocation of starch metabolism to chloroplasts. *Genetics*, 178, pp.2373–2387.

Deschamps, P. Haferkamp, I., Dauvillée, D., Haebel, S., Steup, M., Buléon, A., Putaux, J.L., Colleoni, C., d'Hulst, C., Plancke, C., Gould, S., Maier, U., Neuhaus, H.E., Ball, S. 2006. Nature of the Periplastidial Pathway of Starch Synthesis in the Cryptophyte *Guillardia theta*. *Eukaryotic Cell*, 5, pp.954–963.

EC 2.4.1.183, E.S.D., 2015. Alpha-1,3-glucan synthase, EC 2.4.1.183. Available at: https://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/enzymes/GetPage.pl?ec_number=2.4.1.183 [Accessed November 24, 2015].

EC 2.4.1.21, E.S.D., 2015. EC 2.4.1.21 - Starch synthase. Available at: https://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/enzymes/GetPage.pl?ec_number=2.4.1.21 [Accessed November 24, 2015].

Edner, C. et al. 2007. Glucan, Water Dikinase Activity Stimulates Breakdown of Starch Granules by Plastidial β-Amylases. *Plant Physiology*, 145, pp.17–28.

Edwards, A.. Borthakur, A., Bornemann, S., Venail, J., Denyer, K., Waite, D., Fulton, D., Smith, A., Martin, C. 1999. Specificity of starch synthase isoforms from potato. *The FEBS Journal*, 736, pp.724–736.

Ellis, R.P., Cochrane, M.P., Dale, M.F.B., Duffus, C.M., Lynn, A., Morrison, I.M., Prentice, R.D.M., Swanston, J.S., Tiller, S.A. 1998. Starch Production and Industrial Use. *Journal of the Science of Food and Agriculture*, 77, pp.289–311.

Franklin, S., Vondriska, T.M., 2011. Genomes, Proteomes, and the Central Dogma. *Circulation: Cardiovascular Genetics*, 4, pp.576–576.

Fujita, N., Satoh, R., Hayashi, A., Kodama, M., Itoh, R., Aihara, S., Nakamura, Y. 2011. Starch biosynthesis in rice endosperm requires the presence of either starch synthase i or IIIa. *Journal of Experimental Botany*, 62, pp.4819–4831.

Grada, A., Weinbrecht, K., 2013. Next-Generation Sequencing : Methodology and Application. *Journal of Investigative Dermatology*, 133, pp.1–4.

Greene, T.W., Hannah, L.C., 1998. Enhanced stability of maize endosperm ADP-glucose pyrophosphorylase is gained through mutants that alter subunit interactions. *Proceedings of the National Academy of Sciences of the United States of America*, 95, pp.13342–13347.

Greener A, C.M., 1994. XL1-RED: A highly efficient random mutagenesis strain. *Strategies*, 7, pp.32–34.

Greener A, Callahan M, J.B., 1997. An efficient random mutagenesis technique using an E. coli mutator strain. *Molecular Biotechnology*, 7, pp.185–195.

Greener, C.M., 1994. XL1-RED: a highly efficient random mutagenesis strain. *Strategies*, 7, pp.32–34.

Grimaud, F., Rogniaux, H., James, M.G., Myers, A.M., Planchot, V. 2008. Proteome and phosphoproteome analysis of starch granule-associated proteins from normal maize and mutants affected in starch biosynthesis. *Journal of Experimental Botany*, 59, pp.3395–3406.

Hädrich, N., Hendriks, J.G., Kotting, O., Arrivault, S., Feil, R., Zeeman, S.C., Gibon, Y., Schulze, W.X., Stitt, M., Lunn, J.E. 2012. Mutagenesis of cysteine 81 prevents dimerization of the APS1 subunit of ADP-glucose pyrophosphorylase and alters diurnal starch turnover in *Arabidopsis thaliana* leaves. *The Plant Journal*, 70, pp.231–242.

Hansen, P.I., Spraul, M., Dvortsak, P., Larsen, F.H., Blennow, A., Motawia, M.S., Engelsen, S.B. 2009. Starch phosphorylation--maltosidic restrains upon 3'- and 6'-phosphorylation investigated by chemical synthesis, molecular dynamics and NMR spectroscopy. *Biopolymers*, 91, pp.179–93.

Hirose, T., Aoki, N., Harada, Y., Okamura, M., Hashida, Y., Ohsugi, R., Akio, M., Hirochika, H., Terao, T. 2013. Disruption of a rice gene for α-glucan water dikinase, OsGWD1, leads to hyperaccumulation of starch in leaves but exhibits limited effects on growth. *Frontiers in Plant Science*, 4, pp.1–9.

Holse, M., Larsen, F.H., Hansen, A., Engelsen, S.B. 2011. Characterization of marama bean (*Tylosema esculentum*) by comparative spectroscopy: NMR, FT-Raman, FT-IR and NIR. *Food Research International*, 44, pp.373–384.

Holse, M., Husted, S. & Hansen, Å., 2010. Chemical composition of marama bean (*Tylosema esculentum*)-A wild African bean with unexploited potential. *Journal of Food Composition and Analysis*, 23, pp.648–657.

Inoue, H., Nojima, H. & Okayama, H., 1990. High efficiency transformation of *Escherichia coli* with plasmids. *Gene*, 96, pp.23–28.

Johnson, P.E., Patron, N.J., Bottrill, A.R., Dinges, J.R., Fahy, B.F., Parker, M.L., Waite, D.N., Denyer, K. 2003. A low-starch barley mutant, Risø 16, lacking the cytosolic small subunit of ADP-glucose pyrophosphorylase, reveals the importance of the cytosolic isoform and the identity of the plastidial small subunit1. *Plant Physiology*, 131, pp.684–696.

Keeling, P.L. & Myers, A.M. 2010. Biochemistry and genetics of starch synthesis. *Annual Review of Food Science and Technology*, 1, pp.271–303.

Keurentjes, J.J., Koornneef, M. & Vreugdenhil, D., 2008. Quantitative genetics in the age of omics. *Current Opinion in Plant Biology*, 11, pp.123–128.

Kleczkowski, L., Geisler, M., Ciereszko, I., Johansson, H. 2004. UDP-glucose pyrophosphorylase. An old protein with new tricks. *Plant Physiology*, 134, pp.912–918.

Koornneef, M. & Meinke, D., 2010. The development of Arabidopsis as a model plant. *The Plant journal : for cell and molecular biology*, 61, pp.909–21.

Kossmann, J., Abel, G.J., Springer, F., Lloyd, J.R., Willmitzer, L. 1999. Cloning and functional analysis of a cDNA encoding a starch synthase from potato (*Solanum tuberosum* L.) that is predominantly expressed in leaf tissue. *Planta*, 208, pp.503–511.

Kötting, O., Kossmann, J., Zeeman, S.C., Lloyd, J.R. 2010. Regulation of starch metabolism: the age of enlightenment? *Current Opinion in Plant Biology*, 13, pp.320–328.

Lorberth, R., Ritte, G., Willmitzer, L., Kossmann, J. 1998. Inhibition of a starch-granule-bound protein leads to modified starch and repression of cold sweetening. *Nature Biotechnology*, 16, pp.473–477.

Martin, J.A. & Wang, Z. 2011. Next-generation transcriptome assembly. *Nature reviews. Genetics*, 12, pp.671–682.

Mazimba, O., Majinda, R.T., Modibedi, C., Masesane, I.B., Cencic, A., Chingwaru, W. 2011. *Tylosema esculentum* extractives and their bioactivity. *Bioorganic & Medicinal Chemistry*, 19, pp.5225–5230.

Mitchell, R.A., Keys, A.J., Madqwick, P.J., Parry, M.A., Lawlor, D.W. 2005. Adaptation of photosynthesis in marama bean *Tylosema esculentum* (Burchell A. Schreiber) to a high temperature, high radiation, drought-prone environment. *Plant Physiology and Biochemistry*, 43, pp.969–976.

Morell, M.K., Rahman, S., Abrahams, S.L., Appels, R. 1995. The Biochemistry and Molecular Biology of Starch Synthesis in Cereals. *Functional Plant Biology*, 22, pp.647–660.

Palovaara, J., Saiga, S. & Weijers, D., 2013. Transcriptomics approaches in the early Arabidopsis embryo. *Trends in Plant Science*, 18, pp.514–521.

Patron, N.J. & Keeling, P.J., 2005. Common Evolutionary Origin of Starch Biosynthetic Enzymes in Green and Red Algae1. *Journal of Phycology*, 41, pp.1131–1141.

Quagen, 2015. CLC Genomics Workbench. , pp.1–4. Available at: http://www.clcbio.com/products/clc-genomics-workbench/ [Accessed November 24, 2015.

Ral, J.-P., 2006. Circadian Clock Regulation of Starch Metabolism Establishes GBSSI as a Major Contributor to Amylopectin Synthesis in *Chlamydomonas reinhardtii*. *Plant Physiology*, 142, pp.305–317.

Roldán, I. et al. 2007. The phenotype of soluble starch synthase IV defective mutants of *Arabidopsis thaliana* suggests a novel function of elongation enzymes in the control of starch granule formation. *Plant Journal*, 49, pp.492–504.

Sakulsingharoj, C. et al. 2004. Engineering starch biosynthesis for increasing rice seed weight: The role of the cytoplasmic ADP-glucose pyrophosphorylase. *Plant Science*, 167, pp.1323–1333.

Sambrook, J. & Russell, D., 2001. Molecular Cloning: A Laboratory Manual. *Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY*, p.999.

Sánchez-Pla, A., Reverter, F., Ruiz de Villa, M.C., Comabella, M. 2012. Transcriptomics: mRNA and alternative splicing. *Journal of Neuroimmunology*, 248, pp.23–31.

Scheidig, A. Frohlich, A., Schulze, S., Lloyd, J.R., Kossmann, J. 2002. Downregulation of a chloroplast-targeted β-amylase leads to a starch-excess phenotype in leaves. *Plant Journal*, 30, pp.581–591.

Schwall, G.P. et al. 2000. Production of very-high-amylose potato starch by inhibition of SBE A and B. *Nature biotechnology*, 18, pp.551–554.

Schwartz, M., 1966. Location of the maltose A and B loci on the genetic map of *Escherichia coli. Microbiology*, 92, pp.1083–1089.

Severin, A.J. et al. 2010. RNA-Seq Atlas of *Glycine max*: a guide to the soybean transcriptome. *BMC Plant Biology*, 10, p.160.

Shelembe, J.S., Cromarty, D., Bester, M.J., Minnaar, A., Duodu, K.G. 2012. Characterisation of phenolic acids, flavonoids, proanthocyanidins and antioxidant activity of water extracts from seed coats of marama bean [*Tylosema esculentum*] - an underutilised food legume. *International Journal of Food Science & Technology*, 47, pp.648–655.

Shimonaga, T. Fujiwara, S., Kaneko, M., Izumo, A., Nihei, S., Francisco, P.B jr., Satoh, A., Fujita, N., Nakamura, Y., Tsuzuki, M. 2007. Variation in storage alpha-polyglucans of red algae: amylose and semi-amylopectin types in Porphyridium and glycogen type in Cyanidium. *Marine biotechnology*, 9, pp.192–202.

Smidansky, E.D., Clancy, M., Meyer, F.D., Lanning, S.P., Blake, N.K., Talbert, L.E., Giroux, M.J. 2002. Enhanced ADP-glucose pyrophosphorylase activity in wheat endosperm increases seed yield. *Proceedings of the National Academy of Sciences of the United States of America*, 99, pp.1724–1729.

Smidansky, E.D., Meyer, F.D., Blakeslee, B., Weglarz, T.E., Greene, T.W., Giroux, M.J. 2007. Expression of a modified ADP-glucose pyrophosphorylase large subunit in wheat seeds stimulates photosynthesis and carbon metabolism. *Planta*, 225, pp.965–76..

Sriroth, K., Piyachomkwan, K., Wanlapatit, S, Oates, C.G. 2000. Cassava starch technology: the Thai experience. *Starch - Stärke*, 52, pp.439–449.

Stadnichuk, I.N., Semenova, L.R., Smirnova, G.P., Usov, A.L. 2007. A highly branched storage polyglucan in the thermoacidophilic red microalga *Galdieria maxima* cells. *Applied Biochemistry and Microbiology*, 43, pp.78–83.

Stiller, J.W. & Hall, B.D., 1997. The origin of red algae: implications for plastid evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 94, pp.4520–4525.

Stitt, M. & Zeeman, S.C., 2012. Starch turnover: pathways, regulation and role in growth. *Current Opinion in Plant Biology*, 15, pp.282–92.

Streb, S., Delatte, T., Umhang, M., Eicke, S., Schorderet, M., Reinhardt, D., Zeeman, S.C. 2008. Starch granule biosynthesis in Arabidopsis is abolished by removal of all debranching enzymes but restored by the subsequent removal of an endoamylase. *The Plant cell*, 20, pp.3448–3466.

Swart, C., 2013. *Production of Libraries to Study Biopolymer Metabolism in Arabidopsis thaliana and Tylosema esculentum*. University of Stellenbosch [Accessed November 24, 2015].

Szydlowski, N. et al. 2009. Starch granule initiation in Arabidopsis requires the presence of either class IV or class III starch synthases. *The Plant cell*, 21, pp.2443–2457.

Takundwa, M., Nepolo, E., Chimwamurombe, P.M., Cullis, A.C., Kandawa-Schulz, M.A., Kunert, K. 2010. Development and use of microsatellites markers for genetic variation analysis, in the Namibian germplasm, both within and between populations of marama bean (*Tylosema esculentum*). *Journal of Plant Breeding and Crop Science*, 2, pp.233–242.

Tatge, H., Marshall, J., Martin, C., Edwards, E.A., Smith, A.M. 1999. Evidence that amylose synthesis occurs within the matrix of the starch granule in potato tubers. *Plant, Cell and Environment*, 22, pp.543–550.

Tetlow, I.J., Beisei, K.G., Cameron, S., Makhmoudova, A., Lui, F., Bresolin, N.S., Wait, R., Morell, M.K., Emes, M.J. 2008. Analysis of protein complexes in wheat amyloplasts reveals functional interactions among starch biosynthetic enzymes. *Plant Physiology*, 146, pp.1878–1891.

Tiessen, A., Hendriks, J.H.M., Stitt, M., Branscheid, A., Gibon, Y., Farre, E.M., Geigenberger, P. 2002. Starch synthesis in potato tubers is regulated by post-translational redox modification of ADP-glucose pyrophosphorylase: a novel regulatory mechanism linking starch synthesis to the sucrose supply. *The Plant cell*, 14, pp.2191–213.

Tomlinson, K.L., Lloyd, J.R. & Smith, A.M., 1997. Importance of isoforms of starch-branching enzyme in determining the structure of starch in pea leaves. *Plant Journal*, 11, pp.31–43.

Viola, R., Nyvall, P. & Pedersén, M., 2001. The unique features of starch metabolism in red algae. *Proceedings. Biological sciences / The Royal Society*, 268, pp.1417–22.

Vriet, C. et al. 2010. A suite of *Lotus japonicus* starch mutants reveals both conserved and novel features of starch metabolism. *Plant Physiology*, 154, pp.643–655.

Wong, T.S., Roccatano, D., Zacharias, M., Schwaneberg, U. 2006. A statistical analysis of random mutagenesis methods used for directed protein evolution. *Journal of Molecular Biology*, 355, pp.858–871.

Yu, S. Blennow, A., Bojko, M., Madsen, F., Olsen, C.E., Engelsen, S.B. 2002. Physico-chemical characterization of floridean starch of red algae. *Starch - Starke*, 54, pp.66–74.

Zeeman, S.C., Umemoto, T., Lue, W.L., Au-Yeung, P., Martin, C., Smith, A.M., Chen, J. 1998. A mutant of Arabidopsis lacking a chloroplastic isoamylase accumulates both starch and phytoglycogen. *The Plant cell*, 10, pp.1699–1712.

Zeeman, S.C., Kossmann, J. & Smith, A.M., 2010. Starch: its metabolism, evolution, and biotechnological modification in plants. *Annual Review of Plant Biology*, 61, pp.209–234.

Zeeman, S.C., Smith, S.M. & Smith, A.M., 2007. The diurnal metabolism of leaf starch. *Biochemical Journal*, 401, pp.13–28.

Zhou, G., Willett, J.L. & Carriere, C.J., 2000. Temperature dependence of the viscosity of highly starch- filled poly(hydroxy ester ether) biodegradable composites. *Rheologica Acta*, 39, pp.601–606.