# Functional identification of a putative stachyose synthase (StaS, Medtr7g106910.1) from *Medicago truncatula*, by overexpression in the *Arabidopsis* stachyose deficient double mutant *atrs4/atrs5*

**by**

**Melt Hugo**

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




**Supervisor:** Dr. S.W. Peters
**Co-supervisors:** Dr. B. Loedolff & Dr. D.T. Guzha

Department of Genetics

**December 2018**

## 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.

Melt Hugo

Date**: December 2018**

## Abstract

The Raffinose Family Oligosaccharides (RFOs; Suc-Galn, 13 < n ≥ 1) are α1,6-galactosyl extensions of sucrose occurring exclusively in the plant kingdom and some photoautotrophic bacteria. This unique group of sugars are widely implicated as storage compounds in sink tissues, phloem translocates, and as target molecules that help combat abiotic and biotic plant stresses in various species. The RFO biosynthetic pathway is well characterised and RFOs are synthesised from sucrose by the successive addition of galactose moieties by α-1,6 galactosyltransferases viz. galactinol synthase (GolS, EC 2.4.1.123), raffinose synthase (RafS, EC 2.4.1.82), and stachyose synthase (StaS, EC 2.4.1.67). Amino acid sequence alignments between functionally identified RafS and StaS proteins indicate that the major difference between them is the presence of a conserved motif between amino acid positions 340 to 420 (absent in RafS proteins). The predicted protein sequence of StaS from the model legume - *Medicago trancatula,* Medtr7g106910.1 (designated MtStaS) contains this motif. To explore the functional identity of these carbohydrates in legumes, cDNA encoding stachyose synthase (StaS) which transfers a galactosyl moiety from galactinol to the $C_6$ position of the galactose moiety in raffinose (Raf) to yield the tetrasaccharide stachyose (Sta), was identified and cloned from *M. truncatula.* As part of a multipronged strategy to functionally characterise MtStaS, we performed the following experiments. We **(i)** identified a candidate *MtStaS* gene through rudimentary bioinformatic analyses. We then examined *MtStaS* transcript abundance in a variety of *M. truncatula* organs and concluded that *MtStaS* expression is tissue-specific **ii)** cloned the candidate gene into a binary vector *pMDC32* (dual CaMV35s promoter) and transformed this construct into the *Arabidopsis thaliana atrs4* (devoid of detectable Sta) *and atrs4.atrs5* (devoid of detectable Raf and Sta) T-DNA insertion mutants in an attempt to restore the RFO metabolism. We confirmed that MtStaS is able to recover ablated Sta in *atrs4* mature seeds and **(iii)** cloned *MtStaS* and subsequently characterised it in the dimorphic fungus - *Yarowia lipolytica*. We established that it is a *bona fide* StaS that possess no bifunctionality in synthesising both Raf and Sta, contradictory to StaS from *Arabidopsis thaliana* (AtStaS) which can synthesise both.

## Opsomming

Raffinose Familie Oligosakkariede (RFOs; Suc-Galn, 13 < n ≥ 1) is α1,6-galaktosiel verlengings van sukrose wat slegs in die plant koninkryk en ‘n seleksie van foto-outotrofiese bakterieë voorkom. Die unieke suikers vervul kritiese rolle in plante en dien as storing molekule in die sink weefsel, floëem translokasie molekules, en as teiken molekule wat abiotiese en biotiese spanning teen te veg. Die RFO biosintetiese padweg is goed gekarakteriseer en RFOs word geproduseer vanaf sucrose deur die opeenvolgende byvoeging van galaktosiel molekules deur α-1,6 galaktosielltransferases viz. galaktinol sintase (GolS, EC 2.4.1.123), raffinose sintase (RafS, EC 2.4.1.82), en stachyose sintase (StaS, EC 2.4.1.67). Vergelykings tussen die aminosuur volgordes van funksionele RafS en StaS proteïne toon groot ooreenstemming behalwe vir ‘n bewaarde gebied tussen posisie 330 en 410 wat afwesig is van RafS proteïne. Die gebied is teenwoordig in die vermeende proteïen volgorde van die StaS van *Medicago trancatula,* Medtr7g106910.1 (aangewese MtStaS). Stachyose sintase (StaS) word vermoed om tetrasakariede stachyose (Sta) te produseer deur die oordrag van ‘n galaktosiel molekule vanaf galaktinol na raffinose (Raf) te kataliseer. Om die funksionele rol van die suikers in peulplante verder te ondersoek was stachyose sintase (StaS) vanaf die kDNS van *M. truncatula* geïdentifiseer en gekloneer. Om MtStaS ten volle funksioneel te karakteriseer was die volgende eksperimente uitgevoer. Ons het (i) die kandidaat *MtStaS* geen deur rudimentêre bioinformatiese analise geïdentifiseer. Die hoeveelheid *MtStaS* geen uitdrukking in verskeie *M. truncatula* organe was ook bepaal, wat aangedui het *MtStaS* geen uitdrukking is weefsel spesifiek. Daarna is (ii) *MtStaS* gekloneer in *pMDC32* (dubbele CaMV35s promotor) en die konstruk getransformeer in *Arabidopsis thaliana atrs4* (geen waarneembare Sta) en *atrs4.atrs5* (geen waarneembare Raf en Sta) om sodoende te bepaal of RFO metabolisme herstel kon word. Ons het bevestig dat *atrs4* mutante wat gekomplimenteer is met MtStaS wel Sta in sade kan produseer. Laastens was (iii) MtStaS ook in uitgedruk en gekarakteriseer in *Yarowia lipolytica*. Ons het bevestig dat MtStaS ‘n *bona fide* StaS is wat slegs Sta produseer en nie bifunksionaliteit toon soos die StaS van *Arabidopsis thaliana* (AtStaS).

## Acknowledgments

I would like to sincerely acknowledge and thank everyone who has supported me throughout this endeavour.

Firstly, Dr Shaun Peters, my supervisor, who has proven to be the golden thread in my tertiary education development. The support on all fronts have been unparalleled – from leadership to scientific advice. Thank you for imparting your wellspring of knowledge, experience and ideas to me. Your unwavering belief and support in me have enabled me to grow as a scientist and as a leader. Thank you for showing me that this journey goes beyond science. To Dr Loedolff for her around the clock support throughout the yeast and LC-MS/MS work.

To Mikey and Janto for all the breakaway jogs in Jonkershoek whenever I needed it most. Tapiwa for the robust and critical dialogues that shaped me into a lateral thinker. Henry for providing me with the scientific enthusiasm, technical support and goodwill needed to get me through this journey.

To everyone that made IPB my temporary home, provided inspirational discussions and made the long hours seem like a short stay. To (Dr) Simo Maduna for the laughs, road trips and friendship. To Gaby for her unrelenting support and understanding when the skies got especially dark.

To the leadership structures that I have had the honour to serve on during my candidature, in particular, Student Parliament, Faculty of Sciences' Student Committee, the Academic Affairs Council and the Students' Representative Council – thank you for understanding whenever I needed to prioritize my academic obligations.

And finally, to my parents, for everything.

*"But it was only fantasy. The wall was too high, as you can see.*
*No matter how he tried, he could not break free"* – Pink Floyd, the Wall, 1979.

# Table of Contents

# List of figures

## Chapter I



## Chapter II



## Chapter III

# List of tables

## Chapter I



## Chapter II



## Chapter III



## Chapter IV

## List of abbreviations

| | |
|---|---|
| **AtGolS** | *Arabidopsis thaliana* galactinol synthase |
| **α-Gal** | α-galactosidase enzyme |
| **BLAST** | Basic Local Alignment Search Tool |
| **bp** | Base pair |
| **cDNA** | Complementary DNA |
| **Col-0** | *Arabidopsis thaliana* ecotype Columbia-0 |
| **DTT** | Dithiothreitol |
| **Gal** | Galactose |
| **g-DNA** | Genomic DNA |
| **GGT** | Galactan:galactan galactosyltransferase |
| **Gol** | Galactinol |
| **GolS** | Galactinol synthase |
| **ha** | hectares |
| **Ino** | *myo*-inositol |
| **kb** | Kilobase |
| **MS** | Murashige and Skoog |
| **N** | Nitrogen |
| **PCR** | Polymerase-chain-reaction |
| **PMSF** | Phenylmethylsulfonylfluoride |
| **PVP** | Polyvinylpyrrolidone |
| **Raf** | Raffinose |
| **RafS** | Raffinose synthase |
| **RFOs** | Raffinose family oligosaccharides |
| **RNA** | Ribonucleic acid |
| **qPCR** | Quantitative polymerase-chain-reaction |
| **SDS** | Sodiumdodecylsulphate |
| **SEM** | Standard error of mean |
| **Sf9/21** | Insect cell line from *Spodoptera frugiperda* |

| | |
|---|---|
| **SIP** | Seed imbibition proteins |
| **Sta** | Stachyose |
| **StaS** | Stachyose synthase |
| **Suc** | Sucrose |
| **T-DNA** | Transferred DNA |
| **v** | Volume |
| **w** | Weight |
| **v/v** | Volume-to-volume solution |
| **Ver** | Verbascose |
| **WSC** | Water soluble carbohydrates |

# Chapter I: General introduction, Literature Review, Research Aims and Objectives

## 1.1 An introduction to *Medicago truncatula*

### 1.1.1 Classification, Phylogeny and Evolutionary History of *Medicago*

*Medicago* (Latin: Medica, or “from Media” a former civilisation in the south western region of Asia) consists of a genus comprised of 87 species (Allen and Allen 1981; Steele *et al.* 2010). These species evolved from their perennial ancestors throughout the tertiary era (Lesins and Lesins 1979) . The species are further classified into the sections of Buceras, Cartiensae, Dendrotelis, Geocarpae, Heynianae, Hymenocarpos, Lunatae, Lupularia, Medicago, Orbicularus, Platycarpae and Spirocarpos (Steele *et al.* 2010). In the broader *Leguminosae* family, *Medicago* groups as part of the Trifolieae tribe. The tribe is located on the IRLC (inverted repeat-lacking clade) of the Papilionoideae subfamily (Lewis 2005; Figure 1).

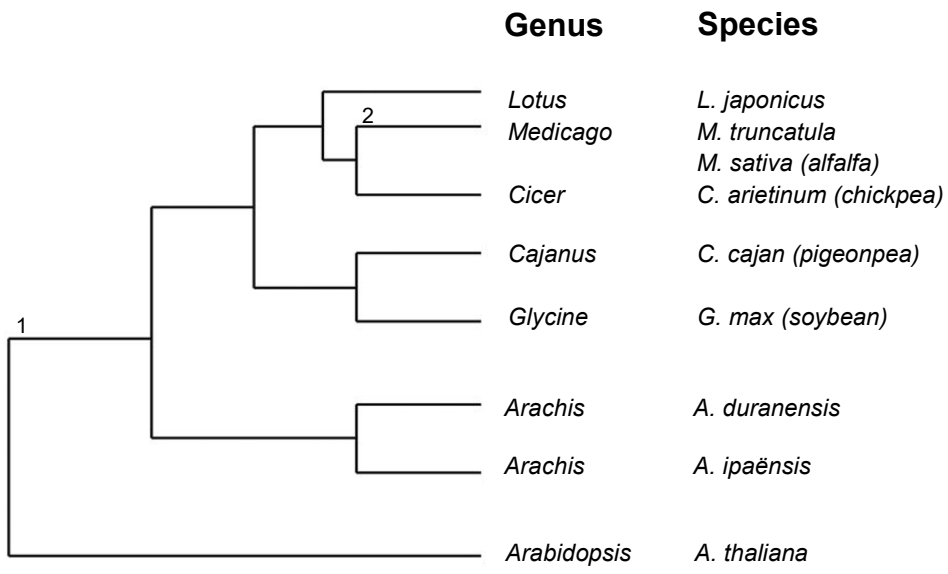


**Figure 1.** Phylogenetic tree of seven major legumes 1) Papilionoideae subfamily and 2) Trifolieae tribe. Adapted and modified from Song *et al.* 2016.

Generally, *Medicago* species are considered to exhibit exceptional tolerance to abiotic stresses such as salinity or drought, enabling it to survive in soils under severe environmental stresses. Furthermore, contributing to its high agronomical value is its potential as an excellent forage crop and the mutualistic capacity to associate with *Rhizobium* and other soil bacteria which form specialised root nodules on the root surfaces (Lesins and Lesins 1979). With the ability to reduce atmospheric nitrogen to ammonia ($N_2$ to $NH_3$), the bacteria housed in these nodules facilitate the growth of *Medicago* in nitrogen-deprived soils (Graham *et al.* 2003).

**Table 1.** Classification of *Medicago truncatula* (Lesins and Lesins 1979).

| | |
|---|---|
| **Kingdom:** | *Plantae* |
| **Phylum:** | Tracheophyta |
| **Class:** | Magnoliopsida |
| **Order** | Fabales |
| **Family:** | Leguminosae |
| **Taxon Name:** | *Medicago truncatula Gaertn.* |
| **Common Name/s:** | English – Barrel Clover, Barrel Medic, Strong-Spined Medick; French - Luzerne tronquée |
| **Synonym/s:** | *Medicago tentaculata Willd.*<br>*Medicago tribuloides Desr.*<br>*Medicago uncinata Willd.* |

### 1.1.2 Distribution, Ecology and Commercial Importance

*Medicago truncatula* is a species that is native to the Mediterranean region, characterised by a climate that yields hot, low rainfall summers, and wet winters (Perry 1997; Kottek *et al.* 2006). Winter seasons in Mediterranean climates are further characterised into rainfall that fluctuates between 250 mm and 480 mm per annum whereas summer rainfall is approximately a third of that (Quinlivan 1965; Perry 1997).

Heavy rains, hail, snow and gale storms are common occurrences during winter, whilst in contrast, summer seasons are hot and dry (Perry 1997; Kottek *et al.* 2006).

Indigenous populations have been sampled in Europe (Spain, Italy, Greece, France, Ukraine), Asia-temperate (Cyprus, Jordan, Turkey, Georgia, Armenia), North Africa (Libya, Egypt, Tunisia, Algeria, Morocco) as well as on the Capri, Crete, Corsica, Malta, Sardinia and Sicily islands (Lesins and Lesins, 1979; Delalande *et al.*, 2007; Figure 2). *M. truncatula* has been found to be well adapted to the Mediterranean climatic conditions in all these countries and has been found to survive and acclimate in different regions of the world that possess analogous conditions.

**Figure 2.** Global native distribution of *Medicago truncatula*. Adapted from Lesins and Lesins 1979; Delalande *et al.* 2007.

Considered as a small herbaceous annual that can grow 150-800 mm long, *M. truncatula* is also identified by the leaves (Figure 3B). As characteristic of the Trifolieae tribe, leaves are made up of three smaller obovate (ovate with the narrower end towards the base) shaped leaflets which can grow approximately 7-21 mm wide and 8-27 mm long (Lesins and Lesins 1979). The apical segment of each leaflet is jagged with wedge-shaped teeth interspersed between large and small triangular teeth (Figure 3C). An anthocyanin pigment may form in the lamina's centre of each leaflet (Bucciarelli *et al.* 2006). Compact, spiky pods house 5-8 seeds on average dependent on growth conditions. Seeds are further protected by a tough coating that must to be scarified for germination and imbibition purposes (Garcia *et al.* 2006; Figure 3A).


**Figure 3.** Morphological features of *Medicago truncatula*. A) seed pod B) stem with trifoliate leaves C) sink, intermediate and source trifoliate leaves.

Leguminous plants are widely incorporated in today's cultivated pastures because of the duality that they provide in serving as a forage crop for farm livestock and providing the soil with a reliable and sustainable source of nitrogen (Swanepoel and Tshuma 2017). Legume pastures have been shown to fix from 35 to 165 kg of nitrogen per hectare yearly (Peoples *et al.* 2001; Smýkal *et al.* 2015). On an annual basis, it is assessed that cultivated legumes fix an estimated 40-60 million tonnes of N, which permits savings of nearly USD 7-10 billion spent on nitrogen fertiliser (Smil 1999; Graham *et al.* 2003). It is unsurprising then that some native rangelands around the world and in particular – South Africa are increasingly being converted into cultivated pastures with the intent to improve livestock productivity and contribute to increased food security (Swanepoel *et al.* 2016).

Of the *Medicago* genus, the most economically important species remains the perennial *Medicago sativa* (Gholami *et al.* 2014). Estimated to be grown on 32 million hectares of land, globally, *M. sativa* is regarded as one of the most vital forage crops in temperate regions around the world (Yuegao and Cash 2009; Vasileva and Kostov 2015). *M. sativa* is also regarded as the third highest valued crop in North America, where it is grown on an estimated 11.9 million ha of land. In South America, it is grown on 7.0 million ha of land with Europe marginally more at 7.12 million ha (Yuegao and Cash 2009).

In Australia and South Africa, *M. truncatula* serves a dual purpose as an important forage and crop rotation plant (Puckridge and French 1983; O'Neill and Bauchan 2000; Choi *et al.* 2004). When traditional supplies are insufficient, *M. truncatula* can be utilised to supply forage and soil nitrogen, in addition it is often used in a rotational system with cash crops including oats (*Avena sativa*), canola (*Brassica napus*), barley (*Hordeum vulgare*), and wheat (*Triticum aestivum*; Swanepoel *et al.* 2016). In the Western Cape, *M. truncatula* and *M. polymorpha* are the two most well adapted species of *Medicago* and are considered to play key roles in conservation agriculture in the province because of its utilisation and self-re-establishment as a pasture crop (Nichols *et al.* 2007).

In South Africa, *M. truncatula* can be grazed by cattle but are most effectively grazed by sheep because of their nimble mouths – allowing for a more effective seedpod retrieval from the soil (Swanepoel and Tshuma 2017). Grazing also causes some seedpods to be trampled into the soil, effectively assisting in re-establishment of the crop (Puckridge and French 1983).

### 1.1.3 *Medicago truncatula* is a model organism

The term "model organisms" refers to refer to organisms that have been used in numerous studies because of their advantageous characteristics, have increased the efficiency of experimental and laboratory research and are considered to have directly benefited humankind (Hedges 2002).

Despite *M. sativa* being significantly and economically more valuable than *M. truncatula*, various challenges made the use of *M. sativa* difficult. Genetic and physiological studies for key traits are greatly complicated by allogamy, autotetraploidy and a large genome size contribute to these challenges (Thoquet *et al.* 2002). Considered as a model organism, *M. truncatula* exhibits several advantageous features over *M. sativa*, including its small diploid (2n=2x=16) sequenced genome size (Blondon *et al.*, 1994), regeneration by somatic embryogenesis and the ease of genetic transformation by *Agrobacterium tumefaciens* (Barker *et al.* 1990; Chabaud *et al.* 1996). *M. truncatula*'s short generation time, self-fertilisation and rapid seed production have also been regarded as significant advantages (Cook 1999).

Higher levels of nucleotide sequence conservation and similar genetic organisation to *M. sativa*, *Cicer arietinum* (chickpea) and *Trifolium subterraneum* (Subterranean Clover) allows for the transfer of *M. truncatula* genomic information to these species possible (Thoquet *et al.* 2002). Therefore, it serves as the best candidate for the new era of 'Omics studies in legumes (Gholami *et al.* 2014). The latest sequenced and refined *M. truncatula* genome comprising of approximately 465 megabases (Mb) is freely accessible (Mt4.0; Tang *et al.* 2014).

Bloat in ruminants is a common and a highly undesirable phenomenon found in legume pastures and animal feeds (Martínez-Villaluenga *et al.* 2008). A class of complex carbohydrates known as the Raffinose Family Oligosaccharides (RFOs) are responsible for causing the bloat (Coon *et al.* 1990). Presence of RFOs in diets interfere with the digestion of other nutrients as well as reduce available dietary energy (Martínez-Villaluenga *et al.* 2008). Soybean RFOs are kept at a low level because of the digestive instabilities that is created in chickens, baby pigs and even dogs (Hartwig *et al.* 1997). Significant improvement in the digestion of all amino acids and overall nutritional value of a lupin diet (Family: Fabaceae) in both poultry (van Barneveld 1999) and rainbow trout (Glencross *et al.* 2003) is observed with the extraction of RFOs. Reducing RFO concentration would accordingly lead to a rise in true metabolizable energy available from legume pastures and seed meals for livestock and is highly sought after by the feed industry. *M. truncatula* lays the foundation as the preferred model organism for many significantly more economically important legumes (*M. sativa, G. max* etc.).

## 1.2 Raffinose Family Oligosaccharides (RFOs): are plant specific α-1, 6 galactosyl extensions of sucrose

### 1.2.1 The RFOs play important roles in carbon translocation, storage and stress tolerance

Typically, plants store fixed carbon as starch and/or translocate it as sucrose. However, many plant species also inherently possess the molecular architecture to synthesise other classes of carbohydrates, with the raffinose family of oligosaccharides (RFOs) being one of the most abundant (Kandler and Hopf 1982; Keller and Pharr 1996). Integral to plant growth and development, RFOs play an important role in plant growth/development and are implicated in various physiological mechanisms such as seed germination, photoassimilate translocation, biotic and abiotic stress tolerance, seed storability, and seed desiccation tolerance (Horbowicz and Obendorf 1994; Blöchl *et al.* 2007; Martínez-Villaluenga *et al.* 2008; Nishizawa-Yokoi *et al.* 2008; Dinant and Lemoine 2010). RFOs are found in various plant organs such as leaves, stems, tubers, bulbs, fruit, seeds and is also widely implicated in carbon transport (Keller and Pharr 1996). RFOs serve as long- or short-term storage reserves. In over 500 species, belonging to almost 100 families, sucrose is found in all phloem exudates, and RFOs were in two thirds of these (Peters *et al.* 2007; Dinant and Lemoine 2010). In some species tested, RFOs are the main transport sugars and their synthesis were proposed to be part of the polymer trap mechanism for symplasmic phloem loading (Turgeon 1996). RFOs are also proposed to play a key role in reducing solute leakage during long-distance transport (Ayre *et al.* 2003).

In both natural and agricultural conditions, plants are regularly exposed to environmental stresses. Stress tolerance is therefore the plant's fitness mechanisms to cope with an unfavourable environment. The prominent accumulation of RFOs throughout seed development and maturation is believed to serve a critical role as osmoprotectants in providing desiccation tolerance, longevity in the dehydrated state and vigour upon ensuing germination (Blackman *et al.* 1992; Corbineau *et al.* 2000; Downie *et al.* 2003; Pukacka *et al.* 2009; Angelovici *et al.* 2010). The stress-inducible galactinol synthase (GolS) is the first committed enzyme in the biosynthesis

pathway of RFOs and is responsible for the accumulation of galactinol (Gol) and raffinose (Raf) when exposed to a myriad of abiotic stresses (Taji *et al.* 2002). Galactinol is further implicated as a signalling molecule upon wounding and pathogenetic attacks (Kim *et al.* 2008).

### 1.2.2 Biosynthesis of RFOs is a multi-enzymatic process dependent on galactose donor molecules

Raffinose Family Oligosaccharides (RFOs) are synthesised from sucrose by the successive addition of galactose moieties by α-1,6 galactosyltransferases viz. galactinol synthase (GolS, EC 2.4.1.123; raffinose synthase (RafS, EC 2.4.1.82) and stachyose synthase (StaS, EC 2.4.1.67), respectively. GolS, RafS, and StaS are all cytosolic, while the subsequent galactinol-independent RFO chain elongation appears to take place in the vacuole and is catalysed by a different enzyme - galactan:galactan galactosyltransferase GGT (Bachmann *et al.* 1994).

#### 1.2.2.1 Galactinol synthase (GolS; EC 2.4.1.123)

The galactosyl residue of UDP-D-galactose is transferred to *myo*-inositol, yielding Gol, which is the specific galactosyl donor to the RFOs, and as such the biosynthesis of galactinol is considered the crucial regulatory step in its synthesis (Keller and Pharr 1996). Most of the studies that have biochemically characterised galactinol synthases (GolSs) are originally from cucurbit leaves and legume seeds (Keller and Pharr 1996), though more recently, GolS has been isolated and described from an assortment of additional plants, such as *Glycine max* (soybean), *Phaseolus vulgaris* (kidney bean), *Cucurbita pepo* (zucchini squash), *Cucumis sativus* (cucumber) and *Ajuga reptans* (common bugle) (Smith *et al.* 1991; Bachmann *et al.* 1994; Liu *et al.* 1995; Ribeiro *et al.* 2000; Wakiuchi *et al.* 2003)

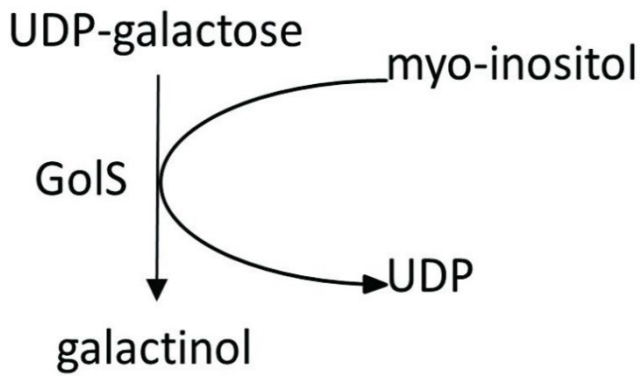


**Figure 4.1.** UDP-Galactose + *myo*-inositol → galactinol in the presence of galactinol synthase (GolS; EC 2.4.1.123).

### 1.2.2.2 Raffinose synthase (RafS; EC 2.4.1.82)

Raffinose synthase catalyses the reversible galactosylation of sucrose from Gol, resulting in Raf and *myo*-inositol. RafSs have successfully been isolated and characterised from a range of plants including *Cucumis sativus* (CsRafS; ABD72603.1), *Pisum sativum* (PsRafS; CAD20127), *Arabidopsis thaliana* (AtRafS; BAB11595), *Oryza sativa* (OsRafS; XP_015621501; Peterbauer *et al.* 2002b; Li *et al.* 2007; Sui *et al.* 2012; Gangl *et al.* 2015).

The OsRafS from *Oryza sativa* (Asian rice) was cloned and expressed in *Escherichia coli,* exhibiting optimal activity at 45°C, a pH of 7.0 and has a requirement for a sulfhydryl group for suitable catalytic activity (Li *et al.* 2007). Possessing similar biochemical requirements, PsRafS was cloned and inserted in *Spodoptera frugiperda* (Sf21) and shown to exhibit an optimal pH of also 7.0 and possesses similar kinetic properties (Peterbauer *et al.* 2002b).

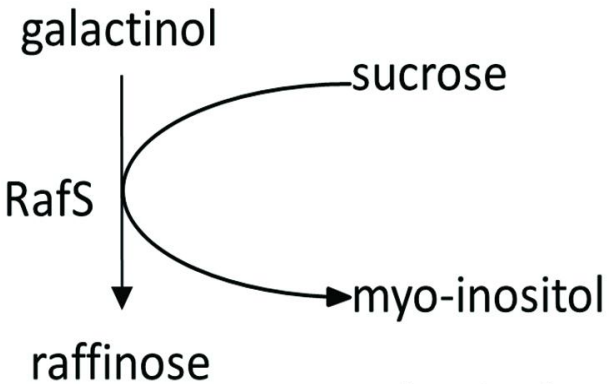


**Figure 4.2.** Galactinol + sucrose → raffinose in the presence of raffinose synthase (RafS; EC 2.4.1.82).

### 1.2.2.3 Stachyose synthase (StaS; EC 2.4.1.67)

Stachyose is also found to be the principal RFO in most plant species (Sosulski *et al.* 1982; Quemener and Brillouet 1983; Andersen *et al.* 2005; Martínez-Villaluenga *et al.* 2008; Martín-Cabrejas *et al.* 2008; Huynh *et al.* 2008; Dilis and Trichopoulou 2009; Wang *et al.* 2010). Stachyose synthases have previously been isolated purified and characterised from *Pisum sativum* (PsStaS; XP_013450269), *Vigna angularis* (VaStaS; CAB64363), *Cucumis melo* (CmStaS; XP_008451468), *Arabidopsis*

*thaliana* (AtStaS; NP_192106) (Holthaus and Schmitz 1991; Hoch *et al.* 1999; Peterbauer *et al.* 1999, 2002b; Gangl *et al.* 2015). From these, two have been identified to possess additional biosynthetic capacity (*in vitro*). StaS from *Arabidopsis thaliana* has been reported to be a sequential bifunctional (Gol-dependent) RafS and a high affinity StaS (Gangl *et al.* 2015). AtStaS utilises substrates Suc and Gol to produce raffinose and stachyose, whereas substrates Raf and Gol produced stachyose, only.

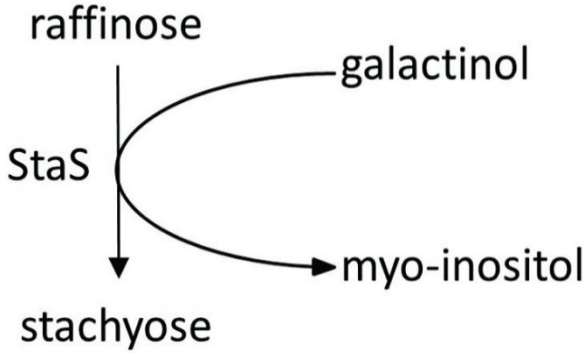


**Figure 4.3.** Raffinose + galactinol → stachyose in the presence of stachyose synthase (StaS; EC 2.4.1.67).

### 1.2.3 Impact of RFOS on Human and Animal Health

Intestinal mucosa in monogastric animals and humans are unable to digest RFOs because they lack α-galactosidase (hydrolytic enzyme; Reddy *et al.* 1984; Kumar *et al.* 2010). A lower intestinal pH, large amounts of hydrogen, carbon dioxide, small quantities of short chain fatty acid and methane are produced by the large intestinal microflora when metabolising RFOs (Krause *et al.* 1994; Naczk *et al.* 1997). In the small intestines, however, RFOs are able to escape digestion and absorption (Saunders and Wiggins 1981). Unwanted gut cramps, bloating stomach, eructation and abdominal pain in human and monogastric animals lacking α-galactosidase are some of the unwanted symptoms caused by deleterious gases produced by bacteria which make up an approximate ¾ of flatulence (Kurbel *et al.* 2006; Swennen *et al.* 2006).

Using the Gumbmann and Williams procedure, a positive association was found between hydrogen production and RFO content (specifically Raf and Sta) in the California Small White beans (CSW; *Phaseolus vulgaris*; Family Fabaceae) in rats

(Gumbmann and Williams 1971). Unusually, increased levels of Raf fed at levels greater than 6.7% resulted in a decline in hydrogen production and is the result of osmotic catharsis. This led to a pressure imbalance within the small intestines before the conclusion of fermentation and hydrolysis by the intestinal microflora (Wagner *et al.* 1976). A decrease in the absorption capacity of the small intestine could be caused by this osmotic pressure imbalance (Saunders and Wiggins 1981). In human subjects, osmotic catharsis induced by raffinose have also been previously observed (Gitzelmann and Auricchio 1965).

### 1.3 T-DNA Insertional Mutagenesis

Gene function is often studied by creating genetic mutations in genes and subsequently studying the phenotype (O'Malley and Ecker 2010). Insertional mutagenesis, TILLING (Targeting Induced Local Lesions in Genomes)/chemical mutagenesis are popular techniques employed to achieve loss of gene function. The most desirable approach to understanding gene function is through complete inactivation of the gene of interest that in turn results in an alteration of the phenotype (Bouché and Bouchez 2001). Successful point mutations have been introduced into the *Arabidopsis* genome through the use of ethyl methanesulfonate (EMS) by chemical mutagenesis (Greene *et al.* 2003). The drawback of this method, however, lies in the shortage of cost effective and direct techniques of screening for mutations in the specific gene(s) of interest from a mutagenized population (Gilchrist and Haughn 2010).

Disruption in gene function and subsequent loss of function can also be achieved through the insertion of foreign T-DNA (Transferred DNA) into the specific gene of interest. The use of T-DNA or other transposable elements are frequently used to study gene function in *Arabidopsis* (Krysan *et al.* 1999; Alonso *et al.* 2003). T-DNAs offer advantages over transposons in that they are both physically and chemically stable throughout multiple generations, and will not transpose within the genome, subsequent to their integration (Krysan *et al.* 1999). The main purpose of T-DNA techniques is two-fold. Primarily, the functioning and expression of the gene(s) of interest should be disrupted. Secondary, mutations can be identified through T-DNA's role as a marker or "tag". This means that T-DNA insertions within the coding region

of a gene have a high probability in completely abolishing gene expression and functioning dependent on the precise insertion site (Bolle *et al.* 2011). Insertions within genes that are essential to the plant will often result in lethality which would make examining these genes more difficult (Gilchrist and Haughn 2010).

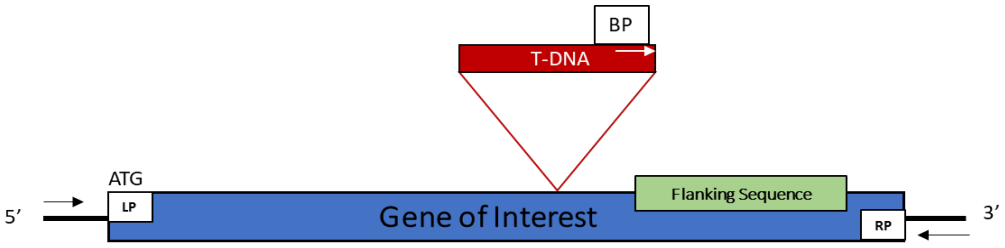


**Figure 5.1. Simplified depiction of T-DNA insertion in the gene of interest and the respective primers needed for genotyping.** A specific primer pair combination is used to genotype (Figure 5.2) for the presence of either homozygous or heterozygous wild-type/T-DNA alleles. LP: Left genomic primer of gene of interest; RP: Right genomic primer of gene of interest; BP: T-DNA border primer. The figure was redrawn based on http://signal.salk.edu/tdnaprimers.2.html (last retrieved: April 2018)

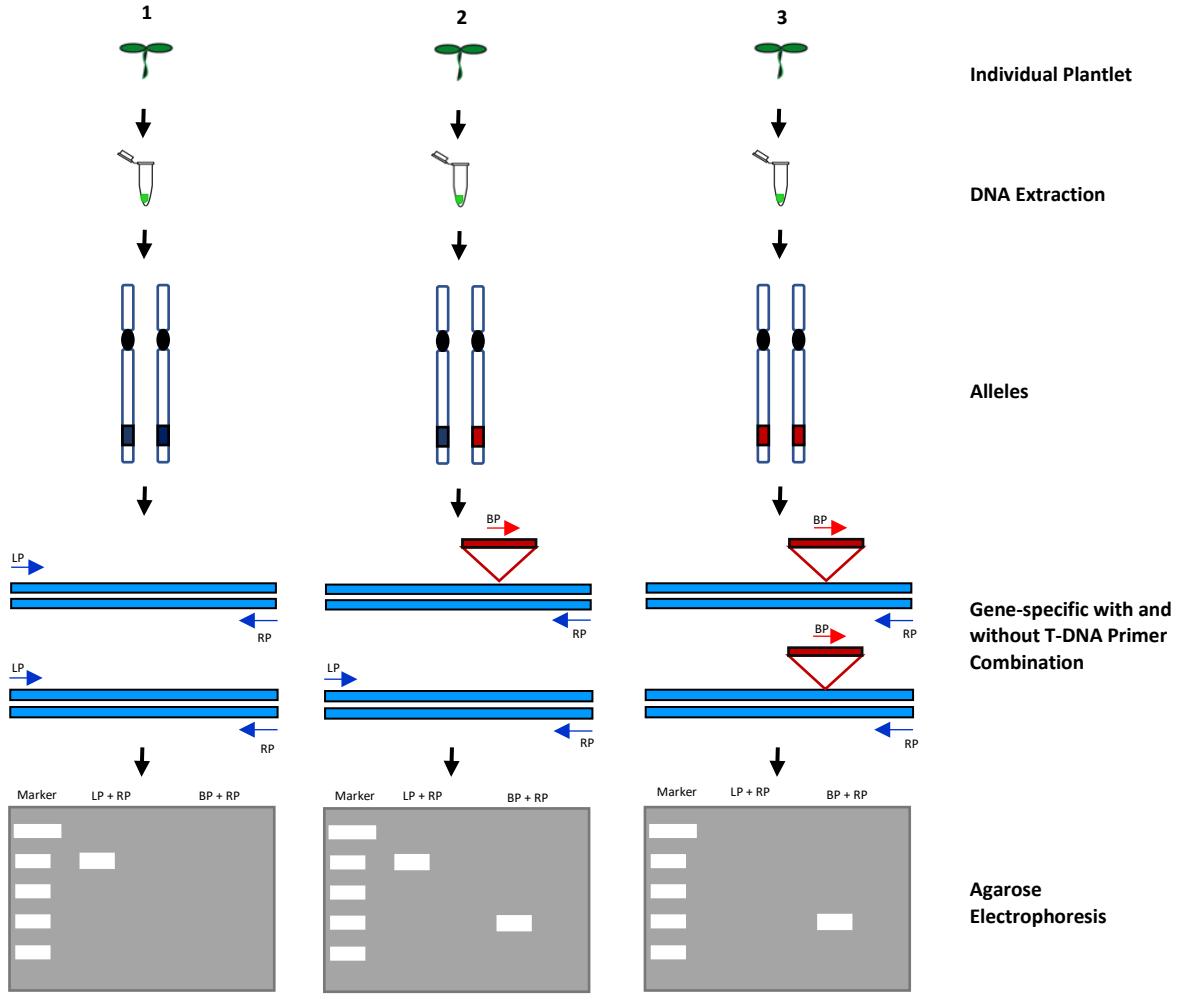


**Figure 5.2. Genotyping segregating T-DNA insertion lines to identify homozygous plantlets.** Here, an example of three individual plants were examined and labelled (1, 2, and 3). A single leaf is cut from multiple progeny seedlings of a T-DNA line and transferred to separate individual tubes. DNA is then extracted from these leaves (I) PCR (II) and gel (agarose) electrophoresis (III) is subsequently performed to genotype the individual DNA samples. A primer pair specific to regions flanking the insertion site is used to check for the presence of a wild type, undisrupted allele of the gene (Figure 3). A separate PCR reaction using T-DNA-specific (red arrow, BP + RP) and gene-specific primer pairs (blue arrow, LP + RP) are used to test for the presence of a T-DNA insertion in the gene. A wild type plantlet (depicted chromosome with two blue loci, individual: 1) will produce a gene-specific product only (size ±1 kb). A homozygous plant (depicted chromosome with two blue loci, individual: 3) will produce a T-DNA insertion product (size ±0.5 kb), but no wild type product. A heterozygous plantlet (depicted in chromosome as both red and blue loci, individual 2) will produce both a T-DNA insertion product and a wild type product. This figure was remodelled based on O'Malley and Ecker (2010)

## 1.4 The state of RFOs in *Arabidopsis thaliana*

Accumulation of RFOs throughout seed development and maturation is believed to serve a critical role as osmoprotectants in providing longevity in the dehydrated state, desiccation tolerance, and vigour upon subsequent germination (Downie *et al.* 2003; Salvi *et al.* 2016; Li *et al.* 2017). Gol, Raf and Sta have all been reported to accumulate during *Arabidopsis* seed development (Ooms *et al.* 1993; Bentsink *et al.* 2000; Nishizawa *et al.* 2008; Gangl and Tenhaken 2016). Whereas Gol and Raf but never Sta have been shown to accumulate in the leaves of *Arabidopsis* (Taji *et al.* 2002; Egert *et al.* 2013; Gangl and Tenhaken 2016).

Ten Galactinol Synthase isoforms (*GolS*) have been reported in *Arabidopsis thaliana* of which seven putative *GolS* (*referred to as AtGolS1*, *2*, *3*, *4*, *5*, *6*, and *7*) isoforms have been defined by their common signature amino acid site characterised by 'APSAA' (Taji *et al.* 2002; Nishizawa *et al.* 2008). In an attempt to functionally characterise and elucidate their physiological roles, three of these isoforms have been implicated in response to abiotic stresses such as drought, high salt and cold stress, respectively (*AtGolS1, AtGolS2, AtGolS3;* Taji *et al.* 2002; Nishizawa *et al.* 2006). Interestingly, *AtGols1* is also implicated as a possible signalling molecule in synthesising Gol linked to pathogen induced-responses (Kim *et al.* 2004, 2008; Cho *et al.* 2010). The same authors demonstrated that the overexpression of the *AtGolS2* in *Arabidopsis* augmented the levels of galactinol and raffinose, and improved drought tolerance.

In *Arabidopsis*, five members make up the *RafS* gene family (*atrs1*, *atrs2*, *atrs4*, *atrs5* and *atrs6), atrs3* was initially included in this gene family but has since been regarded as a pseudogene. The common name given to any enzyme that biosynthesises Raf is called RafS, suggesting a common biochemical function (Knaupp *et al.* 2011). Numerous biochemical investigations into the *Arabidopsis* RafS family, however, prove this practice highly incorrect. For example, *AtRS2* possesses no RafS capability but rather α-galactosidase activity (Egert *et al.* 2013). Total loss of abiotic stress induced accumulation of Raf was observed in mutant *Arabidopsis* plants with an *atrs5* knock-out (Zuther *et al.* 2004). This led to the conclusion that *atrs5* is the only RafS that is able to biosynthesise Raf under a range

of different abiotic stresses. Unpredictably, Raf content only decreased by ~50% in *atrs5* gene knock-out seeds, suggesting that at least a second RafS was implicated in synthesising the remaining Raf (Egert *et al.* 2013). This remained unproven, until recently when double knock-out mutants were created that led to the conclusion that atrs4 (later renamed to AtStaS) is a bifunctional RFO synthase with both StaS and RafS capabilities under certain conditions. It was shown that *atrs4.atrs5* knock-out seeds were unable to synthesise both Raf and Sta (Loedolff *et al.* 2015; Gangl *et al.* 2015). When studied *in vitro*, atrs4 proved to be a bifunctional StaS with RafS capacity (Gangl *et al.* 2015). A clear distinction exists amongst the various *RafS* gene family isoforms. In conclusion, the common RafS name can be ambiguous as to the exact biochemical function which each of the five RafS possess and that caution should be taken when assuming RafS involvement when making predictions based upon gene sequence data from other *atrs* genes. It remains imperative that we do not solely rely on bioinformatic predictions and annotations of gene functions but rather as a tool for identification to use biochemical and genetic experiments to further study gene functions *in vivo* and *in vitro* and either complement or refute the putative bioinformatic annotations.

## 1.5 Research Aims and Objectives

Stachyose is a major sucrose-oligosaccharide occurring in *Medicago spp*. Despite the availability of a genome sequence, there are no reports on the molecular identities of the genes which lead to RFO accumulation in *M. truncatula*. This study sought to identify and functionally characterise a *M. truncatula stachyose synthase* (*MtStaS*).

To this end, the study aimed to **(i)** identify a candidate gene through rudimentary bioinformatic analyses using known StaSs from *Arabidopsis thaliana* (AtStaS, At4g01970) and pea (PsStaS, genbank acc. CAC38094), **(ii)** clone the candidate gene into the binary vector *pMDC32* (dual CaMV35s promoter) and transform this construct into the *A. thaliana atrs4* (compromised in Sta accumulation) *and atrs4.atrs5* (compromised in Raf and Sta accumulation) T-DNA insertion mutants and **(iii)** heterologously express MtStaS in the dimorphic fungus *Yarowia lipolytica* in order to biochemically characterise the recombinant protein, and particularly to ascertain if MtStaS has bifunctionality in synthesising RFOs.

## 1.6 Scientific Contributions during Masters Candidature (2017 - 2018)

### 1.6.1 Publications

° Hugo, M., Loedolff, B., Guzha, DT., and Peters, SW. 2018. Functional identification of a putative stachyose synthase (StaS, Medtr7g106910.1) from *Medicago truncatula*. *In preparation.*

### 1.6.2 Presentations (Presenting author is underlined)

° Hugo, M, Loedolff, B, Guzha, DT, Peters, SW. 2018. Functional identification of a putative stachyose synthase (StaS, Medtr7g106910.1) from *Medicago truncatula*, by overexpression in the *Arabidopsis* stachyose deficient double mutant *atrs4/atrs5*. The South African Academy for Science and Arts symposium, 2-3 November 2017 (University of Pretoria). *Best presentation award.*

# Chapter II: Phylogenetic and gene expression analysis of *Medicago truncatula* α1,6-galactosyltransferases

## 2.1 INTRODUCTION

Raffinose family oligosaccharides (RFOs; Suc-Gal$_n$, 13 < n ≥ 1) are exclusive to the plant kingdom and are synthesised by α1,6-galactosyltransferases (Kandler and Hopf 1982). Catalysing the transfer of galactosyl moieties by means of the uncommon cyclitol-carbohydrate hybrid galactinol (Gol) serving as the galactosyl donor. The biosynthesis of Gol is characterised by galactinol synthase (GolS; inositol 3-α-galactosyltransferase; EC: 2.4.1.123), using *myo*-inositol and UDP-Gal as substrates (Keller and Pharr 1996). Raffinose (Raf, Suc-Gal$_1$) and Stachyose (Sta, Suc-Gal$_2$) are synthesised by raffinose synthase (RafS; galactinol-sucrose galactosyltransferase; EC 2.4.1.82) and stachyose synthase (StaS; galactinol-raffinose galactosyltransferase; EC 2.4.1.67) in a Gol-dependent manner, respectively.

Moreover, in the framework of RFO physiology, Sta is the highest detectable RFO oligomer in *Arabidopsis* seeds (Bentsink *et al.* 2000). Together with Raf, these two RFOs occur in high abundance only in mature seeds and are rapidly mobilised during the germination process (Downie *et al.* 2003; Salvi *et al.* 2016; Li *et al.* 2017). RFOs (Raf) are detectable in only trace amounts, in vegetative tissue of plants. However, upon exposure to an array of abiotic stresses, very noticeable mass increases in both Gol and Raf has been reported to occur in leaves (Taji *et al.* 2002; Panikulangara *et al.* 2004; Nishizawa *et al.* 2008). Seemingly, these increases are attributed to an elevated activity of RafS enzyme(s). Most emphasis, however, has been directed towards GolS in relation to stress-induced Raf mass increases, possibly because GolS is considered to be the rate limiting step in RFO biosynthesis (Kaplan *et al.* 2004).

Raffinose family oligosaccharides fulfil various critical roles in abiotic stress tolerance, seed development and desiccation tolerance (Downie *et al.* 2003). Their expression profiles, however, remain largely unknown in *M. truncatula.* This chapter describes the rudimentary bioinformatic analysis, which led to the identification of genes involved in RFO accumulation. With the focal point on the putative *StaS* from *M. truncatula* (*MtStaS*; Medtr7g106910.1) - a single gene annotated as either a *RafS* or a *StaS*, identified at the genome-wide scale.

This work also lays a foundation for understanding tissue-specificity and localisation of RFO gene expression profiles in different organs and provides valuable information for the identification of candidate genes: *MtGolS, MtRafS* and *MtStaS* for further functional analysis.

## 2.2 MATERIALS AND METHODS

Unless specified otherwise, chemicals used throughout this study were obtained from Sigma-Aldrich® (www.sigmaaldrich.com/south-africa.html) or MERCK® (Modderfontein, South Africa). The Oligo explorer® software (V1.4 BETA) was used to design the primers used and were subsequently synthesised by Inqaba Biotech®. All enzymes used in this study were obtained from New England Biolabs® (NEB, Inqaba Biotechnical Industries (Pty) Ltd, South Africa), unless stated otherwise.

### 2.2.1 Identification and phylogenetic analysis of *M. truncatula* stachyose synthase (MtStaS)

To identify the putative *M. truncatula StaS* (*MtStaS*; Medtr7g106910.1) - four protein sequences reported to have shown StaS activity *in vitro* (PsStaS, VaStaS, CmStaS, and AtStaS; Holthaus and Schmitz 1991; Peterbauer *et al.* 1999a; Pluskota *et al.* 2015; Gangl *et al.* 2015, respectively) were used as BLAST queries against the *M. truncatula*, Mt4.0 cDNA database (Tang *et al.* 2014). Amino acid sequence alignments were then performed and aligned using ClustalW (Thompson *et al.* 1994) in the Geneious software package [version 11.1.4 available from www.geneious.com (Biomatters, Ltd.)] using the default parameters.

The phylogeny of MtStaS was determined using ten StaSs, five RafSs and rooted against the known alkaline α-Galactosidase AtSIP2 (*A. thaliana*; NP_191311; Peters *et al.* 2010). The phylogenetic tree was constructed using the Neighbor-Joining method (Saitou and Nei 1987). The robustness of resultant trees in which the associated taxa are clustered together in the bootstrap test (1 x $10^6$ replicates) are indicated next to each branch (Felsenstein 1985). Branches consistent with partitions reproduced in <50% bootstrap replicates were collapsed. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to

infer the phylogenetic tree. The evolutionary distances were computed using the p-distance method (Nei and Kumar 2000) and are in the units of the number of amino acid differences per site. Overall, the analysis involved 16 protein sequences. Evolutionary analyses were conducted in MEGA X (Kumar *et al.* 2018).

### 2.2.2 Plant material and growth conditions

Seeds of *M. truncatula* (2n=2x=16, 1C value = 0.48 pg) cv. Jemalong line J5 (A17) (The Samuel Roberts Noble Foundation, USA) served as the source material. Chemical scarification was performed as defined in the *M. truncatula* Handbook (Garcia *et al.* 2006). Briefly, seeds were immersed in anhydrous sulphuric acid ($H_2SO_4$) with intermittent agitation (10 min) to remove the seed coating. Seeds were then rinsed in sterile deionised water before being surface sterilised by soaking in sodium hypochlorite (15% v/v, 5 min), followed by five rinses with sterile deionised water. Seeds were vernalised for 2 days at 4°C in darkness and germinated on hormone-free MS (Duchefa, Labretoria, South Africa) media, adjusted to pH 5.7 prior to the addition of 0.8% agar. Plantlets were transferred onto peat disks (Jiffy™ no.7, South Africa) and propagated under controlled environmental conditions in growth chambers (Snijders Labs, Economic deluxe; 8 h light, 120 µmol photons $m^{-2}$ $s^{-1}$, 22°C, 16 h dark, 18°C, 60% relative humidity).

### 2.2.3 RNA extraction and cDNA synthesis

Total RNA was extracted from 80 mg of plant and seed material using the Maxwell® 16 LEV simplyRNA Purification Kit in the Maxwell® 16 Instrument (AS2000; Promega, Anatech, South Africa), according to the manufacturer's instruction. Vegetative organs (roots, stems, sink, intermediate and source trifoliate leaves) were harvested 25 days after planting. Roots were harvested 1.5 cm below the hypocotyl and consisted of the entire root system including the laterals. Leaf material did not contain any petioles and buds were excluded from the stems. Briefly, RNA was harvested in parallel from three biological replicates. Each biological replicate consisted out of a pool of tissues that were harvested from three independent plants as previously described (Laurie *et al.* 2011; Yeoh *et al.* 2013).

The quantity and purity of RNA was measured using a NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific, Inqaba biotech, South Africa). RNA quality was evaluated on denaturing (37% formaldehyde [v/v]) electrophoretic gels stained with ethidium bromide (1.0%; w/v; 60 V). Purified RNA was subsequently stored at -80°C until further use. The complementary DNA (cDNA) template was synthesised by reverse transcription of 1 µg total RNA with an oligo ($dT_{15}$) primer and M-MLV ($H^-$) reverse transcriptase (Promega) following the manufacturer's protocol.

### 2.2.4 Transcript analysis

The templates used in the qPCR experiments represented 1:10 volumetric dilutions of first strand cDNA. All samples evaluated were prepared in parallel and three independent experiments were executed. The qPCR reactions were conducted using the PowerUp™ SYBR™ Green Master Mix (Applied Biosystems, Life Technologies, South Africa) that exhibits a SYBR™ green dye fluorescent signal (excitation at 497nm and emission at 520nm). Using the QuantStudio 3 Real-Time PCR System (Applied Biosystems), qPCR reactions were conducted in 10 µl reactions (1 µl cDNA, 5 µl qPCR master mix and 0.3 µmol of each primer). Cycling conditions were performed according to the recommended thermal profile: initial denaturation step at 95°C for 10 min, followed by 40 cycles of a two-step denaturation/annealing process (95°C, 15 s/60°C, 1 min). Primers were designed with the following parameters: temperature optimums of 60°C and amplicon lengths of 87 to 97 bp, yielding 20 nucleotide primer sequence lengths. Primers were also designed to span exon-exon junctions towards the 3' ends of genes where possible. Melt curves were examined to ensure primer specificity and all primer combinations showed efficiencies greater than 1.8.

Three reference genes (Table 1; Kakar *et al.* 2008; Sinha *et al.* 2015), *UBC*, *ACT2* and *18S* were used in the analyses. The threshold cycle number ($\Delta C_T$) was used to calculate relative fold change with the $\Delta\Delta C_T$ method, using the intermediate trifoliate leaf as the calibrator sample (Livak and Schmittgen 2001). The mean $C_T$ value of three technical replicates were analysed for every biological replicate. *18S* was utilised as the reference gene in all analyses, and changes in mRNA levels relative to *18S* were confirmed using *UBC* and *ACT2* as alternate reference genes in

independent experiments. All qPCR experiments were conducted in accordance with the “Minimum Information for Publication of Quantitative Realtime PCR Experiments” (MIQE, Bustin *et al.* 2009).

### 2.2.5 Statistical analysis

Statistical analyses were performed using GraphPad Prism software (GraphPad Prism version 7.04 for Mac OS X, GraphPad Software, La Jolla California USA, www.graphpad.com). Experimental values are expressed as the mean ± standard error of mean (SEM) of three independent experiments. Significant differences between two groups were identified using the nonparametric one-way ANOVA on ranks (Kruskal–Wallis test) followed by Dunnett’s multiple comparisons test. Mean differences were considered significant at $P < 0.05$.

**Table 1.** Primer pairs used for quantitative real-time PCR (qPCR) analyses. Three independent reference genes, *UBC*, *ACT2* and *18S* were selected for normalisation of the data.

| Name | Designed target / annotation | Sequence (5’ - 3’) | Amplicon size (bp) | Reference |
|---|---|---|---|---|
| UBC_Q_F<br>UBC_Q_R | ubiquitin-conjugating enzyme | CTGACAGCCCACTGAATTGTGA<br>TTTTGGCATTGCTGCAAGC | 100 | Kakar *et al.* 2008 |
| ACT2_Q_F<br>ACT2_Q_R | Actin-2 | TCAATGTGCCTGCCATGTATGT<br>ACTCACACCGTCACCAGAATCC | 100 | Kakar *et al.* 2008 |
| 18S_Q_F<br>18S_Q_R | 18Sr-RNA | CCACTTATCCTACACCTCTC<br>ACTGTCCCTGTCTACTATCC | 102 | Sinha *et al.* 2015 |
| MtStaS_Q_F<br>MtStaS_Q_R | *M. truncatula* Stachyose synthase | AGGTGGTGGGAATTTCCTTG<br>TTTCCATCACCTAGCCACTC | 97 | This study |
| MtRafS_Q_F<br>MtRafS_Q_R | *M. truncatula* Raffinose synthase | ATGCTAAACAATGGTGGGGC<br>AGACCCTCATCTCACCAGCA | 90 | This study |
| MtGolS_Q_F<br>MtGols_Q_R | *M. truncatula* Galactinol synthase (Accession XM_003625957) | TTGGCCAAAGGAAATGGGTC<br>ACGACCTACCTCAGACAATGC | 87 | This study |

## 2.3 RESULTS

### 2.3.1 Functionally expressed α1,6-galactosyltransferases share identities with MtStaS

To identify *M. truncatula* StaS (MtStaS), A BLASTp (www.ncbi.nlm.nih.gov/BLAST/) search was performed using AtStaS, VaStaS, PsStaS, and CmStaS as queries against the *M. truncatula* Mt4.0 cDNA database. These StaS protein sequences showed high homology to a single *M. truncatula* protein (Medtr7g106910.1), and this protein shared a greater evolutionary relationship with StaSs than RafSs from other species (Figure 1B). A reciprocal BLAST was performed and the search returned top hit homologous sequences in *Pisum sativum, Cucumis melo*, *Alonsoa meridionalis*, *Vigna angularis* and *Arabidopsis thaliana* of which all code for StaSs, in addition to homologous sequences coding for RafSs in *P. sativum*, *M. truncatula*, *Oryza sativa* and *Glycine max*. An alignment of MtStaS against these functionally confirmed StaSs and RafSs, revealed that the protein shared the highest amino acid identity (84%) with *Pisum sativum* (PsStaS) and the lowest (40%) with *Oryza sativa* (OsRafS). A section of the sequence alignment revealed an 80 amino acid long sequence conserved motif, shared in StaSs but markedly absent in RafSs (Figure 1A).

### 2.3.2 Phylogenetic Analysis of MtStaS

Based upon the alignment of RafS and StaS proteins, a phylogenetic tree was subsequently constructed in order to deduce evolutionary relationships existing amongst these proteins (Figure 1B). The tree was rooted against the known alkaline α-Galactosidase (AtSIP2; Peters *et al.* 2010). The evolutionary distances were computed using the p-distance method (Nei and Kumar 2000). Evolutionary distances were also computed using the Jukes-Cantor model (Jukes and Cantor 1969) yet, yielded similar results (data not shown). The putative MtStaS protein shared a greater and more common evolutionary relationship with StaSs as opposed to RafSs from other species (Figure 1B), thus further indicating that MtStaS most likely encodes a StaS rather than a RafS.

### 2.3.3 Expression analysis of RFO synthases in *M. truncatula*

In order to understand StaS as well as the contextualisation of the physiological role it may play in RFO metabolism of *M. truncatula,* the accumulation of *GolS*, *RafS* and *StaS* transcripts have been studied and compared using quantitative Realtime PCR analyses. Total RNA was extracted from a range of tissues (Figure 2A) obtained from *M. truncatula* plants. Expression profiles were analysed in seeds, roots, stems, sink, intermediate and source trifoliate leaves. The two additional reference genes used (*UBC* and *ACT2*) for the purpose of transcript normalisation, yielded similar results. High levels of *GolS* and *StaS* transcript abundance was observed in seed tissue with *StaS* transcript abundance in root tissue approximately a tenth of that found in the seeds. In contrast, *RafS* transcript levels were deficient in seeds.

In root tissue, both *GolS* and *StaS* levels are approximately the same, but interestingly, high levels of *RafS* is observed. *GolS* and *RafS* transcript levels in stem tissue are relatively equal with *RafS* levels being slightly more, however, almost no *StaS* levels are detected in this tissue. In remaining tissues (sink and source trifoliate leaves), *GolS*, *RafS* and *StaS* transcript levels are non-significant in comparison with each other (Figure 2B).

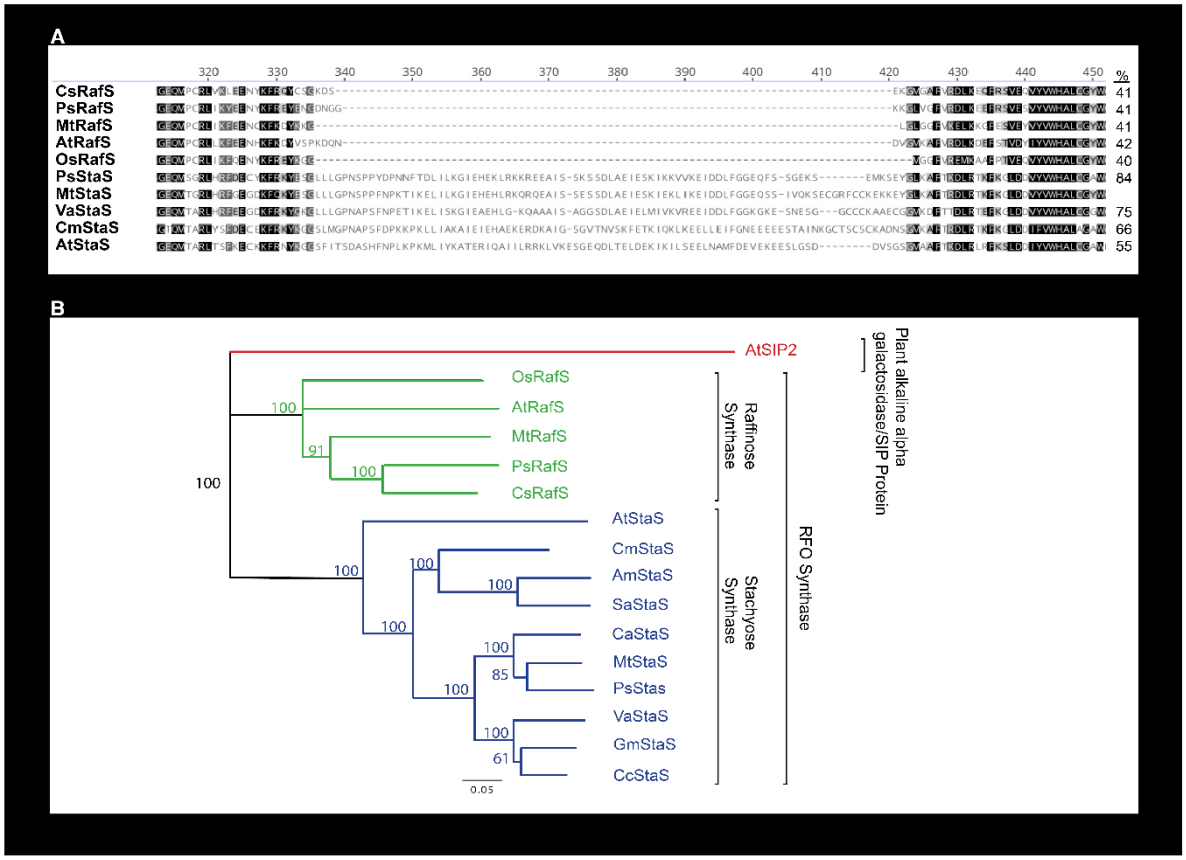


**Figure 1. Sequence and phylogeny of a putative StaS from *Medicago truncatula*** (A) Amino acid alignment section of MtStaS against functionally identified raffinose synthases from *Cucumis sativus* (CsRafS; ABD72603.1), *Pisum sativum* (PsRafS; CAD20127), *M. truncatula* (MtRafS; KEH31804 *putatively annotated), *Arabidopsis thaliana* (AtRafS; BAB11595), *Oryza sativa* (OsRafS; XP_015621501) and stachyose synthases from *P. sativum* (PsStaS; CAC38094), *M. truncatula* (MtStaS; XP_013450269), *Vigna angularis* (VaStaS; CAB64363), *Cucumis melo* (CmStaS; XP_008451468) and *A. thaliana* (AtStaS; NP_192106) demonstrating a conserved 80 amino acid long sequence shared by StaSs but, not present in RafSs. Identical amino acids are highlighted in black and similar ones in grey. Identities (%) are presented against the predicted amino acid sequence of MtStaS. Alignments were performed and aligned using ClustalW (Thompson *et al.* 1994) in the Geneious software package using the default parameters. (B) Evolutionary relationships of StaS and RafS enzymes rooted against the known alkaline α-Galactosidase AtSIP2 (*A. thaliana*; NP_191311; Peters *et al.* 2010). The phylogenetic tree was constructed using the Neighbor-Joining method (Saitou and Nei 1987). Additional StaSs were included: AmStaS (*Alonsoa meridionalis*; CAD31704), SaStaS (*Stachys affinis*; CAC86963), CaStaS (*Cicer arietinum*; XP_004494437), GmStaS (*Glycine max*; NP_001341802) and CcStaS (*Cajunus cajan*; XP_020234864). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test ($1 \times 10^6$ replicates) are shown next to the branches (Felsenstein 1985). Branches corresponding to partitions reproduced in <50% bootstrap replicates were collapsed. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the p-distance method (Nei and Kumar 2000) and are in the units of the number of amino acid differences per site. Overall, the analysis involved 16 protein sequences. Evolutionary analyses were conducted in MEGA X (Kumar *et al.* 2018).

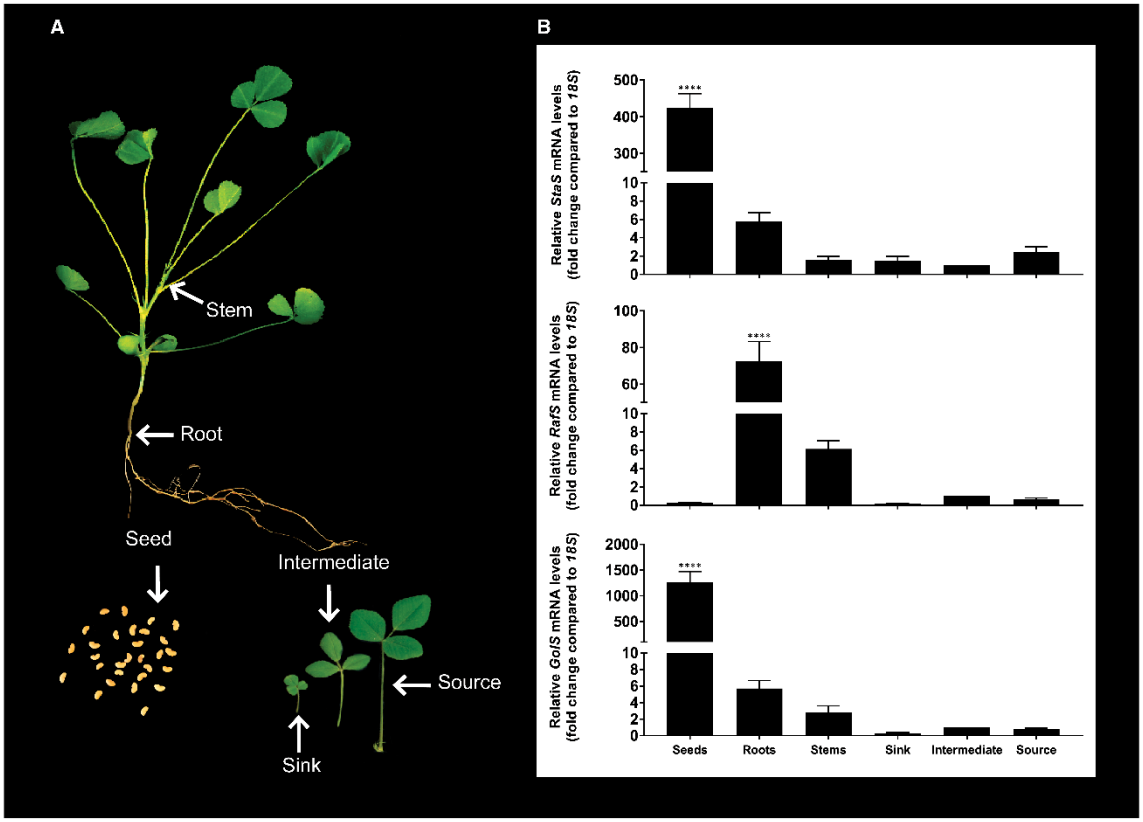


**Figure 2. Transcript levels of *GolS*, *RafS* and *StaS* in different *Medicago* tissue** (A) Arrowheads indicate the tissues harvested to investigate transcript abundance. Harvesting was done at 2 h after dawn. (B) Transcription profiles of *MtGolS*, *MtRafS* and *MtStaS* in various tissue types as shown in A. The threshold cycle number ($\Delta C_T$) was used to calculate relative fold change with the $\Delta\Delta C_T$ method, using Intermediate as the calibrator sample (Livak and Schmittgen 2001). All qPCR experimentation was conducted in compliance with the "Minimum Information for Publication of Quantitative Real-Time PCR Experiments" (MIQE, Bustin *et al.* 2009). Data were normalized to *18S* mRNA and relative mRNA levels are represented graphically as fold change compared to calibrator sample. A value of 1.0 represents no expression//transcript deficiency. Data represents mean ± SEM; n = 3 (each analyzed in triplicate); ****$p < 0.0001$.

## 2.4 DISCUSSION

*Medicago truncatula* is a species that is native to the Mediterranean region, characterised by a climate that yields hot, dry summers, and wet winters (Perry 1997; Kottek *et al.* 2006). RFOs play an integral role in numerous physiological processes such as abiotic stress tolerance, translocation of photoassimilates and developing seed desiccation tolerance (Madore *et al.* 1988; Joersbo *et al.* 1999; Taji *et al.* 2002; Downie *et al.* 2003; Zuther *et al.* 2004; Nishizawa *et al.* 2008). RFOs are also implicated as signalling molecules upon wounding and pathogenetic attacks (Stevenson *et al.* 2000; Couée *et al.* 2006; Xue *et al.* 2007; Kim *et al.* 2008). However, a lesser amount of information is available on the *GolS*, *RafS* and *StaS* gene families in *M. truncatula* despite RFO related synthases (*GolS*, *RafS* and *StaS*) being well characterised and studied in an assortment of plants (Zhu *et al.* 1998; Peterbauer and Richter 2001; Taji *et al.* 2002; Nishizawa *et al.* 2008; Peters and Keller 2009; Zhuo *et al.* 2013; Egert *et al.* 2013; Gangl *et al.* 2015; Li *et al.* 2017).

Herein, a *RafS* and *StaS* were genome-wide identified from *M. truncatula*, which were classified into two distinct clades, respectively, according to their phylogenetic relationship (Figure 1B). A classification brought about based upon the existence of an insertion characteristic of *StaS*s, clearly distinguishing RafSs from StaSs and is in agreement with previous studies of StaS family from *Pisum sativum, Cucumis melo*, *Alonsoa meridionalis*, *Vigna angularis* and *A. thaliana* (Holthaus and Schmitz 1991; Peterbauer *et al.* 1999a; Pluskota *et al.* 2015; Gangl *et al.* 2015, respectively). This conserved 80 amino acid long sequence, shared in StaS but not present in RafS amino acid sequences, appears to have become a reliable diagnostic for StaS identification (Peterbauer *et al.* 2002; Gangl *et al.* 2015; Li *et al.* 2017; Figure 1B). This conserved motif, the α-amylase catalytic (AC) domain, is a characteristic indication of an enzyme belonging to the α-amylase family that forms part of the family of glycosyl hydrolases. Enzymes belonging to this family catalyse the hydrolysis of α-1,4- and α-1,6-glucosidic linkages. Sequence comparison shows that the α-amylase catalytic (AC) domain of all aligned StaSs were missing in all RafSs (Figure 1A).

The MtStaS amino acid sequence also shares high identity (42%, 41%, 41%, 40% identity, Figure 1A) with AtRafS, CsRafS, PsRafS and OsRafS, respectively –

biochemically characterised enzymes responsible for the biosynthesis of Raf (Peterbauer *et al.* 2002b; Li *et al.* 2007; Sui *et al.* 2012; Gangl *et al.* 2015). Utilising Gol as a galactosyl donor, RafS and StaS are both able to transfer Gol to acceptors differing in a single galactosyl unit. In spite of sharing various biochemical and molecular similarities, StaSs are completely inactive on sucrose (Hoch *et al.* 1999; Peterbauer *et al.* 1999) as a substrate and in stark contrast, RafSs are inactive on raffinose substrates (Lehle *et al.* 1970).

A distinct subclade was formed by AtStaS (Figure 1B) – the only known bifunctional StaS with RafS capacity (Gangl *et al.* 2015). The putative MtStaS formed a subclade with PsStaS – a known bifunctional StaS possessing Gol-independent activity in synthesising Verbascose (utilising Raf and Sta) as well as Gol-dependent ability in synthesising Stachyose (utilising Raf and Gol). Species belonging to these subclades could be indicative of two classes of StaS - mono- and bifunctional. Interestingly, PsStaS also displayed RFO hydrolase activity, but was unable to initiate RFO biosynthesis using Suc and Gol as substrates (Peterbauer *et al.* 2002b). From this, it can be hypothesised that MtStaS may present similar biochemical characteristics and *in vitro* enzymatic assays can ascertain whether MtStaS exhibits any RFO hydrolase- or Gol-independent RFO synthase activities.

We conducted comprehensive qPCR analyses on the tissue-specific expression profiles of *MtGolS, MtRafS* and *MtStaS.* We found that all the candidate genes - *Gol, Raf* and *Sta* synthases appear to exhibit high tissue-specific expression patterns. For example, *MtStaS* and *MtGolS* exhibits greater mRNA transcript levels in seeds than in roots and leaves (Figure 2B). Whereas *MtRafS* displays significantly greater transcript levels in roots than in seeds. Stachyose is also found to be the principal RFO in most legumes and plant species due to its high content in seeds (Sosulski *et al.* 1982; Quemener and Brillouet 1983; Andersen *et al.* 2005; Martínez-Villaluenga *et al.* 2008; Martín-Cabrejas *et al.* 2008; Huynh *et al.* 2008; Dilis and Trichopoulou 2009; Wang *et al.* 2010). The total soluble sugars in dry mature *M. truncatula* seeds consists of over 90% stachyose – accounting for 12% of the total dry weight of a single seed (Rosnoblet *et al.* 2007). It is therefore not surprising that detected *StaS* transcript levels in seeds are significantly higher than in any of the other organs.

Interestingly, incredibly low levels of *RafS* transcript was detected in seeds, despite reports indicating substantial Raf content present in *M. truncatula* seeds (Vandecasteele *et al.* 2011). It can be suggested that *MtStaS* may possess bifunctional capabilities in synthesising Raf - explaining the absence of *RafS*-specific transcripts in spite of reports on known Raf content in the seeds. *In sillico* identification within the *M. truncatula* genome, however, revealed seven distinct *RafSs* (Vandecasteele *et al.* 2011). These *RafS* isoforms are hypothetically responsible for the Raf accumulation in seeds in *M. truncatula*. The *RafS* identified in this chapter may also exclusively be expressed in root tissue, only.

Further studies should seek to test for *MtStaS* bifunctionality in synthesising Raf *in vitro* and additional experiments should include the other six (*in sillico* identified) in transcript analysis studies across the different organs to determine the *RafS* responsible for Raf accumulation in mature seeds if biochemical characterisation proves *MtStaS* to not possess RafS bifunctional capabilities in biosynthesising RFOs.

# Chapter III: Functional identification of *MtStaS* as a *bona fide* stachyose synthase

## 3.1 INTRODUCTION

Gene knockout strategies (loss-of-function) are regarded to be a key element of the functional genomics toolbox and serves as an essential component in revealing and characterising gene functions following large scale genome sequencing initiatives (*Arabidopsis* Genome Initiative 2000; Tang *et al.* 2014). While such reverse genetic strategies are commonplace in the *Arabidopsis* model with a number of T-DNA insertion mutants available for almost the entire set of coding sequences in the genome, very few have focused on genes involved in RFO metabolism. T-DNA insertion mutants (*atsip2*) were able to dispute the putative function of AtRS2 (subsequently renamed to AtSIP2; At3g57520) as a genuine alkaline α-galactosidase with a distinctive substrate specificity for Raf, and not a RafS. The *atsip2* knockouts contained more Raf 24 h after water deficit relief when compared to wild-type plants. Interestingly, no differences were observed in leaf Raf content between mutant and wild-type plants after de-acclimation from 4°C. (Peters *et al.* 2010). Implicating AtSIP2 in the hydrolysis of water deficit-induced Raf accumulation subsequent to stress relief.

Gene knockout strategies, using dedicated RFO biosynthetic enzymes RafS and StaS, have been employed to study RFO accumulation *in vivo* (Zuther *et al.* 2004; Egert *et al.* 2013; Gangl *et al.* 2015; Gangl and Tenhaken 2016). The *atrs5* T-DNA insertion mutants lack temperature-induced accumulation of Raf, and it was subsequently concluded that AtRafS (At5g40390) was responsible for the induced accumulation (Zuther *et al.* 2004). AtRafS was thereafter, further characterised to contribute to Raf accumulation in *Arabidopsis* seeds, however unexpectedly, *atrs5* mutants did not display complete ablation of Raf in seeds, postulating the involvement of at least another RafS (Egert *et al.* 2013). It was only when a double knock-out mutant was created for *atrs4.atrs5* when total ablation of Raf content in *Arabidopsis* seeds were observed (Gangl *et al.* 2015). This led to the *in vitro* biochemical characterisation of AtStaS as a bifunctional StaS with RafS capacity.

We undertook to functionally express and characterise the *MtStaS* gene (*Medtr7g106910.1*) which its deduced protein sequence exhibits 84% amino acid identity to the *Pisum sativum* StaS (PsStaS; genbank acc: CAD20127; Peterbauer *et*

*al.* 2002b). Furthermore, if MtStaS proves to be a *bona fide* StaS, it would be heterologously expressed in the dimorphic fungus, *Yarowia lipolytica*, in order to biochemically characterise the recombinant protein, confirm *in planta* results, and particularly to ascertain if it possesses any bifunctionality in synthesising RFOs. In keeping with the theme of this thesis, functionally identifying MtStaS, we were interested in an experimental methodology that would allow *Arabidopsis* single and double mutant plants (*RafS* and *StaS* insertional knock-out) to have the missing RFO pathway complemented by MtStaS and to study its *in vivo* function as a putative RFO synthase.

To this end, we envisaged a strategy that would include the constitutive expression of the putative MtStaS in an *Arabidopsis* stachyose deficient single mutant (*atrs4*), in addition to a stachyose *and* raffinose deficient double mutant (*atrs4.atrs5*). We set out to (i) keep the RFO biosynthetic pathway intact in controls (Wild-type, Col-0), (ii) create a binary vector construct where *MtStaS* expression is driven by a constitutive promotor (dual 35S CamV promotor), (iii) transform the *atrs4* and *atrs4.atrs5* backgrounds with this construct, (iv) analyse transgene expression levels, and (v) analyse the leaves and seeds of the transgenic lines for RFOs (Raf and Sta) using an LC-MS/MS based methodology. We hypothesised that expressing *MtStaS* in single (*atrs4)* and double (*atrs4.atrs5)* mutant lines would accumulate Sta to varying degrees depending on tissue-specificity and bifunctional enzymatic capabilities.

This chapter describes the functional identification of a putative stachyose synthase from *M. truncatula* (MtStaS; Medtr7g106910.1) by constitutive expression in an *Arabidopsis* single and double mutant (*atrs4* and *atrs4.atrs5*). This mutant has been generated and represents a *RafS* and *StaS* insertional knock-out. As such, *atrs4* is deficient of Sta accumulation in the seeds and leaves, presenting no StaS activities in these, whereas *atrs4.atrs5* is deficient of Raf and Sta accumulation in the leaves and seeds, presenting neither RafS nor StaS activities in these. This is the first time, to the best of our knowledge, where mutant *Arabidopsis* plants were used as a novel heterologous platform to study RFO biosynthesis, effectively, serving as an excellent functional screening system to identify RFO genes in their ability to compliment missing RFO metabolism.

## 3.2 MATERIALS AND METHODS

Unless specified otherwise, chemicals used throughout this study were obtained from Sigma-Aldrich® (www.sigmaaldrich.com/south-africa.html) or MERCK® (Modderfontein, South Africa). The Oligo explorer® software (V1.4 BETA) was used to design the primers which were subsequently synthesised by Inqaba Biotech®. All enzymes used in this study were obtained from New England Biolabs® (NEB, Inqaba Biotechnical Industries (Pty) Ltd, South Africa), unless stated otherwise. Vectors and primers used in this study are summarised in Table 1.

### 3.2.1 cDNA isolation of *M. truncatula* stachyose synthase (*MtStaS*)

The nucleotide sequence of the putative stachyose synthase (*MtStaS*; *Medtr7g106910.1*) was obtained in chapter 2 (2.2.1). Total RNA was extracted from 80 mg of seed material from the experimental *M. truncatula* (A17) line using the Maxwell® 16 LEV simplyRNA Purification Kit in the Maxwell® 16 Instrument (AS2000; Promega, Anatech, South Africa), according to the manufacturer's instruction. Complementary DNA (cDNA) was synthesised as outlined in chapter 2 (2.2.3). The coding sequence (CDS) was amplified using Q5® High-Fidelity DNA Polymerase (New England Biolabs®) *via* PCR according to the manufacturer's instructions, using the *MtStaS_CDS_F* forward and *MtStaS_CDS_R* reverse primers (Table 1). An amplicon (~2.5 kb) was identified and subsequently purified using the Wizard® SV Gel and PCR Clean-up System (Promega) in compliance with the manufacturer's protocol by means of gel electrophoresis (0.8%; w/v; 60 V).

### 3.2.2 Gateway® cloning strategy

#### 3.2.2.1 Generation of entry vector – *pCR™8::MtStaS*

The *MtStaS* CDS was then cloned into the pCR™8 cloning vector using the pCR™8 /GW/TOPO®TA Cloning kit (Invitrogen, Life technologies, South Africa) following the manufacturer's protocol. Entry clones were transformed using a conventional heat shock method into OneShot® Competent *Escherichia coli* cells (Invitrogen; Table 1). Antibiotic resistant colonies ($^{100}$Spec.) were subjected to a colony PCR using the

*MtStaS_CDS_F* forward and *T7_R* (pCR™8 specific; Table 1) reverse primers to identify colonies where *MtStaS* had inserted into pCR™8 in 5'-3' orientation. Plasmid minipreparations were obtained using the Wizard® *Plus* SV Minipreps DNA Purification System (Promega), following the manufacturer's protocol. Subsequently, inserts were sequenced (Central Analytical Facility, Stellenbosch University, South Africa), using the *M13_F* forward and *M13_R* reverse primers (Table 1). Thereafter, a single clone was selected for the generation of a Gateway® expression vector.

#### 3.2.2.2 Generation of plant expression vector – *pMDC32::MtStaS*

The pMDC32 binary vector (Curtis and Grossniklaus 2003) was used to create a construct *via* a conventional LR clonase™ reaction (Invitrogen) to obtain *pMDC32::MtStaS* (constitutive *MtStaS* expression driven by dual CamV 35S promoter). Briefly a Gateway® recombination cloning strategy was employed to transfer *MtStaS* from the pCR™8 entry vector into *pMDC32*, following the manufacturer's protocol (Invitrogen). Clonase™ reactions were transformed in One Shot® OmniMAX™ 2 T1 PhageResistant Cells (Invitrogen; Table 1) using a conventional heat shock method. Antibiotic resistant colonies ($^{50}$Kan.) were subjected to a final colony PCR using the *MtStaS* primer (*MtStaS_CDS_F*; Table 1) and the pMDC32 specific primer (*pMDC32_R;* Table 1) to confirm that *MtStaS* had transposed into the vector. Plasmid minipreparations were obtained from positive colonies, and used in *Agrobacterium tumefaciens* transformations.

### 3.2.3 Plant transformation

#### 3.2.3.1 *Agrobacterium tumefaciens* transformation

The plant expression vector harbouring the *pMDC32::MtStaS* construct was introduced into electro-competent *Agrobacterium tumefaciens* (strain GV3101; Koncz and Schell 1986; Table 1) cells by means of electroporation. Plasmid DNA (500 ng) was added to 100 µl of *A. tumefaciens* cells and electroporated using a Genepulser® system (Bio-Rad, Bio Rad Laboratories, South Africa) set to 1.8 kV, 100 Ω and 25 µFD in a 2 mm cuvette. Transformants were selected on LB plates supplemented with the appropriate antibiotics ($^{50}$Rif., $^{25}$Gent., $^{50}$Kan.) after incubation at 28°C for 48 h. Positive clones were confirmed by means of a colony PCR using the *MtStaS_CDS_F* forward and *pMDC32_R* reverse primers (Table 1).

### 3.2.3.2 Plant Material and growth conditions

All *Arabidopsis* T-DNA insertion mutants used in this study were originally obtained from the Salk Institute's T-DNA insertion mutant collection in the Col-0 background (Alonso *et al.* 2003). A single insertion mutant for At4g01970 (Salk_088817; Sta deficient), designated *atrs4* carrying a T-DNA insertion in the 5' UTR region was used in this study along with a double insertion mutant designated *atrs4.atrs5* (Sta and Raf deficient) carrying a T-DNA insertion in the third exon that was created using At4g01970 (Salk_026853; Sta deficient) and an insertion mutant in the fourth exon for At5g40390 (Salk_049583; Raf deficient), respectively, as part of a previous research project (Table 2; Loedolff *et al.* 2015).

*Arabidopsis thaliana* wild-type and mutant (*atrs4*, *atrs4.atrs5*) seeds were surface sterilised and subsequent to seed stratification (24 h, 4°C), were sown onto peat disks (Jiffy™ no.7, South Africa) and propagated under controlled environmental conditions in short-day (SD) growth chambers (Snijders Labs, Economic deluxe; 8 h light, 120 µmol photons $m^{-2}$ $s^{-1}$, 22°C, 16 h dark, 18°C, 60% relative humidity) for approximately 60 days, until the plants developed multiple inflorescences. Mutant plants (*atrs4*, *atrs4.atrs5*) were then selected for floral transformations.

### 3.2.3.3 Plant transformation and selection

*Arabidopsis thaliana* (*atrs4 and atrs4.atrs5*) mutant plants were transformed using a modified floral inoculation protocol (Narusaka *et al.* 2010). Single *A. tumefaciens* colonies harbouring the *pMDC32::MtStaS* construct were selected and inoculated into 5 ml LB cultures supplemented with the appropriate antibiotics ($^{50}$Rif., $^{25}$Gent., $^{50}$Kan.) and incubated (28°C) with shaking (200 rpm) until mid-log phase ($OD_{600}$ = ~1.2) was reached. An aliquot of the culture (1.5 ml) was centrifuged (7000 *g*, 10 min), the supernatant was removed, and the pellet resuspended in 1 ml, 5% (w/v) sucrose. Silwet L-77 was supplemented to a concentration of 0.02% (v/v) and vortexed preceding floral inoculation. Closed flower buds were inoculated with 5 µl of *A. tumefaciens* inoculum. Inoculated plants were incubated in the dark (16 h, 90% relative humidity) prior to growth under controlled environmental conditions.

Seeds (T1) from the transformed plants (T0) were collected, sterilised and plated onto half-strength MS (Duchefa, Labretoria, South Africa) media containing 5% (*w/v*)

sucrose and $^{17.5}$Hyg. for selection. Plates were stratified (4°C, 24 h) and then maintained in the controlled environment chamber described above for two weeks. Positive transformants (seedlings which resisted hygromycin had green, open, expanded cotyledons with long hypocotyls) were selected and transferred to Jiffy peat disks (Jiffy™ nr. 7, South Africa) and maintained in the same chamber described as above. Seeds (T2) were collected and plants representing the T2 generation were used for further characterisation.

### 3.2.4 Characterisation of transgenic lines

Plants representing the T2 generation after transformation were first confirmed homozygous for T-DNA knockout mutant alleles, genotyped for *pMDC32::MtStaS* integration into the genome and expression of *MtStaS* using qPCR.

#### 3.2.4.1 Genotyping *atrs4* and *atrs4.atrs5* lines

Genetic identity was determined *via* PCR from genomic DNA (gDNA) extracted from young *A. thaliana* source leaves (Edwards *et al.* 1991). Wild-type (Col-0), homozygous *atrs4* and *atrs4.atrs5* plants were confirmed using the primers listed in Table 1. PCR cycles were subjected to the following conditions: initial denaturation (95°C) for 3 min, proceeded by repeated denaturation (95°C) for 30 s, annealing (60°C) for 60 - 120 s, and extention (72°C) for 60 s for 25 cycles. The final elongation step was performed at 72°C for 10 min in a thermal cycler (T100™ Thermal Cycler; Bio-Rad).

#### 3.2.4.2 RNA extraction and cDNA synthesis

Total RNA was extracted from 80 mg of leaf material using the Maxwell® 16 LEV simplyRNA Purification Kit in the Maxwell® 16 Instrument (AS2000; Promega, Anatech, South Africa), following the manufacturer's instruction. Source leaves were harvested 32 days after planting. Leaf material did not include petioles. Harvesting was done 2 h after dawn. Briefly, RNA was harvested in parallel from three biological replicates. Each biological replicate consisted out of a pool of tissues that were harvested from three independent plants. The quantity and purity of RNA was measured using a NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific, Inqaba biotech, South Africa). RNA quality was evaluated on denaturing electrophoretic gels stained with ethidium bromide (1.0%; w/v; 60 V). Purified RNA

was subsequently stored at -80°C until further use. Complementary DNA (cDNA) was synthesised as outlined in chapter 2 (2.2.3).

### 3.2.4.3 Transcript analysis

Plants representing the T2 generation after transformation were tested for expression of *MtStaS* using qPCR as outlined in Chapter 2 (2.2.4). Three reference genes (Table 1; Czechowski *et al.* 2005) *ACT2, UBC21* and *CBP20* were used in the analyses. The threshold cycle number ($\Delta C_T$) was used to calculate relative fold change with the $\Delta\Delta C_T$ method, using the wild-type (Col-0) as the calibrator sample (Livak and Schmittgen 2001). The mean $C_T$ value of three technical replicates were analysed for every biological replicate. *ACT2* served as the reference gene in all analyses, and changes in mRNA levels relative to *ACT2* were confirmed using *UBC21* and *CBP20* as alternate reference genes in independent experiments. All qPCR experiments were conducted in accordance with the "Minimum Information for Publication of Quantitative Realtime PCR Experiments" (MIQE, Bustin *et al.* 2009).

### 3.2.5 Water soluble carbohydrate (WSC) extractions

WSCs extractions from freeze-dried seeds (50 mg) and macerated leaf (50 mg) material were conducted as previously described (Peters *et al.* 2007; Peters and Keller 2009; Egert *et al.* 2013), with minor modifications. Extractions were conducted in a three-step sequential series (1 ml 80% (v/v) EtOH, 1 ml 50% EtOH (v/v) and 1 ml de-ionised $H_2O$ ($dH_2O$). Every consecutive extraction was conducted twice at 80°C for 10 min, centrifuged (13 000 *g*, 10 min, RT) and supernatants transferred to a new Eppendorf tube prior to the next step in the series. The supernatants for each individual extraction were pooled, concentrated in a vacuum centrifuged and resuspended in $dH_2O$ to a final volume of 200 µl. Samples were then de-ionised and de-phenolised as previously described (Peters *et al.* 2007; Peters and Keller 2009; Egert *et al.* 2013), before LC-MS/MS analysis.

### 3.2.6 LC-MS/MS analysis

LC-MS/MS analysis was performed at the Central Analytical Facility, Stellenbosch University, South Africa, using a Waters Synapt G2 quadrupole time-of-flight mass spectrometer (Waters Corporation, Milford, MA, USA) equipped with a Waters Acquity UPLC. Samples were separated on a Waters UPLC BEH Amide column (2.1 x 100 mm; 1.7 µm) at a flow rate of 0.17 ml/min at 35°C. Solvent A consisted of acetonitrile/water (30:70) containing 0.1% ammonium hydroxide and solvent B was acetonitrile/water (80:20) containing 0.1% ammonium hydroxide. The mobile phase gradient was from 0% to 60% solvent A over 5 min, maintained for 2 min at 60% solvent A before the column was re-equilibrated to the initial conditions. Electrospray ionization was applied in the negative mode and the scan range was from m/z 150 to 1500. The capillary voltage was set a 2.5 kV, the cone voltage was 15 V, the source temperature 120°C and the desolvation temperature was 275°C. The desolvation gas and cone gas flows were 650 L/h and 50 L/h, respectively. All WSCs were monitored using their deprotonated quasi-molecular ions and quantified with the TargetLynx application manager (Waters MassLynx V4.1V software).

### 3.2.7 Statistical analysis

Statistical analyses were performed as outlined in chapter 2 (2.2.5).

**Table 1. Strains, vectors, and primers used in this study**

| Name | Characteristics | Use | Supplier/Reference |
|---|---|---|---|
| **Strains** | | | |
| *Escherichia coli* One Shot® TOP10 | Genotype: *F- mcrA Δ( mrr-hsdRMS-mcrBC) Φ80lacZΔM15 Δ lacX74 recA1 araD139 Δ( ara-leu)7697 galU galK rpsL (StrR) endA1 nupG* | Host for cloning, vector propagation, assembly of complete vector | Invitrogen |
| *Escherichia coli* One Shot® Omni-MAX™ 2 T1$^{R}$ | Genotype: *F´ {proAB lacI$^{q}$ lacZΔM15 Tn10(Tet$^{R}$ ) Δ(ccdAB)} mcrA Δ(mrr hsdRMS-mcrBC) Φ 80(lacZ)ΔM15 Δ(lacZYA-argF)U169 endA1 recA1 supE44 thi-1 gyrA96 relA1 tonA panD* | Habouring destination vector in Gateway® reaction | Invitrogen |
| *Agrobacterium tumefaciens* GV3101 | Rifampicin$^{r}$, Gentamycin$^{r}$ | *Agrobacterium tumefaciens* mediated genetic transformation harbouring plant expression vector | Koncz and Schell 1986 |
| **Vectors** | | | |
| pCR™8/GW/ TOPO® | https://assets.thermofisher.com/TFS-Assets/ LSG/manuals/pcr8gwtopo_man.pdf | Entry vector in Gateway® reaction | Invitrogen |
| pMDC32 | Hygromycin$^{r}$, Kanamycin$^{r}$, dual 35S CaMV promoter | Destination vector in Gateway® reaction | Curtis and Grossniklaus 2003 |
| **Primers (5' - 3')** | | | |
| MtStaS_CDS_F MtStaS_CDS_R | ATGGCTCCACCGAATTCCACAAACCT GGTGTTTCTGATTTGGCAATTTTCTTCTAG | *MtStaS* gene amplification and orientation screening | This study |
| M13_F M13_R | GTAAAACGACGGCCAG CAGGAAACAGCTATGAC | Sequencing | This study |
| T7_R | CCCTATAGTGAGTCGTATTA | Orientation screening | This study |
| pMDC32_R | TAGAGGATCCCCGGGTACC | Orientation screening | This study |
| atrs4_LP atrs4_RP | TCGAATACGCCATGAATCTTC CAGAAGAACATGGAGGACGAG | Zygosity determination of T-DNA insertion lines | SALK institute |
| atrs5_LP atrs5_RP | CTCTTCTTGAAGGCTCCTTCC ATGACATCAACTTTAACGCCG | Zygosity determination of T-DNA insertion lines | SALK institute |
| atrs4-1_LP atrs4-1_RP | GAGCCACTCTCTGCACAAATC GCATCATAGTTTGCCAAGTAGC | Zygosity determination of T-DNA insertion lines | SALK institute |
| LBb1.3 | ATTTTGCCGATTTCGGAAC | Zygosity determination of T-DNA insertion lines | SALK institute |
| ACT2_Q_F ACT2_Q_R | CTTGCACCAAGCAGCATGAA CCGATCCAGACACTGTACTTCCTT | qPCR transcript analysis | Czechowski *et al.* 2005 |
| CBP20_Q_F CBP20_Q_R | CCTTGTGGCTTTTGTTTCGTC ACACGAATAGGCCGGTCATC | qPCR transcript analysis | Czechowski *et al.* 2005 |
| UBC21_Q_F UBC21_Q_R | CTGCGACTCAGGGAATCTTCTAA TTGTGCCATTGAATTGAACCC | qPCR transcript analysis | Czechowski *et al.* 2005 |
| MtStaS_Q_F MtStaS_Q_R | AGGTGGTGGGAATTTCCTTG TTTCCATCACCTAGCCACTC | qPCR transcript analysis | This study |

## 3.3 RESULTS

### 3.3.1 Transgene integration and zygosity determination of T-DNA insertion lines

To better understand the functional role of *MtStaS*, we set out to utilise *Arabidopsis* mutants that lacked functional *RafS* and *StaS* genes. Homozygous T-DNA insertion lines were obtained from a previous research project where single stachyose deficient (*atrs4*) and double stachyose and raffinose deficient (*atrs4.atrs5*) mutants were created (Figure 1A; Figure 2A; Loedolff *et al.* 2015). These insertion lines were in the Col-0 background (Table 2). Homozygous plants were confirmed *via* genomic DNA PCRs to discriminate between wild-type and mutant alleles (Figure 1B; Figure 2B).

Following selection of hygromycin resistant T2 transgenic plants, plants were then genotyped for integration of *pMDC32::MtStaS* into the *Arabidopsis* genome *via* genomic DNA PCRs conducted on DNA isolated from the leaves of plants (T2). *MtStaS* amplification was absent in Col–0 and untransformed mutant lines (*atrs4* and *atrs4.atrs5*), but detected in all transformed lines (1-3) for both *atrs4/MtStaS* and *atrs4.atrs5/MtStaS* (Figure 1C; Figure 2C).

**Table 2. *Arabidopsis* insertion lines used in this study**

| Locus | Gene | Insertion line | Insertion site | Name of mutant |
|---|---|---|---|---|
| At4g01970<br>At5g40390 | *AtStaS*<br>*AtRafS* | Salk_026853<br>Salk_049583 | Third exon<br>Fourth exon | *atrs4.atrs5* |
| At4g01970 | *AtStaS* | Salk_088817 | 5' UTR | *atrs4* |

### 3.3.2 *MtStaS* is constitutively expressed in transgenic *Arabidopsis* mutant (T2) lines (*atrs4/MtStaS* and *atrs4.atrs5/MtStaS*)

Positive transgenic plants were analysed for expression of *MtStaS via* qPCR. *MtStaS* transcripts were absent in Col-0 and untransformed mutant lines (*atrs4* and *atrs4.atrs5*), but detected in all (1-3) transformed lines for both *atrs4/MtStaS* and *atrs4.atrs5/MtStaS. MtStaS* transcripts occurred in varying abundance in the independent transgenic lines (Figure 1D; Figure 2D).

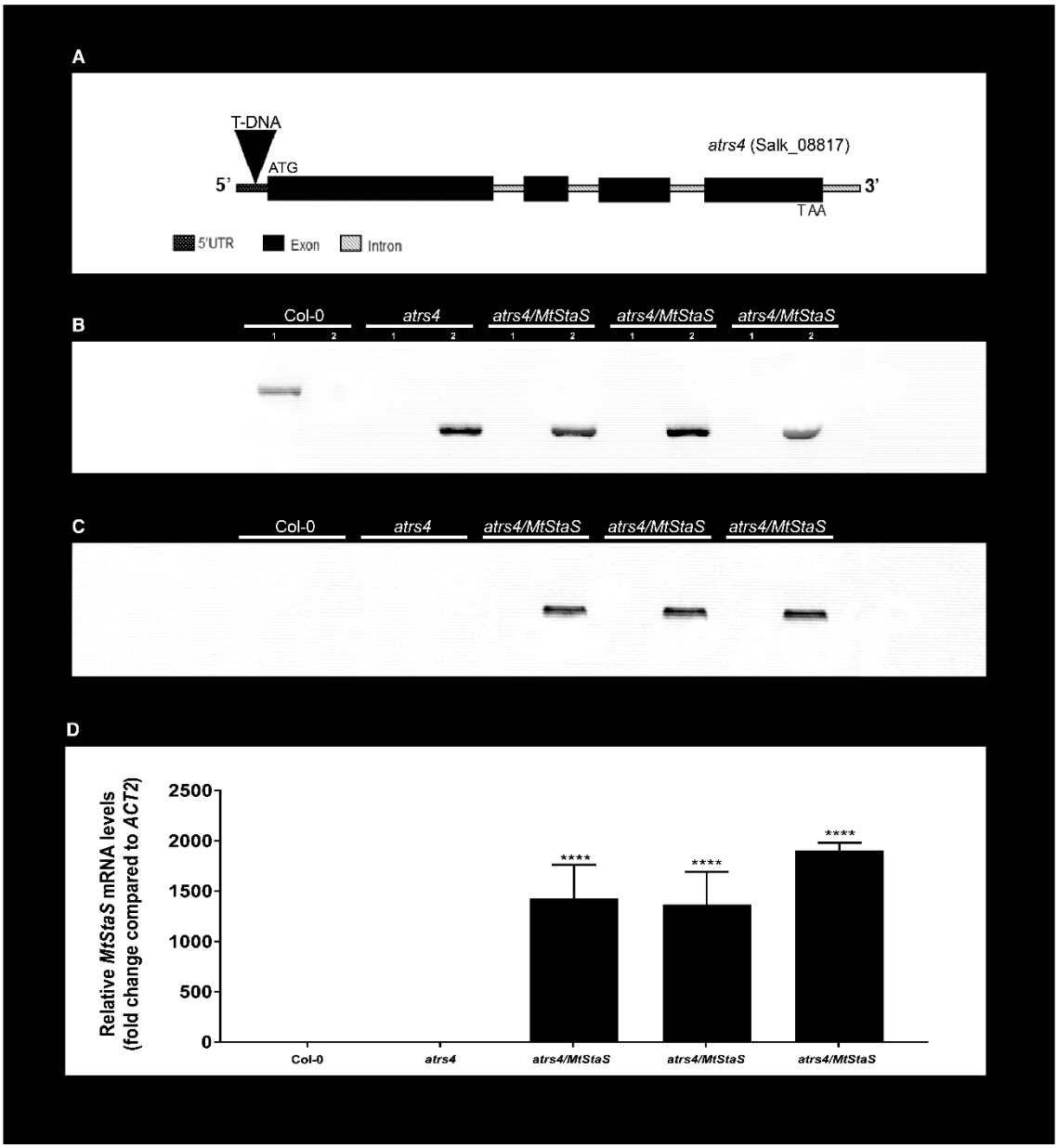


**Figure 1. Analysis of *Arabidopsis atrs4* T-DNA insertion lines** (A) Genetic organisation of *atrs4* illustrating localisation of T-DNA insertion site. (B) Image identifying the homozygous single mutant. Genomic DNA PCRs were conducted on DNA isolated from the leaves of plants (T2). Primer pairs amplified the wild-type (LP+RP) and mutant (RP+LBb1.3) alleles. The following primer pairs were used: 1 – *atrs4-1_LP* + *atrs4-1_RP*; 2 – *atrs4-1_RP* + *LBb1.3* (Table 1). (C) Image identifying the successful integration of *pMDC32::MtStaS* into the *Arabidopsis* genome. Genomic DNA PCRs were conducted on DNA isolated from the leaves of plants (T2). Primer pairs amplified *MtStaS*. (D) Expression levels of *MtStaS* were determined by quantitative real-time PCR (qPCR) in Col-0 and *atrs4* insertion and transformed mutants. The threshold cycle number ($\Delta C_T$) was used to calculate relative fold change with the $\Delta\Delta C_T$ method, using Col-0 as the calibrator sample (Livak and Schmittgen 2001). All qPCR experimentation was conducted in compliance with the "Minimum Information for Publication of Quantitative Real-Time PCR Experiments" (MIQE, Bustin *et al.* 2009). Data were normalised to *ACT2* mRNA and relative mRNA levels are represented graphically as fold change compared to calibrator sample. A value of 1.0 represents no expression//transcript deficiency. Data represents mean ± SEM; n = 3 (each analysed in triplicate); ****$p < 0.0001$.

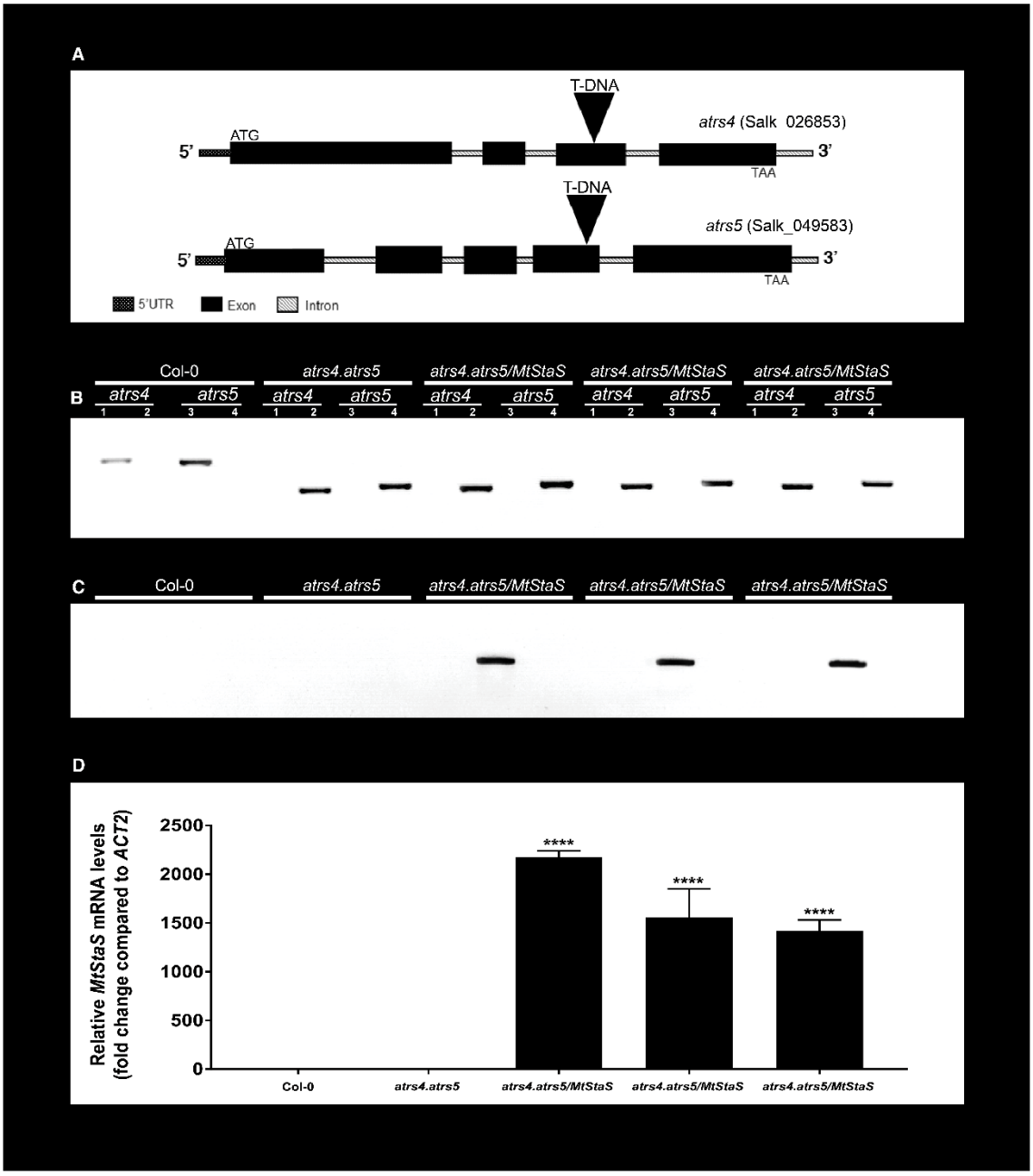


**Figure 2. Analysis of *Arabidopsis atrs4.atrs5* double T-DNA insertion lines** (A) Genetic organisation of *atrs4.atrs5* illustrating localisation of T-DNA insertion site(s). (B) Image identifying the homozygous double mutant. Genomic DNA PCRs were conducted on DNA isolated from the leaves of plants (T2). Primer pairs amplified the wild-type (LP+RP) and mutant (RP+LBb1.3) alleles. The following primer pairs were used: 1 – *atrs4_LP* + *atrs4_RP*; 2 – *atrs4_RP* + *LBb1.3*; 3 – *atrs5_LP* + *atrs5_RP*; 4 – *atrs5_RP* + *LBb1.3* (Table 1). (C) Image identifying the successful integration of *pMD-C32::MtStaS* into the *Arabidopsis* genome. Genomic DNA PCRs were conducted on DNA isolated from the leaves of plants (T2). Primer pairs amplified *MtStaS*. (D) Expression levels of *MtStaS* were determined by quantitative real-time PCR (qPCR) in Col-0 and *atrs4.atrs5* insertion and transformed mutants. The threshold cycle number ($\Delta C_T$) was used to calculate relative fold change with the $\Delta\Delta C_T$ method, using Col-0 as the calibrator sample (Livak and Schmittgen 2001). All qPCR experimentation was conducted in compliance with the "Minimum Information for Publication of Quantitative Real-Time PCR Experiments" (MIQE, Bustin *et al.* 2009). Data were normalised to *ACT2* mRNA and relative mRNA levels are represented graphically as fold change compared to calibrator sample. A value of 1.0 represents no expression//transcript deficiency. Data represents mean ± SEM; n = 3 (each analysed in triplicate); ****$p < 0.0001$.

### 3.3.3 Constitutive expression of *MtStaS* leads to the accumulation of Sta in *atrs4* leaves

Water soluble carbohydrates (WSCs) were extracted from the leaves of Col-0, *atrs4, atrs4/MtStaS* plants, and subsequently analysed by LC-MS/MS. Unsurprisingly, both Gol and Raf were detected in the leaves of Col-0 and *atrs4* plants. Constitutive expression of *MtStaS* in *atrs4,* however, accumulated Gol, Raf and Sta in the leaves (Figure 3).

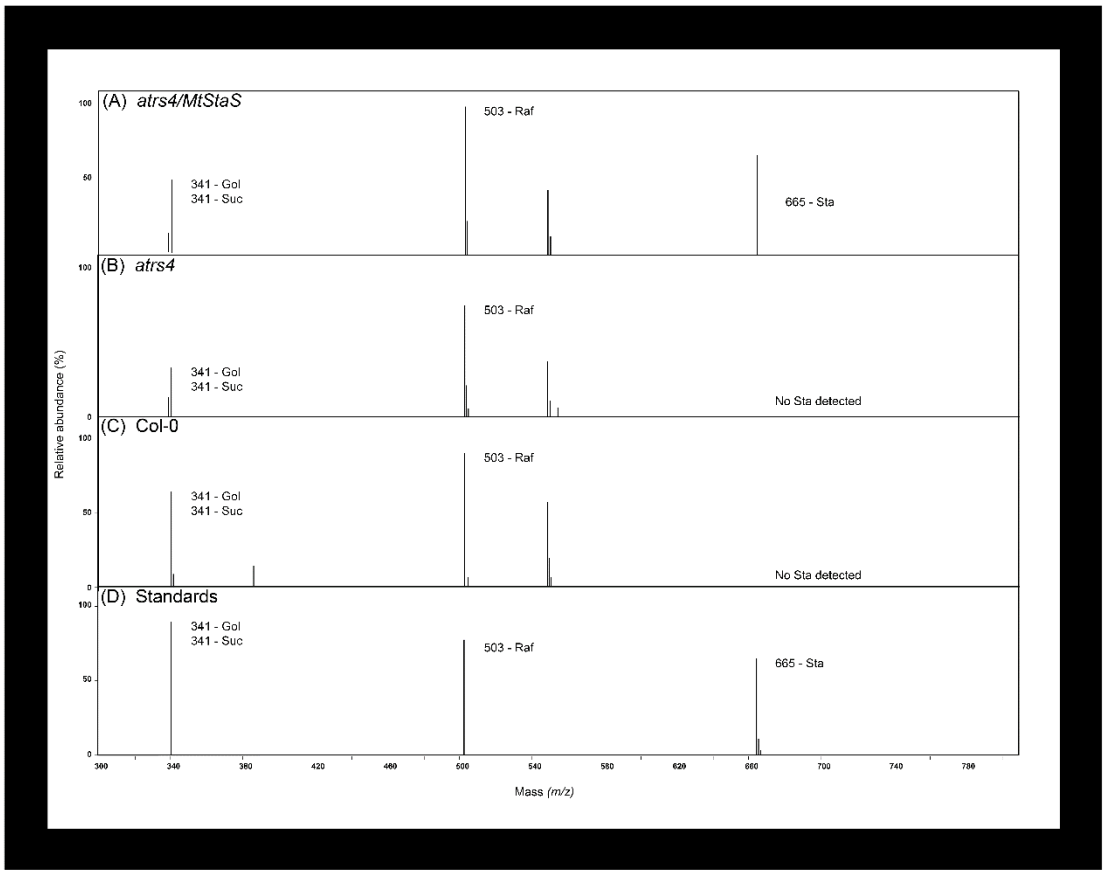


**Figure 3. Water soluble carbohydrate (WSC) profiles in the leaves of Col-0, *atrs4* and *atrs4/MtStaS* lines** Mass spracta representing water soluble carbohydrates (WSCs) extracted from (A) *atrs4/MtStaS* mutant transgenic leaves, (B) *atrs4* mutant leaves, (C) Col-0 leaves and (D) 2 mM galactinol (Gol), raffinose (Raf), and stachyose (Sta) serve as reference compounds. Mass spracta shows a total gain of Sta in *atrs4/MtStaS* mutant leaves.

### 3.3.4 Constitutive expression of *MtStaS* recovers ablated Sta in mature seeds of *atrs4*

Water soluble carbohydrates (WSCs) were extracted from mature seeds of Col-0, *atrs4, atrs4/MtStaS* mutants, and subsequently analysed by LC-MS/MS. RFOs present in Col-0 mature seeds were Gol, Raf and Sta. RFOs present in mature seeds from *atrs4* were only Gol and Raf. Markedly absent, however, was Sta (a total loss of detectable Sta). The *atrs4/MtStaS* accumulated Gol, Raf and Sta in mature seeds – recovering the ablated Sta in the *atrs4* phenotype and resulting in a total gain of detectable Sta (Figure 4).

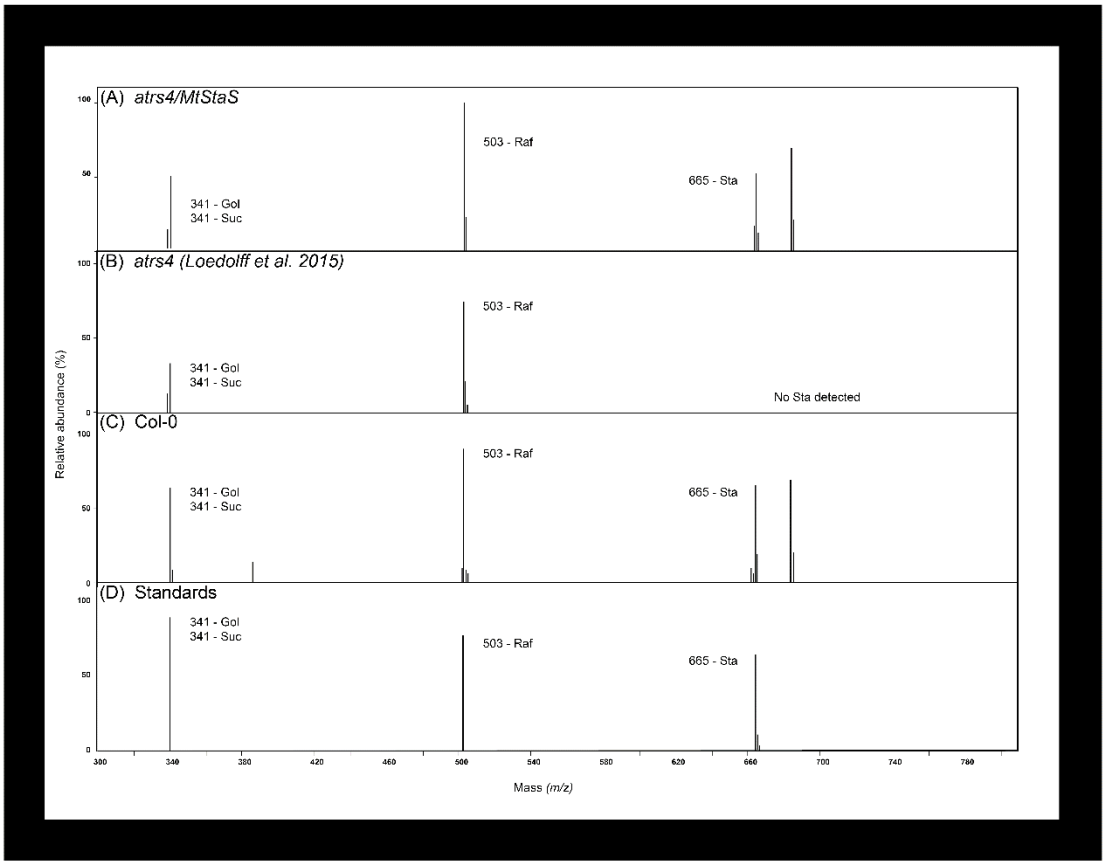


**Figure 4. Water soluble carbohydrate (WSC) profiles in the seeds of Col-0, *atrs4* and *atrs4/MtStaS* lines.** Mass spracta representing water soluble carbohydrates (WSCs) extracted from (A) *atrs4/MtStaS* mutant transgenic seeds, (B) *atrs4 (Loedolff et al. 2015)* mutant seeds, (C) Col-0 seeds and (D) 2 mM galactinol (Gol), raffinose (Raf), and stachyose (Sta) as reference compounds. Mass spracta shows a total loss of detectable Sta in mutant *atrs4* seeds and the phenotype is recovered in *atrs4/MtStaS.*

### 3.3.5 Constitutive expression of *MtStaS* does not lead to the accumulation of Sta in *atrs4.atrs5* leaves

Water soluble carbohydrates (WSCs) were extracted from the leaves of Col-0, *atrs4.atrs5, atrs4.atrs5/MtStaS* double mutants, and analysed by LC-MS/MS. The only RFOs present in Col-0 leaves were Gol and Raf. In contrast, the only RFO present in *atrs4.atrs5* is Gol and a total loss of detectable Raf is observed. Constitutive expression of *MtStaS* is unable to accumulate either Raf or Sta in *atrs4.atrs5/MtStaS* leaves (Figure 5).

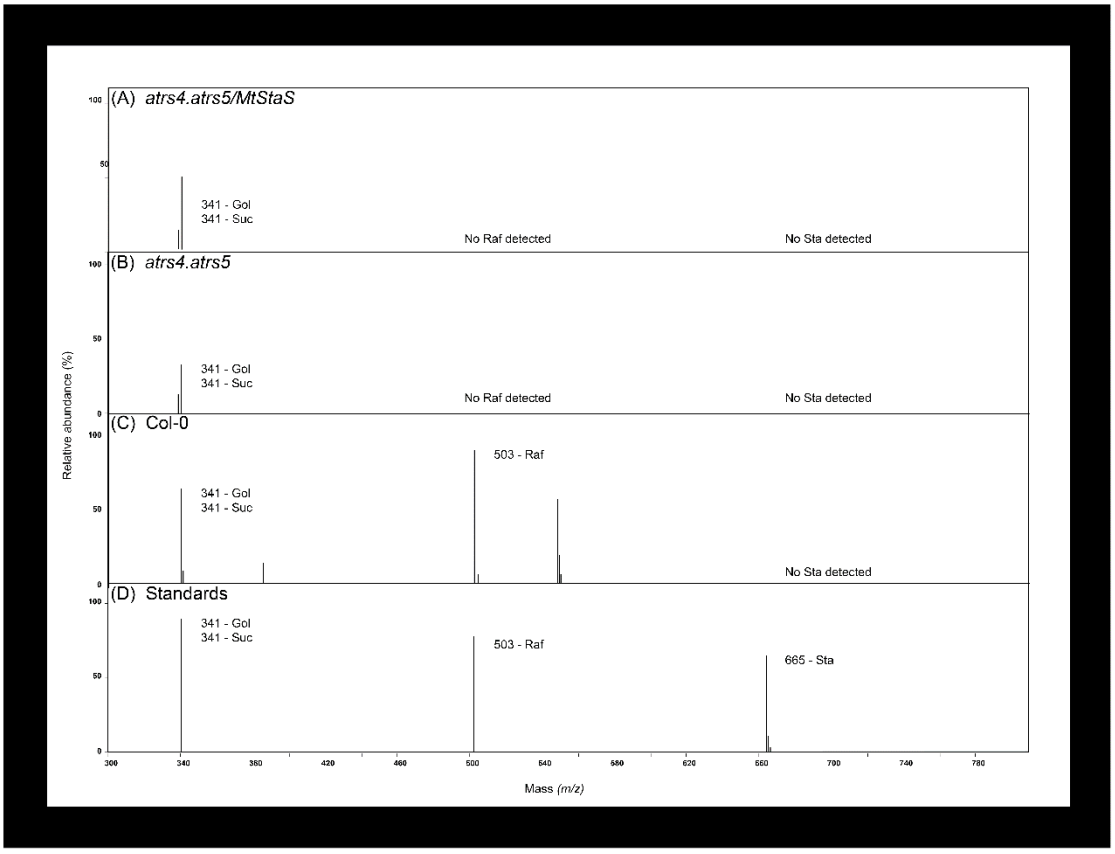


**Figure 5. Water soluble carbohydrate (WSC) profiles in the leaves of Col-0, *atrs4.atrs5* and *atrs4.atrs5/MtStaS* lines.** Mass spracta representing water soluble carbohydrates (WSCs) extracted from (A) *atrs4.atrs5/MtStaS* mutant transgenic leaves, (B) *atrs4.atrs5* mutant leaves, (C) Col-0 leaves and (D) 2 mM galactinol (Gol), raffinose (Raf), and stachyose (Sta) as reference compounds. Mass spracta shows no detectable gain of Sta in any lines.

### 3.3.6 Constitutive expression of *MtStaS* is unable to recover ablated Sta in mature seeds of *atrs4.atrs5*

Water soluble carbohydrates (WSCs) were extracted from mature seeds of Col-0, *atrs4.atrs5, atrs4.atrs5/MtStaS* mutants, and analysed by LC-MS/MS. RFOs present in Col-0 mature seeds were Gol, Raf and Sta. The only RFO present in mature seeds from *atrs4.atrs5* was Gol and a total loss of detectable Raf and Sta is observed. Noticeably missing, was both Raf and Sta. Constitutive expression of *MtStaS* in *atrs4.atrs5,* however, was unable to recover ablated Raf and Sta in mature seeds (Figure 6).

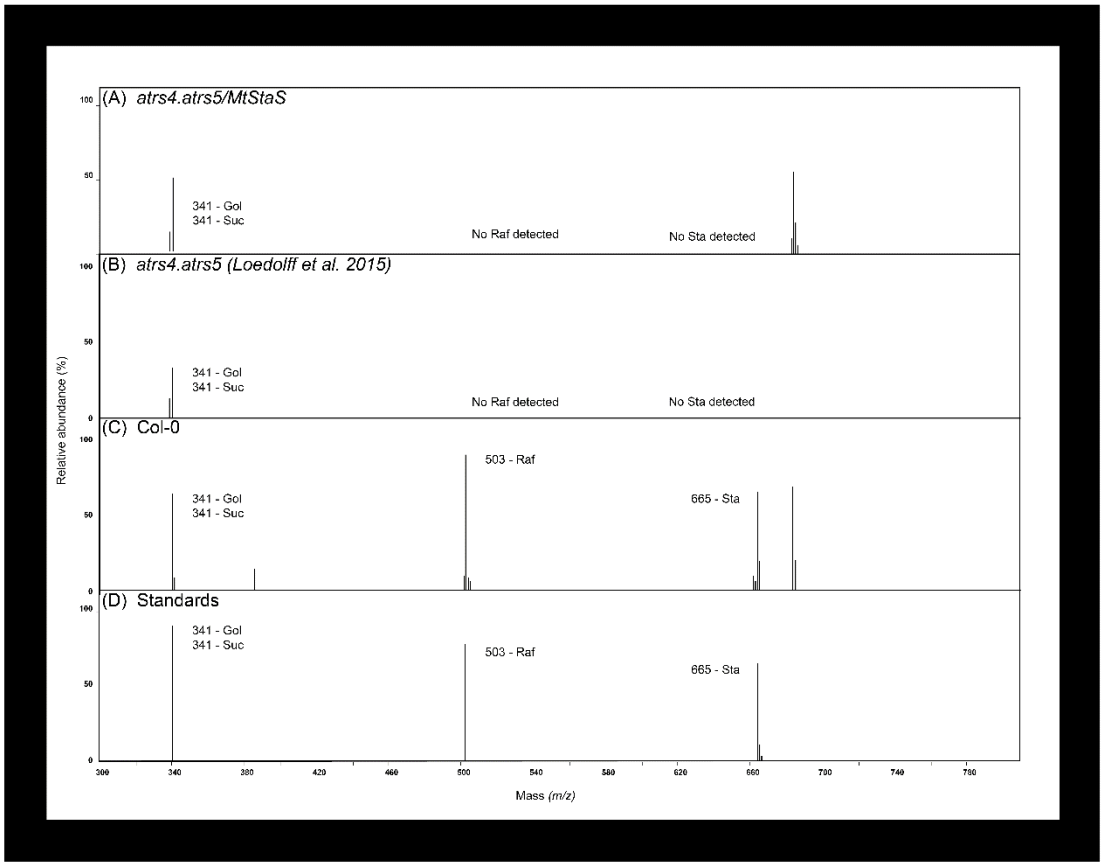


**Figure 6. Water soluble carbohydrate (WSC) profiles in the seeds of Col-0, *atrs4.atrs5* and *atrs4.atrs5/ MtStaS* lines.** Mass spracta representing water soluble carbohydrates (WSCs) extracted from (A) *atrs4. atrs5/MtStaS* mutant transgenic seeds, (B) *atrs4.atrs5* mutant seeds, (C) Col-0 seeds and (D) 2 mM galactinol (Gol), raffinose (Raf), and stachyose (Sta) as reference compounds. Mass spracta shows a total loss of Sta and Raf in mutant *atrs4.atrs5* seeds and the phenotype is unrecoverable in *atrs4.atrs5/MtStaS.*

## 3.4 DISCUSSION

To date only a few StaSs have been functionally identified and these are from *Cucurbita pepo, Cucumis melo, Vigna angularis, Lens culinaris, Pisum sativum* and *Arabidopsis thaliana* (Gaudreault and Webb 1981; Huber *et al.* 1990; Holthaus and Schmitz 1991; Peterbauer and Richter 1998; Hoch *et al.* 1999; Pluskota *et al.* 2015; Gangl *et al.* 2015). Of these, two have demonstrated to present additional biosynthetic capacity (*in vitro*). StaS from *A. thaliana* has reported to be a sequential bifunctional (Gol-dependent) RafS and a high affinity StaS (Gangl *et al.* 2015). Using the substrates Suc and Gol, AtStaS was able to biosynthesise Raf and Sta, whereas substrates Raf and Gol produced Sta, only.

However, α1,6-galactosyltransferases with multi-functional capabilities are not unprecedented. The multi-functional StaS from *Pisum sativum* (PsStaS; Peterbauer *et al.* 2002) is able to synthesise verbascose (Ver, Suc-$Gal_3$) *in vitro via* a Gol-independent manner (utilising Raf and Sta) but also able to synthesise Sta *via* a Gol-dependent manner (utilising Raf and Gol). Long-chain sucrosyl oligosaccharides (≤ Suc-$Gal_{13}$) are reported to accumulate in *Ajuga reptans,* a labiate with a high degree of freeze tolerance (Bachmann *et al.* 1994; Bachmann and Keller 1995; Peters and Keller 2009) and have led to the suggestion that long-chain RFOs further facilitate abiotic stress tolerance. These RFOs are catalysed in a Gol-independent fashion *via* galactan:galactan galactosyl transferase (GGT) – a unique chain elongation enzyme possessing the ability to use RFOs as both galactosyl donors and acceptors (Haab and Keller 2002).

Since a functional and biochemical role for AtStaS and AtRafS was recently described in RFO accumulation of *Arabidopsis* (Gangl *et al.* 2015; Gangl and Tenhaken 2016), we opted for a strategy of using previously described *atrs4* single and *atrs4.atrs5* double T-DNA insertion mutants in order to understand the functional role of *MtStaS* (Loedolff *et al.* 2015). Once *MtStaS* had been bioinformatically identified in chapter 2, we generated a single expression construct where *MtStaS* expression was constitutive (CaMV35S promoter, pMDC32, Curtis and Grossniklaus 2003). Together with the *atrs4* and *atrs4.atrs5* T-DNA insertion mutants, we examined the functional contribution MtStaS might have on Raf and Sta accumulation in the leaves and seeds

of these plants. The double mutant (*atrs4.atrs5*) was included to test for enzyme bifunctionality.

*Arabidopsis thaliana* is characterised by the absence of Sta in the leaves (Taji *et al.* 2002; Egert *et al.* 2013; Gangl *et al.* 2015). In seeds however, RFOs accumulate up to Sta in substantial quantities (Ooms *et al.* 1993; Bentsink *et al.* 2000; Nishizawa-Yokoi *et al.* 2008; Egert *et al.* 2013; Gangl and Tenhaken 2016). Accumulation of RFOs throughout seed development and maturation is believed to serve a critical role as osmoprotectants in providing desiccation tolerance, longevity in the dehydrated state and vigour upon subsequent germination. Sta is believed to contribute significant energy during seed germination (Downie *et al.* 2003; Salvi *et al.* 2016; Li *et al.* 2017).

Mature seeds from *atrs4* exhibited a total loss of detectable Sta but have also been shown to have increased levels of Raf (Loedolff *et al.* 2015; Gangl *et al.* 2015). It was therefore interesting when it was reported that *atrs4* mature seeds germinated to marginally earlier than wild-type, suggestive of Sta having almost no effect on the germination time period or that additional Raf present contributed towards a quicker germination time period. The *atrs4/MtStaS* is able to catalyse the synthesis of Sta from Raf and Gol in both mature seeds and leaves. Accumulation of Sta in leaves is rare, however, not unprecedented where its purpose is believed to serve as a long distance translocate and its presence in leaves have been reported for *Cucumis melo* and *Cucurbita pepo* (Gaudreault and Webb 1981; Holthaus and Schmitz 1991).

This work has demonstrated Sta detection in the leaves and mature seeds of *atrs4/MtStaS*, leading to a recovery of the ablated Sta in *atrs4* in the seeds. Future work should consider including germination experiments using wild-type, *atrs4* and *atrs4/MtStaS* seeds to conclude whether *atrs4/MtStaS* with the recovered ablated Sta is able to germinate at the same pace as wild-type or potentially slower/faster. Furthermore, the absence of any Sta is observed in the double T-DNA insertion mutant (*atrs4.atrs5/MtStaS*) seeds and leaves. This is due to the requirement for Raf (missing in *atrs4.atrs5*) and Gol by StaS in order to produce Sta (Kandler and Hopf 1982). With no Raf present in *atrs4.atrs5/MtStaS* mutants, Sta is unable to be generated in transgenic leaves and mature seeds provided that MtStaS is not bifunctional. To a large extent, this eludes to a non-bifunctional MtStaS (with regards

to both RafS and StaS capability) and currently, *in sillico* reports suggest that only a single *MtStaS* isoform exists (Vandecasteele *et al.* 2011). Therefore, failure to synthesise Sta in *atrs4.atrs5/MtStaS* mature seeds and leaves sheds light on the inability of MtStaS to behave bifunctionally in biosynthesising both Raf and Sta as observed with AtStaS (Gangl *et al.* 2015). Future work should include leaf crude enzyme assays from *atrs4.atrs5/MtStaS* lines. After purification, enzyme extracts should be incubated in substrates: Galactinol (Gol) and Sucrose (Suc) to assay for potential RafS capability and Gol and Raf to confirm StaS ability, *in vitro.*

Additionally, future experimentation should focus on quantifying the accumulation of WSC content within the various lines to be able to compare whether *atrs4/MtStaS* is able to recover Sta content to level as found within the wild-type. Furthermore, it would be interesting to inspect if leaf *atrs4/MtStaS* Sta content would be the same level as seed *atrs4/MtStaS* content.

A recent, novel finding demonstrated mature seeds from *atrs4.atrs5* double mutant displayed a five-day delayed germination phenotype in darkness compared to control. This phenotype was recovered either in light or partly recovered by supplementing the media with galactose in darkness. This signifies that rapid seed germination in the dark require RFOs which also serve as an essential source of galactose in seeds (Gangl and Tenhaken 2016).

In this study, it is the first time to the best of our knowledge, to functionally identify any recombinant StaS *in vivo* using *A. thaliana* as a heterologous platform. In conclusion, we have demonstrated using a single T-DNA insertion mutant (*atrs4)* that is completely devoid of detectable Sta in seeds, is able to recover ablated Sta and complement the missing pathway when expressing recombinant MtStaS. We have also demonstrated using a double T-DNA insertion mutant (*atrs4.atrs5)* that is completely devoid of Raf and Sta accumulation in seeds, is unable to recover ablated Sta when expressing recombinant MtStaS and does not possess any RafS capacity. We therefore conclude that MtStaS is a *bona fide* stachyose synthase.

# Chapter IV: Heterologous expression of *MtStaS* in *Yarrowia lipolytica*

## 4.1 INTRODUCTION

Commonly referred to as the workhorse of molecular biology, *Escherichia coli* was also the first microbial host exploited as a heterologous expression platform. This prokaryotic system offers numerous advantages attributed mainly to its tremendously well-studied physiology, biochemistry and genetics (Rosano and Ceccarelli 2014). Some advantages include its short generation time (20-30 minutes), ability to grow on minimal media, high protein production rate and a wide choice in available promoter and regulatory sequences (Gopal and Kumar 2013). *E. coli* microbial heterologous expression strategies, using the RFO biosynthetic enzymes GolS, RafS and to a lesser extent StaS, have been employed extensively to biochemically characterise RFO synthases *in vitro* (Hoch *et al.* 1999; Peterbauer *et al.* 2002b; Li *et al.* 2007; Peters *et al.* 2007; Pillet *et al.* 2012; Sui *et al.* 2012; Egert *et al.* 2013; Pluskota *et al.* 2015; Gangl *et al.* 2015; Salvi *et al.* 2016).

However, bacteria are of a simpler design than their eukaryotic counterparts. Other than disulphide bond formation, more advanced post-translational modifications, including fatty acid acylation, phosphorylation, amidation and glycosylation are not native to bacteria. Protein production which necessitate such post-translational modifications would require *in vitro* manipulations or different expression hosts (Frommer and Ninnemann 1995; Gupta *et al.* 2013). *E. coli* does not possess the required machinery in order to remove introns from transcripts which possess a massive problem in the expression of foreign genes containing introns (Gomes *et al.* 2016).

Certain proteins require to be processed in eukaryotic host cells in order to retain their biological activity and various alternate expression platforms are able to address the issue of post-translational modification. Insect cell systems are extensively utilised as alternative cell lines to express recombinant proteins, viral pesticides and vaccines which require post-translational modification (Unger and Peleg 2012; Cox 2012). Baculoviruses are usually propagated in insect cell lines derived from *Spodoptera frugiperda* (fall armyworm). Frequently used and commercially available insect cell lines (Sf9, Sf21) have been utilised for the expression of recombinant proteins through molecular cloning of baculovirus vectors (Altmann *et al.* 1999; Smagghe *et al.* 2009).

In addition to the capacity to co-translate and post-translate desired recombinant proteins, insect cell lines are comparatively cheap to maintain and can be scaled up with relative ease (Wu *et al.* 1992; King *et al.* 1992; Shi and Jarvis 2007). The Sf9/21 insect cell systems have also previously been employed to study the *in vitro* biochemical characterisation of *V. angularis StaS*, *P. sativum StaS* and *A. thaliana AtSIP2* (Peterbauer *et al.* 1999, 2002b; Peters *et al.* 2010).

In terms of low requirements needed and ease of cultivation, the only other eukaryotic host that draws near *E. coli*, however, is yeast. With the ability to execute many post-translational modifications required from higher eukaryotes, yeast is also able to efficiently secrete proteins into the culture and uniquely allows for site-specific integration (Glick *et al.* 2010; Mattanovich *et al.* 2012; Gaillardin *et al.* 2013).

In recent years, microbiological research has focussed more on the "non-conventional" yeasts other than *Saccharomyces cerevisiae*. Amongst this list, *Yarrowia lipolytica* is receiving much more attention and is fast becoming one of the most attractive and extensively studied model organisms for its genetic and physiological versatility. *Y. lipolytica* is widely considered by some to be one of the most attractive host organisms for heterologous protein production (Barth and Gaillardin 1996; Barth *et al.* 1997; Gonçalves *et al.* 2014). This can be mainly attributed to its capability in secreting high molecular weight proteins into the medium at high levels (Domînguez *et al.* 1998; Müller *et al.* 1998; Madzak *et al.* 2004; Nicaud 2012). Additionally, genetically modified strains and various expression vectors have been developed (Madzak *et al.* 2000, 2004; Nicaud *et al.* 2002; Yue *et al.* 2008). Moreover, heterologous gene expression is ideal in *Y. lipolytica* because it natively secretes several proteins like proteases, lipases, esterases and RNase in media, simplifying the production of a mass-produced protein of interest (Nicaud *et al.* 2002; Nicaud 2012).

This chapter describes the heterologous expression of a StaS from *M. truncatula* (MtStaS; Medtr7g106910.1) in *Y. lipolytica* – identified *in sillico* in chapter 2 to be a StaS and functionally identified *in vivo* in chapter 3 to be a StaS. This work lays the foundation for future use of *Y. lipolytica* as an ideal heterologous platform to study

RFO biosynthesis *in vitro* and provides valuable information for further biochemical characterisation of MtStaS.

## 4.2 MATERIALS AND METHODS

Unless specified otherwise, chemicals used throughout this study were obtained from Sigma-Aldrich® (www.sigmaaldrich.com/south-africa.html) or MERCK® (Modderfontein, South Africa). The Oligo explorer® software (V1.4 BETA) was used to design the primers which were subsequently synthesised by Inqaba Biotech®. All enzymes used in this study were obtained from New England Biolabs® (NEB, Inqaba Biotechnical Industries (Pty) Ltd, South Africa), unless stated otherwise. Vectors and primers used in this study are summarised in Table 1.

### 4.2.1 Heterologous expression of MtStaS

#### 4.2.1.1 Microorganism and vector

*Yarrowia lipolytica* (Po1g strain, Table 1) and the expression vector pYLEX used in this study were obtained from Yeastern Biotech Co. (Taipei, Taiwan). The *Y. lipolytica* expression vector (pYLEX, 7259 bp) contains the hybrid promoter (hp4d) and a transcription terminator signal. The vector also contains *LEU2* (leucine selection marker gene), effectively complementing the deletion of the *LEU2* gene in the Po1g parent strain.

#### 4.2.1.2 Cloning MtStaS into pYLEX expression cassette

The *MtStaS* gene was amplified from *M. truncatula* cDNA as outlined in chapter 2 (2.2.3) as the template with the use of primers that were designed to introduce a *Pml*I site (AATG) upstream of the first codon, and a *Kpn*I site (GGTACC) after the stop codon of the mature peptide. The gene was amplified using Q5® High-Fidelity DNA Polymerase (New England Biolabs®) *via* PCR according to the manufacturer's instructions. An amplicon (~2.5 kb) was recovered and subsequently purified using the Wizard® SV Gel and PCR Clean-up System (Promega) following the manufacturer's instructions by means of gel electrophoresis (0.8%; w/v; 60 V). The amplicon was then cloned into pYLEX digested with *Pml*I and *Kpn*I according to manufacturer's protocol.

Plasmid propagation throughout the expression work was achieved with *E. coli* One Shot® TOP10 (Invitrogen). Transformed colonies were screened for the sense orientation (5' to 3') by restriction analysis and subsequently sequenced (Central Analytical Facility, Stellenbosch University, South Africa). Linearization of the corresponding vector, *pYLEX::MtStaS,* was achieved with the *Not*I restriction enzyme, prior to transformation of *Y. lipolytica* Po1g by use of the lithium acetate method (Xuan *et al.* 1988). *Y. lipolytica* cells were grown in YPD medium (pH 4; 20 h) and thereafter incubated in the YLOS cocktail buffer (Yeastern Biotech) with approximately 90 ng of the linearized vector (39°C; 60 min). Transformants of the *Y. lipolytica* Po1g were selected for Leu+ prototrophy on YNB plates without leucine, and grown for three days at 28°C.

### 4.2.2 DNA extraction, RNA extraction and cDNA synthesis

The total genomic DNA was extracted from *Y. lipoytica* employing the 'bust n' grab' protocol as previously described (Harju *et al.* 2004). Total RNA was prepared from 50 ml of transformed and untransformed *Y. lipolytica* cultures using the RNeasy® Mini kit (Qiagen, Whitehead Scientific, South Africa), according to the manufacturer's instruction. The quantity and purity of RNA was measured using a NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific, Inqaba biotech, South Africa). RNA quality was evaluated on denaturing electrophoretic gels stained with ethidium bromide (1.0%; w/v; 60 V). Purified RNA aliquots were subsequently stored at -80 °C until further use. Complementary DNA (cDNA) was synthesised as outlined in chapter 2 (2.2.3).

### 4.2.3 Transcript analysis

Transcript analysis was conducted as outlined in Chapter 2 (2.2.4).

*ACT* served as the reference gene in the analyses (Table 1; Rzechonek *et al.* 2017). The threshold cycle number ($\Delta C_T$) was used to calculate relative fold change with the $\Delta\Delta C_T$ method, using the untransformed *Y. lipolytica* as the calibrator sample (Livak and Schmittgen 2001). The mean $C_T$ value of three technical replicates were analysed for every biological replicate. *ACT* served as the reference gene in all analyses. All qPCR experiments were conducted in accordance with the "Minimum

Information for Publication of Quantitative Realtime PCR Experiments" (MIQE, Bustin *et al.* 2009).

### 4.2.4 SDS-PAGE analysis

Proteins were separated by SDS-PAGE using a 12% gel prepared as previously described by Laemmli (1970). Crude protein aliquots of 20 µl were isolated as described below (4.2.5). Extractions were mixed with 10 µl of 2X SDS-PAGE sample buffer (1.5 M Tris-Cl pH 6.8, 20% SDS, 30% glycerol, 10% β-mercaptoethanol and 1.8 mg bromophenol blue). The protein samples were denatured by boiling for 2 min, and then loaded on an SDS-PAGE gel for electrophoresis using the Mini-PROTEAN Tetra Cell system (Bio-Rad). Electrophoresis was conducted using SDS running buffer (25 mM Tris-HCl, 200 mM glycine and 0.1% [w/v] SDS). The samples were then electrophoresed at 200 V for 1 h. The SDS-PAGE gel was thereafter incubated in fixing solution (25% isopropanol and 10% acetic acid) for 20 min and visualised by staining the gel with Coomassie Brilliant Blue (10% [v/v] acetic acid, 0.003% [w/v] Coomassie Brilliant Blue G, 10% [v/v] isopropanol, Sigma-Aldrich®) for 10 hours at room temperature and subsequently washed with destaining solution (5% methanol, 10% [v/v] acetic acid) for 30 min, then repeated until necessary to remove traces of Coomassie Blue stain until protein bands were visible. The PageRuler™ prestained protein ladder (ThermoFischer) served as the size standards.

### 4.2.5 Recombinant protein expression and enzyme assays

All the operations were carried out at 4°C. Total protein extracts were harvested from transformed and untransformed cells during mid-log phase ($OD_{600}$ ~ 0.6). Cultures (50 ml) were centrifuged (4500 *g* for 20 min), pellets were resuspended in 2 ml extraction buffer (50 mM HEPES/KOH pH 7.0, 1 mM EDTA, 20 mM DTT, 0.1% (v/v) Triton X-100, 1 mM benzamidine, 1 mM PMSF, 50 mM Na ascorbate, 2% (w/v) PVP) containing 2 g of glass beads (1–1.5 mm, Sigma-Aldrich®). Lysis was performed by vortexing (4 cycles, 30 s each, followed by cooling on ice for 30 s) and supernatants were recovered by centrifugation at 6000 *g* for 5 min. 200 µl of the supernatant was subsequently desalted *via* gel filtration (1400 *g*, 2 min) through a Sephadex G-25 (fine, Sigma-Aldrich®) column (final bed volume of 3 ml). Sephadex columns were pre-equilibrated twice by centrifugation (1400 *g*, 2 min) with 2 ml of assay buffer (50 mM HEPES/KOH pH7.0, 10 mM DTT). Enzyme activities were assayed, using 50 µl

aliquots of the desalted extracts, in a final volume of 100 µl assay buffer containing either 100 mM Raf and 10 mM Gol for StaS activity or 100mM Suc and 10mM Gol for RafS activity. Assays were performed for 1 h at 30°C and reactions were subsequently stopped by boiling (5 min, 80°C). Samples were desalted as previously described (Peters *et al.* 2007; Peters and Keller 2009; Egert *et al.* 2013) prior to LC-MS/MS analysis as outlined in chapter 3 (3.2.6).

### 4.2.6 Statistical analyses

Statistical analyses were performed as outlined in chapter 2 (2.2.5).

**Table 1. Strains, vectors, and primers used in this study**

| Name | Characteristics | Use | Supplier/Reference |
|---|---|---|---|
| **Strains** | | | |
| *Yarrowia lipolytica* Po1g | Genotype: *MatA, leu2-270, ura3-302::URA3, xpr2-332, axp-2*; phenotype: *Leu −, ΔAEP, ΔAXP, Suc+, pBR platform* | Parental strain-host for expression of recombinant *MtStaS* gene | Yeastern Biotech Co., Ltd., Taiwan |
| *Escherichia coli* One Shot® TOP10 | Genotype: *F- mcrA Δ( mrr-hsdRMS-mcrBC) Φ80lacZΔM15 Δ lacX74 recA1 araD139 Δ( araleu)7697 galU galK rpsL (StrR) endA1 nupG* | Host for routine cloning, vector propagation, assembly of complete vector | Invitrogen |
| **Vectors** | | | |
| pGEM-T-Easy | https://worldwide.promega.com/resources/protocols/technical-manuals/0/pgem-t-and-pgem-t-easy-vector-systems-protocol/ | Subcloning of the *MtStaS* gene, sequencing | Promega |
| pYLEX | pBR322 backbone, hybrid promoter (hp4d), leucine gene (LEU2)—selection marker; *MtStaS* gene was cloned in *PmlI/KpnI* sites | Expression cassette used for cloning and transformation of *Y. lipolytica* Po1g strain | Yeastern Biotech Co., Ltd., Taiwan |
| **Primers (5' - 3')** | | | |
| MtStaS_PmlI_F MtStaS_Kpn_R | CCACAATG ATGGCTCCACCGA AAGGTACC GGTGTTTCTGATTT | Amplification of the respective DNA fragmentfor for construction into the pYLEX vector | This study |
| MtStaS_CDS_F MtStaS_CDS_R | ATGGCTCCACCGAATTCCACAAACCT GGTGTTTCTGATTTGGCAATTTTCTTCTAG | *MtStaS* gene amplification in the pYLEX vector | This study |
| MtStaS_Q_F MtStaS_Q_R | AGGTGGTGGGAATTTCCTTG TTTCCATCACCTAGCCACTC | Quantitative real-time PCR (qPCR) analyses | This study |
| ACT_Q_F ACT_Q_R | CGAGCGAATGCACAAGGA GAGCGGTGATCTTGACCTTGA | Quantitative real-time PCR (qPCR) analyses | Rzechonek *et al.* 2017 |

## 4.3 RESULTS

The full-length open reading frame of *MtStaS* cDNA was cloned into the *Y. lipolytica* expression vector, pYLEX, to confirm biochemical function. MtStaS encodes 860 amino acids with a calculated molecular mass of approximately 95.97 kDa.

### 4.3.1 *Yarrowia lipolytica* transformation and screening of transformants

The *Y. lipolytica* Po1g strain was transformed using *Not*I linearized expression cassettes. The transformants of the *Y. lipolytica* Po1g were selected for Leu+ prototrophy on YNB agar medium supplemented with all amino acids except leucine. The primer pair *MtStaS_CDS_F and MtStaS_CDS_R* (Table 1) specific to *MtStaS* were used in PCR amplification to verify the integration of the cassette within the Leu+ prototrophic transformants of the *Y. lipolytica* yeast genome. PCR amplification was conducted as outlined in chapter 3 (3.2.4.1). The PCR product of approximately 2.5 kb confirmed the integration of the expression cassette (*pYLEX::MtStaS*; Figure 1A).

### 4.3.2 *MtStaS* expression confirmed *via* qPCR

Following confirmation of successful expression cassette (*pYLEX::MtStaS*) integration into the host genome (Section 4.3.1), *MtStaS* expression was analysed by qPCR through the use of primers listed in Table 1. *MtStaS* transcripts were absent in untransformed cells but detected in all transformed cells (Figure 1C).

### 4.3.3 MtStaS production confirmed *via* SDS-PAGE analysis

Crude cell extracts obtained from *Y. lipolytica* cells transformed with the expression cassette (*pYLEX::MtStaS*) were used for SDS-PAGE analysis. MtStaS protein production was confirmed by the presence of a ~95 kDa band in SDS-PAGE gel (Figure 1B). An untransformed *Y. lipolytica* control had no band present. The identity of the band could not be confirmed by mass spectrometry, however, since the protein of interest cannot be purified because it does not contain a polyhistidine-tag.

### 4.3.4 Recombinant MtStaS synthesises Sta using Raf and Gol *in vitro*

Crude cell extracts obtained from *Y. lipolytica* cells transformed with the expression cassette (*pYLEX::MtStaS*) were incubated in the presence of 100 mM raffinose (Raf) and 10 mM galactinol (Gol) for 1 h at pH 7, 30°C. The reactions were boiled, desalted and analysed by LC-MS/MS. Transformed *Y. lipolytica* was evidently able to synthesise a compound which eluted at the same retention time as commercial stachyose (Sta) standard (Figure 1D). In contrast, crude cell extracts from untransformed *Y. lipolytica* showed no Sta synthesis. Independent experiments were conducted using two individual *Y. lipolytica* colonies for enzymatic assays, and enzyme activities were measured in triplicate for each experiment.

### 4.3.5 Recombinant MtStaS is unable to synthesise Raf using Gol and Suc *in vitro*

Crude cell extracts obtained from *Y. lipolytica* cells transformed with the expression cassette (*pYLEX::MtStaS*) were incubated in the presence of 10 mM galactinol (Gol) and 100 mM sucrose (Suc) for 1 h at pH 7, 30°C. The reactions were boiled, desalted and analysed by LC-MS/MS. Transformed *Y. lipolytica* was unable to synthesise a compound which eluted at the same retention time as commercial raffinose (Raf) standard (data not shown). Similarly, crude cell extracts from untransformed *Y. lipolytica* showed no Raf synthesis. Independent experiments were conducted using two individual *Y. lipolytica* colonies for enzymatic assays, and enzyme activities were measured in triplicate for each experiment. MtStaS therefore does not possess any RafS bifunctional capabilities.

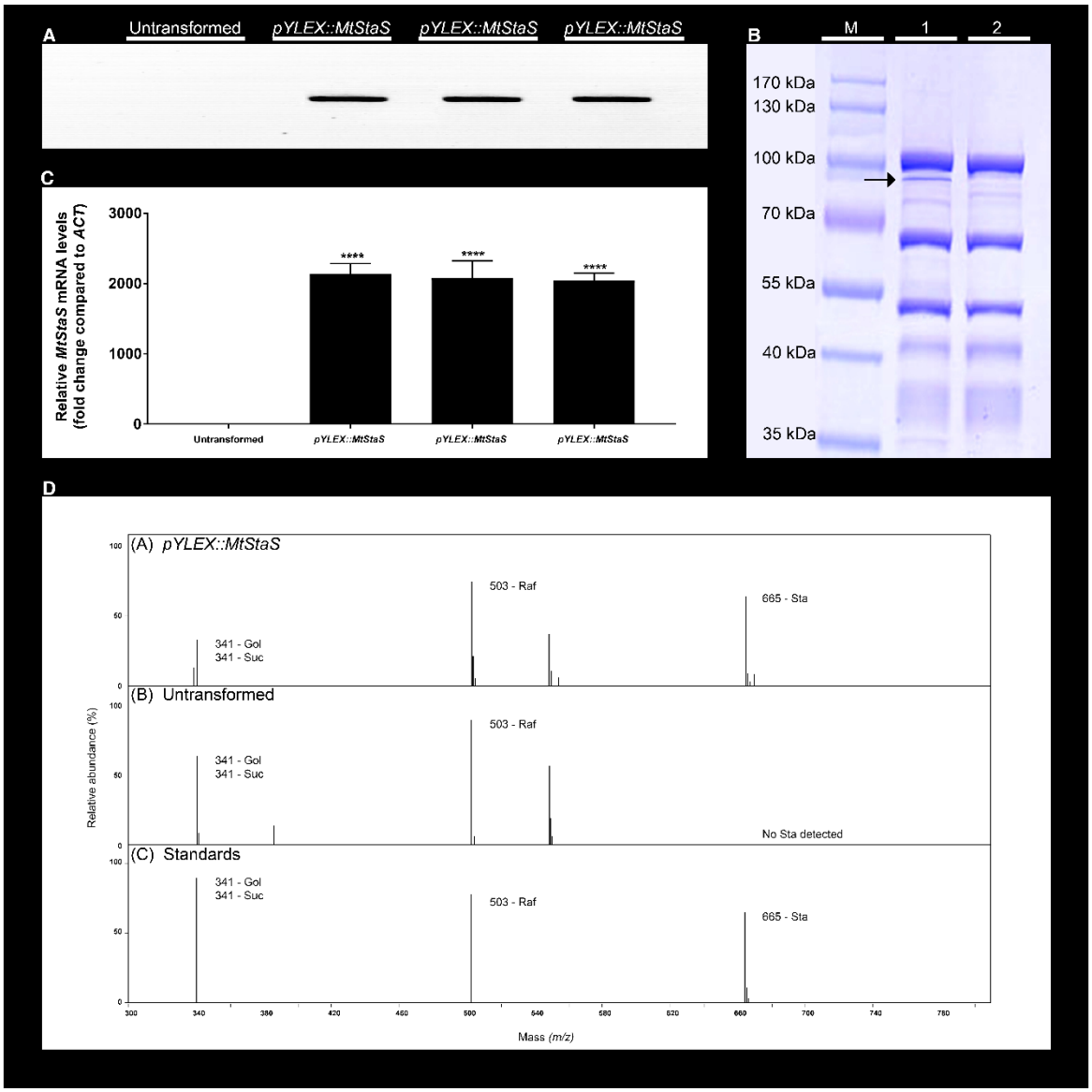


**Figure 1. Heterologous expression of *MtStaS* in *Y. lipolytica*.** (A) Image identifying the successful integration of the *pYLEX::MtStaS* expression cassette into the *Y. lipolytica* genome. Genomic DNA PCRs were conducted on DNA isolated from transformed and untransformed colonies of *Y. lipolytica*. Primer pairs amplified *MtStaS*. (B) SDS-PAGE analysis of *Y. lipolytica* lysates expressing the recombinant *MtStaS* gene products. Lanes; M – PageRuler™ prestained protein ladder. Molecular weight standards are indicated on the left, 1; crude protein extracted from *Y. lipolytica* cells expressing recombinant *M. truncatula* stachyose synthase (*MtStaS*), 2; crude protein extracted from untransformed *Y. lipolytica* cells (control). An arrowhead points to the MtStaS product only found in lane 1. Proteins were separated on a 12% SDS-PAGE gel and visualised with Coomassie Brilliant Blue stain. (C) Expression levels of *MtStaS* were determined by quantitative real-time PCR (qPCR) in transformed and untransformed cells. The threshold cycle number ($\Delta C_T$) was used to calculate relative fold change with the $\Delta\Delta C_T$ method, using untransformed cells as the calibrator sample (Livak and Schmittgen 2001). All qPCR experimentation was conducted in compliance with the "Minimum Information for Publication of Quantitative Real-Time PCR Experiments" (MIQE, Bustin *et al.* 2009). Data were normalised to *ACT* mRNA and relative mRNA levels are represented graphically as fold change compared to calibrator sample. A value of 1.0 represents no expression//transcript deficiency. Data represents mean ± SEM; n = 3 (each analysed in triplicate); ****$p < 0.0001$. (D) Mass spracta representing an *in vitro* Sta synthesis reaction conducted in the presence of 100 mM raffinose (Raf) and 10 mM galactinol (Gol). All the enzymatic assays were performed using crude *Y. lipolytica* cell lysates containing recombinant *pYLEX::MtStaS* or untransfomed, in 50 mM HEPES-KOH buffer pH 7.0, and incubated for 1 h at 30°C. The reactions were boiled, desalted and analysed by LC-MS/MS analysis. Independent experiments were conducted using two individual *Y. lipolytica* colonies for enzymatic assays, and enzyme activities were measured in triplicate for each experiment.

## 4.4 DISCUSSION

This study makes use of an expression vector originally constructed by Madzak *et al.* 2000. The vector is especially suited for this study as it carries a powerful hybrid hp4d promoter-based system that is able to continually direct protein expression without various influences (changes to nitrogen/carbon sources and pH values) that would otherwise decrease protein production in other *Yarrowia* promoter (e.g. pXPR2) systems (Madzak *et al.* 2000, 2004). Interestingly, with hp4d promoter-based expression systems, sustained protein accumulation is observed during the stationary phase as opposed to a peak value subsequently followed by a reduction in activity (Madzak *et al.* 2005; Yang *et al.* 2010). Studies have even shown hp4d-driven heterologous gene expression occurring predominantly when the organism enters the stationary phase (Nicaud *et al.* 2002; Celińska *et al.* 2015).

The hp4d promoter was engineered into an expression cassette followed directly downstream by *MtStaS* - contained within a single cassette. Once linearized, through homologous recombination, this cassette was stably integrated into the host genome of the genetically modified strain of *Y. lipolytica* (Po1g) and offers numerous significant advantages over traditional episomal expression platforms. Greater stability and effectively eliminating the requirement for selection pressure maintenance (particularly of substantial importance throughout complex large-scale cultivations) are amongst the advantages that integrative expression cassettes are offering (Celińska *et al.* 2015). Various heterologous proteins including laccases (Madzak *et al.* 2004; Jolivalt *et al.* 2005), lipase (Nicaud *et al.* 2002), cytokinin oxidase (Kopečný *et al.* 2005), prochymosin, and β-galactosidase (Madzak *et al.* 2000) have been produced by applying this effective hp4d promoter-based expression cassette system.

Owing to complications in expression and purification of functional StaS recombinant proteins, StaSs are largely biochemically uncharacterised (Gangl *et al.* 2015). Biochemical characterisation of StaS have previously relied on purified or crude enzyme extracts from leaf material or transformed Sf21 insect cell lysates (Peterbauer and Richter 1998; Hoch *et al.* 1999; Peterbauer *et al.* 1999). The first microbial

heterologously expressed StaS was only performed in 2015, owing to the difficulty of this endeavour (Gangl *et al.* 2015).

In this study, *Y. lipolytica* was exploited as the preferred host for the expression of a putative plant (*M. truncatula*) StaS to determine its biochemical identity and the host's potential worth in studying RFO biosynthesis. This study also conclusively proves to be the first time to our knowledge to purify any recombinant RFO using a yeast sytem. In utilising a eukaryotic microbial system (*Y. lipolytica*), we have effectively layed the foundation for future studies using this platform to study RFO biosynthesis, offering the possibility to surpass previous constraints experienced with bacterial systems (Loedolff *et al.* 2015).

Successive steps were taken to ensure that the expression cassette containing *MtStaS* (*pYLEX::MtStaS*) was stably integrated into host - *Y. lipolytica*. First of which was to screen transformants on selective medium and then *via* genomic DNA PCR amplification. Once confirmed, we tested for *MtStaS* transcript production and MtStaS protein production using an SDS-PAGE gel. The 12% SDS-PAGE gel analysis demonstrated that MtStaS proteins are indeed expressed in *Y. lipolytica* and are in fairly good agreement with the deduced molecular mass of 95.97 kDa as calculated from MtStaS cDNA amino acid sequence. Other characterised StaS are very similar in size: *Pisum sativum* – 95 kDa (Peterbauer *et al.* 2002a), *Vigna angularis* – 90 kDa (Peterbauer *et al.* 1999), *Cucumis melo* – 95 kDa *(Holthaus and Schmitz 1991)*, *Lens culinaris* – 88.6 kDa (Hoch *et al.* 1999), *Arabidopsis thaliana* – 100 kDa (Gangl *et al.* 2015).

The positive results from the enzymatic activity assays demonstrated that the recombinant *MtStaS* was successfully expressed in the *Y. lipolytica* system and characterised by LC-MS/MS to synthesise Sta from Raf and Gol *in vitro*. We can also confirm that MtStaS does not possess bifunctional capabilities in synthesising Raf *in vitro* as demonstrated when incubated in substrates Gol and Suc. We intend on determining whether MtStaS possesses Gol-independent activity in synthesising verbascose (utilising Raf and Sta) like *Pisum sativum* StaS (PsStaS; genbank acc: CAD20127; Peterbauer *et al.* 2002b) with which MtStaS shares 84% amino acid identity with (Chapter 2) but experimentation could not reach its final completion at the time of writing.

# Chapter V: General summary, conclusions and outlook

## 5.1 General summary, conclusions and outlook

The raffinose family oligosaccharides (RFOs; Suc-$Gal_n$, 13 < n ≥ 1) are sucrosyl-galactosyl oligosaccharides occurring exclusively in the plant kingdom and in some photoautotrophic algae serving critical roles in phloem translocation and carbon storage. RFO biosynthesis occurs linearly, involving GolS, RafS and StaS to produce Gol, Raf and Sta, respectively and sequentially.

In leguminous seeds, Sta is considered to be the major water soluble carbohydrate after sucrose (Wang *et al.* 2010). Despite its ambiguous role in serving as an essential source of carbon during germination, RFOs are implicated in a multitude of abdominal disorders upon ingestion in human and monogastric animals and can lead to severe health complications (Kumar *et al.* 2010).

*Medicago truncatula* is widely considered to be the choice model organism amongst all commercially important legumes (Gholami *et al.* 2014). Despite full sequence availability of the *M. truncatula* genome, it remains largely unannotated (Tang *et al.* 2014). In particular its RFO physiology, in the framework of identified biosynthetic genes, is almost completely unknown. In the genome of *M. truncatula*, a single putative *StaS*, seven *RafS*s and four *GolS*s were reported *in sillico*, however, none have been functionally characterised, their tissue specific localisation/s or their contribution to RFO physiology (Vandecasteele *et al.* 2011). As seen in *Arabidopsis, in sillico* identification of the *RafS* gene family (*atrs1*, *atrs2*, *atrs3*, *atrs4*, *atrs5* and *atrs6)* led to the discovery (amongst others) of a bifunctional StaSs, a pseudogene, a α-galactosidase and a monofunctional RafS. With seven RafSs reported in *M truncatula*, experiments should be conducted to definitively prove the role of each of these RafSs through *in planta* and *in vitro* studies.

In the work presented, we embarked on a multi-pronged approach to functionally characterise a *StaS* from *M. truncatula*. We have described experimental approaches to **(i)** identify a candidate gene through rudimentary bioinformatic analyses **ii)** Clone the candidate gene into a binary vector *pMDC32* (dual CaMV35s promoter), transform this construct into the *Arabidopsis thaliana atrs4 and atrs4.atrs5* T-DNA insertion mutants and **(iii)** heterologously express MtStaS in the dimorphic fungus, *Yarowia*

*lipolytica,* in order to biochemically characterise the recombinant protein, and particularly to ascertain if MtStaS has bifunctionality in synthesising RFOs.

The major findings of this work are highlighted below,

### 5.1.1 *In sillico* identification of *Medtr7g106910.1 (MtStaS)* eludes strongly to a *StaS*

We identified a candidate gene through rudimentary bioinformatic analysis using previously biochemically characterised StaS from *P. sativum* (PsStaS; CAC38094), *Vigna angularis* (VaStaS; CAB64363), *Cucumis melo* (CmStaS; XP_008451468) and *Arabidopsis thaliana* (AtStaS; NP_192106) demonstrating a conserved 80 amino acid long hallmark sequence shared by StaSs but, distinctly missing in RafSs. Medtr7g106910.1 showed high homology percentage to known StaSs, in particular, 84% amino acid identity to PsStaS.

### 5.1.2 RFO transcript analysis in *M. trunctula* is localised to specific tissue

We conducted comprehensive qPCR analyses on the expression profiles of *MtGolS*, *MtRafS* and *MtStaS* in various *M. truncatula* tissues (seeds, roots, stems and leaves). We found that all the candidate genes - *GolS*, *RafS* and *StaS* appear to exhibit high tissue-specific expression patterns. Interestingly, no significant *RafS* mRNA levels were detected in seeds in spite of various reports of Raf accumulation in *M. truncatula* seeds - hypothesising the involvement of at least another RafS or a bifunctional StaS.

### 5.1.3 Functional identification of *Medtr7g106910.1* - a *bona fide StaS* (*MtStaS*)

We employed a transgenic expression strategy to manipulate Sta concentrations *in planta* using previously described *Arabidopsis atrs4* (compromised in Sta accumulation) and *atrs4.atrs5* (compromised in Sta and Raf accumulation) T-DNA insertion mutants as novel heterologous platforms to dissect the functionality of the enzyme. Using the newly identified *MtStaS*, we created a plant overexpression construct where *MtStaS* expression is driven by dual CaMV35s promoters (*pMDC32::MtStaS*). We have demonstrated that *Arabidopsis* mutant plants (F1)

transformed with *pMDC32::MtStaS* exhibits constitutive expression of *MtStaS.* We also report on the successful functional identification of *M. truncatula StaS (Medtr7g106910.1*) and demonstrated that it was MtStaS responsible for the accumulation of Sta in transgenic *atrs4/MtStaS* leaves and recovering ablated Sta in mature seeds contrary to untransformed *atrs4* controls. The course of our investigation using the double knockout *atrs4.atrs5,* we determined MtStaS unable to synthesise Sta in the absence of Raf.

### 5.1.4 MtStaS is a α1,6-galactosyltransferase capable of biosynthesising Sta in the presence of Gol and Raf

Using the dimorphic fungus *Yarrowia lipolytica* as a heterologous expression system, we functionally expressed the *MtStaS* cDNA demonstrating that it showed a requirement for Gol and Raf to biosynthesise Sta *in vitro*. The recombinant MtStaS showed no RafS activity when incubated in substrates Gol and Suc, refuting any potential bifunctional capacity in biosynthesising Raf and maintains consistency with *in vivo* reports and deductions from chapter 3 using *atrs4.atrs5/MtStaS.* Work on the full biochemical characterisation of MtStaS could not be brought to its final conclusion at the time of writing due to time constrains but, we anticipate this data to be included in our publication submission.

## REFERENCE LIST

**Allen O., and E. Allen**, 1981 *The Leguminosae: A source book of characteristics, uses and nodulation.* Macmillan.

**Alonso J. M., A. N. Stepanova, T. J. Leisse, C. J. Kim, H. Chen, P. Shinn, D. K. Stevenson, J. Zimmerman, P. Barajas, R. Cheuk, C. Gadrinab, C. Heller, A. Jeske, E. Koesema, C. C. Meyers, H. Parker, L. Prednis, Y. Ansari, N. Choy, H. Deen, M. Geralt, N. Hazari, E. Hom, M. Karnes, C. Mulholland, R. Ndubaku, I. Schmidt, P. Guzman, L. Aguilar-Henonin, M. Schmid, D. Weigel, D. E. Carter, T. Marchand, E. Risseeuw, D. Brogden, A. Zeko, W. L. Crosby, C. C. Berry, and J. R. Ecker**, 2003 Genome-wide insertional mutagenesis of Arabidopsis thaliana. Science (80-. ). 301: 653–657. https://doi.org/10.1126/science.1086391

**Altmann F., E. Staudacher, I. B. H. Wilson, and L. März**, 1999 Insect cells as hosts for the expression of recombinant glycoproteins. Glycoconj. J. 16: 109–123. https://doi.org/10.1023/A:1026488408951

**Andersen K. E., C. Bjergegaard, P. Møller, J. C. Sørensen, and H. Sørensen**, 2005 Compositional variations for α-galactosides in different species of Leguminosae, Brassicaceae, and barley: A chemotaxonomic study based on chemometrics and high-performance capillary electrophoresis. J. Agric. Food Chem. 53: 5809–5817. https://doi.org/10.1021/jf040471v

**Angelovici R., G. Galili, A. R. Fernie, and A. Fait**, 2010 Seed desiccation: a bridge between maturation and germination. Trends Plant Sci. 15: 211–218. https://doi.org/10.1016/j.tplants.2010.01.003

**Arabidopsis Genome Initiative**, 2000 Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408: 796–815. https://doi.org/10.1038/35048692

**Ayre B. G., F. Keller, and R. Turgeon**, 2003 Symplastic Continuity between Companion Cells and the Translocation Stream: Long-Distance Transport Is Controlled by Retention and Retrieval Mechanisms in the Phloem. PLANT

Physiol. 131: 1518–1528. https://doi.org/10.1104/pp.012054

**Bachmann M., P. Matile, and F. Keller**, 1994 Metabolism of the raffinose family oligosaccharides in leaves of Ajuga reptans L.(cold acclimation, translocation, and sink to source transition: discovery of chain elongation enzyme). Plant Physiol. 105: 1335–1345.

**Bachmann M., and F. Keller**, 1995 Metabolism of the Raffinose Family Oligosaccharides in Leaves of Ajuga reptans L. (Inter- and Intracellular Compartmentation). Plant Physiol. 109: 991–998.

**Barker D. G., S. Bianchi, F. Blondon, Y. Dattée, G. Duc, S. Essad, P. Flament, P. Gallusci, G. Génier, P. Guy, X. Muel, J. Tourneur, J. Dénarié, and T. Huguet**, 1990 Medicago truncatula, a model plant for studying the molecular genetics of the Rhizobium-legume symbiosis. Plant Mol. Biol. Report. 8: 40–49. https://doi.org/10.1007/BF02668879

**Barneveld R. J. van**, 1999 Understanding the nutritional chemistry of lupin ( Lupinus spp.) seed to improve livestock production efficiency. Nutr. Res. Rev. 12: 203. https://doi.org/10.1079/095442299108728938

**Barth G., and C. Gaillardin**, 1996 Yarrowia lipolytica, pp. 313–388 in *Nonconventional Yeasts in Biotechnology*, Springer Berlin Heidelberg, Berlin, Heidelberg.

**Barth G., C. Gaillardin, T. R., H. H., M. I.A., Y. D., B. J., Y. T.W., *et al.***, 1997 Physiology and genetics of the dimorphic fungus *Yarrowia lipolytica*. FEMS Microbiol. Rev. 19: 219–237. https://doi.org/10.1111/j.1574-6976.1997.tb00299.x

**Bentsink L., C. Alonso-Blanco, D. Vreugdenhil, K. Tesnier, S. P. C. Groot, and M. Koornneef**, 2000 Genetic analysis of seed-soluble oligosaccharides in relation to seed storability of Arabidopsis. Plant Physiol. 124: 1595–1604. https://doi.org/10.1104/pp.124.4.1595

**Blackman S. A., R. L. Obendorf, and A. C. Leopold**, 1992 Maturation proteins and sugars in desiccation tolerance of developing soybean seeds. Plant

Physiol. 100: 225–30. https://doi.org/10.1104/pp.100.1.225

**Blöchl A., T. Peterbauer, and A. Richter**, 2007 Inhibition of raffinose oligosaccharide breakdown delays germination of pea seeds. J. Plant Physiol. 164: 1093–1096. https://doi.org/10.1016/j.jplph.2006.10.010

**Blondon F., D. Marie, S. Brown, and a Kondorosi**, 1994 Genome size and base composition in Medicago sativa and M. truncatula species. Genome 37: 264–270. https://doi.org/10.1139/g94-037

**Bolle C., A. Schneider, and D. Leister**, 2011 Perspectives on Systematic Analyses of Gene Function in Arabidopsis thaliana: New Tools, Topics and Trends. Curr. Genomics 12: 1–14. https://doi.org/10.2174/138920211794520187

**Bouché N., and D. Bouchez**, 2001 Arabidopsis gene knockout: Phenotypes wanted. Curr. Opin. Plant Biol. 4: 111–117. https://doi.org/10.1016/S1369-5266(00)00145-X

**Bucciarelli B., J. Hanan, D. Palmquist, and C. P. Vance**, 2006 A standardized method for analysis of Medicago truncatula phenotypic development. Plant Physiol. 142: 207–19. https://doi.org/10.1104/pp.106.082594

**Bustin S. A., V. Benes, J. A. Garson, J. Hellemans, J. Huggett, M. Kubista, R. Mueller, T. Nolan, M. W. Pfaffl, G. L. Shipley, J. Vandesompele, and C. T. Wittwer**, 2009 The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. Clin. Chem. 55: 611–622. https://doi.org/10.1373/clinchem.2008.112797

**Celińska E., W. Białas, M. Borkowska, and W. Grajek**, 2015 Cloning, expression, and purification of insect (Sitophilus oryzae) alpha-amylase, able to digest granular starch, in Yarrowia lipolytica host. Appl. Microbiol. Biotechnol. 99: 2727–2739. https://doi.org/10.1007/s00253-014-6314-2

**Chabaud M., C. Larsonneau, C. Marmouget, and T. Huguet**, 1996 Transformation of barrel medic (Medicago truncatula Gaertn.) by Agrobacterium tumefaciens and regeneration via somatic embryogenesis of transgenic plants with the MtENOD12 nodulin promoter fused to the gus reporter gene. Plant Cell

Rep. 15: 305–310. https://doi.org/10.1007/BF00232361

**Cho S. M., E. Y. Kang, M. S. Kim, S. J. Yoo, Y. J. Im, Y. C. Kim, K. Y. Yang, K. Y. Kim, K. S. Kim, Y. S. Choi, and B. H. Cho**, 2010 Jasmonate-dependent expression of a galactinol synthase gene is involved in priming of systemic fungal resistance in *Arabidopsis thaliana*. Botany 88: 452–461. https://doi.org/10.1139/B10-009

**Choi H. K., D. Kim, T. Uhm, E. Limpens, H. Lim, J. H. Mun, P. Kalo, R. V. Penmetsa, A. Seres, O. Kulikova, B. A. Roe, T. Bisseling, G. B. Kiss, and D. R. Cook**, 2004 A Sequence-Based Genetic Map of Medicago truncatula and Comparison of Marker Colinearity with M. sativa. Genetics 166: 1463–1502. https://doi.org/10.1534/genetics.166.3.1463

**Cook D. R.**, 1999 Medicago truncatula - A model in the making! Curr. Opin. Plant Biol. 2: 301–304. https://doi.org/10.1016/S1369-5266(99)80053-3

**Coon C. N., K. L. Leske, O. Akavanichan, and T. K. Cheng**, 1990 Effect of oligosaccharide-free soybean meal on true metabolizable energy and fiber digestion in adult roosters. Poult. Sci. 69: 787–793. https://doi.org/10.3382/ps.0690787

**Corbineau F., M. A. Picard, J.-A. Fougereux, F. Ladonne, and D. Côme**, 2000 Effects of dehydration conditions on desiccation tolerance of developing pea seeds as related to oligosaccharide content and cell membrane properties. Seed Sci. Res. 10: 329–339. https://doi.org/10.1017/S0960258500000374

**Couée I., C. Sulmon, G. Gouesbet, and A. El Amrani**, 2006 Involvement of soluble sugars in reactive oxygen species balance and responses to oxidative stress in plants. J. Exp. Bot. 57: 449–459. https://doi.org/10.1093/jxb/erj027

**Cox M. M. J.**, 2012 Recombinant protein vaccines produced in insect cells. Vaccine 30: 1759–1766. https://doi.org/10.1016/j.vaccine.2012.01.016

**Curtis M. D., and U. Grossniklaus**, 2003 A gateway cloning vector set for high-throughput functional analysis of genes in planta. Plant Physiol. 133: 462–9. https://doi.org/10.1104/pp.103.027979

**Czechowski T., M. Stitt, T. Altmann, M. K. Udvardi, and W.-R. Scheible**, 2005 Genome-Wide Identification and Testing of Superior Reference Genes for Transcript Normalization in Arabidopsis. PLANT Physiol. 139: 5–17. https://doi.org/10.1104/pp.105.063743

**Delalande M., S. Greene, S. Hughes, R. Nair, T. Huguet, M. E. Aouani, and J. M. Prosperi**, 2007 Wild accessions / populations Table of contents : Medicago truncatula Handb. 1–27.

**Dilis V., and A. Trichopoulou**, 2009 Nutritional and health properties of pulses. Med. J. Nutrition Metab. 1: 149–157. https://doi.org/10.1007/s12349-008-0023-2

**Dinant S., and R. Lemoine**, 2010 The phloem pathway: New issues and old debates. Comptes Rendus - Biol. 333: 307–319. https://doi.org/10.1016/j.crvi.2010.01.006

**Domînguez A., E. Ferminân, M. Sânchez, F. J. Gonzalez, F. M. Pérez-Campo, S. Garcia, A. B. Herrero, A. S. Vicente, J. Cabello, M. Prado, F. J. Iglesias, A. Choupina, F. J. Burguillo, L. Fernândcz-Lago, and M. Carmen Löpez**, 1998 Non-conventional yeasts as hosts for heterologous protein production. Int. Microbiol. 1: 131–142.

**Downie B., S. Gurusinghe, P. Dahal, R. R. Thacker, J. C. Snyder, H. Nonogaki, K. Yim, K. Fukanaga, V. Alvarado, and K. J. Bradford**, 2003 Expression of a GALACTINOL SYNTHASE gene in tomato seeds is up-regulated before maturation desiccation and again after imbibition whenever radicle protrusion is prevented. Plant Physiol. 131: 1347–59. https://doi.org/10.1104/pp.016386

**Edwards K., C. Johnstone, and C. Thompson**, 1991 A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. Nucleic Acids Res. 19: 1349. https://doi.org/10.1093/nar/19.6.1349

**Egert A., F. Keller, and S. Peters**, 2013 Abiotic stress-induced accumulation of raffinose in Arabidopsis leaves is mediated by a single raffinose synthase (RS5, At5g40390). BMC Plant Biol. 13: 218. https://doi.org/10.1186/1471-2229-13-218

**Felsenstein J.**, 1985 Confidence Limits on Phylogenies: An Approach Using the Bootstrap. Evolution (N. Y). 39: 783. https://doi.org/10.2307/2408678

**Frommer W. B., and O. Ninnemann**, 1995 Heterologous Expression of Genes in Bacterial, Fungal, Animal, and Plant Cells. Annu. Rev. Plant Physiol. Plant Mol. Biol. 46: 419–444. https://doi.org/10.1146/annurev.pp.46.060195.002223

**Gaillardin C., M. Mekouar, and C. Neuvéglise**, 2013 Comparative Genomics of Yarrowia lipolytica, pp. 1–30 in Springer, Berlin, Heidelberg.

**Gangl R., R. Behmüller, and R. Tenhaken**, 2015 Molecular cloning of AtRS4, a seed specific multifunctional RFO synthase/galactosylhydrolase in Arabidopsis thaliana. Front. Plant Sci. 6: 789. https://doi.org/10.3389/fpls.2015.00789

**Gangl R., and R. Tenhaken**, 2016 Raffinose Family Oligosaccharides Act As Galactose Stores in Seeds and Are Required for Rapid Germination of Arabidopsis in the Dark. Front. Plant Sci. 7: 1115. https://doi.org/10.3389/fpls.2016.01115

**Garcia J., D. G. Barker, and E.-P. Journet**, 2006 Medicago truncatula handbook Seed storage and germination 1 Introduction A -Seed storage and extraction from pods B -Germination of non-dormant seeds C -Breaking seed embryo dormancy D – Recovery and germination of immature embryos Appendix References Fig

**Gaudreault P. R., and J. A. Webb**, 1981 Stachyose synthesis in leaves of Cucurbita pepo. Phytochemistry 20: 2629–2633. https://doi.org/10.1016/0031-9422(81)85257-0

**Gholami A., N. De Geyter, J. Pollier, S. Goormachtig, and A. Goossens**, 2014 Natural product biosynthesis in Medicago species. Nat. Prod. Rep. 31: 356. https://doi.org/10.1039/c3np70104b

**Gilchrist E., and G. Haughn**, 2010 Reverse genetics techniques: Engineering loss and gain of gene function in plants. Briefings Funct. Genomics Proteomics 9: 103–110. https://doi.org/10.1093/bfgp/elp059

**Gitzelmann R., and S. Auricchio**, 1965 The handling of soya alpha-galactosides

by a normal and a galactosemic child. Pediatrics 36: 231–235.

**Glencross B. D., T. Boujard, and S. J. Kaushik**, 2003 Influence of oligosaccharides on the digestibility of lupin meals when fed to rainbow trout, Oncorhynchus mykiss. Aquaculture 219: 703–713. https://doi.org/10.1016/S0044-8486(02)00664-6

**Glick B. R., J. J. Pasternak, and C. L. Patten**, 2010 *Molecular biotechnology : principles and applications of recombinant DNA*. ASM Press.

**Gomes A. R., S. M. Byregowda, B. M. Veeregowda, and V. Balamurugan**, 2016 An Overview of Heterologous Expression Host Systems for the Production of Recombinant Proteins. Adv. Anim. Vet. Sci. 4: 346–356. https://doi.org/10.14737/journal.aavs/2016/4.7.346.356

**Gonçalves F. A. G., G. Colen, and J. A. Takahashi**, 2014 Yarrowia lipolytica and its multiple applications in the biotechnological industry. Sci. World J. 2014: 476207. https://doi.org/10.1155/2014/476207

**Gopal G. J., and A. Kumar**, 2013 Strategies for the production of recombinant protein in escherichia coli. Protein J. 32: 419–425. https://doi.org/10.1007/s10930-013-9502-5

**Graham P. H., C. P. Vance, and T. R. McDermott**, 2003 Legumes: importance and constraints to greater use. Plant Physiol. 131: 872–7. https://doi.org/10.1104/pp.017004

**Greene E. A., C. A. Codomo, N. E. Taylor, J. G. Henikoff, B. J. Till, S. H. Reynolds, L. C. Enns, C. Burtner, J. E. Johnson, A. R. Odden, L. Comai, and S. Henikoff**, 2003 Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in Arabidopsis. Genetics 164: 731–40.

**Gumbmann M. R., and S. N. Williams**, 1971 The Quantitative Collection and Determination of Hydrogen Gas From the Rat and Factors Affecting Its Production. Exp. Biol. Med. 137: 1171–1175. https://doi.org/10.3181/00379727-137-35749

**Gupta S., N. Adlakha, and S. S. Yazdani**, 2013 Efficient extracellular secretion of

an endoglucanase and a β-glucosidase in E. coli. Protein Expr. Purif. 88: 20–25. https://doi.org/10.1016/j.pep.2012.11.006

**Haab C. I., and F. Keller**, 2002 Purification and characterization of the raffinose oligosaccharide chain elongation enzyme, galactan:galactan galactosyltransferase (GGT), from Ajuga reptans leaves. Physiol. Plant. 114: 361–371. https://doi.org/10.1034/j.1399-3054.2002.1140305.x

**Harju S., H. Fedosyuk, and K. R. Peterson**, 2004 Rapid isolation of yeast genomic DNA: Bust n' Grab. BMC Biotechnol. 4: 8. https://doi.org/10.1186/1472-6750-4-8

**Hartwig E. E., T. M. Kuo, and M. M. Kenty**, 1997 Seed protein and its relationship to soluble sugars in soybean. Crop Sci. 37: 770–773. https://doi.org/10.2135/cropsci1997.0011183X003700030013x

**Hedges S. B.**, 2002 The origin and evolution of model organisms. Nat. Rev. Genet. 3: 838–849. https://doi.org/10.1038/nrg929

**Hoch G., T. Peterbauer, and A. Richter**, 1999 Purification and characterization of stachyose synthase from lentil (Lens culinaris) seeds: Galactopinitol and stachyose synthesis. Arch. Biochem. Biophys. 366: 75–81. https://doi.org/10.1006/abbi.1999.1212

**Holthaus U., and K. Schmitz**, 1991 Stachyose synthesis in mature leaves of Cucumis melo. Purification and characterization of stachyose synthase (EC 2.4.1.67). Planta 184: 525–531. https://doi.org/10.1007/BF00197902

**Horbowicz M., and R. L. Obendorf**, 1994 Seed desiccation tolerance and storability: Dependence on flatulence-producing oligosaccharides and cyclitols—review and survey. Seed Sci. Res. 4: 385–405. https://doi.org/10.1017/S0960258500002440

**Huber J. L. A., D. Mason Pharr, and S. C. Huber**, 1990 Partial purification and characterization of stachyose syynthase in leaves of Cucumis sativus and Cucumis melo: utilization of a rapid assay for myo-inositol. Plant Sci. 69: 179–188. https://doi.org/10.1016/0168-9452(90)90116-6

**Huynh B. L., L. Palmer, D. E. Mather, H. Wallwork, R. D. Graham, R. M. Welch, and J. C. R. Stangoulis**, 2008 Genotypic variation in wheat grain fructan content revealed by a simplified HPLC method. J. Cereal Sci. 48: 369–378. https://doi.org/10.1016/j.jcs.2007.10.004

**Joersbo M., S. Guldager Pedersen, J. E. Nielsen, J. Marcussen, and J. Brunstedt**, 1999 Isolation and expression of two cDNA clones encoding UDP-galactose epimerase expressed in developing seeds of the endospermous legume guar. Plant Sci. 142: 147–154. https://doi.org/10.1016/S0168-9452(99)00012-6

**Jolivalt C., C. Madzak, A. Brault, E. Caminade, C. Malosse, and C. Mougin**, 2005 Expression of laccase IIIb from the white-rot fungus Trametes versicolor in the yeast Yarrowia lipolytica for environmental applications. Appl. Microbiol. Biotechnol. 66: 450–456. https://doi.org/10.1007/s00253-004-1717-0

**Jukes T. H., and C. R. Cantor**, 1969 Evolution of Protein Molecules. Mamm. Protein Metab. 21–132. https://doi.org/10.1016/B978-1-4832-3211-9.50009-7

**Kakar K., M. Wandrey, T. Czechowski, T. Gaertner, W. R. Scheible, M. Stitt, I. Torres-Jerez, Y. Xiao, J. C. Redman, H. C. Wu, F. Cheung, C. D. Town, and M. K. Udvardi**, 2008 A community resource for high-throughput quantitative RT-PCR analysis of transcription factor gene expression in Medicago truncatula. Plant Methods 4: 18. https://doi.org/10.1186/1746-4811-4-18

**Kandler O., and H. Hopf**, 1982 Oligosaccharides Based on Sucrose (Sucrosyl Oligosaccharides), pp. 348–383 in *Plant Carbohydrates I SE - 8*, Springer Berlin Heidelberg, Berlin, Heidelberg.

**Kaplan F., J. Kopka, D. W. Haskell, W. Zhao, K. C. Schiller, N. Gatzke, D. Y. Sung, and C. L. Guy**, 2004 Exploring the Temperature-Stress Metabolome of Arabidopsis. Plant Physiol. 136: 4159–4168. https://doi.org/10.1104/pp.104.052142

**Keller F., and D. M. Pharr**, 1996 Metabolism of carbohydrates in sinks and sources: galactosyl-sucrose oligosaccharides., pp. 157–183 in *Photoassimilate distribution in plants and crops: source-sink relationships.*,.

**Kim M. S., Y. C. Kim, and B. H. Cho**, 2004 Gene expression analysis in cucumber leaves primed by root colonization with Pseudomonas chlororaphis O6 upon challenge-inoculation with Corynespora cassiicola. Plant Biol. 6: 105–108. https://doi.org/10.1055/s-2004-817803

**Kim M. S., S. M. Cho, E. Y. Kang, Y. J. Im, H. Hwangbo, Y. C. Kim, C.-M. Ryu, K. Y. Yang, G. C. Chung, and B. H. Cho**, 2008 Galactinol Is a Signaling Component of the Induced Systemic Resistance Caused by *Pseudomonas chlororaphis* O6 Root Colonization. Mol. Plant-Microbe Interact. 21: 1643–1653. https://doi.org/10.1094/MPMI-21-12-1643

**King G. A., A. J. Daugulis, P. Faulkner, and M. F. A. Goosen**, 1992 Recombinant β-Galactosidase Production in Serum-Free Medium by Insect Cells in a 14-L Airlift Bioreactor. Biotechnol. Prog. 8: 567–571. https://doi.org/10.1021/bp00018a015

**Knaupp M., K. B. Mishra, L. Nedbal, and A. G. Heyer**, 2011 Evidence for a role of raffinose in stabilizing photosystem II during freeze-thaw cycles. Planta 234: 477–486. https://doi.org/10.1007/s00425-011-1413-0

**Koncz C., and J. Schell**, 1986 The promoter of TL-DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by a novel type of Agrobacterium binary vector. MGG Mol. Gen. Genet. 204: 383–396. https://doi.org/10.1007/BF00331014

**Kopečný D., C. Pethe, M. Šebela, N. Houba-Hérin, C. Madzak, A. Majira, and M. Laloue**, 2005 High-level expression and characterization of Zea mays cytokinin oxidase/dehydrogenase in Yarrowia lipolytica. Biochimie 87: 1011–1022. https://doi.org/10.1016/j.biochi.2005.04.006

**Kottek M., J. Grieser, C. Beck, B. Rudolf, and F. Rubel**, 2006 World map of the Köppen-Geiger climate classification updated. Meteorol. Zeitschrift 15: 259–263. https://doi.org/10.1127/0941-2948/2006/0130

**Krause D. O., R. A. Easter, and R. I. Mackie**, 1994 Fermentation of stachyose and raffinose by hind-gut bacteria of the weanling pig. Lett. Appl. Microbiol. 18: 349–352. https://doi.org/10.1111/j.1472-765X.1994.tb00887.x

**Krysan P. J., J. C. Young, and M. R. Sussman**, 1999 T-DNA as an insertional mutagen in Arabidopsis. Plant Cell 11: 2283–90.

**Kumar V., A. Rani, L. Goyal, A. K. Dixit, J. G. Manjaya, J. Dev, and M. Swamy**, 2010 Sucrose and raffinose family oligosaccharides (RFOs) in soybean seeds as influenced by genotype and growing location. J. Agric. Food Chem. 58: 5081–5085. https://doi.org/10.1021/jf903141s

**Kumar S., G. Stecher, M. Li, C. Knyaz, and K. Tamura**, 2018 MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms, (F. U. Battistuzzi, Ed.). Mol. Biol. Evol. 35: 1547–1549. https://doi.org/10.1093/molbev/msy096

**Kurbel S., B. Kurbel, and A. Včev**, 2006 Intestinal gases and flatulence: Possible causes of occurrence. Med. Hypotheses 67: 235–239. https://doi.org/10.1016/j.mehy.2006.01.057

**Laurie R. E., P. Diwadkar, M. Jaudal, L. Zhang, V. Hecht, J. Wen, M. Tadege, K. S. Mysore, J. Putterill, J. L. Weller, and R. C. Macknight**, 2011 The Medicago FLOWERING LOCUS T Homolog, MtFTa1, Is a Key Regulator of Flowering Time. Plant Physiol. 156: 2207–2224. https://doi.org/10.1104/pp.111.180182

**Lehle L., W. Tanner, and O. Kandler**, 1970 myo-Inositol, a Cofactor in the Biosynthesis of Raffmose. Hoppe. Seylers. Z. Physiol. Chem. 351: 1494–1498. https://doi.org/10.1515/bchm2.1970.351.2.1494

**Lesins K. A., and I. Lesins**, 1979 *Genus Medicago (Leguminosae): A Taxogenetic Study*. Springer Netherlands, Dordrecht.

**Lewis G. P.**, 2005 *Legumes of the world*. Royal Botanic Gardens, Kew.

**Li S., T. Li, W. D. Kim, M. Kitaoka, S. Yoshida, M. Nakajima, and H. Kobayashi**, 2007 Characterization of raffinose synthase from rice (Oryza sativa L. var. Nipponbare). Biotechnol. Lett. 29: 635–640. https://doi.org/10.1007/s10529-006-9268-3

**Li T., Y. Zhang, D. Wang, Y. Liu, L. M. A. Dirk, J. Goodman, A. B. Downie, J. Wang, G. Wang, and T. Zhao**, 2017 Regulation of Seed Vigor by Manipulation

of Raffinose Family Oligosaccharides in Maize and Arabidopsis thaliana. Mol. Plant 10: 1540–1555. https://doi.org/10.1016/j.molp.2017.10.014

**Liu J. J., W. Odegard, and B. O. de Lumen**, 1995 Galactinol synthase from kidney bean cotyledon and zucchini leaf. Purification and N-terminal sequences. Plant Physiol. 109: 505–511. https://doi.org/10.1104/pp.109.2.505

**Livak K. J., and T. D. Schmittgen**, 2001 Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. Methods 25: 402–408. https://doi.org/10.1006/meth.2001.1262

**Loedolff B., S. Peters, and J. Kossmann**, 2015 Functional roles of raffinose family oligosaccharides: Arabidopsis case studies in seed physiology, biotic stress and novel carbohydrate engineering

**Madore M. A., D. E. Mitchell, and C. M. Boyd**, 1988 Stachyose Synthesis in Source Leaf Tissues of the CAM Plant Xerosicyos danguyi H. Humb. Plant Physiol. 87: 588–591.

**Madzak C., B. Tréton, and S. Blanchin-Roland**, 2000 Strong hybrid promoters and integrative expression/secretion vectors for quasi-constitutive expression of heterologous proteins in the yeast Yarrowia lipolytica. J. Mol. Microbiol. Biotechnol. 2: 207–216.

**Madzak C., C. Gaillardin, and J. M. Beckerich**, 2004 Heterologous protein expression and secretion in the non-conventional yeast Yarrowia lipolytica: A review. J. Biotechnol. 109: 63–81. https://doi.org/10.1016/j.jbiotec.2003.10.027

**Madzak C., L. Otterbein, M. Chamkha, S. Moukha, M. Asther, C. Gaillardin, and J. M. Beckerich**, 2005 Heterologous production of a laccase from the basidiomycete Pycnoporus cinnabarinus in the dimorphic yeast Yarrowia lipolytica. FEMS Yeast Res. 5: 635–646. https://doi.org/10.1016/j.femsyr.2004.10.009

**Martín-Cabrejas M. A., M. F. Díaz, Y. Aguilera, V. Benítez, E. Mollá, and R. M. Esteban**, 2008 Influence of germination on the soluble carbohydrates and dietary fibre fractions in non-conventional legumes. Food Chem. 107: 1045–

1052. https://doi.org/10.1016/j.foodchem.2007.09.020

**Martínez-Villaluenga C., J. Frias, and C. Vidal-Valverde**, 2008 Alpha-galactosides: Antinutritional factors or functional ingredients? Crit. Rev. Food Sci. Nutr. 48: 301–316. https://doi.org/10.1080/10408390701326243

**Mattanovich D., P. Branduardi, L. Dato, B. Gasser, M. Sauer, and D. Porro**, 2012 Recombinant Protein Production in Yeasts, pp. 329–358 in *Methods in molecular biology (Clifton, N.J.)*,.

**Müller S., T. Sandal, P. Kamp-Hansen, and H. Dalbøge**, 1998 Comparison of expression systems in the yeasts Saccharomyces cerevisiae, Hansenula polymorpha, Klyveromyces lactis, Schizosaccharomyces pombe and Yarrowia lipolytica. Cloning of two novel promoters from Yarrowia lipolytica. Yeast 14: 1267–1283. https://doi.org/10.1002/(SICI)1097-0061(1998100)14:14<1267::AID-YEA327>3.0.CO;2-2

**Naczk M., R. Amarowicz, and F. Shahidi**, 1997 α-Galactosides of Sucrose in Foods: Composition, Flatulence-Causing Effects, and Removal, pp. 127–151 in *Antinutrients and Phytochemicals in Food*,.

**Narusaka M., T. Shiraishi, M. Iwabuchi, and Y. Narusaka**, 2010 The floral inoculating protocol: A simplified Arabidopsis thaliana transformation method modified from floral dipping. Plant Biotechnol. 27: 349–351. https://doi.org/10.5511/plantbiotechnology.27.349

**Nei M., and S. Kumar**, 2000 *Molecular evolution and phylogenetics*. Oxford University Press.

**Nicaud J.-M., C. Madzak, P. van den Broek, C. Gysler, P. Duboc, P. Niederberger, C. Gaillardin, P. Van Den Broek, C. Gysler, P. Duboc, P. Niederberger, and C. Gaillardin**, 2002 Protein expression and secretion in the yeast Yarrowia lipolytica. FEMS Yeast Res. 2: 371–9. https://doi.org/10.1111/j.1567-1364.2002.tb00106.x

**Nicaud J. M.**, 2012 Yarrowia lipolytica. Yeast 29: 409–418. https://doi.org/10.1002/yea.2921

**Nichols P. G. H., A. Loi, B. J. Nutt, P. M. Evans, A. D. Craig, B. C. Pengelly, B. S. Dear, D. L. Lloyd, C. K. Revell, R. M. Nair, M. A. Ewing, J. G. Howieson, G. A. Auricht, J. H. Howie, G. A. Sandral, S. J. Carr, C. T. de Koning, B. F. Hackney, G. J. Crocker, R. Snowball, S. J. Hughes, E. J. Hall, K. J. Foster, P. W. Skinner, M. J. Barbetti, and M. P. You**, 2007 New annual and short-lived perennial pasture legumes for Australian agriculture-15 years of revolution. F. Crop. Res. 104: 10–23. https://doi.org/10.1016/j.fcr.2007.03.016

**Nishizawa-Yokoi A., Y. Yabuta, and S. Shigeoka**, 2008 The contribution of carbohydrates including raffinose family oligosaccharides and sugar alcohols to protection of plant cells from oxidative damage. Plant Signal. Behav. 3: 1016–1018. https://doi.org/10.4161/psb.6738

**Nishizawa A., Y. Yabuta, E. Yoshida, T. Maruta, K. Yoshimura, and S. Shigeoka**, 2006 Arabidopsis heat shock transcription factor A2 as a key regulator in response to several types of environmental stress. Plant J. 48: 535–547. https://doi.org/10.1111/j.1365-313X.2006.02889.x

**Nishizawa A., Y. Yabuta, and S. Shigeoka**, 2008 Galactinol and Raffinose Constitute a Novel Function to Protect Plants from Oxidative Damage. Plant Physiol. 147: 1251–1263. https://doi.org/10.1104/pp.108.122465

**O'Malley R. C., and J. R. Ecker**, 2010 Linking genotype to phenotype using the Arabidopsis unimutant collection. Plant J. 61: 928–940. https://doi.org/10.1111/j.1365-313X.2010.04119.x

**O'Neill N. R., and G. R. Bauchan**, 2000 Sources of Resistance to Anthracnose in the Annual *Medicago* Core Collection. Plant Dis. 84: 261–267. https://doi.org/10.1094/PDIS.2000.84.3.261

**Ooms J., K. M. Leon-Kloosterziel, D. Bartels, M. Koornneef, and C. M. Karssen**, 1993 Acquisition of Desiccation Tolerance and Longevity in Seeds of Arabidopsis thaliana (A Comparative Study Using Abscisic Acid-Insensitive abi3 Mutants). Plant Physiol. 102: 1185–1191. https://doi.org/10.1104/pp.102.4.1185

**Panikulangara T. J., G. Eggers-Schumacher, M. Wunderlich, H. Stransky, and F. Schöffl**, 2004 Galactinol synthase1. A Novel Heat Shock Factor Target

Gene Responsible for Heat-Induced Synthesis of Raffinose Family Oligosaccharides in Arabidopsis. PLANT Physiol. 136: 3148–3158. https://doi.org/10.1104/pp.104.042606

**Peoples M. B., A. M. Bowman, R. R. Gault, D. F. Herridge, M. H. McCallum, K. M. McCormick, R. M. Norton, I. J. Rochester, G. J. Scammell, and G. D. Schwenke**, 2001 Factors regulating the contributions of fixed nitrogen by pasture and crop legumes to different farming systems of eastern Australia. Plant Soil 228: 29–41. https://doi.org/10.1023/A:1004799703040

**Perry A.**, 1997 The Mediterranean : environment and society / edited by Russell King, Lindsay Proudfoot and Bernard Smith., pp. 30–44 in *The Mediterranean: Environment and Society*, A Hodder Arnold Publication. Arnold.

**Peterbauer T., and A. Richter**, 1998 Galactosylononitol and stachyose synthesis in seeds of adzuki bean. Purification and characterization of stachyose synthase. Plant Physiol. 117: 165–72. https://doi.org/10.1104/pp.117.1.165

**Peterbauer T., J. Mucha, U. Mayer, M. Popp, J. Glössl, and A. Richter**, 1999 Stachyose synthesis in seeds of adzuki bean (Vigna angularis): Molecular cloning and functional expression of stachyose synthase. Plant J. 20: 509–518. https://doi.org/10.1046/j.1365-313X.1999.00618.x

**Peterbauer T., and A. Richter**, 2001 Biochemistry and physiology of raffinose family oligosaccharides and galactosyl cyclitols in seeds. Seed Sci. Res. 11: 185–197. https://doi.org/10.1079/SSR200175

**Peterbauer T., J. Mucha, L. Mach, and A. Richter**, 2002a Chain elongation of raffinose in pea seeds. Isolation, characterization, and molecular cloning of a multifunctional enzyme catalyzing the synthesis of stachyose and verbascose. J. Biol. Chem. 277: 194–200. https://doi.org/10.1074/jbc.M109734200

**Peterbauer T., L. Mach, J. Mucha, and A. Richter**, 2002b Functional expression of a cDNA encoding pea (Pisum sativum L.) raffinose synthase, partial purification of the enzyme from maturing seeds, and steady-state kinetic analysis of raffinose synthesis. Planta 215: 839–846. https://doi.org/10.1007/s00425-002-0804-7

**Peters S., S. G. Mundree, J. A. Thomson, J. M. Farrant, and F. Keller**, 2007 Protection mechanisms in the resurrection plant Xerophyta viscosa (Baker): Both sucrose and raffinose family oligosaccharides (RFOs) accumulate in leaves in response to water deficit. J. Exp. Bot. 58: 1947–1956. https://doi.org/10.1093/jxb/erm056

**Peters S., and F. Keller**, 2009 Frost tolerance in excised leaves of the common bugle (Ajuga reptans L.) correlates positively with the concentrations of raffinose family oligosaccharides (RFOs). Plant, Cell Environ. 32: 1099–1107. https://doi.org/10.1111/j.1365-3040.2009.01991.x

**Peters S., A. Egert, B. Stieger, and F. Keller**, 2010 Functional identification of Arabidopsis ATSIP2 (At3g57520) as an alkaline α-galactosidase with a substrate specificity for raffinose and an apparent sink-specific expression pattern. Plant Cell Physiol. 51: 1815–1819. https://doi.org/10.1093/pcp/pcq127

**Pillet J., A. Egert, P. Pieri, F. Lecourieux, C. Kappel, J. Charon, E. Gomès, F. Keller, S. Delrot, and D. Lecourieux**, 2012 VvGOLS1 and VvHsfA2 are Involved in the Heat Stress Responses in Grapevine Berries. Plant Cell Physiol. 53: 1776–1792. https://doi.org/10.1093/pcp/pcs121

**Pluskota W. E., J. Szablińska, R. L. Obendorf, R. J. Górecki, and L. B. Lahuta**, 2015 Osmotic stress induces genes, enzymes and accumulation of galactinol, raffinose and stachyose in seedlings of pea (Pisum sativum L.). Acta Physiol. Plant. 37: 156. https://doi.org/10.1007/s11738-015-1905-9

**Puckridge D. W., and R. J. French**, 1983 The annual legume pasture in cereal-Ley farming systems of southern Australia: A review. Agric. Ecosyst. Environ. 9: 229–267. https://doi.org/10.1016/0167-8809(83)90100-7

**Pukacka S., E. Ratajczak, and E. Kalemba**, 2009 Non-reducing sugar levels in beech (Fagus sylvatica) seeds as related to withstanding desiccation and storage. J. Plant Physiol. 166: 1381–1390. https://doi.org/10.1016/j.jplph.2009.02.013

**Quemener B., and J. M. Brillouet**, 1983 Ciceritol, a pinitol digalactoside form seeds of chickpea, lentil and white lupin. Phytochemistry 22: 1745–1751.

https://doi.org/10.1016/S0031-9422(00)80263-0

**Quinlivan B. J.**, 1965 The naturalised and cultivated annual medics of Western Australia. J. Dep. Agric. West. Aust. 6.

**Reddy N. R., M. D. Pierson, S. K. Sathe, and D. K. Salunkhe**, 1984 Chemical, nutritional and physiological aspects of dry bean carbohydrates—A review. Food Chem. 13: 25–68. https://doi.org/10.1016/0308-8146(84)90026-8

**Ribeiro M., C. R. Felix, and S. D. P. Lozzi**, 2000 Soybean seed galactinol synthase activity as determined by a novel colorimetric assay. Rev. Bras. Fisiol. Veg. 12: 203–212. https://doi.org/10.1590/S0103-31312000000300004

**Rosano G. L., and E. A. Ceccarelli**, 2014 Recombinant protein expression in Escherichia coli: advances and challenges. Front. Microbiol. 5: 172. https://doi.org/10.3389/fmicb.2014.00172

**Rosnoblet C., C. Aubry, O. Leprince, B. L. Vu, H. Rogniaux, and J. Buitink**, 2007 The regulatory gamma subunit SNF4b of the sucrose non-fermenting-related kinase complex is involved in longevity and stachyose accumulation during maturation of Medicago truncatula seeds. Plant J. 51: 47–59. https://doi.org/10.1111/j.1365-313X.2007.03116.x

**Rzechonek D. A., C. Neuvéglise, H. Devillers, W. Rymowicz, and A. M. Mirończuk**, 2017 EUF1-A newly identified gene involved in erythritol utilization in Yarrowia lipolytica. Sci. Rep. 7: 12507. https://doi.org/10.1038/s41598-017-12715-7

**Saitou N., and M. Nei**, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4: 406–425. https://doi.org/10.1093/oxfordjournals.molbev.a040454

**Salvi P., S. C. Saxena, B. P. Petla, N. U. Kamble, H. Kaur, P. Verma, V. Rao, S. Ghosh, and M. Majee**, 2016 Differentially expressed galactinol synthase(s) in chickpea are implicated in seed vigor and longevity by limiting the age induced ROS accumulation. Sci. Rep. 6: 35088. https://doi.org/10.1038/srep35088

**Saunders D. R., and H. S. Wiggins**, 1981 Conservation of mannitol, lactulose, and

raffinose by the human colon. Am. J. Physiol. 241: G397-402. https://doi.org/10.1152/ajpgi.1981.241.5.G397

**Shi X., and D. Jarvis**, 2007 Protein N-Glycosylation in the Baculovirus-Insect Cell System. Curr. Drug Targets 8: 1116–1125. https://doi.org/10.2174/138945007782151360

**Sinha P., V. K. Singh, V. Suryanarayana, L. Krishnamurthy, R. K. Saxena, and R. K. Varshney**, 2015 Evaluation and validation of housekeeping genes as reference for gene expression studies in pigeonpea (cajanus cajan ) under drought stress conditions, (G. K. Pandey, Ed.). PLoS One 10: e0122847. https://doi.org/10.1371/journal.pone.0122847

**Smagghe G., C. L. Goodman, and D. Stanley**, 2009 Insect cell culture and applications to research and pest management. Vitr. Cell. Dev. Biol. - Anim. 45: 93–105. https://doi.org/10.1007/s11626-009-9181-x

**Smil V.**, 1999 Nitrogen in crop production: An account of global flows. Global Biogeochem. Cycles 13: 647–662. https://doi.org/10.1029/1999GB900015

**Smith P. T., T. M. Kuo, and C. G. Crawford**, 1991 Purification and characterization of galactinol synthase from mature zucchini squash leaves. Plant Physiol. 96: 693–8. https://doi.org/10.1104/pp.96.3.693

**Smýkal P., C. J. Coyne, M. J. Ambrose, N. Maxted, H. Schaefer, M. W. Blair, J. Berger, S. L. Greene, M. N. Nelson, N. Besharat, T. Vymyslický, C. Toker, R. K. Saxena, M. Roorkiwal, M. K. Pandey, J. Hu, Y. H. Li, L. X. Wang, Y. Guo, L. J. Qiu, R. J. Redden, and R. K. Varshney**, 2015 Legume Crops Phylogeny and Genetic Diversity for Science and Breeding. CRC. Crit. Rev. Plant Sci. 34: 43–104. https://doi.org/10.1080/07352689.2014.897904

**Song H., P. Wang, C. Li, S. Han, J. Lopez-Baltazar, X. Zhang, and X. Wang**, 2016 Identification of lipoxygenase (LOX) genes from legumes and their responses in wild type and cultivated peanut upon *Aspergillus flavus* infection. Sci. Rep. 6: 35245. https://doi.org/10.1038/srep35245

**Sosulski F. W., L. Elkowicz, and R. D. Reichert**, 1982 Oligosaccharides in Eleven

Legumes and Their Air-Classified Protein and Starch Fractions. J. Food Sci. 47: 498–502. https://doi.org/10.1111/j.1365-2621.1982.tb10111.x

**Steele K. P., S. M. Ickert-Bond, S. Zarre, and M. F. Wojciechowski**, 2010 Phylogeny and character evolution in *Medicago* (Leguminosae): Evidence from analyses of plastid trnK/matK and nuclear GA3ox1 sequences. Am. J. Bot. 97: 1142–1155. https://doi.org/10.3732/ajb.1000009

**Stevenson J. M., I. Y. Perera, I. Heilmann, S. Persson, and W. F. Boss**, 2000 Inositol signaling and plant growth. Trends Plant Sci. 5: 252–258. https://doi.org/10.1016/S1360-1385(00)01652-6

**Sui X. L., F. Z. Meng, H. Y. Wang, Y. X. Wei, R. F. Li, Z. yu Wang, L. P. Hu, S. H. Wang, and Z. X. Zhang**, 2012 Molecular cloning, characteristics and low temperature response of raffinose synthase gene in Cucumis sativus L. J. Plant Physiol. 169: 1883–1891. https://doi.org/10.1016/j.jplph.2012.07.019

**Swanepoel P. A., J. Labuschagne, and M. B. Hardy**, 2016 Historical development and future perspective of conservation agriculture practices in crop-pasture rotation systems in the Mediterranean region of South Africa. Ecosyst. Serv. socio-economic benefits Mediterr. grasslands. https://doi.org/10.1016/S0376-7361(09)70018-4

**Swanepoel P. A., and F. Tshuma**, 2017 Soil quality effects on regeneration of annual *Medicago* pastures in the Swartland of South Africa. African J. Range Forage Sci. 34: 201–208. https://doi.org/10.2989/10220119.2017.1403462

**Swennen K., C. M. Courtin, and J. A. Delcour**, 2006 Non-digestible oligosaccharides with prebiotic properties. Crit. Rev. Food Sci. Nutr. 46: 459–471. https://doi.org/10.1080/10408390500215746

**Taji T., C. Ohsumi, S. Iuchi, M. Seki, M. Kasuga, M. Kobayashi, K. Yamaguchi-Shinozaki, and K. Shinozaki**, 2002 Important roles of drought- and cold-inducible genes for galactinol synthase in stress tolerance in Arabidopsis thaliana. Plant J. 29: 417–426. https://doi.org/10.1046/j.0960-7412.2001.01227.x

**Tang H., V. Krishnakumar, S. Bidwell, B. Rosen, A. Chan, S. Zhou, L. Gentzbittel, K. L. Childs, M. Yandell, H. Gundlach, K. F. X. Mayer, D. C. Schwartz, and C. D. Town**, 2014 An improved genome release (version Mt4.0) for the model legume Medicago truncatula. BMC Genomics 15: 312. https://doi.org/10.1186/1471-2164-15-312

**Thoquet P., M. Ghérardi, E. P. Journet, A. Kereszt, J. M. Ané, J. M. Prosperi, and T. Huguet**, 2002 The molecular genetic linkage map of the model legume Medicago truncatula: An essential tool for comparative legume genomics and the isolation of agronomically important genes. BMC Plant Biol. 2: 1. https://doi.org/10.1186/1471-2229-2-1

**Turgeon R.**, 1996 Phloem loading and plasmodesmata. Trends Plant Sci. 1: 418–423. https://doi.org/10.1016/S1360-1385(96)10045-5

**Unger T., and Y. Peleg**, 2012 Recombinant protein expression in the baculovirus-infected insect cell system, pp. 187–199 in *Methods in Molecular Biology*,.

**Vandecasteele C., B. Teulat-Merah, M. C. Morère-Le Paven, O. Leprince, B. Ly Vu, L. Viau, L. Ledroit, S. Pelletier, N. Payet, P. Satour, C. Lebras, K. Gallardo, T. Huguet, A. M. Limami, J. M. Prosperi, and J. Buitink**, 2011 Quantitative trait loci analysis reveals a correlation between the ratio of sucrose/raffinose family oligosaccharides and seed vigour in Medicago truncatula. Plant, Cell Environ. 34: 1473–1487. https://doi.org/10.1111/j.1365-3040.2011.02346.x

**Vasileva V., and O. Kostov**, 2015 Effect of Alfalfa Grown for Forage on Soil Fertility Related to Mineral and Organic Fertilization. Emirates J. Food Agric. 27: 678. https://doi.org/10.9755/ejfa.2015.05.288

**Wagner J. R., R. Becker, M. R. Gumbmann, and A. C. Olson**, 1976 Hydrogen Production in the Rat Following Ingestion oÃ-Rafiinose , Stachyose and Bean Residue. J. Nutr. 106: 466–470. https://doi.org/10.1093/jn/106.4.466

**Wakiuchi N., R. Shiomi, and H. Tamaki**, 2003 Production of galactinol from sucrose by plant enzymes. Biosci. Biotechnol. Biochem. 67: 1465–1471. https://doi.org/10.1271/bbb.67.1465

**Wang N., D. W. Hatcher, R. T. Tyler, R. Toews, and E. J. Gawalko**, 2010 Effect of cooking on the composition of beans (Phaseolus vulgaris L.) and chickpeas (Cicer arietinum L.). Food Res. Int. 43: 589–594. https://doi.org/10.1016/j.foodres.2009.07.012

**Wu J., G. A. King, A. J. Daugulis, P. Faulkner, and M. F. A. Goosen**, 1992 Recombinant protein production in insect cell cultures infected with a temperature-sensitive baculovirus. Cytotechnology 9: 141–147. https://doi.org/10.1007/BF02521741

**Xuan J. W., P. Fournier, and C. Gaillardin**, 1988 Cloning of the LYS5 gene encoding saccharopine dehydrogenase from the yeast Yarrowia lipolytica by target integration. Curr. Genet. 14: 15–21. https://doi.org/10.1007/BF00405848

**Xue H., X. Chen, and G. Li**, 2007 Involvement of phospholipid signaling in plant growth and hormone effects. Curr. Opin. Plant Biol. 10: 483–489. https://doi.org/10.1016/j.pbi.2007.07.003

**Yang C. H., Y. C. Huang, C. Y. Chen, and C. Y. Wen**, 2010 Heterologous expression of Thermobifida fusca thermostable alpha-amylase in Yarrowia lipolytica and its application in boiling stable resistant sago starch preparation. J. Ind. Microbiol. Biotechnol. 37: 953–960. https://doi.org/10.1007/s10295-010-0745-2

**Yeoh C. C., M. Balcerowicz, L. Zhang, M. Jaudal, L. Brocard, P. Ratet, and J. Putterill**, 2013 Fine Mapping Links the FTa1 Flowering Time Regulator to the Dominant Spring1 Locus in Medicago, (M. A. Blazquez, Ed.). PLoS One 8: e53467. https://doi.org/10.1371/journal.pone.0053467

**Yue L., Z. Chi, L. Wang, J. Liu, C. Madzak, J. Li, and X. Wang**, 2008 Construction of a new plasmid for surface display on cells of Yarrowia lipolytica. J. Microbiol. Methods 72: 116–123. https://doi.org/10.1016/j.mimet.2007.11.011

**Yuegao H., and D. Cash**, 2009 *Global Status and Development Trends of Alfalfa*. Rome.

**Zhu Y., C. C. Sheaffer, M. P. Russelle, and C. P. Vance**, 1998 Dry Matter

Accumulation and Dinitrogen Fixation of Annual Medicago Species. Agron. J. 90: 103. https://doi.org/10.2134/agronj1998.00021962009000010019x

**Zhuo C., T. Wang, S. Lu, Y. Zhao, X. Li, and Z. Guo**, 2013 A cold responsive galactinol synthase gene from Medicago falcata (MfGolS1) is induced by myo-inositol and confers multiple tolerances to abiotic stresses. Physiol. Plant. 149: 67–78. https://doi.org/10.1111/ppl.12019

**Zuther E., K. Büchel, M. Hundertmark, M. Stitt, D. K. Hincha, and A. G. Heyer**, 2004 The role of raffinose in the cold acclimation response of Arabidopsis thaliana. FEBS Lett. 576: 169–173. https://doi.org/10.1016/j.febslet.2004.09.006