Roles of a Glutamyl-tRNA Synthetase in Controlling Early Anther Development in Rice

By

Xiujuan Yang

Thesis by a combination of conventional and publication formats submitted to

The University of Adelaide for the degree of Doctor of Philosophy

Shanghai Jiao Tong University–University of Adelaide Joint Centre for Agriculture and Health

School of Agriculture, Food and Wine

The University of Adelaide

August 2018



Table of Contents

ABSTRACT ..... **I**

DECLARATION ..... **V**

ACKNOWLEDGEMENTS ..... **VI**

CHAPTER 1 INTRODUCTION **7**

1.1 Anther development and pollen formation in rice 7

1.2 Regulation of early anther development 9

1.2.1 LRR-RLK pathway determines anther cell fate 10

1.2.2 Redox status specifies germinal cells and regulates somatic layers in anthers 13

1.2.3 Regulatory network for early anther development 15

STATEMENT OF THE CONTRIBUTIONS OF JOINTLY AUTHORED MANUSCRIPT **19**

CHAPTER 2

Genome-wide Identification and Characterisation of Aminoacyl-tRNA Synthetases in Model Plants **20**

Abstract 21

Introduction 22

Results 24

Identification of 141 AaRSs in Three Model Plants 24

Diverse Origins of AaRS Family 26

Conserved Motifs in Catalytic Domains 26

New Domains in Plant AaRSs 28

The Tertiary Structures of Plant AaRSs 31

The Expression Patterns of Plant AaRSs during Development 32

Discussion 33

Materials and Methods 35

References 37

Figures and Legends 47

STATEMENT OF THE CONTRIBUTIONS OF JOINTLY AUTHORED MANUSCRIPT 60

CHAPTER 3

A Rice Glutamyl-tRNA Synthetase Modulates Early Anther Cell Division and Patterning 62

Supplemental Information 80

CHAPTER 4 TOWARDS THE FUNCTIONAL CHARACTERISATION OF BARLEY GLUTAMYL-TRNA SYNTHETASES 94

4.1 Abstract 94

4.2 Introduction 94

4.3 Results and Discussion 96

4.3.1 Characterisation of barley *Glutamyl-tRNA Synthetase* genes 96

4.3.2 Protein modelling of rice and barley GluRSs 98

4.3.3 Expression pattern of *HvERS* genes 100

4.3.4 *cis*-Acting regulatory elements in *GluRS* genes' promoters 102

4.3.5 Creating knockout mutants for *HvERS1* and *HvERS2* 104

4.4 Materials and Methods 107

4.4.1 Sequence search and phylogenetic analysis 107

4.4.2 RT-qPCR 107

4.4.3 Protein modelling 107

4.4.4 CRISPR-Cas9-HvERS-sgRNA constructions 108

4.4.5 Barley transformation and identification of transgenic knockout lines 109

CHAPTER 5 SUMMARY AND FUTURE DIRECTIONS 112

5.1 Summary of outcomes 112

5.1.1 Identification and characterisation of plant aminoacyl-tRNA synthetases 112

5.1.2 Roles of a glutamyl-tRNA synthetase in rice early anther development 113

5.1.3 Characterisation of glutamyl-tRNA synthetases in barley 113

5.2 Future directions 114

5.2.1 Exploration of plant multi-synthetase complex 114

5.2.2 Structural basis of plant aminoacyl-tRNA synthetases 115

5.2.3 Mechanisms of GluRS in plant development and signalling 115

ABBREVIATIONS 117

APPENDICES 118

Appendix 1 118

Appendix 2 123

Appendix 3 126

Appendix 4 131

BIBLIOGRAPHY 132

ABSTRACT

In seed-propagated plants, the formation and development of the male reproductive organ, the anther, is indispensable for plant propagation. In grain crops, crop productivity is highly dependent on male fertility and thereafter successful fertilisation between male and female gametophytes. Understanding the molecular mechanisms determining anther development and formation of pollen grains in important crops such as rice (*Oryza sativa*) and barley (*Hordeum vulgare*) is important for fundamental biology and agricultural practice. In flowering plants, the male development starts from the formation of anther primordia, and subsequent cell division and differentiation establish the appropriate anther cell organisation. After morphogenesis, the anther in higher plants mostly contains centrally localised germinal cells surrounded by four layers of anther wall composed by the epidermis, endothecium, middle layer and tapetum, from the outer to the inner. The mature pollen grain is formed via meiosis and mitosis supported by the degeneration of anther wall layers such as the innermost tissue tapetum as well as the middle layer. Although previous investigations uncovered that various regulators such as leucine-rich repeat receptor-like kinases (LRR-RLKs), glutaredoxins, transcription factors, hormones and small RNAs are involved in anther cell fate specification, the role of genes responsible for physiological homeostasis in early male development remains unknown. Plant aminoacyl-tRNA synthetases (aaRSs) are enzymes that catalyse the attachment of amino acids onto their cognate transfer RNAs (tRNAs), playing a central role during the translation of genetic information from messenger RNA to protein. However, little is known about their molecular characteristics and roles in plant development. In the research for this thesis bioinformatics analysis was conducted and 141 aaRSs sequences were obtained from *Arabidopsis* (*Arabidopsis thaliana*), rice and *Physcomitrella patens*. In these sequences, beside the conserved motifs such as HIGH, KMSKS, Motif 2, Motif 3 required for the enzymatic functions, additional new domains such as GST_C (glutathione S-transferase C terminus), WHEP and OB (oligonucleotide binding motif) fold were observed. These may function in

tRNA binding or protein-protein interaction. Sequence prediction showed that these plant aaRSs have subcellular localisation in cytosol, mitochondria and chloroplasts, which may be related with evolutionary origin, signal peptide, and intra/extracellular stimulations. Notably, the *aaRS* genes encoding cytosolic ones are more active in reproductive tissues while genes encoding organellar ones are more expressed in vegetative tissues. Together, we provide an informative source for future research and highlight the essential roles of aaRSs in plant.

To understand the function of aaRS in plant male development, one male sterile mutant called *osers1* (*oryza sativa glutamyl-trna synthetase 1*) was isolated from a mutant library and subjected to functional analysis. Although *osers1* shows normal vegetative development, approximate 50% of *osers1* anthers displayed fused lobes, disarranged anther wall layers and increased number of germinal cells. The expression level of mitotic marker genes *H4* and *CDKB2;1* was also shown to be increased in *osers1* anthers, suggesting the uncontrolled cell division during the early anther development in the mutant. These observations suggest that *OsERS1* is required for early anther cell division and organisation. To identify the *OsERS1* gene, map-based cloning was carried out, and an insertion of one nucleotide was discovered in the fourth exon of the gene LOC_Os10g22380, leading to a truncated protein. Complementation experiments using the wild-type genomic *OsERS1* confirmed that LOC_Os10g22380 is *OsERS1*.

OsERS1 encodes a putative glutamyl-tRNA synthetase (GluRS), containing an N terminal GST_C domain and an aminoacylation domain. Subcellular localisation analysis using GFP-OsERS1 fusion showed that OsERS1 localises in the cytosol and mitochondria. Expression analysis using RT-qPCR and *in situ* hybridisation revealed that *OsERS1* is preferentially expressed before meiosis in anther L2-derived (L2-d) cells that undergo active cell division for the formation of anther cell layers, which is consistent with the defects of *osers1* anthers. To biochemically characterise OsERS1, *in vitro* enzyme activity analysis was conducted. The results demonstrated that OsERS1 is able to catalyse glutamyl-

tRNA synthesis using ATP-pyrophosphate (PPi) exchange and glutamylation assays. It was verified that Arg209, Glu213, His219, His222, Ser442, and Lys443 amino acid residues from three key motifs RFAPE, HIGH, LLSKR are critical for OsERS1's function. Further, the GST_C domain mediated interaction between OsERS1 and its putative cofactor rice Aminoacyl-tRNA Synthetase Cofactor (OsARC) was verified by yeast-2-hybrid (Y2H), pull-down and bimolecular fluorescent complementation (BiFC). The biological significance is that the protein complex has a higher enzymatic activity compared with that of OsERS1 alone, or truncated OsERS1 without interacting domain.

To investigate how OsERS1 affects the early anther development, non-targeted metabolomics analysis was conducted for wild-type and *osers1* anthers. A total of 297 metabolites with known structures were identified, including amino acids and their derivatives, carbohydrates, lipids, cofactors, and nucleotides and secondary metabolites. These metabolites cover most of the central metabolic pathways. Notably, the abundance of glutamate and its derivatives, intermediates of TCA cycle and glycolysis chemicals differed between *osers1* and the wild type. Expression changes of metabolic genes were consistent with the metabolites changes.

An increased level of hydrogen peroxide (H_2O_2) was observed in the mutant anthers, indicating that the defects of *osers1* could be caused by the altered redox status. To test this hypothesis, H_2O_2 was injected into wild-type anthers and it was observed that the treated plants showed disarranged anther cell wall layers and L2-d cells over-proliferation, mimicking that of *osers1* anthers. Collectively, these findings provide evidence of aaRSs-mediated metabolism modulating the development of the male organ in plants, likely through affecting metabolic homeostasis and redox status.

To investigate whether the function of GluRS is conserved in cereal crops, three homologs of *OsERS1* were identified in barley: *HvERS1*, *HvERS2* and *HvERS3*. HvERS1 and HvERS2 have high similarity with OsERS1 in catalytic motifs, preferential expression in reproductive tissues, and putative

tertiary structure. To verify the possible roles of *HvERS1* and *HvERS2* in barley anther development, CRISPR/Cas9 vectors were designed and transformed into a barley cultivar with the objective of creating the knockout mutants by targeted gene editing. Although creating knockout mutants came to a failure, we contributed useful experiences for future gene editing applications in barley.

In summary, *OsERS1* was shown to have roles in modulating metabolism, redox status therefore anther cell behaviour and organ development. With *OsERS1* as an entry point, the aminoacyl-tRNA synthetase family was investigated and three barley homologs were discovered. This research add new understanding of how physiological homeostasis affects early anther formation and extended non-canonical roles of housekeeping proteins.

DECLARATION

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I acknowledge that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I give consent to this copy of my thesis when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Xiujuan Yang	26 April 2018
	Date

ACKNOWLEDGEMENTS

I wish to acknowledge my supervisors Professor Dabing Zhang and Professor Diane Mather. I would like to thank Prof. Zhang for his valuable suggestions and support throughout my PhD candidature. Your guidance and trust have helped transform me into a confident and independent researcher. To my co-supervisor Prof. Mather, thank you for your kind support throughout my study, especially in planning and organising my research work. I also sincerely thank my PhD coordinator, Associate Professor Kenneth Chalmers for your encouragement, comfort, and support throughout my PhD.

I would like to thank the members of Prof. Zhang's group. Special thanks go to Dr. Gang Li, a wonderful colleague, for your invaluable discussion about my research and manuscript writing. To Yu Song, for your help in cytological experiments. To Hendrik Kuijer, for your help in barley transformation. To Deborah Devis and Cindy Callens, for the helpful and relaxing chatting time with you. I also would like to appreciate colleagues in Plant Genomics Centre, Jan Nield, Mellissa Pickering, Alison Hay, Anzu Okada, Patricia Warner, for your great support in lab management. My participation in the Lab Managers Meeting is a truly enjoyable memory. I'm also grateful for the help from my independent advisor Professor Stephen Tyerman, Dr. Gwenda Mayo and Associate Professor Matthew Tucker.

I would like to appreciate colleagues in Shanghai Jiao Tong University including Professor Wanqi Liang, Mingjiao Chen, Zhijing Luo, Lu Zhu, Feiyang Xue and Ruifeng Fu.

To my lovely friends, Hui Zhou, Yuan Li, Yichao Huang, Weixiu Zhang, Yong Jia and Huihui Bi in Waite Campus, thank you for sharing laughter and tears and making my PhD study colourful.

Finally, to my dearest family, my parents, my brother, my beloved husband and my son. You have always been my solid source of support and my strong motivation to improve myself.

CHAPTER 1 INTRODUCTION

Rice (*Oryza sativa* L.) is one of the most widely consumed staple foods by a large proportion of world population, especially in Asia. During last few decades, the yield production of rice has been significantly increased because of the application of hybrid rice technology. To use heterosis in self-pollinating plants, manipulating male fertility is a critical step for producing hybrid seeds. Therefore, development of male sterile lines enables a low-cost, high-efficiency approach for seed production. Understanding the molecular mechanisms underpinning the male reproductive organ and identifying key regulators of male fertility are significant for producing suitable male sterile lines. With a relatively small genome and good chromosomal collinearity with other *Poaceae* species, rice has become an important model crop plant and facilitates research in other cereal crops such as maize (*Zea mays*), wheat (*Triticum aestivum*) and barley (*Hordeum vulgare*).

1.1 Anther development and pollen formation in rice

In rice, the male reproductive organ, called the stamen, is composed of the anther producing male germinal cells, and the filament connecting the anther to the flower. The anther has four connected lobes supported by vascular tissues. Each lobe has a locule and four-layer anther wall (from outer to inner: epidermis, endothecium, middle layer, and tapetum). The microspores, which are generated within the locule, develop into mature pollen grains (Zhang et al., 2011).

The development of a rice anther starts with formation of an oval anther primordium, which consists of three meristematic cell layers: L1, L2 and L3. The anther lobes develop at the four ‘corners’ of the anther primordium. The L1 layer forms the epidermis, which protects inner cells from various stresses. The L3 layer divides and differentiates into the connective structure and vascular tissue. The L2 layer undergoes the most complicated differentiation processes to form germ cells in the centre of each anther lobe and three surrounding supportive somatic cell layers: the endothecium, the middle layer and the tapetum. The whole process of male reproductive development in rice can be divided into 14

stages, based on significant changes and events (Zhang et al., 2011). A large number of regulators have been identified as components of this complicated process (Zhang and Liang, 2016; Shi et al., 2015).

Stages 1−6 are pre-meiotic and involve the cell fate specification of pollen mother cells (PMCs) and anther wall cells via cell division and differentiation from the anther primordium.

Stage 1: an egg-shaped anther primordium emerges from floral meristem showing three layer of cells, L1, L2 and L3.

Stage 2: L1 layer and L1-derived (L1-d) cells form the epidermis after anticlinal division; with rapid division of L2-derived (L2-d) cells, the primordium becomes quadrangular and enlarged presumptive archesporial cells appear at the four corners; L3 layer forms connective cells and vascular tissues.

Stage 3: the central archesporial cells display distinctive enlarged shape surrounded by the neighboring primary parietal cells (PPCs).

Stage 4: PPCs form the endothecium and secondary parietal cells (SPCs) via asymmetric periclinal cell division and archesporial cells produce sporogenous cells. And the whole anther develops four clearly individual lobes.

Stage 5: the sporogenous cells continue the division to prepare for meiosis and the SPC layer symmetrically divides and forms the middle layer and tapetum. At this stage, the anther wall layers have been fully developed.

Stage 6: the sporogenous cells further develop into PMCs (or microspore mother cells).

Stages 7−10 are the development of microspores including a series of events, such as meiosis of PMCs, cell death of anther wall layers, and formation of pollen wall.

Stage 7: PMCs initiate meiosis and start to secrete callose. Tapetal cells accumulate high-density cytosolic materials and middle layer becomes thinner.

Stage 8: each PMC produces four haploid microspores enclosed in one tetrad after meiotic division. The tapetum becomes dark-stained and initiates programmed cell death (PCD) characterised by

nuclear DNA fragmentation and subcellular degradation.

Stages 9−10: the microspores are released from tetrads, and pollen wall materials are accumulated on the surface of microspores to form two-layer wall structure. The middle layer becomes degenerated, and the tapetum continues the degradation, and generates Ubisch bodies.

Stage 11: vacuolated microspores undergo the first asymmetric mitotic division to generate a small generative cell and a large vegetative cell. The tapetum becomes nearly invisible, only leaving Ubisch bodies and debris observed within the locule.

Stage 12: the generative cell undergoes the second round of mitosis to form two germ cells, making the mature pollen grain containing three nuclei. At this stage, anther wall layers only contain epidermis and endothecium.

Stages 13−14: mature pollen grains are released as the elongation of filament, unlocking of palea and lemma, and dehiscence of anther wall.

In conclusion, each event within these stages is essential for normal anther formation, pollen development and releasing, and final fertilisation.

1.2 Regulation of early anther development

Within the meristematic tissue anther primordium, cell division and differentiation are under strict control to guarantee patterning and function of the anther. The L2-d cells firstly form centrally localised archesporial cells and surrounding PPC layer at stage 3. After two rounds of periclinal cell division, the PPC layer generates the endothecium, the middle layer and the tapetum to establish anther wall layers pattern together with L1-derived epidermis at stage 4−5. In *Arabidopsis*, one commonly accepted view on anther wall development concludes that the formation of endothecium, middle layer and tapetum are the consequences of two rounds of asymmetric periclinal cell division. The cell layer surrounding sporogenous cells undergoes the first round of asymmetric periclinal cell division to form the outer endothecium and inner layer, which further generates outwards the middle layer and inwards

the tapetum after the second round of asymmetric periclinal division (Ma, 2005; Feng et al., 2013). In monocots, similarly, PPCs divide into endothecium and SPCs after the first round of asymmetric division. However, Kelliher and Walbot (2011) reconstructed the anther cell division pattern in maize using 5-ethynyl-2-deoxyuridine (EdU, S-phase label) labeling assay, 3D observation and statistical analysis. They raised that the second round of periclinal cell division is symmetric, and cell fate determinant signals may come from neighbouring tissue for further differentiations (Fig. 1.1) (Kelliher and Walbot, 2011). In last few decades, research in *Arabidopsis*, rice and maize has identified several pivotal regulators of early anther development, including leucine-rich repeat receptor-like kinases (LRR-RLKs) pathway, redox state determination, transcription factors etc. (Walbot and Egger, 2016).

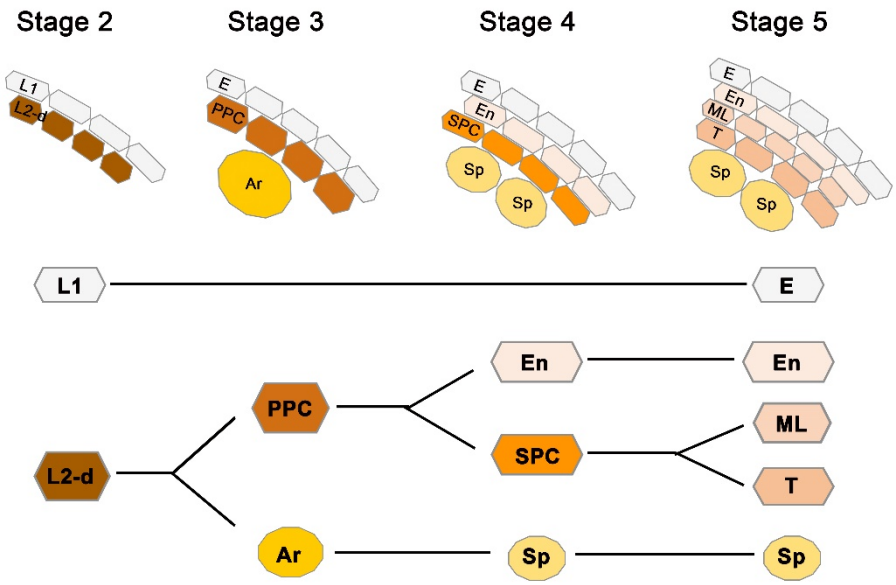


Figure 1.1 Diagrams of anther cell wall pattern formation in high plants

Cell lineage model of early anther development. L1 develops into epidermis. L2-d cells generate somatic endothecium, middle layer, tapetum and germinal sporogenous cells. L2-d, L2-derived cell; PPC, primary parietal cell; SPC, secondary parietal cell; T, tapetum; ML, middle layer; EN, endothecium; E, epidermis; Ar, archesporial cell; Sp, sporogenous cell.

1.2.1 LRR-RLK pathway determines anther cell fate

A typical plant LRR-RLK contains an extra-cytoplasmic LRR motif with the consensus sequence and conserved leucine signature, transmembrane domain and cytoplasmic protein kinase domain. In *Arabidopsis*, over 200 members of the LRR-RLK family participate in plant hormone signalling, defense response, meristem development, and many other biological events (Torii, 2004; Shiu et al., 2004).

Research in *Arabidopsis* has identified a series of LRR-RLKs determining early anther cell fate. *BARELY ANY MERISTEM 1 (BAM1)* and *BAM2* encode CLAVATA1-like LRR-RLKs (Hord et al., 2006). Double mutant *bam1 bam2* shows normal emergence of archesporial cells, but fails to form all three somatic cell layers. Instead, all cells within *bam1 bam2* anther lobes become archesporial-like cells, which are able to initiate meiosis but are degraded later due to lack of support from somatic cells. In early-stage anthers, *BAM1* and *BAM2* are redundantly expressed in the innermost positions of L2-derived cells group: archesporial cells at stage 2, sporogenous cells at stage 3 and 4, PMCs and weakly in tapetum at stage 5. BAM1 and BAM2 may provide cell-to-cell communication to promote the cells surrounding germinal cells to develop into somatic cell layers (Hord et al., 2006). In addition, EXCESS MICROSPOROCYTES 1 (EMS1, alternatively called EXTRA SPOROGENOUS CELLS [EXS]), a cellular surface-localised LRR-RLK, binds to its ligand TAPETAL DETERMINANT 1 (TPD1), and specifies anther cell fate layer (Zhao et al., 2002; Canales et al., 2002). Loss of function of *EMS1/EXS* results in extra sporogenous cells and absence of tapetum, while the mutant has normal endothecium and middle layer. *EMS1* is expressed in L2-d cells from appearance of archesporial cells till meiosis. At stage 5, *EMS1* shows stronger expression in tapetum than in sporogenous cells, indicating its association with determining tapetal cell fate and controlling germinal cell number (Zhao et al., 2002). *TPD1* encodes an unknown small protein with 176 amino acids and the *tpd1* mutant displays a similar phenotype to *ems1/exs* mutant: increased number of sporogenous cells and lack of tapetum (Yang et al., 2003). *In situ* hybridisation revealed that the expression of *TPD1* overlaps with *EMS1* during stage

2–4. At stage 5, when expression of *EMS1* is increased in tapetum and weakened in sporogenous cells, *TPD1* is predominantly expressed in sporogenous cells and weakly in tapetum, suggesting that *TPD1* and *EMS1* have coordinated expression during development of tapetum and sporogenous cells. Further, in the *tpd1* mutant, the expression of *EMS1* is not affected (Yang et al., 2003). TPD1 is able to interact with EMS1 *in vitro* and *in vivo*, and induce the phosphorylation of EMS1 (Jia et al., 2008), demonstrating that their coordination is at physical protein level as a receptor-like-kinase-ligand pair. Two more LRR-RLKs SOMATIC EMBRYOGENESIS RECEPTOR LIKE KINASE 1 (SERK1) and SERK2 were also shown to be involved in early anther development (Colcombet et al., 2005; Albrecht et al., 2005). The *serk1 serk2* double mutant is phenotypically similar to *ems1/exs* and *tpd1* and the expression of *SERK1* and *SERK2* is seen in both germinal cells and somatic cell layers during stage 4–5 and later in the tapetum and middle layer. However, it remains to be determined whether the SERK1 and SERK2 proteins act within the EMS1/TPD1 pathway or a separate signal pathway.

In the model monocot rice, MULTIPLE SPOROCYTE 1 (MSP1) and OsTDL1A (TPD1-like 1A)/MICROSPORELESS 2 (MIL2) have been identified as the orthologs of EMS1/EXS and TPD1, respectively (Nonomura et al., 2003; Zhao et al., 2008; Hong et al., 2012a; Yang et al., 2016). Compared with *ems1/exs, tpd1* and *ems1 tpd1*, which can form endothecium and middle layer without tapetum, rice *msp1, mil2*/*ostdl1a* merely generate the endothecium, lacking the middle layer and tapetum, and the inner space of anther lobe is occupied by extra sporogenous cells. These phenotypic data indicate that the MSP1-OsTDL1A/MIL2 pathway in rice anther starts the regulatory roles earlier than EMS1/EXS-TPD1 pathway in *Arabidopsis* anther. Further, the expression pattern of *MSP1* and *MIL2* overlaps in somatic cell layers–PPC at stage 3, endothecium and SPC at stage 4, middle layer and tapetum at stage 5. However, *MIL2* is expressed in archesporial cells and later in sporogenous cells, *MSP1* is only strongly expressed in archesporial cells at stage 3 (Hong et al., 2012a). Additionally, Yang et al. (2016) confirmed the protein-protein interaction between MSP1 and OsTDL1A *in vitro* and

in vivo. These results suggest that the MSP1-OsTDL1A/MIL2 complex likely regulates the development of PPC to SPC and limits proliferation of sporogenous cells.

In maize, however, only the ortholog of TPD1 and OsTDL1A, MULTIPLE ARCHESPORIAL CELLS 1 (MAC1) has been reported. *MAC1* encodes a small secretory protein, which is predominantly accumulated in germinal cells (archesporial cells and later sporogenous cells) with a decreased radial pattern in surrounding somatic cell layers, differing from those of *TPD1* and *OsTDL1A*. The *mac1* anther contains only one integral somatic layer and significantly excess archesporial cells (Wang et al., 2012). These data indicate that MAC1 may mainly be implicated in formation of SPCs and limitation of archesporial cells proliferation, similar to that in rice. Although the maize homolog of MSP1 remains to be determined, a few LRR-RLK candidates have been found to be involved in maize anther development (Ma et al., 2007). Collectively, this LRR-RLK-ligand signalling pathway has conserved components of the complex among *Arabidopsis*, rice and maize, but those components show different expression pattern, temporal regulation and anther wall cell fate.

1.2.2 Redox status specifies germinal cells and regulates somatic layers in anthers

Different from most animals, who reserve germinal stem cells as early as embryogenesis (Juliano and Wessel, 2010), higher plants have no germinal cells until the growth phase transition from vegetative to reproductive (Ma, 2005). The cellular redox status is considered as an essential factor in germinal cell fate acquisition (Kelliher and Walbot, 2012). Glutaredoxins (GRXs) are important modulators of cellular redox status. GRXs are small redox enzymes of approximately one hundred amino-acid residues that use glutathione as a cofactor (Fernandes and Holmgren, 2004). GRXs are oxidised by substrates, and reduced non-enzymatically by glutathione. Oxidised glutathione is then regenerated by glutathione reductase. Together, these components comprise the glutathione system.

The plant GRXs are classified in three groups based on amino acid sequences of redox centres: CGPC type, CGSF type and CC (CCxC/S) type (Meyer et al., 2009). *Arabidopsis* ROXY1 and ROXY2

are CC type GRXs, redundantly control early anther development (Xing and Zachgo, 2008). The double mutant *roxy1 roxy2* fails to generate any germinal cells or somatic cell layers in the two adaxial-side lobes, leaving two groups of unidentified somatic cells. In the abaxial-side lobes, PMCs and four anther wall layers can develop, but germinal cells are in clusters and degenerate without further development. *ROXY1* and *ROXY2* are expressed in a similar pattern: in all L2-d germinal and somatic cells during stage 2–4 and confined to PMCs and tapetum before meiosis. The glycine^{110} residue of the putative glutathione binding site is critical for ROXY1 function, indicating that ROXY1 regulates anther lobe development by modulating redox status (Xing and Zachgo, 2008).

Rice MICROSPORELESS1 (MIL1) is also a CC type GRX, acting as a switch of meiotic fate of male germ cells (Hong et al., 2012b). In *mil1* anthers, after stage 3, archesporial cells divide into supernumerary sporogenous cell progenies which cannot enter into meiosis, while the PPC layer generates disarranged somatic layers without clear cell properties. At the final stage, all cells within the anther lobe develop into tightly packed somatic cells. *MIL1* is initially expressed at four anther corners where archesporial cells localised, later in parietal cells and sporogenous cells and restricted in the tapetum when meiosis starts (Hong et al., 2012b). These discoveries suggest that MIL1-mediated maintenance of reductive status in L2-d cells is critical for germinal cell fate acquisition and somatic cell layer differentiation.

Maize GRX mutant *male sterile converted anther1 (msca1)* displays a phenotype similar to that of *mil1* anthers: failure to acquire germ cells and final formation of somatic anther lobes (Chaubal et al., 2003). Considering that anther primordium development is accomplished within a naturally formed hypoxia niche, MSCA1 is likely to mediate processes related to redox status. Kelliher and Walbot (2012) induced maize *msca1* to restart archesporial cell specification in reductive conditions by repressing reactive oxygen species (ROS) level. Interestingly, decreasing O_2 or H_2O_2 concentration caused an increase of germinal cell numbers, and stimulated the formation of ectopic germinal cells in

epidermal and subepidermal layers. By contrast, increasing O_2 resulted in distortion of localisation of archesporial cells, which moved to deeper vascular tissue (Kelliher and Walbot, 2012). These data indicate that the maize anther prefers hypoxic or reductive conditions to specify germinal cells; early anther cells have the capacity to differentiate into somatic or germinal cells, depending on environmental cues; hypoxia in the central position of the anther lobe triggers MSCA1-mediated events to specify archesporial cells, which further secrete MAC1 proteins to limit self-proliferation and provide signals to organise somatic cell layers.

Comparing the mutant phenotypes and expression patterns of *Arabidopsis* ROXY1/ROXY2, rice MIL1 and maize MSCA1, and considering the large number of GRX family members, we speculate GRXs play essential and diverse roles in both early somatic and germinal cell lineages. In the naturally arisen hypoxia niche of anthers, GRX may be implicated in early development in two ways: as a component of the glutathione system to maintain reductive status of early anther cells and/or as a transductor of hypoxic signals to modify target proteins.

1.2.3 Regulatory network for early anther development

There are various regulators beyond the LRR-RLK-ligand pathway and GRXs involved in early anther development (Table 1.1), and many of the regulators have been found to be connected and organised into a network (Fig. 1.2).

Table 1.1 List of regulators of early anther development in *Arabidopsis*, Rice and Maize

Stages	*Arabidopsis*	Rice	Maize	Annotation	Events	Ref(s)
1	AG	MADS3, MADS58		MADS-box transcription factor	Stamen identification	Ito et al., 2004; Yamaguchi et al., 2006
2	BAM1, BAM2			LRR-RLK	Ar cells identification and somatic cells differentiation	Hord et al., 2006
		MIL1	MSCA1	GRX	Ar cells differentiation and somatic cells organisation	Hong et al., 2012b; Chaubal et al., 2003; Kelliher and Walbot, 2012
2	ROXY1, ROXY2			GRX	Ar cells specification (adaxial)	Xing and Zachgo, 2008
	TGA9, TGA10			bZIP transcription factor		Murmu et al., 2010
3	SPL/NZZ			MADS-box transcription factor	Ar cells differentiation, somatic cell layers	Schiefthaler et al., 1999; Yang et al., 1999; Liu et al.,

					development	2009
3 (monocot) 5 (dicot)	TPD1	TDL1A/ MIL2	MAC1	Small secreted protein ligand	Somatic cell layers specification, limiting Ar cells proliferation	Jia et al., 2008; Zhao et al., 2008; Hong et al., 2012a; Wang et al., 2012; Yang et al., 2016
	EMS1/EXS	MSP1		LRR-RLK		Canales et al., 2002; Zhao et al., 2012; Nonomura et al., 2003
4			OCL4	HD-ZIP IV transcription factor	Periclinal division of subepidermal layers	Vernoud et al., 2009
5	SERK1, SERK2			LRR-RLK	SPC periclinal division	Albrecht et al., 2005; Colcombet et al., 2005
		TIP2		bHLH transcription factor	Somatic cell layers differentiation after periclinal division	Fu et al., 2014

In *Arabidopsis*, BAM1 and BAM2, two LRR-RLKs, define the most inner small group of L2-d cells as archesporial cells. SPOROCYTELESS/NOZZLE (SPL/NZZ), a MADS-box transcription factor, whose spatial expression is limited by BAM1/BAM2, supports archesporial cells differentiation and somatic cell layers development (Schiefthaler et al., 1999; Yang et al., 1999; Hord et al., 2006). At the same time, *SPL* is under the regulation of homeotic protein AGAMOUS (AG) and ectopic expression of *SPL/NZZ* prompts transformation from petal to stamen (Ito et al., 2004; Liu et al., 2009), indicating the role of *SPL/NZZ* in stamen identification. Also, plant hormone brassinosteroid (BR) is involved in the regulation of *SPL.* BR signalling component transcription factor BES1 (BRI1-EMS-suppressor 1) can directly bind to promoter region of *SPL* (Ye et al., 2010). As the downstream regulators of *BAM1/BAM2* and *SPL/NZZ*, two basic leucine-zipper (bZIP) transcription factor genes *TGA9* and *TGA10* have a high overlapping expression with *ROXY1* and *ROXY2*, and TGA9/TGA10 proteins can physically interact with ROXY1/ROXY2, suggesting that these bZIP transcription factors are part of GRX-triggered biological events (Murmu et al., 2010).

In the rice *ostdl1a msp1* double mutant, a defect in the LRR-RLK-ligand (MSP1-OsTDL1A) pathway has diverse direct and indirect impacts on early anther development, according to genome-wide expression profiling (Yang et al., 2016). Firstly, a TGA-GRX pair is identified as downstream of MSP1-OsTDL1A signalling. The expression levels of *OsTGA10* and *OsGrx_I1* are both down-regulated in *ostdl1a msp1*. *OsTGA10* is expressed in L2-d cells at stage 3–6, and OsTGA10 interacts with OsGrx_I1 in yeast, suggesting the involvement of this TGA-GRX pair in redox status and gene

expression during early rice anther development. Secondly, altered expression of large group of genes encoding basic helix-loop-helix (bHLH) and bZIP transcription factors indicates profound roles of MSP1-OsTDL1A pathway. Among these genes, *OsbHLH142 (TIP2, TDR INTERACTING PROTEIN2)* is significantly down-regulated and this gene defines the developmental fate of three somatic cell layers (Fu et al., 2014). In addition, epigenetic regulators are also involved in MSP1-OsTDL1A pathway. In *msp1* and *ostdl1a* anthers, Fei et al. (2016) found great change in miRNAs, phased secondary siRNAs (phasiRNAs), and PHAS transcripts, especially the absence of miR2275. Accumulation of miR2275 in anthers can prompt the production of phasiRNAs. As a consequence, *msp1* and *ostdl1a* mutants fail to produce precursor *PHAS* mRNA and phasiRNAs. Co-expression network analysis added three more *Argonaute* genes into this process: *OsAGO1d*, *OsAGO2b*, and *OsAGO18* (Fei et al., 2016). These studies suggest the dominant position of MSP1-OsTDL1A pathway, leading to transcriptional and post-transcriptional control of early anther development. Additionally, Kelliher and Walbot (2012) proposed that the naturally formed hypoxia niche of anther lobes contributes to the germinal cell specification. The hypoxic or reductive condition seems to be a very early signal to trigger MSCA1-mediated meiotic fate acquisition.

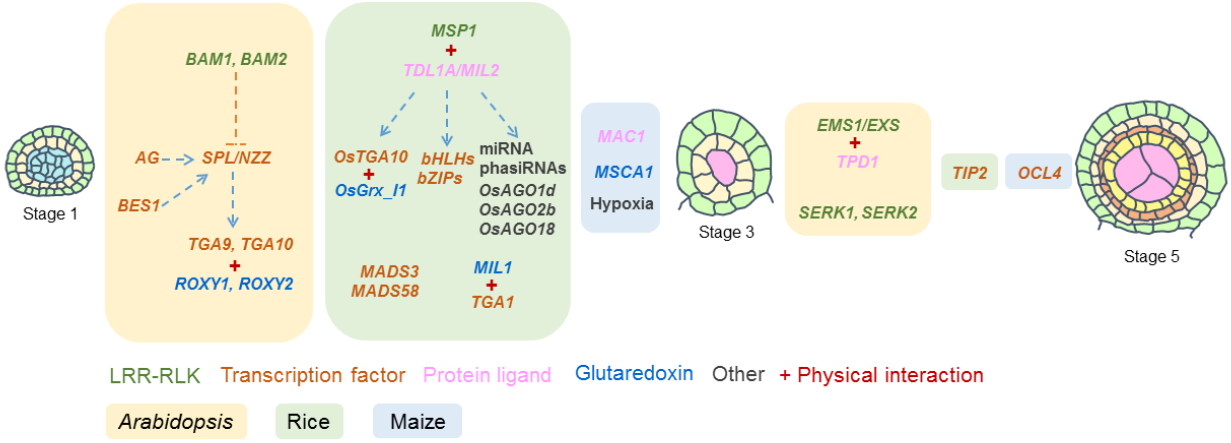


Figure 1.2 Regulatory network in early anther development

The associations among regulators during early anther development in *Arabidopsis*, rice and maize. LRR-RLK-ligand pairs,

transcription factors, redox status, epigenetic control, phytohormone, and environmental cues are involved in multiple events including cell division, cell patterning and cell differentiation.

As shown in Figure 1.2, during early anther primordium development, environmental cues, cellular physiological status, kinase-ligand signalling, transcription control are all intertwined to form a network that guarantees sufficient preparation for meiosis. Significantly, the rice MSP1-TDL1A pair occupies a dominant position in the regulation of early anther development by exerting effects on transcription factors, redox status, and production and editing of non-coding RNA molecules. Many mutants described here show disruptions in the formation of the cell layers of anther walls, indicating that the anticlinal or periclinal cell division in anthers is under strict and elaborate control. Still, there are gaps between these known regulators, for examples, how hypoxia triggers *MSCA1* and what components mediate the communications between adjacent cells. Other issues in early anther development also need to be addressed. The lack of cell type marker genes restricts understanding of anther cell differentiation. Transportation and metabolism of nutrients in anther cells have not been studied much. Multiple omics studies for specific cell types and developmental stages are required to be further explored.

STATEMENT OF THE CONTRIBUTIONS OF JOINTLY AUTHORED MANUSCRIPT

Statement of Authorship

Title of Paper	Genome-wide Identification and Characterisation of Aminoacyl-tRNA Synthetases in Model Plants
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ■ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	Prepared for submission to *Journal of Experimental Botany*.

Principal Author

Name of Principal Author (Candidate)	Xiujuan Yang		
Contribution to the Paper	X.Y. conceived this study, performed all bioinformatics analysis, organised data and figures, and drafted manuscript.		
Overall percentage (%)	85%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	26/04/2018

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);
ii. permission is granted for the candidate in include the publication in the thesis; and
iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Gang Li		
Contribution to the Paper	G.L. contributed to discussion and editing.		
Signature		Date	26/04/2018

Name of Co-Author	Dabing Zhang		
Contribution to the Paper	D.Z. contributed to conceiving this study and editing. D.Z. acted as corresponding author.		
Signature		Date	26/04/2018

CHAPTER 2

Genome-wide Identification and Characterisation of Aminoacyl-tRNA Synthetases in Model Plants

Authors: Xiujuan Yang[a,b], Gang Li[a,b], Dabing Zhang[a,b]

Affiliations:

[a]School of Agriculture, Food and Wine, University of Adelaide, South Australia 5064, Australia.

[b]Joint International Research Laboratory of Metabolic & Developmental Sciences, Shanghai Jiao Tong University–University of Adelaide Joint Centre for Agriculture and Health, State Key Laboratory of Hybrid Rice, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China.

*Correspondence should be addressed to D.Z.

(E-mail: dabing.zhang@adelaide.edu.au, Tel: +61 08 83137167)

Abstract

Aminoacyl-tRNA synthetases (aaRSs), catalysing the combination of amino acid and cognate transfer RNA (tRNA), are essential for protein synthesis in organisms. In mammals, some aaRSs are shown to form a multi-tRNA synthetase complex and many of them are involved in metabolic regulation, organ development and other non-canonical functions. Plant aaRSs are integral parts of aaRS family, and investigation of mutants of aaRSs indicates their indispensable roles in plant life, however the underlying mechanisms remain largely limited. Here, we identify a collection of 141 aaRS members from three model plants, *Arabidopsis thaliana* (*Arabidopsis*), *Oryza sativa* (rice) and *Physcomitrella patens* (*P. patens*) and analyse their molecular features. The plant *aaRS* genes are predicted to have originated from both eukaryote and organelle ancestors. Phylogenetic analysis indicates gene duplication and loss events within *aaRS* family. We also observed the conserved motifs essential for catalytic function and proposed their evolvement and differentiation from prokaryotes. Functional prediction of the additional domains in plant aaRSs such as WHEP, oligonucleotide binding motif and tRNA binding domain, indicates potential roles beyond translation such as protein-protein interaction and nucleic acid behaviours. Subcellular localisation prediction shows localisation of plant aaRSs in cytosol, and/or mitochondria, and/or chloroplast, suggesting a commonly used dual-targeting strategy. Protein modelling of cytosolic and organellar aaRS shows similar/distinguished structural characteristics, consistent with their common/different origins. Collectively, this work provides an evolutionary and functional insight into plant aaRSs.

Key words: aminoacyl-tRNA synthetase, evolution, protein modelling, subcellular localisation

Introduction

Aminoacyl-tRNA synthetases (aaRSs) comprise a family of ancient enzymes that are ubiquitous in all cells requiring protein synthesis. They catalyse the aminoacylation reaction to combine amino acids and their cognate tRNAs during translation of genetic information from messenger RNA to peptide sequence (Schimmel, 1987). Generally, the biochemical function of aaRSs includes two reactions. Firstly, amino acids are activated by ATP, forming high-energy intermediate aminoacyl adenylates. Then aminoacyl adenylates are transferred to the 2' or 3'- OH of cognate tRNA to form final product (Fig. 1) (Ibba and Söll, 2000). The abilities of aaRSs to recognise specific amino acids and tRNAs ensure the translation fidelity in all species from major kingdoms of life.

The characteristics and evolution of the aaRS family have been investigated in many important microorganisms and animals (Woese et al., 2000). Some surprising facts about aaRSs have been discovered, revealing non-canonical functions in translation control, metabolism, development, angiogenesis, tumorigenesis and immunity (Guo and Schimmel, 2013). In mammals, Glu-Pro-RS, IleRS, LeuRS, MetRS, GlnRS, ArgRS, LysRS and AspRS form a multi-tRNA synthetase complex (MSC) together with three auxiliary factors p43, p38 and p18 (Hausmann and Ibba, 2008). The conjunctive factors p43, p38, p18 are multifunctional proteins involved in dermal tissue regeneration (Park et al., 2005b), lung development (Kim et al., 2003) and DNA damage response (Park et al., 2005a), respectively. It is not clear how these components are released from MSC, nor how the whole MSC functions. One important trait of aaRSs is the amino acid binding pocket, which provides a chance for aaRSs to become a link in amino acid signalling (Arnez and Moras, 1997). Amino acids are not only nutrient materials for cells, but also regulatory signals of development and metabolism (Hoffman et al., 2011). Glutamine can suppress apoptosis by mediating the interaction of GlnRS and apoptosis signal regulating kinase 1 (ASK1), and GlnRS acts as a link in inhibition of apoptosis (Ko et al., 2001). LeuRS plays a pivotal role in activating the mammalian target of rapamycin complex

(mTORC). There is sufficient evidence showing that LeuRS serves as an intracellular leucine concentration sensor by combination with Rag GTPase (Han et al., 2012). TrpRS is a potential anti-angiogenesis factor acting via inhibition of vascular endothelial-cadherin (Tzima et al., 2005). Phosphorylated LysRS is relocated into the nucleus and functions as a gene expression regulator in immune response (Yannay-Cohen et al., 2009). However, the non-translation functions of TrpRS and LysRS are independent of their cognate amino acids, suggesting more mechanisms of amino acid binding sites of aaRSs.

Functional understanding about aaRSs in the plant kingdom is still insufficient. Research conducted over the past two decades has uncovered important roles of aaRSs in various biological procresses, such as organelle biogenesis, development, and immunity responses. In *Nicotiana benthamiana*, silencing *GluRS* and *SerRS* in leaves leads to severe yellowing (Kim et al., 2005) . With transmission electron micrography (TEM) it was determined that silencing of these two aaRSs disrupts chloroplast division and thylakoid membrane development and reduces the numbers of both chloroplasts and mitochondria. In rice, a chloroplast defective mutant *chlorophyll deficient 1 (cde1)* caused by the mutation in a *GluRS* gene, showed lower levels of two types of chlorophyll and abnormal thylakoid membrane development aggravated with increasing temperature (Liu et al., 2007). Rice ValRS has been also demonstrated to be a regulator of early chloroplast development and chloroplastic ribosome biogenesis (Wang et al., 2016). Mutation of *OsValRS2* in rice *white panicle 1* (*wp1*) mutant resulted in albino seedlings and panicles with abnormal chloroplasts. Expressions of genes involved in photosynthesis and protein synthesis in chloroplasts were repressed in *wp1*, indicating the essential roles of OsValRS2 in chloroplast development (Wang et al., 2016). Out of 46 aaRSs of *Arabidopsis*, 21 enzymes were investigated using 50 insertional mutants. Disrupted *ValRS*, *LeuRS*, *ArgRS*, *AsnRS*, *GlyRS*, *TyrRS*, *ThrRS*, and *AlaRS* lead to defective embryogenesis mainly due to disturbed protein translation in chloroplasts. Mutants of *MetRS*, *IleRS*, *GluRS*, *TrpRS*, *LysRS*, *ProRS*, *SerRS*, *AsnRS*,

GlnRS show abortion in the ovule because of disturbed protein translation in mitochondria (Berg et al., 2005). A further study demonstrates that a mitochondrial TyrRS is essential in expansion of the embryo sac and proliferation of endosperm nucleolus (Jiang et al., 2012). An *Arabidopsis CysRS* that is specifically expressed in the central cell of female gametophytes determines the cell fate of adjacent cells (Kägi et al., 2010). Some aaRSs show multi-dimensional regulation in plant development. GlyRS participates in leaf dorso-ventral polarity determination, together with the MYB transcription factor ASYMMETRIC LEAVES1 (AS1). Other abaxial fate genes and auxin signalling response components are also involved in the regulatory network of GlyRS (Moschopoulos et al., 2012). Plant aaRS is also a component of immune system. In *Arabidopsis*, β-aminobutyric acid (BABA) is a broad-spectrum protective molecule against various stresses and diseases (Luna et al., 2014). An AspRS mediates sensing BABA that competitively binds to AspRS when cells are attacked by pathogens and then activate defense reactions. These observations indicate broad prospects for expanding non-canonical functions of plant aaRSs.

In this study, we identified 141 aaRSs in three model plants, and sequence analysis revealed their possible gene duplication and loss events during evolution. Furthermore, we dissected their molecular features including catalytic motifs alignment, predicted additional domain analysis, subcellular localisation and tertiary structures modelling. Their subcellular localisation, and tertiary structure are also analysed. This systematic understanding of plant aaRSs will benefit future functional research.

Results

Identification of 141 AaRSs in Three Model Plants

To fulfill protein synthesis, plants genomes must encode an entire set of aaRSs for all 20 proteinogenic amino acids. *Arabidopsis thaliana* (*Arabidopsis*), *Oryza sativa* (rice) and *Physcomitrella patens* (*P. patens*) are well studied model species representing dicots, monocots and mosses respectively and the deciphering of their genome sequences allows searches for genes with specific features and functions

(*Arabidopsis* Genome Initiative, 2000; Goff et al., 2002; Yu et al., 2002; Rensing et al., 2008). The aaRSs family from *P. patens* is expected to provide evolutionary insights as mosses diverge long ago from vascular plants.

Using BLASTp alignment of the reported aaRSs amino acid sequences as queries, we identified 46, 53, and 42 aaRSs, as well as gene loci encoding aaRSs in *Arabidopsis*, rice and *P. patens*, respectively, from the databases of National Center for Biotechnology Information (NCBI) and Universal Protein Resource (UniProt) (Table 1). In each species, aaRSs are divided into two classes and further into six subclasses, as previously defined (Cusack, 1995). We speculate that the angiosperms are prone to generate more aaRSs than non-flowerings plants possibly to adapt more complex developmental events and environmental stimulates. Gene duplication and loss are common in the *aaRS* family. Close paralogues produced by gene duplication of *CysRS*, *TyrRS*, *AsnRS* and *AspRS* in *Arabidopsis*, have been recognised from previous phylogenetic analysis (Table 1) (Peeters et al., 2000; Duchêne et al., 2005).

Furthermore, gene duplication events may occur after monocot-dicot divergence, leading to multiple-copy *aaRS* genes, such as *GluRS* (Yang et al., 2018, Chapter 3) and *SerRS* (Fig. 2) in rice compared with *Arabidopsis*. Copy numbers of homologs also imply similar duplication events may also occur in rice *MetRS*, *LeuRS*, *ValRS*, and *GlnRS* genes (Table 1). Notably, one TyrRS (F4HWL4), which is a fused enzyme containing two catalytic domains, was possibly produced by gene duplication and fusion in *Arabidopsis* during evolution. Besides duplication, gene loss also contributes to copy variations. *Arabidopsis* seems to have lost one *SerRS* gene copy after divergence from *P. patens* (Fig. 2). Similar loss events are also predicted for *Arabidopsis GluRS* genes. Novel gene production also contributes members to aaRS family. One ThrRS in *Arabidopsis* (Q8GZ45) has no close homolog in rice or *P. patens*, but has homologs among dicots. One AlaRS in rice (B9FWG9) is exclusively found in rice but not in other monocots or in any dicots (Table 1). We thus speculate that the genes encoding

these two proteins are likely novel members generated after species specification. Therefore, sequence analysis of plant aaRSs indicated that aaRS family has both conserved and diversified aspects during evolution.

Diverse Origins of AaRS Family

Based on amino acid sequences comparison, most of aaRSs in the three model plants we investigated were divided into cytosolic and organellar subclasses, suggesting differentiated evolutionary origins (Table 1). According to phylogenetic analysis for GluRSs and SerRSs (Fig. 2) (Yang et al., 2018, Chapter 3), we speculate that most organellar aaRSs are similar to bacterial aaRSs, while cytosolic aaRSs are close to eukaryotic aaRSs. The observation of plant GlyRS is also consistent with this point (Nagel et al., 1984; Toth and Schimmel, 1990; Moschopoulos et al., 2012). Plant GlyRSs are divided into two subclasses: Class IIa GlyRS contains one α subunit and putatively form α_2 dimer, while Class IIc GlyRS contains α and β subunits and may form $\alpha_2\beta_2$ dimer. Intriguingly, GlyRS (IIa) is universal in most eukaryotes such as *Bombyx mori*, human and yeast as well as in several prokaryotes like *T. thermophilus*, whereas GlyRS (IIc) is only found in bacteria like *Escherichia coli* (*E. coli*) (Nagel et al., 1984; Toth and Schimmel, 1990) and in *Arabidopsis* (Moschopoulos et al., 2012). The α and β subunits have homologs in cyanobacteria and chlorophyta, indicating that this prokaryote-originated GlyRS was likely integrated into nuclear genome from ancient chloroplasts after endosymbiosis, and remained in most modern plants. LysRS (Q9LJE2 and its homologs in Table 1) is possibly another example of gene transferring from chloroplasts to nucleus, as it has no homolog in human, yeast or *E.coli*, but in cyanobacteria and chlorophyta. In short, plant *aaRS* gene family may have diverse origins during evolution which may be linked with their localised compartments.

Conserved Motifs in Catalytic Domains

AaRSs are characterised by an aminoacylation domain, which contains the active sites to bind amino

acids, ATP, and tRNAs (Fig. 1). Using plant aaRS protein sequences, we identified the pivotal conserved motifs in Class I and Class II aaRS enzymes (Fig. 3).

In the catalytic region, Class I aaRS contains a typical Rossman fold characterised by HIGH motif and a conserved KMSKS motif (Ibba and Söll, 2000). Within HIGH motif, the first histidine is conserved among the three model plants, while the glycine that is conserved in prokaryote aaRSs has an alternative of alanine in plants (Chaliotis et al., 2016). The KMSKS motif of Class I was less conserved, with several substitutions such as methionine replaced by hydrophobic leucine or phenylalanine (Fig. 3, upper). Unlike mammals (KFKTR in human ArgRS) and prokaryotes (most likely KMSTR) (Chaliotis et al., 2016), plant ArgRS shows an unusual RFRTR sequence. In each aaRS, however, the motif sequences may differ among various organelles and species. For instance, cytosolic GluRSs contained HIGH and LLSKR, while organellar GluRSs showed HVGG and KLSKR (Supplemental Table 1). The variations of key motif sequences within one kind of aaRS suggest varied origins of plant *aaRS* genes.

Class II aaRSs have three conserved motifs: Motif 1 is essential for dimer formation; Motif 2 and Motif 3 are required for the catalytic function (Ibba and Söll, 2000). Consistent with the three subclasses within Class II, Motif 1 sequences were also divided into three classes (Fig. 3, lower). Motif 1(IIa) gxxxΦxxPxΦ (Φ, hydrophobic; capital, absolutely conserved; lowercase, highly conserved; x, any residue), present in Class IIa plant aaRSs, had two hydrophobic sites substituted by other residues compared with prokaryotic motif 1 gΦxxΦxxpΦΦ (Chaliotis et al., 2016). Motif 1(IIb) PΦΦxx, present in Class IIb plant aaRSs, showed more flexibility in residue choices. Of note, the highly conserved proline appears pivotal for Class IIa and IIb aaRSs in dimer interface (Fig. 3, lower). Motif 1(IIa) and motif 1(IIb) may have difference ways to form dimers which remains to be elucidated by further structural evidence (Fig. 3). The Class IIc aaRSs are so called ‘odd ones’, forming heterodimers (Cusack, 1995). Accordingly, the Motif 1(IIc) sequences showed high variation and non-conserved

patterns (Supplemental Table 1). In the catalytic core, Motif 2 fRxe is conserved with prokaryotic Motif 2 rather than Motif 2B, narrowing down the range of possible ancestors (Chaliotis et al., 2016). And Motif 3 gΦgΦgΦd/eRΦΦΦΦ showed highly similar sequences to that of prokaryotes. Comparison of conserved motifs and substitutions of residues may provide insights to trace the evolutionary origins of plant aaRSs.

New Domains in Plant AaRSs

Besides the essential aminoacylation domain, we recognised that plant aaRSs have incorporated a diverse range of additional domains with or without the possible association with their catalytic activities. Based on sequence similarity, we predicted the additional domains of plant aaRSs in three model plants (Supplemental Table 1). Briefly, the additional domains could be located at N terminus or C terminus of the aminoacylation domain, and even unusually within the catalytic module (Fig. 4), indicating diverse recruitment events in plant aaRSs.

Several aaRSs have specific domains such as the ArgRS_ N terminal domain for recognising tRNA, SAD (the second addition domain) to interact with tRNA, tRNA-binding arm, B3/B4 tRNA binding domain with editing sites, and DALR domain to bind the anticodon of tRNA, which may enhance their affinities to substrates (Fig. 4). GST (glutathione S-transferase) and a helix-turn-helix motif called WHEP have been reported to be involved in the assembly of human MSC (Mirande, 2017), which makes GluRS with a GST domain and GlyRS/AsnRS with a WHEP domain potential members of plant MSC. In rice, the interaction between OsERS1 (GluRS) and its cofactor OsARC (Aminoacyl-tRNA Synthetase Cofactor) has unveiled the first two components of plant MSC (Yang et al., 2018, Chapter 3). Some additional domains may provide functions beyond translation. OB fold (oligonucleotide binding motif) commonly exists in single-stranded DNA (ssDNA)-binding, cold shock proteins, DNA ligases, RNA-capping enzymes, DNA replication initiators and RNA polymerase subunit and has also been reported to be involved in protein-protein interaction (Theobald et al., 2003).

Plant AsnRS, AspRS, LysRS, and MetRS containing the OB fold possibly have functions associated with DNA behaviour, cold response or transcription control (Fig. 4). We also observed several special and intriguing domains such as ubiquitin domain in SerRS (Q8RWT8), MINDY deubiquitinase domain in AlaRS (B9FWG9), and a ligand-binding TGS domain in ThrRS, which may extend those aaRSs' functions in signalling pathways (Fig 4). The additions of new domains or motifs of aaRSs usually are progressive as the life tree ascends (Guo et al., 2010), which was also reflected in plant aaRSs. The absence of a WHEP domain in *P. patens* GlyRS and the presence of a WHEP domain in *Arabidopsis* and rice GlyRSs suggests that the addition of WHEP domain to GlyRSs may be associated with plant landing during evolution. Rice AlaRS (B9FWG9), which contains a MINDY domain, shares no homolog with moss or dicots or even other monocots, implying a unique recruitment event in Oryzeae species (Table 1).

Protein synthesis occurs in the plant cytosol, mitochondria and chloroplasts. Each compartment needs a whole set of tRNAs and corresponding aaRSs to fulfill translation. The dual or triple targeting of some aaRSs is common in many organisms and has been verified in *Arabidopsis* (Duchêne et al., 2005, Berglund et al., 2009), *Phaseolus vulgaris* (Duchêne et al., 2001), maize (Rokov-Plavec et al., 2008), diatoms and cryptophytes (Gile et al., 2015).

Based on application of the targeting prediction tool PREDOTAR (Small et al., 2004) and previously reported experimental data, we listed the subcellular distribution of aaRSs of *Arabidopsis*, rice and *P. patens* (Fig. 5; Table 1; Supplemental Table 1). Based on the localisation, we concluded two ways to dispatch aaRSs in plant cells. One is to allocate aaRS enzymes to cytosol and organelles separately. In this case, cytosol has specific copy or copies of a specific type of aaRS, while the other(s) are localised in mitochondria and chloroplasts, such as CysRSs, MetRSs and TrpRSs of *Arabidopsis*, ArgRSs, LeuRSs and ValRSs of rice, and MetRSs and TyrRSs of *P. patens* (Fig. 5). The second way is to share the aaRS enzymes between cytosol and organelles. Notably, aaRSs are more likely to be co-localised

in cytosol and mitochondria such as GluRSs of rice and SerRS of *Arabidopsis*, while only few are predicted to be shared between cytosol and chloroplasts, such as HisRS of *Arabidopsis* and GlyRS of rice. Although several aaRSs were predicted as triple-targeting protein, especially in rice, only one AlaRS (F4I4Z2) of *Arabidopsis* has been verified by *in vivo* observation (Fig. 5; Table 1) (Mireau et al., 1996; Duchêne et al., 2005). The aaRSs of *P. patens* were mostly predicted as cytosolic localisation, however it still remains debatable whether moss has specific ways to dispatch aaRSs.

Of note, the *Arabidopsis* genome only contains one copy of *GlnRS* (versus two highly conserved copies in the rice genome), and this GlnRS is predicted as a cytosolic protein. The organellar Gln-tRNA^{Gln} synthesis uses a transamidation pathway to convert Glu-tRNA^{Gln}, synthesised by non-discriminating organellar GluRS, to Gln-tRNA^{Gln} (Pujol et al., 2008). This activity is very likely inherited from the prokaryotic ancestors of mitochondria and chloroplasts, since GlnRS is absent in the majority of Bacteria species (Ibba and Söll, 2000). In contrast, although Asn-tRNA^{Asn} can be synthesised in a similar transamidation pathway, we found that plant mitochondria and chloroplasts had their own AsnRSs. Correspondingly, AsnRS is present in most Bacteria species (Ibba and Söll, 2000). In *P. patens*, we identified two single-copied aaRSs, ArgRS and CysRS (Table 1). Despite their predicted cytosolic localisation, they are very likely shared by organelles. In *Arabidopsis* and rice, four ArgRSs showed a high sequence similarity (identity between 68%~78%), as did six CysRSs (sequence identity between 52%~69%). Three of the CysRSs show similar catalysing domains while one (F4IPY2) contains an N terminal signal for organellar localisation (Peeter et al., 2000). This indicates that the ArgRSs and CysRSs in higher plants may have a common origin, with extra signal peptides added to some copies to guide organellar localisation.

Generally, although plant aaRSs are coarsely divided into cytosolic ones and organellar ones, dual or even triple targeting is a common feature. This mechanism may ensure backups to maintain biochemical functions when damage occurs.

The Tertiary Structures of Plant AaRSs

Protein crystallography has dissected the tertiary structures of most aaRSs in human, *Escherichia coli*, and many other organisms, providing insights into evolutionary comparison, target based drug design, and protein complex formation (Giege and Springer, 2016; Fang and Guo, 2017). In contrast, tertiary structure was not known for any plant aaRSs. Here we used a protein modelling (Biasini et al., 2014) approach to predict the structures of six representative kinds of aaRSs from *Arabidopsis* and rice (Fig. 6). In each group of ArgRSs, CysRSs and GlnRSs, cytosolic proteins and organellar proteins showed very similar tertiary structures (Fig. 6A), suggesting a common evolutionary origin for cytosolic and organellar aaRSs in these cases, which is also implied by single-copied ArgRS, CysRS and GlnRS in *P. patens* (Table 1). In contrast, IleRSs, LysRSs and MetRSs represent the diverse-origin pattern, in which cytosolic proteins and organellar proteins showed divergent tertiary structures, likely originating from different ancestors (Fig. 6B). These observations are also consistent with the grouping of homologs (Table 1) and subcellular localisation predictions (Fig. 5; Supplemental Table 1).

To further provide supportive information the diverse-origins pattern, which appears more common (Table 1), we investigated *Arabidopsis* LeuRSs (Fig. 7), whose evolution follows a standard model, in which Archaea is nearer to Eucarya, and organellar aaRSs are more similar with bacterial ones (Doolittle and Handy, 1998). In *Arabidopsis*, two LeuRSs shared low similarities between primary sequences (34% identity), and predicted to be localised in cytosol and organelles, respectively. $LeuRS_{cyto}$ and $LeuRS_{organ}$ have same main functional domains: active site containing Rossmann fold and other accessory structures, CP1 editing domain, and $tRNA^{Leu}$ binding domain. But $LeuRS_{cyto}$ displays eukaryotic and archaeal characteristics. Its CP2 domain is located after CP1 domain, in contrast with the $LeuRS_{organ}$/Bacterial CP2 domain which precedes the CP1 domain (Zhou et al., 2008); eukaryotic extended LSD 1 (leucine-specific domain 1) and LSD2 are located around Class I signature

sequence KMSKS (Fukunaga and Yokoyama, 2005, Zhou et al., 2010). The orientation of the CP1 editing domain significantly varies between the two LeuRSs, possibly determined by the insertion point between other secondary structures and the elasticity of the bridge structure CP1 hairpin (Fig. 7) (Fukunaga and Yokoyama, 2005; Huang et al., 2014). The protein modelling of *Arabidopsis* LeuRSs provides more supportive evidence to understand the evolutionary classification, origins, and subcellular localisation of aaRSs.

The Expression Patterns of Plant AaRSs during Development

As housekeeping proteins, aaRSs have varied expression pattern spatially and temporally rather than a ubiquitous, unchanged pattern. Researches in animals and human provide plenty of examples. A TrpRS is highly expressed in *Drosophila* salivary gland during the whole embryogenesis (Seshaiah and Andrew, 2009). A TyrRS is localised in axon terminals in differentiating primary motor neuron, and disturbing this specific distribution pattern causes degenerative nerve disease (Jordanova et al., 2006). These findings indicate the appropriate expression patterns of aaRSs are essential for some specific developmental events. For plants, researches in the past two decades uncover important roles of aaRSs in various biological progresses, such as organelles biogenesis (Kim et al., 2005; Wang et al., 2016), organ development (Berg et al., 2005; Jiang et al., 2012; Kägi et al., 2010; Yang et al., 2018), and possible amino acid signalling pathways (Luna et al., 2014). Here, we extracted the expression raw data of aaRSs of *Arabidopsis* and rice from public databases and visualised their spatial and temporal patterns using heatmaps (Fig. 8). In *Arabidopsis*, based on the expression profiling the *aaRS* genes were obviously grouped into two major groups (with three exceptional aaRSs) (Fig. 8A), which were intriguingly in accord with the cytosolic/eukaryotic-originated and organellar/prokaryotic-originated groups that we described above (Table 1). The cytosolic *aaRS* genes (in green) were preferentially expressed in reproductive tissues such as young/mature carpels, young anthers, and young seeds. The young roots especially the apexes surprisingly accumulated transcripts of cytosolic

aaRS genes. In contrast, the transcripts of organellar *aaRS* genes (in black) were more accumulated in vegetative tissues like leaves, pedicels and seedlings. Notably, this group was also preferentially expressed in some floral organs that are not directly involved in the reproduction progress, such as mature petals and sepals (Fig. 8A). Similar grouping was also shown in the expression heatmap of rice aaRSs genes (Fig. 8B). The expressions of cytosolic aaRSs were more active in reproductive tissues including anthers, ovaries, pistils, and young inflorescence, whereas the organellar ones were highly expressed in vegetative tissues like leaves, and supportive floral organs like mature lemmas and paleas. Two extraordinary peaks of transcripts were a *CysRS* in anthers between 1.6mm to 2.0mm and an *AlaRS* in leaf blades, suggesting potential importance of these two genes in rice leaves and anthers respectively.

Discussion

AaRSs present amino acid substrates to ribosomes after accurately adding cognate tRNA molecules. This ancient housekeeping enzyme family exists in all organisms synthesising proteins, yet has not been systematically understood in the plant kingdom. Here we collected aaRS members in three model plants *Arabidopsis*, rice and *P. patens*, analysed key motifs and additional domains and predicted subcellular localisation and tertiary structures.

A total of 46, 53, and 42 aaRSs were identified from *Arabidopsis*, rice and *P. patens*, respectively. As the 20 proteinogenic amino acids require only 20 corresponding aaRSs, this means that plant genomes encode more than one set of aaRSs. Some of the extra copies may result from large-scale gene duplication (Bowers, 2003), when land plants began to diverge from their common ancestor with bryophytes 450 million years ago (Ma), explaining why *Arabidopsis* has more copies of *ArgRS*, *CysRS*, *TyrRS*, *GlyRS*, *AsnRS* and *AspRS* than *P. patens*. Generally, each kind of plant aaRS has both cytosolic and organellar versions, indicating different evolutionary origins. The presence of GlyRS (IIc) and LysRS supports that some organellar aaRSs are likely originated from chloroplast ancestor. Besides,

novel *aaRS* genes have evolved after species emergence, such as one *ThrRS* (protein Q8GZ45) in *Arabidopsis* and one *AlaRS* (protein B9FWG9) in rice. Thus, current plant aaRSs are the consequences of various origins, gene duplication/loss, and novel gene production. The alignment of key motif sequences also demonstrates the evolutionary differences between cytosolic and organellar aaRSs. Notably, the subclasses of motif 1(IIb) for Class II aaRSs are not observed in prokaryotes (Chaliotis et al., 2016), indicating more active evolution of AsnRS, AspRS and LysRS in plants.

AaRSs seem to enable multitasking beyond translation by developing novel functions for existing domains, or by interacting with other proteins in regulatory networks (Guo and Schimmel, 2013; Laporte et al., 2014). In this study we predicted various additional domains of plant aaRSs, which are likely involved in enhancing enzymatic activities, mediating protein-protein interactions, and other potential roles beyond translation. For instance, one rice MetRS has a specialised OB fold annotated as EMAP II (endothelial-monocyte-activating polypeptide II), which has been reported as an enhancer of aminoacylation reaction (Kaminska et al., 2000). The existence of MSC is universal, but in plants, only the interaction between cytosolic GluRS and its cofactor has been demonstrated in rice (Yang et al., 2018, Chapter 3).

Generally, the subcellular localisation of aaRSs is associated with three factors. The dominant determinant for the subcellular localisation of aaRSs is the evolutionary origin. For example, we found that the *Arabidopsis* organelle-localised $AspRS_{m,ch}$ (F4JJT9) shares higher similarity (sequence identities 62%) to homologs in cyanobacteria than that of its paralogs (Q9M084 and Q8H104) (sequence identities 33%), which are likely eukaryotically originated and cytosol-localised. The second determinant for the subcellular localisation of aaRSs is their targeting signal peptides, which guide the import of aaRSs into mitochondria and/or chloroplasts, and those peptides usually help organise dual targeting (Berglund et al., 2009). For example, the first 2-20 residues of ProRS (Q9FYR6) are essential for chloroplastic localisation rather than mitochondrial localisation, whereas 2-16 residues

of AspRS (F4JJT9) are more important for importing to mitochondria than chloroplasts (Berglund et al., 2009). Finally, the subcellular localisation of aaRSs may be sorted by their cofactors. In yeast (*Saccharomyces cerevisiae*), the cytosolic GluRS is also detectable in mitochondria. In the cytosol, a cofactor of GluRS, Arc1p forms complexes with GluRS to maintain their cytosolic localisation. Mutation of Arc1p causes the release of more GluRS into mitochondria possibly under the guidance of the targeting peptide embedded between the N terminal GST_C domain and catalytic domain of GluRS (Frechin et al., 2009). One unusual nuclear localisation event was identified in wheat TaMetRS (Zuo et al., 2016). Under mycotoxin treatment, TaMetRS is clearly concentrated in nucleus, which is observed by fusion with green fluorescence protein, indicating a link between aaRS and plant immune responses. However, the dynamics of aaRSs, the timing and the extent of distributing aaRSs to different compartments still needs to be further elucidated.

Surprisingly, the spatial expression patterns of *aaRS* genes are linked to the subcellular sorting of aaRS proteins. In *Arabidopsis* and rice, the genes encoding cytosolic aaRSs are more actively expressed in reproductive tissues while those encoding organellar aaRSs are preferentially expressed in vegetative tissues. This suggests that the *aaRS* genes and their upstream regulators might be evolved in a coordinated pattern to preserve the regulation relationship.

In short, our results indicate that plant aaRSs have multiple evolutionary origins, and more potential functions beyond aminoacylation to be investigated. Our study provides an informative resource for future functional studies in plant aaRSs.

Materials and Methods

Database Search for Plant Aminoacyl-tRNA Synthetases, Motifs Alignment, Additional Domain and Subcellular Localisation Prediction

To identify aaRS family members in *Arabidopsis,* rice and *P. patens,* BLASTP searches and protein name queries were performed in NCBI (http://www.ncbi.nlm.nih.gov), TAIR

(https://www.arabidopsis.org), Rice Genome Annotation Project (http://rice.plantbiology.msu.edu) and UniProt (http://www.uniprot.org). To search for the genetic loci encoding aaRSs in *Arabidopsis* and rice genomes, the protein sequences were used as query and TBLASTN program was used in TAIR and Rice Genome Annotation Project. The default parameters were used for BLAST searches. Redundant or truncated sequences were excluded. Motif sequences were isolated from aaRS sequences based on annotations from InterPro (https://www.ebi.ac.uk/interpro). The sequences of 66 HIGH motif, 66 KMSKS motif, 31 Motif 1(IIa), 24 Motif 1(IIb), 75 Motif 2, and 75 Motif 3 were analysed by Logos (http://weblogo.berkeley.edu/logo.cgi), respectively (Supplemental Table 1). The annotations of additional domains were determined according to searches in NCBI and InterPro. The subcellular localisation of aaRSs was predicted by PREDOTAR (https://urgi.versailles.inra.fr/predotar/) (Small et al., 2004).

Phylogenetic Analysis

A phylogenetic tree for SerRS was constructed with aligned full-length sequences of homologs of OsERS1. MEGA (version 7.0) and the neighbor-joining methods were used with a p-distance model and pairwise deletion and bootstrap (Gap Ktuple, 2; length penalty, 5; Window, 4; Diagonals, 4; 1000 replicates). UniProt sequences ID used for this analysis were *A. thaliana*, Q8RWT8, Q39230; *O. sativa*, Q2R1C1, Q5JKQ3, Q10QF5; *P. patens*, A9TH40, A9RJR8, A9SZI6; *E. coli*, P0A8L1; *H.sapiens*, P49591, Q9NP81; *S. cerevisiae*, P07284, P38705; *T. thermophilus*, P34945; *Z. mays*, B4FQC4, B6TBG7; *S. lycopersicum*, K4B0S5, K4BU01.

Protein Modelling

Since no plant LeuRS structure has been reported, modelling for *Arabidopsis* LeuRSs using SWISS-MODEL (https://swissmodel.expasy.org) (Biasini et al., 2014) was performed. Models of $LeuRS_{cyto}$ (UniProt ID: Q9XEA0) and $LeuRS_{organ}$ (Uniprot ID: F4I116) structures were generated with two templates (SMTL ID: 2bte.2 and 1wz2.1). Other templates used in protein modelling were: ArgRSs

(4r3z.1 for Q9C713, O23247, XP_015639785; 4zaj.1 for Q5ZEM0); CysRSs (1u0b.1 for Q9LYL3, B3LFA4, F4IPY2, Q0IZQ2, Q7G2N5, Q10RX2); GlnRSs (4ye6.1 for Q5VRX8, Q8W4F3; 4ye8.1 for Q65WY6); IleRSs (1jzq.1 for F4JLM5, B9FQ68; 1ffy.1 for Q8RXK8, Q6ZGC2); LysRSs (3bju.1 for Q9ZPI1, Q6F2U9; 3a74.1 for Q9LJE2; 4ex5.1 for Q6K1T2); MetRSs (1rqg.1 for Q9SVN5, Q9ZTS1; 1a8h.1 for Q9M2T9, Q338J4, Q10Q53).

Expression Heatmaps

The raw expression data of *aaRS* genes in *Arabidopsis* and rice were extracted from TAIR (https://www.arabidopsis.org) and RiceXPro (http://ricexpro.dna.affrc.go.jp), respectively. The raw data were input to ClustVis (https://biit.cs.ut.ee/clustvis/) for visualisation.

Accession Numbers

Accession number of genes and proteins are shown in Table 1 and Supplemental Table 1.

Acknowledgements

The contributions of D.Z. and G.L. was supported by Startup Funding from School of Agriculture, Food and Wine, University of Adelaide, and the contributions of X.Y. by a Beacon Scholarship, University of Adelaide. This work was supported by an Australia-China Science and Research Fund Joint Research Centre grant ACSRF48187.

Author Contributions

X.Y. and D.Z. conceived this study. X.Y. performed all bioinformatics analysis. X.Y., G.L. and D.Z. wrote the manuscript.

References

1. *Arabidopsis* Genome Initiative, 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, 408(6814), p.796.

2. Akashi, K., Grandjean, O. and Small, I., 1998. Potential dual targeting of an *Arabidopsis* archaebacterial-like histidyl-tRNA synthetase to mitochondria and chloroplasts. *FEBS letters*, 431(1), pp.39-44.

3. Arnez, J.G. and Moras, D., 1997. Structural and functional considerations of the aminoacylation reaction. *Trends in biochemical sciences,* 22(6), 211-216.

4. Beale, S.I., 1999. Enzymes of chlorophyll biosynthesis. *Photosynthesis research*, 60(1), pp.43-73.

5. Berg, M., Rogers, R., Muralla, R. and Meinke, D., 2005. Requirement of aminoacyl-tRNA synthetases for gametogenesis and embryo development in *Arabidopsis*. *The Plant Journal*, 44(5), pp.866-878.

6. Berglund, A.K., Pujol, C., Duchene, A.M. and Glaser, E., 2009. Defining the determinants for dual targeting of amino acyl-tRNA synthetases to mitochondria and chloroplasts. *Journal of molecular biology*, 393(4), pp.803-814.

7. Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T.G., Bertoni, M., Bordoli, L. and Schwede, T., 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic acids research*, 42(W1), pp.W252-W258.

8. Bowers, J.E., Chapman, B.A., Rong, J. and Paterson, A.H., 2003. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature*, 422(6930), p.433.

9. Chaliotis, A., Vlastaridis, P., Mossialos, D., Ibba, M., Becker, H.D., Stathopoulos, C. and Amoutzias, G.D., 2016. The complex evolutionary history of aminoacyl-tRNA synthetases. *Nucleic acids research*, 45(3), pp.1059-1068.

10. Cotelle, V., Meek, S.E., Provan, F., Milne, F.C., Morrice, N. and MacKintosh, C., 2000. 14-3-3s

regulate global cleavage of their diverse binding partners in sugar-starved *Arabidopsis* cells. *The EMBO journal*, 19(12), pp.2869-2876.

11. Cusack, S., 1995. Eleven down and nine to go. *Nature Structural and Molecular Biology*, 2(10), p.824.
12. Doolittle, R.F. and Handy, J., 1998. Evolutionary anomalies among the aminoacyl-tRNA synthetases. *Current opinion in genetics & development*, 8(6), pp.630-636.
13. Duchêne, A.M., Giritch, A., Hoffmann, B., Cognat, V., Lancelin, D., Peeters, N.M., Zaepfel, M., Maréchal-Drouard, L. and Small, I.D., 2005. Dual targeting is the rule for organellar aminoacyl-tRNA synthetases in *Arabidopsis* thaliana. *Proceedings of the National Academy of Sciences*, 102(45), pp.16484-16489.
14. Duchêne, A.M., Peeters, N., Dietrich, A., Cosset, A., Small, I.D. and Wintz, H., 2001. Overlapping destinations for two dual targeted glycyl-tRNA synthetases in *Arabidopsis* thaliana and Phaseolus vulgaris. *Journal of Biological Chemistry*, 276(18), pp.15275-15283.
15. Fang, P. and Guo, M., 2017. Structural characterization of human aminoacyl-tRNA synthetases for translational and nontranslational functions. *Methods*, 113, pp.83-90.
16. Frechin, M., Senger, B., Brayé, M., Kern, D., Martin, R.P. and Becker, H.D., 2009. Yeast mitochondrial Gln-tRNAGln is generated by a GatFAB-mediated transamidation pathway involving Arc1p-controlled subcellular sorting of cytosolic GluRS. *Genes & development*, 23(9), pp.1119-1130.
17. Fukunaga, R. and Yokoyama, S., 2005. Crystal structure of leucyl-tRNA synthetase from the archaeon Pyrococcus horikoshii reveals a novel editing domain orientation. *Journal of molecular biology*, 346(1), pp.57-71.
18. Giege, R. and Springer, M., 2016. Aminoacyl-tRNA Synthetases in the Bacterial World. *EcoSal Plus*, 7(1).

19. Gile, G.H., Moog, D., Slamovits, C.H., Maier, U.G. and Archibald, J.M., 2015. Dual organellar targeting of aminoacyl-tRNA synthetases in diatoms and cryptophytes. *Genome biology and evolution*, 7(6), pp.1728-1742.

20. Goff, S.A., Ricke, D., Lan, T.H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H. and Hadley, D., 2002. A draft sequence of the rice genome (*Oryza sativa* L. *ssp. japonica*). *Science*, 296(5565), pp.92-100.

21. Guo, M. and Schimmel, P., 2013. Essential nontranslational functions of tRNA synthetases. *Nature chemical biology*, 9(3), p.145.

22. Guo, M., Yang, X.L. and Schimmel, P., 2010. New functions of aminoacyl-tRNA synthetases beyond translation. Nature reviews *Molecular cell biology*, 11(9), p.668.

23. Han, J.M., Jeong, S.J., Park, M.C., Kim, G., Kwon, N.H., Kim, H.K., Ha, S.H., Ryu, S.H. and Kim, S., 2012. Leucyl-tRNA synthetase is an intracellular leucine sensor for the mTORC1-signalling pathway. *Cell,* 149(2), pp.410-424.

24. Hausmann, C.D. and Ibba, M., 2008. Aminoacyl‐tRNA synthetase complexes: molecular multitasking revealed. *FEMS microbiology reviews*, 32(4), pp.705-721.

25. Hoffman, R.M., 2011. Tumor-seeking Salmonella amino acid auxotrophs. *Current opinion in biotechnology*, 22(6), pp.917-923.

26. Huang, Q., Zhou, X.L., Hu, Q.H., Lei, H.Y., Fang, Z.P., Yao, P. and Wang, E.D., 2014. A bridge between the aminoacylation and editing domains of leucyl-tRNA synthetase is crucial for its synthetic activity. *RNA*, 20(9), pp.1440-1450.

27. Huber, S.C., MacKintosh, C. and Kaiser, W.M., 2002. Metabolic enzymes as targets for 14-3-3 proteins. *Plant molecular biology*, 50(6), pp.1053-1063.

28. Ibba, M. and Söll, D., 2000. Aminoacyl-tRNA synthesis. *Annual review of biochemistry*, 69(1), pp.617-650.

29. Ibba, M., Becker, H.D., Stathopoulos, C., Tumbula, D.L. and Söll, D., 2000. The adaptor hypothesis revisited. *Trends in biochemical sciences*, 25(7), pp.311-316.

30. Jentsch, S. and Pyrowolakis, G., 2000. Ubiquitin and its kin: how close are the family ties? *Trends in cell biology*, 10(8), pp.335-342.

31. Jewell, J.L., Russell, R.C. and Guan, K.L., 2013. Amino acid signalling upstream of mTOR. *Nature reviews Molecular cell biology*, 14(3), p.133.

32. Jiang, L., Wang, S., Li, H., Zhang, G. and Li, H., 2012. EMBRYONIC FACTOR 31 encodes a tyrosyl-tRNA synthetase that is essential for seed development. *Molecular biology reports*, 39(8), pp.8297-8305.

33. Jordanova, A., Irobi, J., Thomas, F.P., Van Dijck, P., Meerschaert, K., Dewil, M., Dierick, I., Jacobs, A., De Vriendt, E., Guergueltcheva, V. and Rao, C.V., 2006. Disrupted function and axonal distribution of mutant tyrosyl-tRNA synthetase in dominant intermediate Charcot-Marie-Tooth neuropathy. *Nature genetics*, 38(2), p.197.

34. Kägi, C., Baumann, N., Nielsen, N., Stierhof, Y.D. and Groß-Hardt, R., 2010. The gametic central cell of *Arabidopsis* determines the lifespan of adjacent accessory cells. *Proceedings of the National Academy of Sciences*, 107(51), pp.22350-22355.

35. Kaminska, M., Deniziak, M., Kerjan, P., Barciszewski, J. and Mirande, M., 2000. A recurrent general RNA binding domain appended to plant methionyl-tRNA synthetase acts as a cis-acting cofactor for aminoacylation. *The EMBO journal*, 19(24), pp.6908-6917.

36. Kim, M.J., Park, B.J., Kang, Y.S., Kim, H.J., Park, J.H., Kang, J.W., Lee, S.W., Han, J.M., Lee, H.W. and Kim, S., 2003. Downregulation of FUSE-binding protein and c-myc by tRNA synthetase cofactor p38 is required for lung cell differentiation. *Nature genetics*, 34(3), p.330.

37. Kim, Y.K., Lee, J.Y., Cho, H.S., Lee, S.S., Ha, H.J., Kim, S., Choi, D. and Pai, H.S., 2005. Inactivation of organellar glutamyl-and seryl-tRNA synthetases leads to developmental arrest of

chloroplasts and mitochondria in higher plants. *Journal of Biological Chemistry*, 280(44), pp.37098-37106.

38. Ko, Y.G., Kang, Y.S., Kim, E.K., Park, S.G. and Kim, S., 2000. Nucleolar localisation of human methionyl–tRNA synthetase and its role in ribosomal RNA synthesis. *The Journal of cell biology*, 149(3), pp.567-574.
39. Ko, Y.G., Kim, E.K., Kim, T., Park, H., Park, H.S., Choi, E.J. and Kim, S., 2001. Glutamine-dependent antiapoptotic interaction of human glutaminyl-tRNA synthetase with apoptosis signal-regulating kinase 1. *Journal of Biological Chemistry*, 276(8), pp.6030-6036.
40. Laporte, D., Huot, J.L., Bader, G., Enkler, L., Senger, B. and Becker, H.D., 2014. Exploring the evolutionary diversity and assembly modes of multi-aminoacyl-tRNA synthetase complexes: Lessons from unicellular organisms. *FEBS letters*, 588(23), pp.4268-4278.
41. Liu, W., Fu, Y., Hu, G., Si, H., Zhu, L., Wu, C. and Sun, Z., 2007. Identification and fine mapping of a thermo-sensitive chlorophyll deficient mutant in rice (*Oryza sativa* L.). *Planta*, 226(3), pp.785-795.
42. Luna, E., Van Hulten, M., Zhang, Y., Berkowitz, O., López, A., Pétriacq, P., Sellwood, M.A., Chen, B., Burrell, M., Van De Meene, A. and Pieterse, C.M., 2014. Plant perception of β-aminobutyric acid is mediated by an aspartyl-tRNA synthetase. *Nature chemical biology*, 10(6), p.450.
43. Menand, B., Marechal-Drouard, L., Sakamoto, W., Dietrich, A. and Wintz, H., 1998. A single gene of chloroplast origin codes for mitochondrial and chloroplastic methionyl–tRNA synthetase in *Arabidopsis* thaliana. *Proceedings of the National Academy of Sciences*, 95(18), pp.11014-11019.
44. Mirande, M., 2017. The aminoacyl-tRNA synthetase complex. *In Macromolecular Protein Complexes* (pp. 505-522). Springer, Cham.

45. Mireau, H., Lancelin, D. and Small, I.D., 1996. The same *Arabidopsis* gene encodes both cytosolic and mitochondrial alanyl-tRNA synthetases. *The Plant Cell*, 8(6), pp.1027-1039.

46. Moschopoulos, A., Derbyshire, P. and Byrne, M.E., 2012. The *Arabidopsis* organelle-localised glycyl-tRNA synthetase encoded by EMBRYO DEFECTIVE DEVELOPMENT1 is required for organ patterning. *Journal of experimental botany*, 63(14), pp.5233-5243.

47. Nagel, G.M., Cumberledge, S., Johnson, M.S., Petrella, E. and Weber, B.H., 1984. The β subunit of E. coil glycyl-tRNA synthetase plays a major role in tRNA recognition. *Nucleic acids research*, 12(10), pp.4377-4384.

48. Park, B.J., Kang, J.W., Lee, S.W., Choi, S.J., Shin, Y.K., Ahn, Y.H., Choi, Y.H., Choi, D., Lee, K.S. and Kim, S., 2005a. The haploinsufficient tumor suppressor p18 upregulates p53 via interactions with ATM/ATR. Cell, 120(2), pp.209-221.

49. Park, S.G., Shin, H., Shin, Y.K., Lee, Y., Choi, E.C., Park, B.J. and Kim, S., 2005b. The novel cytokine p43 stimulates dermal fibroblast proliferation and wound repair. *The American journal of pathology*, 166(2), pp.387-398.

50. Peeters, N.M., Chapron, A., Giritch, A., Grandjean, O., Lancelin, D., Lhomme, T., Vivrel, A. and Small, I., 2000. Duplication and quadruplication of *Arabidopsis* thaliana cysteinyl-and asparaginyl-tRNA synthetase genes of organellar origin. *Journal of molecular evolution*, 50(5), pp.413-423.

51. Povey, S., Lovering, R., Bruford, E., Wright, M., Lush, M. and Wain, H., 2001. The HUGO gene nomenclature committee (HGNC). *Human genetics*, 109(6), pp.678-680.

52. Pujol, C., Bailly, M., Kern, D., Maréchal-Drouard, L., Becker, H. and Duchêne, A.M., 2008. Dual-targeted tRNA-dependent amidotransferase ensures both mitochondrial and chloroplastic Gln-tRNAGln synthesis in plants. *Proceedings of the National Academy of Sciences*, 105(17), pp.6481-6485.

53. Rensing, S.A., Lang, D., Zimmer, A.D., Terry, A., Salamov, A., Shapiro, H., Nishiyama, T., Perroud, P.F., Lindquist, E.A., Kamisugi, Y. and Tanahashi, T., 2008. The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. *Science*, 319(5859), pp.64-69.

54. Rokov-Plavec, J., Dulic, M., Duchêne, A.M. and Weygand-Durasevic, I., 2008. Dual targeting of organellar seryl-tRNA synthetase to maize mitochondria and chloroplasts. *Plant cell reports*, 27(7), pp.1157-1168.

55. Schimmel, P., 1987. Aminoacyl tRNA synthetases: general scheme of structure-function relationships in the polypeptides and recognition of transfer RNAs. *Annual review of biochemistry*, 56(1), pp.125-158.

56. Seshaiah, P. and Andrew, D.J., 1999. WRS-85D: a tryptophanyl-tRNA synthetase expressed to high levels in the developing Drosophila salivary gland. *Molecular biology of the cell*, 10(5), pp.1595-1608.

57. Small, I., Peeters, N., Legeai, F. and Lurin, C., 2004. Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics*, 4(6), pp.1581-1590.

58. Souciet, G., Menand, B., Ovesna, J., Cosset, A., Dietrich, A. and Wintz, H., 1999. Characterization of two bifunctional Arabdopsis thaliana genes coding for mitochondrial and cytosolic forms of valyl-tRNA synthetase and threonyl-tRNA synthetase by alternative use of two in-frame AUGs. *The FEBS Journal*, 266(3), pp.848-854.

59. Theobald, D.L., Mitton-Fry, R.M. and Wuttke, D.S., 2003. Nucleic acid recognition by OB-fold proteins. *Annual review of biophysics and biomolecular structure*, 32(1), pp.115-133.

60. Toth, M.J. and Schimmel, P., 1990. A mutation in the small (alpha) subunit of glycyl-tRNA synthetase affects amino acid activation and subunit association parameters. *Journal of Biological Chemistry*, 265(2), pp.1005-1009.

61. Tzima, E., Reader, J.S., Irani-Tehrani, M., Ewalt, K.L., Schwartz, M.A. and Schimmel, P., 2005.

VE-cadherin links tRNA synthetase cytokine to anti-angiogenic function. *Journal of Biological Chemistry*, 280(4), pp.2405-2408.

62. Wakasugi, K. and Schimmel, P., 1999. Two distinct cytokines released from a human aminoacyl-tRNA synthetase. *Science*, 284(5411), pp.147-151.
63. Wang, Y., Wang, C., Zheng, M., Lyu, J., Xu, Y., Li, X., Niu, M., Long, W., Wang, D., Wang, H. and Terzaghi, W., 2016. WHITE PANICLE1, a Val-tRNA synthetase regulating chloroplast ribosome biogenesis in rice, is essential for early chloroplast development. *Plant physiology*, pp.pp-01949.
64. Woese, C.R., Olsen, G.J., Ibba, M. and Söll, D., 2000. Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiology and Molecular Biology Reviews*, 64(1), pp.202-236.
65. Wolf, Y.I., Aravind, L., Grishin, N.V. and Koonin, E.V., 1999. Evolution of aminoacyl-tRNA synthetases—analysis of unique domain architectures and phylogenetic trees reveals a complex history of horizontal gene transfer events. *Genome research*, 9(8), pp.689-710.
66. Yang, X., Li, G., Tian, Y., Song, Y., Liang, W. and Zhang, D., 2018. A Rice Glutamyl-tRNA Synthetase Modulates Early Anther Cell Division and Patterning. *Plant physiology*, pp.pp-00110.
67. .
68. Yannay-Cohen, N., Carmi-Levy, I., Kay, G., Yang, C.M., Han, J.M., Kemeny, D.M., Kim, S., Nechushtan, H. and Razin, E., 2009. LysRS serves as a key signaling molecule in the immune response by regulating gene expression. *Molecular cell*, 34(5), pp.603-611.
69. Yu, J., Hu, S., Wang, J., Wong, G.K.S., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X. and Cao, M., 2002. A draft sequence of the rice genome (*Oryza sativa* L. *ssp. indica*). *Science*, 296(5565), pp.79-92.
70. Zhou, X.L., Wang, M., Tan, M., Huang, Q., Eriani, G. and Wang, E.D., 2010. Functional

characterization of leucine-specific domain 1 from eukaryal and archaeal leucyl-tRNA synthetases. *Biochemical Journal*, 429(3), pp.505-513.

71. Zhou, X.L., Zhu, B. and Wang, E.D., 2008. The CP2 domain of leucyl-tRNA synthetase is crucial for amino acid activation and post-transfer editing. *Journal of Biological Chemistry*, 283(52), pp.36608-36616.

72. Zuo, D.Y., Yi, S.Y., Liu, R.J., Qu, B., Huang, T., He, W.J., Li, C., Li, H.P. and Liao, Y.C., 2016. A deoxynivalenol-activated methionyl-tRNA synthetase gene from wheat encodes a nuclear localised protein and protects plants against Fusarium pathogens and mycotoxins. *Phytopathology*, 106(6), pp.614-623.

Figures and Legends

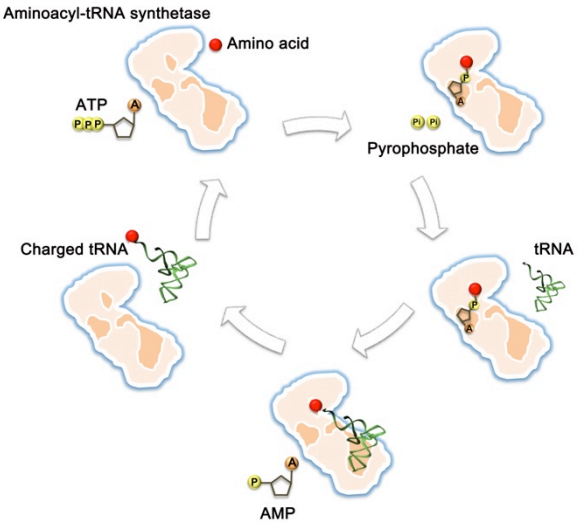


Figure 1. Biochemical function of aaRSs.

Aminoacyl-tRNA synthetases (aaRSs) produce the direct substrates for protein synthesis. AaRSs catalyse a key step in process of translation, where amino acids (aa) are attached to their cognate tRNA bearing the triplet anticodon code. This attachment is carried out in a two-step reaction:

$$aa + ATP + aaRS \rightarrow aaRS(aa–AMP) + PPi$$

$$aaRS(aa–AMP) + tRNA \rightarrow aa–tRNA + AMP + aaRS$$

			Dicot *Arabidopsis thaliana* (Arabidopsis)					Monocot *Oryza sativa* (Rice)					Moss *Physcomitrella patens (P. patens)*		
			Gene Locus	Name	Protein ID	Length	Localisation	Gene Locus	Name	Protein ID	Length	Localisation	Protein ID	Length	Localisation
Class I	Ia	ArgRS	AT1G66530		Q9C713	590	Cyto	LOC_Os01g06510		Q5ZEM0	588	Cyto	A9S0A2	606	Cyto
			AT4G26300		O23247	642	Cyto/Chloro*[1]	LOC_Os05g07030		XP_015639785	664	Mito/Chloro			
		CysRS	AT2G31170		F4IPY2	563	Mito/Chloro*[2]	LOC_Os09g38420		Q0IZQ2	569	Cyto/Mito	A9U4A4	491	Cyto
			AT3G56300		Q9LYL3	489	Cyto	LOC_Os10g32570		Q7G2N5	491	Cyto			
			AT5G38830		B3LFA4	511	Cyto	LOC_Os03g04960		Q10RX2	528	Cyto			
		IleRS	AT5G49030	*OVA2*	Q8RXK8	1,093	Mito/Chloro*[1]	LOC_Os02g53770		Q6ZGC2	1,058	Cyto/Chloro	A9RSP9	1,002	Cyto
			AT4G10320		F4JLM5	1,190	Cyto	LOC_Os06g43760		B9FQ68	1,157	Cyto	A9TB11	1,175	Cyto
		LeuRS	AT4G04350	*EMB2369*	Q9XEA0	973	Chloro*[1]	LOC_Os01g03020		XP_015621237	946	Mito/Chloro	A9SG46	906	Cyto
			AT1G09620		F4I116	1091	Cyto	LOC_Os09g32650		Q0J0N6	1095	Cyto	A9RZT5	1,104	Cyto
								LOC_Os09g21110		Q6H4P7	1094	Cyto			
		MetRS	AT3G55400	*OVA1*	Q9M2T9	616	Mito/Chloro*[3]	LOC_Os03g11120		Q10Q53	603	Cyto/Mito	A9SCS9	640	Mito/Chloro
			AT4G13780		Q9SVN5	797	Cyto	LOC_Os06g31210		Q9ZTS1	804	Cyto	A9SRC9	811	Cyto
								LOC_Os10g26050		Q338J4	633	Cyto			
		ValRS	AT1G14610	*TWN2*	P93736	1,108	Cyto/Mito*[1,5]	LOC_Os03g48850		Q851K9	1,054	Cyto*[9]	A9SJ82	975	Cyto
								LOC_Os10g36210		Q7G2B1	1,056	Cyto			
								LOC_Os03g02100		Q10ST2	943	Cyto			
			AT5G16715	*EMB2247*	F4KE63	974	Mito/Chloro*[1,4]	LOC_Os07g06940	*WP1*	Q7EZJ3	945	Mito/Chloro*[9]	A9SEL3	906	Cyto
	Ib	GlnRS	AT1G25350	*OVA9*	Q8W4F3	795	Cyto	LOC_Os01g09000		Q5VRX8	796	Cyto	A9S2J6	798	Cyto
								LOC_Os05g08990		Q65WY6	796	Cyto			
		GluRS	AT5G26710		O82462	719	Cyto	LOC_Os10g22380	*OsERS1*	Q8S5M6	713	Cyto/Mito*[10]	A9S8X8	659	Cyto
								LOC_Os01g16520		Q9LIZ8	715	Cyto/Chloro	A9S3X4	671	Cyto
			AT5G64050	*OVA3*	Q9FEA2	570	Mito/Chloro*[1]	LOC_Os02g02860	*CDE1*	Q6YUS0	567	Cyto/Mito	A9SYW3	586	Cyto
	Ic	TrpRS	AT2G25840	*OVA4*	Q8RXE9	408	Mito/Chloro*[1]	LOC_Os01g54020		Q5JKV2	399	Cyto/Mito	A9REV4	415	Cyto
			AT3G04600		Q9SR15	402	Cyto	LOC_Os12g35570		Q2QP62	408	Cyto	A9RQM2	397	Cyto
		TyrRS	AT2G33840		Q8S9J2	385	Cyto	LOC_Os08g05490		Q84PX0	394	Cyto	A9RKG5	388	Cyto
			AT1G28350		F4HWL4	748	Cyto	LOC_Os08g23110		B9F0V7	380	Cyto			
			AT3G02660	*EMB2768*	Q9M876	511	Mito/Chloro*[4]	LOC_Os01g31610		Q943W5	497	Cyto/Mito/Chloro	A9RBJ1	502	Mito/Chloro
Class II	II a	GlyRS	AT1G29880		O23627	729	Cyto/Mito*[6]	LOC_Os04g32650		Q7XLG5	673	Cyto/Mito/Chloro	A9TSY1	623	Cyto/Mito
								LOC_Os08g42560		Q6ZD35	740	Cyto			
		HisRS	AT3G46100	*HRS1*	O82413	486	Mito/Chloro*[7]	LOC_Os02g51830		Q6Z695	487	Cyto/Mito/Chloro	A9S048	426	Mito
			AT3G02760		F4IYF8	883	Cyto/Chloro	LOC_Os05g05840		P93422	788	Cyto/Chloro	A9U0U2	431	Mito
		ProRS	AT3G62120		Q9M1R2	530	Cyto	LOC_Os12g25710		Q2QS11	510	Cyto	A9RFZ3	544	Mito
			AT5G52520	*OVA6*	Q9FYR6	543	Mito/Chloro*[4]	LOC_Os07g07060		Q7XID8	545	Cyto/Mito/Chloro	A9TQ78	536	Cyto/Mito
		ThrRS	AT5G26830		O04630	709	Cyto/Mito*[1,5]	LOC_Os08g19850		Q6Z0N5	745	Mito	A9U0D5	685	Cyto
			AT2G04842	*EMB2761*	F4IFC5	650	Mito/Chloro*[4]	LOC_Os02g33500	*LAS*	Q6ER90	674	Cyto/Mito/Chloro	A9SJF5	615	Cyto
													A9RXJ2	613	Cyto
			AT1G17960		Q8GZ45	458	Cyto								
		SerRS	AT1G11870	*OVA7*	Q8RWT8	514	Mito/Chloro*[1]	LOC_Os11g39670		Q2R1C1	509	Cyto/Mito/Chloro	A9TH40	461	Cyto
			AT5G27470		Q39230	451	Cyto	LOC_Os01g37837		Q5JKQ3	425	Cyto	A9RJR8	464	Cyto
								LOC_Os03g10190		Q10QF5	445	Cyto	A9SZI6	451	Cyto
	II b	AsnRS	AT3G07420		Q9SW95	638	Cyto	LOC_Os12g22600		Q2QSX5	615	Cyto			
			AT4G17300	*OVA8*	O48593	567	Mito/Chloro*[1,2]						A9RTZ3	501	Cyto
			AT5G56680		Q9SW96	572	Cyto	LOC_Os01g27520		Q93WM3	559	Cyto	A9S8S1	554	Cyto/Mito
			AT1G70980		Q9SSK1	571	Cyto						A9S8S6	493	Cyto/Mito
		AspRS	AT4G33760		F4JJT9	664	Mito/Chloro*[1,4]	LOC_Os01g06020		Q5ZD71	604	Mito/Chloro	XP_001753365	683	Cyto/Mito
			AT4G31180	*IBI1*	Q9M084	558	Cyto	LOC_Os02g46130		Q6ZHC3	549	Cyto	A9TI29	525	Cyto
			AT4G26870		Q8H104	532	Cyto	LOC_Os02g04700		Q6Z2Y0	508	Cyto			
		LysRS	AT3G11710	*ATKRS-1*	Q9ZPI1	626	Cyto	LOC_Os03g38980		Q6F2U9	602	Cyto	A9SRN4	619	Cyto
			AT3G13490	*OVA5*	Q9LJE2	602	Mito/Chloro*[1]	LOC_Os02g41470		Q6K1T2	586	Cyto/Mito/Chloro	A9SZV3	609	Cyto
	IIc	AlaRS	AT5G22800	*EMB86*	Q9FFC7	978	Mito/Chloro*[1]	LOC_Os06g13660		B9FSH5	996	Mito/Chloro	A9RG15	927	Cyto/Mito
			AT1G50200		F4I4Z2	1,011	Cyto/Mito/Chloro*[1,8]	LOC_Os10g10244		Q10A14	999	Mito	A9T4K7	937	Cyto
								LOC_Os07g15440		B9FWG9	1,329	Cyto			
		PheRS	AT4G39280		Q9T034	485	Cyto	LOC_Os10g26130		Q338I8	493	Cyto	A9TXE7	491	Cyto
			AT1G72550		Q9SGE9	598	Cyto	LOC_Os05g48510		Q6AT40	596	Cyto	A9SET8	599	Cyto
			AT3G58140		Q94K73	429	Mito/Chloro*[1]	LOC_Os12g34860		Q2QPD4	448	Mito	A9T5K9	450	Cyto
		GlyRS	AT3G48110	*EDD1*	Q8L785	1,067	Mito/Chloro*[6]	LOC_Os06g01400		Q5VRH2	1,068	Cyto/Chloro	A9TR89	1,124	Cyto/Mito/Chloro

Table 1. Predicted aaRSs and classification in model plants.

Arabidopsis, rice and *P. patens* have 46, 53 and 42 predicted aaRSs proteins, respectively. Protein IDs are retrieved from Uniprot (http://www.uniprot.org) or NCBI (https://www.ncbi.nlm.nih.gov). Gene locus for *Arabidopsis* and rice aaRSs are retrieved from TAIR (https://www.arabidopsis.org) and Rice Genome Annotation Project (http://rice.plantbiology.msu.edu). Subcellular localisation are based on prediction tool PREDOTAR (https://urgi.versailles.inra.fr/predotar/) and/or experimental data (indicated by asterisks). In each aaRS subfamily, same background colour indicates homology and the two aaRSs in coral are unique ones without homologs in this review. References are [1] Duchêne et al., 2005; [2] Peeters et al., 2000; [3] Menand et al., 1998; [4] Berglund et al., 2009; [5] Souciet et al., 1999; [6] Duchêne et al., 2001; [7] Akashi et al., 1998; [8] Mireau et al., 1996; [9] Wang et al., 2016; [10] Yang et al., 2018, Chapter 3.

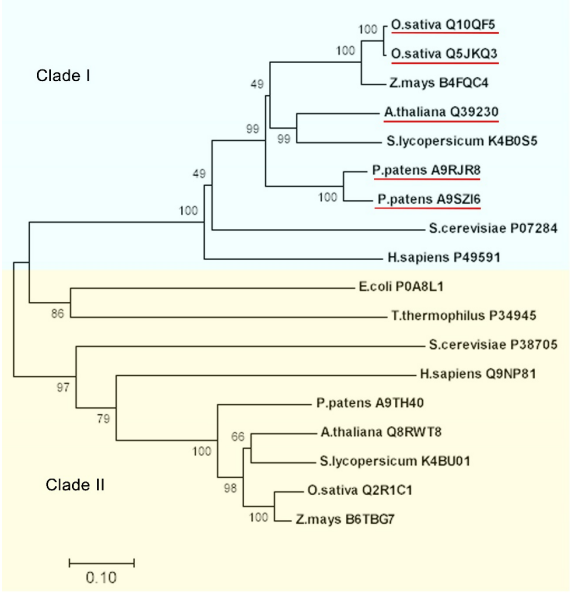


Figure 2. Phylogenetic analysis of SerRSs.

A neighbor-joining tree based on amino acid sequences. In Clade I, only one SerRS is shown in *Arabidopsis*, possibly due to gene loss after divergence with lower plants. Later, after monocots emerge, gene duplication in rice produces two SerRSs. SerRS is an excellent model to display how copies of *aaRS* genes changed during species evolution in plants.

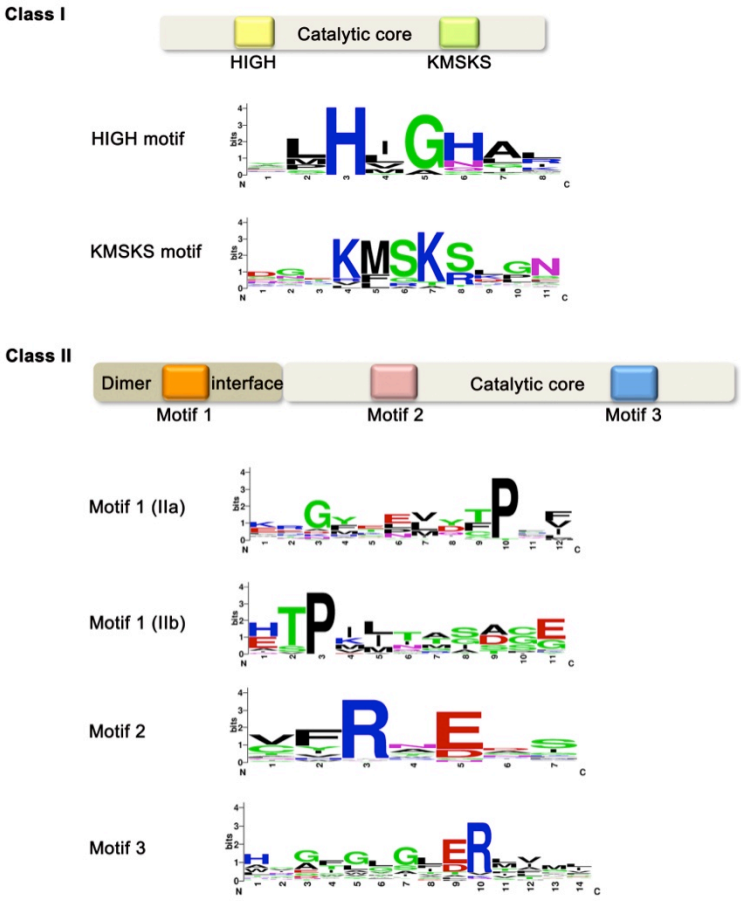


Figure 3. Conserved motifs of Class I and Class II plant aaRSs.

Motifs and their flanking sequences are identified from aaRSs to generate Logos of conserved amino acids. Class I aaRSs contains HIGH and KMSKS signatures, which are identified in all aaRSs in this study. In contrast, motifs of Class II usually contain one or two highly conserved residues, allowing more flexibility in other sites.

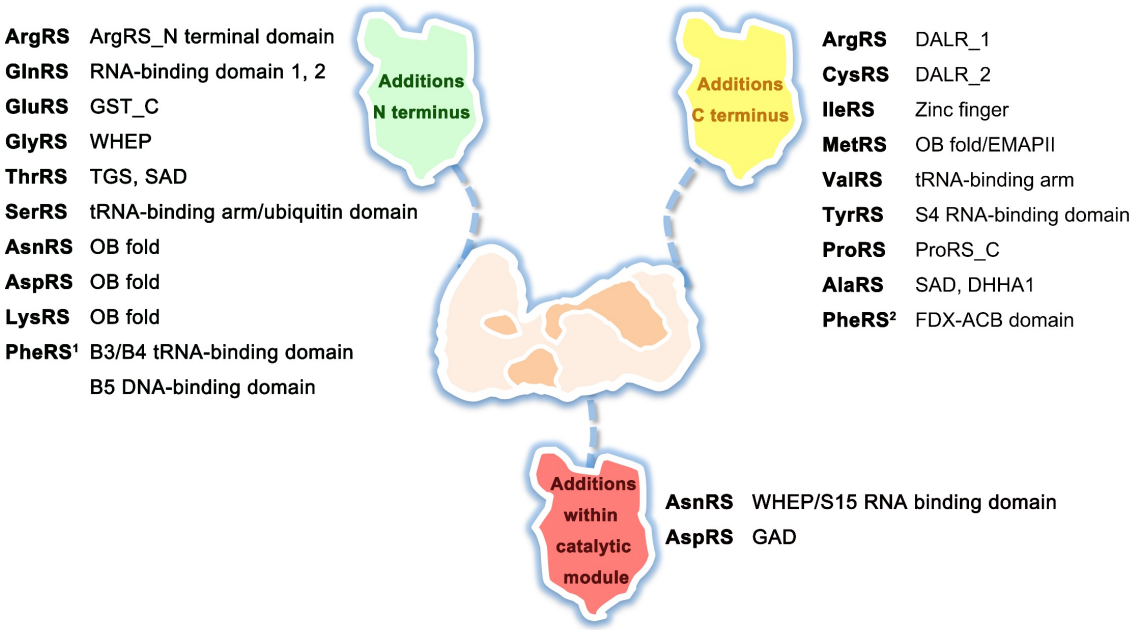


Figure 4. Summary of additional domains in plant aaRSs.

GST_C, glutathione S-transferase C terminal domain; WHEP, a conserved domain of 46 amino acids exists in a number of higher eukaryote aminoacyl-transfer RNA synthetases; TGS, threonyl-tRNA synthetase (ThrRS), GTPase, and guanosine 3', 5'-bis(diphosphate) 3'-pyrophosphohydrolasea (Spot), small domain that consists of ~50 amino acid residues, have potential ligand-binding and regulatory role (Wolf, 1999); SAD, the second additional domain, involves in formation of aaRS dimer; Ubiquitin domain, kin of ubiquitin, has potential in recruitment of ubiquitylated substrates to the proteasome (Jentsch 2000); OB fold, oligonucleotide/oligosaccharide binding motif, involves in multiple nucleotide activities; EMAPII, endothelial monocyte activating polypeptide II; B3/B4 tRNA-binding domain, has post-transfer editing function in PheRS beta subunit; B5 DNA-binding domain, may enable PheRS have effect on DNA behaviors; DALR_1/2, all alpha helical domain within the anticodon binding; Zinc finger, binds zinc and maintain aminoacylation activity; S4 RNA-binding domain, a small domain consisting of 60-65 amino acid residues that is universal in RNA-process and translation regulation related proteins; DHHA1, DHH associated domain, may be involved in DNA repair; MINDY, MIU (motif interacting with ubiquitin)-containing novel DUB (deubiquitinating), binds to ubiquitin; FDX-ACB domain, ferredoxin-fold anticodon binding domain, forms a RNA recognition fold; S15 RNA binding domain, is detected in ribosomal protein S15 and other RNA binding proteins; GAD, an insertion domain found in the glutamyl-tRNA aminotransferase subunit E and some AspRSs.

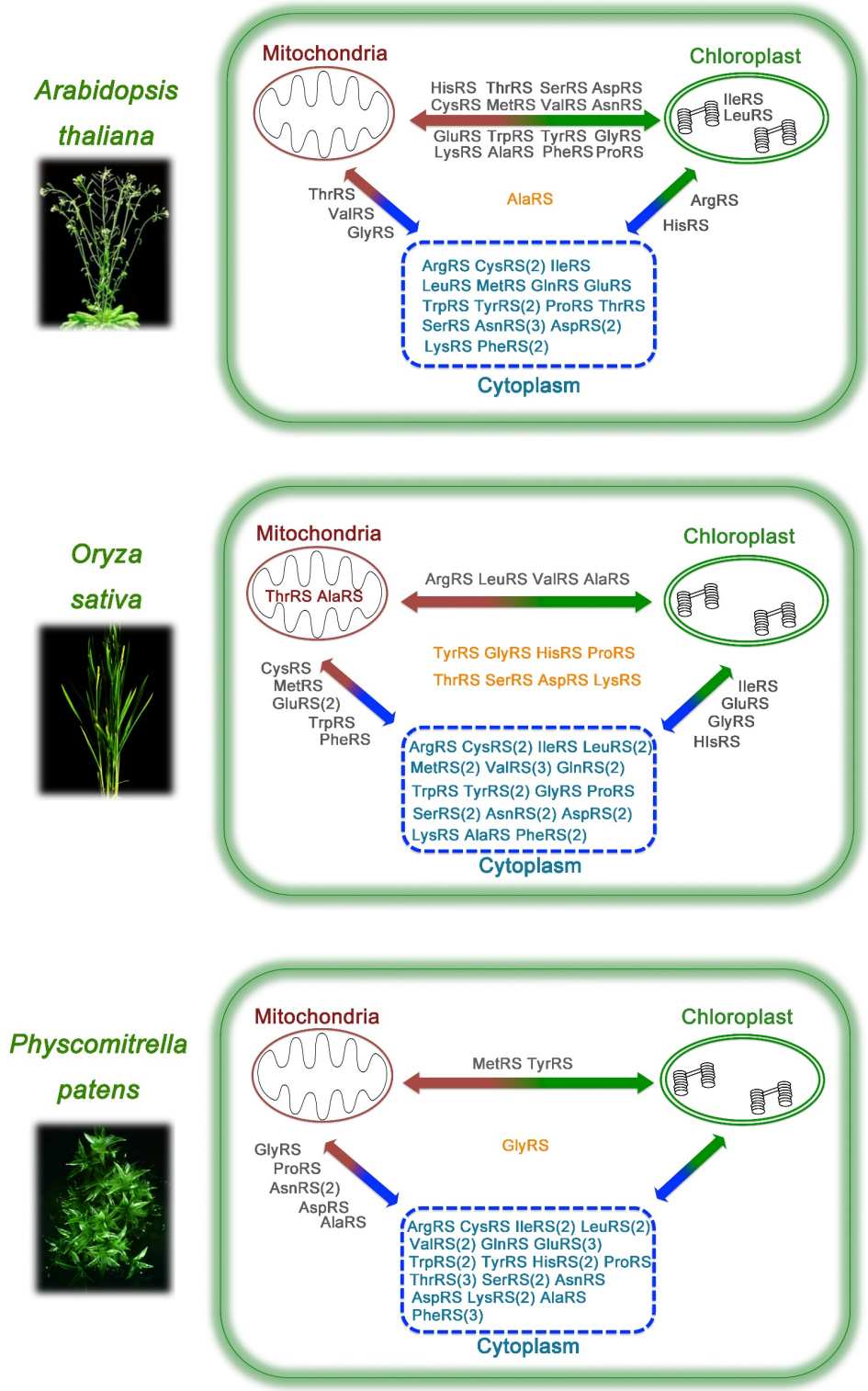


Figure 5. Subcellular localisation of plant aaRSs.

The putative distribution of aaRSs in three model plants. The arrows with two colours indicate co-localisation of aaRS between organelles or organelles and cytosol. The aaRSs in orange indicate triple-localisation.

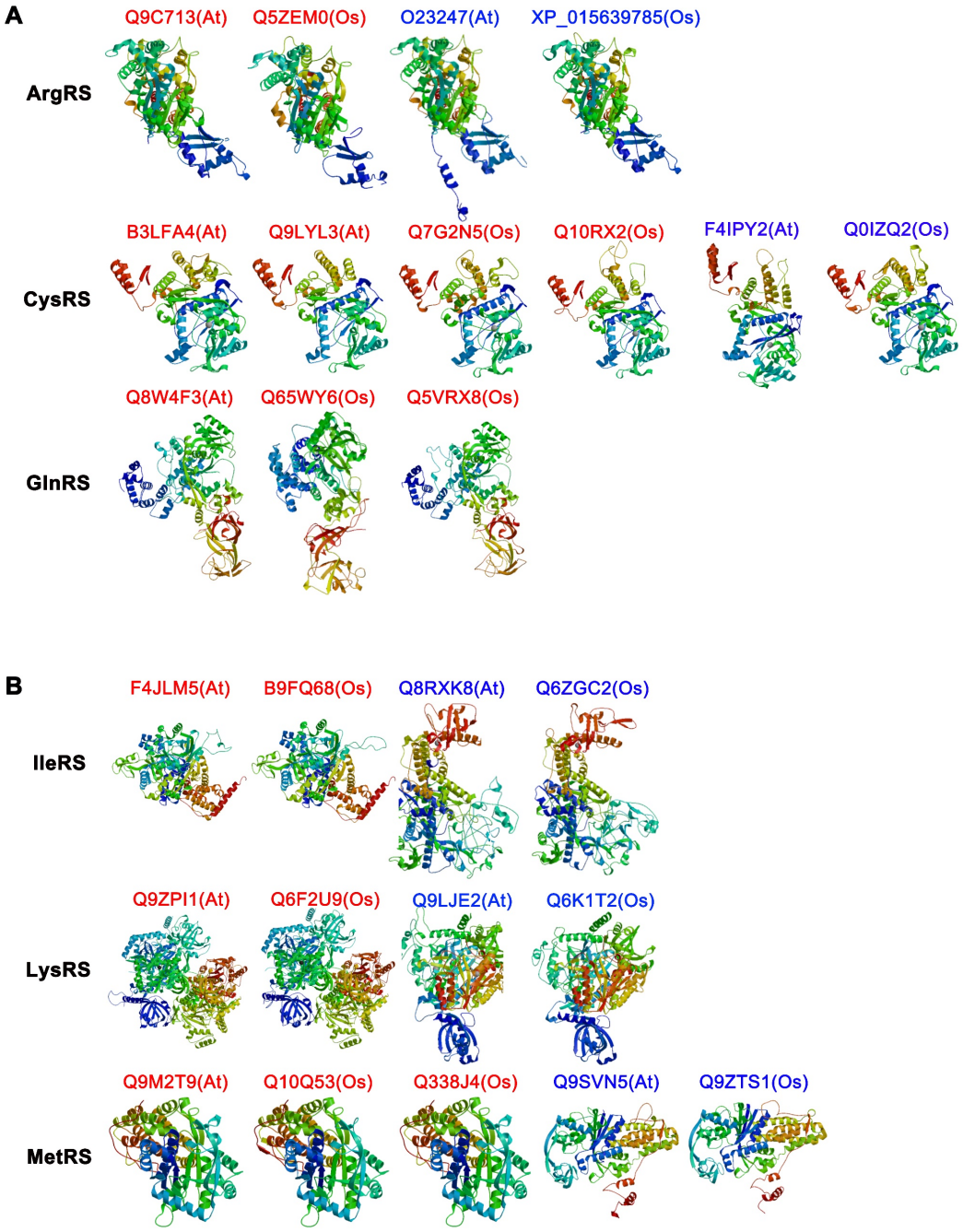


Figure 6. Comparison of tertiary structures of cytosolic and organellar aaRSs.

(A) Protein modelling of ArgRSs, CysRSs, and GlnRSs from *Arabidopsis* and rice. In each kind, the structures of cytosolic ones (red) and organellar ones (blue) are highly similar, indicating single evolutionary origin.

(B) Protein modelling of IleRSs, LysRSs, and MetRSs from *Arabidopsis* and rice. In each kind, the structures of cytosolic ones (red) and organellar ones (blue) are highly divergent, indicating different evolutionary origins. Cytosolic LysRSs are predicted to form dimers, and organellar LysRSs form tetramers. At, *Arabidopsis* proteins; Os, rice proteins.

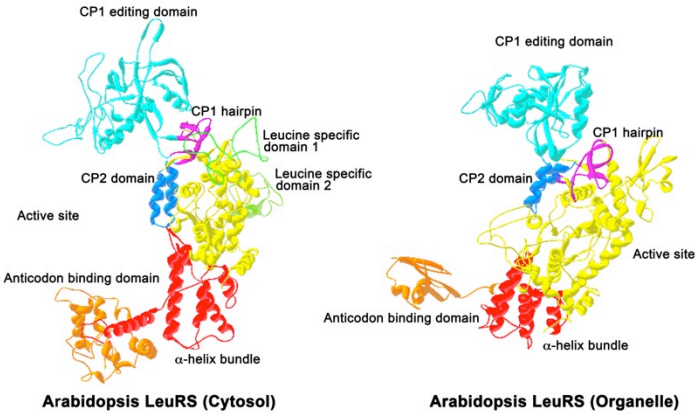
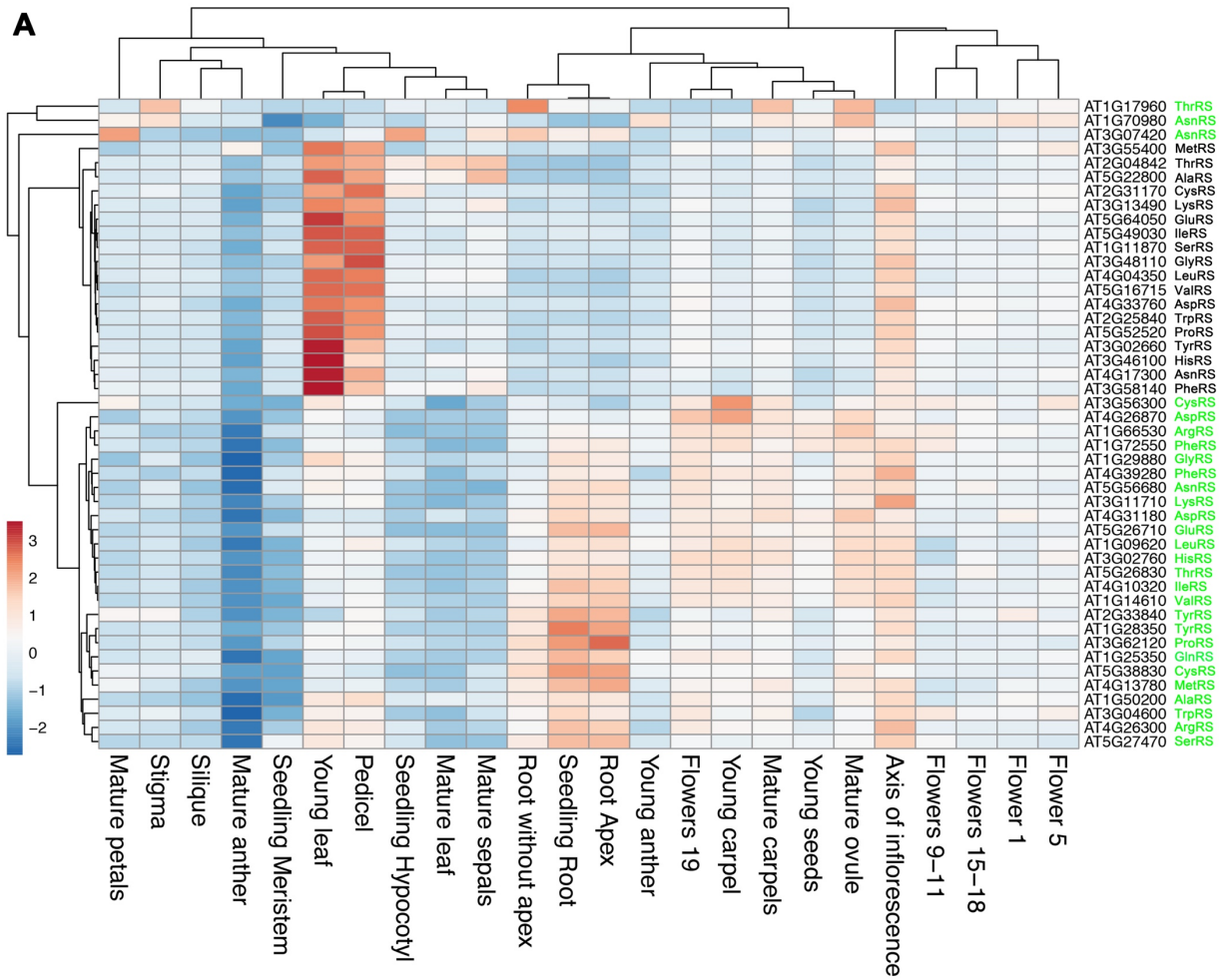


Figure 7. Dissection of structures of *Arabidopsis* LeuRSs.

Left, overall structure of cytosolic LeuRS. Right, overall structure of organellar LeuRS. The Rossmann-fold and accessory domains (yellow), CP1 editing domain (cyan), CP2 domain (aqua), CP1 hairpin (magenta), α-helix bundle (red), anticodon-binding domain (orange) and leucine specific domain 1 and 2 (green) were shown.


A
AT1G17960 ThrRS
AT1G70980 AsnRS
AT3G07420 AsnRS
AT3G55400 MetRS
AT2G04842 ThrRS
AT5G22800 AlaRS
AT2G31170 CysRS
AT3G13490 LysRS
AT5G64050 GluRS
AT5G49030 IleRS
AT1G11870 SerRS
AT3G48110 GlyRS
AT4G04350 LeuRS
AT5G16715 ValRS
AT4G33760 AspRS
AT2G25840 TrpRS
AT5G52520 ProRS
AT3G02660 TyrRS
AT3G46100 HisRS
AT4G17300 AsnRS
AT3G58140 PheRS
AT3G56300 CysRS
AT4G26870 AspRS
AT1G66530 ArgRS
AT1G72550 PheRS
AT1G29880 GlyRS
AT4G39280 PheRS
AT5G56680 AsnRS
AT3G11710 LysRS
AT4G31180 AspRS
AT5G26710 GluRS
AT1G09620 LeuRS
AT3G02760 HisRS
AT5G26830 ThrRS
AT4G10320 IleRS
AT1G14610 ValRS
AT2G33840 TyrRS
AT1G28350 TyrRS
AT3G62120 ProRS
AT1G25350 GlnRS
AT5G38830 CysRS
AT4G13780 MetRS
AT1G50200 AlaRS
AT3G04600 TrpRS
AT4G26300 ArgRS
AT5G27470 SerRS
3
2
1
0
−1
−2
Mature petals
Stigma
Silique
Mature anther
Seedling Meristem
Young leaf
Pedicel
Seedling Hypocotyl
Mature leaf
Mature sepals
Root without apex
Seedling Root
Root Apex
Young anther
Flowers 19
Young carpel
Mature carpels
Young seeds
Mature ovule
Axis of inflorescence
Flowers 9–11
Flowers 15–18
Flower 1
Flower 5

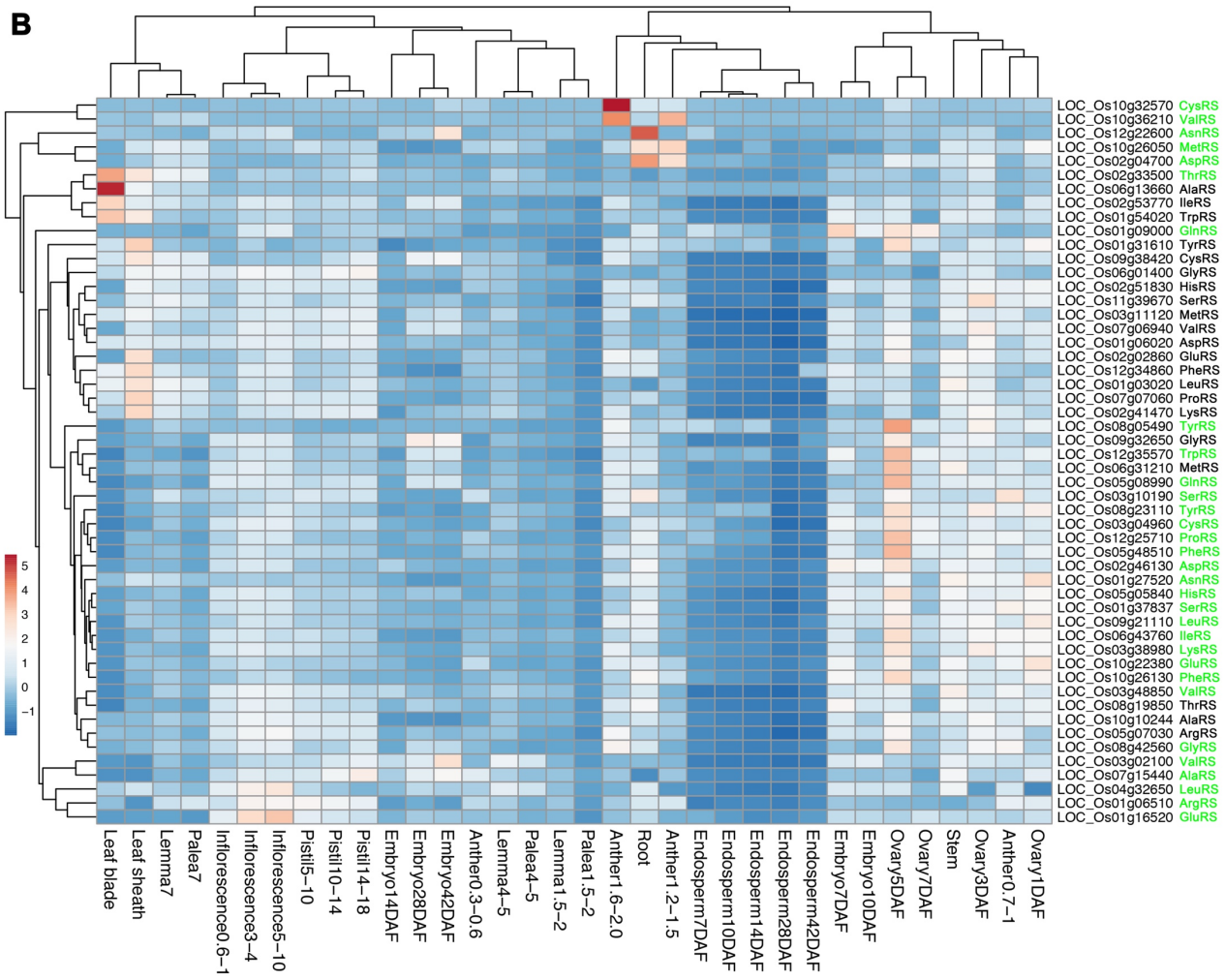


Figure 8. Spatial and temporal transcript accumulation patterns of aaRS genes in *Arabidopsis* and rice.

(A) Heatmap of *aaRS* genes expression profiling in various tissues and developmental stages in *Arabidopsis.*

(B) Heatmap of *aaRS* genes expression profiling in various tissues and developmental stages in rice.

The colour keys represent relative expression rank normalised from raw expression value.

The aaRSs in green indicate cytosol-localised ones or eukaryotically originated ones.

Supplemental Table 1

Subcellular localisation predicted by Predotar, key motif sequences and additional domains of aaRSs in *Arabidopsis*, rice and *P. patens*, respectively.

Arabidopsis thaliana				Predotar				Key motifs			Additional domain
				Mito	Plastid	ER	Elsewhere	HIGH motif	KMSKS motif		
Class I	Ia	ArgRS	Q9C713	0,02	0,06	0,01	0,92	HVGH	RFRTR		DALR_1
			O23247	0,09	0,03	0,04	0,84	HVGH	RFRTR		DALR_1
		CysRS	F4IPY2	0,21	0,70	0,00	0,23	HIGH	KMSKS		DALR-2
			Q9LYL3	0,01	0,01	0,00	0,98	HIGH	KMGKS		
			B3LFA4	0,01	0,00	0,00	0,99	HIGH	KMAKS		
		IleRS	Q8RXK8	0,07	0,52	0,00	0,44	HMGH	KMSKS		ZF-FPG-IleRS
			F4JLM5	0,01	0,01	0,00	0,99	HYGH	KMAKK		
		LeuRS	Q9XEA0	0,01	0,03	0,03	0,93	HVGH	KMSKS		
			F4I116	0,01	0,00	0,01	0,99	HIGH	KMSKS		
		MetRS	Q9M2T9	0,35	0,93	0,00	0,05	HMGS	KMGKS		
			Q9SVN5	0,01	0,00	0,04	0,95	HLGN	KFSKS		OB fold
		ValRS	P93736	0,88	0,04	0,09	0,11	HIGH	KMSKS		Lys-rich
			F4KE63	0,79	0,60	0,01	0,08	HMGH	KMSKS		
	Ib	GlnRS	Q8W4F3	0,01	0,00	0,00	0,99	HIGH	VMSKR		Lys-rich
		GluRS	O82462	0,01	0,05	0,09	0,86	HIGH	LLSKR		GST_C
			Q9FEA2	0,37	0,07	0,01	0,58	HVGG	KLSKR		
	Ic	TrpRS	Q8RXE9	0,18	0,39	0,03	0,48	HLGN	KMSKS		
			Q9SR15	0,01	0,00	0,00	0,99	HLGH	KMSAS		
		TyrRS	Q8S9J2	0,01	0,01	0,00	0,99	HIAQ	KMSKS		
			Q9M876	0,13	0,84	0,02	0,13	HLGN	KFGKS		RNAbinding S4
			F4HWL4	0,01	0,02	0,00	0,98	HIAQ	KMSKS		double
Class II								Motif 1	Motif 2	Motif 3	
	IIa	GlyRS	O23627	0,41	0,01	0,03	0,56	LEVDCPCV	FRNE	IADR	WHEP
		HisRS	O82413	0,50	0,01	0,16	0,42	YEEVDYPVL	WRYER	GFGFGxxV	
			F4IYF8	0,01	0,19	0,01	0,80	ATALDTPVF	YRRDN	GMSLGxxR	
		ProRS	Q9M1R2	0,01	0,00	0,00	0,99	VKNCYFPL	VRWE	AYSTR	ProRS-C_1
			Q9FYR6	0,36	0,19	0,31	0,36	HSNMYFPQ	TRWE	AVSTR	ProRS-C_1
		ThrRS	Q8GZ45	0,01	0,00	0,01	0,98		HRNE	GSIER	TGS
			F4IFC5	0,18	0,96	0,00	0,03	DLIYTPHV	YRYE	GSLER	tRNA_SAD
			O04630	0,91	0,00	0,02	0,09	TEVITPNM	HRNE	GSVER	TGS like, tRNA SAD
		SerRS	Q8RWT8	0,48	0,00	0,48	0,27	TPLTTPEI	FRTE	ACAxxR	Ubiquitin_dom,tRNA-bd_arm
			Q39230	0,01	0,00	0,00	0,99	TGLQPPFF	FRKE	LTAxxR	tRNA-bd_arm
	IIb	AsnRS	Q9SW95	0,01	0,01	0,00	0,98	PVITT	FIAD	SMEQ	
			O48593	0,29	0,97	0,02	0,02	PIITA	FRAE	GFER	OB N-terminal
			Q9SW96	0,01	0,02	0,00	0,97	PIITT	FRAE	GFER	WHEP OB
			Q9SSK1	0,01	0,00	0,00	0,99	PIITT	FRAE	GFER	WHEP OB
		AspRS	Q9M084	0,01	0,01	0,00	0,98	PKLMA	FRAE	GLER	OB
			F4JJT9	0,50	0,01	0,06	0,47	PILSR	FRDE	GLDR	OB GAD
			Q8H104	0,01	0,01	0,00	0,99	PKLIA	FRAE	GLER	OB
		LysRS	Q9ZPI1	0,01	0,01	0,00	0,98	PMMNM	FRNE	GIDR	OB fold
			Q9LJE2	0,26	0,56	0,05	0,31	PVLQG	FRNE	GIDR	OB fold
	IIc	AlaRS	Q9FFC7	0,04	0,00	0,12	0,85	VLPSS	IRT	GLGLER	tRNA SAD, DHHA1
			F4I4Z2	0,69	0,02	0,02	0,30	FWPSS	IRA	GMGFER	tRNA SAD, DHHA1
		PheRS	Q9SGE9	0,02	0,00	0,02	0,96	YTEVVTW	VML	GLLK	B3_4,B5,DNA-bd_dom_put
			Q94K73	0,39	0,48	0,02	0,31	FEKFEDxxS	YRRD	GLGLER	Ferredoxin-fold anticodon binding domain
			Q9T034	0,01	0,02	0,00	0,98	FEEMPNN	FRNE	GLSLER	
		GlyRS	Q8L785	0,09	0,15	0,96	0,03	CAVMQPSN	IRPD	GLER	

Oryza sativa				Predotar				Key motifs			Additional domain
				Mito	Plastid	ER	Elsewhere	HIGH motif	KMSKS motif		
Class I	Ia	ArgRS	Q5ZEM0	0,01	0,00	0,00	0,99	HVGH	RFRTR		DALR_1
			XP_015639785.1	0,56	0,63	0,08	0,15	HVGH	RFRTR		DALR_1
		CysRS	Q0IZQ2	0,39	0,03	0,60	0,24	HIGH	KMSKS		DALR-2
			Q10RX2	0,01	0,03	0,01	0,95	HIGH	KMSKS		
			Q7G2N5	0,01	0,00	0,00	0,99	HVGH	KMSDD		
		IleRS	Q6ZGC2	0,08	0,30	0,00	0,64	HMGH	KMSKS		ZF-FPG-IleRS
			B9FQ68	0,01	0,00	0,01	0,98	HYGH	KMSKS		
		LeuRS	XP_015621237.1	0,36	0,89	0,02	0,07	HVGH	KMSKS		
			Q0J0N6	0,01	0,00	0,00	0,99	HLGH	KMSKS		
			Q6H4P7	0,01	0,00	0,01	0,98	HLGH	KMSKS		
		MetRS	Q10Q53	0,80	0,01	0,05	0,19	HMGS	KMGKS		
			Q338J4	0,01	0,01	0,05	0,94	HLGN	KFSKT		
			Q9ZTS1	0,08	0,06	0,04	0,82	HLGN	KFSKS		EMAPII/OB
		ValRS	Q851K9	0,01	0,01	0,00	0,98	HIGH	KMSKS		
			Q7G2B1	0,01	0,00	0,00	0,99	HIGH	KMAKC		
			Q7EZJ3	0,81	0,63	0,08	0,06	HMGH	KMSKT		
			Q10ST2	0,01	0,00	0,01	0,98	HIGH	KMSKS		
	Ib	GlnRS	Q5VRX8	0,01	0,00	0,32	0,67	HIGH	VMSKR		
			Q65WY6	0,01	0,02	0,00	0,97	HIGH	VMSKR		
		GluRS	Q8S5M6	0,01	0,54	0,00	0,45	HIGH	LLSKR		GST_C
			Q9LIZ8	0,01	0,30	0,04	0,66	HIGH	VLSKR		GST_C
			Q6YUS0	0,25	0,02	0,05	0,70	HVGG	KLSKR		
	Ic	TrpRS	Q2QP62	0,01	0,00	0,00	0,99	HLGH	KMSAS		
			Q5JKV2	0,43	0,09	0,01	0,51	HLGN	KMSKS		
		TyrRS	Q84PX0	0,01	0,01	0,00	0,99	HIAQ	KMSKS		
			B9F0V7	0,01	0,01	0,00	0,98	HIAQ	KMSKS		
			Q943W5	0,79	0,22	0,05	0,16	HLGN	KFGKT		
Class II								Motif 1	Motif 2	Motif 3	
	II a	GlyRS	Q6ZD35	0,01	0,02	0,00	0,98	LEVDCPCV	FRNE	IADR	WHEP
			Q7XLG5	0,64	0,12	0,00	0,32	LEVDCPCV	FRNE	IADR	
		HisRS	Q6Z695	0,73	0,47	0,03	0,14	FEEVDFPVL	WRYER	GFGFGdaV	
			P93422	0,01	0,75	0,05	0,24	AVSLDTPVF	YRRDN	GVSLGieR	L-Aspartase-like.
		ProRS	Q7XID8	0,84	0,25	0,01	0,12	HSNMYFPQ	TRWE	AISTR	ProRS-C_1
			Q2QS11	0,01	0,00	0,01	0,98	LKPYYFPL	VRWE	AYTTR	ProRS-C_1
		ThrRS	Q6ER90	0,61	0,57	0,01	0,16	DLLYTPHV	YRYE	GSLER	tRNA_SAD
			Q6Z0N5	0,91	0,00	0,03	0,08	QEVLSPNI	HRNE	GSVER	TGS
		SerRS	Q5JKQ3	0,01	0,00	0,01	0,99	KPMQTPFF	FRKE	LTAteR	tRNA-bd_arm
			Q10QF5	0,01	0,00	0,01	0,99	KPMQTPFF	FRKE	LTAteR	tRNA-bd_arm
			Q2R1C1	0,44	0,47	0,04	0,29	TPLITPE	FRTE	AVAvpR	tRNA-bd_arm
	II b	AsnRS	Q2QSX5	0,01	0,04	0,00	0,95	PEAVR	FQAE	DLES	
			Q8GTK9	0,91	0,02	0,03	0,08	PIITA	FRAE	GFER	
			Q93WM3	0,01	0,02	0,00	0,98	PIITT	FRAE	GFER	OB, WHEP, S15_NS1
		AspRS	Q5ZD71	0,80	0,42	0,04	0,11	PITTA	FRDE	GLDR	GAD, OB
			Q6Z2Y0	0,01	0,01	0,00	0,99	PKLIG	FRAE	GLER	OB
			Q6ZHC3	0,01	0,01	0,00	0,98	PKLIA	FRAE	GLER	OB
		LysRS	Q6F2U9	0,01	0,01	0,00	0,99	PMMNM	FRNE	GIDR	OB
			Q6K1T2	0,43	0,36	0,01	0,36	PVLQG	FRNE	GIDR	OB
	IIc	AlaRS	Q10A14	0,92	0,01	0,01	0,08	RWPSS	IRA	GMGFER	tRNA SAD, DHHA1
			B9FSH5	0,11	0,96	0,02	0,03	ILPSS	IRT	GMGLER	tRNA SAD, DHHA1
			B9FWG9	0,01	0,00	0,01	0,97	RCSPS	Ipv	GLNFEQ	MINDY_domain deubiquitinase,DUF544
		PheRS	Q2QPD4	0,89	0,00	0,02	0,11	FNKFDDL	YRRD	GLGLER	FDX-ACB
			Q338I8	0,01	0,02	0,00	0,97	FTEMTNN	FRNE	GLSLER	
			Q6AT40	0,04	0,00	0,04	0,92	YMEVLTF	VTL	CLLK	B3 B4 B5, DNA-bd_dom_put
		GlyRS	Q5VRH2	0,02	0,23	0,96	0,03	CAVMQCSN	IRPD	GLER	

Physcomitrella patens				Predotar				Key motifs			Additional domain
				Mito	Plastid	ER	Elsewhere	HIGH motif	KMSKS motif		
Class I	Ia	ArgRS	A9S0A2	0,01	0,00	0,02	0,97	HVGH	RFRTR		DALR_1
		CysRS	A9U4A4	0,01	0,00	0,06	0,93	HIGH	KMSKS		DALR_2
		IleRS	A9RSP9	0,01	0,00	0,00	0,99	HIGH	KMSKS		Znf_FPG/IleRS
			A9TB11	0,01	0,01	0,00	0,99	HYGH	KMSKR		
		LeuRS	A9SG46	0,01	0,00	0,00	0,99	HVGH	KMSKS		
			A9RZT5	0,01	0,00	0,00	0,99	HLGH	KMSKS		
		MetRS	A9SCS9	0,48	0,91	0,01	0,04	HMGS	KMGKS		
			A9SRC9	0,01	0,00	0,01	0,98	HLGN	KFSKS		NA-bd_OB-fold
		ValRS	A9SJ82	0,01	0,00	0,00	0,99	HIGH	KMSKS		
			A9SEL3	0,01	0,00	0,00	0,99	HMGH	KMSKS		
	Ib	GlnRS	A9S2J6	0,01	0,00	0,00	0,99	HIGH	VMSKR		
		GluRS	A9S8X8	0,01	0,02	0,07	0,91	HIGH	LLSKR		GST_C
			A9S3X4	0,01	0,00	0,02	0,97	HIGH	LLSKR		GST_C
			A9SYW3	0,01	0,00	0,01	0,98	HVGG	KLSKR		
	Ic	TrpRS	A9REV4	0,07	0,03	0,09	0,82	HLGN	KMSKS		
			A9RQM2	0,01	0,00	0,00	0,99	HLGH	KMSAS		
		TyrRS	A9RKG5	0,01	0,00	0,00	0,99	HIAQ	KMSKS		
			A9RBJ1	0,28	0,97	0,01	0,02	HLGN	KFGKS		
Class II								Motif 1	Motif 2	Motif 3	
	IIa	GlyRS	A9TSY1	0,22	0,00	0,02	0,76	WEIRCPCV	YRNE	CADR	
		HisRS	A9S048	0,01	0,00	0,00	0,99	FEEVDFPLV	WRYER	GFGFGxxV	
			A9U0U2	0,01	0,00	0,00	0,99	FEEVDFPMV	WRYER	GFGFGxxV	
		ProRS	A9RFZ3	0,01	0,00	0,00	0,99	TENTYFPL	VRWE	GLTTR	ProRS-C_1
			A9TQ78	0,46	0,07	0,00	0,50	HSNMYFPQ	HRWE	GMSTR	ProRS-C_1
		ThrRS	A9U0D5	0,01	0,01	0,00	0,99	EEVVSPNM	HRNE	GSVER	TGS-like
			A9SJF5	0,01	0,00	0,00	0,99	ELLYTPHI	YRYE	GSIER	tRNA_SAD
			A9RXJ2	0,01	0,03	0,00	0,97	ELLYTPHI	YRYE	GSIER	tRNA_SAD
		SerRS	A9TH40	0,01	0,00	0,00	0,99	IPMSTPDL	FRTE	ACAxxR	tRNA-bd_arm
			A9RJR8	0,01	0,00	0,02	0,98	TLLQTPFF	FRKE	LTAxxR	tRNA-bd_arm
			A9SZI6	0,01	0,00	0,02	0,98	TLLQTPFF	FRKE	LTAxxR	tRNA-bd_arm
	IIb	AsnRS	A9RTZ3	0,01	0,00	0,00	0,99	PIITA	FRAE	GFER	NA-bd_OB-fold
			A9S8S1	0,31	0,00	0,03	0,67	PIITT	FRAE	GFER	NA-bd_OB-fold
			A9S8S6	0,31	0,00	0,03	0,67	PIITT	FRAE	GFER	NA-bd_OB-fold
		AspRS	A9TI29	0,80	0,04	0,00	0,19	PKLIA	FRAE	GLER	NA-bd_OB-fold
			XP_001753365.1	0,01	0,00	0,00	0,99	PILTK	FRDE	GLDR	NA-bd_OB-fold, GAD
		LysRS	A9SRN4	0,01	0,01	0,00	0,99	PILTK	FRDE	GLDR	NA-bd_OB-fold
			A9SZV3	0,01	0,00	0,00	0,99	PMMNM	FRNE	GIDR	NA-bd_OB-fold
	IIc	AlaRS	A9RG15	0,11	0,04	0,09	0,77	VLPSS	VRT	GLGLER	tRNA_SAD,DHHA1
			A9T4K7	0,01	0,01	0,00	0,98	FVPSS	IRA	GLGFER	tRNA_SAD,DHHA1
		PheRS	A9SET8	0,01	0,00	0,03	0,97	FTEVVTW	VRL	GMLK	B3/B4_tRNA-bd,B5,DNA-bd_dom_put
			A9T5K9	0,05	0,05	0,06	0,85	FKKFDNL	YRRD	GLGLER	Ferrodoxin-fold anticodon-binding domain
			A9TXE7	0,01	0,09	0,00	0,90	FEEMPxNN	FRNE	GLSLER	
		GlyRS	A9TR89	0,56	0,50	0,03	0,21	CVVMQSSN	VRPD	GLER	

STATEMENT OF THE CONTRIBUTIONS OF JOINTLY AUTHORED MANUSCRIPT

Statement of Authorship

Title of Paper	A Rice Glutamyl-tRNA Synthetase Modulates Early Anther Cell Division and Patterning
Publication Status	☐ Published ■ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	This manuscript was accepted for publication in *Plant Physiology* on 20 April 2018.

Principal Author

Name of Principal Author (Candidate)	Xiujuan Yang		
Contribution to the Paper	X.Y. designed and completed experiments including cytological observations, *in situ* hybridisation, RT-qPCR, vector constructions, protein expression and purification, 20% enzyme kinetics analysis, Y2H, BiFC, pull-down, metabolomics analysis, ROS measurement and chemicals treatments etc. X.Y. analysed data, organised figures and drafted the manuscript.		
Overall percentage (%)	60%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	26/04/2018

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Gang Li		
Contribution to the Paper	G.L. performed 80% enzyme kinetics analysis, and contributed to discussion and editing.		
Signature		Date	26/04/2018

Name of Co-Author	Yuesheng Tian		
Contribution to the Paper	Y.T. performed map-based cloning and complementary experiment.		
Signature		Date	26/04/2018

Statement of Authorship

Name of Co-Author	Yu Song		
Contribution to the Paper	Y.S. fixed and embedded samples for *in situ* hybridisation analysis.		
Signature		Date	26/04/2018

Name of Co-Author	Wanqi Liang		
Contribution to the Paper	W.L. contributed to conceiving the project, provided supervision and were financially responsible for the work.		
Signature		Date	26/04/2018

Name of Co-Author	Dabing Zhang		
Contribution to the Paper	D.Z. contributed to conceiving the project, revised manuscript, provided supervision and were financially responsible for the work. D.Z. acted as corresponding author.		
Signature		Date	26/04/2018

CHAPTER 3

A Rice Glutamyl-tRNA Synthetase Modulates Early Anther Cell Division and Patterning

This chapter was published in *Plant Physiology* in June 2018.
www.plantphysiol.org/cgi/doi/10.1104/pp.18.00110

Check for updates

A Rice Glutamyl-tRNA Synthetase Modulates Early Anther Cell Division and Patterning[1][OPEN]

Xiujuan Yang,[a,b,2] Gang Li,[a,b,2] Yuesheng Tian,[b] Yu Song,[b] Wanqi Liang,[b] and Dabing Zhang[a,b,3]

[a]School of Agriculture, Food, and Wine, University of Adelaide, Adelaide, South Australia 5064, Australia
[b]Joint International Research Laboratory of Metabolic and Developmental Sciences, Shanghai Jiao Tong University-University of Adelaide Joint Centre for Agriculture and Health, State Key Laboratory of Hybrid Rice, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China

Aminoacyl-tRNA synthetases (aaRSs) have housekeeping roles in protein synthesis, but little is known about how these aaRSs are involved in organ development. Here, we report that a rice (*Oryza sativa*) glutamyl-tRNA synthetase (OsERS1) maintains proper somatic cell organization and limits the overproliferation of male germ cells during early anther development. The expression of *OsERS1* is specifically detectable in meristematic layer 2-derived cells of the early anther, and *osers1* anthers exhibit overproliferation and disorganization of layer 2-derived cells, producing fused lobes and extra germ cells in early anthers. The conserved biochemical function of OsERS1 in ligating glutamate to $tRNA^{Glu}$ is enhanced by its cofactor aaRS OsARC. Furthermore, metabolomics profiling revealed that OsERS1 is an important node for multiple metabolic pathways, indicated by the accumulation of amino acids and tricarboxylic acid cycle components in *osers1* anthers. Notably, the anther defects of the *osers1* mutant are causally associated with the abnormal accumulation of hydrogen peroxide, which can reconstitute the *osers1* phenotype when applied to wild-type anthers. Collectively, these findings demonstrate how aaRSs affect male organ development in plants, likely through protein synthesis, metabolic homeostasis, and redox status.

Reproductive cells in flowering plants link the dominant diploid sporophytic generation and the short haploid gametophytic generation. Male germ cells are produced within the sporophytic tissue called the anther (Ma, 2005). The anther primordium that forms from the floral meristem usually contains three layers: L1 to L3. L1 and L3 cells subsequently differentiate into the epidermis and connective tissue, respectively, whereas the L2 layer forms germ cells in the center of anther lobes and three surrounding supportive somatic cell layers. The early stages of anther development (stages 1–5) determine the formation of anther cell layers. At stage 2, the L2 layer undergoes rapid cell division and generates L2-derived (L2-d) cells. At stage 3, the central archesporial cells show a distinctive enlarged shape surrounded by the neighboring primary parietal cells (PPCs). At stage 4, PPCs form endothecium and secondary parietal cells (SPCs) via asymmetric cell division, and archesporial cells differentiate into sporogenous cells. Following this, the sporogenous cells continue to divide for subsequent meiosis and the SPC layer symmetrically divides to form the middle layer and tapetum at stage 5 (Ma, 2005; Wilson and Zhang, 2009; Zhang and Yang, 2014; Walbot and Egger, 2016).

One critical event in premeiotic anther development is rapid periclinal cell division and strict cell lineage specification. However, the underlying mechanism controlling this process remains largely unknown. Previous studies revealed the critical functions of a few regulators, such as receptor-like protein kinases, transcription factors, redox status, glycoprotein, and hormones, in specifying cell lineage and the fate of germ and somatic cells in Arabidopsis (*Arabidopsis thaliana*), rice (*Oryza sativa*), and maize (*Zea mays*; Wilson and Zhang, 2009; Zhang and Yang, 2014; Walbot and Egger, 2016).

Glutaredoxins (GRXs) are small oxidoreductases that modulate various cellular events and responses to oxidative stress (Holmgren, 2000). Rice MICROSPORELESS1 (MIL1) and maize Male Sterile Converted Anther1 (MSCA1) are homologs to two Arabidopsis GRXs, ROXY1 and ROXY2. These GRXs have been shown to regulate abaxial-adaxial anther lobe formation and the differentiation of microsporocytes, suggesting that GRXs trigger archesporial fate determination by controlling redox status. Further studies show that GRXs interact with TGA (TGACGTCA cis-element-binding protein) transcription factors to transcriptionally regulate targets (Chaubal et al., 2003; Xing and Zachgo, 2008; Hong et al., 2012b; Kelliher and Walbot, 2012; Yang et al., 2016).

[1]The contributions of D.Z. and G.L. were supported by Startup Funding from the School of Agriculture, Food, and Wine, University of Adelaide, and the contributions of X.Y. by a Beacon Scholarship, University of Adelaide. This work was supported by the National Key Research and Development Program of China (2016YFD0100804); the Innovative Research Team, Ministry of Education, and 111 Project (grant no. B14016); and Australia-China Science and Research Fund Joint Research Centre grant ACSRF48187.

[2]These authors contributed equally to the article.
[3]Address correspondence to dabing.zhang@adelaide.edu.au.
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Dabing Zhang (dabing.zhang@adelaide.edu.au).

D.Z. and W.L. conceived and designed the project; X.Y., G.L., Y.T., and Y.S. performed the experiments; X.Y., G.L., W.L., and D.Z. performed the analysis; W.L. and D.Z. provided supervision and were financially responsible for the work; X.Y., G.L., and D.Z. wrote the article.

[OPEN]Articles can be viewed without a subscription.
www.plantphysiol.org/cgi/doi/10.1104/pp.18.00110

Downloaded from on August 5, 2018 - Published by www.plantphysiol.org
Copyright © 2018 American Society of Plant Biologists. All rights reserved.

Recent studies have shown the function of cell surface-localized leucine-rich repeat receptor-like kinases (LRR-RLKs) and their ligands, such as rice MULTIPLE SPOROCYTE1 (MSP1) and TAPETAL DETERMINANT1-LIKE A (OsTDL1A)/MIL2, in determining early anther cell fate. These are orthologous to Arabidopsis EXCESS MICROSPOROCYTES1/EXTRA SPOROGENOUS CELLS and TAPETAL DETERMINANT1, respectively, and function in specifying anther cell identity and controlling cell numbers of tapetum and microsporocytes (Canales et al., 2002; Zhao et al., 2002, 2008; Nonomura et al., 2003; Yang et al., 2003; Hong et al., 2012a; Zhang and Yang, 2014). MSP1 and OsTDL1A were suggested to exert their effects in early anther cell differentiation via modulating redox status in rice (Yang et al., 2016).

In maize, early-stage anthers also require hypoxia to maintain proper division and determine cell fate. The maize GRX MSCA1 may maintain growth-generated hypoxic conditions to induce archesporial cells. Also, archesporial cells might silence mitochondria to lower the concentration of reactive oxygen species (ROS) and maintain reducing conditions (Kelliher and Walbot, 2012).

Aminoacyl-tRNA synthetases (aaRSs) are found in all protein-synthesizing organisms and can catalyze the ligation of amino acids and their cognate tRNA, an essential step for translation and amino acid dynamics (Ibba and Söll, 2000; Yamakawa and Hakata, 2010). Studies in mammals showed noncanonical and surprising roles of aaRSs in amino acid signaling, transcriptional control, and antiapoptosis (Guo et al., 2010). As housekeeping proteins, aaRSs are indispensable for plant development, but their detailed function in plant growth remains largely unclear. In *Nicotiana benthamiana*, silencing of *Glutamyl-tRNA Synthetase* (*GluRS*) and *Seryl-tRNA Synthetase* causes a severe yellowing phenotype in leaves due to significant reduction in the number of organelles (Kim et al., 2005). Mutation of a rice ValRS results in a reduced level of chlorophyll and abnormal chloroplast ribosome biogenesis (Wang et al., 2016). Among the 45 aaRSs in Arabidopsis, 21 are required for ovule development and embryogenesis (Berg et al., 2005). In addition, an Arabidopsis *Cysteinyl-tRNA Synthetase* (*CysRS*) expressed in the central cell of the female gametophyte may specify the fate of adjacent cells (Kägi et al., 2010). Noncanonical and surprising roles of aaRSs have been uncovered, such as amino acid signaling, transcriptional control, and antiapoptosis. Furthermore, eight members of aaRSs were found to exist in mammalian cytoplasm as a multitRNA synthetases complex (MSC) whose roles and mechanisms of formation are still unclear (Guo and Schimmel, 2013). However, the detailed role of aaRSs in plant organs and development is limited. Here, we report the critical function of the cytosol- and mitochondria-localized GluRS, OsERS1, in early anther cell proliferation and organization by affecting protein synthesis, amino acid-derived metabolism, and cellular redox status.

RESULTS

OsERS1 Limits L2-d Cell Proliferation and Maintains Anther Cell Organization

To explore the mechanisms underlying anther development in the cereal model plant rice, we identified a new male-sterile mutant, *oryza sativa glutamyl-tRNA synthetase* (*osers1*), because of the mutation in *Glutamyl-tRNA Synthetase1* (see below). The *osers1* mutant displayed no obvious phenotypic changes from the wild type in vegetative growth, inflorescence architecture, female reproductive organ, or seed shape (Supplemental Fig. S1, A and B). However, during the male reproductive stage, the mutant developed defective anthers that were pale and had fused lobes, resulting in the production of ~20% viable pollen grains and an ~10% seed-setting rate (Supplemental Fig. S1, C–F). The *osers1* mutant showed a more tightly woven texture in the anther wall, and the pollen grains were shrunken and displayed rough, lumpy extine, in contrast to the wild type, which had a well-arranged cuticle on the exterior of the anther wall and mature pollen grains with smooth extine (Supplemental Fig. S1, G–N). All F1 plants of reciprocal crosses between *osers1* and the wild type were normal, and the F2 offspring had an ~3:1 segregation of normal versus abnormal anthers (116:42; $\chi^2 = 0.58$, $P > 0.05$), suggesting that the *osers1* mutants contain a unifactorial and recessive mutation.

From stage 3 to 5 of anther development, ~40% of *osers1* anthers had adaxial fused lobes, which was distinct from the butterfly-shaped four lobes in wild-type anthers (Fig. 1, A and B; Supplemental Fig. S2, A–F and I). At stage 3, transverse section analysis showed that each wild-type anther formed ~6.3 archesporial cells ($n = 11$) enringed by a single layer of PPCs, while each *osers1* anther produced ~10 archesporial-like cells ($n = 12$) enringed by PPCs organized in an irregular mosaic pattern (Fig. 1, C, D, I, and K). At stage 5, the wild-type archesporial cells divided further into ~11 sporogenous cells ($n = 15$) surrounded by three circular inner layers derived from PPCs (Fig. 1, E, G, I, and K). However, each *osers1* anther produced ~20 sporogenous-like cells ($n = 13$), and ~50% of the anthers had disordered wall layers with irregular boundaries (Fig. 1, F, H, I, and K; Supplemental Fig. S2I). Notably, the total number of somatic and germ cells derived from the L2 layer in the *osers1* mutant was increased during stages 3 to 5 (Fig. 1J). Some sporogenous-like cells were ectopically embedded in subepidermal layers (Fig. 1I; Supplemental Fig. S2, G and H).

The increased numbers of anther cells in the *osers1* mutant prompted us to compare cell division activity between the wild type and the *osers1* mutant. In situ hybridization showed stronger expression signals for *HISTONE4* (*H4*), a cell-division marker for the S-phase of mitosis, and *CYCLIN-DEPEDENT KINASE B2-1* (*CDKB2;1*), a marker of the G2/M phase transition (Umeda et al., 1999), in both somatic and germ cells of *osers1* lobes, compared with the wild type (Fig. 1, L–Q).

Downloaded from on August 5, 2018 - Published by www.plantphysiol.org
Copyright © 2018 American Society of Plant Biologists. All rights reserved.

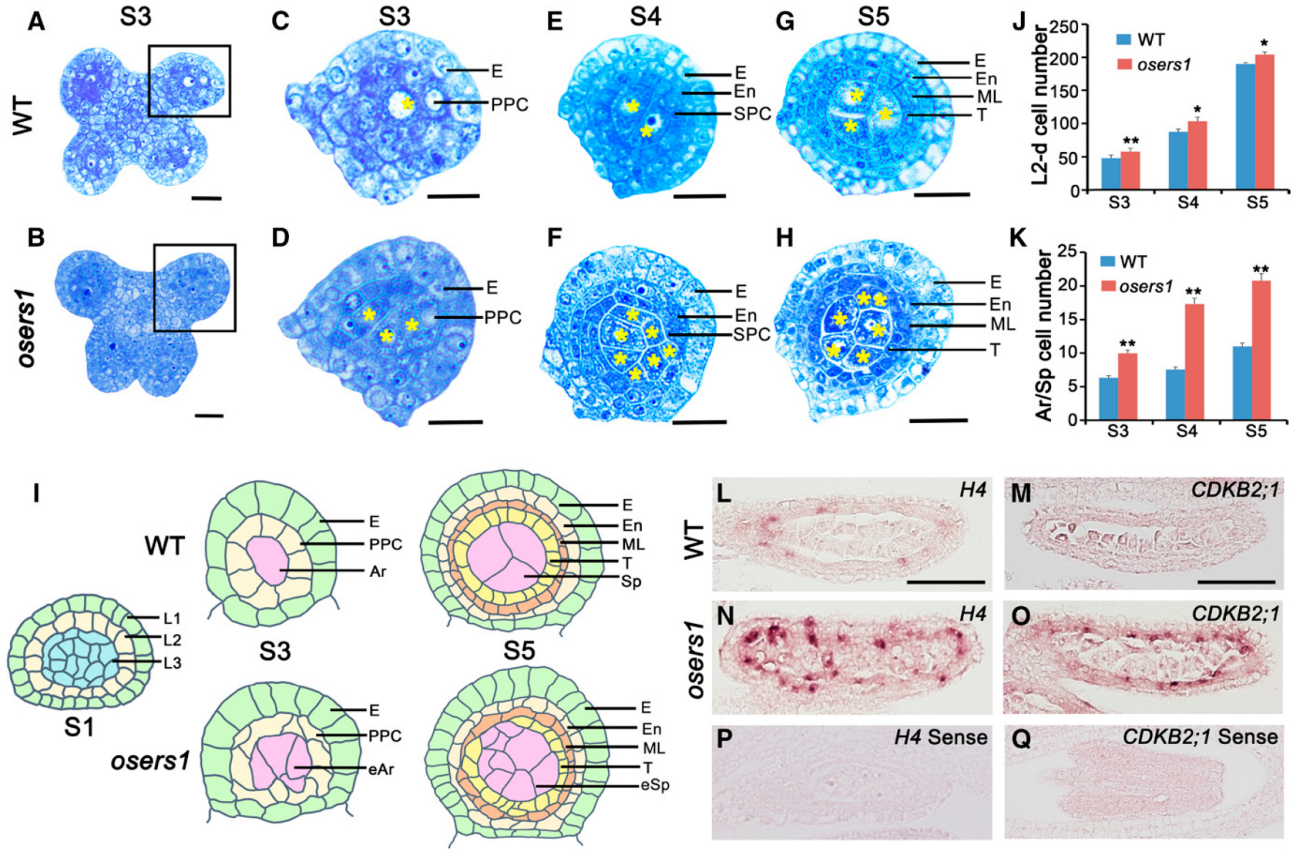


Figure 1. OsERS1 is essential for cell proliferation and organization in early rice anther development. A to H, Cytological comparison of anther development between the wild type (WT) and the *osers1* mutant. S3 to S5 indicate anther developmental stages. A and B show anthers in stage 3, C and D show magnifications of the boxed regions in A and B, respectively, E and F show anther lobes in stage 4, and G and H show anther lobes in stage 5. Yellow asterisks indicate archesporial cells at S3 and sporogenous cells at S4 to S5. Bars = 15 μm. I, Diagrams of anther development from stages 1 to 5 in the wild type and the *osers1* mutant. Disorganization of somatic cell layers and excess archesporial/sporogenous-like cells are shown in the mutant. J and K, Quantification of cell overproliferation in *osers1* anther. J, Average L2-d cell counts of anther cross sections. K, Average archesporial/sporogenous cell counts of the anther cross sections. All error bars represent SD. S3 wild type, $n = 11$; S4 wild type, $n = 13$; S5 wild type, $n = 15$; S3 *osers1*, $n = 12$; S4 *osers1*, $n = 14$; S5 *osers1*, $n = 13$. Student's *t* test: *, $P < 0.05$ and **, $P < 0.01$. Ar, Archesporial cell; E, epidermis; eAr, excess archesporial-like cell; En, endothecium; eSp, excess sporogenous-like cell; ML, middle layer; Sp, sporogenous cell; T, tapetum. L to Q, In situ hybridization of cell division marker genes in stage 4 anthers. Bars = 50 μm.

Reverse transcription-quantitative PCR (RT-qPCR) analysis also showed increased expression of *H4* and *CDKB2;1* in *osers1* anthers (Supplemental Fig. S3B). Due to the lack of marker genes for cell lineage during early anther development, we chose *MSP1* to examine the relationship of the *MSP1* pathway and *OsERS1*. In situ hybridization showed similar distribution patterns of *MSP1* in both wild-type and *osers1* anthers: that is, preferential accumulation on anther wall cells rather than in central sporogenous cells (Supplemental Fig. S3A). RT-qPCR showed no significant difference in the expression of *MSP1*, as well as in that of two other genes involved in anther development, *MIL1* and *MIL2*, between wild-type and *osers1* anthers (Supplemental Fig. S3B). Unlike *osers1*, mutations of *MSP1*, *MIL2*, and *MIL1* led to misidentified anther wall cell layers and complete male sterility (Nonomura et al., 2003; Hong et al., 2012a,b). These results indicate that the cell properties are still maintained in *osers1* anthers, and OsERS1 is strongly implicated in cell proliferation and organization in the rice anther.

OsERS1 Encodes a Glutamyl-tRNA Synthetase

Using map-based cloning combined with the whole-genome sequencing-based MutMap method (Abe et al., 2012; Lü et al., 2015), we identified the mutated *OsERS1* gene as having an insertion of A after A^{539} in the LOC_Os10g22380 coding region (also known as Os10g0369000; see "Materials and Methods"); this resulted in a premature stop codon in the open reading frame (Fig. 2, A and B). The identity of *OsERS1* was

Downloaded from on August 5, 2018 - Published by www.plantphysiol.org
Copyright © 2018 American Society of Plant Biologists. All rights reserved.

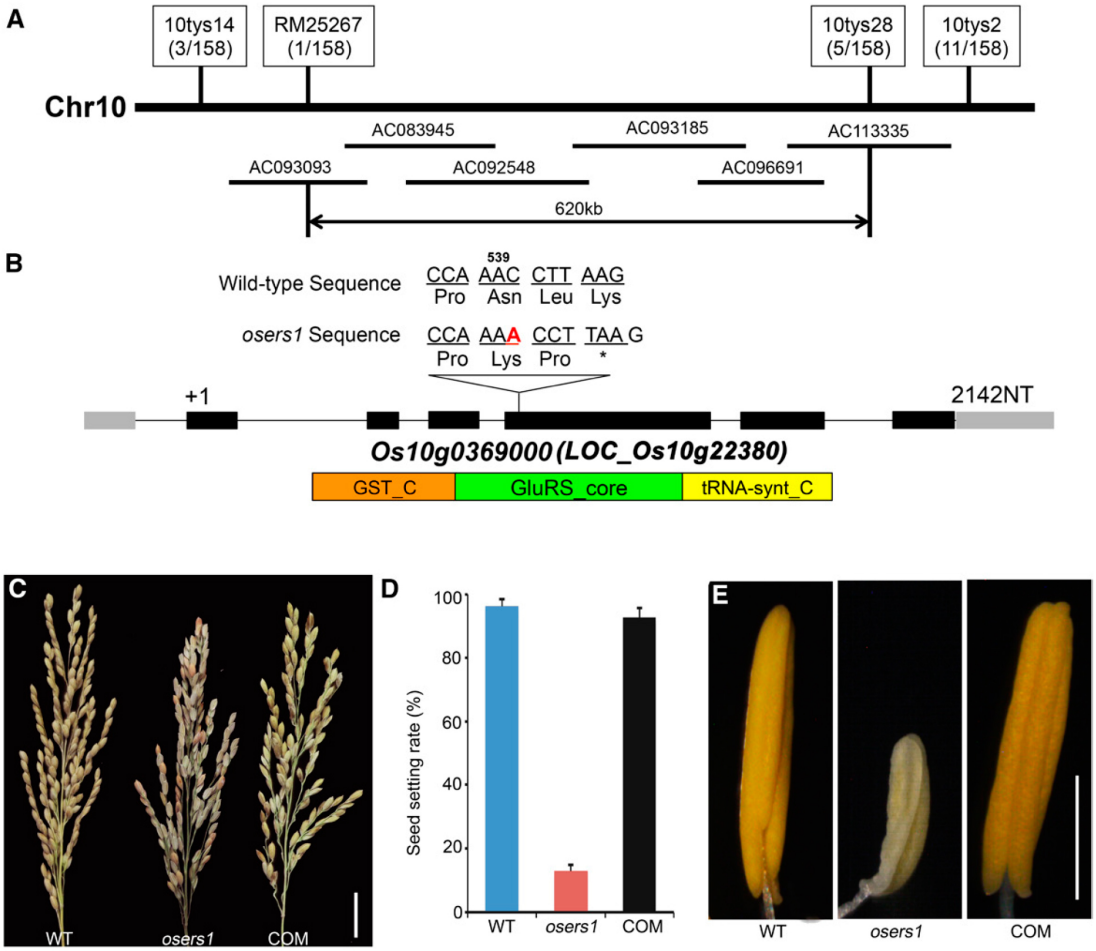


Figure 2. Molecular identification of *OsERS1*. A, Fine-mapping of *OsERS1* to chromosome 10. The molecular markers used for mapping are indicated. Numbers in parentheses indicate the number of recombinants. AC numbers are rice genomic DNA accession numbers of bacterial artificial chromosome clones. B, Schematic representation of *OsERS1* and functional domains. The red A indicates an insertion after A^{539} in *oser1*. NT, Nucleotides. Black rectangles represent exons. The asterisk means a stop codon. GST_C, Glutathione *S*-transferase C-terminal domain; GluRS_core, catalytic core domain of glutamyl-tRNA synthetase; tRNA-synt_C, tRNA synthetases class I anticodon binding domain. C to E, Complementation of the *osers1* mutant. C, Inflorescences with mature seeds. D, Average seed-setting rates in panicles (n = 20 for each set). Error bars represent SD. E, Mature anthers. COM, Complementation; WT, wild type. Bars = 2 cm in C and 1 mm in E.

confirmed by the complementary assay, in which the male-sterile phenotype in the *osers1* mutant was rescued by the expression of a genomic fragment that contained the wild-type *OsERS1* gene (Fig. 2, C–E).

OsERS1 encodes a putative glutamyl-tRNA synthetase (GluRS) that contains an N-terminal GST_C (C terminus of GST) domain, a glutamyl-tRNA synthetase catalytic core, and a C-terminal tRNA-synt_C domain (Fig. 2B). RT-qPCR showed that *OsERS1* was expressed in the wild type at low levels in mature organs but at high levels in young tissues, including seedlings and young inflorescences (Fig. 3A). In the anther, *OsERS1* transcript levels were relatively high during stages 3 to 6 and decreased in later stages (Fig. 3, B and C). In situ hybridization detected *OsERS1* mRNA in presumptive archesporial cells and neighboring L2-d cells before stage 3. During stages 3 to 5, when anther cells undergo rapid cell division, *OsERS1* mRNA was detected in PPC and archesporial cells of stage 3, the SPC and sporogenous cells of stage 4, and the tapetum, middle layer, and sporogenous cells of stage 5. The expression of *OsERS1* in germ cells decreased significantly after stage 5 (Fig. 3C; Supplemental Fig. S4A). The rice genome has two *OsERS1* homologs, LOC_Os01g16520 and LOC_Os02g02860, but these homologs had low expression in the anther and did not appear to have altered expression in the *osers1* mutant, suggesting that the phenotype of the *osers1* mutant was caused by aberrant *OsERS1* in the anther as opposed to the two homologs (Supplemental Fig. S4B). Considering that protein synthesis is essential in the cytosol, chloroplast, and mitochondria, we performed protein

Downloaded from on August 5, 2018 - Published by www.plantphysiol.org
Copyright © 2018 American Society of Plant Biologists. All rights reserved.

localization studies in rice protoplast cells and onion (*Allium cepa*) cells by coexpressing OsERS1-eGFP and the mitochondrial marker ScCOX4-mCherry (carried by vector *CD3-991*; Nelson et al., 2007). Our results showed that OsERS1 is localized mainly in the cytoplasm and is partially shared by the mitochondria (Fig. 3D; Supplemental Fig. S4C).

Phylogenetic analysis of 40 OsERS1 homologs from 21 species from diverse kingdoms clearly divided these proteins into two clades: the eukaryote-specific GluRSs (clade I) and the prokaryote-originated GluRSs (clade II; Fig. 3E; Supplemental Table S1). Intriguingly, most eukaryotes contain both types. Clade II GluRS contained only the catalytic domain, while clade I GluRSs showed more complex structures. The progressive appearance of novel domains, such as GST_C, tRNA-synt_C, repeats of the WHEP domain (initially discovered in human tryptophanyl-tRNA synthetase [TrpRS; W], histidyl-tRNA synthetase [H], and GluRS-prolyl-tRNA synthetase [ProRS; EP]), and chimeric GluRS-ProRS in clade I, indicated multiple domain-recruiting events during evolution (Gao et al., 2010). In grass crops such as rice and maize, there seemed to be a duplication of GST_C containing GluRS, which we did not observe in other species (Fig. 3E), indicating that GluRS might have an expanded function for evolutionary adaption in higher organisms.

OsERS1 Functions as a Ligase between Glu and tRNA^{Glu}

The synthesis of glutamyl-tRNA is a two-step reaction: activation of Glu by ATP and glutamylation of tRNA^{Glu}. In the catalytic domain, GluRS has a core active site (RFAPE in OsERS1) that binds Glu and two key motifs called HIGH and KMSKS (LLSKR in OsERS1) that are involved in ATP binding and tRNA binding, respectively (Freist et al., 1997; Ibba and Söll, 2000). Sequence alignment showed that these three motifs are highly conserved in the 40 GluRS homologs (Fig. 4A; Supplemental Fig. S5).

To confirm the function of OsERS1 in glutamyl-tRNA synthesis, we performed an ATP-pyrophosphate (PPi) exchange assay, which showed that OsERS1 has PPi exchange activity and that its reaction rate is proportional to the concentration of tRNA^{Glu} (Fig. 4B). An aminoacylation assay revealed that OsERS1 has the catalytic activity of aminoacylation of tRNA^{Glu} with Glu (Fig. 4, C and D), yielding a K_m value (tRNA^{Glu} as substrate) of 2.2 μM and a K_{cat}/K_m value of 0.977 s^{-1} μM^{-1} (Fig. 4E). OsERS1's affinity for tRNA^{Glu} is similar to that of the *Thermus thermophilus* GluRS but lower than that of *Rattus* spp. liver and *Escherichia coli* (Freist et al., 1997). Also, OsERS1's K_m value is lower than that of wheat (*Triticum aestivum*) GluRSs, possibly because the latter function as dimers (Thomes et al., 1983). However, it is still not clear whether GluRS can form dimers in rice.

Mutations of the three conserved motifs of OsERS1s, Arg-209 and Glu-213 in the active site region, His-219 and His-222 in the HIGH motif, and Ser-442 and Lys-443 in the KMSKS motif (Fig. 4A; Supplemental Fig. S5), caused significantly reduced glutamylation activity and enzyme efficiency (K_{cat}/K_m; Fig. 4, C–E). These results suggest that OsERS1 has catalytic activity in tRNA^{Glu} glutamylation and that the RFAPE, HIGH, and KMSKS motifs are required for OsERS1's enzymatic function.

The Cofactor OsARC Stimulates the Catalytic Activity of OsERS1

The GST_C domain of GluRS is noncatalytic because it lacks thioredoxin, but it was shown to link the enzyme to protein complexes (Dixon et al., 2002). In yeast and mammals, GluRS/GluRS-ProRS interacts with the aaRS cofactor Arc1p/p18 to form a complex and modulate enzyme activities (Galani et al., 2001; Graindorge et al., 2005; Gao et al., 2010). Rice has a single homolog of yeast Arc1p, OsARC (encoded by LOC_Os01g60660), which contains a heterogeneous GST_C domain (Fig. 5A). To examine whether OsERS1 can form a complex with OsARC, we conducted yeast two-hybrid and in vitro pull-down assays; these confirmed that OsERS1 can interact with OsARC and that the interaction requires the GST_C domains (Fig. 5, B and C;. Bimolecular fluorescence complementation (BiFC) showed that only the cells expressing full-length OsERS1-nYFP and OsARC-cYFP displayed yellow fluorescence signals (Fig. 5D), verifying this interaction in vivo. OsERS1-eGFP and OsARC-mCherry were highly colocalized when coexpressed in onion cells (Fig. 5E), supporting our hypothesis that OsERS1 and OsARC can form a complex. Moreover, compared with that of OsERS1 alone or OsERS1-ΔOsARC (truncated OsARC protein without the GST_C domain; Fig. 5, F and G), the OsERS1-OsARC complex had ~2.5 times the ATP-PPi exchange rate of Glu and ~1.4 times the tRNA^{Glu} aminoacylation efficiency in the plateau stage, indicating an enhancement effect promoted by OsARC. In addition, the GluRS and cofactor pairs are conserved components of MSCs in various organisms (Laporte et al., 2014). However, the structure and composition of plant MSC have not been dissected.

Mutation of *OsERS1* Leads to Metabolic Alterations in Anthers

The substrate for OsERS1, Glu, occupies a central position that links key metabolic pathways (Newsholme et al., 2003). Loss of function of *OsERS1* may cause an accumulation of free Glu that is not used for protein synthesis in the anther. To test this possibility, we conducted nontargeted metabolomics profiling in wild-type and *osers1* anthers. A total of 297 metabolites were identified, covering all major primary and several secondary metabolism pathways (Supplemental Table S2). Consistent with the deficiency in ligating Glu with tRNA^{Glu}, the *osers1* mutant had 2.44 times the level of Glu compared with the wild type. With the exception of glutathione in both reduced (GSH) and

Downloaded from on August 5, 2018 - Published by www.plantphysiol.org
Copyright © 2018 American Society of Plant Biologists. All rights reserved.

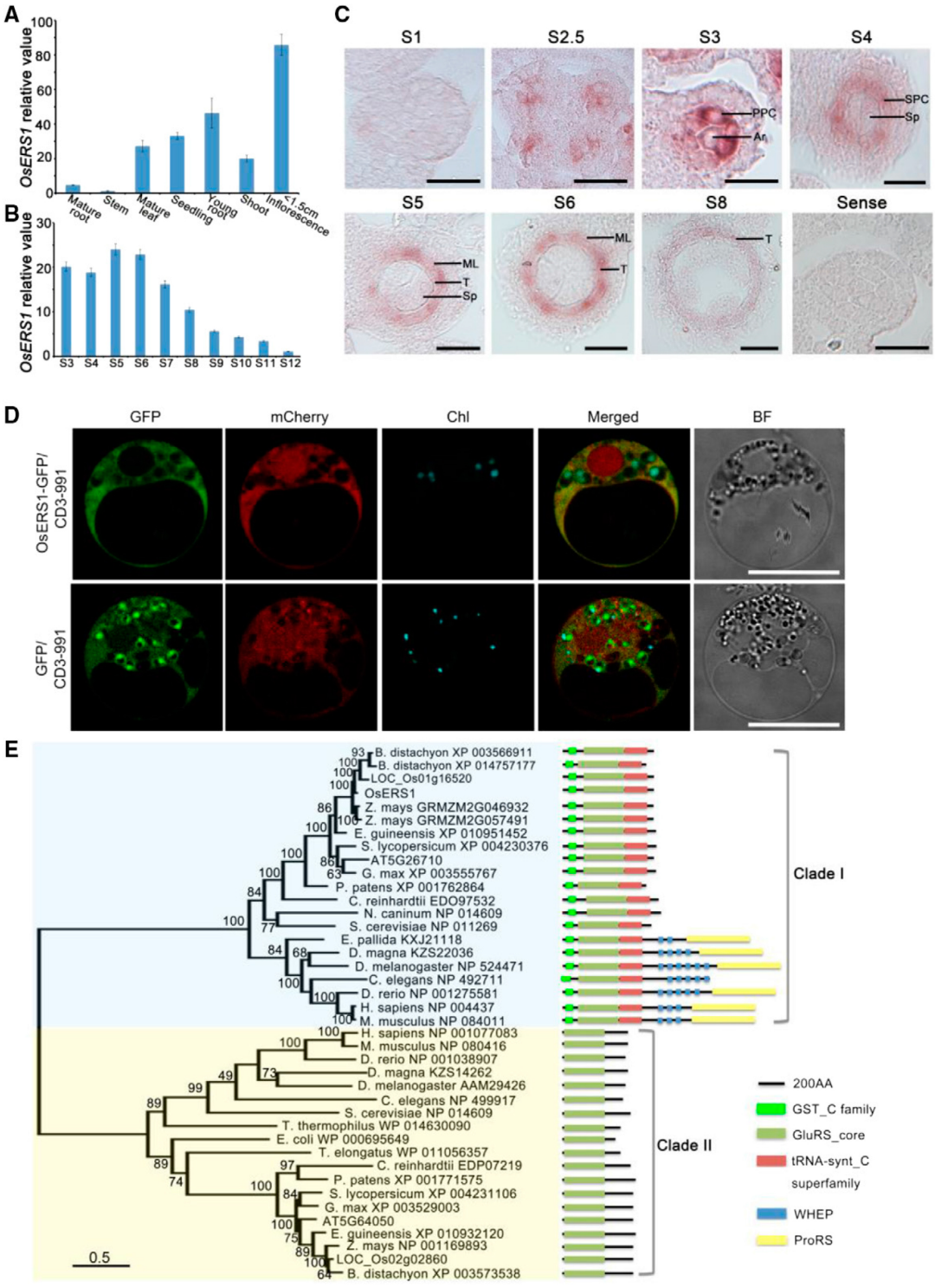
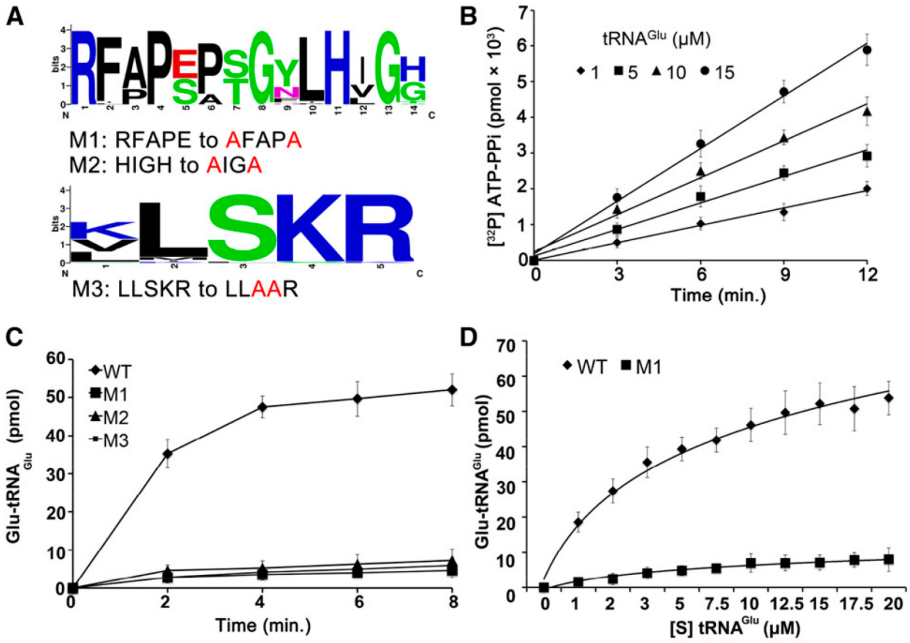


Figure 3. Expression pattern and phylogenetic analysis of OsERS1. A and B, RT-qPCR data showing relative expression levels of OsERS1 in various tissues (A) and anthers at various stages (B). For B, total RNA was extracted from whole rice flowers before stage 6 and from anthers after stage 7. All error bars represent SD (n = 3). C, In situ hybridization of OsERS1 in the wild-type anther at different stages. A sense probe was used as the control. Ar, Archesporial cell; ML, middle layer; Sp, sporogenous cell; T, tapetum. Bars = 15 μm. D, Subcellular localization of OsERS1 in rice protoplasts. GFP was used as a control. Vector CD3-991 containing the mitochondria-specific protein ScCOX4 fused with mCherry served as the organelle marker. Chl, Chloroplast; BF, bright field. Bars = 10 μm. E, Neighbor-joining tree of OsERS1 and its homologs showing the progressive recruitment of

Downloaded from on August 5, 2018 - Published by www.plantphysiol.org
Copyright © 2018 American Society of Plant Biologists. All rights reserved.



E

OsERS1	Km (μM)	Kcat (s^{-1})	Kcat/Km (s^{-1}·μM^{-1})
WT	2.20±0.48	2.15±0.11	0.977
M1	5.16±1.53	0.37±0.03	0.072
M2	5.62±1.72	0.36±0.04	0.064
M3	3.21±1.88	0.46±0.08	0.143

Figure 4. OsERS1 functions as a ligase of Glu and $tRNA^{Glu}$. A, Conserved sites in the catalytic domain of OsERS1 used for site-directed mutagenesis. The figure was generated from 40 homologs of OsERS1 by WebLogo (http://weblogo.berkeley.edu). B, ATP-PPi exchange rate in the presence of increased concentrations of $tRNA^{Glu}$. C, $tRNA^{Glu}$ aminoacylation by wild-type (WT) and mutant (M1–M3) OsERS1. D, Kinetics of $tRNA^{Glu}$ aminoacylation by wild-type OsERS1 and OsERS1 mutant M1 under increased $tRNA^{Glu}$ concentrations. Wild-type OsERS1, $R^2 = 0.99$; OsERS1 mutant M1, $R^2 = 0.97$. E, Kinetic constants determined by $tRNA^{Glu}$ charging assays. Kinetics of the OsERS1 and OsERS1 M1 to M3 aminoacylation reaction was measured using varied concentrations of $tRNA^{Glu}$, and reactions were conducted for 10 min under standard conditions. Reciprocal initial velocity was plotted against the reciprocal $tRNA^{Glu}$ concentration according to Lineweaver-Burk to calculate the corresponding K_m values. All error bars represent SD ($n = 3$).

oxidized (GSSG) status, the levels of Glu family members, including Gln, α-ketoglutarate (α-KG), His, and Arg, were increased in the *osers1* mutant. One Glu derivative, α-KG, also is an intermediate of the tricarboxylic acid cycle and links the metabolism of amino acids and carbohydrates (Jeong et al., 2013). As a result, most metabolites of the tricarboxylic acid cycle, such as oxaloacetate (OAA) and malate, were increased in the *osers1* mutant (Fig. 6A; Supplemental Fig. S6B; Supplemental Table S2). This is also consistent with changes in the expression of genes involved in Glu metabolism and the tricarboxylic acid cycle, such as genes encoding Gln synthetases (*GS2* and *GS3*), Glu dehydrogenase (*GDH1*), and oxoglutarate dehydrogenase (*ODH2*; Supplemental Table S3). Besides the Glu family, the Leu, Ser, and Asp families derived from pyruvate, phosphoglycerate, and OAA also were increased in the *osers1* mutant. We propose that increased levels of amino acids in the mutant may have been caused by the accumulation of Glu, OAA, and pyruvate, which are important precursors for amino acid synthesis (Fig. 6; Supplemental Fig. S6A; Supplemental Table S2).

In addition, we detected the expression level of 12 *aaRS* genes highly expressed in reproductive organs. Most of the *aaRS* genes showed subtle expression changes between the wild type and the *osers1* mutant,

Figure 3. (*Continued.*)
novel functional domains. AA, Amino acids; GST_C, C terminus of GST; GluRS_core, glutamyl-tRNA synthetase catalytic core; tRNA-synt_C, tRNA synthetase class I anticodon binding domain; ProRS, prolyl-tRNA synthetase.

Downloaded from on August 5, 2018 - Published by www.plantphysiol.org
Copyright © 2018 American Society of Plant Biologists. All rights reserved.

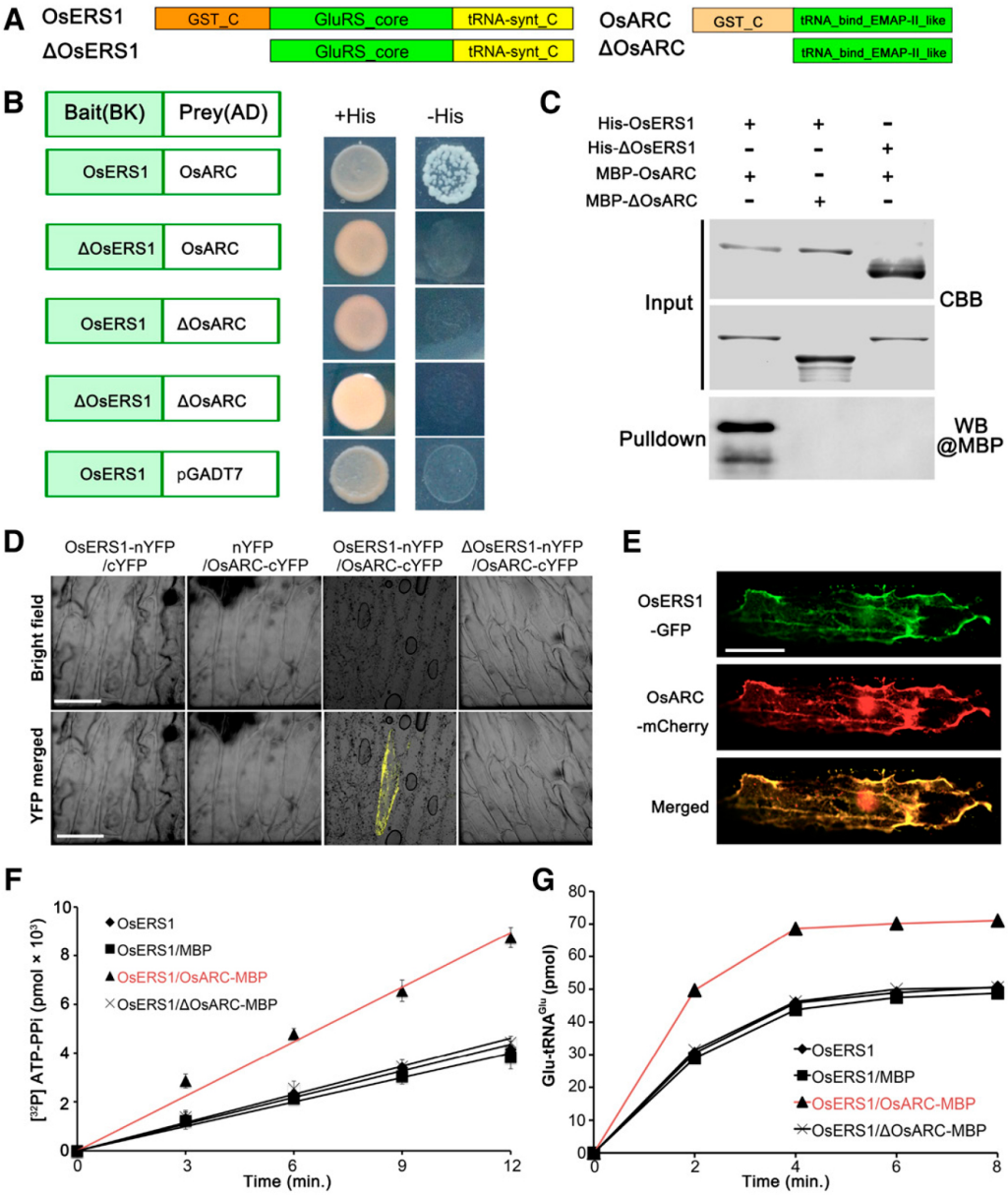


Figure 5. The complex of OsERS1 and OsARC enhances OsERS1's catalytic activity. A, Domain structures of OsERS1 and OsARC and their truncations. B, OsERS1 and OsARC interact in yeast. Interaction is verified by growth of the yeast strain in selective medium. C, Pull-down assays between His-OsERS1 and MBP-OsARC in vitro. The precipitated MBP-OsARC was detected by α-MBP antibody. MBP-tagged ΔOsAR and His-tagged ΔOsERS1 were used as a control to show the interaction in a GST_C domain-dependent manner. CBB, Coomassie Brilliant Blue; WB, western blot. D, BiFC assays in onion epidermal cells showing the interaction between OsERS1 and OsARC. cYFP, C-terminal split of YFP; nYFP, N-terminal split of YFP. Bars = 50 μm. E, Colocalization of OsERS1-eGFP and OsARC-mCherry in onion epidermal cells. Bar = 50 μm. F, Influence of the OsERS1-OsARC complex on ATP-PPi exchange. A total of 15 μM $tRNA^{Glu}$ was used as the substrate. G, Influence of the OsERS1-OsARC complex on $tRNA^{Glu}$ aminoacylation. All error bars represent SD (n = 3).

except for *CysRS1* and *TrpRS2*, which had lower expression in the *osers1* mutant (Supplemental Fig. S7), implying that the amino acid changes in the mutant are caused mainly by the loss of function of *OsERS1*. By contrast, we observed decreased levels of glycolysis intermediates, including Glc, Glc-6-P, and Fru-6-P, in the *osers1* mutant (Fig. 6; Supplemental Fig. S6B); this indicates a higher consumption of these compounds in the *osers1* mutant. Decreased Glc supply and less lactate production suggest that pyruvate was likely used by the tricarboxylic acid cycle and for amino acid biosynthesis. Increased expression of PPP genes and the

Downloaded from on August 5, 2018 - Published by www.plantphysiol.org
Copyright © 2018 American Society of Plant Biologists. All rights reserved.

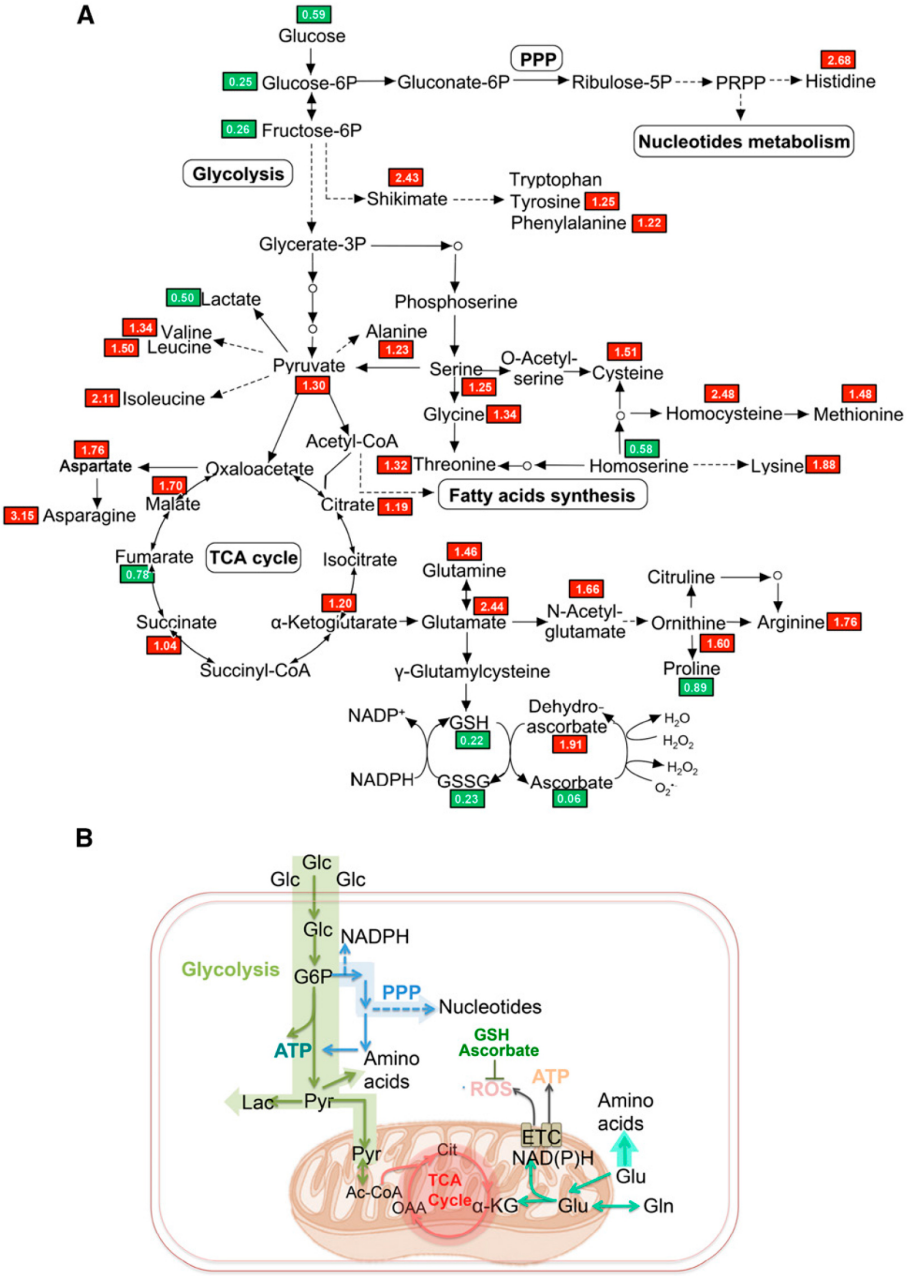


Figure 6. Metabolic rewiring in the *osers1* mutant. A, The identified important metabolites are mapped onto a simplified metabolic network. Red and green blocks indicate metabolites with higher and lower average levels, respectively, in the *osers1* mutant compared with the wild type. All values are averages of three independent trials. B, Putative metabolism in *osers1* anthers. Major metabolic pathways are shown, including glycolysis, pentose phosphate pathway (PPP), tricarboxylic acid (TCA) cycle, and potential ROS producers and scavengers. Ac-CoA, Acetyl-coenzyme A; Cit, citrate; ETC, electron transport chain; α-KG, α-ketoglutarate; Lac, lactate; Pyr, pyruvate.

accumulation of purine and pyrimidine indicate that more intermediates of glycolysis were used to support nucleotide biosynthesis in *osers1* anthers (Supplemental Fig. S6C; Supplemental Tables S2 and S3). Metabolic analysis showed abnormal fatty acid synthesis in the *osers1* mutant, such as cutin and wax, which is consistent with the scanning electron microscopic analysis of *osers1* anther wall and pollen grain (Supplemental Fig. S1, G–N, and S6D; Supplemental Table S2).

Overall, the *osers1* anthers show more active amino acid and carbohydrate metabolism with aerobic characteristics, such as an active tricarboxylic acid cycle

Downloaded from on August 5, 2018 - Published by www.plantphysiol.org
Copyright © 2018 American Society of Plant Biologists. All rights reserved.

and less lactate production. The increased Glu family amino acids, as well as those derived from OAA and pyruvate, contribute to the significant changes in biosynthesis of the amino acid pathway in *osers1* anthers (Fig. 6B; Supplemental Fig. S6A; Supplemental Table S2). Notably, Glu may provide mitochondrial anaplerosis to the tricarboxylic acid cycle to stimulate mitochondrial activities, which produces ROS as an inevitable by-product. Consistent with this, decreased cellular antioxidants (GSH/GSSG and ascorbate) suggest a disturbed redox status in *osers1* anthers (Fig. 6B). Corresponding with the activated cell division in *osers1* anthers, more carbohydrates are consumed in glycolysis to generate energy and provide more intermediates for the PPP, which enhances the synthesis of nucleotides and macromolecules. These results reveal that the metabolic homeostasis is disturbed in *osers1* anthers and that OsERS1 plays key roles in amino acid-related metabolism.

Altered Redox Status Affects Early Anther Development

Global metabolomics profiling suggested that ROS might be produced by mitochondrial activities in the *osers1* mutant. To verify this, we first measured hydrogen peroxide (H_2O_2), a stable ROS molecule implicated in cell proliferation and differentiation (Møller et al., 2007; Tsukagoshi et al., 2010). Fluorometric quantitation assays showed that H_2O_2 content in the *osers1* mutant at stages 4 to 5 was ~900 pmol mg^{-1}, twice that in wild-type anthers at the same stage (Fig. 7A, left). Also, the precursor of H_2O_2, superoxide radical, was increased slightly in the *osers1* mutant (Fig. 7A, right), which contributed to the redox status change in the *osers1* mutant. In addition, 3,3′-diaminobenzidine (DAB) staining of wild-type anthers showed that the H_2O_2 signal was preferentially present in middle layer and endothecium cells until stage 5 (Fig. 7, B–E). However, *osers1* anthers displayed detectable H_2O_2 in the anther primordium before stage 3 (Fig. 7F) and strong H_2O_2 signals in anther cells at stages 4 to 5 (Fig. 7, G–I). In later stages, even though the distribution of H_2O_2 was similar in both wild-type and *osers1* anther wall layers, stronger H_2O_2 signals were seen in *osers1* anther wall layers (Supplemental Fig. S8). The lower expression levels of genes encoding ascorbate peroxidases in the mutant (Supplemental Table S3) may have contributed to the accumulation of ROS.

To test whether the increased level of H_2O_2 in the *osers1* mutant causes the abnormal division and patterning of L2-d cells, we injected 100 μM H_2O_2 and 1 mM of the ROS-removal reagent potassium iodide (KI) into young inflorescences of the wild type and the *osers1* mutant (Supplemental Fig. S9). Transverse section observations revealed that each stage 4 wild-type anther treated with water showed a slight increase in the number of sporogenous-like cells, from 7.5 (n = 16) to 9.2 (n = 12), which may have been triggered by wounding (Fig. 7, J, K, and S). Remarkably, the number of L2-d cells in each wild-type anther treated with H_2O_2 increased from 109 (n = 16) in untreated anthers to 135.4 (n = 18), and sporogenous-like cell numbers increased from 7.5 to 15.4. Therefore, treatment with 100 μM H_2O_2 in the wild type led to overproliferation of L2-d cells, ectopic sporogenous-like cells, and disordered parietal cell layers, phenocopying the defects of the *osers1* mutant (Fig. 7, L, R, and S). Wild-type anthers treated with KI also showed increases in the number of L2-d cells to 132.7 and sporogenous-like cells to 16.8 (n = 17), while parietal cells were organized differently compared with those treated with H_2O_2 (Fig. 7, M, R, and S). Compared with that of the untreated anther (139.3 L2-d cells and 18.5 sporogenous-like cells; n = 14), in the *osers1* mutant, each anther treated with water showed a slight decrease of L2-d cell numbers to 131.5 and sporogenous-like cell numbers to 14.5 (n = 13; Fig. 7, N, O, R, and S). However, each *osers1* anther treated with H_2O_2 and KI showed a reduced number of L2-d cells and sporogenous-like cells (H_2O_2, 102.9 and 12.7, respectively [n = 16]; KI, 110.5 and 14.5, respectively [n = 15]). Aside from changes in cell numbers, we observed highly vacuolated cells in impaired *osers1* anthers (Fig. 7, P–S). We reasoned that elevations of ROS by H_2O_2 injection and neutralization of ROS by KI injection both disturbed ROS homeostasis, which may contribute to repressed cell division, and that the *osers1* anthers appear vulnerable to changes in redox status. Together, these results suggest that changed redox status disturbs early cell division in rice anthers.

DISCUSSION

Early anther development is a complex event characterized by rapid cell division and fate specification to support the development of functional anther layers that are properly organized within the anther (Zhang and Yang, 2014; Walbot and Egger, 2016). In this study, we revealed that OsERS1, a glutamyl-tRNA synthetase, functions in early anther cell division and patterning, likely through maintaining physiological homeostasis of amino acids and carbohydrate metabolism as well as redox status (Fig. 8).

Our genetic and cytological data demonstrate that loss of function of *OsERS1* results in fused anther lobes with supernumerary L2-d cells and disarranged anther wall cell layers. In *osers1* anthers, L2-d cell division was accelerated, but the L1 layer remained normal and formed a rigid enclosure, which aggravated the disarranged cell layers and created abaxial fused lobes. In spite of its putative housekeeping function, expression of *OsERS1* shows spatial-temporal characteristics rather than a constitutive pattern. In the anther, *OsERS1* transcripts are initially expressed in cells of presumptive lobes at the four primordial corners. Then, *OsERS1* mRNA is detected in both parietal and germ cells, preferentially in those undergoing periclinal division or newly formed cell layers (Fig. 3C). After stage 5, the germ cells develop into microspores via meiosis,

Downloaded from on August 5, 2018 - Published by www.plantphysiol.org
Copyright © 2018 American Society of Plant Biologists. All rights reserved.

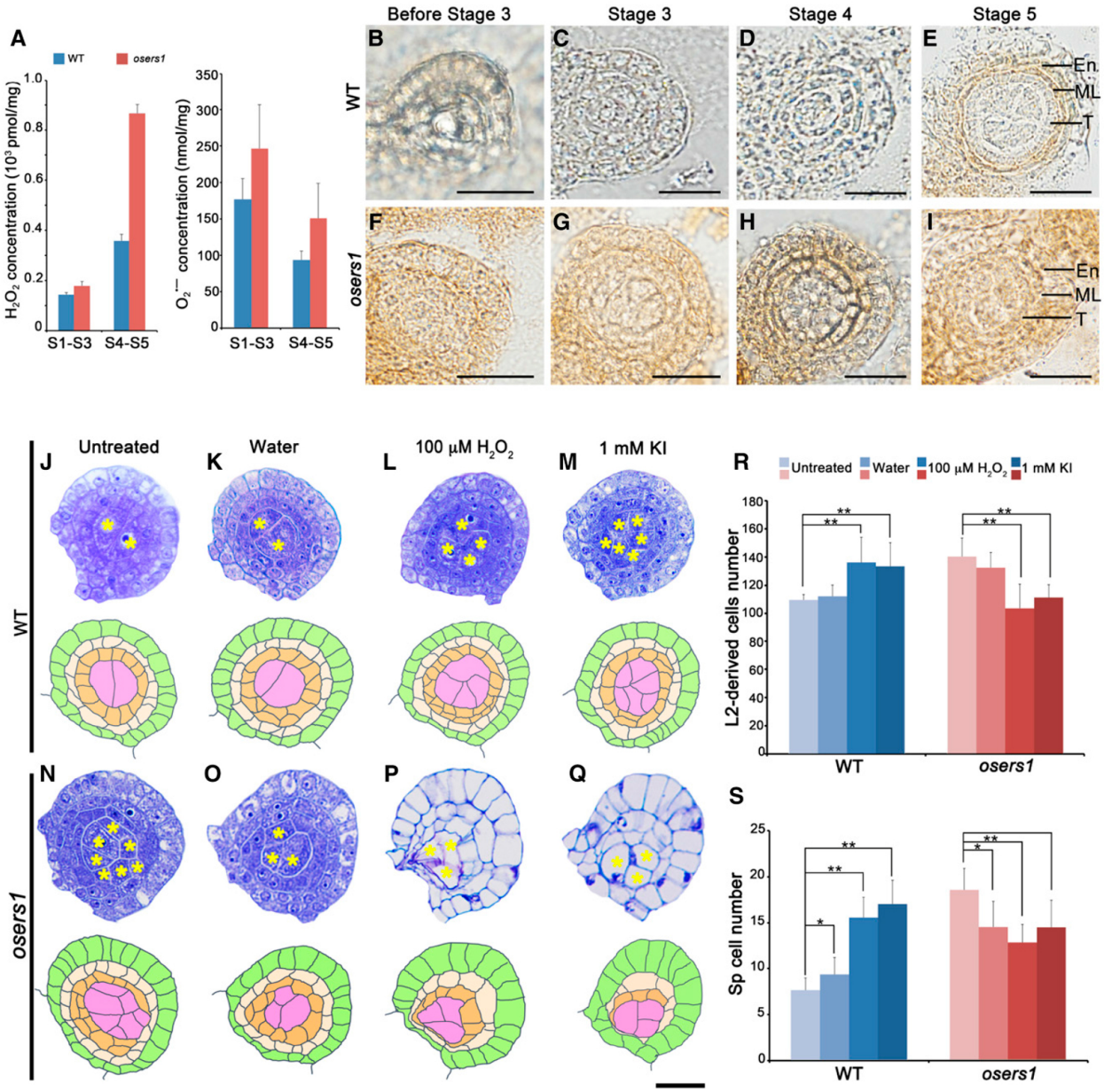


Figure 7. Redox status controls L2-d cell division and organization. A, Measurement of H_2O_2 and superoxide radicals ($O_2^{\bullet-}$) in young spikelets. All error bars represent SD (n = 3). S1–S3, Stages 1 to 3; S4–S5, stages 4 to 5. B to I, DAB staining analysis for H_2O_2 in anther lobes. En, Endothecium; ML, middle layer; T, tapetum; WT, wild type. Bars = 15 μm. J to Q, Cross sections (top rows) and diagrams (bottom rows) of wild-type and *osers1* anther lobes of stage 4 under treatments. Yellow asterisks indicate sporogenous cells. Bar = 15 μm. R and S, Average counts of L2-d cells (R) and sporogenous(-like) cells (S) in cross sections from anthers of stage 4 under the indicated treatments. All error bars represent SD (wild type, n = 16; wild type-water, n = 12; wild type-H_2O_2, n = 18; wild type-KI, n = 17; *osers1*, n = 14; *osers1*-water, n = 13; *osers1*-H_2O_2, n = 16; *osers1*-KI, n = 15). Student's t test: *, $P < 0.05$ and **, $P < 0.01$.

and the abundance of *OsERS1* mRNA in germ cells decreased to a very low level, indicating that *OsERS1* mainly affects early cell division and patterning. This expression pattern is in agreement with the phenotype of increased L2-d cells and disarranged cell layers in the *osers1* mutant.

Phenotypically, mutations of the LRR-RLK-ligand pair *MSP1-MIL2* or the GRX *MIL1* lead to misidentified anther wall cell layers and complete male sterility (Nonomura et al., 2003; Zhao et al., 2008; Hong et al., 2012a,b), while the *osers1* anther still produces a low proportion of viable pollen grains (Supplemental Fig. S1).

Downloaded from on August 5, 2018 - Published by www.plantphysiol.org
Copyright © 2018 American Society of Plant Biologists. All rights reserved.

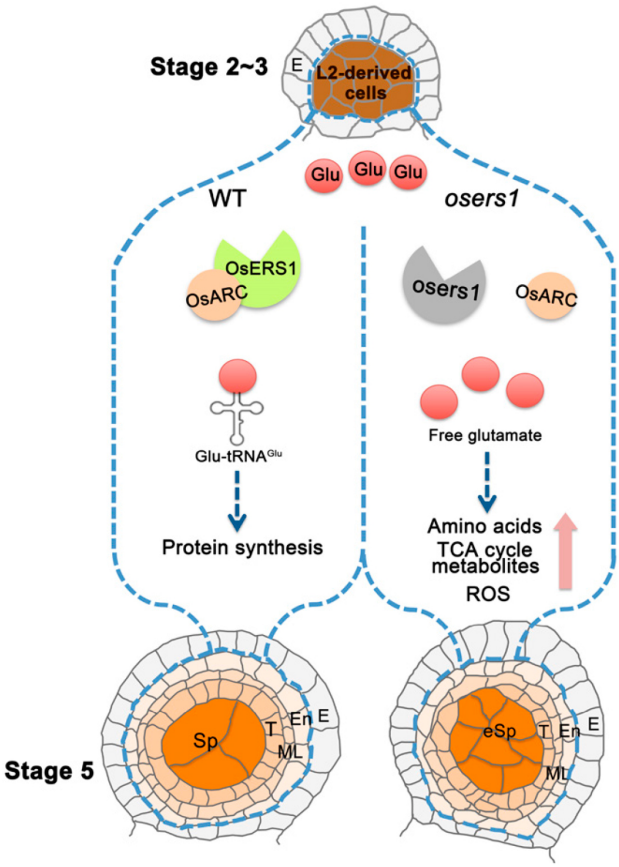


Figure 8. Working model for OsERS1 function in rice anthers. In rice anthers, L2-d cells generate somatic endothecium, middle layer, tapetum, and germinal sporogenous cells. *OsERS1* is expressed specifically in L2-d cells at high levels during stage 2 to 3 and later in cells differentiated from L2-d cells. In the wild type (WT), OsERS1 combined with OsARC are required for protein synthesis and to maintain L2-d cell number and patterning during early anther development. Loss of function of *OsERS1* resulted in the accumulation of free Glu and changes in amino acids and their derived metabolic pathways. Subsequently, ROS produced by metabolic activities affected redox status and cell behavior, leading to changes in anther development. E, Epidermis; En, endothecium; eSp, excess sporogenous-like cell; ML, middle layer; Sp, sporogenous cell; T, tapetum; TCA, tricarboxylic acid.

Anther defects in the *osers1* mutant occur earlier than those in the reported rice mutants with abnormal early anther development, such as *gamyb-4*, *udt1* (*undeveloped tapetum1*), and *tip2* (*tdr interacting protein2*; Jung et al., 2005; Liu et al., 2010; Fu et al., 2014). We showed that there are no significant changes in the expression of *MSP1*, *MIL2*, and *MIL1* in the *osers1* mutant compared with the wild type, and previous studies showed that *OsERS1* has no obviously altered expression in *msp1*, *mil2*, *gamyb-4*, *udt1*, or *tip2* mutants (Jung et al., 2005; Aya et al., 2009; Liu et al., 2010; Fu et al., 2014; Yang et al., 2016). Therefore, we propose that *OsERS1* plays an independent role with known regulators in early anther development. Although loss of function of *OsERS1* causes abnormal anther morphogenesis and decreased pollen vitality, the mutant still can form a small number of seeds with normal shape, suggesting that fertilization is not completely disturbed. However, we cannot exclude the possibility that female gametophyte development also is affected.

We confirmed the in vitro activity of OsERS1 in catalyzing the ligation between Glu and $tRNA^{Glu}$. OsERS1 is dominantly localized in the cytosol and partially shared by mitochondria. The product, glutamyl-tRNA, is an essential substrate for protein synthesis. Compared with the prokaryotic (*E. coli*) GluRS (Freist et al., 1997), OsERS1 possesses a relatively lower affinity to $tRNA^{Glu}$. One explanation is that the additional GST_C domain may change space conformation and reduce substrate affinity. However, the interaction between OsERS1 and its cofactor OsARC overcomes this disadvantage by enhancing the activity of OsERS1 for synthesizing glutamyl-tRNA (Fig. 5). In the early anther, loss of function of *OsERS1* may impair translational processes in the cells that normally accumulate high *OsERS1*. However, the *osers1* anthers maintain partial fertility and produce viable pollen, suggesting that

Downloaded from on August 5, 2018 - Published by www.plantphysiol.org
Copyright © 2018 American Society of Plant Biologists. All rights reserved.

other components may remedy the absence of OsERS1 to maintain protein synthesis in the anther.

Metabolic status is associated with tumorigenesis, stem cell renewal, pluripotency maintenance, and differentiation (Cairns et al., 2011; Schuster et al., 2014; Sperber et al., 2015; Ito and Ito, 2016). In plants, glycolysis and mitochondrial metabolism activate the Arabidopsis root meristem via Glc-target of rapamycin signaling (Xiong et al., 2013). In maize, the anther primordium is more likely glycolysis dependent, as the enclosed niche provides hypoxic conditions for archesporial cells (Kelliher and Walbot, 2012). We speculate that the hypoxia status specifying cell identity and cell division must be well controlled by the specific metabolic pathway within anther cells. In this study, we showed that, as an important metabolic regulatory node, OsERS1 is involved in intracellular Glu dynamics, and Glu could refuel the tricarboxylic acid cycle that occupies a central position in multiple metabolic pathways. Within this context, the Arabidopsis cytosolic GluRS is an important target of 14-3-3 proteins in carbon/nitrogen metabolism and sugar-sensing pathways. GluRS is degraded when sugar starvation occurs in order to release Glu and support cell survival (Cotelle et al., 2000; Coruzzi and Zhou, 2001). In this work, we showed that the mutation of *OsERS1* disturbs the homeostasis of the metabolic network, as evidenced by increased amounts of amino acids, tricarboxylic acid components, and active aerobic metabolism. We speculate that such a metabolic shift acts as a signal to remove the limitation on cell proliferation in the anther lobe. Consistent with this, some mutants of 21 Arabidopsis aaRSs showed defective female gametogenesis and embryo development but no obvious defect during male gametogenesis except for in the *ova9* mutant, which had a mutation in *GlnRS* (Berg et al., 2005). However, how these specific aaRSs are involved in reproductive development remains unclear. Although we observed no obvious change of other *aaRS* genes in the *osers1* mutant (Supplemental Fig. S7), we cannot rule out the function of other *aaRSs* in male reproduction.

ROS exert dose-dependent impacts on cells. A lower or moderate basal level of ROS is essential for cell proliferation and differentiation, but higher levels of ROS are cytotoxic, causing cell damage and death (Sena and Chandel, 2012; Schieber and Chandel, 2014; Mittler, 2017). In rice anthers, the ROS molecule H_2O_2 is initially accumulated in the middle layer at stage 5; this may act as a signal that triggers cell differentiation (Zhang and Yang, 2014; Yang et al., 2016). In this study, we observed the production of H_2O_2 in *osers1* anthers as early as the formation of archesporial cells, which may cause the overproliferation of anther cells and disorganized cell layers (Fig. 7). In support, disruption of the redox status in wild-type anthers by injecting low concentrations of H_2O_2 resulted in a mimic of the phenotype of the *osers1* mutant (Fig. 7). Similarly, in the shoot apical meristem of Arabidopsis, elevating H_2O_2 dramatically enlarges meristem size by increasing cell numbers in the peripheral zone (Zeng et al., 2017). In maize, treatment with a high concentration of H_2O_2 (1 mM) on the anther inhibits cell division in both somatic and germ cells (Kelliher and Walbot, 2012), demonstrating multifaceted regulatory aspects of ROS in cell division. The study in maize also indicates that the anther primordium may use glycolysis as a major energy source, rather than through Glc oxidation in mitochondria to avoid ROS pressure and maintain genomic integrity (Kelliher and Walbot, 2012). In this study, we demonstrated that increased ROS causes the overproliferation of early anther cells in the mutant, suggesting that reductive redox status is essential to maintain normal cell division and organization during rice anther development.

In summary, we discuss the role of a conserved glutamyl-tRNA synthetase, OsERS1, in limiting anther cell division and maintaining cell patterning (Fig. 8). The catalytic activity of OsERS1 in ligating Glu to $tRNA^{Glu}$ can be enhanced by its cofactor OsARC. Loss of function of *OsERS1* causes changes in a variety of metabolites, such as the accumulation of amino acids and tricarboxylic acid cycle intermediates, and leads to aerobic metabolic features in the mutant. Together with the H_2O_2 assays, we concluded that OsERS1 maintains normal early anther cell division and patterning, likely through affecting protein synthesis, amino acid homeostasis, and redox status in rice. Our study reveals how the aaRSs control the cell division and organ development derived from meristematic cells in plants.

MATERIALS AND METHODS

Materials, Plant Growth, and Phenotypic Analysis

The *osers1* mutant was isolated from a rice (*Oryza sativa* ssp. *japonica*) mutant library in the cv 9522 background (Chen et al., 2006). Rice plants were grown in the artificial atmospheric chambers on the Waite Campus, University of Adelaide, and in a paddy field at Shanghai Jiao Tong University. Plants and flowers were photographed with a Nikon E995 digital camera and a Motic K400 dissecting microscope. Photography of plant materials and observation of anthers were done as described by Li et al. (2006) and imaged with a Nikon Ni-E optical microscope and a Philips XL20 scanning electron microscope. Cell counts were determined based on cross sections from more than 10 anthers, and statistical analysis included all four lobes of whole anthers.

Map-Based Cloning and Complementation of the Mutant

For fine-mapping, we identified 158 mutants from an F2 population that was generated from a cross between the *osers1* mutant (*japonica*) and cv 9311 (*indica*). To map the *OsERS1* locus, we used bulked segregation analysis and developed insertion-deletion molecular markers based on the sequence differences between rice cultivars, described in the National Center for Biotechnology Information. The *OsERS1* gene was initially located to chromosome 10 between two insertion-deletion molecular markers, 10tys14 and 10tys2, and finer mapping narrowed down the gene to the region between RM25267 and 10tys28. Mapping combined with genome sequencing was performed using the previously published method (Abe et al., 2012; Lü et al., 2015).

For functional complementation, an 8.3-kb genomic DNA fragment containing the entire *OsERS1* coding region, a 3.2-kb upstream sequence, and a 1.7-kb downstream sequence was amplified from a BAC clone and subcloned through *Eco*RI and *Bst*EII into the binary vector *pCAMBIA1301* (*p1301*), which carries a hygromycin resistance marker, to generate *p1301-OsERS1*. We induced calli from young panicles of the homogenous *osers1* mutant and transformed the calli with *Agrobacterium tumefaciens* EHA105, which carried the *p1301-OsERS1*

Downloaded from on August 5, 2018 - Published by www.plantphysiol.org
Copyright © 2018 American Society of Plant Biologists. All rights reserved.

plasmid. At least 25 independent transgenic lines were obtained after PCR genotyping using the primers listed in Supplemental Table S4.

RT-qPCR

We extracted total RNA from plants using the TRIZOL reagent (Invitrogen), following the manufacturer's instructions. cDNA was synthesized from 4 μg of total RNA using a reverse transcriptase kit (Bio-Rad). SYBR Green Mix (Kapa) was used for RT-qPCR on a QuantStudio Flex 6 (Life Technologies) machine. The level of *ACTIN2* mRNA was used as the internal control. For expression analysis of *OsERS1*, *aaRSs*, metabolism-related genes, or anther developmental genes, we denatured the DNA at 95°C for 10 min, followed by 42 to 45 cycles at 95°C for 10 s, 56°C for 15 s, and 72°C for 15 s. The RT-qPCR results shown are from representatives of experiments repeated at least three times with similar results. Primer sequences are listed in Supplemental Table S4.

RNA in Situ Hybridization

We amplified *OsERS1*-, *MSP1*-, *H4*-, and *CDKB2;1*-specific fragments of the exon region (*OsERS1*, 70–391 bp; *MSP1*, 3,462–3,841 bp; *H4*, full length; *CDKB2;1*, 415–739 bp) by PCR using specific primers fused with the *T7* promoter (sequences are listed in Supplemental Table S4). Digoxigenin-labeled antisense and sense probes were transcribed using an in vitro transcription kit (Roche) according to the manufacturer's instructions. In situ hybridization was performed according to a published protocol (Zeng et al., 2017).

Subcellular Localization

35Spro::OsERS1-eGFP (full-length *OsERS1* cDNA fused with e*GFP*) was generated in *pCAMBIA1301* and expressed in rice protoplasts and onion (*Allium cepa*) epidermal cells using PEG-mediated transfections and a particle delivery system, respectively. The construction of the organelle marker (*CD3-991*) was described by Nelson et al. (2007). For rice protoplast cells, stem tissues of 14-d-old seedlings were processed following the protocol described by Zhang et al. (2011). For onion epidermal cells, the onion was cut into 2- × 2-cm squares and placed on a plate for 4 h with Murashige and Skoog medium containing 40 g L^{-1} D-mannitol. For transformation, 5 μg of each plasmid DNA was coated with 1-μm gold particles (Bio-Rad) and delivered into onion epidermal cells using a PDS-1000/He particle delivery system (Bio-Rad). Fluorescence microscopy was performed using a Leica TCS SP5 laser scanning confocal microscope. Fluorescence signals for mCherry (excitation, 587 nm; emission, 580–630 nm; 15% power), eGFP (excitation, 483 nm; emission, 500–530 nm; 15% power), and chloroplasts (excitation, 504 nm; emission, 650–750 nm; 15% power) were detected with a 20× (dry) objective for onion cells and a 63× (oil) objective for protoplast cells.

Phylogenetic Analysis and Motif Alignment

We constructed a phylogenetic tree with aligned full-length sequences of homologs of OsERS1. MEGA (version 7.0) and the neighbor-joining methods were used with a p-distance model and pairwise deletion and bootstrap (Gap Ktuple, 2; length penalty, 5; Window, 4; Diagonals, 4; 1,000 replicates). The maximum parsimony method of MEGA also was used to support the neighbor-joining tree using the default parameter. Amino acid sequences from regions 199 to 228 and 434 to 454 in OsERS1 were used for motif alignment by MEGA.

Recombinant Protein Purification

Wild-type and mutant *OsERS1* were cloned into the *pET32(a)* vector for expression in the BL21 (DE3) *Escherichia coli* strain. Expression of the wild type and its variants was induced in Luria-Bertani broth with 0.5 mM isopropyl *β*-D-1-thiogalactopyranoside for 4 h at 23°C. We purified the proteins with Ni-NTA affinity resin (Qiagen) according to standard procedures and dialyzed against storage buffer containing 50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 10 mM 2-mercaptoethanol, 3 mM $MgCl_2$, and 50% (v/v) glycerol. The coding sequences of *OsARC* and *ΔOsARC* (deletion of GST_C) were fused to the N terminus of maltose-binding protein (MBP) in the vector *pMAL2c-x*. We expressed the fusion proteins in *E. coli* (strain BL21) and purified them using amylose resin (New England Biolabs) according to the manufacturer's instructions. Supplemental Table S4 lists the primer sequences for the constructs.

tRNA Transcription and Purification

The DNA fragment covering the *T7* promoter and the $tRNA^{Glu}$ gene (http://gtrnadb.ucsc.edu/#eukarya; chr1.trna78-GluTTC; 5′-GCCCCTATCGTCTAGTGGTTCAGGACATCTCTCTTTCAAGGAGGCAGCGGGGATTCGACTTCCCCTGGGGGTA-3′) was synthesized as a double-stranded molecule by Invitrogen, phosphorylated, and ligated into *pUC19* (precleaved by *Eco*RI and *Bam*HI) to construct the plasmid *pUC19-tRNA^{Glu}*. *T7* in vitro transcription was carried out at 37°C in a reaction mixture containing 40 mM Tris-HCl (pH 8), 22 mM $MgCl_2$, 1 mM spermidine, 5 mM DTT, 0.5% (v/v) Triton X-100, 60 to 80 ng μL^{-1} tDNA template, 5 mM NTP (each), 0.8 units μL^{-1} RNase inhibitor, 20 mM GMP, 500 units μL^{-1} T7 RNA polymerase, and 1 unit mL^{-1} pyrophosphatase for 3 h. Subsequently, 5 units mL^{-1} *DNase* I (*RNase* I free) was added and incubated for 1 h to digest the transcription template. We loaded the transcripts onto a 15% (w/v) PAGE-8M urea gel with 1 mm thickness and 40 cm length and ran the gel at a constant 100 W for 4 h. The tRNA band was cut from the gel and eluted with 0.5 M NaAc (pH 5.2) at room temperature (three times), ethanol precipitated at 20°C after two phenol/chloroform extractions, and dissolved in 5 mM $MgCl_2$. The tRNA was denatured at 80°C for 5 min and slowly cooled to room temperature (Fang et al., 2014).

ATP-PPi Exchange and Aminoacylation

The ATP-PPi exchange reaction mixture contains 100 mM Tris-HCl (pH 7.8), 12 mM $MgCl_2$, various concentrations of $tRNA^{Glu}$, 5 mM L-Glu, 4 mM ATP, 2 mM [^{32}P]tetrasodium pyrophosphate (PerkinElmer), 0.1 mg mL^{-1} BSA, and 100 nM His-OsERS1 with or without 100 nM MBP-OsARC (or MBP-ΔOsARC, deleted for GST_C). The formation of [^{32}P]ATP at 37°C was followed at various time intervals in 9-μL aliquots. Samples at specific time points were terminated using 1 mL of quenching buffer containing 2% (w/v) activated charcoal, 3.5% (v/v) $HClO_4$, and 50 mM tetrasodium pyrophosphate. The charcoal suspension was put through a Whatman GF/A filter, washed four times with 5 mL of water, and rinsed with 10 mL of 100% ethanol. The charcoal powder on the filters was dried, and the synthesized [^{32}P]ATP was quantified using a scintillation counter (Beckman Coulter; Chen et al., 2000).

We performed aminoacylation reactions in a mixture of 100 mM Tris-HCl (pH 7.8), 30 mM KCl, 12 mM $MgCl_2$, 2 mM DTT, 4 mM ATP, 20 mM [^{3}H]Glu (15 Ci mM^{-1}; PerkinElmer), 15 μM $tRNA^{Glu}$, and 100 nM His-OsERS1 with or without 100 nM MBP-OsARC (or MBP-ΔOsARC, deleted for GST_C) at 37°C. We used 0.5 to 20 μM $tRNA^{Glu}$ for kinetic constant analyses of OsERS1 and its variants (mutations of three key motifs). Reactions were initiated with the addition of the enzyme and were conducted in a 37°C heat block. Aliquots (9 μL) were taken at different time points and quenched on Whatman filter pads that had been presoaked with 5% (w/v) TCA. The pads were washed three times for 10 min each time with cold 5% (w/v) TCA and once with cold 100% ethanol; the washed pads were then dried. Radioactivity was quantified in the scintillation counter. The mixture of OsERS1 and OsARC was incubated at room temperature for 30 min to allow complexation before ATP-PPi and aminoacylation assays (Chen et al., 2001).

Yeast Two-Hybrid, Pull-Down, BiFC, and Colocalization Assays

We performed yeast two-hybrid experiments according to the manufacturer's instructions (Clontech). The coding sequences of *OsERS1* and *ΔOsERS1* (deleted for GST_C) were fused in frame with the *GAL4* DNA-binding domain of the bait vector *pGBKT7*. The coding sequences of *OsARC* and *ΔOsARC* were cloned in the prey vector *pGADT7*. Each bait-prey pair was cotransformed into the yeast strain AH109. To analyze protein-protein interactions using His auxotrophy assays, yeast colonies were patched in duplicate onto His– and His+ plates and kept at 30°C for 2 to 3 d.

Amylose resin beads containing MBP, MBP-OsARC, and MBP-ΔOsARC were incubated with His-OsERS1 in pull-down buffer (15 mM HEPES-NaOH at pH 7.9, 50 mM potassium glutamate, 5 mM magnesium chloride, 5% glycerol, 0.1% Nonidet P-40, 1 μg μL^{-1} BSA, and 1 mM DTT). We rotated the mixture at 4°C for 1 h and washed the beads five times with wash buffer (15 mM

Downloaded from on August 5, 2018 - Published by www.plantphysiol.org
Copyright © 2018 American Society of Plant Biologists. All rights reserved.

HEPES-NaOH at pH 7.9, 50 mM potassium glutamate, 5 mM magnesium chloride, and 0.1% Nonidet P-40). Proteins were eluted from the beads by boiling in SDS-PAGE sample buffer, and the eluate was analyzed by immunoblot detected with anti-MBP (1:1,000 dilution) antibodies.

The cDNAs of *OsERS1*, *ΔOsERS1*, and *OsARC* were cloned individually into *pSAT1-nEYFP-N1* and *pSAT1-cEYFP-C1-B* vectors that contain either N- or C-terminal eYFP fragments. The method for BiFC assays using onion epidermal cells is described above ("Subcellular Localization"). YFP fluorescence was imaged at an excitation wavelength of 514 nm and emission wavelengths of 525 to 546 nm.

For colocalization between OsERS1 and OsARC1, we constructed *35Spro::OsERS1-eGFP* and *35Spro::OsARC-mCherry* by cloning cDNA sequences of *OsERS1* and *OsARC* into *p1301*. The method for colocalization assays using onion epidermal cells is described above ("Subcellular Localization").

Metabolomics Analysis

Over 120 mg of rice anther from wild-type and *osers1* plants was collected, frozen with liquid nitrogen, and stored at –80°C until metabolomics analysis. We ground samples into a fine powder and analyzed methanol extracts from a 40-mg sample using the analytical platform composed of ultra-HPLC-tandem mass spectrometry and gas chromatography-mass spectrometry. METABOLON performed the experiments and raw data analysis (http://www.metabolon.com). Metabolites were named after standard references (http://www.kegg.jp).

ROS Measurement and H_2O_2 Staining

We collected rice spikelets from wild-type and *osers1* plants and sorted them according to spikelet length. Samples were ground into a fine powder and processed according to the instructions of the Hydrogen Peroxide Assay Kit (abcom; catalog no. ab102500). Fluorometric readings were collected from a POLARstar OPTIMA device (BMG Labtech). Three independent experiments were performed to calculate mean values and SD. Aliquots of sample powder were suspended in 65 mM potassium phosphate buffer (pH 7.8) to extract superoxide radicals. Following a modified method (Wang and Jiao, 2000), the reaction mixture (0.5 mL of extract, 0.7 mL of potassium phosphate buffer, and 0.3 mL of 10 mM hydroxylammonium chloride) was incubated at 25°C for 60 min. Then, 0.5 mL of the above reaction mixture was added to 0.5 mL of 19 mM sulfanilic acid and 0.5 mL of 1% α-naphthylamine. After a 20-min coloring reaction at 25°C, the optical density was determined at 530 nm in a spectrophotometer. The linear standard curve was determined by 0 to 5 μM $NaNO_2$ reacted with sulfanilic acid and α-naphthylamine.

Freshly collected anthers were carefully immersed in the fixative solution O.C.T. Compound (Sakura) and frozen in liquid nitrogen. Embedded fresh tissues were sectioned into 9-μm-thick sections using a Leica CM1800 cryostat. Production of H_2O_2 was performed by incubating sectioned anthers in the DAB solution (Vector Laboratories) for 3 to 4 min, optimized from the manufacturer's instructions, and photographed with a Nikon Ni-E optical microscope.

Anther Treatments

We removed the outermost whorls of leaves from wild-type and *osers1* stems to locate the inflorescences. Next, 1 mM KI and 100 μM H_2O_2 solutions were freshly prepared and injected through leaf whorls using an 18-gauge needle into the airspace about 5 cm above the first stem node, where the young inflorescence was developed (Supplemental Fig. S9). Water injection served as a control. After 72 h of treatment, rice spikelets were fixed, embedded in resin, and sectioned as described above.

Accession Numbers

Accession numbers are as follows: *OsERS1*, LOC_Os10g22380; *OsERS2*, LOC_Os01g16520; *OsERS3*, LOC_Os02g02860; *OsARC*, LOC_Os01g60660; *MSP1*, LOC_Os01g68870; *H4*, LOC_Os09g38020; *CDKB2;1*, LOC_Os08g40170; *AlaRS3*, LOC_Os10g10244; *ArgRS2*, LOC_Os05g07030; *AsnRS1*, LOC_Os01g27520; *AspRS3*, LOC_Os02g46130; *CysRS1*, LOC_Os03g04960; *GlnRS1*, LOC_Os01g09000; *GlyRS3*, LOC_Os08g42560; *LysRS2*, LOC_Os03g38980; *MetRS2*, LOC_Os06g31210; *ProRS2*, LOC_Os12g25710; *ThrRS2*, LOC_Os08g19850; and *TrpRS2*, LOC_Os12g35570.

Supplemental Data

The following supplemental materials are available.

Supplemental Figure S1. Phenotypic analysis of the *osers1* mutant.

Supplemental Figure S2. Cytological analysis of wild-type and *osers1* anthers.

Supplemental Figure S3. Expression of genes that encode regulators of early anther development in *osers1* anthers.

Supplemental Figure S4. Molecular characteristics of *OsERS1*.

Supplemental Figure S5. Alignment of conserved motifs among GluRSs.

Supplemental Figure S6. Shift of major metabolic pathways in *osers1* anthers.

Supplemental Figure S7. RT-qPCR detection of expression levels of aaRSs genes.

Supplemental Figure S8. DAB staining analysis for H_2O_2 in anther lobes.

Supplemental Figure S9. Chemical treatments to anthers.

Supplemental Figure S10. Full scan figures of the gel and western blot.

Supplemental Table S1. Alignment of GluRS amino acid sequences.

Supplemental Table S2. Heat map of metabolite changes between the wild type and the *osers1* mutant.

Supplemental Table S3. RT-qPCR detection of the expression changes of genes in metabolic pathways between the wild type and the *osers1* mutant.

Supplemental Table S4. Sequences of primers used in this study.

ACKNOWLEDGMENTS

We thank Zhijing Luo and Mingjiao Chen for performing *osers1* mutant screens and rice crosses; Enduo Wang and Dr. Yong Wang (Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences) for help with enzyme activity measurements; Chi-kuang Wen (Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences) for providing the *CD3-991* vector; Dr. Sheng Quan, Dr. Guorong Qu, Qian Luo, and Jin Zhou (Shanghai Jiao Tong University) for support in metabolomics analysis; Dr. Gwenda Mayo (Adelaide Microscopy) for technical support; Jianping Hu (Michigan State University) for valuable comments; Deborah Devis (University of Adelaide) for revision of the article; and Diane Mather (University of Adelaide) for advice and comments.

Received January 25, 2018; accepted April 20, 2018; published May 2, 2018.

LITERATURE CITED

Abe A, Kosugi S, Yoshida K, Natsume S, Takagi H, Kanzaki H, Matsumura H, Yoshida K, Mitsuoka C, Tamiru M, (2012) Genome sequencing reveals agronomically important loci in rice using MutMap. Nat Biotechnol **30**: 174–178

Aya K, Ueguchi-Tanaka M, Kondo M, Hamada K, Yano K, Nishimura M, Matsuoka M (2009) Gibberellin modulates anther development in rice via the transcriptional regulation of GAMYB. Plant Cell **21**: 1453–1472

Berg M, Rogers R, Muralla R, Meinke D (2005) Requirement of aminoacyl-tRNA synthetases for gametogenesis and embryo development in Arabidopsis. Plant J **44**: 866–878

Cairns RA, Harris IS, Mak TW (2011) Regulation of cancer cell metabolism. Nat Rev Cancer **11**: 85–95

Downloaded from on August 5, 2018 - Published by www.plantphysiol.org
Copyright © 2018 American Society of Plant Biologists. All rights reserved.

Canales C, Bhatt AM, Scott R, Dickinson H (2002) *EXS*, a putative LRR receptor kinase, regulates male germline cell number and tapetal identity and promotes seed development in *Arabidopsis*. Curr Biol **12**: 1718–1727

Chaubal R, Anderson JR, Trimnell MR, Fox TW, Albertsen MC, Bedinger P (2003) The transformation of anthers in the *msca1* mutant of maize. Planta **216**: 778–788

Chen JF, Guo NN, Li T, Wang ED, Wang YL (2000) CP1 domain in *Escherichia coli* leucyl-tRNA synthetase is crucial for its editing function. Biochemistry **39**: 6726–6731

Chen JF, Li T, Wang ED, Wang YL (2001) Effect of alanine-293 replacement on the activity, ATP binding, and editing of *Escherichia coli* leucyl-tRNA synthetase. Biochemistry **40**: 1144–1149

Chen L, Chu HW, Yuan Z, Pan AH, Liang WQ, Huang H, Shen MS, Zhang DB, Chen L (2006) Isolation and genetic analysis for rice mutants treated with 60Co γ-ray. J Xiamen Univ **45**: 81–85

Coruzzi GM, Zhou L (2001) Carbon and nitrogen sensing and signaling in plants: emerging 'matrix effects.' Curr Opin Plant Biol **4**: 247–253

Cotelle V, Meek SE, Provan F, Milne FC, Morrice N, MacKintosh C (2000) 14-3-3s regulate global cleavage of their diverse binding partners in sugar-starved Arabidopsis cells. EMBO J **19**: 2869–2876

Dixon DP, Lapthorn A, Edwards R (2002) Plant glutathione transferases. Genome Biol **3**: S3004

Fang ZP, Wang M, Ruan ZR, Tan M, Liu RJ, Zhou M, Zhou XL, Wang ED (2014) Coexistence of bacterial leucyl-tRNA synthetases with archaeal tRNA binding domains that distinguish tRNALeu in the archaeal mode. Nucleic Acids Res **42**: 5109–5124

Freist W, Gauss DH, Söll D, Lapointe J (1997) Glutamyl-tRNA sythetase. Biol Chem **378**: 1313–1329

Fu Z, Yu J, Cheng X, Zong X, Xu J, Chen M, Li Z, Zhang D, Liang W (2014) The rice basic helix-loop-helix transcription factor TDR INTERACTING PROTEIN2 is a central switch in early anther development. Plant Cell **26**: 1512–1524

Galani K, Grosshans H, Deinert K, Hurt EC, Simos G (2001) The intracellular location of two aminoacyl-tRNA synthetases depends on complex formation with Arc1p. EMBO J **20**: 6889–6898

Graindorge JS, Senger B, Tritch D, Simos G, Fasiolo F (2005) Role of Arc1p in the modulation of yeast glutamyl-tRNA synthetase activity. Biochemistry **44**: 1344–1352

Guo M, Schimmel P (2013) Essential nontranslational functions of tRNA synthetases. Nat Chem Biol **9**: 145–153 10.1038/nchembio.115823416400

Guo M, Yang XL, Schimmel P (2010) New functions of aminoacyl-tRNA synthetases beyond translation. Nat Rev Mol Cell Biol **11**: 668–674

Holmgren A (2000) Antioxidant function of thioredoxin and glutaredoxin systems. Antioxid Redox Signal **2**: 811–820

Hong L, Tang D, Shen Y, Hu Q, Wang K, Li M, Lu T, Cheng Z (2012a) MIL2 (MICROSPORELESS2) regulates early cell differentiation in the rice anther. New Phytol **196**: 402–413

Hong L, Tang D, Zhu K, Wang K, Li M, Cheng Z (2012b) Somatic and reproductive cell development in rice anther is regulated by a putative glutaredoxin. Plant Cell **24**: 577–588

Ibba M, Söll D (2000) Aminoacyl-tRNA synthesis. Annu Rev Biochem **69**: 617–650

Ito K, Ito K (2016) Metabolism and the control of cell fate decisions and stem cell renewal. Annu Rev Cell Dev Biol **32**: 399–409

Jeong SM, Xiao C, Finley LW, Lahusen T, Souza AL, Pierce K, Li YH, Wang X, Laurent G, German NJ, (2013) SIRT4 has tumor-suppressive activity and regulates the cellular metabolic response to DNA damage by inhibiting mitochondrial glutamine metabolism. Cancer Cell **23**: 450–463

Jung KH, Han MJ, Lee YS, Kim YW, Hwang I, Kim MJ, Kim YK, Nahm BH, An G (2005) *Rice Undeveloped Tapetum1* is a major regulator of early tapetum development. Plant Cell **17**: 2705–2722

Kägi C, Baumann N, Nielsen N, Stierhof YD, Gross-Hardt R (2010) The gametic central cell of Arabidopsis determines the lifespan of adjacent accessory cells. Proc Natl Acad Sci USA **107**: 22350–22355

Kelliher T, Walbot V (2012) Hypoxia triggers meiotic fate acquisition in maize. Science **337**: 345–348

Kim YK, Lee JY, Cho HS, Lee SS, Ha HJ, Kim S, Choi D, Pai HS (2005) Inactivation of organellar glutamyl- and seryl-tRNA synthetases leads to developmental arrest of chloroplasts and mitochondria in higher plants. J Biol Chem **280**: 37098–37106

Laporte D, Huot JL, Bader G, Enkler L, Senger B, Becker HD (2014) Exploring the evolutionary diversity and assembly modes of multi-aminoacyl-tRNA synthetase complexes: lessons from unicellular organisms. FEBS Lett **588**: 4268–4278

Li N, Zhang DS, Liu HS, Yin CS, Li XX, Liang WQ, Yuan Z, Xu B, Chu HW, Wang J, (2006) The rice *tapetum degeneration retardation* gene is required for tapetum degradation and anther development. Plant Cell **18**: 2999–3014

Liu Z, Bao W, Liang W, Yin J, Zhang D (2010) Identification of *gamyb-4* and analysis of the regulatory role of *GAMYB* in rice anther development. J Integr Plant Biol **52**: 670–678

Lü Y, Cui X, Li R, Huang P, Zong J, Yao D, Li G, Zhang D, Yuan Z (2015) Development of genome-wide insertion/deletion markers in rice based on graphic pipeline platform. J Integr Plant Biol **57**: 980–991

Ma H (2005) Molecular genetic analyses of microsporogenesis and microgametogenesis in flowering plants. Annu Rev Plant Biol **56**: 393–434

Mittler R (2017) ROS are good. Trends Plant Sci **22**: 11–19

Møller IM, Jensen PE, Hansson A (2007) Oxidative modifications to cellular components in plants. Annu Rev Plant Biol **58**: 459–481

Nelson BK, Cai X, Nebenführ A (2007) A multicolored set of *in vivo* organelle markers for co-localization studies in Arabidopsis and other plants. Plant J **51**: 1126–1136

Newsholme P, Procopio J, Lima MMR, Pithon-Curi TC, Curi R (2003) Glutamine and glutamate: their central role in cell metabolism and function. Cell Biochem Funct **21**: 1–9

Nonomura K, Miyoshi K, Eiguchi M, Suzuki T, Miyao A, Hirochika H, Kurata N (2003) The MSP1 gene is necessary to restrict the number of cells entering into male and female sporogenesis and to initiate anther wall formation in rice. Plant Cell **15**: 1728–1739

Schieber M, Chandel NS (2014) ROS function in redox signaling and oxidative stress. Curr Biol **24**: R453–R462

Schuster C, Gaillochet C, Medzihradszky A, Busch W, Daum G, Krebs M, Kehle A, Lohmann JU (2014) A regulatory framework for shoot stem cell control integrating metabolic, transcriptional, and phytohormone signals. Dev Cell **28**: 438–449

Sena LA, Chandel NS (2012) Physiological roles of mitochondrial reactive oxygen species. Mol Cell **48**: 158–167

Sperber H, Mathieu J, Wang Y, Ferreccio A, Hesson J, Xu Z, Fischer KA, Devi A, Detraux D, Gu H, (2015) The metabolome regulates the epigenetic landscape during naive-to-primed human embryonic stem cell transition. Nat Cell Biol **17**: 1523–1535

Thomes JC, Ratinaud MH, Julien R (1983) Dimeric glutamyl-tRNA synthetases from wheat: kinetic properties and functional structures. Eur J Biochem **135**: 479–484

Tsukagoshi H, Busch W, Benfey PN (2010) Transcriptional regulation of ROS controls transition from proliferation to differentiation in the root. Cell **143**: 606–616

Umeda M, Umeda-Hara C, Yamaguchi M, Hashimoto J, Uchimiya H (1999) Differential expression of genes for cyclin-dependent protein kinases in rice plants. Plant Physiol **119**: 31–40

Walbot V, Egger RL (2016) Pre-meiotic anther development: cell fate specification and differentiation. Annu Rev Plant Biol **67**: 365–395

Wang SY, Jiao H (2000) Scavenging capacity of berry crops on superoxide radicals, hydrogen peroxide, hydroxyl radicals, and singlet oxygen. J Agric Food Chem **48**: 5677–5684

Wang Y, Wang C, Zheng M, Lyu J, Xu Y, Li X, Niu M, Long W, Wang D, Wang H, (2016) WHITE PANICLE1, a Val-tRNA synthetase regulating chloroplast ribosome biogenesis in rice, is essential for early chloroplast development. Plant Physiol **170**: 2110–2123

Wilson ZA, Zhang DB (2009) From Arabidopsis to rice: pathways in pollen development. J Exp Bot **60**: 1479–1492

Xing S, Zachgo S (2008) *ROXY1* and *ROXY2*, two Arabidopsis glutaredoxin genes, are required for anther development. Plant J **53**: 790–801

Xiong Y, McCormack M, Li L, Hall Q, Xiang C, Sheen J (2013) Glucose-TOR signalling reprograms the transcriptome and activates meristems. Nature **496**: 181–186

Yamakawa H, Hakata M (2010) Atlas of rice grain filling-related metabolism under high temperature: joint analysis of metabolome and transcriptome demonstrated inhibition of starch accumulation and induction of amino acid accumulation. Plant Cell Physiol **51**: 795–809

Yang L, Qian X, Chen M, Fei Q, Meyers BC, Liang W, Zhang D (2016) Regulatory role of a receptor-like kinase in specifying anther cell identity. Plant Physiol **171**: 2085–2100

Yang SL, Xie LF, Mao HZ, Puah CS, Yang WC, Jiang L, Sundaresan V, Ye D (2003) *Tapetum determinant1* is required for cell specialization in the *Arabidopsis* anther. Plant Cell **15**: 2792–2804

Downloaded from on August 5, 2018 - Published by www.plantphysiol.org
Copyright © 2018 American Society of Plant Biologists. All rights reserved.

Zeng J, Dong Z, Wu H, Tian Z, Zhao Z (2017) Redox regulation of plant stem cell fate. EMBO J **36**: 2844–2855

Zhang D, Yang L (2014) Specification of tapetum and microsporocyte cells within the anther. Curr Opin Plant Biol **17**: 49–55

Zhang Y, Su J, Duan S, Ao Y, Dai J, Liu J, Wang P, Li Y, Liu B, Feng D, Wang J, Wang H (2011) A highly efficient rice green tissue protoplast system for transient gene expression and studying light/chloroplast-related processes. Plant Methods **7**: 30

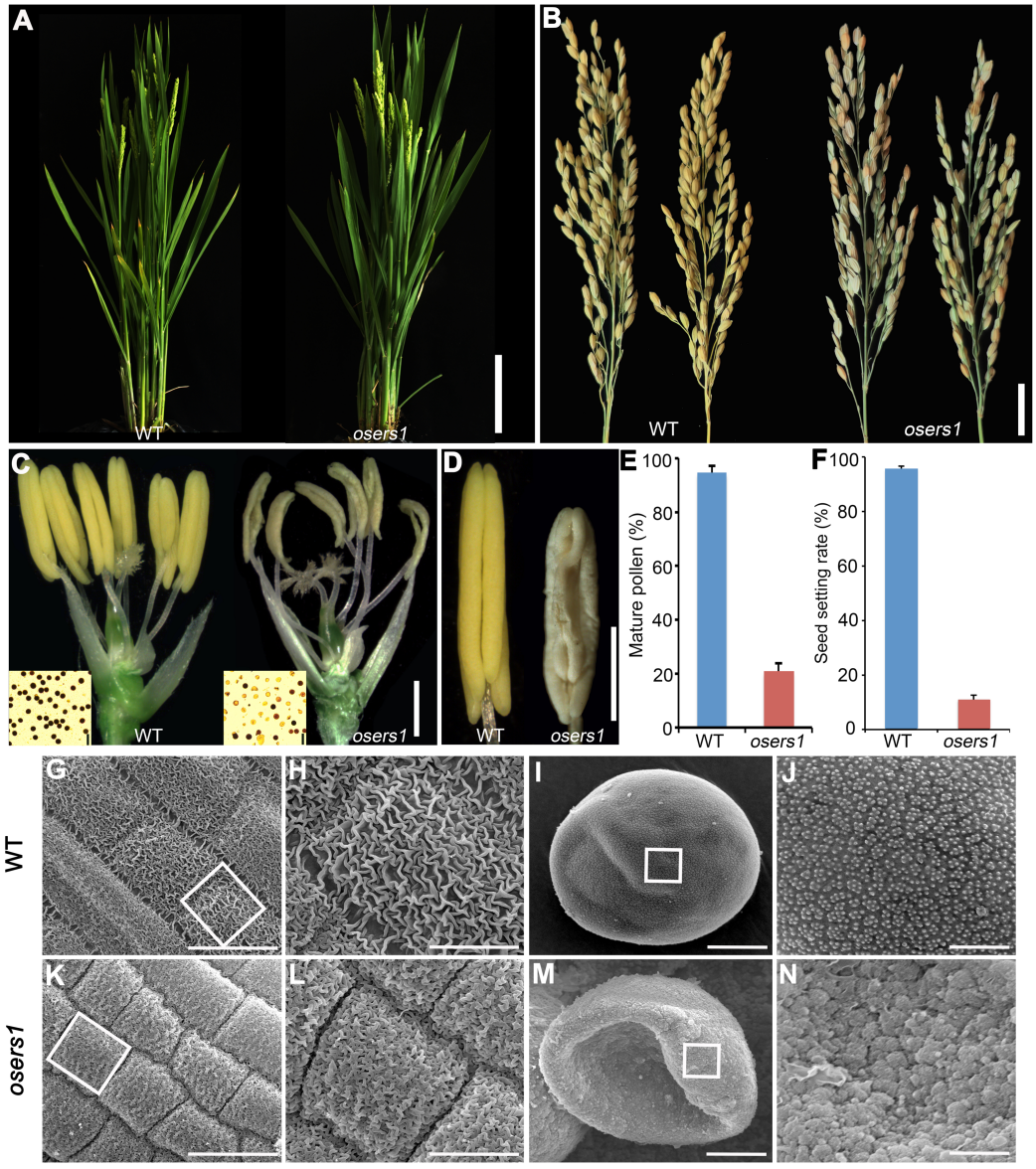
Zhao DZ, Wang GF, Speal B, Ma H (2002) The *excess microsporocytes1* gene encodes a putative leucine-rich repeat receptor protein kinase that controls somatic and reproductive cell fates in the *Arabidopsis* anther. Genes Dev **16**: 2021–2031

Zhao X, de Palma J, Oane R, Gamuyao R, Luo M, Chaudhury A, Hervé P, Xue Q, Bennett J (2008) OsTDL1A binds to the LRR domain of rice receptor kinase MSP1, and is required to limit sporocyte numbers. Plant J **54**: 375–387

Downloaded from on August 5, 2018 - Published by www.plantphysiol.org
Copyright © 2018 American Society of Plant Biologists. All rights reserved.

Supplemental Information

A Rice Glutamyl-tRNA Synthetase Modulates Early Anther Cell Division and Patterning



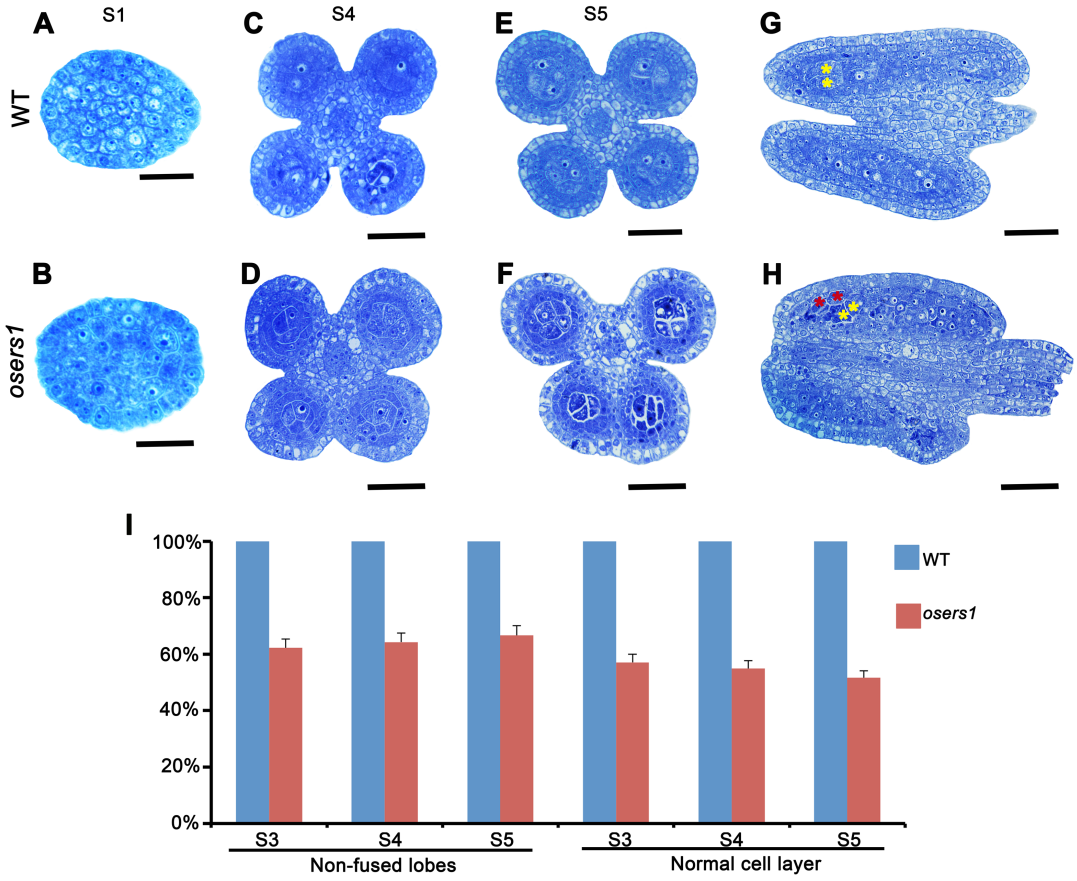
Supplemental Figure S1. Phenotypic analysis of the *osers1* mutant.

(A) Whole plants after heading. Bar = 10 cm.

(B) Inflorescences and mature seeds. Bars = 2 cm (inflorescence), 2 mm (seeds).

(C) to **(F)** Defective male fertility of the *osers1* mutant. **(C)** Comparison between wild-type and *osers1* anthers and (insets) pollen grains stained by iodine-potassium iodide (I_2-KI). Bars, 1 mm for the anthers and 100 μm for the pollens. **(D)** Comparison between wild-type (left) and pale, lobe-fused *osers1* (right) anthers. Bar = 1 mm. **(E)** *osers1* showed a lower percentage of mature pollen grains. Error bars represent SD from 3 biological replicates (n>700 pollen grains for each set). **(F)** *osers1* inflorescences showed lower seed setting rate. Error bars represent SD (n>15 for each set).

(G) to **(N)** Scanning electron micrographs of outer surface of anther wall and pollen grains. **(H)** and **(L)** are magnifications of boxed regions in anther wall cells in **(G)** and **(K)**; **(J)** and **(N)** are magnifications of boxed regions in pollen grain surface in **(I)** and **(M)**. *osers1* anther showed a more tightly woven texture on outer surface, producing shrunken pollen grains with abnormal outer pollen wall (especially exine), which differs from the wild type that showed well-arranged cuticle on the exterior of the anther and smooth outer surface of the mature pollen grain. Bars = 20 μm in **(G)** and **(K)**, 10 μm in **(H)**, **(I)**, **(L)** and **(M)** and 2 μm in **(J)** and **(N)**.

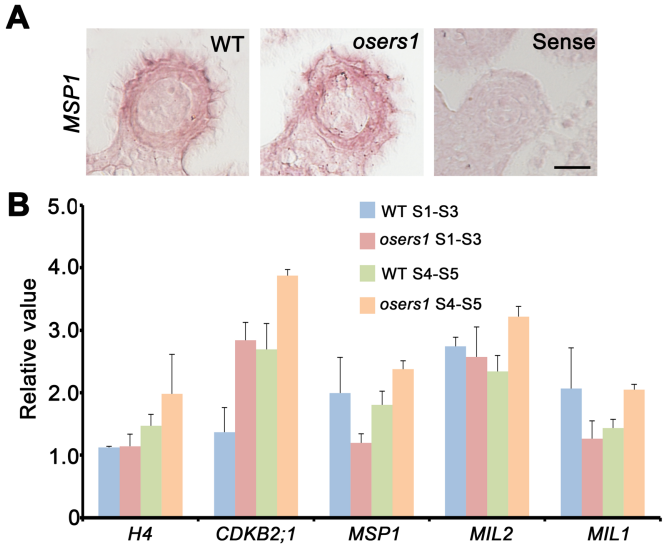


Supplemental Figure S2. Cytological analysis of wild-type and *osers1* anthers.

(A) to **(F)** Cross sections of anthers at various stages.

(G) and **(H)** Longitudinal sections of wild-type **(G)** and *osers1* **(H)** anthers. Yellow asterisks indicate sporogenous-like cells in normal position, and red asterisks indicate ectopic sporogenous-like cells in *osers1*. Bars = 50 μm.

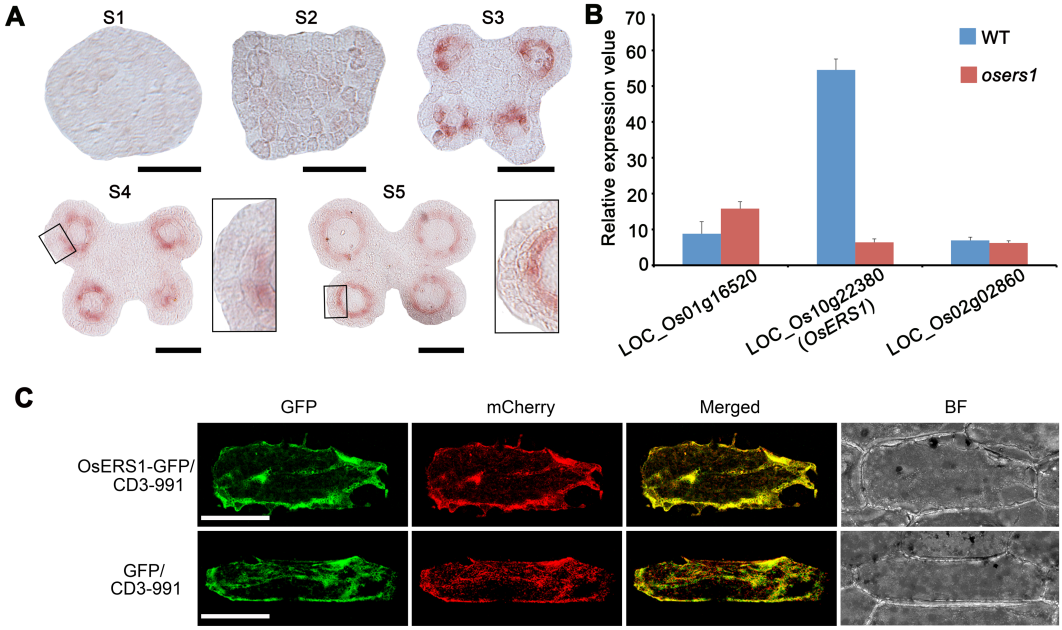
(I) Comparison of normally developed lobes and normal cell layers in wild-type and *osers1* anthers. Error bars represent SD (n=48 for each genotype).



Supplemental Figure S3. Expression of genes that encode regulators of early anther development in *osers1* anthers.

(A) *MSP1* mRNA hybridization by anti-sense probe in wild-type and *osers1* anthers with sense probe as the control. Bar = 15 μm.

(B) RT-qPCR analysis of *H4, CDKB2;1, MSP1, MIL2* and *MIL1* mRNA accumulation at different anther developmental stages. Error bars represent SD of three independent experiments.

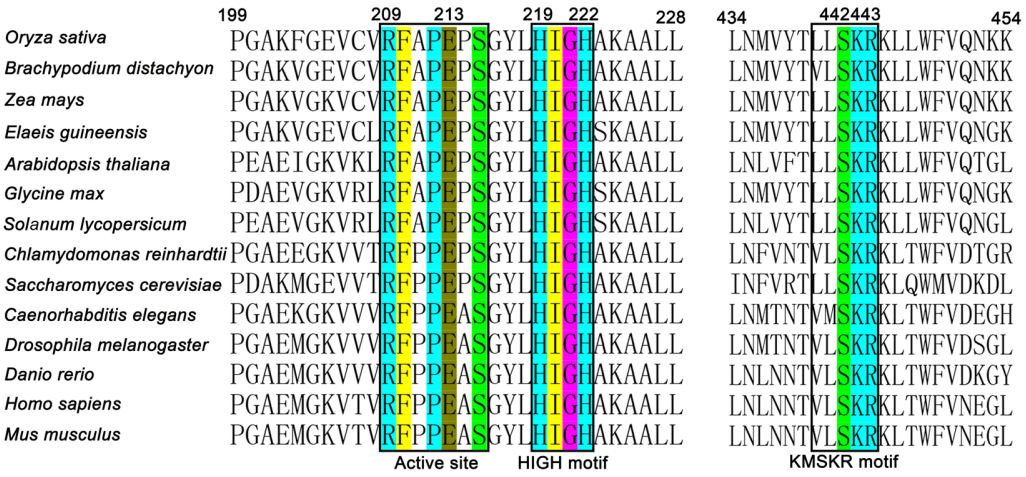


Supplemental Figure S4. Molecular characteristics of *OsERS1*.

(A) *In situ* hybridization of *OsERS1* in wild-type anther from stage 1 to stage 5 with magnifications of anther wall layers. Bars = 15 μm.

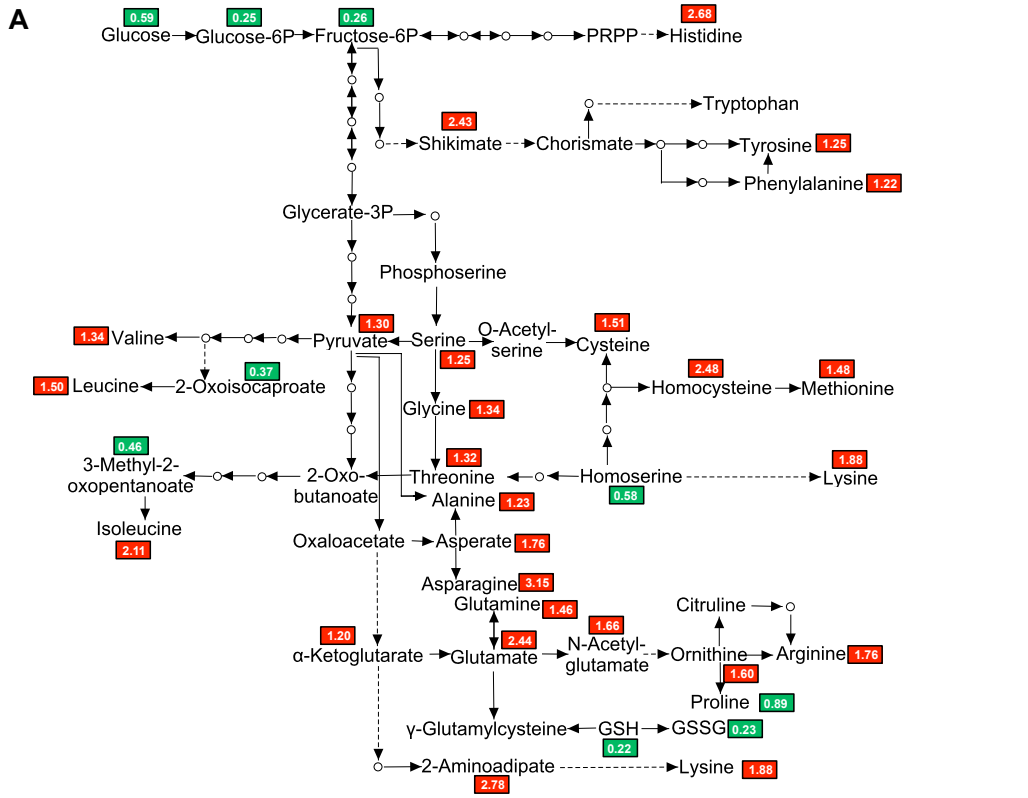
(B) RT-qPCR analysis for expression levels of three rice *ERS* genes. Error bars represent SD (n=3).

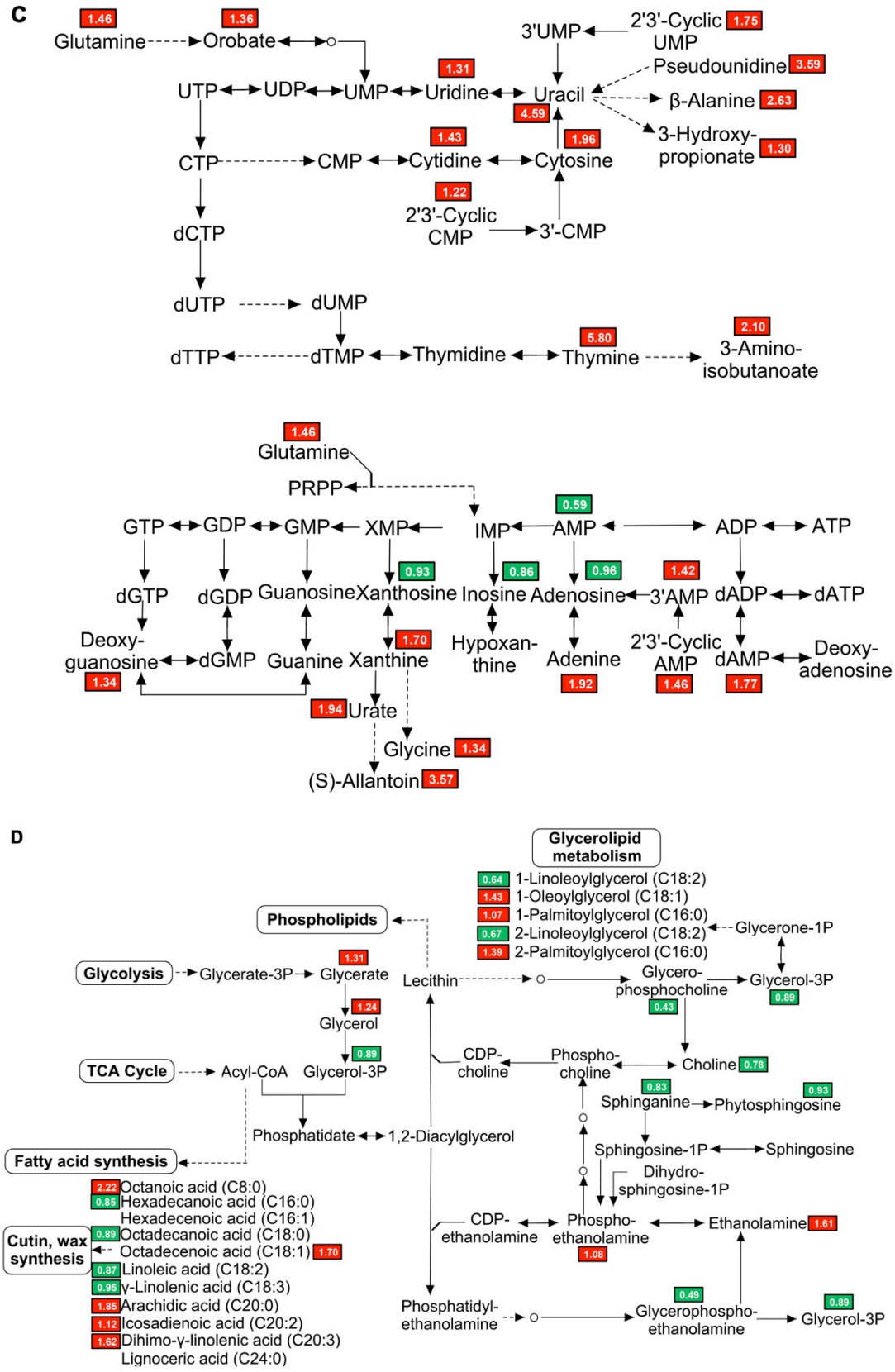
(C) Subcellular localization of OsERS1-GFP in onion epidermal cells. ScCOX4-mCherry serves as the mitochondrial marker. BF, bright field. Bars = 50 μm.



Supplemental Figure S5. Alignment of conserved motifs among GluRSs.

Amino acid sequence alignment of two conserved regions from GluRSs from various species. Three conserved motifs RFAPEPS, HIGH and KMSKR, which are important for ATP binding and activation, are boxed. Conserved amino acids are colored. Numbers above amino acids indicate positions in OsERS1. Sequence accession numbers are: *Oryza sativa*, OS10T0369000-01; *Brachypodium distachyon*, XP_003566911.1; *Zea mays,* GRMZM2G046932_T01; *Elaeis guin eensis*, XP_010941452.1; *Arabidopsis thaliana*, AT5G26710.1; *Glycine max*, XP_003555767.1; *Solanum lycopersicum*, XP_004230376.1; *Chlamydomonas reinhardtii*, EDO97532; *Saccharomyces cerevisiae*, NP_011269.2; *Caenorhabditis elegans*, NP_492711.1; *Drosophila melanogaster*, NP_524471.2; *Danio rerio*, NP_001275581.1; *Homo sapiens*, NP_004437.2; *Mus musculus*, NP_084011.1.


A
Glucose
Glucose-6P
Fructose-6P
PRPP
Histidine
Tryptophan
Shikimate
Chorismate
Tyrosine
Phenylalanine
Glycerate-3P
Phosphoserine
Valine
Pyruvate
Serine
O-Acetyl-serine
Cysteine
Leucine
2-Oxoisocaproate
Glycine
Homocysteine
Methionine
3-Methyl-2-oxopentanoate
2-Oxo-butanoate
Threonine
Homoserine
Lysine
Alanine
Isoleucine
Oxaloacetate
Asperate
Asparagine
Glutamine
Citruline
α-Ketoglutarate
Glutamate
N-Acetyl-glutamate
Ornithine
Arginine
Proline
γ-Glutamylcysteine
GSH
GSSG
2-Aminoadipate
Lysine
B
Glucose
Sorbitol
Fructose
Mannitol
Mannitol-1P
Glucose
Gluconate
Mannose
Mannose-6P
Fructose-6P
Glucose-6P
Gluconate-6P
Fructose-1,6P2
ß-Glucose-6P
Ribose-5P
Ribulose-5P
Arbutin
Glycolysis
Glyceraldehyde-3P
PRPP
PPP
Nucleotides metabolism
Glycerate-1,3P2
Serine
Glycerate-3P
Ribulose-1,5P2
Glycerate
Glycerate-2P
Glycolate
Asperate
Alanine
Pyruvate
Lactate
Malate
Oxaloacetate
Citrate
Acetyl-CoA
Fatty acid synthesis
Fumarate
TCA Cycle
Isocitrate
Nucleotides metabolism
Succinate
Succinyl-CoA
α-Ketoglutarate
Glutamate
Glutamine



Supplemental Figure S6. Shift of major metabolic pathways in *osers1* anthers.

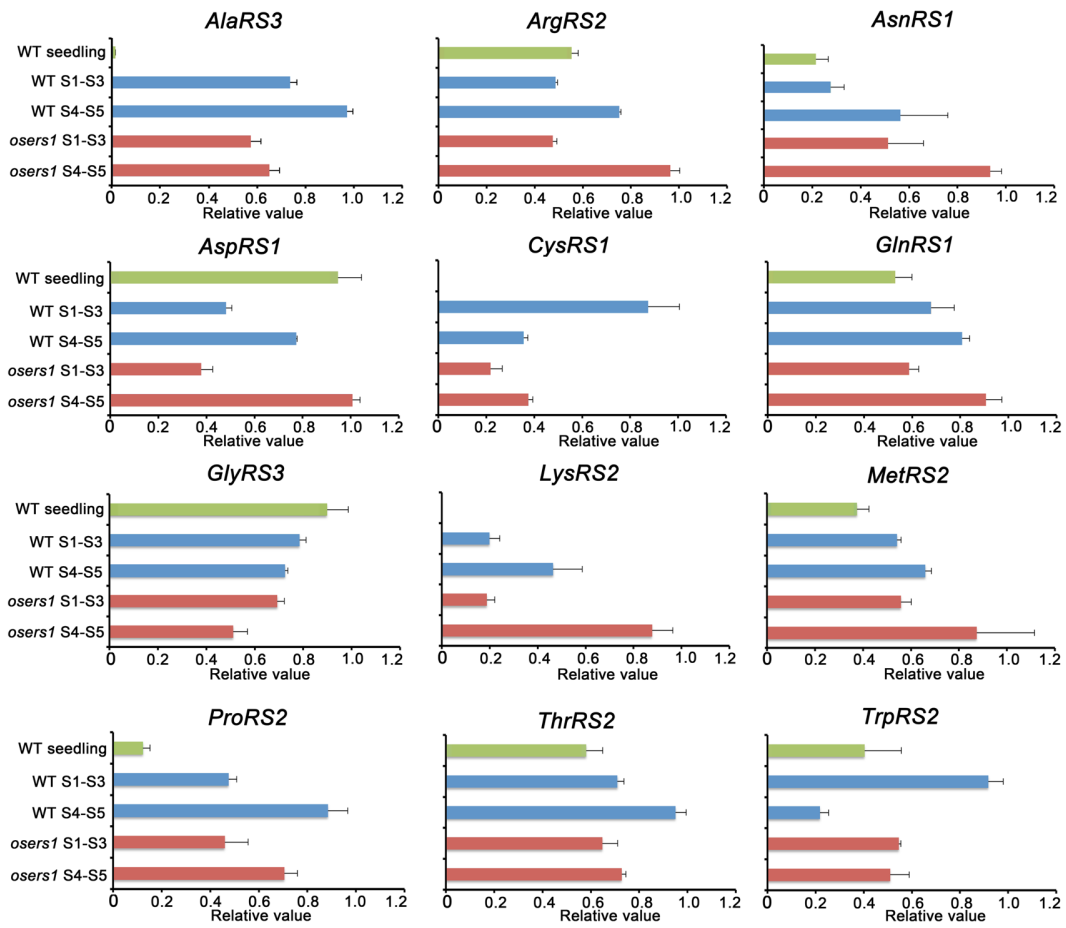
(A) Amino acid metabolism in *osers1* anthers. There is a trend of global increase in the

levels of amino acids and their derivatives in the mutant. The 20 essential amino acids were mostly increased except for proline. Derivatives from three key metabolites, pyruvate, oxaloacetate and α-ketoglutarate, attributed the most to changes in amino acids levels.

(B) Carbohydrate metabolism in *osers1* anthers. Decreased glucose, fructose and other intermediates indicated the supply of glycolysis was consumed more in *osers1* anthers, causing an increase in the final product pyruvate. In *osers1* anthers, pyruvate was used more to support amino acid synthesis and TCA cycle instead of to produce lactate. On the other hand, higher level of α-ketoglutarate was an entry point for amino acids into carbohydrate metabolism. TCA cycle in *osers1* anthers was promoted.

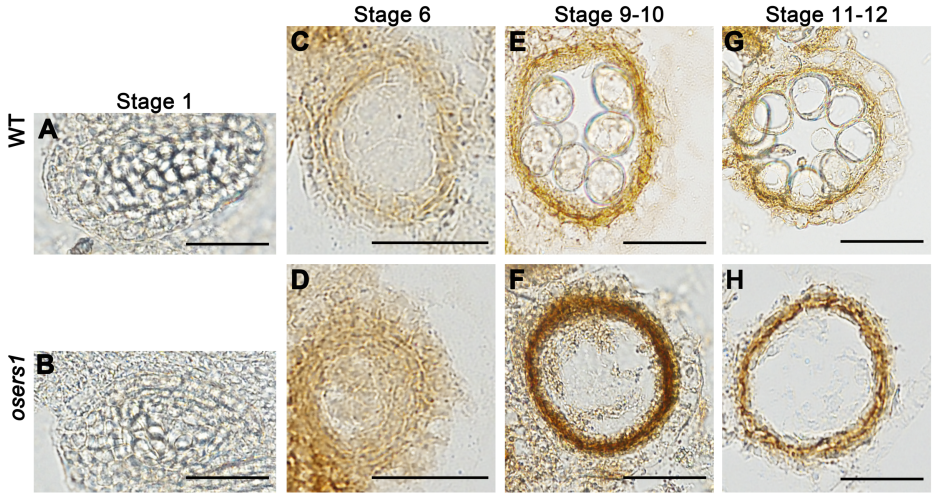
(C) Purine and pyrimidine biosynthesis in *osers1* anthers. Both purine and pyrimidine were synthesized in a larger amount, providing essential resources to support cell proliferation in *osers1* anthers. The carbon and nitrogen for this pathway were likely supplied by intermediates from glycolysis and glutaminolysis.

(D) Fatty acid metabolism in *osers1* anthers. Fatty acid biosynthesis was disturbed in *osers1* anthers. Notably, the disorder in cutin and wax synthesis was consistent with defective anther/pollen wall of *osers1*. Red and green squares respectively indicate metabolites with higher and lower average levels in *osers1* mutant compared with wild type.



Supplemental Figure S7. RT-qPCR detection of expression levels of aminoacyl-tRNA synthetase genes.

WT, wild type; S1−S3, stage 1 to stage 3; S4−S5, stage 4 to stage 5. All error bars represent SD (n=3).



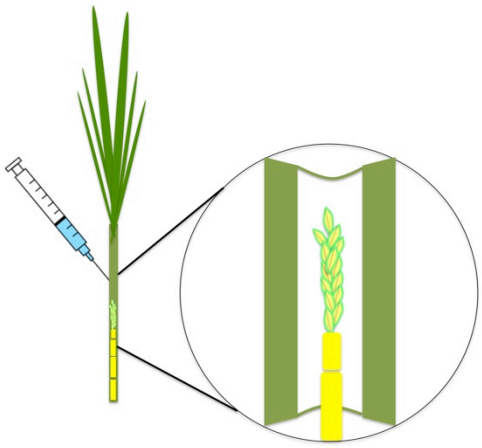
Supplemental Figure S8. DAB staining analysis for hydrogen peroxide in anther lobes.

(A) and **(B)** anther primordia in stage 1.

(C) and **(D)** anther lobes in stage 6. H_2O_2 was increased in *osers1*.

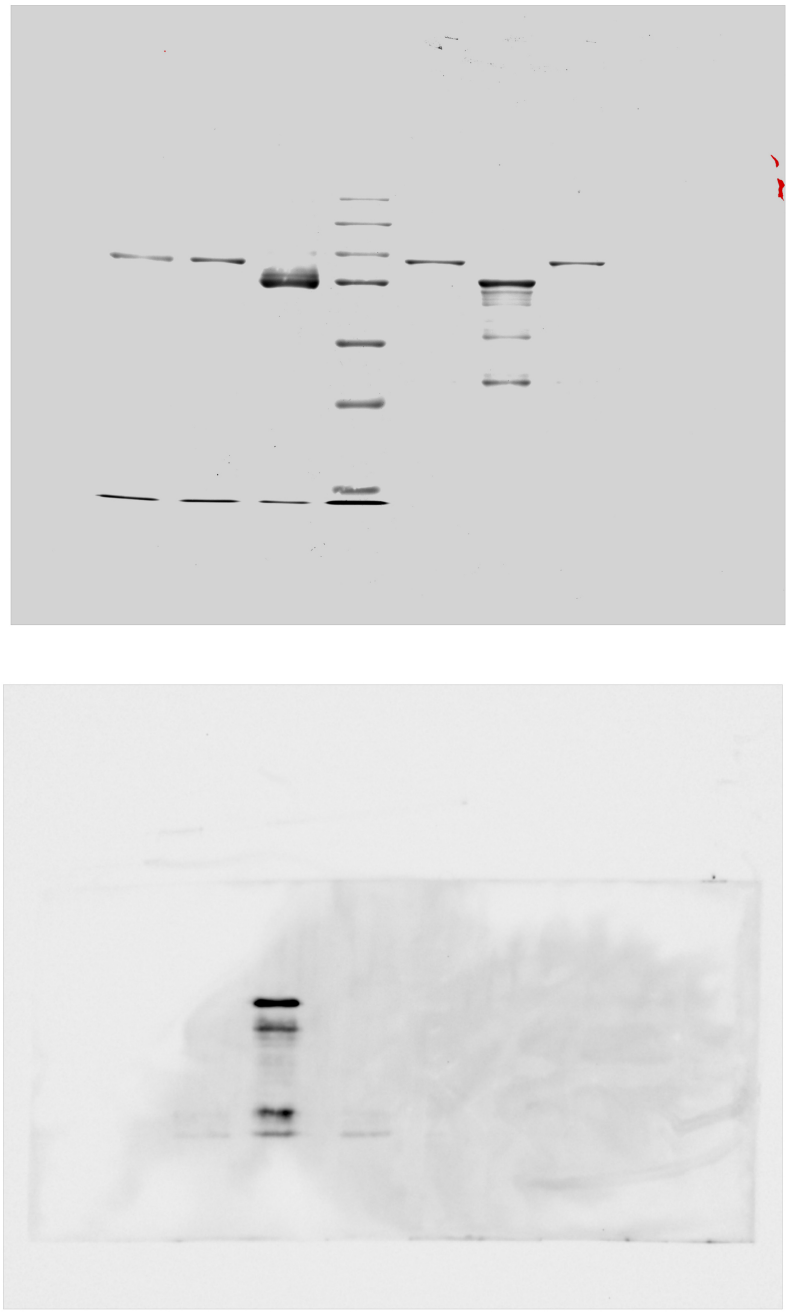
(E) and **(F)** anther lobes in stage 9−10. H_2O_2 was mainly localized in degrading anther inner wall layers.

(G) and **(H)** anther lobes in stage 11−12. H_2O_2 was decreased as anther inner wall layers disappeared. Bars = 50 μm.



Supplemental Figure S9. Chemical treatments to anthers.

Chemicals dissolved in water were injected using a sterile syringe and needle into the airspace about 5 cm above the first stem node, where the young inflorescence was developed.



Supplemental Figure S10. Full scan figures of the gel and western blot.

The original images of gel and western blot of Figure 5C.

Supplemental Table S3. RT-qPCR detection of the expression changes of genes in metabolic pathways between the wild type and the *osers1* mutant.

Pathway	Gene	Locus	Relative value			
			WT S1-S3	*osers1* S1-S3	WT S4-S5	*osers1* S4-S5
TCA cycle	*CS1*	LOC_Os02g10070	1.214±0.153	5.450±0.591	11.531±0.250	15.743±1.580
	CS2	LOC_Os02g13840	1.908±0.028	1.323±0.112	5.116±0.282	1.014±0.012
	CS3	LOC_Os11g33240	1.345±0.251	3.303±0.377	5.145±0.243	7.594±1.103
	ODH1	LOC_Os04g32020	1.768±0.048	1.330±0.100	2.923±0.207	1.078±0.065
	ODH2	LOC_Os04g32330	1.118±0.085	5.548±1.217	2.896±0.492	9.365±0.338
	ODH3	LOC_Os07g49520	1.796±0.092	1.389±0.183	2.589±0.118	1.050±0.043
	MDH2	LOC_Os01g61380	1.181±0.031	2.696±1.510	1.150±0.118	9.663±1.680
	MDH9	LOC_Os08g33720	1.508±0.022	1.201±0.142	1.583±0.018	1.176±0.041
	MDH11	LOC_Os10g33800	2.641±0.805	1.882±0.145	1.257±0.258	8.018±0.654
Glutamate metabolism	*GS2*	LOC_Os03g12290	1.555±0.395	3.923±1.321	3.212±0.479	4.291±0.208
	GS3	LOC_Os03g50490	1.085±0.061	3.547±0.326	5.118±0.767	6.743±2.266
	GS4	LOC_Os04g56400	1.051±0.055	1.792±0.140	2.562±0.230	1.752±0.164
	GDH1	LOC_Os01g37760	1.117±0.130	3.336±1.042	1.638±0.479	4.998±1.066
	GDH3	LOC_Os03g58040	2.411±0.107	2.114±0.029	1.451±0.323	1.749±0.107
Pentose phosphate pathway	*G6PDH3*	LOC_Os03g29950	1.386±0.357	9.227±1.525	1.785±0.227	21.765±2.386
	G6PDH5	LOC_Os07g22350	1.269±0.270	3.686±1.601	3.898±0.665	2.363±0.022
ROS scavenging	*APX1*	LOC_Os03g17690	2.442±0.214	1.491±0.048	3.020±0.032	1.192±0.149
	APX2	LOC_Os07g49400	1.689±0.065	1.018±0.013	4.250±0.223	1.247±0.072
	APX3	LOC_Os04g14680	1.077±0.068	1.403±0.133	2.415±0.143	1.759±0.105
	APX4	LOC_Os08g43560	1.488±0.051	1.573±0.129	2.700±0.148	1.067±0.054
	APX5	LOC_Os12g07830	1.136±0.067	1.343±0.136	1.781±0.135	1.074±0.061
	APX6	LOC_Os12g07820	1.546±0.044	1.469±0.060	1.966±0.090	1.028±0.020
	GPX1	LOC_Os02g44500	1.047±0.034	1.271±0.089	2.150±0.112	1.308±1.047
	GPX2	LOC_Os03g24380	1.085±0.115	1.406±0.290	2.362±0.072	1.047±0.034
	GPX3	LOC_Os04g46960	1.057±0.040	1.623±0.025	2.754±0.054	1.555±0.070
	GPX4	LOC_Os11g18170	1.029±0.027	1.086±0.085	2.130±0.099	1.137±0.078
	CAL	LOC_Os02g02400	1.464±0.470	4.382±0.417	5.804±0.571	24.368±1.883

Citrate synthase, CS; Oxoglutarate dehydrogenase, ODH; Malate dehydrogenase, MDH; Glutamine synthetase, GS; Glutamate dehydrogenase, GDH; Glucose-6-phosphate dehydrogenase, G6PDH; Ascorbate peroxidase, APX; Glutathione peroxidase, GPX; Catalase, CAL.

CHAPTER 4 TOWARDS THE FUNCTIONAL CHARACTERISATION OF BARLEY GLUTAMYL-TRNA SYNTHETASES

4.1 Abstract

To investigate whether roles of glutamyl-tRNA synthetase are conserved in the anther development among cereals, I identified the homologs of OsERS1 (Glutamyl-tRNA Synthetase 1) in another important crop, barley (*Hordeum vulgare* L.). Sequence analysis showed that barley genome has three genes encoding GluRS, named *HvERS1*, *HvERS2* and *HvERS3* with 81-87% amino acid identity to OsERS1. Phylogenetic analysis showed that HvERS1 and HvERS2 are grouped in the same clade with OsERS1 and OsERS2, while HvERS3 and OsERS3 are in another clade, suggesting the duplication and diversification of OsERS1 during evolution. HvERS1 and HvERS2 share some common features with OsERS1, such as containing a GST_C (C terminus of glutathione S-transferase) domain at the N terminus, β-barrel anticodon binding domain, and predicted cytosolic localisation. RT-qPCR analysis showed that *HvERS1* and *HvERS2* are highly expressed in young inflorescences, flowers, anthers and developing grains, implying their role during reproduction, which is similar to that of *OsERS1* in rice. Consistently, the promoter sequence of *HvERS1* contains the *cis*-elements potentially active during reproductive stage. To understand the genetic role of *HvERS1* and *HvERS2,* I attempted CRISPR/Cas9-based gene editing for these two genes. This work is useful for elucidating the role of glutamyl-tRNA synthetase in male reproduction.

4.2 Introduction

Barley (*Hordeum vulgare* L.) is one of the most important staple crops. Australia produces high quality barley, with annual production averaging around eight million metric tonnes/year (data source http://aegic.org.au/publications/australian-grains/barley/). It is a widely grown crop and occupies a large geographic area from southern Queensland to Western Australia. The grain yield of barley is largely determined by spikelet fertility (Del Moral et al., 2002). It has long been realised that fertilised

grain number per spike is determined by the flower organ development and pollen activity, so that the reproductive phase is crucial for increasing the yield. Therefore, deep understanding about the regulators of barley reproductive development will provide molecular mechanisms to improve the yield. Anthers, the male reproductive organs, are essential for fertilisation and grain formation. As a member of the grass family, barley shares common cytological characteristics with rice during anther development (see Chapter 1). Similarly, the barley male reproductive development can also be divided into 14 stages (Abiko et al., 2005; Gómez and Wilson, 2012).

Compared with rice, fewer regulators of anther and pollen development have been reported in barley. One known regulator of barley anther development is HvGAMYB (Murray et al., 2003), a homolog of rice GAMYB (Liu et al., 2010), which is highly expressed when the four layers of anther wall are formed, and throughout anther development. During pre-meiotic stages, HvGAMYB was localised in the nuclei of four anther wall layers: epidermis, endothecium, middle layer and tapetum, indicating its role in anther wall development (Murray et al., 2003). In late stages of barley anther development, HvGAMYB protein still remained in anther wall cells. Over-expression of *HvGAMYB* resulted in shortened and indehiscent anthers in transgenic barley lines. In addition, the expression of *HvGAMYB* was induced by GA_3 (gibberellin), following a conserved pathway in cereal aleurone (Murray et al., 2003; Gubler et al., 1999). During stages 8−10, when the tapetum is undergoing programmed degradation, a PHD (plant homeodomain) transcription factor HvMS1 (MALE STERILITY1), a homolog of *Arabidopsis* MS1 (Wilson et al., 2001) and rice OsPTC1 (PERSISTENT TAPETAL CELL1) (Li et al., 2011), was shown to be specifically expressed in the tapetum. Silencing *HvMS1* and overexpression of *HvMS1* in barley both reduced male fertility (Gómez and Wilson, 2014). As for pollen development, a homolog of *Arabidopsis* DEFECTIVE IN EXINE FORMATION1 (DEX1) has been identified in barley based on sequence alignment, yet its function remains unclear (Ma et al.,

2013).

The molecular mechanism of barley male sterility remains largely unclear due to lack of fundamental genomic sequence. The recent release of a high-quality barley reference genome will definitely speed up the progress on the research on barley reproduction (International Barley Genome Sequencing Consortium, 2012; Mascher et al., 2017). Comparative genomics analysis, combined with newly developed CRISPR-Cas9 gene editing technology in barley (Lawrenson et al., 2015), will prompt both basic research and precisely targeted breeding.

This chapter aimed to investigate the molecular characteristics of barley glutamyl-tRNA synthetase (GluRS) genes, and explore the roles of barley GluRSs in male reproductive organ development. Barley homologs of three rice GluRS (OsERS1, OsERS2 and OsERS3) were searched by BLASTp program and a phylogenetic analysis was performed. Their expression pattern was investigated by RT-qPCR combined with sequencing data from IPK database (http://webblast.ipk-gatersleben.de/barley_ibsc/). I performed protein modelling for rice and barley GluRSs, respectively, to compare their tertiary structures. Finally, I transformed a popular barley cultivar 'Golden Promise' using CRISPR-Cas9-HvERS-sgRNA constructions to create knockout mutants of *HvERS* genes. Previously developed CRISPR-Cas9 gene editing technology in barley by Lawrenson et al. (2015) used the *Cas9* sequence from bacteria *Streptococcus pyogenies*, lacking a feature of the 5' GC enrichment in Gramineae genomes (Wong et al., 2002). Here I chose the CRISPR/Cas9 Multi-Targeting Vector containing a plant codon optimised *Cas9* gene with higher GC content (Ma et al., 2015). This system has been tested to be efficient in creating mutants in *Arabidopsis*, rice (Ma et al., 2015; Wang et al., 2015; Liang et al., 2016), populus (Fan et al., 2015), tomato (Pan et al., 2016) etc.

4.3 Results and Discussion

4.3.1 Characterisation of barley *Glutamyl-tRNA Synthetase* genes

A homolog search based on BLASTp using OsERS1 as a query identified three GluRSs in barley,

encoded by *HvERS1* (HORVU3Hr1G038360), *HvERS2* (HORVU1Hr1G078370), and *HvERS3* (HORVU6Hr1G012710). The open reading frames (ORF) of *HvERS1*, *HvERS2* and *HvERS3* are 2,160 bp, 2,163 bp and 1,722 bp, encoding 719, 720 and 573 amino acid residues (aa), respectively. Notably, both HvERS1 and HvERS2 showed higher sequence similarity with OsERS2 (87% identity) compared with that of OsERS1 (81% identity). The phylogenetic analysis of 24 homologs from other monocots, dicots, moss and green algae indicated that OsERS1-homologs were highly divergent and were grouped into two clades based on the protein sequences (same with results from Chapter 3, Fig 3; Chapter 4, Fig 4.1A). Similar to OsERS1 and OsERS2, HvERS1 and HvERS2 are grouped into Clade I GluRS, with the typical GST_C domain, GluRS catalytic core and tRNA-synt_C (tRNA synthetase class I anti-codon binding domain). In addition, the barley HvERS1 and HvESR2 proteins were more similar to those in *Brachypodium distachyon* (e.g. XP014757177 and XP003566911) than those in rice and maize (Fig 4.1A). HvERS3 is in Clade II GluRS, with only catalytic core, which was grouped in the same clade of OsERS3, showing only catalytic domains (Fig 4.1A; Chapter 3, Fig 3). These results indicated that gene number and evolutionary pattern of *GluRS* family genes are conserved between rice and barley.

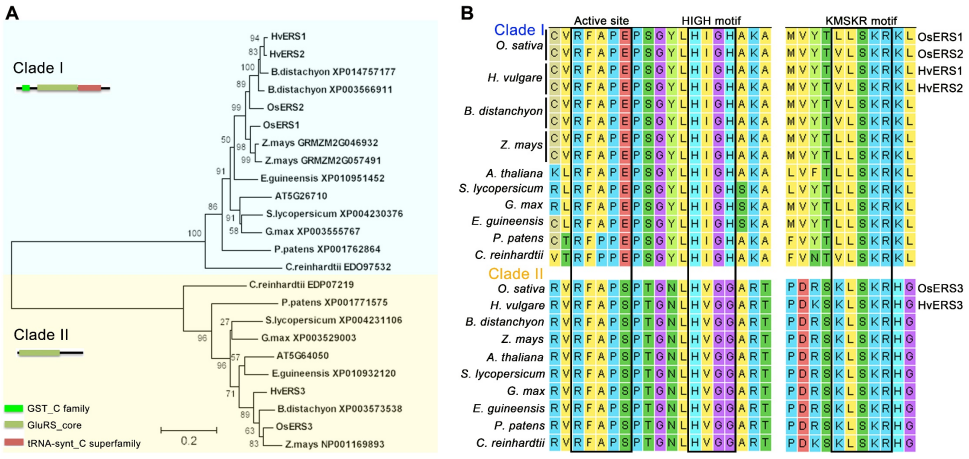


Figure 4.1 Phylogenetic and motifs analyses of plant GluRSs

(A) A rooted phylogenetic tree of a selection of plant GluRSs. The topologic structure was shown and distances were indicated. Sequences used for this analysis were from *C. reinhardtii*, EDP07219, EDO97532; *P. patens*, XP_001771575.1,

XP_001762864.1; *A. thaliana*, AT5G64050.1, AT5G26710.1; *S. lycopersicum*, XP_004231106.1, XP_004230376.1; *G. max*, XP_003555767.1, XP_003529003.1; *O. sativa*, Q8S5M6 (OsERS1), Q9LIZ8 (OsERS2), Q6YUS0 (OsERS3); *B. distachyon*, XP_003573538.1, XP_003566911.1, XP_014757177.1; *Z. mays*, GRMZM2G046932_T01, GRMZM2G057491_T01, NP_001169893.2; *E. guineensi*, XP_010932120.1, XP_010941452.1; *H. vulgare*, M0XNG3 (HvERS1), F2EEU4 (HvERS2) and Q43768 (HvERS3).

(B) The MUSCLE alignment of three key motifs of plant GluRSs. The active site RFAPE, HIGH motif and KMSKS motif were highlighted. The divergent motif sequences between two clades were consistent with the phylogenetic analysis.

My alignment of three key motifs, one active site RFAPE (RFAPS in Clade II), two essential catalytic motifs HIGH (HVGG in Clade II) and KMSKR (V/LLSKR for 14 homologs in Clade I and KLSKR for 10 homologs in Clade II), showed highly conserved patterns within each clade (Fig 4.1B). Notably, in Clade I, the active site and HIGH motif exhibited 100% identity among rice, barley, *B. distachyon* and maize, and the substitution of valine from lysine in the KMSKR motif happened in OsERS2, GluRSs of barley and *B. distanchyon*, but not in maize or dicots (Fig 4.1B, upper). In Clade II, the active site and two catalytic motifs showed 100% identity both in monocots and dicots (Fig 4.1B, lower). In short, three barley GluRSs exhibit high amino acid similarity and evolutionary conservation with rice GluRSs, suggesting their conserved and critical biochemical functions during evolution.

4.3.2 Protein modelling of rice and barley GluRSs

To further understand the molecular characteristics of GluRSs at protein level, I performed structural modelling for six GluRSs by SWISS-MODEL (Fig 4.2). Due to lack of proper eukaryotic template, templates from *Escherichia coli* (SMTL ID: 1qru.1 and 1qrt.1) and *Thermotoga maritima* (SMTL ID: 2o5r.1), which shared 37 ~ 41% sequence similarity with cereal GluRSs, were selected based on homology comparison. The modellings of four Clade I GluRSs, including OsERS1, OsERS2, HvERS1 and HvERS2, displayed catalytic domains (Fig 4.2), but failed to create the structure of N-terminal GST_C domain because of the lack of this domain in the templates.

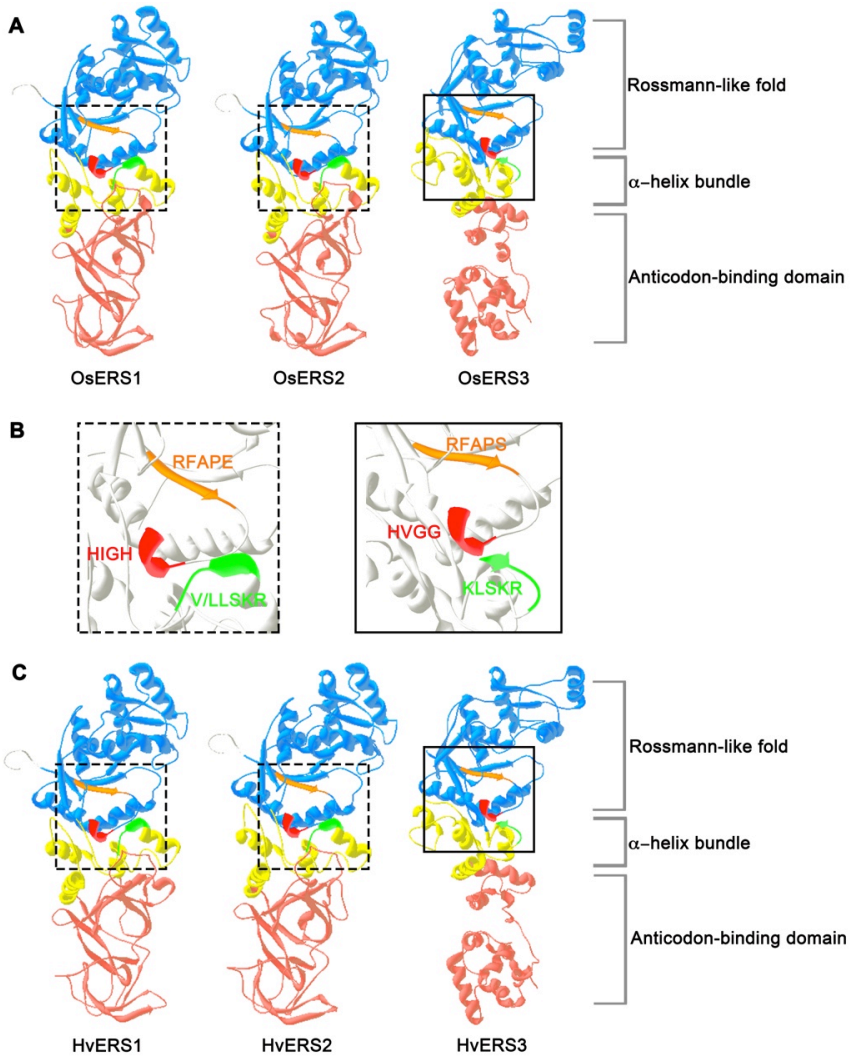


Figure 4.2 Protein modelling of GluRSs in rice and barley

(A) Structures of OsERS1, OsERS2 and OsERS3.

(B) Amplification of catalytic centres. The square of dashed lines indicates catalytic centre from cytoplasmic GluRSs, and the square of solid lines indicates the catalytic centre from organellar GluRSs.

(C) Structures of HvERS1, HvERS2 and HvERS3.

Areas in squares are catalytic centres with three key motifs. Aqua, Rossmann-like fold; yellow, α-helix bundle; coral, anticodon-binding domain; orange, RFAPE motif; red, HIGH motif; green, KMSKR motif.

As a typical Class I aminoacyl-tRNA synthetase, GluRS has a Rossmann-like fold and HIGH and KMSKR signatures in its catalytic domain (Ibba and Söll, 2000). Consistent with the phylogenetic analysis of primary amino acid sequences, the tertiary structures of the six GluRSs were divided into

two subgroups. The four GST_C-domain-containing GluRSs showed similar structures, with β-barrel type anticodon-binding domains, while OsERS3 and HvERS3 formed α-helix type anticodon-binding domains (Fig 4.2A and C). The β-barrel type anticodon-binding domain was characterised in deciphering mechanism of anticodon loop recognition in GlnRS (glutaminyl-tRNA synthetase) (Rould et al., 1991) and archaeal GluRS (Nureki et al., 2010). Supportively, plant Clade I GluRS and other eukaryotic GluRSs are evolutionarily close to GlnRS of *E.coli* (Woese et al., 2000). The α-helix type anticodon-binding domain is a common feature of both discriminating GluRS and non-discriminating GluRS in bacteria, indicating the different origin of plant Clade II GluRS than Clade I (Sekine et al., 2001; Schulze et al., 2006). Despite the distinguishing pattern of anticodon-binding domains, the two subgroups displayed similar catalytic centres, in which RFAPE, HIGH and KMSKR formed cavities, with a subtle difference in orientation between the two subgroups (Fig 4.2B). The subgroups of putative tertiary structures are consistent with phylogenetic analysis, motif alignment and subcellular localisation prediction, all of which suggested that cytosolic GluRSs and organellar GluRSs are differently originated.

4.3.3 Expression pattern of *HvERS* genes

To understand the roles of *HvERS* genes in barley, I investigated expression patterns for the three genes. Searching RNA-sequencing data provided by IPK Barley BLAST Server (http://webblast.ipk-gatersleben.de/barley_ibsc/), I extracted the expression pattern of three *HvERS* genes in various tissues, including embryo, seedling, the developing inflorescence and flower organs (Fig 4.3A). *HvERS3* showed its expression peak in etiolated seedling, and HvERS3 protein was predicted to localise in chloroplasts (Table 4.1), which are in agreement with the role of GluRS in catalysing the initial step of chlorophyll biosynthesis (Schön and Söll, 1988; Von Wettstein et al., 1995). HvERS1 and HvERS2 were more likely localised in cytosol (Table 4.1), and RNA expression peaks of *HvERS1* and *HvERS2* were present in developing inflorescence and grain (Fig 4.2A), indicating more involvement of these

two GluRSs in reproductive development.

Table 4.1 Putative subcellular localisation of HvERS proteins

		Predotar				TargetP			
	Length(aa)	Mito	Chl	ER	Elsewhere	Chl	Mito	SP	other
HvERS1	719	0.01	0.46	0.11	0.48	0.099	0.038	0.596	0.348
HvERS2	720	0.01	0.59	0.00	0.41	0.334	0.076	0.233	0.438
HvERS3	560	0.36	0.30	0.05	0.42	0.733	0.616	0.005	0.005

Predotar (https://urgi.versailles.inra.fr/predotar/) and TargetP (http://www.cbs.dtu.dk/services/TargetP/) were used to predict subcellular localisation. Mito, mitochondria; Chl, chloroplast; SP, secretary pathway; ER, endoplasmic reticulum.

To validate the RNA-sequencing data, RT-qPCR was performed to further investigate the expression characteristics of *HvERS* genes, especially in reproductive stages (Fig 4.3B). I extracted total RNA from vegetative tissues including leaf, stem, shoot and root, and reproductive organs including flower and the developing spike meristems of barley cultivar, Golden Promise. Synthesised cDNA was used for RT-qPCR and *HvActin7* (HORVU5Hr1G039850) served as an internal control. These three *HvERS* genes showed a relatively lower expression level in vegetative tissues, compared with that of reproductive organs (Fig 4.3B). During the early spike development stages W3−6 (Waddington scale, Waddington et al., 1983), *HvERS1* and *HvERS2* showed similar expression levels, but higher than that of *HvERS3* (Fig 4.3B), indicating that *HvERS1* and *HvERS2* are likely to be involved in early reproductive development. In later flower development, *HvERS1* exhibited dominantly high expression level in young and mature flowers, suggesting its roles in reproductive events. In anthers, *HvERS1* and *HvERS2* also showed similar levels, suggesting that the two genes may have redundant functions in barley male organ formation. In addition, although *HvERS3* was highly expressed in green spike (Fig 4.3B), it was more likely demanded my chlorophyll biosynthesis in the outer whorls of floral organs rather than participating in reproductive development. Therefore, these results suggest that *HvERS* genes have roles in barley reproductive events and male fertility.

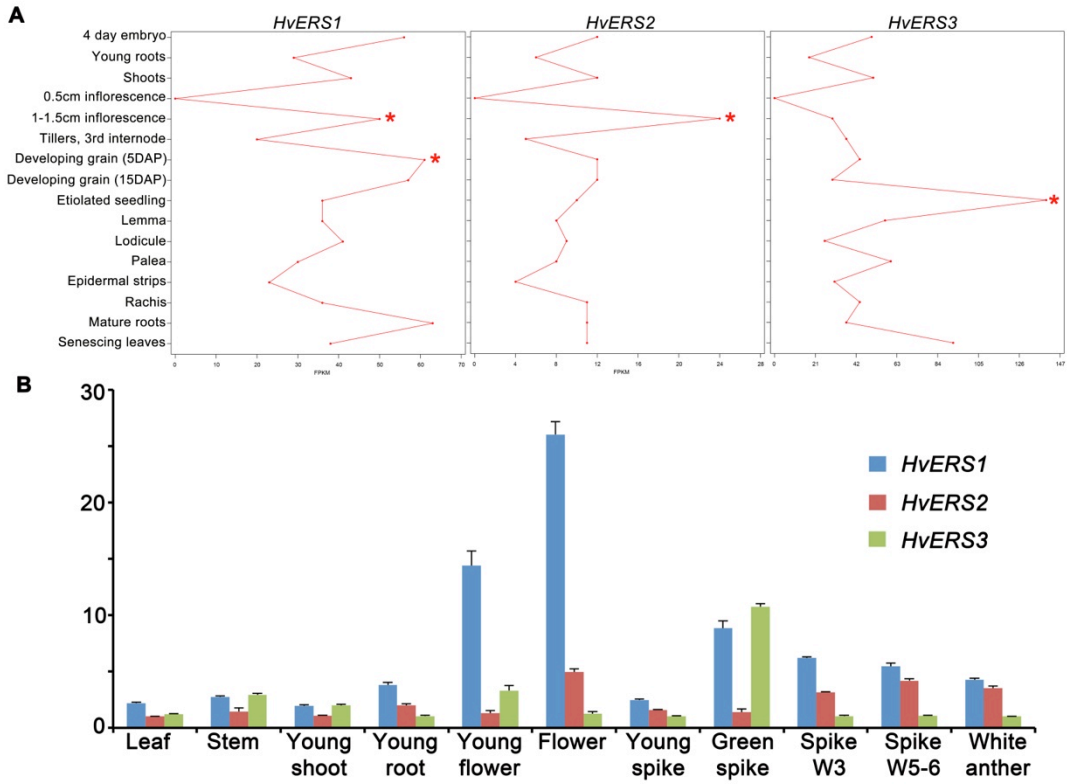


Figure 4.3 Expression pattern analysis of *HvERS* genes

(A) Expression levels of *HvERS* genes from IPK. Red asterisks indicate significant expression peaks.

(B) RT-qPCR analysis for *HvERS* genes. Spike W3 and W5−6 stages were defined by Waddington scales (Waddington et al., 1983).

4.3.4 *cis*-Acting regulatory elements in *GluRS* genes' promoters

cis-Acting regulatory elements are important molecular switches, which are involved in the transcriptional regulation of gene expression networks and activities controlling various biological processes in plants, including abiotic stress responses, hormone responses and developmental processes (Lescot et al., 2002). The differences of expression pattern between GluRS genes are likely due to different transcriptional regulation. To compare the potential regulatory *cis*-elements within promoters of *OsERS* and *HvERS* genes, 1.5-kbp upstream sequence to the start codons of *OsERS1, OsERS2, OsERS3, HvERS1, HvERS2* and *HvERS3* were used *in silico* analysis by PlantCARE (Lescot et al., 2002) (Fig 4.4). Motifs putatively involved in plant hormone responses, defense and stress

responses and development-specific activities were found in those promoter regions. The promoters of *HvERS* genes enriched more methyl jasmonate (MeJA) responsiveness elements while *OsERS* genes' promoters have more gibberellin (GA) responsiveness elements, indicating their possible roles as responsive components in JA signalling and GA signalling pathway respectively. Both *OsERS2* and *HvERS2* promoters contain motifs of heat-stress and low-temperature responsiveness, suggesting their potential involvement in adaption of thermal change.

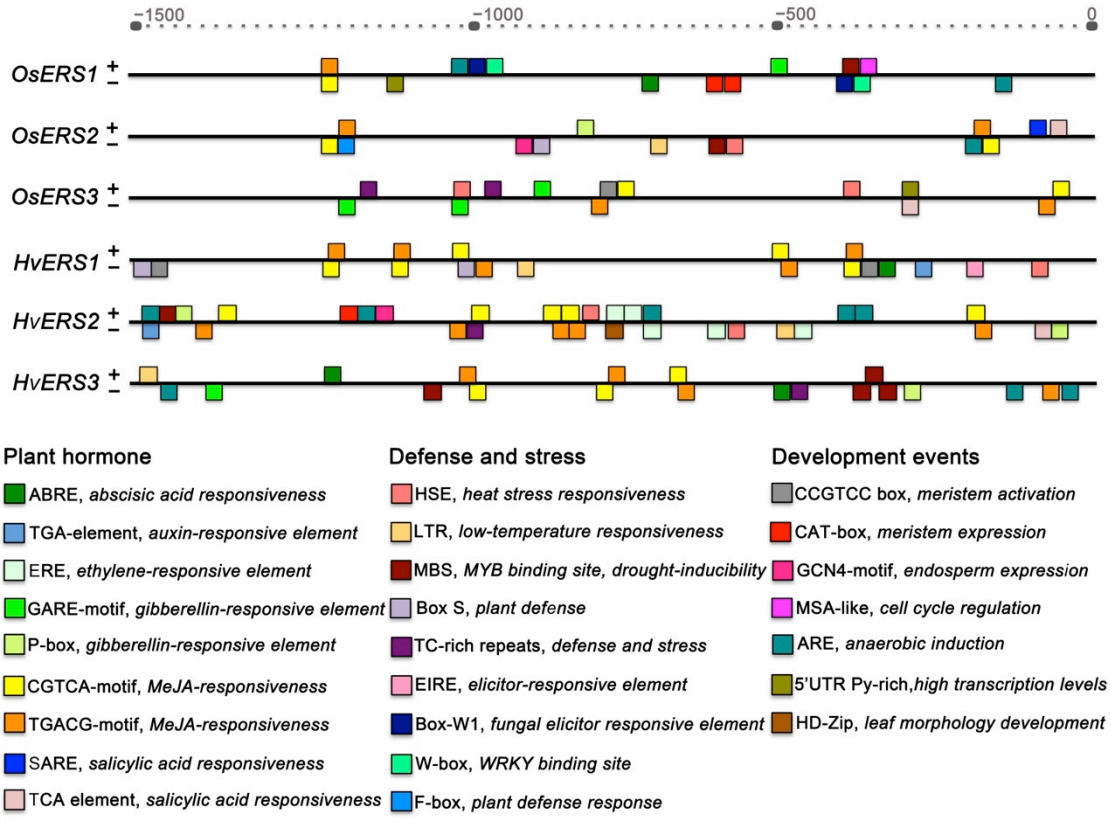


Figure 4.4 The conserved motifs distributed in promoter area of GluRS genes.

DNA motifs are identified by PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare). Each motif is indicated by a specific colour. The size of the coloured cubes does not indicate the length of motifs.

Importantly, as shown in Figure 4.4, the isolated promoters contain various well-known developmental event sequence motifs. Some of these *cis*-acting elements were shown to be required for meristem- and endosperm-specific expression, including CCGTCC-box, CAT-box and GCN4

motif. One anaerobic induction motif, one meristem expression motif and one cell cycle regulation motif are identified in *OsERS1* promoter, which is agreement with its roles in early anther cell division and patterning within a hypoxic niche (Chapter 3). One meristem activation box (CCGTCC) and meristem expression box (CAT) are also identified in *HvERS1* and *HvERS2* promoter regions, respectively, indicating the regulatory activities of expression in reproductive tissues, which is consistent with high expression level of *HvERS1* and *HvERS2* in young root and spike meristem.

4.3.5 Creating knockout mutants for *HvERS1* and *HvERS2*

To further uncover functions of its homologs HvERS1 and HvERS2 at barley reproductive development, I designed target-specific knockouts using the CRISPR/Cas9 system. The high sequence similarity (94% identity) between *HvERS1* and *HvERS2* allowed very limited choices to specifically knock out *HvERS1*. Also, considering the non-catalytic GST_C domain, I chose one target sequence close to active sites specific for *HvERS1*, and a second target sequence for both *HvERS1* and *HvERS2* (Fig 4.9). The two targets were synthesised in chimeric PCR primers and cloned into binary vector pYLCRISPR/Cas9P$_{Ubi}$-H as described in Materials and Methods (Ma et al., 2015). Vectors verified by Sanger sequencing were used to transform immature embryos from the barley cultivar 'Golden Promise' (Fig 4.5).

pYLCRISPR/Cas9P$_{Ubi}$-H -T1
TGAGGTGGGTTTTTATGATTAGAGTCCCGCAATTATACATTTAATACGCGATAGAAAACAAAATATAGCGCGCAAACTAGGATAAA
TTATCGCGCGCGGTGTCATCTATGTTACTAGATCGGGAGCACCGGTAAGGCGCGCCGTAGTGCTCGACTAGTATGGAATCGGCA
GCAAAGGACGCGTTGACATTGTAGGACTATATTGCTCTAATAAAGGAGGCAGCTATGCTGGCCGTCGTTTTACAACGTCGTGACT
GGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCA
CCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCTAATTTTTTCCTGTAGTTTTCCCACAACCATTTTTTACCATCCGAATG
ATAGGATAGGAAAAATATCCAAGTGAACAGTATTCCTATAAAATTCCCGTAAAAAGCCTGCAATCCGAATGAGCCCTGAAGTCTGA
ACTAGCCGGTCACCTGTACAGGCTATCGAGATGCCATACAAGAGACGGTAGTAGGAACTAGGAAGACGATGGTTGATTCGTCAG
GCGAAATCGTCGTCCTGCAGTCGCATCTATGGGCCTGGACGGAATAGGGGAAAAAGTTGGCCGGATAGGAGGGAAAGGCCCAG
GTGCTTACGTGCGAGGTAGGCCTGGGCTCTCAGCACTTCGATTCGTTGGCACCGGGGTAGGATGCAATAGAGAGCAACGTTTAG
TACCACCTCGCTTAGCTAGAGCAAACTGGACTGCCTTATATGCGCGGGTGCTGGCTTGGCTGCCGTTTTGCTCCAGAGCCTAGC
GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTCAA
GAGCTTGGAGTGGATGGATACGCGTCGGTATCATTGGCGCGCCTCTCGAGCTAGCGGCCGCATGCATCGATCTCCTACATCGTT
ATGCCCCGC

pYLCRISPR/Cas9P$_{Ubi}$-H -T2
CATATAATTTCTGTTGAATTACGTTAAGCATGTAATAATTAACATGTAATGCATGACGTTATTTATGAGGTGGGTTTTTATGATTAGA
GTCCCGCAATTATACATTTAATACGCGATAGAAAACAAAATATAGCGCGCAAACTAGGATAAATTATCGCGCGCGGTGTCATCTAT
GTTACTAGATCGGGAGCACCGGTAAGGCGCGCCGTAGTGCTCGACTAGTATGGAATCGGCAGCAAAGGACGCGTTGACATTGTA
GGACTATATTGCTCTAATAAAGGAGGCAGCTATGCTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAA
CTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTG
CGCAGCCTGAATGGCTAATTTTTTCCTGTAGTTTTCCCACAACCATTTTTTACCATCCGAATGATAGGATAGGAAAAATATCCAAGT
GAACAGTATTCCTATAAAATTCCCGTAAAAAGCCTGCAATCCGAATGAGCCCTGAAGTCTGAACTAGCCGGTCACCTGTACAGGC
TATCGAGATGCCATACAAGAGACGGTAGTAGGAACTAGGAAGACGATGGTTGATTCGTCAGGCGAAATCGTCGTCCTGCAGTCG
CATCTATGGGCCTGGACGGAATAGGGGAAAAAGTTGGCCGGATAGGAGGGAAAGGCCCAGGTGCTTACGTGCGAGGTAGGCCT
GGGCTCTCAGCACTTCGATTCGTTGGCACCGGGGGTAGGATGCAATAGAGAGCAACGTTTAGTACCACCTCGCTTAGCTAGAGCA
AACTGGACTGCCTTATATGCGCGGGTGCTGGCTTGGCTGCCGCTGAAAGTTTGATTAAGCAGTTTTAGAGCTAGAAATAGCAAGT
TAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTCAAGAGCTTGGAGTGGATGGATACGC
GTCGGTATCATTGGCGCGCCTCTCGAGCTAGCGGCCGCATGCATCG

Figure 4.5 Sequences of target sgRNA expression cassettes.

Two target sgRNA expression cassettes were cloned into pYLCRISPR/Cas9PUbi-H and sequenced. Target 1 and Target 2 are green shaded and yellow shaded, respectively. The red 'G' before targets is transcription starting site for *OsU6* promoter, which is gray shaded. The cyan shaded sequence is *LacZ*.

Agrobacterium AGL1-mediated transformation was performed as described previously (Harwood, 2014). Callus induced from at least 60 immature embryos were further selected by hygromycin-containing medium. T_0 transgenic seedlings were regenerated from hygromycin-resistant callus (Fig 4.6). In total, 38 and 45 individual T_0 transgenic lines were created for Target 1 and Target 2, respectively. These lines were aimed to contain deletion mutations in target sequences of *HvERS1* and *HvERS2*. However, no mutation was identified in all T_0 transgenic lines, indicating the failure to knock out target genes (Fig 4.7).

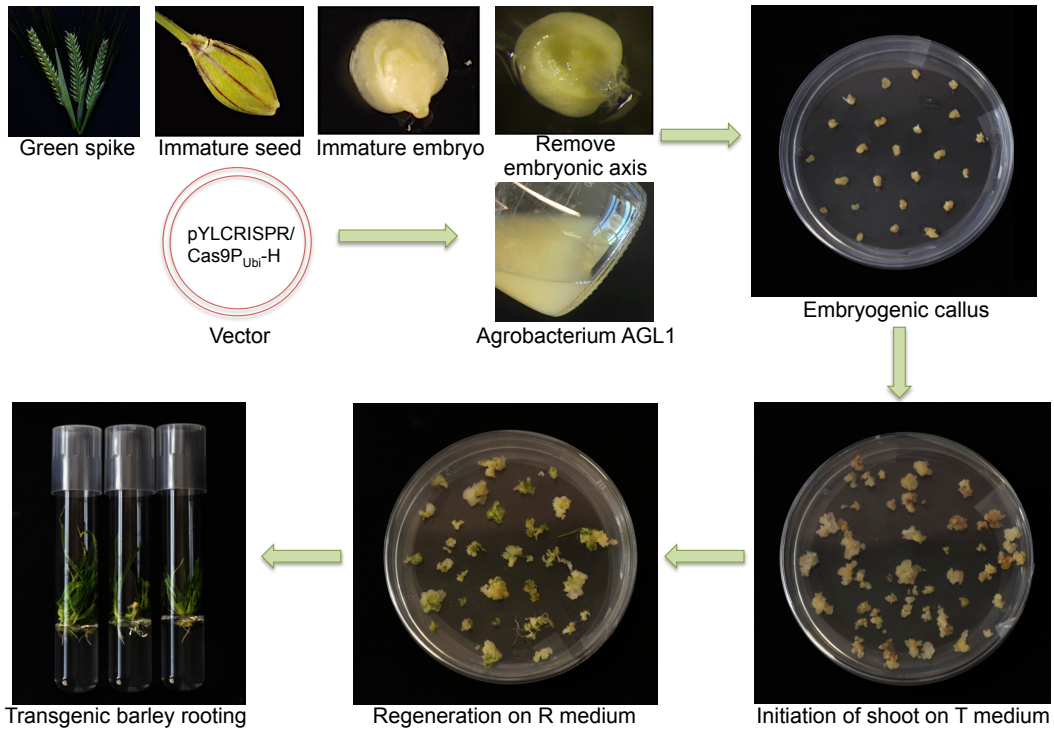


Figure 4.6 Production of T_0 transgenic lines (photos were taken during transformation progress**)**

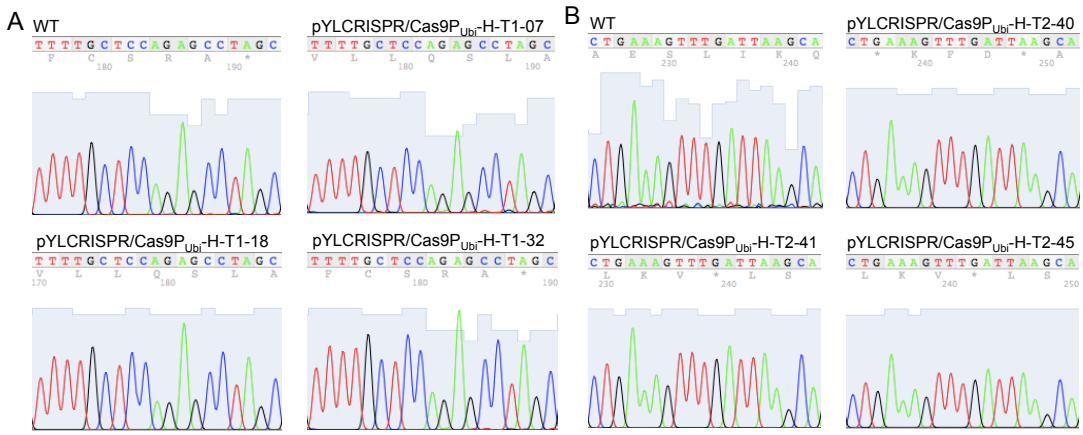


Figure 4.7 Sanger sequencing results of T_0 transgenic barley

(A) Examples of Sanger sequencing results of pYLCRISPR/Cas9P$_{Ubi}$-H-T1transformed barley.

(B) Examples of Sanger sequencing results of pYLCRISPR/Cas9P$_{Ubi}$-H-T2 transformed barley.

The possible reasons for the pYLCRISPR/Cas9 system showing no editing in *HvERS1* and *HvERS2* are summarised. The major one is target sequence selection. The developers of pYLCRISPR/Cas9 system observed that target sequences with higher GC contents (50−70%) were edited with higher efficiency (Ma et al., 2015). In Gramineae genomes, sequences with high GC content sequences tend to be in the 5' terminal regions of genes (Wong et al., 2002). The targets selected for *HvERS1* and *HvERS2* are 1.1−1.4 kbp downstream of the start codon and have GC contents below 50%. This could have led to failure in creating mutations. In experiments conducted by other members of our laboratory using the same pYLCRIPSR/Cas9 system, target sequences close to or even including the start codon ATG of barley *MADS*-box genes have very high editing efficiency (80−90%) (data provided by Dr. G. L. and H. NJ. K.). Another determinant of editing efficiency is the expression level of sgRNA (Ma et al., 2015). I used *LacZ*-containing sgRNA expression cassettes, while *LacZ*-free cassettes were used in barley *MADS*-box gene editing mentioned above. It is still unclear whether the *LacZ* has any negative effects on sgRNA expression level in barley, but as an obvious foreign genetic element, *LacZ* might attract gene silencing factors (Vaucheret et al., 1998; Matzke and Matzke, 1998).

4.4 Materials and Methods

4.4.1 Sequence search and phylogenetic analysis

The OsERS1 full-length amino acid sequence was used as a query for BLASTp against the EnsemblPlant database (http://plants.ensembl.org), NCBI (https://www.ncbi.nlm.nih.gov) and UniProt (http://www.uniprot.org/) to obtain amino acid sequences of 24 OsERS1 homologs from barley (*Hordeum vulgare*), maize (*Zea mays*), soybean (*Glycine max*) etc. MEGA (version 7.0) and the neighbor-joining methods were used with a p-distance model and pairwise deletion and bootstrap. The max parsimony method of MEGA was also used to support the NJ tree, using the default parameters. Motifs alignment used the MUSCLE tool and default parameters (Edgar, 2004).

4.4.2 RT-qPCR

Total RNA from barley various tissues was extracted using the TRIZOL reagent (Invitrogen) (Simms et al., 1993). Single-stranded cDNA was synthesised from 2 µg of total RNA by iScript Select cDNA Synthesis Kit (Bio-Rad). SYBR Fast Universal qPCR Kit (KAPA) was used for RT-qPCR on a QuantStudio Flex 6 (Life Technologies) machine. The expression level of *ACTIN* (*ACT*) mRNA was used as the internal control. For expression analysis of *HvERS1*, *HvERS2* and *HvERS3*, cDNA was denatured at 95 °C for 10 minutes (min), followed by 42 cycles at 95 °C for 10 s, 60°C for 10 s, and 72 °C for 10 s. The RT-qPCR results shown are from representatives of experiments repeated at least three times with similar results. The primers used are listed in Appendix 3.

4.4.3 Protein modelling

Since no plant GluRS structure has been reported, I performed modelling for GluRSs of rice and barley using SWISS-MODEL (https://swissmodel.expasy.org) (Biasini et al., 2014). Models of OsERS1 (Uniprot ID: Q8S5M6), OsERS2 (Uniprot ID: Q9LIZ8), OsERS3 (Uniprot ID: Q6YUS0), HvERS1 (Uniprot ID: M0XNG3), HvERS2 (Uniprot ID: F2EEU4) and HvERS3 (Uniprot ID: Q43768) structures were generated with two templates from *Escherichia coli* (SMTL ID: 1qru.1 and 1qrt.1) and

one template from *Thermotoga maritima* (SMTL ID: 2o5r.1).

4.4.4 CRISPR-Cas9-HvERS-sgRNA constructions

The pYLsgRNA-OsU6/LacZ and pYLCRISPR/Cas9P$_{Ubi}$-H vectors are provided by Prof. Yaoguang Liu in South China Agricultural University, China. The primers used are listed in Appendix 3.

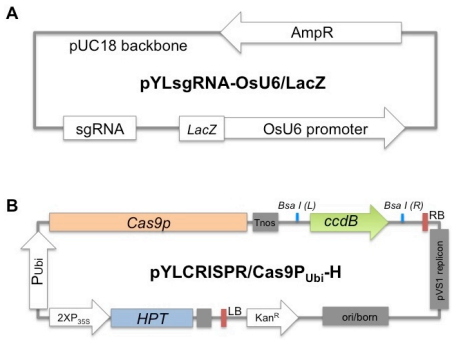


Figure 4.8 Vectors used in CRISPR/Cas9 gene editing system

(A) Structure of pYLsgRNA-OsU6/LacZ. This vector is used as PCR template for sgRNA (single guide RNA) expression cassette. *LacZ* can be used for selection in *Escherichia coli* for positive clones. The *U6* promoter sequence is from rice to drive small nuclear RNA expression.

(B) Structure of binary vector pYLCRISPR/Cas9PUbi-H. This vector is used for barley transformation. *Cas9* is driven by the maize *ubiquitin* promoter (P$_{Ubi}$). This vector contains a hygromycin selection marker and two *Bsa* I sites in flanking sequence of a *ccdB* gene.

i) Target sequences to knock out *HvERS1* and *HvERS2*

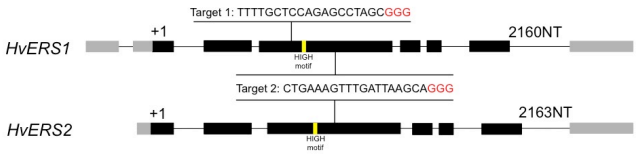


Figure 4.9 Target sequences selection for CRISPR/Cas9 gene editing system

Target 1 was selected specifically for *HvERS1*, and Target 2 was for both *HvERS1* and *HvERS2*. The GGG (red) was protospacer adjacent motif for Cas9 enzyme recognition. The yellow areas are sequences encoding HIGH motifs.

ii) Production of sgRNA expression cassette

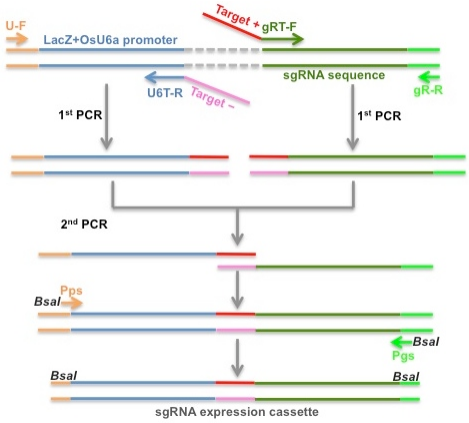


Figure 4.10 Production of sgRNA expression cassette

The 1st PCR was performed in one reaction with U-F/U6T-R and gRT-F/gR-R primer pairs, to generate two fragments, *LacZ* + *U6* prmoter + Target and Target + sgRNA. The 2nd PCR was overlapping PCR to produce the whole sgRNA expression cassette with two flanking *BsaI* sites.

iii) Construction of CRISPR-Cas9-HvERS-sgRNA.

The sgRNA expression cassettes containing Target 1 and Target 2 sequences were ligated into pYLCRISPR/Cas9PUbi-H using two different *Bsa* I sites (L and R, Fig 4.9). The cutting-ligating reaction system was:

Reagent	**Volume (μl)**	**Concentration**
10×CutSmart Buffer (NEB)	1.5	1×
pYLCRISPR/Cas9	60-80 ng	4~6 ng/μl
sgRNA expression cassette	10-15 ng	
Bsa I-HF (NEB)	10 U	~1 U/μl
T4 DNA ligase (NEB)	35 U	2~3 U/μl
10×T4 DNA ligase buffer (NEB)	1.5	1×
H_2O	up to 15μl	

37 °C 5 min
10 °C 5 min } 15 cycles
20 °C 5 min
37 °C 5 min

Then the ligation product was transformed to *E. coli* competent cells (DH5α) for positive clone selection. PCR-detected positive clones were further sequenced.

4.4.5 Barley transformation and identification of transgenic knockout lines

The barley transformation procedure followed a standard protocol (Harwood, 2014), and all media and stock solutions were prepared exactly same as the formulas in this protocol.

i) Isolation of barley immature embryo and *Agrobacterium* cocultivation

Plants of the barley cultivar 'Golden Promise' were grown in controlled-environment growth chambers in Waite Campus, University of Adelaide, with 15 °C daytime (16 hours) and 10 °C night (8 hours), 70~80% relative humidity. Barley spikes were collected from plants usually two weeks after flowering when the immature embryo size reached ∅1.5−2 mm. All the sterilisation and isolation steps were performed in a laminar flow hood. The young seeds were removed from spike and sterilised by 70% ethanol solution for 1 min and rinsed three times in sterile distilled water. Then 10% sodium hypochlorite stock solution was used for further 15 mins sterilisation followed by four-six times water rinses. A pair of fine forceps was used to remove the seed coat and the embryonic axis from exposed embryo. About 100 isolated embryos were placed in a 10×10 array on CI (callus induction) medium. 10 ml overnight culture of *Agrobacterium* clone with correct vector (CRISPR-Cas9-HvERS1/2-sgRNA) was prepared for cocultivation. Each embryo was covered by a small drop of culture and transferred to fresh CI medium after excess culture liquid had dried. The cocultivation lasts 3 days at 21 °C in dark growth cabinet.

ii) Selection and regeneration of transgenic lines

After 3 days of cocultivation, embryos were transferred to fresh CI medium with hygromycin and timentin for induction of callus and selective growth. Calli were transferred to fresh medium at 2 weeks and 4 weeks. At 6 weeks, the calli were transferred to T (transition) medium with hygromycin and timentin, and kept at 21 °C in low light. After 2 weeks, the calli, were transferred to R (regeneration) medium for shoot induction, and kept at 21 °C in light until new strong roots grew in hygromycin. The regenerated plants were transferred to Root medium in tubes, and to cocopeat soil in small pots and large pots as transgenic lines grew.

iii) Identification of transgenic lines

A round shaped leaf sample ∅5 mm was collected from each transgenic plant after one-week growth in soil. The leaf samples were lysed to release DNA for PCR detection by Phire Plant Direct PCR Kit (Thermo Scientific). The PCR products were identified by Sanger-sequencing to find out any knockout mutations. The primers used are listed in Appendix 3.

CHAPTER 5 SUMMARY AND FUTURE DIRECTIONS

5.1 Summary of outcomes

Possessing the ability to ligate amino acid to cognate tRNA, aminoacyl-tRNA synthetases (aaRSs) are indispensable components for genetic information translation from mRNA to protein. Widely existing in three kingdoms of organisms, aaRSs have been researched in multiple facets: biochemical activity, structure, evolution, non-canonical functions, and novel antibiotics development (Carter, 1993; Cusack, 1997; Ibba and Söll, 2000; Guo et al., 2010; Vondenhoff and Van Aerschot, 2011; Guo and Schimmel, 2013). Compared with the research progresses of aaRSs in bacteria and mammals, plant aaRSs have been less understood.

5.1.1 Identification and characterisation of plant aminoacyl-tRNA synthetases

As reported in Chapter 2, we identified all aaRSs proteins of *Arabidopsis*, rice and *P. patens* (Chapter 2, Table 1). Based on phylogenetic analysis and paralogue grouping, we speculated that the plant *aaRS* genes may have originated from both eukaryote and organelle ancestors, and that gene duplication/loss events within the *aaRS* family contributed to the changes of members among species (Chapter 2, Fig. 2; Table 1). We observed conserved motifs that are essential for catalytic function and proposed how they may have evolved from prokaryotes (Chapter 2, Fig. 3). In addition to the aminoacylation domain, we identified domains such as WHEP, oligonucleotide binding motif and tRNA binding domain, which may enhance or expand the functions of aaRSs (Chapter 2, Fig. 4). The plant aaRSs are dispatched to cytosol, mitochondria and chloroplast using dual targeting strategies (Chapter 2, Fig. 5). Protein modelling of cytosolic and organellar aaRS shows distinct structural characteristics, consistent with their different origins (Chapter 2, Fig. 6 and Fig. 7). The expression profiling of *aaRS* genes from *Arabidopsis* and rice also shows diverged expression preference between cytosolic and organellar ones. This study provides an informative resource of plant aaRSs and insights for evolutionary history for this family in plants.

5.1.2 Roles of a glutamyl-tRNA synthetase in rice early anther development

In higher plants, stem cells in the anther primordium differentiate into four layers of somatic cells that constitute the anther wall and centrally localised germinal cells. Although a few regulators controlling the fates of germinal and somatic cells have been reported, how plants maintain metabolic homeostasis in organ primordial niche remains largely unknown.

In the research reported in Chapter 3, we investigated the roles of a rice GluRS, OsERS1, in early anther development. The *osers1* anthers displayed over-proliferation of somatic and germinal cells and disarranged anther wall layers (Chapter 3, Fig 1). *OsERS1* was expressed in anther cells undergoing rapid periclinal division, supporting the phenotype of *osers1* anthers (Chapter 3, Fig 3). Then we verified the conserved activity of OsERS1 in ligating glutamate to $tRNA^{Glu}$ (Chapter 3, Fig 4). The enzymatic activity can be enhanced by complexation with the cofactor OsARC (aminoacyl-tRNA synthetase co-factor) (Chapter 3, Fig 5). Non-targeted metabolomics analysis showed that OsERS1 is important for multiple metabolic pathways including intracellular glutamate dynamics, TCA cycle, and *osers1* anthers show the accumulation of all amino acids and TCA components and active aerobic metabolites such as H_2O_2. Notably, treatment of H_2O_2 reconstituted the *osers1*-characterised phenotype in wild type (Chapter 3, Fig 6 and Fig 7). Our findings reveal novel roles of OsERS1 in modulating plant organ development, likely through the maintenance of metabolic homeostasis and redox status (Chapter 3, Fig 8).

5.1.3 Characterisation of glutamyl-tRNA synthetases in barley

Although barley and rice share similar anther development progresses and numerous homologous genes, the molecular basis of these processes has hardly been researched in barley due to the limited availability of reference genomic sequences. With release of barley genome assemblies and with the availability of new gene editing technology, both fundamental research and practical breeding will be greatly prompted in barley. Based on our studies of OsERS1, I investigated its homologs from barley.

As reported in Chapter 4, there are three barley GluRSs, encoded by *HvERS1*, *HvERS2*, and *HvERS3*. These form two clades: HvERS1 and HvERS2, which have additional GST_C domains and are putatively mainly localised in cytosol, and HvERS3, which has only one catalytic domain and is putatively localised in mitochondria and chloroplast (Chapter 4, Fig. 4.1 and Table 4.1). The differentiation between two clades of GluRSs is also reflected in the results of protein modelling analysis (Chapter 4, Fig. 4.2). *HvERS3* is predominantly expressed where chlorophyll synthesis is required, while *HvERS1* and *HvERS2* are more likely involved in reproductive development (Chapter 4, Fig. 4.3). The *cis*-elements in promoters link these rice and barley GluRS genes to multiple regulatory pathways, such as plant hormone responses, defense system, and development processes (Chapter 4, Fig. 4.4). To explore possible conserved roles of *HvERS1* and *HvERS2* in anther development, I tried to create knockout mutants using CRISPR/Cas9 gene editing technology (Chapter 4, Fig. 4.5 to 4.7). Although the trial failed due to inappropriate target sequences and other uncertain issues, the experiences will contribute to further application of gene editing in barley.

5.2 Future directions

5.2.1 Exploration of plant multi-synthetase complex

In many organisms ranging from Archaea to Eukarya, aaRSs are assembled as a high-molecular-weight complex, so called multi-synthetase complex (MSC) (Kim et al., 2013; Laporte et al., 2014). For example, human MSC is formed by MetRS, LeuRS, GluRS-ProRS, IleRS, ArgRS, LysRS, GlnRS, AspRS, and three scaffold proteins p43, p38, p18 (Guo and Schimmel, 2013); yeast MSC contains GluRS, MetRS and cofactor Arc1p (Simos et al., 1996). In higher plants, the existence of MSC and its components are still unclear. We have verified the interaction between rice GluRS and its cofactor OsARC (homolog of yeast Arc1p and human p18) in Chapter 3, which are potential members of plant MSC. The aaRSs with additional domains like WHEP and GST_C are more likely to be recruited. By protein technology like yeast-2-hybrid and co-immunoprecipitation, uncovering the structure of plant

MSC will provide insights to protein synthesis and other intracellular events in the plant kingdom.

5.2.2 Structural basis of plant aminoacyl-tRNA synthetases

Crystallography technology greatly prompts studies of aaRSs and their complexation with substrates and even interactive proteins, detailing how these enzymes function in the molecular level (Cusack, 1997; Jeong et al., 2000). However, the structure has not yet been solved for any plant aaRS. This limits the understanding about the aaRS family in plants. Crystal structures of some unique aaRSs of plants will fill this gap and provide more evidence for comparative biology and evolutionary analysis.

5.2.3 Mechanisms of GluRS in plant development and signalling

As we demonstrated in Chapter 3, the putative housekeeping gene *OsERS1* is expressed in specific cell groups of rice anther primordia, regulating cell behaviour through metabolic homeostasis and redox status. This suggests that expression of *OsERS1* is under the regulation of specific transcription factors during the anther development. Showing no significant change of expression level in reported mutants like *msp1-4, ostdl1a, gamyb-4, tip2*, *OsERS1* seems to be in an independent pathway. Considering the anaerobic induction motif, meristem expression motif and cell cycle regulation motif identified in its promoter sequence, *OsERS1* is likely to respond to physiological status such as oxygen concentration and/or positional cues. Using yeast-1-hybrid and transcriptome data, we may narrow down the candidates and expand the mechanisms of GluRS in the anther development. In addition, GluRS is an important target of metabolic regulation in plant cells (Cotelle et al., 2000; Coruzzi and Zhou, 2001). Increasing research findings support the idea that metabolism can actively affect cell fate and organ development (Zhang et al., 2018). OsERS1 could be further investigated as an entry point for how metabolism modulates anther development.

OsERS1 is also highly expressed in young roots (Chapter 3, Fig. 3). Under L-glutamate treatment, wild type root growth was repressed, while *osers1* root seemed insensitive to extra glutamate (data not shown). Further study of OsERS1 may reveal more roles of GluRS in areas like glutamate signalling,

cell division/differentiation, and meristematic activities.

ABBREVIATIONS

AaRS	aminoacyl-tRNA synthetase
α-KG	α-ketoglutarate
ATP	adenosine triphosphate
BiFC	bimolecular fluorescence complementation
bp	base pair
BSA	bovine serum albumin
cDNA	complementary DNA
DAB	3,3'-diaminobenzidine
DTT	dithiothreitol
eGFP	enhanced green fluorescent protein
GSH/GSSG	glutathione reduced/oxidised
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
Kb	kilo base
KI	potassium iodide
MBP	maltose binding protein
mRNA	messenger RNA
MSC	multi-synthetase complex
OAA	oxaloacetate
PAGE	polyacrylamide gel electrophoresis
PMC	pollen mother cell
PPC	primary parietal cell
PPP	pentose phosphate pathway
RT-qPCR	Reverse transcription quantitative polymerase chain reaction
ROS	reactive oxygen species
SDS	sodium dodecyl sulfate
SPC	secondary parietal cell
TCA	tricarboxylic acid
YFP	yellow fluorescent protein
Y2H	yeast-two-hybrid

APPENDICES

Appendix 1

Supplemental Table S1 (Chapter 3)

Sequences for the phylogenetic analysis of GluRSs.

>E. coli WP_000695649
MKIKTRFAPSPTGYLHVGGARTALYSWLFARNHGGEFVLRIEDTDLERSTPEAIEAIMDGMNWLSLEWDEGPYYQTKRFDRYNAVIDQMLEEGTAYKCYCSKERLE
ALREEQMAKGEKPRYDGRCRHSHDHHADDEPCVVRFANPQEGSVVFDDQIRGPIEFSNQELDDLIIRRTDGSPTYNFCVVVDDWDMEITHVIRGEDHINNTPRQINI
LKALKAPVPVYAHVSMINGDDGKKLSKRHGAVSVMQYRDDGYLPEALLNYLVRLGWSHGDQEIFTREEMIKYFTLNAVSKSASAFNTDKLLWLNHHYINALPPE
YVATHLQWHIEQENIDTRNGPQLADLVKLLGERCKTLKEMAQSCRYFYEDFAEFDADAAKKHLRPVARQPLEVVRDKLTAITDWTAENVHHAIQATADELEVGMG
KVGMPLRVAVTGAGQSPALDVTVHAIGKTRSIERINKALAFIAERENQQ
>T. elongatus WP_011056357
MTVRVRLAPSPTGNLHIGTARTAVFNWLYARHRGGKFILRIEDTDRERSRPEYTENILEGLQWLGLTWDEGPYFQSDRLDLYRQAIQTLLDKGLAYYCYCTPEELEA
LRAEQKAKGQAPRYDNRHRHLTPEEQAAFEAAGRTPVIRFKIEDDRQIEWQDLVRGRVSWQGADLGGDMVIARAAPRGEIGYPLYNLVVVVDDIAMGITDVIRGE
DHIGNTPKQILLYEALGATPPNFAHTPLILNSTGQKLSKRDGVTSISDFRAMGYLAPALANYMTLLGWSPPEGVGELFTLDLAAKHFSFERINKAGARFDWDKLNW
LNRQYIQQLEPEEFLAELIPLWQGAGYAFDEERDRPWLFDLAQLLQPGLNTLREAIDQGAVFFIPSVTFDSEAMAQLGQPQSATILAYLLEHLPAEPALTVAMGQQLI
QQAAKAAGVKKGATMRTLRAALTGAVHGPDLMAAWQILHQRGWDEPRLAAALKQAQTTS
>T. thermophilus WP_014630090
MVVTRIAPSPTGDPHVGTAYIALFNYAWARKNGGRFIVRIEDTDRARYVPGAEERILAALKWLGLSYDEGPDVGGPHGPYRQSERLPLYRRYAEELLKRGWAYRAF
ETPEELEQIRKEKGGYDGRARNIPPEEAEERARRGEPHVIRLKVPRPGTTEVKDELRGVVVYDNQEIPDVVLLKSDGYPTYHLANVVDDHLMGVTDVIRAEEWLV
STPIHVLLYRAFGWETPRFYHMPLLRNPDKTKISKRKNHTSLDWYKAEGFLPEALRNYLCLMGFSMPDGREIFTLEEFIEAFTWERVSLGGPVFDLEKLRWMNGKY
IREVLSLEEVAERVKPFLREAGLSWESEAYLRRAVELMRPRFDTLKEFPEKARYLFTEDYPVSEKAQRKLEEGLPLLRELYPRLRAQEEWTEAALEALLRGFAAEKG
VKLGQVAQPLRAALTGSLETPGLFEILALLGKERALRRLERALA
>N. caninum NP_014609
MLRVRYSPQHPPLATLCVSVLLERGLAACVPDNKLQGSLLIVQKSDNAPEEFTTDVAAAKILCALSPKTLLLYPAAVPSPSFEDACATACIDAMLSLAATLQAASSLSP
KQLGSLSSHLQLRTFLCGYHLTLADLAVYAQLRRLACTEQHAGAPPAGWKDKFVHVSRWYSFLHGQQSTANAVGVALRKAGSTATRAESQKNEDGKKKAERANA
NPSSPTAGDKKGTASHASYEGKLEGAVQGKVVTRFPPEPSGYLHIGHAKAALLNSYFAKKYNGKMLFRFDDTNPARENEEFESSIAEDLRLLNVEWAKISHTSDYFE
EMQSLCERLIKEGKFYVDDTPAEVMREQRAEGVESCRRNASVEESLQRWAEMLKGSAEGQKCCVRAKMDMQSKNKCMRDPVMYRCVADCPHHRHGDRFKAYP
TYDFACPVVDSIEGVTHALRTNEYADRIPQYQWVQQAAGLAPVHIYEFSRLCFVKTLLSKRKLKQFVDTGLVEGWDDPRMPTVRGIRRRGLQVEALLEFILEQGPS
KAGNLMEWDKLWTKNKQIIDPIVPRFMAVGKDAVPVFIKGGPETVESKKRRMHAKNESLGEADLLLLNKILIDRDDAAMCENGEEVTLMHWGNCIFDTVVKNAA
GEVTEIQATLHLEGDFRKTKKKLHWLANLSGLSAAPAQSAELVLRELDHLITVDKLDQEEGDWEKFINPQTLFDTPAIGDPLLRQLREGDLLQLERRGYFRVDQAG
DRLVLIKIPDGRSKAMSAVGTKVDAAKLSGAKITGKK
>S. cerevisiae NP_014609
MIMLRIPTRSYCSPSKLIKGVGLSPLKKSLLSKKIKEDIHPSLPVRTRFAPSPTGFLHLGSLRTALYNYLLARNTNGQFLLRLEDTDQKRLIEGAEENIYEILKWCNINY
DETPIKQSERKLIYDKYVKILLSSGKAYRCFCSKERLNDLRHSAMELKPPSMASYDRCCAHLGEEEIKSKLAQGIPFTVRFKSPERYPTFTDLLHGQINLQPQVNFND
KRYDDLILVKSDKLPTYHLANVVDDHLMGITHVIRGEEWLPSTPKHIALYNAFGWACPKFIHIPLLTTVGDKKLSKRKGDMSISDLKRQGVLPEALINFCALFGWSP
PRDLASKKHECFSMEELETIFNLNGLTKGNAKVDDKKLWFFNKHFLQKRILNPSTLRELVDDIMPSLESIYNTSTISREKVAKILLNCGGSLSRINDFHDEFYYFFEKP
KYNDNDAVTKFLSKNESRHIAHLLKKLGQFQEGTDAQEVESMVETMYYENGFSRKVTYQAMRFALAGCHPGAKIAAMIDILGIKESNKRLSEGLQFLQREKK
>S. cerevisiae NP_011269
MPSTLTINGKAPIVAYAELIAARIVNALAPNSIAIKLVDDKKAPAAKLDDATEDVFNKITSKFAAIFDNGDKEQVAKWVNLAQKELVIKNFAKLSQSLETLDSQLNLRT
FILGGLKYSAADVACWGALRSNGMCGSIIKNKVDVNVSRWYTLLEMDPIFGEAHDFLSKSLLELKKSANVGKKKETHKANFEIDLPDAKMGEVVTRFPPEPSGYL
HIGHAKAALLNQYFAQAYKGKLIIRFDDTNPSKEKEEFQDSILEDLDLLGIKGDRITYSSDYFQEMYDYCVQMIKDGKAYCDDTPTEKMREERMDGVASAIRDRSV
EENLRIFTEEMKNGTEEGLKNCVRAKIDYKALNKTLRDPVIYRCNLTPHHRTGSTWKIYPTYDFCVPIVDAIEGVTHALRTIEYRDRNAQYDWMLQALRLRKVHIW
DFARINFVRTLLSKRKLQWMVDKDLVGNWDDPRFPTVRGVRRRGMTVEGLRNFVLSQGPSRNVINLEWNLIWAFNKKVIDPIAPRHTAIVNPVKIHLEGSEAPQEP
KIEMKPKHKKNPAVGEKKVIYYKDIVVDKDDADVINVDEEVTLMDWGNVIITKKNDDGSMVAKLNLEGDFKKTKHKLTWLADTKDVVPVDLVDFDHLITKDRLE
EDESFEDFLTPQTEFHTDAIADLNVKDMKIGDIIQFERKGYYRLDALPKDGKPYVFFTIPDGKSVNKYGAKK
>E. pallida KXJ21118
MWTWSSLLVIPVRSHKLTLTAIIQARILCTTTASSTMKTYNVVASSTKPPIATLLTVEHLKSSGLVSLQHGDSTTLKLNDNFTITGDHNVARHLARVATGKGLYGNSPL
QASEVDHWLEFSXRCQGNDXQXEKTLQQLNKNLAPVTYLVGSSVTLADFAVWGMIKGNGNLHNLLNNSTSFQHLTRWFNFCCSQSPFSVVSKDLPKTDXGNDSG
ENIMGTKMKQEGKYVDLPGAEMGKVVTRFPPEASGYLHIGHAKAALLNQYYKNIYNGKLIMRFDDTNPAKESSEFEQVILKDLEMLDFKFDMFSHTSDHFDLLLE
YAEKMLKKGQAYVDDTPTETMREEREQRIKSKNRDNSVEKNLQDWEEMKKGSERGTKCCVRAKIDYTSLNGALRDPTIYRCKPESHVKTGTKYKVYPVYDFACP
IVDSIEGVTHALRTMEYHDRDPQYYWFIETLGLRKVQIYEFSRLNLQNTVMSKRKLTYLVDQGYVDGWDDPRFPTVRGVLRRGMTVEGLREFIILQGSSRATVQM
EWDKIWSINKKVIDPVAPRYTALKKDTIVPYNLPGAKEESQEQPLHPKNADVGTKKVWYSPKVFIEGEDAETLKVGDMVTLMNWGNTKVTKINRNSKGKIDSIDG
ELHLDNKDYKKTTKLTWLGDTSNAPLLPVVMVHFDNVISKAILTKTDDFKDHINKNSKTETSLLGESELAKLKKGDIIQLQRRGYFICDQPYEPPSRHTGRASPCVL
FSIPDGHTKSQAQPAAKSTKNESKKNEKGGKKEKPSTTPQAKATASPEEVEKLKADIAKQGDKVRELKTSGAAKDQVTAEVAALLKLKASYKELTGEDFAPAGKG
KQDKKKEKMGDNQAAKPVSVAASSKEAEKLKEAITKQGDKVRELKTSGAAKDQVDAEVQTLLSLKAQYKELTGEDLAPAGKGKKDKKKDEKKQEKKTNNSKK
SEEAAKSATADDGDSKKITRLCLEAKKAENLSDWYSQVITKSEMIEYYDVSGCYILRAWSYAIWEKIKDFFDAEIKKLGVENVYFPIFVSQAALEKEKDHIADFAPE
VAWVTKSGQSELAEPIAIRPTSETVMYPAYARWVQSHRDLPLKYNQWNNVVRWEFKHPQPFLRTREFLWQEGHTAYADKEGACDEVYTILDLYAKVYTDLLAIPV
VKGRKTEKEKFAGADFTTTVEAYISASGRGIQGATSHHLGQNFSKMFDISFEDPDPSVQQKRYVYQNSWGLTTRTIGVMCMVHGDDQGLVLPPRVASVQVVIVPCG
LTVNLSDEEKDVLMKRCAEYETALQEAGIRVKGDFRENYSPGWKFYHWELKGVPIRLEVGPRDVKNNQFVLVKRNDRTKKIVKSEDLVNDVQTTLEKIHTEMLA
KAQSDLDAHLGVTTSWQEFLSMLDNKKIIQAPFCGDIPCEDKIKKESTREESAEPGAPSMGAKALCIPFDQPAEITSECKCVCPECEKPAKFYTLFGRSY
>C. elegans NP_499917
MKLTGFLKQNVRVRFAPSPTGHLHIGGLRTAFFNYLFAKKYGGDFILRIEDTDRTRFVDDAQDQIYSSLNFYNLLPDEGPREGGKFGPYEQSKRLEIYRNAAYRLIDS
GHAYRCFCSENRLDLLRKTAEKRGEIPKYDRKCANLSSRDAVKMEQNGEKFVIRFKLDKQNVQFHDEVFGSVNQFIDESDPVLLKSDGFPTYHLANVIDDRKMEIS
HVIRGMEWLSSTGKHTILYKAFNWTPPKFVHLSLIMRSATKKLSKRDKDAFVSYYSEQLGALPEAVLNLMIRNGAGIRNFDAEHFYSLDEMIEQFDLSLLGRRNLLL
DSDVLQKYSRMAFQKSDFKELYPRIIDILNKKSNYSTSREDIQKIVTFLKAKEENFGFLSSLSTEFSWFFTRPQSSQLLKESHPNVDLRNILNSLLEIEVFNSESLEYLA
KNHQLNLAKAMGIVRISLIGSKKGPPISELVEFFGMTECHRRIRIMQELL
>C. elegans NP_492711
MMITRELVLKANREQPPYASILALAASGFSLEKSVQFSEKQQLALNLDGELLSNDVEIARIIAQSTDAADSLLGSSIIDFAVVDDLTNFIADAVTKNDYSLLLGKDFPT
KILDNSLTVADFAIFSVAHNNPQLKAKFSGIIDKVLKEPTLAAAHNFVGLYKSAAAPTATASTGKEKKKDEGKFVELPGAEKGKVVVRFPPEASGYLHIGHAKAALL
NQYYQQAFEGQLIMRFDDTNPAKENAHFEHVIKEDLSMLNIVPDRWTHSSDHFEMLLTMCEKLLKEGKAFVDDTDTETMRNEREQRQDSRNRSNTPEKNLQLWE
EMKKGSPKGLTCCVRMKIDMKSNNGAMRDPTIYRCKPEEHVRTGLKYKVYPTYDFTCPIVDSVEGVTHALRTTEYHDRDDQYYFICDALGLRRPHIWEYARLNM
TNTVMSKRKLTWFVDEGHVEGWDDPRLPTVRGVMRRGLTVEGLKQFIVAQGGSRSVVMMEWDKIWAFNKKVIDPVAPRYTALDSTSPLVSIELTDSISDDTSNVSL
HPKNAEIGSKDVHKGKKLLLEQVDAAALKEGEIVTFVNWGNIKIGKIEKKGAVITKISATLQLDNTDYKKTTKVTWLGDVKAEAGKTIPVVTADYDHIISKAIIGKD
EDWKQFINFDSVHYTKMVGEPAIKNVKKGDIIQIQRKGFYIVDQPYNPKSELSGVETPLLLIAIPDGHTGKEAEKPPKGAPVSAAASSGGSDALQLYSNIEQQGGIVR
DLKTKDAKSQATKDAIAKLLDLKKKYKELTGSDHKPGAPPSAPAPTTPSSSESNSALNIYNLIEAQGLLVRELKGKDAKSQATKDAIAKLLELKKQYKEVSGSDHKP
GVPPAASVASAPAPAPAAGSTNALDIYHQIESQGALVRELKGKDAKSQATKDAIAKLLALKKQYKEVTGSDHKPGVVPVSAPAPTPLAGGADIAGQIEAQGVLVRD
LKNKDSKSQETKDAIAKLLELKKNYKEVTGSDYKPGPAPAAAPAKVTVPAPSVSGGDTLSKQIDDQALLVRELKMKDAKSQETKDAIAKLLQLKKQYKDATGSDY
KPAPAQLAASAAAAPAPAPAFDEAALLKEIEEQGAVVRDAKSKDPKSQDSADAINKLLALKANFKKATGKDFPAPSNAQSKKGKKK

>D. magna KZS14262
MFSKFSRRFFHLSCRTRNVRVRFAPSPTGYIHLGGLRTALYNFLFAKANNGQFILRIEDTDQKRLVPGAANQLEMDLDWMGLSPDESPMKNGTFGPYTQSERLCLYN
QFVVELIKSGHAYHCFCSERRLEFVKKEAIREGAIPRYDNRCRQLSSSQVEEKKQHGMTSCIRFKLSSLEEPLQDLVYGDVDRNAVLHEGDPVIMKQDGYPTYHLA
NVIDDHLMGITHVLRGTEWLVSTAKHVLLYKAFGWHPPRFAHLPLITNADGTKLSKRQDDIRLDYFRSQGFYPEAIIGLLCLVGGGFSSSIDIQQNLHSLPELAKSFN
LGCVTTHSGRLDMTRLDNFNRLALQNRISQPEKMDEMRKQLSHLVYENIPNSSIDKEMLQDASLLKVLHWASSEGRIHLLRQLTDPHMGFLWSSQRNPIISKIDPST
ARNVIVRVVEILNQYDRIEKKEVSAMLRQIALEHNLKVNEVMKLMRSSLTGLKEGPPVSEIIEILGPKICSCRLVSSLNLIV
>D. magna KZS22036
MAVTLSVSKNNVSIGVFLLSEYVKPTLPVNLTWGKDFSLKINNDLILNSPSSCCRYFARTSGSLNLYGITLLERTEVDHWISFASCMQGNDKDFETIEKALHMATYLV
GEKISAADLVLWGCLFGSDQWQSKKQNSDNYVNIKRWFKLVESQPLVMKAIKSLPDDVKVNNRASPSHVATTLGRKEEGKFVDLPGAEEGKVVVRFPPEASGYLH
IGHAKAALLNQYYQQLFKGTLIMRFDDTNPAKENAEFEKVILEDLEMLQIKYDKLTHTSDHFDLIERYCEQMLKEGKAYVDDTDPEIMKQEREQKTESKNRGNSVE
KNFTLWKEMKNGTDLGQKCCVRAKIDMQSVNGCMRDPTMYRCKPEPHVRTGYKYKVYPTYDFACPIVDSIEGVTHALRTTEYHDRDDQYYWFIDALKLRKPYI
WEYARLSMINTVLSKRKLTWFVQNGHVDGWDDPRFPTVRGVLRHGMTVEGLKQFIIAQGSSRSVVQMEWDKIWSLNKKIIDPIAPRYTALERNTAKVHVRGVKET
CTTAALHPKNPDVGSKDVWVSDQILIDLADVESLRAGENATFINWGNIRITSVNRNKDEVSIEAEPNLEDKDYKKTPKLTWIAVTEKAKPTPCVCVFFDHIISKPVLA
KDEDFKQYINTNTRTEVVMLGDPELVKVRKGDIIQLQRRGFFICDSPYEPISAHSCREAPIVLFSIPDGSKPAEAAKSTVAATSKKGPVKSEASPKTKTVEKAAVSKEI
DVSFLDEQIGKQGEKVRELKASKADKSQIDSAVKELLSLKSDFKLATGMDWKAGLTVPNSSEKNDEKPKTEQLDINKLDGDIKTCGDKVRDLKTKKADKVQIDAA
VKELLSLKAAFKESFGIEWGPNVASSLPQANDAKPESSVTNEESLSSQIKECGDKVRDLKAKKAGKSEIDSAVKALLDLKATYKAATGKEWTPQAVTKPVNESPPIA
PATVSASTGKNEDILCAQITECGNKVRDLKTKKAEKSEIDAAVKDLLSLKAAYKTATGLDWTPTATATKENKPAATTAIKKDSTDKKTEKKKEAKPKDKKVEEKKV
EETTSGLKKQTRLGLEFKKDENLSDWYSQVITKAEMIEYYDVSGCYILRPWAYAIWESIKDFFDAEIKKLGVQNCYFPMFVSHQALEKEKTHIADFAPEVAWVTKS
GNSDLAEPIAIRPTSETVMYPAYAKWIQSHRDLPLKLNQWNNVVRWEFKHPQPFLRTREFLWQEGHTAFATAKEAEEEVYQILELYSQVYTHLMAIPVIKGRKTEKE
KFAGGDFTTTVEAFISASGRAIQGATSHHLGQNFSKMFEIVVEDPETKEKKHIYQNSWGITTRTIGVMVMVHSDDKGLVLPPNVASVQVVIVPCGINVNVTEDEKSN
LMAACMEYEKVLSAAGIRVKGDYRDNYSPGWKFNHWELKGVPLRLEIGPRDVKQGEYVAVRRDTSQKLIQSKASACEDIKQLLEAIQKALLDKATQELENHTVIV
ENFRDFIPELDKKNIILSPFCGEISCEEKIKKESTRDDGAEPGAPAMGAKSLCIPFQQPAELNSNATCIYPGCPNKPKYYCLFGRSY
>D. melanogaster AAM29426
MLKINPRMLMQQCLRHAHSQVRVRFAPSPTGYLHLGGLRTALYNFLYARHLGGKFLLRIEDTDQTRLVPGASERLVEDLLWAGIEIDEGPGFGGQLGPYVQSQRTEI
YSKAIEELLHNGTAYRCFCTERRLDLLRKEALRTRQVPRYDNKCRHLTTSEVDSLLAKGTPYCIRFKLTDHEEPLDDLIYGKVHHNVSDNEGDPVVMKSDQFPTYH
FANVVDDHLMGITHVLRGVEWQISTTKHLLLYKAFGWQPPKFGHLPLLVNADGTKLSKRQGDIGIQHFRERGYFPQALVNYVVSAGGGFEHRTNAKQQLLSMQA
LHEQFHIERVNSHPSRLNPELLDDLNRLEIVQRLSGNESRTKLVNEVQQLVKEAYPQHSNLDLAEEHILDVLNWSSKRLTLLQDLTSSKLSFLWVKPSDFQIKDLTAE
QLAHLLQLINAIQDFHKDELNVKLKSFAQKENIKFPLMMKTLRAALSGLKEGPGVAEMMEILGKSVTLERLKEALPKESLHLEEDQ
>D. melanogaster NP_524471
MSIKLKANLNNPPISGLATAHLINGTVPVEIVWSKEETSLQFPDNRLLVCHSNNDVLRALARAAPDYKLYGETAIERTQIDHWLSFSLTCEDDISWALSFLDKSIAPVT
YLVANKLTIADFALFNEMHSRYEFLAAKGIPQHVQRWYDLITAQPLIQKVLQSLPEDAKVKRSPQSSKEQTPAKTGERKQEGKFVDLPGAEMGKVVVRFPPEASGY
LHIGHAKAALLNQYYALAFQGTLIMRFDDTNPAKETVEFENVILGDLEQLQIKPDVFTHTSNYFDLMLDYCVRLIKESKAYVDDTPPEQMKLEREQRVESANRSNS
VEKNLSLWEEMVKGSEKGQKYCVRAKIDMSSPNGCMRDPTIYRCKNEPHPRTGTKYKVYPTYDFACPIVDAIENVTHTLRTTEYHDRDDQFYWFIDALKLRKPYI
WSYSRLNMTNTVLSKRKLTWFVDSGLVDGWDDPRFPTVRGIIRRGMTVEGLKEFIIAQGSSKSVVFMNWDKIWAFNKKVIDPIAPRYTALEKEKRVIVNVAGAKVE
RIQVSVHPKDESLGKKTVLLGPRIYIDYVDAEALKEGENATFINWGNILIRKVNKDASGNITSVDAALNLENKDFKKTLKLTWLAVEDDPSAYPPTFCVYFDNIISKA
VLGKDEDFKQFIGHKTRDEVPMLGDPELKKCKKGDIIQLQRRGFFKVDVAYAPPSGYTNVPSPIVLFSIPDGHTKDVPTSGLKVNAPDAKATKKASSPVSSSGQASE
LDSQISQQGDLVRDLKSKKAAKDQIDVAVKKLLALKADYKSATGKDWKPGQTSASSAPVPAASSSSANDAVSVNASIVKQGDLVRDLKGKKASKPEIDAAVKTLLE
LKAQYKTLTGQDWKPGTVPTTAAPSASAAPSVGVNDSVAQILSQITAQGDKVRELKSAKADKATVDAAVKTLLSLKADYKAATGSDWKPGTTAPAPAAAPVKVK
QEKNPDPASVLTVNTLLNKIAQQGDKIRQLKSAKSEKSLVEAEVKLLLALKTDYKSLTGQEWKPGTVAPAPTTVNVIDLTGGDSGSDVGSVLSKIQAQGDKIRKLKS
EKAAKNVIDPEVKTLLALKGEYKTLSGKDWTPDAKSEPAVVKKEASPVSMASPAKDELTQEINAQGEKVRAAKGNKAAKEVIDAEVAKLLALKAKYKEVTGTDF
PVAGRGGGGGGGSAKKAPKEAQPKPAKPVKKEPAADASGAVKKQTRLGLEATKEDNLPDWYSQVITKGEMIEYYDVSGCYILRQWSFAIWKAIKTWFDAEITRM
GVKECYFPIFVSKAVLEKEKTHIADFAPEVAWVTKSGDSDLAEPIAVRPTSETVMYPAYAKWVQSYRDLPIRLNQWNNVVRWEFKQPTPFLRTREFLWQEGHTAFA
DKEEAAKEVLDILDLYALVYTHLLAIPVVKGRKTEKEKFAGGDYTTTVEAFISASGRAIQGATSHHLGQNFSKMFEIVYEDPETQQKKYVYQNSWGITTRTIGVMIM
VHADNQGLVLPPHVACIQAIVVPCGITVNTKDDERAQLLDACKALEKRLVGGGVRCEGDYRDNYSPGWKFNHWELKGVPLRLEVGPKDLKAQQLVAVRRDTVEK
ITIPLADVEKKIPALLETIHESMLNKAQEDMTSHTKKVTNWTDFCGFLEQKNILLAPFCGEISCEDKIKADSARGEEAEPGAPAMGAKSLCIPFDQPAPIAASDKCINP
SCTNKPKFYTLFGRSY
>D. rerio NP_001038907
MKILRGVSRQMCTSRPEVRVRFAPSPTGFLHLGGLRTALYNFLFSRQRRGVFILRLEDTDQKRLVPGAAEHIEDMLEWAGIPPDESSRRGGDYGPYVQSERLHLYTE
AASSLLNTGHAYYCFCSNQRLELLKKEAQRSGHAPRYDNRCRRLQPQQVEQKLAAGVPAVVRFKLHTGTEEFQDLVFGWTGHAVGAVEGDPVILKADGYPTYHL
ASVVDDHHMRISHVLRGCEWLISSAKHLQLYRALRWTPPTYAHLPLLLNRDGSKLSKRQGDIFLQSFRDRGVLPETLLDLVTHAGSGFSDNRMGRRLDELIRDFNIS
KITTHSALLDLDKLEEFSRLHLQRRIEDPQQCVWLCEELKQMVKHTHSSEISAAAVLEPEYIERVLQLRKGHISSLQDLLSSTHSYLWVRPRVSQTQLQSESAHAKDI
ATAVMQMVLAGGSLVSMERLSSELKQISSRTNSTHSSVMKVLRLLLSAQQRGPSVAEMMLSLGEQEVCVRLQKALEL
>D. rerio NP_001275581
MALNLTFNSSNPPLGALLTAEHVKGSVNLSVEEGKDTMLHVSDQVQFSDVNSITRYLARVAPALGLYGSNMMEQTEVDHWLEFSARRLCGQSDLSSALADLDKAL
ALRTFLVGRSVTLADLCIWAALKGNGESQAKPSSHPHLCRWFSFLSSQVPFSSVGSKWASKVSASKAASVEKDKKQDLGKFVELPGAEMGKVVVRFPPEASGYLHI
GHAKAALLNQHYQVTFKGKLIMRFDDTNPEKEKEDFEKVILEDVAMLQIKPDQFTYTSDHFPTILRMGEKLLQEGKAYIDDTPPDIMKQEREQRAESRNRNNSVEK
NMQMWEEMKKGTEFGQTCCMRAKIDMNSNNGCLRDPALFRCKNAAHPRTGTTYKVYPTYDFACPIVDSIEGVTHALRTTEYHDRDEQFYWVIEALGLRKPYIWE
YARLNLNNTVLSKRKLTWFVDKGYVDGWDDPRFPTVRGVLRRGMTVDGLKQFIAAQGGSRSVVNMEWDKIWAFNKKVIDPVAPRFTALLSSQVIPVCVSEAKET
MKEVPKHPKNADVGLKQVWYGPKVFIEGADAETFTEGEIVTFINWGNIIITKIHRDSSGAILSLDGHLNLENTDYKKTTKITWLAESSRASFVPTVCVNYQHLITKP
VLGKDDDFKDYINKNSKIEEKMIGDPCLKDLKKGDIIQLQRRGFYICDQPYEPISPHSCRESPCVLLYIPDGHTKEMPTAGSKDKSKSQPAKPVKAEPAPPSKAKSAPS
SNPTPSAPGDASVILSNIMAQGDAVRQLKTEKAPKEQVDAAVKQLLALKSEFKQLTGQEYKPGMTLPASAPVQKSSAPTLAAAPGPASTGCPFTRVSEQGELVRKLK
AEKAPKEQIDAAVKQLLAFKAEFKQLTGQDYKPGMAPPTASPPEPAPSSNSSPAGIYERVSEQGELVRKLKAEKAPKDQVDAAVKQLLALKAEYKQGTGQDYKPG
VPTAKASPASSSSPAQLTVASTPAASAPSDSSPQAIYDRVSQQGDLVRKLKTEKAPKDQVDAAVKQLLNLKAEYKQQTGQEYKPGAPPSSGSATAQSNPTPPQSSSSQ
PQSSSSPQAQALFSDIAQQGEQVRRLKTEKAPKEQVDKAVKTLLELKGQYKALTGEEYKPVTTPGNSGGEDKSRKERENKSEKQGGGGGGRKQQKGGEKPQQG
QESGAGGAGEGQGPKKQTRLGLEAKKEENLADWYSQVITKAEMIEYYDVSGCYVLRPWSYSIWDAIKEFFDREIKKLGVENCYFPMFVSQAALEKEKTHIADFAP
EVAWVTRSGKTELAEPIAVRPTSETVMYPAYAKWVQSHRDLPIKLNQWCNVVRWEFKHPQPFLRTREFLWQEGHTAFATKEEATEEVLQILDLYARVYEELMAIPVV
KGRKTEKEKFAGGDYTTTVEAFISASGRAIQGATSHHLGQNFSRMFEIVFEDPKKPGEKQLAYQNSWGITTRTIGVLTMVHGDNMGLVLPPRVACLQVIIIPCGITATL
PEAEKDLLLDQCSKYLSRLQKADIRVKADLRDNYSPGWKFNHWELKGVPIRFEVGPKDLKQGQFVAVRRDTGEKITVKEADAEKKIPNLLEEIQNNLFKRASDDLN
KHMVVADTMEQFQKDLDQGRIVQIPFCGGIECEDWIKKTTAKDQDLEPGAPSMGAKSLCIPFKPLKNLQAGQMCVSGKEPAQYYTLFGRSY
>H. sapiens NP_001077083
MAALLRRLLQRERPSAASGRPVGRREANLGTDAGVAVRVRFAPSPTGFLHLGGLRTALYNYIFAKKYQGSFILRLEDTDQTRVVPGAAENIEDMLEWAGIPPDESPR
RGGPAGPYQQQSQRLELYAQATEALLKTGAAYPCFCSPQRLELLKKEALRNHQTPRYDNRCRNMSQEQVAQKLAKDPKPAIRFRLEQVVPAFQDLVYGWNRHEVAS
VEGDPVIMKSDGFPTYHLACVVDDHHMGISHVLRGSEWLVSTAKHLLLYQALGWQPPHFAHLPLLLNRDGSKLSKRQGDVFLEHFAADGFLPDSLLDIITNCGSGF
AENQMGRTLPELITQFNLTQVTCHSALLDLEKLPEFNRLHLQRLVSNESQRRQLVGKLQVLVEEAFGCQLQNRDVLNPVYVERILLLRQGHICRLQDLVSPVYSYL
WTRPAVGRAQLDAISEKVDVIAKRVLGLLERSSMSLTQDMLNGELKKLSEGLEGTKYSNVMKLLRMALSGQQQQGPPVAEMMLALGPKEVRERIQKVVSS
>H. sapiens NP_004437
MATLSLTVNSGDPPLGALLAVEHVKDDVSISVEEGKENILHVSENVIFTDVNSILRYLARVATTAGLYGSNLMEHTEIDHWLEFSATKLSSCDSFTSTINELNHCLSLRT
YLVGNSLSLADLCVWATLKGNAAWQEQLKQKKAPVHVKRWFGFLEAQQAFQSVGTKWDVSTTKARVAPEKKQDVGKFVELPGAEMGKVTVRFPPEASGYLHIG
HAKAALLNQHYQVNFKGKLIMRFDDTNPEKEKEDFEKVILEDVAMLHIKPDQFTYTSDHFETIMKYAEKLIQEGKAYVDDTPAEQMKAEREQRIDSKHRKNPIEKN
LQMWEEMKKGSQFGQSCCLRAKIDMSSNNGCMRDPTLYRCKIQPHPRTGNKYNVYPTYDFACPIVDSIEGVTHALRTTEYHDRDEQFYWIIEALGIRKPYIWEYSR
LNLNNTVLSKRKLTWFVNEGLVDGWDDPRFPTVRGVLRRGMTVEGLKQFIAAQGSSRSVVNMEWDKIWAFNKKVIDPVAPRYVALLKKEVIPVNVPEAQEEMKE
VAKHPKNPEVGLKPVWYSPKVFIEGADAETFSEGEMVTFINWGNLNITKIHKNADGKIISLDAKLNLENKDYKKTTKVTWLAETTHALPIPVICVTYEHLITKPVLG
KDEDFKQYVNKNSKHEELMLGDPCLKDLKKGDIIQLQRRGFFICDQPYEPVSPYSCKEAPCVLIYIPDGHTKEMPTSGSKEKTKVEATKNETSAPFKERPTPSLNNN
CTTSEDSLVLYNRVAVQGDVVRELKAKKAPKEDVDAAVKQLLSLKAEYKEKTGQEYKPGNPPAEIGQNISSNSSASILESKSLYDEVAAQGEVVRKLKAEKSPKAKI
NEAVECLLSLKAQYKEKTGKEYIPGQPPLSQSSDSSPTRNSEPAGLETPEAKVLFDKVASQGEVVRKLKTEKAPKDQVDIAVQELLQLKAQYKSLIGVEYKPVSATG
AEDKDKKKKEKENKSEKQNKPQKQNDGQRKDPSKNQGGGLSSSGAGEGQGPKKQTRLGLEAKKEENLADWYSQVITKSEMIEYHDISGCYILRPWAYAIWEAIK
DFFDAEIKKLGVENCYFPMFVSQSALEKEKTHVADFAPEVAWVTRSGKTELAEPIAIRPTSETVMYPAYAKWVQSHRDLPIKLNQWCNVVRWEFKHPQPFLRTREF

LWQEGHSAFATMEEAAEEVLQILDLYAQVYEELLAIPVVKGRKTEKEKFAGGDYTTTIEAFISASGRAIQGGTSHHLGQNFSKMFEIVFEDPKIPGEKQFAYQNSWGL
TTRTIGVMTMVHGDNMGLVLPPRVACVQVVIIPCGITNALSEEDKEALIAKCNDYRRRLLSVNIRVRADLRDNYSPGWKFNHWELKGVPIRLEVGPRDMKSCQFVA
VRRDTGEKLTVAENEAETKLQAILEDIQVTLFTRASEDLKTHMVVANTMEDFQKILDSGKIVQIPFCGEIDCEDWIKKTTARDQDLEPGAPSMGAKSLCIPFKPLCEL
QPGAKCVCGKNPAKYYTLFGRSY
>M. musculus NP_080416
MAAPLKRLLLAEPHVVALGHRVGRREASLGPDPGAPVRVRFAPSPTGFLHLGGLRTALYNYIFAKKHQGSFILRLEDTDQSRLVPGAAESIEDMLEWAGIPPDESPRR
GGPAGPYCQSQRLALYAQATEALLRSGAAYPCFCLPQRLELLKKEALRSRQTPRYDNRCRNLSQAQVAQKLAVDPKPAIRFRLEEAVPAFQDLVYGWTQHEVASVE
GDPVILKSDGFPTYHLACVVDDHHMSISHVLRGSEWLVSTSKHLLLYQALGWQPPRFAHLPLLLNRDGSKLSKRQGDIFLEHFAATGFLPEALLDIITNCGSGFAENQ
MGRTLPELITQFDLTRITCHSALLDLEKLPEFNRLHLRRLVSSETQRPQLVEKLQGLVKEAFGSELQNKDVLDPAYMERILLLRQGHISRLQDLVSPVYSYLWTRPAVH
RSELGASSENVDVIAKRLLGLLERPGLSLTQDVLNRELKKLSEGLEGAKHSSVMKLLRMALSGQLQGPPVAEMMVSLGPKEVRERIQKVLSS
>M. musculus NP_084011
MAALCLTVNAGNPPLEALLAVEHVKGDVSISVEEGKENLLRVSETVAFTDVNSILRYLARIATTSGLYGTNLMEHTEIDHWLEFSATKLSSCDRLTSAINELNHCLSL
RTYLVGNSLTLADLCVWATLKGSAAWQEHLKQNKTLVHVKRWFGFLEAQQAFRSVGTKWDVSGNRATVAPDKKQDVGKFVELPGAEMGKVTVRFPPEASGYLH
IGHAKAALLNQHYQVNFKGKLIMRFDDTNPEKEKEDFEKVILEDVAMLHIKPDQFTYTSDHFETIMKYAEKLIQEGKAYVDDTPAEQMKAEREQRTESKHRKNSVE
KNLQMWEEMKKGSQFGQSCCLRAKIDMSSNNGCMRDPTLYRCKIQPHPRTGNKYNVYPTYDFACPIVDSIEGVTHALRTTEYHDRDEQFYWIIEALGIRKPYIWEY
SRLNLNNTVLSKRKLTWFVNEGLVDGWDDPRFPTVRGVLRRGMTVEGLKQFIAAQGSSRSVVNMEWDKIWAFNKKVIDPVAPRYVALLKKEVVPVNVLDAQEE
MKEVARHPKNPDVGLKPVWYSPKVFIEGADAETFSEGEMVTFINWGNINITKIHKNADGKITSLDAKLNLENKDYKKTTKITWLAESTHALSIPAVCVTYEHLITKP
VLGKDEDFKQYINKDSKHEELMLGDPCLKDLKKGDIIQLQRRGFFICDQPYEPVSPYSCREAPCILIYIPDGHTKEMPTSGSKEKTKVEISKKETSSAPKERPAPAVSS
TCATAEDSSVLYSRVAVQGDVVRELKAKKAPKEDIDAAVKQLLTLKAEYKEKTGQEYKPGNPSAAAVQTVSTKSSSNTVESTSLYNKVAAQGEVVRKLKAEKAPK
AKVTEAVECLLSLKAEYKEKTGKDYVPGQPPASQNSHSNPVSNAQPAGAEKPEAKVLFDRVACQGEVVRKLKAEKASKDQVDSAVQELLQLKAQYKSLTGIEYKP
VSATGAEDKDKKKKEKENKSEKQNKPQKQNDGQGKDSSKSQGSGLSSGGAGEGQGPKKQTRLGLEAKKEENLAEWYSQVITKSEMIEYYDVSGCYILRPWSYSI
WESIKDFFDAEIKKLGVENCYFPIFVSQAALEKEKNHIEDFAPEVAWVTRSGKTELAEPIAIRPTSETVMYPAYAKWVQSHRDLPVRLNQWCNVVRWEFKHPQPFLR
TREFLWQEGHSAFATFEEAADEVLQILELYARVYEELLAIPVVRGRKTEKEKFAGGDYTTTIEAFISASGRAIQGATSHHLGQNFSKMCEIVFEDPKTPGEKQFAYQCS
WGLTTRTIGVMVMVHGDNMGLVLPPRVASVQVVVIPCGITNALSEEDREALMAKCNEYRRRLLGANIRVRVDLRDNYSPGWKFNHWELKGVPVRLEVGPRDMK
SCQFVAVRRDTGEKLTIAEKEAEAKLEKVLEDIQLNLFTRASEDLKTHMVVSNTLEDFQKVLDAGKVAQIPFCGEIDCEDWIKKMTARDQDVEPGAPSMGAKSLCI
PFNPLCELQPGAMCVCGKNPAKFYTLFGRSY
>C. reinhardtii EDP07219
MLLAQVSVSRRPVVTYATAAVQEAPAATTAPAKGSQVRVRFAPSPTGNLHVGGARTALFNWLYARKMGGKFILRVEDTDTARSTRASEEAMVRDLKWLGLDWDE
GPDVGGAHGPYRQSERTAIYKEMADKLVAEGKAYPCFCTDEELEAMKKDAEEKKLPPIYRGKWARATAAEVEAAKAQGLPYCYRFRVPSNQEVKIHDLIRGEVSW
NTDTLGDFVLLRSNGMPVYNFCVAVDDALMGITHVLRAEEHLPNTLRQMLIYQALGFPVPTFGHVSLILAPDKSKLSKRHGATSVGEFRTQGYLAPAMINFLSLLG
WNDGTEQEIFSVEELTTKFSLDRITKSAAVFDKVKLGWMNGQHLKQLPEEEAAAMAGQHLVACGLLASADSAFAKAAVKLVSKSMELVTDVEPEVRKLLAYPLAE
TLAAPEAKAVIDDGFAEVAGAVLAAYESGELAAAMAQGHDGYKAWINGVGKAQKRKGKKLFMPARVAITGRMAGPDVGDQLEVLGALAGAADVAEAGAYVPL
ADRMAQLKAWLEANKA
>C. reinhardtii EDO97532
MSATLTFWDRNPPYAAVALARLANVPVAATHDPKATKETVPTLSVSGETFTGTMILKYIARASAQRDELYGGSDAVAACQVDQWLDQAGAITPAALEAQAAALND
FLQLRTFLVGYSLTLADLAVWGALQASPLWNKVRGSGKVPHLARWFAFCSEVPQVRAATEELAALAGKKPGAPAAGGADKKGAGAGAAGDNKKGGAAGGAAG
AAAGGAEGGDAGGSFDIGLPGAEEGKVVTRFPPEPSGYLHIGHAKAALLNQYFADMYKGKLLVRFDDTNPSKEKDEFVENIMKDIADLGLRYEKLTYTSDYFPQL
LDLGERMIRAGLMYADDTPVEAMREQRGRGEESACRGRSVEENLRLWEEMKAGSEVGLANAMRFKIDMKSVNGTMRDPVAFRCNLTHHWRTGTKYKLYPTYD
CACPFVDAVEGVTHALRTSEYRDREEQYYWILSAYQKVWPGGLPHVHIWDYSRLNFVNTVLSKRKLTWFVDTGRVDGWADPRMPTVQGILRRGLRIEALKEFILS
QGASKNITFQEWDKIWTINKKLIDPVCPRHTAVDVEGKVLLTLEDGPSPPEVVTVPRHLKYPPAGKKAVTRGRTLWLEQADAAAISEGEEVTLMSWGNAVITAVTR
DAATGAVTALAGKLNLAGDVKKTKLKLTWLAAVPECETVELSLLDFDYLITKKKVEEDDNFMDLVNPVTKFEKAALGDPNMRSLQKGDVIQLERKGYYIVDEPH
GVKGAGKPMVLFAIPDGRTKNMTKPGAAAPAASS
>P. patens XP_001771575
MLYSCGAPERVSSVRWVGSFLNPLEKLALPTLRNSGNVAPRVLLSARASTKGAVQGDGTGNLKGSEVRVRFAPSPTGNLHVGGARTALFNYLYAKSLGGKFILRIED
TDLERSTKESEEAVIRDLQWLGLEWDEGPDKEGECGPYRQSERTHIYQAYVEKLLKTGHVYRCFCTDEELEAMREEAKSKNMPPRYMGKWGSASQAEIDAELEK
GTPFTYRFRVPQEGSVTIHDLIRGEVSWKKDTLGDFVVLRSNGQPVYNFCVAVDDATMKISHVIRAEEHLPNTLRQALIYNALGFSMPQFGHVSLILAPDRSKLSKR
HGATSVGQFREMGYLAPAMVNYLALLGWNDGSEDEIFTIEQIIEKFSIDRITKSGAIFDQTKLGWMNGQHLRLMPTEDVATMFREQWLKAGIVSESVSGSFIDVAVE
LLKNGTELVADSENHLLSILHYPLNETLTSSEVKDVVDDNISEVADAIVSAYDNGELQAAIEGGHDTWSKWIKAMGKAFKRKGKRLFMPVRVLLTGSMKGPEVGA
VLDLVREGETTGSVTEKAGLVPLDARIKILREVDWAAVTSKLAESASQETEPALSH
>P. patens XP_001762864
MSDFRRQNPKSFSNVGQVDDWLKFASSFGVGSEFEGACAYANSYLTLRTFLVGNELTIADITVVAYLAGAGPRWESFRKSSKFPNLVRWYNCILAQYAAIADVLPRR
GSGKPAAEVSKPSSTAAAPDEAKEKVDGSFEVDLPGAEMGKVCTRFPPEPSGYLHIGHAKAALLNQFFAQKYKGRLIIRFDDTNPAKEKDEFVESILHDLETLGVKG
DVITYTSDYFPQLMEMCEKLIKEGKAYVDDTEREQMQTERKAMIESKCRNQSVEENLILWKEMIAGSPRGVQCCVRAKMDMSNPNASLRDPVYYRCNPAPHLRV
GSKYKVYPTYDFACPYVDSIEGVTHALRSSEYHDRNDQYYRVLEDMQLRKVHVWDFSRLNFVYTLLSKRKLQWFVDNGRVEGWNDPRFPTVQGIIRRGLTIEALK
QFILSQGASKNLNLMEWDKLWTINKKIIDPICPRHTAVLSEGKVLLDLSNGPETPFTRVILRHKKYEPAGKKASVFTKRIWLDHADATAVSEGEEVTLMDWGNCFIR
KVERDQYGLVTRLQGELHVEGSVKTTKLKLTWLPEFSELVPVTLVDLDHLIVKKKLEENDDFVSVLNPVTRIETSALGDSNMRNLQRGEIIQLERKGYYICDVPFLR
SSKPIVLISIPDGRQQKPPGPQ
>AT5G64050
MASLVYGTPWLRVRSLPELAPAFLRRRQSSLFYCSRRSFAVVACSTPVNNGGSVRVRFAPSPTGNLHVGGARTALFNYLFARSKGGKFVLRIEDTDLERSTRESEAAV
LQDLSWLGLDWDEGPGVSGDFGPYRQSERNALYKQYAEKLLESGHVYRCFCSSEELVKMKENAKLKQLPPVYTGKWATASDAEIEQELEKGTPFTYRFRVPKEGS
LKINDLIRGEVCWNLDTLGDFVVMRSNGQPVYNFCVTVDDATMAISHVIRAEEHLPNTLRQALIYKALKFPMPQFAHVSLILAPDRSKLSKRHGATSVGQYREMG
YLPQGMVNYLALLGWGDGTENEFFTLEDLVEKFSIERVNKSGAIFDSTKLRWMNGQHLRALPNEKLTKLVGERWKSAGILTESEGSFVNEAVELLKDGIELVTDSD
KVLLNLLSYPLHATLASPEAKPAVEDKLHEVAASLIAAYDSGEIPSALEEGQGAWQKWVKAFGKSLKRKGKSLFMPLRVLLTGKLHGPEMGTSIVLIYKAGSPGIVV
PQAGFVSMEERFKILREIDWEALNKDESVPLESTATVST
>AT5G26710
MDGMKLSFPPESPPLSVIVALSLSASPVTIDSSAAATTVPSFVFSDGRKLNGATVLLRYVGRSAKKLPDFYGNNAFDSSQIDEWVDYASVFSSGSEFENACGRVDKYL
ESSTFLVGHSLSIADVAIWSALAGTGQRWESLRKSKKYQSLVRWFNSILDEYSEVLNKVLATYVKKGSGKPVAAPKSKDSQQAVKGDGQDKGKPEVDLPEAEIGKV
KLRFAPEPSGYLHIGHAKAALLNKYFAERYQGEVIVRFDDTNPAKESNEFVDNLVKDIGTLGIKYEKVTYTSDYFPELMDMAEKLMREGKAYVDDTPREQMQKER
MDGIDSKCRNHSVEENLKLWKEMIAGSERGLQCCVRGKFNMQDPNKAMRDPVYYRCNPMSHHRIGDKYKIYPTYDFACPFVDSLEGITHALRSSEYHDRNAQYF
KVLEDMGLRQVQLYEFSRLNLVFTLLSKRKLLWFVQTGLVDGWDDPRFPTVQGIVRRGLKIEALIQFILEQGASKNLNLMEWDKLWSINKRIIDPVCPRHTAVVAER
RVLFTLTDGPDEPFVRMIPKHKKFEGAGEKATTFTKSIWLEEADASAISVGEEVTLMDWGNAIVKEITKDEEGRVTALSGVLNLQGSVKTTKLKLTWLPDTNELVNL
TLTEFDYLITKKKLEDDDEVADFVNPNTKKETLALGDSNMRNLKCGDVIQLERKGYFRCDVPFVKSSKPIVLFSIPDGRAAK
>S. lycopersicum XP_004231106
MATVAAAPWFRARLIPELKSRQSLLYCRNGPRRYPQSLCFRRRSFSIYASAGDEGAVRVRFAPSPTGNLHVGGARTALFNYLYARSKGGKFVLRIEDTDLERSTRESE
EAVLRDLSWLGLAWDEGPGIGGNYGPYRQSERNALYKEFAEKLLQSGHVYRCFCSNEELEKMKEMAKLKQLPPVYTGRWASATEQEVLEELTKGTPYTYRFRVPK
EGSLKIDDLIRGEVSWNFDTLGDFVIMRSNGQPVYNFCVTVDDATMAISHVIRAEEHLPNTLRQALIYKALGFPMPHFAHVSLILAPDRSKLSKRHGATSVGQFRDM
GYLPQAMVNYLALLGWGDGTENEFFSLEQLVEKFTIERVNKSGAIFDSTKLRWMNGQHLRSLPSEELNQIIGERWKDADIVTESQGVFVQDAVLLLKDGIDLIPDSE
KALSSLLSYPLYETLTSTEGKLILEDGVSEVAKSLLAAYDSGELCGALAEGQPGWRNWVKKFGKLLKRKGKSLFMPLRVLLTGKLHGPDIGATTVLLYKAGTSGSV
APQAGFVTFDERFKILREVQWESFSSNVPLSAGVVTH
>S. lycopersicum XP_004230376
MELKLSFPADSPPLAVIAAAKIAGVPISTDPTLTTGSTPIMHIDNGLKLRGNFVLLRYISRVANIPDLYQRDACESSQIDEWLEYAPIFASGSQFEEACGYVDGYLLQHT
FLVGHSLSIADIAVWSGLAGTGKRWESLRSSKKYQNLARWFNSILTEYGVLNEVTATYVGKKGPGKPTASKVKEQQGSNANLAKVNGDAADKEKEGRKPTSEVDL
PEAEVGKVRLRFAPEPSGYLHIGHSKAALLNQYFAERYQGEVIIRFDDTNPDKESNEFVDNLLKDIETLGIKYRTVTYTSDYFPQLMEMAEKLIREGKAYVDDTPRE
QMQKERMDGIESRCRNNSVEENLKLWKEMIAGSERGTMCCVRGKLDMQDPNKSVRDPVYYRCNQTPHHRIGAKYKVYPTYDFACPFVDAFEGITHALRSSEYH
DRNDQYYWIQTDMGFSKVHIYEFSRLNLVYTLLSKRKLLWFVQNGLVEGWDDPRFPTVQGIVRRGLKIEALIQFILEQGASKNLNLMEWDKLWATNKKIVDPVCP
RHTAVIEERRVLLTLSNGPDDPFVRIVPKHKKYAGAGEKATTYTKRIWIDYDDAVSISVNEEVTLMDWGNAIVKEIRKDQEGNVTHLSGILHLEGSVKTTKLKLTWL
PESDELVKLSVVDFDYLITKKKLEEDENFVDVVNPCTRKETPALGDSNMRNLQRGDVLQLERKGYFRCDVPFLRPQKRIVLFAIPDGKQQPVLRFAVPDGKTK

>G. max XP_003555767
MDIKTFAFAPSSPPLSVIAAAKLAGISPTIDTALSPDSPPSFIFSDGLKLHGPYALLRYIGRVASLPNFYGQNPFESSQIDEWLDYAPILSLGPAFENACKYIDEYLEKRTF
LVGYSLSIADLAIWAGLAGSGKRWESLRKSKKYPNLARWFNSIVEEHGTALNEVTSVYAGKKGLGEPTATKSKEQSVNTDKVKKVNGDVSENNKGGSKPSAEIDLP
DAEVGKVRLRFAPEPSGYLHIGHSKAALLNKYFAERYQGQVIVRFDDTNPAKESNEFVDNLIKDIDTLGIKYEQITYTSDYFPELMEMAEKLIREGKTYVDDTPREQ
MQKERMDGIESKCRNNIVEENLKLWKEMIAGTERGLQCCVRGKLDMQDPNKSLRDPVYYRCNPMPHHRIGSKYKVYPTYDFACPYVDAIEGITHALRSSEYHDR
NAQYYRIQEDMGLRKVLIYEFSRLNMVYTLLSKRKLLWFVQNGKVDGWDDARFPTVQGIVRRGLKIEALIQFIVEQGASKNLNLMEWDKLWTINKKIIDPVCPRH
TAVIADRRVLLTLTDGPEKPFVRIIPRHKKYEAAGDKVTTYTKRIWLDHADAVSLSAGEEVTLMDWGNAIVKEIEKDQDGNITGLSGALHLEGSVKTTKLKLTWLPE
IDELVCLTLVEFDYLITKKKLEEEDFIDVLNPCTKKETLAYGDSNMRNLQRGDVLQLERKGYFRCDLPFIRPSQPIVLYAIPDGRQQTSLK
>G. max XP_003529003
MALLCGGTPWSKVTISPIFHHSHTPRTFFFQRRRFSVSAAFSEQQPPPPVRVRFAPSPTGNLHVGGARTALFNYLFARSKGGKFVLRIEDTDLERSTRESEEAMLKDLS
WLGLDWDEGPGVGGDYGPYRQSDRNSLYKQFADNLHQSGHVYRCFCSNEELEKMKEDAKLKQLPPVYTGKWASATNEEVEEELAKGTPYTYRFRVPKGSLKIN
DQIRGEVSWNLDTLGDFVIMRSNGQPVYNFCVTVDDATMAISHVIRAEEHLPNTLRQALIYKALGFPMPYFAHVSLILAPDRSKLSKRHGATSVGQFREMGYLPEA
MVNYLALLGWGDGTENEFFTLDQLVEKFTIERVNKSGAIFDSTKLRWMNGQHLRARPSEELTKLIGEYWKTSGLLTVSDGSFIDEAVQLLKDGIDLINDADKALAN
LLSYPVYSTLQSDEAESVLQDNLSEFAASLLAAYDSGDLVAALEEGHAGWQKWVKGFGKLLKRKGKSLFMPLRLLLTGKIHGPDMGASVVLLYKAGTSNAVAPDA
GFVTLDERFKILGQINWETLSKDQTVKETADPV
>Os02g0121000
MMAAAMGSPWLRIRLLPEVPPRLLRPHLRRTLSVRASASASASPDGAGGPVRVRFAPSPTGNLHVGGARTALFNYLFARSKGGKFVLRIEDTDLERSTKKSEEAVLS
DLAWLGLDWDEGPDVGGEFGPYRQSERNSMYKQYAEKLMESGAVYQCFCSSEELEQMKETAKQMQLPPVYMGKWGTASDAEIQQELEKGTPYTYRFRVPKEGS
LKINDLIRGEVSWNLDTLGDFVIMRSNGQPVYNFCVTVDDATMRISHVIRAEEHLPNTLRQALIYKALGFPMPSFAHVSLILAPDRSKLSKRHGATSVGQYKEMGYL
PQAMVNYLALLGWGDGTENEFFTIDDLVEKFTINRVNKSGAVFDAVKLKWMNGQHLRSFPPDVLIKSFEDRWKDTGILQESESGFAKEAAELLKDGIDLITDADAA
LSNLLSYPLHATLSSDEAKSVVQDKLSEVASGLISAYDSGELCQALAEGRDGWQKWVKIFGKSLKRKGKSLFMPLRVLLTGKLHGPDMGGTVVLIHKAGTCGAVT
QQSGFVNLDERFRILKEVEWESLVQEQESPAETAVPASR
>OsERS1
MEAKLSFSQDSPPISIISAAKVAGVSLSIDPSLAAGSAPVLCFSSGESLRGINPILEYIAQSSPSLHGRDAIESGHVVEWLEYAPTFLLGSEFEVACSFVDGYLMSRTFLV
GHVLTIADITVWSNLAGIGQRWESLRKSKKYQNLVRWFNGIDSDYKDTLNEVIAAYVGKRGIGKSPAPNLKEKVNDSKDPSAPEVDLPGAKFGEVCVRFAPEPSGY
LHIGHAKAALLNKYFAERYQGRLIVRFDDTNPSKESNEFVENLLKDIETLGIKYDAVTYTSDYFPKLMEMAEKLIKQGKAYVDDTPKEQMRSERMDGVESKCRNN
TVEENLSLWKEMINGSERGMQCCVRGKLDMQDPNKSLRDPVYYRCNTDPHHRIGSKYKVYPTYDFACPFVDALEGVTHALRSSEYHDRNAQYYRILQDMGMRR
VEIYEFSRLNMVYTLLSKRKLLWFVQNKKVEDWTDPRFPTVQGIVRRGLKVEALIQFILQQGASKNLNLMEWDKLWTINKKIIDPVCARHTAVLKDQRVIFRLTNG
PEKSFVQILPRHKKFDGAGKKATTFTNRIWLDYADASAISKGEEVTLMDWGNAIIKEIKMENGVITELIGELHLEGSVKTTKLKVTWLPDIDDLVPLSLVEFDYLISK
KKLEEDENFLDNLNPCTRKETLALGDANMRNLLRGEVIQLERKGYYRCDAPFVRSSKPVVLFAIPDGRQQASLN
>Os01g0271200
MEPKLAFPQDSPPLAILCAAKVAGVSLTLDPKLASGSAPTLHLGSGDFIHGVNTILRYIARVASVTSFYGQDAIQAAYVDQWLDYAPVILSGSEFEAACSFLDGYLAS
RTFLVSYGLSIADIVVWSNLAGTGQRWESLRRSKKYQNLVRWFNSIADYSDTLDEVVSAYVGKRGIGKSPAPSLKEKLPDSKQNISIPEVDLPGAKVGEVCVRFAPE
PSGYLHIGHAKAGLLNKYFAERYKGRLIVRFDDTNPSKESNEFVENVLKDVETLGIKYDIVTYTSDYFPQLMEMAENLIKQGKAYVDDTPKEQMRSERMDGVESK
CRNNTVQDNLSLWKEMINGTERGMQCCVRGKLDMQDPNKSLRDPVYYRCNTDPHHRIGSKYKVYPTYDFACPFVDALEGVTHALRSSEYHDRNAQYYRILQEM
GLRRVEIYEFSRLNMVYTVLSKRKLLWFVQNKKVEDWTDPRFPTVQGIVRRGLKVEALVQFILEQGASKNLNLMEWDKLWTINKKIIDPVCGRHTAVLKDQRVLF
TLTNGPEEPFIRVLPRHKKYEGAGKKATTFTNKIWLESADASVISIGEEVTLMDWGNAIIKEIKTQNGIITELLGELHLEGSVKMTKLKLTWLPDIEDLVSLSLVEFDY
LIKKKKLEEDDNFLDNLNPCTCQEFPALGDANMRNLKQGEIIQLERKGYYRCDAPFIRSSKPIVLFAIPDGRQKSATK
>B. distachyon XP_003573538
MAASTFMGSPWLRIRLLPSAPAQRLLRRTLSVRASASAGDDGGPVRVRFAPSPTGNLHVGGARTALFNYLFARSRGGKFVLRVEDTDLQRSTKESEDAVLSDLSWL
GLDWDEGPDVGGEFGPYRQSERNSLYKQHAEKLMESDAVYRCFCSSEELEKMKETANRMKLPPVYMGKWATASDAEVQQELEKGTPYTYRFRVPKEGSLKINDL
IRGEVSWNLDTLGDFVIMRSNGQPVYNFCVTVDDATMRISHVIRAEEHLPNTLRQALIYKALGFPMPLFAHVSLILAPDRSKLSKRHGATSVGQYKEMGYLPEAMV
NYLALLGWGDGTENEFFTIDDLVEKFTINRVNKSGAVFDGTKLKWMNGQHLRSLPSDLLIKNFEDRWKNTGILLESESGFAKEAAELLKEGIDLITDADAALSNLLS
YPLHDTLSSDEAKPVVEDKLSEVASGLISAYDSGELCQAIAEGHDGWKKWVKSFGKSHKRKGKSLFMPLRVLLTGKLHGPEMASAIILVHKAGICRAVAPQSGFVN
LDERFRILKEVDWESLHKNQESPAESAAAVPAAS
>B. distachyon XP_003566911
MEAKLAFPQDSPPLSILSAAKIAGVPLTIDPTLASGSVPTLHLSSGDFIHGVNTILRYIARAASVSSFYGQDAIQAAHVDQWLEYAPLILSGSEFEAACSFIDGYLASRT
FLVGYGLSIADIVVWSNLTGTGQRWESLRKSKKYQNLVRWFNSIAVDYSDALDEVTTAYVGKRGIGKSPAPSLKEKVPGLKENISGHEIDLPGAKVGEVCVRFAPEP
SGYLHIGHAKAALLNKYFAERYKGRLIVRFDDTNPSKESNEFVENVLKDIETLGVKYDVVTYTSDYFPKLMEMAESLIKQGKAYVDDTPKEQMRSERMDGVESKC
RNSTVQENLLLWSEMVKGTQRGTQCCVRGKLDMQDPNKSLRDPVYYRCNPDPHHRVGSKYKVYPTYDFACPFVDALEGVTHALRSSEYHDRNAQYYRVLQDM
GLQRVEIYEFSRLNMVYTVLSKRKLLWFVQNKKVEDWTDARFPTVQGIVRRGLKIEALIQFILEQGASKNLNLMEWDKLWTINKKIVDPVCGRHTAVLKDQCVLL
TLTNGPEEPFVRILPRHKKYEGAGKKATTFTNRIWLEYADASVISVGEEVTLMDWGNAIIKEVKTDNGAITQLVGELHLDGSVKMTKLKLTWLADIEDLVSLSLVDF
DYLINKKKLEEDDNFLDNLNPCTRREALALGDPNMRNIKQGEVIQLERKGYYRCDVPFVRSSKPVVLFAIPDGRQKSTSIVSGA
>B. distachyon XP_014757177
MARTLFRRHIYVLVDQWLEYAXLILSGSGFEAACLFLDGYLVSRTFLVGYALSIADIVVWSNLTGTGQRWESLRKSKKYQNLVRWFNSVSVDYADVLDEVTAAYVG
KRGIGKSPAPSLKDKEPGLKENASGQEIDLPGAKVGEVCVRFAPEPSGYLHIGHAKAALLKXFAERYKGRLIVRFDDTNPSKESNEFVENVLKDIETLGVKYDVVTY
TSDYFPKLMEMAESLIKQGKAYVDDTPKLQMRNERMDGVESKCRNSTVEENLSSWNEMVSGTKRGTECCVRGKLDMQDPNKSLRDPVYYRCNPDPHHRVGSK
YKVYPTYDFACPFVDALEGVTHALRSSEYHDRNAQYYRILQDMGLRRVEIYEFSRLNMVYTVLSKRKLLWFVQNKKVEDWTDARFPTVQGIVRRGLKIEALIQFIL
EQGASKNLNLMEWDKLWTINKKIVDPVCGRHTAVLKEQRVLLTLTNGPEEPLVRILPRHKKYEGAGKKATTFTDKIWLEYADASVISVGEEVTMMDWGNAIIREIK
TDNGTTTQLVGELHLDGSVKMTKLKLTWLPDIEDLVSLSLVEFDYLINKKKLEEDEDFLDNLNPCTRREALALGDPNMRNVKQGEVIQLERKGYYRCDVPFLRPSK
PIVLFAIPDGRQKSTSIVSGA
>Z. mays GRMZM2G046932
MEAALSFSKDSPPISIICAAKIVGLPLTINHSLAAGSAPTLQFASGESLHGVNPIILYIARGASIASLSGKNDIEFGHVVEWLEYAPTFLSGSEFENACLFVDGFLASRTFL
VGHGLTIADIAVWSNLAGIGQRWESLRKSKKYQNLVRWFNSIDSEYKEALNEVVAAFVGKRGIGKSPAPSLKEKVHDSKDPSAPEVDLPGAKVGKVCVRFAPEPSG
YLHIGHAKAALLNKYFAERYQGRLIVRFDDTNPSKESNEFVENLLKDIETLGIKYDAVTYTSDYFPKLMEMAESLIKQGKAYIDDTPKEQMRKERMDGIESRCRNN
TVEENLSLWKEMVNGTERGMQCCVRGKLDMQDPNKSLRDPVYYRCNTDPHHRVGSKYKVYPTYDFACPFVDALEGVTHALRSSEYHDRNAQYYRILQDMGLR
RVEIYEFSRLNMVYTLLSKRKLLWFVQNKKVEDWTDPRFPTVQGIVRRGLKVDALIQFILQQGASKNLNLMEWDKLWTINKKIIDPVCARHTAVLKDQRVIFTLTN
GPEEPFVRILPRHKKFEGAGKKATTFANRIWLDYADAAAISKGEEVTLMDWGNAIVKEIKVESGVITELVGELHLEGSVKTTKLKITWLADIEELVPLSLVEFDYLIS
KKKLEEDEDFLDNLNPCTRREIPALGDANMRNIKRGEVIQLERKGYYRCDAPFIRSSKPVVLFAIPDGRQQASLS
>Z. mays GRMZM2G057491
MEATLAFSQDSPPISIICAAKFVGLPLTIDPSLASGLAPTLRFASGESLHGVNPILLYIASGATIPCFSGKNDIEFGHVVEWLEYVPIFLSGSEFENACLFVDGFLASRTFL
VGHGLTIADIVVWSNLAGIGQRWESLRKSKKYQNLVRWFNSIDSEYSDTLNEVLAAYVGKRGIGKSPAPSLKEKVHVSKDPSAPEVDLPGAKVGKVCVRFAPEPSG
YLHIGHAKAGLLNKYFAERYQGRLIVRFDDTNPSKESNEFVENVLKDIETLGIKYDAVTYTSDYFPKLMEMAGSLIKQGKAYIDDTPKEQMRKERMDGIESRCRNS
TVEENLSLWNEMVNGTERGMQCCVRGKLDMQDPNKSLRDPVYYRCNIDPHHRVGSKYKVYPTYDFACPFVDALEGVTHALRSSEYHDRNAQYYRILQDMGLRR
VEIYEFSRLNMVYTLLSKRKLLWFVQNKKVEDWTDPRFPTVQGIVRRGLKIEALIQFILQQGASKNLNLMEWDKLWTINKKIIDPLCARHTAVLKDQRVIFTLTNGP
EKPFVRILPRHKKFVGAGKKATTFANRIWLDYADAAAISKGEEVTLMDWGNAIVKEIKVESGVITELVGELHLEGSVKTTKLKITWLADIEELVPLSLVEFDYLISKK
KLEEDDDFLENLNPCTRRENPALGDANMMNLKRGEIIQLERKGYYRCDAPFIRSLKPVVLFAIPDGRQQALLN
>Z. mays NP_001169893
MAASALLTGSPWLRMRLLPDAPARPFRHLHLRRSFSVRASGADGSPAPVRVRFAPSPTGNLHVGGARTALFNYLFARSKGGRFVLRVEDTDLERSTRKSEEAVLAD
LAWLGLDWDEGPDVGGDFGPYRQSERNSLYKQYAEKLLDSGAVYRCFCSSEELEQMKEVAKQRQLPPVYMGKWASASDAEVQQELEKGTPYTYRFRVPKEGSLK
ISDLIRGEVSWNLDTLGDFVIMRSNGQPVYNFCVTVDDATMQISHVIRAEEHLPNTLRQALIYKALGFAMPSFAHVSLILAPDRSKLSKRHGATSVGQYKNMGYLP
QAMVNYLALLGWGDGTENEFFTIDDLVQKFTINRVNKSGAVFDATKLKWMNGQHLRSFPHDELIKAFEDRWTDTGILQESESGFAKEAAELLKDGIDLITDADAAL
TNLLSYPLHTTLSSEEAKPVVQDKISEVAWSLISAYDSGELGQALAEGRDGWQKWVKGFGKSIKRKGKGLFMPLRVLLTGKLHGPDMGGSIALIHKAGVCCAVTPQ
SNFVTLDERFRMLKDVDWESLAKEQETPAVPAAAS
>E. guineensis XP_010932120
MASIVGSPWLRIRVRPDLPPSFLRPCLVSKPYVSVRRRILSVSASSAASPPDGEVRVRFAPSPTGNLHVGGARTALFNYLFARSKGGKFVLRIEDTDLERSTRKSEEAM
LNDLSWLGLDWDEGPGVVGDYGPYRQSERNSLYKQYAEKLLNSGAVYRCFCSNEELEQMKETAKQMQLPPVYMGKWATASDEEIEEELAKGTPYTYRFRVPKEG

TLKISDLIRGEVGWSLDTLGDFVIMRSNGQPVYNFCVTVDDATMCISHVIRAEEHLPNTLRQALIYKALGFPMPSFAHVSLILAPDRSKLSKRHGATSVGQFREMGY
LPQAMVNYLALLGWGDGTENEFFTIDQLVEKFSINRVNKSGAIFDSTKLRWMNGQHLRALPSEKLIEIFGDRWKGTGILLESKSKFVEDATELLKDGIDLVTDADTA
LLNLLSYPLDSTLASAECKPVVQDKLSEVAASLISAYDSGELREALEEGHAGWQKWVKSFGKLLKRKGRSLFMPIRLLLTGKLHGPDMGASIVLIYKAGTCGAVSP
QAGFVTLDERFKILRELRWEALNKEAEPHAESVATVTQ
>E. guineensis XP_010951452
MEIAVAFSADSPPLPVIAAARIAGVPLSTDPSLPPGSAPVFRFDSGLKLQGTNTLLRYIGRVASIPNFYGQDALESGQIDEWLDYAPIFLSGSEFEGACSDVDGYLSLRT
FLVGYSLSIADIAVWSGLAGTGQRWESLRKSKKYQNLVRWFNSIAVEYGHALNEVTAAFVGKRGLGKPAAKPTGVVSGSNPDLKGKETNGPVVSKANTAGFEVDL
PGAKVGEVCLRFAPEPSGYLHIGHSKAALLNQYFAQRYQGRLIIRFDDTNPSKESNEFVENLLKDIETLGIKYDAVTYTSDYFPQLMKMAENLILKGKAYVDDTPRE
QMQKERMDGIESKCRNNTVDENLALWKEMIAGSERGMQCCLRGKLDMQDPNKSLRDPVYYRYNPDPHHRIGSKYKVYPTYDFACPFVDSIEGVTHALRSSEYHD
RNAQYLQILEDMGLRKVQIYEFSRLNMVYTLLSKRKLLWFVQNGKVEGWDDPRFPTVQGIVRRGLKIEALIQFILQQGASKNLNLMEWDKLWTINKKIIDPICPRH
TAVLEENRVMLLLTDGPDGPEKPFVRIMPRHKKFEGAGTKATTYTKRIWIDYADALSITEGEEVTLMDWGNAIIKEIKKDNAKITHLIGVLHLEGSVKTTKLKLTWL
PEMDELVHLSLVEFDYLIKKKKLEEGEDFLDNLNPCTRRETAALGDSNMRNLKRGEILQIERKGYYRCDVPFLRPSKPIVLFAIPDGRQQTSLN
>HvERS1
MELKLAFPQDSLPLSIISAAKIAGVPLIIDPTLASGSVPTLHFSSGDFIHGVNTILRYIARAASVSSFYGQDHIQASHVDQWLEYAPLILSGSEFEAACSFLDGYLSSRTF
LVGYGLSIADIVVWSNLTGTGQRWESLRRSKKYQSLVRWFNSVAADYADALDEVTSAYVGKRGIGKSPAPSLKEKVPGLKENTSGHEIDLPGATVGEVCVRFAPE
PSGYLHIGHAKAALLNKYFAERYKGRLIVRFDDTNPSKESNEFVENVLKDIETLGVKYDVVTYTSDHFPKLMEMAESLIKQGKAYVDDTPKEQMRSERMDGVES
KRRNSTVEENLSLWKEMVNGTKRGTECCVRGKLDMQDPNKSLRDPVYYRCNPDPHHRVGSKYKVYPTYDFACPFVDALEGVTHALRSSEYHDRNAQYYRILQD
MGLRRVEIYEFSRLNMVYTVLSKRKLLWFVQNRMVEDWTDARFPTVQGIVRRGLKIEALIQFILEQGASKNLNLMEWDKLWTINKKIVDPVCGRHTAVLKDKRV
LLTLTNGPEEPFVRILPRHKKHEGAGKKATTFANRIWLEYADASVVSVGEEVTLMDWGNAIIREIKTNGTVTQLVGELHLEGSVKMTKLKLTWLSDIEDLVSLSLV
EFDYLINKKKLEEDEDFLDNLNPCTRREALALGDPNMRSVKKGEVIQLERKGYYRCDVPFVRSSKPIMLFAIPDGRQKSTSIVTGA
>HvERS2
MEARLAFPQESPPLSIISAAKIAGVSIIIDPTLASGSVPTLHFSSGDFIHGVNTILRYIARAASVSSFYGQDDIQAAHVDQWLEYASVILSGSEFEVACSFIDGYLATRTF
LVGYGLSIADIVVWSNLTGTGQRWESLRRSKKYQSLVRWFNSVAADYADTLDEVTSAYVGKRGIGKSPAPSLKEKMPGLKENASGHEIDLPGAKVGDVCVRFAP
EPSGYLHIGHAKAALLNKYFAERYKGRLIVRFDDTNPSKESNEFVENVLKDIETLGVKYDVVTYTSDYFPKLMEMAESLIKQGKAYIDDTPKEQMRSERMDGVES
KCRNSTVEENLLLWKEMVNGTKRGTQCCVRGKLDMQDPNKSLRDPVYYRCNPDPHHRVGSKYKVYPTYDFACPFVDALEGVTHALRSSEYHDRNAQYYRILQ
DMGLRRVEIYEFSRLNMVYTVLSKRKLLWFVQNKKVEDWTDPRFPTVQGIVRRGLKIEALIQFILEQGASKNLNLMEWDKLWTINKKIVDPVCGRHTAVLKDKC
VLLTLTNGPEEPFVRILPRHKKHEGAGKKATTFANRIWLEYADASVISVGEEVTLMDWGNAIIREIKTDNGTVTQLVGELHLEGSVKMTKLKLTWLSDIEDLVSLS
LVDFDYLINKKKLEEDEDFLDNLNPCTRREALALGDPNMRNVKKGEVIQLERKGYYRCDVPFVRSSKPIMPFAIPDGRQKSTSIVTGA
>HvERS3
MAASNFMGSSARLRVGLLPSVTPRLSRRALATRASADSGGSGPVRVRFAPSPTGNLHVGGARTALFNYLFARSRGGKFVLRVEDTDLERSTKKSEEAVLTDLSWL
GLDWDEGPDIGGDFGPYRQSERNALYKEHAQKLMESGAVYRCFCSNEELEKMKETANRMKIPPVYMGKWATASDAEVQQELEKGTPYTYRFRVPKEGSLKIND
LIRGEVSWNLNTLGDFVIMRSNGQPVYNFCVTVDDATMRISHVIRAEEHLPNTLRQALIYKALGFAMPLFAHVSLILAPDKSKLSKRHGATSVGQYKEMGYLPQA
MVNYLALLGWGDGTENEFFTIDDLVEKFTIDRVNKSGAVFDATKLKWMNGQHLRSLPSDLLIKDFEDQWRSTGILLESESGFAKEAAELLKEGIDLITDADAALCK
LLSYPLHETLSSDEAKSVVEDKLSEVASGLISAYDSGELDQALAEGHDGWKKWVKSFGKTHKRKGKSLFMPLRVLLTGKLHGPAMDSTVILVHKAGTSGAVAPQ
SGFVSLDERFKILKEVNWESLQKQQESPVESAVPAAS

Appendix 2

Supplemental Table S2 (Chapter 3)

Heat map of metabolites changes between wild type and *osers1*.

Supplemental Table S1. Heat map of metabolites change between wild type and *osers1*.

Red and green shaded cells indicate ps0.05 (red indicates that the mean values are significantly higher for that comparison; green values significantly lower). Light red and light green shaded cells indicate 0.05<p<0.10 (light red indicates that the mean values trend higher for that comparison; light green values trend lower).

								MUTANT EFFECTS		
								Fold of Change (Ratios of Group Means in Scaled Imputed Data)	Statistical Results	
									osers1 / Wild type	
Super Pathway	Sub Pathway	Biochemical Name	Platform	Comp ID	KEGG	PlantCyc	PUBCHEM	*osers1* / Wild type	p-value	q-value
Amino acid	Serine family (phosphoglycerate derived)	cysteine	GC/MS	31453	C00097	CYS	5862	1.1	0.4122	0.1076
		cystine	GC/MS	31454	C00491	CYSTINE	67678	4.08	6.38E-06	1.88E-05
		glycine	GC/MS	11777	C00037	GLY	750	1.34	0.0075	0.0035
		homocysteine	GC/MS	40266	C00155	HOMO-CYS	778	2.48	0.0041	0.0022
		N-acetylserine	GC/MS	37076		ACETYLSERINE	65249	1.59	0.0011	0.0007
		O-acetylserine	GC/MS	15947	C00979	ACETYLSERINE	99478	1.51	0.0156	0.0068
		serine	GC/MS	1648	C00065	SER	5951	1.25	0.0002	0.0002
		S-methylcysteine	GC/MS	39592		S-METHYL-L-CYS	24417	1.7	0.0007	0.0005
	Aromatic amino acid metabolism (PEP derived)	3-(4-hydroxyphenyl)lactate	LC/MS neg	32197	C03672	CPD-2222	9378	1		
		phenylacetate	LC/MS neg	15958	C07086		999	1.35	0.1325	0.0406
		phenylalanine	LC/MS pos	64	C00079	PHE	6140	1.22	0.0426	0.0155
		quinate	GC/MS	18335	C00296	QUINATE	6508	1.94	1.88E-06	1.01E-05
		serotonin	LC/MS pos	2342	C00780	SEROTONIN	5202	1.26	0.3148	0.0836
		shikimate	GC/MS	15144	C00493	SHIKIMATE	8742	2.43	5.02E-06	1.71E-05
		tryptophan	LC/MS pos	54	C00078	TRP	6305	0.98	0.8478	0.1961
		tyrosine	LC/MS pos	1299	C00082	TYR	6057	1.25	0.0481	0.0169
		N-formylphenylalanine	LC/MS neg	48433			759256	1.3	0.132	0.0406
	Aspartate family (OAA derived)	2-aminoadipate	GC/MS	6146	C00956	CPD-468	469	2.78	1.09E-07	1.93E-06
		alanine	GC/MS	1126	C00041	L-ALPHA-ALANI	5950	1.23	0.0062	0.003
		allo-threonine	GC/MS	15142	C05519	CPD-10303	99289	1.72	0.0036	0.002
		asparagine	GC/MS	11398	C00152	ASN	6267	3.15	3.63E-07	4.70E-06
		aspartate	GC/MS	15996	C00049	L-ASPARTATE	5960	1.76	9.86E-07	9.12E-06
		homoserine	GC/MS	23642	C00263	HOMO-SER	12647	0.58	3.38E-06	1.37E-05
		homoserine lactone	GC/MS	34458	C02926	HOMO-SER	445963	1.92	0.0002	0.0002
		lysine	LC/MS pos	1301	C00047	LYS	5962	1.88	0.0003	0.0002
		methionine	LC/MS pos	1302	C00073	MET	6137	1.48	0.002	0.0012
		methionine sulfoxide	LC/MS pos	18374	C02989	L-Methionine-S	158980	1.61	0.0295	0.0114
		N6-acetyllysine	LC/MS pos	36752	C02727		92832	1.26	0.0298	0.0114
		N-6-trimethyllysine	LC/MS pos	1498	C03793		440120	1.28	0.0736	0.0243
		N-acetylmethionine	LC/MS neg	1589	C02712		448580	1.88	0.0031	0.0017
		pipecolate	LC/MS pos	1444	C00408		849	2.14	4.22E-06	1.61E-05
		S-adenosylhomocysteine (SAH)	LC/MS neg	15948	C00021	ADENOSYL-HOMO	439155	1.04	0.7211	0.1698
		threonine	GC/MS	1284	C00188	THR	6288	1.32	7.25E-05	8.76E-05
		6-oxopiperidine-2-carboxylic acid	GC/MS	43231			3014237	1.69	0.0758	0.0249
		2-hydroxybutyrate (AHB)	GC/MS	21044	C05984	Hydroxy-Butyra	440864	1.3	0.2797	0.0771
		pyrraline	LC/MS pos	48428				1		
		S-methylmethionine	LC/MS pos	38127	C05319	CPD-397	458	1.37	8.61E-05	9.80E-05
	Glutamate family (alpha-ketoglutarate derived)	1,3-diaminopropane	GC/MS	1654	C00986	CPD-313	428	0.89	0.6012	0.144
		2-aminobutyrate	GC/MS	1577	C02261	2-Aminobutyrat	439691	1.06	0.4354	0.1119
		2-pyrrolidinone	GC/MS	31675			12025	0.54	0.0183	0.0078
		4-hydroxybutyrate (GHB)	GC/MS	34585	C00989	4-HYDROXY-BUT	10413	0.88	0.0268	0.0108
		arginine	LC/MS pos	1638	C00062	ARG	232	1.76	0.0002	0.0002
		argininosuccinate	GC/MS	15497	C03406	L-ARGININO-SUC	16950;828	2.63	0.0021	0.0012
		citrulline	LC/MS pos	2132	C00327	L-CITRULLINE	9750	0.85	0.2304	0.0649
		dimethylarginine (SDMA + ADMA)	LC/MS pos	36808	C03626		123831	1.16	0.0056	0.0028
		gamma-aminobutyrate (GABA)	GC/MS	1416	C00334	4-AMINO-BUTYRA	119	1.07	0.5445	0.1336
		glutamate	LC/MS pos	57	C00025	GLT	611	2.44	0.0031	0.0017
		glutamine	LC/MS pos	53	C00064	GLN	5961	1.46	2.79E-05	5.31E-05
		histidine	LC/MS neg	59	C00135	HIS	6274	2.68	1.19E-07	1.93E-06
		N-acetylglutamate	LC/MS pos	15720	C00624	ACETYL-GLU	70914	1.66	0.0526	0.0182
		ornithine	GC/MS	1493	C00077	L-ORNITHINE	6262	1.6	0.0002	0.0002
		proline	LC/MS pos	1898	C00148	PRO	145742	0.89	0.0051	0.0026
		pyroglutamine*	LC/MS pos	32672			134508	1.61	1.86E-06	1.01E-05
		trans-4-hydroxyproline	GC/MS	1366	C01157	Protein-hydrox	5810	0.96	0.283	0.0772
		4-hydroxy-2-oxoglutaric acid	GC/MS	40062	C01127		599	1.04	0.9059	0.2052
		4-hydroxyglutamate	GC/MS	40499	C03079		439902	2.54	0.0072	0.0034
		N-delta-acetylornithine*	LC/MS pos	43249			9920500	1.38	0.0014	0.0009
		4-guanidinobutanoate	LC/MS pos	15681	C01035		500	1.39	0.0122	0.0054
		guanidine	GC/MS	22287	C17349		3520	2.8	0.0008	0.0006
		1-propanoylproline	LC/MS pos	43505			3614637	1.32	0.3986	0.1045
	Branched Chain Amino Acids (OAA derived)	isoleucine	LC/MS pos	1125	C00407	ILE	6306	2.11	7.08E-05	8.76E-05
		allo-isoleucine	GC/MS	46552			6950182;99288	1.07	0.9101	0.2054
	Branched Chain Amino Acids (pyruvate derived)	2,3-dihydroxyisovalerate	GC/MS	38276	C04039	DIOH-ISOVALERA	677	0.87	0.2837	0.0772
		norvaline	GC/MS	37426	C01826	L-2-AMINOPENTA	824	1.97	0.0004	0.0003
		3-methyl-2-oxovalerate	LC/MS neg	15676	C00671	2-KETO-3-METHY	47	0.46	0.0297	0.0114
		4-methyl-2-oxopentanoate	LC/MS neg	22116	C00233	2K-4CH3-PENTAN	70	0.37	0.0344	0.0127
		alpha-hydroxyisovalerate	GC/MS	33937			99823	0.67	0.088	0.0282
		leucine	LC/MS pos	60	C00123	LEU	6106	1.5	0.0004	0.0003
		methylsuccinate	GC/MS	15745		CPD-2461	10349	1.95	0.0021	0.0012
		N-acetylvaline	LC/MS pos	1591			66789	1		
		valine	LC/MS pos	1649	C00183	VAL	6287	1.34	0.0057	0.0028
		acetoacetate	GC/MS	33963	C00164	3-KETOBUTYRATE	96	0.79	0.1553	0.0461
	Amines and polyamines	5-methylthioadenosine (MTA)	LC/MS pos	1419	C00170	5-METHYLTHIOAD	439176	1.08	0.4888	0.1232
		agmatine	GC/MS	15496	C00179	AGMATHINE	199	1.27	0.0457	0.0163
		putrescine	GC/MS	1408	C00134	PUTRESCINE	1045	1.02	0.9553	0.2133
		spermidine	LC/MS pos	485	C00315	SPERMIDINE	1102	1.2	0.0447	0.0161
		nicotianamine	GC/MS	43026	C05324		9882882	1.53	0.0009	0.0006
	Glutathione metabolism	5-oxoproline	LC/MS neg	1494	C01879	5-OXOPROLINE	7405	2.19	0.0003	0.0002
		cysteine-glutathione disulfide	LC/MS pos	35159			4247235	1.55	0.0002	0.0002
		glutathione, oxidized (GSSG)	LC/MS pos	27727	C00127	GLUTATHIONE	65359	0.23	4.42E-05	7.16E-05
		glutathione, reduced (GSH)	LC/MS pos	2127	C00051	GLUTATHIONE	124886	0.22	0.0002	0.0002
		ophthalmate	LC/MS pos	34592			7018721	0.74	0.0001	0.0001
		norophthalmate*	LC/MS pos	39251			5489007	0.67	0.0023	0.0014
	Glycolysis	fructose-6-phosphate	GC/MS	12021	C05345	FRUCTOSE-6P	69507	0.26	2.69E-06	1.24E-05
		glucose	GC/MS	20488	C00031	CPD-3607	79025	0.59	0.0002	0.0002
		glucose 6-phosphate	GC/MS	31260	C00668	ALPHA-GLC-6-P	5958	0.25	0.0003	0.0003
		lactate	GC/MS	527	C00186	L-LACTATE	612	0.5	0.011	0.005
		pyruvate	GC/MS	599	C00022	PYRUVATE	1060	1.3	0.0194	0.0081
	TCA cycle	alpha-ketoglutarate	GC/MS	33453	C00026	2-ketoglutarat	51	1.2	0.0389	0.0158
		citrate	GC/MS	1564	C00158	CIT	311	1.19	0.0051	0.0026
		fumarate	GC/MS	1643	C00122	FUM	444972	0.78	0.0117	0.0052
		malate	GC/MS	1303	C00149	MAL	525	1.7	6.05E-05	8.71E-05
		mesaconate (methylfumarate)	GC/MS	18493	C01732		638129	1.39	0.0171	0.0074
		succinate	GC/MS	1437	C00042	SUC	1110	1.04	0.535	0.1328
	Photorespiration	glycolate (hydroxyacetate)	GC/MS	15737	C00160	GLYCOLLATE	757	0.91	0.0485	0.017

	Photorespiration	glycerate	GC/MS	1572	C00258	GLYCERATE	752	1.31	0.0006	0.0004
Carbohydrate	Amino sugar and nucleotide sugar	arabinose	GC/MS	575	C00216	D-ARABINOSE	66308	0.91	0.4184	0.1088
		arabitol	GC/MS	38075	C01904		94154	2.42	4.61E-05	7.29E-05
		arabonate	GC/MS	37516	C00878		122045	1.86	7.87E-08	1.93E-06
		erythritol	GC/MS	20699	C00503	2-C-METHYL-D-E	222285	1.44	5.98E-05	8.71E-05
		erythronate*	GC/MS	33477			2781043	1.46	5.34E-05	8.23E-05
		glucosaminate	GC/MS	27681	C03752		73563	1.24	0.0628	0.0212
		glucosamine	GC/MS	18534	C00329	GLUCOSAMINE	441477	1.27	0.3172	0.0839
		N-acetylglucosamine	GC/MS	15095	C00140	N-acetyl-D-glu	24139	1.92	0.0002	0.0002
		ribose	GC/MS	12080	C00121	RIBOSE	5779	2.19	9.64E-05	0.0001
		ribulose	GC/MS	35855	C00309	Ribuloses	151261	1.17	0.3041	0.0819
		threitol	GC/MS	35854	C16884	CPD-6995	169019	1.73	5.89E-05	8.71E-05
		xylonate	GC/MS	35638	C05411		6602431	1.64	1.10E-05	2.45E-05
		xylose	GC/MS	15835	C00181	XYLOSE	135191	1.54	0.0039	0.0021
		mannosamine	GC/MS	40517	C03570		440049	0.84	0.8129	0.1894
	Inositol metabolism	inositol 1-phosphate (I1P)	GC/MS	1481	C04006	1-L-MYO-INOSIT	440194	0.76	0.018	0.0077
		myo-inositol	GC/MS	19934	C00137	MYO-INOSITOL	892	0.5	0.0078	0.0037
	Sucrose, glucose,fructose metabolism	1-kestose	GC/MS	22013	C03661	1-KESTOTRIOSE	440080	2.02	0.0365	0.0134
		3-deoxyoctulosonate	GC/MS	15942		KDO	4636210	1.4	0.0188	0.008
		fructose	GC/MS	577	C00095	Fructose	5984	0.57	0.0002	0.0002
		galactinol	GC/MS	21034	C01235	CPD-458	439451	1.37	0.029	0.0113
		galactitol (dulcitol)	GC/MS	1117	C01697		11850	1.86	0.0001	0.0001
		galactose	GC/MS	12055	C01582	L-GALACTOSE	3037556	1.87	0.0031	0.0017
		maltose	GC/MS	15806	C00208	MALTOSE	10991489	0.28	0.0005	0.0004
		maltotriose	GC/MS	27723	C01835	MALTOTRIOSE	439586	0.77	0.2158	0.0616
		mannitol	GC/MS	15335	C00392	MANNITOL	6251	1.13	0.1609	0.0475
		mannose-6-phosphate	GC/MS	1470	C00275	MANNOSE-6P	439198	0.44	3.80E-05	6.47E-05
		raffinose	LC/MS neg	586	C00492	CPD-1099	10542	2.16	0.0065	0.0031
		rhamnose	GC/MS	15826	C00507	L-rhamnose	25310	1.56	0.0006	0.0004
		sorbitol	GC/MS	15053	C00794	SORBITOL	5780	0.82	0.0597	0.0203
		sucrose	LC/MS neg	1519	C00089	SUCROSE	5988	1.97	0.0026	0.0015
		trehalose	GC/MS	15573	C01083	TREHALOSE	7427	1.28	0.0473	0.0168
		gluconate	GC/MS	587	C00257	GLUCONATE	10690	0.54	0.004	0.0021
	C5 branched dibasic acid metabolism	citramalate	GC/MS	22158	C00815		1081	1.71	1.23E-05	2.66E-05
Lipids	Free fatty acid	10-heptadecenoate (17:1n7)	LC/MS neg	33971			5312435	1.5	0.0559	0.0193
		10-nonadecenoate (19:1n9)	LC/MS neg	33972			5312513	1.95	0.0095	0.0044
		2-hydroxypalmitate	LC/MS neg	35675			92836	1.68	0.0011	0.0007
		2-hydroxystearate	LC/MS neg	17945	C03045		69417	1.81	0.003	0.0017
		3-hydroxypropanoate	GC/MS	42103	C01013	3-HYDROXY-PROP	68152	1.3	0.0366	0.0134
		8-hydroxyoctanoate	LC/MS neg	21239			69820	1.38	0.1999	0.0573
		arachidate (20:0)	LC/MS neg	44679	C06425	ARACHIDIC_ACID	10467	1.58	0.034	0.0127
		caproate (6:0)	LC/MS neg	32489	C01585		8892	1.14	0.4758	0.1208
		caprylate (8:0)	LC/MS neg	32492	C06423	CPD-195	379	2.22	0.09	0.0287
		cis-vaccenate (18:1n7)	GC/MS	33970	C08367	CPD-9247	5282761	1.28	0.0046	0.0024
		dihomo-linoleate (20:2n6)	LC/MS neg	17805	C16525		6439848	1.12	0.6025	0.144
		dihomo-linolenate (20:3n3 or n6)	LC/MS neg	35718	C03242	CPD-8120	5280581	1.62	0.0246	0.01
		eicosenoate (20:1)	LC/MS neg	33587		CPD-12189	5282768	1.73	0.0142	0.0063
		hexacosanoate (26:0)	GC/MS	21143			10469	0.99	0.9551	0.2133
		isovalerate	LC/MS neg	34732	C08262		10430	1.64	0.1507	0.0452
		lignocerate (24:0)	GC/MS	1364	C08320	TETRACOSANOATE	11197	1.02	0.6753	0.1596
		linoleate (18:2n6)	LC/MS neg	1105	C01595	LINOLEIC_ACID	5280450	0.87	0.2299	0.0649
		linolenate [alpha or gamma; (18:3n3 or 6)]	LC/MS neg	34035	C06426	CPD-8117	5280934	0.95	0.6097	0.1452
		margarate (17:0)	LC/MS neg	1121			10465	1.17	0.5576	0.1358
		mead acid (20:3n9)	LC/MS neg	35174			5312531	1.55	0.1324	0.0406
		nonadecanoate (19:0)	LC/MS neg	1356	C16535		12591	0.97	0.9015	0.2049
		oleate (18:1n9)	GC/MS	1359	C00712	OLEATE-CPD	445639	1.7	0.0001	0.0001
		palmitate (16:0)	LC/MS neg	1336	C00249	PALMITATE	985	0.85	0.3048	0.0819
		palmitoleate (16:1n7)	LC/MS neg	33447	C08362	CPD-9245	445638	1.02	0.8991	0.2049
		pelargonate (9:0)	LC/MS neg	12035	C01601		8158	1.23	0.282	0.0772
		pentadecanoate (15:0)	LC/MS neg	1361	C16537		13849	0.92	0.5547	0.1356
		stearate (18:0)	LC/MS neg	1358	C01530	STEARIC_ACID	5281	0.89	0.4356	0.1119
		stearidonate (18:4n3)	LC/MS neg	33969	C16300	CPD-12653	5312508	0.98	0.8687	0.2002
		17-methylstearate	LC/MS neg	38296			3083779	1.2	0.1477	0.0445
	Fatty acid amide	oleic ethanolamide	LC/MS neg	38102			5283454	3.59	0.0001	0.0001
		palmitoyl ethanolamide	LC/MS neg	38165	C16512		4671	2.95	0.0001	0.0001
		stearamide	LC/MS pos	37487	C13846		31292	1.33	0.2261	0.0642
		palmitic amide	LC/MS pos	42095			69421	1.11	0.541	0.1333
	Oxylipins	13-HODE + 9-HODE	LC/MS neg	37752			43013	3.1	2.65E-05	5.31E-05
		12,13-DiHOME	LC/MS neg	38395	C14829	CPD-13086	10236635	3.37	0.0006	0.0004
		9,10-epoxystearate	LC/MS neg	39627	C19418	CPD-10514	15868	3.25	0.0004	0.0003
	Fatty acid ester	15-methylpalmitate (isobar with 2-methylpalmitate)	LC/MS neg	38768			17903417	0.86	0.1916	0.0557
	Glycerolipids	1-linoleoylglycerol (18:2)	LC/MS neg	27447			5283469	0.64	0.0433	0.0157
		1-oleoylglycerol (18:1)	LC/MS neg	21184			5283468	1.43	0.1546	0.0461
		1-palmitoylglycerol (16:0)	LC/MS neg	21127			14900	1.07	0.5974	0.1438
		2-linoleoylglycerol (18:2)	LC/MS neg	32506			5365676	0.67	0.0865	0.0279
		2-palmitoylglycerol (16:0)	LC/MS neg	33419			123409	1.39	0.6731	0.1596
		glycerol	GC/MS	15122	C00116	GLYCEROL	753	1.24	1.64E-06	1.01E-05
		glycerophosphoglycerol	GC/MS	48857	C03274		439964	0.78	0.0111	0.005
	Phospholipids	1-linoleoyl-GPC (18:2)	LC/MS pos	34419	C04100		11988421	0.5	0.424	0.1098
		1-linoleoyl-GPE (18:2)*	LC/MS neg	32635			52925130	0.43	0.0046	0.0024
		1-linoleoyl-GPI (18:2)*	LC/MS neg	36594				0.72	0.0969	0.0306
		1-oleoyl-GPC (18:1)	LC/MS pos	33960			16081932	0.77	0.7282	0.1709
		1-oleoyl-GPE (18:1)	LC/MS neg	35628			9547071	0.92	0.5065	0.1267
		1-palmitoyl-GPA (16:0)	LC/MS neg	34428	C04036		6419701	1.14	0.5414	0.1333
		1-palmitoyl-GPC (16:0)	LC/MS pos	33955		2-Acylglycero	86554	0.44	0.1994	0.0573
		1-palmitoyl-GPE (16:0)	LC/MS neg	35631		2-Lysophospha	9547069	0.44	0.0018	0.0011
		1-palmitoyl-GPI (16:0)*	LC/MS neg	35305				0.38	0.002	0.0012
		1-stearoyl-GPC (18:0)	LC/MS pos	33961			497299	0.51	0.842	0.1955
		2-palmitoyl-GPC (16:0)*	LC/MS pos	35253			15061532	0.7	0.4819	0.1219
		glycerol 3-phosphate	GC/MS	15365	C00093	GLYCEROL-3P	754	0.89	0.0771	0.0252
		glycerophosphoethanolamine	GC/MS	37455	C01233		123874	0.49	1.05E-05	2.44E-05
		glycerophosphorylcholine (GPC)	LC/MS pos	15990	C00670	SN-3-GLYCEROP	71920	0.43	5.66E-06	1.78E-05
		phosphoethanolamine	GC/MS	12102	C00346	PHOSPHORYL-ETH	1015	1.08	0.4386	0.1123
		1-linolenoyl-GPC (18:3)*	LC/MS pos	44562				0.47	0.1615	0.0475
		1-palmitoyl-GPG (16:0)*	LC/MS neg	45970			3300276	0.62	0.0171	0.0074
	Choline metabolism	choline	LC/MS pos	15506	C00114	CHOLINE	305	0.78	2.68E-05	5.31E-05
		ethanolamine	GC/MS	1497	C00189	ETHANOL-AMINE	700	1.61	0.0213	0.0088
	Sphingolipid	phytosphingosine	LC/MS pos	1510	C12144	PHYTOSPINGOSIN	122121	0.93	0.9436	0.2122
		sphinganine	LC/MS pos	17769	C00836	DIHYDRO-SPHING	3126	0.83	0.2594	0.0718
		sphingosine	LC/MS pos	17747	C00319	SPHINGOSINE	5353955	1		
	Sterols	beta-sitosterol	GC/MS	27414	C01753	CPD-4143	222284	1.23	0.0011	0.0007
		campesterol	GC/MS	39511	C01789	CPD-707	173183	1.6	0.005	0.0026
		stigmasterol	GC/MS	21183	C05442	CPD-4162	5280794	1.28	3.92E-05	6.51E-05

	Fatty acid, Dicarboxylate	**azelate (nonanedioate)**	LC/MS neg	18362	C08261		2266	3.34	0.0006	0.0004
		docosadioate	LC/MS neg	39837			244872	0.74	0.1064	0.0333
		dodecanedioate	LC/MS neg	32388	C02678		12736	2.17	0.0036	0.002
		sebacate (decanedioate)	LC/MS neg	32398	C08277		5192	1.08	0.0727	0.0242
		suberate (octanedioate)	LC/MS pos	15730	C08278	CPD0-1264	10457	1.76	0.0306	0.0116
		undecanedioate	LC/MS neg	35671			15816	1.69	0.0019	0.0012
		2-hydroxyglutarate	GC/MS	37253	C02630	CPD-381	43	1.99	6.20E-05	8.72E-05
		adipate	LC/MS pos	21134	C06104	ADIPATE	196	1.94	0.0301	0.0115
Cofactors, Prosthetic Groups, Electron Carriers	CoA metabolism	**pantothenate**	LC/MS pos	1508	C00864	PANTOTHENATE	6613	1.15	0.0282	0.0111
	Nicotinate and nicotinamide metabolism	**nicotinamide**	LC/MS pos	594	C00153	NIACINAMIDE	936	1.44	0.0217	0.0089
		nicotinamide adenine dinucleotide (NAD+)	LC/MS pos	5278	C00003	NAD	5893	0.37	3.09E-05	5.71E-05
		nicotinamide ribonucleotide (NMN)	LC/MS pos	22152	C00455	NICOTINAMIDE N	14180	0.76	0.0042	0.0022
		nicotinamide riboside*	LC/MS pos	33013	C03150	NICOTINAMIDE R	439924	1.51	6.82E-05	8.76E-05
		nicotinate	LC/MS pos	1504	C00253	NIACINE	938	1.86	8.62E-05	9.80E-05
		nicotinate ribonucleoside*	LC/MS pos	33471	C05841	NICOTINAMIDE R	161234	1.28	0.0002	0.0002
		trigonelline (N'-methylnicotinate)	LC/MS pos	32401	C01004	METHYLNICOTINA	5570	1.33	7.18E-05	8.76E-05
	Oxidative phosphorylation	**methylphosphate**	GC/MS	37070			13130	0.88	0.1665	0.0488
		phosphate	GC/MS	11438	C00009	Pi	1061	1.12	0.0274	0.0109
	Riboflavin and FAD metabolism	**flavin adenine dinucleotide (FAD)**	LC/MS neg	2134	C00016	FAD	643975	1.07	0.3111	0.0829
		flavin mononucleotide (FMN)	LC/MS neg	15797	C00061	FMN	710	1.13	0.1417	0.0429
	Ascorbate metabolism	**ascorbate (Vitamin C)**	GC/MS	1640	C00072	ASCORBATE		0.06	6.64E-05	8.76E-05
		dehydroascorbate	GC/MS	1659	C05422	L-DEHYDRO-ASC	835	1.91	0.0516	0.018
		threonate	GC/MS	27738	C01620	L-THREONATE	151152	0.85	0.0967	0.0306
	Tocopherol metabolism	**alpha-tocopherol**	GC/MS	1561	C02477	ALPHA-TOCOPHE	14985	1.11	0.1793	0.0523
		alpha-tocotrienol	GC/MS	37502	C14153		5282347	0.8	0.0479	0.0169
		gamma-tocopherol	GC/MS	33420	C02483	GAMA-TOCOPHERO	14986	1.13	0.8866	0.2029
	Vitamin B metabolism (B6 or B12)	**pyridoxal**	LC/MS pos	1651	C00250	PYRIDOXAL	1050	1.29	0.0319	0.012
		pyridoxate	LC/MS neg	31555	C00847		6723	1.6	0.1343	0.0408
		pyridoxine (Vitamin B6)	LC/MS pos	608	C00314	PYRIDOXINE	1054	1.72	0.0034	0.0019
	Chlorophyll and heme metabolism	**pheophorbide A**	LC/MS pos	35879	C18021	CPD-7061	5323510	0.92	0.4749	0.1208
Nucleotide	Purine metabolism	**2'-deoxyadenosine**	LC/MS pos	1553	C00559	DEOXYADENOSINE	13730	1.77	0.0078	0.0037
		2'-deoxyguanosine	LC/MS neg	1411	C00330	DEOXYGUANOSINE	187790	1.34	0.0265	0.0107
		adenine	LC/MS pos	554	C00147	ADENINE	190	1.92	3.35E-06	1.37E-05
		adenosine	LC/MS pos	555	C00212	ADENOSINE	60961	0.96	0.4982	0.1251
		adenosine 2'-monophosphate (2'-AMP)	LC/MS neg	36815	C00946	CPD-3705	94136	1.77	8.14E-06	2.20E-05
		adenosine 3'-monophosphate (3'-AMP)	LC/MS pos	35142	C01367	CPD-3706	41211	1.42	0.0632	0.0212
		adenosine 5'-monophosphate (AMP)	LC/MS pos	32342	C00020	AMP	6083	0.59	0.0002	0.0002
		adenosine-2',3'-cyclic monophosphate	LC/MS pos	37467	C02353		2024	1.46	0.0206	0.0085
		allantoin	GC/MS	1107	C02350	ALLANTOIN	204	3.57	1.64E-06	1.01E-05
		guanine	GC/MS	418	C00242	GUANINE	764	1.7	0.0067	0.0032
		guanosine	LC/MS pos	1573	C00387	GUANOSINE	6802	0.93	0.1331	0.0406
		guanosine-2',3'-cyclic monophosphate	LC/MS neg	37139	C06194		92823	2.52	1.22E-06	9.85E-06
		inosine	LC/MS neg	1123	C00294	INOSINE	6021	0.86	0.5227	0.1302
		N1-methyladenosine	LC/MS pos	15650	C02494		27476	1.57	0.0108	0.005
		urate	GC/MS	1604	C00366	URATE	1175	1.94	0.1121	0.0349
		xanthine	GC/MS	3147	C00385	XANTHINE	1188	0.83	0.3066	0.082
		2'-O-methyladenosine	LC/MS pos	42607			102213	0.89	0.5768	0.1399
	Pyrimidine metabolism	**3-aminoisobutyrate**	GC/MS	1566	C05145	3-AMINO-ISOBUT	64956	2.1	9.66E-06	2.38E-05
		cytidine	LC/MS pos	514	C00475	CYTIDINE	6175	1.43	0.0004	0.0003
		cytosine	GC/MS	573	C00380	CYTOSINE	597	1.96	0.0003	0.0003
		cytidine 2',3'-cyclic monophosphate	LC/MS neg	37465	C02354		417654	1.22	0.0816	0.0264
		orotate	GC/MS	1505	C00295	OROTATE	967	1.36	0.0326	0.0122
		pseudouridine	LC/MS pos	33442	C02067		15047	3.59	6.60E-05	8.76E-05
		thymine	GC/MS	604	C00178	THYMINE	1135	5.8	9.06E-06	2.35E-05
		uracil	GC/MS	605	C00106	URACIL	1174	4.59	9.74E-07	9.12E-06
		uridine	LC/MS neg	606	C00299	URIDINE	6029	1.31	0.0007	0.0005
		uridine-2',3'-cyclic monophosphate	LC/MS neg	37137	C02355		439715	1.75	7.08E-05	8.76E-05
		beta-alanine	GC/MS	55	C00099	B-ALANINE	239	2.63	0.0038	0.002
Peptide	gamma-glutamyl	**gamma-glutamylglutamine**	LC/MS pos	2730	C05283		150914	2.23	2.79E-05	5.31E-05
		gamma-glutamylisoleucine*	LC/MS pos	34456			14253342	2.17	9.91E-06	2.38E-05
		gamma-glutamylleucine	LC/MS pos	18369			151023	1.52	0.0002	0.0002
		gamma-glutamylphenylalanine	LC/MS pos	33422			111299	1.56	0.0003	0.0002
		gamma-glutamyltryptophan	LC/MS pos	33947			3989307	0.7	0.0061	0.003
		gamma-glutamyltyrosine	LC/MS pos	2734			94340	1.31	7.30E-05	8.76E-05
		gamma-glutamylvaline	LC/MS pos	32393			7015683	1.24	0.0115	0.0052
	Dipeptide	**valylglycine**	LC/MS pos	40475			136487	1.08	0.2928	0.0794
		alanylleucine	LC/MS pos	37093			259583	1.73	7.84E-05	9.23E-05
		glycylisoleucine	LC/MS pos	36659			88079	1.6	0.0286	0.0112
		glycylleucine	LC/MS pos	34398	C02155		92843	1.78	3.72E-05	6.47E-05
		glycylvaline	LC/MS pos	18357			97417	1.73	0.0017	0.0011
		isoleucylglycine	LC/MS pos	40008			342532	1	0.9907	0.2205
		threonylphenylalanine	LC/MS pos	31530			4099799;4099798	2.03	0.0002	0.0002
		lysylleucine	LC/MS pos	40020			4682588	1.91	4.72E-06	1.70E-05
		valylleucine	LC/MS pos	39994			352039	1.1	0.0584	0.02
		glutamine-leucine	LC/MS pos	42994				1.96	3.78E-05	6.47E-05
Hormone metabolism	Auxin metabolism	**indoleacetate**	LC/MS pos	27513	C00954	INDOLE_ACETATE	802	0.97	0.7924	0.1853
	Ethylene metabolism	**cyano-alanine**	GC/MS	35660	C02512	CPD-603	13538	3.13	0.0006	0.0004
Secondary metabolism	Benzenoids	**2,4,6-trihydroxybenzoate**	LC/MS neg	35892			66520	1.93	0.1961	0.0567
		4-hydroxybenzoate	GC/MS	41834	C00156	4-hydroxybenzo	135	3.94	3.50E-08	1.93E-06
		benzoyl-O-glucose	LC/MS pos	38302			314421	0.28	2.69E-06	1.24E-05
		hydroquinone beta-D-glucopyranoside	LC/MS neg	38160	C06186		346	1.59	0.02	0.0083
		salicylate	LC/MS neg	1515	C00805	CPD-110	338	1.23	0.245	0.0684
	Flavonoids	**isovitexin**	LC/MS neg	37159	C01714	ISOVITEXIN	162350	1.25	0.103	0.0324
		kaempferol	LC/MS neg	17785	C05903	CPD1F-90	5280863	1.16	0.8811	0.2024
		kaempferol 3-O-beta-glucoside	LC/MS pos	38145				0.79	0.0017	0.0011
		luteolin 6-C-glucoside	LC/MS pos	37473				1.34	0.0629	0.0212
		naringenin	LC/MS neg	21182	C00509	NARINGENIN-CHE	932	4.55	0.0003	0.0002
		quercetin 3-galactoside	LC/MS neg	34114	C10073	CPD1F-437	5281643	0.8	0.2429	0.0681
		chrysoeriol	LC/MS neg	40474	C04293	5?4-TRIHYDROX	5280666	1.54	0.0681	0.0227
	Phenylpropanoids	**4-hydroxycinnamate**	LC/MS neg	36770	C00811	COUMARATE	637542	0.6	0.0001	0.0001
		ferulate	LC/MS neg	35630	C01494	FERULIC-ACID	445858	0.97	0.5913	0.1429
		sinapate	LC/MS neg	21150	C00482	SINAPATE	637775	2.25	7.98E-06	2.20E-05
		coumaroylquinate	LC/MS neg	43599	C10441			1.49	0.0778	0.0253
		coniferin	GC/MS	43500	C00761		5280372	1.62	0.0271	0.0108
		vanillate	GC/MS	35639	C06672		8468	3.8	5.79E-06	1.78E-05
		gentisic acid-5-glucoside	LC/MS neg	48729			10914066	2.68	0.0009	0.0006
	Terpenoids	**squalene**	GC/MS	1518	C00751	SQUALENE	638072	0.93	0.3548	0.0934

Appendix 3

Sequences of primers used in Chapter 3 and 4.

Supplemental Table S4 (Chapter 3)

Supplemental Table S5. Sequences of primers used in the study		
Primer names	**Sequence(5'-3')**	**Type of analysis**
OS1002F	CGGGTCCCAAAACAGGC	Map based cloning
OS1002R	GTTCCAGCTCCGCCACT	
10tys2-R	CATCATTGCCAAGAACCCTA	
10tys6-F	TTAAGGGCCAAGGCCAATTC	
10tys6-R	GTCTCGTGGGTAAACGGAAG	
10tys7-F	ACAAGTCATAATCGGACTCTAA	
10tys7-R	TCTTTAGTGGATTCGGTTAA	
10tys8-F	TGGGTACACCATCGCTTCAT	
10tys8-R	GAAAGTTCTATTCAACGCTGTT	
10tys14-F	CAATGATTACCCTTATTAGTTCCG	
10tys14-R	GCTCACTAGTTTAATTGCTAGCTG	
10tys26-F	CTCGCTCCTAGTTGTTTGATTG	
10tys26-R	CAAGTGGAGTTCGTATGAATTAG	
10tys27-F	TAAAAGATGAAAGAATGGCAAG	
10tys27-R	GGGCATGTAACCTGAGTCAAAC	
10tys28-F	TTTATATAGCGTCCCCGGAC	
10tys28-R	CAAACACACCTATATTTCACGG	
10tys29-F	CCAAAGATGTTTTTATCCTGTG	
10tys29-R	TCTGTCAGAAACATTCCAGATC	
10tys30-F	CCACTACTTTTGCACGTTCTAC	
10tys30-R	TGTGGAGTGCACATTTAAATCT	
OS1003F	TAGAGGCATGTAAGGGTAGTGAG	
OS1003R	CCCCATAAATGTTGCAGACT	
HF	CCATGATTACGAATTCTGGGTTCATCCCATCCCATC	Genetic complement
HR	ATTCGAGCTGGTCACCAAAATACCATCTTTCGGGACTG	
ERS-eGFP-F	CATGCCATGGCCATGGAGGCGAAGTTATCGTTTTC	Subcellular localisation
ERS_G-eGFP-F	CATGCCATGGCCATGGGCATAGATTCAGACTACA	
ERS_204-eGFP-F	CATGCCATGGCCATGGAGGTGTGTGTTCGCTTTGCC	
ERS-eGFP-R	CCACTAGTGTTCAAACTGGCCTGCTGTCGAC	
ARC-mcherry-F	TGACCATGGTAGATCTCATGGCGGCGGCGGCGGCGGACT	
ARC-mcherry-R	TGCTCACCATACTAGTCTTAATGCTCGCATTAGGAACAG	
ARC-Y2H-inF	GGAGGCCAGTGAATTCATGGCGGCGGCGGCGGCGGACTT	Yeast two hybrid
ARC_G-Y2H-inF	GGAGGCCAGTGAATTCATGGCGACCAAATCTGTCTTCGATCC	
ARC-Y2H-inR	TCATCTGCAGCTCGAGCTTAATGCTCGCATTAGGAACAG	
ERS-Y2H-F	AAAAGGCCAATCCGGCCATGGAGGCGAAGTTATCGTTTTC	
ERS_G-Y2H-F	AAAAGGCCAATCCGGCCATGGCGACGGCGGAGGTCGGC	
ERS-Y2H-R	CGCGGGCCTTAGAGGCCGTTCAAACTGGCCTGCTGTCGAC	
ERS-nYFP-F	CGGACTCAGATCTCGAGCTCAAATGGAGGCGAAGTTATCGTTT	BiFC

ERS_G-nYFP-F	CGGACTCAGATCTCGAGCTCAAATGGGCATAGATTCAGACTACA	
ERS-nYFP-R	GGGCCCGCGGTACCGTCGACGTTCAAACTGGCCTGCTGTCGA	
ARC-cYFP-F	CGAGAGCTCAAATGGCGGCGGCGGCGGCGGACTT	
ARC-cYFP-R	GCAGTCGACCTTAATGCTCGCATTAGGAACAG	
ERS-his-F	AAAGAGCTCATGGAGGCGAAGTTATCG	Protein expression and pull-down
ERS_G-his-F	AAAGAGCTCGGCATAGATTCAGACTACA	
ERS_G-his-R	AAACTCGAGGTTCAAACTGGCCTGCTGTC	
ARC-MBP-F	AAAGGATCCATGGCGGCGGCGGCGGCGGACTT	
ARC_G_MBP-F	AAAGGATCCACCAAATCTGTCTTCGATCC	
ARC-MBP-R	AAAAAGCTTGCCTTAATGCTCGCATTAGGAA	
ERSm1-F	TGTGTTGCCTTTGCCCCAGCGCCCAGTG	Site-directed mutagenesis
ERSm1-R	CACTGGGCGCTGGGGCAAAGGCAACACA	
ERSm2-F	TTACCTCGCTATAGGTGCTGCAAAGG	
ERSm2-R	CCTTTGCAGCACCTATAGCGAGGTAA	
ERSm3-F	ATACTCTTCTTGCCGCGCGGAAGCT	
ERSm3-R	AGCTTCCGCGCGGCAAGAAGAGTAT	
H4-ISH-AF	ATGTCTGGGAGAGGCAAGGG	in situ hybridisation
H4-ISH-AR	TAATACGACTCACTATAGGGCTAGCCGCCGAATCCGTAGA	
H4-ISH-SF	TAATACGACTCACTATAGGGATGTCTGGGAGAGGCAAGGG	
H4-ISH-SR	CTAGCCGCCGAATCCGTAGA	
MSP-ISH-AF	GCTCAAGATGACCCACGGCA	
MSP-ISH-AR	TAATACGACTCACTATAGGGACAGCACATTAAGCACTCAC	
MSP-ISH-SF	TAATACGACTCACTATAGGGGCTCAAGATGACCCACGGCA	
MSP-ISH-SR	ACAGCACATTAAGCACTCAC	
CDKB2-ISH-AF	CTGATGTACCAGCTCTGCAA	
CDKB2-ISH-AR	TAATACGACTCACTATAGGGGCAGAAGCTGCTGAACCTC	
CDKB2-ISH-SF	TAATACGACTCACTATAGGGCTGATGTACCAGCTCTGCAA	
CDKB2-ISH-SR	GCAGAAGCTGCTGAACCTC	
GST-F	CTTGAGCGAGTCCCTCGTC	RT-qPCR
GST-R	CGAAACAATCCTGCATGTTG	
HSP-F	AGATCACCGTGTGCTTCGAC	
HSP-R	TCCGCCTTGTACTTCTCCGC	
MAP3K-F	GAGCGTGCTCATCACCTCCA	
MAP3K-R	CAAGATCAACATGGATGGTC	
MAPKC-F	CAGACAACTGAGTGTGAGAG	
MAPKC-R	ACGCTATGCAATCTATGAGC	
MIL1-F	ATGATGTCGGTGTACGAGAG	
MIL1-R	CGAGCTGGTCGAGCTCGTAG	
PDC2-F	CGGTAGGTGCAACTCTGGGA	
PDC2-R	ATCTCCACCTCGATGGTGTA	
APX1-F	CGGTGCGCAGATGGGCTTGA	
APX1-R	CCTCCTTGTCACCACTCAGA	
APX2-F	TCAGGCTTCTCGACCCCATC	
APX2-R	TTCAGGAGGAGGCTCAGGCT	

APX3-F	CCAGTACTGCTGATCCTGCT	
APX3-R	AGATGAAGATGGCTGTGGCT	
APX4-F	GCATGGACTCAGGAGCCTCT	
APX4-R	TTGTGCGATTCAGCGTAGTC	
APX5-F	TCGAGCTAGACCAGAGCGTA	
APX5-R	AGAACTAGAAGATCCTCATC	
APX6-F	CAGTGCTACAGCCATTGAGG	
APX6-R	TTCCTTGTCACTCAGGCCCA	
GPX1-F	GAGCCAGTTGTATGAGAAGT	
GPX1-R	TGGATACTCGGCCTTGAAGC	
GPX2-F	GCTGACGGAGCTCTATCAGA	
GPX2-R	CATTTACACGCACCTTCTGA	
GPX3-F	GACAAGGTTGATGTCAACGG	
GPX3-R	CATAGCGATCCACCACGCGA	
GPX4-F	TGAAGAGATTGAGCAAACAG	
GPX4-R	TGATGCCATCACCAAGGAAT	
CAL-F	CGTCAGCTGCTTGATGGACG	
CAL-R	GGTGTCATCCAGCGGGTCGA	
GDH1-F	AGGCTCTTGCAATAATTAAC	
GDH1-G	GTTCAGTTCGAGTTCTCCAA	
GDH3-F	GATGCAATCGACCCACGCTC	
GDH3-R	GGAGTTCGCGAGGATGTCTG	
GS2-F	TACGCCGGGATCAACATCAG	
GS2-R	CAAGGGAGAGCACGACTCCG	
GS3-F	GAGTTCCAGATTGGCCCTGT	
GS3-R	CCTCATCGACTTGGTGCTGT	
GS4-F	ACGGAGCAGGCTGGTGTAGT	
GS4-R	ATGCAAGTCATGGCGAAGTG	
FBP1-F	ACATTAAGCCTACTCTCGGC	
FBP1-R	GCCTGTGAAAGACTGGCCTC	
FBP2-F	CCAGGGACCAGAAGAGCAAG	
FBP2-R	ATGTACAGCGGCACTCTCTG	
FBP4-F	CTGTACGAAGTGTTCCCAAT	
FBP4-R	GACTCTGAAGCGTACAATGC	
G6PDH3-F	CGCAACGAGAAGGTGAAAGT	
G6PDH3-R	CGAACGTCGGCGTGATGCTG	
G6PDH5-F	TATGCAGCAAGGTACTCCAA	
G6PDH5-R	CCATAAGGGTATAGCTCAGG	
ODH1-F	TGCAAGGTTGGAGCGTTTCC	
ODH1-R	CTGAAGTCCCGGTGAATCTG	
ODH2-F	ATGAAGTTGACATGACTAAT	
ODH2-R	GTATATGATGTCATCACCAT	
ODH3-F	GTGAACGTGACAACGCCTGC	
ODH3-R	GCTTATCGAAACCTGGATGG	

CS1-F	GGACATCTGGTTGGAAGTGC	
CS1-R	TGAGTTGATCAGTTGTAACG	
CS2-F	GCTAGCAGAGGAGGTCTTCT	
CS2-R	AGAACTCTGTAGGGAATCCC	
CS3-F	ATGCGACTATATGTGACAAT	
CS3-R	GATCCACAACAGCACTTCCT	
MDH2-F	GACAATTGTCCAGAAGCCTT	
MDH2-R	TCTGAGCTATAAATGTGTTA	
MDH9-F	ACCACCCTAGATGTTGTTAG	
MDH9-R	TCATCTGTGAAGGTGACAGA	
MDH11-F	TACGGAATTCATCTCTACCG	
MDH11-R	TCAGAGTACACACCCATGGA	
H4-F	ACAACATCCAGGGGATCACG	
H4-R	AGGAAGATCTTGAGCACGCC	
CDKB2;1-F	CCAGGAGTGAGCAAGCTACC	
CDKB2;1-R	TCTTCGCAGAGATCCGCTTC	
MSP1-F	CGTCTCCAGGGACATGTAAT	
MSP1-R	GCACATTAAGCACTCACATT	
MIL2-F	TACACGGTGGACGTGATGA	
MIL2-R	CTAGTAAGGCAAGAGATGTCACTAG	
AlaRS3-F	GTTGCTACCCAGATAGCTTC	
AlaRS3-R	CGTACAGAAGCGTATCGGTA	
ArgRS2-F	GAGCTGCTCGGAATTACACC	
ArgRS2-R	ATACAGCAACCAAGACTAGA	
AsnRS1-F	GCTGTCCTGTTAGCTCCTCA	
AsnRS1-R	CGTTCATATGCTGTAGTGGC	
AspRS3-F	CTCGTGGTCTTCAACATCCA	
AspRS3-R	TCTAATGGTGCAATGACATC	
CysRS1-F	CCTCATGGATGAACCTACTG	
CysRS1-R	CAATTCAGCAGATGTGCTTG	
GlnRS1-F	TCGCAGTGGACACAGACTCG	
GlnRS1-R	ACTGTGTAACCTCAGTGCAC	
GlyRS3-F	ACAGATGAGATCGGTGTTCC	
GlyRS3-R	ACCAATATGGCTGGGATACT	
LysRS2-F	CGGCTATGAAGCCTCAAGAT	
LysRS2-R	CATGGAAACCTCACATGCAC	
MetRS2-F	GCTATCAGCTGACCTGCACA	
MetRS2-R	CAAGTCACTCGAGTATATCC	
ProRS2-F	GAGCTGCCAAGACATTGTGT	
ProRS2-R	CGGCTTCGCCAGCAGGAATG	
ThrRS2-F	TCATGTAGCCATCGATGAGA	
ThrRS2-R	GCTCACTGTGGCTAGTTCAG	
TrpRS2-F	GCAGTTCTATCCGAACTGGT	
TrpRS2-R	CTAGAGATCACTTCGGACAT	

Sequences of primers used in Chapter 4.

HVERS1-F	TAATGGAACAGTTACTCAGC	RT-qPCR of *HvERS* genes
HVERS1-R	CGCTCCGCATGTTCGGGTCTC	
HVERS2-F	AGATAATGGAACAGTTACTC	
HVERS2-R	GACGTTCCGCATGTTTGGGTC	
HVERS3-F	TATGATAGTGGTGAGCTTGA	
HVERS3-R	TGTGCACGAGGATTACGGTA	
HvACTIN7-F	CGTGTTGGATTCTGGTGATG	
HvACTIN7-R	AGCCACATATGCGAGCTTCT	
U-F	CTCCGTTTTACCTGTGGAATCG	CRISPR-Cas9 construction
gR-R	CGGAGGAAAATTCCATCCAC	
Pps	AAGGTCTCGCTCGATGGAATCGGCAGCAAAGG	
Pgs	AAGGTCTCTACCGCTCCATCCACTCCAAGCTC	
gR- T1-F	TTTTGCTCCAGAGCCTAGCGTTTTAGAGCTAGAAAT	
U6- T1-R	GCTAGGCTCTGGAGCAAAACGGCAGCCAAGCCAGCA	
gR- T2-F	CTGAAAGTTTGATTAAGCAGTTTTAGAGCTAGAAAT	
U6- T2-R	TGCTTAATCAAACTTTCAGCGGCAGCCAAGCCAGCA	
KO-HvERS-F1	CCTTGTCCGCTGGTTCAACAGT	Sequencing
KO-HvERS-F2	CAAAAGTTGGGGATGTTTGTGT	
KO-HvERS-R	TGTTTGGGTCCTGCATGTCAAGT	

Appendix 4

Other Publications

1. Li, G*., **Yang, X*.** (co-first author), Zhang, X., Song, Y., Liang, W., and Zhang, D. 2018. Rice Morphology Determinant-mediated Actin Filament Organization Contributes to Pollen Tube Growth. *Plant Physiol*. (DOI: https://doi.org/10.1104/pp.17.01759).

2. Liu, Z., Lin, S., Shi, J., Yu, J., Zhu, L., **Yang, X.**, Zhang, D. and Liang, W., 2017. *Rice No Pollen 1* (*NP1*) is required for anther cuticle formation and pollen exine patterning. *Plant J.* (DOI: https://doi.org/10.1111/tpj.13561).

BIBLIOGRAPHY

(Chapter 1, 4 and 5)

Abiko, M., Akibayashi, K., Sakata, T., Kimura, M., Kihara, M., Itoh, K., Asamizu, E., Sato, S., Takahashi, H. and Higashitani, A., 2005. High-temperature induction of male sterility during barley (*Hordeum vulgare* L.) anther development is mediated by transcriptional inhibition. *Sexual Plant Reproduction*, 18(2), pp.91-100.

Albrecht, C., Russinova, E., Hecht, V., Baaijens, E. and de Vries, S., 2005. The *Arabidopsis thaliana* SOMATIC EMBRYOGENESIS RECEPTOR-LIKE KINASES1 and 2 control male sporogenesis. *The Plant Cell*, 17(12), pp.3337-3349.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T.G., Bertoni, M., Bordoli, L. and Schwede, T., 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic acids research*, 42(W1), pp.W252-W258.

Canales, C., Bhatt, A.M., Scott, R. and Dickinson, H., 2002. EXS, a putative LRR receptor kinase, regulates male germline cell number and tapetal identity and promotes seed development in *Arabidopsis*. *Current Biology*, 12(20), pp.1718-1727.

Carter Jr, C.W., 1993. Cognition, mechanism, and evolutionary relationships in aminoacyl-tRNA synthetases. *Annual review of biochemistry*, 62(1), pp.715-748.

Chaubal, R., Anderson, J.R., Trimnell, M.R., Fox, T.W., Albertsen, M.C. and Bedinger, P., 2003. The transformation of anthers in the *msca1* mutant of maize. *Planta*, 216(5), pp.778-788.

Colcombet, J., Boisson-Dernier, A., Ros-Palau, R., Vera, C.E. and Schroeder, J.I., 2005. *Arabidopsis* SOMATIC EMBRYOGENESIS RECEPTOR KINASES1 and 2 are essential for tapetum development and microspore maturation. *The Plant Cell*, 17(12), pp.3350-3361.

Coruzzi, G.M. and Zhou, L., 2001. Carbon and nitrogen sensing and signalling in plants: emerging 'matrix effects'. *Current opinion in plant biology*, 4(3), pp.247-253.

Cotelle, V., Meek, S.E., Provan, F., Milne, F.C., Morrice, N. and MacKintosh, C., 2000. 14-3-3s regulate global cleavage of their diverse binding partners in sugar-starved *Arabidopsis* cells. *The EMBO journal*, 19(12), pp.2869-2876.

Cusack, S., 1997. Aminoacyl-tRNA synthetases. *Current opinion in structural biology*, 7(6), pp.881-889.

Del Moral, L.G., Miralles, D.J. and Slafer, G., 2002. Initiation and appearance of vegetative and reproductive structures throughout barley development. *Barley science*: recent advances from molecular biology to agronomy of yield and quality'.(Eds GA Slafer, JL Molina-Cano, R Savin, JL Araus and I Romagosa), pp.243-267.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, 32(5), pp.1792-1797.

Fan, D., Liu, T., Li, C., Jiao, B., Li, S., Hou, Y. and Luo, K., 2015. Efficient CRISPR/Cas9-mediated targeted

mutagenesis in *Populus* in the first generation. *Scientific reports*, 5, p.12217.

Fei, Q., Yang, L., Liang, W., Zhang, D. and Meyers, B.C., 2016. Dynamic changes of small RNAs in rice spikelet development reveal specialized reproductive phasiRNA pathways. *Journal of experimental botany*, 67(21), pp.6037-6049.

Feng, X., Zilberman, D. and Dickinson, H., 2013. A conversation across generations: soma-germ cell crosstalk in plants. *Developmental cell*, 24(3), pp.215-225.

Fernandes, A.P. and Holmgren, A., 2004. Glutaredoxins: glutathione-dependent redox enzymes with functions far beyond a simple thioredoxin backup system. *Antioxidants and Redox Signalling*, 6(1), pp.63-74.

Fu, Z., Yu, J., Cheng, X., Zong, X., Xu, J., Chen, M., Li, Z., Zhang, D. and Liang, W., 2014. The rice basic helix-loop-helix transcription factor TDR INTERACTING PROTEIN2 is a central switch in early anther development. *The Plant Cell,* 26(4), pp.1512-1524.

Gómez, J.F. and Wilson, Z.A., 2012. Non-destructive staging of barley reproductive development for molecular analysis based upon external morphology. *Journal of experimental botany*, 63(11), pp.4085-4094.

Gómez, J.F. and Wilson, Z.A., 2014. A barley PHD finger transcription factor that confers male sterility by affecting tapetal development. *Plant biotechnology journal*, 12(6), pp.765-777.

Gubler, F., Raventos, D., Keys, M., Watts, R., Mundy, J. and Jacobsen, J.V., 1999. Target genes and regulatory domains of the GAMYB transcriptional activator in cereal aleurone. *The Plant Journal*, 17(1), pp.1-9.

Guo, M. and Schimmel, P., 2013. Essential nontranslational functions of tRNA synthetases. *Nature chemical biology,* 9(3), p.145.

Guo, M., Yang, X.L. and Schimmel, P., 2010. New functions of aminoacyl-tRNA synthetases beyond translation. *Nature reviews Molecular cell biology*, 11(9), p.668.

Harwood, W.A., 2014. A protocol for high-throughput Agrobacterium-mediated barley transformation. In *Cereal Genomics* (pp. 251-260). Humana Press, Totowa, NJ.

Hong, L., Tang, D., Shen, Y., Hu, Q., Wang, K., Li, M., Lu, T. and Cheng, Z., 2012a. MIL2 (MICROSPORELESS2) regulates early cell differentiation in the rice anther. *New Phytologist,* 196(2), pp.402-413.

Hong, L., Tang, D., Zhu, K., Wang, K., Li, M. and Cheng, Z., 2012b. Somatic and reproductive cell development in rice anther is regulated by a putative glutaredoxin. *The Plant Cell*, 24(2), pp.577-588.

Hord, C.L., Chen, C., DeYoung, B.J., Clark, S.E. and Ma, H., 2006. The BAM1/BAM2 receptor-like kinases are important regulators of *Arabidopsis* early anther development. *The Plant Cell*, 18(7), pp.1667-1680.

Ibba, M. and Söll, D., 2000. Aminoacyl-tRNA synthesis. *Annual review of biochemistry*, 69(1), pp.617-650.

International Barley Genome Sequencing Consortium, 2012. A physical, genetic and functional sequence assembly of the barley genome. *Nature*, 491(7426), p.711.

Ito, T., Wellmer, F., Yu, H., Das, P., Ito, N., Alves-Ferreira, M., Riechmann, J.L. and Meyerowitz, E.M., 2004. The homeotic protein AGAMOUS controls microsporogenesis by regulation of SPOROCYTELESS. *Nature*, 430(6997),

pp.356-360.

Jeong, E.J., Hwang, G.S., Kim, K.H., Kim, M.J., Kim, S. and Kim, K.S., 2000. Structural analysis of multifunctional peptide motifs in human bifunctional tRNA synthetase: identification of RNA-binding residues and functional implications for tandem repeats. *Biochemistry*, 39(51), pp.15775-15782.

Jia, G., Liu, X., Owen, H.A. and Zhao, D., 2008. Signalling of cell fate determination by the TPD1 small protein and EMS1 receptor kinase. *Proceedings of the National Academy of Sciences*, 105(6), pp.2220-2225.

Juliano, C. and Wessel, G., 2010. Versatile germline genes. *Science*, 329(5992), pp.640-641.

Kaminska, M., Deniziak, M., Kerjan, P., Barciszewski, J. and Mirande, M., 2000. A recurrent general RNA binding domain appended to plant methionyl-tRNA synthetase acts as a cis-acting cofactor for aminoacylation. *The EMBO journal*, 19(24), pp.6908-6917.

Kelliher, T. and Walbot, V., 2011. Emergence and patterning of the five cell types of the Zea mays anther locule. *Developmental biology*, 350(1), pp.32-49.

Kelliher, T. and Walbot, V., 2012. Hypoxia triggers meiotic fate acquisition in maize. *Science*, 337(6092), pp.345-348.

Kim, J.H., Han, J.M. and Kim, S., 2013. Protein–protein interactions and multi-component complexes of aminoacyl-tRNA synthetases. In *Aminoacyl-tRNA Synthetases in Biology and Medicine* (pp. 119-144). Springer Netherlands.

Laporte, D., Huot, J.L., Bader, G., Enkler, L., Senger, B. and Becker, H.D., 2014. Exploring the evolutionary diversity and assembly modes of multi-aminoacyl-tRNA synthetase complexes: Lessons from unicellular organisms. *FEBS letters*, 588(23), pp.4268-4278.

Lawrenson, T., Shorinola, O., Stacey, N., Li, C., Østergaard, L., Patron, N., Uauy, C. and Harwood, W., 2015. Induction of targeted, heritable mutations in barley and Brassica oleracea using RNA-guided Cas9 nuclease. *Genome biology*, 16(1), p.258.

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., Rouzé, P. and Rombauts, S., 2002. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic acids research*, 30(1), pp.325-327.

Li, H., Yuan, Z., Vizcay-Barrena, G., Yang, C., Liang, W., Zong, J., Wilson, Z.A. and Zhang, D., 2011. PERSISTENT TAPETAL CELL1 encodes a PHD-finger protein that is required for tapetal cell death and pollen development in rice. *Plant physiology*, 156(2), pp.615-630.

Liang, G., Zhang, H., Lou, D. and Yu, D., 2016. Selection of highly efficient sgRNAs for CRISPR/Cas9-based plant genome editing. *Scientific reports*, 6, p.21451.

Liu, X., Huang, J., Parameswaran, S., Ito, T., Seubert, B., Auer, M., Rymaszewski, A., Jia, G., Owen, H.A. and Zhao, D., 2009. The SPOROCYTELESS/NOZZLE gene is involved in controlling stamen identity in *Arabidopsis*. *Plant physiology*, 151(3), pp.1401-1411.

Liu, Z., Bao, W., Liang, W., Yin, J. and Zhang, D., 2010. Identification of gamyb-4 and Analysis of the Regulatory Role

of GAMYB in Rice Anther Development. *Journal of integrative plant biology*, 52(7), pp.670-678.

Ma, H., 2005. Molecular genetic analyses of microsporogenesis and microgametogenesis in flowering plants. *Annual review of plant biology,* 56, pp.393-434.

Ma, J., Duncan, D., Morrow, D.J., Fernandes, J. and Walbot, V., 2007. Transcriptome profiling of maize anthers using genetic ablation to analyse pre-meiotic and tapetal cell types. *The Plant Journal,* 50(4), pp.637-648.

Ma, L., Yang, Z. and Zhang, S., 2013. DEX1, a plasma membrane-localised protein, functions in microspore development by affecting CalS5 expression in *Arabidopsis thaliana*. *Chinese Science Bulletin*, 58(23), pp.2855-2861.

Ma, X., Zhang, Q., Zhu, Q., Liu, W., Chen, Y., Qiu, R., Wang, B., Yang, Z., Li, H., Lin, Y. and Xie, Y., 2015. A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. *Molecular plant*, 8(8), pp.1274-1284.

Mascher, M., Gundlach, H., Himmelbach, A., Beier, S., Twardziok, S.O., Wicker, T., Radchuk, V., Dockter, C., Hedley, P.E., Russell, J. and Bayer, M., 2017. A chromosome conformation capture ordered sequence of the barley genome. *Nature*, 544(7651), p.427.

Matzke, A.J. and Matzke, M.A., 1998. Position effects and epigenetic silencing of plant transgenes. *Current opinion in plant biology*, 1(2), pp.142-148.

Meyer, Y., Buchanan, B.B., Vignols, F. and Reichheld, J.P., 2009. Thioredoxins and glutaredoxins: unifying elements in redox biology. *Annual review of genetics*, 43, pp.335-367.

Murmu, J., Bush, M.J., DeLong, C., Li, S., Xu, M., Khan, M., Malcolmson, C., Fobert, P.R., Zachgo, S. and Hepworth, S.R., 2010. *Arabidopsis* basic leucine-zipper transcription factors TGA9 and TGA10 interact with floral glutaredoxins ROXY1 and ROXY2 and are redundantly required for anther development. *Plant Physiology*, 154(3), pp.1492-1504.

Murray, F., Kalla, R., Jacobsen, J. and Gubler, F., 2003. A role for *HvGAMYB* in anther development. *The Plant Journal*, 33(3), pp.481-491.

Nonomura, K.I., Miyoshi, K., Eiguchi, M., Suzuki, T., Miyao, A., Hirochika, H. and Kurata, N., 2003. The *MSP1* gene is necessary to restrict the number of cells entering into male and female sporogenesis and to initiate anther wall formation in rice. *The Plant Cell*, 15(8), pp.1728-1739.

Nureki, O., O'Donoghue, P., Watanabe, N., Ohmori, A., Oshikane, H., Araiso, Y., Sheppard, K., Söll, D. and Ishitani, R., 2010. Structure of an archaeal non-discriminating glutamyl-tRNA synthetase: a missing link in the evolution of Gln-$tRNA^{Gln}$ formation. *Nucleic acids research*, 38(20), pp.7286-7297.

Pan, C., Ye, L., Qin, L., Liu, X., He, Y., Wang, J., Chen, L. and Lu, G., 2016. CRISPR/Cas9-mediated efficient and heritable targeted mutagenesis in tomato plants in the first and later generations. *Scientific reports*, 6, p.24765.

Rould, M.A., Perona, J.J. and Steitz, T.A., 1991. Structural basis of anticodon loop recognition by glutaminyl-tRNA synthetase. *Nature*, 352(6332), pp.213-218.

Sakata, T., Takahashi, H., Nishiyama, I. and Higashitani, A., 2000. Effects of high temperature on the development of pollen mother cells and microspores in barley *Hordeum vulgare* L. *Journal of plant research*, 113(4), pp.395-402.

Schiefthaler, U., Balasubramanian, S., Sieber, P., Chevalier, D., Wisman, E. and Schneitz, K., 1999. Molecular analysis of *NOZZLE*, a gene involved in pattern formation and early sporogenesis during sex organ development in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences*, 96(20), pp.11664-11669.

Schön, A. and Söll, D., 1988. tRNA specificity of a mischarging aminoacyl-tRNA synthetase: glutamyl-tRNA synthetase from barley chloroplasts. *FEBS Letters*, 228(2), pp.241-244.

Schulze, J.O., Masoumi, A., Nickel, D., Jahn, M., Jahn, D., Schubert, W.D. and Heinz, D.W., 2006. Crystal structure of a non-discriminating glutamyl-tRNA synthetase. *Journal of molecular biology*, 361(5), pp.888-897.

Sekine, S.I., Nureki, O., Shimada, A., Vassylyev, D.G. and Yokoyama, S., 2001. Structural basis for anticodon recognition by discriminating glutamyl-tRNA synthetase. *Nature Structural and Molecular Biology*, 8(3), p.203.

Shi, J., Cui, M., Yang, L., Kim, Y.J. and Zhang, D., 2015. Genetic and biochemical mechanisms of pollen wall development. *Trends in plant science*, 20(11), pp.741-753.

Shiu, S.H., Karlowski, W.M., Pan, R., Tzeng, Y.H., Mayer, K.F. and Li, W.H., 2004. Comparative analysis of the receptor-like kinase family in *Arabidopsis* and rice. *The Plant Cell*, 16(5), pp.1220-1234.

Simms, D., Cizdziel, P.E. and Chomczynski, P., 1993. TRIzol: A new reagent for optimal single-step isolation of RNA. *Focus*, 15(4), pp.532-535.

Simos, G., Segref, A., Fasiolo, F., Hellmuth, K., Shevchenko, A., Mann, M. and Hurt, E.C., 1996. The yeast protein Arc1p binds to tRNA and functions as a cofactor for the methionyl-and glutamyl-tRNA synthetases. *The EMBO journal*, 15(19), pp.5437-5448.

Torii, K.U., 2004. Leucine-rich repeat receptor kinases in plants: structure, function, and signal transduction pathways. *International review of cytology*, 234, pp.1-46.

Vaucheret, H., Béclin, C., Elmayan, T., Feuerbach, F., Godon, C., Morel, J.B., Mourrain, P., Palauqui, J.C. and Vernhettes, S., 1998. Transgene-induced gene silencing in plants. *The Plant Journal*, 16(6).

Vernoud, V., Laigle, G., Rozier, F., Meeley, R.B., Perez, P. and Rogowsky, P.M., 2009. The HD-ZIP IV transcription factor OCL4 is necessary for trichome patterning and anther development in maize. *The Plant Journal*, 59(6), pp.883-894.

Von Wettstein, D., Gough, S. and Kannangara, C.G., 1995. Chlorophyll biosynthesis. *The Plant Cell*, 7(7), p.1039.

Vondenhoff, G.H. and Van Aerschot, A., 2011. Aminoacyl-tRNA synthetase inhibitors as potential antibiotics. *European journal of medicinal chemistry*, 46(11), pp.5227-5236.

Waddington, S.R., Cartwright, P.M. and Wall, P.C., 1983. A quantitative scale of spike initial and pistil development in barley and wheat. *Annals of Botany*, 51(1), pp.119-130.

Walbot, V. and Egger, R.L., 2016. Pre-meiotic anther development: cell fate specification and differentiation. *Annual review of plant biology,* 67, pp.365-395.

Wang, C.J.R., Nan, G.L., Kelliher, T., Timofejeva, L., Vernoud, V., Golubovskaya, I.N., Harper, L., Egger, R., Walbot, V. and Cande, W.Z., 2012. Maize *multiple archesporial cells 1* (*mac1*), an ortholog of rice *TDL1A*, modulates cell proliferation and identity in early anther development. *Development*, 139(14), pp.2594-2603.

Wang, Z.P., Xing, H.L., Dong, L., Zhang, H.Y., Han, C.Y., Wang, X.C. and Chen, Q.J., 2015. Egg cell-specific promoter-controlled CRISPR/Cas9 efficiently generates homozygous mutants for multiple target genes in *Arabidopsis* in a single generation. *Genome biology*, 16(1), p.144.

Wilson, Z.A., Morroll, S.M., Dawson, J., Swarup, R. and Tighe, P.J., 2001. The *Arabidopsis MALE STERILITY1* (*MS1*) gene is a transcriptional regulator of male gametogenesis, with homology to the PHD-finger family of transcription factors. *The Plant Journal*, 28(1), pp.27-39.

Woese, C.R., Olsen, G.J., Ibba, M. and Söll, D., 2000. Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiology and Molecular Biology Reviews*, 64(1), pp.202-236.

Wong, G.K.S., Wang, J., Tao, L., Tan, J., Zhang, J., Passey, D.A. and Yu, J., 2002. Compositional gradients in Gramineae genes. *Genome research*, 12(6), pp.851-856.

Xing, S. and Zachgo, S., 2008. ROXY1 and ROXY2, two *Arabidopsis* glutaredoxin genes, are required for anther development. *The Plant Journal*, 53(5), pp.790-801.

Yamaguchi, T., Lee, D.Y., Miyao, A., Hirochika, H., An, G. and Hirano, H.Y., 2006. Functional diversification of the two C-class MADS box genes *OSMADS3* and *OSMADS58* in *Oryza sativa*. *The Plant Cell*, 18(1), pp.15-28.

Yang, L., Qian, X., Chen, M., Fei, Q., Meyers, B.C., Liang, W. and Zhang, D., 2016. Regulatory role of a receptor-like kinase in specifying anther cell identity. *Plant physiology*, 171(3), pp.2085-2100.

Yang, S.L., Xie, L.F., Mao, H.Z., San Puah, C., Yang, W.C., Jiang, L., Sundaresan, V. and Ye, D., 2003. TAPETUM DETERMINANT1 is required for cell specialization in the *Arabidopsis* anther. *The Plant Cell*, 15(12), pp.2792-2804.

Yang, W.C., Ye, D., Xu, J. and Sundaresan, V., 1999. The *SPOROCYTELESS* gene of *Arabidopsis* is required for initiation of sporogenesis and encodes a novel nuclear protein. *Genes & Development*, 13(16), pp.2108-2117.

Ye, Q., Zhu, W., Li, L., Zhang, S., Yin, Y., Ma, H. and Wang, X., 2010. Brassinosteroids control male fertility by regulating the expression of key genes involved in *Arabidopsis* anther and pollen development. *Proceedings of the National Academy of Sciences*, *107*(13), pp.6100-6105.

Zhang, D. and Liang, W., 2016. Improving food security: using male fertility for hybrid seed breeding. *Science*, pp.45-48.

Zhang, D. and Yang, L., 2014. Specification of tapetum and microsporocyte cells within the anther. *Current opinion in plant biology*, 17, pp.49-55.

Zhang, D., Luo, X. and Zhu, L., 2011. Cytological analysis and genetic control of rice anther development. *Journal of Genetics and Genomics*, 38(9), pp.379-390.

Zhang, J., Zhao, J., Dahan, P., Lu, V., Zhang, C., Li, H. and Teitell, M.A., 2018. Metabolism in Pluripotent Stem Cells

and Early Mammalian Development. *Cell Metabolism*, 27(2), pp.332-338.

Zhao, D.Z., Wang, G.F., Speal, B. and Ma, H., 2002. The *EXCESS MICROSPOROCYTES1* gene encodes a putative leucine-rich repeat receptor protein kinase that controls somatic and reproductive cell fates in the *Arabidopsis* anther. *Genes & Development*, 16(15), pp.2021-2031.

Zhao, X., De Palma, J., Oane, R., Gamuyao, R., Luo, M., Chaudhury, A., Hervé, P., Xue, Q. and Bennett, J., 2008. OsTDL1A binds to the LRR domain of rice receptor kinase MSP1, and is required to limit sporocyte numbers. *The Plant Journal*, 54(3), pp.375-387.