Roles of Post-transcriptional Modifications in *Solanum lycopersicum*

Cameron Grundy BSc (Hons)

November 2021

Thesis submitted to the University of Nottingham for Degree of Doctor of Philosophy

Supervised by: Professor Rupert Fray

Professor Graham Seymour

Acknowledgements

Firstly, I would like to express my appreciation for Professor Rupert Fray for his mentorship and support throughout my PhD. His passion for science is reflected in the supervision he provides to his students. He gave me the opportunity to study and without his ideas and problem solving this thesis would not have been possible.

It is also important for me to thank Professor Graham Seymour, Dr Peter Wittich and Charles Baxter, without their ongoing support throughout this project, I would have never had this opportunity. I hope that you can utilise some of this work in the future.

I would like to thank the members of A49 for their assistance throughout my projects. A49 was always a busy lab, full of people to bounce ideas off and to stave off the boredom of endless screening. Special thanks need to go to Dr Zsuzsa Bodi and Dr Eleanor Bellows. Zsuzsa taught me many techniques which are used throughout this work and Eleanor was the first person to ask if I needed any help, no matter how busy she was with her own projects. I would also like to thank Dr Frank Baron, who produced the random oligos used for the *in vitro* selection protocol.

I would like to thank Sophie Cowling for pushing me to take the PhD, then listening to me complain about it for 4+ years. I would also like to express my absolute love and gratitude to my family, especially my parents, Heather and Simon for their unquestioning belief in me.

Finally, I would like to thank my partner, Giorgia. She was (and still is) full of unwavering love and guidance both in and out of the lab. We shared all aspects of the experience and she helped me through every single difficult moment. I can honestly say I would not have finished this thesis without her.

Declaration

Unless otherwise acknowledged, the work presented within this thesis is my own. No part has been submitted for another degree at the University of Nottingham or any other institution of learning.

Table of Contents

Acknowledgements i

Declaration ii

Table of Contents iii

List of Figures xi

List of Tables xiv

Abbreviations xv

Abstract 1

Chapter 1 - Introduction 2

1.1 Ribonucleic acid species 2

1.2 mRNA biogenesis 3

1.2.1 Initiation 4

1.2.2 Elongation 4

1.2.3. Termination 5

1.3 Co-transcriptional processing 6

1.3.1 Capping 7

1.3.1 Splicing and the spliceosome 9

1.3.1.1 Splicing in plants 10

1.3.2 Polyadenylation 11

1.3.2.1 Alternative polyadenylation 12

1.4 Post transcriptional modifications 13

1.5 N6-metyladenosine 15

1.5.1 Mapping m^6A 16

1.5.2 Writing m^6A 17

1.5.2.1 METTL3 21

1.5.2.2 METTL14 21

1.5.2.3 WTAP 22

1.5.2.4 Virilizer/KIAA1429 23

1.5.2.5 RBM15/15B 23
1.5.2.6 ZC3H13-WTAP 24
1.5.2.7 HAKAI 25
1.5.2.8 METTL16 25
1.5.3 Erasing m^{6}A 26
1.5.3.1 FTO 26
1.5.3.2 ALKBH5 27
1.5.3.3 Demethylation in plants 28
1.6 Reading m^{6}A 28
1.6.1 YTH Domain proteins 29
1.6.2 Molecular and physiological roles of YTH domain proteins in animals 31
1.6.2.1 Alternate splicing 31
1.6.2.2 RNA stability 32
1.6.2.3 Translational initiation and control 32
1.6.2.4 Sex determination 33
1.6.2.5 Gametogenesis 33
1.6.2.6 Neuronal development 34
1.6.2.7 Abiotic Stress responses 34
1.6.2.8 Cancer 35
1.6.2.9 Immune response and viral infection 36
1.6.3 Molecular and physiological roles of YTH domain proteins in plants 37
1.6.3.1 YTHDF-type proteins 37
1.6.3.2 YTHDC-type proteins 38
Chapter 2 - Materials and methods 41
2.1 Plant materials, media and growth conditions 41
2.1.1 Plant materials 41
2.1.2 Plant media 41

2.1.2.1 MSR3 solid growth medium 41
2.1.2.2 MSR3ZA callusing medium 41
2.1.2.3 MSR3BA callusing medium.....41
2.1.2.4 MSR3 rooting medium.....42
2.1.2.5 Liquid MS.....42
2.1.3 Growth conditions.....42
2.1.4 Tomato seed collection.....42
2.2 Bacterial growth, media and strains 43
2.2.1 Bacterial strains 43
2.2.2 Media 44
2.2.2.1 Luria-Bertani medium.....44
2.2.2.2 Hannahan's Broth (SOB and SOC medium).....44
2.2.3 Culture growth 44
2.2.4 Bacterial storage.....45
2.3 General chemicals and buffers.....45
2.3.1 DH5alpha competent cell production buffers 45
2.3.2 Nickle column protein purification buffers.....45
2.3.3 Poly A buffers.....46
2.3.4 Protein pull down 1 buffers.....46
2.3.5 Protein pull down 2 buffers.....46
2.3.6 RNA binding affinity buffers.....47
2.3.7 Western blot buffers 47
2.3.8 TLC Running buffers.....47
2.3.9 Stock Solutions 47
2.4 DNA manipulation and amplification.....50
2.4.1 Plasmids.....50
2.4.2 Isolation of plasmid DNA 55
2.3.3 Oligonucleotide Primers 56

2.3.4 Isolation of plant genomic DNA 62
2.3.5 Polymerase chain reaction (PCR) 63
2.3.6 Restriction analysis of DNA 63
2.3.7 A-tailing PCR products 64
2.3.8 Agarose gel electrophoresis 64
2.3.9 Purification of PCR and DNA products 64
2.3.9.1 Purification of DNA and PCR products via kit extraction ..64
2.3.9.2 Purification of PCR products via gel extraction 64
2.3.10 DNA quantification 64
2.3.11 Golden gate ligation assembly 65
2.3.12 HiFi and Gibson DNA assembly 65
2.3.13 Sequencing of DNA 65
2.4 Transformation of bacterial strains 66
2.4.1 Transformation of *Escherichia coli* chemically competent cells. 66
2.4.2 Electroporation of *Agrobacterium tumefaciens* electro-competent cells 66
2.4.3 Preparation of *E.coli* chemically competent cells 66
2.3.4 Preparation of *A.tumefaciens* electrocompetent cells 67
2.5 Plasmid construction 68
2.5.1 Construction of CRISPR-Cas9 disruption constructs 68
2.5.1.1 Selection of CRISPR disruption targets 68
2.5.1.2 Assembly of level 1 CRISPR-Cas9 disruption constructs. ... 68
2.5.1.3 Assembly of level 2 CRISPR-Cas9 disruption constructs 68
2.5.2 Construction of constitutive expression constructs 69
2.5.3 Construction of ECT promoter GUS expression constructs 69
2.5.4 Construction of ECT1-3 protein MBP-fusion constructs. 69
2.5.5 Construction of ECT4 protein MBP-fusion constructs. 70
2.6 RNA manipulation 70

2.6.1 Total RNA extraction....70
2.6.2 Poly-A purification....71
2.6.2.1 Bead regeneration....72
2.6.3 Reverse transcription....72
2.7 Plant transformation via tissue culture....72
2.7.1 Excision and precallusing....72
2.7.2 Agrobacterial growth and tissue incubation....73
2.7.3 Selection and rooting....73
2.7.4 Screening transgenic lines....74
2.8 Protein purification....74
2.8.1 Pilot experiments....74
2.8.2 Denaturing polyacrylamide SDS-PAGE gel electrophoresis....75
2.8.3 Protein production and purification....75
2.8.4 Further protein purification and buffer exchange....76
2.9 Protein-RNA binding affinity assay....76
2.9.1 Production of RNA oligoes....77
2.9.2 Binding to amylose beads....77
2.9.3 Kinasing of RNA oligoes....77
2.9.4 Application of RNA to protein....78
2.9.5 DNase treatment....78
2.9.6 Proteinase K treatment....78
2.9.7 Pull-down RNA library synthesis....79
2.9.8 Small RNA library sequencing....79
2.9.9 RNA sequencing analysis....79
2.10 Western Blot....80
2.10.1 Transfer....80
2.10.2 Blocking and antibody incubation....80
2.10.3 Incubation with substrate....81

2.10.4 Detection 81

Chapter 3 - Genome-wide bioinformatic identification of *Solanum lycopersicum* YTH proteins 83

3.1 Introduction 83

3.2 Results 85

3.2.1 Phylogenetic comparison of YT521-B domain containing proteins 85

3.2.2 Genome-wide identification of plant YT521-B domain containing proteins 88

3.2.3 High conservation of YT521-B domain 90

3.2.4 Analysis and comparison of the DC-type YTH protein family within Plant species 93

3.2.6 YT521-B domain containing proteins in *Solanum lycopersicum* 104

3.2.7 DF-type YT521-B domain containing proteins in *Solanum lycopersicum* 106

3.2.7.1 Intrinsic disorder of YT521-B domain containing proteins in *Solanum lycopersicum* 109

3.3 Discussion 112

Chapter 4 - Characterisation of YT521-B proteins in *Solanum lycopersicum* 116

4.1 Introduction 116

4.2 Results 117

4.2.1 YTHDF gRNA and vector design 117

4.2.2 Generation of single YTHDF gene KOs 120

4.2.3 Characterization of ythdf loss-of-function mutants 124

4.2.3.1 Physiological characterisation of ythdf null mutants in nutrient solution 124

4.3.3.2 Germination rate of *ythdf* loss-of-function mutants 129

4.2.3.3 Effect of salt stress on *ythdf* loss-of-function mutants 130

4.2.3.4 Loss of *ythdf1* or *ythdf2* does not affect ambient m^{6}A levels 135

4.3 Discussion 136

4.3.1 Development of CRISPR/Cas9 mediated gene disruption in YTHDF proteins 137

4.3.2 Analysis of *ythdf* loss-of-function mutant characterisation.....137

4.3.3 Analysis of effect of salinity stress of *ythdf* loss-of-function mutants.....140

4.3.4 Measuring m^{6}A levels in *ythdf* loss-of-function mutants.....141

4.3.5 Future Work 142

Chapter 5 - Characterisation of YT521-B protein double loss-of-function mutants in Solanum lycopersicum 145

5.1 Introduction 145

5.2 Results 146

5.2.1 *ythdf1/ythdf4* homozygous plants are embryo lethal.....147

5.2.1.1 Treatment of *ythdf1/ythdf4* seeds with gibberellic acid....148

5.2.2 Generation of *ythdf1/ythdf2* and *ythdf2/ythdf4* double mutants. 150

5.2.4 Characterisation of ythdf double mutants 157

5.2.4.1 Physiological characterisation of *ythdf1/ythdf3* double homozygous mutant lines.....157

5.2.4.2 Physiological characterisation of *ythdf2/ythdf3* double homozygous mutant lines.....163

5.2.4.3 Physiological characterisation of *ythdf3/ythdf4* double homozygous mutant lines.....166

5.3 Discussion 169

5.3.1 Future Work 173

Chapter 6 - *In-vitro* selection with specifically modified RNA library for the identification of YTHDF binding preferences in *Solanum lycopersicum*..176

6.1 Introduction 176

6.2 Results 178
6.2.1 *in Vitro* selection strategy 178
6.2.2 Purification of tomato YTHDF proteins 179
6.2.3 Synthesis of Random RNA libraries 181
6.2.4 Utilisation of *in vitro* Selection Platform 181
6.2.5 Sequencing analysis of YTHDF-protein RNA selection 184
6.2.5.1 Characterisation of Input libraries 184
6.2.5.2 Analysis of YTHDF1 *in vitro* selection libraries 186
6.2.5.3 Analysis of YTHDF2 *in vitro* selection libraries 189
6.2.5.4 Analysis of YTHDF3 *in vitro* selection libraries 192
6.2.5.5 Analysis of YTHDF4 *in vitro* selection libraries 195
6.2.5.6 Relative abundance of sequences enriched upon *in vitro* selections 197
6.3 Discussion 202
7.1 General Discussion 205
7.2 Future Work 216
7.3 Concluding Remarks 219
Appendix 220
References 238

List of Figures

Figure 1.1: Basic Dogma of Biology..... 3
Figure 1.2: Structure of cap 0, cap 1 and m6Am first nucleotides. 9
Figure 1.3: Transition from Adenoside to N^6-Methyladenosine......16
Figure 1.4: Readers, writers and erasers of m^6A modifications. 18
Figure 1.5: Hydrophobic aromatic cage in the YTH domain.. 30

Figure 3.1: Schematic tree of similarity of YTHDF and YTHDC proteins. 87
Figure 3.2: Multiple sequence alignment of YTHDF and YTHDC proteins from *Solanum lycopersicum, Arabidopsis thaliana* and *Homo sapiens*.....92
Figure 3.3: Schematic tree of similarity of YTHDC proteins. Comparisons between model Planta species. 94
Figure 3.4: Schematic tree of similarity of YTHDF proteins.....102
Figure 3.5: Intron-Exon distribution of all YTH genes found in *Solanum lycopersium*.....105
Figure 3.6: Multiple sequence alignment of YTHDF proteins from *Solanum lycopersicum*. 108
Figure 3.7: Intrinsic disorder region predictions of YTHDF1, YTHDF2, YTHDF3 and YTHDF4. 111

Figure 4.1: sgRNA DNA target sites for CRISPR/CAs9 gene disruption systems within four YTHDF genes in Solanum lycopersium. L. 119
Figure 4.2: Mutation detection and consensus sequences of CRISPR/CAs9 mediated mutations. 122
Figure 4.3: Wild-type and *ythdf* single mutant lines at 4 weeks.....127
Figure 4.4: Characterisation of *ythdf1-4* single mutant KO lines. Histogram representation of key plant growth measurements. 128
Figure 4.5: Germination rate observed of WT and *ythdf* loss-of-function lines. 130
Figure 4.6: Wild-type and *ythdf* single mutant lines at 4 weeks, comparison of high salt treatment and standard conditions.. 133

Figure 4.7: Characterisation of *ythdf1-4* single mutant KO lines in high salt and standard conditions. 134
Figure 4.8: Two dimensional TLC detection of relative global abundance of m^6A in WT, *ythdf1* and *ythdf2.* 136

Figure 5.1: Percentage of seeds germinated in *ythdf1, ythdf4* and *ythdf1/4* lines. 149
Figure 5.2: Mutation detection and consensus sequences of *ythd1/ythdf2* double mutant generation. 152
Figure 5.3: Mutation detection and consensus sequences of *ythdf2/ythdf4* double mutant generation. 153
Figure 5.4: Wild-type and *ythdf* single mutant lines at 4 weeks.. 155
Figure 5.5: Characterisation of *ythdf1-4* single mutant KO lines. Histogram representation of key plant growth measurements. 156
Figure 5.6: Wild-type and *ythdf1, ythdf3* single mutant lines and *ythdf1/ythdf3* double homozygous mutant at 4 weeks. 161
Figure 5.7: Characterisation of *ythdf1, ythdf3* single mutant lines and *ythdf1/ythdf3* double homozygous mutant. Histogram representation of key plant growth measurements. 162
Figure 5.8: Wild-type and *ythdf2, ythdf3* single mutant lines and *ythdf2/ythdf3* double homozygous mutant at 4 weeks. 164
Figure 5.9: Characterisation of *ythdf2, ythdf3* single mutant lines and *ythdf2/ythdf3* double homozygous mutant. Histogram representation of key plant growth measurements. 165
Figure 5.10: Wild-type and *ythdf3, ythdf4* single mutant lines and *ythdf3/ythdf4* double homozygous mutant at 4 weeks. 167
Figure 5.11: Characterisation of *ythdf3, ythdf4* single mutant lines and *ythdf3/ythdf4* double homozygous mutant. Histogram representation of key plant growth measurements. 168

Figure 6.1: *In vitro* selection strategy utilised to identify binding preferences of *Solanum lycopersicum* YTHDF RNA binding proteins. 179
Figure 6.2: Purification of *Solanum lycoperisum* YTHDF proteins. 180

Figure 6.3: Electropherograms from high sensitivity DNA assay of YTHDF *in-vitro* RNA selection libraries after NEBNext Small RNA library production.....183
Figure 6.4: Characterisation of random sequencing libraries present in input. 185
Figure 6.5: Characterisation of random sequencing libraries present in YTHDF1-1 and YTHDF1-2 *in vitro* selection libraries. 188
Figure 6.6: Characterisation of random sequencing libraries present in YTHDF2-1 and YTHDF2-2 *in vitro* selection libraries. 191
Figure 6.7: Characterisation of random sequencing libraries present in YTHDF3-1 and YTHDF3-2 *in vitro* selection libraries:. 194
Figure 6.8: Characterisation of random sequencing libraries present in YTHDF4-1 and YTHDF4-2 *in vitro* selection libraries. 196

Appendix 6.1: Purification of *Solanum lycoperisum* YTHDF proteins...236
Appendix 6.2: Western blot from the maltose elution troubleshooting assay 237

List of Tables

Table 1.1: m^{6}A methylation complex components and eraser proteins in *Arabidopsis,* Mammals and *Drosophila* 43

Table 2.1: List of strains used in this study..... 43
Table 2.2: List of plasmids used in this study. 50
Table 2.3: List of oligonucleotide primers used in this study.. 56
Table 2.4: List of growth conditions used for protein purification pilot studies. 74

Table 3.1: YTHDF and YTHDC proteins in plant species. 88
Table 3.2: YTH proteins of *Solanum lycopersicum*..... 104

Table 4.1: Description of all generated single mutant lines of YTHDF proteins in *Solanum lycopersicum* 123

Table 5.1: Genotyping results during *ythdf1/ythdf4* double mutant generation. 147
Table 5.2: Average gemination rates of WT, *ythdf1, ythdf4, . ythdf1--/ythdf4*+- and *ythdf1*-+/*ythdf4*— lines. 149

Table 6.1: Relative abundance of A-centred 9-mers enriched via *in vitro* selection strategy for YTHDF1, YTHDF2, YTHDF3, and YTHDF4. 200
Table 6.2: Relative abundance of m^{6}A-centred 9-mers enriched via *in vitro* selection strategy for YTHDF1, YTHDF2, YTHDF3, and YTHDF4. 201

Appendix 3.1: Accession numbers for proteins used in comparison trees 220

Abbreviations

µL	Microliter
µm	Micrometer
A	Adenosine
ALKBH1	Alpha-ketoglutarate-dependent dioxygenase B (AlkB) Homolog 1
ALKBH2	Alpha-ketoglutarate-dependent dioxygenase B (AlkB) Homolog 2
ALKBH3	Alpha-ketoglutarate-dependent dioxygenase B (AlkB) Homolog 3
ALKBH4	Alpha-ketoglutarate-dependent dioxygenase B (AlkB) Homolog 4
ALKBH5	Alpha-ketoglutarate-dependent dioxygenase B (AlkB) Homolog 5
ALKBH6	Alpha-ketoglutarate-dependent dioxygenase B (AlkB) Homolog 6
ALKBH7	Alpha-ketoglutarate-dependent dioxygenase B (AlkB) Homolog 7
ALKBH8	Alpha-ketoglutarate-dependent dioxygenase B (AlkB) Homolog 8
Am	2′O-ribose methyladenosine
ANOVA	Analysis of variance
ATP	Adenosine triphosphate
C	Cytosine
cap0	m^7GpppN structure
cap1	m^7GpppNm structure
cap2	m^7GpppNmNm structure
cDNA	complementary DNA
CPSF	Cleavage and polyadenylation specificity factor
CstF	Cleavage stimulation factor
DNA	Deoxyribonucleic acid
ECT	Evolutionarily conserved C-terminal domain containing protein
FTO	Fat mass and obesity-associated protein
G	Guanine
GA	Gibberellic Acid
HNRNPA2B1	Heterogeneous Nuclear Ribonucleoprotein A2/B1
HNRNPC	Heterogeneous Nuclear Ribonucleoprotein C
IDR	Intrinsically disordered region
IPTG	Isopropyl-ß-D-thiogalactoside
kDa	KiloDaltons
KIAA1429	Vir Like M^6A Methyltransferase Associated
LB	Lysogeny broth
lncRNA	Long no coding RNA

M	Molar
m^6A	N^6-methyladenosine
m^6Am	N^6-2′O-dimethyladenosine
me-RIP-seq	m^6A methylation RNA immunoprecipitation chip sequencing
METTL14	Methyltransferase Like 14
METTL3	Methyltransferase Like 3
mi-CLIP	m^6A individual-nucleotide-resolution cross-linking and immunoprecipitation
min	Minute
mL	Millilitre
mm	Millimetre
mM	Millimolar
miRNA	Micro RNA
mRNA	Messenger RNA
MTA	N6-adenosine-methyltransferase MT-A70-like
MTB	N6-adenosine-methyltransferase non-catalytic subunit MTB
PAP	Polyadenylate polymerase
PARP1	Poly(ADP-Ribose) Polymerase 1
PCR	Polymerase chain reaction
pmol	Picomoles
PNK	Poly nucleotide kinase
RBM15	RNA Binding Motif Protein 15
RIP-seq	RNA immunoprecipitation chip sequencing
RNA	Ribonucleic acid
RNA pol II	RNA polymerase II
RNA pol II CTD	RNA polymerase II C terminal binding domain
RppH	RNA 5' Pyrophosphohydrolase
rRNA	Ribosomal RNA
SAM	Shoot apical meristem
SDS-PAGE	Sodium dodecyl sulphate-Polyacrylamide gel electrophoresis
SCARLET	Site-specific cleavage and radioactive-labelling followed by ligation-assisted extraction and thin-layer chromatography
SEM	Standard error of the mean
snRNA	Small nuclear RNA
SRSF10	Serine And Arginine Rich Splicing Factor 10
SRSF3	Serine And Arginine Rich Splicing Factor 3
ssDNA	Single stranded DNA
ssRNA	Single stranded RNA
T	Tyrosine
TLC	Thin layer chromatography
tRNA	Transfer RNA
TSS	Transcription start site
U	Uridine

Um	2′O-ribose methyluridine
WTAP	Wilms Tumour 1-Associating Protein
XRN1	5′-3′ Exoribonuclease 1
XRN4	5′-3′ Exoribonuclease 4
YTH	YT521-B homology domain
YTHDC1	YTH Domain Containing 1
YTHDC2	YTH Domain Containing 2
YTHDF1	YTH N6-Methyladenosine RNA Binding Protein 1
YTHDF2	YTH N6-Methyladenosine RNA Binding Protein 2
YTHDF3	YTH N6-Methyladenosine RNA Binding Protein 3
ZC3H13	Zinc Finger CCCH-Type Containing 13

Abstract

Methylation at the N^6 position of internal adenosines (m^6A) is known as the most prevalent and widespread modification in mRNA. In plants, m^6A has roles in post-transcriptional gene control, influencing growth, development and reproduction. The function of these modifications is often mediated by RNA binding proteins containing a YT521-B homology (YTH) domain. These domains contain a hydrophobic pocket that can directly accommodate m^6A residues. Despite studies demonstrating that m^6A is crucial for plant development and multiple studies on mammalian YTH proteins, very little is known about YTH proteins in plants.

Here, we have identified that *Solanum lycopersicum* have some of the lowest number of YTH proteins in *Plantae*. We show that plants lacking YTHDF1, YTHDF2, YTHDF3 or YTHDF4 from *Solanum lycopersicum*, exhibit vegetative growth defects, including phenotypes which overlap with those identified in low m^6A plant lines and respond differently to salt stress. In addition, we have produced several YTHDF double mutants, which demonstrate that each YTHDF protein have both redundant and specific functions *in vivo*. Finally, we demonstrate a novel *in vitro* randomised RNA-binding protein selection strategy, which has the potential to be used to identify key consensus sequences to further our understanding of complex processes. This strategy allowed us to identify that just like their phenotypes, each YTHDF protein in tomato show affinity for both unique and overlapping sequences *in vitro*. This work provides a platform for future advances in our understanding of these complex proteins.

Chapter 1 - Introduction

1.1 Ribonucleic acid species

The central dogma of molecular biology describes a two-step process. Deoxyribose nucleic acid (DNA) transcribed into ribonucleic acid (RNA) which is in turn translated into protein. These processes are the means by which information within biological systems translate into function (F. H. Crick, 1958) (Fig.1.). Early models followed these basic principles but RNA has since been shown to be more than just a passive intermediate. Recent studies have highlighted the catalytic role of RNA from RNA interference (RNAi) to riboswitches, and demonstrated enzymatic activities once thought to be exclusive to proteins (Krupp *et al.*, 1996; Nilsen, 2007). The model has therefore been revised and even inspired 'the RNA world' hypothesis which proposed the first replicating system with RNA as the medium for storing information, as well as "enzyme functions" for self-replications (Gilbert Walter, 1986).

RNA is a polymeric molecule comprised of pyrimidine and purine nucleotides. These nucleotides form phosphodiester linkages between the 3' carbon of one nucleotide and the 5' carbon of the next. Each nucleotide (nt) is composed of i) a nitrogenous base ii) a triphosphate iii) a pentose sugar (in this case ribose) (Figure 1.1). The nitrogenous bases which make up RNA are the pyrimidines; cytosine (C) and uracil (U) and the purines; adenine (A) and guanine (G) (Figure 1.1). RNA differs from DNA in a number of ways. 1) RNA is commonly found as a single stranded (ss) molecule in most of its biological roles. 2) In DNA A forms Watson-crick base pairing with thymine (T), whereas in RNA T is replaced by an unmethylated form, U. 3) In RNA the 2' hydroxyl group (2' OH) in the pentose sugar is replaced with a 2' hydrogen (2' H). The linear sequence of RNA, which can also be referred to as primary structure, predominately follows Watson-Crick (canonical) base

pairing via hydrogen bonds. G-C with three hydrogen bonds or A-U with two hydrogen bonds. Non-Watson-Crick (non-canonical) interactions are also possible (Xin and Olson, 2008). Both types of intramolecular base contribute to RNA folding and secondary structure and can thus influence the production of RNA-protein complexes to convey the function of various RNAs (Miao and Westhof, 2017). These secondary and tertiary structures are crucial to the stability and function of all RNAs (Bevilacqua *et al.*, 2016; Vandivier *et al.*, 2016).

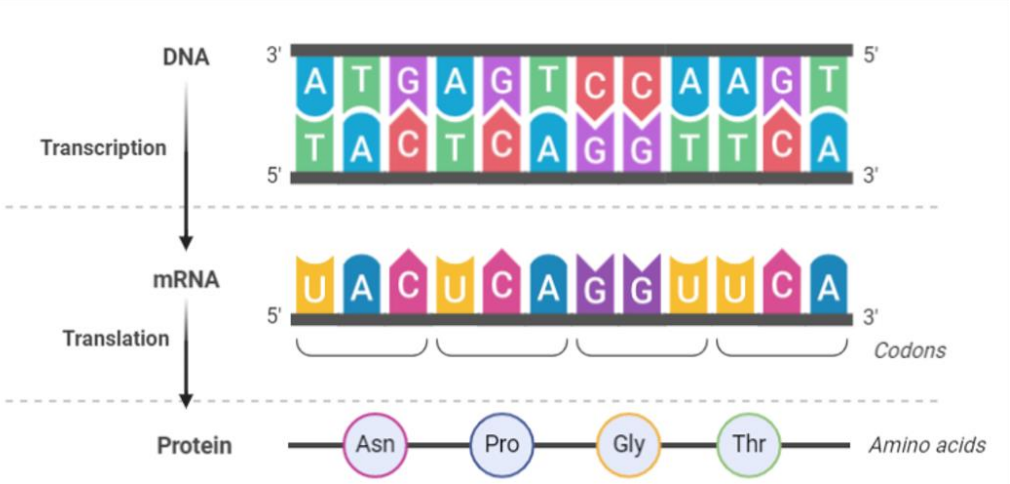


Figure 1.1: Basic Dogma of Biology. DNA into RNA into Protein.

1.2 mRNA biogenesis

Transcription of Messenger RNA (mRNA) from DNA is the first major step of translating any genetic information into functional protein. RNA polymerases (RNAP) carry out RNA synthesis. Within Eukaryotes, a large protein complex known as RNA polymerase II (Pol II) transcribes all protein-coding genes within the nucleus. During this process a 5' guanine cap, 3' polyadenylated tail and the removal of introns result in stable mature mRNA. Mature mRNA can then translocate from the nucleus to the cytoplasmic ribosomes for translation.

1.2.1 Initiation

To initiate transcription Pol II must be recruited to the promoter of the target gene to form a pre-initiation complex (PIC). There are six general transcription factors (GTFs) which are essential for the recognition of these promoters. GTFs; TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH along with Pol II form the PIC (Roeder, 1996). In most cases TFIID subunits locate the TATA box within promoters and the promoter structure is unwound and the template strand is loaded onto the active site of Pol II (Buratowski *et al.*, 1989). The PIC then recognises the transcription start site (TSS) which leads to the catalysis of the first phosphodiester bond. This forms the mechanism for basal transcription. GTFs have a number of interacting partners known as general cofactors, and each of these proteins can interact with different promoter elements. Each promoter can have a range of promoter elements both upstream and downstream of the TSS and can often be recognised by more than one GTF. The regulation between elements, GTFs and cofactors is highly complex and not yet fully understood (Deng and Roberts, 2005; Gershenzon and Ioshikhes, 2005; Louder *et al.*, 2016). Functional repression of basal transcriptional activation or catalysis of assembly greatly influence gene expression. The range of GTFs, cofactors and promoter element lend flexibility to the assembly of the PIC, increasing levels of regulation and assisting Pol II in the initiation, in a site and gene specific manner (Thomas and Chiang, 2006).

1.2.2 Elongation

Transcript elongation involves the sequential addition of NTPs via phosphodiester bonding. During elongation Pol II functions as part of a ternary elongation complex (also known as transcription elongation complex- TEC) with the target DNA and the nascent transcript. Similar to the PIC, the TEC is complex and is tightly regulated. The machinery adapts to signals from regulatory proteins, as well as decoding its DNA template (Daube and von Hippel, 1992; Nudler, 1999; Gnatt *et al.*, 2001). One of the most influential mechanisms in elongation regulation is sequence-

dependant pausing. Disruption occurs when normal catalytic activity of RNAP is disrupted by either changed structural orientations of the TEC (due to nucleic acid disruption) or specific interaction with external protein factors (Landick, 2006; Adelman and Lis, 2012). Areas of nucleic acids within the TEC that result in pausing are known as pause signals, and a given signal will only affect a certain number of RNAP molecules. In addition, the length of the pause also varies between signals (Nierman, Wiggs and Neff, 1981; Neuman *et al.*, 2003). Elongation can also be modified by two classes of elongation factors i) Groups of transcription factors which directly target the TEC to regulate its function ii) Chromatin modifying and remodelling proteins, and histone chaperones which alter chromatin structure to adapt elongation (Conaway and Conaway, 2019). Pausing is also an essential mechanism involved in termination. Pausing has been shown to allow maximum binding of termination and 3' end processing factors to Pol II (Glover-Cutter *et al.*, 2008). To what extent these factors affect elongation varies within different conditions, but elongation remains an important step in transcript biogenesis and gene regulation.

1.2.3. Termination

Termination of transcription elongation, as well as post-transcriptional regulation, are essential to sort functional transcripts from the numerous transcripts which appear to have no role. It also allows independent downstream regulation post elongation (Wyers *et al.*, 2005; Davis and Ares, 2006; Jensen, Jacquier and Libri, 2013). Pol II termination is linked with pre-mRNA 3' end processing and so has a large influence on final product lengths. This in turn influences the localisation, stability and eventual function of the transcripts (Edwalds-Gilbert, Prescott and Falck-Pedersen, 1993; West and Proudfoot, 2009; Mapendano *et al.*, 2010). The TEC is inherently very stable, this allows it to transcribe efficiently, however, this means that disruption of the TEC heteroduplex is necessary for termination. In Eukaryotes, Pol II termination can be initiated by a number of different pathways. Two of the most studied pathways, Polyadenylation (Poly(A)) dependent and Sen1-dependent pathways, provide a significant portion of

termination regulation. The Sen1-dependent pathway is involved with non-coding RNA transcription. The 3' end processing of protein-coding transcripts can be broken down into two steps. Transcription of a poly(A) site within the DNA template, pausing of Pol II and endoribonucleolytic cleavage is followed by polyadenylation of the upstream cleavage product. The remaining downstream transcript is degraded. This is catalysed by RNA binding proteins which recognise the cis-acting elements within a poly(A) site (Gu, Gupta and Schoenberg, 1999). The strength of the poly(A) site can be used to predict the efficiency of termination (Edwalds-Gilbert, Prescott and Falck-Pedersen, 1993). The carboxy-terminal domain (CTD) of the largest subunit of Pol II and post-translational modifications associated with it, is a strong platform for recruitment of termination factors. The CTD of Pol II is made up of a series of heptad repeats (Y1-S2-P3-T4-S5-P6-S7) (Corden, 1990). Phosphorylation of the CTD changes throughout transcription, altering binding affinity of its binding partners, regulating its function (Aviv, 2003). The current model of Poly(A) dependent termination involves cleavage and polyadenylation specificity factor (CPSF or CPF in yeast) recognising the transcribed AAUAAA sequence, then re-localising as part of a complex to the CTD on Pol II, resulting in CPSF-mediated release of Pol II (Proudfoot and Brownlee, 1976; Zhao, Hyman and Moore, 1999). This mechanism for release of Pol II also requires a 5'-3' exoribonuclease protein (XRN); XRN2 in mammals, XRN2, 3 in plants or Rat1 in yeast in a processing complex. After CPSF mediated cleavage of the nascent transcript, the XRN duplex 'nibbles' its way along the RNA tether remaining on Pol II 5' to 3'. It has been proposed that contact with Pol II at this point creates the force necessary for termination (Kim *et al.*, 2004; West, Gromak and Proudfoot, 2004).

1.3 Co-transcriptional processing

As previously mentioned, during transcription, pre-mRNAs are processed by the addition of a 5' guanine cap structure, a 3' poly(A) tail and the removal of introns. Processing allows, among other consequences, nuclear export, structural maintenance and the production of the correct coding sequence

for translation (Ramanathan, Robb and Chan, 2016). The separation of transcription and translation in the nuclear and cytoplasmic compartments respectively, allows the processing of pre-mRNA before translation. This also ensures that only mature mRNA ready for translation reaches the machinery for protein production.

1.3.1 Capping

Capping is the first co-transcriptional modification during Pol II transcription. All eukaryotic mRNAs contain a cap structure at its 5' end. This is an N^7-methylated guanosine (m^7G cap-also known as cap 0) (Figure 1.2). This cap has many functions including cap-dependent translation initiation, protective cap from 5'-3' cleavage by exonucleases and roles as a platform for recruitment of protein factors involved in pre-mRNA processing, polyadenylation and export from the nucleus (Ramanathan, Robb and Chan, 2016). During transcription, phosphorylation of serine 2 and 5 in the CTD heptad repeat regulates the recruitment of effector proteins. Phosphorylation of serine 5 occurs shortly after the onset of transcription, resulting in the recruitment of 3 enzymes: RNA triphosphatase (TPase), RNA guanylyltransferase (GTase) and guanine-N7 methyltransferase (guanine-N7 MTase), which are conserved across eukaryotes. RNA TPase removes a phosphate from the 5'triphospahte. GTase transfersing a guanosine monophosphate (GMP) from guanosine triphosphate (GTP), the newly formed 5'diphosphate. Finally, the guanine-N7 MTase then adds a methyl group to the N in the 7^{th} position of the amine in the guanine cap, forming the cap 0 structure (Ghosh and Lima, 2010).

The first two nucleotides of the mRNA transcript can be methylated on the ribose sugar, irrespective of the nucleotide forming 2′-O-ribose methyladenosine, 2′-O-ribose methylcytosine, 2′-O-ribose methylguanine, 2′-O-ribose methyluridine (Am, Cm, Gm, Um), this is known as cap1 on the first nucleotide and cap2 on the second nucleotide. If the cap1 nucleotide is adenosine, the nitrogenous base can be further modified to N^6,2′-O-dimethyladenosine (m^6Am) (Wei, Gershowitz and Moss, 1975). Studies have

revealed that 2'O methylation of the second nucleotide (+1), known as cap 1, is necessary for recognition of non-self, foreign RNA integral to immune responses (Daffis *et al.*, 2010). The function of m^6Am is largely unknown. In 2016, it was shown that fat mass and obesity-associated protein (FTO) thought by some to be an m^6A demethylase could demethylate m^6Am (Mauer *et al.*, 2017). Since then, the protein PCIF1 has been found to be a m^6Am specific methylase (Akichika *et al.*, 2019). Little work has focused on the cap1 modification and currently there is no sequencing method available to detect which transcripts are cap1 modified.

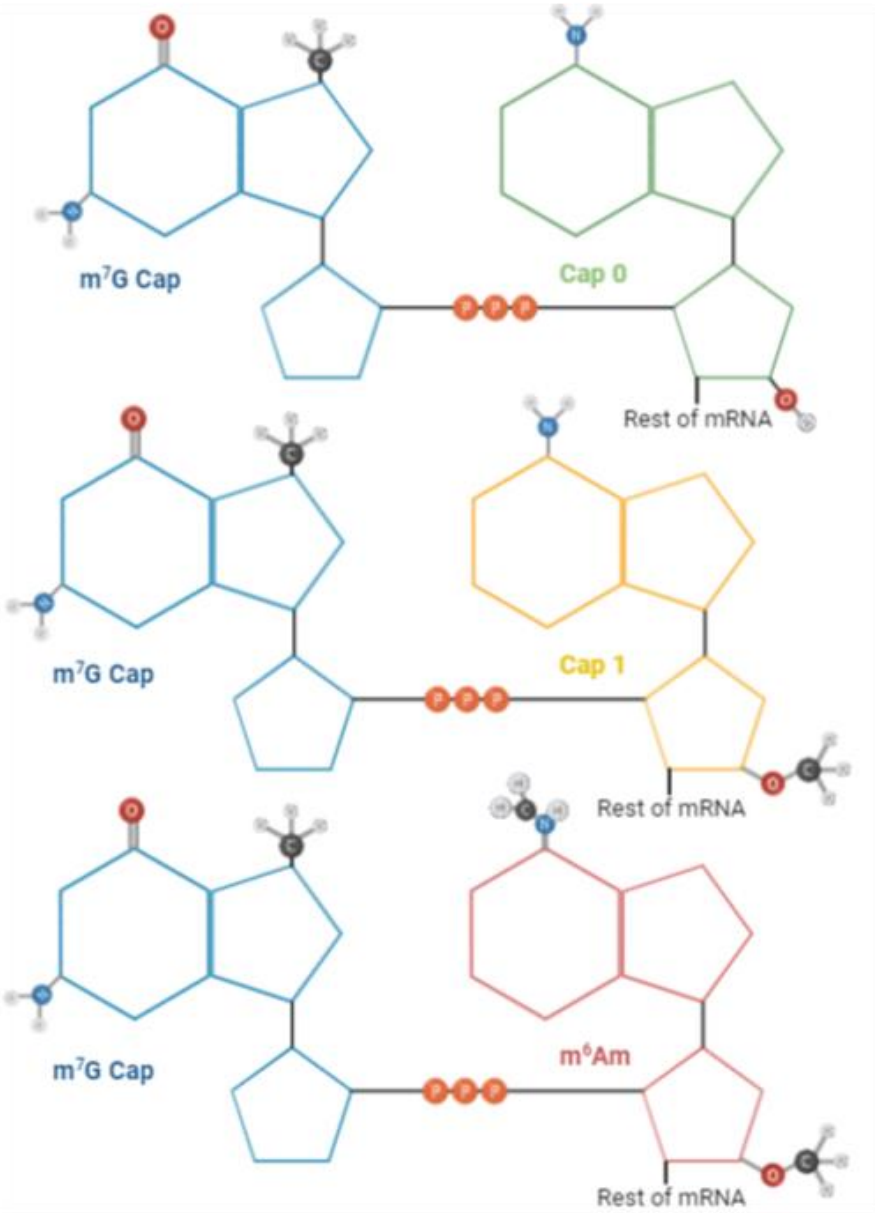


Figure 1.2: Structure of cap 0, cap 1 and m6Am first nucleotides. The guanine cap (light blue) is attached to the first nucleotide through a 5′-5′ linkage. Cap0 has no modifications on the first nucleotide, in this diagram the first nucleotide is adenosine (Green); Cap 1 has a methylation on the ribose sugar of the first nucleotide, in this diagram the first nucleotide is Am (Yellow); and m6Am has two methylations, one on the ribose sugar and the other on the nitrogenous base and can only occur on adenosine (Red).

1.3.1 Splicing and the spliceosome

Protein-coding sequences within eukaryotic genes often contain regions of non-coding introns. During co-transcriptional processing these introns need to be removed from pre-mRNA, and the remaining coding regions, known as introns, must be spliced together to from mature mRNA. This is known as pre-mRNA splicing. In humans for example, there are around 20000 protein-coding genes and each of these contain 8 introns, on average (Hong, Scofield and Lynch, 2006). In these genes, up to 95% have alternate splicing (Wang *et al.*, 2008), meaning the 20000 genes translate into the vast complexity of most eukaryotic organisms. Although, some would argue that not all these splice forms produce functional protein.

Splicing reactions are catalysed by a large complex, termed the spliceosome, comprised of small nuclear ribonucleoproteins (snRNPs) U1, U2, U4/U6 and U5, snRNAs and ~150 associated proteins (Wahl, Will and Lührmann, 2009). Seven homologous Sm proteins, from a ring around an Sm site at the 3' end of U1, U2, U4 and U5 (Bringmann and Luhrmann, 1986). The U6 snRNA, is instead threaded through a ring of seven LSm proteins (Seraphin, 1995; Achsel *et al.*, 1999). Each snRNA binds to a specific set of proteins to form the snRNPs. Finally, the complex is completed by a host of non-snRNP proteins, complexes, splicing factors and eight ATP-dependent helicases (Wilkinson, Charenton and Nagai, 2019).

The spliceosome must recognise splice sites prior to the reaction. The current model describes recognition of the 5' splice site (5'SS) and the branch point sequence (BP) by U1 and U2 snRNPs, to form the pre-spliceosome. The

pre-spliceosome then associates with a preassembled U4/U6/U5 snRNP complex, completing the spliceosome. The U6 snRNP folds precisely to from the active site of spliceosome. The DEAD box helicase Prp28 releases the 5'SS fromU1 snRNP and transfers it to a ACAGAGA recognition site in the newly folded U6 snRNP (Staley and Guthrie, 1999; Kastner *et al.*, 2019). An RNA helicase known as Brr2, cleaves U4 from U6, which allows the U6 bound to the ACAGAGA box to associate with U2, creating the active site. This conformational change allows the BP to dock into the active site, resulting in a branching reaction, cleaving the 5' exon and an intron intermediate (lariat-intron). The BP leaves the active site, allowing the 3'SS to move into the active site for exon ligation. Once the exons are ligated together, forming the mRNA, the RNA is released from the spliceosome (Galej *et al.*, 2016; Wilkinson, Charenton and Nagai, 2019). Release is mediated by Prp-22. Post release the lariat intron is released and targeted for degradation, the snRNAs and associated proteins are then recycled for subsequent splicing reactions (Fourmann *et al.*, 2013).

1.3.1.1 Splicing in plants

The spliceosome machinery in plants is not well understood due to the lack of a suitable in-vitro model. Plants have evolved numerous mechanisms and developmental processes to control and respond to variable and extreme conditions, that by their nature, they cannot avoid. Although splicing and alternate splicing is present, and highly conserved between metazoans and plants, there is evidence that plant specific regulation of alternate splicing occurs (Irimia and Roy, 2014; WANG *et al.*, 2015). Intron-retention is the most predominant method of alternate splicing in plants, whereas exon-skipping is the major type in mammals (Sammeth, Foissac and Guigó, 2008). Intron retention commonly results in non-sense mRNAs, often resulting in nonsense mediated decay (NMD) of these transcripts. This finding fits with many studies into plant stress responses and programming of circadian rhythm, where AS coupled to NMD was shown as an essential mechanism for regulating gene expression, in different developmental stage

and environmental stress (Kalyna *et al.*, 2012; Drechsel *et al.*, 2013; Filichkin *et al.*, 2015).

1.3.2 Polyadenylation

The addition of a poly(A) tail is tightly linked to transcript cleavage and transcription termination. Poly(A) tails, as well as the 5' cap, have roles in the initiation of translation. Protein mediated interactions with the 5' cap and poly(A) tail, assists in the linking of the 40S ribosomal subunit into the coding transcript. The poly(A) tail is essential for a mRNAs ability to be exported from the nucleus. In addition, poly(A) has been shown to have a strong role in the stability of transcripts, after translocation out of the nucleus. Deadenylation, removal of the poly(A) tail, is the primary rate limiting step and regulator of mRNA decay in the cytoplasm. The polyadenylation site (PAS), together with the stop codon upstream, define the 3' UTR of any given transcript, selection of this site is an integral mechanism in regulation of gene expression

The processing of the 3' end in mammals, is carried out by a conserved set of proteins, known as the 3'-end processing machinery. The 3'-end processing complex, co-transcriptionally detects the PAS, and upon recognition, mediates addition of the poly(A) tail (Zhao, Hyman and Moore, 1999). In mammals the PAS is a highly conserved hexanucleotide AAUAAA found ~ 10-30 nts upstream of the eventual cleavage site. Downstream of this cleavage site is a U-rich element (known as a downstream element (DSE)) which also determines the position for poly(A) addition, is located approximately 30 nts downstream of the poly(A) site. The position for the poly(A) site is determined by the distance between the PAS and the DSE, normally at an A (Sheets, Ogg and Wickens, 1990).

CPSF, catalyses the cleavage of the 3' end of nascent transcripts. Recognition of the PAS is carried out by CPSF. Interactions of cleavage stimulation factor (CstF) with the DSE, and cleavage factor 1 (CF1) with a UGUAA-like sequence, to further increase the specificity of recognition. The binding of CstF and CF1 to these regions, can recruit CPSF even if the USE

is missing (Venkataraman, Brown and Gilmartin, 2005; Millevoi *et al.*, 2006)

When the RNA is cleaved, polyadenylation starts, catalysed by a poly(A) polymerase (PAP). In animals, there are two forms of PAP, resulting from different splicing patterns of the same gene, PAP I and PAP II. PAP has specific biochemistry for utilisation of ATP and is recruited by CPSF. These are the minimum factors necessary for the addition of poly(A), however another factor is necessary for rapid elongation and tail length regulation, poly(A)-binding protein II (PAB II). (Bienroth, Keller and Wahle, 1993). Many protein coding genes contain multiple PAS, allowing a gene to code for more than one mRNA. This further increases the potential for multiple levels of gene regulation. Poly(A) factors, extracellular stimuli, RNPs and post-transcriptional modifications have all been shown to effect alternative polyadenylation (Tian and Manley, 2016).

Plants encode four isoforms of the poly(A) polymerase. Three located in the nucleus (PAPS1, PAPS2 and PAPS4) and one is cytoplasmic PAPS3. Differential mutant phenotypes have demonstrated that these four genes are functionally specialised (Kappel *et al.*, 2015). 3' end formation is less well understood in plants. There are three classes of *cis*-elements; near-upstream element (NUE), far-upstream element (FUE) and a cleavage element (CE - PAS adjacent to a U-rich element). While the FUE and U-rich elements in plants are similar to those in mammals, the NUE is an A-rich element between 10 and 40 nts upstream from the plant PAS (Gilmartin, 2005; Shen *et al.*, 2008). The protein factors and poly(A) signals involved in the poly(A) complex are also not well characterised. Homologues of key proteins have been studied but their functions within 3' end formation and position in the complex is largely unknown (Reviewed in (Hunt, Xing and Li, 2012)).

1.3.2.1 Alternative polyadenylation

A single gene can contain multiple PASs. Usage of different PASs can lead to sequentially and functionally distinct isoforms, also known as alternate polyadenylation (APA). This occurs frequently and in a large proportion of

genes (Gautheret *et al.*, 1998; Tian *et al.*, 2005). Most APA sites are present in the 3' UTR and can result in significantly different isoform lengths, with large effects on post-transcriptional modification, mRNA stability and mRNA nuclear export and localisation (Reviewed in (Tian and Manley, 2016). APA is largely tissue specific and have specific patterns in different cells and under different abiotic conditions, as such APA might be tightly linked to responses to cellular and extracellular changes (Lianoglou *et al.*, 2013). A large number of APA-regulatory factors have been characterised, ranging from global regulators to gene specific interactions. Characterising the roles of these factors in different mechanisms will greatly enhance our understanding of the importance of APA and could have significant implications.

1.4 Post transcriptional modifications

Chemical modifications of DNA and proteins have long been established as important functional regulators of expression, development and reactionary responses. More recently however, chemical modifications on RNA have been given the same level of interest and influence. This new layer of gene regulation, termed the epitranscriptome, is involved in many processes and modifications are present on all forms of RNA and across all kingdoms.

The presence of RNA modifications was first observed by two pioneering groups (Desrosiers, Friderici and Rottman, 1974; Perry and Kelley, 1974). Since then, developments in high throughput sequencing and detection methods have led to many new revelations about the epitranscriptome. Previously characterisation and discovery of modifications were limited to highly abundant RNAs: ribosomal RNAs (rRNA), transfer RNAs (tRNAs) and small nuclear RNAs (snRNA). These new methods have allowed the study of modifications found on even lowly expressed mRNA (Karijolich, Kantartzis and Yu, 2010; Yacoubi, Bailly and de Crécy-Lagard, 2012). These methods have discovered 172 types of post-transcriptional modifications (http://modomics.genesilico.pl/, accessed 06/04/20) (Machnicka *et al.*, 2013; Boccaletto *et al.*, 2018). Modifications introduce a significant increase in

functional diversity of nucleotides. Like amino acids, the modifications can change the characteristics of nucleotides including polarity, charge or conformational changes which effect base pairing. This can result in changes in RNA structure, interactions with RNA binding proteins (RBPs), leading to widespread influential changes in gene expression, splicing, maturation and developmental mechanisms (Fustin *et al.*, 2013; Haussmann *et al.*, 2016)

Early digests of tRNA from calf liver samples, showed mononucleotides which didn't match the four canonical bases found in RNA (Cohn, W.E., Volkin, 1951). The modifications were characterised over subsequent years. Today, 93 have been identified in tRNA, with a large majority of these conserved across species, while a minority are species specific.

In rRNAs, the 80S ribosomal RNA subunit has been shown to be heavily modified. It has been suggested that more than 14 modifications can be found on one subunit, stabilising the structure and increasing protein production (Taoka *et al.*, 2018). Many modifications on tRNA facilitate tertiary structure essential for its function. M^1A deposits at position 58 in yeast are essential, conserved across all domains. Removal of this modification is lethal, resulting in tRNA depletion (Anderson *et al.*, 1998). Pseudouridine (ψ) found in position 55 (ψ55) is universal across all life. ψ55 is located in the T-arm of tRNAs, with this residue the T-arm interacts with the D-arm, facilitating proper tertiary structure (Robertus *et al.*, 1974). Additional modifications occur after tRNA translocation out of the nucleus, hypomodified tRNA in the cytoplasm is quickly targeted for degradation via the 'rapid tRNA decay' pathway (Chernyakov *et al.*, 2008).

Our understanding of the biological functions of mRNA modifications is rapidly evolving. Until recently, high throughput methods for the identification or sequencing of mRNA modifications were lacking. Therefore, few modifications have been intensively studied. N^1-methyladenosine (m^1A) initially known to regulate structure and stability of tRNA and rRNAs (Dunn, 1961), is now known to be found in eukaryotic mRNAs with a distinct pattern of occurrence, although some scholars speculate that this could be an artifact. m^1A is predominantly found near translation start sites and the first SS (Dominissini *et al.*, 2016). This, in combination with the positive

charge (Zhou *et al.*, 2016), suggest a potentially strong role of m^1A (likely in translation regulation) which has yet to be uncovered. 5-Methylcytosine found initial interest in DNA methylation, now studies are involving m^5C in mRNA as well. Levels of m5C are much lower in eukaryote mRNAs than other modifications, making study difficult (Dubin and Taylor, 1975). Today, m^5C has a relatively well characterised writer complex, but the function of this modification remains elusive (Khoddami and Cairns, 2013). Pseudouridine (ψ) is the most abundant modification in RNA in Eukaryota, and in mainly found in rRNA and tRNA. Ψ is formed by isomerisation of uridine is common in tRNA and rRNAs (Robertus *et al.*, 1974; Bass, 2013). Single base resolution mapping of ψ has been carried out (Carlile *et al.*, 2014), and it has a strong ability to alter base paring, resulting in large charges in RNA structure and mRNA coding (Fernández *et al.*, 2013). It has even been reported that conversion of U to ψ in nonsense codons can pause translation. It is therefore possible that ψ, is an important regulatory element, in RNA structure and in the potential novel nonsense suppression, or to expand the genetic code (Karijolich and Yu, 2011). There are a host of pseudouridine synthase enzymes shown to be responsible for the isomerization modification in tRNA, so it is likely that these enzymes are also responsible for the modification in mRNA as well (Karijolich, Kantartzis and Yu, 2010). Finally, 2'-O-methylation (2'OMe) is a common RNA modification of all four ribonucleotides and can be found on all eukaryotic RNAs (Boccaletto *et al.*, 2018). Although sequencing and quantification methods exist, the function of 2'OMe is unclear (Kellner, Burhenne and Helm, 2010).

1.5 N6-metyladenosine

N^6-metyladenosine (m^6A) is the addition of a methyl group ($-CH_3$) to the 6^{th} position of an adenosine nucleotide in any internal position along an RNA transcript (Zhong *et al.*, 2008) (Figure 1.3). It is the most prevalent internal modification of mRNA, typically found at a ratio of 1.8 nts per 1000 bases (Dubin and Taylor, 1975). m^6A has been described as a 'dynamic' modification, with a complex system involving writer, reader and eraser

proteins characterised in many organisms including mammals, viruses, bacteria, fungi and plants (Zhong et al., 2008; Liu et al., 2015; Patil, Pickering and Jaffrey, 2017; Roundtree et al., 2017). Antibody immunoprecipitation pulldowns have enabled specific m^{6}A mapping, detecting peaks of localised m^{6}A marks at ~100 nt resolution (Dominissini *et al.*, 2012; Meyer *et al.*, 2012). This method has since been improved to allow single nucleotide mapping (Linder *et al.*, 2015a) greatly enhancing advances in the functional roles of m^{6}A mediated post-transcriptional regulation.

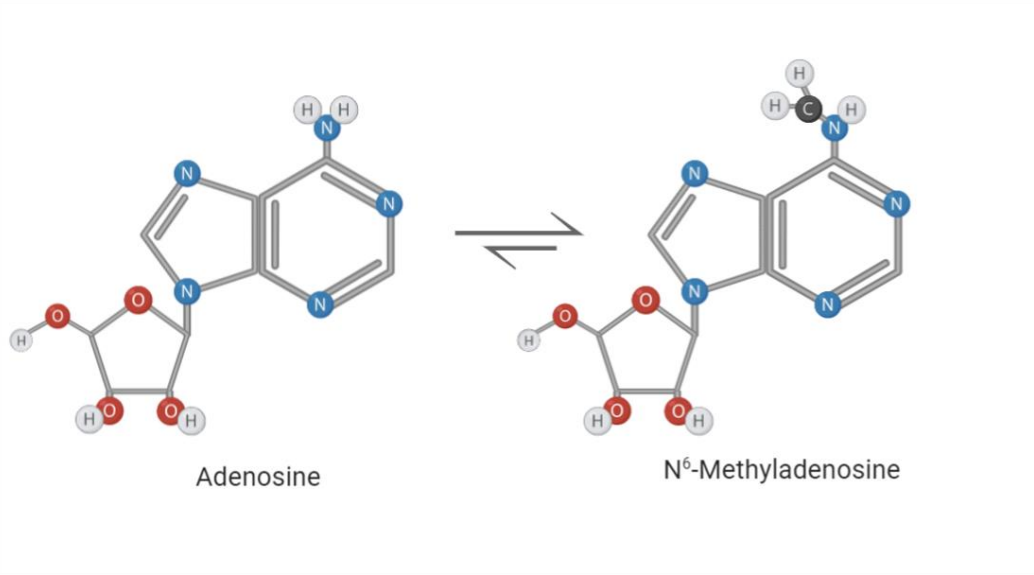


Figure 1.3: Transition from Adenoside to N^6-Methyladenosine. Key molecules carbon (Black), Nitrogen (Blue), Hydrogen (Grey) and Oxygen (Red).

1.5.1 Mapping m^{6}A

The first transcriptome wide method for identification of m^{6}A sites within individual RNAs was described in 2012. Two groups independently developed methods which could be utilised for global m^{6}A mapping (Dominissini *et al.*, 2012; Meyer *et al.*, 2012). MeRIP-seq, as this method was termed, utilises specific antibodies, which bind to and immunoprecipitate m^{6}A residues on a transcript. The isolated transcripts are then sequenced, before aligning to a known transcriptome. The data collected from these

studies confirmed previous suggestions of a GAC/AAC sequence for m^{6}A (Wei, Gershowitz and Moss, 1976), and shown a large enrichment of m^{6}A within the 3'UTR and around stop codons. MeRIP-seq data also indicates that m^{6}A can be found throughout mRNAs, including intronic sequences, which indicates that it might be added co-transcriptionally (Carroll, Narayan and Rottman, 1990). This combined with the fact that the methyltransferase complex is found localised in nuclear speckles and methylation can be achieved within 10 minutes, support this theory (Bokar *et al.*, 1997; Ping *et al.*, 2014).

Since this pioneering method, new techniques have emerged. Single nucleotide resolution mapping is now possible with a technique known as m^{6}A individual nucleotide resolution crosslinking and immunoprecipitation (miCLIP) (Linder *et al.*, 2015b). Instead of relying on RNA fragmentation, miCLIP utilises crosslinking induced mutations during reverse transcription, directly as the sites of m^{6}A specific immunoprecipitation. Although this is a marked improvement on previous methods, there is still debate on the specificity on the antibodies, and the range of antibodies available. In addition, although meRIP-seq and miCLIP can give useful estimates of m^{6}A abundance and a map of methylated transcripts, the stoichiometry of the methylation remains unknown. Advances in nanopore sequencing have allowed the mapping of m^{6}A in full-length mRNAs, transcriptome wide. This rapid method can refine *Arabidopsis* genome annotation and perhaps increase the ease of studying further plant species.

1.5.2 Writing m^{6}A

Initial studies into the enzymes which facilitate m^{6}A formation utilised synthetic mRNAs, while detecting incorporation of ^{3}H from labelled SAM as a measure of m^{6}A addition (Harper *et al.*, 1990). This pioneering study reported for the first time, a consensus sequence for the recognition of m^{6}A sites (GGACU). Several years later, the Rottman lab, isolated nuclear fractions and tested the methylation activity of each. This led to the identification of a 70KDa fraction, containing a protein with S-adenosylmethionine (SAM) activity (Bokarso *et al.*, 1994) This eventually

lead to the identification and sequence of the first m^{6}A methyltransferase, MT-A70 (Bokar *et al.*, 1997), now known as METTL3. The complex is now known to contain many factors in a highly ordered manner (Figure 1.4). The components in *Arabiposis,* Mammals and *Drosophila* are compared in Table 1.1.

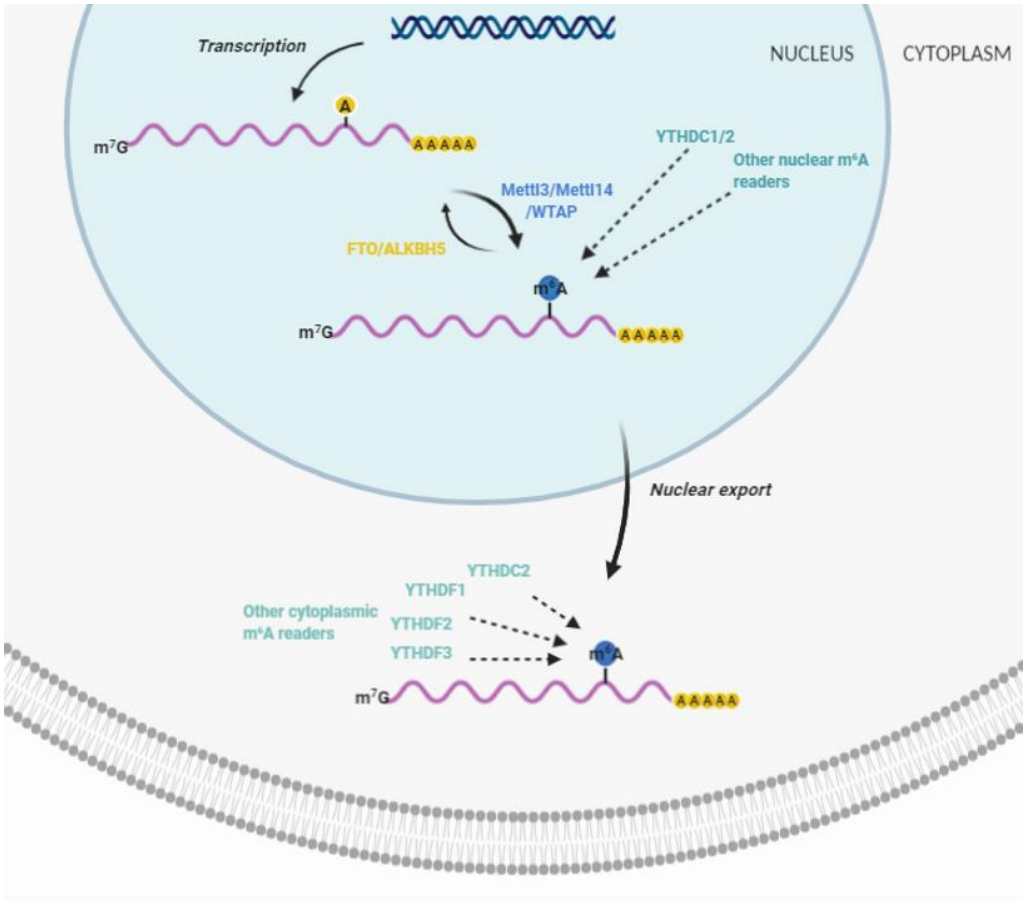


Figure 1.4: Readers, writers and erasers of m^{6}A modifications. The m^{6}A writer complex which is comprised of the core methylase METTL3 and its adaptors are localised to the nucleus. Erasers ALKBH-like erasers are aslo commonly found in the nucleus. ALKBH5 is the main eraser acting on m^{6}A. While in the nucleus m^{6}A mRNAs can be affected by nuclear reader proteins YTHDC. In the cytoplasm, m^{6}A readers YTHDF1-3 directly affect the exported mRNAs, along with other reader proteins.

Arabidopsis	**Mammals**	***Drosophila***	**Linked Phenotypes**	**Reference**
MTA (A14g10760)	METTL3	METTL3	Embryo lethality (arrested at globular stage).	(Zhong *et al.*, 2008)
			Reduced apical dominance, abnormal organ definition and increased number of trichome branches.	(Bodi *et al.*, 2012a)
			Reduced root growth and defects in protoxylem and gravitropism.	(Růžička *et al.*, 2017)
			Abnormal transcription termination.	(Pontier *et al.*, 2019)
			SAM overproliferation, delayed growth and abnormal leaf shape.	(Shen *et al.*, 2016)
MTB (A14g09980)	METTL14	METTL14	Embryo lethality.	
			Reduction in m^6A/A ratio to 50% Delayed growth, defects in apical dominance, protoxylem, root growth and gravitropism.	(Růžička *et al.*, 2017)
FIP37 (At3g54170)	WTAP	Fl(2)d	Delay in the pace of the cell cycle in embryo and endosperm. Embryo arrest at midglobular stage.	(Zhong *et al.*, 2008)
			More numerous, highly branched trichomes.	
			Arrested seeds and failed to produce viable homozygous seeds.	(Růžička *et al.*, 2017)
			Viable with stunted growth, reduction in 5-15% m^6A level, defects in apical dominance, protoxylem, and gravitropism.	(Růžička *et al.*, 2017)
			SAM overproliferation, delayed growth and abnormal leaf shape.	(Shen *et al.*, 2016)
			Abnormal transcription termination.	(Pontier *et al.*, 2019)
	KIAA1429	VIRILIZER	Embryo lethality.	

VIR (At3g05680)			Reduction in m^{6}A/A ratio to 5-15% Stunted growth, defects in apical dominance, protoxylem, lateral root and gravitropism.	(Růžička *et al.*, 2017)
HAKAI (At5g01160)	CBLL1	CBLL1	Lengthening of the circadian period. Altered PAS usage. Reduction in m^{6}A levels to 65% Viable, morphologically wild type phenotype.	(Parker *et al.*, 2020) (Růžička *et al.*, 2017)
HIZ2 (At5g53440)	ZC3H13	SPENITO	85% reduction in m6A abundance and severe developmental defects.	(Zhang *et al.*, 2022)
FPA? (At2g43410)	RBM15	RBM15	Delayed flowering.	(Schomburg *et al.*, 2001a)
FIONA1? (At2g21070)	METTL16	METTL16	Longer circadian period.	(Kim *et al.*, 2008)
ALKBH10B (At4g02940)	ALKBH5	ALKBH5	KO results in elevated m^{6}A/A ratio in total mRNA. Late flowering. Overexpression results in reduced m^{6}A/A ratio in total mRNA. Early flowering	(Duan *et al.*, 2017)
ALKBH9B (At2g17970)	ALKBHS	ALKBHS	No significant changes in m^{6}A/A ratio of endogenous mRNAs. Elevated m^{6}A/A ratio in viral RNA of AMV-infected plants. Increased resistance to Alfalfa Mosaic Virus (AMV).	(Duan *et al.*, 2017)

Table 1.1: m^{6}A methylation complex components and eraser proteins in *Arabidopsis*, Mammals and *Drosophila*. Phenotypes of *Arabidopsis* KO and overexpression lines summarised.

1.5.2.1 METTL3

METTL3, is part of a highly conserved family of SAM-dependent methyltransferases. It is an essential component of methylation, with numerous studies showing almost complete loss of m^6A levels, after genetic knockout of METTL3 (Zhong *et al.*, 2008; Ke *et al.*, 2017). In plants, homologs of METTL3, was identified and named MTA (methyltransferase A). Null mutants of MTA are embryo lethal, which indicates that m^6A is essential for plant survival (Zhong *et al.*, 2008) (Table 1.1). Expression of MTA under the control of an embryo specific promoter (ABI3) that is only weakly expressed post germination, rescues this lethality (Bodi *et al.*, 2012b). This system reduces the levels of m^6A to 5-15% allowing the study of m^6A related phenotypes and functions in depth. These lines exhibit reduced apical dominance, increased number of trichomes, defects in protoxylem and gravitropism alongside reduced root growth (Zhong *et al.*, 2008, Bodi *et al.*, 2012b, Růžička *et al.*, 2017). In addition to observable phenotypes MTA KO also leads to abnormal transcription termination as a result of reduced control of mRNA chimera formation (Pontier *et al.*, 2019).

1.5.2.2 METTL14

After identification of METTL3, phylogenetic analysis of the MT-A70 family of methyltransferases, suggested METTL14, which shares 43% identity with METTL3, as a homolog of METTL3. This led He and colleagues to induce METTL14 knockdown, using a siRNA system. 80% knockdown of METTL14, results in between 35%-40% reduction of m^6A in mammalian cell lines (Liu *et al.*, 2014). In the same study, tagged METTL3 and METTL14 recombinant proteins formed a stable complex in both gel filtration analysis and SDS-PAGE, with a 1:1 stoichiometry. Despite the evidence clearly demonstrating the essential role of METTL14, the mammalian protein has a degenerate SAM-binding domain which is non-functional (Wang *et al.*, 2016). METTL14 has since been identified as a methyltransferase complex

adaptor essential for METTL3/m^{6}A activity, but is not a functional methyltransferase itself, although this has not been identified in the plant homolog, MTB. Although METTL14 does transfer the methyl group, it does exhibit RNA binding properties and orientates the methyl group and transcript for transfer via METTL3 (Śledź and Jinek, 2016) (Table 1.1). The plants homolog of METTL14, MTB exhibits similar phenotypes to its counterpart MTA upon KO. These include reduced apical dominance and similar root growth phenotypes, include defects in root gravitropism (Růžička *et al.*, 2017).

1.5.2.3 WTAP

Initial work in our lab, uncovering the plant methylation complex led to the discovery of a number of novel components of the methyltransferase machinery (Zhong *et al.*, 2008). Yeast-two hybrid analysis of MTA-binding proteins, resulting in the identification of FIP37, the plant homolog of WTAP and female –lethal(2)d in mammals and Drosophila respectively. FIP37 was confirmed to bind to MTA, and co-localise in the nucleus, with other components of these complex. Several years later, Agarwala and colleges, showed WTAP/FIP37 yeast homolog mum2, binds to METTL3 homolog Ime4 and is essential to methylation (Agarwala *et al.*, 2012), later confirmed in mammalian lines (Liu *et al.*, 2014; Schwartz, Maxwell R. Mumbach, *et al.*, 2014). Interestingly, reduction of WTAP expression causes a drop in levels of the complex within nuclear speckles, suggesting WTAP assists in the localisation of the methyltransferase machinery (Table 1.1). Again, like MTA and MTB, in plants *fip37* null mutants exhibit phenotypes now associated with loss of m6A. Unlike MTA and MTB KO, *fip37* mutant lines were shown to be viable with around 5-15% m^{6}A levels. Phenotypes include delay in the cell cycle, increased trichome branching, inability to produce homozygous seeds, as well as abnormal transcription termination again suggested to be down to reduced cellular control of mRNA chimeric formation at transcription (Zhong *et al.*, 2008; Růžička *et al.*, 2017; Pontier *et al.*, 2019). Overexpressed fip37 lines exhibit SAM overproliferation, delayed growth display and serrated leaves (Shen *et al.*, 2016).

1.5.2.4 Virilizer/KIAA1429

After the discovery of WTAP, a proteomics analysis hinted at the presence of another member of the complex. KIAA1429 (also known as vir-like m^{6}A methyltransferase associated protein- VIRMA) is a strong interactor of WTAP and is the homolog of Drosophila protein Virilizer.(Schwartz, Maxwell R Mumbach, *et al.*, 2014). This interaction has also now been shown in Drosophila and plants, in addition to interactions with Ime4/MTA, and FL(2)d/FIP37 (Lence *et al.*, 2016; Růžička *et al.*, 2017). As with other members of the complex, knockdown of Virilizer results in significantly lower levels of m^{6}A on poly(A) transcripts. Recent work has implicated m^{6}A in the role of Virilizer and FL(2)d in sex specific splicing of Sxl in Drosophila. Absence of m^{6}A in Drosophila, results in flightlessness and male sex bias. Female specific splicing of Sxl requires m^{6}A found in 5'UTR (Haussmann *et al.*, 2016; Lence *et al.*, 2016). The exact role of Virilizer and its homologs in the formation of the complex remains unknown (Table 1.1). In plants, loss of VIR is embryo lethal with embryos arresting at the globular stage of development. Reduced ABI3:vir expression lines results in 5-15% m^{6}A level plants with stunted growth, reduced apical dominance and root growth defects (Růžička *et al.*, 2017).

1.5.2.5 RBM15/15B

Like Virilizer, proteomic studies of WTAP identified RBM15 and RBM15B, paralogous RNA-binding proteins, as strong interactors of WTAP (Horiuchi *et al.*, 2013a). This led to speculation of potential roles within the methylation complex. Co-immunoprecipitation revealed that RBM15 and RBM15B, interact with METTL3, but this is WTAP-dependent (Patil *et al.*, 2016). Patil et al, knocked-out both RBM15 and RBM15B, resulting in significant reductions of m^{6}A. This has been confirmed in Drosophila homolog, Spenito (Nito), which also reduces m^{6}A levels when disrupted (Lence *et al.*, 2016). RBM15 and RBM15B has essential roles in the

recognition of key consensus sequences. Both bind to U rich regions, before recruiting the METTL3/WTAP complexes. Not all DRACH motifs are located in uridine enriched regions, so it is possible that undiscovered co-factors mediate recognition in these transcripts.

A specific role of RBM15/15B is in the lncRNA XIST. X-inactivation during development is mediated by this lncRNA which requires methylation. This XIST-mediated silencing requires RBM15/15B (Patil *et al.*, 2016).

Interestingly, the closest homolog of RBM15A/B is FLOWERING LOCUS PA (FPA). FPA is an RNA-binding protein required for distinct control of flowering time. The overall identify is low, however. The involvement of FPA in the methylation complex in plant has not been confirmed, however partial loss-of-function mutations of FPA do show transcriptional read-through and chimeric RNA formation, similar to phenotypes of other m^{6}A methylase components (Duc *et al.*, 2013; Pontier *et al.*, 2019). FPA *null* mutants do not show embryo lethality, so in contrast to other methylase components, FPA is not essential to m^{6}A methylation (Schomburg *et al.*, 2001b) (Table 1.1).

1.5.2.6 ZC3H13-WTAP

Zinc finger CCCH domain-containing protein 13 (ZC3H13) and its Drosophila homolog Fl(2)d-associated complex component (Flacc) were identified through interactome analysis performed on mouse and Drosophila cell extracts. In this study, Flacc disruption results in aberrant splicing of Sxl, in line with m^{6}A dependent splicing, previously demonstrated in collaboration with our group (Knuckles *et al.*, 2018). Additionally the Roignant group, also identified Flacc as an intermediate between the interactions of RBM15/15B (Nito) and WTAP (Fl(2)d), therefore a member of the rapidly growing complex. Recently, our lab identified two HAKAI interacting zinc finger proteins HIZ1 and HIZ2, both components of the methylation complex. While null mutants of hiz1 had no observable phenotypes, *hiz2* mutants exhibit 85% WT m^{6}A levels with severe developmental defects. This observation along with HIZ2 interaction with

HAKAI lead them to suggest that HIZ2 is the likely equivalent of ZC3H13 in plants (Zhang *et al.,* 2022).

1.5.2.7 HAKAI

HAKAI, an E3 ubiquitin ligase, was first identified by its role mediating the post-translational down-regulation of the E-cadherin complex (Fujita *et al.*, 2002). Its presence in interactomes with other m^{6}A writer complex members (Horiuchi *et al.*, 2013b), assisted in the discovery of its role in the complex, this time in plants. Work in our lab simultaneously identified the conserved role of HAKAI in Eukaryotic methylation complexes, and the homologs of mammalian and Drosophila complex member VIR, in plant systems (Růžička *et al.*, 2017)(Table 1.1). Interestingly, disruption of HAKAI in *Arabidopsis* is non-lethal, unlike MTA, MTB, FIP37 and VIR, which do not process past the globular stage. m^{6}A levels in *hakai* lines are reduced by around 40%, which is relatively low when compared to *ABI3:MTA*, *ABI3:FIP37* or *ABI3:VIR,* suggesting a potentially unique function of Hakai in methylation. Hakai knockout doubles with *ABI3:MTA* and *ABI3:FIP37* dramatically increases the severity of the deleterious phenotypes observed in the single ABI3 lines, confirming its role as a *bona fide* component of the methyltransferase machinery. Although Hakai mutants do not display any aberrant phenotypes associated with the reduction in m^{6}A levels in plants, recent findings have identified specific transcripts which have altered m^{6}A levels. Indicating roles in root hair morphological changes and gravitropism (Zhang *et al.*, 2022).

1.5.2.8 METTL16

METTL16 was shown to have m^{6}A-forming function during studies into intron retention in methionine adenosyltransferase 2A (MAT2A), which is the major SAM forming enzyme. Observations noted that the hairpin

formed during this intron retention were highly similar to the RNA hairpin found in the U6 snRNA (Pendleton *et al.*, 2017). Both have m^{6}A at the same residue in these hairpins. As m^{6}A methylation requires SAM, it was unlikely that this methylation was inconsequential in the formation of the hairpin and the SAM-dependent splicing of the MAT2A intron. However, METTL3 is not responsible for the methylation on this transcript or in the u6 snRNA. Similar studies in *Schizosaccharomyces pombe* suggest that METTL16 may be responsible. Indeed, Conrad and colleagues showed that human METTL16 synthesises m^{6}A in both MAT2A intron retention and u6 snRNA hairpins (Pendleton *et al.*, 2017). Further investigation may identify further targets of METTL16 mediated methylation (Table 1.1).

1.5.3 Erasing m^{6}A

A key discovery in the study of m^{6}A was the identification of proteins which catalyse the removal of m^{6}A. Two of these demethylases were originally proposed to 'erase' m^{6}A; FTO and ALKBH5 (Jia *et al.*, 2011; Zheng *et al.*, 2013). Although original discoveries indicated that erasers were integral in the function and fine tuning of m^{6}A, allowing dynamic removal of methylation (Wang *et al.*, 2015) it now appears to have a limited role under standard physiological conditions. This change in stance was primarily led by the discovery that m^{6}A are stable throughout the life cycle of a cell (Darnell, Ke and Darnell, 2018).

1.5.3.1 FTO

Fat mass and obesity-associated protein (FTO) sequences predicted functional homology with the AlkB subfamily of Fe(II)/ α-Ketoglutarate-dependent dioxygenases superfamily that has 9 members (ALKBH1-8 and FTO) in humans (Xu, Liu, *et al.*, 2014). The function of the ALKB family to demethylate DNA and RNA had long been studied (Ougland *et al.*, 2004). FTO displays demethylase activity on a variety of methylated nucleotides and was subsequently identified as a legitimate m^{6}A demethylase (Jia *et al.*,

2011). These studies were carried out in vitro and only subtle changes are observed in m^{6}A levels within poly(A) mRNA, with FTO depletion or when overexpressed (Jia *et al.*, 2011). Later studies involving mammalian systems, showed that very little observable differences in global m^{6}A levels occur in *fto* compared with WT. In depth analysis via meRIP-seq shows change in only a limited subset of methylated transcripts (Hess *et al.*, 2013). Further analysis showed that transcripts which display elevated m^{6}A in *fto* lines had m^{6}A within the 5'UTR not the 3'UTR (Mauer *et al.*, 2017). The Jaffrey lab has since proposed that this elevation in certain 5'UTR m^{6}A marks is due to the function of FTO in the demethylation of cap specific m^{6}Am, highly similar to m^{6}A. Jaffery continues to show that FTO has nearly 100 times the catalytic activity for m^{6}Am over m^{6}A, and that physiologically significant function of FTO is in the context of the m7G cap (Mauer *et al.*, 2017). Therefore, it is possible that FTO is the eraser for m^{6}Am and not m^{6}A. Due to the significantly higher abundance of m^{6}A over m^{6}Am, m^{6}A is the predominant substrate for FTO *in vivo*, despite higher binding preference for m^{6}Am (J. Wei *et al.*, 2018). These findings, along with structural analysis suggesting an actual preference for m^{6}A not m^{6}Am have heated up this debate (Zhang *et al.*, 2019). Contrasting hypotheses around the context of any functions of FTO remain.

1.5.3.2 ALKBH5

ALKBH5 was initially identified in a screen of demethylase proteins for m^{6}A specific activity (Zheng *et al.*, 2013). Similar to FTO, ALKBH5 knockdown only results in small increases in cellular levels of m^{6}A, but this suggests that ALKBH5 may be a bona fide m^{6}A demethylase (Zheng *et al.*, 2013). Unlike FTO however, ALKBH5 exhibits no preference for m^{6}Am (Mauer *et al.*, 2017). ALKBH5 expression is limited to the nucleus, due to mature mRNA export from the nucleus post-transcription, it is unlikely that ALKBH5 demethylases mature mRNA. Numerous studies have observed specific changes in m^{6}A peaks, after ALKBH5 disruption. Regulatory expression under hypoxia, tumorigenesis, viral infection as well as enrichment in male and female germlines, suggest significant and specific

functions of ALKBH5 but how this relates the m^{6}A pathway remains unclear (Thalhammer *et al.*, 2011; Tang *et al.*, 2017; Zhang *et al.*, 2017; Rubio *et al.*, 2018).

1.5.3.3 Demethylation in plants

In *Arabidopsis*, there are thirteen AlkB family proteins, based on sequence alignments and functional predictions five of these (ALKBH9A, ALKBH9B, ALKBH9C, ALKBH10A and ALKBH10B) are likely homologs of the mammalian demethylase ALKBH5 (Duan *et al.*, 2017). Homologs of these proteins can also be found in other key plant model species including *Solanum lycopersicum*. No homolog of the FTO demethylase can be found in plant genomes, this may correlate with the absence of m^{6}Am in higher plants (Unpublished data). Many of the suggested demethylase genes in *Arabidopsis* show high expression in a range of tissues. Recent experiments have shown that ALKBH9B and ALKBH10B both have m^{6}A demethylation activities *in-vitro* and ALKBH10B on *in-vivo* transcripts (Duan *et al.*, 2017; Martínez-Pérez *et al.*, 2017). Duan *et al.*, demonstrated that ALKBH10B is essential for the proper transition from vegetative to reproductive stages, mediated by stabilising of key floral transition transcripts via demethylation (Duan *et al.*, 2017). Although ALKBH9B has not been observed influencing m^{6}A levels in planta, it does influence m^{6}A abundance on the viral genome of Alfalfa mosaic virus (AMV) and this influences its infectivity (Martínez-Pérez *et al.*, 2017).

1.6 Reading m^{6}A

The m^{6}A mark confers its function in a number of ways, it can act to repel RBPs (Arguello, Deliberto and Kleiner, 2017), or the recruitment of m^{6}A binding proteins known as 'readers'. After recruitments to their targets these proteins themselves can directly control mRNA fate or indirectly by further recruitment of effector proteins. There are a number of RBP families whose roles are emerging.

1.6.1 YTH Domain proteins

YT521-B homology domain containing proteins (YTH proteins) are the predominant reader proteins in Eukaryota. The domain was first identified on the human splicing factor YT521-B (YTHDC1) and then was later categorised as a highly structured and conserved RBP (Hartmann *et al.*, 1999; Stoilov, Rafalska and Stamm, 2002). YTH domain proteins fall into two major categories. The nuclear YTH proteins, termed YTH domain containing (YTHDC) and the cytoplasmic, termed YTH domain family (YTHDF) (Scutenaire, J.-M. Deragon, *et al.*, 2018). The canonical YTH domain is around 150aa The YTH domain confers a tertiary protein structure which is highly conserved and has been confirmed biochemically in a number of studies (D. Li *et al.*, 2014; F. Li *et al.*, 2014; Shukun Luo and Tong, 2014; Xu, Wang, *et al.*, 2014; Zhu *et al.*, 2014). All YTH proteins appear to adopt a specific fold pattern consisting of three (alpha) helices and six (beta) sheets. The (beta) sheets are arranged in a barrel conformation, with the three (alpha) helices tightly packed against it. This forms a canonical hydrophobic aromatic cage within a core (Figure 1.5). In all studies, this aromatic cage (formed by the C-termini of beta 1 and 2 with alpha 1, and the N-terminus of alpha 2 with beta sheets 4 and 5) consists of two conserved tryptophan residues and either a third tryptophan (WWW, most YTHDF proteins (F. Li *et al.*, 2014), a tyrosine (WWY, in yeast (Zhu *et al.*, 2014)) or a leucine (WWL, found in most YTHDCs (Xu, Wang, *et al.*, 2014)). The nucleotides surrounds the m^6A within the consensus site (residues -2,-1, +1, +2) have significant contribution to the integrations within the aromatic cage, but they do not interact directly with the YTH domain (Xu, Wang, *et al.*, 2014). Although these studies all demonstrated a selective preference for m^6A, a number of the YTH proteins analysed in mammals and yeast, also display general RNA binding properties to A, however this is to a much lower affinity (Shukun Luo and Tong, 2014).

There is an interesting divergence in the number of YTH domain proteins present in different evolutionary groups. Mammalian genomes code for five YTH domain proteins, two in the nuclear DC group (YTHDC1-2) and three

in the cytoplasmic DF group (YTHDF1-3) (Patil, Pickering and Jaffrey, 2017). *Drosophila melanogaster* only have two YTH proteins; YT521-B, renamed YTHDC1 and CG6422, renamed YTHDF (Haussmann *et al.*, 2016). *Saccharomyces cerevisiae* genomes encode for a single YTH domain protein, Pho92 , of the YTHDF type (Kang *et al.*, 2014). Finally, plant genomes contain largely expanded families of both types of YTH domain proteins. *Arabidopsis* genomes, for examples contain two YTHDC readers (ECT12 and CPSF30-L), and eleven highly similar YTHDF proteins (ECT1-ECT11) (Ok *et al.*, 2005; D. Li *et al.*, 2014; Li *et al.*, 2017). It is unclear whether all of these proteins actively bind m^{6}A. In fact, mammalian YTHDC2 did not show m^{6}A binding properties during iCLIP assays, although it has been observed to bind m^{6}A *in vitro* (Patil *et al.*, 2016). It may be the case that certain YTH proteins only bind m^{6}A under certain conditions or in certain cells.

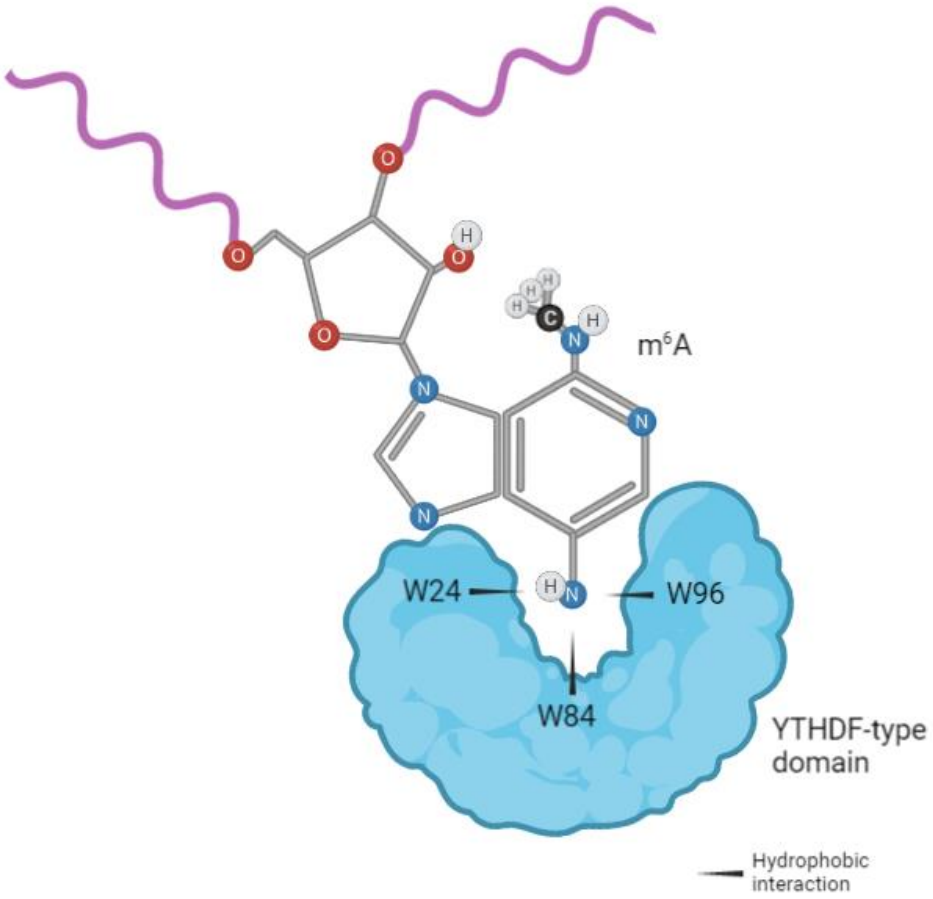


Figure 1.5: Hydrophobic aromatic cage in the YTH domain. An example diagram demonstrating the WWW-type aromatic cage found within

plant YTHDF domain containing proteins. Tryptophan residues W24 and W96 within with YTH domain sandwich the 6-methyladenine group, via hydrophobic interactions. This has been shown to orientate the methyl group to a third tryptophan (W84), forming the remainder of the cage. The amino acid positions are highly conserved across species (F. Li *et al.*, 2014).

1.6.2 Molecular and physiological roles of YTH domain proteins in animals

The YTH proteins in mammals are the most intensively studied m^{6}A reader proteins. Although the most prevalent role of m^{6}A is related to translation promotion and mRNA stability, a large group of methylated transcripts are involved in complex interactions with YTH domain proteins and their effectors. This has been observed in several roles and mechanisms within mammals. To what extent these physiological roles are conserved in other key model species is not yet clear.

1.6.2.1 Alternate splicing

The original characterisations of YTH domain proteins before further characterisations as m^{6}A reader proteins, gave hints into the m^{6}A dependent roles within the cell. YTHDC1 was first identified as a regulator of alternate splicing, this was later demonstrated as YTHDC1-m^{6}A dependent (Stoilov, Rafalska and Stamm, 2002; Xiao *et al.*, 2016). YTHDC1 acts a recruiter for pre-mRNA splicing factor SRSF3, and blocks SRSF10 from binding the nascent transcript (Xiao *et al.*, 2016). The result is a promotion of exon inclusion alternate splicing. This is mediated by specific YTHDC1 controlled localisation of SPSF3 and SPSF10 in nuclear speckles, with enrichment and repression of SPSF3 and SPSF10 respectively. Interestingly, not all SPSF3 and SPSF10 targets contain m^{6}A or are bound by YTHDC1, suggesting this regulation is transcript specific (Xiao *et al.*, 2016). SPSF3 and YTHDC1 also have roles in nuclear export of mRNAs. Hypophosphorylation of SPSF3 was shown to be essential in the export of a

number of YTHDC1 targets, through receptor NXF1, depending on the extent of phosphorylation (Roundtree and He, 2016).

1.6.2.2 RNA stability

YTHDF2, was the first YTH protein to be functionally characterised. Like YTHDC1, YTHDF2 has been shown to selectively bind m^{6}A-contianing RNAs (Wang *et al.*, 2013) YTHDF2 deletion results in a significant increase in a large number of target transcripts. This was shown to be linked to the deadenylation step of mRNA. Moreover, targets of YTHDF2, tend to have shorter poly(A) tails than the mean, supporting this discovery (Wang *et al.*, 2013). METLL3 depletion, reduces deadenylation during *in vitro* assays, suggesting that YTHDF2 mediated deadenylation is itself mediated by m^{6}A (Du *et al.*, 2016). The CCR4-NOT complex is directly recruited by YTHDF2, specifically at the N-terminus. This interaction is essential for deadenylation of m^{6}A containing RNAs. Further potential roles of YTHDF2 in the cell, revolve around YTHDF2 mediated mRNA/mRNP aggregations in processing bodies (p-bodies). This activity was initially proposed as a mechanism of YTH directed decay of targeted transcripts (Sheth and Parker, 2003; Wang *et al.*, 2015). Since P-bodies are not required for mRNA decay (Eulalio *et al.*, 2007), have recently been shown to have very little degradation activity, therefore the functional role of these aggregations need further elucidation YTHDF1 and YTHDF3 have been shown to target methylated transcripts and recruit YTHDF2 for degradation (Shi *et al.*, 2017).

1.6.2.3 Translational initiation and control

YTHDF1 mediates the m^{6}A pathway's role in translation promotion, and this has been observed in a number of cells and species (Wang *et al.*, 2015; Weng *et al.*, 2018a). YTHDF1 actively recruits eukaryotic initiation factor 3 (eiF3). eiF3 then binds mRNAs and associated ribosomal complexes, to facilitate organisation of associated translation factors and machinery

(Wang *et al.*, 2015). This is often with the assistance of YTHDF3 (Shi *et al.*, 2017). YTHDF3 has a similar role, however it acts as a part of an interaction complex to assist translation, sharing targets with both YTHDF1 and YTHDF2.

It has been suggested that YTHDC2 participates in translational control within mouse testes, however only minimal changes in m^6A levels were observed. YTHDC2 contains numerous domains (Phillip J Hsu *et al.*, 2017) , whether this is an m^6A dependent phenotype has yet to be conclusively proven.

1.6.2.4 Sex determination

YTHDC1 and its homologs are involved in sex determination in a number of species. YTHDC1 in Drosophila is essential for the sex specific splicing of *Sxl* pre-mRNA. M^6A is necessary for the exclusion of the male specific exon in *Sxl*. Depletion of YTHDC1 results in male bias, whereas overexpression is "male lethal", with only females produced (Haussmann *et al.*, 2016; Lence *et al.*, 2016). Meanwhile in mammalian systems, XIST mediated, X-chromosome dosage compensation, can be promoted by YTHDC1. This activity has been shown to be driven by several m^6A residues found on XIST lncRNA (Patil *et al.*, 2016).

1.6.2.5 Gametogenesis

Mouse nuclear YTH proteins, YTHDC1 and YTHDC2 are both essential in viability and development of both male and female germlines. Depletion of YTHDC1, results in aberrant splicing of a large number of transcripts, as well cell specific changes of target poly(A) and 3'UTR lengths. Spermatogonial development and embryo viability are both impaired (Kasowitz *et al.*, 2018). YTHDC2 depletion causes an increase in transcripts encoding meiotic regulators. These functions are as part of a complex with

XRN1 suggesting YTHDC2 is involved in meiosis specific transcript destabilisation (Phillip J Hsu *et al.*, 2017; Wojtas *et al.*, 2017).

1.6.2.6 Neuronal development

In Drosophila, which are depleted of YTHDC1, several behavioural defects are observed, with abnormal flying and inability to climb (Haussmann *et al.*, 2016; Lence *et al.*, 2016). Also in flies, YTHDF controls neuronal development in mushroom bodies and neuro-muscular junctions, through translational repression of two microtubule associated transcripts, *Chickadee* and *Futsch.* This is in direct association with fragile mental retardation protein homolog Fmr1 (Soldano *et al.*, 2020). In mice, YTHDF1 is required for repair and regeneration of damaged axons (Weng *et al.*, 2018b). Additionally essential mRNAs for correct learning and memory in adult mice hippocampi, contain m^6A. YTHDF1 binds to these transcripts in reaction to certain stimuli, localising transcripts in preparation for translation (Shi *et al.*, 2018).

1.6.2.7 Abiotic Stress responses

A growing list of studies have implicated YTH domain proteins in the cellular response to stressful conditions. In mammalian cell lines, m^6A is additionally added to multiple translationally active mRNAs under stress (oxidative and heat). This induces 'triaging' to foci known as stress granules (SGs). The localisation to these foci is mediated by YTHDF3. It was also noted that the mRNAs present in these stress granules exhibit a distinct pattern of m^6A (Anders *et al.*, 2018). It is possible that this mechanism is part of an innate m^6A dependant response to stress. The formation of these granules was later shown to be reliant on all three YTHDF proteins (DF1-3). Both the disordered N-terminus and the C-terminal YTH domain are crucial for SG formation (Fu and Zhuang, 2019). This YTHDF dependent

phase separation is facilitated by m^{6}A (Liu *et al.*, 2020). Further to this, in animal cells, transcripts encoding key heat stress response proteins are specifically methylated in their 5'UTR in response to heat stress. It was proposed that YTHDF2 indirectly promotes their cap-independent translation, through recruitment of eIF3 (Zhou *et al.*, 2015). It is likely that m^{6}A also facilitates responses to other stresses, however these require further investigation.

1.6.2.8 Cancer

M^{6}A and its conserved complexes have been implicated in numerous oncogenic responses and cancer progressions (Reviewed in (He *et al.*, 2019). Reader proteins are also involved in these functions.

YTHDC2 mRNA levels are significantly higher in several cancerous cell lines. Depletion of YTDC2, reduces the mobility and number of metastases, as well as cell proliferation in hepatocellular carcinoma, in colon tumour cell lines (Tanabe *et al.*, 2016). Colon tumour cells, exhibiting YTHDC2 depletion, were inserted into mouse spleens. In these cells the translation of around three hundred YTHDC2 target genes was significantly increased (Tanabe *et al.*, 2016). This suggests that YTHDC2 regulates cancer proliferation through translational control of its targets.

High YTHDF1 mRNA levels have been associated with malignancy in tumours, and correlations have been drawn between these and a poor chance of survival in cancer patients. This is supported by a reduction in proliferation and increased sensitivity to chemotherapy, post YTHDF1 knockdown (Nishizawa *et al.*, 2018).

A similar pattern is observed for YTHDF2 in hepatocellular carcinoma cells. In fact, YTHDF2 has been shown to bind to anti-tumour cytokine signalling 2 (*SOCS2)* transcripts, which is modified by a number of m^{6}A marks. This promotes degradation of *SOCS2,* increasing proliferation (Chen *et al.*, 2018).

These roles are likely only the beginning, with future developments unearthing the overall mechanisms involved in m^{6}A regulated oncogenes is.

1.6.2.9 Immune response and viral infection

The presence of m^6A on viral RNA has been observed in multiple viruses, however the function behind its addition is unclear. HIV-1, Influenza A, Hepatitis C and Zika, are all examples of viruses with m^6A (Gokhale *et al.*, 2016; Kennedy *et al.*, 2016; Lichinchi *et al.*, 2016; Courtney *et al.*, 2018). Outside of animal systems, links between the m^6A pathway and viruses has also been observed in plant viral infection and immune response (Conti *et al.*, 2017; Martínez-Pérez *et al.*, 2017; Jiang *et al.*, 2018).

Interestingly, reader proteins have been implicated in both promotion and repression of viral infection, this appears to be dependent on the type of virus.

In Influenza A infection, YTHDF2 alone, promotes viral infection, through increased viral replication. Furthermore, reduction in global m^6A inhibits influence A gene expression and replication (Courtney *et al.*, 2018).

The opposite was observed in both hepatitis C and ZIKA viruses, where YTHDF1, 2 and 3 all bind methylated viral RNA, antagonising infectious RNA. In ZIKA infection, YTHDF2 has by far the strongest response, mirroring influenza infections. In these cases, the proliferation of the viruses themselves were unaffected (Gokhale *et al.*, 2016; Lichinchi *et al.*, 2016). This suggest that the mechanism relies on specific interaction and processing of the viral RNAs themselves, in a rapid manner.

In the HIV-1 virus, all three YTHDF proteins bind to m^6A modified, HIV-1 transcripts. Conflicting findings have been obtained from two independent labs. Lu *et al.*,, demonstrated that YTHDF1-3 all antagonise HIV-1 infection when overexpressed, whereas knockdown, promoted viral infection (Lu *et al.*, 2018). Conversely, Kennedy *et al.,* observed increased HIV-1 infection, under YTHDF1-3 overexpression (Kennedy *et al.*, 2016). Any explanations for the conflict, have yet to be explained.

1.6.3 Molecular and physiological roles of YTH domain proteins in plants

The YTH domain in plants was originally characterised in *Arabidopsis*, in a study identifying two proteins which bind to calcinurin B-like interacting protein kinase (CIPK1). The discovery of these two proteins led to the classification of eleven proteins, as members of the same family (Ok *et al.*, 2005). These were termed **E**volutionarily conserved **C**-**T**erminal domain containing proteins 1-11 (ECT1-11) and later classified as YTHDF-type YTH domain containing proteins. Subsequent searches for YTH domain containing proteins in *Arabidopsis* has identified two additional proteins: ECT12, a nuclear YTHDC-type YTH domain protein (Scutenaire, J.-M. Deragon, *et al.*, 2018) and CSPF30-L, a long isoform of the splicing factor CPSF30 (Addepalli and Hunt, 2007; Li *et al.*, 2017).

1.6.3.1 YTHDF-type proteins

Very limited functional evidence is available, indicating roles for the YTHDF-type ECT proteins.

To date, only ECT2 has been demonstrated as a definitive m^6A-binding protein, and this binding requires its aromatic cage within the YTH domain protein (Arribas-Hernández *et al.*, 2018; L.-H. H. Wei *et al.*, 2018; Scutenaire, J.-M. Deragon, *et al.*, 2018). The large number of ECT proteins, suggests some level of redundancy. Indeed single mutants *ect2, ect3* or *ect4* display no obvious phenotypes. Double mutants *ect2;ect3* are involved in timing of leaf emergence and correct morphological development of leaves (L.-H. H. Wei *et al.*, 2018).

Triple mutant *ect2;ect3;ect4* display similar, albeit less severe phenotypes to those observed in low methylation *Arabidopsis* lines (Zhong *et al.*, 2008; Růžička *et al.*, 2017; L.-H. H. Wei *et al.*, 2018). *Ect2* and *ect3* loss-of-function mutants, also exhibit increased trichome branching (branched single cells

found on the epidermis of *Arabidopsis*) (Arribas-Hernández *et al.*, 2018; L.-H. H. Wei *et al.*, 2018; Scutenaire, J.-M. Deragon, *et al.*, 2018). This was shown to be mediated by degradation and down-regulation of three trichome specific transcripts, in *ect2* disruption lines (Arribas-Hernández *et al.*, 2018). This suggests that ECT2 has a functional role in mRNA stability during development. Intriguingly, ECT2 localises in both the cytoplasm and the nucleus. ECT2 could therefore regulate pre-mRNA synthesis to influence gene specific regulation.

Similar to YTHDF2, ECT2 relocalised to stress granules during heat-stress (Arribas-Hernández *et al.*, 2018). This is also observed in ECT4, but not in ECT3. Scutenaire *et al.*, predicted that this can be explained due to the presence of YPQ-rich regions within both ECT2 and ECT4, and the absence of such a region in ECT3 (Scutenaire, J.-M. Deragon, *et al.*, 2018). YPQ-rich regions are also present in ECT5-8. Conversely, the Brodersen lab suggested the ability of a YTH domain protein to localise in specific foci can be predicted by the level of disorder outside of the YTH domain itself (Arribas-Hernández *et al.*, 2018).

1.6.3.2 YTHDC-type proteins

The *Arabidopsis* genome encodes two YTHDC-type proteins, ECT12 and CPSF30-L. CPSF30, in the homolog of the 30KDa subunit of both human and yeast CPSF. Interestingly, in plants, CPSF30 codes for two isoforms CPSF30-S and CPSF30-L, with the later containing an YTH domain. While CPSF30-S regulates the choice of PAS in a large number of genes, the role of the YTH domain in CPSF30-L is largely unknown (Thomas *et al.*, 2012).

A recent study screening for factors involved in nitrate signalling uncovering an interesting role for CPSF30-L. CPSF30-L expression can rescue the aberrant nitrate signalling phenotypes associated with a point mutation in CPSF30 (*cpsf30.2).* Normal signalling cannot be restored with expression of CPSF30-S (Li *et al.*, 2017). This led Li *et al.*, to discover that CPsf30-L controls the expression of several nitrate responsive genes. The transcript of essential nitrate signalling gene NRT1.1, is present in an

isoform with a significantly shorter 3'UTR in *cpsf30.2* lines (Li *et al.*, 2017). This is potentially due to alternate PAS. This represents the first role of CPSF30-L and suggests a role in 3' end processing of key transcripts, potentially in conjunction with CPSF30-S. CPSF30-L could still be characterised as a key regulator in pre-mRNA processing.

Any potential functions of YTHDC protein ECT12, remain to be discovered

1.7 Thesis aims

The main scope of this thesis is to investigate the role of m^6A RNA binding proteins, containing a YTH domain have on plant development. Previous studies have identified that YTH-domain containing proteins in non-plant species have a key role in reading and mediating the function of the m^6A modification. These are split into two functional groups known as YTHDC and YTHDF proteins, both proposed to contain functionally distinct YTH domains.

The presence of YTH proteins within plant genomes suggest that a similar role as the primary m^6A readers would be observed there. Plant genomes contain a similar number of DC-type YTH proteins and an expanded group of DF-type proteins. *Arabidopisis* genomes for example, encode 13 YTH proteins; 12 previously described as Evolutionarily Conserved C-terminus (ECT) proteins and one a homolog of Cleavage and Polyadenylation Specificity Factor 4 (CPSF4), named CPSF30. Of these, ECT1 to 11 are YTHDF-type proteins, while CPSF30 and ECT12 are YTHDC-type proteins. The observed expansion of the YTHDF-type proteins in plants increases the difficultly of their study.

The overall aim of this work is to collect and analyse the expanded DF-type group of plant YTH proteins. Through genome-wide analysis of the protein sequences, we will be able to identify and characterise a suitable model for the *in-vivo* study of the YTH protein family, with less YTHDF-type proteins. Our hypothesis for this initial set up is that any plant species with fewer YTHDF proteins would exhibit reduced genetic redundancy and greater

ease of study. Once identified we will be able to produce mutant lines to develop a permanent platform for the future study of this key protein family.

The first studies into YTH domain proteins in plants, using micro-array data, found that YTH genes are induced in both abiotic and biotic stresses conditions (D. Li *et al.*, 2014). Although further clues into YTH protein specific roles have been uncovered, with ECT2, ECT3 and ECT4 localising into P-bodies and ECT2 relocalising into stress granules under heat stress in *Arabidopsis* (Arribas-Hernández *et al.*, 2018; Scutenaire, J.-M. Deragon, *et al.*, 2018), few advances have been made since. A recently discoveries demonstrated that hypermethylation of key salt stress transcripts resulted in selective stabilization (Anderson *et al.*, 2018). Collectively these findings suggest many overlapping roles of YTH proteins in plants response to abiotic stress. A further aim is utilisation of our YTHDF model to further our understanding of any functions in this area.

Chapter 2 - Materials and methods

2.1 Plant materials, media and growth conditions

2.1.1 Plant materials

Tomato seeds (Lycopersicon esculentum var Ailsa Craig) were obtained from fruit of a homozygous line maintained at Sutton Bonington since 1979.

2.1.2 Plant media

2.1.2.1 MSR3 solid growth medium

MSR3 medium was used for the germination and growth of seeds under sterile conditions. It consisted of 30g/L Sucrose, 4.3g/L Murashige and skoog salts, 1mL/L R3 vitamins. Adjusted to Ph 5.8 with 1M NaCl before adding 6.25g/L Agar and sterilised by autoclaving at 121°C.

2.1.2.2 MSR3ZA callusing medium

MSR3ZA callusing medium is made as per MSR3 media with the addition of 1.5mg/L Zeatin and 0.1mgL IAA once the media reaches 60°C. If required selective antibiotics were also added at the 60°C.

2.1.2.3 MSR3BA callusing medium

MSR3BA callusing medium is made as per MSR3 media with the addition of 1.5mg/L BAP and 0.1mgL IAA once the media reaches 60°C. If required selective antibiotics were also added at the 60°C.

2.1.2.4 MSR3 rooting medium

MSR3 rooting medium was used for the induction of roots in excised tomato shoots. It consisted of 15g/L Sucrose, 2.2g/L Murashige and skoog salts, 1mL/L R3 vitamins. Adjusted to Ph 5.8 with 1M NaCl before adding 6.25g/L Agar and sterilised by autoclaving at 121°C. 0.1mg/L IAA was added once the media reaches 60°C. If required selective antibiotics were also added at the 60°C.

2.1.2.5 Liquid MS

Liquid MS media is used during transformation via tissue culture of plants. It consists of 30g/L Sucrose, 2.2g/L Murashige and skoog salts, 1mL/L R3 vitamins. Adjusted to Ph 5.8 with 1M NaCl. Sterilised by autoclaving at 121°C.

2.1.3 Growth conditions

Seeds were surface sterilised by washing for 20 minutes using 50% (w/v) sodium chlorate. Seeds were then washed a minimum 5 times with sterile dH_2O. Sterile seeds were either sown in Levington M3 compost or in sterile T16 plastic containers containing 60mL MSR3 media. Plants were grown in growth or glasshouse rooms at 16 hours at 23°C and 8 hours at 18°C. Mercury vapour lamps were used to provide supplementary lighting to ensure these day lengths for glasshouse compartments. Tomato fruits were collected at red-ripening (RR) stage.

2.1.4 Tomato seed collection

Seed from fruit of each parent plant and placed in plastic bags. Seeds were covered in 0.1M HCl and left for at least 16 hours, to remove gel. Once

soaked, seeds were thoroughly rinsed with water before drying on labelled filter paper. After drying for 24 hours at room temperature, seeds were stored in labelled brown envelopes.

2.2 Bacterial growth, media and strains

2.2.1 Bacterial strains

The bacterial strains used in this study are listed in Table 1 and include modified *E.coli* and *A.tumefaciens* strains.

Table 2.1: List of strains used in this study

Bacterial strain	Genotype/Description	Function	Source
E.coli			
DH5α	F$^-$ φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F)U169 *deo*R *rec*A1 *end*A1 *hsd*R17(rk$^-$, mk$^+$) *pho*A *sup*E44 *thi*-1 *gyr*A96	Standard cloning	Fray lab strain collection
SoluBL21	F- ompT hsdSB (rB - mB -) gal dcm (DE3)†	Optimisation of soluble protein expression and production	Genlantis (San Diego, California, USA)
A.tumefaciens			
EHA105	C58 (rif R) Ti pEHA105 (pTiBo542DT-DNA) Succinamopine	Tomato specfic *in planta* DNA transfer	Fray lab strain collection

2.2.2 Media

2.2.2.1 Luria-Bertani medium

Standard Luria-Bertani media (LB) was used for general procedures in *E. coli* and *A.tumefaciens*. It consisted of 10 g/L tryptone, 5 g/L yeast extract , 10 g/L NaCl and NaOH to pH 7.0. LB agar was prepared by addition of 15 g/L Bactoagar No. 3 to the broth. LB media and agar were sterilised by autoclaving at 121°C.

2.2.2.2 Hannahan's Broth (SOB and SOC medium)

SOB medium was used for the preparation of DH5α chemically competent cells. It was reconstituted following manufacturer's instructions. The medium consists of 20 g/L tryptone, 5 g/L yeast extract, 2.4 g/L $MgSO_4$, 0.5 g/L NaCl and 0.186 g/L KCl and sterilised by autoclaving at 121°C. SOC medium was made in the same manner with the addition of 20mM sterile glucose, post autoclaving.

2.2.3 Culture growth

Unless stated otherwise, E. coli and A.tumefaciens strains were grown aerobically in a shaking incubator at 37 or 28 °C, respectively, with shaking at 200 rpm. 50mL centrifuge tubes (Corning) and shake flaks (Duran® baffled flask 250,500 mL 1 Litre or 2 Litre) were filled to a 10 % of their total volume to allow for good oxygenation. Culture growth was monitored by measuring optical density at 600 nm (OD600) BioMate™ 3S Spectrophotometer (Thermo Fisher Scientific).

2.2.4 Bacterial storage

All *E. coli* and *A.tumefaciens* strains were stored at -80 °C using 2mL safe-lock Eppendorf tubes. A liquid culture of the strain of interest grown overnight (O/N) was used to inoculate three tubes in 25% glycerol (final concentration). For recovery, Stocks were used to directly streak on to solid medium or dropped into an appropriate liquid medium.

2.3 General chemicals and buffers

AE buffer- 50mM NaOAc pH, 5.3; 10mM EDTA

Edwards DNA extraction buffer- 200mM Tris-HCl- Ph 7.5, 200mM NaCl, 25mM EDTA, 0.5% SDS

2x Lamelli buffer- 62.5mM Tris-HCL, pH 6.8; 25% glycerol; 0.01% (w/v) bromophenol blue; 2% SDS; 5% ß-mercaptoethanol

10x SDS running buffer- 250mM Tris base; 1.92mM glycine, 1% SDS

2x SDS page loading buffer- 62.5mM Tris-HCl, 25% Glycerol, 0.01% Bromophenol Blue, 2% SDS

P1 buffer- 30mM NaOAc

2.3.1 DH5alpha competent cell production buffers

RF1- 100mM RbCl, 50mM $MnCl_2$, 30mM Potassium acetate, 10mM $CaCl_2.2H_2O$, 15% w/v Glycerol: Adjusted to Ph 5.8- 0.2M Acetic acid

RF2- 10mM MOPS, 10mM RbCl, 75mM $CaCl_2.2H_2O$, 15% w/v Glycerol

2.3.2 Nickle column protein purification buffers

Ni column Elution buffer- 20mM Na_3PO_4; 300mM NaCl; 250mM Imidazole; adjust to pH 7.4, do not autoclave imidazole

Ni column equilibration buffer- 20mM Na_3PO_4; 300mM sodium chloride; 10mM Imidazole; adjust to pH 7.4, do not autoclave imidazole

Ni column MES buffer- 20mM 2-(N-morpholine)-ethanesuulfonic acid; 100mM NaCl, adjust to pH 5 45

Ni column wash buffer- 20mM Na_3PO_4; 300mM NaCl; 25mM Imidazole; adjust to pH 7.4, do not autoclave imidazole

2.3.3 Poly A buffers

RNA binding buffer- 1M LiCl; 40mM Tris-HCL, pH 7.5; 2mM EDTA; 0.1% NP40

RNA wash buffer- 150mM LiCl; 20mM Tris-HCL, pH 7.5; 1mM EDTA; 0.01% NP40

Oligo d(T) storage and equilibrium buffer- PBS; 0.1% Tween 20

2.3.4 Protein pull down 1 buffers

IP Lysis Buffer- 25mM Tris-HCL, pH 7.4; 150mM NaCl; 1% NP40; 1mM EDTA; 5% glycerol

RNA Capture Buffer- 20mM Tris, pH 7.5; 1M NaCl; 1mM EDTA

Protein RNA binding buffer (10x)- 200mM Tris, pH 7.4; 500mM NaCl; 20mM $MgCl_2$; 1%Tween 20

Wash Buffer- 20mM Tris, pH 7.4; 10mM NaCl; 0.1% Tween 20

2.3.5 Protein pull down 2 buffers

RNA binding Buffer- 50mM Hepes-HCL, pH 7.5; 150mM NaCl; 0.5% (v/v) Np-40; 10mM $MgCl_2$

RNA Wash buffer- 50mM Hepes-HCL, pH 7.5; 250mM NaCl; 0.5% (v/v) Np-40; 10mM$MgCl_2$

Protein incubation buffer- 10mM Tris-HCL, Ph 7.5; 150mM KCl; 1.5mM $MgCl_2$; 0.1% (v/v) Np-40; 0.5mM DTT, Complete Protease Inhibitor

Elution Buffer- 50mM Tris-HCL pH 8.5; 2M Urea; 10mM DTT

2.3.6 RNA binding affinity buffers

Exchange buffer- 20mM HEPES (pH 7.4), 220mM NaCl, 1mM DTT

MBP binding protein- 200mM NaCl, 20mM Tris-HCl, 0.05% NP-40 (pH 7.4)

Selection buffer- 200mM NaCl, 20mM Tris-HCl, 0.05% NP-40, 100ug/mL Salmon sperm DNA (pH 7.4)

2.3.7 Western blot buffers

Transfer buffer- 25mM Tris, 192mM Glycine, 20% Methanol, 0.01% SDS.

2.3.8 TLC Running buffers

First dimension buffer- 5:3, Isobutyric acid:0.5M NH_4OH

Second dimension buffer- 70:15:15, Isopropanol:HCL:Water

2.3.9 Stock Solutions

Stock solution used to make up buffers are detailed below:

50mM Ammonium Hydrogen carbonate (ABC) buffer- Dissolve 1.58g in 400ml sterile distilled water.

Ampicillin (100 mg/mL)- Dissolve 1g Ampicillin in 10ml water, filter sterilise and store at -20°c.

Cefotaxime (500mg/mL) Dissolve 5g Cefotaxime in 10mL water, filter sterilse and and store at -20°c.

Chloramphenicol (25 mg/mL)- Dissolve 0.26g in 10ml Ethanol, filter sterilise and store at -20°c.

100mM $CaCl_2$- Dissolve 1.11g of $CaCl_2$ into 100ml of water. Autoclave to sterilise.

1M DTT- Dissolve 1.542g DTT in 10ml of 50mM ABC Buffer, filter sterilise and store at -20°c.

100mM EDTA pH8- Dissolve 29.224g EDTA in 400ml, adjust to pH8 with NaOH pellets. Make up to 500ml, autoclave to sterilise.

500mM EDTA pH8- Dissolve 93.05g EDTA in 400ml, adjust to pH8with NaOH pellets. Make up to 500ml, autoclave to sterilise.

50% Glycerol- Mix 20ml of Glycerol and 20ml of water, place on rollers till mixed. Filter sterilise.

0.5M Hepes-HCL pH 7.5- Dissolve 23.83g of Hepes in 60ml of distilled water, adjust pH to 7.5 with NaOH. Make up to 100ml with distilled water. Filter sterilise and store at -20°c

0.5M IAA- Dissolve 0.925g IAA in 10 ml of 50mM ABC buffer, filter sterilise and store at -20°c.

0.1M IPTG- Dissolve 23.83g of IPTG in 100ml of water. Filter sterilise and store at -20°c.

Kanamycin (50mg/ml)- Dissolve 0.5g of Kanamycin in 10ml of water. Filter sterilise and store in freezer

1M KCl- Dissolve 7.45g of KCl in 100ml of distilled water. Autoclave to sterilise

3M KOAc pH4.8- Dissolve 9.81g of KOAc in 60ml of water, adject pH using Glacial Acetic Acid to pH 4.8. Autoclave to sterilise.

2M LiCl- Dissolve 8.46g of LiCl in 100ml of distilled water. Autoclave to sterilise.

200mM $MgCl_2$- Dissolve 1.9g of $MgCl_2$ in 100ml of distilled water. Autoclave to sterilise.

1M $MgS0_4$- Dissolve 12.03g of MgSO4 in 100ml of distilled water. Autoclave to sterilise.

20mM $MnCl_2$- Dissolve 2.51g of $MnCl_2$ in 100ml of distilled water. Autoclave to sterilise.

2M NaCl- Dissolve 11.68 of NaCl in 100ml of distilled water. Autoclave to sterilise.

5M NaCl- Dissolve 29.22g of NaCl in 100ml of distilled water. Autoclave to sterilise.

3M NaOAc pH5.3- Dissolve 24.61g of NaOAc in 60ml of water, adjust pH using glacial Acetic Acid to pH 4.8. Autoclave to sterilise.

100mM NaOAc pH5.3- Dissolve 0.82g of NaOAc in 60ml of water, adjust pH using glacial Acetic Acid to pH5.3. Autoclave to sterilise.

0.1N NaOH- Dissolve 1g NaOH in 250ml.Filter sterilise.

10% NP40- Mix 1ml of NP40 with 9ml of water. Mix until dissolved, filter sterilise.

1x PBS- Put 1 tablet in 500ml water. Autoclave to sterilise.

10% Sarkosyl solution- Dissolve 1g in 10ml of distilled water. Filter sterilise.

10% SDS- Mix 10ml of SDS into 90ml distilled water. Filter sterilise.

Spectinomycin (100mg/mL) Dissolve 1g Spectinomycin in 10ml water, filter sterilise and store at -20°c.

1x TBE- Mix 100ml of 10x TBE into 900ml distilled water

Tetracycline (12.5mg/ml)- Dissolve 0.125g tetracycline in 10ml distilled water. Filter sterilise.

1M Tris-HCL- Dissolve 60.5g of Tris base in 400ml of distilled water. Adjust the PH accordingly (ph7.5, 8.5, 9.5) with HCL. Autoclave to sterilise.

2.4 DNA manipulation and amplification

2.4.1 Plasmids

All plasmids used in this study are listed in Table 2.2.

Table 2.2: List of plasmids used in this study. Name, functional description and sources labelled.

Plasmid	Description	Reference/Source
pICH86966	Derived from pUC19 and pPZP200. LacZ alpha and KanR selectable markers. Golden gate vector system. Template for PCR amplification of AtU6p::sgRNA.	Weber et al., 2011
pICSL01009	Derived from pUC19 and pPZP200. LacZ alpha and SpecR selectable markers. Golden gate vector system. Level 1 vector AtU6p module.	Weber et al., 2011
pICH47751	Derived from pUC19 and pBIN19. LacZ alpha and AmpR selectable markers. Golden gate vector system. Level 1 destination vector.	Weber et al., 2011
pAGM4723	Unspecified derivation. Canthaxanthin operon and KanR selectable markers. Golden gate	Weber et al., 2011

	vector system. Level 2 destination vector.	
pICH47732	Derived from pUC19 and pBIN19. LacZ alpha and AmpR selectable markers. Golden gate vector system. Level 1 NOSp::NPTII-OCST module	Weber et al., 2011
pICH47742	Derived from pUC19 and pBIN19. LacZ alpha and AmpR selectable markers. Golden gate vector system. Level 1 35sp::Cas9-NOST module	Weber et al., 2011
pICH41766	Derived from pUC19. SpecR selectable marker. Golden gate vector system. Level 2 linker module	Weber et al., 2011
pCGECT1-NPTII-Cas9-sgRNA	Derived from completed ligation of above modules. For Agrobacterium mediated expression of CRISPR-Cas9 disruption in ECT1 (Solyc01g028860.2.1.) KanR selectable marker	This study
pCGECT2-NPTII-Cas9-sgRNA	Derived from completed ligation of above modules. For Agrobacterium mediated expression of CRISPR-Cas9 disruption in ECT2 (Solyc05g032850.2.1.) KanR selectable marker	This study
pCGECT3-NPTII-Cas9-sgRNA	Derived from completed ligation of above modules. For Agrobacterium	This study

	mediated expression of CRISPR-Cas9 disruption in ECT3 (Solyc12g099090.1.1.) KanR selectable marker	
pCGECT4-NPTII-Cas9-sgRNA	Derived from completed ligation of above modules. For Agrobacterium mediated expression of CRISPR-Cas9 disruption in ECT4 (Solyc01g103540.2.1.) KanR selectable marker	This study
pJET 1.2	Positive selection cloning vector. AmpR and eco47IR selectable markers. T7 promoter for in-vivo or in-vitro transcription. Used for cloning sequences for amplification or sequencing.	Thermo Fisher Scientific (Thermo Fisher Scientific, Altrincham, England)
pCR8-GW-TOPO	TA cloning gateway entry vector. SpecR selectable marker. *Att* sites for gateway recombination. EcoR1 insert excision. Used for gatway recombination into gateway destination vectors.	Thermo Fisher Scientific (Thermo Fisher Scientific, Altrincham, England)
pGKPGWG	Gateway destination vector for constituative expression of GFP fusions. KanR and CmR selectable markers. 35s promoter. GFP	Zhong et al., 2008
pGKGWG	Gateway destination vector for cloned promoter expression of GFP fusions. KanR and CmR selectable markers. No promoter. GFP	Zhong et al., 2008

pGWB3	Gateway destination vector for C-GUS fusions. KanR and HyR selectable markers. No promoter. GUS	This study
pGKP-cECT1-G	*In Planta* Constituative expression vector for the ubiquitious expression of ECT1-GFP (Solyc01g028860.2.1) fusion protein. 35s promoter. KanR resistance. NPTII. eGFP	This study
pGKP-cECT2-G	*In Planta* Constituative expression vector for the ubiquitious expression of ECT2-GFP (Solyc05g032850.2.1) fusion protein. 35s promoter. KanR resistance. NPTII. eGFP	This study
pGKP-cECT3-G	*In Planta*_Constituative expression vector for the ubiquitious expression of ECT3-GFP (Solyc12g099090.1.1) fusion protein. 35s promoter. KanR resistance. NPTII. eGFP	This study
pGKP-cECT4-G	*In Planta* Constituative expression vector for the ubiquitious expression of ECT4-GFP (Solyc01g103540.2.1.) fusion protein. 35s promoter. KanR resistance. NPTII. eGFP	This study
pGW-pECT1-B3	*In Planta* expression vector for ECT1 (Solyc01g028860.2.1.) promoter	This study

	driven C-terminal GUS expression. KanR and HyR selectable markers.	
pGW-pECT2-B3	*In Planta* expression vector for ECT2 (Solyc05g032850.2.1.) promoter driven C-terminal GUS expression. KanR and HyR selectable markers.	This study
pGW-pECT3-B3	*In Planta* expression vector for ECT3 (Solyc12g099090.1.1.) promoter driven C-terminal GUS expression. KanR and HyR selectable markers.	This study
pGW-pECT4-B3	*In Planta* expression vector for ECT4 (Solyc01g103540.2.1.) promoter driven C-terminal GUS expression. KanR and HyR selectable markers.	This study
pMAL-c5x	Protein expression vector for the production of maltose-binding protein fusions. Cleaved via Factor Xa. XmnI site insertion. AmpR selectable marker.	New England Biolabs (Hitchin, England)
pMAL-c5x-cECT1	Protein expression vector for the production of ECT1-maltose-binding protein fusions. IPTG inducable promoter AmpR selectable marker.	This study
pMAL-c5x-cECT2	Protein expression vector for the production of ECT2-maltose-binding protein fusions. IPTG inducable promoter AmpR selectable marker.	This study

pMAL-c5x-cECT3	Protein expression vector for the production of ECT3-maltose-binding protein fusions. IPTG inducable promoter AmpR selectable marker.	This study
pMAL-c5x-cECT4	Protein expression vector for the production of ECT4-maltose-binding protein fusions. IPTG inducable promoter AmpR selectable marker.	This study

2.4.2 Isolation of plasmid DNA

Plasmid DNA isolation was performed using GeneJET plasmid miniprep kit (Thermo Fisher Scientific) in accordance with manufacturers protocol. 5mL cultures were centrifuged at 8000g and plasmid DNA extracted.

2.3.3 Oligonucleotide Primers

All primers used in this study are listed in Table 2.3.

Table 2.3: List of oligonucleotide primers used in this study. Name, functional description and sources labelled.

Primer	Sequence (5'-3')	Function
TomECT1	TGTGGTCTCAATTGTGTTCGTTCTCCTATTCCAGTTTTAGAGCTAGAAATAGCAAG	Amplification of ECT1 SgRNA guide from pICH86966
TomECT2	TGTGGTCTCAATTGCTTTCAGGAAGCAGCAACAGTTTTAGAGCTAGAAATAGCAAG	Amplification of ECT2 SgRNA guide from pICH86966
TomECT3	TGTGGTCTCAATTGGTGCCGTAGTAGGAGCTGAGTTTTAGAGCTAGAAATAGCAAG	Amplification of ECT3 SgRNA guide from pICH86966
TomECT4	TGTGGTCTCAATTGGCTATAATCCTTACAGTGCGTTTTAGAGCTAGAAATAGCAAG	Amplification of ECT4 SgRNA guide from pICH86966
Cas9-F	TACGCCGGATACATTGACGG	Screening for presence of Cas9 t-DNA insertion
Cas9-R	CGGTTGATCTATCTCGCGCT	
pICH47751-pcr	GGAGATCCTGATCTGATCTGACTGA	Level 1-CRISPR construct colony PCR
CRISPrev	TGTGGTCTCAAGCGTAATGCCAACTTTGTAC	Level 1 and Level 2 CRISPR construct colony PCR-Downstream of guide
NPTII-F	CTTCTGAGCGGGACTCTGGG	Level 2-CRISPR construct colony PCR. 3' end of NPTII module

Cas-R	GCCTTTCCTTTATCGCAATG	Level 2-CRISPR construct colony PCR. 5' end of Cas9 module
CRISPRseq-Fwd2	GCCCATCAGGGAGCAGGCAG	CRISPR level 2 construct guide sequencing
CRISPRseq-Rev2	GCAGCCGCACCATTTGTATGTG	
pICH47751-seq-R	CCGCCAATATATCCTGTCAAAC	CRISPR level 1 construct guide sequencing
ECT1-screen-Fwd	CTTCGCCTGTCCCTACTGTG	ECT1 mutation genotyping screen and sequencing-
ECT1-screen-Rev	GACCTGTAGCGCGAGTATCA	
ECT2-screen-Fwd	CTTGCCCCATGTTACATCAG	ECT2 mutation genotyping screen and sequencing
ECT2-screen-Rev	GGATGAGTCCAGGGCTACATTGC	
ECT3-screen-Fwd	AGCCCCTGAGTTTGTTGTTGA	ECT3 genome screen and sequencing
ECT3-screen-Rev	ATCGGATCTGCAGCATTCTC	
ECT4-screen-Fwd	CTGTGGACTTGTATGTGTTTGCC	ECT4 mutation genotyping screen and sequencing
ECT4-screen-Rev	AAGGCTGAGCTTGATGGAAAG	
ECT2 3.24-34BP-fwd	GCTTTCAGGAAAGTTGGATCTG	ECT2 3.24.4 PCR screen
ECT2 34BP DEL 2	GTGAGCTTTCAGGAAAGTTG	Used with ECT2-screen-Rev
ECT2 3.24-6BP-fwd	GTGAGCTTTCAGGAAACATGG	ECT2 3.24.3 PCR screen Used with ECT2-screen-Rev

ECT1 guide seq F	GTCATGTGAATGGTGGTGTAAAG	CRISPR ECT1 sgRNA guide sequencing
ECT2 guide seq F	CCCATGTTACATCAGCTCTTCC	CRISPR ECT2 sgRNA guide sequencing
ECT3 guide seq F	TTTCCCTTCATTTCAGGATGC	CRISPR ECT3 sgRNA guide sequencing
ECT4 guide seq F	AATAACCCGTTCCATGATAGTG	CRISPR ECT4 sgRNA guide sequencing
cdECT1fwd	ATGGCTGCTGGTGCTCCTC	ECT1 transcript amplification
cdECT1rev	GCATCCATTAGAAGCCCCAT	
cdECT2fwd	ATGGAGGATACCAATCAGCAGAG	ECT2 transcript amplification
cdECT2rev	AGGCAACTTCTTGTTGCCTT	
cdECT3fwd	ATGTATAATGAAGGAGCCCCTG	ECT3 transcript amplification
cdECT3rev	CTTAAGGACCTTTTTTCCAGAC	
cdECT4fwd	ATGGCTGGTGAGAAGATTATAGAA	ECT4 transcript amplification
ECT4 cDNA Fwd 2	ATGGCTGGTGAGAAGATTAT	
cdECT4rev	GACACTACTTTTGAATGGCCTTG	
ECT1-cDNA-INT	ATGGCTCTTATGGACCTGGC	ECT1 transcript amplification-High percentage PCR check
ECT2-cDNA-INT	GGCTTGCCCCATGTTACATC	ECT2 transcript amplification-High percentage PCR check
ECT3-cDNA-INT	AAGATCAATGCTGCATCGG	ECT3 transcript amplification-High percentage PCR check
ECT4-INT-F	GGGTTTCATCTGGTTGGGGA	ECT4 transcript amplification-High percentage PCR check

ECT4-INT-R	GTGGGAAGAACCCACTCCAC	ECT4 transcript amplification-High percentage PCR check
ACTIN-F	GTGGCTCCTTGACTACGAGCA	ACTIN cDNA positive control
ACTIN-R	CCGGGAAACAGACAGGACAC	
ECT1-His-R	TTAATGGTGATGGTGATGGTGGCATCCATTAGAAGCCCCAT	Addition of Histadine tag to 3' end of ECT1 cDNA
ECT2-His-R	TCAATGGTGATGGTGATGGTGAGGCAACTTCTTGTTGCCTT	Addition of Histadine tag to 3' end of ECT2 cDNA
ECT3-His-R	TCAATGGTGATGGTGATGGTGCTTAAGGACCTTTTTTCCAGAC	Addition of Histadine tag to 3' end of ECT3 cDNA
ECT4-His-R	TCAATGGTGATGGTGATGGTGGACACTACTTTTGAATGGCC	Addition of Histadine tag to 3' end of ECT4 cDNA
ECT1Gibson-F	ACCTCGGGATCGAGGGAAGGATGGCTGCTGGGTGCTCCT	Overhangs for gibson cloning
ECT1Gibson-R	CCCATGGACATATGTGAAATTTAATGGTGATGGTGATGGTGGC	
ECT2Gibson-F	ACCTCGGGATCGAGGGAAGGATGGAGGATACCAATCAG	
ECT2Gibson-R	CCCATGGACATATGTGAAATT	
ECT3Gibson-F	ACCTCGGGATCGAGGGAAGGATGG	
ECT3Gibson-R	CCCATGGACATATGTGAAATT	

ECT4Gibson-F	ACCTCGGGATCGAGGGAAGGATGG	
ECT4Gibson-R	CCCATGGACATATGTGAAATT	
ECT4 cDNA Gibson Fwd	ACCTCGGGATCGAGGGAAGGATGGATGGCTGGTGAGAAGATTAT	
CG2_ect4_F	TCGGGATCGAGGGAAGGATTTCACATATGGCTGGTGAGAAGATTATAG	
CG2_ect4_R	GAAGCTTATTTAATTACCTGCAGGGAATTCGTGATGGTGATGGTGATGG	
Cas9-seq-1	GCTGCCCCCGAGAATTATGC	Cas9 module sequencing
Cas9-seq-2	CATTTCATTTGGAGAGGACA	
Cas9-seq-3	GAACGGCCTGTTTGGTAATC	
Cas9-seq-4	AAGGTGCTTCCTAAACACTC	
Cas9-seq-5	CCCGAGAATATCGTTATCGAGA	
Cas9-seq-6	GATGCCTACCTGAATGCAGTGG	
Cas9-seq-7	CGAGCTGCAGAAAGGTAACG	
GW1	GTTGCAACAAATTGATGAGCAATGC	Sequencing pCR8-GW-TOPO inserts
GW2	GTTGCAACAAATTGATGAGCAATTA	
GFP F	AACTCCAGCAGGACCATGTG	GFP tDNA screen
GFP R	CGTAAACGGCCACAAGTTCAG	

ECT1 XbaI Fwd 3	CTAGTCTAGATCCGAGCTAG ATCTTCAACAG	ECT1 promoter amplification- Addition of Xbal restriction site for cloning into pGKGWG and pGWB3
ECT1 SacI Rev 3	CTAGGAGCTCGAAACAGAG AAGTGAGTTTCTC	ECT1 promoter amplification Addition of SacI restriction site for cloning into pGKGWG and pGWB3
ECT2 HindIII Fwd 3	CTAGAAGCTTCAATTGAACT GAATTGAAAATG	ECT2 promoter amplification Addition of HindIII restriction site for cloning into pGKGWG and pGWB3
ECT2 XbaI Rev 3	CTAGTCTAGACTAAATAGCA AACCCACAA	ECT2 promoter amplification Addition of Xbal restriction site for cloning into pGKGWG and pGWB3
ECT3 HindIII Fwd 3	CTAGAAGCTTAGCTTGAGAG TTCTGAAGCAGGG	ECT3 promoter amplification Addition of HindIII restriction site for cloning into pGKGWG and pGWB3
ECT3 XbaI Rev 3	CTAGTCTAGAGGGTTGAGG GGTGTGGTAGTTG	ECT3 promoter amplification Addition of Xbal restriction site for cloning into pGKGWG and pGWB3
ECT4 SpeI Fwd 3	CTAGACTAGTTATCAAGCTT CAAAAGCCAAT	ECT4 promoter amplification Addition of SpeI restriction site for cloning into pGKGWG and pGWB3
ECT4 XbaI Rev 3	CTAGTCTAGAGAACATTTTT CTGTTATTG	ECT4 promoter amplification Addition of Xbal restriction site for cloning into pGKGWG and pGWB3
pGKPGWG Start	ATCGGAAATCCAGAAGCCC GAG	pGKPGWG sequencing

pGKPGWG FOR1	TGCGTGATCTGATCCTTCAA
pGKPGWG FOR2	CTCATGGTGATTTCTCACTT GAT
pGKPGWG FOR3	GGAGAAAGGCGGACAGGTA T
pGKPGWG FOR4	TCCCATGGAGTCAAAAATTC A
pGKPGWG FOR5	GTGACACCTGTGTTGCCAGA
pGKPGWG FOR6	TCTTCAAGTCCGCCATGC
pGKPGWG FOR7	AAGCCTGTTACGAACAGCA GA
pGKPGWG FOR8	TCTTCAAGTCCGCCATGC
pGKPGWG FOR9	CGTATTACAATTCACTGGCC G
pGKPGWG FOR10	GAAGGGACTGGCTGCTATT G
pGKPGWG FOR11	CGCGATAGAAAACAAAATAT AGCG

2.3.4 Isolation of plant genomic DNA

DNA extraction protocol modified from the method described by Shure et al. (1983). Fresh young leaf tissue was ground to a fine powder either in liquid nitrogen using a prechilled mortar and pestle or freeze died and lysed using a QIAGEN TissueLyser II. Lysed tissue transferred to a sterile, capped centrifuge tube and 700 µL of Edwards buffer (200mM Tris-HCl- Ph 7.5, 200mM NaCl, 25mM EDTA, 0.5% SDS) was added per 1g tissue. Powder and buffer vortexed, before centrifugation at max speed for 10 minutes.

Supernatant was collected and transferred to a fresh Eppendorf tube. Extracted supernatant was mixed 1:1 with isopropanol and precipitated overnight at -20°c. After precipitation, samples were centrifuged at maximum speed for 10 minutes. Supernatant is collected, without disturbing pellet, and discarded. Pellets are washed with 70% and dried in a vacuum for 2-4 minutes, before resuspending in sterile dH2O. Stored at -20°c. When needed, high-purity DNAs were extracted using GenElute Plant Genomic DNA Miniprep Kit (Sigma-Aldrich) according to the manufacturer's instructions.

2.3.5 Polymerase chain reaction (PCR)

Oligonucleotide primers from Sigma-Aldrich and Eurofins Scientific were used unmodified for PCR amplification. Primers were provided lyophilised and resuspended in nuclease-free H_2O (Ambion) to a final concentration of 100 pmol/µL.

PCRs were performed using Standard Taq Polymerase, Q5® DNA Polymerase, Phusion® DNA Polymerase (NEB) following manufacturers' protocols. Standard Taq Polymerase was used for screening of clones and plant genomic DNA. Phusion and Q5 DNA Polymerases were used for amplification of sequences for cloning. Annealing temperatures were calculated using NEB Tm calculator (https://tmcalculator.neb.com/#!/main). PCR amplifications were carried out in Eppendorf 5332 Mastercycler Gradient Thermal Cyclers.

2.3.6 Restriction analysis of DNA

Restriction enzymes used in this study were used according to manufacturer's protocols. Reactions generally contained 0.100 – 1 µg DNA, 0.5 Units of each enzyme (approximately 1 µL), 2 µL of 10x Buffer and H_2O up to 20 µL. The digestions were performed at 37 °C for 1.5 h and analysed by agarose gel electrophoresis. (Unit definition: One unit is defined as the amount of enzyme required to digest 1 µg of λ DNA in 1 h at 37 °C in a total reaction volume of 50 µL).

2.3.7 A-tailing PCR products

A tailing of PCR products entails the addition of a single overhanging A at both the 5' and 3' ends of each PCR product. This allows cloning into TA cloning vectors. This was achieved by producing the following reaction mix on ice:10µL purified PCr product, 5.25 µL dH_2O, 2 µL Taq polymerase buffer, 0.5 µL $MgCl_2$, 1.25 µL 1mM dATPs, 1 µL Taq DNA Polymerase. This was incubated at 73°C for 30 minutes. This was then used in further downstream reactions.

2.3.8 Agarose gel electrophoresis

DNA electrophoresis was performed in 0.7-3% (w/v) agarose gels (with 5µL/100mL) cast with combs to produce wells. 5µL Hyperladder 1kb (0.2-10kb, Bioline) was used as a DNA molecular weight marker. Up to 20µL was pipetted in each well. Gels were run in 1X TBE (section 2.3.9) at 100V for approximately 1 hour and visualised using a Gel Doc ™ XR+(Bio-Rad).

2.3.9 Purification of PCR and DNA products

2.3.9.1 Purification of DNA and PCR products via kit extraction

For clean, purified PCR products extractions were performed using the Monarch® PCR & DNA Cleanup Kit (5 µg) (New England Biolabs), as per manufacturer protocols. This was only carried out if only a single band was present on agarose gel.

2.3.9.2 Purification of PCR products via gel extraction

PCR amplicons and relevant digestion bands were excised from 1% (w/v) agarose gels when needed and recovered with the Monarch® DNA Gel Extraction kit following manufacturer's protocol.

2.3.10 DNA quantification

DNA concentration and purity was assessed using the Thermo Fisher Scientific Nanodrop 2000 Spectrophotometer.

2.3.11 Golden gate ligation assembly

The assembly of all CRISPR disruption constructs was achieved via Golden gate ligation assembly. Concentrations of components were optimised according to (Engler, Kandzia and Marillonnet, 2008; Weber *et al.*, 2011; Mali *et al.*, 2013). Use of specific restriction enzymes was dependent on functional modules.

2.3.12 HiFi and Gibson DNA assembly

The assembly of certain constructs was achieved using HiFi DNA Assembly or Gibson assembly (NEB) according to manufacturer's protocol. As recommended, a total of 0.03-0.2 pmols of DNA fragments was used during the assembly of 1 or 2 fragments. No constructs required 4-6 fragments. Optimisation of assembly conditions was calculated based on fragment length and molecular weight. This was performed using the NEBiocalculator tool (https://nebiocalculator.neb.com/#!/ligation) to calculate the following formula:

$$Pmols = \frac{(molecular\ weight\ in\ ng) x\ 1000}{base\ pairs\ x\ 650\ daltons}$$

As suggested, DH5-α competent *E. coli* cells were transformed with 2-5 µL of the assembled product, following the transformation protocol (section 2.4.1).

2.3.13 Sequencing of DNA

DNA sequencing was performed by the Eurofins MWG Custom DNA Tube Sequencing Service using the cycle sequencing technology (dideoxy chain termination / cycle sequencing) on ABI 3730XL sequencing machines.

2.4 Transformation of bacterial strains

2.4.1 Transformation of *Escherichia coli* chemically competent cells

Cloning reactions were transformed to competent *E. coli* cells as follows. 10 µl (or less, depending on the reaction volume) of the cloning reaction was added to a vial of chemically competent *E. coli* cells (DH5α or TOP10) and mixed gently. Mixtures were incubated on ice for 30 min and then heat-shocked for 30 seconds at exactly 42°C. The vials were immediately transferred and incubated on ice for 3 min. 700 µl of liquid LB medium was added into each vial and transferred into a 1.5 ml Eppendorf tube. The tube was incubated in a 37 °C shaker (200 rpm) for 1 h and 30 min. 150 µl or more (depending on the transformation efficiency) of *E. coli* culture was spread on solid LB plates containing the appropriate antibiotic. Low copy number plasmids were pelleted before resuspension in 150 µl of LB and plated. Plates were incubated upside down at 37°C overnight. For long term storage: Aliquots of 500 µL were prepared in 1.5 mL 1:1 Eppendorf tubes with 50% glycerol and stored at -80 °C.

2.4.2 Electroporation of *Agrobacterium tumefaciens* electro-competent cells

30ng of purified plasmid DNA was added to a 50 µL aliquot of electrocompetent *A.tumefaciens* cells. Once mixed, cells were transferred into pre-chilled 2 mm electroporation cuvettes (Scientific Laboratory Supplies, SLS). (BioRad MicroPulser- Agr preset). Samples were then incubated at 28°C at 200rmp for 2-3 hours. Incubated cells were then spread (150-300µL) on selective LB agar plates and incubated overnight at 28°C.

2.4.3 Preparation of *E.coli* chemically competent cells

A 10mL culture of LB was inoculated with a heavy loop of fresh, plate grown *E.coli* cells and vortexed for resuspension. The culture was incubated at 37°C and 200rpm overnight (12-16 hours). 100mL of SOC media was inoculated with 10mL of initial overnight. The culture was then incubated at 37°C and 200rpm to an OD_{600} of 0.4-0.6. Once the correct OD_{600} was reached the culture was incubated in a sterile, pre-chilled conical flask and incubated on ice for 15 minutes. Cultures were then transferred to pre-chilled 50mL falcon tubes and centrifuged at 3000 rpm for 12 minutes at 4°C in a refrigerated centrifuge. The supernatants were decanted, and the pellets were resuspended in 1/3 pellet volume of RF1 buffer (100mM RbCl, 50mM $MnCl_2$, 30mM Potassium acetate, 10mM $CaCl_2.2H_2O$, 15% w/v Glycerol: Adjusted to Ph 5.8- 0.2M Acetic acid). Cells were then chilled for 15 minutes on ice before centrifugation as before. The supernatant was decanted, and pellets were resuspended in 1/12.5 original media volume of RF2 buffer (10mM MOPS, 10mM RbCl, 75mM $CaCl_2.2H_2O$, 15% w/v Glycerol). Cells were then prepared in aliquots of 100 µL, flash frozen in liquid nitrogen and stored at -80°C.

2.3.4 Preparation of *A.tumefaciens* electrocompetent cells

A 10mL culture of LB was inoculated with a heavy loop of fresh, plate grown *A.tumefaciens* cells and vortexed for resuspension. The culture was incubated at 28°C and 200rpm overnight (16-24 hours). The following day, 150-300mL of LB was inoculated with 2-4mL of initial overnight culture into a baffled flask and grown at 28°C and 200rpm to an OD_{600} of 0.5-1.0. Once the correct OD_{600} was reached the culture was incubated in a sterile, pre-chilled conical flask and incubated on ice for 30 minutes. Cultures were then transferred to pre-chilled 50mL falcon tubes and centrifuged at 5000 rpm for 9 minutes at 4°C in a refrigerated centrifuge. The supernatants were decanted, and the pellets were resuspended in 50mL of pre-chilled sterile dH_2O. Centrifugation was repeated before decanting the supernatant and resuspending in 25mL of pre-chilled sterile dH_2O. Cells were centrifuged at >6000rpm for 10 minutes at 4°C in a refrigerated centrifuge. Supernatants were decanted and pellets were resuspended in 10mL of pre-chilled sterile

10% glycerol. This step was repeated an additional two times before resuspending the final pellet in 1mL pre-chilled sterile 10% glycerol. Each of the 1mL competent cell cultures were divided into 0.5mL Eppendorf tubes in 50 µL aliquots, immediately flash frozen in liquid nitrogen and stored at -80°C.

2.5 Plasmid construction

2.5.1 Construction of CRISPR-Cas9 disruption constructs

2.5.1.1 Selection of CRISPR disruption targets

Genes of interest were screened for 20bp sequences ending in a PAM (NGG) site following (Mali *et al.*, 2013) GNNNNNNNNNNNNNNNNNNNGG. Any potential target sequences were checked for position completely within exons. Specific target amplification primers were then designed for PCR amplification.

2.5.1.2 Assembly of level 1 CRISPR-Cas9 disruption constructs.

Gene specific sgRNA guide fragments were amplified from pICH86966 using specifically designed amplification primers (Table 3.) as described in section 2.3.5. The PCR products were separated and purified, as described in section 2.3.9.2. Concentration of extracted DNA was quantified by direct comparison to the DNA marker using a Gel Doc ™ XR+and imagelab 6.0 (BioRad), before diluting to 20ng/µL. 20ng/µL of each PCR product was used in a BsaI golden gate reaction with pICH47751 and pICSL01009, following the protocol and concentration described in section 2.3.11.. Ligated level 1 plasmids were transformed into DH5α as described in section 2.4.1, on solid plates containing 100mg/L Ampicilin, 200mg/L X-gal and 1mM IPTG. White (not blue) colonies were selected and screened for correct assembly.

2.5.1.3 Assembly of level 2 CRISPR-Cas9 disruption constructs

Separate BpiI golden gate reactions with each screened level 1 construct (pICH47751-ECTx-sgRNA), pICH47732, pICH47742, pICH41766 and ligated into pAGM4723 were performed as described in section 2.3.11. Ligated products were transformed into DH5α as described in section 2.4.1, on solid plates containing 50mg/L Kanamycin. White (not red) colonies were selected and screened for correct assembly. Colonies were subsequently sequenced as described in section 2.3.13 to ensure correct assembly.

2.5.2 Construction of constitutive expression constructs

Prior to cloning, cDNA was produced from AC++ whole RNA as described in section 2.6.1. Full length ECT1-4 cDNA was amplified via PCR (section 2.3.5) using transcript specific primers starting from the ATG (Table 3). Amplified cDNA PCR products were purified as described in section 2.3.9.1. Purified PCR products were then A-tailed as described in section 2.3.7, before being transformed into pCR8-GW-TOPO entry vector (Table 2) following the manufacturer's protocol. Colonies were screen for correct orientation of insert before recombining into pGKPGWG using Gateway LR clonase II enzyme mix, following manufacturer's protocols. Recombined plasmids were transformed into DH5α as described in section 2.4.1, on solid plates containing 50mg/L Kanamycin. Colonies were sequenced as described in section 2.3.13 to ensure correct assembly.

2.5.3 Construction of ECT promoter GUS expression constructs

Prior to cloning, promoter regions of ECT1-4 were amplified using gene specific primers 2000bp upstream of the first open reading frame. Amplified promoters were cloned into pCR8-GW-TOPO as described in above section 2.5.2. Using the entry vector pCR8-GW-TOPO, promoter regions were recombined into pGWB3 as described above.

2.5.4 Construction of ECT1-3 protein MBP-fusion constructs.

Full length cDNA of ECT1-4 was amplified from total cDNA as described in section 2.5.2. A Histidine-tag consisting of 6 sequential histidine codons were added to the 3' end of each cDNA, using overhanging primers (Table 3)

using PCR (2.3.5) and purified (2.3.9.1). Each His-tagged cDNA was reamplified with specifically designed Gibson overhang primers (Table 3) and purified (2.3.9.2). pMAL-c5x was digested using XmnI (section 2.3.6) before purifying the linearised plasmid using a kit (2.3.9.1). Gibson assembly reactions were then performed into pMAL-c5x as described in section 2.3.12. Assembled plasmids were transformed into DH5α as described in section 2.4.1, on solid plates containing 100mg/L Ampicillin. Then screened by colony PCR before sequencing to confirm correct assembly (2.3.13).

2.5.5 Construction of ECT4 protein MBP-fusion constructs.

Specific difficulty was had in assembly of ECT4 cDNA into pMAL-c5x as described above. In this case ECT cDNA assembly into pMAL-c5x was carried out by HiFi assembly via EcoRI and NdeI (2.3.12). Transformed into DH5α (section 2.4.1), on solid plates containing 100mg/L Ampicillin. Screening of correct assembly of constructs was performed by restriction digest (2.3.6) using KpnI and XhoI, before sequencing to confirm assembly (2.3.13).

2.6 RNA manipulation

2.6.1 Total RNA extraction

Total RNA was prepared following the method of total RNA extraction from yeast with minor modifications (Schmitt et al., 1990). Plant samples were ground into powder using chilled mortar and pestle and liquid nitrogen. Ground samples were divided into several tubes. In each tube, 400 µl of AE buffer (50 mM sodium acetate pH 5.2, 10 mM EDTA), 40 µl of 10% (w/v) SDS and 400 µl of AE saturated phenol were added. Tubes were vortexed and incubated at 65°C for 4 min. Tubes were rapidly chilled in liquid nitrogen for 3 seconds. This was repeated 4-5 times until phenol crystals appear.

Tubes were then centrifuged at maximum speed for 5 min and the aqueous phase was transferred to a new tube. An equal volume of phenol/chloroform was added to the new tube and mixed by vortexing. After centrifuging for 5 min, the aqueous phase was again transferred to a new tube and precipitated by adding 1/10 volume of 3 M sodium acetate (pH 5.2) and 3 volumes of absolute ethanol. The mixture was vortexed and kept at -20°C overnight. The following day, tubes were centrifuged at maximum speed for 10 min and then the RNA pellet was washed with 70% (v/v) ethanol. After drying, the RNA pellet was resuspended in sterile water.

2.6.2 Poly-A purification

Poly-A purification was based on the NEBNext® Poly(A) mRNA Magnetic Isolation Module (NEB). The oligo d(T) beads were bought separately, and the buffers made up in the lab according to the NEBNext® manufacturer's instructions. 5µg of total RNA was diluted up to 50µl with sterile distilled water. 20µl of beads were aliquoted into a 1.5ml Eppendorf and left to stand on the magnetic rack for 2 minutes. The supernatant was then discarded and the beads re-suspended in 100µl RNA binding buffer (1M LiCl; 40mM Tris-HCL, pH 7.5; 2mM EDTA; 0.1% NP40). The tube was placed back on the rack and the supernatant removed. This was repeated once more. The beads were re-suspended in 50µl of RNA binding buffer and the 50µl of RNA was added. The tube was placed in a heat block at 65°C for 5 minutes and cooled on ice. The beads were pipetted up and down 6 times before being put on rollers at room temperature for 20 minutes. The tube was put back on the magnetic rack for two minutes. The supernatant was removed which was the –poly(A) RNA. The beads were washed twice with 200µl wash buffer (150mM LiCl; 20mM Tris-HCL, pH 7.5; 1mM EDTA; 0.01% NP40). 50µl of pre heated 80°C water was added, and the beads were resuspended and incubated at 80°C for two minutes. The tube was cooled at room temperature. 50µl of RNA binding buffer was added and the beads were pipetted up and down six times and incubated on the rollers for 15 minutes at room temperature. The supernatant was discarded, and the beads were washed with 200µl of wash buffer. The beads were resuspended in 10µl of

pre warmed 80°C water, incubated for 2 minutes at 80°C, and placed on to the magnetic rack. The supernatant containing the poly(A) RNA was transferred into a clean tube.

2.6.2.1 Bead regeneration

The oligo d(T) beads were regenerated so they could be used again. Beads were washed in 200µl 0.1N NaOH on the rollers for 5 minutes. The beads were washed with distilled water until a neutral PH was achieved; and then equilibrated and stored in PBS Tween buffer.

2.6.3 Reverse transcription

Reverse transcription was performed with Protoscript II reverse transcriptase (Thermo Fischer scientific) and both 100ng-1 µg or total RNA or 100ng of poly-A RNA, according to manufacturer protocols. 4 µL of resulting cDNA was used in each downstream reaction or PCR.

2.7 Plant transformation via tissue culture

The following protocols were used for both tomato and tobacco transformation and tissue culture.

2.7.1 Excision and precallusing

Seeds were surfaced sterilised (2.1.3) and sown on sterile MSR3 media (2.1.2.1). After seedlings reached 7-10 days (prior to first true leaves), cotyledons and hypocotyls were excised using a sterile scalpel blade (Bird *et al.*, 1988). Excisions were performed with the seedlings in dH_2O to minimise damage and osmotic water loss. Both the cotyledons and hypocotyls were

placed on MSR3ZA plates (2.1.2.2), later MSR3BA were found to be more efficient (2.1.2.3). Plates were sealed with micropore tape (3M) and placed in growth rooms under standard growing conditions with no light for a minimum of 48 hours, to pre-callus.

2.7.2 Agrobacterial growth and tissue incubation.

After 24 hours of pre-callusing, colonies from *Agrobacterium* strains, containing transformation constructs of interest were inoculated into 5mL of liquid LB media and grown under standard conditions (section 2.1.5). On the day of co-cultivation, 200mL of LB with appropriate antibiotic were inoculated with 5mL of *Agrobacterium* O/N into a baffled flask. Cultures were grown under standard conditions to an OD^{600} of 0.2-0.3. Cultures were then transferred into 50mL falcon tubes and centrifuged at 5000rpm for 5 minutes. The supernatant was decanted and gently resuspended in liquid MS (2.1.2.5).

To inoculate excised plant tissues, cotyledons and hypocotyls were placed in agrobacterium liquid MS cultures and incubated at room temperature for 20 minutes. Tissues were blotted onto sterile filiter paper to remove excess cell culture before replacing the tissues on MSR3ZA/BA plates. The plates were returned to the dark growth room for a further 24-48 hours.

2.7.3 Selection and rooting

Tissues were then placed onto selective MSR3ZA/BA plates containing 50mg/L Kanamycin and 500mg/L Cefotaxime and grown under normal growth conditions (2.1.3). Selection plates were refreshed every 10-14 days, until meristematic shoots developed from the callusing tissues. These were then excised and placed on MSR3 rooting media pots (2.1.2.4) containing 25mg/L Kanamycin. Once branched roots appeared, shots were removed from eh growth rooms and transplanted into

(COMPOST) and placed in bags to maintain humidity. These plants were then treated as normal.

2.7.4 Screening transgenic lines

Presence of t-DNA insertions was performed via PCR of genomic DNA (section 2.3.4) and detection of correct bands on agarose electrophoresis gels.

2.8 Protein purification

2.8.1 Pilot experiments

To check that the expressed protein was soluble and to further optimise soluble protein production pilot experiments were performed according to the protocol from the pMAL protein fusions and purification system manual (NEB). Details are provided in Table 4. For all the pilot experiments *E. coli* was grown at 37°C until the cells reached an OD A_{600} of 0.5 and then protein production was induced with IPTG.

Table 2.4: List of growth conditions used for protein purification pilot studies.

Cells	Media	IPTG (mM)	Temperature (°C)	Induction time
DH5α	SOC	0.3	37	3 hours
DH5α	SOC	0.3	37	2 hours
DH5α	LB	0.15	20	2 hours
SoluBL21	LB	0.4	20	20 hours

SoluBL21	LB + ethanol	2%	0.4	12	20 hours
SoluBL21	LB + ethanol	2%	0.4	8	20 hours

2.8.2 Denaturing polyacrylamide SDS-PAGE gel electrophoresis

Analysis of protein expression and solubility was performed using denaturing SDS-polyacrylamide gel electrophoresis (SDS-PAGE). 10µl of sample was mixed with the same volume of lamelli buffer (- 62.5mM Tris-HCL, pH 6.8; 25% glycerol; 0.01% (w/v) bromophenol blue; 2% SDS; 5% ß-mercaptoethanol) and heated to 90°C for 5 minutes. The samples were loaded on to Biorad TGX precast gels along with 7µl of protein ladder (Color-coded Prestained Protein Marker, Broad Range (11-250 kDa) – Cell signalling technologies). The gels were run for 1 hour at 110V in SDS running buffer (250mM Tris base; 1.92mM glycine, 1% SDS). After running, the gels were washed 3 times in dH_2O for 10 minutes and, stained in page blue or imperial stain for 1hour on a rocker. The stain was removed, and the gel was washed for 60 minutes in dH_2O, changing the dH_2O every 10 minutes. The gel was left in dH_2O overnight before being imaged using a Biorad GS800 scanner.

2.8.3 Protein production and purification

To grow SoluBL21 strains for protein production 1 Litre of LB media was inoculated with a 10ml culture and100mg/L Ampicillin. The culture was grown at 37°C until an OD^{600} of 0.5 and then induced with 2% ethanol and 0.4mM IPTG and grown for 20 hours at 8°C. The cells were pelleted by centrifugation at 4000rpm for 20 minutes. The pellets were then resuspended in equilibration buffer (20mM Na3PO4; 300mM sodium

chloride; 10mM Imidazole; adjust to pH 7.4, do not autoclave imidazole) and stored at -20°C overnight to lyse the cells.

Sonication was carried out following manufacturer guidelines described in section 2.7.1.

All proteins were purified using the gravity flow column method in the HisPUr Ni-NTA resin manual. 1ml of Ni-NTA resin was put into a gravity flow column and storage buffer was allowed to drain out. The resin was washed with 2 bed volumes of equilibration buffer. The protein-equilibration buffer samples were put through the column. The column was washed with two bed volumes of wash buffer (20mM Na_3PO_4; 300mM NaCl; 25mM Imidazole; adjust to pH 7.4, do not autoclave imidazole), and repeated until the absorbance of the flow through at 280nm approached baseline. The first and last wash were collected to run on a protein gel later. The protein was eluted using two volumes of elution buffer, repeated twice. The first 0.5ml was collected separately, this was the last of the column buffer. From each elution a 10µl aliquot was taken to run on a protein gel. The rest was mixed with 50% glycerol and an aliquot of 50µl was removed to detect the protein concentration. The remainder of the elution was frozen in liquid nitrogen. The concentration was detected using Pierce™ BCA Protein Assay Kit, according to manufacturer's instructions.

2.8.4 Further protein purification and buffer exchange

Exchange into appropriate buffer was achieved using Slide-A-Lyzer MINI Dialysis devices (Thermo Fischer scientific) following manufacturer guidelines. Further purification was then performed via size selection using Amicon Ultra-2 Centrifugal Filters (50KDa) (Merck) according to manufacturer protocols.

2.9 Protein-RNA binding affinity assay

All protein samples were exchanged into exchange buffer (20mM HEPES (pH 7.4), 220mM NaCl, 1mM DTT and purified by size selection (section 2.8.4).

2.9.1 Production of RNA oligoes

A student in our lab first synthesised N^6-methyladenosine using the nucleophilic displacement method (Ross, Bruno and Petersen, 1963). A method developed by Kruse et al demonstrated that Aritomo's phase transfer catalysis method could be modified to produce m^6A phosphoramidites (Kruse *et al.*, 2011). These m^6A phosphoramidites were then incorporated into random oligonucleotide sequences in a ratio of A/C/G/U = 0.30:0.25:0.23:0.20.

2.9.2 Binding to amylose beads

NEB Amylose magnetic beads were vortexed to thoroughly resuspend. 200µL of bead suspension was aliquoted into a sterile Eppendorf tube, for each prtein sample. 500µL of MBP column buffer (200mM NaCl, 20mM Tris-HCl, 0.05% NP-40) was added to each tube and vortexed to wash. A magnet rack (NEB) was used to pull beads aside, buffer was removed, and the above step repeated. 286 pmol of purified protein was added to each sample. Beads-protein mixes were incubated at 4°C with agitation for 1 hour. Subsequent protein-bound beads were washed 2 times with MBP column buffer as described above, and a final wash with 500 µL selection buffer (200mM NaCl, 20mM Tris-HCl, 0.05% NP-40, 100 µg/µL Salmon sperm DNA).

2.9.3 Kinasing of RNA oligoes

PNK was used to add a 5′ phosphate to the first nucleotide of the RNA. PNK reaction were scaled up to 5X. Each reaction consisted of: 25µL PNK buffer, Mix 1 (2.5 µg 1agAgc, 2.5 µg 2ga6cg) or Mix 2 (2.5 µg 3a6gc, 2.5 µg 4gaAcg), 2.5µL rATP, 5µL PNK, dH_2O Up to 250µL. This was incubated at 37°C for 1 hour.

2.9.4 Application of RNA to protein

Kinased oligoes produced in section 2.9.3 were mixed with protein bound bead produced in section 2.9.2. A magnetic rack was used to remove selection buffer buffer from each protein-bead sample. 300µL of selection buffer and 8.75 RNase inhibitor murine (New England Biolabs) were mixed with protein-beads, before the addition of 50µL of kinased oligo Mix 1 or Mix 2 per protein sample. Tubes were parafilmed to prevent contamination and agitated at room temperature (25°C) for 1 hour.

2.9.5 DNase treatment

Each sample was DNase treated to remove salmon sperm DNA before library synthesis. Each sample was washed in 500µL DNase buffer (Thermo Fisher Scientific TURBO DNase). RNA-protein bound beads were resuspended in 200µL of DNase buffer and 3µL of TURBO DNase (Thermo Fisher Scientific) and incubated at 37°C for 30 minutes.

2.9.6 Proteinase K treatment

RNA was eluted of beads using Proteinase K. Beads were washed in 500µL Proteinase K buffer. Then re-suspended in 45µL of Proteinase K buffer with 5µL Proteinase K. Incubated at 50°C for 1 hour. Beads were pulled away and the supernatant was kept and stored at -20°C.

2.9.7 Pull-down RNA library synthesis.

Eluted RNA (section 2.9.2-2.9.6) was cleaned and concentrated to 11µL using a RNA clean and concentrate-5 kit (Zymo), according to manufacturer protocols. Libraries were produced with NEBNext Multiplex Small RNA Library Prep Set (New England Biolabs). Oligoes must be kinased for liagtion of the NEBNext 3' SR Adaptor but no ATP can be present for the ligation of the NEBNext 5' SR Adaptor.

The 3' adaptor was ligated to the eluted RNA following manufacturer protocols. RNA oligoes ligated to the 3' adaptor was then kinased using PNK: 5µL PNK buffer, 10µL Oligo mix, 0.5µL rATP, 1µL PNK, 3.5µL dH_2O. Kinased oligoes were then reconcentrated as described above. The remainder of the protocol from the NEBNext Multiplex Small RNA Library Prep Set was followed according to manufacturer guidelines. After production of PCR products, these were treated with ExoI (Thermo Fisher Scientific) before purifying.

2.9.8 Small RNA library sequencing

Produced small RNA libraries were sent for next generation sequencing (Deep Seq - University of Nottingham). Prior to sequencing all libraries were further purified and size selected utilising BluePippin (Sage Science), collecting all DNA fragments between 147-151bp. The 9 samples were split into 4 lanes and sequenced using an Illumina NextSeq500 system.

2.9.9 RNA sequencing analysis

Custom DAX and bash scripts were developed to extract the 9-mer from each oligonucleotide and label them as A containing or m^6A containing sequences. Additional scripts were then used to count the abundance of each unique sequence and compare these to the relative abundance of the input. The relative abundance of each sequence was normalised to account for slight

variation in the abundance of each unique sequence in the input. The relative abundance of each base at each position was also calculated for every sample.

2.10 Western Blot

Prior to western blotting Protein gels, containing target samples were run as described in section 2.8.2.

2.10.1 Transfer

Transferring was carried out following instructions for "Mini Trans-Blot® Electrophoretic Transfer Cell" (Bio-Rad). The gel was equilibrated using cold transfer buffer (25 mM Tris, 192 mM glycine and 20% [v/v] methanol, pH 8.3) for 30 min prior to transferring. Amersham™ Protran® Western blotting membrane (nitrocellulose, pore size: 0.2 µm), filter paper and fibre pads were also soaked in cold transfer buffer before transferring. The transfer cassette was assembled according the manufacturer's instructions and the transfer was run at 120 V for 1 h.

2.10.2 Blocking and antibody incubation

The blocking and the following steps were carried out according to the protocol for "WesternBreeze® Chemiluminescent Western Blot Immunodetection" Kit. After blocking with 10 ml of Blocking Solution for 1 h, the membrane was rinsed with 20 ml of sterile water. Subsequently, the membrane was incubated with the primary antibody solution at room temperature for 1 h. The primary antibody was diluted according to the manufacturer's instructions. After 4 washes with Antibody Wash Solution, the membrane was incubated with Secondary Antibody Solution for 30 min.

Afterwards, the membrane was washed 4 times with Antibody Wash Solution and then washed twice with sterile water.

2.10.3 Incubation with substrate

The Chemiluminescent Substrate was prepared following the manufacturer's instructions. The substrate was evenly applied onto the membrane and the reaction was developed at room temperature for 5 min. The extra substrate was removed by filter paper

2.10.4 Detection

The membrane from the last step was placed between two pieces of transparency plastic and was then put into a cassette. In the dark room, a film was put on top of the membrane and the exposure was kept for 30 seconds or longer (depending on the specificity of the primary antibody). Afterwards, the film was developed in Carestream® Kodak® autoradiography GBX developer/replenisher (Sigma-Aldrich) and then fixed by Carestream® Kodak® autoradiography GBX fixer/replenisher (Sigma-Aldrich). Finally, the film was scanned by GS-800™ Calibrated Imaging Densitometer (Bio-Rad) in combination with Quantity One software.

2.11 Thin Layer Chromatography

2.11.1 Internal TLC

m^6A was detected in the consensus sequence GAC. The T1 enzyme used in the TLC experiments cuts RNA after G exposing the next nucleotide including m^6A modified adenosine. The reaction mixture used consisted of 25ng of RNA, 1µl PNK buffer A, 1µl T1 and water up to 8.5µl and was incubated for 15 minutes at 37°C. 1µl PNK and 0.5µl 32pATP was then added to label the exposed nucleotides. The PNK reaction was incubated for 1hr at 37°C. The reaction mixture was subsequently ethanol precipitated overnight at -20°C with the addition of 1µl 3M NaOAc ph5 and 40µl ethanol. Labelled RNA was precipitated by centrifugation at 13,000rpm for 20 minutes. The supernatant was removed and the pellet was dried for 5 minutes in a 37°C hot block. The RNA was digested using P1 enzyme in a reaction mixture consisting of the RNA pellet, 1µl 10x P1 buffer, 0.5µl P1, 8.5µl water and this was incubated for 2 hours at 37°C.

2.11.2 Running a TLC

1µl of the labelled nucleotides were added to cellulose TLC plates and placed in the TLC first buffer overnight and left to dry before the plate was turned and placed into the second buffer overnight. The plates were dried and placed on a phospho-screen and imaged using a FX scanner. The TLCs were cellulose TLC plates.

2.11.3 Imaging a TLC screen

The TLC screens were imaged using a FX scanner (Bio-Rad). Quantity One software (Bio-Rad) was used to analyse the TLC images. The software detects the saturation of the pixels in a given area. A ratio was created between the saturation of the A spot the m^6A spot.

Chapter 3 - Genome-wide bioinformatic identification of *Solanum lycopersicum* YTH proteins

3.1 Introduction

The epitranscriptomic mark N^6-methyladenosine (m^6A) is the most prevalent internal modification of eukaryotic mRNAs. Recent years have seen numerous studies on this mark, elucidating roles in gene regulation and RNA processing. The majority of research on m^6A has been on mammalian systems and focussed on the multi-protein writer complexes which make this mark so prevalent (Liu *et al.*, 2014; Schwartz, Maxwell R. Mumbach, *et al.*, 2014; Y. Wang *et al.*, 2014; Ke *et al.*, 2017; Yue *et al.*, 2018).

While less is known about m^6A methylation in plants, several components of the methylation complex have been identified. However, little is known about the RNA-binding proteins (RBPs) which 'read' this mark and convey their function (Zhong *et al.*, 2008; Růžička *et al.*, 2017; Zhao *et al.*, 2019).

The characterization of m^6A readers is essential for improving our understanding of how m^6A functions and to provide avenues to reveal the diverse roles of this methylation. Many of the effects of m^6A are mediated by RBPs which contain an YT521-B homology (YTH) domain (Hartmann *et al.*, 1999; Stoilov, Rafalska and Stamm, 2002). A few studies on the fundamentals of mammalian epitranscriptiomics have established that in human and mouse there are 5 YTH proteins and these are split into two functional groups named: YTH Domain Containing (YTHDC1 and DC2) and YTH domain Family (YTHDF1 to 3) (N. Wang *et al.*, 2014). Confirmed roles of these proteins include: mRNA degradation, nuclear export, translational control and pre-mRNA splicing (N. Wang *et al.*, 2014; Meyer *et al.*, 2015; Xiao *et al.*, 2016; Phillip J Hsu *et al.*, 2017; Roundtree, Luo, *et al.*, 2017; Shi *et al.*, 2017). However, to what extent m^6A and their 'reader' proteins mediate these roles is unclear.

The presence of YTH proteins within plant genomes suggest that a similar role as the primary m^{6}A readers would be observed there. Plant genomes contain a similar number of DC-type YTH proteins and an expanded group of DF-type proteins. *Arabidopisis* genomes for example, encode 13 YTH proteins; 12 previously described as Evolutionarily Conserved C-terminus (ECT) proteins and one a homolog of Cleavage and Polyadenylation Specificity Factor 4 (CPSF4), named CPSF30. Of these, ECT1 to 11 are YTHDF-type proteins, while CPSF30 and ECT12 are YTHDC-type proteins. The observed expansion of the YTHDF-type proteins in plants increases the difficultly of their study. In fact, the chances of redundant functions within the 11 ect genes is high, while the concurrent knockout of all genes is not practical. Recent studies on the ECT proteins in *Arabidopsis* have highlighted this issue (Arribas-Hernández *et al.*, 2018; L.-H. H. Wei *et al.*, 2018; Scutenaire, J.-M. Deragon, *et al.*, 2018). Where, the simultaneous knockout of ECT2, 3 and 4 together was necessary to observe any aberrant phenotypes.

The overall aim of this work is to collect and analyse the expanded DF-type group of plant YTH proteins. Through genome-wide analysis of the protein sequences, we will be able to identify and characterise a suitable model for the *in-vivo* study of the YTH protein family.

Overall, the data presented here suggests *Solanum lycopersicum* (from now on addressed as Tomato) as an ideal model for the elucidation of the cytoplasmic YTHDF proteins. The establishment of this model for m^{6}A reader research should further expand our knowledge of m^{6}A and provide a platform for future applications. Tomato's current role as a key plant model for fruit biology, and its ranking as the fourth largest crop in terms of global cash value, give additional significance to these studies.

3.2 Results

3.2.1 Phylogenetic comparison of YT521-B domain containing proteins

The presence of these YTH proteins in almost all Eukaryote species raises many questions, regarding their presence and abundance in plant species and about their functional in-vivo roles compared to their relatives in metazoans. How many of these proteins can be found in plant species? How closely related are they to studied YTH proteins in mammalian species? Do they have similar functional roles *in vivo*? To gain insight into the functional role of YTH domain proteins, the amino-acidic sequences were compared across key evolutionary species. Representatives of Mammals (*Homo sapiens*), Birds (*Gallus gallus*), Amphibians (*Xenopus laevis*), Fish (*Danio rario*), Insects (*Drosophila melanogaster*) and Yeast (*Saccharomyces cerevisiae*) were aligned against all YTH proteins from *Arabidopsis*, Rice and Tomato.

Proteins were aligned using MUSCLE 3.8.4 (Edgar, 2004). Alignments were then passed through MEGA 10.1 standard parameters with 500 bootstrap replications. (Figure 3.1)

From this study we observed that the two classes of YTH protein described in the literature, YTHDF and YTHDC are consistent across Eukaryotes, visible as two groups, a monophyletic clade of YTHDC proteins and the remaining YTHDF proteins. It is important to note that the bootstrap value of 63 at this split is not a high level of significance (Figure 3.1). The resolution of this tree is likely harmed by a high degree of noise or convergent evolution from the residues outside of the YTH domain. A clearer tree was observed with the YTH domain sequences only in a concurrent analysis (Scutenaire, J.-M. Deragon, *et al.*, 2018).

Within each group, the animal YTH proteins are distanced from those in plants. In fact, all plant representatives diverge from those within the animal kingdom clustering together, separate from those within the animal kingdom. Interestingly, there is a single plant YTH protein (OsYTHDC1), which is placed within the Animalia YTHDC clade. It is important to note

that the plant YTHDC proteins are more similar to YTHDC1, than YTHDC2. Whether this protein is functionally similar to the YTHDC2 proteins has yet to be studied. Interestingly, OsYTHDF9, previously described as a DF-type protein, falls within the DC clade.

This would suggest that this protein could be re-classified into the YTHDC group.

It is of note that the YTHDF proteins of Animalia and Yeast are more similar in sequence to plant YTHDC proteins than the plant YTHDF proteins are to proteins within their species. This observation and the clear evolutionary expansion of plant YTHDF-type proteins could suggest a more divergent role of plant YTHDF proteins than previously discussed. This subgroup of YTH proteins are likely to be functionally distinct to the YTHDF proteins previously characterised in animal systems (Patil, Pickering and Jaffrey, 2017). For this reason, the interest in the plant YTH protein family is growing with each published study.

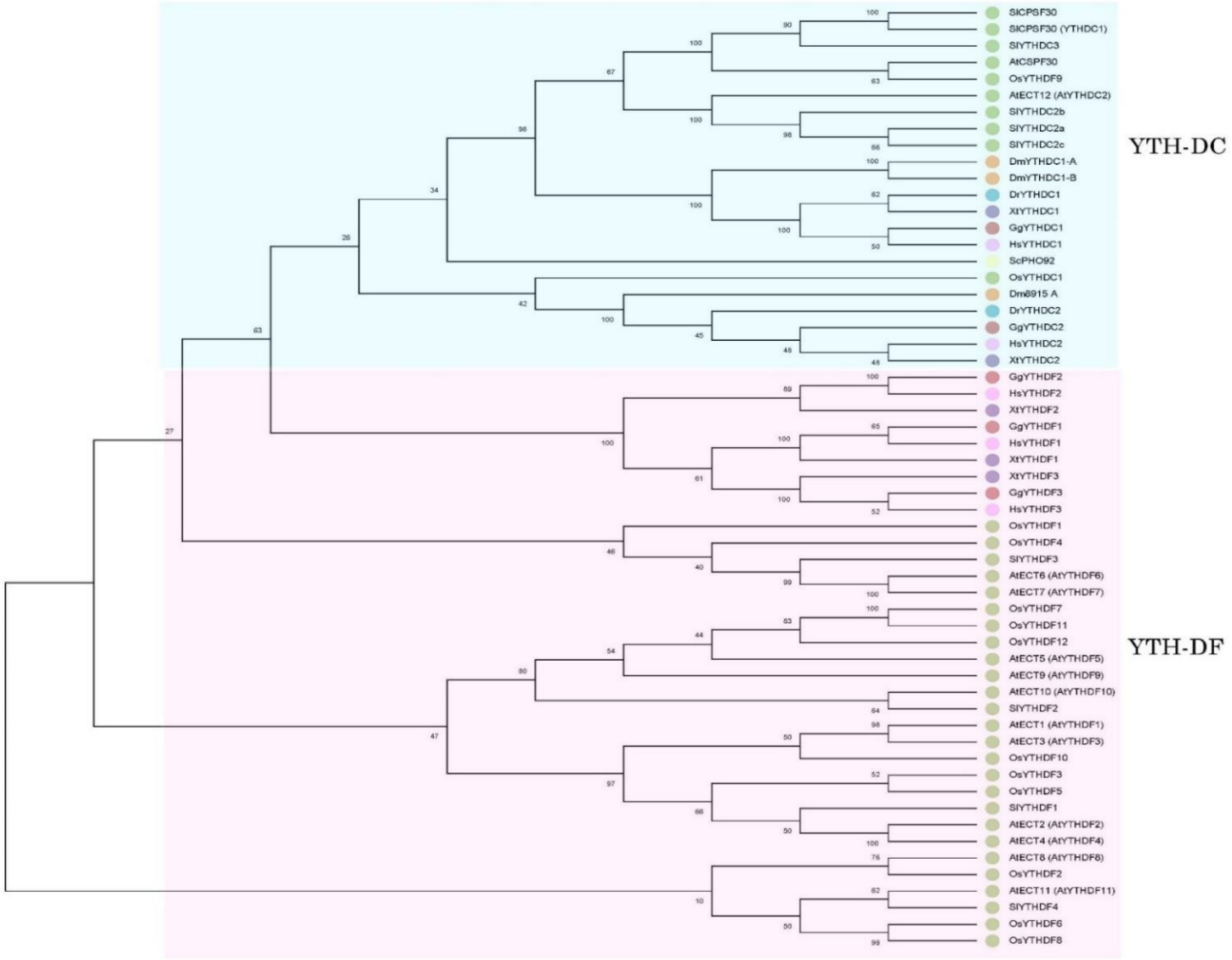


Figure 3.1: Schematic tree of similarity of YTHDF and YTHDC proteins. Comparisons between model Planta (Green), Animalia (Mixed) and Yeast (Yellow) species. At- *Arabidopsis thaliana,* Dr- *Danio rario,* Dm- *Drosophila melanogaster,* Gg- *Gallus gallus,* Hs- *Homo sapiens,* Os- *Oryza sativa,* Sc- *Saccharomyces cerevisiae* Sl- *Solanum lycopersicum,* Xt- *Xenopus tropicalis.* .Protein sequences analysed using aligned using MUSCLE 3.8.4 and analysed using MEGA 10.1 standard parameters with 500 bootstrap replications.

3.2.2 Genome-wide identification of plant YT521-B domain containing proteins

Many plant YTH proteins have been computationally predicted, but the vast majority have not been studied biologically. We surmised that the evolutionary expansion of the plant YTH protein is unlikely to be insignificant. Therefore, we set out to gain a better understanding of this protein family throughout Plantae. Genome-wide bioinformatic approaches were used for the collection of all YTH proteins present in key evolutionary models. The model species were chosen based on their position within evolutionary and agriculturally relevant clades. The YTH domain (PFAM: PF04146 was used as a query in genomic databases GenBank, Phytozome (v12.1) utilising both BLASTP and HMM searches, each result was then utilised as a query until no further novel proteins were returned. Using this approach, we have identified 286 YTH domain proteins within 29 plant species (Table 3.1). Full sequences can be found in the Appendix (Appendix 3.1).

Table 3.2: YTHDF and YTHDC proteins in plant species. Numbers of YTHDF and YTHDC proteins identified in each of the representative plant species.

Species	Initials	YTHDC	YTHDF	Total
Ananas comosus	*Aco*	1	8	9
Aquilegia caeruea	*Ac*	3	4	7
Amaranthus hypochondriacus	*Ah*	2	7	9
Amborella trichopoda	*Atr*	2	5	7
Arabidopsis thaliana	*At*	2	11	13
Brassica oleracea	*Bo*	2	12	14
Capsella rubella	*Cr*	2	9	11
Citrus sinensis	*Ci*	3	7	10
Eucalyptus grandis	*Eg*	2	4	6
Gossypium raimondii	*Gr*	2	11	13
Manihot esculenta	*Me*	2	11	13
Marchantia polymorpha	*Mpo*	2	1	3
Musa acuminata	*Ma*	2	16	18
Mimulus glutinosus	*Mg*	3	6	9
Medicago truncatula	*Mt*	1	8	9
Oropetium thomaeum	*Ot*	1	7	8
Oriza sativa	*Os*	1	12	13
Panicum hallii	*Ph*	1	11	12
Prunus persica	*Pp*	4	8	12

Populus trichocarpa	*Pt*	3	14	17
Senna italica	*Si*	1	11	12
Selaginella moellendorffii		1	1	2
Sphgnum fallax	*Sf*	1	5	6
Solanum lycopersicum	*Sl*	5	4	9
Theobroma cacao	*Tc*	2	6	8
Vitus vinifera	*Vv*	2	8	10
Volvox carteri	*Vc*	1	0	1
Zea mays	*Zm*	1	14	15
Zostera marina	*Zma*	1	7	8

All plant species analysed have at least one YTHDC (Mean: 1.93) protein and all but Green algae (*Volvox carteri*) have at least one YTHDF protein (Mean: 7.86). On average plants have a significantly higher ($p=<0.0001$) number of YTHDF type proteins than the nuclear YTHDC type. Interestingly, Tomato is the only higher plant species as yet identified to have more YTHDC proteins than YTHDF proteins. They also have far fewer YTHDF proteins than other dicotyledonous model species, providing greater potential for reverse-genetic studies.

When comparing the number of YTH proteins between monocot, dicot and non-flowering plant groups, differences can be observed. Dicot plants have significantly more DC-type YTH proteins than dicot plants and monocot plants have more DF-type proteins than both dicots and the non-flowering plants studied. Both dicots and monocots have more YTH proteins than their non-flowering cousins which suggests that the expansion of both groups occurred after these plants and metazoans diverged. As mentioned above, tomato seems to be unique in its distribution of YTH proteins. In fact, tomato has significantly more YTHDC proteins than all other plant groups, including its fellow dicots, with 5 proteins. However, this is primarily due to a local YTHDC gene triplication event at a single locus.

3.2.3 High conservation of YT521-B domain

All previous studies have highlighted the high level of conservation of the YTH domain, even between divergent evolutionary groups (Ok *et al.*, 2005; Shukun Luo and Tong, 2014; Zhu *et al.*, 2014). Comparison of plant YTH domains (*Arabidopsis, Solanum* and Human) identify 18 identical sites which are 100% conserved across all 27 protein sequences (9.9% of the total protein), with an overall pairwise identity (BLOSUM62) of 63.3%. This rises to 29 identical sites (16.1% of the total protein) with 76.0% pairwise identity and 20 identical sites (11.3% of the total protein) with 66.0% pairwise identity between YTHDF and YTHDC proteins groups, respectively. This high level of conservation (Figure 3.2) not only indicates the functional importance of the conserved residues within these domains for their roles *in vivo*, but also lead us to hypothesise that any differences in the role of individual YTH protein, especially within the expanded YTHDF proteins, are likely to be mediated by sequence differences outside of the domain. This is only further supported by the lack of additional domains (with a few exceptions) within the YTH family (D. Li *et al.*, 2014).

A number of published works have biochemically confirmed that the predicted aromatic cage which directly accommodates the m^6A is formed by either three Tryptophan (W) residues or two Tryptophan and a Leucine (L) residue (Xu, Wang, *et al.*, 2014; Zhu *et al.*, 2014). Two W residues have been shown to be essential in the binding of YTH proteins to m^6A (Figure 3.2; W^{24}, W^{85}) (Xu, Wang, *et al.*, 2014). At position 96 in figure 3.2 corresponding to the 3rd crucial residues W or L, most plant YTH proteins contain a conserved W^{94}, however the CPSF30 type proteins share a Tyrosine (Y) residue in place of the L observed in HsYTHDC1/2. It is still possible that this Y has a role in the domain's recognition of m^6A, in fact Y is the third essential residue in the m^6A recognition site in yeast YTH protein, ScPho92 (Shukun Luo and Tong, 2014). The presence of Y^{94} in SlYTHDC2c and not W^{94} found in SlYTHDC2a/b would be a great subject for future work. SlYTHDC2 went through a triplication which is not observed in other plant species (Figure 3.1). The reason for this difference in such a key residue is unknown.

We therefore hypothesised that the conservation of these residues foreshadows roles for plant YTH proteins as *bone fide* m^{6}A 'readers'. And we predicted the variation in residues in YTH domain position 94, to be of importance in their role as m^{6}A binding proteins, but whether this is true and to what extent is not a subject of this study.

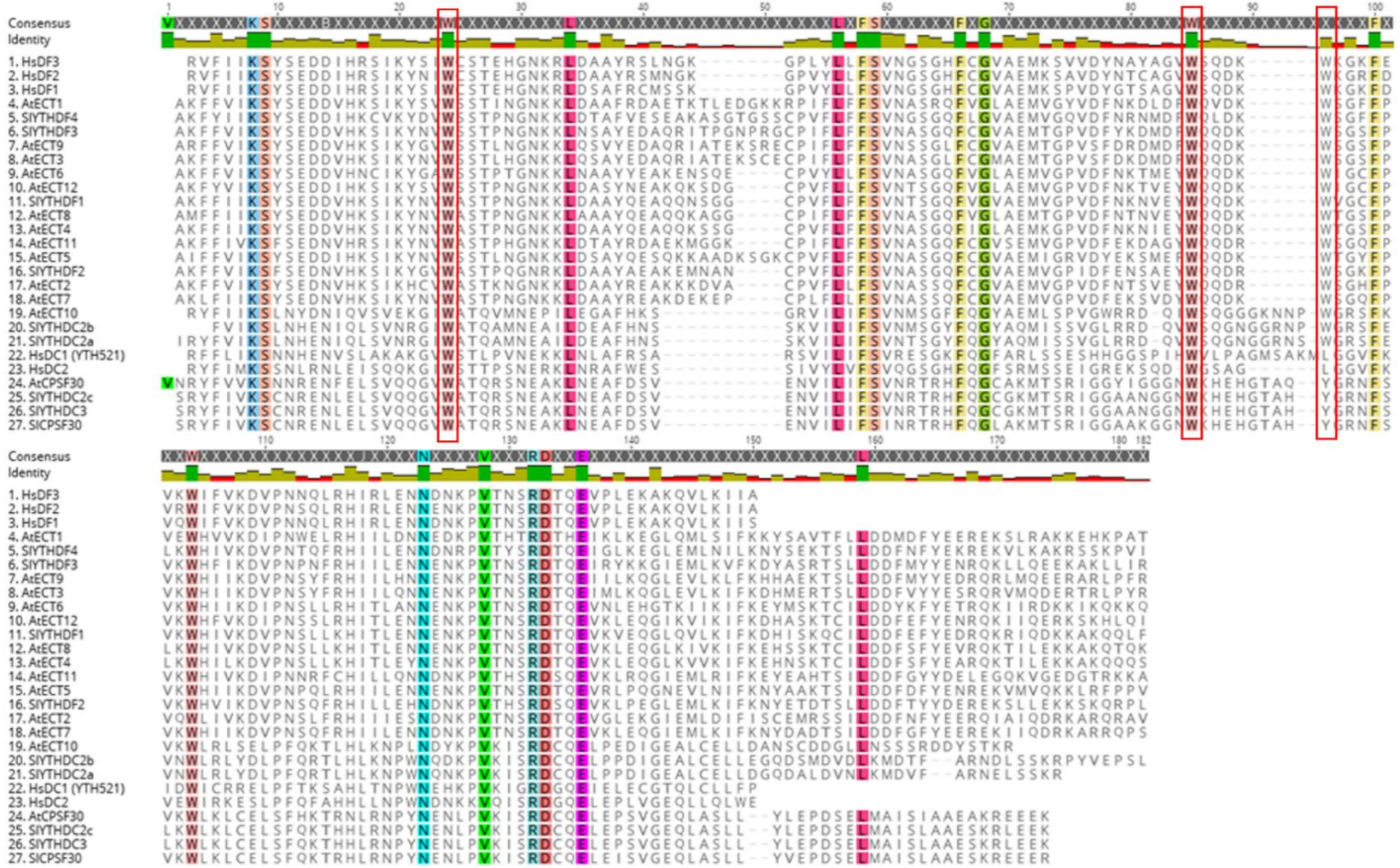


Figure 3.2: Multiple sequence alignment of YTHDF and YTHDC proteins from *Solanum lycopersicum, Arabidopsis thaliana* and *Homo sapiens.* Aligned using MUSCLE 3.8.4. Residues with 100% conservation highlighted in block colour. Key residues for YTH domain binding at positions 24, 84 and 96, indicated.

3.2.4 Analysis and comparison of the DC-type YTH protein family within Plant species

Comparison of the similarity in YTHDC proteins found throughout plant species can suggest avenues for further research and aid in our selection of an appropriate model for further study. As above all YTHDC proteins were aligned using MUSCLE 3.8.4 (Edgar, 2004). Alignments were passed through MEGA 10.1 standard parameters with 500 bootstrap replications (Figure 3.3).

Like the YTHDC clades observed in figure 3.1, a large monophyletic clade containing CPSF30 like proteins and other DC1 proteins can be observed. Every plant species has at least one YTHDC1 protein and this observation is consistent even in outgroups to flowering plants.

This CPSF30-like/DC1 group is the only group from plant YTH protein to have a representative protein from every single species analysed in this study. This suggests that these proteins are essential for development and/or survival of these plants. Plants are the only group to have a YTH domain within these polyadenylation factors (Addepalli and Hunt, 2007). Recent advances in our understanding of *Arabidopsis* CPSF30-L has identified that the YTH-domain in CPSF30 is required for floral transition and abscisic acid responses. The ability of CPSF30-L to bind m^6A, encourages the formation of nuclear assemblages. In these assemblages, CSPF30-L was shown to bind m^6A-modified far-upstream elements controlling polyadenylation site choice (Song *et al.*, 2021).

The other clusters of this predominately nuclear YTHDC group are almost exclusively from dicot plant species (with the notable exception of *Amborella*). This result could indicate that this group of YTHDC proteins originated in plants before the first divergence of basal angiosperms and subsequently was lost in the monocots after they diverged from present day eudicots. Alternatively, basal dicots could have evolved this branch independently.

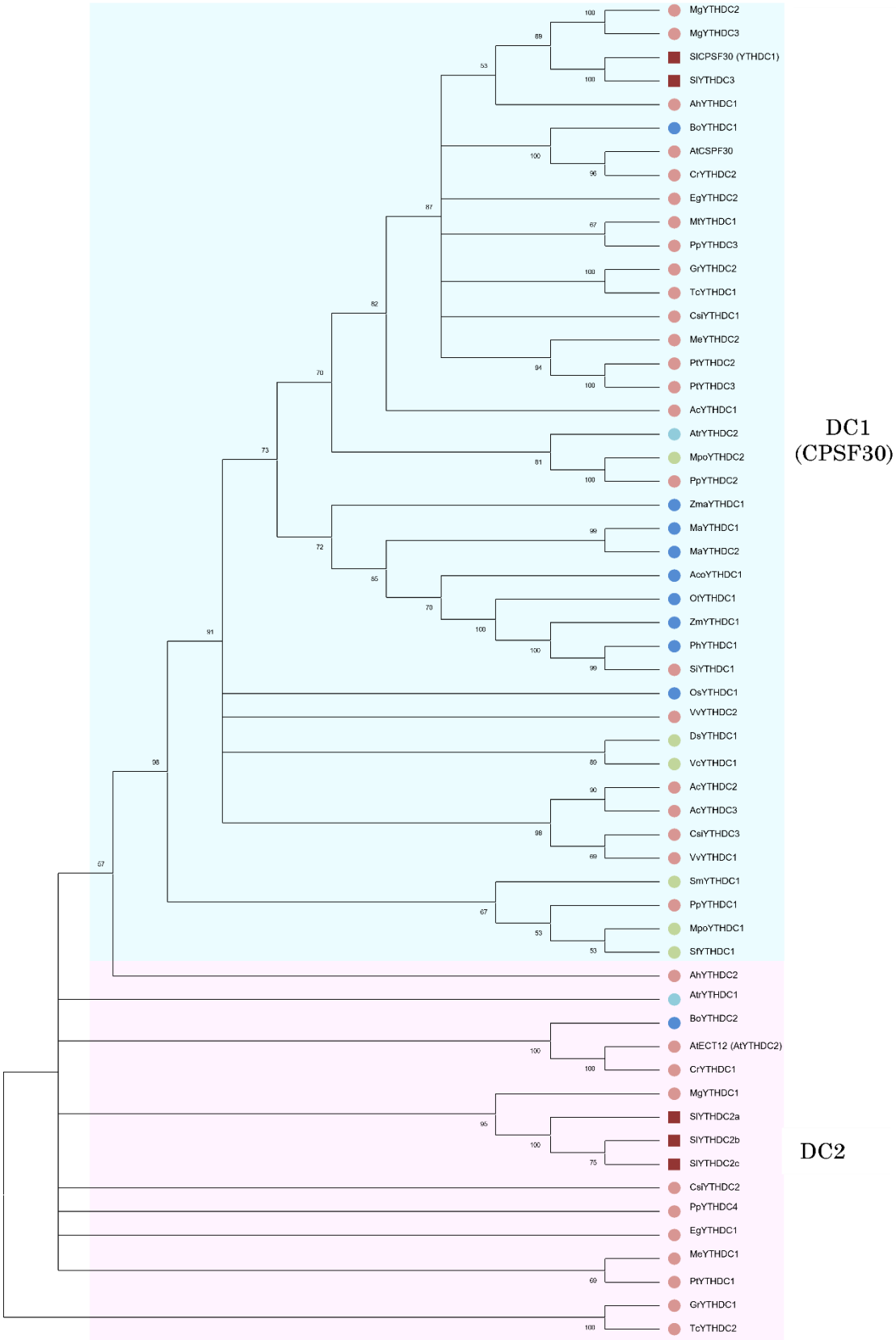


Figure 3.3: Schematic tree of similarity of YTHDC proteins. Comparisons between model Planta species. Evolutionary groups indicated: Monocots (Blue), Dicots (Red), Amborella (Light Blue) and Outgroup (Green). *Solanum lycopersicum* indicated separately (Red Square).Aco- *Ananas comosus*, Ac- *Aquilegia caeruea*, Ah- *Amaranthus*

hypochondriacus, Atr- *Amborella trichopoda*, At-*Arabidopsis thaliana*, Bo- *Brassica oleracea*, Cr- *Capsella rubella*, Ci- *Citrus sinensis*, Eg- *Eucalyptus grandis*, Gr- *Gossypium raimondii*, Me- *Manihot esculenta*, Mpo- *Marchantia polymorpha*, Ma- *Musa acuminantane*, Mg- *Mimulus glutinosus*, Mt- *Medicago trunculata*, Ot- *Oropetium thomaeum*, Os- *Oriza sativa*, Ph- *Panicum hallii*, Pp- *Prunus persica*, Pt- *Populus trichocarpa*, Si- *Senna italica*, Sm- *Selaginella moellendorffii*, Sf- *Sphgnum fallax*, Sl- *Solanum lycopersicum*, Tc- *Theobroma cacao*, Vv- *Vitus vinifera*, Vc- *Volvox carteri*, Zm- *Zea mays*, Zma- Zostera marina .Protein sequences analysed using aligned using MUSCLE 3.8.4 and analysed using MEGA 10.1 standard parameters with 500 bootstrap replications.3.2.5 Analysis and comparison of the DF-type YTH protein family within Plant species.

As previously mentioned, *Solanum lycopersicum* have a triplicated YTHDC2. This is the only example of this currently identified and only adds to the interest of Tomato as an ideal model for the study of YTH RNA binding proteins.

3.2.5 Analysis and comparison of the DF-type YTH protein family within Plant species

Given the expansion and complexity found within the DF group of YTH proteins it is understandable that current research into this family has been limited and slow. High numbers of YTHDF proteins are present within the genomes of higher plant species, including *Arabidopsis.* During this study several articles were published looking into this group. Each was only able to focus on a small number of *Arabidopsis* YTHDF proteins (typically AtYTHDF2, 3, 4). For an alternative model to be identified. Previously,

Again, all YTHDF proteins were aligned using MUSCLE 3.8.4 (Edgar, 2004). Alignments were passed through MEGA 10.1 standard parameters with 500 bootstrap replications (Figure 3.4).

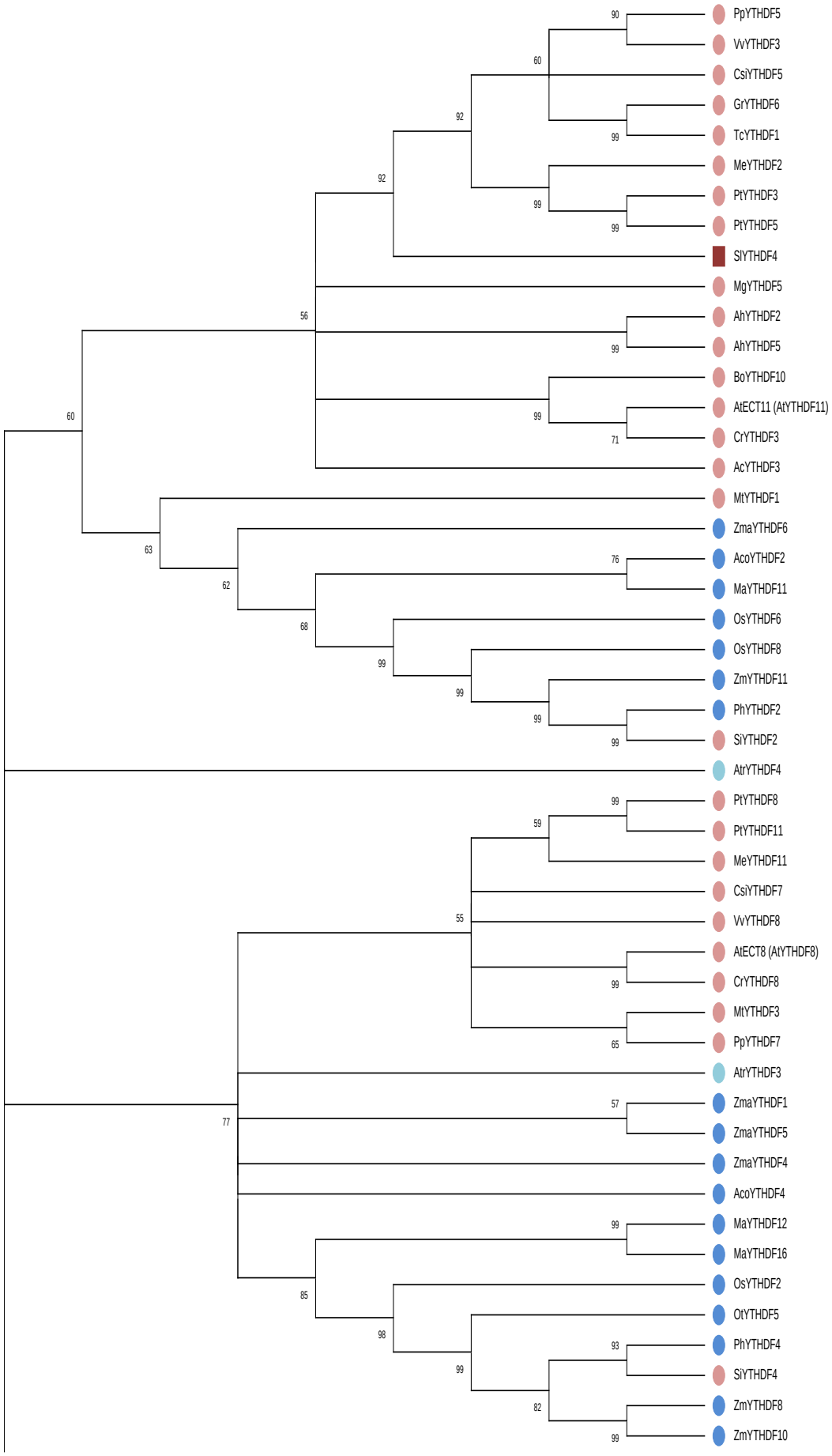
Due to the large expansion of the YTHDF protein family, any potential species, with fewer YTH proteins, would ideally have proteins which fall into each of the predicted functional groups. A bioinformatics study into this group was conducted during this project (Scutenaire, J. M. Deragon, *et al.*, 2018). This paper postulated that there were three major YTHDF groups (DFA, B and C), this analysis was performed using alignments only of the YTH domain of each individual protein. As mentioned above the YTH

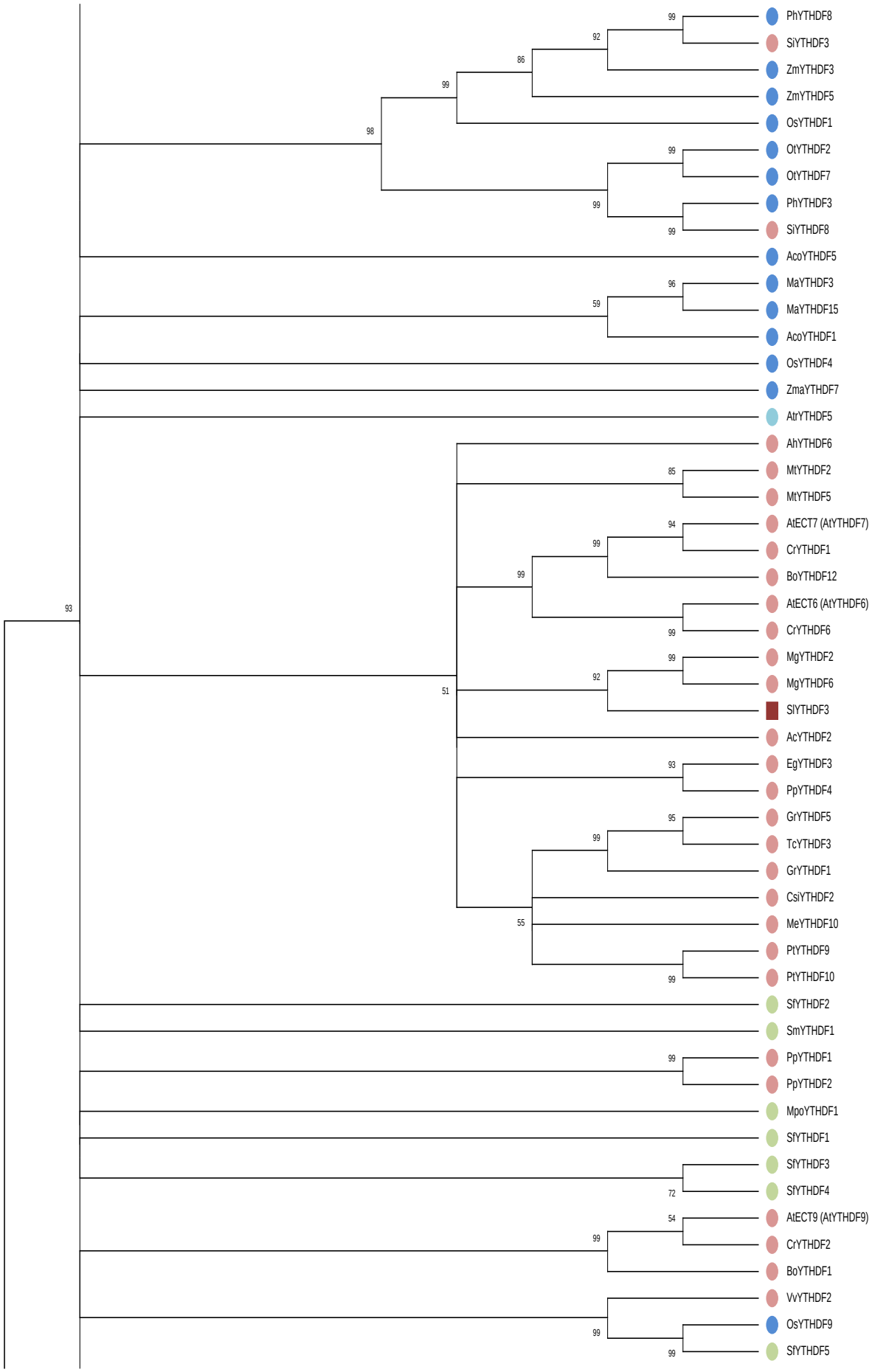
domain is conserved to an extent were leaving out the full-length protein could potentially exclude necessary regions or residues which confer the eventually function of the YTH protein family.

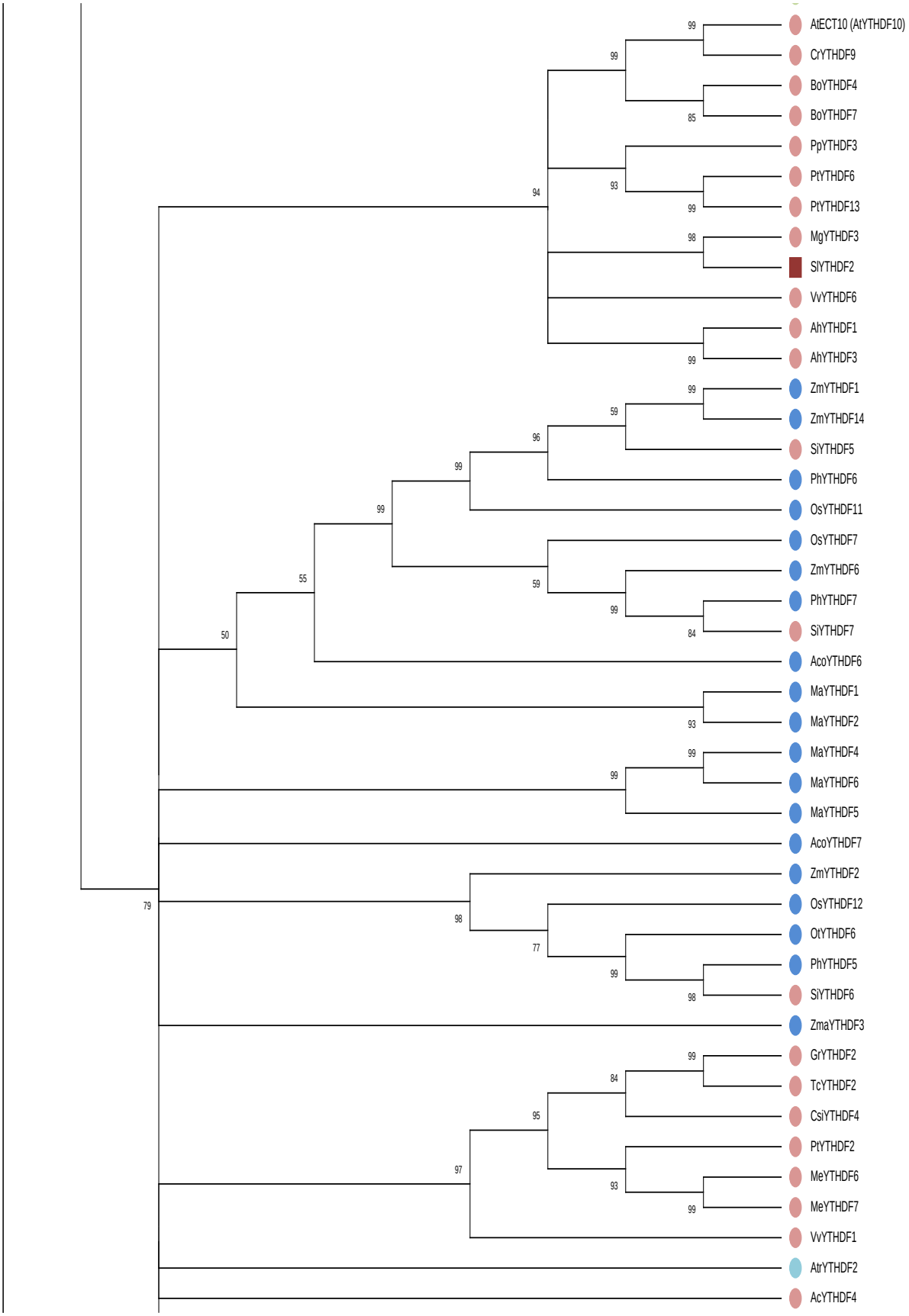
Our analysis of full length YTH proteins is more representative of the plant population. However, as previously observed in the YTHDC analysis (Figure 3.3) the regions outside of the YTH domain add evolutionary noise to the analysis leading to a poorly resolved tree. The tree does allow us to identify the closest homolog of each tomato YTHDF protein in *Arabidopsis,* since almost all papers on plant YTHDF proteins have focussed on this model this could be important for future comparisons. SlYTHDF1 is likely the closest homolog of AtECT1, AtECT2, AtECT3 and AtECT4, the focus of the predominant studies in this family (Arribas-Hernández et al., 2018; Arribas-Hernández et al., 2021; L.-H. H. Wei et al., 2018; Scutenaire, J.-M. Deragon, et al., 2018). SlYTHDF2 falls into a monophyletic subgroup with homologs AtECT10 and AtECT5. SlYTHDF3 appears to be most closely related to AtECT6 and AtECT7. Finally, SlYTHDF4 closest homolog in *Arabidopisis* is AtECT11. Although these suggestions have very low statistical significance, they do follow the findings of the previous phylogenetic analysis performed by the Bousquet-Antonelli Lab (Scutenaire, J. M. Deragon, *et al.*, 2018). That analysis using only the YTH domains of key plant models suggested that SlYTHDF1 (SlDF1A) fell into group DFA with AtECT1, AtECT2, AtECT3 and AtECT4, SlYTHDF2 (SlDF1B) with AtECT5, AtECT10 and AtECT9 in group DFB and SlYTHDF3 (SlDF1C) / SlYTHDF4 (SlDF2C) with AtECT6, AtECT7, AtECT8 and AtECT11 in group DFC. Each of these groups includes additional sub-clades which tend to cluster in monocot, dicot and non-flowering plants groups (Figure 3.4).

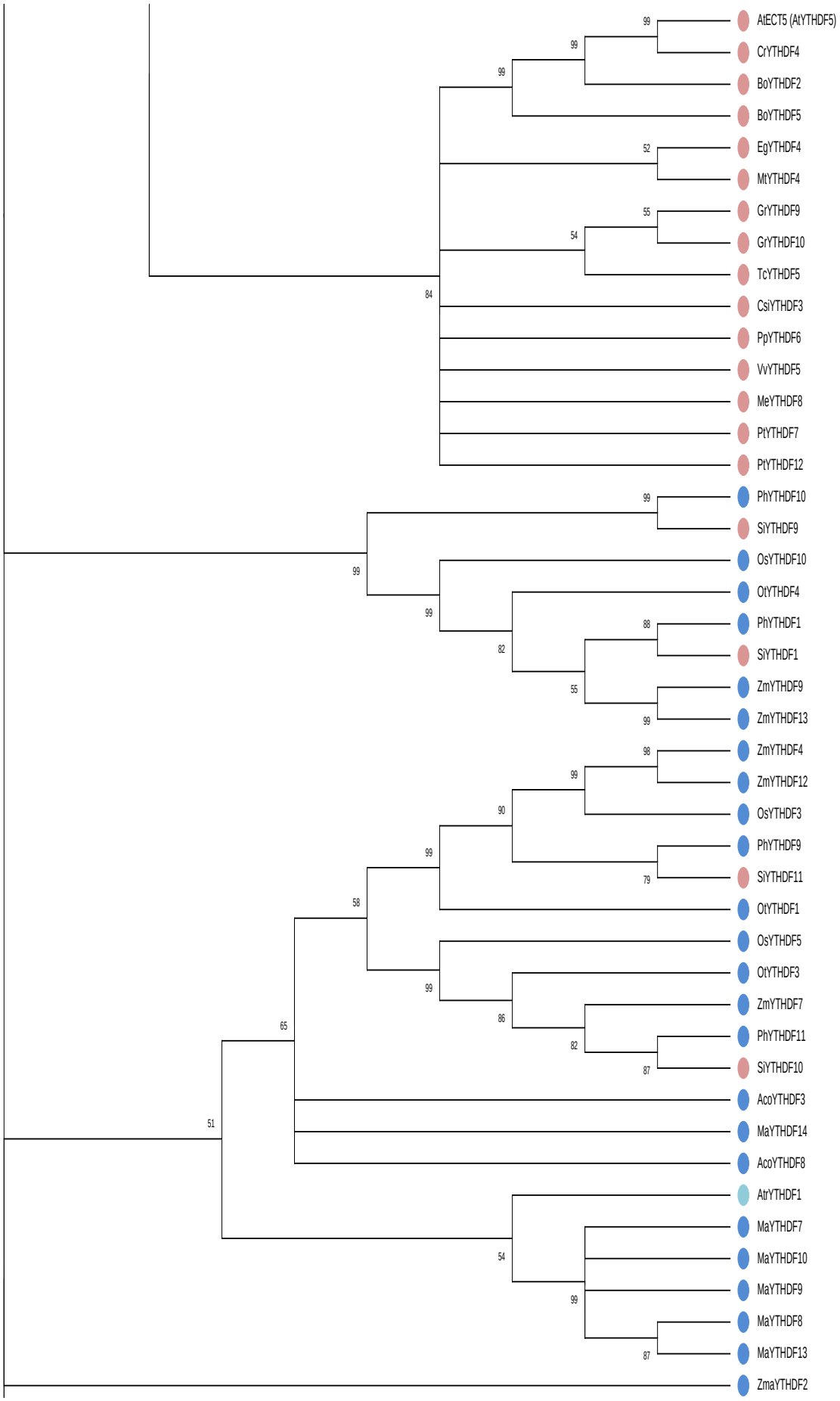
Unfortunately, although it is perhaps desirable to include the regions outside of the YTH domain in trees of similarity, the lack of any ordered domains in YTH domain containing proteins the resolution and significance of the results. However, it still may be beneficial in the future to revisit this analysis. With the number publication of studies into this family growing, and the prevalence of *Arabidopsis* models, identifying the closest homologs to each YTHDF protein could be key into understanding overlapping functions and connection between species. Overall the unusually low

number of YTHDF proteins (4) and its position as an economically important crop species make tomato a suitable model for the study of YTH domain containing proteins within Planta.


PpYTHDF5
VvYTHDF3
CsiYTHDF5
GrYTHDF6
TcYTHDF1
MeYTHDF2
PtYTHDF3
PtYTHDF5
SlYTHDF4
MgYTHDF5
AhYTHDF2
AhYTHDF5
BoYTHDF10
AtECT11 (AtYTHDF11)
CrYTHDF3
AcYTHDF3
MtYTHDF1
ZmaYTHDF6
AcoYTHDF2
MaYTHDF11
OsYTHDF6
OsYTHDF8
ZmYTHDF11
PhYTHDF2
SiYTHDF2
AtrYTHDF4
PtYTHDF8
PtYTHDF11
MeYTHDF11
CsiYTHDF7
VvYTHDF8
AtECT8 (AtYTHDF8)
CrYTHDF8
MtYTHDF3
PpYTHDF7
AtrYTHDF3
ZmaYTHDF1
ZmaYTHDF5
ZmaYTHDF4
AcoYTHDF4
MaYTHDF12
MaYTHDF16
OsYTHDF2
OtYTHDF5
PhYTHDF4
SiYTHDF4
ZmYTHDF8
ZmYTHDF10


PhYTHDF8
SiYTHDF3
ZmYTHDF3
ZmYTHDF5
OsYTHDF1
OtYTHDF2
OtYTHDF7
PhYTHDF3
SiYTHDF8
AcoYTHDF5
MaYTHDF3
MaYTHDF15
AcoYTHDF1
OsYTHDF4
ZmaYTHDF7
AtrYTHDF5
AhYTHDF6
MtYTHDF2
MtYTHDF5
AtECT7 (AtYTHDF7)
CrYTHDF1
BoYTHDF12
AtECT6 (AtYTHDF6)
CrYTHDF6
MgYTHDF2
MgYTHDF6
SlYTHDF3
AcYTHDF2
EgYTHDF3
PpYTHDF4
GrYTHDF5
TcYTHDF3
GrYTHDF1
CsiYTHDF2
MeYTHDF10
PtYTHDF9
PtYTHDF10
SfYTHDF2
SmYTHDF1
PpYTHDF1
PpYTHDF2
MpoYTHDF1
SfYTHDF1
SfYTHDF3
SfYTHDF4
AtECT9 (AtYTHDF9)
CrYTHDF2
BoYTHDF1
VvYTHDF2
OsYTHDF9
SfYTHDF5


AtECT10 (AtYTHDF10)
CrYTHDF9
BoYTHDF4
BoYTHDF7
PpYTHDF3
PtYTHDF6
PtYTHDF13
MgYTHDF3
SlYTHDF2
VvYTHDF6
AhYTHDF1
AhYTHDF3
ZmYTHDF1
ZmYTHDF14
SiYTHDF5
PhYTHDF6
OsYTHDF11
OsYTHDF7
ZmYTHDF6
PhYTHDF7
SiYTHDF7
AcoYTHDF6
MaYTHDF1
MaYTHDF2
MaYTHDF4
MaYTHDF6
MaYTHDF5
AcoYTHDF7
ZmYTHDF2
OsYTHDF12
OtYTHDF6
PhYTHDF5
SiYTHDF6
ZmaYTHDF3
GrYTHDF2
TcYTHDF2
CsiYTHDF4
PtYTHDF2
MeYTHDF6
MeYTHDF7
VvYTHDF1
AtrYTHDF2
AcYTHDF4


AtECT5 (AtYTHDF5)
CrYTHDF4
BoYTHDF2
BoYTHDF5
EgYTHDF4
MtYTHDF4
GrYTHDF9
GrYTHDF10
TcYTHDF5
CsiYTHDF3
PpYTHDF6
VvYTHDF5
MeYTHDF8
PtYTHDF7
PtYTHDF12
PhYTHDF10
SiYTHDF9
OsYTHDF10
OtYTHDF4
PhYTHDF1
SiYTHDF1
ZmYTHDF9
ZmYTHDF13
ZmYTHDF4
ZmYTHDF12
OsYTHDF3
PhYTHDF9
SiYTHDF11
OtYTHDF1
OsYTHDF5
OtYTHDF3
ZmYTHDF7
PhYTHDF11
SiYTHDF10
AcoYTHDF3
MaYTHDF14
AcoYTHDF8
AtrYTHDF1
MaYTHDF7
MaYTHDF10
MaYTHDF9
MaYTHDF8
MaYTHDF13
ZmaYTHDF2

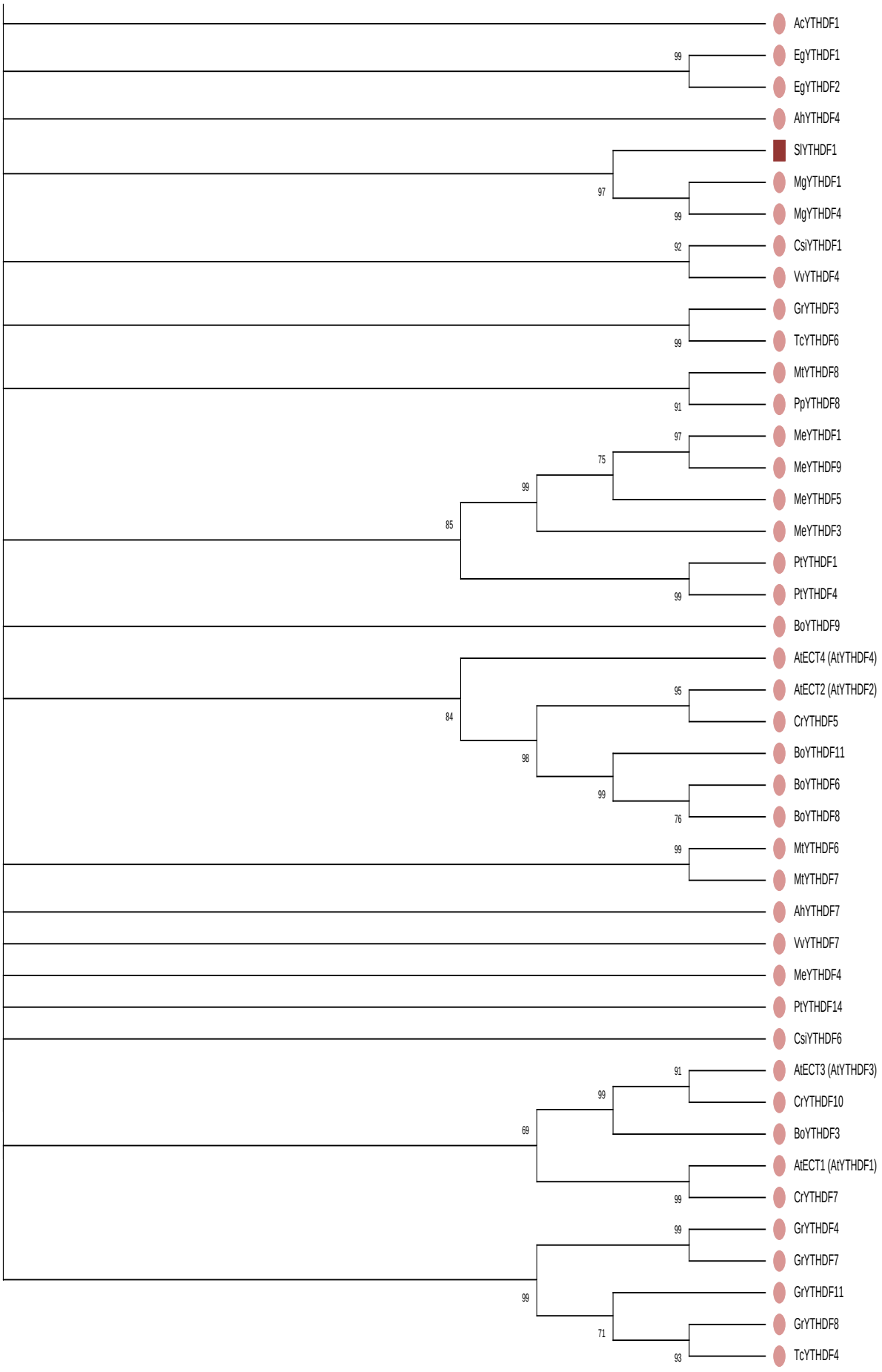

AcYTHDF1
EgYTHDF1
EgYTHDF2
AhYTHDF4
SlYTHDF1
MgYTHDF1
MgYTHDF4
CsiYTHDF1
VvYTHDF4
GrYTHDF3
TcYTHDF6
MtYTHDF8
PpYTHDF8
MeYTHDF1
MeYTHDF9
MeYTHDF5
MeYTHDF3
PtYTHDF1
PtYTHDF4
BoYTHDF9
AtECT4 (AtYTHDF4)
AtECT2 (AtYTHDF2)
CrYTHDF5
BoYTHDF11
BoYTHDF6
BoYTHDF8
MtYTHDF6
MtYTHDF7
AhYTHDF7
VvYTHDF7
MeYTHDF4
PtYTHDF14
CsiYTHDF6
AtECT3 (AtYTHDF3)
CrYTHDF10
BoYTHDF3
AtECT1 (AtYTHDF1)
CrYTHDF7
GrYTHDF4
GrYTHDF7
GrYTHDF11
GrYTHDF8
TcYTHDF4
99
97
99
92
99
91
97
75
99
85
99
95
84
98
99
76
99
91
99
69
99
99
99
71
93

Figure 3.4: Schematic tree of similarity of YTHDF proteins. Comparisons between model Planta species. Evolutionary groups indicated: Monocots (Blue), Dicots (Red), Amborella (Light Blue) and Outgroup (Green). *Solanum lycopersicum* indicated separately (Red Square *Solanum lycopersicum* indicated separately (Red Square).Aco- *Ananas comosu*s, Ac- *Aquilegia caeruea*, Ah- *Amaranthus hypochondriacus*, Atr- *Amborella trichopoda*, At- *Arabidopsis thaliana*, Bo- *Brassica oleracea*, Cr- *Capsella rubella*, Ci- *Citrus sinensis*, Eg- *Eucalyptus grandis*, Gr- *Gossypium raimondii*, Me- *Manihot esculenta*, Mpo- *Marchantia polymorpha*, Ma- *Musa acuminantane*, Mg- *Mimulus glutinosus*, Mt- *Medicago trunculata*, Ot- *Oropetium thomaeum*, Os- *Oriza sativa*, Ph- *Panicum hallii*, Pp- *Prunus persica*, Pt- *Populus trichocarpa*, Si- *Senna italica*, Sm- *Selaginella moellendorffii*, Sf- *Sphgnum fallax*, Sl- *Solanum lycopersicum*, Tc- *Theobroma cacao*, Vv- *Vitus vinifera*, Vc- *Volvox carteri*, Zm- *Zea mays*, Zma- Zostera marina .Protein sequences analysed using aligned using MUSCLE 3.8.4 and analysed using MEGA 10.1 standard parameters with 500 bootstrap replications

3.2.6 YT521-B domain containing proteins in *Solanum lycopersicum*

To establish *Solanum lycopersicum* as a novel model for the study of the YTH family of RNA binding proteins we need a greater understanding of the characteristics of these proteins and their positions within established literature. We found 9 putative YTH proteins within the tomato genome. Research focussing on tomato species, or these proteins has not been published

Table 3.2 below summarises all YTH proteins found within the Tomato genome. The majority of the proteins are 550-710 amino acids (aa) in length, with the triplicated nuclear YTHDCa-c protein between 299 and 400aa. The molecular weight of the proteins range between 33.86kDa and 77.22kDa. The genes encoding the proteins are distributed on chromosomes I, II, V, VIII and XII. Unlike *Arabidopsis,* with the exception of SlYTHDC2a,b, c which are in tandem due to a triplication event, none of the YTH genes are distributed in pairs (D. Li *et al.*, 2014).

Intron and exon distribution within gene families has been suggested as a mechanism for predicting evolutionary and functional links between YTH proteins (N. Wang *et al.*, 2014). Again, each YTH protein has a unique exon-intron arrangement (Figure 3.5), further indication of the unique evolution of each of these proteins and the lack of expansion observed within this family.

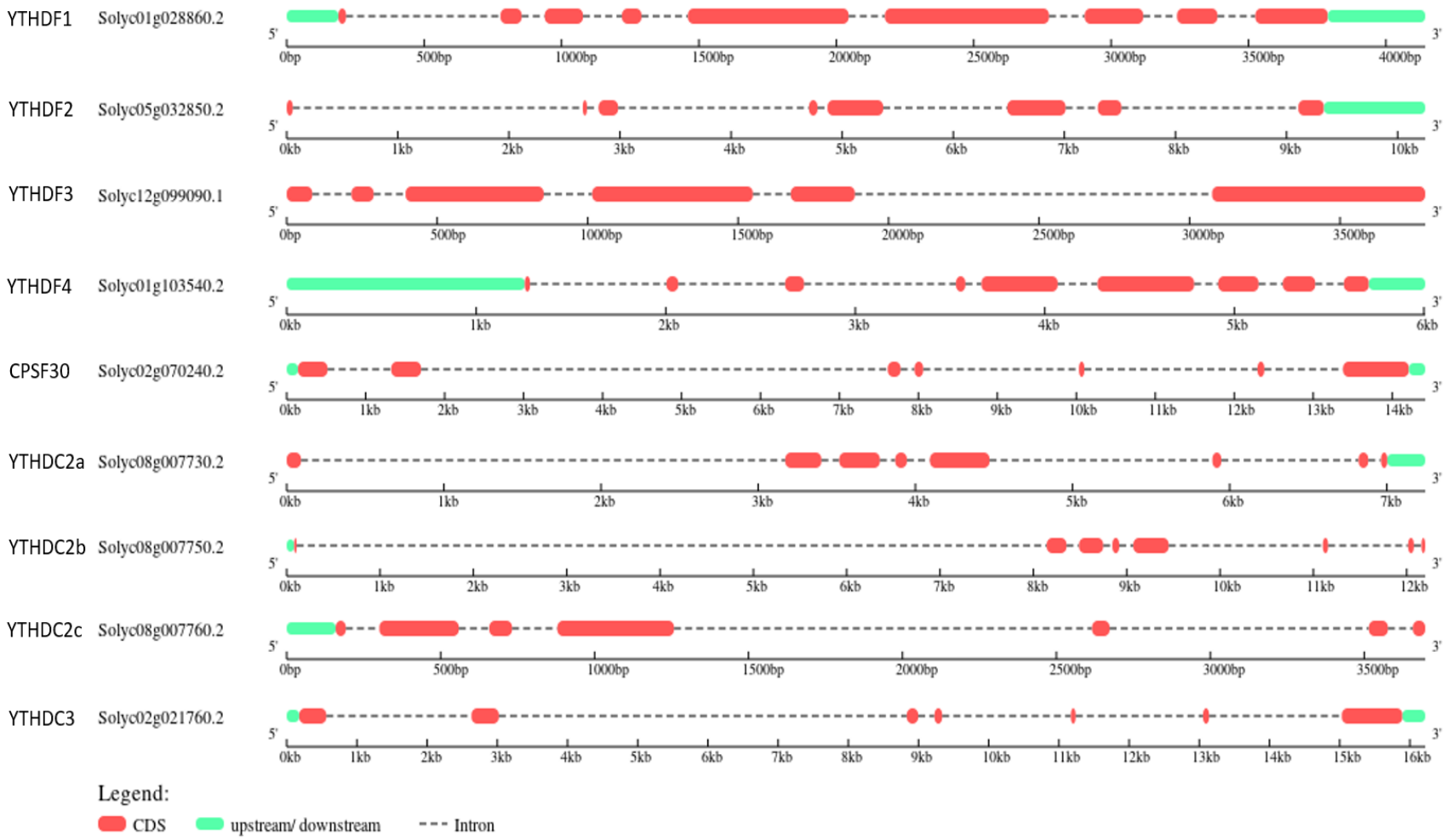


Figure 3.5: Intron-Exon distribution of all YTH genes found in *Solanum lycopersium*. UTRs (Green), CDS (Red) and Introns (Dotted lines).

Table 3.3: YTH proteins of *Solanum lycopersicum*. Characteristics and gene codes for all YTH proteins identified in the genome of tomato.

Genes	Loci	Size(aa)	Genomic positions	MW (Da)	p*I*
SlYTHDF1	Solyc01g028860.2.1	706	SL2.50ch01:41111999..41116141	77223.5	6.95
SlYTHDF2	Solyc05g032850.2.1	604	SL2.50ch05:43986183..43996430	65970.9	5.31
SlYTHDF3	Solyc12g099090.1.1	690	SL2.50ch12:66365360..66369142	75321.5	6.45
SlYTHDF4	Solyc01g103540.2.1	555	SL2.50ch01:92127117..92133123	61494.5	8.65
SlCPSF30	Solyc02g070240.2.1	671	SL2.50ch02:39989910..40004327	73772.3	6.1
SlYTHDC2a	Solyc08g007730.2.1	395	SL2.50ch08:2230630..2237873	44219.3	6.39
SlYTHDC2b	Solyc08g007750.2.1	369	SL2.50ch08:2251358..2271383	41090.1	6.52
SlYTHDC2c	Solyc08g007760.2.1	299	SL2.50ch08:2275280..2278977	33856.8	6.32
SlYTHDC3	Solyc02g021760.2.1	689	SL2.50ch02:23611480..23627697	75934.3	6.24

3.2.7 DF-type YT521-B domain containing proteins in *Solanum lycopersicum*

The value of using *Solanum lycopersicum* as a model for the study of YTH proteins, comes from the smaller number of YTHDF proteins found within its genome. This means that redundancy between the proteins is far less likely and studies involving knockouts become much simpler. The focus of this thesis will be these proteins.

Alignment of the YTHDF proteins (Figure 3.6) show the high conservation of the YTH domain (residues 487- 657) making the YTH domain in tomato around 170aa in length. There are 102 absolute conserved residues (12.1%) with a pairwise identity of 39.2% with 83 of these being found within the YTH domain (BLOSUM62 pairwise identity 82.1%). This extremely high level of conservation within the YTH domain and relatively low level of conservation outside of this domain, is indicative of the YTH protein family. As has been noted by numerous studies, the YTH domain is essential to the protein m^6A binding ability, but it is the remainder of the protein which mediates is localisation, function and structure (Zhu *et al.*, 2014; Arribas-

Hernández *et al.*, 2018; Scutenaire, J. M. Deragon, *et al.*, 2018; Berlivet *et al.*, 2019). The conserved residues which form the aromatic m^{6}A binding cage (section 3.2.3), are located at residues 509, 570 and 575 and the variable residue at position 575 are all Tryptophan residues making the aromatic cage comprised of WWW. This is the most common combination for YTHDF YTH domains (Berlivet *et al.*, 2019).

Overall, all four proteins are divided into two regions, highly conserved YTH domain and the remainder of the proteins, which, exhibit a very low level of similarity. All four YTHDF proteins have been predicted to utilise the same W residues for m^{6}A binding and therefore are likely to be *bone fide* m^{6}A binding proteins

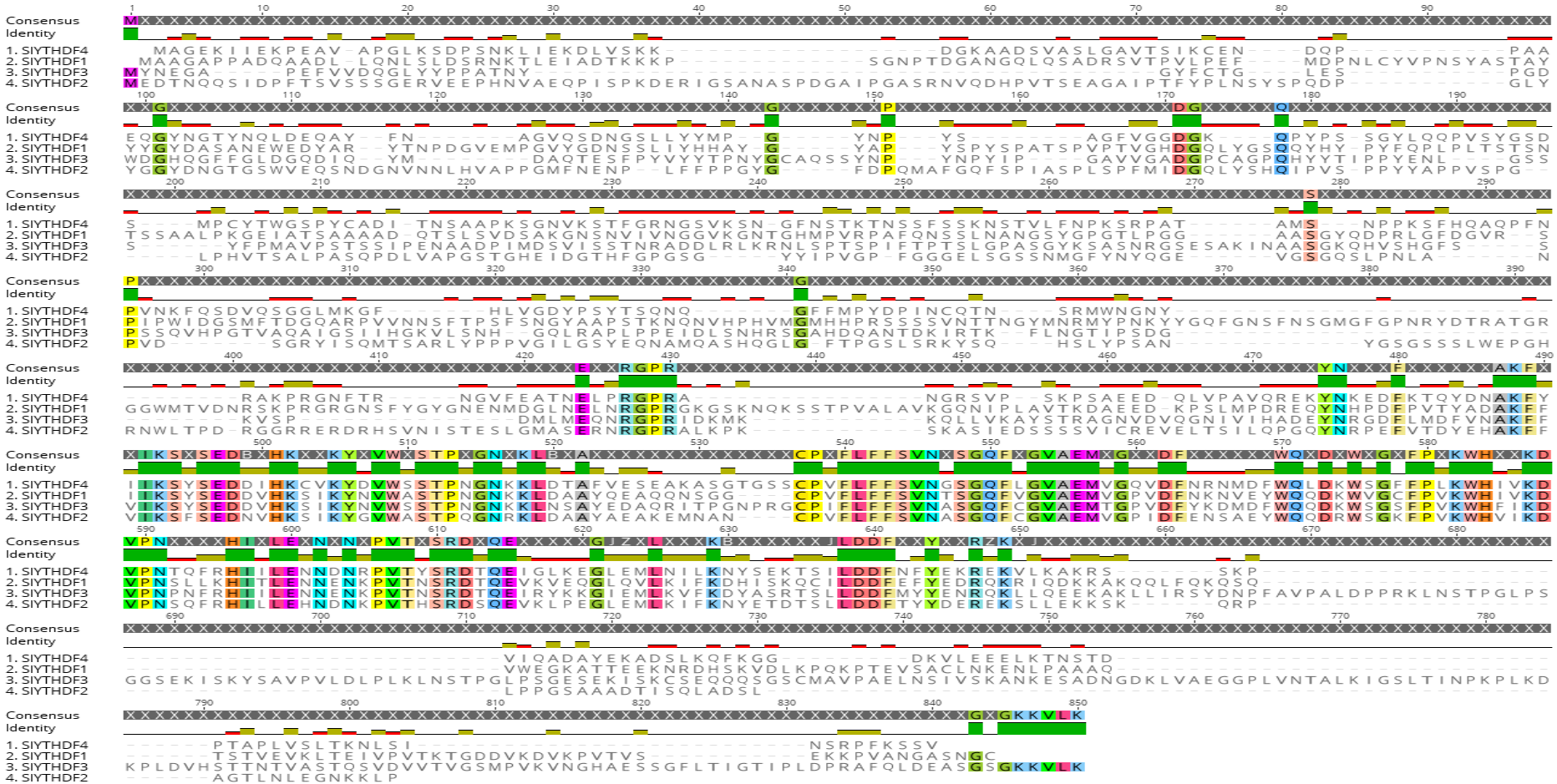


Figure 3.6: Multiple sequence alignment of YTHDF proteins from *Solanum lycopersicum*. Aligned using MUSCLE 3.8.4. Residues with 100% conservation highlighted

3.2.7.1 Intrinsic disorder of YT521-B domain containing proteins in *Solanum lycopersicum*

As in mammalian YTHDF proteins, The N-terminus of ECT2, ECT3 ECT4 in *Arabidopsis* contains a large intrinsically disordered region (IDR). Studies on the human YTHDF2 have suggested the two regions of the protein (IDR and YTH domain) operate independently functioning as regulation and m^6A binding respectively (Zhu *et al.*, 2014; Patil, Pickering and Jaffrey, 2017). These IDRs present in YTHDF proteins could also function or participate in RNA binding. The ability of YTHDF2 to degrade translationally repressed mRNA within the cytoplasm often occurs within processing bodies (P-bodies). The IDR of YTHDF2 is sufficient to drive YTHDF2 into these P-bodies (Zhu *et al.*, 2014; Du *et al.*, 2016).

This reflects work which demonstrated that low complexity sequences such as the IDRs repeatedly localise in RNA granules, but whose function are currently unknown (Kato *et al.*, 2012).

We analysed the potential of tomato YTHDF proteins to localise into these granules or the ECT2/ECT4 foci observed by the Broderson lab (Arribas-Hernández *et al.*, 2018). Protein sequences were fed into a prediction algorithm PONDR (Predictor of Naturally Disordered Regions- VL3) (Figure 3.7).

Like IDRs in *Arabidopsis* YTH proteins, the tomato YTHDF proteins have variable amounts of disorder, which affect predictions. The distribution of disorder across the sequence of each YTHDF protein, changes for each sequence. The YTH domain (indicated) has a high level of order across the entirety of its length. The remainder of the proteins exhibit a relatively high level of disorder. Taking an average of the proteins termed the cumulative distribution function (CDF) (Figure 3.7-B) can be used to predict whether these proteins are likely to aggregate into foci or droplets, which can mediate the protein's function and localisation under varying conditions. Proteins which have CDF curves above the boundary are disordered and those below predicted as ordered (Xue *et al.*, 2010; Kato *et al.*, 2012; Arribas-Hernández *et al.*, 2018). YTHDF1 and YTHDF2 are both highly disordered, with a level of confidence in both the YTH domain and the disordered regions (FIGURE

3.8-A) of YTHDF1 and YTHDF2 confirm that they are in fact disordered proteins and are thus likely to form assemblages similar to ECT2 (Arribas-Hernández *et al.*, 2018). YTHDF3 has one disordered region outside of the YTH domain but the N and C termini are ordered, albeit with low confidence levels. Overall, the CDF is only slightly above the line so is considered both an ordered and disordered protein. YTHDF4 has three disordered regions outside of the YTH domain. These have low confidence, and the ordered regions cover the majority of the protein sequence. CDF predictions place its score below the boundary line, predicting that it is unlikely to be found in foci independently.

Overall, the role of YTH proteins in the diverse range of droplets, foci and assemblages, within plant species is almost completely unknown. These predictions can provide guidance into the study of the aggregations. It is likely that YTHDF1 and YTHDF2 both localise into these aggregations under specific conditions. The growing number of examples of the relevance of m^6A and YTH proteins in cellular aggregations, foreshadow interesting roles for these proteins in tomato cellular mechanisms.

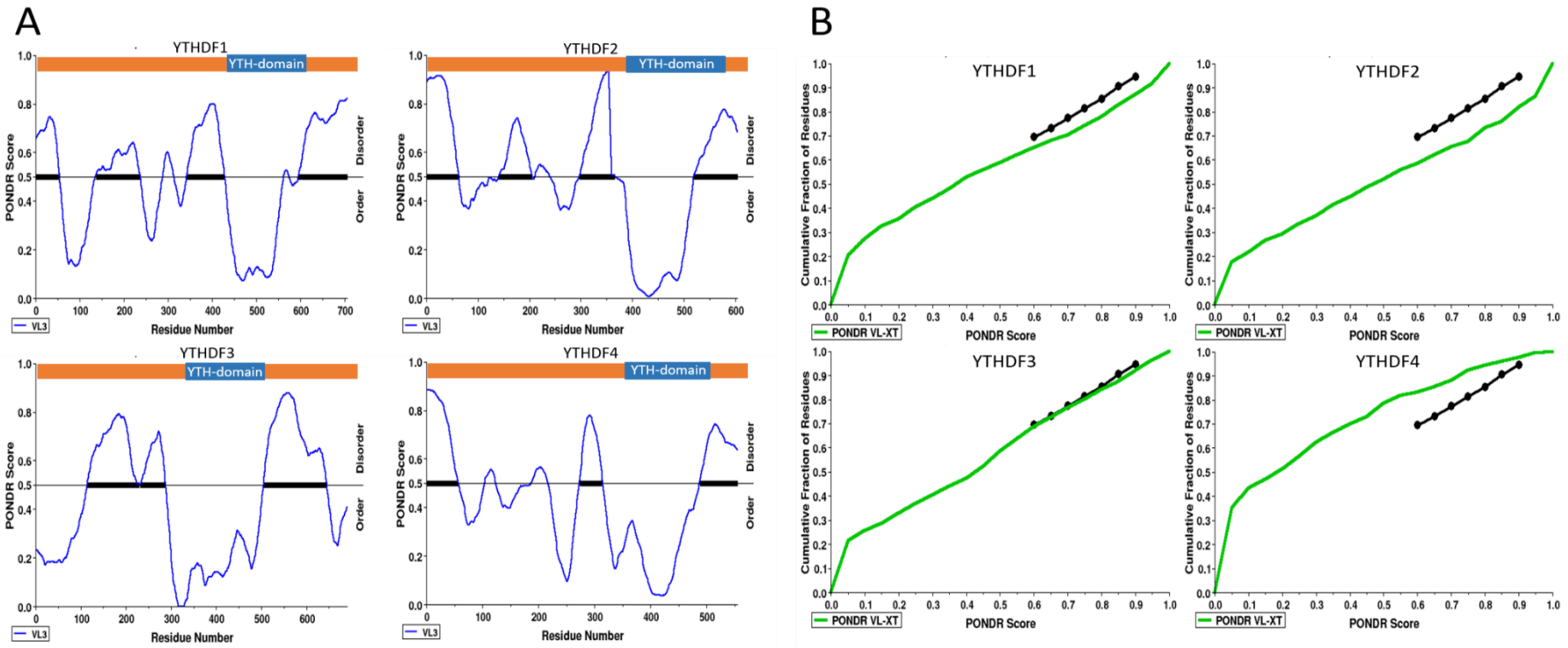


Figure 3.7: Intrinsic disorder region predictions of YTHDF1, YTHDF2, YTHDF3 and YTHDF4. (A) Disorder distributions PONDER-VL3 (B) Cumulative Distribution Functions (CDF). Predictions generated by PONDR-VL3 (predictor of Naturally Disordered Regions-VL3). The Cumulative Distribution Function (CDF) applied to PONDR® VL-XT predictions is a cumulative histogram of the PONDR® scores for a given protein (Green line). Proteins that have CDF curves (Green line) above the boundary (Black line) (i.e., a high proportion of low scores) were predicted to be ordered, proteins with curves below the boundary (i.e., a high proportion of high scores) are predicted to be disordered

3.3 Discussion

YTH proteins within *Plantae* are diverse and highly expanded. The large number of these protein, the YTHDF group, and the high level of conservation within each species, increases the potential for redundancy and overlapping functions of each protein (Berlivet *et al.*, 2019). The variation in the number of YTHDF and YTHDC type proteins lead us to believe that there was likely to be species that were better suited for comprehensive analysis of the family and its overall functions. Information on the YTH family and their biological functions is limited, any advance in our understanding would be of great value to future studies.

The aim of this chapter was to collect and analyse all YTH proteins found, within model plant species, across *Plantae.* The data collected would then be searched to find optimal candidates for a future m^6A-YTH protein model. 286 YTH proteins were identified across 29 plant species, with an average of 1.93 YTHDC proteins and 7.86 YTHDF proteins (Table 3.1) This analysis highlighted the potential of Tomato (*Solanum lycopersicum)* as an ideal model for the study of YTHDF proteins. It also contains an YTHDC group of high interest. Tomato genomes contain 5 YTHDC proteins and 4 YTHDF proteins. The 4 YTHDF proteins is significantly lower than that seen in other key model species, such as, *Arabidopsis thaliana* (11- YTHDF) and *Oryza sativa* (12- YTHDF).

The YTHDC group in tomato seems to be unique among plant species in having more members than the YTHDF group. In addition, it is the only species observed in our study to have a triplicated YTHDC2 gene. Although, expression has not been quantified and it is possible only one or two of these genes are functionally expressed.

The YTHDF group was aligned, and a tree of similarity was developed to predict the closest homologs of each tomato YTHDF protein. (Figure 3.4). The high level of conservation of YTH domains across Eukaryota has led many to speculate the functional significance of the protein residues and sequences outside of the YTH domains (S. Luo and Tong, 2014; L.-H. H. Wei *et al.*, 2018). In fact, these external regions are less conserved even when

analysed together with protein from the same species. These observations mean that the placement of the YTH proteins on the tree are predominantly as a result of the residues outside of the YTH domain, increasing the importance of these residues. Although this analysis did allow us to predict the closest homologs of tomato in *Arabidopsis,* somewhat confirming those found in a previous study (Scutenaire, J. M. Deragon, *et al.*, 2018) the evolutionary noise caused by regions outside of the YTH domain reduce the resolution of the tree to a degree where we cannot observe any functional groups.. The extremely high level of conservation within the YTH domain of tomato proteins (Figure 3.2) and relatively low level of conservation outside of this domain, is characteristic of the YTH protein family. This has been noted by numerous studies, the YTH domain is essential to the protein m^6A binding ability, but it is the remainder of the protein which mediates its localisation, function and structure (Zhu *et al.*, 2014; Arribas-Hernández *et al.*, 2018; Scutenaire, J. M. Deragon, *et al.*, 2018; Berlivet *et al.*, 2019). We agree with the hypotheses that the regions outside of the YTH domain, mediate the specific functions of the YTH proteins, while the YTH domain is essential for the m^6A affinity (D. Li *et al.*, 2014; S. Luo and Tong, 2014; Theler *et al.*, 2014; Xu, Liu, *et al.*, 2014; Zhu *et al.*, 2014). Previous attempts to characterise plant YTH domain proteins and bioinfomatically predict their functionality have been based on sequence similarities of the YTH domain alone (D. Li *et al.*, 2014; L.-H. H. Wei *et al.*, 2018; Scutenaire, J.-M. Deragon, *et al.*, 2018). It is possible these analyses we limited to the YTH domain for the same reason that our tree gives limited information. It could be of value to revisit this analysis in future work.

The four YTHDF proteins are located on chromosomes I, V and XII. The singly distributed YTHDF3 and YTHDF2 have likely evolved independently, and therefore are also likely to have divergent functions. The intron-exon distribution is also unique between YTHDF2, YTHDF3 and YTHDF1/4. YTHDF1 and YTHDF4 are both located on chromosome I, although they are located 50kb apart, due to some similarities in intron-exon distribution it is possible that these two proteins arose from segmental duplication and have since evolved separately. The triplicated YTHDC2a-c proteins are clustered within 2kb of each other. As expected, the intron-exon distribution of YTHDCa and YTHDCb are highly similar, unexpectedly

however, YTHDC2c has an intron-exon distribution with reduced intron lengths (Figure 3.5). Analysis of the absolute conserved residues found in Tomato YTH domains, indicate the YTHDF YTH domains likely rely on the WWW aromatic cage common to almost all YTHDF proteins (F. Li *et al.*, 2014). Interestingly, YTHDC proteins SlYTHDC2a and DC2b also have the WWW residues, which is different to the L present in Human YTHDC1 and YTHDC2 (Xu *et al.*, 2015). Even more surprisingly, SlCPSF30, YTHDC3 and YTHDC2c are predicted to utilise the Y, similar to the aromatic cage in Yeast m^{6}A protein PHO92 (Figure 3.2) (Kang *et al.*, 2014) The difference in this highly important 3rd residue in the triplicated YTHDC2a, b and c proteins has never been reported in the literature and this, coupled with the differences in intron lengths and distribution could suggest that the proteins have diverse functions *in vivo*. The presence of these essential residues means the tomato YTH proteins are likely to bind m^{6}A. The significance of any differences in residues between plant YTH domains is yet to be uncovered. Confirmation of the functional importance of these residues would confirm that any phenotypes observed, require the protein's ability to bind m^{6}A. Whether these proteins have any roles which do not require the binding of m^{6}A has never been reported. In addition, whether these domains in tomato YTH proteins exhibit affinity to other RNA modifications or even DNA modifications would also be a great subject for future study.

Understanding cellular localisation of RNA binding proteins is essential to understand their function. A growing list of studies have implicated YTH domain proteins in the cellular response to stressful conditions. In mammalian cell lines, m^{6}A is additionally added to multiple translationally active mRNAs under stress (oxidative and heat). This induces 'triaging' to foci known as stress granules (SGs). The formation of these granules was later shown to be reliant on all three YTHDF proteins (DF1-3). Similar induction of foci has been observed in *Arabidopsis,* and similarly linked to IDR of YTH protein N-termini (Arribas-Hernández *et al.*, 2018). We have discovered that all four YTHDF proteins in tomato have IDRs but only YTHDF1 and YTHDF2 have been predicted to form assemblages *in vivo* (Figure 3.7). The importance of these assemblages for numerous biological processes are likely shared by tomato (and *Arabidopsis*) systems.

Overall, these initial observations demonstrate the potential of *Solanum lycopersicum* as a novel and highly effective model for the study of YT521-B proteins. Tomato YTH proteins share many of the characteristics which make YTH proteins such an exciting family to study. These characteristics also can make the study challenging, with such a vast range of functions *in vivo*. It's this level of complexity which leads us to choose tomato to uncover these roles.

Chapter 4 - Characterisation of YT521-B proteins in *Solanum lycopersicum*

4.1 Introduction

Tomato (*Solanum lycopersicum*) is an ideal candidate for any genomic, transcriptomic or proteomic model. Through its importance as a globally relevant food crop (fresh and processed), the availability of efficient transformation methodology (van Eck, 2018), the diploidy of its genome and an increasingly high-quality sequence database (Tomato Genome Consortium, 2012), tomato is recognised as key model crop for plant development studies.

Development of transgenic plants is a widely used technique for the investigation of specific gene functions. Genome editing strategies have been developing rapidly in both efficiency and flexibility. The most prevalent of these is known as Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/Cas9 (Belhaj *et al.*, 2013).

CRISPR/Cas9 systems have been developed from prokaryotic adaptive immune systems, the biological function being to confer resistance to invasive nucleic acids from viral genomes or plasmids (Jinek *et al.*, 2012). The CRISPR/Cas9 system has 3 main components: CRISPR-associated9 endonuclease, a CRISPR RNA (crRNA) and a trans-activating crRNA (tracrRNA). The crRNA contains a short spacer sequence, which in nature is captured from invading nucleic acids and subsequently stored. This 'guide' hybridises with the tracrRNA before forming a complex with the Cas9 (Jinek *et al.*, 2012). The recognition of nucleic acids which is utilised in CRISPR/Cas9 mediated genome editing is dependent on pairing between the crRNA and a corresponding 'protospacer' sequence and a protospacer adjacent motif (PAM) canonically NGG, immediately upstream of the protospacer. This recognition is followed by Cas9 cleavage of the paired DNA and subsequent dissociation (Brooks *et al.*, 2014)

The repurposed CRISPR/Cas9 system utilised for genome editing, contains a single synthetic guide RNA (gRNA), formed from fused crRNA and tracrRNAs (Jinek *et al.*, 2012). Targeted mutagenesis is achieved by producing double strand breaks within the targeted DNA. These double strand breaks are typically generated 3bp upstream of the PAM site. The error prone non-homologous end joining (NHEJ) pathway often leads to a small insertion or deletion mutation. Such mutations can shift the open reading frame of a gene coding sequence, potentially introducing premature stop codons, resulting in gene knockouts.

In plants, CRISPR/Cas9 has been utilised in the majority of plant models and was demonstrated in *Solanaceae*, as early as 2013 (van Eck, 2018). Using *Agrobacterium tumefaciens*-mediated transformation, we utilised CRISPR-Cas9 genome editing in an attempt to generate gene knockout lines for each of the four YTHDF proteins in Tomato. Confirmed mutant lines were then screened for the presence of the T-DNA, and selfed to remove the inserted CRISPR/Cas9 cassette. Subsequent lines will inform on any potential phenoptypes visible in these gene knockouts. Observations of any adverse effects on morphology, physiology or responses to abiotic stress, will uncover potential roles for these four proteins. The advances in our knowledge in this area will shed light on underlying mechanisms through which m^6A functions, within plant species.

4.2 Results

4.2.1 YTHDF gRNA and vector design

Using full-length genomic sequences (NCBI- Solyc01g028860.2.1, Solyc05g032850.2.1, Solyc12g099090.1.1, Solyc01g103540.2.1) we searched for DNA targets as close to the 5' end of each gene as possible and searching solely within exons. Each target was of the consensus $(G)_{20}$-NGG (Figure 4.1). Finding targets closer to the 5' end of each gene increases the chance of full gene KOs, after CRISPR/Cas9 induced frame-shift mutations (Brooks *et al.*, 2014). The U6 promoters utilised in the CRISPR/Cas9 constructs,

have a defined transcription start site, which is a "G". Finding targets which start with a G, increases the efficiency of transcription and therefore increases the likelihood of mutations (Belhaj *et al.*, 2013). Constructs were assembled using a combination of PCR and golden gate cloning, using a system developed by the Marillonnet and Nekrasov labs (Weber *et al.*, 2011; Nekrasov *et al.*, 2013). It is advantageous to find targets which have restriction sites overlapping the Cas9 cut site, known to be ~3bp upstream of the PAM site (NGG). Only two of the four suitable targets (YTHDF1 and YTHDF4) contained restriction sites in a position which overlaps this region. In instances where these restriction sites are present, repair of double-strand breaks incorrectly repairs by NHEJ, would result in disruption of the restriction sites. Amplification across this region and digestion of subsequent amplicons can identify induced mutations. Finally each selected DNA target was processed to determine the likelihood of off-target effects, using Geneious 10.5.1.

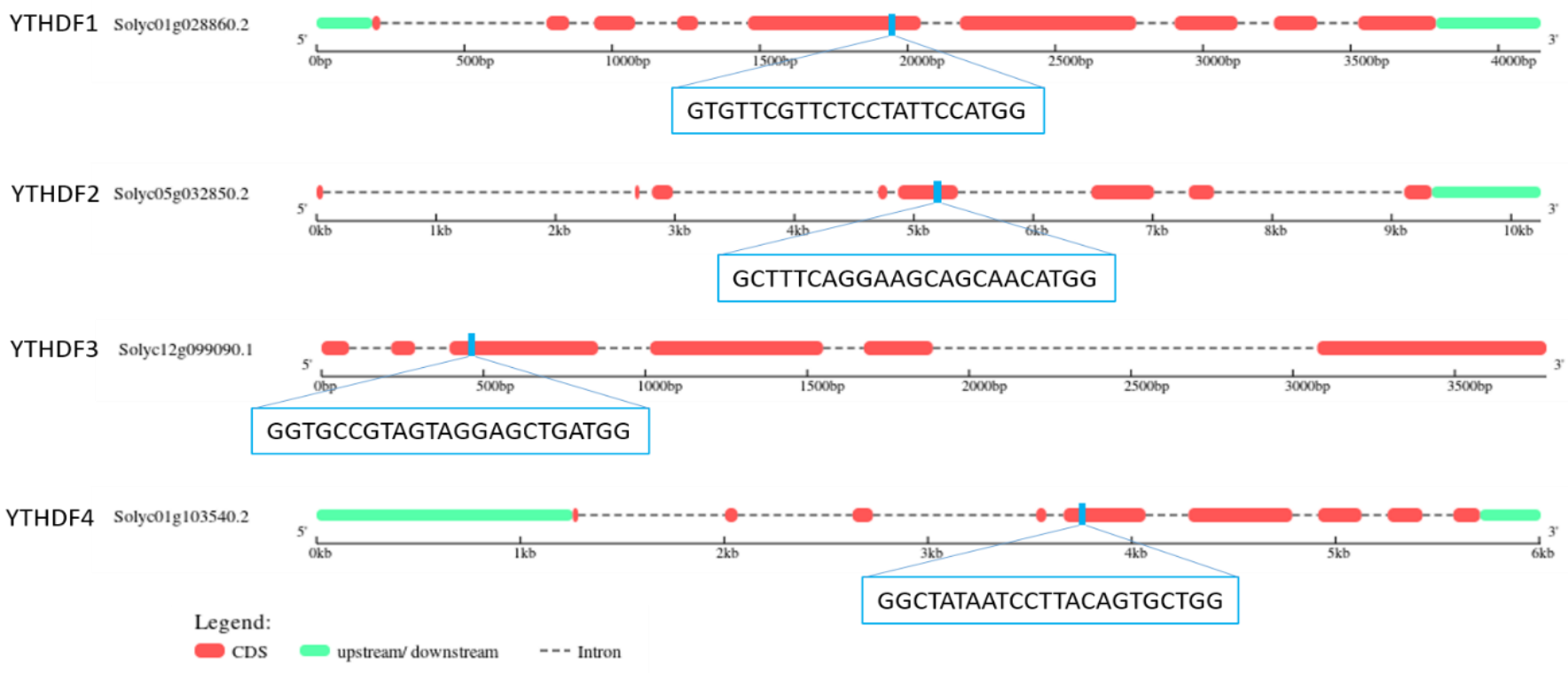


Figure 4.1: sgRNA DNA target sites for CRISPR/CAs9 gene disruption systems within four YTHDF genes in Solanum lycopersium.L. Targets located in CDS of each gene (Red) close to 5'UTR (Green). Dotted lines indicate introns.

Selected targets were cloned into an efficient single guide CRISPR/Cas9 vector platform (Methods 2.5). SgRNAs were amplified using the template described in Mali *et al.* as a template. Each of these guide amplicons were then cut-ligated (BsaI) into level 0 vector pICSL01009::AtU6p and level 1 vector pICH47751 (Methods 2.5.1.2). Level 1 vectors pICH47732::NOSp:NPTII, pICH47742::35Sp::Cas9-NOST, pICH41766 and sgRNA vectors AtU6p:sgRNA were cut ligated into level 2 vector pAGM4723 to produce pAGM4723::AtU6p::sgRNA1::Cas9::NOST with NPTII selection (Methods 2.5.1.3). The resulting plasmids for YTHDF1-4 were named pCGYTHDF(1-4)-NPTII-Cas9-sgRNA (2.4.1).

4.2.2 Generation of single YTHDF gene KOs

For each of the four plasmids described above, we transformed more than 500 cotyledons of *Solanum lycopersicum. L* 'Ailsa Craig', generating between 16 and 31 transgenic lines. Each transgenic line was screened for the presence of the inserted t-DNA to confirm transformation before collecting T_1 seed. T_0 lines were screened for CRISPR/Cas9 induced mutations using a combination of PCR and amplicon restriction or PCR and fragment sequencing. DNA from lines with potential mutations were cloned into pJET vectors in *E.coli*, purified and resulting colonies sequenced to confirm mutation.

To ensure no mutations were missed, seeds from all T_0 plants were collected and re-screened in the T_1 generation. Confirmed mutations from T_0 were also rescreened to confirm heritability of the produced mutations. Of the four YTHDF CRISPR/Cas9 transgenic lines, at least one confirmed mutation was produced (ythdf1-1, ythdf2-1, ythdf2-2, ythdf3-1, ythdf3-2, ythdf4-1 and ythdf4-2) (Table 4.1; Figure 4.2). All induced mutations produced a frame-shift mutation, as a result of either single base insertions or base deletions (not divisible by 3). Translation predictions show that all generated lines result in truncated proteins resulting from a mutation induced premature stop codon (Table 4.1). Confirmed YTHDF disruption lines were carried into T_2. T_2 plants were sown on MSR3 media containing 50mg/L Kanamycin. Seedlings visibly affected by the kanamycin (and

therefore potentially with the T-DNA crossed out) were rescued and transplanted into compost to recover. Surviving plants were then screened for the presence of the t-DNA carrying the Cas9-NPTII. Plants no longer carrying the t-DNA were kept for the collection of T_3 seed. Only clean T_3 or T_4 seed, absent of any homozygous or hemizygous t-DNA insertions were used for subsequent experiments. To the best of our knowledge these mutant lines represent the first instance of all YTHDF null mutants of a single plant species. This is a significant milestone in growing our understanding of this important protein sub-family.

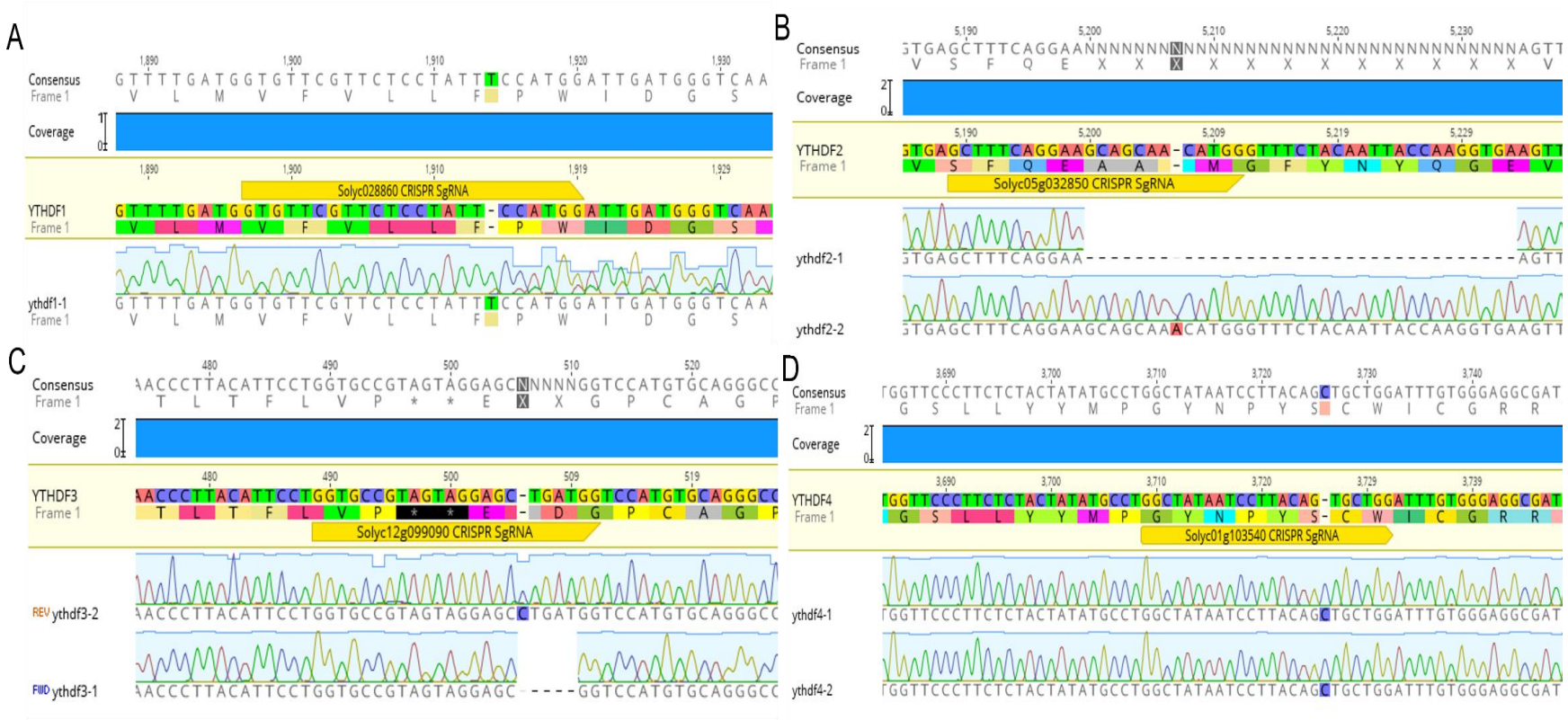


Figure 4.2: Mutation detection and consensus sequences of CRISPR/CAs9 mediated mutations. (A) Sanger sequence read for ythdf1-1 line (B) Sanger sequence reads for ythdf2-1, ythdf2-2 mutant lines (C) Sanger sequence reads for ythdf3-1, ythdf3-2 mutant lines (D) Sanger sequence reads for ythdf4-1, ythdf4-2 mutant lines. sgRNA target guides indicated (yellow arrow).

Gene Name	Loci	Length (bp)	Line Name	Mutation	Position	Wild Type Protein Length (aa)	Predicted Protein Length (aa)	Description
SlYTHDF1	Solyc01g028860.2.1	4144	ythdf1-1	T insertion	1913-1914	706	259	Single T insertion between positions 1913 and 1914. Resulting in a frame-shift and premature stop codon.
SlYTHDF2	Solyc05g032850.2.1	10214	ythdf2-1	34bp deletion	5200-5234	604	235	34bp deletion between positions 5200 and 5234. Resulting in a frame-shift and premature stop codon.
SlYTHDF2	Solyc05g032850.2.2	10249	ythdf2-2	A insertion	5206-5207	604	236	Single A insertion between positions 5206 and 5207. Resulting in a frame-shift and premature stop codon.
SlYTHDF3	Solyc12g099090.1.1	3784	ythdf3-1	C insertion	505-506	690	90	Single C insertion between positions 505 and 506. Resulting in a frame-shift and premature stop codon.
SlYTHDF3	Solyc12g099090.1.1	3779	ythdf3-2	4bp deletion	505-509	690	131	4bp deletion between positions 505 and 509. Resulting in a frame-shift and premature stop codon.
SlYTHDF4	Solyc01g103540.2.1	6008	ythdf4-1	C insertion	3726-3727	555	140	Single C insertion between positions 3726 and 3727. Resulting in a frame-shift. and premature stop codon.
SlYTHDF4	Solyc01g103540.2.1	6008	ythdf4-2	C insertion	3726-3727	555	140	Single C insertion between positions 3726 and 3727. Resulting in a frame-shift and premature stop codon.

Table 4.1: Description and resulting protein length prediction of all generated single mutant lines of YTHDF proteins in *Solanum lycopersicum*

4.2.3 Characterization of ythdf loss-of-function mutants

The developed mutant lines described above provide a platform for the identification of growth or developmental phenotypes linked to ythdf1-4 loss-of-function mutants. As potential m^6A readers, it is likely that the YTHDF proteins perform some physiological roles *in vivo*. *Arabidopsis* reduced m^6A lines, exhibit phenotypic changes in leaf and flower structures, consistent with reduced apical dominance and changed endo-reduplication (Bodi *et al.*, 2012a). With the discovery triple mutant *ect2;ect3;ect4* in *Arabidopsis* display similar, albeit less severe phenotypes to those observed in low methylation *Arabidopsis* lines (Zhong *et al.*, 2008; Růžička *et al.*, 2017; L.-H. H. Wei *et al.*, 2018) *Ect2* and *ect3* loss-of-function mutants, also exhibit increased trichome branching (branched single cells found on the epidermis of *Arabidopsis*) (Arribas-Hernández *et al.*, 2018; L.-H. H. Wei *et al.*, 2018; Scutenaire, J.-M. Deragon, *et al.*, 2018). This was shown to be mediated by degradation and down-regulation of three trichome specific transcripts, in *ect2* disruption lines (Arribas-Hernández *et al.*, 2018). It is reasonable to expect that ythdf loss-of-function mutants will exhibit a range of similar phenotypes.

4.2.3.1 Physiological characterisation of ythdf null mutants in nutrient solution

Each ythdf line was grown to 4 weeks, with a minimum of 20 repeats. Plants were grown in a growth room under controlled conditions in modified Hoagland's solution (Methods 2.1.3). Use of this growth method allows for repeatable experimental set up and provides the opportunity to modify the solution in future abiotic response experiments. At the end of 4 weeks, measurements were made for key growth metrics (Figure 4.4).

Ythdf null mutants exhibit a range of physiological and morphological difference when compared with wild-type (WT) (Figure 4.3, 4.4). At 4 weeks, *ythdf1, ythdf2* and *ythdf4* all exhibit significantly reduced height ($p<0.0001$, $p<0.0001$ and $p<0.0005$, respectively), with *ythdf2* KO having the most pronounced effect. Comparisons to WT show a 23.6% decrease in height for

ythdf1, a 39.1% decrease in *ythdf2,* a 7.4% increase in *ythdf3* lines and a 29.6% decrease in the height of *ythdf4. Ythdf2* is also significantly shorter than *ythdf1* (p=0.0086) but not *ythdf4.* The increase in height of *ythdf3* has a mean slightly higher than WT, but this is not significant. The root lengths are also affected in the null mutants.

Root lengths are significantly reduced in both *ythdf1* (-19.9%) and *ythdf2 (*-18.7%), when compared to WT ($p<0.0001$ and $p<0.005$, respectively). Interestingly *ythdf4,* does not exhibit significantly reduced root length despite having clear stunted growth in vegetative tissues. *Ythdf3,* again exhibits similar root lengths to WT.

The length of leaves 3 and 4, for *ythdf1* (-18.7%), *ythdf2* (-47.8%) and *ythdf4* (-24.6%) are all significantly lower than the leaves of WT ($p<0.0001$, $p<0.0001$, $p<0.0005$, respectively). *Ythdf2* leaves are significantly shorter than *ythdf1* leaves but not *ythdf4.* Ythdf3 null lines showed significantly longer true leaves with leaves being 21.7% longer (p=0.0075). Leaf widths show that again *ythdf1, ythdf2* and *ythdf4* leaves are significantly less than WT ($p<0.05$, $p<0.0001$ and $p<0.05$). *ythdf3* leaf widths do not differ significantly from WT. Overall, all *ythdf* mutant lines show physiological differences in leaf size, to varying degrees. *Ythdf1, 2* and *ythdf4* with reduced leaf size and *ythdf3* with an overall increase, but this is mostly due to increased leaf length, although there are observable differences in leaf width, these figures were not found to be significant (Figure 4.3, 4.4).

The number of lateral branches (formed from auxillary meristems) in plants at 4 weeks, can be used to suggest any changes in apical dominance when compared to WT. *ythdf1, ythdf2* and *ythdf4* mutant lines exhibit some degree of increase in lateral branching. This is most obvious is *ythdf2* (Figure 4.3, 4.4) Only *ythdf1, ythdf2* and *ythdf4* have significantly more lateral branching when compared to WT ($p<0.0001$, $p<0.0001$ and $p<0.005$). The number of compound leaves, and the subsequent number of nodes on each plant can also hint at any roles of these proteins in development. In this case, *ythdf2, ythdf3* and *ythdf4* have significantly more leaves than WT with increases of 59.2%, 34.1% and 32.0%, respectively ($p<0.0001$, $p<0.005$ and $p<0.0001$, respectively). Linked with the number of nodes in the distance between each node, this along with height can be used to indicate

stunted growth physiologies associated with the KO of each single gene. As expected, due to observed loss of height in *ythdf1, ythdf2* and *ythdf4*, as well as increased number of nodes in all mutant lines, internode length are reduced in all lines with 36.3%, 76.8%, 29.8% and 61.1% reductions in internode lengths observed in *ythdf1, ythdf2, ythdf3* and *ythdf4,* respectively ($p<0.0001$, $p<0.0001$ and $p<0.0001$).



Figure 4.3: Wild-type and *ythdf* single mutant lines at 4 weeks. Representative images of wild-type and *ythdf1-4* single mutant lines, grown in modified Hoagland's solution and photographed at 4 weeks. Scale bars at 10cm.

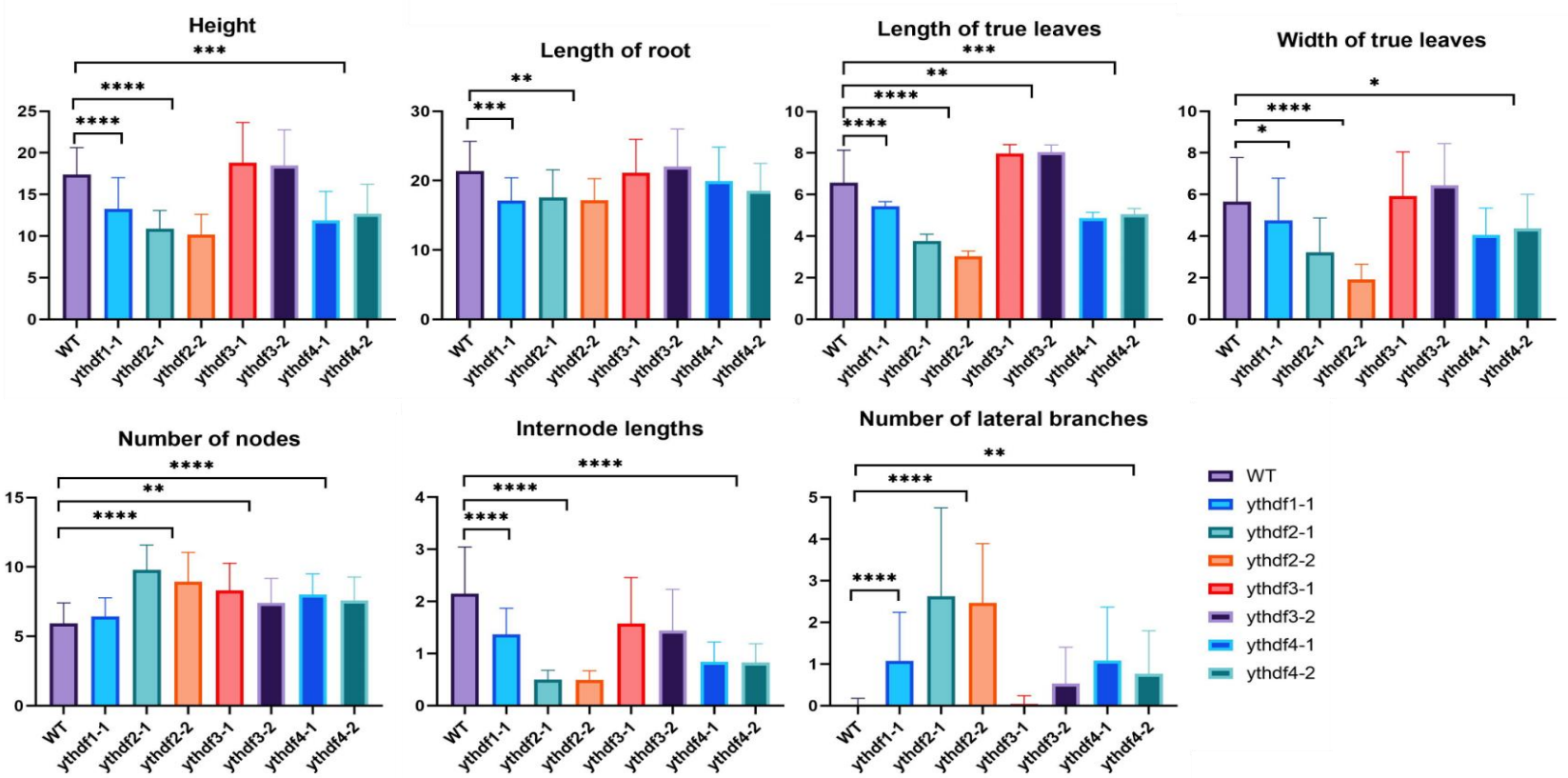


Figure 4.4: Characterisation of *ythdf1-4* single mutant KO lines. Histogram representation of key plant growth measurements. Plants grown in modified Hoagland's solution. Measurements taken at 4 weeks in cm. Error bars representing SD values from 15-20 replicates (n= 15-20). P values were obtained using mixed effects model and Dunnett's multiple comparisons test to determine whether mutant values are significantly different from wild-type values. P value significance bars: * (<0.05), ** (<0.005), *** (<0.0005), **** (<0.0001). No significance bars indicate no significance to WT.

4.3.3.2 Germination rate of *ythdf* loss-of-function mutants

Generation of single mutant lines allowed informal observations to be made about the potential phenotypes associated with loss-of-function *ythdf* mutants. One of these initial observations was changes in the germination rate of seeds of each of the mutant lines. Seed germination is controlled by several mechanisms and is an essential step in the growth of development of plant embryos. Factors controlling germination rate include plant hormones and complex interactions between plant hormones and effector genes can have large effects on germination (Miransari and Smith, 2014). Any changes in germination rate in loss-of-function mutants can be an effective indicator of potential roles, including potential interactions with a number of plant hormones.

To investigate germination rate, 40 seeds (10 seeds per plate) of each line were sown on filter paper soaked with distilled water. Plates were sealed with micropore tape to prevent contamination and placed in standard growth conditions (0 Hr light). All seeds were collected from parent lines of the same age and seeds were harvested on the same day. At the end of 10 days the number of seeds which has penetrated the endosperm cap of the seed coat were counted and germination percentages were calculated (Figure 4.5).

Of the four *ythdf* lines, *ythdf1* and *ythdf4* show significant reductions in germination rate when compared to WT ($p<0.0005$ and $p<0.0001$, respectively). Loss of *ythdf1* and *ythdf4* result in a 18.75% and 36.25% reduction in the percentage of seeds germinated compared to WT seeds. Neither *ythdf2* or *ythdf3* show any significant drop in germination rate in any of the tests or replicates.

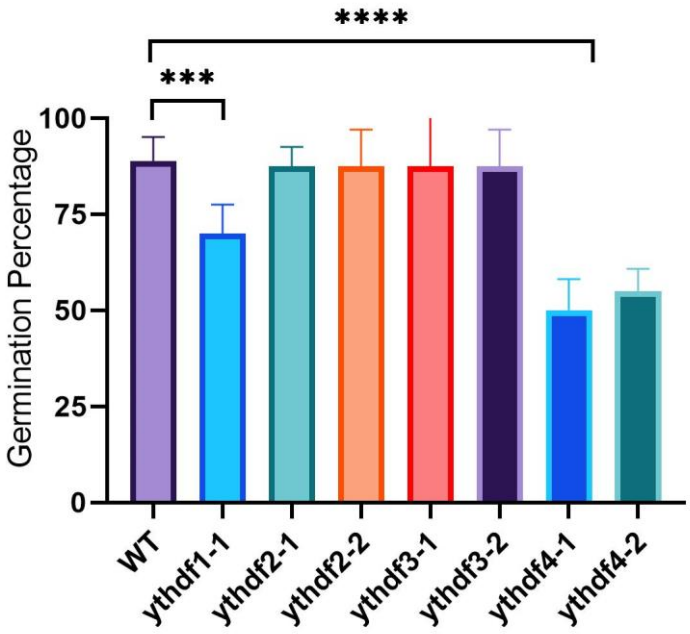


Figure 4.5: Germination rate observed of WT and *ythdf* loss-of-function lines. Seeds germinated on filter paper under standard conditions. Error bars representing SD values from 40 repeats. P values were obtained using ordinary one-way ANOVA and Sidak's multiple comparisons test to determine whether mutant values are significantly different from wild-type values. P value significance bars: * (<0.05), ** (<0.005), *** (<0.0005), **** (<0.0001). No significance bars indicate no significance to WT.

4.2.3.3 Effect of salt stress on *ythdf* loss-of-function mutants

Epitranscriptomic marks within plants have recently been linked to both abiotic and biotic stress responses, from heat stress to regulating responses to viral pathogens (Alaguero-Cordovilla *et al.*, 2018; Scutenaire, J.-M. Deragon, *et al.*, 2018). Even more recently, m^{6}A has been shown to be dynamically deposited on transcripts encoding proteins essential for salt and osmotic stress responses. This m^{6}A deposition increases the stability of these key transcripts (Anderson *et al.*, 2018). The function of m^{6}A is known to be primarily mediated by YTH RNA binding proteins, so it is likely that YTHDF1-4 in Tomato share some role in this response mechanism.

To investigate this, 20 plants per line were grown as above in controlled conditions. As the phenotypes of each independent lines for each YTHDF KO was consistent, seeds were pooled for this and subsequent experiments.

Plants were grown in Hoagland's solution for 2 weeks, at the end of the second week the concentration of NaCl was increased to 50mM. Three days later this was increased to 100mM, and further increased to 150mM after another 3 days. The solution was then refreshed every 3 days until the end of 4 weeks. At 4 weeks, plants were measured for key growth metrics (Figure 4.6, 4.7).

At the end of the salt treatment, *ythdf1, ythdf2* and *ythdf3* heights differ significantly from the untreated measurements (Figure 4.7). *Ythdf1* and *ythdf3* display a reduction in height ($p<0.05$ and $p<0.005$, respectively), the opposite is true of *ythdf2* which shows significantly higher plant height in treated when compared to untreated ($p<0.05$). This is of greater significance since WT grown in high salt does not differ from untreated WT. Similarly, to *ythdf2, ythdf4* shows higher growth in treated plants, but this difference was not found to be significant. Treated y*thdf4* also does not show any significant difference to WT.

Only *ythdf2* indicates any significant difference between treated and untreated samples, with a slight increase in root length under salt stress ($p<0.05$). This is the only line to display longer roots in high salt solution, than under normal conditions. *ythdf1, ythdf3,* and *ythdf4* all show similar root lengths in both solutions.

The size of the 3rd and 4th true leaves indicate some effect of the treatment on vegetative growth of the null mutants in high salt. Lengths of the leaves on *ythdf3* and *ythdf4* lines show opposite reactions to salt treatment (Figure 4.7). *Ythdf3* leaves are significantly shorter in treated plants ($p<0.0001$), while *ythdf4* exhibit longer leaves compared to leaves in untreated conditions ($p<0.0005$). Again, WT does not display any significant changes in leaf lengths, despite slightly longer leaves in treated plants. *Ythdf1* and *ythdf2,* also do not differ between the two data sets. Widths of leaves in treated and untreated WT plants, do not differ from one another. *Ythdf2* exhibits significantly wider leaves post treatment when compared to standard conditions ($p<0.05$)

There were no significant differences between number of nodes in any of the null lines or WT. The average observed are consistent between both treatments. Due to difference in heights of the plants between treated and

untreated, *ythdf2* displayed significantly longer average internode lengths ($p<0.005$), despite no differences in the number of nodes. However even though *ythdf1* and *ythdf3* were significantly shorter after salt treatment, there was no significant difference in the internode lengths (Figure 4.6, 4.7).

Perhaps, most interestingly, the number of lateral branches per plant seems to change under salt stress. *Ythdf1, ythdf2* and *ythdf4*, which exhibit increased lateral branching at 4 weeks under standard conditions, all displayed a decrease to zero lateral branching in treatment conditions. Despite these changes only differences observed between treated and untreated *ythdf2* lines are significant ($p<0.0001$). WT and *ythdf3,* which do not branch at 4 weeks, displayed a low level of branching. Unfortunately, this is not of any significance with the collected data (Figure 4.7).

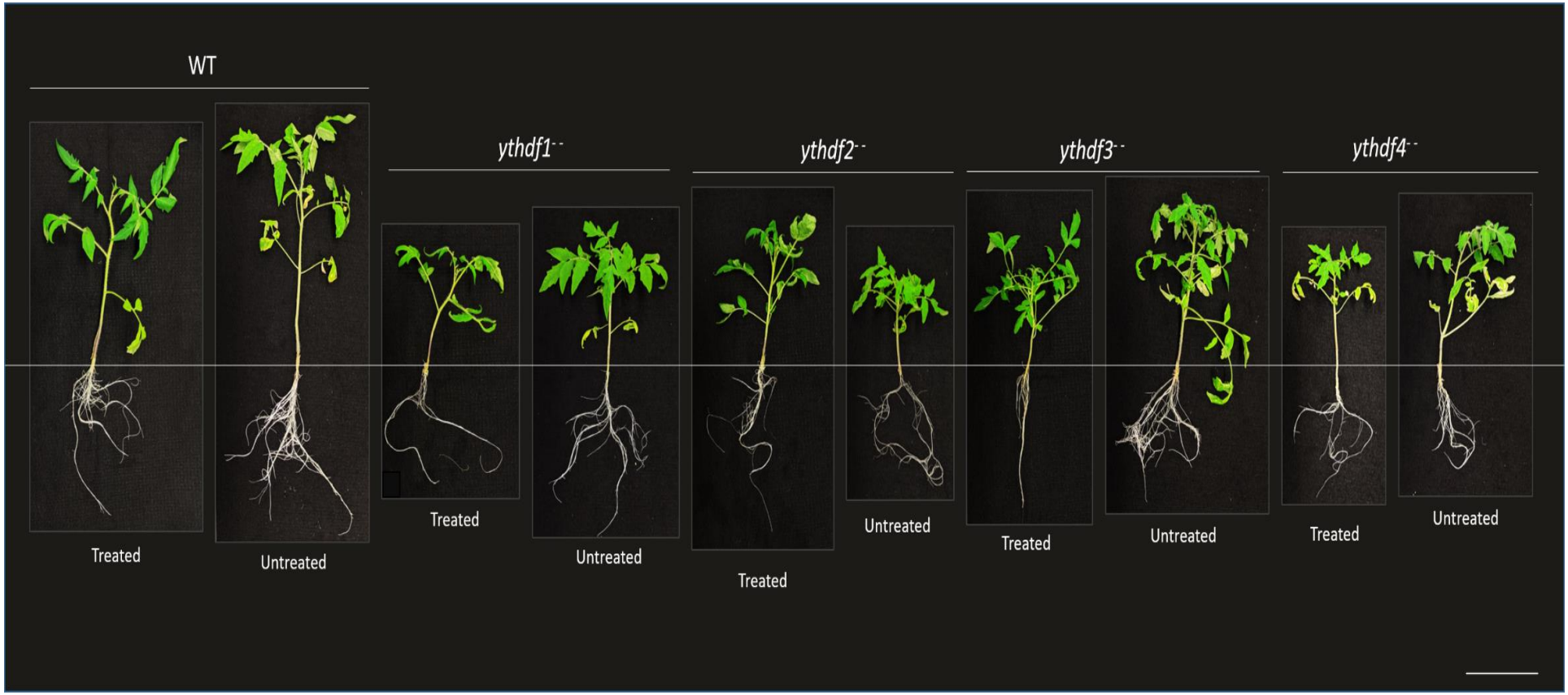


Figure 4.6: Wild-type and *ythdf* single mutant lines at 4 weeks, comparison of high salt treatment and standard conditions. Representative images of wild-type and *ythdf1-4* single mutant lines, grown in modified Hoagland's solution 50-150mM NaCl and Standard) and photographed at 4 weeks. Scale bars at 10cm.

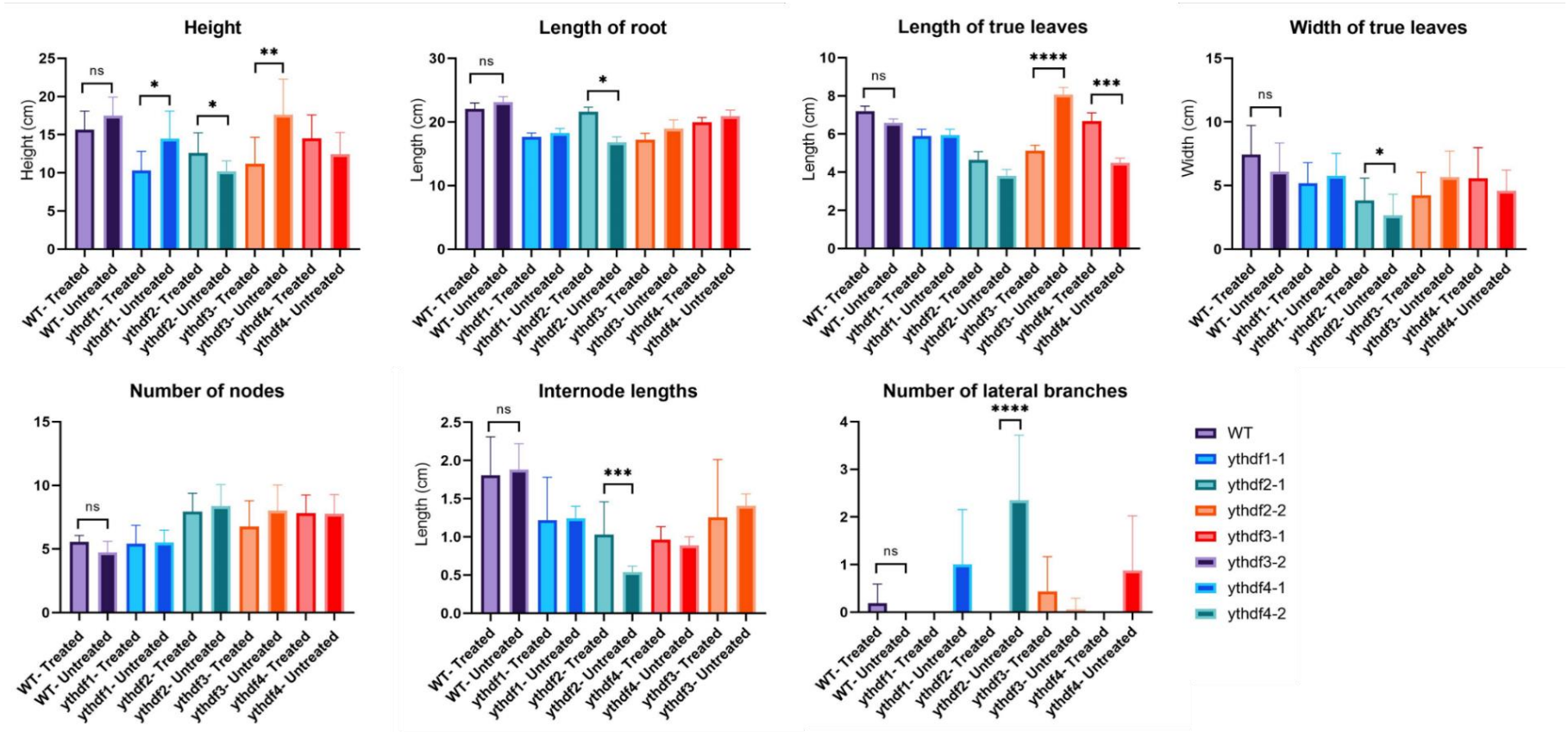


Figure 4.7: Characterisation of *ythdf1-4* single mutant KO lines in high salt and standard conditions. Histogram representation of key plant growth measurements. Plants grown in modified Hoagland's solutions (50-150mM NaCl and Standard). Measurements taken at 4 weeks. Error bars representing SD values from 15-20 replicates (n= 15-20). P values were obtained using mixed effects model and Dunnett's multiple comparisons test to determine whether mutant values are significantly different from wild-type values. P value significance bars: * (<0.05), ** (<0.005), *** (<0.0005), **** (<0.0001). No significance bars indicate no significance to WT.

4.2.3.4 Loss of *ythdf1* or *ythdf2* does not affect ambient m^{6}A levels

Studies of *ythdf* loss-of-function mutants in mammalian systems have indicated a range of roles of YTHDF proteins in RNA metabolism, including alternate splicing and RNA stability (Wang *et al.*, 2013; Du *et al.*, 2016). It is possible that YTHDF protein in plants, mirror these roles *in vivo.* Any changes in RNA stability, if present, could affect the levels of m^{6}A in the cells. mRNA fragments containing m^{6}A, could increase or decrease in line with any increase or decrease in their stability. A recent study by the Brodersen lab, did indeed find suggestions that SlYTHDF1 homologs ECT2 and ECT3, regulate RNA stability. Loss of ECT2/3/4 results in a decrease in ECT2/3 mRNA targets, suggesting the YTH proteins protect their targets from degradation (Arribas-Hernández *et al.*, 2021).

Total RNA was extracted from 3 seedlings of WT, *ythdf1* and *ythdf2* lines at 2 weeks. Total RNA used to produce poly(A) RNA by oligo d(T) magnetic beads. poly(A) fractions were then pooled before being subjected to ribonuclease T1 digestion, 5' end labelling, P1 nuclease digestion and TLC analysis (Zhong *et al.*, 2008) (Figure 4.8).

TLC detection of m^{6}A:A ratio in *Solanum lycopersicum. L* showed that whole plant basal ratios of A:m^{6}A in two-week-old seedlings is 1.85%. When compared to *ythdf1* and *ythdf2* mutant lines, the ratios are 1.81% and 1.89% respectively. This is well within the range of normal variation commonly observed within plant mRNAs (Bodi *et al.*, 2012a). These initial findings do not indicate any significant change in m^{6}A levels in these two mutants.

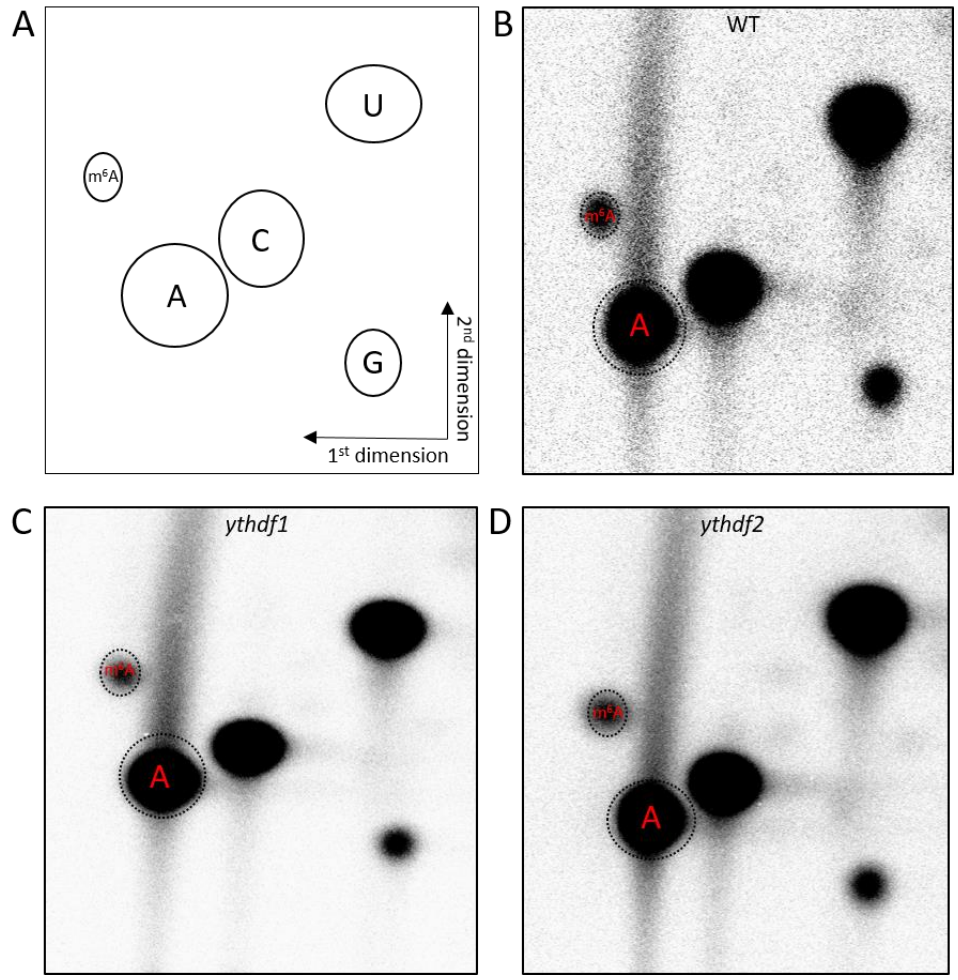


Figure 4.8: Two-dimensional TLC detection of relative global abundance of m⁶A in WT, *ythdf1* and *ythdf2*. (A) Schematic diagram of the relative positions of single nucleotide spots from TLC method. (B) Two-dimensional TLC analysis of poly(A) RNA from 2-week-old *Solanum lycopersicum* WT. (C) Two-dimensional TLC analysis of poly(A) RNA from 2-week-old *Solanum lycopersicum ythdf1* loss-of-function mutant. (D) Two-dimensional TLC analysis of poly(A) RNA from 2-week-old *Solanum lycopersicum ythdf1* loss-of-function mutant.

4.3 Discussion

In this study, CRISPR/Cas9 DNA targets were designed for efficient disruption of all four YTHDF genes in *Solanum lycopersum*. sgRNA targets were cloned into a modular CRISPR/Cas9 and subsequently used for Agrobacterium-directed transformation using wild-type cotyledons. From more than 500 cotyledons, a minimum of 16 transgenic lines were generated

and screened for the presence of induced mutations. From the confirmed mutated lines, two were selected for the characterisation of the mutated genes. Despite numerous attempts and downstream screening, unfortunately only a single mutated *ythdf1* line was generated. The produced loss-of-function *ythdf* lines, were characterised for any physiological differences resulting from the induced mutations. Any differences would suggest potential roles for these novel proteins, *in vivo*.

4.3.1 Development of CRISPR/Cas9 mediated gene disruption in YTHDF proteins

Although the gene disruption system chosen for this study has been used for numerous studies, with great success. The apparent efficiency of mutations generated by the Cas9, was low. From between 16 and 31 transgenic lines, post transformation between 1 and 3 of these had any stable mutations.

Numerous studies have utilised two single sgRNA guides for the generation of large deletions within target genes. This not only increases the efficiency of gene mutagenesis, but often allows for a greater ease of screening (Li *et al.*, 2013; van Eck, 2018). Large deletions, increase the likely hood of disrupting common restriction sites, or allow for size specific PCR amplicon screening. Future attempts to generate an additional *ythdf1* loss-of-function line will likely utilise a combination of these approaches to increase the efficiency of induced mutations. It is possible that use of two single sgRNA guides from the outset would have aided in this investigation.

4.3.2 Analysis of *ythdf* loss-of-function mutant characterisation

The development of gene disruption lines for all four YTHDF genes in *Solanum lycopersicum*, is the first key step in creating a new model for the study of YTH proteins and their function in tomato. Confirmation of these lines provide a platform for all future research in this area. Stable lines have been generated, to allow this work to continue, not only in our lab but within

future collaborations. As previously mentioned, only a single *ythdf1* loss-of-function line was obtained. This is the primary limitation of this study, however any data provided by this line assists in painting the picture along with the YTHDF2, YTHDF3 and YTHDF4.

The phenotypes associated with the disruption of the target genes, suggest key roles for the YTHDF genes in tomato (Figure 4.3, 4.5). Physiological differences are apparent at a macroscopic scale. This observation supports our decision to utilise *S. lycopersicum* for the study of these proteins. All studies to date into the roles YTHDF proteins in plants, has required at KO of at least three YTHDF genes concurrently to observe macroscopic morphological changes (Wang, Guo, Wang, *et al.*, 2017; Arribas-Hernández *et al.*, 2018, 2020, 2021; Scutenaire, J.-M. Deragon, *et al.*, 2018).

Vertical growth of *ythdf1, ythdf2* and *ythdf4* lines were all significantly reduced. 23.6% decrease in height for *ythdf1,* a 39.1% decrease in *ythdf2* lines and a 29.6% decrease in the height of *ythdf4* (Figure 4.7). The size of the 2nd, 3rd and 4th compound leaves of each plant also showed defects associated with leaf development. All four lines exhibited significantly different leaf lengths with *ythdf1, ythdf2* and *ythdf4* showing reduced growth and *ythdf3* with significantly longer leaves. *ythdf1, ythdf2* and *ythdf4* also displayed reduced leaf widths (Figure 4.4) Overall, growth rates of *ythdf1, ythdf2* and *ythdf4* are significantly reduced in loss-of-function lines. Interestingly *ythdf3* does not seem to show decreased growth rate, the average vertical height of the measured plants was higher than WT with a 7.4% increase, and significantly larger leaves. These observations clearly indicate that correct timing and development of vegetative tissues requires YTHDF RBPs. This replicates the previous findings of three concurrent studies on ECT protein in *Arabidopsis* (Arribas-Hernández *et al.*, 2018; L.-H. H. Wei *et al.*, 2018; Scutenaire, J.-M. Deragon, *et al.*, 2018), and confirmed by the Brodersen lab in 2020/2021 (Arribas-Hernández *et al.*, 2020, 2021). m^{6}A has already been linked to growth rate. Low m^{6}A lines exhibit slower growth in *Arabidiopsis* and abnormal organ definition, here our observations show that YTHDF proteins in tomato overlap with some of these m^{6}A related phenotypes, suggesting the aberrant morphologies may be due to YTH mediation of m^{6}A functions.

SEVERAL studies have associated m^{6}A and its binding proteins to root development. Hypomorphic mutants exhibiting low levels of m^{6}A in *Arabidopsis* result in defective gravitropic responses and altered development of lateral roots and vasculature (Bodi *et al.*, 2012a; Růžička *et al.*, 2017). Root length was the only metric recorded in this study. Root lengths of *ythdf1* and *ythdf2* were both reduced, by -19.9% and -18.7% respectively. No significantly differences were observed in *ythdf3* or *ythdf4* lines. However, *Ythdf4* does show longer roots than would be expected with such marked reductions in vegetative growth (Figure 4.3, 4.4). m^{6}A's role in root development and responses hint at further discoveries to be made. A closer look into gravitropic responses, root vasculature and lateral branching could be all that is needed to uncover if m^{6}A functions in plant roots are mediated by YTHDF proteins.

The number of compound leaves (nodes) and the distance between them can be a key indicator of genetic defects. Internode length in particular has been used as a diagnostic index, and is particularly sensitive to abiotic stresses including drought, temperature changes, lowlight and excess nitrogen (Grimstad, 1993). Loss of YTHDF2, YTHDF3 and YTHDF4 greatly increase the number of nodes produced, this combined with reduction in vertical height results in significantly reduced internode lengths in *ythdf1, ythdf2* and *ythdf4. Ythdf3* has significantly more leaves/nodes, but the internode length of the plants is not significantly reduced as in the other mutant lines, showing that *ythdf3* does not exhibit any dwarf phenotypes. *Ythdf1, ythdf2* and *ythdf4* on the other hand all show traits associated with dwarfism (Figure 4.3, 4.4). The contrasting phenotypes of YTHDF3; increased leaf expansion and internode elongation also hint at a connection between m^{6}A and GA. This raises some very interesting questions. Could YTHDF proteins be involved in the regulation of the GA-biosynthesis pathway? Are YTHDF1, 2 and YTHDF4 in competition with YTHDF3 in these roles? Interestingly, *ythdf1* and *ythdf4* have a significantly reduced germination rate, when compared to WT. This could also support a link between GA and m^{6}A/YTHDF proteins. *Ythdf2,* exhibits the strongest aberrant phenotypes measured in this study. Despite this, the germination rate of its seeds do not differ significantly from WT (Figure 4.5). The aberrant phenotypes resulting from loss of YTHDF proteins show a large degree of overlap, with the

exception of *ythdf3,* but the consistent differences also hint at unique targets and subsequent roles for each of the protein *in vivo.*

The ability of angiosperms to multiply their shoot apical meristems (SAM) and subsequently distribute these in the form of auxiliary meristems, is an essential process. In seed plants, like tomato, shoot branching is initiated at the shoot apex and the meristems derived from these are very similar in structure to the SAM. The inhibitory effect of the shoot apex on the development of these lateral branches, prevents lateral branching in tomato until later in the plant life cycle (Cline, 1997). In gene disruption lines *ythdf1, ythdf2* and *ythdf4,* the number of lateral branches is increased significantly. This is especially significant due to the complete lack of visibly active auxiliary meristems in WT plants at 4 weeks. This indicates a significant loss of apical dominance. Loss of apical dominance is also observed in reduced methylation *Arabidopsis* lines (Bodi *et al.*, 2012a). The transcript of key transcription factors (E.g., WUSCHEL and SHOOT MERISTEMLESS), are methylated to a significant level. Mutants lacking methylation complex member FIP37 exhibit massive over-proliferation of shoot meristems. The transcripts of these transcription factors have increased stability in the absence of m^6A, resulting in the over-proliferation of the auxiliary meristems (Shen *et al.*, 2016). Outside of plants, m^6A has been shown to regulate stem cell differentiation in mammals (Batista *et al.*, 2014; Ke *et al.*, 2017), it is possible that m^6A is involved in this function in plants, but no evidence of this in plants has been published to date. If YTHDF proteins are involved in the mediation of meristematic control, and if so to what extent is currently unknown. Our findings certainly suggest that at least some of the YTHDF proteins are likely to have some role to play in meristematic function.

4.3.3 Analysis of effect of salinity stress of *ythdf* loss-of-function mutants

Mechanisms of posttranscriptional gene regulation are often utilised to mediate important or rapid changes in gene expression or transcriptome profiles. The flexibility and speed of these responses make these

mechanisms integral to stress adaptation (Sunkar *et al.*, 2007). The first studies into YTH domain proteins in plants, using micro-array data, found that YTH genes are induced in both abiotic and biotic stresses conditions (D. Li *et al.*, 2014). Although further clues into YTH protein specific roles have been uncovered, with ECT2, ECT3 and ECT4 localising into P-bodies and ECT2 relocalising into stress granules under heat stress in *Arabidopsis* (Arribas-Hernández *et al.*, 2018; Scutenaire, J.-M. Deragon, *et al.*, 2018), few advances have been made since. A recent collaboration by the Fray lab, identified a set of key salt stress transcripts resulted in selective stabilization upon hypermethylation of m^{6}A (Anderson *et al.*, 2018). Here we show that YTHDF proteins have a more significant response to high salinity than WT plants. *Ythdf1* and *ythdf3* show reduced growth rate, while y*thdf2* and *ythdf4* show significantly increased growth rate. Again, *ythdf2* shows the most pronounced differences, with significantly increased internode lengths compared to untreated plants, and a large decrease in activation of auxiliary meristems from an average of 2.55 to 0 (Figure 4.7). These findings could indicate opposing roles of *ythdf1/ythdf3* to *ythdf2/ythdf4,* with potential interplay between them. Of course, it is hard to compare responses of plants that already show aberrant phenotypes in standard conditions, but there was no significant differences between WT physiologies in either condition. Mechanistic investigations utilising molecular approaches would provide more concrete evidence and would be necessary before any solid conclusions are drawn. The picture of YTHDF functions in stress response and adaptation is slowly appearing, the utilisation of our novel tomato model could be integral to future developments.

4.3.4 Measuring m^{6}A levels in *ythdf* loss-of-function mutants

We measured the global m^{6}A levels in *Solanum lycopersium. L* to assess the basal levels in whole plants. TLC detection analysis developed by the Fray Lab (Zhong *et al.*, 2008), showed that there is a m^{6}A:A ratio of 1.85% in 2-week-old seedlings. This is the first-time levels of m^{6}A in whole tomato plants have been measured. Interestingly, an article in 2019 looking into the role of m^{6}A in fruit ripening published a m^{6}A:A ratio of between 0.5% and

0.8% through 17- and 52-days post anthesis. This suggests that m^{6}A levels differ to a large extent in difference tissues (Zhou, Tian and Qin, 2019). Mirroring dynamic DNA methylation already described in tomato fruit ripening (Zhong *et al.*, 2013). Similar differences in m^{6}A are found between *Arabidopsis* tissues (Bodi *et al.*, 2012a).

Comparisons of *ythdf1* and *ythdf2* m^{6}A levels, suggest that there is no significant changes in the levels of m^{6}A upon loss of either of these proteins (Figure 4.8). While this does show that these proteins do not regulate m^{6}A levels *in vivo*, it does not rule out roles in RNA biogenesis, stability or degradation observed in mammals (Zhu *et al.*, 2014; N. Liu *et al.*, 2015a; Du *et al.*, 2016; Shi *et al.*, 2017). These TLC detection assays were performed with a single pooled repeat, however repeats were not performed for *ythdf1* or *ythdf2*. In addition to this, levels of *ythdf3* and *ythdf4* are yet to be measured. These limitations do not allow any conclusions to be draw from this study. Normal basal levels of m^{6}A are essential for correct development, growth patterns, responses to stress and hormonal regulation. Clearly, more work is needed to aid in our understanding of m^{6}A in tomato.

4.3.5 Future Work

The work laid out in this chapter will provide a platform for future investigations into the highly interesting YT521-B domain containing family of RBPs. Previous studies into m^{6}A in plants have relied upon the hypo-methylated lines (Zhong *et al.*, 2008; Bodi *et al.*, 2012a; Růžička *et al.*, 2017). The phenotypes associated with low methylation has been utilised as a background to aid in the comparison of any m^{6}A related studies, outside of the methylation complex. This includes YTH proteins. This raises some important questions and highlights some limitations of the current m^{6}A research model in *Solanum lycopersicum*. Are there any differences in the methylation within tomato? Do tomato low methylation lines exhibit the same hypomorphic phenotypes as those of Arabidopsis? Are all YTHDF roles linked to m^{6}A? Development of low methylation tomato lines and robust phenotyping would allow direct comparison to our YTHDF null lines. This is especially important due to the slight overlap of initial findings in this

study. To this end, RNAi and CRISPR/Cas9 constructs have been designed and constructed to target key components of the methylation complex in tomato, this work will continue in the future.

As previously mentioned, the lack of additional *ythdf1* mutants is a limitation of this work (Figure 4.2). Replication of the phenotyping of *ythdf1-1* in another line would add weight to any findings in this study. To mitigate this limitation, we have designed antibodies (Agrisera AB, Vännäs) for all YTHDF1 and YTHDF2 proteins in tomato. Once produced these will allow western blot analysis to confirm the lack of YTHDF proteins in these null lines. Confirmation of this would allow more confidence in assigning measured phenotypes to genomic mutation and not an unknown effect, especially for YTHDF1. Ideally another *ythdf1* line would be generated. Future attempts at this would utilise two single sgRNA guides, which is proved to increase the efficiency of CRISPR/Cas9 mediated gene disruption (Li *et al.*, 2013; van Eck, 2018).

The physiological phenotypes described in this study all use basic growth indices, but these have barely scratched the surface. There are many avenues which we have yet to explore. m^6A RNA precipitation and sequencing (m^6A-seq) on WT tomato will give a library of m^6A peaks providing a database for future comparisons. Knowledge of which transcripts are enriched for m^6A would be invaluable for future research. Repeating the m^6A-seq in *ythdf1, ythdf2, ythdf3* and *ythdf4* lines could elucidate whether the loss of each YTHDF protein affects the stability of any of the transcripts with the m^6A mark.

More in-depth characterisation of phenotypes resulting from loss-of-function mutations of YTHDF genes, could be informative. Further investigation into gravitropic responses and lateral branching in roots would be necessary to replicate phenotypes associated with low methylation lines in Arabidopsis. Observations of floral patterning and timing, fruit development and meristem maintenance would also be of interest. It will be important to determine patterns and timings of expression and to test whether the different "readers" act in an independent, overlapping, synergistic or antagonistic fashion. Some of these questions could be addressed by looking

at phenotypes of double mutant plants and this will be discussed in the next chapter.

Chapter 5 - Characterisation of YT521-B protein double loss-of-function mutants in Solanum lycopersicum

5.1 Introduction

Tomato (*Solanum lycopersicum*) is an ideal model system for any YTH-related genomic, transcriptomic or proteomic study. We have developed the beginnings of a robust plant model for developing an in-depth understanding of YTH proteins and their role within plant systems. The phylogeny of YTH domain and YTH proteins is classified into two subsets: predominantly nuclear YTHDC and the predominantly cytoplasmic YTHDF proteins (Zaccara, Ries and Jaffrey, 2019). Plant genomes encode a largely expanded set of YTHDF proteins. In *Arabidopsis* these have been named EVOLUTIONARILY CONSERVED C-TERMINUS (ECT) proteins, 11 of which fall into the YTHDF classification (D. Li *et al.*, 2014; Scutenaire, J.-M. Deragon, *et al.*, 2018). However, to date only *in vitro* studies of AtECT2 has demonstrated binding to m^{6}A (L.-H. H. Wei *et al.*, 2018).

In plants, m^{6}A is not evenly distributed across mRNA molecules or the transcriptome. Typically, around 1.5% of the adenosines found in mRNA molecules (outside of the polyA tail) are m^{6}A. m^{6}A is deposited in the RRm6ACH (R=G/A, H: U/A/C) motif, predominantly Gm^{6}AC. They are also found almost exclusively on exons, with a preferential distribution in terminal exon and 3'UTRs (Zhong *et al.*, 2008; Bodi *et al.*, 2012b; Dominissini *et al.*, 2012; Meyer *et al.*, 2012; Schwartz *et al.*, 2013; Schwartz, Maxwell R. Mumbach, *et al.*, 2014; K. Liu *et al.*, 2015). There is some debate around the consensus sequence and its context for YTH proteins in plants, with studies finding contrasting binding sequences for AtECT2. A plant specific motif URU(m^{6}A)Y (L.-H. H. Wei *et al.*, 2018) was later contradicted by the Brodersen lab in 2020 who provided evidence that AtECT2 preferentially binds the traditional RRACH motif, but suggested that U rich

sequences were frequently nearby (Arribas-Hernández and Brodersen, 2020a; Arribas-Hernández *et al.*, 2021). While there seems to be an emerging role for pyrimidine-rich motifs in m^{6}A reading and writing, demonstrated by consistent conservation across plants (Y. Li *et al.*, 2014; Zhou *et al.*, 2018; Luo *et al.*, 2020; Miao *et al.*, 2020b), functions of these motifs are still unknown. What is clear is that YTHDF proteins in plants are likely to have a range of binding sites in an m^{6}A context. Whether YTHDF proteins have function outside of their potential ability to bind m^{6}A is an interesting question for the future.

The expansion of YTHDF proteins has been shown to result in significant levels of genetic redundancy. In *Arabidopsis*, concurrent KO of ect2 and ect3 was necessary to induce any observable phenotypic changes (Arribas-Hernández *et al.*, 2018; Scutenaire, J.-M. Deragon, *et al.*, 2018). Further KO of ect4 genes only induces minor aberrant phenotypes (Arribas-Hernández *et al.*, 2020). Tomato genomes encode significantly fewer YTHDF proteins than the average for monocot or dicot plants (4 proteins, section 3.2.2), resulting in potentially lower genetic redundancy and improving the efficiency of generating all double *ythdf* null mutants. The aim of this chapter is to generate and characterize the homozygous double mutant. It is hoped that such an approach will allow the identification of ythdf class-specific and redundant functions and interactions.

5.2 Results

Construction of double mutants within the same gene family can be a useful tool in further identifying roles of target genes, as well as uncovering any genetic redundancy between them. Constitutive knockout of two similar genes can also identify redundant functions even if the single mutant knockouts cause visible phenotypic effects. We attempted to generate double mutant knockouts for all possible combinations of the four *ythdf* loss-of-function mutants. Multiple crosses were attempted for all possible combinations with subsequent fruit collected for downstream screening of both genes on both alleles. *Ythdf1/ythdf3, ythdf2/ythdf3,* and

ythdf3/ythdf4 lines were all successfully produced, and seed bulked for subsequent investigations.

5.2.1 *ythdf1/ythdf4* homozygous KOs are lethal

At least 20 fruits were generated from crossing *ythdf1* and *ythdf4* single mutant plants, producing F1 seeds. As expected, these were heterozygous for both alleles. The F1 generation was then selfed to produce populations of F2 seeds. We would expect that 1 in 16 of these seeds would be homozygous for both *ythdf1* and *ythdf4.* 75 of these F2 seedlings were genotyped with 0 returning homozygous for *ythdf1* and *ythdf2.* However, we did successfully identify *ythdf1--/ythdf4*+- and *ythdf1*-+/*ythdf4*-- plants. Both ythdf*1--/ythdf4*+- and *ythdf1*-+/*ythdf4*—plants were selfed top produce the F3 generation. In this generation we would expect that 1 in 4 would be the double homozygous mutant. 20 of the F3 seedlings from *ythdf1--/ythdf4*+- and *ythdf1*-+/*ythdf4*—parents were genotyped with 0 returning double homozygous mutants of double homozygous mutants of ythdf1 and ythdf4. This suggested that a homozygous double mutant line would not be generated via normal screening methods (Table 5.1)

Table 5.1: Genotyping results during *ythdf1/ythdf4* double mutant generation.

Generation	Parent Lines	Expected Ratio of Double mutant	Number Screened	Number of Double mutant plants
F2	*ythdf4-- and ythdf1--*	1:16	75	0
F3	*ythdf1--/ythdf4+- selfed*	1:4	20	0
F3	*ythdf1-+/ythdf4-- selfed*	1:4	20	0

5.2.1.1 Treatment of *ythdf1/ythdf4* seeds with gibberellic acid

Both *ythdf1* and *ythdf4* single mutants exhibited significantly reduced germination rates, when compared to WT and *ythdf2* and *ythdf3* single mutant seeds (Figure 4.5). To investigate the possibility that the lack of homozygous double *ythdf1/ythdf4* found in the screened lines is due to the seeds' inability to germinate, we tested the germination efficiency of the *ythdf1--/ythdf4*+- and *ythdf1*+-*/ythdf4*-- lines. We hypothesised that the cause of this reduction in germination efficiency as well as some of the vegetative phenotypes might be due to roles of functional *ythdf1* and *ythdf4* in the gibberellic acid biosynthesis (or response) pathway. Treatment with GA3 has been shown to increase the germination efficiency of seeds, especially in GA insensitive mutants (Groot and Karssen, 1987; de Castro *et al.*, 2000). Each line was treated with 150mg/L of gibberellic acid to induce germination. As previously observed (Figure 4.5), both *ythdf1* and both *ythdf1* and *ythdf4* lines exhibit significantly reduced germination efficiency compared to WT. Treatment with GA3 significantly increased the germination rate of both lines ($p<0.005$ and $p<0.0001$, respectively) (Figure 5.1). This demonstrated the potential of treatment with GA to induce germination of the target *ythdf1/ythdf4* mutant seeds. Under normal conditions, germination of seeds from *ythdf1--/ythdf4*+- lines were reduced but this difference was not significantly different from WT. *ythdf1*+-*/ythdf4*-- however did show significantly reduced germination rate. Upon treatment with GA_3, only *ythdf1*+-*/ythdf4*-- seed germination increased significantly ($p<0.005$) while there was an observable change in *ythdf--/ythdf4*+- seeds, this was not significantly higher than untreated *ythdf1--/ythdf4*+- seeds (Table 5.2)

Screening of 40 seeds of the resulting *ythdf1--/ythdf4*+- and *ythdf1*+-*/ythdf4*—lines again did not yield any double *ythdf1/ythdf4* plants. It is therefore likely that double knockout of both ythdf1 and ythdf4 genes, is lethal in tomato. The increase in germination rate to WT levels, post-GA_3 treatment could be an indication that *ythdf1/ythdf4* loss-of-function seeds

do not form or abort in the ovaries, further investigation is necessary confirm this hypothesis.

Table 5.2: Average gemination rates of WT, *ythdf1, ythdf4, . ythdf1–/ythdf4+-* and *ythdf1-+/ythdf4—* lines. Numbers generated from 4 repeats of 10 seeds each repeat. Normal treatment: Water. A treatment: 150mg/L GA3.

	Normal Treatment	GA Treatment	p-value
Wild-Type	0.900	0.950	ns
ythdf1	0.675	0.900	p<0.005
ythdf4	0.500	0.950	p<0.0001
ythdf1+-/ythdf4--	0.700	0.820	p<0.005
ythdf1--/ythdf4+-	0.800	0.867	ns

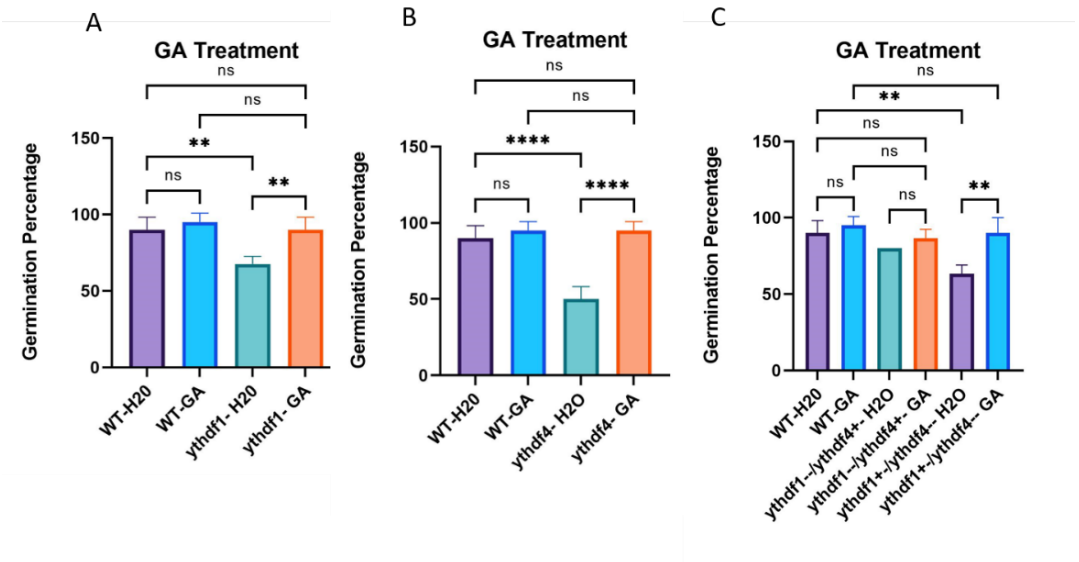


Figure 5.1: Percentage of seeds germinated in *ythdf1, ythdf4* and *ythdf1/4* lines. Seeds germinated on filter paper under standard conditions or following addition of 150mg/L GA_3. Error bars representing SD values from 30-40 replicates (n= 30-40). P values were obtained using ordinary one-way ANOVA and Tukey's multiple comparisons test to determine whether mutant values are significantly different from wild-type values. P value significance bars: * (<0.05), ** (<0.005), *** (<0.0005), **** (<0.0001) ns (No significance).

5.2.2 Generation of *ythdf1/ythdf2* and *ythdf2/ythdf4* double mutants.

Multiple crosses between *ythdf1* and *ythdf2*, and *ythdf2* and *ythdf4* were performed producing F2 seeds which should be heterozygous for both KO alleles in each case. Unfortunately, due to the Covid-19 pandemic we were unable to confirm the generation of either *ythdf1/ythdf2* and *ythdf2/ythdf4*, however we were able to successfully identify *ythdf1--/ythdf2*+- and *ythdf1-+/ythdf2*—plants, in the case of *ythdf1/ythdf2* and *ythdf2--/ythdf4*+- and *ythdf2-+/ythdf4*—plants, in the case of *ythdf2/ythdf4* (Figure 5.2; 5.3). This provides a platform for future investigations.

It is possible that, like *ythdf1/ythdf4*, *ythdf1/ythdf2* and *ythdf2/ythdf4* double mutants are lethal, therefore not possible to generate using standard crossing and genotyping techniques. Further investigations would be required to confirm this. 5.2.3 Physiological characterisation of ythdf null mutants in compost

To allow for large scale comparisons of all the produced double mutant lines, multiple repeats of all single and double mutant lines must be observed together under the same conditions. Previous phenotyping of the single ythdf null mutants was carried out in controlled growth conditions in a contained growth room. This environment did not have the required space for bulk phenotyping of the generated ythdf double mutant lines. All phenotyping of double mutant lines was therefore carried out in glasshouse conditions. A minimum of 10 seedlings per line were transferred into 9cm pots and grown at standard conditions (Section 2.1.3). Measurement was made at 4 weeks. Unlike measurements made in nutrient solution, only *ythdf1-1, ythdf2-1, ythdf3-1* and *ythdf4-1* were used in this study, hereby simply called *ythdf1, ythdf2, ythdf3* and *ythdf4.*

Similarly, to lines previously grown in nutrient solution, ythdf null mutants exhibit a range of physiological and morphological differences when

compared to WT (Figure 5.4). At 4 weeks, *ythdf1, ythdf2* and *ythdf4* exhibit significantly reduced height ($p<0.0001$, $p<0.0001$ and $p<0.005$, respectively).

The lengths of the 3rd and 4th true leaves are significantly shorter in *ythdf1* and *ythdf2* lines ($p<0.0001$ and $p<0.0001$). *ythdf4* does show a reduction similar to that observed in nutrient solution but this is not significant in this case. Leaf widths show a similar picture with *ythdf1* and *ythdf2* again showing significant reductions in growth ($p<0.0001$ and $p<0.0001$, respectively) (Figure 5.5).

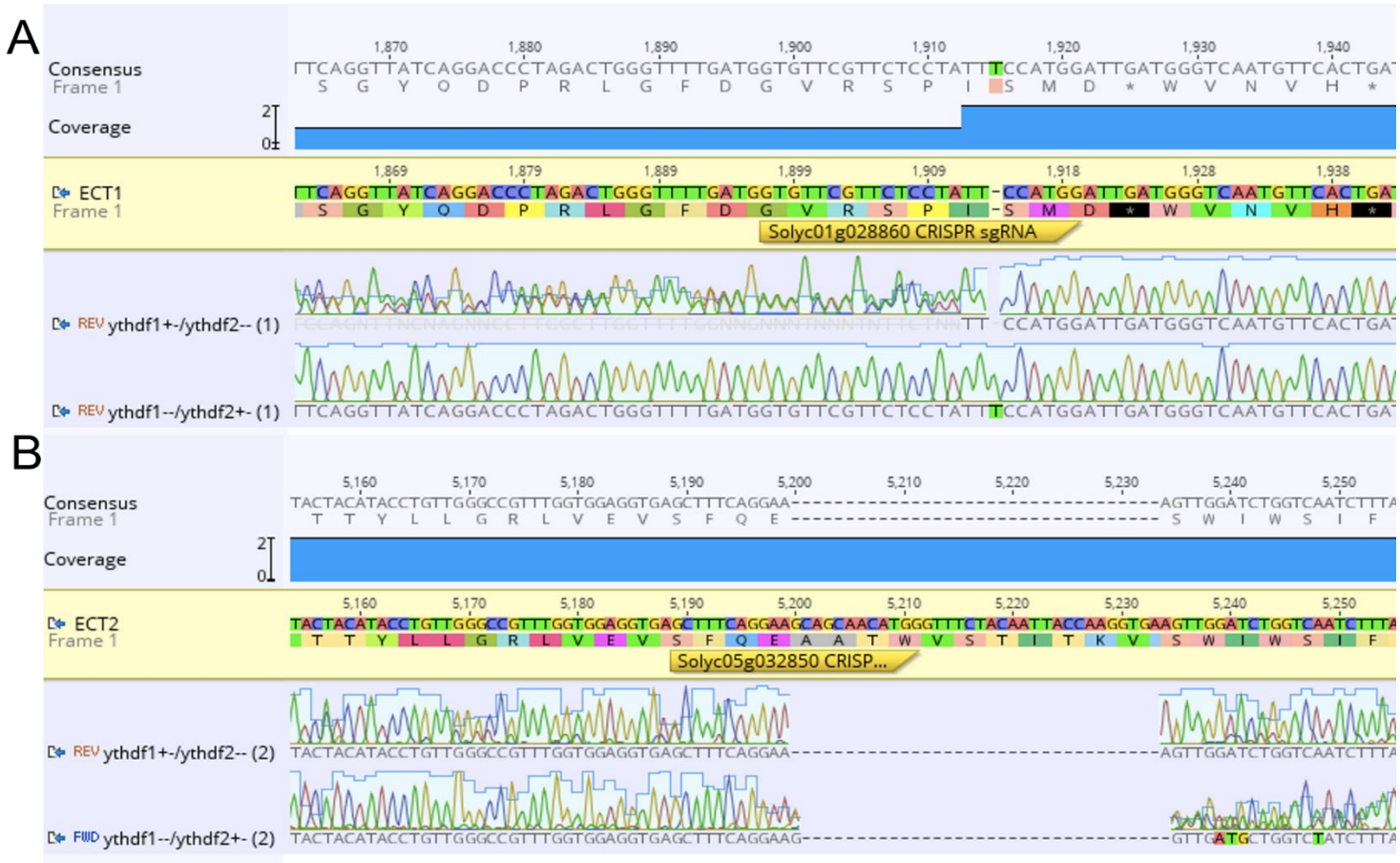


Figure 5.2: Mutation detection and consensus sequences of *ythd1/ythdf2* double mutant generation. (A) Sanger sequence read for ythdf1 in *ythdf1/ythdf2* homozygous/heterozygous lines. (B) Sanger sequence read for ythdf2 in *ythdf1/ythdf2* homozygous/heterozygous lines. sgRNA target guides indicated (yellow arrow).

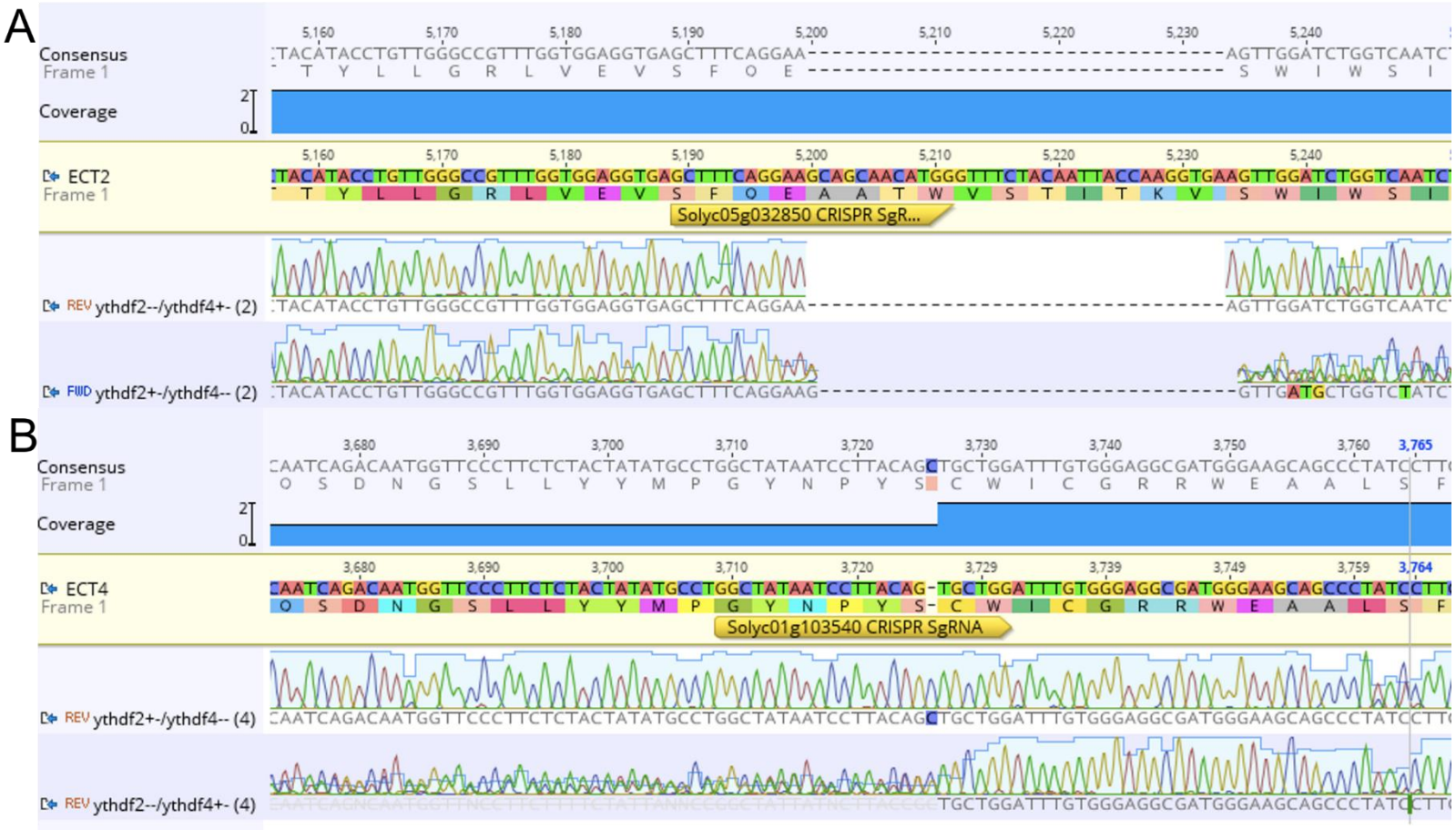


Figure 5.3: Mutation detection and consensus sequences of *ythdf2/ythdf4* double mutant generation. (A) Sanger sequence read for ythdf2 in *ythdf2/ythdf4* homozygous/heterozygous lines. (B) Sanger sequence read for ythdf4 in *ythdf2/ythdf4* homozygous/heterozygous lines. sgRNA target guides indicated (yellow arrow).

The number of nodes on each of the four *ythdf* lines in these conditions is the same as previous investigations. *Ythdf2, ythdf3* and *ythdf4* all show a significant increase in the number of compound leaves at 4 weeks ($p<0.0001$, $p<0.0005$ and $p<0.05$, respectively). Again, ythdf1 null mutant plants do not have a significantly different number of nodes at 4 weeks. However, the distance between each of these nodes, a key metric for the measurement of vertical growth rate, is significantly reduced in all four *ythdf* lines ($p<0.0001$, $p<0.0001$, $p<0.0001$ and $p<0.0005$, respectively) (Figure 5.5). *Ythdf3* exhibits a reduction in internode length despite a consistent slight increase in height, due to the increase in the number of compound leaves developed at this stage. Whereas *ythdf1, ythdf2* and *ythdf4* internode length reductions is to be expected as a combination of reduced height and significantly more leaves. *Ythdf1,* does not indicate an increase in the number of nodes at 4 weeks, but the significant reduction in internode length reflected a significantly reduced growth rate in the absence of YTHDF1 proteins.

Finally, the number of lateral branches (formed from auxiliary meristems) is visibly increased in *ythdf1, ythdf2* and *ythdf4* as previously observed in nutrient growth media. In this case, only *ythdf2* and *ythdf4* branching is significantly increased when compared to WT ($p<0.0001$ and $p<0.005$, respectively) (Figure 5.5). As mentioned, *ythdf1* does show an increase but the variability of the loss of apical dominance means it is not significant within this specific study. Taken together with other experiments, there is a strong indication that *ythdf1* also has a significant reduction in apical dominance due to the induced mutation.

Overall, these measurements indicate that the physiological changes induced by the generated mutants are consistent across different growth medium. Any differences in significance are likely as a variation derived from slight changes in conditions and the higher variability of glasshouse conditions.

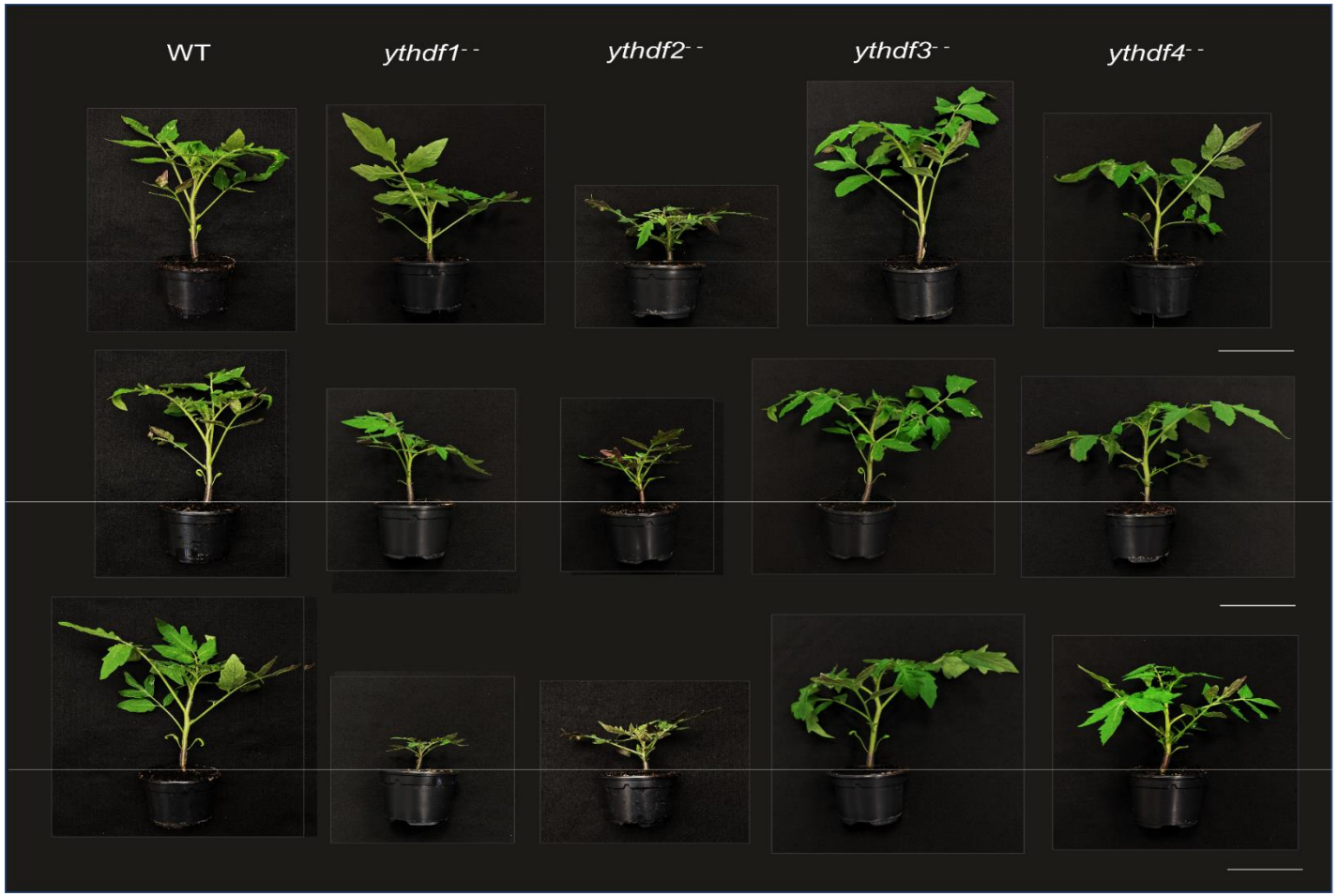


Figure 5.4: Wild-type and *ythdf* single mutant lines at 4 weeks. Representative images of wild-type and *ythdf1-4* single mutant lines, grown in Levington M3 and photographed at 4 weeks. Scale bars at 10cm.

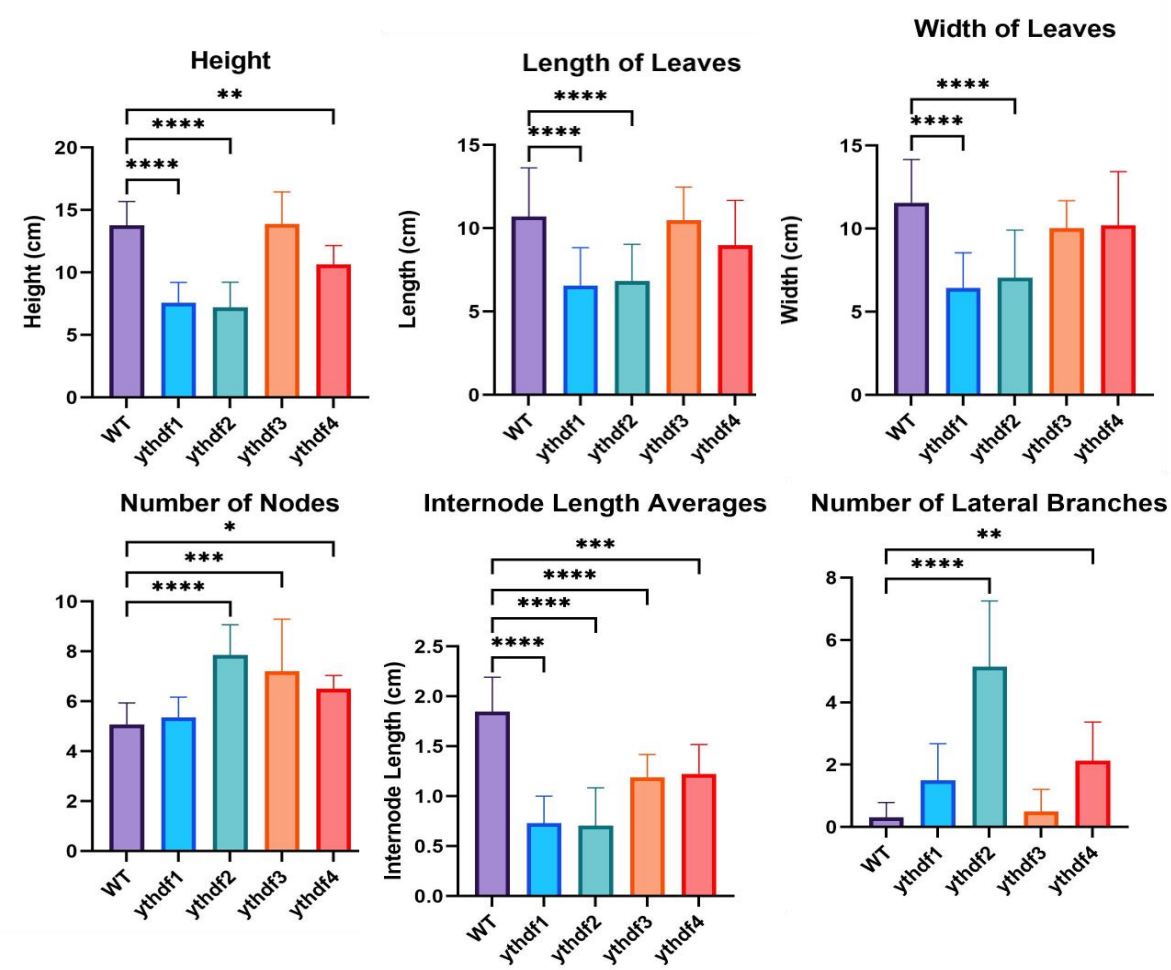


Figure 5.5: Characterisation of *ythdf1-4* single mutant KO lines. Histogram representation of key plant growth measurements. Plants grown in Levington M3 solution. Measurements taken at 4 weeks in cm. Error bars representing SD values from 10-15 10-15 replicates (n=10-15). P values were obtained using mixed effects model and Dunnett's multiple comparisons test to determine whether mutant values are significantly different from wild-type values. P value significance bars: * (<0.05), ** (<0.005), *** (<0.0005), **** (<0.0001). No significance bars indicate no significance to WT

5.2.4 Characterisation of ythdf double mutants

We successfully generated stable double mutant lines for *ythdf1/ythdf3, ythdf2/ythdf3* and *ythdf3/ythdf4*. The covid-19 pandemic prevented the completion of genotyping for *ythdf1/ythdf2* and *ythdf2/4*. Although we do not have 3 of the double mutants, the double mutants created provide a platform for more in-depth investigation into the roles of tomato YTHDF proteins. The knockout of two or more ythdf genes simultaneously can allow any genetic redundancy to present to be revealed. This is even more important in large protein families. As before, all phenotyping of double mutant lines was therefore carried out in glasshouse conditions. A minimum of 10 seedlings per line were transferred into 9cm pots and grown under standard conditions (Section 2.1.3). Measurement was made at 4 weeks. Vertical growth was assessed by measuring to the shoot apical meristem.

5.2.4.1 Physiological characterisation of *ythdf1/ythdf3* mutant lines

Measurements taken at 4 weeks show that *ythdf1/ythdf3* lines have significantly more vertical growth than *ythdf1* single mutants ($p<0.0001$) but no significant difference to heights of *ythdf3* plants, despite a lower average. *Ythdf1/ythdf3* lines do show a significant reduction in vertical growth when compared to WT plants, however. This could suggest that loss of *ythdf3* does indeed result in increased vertical growth and *ythdf1* results in reduced vertical growth (Figure 5.7) The growth of vegetative tissues indicated by the size of the true leaves is a key metric for understanding protein function. *Ythdf1/ythdf3* lines have significantly longer leaves than *ythdf1* plants ($p<0.0001$). The average lengths of the 3rd and 4th leaves measured in this experiment were higher in the double mutant plants, but not significantly longer than either WT or *ythdf3*. The widths of the leaves again show a significant increase in leaf width upon KO of *ythdf3* in the *ythdf1* background ($p<0.0001$). The widths of leaves of *ythdf1/ythdf3* were significantly higher than *ythdf3* ($p<0.05$) but not WT (Figure 5.6, 5.7). This could suggest overall larger leaves in the double mutant but this was not measured in this study.

The number of nodes in *ythdf1/ythdf3* plants was significantly lower than *ythdf3 (*$p<0.005$), displaying similar numbers of compound leaves to WT and *ythdf1*. Since *ythdf3* has a significantly higher number of leaves than WT, this reduction possibly conflicts with other metrics. The subsequent internode lengths resulting from the number of leaves show a complicated picture. *Ythdf1/ythdf3* plants have greater distance between each node on average than both *ythdf1* and *ythdf3* ($p<0.0001$ and $p<0.05$), but still significantly less than WT ($p<0.05$) (Figure 5.6, 5.7). The average height of *ythdf1/ythdf3* plants is higher than *ythdf3*, so this result increases interest in the interplay between *ythdf1* and *ythdf3* physiological roles.

significantly lower in *ythdf2/ythdf3* plants than WT ($p<0.005$) (Figure 5.8, 5.9). Internode lengths were also significantly higher than *ythdf2* ($p<0.0005$) but not *ythdf3*. *Ythdf2/ythdf3* plants are significantly shorter than *ythdf3* plants, the slightly higher internode average of double KO lines would therefore indicate that the number of nodes is potentially of significant so a repeat of these experiments would be advantageous.

The observed number of induced auxiliary meristems is higher in *ythdf2/ythdf3* plants than *ythdf3* and WT ($p<0.0001$ and $p<0.0001$, respectively) (Figure 5.8, 5.9). These numbers do not differ significantly from *ythdf2*. Similar to *ythdf1/ythdf3* plants, this increase in lateral branching compared to plants lacking ythdf3 is likely down to the loss of ythdf2.

Upon KO of ythdf3 there is no obvious increase in the amount of lateral branching when the plants are at an age of 4 weeks. All other *ythdf* lines show an increase in the number of auxiliary meristems formed at 4 weeks, including *ythdf1*. *Ythdf1/3* plants show significantly higher lateral branching when compared to *ythdf3* and to WT ($p<0.0001$ and $p<0.0001$, respectively). The lateral branching of these lines is not higher than *ythdf1* however (Figure 5.6, 5.7). Since ythdf3 has not shown any role in the maintenance of apical dominance, the increase observed is likely down to the knockout of ythdf1.

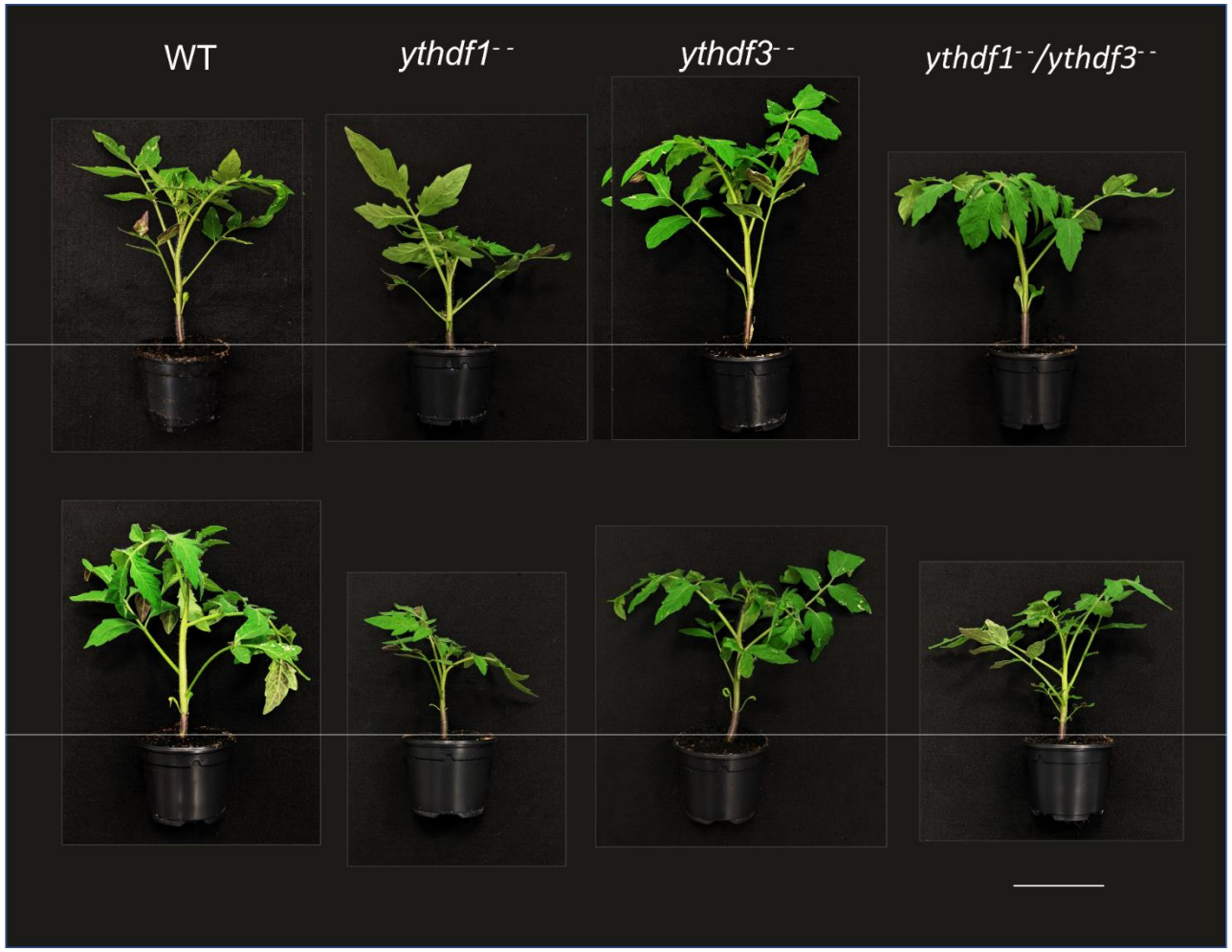


Figure 5.6: Wild-type and *ythdf1, ythdf3* single mutant lines and *ythdf1/ythdf3* mutant at 4 weeks. Representative images of wild-type and *ythdf1, ythdf3, ythdf1/ythdf3*, grown in Levington M3 and photographed at 4 weeks. Scale bars at 10cm.

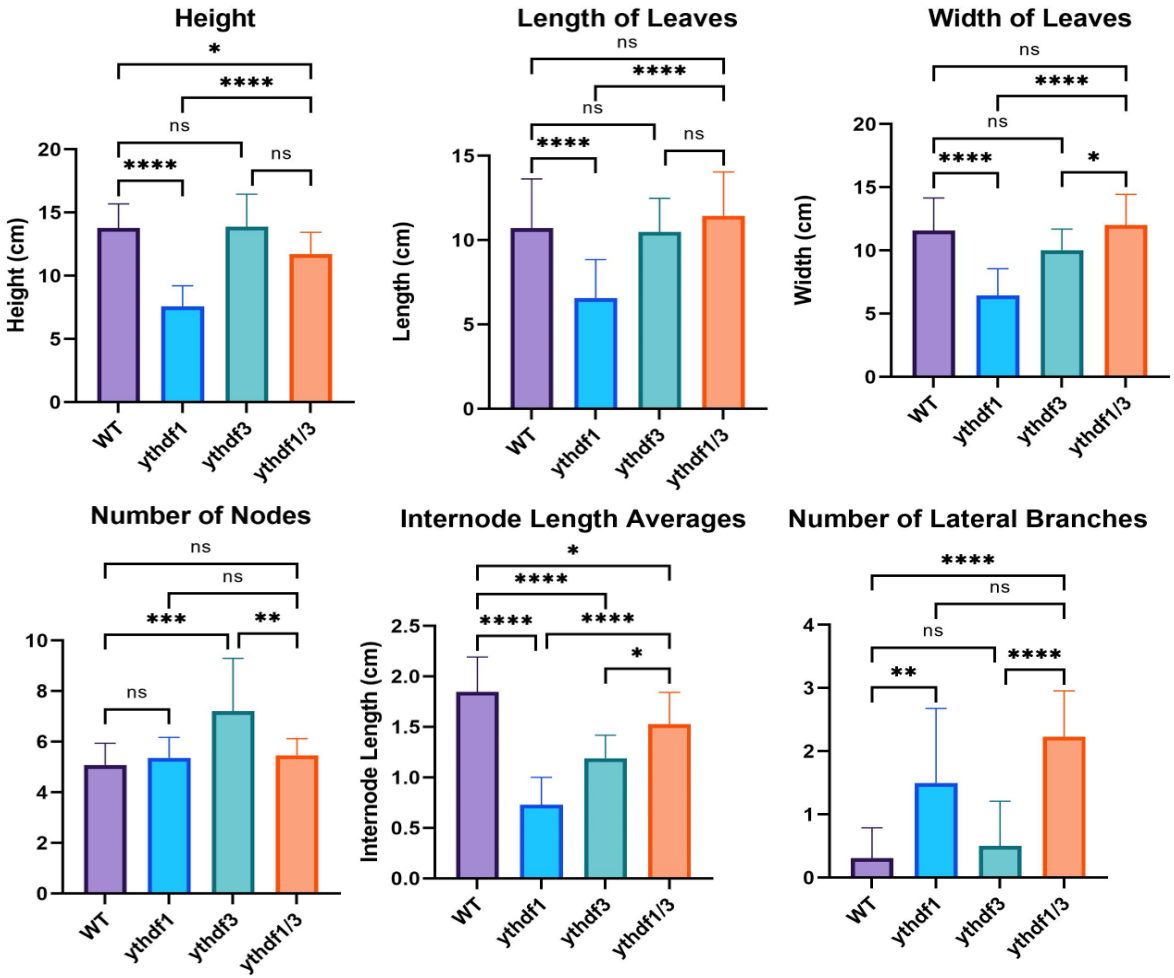


Figure 5.7: Characterisation of *ythdf1, ythdf3* single mutant lines and *ythdf1/ythdf3* mutant. Histogram representation of key plant growth measurements. Plants grown in Levington M3. Measurements taken at 4 weeks in cm. Error bars representing SD values from 10-15 10-15 replicates (n=10-15). P values were obtained using mixed effects model and Tukey's multiple comparisons test to determine whether mutant values are significantly different from wild-type values. P value significance bars: * (<0.05), ** (<0.005), *** (<0.0005), **** (<0.0001), ns (No significance).

5.2.4.2 Physiological characterisation of *ythdf2/ythdf3* mutant lines

Vertical growth of *ythdf2/ythdf3* double mutants were measured and compared to *ythdf2* and *ythdf3* single mutant lines. *Ythdf2/ythdf3* plants are significantly taller than *ythdf2* lines ($p<0.0005$) and shorter than *ythdf3* ($p<0.05$). The double mutants also have less vertical growth than WT ($p<0.05$). This shows that KO of *ythdf2* and *ythdf3* both increase and decrease the height of plants at 4 weeks (Figure 5.8, 5.9).

Vegetative growth of compound leaves 3 and 4 are significantly longer in *ythdf2/ythdf3* plants when compared to *ythdf2* ($p<0.0005$). The lengths of these leaves were not found to significantly different from either WT or *ythdf3*. As expected, the widths of the leaves show the same differences with wider leaves of *ythdf2/ythdf3* then *ythdf2* ($p<0.0005$), but no changes to WT or *ythdf3* (Figure 5.8, 5.9).

The number of compound leaves found on each plant of *ythdf2/ythdf3* may be less than *ythdf2* and *ythdf3.* The average number of nodes was found to be lower, but this was also found to be insignificant. Despite this, *ythdf2* and *ythdf3* both exhibit a higher number of leaves than WT ($p<0.005$ and $p<0.005$, respectively) but *ythdf2/ythdf3* plants were not found to have significantly higher node count than WT. These measurements would require repeating to achieve a greater degree of confidence in the resulting measurements. The average length between each of these nodes was still



Figure 5.8: Wild-type and *ythdf2, ythdf3* single mutant lines and *ythdf2/ythdf3* mutant at 4 weeks. Representative images of wild-type and *ythdf2, ythdf3, ythdf2/ythdf3* mutant lines, grown in Levington M3 and photographed at 4 weeks. Scale bars at 10cm.

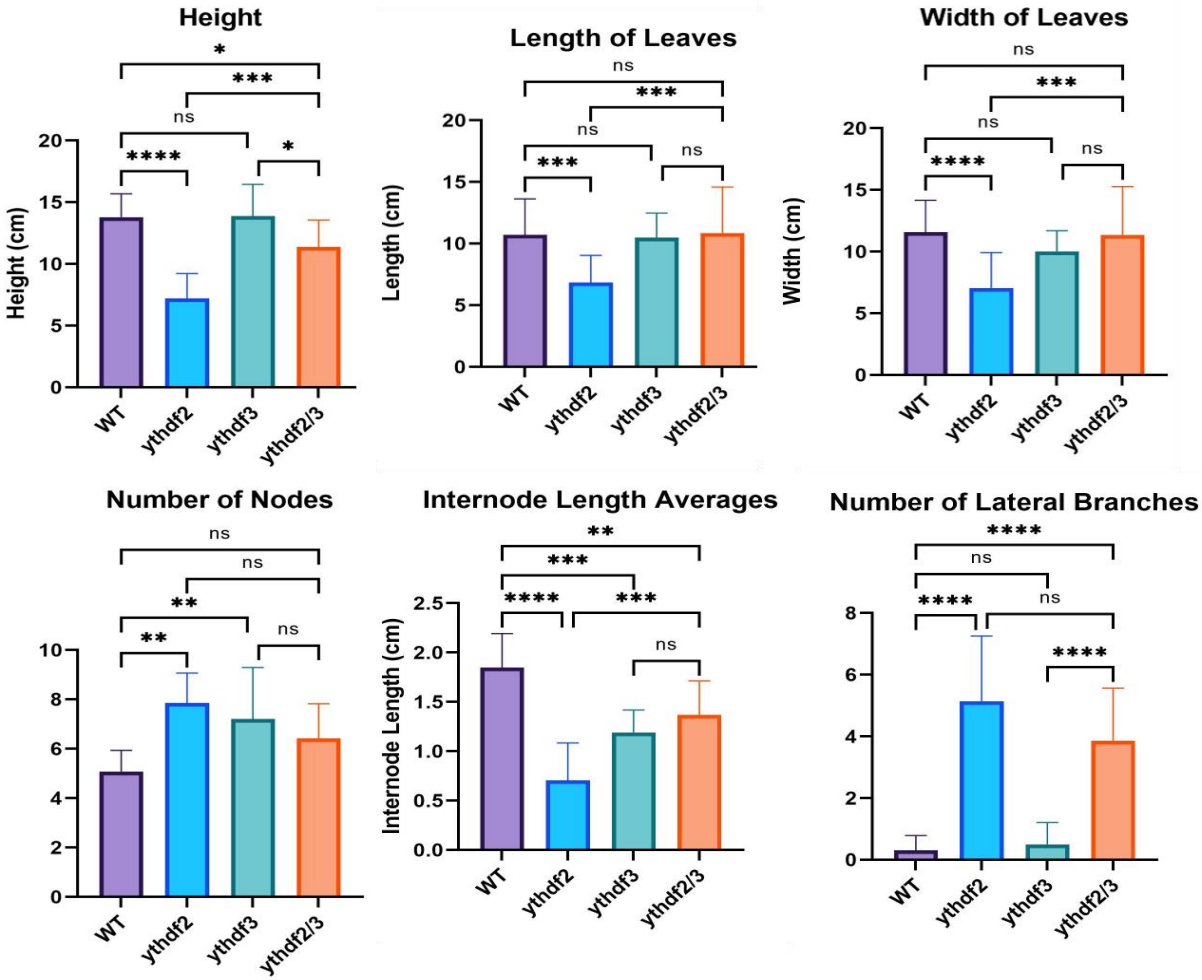


Figure 5.9: Characterisation of *ythdf2, ythdf3* single mutant lines and *ythdf2/ythdf3* mutant. Histogram representation of key plant growth measurements. Plants grown in Levington M3. Measurements taken at 4 weeks in cm. Error bars representing SD values from 10-15 10-15 replicates (n=10-15). P values were obtained using mixed effects model and Tukey's multiple comparisons test to determine whether mutant values are significantly different from wild-type values. P value significance bars: * (<0.05), ** (<0.005), *** (<0.0005), **** (<0.0001), ns (No significance)

5.2.4.3 Physiological characterisation of *ythdf3/ythdf4* mutant lines

Vertical growth in *ythdf3/ythdf4* is significantly lower than figures observed of *ythdf3* tomato plants ($p<0.005$). There are no significant changes in *ythdf3/ythdf4* when compared to the *ythdf4* background. Since *ythdf3* shows no significant changes when compared to WT, the KO of *ythdf4* results in a significant reduction in *ythdf3* backgrounds (Figure 5.10, 5.11)

There are no significant changes in leaf length among any of the single or double mutants *ythdf3, ythdf4* or *ythdf4.* The width of the 3rd and 4th leaves in *ythdf3/ythdf4* plants is significantly less than WT ($p<0.05$) (Figure 5.10, 5.11) suggesting that vegetative growth in *ythdf3/ythdf4* is reduced when compared to *ythdf3* and *ythdf4.*

The number of nodes is significantly higher in *ythdf3/ythdf4* than WT ($p<0.05$) but there is no differences in the number of nodes when compared to *ythdf3* or *ythdf4.* The same is true of the average internode lengths of *ythdf3.ythdf4* plants, with no differences between *ythdf3, ythdf4* or the double KO. All three however are significantly lower than WT ($p<0.0001$, $p<0.0005$ and $p<0.0001$) (Figure 5.10, 5.11).

The number of lateral branches which were initiated in 4-week-old plants were higher in *ythdf3/ythdf4* than *ythdf3* and WT ($p<0.0001$ and $p<0.0001$, respectively) showing that KO of ythdf4 in a *ythdf3* background reduces apical dominance. There was a measurably higher number of auxiliary meristems in *ythdf3/ythdf4* than *ythdf4*, however this was not found to be significant (Figure 5.10, 5.11). Further investigation may find a further reduction of apical dominance in *ythdf3/ythdf4* lines.



Figure 5.10: Wild-type and *ythdf3, ythdf4* single mutant lines and *ythdf3/ythdf4* mutant at 4 weeks. Representative images of wild-type and *ythdf3, ythdf4, ythdf3/ythdf4* mutant lines, grown in Levington M3 and photographed at 4 weeks. Scale bars at 10cm.

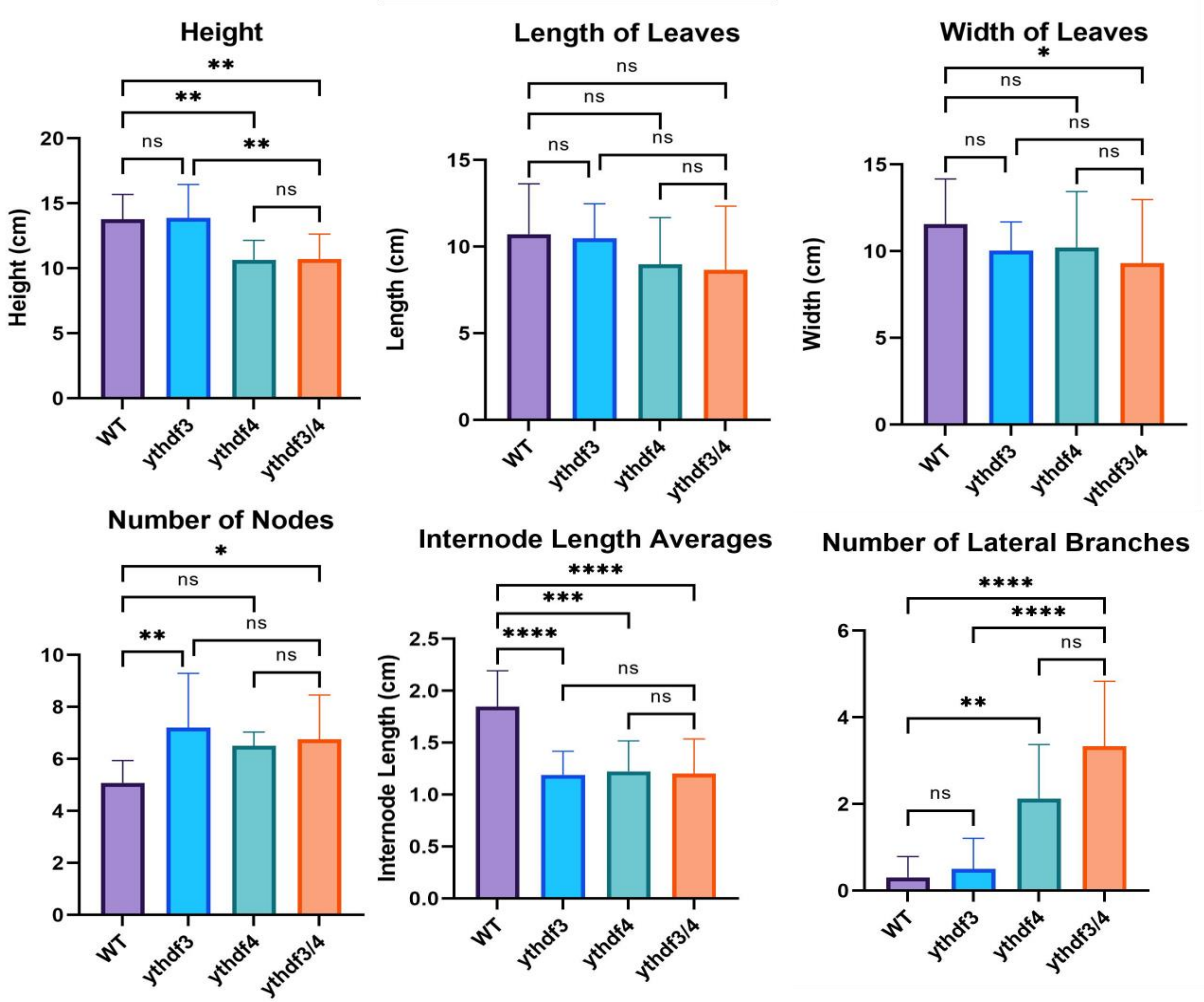


Figure 5.11: Characterisation of *ythdf3, ythdf4* single mutant lines and *ythdf3/ythdf4* mutant. Histogram representation of key plant growth measurements. Plants grown in Levington M3. Measurements taken at 4 weeks in cm. Error bars representing SD values from 10-15 replicates (n=10-15). P values were obtained using mixed effects model and Tukey's multiple comparisons test to determine whether mutant values are significantly different from wild-type values. P value significance bars: * (<0.05), ** (<0.005), *** (<0.0005), **** (<0.0001), ns (No significance)

5.3 Discussion

The focus of this chapter was to evaluate how concurrent KO of two YTHDF proteins, influence the phenotypes we observed from the generation of single ythdf gene mutants. Firstly, we generated F1 plants from the crossing of every single mutant to every other single mutant. These F1 plants were heterozygous for ythdf gene mutations. These F1 plants were self-pollinated and the F2 populations were screened for double homozygous mutant lines. Where we didn't identify the double mutants in the F2 population, we selected F2 plants which were homozygous for one KO allele and heterozygous for the other. These were then selfed to produce F3 seeds, which, were then screened for double mutant genotypes. We have subsequently confirmed homozygous double mutants for *ythdf1/ythdf3, ythdf2/ythdf3,* and *ythdf3/ythdf4.* To the best of our knowledge, this is the first time double homozygous mutant lines have been generated for any plant species outside of *Arabidopsis*, and the double mutant lines between the three/four predicted YTHDF groups (YTHDF1-4) (Scutenaire, J.-M. Deragon, *et al.*, 2018). This represents a significant advance for the study of this protein family *in planta.*

Despite, numerous attempts to identify an *ythdf1/ythdf4* homozygous double mutant line, we were unable to generate a single plant homozygous for both *ythdf1* and *ythdf4.* At least 20 fruits were generated from crossing *ythdf1* and *ythdf4* single mutant plants, producing F1 seeds. The F1 generation was then selfed to produce populations of F2 seeds. We expected that 1 in 16 of these seeds would be homozygous for both *ythdf1* and *ythdf4.* 75 of these F2 seedlings were genotyped with 0 returning homozygous for *ythdf1* and *ythdf2.* However, we did successfully identify *ythdf1--/ythdf4*+- and *ythdf1*-+*/ythdf4*-- plants. Both *ythdf1--/ythdf4*+- and *ythdf1*-+*/ythdf4*—plants were selfed top produce the F3 generation. In this generation we expected that 1 in 4 would be the double mutant. 20 of the F3 seedlings from *ythdf1--/ythdf4*+- and *ythdf1*-+*/ythdf4*—parents were genotyped with 0 returning double homozygous mutants of double homozygous mutants of ythdf1 and ythdf4. For future investigations we generated stable *ythdf1--/ythdf4*+- and *ythdf1*+-*/ythdf4*-- lines. Previous

observations (Figure 4.5) identified significantly reduced germination rates in *ythdf1* and *ythdf4* lines, when compared to WT, *ythdf2* and *ythdf3* single mutant seeds. Discussions on the potential causes of this reduction in germination efficiency led to a hypothesis that *ythdf1* and *ythdf4* aberrant phenotypes may be due to roles of YTHDF1 and YTHDF4 in the gibberellic acid biosynthesis pathway. To investigate the possibility that the lack of homozygous double *ythdf1/ythdf4* is due to the seeds inability to germinate we tested the germination efficiency of the *ythdf1--/ythdf4*+- and *ythdf1*+-*/ythdf4*-- lines. Indeed, both *ythdf1--/ythdf4*+- and *ythdf1*+-*/ythdf4*-- lines showed reduced germination efficiency when compared with WT seeds of a similar age. Despite, treatment with 150mg/L of gibberellic acid increasing the germination rate of both lines (Figure 5.1) close to WT levels. Subsequent screening of the resulting seedlings did not yield any double homozygous plants. These findings suggest that simultaneous KO of both ythdf1 and ythdf4 is lethal in tomato, with seeds not forming or aborting in the ovaries during fruit development. It is possible that this lethality is due to genetic redundancy between ythdf1 and ythdf4 genes. Concurrent KO of both ythdf1 and ythdf4 could result in a breakdown of a crucial function for embryo development. In this case we would expect to see expression of both ythdf1 and ythdf4 during embryo development. If this were true, YTHDF2 and YTHDF3 would not have crucial roles in germination or embryo development. Further investigation would be necessary to evaluate and elucidate these roles. Perhaps, expression of ythdf2 ythdf3, under the power of an ythdf1 or ythdf4 promoter could complement the aberrant development of these lines.

It is unfortunate that due to the Covid-19 pandemic we were unable to confirm the generation of either *ythdf1/ythdf2* and *ythdf2/ythdf4*. Simply knowing whether the double null mutants of these YTHDF genes are viable would provide valuable insight into the overlapping roles of these proteins. The difficulty of finding double mutants for *ythdf1/ythdf2* and *ythdf2/ythdf4*, could suggest that it is indeed impossible to generate these lines using normal generation methods. We were able to successfully identify *ythdf1--/ythdf2*+- and *ythdf1-+/ythdf2*—plants, in the case of *ythdf1/ythdf2* and *ythdf2--/ythdf4*+- and *ythdf2-+/ythdf4*—plants, in the case of *ythdf2/ythdf4* (Figure 5.2; 5.3). This will provide a platform for

confirmation of this like *ythdf1/ythdf4.* We showed that YTHDF1 and YTHDF2 share the greatest homology between the four YTHDF proteins in tomato. It would not be unreasonable to suggest that there is redundancy shared between the two, resulting in lethality. Any overlap of YTHDF1 and YTHDF2 function, during embryo development could explain the inviability in *ythdf2/ythdf4* double mutants, as *ythdf1/ythdf4* are lethal. Our development of YTHDF promoter- GUS fusion constructs would provide a better understanding of expression that would aid in this study.

Measurable physiological changes were observed in all confirmed double homozygous lines. Comparison of these measurements to those obtained in single mutants would allow the identification of any redundancy or interaction between each of the ythdf genes *in vivo* in tomato. Initial phenotypic observation of the ythdf single mutants were carried out in modified Hoagland's solution and plants were grown in a closed growth compartment. To allow for phenotyping of all generated double mutants the phenotyping was performed in compost in controlled glasshouse conditions. Comparison of ythdf single mutant phenotypes to those observed in nutrient solution found that the phenotypes conserved were indeed consistent between these conditions and media. However, due to a low number of seeds and time constraints due to the coronavirus pandemic, only *ythdf1-1, ythdf2-1, ythdf3-1* and *ythdf4-1* lines were used in comparisons with the generated double mutants. Although *ythdf2-2, ythdf3-2* and *ythdf4-2* were shown to have phenotypes like *ythdf2-1, ythdf3-1* and *ythdf4-1,* the lack of comparison in compost potentially affects the validity of these results. Repetition of these experiments with all single mutant lines and double mutants generated from both parent lines would alleviate any concerns around the reproducibility of the observed phenotypes.

Concurrent KO of ythdf1 and ythdf3 provides further evidence to aid in our understanding of potential role of the YTHDF proteins. KO of ythdf3 in the *ythdf1* background shows a partial rescue of *ythdf1* mediated phenotypes. Vertical growth, leaf length and width are all significantly increased in *ythdf1/ythdf3* plants compared to *ythdf1* plants (Figure 5.6, 5.7). This could be considered further evidence that KO of ythdf3 genes results in an increase in vegetative growth. Vertical growth and leaf size is significantly

less than WT, showing the KO of ythdf3 only partially rescues these *ythdf1* phenotypes. The increase in the number of nodes observed after generating ythdf3 mutations are rescued after ythdf1 mutation, returning the number of nodes to WT levels. The decrease in nodes compared to *ythdf3* and the increase in height compared to *ythdf1*, displayed by *ythdf1/ythdf3* plants results in a rescuing of average distance between each node, although internode lengths do not return to WT levels (Figure 5.7). Finally, the number of induced auxiliary meristems increases in *ythdf1/ythdf3*, this could suggest that ythdf1 and ythdf3 both function in a way to regulate apical dominance. Overall, our data suggest that ythdf1 and ythdf3 do not have many overlapping functions. KO of ythdf3 partially rescues ythdf1 growth phenotypes, suggesting that ythdf1 and ythdf3 have contrasting functions in regulating growth. The increase in lateral branching in *ythdf1/ythdf3* plants suggests that YTHDF1 and YTHDF3 proteins have some functional redundancy. Low methylation plants have been shown to exhibit loss of apical dominance, and this was shown to be due (at least in part) to increased stability of WUSCHEL and SHOOT MERISTEMLESS transcripts (Shen *et al.*, 2016). The changes in apical dominance observed in *ythdf1/ythdf3* double homozygous mutant lines could be as a result of YTHDF recognition of these transcripts. Since m^6A was shown to be essential in the regulation of transcript stability, this may point at functions of both YTHDF1 and YTHDF3 in the m^6A mediated degradation of WUSCHEL and SHOOT MERISTEMLESS mRNA.

Like *ythdf1/ythdf3*, generation of a mutant of ythdf2 and ythdf3 shows a significant increase in vertical growth, leaf width and leaf length when compared to *ythdf2* (Figure 5.8, 5.9). Vertical growth of *ythdf2/ythdf3* plants remains less than *ythdf3* and WT levels. In *ythdf2/ythdf3* plants however, the lengths and widths of the 3rd and 4th leaves do not differ from WT, suggesting a more complete rescue than that observed in *ythdf1/ythdf3.* The lack of any significant changes in the number of leaves indicates that there is no redundancy between ythdf2 and ythdf3, however the increase in height in the double mutant compared to *ythdf2* results in an increase in internode length average in *ythdf2/ythdf3*. KO of *ythdf2,* in *ythdf3* lines reduces apical dominance resulting in an increase in lateral branches at 4 weeks. Unlike *ythdf1/ythdf3,* this does not increase more than *ythdf2*. The function of

YTHDF2 proteins in regulating the SAM may therefore be independent of YTHDF3. From these findings we can speculate that there is little to no genetic redundancy between ythdf2 and ythdf3 genes.

Finally, KO of ythdf3 in *ythdf4* plants does not increase the vertical growth in *ythdf3/ythdf4* plants when compared to *ythdf4,* with the double mutant exhibiting significantly shorter vertical growth than *ythdf3.* This is in contrast to *ythdf1/ythdf3* and *ythdf2/ythdf3* plants. Whether this suggests that the increase in vertical growth observed upon mutation of ythdf3 genes is mediated by active YTHDF4, would require further investigation. The length and widths of leaves show no significant differences to either *ythdf3* or *ythdf4* (Figure 5.11). The number of nodes found on *ythdf3/ythdf4* tomatoes does not differ from *ythdf3* or *ythdf4*, however *ythdf3/ythdf4* lines do show significantly increased number of leaves when compared to WT. Internode length averages also do not differ significantly in *ythdf3/ythdf4* from *ythdf3* or *ythdf4,* the lack of any differences suggests that *ythdf3*, which has higher vertical growth, still has slightly higher number of nodes, than *ythdf3/ythdf4.* The increase in lateral branches per plant in *ythdf3/ythdf4* plants, suggests that there are some redundant functions of YTHDF3 and YTHDF4 *in vivo* (Figure 5.10, 5.11)*,* although the increase is not significant from *ythdf4*, and would require further confirmation.

The overlapping roles of YTHDF1, YTHDF2, YTHDF3 and YTHDF4 paint a complicated picture of YTH RNA binding proteins. Each protein likely has several unique functions but some redundancy between each protein is present. In *Arabidopsis*, concurrent KO of ect2 and ect3 was necessary to induce any observable phenotypic changes (Arribas-Hernández *et al.*, 2018; Scutenaire, J.-M. Deragon, *et al.*, 2018). Further KO of ect4 genes only induces minor additional aberrant phenotypes. This level of redundancy is not observed in tomato genomes, increasing the benefits of our YTHDF model.

5.3.1 Future Work

The findings presented within this chapter identify many avenues for future research. Many of the suggested experiments discussed in section 4.3 are also relevant here. Namely the importance of developing low methylation tomato lines to better understand the phenotypic characteristics of the double mutants generated by this work. As previously, RNAi and CRISPR/Cas9 constructs have been designed and constructed to target key components of the methylation complex in tomato. Other initial experiments which would be invaluable would-be further characterisation of both single and double mutant lines in tomato. One of the limitations of this study is the narrow window of observation. Broadening our window of observation is necessary for future studies. Many of the phenotypes associated with low methylation (Zhong *et al.*, 2008; Bodi *et al.*, 2012b; Arribas-Hernández *et al.*, 2020) are similar to those caused by impaired auxin function, such as aberrant phyllotaxis, gravitropism, leaf morphologies or changes in floral development (Heisler *et al.*, 2005; Su *et al.*, 2017; Bhatia and Heisler, 2018). Additional emphasis should therefore be placed on the relationship between our observed phenotypes and mis-regulation of hormone biosynthesis factors, transporters or response factors, specifically auxin and gibberellic acid (Section 4.3).

The lack of additional *ythdf1* mutants is a limitation of this work, this carries forward to the generated double mutants. Replication of the phenotyping of *ythdf1-1* in another line would add weight to any findings in this study. To mitigate this limitation, we have designed antibodies (Agrisera AB, Vännäs) for all four YTHDF proteins in tomato. Once produced these will allow western blot analysis to confirm the lack of YTHDF proteins in each of the null lines and (hopefully) to allow RIPSeq experiments to identify target mRNAs that they bind.

A complete understanding of expression of each of the ythdf genes, would aid in the discussion of these double mutant phenotypes. Until we have this information the hypotheses laid out in this chapter will remain conjecture. Even a basic understanding of developmental expression of YTHDF proteins would give invaluable direction for our future work on this model. qRT-PCR or SAGE could be utilised in this regard, however our development of

YTHDF promoter- GUS fusion constructs would provide a better understanding of expression that would aid this study.

Perhaps the most elucidating next step would be to increase our understanding of which transcripts in tomato are methylated and which of those transcripts each of the YTHDF proteins interact with. meRIP-Seq, a combination of ChIP-seq and RNA-Seq, could be used to detect transcriptome-wide m^6A modifications (Dominissini *et al.*, 2012; Meyer *et al.*, 2012). This method would require a quality reference transcriptome to be compared with the generated methylome. A limitation of this approach would be that meRIP-Seq only provides insight into a single developmental stage. Although, repeats of samples from different developmental stages will alleviate this. Development of a transcriptome-wide library of m^6A modification would also allow us to compare the transcripts with enrichment for m^6A, to transcripts which are shown to bind to each YTH protein. The complexity of mechanisms involving m^6A only increase the importance of future developments in this field.

Chapter 6 - *In-vitro* selection with specifically modified RNA library for the identification of YTHDF binding preferences in *Solanum lycopersicum*

6.1 Introduction

YTHDF proteins are typically cytoplasmic proteins, many of which contain a long intrinsically disordered region (IDR) near the N-terminus, followed by the YTH domain itself (Arribas-Hernández *et al.*, 2018). YTHDF proteins have often been suggested to control mRNA, via decay or translational activation, however the mechanisms involved in these processes remain unclear. Answers to the open questions requires systematic characterisation of mRNA-YTHDF binding preferences. In plants, there has been some debate as to the correct recognition sequence. The high level of conservation between the eukaryotic catalytic core of the methylase complex, suggests that plant YTHDF proteins would recognised the RR(m^{6}A)CH context in which m^{6}A is deposited in all currently investigated (Dominissini *et al.*, 2012; Meyer *et al.*, 2012; Schwartz, Maxwell R. Mumbach, *et al.*, 2014; S. Luo and Tong, 2014; Chen *et al.*, 2015; Lence, Soller and Roignant, 2017; Anderson *et al.*, 2018; L.-H. H. Wei *et al.*, 2018; Miao *et al.*, 2020a; Parker *et al.*, 2020). The Jia lab has suggested that *Arabidopsis* YTHDF protein, ECT2, binds to a plant specific motif of URU(m^{6}A)Y (L.-H. H. Wei *et al.*, 2018). This however has likely been explained by inaccuracies derived from protein-protein interactions, not protein-RNA interactions, resulting from formaldehyde cross-linking (Arribas-Hernández *et al.*, 2021) Indeed, recent improvements in nanopore-based sequencing have allowed single resolution mapping of m^{6}A. Data from these studies do not support URUAY-ECT2 binding, as little URUAY-sites were found to contain even m^{6}A (Parker *et al.*, 2020). In fact, recent investigations by the Brodersen lab found that URUAY motifs are more enriched near to m^{6}A sites, with no observable evidence of AtECT2 binding. This suggests that URUAY repels AtECT2 to

some extent (Arribas-Hernández *et al.*, 2021). YTHDF-mRNA binding is a highly complex area and represents a barrier in furthering our understanding of these proteins. Any developments within plant models would be highly beneficial in advancing our knowledge of m^6A functions in plant species.

6.2 Results

6.2.1 *in Vitro* selection strategy

In order to assess and profile the binding preferences for each of the four YTHDF protein *in vitro* we adapted a protocol recently developed by the Kleiner lab (Arguello, Leach and Kleiner, 2019). This method utilises synthetically synthesised RNA oligonucleotides and tags and purified RNA binding proteins. Since YTHDF proteins have been shown to recognize short sequence motifs (Gerstberger, Hafner and Tuschl, 2014; L.-H. H. Wei *et al.*, 2018; Arribas-Hernández *et al.*, 2021) we designed the experiment based around 9 bp variable sequences with a single A or m^{6}A residue at the central position 5. These were flanked on either side by 10 nucleotides of defined sequence. Identification of the binding preferences, without bias requires a pool of oligo nucleotides with random incorporation of bases at each position, creating a population of equally represented consensus sequences around both A and m^{6}A residues. Each oligonucleotide was synthesised with the 10 nucleotides of defined sequence containing unique identifiers for either A or m6A (GUGAGUCUUCNNNN-A-NNNNUCCUGCAGCG or GUGGAUCUUCNNNN-m6A-NNNNUCCUCGAGCG) (Figure 6.1-A). This is necessary as m^{6}A is recognised as A during reverse transcription. The resulting oligonucleotide mix will be comprised of 4^8 (~65500) unique, A containing sequences and an equal number of m^{6}A containing sequences. Tagged proteins could then be incubated with the oligo pools, washes to remove sequences with lower affinity for each protein, before releasing the RNA digesting the RNA off the bound protein. small RNA libraries were produced for each sample of enriched RNA before next-generation sequencing of each library (Figure 6.1-B). This adapted approach would allow us to not only identify any specific affinity to a particular sequence of consensus sequence, but the novel approach to mix oligonucleotides also containing both A and m^{6}A would provide an opportunity to detect any preference of the YTHDF protein for A or m^{6}A, giving support to their anticipated roles as *bona fide* m^{6}A binding proteins.

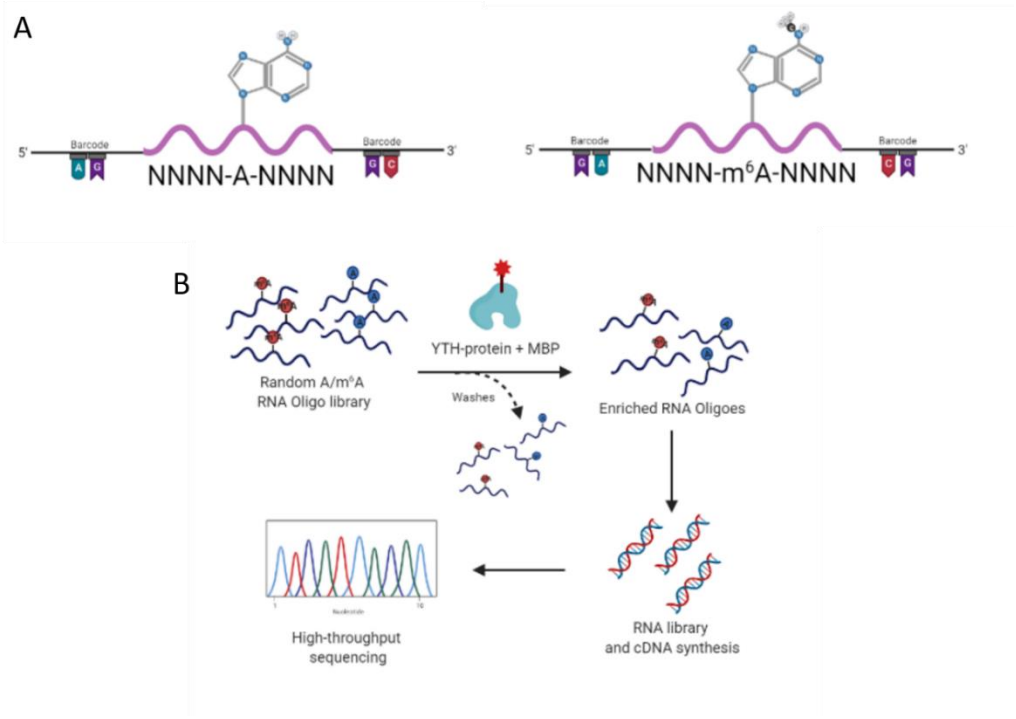


Figure 6.1: *In vitro* selection strategy utilised to identify binding preferences of *Solanum lycopersicum* YTHDF RNA binding proteins. (A) Make up of random RNA oligonucleotide sequence pools. (B) Diagrammatic protocol flow of selection strategy.

6.2.2 Purification of tomato YTHDF proteins

Bacterial expression and purification of each YTHDF protein utilised the pMAL-c5x system (New England Biolabs). Reverse transcription of tomato poly(A) RNA (Methods 2.6.2) followed by gene-specific amplification with primers. Each 3' primer introduced 6 histidine codons immediately before the stop codon. These amplified cDNAs were then cloned into pMAL-c5x, via XmnI and SbfI digestion and Gibson assembly (Methods 2.3.12). Completed contructs were verified by vector sequencing. Cloning cDNA into pMAL-c5x results in proteins fused to maltose binding protein 5 (MBP5) at their amino terminus, which was subsequently used as a tag for the affinity purification experiments. Getting purified proteins to be soluble from the cultures proved difficult, Repeated attempts to purify the proteins, resulted in the YTHDF proteins being found in the insoluble pellet (Figure 6.2-B). This is likely due to the intrinsic nature of YTHDF proteins to form aggregation *in*

vivo (Arribas-Hernández *et al.*, 2018, 2021; Scutenaire, J. M. Deragon, *et al.*, 2018). We tried several temperatures and conditions and found that, a multi-faceted strategy of growth at 12◦C and 200rpm and the addition of 2% Ethanol to the media allowed the production of soluble YTHDF proteins using this method (Figure 6.2-C). Figure 6.2-C shows a representative SDS-PAGE gel image from YTHDF1 bulk purifications (Images of YTHDF2, YTHDF3 and YTHDF4 bulk purification: Appendix 6.1). Bulk purifications with nickel resin affinity columns were performed before each protein was further purified with Amicon 50kDa size selection filtration columns resulting in pure samples of each YTHDF protein (Figure 6.2- D).

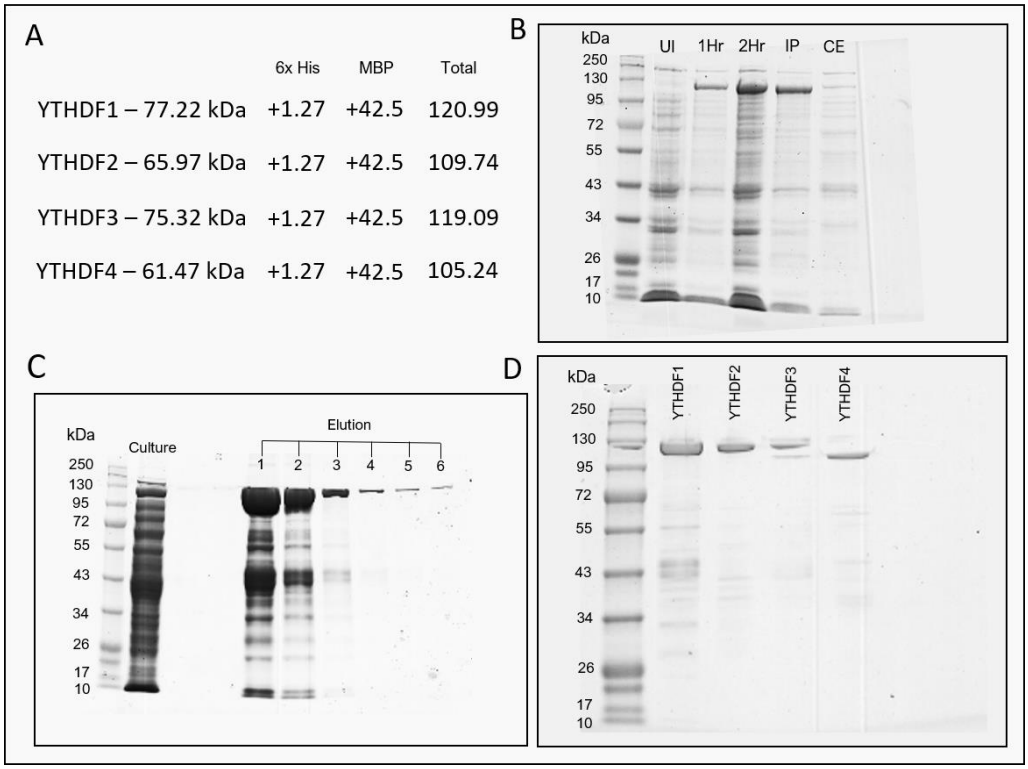


Figure 6.2: Purification of *Solanum lycoperisum* YTHDF proteins (A) Molecular weight of YTHDF1, YTHDF2, YTHDF3 and YTHDF4 before and after addition of His-tag and Maltose Binding Protein 5 (MBP) (kDa) (B) SDS-PAGE protein gel from Pilot experiment of YTHDF1 protein production construct, DH5a *E.coli* grown at 37◦c and 200rpm. UI-Uninduced, 1Hr- 1 Hr post-induction, 2Hr- 2Hr post-induction, IP- Insoluble pellet, CE-Crude soluble extract. (C) SDS-PAGE protein gel of YTHDF1 post purification at 12◦c and 200rpm. (D) SDS-PAGE protein gel of all four YTHDF proteins after dialysis and amicon 50kDa size selection filtration columns.

6.2.3 Synthesis of Random RNA libraries

The production of equally distributed RNA oligonucleotide libraries is an essential component of our experiment. Dr Francis Baron (then a student in our lab) first synthesised N^6-methyladenosine using the nucleophilic displacement method (Ross, Bruno and Petersen, 1963). A method developed by Kruse et al demonstrated that Aritomo's phase transfer catalysis method could be modified to produce m^6A phosphoramidites (Kruse *et al.*, 2011). These m^6A phosphoramidites were then incorporated into random oligonucleotide sequences in a ratio of A/C/G/U = 0.30:0.25:0.23:0.20. These ratios reflect the different coupling efficiencies of the phosphoramidites, resulting in a 1:1:1:1 ratio at each "random" position within the oligonucleotide (Frank Baron, *Unpublished*).

6.2.4 Utilisation of *in vitro* Selection Platform

The ability of each YTHDF protein as a potential m^6A reader protein was evaluated using the selection platform described above. Firstly, we calculated that a minimum of 143 pmol was required to saturate the amylose-magnetic beads. For each sample double this concentration (286pmol) of each YTHDF-MBP protein fusion was immobilised onto amylose-magnetic beads (NEB) to ensure complete enrichment of 200ul of beads. Equal concentrations of both the A and m^6A random oligo nucleotide mixes were incubated with the immobilised proteins. Stringent washes were performed to remove any unbound RNA. The RNA was released off the beads by digesting the YTHDF proteins using proteinase K (Sigma-Aldrich). This step was necessary due to the inability of standard approaches using maltose to elute the proteins off the beads (Supplementary 6.2). The resulting samples containing RNA and peptide fragments was cleaned and concentrated to 6ul (Zymo RNA Clean and Concentrate Kit). 5' and 3' adapters were then ligated onto each small RNA oligonucleotide and RT-PCR utilised to produce multiplex compatible small RNA libraries (NEBNext Small RNA library kit). After adaption and production of cDNA libraries, each oligo nucleotide should be a total of ~ 150bp. The binding

affinity experiment was repeated twice for each of the four YTHDF proteins. The library production was also repeated with the RNA oligonucleotide mix to create an input library for sequencing. Analysis of the produced libraries using Agilent High Sensitivity DNA assay was performed (Figure 6.3). Figure 6.3 shows (with the exception of Input 1) pure peaks at around 150bp, suggesting the successful production of the correct DNA libraries. Input 1 lacked any recognisable peaks at 150bp, indicating that no DNA library was produced for this sample. Unfortunately, due to the Covid-19 outbreak, we had to proceed with only a single sample of input for this sample.

All 9 libraries were sent for next generation sequencing (Deep Seq - University of Nottingham). Prior to sequencing all libraries were further purified and size selected utilising BluePippin (Sage Science), collecting all DNA fragments between 147-151bp. The 9 samples were split into 4 lanes and sequenced using an Illumina NextSeq500 system.

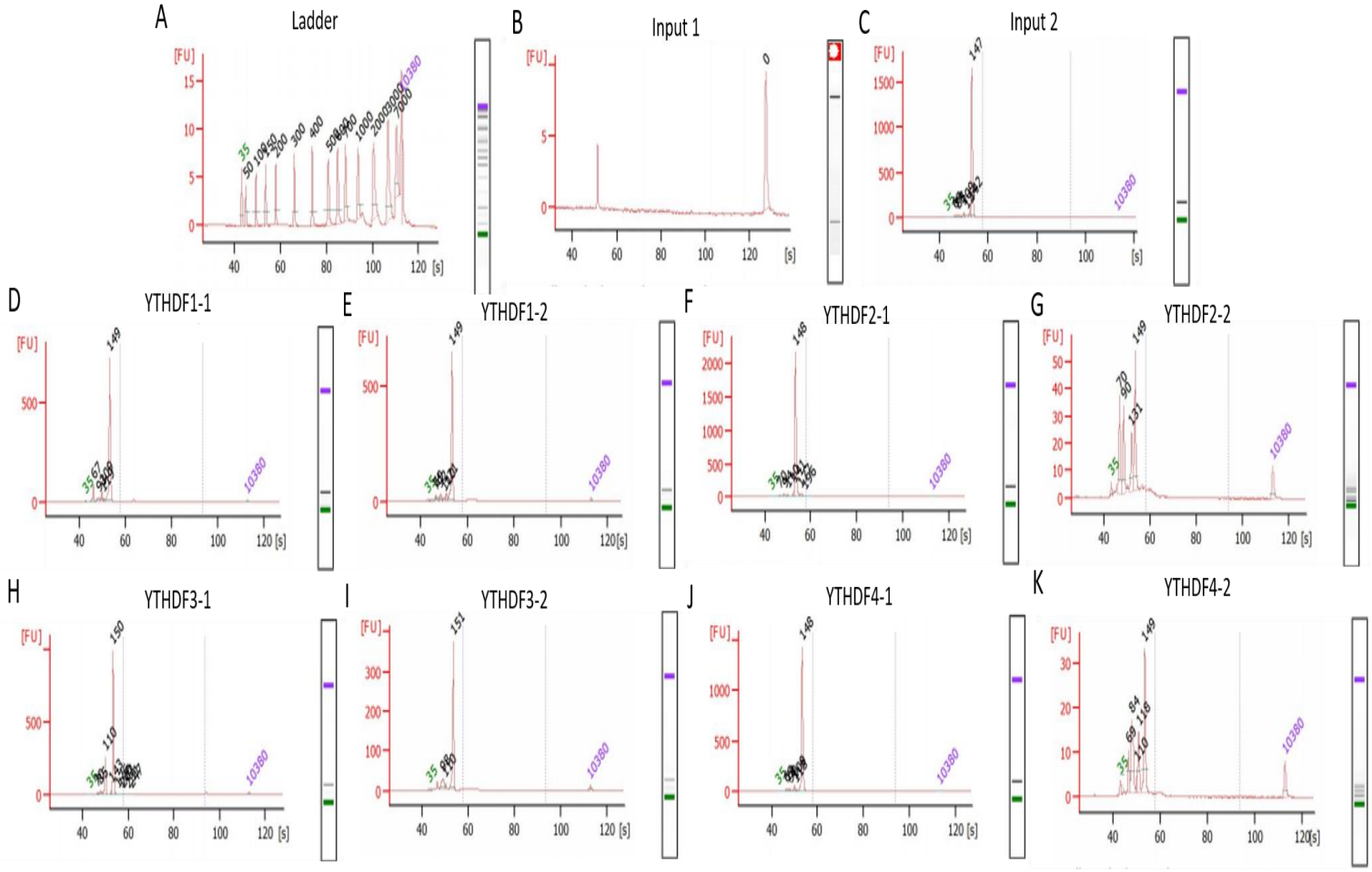


Figure 6.3: Electropherograms from high sensitivity DNA assay of YTHDF *in-vitro* RNA selection libraries after NEBNext Small RNA library production. (A) Ladder control (B-C) Input1 and Input 2 DNA libraries from RNA selection (D-E) YTHDF1-1 and YTHDF1-2 DNA libraries from RNA selection (F-G) YTHDF2-1 and YTHDF2-2 DNA libraries from RNA selection (H-I) YTHDF3-1 and YTHDF3-2 DNA libraries from RNA selection (J-K) YTHDF4-1 and YTHDF4-2 DNA libraries from RNA selection.

6.2.5 Sequencing analysis of YTHDF-protein RNA selection

In order to analyse the sequences resulting from the *in vitro* selection using YTHDF1, YTHDF2, YTHDF3 and YTHDF4, custom DAX and bash scripts were developed to extract the 9-mer from each oligonucleotide and label them as A containing or m^6A containing sequences. Additional scripts were then used to count the abundance of each unique sequence and compare these to the relative abundance of the input. The relative abundance of each sequence was normalised to account for slight variation in the abundance of each unique sequence in the input. The relative abundance of each base at each position was also calculated for every sample.

The Illumina based sequencing method generated between 1.5 and 6.5 million reads per sample. 6.2.5.1 Characterisation of Input libraries

The sequences from the input were analysed to assess the comparable distribution of all bases and of both A and m^6A containing sequences within

the mix. The single input library's 1516232 reads from this study demonstrates a 48.3% to 51.7% A:m^6A ratio, the inputs also contain 65149 and 65128 of the possible 65536 combinations containing A and m^6A respectively. Any comparisons of enrichments of A vs m^6A was compared to this distribution. The comparable distribution of each of the 9 bases was analysed to ensure equal incorporation of the four bases in positions 1-4 and 6-9, and exclusive incorporation of A at the centre of each oligonucleotide fragment (Figure 6.4-A/C). The distribution of each base in the positions show that there is an uneven distribution of the four bases but an almost complete incorporation of A at position 5 for both A and m^6A. Due to the depth of sequencing this variation is likely present in the synthesised oligonucleotide populations and as such will be consistent across the YTHDF affinity libraries. To improve the accuracy of comparisons the oligonucleotides sequenced from the four YTHDF *in vitro* selection libraries were normalised against the expected proportions in the input. The ratios of each base at each position were also normalised against the input. Normalising the input using this method results in a more evenly balanced distribution (Figure 6.4-B/D).

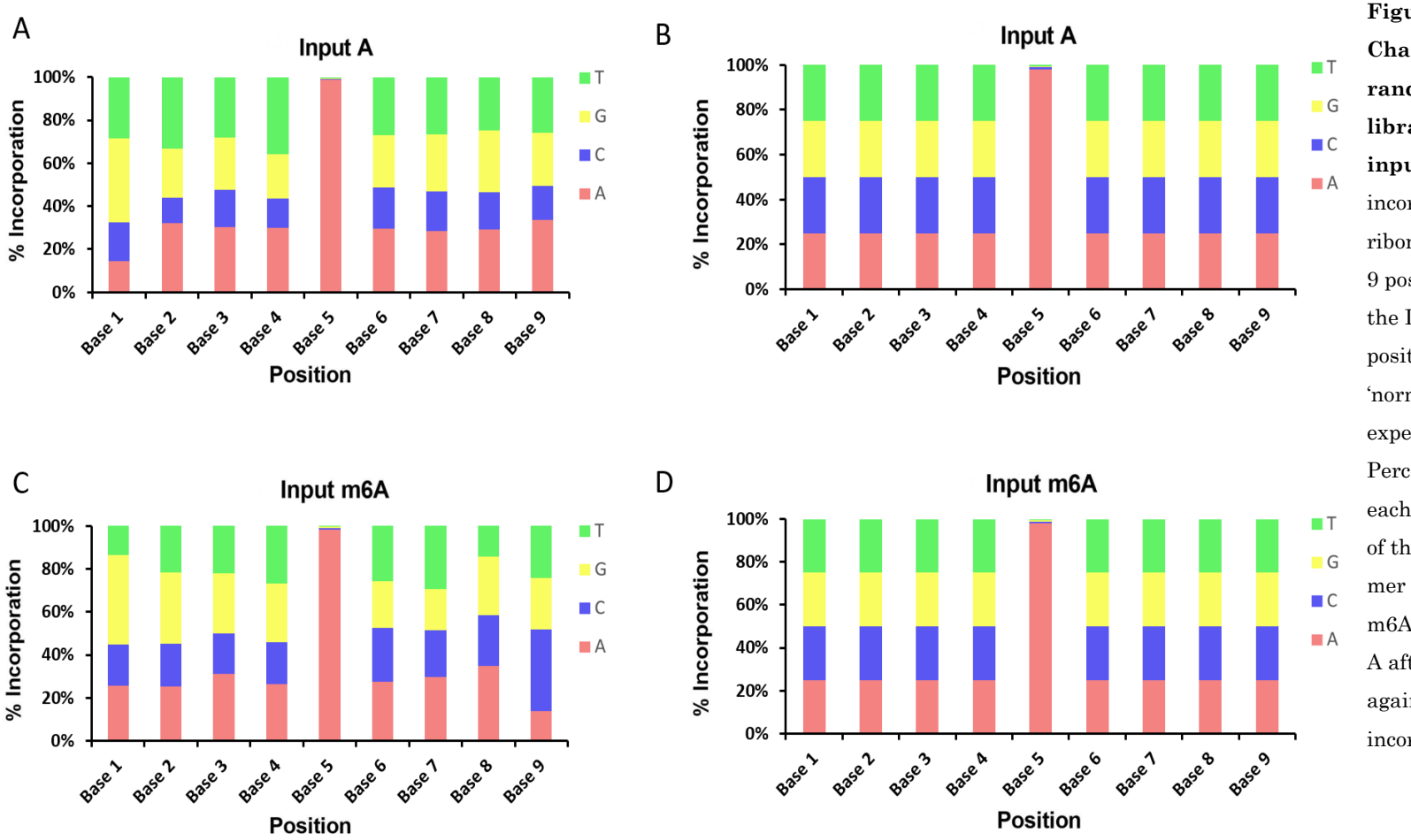


Figure 6.4: Characterisation of random sequencing libraries present in input: (A) Percentage incorporation of each ribonucleotide at each of the 9 position in the 9-mer in the Input containing A at position 5. (B) Input A after 'normalisation' against expected incorporation (C) Percentage incorporation of each ribonucleotide at each of the 9 position in the 9-mer in the Input containing m6A at position 5 (D) Input A after 'normalisation' against expected incorporation.

6.2.5.2 Analysis of YTHDF1 *in vitro* selection libraries

YTHDF1-1 and YTHDF1-2 sample libraries were cleaned and analysed using the same custom scripts developed for the input sequence library. After removal of all ambiguous 9-mer sequences, or any sequences which contained a C, G or T at position 5, there were 1955872 and 2800317 sequence reads for oligonucleotides containing A, and 1829392 and 1682406 for oligonucleotides containing m^6A for YTHDF1-1 and YTHDF1-2, respectively. This gives a 51.7% to 48.3% and 62.5% to 37.5% A:m^6A ration for YTHDF1-1 and YTHDF1-2, with an average of 57.1% to 42.9%. Once compared to the relative abundance of A vs m^6A in the input mix, YTHDF1 shows an average enrichment of +8.77% for A over m^6A (which is the opposite of what we were anticipating).

We next chose to focus on the changes in percentage representation of each base at base positions 1-4 and 6-9. This would give a suggestion of any preferred bases within the consensus sequence (Figure 6.5). Both YTHDF1-1 and YTHDF1-2 A-containing sequences show little variation at each position, percentage changes were averaged for both YTHDF1-1 and YTHDF1-2. The largest preference is shown to be an average 8.94% increase in C at position 9 (+8.49%) and the least preferred base at position 1 being G (-5.28%). Oligonucleotides containing m^6A on the other hand display a large degree of variation. Again, G is the least preferred base at position 1 (-7.07%) with both C and T preferred over A (+6.14% and +6.97%, respectively). A slight increase of A (+2.68%) is observable at position 2. Position 3 shows a decrease in C (-2.35%) and a preference for T over A and G (+1.01%, +0.23% and +0.53%, respectively), this loosely follows the consensus DR(m^6A)CH (D = A/G/U). Position 4, adjacent to the m^6A shows an increase of A and C (+1.30% and +3.53%, respectively), with slight decrease of G and a -3.28% decrease of the presence of T. This does not follow the published consensus of DR(m^6A)CH, where R = purines A/G. Downstream of the m^6A residue, there is a slight enrichment of C and T (0.34% and 1.52%), with a reduction in A and G, again not falling into published consensus sequences for m^6A deposition. Position 7 displays an increase in C and G (+1.52% and +1.18%) and a decrease of A (-2.11%).

Position 8 has a large increase in the proportion of T of 24.4% and an increase of +1.56% of A, when compared to the input with a decrease of A and G (-3.90% and -9.28%, respectively). Finally, position 9 is enriched for C (+5.57%) and a decrease in A, G and T (-3.93%, -6.5% and -0.08%, respectively). Overall these results show that the overall sequences selected do not follow the DR(m^6A)CH consensus. It also suggests that YTHDF1 proteins may have a higher binding affinity for sequences containing A over m^6A. However, there does appear to be more sequence specificity in the m^6A context than in an A context.

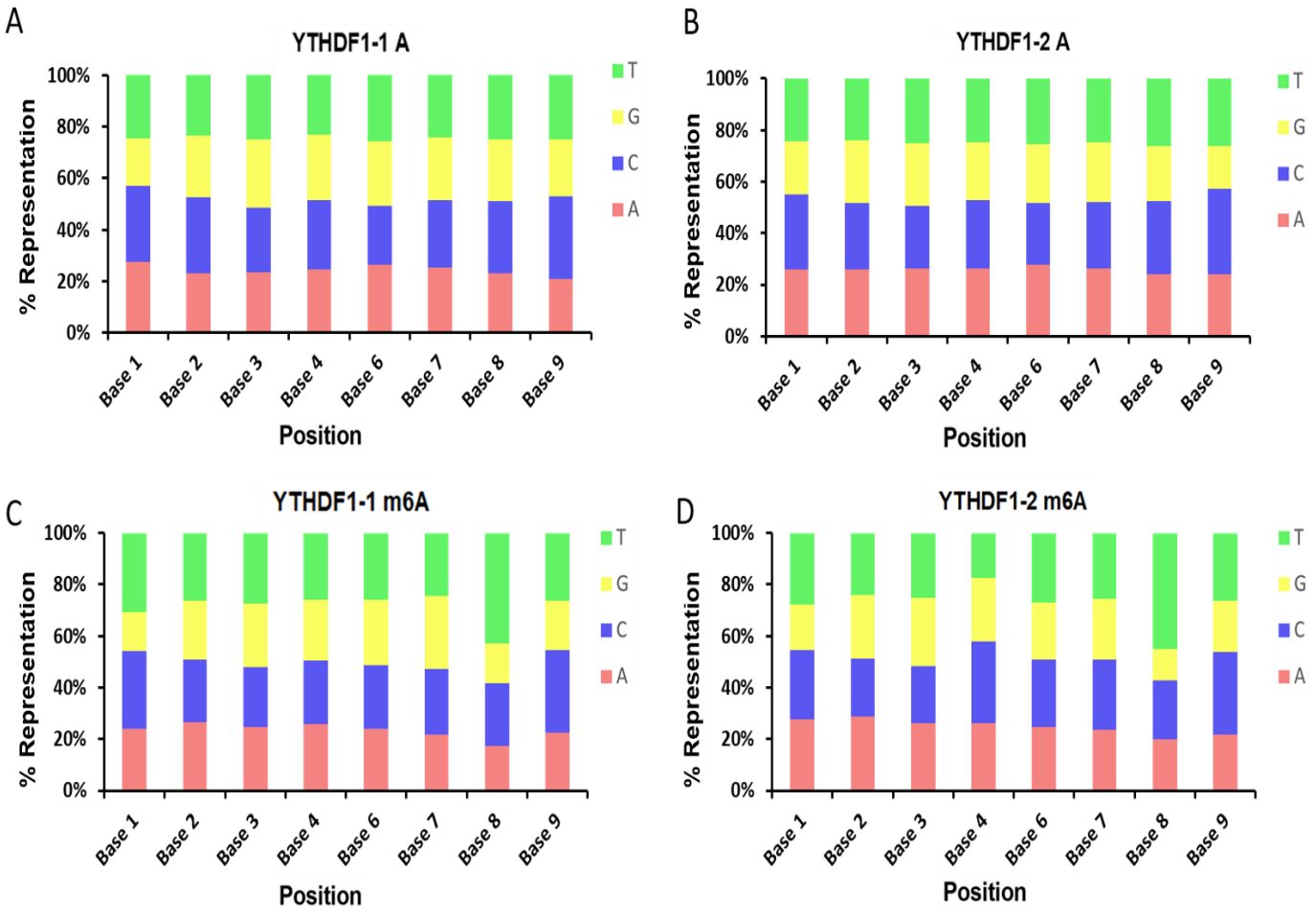


Figure 6.5: Characterisation of random sequencing libraries present in YTHDF1-1 and YTHDF1-2 *in vitro* selection libraries: (A) Percentage incorporation of each ribonucleotide at each of the 9 position in the 9-mer in sample YTHDF1-1 containing A at position 5. (B) Percentage incorporation of each ribonucleotide at each of the 9 position in the 9-mer in sample YTHDF1-2 containing A at position 5 (C) Percentage incorporation of each ribonucleotide at each of the 9 position in the 9-mer in sample YTHDF1-1 containing m^6A at position 5 (D) Percentage incorporation of each ribonucleotide at each of the 9 position in the 9-mer in sample YTHDF1-2 containing m^6A at position 5. All percentages normalised against input.

6.2.5.3 Analysis of YTHDF2 *in vitro* selection libraries

YTHDF2-1 and YTHDF2-2 sample libraries were cleaned and analysed as above. After removal of all ambiguous 9-mer sequences or any sequences which contained a C, G or T at position 5, there were 1534082 and 722246 sequence reads for oligonucleotides containing A, and 1820142 and 2599613 oligonucleotides containing m^6A for YTHDF2-1 and YTHDF2-2, respectively. This gives a 45.7% to 54.3% and 21.7% to 78.3% A:m^6A ratio for YTHDF2-1 and YTHDF2-2, with an average of 33.7% to 66.3%. Once compared to the relative abundance of A vs m^6A in the input mix, YTHDF2 shows an average enrichment of +14.56% for m^6A over A, contrasting the result observed in YTHDF1-1 and YTHDF1-2 and suggesting that YTHDF2 is a *bone fide* m^6A binding protein.

Unlike YTHDF1-1 and YTHDF1-2, percentage representation of the four bases at each position of sequences containing A shows more variation when compared to the input (Figure 6.6-A/B). Position 1 shows a +15.78% increase of C, with reductions in A, G and T (-1.41%, -6.70% and -0.01%, respectively). At position 2, C and G are increased (+3.19% and +4.17%) and A and T decreased (-2.37% and -1.675). There is little difference between the bases at position 3, with a slight preference for G (+0.94). Position 4 shows a large increase in C and G (+12.25% and +6.14%) and a reduction in T (-8.00%). At position 6, a slight preference for G and T over A and C can be observed. Followed by a preference for C (+2.88%) at position 7. A preference for C can also be seen at positions 8 and 9 (+5.60% and +6.20%, respectively) and slight preference for T (+1.08% and +0.85%, respectively). A and G are both reduced. This demonstrates a higher level of specificity than observed in YTHDF1 libraries.

When nucleotide representation is compared in m^6A sequences selected we also see changes when compared to input (Figure 6.6- C/D). Position 1 is enriched for both C and T (+6.33% and +4.56%) with a reduction in selection for G (-4.56%). Position 2 base incorporations show an increase of purine bases A and G (+1.29% and +1.83%) and a reduction in C and T (-3.47% and -1.205, respectively). There is little difference in percentage representation at position 3. Position 4 displays an increase in C and T (+1.53 and +2.76%, respectively), and a reduction in G of -3.98%. Just downstream of the m^6A there is a slight preference for sequences containing T (+1.61%) with a -1.73% decrease in A. Sequences containing a G or a T at position 7 are slightly enriched (+2.37% and +1.23%)

with again a -3.36% reduction in sequences containing A. Like YTHDF1 libraries, T is significantly enriched (+22.89%) and C slightly increased (+1.01%), with both purines A and G reduced (-5.65% and -5.75%, respectively). Again, sequences ending in C are enriched in the context of m^6A (+7.37%) with A, G and T reduced (-6.13%, -6.24% and -1.91%, respectively). Overall these results show that the overall sequences selected do not follow the DR(m^6A)CH consensus. m^6A is enriched over A. There is also more sequence specificity in the A context than observed in YTHDF1 *in vitro* selection samples.

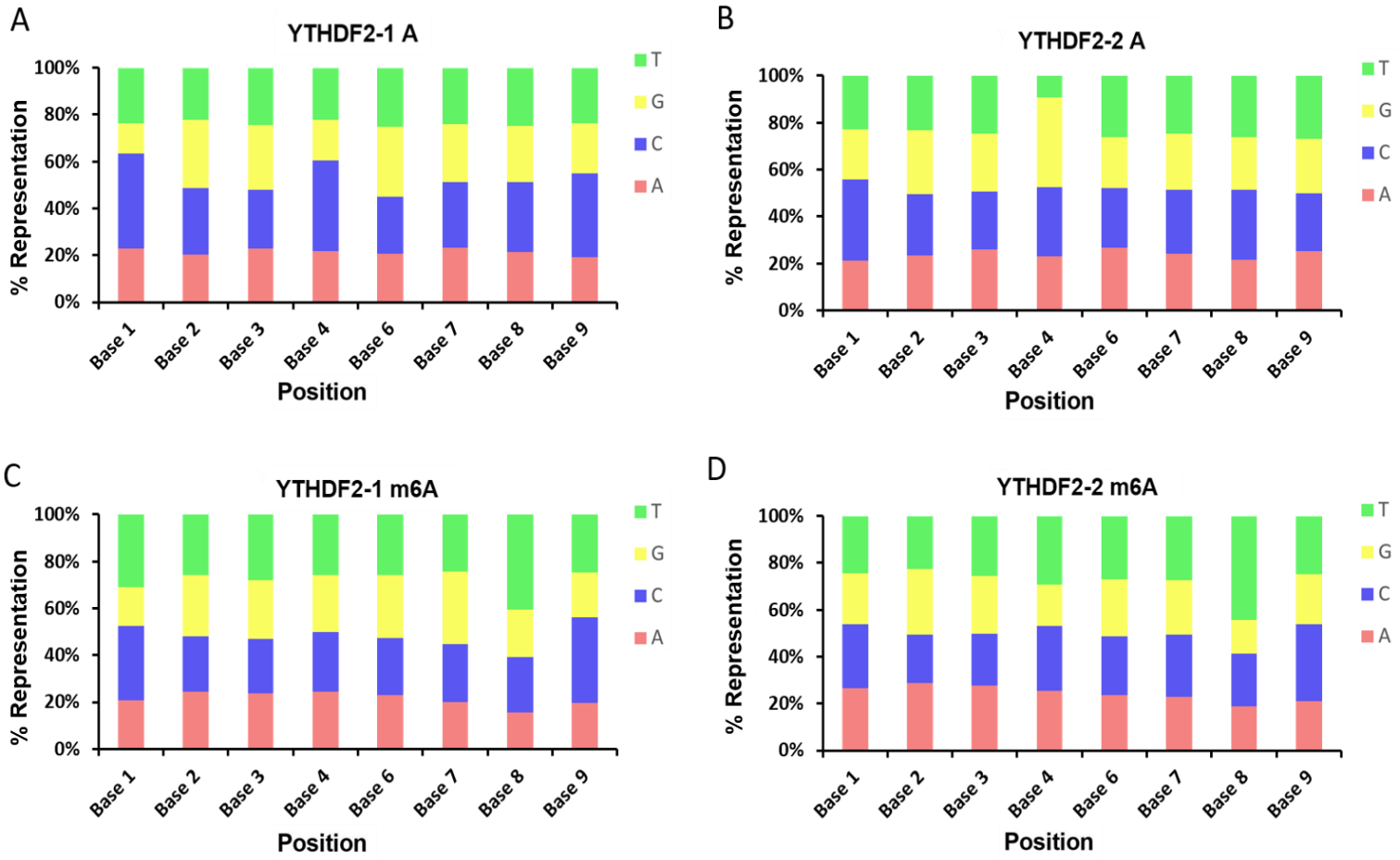


Figure 6.6: Characterisation of random sequencing libraries present in YTHDF2-1 and YTHDF2-2 *in vitro* selection libraries: (A) Percentage incorporation of each ribonucleotide at each of the 9 position in the 9-mer in sample YTHDF2-1 containing A at position 5. (B) Percentage incorporation of each ribonucleotide at each of the 9 position in the 9-mer in sample YTHDF2-2 containing A at position 5 (C) Percentage incorporation of each ribonucleotide at each of the 9 position in the 9-mer in sample YTHDF2-1 containing m^{6}A at position 5 (D) Percentage incorporation of each ribonucleotide at each of the 9 position in the 9-mer in sample YTHDF2-2 containing m^{6}A at position 5. All percentages normalised against input.

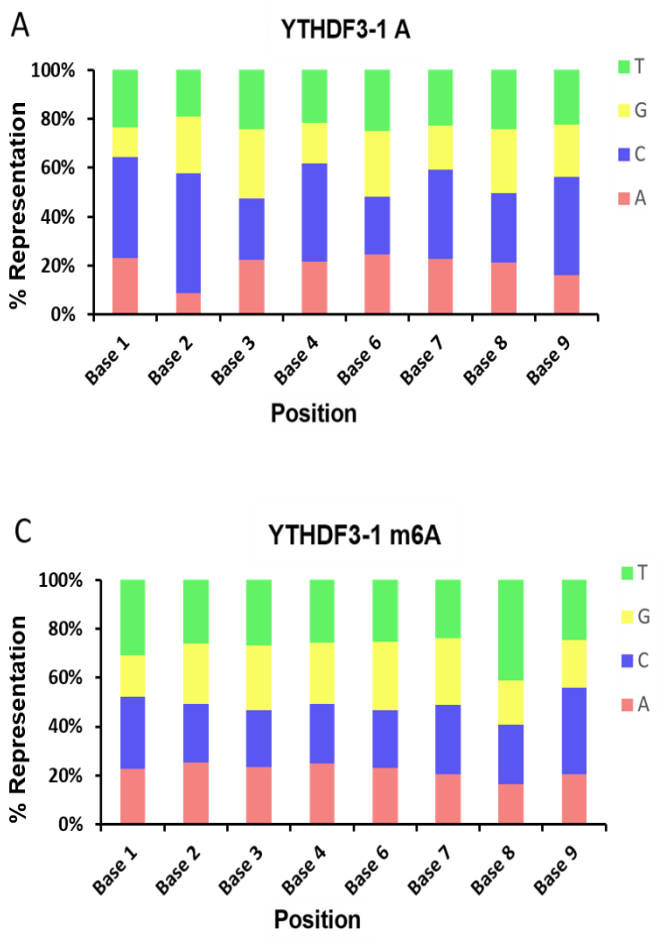
6.2.5.4 Analysis of YTHDF3 *in vitro* selection libraries

YTHDF3-1 and YTHDF3-2 sample libraries were analysed as above. After removal of all ambiguous 9-mer sequences or any sequences which contained a C, G or T at position 5, there were 3044876 and 1620955 sequence reads for oligonucleotides containing A, and 3392676 and 2137372 oligonucleotides containing m^6A for YTHDF3-1 and YTHDF3-2, respectively. This gives a 47.3% to 52.7% and 43.1% to 56.9% A:m^6A ration for YTHDF3-1 and YTHDF3-2, with an average of 45.2% to 54.8%. Once compared to the relative abundance of A vs m^6A in the input mix YTHDF3 shows an average enrichment of +3.09% for m^6A over A.

Like YTHDF2, there are wide ranging changes in the base representation observed in the sequences from high-throughput sequencing (Figure 6.7- A/B) In sequences centred around A, the consensus sequences show a preference for C at position 1 (+5.29%) over A and T and a reduction in G (-2.46%). Position 2 shows large scale increases for C with a 32.60% increase in sequences containing C, with reductions in A and T (-1.97% and -2.07%, respectively) and large reduction in G (-10.88%). Position 3 does not show the same level of preference for specific bases, however there is a slight increase in sequences containing C and G over A and T. A similar picture is observed for position 4 with a preference for C and G (+2.43% and 1.48%, respectively) over A and T. There is very little difference observed after the A at position 5. Position 7, 8 and 9 show a preference for C over the other bases, with +7.13%, +5.58% and +6.24% increases when compared to the input for positions 7, 8 and 9, respectively. A, G and T show reductions for these positions. There seems to be an overall preference for C across YTHDF3 A containing sequences, with a preference observed in all positions.

Base proportions in sequences centred around m^6A also vary from position to position (Figure 6.7 – C/D). The start of the sequence is more likely to be a C or a T (+6.14% and +5.18%), with also a slight preference for A. G is selected against with a reduction of 5.18%. There is a slight preference for A at position 2 (+1.46%), with C slightly reduced (-1.76%). Position 3 is similar with a reduction of 2.15% for sequences containing C. Position 4 shows a slight preference for A or T (+0.80% and +1.54%, respectively). After the m^6A residue, there is a higher preference for G or T, over A or C (+4.19% and +1.01%). Position 7 shows a preference for C (+2.96%) and a reduction in sequences containing A (-2.73%). Like both YTHDF1 and YTHDF2, there is large bias for sequences containing T at position 8, when compared to the input sequences (+22.92%) and a reduction in A, C and

G (-4.69%, -2.81% and -3.67%, respectively). Finally, the sequences are again more likely to end in a C than A, G and T (+6.74%), and more likely to end in T than A or G (-1.78%, -5.69% and -5.61%, respectively). These results again show that the overall sequences selected do not follow the DR(m^{6}A)CH consensus. m^{6}A is slightly enriched over A. In the context of A there is a significant preference for C across the whole consensus with specific preferences at all positions except position 6. This is not the same in the context of m^{6}A, with varied specificity across the 8 bases, with some similarities to YTHDF1 and YTHDF2.



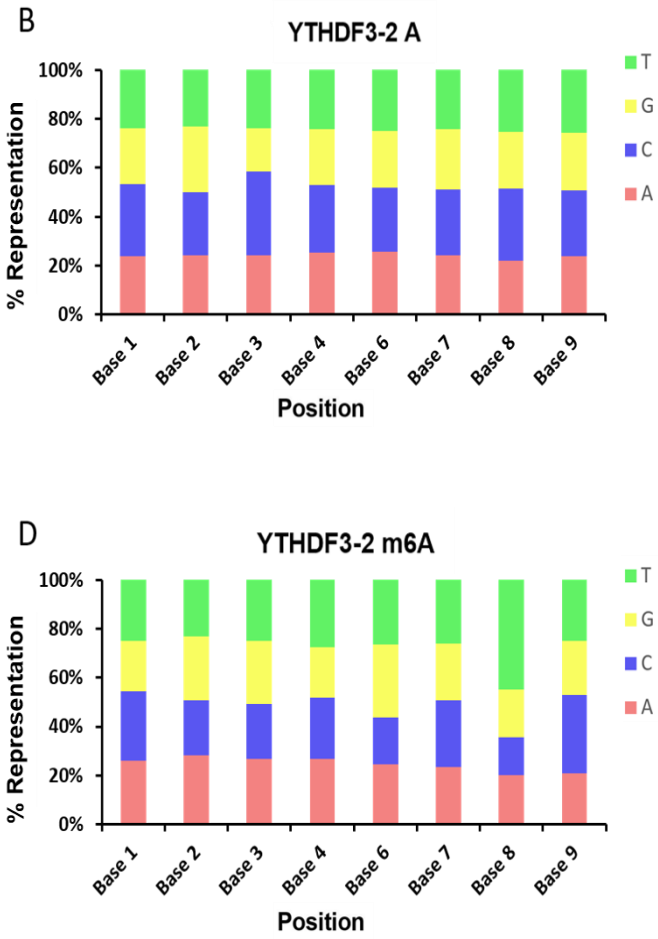


Figure 6.7: Characterisation of random sequencing libraries present in YTHDF3-1 and YTHDF3-2 *in vitro* selection libraries: (A) Percentage incorporation of each ribonucleotide at each of the 9 position in the 9-mer in sample YTHDF3-1 containing A at position 5. (B) Percentage incorporation of each ribonucleotide at each of the 9 position in the 9-mer in sample YTHDF3-2 containing A at position 5 (C) Percentage incorporation of each ribonucleotide at each of the 9 position in the 9-mer in sample YTHDF3-1 containing m^6A at position 5 (D) Percentage incorporation of each ribonucleotide at each of the 9 position in the 9-mer in sample YTHDF3-2 containing m^6A at position 5. All percentages normalised against input.

6.2.5.5 Analysis of YTHDF4 *in vitro* selection libraries

Analysed YTHDF4-1 and YTHDF4-2 sample libraries contained 2381210 and 1389329 sequences centred around A and 2582834 and 1349826 sequences containing m^6A , respectively. This gives a 47.9% to 51% and 50.7% to 49.3% A:m^6A ration for YTHDF4-1 and YTHDF4-2, with an average of 49.3% to 50.7%. Once compared to the relative abundance of A vs m^6A in the input mix YTHDF3 shows an average enrichment of +1.04% for m^6A over A. Demonstrating YTHDF4 has very little preference of either A or m^6A over the other.

Following the trend observed in the other selected libraries the first base in the sequences in the context of A is more likely to be a C than A, G or T (+5.29%) (Figure 6.8- A/B). The is a very strong preference for C at position 2 (+32.6%) with reductions in A, G and T (-1.97%, -10.88% and -2.07%, respectively). The is a slight preference for G at position 3 (+1.34%). Adjacent to the A, there is a preference for C and G (+2.43% and +1.48%), but little difference between base representation at position 6. Similar to YTHDF3, position 7, 8 and 9 show a preference for C (+7.13%, +5.58% and 6.24%) with reductions in sequences containing A, G and T at these positions. Like the other three YTHDF proteins, in the context of A there is no suggestion of preference for any of the published consensus sequences.

The variation in base proportions found in YTHDF4 *in vitro* selection libraries differs in the context of m^6A (Figure 6.8 – C/D). The first base is more likely to be a pyrimidine, C or T (+14.69% and 6.90%) with a 9.51% reduction of sequences starting with a G. There is also a preference for C at position 2 (+7.02) with a reduction in sequences containing G or T at position 2 (-2.94% and -2.89%). C is preferred in positions 3 and 4 also (+5.72% and +3.65%, respectively). T is reduced in position 3 (-3.26%) and G is reduced in position 4 (-3.57%). At position 6, G, C and T are slightly preferred over A (0.87%, 1.69% and 0.21% vs -2.31%). G is enriched at position 7 (+5.77%) and A and C are reduced (-2.39% and -2.01%). Like the other three YTHDF proteins T is significantly enriched at position 8 (+20.89%) with a slight increase in C (+1.50%), meanwhile the proportion of sequences containing purines A or G are reduced (-4.57% and -6.49%). Finally, C is again enriched at the last position (+8.025) and all other bases are reduced in enrichment (-6.82%, -6.49% and -2.29%, for A, G and T, respectively). Overall, this shows a different consensus preference in the context of A vs m^6A. The preferred consensus is unclear.

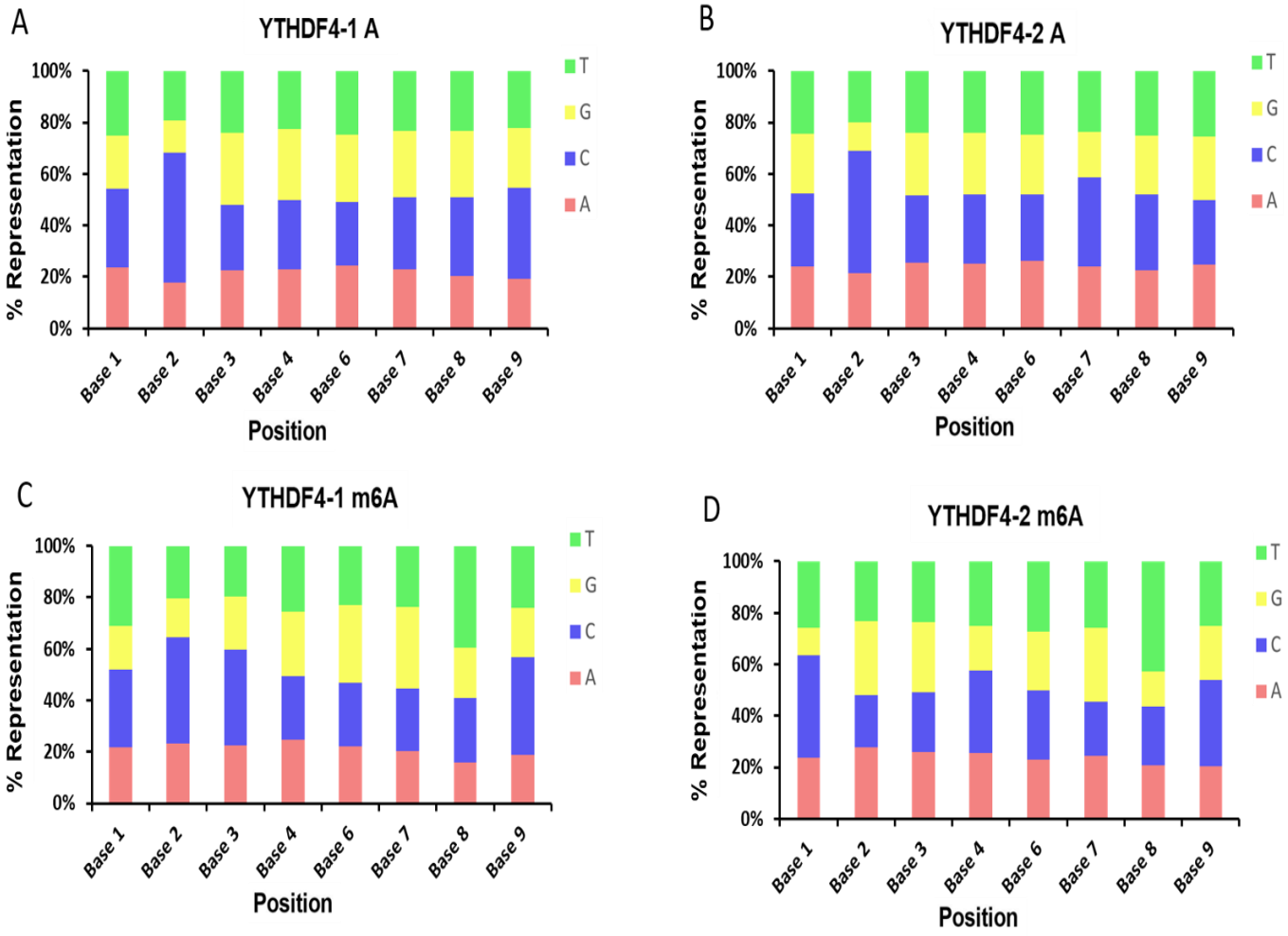


Figure 6.8: Characterisation of random sequencing libraries present in YTHDF4-1 and YTHDF4-2 *in vitro* selection libraries: (A) Percentage incorporation of each ribonucleotide at each of the 9 position in the 9-mer in sample YTHDF4-1 containing A at position 5. (B) Percentage incorporation of each ribonucleotide at each of the 9 position in the 9-mer in sample YTHDF4-2 containing A at position 5 (C) Percentage incorporation of each ribonucleotide at each of the 9 position in the 9-mer in sample YTHDF4-1 containing m^6A at position 5 (D) Percentage incorporation of each ribonucleotide at each of the 9 position in the 9-mer in sample YTHDF4-2 containing m^6A at position 5. All percentages normalised against input.

6.2.5.6 Relative abundance of sequences enriched upon *in vitro* selections

Next, we chose to focus on the 20 most enriched sequences and compare the found sequences for each YTHDF selection and to observe differences in sequences context for A (Table 6.1) and m^{6}A (Table 6.2).

For YTHDF1 the top 20 sequences by proportion of sequences centred around A comprise 0.29% of the *in vitro* selection library, the highest of any YTHDF protein. This is lower than the 0.54% comprised by the top 20 m^{6}A containing sequences. The higher proportion of total despite lower levels of m^{6}A in the total library is enlightening. This can be explained by higher sequence specificity in the context of m^{6}A, like what was observed in the base proportion analysis (Section 6.3.5.2). 3 of the top 4 A containing sequences are also shared with YTHDF2, YTHDF3 and YTHDF4. 5 YTHDF1 A containing sequences are also present in the top 20 of YTHDF2, AND 1 is shared with YTHDF3 and 1 with YTHDF4. The remaining 10 sequences are unique to the top 20 of YTHDF1. YTHDF1 appears to have a preference for A in the centre of the sequence but only in the context of A, several the sequences also contain CTT/CTTT 2 of the m^{6}A containing sequences are all also found in the top 20 of YTHDF2, YTHDF3 and YTHDF4. 5 sequences are shared with YTHDF2 and 2 are shared with both YTHDF3 and YTHDF4. Interestingly, in contrast to the A containing sequences, none of the top 20 sequences are shared with YTHDF2. Many of the top 20 YTHDF1 selected sequences by proportion compared to the input contain CTT at the 3' end with also ending in a C or a T. The preference for A observed in A-centred sequences are not observed here.

For YTHDF2 the top 20 sequences by enrichment centred around A comprise 0.18% of the total sequences vs 2.76% of m^{6}A-centred sequences. This reflects the higher proportion of the m^{6}A sequences in the total samples, however this still displays a higher specificity in the m^{6}A context. 3 of the 20 sequences are shared across all YTHDF proteins. 2 sequences are shared with YTHDF3/YTHDF4. 5 sequences are shared with YTHDF1 alone and 2 are shared with YTHDF3. None are shared with YTHDF4 alone. Similar to YTHDF1, many of the top 20 contain strings of A, suggesting an overall preference for A in the A-centred context. Two m^{6}A sequences are shared between all proteins. 1 sequence is shared between YTHDF2, YTHDF3 and YTHDF4. 2 of the top 3 sequences are shared with YTHDF3. 5 out the 20 sequences overlap with the other proteins, less than the 12

A sequences shared with the other YTHDF proteins. Overall, there is a more consistent consensus in the sequences starting DDDWAD with 19 of the 20 sequences ending in TGC.

The top 20 A-containing sequences in YTHDF3 libraries comprise 0.27% of the total vs 1.03% comprised by the top 20 m^6A containing sequences. The specificity of binding in the m^6A context again appears to be higher for YTHDF3. YTHDF3 A-centred sequences have a large degree of overlap with the other YTHDF A sequences. Other than the 3 global sequences found in this section, YTHDF3 shares 1 sequence with YTHDF1/YTHDF4 and one with YTHDF2/YTHDF4. 1 sequence is shared with YTHDF1, 2 with YTHDF2 and 7 out of the 20 is shared with YTHDF4. Overall nearly half of the YTHDF3 A sequences are found in YTHDF4 top 20. Only 5 of the sequences are unique to YTHDF3. The sequences tend to start and end with pyrimidine bases with a preference for A in the centre. In the m^6A sequences, unlike YTHDF1, YTHDF2 and YTHDF4, there is also a large degree of overlap in the m^6A containing sequences shared by YTHDF3. 2 are shared with YTHDF1/YTHDF4 and 2 are shared with YTHDF2/YTHDF4. 5 of the 20 are shared with YTHDF1, 2 with YTHDF2 and 5 with YTHDF4. With 3 YTHDF3 has the lowest number of unique m^6A sequences. This overlap can also be seen in the consensus sequences of YTHDF3 m^6A sequences, the m^6A centred sequences are also capped by pyrimidine bases.

The top 20 A-containing sequences in YTHDF4 libraries comprise 0.198% of the total vs 0.96% comprised by the top 20 m^6A containing sequences. This means all YTHDF protein show a higher level of specificity in the top 20 selected sequences. YTHDF4 selects for 3 A sequences shared between all proteins, 2 with YTHDF2/YTHDF3, 1 with YTHDF1 and 7 with YTHDF3. Leaving 7 unique A containing sequences. The top 2 m^6A sequences are shared with YTHDF1, YTHDF2 and YTHDF3. 2 more sequences are shared with YTHDF1/YTHDF3. And 1 with YTHDF2/YTHDF3. A s previously mentioned none of the sequences are shared with YTHDF1 or YTHDF2 alone, but a quarter of the 20 sequences are shared with YTHDF3. That leaves half of the 20 as unique to the top 20 of YTHDF4. The majority of the m^6A centred sequences selected by YTHDF4, start with TG and end with GC.

Overall, observing the sequences with the highest proportion enrichment when compared to the input can provide some interesting insights, but the biological context or relevance

of these preferences observed using this system is unknown. The top 20 sequences represent a rather small subset of the total library. It does show however that m^6A sequences have a higher enrichment to A sequences and generally a more consistent sequence consensus. Without analysis of more sequences and a comparison to m^6A-seq data, and RIP-seq with tagged YTHDF proteins, these results only represent very slight advances in our knowledge of these proteins. Furtheradvances will increase the relevance of this data.

Table 6.1: Relative abundance of A-centred 9-mers enriched via *in vitro* selection strategy for YTHDF1, YTHDF2, YTHDF3, and YTHDF4. Proportion of total sequences average across both repeats for each protein. Highlights indicate sequences shared between proteins.

YTHDF1	Proportion	YTHDF2	Proportion	YTHDF3	Proportion	YTHDF4	Proportion
CTTTAACGT	0.0244	CTTTAACGT	0.0122	CTTTAACGT	0.0184	TGATACGTC	0.0136
TCGTAAAAA	0.0212	CTTCACGCT	0.0116	TGACATGGC	0.0174	TGAAAATCT	0.0117
CGACAACGC	0.0180	ACGAACTTC	0.0107	GTAGAAGGT	0.0164	TGATATTGC	0.0105
CTTAATGCT	0.0166	CGTAAATCT	0.0106	TGGGAAGTC	0.0152	CTTTAACGT	0.0103
CTTTAGCGT	0.0163	TTAAAATCT	0.0104	TGATACGTC	0.0145	CGGTATTGC	0.0100
TTAAAATCT	0.0150	GTAAAATCT	0.0100	TGAAATTGC	0.0142	TGACACGTC	0.0098
GTAAAATCT	0.0149	TCGTAAAAA	0.0100	TGAAAATCT	0.0141	CAGTATGTC	0.0098
TAAAAATCT	0.0149	TTTAAATCT	0.0098	CTATAGGTC	0.0133	GAACATCCC	0.0097
AAGAAATAG	0.0141	ACGTATATC	0.0084	CTTAATGCT	0.0131	TCGTAAAAA	0.0097
CGAGACGTG	0.0130	TGAAAATCT	0.0084	TGAAACGTC	0.0129	TGATATGTC	0.0097
CTTCAGCGT	0.0128	TGAAATTGC	0.0084	TCAAAAACG	0.0130	CGAGACGTG	0.0096
CGTAAAATC	0.0127	CGTAAAATT	0.0082	TGAAAGGTC	0.0120	GTCTATGCC	0.0096
TCAAAAACG	0.0122	CTTAATGCT	0.0079	CTGAAGGTC	0.0119	TGAAACGTC	0.0095
TCGTAAATA	0.0121	CCATATGCG	0.0078	TCGTAAAAA	0.0119	GTAGAAGGT	0.0093
CGTAAATCT	0.0120	TGATATTGC	0.0078	TCGGATCGC	0.0118	TGGGAAGTC	0.0093
CTTTAGCCT	0.0117	CTTAAATCT	0.0078	TGATATGTC	0.0118	TTGTATGCT	0.0093
CTAAAAACG	0.0116	CGACAACGT	0.0076	ACGAACTTC	0.0117	CGGTAGCCT	0.0093
CTTTAAACG	0.0114	CGTAAAATC	0.0075	TGATATTGC	0.0116	CGCGAGTGC	0.0093
CGTAAAATT	0.0114	CGTAAAAAA	0.0075	TTGTATGCT	0.0114	CTTAATGCT	0.0092
TGATAAATC	0.0113	TGTTACTCC	0.0072	TCGAACTCT	0.0113	CTATAGGTC	0.0091
TOTAL	**0.2876**	**TOTAL**	**0.1798**	**TOTAL**	**0.2679**	**TOTAL**	**0.1983**

DF1, DF2, DF3 + DF4 | DF1, DF3 + DF4 | DF2, DF3 + DF4 | DF1 + DF2 | DF1 + DF3 | DF1 + DF4 | DF2 + DF3 | DF3 + DF4

Table 6.2: Relative abundance of m^{6}A-centred 9-mers enriched via *in vitro* selection strategy for YTHDF1, YTHDF2, YTHDF3, and YTHDF4. Proportion of total sequences average across both repeats for each protein. Highlights indicate sequences shared between proteins.

YTHDF1¤	Proportion¤	YTHDF2¤	Proportion¤	YTHDF3¤	Proportion¤	YTHDF4¤	Proportion¤
CTCTAACCC¤	0.0654¤	GAAAAGTGC¤	0.2188¤	TTAGAGTGC¤	0.0711¤	TGTTAGTGC¤	0.0815¤
CAGCACTTC¤	0.043¤	TGTTAGTGC¤	0.2062¤	CTCTAGCCC¤	0.0701¤	TTAGAGTGC¤	0.0653¤
CTCTAGCCC¤	0.0379¤	AGATAATGC¤	0.1784¤	CTCTAACCC¤	0.0701¤	TGCTATTGC¤	0.0644¤
CTTAACTCT¤	0.0356¤	AAGTATTGC¤	0.1524¤	CAGCACTTC¤	0.0611¤	TGATATTGC¤	0.0633¤
CTTTACTCT¤	0.0287¤	AGATAGTGC¤	0.1524¤	TGCTATTGC¤	0.0581¤	TGATATGGC¤	0.0579¤
CTTTAGGCT¤	0.0262¤	TGATATTGC¤	0.1486¤	CAGAAAGGC¤	0.0566¤	TGAAATGGC¤	0.0513¤
TAAAATTGC¤	0.0243¤	GATAAGTGC¤	0.1385¤	TGTTAGTGC¤	0.0557¤	TGAAATTGC¤	0.051¤
TTAGAGTGC¤	0.0241¤	AACTAGTGC¤	0.1379¤	CTTAACTCT¤	0.0547¤	TGAGAGTGC¤	0.0482¤
CTCGACACC¤	0.0237¤	GAAAATTGC¤	0.1364¤	AGATAAGGC¤	0.0508¤	TGATAAGGC¤	0.0478¤
CTCTAGTGC¤	0.0233¤	AGATATTGC¤	0.1327¤	CTTTAGGCT¤	0.0488¤	TGAAACGTC¤	0.0418¤
CTTGACTCT¤	0.0225¤	TAGTATTGC¤	0.1225¤	CTCTATGCC¤	0.0484¤	CTTAAGTGC¤	0.0412¤
CTTTAGCCT¤	0.0225¤	GGAGAAAGT¤	0.1199¤	TATGATGGC¤	0.0467¤	TCCGATGGC¤	0.0407¤
CTCTAGGCC¤	0.0216¤	GTAAAGTGC¤	0.1187¤	TGATATGGC¤	0.0462¤	TGATAGTGC¤	0.0407¤
CTCTATGCC¤	0.0215¤	AGAAAGTGC¤	0.1169¤	TGATATTGC¤	0.045¤	TATGATGGC¤	0.0396¤
TGTTAGTGC¤	0.0208¤	GACTAGTGC¤	0.1151¤	CTTTAGTGC¤	0.0423¤	TGATAGGGC¤	0.0393¤
CTTTACAGC¤	0.0202¤	TTAGAGTGC¤	0.1146¤	GAAAAGTGC¤	0.0423¤	CAGAAAGGC¤	0.0389¤
CTTAAGGCT¤	0.0202¤	GAATATTGC¤	0.1144¤	CCCTAAGGC¤	0.0423¤	CTCTAGCCC¤	0.0389¤
CTTCAGGCT¤	0.0202¤	AAATAGTGC¤	0.1139¤	TAAAATTGC¤	0.0409¤	TGAGATGGC¤	0.0377¤
CTCTATCTC¤	0.0200¤	AGACAGTGC¤	0.113¤	AGATAATGC¤	0.0405¤	CAGCACTTC¤	0.0375¤
CTCAAGGCC¤	0.0188¤	GTAAATTGC¤	0.1071¤	TATCATGGC¤	0.0403¤	AGATAAGGC¤	0.0364¤
TOTAL¤	**0.5405¤**	**TOTAL¤**	**2.7584¤**	**TOTAL¤**	**1.032¤**	**TOTAL¤**	**0.9634¤**

DF1, DF2, DF3 + DF4 | DF1, DF3 + DF4 | DF2, DF3 + DF4 | DF1 + DF2 | DF1 + DF3 | DF1 + DF4 | DF2 + DF3 | DF3 + DF4

6.3 Discussion

In this chapter we have developed and modified an *in vitro* selection strategy for the comparison of biding affinities to RNA oligonucleotides of RNA binding proteins. The strategy we used here is similar to that demonstrated by the Kleiner lab in 2019 (Arguello, Leach and Kleiner, 2019), however the idea and development of our protocol was independent to this paper. Despite this there are several techniques demonstrated in the Kleiner paper, that could inform improvements to our process.

One of the major limitations during this process was the random RNA library itself. The incorporation of random nucleotides at different positions was found to be uneven (Figure 6.4 – A/C). This could potentially introduce a bias into the subsequent libraries. It also increases the complexity of the analysis with all counts and sequences normalised against the input (random oligonucleotide pre-*in vitro* selection). Previous studies have utilised custom phosphoramidite mixes to achieve equal incorporations of A/C/G/U. Bartel and Colleagues suggested (A/C/G/T = 0.26:0.25:0.29:0.20) to produce random DNA sequences (Unrau and Bartel, 1998). However, it has since been suggested that for RNA oligonucleotide libraries this ratio is not sufficient to achieve equal incorporation (Arguello, Leach and Kleiner, 2019). RNA phosphoramidites likely have different relative reactivities to the DNA monomers used in that study. We utilised a ratio of A/C/G/U = 0.30:0.25:0.23:0.20. Adjusting this ratio could yield a more balanced random library for our applications. A further limitation which, could have potentially compounded the bias in the random oligonucleotide library is the lack of an additional input library for reference. All analysis is based on a single library with no repeats. Future work would involve an additional input to aid in the analysis of the results.

Interactions between m^6A-containing mRNA and 'reader' proteins such as YTHDF1-4 invoke a diverse range of functions including stress response, nutrient uptake, leaf and trichome morphogenesis in plants and neuronal development, cell differentiation and sex determination in animals (Ok *et*

al., 2005; Haussmann *et al.*, 2016; Aguilo and Walsh, 2017; Li *et al.*, 2017; Phillip J. Hsu *et al.*, 2017; Arribas-Hernández *et al.*, 2018). The mechanism controlling the localisation, recognition and subsequent function of the YTH-domain proteins and their interaction with m^{6}A is essential to these functions.

Our study has demonstrated that YTHDF1, YTHDF2, YTHDF3 and YTHDF4 in tomato have different affinities to consensus sequences in an A or an m^{6}A context. We had hoped that unbiased application of equal amounts of A and m^{6}A sequences to each YTHDF protein would clearly demonstrate a stronger enrichment of m^{6}A over A containing sequences. While this was proven to be true in the case of YTHDF2, with significantly more m^{6}A-centred sequences present in the selection libraries, without further analysis this cannot be definitively proven for YTHDF1, YTHDF3 or YTHDF4. In fact, our analysis shows that it is more likely that YTHDF1 has a preference for A over m^{6}A. Of course, biological relevance of the enrichment of A cannot be ruled out for any of the YTHDF proteins and may yet prove crucial to their function *in vivo*. Past studies on plant YTHDF proteins have observed post-translational phosphorylation of tomato YTHDF1 homolog, AtECT2. It is likely that understanding of post-translational modification of YTH proteins, in addition to their binding affinities will be essential to fully understanding their role in the mediation of m^{6}A.

Analysis of the proportion of each base at each position of the 9-mer also highlighted the difference in specificity in an A context to that in an m^{6}A context. Different bases increase or decrease in enrichment for each protein. It must be said that the analysis carried out was performed on the libraries. Results may differ if focussed on a top subset of each *in vitro* selection. In fact, brief analysis of the top 20 sequences enriched in our sample libraries show that these sequences do not follow the overall suggestions found by our base incorporation analysis. It may provide more insight if this were to be repeated on a smaller and more selective subset, perhaps more representative of the preferred sequences.

Comparing the top 20 sequences containing both A and m^{6}A that were enriched relative to the input did yield some interesting results. Firstly,

there was some overlap between the top sequences for each of the four proteins. Interestingly, there was more overlap between proteins closer in overall similarity in the context of A than in the sequences containing m^{6}A (Table 6.1, 6.2). For example, there are 5 shared between YTHDF1 and YTHDF2 in the A set, but none shared in the m^{6}A top 20 by proportion. Also of note, is the high level of overlap observed in the YTHDF3 selected m^{6}A sequences, with 3 unique to YTHDF3 vs 11, 15 and 10 unique sequences in YTHDF1, YTHDF2 and YTHDF4, respectively. There is no suggestion however that any of the YTHDF proteins in tomato have any strong preference for the canonical motifs previously identified in other species (DR(m^{6}A)CH, URU(m^{6}A)Y, RR(m^{6}A)CH) (Bodi *et al.*, 2012a; Xu *et al.*, 2015; L.-H. H. Wei *et al.*, 2018; Woodcock *et al.*, 2020) in either context. Interestingly, similar work also came to this conclusion for YTHDF proteins in mammalian systems, despite strong evidence that YTHDC proteins exhibit these preferences (Arguello, Leach and Kleiner, 2019). Perhaps the remaining consensus sequences preferred by YTHDF proteins remain to be discovered, or conversely no specific preferences exist.

This method represents a potentially unbiased view of binding selectivity of RNA-binding proteins. All the analysis here is from results *in vitro* and with protein synthesised in a bacterial system. It is likely that only a subset of the enriched sequences exists *in vivo,* with or without m^{6}A, and all 'reader' protein activity is altered by the binding preference of other components of the m^{6}A mechanism ('Writers' and 'Erasers'). However, our findings here can be a useful resource in future work. Expanded data on the tomato methylomes and transcriptomes will only increase the relevance of this data. It is also possible that the utilisation of a dual modified/unmodified RNA oligonucleotide population could be advantageous for additional RNA binding proteins or RNA modifications.

Chapter 7 – *General Discussion*

7.1 General Discussion

As emphasised repeatedly throughout this work, YTH proteins within *Plantae* are diverse and highly expanded. The large number of these protein, in particular the YTHDF group, and the high level of conservation within each species, increases the potential for redundancy and overlapping functions of each protein (Berlivet *et al.*, 2019). There is large variation in the number of YTHDF and YTHDC type proteins. We set out to identify a suitable plant model for the comprehensive analysis of the YTHDF family of RNA-binding proteins. Plant based studies of the YTH family, and their biological functions is limited, any advance in our understanding would be of great value to future studies.

In this thesis, we collected and evaluated all YTH proteins found within a number of key model plant species, in order to identify a more ideal model for the study of YTHDF type proteins. Bioinformatic analysis of 286 YTH proteins identified an average of 1.93 YTHDC proteins and 7.86 YTHDF proteins (Table 3.1). From this analysis we identified the potential of Tomato (*Solanum lycopersicum)* as an ideal model species for the study of the predominantly cytoplasmic YTHDF protein group. With 4 YTHDF proteins, tomato represents a significant improvement in the potential efficiency of forward genetic studies of YTHDF biological roles and functions, then that seen in other key model species, such as, *Arabidopsis thaliana* (11- YTHDF) and *Oryza sativa* (12- YTHDF).

If we are successful in establishing this model for m^{6}A reader research, the potential for furthering our knowledge of m^{6}A and the mechanisms of their role could be high. Tomato's current role as a key plant model for fruit biology, and its ranking as the fourth largest crop in terms of global value, add additional significance to any future studies.

The YTHDC group in tomato seems to be unique among plant species in having more members than the YTHDF group. In addition, it is the only

species observed in our study to have a triplicated YTHDC2 gene. Although, expression has not been quantified and it is possible only one or two of these genes are functionally expressed.

It has been noted by numerous studies, the YTH domain is essential to the protein m^{6}A binding ability, but it is the remainder of the protein which mediates its localisation, function and structure (Zhu *et al.*, 2014; Arribas-Hernández *et al.*, 2018; Scutenaire, J. M. Deragon, *et al.*, 2018; Berlivet *et al.*, 2019). We therefore hypothesised that the regions outside of the YTH domain, mediate specific functions of the YTH proteins, while the YTH domain is essential for the m^{6}A affinity (D. Li *et al.*, 2014; S. Luo and Tong, 2014; Theler *et al.*, 2014; Xu, Liu, *et al.*, 2014; Zhu *et al.*, 2014; Arribas-Hernández *et al.*, 2021)

We aligned the YTHDF group and generated a tree of similarity predict the closest homologs of each tomato YTHDF protein. The high level of conservation of YTH domains across Eukaryota has led many to speculate the functional significance of the protein residues and sequences outside of the YTH domains (S. Luo and Tong, 2014; L.-H. H. Wei *et al.*, 2018). Previous attempts to characterise plant YTH domain proteins and bioinfomatically predict their functionality have been based on sequence similarities of the YTH domain alone (D. Li *et al.*, 2014; L.-H. H. Wei *et al.*, 2018; Scutenaire, J.-M. Deragon, *et al.*, 2018). Our analysis attempted to provides a more comprehensive, therefore potentially more representative of the evolutionary roles *in vivo*. As previously mentioned the analysis was not of high enough statistical significance or clarity to yield any further prediction of plant YTHDF functional groups. It is clear that the high level of disorder observed outside of the YTH domains in plant species make this type of analysis difficult.

We analysed the absolute conserved residues found in Tomato YTH domains, indicate the YTHDF YTH domains likely rely on the WWW aromatic cage. This follows previous work showing the WWW residues are common to almost all YTHDF proteins (F. Li *et al.*, 2014). Interestingly, YTHDC proteins SlYTHDC2a and DC2b also have the WWW residues, which is different to the L present in Human YTHDC1 and YTHDC2 (Xu *et al.*, 2015). Even more surprisingly, SlCPSF30, YTHDC3 and YTHDC2c are

predicted to utilise the Y, similar to the aromatic cage in Yeast m^{6}A protein PHO92 (Figure 3.2) (Kang *et al.*, 2014). This difference in this highly important 3rd residue in the triplicated YTHDC2a, b and c proteins has never been reported in the literature and this, coupled with the differences in intron lengths and distribution could suggest that the proteins have diverse functions *in vivo*. Any confirmation of functional importance of these residues would confirm that any phenotypes observed, require the protein's ability to bind m^{6}A.

In mammalian cell lines, m^{6}A is additionally added to multiple translationally active mRNAs under stress (oxidative and heat). This induces 'triaging' to foci known as stress granules (SGs). The formation of these granules was later shown to be reliant on all three YTHDF proteins (DF1-3). Similar induction of foci has been observed in *Arabidopsis,* and similarly linked to IDR of YTH protein N-termini (Arribas-Hernández *et al.*, 2018). We have discovered that all four YTHDF proteins in tomato have IDRs but only YTHDF1 and YTHDF2 have been predicted to form assemblages *in vivo* (FIGURE 3.7). The importance of these assemblages for numerous biological processes are likely shared by tomato (and *Arabidopsis*) systems.

Overall, these initial observations demonstrate the potential of *Solanum lycopersicum* as a novel and highly effective model for the study of YT521-B proteins. Tomato YTH proteins share many of the characteristics which make YTH proteins such an exciting family to study. These characteristics also can make the study challenging, with such a vast range of functions *in vivo*. It's this level of complexity which leads us to choose tomato to uncover these roles. We hypothesise that the ythdf gene family will have a lower level of genetic redundancy than observed in *Arabidopsis* and therefore provide an opportunity to further advance our understanding of this highly interesting and complex protein family.

In order to effectively study the four YTHDF proteins in tomato, we developed a CRISPR/Cas9 gene disruption system for all four YTHDF genes in *Solanum lycopersum*. Although the gene disruption system we utilised for this study had been successfully used previously (Weber *et al.*, 2011; Wu *et al.*, 2018), the overall efficiency of stable mutation generation was low.

Unfortunately, only a single *ythdf1* null mutant line was generated. Advances since the advent of CRISPR/Cas9 have been significant. To date up to six genes in tomato have been targeted using a single system, inducing mutations in four genes simultaneously (Zsögön *et al.*, 2018). The original Cas9 derived from *Streptococcus pyogenes*, has a relatively high level of off target effects, Cas9 enzymes engineered from other species (E.g., *N. meningitides* and *Streptococcus thermophiles*) with different length PAM sites, have been proven to have reduced off-target effects and higher specificity (Lee, Cradick and Bao, 2016; Hu *et al.*, 2018). No off-target effects were predicted using the designed sgRNA target sequences in this study, but any future utilisation of CRISPR/Cas9 technology would benefit from these advances. It is likely that utilisation of these advances would have yielded more agreeable results. Any future attempts to generate a second *ythdf1* line will use an alternative system.

Phenotypic differences between our generated single mutant lines and wild-type tomato were used to suggest the functional roles of the YTHDF protein *in vivo. Ythdf1, ythdf2* and *ythdf4* plants all exhibited significant reductions in vertical growth, shorter leaves and significantly reduced leaf widths. These measurements indicate a significant reduction in vegetative growth rates in the absence of YTHDF1, YTHDF2 and YTHDF4. Interestingly the converse is seen in the absence of YTHDF3. *Ythdf3* lines display a slight increase of vertical height, significantly longer leaves and an increase in leaf widths (Figure 4.3, 4.4). These findings suggest that the correct timing of vegetative tissues in tomato require the function of all four YTHDF proteins. Perhaps this is no surprise due to recent findings by three concurrent studies into ECT proteins within *Arabidopsis,* which demonstrated that YTHDF1 homologs AtECT2 and AtECT3 are involved in correct timing of emergence and morphological development of leaves (Arribas-Hernández *et al.*, 2018; L.-H. H. Wei *et al.*, 2018; Scutenaire, J.-M. Deragon, *et al.*, 2018). It is worth noting that single mutants of single null mutants of AtECT2, AtECT3 and AtECT4 do not show any obvious phenotypes, further supporting tomato as an effective model.

Root length was reduced in both *ythdf1* and *ythdf2,* with no significant differences in *ythdf3* or *ythdf4* root lengths (Figure 4.3, 4.4). Hypomorphic

mutants resulting from low m^{6}A levels in *Arabidopsis* exhibit defective gravitropic responses and development of vasculature and lateral root branching. Future phenotyping of the single mutants could take a closer look into gravitropic responses, hormone responses, root vasculature and lateral branching. Observable differences could highlight if any m^{6}A functions in root development are mediated by YTHDF proteins in tomato.

Loss of functional YTHDF2, YTHDF3 and YTHDF4 proteins, results in a significant increase in the number of nodes on each plant. This in combination with the significantly reduces vertical height of *ythdf1, ythdf2* and *ythdf4,* results in a reduction in internode lengths for *ythdf1, ythdf2* and *ythdf4*. Internode lengths has been used as a key indicator for phenotypic changes and is particularly sensitive to abiotic stresses including low and high temperature, drought and responses to nitrogen. *Ythdf3* plants do not exhibit any dwarf phenotypes. *Ythdf1, ythdf2* and *ythdf4* on the other hand all show traits associated with dwarfism (Figure 4.3, 4.4). The contrasting phenotypes of *ythdf3 (*increased leaf expansion and internode elongation) to *ythdf1, ythdf2* and *ythdf4*; also hint at a connection between m^{6}A and GA. The endogenous phyto-hormone gibberellin (GA) is heavily involved in plant growth and development, influencing seed germination, stem elongation and the induction of flowering to name a few. They are also heavily involved in the mediation of responses to environmental stimuli, through the GA-biosynthetic pathway (Hedden and Kamiya, 1997). The phenotypes observed in *ythdf1, ythdf2* and *ythdf4* all mirror phenotypes associated with GA insensitivity mutants to varying degrees, with loss resulting in both restricted internode elongation and restricted leaf expansion (Greco *et al.*, 2012; Yoshida, Mogami and Yamaguchi-Shinozaki, 2014). To date connections between m^{6}A/YTH proteins and suggested hormone related phenotypes, have been limited to low m^{6}A lines in *Arabidopsis* (Bodi *et al.*, 2012a). Perhaps these responses are mediated by YTHDF proteins *in-vivo.* Many questions remain on how the YTHDF protein would mediate these functions and how they interact between one another.

Loss of YTHDF1, YTHDF2 and YTHDF4 results in an increase in the number activated auxiliary meristems observed at 4 weeks. This indicates that these proteins have some role in the maintenance of the shoot apical

meristems (SAM). The inhibitory effect the SAM has on the auxiliary meristems would normally prevent these meristems from activating at 4 weeks. Loss of apical dominance is also associated with low methylation lines in *Arabidopsis*. Bodi et al discovered that transcripts of key transcription factors; WUSCHEL and SHOOT MERISTEMLESS, are significantly methylated under normal conditions (Bodi *et al.*, 2012a). Knockout of FIP37, a key component of the plant methylation complex, results in over-proliferation of shoot meristems. Later, this was shown to be due to an increased stability of these transcripts in the absence of m^6A (Shen *et al.*, 2016). This could suggest that in plants the YTHDF proteins 'read' the m^6A marks, deposited on these transcripts. Therefore, lack of YTHDF1, YTHDF2 and YTHDF4 has a similar effect to a reduction in levels of m^6A. Whether the similar phenotypes demonstrated by *ythdf1, ythdf2* and *ythdf4* plants are due to an overlap of mRNA target or different mechanisms altogether is the next big question. Outside of plants, m^6A has a role in maintenance of stem cell differentiation, with confirmed links between methylation complex components and pluripotency genes, developmental regulators and stem cell RNA stability (Batista *et al.*, 2014; Ke *et al.*, 2017; Zhang *et al.*, 2020). This work represents significant evidence that YTHDF proteins have important roles in stem cell and SAM maintenance in plants. To what extent these proteins are involved in these essential processes in unknown.

We have demonstrated that plants lacking YTHDF proteins exhibit a more dramatic phenotypic response to high salinity than WT plants. Conversely to normal conditions *ythdf3* and *ythdf1* plants show significantly reduced growth rate, while *ythdf2* and *ythdf4* show increased growth rate (Figure 4.7). This could suggest contrasting role of YTHDF1/YTHDF3 and YTHDF2/4 in response to high salt concentrations. These results are hard to confirm due to measuring responses of plants which already display aberrant phenotypes under normal conditions. However, they are supported by no significant differences in growth observed in WT plants. The first study into YTH domain proteins in plants, discovered that YTH genes are induced by abiotic and biotic stress conditions. Recent studies into the

involvement of m^{6}A in the response to salt stress found that key salt stress transcripts are selectively stabilised via hypermethylation of m^{6}A under high salinity (Anderson *et al.*, 2018). Additionally, AtECT3 and AtECT4 were shown to relocalise into processing bodies and AtECT2 into processing bodies and stress granules under heat stress in *Arabidopsis* (Arribas-Hernández *et al.*, 2018; Scutenaire, J. M. Deragon, *et al.*, 2018). This work provides further evidence to add to the growing list of YTHDF functions in abiotic stress responses. Utilisation of this model could yet yield more concrete evidence to increase our understanding of YTHDF protein mediation of abiotic and biotic stress responses *in vivo*.

No differences were identified in m^{6}A levels in 2 weeks old *ythdf1* or *ythdf2* plants, this shows that YTHDF1 and YTHDF2 do not regulate m^{6}A levels directly, but no further conclusions could be drawn from this result. It is still possible that YTHDF proteins has functions in RNA stability, biogenesis and degradation, as these studies involving YTHDF proteins did not observed altered global levels of m^{6}A in mammals (Zhu *et al.*, 2014; K. Liu *et al.*, 2015; Xu *et al.*, 2015; Shi *et al.*, 2017). The basal A:m^{6}A ratio in 2-week-old seedlings was found to be 1.85%. This is the first example of whole tomato m^{6}A measurements. This differs from levels of 0.5% and 0.8% in 17- and 52-days post anthesis published by Zhou et al., in 2019 (Zhou, Tian and Qin, 2019). This demonstrates that tomato tissues exhibit a range of m^{6}A levels between different tissues and throughout development. This has also been demonstrated in *Arabiopsis* (Bodi *et al.*, 2012a). It is clear, further work is required to understand these differences in tomato.

After phenotyping of the single mutants, we decided to evaluate how concurrent Ko of two YTHDF proteins, further influenced observable phenotypes. We successfully generated stable lines for *ythdf1/ythdf3, ythdf2/ythdf3* and *ythdf3/ythdf4*. This is the first published work to demonstrate double YTHDF protein mutants in any plant species outside of *Arabidopsis*. It is also the first published work to confirm double mutants outside of the first predicted functional group (Scutenaire, J. M. Deragon, *et al.*, 2018). Unfortunately, we were unable to generate double mutants for

ythdf1/ythdf2 or *ythdf2/ythdf4*. This could suggest that concurrent KO of these genes is lethal in tomato but due the Covid-19 pandemic we were unable to test this hypothesis. We were able to successfully identify *ythdf1--/ythdf2*+- and *ythdf1-+/ythdf2*—plants, in the case of *ythdf1/ythdf2* and *ythdf2--/ythdf4*+- and *ythdf2-+/ythdf4*—plants, in the case of *ythdf2/ythdf4* (Figure 5.2; 5.3). This will aid in further attempts to generate double mutants of these ythdf genes or provide a platform to confirm their inviability.

We were unable to produce a single *ythdf1/ythdf4* homozygous double mutant line. We successfully generated stable *ythdf1--/ythdf4*+- and *ythdf1-+/ythdf4*-- plants. Both *ythdf1--/ythdf4*+- and *ythdf1-+/ythdf4*—plants were selfed top produce the F3 generation. These populations were used to confirm that double homozygous mutants were not possible to identify using normal screening and germination methods. As previously discussed, observations (Figure 4.5), identified significantly reduced germination rates in *ythdf1* and *ythdf4* lines, when compared to WT, *ythdf2* and *ythdf3* single mutant seeds. Discussions on the potential causes of this reduction in germination efficiency led to a hypothesis that *ythdf1* and *ythdf4* aberrant phenotypes may be due to roles of YTHDF1 and YTHDF4 in the gibberellic acid biosynthesis pathway. Indeed, both *ythdf1--/ythdf4*+- and *ythdf1*+-*/ythdf4*-- lines showed reduced germination efficiency when compared with WT seeds of a similar age. Treatment with gibberellic acid did increase the germination efficiency of both of these lines. The absence of *ythdf1/ythdf4* plants suggest that lack of YTHDF1 and YTHDF4, simultaneously, cause the seeds to abort in the varies during fruit development, and thus is lethal. If both YTHDF1 and YTHDF4 are essential for correct embryo and seed development, we would expect to see expression of these genes during embryo development. Proof of this would aid in confirming our hypothesis. This also suggests that YTHDF2 and YTHDF3 do not have crucial roles in embryo development.

KO of ythdf1 and ythdf3 resulted in an increase in vegetative growth when compared to *ythdf1* plants. This is only a partial rescue of *ythdf1* phenotypes Figure 5.6, 5.7). This is further evidence that YTHDF3 may have a role in correct control of vegetative growth. The number of auxiliary meristems

increases in *ythdf1/ythdf3* when compared to both *ythdf1* and *ythdf3* single mutant plants. This does suggest that YTHDF1 and YTHDF3 protein have functional redundancy in the regulation of meristematic function. *Ythdf2/ythdf3* double mutants show a significant increase in vertical growth, leaf width and length when compared to *ythdf2* single mutants (Figure 5.8, 5.9). Again, lack of YTHDF3 partially rescuing the phenotypes caused by removing YTHDF2. This increase in height increases the average internode length in *ythdf2/ythdf3* plants. The number of lateral branches observed in *ythdf2/ythdf3* plants does not differ significantly from *ythdf2* but is higher than *ythdf3*. Unlike, *ythdf1/ythdf3* this could suggest that YTHDF2 and YTHDF3 have independent roles in the regulation of apical dominance. Overall, there appears to be little genetic redundancy between ythdf2 and ythdf3. Lack of YTHDF3 and YTHDF4, contrasts with *ythdf1/ythdf3* and *ythdf2/3* plants. No increase in vertical growth is observed when compared to *ythdf4* single mutant plants and leaf growth does not differ significantly from *ythdf3* or *ythdf4* (Figure 5.10, 5.11). This could suggest that the visible increase of vegetative growth observed in *ythdf3* plants could be mediated by active YTHDF4. Investigations into the expression patterns of the YTHDF proteins, could suggest further avenues for study. There are also no observable changes in the number of nodes or the internode length with the simultaneous KO of ythdf3 and ythdf4. However, we do see an increase in lateral branching in *ythdf3/ythdf4* compared to *ythdf3, ythdf4* and WT. Again, this suggests that there is some degree of functional redundancy between YTHDF3 and YTHDF4 proteins. This adds to the complex interactions demonstrated by the four YTHDF proteins and meristematic regulation.

Overall, the four YTHDF proteins display less redundancy between them than displayed in *Arabidopsis*. In *Arabidopsis*, KO of ect2 and ect3 was necessary for observable phenotypic changes (Arribas-Hernández *et al.*, 2018; Scutenaire, J.-M. Deragon, *et al.*, 2018). Additional KO of Atect4 only causes minor aberrant phenotypes. We believe that this alone is clear evidence of the potential efficacy of our model.

All YTHDF functions studied to date appear to be as a result of regulation of key mRNAs. AtECT2 (homolog of SlYTHDF1) regulates mRNA levels of

proteosome regulator PTRE1 and other 20S proteosome units, at different developmental stages. Here, overexpression of ect2 increases the stability of PTRE1 transcripts, while *ect2* null mutants have reduced levels of PTRE1 mRNAs, although these finding were somewhat inconsistent. (Wu, Peled-Zehavi and Galili, 2020). KO of ect2 was also shown to decrease the stability of transcripts related to trichome morphology (*TTG1*, *ITB1* and *DIS2*), in an m^{6}A dependent manner (L.-H. H. Wei *et al.*, 2018). Both show a role of YTHDF1 homolog, AtECT2 in stabilising mRNA contrasting with the proposed function of Human YTHDF2 in mRNA destabilisation (Wang *et al.*, 2013; Du *et al.*, 2016). It would therefore be possible that the phenotypes associated with single and double KO of ythdf genes we have demonstrated in tomato, are a result of changes in specific transcript levels *in vivo.* Many of the YTHDF proteins investigated bind to a subset of sequenced m^{6}A containing transcripts, observed in HsYTHDF1 (Xu *et al.*, 2015) HsYTHDF2 (Du *et al.*, 2016), AtECT2 and AtECT3 (L.-H. H. Wei *et al.*, 2018; Arribas-Hernández *et al.*, 2021), HsYTHDC1 (Xu, Wang, *et al.*, 2014). If the structure of the YTH proteins outside of the YTH domain confer different functions, any overlap in the pools of mRNA each protein binds could explain the complex difference we observe here.

It has long been suggested that m^{6}A is co-transcriptionally deposited at DRACH (D= A/G/U, R=A/G, H=A/C/U) consensus sites by the methylation complex (Shimba *et al.*, 1995; Bokar *et al.*, 1997; Zhong *et al.*, 2008; Schwartz *et al.*, 2013; Liu *et al.*, 2014). This has since been shown in plants and fungi (Schwartz, Maxwell R Mumbach, *et al.*, 2014; S. Luo and Tong, 2014; Wan *et al.*, 2015; Duan *et al.*, 2017; Anderson *et al.*, 2018). There is some debate around the consensus sequence and its context for YTH proteins in plants, with contrasting binding sequences for AtECT2 of RR(m^{6}A)CH (Arribas-Hernández and Brodersen, 2020a) vs plant specific motif URU(m^{6}A)Y (L.-H. H. Wei *et al.*, 2018). While there seems to be an emerging role for pyrimidine-rich motifs in m^{6}A reading and writing, demonstrated by consistent conservation across plants (Y. Li *et al.*, 2014; Zhou *et al.*, 2018; Luo *et al.*, 2020; Miao *et al.*, 2020b), currently functions of these motifs are still unknown. What is clear from these articles is that YTH proteins in plants are likely to have a range of binding sites in an m^{6}A context, the

binding sites on which each YTH protein acts could confer its subsequent functions.

If we could identify which consensus sequences each of the proteins have affinity for, we may gain a key resource for the suggestion of YTHDF functions. In this work, we developed and modified a strategy for *in vitro* selection of RNA binding proteins. The strategy we used here is like that demonstrated by the Kleiner lab in 2019 (Arguello, Leach and Kleiner, 2019), however the idea and development of our protocol was independent to this paper. Despite this there are several techniques demonstrated in the Kleiner paper, that could inform improvements to our process. As previously mentioned, the random RNA library used throughout our study did not have equal incorporation of each base at each position (Figure 6.4 – A/C). Our libraries were based on work by Unrau and Bartel, producing random DNA sequences (Unrau and Bartel, 1998). It appears that the ratio of each phosphoramidite used to produce our random RNA libraries were not correct. Adjustments to our ratio of 0.30:0.25:0.23:0.20 (A/C/G/U) for future utilisation of this method, could yield more even distributions of each base.

We hypothesised that unbiased application of equal amounts of A and m^{6}A sequences to each YTHDF protein would clearly demonstrate a stronger enrichment of m^{6}A over A containing sequences. While this was proven to be true in the case of YTHDF2, with significantly more m^{6}A-centred sequences present in the selection libraries, without further repeats this cannot be proven for YTHDF1, YTHDF3 or YTHDF4. In fact, our analysis shows that it is more likely that YTHDF1 has a preference for A over m^{6}A. This result does not suggest that YTHDF1 is in fact not an m^{6}A binding protein. It remains a possibility that the YTHDF proteins in this study have roles outside of any ability to mediate the functions of the m^{6}A mark. These results do suggest that further repeats, addressing the limitations of the method we developed are indeed necessary. We compared the top 20 sequences represented by the RNA sequence selection. The sequences returned by their affinity to the YTHDF proteins, did exhibit some overlap. More overlap was observed in proteins which are more similar in sequence (E.g., YTHDF1 and YTHDF2). YTHDF3 exhibited the greatest degree of overlap with the other 3 proteins (Table 6.1, 6.2). There is no suggestion

however that any of the YTHDF proteins in tomato have any strong preference for the canonical motifs identified in other species (DR(m^{6}A)CH, URU(m^{6}A)Y or RR(m^{6}A)CH) (Bodi *et al.*, 2012a; Xu *et al.*, 2015; L.-H. H. Wei *et al.*, 2018; Woodcock *et al.*, 2020) in either an A or m^{6}A context. Again, further investigations into these initial suggestions would be necessary to conclusively confirm any observations. For now, it does suggest that there may be a greater degree of variation to the consensus sequences identified by m^{6}A sequence analysis. Obviously, we have only discussed the top 20 sequences out of thousands of potentially relevant motifs. It is likely that only a subset of the enriched sequences exists *in vivo,* with or without m^{6}A, and all 'reader' protein activity is altered by the binding preference of other components of the m^{6}A mechanism ('Writers' and 'Erasers'). Expanded data on the tomato epitranscriptomic methylomes and transcriptomes will only increase the potential usefulness of this or similar methods. We will continue to search for these answers.

7.2 Future Work

We have developed a platform that can be used as a valuable tool for future investigations into plant YTHDF proteins. The majority of studies into these proteins, and m^{6}A in plants have been performed in *Arabidopsis*. *Arabidopsis* has the clear advantage of established low methylation lines lines (Zhong *et al.*, 2008; Bodi *et al.*, 2012a; Růžička *et al.*, 2017). The phenotypes associated with low methylation has can be utilised as a background to aid in the comparison of any m^{6}A related studies, outside of the methylation complex, including YTH proteins. Development of low methylation tomato lines and robust phenotyping is essential for us to gain a fuller understanding of our model. This is especially important due to the overlap in aberrant phenotypes identified in this, and other studies (Arribas-Hernández *et al.*, 2018; Scutenaire, J.-M. Deragon, *et al.*, 2018). To this end, RNAi and CRISPR/Cas9 constructs have been designed and constructed to target key components of the methylation complex in tomato.

There are several additional limitations directly related to the work laid out in this thesis. These include a lack of additional *ythdf1* lines, as well as incomplete screening of *ythdf1/ythdf2* and *ythdf2/ythdf4* double mutant lines. These need to be addressed before we can have complete confidence in our findings. As previously mentioned, we have designed antibodies (Agrisera AB, Vännäs) for YTHDF1 and YTHDF2 proteins in tomato. Once produced these will allow western blot analysis to confirm the lack of YTHDF proteins in these null lines. Confirmation of this would add weight to all analysis and observation in this work.

Many studies have uncovered a role of m^{6}A in the stability of transcripts with the m^{6}A mark, with both increased stability and destabilisation being described (Shen *et al.*, 2016; Anderson *et al.*, 2018; Arribas-Hernández and Brodersen, 2020a). Production of a tomato m^{6}A library via m^{6}A-seq would be invaluable for us. Once produced we could repeat in our single mutant *ythdf* lines. Any changes in peak levels will guide more in-depth investigations in the future.

A recurring theme of YTH proteins in plants is the localisation of the proteins through the cellular response to stressful conditions Understanding cellular localisation of RNA binding proteins is essential in understand their function. In mammalian cell lines, m^{6}A is added to multiple translationally active mRNAs under stress (oxidative and heat). This induces 'triaging' to cytoplasmic assemblages known as stress granules (SGs). The formation of these granules was later shown to be reliant on all three YTHDF proteins (DF1-3) (Anders *et al.*, 2018). Similar induction of foci has been observed in *Arabidopsis,* and similarly linked to IDR of YTH protein N-termini (Arribas-Hernández *et al.*, 2018). All four YTHDF proteins in tomato have IDRs, but only YTHDF1 and YTHDF2 have been predicted to form assemblages *in vivo* (Figure 3.7). The importance of these assemblages for numerous biological processes are likely shared by tomato (and *Arabidopsis*) systems. Loss-of-function mutants may affect the formation of these granules/foci, this has been observed in mammalian systems (Anders *et al.*, 2018), it is possible YTHDF protein are essential for their formation as well.

A richer knowledge of expression profiles of YTHDF1-4 would provide invaluable direction for this study. We have developed GUS fusion constructs, powered by promoters of all four YTHDF proteins. We intend to use these to identify areas for future study. Identifying which transcripts bind to RNA binding proteins, in a protein centric manner, can be achieved in a number of ways. *In-vivo* UV crosslinking to crosslink proteins to bound RNAs, immunoprecipitation of protein-RNA complexes followed by radioactive labelling and high-throughput sequencing of purified RNA (iCLIP) (Huppertz *et al.*, 2014) would allow identification of YTHDF protein binding targets to a nucleotide resolution. This approach has already been utilised in the study of plant YTH proteins (Arribas-Hernández *et al.*, 2021), however IP methods have traditionally low efficiency which can make generation of sufficient RNA difficult (Darnell, 2010). During this study, we produced YTHDF-GFP fusion constructs which could be used for the immunoprecipitation step in this method. Another potential method could be the recently developed HyperTRIBE method (Rahman *et al.*, 2018). HyperTRIBE fuses the catalytic domain of modification enzyme ADAR, to RBPs of interest, inducing A-to-I editing in target RNAs . Inosine is recognised as guanosine by cellular enzymes and can therefore be detected at A-to-G mutations by high-throughput sequencing.

Further utilisation of our *in-vitro* RNA selection method in combination with any advances in tomato RNA methylomes could become an essential resource for future studies. This method can be used with any proteins, we look forward to similar studies in *Arabidopsis* and beyond.

Finally, more in-depth characterisation of all single and double mutant lines produced could be highly informative. All our findings are centred around plants at 4 weeks, and while it would not be practical to evaluate phenotypes at every developmental stage, investigations into fruit development, flowering time, root development and responses to key abiotic and biotic stresses have all yielded interesting findings in relation to m^6A (Zhong *et al.*, 2008; Bodi *et al.*, 2012b; Martínez-Pérez *et al.*, 2017; Růžička *et al.*, 2017; Wang, Guo, Sun, *et al.*, 2017; Anderson *et al.*, 2018; Arribas-Hernández and Brodersen, 2020a; Arribas-Hernández *et al.*, 2020). As previously mentioned, many of the aberrant phenotypes present in hypomethylation

lines overlap with those caused by impaired auxin function (Heisler *et al.*, 2005; Zhong *et al.*, 2008; Su *et al.*, 2017; Bhatia and Heisler, 2018; Arribas-Hernández and Brodersen, 2020b). This would likely be a fruitful source of discovery.

7.3 Concluding Remarks

To conclude, this thesis set out to develop a model for the study of YTHDF proteins and their role within plant species. We have successfully produced single mutant lines of all four YTHDF proteins in *Solanum lycopersicum*. These were subsequently crossed to produced double mutant lines for further study. To date, we have *ythdf1/ythdf3, ythdf2/ythdf3* and *ythdf3/ythdf4* double mutant lines, in addition we have proved that simultaneous KO of ythdf1 and ythdf4 is lethal.

We have begun the task of comprehensively phenotyping these lines to identify potential roles of these proteins *in vivo* and any link to the functions of m^6A. These finding show that not only are YTHDF proteins involved in many essential processes, that many of these processes overlap with those identified in hypomethylation lines. In addition, we have also demonstrated that our hypothesis that tomato would make a suitable model for this investigation was correct. Distinctly less redundancy was observed between YTHDF1, YTHDF2, YTHDF3 and YTHDF4, than demonstrated in *Arabidopsis*. Our hope is that the initial work here, we increase the ease and speed of future discoveries.

Furthermore, we developed and tested a *in vitro* selection strategy which can be utilised to identify binding affinities for RNA sequences. This could be invaluable for identifying key transcripts of any RNA binding protein.

All these findings improve our understanding of the complex mechanisms surrounding m^6A reader proteins and more importantly provide a firm foundation, increasing the scope of our investigations in the future.

Appendix

Appendix 3.1: Accession numbers for proteins used in comparison trees

Species	Accession number	Protein name DC/DF	Protein name Original
Ananas Comosus	Aco001739.1	AcoDF1C	AcoYTHDF1
	Aco003339.1	AcoDF4C	AcoYTHDF2
	Aco004279.1	AcoDF1A	AcoYTHDF3
	Aco004441.1	AcoDF2C	AcoYTHDF4
	Aco007343.1	AcoDF3C	AcoYTHDF5
	Aco007480.1	AcoDF2B	AcoYTHDF6
	Aco008320.1	AcoDF1B	AcoYTHDF7
	Aco008423.1	AcoDC1A	AcoYTHDC1
	Aco009712.1	AcoDF2A	AcoYTHDF8
Amaranthus hypochondriacus	AHYPO_001321-RA	AhDF1B	AhYTHDF1
	AHYPO_001693-RA	AhDF3C	AhYTHDF2
	AHYPO_003351-RA	AhDF2B	AhYTHDF3
	AHYPO_010570-RA	AhDF1A	AhYTHDF4
	AHYPO_011235-RA	AhDF2C	AhYTHDF5
	AHYPO_012466-RA	AhDC1A	AhYTHDC1
	AHYPO_014122-RA	AhDF1C	AhYTHDF6

	AHYPO_014607-RA	AhDF2A	AhYTHDF7
	AHYPO_015768-RA	AhDC1B	AhYTHDC2
Aquilegia caerluea	Aqcoe1G060600.1	AcDF1A	AcYTHDF1
	Aqcoe1G183700.1	AcDC1A	AcYTHDC1
	Aqcoe2G257600.1	AcDF1C	AcYTHDF2
	Aqcoe5G227200.1		AcYTHDC2
	Aqcoe5G460600.1		AcYTHDC3
	Aqcoe6G010000.1	AcDF3C	AcYTHDF3
	Aqcoe7G241300.1	AcDF1B	AcYTHDF4
Arabidopsis thaliana	AT1G09810	AtDF1	AtECT1 (AtYTHDF1)
	AT1G27960	AtDF2	AtECT2 (AtYTHDF2)
	AT1g48110	AtDF3	AtECT3 (AtYTHDF3)
	AT1G55500	AtDF4	AtECT4 (AtYTHDF4)
	AT1G79270	AtDF5	AtECT5 (AtYTHDF5)
	AT3G03950	AtDF6	AtECT6 (AtYTHDF6)
	AT3G13060	AtDF7	AtECT7 (AtYTHDF7)
	At3g13460	AtDF8	AtECT8 (AtYTHDF8)
	AT3G17330	AtDF9	AtECT9 (AtYTHDF9)

	AT4G11970	AtDF10	AtECT10 (AtYTHDF10)
	AT5G58190	AtDF11	AtECT11 (AtYTHDF11)
	AT5G61020	AtDC2	AtECT12 (AtYTHDC2)
Brassica oleracea	Bol008438	BoDF3B	BoYTHDF1
	Bol010247	BoDF1B	BoYTHDF2
	Bol012589	BoDF4A	BoYTHDF3
	Bol015553	BoDF4B	BoYTHDF4
	Bol015779	BoDF2B	BoYTHDF5
	Bol015814	BoDF3A	BoYTHDF6
	Bol016106	BoDC1A	BoYTHDC1
	Bol025994	BoDF5B	BoYTHDF7
	Bol035532	BoDF1A	BoYTHDF8
	Bol035536	BoDF5A	BoYTHDF9
	Bol036797	BoDF2C	BoYTHDF10
	Bol040382	BoDF2A	BoYTHDF11
	Bol044954	BoDC1B	BoYTHDC2
	Bol045155	BoDF1C	BoYTHDF12
Capsella rubella	Carubv10003808m	CrDC1B	CrYTHDC1
	Carubv10008560m	CrDF1C	CrYTHDF1
	Carubv10008570m	CrDC1A	CrYTHDC2
	Carubv10008725m	CrDF3B	CrYTHDF2

	Carubv10011018m	CrDF4C	CrYTHDF3
	Carubv10013132m	CrDF1B	CrYTHDF4
	Carubv10013133m	CrDF1A	CrYTHDF5
	Carubv10013212m	CrDF3C	CrYTHDF6
	Carubv10013762m	CrDF2A	CrYTHDF7
	Carubv10021946m	CrDF2C	CrYTHDF8
	Carubv10026226m	CrDF2B	CrYTHDF9
	Carubv10026294m	CrDF3A	CrYTHDF10
Dunaliella salina	Dusal.0106s00010.1.p	DsDC1	DsYTHDC1
Eucalyptus grandis	Eucgr.B03590.1.p	EgDF1A	EgYTHDF1
	Eucgr.B03591.1.p	EgDF2A	EgYTHDF2
	Eucgr.D01949.1.p	EgDC1B	EgYTHDC1
	Eucgr.E02496.1.p	EgDC1A	EgYTHDC2
	Eucgr.H04096.1.p	EgDF1C	EgYTHDF3
	Eucgr.J03195.1.p	EgDF1B	EgYTHDF4
Amborella trichopoda	evm_27.model.AmTr_v1.0_scaffold00001.426	AtrDC1B	AtrYTHDC1
	evm_27.model.AmTr_v1.0_scaffold00004.54	AtrDF1A	AtrYTHDF1

	evm_27.model.AmTr_v1.0_scaffold000025.287	AtrDF1B	AtrYTHDF2
	evm_27.model.AmTr_v1.0_scaffold000068.160	AtrDF2C	AtrYTHDF3
	evm_27.model.AmTr_v1.0_scaffold000071.175	AtrDF3C	AtrYTHDF4
	evm_27.model.AmTr_v1.0_scaffold000096.42	AtrDF1C	AtrYTHDF5
	evm_27.model.AmTr_v1.0_scaffold000155.52	AtrDC1A	AtrYTHDC2
Gossypium raimondii	Gorai.002G031800.1	GrDF2C	GrYTHDF1
	Gorai.002G182900.1	GrDF3B	GrYTHDF2
	Gorai.004G020600.1	GrDF1A	GrYTHDF3
	Gorai.004G205800.1	GrDC1B	GrYTHDC1
	Gorai.004G244600.1	GrDF3A	GrYTHDF4
	Gorai.005G070500.1	GrDF1C	GrYTHDF5
	Gorai.005G265300.1	GrDF3C	GrYTHDF6
	Gorai.007G288900.1	GrDF2A	GrYTHDF7

	Gorai.007G289200.1	GrDF4A	GrYTHDF8
	Gorai.008G046800.1	GrDC1A	GrYTHDC2
	Gorai.009G120700.1	GrDF2B	GrYTHDF9
	Gorai.011G001700.1	GrDF1B	GrYTHDF10
	Gorai.013G123200.1	GrDF5A	GrYTHDF11
Musa acuminata	GSMUA_Achr1P10560_001	MaDF5B	MaYTHDF1
	GSMUA_Achr1P10590_001	MaDF4B	MaYTHDF2
	GSMUA_Achr1P24980_001	MaDF3C	MaYTHDF3
	GSMUA_Achr4P04600_001	MaDF1B	MaYTHDF4
	GSMUA_Achr4P11560_001	MaDF2B	MaYTHDF5
	GSMUA_Achr4P12480_001	MaDF3B	MaYTHDF6
	GSMUA_Achr4P21380_001	MaDF1A	MaYTHDF7
	GSMUA_Achr4P22630_001	MaDF2A	MaYTHDF8
	GSMUA_Achr4P28430_001	MaDF4A	MaYTHDF9
	GSMUA_Achr5P02830_001	MaDF3A	MaYTHDF10

	GSMUA_Achr6P12810_001	MaDF5C	MaYTHDF11
	GSMUA_Achr6P19820_001	MaDC2A	MaYTHDC1
	GSMUA_Achr6P26690_001	MaDF2C	MaYTHDF12
	GSMUA_Achr7P01490_001	MaDF5A	MaYTHDF13
	GSMUA_Achr7P27220_001	MaDF6A	MaYTHDF14
	GSMUA_Achr8P29890_001	MaDF4C	MaYTHDF15
	GSMUA_Achr9P06180_001	MaDF1C	MaYTHDF16
	GSMUA_Achr9P30440_001	MaDC1A	MaYTHDC2
Vitis vinifera	GSVIVT01010494001	VvDF2B	VvYTHDF1
	GSVIVT01017649001	VvDF1C	VvYTHDC1
	GSVIVT01019895001	DF	VvYTHDF2
	GSVIVT01023248001	VvDF3C	VvYTHDF3
	GSVIVT01025494001	VvDF1A	VvYTHDF4
	GSVIVT01029460001	VvDF1B	VvYTHDF5
	GSVIVT01032236001	VvDF3B	VvYTHDF6

	GSVIVT01033395001	VvDF2A	VvYTHDF7
	GSVIVT01034843001	VvDC1B	VvYTHDC2
	GSVIVT01036696001	VvDF2C	VvYTHDF8
Manihot esculenta	Manes.02G029400.1.p	MeDC1B	MeYTHDC1
	Manes.06G172700.1.p	MeDF3A	MeYTHDF1
	Manes.07G024100.1.p	MeDF3C	MeYTHDF2
	Manes.07G081200.1.p	MeDF2A	MeYTHDF3
	Manes.09G141700.1.p	MeDF5A	MeYTHDF4
	Manes.10G067800.1.p	MeDF4A	MeYTHDF5
	Manes.12G010500.1.p	MeDF3B	MeYTHDF6
	Manes.13G011300.1.p	MeDF2B	MeYTHDF7
	Manes.14G000200.1.p	MeDF1B	MeYTHDF8
	Manes.14G012900.1.p	MeDF1A	MeYTHDF9
	Manes.14G122500.1.p	MeDC1A	MeYTHDC2
	Manes.15G003300.1.p	MeDF1C	MeYTHDF10

	Manes.16G048600.1.p	MeDF2C	MeYTHDF11
Marchantia polymorpha	Mapoly0007s0039.1.p	MpoDC1B	MpoYTHDC1
	Mapoly0056s0030.1.p	MpoDC1A	MpoYTHDC2
	Mapoly0103s0072.1.p	MpoDF1	MpoYTHDF1
Medicago truncatula	Medtr2g005840.1	MtDF4C	MtYTHDF1
	Medtr2g029540.1	MtDC1A	MtYTHDC1
	Medtr2g032660.1	MtDF3C	MtYTHDF2
	Medtr2g063460.1	MtDF2C	MtYTHDF3
	Medtr4g005110.1	MtDF1B	MtYTHDF4
	Medtr4g123630.1	MtDF1C	MtYTHDF5
	Medtr6g032970.1	MtDF3A	MtYTHDF6
	Medtr7g085500.1	MtDF2A	MtYTHDF7
	Medtr8g088390.1	MtDF1A	MtYTHDF8
Mimulus glutinosus	Migut.A00893.1.p	MgDC1B	MgYTHDC1
	Migut.D02575.1.p	MgDF2A	MgYTHDF1
	Migut.G00015.1.p	MgDF2C	MgYTHDF2
	Migut.H01337.1.p	MgDC1A	MgYTHDC2
	Migut.J00855.1.p	MgDF1B	MgYTHDF3
	Migut.M01886.1.p	MgDC2A	MgYTHDC3
	Migut.N00538.1.p	MgDF1A	MgYTHDF4
	Migut.N00839.1.p	MgDF3C	MgYTHDF5
	Migut.O00389.1.p	MgDF1C	MgYTHDF6

Citrus sinensis	orange1.1g005158m	CsiDF1A	CsiYTHDF1
	orange1.1g005338m	CsiDC1A	CsiYTHDC1
	orange1.1g005441m	CsiDF1C	CsiYTHDF2
	orange1.1g005934m	CsiDF1B	CsiYTHDF3
	orange1.1g006798m	CsiDF3B	CsiYTHDF4
	orange1.1g008255m	CsiDF2C	CsiYTHDF5
	orange1.1g010141m	CsiDF2A	CsiYTHDF6
	orange1.1g015749m	CsiDC1B	CsiYTHDC2
	orange1.1g038459m	CsiDF2B	CsiYTHDC3
	orange1.1g048663m	CsiDF3C	CsiYTHDF7
Oropetium thomaeum	Oropetium_20150105_00131A	OtDF1A	OtYTHDF1
	Oropetium_20150105_03021A	OtDF2C	OtYTHDF2
	Oropetium_20150105_07558A	OtDF2A	OtYTHDF3
	Oropetium_20150105_09547A	OtDF3A	OtYTHDF4
	Oropetium_20150105_11598A	OtDF1C	OtYTHDF5

	Oropetium_20150105_14508A	OtDF1B	OtYTHDF6
	Oropetium_20150105_15590A	OtDC1A	OtYTHDC1
	Oropetium_20150105_26935A	OtDF3C	OtYTHDF7
Oryza sativa	OsYTH1		OsYTHDF1
	OsYTH2		OsYTHDF2
	OsYTH3		OsYTHDF3
	OsYTH4		OsYTHDF4
	OsYTH5		OsYTHDF5
	OsYTH6		OsYTHDF6
	OsYTH7		OsYTHDF7
	OsYTH8		OsYTHDF8
	OsYTH9		OsYTHDC2
	OsYTH10		OsYTHDF9
	OsYTH11		OsYTHDF10
	OsYTH12		OsYTHDF11
	OsYTH13		OsYTHDF12
Panicum halli	Pahal.B00424.1	PhDF2A	PhYTHDF1
	Pahal.C00997.1	PhDF4C	PhYTHDF2
	Pahal.C01447.1	PhDF3C	PhYTHDF3
	Pahal.D00506.1	PhDC1A	PhYTHDC1
	Pahal.E02120.1	PhDF1C	PhYTHDF4
	Pahal.F00398.1	PhDF1B	PhYTHDF5
	Pahal.F01360.1	PhDF2B	PhYTHDF6
	Pahal.G02347.1	PhDF3B	PhYTHDF7

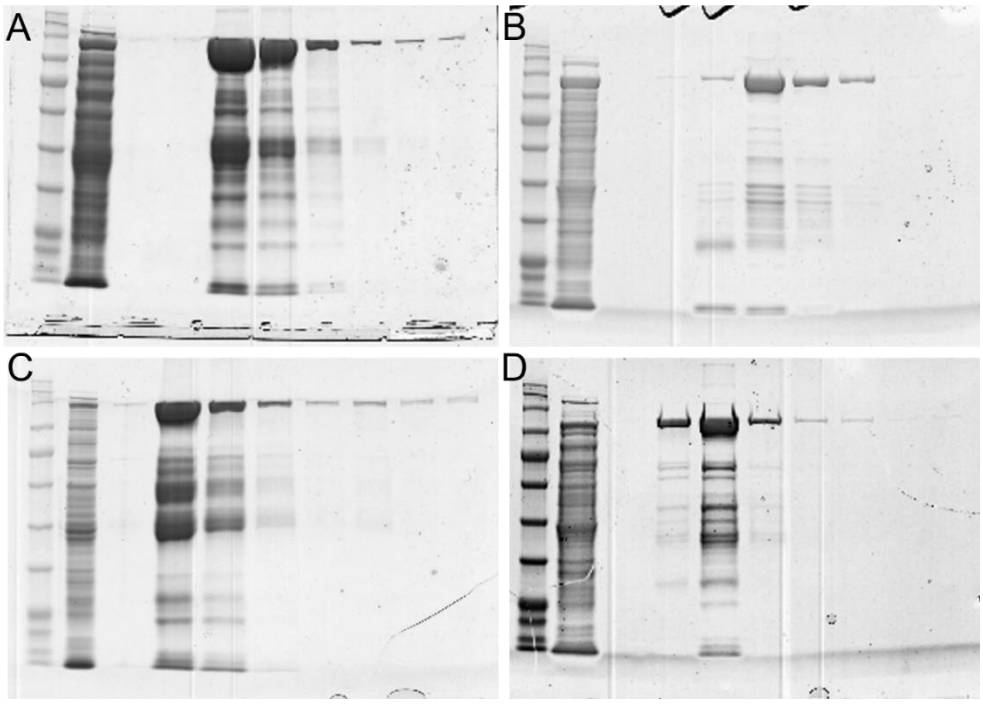
	Pahal.H01140.1	PhDF2C	PhYTHDF8
	Pahal.I00478.1	PhDF1A	PhYTHDF9
	Pahal.I01354.1	PhDF4A	PhYTHDF10
	Pahal.I01743.1	PhDF3A	PhYTHDF11
Populus trichocarpa	Potri.001G002000.1	PtDF2A	PtYTHDF1
	Potri.001G056100.1	PtDF5B	PtYTHDF2
	Potri.001G113500.1	PtDC1B	PtYTHDC1
	Potri.001G357800.1	PtDC1A	PtYTHDC2
	Potri.003G008400.1	PtDF6C	PtYTHDF3
	Potri.003G222700.1	PtDF1A	PtYTHDF4
	Potri.004G223800.1	PtDF5C	PtYTHDF5
	Potri.006G079900.1	PtDF3B	PtYTHDF6
	Potri.007G002800.1	PtDF2B	PtYTHDF7
	Potri.008G080800.1	PtDF4C	PtYTHDF8
	Potri.008G100200.1	PtDF2C	PtYTHDF9
	Potri.010G152300.1	PtDF1C	PtYTHDF10
	Potri.010G175500.1	PtDF3C	PtYTHDF11

	Potri.011G089800.1	PtDC2A	PtYTHDC3
	Potri.014G001000.1	PtDF1B	PtYTHDF12
	Potri.018G149800.1	PtDF4B	PtYTHDF13
	Potri.019G034300.1	PtDF3A	PtYTHDF14
Physcomitrella patens	Pp3c3_33560V3.1.p	PpaDF1	PpYTHDF1
	Pp3c8_11880V3.1.p	PpaDC1B	PpYTHDC1
	Pp3c10_17810V3.1.p	PpaDF2	PpYTHDF2
	Pp3c20_9290V3.1.p	PpaDC1A	PpYTHDC2
Prunus persica	Prupe.1G002400.1.p	PpDF2B	PpYTHDF3
	Prupe.1G109700.1.p	PpDF1C	PpYTHDF4
	Prupe.1G579100.1.p	PpDF3C	PpYTHDF5
	Prupe.3G027100.1.p	PpDF1B	PpYTHDF6
	Prupe.4G089500.1.p	PpDC1A	PpYTHDC3
	Prupe.4G124900.1.p	PpDF2C	PpYTHDF7
	Prupe.5G086300.1.p	PpDC1B	PpYTHDC4

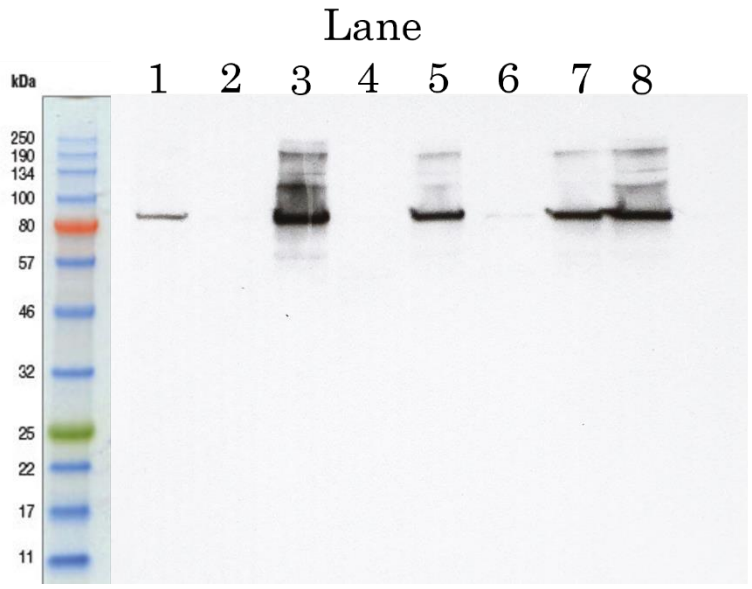
	Prupe.6G002200.1.p	PpDF1A	PpYTHDF8
Senna italica	Seita.2G046200.1.p	SiDF3A	SiYTHDF1
	Seita.3G070000.1.p	SiDF3C	SiYTHDF2
	Seita.4G262400.1.p	SiDC1A	SiYTHDC1
	Seita.5G156800.1.p	SiDF2C	SiYTHDF3
	Seita.5G271100.1.p	SiDF1C	SiYTHDF4
	Seita.6G022500.1.p	SiDF1B	SiYTHDF5
	Seita.6G246600.1.p	SiDF2B	SiYTHDF6
	Seita.7G231300.1.p	SiDF3B	SiYTHDF7
	Seita.7G270400.1.p	SiDF4C	SiYTHDF8
	Seita.9G059600.1.p	SiDF4A	SiYTHDF9
	Seita.9G092900.1.p	SiDF2A	SiYTHDF10
	Seita.9G534000.1.p	SiDF1A	SiYTHDF11
Sphagnum fallax	SfYTH1	F	SfYTHDF1
	SfYTH2	F	SfYTHDF2
	SfYTH3	F	SfYTHDF3
	SfYTH4	C	SfYTH04

	SfYTH5	C	SfYTH05
	SfYTH6		SfYTH06
Selaginella moellendorffii	SmYTH1		SmYTH01
	SmYTH2		SmYTH02
Solanum lycopersicum	SolyCPSF30		SlCPSF30
	SolyECT1		SlYTHDF1
	SolyECT2		SlYTHDF2
	SolyECT3		SlYTHDF3
	SolyECT4		SlYTHDF4
	SolyYTHa		SlYTHDC1a
	SolyYTHb		SlYTHDC1b
	SolyYTHc		SlYTHDC1c
Theobroma cacao	Thecc1EG006095t1	TcDC1A	TcYTHDC1
	Thecc1EG006167t1	TcDF2C	TcYTHDF1
	Thecc1EG008783t1	TcDF2B	TcYTHDF2
	Thecc1EG016131t1	TcDC1B	TcYTHDC2
	Thecc1EG021457t1	TcDF1C	TcYTHDF3
	Thecc1EG026461t1	TcDF2A	TcYTHDF4
	Thecc1EG027659t6	TcDF1B	TcYTHDF5

	Thecc1EG039865t1	TcDF1A	TcYTHDF6
Volvox carteri	Vocar.0006s0335.1.p	DC1	VcYTHDC1
Zea mays	ZmYTH1		ZmYTH01
	ZmYTH2		ZmYTH02
	ZmYTH3		ZmYTH03
	ZmYTH4		ZmYTH04
	ZmYTH5		ZmYTH05
	ZmYTH6		ZmYTH06
	ZmYTH7		ZmYTH07
	ZmYTH8		ZmYTH08
	ZmYTH9		ZmYTH09
	ZmYTH10		ZmYTH10
	ZmYTH11		ZmYTH11
	ZmYTH12		ZmYTH12
Zostera marina	Zosma14g00600.1	ZmaDF4C	ZmaYTHDF1
	Zosma39g00650.1	ZmaDF1A	ZmaYTHDF2
	Zosma91g00280.1	ZmaDF1B	ZmaYTHDF3
	Zosma103g00140.1	ZmaDF3C	ZmaYTHDF4
	Zosma105g00190.1	ZmaDF2C	ZmaYTHDF5
	Zosma105g00770.1	ZmaDF5C	ZmaYTHDF6
	Zosma158g00160.1	ZmaDC1A	ZmaYTHDC1
	Zosma467g00180.1	ZmaDF1CA	ZmaYTHDF7



Appendix 6.1: Purification of *Solanum lycoperisum* YTHDF proteins (A) SDS-PAGE protein gel from the purification of YTHDF1 (B) SDS-PAGE protein gel from the purification of YTHDF2 (C) SDS-PAGE protein gel from the purification of YTHDF3 (D) SDS-PAGE protein gel from the purification of YTHDF4



Appendix 6.2: Western blot from the maltose elution troubleshooting assay (1) Non-eluted amylose beads bound to YTHDF1 (2) Non-eluted amylose beads bound to YTHDF1 + Random RNA oligonucleotide library (3) Supernatant from 4◦c elution (4) Beads from from 4◦c elution (5) Supernatant from room temp elution (6) Beads from room temp elution (7) Supernatant from room temp elution + (3o mins rest) (8) Beads from room temp elution + (3o mins rest).

References

Achsel, T. *et al.* (1999) *A doughnut-shaped heteromer of human Sm-like proteins binds to the 3-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro*, *The EMBO Journal*.

Addepalli, B. and Hunt, A.G. (2007) "A novel endonuclease activity associated with the Arabidopsis ortholog of the 30-kDa subunit of cleavage and polyadenylation specificity factor," *Nucleic Acids Research*, 35(13), pp. 4453–4463. doi:10.1093/nar/gkm457.

Adelman, K. and Lis, J.T. (2012) "Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans.," *Nature reviews. Genetics*, pp. 720–731. doi:10.1038/nrg3293.

Agarwala, S.D. *et al.* (2012) "RNA Methylation by the MIS Complex Regulates a Cell Fate Decision in Yeast," *PLoS Genet*, 8(6), p. 1002732. doi:10.1371/journal.pgen.1002732.

Aguilo, F. and Walsh, M.J. (2017) "The N6-Methyladenosine RNA modification in pluripotency and reprogramming," *Current Opinion in Genetics and Development*, 46, pp. 77–82. doi:10.1016/j.gde.2017.06.006.

Akichika, S. *et al.* (2019) "Cap-specific terminal N 6 -methylation of RNA by an RNA polymerase II–associated methyltransferase," *Science*, 363(6423). doi:10.1126/science.aav0080.

Alaguero-Cordovilla, A. *et al.* (2018) "Morphological Characterization of Root System Architecture in Diverse Tomato Genotypes during Early Growth," *International Journal of Molecular Sciences Article* [Preprint]. doi:10.3390/ijms19123888.

Anders, M. *et al.* (2018) "Dynamic m 6 A methylation facilitates mRNA triaging to stress granules." doi:10.26508/lsa.201800113.

Anderson, J. *et al.* (1998) *The essential Gcd10p-Gcd14p nuclear complex is required for 1-methyladenosine modification and maturation of initiator methionyl-tRNA*. Available at: www.genesdev.org (Accessed: April 8, 2020).

Anderson, S.J. *et al.* (2018) "N6-Methyladenosine Inhibits Local Ribonucleolytic Cleavage to Stabilize mRNAs in Arabidopsis," *Cell Reports*, 25(5), pp. 1146-1157.e3. doi:10.1016/j.celrep.2018.10.020.

Arguello, A.E., Deliberto, A.N. and Kleiner, R.E. (2017) "RNA Chemical Proteomics Reveals the N6-Methyladenosine (m6A)-Regulated Protein-RNA Interactome," *Journal of the American Chemical Society*, 139(48), pp. 17249–17252. doi:10.1021/jacs.7b09213.

Arguello, A.E., Leach, R.W. and Kleiner, R.E. (2019) " In Vitro Selection with a Site-Specifically Modified RNA Library Reveals the Binding Preferences of N 6 -Methyladenosine Reader Proteins ," *Biochemistry*, 58(31), pp. 3386–3395. doi:10.1021/acs.biochem.9b00485.

Arribas-Hernández, L. *et al.* (2018) "An m6A-YTH Module Controls Developmental Timing and Morphogenesis in Arabidopsis," *The Plant Cell*, p. tpc.00833.2017. doi:10.1105/tpc.17.00833.

Arribas-Hernández, L. *et al.* (2020) "Recurrent requirement for the m6A-ECT2/ECT3/ECT4 axis in the control of cell proliferation during plant organogenesis," *Development (Cambridge)*, 147(14). doi:10.1242/dev.189134.

Arribas-Hernández, L. *et al.* (2021) "Principles of mRNA targeting and regulation via the Arabidopsis m^{6}A-binding proteins ECT2 and ECT3," *bioRxiv*, p. 2021.04.18.440342. Available at: http://biorxiv.org/content/early/2021/04/19/2021.04.18.440342.abstract (Accessed: April 21, 2021).

Arribas-Hernández, L. and Brodersen, P. (2020a) "Update on Covalent Modification of Plant mRNA Occurrence and Functions of m 6 A and Other Covalent Modifications in Plant mRNA 1[OPEN]." doi:10.1104/pp.19.01156.

Arribas-Hernández, L. and Brodersen, P. (2020b) "Update on Covalent Modification of Plant mRNA Occurrence and Functions of m 6 A and Other Covalent Modifications in Plant mRNA 1[OPEN]." doi:10.1104/pp.19.01156.

Aviv, A.K. (2003) "N E W S A N D V I E W S The CTD code How does the C-terminal domain (CTD) of RNA polymerase II interact specifically with multiple targets?," *Proc. Natl. Acad. Sci. USA*, 11(9), pp. 432–438. Available

at: http://www.nature.com/naturestructuralbiology (Accessed: April 3, 2020).

Bass, B.L. (2013) "RNA Editing," in *Brenner's Encyclopedia of Genetics: Second Edition*, pp. 277–279. doi:10.1016/B978-0-12-374984-0.01344-9.

Batista, P.J. *et al.* (2014) "M^{6}A RNA modification controls cell fate transition in mammalian embryonic stem cells," *Cell Stem Cell*, 15(6). doi:10.1016/j.stem.2014.09.019.

Belhaj, K. *et al.* (2013) "Plant genome editing made easy: Targeted mutagenesis in model and crop plants using the CRISPR/Cas system," *Plant Methods*. doi:10.1186/1746-4811-9-39.

Berlivet, S. *et al.* (2019) "Readers of the m 6 A epitranscriptomic code," *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1862(3), pp. 329–342. doi:10.1016/j.bbagrm.2018.12.008.

Bevilacqua, P.C. *et al.* (2016) "Genome-Wide Analysis of RNA Secondary Structure." doi:10.1146/annurev-genet-120215-035034.

Bhatia, N. and Heisler, M.G. (2018) "Self-organizing periodicity in development: Organ positioning in plants," *Development (Cambridge)*. doi:10.1242/dev.149336.

Bienroth, S., Keller, W. and Wahle, E. (1993) *Assembly of a processive messenger RNA polyadenylation complex*, *The EMBO Journal*.

Bird, C.R. *et al.* (1988) *The tomato polygalacturonase gene and ripening-specific expression in transgenic plants*, *Plant Molecular Biology*.

Boccaletto, P. *et al.* (2018) "MODOMICS: A database of RNA modification pathways. 2017 update," *Nucleic Acids Research*, 46(D1), pp. D303–D307. doi:10.1093/nar/gkx1030.

Bodi, Z. *et al.* (2012a) "Adenosine methylation in Arabidopsis mRNA is associated with the 3′ end and reduced levels cause developmental defects," *Frontiers in Plant Science*, 3(MAR), p. 48. doi:10.3389/fpls.2012.00048.

Bodi, Z. *et al.* (2012b) "Adenosine Methylation in Arabidopsis mRNA is Associated with the 3′ End and Reduced Levels Cause Developmental

Defects," *Frontiers in Plant Science*, 3(March), pp. 1–10. doi:10.3389/fpls.2012.00048.

Bokar, J.A. *et al.* (1997) "Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase," *Rna*, 3(11), pp. 1233–1247.

Bokarso, J.A. *et al.* (1994) *THE JOURNAL OF B r o m r c ~ ~ CHEMISTRY Characterization and Partial Purification of mRNA NG-Adenosine Methyltransferase from HeLa Cell Nuclei INTERNAL mRNA METHYLATION REQUIRES A MULTISUBUNIT COMPLEX**.

Bringmann, P. and Luhrmann, R. (1986) *Purification of the individual snRNPs Ul, U2, U5 and U4/U6 from HeLa cells and characterization of their protein constituents*, *The EMBO Journal*.

Brooks, C. *et al.* (2014) "Efficient Gene Editing in Tomato in the First Generation Using the Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-Associated9 System," *Plant Physiology*, 166(3), pp. 1292–1297. doi:10.1104/pp.114.247577.

Buratowski, S. *et al.* (1989) "Five intermediate complexes in transcription initiation by RNA polymerase II," *Cell*, 56(4), pp. 549–561. doi:10.1016/0092-8674(89)90578-3.

Carlile, T.M. *et al.* (2014) "Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells," *Nature*, 515(7525), pp. 143–146. doi:10.1038/nature13802.

Carroll, S.M., Narayan, P. and Rottman, F.M. (1990) *N6-Methyladenosine Residues in an Intron-Specific Region of Prolactin Pre-mRNA*, *MOLECULAR AND CELLULAR BIOLOGY*.

de Castro, R.D. *et al.* (2000) "Cell division and subsequent radicle protrusion in tomato seeds are inhibited by osmotic stress but DNA synthesis and formation of microtubular cytoskeleton are not," *Plant Physiology*, 122(2), pp. 327–335. doi:10.1104/pp.122.2.327.

Chen, M. *et al.* (2018) "RNA N6-methyladenosine methyltransferase-like 3 promotes liver cancer progression through YTHDF2-dependent

posttranscriptional silencing of SOCS2," *Hepatology*, 67(6), pp. 2254–2270. doi:10.1002/hep.29683.

Chen, T. *et al.* (2015) "M6A RNA methylation is regulated by microRNAs and promotes reprogramming to pluripotency," *Cell Stem Cell*, 16(3). doi:10.1016/j.stem.2015.01.016.

Chernyakov, I. *et al.* (2008) "Degradation of several hypomodified mature tRNA species in Saccharomyces cerevisiae is mediated by Met22 and the 5-3 exonucleases Rat1 and Xrn1." doi:10.1101/gad.1654308.

Cline, M.G. (1997) "Concepts and terminology of apical dominance," *American Journal of Botany*, 84(8), pp. 1064–1069. doi:10.2307/2446149.

Cohn, W.E., Volkin, E. (1951) "Nucleoside-5'-Phosphate from Ribonucleic Acid," *Nature*, 167(1), pp. 483–484. doi:10.1021/jf60001a611.

Conaway, R.C. and Conaway, J.W. (2019) "The hunt for RNA polymerase II elongation factors: a historical perspective," *Nature Structural and Molecular Biology*, pp. 771–776. doi:10.1038/s41594-019-0283-1.

Conti, G. *et al.* (2017) "TMV induces RNA decay pathways to modulate gene silencing and disease symptoms," *Plant Journal*, 89(1), pp. 73–84. doi:10.1111/tpj.13323.

Corden, J.L. (1990) "Tails of RNA polymerase II," *Trends in Biochemical Sciences*, 15(10), pp. 383–387. doi:10.1016/0968-0004(90)90236-5.

Courtney, D.G. *et al.* (2018) "Epitranscriptomic enhancement of influenza A virus gene expression and replication," *Cell Host Microbe*, 22(3), pp. 377–386. doi:10.1016/j.chom.2017.08.004.Epitranscriptomic.

Daffis, S. *et al.* (2010) "29-O methylation of the viral mRNA cap evades host restriction by IFIT family members," *Nature*, 468. doi:10.1038/nature09489.

Darnell, R.B. (2010) "HITS-CLIP: Panoramic views of protein-RNA regulation in living cells," *Wiley Interdisciplinary Reviews: RNA*, 1(2), pp. 266–286. doi:10.1002/wrna.31.

Darnell, R.B., Ke, S. and Darnell, J.E. (2018) "Protein–RNA interactions: structural characteristics and hotspot amino acids," *RNA*, 24, pp. 1457–1465. doi:10.1261/rna.

Daube, S.S. and von Hippel, P.H. (1992) "Functional transcription elongation complexes from synthetic RNA-DNA bubble duplexes," *Science*, 258(5086), pp. 1320–1324. doi:10.1126/science.1280856.

Davis, C.A. and Ares, M. (2006) *Accumulation of unstable promoter-associated transcripts upon loss of the nuclear exosome subunit Rrp6p in Saccharomyces cerevisiae*, *PNAS*. Available at: www.pnas.orgcgidoi10.1073pnas.0507783103 (Accessed: April 3, 2020).

Deng, W. and Roberts, S.G.E. (2005) "A core promoter element downstream of the TATA box that is recognized by TFIIB." doi:10.1101/gad.342405.

Desrosiers, R., Friderici, K. and Rottman, F. (1974) *Identification of Methylated Nucleosides in Messenger RNA from Novikoff Hepatoma Cells (RNA methylation/RNA processing/methylnucleoside composition)*.

Dominissini, D. *et al.* (2012) "Topology of the human and mouse m 6 A RNA methylomes revealed by m 6 A-seq," *Nature* [Preprint]. doi:10.1038/nature11112.

Dominissini, D. *et al.* (2016) "The dynamic N1 -methyladenosine methylome in eukaryotic messenger RNA," *Nature*, 530(7591), pp. 441–446. doi:10.1038/nature16998.

Drechsel, G. *et al.* (2013) "Nonsense-mediated decay of alternative precursor mRNA splicing variants is a major determinant of the Arabidopsis steady state transcriptome," *Plant Cell*, 25(10), pp. 3726–3742. doi:10.1105/tpc.113.115485.

Du, H. *et al.* (2016) "YTHDF2 destabilizes m 6 A-containing RNA through direct recruitment of the CCR4-NOT deadenylase complex," *Nature Communications*, 7. doi:10.1038/ncomms12626.

Duan, H.-C. *et al.* (2017) "ALKBH10B Is an RNA N^6 -Methyladenosine Demethylase Affecting Arabidopsis Floral Transition," *The Plant Cell*, 29(12), pp. 2995–3011. doi:10.1105/tpc.16.00912.

Dubin, D.T. and Taylor, R.H. (1975) *The methylation state of poly A-containing-messenger RNA from cultured hamster cells*, *Nucleic Acids Research*.

Duc, C. *et al.* (2013) "Transcription Termination and Chimeric RNA Formation Controlled by Arabidopsis thaliana FPA," *PLoS Genetics*, 9(10). doi:10.1371/journal.pgen.1003867.

Dunn, D.B. (1961) "The occurence of 1-methyladenine in ribonucleic acid," *BBA - Biochimica et Biophysica Acta*, 46(1), pp. 198–200. doi:10.1016/0006-3002(61)90668-0.

van Eck, J. (2018) "Genome editing and plant transformation of solanaceous food crops," *Current Opinion in Biotechnology*, 49, pp. 35–41. doi:10.1016/j.copbio.2017.07.012.

Edgar, R.C. (2004) "MUSCLE: Multiple sequence alignment with high accuracy and high throughput," *Nucleic Acids Research*, 32(5), pp. 1792–1797. doi:10.1093/nar/gkh340.

Edwalds-Gilbert, G., Prescott, J. and Falck-Pedersen, E. (1993) *3' RNA Processing Efficiency Plays a Primary Role in Generating Termination-Competent RNA Polymerase II Elongation Complexes*, *MOLECULAR AND CELLULAR BIOLOGY*.

Engler, C., Kandzia, R. and Marillonnet, S. (2008) "A one pot, one step, precision cloning method with high throughput capability," *PLoS ONE*, 3(11). doi:10.1371/journal.pone.0003647.

Eulalio, A. *et al.* (2007) "P-Body Formation Is a Consequence, Not the Cause, of RNA-Mediated Gene Silencing †," *MOLECULAR AND CELLULAR BIOLOGY*, 27(11), pp. 3970–3981. doi:10.1128/MCB.00128-07.

F. H. Crick (1958) "On protein synthesis," *Symposia of the Society for Experimental Biology*, 12, pp. 138–163. doi:10.1038/227561a0.

Fernández, I.S. *et al.* (2013) "Unusual base pairing during the decoding of a stop codon by the ribosome HHS Public Access," *Nature*, 500(7460), pp. 107–110. doi:10.1038/nature12302.

Filichkin, S.A. *et al.* (2015) "Environmental Stresses Modulate Abundance and Timing of Alternatively Spliced Circadian Transcripts in Arabidopsis," *Mol. Plant*, 8, pp. 207–227. doi:10.1016/j.molp.2014.10.011.

Fourmann, J.B. *et al.* (2013) "Dissection of the factor requirements for spliceosome disassembly and the elucidation of its dissociation products using a purified splicing system," *Genes and Development*, 27(4), pp. 413–428. doi:10.1101/gad.207779.112.

Fu, Y. and Zhuang, X. (2019) "m6A-binding YTHDF proteins promote stress granule formation by modulating phase separation of stress granule proteins," *bioRxiv*, p. 694455. doi:10.1101/694455.

Fujita, Y. *et al.* (2002) "Hakai, a c-Cbl-like protein, ubiquitinates and induces endocytosis of the E-cadherin complex," *NATURE CELL BIOLOGY*, 4. doi:10.1038/ncb758.

Fustin, J.M. *et al.* (2013) "XRNA-methylation-dependent RNA processing controls the speed of the circadian clock," *Cell*, 155(4), pp. 793–806. doi:10.1016/j.cell.2013.10.026.

Galej, W.P. *et al.* (2016) "Cryo-EM structure of the spliceosome immediately after branching," *Nature*, 537(7619), pp. 197–201. doi:10.1038/nature19316.

Gautheret, D. *et al.* (1998) "Alternate polyadenylation in human mRNAs: A large-scale analysis by EST clustering," *Genome Research*, 8(5), pp. 524–530. doi:10.1101/gr.8.5.524.

Gershenzon, N.I. and Ioshikhes, I.P. (2005) "Synergy of human Pol II core promoter elements revealed by statistical sequence analysis," *BIOINFORMATICS DISCOVERY NOTE*, 21(8), pp. 1295–1300. doi:10.1093/bioinformatics/bti172.

Gerstberger, S., Hafner, M. and Tuschl, T. (2014) "A census of human RNA-binding proteins," *Nature Publishing Group*, 15, p. 829. doi:10.1038/nrg3813.

Ghosh, A. and Lima, C.D. (2010) "Enzymology of RNA cap synthesis NIH Public Access," *Wiley Interdiscip Rev RNA*, 1(1), pp. 152–172. doi:10.1002/wrna.19.

Gilbert Walter (1986) "The RNA world," *Nature*, 319, p. 618.

Gilmartin, G.M. (2005) “Eukaryotic mRNA 3′ processing: A common means to different ends,” *Genes and Development*, 19(21), pp. 2517–2521. doi:10.1101/gad.1378105.

Glover-Cutter, K. *et al.* (2008) “RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes,” *Nature Structural and Molecular Biology*, 15(1), pp. 71–78. doi:10.1038/nsmb1352.

Gnatt, A.L. *et al.* (2001) “Structural basis of transcription: An RNA polymerase II elongation complex at 3.3 Å resolution,” *Science*, 292(5523), pp. 1876–1882. doi:10.1126/science.1059495.

Gokhale, N.S. *et al.* (2016) “N6-Methyladenosine in Flaviviridae Viral RNA Genomes Regulates Infection,” *Cell Host and Microbe*, 20(5), pp. 654–665. doi:10.1016/j.chom.2016.09.015.

Greco, M. *et al.* (2012) “In Posidonia oceanica cadmium induces changes in DNA methylation and chromatin patterning,” *Journal of Experimental Botany*, 63(2), pp. 695–709. doi:10.1093/jxb/err313.

Grimstad, S.O. (1993) “The effect of a daily low temperature pulse on growth and development of greenhouse cucumber and tomato plants during propagation,” *Scientia Horticulturae*, 53(1–2), pp. 53–62. doi:10.1016/0304-4238(93)90137-F.

Groot, S.P.C. and Karssen, C.M. (1987) “Gibberellins regulate seed germination in tomato by endosperm weakening: a study with gibberellin-deficient mutants,” *Planta*, 171(4), pp. 525–531. doi:10.1007/BF00392302.

Gu, H., Gupta, D. and Schoenberg, D.R. (1999) *The poly(A)-limiting element is a conserved cis-acting sequence that regulates poly(A) tail length on nuclear pre-mRNAs*, *Biochemistry*. Available at: www.pnas.org. (Accessed: April 3, 2020).

Harper, J.E. *et al.* (1990) *Sequence specificity of the human mRNA N6-adenosine methylase in vitro*, *Nucleic Acids Research*. doi:10.1093/nar/18.19.5735.

Hartmann, A.M. *et al.* (1999) *The Interaction and Colocalization of Sam68 with the Splicing-associated Factor YT521-B in Nuclear Dots Is Regulated by the Src Family Kinase p59 fyn*, *Molecular Biology of the Cell*.

Haussmann, I.U. *et al.* (2016) "m6A potentiates Sxl alternative pre-mRNA splicing for robust Drosophila sex determination," *Nature*, 540(7632), pp. 301–304. doi:10.1038/nature20577.

He, L. *et al.* (2019) "Functions of N6-methyladenosine and its role in cancer," *Molecular Cancer*. doi:10.1186/s12943-019-1109-9.

Hedden, P. and Kamiya, Y. (1997) *GIBBERELLIN BIOSYNTHESIS: Enzymes, Genes and Their Regulation*, *Annu. Rev. Plant. Physiol. Plant. Mol. Biol.* Available at: www.annualreviews.org (Accessed: April 22, 2021).

Heisler, M.G. *et al.* (2005) "Patterns of auxin transport and gene expression during primordium development revealed by live imaging of the Arabidopsis inflorescence meristem," *Current Biology*, 15(21), pp. 1899–1911. doi:10.1016/j.cub.2005.09.052.

Hess, M.E. *et al.* (2013) "The fat mass and obesity associated gene (Fto) regulates activity of the dopaminergic midbrain circuitry," *Nature Neuroscience*, 16(8), pp. 1042–1048. doi:10.1038/nn.3449.

Hong, X., Scofield, D.G. and Lynch, M. (2006) "Intron size, abundance, and distribution within untranslated regions of genes," *Molecular Biology and Evolution*, 23(12), pp. 2392–2404. doi:10.1093/molbev/msl111.

Horiuchi, K. *et al.* (2013a) "Identification of Wilms' tumor 1-associating protein complex and its role in alternative splicing and the cell cycle," *Journal of Biological Chemistry*, 288(46), pp. 33292–33302. doi:10.1074/jbc.M113.500397.

Horiuchi, K. *et al.* (2013b) "Identification of Wilms' Tumor 1-associating Protein Complex and Its Role in Alternative Splicing and the Cell Cycle * □ S." doi:10.1074/jbc.M113.500397.

Hsu, Phillip J *et al.* (2017) "Ythdc2 is an N 6-methyladenosine binding protein that regulates mammalian spermatogenesis," *Nature Publishing Group*, 27, pp. 1115–1127. doi:10.1038/cr.2017.99.

Hsu, Phillip J. *et al.* (2017) "Ythdc2 is an N6 -methyladenosine binding protein that regulates mammalian spermatogenesis," *Cell Research*, 27(9), pp. 1115–1127. doi:10.1038/cr.2017.99.

Hu, X. *et al.* (2018) "Increasing the efficiency of CRISPR-Cas9-VQR precise genome editing in rice," *Plant Biotechnology Journal*, 16(1), pp. 292–297. doi:10.1111/pbi.12771.

Hunt, A.G., Xing, D. and Li, Q.Q. (2012) "Plant polyadenylation factors: Conservation and variety in the polyadenylation complex in plants," *BMC Genomics*, 13(1). doi:10.1186/1471-2164-13-641.

Huppertz, I. *et al.* (2014) "iCLIP: Protein-RNA interactions at nucleotide resolution," *Methods*, 65(3), pp. 274–287. doi:10.1016/j.ymeth.2013.10.011.

Irimia, M. and Roy, S.W. (2014) "Origin of spliceosomal introns and alternative splicing," *Cold Spring Harbor Perspectives in Biology*, 6(6), pp. 16071–16072. doi:10.1101/cshperspect.a016071.

Jensen, T.H., Jacquier, A. and Libri, D. (2013) "Dealing with pervasive transcription," *Molecular Cell*, pp. 473–484. doi:10.1016/j.molcel.2013.10.032.

Jia, G. *et al.* (2011) "N6-Methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO," *Nature Chemical Biology*, 7(12), pp. 885–887. doi:10.1038/nchembio.687.

Jiang, S. *et al.* (2018) "Over-expression of Oryza sativa Xrn4 confers plant resistance to virus infection," *Gene*, 639(October 2017), pp. 44–51. doi:10.1016/j.gene.2017.10.004.

Jinek, M. *et al.* (2012) "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity," *Science*, 337(6096), pp. 816–821. doi:10.1126/science.1225829.

Kalyna, M. *et al.* (2012) "Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in Arabidopsis," *Nucleic Acids Research*, 40(6), pp. 2454–2469. doi:10.1093/nar/gkr932.

Kang, H.-J. *et al.* (2014) "A novel protein, Pho92, has a conserved YTH domain and regulates phosphate metabolism by decreasing the mRNA

stability of PHO4 in Saccharomyces cerevisiae," *Biochem. J*, 457, pp. 391–400. doi:10.1042/BJ20130862.

Kappel, C. *et al.* (2015) "Genome-Wide Analysis of PAPS1-Dependent Polyadenylation Identifies Novel Roles for Functionally Specialized Poly(A) Polymerases in Arabidopsis thaliana," *PLoS Genet*, 11(8), p. 1005474. doi:10.1371/journal.pgen.1005474.

Karijolich, J., Kantartzis, A. and Yu, Y.T. (2010) "RNA modifications: a mechanism that modulates gene expression.," *Methods in molecular biology (Clifton, N.J.)*, pp. 1–19. doi:10.1007/978-1-60761-657-3_1.

Karijolich, J. and Yu, Y.T. (2011) "Converting nonsense codons into sense codons by targeted pseudouridylation," *Nature*, 474(7351), pp. 395–399. doi:10.1038/nature10165.

Kasowitz, S.D. *et al.* (2018) "Nuclear m6A reader YTHDC1 regulates alternative polyadenylation and splicing during mouse oocyte development," *PLoS Genetics*, 14(5), pp. 1–28. doi:10.1371/journal.pgen.1007412.

Kastner, B. *et al.* (2019) "Structural insights into nuclear pre-mRNA splicing in higher eukaryotes," *Cold Spring Harbor Perspectives in Biology*, 11(11). doi:10.1101/cshperspect.a032417.

Kato, M. *et al.* (2012) "Cell-free formation of RNA granules: Low complexity sequence domains form dynamic fibers within hydrogels," *Cell*, 149(4), pp. 753–767. doi:10.1016/j.cell.2012.04.017.

Ke, S. *et al.* (2017) "m6A mRNA modifications are deposited in nascent pre-mRNA and are not required for splicing but do specify cytoplasmic turnover," *Genes and Development*, 31(10), pp. 990–1006. doi:10.1101/gad.301036.117.

Kellner, S., Burhenne, J. and Helm, M. (2010) "Detection of RNA modifications," *RNA Biology*. Taylor and Francis Inc., pp. 237–247. doi:10.4161/rna.7.2.11468.

Kennedy, E.M. *et al.* (2016) "Posttranscriptional m6A Editing of HIV-1 mRNAs Enhances Viral Gene Expression," *Cell Host and Microbe*, 19(5), pp. 675–685. doi:10.1016/j.chom.2016.04.002.

Khoddami, V. and Cairns, B.R. (2013) *Identification of direct targets and modified bases of RNA cytosine methyltransferases*, *Nat Biotechnol*. Available at: http://www.nature.com/authors/editorial_policies/license.html#terms (Accessed: April 15, 2020).

Kim, J. *et al.* (2008) "FIONA1 is essential for regulating period length in the Arabidopsis circadian clock," *Plant Cell*, 20(2), pp. 307–319. doi:10.1105/tpc.107.055715.

Kim, M. *et al.* (2004) *The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II*. Available at: www.nature.com/nature (Accessed: April 6, 2020).

Knuckles, P. *et al.* (2018) "Zc3h13/Flacc is required for adenosine methylation by bridging the mRNA-binding factor RbM15/spenito to the m6 a machinery component Wtap/Fl(2)d," *Genes and Development*, 32(5–6), pp. 415–429. doi:10.1101/gad.309146.117.

Krupp, G. *et al.* (1996) "Enzymatic RNA synthesis and RNase P Evolutionary aspects T stem 54 Acceptor stem T stem T loop Accept (~ r stem D loop Variable , oop D loop ~ i ARticodon stem ~~ i : Anticodon loop J ~ Antl Anti codon ANticodon stem Anticodon loop I ~ l j," (Figure 1), pp. 177–180.

Kruse, S. *et al.* (2011) "A novel synthesis and detection method for cap-associated adenosine modifications in mouse mRNA," *Scientific Reports*, 1, pp. 1–6. doi:10.1038/srep00126.

Landick, R. (2006) "The regulatory roles and mechanism of transcriptional pausing," *Biochemical Society Transactions*, 34(6), pp. 1062–1066. doi:10.1042/BST0341062.

Lee, C.M., Cradick, T.J. and Bao, G. (2016) "The Neisseria meningitidis CRISPR-Cas9 System Enables Specific Genome Editing in Mammalian Cells." doi:10.1038/mt.2016.8.

Lence, T. *et al.* (2016) "M6A modulates neuronal functions and sex determination in Drosophila," *Nature*, 540(7632), pp. 242–247. doi:10.1038/nature20568.

Lence, T., Soller, M. and Roignant, J.Y. (2017) "A fly view on the roles and mechanisms of the m6A mRNA modification and its players," *RNA Biology*, 14(9), pp. 1232–1240. doi:10.1080/15476286.2017.1307484.

Li, D. *et al.* (2014) "Genome-Wide Identification, Biochemical Characterization, and Expression Analyses of the YTH Domain-Containing RNA-Binding Protein Family in Arabidopsis and Rice," *Plant Molecular Biology Reporter*, 32(6), pp. 1169–1186. doi:10.1007/s11105-014-0724-2.

Li, F. *et al.* (2014) "Structure of the YTH domain of human YTHDF2 in complex with an m 6 A mononucleotide reveals an aromatic cage for m 6 A recognition," *Nature Publishing Group* [Preprint]. doi:10.1038/cr.2014.153.

Li, J.F. *et al.* (2013) "Multiplex and homologous recombination-mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9," *Nature Biotechnology*, pp. 688–691. doi:10.1038/nbt.2654.

Li, Y. *et al.* (2014) "Transcriptome-wide N6-methyladenosine profiling of rice callus and leaf reveals the presence of tissue-specific competitors involved in selective mRNA modification," *RNA Biology*, 11(9), pp. 1180–1188. doi:10.4161/rna.36281.

Li, Z. *et al.* (2017) "The *Arabidopsis CPSF30-L* gene plays an essential role in nitrate signaling and regulates the nitrate transceptor gene *NRT1.1*," *New Phytologist* [Preprint]. doi:10.1111/nph.14743.

Lianoglou, S. *et al.* (2013) "Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression," *Genes and Development*, 27(21), pp. 2380–2396. doi:10.1101/gad.229328.113.

Lichinchi, G. *et al.* (2016) "Dynamics of Human and Viral RNA Methylation during Zika Virus Infection," *Cell Host and Microbe*, 20(5), pp. 666–673. doi:10.1016/j.chom.2016.10.002.

Linder, B. *et al.* (2015a) "Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome," *Nature Methods*, 12(8), pp. 767–772. doi:10.1038/nmeth.3453.

Linder, B. *et al.* (2015b) "Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome," *Nature Methods*, 12(8), pp. 767–772. doi:10.1038/nmeth.3453.

Liu, J. *et al.* (2014) "A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation.," *Nature chemical biology*, 10(2), pp. 93–95. doi:10.1038/nchembio.1432.

Liu, K. *et al.* (2015) "Structural Basis for the Discriminative Recognition of N 6 -Methyladenosine RNA by the Human YT521-B Homology Domain Family of Proteins *," 290(41), pp. 24902–24913. doi:10.1074/jbc.M115.680389.

Liu, N. *et al.* (2015a) "N6 -methyladenosine-dependent RNA structural switches regulate RNA-protein interactions," *Nature*, 518(7540), pp. 560–564. doi:10.1038/nature14234.

Liu, N. *et al.* (2015b) "N6-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions," *Nature*, 518(7540). doi:10.1038/nature14234.

Liu, S. *et al.* (2020) "m ^{6} A facilitates YTHDF-independent phase separation," *Journal of Cellular and Molecular Medicine*, 24(2), pp. 2070–2072. doi:10.1111/jcmm.14847.

Louder, R.K. *et al.* (2016) "Structure of promoter-bound TFIID and model of human pre-initiation complex assembly," *Nature*, 531(7596), pp. 604–609. doi:10.1038/nature17394.

Lu, W. *et al.* (2018) "N6-Methyladenosine– binding proteins suppress HIV-1 infectivity and viral production," *Journal of Biological Chemistry*, 293(34), pp. 12992–13005. doi:10.1074/jbc.RA118.004215.

Luo, J.H. *et al.* (2020) "Natural variation in RNA m6a methylation and its relationship with translational status," *Plant Physiology*, 182(1), pp. 332–344. doi:10.1104/pp.19.00987.

Luo, Shukun and Tong, L. (2014) "Molecular basis for the recognition of methylated adenines in RNA by the eukaryotic YTH domain," *Proceedings*

of the National Academy of Sciences of the United States of America, 111(38), pp. 13834–13839. doi:10.1073/pnas.1412742111.

Luo, S. and Tong, L. (2014) "Molecular basis for the recognition of methylated adenines in RNA by the eukaryotic YTH domain," *Proceedings of the National Academy of Sciences*, 111(38), pp. 13834–13839. doi:10.1073/pnas.1412742111.

Machnicka, M.A. *et al.* (2013) "MODOMICS: A database of RNA modification pathways - 2013 update," *Nucleic Acids Research*, 41(D1). doi:10.1093/nar/gks1007.

Mali, P. *et al.* (2013) "RNA-Guided Human Genome," *Science*, (February), pp. 823–827.

Mapendano, C.K. *et al.* (2010) "Crosstalk between mRNA 3′ End Processing and Transcription Initiation," *Molecular Cell*, 40(3), pp. 410–422. doi:10.1016/j.molcel.2010.10.012.

Martínez-Pérez, M. *et al.* (2017) "Arabidopsis m6A demethylase activity modulates viral infection of a plant virus and the m6A abundance in its genomic RNAs," *Proceedings of the National Academy of Sciences of the United States of America*, 114(40), pp. 10755–10760. doi:10.1073/pnas.1703139114.

Mauer, J. *et al.* (2017) "Reversible methylation of m6 Am in the 5′ cap controls mRNA stability," *Nature*, 541(7637), pp. 371–375. doi:10.1038/nature21022.

Meyer, K.D. *et al.* (2012) "Comprehensive Analysis of mRNA Methylation Reveals Enrichment in 3' UTRs and Near Stop Codons." doi:10.1016/j.cell.2012.05.003.

Meyer, K.D. *et al.* (2015) "Transcriptome-wide high-throughput deep m6A-seq reveals unique differential m6A methylation patterns between three organs in Arabidopsis thaliana," *Genome Biology*, 16(1), p. 272. doi:10.1186/s13059-015-0839-2.

Miao, Z. *et al.* (2020a) "Evolution of the RNA N 6-Methyladenosine Methylome Mediated by Genomic Duplication 1[OPEN]." doi:10.1104/pp.19.00323.

Miao, Z. *et al.* (2020b) "Evolution of the RNA N6-methyladenosine methylome mediated by genomic duplication," *Plant Physiology*, 182(1), pp. 345–360. doi:10.1104/pp.19.00323.

Miao, Z. and Westhof, E. (2017) "RNA Structure: Advances and Assessment of 3D Structure Prediction," *Annu. Rev. Biophys*, 46, pp. 483–503. doi:10.1146/annurev-biophys.

Millevoi, S. *et al.* (2006) "An interaction between U2AF 65 and CF Im links the splicing and 3′ end processing machineries," *The EMBO Journal*, 25(20), pp. 4854–4864. doi:10.1038/sj.emboj.7601331.

Miransari, M. and Smith, D.L. (2014) "Plant hormones and seed germination," *Environmental and Experimental Botany*, 99, pp. 110–121. doi:10.1016/j.envexpbot.2013.11.005.

Nekrasov, V. *et al.* (2013) "Targeted mutagenesis in the model plant Nicotiana benthamiana using Cas9 RNA-guided endonuclease," *Nature Biotechnology*, 31(8), pp. 691–693. doi:10.1038/nbt.2655.

Neuman, K.C. *et al.* (2003) *Second, pausing repre-1997b; Palangat and Landick, 2001; Reeder and Hawley, sents the first event in transcriptional termination*, *Cell*. Available at: http://www. (Accessed: March 26, 2020).

Nierman, J., Wiggs, W.C. and Neff, J.L. (1981) *A Direct Effect of Guanosine Tetraphosphate on Pausing of Escherichia coli RNA Polymerase during RNA Chain Elongation**, *THE JOURNAL OF BIOLOGICAL CHEMISTRY Printed in U.S.A.*

Nilsen, T.W. (2007) "RNA 1997-2007: A Remarkable Decade of Discovery," *Molecular Cell*. Cell Press, pp. 715–720. doi:10.1016/j.molcel.2007.11.017.

Nishizawa, Y. *et al.* (2018) "Oncogene c-Myc promotes epitranscriptome m6A reader YTHDF1 expression in colorectal cancer," *Oncotarget*, 9(7), pp. 7476–7486. doi:10.18632/oncotarget.23554.

Nudler, E. (1999) “Transcription elongation: Structural basis and mechanisms,” *Journal of Molecular Biology*, 288(1), pp. 1–12. doi:10.1006/jmbi.1999.2641.

Ok, S.H. *et al.* (2005) “Novel CIPK1-associated proteins in Arabidopsis contain an evolutionarily conserved C-terminal region that mediates nuclear localization.,” *Plant physiology*, 139(1), pp. 138–50. doi:10.1104/pp.105.065649.

Ougland, R. *et al.* (2004) “AlkB restores the biological function of mRNA and tRNA inactivated by chemical methylation,” *Molecular Cell*, 16(1), pp. 107–116. doi:10.1016/j.molcel.2004.09.002.

Parker, M.T. *et al.* (2020) “Nanopore direct RNA sequencing maps the complexity of Arabidopsis mRNA processing and m 6 A modification,” *eLife* [Preprint]. doi:10.7554/eLife.49658.

Patil, D.P. *et al.* (2016) “m6A RNA methylation promotes XIST-mediated transcriptional repression,” *Nature*, 537(7620), pp. 369–373. doi:10.1038/nature19342.

Patil, D.P., Pickering, B.F. and Jaffrey, S.R. (2017) “Reading m6A in the Transcriptome: M6A-Binding Proteins,” *Trends in Cell Biology*, 28(2), pp. 113–127. doi:10.1016/j.tcb.2017.10.001.

Pendleton, K.E. *et al.* (2017) “The U6 snRNA m6A methyltransferase METTL16 regulates SAM synthetase intron retention,” 169(5), pp. 824–835. doi:10.1016/j.cell.2017.05.003.The.

Perry, R.P. and Kelley, D.E. (1974) “Existence of methylated messenger RNA in mouse L cells,” *Cell*, 1(1), pp. 37–42. doi:10.1016/0092-8674(74)90153-6.

Ping, X.L. *et al.* (2014) “Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase,” *Cell Research*, 24(2), pp. 177–189. doi:10.1038/cr.2014.3.

Pontier, D. *et al.* (2019) “The m6A pathway protects the transcriptome integrity by restricting RNA chimera formation in plants,” *Life Science Alliance*, 2(3). doi:10.26508/lsa.201900393.

Proudfoot, N.J. and Brownlee, G.G. (1976) “3′ Non-coding region sequences in eukaryotic messenger RNA,” *Nature*, 263(5574), pp. 211–214. doi:10.1038/263211a0.

Rahman, R. *et al.* (2018) “Identification of RNA-binding protein targets with HyperTRIBE,” *Nature Protocols*, 13(8), pp. 1829–1849. doi:10.1038/s41596-018-0020-y.

Ramanathan, A., Robb, G.B. and Chan, S.-H. (2016) “SURVEY AND SUMMARY mRNA capping: biological functions and applications,” *Nucleic Acids Research*, 44(16), pp. 7511–7526. doi:10.1093/nar/gkw551.

Robertus, J.D. *et al.* (1974) “Structure of yeast phenylalanine tRNA at 3 Å resolution,” *Nature*, 250(5467), pp. 546–551. doi:10.1038/250546a0.

Roeder, R. (1996) “The role of general initiation factors in transcription by RNA polymerase II,” *Trends in Biochemical Sciences*, 21(9), pp. 327–335. doi:10.1016/s0968-0004(96)10050-5.

Ross, S.D., Bruno, J.J. and Petersen, R.C. (1963) “Nucleophilic Displacement Reactions in Aromatic Systems. VIII. Rates and Mechanism for the Reaction of 2,4-Dinitrochlorobenzene with 1,4-Diazabicyclo[2.2.2]octane in Acetonitrile,” *Journal of the American Chemical Society*, 85(24), pp. 3999–4003. doi:10.1021/ja00907a020.

Roundtree, I.A., Evans, M.E., *et al.* (2017) “Dynamic RNA Modifications in Gene Expression Regulation,” *Cell*. doi:10.1016/j.cell.2017.05.045.

Roundtree, I.A., Luo, G.Z., *et al.* (2017) “YTHDC1 mediates nuclear export of N6-methyladenosine methylated mRNAs,” *eLife*, 6, pp. 1–28. doi:10.7554/eLife.31311.

Roundtree, I.A. and He, C. (2016) “Nuclear m6A Reader YTHDC1 Regulates mRNA Splicing,” *Trends in Genetics*, 32(6), pp. 320–321. doi:10.1016/j.tig.2016.03.006.

Rubio, R.M. *et al.* (2018) “RNA m 6 A modification enzymes shape innate responses to DNA by regulating interferon ß,” *Genes and Development*, 32(23–24), pp. 1472–1484. doi:10.1101/gad.319475.118.

Růžička, K. *et al.* (2017) "Identification of factors required for m6A mRNA methylation in Arabidopsis reveals a role for the conserved E3 ubiquitin ligase HAKAI," *New Phytologist*, 215(1), pp. 157–172. doi:10.1111/nph.14586.

Sammeth, M., Foissac, S. and Guigó, R. (2008) "A General Definition and Nomenclature for Alternative Splicing Events," *PLoS Comput Biol*, 4(8), p. 1000147. doi:10.1371/journal.pcbi.1000147.

Schomburg, F.M. *et al.* (2001a) *FPA , a Gene Involved in Floral Induction in Arabidopsis, Encodes a Protein Containing RNA-Recognition Motifs*, *The Plant Cell*. Available at: www.plantcell.org.

Schomburg, F.M. *et al.* (2001b) "FPA , a Gene Involved in Floral Induction in Arabidopsis, Encodes a Protein Containing RNA-Recognition Motifs," *The Plant Cell*, 13, pp. 1427–1436. Available at: www.plantcell.org (Accessed: July 18, 2021).

Schwartz, S. *et al.* (2013) "High-Resolution mapping reveals a conserved, widespread, dynamic mRNA methylation program in yeast meiosis," *Cell*, 155(6), pp. 1409–1421. doi:10.1016/j.cell.2013.10.047.

Schwartz, S., Mumbach, Maxwell R., *et al.* (2014) "Perturbation of m6A writers reveals two distinct classes of mRNA methylation at internal and 5' sites," *Cell Reports*, 8(1). doi:10.1016/j.celrep.2014.05.048.

Schwartz, S., Mumbach, Maxwell R, *et al.* (2014) "Perturbation of m6A writers reveals two distinct classes of mRNA methylation at internal and 5' sites," *Cell Rep*, 8(1), pp. 284–296. doi:10.1016/j.celrep.2014.05.048.

Scutenaire, J., Deragon, J.-M., *et al.* (2018) "The YTH Domain Protein ECT2 is an m6A Reader Required for Normal Trichome Branching in Arabidopsis," *The Plant Cell*, p. tpc.00854.2017. doi:10.1105/tpc.17.00854.

Scutenaire, J., Deragon, J.M., *et al.* (2018) "The YTH domain protein ECT2 is an m6A reader required for normal trichome branching in arabidopsis," *Plant Cell*, 30(5), pp. 986–1005. doi:10.1105/tpc.17.00854.

Seraphin, B. (1995) *Sm and Sm-like proteins belong to a large family: identification of proteins of the U6 as well as the UI, U2, U4 and U5 snRNPs*,

The EMBO Journal. Available at: https://watermark.silverchair.com/55-1-49.pdf?token=AQECAHi208BE49Ooan9kkhW_Ercy7Dm3ZL_9Cf3qfKAc485ysgAAAncwggJzBgkqhkiG9w0BBwagggJkMIICYAIBADCCAlkGCSqGSIb3DQEHATAeBglghkgBZQMEAS4wEQQMV2uEsMHhcN6u5ChIAgEQgIICKpeDlr1PoG0MfOQX9w386gIDAAatTqjcqelRm4b1QoEHqL7 (Accessed: April 10, 2020).

Sheets, M.D., Ogg, S.C. and Wickens, M.P. (1990) "Point mutations in AAUAAA and the poly (A) addition site: Effects on the accuracy and efficiency of cleavage and polyadenylation in vitro," *Nucleic Acids Research*, 18(19), pp. 5799–5805. doi:10.1093/nar/18.19.5799.

Shen, L. *et al.* (2016) "N6-Methyladenosine RNA Modification Regulates Shoot Stem Cell Fate in Arabidopsis," *Developmental Cell*, 38(2), pp. 186–200. doi:10.1016/j.devcel.2016.06.008.

Shen, Y. *et al.* (2008) "Genome level analysis of rice mRNA 3'-end processing signals and alternative polyadenylation," *Nucleic Acids Research*, 36(9), pp. 3150–3161. doi:10.1093/nar/gkn158.

Sheth, U. and Parker, R. (2003) "Decapping and decay of messenger RNA occur in cytoplasmic processing bodies," *Science*, 300(5620), pp. 805–808. doi:10.1126/science.1082320.

Shi, H. *et al.* (2017) "YTHDF3 facilitates translation and decay of N 6-methyladenosine-modified RNA," *Nature Publishing Group*, 27, pp. 315–328. doi:10.1038/cr.2017.15.

Shi, H. *et al.* (2018) "m6A facilitates hippocampus-dependent learning and memory through YTHDF1," *Nature*, pp. 249–253. doi:10.1038/s41586-018-0666-1.

Shimba, S. *et al.* (1995) *Accurate and efficient N-6-adenosine methylation in spliceosomal U6 small nuclear RNA by HeLa cell extract in vitro*, *Nucleic Acids Research*.

Śledź, P. and Jinek, M. (2016) "Structural insights into the molecular mechanism of the m6A writer complex," *eLife*, 5(September), pp. 1–16. doi:10.7554/eLife.18434.

Soldano, A. *et al.* (2020) "The m 6 A reader Ythdf restricts axonal growth in Drosophila through target selection modulation of the Fragile X mental retardation protein." doi:10.1101/2020.03.04.976886.

Song, P. *et al.* (2021) "Arabidopsis N6-methyladenosine reader CPSF30-L recognizes FUE signals to control polyadenylation site choice in liquid-like nuclear bodies," *Molecular Plant*, 14(4), pp. 571–587. doi:10.1016/j.molp.2021.01.014.

Staley, J.P. and Guthrie, C. (1999) *An RNA Switch at the 5' Splice Site Requires ATP and the DEAD Box Protein Prp28p*, *Molecular Cell*. Available at: https://ublearns.buffalo.edu/bbcswebdav/pid-4941732-dt-content-rid-22635287_1/courses/2191_12832_PsC/1999 Staley and Guthrie Prp28.pdf (Accessed: April 10, 2020).

Stoilov, P., Rafalska, I. and Stamm, S. (2002) "YTH: A new domain in nuclear proteins," *Trends in Biochemical Sciences*, 27(10), pp. 495–497. doi:10.1016/S0968-0004(02)02189-8.

Su, S.H. *et al.* (2017) "Molecular Mechanisms of Root Gravitropism," *Current Biology*, 27(17), pp. R964–R972. doi:10.1016/j.cub.2017.07.015.

Sunkar, R. *et al.* (2007) "Small RNAs as big players in plant abiotic stress responses and nutrient deprivation," *Trends in Plant Science*, 12(7), pp. 301–309. doi:10.1016/j.tplants.2007.05.001.

Tanabe, A. *et al.* (2016) "RNA helicase YTHDC2 promotes cancer metastasis via the enhancement of the efficiency by which HIF-1α mRNA is translated," *Cancer Letters*, 376(1), pp. 34–42. doi:10.1016/j.canlet.2016.02.022.

Tang, C. *et al.* (2017) "ALKBH5-dependent m6A demethylation controls splicing and stability of long 3'-UTR mRNAs in male germ cells," *Proceedings of the National Academy of Sciences of the United States of America*, 115(2), pp. E325–E333. doi:10.1073/pnas.1717794115.

Taoka, M. *et al.* (2018) "Landscape of the complete RNA chemical modifications in the human 80S ribosome," *Nucleic Acids Research*, 46(18), pp. 9289–9298. doi:10.1093/nar/gky811.

Thalhammer, A. *et al.* (2011) "Human AlkB Homologue 5 Is a Nuclear 2-Oxoglutarate Dependent Oxygenase and a Direct Target of Hypoxia-Inducible Factor 1a (HIF-1a)," *PLoS ONE*, 6(1), p. 16210. doi:10.1371/journal.pone.0016210.

Theler, D. *et al.* (2014) "Solution structure of the YTH domain in complex with N6-methyladenosine RNA: A reader of methylated RNA," *Nucleic Acids Research*, 42(22). doi:10.1093/nar/gku1116.

Thomas, M.C. and Chiang, C.M. (2006) "The general transcription machinery and general cofactors.," *Critical reviews in biochemistry and molecular biology*, 41(3), pp. 105–178. doi:10.1080/10409230600648736.

Thomas, P.E. *et al.* (2012) "Genome-wide control of polyadenylation site choice by CPSF30 in arabidopsisC W OA," *Plant Cell*, 24(11), pp. 4376–4388. doi:10.1105/tpc.112.096107.

Tian, B. *et al.* (2005) "A large-scale analysis of mRNA polyadenylation of human and mouse genes," *Nucleic Acids Research*, 33(1), pp. 201–212. doi:10.1093/nar/gki158.

Tian, B. and Manley, J.L. (2016) "Alternative polyadenylation of mRNA precursors," *Nature Reviews Molecular Cell Biology*, pp. 18–30. doi:10.1038/nrm.2016.116.

Unrau, P. and Bartel, D. (1998) "RNA-catalysed nucleotide synthesis," *Nature*, (1), pp. 430–439. Available at: https://linkinghub.elsevier.com/retrieve/pii/B012226765600013X.

Vandivier, L.E. *et al.* (2016) "The Conservation and Function of RNA Secondary Structure in Plants," *Annual Review of Plant Biology*, 67(1), pp. 463–488. doi:10.1146/annurev-arplant-043015-111754.

Venkataraman, K., Brown, K.M. and Gilmartin, G.M. (2005) "Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition," *Genes and Development*, 19(11), pp. 1315–1327. doi:10.1101/gad.1298605.

Wahl, M.C., Will, C.L. and Lührmann, R. (2009) "The Spliceosome: Design Principles of a Dynamic RNP Machine," *Cell*, pp. 701–718. doi:10.1016/j.cell.2009.02.009.

Wan, Y. *et al.* (2015) "Transcriptome-wide high-throughput deep m6A-seq reveals unique differential m6A methylation patterns between three organs in Arabidopsis thaliana," *Genome Biology*, 16(1), p. 272. doi:10.1186/s13059-015-0839-2.

Wang, E.T. *et al.* (2008) "Alternative Isoform Regulation in Human Tissue Transcriptomes HHS Public Access," *Nature*, 456(7221), pp. 470–476. doi:10.1038/nature07509.

Wang, N. *et al.* (2014) "Genome-wide identification of members in the YTH domain-containing RNA-binding protein family in apple and expression analysis of their responsiveness to senescence and abiotic stresses," *Gene*, 538(2), pp. 292–305. doi:10.1016/j.gene.2014.01.039.

Wang, N., Guo, T., Sun, X., *et al.* (2017) "Functions of two Malus hupehensis (Pamp.) Rehd. YTPs (MhYTP1 and MhYTP2) in biotic- and abiotic-stress responses," *Plant Science*, 261(March), pp. 18–27. doi:10.1016/j.plantsci.2017.05.002.

Wang, N., Guo, T., Wang, P., *et al.* (2017) "MhYTP1 and MhYTP2 from Apple Confer Tolerance to Multiple Abiotic Stresses in Arabidopsis thaliana," *Frontiers in Plant Science*, 8(August), pp. 1–9. doi:10.3389/fpls.2017.01367.

Wang, X. *et al.* (2013) "N6-methyladenosine-dependent regulation of messenger RNA stability," *Nature*, 505(7481), pp. 117–120. doi:10.1038/nature12730.

Wang, X. *et al.* (2015) "N^6-methyladenosine modulates messenger RNA translation efficiency," *Cell*, 161(6), pp. 1388–1399. doi:10.1016/j.cell.2015.05.014.

Wang, X. *et al.* (2016) "Structural basis of N6-adenosine methylation by the METTL3-METTL14 complex," *Nature*, 534(7608), pp. 575–578. doi:10.1038/nature18298.

Wang, Y. *et al.* (2014) "N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells," *Nature Cell Biology*, 16(2), pp. 191–198. doi:10.1038/ncb2902.

WANG, Y. *et al.* (2015) "Mechanism of alternative splicing and its regulation," *Biomedical Reports*, 3(2), pp. 152–158. doi:10.3892/br.2014.407.

Weber, E. *et al.* (2011) "A Modular Cloning System for Standardized Assembly of Multigene Constructs," *PLoS ONE*, 6(2), p. 16765. doi:10.1371/journal.pone.0016765.

Wei, C.M., Gershowitz, A. and Moss, B. (1975) "N6, O2'-dimethyladenosine a novel methylated ribonucleoside next to the 5' terminal of animal cell and virus mRNAs," *Nature*, 257(5523), pp. 251–253. doi:10.1038/257251a0.

Wei, C.M., Gershowitz, A. and Moss, B. (1976) "5′-Terminal and Internal Methylated Nucleotide Sequences in Hela Cell mRNA," *Biochemistry*, 15(2), pp. 397–401. doi:10.1021/bi00647a024.

Wei, J. *et al.* (2018) "Differential m 6 A, m 6 A m , and m 1 A Demethylation Mediated by FTO in the Cell Nucleus and Cytoplasm," *Molecular Cell*, 71(6), pp. 973-985.e5. doi:10.1016/j.molcel.2018.08.011.

Wei, L.-H.H. *et al.* (2018) "The m6A Reader ECT2 Controls Trichome Morphology by Affecting mRNA Stability in Arabidopsis," *The Plant Cell*, 16(5), p. tpc.00934.2017. doi:10.1105/tpc.17.00934.

Weng, Y.L. *et al.* (2018a) "Epitranscriptomic m6A Regulation of Axon Regeneration in the Adult Mammalian Nervous System," *Neuron*, 97(2), pp. 313-325.e6. doi:10.1016/j.neuron.2017.12.036.

Weng, Y.L. *et al.* (2018b) "Epitranscriptomic m6A Regulation of Axon Regeneration in the Adult Mammalian Nervous System," *Neuron*, 97(2), pp. 313-325.e6. doi:10.1016/j.neuron.2017.12.036.

West, S., Gromak, N. and Proudfoot, N.J. (2004) "Human 5′ → 3′ exonuclease Xm2 promotes transcription termination at co-transcriptional cleavage sites," *Nature*, 432(7016), pp. 522–525. doi:10.1038/nature03035.

West, S. and Proudfoot, N.J. (2009) “Transcriptional Termination Enhances Protein Expression in Human Cells,” *Molecular Cell*, 33(3), pp. 354–364. doi:10.1016/j.molcel.2009.01.008.

Wilkinson, M.E., Charenton, C. and Nagai, K. (2019) “Annual Review of Biochemistry RNA Splicing by the Spliceosome.” doi:10.1146/annurev-biochem-091719.

Wojtas, M.N. *et al.* (2017) “Regulation of m6A Transcripts by the 3′→5′ RNA Helicase YTHDC2 Is Essential for a Successful Meiotic Program in the Mammalian Germline,” *Molecular Cell*, 68(2), pp. 374-387.e12. doi:10.1016/j.molcel.2017.09.021.

Woodcock, C.B. *et al.* (2020) “Biochemical and structural basis for YTH domain of human YTHDC1 binding to methylated adenine in DNA,” *Nucleic Acids Research*, 48(18), pp. 10329–10341. doi:10.1093/nar/gkaa604.

Wu, J., Peled-Zehavi, H. and Galili, G. (2020) “The m6A reader ECT2 post-transcriptionally regulates proteasome activity in Arabidopsis,” *New Phytologist*, 228(1), pp. 151–162. doi:10.1111/nph.16660.

Wu, R. *et al.* (2018) “An efficient CRISPR vector toolbox for engineering large deletions in Arabidopsis thaliana,” *Plant Methods*, 14(1), pp. 1–9. doi:10.1186/s13007-018-0330-7.

Wyers, F. *et al.* (2005) “Cryptic Pol II Transcripts Are Degraded by a Nuclear Quality Control Pathway Involving a New Poly(A) Polymerase,” *Cell*, 121, pp. 725–737. doi:10.1016/j.cell.2005.04.030.

Xiao, W. *et al.* (2016) “Nuclear m 6 A Reader YTHDC1 Regulates mRNA Splicing.” doi:10.1016/j.molcel.2016.01.012.

Xin, Y. and Olson, W.K. (2008) “BPS: a database of RNA base-pair structures,” *Nucleic Acids Research*, 37, pp. 83–88. doi:10.1093/nar/gkn676.

Xu, C., Wang, X., *et al.* (2014) “Structural basis for selective binding of m6A RNA by the YTHDC1 YTH domain,” *Nature Chemical Biology*, 10(11), pp. 927–929. doi:10.1038/nchembio.1654.

Xu, C., Liu, K., *et al.* (2014) “Structures of human ALKBH5 demethylase reveal a unique binding mode for specific single-stranded N6-

methyladenosine RNA demethylation," *Journal of Biological Chemistry*, 289(25), pp. 17299–17311. doi:10.1074/jbc.M114.550350.

Xu, C. *et al.* (2015) "Structural basis for the discriminative recognition of N6-Methyladenosine RNA by the human YT521-B homology domain family of proteins," *Journal of Biological Chemistry*, 290(41), pp. 24902–24913. doi:10.1074/jbc.M115.680389.

Xue, B. *et al.* (2010) "PONDR-FIT: A Meta-Predictor of Intrinsically Disordered Amino Acids," *Biochim Biophys Acta*, 1804(4), pp. 996–1010. doi:10.1016/j.bbapap.2010.01.011.

Yacoubi, B. el, Bailly, M. and de Crécy-Lagard, V. (2012) "Biosynthesis and Function of Posttranscriptional Modifications of Transfer RNAs." doi:10.1146/annurev-genet-110711-155641.

Yoshida, T., Mogami, J. and Yamaguchi-Shinozaki, K. (2014) "ABA-dependent and ABA-independent signaling in response to osmotic stress in plants," *Current Opinion in Plant Biology*, 21, pp. 133–139. doi:10.1016/j.pbi.2014.07.009.

Yue, Y. *et al.* (2018) "VIRMA mediates preferential m 6 A mRNA methylation in 3′UTR and near stop codon and associates with alternative polyadenylation," *Cell Discovery*, 4, p. 10. doi:10.1038/s41421-018-0019-0.

Zaccara, S., Ries, R.J. and Jaffrey, S.R. (2019) "Reading, writing and erasing mRNA methylation," *Nature Reviews Molecular Cell Biology*, pp. 608–624. doi:10.1038/s41580-019-0168-5.

Zhang, M. *et al.* (2020) "Roles of N6-Methyladenosine (m6A) in Stem Cell Fate Decisions and Early Embryonic Development in Mammals," *Frontiers in Cell and Developmental Biology*, 0, p. 782. doi:10.3389/FCELL.2020.00782.

Zhang, M. *et al.* (2022) "Two zinc finger proteins with functions in m6A writing interact with HAKAI," *Nature Communications*, 13(1). doi:10.1038/s41467-022-28753-3.

Zhang, S. *et al.* (2017) "The m 6 A Demethylase ALKBH5 Maintains Tumorigenicity of Glioblastoma Stem-Like Cells by Sustaining FOXM1

Expression and Cell Proliferation Program HHS Public Access," *Cancer Cell*, 31(4), pp. 591–606. doi:10.1016/j.ccell.2017.02.013.

Zhang, X. *et al.* (2019) "Structural insights into FTO's catalytic mechanism for the demethylation of multiple RNA substrates," *Proceedings of the National Academy of Sciences of the United States of America*, 116(8), pp. 2919–2924. doi:10.1073/pnas.1820574116.

Zhao, J., Hyman, L. and Moore, C. (1999) *Formation of mRNA 3 Ends in Eukaryotes: Mechanism, Regulation, and Interrelationships with Other Steps in mRNA Synthesis*, *MICROBIOLOGY AND MOLECULAR BIOLOGY REVIEWS*.

Zhao, Y.L. *et al.* (2019) "Understanding m 6 A Function Through Uncovering the Diversity Roles of YTH Domain-Containing Proteins," *Molecular Biotechnology*, 61(5), pp. 355–364. doi:10.1007/s12033-018-00149-z.

Zheng, G. *et al.* (2013) "ALKBH5 Is a Mammalian RNA Demethylase that Impacts RNA Metabolism and Mouse Fertility," *Molecular Cell*, 49(1), pp. 18–29. doi:10.1016/j.molcel.2012.10.015.

Zhong, S. *et al.* (2008) "MTA Is an Arabidopsis Messenger RNA Adenosine Methylase and Interacts with a Homolog of a Sex-Specific Splicing Factor," *the Plant Cell Online*, 20(5), pp. 1278–1288. doi:10.1105/tpc.108.058883.

Zhong, S. *et al.* (2013) "Single-base resolution methylomes of tomato fruit development reveal epigenome modifications associated with ripening," *Nature Biotechnology*, 31(2), pp. 154–159. doi:10.1038/nbt.2462.

Zhou, C. *et al.* (2018) "Identification and analysis of adenine N 6-methylation sites in the rice genome," *Nature Plants*, pp. 554–563. doi:10.1038/s41477-018-0214-x.

Zhou, H. *et al.* (2016) "M1A and m1G disrupt A-RNA structure through the intrinsic instability of Hoogsteen base pairs," *Nature Structural and Molecular Biology*, 23(9), pp. 803–810. doi:10.1038/nsmb.3270.

Zhou, J. *et al.* (2015) "Dynamic m 6 A mRNA methylation directs translational control of heat shock response HHS Public Access," *Nature*, 526(7574), pp. 591–594. doi:10.1038/nature15377.

Zhou, L., Tian, S. and Qin, G. (2019) "RNA methylomes reveal the m6A-mediated regulation of DNA demethylase gene SlDML2 in tomato fruit ripening," *Genome Biology*, 20(1). doi:10.1186/s13059-019-1771-7.

Zhu, T. *et al.* (2014) "Crystal structure of the YTH domain of YTHDF2 reveals mechanism for recognition of N6-methyladenosine," *Nature Publishing Group*, 24. doi:10.1038/cr.2014.152.

Zsögön, A. *et al.* (2018) "De novo domestication of wild tomato using genome editing," *Nature Biotechnology*, 36(12), pp. 1211–1216. doi:10.1038/nbt.4272.