Transcriptome Dynamics in the *Arabidopsis thaliana* Male Germ Unit: From Mature Pollen Grain to Pollen Tube

Chandra Shekhar Misra

Dissertation presented to obtain the Ph.D degree in Plant Sciences

Instituto de Tecnologia Química e Biológica António Xavier | Universidade Nova de Lisboa

Oeiras,
July, 2021

UNIVERSIDADE NOVA DE LISBOA

Transcriptome Dynamics in the *Arabidopsis thaliana* Male Germ Unit: From Mature Pollen Grain to Pollen Tube

Chandra Shekhar Misra

Dissertation presented to obtain the Ph.D degree in Plant Sciences

Instituto de Tecnologia Química e Biológica António Xavier | Universidade Nova de Lisboa

Research work coordinated by: INSTITUTO GULBENKIAN DE CIÊNCIA

Oeiras, July, 2021

UNIVERSIDADE NOVA DE LISBOA

Financial support from Fundação para a Ciência e a Tecnologia, Portugal, through grant PD/BD/114362/2016 awarded to Chandra Shekhar Misra

Supervisor: Dr. Jörg Dieter Becker

Declaration

The author of this PhD thesis, Chandra Shekhar Misra, hereby declares that all data and results herein contained have been an outcome of my own work conducted under the guidance of Doctor Jörg Dieter Becker at the Instituto Gulbenkian de Ciência in Oeiras, Portugal. Chandra Shekhar Misra was supported by PhD fellowship PD/BD/114362/2016 awarded by Fundação para a Ciência e Tecnologia (FCT), FCT ERA-CAPS project EVOREPRO (ERA-CAPS-0001-2014) and FCT project PTDC/BIA-FBT/28484/2017. All relevant contributions and help by collaborators in each chapter/publication have been duly acknowledged.

The following manuscripts are either published or under preparation:

- Misra, C.S., Santos, M.R., Rafael-Fernandes, M., Martins, N.P., Monteiro, M. and Becker, J.D., 2019. Transcriptomics of Arabidopsis sperm cells at single-cell resolution. Plant reproduction, 32(1), 29-38.
- Misra C.S., Barros, P.M., Sousa, A.G., Becker, J.D. Cell-type specific alternative splicing in the Arabidopsis germline (*Under preparation*).
- Misra C.S., Sousa, A.G., Becker, J.D. Transcriptome reprogramming in the Arabidopsis male germline during the progamic phase (*Under preparation*).

Acknowledgement

The journey of a thousand miles begins with a single step, and my journey of this enriching PhD experience began five years ago. This journey was not of mine alone but of so many people without whose help, support, encouragement and motivation, I would not have been able to finish this task that I had embarked upon to take. First and foremost, I would like to express my immense gratitude to my supervisor, Dr. Jörg Becker for giving me an opportunity to work on this very interesting research project. I thank him for the timely and valuable guidance, immense motivation, untiring and constant help and most importantly his positive attitude towards my capabilities which made the whole process of achievement of this goal a highly rewarding, enjoyable, enriching and stimulating experience. Many a times, when experiments failed, he always encouraged and motivated me to not only perform and not be scared of making mistakes but also to learn from the mistakes and march ahead. Despite my limited computational knowledge, he gave me the full freedom to learn at my pace, collaborate if necessary, which made the whole PhD experience worthwhile. I will always be grateful to him for kindly accepting me in his group and allowing me to work on this challenging, yet very interesting project. I don't have enough words to explain how his support during some of the most challenging times in the last two years of my PhD allowed me to remain focussed and keep the momentum going.

When I started my PhD, I did not have any background on pollen or Arabidopsis. I learnt everything from scratch, and I can't thank enough Dr. Leonor Boavida, then postdoc in our lab and now Assistant Professor at Purdue University, who took time out of her busy schedule to teach me the nuances of how to navigate through research. She helped me with my initial experiments that I was still trying to optimize, particularly pollen germination and pollen tube growth assays. The interesting conversation during lunch

breaks and or some late evening microscopy experiments that she was finishing, were very useful later for my own project.

My PhD project was supposed to have a lot of bioinformatics, which I realised very early. Learning it from scratch, specially without prior knowledge, was a daunting task. But I would like to thank Dr. Ann-Cathrin Linder, postdoc in our lab, for taking up this challenge of teaching me the analysis of RNA-seq and bisulphite data. She was my go-to person whenever I had any issues during the analysis.

I always feel very fortunate to have such great colleagues with whom I had the opportunity to learn so much. I am grateful to Dr. Paulo Navarro Costa for so many intellectual discussions over my project. He took out time from his busy schedule to provide all the valuable feedback that he gave me, which later helped me in shaping my thesis. He also provided some constructive feedback during the lab meetings that opened up more ideas for my PhD. Apart from scientific discussion, he always made sure to check if I was fine specially during the last two years which have been personally very challenging.

I would like to thank my thesis committee members Professor Dr. Margarida Oliveira and Dr. Ivo Chelo for their continuous feedback and support during the entire duration of my PhD.

During my PhD, I had to depend on almost all the facilities at IGC. One of them I would have annoyed the most was the Genomics Facility. I would like to thank Sara and Joao Sobral during the initial years of my PhD for being so patient with all my sequencing requests. Later on, new members joined the team and I will be forever grateful to Susana and Cathy for being so kind enough to have time and discuss with me the very peculiar needs of my project. Whenever I reached out to them, they were always ready to accommodate my request and I am really thankful to both of them. Of course,

all this could not be possible without the support of Ricardo, who is the head of the facility.

I would like to also thank Vera, head of the Plant Facility for taking so good care of our plants. Last one year especially, since we were restricted to our home due to the pandemic and plants still needed care and maintenance. She was going to the IGC regularly to make sure, our plants don't die. She was not only a plant technician, but a very close friend of mine too. If I may say, she was perhaps the only person in IGC, I could share everything with- from failed experiments to my personal stories. She was the one person, over the last two years I could always count on for her moral and emotional support during the times I needed the most.

I would like to thank the Flow Cytometry Facility for being so supportive and helpful. Special thanks to Marta and Mariana for being there for so many insightful discussions for optimization of protocols for sorting sperm cells from growing pollen tubes. This was perhaps one of the most challenging tasks of my PhD, but their fruitful discussions and at times patience made sure I was able to sort even from such a small number of cells from growing pollen tubes. Later on, the help by Ana, Denise could also not be forgotten. All the members of FACS facility were ready to help me, even on days when the sorting was going very late in the evening.

I would like to thank all my lab colleagues: Mario, Anton, Sonia, Carmen, Clement, Miguel, Joana and Rion for all the discussions and help over the last years. I hope I have not forgotten anyone.

I would like to thank Elio for formally including me into the IGC´s PhD program. I cannot forget the help and assistance provided by Ana Aranda during the last four years of my PhD that helped me to remain focus on my main goal of obtaining the PhD degree. Her exemplary support especially in dealing and navigating the notorious bureaucratic process made my stay in

Portugal as well as in IGC as convenient as possible. I can’t stress enough how she was always a call or an email away. I would like to also thank Jorge Carneiro for being so supportive during the last part of my PhD specially formalising all the process for my PhD extension and making sure I continue to receive all benefits.

I would like to thank my friends Anurag and Devesh for continuously motivating me, and for all the late night talks and discussion over the weekends, both scientific and otherwise.

Finally, last but not the least, I will always be indebted to my parents, who despite all the challenges, gave me full freedom to pursue my career in science miles aways from them. My Grandmother, without whose blessing, love and affection I would never have had acheived this milestone of finishing the PhD.

“Ask the right questions, and nature will open the doors to her secrets”- *C.V. Raman*

Resumo

O desenvolvimento do pólen em plantas com flor começa no interior das anteras, onde as células progenitoras dos micrósporos sofrem meiose para formar micrósporos haplóides. O micrósporo haplóide sofre divisão assimétrica (mitose do pólen I) para formar uma célula vegetativa maior, que abriga uma célula geradora menor, que sofre uma segunda divisão mitótica (mitose do pólen II) para formar duas células espermáticas. Enquanto a célula vegetativa impulsiona o crescimento do tubo polínico, as células espermáticas são transportadas passivamente dentro dos tubos polínicos até ao saco embrionário para a fertilização dupla. O desenvolvimento dos gametófitos masculinos em plantas com flores tem sido estudado extensivamente nas últimas décadas. Abordagens transcriptómicas têm sido utilizadas para analisar não apenas a dinâmica da expressão génica durante o desenvolvimento do pólen, mas também ao nível da linha germinativa masculina em Arabidopsis, arroz, milho e tomate. Até agora, uma análise sistemática dos transcriptomas de células espermáticas e núcleos vegetativos de pólen maduro, e de tubos polínicos que cresceram através do pistilo de Arabidopsis nunca foi realizada, em parte devido a limitações técnicas no isolamento e processamento dessas amostras. Esse estudo, no entanto, promete identificar as mudanças transcricionais desencadeadas nas células espermáticas durante as interações pólen-pistilo, potencialmente importantes para o sucesso da fertilização do óvulo e da célula central.

Apesar de Arabidopsis ser extensivamente estudada, uma análise aprofundada do transcriptoma de células espermáticas, e particularmente de núcleos vegetativos isolados de grãos de pólen maduros, ainda não estava disponível. Consequentemente, o capítulo 2 da minha tese teve como objetivo preencher essa lacuna de conhecimento, através de sequenciação de RNA de larga escala em células espermáticas e núcleos vegetativos isolados por separação de células ativadas por fluorescência (FACS). Os

resultados revelaram 7445 genes a serem expressos em células espermáticas, sendo mais 2788 do que se conhecia anteriormente. Pela primeira vez, pudemos mostrar a expressão de 9200 genes em núcleos vegetativos. Comparámos a célula espermática e o núcleo vegetativo e encontrámos 6906 genes expressos diferencialmente entre eles. A análise da ontologia génica mostrou que os genes enriquecidos em células espermáticas estavam associados à regulação do ciclo celular, processos metabólicos relativos ao DNA e organização da cromatina, enquanto os genes regulados negativamente nas células espermáticas (regulados positivamente no núcleo vegetativo) estavam associados à germinação do pólen e crescimento do tubo, e ao transporte vesicular. Os nossos dados do transcriptoma permitiram decifrar as mudanças não apenas ao nível de expressão génica, mas também ao nível de isoforma. Ao compararmos a célula espermática e o núcleo vegetativo, encontrámos 468 genes principalmente envolvidos no processamento do mRNA, na modificação da cromatina e localização de proteínas, processados diferencialmente ao nível do *splicing*. Os nossos resultados também revelaram que os transcriptomas de células espermáticas e núcleos vegetativos são regulados transcricional e post-transcricionalmente, uma vez que 295 fatores de transcrição e 305 fatores de *splicing* foram diferencialmente regulados tanto na expressão génica como no *splicing* .

A interação pólen-pistilo provoca mudanças nos tubos polínicos conforme eles viajam dentro dos tecidos femininos, mas até que ponto essas mudanças se refletem a uma resolução de células espermáticas e núcleo vegetativo não era conhecido anteriormente. No capítulo 3 desta tese otimizei um método que nos permitiu analisar o transcriptoma mesmo de um pequeno número de células espermáticas isoladas, bem como de núcleos vegetativos, a partir de tubos polínicos em crescimento. Realizámos análises de larga escala de transcriptoma de células espermáticas (SC) de grãos de pólen maduros (MPG), de tubos polínicos cultivados em placa (*in vitro*) e de

tubos polínicos cultivados semi- *in vivo* (SIVPT). Os nossos resultados indicaram que as SC de tubos polínicos eram transcricionalmente muito distintas das provenientes de grãos de pólen maduros, com pequenas diferenças detectáveis entre SC *in vitro* e SC SIVPT. O que sugere que o crescimento do tubo polínico *in vitro* ou em SIVPT pode reprogramar o transcriptoma das células espermáticas, e os sinais femininos podem servir apenas para ajustar essa reprogramação. A tripla comparação dos transcriptomas de células de esperma revelou 1374, 1132, 237 genes diferencialmente expressos entre SC *in vitro* e SC MPG, entre SC SIVPT e SC MPG, e entre SC *in vitro* e SC SIVPT, respectivamente. O enriquecimento funcional de genes regulados positivamente em células espermáticas quando passaram pelo pistilo revelou papéis associados à resposta ao choque térmico, *stress* oxidativo e *folding* e modificação de proteínas, indicando que as células espermáticas sofrem maturação antes da fusão de gâmetas à medida que viajam dentro do tubo polínico para os óvulos. Dos 52 genes que foram regulados positivamente na comparação SC SIVPT vs SC *in vitro*, vários podem ter uma função importante na diferenciação de esperma e fusão de gâmetas (por exemplo, RALF18). Também realizámos a análise do transcriptoma em núcleos vegetativos de MPG e de tubos polínicos cultivados *in vitro*, e usámos tubos polínicos intactos cultivados semi-*in vivo* (referidos aqui como PT) como um *proxy* para um transcriptoma de núcleos vegetativos de SIVPT. A nossa análise indicou 3802, 5068 e 3213 genes a serem diferencialmente expressos em comparação VN *In vitro* vs VN MPG, PT vs VN MPG, PT vs VN *In vitro*, respectivamente. Os nossos resultados indicam que os transcriptomas do núcleo vegetativo sofrem alterações mais dinâmicas durante o crescimento do tubo polínico em comparação com as células espermáticas. Quando observámos os genes que foram regulados positivamente nos núcleos vegetativos durante o crescimento do tubo polínico, descobrimos que a função associada ao crescimento da extremidade da célula e ao transporte mediado pela vesícula era enriquecida, independentemente dos tubos polínicos terem crescido *in*

vitro ou através do pistilo. Esses resultados indicam que os núcleos vegetativos sofrem maioritariamente alterações semelhantes ao nível da expressão génica durante o crescimento do tubo polínico, e as alterações que ocorrem exclusivamente durante o SIVPT são sobretudo devido à presença de sinais femininos que guiam o tubo polínico. A comparação direta de PT vs VN *in vitro* leva a uma dramática regulação positiva de genes enriquecidos para funções associadas à regulação do ciclo celular, reparação e replicação do DNA, indicando o seu potencial papel durante o crescimento do tubo polínico, particularmente à medida que viajam no interior dos tecidos femininos.

As células espermáticas na maioria das plantas com flores são isomórficas. No entanto, algumas espécies têm células espermáticas dimórficas causadas por uma divisão celular assimétrica da célula geradora. O exemplo mais bem estudado de células espermáticas dimórficas corresponde provavelmente à planta *Plumbago zeylanica*, na qual as duas células espermáticas diferem nos níveis de expressão génica. No entanto, a questão de se plantas como Arabidopsis, com gâmetas masculinos isomórficos, mostram diferenças ao nível de expressão génica entre as duas células espermáticas, permanece. Para responder a essa questão, no Capítulo 4 usamos o nosso método otimizado de tubo polínico semi *in vivo* para estudar o transcriptoma de células espermáticas com resolução de célula única. A análise de sequenciação do RNA de 80 células individuais indicou que 208 genes são potencialmente expressos diferencialmente entre dois grupos de células espermáticas. As funções prováveis desses genes enquadram-se nas categorias de transcrição, metabolismo de proteínas, transporte e transdução de sinal. No entanto, análises adicionais com um número muito maior de células são necessárias para confirmar estas observações.

Para concluir, otimizei um método que me permitiu analisar o transcriptoma de células espermáticas e núcleos vegetativos separadamente, não apenas

ao nível de expressão do gene, mas também ao nível de *splicing* alternativo. Além disso, a análise de células espermáticas e núcleos vegetativos de tubos polínicos em crescimento permitiu-me desvendar a dinâmica do transcriptoma, particularmente durante o crescimento do tubo polínico nos tecidos femininos, com uma resolução sem precedentes. No geral, o meu trabalho de doutoramento fornece recursos abundantes para pesquisas futuras no campo da reprodução das plantas, com muitos dos genes identificados sendo os principais candidatos para uma futura caracterização funcional detalhada.

Abstract

Pollen development in flowering plants initiates within anthers, where microspore mother cells undergo meiosis to form haploid microspores. The haploid microspore undergoes asymmetric division (pollen mitosis I) to form a larger vegetative cell harbouring a smaller generative cell that undergoes a second mitotic division (pollen mitosis II) to form two sperm cells. While the vegetative cell drives pollen tube growth, the two sperm cells are passively transported within pollen tubes to the embryo sac for double fertilization. Male gametophyte development in flowering plants has been studied extensively over the last decades. Transcriptomic approaches have been used to analyse not only gene expression dynamics during pollen development, but also at the level of the male germline in Arabidopsis, rice, maize and tomato. A systematic analysis of the transcriptomes of Arabidopsis sperm cells and vegetative nuclei from mature pollen and from pollen tubes that have grown through the pistil had never been undertaken, partially due to technical limitations with isolation and processing of such samples. Such a study however holds the promise to identify transcriptional changes triggered in the sperm cells during pollen-pistil interactions, potentially important for successful fertilization of egg and central cell, respectively.

Despite Arabidopsis being extensively studied, an in-depth transcriptome analysis of sperm cells and particularly vegetative nuclei isolated from mature pollen grains was still not available. Chapter 2 of my thesis therefore aimed to bridge this knowledge gap by performing high throughput RNA-seq on FACS sorted sperm cells and vegetative nuclei. The results revealed 7445 genes to be expressed in sperm cells, more 2788 in addition to what was previously known. For the first time, we were able to show the expression of 9200 genes in vegetative nuclei. We compared sperm cell and vegetative nucleus and found 6906 genes to be differentially expressed between sperm cell and vegetative nucleus. Gene Ontology analysis showed that genes

enriched in sperm cells were associated with cell cycle regulation, DNA metabolic process, and chromatin organisation, whereas genes downregulated in sperm cells (up-regulated in vegetative nucleus) were associated with pollen germination and tube growth, and vesicle-mediated transport. Our transcriptome data helped to decipher changes not only at gene expression level, but also at the isoform level. When we compared sperm cell and vegetative nucleus, we found 468 genes to be differentially spliced and these genes were enriched for mRNA splicing, chromatin modification, and protein localization. Our results revealed that transcriptomes of sperm cells and vegetative nuclei are regulated both at transcriptional level as well as splicing level, as 295 transcription factors and 305 splicing factors were differentially expressed both at gene level as well as isoform level.

Pollen-pistil interaction elicits changes in the pollen tubes as they travel within the female tissues, but to what extent these changes reflect at the resolution of sperm cells and vegetative nucleus was not known previously. In chapter 3 of this thesis I optimised a method that enabled us to analyse the transcriptome even from a small number of isolated sperm cells as well as vegetative nuclei from growing pollen tubes. We performed high throughput transcriptome analysis of sperm cells (SC) from mature pollen grains (MPG), pollen tubes grown on plate (*in vitro*) and semi *in vivo* grown pollen tubes (SIVPT). Our results indicated that sperm cells from pollen tubes were highly transcriptionally distinct when compared to sperm cells from mature pollen grains, with minor differences detectable between SC in vitro and SC SIVPT. This suggests that both *in vitro* and SIVPT pollen tube growth may reprogram the transcriptome of sperm cells, and female cues may only serve to fine tune this reprogramming. Three-way comparison of the sperm cell transcriptomes revealed 1374, 1132, 237 genes to be differentially expressed in SC in vitro vs SC MPG, SC SIVPT vs SC MPG, and SC in vitro vs SC SIVPT comparisons, respectively. Functional enrichment of genes that were

upregulated in sperm cells when they passed through the pistil were found to be associated with heat shock response, oxidative stress, and protein folding and modification, indicating that sperm cells undergo sperm maturation before gamete fusion as they travel within the pollen tube to the ovules. Out of 52 genes that were up-regulated in SC SIVPT vs SC In vitro comparison, several may have an important function in sperm differentiation and gamete fusion (e.g. RALF18). We also performed transcriptome analysis on vegetative nuclei from MPG and from In vitro grown pollen tubes, and used whole semi *in vivo* grown pollen tubes (referred here as PT) as a proxy for a vegetative nuclei transcriptome from SIVPT. Our analysis indicated 3802, 5068 and 3213 genes to be differentially expressed in VN In vitro vs VN MPG, PT vs VN MPG, PT vs VN Invitro comparison respectively. Our results indicate that vegetative nuclei transcriptomes undergo more dynamic changes during pollen tube growth compared to the sperm cells. When we looked at the genes that were upregulated in the vegetative nuclei during pollen tube growth, we found function associated with cell tip growth and vesicle mediated transport to be enriched irrespective of whether pollen tubes were grown in vitro or through the pistil. These results indicate that vegetative nuclei largely undergo similar changes at the gene expression level during pollen tube growth and the changes that occur exclusively during SIVPT are mainly due to the presence of female cues guiding the pollen tube. The direct comparison of PT vs VN In vitro, leads to dramatic upregulation of genes enriched for functions associated with cell cycle regulation, DNA repair and replication, indicating their potential role during pollen tube growth particularly as they travel within the female tissues.

Sperm cells in most flowering plants are isomorphic. However, a few plant species have dimorphic sperm cells caused by an asymmetrical cell division of the generative cell. Probably the best studied example for dimorphic sperm cells is *Plumbago zeylanica*, in which the two sperm cells differ at the gene expression level. However, the question remains, whether plants such as

Arabidopsis with isomorphic male gametes do show differences at the gene expression level between the two sperm cells. To address this question, in Chapter 4, we used our optimised semi *in vivo* pollen tube method to study the transcriptome of sperm cells at single cell resolution. RNA-Seq analysis of 80 single cells indicated 208 genes being potentially to be differentially expressed between two sperm cell clusters. Potential functions of these genes fell into the categories of transcription, protein metabolism, transport and signal transduction. However, additional analyses with much larger numbers of cells are needed to confirm our observations.

To conclude, I have optimised a method that allowed me to analyse the transcriptome of sperm cells and vegetative nuclei separately, not only at gene expression level but also at the level of alternative splicing. Moreover, analysis of sperm cells and vegetative nuclei from growing pollen tubes allowed me to unravel transcriptome dynamics particularly during pollen tube growth within female tissues at an unprecedented resolution. Overall, my PhD work provides rich resources for future research in the field of plant reproduction, with many of the genes identified being prime candidates for future in-detail functional characterization.

Abbreviations

AA	Alternative acceptor
AD	Alternative donor
AGL	Agamous like
AP2/EREBP	APETALA2/ethylene-responsive element binding protein
AS	Alternative splicing
At	Arabidopsis thaliana
AtVLN1	*Arabidopsis thaliana* villin1
BCP	Bicellular pollen
bHLH	Basic helix-loop-helix
bp	Base pair
bZIP	Basic leucine zipper
CaCl2	Calcium chloride
cDNA	Complementary Deoxyribonucleic acid
CMT2	Chromomethylase 2
CPM	Count per million
DAS	Differentially alternatively spliced
DAZ1	DUO1-Activated Zinc Finger 1
DAZ2	DUO1-Activated Zinc Finger 2
DEG	Differentially expressed genes
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleoside-triphosphate
DTT	Dithiothreitol
EC	Egg cell
ERF	Ethylene response factor
ES	Exon Skipping
EST	Expressed sequence tag
FACS	Fluorescence activated cell sorting
FDR	False Discovery Rate
FSC	Forward scatter

gDNA	Genomic deoxyribonucleic acid
GFP	Green Fluorescent Protein
GO	Gene Ontology
GP	Germinating Pollen
GRO-seq	Global Run-On sequencing
H_2O	Water
H_3BO_3	Boric acid
HMG	High Mobility Group
IR	Intron retention
IVPT	*In vivo* grown pollen tubes
kb	Kilobase pair
KH_2PO_4	Monopotassium phosphate
LRR_RLKs	Leucine-rich repeats receptor-like kinases
LUG	Leunig
MEF2-type	Myocyte Enhancer Factor 2 type
Mei	Meiocytes
$MgCl_2$	Magnesium chloride
MgSO4	Magnesium sulphate
MGU	Male Germ Unit
MLO	Mildew Resistance Locus O
mM	Milli Molar
MOPS	3-(*N*-morpholino)propanesulfonic acid)
MPG	Mature pollen grain
mRNA	messenger RNA
MYA	Million years ago
MYB	Myeloblastosis
NET-seq	Native elongating transcript sequencing
Neu-seq	Nascent 5-EU–Labeled RNA Sequencing
NGS	Next generation sequencing
PCR	Polymerase chain reaction

PMC	Pollen mother cell
PMI	Pollen Mitosis I
PMII	Pollen Mitosis II
PS	Percent spliced
RALF	Rapid Alkalinization Factor
RdDM	RNA-directed DNA methylation
REM	Reproductive Meristem
RFP	Red Fluorescent Protein
RNA	ribonucleic acid
RNAse	Ribonuclease
RNA-seq	ribonucleic acid-sequencing
ROS1	Repressor of silencing1
RPM	Revolutions per minute
rRNAs	Ribosomal ribonucleic acid
RT	Reverse transcription
RT-PCR	Reverse transcription polymerase chain reaction
SC	Sperm cell
scATAC-seq	Single cell assay for transposase-accessible chromatin using sequencing
siRNA	Small interfering ribonucleic acid
SIV-PG	Semi *in vivo* guided pollen tube growth
SIVPT	Semi *in vivo* pollen tube
Spm	specificity measure
SR	serine/arginine-rich
SSC	side scatter
TAIR	The Arabidopsis Information Resource
TCP	Tricellular pollen
TFs	Transcription factors
TMM	Trimmed mean of M-values
TPM	Transcripts per million

tRNAs	Transfer ribonucleic acid
TSO	Template switching oligo
UNM	Uninucleate microspore
VN	Vegetative nucleus
µl	Micro liter
µm	micrometer
µM	Micro molar

Table of Content

Chapter 1
Introduction

1.1. Introduction

Plant sexual reproduction is a complex process involving interaction between the male gametophyte (pollen grain), female gametophyte (embryo sac) and other female reproductive sporophytic tissues that facilitate pollen tube growth and guidance. Extensive research has been conducted over the past two decades to understand this vital process, significantly improving our understanding of both male and female gametophyte and how they interact during gamete fusion. The need to study this fundamental process in plants has been exacerbated by the ever-growing population coupled with the changing global climate.

The effect of climate change poses a serious risk to agriculture, impacting future food security (Vermeulen et al., 2012). Increased temperature, drought, and shifting of seasonal rainfall all affect plant growth and development (Hedhly et al., 2009). But the most affected of all is plant reproduction, particularly the male gametophyte (Zinn et al., 2010; Hedhly, 2011). Elevated temperatures pose major risk to pollen development and all stages of pollen development from anther dehiscence to pollen germination and tube growth are equally hampered (Herrero and Johnson, 1980; Kakani et al., 2002; Kakani et al., 2005; Sato et al., 2002; Zinn et al., 2010). Studies have found that high temperature during pollen tube growth results in altered Golgi apparatus, mitochondria and rough endoplasmic reticulum, thus affecting vesicle transport, protein translation and energy metabolism, comprising some of the most critical processes during plant reproduction (Kandasamy and Kristen, 1989).

1.2. Sexual reproduction in flowering plants

The life cycle of flowering plants alternates between dominant and diploid sporophyte phase and highly reduced haploid gametophyte phases. Both male and female gametophyte are the products of meiosis occurring in

progenitor cells inside anthers and carpels, respectively (McCormick 1993). Upon full maturation the female gametophyte consists of an egg cell, the female gamete, a central cell, antipodal cells and two synergids. The male gametes are passively delivered to the embryo sac by a pollen tube that forms after a pollen grain lands on a receptive stigma. The pollen tube grows through the transmitting tissue of the pistil and releases its two sperm cells into a synergid (Figure 1.1). During the ensuing double fertilization event one sperm cell fertilizes the egg cell, giving rise to the diploid embryo, while a second sperm cell fertilizes the diploid central cell, resulting in a triploid endosperm. A schematic representation of sexual reproduction is shown in Figure 1.1.

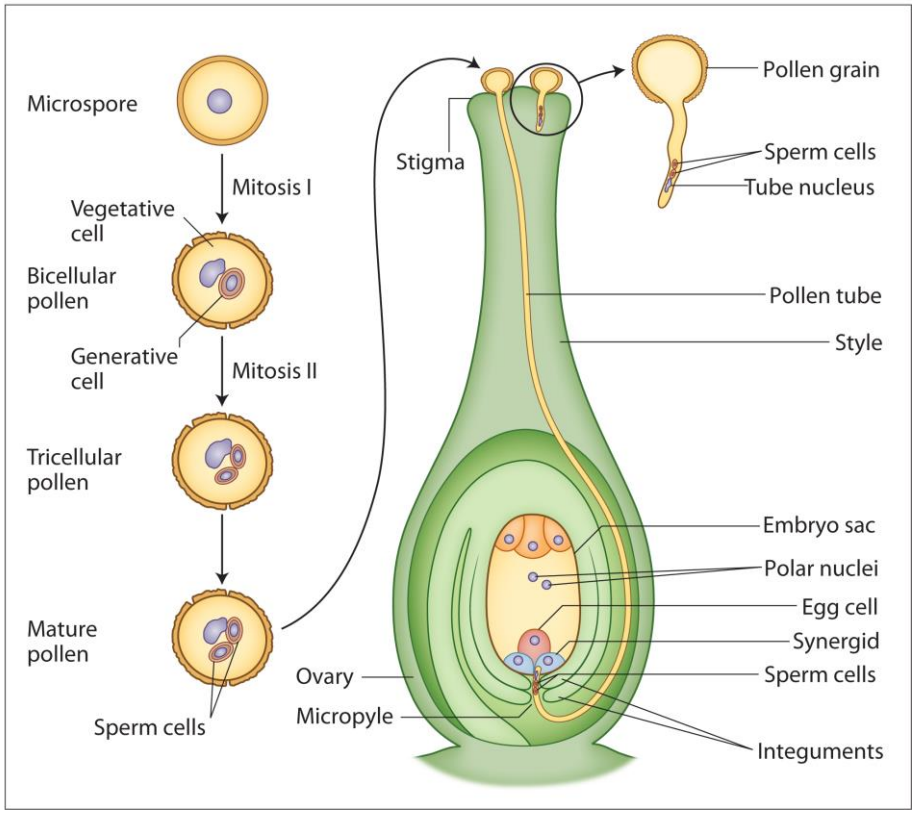


Figure 1.1: Sexual reproduction in flowering plants (Julca et al., 2021).

1.2.1. Male gametophyte development in flowering plants

Male gametophyte (pollen) development starts within the anthers of flowers, in which a pollen mother cell (PMC) undergoes meiosis to form four microsporocytes. The four microspores are connected as a tetrad by callose, which is later digested by the activity of callases releasing the four haploid microspores (McCormick, 1993). The individual microspores then increase in size, vacuolise and their nuclei migrate towards the periphery. The haploid microspores then undergo asymmetric pollen mitosis I (PMI) resulting in a large vegetative cell and a smaller generative cell. The generative cell, which later gives rise to the male gametes, is then engulfed by the vegetative cell. The generative cell undergoes another round of symmetric division, pollen mitosis II (PMII), leading to the formation of two sperm cells. Timing of PMII is a key determinant of whether the discharged pollen is two-celled (bi-cellular) or three celled (tri-cellular) at anthesis (Brewbaker, 1967). In most plant species pollen dehisces at the bicellular pollen stage, but in approximately 30% of plant species, including Arabidopsis, maize and rice, PMII still occurs in the anther, and the tricellular mature pollen grain is then released for pollination (Williams et al., 2014a). A schematic of tricellular pollen development is represented in Figure 1.2.

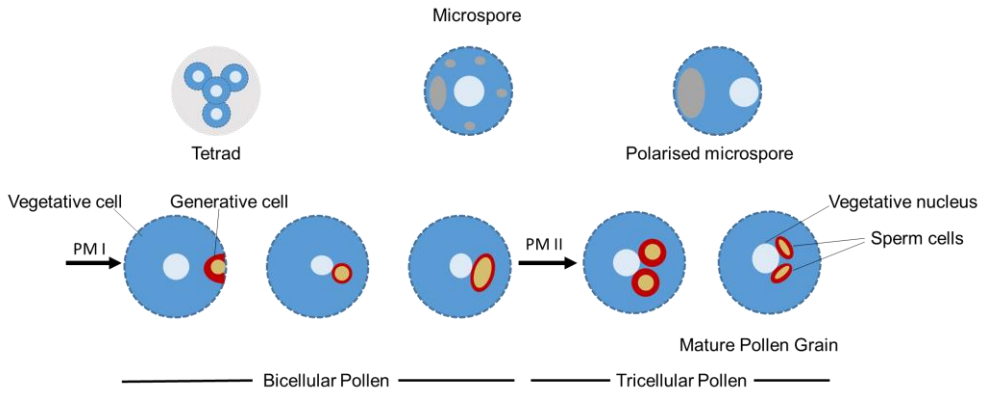


Figure 1.2: Different stages of pollen development in *Arabidopsis thaliana*. A tetrad of 4 haploid microspores are the product of meiotic division of an undifferentiated pollen mother cells. These microspores undergo a first asymmetrical pollen mitosis division (PM I) giving rise to bicellular pollen with a generative cell embedded within a cytoplasm of a vegetative cell. The generative cells undergoes a second pollen mitotic division (PM II) to give rise to tricellular pollen with two sperm cells. The two sperm cells together with the vegetative nucleus form the male germ unit within a mature pollen grain (Adapted from Rutley and Twell, 2015).

1.2.2. Gene expression in the male gametophyte

The past two decades have seen an increase in gene expression studies covering several stages of pollen development. Though initial evidence of gene expression in pollen was obtained at the level of single genes, such as LAT52 from tomato or cloned cDNAs of Adh-1 from maize (Stinson and Mascarenhas, 1985; Twell et al., 1989), more recently genome-wide transcriptome studies have become common, providing unprecedented insights into pollen development.

Early studies in maize and *Tradescantia* estimated approximately 20,000 genes to be expressed in pollen in these species (Willing and Mascarenhas, 1984; Willing et al., 1988). The number of genes expressed in pollen were up to 30% reduced compared to shoots, indicating the small size of the pollen transcriptome. This was not entirely surprising as the male gametophyte is highly reduced compared to sporophytic tissues (Willing and Mascarenhas, 1984; Willing et al., 1988).

Subsequent studies in *Tradescantia* and *Lilium* pollen provided evidence for a developmental regulation of transcription since levels of regulatory RNAs such as tRNAs and rRNAs peaked around pollen mitosis I (Mascarenhas, 1990). Subsequently, a study in tobacco pollen found that these RNAs

continue to increase even after PMI, indicating that each species may have different transcriptional and regulatory mechanisms governing gametogenesis (Schrauwen et al., 1990).

The first large scale study to profile gene expression in pollen was made possible using the 8K Arabidopsis Genome (AG) array from Affymetrix, that had probes for almost 8100 genes encompassing 30% of Arabidopsis genome. Use of this miroarray allowed to identify expression of almost 1000 (Honys and Twell, 2003) and 1600 (Becker et al., 2003) genes, respectively, in two independent studies of mature pollen grains. Later, high-density *Arabidopsis* 'whole genome' oligonucleotide probe arrays were developed, called the Affymetrix ATH1 GeneChip® probe array, encompassing approximately 23750 genes (Redman et al., 2004).

The era of DNA microarrays revolutionized the field of gene expression analysis. The ease of sample preparation, well-established data analysis protocols, as well as time-saving and standardized data mining across studies, favored the use of DNA microarrays for a long time. However, the use of DNA microarrays was not without its share of limitations. For example, GeneChip preparation required microfabrication, limiting the use of the platform to species for which the genome sequence was available and hindering frequent updates of the GeneChip content.

Thus, not surprisingly, recent deep DNA sequencing technologies such as RNA-seq have become the predominant technology to analyze gene expression levels (Loraine et al., 2013). RNA-seq offers several advantages, as it requires little to no information about the genome, enabling transcriptome assembly and gene expression analysis for more non-model species.

The wealth of omics data has advanced our understanding of pollen development, not only in Arabidopsis but also in other species. Of all omics studies done on male gametophytes so far, transcriptomics formed the

dominant approach, including both coding as well as non-coding transcriptome analyses, followed by proteomics, translatome, methylome, phosphoproteome, secretome, and metabolome.

While almost all important plant families are represented in male gametophyte studies, only a handful of plant species are well documented for all pollen developmental stages. The four most represented families of plants include a combination of both monocots and dicots, and plants with pollen that is shed at the bicellular or tricellular stage. These are Brassicaceae (dicot, e.g. *Arabidopsis thaliana*), Poaceae (monocot, eg. *Oryza sativa* and *Zea mays*), Solanaceae (dicot, eg. *Nicotiana tabacum* and *Solanum lycopersicum*), Liliaceae (monocot, eg. *Lilium longiflorum*). These four different families are representative model species studied to understand various aspects of plant physiology and development, including but not limited to male gametophyte development. Some of these model plant species were subjected to a combination of various omics approaches, in particular Arabidopsis, rice, maize, and tomato.

1.2.3. Pollen developmental transcriptome

The first large scale omics study of the male gametophyte was transcriptomic (Becker et al., 2003; Honys and Twell 2003; Lee and Lee 2003) and so far transcriptomic data for at least one pollen developmental stage has been published for 24 plant species (See Table 1.1).

Table 1.1: Summary of transcriptomic studies from different published pollen development stages

Species	M	T	UNM	BCP	TCP	MPG	References
Bicellular Pollen							
Cryptomeria japonica						✓	Futamura et al., 2006, Tsubomura et al., 2016
Lilium longiflorum			✓			✓	Okada et al., 2006; Okada et al., 2007; Obermeyer et al., 2013; Lang et al., 2015
Vitis vinifera						✓	Fasoli et al., 2012
Glycine max						✓	Haerizadeh et al., 2009
Pyrus bretschneideri						✓	Zhou et al., 2016
Nicotiana tobacum			✓	✓	✓	✓	Xin et al., 2011; Hafidh et al., 2012a; Hafidh et al., 2012b; Bokvaj et al., 2015; Conze et al., 2017
Solanum lycopersicum			✓			✓	Frank et al., 2009; Loraine et al., 2015; Liu et al., 2018
Solanum tuberosum						✓	Sanetomo and Hosaka 2013
Petunia inflata						✓	Williams et al., 2014b
Olea europaea						✓	Carmona et al., 2015, Iaria et al., 2016
Solanum demissum						✓	Sanetomo and Hosaka 2013
Tricellular Pollen							
Oryza Sativa	✓	✓	✓	✓	✓	✓	Hirano et al., 2008; Hobo et al., 2008; Russell et al., 2008; Suwabe et al., 2008; Tang et al., 2010; Aya et al., 2011; Luo et al., 2011; Shen et al., 2011; Peng et al., 2012; Russell et al., 2012; Anderson et al., 2013; Wu et al., 2014

Species	M	T	UNM	BCP	TCP	MPG	References
Zea Mays	✓		✓			✓	Engel et al., 2003; Ma et al., 2008; Davidson et al., 2011; Chettoor et al., 2014; Dukowic-Schulze et al., 2014, Zhang et al., 2014; Chen et al., 2017; Warman et al., 2020
Triticale						✓	Tran et al., 2013
Sorghum halepense						✓	Campbell et al., 2015
Phleum pratense						✓	Schulten et al., 2013
Fragaria vesca			✓			✓	Hollender et al., 2014
Arabidopsis thaliana	✓		✓	✓	✓	✓	Becker et al., 2003; Honys and Twell 2003; Lee and Lee 2003; Honys and Twell 2004; Pina et al., 2005; Schmid et al., 2005; Verelst et al., 2007a; Verelst et al., 2007b; Borges et al., 2008; Wang et al., 2008 a; Adamczyk and Fernandez 2009; Gibalová et al., 2009; Qin et al., 2009; Chen et al., 2010; Costa et al., 2013; Loraine et al., 2013; Becker et al., 2014; Lin et al., 2014
Brassica napus			✓			✓	Whittle et al., 2010
Plumbago zeylanica							Gou et al., 2009
Ambrosia artemisiifolia						✓	Kanter et al., 2013; El Kelish et al., 2014; Bordas-Le Floch et al., 2015
Hordeum vulgare	✓						Barakate et al., 2020
Brassica rapa		✓	✓	✓	✓		Golicz et al., 2021

The 2nd- 7th columns column correspond to the development stage of pollen- M meiocytes; T tetrads; UNM uninucleate microspores; BCP Bi-cellular pollen; TCP tri-cellular pollen; MPG mature pollen grain. (Adapted from Fila et al., 2017)

The first study on different stages of pollen development was published by Honys and Twell, 2004. They used density gradient centrifugation to isolate uninucleate microspore (UNM), bicellular pollen (BCP), tricellular pollen (TCP), and mature pollen grains (MPG). The microarray ATH1 platform was used to study the gene expression profile during male gametophyte development. The number of genes detected as expressed decreased from 11565 in UNM to approximately 7235 in MPG. Though the overall transcriptome size and complexity reduced from UNM to MPG, pollen-specific gene expression increased from 6.9% in UNM to 8.6 % in MPG, indicating pollen differentiation and maturation. While gene expression pattern of UNM and BCP (R = 0.96) and TCP and MPG (R = 0.86) showed a strong correlation, the correlation dropped between BCP and TCP (R = 0.54). This indicates that changes in the transcriptome size between BCP and TCP are not just because of reduced gene expression but also due to the activation of a new set of genes that may be involved in pollen maturation. The pairwise comparison between UNM and MPG showed a significant difference indicating pronounced changes occurring during the transition from proliferating microspore to terminally differentiated mature pollen grain. While the early proliferative stage had an overrepresentation of transcription factors and core-cell cycle genes, late maturation stages showed enrichment of signaling and cell wall metabolism genes (Honys and Twell, 2004).

The first study on Arabidopsis pollen using RNA-seq was published by Loraine et al., 2013. Nevertheless, RNA-seq of mature pollen grains did not show a dramatic increase in the number of genes expressed compared to microarrays. RNA-seq identified 6044 genes in pollen, against an average of 6000 (between 3954-7235) identified by microarrays in various studies (Honys and Twell 2004; Pina et al., 2005; Schmid et al., 2005; Borges et al., 2008; Wang et al., 2008a; Qin et al., 2009; Loraine et al., 2013). Overall, the findings of these studies revealed that the pollen has one of the smallest

transcriptomes in the plant expressing just a quarter of its genome. If we compare the Arabidopsis pollen transcriptome to other tissues with reduced numbers of cells such as root hairs with 11696 genes (Becker et al., 2014), guard cells of stomata with 13222 genes (Bates et al., 2012), and trichomes with 13328 genes (Marks et al., 2009), each of these express nearly twice the number of genes than in pollen.

Apart from *Arabidopsis thaliana*, the transcriptomic approach was also used in rice that has served as an excellent model system to study male gametophyte development in monocots. Rice diverged from Arabidopsis almost 200 Million years ago and therefore studying its transcriptome provided an important research tool to understand how the molecular mechanism underlying male gametophyte development evolved and if there is any conservation between Arabidopsis and Rice (Goff et al., 2002). Analysis of the genome from both species shed important light on how 85% of Arabidopsis genes have homologs in rice, while only 41 % of rice genes have homologs in Arabidopsis (Goff et al., 2002).

The pollen transcriptome in rice has been studied in two independent studies. The first study used a combination of laser microdissection of anthers and 44 K rice oligo microarrays (Hobo et al., 2008; Suwabe et al., 2008) while the second one used density gradient centrifugation and the 57 K Rice genome Array (Wei et al., 2010). Both studies showed similar transcriptome progression during pollen development (Hobo et al., 2008; Suwabe et al., 2008). The number of genes expressed decreased from 14590 in UNM to 5939 in MPG (Wei et al., 2010). The rice pollen transcriptome was highly reduced, similar to Arabidopsis, expressing only a third of genes in pollen compared to roots (17383 genes) and leaves (17242 genes). In addition, there was a strong similarity between transcriptomes of early stages (UNM and BCP; R= 0.82), and of the late stages (TCP and MPG; R=0.76), indicating a strong shift in the transcriptome from early proliferating stage to late

maturation stage. This U-type change, as the authors described it, occurred for pollen stage-specific transcripts in both Arabidopsis and rice with the smallest number of genes being expressed at the bicellular stage of pollen development. Although both Arabidopsis and rice showed a decrease in transcriptome complexity during pollen development, the decline was more pronounced in rice from TCP to MPG indicating differences in the onset of large-scale repression of gene expression or a high rate of transcript turnover in rice. Despite global pollen developmental profiles in rice and Arabidopsis being similar, significant differences were reported for GO terms for genes expressed during pollen developmental stages. For instance, while rice had expression of more stage-specific genes associated with defense and stress response in mature pollen, Arabidopsis showed enrichment for signaling at the bicellular stage. Species-specific expression of some transcription factor families was observed, such as LUG, LIM, and HMG in rice and REM, RAV, and Whirly transcription factors expressed only in Arabidopsis pollen (Wei et al., 2010). In fact, out of a total of 722 pollen enriched transcription factors in rice, 73.8 % (533) had no corresponding homologue expressed in the male gametophyte of Arabidopsis. These differences in the expression of transcription factor families imply distinct gene expression networks governing the transcriptome of pollen development in the two species (Wei et al., 2010).

Tobacco has been one of the most studied pollen systems, even before the Arabidopsis genome was sequenced (Rogers et al., 1992; Van Herpen et al., 1992). However, it was not before 2012 that the first transcriptome of tobacco pollen using Agilent 44-K tobacco microarrays was published. Mature pollen, 4 h and 24 h *in vitro* grown pollen tubes were studied to understand gene expression changes during these pollen developmental stages (Hafidh et al., 2012a, Hafidh et al., 2012b). This study was then complemented with the transcriptome profiles at early stages of pollen development including uninucleate microspore (UNM), early and late bicellular pollen (BCP) (Bokvaj

et al., 2015). As observed in Arabidopsis and rice, the tobacco pollen transcriptome also showed a reduced complexity, although the number of transcripts increased from uninucleate microspore (16700) to early bicellular pollen (22872) and decreased in the late bicellular pollen (19750). Taken together, and despite some similarities between Arabidopsis, rice, and tobacco pollen regarding the overall pollen development profile, the difference in the shift of the transcriptome complexity from microspore pollen to bicellular pollen stage reflects changes in the rate of pollen development. Therefore, it would be worthwhile to explore, if this is a general trend for plants with bicellular pollen at anthesis, such as tobacco, versus plants that shed pollen at tricellular stage, such as rice and Arabidopsis.

Maize is another interesting model system that is extensively studied for its male gametophyte development with three studies having been conducted. The first study was conducted by Ma et al., 2008 using microarrays, in which the authors identified 10539 different transcripts in pollen. The study was further extended to two independent RNA-seq experiments in which up to 38% more genes were identified (Davidson et al., 2011; Chettoor et al., 2014). The maize pollen transcriptome expressed only around half (57%) of the average number of genes expressed across the other 12 sporophytic tissues that were compared in the study. Pollen was also the tissue that showed the most extreme range of transcript levels (Davidson et al., 2011). There was a considerable amount of overlap between the pollen transcriptome in Arabidopsis and maize. For instance, out of 4407 homologs of Arabidopsis pollen expressed genes represented on the maize array, 78% or 3444 genes were expressed in maize pollen, with a substantial overlap for highly expressed genes.

The analysis of the pollen transcriptome in soybean (*Glycine max*) was the first described example for a legume and a bicellular pollen species (Haerizadeh et al., 2009). The pollen of soybean expressed 10299 transcripts

representing 37% of those detected in sporophytic tissues, while 7.9% were pollen-specific. Like Arabidopsis, the soybean pollen transcriptome had enrichment of genes for signaling genes, cell-wall modifying enzymes, and transporters. The differences, however, were observed in the number as well as diversity of regulatory proteins, while small RNA pathway proteins were underrepresented in soybean. Soybean pollen had a higher expression of heat-shock protein and heat-shock TFs in contrast to Arabidopsis pollen, in which their expression was limited to germinating pollen, suggesting their involvement as molecular chaperons to facilitate intense physiological activities associated with pollen germination and tube growth.

Apart from some of the species described above, there are many more plant species that have been studied for their male gametophyte development (summarized in Table 1.1). For instance, grapevine is one of the most economically important crops of the Mediterranean climate. The first pollen transcriptome of grapevine was published as a part of a gene expression atlas (Fasoli et al., 2012). As with the other species, mature pollen in *Vitis* also showed a distinct transcriptome including enrichment of cell wall modifying enzymes. However, the pollen transcriptome in *Vitis* was not the most reduced, with stem and berries tendrils showing much lower complexity of the transcriptome, indicating specialized function of specific tissues and the specificity of the tissue studied.

The wealth of knowledge obtained through comparative and developmental transcriptomic studies have laid the foundation for follow up research including analysis of several mutants for pollen transcription factors (Reňák et al., 2012), signaling proteins (Chen et al., 2014), F- box proteins (Ikram et al., 2014), and several other functional studies. Nevertheless, transcriptomic studies on direct comparison between wild-type versus mutant pollen grains are very rare.

1.3. Knowledge gaps in pollen genomics

Despite so many transcriptomic studies, a large majority of studies are from mature pollen, and very few of them included various developmental stages. The data on all or most stages of pollen development is available for at least three species- *Arabidopsis thaliana, Oryza sativa, and Nicotiana tabacum* (Honys and Twell 2004; Wei et al., 2010; Bokvaj et al, 2015; Conze et al., 2017). If we consider species for which at least two pollen development stages are available, we can add several more examples such as *Zea mays* (Xu et al., 2012; Chettoor et al., 2014; Dukowic-Schulze et al., 2014), *Lilium longiflorum* (Okada et al., 2007), *Brassica napus* (Whittle et al., 2010), *Brassica rapa* (Golicz et al., 2021) and *Fragaria vesca* (Hollender et al., 2014). While the earlier studies based on microarrays paved the way for fundamental understanding of pollen development in the model plant Arabidopsis, the use of more advanced techniques such as RNA-seq can now be applied to agriculturally important crops, such as Barley (Barakate et al., 2020). Identification of new transcripts is not the only advantage of RNA-seq. It also confers several benefits in terms of the identification of genome-wide alternative splicing. The first study using RNA-seq for identification of a global alternative splicing landscape in mature pollen was published by Loraine et al., 2013. Here the authors identified at least 30 genes that were differentially spliced between pollen and leaf. Similarly, transposable element-related genes were found to be over-represented in maize pollen, which was possible only through RNA-seq (Chettoor et al., 2014). Also, previous studies on maize pollen through microarrays lacked transcripts encoding for small signaling peptides of the LURE family, as microarrays often did not contain probes for small peptide encoding genes (Chettoor et al., 2014).

1.4. Pollen transcriptome during *in vitro* pollen tube growth and progamic phase

Pollen grains, after landing on a receptive stigma, are hydrated and start to germinate a pollen tube that will then travel within the transmitting tissue of the pistil to reach the ovules. For a long time, it was thought that mature pollen grains contained all transcripts needed for pollen tube growth and it was unclear to which extent de novo transcription was taking place in growing pollen tubes. The first transcriptome study of growing pollen tubes came from Wang et al., 2008a. The study used ATH1 arrays to analyze the transcriptome of mature pollen grains (MPG), *in vitro* germinated pollen after 45 min, and pollen tubes grown *in vitro* after 4h. The number of genes identified during different stages of pollen tube development increased from 3945 in MPG to 4892 in 4h grown pollen tubes.

In terms of percentage, more genes were differentially expressed between pollen germination for 45 min to 4h grown pollen tubes as compared to the transition from mature pollen grain to 45 min germinating pollen. This corroborates with previous observations that stored transcripts drive pollen tube growths during the first hours, even in the presence of transcriptional inhibitors such as actinomycin D, though germination and pollen tube growth were significantly reduced in the presence of the translational inhibitor cycloheximide (Honys and Twell, 2004; Wang et al., 2008a). These results indicate that pollen germination and tube growth do not necessarily depend on transcription but are highly dependent on protein synthesis. The first 45 min of pollen germination showed enrichment of transcripts related to stress response and transcription, whereas during pollen tube growth, signaling and cell metabolism-related genes were found to be overrepresented (Wang et al., 2008a).

Pollen tubes grown on plate (*in vitro*) are devoid of any transcriptomic changes that may be elicited upon interaction with female tissues when pollen

tubes pass through a pistil (Palanivelu and Preuss 2006; Palanivelu and Tsukamoto 2012). The pollen tube is intricately guided by the female sporophytic tissue to reach the ovules, and this journey can take a pollen tube up to 8h and approximately 400 µm to reach the lowest ovule in an Arabidopsis pistil. How this journey of the pollen tube within the female tissues shapes its transcriptome was not known until the study by Qin et al., (2009). Here, the transcriptome of the pollen tube grown through the pistil was compared to *in vitro* grown pollen tubes at 0.5h, 4 h and dry mature pollen. The study showed that though a similar number of genes were found to be expressed in mature pollen and in *in vitro* grown pollen tubes, it was in pollen tubes grown semi *in vivo* (SIVPT) that the highest number of genes (7044 genes) was expressed. These semi *in vivo* grown pollen tubes also showed the largest number of SIVPT-specific genes (1254), suggesting that the growth through the pistil confers significant changes to the pollen tube transcriptome. The authors also identified 383 genes that were unique to semi *in vivo* grown pollen tubes compared to any other pollen condition, and the genes were enriched for signaling, defense responses, and transportation. Many of the genes were found to be pollen-specific and none of these genes were found to be active in pollen tubes grown *in vitro* (Qin et al., 2009). This was the first attempt at identifying genes induced only in the presence of female cues. Of 33 candidate genes induced during semi *in vivo* pollen tube growth that were tested for their function, 5 were found to be indispensable for tube growth while another two exhibited an impaired ovule targeting phenotype.

Apart from this first study on the transcriptome of pollen tubes grown through the pistil, two other studies came out in 2014 by Chen et al., 2014 and Lin et al., 2014. Chen et al., 2014 used a 12X 135 k Arabidopsis gene chip (Roche-NimbleGen) to understand how the transcriptome of the pollen tube changes when grown through the pistil. The authors used a modified version of semi *in vivo* pollen tube growth (Palanivelu and Preuss, 2006), in which they placed

ovules beneath the growing pollen tubes to guide the pollen tubes emerging from the pistils, called semi *in vivo* guided pollen tube growth (SIV-PG). In this study 719 genes were found to be expressed specifically in SIV-PG which were enriched for defensin-like, leucine-rich repeats receptor-like kinases (LRR_RLKs) and TLR-NBS-LRR receptors. Surprisingly there was little overlap between up-regulated genes identified in this study to those obtained from the semi *in vivo* pollen tube system of Qin et al., 2009. The functional screening of 18 identified candidate genes revealed no defects in pollen tube guidance, indicating potential redundancy among large pollen-specifically expressed gene families.

Both studies found genes that were induced upon pistil interaction to be involved in pollen tube growth, signaling, and transcription, including the three transcription factors MYB98, MYB101, and MYB120 (Qin et al., 2009) that were later reported to be responsible for performing an important role in pollen tube differentiation and maturation for sperm release (Leydon et al., 2013; Liang et al., 2013).

Apart from Arabidopsis, more studies have profiled the transcriptome of growing pollen tubes in other species as well (see Table 1.2). For instance, tobacco is used most commonly to study pollen tube tip growth. Two studies on the pollen tube transcriptome of tobacco revealed a 3.2% increase in the number of transcripts during pollen tube growth from 13966 in mature pollen grain to 14420 in 24 h *in vitro* grown pollen tubes, and 3597 transcripts were found to be common between these stages. Up to 4.8 % of transcripts (699) were significantly up-regulated compared to mature pollen grain and 4 h pollen tubes, while 2.2 % of transcripts (320) expressed de novo after 4h, showing the enhanced transcriptional capacity of *in vitro* grown pollen tubes (Hafidh et al., 2012a; Hafidh et al., 2012b).

Table 1.2: Summary of transcriptomic studies from pollen tubes and progamic phase of different species

Species	GP	Plate	SIVPT	IVPT	References
Lilium longiflorum	✓	✓		✓	Huang et al., 2011; Obermeyer et al., 2013; Lang et al., 2015
Oryza sativa	✓				Wei et al., 2010
Pyrus bretschneideri	✓	✓			Zhou et al., 2016
Arabidopsis thaliana	✓	✓	✓		Wang et al., 2008a ; Qin et al., 2009; Chen et al, 2014
Camelia sinesis		✓			Wang et al., 2016
Nicotiana tabacum		✓			Hafidh et al., 2012a, Hafidh et al., 2012b, Conze et al., 2017
Olea europaea		✓		✓	Carmona et al., 2015, Iaria et al., 2016

The 2nd - 5th columns correspond to the different stage of pollen germination or tube growth- GP Germinating pollen; Plate *In vitro* pollen tubes grown at different time points; SIVPT semi-*in vivo* grown pollen tubes; IVPT *in vivo* grown pollen tubes (Adapted from Fila et al., 2017).

To summarize, studies on pollen gene expression dynamics revealed a similar trend during pollen development and tube growth across all species analyzed. Overall, the male gametophyte exhibited reduced transcriptome complexity when compared to other sporophytic tissues. Maximum complexity was reached around the bicellular pollen stage and then reduced thereafter until the mature pollen grain stage (Honys and Twell 2004; Wei et al., 2010; Peng et al., 2012; Bokvaj et al., 2015; Rutley and Twell 2015). Pollen tube transcriptome diversity remained largely stable during the progamic phase, being similar to that of a mature pollen grains or increasing slightly by up to 0.1-3% in growing pollen tubes in comparison to mature pollen grains in *Arabidopsis thaliana, Oryza sativa* and *Nicotiana tabacum* (Qin et al., 2009; Wei et al., 2010; Hafidh et al., 2012a; Hafidh et al., 2012b; Zhou et al., 2016; Conze et al., 2017). However, all these studies have been conducted on the whole pollen tube and the transcriptomic changes at the resolution of sperm cell and vegetative nucleus still remain elusive.

1.5. Male germline transcriptome

The analysis of male gametophytes has provided unprecedented insights into transcriptomic changes occurring during pollen development. However, the studies described so far could not distinguish the relative contribution of the sperm cells and the vegetative nucleus to the pollen transcriptome. Therefore, after analysis of the whole pollen grain, several studies compared the transcriptome of isolated sperm cells with either whole pollen grain or isolated vegetative nuclei, whenever possible. So far, nine studies of sperm cell transcriptomes from 7 species have been published, covering both monocots and dicots, and bicellular or tricellular pollen.

The analysis of the sperm cell transcriptome conducted in maize using EST library sequencing (Engel et al., 2003) helped to identify a diverse set of mRNAs corresponding to at least 2560 genes. Following this, the

transcriptome of Arabidopsis sperm cells was first published by our group using ATH1 microarray (Borges et al., 2008). The Arabidopsis sperm cell transcriptome with 5829 expressed genes was found to be smaller than that of mature pollen grain (7177 genes). Arabidopsis sperm cells showed enrichment of 2400 genes compared to seedlings and pollen, while approx. 642 genes were exclusively expressed in sperm. Among the transcripts enriched in sperm cells, most were encoding proteins involved in cell cycle regulation, ubiquitin-proteasome system, and DNA repair, all of which are necessary for the maintenance of germline integrity (Pina et al., 2005; Borges et al., 2008). The list of the overrepresented terms for mature pollen grains was different (eg. vesicle mediated transport, cell-cell communication and signal transduction), implying that sperm cells have a unique transcriptome.
On the other hand, studies on rice sperm cells using the 57 K microarray showed expression of over 11000 genes, almost twice as many as in Arabidopsis sperm cells, possibly due to a larger size of the rice genome and a more complete representation of the genome than on the Arabidopsis array. Subsequently, an RNA-seq study on rice sperm cells and vegetative cell showed different results compared to 57 K microarray as more genes were expressed in vegetative cells (18611 genes) compared to the sperm cells (16985 genes) (Russell et al., 2012; Anderson et al., 2013). Similar to Arabidopsis, both microarray and RNA-seq data of rice sperm cells showed genes enriched for cell cycle, DNA modification and repair, and ubiquitin pathways.

The first large scale transcriptomic study on maize sperm cells was published by Chen et al., 2017, in which 1000 sperm cells from maize were used to profile their gene expression. The results of this study revealed that genes that were enriched in sperm cells were involved in cell cycle regulation. The authors also compared the transcriptome data of maize sperm cells with rice and found that several predicted rice homologs had similar expression profiles, though many predicted rice homologs were found to have weaker

expression patterns in rice than in maize. This was largely attributed to the difference in the method as well as difficulty in predicting true orthologs if a gene existed in a large superfamily. Moreover, despite some differences, the authors were still able to pinpoint some of the similarities between the sperm cells of the two monocots. For instance, among the top 80 highly expressed genes in maize, 10% encoded for histone and high mobility proteins, an observation also reported for rice sperm cells. These findings attest to the growing evidence that the sperm cell chromatin is highly compacted and condensed in all plant species studied, including Arabidopsis. The lack of genes encoding for translational machinery in maize sperm cells also led to the conclusion that while sperm cells are transcriptionally active, they lack translation, which is in contrast to egg cells which are prepared to undergo translation immediately upon being fertilized. This observation of lack of translation machinery in maize sperm cells was consistent with what was reported in rice.

The most recent plant species for which the transcriptome of the male germline is available is tomato (Liu et al., 2018), in which microspore, generative cell, and sperm cells were used to study their transcriptomes. The microspores had the highest diversity with 30906 genes expressed, followed by generative cell and sperm cell, with sperm cell expressing 16561 genes (Liu et al., 2018). The authors of this study were able to conclude that the development of the sperm cell lineage in tomato was marked by a global reduction in expression of genes involved in pluripotency, somatic development, and metabolism; a molecular signature previously known to be associated with germ cell commitment and gamete development in other plant and mammalian species. Extensive transcriptomic reprogramming occurred particularly after the asymmetric first division, leading to a coordinated upregulation of germ cell/sperm cell preferentially expressed genes (Liu et al., 2018).

Most of the plant species analyzed so far have isomorphic sperm cells. However, one of the most striking studies on sperm dimorphism comes from the only published report on *Plumbago*, in which transcriptomes of these two sperm cells were transcriptionally different, and reflect their destiny to fuse either with egg cell or central cell (Gou et al., 2009).

The number and complexity of sperm transcriptomes from several studies disprove the original assumption that the sperm cells are transcriptionally quiescent. In all the species analyzed, the sperm cell had a different transcriptome not only from the control vegetative/sporophytic tissues but also from the transcriptome of the whole pollen grain/vegetative nucleus. (Borges et al., 2008, Russel et al., 2012). Globally, the sperm cell transcriptome from all the species studied so far was enriched for genes involved in ubiquitin-mediated proteolysis, membrane-associated proteins, cell cycle, signal transduction, and epigenetic modifications (Engel et al., 2003; Okada et al., 2006, Okada et al., 2007; Borges et al., 2008; Xin et al., 2011; Russell et al., 2012; Anderson et al., 2013). On the other hand, genes for small RNA mediated silencing were largely downregulated (Borges et al., 2008; Russel et al., 2012, Anderson et al., 2013).

1.6. Regulation of gene expression during pollen and male germline development

Male gametophyte development is a tightly controlled process at least at the gene expression level. There are two major global transcriptional changes during the pollen developmental program: early and late, and the switching point occurs after the first asymmetric pollen division (PMI) (Twell et al., 2006). The late program is however characterized by maturation of the pollen grain rather than the timing of pollen mitosis II (Hafidh et al., 2012a), supporting the idea of having a unique pollen transcriptome during late development stages (Honys and Twell., 2004).

Both early and late pollen transcriptomes are characterized by genes that are involved in transcriptional regulation and cell cycle control, albeit with differential expression of various transcription factor gene families, such as bHLH, C2H2, bZIP, MYB/MYB related, AP2/EREBP, MADs, WRKY and TCP. On the other hand, protein synthesis/translation-related genes were found to be overrepresented in early pollen developmental stages, whereas genes associated with cytoskeleton, cell wall synthesis, signaling, protein turnover, and localization were found to be upregulated during late pollen development stages, particularly during pollen maturation and pollen tube growth (Twell et al., 2006; Wei et al., 2010; Hafidh et al., 2012a; Costa et al., 2013, Zhou et al., 2016).

Five members belonging to the MIKC subgroup of the MADS-box transcription factor families AGL65, AGL94, AGL30, AGL104, and AGL66 were found to be highly expressed in mature pollen grains (Pina et al., 2005). These genes were also later reported to form heterodimeric complexes preferentially binding MEF2-type CArG-box sequence motifs, as these motifs were also found to be over-represented in the promoter region of late pollen expressed genes (Verelst et al., 2007a). Several combinations of MIKC gene mutants and their transcriptome analysis revealed the complexity of the MADS-box transcription factor network necessary for regulating cellular differentiation and maturation during pollen development (Verelst et al., 2007b; Adamczyk and Fernandez 2009). However, while most of the studies in MIKC MADS-box transcription factors were done on Arabidopsis, the functional conservation of these transcription factor families in rice revealed that their function in male gametophyte development has been largely conserved since the divergence of monocots and eudicots, some 150 MYA (Liu et al., 2013). Several classes of regulatory proteins, such as transcription factors are usually found to be common between rice and Arabidopsis pollen transcriptomes, however, there was a unique set of TFs found only in rice indicating that regulation of gene expression differed in these two species

(Wei et al., 2010). The first evidence to support the conservation of transcription factor networks between rice and Arabidopsis was performed on MIKC MADS-box networks, a family of genes essential for the maturation of pollen grains (Liu et al., 2013). In Arabidopsis, there are five MIKC genes expressed in pollen, while rice only has 3 genes divided between two clades: S (OsMADS62, OsMADS63) and P (OsMADS68). The MIKC mutants belonging to rice P or S clade showed reduced pollen viability, defects in *in vitro* pollen germination, and abnormal starch accumulation similar to what was observed in Arabidopsis (Liu et al., 2013).

Some other classes of transcription factor families, such as basic leucine zipper TFs (bZIP), also have an important function in plant reproduction. For instance, AtbZIP34 is highly expressed during the late stages of pollen development with dual function in both gametophytic and sporophytic tissues. The transcriptome analysis of *atbzip*34 mutant pollen helped to identify a transcriptional network and regulon comprised of genes associated with membrane transport, cell wall synthesis, and lipid metabolism (Gibalová et al., 2009).

Sperm cells of Arabidopsis also express a large number of transcription factors (~ 250 TF), but the functions of only three have been studied to date. These three important transcription factors in Arabidopsis are DUO1, DAZ1, and DAZ2. They have been found to be critical for germ cell division, and differentiation of sperm cells, and they are considered to form an important regulatory hub in the male germline gene network (Rotman et al., 2005; Borg et al., 2014). Specifically, male germline-specific R2R3-MYB transcription factor DUO pollen 1 (DUO1) was found to act as a regulon in sperm cell differentiation (Durbarry et al., 2005; Rotman et al., 2005).

To better understand the function of DUO1, Borg et al., 2011 ectopically expressed DUO1 in seedlings and analyzed the transcriptome, which helped to identify 63 potential targets. In addition, DUO1 was shown to bind to

canonical MYB sites, thereby regulating its target promoters. Subsequently, DAZ1 and DAZ2 identified in the previous studies as two of the DUO1 target genes, were characterized as encoding EAR motif-containing C2H2 type zinc finger proteins that were involved in both generative cell division and germ cell differentiation (Borg et al., 2014). DUO1 is known to be a master regulator of male germ cell division, and in both bicellular and tricellular pollen, the function of DUO1 may be associated with the timing of generative cell division, particularly during the evolution of bicellular or tricellular pollen types (Rotman et al., 2005; Hafidh et al., 2012b; Higo et al., 2018).

1.7. Role of alternative splicing in plant reproduction

Alternative splicing is a process in which pre-messenger RNA is converted into mature mRNAs by removal of introns and joining of exons. Alternative splicing plays a very important role in expanding the coding capacity of genomes, thereby aiding the flow of genetic information from DNA to proteins (Naftelberg et al., 2015). This is an important step for the regulation of gene expression in all eukaryotes to determine cell and tissue specific features, response to external cues and for overall cell functioning. Alternative splicing in plants has largely remained an unexplored area of gene expression until very recently, as this phenomenon used to be considered rare (Reddy, 2007). However, recent studies at genome-wide level have revealed that alternative splicing in plants is far more predominant than previously thought. Most of the genome-wide studies in plants have focussed on plant stress responses. Moreover, in animals, tissue specific alternative splicing data is available (Wang et al., 2008b), but such information in plants is still limited (Martin et al., 2021), more so when it comes to reproductive tissues. For instance, mammalian spermatogenesis is well characterized, including the splicing landscape (Soumilton et al., 2013), but such information of alternative splicing during pollen development is scarce. Recently, genome-wide alternative splicing analysis was performed on five stages (pollen mother cell (PMC),

tetrad, microspore, bicelluar pollen and mature pollen grain) of pollen development in field mustard (*Brassica rapa*). The authors found significant increase in intron retention during meiotic stage transition from PMC to tetrad stage, with many of the genes having intron-retained isoforms resulting in non-functional proteins (Golicz et al., 2021). Moreover, genes with isoforms having intron retention at PMC to tetrad transition were enriched in functions related to mRNA transport and ribosome biogenesis, indicating their possible role in regulating the rate of translation and maintaining the architecture of the nuclear envelope during meiosis (Golicz et al., 2021). Studies on genome-wide alternative splicing in Arabidopsis pollen development are still unavailable due to the fact that until recently, no RNA-seq data for different pollen development stages was available. The first and only data from any plant reproductive cell type in Arabidopsis became available for mature pollen grains (Loraine et al., 2013), in which the authors identified 30 genes as differentially spliced between leaf and pollen grain. One of the genes, AtU2AF65a, was later characterised to have a role in regulating flowering time, while displaying redundant roles in pollen tube growth together with its other isoform AtU2AF65b (Park et al., 2019). Direct evidence to support the role of alternative splicing in plant reproduction is scarce and only some splicing factors have been characterised that may be involved in pollen germination and tube growth. For example, Arabidopsis serine/arginine-rich (SR) protein SR45 was found to alter F-actin dynamics via directly or indirectly affecting alternative splicing of some actin binding proteins, which may affect pollen germination and tube growth (Cao et al., 2013). Also, when it comes to differences between splicing landscapes in animals and plants, it was thought that animals have exon skipping and plants have intron retention as the most predominant form of alternative splicing events. This notion of intron retention to be the major splicing event in plants was contested recently as some studies have shown that intron retention could be overestimated in plants (Martin et al., 2021). While pollen is now known to undergo differential

splicing, it remains to be seen how genome-wide splicing differs at the resolution of sperm cell and vegetative nucleus.

1.8. Transcriptome analysis at single cell level

Single-cell genomics and epigenomics have opened up a wide area of research with enormous potential for plant biology. Several studies in maize and barley, for example, have helped to understand meiotic recombination and cross over in meiocytes at the single-cell level (Dreissig et al., 2015; Li et al., 2015; Dreissig et al., 2017; Luo et al., 2019). Application of these techniques have the potential to unveil the global transcriptome and genomic details at single-cell resolution both during vegetative and reproductive development. Maize is the first crop species whose pollen cells have been studied at the level of single-cell. Nelms and Walbot, recently applied single-cell RNA-seq to 144 early germinal cells in maize anthers and identified changes that occurred during meiosis. This was the first study highlighting the importance of understanding pollen development at single-cell level (Nelms and Walbot 2019). Apart from transcriptomic studies, recently the three-dimensional genome structure of rice gametes (sperm cells, egg cells, and zygotes) were analyzed using chromatin conformation capture and high throughput 3C (Hi-C). The authors in this study provided evidence of global changes in chromatin structure after fertilization (Zhou et al., 2019).

1.9. Epigenetic reprogramming during pollen development

DNA methylation profiles of sporophytic tissues have been extensively studied over the past years but methylomes from pure populations of reproductive cell types such as sperm, vegetative nucleus and egg cell have been limited by technical challenges in obtaining sufficient quantities of starting material. Despite significant progress in overcoming these challenges, DNA methylome data from sperm and pollen vegetative nucleus is only available for Arabidopsis (Calarco et al., 2012; Ibarra et al., 2012), rice

(Hsieh et al., 2016; Kim et al., 2019), and more recently tomato (Lu et al., 2021). The sperm cells in Arabidopsis have significantly reduced CHH methylation relative to the pollen vegetative nucleus, an effect that is primarily attributed to the activity of chromomethyltransferase CMT2 rather than RNA directed DNA methylation (RdDM) (Hsieh et al., 2016; Borges et al., 2018).

DNA methylation (mCHH and mCHG) levels in both rice and Arabidopsis sperm cells at specific loci correlate with reduced mCG methylation at the corresponding loci in the pollen vegetative nucleus. This is consistent with the hypothesis that the reduced DNA methylation in the pollen vegetative nucleus might facilitate the expression of siRNAs, which are then transported into the sperm cell to reinforce silencing of transposable elements through RNA directed DNA methylation in sperm cells (Calarco et al., 2012; Ibarra et al., 2012; Kim et al., 2019). A study of mutants in rice and Arabidopsis lacking ROS1 DNA glycosylase activity revealed a correlation between demethylation in the vegetative nucleus and increased methylation in sperm (Ibarra et al., 2012; Kim et al., 2019). This ROS1 mediated DNA demethylation in the vegetative nucleus was later shown to be essential in ensuring pollen tube progression (Khouider et al., 2021). Early studies of the DNA methylome from reproductive cell types were limited to tricellular pollen. The first study of this type on bicellular pollen was conducted in tomato, in which whole genome bisulphite sequencing was performed from all stages of pollen development: microspore, generative cell, vegetative cell as well as sperm cells (Lu et al., 2021). The authors found millions of differentially methylated sites when they compared: microspore and generative cell, microspore and vegetative nucleus, and vegetative nucleus and a generative cell. However, when they compared methylation between generative cell and sperm cell, only 62 differential sites were found. This indicates that the tomato sperm cell lineage undergoes extensive DNA methylation following asymmetric division, but methylation patterns are largely maintained during the second pollen mitotic division, with the resulting sperm cells inheriting largely the same DNA

methylome (Lu et al., 2021). This leads to the question of whether siRNA mediated transposable element silencing through increased DNA methylation in sperm cells is already pre-established in the generative cell. Some indirect evidence to support this hypothesis has been provided recently, since small RNAs were found to be active during early stages of pollen development (Oliver et al., 2019). Epigenetic reprogramming in the tomato germline is not limited to changes at the methylation level, but occurs also at the level of histone modifications**.** The development of the sperm cell lineage in tomato was found to be marked by the loss of histone marks such as H3K4me3 and H3K9ac responsible for activation of gene expression, whereas repressive histone marks such as H3K9 were found to be accumulated, indicating that both active as well as repressive histone marks are important for the development of the sperm cell lineage in tomato.

To conclude, male gametophyte development is one of the fundamental biological processes occurring in plants. A significant amount of effort has been put into understanding the molecular mechanism underlying this critical process, results of which have been crucial in driving the advancement in the plant reproduction field over the last two decades. This is particularly important as rising temperatures and resulting detrimental effects of climate change have begun to show its effect on pollen development, thereby affecting crop productivity. A wealth of genome-wide data at transcriptome, proteome and methylome level from 24 different angiosperm species has shed light on all stages of pollen and sperm development. This knowledge, first obtained mainly for the model plant *Arabidopsis thaliana*, then paved the way for studies in other agriculturally important crops such as rice, maize, and tomato. This wealth of resources has become the motivation for creating community pooled data bases, accelerating not only research on the molecular mechanisms underlying pollen development, but also on how this fundamental process evolved (Julca et al., 2021). Despite this tremendous

progress, several key facets of male gametophyte development remain unknown.

1.10. Thesis outline

Male gametophyte development in several plant species has been studied extensively. All previous studies though have focussed mostly on the gene expression level and very little information is available at the splicing level. Moreover, most of the studies have focussed on the sperm cells and transcriptomic data for vegetative cells is available only for rice. In order to bridge this knowledge gap, I have carried out this PhD thesis to address some of these missing links and to advance our understanding about transcriptome reprogramming of the sperm cells and vegetative nucleus. Not only this, I also provide for the first time insights into transcriptomic changes triggered during pollen-pistil interaction, and at a sub-cellular resolution.

In chapter 2, I address the question how the transcriptomes of sperm cells and vegetative nuclei isolated from *Arabidopsis thaliana* mature pollen grains differ. In addition, I compared sperm cell and egg cell transcriptomes to decipher differences in the transcriptomes of the two gametes. These analyses were carried out not only at gene level, but also at the level of alternative splicing. Comparison of sperm cell and vegetative nucleus revealed 6906 genes to be differentially expressed between sperm cell and vegetative nucleus. We also found 468 genes that were differentially spliced between sperm cell and vegetative nucleus, including several transcription and splicing factors being regulated both at the transcriptional level as well as at the splicing level. Moreover, a comparative analysis of splicing landscapes in reproductive cells between different species shed light on the extent on divergence between monocot and dicot species.

Chapter 3 of my thesis reveals the changes in the transcriptome of sperm cells and vegetative nucleus while the pollen tube carrying them travels within

the female tissues. I compared the transcriptome of sperm cells from mature pollen grains with that of pollen tubes that had passed through the pistil. Furthermore, in order to understand the changes that are elicited only in the presence of female cues, I also compared with the transcriptome of sperm cells obtained from pollen tubes grown on plate. The results revealed that the presence of female cues evokes unique changes, fine-tuning the transcriptome of sperm cells that may help them prepare for gamete fusion. I also performed similar comparative analysis for the vegetative nucleus and found that the transcriptome of vegetative nuclei from growing pollen tubes is distinct from that of vegetative nuclei from the mature pollen grain.

In chapter 4 I developed a method to analyse the transcriptome of Arabidopsis sperm cells at single cell resolution. I collected single sperm cells from pollen tubes grown through the pistil by FACS and tried to address the question whether the two isomorphic sperm cells differ at the gene expression level. Profiling 80 sperm cells at single cell level pointed towards a separation into two major clusters, however much higher number of sperm cells would have to be analysed to increase statistical power and to ascertain whether the transcriptomes of the two sperm cells differ.

In summary, I have developed a method to analyse the transcriptomes of Arabidopsis sperm cells and vegetative nuclei separately from minute amounts of isolated cells, by combining state-of-the-art FACS protocols with optimized RNA-seq library preparation. This has allowed me to analyse the differences in the transcriptomes of the male germ unit in mature pollen grains, not only at gene expression level, but more importantly at the level of alternative splicing. Moreover, I have revealed the dynamic changes in gene expression in male gametes and the vegetative nucleus undergo during the progamic phase, providing such high resolution for the first time.

My work provides a rich resource for future research in the field of plant reproduction, in particular male germline development, by identifying many candidate genes for in-depth functional characterization.

1.11. References

Adamczyk, B.J. and Fernandez, D.E., 2009. MIKC* MADS domain heterodimers are required for pollen maturation and tube growth in *Arabidopsis*. *Plant Physiology*, *149*(4), pp.1713-1723.

Anderson, S.N., Johnson, C.S., Jones, D.S., Conrad, L.J., Gou, X., Russell, S.D. and Sundaresan, V., 2013. Transcriptomes of isolated *Oryza sativa* gametes characterized by deep sequencing: evidence for distinct sex-dependent chromatin and epigenetic states before fertilization. *The Plant Journal*, *76*(5), pp.729-741.

Aya, K., Suzuki, G., Suwabe, K., Hobo, T., Takahashi, H., Shiono, K., Yano, K., Tsutsumi, N., Nakazono, M., Nagamura, Y. and Matsuoka, M., 2011. Comprehensive network analysis of anther-expressed genes in rice by the combination of 33 laser microdissection and 143 spatiotemporal microarrays. *PloS One*, *6*(10), p.e26162.

Barakate, A., Orr, J., Schreiber, M., Colas, I., Lewandowska, D., McCallum, N., Macaulay, M., Morris, J., Arrieta, M., Hedley, P.E. and Ramsay, L., 2020. Time-resolved transcriptome of barley anthers and meiocytes reveals robust and largely stable gene expression changes at meiosis entry. *bioRxiv*.

Bates, G.W., Rosenthal, D.M., Sun, J., Chattopadhyay, M., Peffer, E., Yang, J., Ort, D.R. and Jones, A.M., 2012. A comparative study of the *Arabidopsis thaliana* guard-cell transcriptome and its modulation by sucrose. *PLOS One*, *7*(11), p.e49641.

Becker, J.D., Boavida, L.C., Carneiro, J., Haury, M. and Feijó, J.A., 2003. Transcriptional profiling of *Arabidopsis* tissues reveals the unique characteristics of the pollen transcriptome. *Plant Physiology*, *133*(2), pp.713-725.

Becker, J.D., Takeda, S., Borges, F., Dolan, L. and Feijó, J.A., 2014. Transcriptional profiling of *Arabidopsis* root hairs and pollen defines an apical cell growth signature. *BMC Plant Biology*, *14*(1), pp.1-14.

Bokvaj, P., Hafidh, S. and Honys, D., 2015. Transcriptome profiling of male gametophyte development in *Nicotiana tabacum*. *Genomics Data*, *3*, pp.106-111.

Bordas-Le Floch, V., Le Mignon, M., Bouley, J., Groeme, R., Jain, K., Baron-Bodo, V., Nony, E., Mascarell, L. and Moingeon, P., 2015. Identification of novel short ragweed pollen allergens using combined transcriptomic and immunoproteomic approaches. *PLoS One*, *10*(8), p.e0136258.

Borg, M., Brownfield, L., Khatab, H., Sidorova, A., Lingaya, M. and Twell, D., 2011. The R2R3 MYB transcription factor DUO1 activates a male germline-specific regulon essential for sperm cell differentiation in *Arabidopsis*. *The Plant Cell*, *23*(2), pp.534-549.

Borg, M., Rutley, N., Kagale, S., Hamamura, Y., Gherghinoiu, M., Kumar, S., Sari, U., Esparza-Franco, M.A., Sakamoto, W., Rozwadowski, K. and Higashiyama, T., 2014. An EAR-dependent regulatory module promotes male germ cell division and sperm fertility in *Arabidopsis*. *The Plant Cell*, *26*(5), pp.2098-2113.

Borges, F., Gomes, G., Gardner, R., Moreno, N., McCormick, S., Feijó, J.A. and Becker, J.D., 2008. Comparative transcriptomics of *Arabidopsis* sperm cells. *Plant Physiology*, *148*(2), pp.1168-1181.

Borges, F., Parent, J.S., Van Ex, F., Wolff, P., Martínez, G., Köhler, C. and Martienssen, R.A., 2018. Transposon-derived small RNAs triggered by miR845 mediate genome dosage response in *Arabidopsis*. *Nature Genetics*, *50*(2), pp.186-192.

Brewbaker, J.L., 1967. The distribution and phylogenetic significance of binucleate and trinucleate pollen grains in the angiosperms. *American Journal of Botany*, *54*(9), pp.1069-1083.

Calarco, J.P., Borges, F., Donoghue, M.T., Van Ex, F., Jullien, P.E., Lopes, T., Gardner, R., Berger, F., Feijó, J.A., Becker, J.D. and Martienssen, R.A., 2012. Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. *Cell*, *151*(1), pp.194-205.

Campbell, B.C., Gilding, E.K., Timbrell, V., Guru, P., Loo, D., Zennaro, D., Mari, A., Solley, G., Hill, M.M., Godwin, I.D. and Davies, J.M., 2015. Total transcriptome, proteome, and allergome of Johnson grass pollen, which is important for allergic rhinitis in subtropical regions. *Journal of Allergy and Clinical Immunology*, *135*(1), pp.133-142.

Cao, L.J., Zhao, M.M., Liu, C., Dong, H.J., Li, W.C. and Ren, H.Y., 2013. LlSR28 is involved in pollen germination by affecting filamentous actin dynamics. *Molecular Plant*, *6*(4), pp.1163-1175.

Carmona, R., Zafra, A., Seoane, P., Castro, A.J., Guerrero-Fernández, D., Castillo-Castillo, T., Medina-García, A., Cánovas, F.M., Aldana-Montes, J.F., Navas-Delgado, I. and Alché, J.D.D., 2015. ReprOlive: a database with linked data for the olive tree (*Olea europaea* L.) reproductive transcriptome. *Frontiers in Plant Science*, *6*, p.625.

Chen, C., Farmer, A.D., Langley, R.J., Mudge, J., Crow, J.A., May, G.D., Huntley, J., Smith, A.G. and Retzel, E.F., 2010. Meiosis-specific gene

discovery in plants: RNA-Seq applied to isolated *Arabidopsis* male meiocytes. *BMC Plant Biology*, *10*(1), pp.1-14.

Chen, J., Strieder, N., Krohn, N.G., Cyprys, P., Sprunck, S., Engelmann, J.C. and Dresselhaus, T., 2017. Zygotic genome activation occurs shortly after fertilization in maize. *The Plant Cell*, *29*(9), pp.2106-2125.

Chen, Y., Zou, M. and Cao, Y., 2014. Transcriptome analysis of the *Arabidopsis* semi-in vivo pollen tube guidance system uncovers a distinct gene expression profile. *Journal of Plant Biology*, *57*(2), pp.93-105.

Chettoor, A.M., Givan, S.A., Cole, R.A., Coker, C.T., Unger-Wallace, E., Vejlupkova, Z., Vollbrecht, E., Fowler, J.E. and Evans, M.M., 2014. Discovery of novel transcripts and gametophytic functions via RNA-seq analysis of maize gametophytic transcriptomes. *Genome Biology*, *15*(7), pp.1-23.

Conze, L.L., Berlin, S., Le Bail, A. and Kost, B., 2017. Transcriptome profiling of tobacco (Nicotiana tabacum) pollen and pollen tubes. *BMC Genomics*, *18*(1), pp.1-11.

Costa, M., Nobre, M.S., Becker, J.D., Masiero, S., Amorim, M.I., Pereira, L.G. and Coimbra, S., 2013. Expression-based and co-localization detection of arabinogalactan protein 6 and arabinogalactan protein 11 interactors in *Arabidopsis* pollen and pollen tubes. *BMC Plant Biology*, *13*(1), pp.1-19.

Davidson, R.M., Hansey, C.N., Gowda, M., Childs, K.L., Lin, H., Vaillancourt, B., Sekhon, R.S., de Leon, N., Kaeppler, S.M., Jiang, N. and Buell, C.R., 2011. Utility of RNA sequencing for analysis of maize reproductive transcriptomes. *The Plant Genome*, *4*(3).

Dreissig, S., Fuchs, J., Cápal, P., Kettles, N., Byrne, E. and Houben, A., 2015. Measuring meiotic crossovers via multi-locus genotyping of single pollen grains in barley. *PLoS One*, *10*(9), p.e0137677.

Dreissig, S., Fuchs, J., Himmelbach, A., Mascher, M. and Houben, A., 2017. Sequencing of single pollen nuclei reveals meiotic recombination events at megabase resolution and circumvents segregation distortion caused by postmeiotic processes. *Frontiers in Plant Science*, *8*, p.1620.

Dukowic-Schulze, S., Sundararajan, A., Mudge, J., Ramaraj, T., Farmer, A.D., Wang, M., Sun, Q., Pillardy, J., Kianian, S., Retzel, E.F. and Pawlowski, W.P., 2014. The transcriptome landscape of early maize meiosis. *BMC Plant Biology*, *14*(1), pp.1-15.

Durbarry, A., Vizir, I. and Twell, D., 2005. Male germ line development in *Arabidopsis*. duo pollen mutants reveal gametophytic regulators of generative cell cycle progression. *Plant Physiology*, *137*(1), pp.297-307.

El Kelish, A., Zhao, F., Heller, W., Durner, J., Winkler, J.B., Behrendt, H., Traidl-Hoffmann, C., Horres, R., Pfeifer, M., Frank, U. and Ernst, D., 2014. Ragweed (*Ambrosia artemisiifolia*) pollen allergenicity: SuperSAGE transcriptomic analysis upon elevated CO_2 and drought stress. *BMC Plant Biology*, *14*(1), pp.1-16.

Engel, M.L., Chaboud, A., Dumas, C. and McCormick, S., 2003. Sperm cells of *Zea mays* have a complex complement of mRNAs. *The Plant Journal*, *34*(5), pp.697-707.

Fasoli, M., Dal Santo, S., Zenoni, S., Tornielli, G.B., Farina, L., Zamboni, A., Porceddu, A., Venturini, L., Bicego, M., Murino, V. and Ferrarini, A., 2012. The grapevine expression atlas reveals a deep transcriptome shift driving the entire plant into a maturation program. *The Plant Cell*, *24*(9), pp.3489-3505.

Fíla, J., Drábková, L.Z., Gibalová, A. and Honys, D., 2017. When simple meets complex: Pollen and the-omics. In *Pollen Tip Growth* (pp. 247-292). Springer, Cham.

Frank, G., Pressman, E., Ophir, R., Althan, L., Shaked, R., Freedman, M., Shen, S. and Firon, N., 2009. Transcriptional profiling of maturing tomato (*Solanum lycopersicum* L.) microspores reveals the involvement of heat shock proteins, ROS scavengers, hormones, and sugars in the heat stress response. *Journal of Experimental Botany*, *60*(13), pp.3891-3908.

Futamura, N., Ujino-Ihara, T., Nishiguchi, M., Kanamori, H., Yoshimura, K., Sakaguchi, M. and Shinohara, K., 2006. Analysis of expressed sequence tags from *Cryptomeria japonica* pollen reveals novel pollen-specific transcripts. *Tree Physiology*, *26*(12), pp.1517-1528.

Gibalová, A., Reňák, D., Matczuk, K., Dupl'áková, N., Cháb, D., Twell, D. and Honys, D., 2009. AtbZIP34 is required for *Arabidopsis* pollen wall patterning and the control of several metabolic pathways in developing pollen. *Plant Molecular Biology*, *70*(5), pp.581-601.

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

Golicz, A.A., Allu, A.D., Li, W., Lohani, N., Singh, M.B. and Bhalla, P.L., 2021. A dynamic intron retention program regulates the expression of several hundred genes during pollen meiosis. *Plant Reproduction*, pp.1-18.

Gou, X., Yuan, T., Wei, X. and Russell, S.D., 2009. Gene expression in the dimorphic sperm cells of *Plumbago zeylanica*: transcript profiling, diversity, and relationship to cell type. *The Plant Journal*, *60*(1), pp.33-47.

Haerizadeh, F., Wong, C.E., Bhalla, P.L., Gresshoff, P.M. and Singh, M.B., 2009. Genomic expression profiling of mature soybean (*Glycine max*) pollen. *BMC Plant Biology*, *9*(1), pp.1-12.

Hafidh, S., Breznenová, K., Růžička, P., Feciková, J., Čapková, V. and Honys, D., 2012a. Comprehensive analysis of tobacco pollen transcriptome unveils common pathways in polar cell expansion and underlying heterochronic shift during spermatogenesis. *BMC Plant Biology*, *12*(1), pp.1-20.

Hafidh, S., Breznenová, K. and Honys, D., 2012b. De novo post-pollen mitosis II tobacco pollen tube transcriptome. *Plant Signaling & Behavior*, *7*(8), pp.918-921.

Hedhly, A., 2011. Sensitivity of flowering plant gametophytes to temperature fluctuations. *Environmental and Experimental Botany*, *74*, pp.9-16.

Hedhly, A., Hormaza, J.I. and Herrero, M., 2009. Global warming and sexual plant reproduction. *Trends in Plant Science*, *14*(1), pp.30-36.

Herrero, M.P. and Johnson, R.R., 1980. High temperature stress and pollen viability of maize 1. *Crop Science*, *20*(6), pp.796-800.

Higo, A., Kawashima, T., Borg, M., Zhao, M., López-Vidriero, I., Sakayama, H., Montgomery, S.A., Sekimoto, H., Hackenberg, D., Shimamura, M. and Nishiyama, T., 2018. Transcription factor DUO1 generated by neo-functionalization is associated with evolution of sperm differentiation in plants. *Nature Communications*, *9*(1), pp.1-13.

Hirano, K., Aya, K., Hobo, T., Sakakibara, H., Kojima, M., Shim, R.A., Hasegawa, Y., Ueguchi-Tanaka, M. and Matsuoka, M., 2008. Comprehensive transcriptome analysis of phytohormone biosynthesis and signaling genes in microspore/pollen and tapetum of rice. *Plant and Cell Physiology*, *49*(10), pp.1429-1450.

Hobo, T., Suwabe, K., Aya, K., Suzuki, G., Yano, K., Ishimizu, T., Fujita, M., Kikuchi, S., Hamada, K., Miyano, M. and Fujioka, T., 2008. Various

spatiotemporal expression profiles of anther-expressed genes in rice. *Plant and Cell Physiology*, *49*(10), pp.1417-1428.

Hollender, C.A., Kang, C., Darwish, O., Geretz, A., Matthews, B.F., Slovin, J., Alkharouf, N. and Liu, Z., 2014. Floral transcriptomes in woodland strawberry uncover developing receptacle and anther gene networks. *Plant Physiology*, *165*(3), pp.1062-1075.

Honys, D. and Twell, D., 2003. Comparative analysis of the *Arabidopsis* pollen transcriptome. *Plant Physiology*, *132*(2), pp.640-652.

Honys, D. and Twell, D., 2004. Transcriptome analysis of haploid male gametophyte development in *Arabidopsis*. *Genome Biology*, *5*(11), pp.1-13.

Hsieh, P.H., He, S., Buttress, T., Gao, H., Couchman, M., Fischer, R.L., Zilberman, D. and Feng, X., 2016. *Arabidopsis* male sexual lineage exhibits more robust maintenance of CG methylation than somatic tissues. *Proceedings of the National Academy of Sciences*, *113*(52), pp.15132-15137.

Huang, J.C., Chang, L.C., Wang, M.L., Guo, C.L., Chung, M.C. and Jauh, G.Y., 2011. Identification and exploration of pollen tube small proteins encoded by pollination-induced transcripts. *Plant and Cell Physiology*, *52*(9), pp.1546-1559.

Iaria, D., Chiappetta, A. and Muzzalupo, I., 2016. De novo transcriptome sequencing of Olea europaea L. to identify genes involved in the development of the pollen tube. *The Scientific World Journal*, *2016*.

Ibarra, C.A., Feng, X., Schoft, V.K., Hsieh, T.F., Uzawa, R., Rodrigues, J.A., Zemach, A., Chumak, N., Machlicova, A., Nishimura, T. and Rojas, D., 2012. Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. *Science*, *337*(6100), pp.1360-1364.

Ikram, S., Durandet, M., Vesa, S., Pereira, S., Guerche, P. and Bonhomme, S., 2014. Functional redundancy and/or ongoing pseudogenization among F-box protein genes expressed in *Arabidopsis* male gametophyte. *Plant Reproduction*, *27*(2), pp.95-107.

Julca, I., Ferrari, C., Flores-Tornero, M., Proost, S., Lindner, A., Hackenberg, D., Steinbachová, L., Michaelidis, C., Pereira, S.G., Misra, C.S., Kawashima, T., Borg, M., Berger, F., Goldberg, J., Johnson, M., Honys, D., Twell, D., Sprunck, S., Dresselhaus, T., Becker, J.D., Mutwil, M., 2020. Comparative transcriptomic analysis reveals conserved transcriptional programs underpinning organogenesis and reproduction in land plants. *Nature Plants*, *7(8),* pp. 1143-1159.

Kakani, V.G., Prasad, P.V.V., Craufurd, P.Q. and Wheeler, T.R., 2002. Response of in vitro pollen germination and pollen tube growth of groundnut (*Arachis hypogaea* L.) genotypes to temperature. *Plant, Cell & Environment*, *25*(12), pp.1651-1661.

Kakani, V.G., Reddy, K.R., Koti, S., Wallace, T.P., Prasad, P.V.V., Reddy, V.R. and Zhao, D., 2005. Differences in in vitro pollen germination and pollen tube growth of cotton cultivars in response to high temperature. *Annals of Botany*, *96*(1), pp.59-67.

Kandasamy, M.K. and Kristen, U., 1989. Ultrastructural responses of tobacco pollen tubes to heat shock. *Protoplasma*, *153*(1-2), pp.104-110.

Kanter, U., Heller, W., Durner, J., Winkler, J.B., Engel, M., Behrendt, H., Holzinger, A., Braun, P., Hauser, M., Ferreira, F. and Mayer, K., 2013. Molecular and immunological characterization of ragweed (*Ambrosia artemisiifolia* L.) pollen after exposure of the plants to elevated ozone over a whole growing season. *PloS One*, *8*(4), p.e61518.

Khouider, S., Borges, F., LeBlanc, C., Ungru, A., Schnittger, A., Martienssen, R., Colot, V. and Bouyer, D., 2021. Male fertility in *Arabidopsis* requires active DNA demethylation of genes that control pollen tube function. *Nature Communications*, *12*(1), pp.1-10.

Kim, M.Y., Ono, A., Scholten, S., Kinoshita, T., Zilberman, D., Okamoto, T. and Fischer, R.L., 2019. DNA demethylation by ROS1a in rice vegetative cells promotes methylation in sperm. *Proceedings of the National Academy of Sciences*, *116*(19), pp.9652-9657.

Lang, V., Usadel, B. and Obermeyer, G., 2015. De novo sequencing and analysis of the lily pollen transcriptome: an open access data source for an orphan plant species. *Plant Molecular Biology*, *87*(1-2), pp.69-80.

Lee, J.Y. and Lee, D.H., 2003. Use of serial analysis of gene expression technology to reveal changes in gene expression in *Arabidopsis* pollen undergoing cold stress. *Plant Physiology*, *132*(2), pp.517-529.

Leydon, A.R., Beale, K.M., Woroniecka, K., Castner, E., Chen, J., Horgan, C., Palanivelu, R. and Johnson, M.A., 2013. Three MYB transcription factors control pollen tube differentiation required for sperm release. *Current Biology*, *23*(13), pp.1209-1214.

Li, X., Li, L. and Yan, J., 2015. Dissecting meiotic recombination based on tetrad analysis by single-microspore sequencing in maize. *Nature Communications*, *6*(1), pp.1-9.

Liang, Y., Tan, Z.M., Zhu, L., Niu, Q.K., Zhou, J.J., Li, M., Chen, L.Q., Zhang, X.Q. and Ye, D., 2013. MYB97, MYB101 and MYB120 function as male factors that control pollen tube-synergid interaction in *Arabidopsis thaliana* fertilization. *PLoS Genetics*, *9*(11), p.e1003933.

Lin, S.Y., Chen, P.W., Chuang, M.H., Juntawong, P., Bailey-Serres, J. and Jauh, G.Y., 2014. Profiling of translatomes of in vivo–grown pollen tubes

reveals genes with roles in micropylar guidance during pollination in *Arabidopsis*. *The Plant Cell*, *26*(2), pp.602-618.

Liu, L., Lu, Y., Wei, L., Yu, H., Cao, Y., Li, Y., Yang, N., Song, Y., Liang, C. and Wang, T., 2018. Transcriptomics analyses reveal the molecular roadmap and long non-coding RNA landscape of sperm cell lineage development. *The Plant Journal*, *96*(2), pp.421-437.

Liu, Y., Cui, S., Wu, F., Yan, S., Lin, X., Du, X., Chong, K., Schilling, S., Theißen, G. and Meng, Z., 2013. Functional conservation of MIKC*-Type MADS box genes in *Arabidopsis* and rice pollen maturation. *The Plant Cell*, *25*(4), pp.1288-1303.

Loraine, A.E., Blakley, I.C., Jagadeesan, S., Harper, J., Miller, G. and Firon, N., 2015. Analysis and visualization of RNA-Seq expression data using RStudio, Bioconductor, and Integrated Genome Browser. In *Plant Functional Genomics* (pp. 481-501). Humana Press, New York, NY.

Loraine, A.E., McCormick, S., Estrada, A., Patel, K. and Qin, P., 2013. RNA-seq of *Arabidopsis* pollen uncovers novel transcription and alternative splicing. *Plant Physiology*, *162*(2), pp.1092-1109.

Lu, Y., Song, Y., Liu, L. and Wang, T., 2021. DNA methylation dynamics of sperm cell lineage development in tomato. *The Plant Journal*, *105*(3), pp.565-579.

Luo, C., Li, X., Zhang, Q. and Yan, J., 2019. Single gametophyte sequencing reveals that crossover events differ between sexes in maize. *Nature Communications*, *10*(1), pp.1-8.

Luo, M., Taylor, J.M., Spriggs, A., Zhang, H., Wu, X., Russell, S., Singh, M. and Koltunow, A., 2011. A genome-wide survey of imprinted genes in rice seeds reveals imprinting primarily occurs in the endosperm. *PLoS Genetics*, *7*(6), p.e1002125.

Ma, J., Skibbe, D.S., Fernandes, J. and Walbot, V., 2008. Male reproductive development: gene expression profiling of maize anther and pollen ontogeny. *Genome Biology*, *9*(12), pp.1-17.

Marks, M.D., Wenger, J.P., Gilding, E., Jilk, R. and Dixon, R.A., 2009. Transcriptome analysis of *Arabidopsis* wild-type and gl3–sst sim trichomes identifies four additional genes required for trichome development. *Molecular Plant*, *2*(4), pp.803-822.

Martín, G., Márquez, Y., Mantica, F., Duque, P. and Irimia, M., 2021. Alternative splicing landscapes in *Arabidopsis thaliana* across tissues and stress conditions highlight major functional differences with animals. *Genome Biology*, *22*(1), pp.1-26.

Mascarenhas, J.P., 1990. Gene activity during pollen development. *Annual Review of Plant Biology*, *41*(1), pp.317-338.

McCormick, S., 1993. Male gametophyte development. *The Plant Cell*, *5*(10), p.1265.

Naftelberg, S., Schor, I.E., Ast, G. and Kornblihtt, A.R., 2015. Regulation of alternative splicing through coupling with transcription and chromatin structure. *Annual Review of Biochemistry*, *84*, pp.165-198.

Nelms, B. and Walbot, V., 2019. Defining the developmental program leading to meiosis in maize. *Science*, *364*(6435), pp.52-56.

Obermeyer, G., Fragner, L., Lang, V. and Weckwerth, W., 2013. Dynamic adaption of metabolic pathways during germination and growth of lily pollen tubes after inhibition of the electron transport chain. *Plant Physiology*, *162*(4), pp.1822-1833.

Okada, T., Bhalla, P.L. and Singh, M.B., 2006. Expressed sequence tag analysis of Lilium longiflorum generative cells. *Plant and Cell Physiology*, *47*(6), pp.698-705.

Okada, T., Singh, M.B. and Bhalla, P.L., 2007. Transcriptome profiling of *Lilium longiflorum* generative cells by cDNA microarray. *Plant cell Reports*, *26*(7), pp.1045-1052.

Oliver, C., Annacondia, M.L., Wang, Z., Slotkin, R.K., Köhler, C. and Martinez, G., 2019. MicroRNA function transitions from regulating developmental genes to transposable elements during the maturation of pollen. *bioRxiv*, p.864017.

Palanivelu, R. and Preuss, D., 2006. Distinct short-range ovule signals attract or repel *Arabidopsis thaliana* pollen tubes in vitro. *BMC Plant Biology*, *6*(1), pp.1-9.

Palanivelu, R. and Tsukamoto, T., 2012. Pathfinding in angiosperm reproduction: pollen tube guidance by pistils ensures successful double fertilization. *Wiley Interdisciplinary Reviews: Developmental Biology*, *1*(1), pp.96-113.

Park, H.Y., Lee, H.T., Lee, J.H. and Kim, J.K., 2019. *Arabidopsis* U2AF65 regulates flowering time and the growth of pollen tubes. *Frontiers in Plant Science*, *10*, p.569.

Peng, H., Chun, J., Ai, T.B., Tong, Y.A., Zhang, R., Zhao, M.M., Chen, F. and Wang, S.H., 2012. MicroRNA profiles and their control of male gametophyte development in rice. *Plant Molecular Biology*, *80*(1), pp.85-102.

Pina, C., Pinto, F., Feijó, J.A. and Becker, J.D., 2005. Gene family analysis of the *Arabidopsis* pollen transcriptome reveals biological implications for cell growth, division control, and gene expression regulation. *Plant Physiology*, *138*(2), pp.744-756.

Qin, Y., Leydon, A.R., Manziello, A., Pandey, R., Mount, D., Denic, S., Vasic, B., Johnson, M.A. and Palanivelu, R., 2009. Penetration of the stigma and style elicits a novel transcriptome in pollen tubes, pointing to genes critical for growth in a pistil. *PLoS Genetics*, *5*(8), p.e1000621.

Reddy, A.S., 2007. Alternative splicing of pre-messenger RNAs in plants in the genomic era. *Annual Review of Plant Biology*, *58*, pp.267-294.

Redman, J.C., Haas, B.J., Tanimoto, G. and Town, C.D., 2004. Development and evaluation of an *Arabidopsis* whole genome Affymetrix probe array. *The Plant Journal*, *38*(3), pp.545-561.

Reňák, D., Dupl'áková, N. and Honys, D., 2012. Wide-scale screening of T-DNA lines for transcription factor genes affecting male gametophyte development in *Arabidopsis*. *Sexual Plant Reproduction*, *25*(1), pp.39-60.

Rogers, H.J., Harvey, A. and Lonsdale, D.M., 1992. Isolation and characterization of a tobacco gene with homology to pectate lyase which is specifically expressed during microsporogenesis. *Plant Molecular Biology*, *20*(3), pp.493-502.

Rotman, N., Durbarry, A., Wardle, A., Yang, W.C., Chaboud, A., Faure, J.E., Berger, F. and Twell, D., 2005. A novel class of MYB factors controls sperm-cell formation in plants. *Current Biology*, *15*(3), pp.244-248.

Russell, S.D., Bhalla, P.L. and Singh, M.B., 2008. Transcriptome-based examination of putative pollen allergens of rice (*Oryza sativa* ssp. japonica). *Molecular Plant*, *1*(5), pp.751-759.

Russell, S.D., Gou, X., Wong, C.E., Wang, X., Yuan, T., Wei, X., Bhalla, P.L. and Singh, M.B., 2012. Genomic profiling of rice sperm cell transcripts reveals conserved and distinct elements in the flowering plant male germ lineage. *New Phytologist*, *195*(3), pp.560-573.

Rutley, N. and Twell, D., 2015. A decade of pollen transcriptomics. *Plant Reproduction*, *28*(2), pp.73-89.

Sanetomo, R. and Hosaka, K., 2013. Pollen transcriptome analysis of *Solanum tuberosum* (2 n= 4 x= 48), *S. demissum* (2 n= 6 x= 72), and their reciprocal F 1 hybrids. *Plant Cell Reports*, *32*(5), pp.623-636.

Sato, S., Peet, M.M. and Thomas, J.F., 2002. Determining critical pre-and post-anthesis periods and physiological processes in *Lycopersicon esculentum* Mill. exposed to moderately elevated temperatures. *Journal of Experimental Botany*, *53*(371), pp.1187-1195.

Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M., Schölkopf, B., Weigel, D. and Lohmann, J.U., 2005. A gene expression map of *Arabidopsis thaliana* development. *Nature Genetics*, *37*(5), pp.501-506.

Schrauwen, J.A.M., De Groot, P.F.M., Van Herpen, M.M.A., van der Lee, V.T., Reynen, W.H., Weterings, K.A.P. and Wullems, G.J., 1990. Stage-related expression of mRNAs during pollen development in lily and tobacco. *Planta*, *182*(2), pp.298-304.

Schulten, V., Greenbaum, J.A., Hauser, M., McKinney, D.M., Sidney, J., Kolla, R., Arlehamn, C.S.L., Oseroff, C., Alam, R., Broide, D.H. and Ferreira, F., 2013. Previously undescribed grass pollen antigens are the major inducers of T helper 2 cytokine-producing T cells in allergic individuals. *Proceedings of the National Academy of Sciences*, *110*(9), pp.3459-3464.

Shen, Y., Venu, R.C., Nobuta, K., Wu, X., Notibala, V., Demirci, C., Meyers, B.C., Wang, G.L., Ji, G. and Li, Q.Q., 2011. Transcriptome dynamics through alternative polyadenylation in developmental and environmental responses in plants revealed by deep sequencing. *Genome Research*, *21*(9), pp.1478-1486.

Soumillon, M., Necsulea, A., Weier, M., Brawand, D., Zhang, X., Gu, H., Barthes, P., Kokkinaki, M., Nef, S., Gnirke, A. and Dym, M., 2013. Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. *Cell Reports*, *3*(6), pp.2179-2190.

Stinson, J. and Mascarenhas, J.P., 1985. Onset of alcohol dehydrogenase synthesis during microsporogenesis in maize. *Plant Physiology*, *77*(1), pp.222-224.

Suwabe, K., Suzuki, G., Takahashi, H., Shiono, K., Endo, M., Yano, K., Fujita, M., Masuko, H., Saito, H., Fujioka, T. and Kaneko, F., 2008. Separated transcriptomes of male gametophyte and tapetum in rice: validity of a laser microdissection (LM) microarray. *Plant and Cell Physiology*, *49*(10), pp.1407-1416.

Tang, X., Zhang, Z.Y., Zhang, W.J., Zhao, X.M., Li, X., Zhang, D., Liu, Q.Q. and Tang, W.H., 2010. Global gene profiling of laser-captured pollen mother cells indicates molecular pathways and gene subfamilies involved in rice meiosis. *Plant Physiology*, *154*(4), pp.1855-1870.

Tran, F., Penniket, C., Patel, R.V., Provart, N.J., Laroche, A., Rowland, O. and Robert, L.S., 2013. Developmental transcriptional profiling reveals key insights into Triticeae reproductive development. *The Plant Journal*, *74*(6), pp.971-988.

Tsubomura, M., Kurita, M. and Watanabe, A., 2016. Determination of male strobilus developmental stages by cytological and gene expression analyses in Japanese cedar (*Cryptomeria japonica*). *Tree Physiology*, *36*(5), pp.653-666.

Twell, D., Oh, S.A. and Honys, D., 2006. Pollen development, a genetic and transcriptomic view. In *The pollen tube* (pp. 15-45). Springer, Berlin, Heidelberg.

Twell, D., Wing, R., Yamaguchi, J. and McCormick, S., 1989. Isolation and expression of an anther-specific gene from tomato. *Molecular and General Genetics Mgg*, *217*(2), pp.240-245.

Van Herpen, M.M.A., De Groot, P.F.M., Schrauwen, J.A.M., Van den Heuvel, K.J.P.T., Weterings, K.A.P. and Wullems, G.J., 1992. In-vitro culture of tobacco pollen: gene expression and protein synthesis. *Sexual Plant Reproduction*, *5*(4), pp.304-309.

Verelst, W., Saedler, H. and Münster, T., 2007a. MIKC* MADS-protein complexes bind motifs enriched in the proximal region of late pollen-specific *Arabidopsis* promoters. *Plant Physiology*, *143*(1), pp.447-460.

Verelst, W., Twell, D., de Folter, S., Immink, R., Saedler, H. and Münster, T., 2007b. MADS-complexes regulate transcriptome dynamics during pollen maturation. *Genome Biology*, *8*(11), pp.1-15.

Vermeulen, S.J., Campbell, B.M. and Ingram, J.S., 2012. Climate change and food systems. *Annual Review of Environment and Resources*, *37*, pp.195-222.

Wang, Y., Zhang, W.Z., Song, L.F., Zou, J.J., Su, Z. and Wu, W.H., 2008a. Transcriptome analyses show changes in gene expression to accompany pollen germination and tube growth in *Arabidopsis*. *Plant Physiology*, *148*(3), pp.1201-1211.

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P. and Burge, C.B., 2008b. Alternative isoform regulation in human tissue transcriptomes. *Nature*, *456*(7221), pp.470-476.

Wang, W., Sheng, X., Shu, Z., Li, D., Pan, J., Ye, X., Chang, P., Li, X. and Wang, Y., 2016. Combined cytological and transcriptomic analysis reveals a nitric oxide signaling pathway involved in cold-inhibited *Camellia sinensis* pollen tube growth. *Frontiers in Plant Science*, *7*, p.456.

Warman, C., Panda, K., Vejlupkova, Z., Hokin, S., Unger-Wallace, E., Cole, R.A., Chettoor, A.M., Jiang, D., Vollbrecht, E., Evans, M.M. and Slotkin, R.K., 2020. High expression in maize pollen correlates with genetic contributions to pollen fitness as well as with coordinated transcription from neighboring transposable elements. *PLoS Genetics*, *16*(4), p.e1008462.

Wei, L.Q., Xu, W.Y., Deng, Z.Y., Su, Z., Xue, Y. and Wang, T., 2010. Genome-scale analysis and comparison of gene expression profiles in developing and germinated pollen in *Oryza sativa*. *BMC Genomics*, *11*(1), pp.1-20.

Whittle, C.A., Malik, M.R., Li, R. and Krochko, J.E., 2010. Comparative transcript analyses of the ovule, microspore, and mature pollen in *Brassica napus*. *Plant Molecular Biology*, *72*(3), pp.279-299.

Williams, J.H., Taylor, M.L. and O'Meara, B.C., 2014a. Repeated evolution of tricellular (and bicellular) pollen. *American Journal of Botany*, *101*(4), pp.559-571.

Williams, J.S., Der, J.P., dePamphilis, C.W. and Kao, T.H., 2014b. Transcriptome analysis reveals the same 17 S-locus F-box genes in two haplotypes of the self-incompatibility locus of *Petunia inflata*. *The Plant Cell*, *26*(7), pp.2873-2888.

Willing, R.P. and Mascarenhas, J.P., 1984. Analysis of the complexity and diversity of mRNAs from pollen and shoots of *Tradescantia*. *Plant Physiology*, *75*(3), pp.865-868.

Willing, R.P., Bashe, D. and Mascarenhas, J.P., 1988. An analysis of the quantity and diversity of messenger RNAs from pollen and shoots of Zea mays. *Theoretical and Applied Genetics*, *75*(5), pp.751-753.

Wu, J., Shahid, M.Q., Guo, H., Yin, W., Chen, Z., Wang, L., Liu, X. and Lu, Y., 2014. Comparative cytological and transcriptomic analysis of pollen

development in autotetraploid and diploid rice. *Plant Reproduction*, *27*(4), pp.181-196.

Xin, H.P., Peng, X.B., Ning, J., Yan, T.T., Ma, L.G. and Sun, M.X., 2011. Expressed sequence-tag analysis of tobacco sperm cells reveals a unique transcriptional profile and selective persistence of paternal transcripts after fertilization. *Sexual Plant Reproduction*, *24*(1), pp.37-46.

Xu, X.H., Chen, H., Sang, Y.L., Wang, F., Ma, J.P., Gao, X.Q. and Zhang, X.S., 2012. Identification of genes specifically or preferentially expressed in maize silk reveals similarity and diversity in transcript abundance of different dry stigmas. *BMC Genomics*, *13*(1), pp.1-17.

Zhang, H., Egger, R.L., Kelliher, T., Morrow, D., Fernandes, J., Nan, G.L. and Walbot, V., 2014. Transcriptomes and proteomes define gene expression progression in pre-meiotic maize anthers. *G3: Genes, Genomes, Genetics*, *4*(6), pp.993-1010.

Zhou, H., Yin, H., Chen, J., Liu, X., Gao, Y., Wu, J. and Zhang, S., 2016. Gene-expression profile of developing pollen tube of *Pyrus bretschneideri*. *Gene Expression Patterns*, *20*(1), pp.11-21.

Zhou, S., Jiang, W., Zhao, Y. and Zhou, D.X., 2019. Single-cell three-dimensional genome structures of rice gametes and unicellular zygotes. *Nature Plants*, *5*(8), pp.795-800.

Zinn, K.E., Tunc-Ozdemir, M. and Harper, J.F., 2010. Temperature stress and plant sexual reproduction: uncovering the weakest links. *Journal of Experimental Botany*, *61*(7), pp.1959-1968.

Chapter 2

Cell-type specific alternative splicing in the Arabidopsis germline

2.1. Abstract

During sexual reproduction in flowering plants, the two haploid sperm cells embedded within the cytoplasm of a growing pollen tube are carried to the embryo sac for double fertilization. Pollen development in flowering plants is a dynamic process that encompasses changes at transcriptome and epigenome level. While the transcriptome of pollen and sperm cells in *Arabidopsis thaliana* is well documented, previous analyses were mostly based on expression at gene level. Moreover, in-depth transcriptome analysis, particularly the extent of alternative splicing at the resolution of sperm cell and vegetative nucleus is still lacking. Therefore, we performed high throughput RNA-seq analysis to generate a spliceome map of Arabidopsis sperm cells and vegetative nuclei isolated from mature pollen grains. Overall, we identified 58039 transcripts, a 20% increase over the reference annotation. Out of the 9681 potential novel transcripts, 2091 and 3600 were expressed in sperm cell and vegetative nucleus, respectively. The differential splicing analysis helped to uncover 468 genes that were regulated both at gene level as well as splicing level. Many of these genes were found to have functions associated with mRNA splicing, chromatin modification, and protein localization. The data in this study provide novel insights into a gamete specific alternative splicing landscape, identifying several potential candidates for further molecular and functional characterization.

2.2. Introduction

Male gametophyte development in flowering plants starts with haploid microspores, which undergo two consecutive pollen mitotic divisions to form mature pollen grains. A mature pollen grain contains two sperm cells, partially embedded within the vegetative cell of the pollen grain. The vegetative nucleus drives pollen tube growth, delivering the two sperm cells to the female gametophyte for double fertilization (reviewed in Berger and Twell, 2011).

Transcriptomes from all stages of pollen have been extensively studied in *Arabidopsis thaliana* (Honys and Twell, 2004; Pina et al., 2005; Borges et al., 2008). These studies were complemented by transcriptome analyses of pollen and sperm cells from other flowering plants, such as rice (Anderson et al., 2013), maize (Chen et al., 2017), and tomato (Liu et al., 2018). It is worth noting that while studies have reported gene expression levels in pollen and sperm, except for rice, there is no transcriptome data available for the vegetative nucleus to date. Initially, studies on gene expression from pollen and sperm were performed based on microarrays, but RNA-seq is the more commonly used technology nowadays. Reduction in cost coupled with high throughput transcriptomic analysis through RNA-seq offers several advantages over traditionally used DNA microarrays and one of them is detection of alternative splice variants that are difficult to detect using arrays.

Alternative splicing (AS) is a mechanism of gene expression in which more than one unique mRNA species is generated from a single gene (Baralle, and Giudice, 2017). This enables cells to generate a vast diversity of proteins necessary for normal growth and development of an organism. Alternative splicing plays an important role in plant growth and development, regulating the circadian clock, flowering and stress responses (Howard et al., 2013; Reddy et al., 2013; Staiger and Brown, 2013). In humans, 95% of intron-containing genes are alternatively spliced (Wang et al., 2008). Recent analyses of splicing in *Arabidopsis thaliana* (Filichkin et al., 2010; Marquez et

al., 2012), cotton (Li et al., 2014), soybean (Shen et al., 2014), maize (Thatcher et al., 2014), and rice (Zhang et al., 2010) have highlighted the prevalence of AS in plants. It is estimated that up to 60% intron containing genes in Arabidopsis undergo alternative splicing (Marquez et al., 2012) and this number varies between plants. For instance, it is ~52% in soybean (Shen et al., 2014), ~40% in cotton (Li et al., 2014) and maize (Thatcher et al., 2014), and ~33% in rice (Zhang et al., 2010). The most predominant AS event in humans is exon-skipping, whereas in plants intron retention is considered to be the most predominant form of AS event, a notion that has recently been contested though (Martín et al., 2021). More and more studies now predict that intron retention could be overestimated in plants, and non-intron retention events, such as exon skipping, could be more prevalent than originally thought (Martin et al., 2021). For instance, alternative donor was found to be the most predominant form of AS during sugarcane smut disease infection (Bedre et al., 2019) and in barley meiocytes (Barakate et al.,2020). In addition, exon skipping formed the most predominant form of splicing event in developing maize ears (Thatcher et al., 2016), implying that tissue specific AS is more diverse than previously considered.

The advent of deep sequencing has revealed in an unbiased manner how alternative splicing is coordinated during plant growth and development, and complementary studies aiming at understanding the functions of alternative splicing at a global level are also being carried out. For instance, genome wide analysis of pollen specific AS in Arabidopsis identified 30 genes to be differentially spliced between leaf and pollen (Loraine et al., 2013). One of the genes, AtU2AF65a was identified to be differentially spliced between leaf and pollen and was later characterised to have a role in regulating flowering time, while displaying redundant roles in pollen tube growth together with its other isoform AtU2AF65b (Park et al., 2019).

Currently there are very few studies that provide direct evidence to support a role of alternative splicing in plant reproduction. Only recently, the functions of some splicing factors have started to be uncovered, particularly those that could be involved in plant reproduction. For instance, in an Arabidopsis *sr45* mutant, pollen germinated earlier than that of wild-type. AS analysis of *sr45* mutant showed that Arabidopsis villin1 (AtVLN1), an actin binding protein responsible for pollen tube germination and growth, was affected. Moreover, in *sr45*, the expression level of AtVLN1 full length transcript was increased, while the transcript that produced a truncated protein decreased (Cao et al., 2013). These results indicate that SR45 might alter F-actin dynamics by directly or indirectly affecting alternative splicing of AtVLN1 and some other actin binding proteins (Cao et al., 2013).

Alternative splicing can occur either transcriptionally or co-transcriptionally, regulating development at specific time points, under stress conditions, or exerting a tissue specific response. Genome wide analysis of AS in different tissues in human is well documented (Xu et al., 2002; Yeo et al., 2004, Pan et al., 2008), whereas in plants analysis of AS is limited to stress responses (Mandadi and Scholthof, 2015; Keller et al., 2017; Calixto et al., 2018; Huang et al., 2020), while tissue specific AS analysis is limited (Martin et al., 2021).

In this study, we have used RNA-seq to study the transcriptome of Arabidopsis sperm cells and vegetative nucleus. While the transcriptome of Arabidopsis sperm cells was available for more than a decade (Borges et al., 2008), this is the first study to describe the transcriptome of the vegetative nucleus in Arabidopsis. More importantly, we used our RNA-seq data to uncover genome wide alternative splicing in Arabidopsis sperm cells and vegetative nuclei. Combining our data with a published dataset for Arabidopsis egg cells (Zhao et al., 2019) helped to uncover the genome wide splicing landscape in gametes, thereby expanding the repertoire of genes

involved in plant reproduction and their possible regulation through alternative splicing.

2.3. Material and Methods

2.3.1. Plant Material and growth Conditions

Seeds of a transgenic marker line harbouring *MGH3p::MGH3-GFP* and *ACT11p::H2B-mRFP* (Borges et al., 2012) were sown on soil in short-day conditions (8-h light at 21–23 °C) for 8 weeks and then transferred to long-day conditions (16-h light) to induce flowering.

2.3.2. RNA-seq library preparation and reference-based transcriptome reassembly

Fluorescence activated cell sorting was used to sort sperm cell and vegetative nucleus from mature pollen grains (Borges et al., 2012; Santos et al., 2017). 500 sperm cells and vegetative nuclei, respectively, were sorted in lysis buffer, and full length Smart-seq2 cDNA libraries were produced as described in Misra et al., 2019.

cDNA libraries were checked on a Fragment Analyser for quality and yield and then converted into Nextera libraries at the Genomics Facility of Instituto Gulbenkian de Ciência, Oeiras, Portugal, and sequenced at Novogene, United Kingdom, with 150 bp paired-end reads on an Illumina Novaseq. The reads were checked for quality and then mapped to the Arabidopsis genome (TAIR 10) assembly using Hisat2 v2.1.0 with command line options “–dta --min-intronlen 60 --max-intronlen 6000 (Kim et al., 2019).

Transcriptomes from each RNA-seq library were reconstructed using reference-based transcriptome assembly by Stringtie v2.0 using the parameters -c 5 -a 15 -j 10 (Pertea et al., 2015). The assembled transcripts were then merged to form a master annotation using the merge option of

Stringtie. The merged annotation was compared to reference annotation using gffcompare and the assembled transcripts were then identified as known or novel isoform based on class code annotation (Pertea et al., 2016). In the new annotation, while the known transcripts retained the names of genes from the reference Araport11 annotation, the novel transcripts were named according to Stringtie´s default naming convention (eg., MSTRG.1090.1). Only transcripts having an expression of TPM>1 were retained for downstream analysis.

2.3.3. Identification of differential gene expression and splicing

Differential expression and alternative splicing analyses were carried out using 3D-RNAseq app (Guo et al., 2020). The new transcriptome assembly was used to quantify transcripts per million using Salmon (version 0.14.1) for each individual sample (Patro et al., 2017). Low expressed genes or transcripts were filtered out and only those genes were retained that had expression of 1 cpm (count per million ≥1) in at least 2 samples. Normalisation of expressed transcripts was performed using trimmed mean of M-values (TMM) method, followed by pair-wise analyses.

Differential gene expression analysis between two groups was performed using limma pipeline with thresholds of adjusted p value <0.01 and fold change log2FC >1. In order to identify the differentially alternatively spliced genes between two contrast group, adjusted p value <0.01 and at least one transcript of the gene with percent spliced-in $\Delta PS > 0.1$ were used.

2.3.4. Gene ontology analysis

Functional enrichment of genes that were differentially expressed or spliced in pair-wise comparisons were determined using an updated version of g:Profiler (Raudvere et al., 2019), with thresholds of adjusted p value <0.001 and <0.05, respectively (significance threshold method: g_SCS).

2.3.5. Detection of alternative splicing events

Different types of alternative splicing events were identified using Suppa2 (Trincado et al., 2018). Major alternative splicing events were classified into intron retention (IR), exon skipping (ES), alternative 3′ splice site or alternative acceptor (A3SS/AA), and alternative 5′ splice site or alternative donor (A5SS/AD). In order to identify isoform switching events, we used default parameters of 3D-RNaseq tool that uses the salmon output to detect significant isoform switches between conditions using inbuilt IsokTSP tool (Sebestyénet al, 2015; Guo et al, 2020).

2.4. Results

2.4.1. Transcriptome of sperm cells and vegetative nucleus

We performed RNA-seq on sperm cells and vegetative nuclei isolated from mature pollen grains by FACS, aiming at an average of 30 million 150bp paired end reads per sample. A summary of FACS sorting, sequencing and data analysis is shown in Figure 2.1. We also used published datasets from egg cells (Zhao et al., 2019) to compare the cell specific transcriptome profiles of the two germ cell types. In order to understand how the transcriptome of sperm cell and vegetative nucleus differed and to visualize overall differences between the male and female germline, we performed hierarchical clustering using all three sample types. The analysis revealed that the transcriptome of male and female germ cells clustered separately, whereas sperm cells and vegetative nuclei are clustered together; an observation that can be explained by sperm cells (SC) and vegetative nucleus (VN) deriving from the same precursor, the microspore (Figure 2.2A).

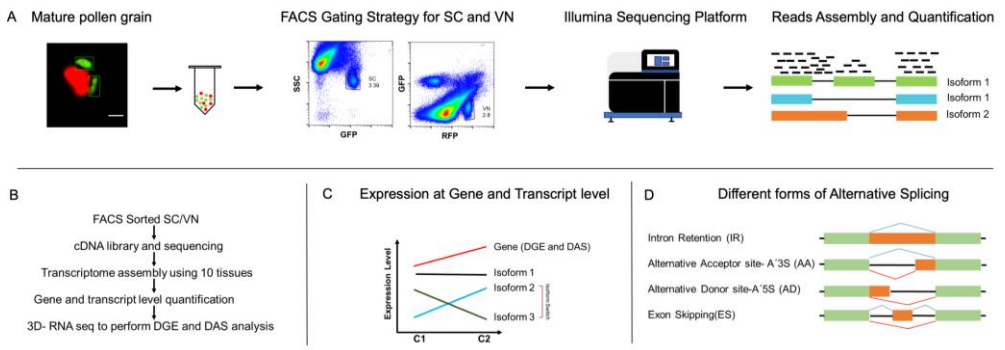


Figure 2.1: Flow diagram summarizing steps for the analysis of genome wide alternative splicing in Arabidopsis sperm cell and vegetative nucleus. A) Scheme showing FACS sorting strategy of GFP positive (*MGH3p::MGH3-eGFP*) sperm and RFP positive (*ACT11p::H2B-mRFP*) vegetative nucleus, followed by RNA-seq library preparation, sequencing, read assembly, and quantification. **B)** Flow chart depicting pipeline for analysis of alternative splicing. **C)** Example of a gene undergoing differential expression at the gene level or alternative splicing level. **D)** Different forms of alternative splicing events. SC sperm cells, VN vegetative nucleus, SSC side scatter, DGE differential gene expression, DAS differential alternative splicing.

Our data suggest that the transcriptome diversity in sperm cells was reduced when compared to both vegetative nuclei and egg cells (Figure 2.2B). The number of expressed genes in the three cell types (average expression TPM>1) ranged from 7445 in sperm cells, over 9310 in vegetative nuclei to 14925 in egg cells (Figure 2.2C). A considerable overlap in the gene expression profiles was observed between the three cell types, but each of them also expressed transcripts unique to them (597 in SC, 1657 in VN, and 6391 in EC, respectively), indicating that these transcripts in particular may encode proteins performing functions specific to the respective cell type (Figure 2.2C).

The transcriptome of Arabidopsis gametes has been known for almost a decade (Borges et al., 2008; Wuest et al., 2010). However, due to limitations posed by the microarray platforms used in these studies, the number of genes that are expressed in sperm cells and egg cell were likely to be underestimated. Some findings based on our analysis of the sperm cell transcriptome using RNA-seq are: 1) 1288 genes represented as probes on the arrays, but called absent in our array study (Borges et al., 2008), have been detected as expressed here (Figure 2.2D), 2) RNA-seq helped to uncover expression of an additional 2788 genes for which there were no probes on the arrays (Figure 2.2D), 729 genes of which with high expression (TPM>10) in sperm cells, including the well-known SC specific transcription factor DUO1 (Rotman et al., 2005; Brownfield et al., 2009a; Brownfield et al., 2009 b) Interestingly, our results also show some genes that are exclusively detected as sperm expressed using arrays, potentially due to some non-protein coding genes that were present on arrays, and/or genes are obsolete and no longer part of the new Araport11 annotation.

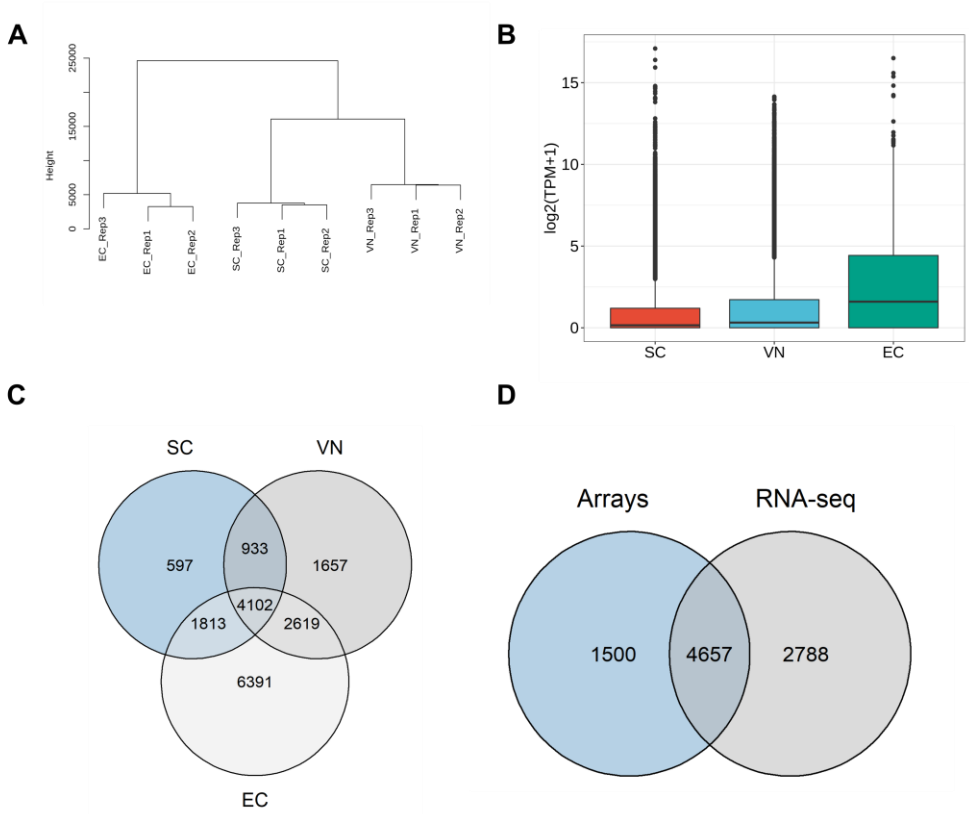


Figure 2.2: Transcriptome of Arabidopsis sperm cell, vegetative nucleus and egg cell. A) Hierarchical clustering of the three different cell types **B)** Box plot representation of gene expression in sperm cell, vegetative nucleus and egg cell. **C)** Venn diagram showing overlap between expressed genes between the three cell types at average TPM>1. **D)** Venn diagram showing comparison of transcriptome profiles of sperm cells obtained from microarrays (Borges et al., 2008) and RNA-seq. SC sperm cell, VN vegetative nucleus, EC egg cell.

2.4.2. Isoform level quantification in Arabidopsis gametes

Transcriptome analysis using RNA-seq facilitates expression quantification not only at gene level, but also at the isoform level. In order to perform isoform level quantification, we first performed a de novo transcriptome assembly

using sperm cell and vegetative nuclei data produced in this study, in combination with published datasets from 10 other tissues, focussing mostly on reproductive tissues (Supplemental Table S2.1).

We did this because we anticipated that the Arabidopsis gametes, both male and female, may potentially express a set of specific transcripts not previously identified in any other tissues studied. In addition, the most recent Arabidopsis Annotation, Araport 11 (Cheng et al., 2017), was based on vegetative tissues and various stress treatment RNA-seq data, and lacks data particularly from reproductive cell types such as sperm, egg cells and vegetative nucleus, as those were unavailable at the time. This suggests some novel isoform might be missing even in the recent most comprehensive Araport annotation. A total of 1.17 billion reads were used to reassemble the Arabidopsis transcriptome (Supplemental Table S2.1).

The transcripts were assembled for each sample using Stringtie and a total of 58039 non-redundant transcripts were generated across all 12 tissue/cell types. These transcripts mapped to 27655 loci. A total of 20861 (75%) gene models contained at least one intron. We recovered a total of 48358 known transcripts and 9681 novel transcripts.

If we focus on the three cell types in this study, the number of expressed transcripts (average TPM>1) ranged from 11113 (7135 loci) in sperm cells, over 15265 (8916 loci) in vegetative nuclei to 25222 (14805 loci) transcripts in egg cells. This new annotation was then used to quantify the isoforms detected. We found that 42% of multi-exonic genes undergo alternative splicing in sperm cells and vegetative nucleus. This number was slightly higher in egg cell (53%) (Supplemental Table S2.2). We found a total of 25023 alternative splicing event in our assembled transcriptome, from 12658 alternatively spliced genes.

Out of these new isoforms predicted, 2091 were expressed in sperm cells, 3600 in vegetative nuclei and 4530 in egg cells. In order to identify how many of these transcripts were specific to each cell type, we calculated a specificity score called specificity measure (spm). Specificity score ranges between 0 and 1 and the larger the spm values, the higher is the tissue specificity. Using the stringent specificity score of SPM≥0.9, we found 1174 (386 novel), 1928 (736 novel), 1813 (568 novel) transcripts to be specific to sperm cell, vegetative nucleus and egg cell, respectively. Sequence analysis of overrepresented splice sites (SS) at alternative 5´SS and 3´SS in the assembled transcriptome revealed that these splice sites were associated with GU and AG dinucleotides (Supplemental Figure S2.1).

The relatively high number of germ cell specific novel transcripts identified here indicates that the Arabidopsis annotation still remains incomplete. It is heavily biased towards vegetative tissues and more condition and/or cell/tissue specific transcripts could be missing.

2.4.3. Cell type specific alternatively spliced genes

In order to determine alternatively spliced genes that were differentially regulated between male and female gametes, we performed differential expression and differential alternative splicing analysis in a pair-wise manner: between sperm cell and vegetative nucleus and between sperm cell and egg cell, both at gene level as well as transcript level. We used limma in 3D-RNA seq pipeline (Guo et al., 2020) to identify the differentially expressed (DE) and differentially alternatively spliced (DAS) genes using pairwise comparisons.
When the transcriptome of sperm cell was compared with the vegetative nucleus, we identified 6906 DE genes (FDR < 0.01 and log2 fold change > 1), and 468 DAS genes (FDR < 0.01 and percent spliced (∆PS) >0.1) (Figure 2.3A-B) (See Supplemental Dataset 2.1 and 2.2). When we compared the genes that were either differentially expressed or spliced between sperm cell and vegetative nucleus, a significant overlap of 310 genes was observed

(Figure 2.3A). These genes might be regulated both at transcriptional as well as splicing level. The 468 differentially spliced genes encompassed 2131 transcripts, out of which 1411 (66.2%) were known transcripts, while 720 (33.8%) comprised novel unannotated transcripts. Similarly, when we compared sperm and egg cell, we identified 10809 DE genes (FDR < 0.01 and log2fold change > 1) and 608 DAS genes (FDR < 0.01 and percent spliced (ΔPS) >0.1) (See Supplemental Dataset 2.3 and 2.4). The overlap of 416 genes showed regulation both at transcription and splicing level (Supplemental Fig S2.2 A-B).

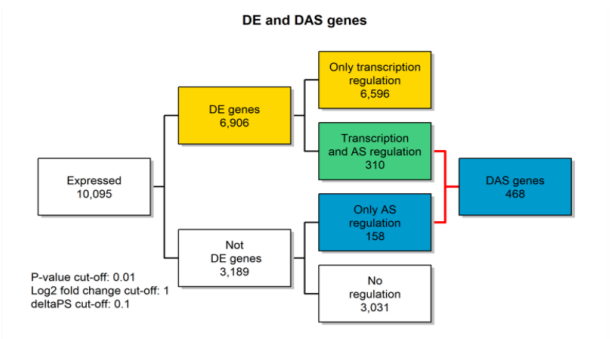
2.4.4. Functional classification of DE genes and transcripts

Next, we asked ourselves what could be the function of these differentially expressed genes. Therefore, we performed gene ontology analysis of the 6906 DE genes in our sperm vs vegetative nucleus comparison. GO categories with highest enrichment scores for the upregulated genes in sperm cells were found to be cell cycle regulation, DNA metabolic process, and chromatin organisation (Figure 2.3C). The genes that were downregulated in sperm cells were associated with pollen germination and tube growth, and vesicle mediated transport. On the other hand, gene set enrichment analysis of genes that were differentially spliced (DAS) between sperm and vegetative nucleus revealed mRNA splicing, chromatin modification, and protein localization and transport as significantly enriched terms (Figure 2.3D).

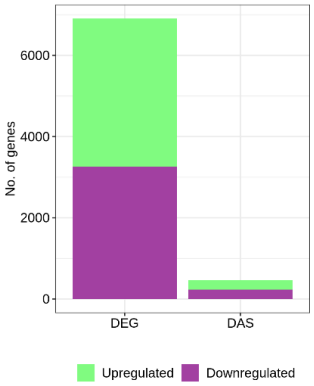
We then tried to understand the differences between sperm cells and egg cells, and we found 10809 genes to be differentially expressed between the two cell types. Our functional enrichment analysis of 3804 genes that were up-regulated in sperm cells revealed chromosome organization, DNA replication, and cell cycle and chromatin organization as significant terms (Supplemental Figure S2.2C). Similarly, analysis of genes that were differentially spliced between sperm and egg cell were found to be functionally

enriched for mRNA processing, RNA splicing, and regulation and metabolic process (Supplemental Figure S2.2D).

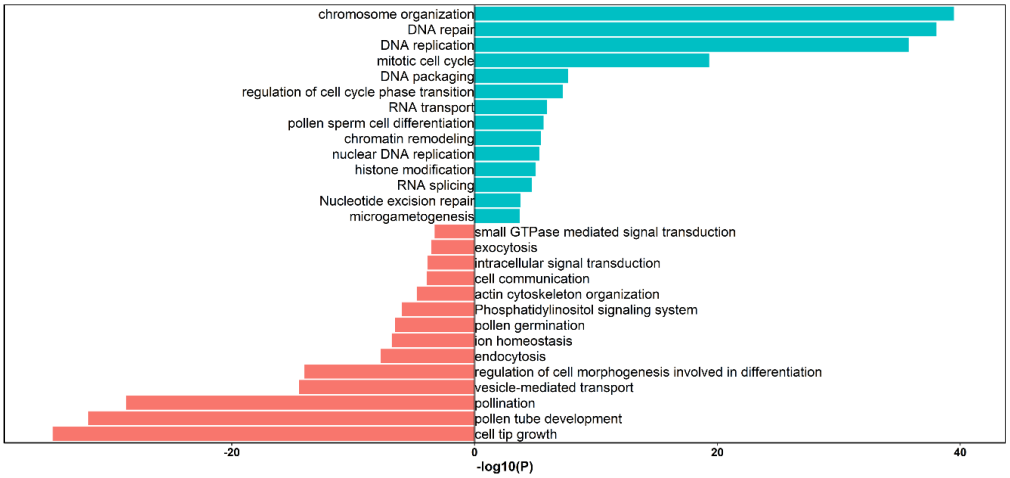
A


DE and DAS genes
Expressed
10,095
DE genes
6,906
Not
DE genes
3,189
Only transcription
regulation
6,596
Transcription
and AS regulation
310
Only AS
regulation
158
No
regulation
3,031
DAS genes
468
P-value cut-off: 0.01
Log2 fold change cut-off: 1
deltaPS cut-off: 0.1

B


No. of genes
6000
4000
2000
0
DEG
DAS
Upregulated
Downregulated

C


chromosome organization
DNA repair
DNA replication
mitotic cell cycle
DNA packaging
regulation of cell cycle phase transition
RNA transport
pollen sperm cell differentiation
chromatin remodeling
nuclear DNA replication
histone modification
RNA splicing
Nucleotide excision repair
microgametogenesis
small GTPase mediated signal transduction
exocytosis
intracellular signal transduction
cell communication
actin cytoskeleton organization
Phosphatidylinositol signaling system
pollen germination
ion homeostasis
endocytosis
regulation of cell morphogenesis involved in differentiation
vesicle-mediated transport
pollination
pollen tube development
cell tip growth
-20
0
20
40
-log10(P)

D

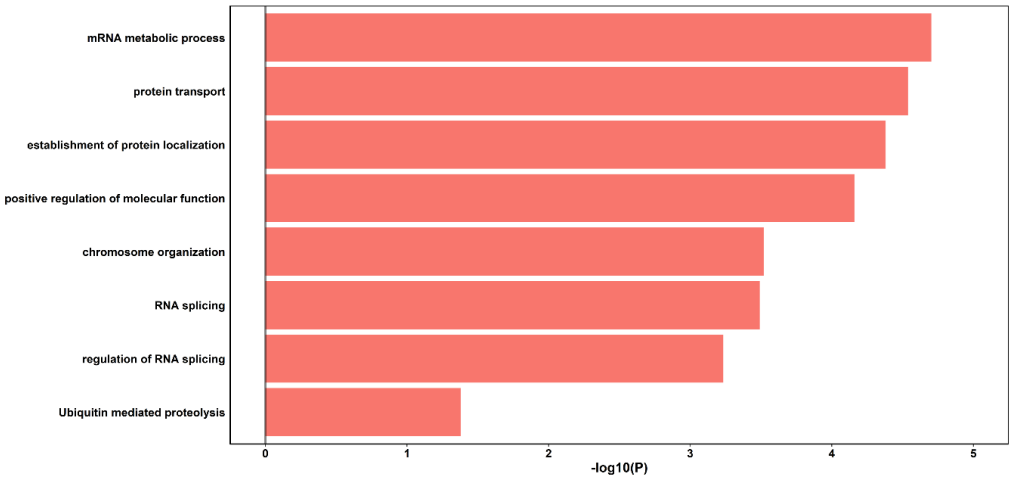
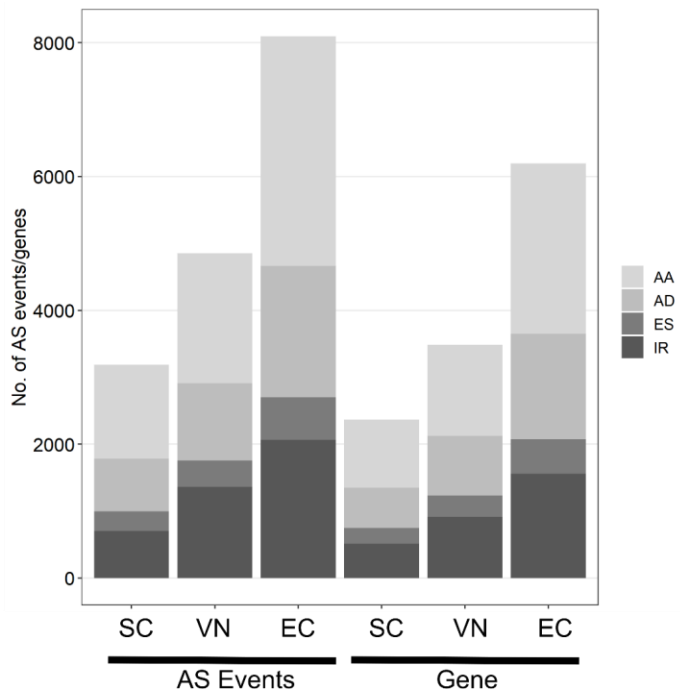

mRNA metabolic process
protein transport
establishment of protein localization
positive regulation of molecular function
chromosome organization
RNA splicing
regulation of RNA splicing
Ubiquitin mediated proteolysis
0
1
2
3
4
5
-log10(P)

Figure 2.3: Differential expression analysis at gene and transcript level. **A)** Flowchart depicting analysis of gene expression and alternative splicing between sperm cells and vegetative nuclei. Numbers of differentially expressed (DE) genes and differential alternative splicing (DAS) genes in sperm cell vs vegetative nucleus comparison. **B)** Bar plots showing the number of genes that are differentially expressed (DEG) and differentially alternatively spliced (DAS), separated by upregulated and downregulated. **C)** Functional enrichment of upregulated (blue bars) and downregulated (red bars) differentially expressed genes between sperm and vegetative nucleus. **D)** Functional enrichment of differentially spliced genes between sperm and vegetative nucleus. For detail GO, see Supplemental Dataset 2.5-2.6.

2.4.5. Sperm cells and vegetative nuclei differ in their splicing landscape

Although alternative splicing occurs in both plants and animals, information about tissue specific alternative splicing is relatively sparse in plants. Also, the relative proportion and frequency of different types of alternative splicing events in plants and animals differ (Chaudhary et al., 2019). To analyse the alternative splicing landscape in germ cells and determine their cell specific pattern in sperm cells, vegetative nuclei, and egg cells, we used Suppa2 tool (Trincado et al., 2018) and classified the splicing events detected into four types: intron retention (IR), exon skipping (ES), alternative acceptor (AA), and alterative donor (AD) (Figure 2.4).

A


8000
6000
4000
2000
0
No. of AS events/genes
AA
AD
ES
IR
SC
VN
EC
SC
VN
EC
AS Events
Gene

B

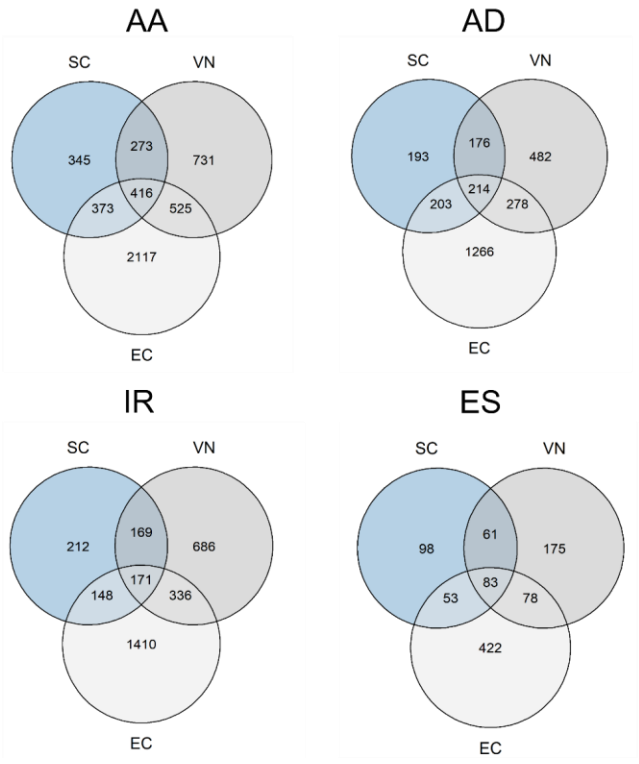

AA
SC
VN
345
273
731
416
373
525
2117
EC
AD
SC
VN
193
176
482
214
203
278
1266
EC
IR
SC
VN
212
169
686
171
148
336
1410
EC
ES
SC
VN
98
61
175
83
53
78
422
EC

Figure 2.4: Sex specific alternative splicing in the Arabidopsis germline. **A)** Bar plot showing different kinds of alternative splicing (AS) events and the number of genes corresponding to those alternative splicing events in SC, VN, and EC respectively. The four predominant forms of AS events identified were: ES, exon skipping; IR, intron retention; AA, alternative acceptor site; AD, alternative donor site. **B)** Venn diagrams depicting the number of overlapping AS events between the three cell types. SC sperm cells, VN vegetative nucleus, EC egg cell.

The alternative acceptor type was found to be the most predominant alternative splicing event in all three cell types, followed by intron retention in vegetative nucleus and egg cell, whereas in sperm cells, alternative donor was the second most prevalent form of splicing events (Figure 2.4A). The least abundant splicing event in all three cell types was found to be exon skipping. We also looked at the common AS events shared between the three cell types and identified 884 common AS events corresponding to 634 genes. The three cell types shared 171 IR events, 416 AA events, 214 AD events, and 83 ES events as common AS events (Figure 2.4B).

Tissue specific alternative splicing data is limited in plants, more so in other reference plants species such as rice and maize. In order to understand whether the alternative splicing landscape in germ cells is different between monocots and dicots, we used published data sets of sperm, vegetative nucleus and egg cells from rice and maize as a reference for monocot species (Anderson et al., 2013; Chen et al., 2017) and reanalysed to quantify the splicing landscape in other species.

Surprisingly, our analysis revealed that the proportion of alternative splicing events in rice and maize is different from that of Arabidopsis (Fig 2.5). We also found exon skipping to be the most predominant form of splicing event in maize egg cells and second most prevalent in sperm cells (Fig 2.5A-B), in contrast to both rice and Arabidopsis. Intron retention is the most predominant

form of splicing event in the rice male germline (Figure 2.5C-D), however alternative acceptor is the most predominant form of splicing event in egg cells of rice (Figure 2.5E), similar to what we observe in Arabidopsis. Taken together our data of these three species indicate that alternative splicing is highly tissue and condition specific, and probably more complex in plants than previously thought.

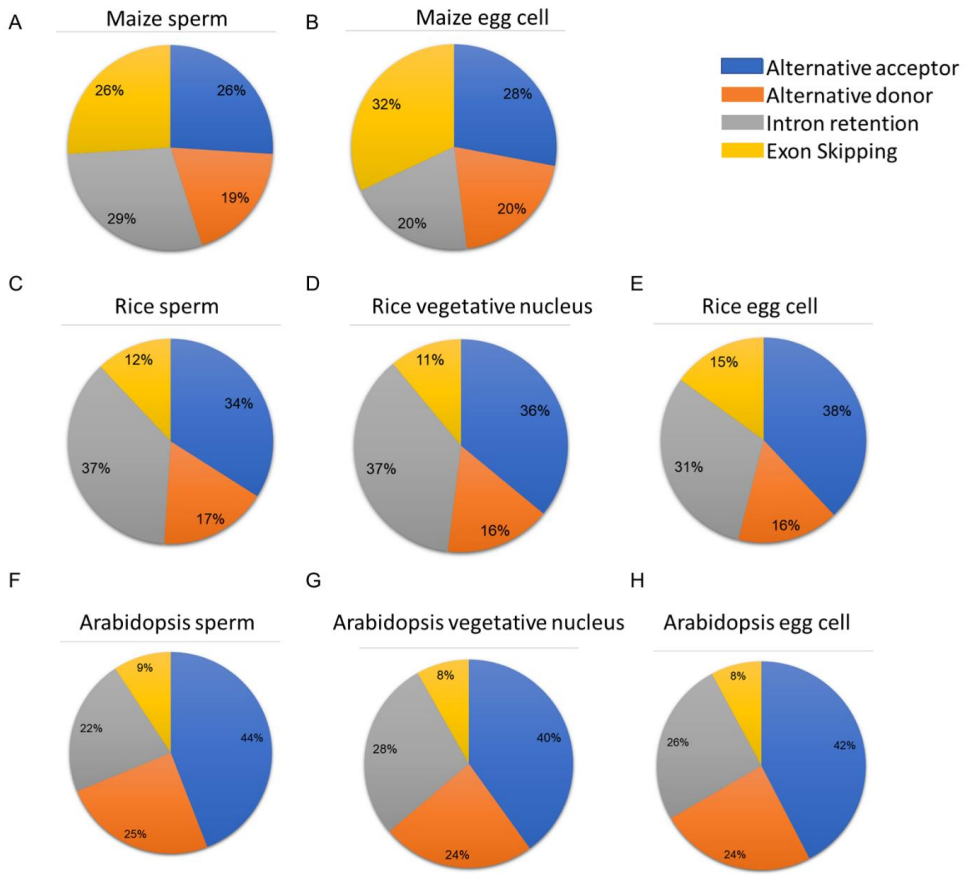


Figure 2.5: Comparison of alternative splicing landscape in germ cells of several plant species. A-B) Alternative splicing landscape in maize sperm and egg cell, respectively. **C-E)** Alternative splicing landscape in rice sperm, vegetative nucleus and egg cell. **F-H)** Alternative splicing landscape in Arabidopsis sperm, vegetative nucleus and egg cell.

2.4.6. Differentially spliced genes show variable expression between sperm cell and vegetative nucleus

Our study on expression analysis both at gene and transcript level is the first such attempt to understand the scale of alternative splicing in sex specific manner and how differences at transcript level may shape the transcriptome reprogramming particularly in sperm cell and vegetative nucleus.

Our detailed comparison of the differentially spliced genes between sperm cells and vegetative nuclei at the isoform level revealed varying expression dynamics ranging from complex splicing patterns (Figure 2.6A-C) to simple isoform switching (Figure 2.6D-F). Although a complex splicing pattern was observed for many genes, we could only detect 26 instances of isoform switching when we performed pairwise comparisons between sperm cell and egg cell, and sperm cell and vegetative nucleus (Supplemental Table S2.3 and S2.4).

Instances of simple isoform switching often showed very little changes in total gene expression despite significant differences in the abundances of individual isoforms. For instance, if we look at the expression at gene level of AT1G32130, with two annotated (AT1G32130.1 and AT1G32130.2) and two novel transcripts (MSTRG.3451.1 and MSTRG.3451.2), no significant difference was observed between sperm and vegetative nucleus (Figure 2.6D). However, measurement of isoform level expression revealed that the canonical isoform (AT1G32130.1) was more expressed in sperm cells and downregulated 5.5-fold in vegetative nuclei, whereas novel transcripts MSTRG.3451.1 and MSTRG.3451.2 were downregulated by more than 100-fold in sperm cells, thereby showing completely opposite expression level in sperm cells and vegetative nucleus. On the other hand, if we look at the gene expression level change of AT5G01990 with one annotated and one predicted isoform, we could see the differences both at gene expression level (downregulated by 1.78-fold in vegetative nucleus) as well as at the isoform

level. Complex splicing often resulted in a dynamic expression of isoforms which in turn resulted in a net shift in the expression of the gene towards the direction of the predominant isoform.

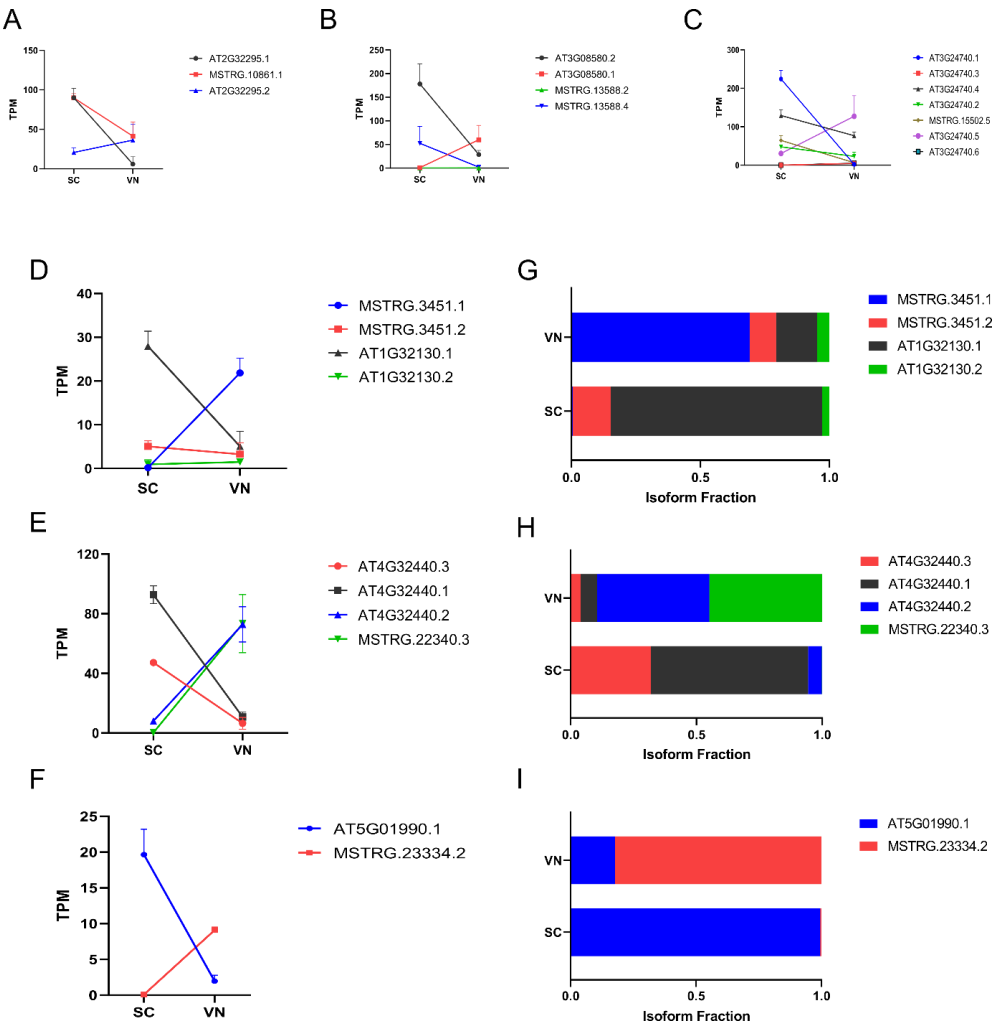


Figure 2.6: Isoform level dynamics in sperm cells and vegetative nucleus. A-C) Expression level (TPM) of individual isoforms that are significantly differentially expressed between sperm cells and vegetative nucleus, representing examples of complex splicing patterns. **D-F)** Expression levels of isoforms of genes depicting simple cases of isoform switching. **G-I)** Changes in the isoform fraction of the corresponding genes showing the isoform switching (D-F).

2.4.7. Sex-specific regulation of transcription and splicing factors

The Arabidopsis male gametophyte undergoes extensive transcriptome reprogramming during pollen development and studies so far suggest that this is tightly regulated at the gene expression level with at least two major successive gene expression reprogramming events occurring during pollen development (reviewed in Rutley and Twell, 2015). However, how alternative splicing shapes this reprogramming during the two successive pollen mitoses remains unknown. Also, knowledge about transcriptional control at the specific level of sperm cells and vegetative nucleus remains limited. In order to bridge this gap, we used our RNA-seq data to identify various transcription factor families expressed in the sperm cells and/or vegetative nucleus and how they compare with expression in the female germline.

Using the Plant Transcription Factor (TF) database (http://planttfdb.gao-lab.org/, Jin et al., 2017), we analysed expression of a total of 1717 annotated transcription factors, belonging to 53 different transcription factor families. The most predominant TF families in sperm cells were C3H (n=21), and ERF (n=18), whereas in vegetative nuclei, the most predominant TF families were ERF (n=33), MYB (n=29), and MYB related (n=28) (Figure 2.7A-B). On the other hand, the most predominant TF family in egg cell were bHLH (n=41) and ERF (n=41) (Supplemental Figure S2.3). Our results also indicate that the expression of several transcription factors was specific to each cell type (Figure 2.7C).

Out of the total 1717 annotated TFs, 295 TF belonging to several different TF families were differentially expressed when we compared sperm cell and vegetative nucleus. At least 17 TFs were also regulated at the splicing level, of which 6 exclusively at the splicing level, suggesting that regulation of the male germline in Arabidopsis occurs not only at transcriptional level, but also to a certain extent at splicing level (Figure 2.7E). We made a similar observation when comparing sperm cell and egg cell. Here, 437 TFs were differentially expressed, whereas 25 TFs were regulated at splicing level, and 9 of those exclusively at the splicing level (Figure 2.7D). Overall, our analysis sheds light on cell type specific transcriptional control of both male and female germlines, with several TFs being exclusively expressed in specific cell types marking them as potential candidates for further functional analyses.

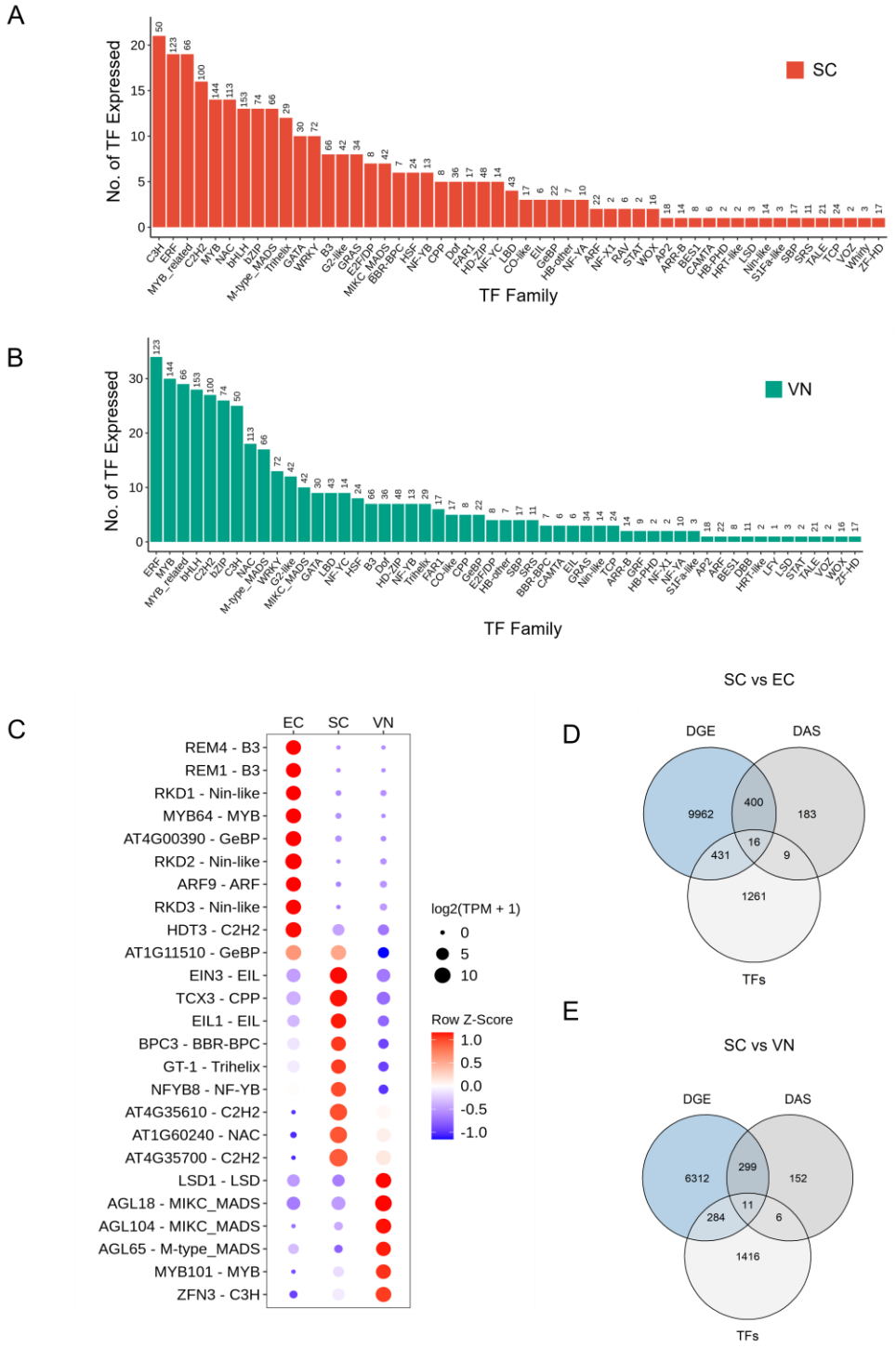

A
No. of TF Expressed
SC
TF Family
B
No. of TF Expressed
VN
TF Family
C
EC
SC
VN
REM4 - B3
REM1 - B3
RKD1 - Nin-like
MYB64 - MYB
AT4G00390 - GeBP
RKD2 - Nin-like
ARF9 - ARF
RKD3 - Nin-like
HDT3 - C2H2
AT1G11510 - GeBP
EIN3 - EIL
TCX3 - CPP
EIL1 - EIL
BPC3 - BBR-BPC
GT-1 - Trihelix
NFYB8 - NF-YB
AT4G35610 - C2H2
AT1G60240 - NAC
AT4G35700 - C2H2
LSD1 - LSD
AGL18 - MIKC_MADS
AGL104 - MIKC_MADS
AGL65 - M-type_MADS
MYB101 - MYB
ZFN3 - C3H
log2(TPM + 1)
Row Z-Score
D
SC vs EC
DGE
DAS
9962
400
183
431
16
9
1261
TFs
E
SC vs VN
DGE
DAS
6312
299
152
284
11
6
1416
TFs

Figure 2.7: Transcriptional regulation in male and female germlines. A-B) Relative frequency of transcription factors (TFs) per family expressed in SC **(A)** and VN **(B).** Number on top of each bar indicates total TFs present in a given family. C) Dot plot showing the expression of 25 most variable TFs across the three cell types. **D-E)** Venn plots to show the number of TFs that are differentially expressed or spliced between SC and EC, and SC and VN, respectively. SC sperm cells, VN vegetative nucleus, EC egg cell.

The important role of splicing in plant stress responses is well known, but much less is known about its role in plant reproduction. Alternative splicing mechanisms involve spliceosome proteins and splicing factors, which in addition might undergo alternative splicing themselves.

In Arabidopsis, there are multiple SR proteins and our results show that many of them are differentially expressed or spliced in a gamete dependent manner. In our sperm cell vs vegetative nucleus comparison, we found that at least 274 splicing factors were differentially expressed, and 42 were differentially spliced (Figure 2.8C). Similarly, when we looked at sperm cell and egg cell, 536 splicing factor genes were differentially expressed and 64 were differentially spliced between sperm and egg cells (Figure 2.8B). Interestingly, 11 and 21 splicing factors were only regulated at the splicing level when we compared sperm and vegetative nucleus or sperm and egg cell, respectively (Figure 2.8B-C).

Similar to what we observed for transcription factors, the splicing machinery seems to be unique to each cell type and several splicing factors were found to be exclusively expressed in one of the cell types (Figure 2.8A), potentially mediating sex specific functions during plant reproduction. Splicing factors are known to undergo splicing themselves to mediate their response in a tissue dependent manner, but so far no study has reported their role in plant germline development and our data could therefore provide a good starting point for such future studies.

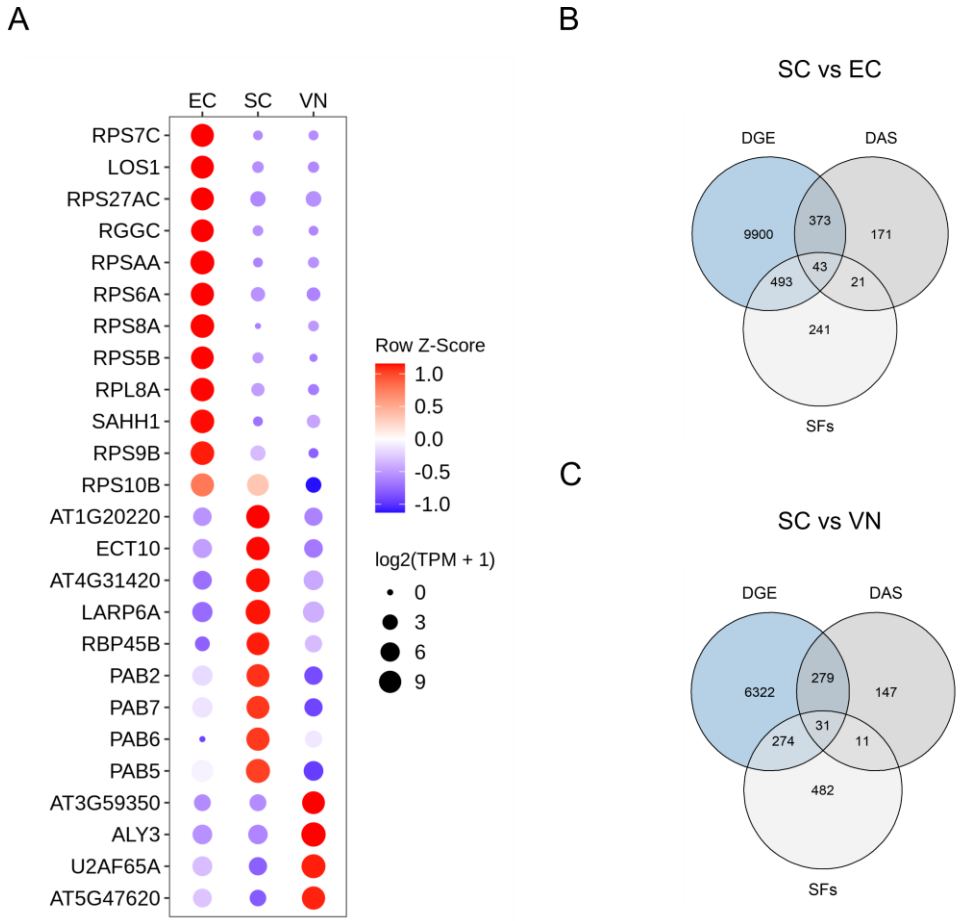


Figure 2.8: Splicing regulation in the three cell types. A) Dot plot showing the expression of 25 most variable splicing factors across the three cell types. **B-C)** Venn plots to show the number of splicing factors that are differentially expressed or spliced between SC and EC and between SC and VN, respectively. SC sperm cells, VN vegetative nucleus, EC egg cell.

2.5. Discussion

Male gametophyte development in flowering plants has been a very active area of research over the last two decades. Extensive transcriptomic and epigenomic data sets have provided unprecedented insights into pollen development (Borges et al., 2008; Slotkin et al., 2009; Calarco et al., 2012; Ibarra et al., 2012). The wealth of data available is not just limited to Arabidopsis, but rice, maize and more recently tomato have also been well studied (Anderson et al., 2013; Chen et al., 2017; Liu et al., 2018). The transcriptomic data available for pollen development varies though between species. For instance, in Arabidopsis data from every stage of pollen development from unicellular microspore to mature pollen grain is available (Honys and Twell, 2004; Julca et al., 2021), while for most other species the data is largely limited to mature pollen grains. Similarly, when it comes to sperm cells, data from Arabidopsis, rice, maize and tomato is available. However, vegetative nucleus data is only available for rice (Anderson et al., 2013). This might be due to the fact that the isolation of vegetative nuclei from pollen grains is technically challenging. Also, it is important to note that, while the Arabidopsis sperm cell transcriptome was based on microarrays, rice, tomato, maize germline transcriptomes are available in the form of RNA-seq data. Only recently, RNA-seq data on Arabidopsis sperm cells was published, but without detailed analysis (Borg et al., 2020). We closed this gap by performing RNA-seq analysis on sperm cells and vegetative nuclei isolated from mature pollen grains of Arabidopsis. This allowed us to identify not only more genes that are expressed in sperm cells and vegetative nucleus than previously known, but also to uncover genome wide alternative splicing patterns in the germline.

The transcriptome data from sperm cells helped to identify 21% more genes as expressed than previously known (Figure 2.2D). Our transcriptomic data from vegetative nuclei also revealed a higher number of genes than

previously reported for whole pollen grains. Thus, overall our study manages to decipher the transcriptome of the individual components of the mature pollen grain at a much higher resolution than can be achieved when isolating RNA from whole pollen grains.

Transcriptome reprogramming during pollen development has been well studied but most of the studies have focussed on changes at gene expression level. However, with RNA-seq it is possible to identify changes not only at gene level but also at transcript level, revealing the scale of alternative splicing in reproductive cell types. The first landmark transcriptomic study to identify changes in pollen at the splicing level was published by Loraine et al., 2013, in which 30 genes were found to be differentially spliced between leaf and pollen. However, this study could not reveal the relative contribution of transcripts coming from sperm cell or vegetative nucleus. Our study therefore served to fill this knowledge gap towards the understanding of genome wide alternative splicing in germ cells. To guide our analysis, we reassembled an Arabidopsis transcriptome, mainly from reproductive tissue samples. We identified 58039 transcripts, an increase of approx. 20% over the number of transcripts predicted in Araport 11, the most recent Arabidopsis annotation comprising 48358 protein coding transcripts. In addition, when we compared the Araport11 annotation with our newly assembled transcriptome, we identified at least 2881 genes that were previously not known to undergo splicing. Comparing to the more recent, updated Arabidopsis annotation AtRTD2 with its 74194 non-redundant protein coding transcripts (Zhang et al., 2017), we were still able to identify 2148 genes that were not predicted to undergo splicing. These data suggest that the Arabidopsis annotation is still far from complete and more tissue and condition specific transcripts might await discovery.

The role of splicing factors and RNA binding proteins in plant reproduction is relatively unexplored. Splicing factors are commonly known to be involved in

stress responses. However, recently some of them have been reported to have a role in reproduction. For instance, SR45 and U2AF65A/B have been implicated in pollen germination and tube growth (Cao et al., 2013; Park et al., 2019). Arabidopsis has 798 known splicing factors/RNA binding proteins. This complex network of splicing factors and RNA binding proteins might have an important function in plant reproduction and could be an area worth exploring for future research. Our transcriptome data from sperm cells also helped to identify the expression of 76 additional splicing factors for which there were no probes on microarrays (Borges et al., 2008).

Information about tissue specific splicing in Arabidopsis is very limited and the only study on Arabidopsis pollen provided no information about the different types of splicing events. Intron retention is thought to be the major splicing event in plants, but this did not hold true in our analysis of reproductive tissue specific splicing. The most predominant form of alternative splicing event was alternative acceptor in all three cell types, followed by intron retention in vegetative nucleus and egg cell, while in sperm cells alternative donor was the second most predominant form of splicing (Figure 2.4). However, a recent study on alternative splicing in 28 different tissues (flowers, shoots, seeds, roots, and leaves) and developmental stages in soybean (Shen et al., 2014) points towards differences in the splicing landscape between monocots and dicots. In order to test this hypothesis, we compared alternative splicing in sperm cells and vegetative nuclei of Arabidopsis with that in two reference monocot species, rice and maize. Our data revealed that the relative proportion of alternative splicing events between monocot and dicot show a high degree of variation, particularly in maize, where exon skipping was found to be the most prevalent (32%) and second most (26%) prevalent form of alternative splicing event in egg cell and sperm cell, respectively (Figure 2.5 A-B). This difference in the proportion of splicing events between maize and the other two species (Arabidopsis and rice) might be plant specific. Recently developing ears in maize were shown

to undergo exon skipping as their most predominant form (27%) of splicing event (Thatcher et al., 2016), a form usually expected to be the least abundant in plants. We also found intron retention to be the most predominant form of splicing in rice sperm and vegetative nucleus, which is in contrast to our results in Arabidopsis.

Interestingly, we did not find intron retention to be the most predominant form of alternative splicing events as is generally expected in plants, particularly in Arabidopsis. Intron retention is thought to be overestimated in plants, a notion that has recently been validated in other studies (Martín et al., 2021) as well as in our analysis. These differences could be explained at least partly, due to the advances in RNA-seq technologies and improvement of the de novo assembly algorithms and genome annotation detail that may help in improving the prediction of transcripts and splicing events more accurately. Nevertheless, how the expression dynamics and splicing patterns of different splicing factors contribute to the overall fertility of plants is unclear and calls for further research.

Our analysis of sperm and vegetative nucleus transcriptome also revealed more than 6000 genes to be differentially expressed. Genes that were upregulated in sperm cells were enriched for DNA replication and packaging, sperm cell differentiation and cell cycle progression (Figure 2.3C). The functional enrichment analysis points towards the role of genes that are involved in sperm cell maturation before gamete fusion and are similar to previous classifications of sperm cell enriched transcripts (Borges et al., 2008). Sperm transcriptome analysis in other plant species like rice, maize, and tomato have also revealed an enrichment of transcripts with similar functions, such as cell-cycle regulation, proteasome mediated protein metabolism, and DNA repair and replication (Anderson et al., 2013; Chen et al., 2017; Liu et al., 2018). This indicates that the process of sperm maturation and differentiation is largely conserved across flowering plants. Genes that

are downregulated in sperm cells relative to the vegetative nucleus are involved in pollen germination and tube growth, cell-cell communication and vesicle mediated transport. These functions are important for pollen tube growth and hence reflect the main function of the vegetative nucleus.

Pollen development is under tight transcriptional control and several transcription factors have been identified to be important for pollen development, germination and ovule targeting (Rutley and Twell., 2015). Our analysis identified at least 295 transcription factors that were differentially expressed between sperm cells and vegetative nucleus (Figure 2.7E). Transcription factors are not only regulated at gene expression level, but also at the splicing level and at least 6 are exclusively regulated at the splicing level, warranting further functional characterization. Several transcriptional factors were found to have cell-type specific expression patterns (Figure 2.7C), some of which have already been studied. For instance, agamous like genes AGL65 and AGL104 belonging to the MIKC subgroup of the MADS-box transcription factor family were found to be upregulated during late stages of pollen development (Pina et al., 2005). Functional characterization of *agl65, agl104, agl66 and agl94* in different combinations of double and triple mutants highlighted the role of MADS-box transcription factor networks for pollen development and maturation (Verelst et al., 2007; Adamczyk and Fernandez 2009).

Extensive studies of male germline development over the past years have highlighted the role of the transcriptional regulon mediated by the MYB transcription factor DUO1. Our analysis of cell specific transcription factors points towards a DAZ3- C2H2 type transcription factor to be highly specific to sperm cells, and it is predicted to be a direct target of DUO1 (Borg et al., 2011). Similarly, the RKD subfamily in Arabidopsis has 5 genes out of which three were found to be highly expressed in egg cells. Interestingly, although several T-DNA insertion alleles of Arabidopsis RKD genes are available, no

egg cell defects have been reported so far for rkd1/2 double mutants, suggesting high levels of functional redundancy among gene family members or lack of effective functional alleles. Use of new gene editing techniques such as CRISPR and deleting all genes to uncover a potential egg cell defect, may help to provide more detailed insights into this predominantly egg cell expressed subfamily (Wuest et al., 2010; Koszegi et al., 2011).

In conclusion, using RNA-seq, we have shown the extent of genome-wide alternative splicing occurring in the sperm cell and vegetative nucleus. In addition, our comparative analysis of the male and female germline helped to shed light on sex specific modulation of alternative splicing, an area not addressed in previous studies. Our new annotation helped to predict at least 20% more transcripts than previously known, including many transcripts that are highly specific to sperm cell, vegetative nucleus or egg cell. Many of these genes are also associated with transcription, splicing factors, chromatin modification and are differentially spliced between sperm and egg cell. These new splice variants potentially contribute to proteome diversity and function in Arabidopsis gametes. Our study therefore provides an overview of genome wide alternative splicing in sperm cell, vegetative nucleus and egg cell and serves as a starting point for further in-depth analysis at functional and molecular level.

2.6. Authors Contributions

Chandra Shekhar Misra and Jörg D. Becker (head of “Plant Reproduction and Evolution Group” at Instituto de Tecnologia Química e Biológica António Xavier (ITQB)), conceptualized the project. Chandra Shekhar Misra performed the experiments for collection of sperm cells and vegetative nuclei using FACS. Chandra Shekhar analysed the RNA-seq data from all the tissues, and assembled the transcriptome with the help of Pedro Barros (Postdoc at Plant Functional Genomics group at ITQB). Chandra Shekhar and Pedro Barros performed the alternative splicing analysis. António Sousa

(Bioinformatician at IGC) prepared the plots with inputs from Chandra Shekhar and Jörg D. Becker. This work was supervised by Jörg D. Becker.

2.7. Acknowledgment

I would like to thank Vera Nunes, technician at the Plant Facility of IGC for plant growth and maintenance. I would also like to acknowedge the help of other facilities at IGC namely Microscopy Facility, Flow Cytometry Unit and Genomics Unit.

2.8. References

Adamczyk, B.J. and Fernandez, D.E., 2009. MIKC* MADS domain heterodimers are required for pollen maturation and tube growth in Arabidopsis*. Plant Physiology*, 149(4), pp.1713-1723.

Anderson, S.N., Johnson, C.S., Jones, D.S., Conrad, L.J., Gou, X., Russell, S.D. and Sundaresan, V., 2013. Transcriptomes of isolated Oryza sativa gametes characterized by deep sequencing: evidence for distinct sex-dependent chromatin and epigenetic states before fertilization. *The Plant Journal*, *76*(5), pp.729-741.

Barakate, A., Orr, J., Schreiber, M., Colas, I., Lewandowska, D., McCallum, N., Macaulay, M., Morris, J., Arrieta, M., Hedley, P.E. and Ramsay, L., 2020. Time-resolved transcriptome of barley anthers and meiocytes reveals robust and largely stable gene expression changes at meiosis entry. *bioRxiv*.

Baralle, F.E. and Giudice, J., 2017. Alternative splicing as a regulator of development and tissue identity. *Nature Reviews Molecular Cell Biology*, *18*(7), p.437.

Bedre, R., Irigoyen, S., Schaker, P.D., Monteiro-Vitorello, C.B., Da Silva, J.A. and Mandadi, K.K., 2019. Genome-wide alternative splicing landscapes

modulated by biotrophic sugarcane smut pathogen. *Scientific Reports*, *9*(1), pp.1-12.

Berger, F. and Twell, D., 2011. Germline specification and function in plants. *Annual Review of Plant Biology*, *62*, pp.461-484.

Birkenbihl, R.P., Kracher, B., Ross, A., Kramer, K., Finkemeier, I. and Somssich, I.E., 2018. Principles and characteristics of the Arabidopsis WRKY regulatory network during early MAMP-triggered immunity. *The Plant Journal*, *96*(3), pp.487-502.

Borg, M., Brownfield, L., Khatab, H., Sidorova, A., Lingaya, M. and Twell, D., 2011. The R2R3 MYB transcription factor DUO1 activates a male germline-specific regulon essential for sperm cell differentiation in Arabidopsis. *The Plant Cell*, *23*(2), pp.534-549.

Borg, M., Jacob, Y., Susaki, D., LeBlanc, C., Buendía, D., Axelsson, E., Kawashima, T., Voigt, P., Boavida, L., Becker, J. and Higashiyama, T., 2020. Targeted reprogramming of H3K27me3 resets epigenetic memory in plant paternal chromatin. *Nature cell Biology*, *22*(6), pp.621-629.

Borges, F., Gardner, R., Lopes, T., Calarco, J.P., Boavida, L.C., Slotkin, R.K., Martienssen, R.A. and Becker, J.D., 2012. FACS-based purification of Arabidopsis microspores, sperm cells and vegetative nuclei. *Plant Methods*, *8*(1), pp.1-9.

Borges, F., Gomes, G., Gardner, R., Moreno, N., McCormick, S., Feijó, J.A. and Becker, J.D., 2008. Comparative transcriptomics of Arabidopsis sperm cells. *Plant Physiology*, *148*(2), pp.1168-1181.

Brownfield, L., Hafidh, S., Borg, M., Sidorova, A., Mori, T. and Twell, D., 2009a. A plant germline-specific integrator of sperm specification and cell cycle progression. *PLoS Genetics,* 5(3), p.e1000430.

Brownfield, L., Hafidh, S., Durbarry, A., Khatab, H., Sidorova, A., Doerner, P. and Twell, D., 2009b. Arabidopsis DUO POLLEN3 is a key regulator of male germline development and embryogenesis. *The Plant Cell*, 21(7), pp.1940-1956.

Calarco, J.P., Borges, F., Donoghue, M.T., Van Ex, F., Jullien, P.E., Lopes, T., Gardner, R., Berger, F., Feijó, J.A., Becker, J.D. and Martienssen, R.A., 2012. Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. *Cell*, *151*(1), pp.194-205.

Calixto, C.P., Guo, W., James, A.B., Tzioutziou, N.A., Entizne, J.C., Panter, P.E., Knight, H., Nimmo, H.G., Zhang, R. and Brown, J.W., 2018. Rapid and dynamic alternative splicing impacts the Arabidopsis cold response transcriptome. *The Plant Cell*, *30*(7), pp.1424-1444.

Cao, L.J., Zhao, M.M., Liu, C., Dong, H.J., Li, W.C. and Ren, H.Y., 2013. LISR28 is involved in pollen germination by affecting filamentous actin dynamics. *Molecular Plant*, *6*(4), pp.1163-1175.

Chaudhary, S., Khokhar, W., Jabre, I., Reddy, A.S., Byrne, L.J., Wilson, C.M. and Syed, N.H., 2019. Alternative splicing and protein diversity: plants versus animals. *Frontiers in Plant Science*, 10, p.708.

Chen, J., Strieder, N., Krohn, N.G., Cyprys, P., Sprunck, S., Engelmann, J.C. and Dresselhaus, T., 2017. Zygotic genome activation occurs shortly after fertilization in maize. *The Plant Cell*, *29*(9), pp.2106-2125.

Cheng, C.Y., Krishnakumar, V., Chan, A.P., Thibaud-Nissen, F., Schobel, S. and Town, C.D., 2017. Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. *The Plant Journal*, *89*(4), pp.789-804.

Crooks, G.E., Hon, G., Chandonia, J.M. and Brenner, S.E., 2004. WebLogo: a sequence logo generator. *Genome Research*, *14*(6), pp.1188-1190.

Filichkin, S.A., Priest, H.D., Givan, S.A., Shen, R., Bryant, D.W., Fox, S.E., Wong, W.K. and Mockler, T.C., 2010. Genome-wide mapping of alternative splicing in Arabidopsis thaliana. *Genome Research*, *20*(1), pp.45-58.

Guo, W., Tzioutziou, N.A., Stephen, G., Milne, I., Calixto, C.P., Waugh, R., Brown, J.W. and Zhang, R., 2020. 3D RNA-seq: a powerful and flexible tool for rapid and accurate differential expression and alternative splicing analysis of RNA-seq data for biologists. *RNA Biology*, pp.1-14.

Hofmann, F., Schon, M.A. and Nodine, M.D., 2019. The embryonic transcriptome of Arabidopsis thaliana. *Plant reproduction*, *32*(1), pp.77-91.
Honys, D. and Twell, D., 2004. Transcriptome analysis of haploid male gametophyte development in Arabidopsis. *Genome Biology*, *5*(11), pp.1-13.

Howard, B.E., Hu, Q., Babaoglu, A.C., Chandra, M., Borghi, M., Tan, X., He, L., Winter-Sederoff, H., Gassmann, W., Veronese, P. and Heber, S., 2013. High-throughput RNA sequencing of pseudomonas-infected Arabidopsis reveals hidden transcriptome complexity and novel splice variants. *PloS One*, *8*(10), p.e74183.

Huang, J., Lu, X., Wu, H., Xie, Y., Peng, Q., Gu, L., Wu, J., Wang, Y., Reddy, A.S. and Dong, S., 2020. Phytophthora effectors modulate genome-wide alternative splicing of host mRNAs to reprogram plant immunity. *Molecular Plant*, *13*(10), pp.1470-1484.

Huang, L., Shi, X., Wang, W., Ryu, K.H. and Schiefelbein, J., 2017. Diversification of root hair development genes in vascular plants. *Plant Physiology*, *174*(3), pp.1697-1712.

Ibarra, C.A., Feng, X., Schoft, V.K., Hsieh, T.F., Uzawa, R., Rodrigues, J.A., Zemach, A., Chumak, N., Machlicova, A., Nishimura, T. and Rojas, D., 2012. Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. *Science*, *337*(6100), pp.1360-1364.

Jin, J., Tian, F., Yang, D.C., Meng, Y.Q., Kong, L., Luo, J. and Gao, G., 2016. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Research*, p.gkw982.

Julca, I., Ferrari, C., Flores-Tornero, M., Proost, S., Lindner, A., Hackenberg, D., Steinbachová, L., Michaelidis, C., Pereira, S.G., Misra, C.S., Kawashima, T., Borg, M., Berger, F., Goldberg, J., Johnson, M., Honys, D., Twell, D., Sprunck, S., Dresselhaus, T., Becker, J.D., Mutwil, M., 2020. Comparative transcriptomic analysis reveals conserved transcriptional programs underpinning organogenesis and reproduction in land plants. *Nature Plants*, *7(8)*, pp. 1143-1159.

Keller, M., Hu, Y., Mesihovic, A., Fragkostefanakis, S., Schleiff, E. and Simm, S., 2017. Alternative splicing in tomato pollen in response to heat stress. *DNA Research*, *24*(2), pp.205-217.

Kim, D., Paggi, J.M., Park, C., Bennett, C. and Salzberg, S.L., 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, *37*(8), pp.907-915.

Koszegi, D., Johnston, A.J., Rutten, T., Czihal, A., Altschmied, L., Kumlehn, J., Wu¨ st, S.E., Kirioukhova, O., Gheyselinck, J., Grossniklaus, U., and Ba¨ umlein, H., 2011. Members of the RKD transcription factor family induce an egg cell-like gene expression program. *The Plant Journal, 67(2)*, pp. 280–291.

Li, Q., Xiao, G. and Zhu, Y.X., 2014. Single-nucleotide resolution mapping of the Gossypium raimondii transcriptome reveals a new mechanism for alternative splicing of introns. *Molecular Plant*, *7*(5), pp.829-840.

Liu, L., Lu, Y., Wei, L., Yu, H., Cao, Y., Li, Y., Yang, N., Song, Y., Liang, C. and Wang, T., 2018. Transcriptomics analyses reveal the molecular roadmap

and long non-coding RNA landscape of sperm cell lineage development. *The Plant Journal*, *96*(2), pp.421-437.

Loraine, A.E., McCormick, S., Estrada, A., Patel, K. and Qin, P., 2013. RNA-seq of Arabidopsis pollen uncovers novel transcription and alternative splicing. *Plant Physiology*, *162*(2), pp.1092-1109.

Mandadi, K.K. and Scholthof, K.B.G., 2015. Genome-wide analysis of alternative splicing landscapes modulated during plant-virus interactions in Brachypodium distachyon. *The Plant Cell*, *27*(1), pp.71-85.

Marquez, Y., Brown, J.W., Simpson, C., Barta, A. and Kalyna, M., 2012. Transcriptome survey reveals increased complexity of the alternative splicing landscape in Arabidopsis. *Genome Research*, *22*(6), pp.1184-1195.

Martín, G., Márquez, Y., Mantica, F., Duque, P. and Irimia, M., 2021. Alternative splicing landscapes in *Arabidopsis thaliana* across tissues and stress conditions highlight major functional differences with animals. *Genome Biology*, *22*(1), pp.1-26.

Mergner, J., Frejno, M., List, M., Papacek, M., Chen, X., Chaudhary, A., Samaras, P., Richter, S., Shikata, H., Messerer, M. and Lang, D., et al., 2020. Mass-spectrometry-based draft of the Arabidopsis proteome. *Nature*, *579*(7799), pp.409-414.

Misra, C.S., Santos, M.R., Rafael-Fernandes, M., Martins, N.P., Monteiro, M. and Becker, J.D., 2019. Transcriptomics of Arabidopsis sperm cells at single-cell resolution. *Plant Reproduction*, *32*(1), pp.29-38.

Pan, Q., Shai, O., Lee, L.J., Frey, B.J. and Blencowe, B.J., 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics*, *40*(12), pp.1413-1415.

Park, H.Y., Lee, H.T., Lee, J.H. and Kim, J.K., 2019. Arabidopsis U2AF65 regulates flowering time and the growth of pollen tubes. *Frontiers in Plant Science*, *10*, p.569.

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A. and Kingsford, C., 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, *14*(4), pp.417-419.

Pertea, M., Kim, D., Pertea, G.M., Leek, J.T. and Salzberg, S.L., 2016. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols*, *11*(9), p.1650.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T. and Salzberg, S.L., 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, *33*(3), pp.290-295.

Pina, C., Pinto, F., Feijó, J.A. and Becker, J.D., 2005. Gene family analysis of the Arabidopsis pollen transcriptome reveals biological implications for cell growth, division control, and gene expression regulation. *Plant Physiology*, *138*(2), pp.744-756.

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H. and Vilo, J., 2019. g: Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research*, *47*(W1), pp.W191-W198.

Reddy, A.S., Marquez, Y., Kalyna, M. and Barta, A., 2013. Complexity of the alternative splicing landscape in plants. *The Plant Cell*, *25*(10), pp.3657-3683.

Rotman, N., Durbarry, A., Wardle, A., Yang, W.C., Chaboud, A., Faure, J.E., Berger, F. and Twell, D., 2005. A novel class of MYB factors controls sperm-cell formation in plants. Current Biology, 15(3), pp.244-248.

Rutley, N. and Twell, D., 2015. A decade of pollen transcriptomics. *Plant Reproduction*, *28*(2), pp.73-89.

Santos, M.R., Bispo, C. and Becker, J.D., 2017. Isolation of Arabidopsis pollen, sperm cells, and vegetative nuclei by fluorescence-activated cell sorting (FACS). In *Plant germline development* (pp. 193-210). Humana Press, New York, NY.

Sebestyén, E., Zawisza, M. and Eyras, E., 2015. Detection of recurrent alternative splicing switches in tumor samples reveals novel signatures of cancer. *Nucleic Acids Research*, *43*(3), pp.1345-1356.

Shen, Y., Zhou, Z., Wang, Z., Li, W., Fang, C., Wu, M., Ma, Y., Liu, T., Kong, L.A., Peng, D.L. and Tian, Z., 2014. Global dissection of alternative splicing in paleopolyploid soybean. *The Plant Cell*, *26*(3), pp.996-1008.

Slotkin, R.K., Vaughn, M., Borges, F., Tanurdžić, M., Becker, J.D., Feijó, J.A. and Martienssen, R.A., 2009. Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell*, *136*(3), pp.461-472.

Staiger, D. and Brown, J.W., 2013. Alternative splicing at the intersection of biological timing, development, and stress responses. *The Plant Cell*, *25*(10), pp.3640-3656.

Tannenbaum, M., Sarusi-Portuguez, A., Krispil, R., Schwartz, M., Loza, O., Benichou, J.I., Mosquna, A. and Hakim, O., 2018. Regulatory chromatin landscape in Arabidopsis thaliana roots uncovered by coupling INTACT and ATAC-seq. *Plant methods*, *14*(1), pp.1-12.

Thatcher, S.R., Danilevskaya, O.N., Meng, X., Beatty, M., Zastrow-Hayes, G., Harris, C., Van Allen, B., Habben, J. and Li, B., 2016. Genome-wide analysis of alternative splicing during development and drought stress in maize. *Plant Physiology*, *170*(1), pp.586-599.

Thatcher, S.R., Zhou, W., Leonard, A., Wang, B.B., Beatty, M., Zastrow-Hayes, G., Zhao, X., Baumgarten, A. and Li, B., 2014. Genome-wide analysis of alternative splicing in Zea mays: landscape and genetic regulation. *The Plant Cell*, *26*(9), pp.3472-3487.

Trincado, J.L., Entizne, J.C., Hysenaj, G., Singh, B., Skalic, M., Elliott, D.J. and Eyras, E., 2018. SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. *Genome Biology*, *19*(1), pp.1-11.

Verelst, W., Twell, D., de Folter, S., Immink, R., Saedler, H. and Münster, T., 2007. MADS-complexes regulate transcriptome dynamics during pollen maturation. *Genome Biology*, *8*(11), pp.1-15.

Walker, J., Gao, H., Zhang, J., Aldridge, B., Vickers, M., Higgins, J.D. and Feng, X., 2018. Sexual-lineage-specific DNA methylation regulates meiosis in Arabidopsis. *Nature Genetics*, *50*(1), pp.130-137.

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P. and Burge, C.B., 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature*, *456*(7221), pp.470-476.

Wuest, S.E., Vijverberg, K., Schmidt, A., Weiss, M., Gheyselinck, J., Lohr, M., Wellmer, F., Rahnenführer, J., von Mering, C. and Grossniklaus, U., 2010. Arabidopsis female gametophyte gene expression map reveals similarities between plant and animal gametes. *Current Biology*, *20*(6), pp.506-512.

Xu, Q., Modrek, B. and Lee, C., 2002. Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. *Nucleic Acids Research*, *30*(17), pp.3754-3766.

Yeo, G., Holste, D., Kreiman, G. and Burge, C.B., 2004. Variation in alternative splicing across human tissues. *Genome Biology*, *5*(10), pp.1-15.

Zhang, G., Guo, G., Hu, X., Zhang, Y., Li, Q., Li, R., Zhuang, R., Lu, Z., He, Z., Fang, X. and Chen, L., 2010. Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome. *Genome Research*, *20*(5), pp.646-654.

Zhang, R., Calixto, C.P., Marquez, Y., Venhuizen, P., Tzioutziou, N.A., Guo, W., Spensley, M., Entizne, J.C., Lewandowska, D., Ten Have, S. and Frei dit Frey, N., 2017. A high quality Arabidopsis transcriptome for accurate transcript-level analysis of alternative splicing. *Nucleic Acids Research*, *45*(9), pp.5061-5073.

Zhao, P., Zhou, X., Shen, K., Liu, Z., Cheng, T., Liu, D., Cheng, Y., Peng, X. and Sun, M.X., 2019. Two-step maternal-to-zygotic transition with two-phase parental genome contributions. *Developmental Cell*, *49*(6), pp.882-893.

Supplemental Data

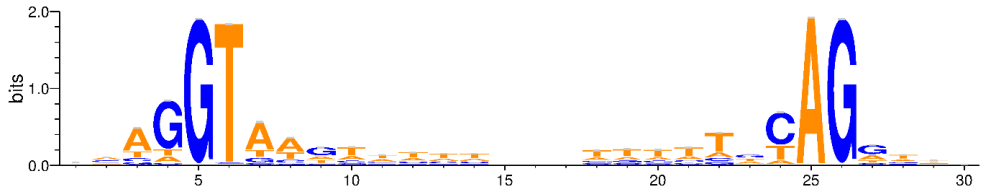
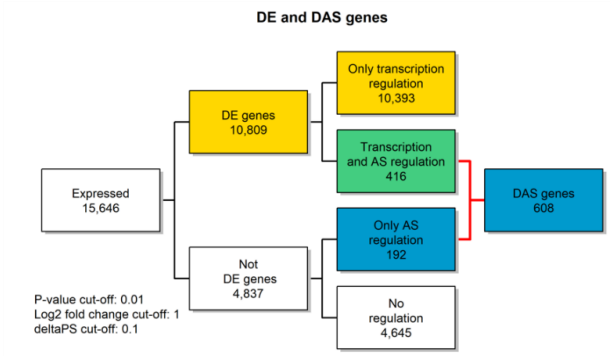
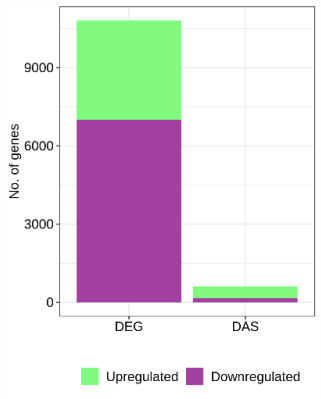


Figure S2.1. Sequence logos of the exonic and the intronic boundaries of the 5´ and 3´ splice sites; logos were created using WebLogo (Crooks et al., 2004).

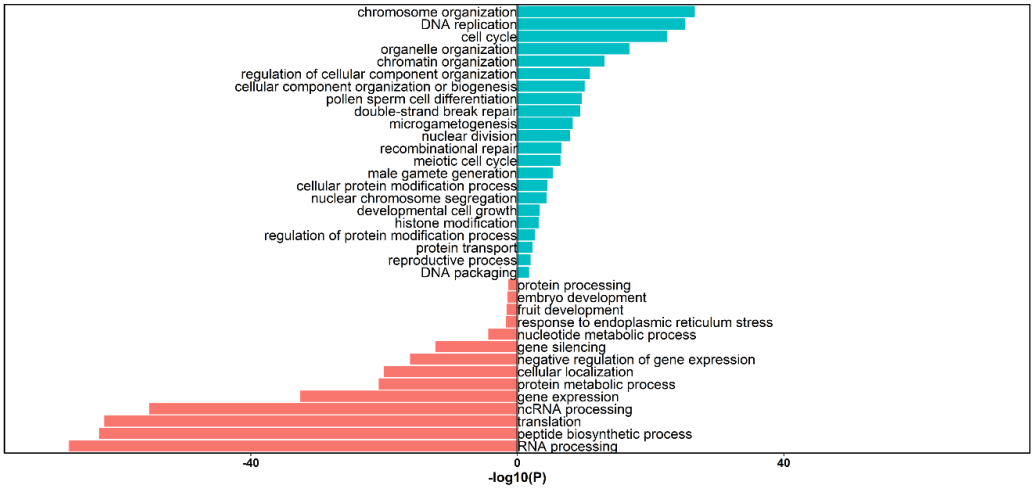
A


DE and DAS genes
Expressed
15,646
DE genes
10,809
Not
DE genes
4,837
Only transcription
regulation
10,393
Transcription
and AS regulation
416
Only AS
regulation
192
No
regulation
4,645
DAS genes
608
P-value cut-off: 0.01
Log2 fold change cut-off: 1
deltaPS cut-off: 0.1

B


No. of genes
9000
6000
3000
0
DEG
DAS
Upregulated
Downregulated

C


chromosome organization
DNA replication
cell cycle
organelle organization
chromatin organization
regulation of cellular component organization
cellular component organization or biogenesis
pollen sperm cell differentiation
double-strand break repair
microgametogenesis
nuclear division
recombinational repair
meiotic cell cycle
male gamete generation
cellular protein modification process
nuclear chromosome segregation
developmental cell growth
histone modification
regulation of protein modification process
protein transport
reproductive process
DNA packaging
protein processing
embryo development
fruit development
response to endoplasmic reticulum stress
nucleotide metabolic process
gene silencing
negative regulation of gene expression
cellular localization
protein metabolic process
gene expression
ncRNA processing
translation
peptide biosynthetic process
RNA processing
-40
0
40
-log10(P)

D

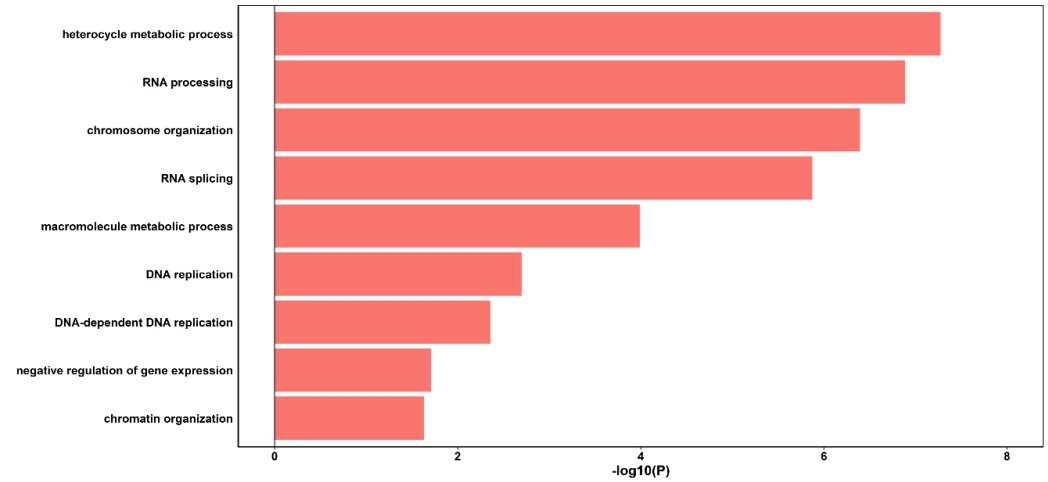

heterocycle metabolic process
RNA processing
chromosome organization
RNA splicing
macromolecule metabolic process
DNA replication
DNA-dependent DNA replication
negative regulation of gene expression
chromatin organization
0
2
4
6
8
-log10(P)

Figure S2.2: Differential expression analysis at gene and transcript level between sperm cell and egg cell. A) Flowchart depicting analysis of gene expression and alternative splicing between sperm cells and egg cell. Numbers of differentially expressed (DE) genes and differential alternative splicing (DAS) genes in sperm cell vs vegetative nucleus comparison. **B)** Bar plots showing the number of genes that are differentially expressed (DEG) and differentially alternatively spliced (DAS), separated by upregulated and downregulated. **C)** Functional enrichment of upregulated (blue bars) and downregulated (red bars) differentially expressed genes between sperm and vegetative nucleus. **D)** Functional enrichment of differentially spliced genes between sperm and egg cell. For detail GO, see Supplemental Dataset 2.7-2.8.

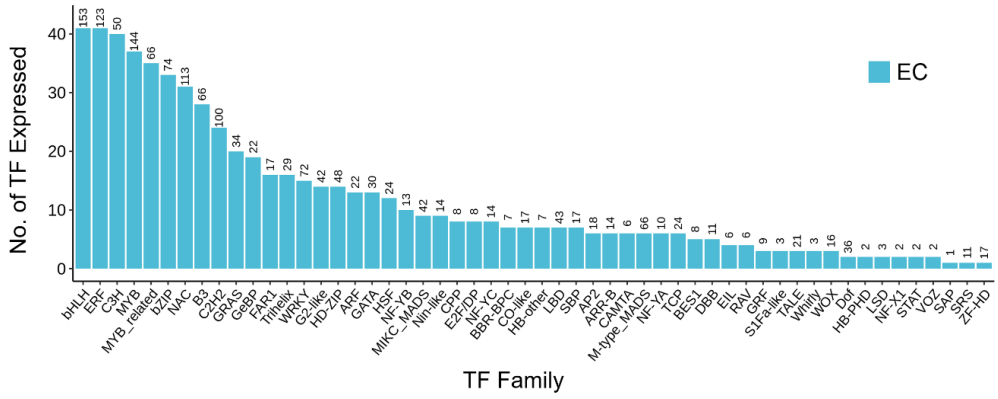


Figure S2.3: Relative frequency of transcription factors (TFs) per family expressed in egg cell (EC).

Supplemental Tables

Table S2.1: Summary of datasets used for transcriptome assembly

Tissue	Total Reads	GEO Dataset	Reference
Root Hair	162253780	GSE85516	Huang et al., 2017
Root	129977077	GSE122772	Tannenbaum et al., 2018
Pollen	52181949	PRJNA194429	Loraine et al., 2013
Meiocyte	217130775	PRJNA342309	Walker et al., 2018
Egg cell	71896362	PRJNA495335	Zhao et al., 2019
Embryo	86389699	GSE121236	Hofmann et al., 2019
Leaf	77668303	PRJEB32665	Mergner et al., 2020
Seedlings	85669601	GSE109150	Birkenbihl et al., 2018
Silique	36049005	PRJEB32665	Mergner et al., 2020
Seed	41916346	PRJEB32665	Mergner et al., 2020
Sperm cell	107395784		This Study
Vegetative nucleus	101586723		This Study
Total Reads	1170115404		

Table S2.2: Quantification of genes undergoing alternative splicing in each cell type

	Sperm cell	Vegetative nucleus	Egg cell
Single-exon genes	921	1226	1613
Multi-exon genes	5859	7041	12041
Multi-transcript genes	2359	3332	6347
Number of gene spliced (%)	40.3	47.3	52.7

Table S2.3: Isoform switch analysis of comparison between sperm cell and vegetative nucleus

Gene_ID	contrast	iso1	iso2	after.FDR
AT5G54940	SC_MPG-VN_MPG	AT5G54940.2	AT5G54940.3	0.006459
AT5G01990	SC_MPG-VN_MPG	AT5G01990.1	MSTRG.23334.2	0.002147
AT4G32440	SC_MPG-VN_MPG	AT4G32440.1	AT4G32440.2	0.006789
AT4G18975	SC_MPG-VN_MPG	MSTRG.20846.7	MSTRG.20846.9	0.002443
AT4G08180	SC_MPG-VN_MPG	AT4G08180.1	AT4G08180.4	0.002587
AT3G03720	SC_MPG-VN_MPG	MSTRG.13057.2	AT3G03720.1	0.006103
AT3G01050	SC_MPG-VN_MPG	AT3G01050.1	AT3G01050.5	0.007262
AT2G21300	SC_MPG-VN_MPG	AT2G21300.5	MSTRG.9585.5	0.001614
AT1G32130	SC_MPG-VN_MPG	MSTRG.3451.1	AT1G32130.1	0.000332
AT1G14685	SC_MPG-VN_MPG	AT1G14685.1	AT1G14685.3	0.009315

Table S2.4: Isoform switch analysis of comparison between sperm cell and egg cell

Gene ID	contrast	iso1	iso2	after.FDR
AT5G66140	SC_MPG-EGG_CELL	AT5G66140.1	MSTRG.29905.4	0.007156
AT5G59710	SC_MPG-EGG_CELL	MSTRG.29158.1	MSTRG.29158.2	0.009021
AT4G36440	SC_MPG-EGG_CELL	MSTRG.22795.3	MSTRG.22795.4	0.008559
AT4G34570	SC_MPG-EGG_CELL	AT4G34570.1	MSTRG.22595.6	0.009559
AT4G34570	SC_MPG-EGG_CELL	AT4G34570.1	AT4G34570.2	0.006552
AT4G31390	SC_MPG-EGG_CELL	MSTRG.22221.1	MSTRG.22221.9	0.000997
AT3G51290	SC_MPG-EGG_CELL	AT3G51290.3	MSTRG.17341.7	0.008745
AT3G22290	SC_MPG-EGG_CELL	AT3G22290.1	MSTRG.15200.1	0.005331
AT3G22290	SC_MPG-EGG_CELL	AT3G22290.1	AT3G22290.2	0.008467
AT3G14890	SC_MPG-EGG_CELL	AT3G14890.2	MSTRG.14336.6	0.008286
AT3G14840	SC_MPG-EGG_CELL	AT3G14840.2	MSTRG.14331.6	0.009032
AT3G08970	SC_MPG-EGG_CELL	AT3G08970.2	MSTRG.13622.10	0.009213
AT1G73350	SC_MPG-EGG_CELL	MSTRG.6968.7	AT1G73350.7	0.003798
AT1G47570	SC_MPG-EGG_CELL	AT1G47570.1	MSTRG.4212.6	0.0068
AT1G09100	SC_MPG-EGG_CELL	AT1G09100.1	MSTRG.918.12	0.008141
AT1G07410	SC_MPG-EGG_CELL	MSTRG.729.1	AT1G07410.1	0.003204

Chapter 3
Transcriptome reprogramming in the Arabidopsis male germline during the progamic phase

3.1. Abstract

A pollen tube's journey through the transmitting tissues of the pistil leads to changes in its gene expression profile. To which extent these changes occur in the vegetative cell (pollen tube) or in the sperm cells it transports has not been addressed so far. This question is of particular importance since sperm cells are believed to acquire competence for fertilization during pollen-pistil interactions. Here, we have compared the transcriptomes of *Arabidopsis thaliana* sperm cells and vegetative nuclei isolated from mature pollen grains and from pollen tubes grown semi *in-vivo.* We used an optimised semi *in-vivo* system and a fluorescent marker line to isolate GFP labelled sperm cells and RFP labelled vegetative nuclei by FACS, followed by a low input RNA-seq protocol to compare the transcriptomes. Moreover, to understand how many genes are elicited in the male gametes and the vegetative nuclei exclusively in the presence of female cues, we also analyzed sperm cells and vegetative nuclei from pollen tubes grown *in vitro*. Our data suggest that sperm cells undergo extensive transcriptomic changes during pollen tube growth, some of which are elicited exclusively when they are passively transported within the pistil, revealing hitherto unidentified transcripts that may be important for sperm maturation and gamete fusion. Similarly, vegetative nuclei undergo even more extensive transcriptome reprogramming during pollen tube growth, mainly through the upregulation of genes associated with pollen tube growth and vesicle mediated transport. Our study provides the most comprehensive overview of transcriptome dynamics of Arabidopsis sperm cells and vegetative nuclei during pollen tube growth to date.

3.2. Introduction

Plant reproduction is a key biological process that starts with the deposition of pollen on female receptive tissue called stigma. The dehydrated pollen then gets hydrated and upon germination a pollen tube extends into the female transmitting tissues of the style until it reaches the ovule funiculus, where it releases the two sperm cells into a synergid (Li et al., 2018).

During double fertilization one sperm cell fuses with the egg cell to give rise to the embryo, while the other fuses with the central cell to form the nourishing endosperm. The journey from pollen germination, tube growth, sperm cell release and finally fertilization involves interaction between at least seven different cell types (synergids, integument, funiculus, septum, transmitting tract, stigma and style). These interactions therefore require multiple changes at transcriptional level as well as physiology (Palanivelu and Tsukamoto, 2012; Johnson, et al., 2019).

Several studies at gene expression level have been conducted to understand the dynamics the transcriptome of growing pollen tubes undergoes (Wang et al., 2008; Qin et al., 2009; Leydon et al., 2018). Arabidopsis pollen tubes can be grown using a suitable pollen germination medium (Boavida and Mccormick, 2007), either on plate (*In vitro*) or by pollinating a pistil and allowing the pollen tubes to grow through the cut end of the pistil, comprising a method called semi *in vivo* pollen tube (SIVPT) growth, mimicking a more natural condition. There are some fundamental characteristics that are shared between pollen tubes grown *in vitro* or through the pistil: The pollen germination and tube growth is highly polarised, extending only at the tip wherein the vegetative nucleus and a pair of sperm cells travel together as a male germ unit localised at the tip of the pollen tube (Palanivelu and Johnson, 2010; Johnson et al., 2019). However, some differences in Arabidopsis pollen tube growth behaviour under these two conditions have also been pointed out. For instance, the rate of pollen tube growth and their length is much

higher when they grow through the pistil than when grown *in vitro* on plate. In addition, pollen tubes grown on plate do not target ovules efficiently, and only when the pollen tubes are first passed through stigma and style can they target ovules more efficiently (Higashiyama et al., 1998; Palanivelu and Preuss, 2006). These results therefore indicate that the journey within the female tissues renders the pollen tubes competent to perceive signals from the ovules. Only recently has been explored how the capacitation of pollen tubes occurs in the presence of female cues in the pistil (Qin et al., 2009; Boavida et al., 2011; Li et al., 2018).

The first documented studies on pollen-pistil interaction pointed towards expression of several transcripts that were induced exclusively when the pollen tubes passed through a pistil (Qin et al., 2009; Boavida et al., 2011; Leydon et al., 2018), including the receptors of ovules attractants and three MYB transcription factors critical for pollen tube burst and release of sperm cells (Leydon et al., 2013).

However, all of these studies were conducted on whole pollen tubes and therefore represent a mix of changes occurring in the two sperm cells and the vegetative cell (the pollen tube cytoplasm and the vegetative nucleus). This is particularly important for changes occurring in the sperm cells, necessitating the study of transcriptomes of sperm cells isolated from growing pollen tubes. One would expect some changes in the sperm cell transcriptome as they are passively transported to the ovules, especially because until recently sperm cells were believed to be under cell cycle control and progress from S phase during anthesis to G2 phase just before fertilization (Friedman, 1999). This notion has been challenged in a recent study suggesting sperm cells are still in G1 phase and complete DNA replication only during zygote development (Liu et al., 2021). Technical challenges with the isolation of sperm cells from growing pollen tubes due to their small size hindered their transcriptome analysis in Arabidopsis, though

studies on sperm cells isolated from *in vitro* grown pollen tubes are available for species that shed pollen at bicellular stage, such as tomato (Liu et al., 2018) and tobacco (Xin et al., 2011).

Despite extensive transcriptomic studies on pollen from several species including Arabidopsis, rice, maize, tobacco and tomato (Reviewed in Rutley and Twell. 2015), transcriptomic data from growing pollen tubes is relatively scarce. Mature pollen grain being the most studied stage in several species including Arabidopsis, large scale transcriptomic data from the progamic phase is only available for Arabidopsis (Qin et al., 2009). Arabidopsis and tobacco are also some of the best studied examples of plant species for which transcriptome data is available for germinating pollen, and pollen tube growth at different time point (Qin et al., 2009; Hafidh et al., 2012; Conze et al., 2017).

Moreover, the transcriptome data on Arabidopsis pollen tubes that have passed through the pistils are based on microarrays (Qin et al., 2009), missing almost 6000 genes from the Arabidopsis transcriptome. The only study analyzing changes in pollen tubes by RNA-seq (Leydon et al., 2018), was performed on whole pistils with the growing pollen tube inside, making it very difficult to identify changes particularly in the sperm cells and vegetative nucleus as they travel within the pollen tube through the pistil.

In order to understand how the transcriptome of the sub-cellular components changes during pollen tube growth, we performed high-throughput RNA-seq on sperm cells and vegetative nucleus isolated from growing pollen tubes - both on plate (*in vitro*) as well as after growth through the pistil (semi *in vivo* pollen tube). For this we have used our optimized protocol (Misra et al., 2019) to collect sperm cells from pollen tubes grown on plate (*in vitro*), hereafter referred to as SC *in vitro* as well as pollen tubes that have passed through the pistil (semi *in vivo* pollen tube), hereafter referred to as SC SIVPT. Similarly, we collected vegetative nuclei from pollen tubes grown on plate (*in vitro*), hereafter referred to as VN *in vitro* and whole pollen tubes that have

passed through the pistil (semi *in vivo* pollen tube), hereafter referred to as PT. We performed low input RNA-seq on these samples providing unprecedented insights into transcriptome dynamics during pollen tube growth, particularly as they travel within the female tissues. We found over 2000 genes to be differentially expressed between sperm cells from mature pollen grain and growing pollen tube, with several genes induced only as sperm cells travelled within pollen tubes within a pistil. Moreover, our analysis of vegetative nuclei and pollen tube transcriptomes revealed over 9000 genes to be differentially expressed across the three pollen conditions.

3.3. Materials and Methods

3.3.1. Plant growth and conditions

Wild type (Col-0) plants and a transgenic marker line harbouring *MGH3p::MGH3-GFP* and *ACT11p::H2B-mRFP* (Borges et al., 2012) were sown on soil in short-day conditions for 8 weeks in 8-h light at 21–23 °C and then transferred to long-day conditions with 16-h light to induce flowering.

3.3.2. In vitro and semi in vivo pollen tube growth assays

Semi *in vivo* pollen tube growth assay was performed as described before (Misra et al., 2019). In short: The Col-0 flowers were emasculated 16h before and were pollinated *in vivo* with the transgenic line for two hours. After two hours, the pistil was excised and cut at the junction of style and ovary and placed on liquid pollen germination medium (Boavida and McCormick 2007) containing 0.01% Boric acid, 5 mM $CaCl_2$, 5 mM KCl, 1 mM $MgSO_4$, 10% sucrose, final volume adjusted to pH 7.5). The pistil was incubated for another 4 hours for the total of 6 h at 22.5 ^oC, for the pollen tubes to emerge from the cut end of the pistil.

For *in vitro* pollen tube growth assay, the flowers were collected and gently shaken with pollen germination medium for 3 minutes, followed by centrifugation at 1811 g for 3 minutes. The supernatant was removed gently and the pellet with released pollen was re-suspended in 1ml of pollen germination medium. The re-suspended pollen pellet in germination media was then spread on the cover slip of a round bottom dish as a thin layer. The dish was covered and incubated in the dark for 6h at 22.5 ^{o}C.

3.3.3. Isolation of sperm cells and vegetative nuclei from pollen tubes

The semi *in vivo* grown pollen tubes and *in vitro* grown pollen tubes were burst using osmotic shock (8% w/v mannitol) and the liquid suspension containing released sperm cells and vegetative nucleus was then immediately transferred to ice-cold sperm extraction buffer containing 1.3 mM H_3BO_3, 3.6 mM $CaCl_2 \cdot 2H_2O$, 0.74 mM KH_2PO_4, 438 mM sucrose, 7 mM MOPS, 0.83 mM $MgSO_4 \cdot 7H_2O$, pH adjusted to 6.0 (Santos et al., 2017).

The liquid suspension medium containing sperm cells and vegetative nuclei was then subjected to FACS as detailed in Santos et al., 2017 and Misra et al., 2019. 500 sperm cells and vegetative nuclei from semi *in vivo* grown pollen tubes and *in vitro* grown pollen tubes were collected and stored at -80 ^{o}C until library preparation.

Since the vegetative nuclei from pollen tubes grown through the pistil (semi *in vivo*) could not be collected, we instead used the whole semi in vivo grown pollen tubes. The whole semi *in vivo* pollen tube growth assay was performed as described above in section 3.3.2., and instead of collecting vegetative nuclei, the whole tubes were harvested as described in detail in Julca et al., 2021.

3.3.4. cDNA library preparation and sequencing analysis

The sorted SC and VN were subjected to library preparation using the optimised Smart-seq2 protocol as described in detail before (Misra et al., 2019). In brief, 500 sorted sperm cells and vegetative nuclei, respectively, were subjected to Smart-seq2 as described earlier (Macaulay et al., 2016) for single cells, with some modifications to improve cDNA yield. For this, a modified bead capture protocol was used, in which 15 ul of oligo(dT) primers attached to Dyna beads were mixed with thawed 500-cell lysates and mixed at 2000 rpm at room temperature for 20 minutes. The tubes containing cell lysate and beads were placed on a magnet for 1 min and supernatant containing gDNA was discarded. The beads were washed twice with gDNA buffer and were then re-suspended with 5 µl of RT mix containing: 0.5 µl of 10 mM dNTP, 1.14 µl of nuclease-free water, 0.06 µl RNAse Inhibitor, 1 µl of 5× Superscript II first-strand buffer, 0.25 µl of 100 mM DTT,1 µl of 5 M Betaine, 0.3 µl of 100 mM $MgCl_2$, 0.5 µl of 10 µM TSO, and 0.25 µl of Superscript II reverse transcriptase. The beads were mixed well with the RT mix at 2000 rpm for 1 min. RT reaction was then carried out by incubating the tubes at 42 °C for 60 min followed by 50 °C for 30 min, and finally enzyme inactivation at 60 °C for 10 min. After RT reaction, PCR amplification was performed using 7.5 µl of PCR mix containing 6.25 µL of 2× KAPA HiFi Hot Start 6 Ready Mix, 0.125 µL of 10 µM ISPCR primer and 1.125 µL nuclease-free water.

PCR amplification was carried out with following conditions: 98 °C for 3 min followed by amplification using 15 cycles (98 °C for 20 s, 67 °C for 15 s, 72 °C for 6 min) and final extension at 72 °C for 5 min. Amplified cDNA was then cleaned with the help of Agencourt AMPure XP SPRI beads. Final sperm cell and vegetative nucleus cDNA libraries were used to confirm size distribution using an NGS high-sensitivity kit on a Fragment Analyser (AATI) at the Genomics Facility of IGC. All cDNA libraries were then converted into Nextera

libraries and sequenced at Novogene, United Kingdom, with 150 bp paired-end reads on an Illumina Novaseq, except whole semi *in vivo* pollen tubes that were sequenced in-house with 75bp single-end reads at the Genomics Facility of IGC.

The raw data was checked for quality and the reads were mapped to Arabidopsis genome TAIR10 using Hisat2 v2.1.0 with command line options "–dta --min-intronlen 60 --max-intronlen 6000 (Kim et al., 2019). The aligned files were then used for gene expression analysis using featurecounts with default parameters (Liao et al., 2014). To identify genes differentially expressed during pollen tube growth, we used DESeq2 (Love et al., 2010). We performed pairwise comparison of sperm cells: SC In vitro vs SC MPG, SC SIVPT vs SC MPG, SC SIVPT vs SC In vitro. Similar analysis for vegetative nuclei data was performed. The differentially expressed genes were then analysed using gprofiler to identify enriched Gene Ontology terms (Raudvere et al., 2019).

3.4. Results

3.4.1 High quality transcriptome data of sperm and vegetative nucleus during pollen tube growth

Transcriptomic studies performed so far on mature pollen grains or on whole growing pollen tubes do not reveal the changes at the resolution of sperm cell and vegetative nuclei that are carried within. Therefore, in order to bridge this knowledge gap, we performed high throughput RNA-seq from sperm cells and vegetative nuclei from 3 pollen development stages: Mature pollen grain (MPG*), in vitro* grown pollen tube and semi *in vivo* pollen tube, as depicted in Figure 3.1.

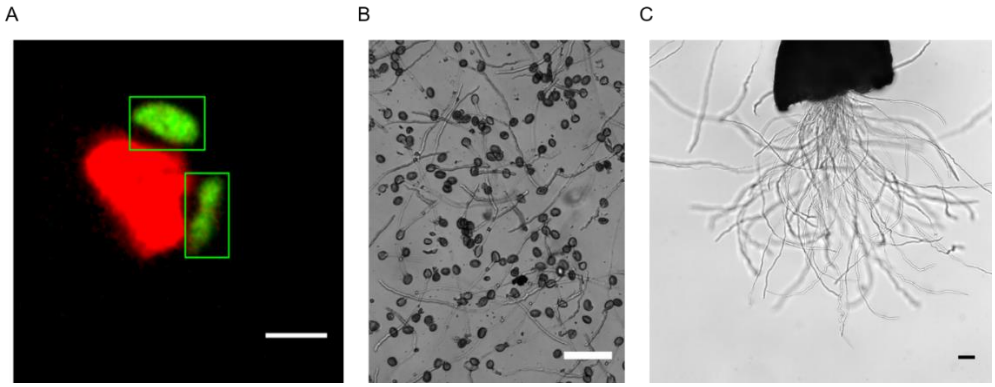


Figure 3.1: Isolation of sperm cells and vegetative nuclei from pollen grain and pollen tube. A) Mature pollen grain with GFP labelled sperm cells and RFP labelled vegetative nucleus. **B)** 6h *in vitro* grown pollen tubes. **C)** Semi *in vivo* grown pollen tubes that have emerged from the cut end of a pistil after 6 hours. Scale bar: **A)** 3um, **B-C)** 100 um.

We have recently optimised a protocol for isolation of pure populations of sperm cells from growing pollen tubes using Fluorescence Activated Cell Sorting (FACS) (Misra et al., 2019). This method allowed us to collect the relatively small number of sperm cells and vegetative nuclei released from growing pollen tubes *in vitro* and/or semi *in vivo* with high efficiency. The gating strategy for FACS is shown in Fig 3.2. Up to 500 sperm cells from mature pollen grains, *in vitro* and semi *in vivo* pollen tubes, and 500 vegetative nuclei from mature pollen grains and *in vitro* pollen tubes, were collected and subjected to an optimised Smart-seq2 protocol (Misra et al., 2019). All libraries were then sequenced with 150bp paired-end reads, except whole semi *in vivo* pollen tubes, which were sequenced at 75bp single-end.

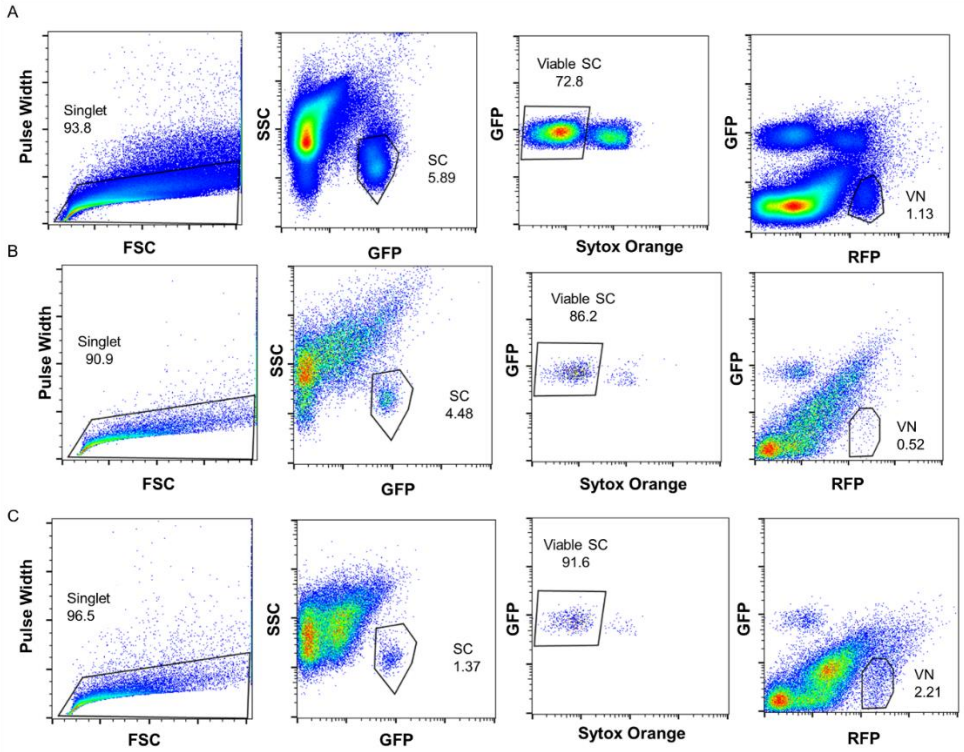


Figure 3.2: Flow cytometry profile of sperm cells and vegetative nucleus: A) Gating strategy established using sperm cells and vegetative nuclei from mature pollen grains, used as template for subsequent sorting. **B)** Gating strategy used for sorting sperm cells and vegetative nuclei from *in vitro* grown pollen tubes. **C)** Gating strategy used for sorting sperm cells and vegetative nuclei from semi *in vivo* grown pollen tubes. Populations of gated sperm cells and vegetative nuclei are marked by enclosed boxes. SSC side scatter, FSC Forward scatter. SC Sperm cells, VN Vegetative nuclei. Sytox orange is used as cell viability dye that stains only non-viable cells.

Since, we could not successfully obtain vegetative nuclei from semi *in vivo* pollen tubes, we used whole semi *in vivo* grown pollen tubes as proxy for VN SIVPT instead. However, in order to have a better approximation of vegetative nuclei expression from semi *in vivo* conditions, we applied two filtering criteria to remove genes in whole tube data that may be contributed by sperm cell expression: 1) Expression of a gene in SC SIVPT ≥ five times

the expression in whole pollen tube. 2) Expression of a gene in SC SIVPT should be ≥ 500 TPM. While the first criterium of selection will make sure that we exclude genes that are expressed in sperm cells and at low level in pollen tubes, hence likely to reflect contribution of the sperm cell transcriptome, the second criterium helps to remove genes that may be expressed at lower level in SCs but still fulfilling the first criterium. Approximately 140 genes fulfilled both the criteria and were therefore excluded from all the downstream analysis involving whole pollen tube data and hereafter this filtered data will be called as PT (Supplemental Dataset 3.6). A summary of RNA-seq statistics is given in Table 3.1.

Table 1: Statistics of RNA-seq libraries from different samples

Sample Type	Number of Reads	Cleaned Reads	Unique aligned
SC MPG Rep1	33365386	29023164	25533293
SC MPG Rep2	53689070	46885797	39770023
SC MPG Rep3	20341328	17670818	14202343
SC In vitro Rep1	19625374	16682965	13370290
SC In vitro Rep2	51985274	45784072	39621065
SC In vitro Rep3	45374222	40783526	35470871
SC SIVPT Rep1	31348698	27129991	23275793
SC SIVPT Rep2	52483069	46220997	39337114
SC SIVPT Rep3	41171322	36563118	32508544
VN MPG Rep1	52826859	43984317	39024275
VN MPG Rep2	21061338	16151810	13179352
VN MPG Rep3	27698526	22255287	17992732
VN In vitro Rep1	21437563	17159458	14259746
VN In vitro Rep2	50996011	41058109	33853929
VN In vitro Rep3	36892598	31549191	26993481
PT Rep1	19654665	19648770	16116008
PT Rep2	20140468	20135710	16958101
PT Rep3	20409335	20402267	16853533

SC Sperm cells, VN Vegetative nuclei, MPG Mature Pollen Grain, SIVPT Semi *in vivo* pollen tube, PT- Whole semi *in vivo* pollen tube.

3.4.2. Arabidopsis sperm cells and vegetative nuclei undergo transcriptomic changes during pollen tube growth

The journey within female tissues elicits changes in the pollen tube transcriptome, but to what extent those changes are contributed by sperm cells remained elusive so far. Therefore, to understand the transcriptome dynamics of sperm cells during pollen tube growth, we compared SC MPG,

SC in vitro and SC SIVPT. Our results indicate that the size and complexity of the sperm cell transcriptome was different in all the three conditions studied. The number of genes expressed (TPM ≥1) in SC MPG, SC in vitro and SC SIVPT were 7872 (539 non-coding), 9224 (1232 non-coding) and 7339 (590 non-coding), respectively (Figure 3.3A). We also noticed a considerable overlap in the number of expressed genes between the three conditions, though small proportions of genes were found to be exclusively expressed in each of them, indicating that sperm cells from the three pollen developmental conditions may express a unique set of genes that may perform specific functions.

We then determined the expression level of protein coding and non-coding genes using TPM in sperm cells across the three pollen conditions. Our results indicate that, overall the expression level of protein coding genes was higher than that of the non-protein coding genes, except in SC In vitro. SC In vitro tend to show higher expression compared to the other two pollen conditions, indicating a potential function of non-coding transcriptomes in sperm cells from these three conditions (Figure 3.3B).

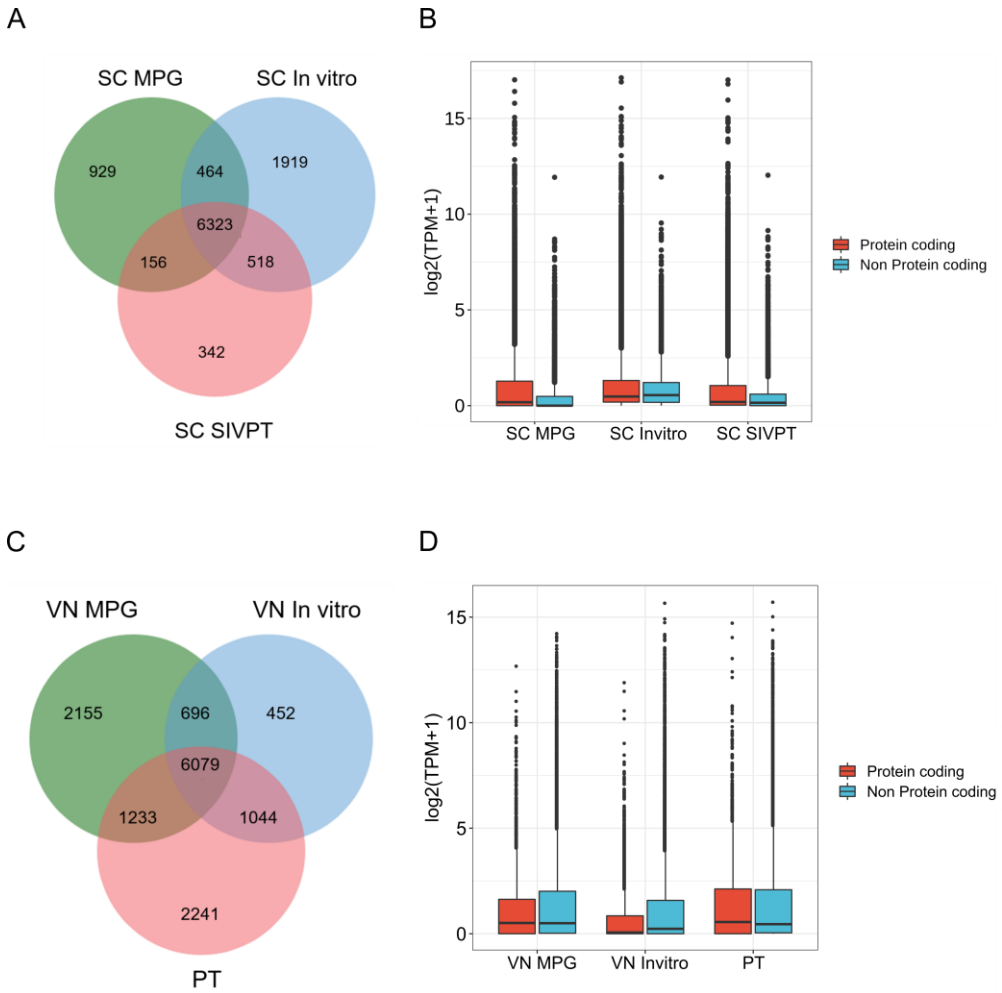


Figure 3.3: Transcriptomes of Arabidopsis sperm cells and vegetative nucleus from three pollen development stages. A) Venn diagram showing overlap between expressed genes in sperm cells between the pollen development stages at average TPM>1. **B)** Box plot representation of protein coding and non-coding gene expression in sperm cells from three pollen development stages. **C)** Venn diagram showing overlap between expressed genes in vegetative nucleus between the pollen development stages at average TPM>1. **D)** Box plot representation of protein coding and non-coding gene expression in vegetative nucleus from three pollen development stages. SC Sperm Cell, VN Vegetative nucleus, MPG Mature Pollen Grain, SIVPT Semi *in vivo* pollen tube, PT- Whole *semi in vivo* pollen tube.

As seen in sperm cells, our data from vegetative nuclei also showed a similar trend of gene expression across three pollen development conditions. While several expressed genes overlapped across the three conditions, each of them also expressed a unique set of genes (Fig 3.3C). Interestingly, the large number of genes exclusively expressed in semi *in vivo* pollen tubes is in contrast to our result in SC SIVPT indicating that the VN transcriptome during semi in vivo growth undergoes larger transcriptomic changes compared to sperm cells. Also, overall the vegetative nuclei showed almost similar expression profiles between protein coding and non-protein coding genes, in contrast to what we observed for sperm cells, underlining the different transcriptomic fates of the two cell types during pollen tube growth (Figure 3.3D).

In order to get an overview of the changes in the transcriptome of the sperm cells from mature pollen grains when compared to sperm cells from the growing pollen tubes, we performed Principal Component Analysis (PCA). The sperm cell transcriptome from mature pollen grains (SC MPG) was significantly different from the transcriptomes of sperm cells from growing pollen tubes (Figure 3.4A – PC1), while the growing pollen tube stages separated also, albeit with less variance, along PC2. This indicates that the sperm cells from *in vitro* pollen tubes as well as semi *in vivo* pollen tubes might share a similar development trajectory and the only difference in the two pollen tube stages might be associated with transcriptional fine tuning that would occur in the presence of female cues.

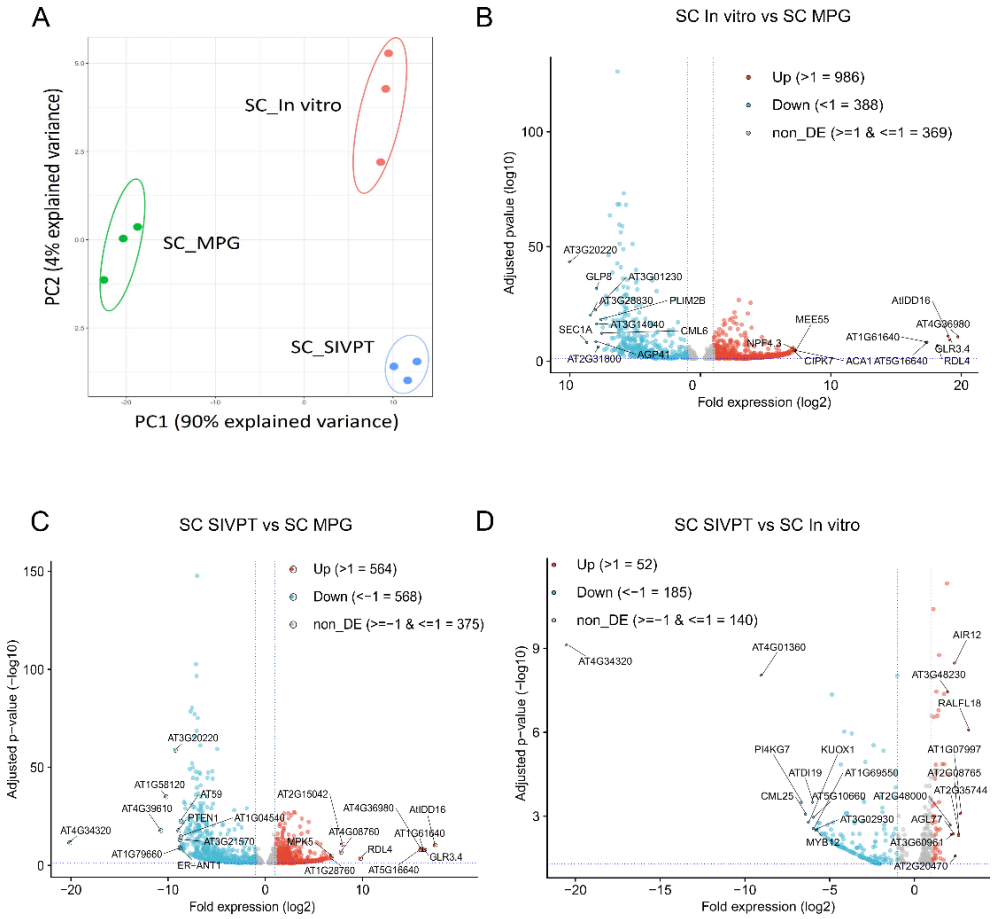


Figure 3.4: Transcriptome changes in sperm cells during pollen tube growth. **A)** Principle component analysis of sperm cell transcriptome data from three pollen development stages: mature pollen grain, pollen tube grown *in vitro* and pollen tube grown through pistil. Principal Component 1 (PC1, x-axis) represents 90 % and PC2 (y-axis) represents 4% of the total variation in the transcriptome data of sperm cells. **B-D)** Volcano plots showing differential gene expression in the three pairwise comparisons: SC In vitro vs SC MPG, SC SIVPT vs SC MPG, and SC SIVPT vs SC In vitro, respectively. MPG: Mature Pollen Grain; SIVPT: Semi *in vivo* pollen tube growth. Only genes with log2FC ≥1 and FDR ≤0.05 are represented. Complete list can be found in Supplemental Dataset 3.1-3.3.

Similarly, when we looked at the transcriptome of vegetative nuclei across the three pollen conditions, our results indicated that the transcriptome of vegetative nuclei from mature pollen grains was significantly different from that of the two pollen tube stages (PC1 - Figure 3.5A). We also noticed that the difference between two pollen tube conditions was more pronounced in vegetative nuclei (PC2) when compared to sperm cells (Figure 3.5A). This could potentially be due to the different kind of datasets used in vegetative nuclei comparisons: whole pollen tubes used for vegetative nuclei comparison vs pure isolated sperm cell from three pollen conditions. Nevertheless, as expected, our data suggests that vegetative nuclei undergo more dynamic transcriptomic changes compared to sperm cells during pollen tube growth.

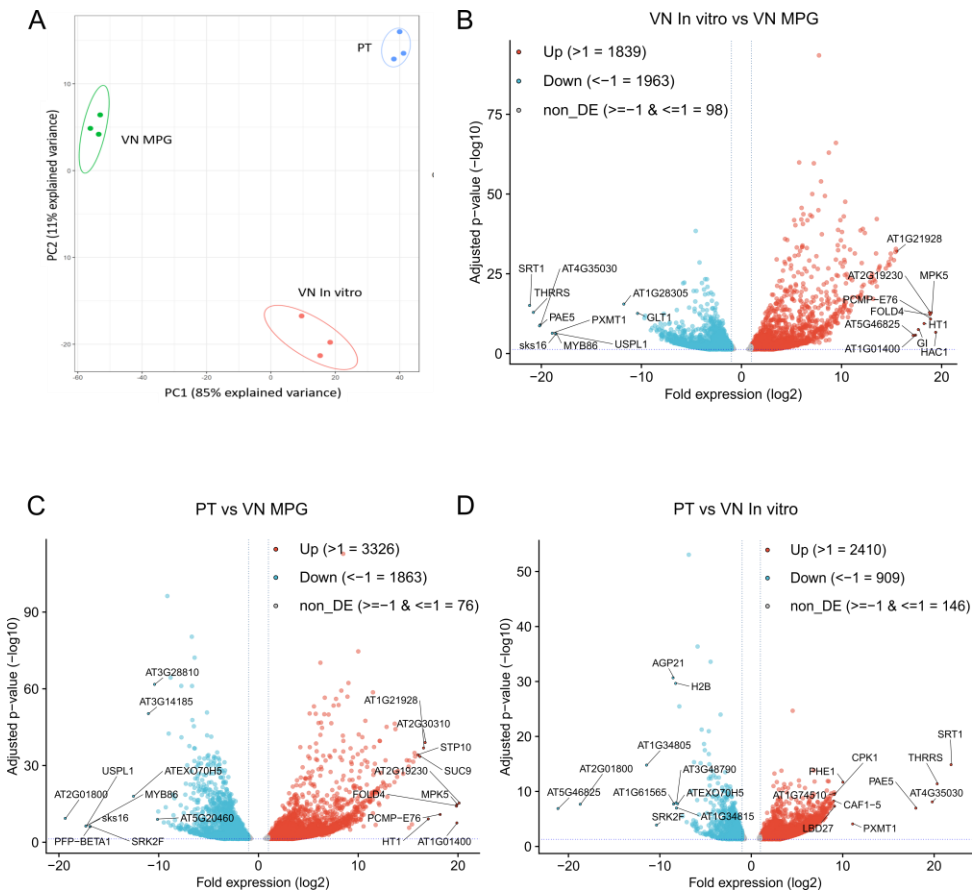


Figure 3.5: Transcriptome changes in vegetative nuclei during pollen tube growth. A) Principle component analysis of vegetative nuclei transcriptome data from three pollen development stages: mature pollen grain, pollen tube grown *in vitro* and the whole pollen tube grown through pistil. Principal Component 1 (PC1, x-axis) represents 85% and PC2 (y-axis) represents 11% of the total variation in the transcriptome data of vegetative nuclei. **B-D)** Volcano plots showing differential gene expression in the three pairwise comparisons: VN In vitro vs VN MPG, PT vs VN MPG, and PT vs VN In vitro, respectively. MPG: Mature Pollen Grain; PT: Whole semi *in vivo* pollen tube growth. Only genes with log2FC ≥1 and FDR ≤0.05 are represented. A complete list can be found in Supplemental Dataset 3.7-3.9.

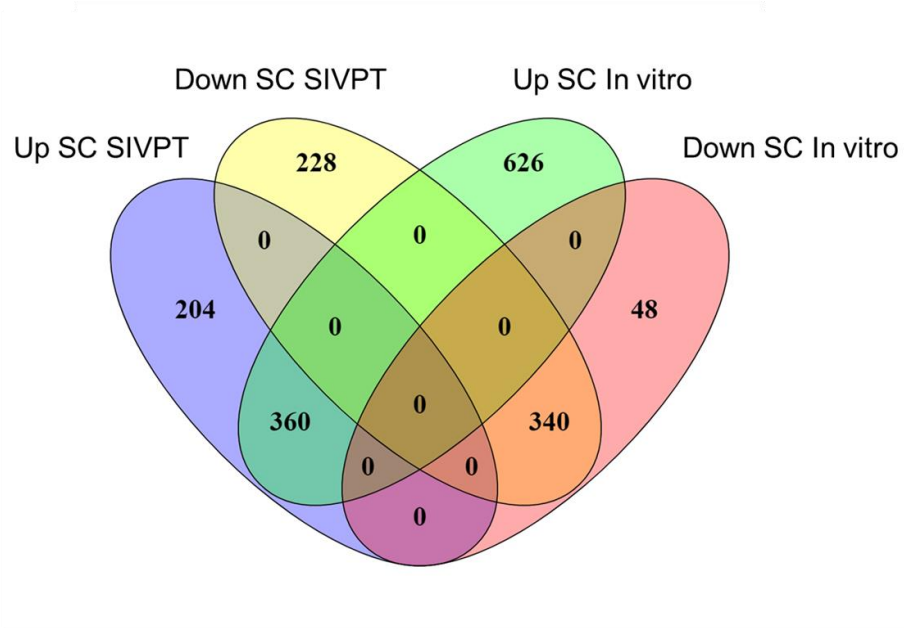
In order to understand how the pollen tube journey within the female tissues affects gene expression levels in the sperm cells, we performed pair-wise comparisons: SC SIVPT vs SC MPG, SC In vitro vs SC MPG, SC SIVPT vs SC In vitro. When we compared SC In vitro vs SC MPG, we found that 986 genes were upregulated while 388 were downregulated in SC In vitro (Figure 3.4B). Similarly, comparison of SC SIVPT vs SC MPG revealed 564 genes to be upregulated and 568 genes to be downregulated in SC SIVPT (Figure 3.4C). Direct comparison of sperm cells from pollen tube data identified 237 genes (52 up and 185 downregulated) to be differentially expressed between SC SIVPT vs SC In vitro (Figure 3.4D).

We then looked at how vegetative nuclei transcriptomes change during pollen tube growth by performing the following pair-wise comparisons: PT vs VN MPG, VN In vitro vs VN MPG, PT vs VN In vitro. When we compared VN In vitro vs VN MPG, we found that 1839 genes were upregulated while 1963 were downregulated in VN In vitro (Figure 3.5B). Similarly, comparison of PT vs VN MPG revealed 3210 genes to be upreguated and 1858 genes to be downregulated in PT (Figure 3.5C). Direct comparison of two pollen tubes (PT vs VN In vitro) revealed a much larger number of genes compared to sperm

cells: 2306 genes upregulated and 907 genes to be downregulated in PT (Figure 3.5D).

When we compared genes that were differentially expressed in sperm cells during pollen tube growth, we found 204 genes to be exclusively upregulated in SC SIVPT (Figure 3.6A). On the other hand, 1651 genes were exclusively upregulated in whole semi *in vivo* pollen tubes (used as a proxy for the expression in VN SIVPT stage), which is significantly higher than in sperm cells indicating that vegetative nuclei undergo far more transcriptomic changes when passed through female tissues (Figure 3.6B).

A


Down SC SIVPT
Up SC In vitro
Up SC SIVPT
Down SC In vitro
228
626
204
48
360
340
0
0
0
0
0
0
0
0
0
0
0

B

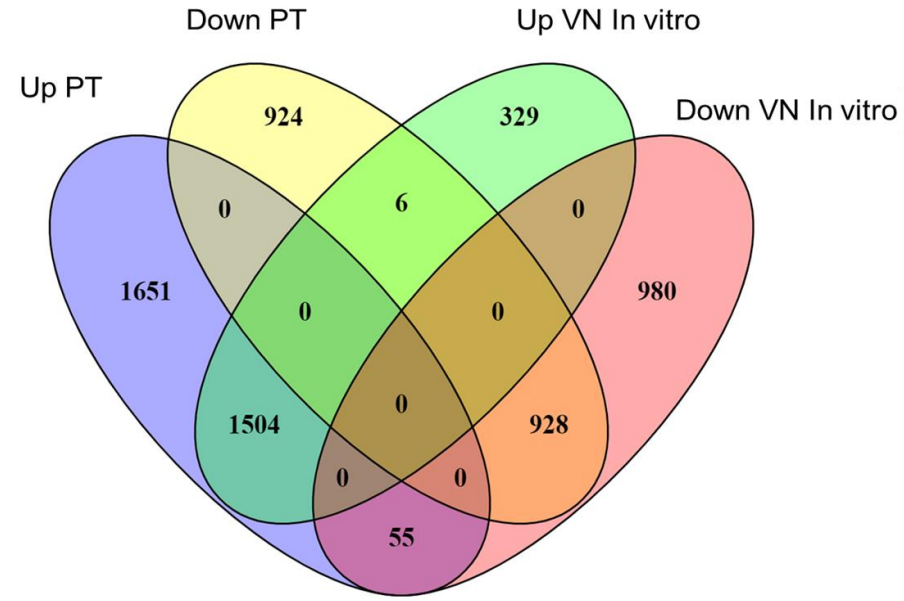

Down PT
Up VN In vitro
Up PT
Down VN In vitro
924
329
1651
980
1504
928
55
6
0
0
0
0
0
0
0
0
0
0

Figure 3.6: Venn diagram showing number of genes that are differentially expressed in sperm cells (A) and vegetative nuclei (B) during pollen tube growth. SIVPT Semi *in vivo* pollen tube growth, PT Whole semi *in vivo* pollen tube growth, SC Sperm cells, VN Vegetative nuclei.

All differentially expressed genes in sperm cells and vegetative nuclei across the three pollen development conditions are shown in Figure 3.7.

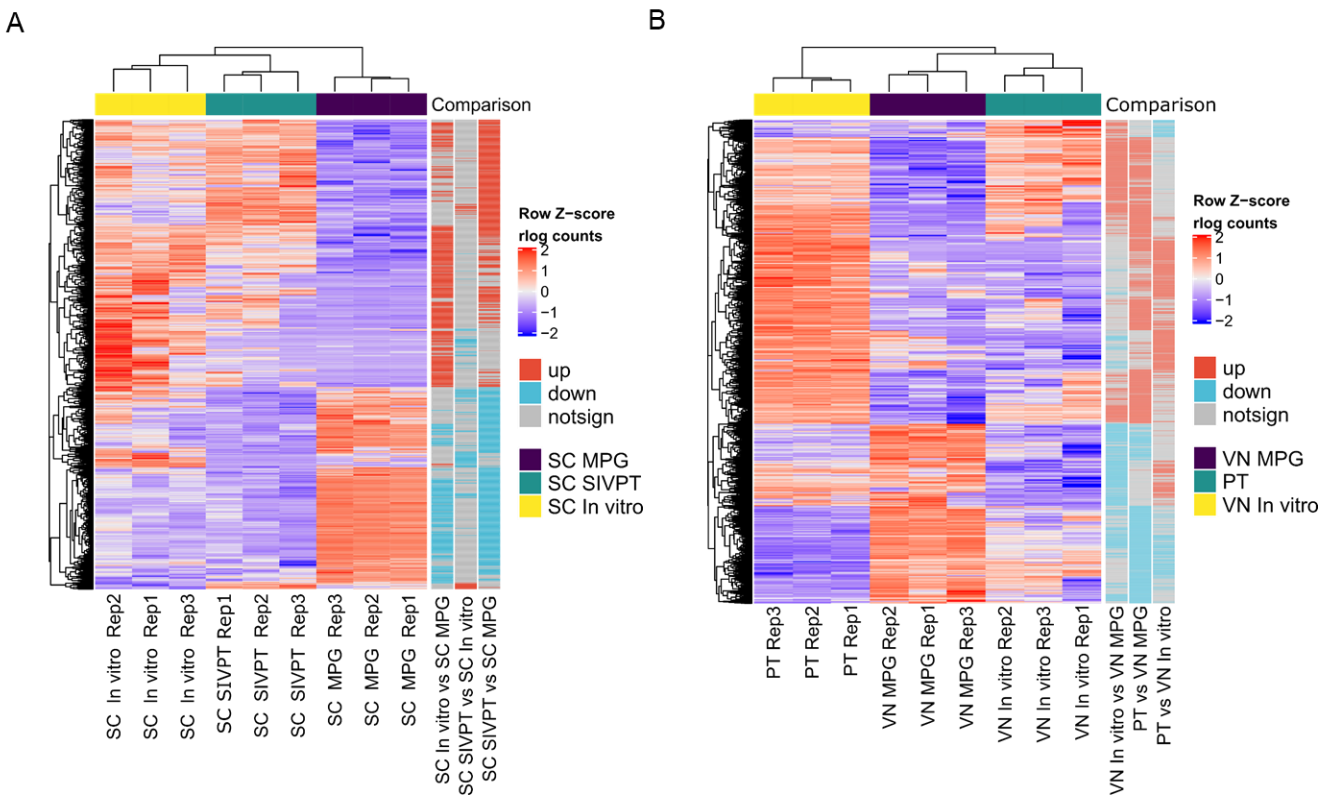


Figure 3.7: Heatmap of the differentially expressed genes (log2FC ≥1 and FDR ≤0.05) in sperm cells (A) and vegetative nuclei (B) from three pollen development conditions: Mature pollen grain (SC MPG and VN MPG), *In vitro* grown pollen tube (SC In vitro and VN In vitro) and semi *in vivo* pollen tube (SC SIVPT and PT).

When we looked at the expression level of the 52 genes, we found many of the genes to be significantly upregulated in SC SIVPT (Figure 3.8). For instance, expression of RALF18 (AT2G33130) was increased 17.98 and 9.46-fold in SC SIVPT compared to SC MPG and SC In vitro, respectively (Fig 3.8A).

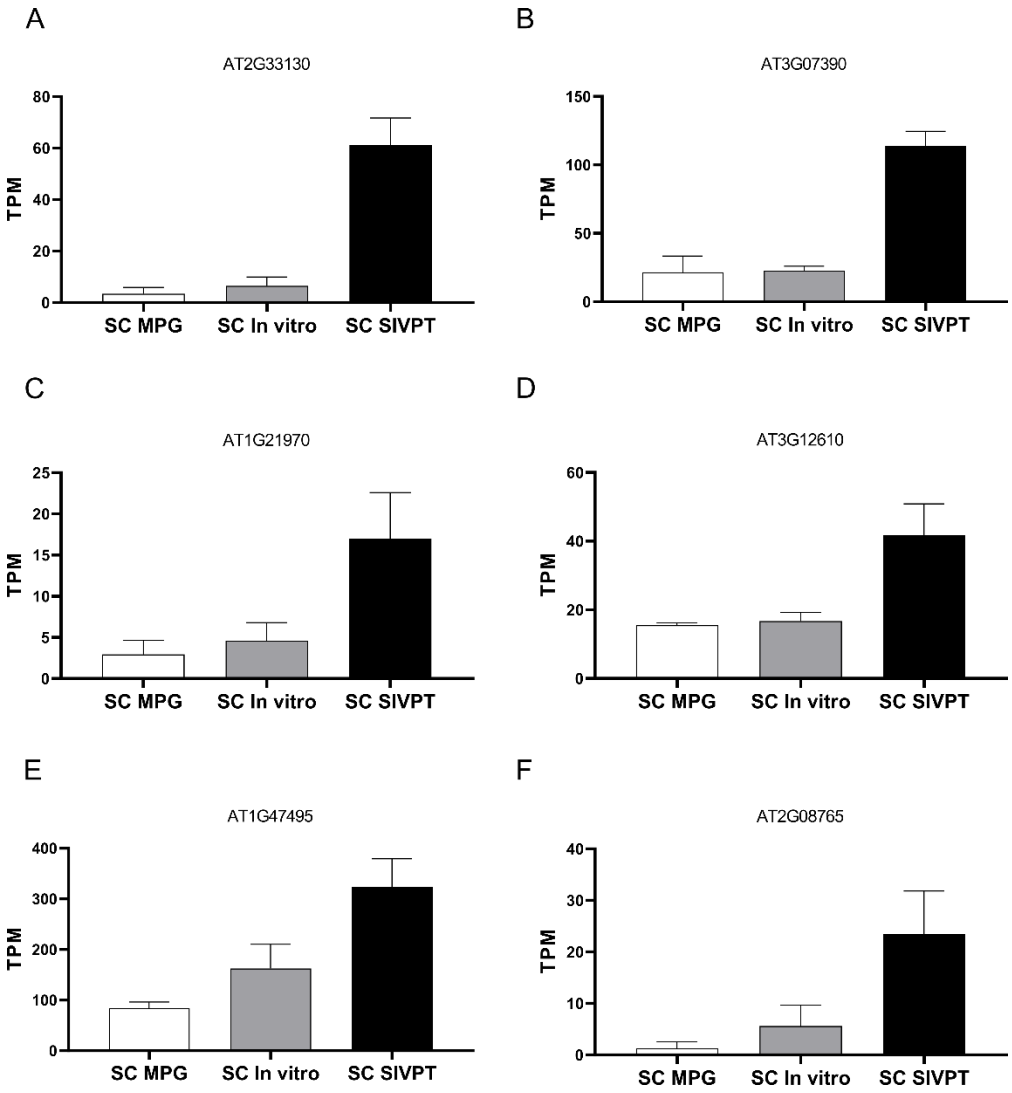


Figure 3.8: Examples of genes that are significantly upregulated during the progamic phase in sperm cells. Significant upregulation was seen for both protein coding genes **(A-E)** and non-protein coding genes **(F).** Y- axis shows expression level of genes as TPM values (means ± SD) of three biological replicates.

3.4.3. Transcriptional regulation of sperm cells and vegetative nuclei during pollen tube growth

Male germline specification during pollen development is under tight transcriptional control and R2R3-MYB type DUO1 mediated sperm cell differentiation is well documented (Durbarry et al., 2005; Rotman et al., 2005; Borg et al., 2011). How this transcriptional control unravels during sperm cells while they are delivered to the female ovules was the question we asked ourselves. When we looked at the transcriptome data of sperm across the three pollen development stages, we found a total of 121 transcription factors belonging to 30 different transcription factor families, that were differentially expressed (Figure 3.9). We found 59 transcription factors to be differentially expressed (55 up and 4 down) in SC In vitro vs SC MPG comparison, 51 transcription factors (43 up and 8 down) in SC SIVPT and SC MPG comparison and 11 transcription factors (3 up and 8 down) in SC SIVPT vs SC In vitro comparison.

Figure 3.9: Dot plot showing fold change of differentially expressed transcription factors in sperm cells from different pollen development stages. Each transcription factor is symbolized by a dot. Red and blue denote up- and downregulation, respectively.

On the other hand, when we looked at the vegetative nuclei transcriptome data, a total of 288 transcription factors belonging to almost all (51) class of transcription factor families were found to be differentially expressed (Figure 3.10). We found 153 transcription factors to be differentially expressed (74 up, 79 down) in VN In vitro vs VN MPG comparison, 213 transcription factors (142 up, 71 down) in PT vs VN MPG comparison, and 125 transcription factors (92 up, 33 down) in PT vs VN In vitro comparison.

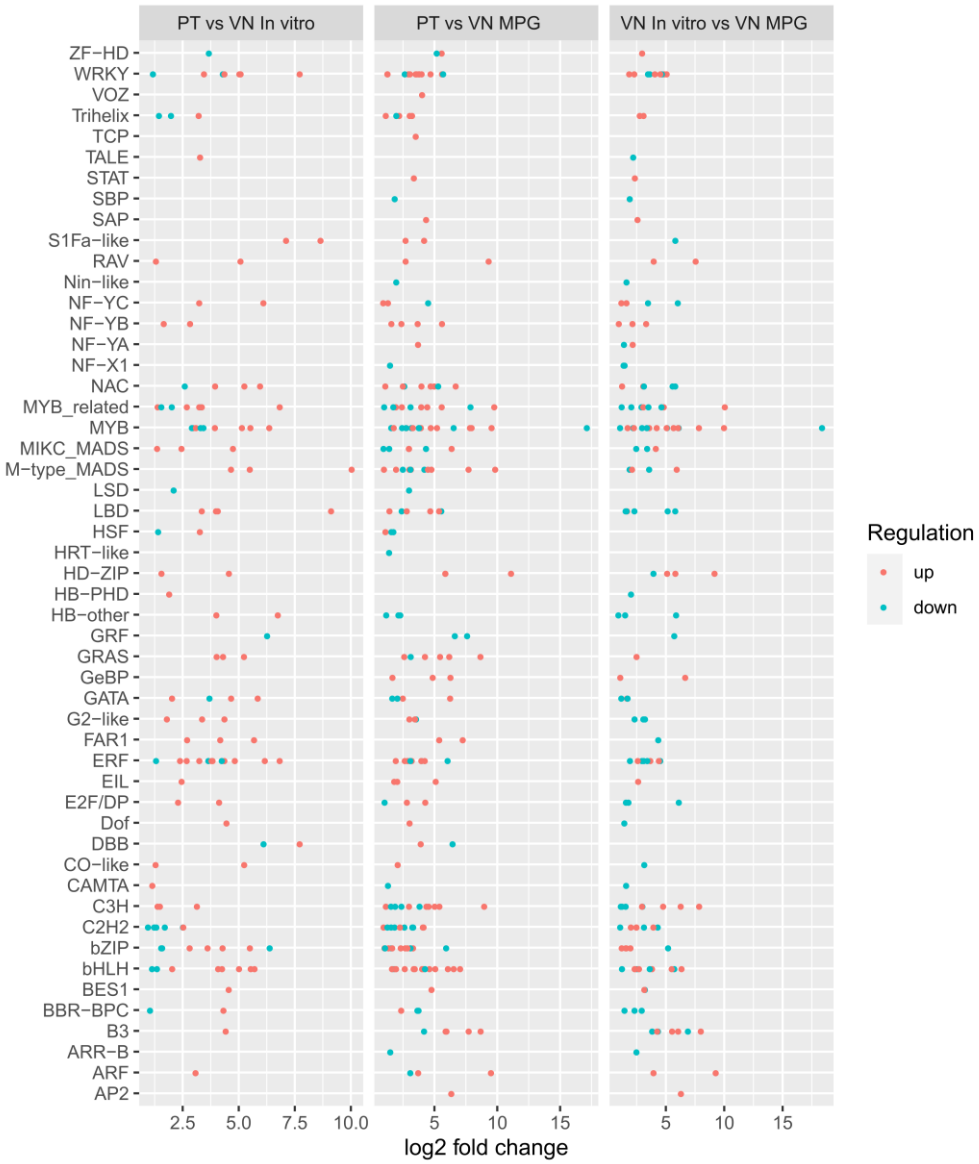


Figure 3.10: Dot plot showing fold change of differentially expressed transcription factors in vegetative nuclei from different pollen development stages. Each transcription factor is symbolized by a dot. Red and blue denote up- and downregulation, respectively.

3.4.4. Functional Categories of differentially expressed genes

Next, we asked ourselves what were the potential functions of the differentially expressed genes. When we looked at the genes that are differentially expressed between SC In vitro and SC MPG, we found upregulated genes to be associated with response to abiotic and chemical stimuli, oxidative stress and protein folding (Figure 3.11A). On the other hand, when we compared SC SIVPT and SC MPG, we found genes associated with heat and oxidative stress, chaperone mediated protein folding to be upregulated (Fig 3.11B). We could not find enrichment for the set of 52 genes that were up regulated during SC SIVPT vs SC In vitro comparison. However, some of those genes might be involved in sperm maturation and capacitation, and are therefore primary targets for future research.

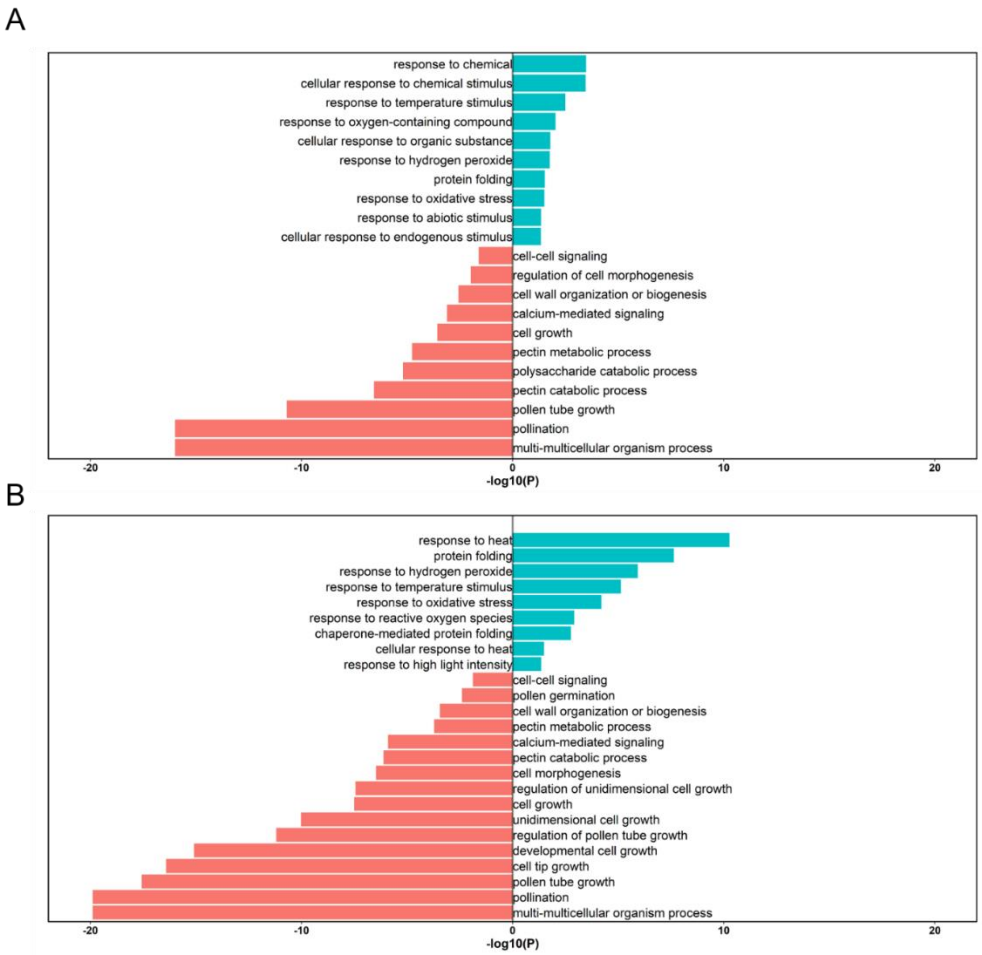


Figure 3.11: Functional enrichment of genes differentially expressed in sperm cells during pollen tube growth. A) GO analysis of genes upregulated or downregulated in comparison of SC In vitro vs SC MPG. **B)** GO analysis of genes upregulated or downregulated in comparison of SC SIVPT vs SC MPG. Blue bar: Upregulated, Red Bar: Downregulated. For detail GO, see Supplemental Dataset 3.4-3.5.

Interestingly, our data on sperm cells from growing pollen tubes revealed that genes associated with pollen tube growths, cell-cell signalling, and pectin metabolic process were downregulated, indicating that the genes that are necessary for pollen germination and tube growth are no longer needed in

the sperm cells (Fig 3.11). This is important as sperm cell and vegetative nucleus are derived from the same progenitor cells, but during germline specification, genes associated with function of each cell type are enriched in sperm cell or vegetative nucleus. Therefore, the downregulation of genes associated with pollen tube functions in sperm cells may be the residual transcripts from the previous developmental stages that continue to be repressed as sperm prepare for the next steps, including gamete fusion.

When we looked at the differentially expressed genes in vegetative nuclei, we found that genes that are involved in cell tip growth, vesicle mediated transport, and response to stimulus were upregulated (Supplemental Figure S3.1), irrespective of whether vegetative nuclei were from pollen tubes grown in vitro (VN In vitro) or in the whole semi *in vivo* (PT) data. These results indicate that overall pollen tube growth triggers similar transcriptomic changes during in vitro or semi *in vivo* pollen tube growth. However, if we look at the genes that are enriched in PT vs VN In vitro, we found genes associated with DNA replication, cell cycle regulation, and translation were upregulated, indicating the role of genes that may be involved specifically in the pollen tube growth (Figure 3.12).

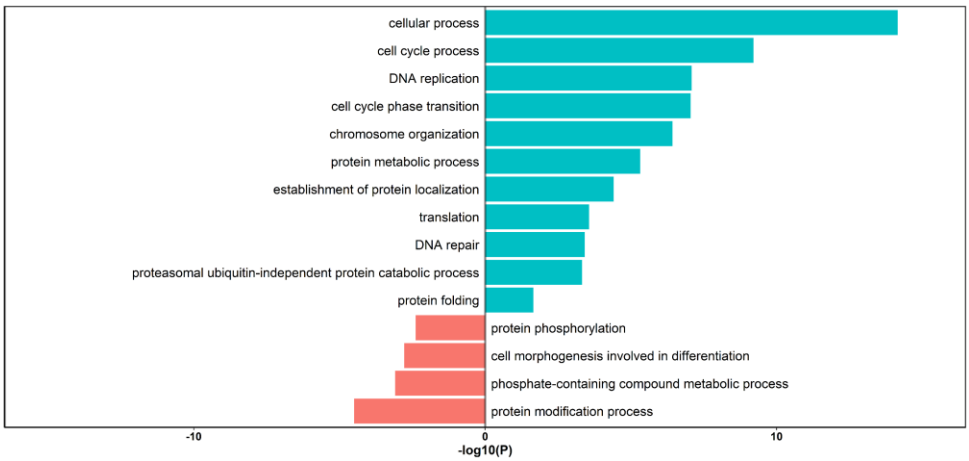


Figure 3.12: Functional enrichment of genes differentially expressed in vegetative nuclei during pollen tube growth. GO analysis of genes upregulated or downregulated in comparison of PT vs VN In vitro. Blue bar: Upregulated, Red Bar: Downregulated. For detail GO, see Supplemental Dataset 3.12.

3.5. Discussion

A pollen tube's journey within the female tissue is a highly transcriptionally active process and 1257 genes had been identified as upregulated during this process (Qin et al., 2009). However, the relative contribution of sperm cells and vegetative nuclei to the growing pollen tube's transcriptome remained to be deciphered. Therefore, we used our optimised method for the collection of sperm cells and vegetative nuclei from growing pollen tubes (Misra et al., 2019). Our FACS based method was very efficient in collecting even small numbers of cells from growing pollen tubes, which were then subjected to high throughput transcriptome analysis.

Our transcriptome data revealed a reduced size of the transcriptome in SC SIVPT compared to the other two pollen development conditions (Figure 3.3B). This might be due to a reduced transcriptional rate in the sperm cells

as they mature and prepare for gamete fusion. On the contrary, the lack of female cues mediating transcriptional fine tuning might also explain the larger number of genes expressed in SC In vitro.

Our analysis of the transcriptome data from three pollen development conditions reveals that the sperm cells undergo transcriptional reprogramming during pollen tube growth as several genes were found to be upregulated (Figure 3.4A). To what extent the presence of female cues reshapes the transcriptome of sperm cells when they travel within the pollen tube through the pistil might help to reveal genes important for either sperm maturation or gamete fusion. Two main observations from our data stand out: Firstly, sperm cells undergo strong transcriptomic changes during pollen tube growth, irrespective of whether they are obtained from pollen tubes grown *in vitro* or passed through a pistil. This indicates that changes at the transcriptomic level occur also in the sperm cells, which was impossible to ascertain from transcriptome data of whole pollen tubes. Secondly, our data also indicate differences between transcriptomes of sperm cells from In vitro and SIVPT stage, suggesting that the two pollen tube growth conditions also affect the transcriptome of sperm cells, albeit less compared to what we observe between mature pollen grain (SC MPG) and pollen tubes (SC In vitro and SC SIVPT).

Comparison of our sperm cell data with the previous study on whole pollen tube data (Qin et al., 2009) also highlights the resolution at which we were able to identify the changes occurring exclusively in sperm cells, as only 189 genes were found to be differentially expressed in pollen grain vs pollen tube data (Qin et al., 2009) and overlapped with our analysis. There are several possible reasons that could explain this difference: first, the previous study was performed using arrays and probes for some of these transcripts might not have been present on arrays. Secondly, we have grown the pollen tubes for 6h in both SIVPT and *in vitro* conditions, while in the previous study only

4h of growth for *in vitro* and SIVPT were used. To which extend this time difference of pollen tube growth can explain the difference in the studies is not clear. Thirdly, and probably the most parsimonious explanation, could be due to the dilution effect of sperm cell transcripts in the data for whole pollen tubes. The changes hence could only be identified when sperm cells were analysed separately as shown in this study.

Sperm cells from mature pollen grains were found to be transcriptionally active, a hypothesis that holds true even when they are passively transported within pollen tubes to the female ovules. Our data indicate 1374 and 1132 genes to be differentially expressed in SC In vitro vs SC MPG and SC SIVPT vs SC MPG comparisons, respectively, indicating that sperm cells are not in a transcriptionally quiescent stage while they are transported (Figure 3.4 B-C).

Our transcriptome data showed that many proteins which are typically associated with stress responses in plants (eg HSP90-1, HSP70-5) were significantly upregulated in sperm cells during pollen tube growth. Recently, the role of HSP90´s have been studied extensively either under stress conditions and/or to understand their diverse role in the regulation of plant growth and development (Tichá et al., 2020). For instance, HSP90 was recently found to be involved in signaling pathways during embryo development as impaired function of HSP90 together with the Yoda kinase pathway leads to defective development of suspensor and embryo proper. This new function of heat shock proteins (e.g. HSP90) to be involved in modulating transcriptional networks during early embryogenesis is adding to a growing body of evidence for their role in plant reproduction (Samakovli et al., 2021). Similarly, heat shock proteins have been extensively studied during mammalian spermatogenesis for their possible in male germ cells differentiation and sperm maturation and capacitation (Reviewed in Abane and Mezger, 2010; Dun et al., 2012), but whether they perform similar

functions during plant gametogenesis and particularly in non-flagellated sperm cells remains to be seen. Moreover, the upregulation of these heat shock proteins in sperm cells during pollen tube growth may also be associated with developmental priming (Chaturvedi et al., 2013; Chaturvedi et al., 2016). This mechanism serves to protect plants against sudden stresses, however it remains to be seen whether changes in these proteins also have some developmental functions in non-stressed plants and might influence sperm specification and maturation during the fertilization process. Based on the upregulation of several heat shock proteins during pollen tube growth, we can surmise that this specific class of proteins might be synthesized in order to meet any metabolic demands or environmental stresses and to prepare the male gametes for double fertilization (Chaturvedi et al., 2013; Ischebeck et al., 2014).

Pollen-pistil interaction and plant-pathogen interaction have received significant attention from researches as they both tend to elicit similar transcriptomic changes (Elleman and Dickinson, 1996; Mondragón-Palomin et al., 2017). Several studies that have analysed the transcriptome during pollen-pistil interaction have reported enrichment of genes linked to stress or defense responses (Tung et al., 2005; Boavida et al., 2011; Leydon et al., 2017; Mondragón-Palomin et al., 2017; Osaka et al., 2019).

Our analysis also revealed that several genes involved in ubiquitin-mediated proteolysis were over-represented in the sperm cells from growing pollen tubes. Several studies have shown enrichment of these genes in sperm cells of rice, Arabidopsis or lily (Okada et al., 2006; Okada et al., 2007; Borges et al., 2008; Russel et al., 2012;). However, these studies were done on sperm cells from mature pollen grains, but our data indicating their upregulation also during pollen tube growth suggests that the proteolysis pathway may be important for specification of sperm cell function.

In order to identify genes that may be induced specifically in the presence of female cues, we compared SC SIVPT with SC In vitro, resulting in a list of 52 upregulated genes. Many of these genes may have potential functions in gamete fusion. For instance, Rapid Alkalinization Factor 18 (RALF18) is significantly upregulated in SC SIVPT (Figure 3.8A). Interestingly RALF18 is one of the 9 RALFs expressed in Arabidopsis pollinated pistils, as shown in a previous study (Ge et al., 2017), but in our study we could show that some if not all of this expression in pistil might be contributed by sperm cells. Very few RALFS have been functionally characterised so far that may be involved in male or female gametophyte development or during fertilization. For instance, out of 37 members of the RALF family, RALF4 and RALF19 are the only pollen RALFs that have been functionally characterised to date to be involved in pollen tube growth while they travel within the pistil (Mecchia et al., 2017). Interestingly, 9 RALF genes including RALF19 were also found to be upregulated in our PT vs VN MPG comparison and VN In vitro vs VN MPG comparison, indicating the potential functions of these RALF proteins during pollen tube growth. Functions of several RALFs remain unknown particularly in plant reproduction, including RALF18 induced in SC SIVPT.

In addition, our data showed upregulation of MILDEW RESISTANCE LOCUS O 6 (MLO6) gene in SC SIVPT. MLOs proteins are 15-member family in Arabidopsis, localised on plasma membrane and first identified in barley for powdery mildew susceptibility (Büschges et al., 1997). Several members (MLO1, MLO5, MLO9) of this family are known to be expressed in pollen and pollen tube, and MLO5 is upregulated after pollen tubes passed through the pistil indicating MLO members may be involved in pollen tube growth (Qin et al., 2009; Meng et al., 2020). Previous studies have shown that MLOs are involved in communication pathways between cells and their external environment (Buschges et al., 1997; Kessler et al., 2010). Given that the sperm are passively transported within the pollen tube, MLO6 is more likely to function redundantly with other MLO members in orchestrating the pollen

tube growth. Whether MLO6 may have any sperm cell specific function such as maturation or differentiation during pollen tube growth remains to be seen. So far, none of the studies could pinpoint the exact function of these stress related genes in male gametes, and would therefore be an ideal candidate for further functional characterization.

We also found that downregulated genes in sperm cells from growing pollen tubes were mainly associated with pollen tube growth and cell-cell signalling, functions mostly associated with the vegetative cell (Figure 3.11). This is expected, as it is already known that sperm cells are passively transported within the female tissues and are not involved in pollen tube growth (Zhang et al., 2017). Therefore, many of the genes downregulated during pollen tube growth might be shut down as they are not necessary anymore and the function is performed mainly by the vegetative nucleus.

The transcriptome of vegetative nuclei showed more changes during pollen tube growth compared to sperm cells and this is in line with their expected role of driving pollen tube growth and of responding to the female cues (Figure 3.5A). Technical challenges made it impossible to collect enough material to prepare cDNA libraries from vegetative nuclei from semi *in vivo* grown pollen tubes. Therefore, we decided to use whole semi *in vivo* pollen tubes as a proxy for expression in vegetative nuclei. It is important here to mention that, although we carefully defined the criteria to exclude very high sperm expressed transcripts (~140) that may contribute to the transcriptome of whole semi *in vivo* grown pollen tubes, we cannot decipher the precise expression of vegetative nuclei from SIVPT stage as there might be genes expressed at a similar level in both sperm cells and vegetative nucleus.

Nevertheless, our transcriptome data from vegetative nuclei helped to identify a larger number of genes upregulated during pollen tube growth than known previously (Qin et al., 2009). For instance, we identified 3210 genes to be up regulated in PT vs VN MPG comparison (Figure 3.5C), while Qin et al., 2009

found 1222 genes to be upregulated in SIVPT vs dry pollen comparison. While the direct comparison of PT vs VN MPG may not be without problems, the fact that we saw almost 1839 genes upregulated in VN in vitro vs VN MPG comparison as against only 186 genes to be up regulated in 4h pollen tube vs dry pollen (Qin et al., 2009), clearly demonstrates the depth and resolution at which we were able to identify the transcriptomic changes in vegetative nuclei during pollen tube growth. Therefore, to get a clearer picture of changes during SIVPT stage, efforts directed towards optimizing and preparing RNA-seq libraries from VN SIVPT would be advantageous. However, overall our transcriptome data from vegetative nuclei corroborate the previous analysis on pollen tubes at different time point (Qin et al., 2009). The differences in the number of genes that were observed between our study and the previous study could probably be attributed to the use of different transcriptome analysis platforms: microarrays vs RNA-seq.

Three-way comparison of the vegetative nuclei from three pollen conditions (MPG, In vitro and PT) revealed similar sets of genes to be upregulated during pollen tube growth, particularly during VN In vitro vs VN MPG and PT vs VN MPG comparison. However, we observed modest differences in the transcriptome of vegetative nuclei between pollen tubes grown *in vitro* vs pollen tubes that have passed through a pistil. The significantly higher number of genes (~2306) that we found to be differentially expressed in our analysis between PT vs VN In vitro comparison (Figure 3.5D) underscores the role of female cues in pollen tube guidance, and how ovular signals leads to dramatic changes in the transcriptome during SIVPT stage.

If we compare the functional enrichment of genes that are upregulated, we found that genes associated with pollen tube growth, vesicle mediated transported, and DNA replication were found to be upregulated during pollen tube growth and the trend was quite similar whether pollen tubes were grown *in vitro* or through the pistil (Figure S3.1). However, the direct comparison of

the PT vs VN in vitro resulted in the significant enrichment of genes involved in cell cycle regulation, DNA repair and DNA replication (Figure 3.12). The vegetative nuclei is known to exit the cell cycle, and therefore changes in the cell cycle and DNA replication genes might be associated with sperm cells which are under cell cycle control (Friedman, 1999; Liu et al., 2021). Surprisingly, we did not find a significant enrichment of cell cycle related genes in our sperm cell data, though some cell cycle related genes were upregulated in SC SIVPT. Whether the vegetative nucleus is still under cell cycle regulation needs to be investigated further.

Our data revealed upregulation of several transmembrane receptors (TIR-NBS-LRR type) during semi *in vivo* pollen tube growth. Out of 90 annotated TIR-NBS-LRR type proteins in Arabidopsis, eight of them were upregulated in PT vs VN MPG comparison. These proteins belong to a large gene family and are known to be involved in the molecular recognition of effector proteins derived from pathogens. The exact functions of many of these proteins remain unknown, however, enrichment of these genes in our data and also in the previous study indicate their diverse function in processes other than plant defense.

Our analysis identified the expression of several transcription factors (e.g. MYB120 and MYB109), homeobox-containing proteins (e.g. HDG3) that were upregulated in PT vs VN MPG comparison. These results are in line with a previous study in which several transcription factors were found to be upregulated during pollen tube growth as they passed through the pistil (Qin et al., 2009). For instance, transcription factors MYB120 (log2 FC > 1.8) and MYB97 were upregulated (log2 FC ~1) in our PT vs VN MPG comparison. These two closely related MYB transcription factors together with MYB101 were found to be involved in pollen tube burst and release of sperm cells (Leydon et al., 2013). These results suggest that the upregulation of transcription factors during pollen tube growth may be a part of the

transcriptional network regulating the mechanism of pollen tube guidance, the function of many of which remains to be determined (Qin et al., 2009).

To conclude, our data on sperm cell and vegetative nuclei from three conditions of pollen development revealed transcriptomic changes at a resolution not achieved so far. First, the genes identified in sperm cells during pollen tube growth could be potential candidate for understanding their role in germline specification and whether they may have any role in plant fertility. Secondly, the genes identified in vegetative nuclei could be characterized for their role in pollen tube growth and integrity during sperm delivery.

3.6. Authors Contributions

Chandra Shekhar Misra and Jörg D. Becker (Head of “Plant Reproduction and Evolution Group” at Instituto de Tecnologia Química e Biológica António Xavier), conceptualized the project. Chandra Shekhar Misra optimized the protocol for collection of sperm cells and vegetative nuclei from *in vitro* and semi *in vivo* grown pollen tubes. Chandra Shekhar Misra prepared the cDNA libraries and analysed the RNA-seq data from the three pollen conditions. António Sousa (Bioinformatician at IGC) prepared the plots in R with inputs from Chandra Shekhar and Jörg D. Becker. This work was supervised by Jörg D. Becker.

3.7. Acknowledgment

I would like to thank Vera Nunes, technician at Plant Facility, IGC, for plant growth and maintenance. I would also like to acknowledge the help of other facilities at IGC, namely Microscopy Facility, Flow Cytometry Unit and Genomics Facility.

3.8. References

Abane, R. and Mezger, V., 2010. Roles of heat shock factors in gametogenesis and development. *The FEBS journal*, *277*(20), pp.4150-4172.

Boavida, L.C. and McCormick, S., 2007. TECHNICAL ADVANCE: Temperature as a determinant factor for increased and reproducible in vitro pollen germination in Arabidopsis thaliana. *The Plant Journal*, *52*(3), pp.570-582.

Boavida, L.C., Borges, F., Becker, J.D. and Feijó, J.A., 2011. Whole genome analysis of gene expression reveals coordinated activation of signaling and metabolic pathways during pollen-pistil interactions in Arabidopsis. *Plant Physiology*, *155*(4), pp.2066-2080.

Borg, M., Brownfield, L., Khatab, H., Sidorova, A., Lingaya, M. and Twell, D., 2011. The R2R3 MYB transcription factor DUO1 activates a male germline-specific regulon essential for sperm cell differentiation in *Arabidopsis*. *The Plant Cell*, *23*(2), pp.534-549.

Borges, F., Gardner, R., Lopes, T., Calarco, J.P., Boavida, L.C., Slotkin, R.K., Martienssen, R.A. and Becker, J.D., 2012. FACS-based purification of Arabidopsis microspores, sperm cells and vegetative nuclei. *Plant Methods*, *8*(1), pp.1-9.

Borges, F., Gomes, G., Gardner, R., Moreno, N., McCormick, S., Feijó, J.A. and Becker, J.D., 2008. Comparative transcriptomics of *Arabidopsis* sperm cells. *Plant Physiology*, *148*(2), pp.1168-1181.

Büschges, R., Hollricher, K., Panstruga, R., Simons, G., Wolter, M., Frijters, A., Van Daelen, R., van der Lee, T., Diergaarde, P., Groenendijk, J. and

Töpsch, S., 1997. The barley Mlo gene: a novel control element of plant pathogen resistance. *Cell*, 88(5), pp.695-705.

Chaturvedi, P., Ghatak, A. and Weckwerth, W., 2016. Pollen proteomics: from stress physiology to developmental priming. *Plant Reproduction*, *29*(1), pp.119-132.

Chaturvedi, P., Ischebeck, T., Egelhofer, V., Lichtscheidl, I. and Weckwerth, W., 2013. Cell-specific analysis of the tomato pollen proteome from pollen mother cell to mature pollen provides evidence for developmental priming. *Journal of Proteome Research*, *12*(11), pp.4892-4903.

Conze, L.L., Berlin, S., Le Bail, A. and Kost, B., 2017. Transcriptome profiling of tobacco (Nicotiana tabacum) pollen and pollen tubes. *BMC Genomics*, *18*(1), pp.1-11.

Dun, M.D., Aitken, R.J. and Nixon, B., 2012. The role of molecular chaperones in spermatogenesis and the post-testicular maturation of mammalian spermatozoa. *Human Reproduction Update*, *18*(4), pp.420-435.

Durbarry, A., Vizir, I. and Twell, D., 2005. Male germ line development in *Arabidopsis*. duo pollen mutants reveal gametophytic regulators of generative cell cycle progression. *Plant Physiology*, *137*(1), pp.297-307.

Elleman, C.J. and Dickinson, H.G., 1996. Identification of pollen components regulating pollination-specific responses in the stigmatic papillae of Brassica oleracea. *New Phytologist*, *133*(2), pp.197-205.

Friedman, W.E., 1999. Expression of the cell cycle in sperm of Arabidopsis: implications for understanding patterns of gametogenesis and fertilization in plants and other eukaryotes. *Development*, *126*(5), pp.1065-1075.

Ge, Z., Bergonci, T., Zhao, Y., Zou, Y., Du, S., Liu, M.C., Luo, X., Ruan, H., García-Valencia, L.E., Zhong, S. and Hou, S., 2017. Arabidopsis pollen tube

integrity and sperm release are regulated by RALF-mediated signaling. *Science*, *358*(6370), pp.1596-1600.

Hafidh, S., Breznenová, K., Růžička, P., Feciková, J., Čapková, V. and Honys, D., 2012. Comprehensive analysis of tobacco pollen transcriptome unveils common pathways in polar cell expansion and underlying heterochronic shift during spermatogenesis. *BMC Plant Biology*, *12*(1), pp.1-20.

Higashiyama, T., Kuroiwa, H., Kawano, S. and Kuroiwa, T., 1998. Guidance in vitro of the pollen tube to the naked embryo sac of Torenia fournieri. *The Plant Cell*, *10*(12), pp.2019-2031.

Ischebeck, T., Valledor, L., Lyon, D., Gingl, S., Nagler, M., Meijón, M., Egelhofer, V. and Weckwerth, W., 2014. Comprehensive cell-specific protein analysis in early and late pollen development from diploid microsporocytes to pollen tube growth. *Molecular & Cellular Proteomics*, *13*(1), pp.295-310.

Johnson, M.A., Harper, J.F. and Palanivelu, R., 2019. A fruitful journey: pollen tube navigation from germination to fertilization. *Annual Review of Plant Biology*, *70*, pp.809-837.

Julca, I., Ferrari, C., Flores-Tornero, M., Proost, S., Lindner, A., Hackenberg, D., Steinbachová, L., Michaelidis, C., Pereira, S.G., Misra, C.S., Kawashima, T., Borg, M., Berger, F., Goldberg, J., Johnson, M., Honys, D., Twell, D., Sprunck, S., Dresselhaus, T., Becker, J.D., Mutwil, M., 2020. Comparative transcriptomic analysis reveals conserved transcriptional programs underpinning organogenesis and reproduction in land plants. *Nature Plants*, *7(8)*, pp. 1143-1159.

Kessler, S.A., Shimosato-Asano, H., Keinath, N.F., Wuest, S.E., Ingram, G., Panstruga, R. and Grossniklaus, U., 2010. Conserved molecular components

for pollen tube reception and fungal invasion. *Science*, 330(6006), pp.968-971.

Kim, D., Paggi, J.M., Park, C., Bennett, C. and Salzberg, S.L., 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, *37*(8), pp.907-915.

Leydon, A.R., Beale, K.M., Woroniecka, K., Castner, E., Chen, J., Horgan, C., Palanivelu, R. and Johnson, M.A., 2013. Three MYB transcription factors control pollen tube differentiation required for sperm release. *Current Biology*, *23*(13), pp.1209-1214.

Leydon, A.R., Weinreb, C., Venable, E., Reinders, A., Ward, J.M. and Johnson, M.A., 2017. The molecular dialog between flowering plant reproductive partners defined by SNP-informed RNA-sequencing. *The Plant Cell*, *29*(5), pp.984-1006.

Li, H.J., Meng, J.G. and Yang, W.C., 2018. Multilayered signaling pathways for pollen tube growth and guidance. *Plant Reproduction*, *31*(1), pp.31-41.

Liao, Y., Smyth, G.K. and Shi, W., 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), pp.923-930.

Liu, L., Lu, Y., Wei, L., Yu, H., Cao, Y., Li, Y., Yang, N., Song, Y., Liang, C. and Wang, T., 2018. Transcriptomics analyses reveal the molecular roadmap and long non-coding RNA landscape of sperm cell lineage development. *The Plant Journal*, *96*(2), pp.421-437.

Liu, X.Q., Shi, J.J., Fan, H., Jiao, J., Gao, L., Tan, L., Nagawa, S. and Wang, D.Y., 2021. Nuclear DNA replicates during zygote development in Arabidopsis and Torenia fournieri. *Plant Physiology*, *185*(1), pp.137-145.

Love, M.I., Huber, W. and Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), pp.1-21.

Macaulay, I.C., Teng, M.J., Haerty, W., Kumar, P., Ponting, C.P. and Voet, T., 2016. Separation and parallel sequencing of the genomes and transcriptomes of single cells using G&T-seq. *Nature Protocols*, *11*(11), pp.2081-2103.

Mecchia, M.A., Santos-Fernandez, G., Duss, N.N., Somoza, S.C., Boisson-Dernier, A., Gagliardini, V., Martínez-Bernardini, A., Fabrice, T.N., Ringli, C., Muschietti, J.P. and Grossniklaus, U., 2017. RALF4/19 peptides interact with LRX proteins to control pollen tube growth in Arabidopsis. *Science*, *358*(6370), pp.1600-1603.

Meng, J.G., Liang, L., Jia, P.F., Wang, Y.C., Li, H.J. and Yang, W.C., 2020. Integration of ovular signals and exocytosis of a Ca 2+ channel by MLOs in pollen tube guidance. *Nature Plants*, 6(2), pp.143-153.

Misra, C.S., Santos, M.R., Rafael-Fernandes, M., Martins, N.P., Monteiro, M. and Becker, J.D., 2019. Transcriptomics of Arabidopsis sperm cells at single-cell resolution. *Plant Reproduction*, *32*(1), pp.29-38.

Mondragón-Palomino, M., John-Arputharaj, A., Pallmann, M. and Dresselhaus, T., 2017. Similarities between reproductive and immune pistil transcriptomes of Arabidopsis species. *Plant Physiology*, *174*(3), pp.1559-1575.

Okada, T., Bhalla, P.L. and Singh, M.B., 2006. Expressed sequence tag analysis of Lilium longiflorum generative cells. *Plant and Cell Physiology*, *47*(6), pp.698-705.

Okada, T., Singh, M.B. and Bhalla, P.L., 2007. Transcriptome profiling of Lilium longiflorum generative cells by cDNA microarray. *Plant Cell Reports*, *26*(7), pp.1045-1052.

Osaka, M., Nabemoto, M., Maeda, S., Sakazono, S., Masuko-Suzuki, H., Ito, K., Takada, Y., Kobayashi, I., Lim, Y.P., Nakazono, M. and Fujii, S., 2019. Genetic and tissue-specific RNA-sequencing analysis of self-compatible mutant TSC28 in Brassica rapa L. toward identification of a novel self-incompatibility factor. *Genes & Genetic Systems*, *94*(4), pp.167-176.

Palanivelu, R. and Johnson, M.A., 2010. Functional genomics of pollen tube–pistil interactions in Arabidopsis. *Biochemical Society Transactions*, *38*(2), pp.593-597.

Palanivelu, R. and Preuss, D., 2006. Distinct short-range ovule signals attract or repel Arabidopsis thaliana pollen tubes in vitro. *BMC Plant Biology*, *6*(1), pp.1-9.

Palanivelu, R. and Tsukamoto, T., 2012. Pathfinding in angiosperm reproduction: pollen tube guidance by pistils ensures successful double fertilization. *Wiley Interdisciplinary Reviews: Developmental Biology*, *1*(1), pp.96-113.

Qin, Y., Leydon, A.R., Manziello, A., Pandey, R., Mount, D., Denic, S., Vasic, B., Johnson, M.A. and Palanivelu, R., 2009. Penetration of the stigma and style elicits a novel transcriptome in pollen tubes, pointing to genes critical for growth in a pistil. *PLoS Genetics*, *5*(8), p.e1000621.

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H. and Vilo, J., 2019. g: Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research*, *47*(W1), pp.W191-W198.

Rotman, N., Durbarry, A., Wardle, A., Yang, W.C., Chaboud, A., Faure, J.E., Berger, F. and Twell, D., 2005. A novel class of MYB factors controls sperm-cell formation in plants. *Current Biology*, *15*(3), pp.244-248.

Russell, S.D., Gou, X., Wong, C.E., Wang, X., Yuan, T., Wei, X., Bhalla, P.L. and Singh, M.B., 2012. Genomic profiling of rice sperm cell transcripts reveals conserved and distinct elements in the flowering plant male germ lineage. *New Phytologist*, *195*(3), pp.560-573.

Rutley, N. and Twell, D., 2015. A decade of pollen transcriptomics. *Plant Reproduction*, *28*(2), pp.73-89.

Samakovli, D., Tichá, T., Vavrdová, T., Závorková, N., Pecinka, A., Ovečka, M. and Šamaj, J., 2021. HEAT SHOCK PROTEIN 90 proteins and YODA regulate main body axis formation during early embryogenesis. *Plant Physiology, 186,* pp. 1526–1544

Santos, M.R., Bispo, C. and Becker, J.D., 2017. Isolation of Arabidopsis pollen, sperm cells, and vegetative nuclei by fluorescence-activated cell sorting (FACS). In *Plant germline development* (pp. 193-210). Humana Press, New York, NY.

Tichá, T., Samakovli, D., Kuchařová, A., Vavrdová, T. and Šamaj, J., 2020. Multifaceted roles of HEAT SHOCK PROTEIN 90 molecular chaperones in plant development. *Journal of Experimental Botany*, *71(14),* pp.3966-3985.

Tung, C.W., Dwyer, K.G., Nasrallah, M.E. and Nasrallah, J.B., 2005. Genome-wide identification of genes expressed in Arabidopsis pistils specifically along the path of pollen tube growth. *Plant Physiology*, *138*(2), pp.977-989.

Wang, Y., Zhang, W.Z., Song, L.F., Zou, J.J., Su, Z. and Wu, W.H., 2008. Transcriptome analyses show changes in gene expression to accompany

pollen germination and tube growth in Arabidopsis. *Plant Physiology*, *148*(3), pp.1201-1211.

Xin, H.P., Peng, X.B., Ning, J., Yan, T.T., Ma, L.G. and Sun, M.X., 2011. Expressed sequence-tag analysis of tobacco sperm cells reveals a unique transcriptional profile and selective persistence of paternal transcripts after fertilization. *Sexual Plant Reproduction*, *24*(1), pp.37-46.

Zhang, J., Huang, Q., Zhong, S., Bleckmann, A., Huang, J., Guo, X., Lin, Q., Gu, H., Dong, J., Dresselhaus, T. and Qu, L.J., 2017. Sperm cells are passive cargo of the pollen tube in plant fertilization. *Nature Plants*, *3*(6), pp.1-5.

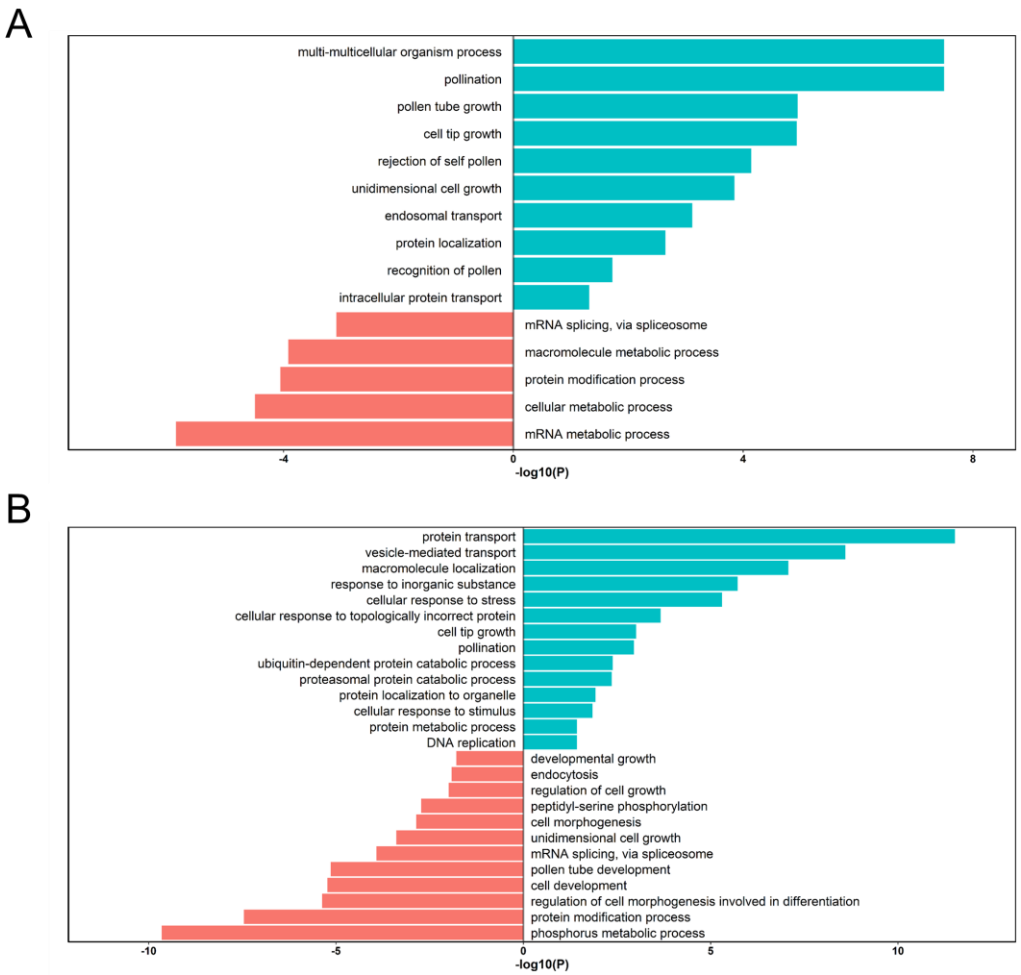


Figure S3.1: Functional enrichment of genes differentially expressed in vegetative nuclei during pollen tube growth. A) GO analysis of genes upregulated or downregulated in comparison of VN In vitro vs VN MPG. **B)** GO analysis of genes upregulated or downregulated in comparison of PT vs VN MPG. Blue bar: Upregulated, Red Bar: Downregulated. For detailed GO, see Supplemental Dataset 3.10-3.11.

Chapter 4

Transcriptomics of Arabidopsis sperm cells at single-cell resolution

This chapter was adapted and updated from:

Misra, C.S., Santos, M.R., Rafael-Fernandes, M., Martins, N.P., Monteiro, M. and Becker, J.D., 2019. Transcriptomics of Arabidopsis sperm cells at single-cell resolution. *Plant Reproduction*, *32*(1), pp.29-38.

4.1. Abstract

Male gametophyte development in flowering plants begins with a microspore mother cell, which upon two consecutive cell divisions forms a mature pollen grain containing a vegetative nucleus and two sperm cells. Pollen development is a highly dynamic process, involving changes both at the transcriptome and epigenome level of vegetative nuclei and the pair of sperm cells that have their own cytoplasm and nucleus. While the overall transcriptome of Arabidopsis pollen development is well documented, studies at single cell level, in particular of sperm cells, are still lacking. Such studies would be essential to understand whether and how the two sperm cells are transcriptionally different, in particular once the pollen tube grows through the transmitting tissue of the pistil. Here we describe a detailed protocol for isolation of single sperm cells from pollen tubes grown through pistils and analysis of their transcriptome. The analyses of our 80 single cell data revealed an average expression of at least 1900 genes per sperm cell. Our preliminary data provides some insights into differences that might exist between the two transcriptionally different clusters of sperm cells, but more detailed analyses with much larger numbers of cells are needed to confirm these results. Nevertheless, our data provides evidence for the potential use of single cell genomics for plant reproductive biology.

4.2. Introduction

The male germline in flowering plants is formed when post-meiotic haploid microspores undergo an asymmetric division, producing a larger vegetative cell that exits the cell cycle, and a smaller generative cell (or male germ cell) that undergoes a second division to produce twin sperm cells (reviewed in Berger and Twell, 2011). The result is a unique “cell within a cell system”, combining two independent cell lineages with different fates (Twell 2011). While the vegetative cell controls pollen tube growth, the two sperm cells delivered by the pollen tube to the embryo sac are required for double fertilization.

Over the last decade, a wealth of transcriptome and methylome data from *Arabidopsis thaliana* pollen (Becker et al., 2003; Hony and Twell, 2004; Pina et al., 2005) and sperm cells (Borges et al., 2008; Slotkin et al., 2009; Calarco et al., 2012; Ibarra et al., 2012) have improved our understanding of this unique system. These studies are complemented by transcriptome analyses of sperm cell samples from other angiosperms, e.g. from maize (Engel et al., 2003), rice (Anderson et al., 2013), and tomato (Liu et al., 2018). Importantly, studies in Arabidopsis have shown that pollen tubes undergo transcriptomic changes as they travel through the transmitting tissue of the pistil (Qin et al., 2009; Lin et al., 2014; Leydon et al., 2017), but to what extent the sperm cells contribute to these changes remains unknown. Though sperm cells are passively transported in the growing pollen tube (Zhang et al., 2017), changes in their transcriptome are to be expected, since sperm cells are believed to progress from S phase of the cell cycle at anthesis to G2 just before fertilization (Friedman 1999).

Sperm cells in most flowering plants are isomorphic. However, a few plant species have dimorphic sperm cells caused by an asymmetrical cell division of the generative cell (Mogensen, 1992; Chen et al., 1995; Saito et al., 2002). Probably the best-studied example is *Plumbago zeylanica*, where the larger

sperm cell containing more mitochondria preferentially fuses with the central cell, and the smaller plastid-rich sperm cell fuses preferentially with the egg cell (Russell, 1985). Differentially expressed genes between dimorphic sperm cells in Plumbago have been identified, but lacked clear homologs in Arabidopsis (Gou et al., 2009).

No study so far has addressed the question whether the isomorphic sperm cells in Arabidopsis differ at gene expression level. Such a study was not possible until now because of the technical challenges involved in collecting sperm cells to study their transcriptomes at single cell level. But with the combination of sperm cell FACS (Borges et al., 2012; Santos et al., 2017) and scRNA-seq (Angermueller et al., 2016; Macaulay et al., 2016), it is now feasible to isolate single sperm cells and study the transcriptome profile of each sperm cell individually.

Single-cell analyses have been extensively performed using animal models to understand various aspects of biological processes, including gene expression dynamics and changes in the epigenetic landscape (Gawad et al., 2016; Macaulay et al., 2017). However, plant single cell studies have been challenging due to difficulties in obtaining single cells from plants. For the isolation of single cells from complex tissues, cell walls have to be removed by maceration or enzymatic digestion to obtain protoplasts (Efroni and Birnbaum, 2016; Libault, et al., 2017). Nevertheless, over the years several plant single cell types have emerged as models, including root hairs (Libault et al., 2010; Qiao and Libault, 2013), trichomes (Marks et al., 2008; Yang and Ye, 2013), stomatal guard cells (Jin et al., 2013) and germline cells (pollen and egg cells) (Reviewed in Rutley and Twell, 2015). Plant single cell type isolation by shearing (e.g. trichomes, cotton fibres or root hairs) or other techniques without the need to prepare protoplasts (e.g. FACS and percoll gradients for pollen developmental stages), have accelerated the analysis of such plant cell types (Becker et al., 2014; Dupl'áková, et al., 2016).

Here we show that using an optimised system for semi *in vivo* pollen tube growth coupled to FACS it is possible to collect individual sperm cells from growing pollen tubes. We confirmed the purity of the sorted sperm cells, and were able to successfully analyse the transcriptome of single sperm cells by using an optimized scRNA-Seq protocol. Our single cell transcriptome analysis of 80 sperm cells helped to identify 208 genes to be differentially expressed between two transcriptionally different clusters. However, due to lack of statistically significant data, more detailed analyses with much larger numbers of cells are needed to confirm these preliminary results.

4.3. Material and Methods

4.3.1. Plant material and growth conditions

Arabidopsis thaliana wild type (Columbia-0) and a transgenic marker line harboring *MGH3p::MGH3-eGFP* and *ACT11p::H2B-mRFP* (Borges et al., 2012) were used in this study. Plants were sown on soil, grown for 8 weeks in short-day conditions (8 h light at 21°C-23°C) and then transferred to long-day conditions (16 h light) to induce flowering.

4.3.2. Semi *in vivo* pollen tube growth and sperm cell release

For semi *in vivo* pollen tube growth, the marker line was used to pollinate WT flower buds, emasculated 16 hours prior to pollination. After 2 hours, the pollinated pistil was excised and placed on double sided tape. The excised pistil was then cut at the junction of style and ovary and placed gently on liquid Arabidopsis pollen germination medium, avoiding the submergence of the papillae (stigma) with pollen grains. The pistil was incubated for an additional 3-4 hours for the pollen tubes to emerge from the cut end of the style (Boavida and McCormick, 2007). The pollen tubes were burst using osmotic shock (8% w/v mannitol). A summary of the procedure is illustrated in Figure 4.1.

After the pollen tubes were burst, the suspension was immediately transferred to ice cold sperm extraction buffer: 1.3 mM H_3BO_3, 3.6 mM $CaCl_2.2H_2O$, 0.74 mM KH_2PO_4, 438 mM Sucrose, 7 mM MOPS, 0.83 mM $MgSO_4.7H_2O$, pH adjusted to 6.0. (Santos et al., 2017). In order to collect only intact cells by FACS, Sytox orange (25nM final concentration) was added to the cell suspension and incubated for 10 minutes on ice (for details see Supplemental Data File S1).

4.3.3. Fluorescence activated cell sorting

Fluorescent activated cell sorting was carried out using a MoFlo Legacy (Beckman Coulter, Fort Collins, USA) with settings as described in Santos et al., 2017 (for details see Supplemental Data File S1). MoFlo settings for forward and side scatter were defined for GFP labelled sperm cells isolated from mature pollen grains. This gating strategy was used as a template for sperm cells obtained after bursting of semi *in vivo* grown pollen tubes. Single cell suspension (sperm extraction buffer containing sperm cells, vegetative nuclei and debris) was then loaded onto the cell sorter and sorting was done using a 100 µm nozzle with sort mode set to single cell. Single cells were sorted into PCR strip tubes or 0.2 ml Eppendorf, prefilled with 2.5 µL of freshly prepared RLT plus lysis buffer supplemented with 1% β-mercaptoethanol. Doublets and dead/compromised cells were excluded based on forward scatter, pulse width and Sytox Orange fluorescence. Sorted single cell lysates were vortexed, spun down and transferred into liquid nitrogen before being stored at -80°C. Tubes with 50 sperm cells or tubes with sorted single beads (Spherotech) were used as positive and negative controls, respectively.

4.3.4. Single-cell RNA-sequencing: reverse transcription

A detailed list of reagents and a step by step protocol is provided in Supplemental Data File S4.1. In brief: single sorted sperm cells were subjected to Smart-seq2 as described earlier (Macaulay et al., 2016), with

some modifications. A modified bead capture protocol was used, in which Dyna Beads MyOne Streptavidin C1 (Invitrogen) were washed and attached to oligo-DT primers (5'AAGCAGTGGTATCAACGCAGAGTACTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN-3', Biomer). 10 µl of conjugated beads were then mixed with thawed single-cell lysates and incubated at room temperature for 20 minutes with constant mixing at 2000 rpm. The tubes were then moved onto a ring magnet (Alpaqua, USA) for 1 minute. Supernatant containing gDNA was discarded. The beads were washed twice with gDNA buffer, with care being taken to avoid loss of beads during washes.

Beads in each tube were then resuspended with 5µl of RT mix containing: 0.5 µl of 10 mM dNTP (Thermo Fisher Scientific), 1.14 µl of nuclease free water , 0.06 µl RNAse Inhibitor (40 U/µl, Takara), 1 µl of 5x Superscript II first-strand buffer (Thermo Fisher Scientific), 0.25 µl of 100 mM DTT (Invitrogen) ,1µl of 5M Betaine (Sigma-Aldrich), 0.3 µl of 100mM $MgCl_2$ (Sigma-Aldrich), 0.5µl of 10µM TSO (5'-AAGCAGTGGTATCAACGCAGAGTACrGrG+G-3', Exiqon), and 0.25 µl of Superscript II reverse transcriptase (200 U/µl, Thermo Fisher Scientific).

Each tube with RT mix was mixed well with the beads for 1 min at 2000 rpm. Reverse transcription was carried out by incubating the tubes at 42^{o}C for 60 min (template switching) followed by 50^{o}C for 30 min (template switching and RT) and finally enzyme inactivation at 60^{o}C for 10 min.

4.3.5. Single-cell RNA-sequencing: PCR pre-amplification and cDNA purification

7.5 µl of PCR mix containing 6.25 µL of 2x KAPA HiFi Hot Start 6 Ready Mix (KAPA Biosystems), 0.125 µL of 10µM ISPCR primer (5'-AAGCAGTGGTATCAACGCAGAGT-3', IDT) and 1.125 µL nuclease-free

water, was added to each PCR tube for a final PCR reaction volume of 12.5 µL.

The reaction was carried out with an initial incubation at 98°C for 3 minutes followed by amplification using 21 cycles (98°C for 20 sec, 67°C for 15 sec, 72°C for 6 min) and final extension at 72ºC for 5 minutes. PCR products were then cleaned by mixing them with an equal volume of Agencourt AMpureXP SPRI beads (Beckman Coulter) followed by incubation at room temperature for 8 minutes. The tubes were placed onto a magnet for 5 minutes before removing the supernatant. Beads were washed twice with 100 µl of 80% freshly prepared ethanol, with care being taken to avoid loss of beads during the washes. All traces of residual ethanol were removed and beads were left to dry at room temperature for 5 minutes. The beads were then resuspended in 17.5 µl of molecular biology grade water and incubated at room temperature for 5 minutes. The tubes were placed on the magnet for 2 minutes prior to transferring the supernatant containing the amplified cDNA to new Eppendorf tubes. The concentration of amplified cDNA was measured using a Qubit dsDNA high sensitivity Assay kit (Thermo Fisher Scientific). Some single cell cDNA libraries were then randomly selected to confirm size distribution using an NGS high sensitivity kit on a Fragment Analyser (AATI).

4.3.6. RT-PCR analysis of sperm-specific transcripts

Quality and purity of single cell RNA-seq libraries prepared by Smart-seq2 were verified by PCR using primers for the sperm specific genes: *MGH3* (AT1G19890) and *DAZ3* (AT4G35700), reported to be highly expressed in sperm cells (Borges et al., 2008). *VEX1* (AT5G62850) specific primers were used to control for contamination with vegetative nuclei (Borges et al., 2012). For this, the beads used in the previous step were incubated for a second time with 5 µl of molecular biology grade water and the eluate used as a template in PCR reactions of 35 cycles. See table S1 for list of primers.

4.3.7. Single-cell RNA-sequencing: library preparation

Based on the results of RT-PCR, only those samples which showed amplification for sperm specific genes were further processed for library preparation and sequencing. Libraries were prepared using a low-volume Nextera transposase-based protocol described in Baym et al., 2015. 2nM of each library was pooled and sequenced on an in-house Illumina Nextseq 500 instrument, using 75 bp single end reads.

4.3.8. RNA-seq data processing

RNA sequencing data from single sperm cells and 50 sperm cells were processed to check for quality of reads and alignment. For each RNA seq data set, raw RNA-seq reads were checked for quality metrics with fastqc, and polyAs and adaptors were trimmed with cutadapt (Martin, 2011) and trimmomatic (Bolger et al., 2014). The reads were then aligned to the Arabidopsis reference genome with recent annotation using Hisat 2 (Kim et al., 2015). HT-seq count (Ander et al., 2015) was then used to count gene features using recent annotation from Araport 11 (Cheng et al., 2017) and the expression matrix of raw counts was generated to be used for further analysis (Ander et al., 2015). Transcript quantification was achieved using Salmon (Patro et al., 2017). A transcript was considered to be expressed if transcript per million (TPM) was more than 1.

4.3.9. Quality control for scRNA-seq data

In order to remove poor quality cells, the following criteria were applied: mapping rate of more than 60%, cells expressing more than 3000 genes. At the gene level, only those genes were considered for downstream analysis which had one read in at least 20 cells. After this filtering steps, 70 out of 80 cells were retained and a total of 2336 genes used for further analysis.

4.3.10. Clustering of scRNA-seq data

The Seurat v.2.3.3 R package (Butler et al., 2018) was used to perform quality-control, normalization with 'NormalizeData()' using the method "LogNormalize" and a scaling factor of 1e6; find variable features ('FindVariableGenes()'), and scaling ('ScaleData()'). Then, clustering was performed by creating a Pearson correlation matrix with the 'cor()' function and performing hierarchical clustering analysis with 'hclust()' which uses the complete linkage method by default, and subsequently passed to the function 'cutreeDynamic()' to prune and find the clusters from the package dynamicTreeCut v.1.63-1 (Langfelder et al., 2008). Finally, positive/upregulated markers ('only.pos = TRUE') for three clusters were determined using the Seurat function 'FindAllMarkers()', which compares each cluster against the remaining ones using the Wilcoxon Rank Sum test, and pairwise comparisons between specific clusters (1 vs 2) was determined with 'FindMarkers()' function using the t-test.

4.4. Results

4.4.1. Semi *in vivo* pollen tube growth in liquid medium

For FACS based collection of single sperm cells, pollen tubes were first grown through pistils. Earlier studies using such semi *in vivo* assays (Qin et al., 2009) were mostly based on agar solidified medium. Here however the use of liquid pollen germination medium was important to allow the efficient collection of sperm cells for FACS. This optimised system for semi *in vivo* pollen tube (SIVPT) growth differs from what was originally described (Palanivelu and Preuss 2006; Qin et al., 2009): First, we pollinated the pistil *in vivo* and let the pollen tubes enter the style for about two hours (Figure 4.1). After that, the pistil was excised and placed on pollen germination medium for 4 hours for the pollen tubes to grow and emerge from the cut end of the style.

These modifications were incorporated to capture the transcriptome of sperm cells from SIVPT under the most natural conditions possible.

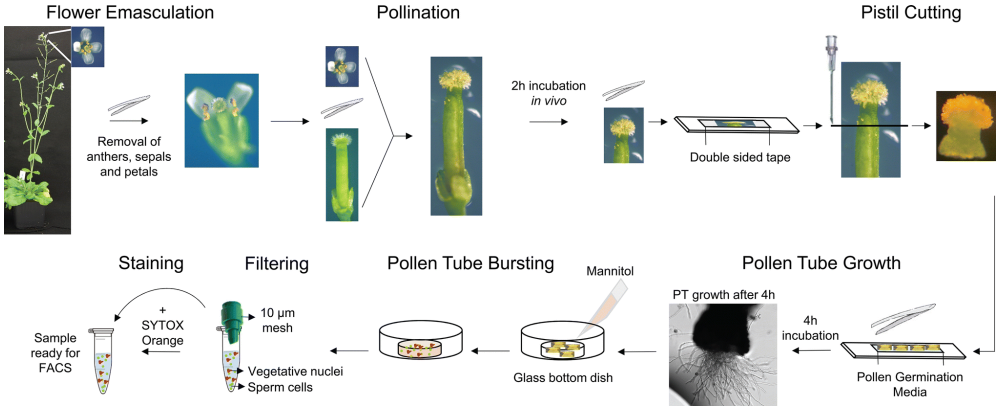


Figure 4.1: Flow diagram of semi *in vivo* pollen tube growth assay and isolation of sperm cells.

The emerging pollen tubes grew efficiently into the liquid pollen germination medium after 4 hours, as shown in Figure 4.2A. Moreover, our method of SIVPT growth on liquid germination medium allowed for the bursting of pollen tubes by osmotic shock and efficient release of intact sperm cells and vegetative nuclei as shown in Figure 4.2B. The solution with released sperm cells and vegetative nuclei (occasionally still as male germ unit as seen in Figure 4.2B) was collected and subjected to FACS.

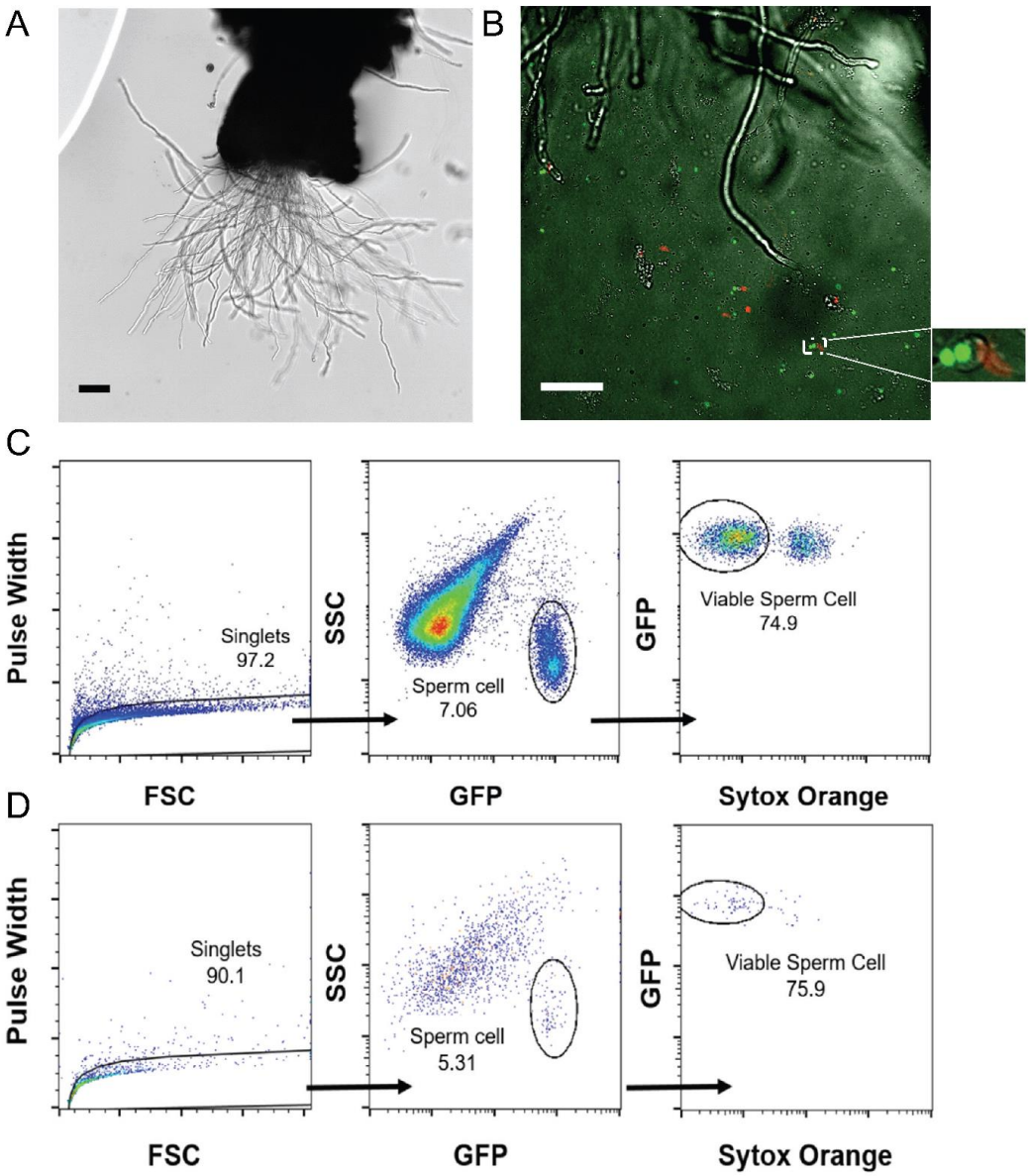

A
B
C
Pulse Width
Singlets
97.2
FSC
SSC
Sperm cell
7.06
GFP
GFP
Viable Sperm Cell
74.9
Sytox Orange
D
Pulse Width
Singlets
90.1
FSC
SSC
Sperm cell
5.31
GFP
GFP
Viable Sperm Cell
75.9
Sytox Orange

Figure 4.2: Isolation of sperm cells from semi *in vivo* grown pollen tubes. A) cut pistil with pollen tubes emerging. Robust pollen tube growth was seen after 4 hrs. **B)** GFP (green) labelled sperm cells and RFP (red) labelled vegetative nuclei floating in liquid medium after bursting of pollen tubes using 8% mannitol solution. **C-D)** Flow cytometry profile of sperm cells for FACS: **C)** Gating strategy template established using sperm cells from mature pollen grains, **D)** Gating strategy used for sorting sperm cells from pollen tubes. Populations of gated sperm cells are marked by circles. SSC-Side Scatter, FSC-Forward Scatter. Bar: A- 100 µm, B- 50 µm

4.4.2. Flow cytometry and sorting of single sperm cells

For FACS sorting, Arabidopsis sperm cells expressing GFP were isolated from mature pollen grains and used to establish the gating strategy (Figure 4.2C), which then served as a template for FACS of sperm cells from pollen tubes (Figure 4.2D). This pre-gating strategy using sperm cells from mature pollen grains was important, because the number of sperm cells collected from pollen tubes would not be sufficient to set the gate for sorting.

The solution containing the released sperm cells from growing pollen tubes was transferred to sperm extraction buffer and used for FACS. Sytox orange was used to distinguish intact sperm cells from dead/compromised cells. Only intact GFP positive cells, identified by GFP signal and absence of sytox-derived fluorescence, were selected during sorting. Forward and side scatter profiles were used to further discriminate sperm cells from debris. Based on this strategy, only a homogeneous population of sperm cells were sorted. In conclusion, we were able to successfully sort this rare population of sperm cells obtained from growing pollen tubes and use it for further transcriptome analysis.

4.4.3. Single sperm cell transcriptome

We developed an optimized protocol for single sperm cell RNA-seq by modifying the Smart-seq2 protocol (Picelli et al., 2014; Macaulay et al., 2016). Optimization of Smart-seq2 was done with single sperm cells along with bulk samples (e.g. 50 sperm cells) in order to assess to which extent a single cell transcriptome would be different from bulk samples with respect to number of transcripts detected, alignment rate etc.

Single cell cDNA was quantified and was found to have a concentration of around 0.3-0.7 ng/µl, with size distribution peaking around 1.5-1.7 kb (Figure 4.3A-C). Since these two parameters were not a sufficient indicator of sperm specificity of the library, we tested if amplification of sperm specific genes (*DAZ3, MGH3*) could be a viable alternative. To limit the loss of cDNA for subsequent steps, an additional step of re-elution of cDNA after the first beads clean-up was introduced and this extra eluted cDNA was used as template for PCR.

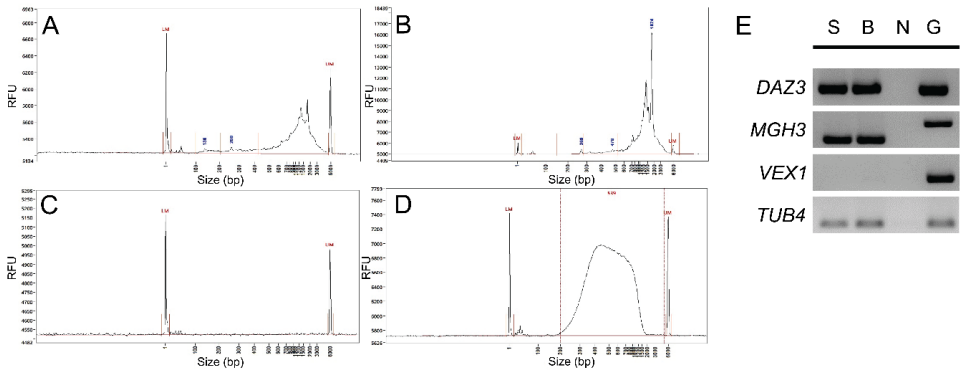


Figure 4.3: Representative electropherogram plots from Fragment Analyzer showing various stages of single cell RNA-seq library preparation. A) Example of a good cDNA amplification from single cell with fragment distribution between 600 and 2,500 bp, and an average size of 1.5–2 kb. **B)** Example of a good cDNA amplification from bulk sample used as a positive control, with a similar fragment distribution. **C)** Example of a negative (no cell) control. **D)** Example of a typical tagmented and indexed Nextera library with a size distribution of 300–800 bp. RFU, Relative fluorescence units. **E)** RT-PCR analysis of sperm specific transcripts (*DAZ3, MGH3*) from independent cDNA libraries obtained from single sperm cells and bulk populations. *VEX1* gene specific primers were used as a reference to show purity (no vegetative nuclei contamination). S- Single sperm cell, B- Bulk population (50 sperm cells), N- Negative control, G- genomic DNA. *TUB4* was used a positive control.

As shown in Figure 4.3E, we were successful in performing PCR on the cDNA obtained from single cells and thus could add this important quality control step before library preparation to our protocol. To check for contamination from vegetative nuclei, we analysed some of our cDNA libraries using gene specific primers for *VEX1* (Supplemental Table S4.1). Only single cell cDNAs that showed PCR amplification for *DAZ3*, were then used for Nextera library preparation (Figure 4.3D).

4.4.4. Number, quality and mapping rate of single cells

We sequenced 80 single cell libraries along with two bulk samples. Our single sperm cell RNAseq protocol produced high quality libraries with uniform coverage along the transcript length, consistent with the data obtained from bulk population (Figure 4.4, Supplemental Figure S4.1).

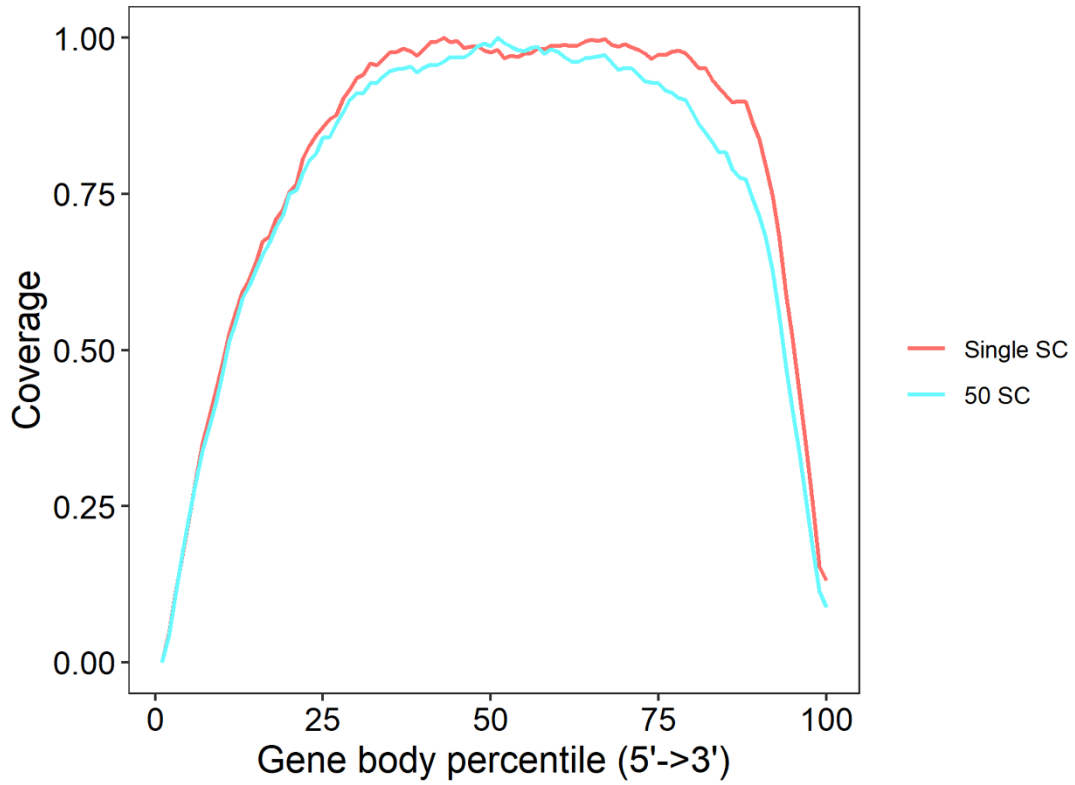


Figure 4.4. Gene body coverage. 3´ and 5´ coverage of sequenced reads from 50 cells and a single cell (SC) along the length of transcripts. Graph was generated using geneBody_coverage.py from the RSeQC package version 2.6.4 (Wang et al., 2012) and reshaped using ggplot2 package from Rstudio (http://www.rstudio.com/).

Approximately 5 million reads per cell were targeted for 80 single cell libraries (Figure 4.5A). The sperm cell libraries sequenced showed an average of 81% mapping rate with maximum reaching 88%, with large number of reads falling into exonic regions (Supplemental Fig S4.2). Our results from single sperm cells suggest that on average around 1900±571 active genes were detected in each sperm cell (Figure 4.5B). And even though some cells were sequenced more deeply, the overall number of genes detected as expressed in these single sperm cells did not increase dramatically. We also estimated the number of transcripts detected for the single cell libraries using the

Salmon tool and the representative result for one single cell and bulk (50 SC) library are shown in Figure 4.6A. A transcript was considered to be expressed, if the transcript per million (TPM) was higher than 1.

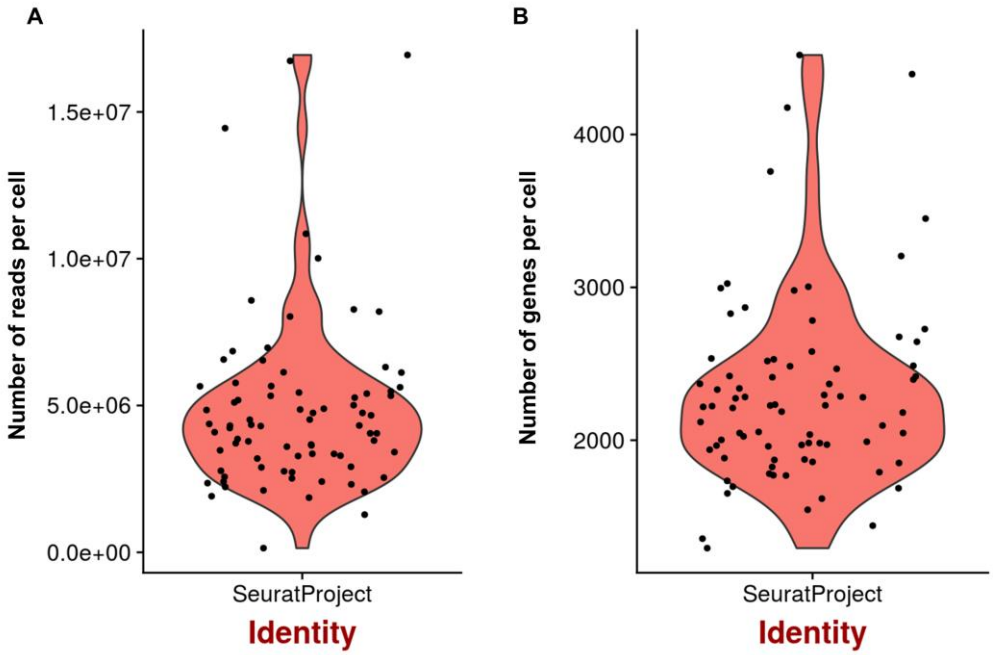


Figure 4.5: Single sperm cell transcriptome data analysis. A) Violin plot showing sequencing depth of each cell. **B)** Number of genes detected as expressed in each cell with at least one read per gene.

4.4.5. Comparison of single cell vs bulk samples

In order to assess if the single cell RNA-seq (scRNA-seq) data were in agreement with results from bulk experiments, the scRNA-seq protocol was performed with samples containing bulk populations of 50 sperm cells (SCs), which were processed in the same batch as single cells and sequenced together. As shown in Figure 4.6A the number of transcripts detected (Transcript per million, TPM>1) was higher in bulk samples (50 SC) than in single sperm cells.

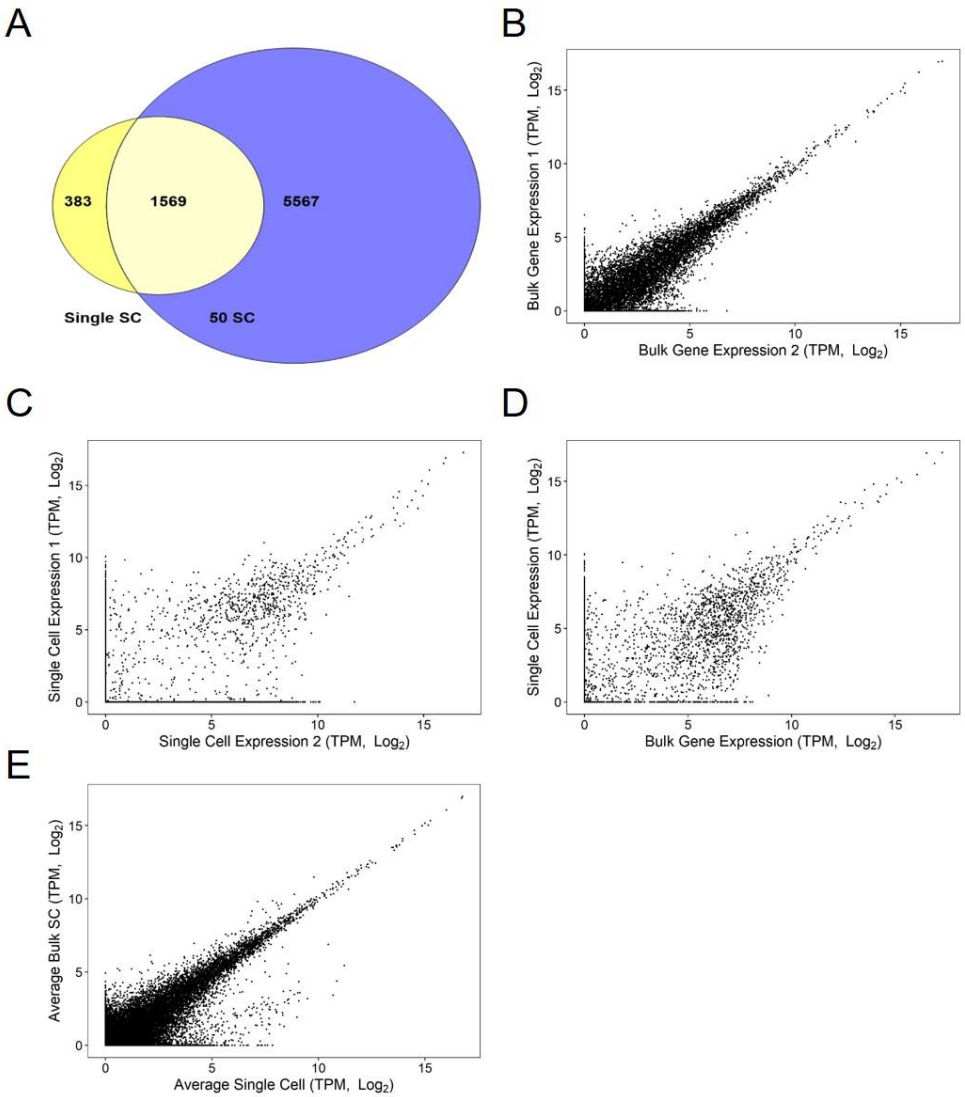


Figure 4.6: Single cell vs bulk gene expression. A) Comparison of gene expression levels of representative single sperm cell (SC) and bulk sample (50 sperm cells). Number of transcripts detected above threshold (TPM>1) are given. **B)** Gene expression correlation (normalised) between two bulk samples (Pearson´s Coefficient r= 0.84); **C)** Correlation between two single cells (Pearson´s Coefficient r= 0.57); **D)** Correlation between bulk sample and gene expression of a single cell (Pearson´s Coefficient r= 0.61); **E)**

Correlation between bulk sample and average of 80 single cell gene expression (Pearson´s Coefficient r= 0.84).

Gene expression data of bulk and single cells were used to draw scatter plots using R studio. Analysis of the data showed very high correlation between RNA seq data obtained from two bulk populations, with Pearson correlation coefficient of 0.84 (Figure 4.6B). This correlation dropped to 0.61 when comparing single cell vs bulk population (Figure 4.6C) and 0.57 when comparing two single cell transcriptomes (Figure 4.6D). We then compared the correlation between the average of 80 cells with the bulk population and strong correlation exists (Figure 4.6E). This showed that all the single cell libraries together are representative of expression in the bulk data.

4.4.6. Single sperm cell transcriptome data cluster into 3 main clusters

Clustering analysis of the single cell data identified 3 main clusters: cluster “0”, and clusters “1” and “2”, as shown in Figure 4.7. We then performed differential gene expression analysis to identify genes that were differentially expressed between cluster 0 and clusters 1 and 2. We found 58 genes to be differentially expressed, but when we looked at the marker genes for cluster 0, we found several genes that are associated with pollen development and tube growth, and were highly expressed (e.g. pectin methylesterase AGP11).

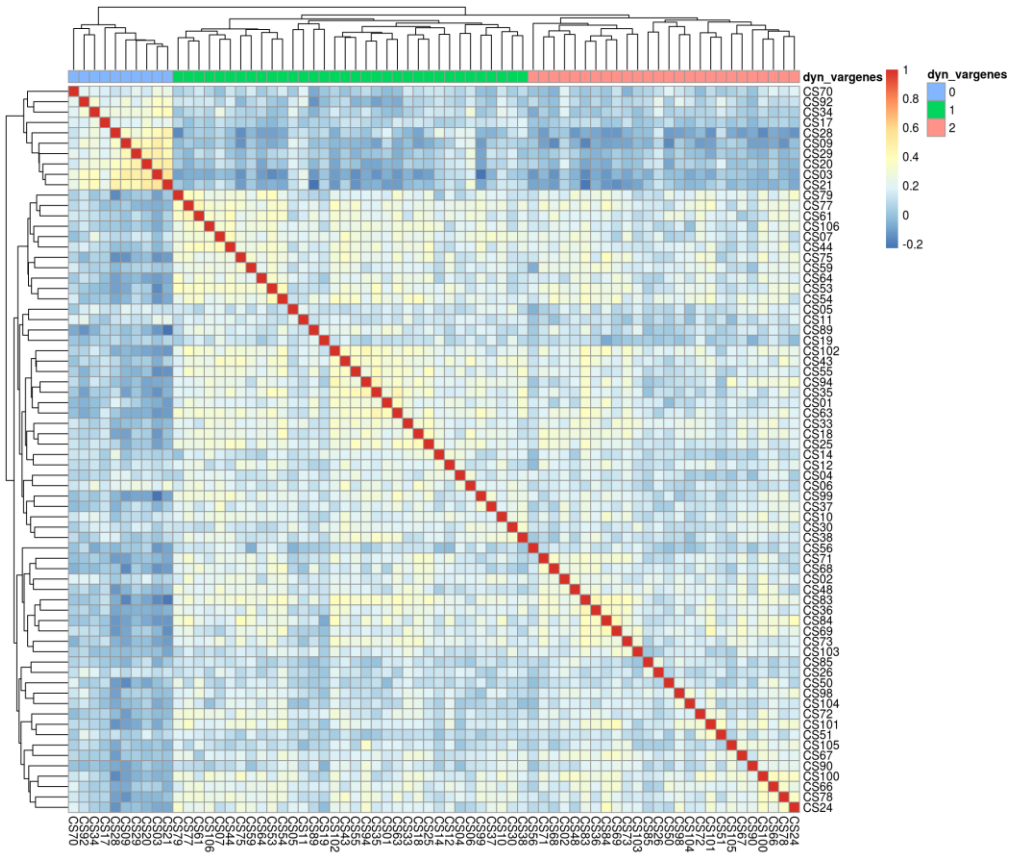


Figure 4.7: Heat map showing clustering of single sperm cell RNA-seq data. The single sperm cell transcriptomic data could be divided into two major cluster (1 and 2) and one minor cluster (0).

Based on the genes that represented cluster 0, we concluded that cell populations forming cluster 0 were most likely contaminated with pollen cytoplasm. Consequently, we focussed our analysis on those genes differentially expressed between cluster 1 and cluster 2. We found 208 genes to be differentially expressed between these cluster. We then manually curated the functional categories of those genes and found that the highest number of genes fell into the categories transcription, protein metabolism, transport and signal transduction (Figure 4.8). Genes related to sperm cell differentiation and cell cycle regulation were also found to be differentially expressed.

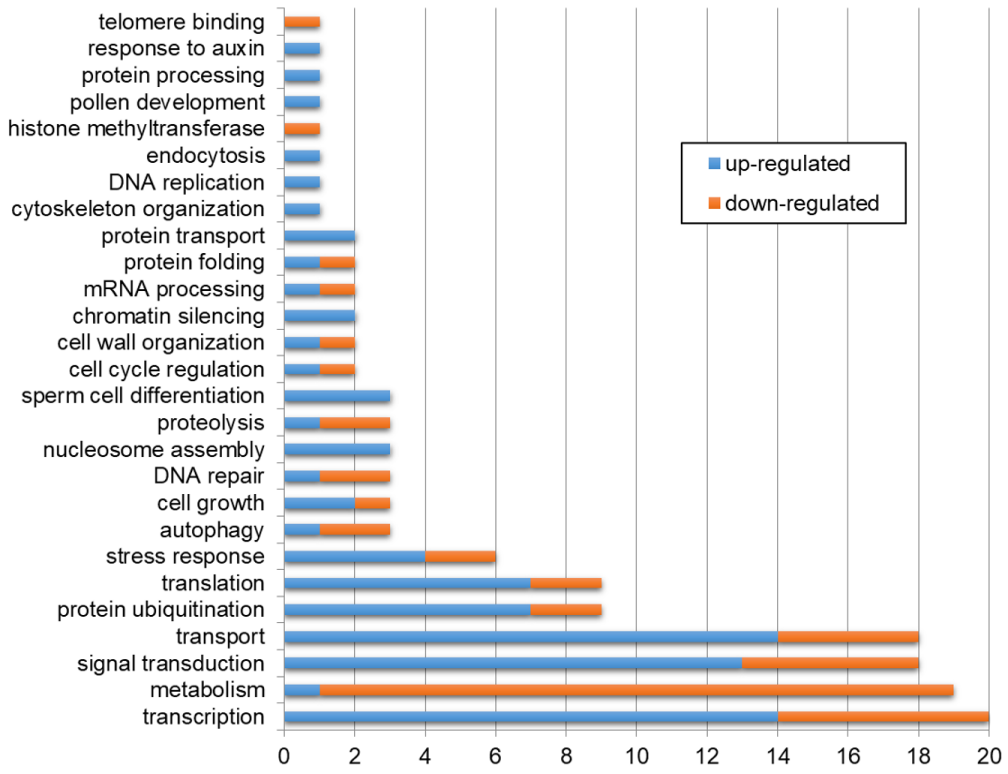


Figure 4.8: Functional classification of differentially expressed genes between the two sperm cell clusters 1 and 2.

4.5. Discussion

Plant male germline development has been an active area of research, with a wealth of transcriptomic and epigenomic data available (see review Rutley and Twell, 2015). All transcriptome studies on pollen (e.g. Hony and Twell, 2004; Pina et al., 2005) and sperm development (Borges et al., 2008) so far have been performed using bulk cell isolations, thus obscuring any potential differences at single cell level. Recent improvements of scRNA-seq protocols, a decrease in cost of high throughput sequencing, and advances in FACS isolation of Arabidopsis sperm cells, however, have allowed us to develop a single sperm cell RNA-seq method that should enable us to address the question, if isomorphic Arabidopsis sperm cells differ in their transcriptomes.

Rather than isolating sperm cells from mature pollen grains, which can be achieved in high numbers using existing protocols (Chumak et al., 2015; Santos et al., 2017), we optimized their isolation from semi *in vivo* grown pollen tubes. It is well known that pollen-pistil interactions elicit changes in the pollen tube transcriptome (Qin et al., 2009; Leydon et al., 2017), but to which extend changes in the sperm cell transcriptome contribute to this observation could not be assessed until now. Our optimized protocol for collection of sperm cells from semi *in vivo* growing pollen tubes removes these limitations. Single cell RNA-seq in general and Smart-seq2 in particular is a method primarily developed for mammalian cells that are much larger (10-100 µm), and thus assumingly with a higher cellular content (including RNA) than Arabidopsis sperm cells with a size of ~2.5 µm and very little cytoplasm. Therefore, there was no guarantee that we would obtain suitable cDNA libraries from single sperm cells of Arabidopsis. A number of intermediate optimization and quality control steps addressed this question. At the cDNA level we introduced a QC step, in which we performed a PCR-based validation of sperm cell specificity and purity (Figure 4.3E). We checked selected single cell libraries for possible contamination from the vegetative cell using primers specific for the VEX1 gene. However, VEX1 is not expressed at very high levels in pollen and it might be useful to check other highly expressed vegetative cell specific genes. The PCR based validation step together with cDNA fragmentation profiles allowed us to select only those cDNA libraries for further processing and sequencing that were of highest quality. Furthermore, we noticed that addition of 1% β-mercaptoethanol in the lysis buffer increased the alignment rate of RNA-seq data from single sperm cells by up to 40%. A growing number of recent single cell studies supports this observation (Puram et al., 2017; Villani et al., 2017).

We compared the transcriptomes of single sperm cells with those of 50 sperm cells (bulk sample). In the bulk sample, we detected more than 7000 genes (TPM>1), which is 20% more genes than the 5829 genes detected using

ATH1 gene arrays (Borges et al., 2008). Such an increase was expected though, given that the array did not cover all genes of the Arabidopsis genome. The results show however that our protocol is well suited to produce high quality transcriptome data from very low number of cells. The number of genes detected in a single cell however is lower, with an average of around 1900 genes (Figure 4.6A). These differences between single cell and bulk libraries are mostly due to genes with low expression levels falling below the detection limit in single cell expression data (Marinov et al., 2014). Technical variation in single cell data may arise due to low genome coverage, lower sensitivity of single cell protocols and high amplification bias, which also might explain how some genes are only expressed in single cell but not in bulk data (Figure 4.6A). This may lead to masking of real biological variation that may exist between single cells (Mantsoki et al., 2016). Our data also reveals that though strong correlation exist (Figure 4.6B) between gene expression data of two bulk population (r=0.84), this correlation drops when we compare two single cells or single cell expression with bulk population (Figure 4.6C-D). How much of the variation observed in our single cell data can be attributed to technical variation and how much represents true biological variation, possibly between sperm cells in the same male germ unit was the next question we wanted to understand.

Clustering analysis separated the single cell transcriptome data into 3 clusters (Cluster 0, 1 and 2). Cluster 0 was enriched for transcripts associated with pollen development and tube growth, clearly indicating possible contamination from pollen debris or pollen tube cytoplasm. This was surprising since we only sorted GFP positive sperm cells from growing pollen tubes, following very stringent criteria for FACS gating strategy. However, the male germ unit is embedded into the cytoplasm of the growing pollen tube, and this cytoplasm also contains proteins encoding transcripts necessary for actively growing pollen tubes. Therefore, despite extreme care, the possibility of some of this cytoplasm containing pollen vegetative nucleus derived

transcripts remaining attached to the surface of sperm cells during sorting cannot be excluded and might explain the expression levels observed.

After excluding cluster 0, we focussed our analysis on the remaining 60 cells forming clusters 1 and 2. When we compared these two clusters, we found 208 genes to be differentially expressed. However, these results do not hold when applying a stringent test for statistical significance (FDR<0.05). This might be due to the relatively low number of cells analyzed not providing sufficient statistical power.

Nevertheless, we tried to understand what could be potential function of these 208 genes, and found a majority falling into the categories transcription, signal transduction, and protein ubiquitination (Figure 4.8). Similar results were obtained in *Plumbago zeylanica,* in the sperm cell closer to the vegetative nucleus (S_{vn}) that expressed a higher number of ubiquitin-related transcripts than the other sperm cell (S_{ua}) (Gou et al, 2009). This might be due to the role of the ubiquitin complex mediated control of protein longevity which are activated in sperm (Singh et al., 2002).

One way to overcome the lack of statistical significance in our data would be to increase the number of cells. However, the Smart-seq2 protocol is expensive and with the current protocol, even a moderate increase in the number of cells would increase the cost of the whole experiment many folds and batch effects would potentially be even more pronounced. The Smart-seq2 protocol has been used successfully for the same number of cells in mammalian studies giving much higher resolution of the transcriptomic landscape at single cell level (Llorens-Bobadilla et al., 2015). This could mean that the potential difference between the two sperm cells in our study were not strong enough to be detected reliably at a statistically significant rate using a low number of cells. Furthermore, to increase the robustness of the analysis, several other changes would be desired: Firstly, instead of using sperm cells from growing pollen tubes, we would propose to use sperm cells

from mature pollen grains. The pollen tubes that emerge from the pistil after 6 hours of growth are not usually of uniform length and such difference can be observed in pollen tubes emerging from the same pistil or between different pistils. How difference in pollen tube length affects the gene expression of the single sperm cells isolated from these pollen tubes is hard to quantify. The use of sperm cells isolated from mature pollen grains would help to overcome this problem. Not only this, sperm cells from growing pollen tube are also a limiting factor when it comes to obtaining enough starting material, as only a limited number of sperm cells could be collected from growing pollen tubes. Hence using mature pollen directly would help to circumvent this additional problem of having sufficient numbers of sperm cells for statistically significant data. Secondly, to obtain high throughput analysis from a large number of cells as proposed, it is desirable to use more advanced single cell sequencing platforms, such as droplet based 10x chromium. The 3' sequencing based 10x chromium platform has recently been implemented successfully in several plant-based studies (Reviewed in Rich-Griffin et al., 2020 and Shaw et al., 2020). However, the application of the droplet method to sperm cells from mature pollen grains wouldn't be without challenge: Arabidopsis sperm cells are very fragile and burst in any medium, except sperm cell buffer. PBS and similar isotonic buffers used commonly for mammalian cells form the base of the droplet method and are not favourable for Arabidopsis sperm cells. In addition, sperm cell buffer being highly viscous makes it harder to concentrate the sperm cells to reach the desired concentration necessary for the encapsulation step (400-1000 cells/µl).

Nevertheless, with more and more advancement in single cell-based techniques, it is expected that cost effective high throughput methods for cell capture and sequencing will be developed that would facilitate an application across a wide spectrum of cell types and buffers. This would not only help to fully address the question whether the two sperm cells in Arabidopsis are transcriptionally different, but also allow the potential application of single cell

technology to the wider plant reproduction field, as seen in the recent analysis of endosperm nuclei (Picard et al., 2021).

4.6. Authors Contributions

Chandra Shekhar Misra and Jörg D. Becker (head of “Plant Reproduction and Evolution Group” at Instituto de Tecnologia Química e Biológica António Xavier (ITQB)), conceptualized the project. Chandra Shekhar Misra and Mario Santos (Lab manager at Plant Reproduction and Evolution Group at ITQB) designed the experiments for collection of single sperm cells using FACS. Chandra Shekhar Misra optimised the protocol for pollen tube bursting and Smart-seq2 RNA-seq library preparation. Marta Monteiro (Head of Flow Cytometry & Antibodies Facility at IGC), Mariana Rafael-Fernandes (then technician at Flow Cytometry Facility at IGC, now head of Flow Cytometry Unit at Institute of Molecular Medicine (IMM), University of Lisbon), and Nuno Pimpão Martins (then microscopy assistant at Advanced Imaging Unit, IGC, now PhD student at Max Planck Institute of Molecular Cell Biology and Genetics) performed the FACS isolation using MoFlo and helped with quality control after FACS. Chandra Shekhar Misra analysed the preliminary scRNA-seq data and António Sousa (Bioinformatician at IGC) helped with the clustering of the data and plots. Chandra Shekhar Misra and Jörg D. Becker discussed the results and wrote the manuscript. This work was supervised by Jörg D. Becker.

4.7. Acknowledgment

I would like to thank Vera Nunes, technician at Plant Facility, IGC for plant growth and maintenance. I would also like to acknowedge the help of other Facilities at IGC namely Microscopy Facility, Flow Cytometry Unit and Genomics Facility.

4.8. References

Anders, S., Pyl, P.T. and Huber, W., 2015. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*(2), pp.166-169.

Anderson, S.N., Johnson, C.S., Jones, D.S., Conrad, L.J., Gou, X., Russell, S.D. and Sundaresan, V., 2013. Transcriptomes of isolated *Oryza sativa* gametes characterized by deep sequencing: evidence for distinct sex-dependent chromatin and epigenetic states before fertilization. *The Plant Journal*, *76*(5), pp.729-741.

Angermueller, C., Clark, S.J., Lee, H.J., Macaulay, I.C., Teng, M.J., Hu, T.X., Krueger, F., Smallwood, S.A., Ponting, C.P., Voet, T. and Kelsey, G., 2016. Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nature Methods*, *13*(3), pp.229-232.

Baym, M., Kryazhimskiy, S., Lieberman, T.D., Chung, H., Desai, M.M. and Kishony, R., 2015. Inexpensive multiplexed library preparation for megabase-sized genomes. *PloS One*, *10*(5), p.e0128036.

Becker, J.D., Boavida, L.C., Carneiro, J., Haury, M. and Feijó, J.A., 2003. Transcriptional profiling of Arabidopsis tissues reveals the unique characteristics of the pollen transcriptome. *Plant Physiology*, *133*(2), pp.713-725.

Becker, J.D., Takeda, S., Borges, F., Dolan, L. and Feijó, J.A., 2014. Transcriptional profiling of Arabidopsis root hairs and pollen defines an apical cell growth signature. *BMC Plant Biology*, *14*(1), pp.1-14.

Berger, F. and Twell, D., 2011. Germline specification and function in plants. *Annual Review of Plant Biology*, *62*, pp.461-484.

Boavida, L.C. and McCormick, S., 2007. TECHNICAL ADVANCE: Temperature as a determinant factor for increased and reproducible in vitro pollen germination in Arabidopsis thaliana. *The Plant Journal*, *52*(3), pp.570-582.

Bolger, A.M., Lohse, M. and Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), pp.2114-2120.

Borges, F., Gardner, R., Lopes, T., Calarco, J.P., Boavida, L.C., Slotkin, R.K., Martienssen, R.A. and Becker, J.D., 2012. FACS-based purification of Arabidopsis microspores, sperm cells and vegetative nuclei. *Plant Methods*, *8*(1), pp.1-9.

Borges, F., Gomes, G., Gardner, R., Moreno, N., McCormick, S., Feijó, J.A. and Becker, J.D., 2008. Comparative transcriptomics of Arabidopsis sperm cells. *Plant Physiology*, *148*(2), pp.1168-1181.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E. and Satija, R., 2018. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology*, *36*(5), pp.411-420.

Calarco, J.P., Borges, F., Donoghue, M.T., Van Ex, F., Jullien, P.E., Lopes, T., Gardner, R., Berger, F., Feijó, J.A., Becker, J.D. and Martienssen, R.A., 2012. Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. *Cell*, *151*(1), pp.194-205.

Chen, G.H., Hu, Z.M., Guo, F.L. and Guan, X.L., 1995. Male gametophyte development inPlumbago zeylanica: cytoplasm localization and cell determination in the early generative cell. *Protoplasma*, *186*(1-2), pp.79-86.

Cheng, C.Y., Krishnakumar, V., Chan, A.P., Thibaud-Nissen, F., Schobel, S. and Town, C.D., 2017. Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. *The Plant Journal*, *89*(4), pp.789-804.

Chumak, N., Mosiolek, M. and Schoft, V.K., 2015. Sample preparation and fractionation of Arabidopsis thaliana sperm and vegetative cell nuclei by FACS. *Bio-protocol*, *5*(22).

Dupl'akova, N., Dobrev, P.I., Reňák, D. and Honys, D., 2016. Rapid separation of Arabidopsis male gametophyte developmental stages using a Percoll gradient. *Nature Protocols*, *11*(10), pp.1817-1832.

Efroni, I. and Birnbaum, K.D., 2016. The potential of single-cell profiling in plants. *Genome Biology*, *17*(1), pp.1-8.

Engel, M.L., Chaboud, A., Dumas, C. and McCormick, S., 2003. Sperm cells of Zea mays have a complex complement of mRNAs. *The Plant Journal*, *34*(5), pp.697-707.

Friedman, W.E., 1999. Expression of the cell cycle in sperm of Arabidopsis: implications for understanding patterns of gametogenesis and fertilization in plants and other eukaryotes. *Development*, *126*(5), pp.1065-1075.

Gawad, C., Koh, W. and Quake, S.R., 2016. Single-cell genome sequencing: current state of the science. *Nature Reviews Genetics*, *17*(3), p.175.

Gou, X., Yuan, T., Wei, X. and Russell, S.D., 2009. Gene expression in the dimorphic sperm cells of Plumbago zeylanica: transcript profiling, diversity, and relationship to cell type. *The Plant Journal*, *60*(1), pp.33-47.

Honys, D. and Twell, D., 2004. Transcriptome analysis of haploid male gametophyte development in Arabidopsis. *Genome Biology*, *5*(11), pp.1-13.

Ibarra, C.A., Feng, X., Schoft, V.K., Hsieh, T.F., Uzawa, R., Rodrigues, J.A., Zemach, A., Chumak, N., Machlicova, A., Nishimura, T. and Rojas, D., 2012. Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. *Science*, *337*(6100), pp.1360-1364.

Jin, X., Wang, R.S., Zhu, M., Jeon, B.W., Albert, R., Chen, S. and Assmann, S.M., 2013. Abscisic acid–responsive guard cell metabolomes of Arabidopsis wild-type and gpa1 G-protein mutants. *The Plant Cell*, *25*(12), pp.4789-4811.

Kim, D., Langmead, B. and Salzberg, S.L., 2015. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, *12*(4), pp.357-360.

Langfelder, P., Zhang, B. and Horvath, S., 2008. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics*, *24*(5), pp.719-720.

Leydon, A.R., Weinreb, C., Venable, E., Reinders, A., Ward, J.M. and Johnson, M.A., 2017. The molecular dialog between flowering plant reproductive partners defined by SNP-informed RNA-sequencing. *The Plant Cell*, *29*(5), pp.984-1006.

Libault, M., Brechenmacher, L., Cheng, J., Xu, D. and Stacey, G., 2010. Root hair systems biology. *Trends in Plant Science*, *15*(11), pp.641-650.

Libault, M., Pingault, L., Zogli, P. and Schiefelbein, J., 2017. Plant systems biology at the single-cell level. *Trends in Plant Science*, *22*(11), pp.949-960.

Lin, S.Y., Chen, P.W., Chuang, M.H., Juntawong, P., Bailey-Serres, J. and Jauh, G.Y., 2014. Profiling of translatomes of in vivo–grown pollen tubes reveals genes with roles in micropylar guidance during pollination in Arabidopsis. *The Plant Cell*, *26*(2), pp.602-618.

Liu, L., Lu, Y., Wei, L., Yu, H., Cao, Y., Li, Y., Yang, N., Song, Y., Liang, C. and Wang, T., 2018. Transcriptomics analyses reveal the molecular roadmap and long non-coding RNA landscape of sperm cell lineage development. *The Plant Journal*, *96*(2), pp.421-437.

Llorens-Bobadilla, E., Zhao, S., Baser, A., Saiz-Castro, G., Zwadlo, K. and Martin-Villalba, A., 2015. Single-cell transcriptomics reveals a population of

dormant neural stem cells that become activated upon brain injury. *Cell Stem Cell*, *17*(3), pp.329-340.

Macaulay, I.C., Ponting, C.P. and Voet, T., 2017. Single-cell multiomics: multiple measurements from single cells. *Trends in Genetics*, *33*(2), pp.155-168.

Macaulay, I.C., Teng, M.J., Haerty, W., Kumar, P., Ponting, C.P. and Voet, T., 2016. Separation and parallel sequencing of the genomes and transcriptomes of single cells using G&T-seq. *Nature Protocols*, *11*(11), pp.2081-2103.

Mantsoki, A., Devailly, G. and Joshi, A., 2016. Gene expression variability in mammalian embryonic stem cells using single cell RNA-seq data. *Computational Biology and Chemistry*, *63*, pp.52-61.

Marinov, G.K., Williams, B.A., McCue, K., Schroth, G.P., Gertz, J., Myers, R.M. and Wold, B.J., 2014. From single-cell to cell-pool transcriptomes: stochasticity in gene expression and RNA splicing. *Genome Research*, *24*(3), pp.496-510.

Marks, M.D., Betancur, L., Gilding, E., Chen, F., Bauer, S., Wenger, J.P., Dixon, R.A. and Haigler, C.H., 2008. A new method for isolating large quantities of Arabidopsis trichomes for transcriptome, cell wall and other types of analyses. *The Plant Journal*, *56*(3), pp.483-492.

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal*, *17*(1), pp.10-12.

Mogensen, H.L., 1992. The male germ unit: concept, composition, and significance. *International Review of Cytology*, *140*, pp.129-147.

Okonechnikov, K., Conesa, A. and García-Alcalde, F., 2016. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*, *32*(2), pp.292-294.

Palanivelu, R. and Preuss, D., 2006. Distinct short-range ovule signals attract or repel Arabidopsis thaliana pollen tubes in vitro. *BMC Plant Biology*, *6*(1), pp.1-9.

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A. and Kingsford, C., 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, *14*(4), pp.417-419.

Picard, C.L., Povilus, R.A., Williams, B.P. and Gehring, M., 2021. Transcriptional and imprinting complexity in Arabidopsis seeds at single-nucleus resolution. *Nature Plants*, *7*(6), pp.730-738.

Picelli, S., Faridani, O.R., Björklund, Å.K., Winberg, G., Sagasser, S. and Sandberg, R., 2014. Full-length RNA-seq from single cells using Smart-seq2. *Nature Protocols*, *9*(1), pp.171-181.

Pina, C., Pinto, F., Feijó, J.A. and Becker, J.D., 2005. Gene family analysis of the Arabidopsis pollen transcriptome reveals biological implications for cell growth, division control, and gene expression regulation. *Plant Physiology*, *138*(2), pp.744-756.

Puram, S.V., Tirosh, I., Parikh, A.S., Patel, A.P., Yizhak, K., Gillespie, S., Rodman, C., Luo, C.L., Mroz, E.A., Emerick, K.S. and Deschler, D.G., 2017. Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell*, *171*(7), pp.1611-1624.

Qiao, Z. and Libault, M., 2013. Unleashing the potential of the root hair cell as a single plant cell type model in root systems biology. *Frontiers in Plant Science*, *4*, p.484.

Qin, Y., Leydon, A.R., Manziello, A., Pandey, R., Mount, D., Denic, S., Vasic, B., Johnson, M.A. and Palanivelu, R., 2009. Penetration of the stigma and style elicits a novel transcriptome in pollen tubes, pointing to genes critical for growth in a pistil. *PLoS Genetics*, *5*(8), p.e1000621.

Rich-Griffin, C., Stechemesser, A., Finch, J., Lucas, E., Ott, S. and Schäfer, P., 2020. Single-cell transcriptomics: a high-resolution avenue for plant functional genomics. *Trends in Plant Science*, *25*(2), pp.186-197.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G. and Mesirov, J.P., 2011. Integrative genomics viewer. *Nature Biotechnology*, *29*(1), pp.24-26.

Russell, S.D., 1985. Preferential fertilization in Plumbago: ultrastructural evidence for gamete-level recognition in an angiosperm. *Proceedings of the National Academy of Sciences*, *82*(18), pp.6129-6132.

Rutley, N. and Twell, D., 2015. A decade of pollen transcriptomics. *Plant Reproduction*, *28*(2), pp.73-89.

Saito, C., Nagata, N., Sakai, A., Mori, K., Kuroiwa, H. and Kuroiwa, T., 2002. Angiosperm species that produce sperm cell pairs or generative cells with polarized distribution of DNA-containing organelles. *Sexual Plant Reproduction*, *15*(4), pp.167-178.

Santos, M.R., Bispo, C. and Becker, J.D., 2017. Isolation of Arabidopsis pollen, sperm cells, and vegetative nuclei by fluorescence-activated cell sorting (FACS). In *Plant germline development* (pp. 193-210). Humana Press, New York, NY.

Shaw, R., Tian, X. and Xu, J., 2020. Single-cell transcriptome analysis in plants: advances and challenges. Molecular Plant. 14 (1), pp. 115-126.

Singh, M.B., Xu, H., Bhalla, P.L., Zhang, Z., Swoboda, I. and Russell, S.D., 2002. Developmental expression of polyubiquitin genes and distribution of ubiquitinated proteins in generative and sperm cells. *Sexual Plant Reproduction*, 14(6), pp.325-329.

Slotkin, R.K., Vaughn, M., Borges, F., Tanurdžić, M., Becker, J.D., Feijó, J.A. and Martienssen, R.A., 2009. Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell*, *136*(3), pp.461-472.

Twell, D., 2011. Male gametogenesis and germline specification in flowering plants. *Sexual Plant Reproduction*, *24*(2), pp.149-160.

Villani, A.C., Satija, R., Reynolds, G., Sarkizova, S., Shekhar, K., Fletcher, J., Griesbeck, M., Butler, A., Zheng, S., Lazo, S. and Jardine, L., 2017. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science*, *356*(6335).

Wang, L., Wang, S. and Li, W., 2012. RSeQC: quality control of RNA-seq experiments. *Bioinformatics*, *28*(16), pp.2184-2185.

Yang, C. and Ye, Z., 2013. Trichomes as models for studying plant cell differentiation. *Cellular and Molecular Life Sciences*, *70*(11), pp.1937-1948.

Zhang, J., Huang, Q., Zhong, S., Bleckmann, A., Huang, J., Guo, X., Lin, Q., Gu, H., Dong, J., Dresselhaus, T. and Qu, L.J., 2017. Sperm cells are passive cargo of the pollen tube in plant fertilization. *Nature Plants*, *3*(6), pp.1-5.

Supplemental Data

File S4.1: Detailed protocol for semi *in vivo* pollen tube growth and cDNA synthesis from single sperm cells

Materials

Semi *in vivo* pollen tube growth:

- Microscopic slides (VWR, catalog no. 6131552)
- Double sided Tape (Scotch)
- Needle 25G (Terumo, catalog no. AN-2516R1)
- Tweezer (Dumont, catalog no. 010255-PO)

Cell Collection and Lysis:

- DNA-OFF (Molecular Bioproducts, catalog no. 7010)
- RLT plus buffer (Qiagen, catalog no. 1053393)
- β-Mercaptoethanol (Sigma Aldrich, catalog no. M3148)
- Ultra Rainbow Fluorescent Particles (Beads used for sorting as a negative control), 3.1µm (Spherotech, catalog no. URFP01-30-2K)

Beads Clean up:

- Nuclease-free water (Lonza, catalog no. 51200)
- 10 M NaOH (Sigma-Aldrich, catalog no. 72068)
- 5 M NaCl (AMRESCO, catalog no. E529-500ml)
- 0.5 M EDTA, pH 8.0 (PROMEGA, catalog no. V4231)
- Trizma pre-set crystals, pH 8.3 (Sigma-Aldrich, catalog no. T8943)
- 2 M KCl (Ambion, catalog no. AM9640G)
- 1 M DTT (AppliChem, catalog no. A3668,0050)
- 1 M Tris buffer, pH 7.5 (AppliChem, catalog no. A4263,0500)
- Tween 20 (VWR, catalog no. 437082Q)
- Dyna Beads MyOne Streptavidin C1 (Invitrogen, catalog no. 65001)

- Biotinylated Oligo-dT30VN (5-Biotin-TEG-5'AAGCAGTGGTATCAACGCAGAGTACTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN-3', Biomer)

Reverse Transcription

- 1 M $MgCl_2$ (Sigma-Aldrich, catalog no. M 1028)
- dNTP mix, 10 mM each (Thermo Fisher Scientific, catalog no. R0192)
- RNAse Inhibitor (40 U/µl, Takara catalog no. 2313A)
- 5x Superscript II first-strand buffer (Thermo Fisher Scientific, catalog no. 18064014)
- 0.1M DTT (Invitrogen, catalog no. 707265)
- 5M Betaine (Sigma-Aldrich, catalog no. B0300)
- Template-switching oligo (TSO; 5-AAGCAGTGGTATCAACGCAGAGTACATrGrG+G-3; Exiqon)
- Superscript II reverse transcriptase (200 U/µl, Thermo Fisher Scientific, catalog no. 18064014).
- 2x KAPA HiFi HotStart PCR Ready Mix (KapaBiosystem, catalog no. KK2602)
- ISPCR oligo (5-AAGCAGTGGTATCAACGCAGAGT-3; IDT)
- Agencourt AMPure XP Beads (Beckman Coulter, catalog no. A63881)
- Ethanol 95–97% (vol/vol) (VWR, catalog no. 20821.330.327)

Nextera Library Preparation

- Nextera DNA Library Preparation Kit (Illumina, catalog no. FC-121-1030)
- Nextera XT index Kit v2 Set A/D (Illumina, catalog no. FC-131-2001/2004)

Equipment

- Fluorescence-activated cell sorter (MoFlo, Beckman Coulter, Fort Collins, USA)
- Fragment Analyzer (AATI)
- Microcentrifuge (VWR MiniStar Silverline, catalog no. 521-2845)
- Shaker Polymax 1040 (Heidolph, catalog no. 543-42210-00)
- 0.2ml Polypropylene PCR tubes (Eppendorf, catalog no. 951010006)
- Centrifuge tubes (15 ml/50ml VWR, catalog no. 525-0604/0610)
- Thermal cycler (Biorad, catalog no. 1861096)
- IKA Mixer Vortex Shaker (Fisher Scientific, catalog no. 12819435)
- Magna-Rack (ThermaFischer Scientific, catalog no. CS15000)
- Ring Magnet (Alpaqua, catalog no. A A001322)
- MixMate Mixer (Eppendorf, catalog no. 5353000014)
- All mixing (Section 2-4) at 2000 rpm was done on this machine.
- Biosphere filter tips 1000 µl (Sarstedt, catalog no. 70.762.211)
- Biosphere filter tips 200 µl (Sarstedt, catalog no. 70.760.211)
- Biosphere filter tips 20 µl (Sarstedt, catalog no. 70.760.213)
- Biosphere filter tips 10 µl (Sarstedt, catalog no. 70.1130.210)
- 0.8 mL Deep-well storage plates (Thermo Scientific Abgene, catalog no. 12194162)
- Adhesive PCR Plate Seals (ThermoFisher, catalog no. AB0558)
- Qubit assay tubes (Axygen, catalog no. PCR-05-C)
- Qubit dsDNA HS assay kit (Invitrogen, catalog no. 32854)
- Qubit 2.0 fluorometer (Invitrogen, catalog no. Q32866)

Methods:

Semi *in vivo* pollen tube growth and sperm cell release

For semi *in vivo* pollen tube growth, the marker line was used to pollinate WT flower buds, emasculated 16 hours prior to pollination. After 2 hours, the

pollinated pistil was excised and placed on double sided tape. The excised pistil was then cut at the junction of style and ovary and placed gently on 150 µl of liquid Arabidopsis pollen germination medium spread uniformly on the thin cover slide. Care has to be taken to avoid submergence of the papillae (stigma) with pollen grains. The pistil was incubated for an additional 3-4 hours for the pollen tubes to emerge from the cut end of the style (Boavida and McCormick, 2007). The pollen tubes were subjected to osmotic shock (8% w/v mannitol) for 15 minutes.

After the pollen tubes were burst, the suspension was immediately transferred to ice cold sperm extraction buffer: 1.3 mM H_3BO_3, 3.6 mM $CaCl_2.2H_2O$, 0.74 mM KH_2PO_4, 438 mM Sucrose, 7 mM MOPS, 0.83 mM $MgSO_4.7H_2O$ with pH adjusted to 6.0. (Santos et al., 2017). The cells were suspended in a final volume of 300 µl of sperm cell buffer containing 6 µl of Sytox orange (25nM final concentration), and subjected to FACS. Sperm extraction buffer should remain on ice until FACS sorting.

FACS Setting:

In order to sort viable sperm cells of high purity and integrity, cell sorting should be run at constant pressure of 30 Psi using a 100 µm nozzle with a frequency of drop formation of 41 kHz.

Before starting to sort, confirm that voltages are optimal using the prepared samples in sperm extraction buffer. SC can be identified as GFP positive in the fluorescence 530/40 nm channel. In order to set the gate for GFP positive, Sytox Orange negative cells, use the channel for RFP (630/75nm). To sort, consider only viable sperm cell gates defined by absence of Sytox Orange staining and/or population with increased Forward scatter (FSC) and a uniform side scatter (SSC profile). This helps to exclude debris and cells which have altered integrity.

Single cell transcriptome of *Arabidopsis* sperm cells

Here we describe a protocol for single cell RNA-seq (scRNA-seq) of Arabidopsis sperm cells based on a modified Smart-seq2 protocol (Picelli et al., 2014; Macaulay et al., 2016). This protocol is described for 15 reactions including one positive control (50 cells) and a negative control. A tube with a single bead (Spherotech) serves as a negative control for FACS as well as Smart-seq2 protocol. The first part of reactions from beads clean up until reverse transcription should be performed in a designated pre-PCR room. If cells were frozen, thaw them on ice and spin down before use. If more than 50 cells are sorted in a tube, it is recommended to briefly vortex the tube before freezing. All the reagents described below should be aliquoted based on planned experiments and stock solutions should be opened as few times as possible to prevent contamination.

Section 1: Preparation of Oligo-dT conjugated beads

This step requires different solutions for beads clean-up, as shown in Table 1. These reagents need to be prepared fresh and can be stored at 4^{o}C up to six months. We recommend aliquoting small volumes depending on the predicted batch sizes, and to discard any remaining volume of an aliquot in use.

Table 1: Composition of beads clean up solutions

Reagents	Component	Stock Concentration(M)	Final Concentration (M)	Final Volume (mL)
DynabeadsA	Sodium Chloride 5M	5	0.05	0.15
	Sodium hydroxide	10	0.1	0.15
	Water			14.7
DynabeadsB	Sodium Chloride 5M	5	0.1	0.3
	Water			14.7
Dyna 2x B&W buffer	Tris Buffer pH 7.5 (1M)	1	0.01	0.15
	EDTA 0.5M	0.5	0.001	0.03
	Sodium Chloride 5M	5	2	6
	Water			8.82
G&T Seq wash buffer	Trizma® Pre-set crystals	0.1	0.05	7.5
	KCl (2 M)	2	0.075	0.562
	$MgCl_2$ Solution	1	0.003	0.045
	DTT 1M	1	0.01	0.15
	Tween 20	100	0.5	0.075
	Water			6.6675

1. Before starting, make sure the Dynabeads are at room temperature for at least 30 minutes. Vortex for 1 minute to make the bead suspension uniform.
2. For 16 reactions, add 12.5 µl of Dynabeads to a 1.5 ml Eppendorf tube and place it on a magnet for 1 minute or until a clear solution is obtained. Then discard the supernatant while keeping the tube on the magnet. The position of the tube may have to be adjusted on the magnet to facilitate complete removal of the supernatant.
3. Remove the tube from the magnet and resuspend the beads in 200 µl of Dynabead solution A. Mix by pipetting up and down slowly up to 10 times and return the tube to the magnet for 1 minute or until the solution is clear. Remove the supernatant while keeping the tube on the magnet. Repeat this step once.
4. Remove the tube from the magnet and resuspend the beads in the 200 µl of Dynabeads solution B. Mix by pipetting up and down slowly up to 10 times and return the tube to the magnet for 1 minute or until the solution is clear. Remove the supernatant while keeping the tube on the magnet.
5. Remove the tube from the magnet and resuspend the beads in 12.5 µl of 2X B&W buffer.
6. Add the same amount (12.5 µl) of 100 µM Biotinylated Oligo-dT30VN to the beads. Gently mix the beads and oligo-dT by setting the pipette to 20 µl.
7. Incubate the mixture for 20 minutes with gentle rotation on polymax shaker at room temperature.
8. Place the Eppendorf tube containing the oligo-dT conjugated beads on the magnet for 1 minute and discard the supernatant while keeping on the magnet.
9. Remove the tube from the magnet and slowly wash the beads with 200 µl of 1X B&W buffer (Dilute 2X B&W buffer with water 1:1). Place the tube on the magnet for one minute and remove the supernatant. Repeat this step three more times. Care has to be taken to avoid beads being lost

during the washing steps. For this, gentle mixing of the buffer and using the same tip for each wash and supernatant removal is recommended.

10. Prepare 200 µl of resuspension buffer by mixing the following reagents as described below in Table 2. Add 200 µl of bead resuspension buffer to the beads and mix the beads by gently pipetting up and down or until the solution looks homogenous.

Table 2: Composition of resuspension buffer for beads

Reagent	**Stock Concentration**	**Amount (µl)**	**Final concentration**
Superscript II first-strand buffer	5×	40	1×
RNAse inhibitor	40 U/µl	4.96	1 U/µl
Nuclease-free water		155.04	
Total		200	

Section 2: Capture of polyadenylated mRNAs

1. Prepare the RT master mix for 16 reactions as described in Table 3. The reagents have to be mixed in the order indicated in the table. The RT mix should be kept on ice until use.

Table 3: Master mix for reverse transcription reaction

Reagent	Stock concentration	Volume/reaction (µl)	Final Concentration	16X
Nuclease-free water		1.1375		18.2
dNTP mix	10 mM	0.5	1 mM	8
TSO	10 µM	0.5	1 µM	8
$MgCl_2$	100 mM	0.3	6 mM	4.8
Betaine	5 M	1	1 M	16
Superscript II first-strand buffer (5×)	5X	1	1 X	16
DTT	100 mM	0.25	5 mM	4
Superscript II reverse transcriptase (200 U/µl)	200 U/µl	0.25	10 U/µl	4
RNAse Inhibitor	40 U/µl	0.0625	0.5 U/µl	1
Total Volume		5		80

2. Prepare the G&T seq wash buffer by mixing buffer with 0.05X RNAse inhibitor at room temperature just before this step.
3. Thaw the tubes containing single cell lysate if stored at -80^{o}C.
4. Add 10 µl of oligo-dT conjugated beads from Step 10 of section 1 to the tubes containing single cell lysate.
5. Incubate the beads with cell lysate at room temperature on the mixer with 2000 rpm for 20 minutes.
6. After 20 minutes, place the tubes on the low elution magnet for 1 minute and remove the supernatant to a new tube in the same order.
7. Wash the beads with 10 µl of G&T seq wash buffer. Resuspend the beads in G&T seq buffer after removing the tube from the magnet.
8. Return the tube to the low elution magnet and allow the beads to settle for 1 minute. Carefully aspirate all the supernatant containing gDNA and transfer to the same tube containing the gDNA fraction from step 6.
9. Repeat the steps (7-8) once again to wash the beads.
10. Now adjust the pipette to 5 µl and remove any leftover buffer to the tube containing gDNA. The gDNA can either be stored at -20^{o}C or at -80^{o}C depending on the further application.
11. Transfer 5 µl of RT master mix to each tube, briefly mix with pipette to resuspend the beads without making bubbles and spin-down quickly to collect the beads and RT mix at the bottom of the tube.
12. Mix the tubes on the mixer at room temperature for 1 min.
13. Place the tubes in the PCR machine with heated lid according to the thermal conditions in Table 4 below:

Table 4: Thermal conditions for reverse transcription reaction

Step	Temperature (°C)	Time (min)	Mixing Conditions*
1	42	2	Mix before next step
2	42	60	Mix 3-4 times in between
3	50	30	Mix once or twice
4	60	10	

*In case the thermal cycler is used without mixer the RT reaction should be stopped every 15 minutes to shake the RT mix with beads for 30 sec at 2000 rpm.

14. Once the RT is complete, spin down the tubes for 1 min to collect the liquid at the bottom of the tubes. Proceed to the next step or store the cDNA at -20°C.

Section 3: PCR amplification of cDNA

1. Prepare the PCR amplification master mix according to Table 5 below:

Table 5: Master mix for PCR amplification of cDNA

Reagents	Stock	Volume/reaction (µl)	16X
Kapa HiFi HotStart Ready mix	2x	6.25	100
ISPCR primer	10 µM	0.125	2
Nuclease free water		1.125	18
Total		7.5	120

2. Add 7.5 µl of PCR reaction master mix directly to the tubes containing cDNA from step 14 of section 2. Spin down briefly to collect the liquid at the bottom of the wells.
3. Mix the beads and the reaction mixture briefly on the shaker at 2000 rpm for 1 minute to ensure complete resuspension of beads.
4. Perform cDNA amplification according to the conditions described below in Table 6 using thermal cycler with heated lid set to 105^0C.

Table 6: Thermal conditions for PCR amplification of cDNA

Step/cycle	Temperature (ºC)	Time (s)
1	98	180
Depending on the starting material (see Table 7)	98	20
	67	15
	72	360
1	72	300
1	4	Hold

5. Number of PCR cycles for amplification need to be adjusted based on the starting input material. The number of PCR cycles is suggested in Table 7 below:

Table 7: Number of cycles for PCR amplification of cDNA

Input Material	PCR cycles
1 cell	21-22
50 cells	18-19
500 cells	15-17

6. After completion, remove the tubes, spin down for 1 minute and proceed to clean-up step or store the amplified cDNA at -20^{o}C.

Section 4: Purification of amplified cDNA

We performed the clean-up in 0.8 mL deep-well storage plates as it is convenient and facilitates efficient recovery of the purified cDNA. If such plates are not available, clean-up on a regular ring magnet that can hold up to 0.2 ml volume should work as well.

1. Remove the Agencourt AMPure beads from the fridge to warm up to room temperature for 30 min. Vortex the beads for 1 minutes to resuspend them completely.
2. Add room temperature acclimatised beads to the amplified cDNA in 1:1 ratio and mix thoroughly up and down 10 times. Allow the mixture to stand at room temperature for 8 minutes.
3. Transfer the plate to a low elution magnet and allow the beads to settle for 2 minutes or until the solution is clear.
4. Once the beads have settled, carefully remove the supernatant without disturbing the beads.
5. Wash the AMPure beads with 100 µl of freshly prepared 80% (Vol/Vol) ethanol for 30 s and then remove the ethanol without disturbing the beads.
6. Repeat step 5 one more time.

7. Take a 20 µl pipette and remove any traces of remaining ethanol from the tubes without disturbing the beads. Allow the beads to dry for ~5 minutes or until cracks appear.
8. Add 17.5 µl of nuclease free water to the Agencourt AMPure beads. Remove the plate from the magnet to resuspend the beads in water by pipetting up and down 10 times.
9. Incubate the tubes for 2 minutes off the magnet at room temperature.
10. Place the plate on the magnet and allow the AMPure beads to settle for 5 minutes or until the solution is clear.
11. Carefully elute 16.5 µl of the supernatant containing the purified cDNA without disturbing the beads and transfer it to the new tubes at room temperature.
12. If any confirmation with PCR is planned, we recommend a second elution (repeat steps 8-11) with an additional 5 µl of nuclease free water, which should yield sufficient cDNA for PCR tests. To prevent excessive use of cDNA for PCR test, we recommend amplification of a highly expressed sperm specific gene (e.g. DAZ3). In addition some libraries should be checked for possible contamination deriving from the vegetative cell (e.g. VEX1).

Quality Control of Amplified cDNA

After the cDNA is purified, check the quality of cDNA using a Fragment Analyzer with NGS High sensitivity kit or similar. A typical cDNA obtained using this protocol should have a fragment size between 0.5 and 2 kb, reaching a maximum around 1.5-1.7 kb.

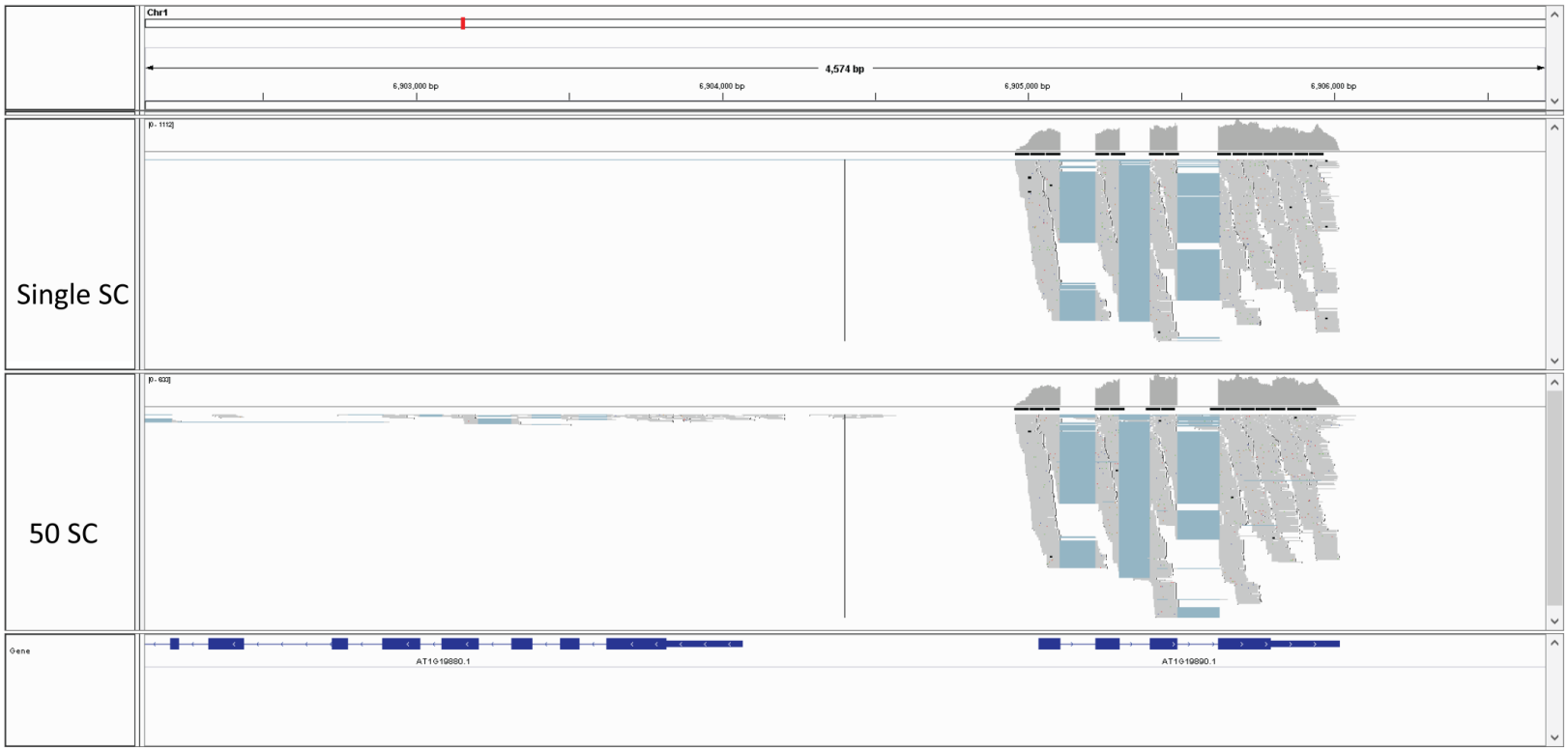


Figure S4.1 Visualization of RNA-seq data on IGV (Robinson et al., 2011) from single sperm cell and 50 sperm cells, depicting the uniform distribution of reads along the gene length and detection limits.

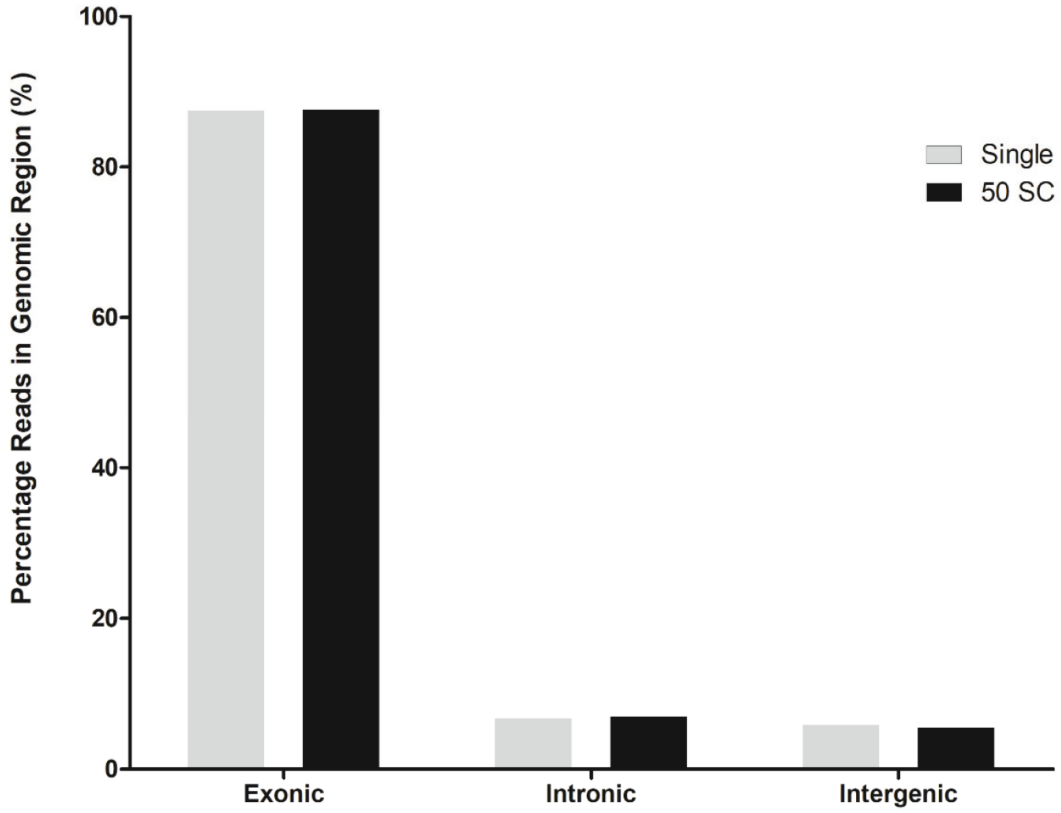


Figure S4.2 Quality control analysis of sperm cell RNA-Seq data. Representative read distribution over exons, introns and intergenic regions in single sperm cell and 50 sperm cells using Qualimap2 (Okonechnikov et al., 2015). N=3 for single SC; N=2 for 50 SC

Table S4.1: List of primers used in this study

Primers	Sequence (5´→3´)
DAZ3_for	AAGTGCTAGTGGTACCGGTG
DAZ3_rev	CACCATCGGAAGAAGCCGTAG
MGH3_for	ACTAGACGACCGTACCGTGGT
MGH3_rev	CGATTCTATCTCACCCATCAA
Tub4_for	ATCCCAAACAACGTCAAGTCC
Tub4_rev	CTCTCCGGCTGTAGCATCTTGGTAC
VEX1_for	ATGGAAGATGAAATCGGTCTC
VEX1_rev	GCAGCATACATGATGACGTTG

Chapter 5

General Discussion and Future Prospects

Male gametophyte development has been an active area of research over the last two decades. Pollen transcriptomes of several plant species have been studied, including both monocots (e.g. rice, maize) and dicots (e.g. *Arabidopsis thaliana*, tobacco), and pollen from species that are shed both at bicellular (e.g. tomato) or tricellular (e.g. *Arabidopsis thaliana*, rice) stage. The transcriptome of at least one stage of pollen development has been published for 24 species and this number continues to increase with the advancement of new techniques. Apart from whole pollen, sperm cell transcriptomes were also published for several species, including Arabidopsis (Borges et al., 2008), rice (Anderson et al., 2013), maize (Chen et al., 2017), and more recently tomato (Liu et al., 2018). Surprisingly, only rice has so far been studied for its vegetative cell and this could be due to the difficulty in collecting vegetative nuclei either from mature pollen or even growing pollen tubes. The first study on Arabidopsis pollen using RNA-seq also provided for the first time (Loraine et al., 2013), cell-type-specific genome-wide alternative splicing in plants, an area that remains largely unexplored to date. While we know that differential splicing occurs in mature pollen grains, what happens at the resolution of sperm cell and vegetative nucleus still remained unknown. Therefore, this thesis aimed to bridge some of these knowledge gaps by analyzing the transcriptome of sperm cells and vegetative nuclei, not only from mature pollen grains but also from growing pollen tubes grown on plates and as they travel within the female tissues. This is particularly important, as so far studies conducted on pollen-pistil interaction were based on whole pollen tubes and changes at the gene expression level in sperm cell and vegetative nucleus could not be determined at a sub-cellular resolution.

In chapter 2, using RNA-seq, we profiled the transcriptome of Arabidopsis sperm cells and vegetative nucleus. Our in-depth analysis of RNA-seq data helped to identify the expression of 21% more genes compared to previously published microarray-based data, 729 of which had very high expression in the sperm cells. Additionally, we also studied the transcriptome of the

vegetative nucleus for the first time in Arabidopsis, identifying 9310 genes, which is more than the average number of genes expressed in the whole pollen (Borges et al., 2008; Loraine et al., 2013). This could be possibly due to the dilution effect for the transcripts that might be expressed at low level in vegetative nucleus and fall below the threshold of detection when considered in the whole pollen grain. While both sperm cell and vegetative nucleus derive from the same progenitor cell, the microspore, transcriptional reprogramming during male gametophyte development decides their ultimate fate and functions of gamete fusion and pollen tube growth, respectively. As a result of these distinctive gene expression changes, we found 6906 genes to be differentially expressed between sperm and vegetative nucleus (Figure 2.3A). Many of the genes differentially expressed between sperm and vegetative nucleus point towards their respective function. For instance, genes that were upregulated in sperm cells were enriched for functions related to chromosome organization, DNA replication, and cell cycle and chromatin organization, pointing towards overall sperm differentiation and maturation (Figure 2.3C). On the other hand, genes that were downregulated in sperm cells (upregulated in the vegetative nucleus) were enriched for functions associated with exocytosis, cell-cell communication, and pollen tube growth, indicating transcriptional reprogramming of the vegetative nucleus to support pollen germination and tube growth. Using our RNA-seq data, not only did we provide an in-depth analysis of Arabidopsis sperm and vegetative nucleus transcriptome, but we also made use of our high-quality data together with data from 10 other tissues to reassemble the Arabidopsis annotation focussing mainly on reproductive tissues. Our hypothesis that the current Arabidopsis annotation is heavily biased towards vegetative tissues or stress-treated tissues, was proven right as we predicted 9681 novel transcripts which are 20% more than the known transcripts. Interestingly, when we compared our new annotation with the most recent and extended Arabidopsis annotation ATRTd2 (Zhang et al., 2017), we were still able to find 2148 genes in our study that were previously not predicted to undergo splicing. This is not

entirely surprising, as recently 1601 new transcripts were found even after using AtRTD2 as reference annotation (Raxwal et al., 2020), indicating that the Arabidopsis annotation is still far from complete and many tissues and/or condition-specific transcripts could be missing. Transcript level analysis identified 468 genes to be differentially spliced between sperm and vegetative nucleus, almost 34 % exclusively regulated at the splicing level. Many of the genes differentially spliced and regulated at the splicing level were enriched for functions associated with splicing machinery, protein transport, and ubiquitin-mediated proteolysis (Figure 2.3D). Our data also provided some evidence that the splicing landscape could be highly tissue-dependent and may vary between species. For instance, intron retention was considered to be the most predominant form of alternative splicing in plants, an observation that was probably based on stress-induced transcripts (Ner-Gaon et al., 2004). This notion is now being challenged in our analysis of reproductive cell-types, and also in some of the recent research that predicts overestimation of intron retention events in plants (Martin et al., 2021). We found alternative acceptor to be the most predominant form of splicing event in all three cell types (sperm, vegetative nucleus and egg cell), though the prevalence of splicing events in reproductive cell types seems to be different in other species. For instance, intron retention was the most predominant form of splicing event in rice sperm and vegetative nucleus, indicating at least to some degree the divergence of splicing landscapes between monocots and dicots, though further studies are needed to ascertain whether the splicing machinery is conserved or not across flowering plants. While genome-wide alternative splicing is well studied in several plant species, particularly as stress responses, studies on the actual function of alternative splicing are still lacking. The function of splicing factors/RNA binding proteins could be a good start as we identified at least 274 splicing factors/RNA binding proteins to be differentially expressed between sperm cells and vegetative nucleus (Figure 2.8C). Additionally, 42 of them were differentially spliced between the two cell types and represent potential candidates for exploring the function of splicing

factors in plant reproduction. While our present study focussed on the protein coding genes and their genome-wide splicing analysis, this data could also be used to study the long non-coding RNAs in Arabidopsis reproductive cell-types. So far, to the best of my knowledge, long non-coding transcriptome analysis of sperm cells is only available for rice (Johnson et al., 2018) and tomato (Liu et al., 2018). Such complementary study of mRNA and non-coding RNA would provide the full picture of the mechanism of regulation of gene expression in the male germline. Though technically challenging, transcriptome data in this study could be complemented with proteomic data for sperm cells, as so far sperm proteomes are only available for rice (Abiko et al., 2013) and lily (Zhao et al., 2013; Zhang et al., 2021). Also, though our data provide a good indication of an occurrence of genome-wide alternative splicing between sperm and vegetative nucleus, the use of single-molecule long-read sequencing will help to better understand the isoform structures and provide robustness to the conclusion of this study. Such a study can then be complemented with nascent transcript sequencing by NET-seq (Churchman and Weissman, 2011), GRO-seq (Hetzel et al., 2016) or Neu-seq (Szabo et al., 2020), and ribosome profiling by RiboTag-seq (Lesiak et al., 2015), that would help to decipher complete transcriptional and translational activities during sperm differentiation and development.

Pollen tubes have to navigate through the female tissue and reach the ovules to deliver the sperm cells for fertilization. This journey of the pollen tube is transcriptionally active and some studies previously have shed light on the changing dynamics of pollen tube transcriptomes as they travel through the pistil (Qin et al., 2009; Boavida et al., 2011; Leydon et al., 2018). However, all of these studies focussed on the whole pollen tube, and therefore the relative contribution of the sperm cell and vegetative nucleus could not be distinguished. This is particularly important, as some of the genes undergoing minor transcriptomic changes in sperm cells while it is passively transported, would remain undetected due to the dilution effect. In addition, previous

studies were done using microarrays that do not reflect the complete Arabidopsis transcriptome (Qin et al., 2009; Boavida et al., 2011). Chapter 3 of this thesis therefore aimed to address this fundamental gap in the field of plant reproduction to understand how the transcriptome of sperm cells and vegetative nucleus changes while they are transported within the pollen tube to the female ovules. For the first time, we provide direct and unambiguous evidence of transcriptome reprogramming at the sub-cellular level of sperm and vegetative nucleus during pollen tube growth. We performed a comparative analysis of sperm cell and vegetative nucleus transcriptomes from mature pollen grains, *in vitro* grown pollen tubes, and pollen tubes grown through pistils. Our results indicate that the sperm cells undergo changes at the transcriptional level and several of these changes occurred exclusively when the sperm cells passed through pistil. Most of the genes that were enriched in the sperm cell passing through the pistil were associated with heat stress response and heat shock transcription factors. Increasing evidence suggests that heat shock proteins mediate protein stability and therefore their upregulation in sperm cells during pollen tube growth could result in protein stabilization, thereby leading to sperm maturation (Wang et al., 2004; Waters and Vierling, 2020). However, what the exact functions of these heat shock proteins in sperm cells are remains elusive and would be worth exploring in the future. Our data indicate that vegetative nuclei undergo more dynamic changes in their transcriptome during pollen tube growth compared to sperm cells. The function of genes that were enriched in vegetative nuclei were associated with cell tip growth and vesicle mediated transport indicating these genes might be important for pollen tube growth. We also found a significantly large number of genes that were exclusively upregulated in the vegetative nuclei when pollen tubes passed through the pistil, indicating the role of female cues in shaping the transcriptome of vegetative nuclei. While in this chapter, we focus on the transcriptome changes occurring in sperm and vegetative nucleus at the gene level, the dataset could also serve to explore genome-wide alternative splicing during pollen tube growth. More importantly,

the transcriptomic data could also be complemented with the methylome of the sperm cell and vegetative nucleus from growing pollen tubes. Such an analysis, though technically challenging due to limitation of starting material, would provide unprecedented insight into epigenetic reprogramming during pollen tube growth, particularly under the influence of female cues. There are indications, at least for the vegetative nucleus, to undergo demethylation during *in vitro* pollen tube growth (Janousek et al., 2000), but two questions remain: firstly, whether female cues could influence the methylation status, particularly of the vegetative nucleus, and secondly, if sperm cells also undergo limited epigenetic reprogramming while they are transported towards the ovules. The mechanism behind epigenetic reprogramming in sperm or vegetative nucleus would be another interesting area to explore. It would also be worth exploring whether small RNA-mediated silencing of transposable elements continue to occur during pollen tube growth. To answer these questions, small RNA transcriptome analysis would be important to help in unravelling the entire regulatory mechanism of transcriptomic and epigenetic reprogramming during pollen tube growth, particularly under the presence of female cues.

A long-standing question in the field of plant reproduction is whether in species such as Arabidopsis the isomorphic sperm cells differ at the gene expression level. To address this question, Chapter 4 of this thesis aimed to understand the transcriptomic changes in Arabidopsis sperm cells at single-cell resolution. We used an optimized method of collecting sperm cells from growing pollen tubes and performed transcriptome analysis on 80 single sperm cells. Despite some limitations, our transcriptome data revealed some potential differences between the two sperm cells. One of the limitations of the current analysis was the lack of statistically significant results and some steps have been proposed to overcome this challenge. One of them is to use sperm cells from mature pollen grains instead of growing pollen tubes as it would help to overcome the problem of limited quantities of sperm cells that

could be collected from the growing pollen tube. Also, our current Smart-seq2 protocol could nowadays be replaced with droplet-based high throughput platforms such as 10x Chromium, which is not only more cost-effective but also a large number of cells could be profiled at the same time. Additionally, while we hypothesize that the two sperm cells are transcriptionally different, the possibility of difference also at the epigenetic level could not be excluded. Therefore, future experiments involving the use of single-cell platforms (e.g. scATAC-seq) to identify changes in the chromatin level between the two sperm cells could provide novel insight into the chromatin status of the individual sperm cells. This might be relatively easy as these protocols only require nuclei and therefore obtaining enough material from mature pollen grains would not be an issue. However, whether these changes already occur in sperm cells from mature pollen grains or are induced during pollen tube growth would be important to determine. Nevertheless, our analysis indicates the potential for the use of single-cell transcriptomic in understanding the differences between the two sperm cells in Arabidopsis as well as other agriculturally important crop species.

In conclusion, during this project, I have developed and optimised a method for collecting sperm cells and vegetative nuclei from growing pollen tubes by FACS, which so far was technically challenging. I was also able to perform RNA-seq on relatively small number of cells, even at the level of single cells. The main outcomes of my PhD project include: 1) For the first time, I was able to show the transcriptomic changes in the Arabidopsis sperm cells and vegetative nuclei occurring which occurred not only at the gene level but also at the splicing level, an aspect of male germline development that was largely unexplored in previous studies. Many of the genes found to be differentially spliced between sperm and vegetative nucleus in this study could be explored in the future for their functions in a cell-type dependent manner. 2) Our optimised protocol for collecting sperm cell and vegetative nucleus helped to collect enough sperm cells and vegetative nucleus to provide unprecedented

insights into transcriptomic changes that occurr during pollen tube growth, particularly changes that were elicited in the presence of female cues. To the best of my knowledge, this is the first study that looks at the changes during pollen tube growth at the resolution of sperm cell and vegetative nucleus. Many of the genes identified in this study could be explored for functional characterization as they could be a good candidate for studying gamete fusion and could widen the limited list of fusogens known so far. 3) Finally, I tried to understand whether two sperm cells in Arabidopsis differ at gene expression level, for which we performed single cell transcriptome analysis. Our preliminary data indicate there might be some differences between the two transcriptionally different cluster, though further studies with higher number of cells are needed to reach firm conclusions. Our data also opens up the avenue for potential use of single cell studies to answer one of the long standing questions in the plant reproductive field, namely whether plants with isomorphic male gametes do show differences at the gene expression level between two sperm cells, as was envisaged in my PhD project.

5.1. References

Abiko, M., Furuta, K., Yamauchi, Y., Fujita, C., Taoka, M., Isobe, T. and Okamoto, T., 2013. Identification of proteins enriched in rice egg or sperm cells by single-cell proteomics. *PloS One*, *8*(7), p.e69578.

Anderson, S.N., Johnson, C.S., Jones, D.S., Conrad, L.J., Gou, X., Russell, S.D. and Sundaresan, V., 2013. Transcriptomes of isolated Oryza sativa gametes characterized by deep sequencing: evidence for distinct sex-dependent chromatin and epigenetic states before fertilization. *The Plant Journal*, *76*(5), pp.729-741.

Boavida, L.C., Borges, F., Becker, J.D. and Feijó, J.A., 2011. Whole genome analysis of gene expression reveals coordinated activation of signaling and

metabolic pathways during pollen-pistil interactions in Arabidopsis. *Plant Physiology*, *155*(4), pp.2066-2080.

Borges, F., Gomes, G., Gardner, R., Moreno, N., McCormick, S., Feijó, J.A. and Becker, J.D., 2008. Comparative transcriptomics of Arabidopsis sperm cells. *Plant Physiology*, *148*(2), pp.1168-1181.

Chen, J., Strieder, N., Krohn, N.G., Cyprys, P., Sprunck, S., Engelmann, J.C. and Dresselhaus, T., 2017. Zygotic genome activation occurs shortly after fertilization in maize. *The Plant Cell*, *29*(9), pp.2106-2125.

Churchman, L.S. and Weissman, J.S., 2011. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature*, *469*(7330), pp.368-373.

Hetzel, J., Duttke, S.H., Benner, C. and Chory, J., 2016. Nascent RNA sequencing reveals distinct features in plant transcription. *Proceedings of the National Academy of Sciences*, *113*(43), pp.12316-12321.

Janousek, B., Zluvova, J. and Vyskot, B., 2000. Histone H4 acetylation and DNA methylation dynamics during pollen development. *Protoplasma*, *211*(1-2), pp.116-122.

Johnson, C., Conrad, L.J., Patel, R., Anderson, S., Li, C., Pereira, A. and Sundaresan, V., 2018. Reproductive long intergenic noncoding RNAs exhibit male gamete specificity and polycomb repressive complex 2-mediated repression. *Plant Physiology*, *177*(3), pp.1198-1217.

Lesiak, A.J., Brodsky, M. and Neumaier, J.F., 2015. RiboTag is a flexible tool for measuring the translational state of targeted cells in heterogeneous cell cultures. *Biotechniques*, *58*(6), pp.308-317.

Leydon, A.R., Weinreb, C., Venable, E., Reinders, A., Ward, J.M. and Johnson, M.A., 2017. The molecular dialog between flowering plant

reproductive partners defined by SNP-informed RNA-sequencing. *The Plant Cell*, *29*(5), pp.984-1006.

Liu, L., Lu, Y., Wei, L., Yu, H., Cao, Y., Li, Y., Yang, N., Song, Y., Liang, C. and Wang, T., 2018. Transcriptomics analyses reveal the molecular roadmap and long non-coding RNA landscape of sperm cell lineage development. *The Plant Journal*, *96*(2), pp.421-437.

Loraine, A.E., McCormick, S., Estrada, A., Patel, K. and Qin, P., 2013. RNA-seq of Arabidopsis pollen uncovers novel transcription and alternative splicing. *Plant Physiology*, *162*(2), pp.1092-1109.

Martín, G., Márquez, Y., Mantica, F., Duque, P. and Irimia, M., 2021. Alternative splicing landscapes in Arabidopsis thaliana across tissues and stress conditions highlight major functional differences with animals. *Genome Biology*, *22*(1), pp.1-26.

Ner-Gaon, H., Halachmi, R., Savaldi-Goldstein, S., Rubin, E., Ophir, R. and Fluhr, R., 2004. Intron retention is a major phenomenon in alternative splicing in Arabidopsis. *The Plant Journal*, *39*(6), pp.877-885.

Qin, Y., Leydon, A.R., Manziello, A., Pandey, R., Mount, D., Denic, S., Vasic, B., Johnson, M.A. and Palanivelu, R., 2009. Penetration of the stigma and style elicits a novel transcriptome in pollen tubes, pointing to genes critical for growth in a pistil. *PLoS Genetics*, *5*(8), p.e1000621.

Raxwal, V.K., Simpson, C.G., Gloggnitzer, J., Entinze, J.C., Guo, W., Zhang, R., Brown, J.W. and Riha, K., 2020. Nonsense-Mediated RNA Decay Factor UPF1 Is Critical for Posttranscriptional and Translational Gene Regulation in Arabidopsis. *The Plant Cell*, *32*(9), pp.2725-2741.

Szabo, E.X., Reichert, P., Lehniger, M.K., Ohmer, M., de Francisco Amorim, M., Gowik, U., Schmitz-Linneweber, C. and Laubinger, S., 2020. Metabolic

labeling of RNAs uncovers hidden features and dynamics of the Arabidopsis transcriptome. *The Plant Cell*, *32*(4), pp.871-887.

Wang, W., Vinocur, B., Shoseyov, O. and Altman, A., 2004. Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. *Trends in Plant Science*, *9*(5), pp.244-252.

Waters, E.R. and Vierling, E., 2020. Plant small heat shock proteins–evolutionary and functional diversity. *New Phytologist*, *227*(1), pp.24-37.

Zhang, R., Calixto, C.P., Marquez, Y., Venhuizen, P., Tzioutziou, N.A., Guo, W., Spensley, M., Entizne, J.C., Lewandowska, D., Ten Have, S. and Frei dit Frey, N., 2017. A high quality Arabidopsis transcriptome for accurate transcript-level analysis of alternative splicing. *Nucleic Acids Research*, *45*(9), pp.5061-5073.

Zhang, Y., You, C., Yang, S., Yao, W., Jin, W., Wang, X., WANG, W., Hu, X. and Yang, H., 2021. Proteomic analysis of generative and vegetative nuclei reveals molecular characteristics of pollen cell differentiation in Lily. *Frontiers in Plant Science*, *12*, p.452.

Zhao, X., Yang, N. and Wang, T., 2013. Comparative proteomic analysis of generative and sperm cells reveals molecular characteristics associated with sperm development and function specialization. *Journal of Proteome Research*, *12*(11), pp.5058-5071.

ITQB-UNL | Av. da República, 2780-157 Oeiras, Portugal
Tel (+351) 214 469 100 | Fax (+351) 214 411 277

www.itqb.unl.pt