IDENTIFICATION OF GENES INVOLVED IN SPIKELET FERTILITY IN RICE USING RNA SEQUENCING

BY

KANOKWAN KAEWMUNGKUN

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
THE DOCTOR OF PHILOSOPHY (BIOTECHNOLOGY)
DEPARTMENT OF BIOTECHNOLOGY
FACULTY OF SCIENCE AND TECHNOLOGY
THAMMASAT UNIVERSITY
ACADEMIC YEAR 2023

THAMMASAT UNIVERSITY
FACULTY OF SCIENCE AND TECHNOLOGY

DISSERTATION

BY

KANOKWAN KAEWMUNGKUN

ENTITLED

IDENTIFICATION OF GENES INVOLVED IN SPIKELET FERTILITY IN RICE USING RNA SEQUENCING

was approved as partial fulfillment of the requirements for
the degree of Doctor of Philosophy program (Biotechnology)

on October 27, 2023

Chairman
(Associate Professor Kittipat Ukoskit, Ph.D.)

Member and Advisor
(Associate Professor Theerachai Thanananta, Ph.D.)

Member and Co-Advisor
(Amorntip Muangprom, Ph.D.)

Member
(Assistant Professor Chareerat Mongkolsiriwatana, Ph.D.)

Member
(Assistant Professor Chanitchote Piyapittayanun, Ph.D.)

Dean
(Associate Professor Supet Jirakajohnkool, Ph.D.)

Dissertation Title	Identification of genes involved in spikelet fertility in rice using RNA sequencing
Author	Kanokwan Kaewmungkun
Degree	Doctor of Philosophy program (Biotechnology)
Department/Faculty/University	Biotechnology Faculty of Science and Technology Thammasat University
Dissertation Advisor	Associate Professor Theerachai Thanananta, Ph.D.
Dissertation Co-Advisor	Amorntip Muangprom, Ph.D.
Academic Year	2023

ABSTRACT

The identification of genes controlling spikelet fertility is directly important for the yield and quality of rice. This study developed two groups of genetically stable Recombinant Inbred Lines (RILs) with significantly different seed set rates. One group had a high seed set rate (≥ 70%) and the other had a low seed set rate (≤ 25%). These populations were created by crossing the Japanese rice variety Nipponbare with the Thai rice variety Khao Dawk Mali 105. To identify genes, the study examined the stages and tissues affecting the different seed set rates in the two groups. In the F_5 generation, the populations with different seed set rates still had complete pollen but varied significantly in seed set rates. The study further observed differences in pollen adhesion and pollen tube growth abnormalities in the low seed set population 4-5 hours after pollination. Samples were collected, and RNA sequencing was conducted at this stage in the F_6 generation. Analysis of the differentially expressed genes (DEGs) between the high and low seed set populations identified 249 DEGs in the high seed set group, with 118 up-regulated and 131 down-regulated genes. The low seed set group had 473 DEGs, with 214 up-regulated and 259 down-regulated genes. Validation of 10 DEGs using qRT-PCR showed consistent results with the RNA sequencing data. GO ontology analysis of DEGs in the high seed set group identified 18 GO terms, while the low seed set group had 16 GO terms. KEGG pathway analysis revealed that 11 DEGs in the high seed set group were involved in metabolic pathways, whereas 6 DEGs in the low seed set group were related to pentose and glucuronate interconversions and

flavonoid biosynthesis pathways. Further study of DEGs interaction with other genes involved in pollen development, using rice databases, found several interactions, including some previously reported genes related to spikelet fertility. This study discovered new genes and gene functions associated with spikelet fertility, such as *Os02g0467600* (cinnamate 4-hydroxylase, CYP73), which is involved in flavonoid synthesis and is expressed more in the high seed set group post-pollination. Additionally, *OsPME35* and *OsPLL9*, which control pectin methylesterase (PME) and pectate lyase (PLL) respectively, showed higher expression in the ovary before pollination in the high seed set group compared to the low seed set group. This indicates their necessity for cell wall formation of the pistil during the interaction between pistil and pollen. The DEGs identified were also located near QTLs related to pollen tube growth and yield components, as previously reported. These findings are valuable as they provide new knowledge about the genetic control of spikelet fertility and can be utilized for developing high-yield rice varieties.

Keywords: Transcriptome; RNA-Seq; Pollinated pistils; Spikelet fertility; Inter-subspecific crosses; Rice

ACKNOWLEDGEMENTS

First of all, this thesis would not be possible without the valuable contributions of many people and BIOTEC, National Science and Technology Development Agency (NSTDA). I would like to express my gratitude to my advisor Associate Professor Dr. Theerachai Thanananta and my co-advisor Dr. Amornthip Muangprom for his excellent advice, whose advices, suggestions, encouragement, insight, patience, kindness, and support for doctoral studies. I also would like to express my appreciation to Assistant Professor Dr. Kanokrat Limpisophon and Dr. Theerawat Dobutr for their kindness in providing technical research advice.

This thesis was supported by the Research promotion fund for international and educational excellence (Grant No. 08/2562). I am grateful to Kittipat Ukoskit, Chareerat Mongkolsiriwatana, Chanitchote Piyapittayanun, and Suriyan Cha-um for helpful suggestions. I also thank Jennifer Luangsa-ard and Alongkorn Amnuaykanjanasin for research facilities.

I would like to express my gratitude to all the teachers in the Biotechnology Department at the Faculty of Science and Technology, Thammasat University, for your invaluable knowledge, guidance, and exceptional care for your students. I also would like to express my gratitude to research promotion fund for international and educational excellence (Full amount grant), Thammasat University.

I am also appreciative of the many individuals, whose names I cannot mention, for their unwavering support during my doctoral studies. It was a privilege to collaborate with them.

Finally, I would like to thank my dearest parent, sister, brother and family for their belief, unconditional love, encouragement and support throughout my whole life.

Kanokwan Kaewmungkun

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Symbols/Abbreviations	Terms
LSSR1	*LOW SEED SETTING RATE1*
WA-CMS	Wild-abortive cytoplasmic male sterility
PTB1	*POLLEN TUBE BLOCKED*
KDML105	Khao Dawk Mali 105
HS	High seed setting rate
LS	Low seed setting rate
amiRNA	artificial microRNA
RNA-Seq	RNA Sequencing
NGS	Next-Generation Sequencing
SE	Single-End
PE	Paired-End
%GC	The percentage of G&C base numbers of total bases
Q20	Q Phred values is greater than 20
Q30	Q Phred values is greater than 30
logFC	Log2 Fold Change
DEGs	Differentially expressed genes
PCA	Principal component analysis
GO	Gene Ontology
BP	Biological Process
MF	Molecular Function
CC	Cellular Component
KEGG	The Kyoto Encyclopedia of Genes and Genomes
DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
mRNA	messenger RNA

cDNA	complementary DNA
ncRNA	noncoding RNA
miRNA	microRNA
tRNA	transfer RNA
lncRNA	long coding RNA
piRNA	piwi-interacting RNA
rRNA-Seq	ribosomal RNA sequencing
scRNA-seq	single-cell RNA sequencing
AS	Alternative Splicing
FFPE	Formalin-Fixed Paraffin-Embedded
CTCs	Circulating Tumor Cells
FACS	Fluorescence-activated cell sorting
PCR	Polymerase Chain Reaction
RT-PCR	Reverse transcription polymerase chain reaction
FDR	False Discovery Rate
TFs	Transcription Factors
RPKM	Reads per kilobase of exon model per million reads
FPKM	Fragments per kilobase of exon model per million mapped reads
TPM	Transcripts per million
TMM	Trimmed Mean of *M*-values
eQTLs	expression Quantitative Trait Loci
GSEA	Gene Set Enrichment Analysis
SNP	Single Nucleotide Polymorphism
MR	Mutual Rank
QTL	Quantitative trait loci
pg	Picogram
µg	Microlite
mm	Millimeters

mg	Milligram
°C	The degree Celsius
Rpm	Revolutions per minute
Min.	Minute
Sec.	Second
Cm	Centimeter
a.m.	Ante Meridiem
v/v	Volume per volume
M	Molar
DAPI	4′,6-diamidino-2-phenylindole
ng/µl	Nanogram per microliter
µg	Microgram
U	Unit
mL	Milliliter
U/ µl	Unit per microlitre
mM	Millimolar
$MgCl_2$	Magnesium chloride
EDTA	Ethylene diamine tetraacetic acid
DI water	Deionized water
I_2-KI	Iodine potassium iodide

CHAPTER 1
INTRODUCTION

1.1 Background and significance of the research

Rice is one of the most important staple foods that widely cultivated worldwide. By 2030, about 40 % increase in rice production will be demanded because of the rapid increasing population (Khush, 2005). The main way to increase rice production is increasing the spikelet number per panicle (Liang et al., 2017; Liu et al., 2022). Spikelet and the inner floral organs are crucial to the filling rate of grains and spikelet fertility (Sekhar et al., 2021), which is associated with pollination and fertilization (Qi & Wu, 2022). Pollen germination and pollen tube elongation are one of main stages for spikelet fertility (Matsui, 2005). The exploitation of useful genes controlling pollen tube growth through intersubspecific crosses has been an important procedure for genetic improvement in rice (C.-M. Lee et al., 2021). The inter-subspecific crosses between *indica* and *japonica* rice varieties generated F_1 having advances in genomics and genetic diversity (Khush, 2001). Utilizing hybrid rice from intersubspecific crosses is one of the most applications for promoting higher performance for a trait than both parents vigor (heterosis) (Birchler, 2015). However, the introgressive hybridization of valuable genes from distant relatives of rice is often prevented by interspecific or intersubspecific reproductive barriers (Oka, 2012).

The F_1 plant is selfed to generate an F_2 population; in the F_2 population, there are maximum segregation and recombination for imposing agronomic traits selection (Birchler, 2015; Fu et al., 2014; Guo et al., 2016; Kwon et al., 2015). The segregation of F_2 or a later generation can be applied to breeding programs in several plants for new cultivars or germplasm with desirable agronomic traits, e.g., in wheat (*Triticum vulgare*), pea (*Pisum sativum*), soybean (*Glycine max* L.) and cotton (*Gossypium hirsutum* L) (Cazzola et al., 2020; Li et al., 2007; Nevhudzholi et al., 2020; Shindo et al., 2003). In rice, the maximum segregation and recombination for imposing agronomic traits were selected to develop high-yield cultivars, such as improving the number of panicle number, grain number and spikelet fertility.

Rice is a good model monocot with the completed genome sequences, making it a valuable tool for studying reproductive development. The molecular mechanisms in reproductive development have a complexity, involving fertile and sterile spikelet in rice. A complex network of regulators controls numerous genes during panicle initiation, flower formation, embryo development, and seed storage (Guo & Liu, 2012; Huang et al., 2017; Itoh et al., 2005; Katara et al., 2020; S. Kim et al., 2021). Spikelet fertility caused height seed-setting rate in rice due to normal floret structures, fertile pollen grains, paired anther dehiscence, pollen tube growth and elongation, and embryo sac fertilization (Matsui, 2005; Yang Xu et al., 2017). Pollen tube elongation is a key step for successful pollination in rice (F. Zhang et al., 2018). The respiration rate, carbohydrate and protein metabolism and phytohormone production are increased to provide the energy during pollination (Kovaleva & Zakharova, 2004; Selinski & Scheibe, 2014; Taylor & Hepler, 1997; F. Zhang et al., 2018). Previously, *LSSR1*, *Rf3 and Rf4* were reported that they regulated the seed setting rate in rice (Pranathi et al., 2016; Xiang et al., 2019). *LSSR1* is a novel gene in rice that is expressed in the anther from the premeiotic stage to the single-cell pollen stage. *Rf3* and *Rf4* were associated with fertility restoration of wild-abortive cytoplasmic male sterility (WA-CMS) in rice. However, several studies reported that the defection of pollen tube growth in pistil could also cause reduced panicle fertility and seed setting. For examples, *POLLEN TUBE BLOCKED* (*PTB1*) regulates rice seed setting rate by controlling pollen tube growth (Li et al., 2013). *OsCNGC13* promotes seed-setting rate by supporting pollen tube growth in stylar tissues (Yang Xu et al., 2017). Future more, the knockdown of *OsPLL3* and *OsPLL4* by artificial microRNA (amiRNA) disrupted normal pollen development and resulted in partial male sterility. *OsUCL8* regulates pollen intine deposition and pollen tube growth (F. Zhang et al., 2018). In addition, the communication between the pollen grain and pistil occurs closely and frequently to complex reproductive success (Yu et al., 2017). Nevertheless, the lack of a large-scale to evaluate information of the regulatory networks and potential genes of the pollen tube growth enhancing reproductive success, in particular, the identification of genes involved in the fertility of spikelet by controlling the growth of the pollen tube in inter-subspecific crosses of rice.

RNA Sequencing (RNA-Seq) is a technique that can examine the sequences of RNAs, including coding and non-coding RNA in the whole genome, using next-generation sequencing (NGS) (Marguerat & Bähler, 2010; Singh et al., 2012). Transcriptome analyses by RNA-Seq have been used to identify active genes at a particular point in time, such as heterosis, drought tolerance, salinity tolerance, and cold tolerance in rice (Dametto et al., 2015; Guo et al., 2017; Lenka et al., 2011; Shankar et al., 2016). Recently, RNA-Seq has been combined with bulked segregant analysis (Bulked Segregant RNA-Seq or BSR-Seq) to identify distinct expression profiles and differentially expressed genes (DEGs) for candidate genes and QTL mapping related to fertility and sterility in various plants. This approach has been applied to rice (Kamara et al., 2021), maize (*Zea mays*), (Shi et al., 2021), foxtail millet (*Setaria italica*) (Gao et al., 2022), and alfalfa (*Medicago sativa*) (Zhou et al., 2022). Furthermore, the mechanism analyzed using Go and KEGG analysis of candidate genes showed that the transport, localization, communication, and pollination-related genes were significantly enriched in the stigma. RNA-Seq profiling was used underlying pollen tube growth at pollination stage. The results indicated that the complex regulatory mechanisms in pollen tube growth at pollination stage were explored in the context of hormone signaling, transcriptional regulations related to abiotic stress, and lipid metabolism.

In this work, we studied the spikelet fertility, pollen and pollen germination in F_5 plants derived from a cross between Nipponbare and KDML105. Then, we identify the critical stage involved in spikelet fertility. RNA-Seq and qRT-PCR techniques were used to compare differentially expressed genes (DEGs) between bulked of the two extreme groups of the F_5 plants, including high seed setting rate, HS ($\geq$ 70%) and low seed setting rate, LS ($\leq$ 25%). Some DEGs were co-localized with previously identified quantitative trait loci (QTL). The results obtained provides valuable information for understanding the molecular mechanism controlling spikelet fertility in rice.

1.2 Objectives

1.2.1 To identify genes involved in rice spikelet fertility using RNA-Seq

1.2.2 To study mechanism of genes regulated in rice spikelet fertility

1.3 Expected benefits

1.3.1 RNA-Seq may help to explain the candidate gene involved in spikelet fertility.

1.3.2 The understanding mechanism of genes regulating rice spikelet fertility will be useful for rice yield production.

CHAPTER 2
REVIEW OF LITERATURE

2.1 Rice

2.1.1 Origin and cultivation

Rice is one of the world's largest food crops, providing the caloric needs of millions of people daily (Sweeney & McCouch, 2007)). More than 90% of the world's rice were grown and consumed in Asia, which was 60% of the world population (Khush, 1997). Rice (*Oryza sativa* L.) belongs to family Poaceae (Gramineae), genus *Oryza*. There are 25 species within the genus *Oryza*, organized into four species complexes: O. *sativa*, O. *officialis*, O. *ridelyi*, and O. *granulata*. However, the two most significant rice species for human nutrition are O. *sativa* and O. *glaberrima*.

O. sativa is distributed with a high concentration in Asia, while *O. glaberrima* is grown in West Africa (Sweeney & McCouch, 2007) (Figure 2.1). The common predecessor of all rice species is regarded as O. *perennis*, with O. *sativa* being derived from O. *rufipogan* and O. *nivara*, while O. *glaberrima* is derived from O. *longistaminate* and O. *barthii*. The schematic representation of the evolutionary chart is shown in Figure 2.1.

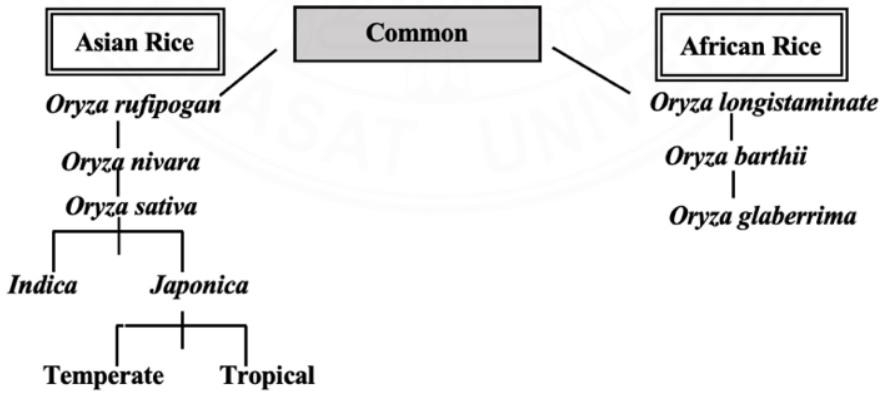

Figure 2.1 The schematic representation of evolutionary chart of Asian and African cultivated rice

The cultivated species *O. sativa* was domesticated approximately 7000-8000 years before the present (BP) in the middle to lower parts of the Yangtze River region (Zong et al., 2007). *O. glaberrima*, the African rice, is grown in West Africa, originated in the Niger River in Mali near the Guinean Coast around 1500 B.C. and 500 years later (Portères, 1956). The genome of *O. sativa* and *O. glaberrima* is composed of diploid 2n=24 (Brar *et al*., 1996). As early 1952, rice has been categorized into three geographical sub-species due to variations in its natural distribution and evolutionary development in different regions. These included O. *sativa* sub sp. *indica*, O. *sativa* sub sp. *japonica* and O. *sativa* sub sp. *Javanica* (Matsuo, 1952).

2.1.2 *Indica* and *japonica* rice

Oryza sativa is further divided into two major groups: *indica* and *japonica*. *Indica* rice varieties are typically grown in tropical regions and have long, slender grains, while *japonica* rice varieties are more commonly found in temperate climates and have shorter, rounder grains.

Indica rice, which is cultivated in tropical and subtropical regions, typically exhibits tall growth with heavy tillering (tillers being secondary stems) and long, droopy, light-green leaves. Generally, indica rice grains are long, and the starch content contains a medium to high amylose level, resulting in cooked rice that is dry, fluffy, and less prone to disintegration (Robert and Chandler, 1975) (Figure 2.2a).

Japonica rice is primarily grown in temperate climate regions, accounting for more than 10 percent of global rice trade. It is predominantly produced in Japan, Korea, and China. Japonica varieties are characterized by greener, more upright leaves and a lower tillering capacity compared to indica varieties (Figure 2.2b). Japonica rice grains are short, and the starch has a lower amylose content, making the cooked rice stickier and glossier, with grains more prone to disintegration if overboiled (Robert and Chandler, 1975).

Figure 2.2 The phenotypes of *indica* and *japonica* rice

2.1.3 Morphology

The rice plant comprises roots, stems (culms), tillers, leaves, and panicles in the mature portion of productive tillers. The plant's height varies depending on the variety and environmental conditions, ranging from 0.4 meters in dwarf varieties to 5 meters in some deep-water floating rice varieties.

2.1.3.1 Root

Rice, as a member of the Gramineae family, belongs to the monocot category and possesses an adventitious root system. Multiple adventitious roots are supported by the primary root, which develops from the extension of the radical root. Typically, taller rice varieties exhibit more extensive perforation compared to shorter varieties. Additionally, varieties with robust tillering habits tend to have well-developed root systems.

2.1.3.2 Culm

The jointed stem of rice, referred to as a culm, consists of solid nodes and hollow internodes. Primary tillers emerge from the lowermost nodes and subsequently produce secondary tillers (Figure 2.3). The extent of tillering is influenced by various factors, including rice varieties, spacing, fertilization, and cultural conditions. Typically, tillering commences approximately 10 days after transplanting and reaches its peak between the 40th and 50th days after transplanting (Arunachalam & Bandyopadhyay, 1984).

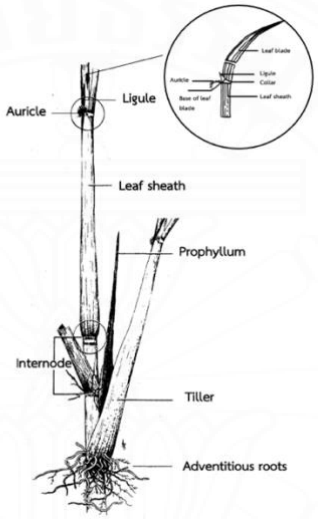

Figure 2.3 Pasts of a primary tiller (Chang and Bardenas, 1965)

2.1.3.3 Leaf

Leaves emerge from the nodes on the culm in two distinct rows. Each leaf consists of two parts: the sheath and the blade (as shown in Figure 2.3 according to IRRI). The leaf sheath is connected seamlessly to the blade and covers the culm above the node. This covering varies in terms of length, shape, and how tightly it wraps around the culm. The point where the leaf sheath and blade meet are referred to as the collar. At the base of this juncture, there are ligules and auricles present on each side of the blade. The ligule's characteristics, including its length, color, and shape, differ among varieties. The main culm bears the highest number of leaves, while the number of leaves on a tiller gradually decreases with higher tillering order. The last leaf on the culm is known as the flag leaf. Tiller growth arises from auxiliary buds and is protected by specialized leaves called prophyllums.

2.1.3.4 Inflorescence

The inflorescence structure in rice takes the form of a terminal panicle, as illustrated in Figure 2.4. The primary axis of the panicle, known as the primary axis, gives rise to multiple secondary rachii. Each secondary rachis then subdivides into tertiary rachii. These tertiary rachii, in turn, give rise to rachillae that are arranged alternately and carry spikelets at their tips. While the primary rachis has hollow internodes, the secondary and tertiary rachii do not. In the majority of rice species, the inflorescence shape is fusiform.

2.1.3.5 Flower

The structure of the spikelet comprises two sterile glumes, also known as empty glumes, and two fertile glumes, namely the outer lemma and the inner palea. At the base of the spikelet, there are two transparent and fleshy structures called lodicules, which represent the reduced perianth parts. Inside the lemma and palea, six stamens measuring 2.0-2.5 mm in length are arranged in two whorls of three each. Additionally, there is a pistil with a purple shading, containing a single ovule, a short slender style, and a bifurcated, plumose stigma. The empty glumes typically have a single nerve each, whereas the lemma is composed of five veins, as is the palea. In certain varieties, the middle vein of the lemma may elongate to form an awn (Figure 2.4).

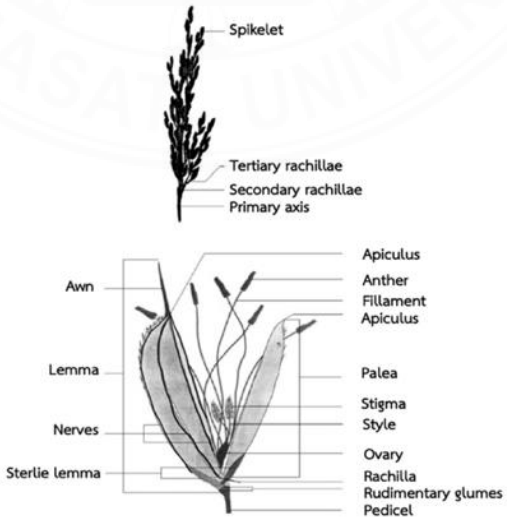

Figure 2.4 Inflorescence, floral parts and grain of rice (Vijay and Roy, 2013)

2.1.3.6 Grain

The rice grain, termed a caryopsis, possesses an oblong shape. This caryopsis is enclosed by the outer hull, which is composed of the lemma and palea. Within the rice grain, there are distinct layers, including the outer layers such as the pericarp, seed coat, and nucellus. The endosperm, in turn, is comprised of the aleurone layer and the endosperm proper, encompassing the subaleurone layer and the starchy endosperm. The aleurone layer envelops the embryo and contains rich pigment, primarily confined to the pericarp. The thickness of the aleurone layer varies, ranging from one to five cell layers.

In *indica* rice, the hull is primarily formed by the lemma, palea, and rachilla, while in japonica rice, the hull is predominantly composed of rudimentary glumes. A solitary grain of rice typically weighs around 10 to 40 mg. The hull contributes to nearly 20% of the total weight of the grain. The dimensions of the grain, including its length and width, exhibit variations across different rice varieties.

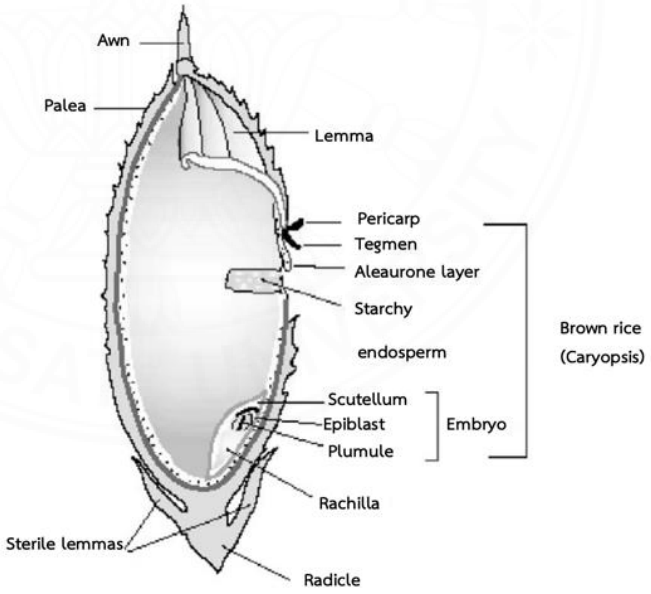

Figure 2.5 Structure of rice grain (Kennedy et al., 2002)

2.1.4 Rice growth and development

Rice is an annual grass belonging to the Poaceae (Gramineae) family. The life cycle of rice is approximately 110 to 150 days from germination to maturity, which varies depending on the variety and environmental conditions. The common cultivated rice plant is an annual that typically grows to a height of half a meter to two meters, although there are certain varieties that can reach much greater heights (6-9 meters). The growth duration of the rice plant ranges from 3 to 6 months, depending on the variety and environment. The growth of the rice plant can be divided into three stages of development: 1) vegetative, 2) reproductive, and 3) grain filling (Figure 2.6).

2.1.4.1 Vegetative stage

The vegetative stage had activation of tillering, increase in plant height and leaf emergence. Some cultivars have a short vegetative growth stage. The seed dormancy is broken when germination starts. The seed absorbs water and germinates at 10 to 40 °C. The grain is submerged to water, the coleoptile emerges ahead of the coleorhizae under anaerobic conditions. The coleorhiza slightly elongates, emerges through the seed coat. The primary seminal root (radical) breaks through the coleorhiza in the soil. Seminal roots are later replaced by the secondary system of adventitious roots. Primary leaf is elongated. The periods from seed germination to seedlings is 14-20 days.

The first tiller is emerged from the axillary bud of the second leaf on the culm. A series of nodes and internodes are made up on each stem of rice. Each upper node has a leaf and a bud, which grows into a tiller. The second tiller develops between the main stem and the third leaf from the base when the sixth leaf emerges. The rice root system comprises two primary types: crown roots and nodal roots. Both of these roots develop from nodes, but crown roots emerge from nodes below the soil surface. Nodal roots are commonly observed in rice cultivars that thrive in water depths exceeding 80 cm.

2.1.4.2 Reproductive stage

The reproductive stage of rice is marked by several key events, including the elongation of the culm, a reduction in the number of tillers, the occurrence of booting, the emergence of the flag leaf, heading, and ultimately flowering. This phase typically spans around 30 days in most rice varieties. It is also known as the internode elongation or jointing stage, and its duration may slightly differ based on the specific rice variety and prevailing weather conditions.

2.1.4.3 Ripening stage

The grain filling and ripening, also known as the maturation stage, follows the fertilization of the ovary and is characterized by the rapid growth of the grain. Throughout this phase, the grain undergoes an increase in both size and weight, while starch and sugars accumulate within it. The color of the grain changes from green to a golden or straw hue as it matures. The intensity of light during this period holds significant importance, as around 60% or more of the carbohydrates essential for grain filling are synthesized through photosynthesis. Temperature also plays a crucial role during this interval. Despite the relative stability of grain weight within a specific variety, it can be influenced by environmental factors. Higher temperatures often lead to a reduction in the grain filling period and can potentially lower grain weight. Conversely, lower temperatures tend to extend the time required for both grain filling and ripening. The ripening process may cease after a significant frost occurs. Steps in the ripening process are:

1. Milk Stage: During this stage, the starch grains within the kernel are in a soft state, and the kernel's interior is filled with a white liquid, reminiscent of milk.

2. Soft Dough Stage: At this stage, the starch within the grain is starting to solidify, although it remains relatively soft in texture.

3. Hard Dough Stage: In this phase, the entire grain achieves a firm consistency and is nearly prepared for harvesting. The moisture content still remains above 22%.

4. Maturity: The grain is now fully mature and has become hard, signaling its readiness for harvesting. This stage typically occurs when the moisture content reaches around 20 to 22%.

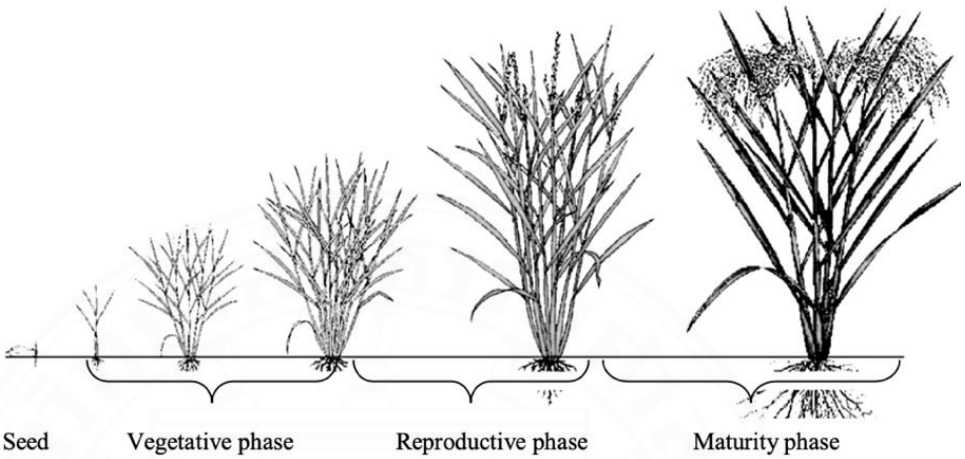

Figure 2.6 Rice crop stages (International Rice Research Institute, 2016)

2.1.4.4 Reproductive biology

Rice seeds are produced through self-pollination (Vijay and Roy, 2013). Anthesis typically commences shortly after the emergence of the panicle from the flag leaf. The flowers within the panicle open sequentially from top to bottom over the course of about one week, culminating in the completion of anthesis. Normally, the spikelet opens around 9-10 am in the morning, though this timing is influenced by the prevailing weather conditions, including temperature and humidity. Each anther contains more than 1000 pollen grains, which disperse upon spikelet opening and promptly germinate on the stigmatic surface within approximately 3 minutes after pollination. Within an hour after pollination, only one pollen tube typically reaches the ovule, initiating the processes of double fertilization and triple fusion. Notably, the maturation of both anthers and ovules occurs without any time gap relative to the spikelet. The germinability of pollen is brief, lasting only a few minutes under favorable temperature and moisture conditions. Conversely, ovules can remain viable for several days after maturation (Vijay and Roy, 2013). Oka and Morishima (1967) reported that cultivated species' pollen viability lasts for 3-5 minutes, while wild species exhibit pollen viability for up to 9 minutes.

Even though rice is basically a self-pollinated crop, natural outcrossing can emerge up to 0.5 to 4% (Oka, 2012). Moreover, Oka (2012) reported that among different ecotypes O. *sativa*, indica have higher percentage of outcrossing than japonica. The production of viable hybrids between indica and japonica has the distant relatives. It difficult to achieve mainly due to the reproductive barriers exiting between them. Their post fertilization barriers are separated into 4 types including F_1 weakness (early embryo abortion), F_1 weakness, F_1 sterility and hybrid breakdown (Oka, 2012).

2.1.4.5 spikelet fertility in rice

The spikelet fertility is the main determine of yield and quality of rice. It is associated with pollination and fertilization, which include four main stages: (i) anther dehiscence, (ii) pollen dispersal, (iii) pollen transport, (iv) pollen germination and fertilization (Hong et al., 2022; Zhou et al., 2017). Abnormal anther development induces to lower spikelet fertility in rice and reducing grain yield (Hu et al., 2021).

(i) After removal of the lemma (Figure 2.7a), the process of anther dehiscence can be separated into two steps: (step1) rupture of the whole septum, and splitting of the locule wall and splitting of the stomium, rising of stomium, (step2) the widening of the slits at the theca and the large locule, and the overflow of the pollen grains from both of the widened slits (Figure 2.7b) (Matsui et al., 1999).

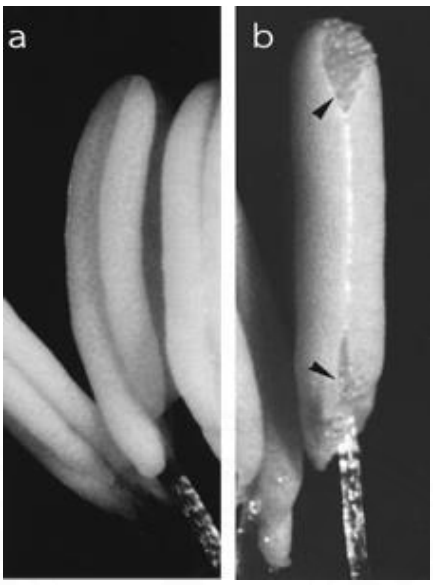

Figure 2.7 The process of anther dehiscence. (a) after removal of the lemma and (b) overflow of pollen grains from both the widened slits at the apical and basal parts (arrowheads) (Matsui et al., 1999).

(ii) Pollen dispersal occurs when pollen grains are released from the anther pores (Fahad et al., 2015). The release of pollen grains depends on factors such as the size and shape of the anther pores, as well as the viscosity of the pollen grains (Santiago & Sharkey, 2019). Moreover, pollen dispersal is one of the main ways of gene flow. Gene flow is an integrated process which includes the pollen of a specific donor diffuse to the stigma of a recipient plant by insects or wind, and then fertilizes to produce seed (Hu et al., 2021).

(iii) Pollen transport refers to the dispersion and deposition of pollen grains onto stigmas. Moreover, the number of pollen grains on the stigma is influenced by factors such as the distance and relative positions between anther pores and the stigma (Liu et al., 2015).

(iv) Pollen germination

The amount of rice pollen grain germination in *vitro*, pollen tube elongation, and embryo sac fertilization was observed with a microscope. Normally, pollination occurs when pollen grains land on stigma. A pollen tube emerges through one of the apertures of the grain, and then it grows through stigma, style, placenta, micropyle, and finally into embryo sac (Figure 2.8) (Li et al., 2013; Zeng et al., 2017).

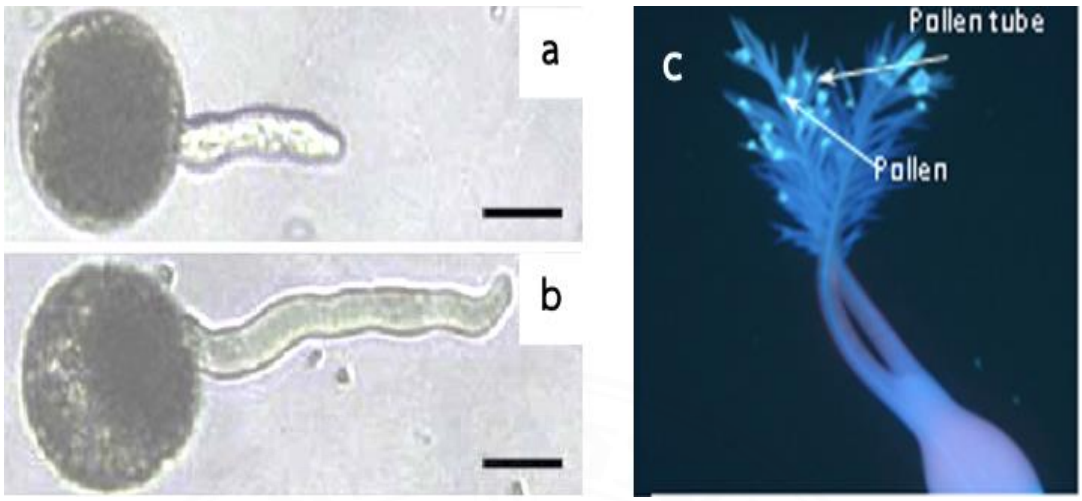

Figure 2.8 The pollen germination (a) pollen tube elongation (b) and pollen grains and pollen germination under florescent microscope (Li et al., 2013 and Zeng et al., 2017)

Chen et al. (2008) studied the growth dynamic of rice pollen tube, the growth rate and the pathway of pollen tube in stigma and ovary with a fluorescence microscopy. In their study, it was observed that rice pollen grains-initiated germination around 2 minutes after pollination. Subsequently, the pollen tube traversed the stigma and entered the style within 5 to 10 minutes. Approximately 30 minutes after pollination, the pollen tube reached the lower part of the ovary. By the 40-minute mark after pollination, only a subset of pollen tubes had successfully reached the embryo sac. The growth rates of the pollen tube within the rice stigma, style, and ovary were measured at 1500 μm/h, 5000 μm/h, and 5400 μm/h, respectively. The seed setting rate exhibited variation depending on the timing of stigma removal after pollination. When the stigma was removed within 10 to 15 minutes after pollination, the seed setting rate was notably low. However, this rate gradually increased when the stigma was removed between 20 to 50 minutes after pollination, reaching over 60% when removal occurred at 50 minutes after pollination. Eventually, the seed setting rate stabilized. It is worth noting that pollen development is influenced by various climatic factors, including rainfall, air temperature, humidity, and light intensity (Chelong & Sdoodee, 2012). These factors contribute to the dynamic nature of pollen tube growth and subsequent seed setting in rice.

2.1.5 The molecular genetic study of pollen tube growth and spikelet fertility

Rice is a good model monocot with the completed genome sequences, making it a valuable tool for studying the molecular mechanism of the spikelet fertility. Spikelet fertility is the main determine of yield and quality of rice. It has a complicated regulator controlled by numerous genes during flower formation, embryo development, and seed storage (Guo & Liu, 2012; Huang et al., 2017; Itoh et al., 2005).

In flowering plants, successful fertilization depends on the delivery of sperm to the egg (Jiang et al., 2005). The penetration of pollen tubes within the stigma is essential for guiding them towards the egg for double fertilization (Kim et al., 2023). Multiple processes and coordinated communication are crucial for transporting sperm nuclei to egg cells within the female tissues (Kim et al., 2023). These processes begin with the adhesion of pollen grains to the stigmatic tissues. The interaction between pollen and stigmatic cells stimulates the hydration and germination of pollen tubes (Jiang et al., 2005). Pollen tubes then penetrate the style and are directed towards the egg embedded in the ovule. During the apex of pollen tube growth, the pollen tube needs to synthesize a large amount of cell wall materials to maintain its structure and balance osmotic pressure (Tang et al., 2020). Several studies have revealed various factors that contribute to this maintenance process, such as lipids on pollen and stigma that act as transporters of water from stigmatic cells to pollen, cell wall-modifying enzymes involved in pollen tube elongation, and the role of reactive oxygen species in cells, which serve as signaling molecules during pollination (Fiebig et al., 2000; Kim et al., 2019; Zhang et al., 2020). Rapid alkalization factors also play a crucial role in transducing signal molecules (Kim et al., 2023). Morphological studies have additionally shown that the disruption of the maintenance process can inhibit pollen tube growth. However, the genetic and molecular mechanisms governing the growth of pollen tubes through stigmatic and stylar tissues remain poorly understood.

Pollination is the process that triggers the germination of pollen grains, leading to the formation of distinctive structures known as pollen tubes (Tran et al., 2023). Several studies have revealed that pollen tubes germinate rapidly at varying rates. In many grass species, the time of pollen germination is less than 5 minutes

(Heslop-Harrison, 1979). A developing rice pollen tube can reach the embryo sac within 40 minutes, while maize pollination allows them to reach ovules within 24 hours (Bedinger, 1992; Chen et al., 2008). This process serves as a widely employed experimental model for exploring the cellular and molecular mechanisms that underlie polar cell expansion, a crucial aspect of morphogenesis during plant development (Kim et al., 2019). In plant species such as tobacco (*Nicotiana tabacum*), lily (*Lilium longiflorum*), and tomato (*Solanum lycopersicum* L.), these are commonly used model systems for studying pollen tube growth (Conze et al., 2017; Lord, 2001; Tran et al., 2023). Similarly, species like Arabidopsis (*Arabidopsis thaliana*) serve as models for gaining insights into the regulation of pollen germination and gamete interaction (Jiang et al., 2005). However, the molecular mechanisms regulating pollen tube growth in rice (*Oryza sativa*) remain largely unknown.

2.2 Intersubspecific crosses between *indica* and *japonica* rice varieties

The Asian cultivated rice (O. *sativa* L.) is separated into two subspecies including *indica* and *japonica*. Since the 1970s, an intra-subspecific hybrid rice or *indica* hybrid rice has being used in China (Zhang, 2020). The adoption of *indica* hybrid rice cultivation practices play a crucial role in enhancing rice yields (Yuan & Virmani, 1988). In China, a significant number of *japonica* hybrid rice varieties have also been developed. However, the utilization of *japonica* hybrid rice remains limited due to its weaker heterosis. Fu et al. (2014) have reported that inter-subspecific crosses exhibit stronger heterosis compared to intra-subspecific crosses. Nonetheless, the presence of hybrid sterility between *indica* and *japonica* rice hampers the utilization of inter-subspecific heterosis. Consequently, extensive research has been conducted to overcome hybrid sterility between *indica* and *japonica* in order to harness the potential of inter-subspecific heterosis between these two rice varieties (Guo et al., 2016).

Heterosis is the vigor and producing yield of F_1 hybrids over either parent (Zhang, 2020). Intersubspecific crosses between *indica* and *japonica* rice varieties generate F_1 having higher heterosis, when compared with intraspecific crosses between *indica* and *indica* or *japonica* and *japonica* (Khush, 2001). Liu and Wu (1998) conducted a study analyzing hybrid sterility in rice using a diallel cross involving 21 parents, including *indica, japonica*, and wide compatibility varieties Their results

showed that inter-subspecific crosses (*indica* x *japonica*) enhanced strong heterosis, increasing the yield potential. However, direct application of these hybrids for high production is hindered by F_1 sterility. *S5* is a major regulator of the reproductive barrier and compatibility of *indica–japonica* hybrids in rice (Chen et al., 2018). Chen et al. (2008) cloned the *S5* gene and discovered that the *indica* (*S5i*) and *japonica* (*S5j*) alleles differed by two nucleotides, resulting in two amino acid replacements in the corresponding protein. This difference affects the sterility of hybrids. Conversely, the *S5n* allele has a loss-of-function mutation, characterized by a 136 bp DNA sequence deletion, which contributes to compatibility (Chen et al., 2008; Ji et al., 2010). This finding made it to mine new rice germplasm with the *S5n* gene using molecular markers, which showed the importance of exploring new germplasm to overcome intersubspecific hybrid sterility in rice. *S5n* had a large DNA sequence deletion, totaling 136 base pairs, in the N-terminus of the predicted S5 protein. This deletion led to subcellular mislocalization of the protein and, consequently, was assumed to render it nonfunctional. Moreover, *S5n* is a neutral allele and produces fertile offspring in combination with either *S5i* or *S5j* (Ji et al., 2010).

The heterosis in F_1 hybrids is a term used in genetics and selective breeding. When the F_1 plant is selfed to generate an F_2 population. Their descendants (F_2 generation) will vary and some F_2 hybrids will be highly homozygous. The extreme phenotypes caused by transgressive segregation are heritably stable. Transgressive segregation produces hybrid progeny phenotypes that exceed the parental phenotypes (Reddyyamini et al., 2019). The F_2 generation, characterized by maximum genetic variation, presents the initial opportunity for selecting individual plants, any of which may eventually lead to the development of a new cultivar. Moreover, there are maximum of segregation and recombination for imposing agronomic traits selection such as yield, number of panicles per plant, grain weight and seed setting rate (Figure 2.9) (Thirugnanakumar et al., 2011). Therefore, the segregation of F_2 or a later generation can be applied to breeding programs improving for new cultivars or germplasm with agronomic desirable traits (Beche et al., 2013).

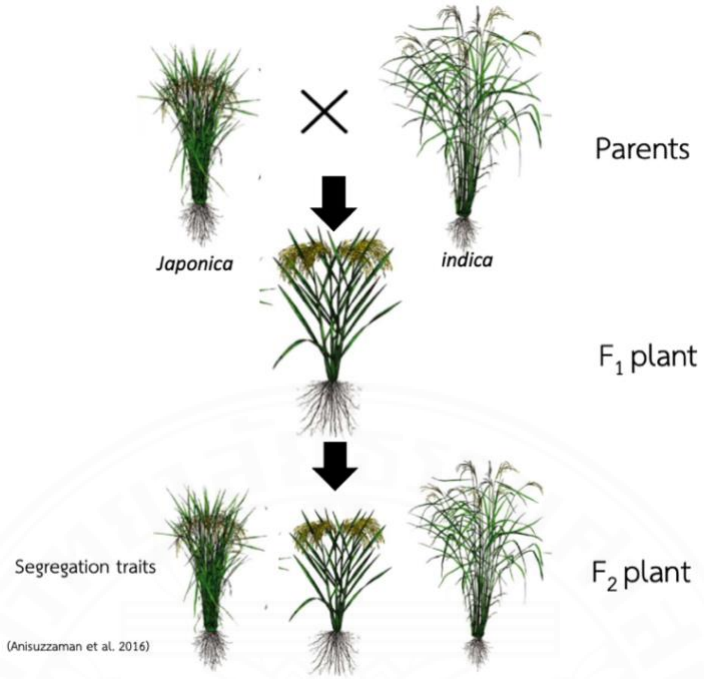

Figure 2.9 Intersubspecific crosses between *japonica* and *indica* rice

2.3 Transcription and translation

The central dogma of molecular biology describes the flow of genetic information stored in genes as DNA, transcribed into RNA, and ultimately translated into proteins (Crick, 1970). The phenotype of an organism results from the extreme expression of this genetic information, which can be modified by environmental factors. The transcription of specific genes into complementary RNA molecules not only specifies a cell's identity but also regulates its biological activities. Consequently, RNA molecules play a crucial role in interpreting the functional elements of the genome and understanding development and disease.

Transcription and translation are the mean of the cells read out, or express, the genetic instruction in their genes. The way of a process transcription is known as the transferring of the information in DNA to a messenger RNA (mRNA) molecule. During transcription, the DNA of the gene is used as a template for complementary base paring, RNA polymerase II catalyzes the formation of a pre-mRNA molecule, which is then processed to form mature mRNA (Figure 2.10). A single-stranded copy of the gene is the result of mRNA, which next must be translated into a protein molecule (Clancy & Brown, 2008).

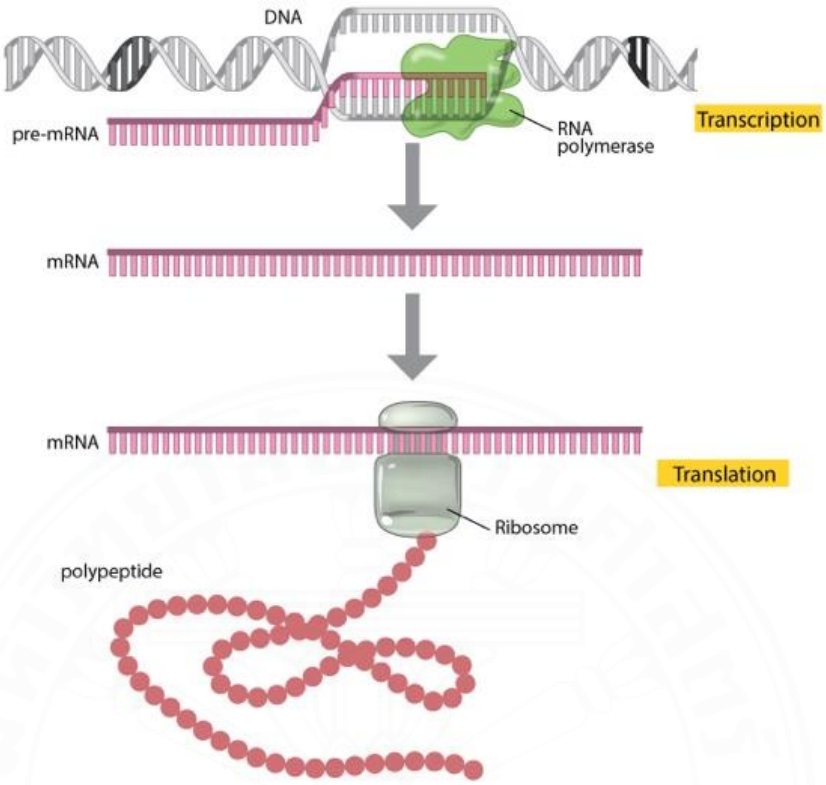

Figure 2.10 A gene is expressed through the processes of transcription and translation. During transcription, the enzyme RNA polymerase (shown in green) utilizes DNA as a template to produce a pre-mRNA transcript (shown in pink). This pre-mRNA undergoes processing to form a mature mRNA molecule, which is subsequently translated to synthesize the protein (polypeptide) encoded by the original gene (Clancy and Brown, 2008)

Segments of DNA sequences are transcribed into RNA nucleotide sequences within a cell, enabling the expression of its genetic instructions (Alberts et al., 2002). Even though the information in RNA is copied into a different chemical form, it remains essentially written in the same language as the DNA sequence. DNA and RNA are linear polymers composed of four distinct types of nucleotide subunits, interconnected through phosphodiester bonds (Figure 2.11). RNA differs from DNA chemically in two respects: 1) the nucleotides in RNA are ribonucleotides containing the sugar ribose and 2) RNA contains the base uracil (U) instead of the thymine (T) in DNA. Since U, like T, can base-pair by hydrogen-bonding with A (Figure 2.12).

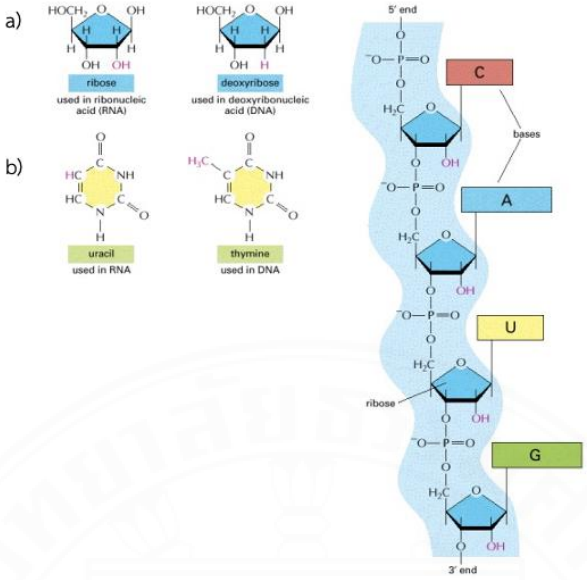

Figure 2.11 The chemical structure of RNA differs from that of DNA. RNA contains the sugar ribose, which sets it apart from DNA's deoxyribose sugar due to the inclusion of an additional -OH group (a). RNA consists of the base uracil, which distinguishes it from DNA where the equivalent base is thymine (b) (Alberts et al., 2002).

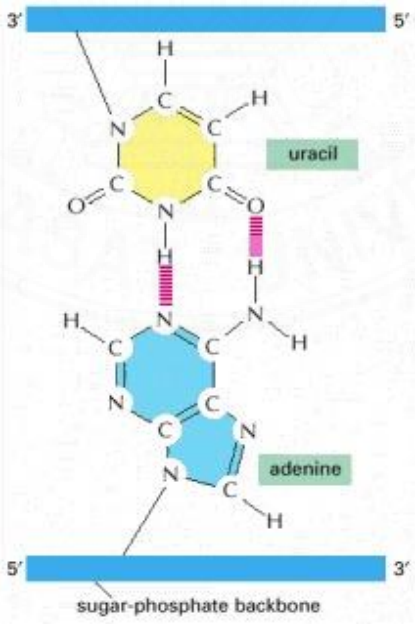

Figure 2.12 Uracil forms base pairs with adenine. The absence of a methyl group in U has no effect on base-pairing; thus, U-A base pairs closely resemble T-A base pairs.

Additionally, while DNA forms a double-stranded helix within the cell, RNA exists as a single-stranded molecule. As a result, RNA chains can fold into a diverse range of configurations, whereas only polypeptide chains fold to attain the ultimate structure of a protein. The capacity to adopt intricate three-dimensional forms enables certain RNA molecules to serve structural and catalytic roles (Alberts et al., 2002).

2.3.1 Transcription produces RNA complementary to one strand of DNA

All the RNA within a cell is generated through DNA transcription. The process of transcription commences with the unwinding of a segment of the DNA double helix, exposing the bases on each DNA strand (Alberts et al., 2002). One of the two DNA strands serves as a template for the synthesis of an RNA molecule. The determination of the nucleotide sequence of the RNA chain is achieved through the process of complementary base-pairing between incoming nucleotides and the DNA template. When a match is established, the incoming ribonucleotide becomes linked to the expanding RNA chain through an enzymatically catalyzed reaction. Thus, the RNA chain generated by transcription. Elongation occurs by adding one nucleotide at a time, and the resulting sequence is complementary to the DNA strand used as the template (Figure 2.13).

However, there are several crucial ways to differ between transcription and DNA replication. The RNA strand does not remain hydrogen-bonded to the DNA template strand. The RNA chain is displaced and the DNA helix re-forms because it behind the region where the ribonucleotides are being added. Thus, the RNA molecules produced by transcription are released from the DNA template as single strands.

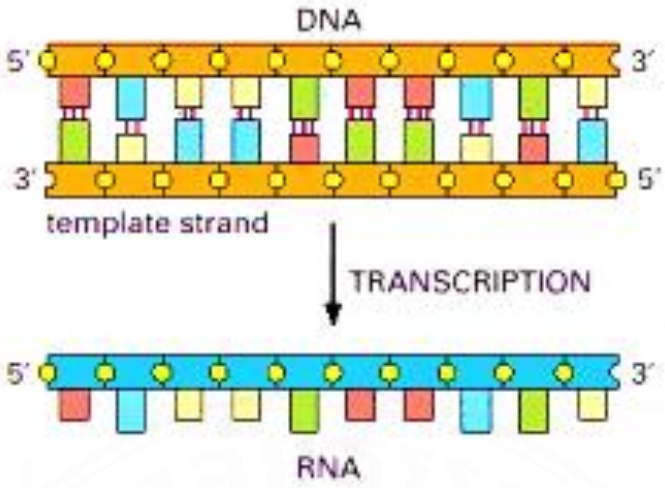

Figure 2.13 DNA transcription produces a single-stranded RNA molecule that is complementary to one strand of DNA (Alberts et al., 2002).

2.4 RNA sequencing

RNA sequencing (RNA-Seq) is a sample's mRNA sequencing approach next-generation sequencing platforms to study the entire transcriptome (Singh et al., 2012). RNA-Seq was developed more than a decade ago, and then has becomes the major tool in molecular biology for understanding of genomic function (Stark et al., 2019). Most of application of RNA-seq is analysis of differential gene expression (DGE). The standard workflow of DGE assay has not changed from the previous publications including RNA extraction, followed by fragment mRNA into short segment, complementary DNA (cDNA) synthesis, preparation of an adaptor-ligated sequencing library, perform NGS sequencing and data analysis. A read depth of 10–30 million reads per sample was used to construct the library for sequencing on a high-throughput platform.

2.4.1 Types of RNA sequencing

There are two types for RNA-seq including on the basis of the formation of cDNA and on the basis of types of RNA-seq. On the basis of the formation of cDNA is separated into two methods, which include; 1) direct RNA-Seq and 2) indirect RNA-Seq.

2.4.1.1 On the basis of the formation of cDNA

2.4.1.1.1 Direct RNA-Seq method

RNA was isolated without prior converting them into cDNA before RNAs are directly sequenced. However, RNAs are unstable as compared to DNA because they have single strand. Thus, they are difficult to handle and work with. Direct RNA sequencing was generated by the nanopore sequencing platform. The concentration of RNA around 500-700 ng of poly(A) used as input for library preparation. The RNA was ligated to a poly (T) adaptor using T4 DNA ligase, then the products of adaptor ligation were purified by RNA Clean beads. Sequencing adaptors preloaded with motor protein were then ligated onto the overhang of the previous adaptor using T4 DNA ligase, and the final library was cleaned up once more (Figure 2.14a). The RNA library was quantified using a Qubit fluorometer. The final RNA libraries were added to flow cells and carried out on the nanopore sequencing platform. Many of the research determined the nucleic acid molecule using nanopore-based sequencing for identity and sequential order of nucleotides of interest (Branton et al., 2008; Nivala et al., 2013). In the context of direct RNA sequencing using the nanopore principle, two stretches have been observed. Firstly, in the process of direct RNA sequencing, a single-stranded RNA stretch is moved through a hemolysin pore within a membrane. During this translocation, the detection of each ribobase is achieved by measuring current changes between the cis (−) and trans (+) compartments, separated by the lipid bilayer (Figure 2.14b). The alternative approach, known as single-stranded RNA exosequencing, involves the cleavage of RNA by polynucleotide phosphorylase (PNPase) (Ayub et al., 2013). In this method, every liberated ribobase can be individually read by measuring the current between the cis (−) and trans (+) compartments (Figure 2.14b) (Saliba et al., 2014).

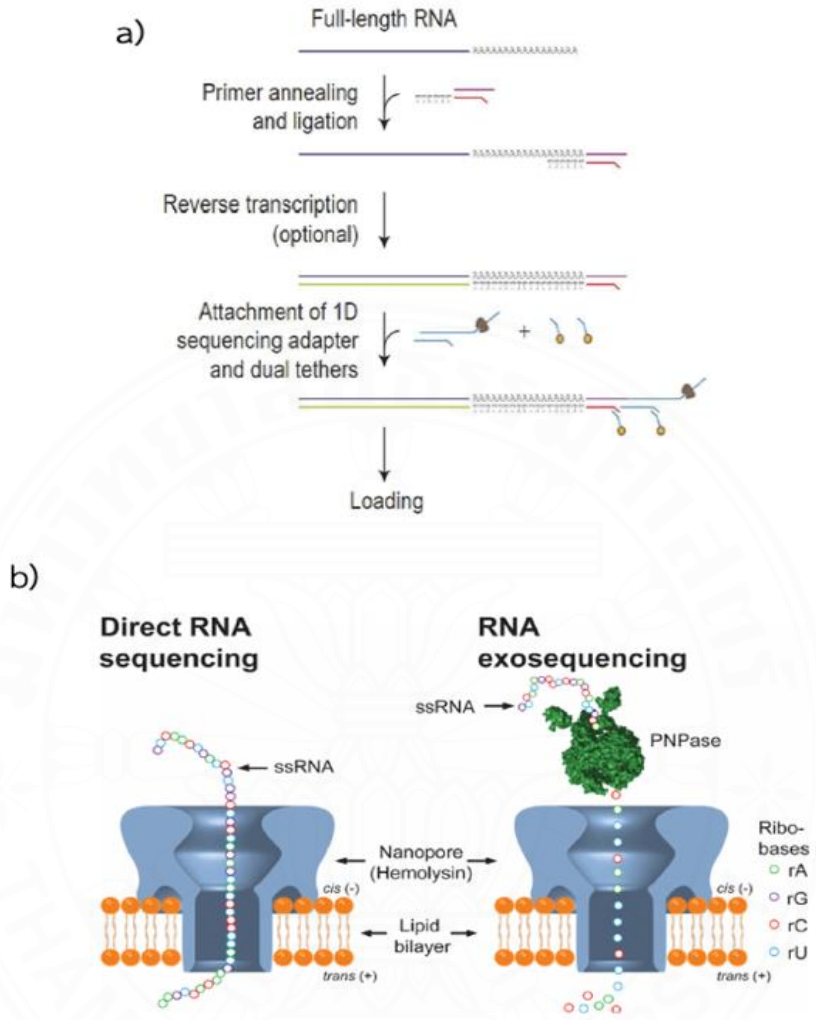

Figure 2.14 Direct RNA-Seq library preparation and Nanopore-based RNA-seq. Source: https://www.cd-genomics.com/nanopore-direct-rna-sequencing.html.

Recently, direct sequencings of RNA are used to enables the detection of RNA modifications **(**RNA methylation), gene function and characterization of full-length transcripts by Illumina HiSEquation 2000 sequencing platform and nanopore (Wongsurawat et al., 2022). In a model plant (*Arabidopsis*), direct sequencings of RNA are used for revealing the new insights into transcriptome complexity. They extracted the purified RNAs with converting them into to mRNA for each library. The mRNA for each library were used with nanopore direct RNA sequencing (Zhang, 2020).

2.4.1.1.2 Indirect RNA Sequencing

In this method, RNAs are first converted into cDNAs, then the cDNAs are sequenced. It is also known as cDNA sequencing. Moreover, the advantage of converting RNA to cDNA makes it more stable and easier to sequence. A typical RNA-Seq experiment consists of extracting RNA, selecting RNA, synthesizing cDNA, preparing libraries, sequencing and analyzing data. The first step of RNA sequencing is cell lysis and RNA extraction, then RNA is selected through several processes such as electrophoresis, filtration (size exclusion) and enzymes. Next, mRNAs are selected and converting them to cDNAs, following by optimizing of fragmentation and size selection of the sequence of cDNAs using chemical and enzymatic. Next, cDNAs adaptors are ligated at the end of fragments, then these adaptors of cDNAs are amplified to increase the concentration. The next generation sequencing (NGS) method is used to sequence the cDNAs library. Finally, the bioinformatic tools are analyzed of RNA sequence (Figure 2.15). However, several experimental details, dependent on objectives of the research, should be considered before performing RNA-Seq (Kumar et al., 2012).

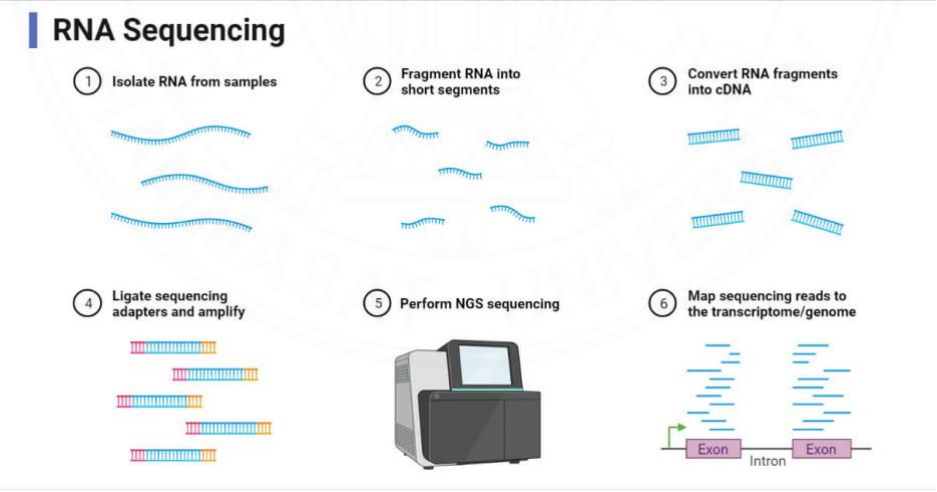

Figure 2.15 RNA-Seq

Source: https://thebiologynotes.com/RNA-sequencing-principle-steps-types-uses/

Indirect RNA sequencing is general method for analysis of gene expression and numerous specific applications, including gene expression profiling, targeted sequencing and single-cell analysis (Hrdlickova et al., 2017). Zhang et al. (2017) converted mRNA into double-stranded cDNA, then emulsion PCR was performed using the cDNA library as a template. They used the Ion PITM chip for sequencing to compared transcriptome of flower heterosis in two soybean F_1 hybrids.

2.4.1.2 On the basis of types of RNA sequenced

Generally, a typical RNA-Seq experiment consists of extracting RNA, selecting RNA, synthesizing cDNA, preparing libraries, sequencing and analyzing data (Koch et al., 2018). However, there are many types of RNA sequence such as total RNA, precursor messenger mRNA (pre-mRNA), rRNA (ribosomal RNA) and mRNA, and a lot of types of noncoding RNA (ncRNA), such as microRNA (miRNA), transfer RNA (tRNA), and long ncRNA (lncRNA; transcripts longer than 200 nucleotides not translated into protein) (Hrdlickova et al., 2017). Thus, RNA sequence types are considered prior to preparing the cDNA for sequencing, depending on research's object (Table 2.1).

Table 2.1 Common types of RNA in typical mammalian cells and the optimal RNA-Seq approaches to study them (Invitrogen, 2019).

RNA species	Biological function	Estimated percent of total RNA*	Total RNA-Seq	mRNA-Seq
Ribosomal RNA (rRNA)	Protein synthesis	80–90	– (ribosomal reduction)	– (ribosomal reduction)
Transfer RNA (tRNA)	Protein synthesis	10–15	–	–
Messenger RNA (mRNA)	Translation	3–7	Yes	Yes
Long intergenic noncoding RNA (lincRNA)	Various (cell identity and differentiation)	<0.2	Yes	No
Small nucleolar RNA (snoRNA)	Post-transcriptional modification of RNA	0.04–0.2	Yes	No
Pseudogenes	Various (transcriptional regulation)	Variable**	Yes	No
MicroRNA (miRNA)	RNA silencing, post-transcriptional regulation	0.02	Yes	No

* Percentage by mass [7].
** Pseudogene transcripts can be counted in both coding and noncoding RNA.

2.4.1.2.1 Whole Transcriptome RNA-Seq (Total RNA-Seq)

The total RNA or whole transcriptome sequencing explored sequences of all types of RNAs in the sample such as coding plus multiple forms of noncoding RNA. A total RNA-Seq experiment consists of isolating RNA, converting it to cDNA, preparing the sequencing library, and sequencing it on an NGS platform, followed by mapping of the sequencing reads to the reference genome or gene sets for gene expression analysis and polymorphism detection (Figure 2.16) (Kumar et al., 2012).

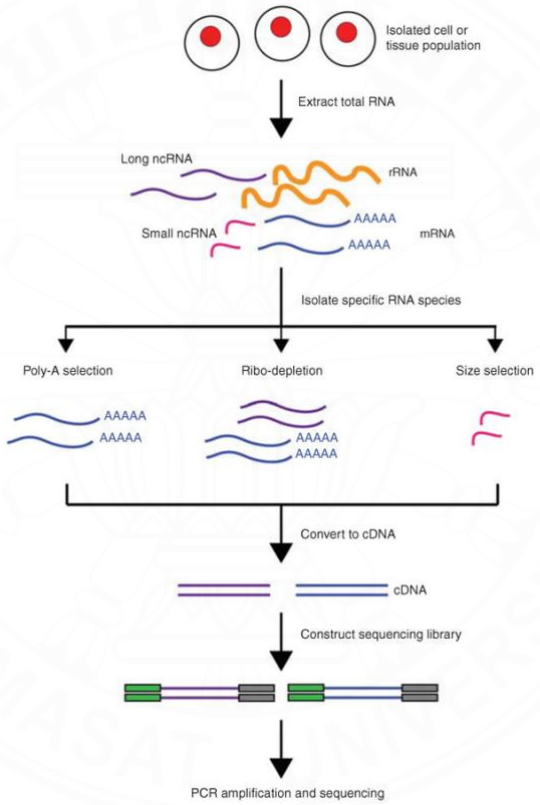

Figure 2.16 An overview of total RNA-seq involves several key steps. Initially, RNA is extracted from cells or tissues, followed by the isolation of RNA molecules using a particular protocol. This could involve the poly-A selection method to enrich for polyadenylated transcripts, or a ribodepletion protocol to eliminate ribosomal RNAs. Subsequently, the RNA is converted into cDNA through the action of reverse transcriptase enzyme, and sequencing adaptors are attached to the ends of the cDNA fragments. The RNA-Seq library is then amplified using PCR. Ultimately, the RNA-Seq library is prepared for sequencing.

2.4.1.2.2 mRNA-Seq

In this sequencing, only mRNAs are sequenced. mRNAs are first isolated from total RNA or tissue using poly-A magnetic beads and forming a poly-A library (Figure 2.16). The 3′ poly-A tail of mRNA molecules is targeted using poly-T oligos that are covalently attached to magnetic beads (Kukurba and Montgomery, 2015). Next, the single-stranded messenger RNAs (mRNAs) are selected and converted to cDNA for library preparation for sequencing. mRNA-Seq can detect quantitative profiling of transcripts in several conditions including different tissues and samples and various treatments. Furthermore, mRNA-Seq can apply to discovery of novel transcripts, alternative splicing (AS), and transcript variations (Novogene, 2023). mRNA molecules were the most frequently studied RNA species because they encoded proteins via the genetic code (Kukurba and Montgomery, 2015).

2.4.1.2.3 rRNA-Seq and tRNA-Seq

In most cell types, the majority of RNA molecules are rRNA accounting for over 80-90% of the total cellular RNA (Table 2.1). rRNA-Seq processes consist of isolation of total RNA from cell or tissue. Next, the total RNA obtained was run in agarose gel. The target rRNA was purified from an agarose gel slice. The hybridization step was required for the ribodepletion was bypassed and only the purification/concentration column module was used from the kit. The libraries were prepared following the sequenced platform (Rosselli et al., 2016). rRNA-Seq was used to identify and classification the information of 16s rRNA bacteria. Several research has been studied the structure of the 16S-rRNA genes by sequencing on SOLiD 5500xl platform (Wang & Hickey, 2002). Because the resulting from rRNA-Seq was highly stable with the known biochemical process. Furthermore, rRNA-Seq is based not only on taxon abundance but on physiological activity (Rosselli et al., 2016).

tRNAs are short, abundant molecules required for translating genetic information into protein sequences. tRNA expression is dynamically regulated in different tissues and during development that estimated 10-15% of the total cellular RNA (Table 2.1) (Dittmar et al., 2006). tRNA utilizes many properties of abundance, modification, and aminoacylation in translational regulation (Hernandez-Alias et al., 2023). Nevertheless, workflows for tRNA library generation lacking strategies to address reverse transcriptase (RT) blocks often result in predominantly short reads due to RT stops at modified sites. An example of this is seen in quantitative mature tRNA sequencing (QuantM-tRNAseq) (Pinkard et al., 2020). While hybridization-based approaches offer a workaround by eliminating the necessity for cDNA synthesis, they are only capable of distinguishing tRNAs that vary by a minimum of eight nucleotides (Dittmar et al., 2006). This limitation becomes problematic due to the extensive sequence similarity found among tRNA transcripts, which can differ by just a single nucleotide even if they encode different codons (Chan & Lowe, 2016).

2.4.1.2.4 Targeted RNA-Seq

Targeted RNA-Seq is the method of sequencing a specific transcript of interest. Targeted RNA-Seq can be achieved via either enrichment or amplicon-based approaches, both of which enable gene expression analysis in a focused set of genes of interest. Enrichment assays also provide the ability to detect both known and novel gene fusion partners in many sample types, including formalin-fixed paraffin-embedded (FFPE) tissue (illumine, 2023). To sequenced the regions of interest, hybridization capture is especial process of targeted NGS method. The biotinylated oligonucleotide baits (probes) were used to hybridize to the regions of interest. The hybridize biotinylated probes to targeted regions were captured using streptavidin beads. Finally, the targeted regions were eluted from bead, then ready to sequence (Figure 2.17).

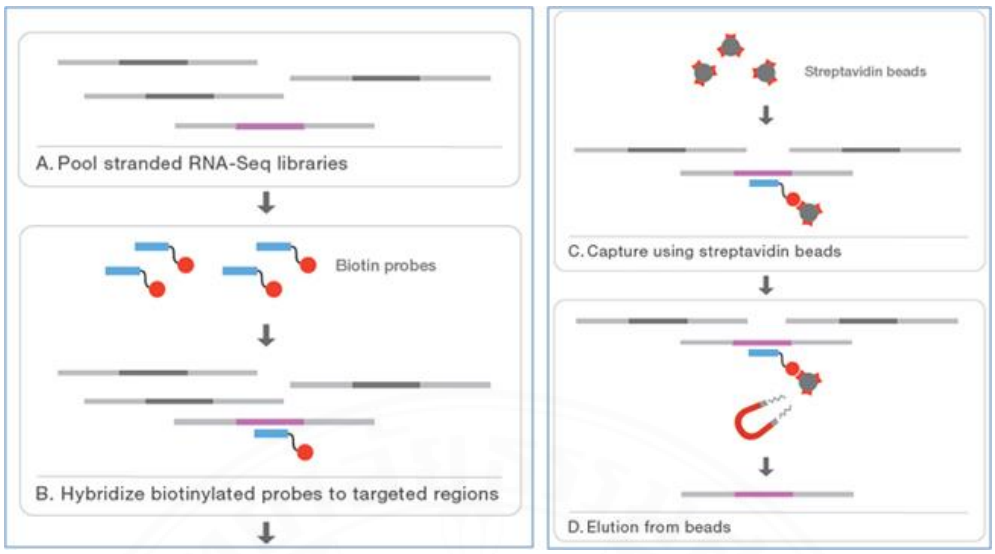

Figure 2.17 Workflow diagram of targeted RNA-Seq
Source: https://www.illumina.com/techniques/sequencing/rna sequencing/targeted-rna-seq.html

2.4.1.2.5 Small RNAs-Seq

In this sequencing type, small non-coding RNAs of a cell are sequenced. The most commonly sequenced small RNAs include miRNA, siRNA, and piwi-interacting RNA (piRNA). Because small RNAs are lowly abundant, short in length (15–30 nt), and lack polyadenylation, a separate strategy is often preferred to profile these RNA species (Morin et al., 2008). The specific protocols have been developed to selectively target small RNA species. It is the key regulators of gene expression. The selective target small RNA was generated in the library preparation process. Most kits involve isolation of small RNAs by size fractionation using gel electrophoresis. Size fractionation of small RNAs requires for running the total RNA on a gel, cutting a gel slice in the 14–30 nucleotide region, purifying the gel slice. The concentrations of small RNAs from gel slice were increased using ethanol precipitation. An alternative to gel electrophoresis is the use of silica spin columns, which bind and elute small RNAs from a silica column. After isolation of small RNAs species from total RNA, the RNA is ready for cDNA synthesis and primer ligation, ready to sequenced.

2.4.1.2.6 Single Cell RNA-Seq

In this method, RNAs extracted from a single cell line/type are sequenced. All the transcripts of a single cell are captured, transcript libraries are developed, and the whole library is sequenced. However, conventional RNA-Seq studies do not capture the transcriptomic composition of individual cells (Kukurba and Montgomery et al., 2015).

The key limiting factors in the detection of transcripts in single cells are cDNA synthesis and PCR amplification. The limits of library preparation are extremely low input RNA. Illumina's TruSeq RNA-Seq kit recommends 400 ng to 1 μg of input RNA material. Various single-cell RNA amplification methods have been developed to accommodate less input RNA (Hashimshony et al., 2012; Tang et al., 2009). Libraries for single-cell RNA sequencing (scRNA-seq) are commonly produced through a series of steps. This involves cell lysis, followed by reverse transcription into first-strand cDNA using uniquely barcoded beads. Subsequently, second-strand synthesis occurs, and the cDNA is then amplified. Hwang et al. (2018) have been developed the techniques for single-cell isolation and library preparation. The initial step in RNA sequencing is the isolation of single cells. A widely-used technique for this purpose is limiting dilution, where individual cells are isolated using pipettes by dilution (Figure 2.18a). Microscope-guided capillary pipettes are also utilized to extract single cells from suspensions (Figure 2.18b). Flow-activated cell sorting (FACS) has become a prevalent strategy for isolating highly purified single cells (Julius et al., 1972) (Figure 2.18c). Laser capture microdissection leverages a laser system assisted by a computer to isolate cells from solid samples (Figure 2.18d). Microfluidic technology for single-cell isolation has gained traction owing to its minimal sample consumption (Figure 2.18e). For instance, to isolate rare circulating tumor cells (CTCs), the Cell Search system, the first clinically validated and FDA-cleared test, was developed to enumerate CTCs in patient blood samples (Figure 2.18f).

The general steps required for generating scRNA-seq libraries encompass cell lysis, reverse transcription into first-strand cDNA, second-strand synthesis, and cDNA amplification. Cells are lysed in a hypotonic buffer, followed by poly(A)+ selection using poly (dT) primers to capture mRNAs (Figure

2.18g). It has been confirmed that due to poison sampling, only 10–20% of transcripts are reverse transcribed at this stage (Islam et al., 2014).

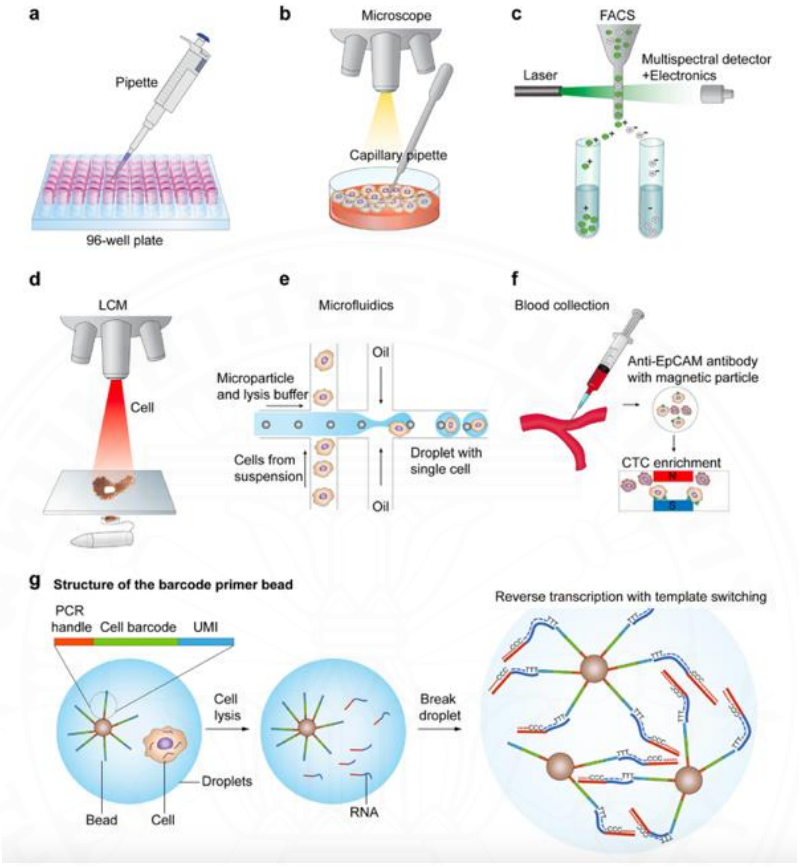

Figure 2.18 Illustrates the process of single-cell isolation and library preparation. The dilution method, shown in (a), allows for isolating individual cells. Microscope-guided capillary pipettes are employed to collect single cells, as depicted in (b). FACS, shown in (c), achieves the isolation of highly purified single cells through the utilization of fluorescent marker proteins for cell tagging. In the case of Laser Capture Microdissection (LCM) (d), a laser system guided by a computer system is used to isolate cells from solid samples. Microfluidic technology has been pivotal for single-cell isolation, as demonstrated in (e). An example is in-house microdroplet-based microfluidics, such as Drop-Seq. The Cell Search system, illustrated in (f), quantifies circulating tumor cells (CTCs) in patient blood samples using a magnet conjugated with CTC-binding antibodies. Figure (g) provides a schematic example of droplet-based library generation (Hwang et al., 2018).

2.4.2 Experimental design for RNA-Seq

To ensure the success of an RNA-seq study, it's crucial that the data generated aligns with the necessary processes for addressing the biological questions at hand. A well-designed experiment involves key decisions like selecting the appropriate library type, sequencing depth, and the number of replicates suitable for the biological system under investigation (Figure 2.19). Additionally, meticulous planning and execution of the sequencing experiment are essential to prevent the introduction of unnecessary biases into the data acquisition process (Conesa et al., 2016).

RNA extraction is a pivotal protocol employed to eliminate highly abundant rRNA, which typically constitutes around 90% of total RNA in a cell, leaving behind the remaining 1-2% that comprises mRNA (Conesa et al., 2016). In the case of eukaryotes, a choice needs to be made regarding whether to enhance mRNA content through poly (A) tail selection or to deplete rRNA. Additionally, there's the consideration of whether to create strand-preserving libraries. In the initial generation of Illumina-based RNA-seq, the random hexamer primer was used to reverse-transcribe poly(A)-selected mRNA. However, this approach failed to retain information from the DNA strand, thus complicating the analysis and quantification of antisense or overlapping transcripts.

In the cDNA synthesis step, various strand-specific protocols are available, including the dUTP method. This approach extends the original protocol by introducing UTP nucleotides before adapter ligation, followed by digestion of the strand containing dUTP (Parkhomchuk et al., 2009). In all instances, the size of the final fragments (typically less than 500 bp for Illumina) plays a crucial role in ensuring suitable sequencing and subsequent analysis. Moreover, sequencing can be carried out with either single-end (SE) or paired-end (PE) reads. Short SE reads are generally sufficient for investigating gene expression levels in well-annotated organisms, whereas longer PE reads are preferable for characterizing transcriptomes with limited annotations (Conesa et al., 2016).

Sequencing depth or library size is another pivotal consideration, denoting the number of sequenced reads for a given sample (Mortazavi et al., 2008). Research by Sims et al. (2014) revealed that around five million mapped reads are generally sufficient to accurately quantify medium to highly expressed genes in most eukaryotic

transcriptomes. However, some studies may opt for sequencing up to 100 million reads to effectively quantify genes and transcripts with low expression levels. Additionally, the number of replicates is a critical aspect of experimental design. It should be determined based on both the technical variability in RNA-seq procedures and the inherent biological variability of the system being studied. This choice significantly influences the ability to detect statistically significant differences in gene expression between experimental groups. Furthermore, the proper planning of sequencing experiments is just as important as sound experimental design, especially when dealing with a large number of samples that need to be processed in multiple batches.

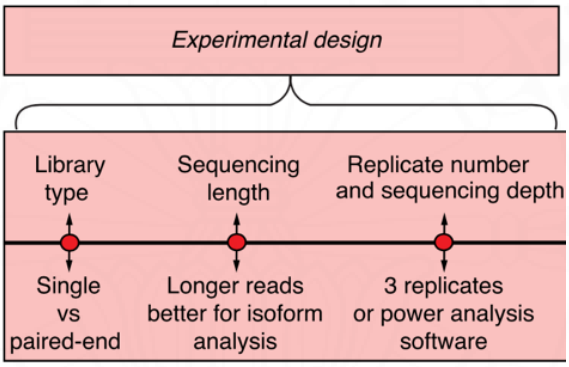

Figure 2.19 Experimental design for sequencing (Conesa et al., 2016).

2.4.3 Illumina next-generation sequencing

Next-generation sequencing (NGS) has rapidly evolved over the past 15 years and new methods are continually being commercialized (Kukurba & Montgomery, 2015). In term “NGS” has implied a next step in the development of DNA sequencing technology after the original Sanger methods has been developed (Sanger et al., 1977). Several research offered that NGS provides the ability to look at the identification of expressed genes, novel transcripts, gene expression level, gene fusion, gene mutations, post-transcriptional modifications and biological pathway changing (Schuster, 2008; Zenoni et al., 2010). Illumina® dye sequencing is one technique of NGS, an invention of Illumina company. This sequencing method is based on as known “bridge amplification” technique that is the clonal array formation and reversible dye-terminators enabling the identification of single bases that are introduced into DNA strands (Morozova & Marra, 2008). The workflow of this techniques includes three basic steps: amplify, sequence and analyze as shown in Figure 2.20. The sequencing process begins with the preparation of DNAs and the ligation them to adapters at both ends fragment. Then, the DNA templates are immobilized onto the flow cell surface. Several million dense clusters of double-stranded DNA are generated by solid-phase bridge amplification. Next, laser excitation, the fluorescent dye is captured to identify the base. The enzymatic cleaved a base out to allow competing of next nucleotide into the template. The sequencing cycles go on to determine the sequence, one base at a time. Finally, the sequence data are collected and aligned to references. This technique prefers highly accurate, covered the large-scale sequencing and rapidly sequencing data. Illumina® sequencing is widely used under many conditions of rice transcriptome study such as photoperiod, temperature, drought tolerance, salt tolerance, inter-subspecific hybrid rice (Jiang et al., 2023).

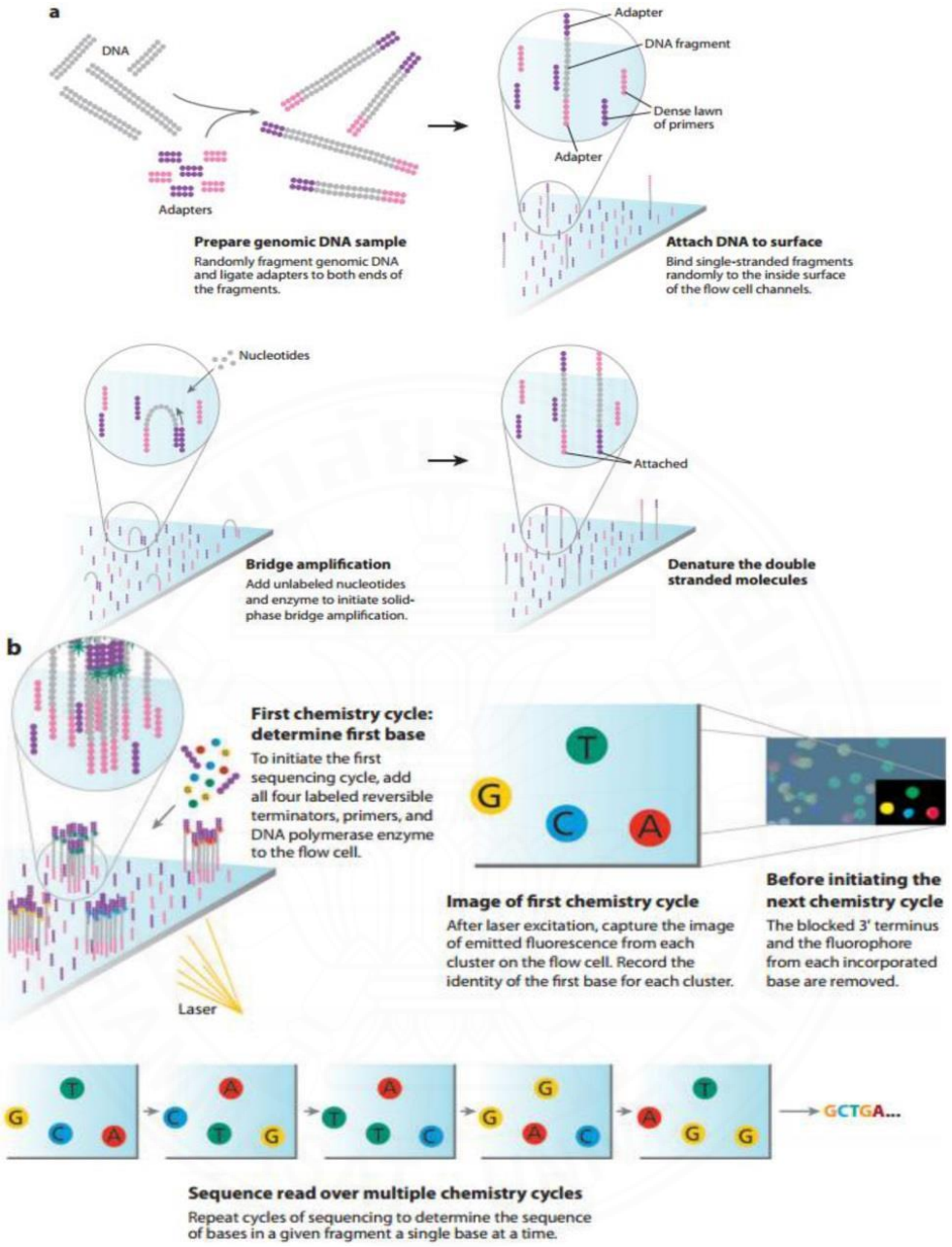

Figure 2.20 Illumina® sequencing; (a) Randomly fragmented DNA is ligated into adapters and bound to the inside surface of the flow cell channels. Unlabeled nucleotides and enzyme are added to initiate solid-phase bridge amplification and generated double-stranded DNA templates. (b) Four labeled reversible terminators nucleotides are added and incorporated into DNA template. After laser excitation, the emitted fluorescent from each cluster is captured. The complementary base is identified, the sequence data is collected and aligned to a reference (Illminar, 2020).

2.4.4 RNA-Seq data analysis

RNA-seq data analysis involves several critical steps, encompassing quality control, read alignment both with and without a reference genome, quantification of gene and transcript expression, and methods for identifying differential gene expression (Conesa et al., 2016). Finally, the results of RNA analysis are used for visualization. RNA-seq data consists of several step for quality-control checkpoints including obtaining raw reads, read alignment and quantification. Each of these steps has specific checks for monitor the quality of the data (Figure 2.21). The quality control of the data begins with the analysis of sequence quality, GC content, the presence of adaptors, the sequencing errors, PCR contaminations. Furthermore, the percentage of mapped reads is one of an important mapping quality parameter which is the overall sequencing accuracy.

During transcript identification, RNA-seq analysis typically involves mapping the reads onto the reference genome or transcriptome to determine which transcripts are expressed. This process can unveil novel transcripts that were previously unknown. Conversely, when an organism lacks a sequenced genome, the analysis pathway involves assembling reads into longer contigs (refer to Figure 2.21). When a complete reference sequence is available, two alternatives exist: mapping reads to the genome or mapping them to the annotated transcriptome. In cases where the reference genome is incomplete, RNA-seq reads can be assembled de novo into a transcriptome using bioinformatics tools such as those developed by Schulz et al. (2012) and Grabherr et al. (2011). In transcript quantification, RNA-seq is most commonly used to estimate gene and transcript expression. Raw read counts are insufficient for comparing expression levels among samples, as they are influenced by transcript length, total number of reads, and sequencing biases. The RPKM (reads per kilobase of exon model per million reads) is a normalization method that helps account for feature length and library size effects within a sample (Mortazavi et al., 2008). Moreover, frequently reported RNA-seq gene expression values include FPKM (fragments per kilobase of exon model per million mapped reads), which is a within-sample normalized transcript expression measure analogous to RPKM, and TPM (transcripts per million) (Pachter, 2011). The final step in a transcription study involves characterizing the molecular functions or pathways in which differentially expressed genes (DEGs) participate.

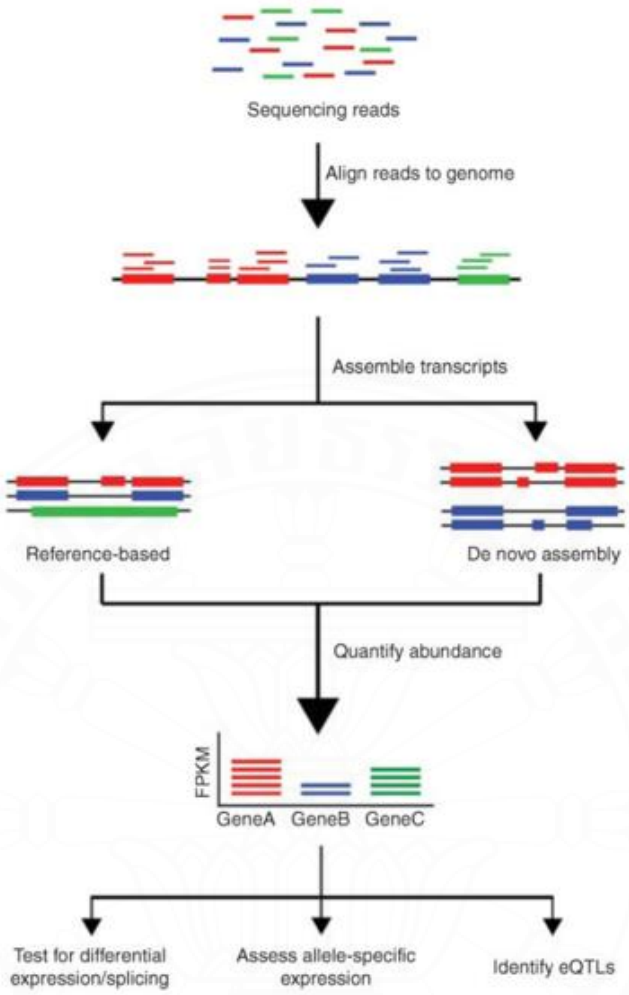

Figure 2.21 RNA-Seq data analysis overview: The RNA-Seq analysis process commences with the alignment of reads to a reference genome. Subsequently, reads can be assembled into transcripts utilizing either reference transcript annotations or de novo assembly techniques. Following this, the expression level of each gene is evaluated by quantifying the reads that align to individual exons or complete transcript sequences. Further analyses employing RNA-Seq data encompass investigating differential expression between samples, identifying allele-specific expression, and detecting expression quantitative trait loci (eQTLs) (Conesa et al., 2016).

2.4.5 Functional profiling with RNA-seq

The final step in a standard transcriptomics study is the characterization of the molecular functions or pathways in which DEGs are involved (Figure 2.21). There are two main approaches to functional characterization that were developed first for microarray technology; 1) comparing a list of DEGs with the rest of the genome for overrepresented functions, and 2) gene set enrichment analysis (GSEA), which is based on ranking the transcriptome following a measurement of differential expression. However, gene length resulted from the RNA-Seq has the biases and complicate the direct applications of these methods for count data, thus RNA-seq-specific tools have been proposed (Conesa et al., 2016).

Go seq is one of GSEA to estimate a bias effect (such as gene length) on differential expression results. Furthermore, it can adapt the traditional statistic used in the functional enrichment test to account for this bias. Differential expression from RNA-Seq can be functionally annotated using the availability of functional annotation data that is necessary for functional analysis of the transcriptome such as gene ontology, bioconductor, DAVID or babelomics (Ashburner et al., 2000). The GO database is a collection of ontologies, that describe in their biological properties. There are three aspects of GO including molecular function, cellular component and biological process (Table 2.2). Nowadays, gene ontology website (http://geneontology.org) provides the core dataset for 7,554,638 functional annotations, 43,008 GO terms, 1,519,515 gene products in 5,291 species. In addition, *Oryza sativa* is one in these core datasets.

The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways is a knowledge database for biological system of high-level functions, such as the cell, the organism and the ecosystem, from genomic and molecular-level information (Kanehisa & Goto, 2000). It is a computer representation of molecular building blocks of genomic information and chemical information that are integrated with the knowledge on molecular wiring diagrams of interaction, reaction and relation networks (systems information). The KEGG database has been in development by Kanehisa Laboratories since 1995.

Nowadays, KEGG is a reference knowledge base for integration of large-scale molecular data sets generated by genome sequencing and other high-throughput experimental technologies (https://www.genome.jp/kegg/kegg1a.html). Several

transcriptome research used GO and KEGG database to investigate the DEGs that were involved and enriched in various metabolic pathways, pathway-based analysis (L. Chen et al., 2017). They used KOBAS software to test the statistical enrichment of DEGs in KEGG pathways based on molecular-level information, particularly large-scale molecular datasets generated by genome sequencing and other high-through put experimental technologies (http://www.genome.jp/ kegg/). Therefore, the GO and KEGG database is appropriate for function profile analysis with RNA-Seq.

Table 2.2 GO system of classification

GO	**Description**
Molecular Function	Molecular function terms describe activities that occur at the molecular level, such as "catalysis" or "transport". GO molecular function terms represent activities rather than the entities (molecules or complexes).
Cellular Component	The locations relative to cellular structures in which a gene product performs a function, either cellular compartments (e.g., mitochondrion), or stable macromolecular complexes of which they are parts (e.g., the ribosome).
Biological Process	The larger processes, or 'biological programs' accomplished by multiple molecular activities. Examples of more specific terms are pyrimidine nucleobase biosynthetic process or glucose transmembrane transport. Note that a biological process is not equivalent to a pathway.

2.4.6 Integration RNA-seq with other data types

The integration of RNA-seq data with other types of genome-wide data can be connect the regulation of gene expression with specific aspects of molecular physiology and functional genomics. Expression Quantitative Trait Loci (eQTL) is the combination of RNA and DNA sequencing can be used for many objects, such as the single nucleotide polymorphism (SNP) discovery and polymorphisms in expression dynamics (Kuroha et al., 2017). Several eQTL studies have reported that the genetic variation affects the expression of most genes (Montgomery et al., 2010; Pickrell et al., 2010). RNA-seq has two major advantages for detecting eQTLs. First, it can identify factor that affect transcript processing. Second, reads that overlap heterozygous SNPs can be mapped to parental chromosomes, enabling quantification of allele-specific expression (Pastinen, 2010). The genetic effect on transcription and a number of computational methods have become available for association mapping. For example, Shabalin (2012) have focused on testing only SNPs in the cis region surrounding the gene using computational methods for mapping of eQTLs genome-wide. Moreover, the combination of RNA-seq and re-sequencing can be used to remove false positives when inferring fusion genes. In rice, eQTL analysis is often used to find genomic regions which is basic gene expression polymorphisms (Kashima et al., 2021). They identified 1,675 eQTLs leading to polymorphisms in expression dynamics under field conditions. In addition, a genomic region on chromosome 11 affects the expression of several defense-related genes in a time-of-day-and scaled-age-dependent manner. Therefore, the integration of transcriptome data and eQTL is necessary the complete set of the QTL mapping population under the given conditions.

The combination of RNA-seq and co-expression network have been widely used to identify the gene modules that regulate several conditions such as agronomic traits, salt tolerance and drought tolerance (Chutimanukul et al., 2021; Zhang et al., 2022). The gene co-expression network (GCN) is a method used in investigating the biological functions of genes using the node degree or hub centrality (Chutimanukul et al., 2021). In addition, coexpressed genes can be selected and used for further analysis, such as GO enrichment test and cis-element analysis. In rice, the Rice Functionally Related Gene Expression Network Database (RiceFREND) version 2.0 was used to analyze a gene coexpression network (http://ricefrend.dna.affrc.go.jp).

RiceFREND is a platform for identifying functionally related genes in various biological pathways and metabolic processes based on an extensive collection of microarray data derived from various tissues/organs at different stages (Khlaimongkhon et al., 2021). The coexpression data for single gene are shown in the mutual rank (MR) which is an index for coexpression as described in ATTED-II (Obayashi et al., 2018). The MR list also contains information on predicted subcellular localization sites of proteins, KEGG pathways and transcription factors (Figure 2.22a). In addition, a graphical representation of the coexpression network is shown in HyperTree (http://philogb.github.com/jit/) format (Figure 2.22b). For multiple guide gene search, two or more guide genes can be selected to initiate coexpression search (Figure 2.22c). Graphical view of the coexpressed gene network in Cytoscape Web format. The guide genes are shown as large circles with yellow outline (Figure 2.22d). The coexpression of some predicted genes will be investigated in several condition.

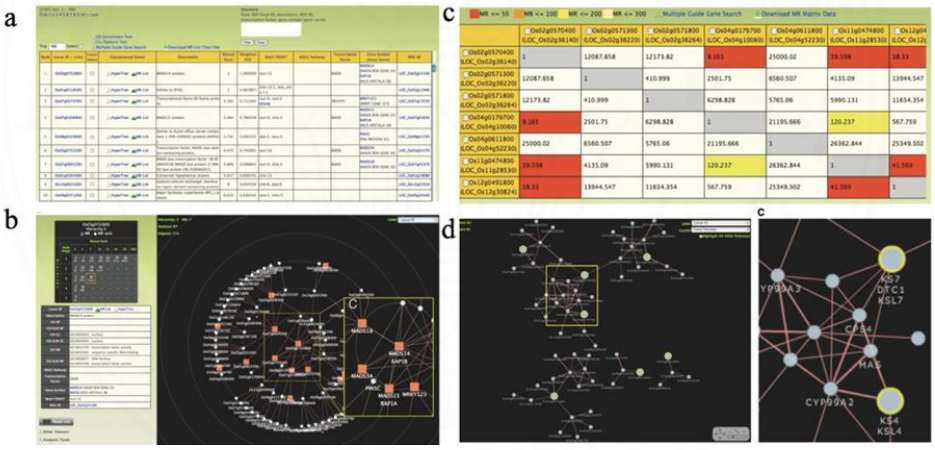

Figure 2.22 Single and Multiple guide gene search results interfaces. (a) Tabular list of coexpressed genes generated from the search with *Os03g0752800* as single guide gene. The coexpressed genes are arranged in the order of decreasing MR value. (b) Graphical view of the coexpressed gene network in HyperTree format, two nodes representing coexpressed genes with MR value below a defined threshold are joined to form an edge. Matrix data of MR values among multiple guide genes. Two or more guide genes can be selected to initiate coexpression search (c). Graphical view of the coexpressed gene network in Cytoscape Web format. The guide genes are shown as large circles with yellow outline (d)

CHAPTER 3
RESEARCH METHODOLOGY

3.1 Plant materials

F_1 to F_3 plants generated from a cross between Nipponbare (*japonica*) x KDML105 (*indica*) were developed at National Center for Genetic Engineering and Biotechnology (BIOTEC), Thailand. In this study, the selected 20 F_3 having high fertility (≥ 70%) and 43 F_3 the low (≤ 25%) fertility were grown one seedling per potted soil, for ten potted soil per line number at BIOTEC in wet season 2019. For each line, one rows, with 10 plants per row were planted. For phenotypic seed setting rate and agronomic trait observation, the rice seeds of two extreme groups of the F_4 population having high (≥ 70%) and the low (≤ 25%) fertility were soaked in water for 24 h. They were then germinated in plastic bag with moistened in the dark for three days at 30 °C. After germination, seedlings were grown in potted soil. Two-week-old plants were transferred in potted soil in a greenhouse under natural light conditions (Figure 3.1). Agronomic important traits, including high plant (cm), days to heading (days), panicle number per plant, grain number per plant, grain weight per plant (g) and spikelet fertility (%) were measured and recorded. The results were used to select F_4 plants, which was selfed to produce F_5 generation. The resulting F_5 plants were selfed to produce F_6 generation, which was used for RNA extraction. The process for population development was shown in Figure 3.2.

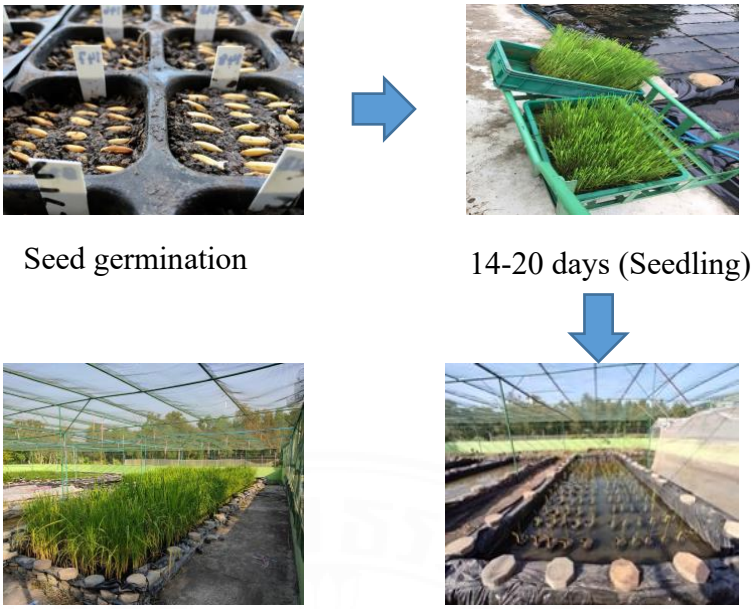

Figure 3.1 The F_4 population were planted in potted soil in a greenhouse under natural light conditions.

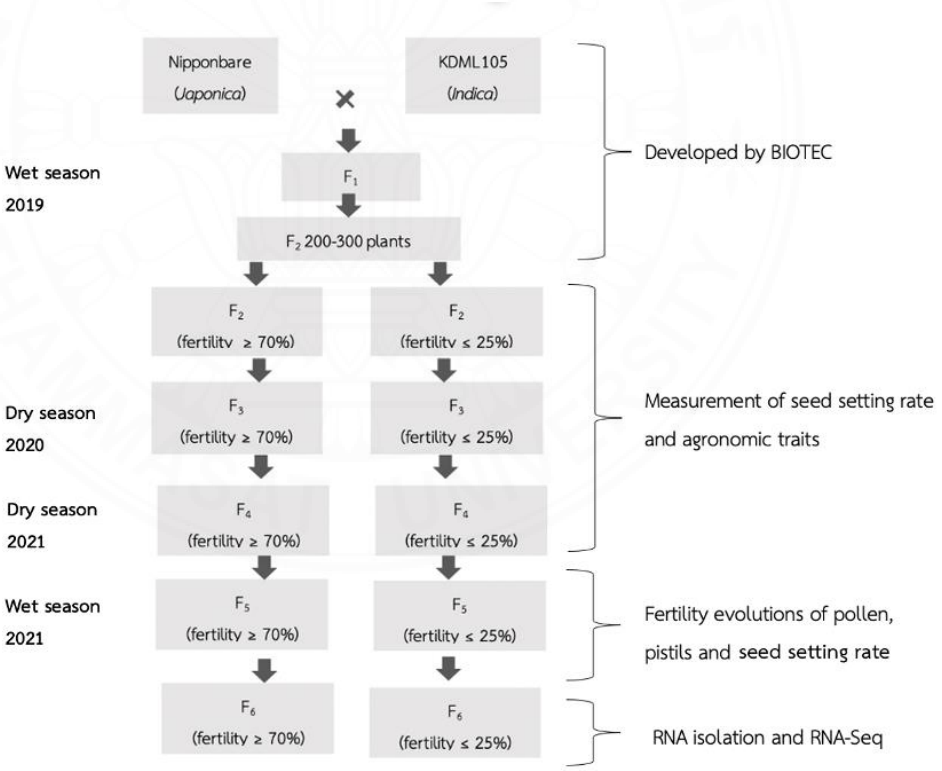

Figure 3.2 The construction of RILs (NB x KDML105) used for identification of gene involved in spikelet fertility in rice using RNA-Seq.

3.2 Measurements of agronomic important traits

For phenotypic and agronomic important traits observation, the selected 27 F_4 plants having high (≥ 70%) and 13 F_4 the low (≤ 25%) fertility were planted for measurement of agronomic important traits. Agronomic traits such as high plant (cm), days to heading (days), panicle number per plant, grain number per panicle, grain weight per plant (g) and seed setting rate were measured using the method of Institute IPGR, 2007 (Institute IPGR, 2007).

Three plants of three replication were chosen for measurements of agronomic traits. High plant was measured from soil surface to tip of the tallest panicle (awns excluded) at 75% after-ripening. Whole numbers was measured. Date on which 75% of the plant are heading, was recorded. The number of panicles per plant were recorded at early ripening stage. For grain number per panicle, one panicle were randomly selected from five plants for counting at post-harvest stage. Whole grains-dried were weigh on a precision balance at post-harvest stage. For evaluation of seed setting rate, three panicles with fully ripened grains were collected from each plant, and the numbers of filled and unfilled spikelets were counted (Wan et al., 1996). Seed setting rate was estimated by the formula:

$$\text{Spikelet fertility (\%)} = \frac{\text{Number of fertile grains}}{\text{Total number of florets}} \times 100$$

3.3 Strategy to identify tissue-specific involved in spikelet fertility

3.3.1 Fertility evaluations pistils, pollen and seed setting rate (%)

To observe the development of the anther and spikelet fertility, three panicles from three replication of HS and LS groups were tested to examine the pollen fertility using the iodine potassium iodide (I_2-KI) (Li, 1980). Ten mature florets 1 day or 2 day prior to anthesis were randomly collected from three panicles of each plant and fixed in 70% (v/v) ethanol. The anthers were broken and stained with 1% I_2-KI solution at 15 min, then rinsed with deionized water (DI water) for three time. The total number of the pollen grains was counted under light microscope. The typical type of abortion indicates pollen grains with irregular shape and no starch accumulation so that no pollen grains are stained by I_2-KI (Li, 1980). The spherical type of abortion indicates spherical shape pollen grains but smaller than normal pollen grains that I_2-KI does not stain due

to the absence of starch accumulation (Tao et al., 2004). The pistils in the spikelet were observed using an optical microscope (Kwon et al., 2015).

3.3.2 Pollen tube germination or pollination state

To evaluate the pollen germinability in *vitro*, the five-teen spikelets from five lines number of F_5 HS and LS population were marked the first day before flowering. More than 100-spikelet from three plants (pots) per lines number of HS and LS groups were collected at 8, 9, 10, 11, 12 am in R4 growth stages (Figure 3.3). The spikelets were collected again, after pollination was generated for 4 to 5 hr. The spikelets sample were collected to fixed in FAA solution (100% ethanol: acetic acid: formalin = 14:1:2) for 16 hr. followed by washing with DI water for three times (H. Chen et al., 2017). The ovaries were extracted from the spikelets for treating with 1 M NaOH for 8 hr. Then, the ovaries were stained with aniline blue (0.15 M aniline blue in 0.1 M K_2HPO_4 buffer, pH 8.2) for 10 min follower by washing with DI water. The ovaries stained with aniline blue were observed under fluorescence microscope with DAPI channel (C. Zhang et al., 2018).

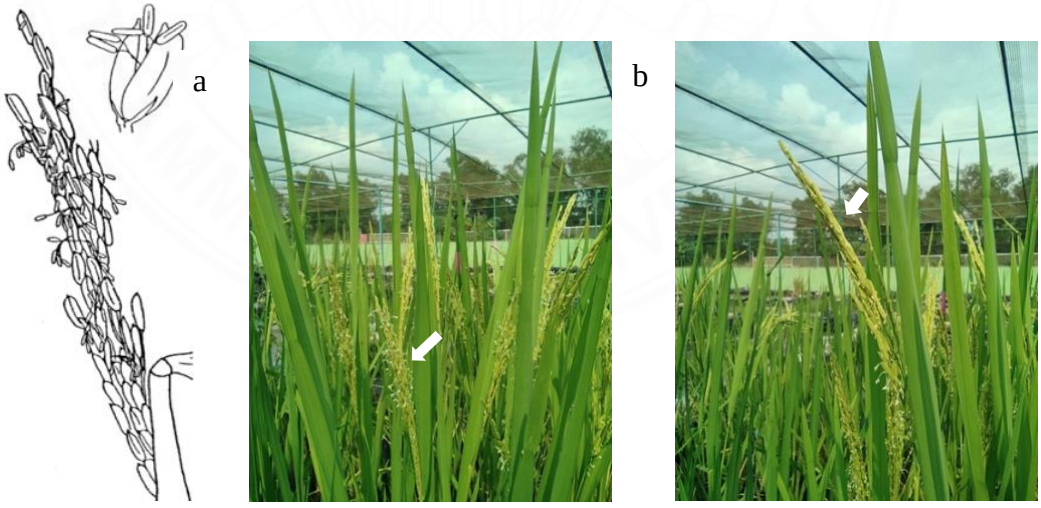

Figure 3.3 Reproductive development R4 growth stages based on discrete morphological criteria, R4 growth stages (a) and after pollination at 4-5 hr. (b).

3.4 RNA-Seq

3.4.1 Extraction and purification of total RNA

For RNA isolation, panicles of HS and LS cultivars were collected after pollination at 4-5 hr. Each sample was wrapped with aluminum foil, frozen in liquid nitrogen, and stored at −80 ◦C until use. Total RNA was extracted from 540 ovary specimens from 18 plants of each HS and LS after pollination was initiated for 4-5 hr. using RNA RNeasy Plant Mini Kit (QIAGEN, Germany). Ovaries were grinded to be powder and then homogenied them in 450 µl buffer RLT in 2 mL microcentrifuge tube. The lysate was transferred into a homogenization column, centrifuged at 14,000 rpm for 2 min. The flow-through was kept for adding 0.5 volume of ethanol (96–100%) to the cleared lysate, and mix immediately by pipetting. The tissue samples were transferred into RNeasy Mini spin column in a 2 ml collection tube, centrifuged at 13,000 rpm for 1 min, discard the flow-through. Seven hundred µl of RW1 was added in the flow-through, then centrifuged at 13,000 rpm for 15 sec, discard the flow-through. Five hundred µl RPE was added to the RNeasy spin column, centrifuged at 13,000 rpm for 15 sec, discard the flow-through. Five hundred µl RPE was added to the RNeasy spin column, centrifuged at 13,000 rpm for 2 min. Thirty to fifty µl of RNase-free water was directly added into the membrane to elute the RNA, then centrifuged at 13,000 rpm for 1 min and stored RNA at -80°C.

Total RNA of nine lines number of HS and LS were mixed and pooled as HS and LS groups, respectively (Figure 3.4). The concentration and purity of each RNA sample were determined using Nanodrop 1000 spectrophotometer (Nanodrop, Wilmington, DE, USA), with the ratio of absorbance at 260 nm and 280 nm greater than 1.8 for RNA-Seq (Table 3.1). Concentrations of RNA in 9 HS-extracted tubes ranged from 44.20 to 111.40 ng/µl, purity of DNA and RNA at A260/A280 ranged from 2.01 to 2.09. The quality of RNA sample consists 18s and 28s ribosomal band ratio, indicating purity of RNA (Figure 3.5). In addition, the concentrations of RNA in 9 LS-extracted tubes ranged from 36.90 to 115.00 ng/µl, and the purity of DNA and RNA at A260/A280 ranged from 2.03 to 2.12 and purity of RNA at A260/A230. The average concentration of NB was 34.0 ng/µl, and the purity of DNA and RNA at A260/A280 was 2.02. Finally, total RNA samples (~ 2 µg) (pooled in equal quantity from three biological replication) were sent to Novogene Co., Ltd. (Beijing, China).

Table 3.1 RNA quantification of HS and LS

No.	HS		LS	
	Conc.(ng/µl)	**A260/280**	**Conc.(ng/µl)**	**A260/280**
1	92.30	2.04	115.00	2.03
2	68.25	2.01	36.90	2.11
3	49.40	2.09	50.50	2.11
4	78.20	2.08	81.60	2.05
5	48.65	2.00	53.40	2.06
6	111.40	2.04	32.10	2.05
7	97.60	2.06	52.60	2.05
8	44.20	2.04	57.30	2.08
9	50.95	2.08	49.70	2.12

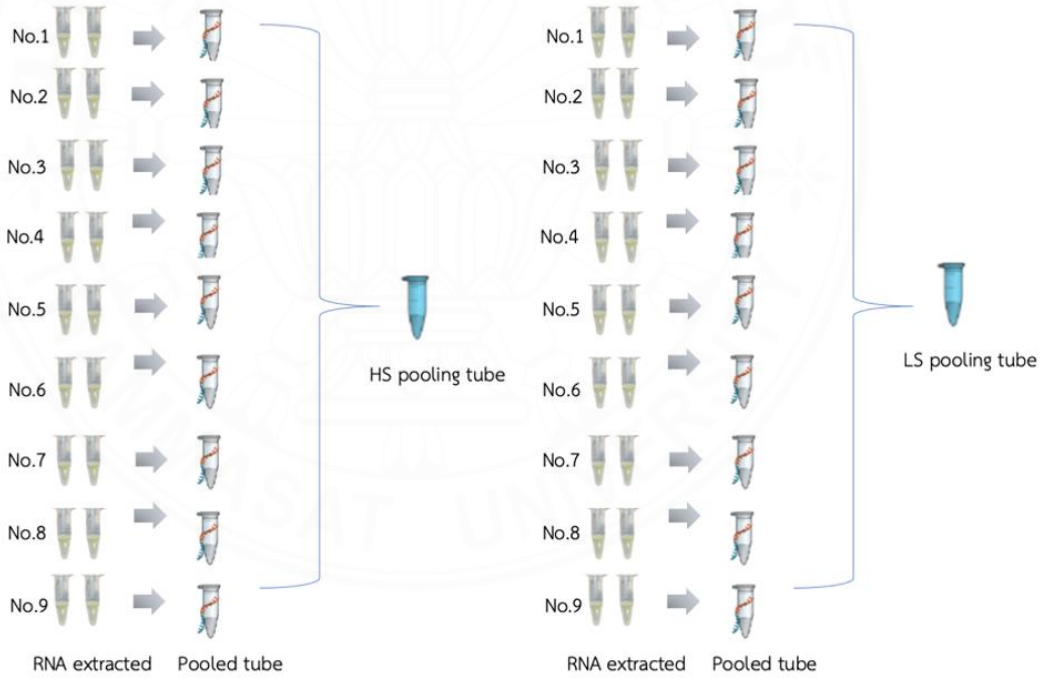

Figure 3.4 Pooling of extracted RNA of HS and LS

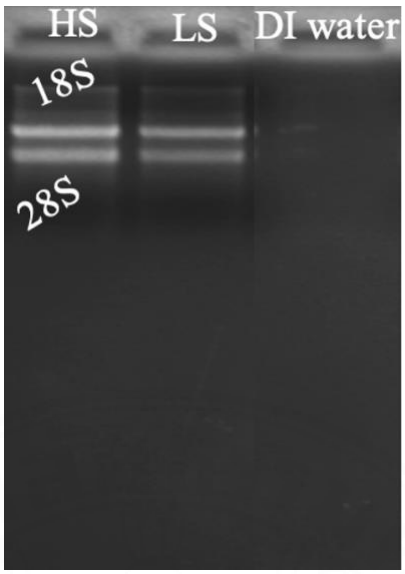

Figure 3.5 Electrophoresis analysis at 1% agarose gels for 15 min to monitor whether its degradation and contamination.

3.4.2 Sample collection and preparation by NOVOGENE

A transcriptome is a set of all the transcripts in one cell or one population of cells at certain status. Transcriptome analysis assists to study the identification of genes that are differentially expressed in distinct cell populations. Researchers can also gain a deeper insight into gene boundary identification, variable cleavage and transcript variation (Wang et al., 2009). RNA sequencing via Illumina platforms, based on the mechanism of SBS (sequencing by synthesis), offers a wide range of benefits on high throughput and high accuracy out of low sample requirements. This technical method can be a powerful tool for researching RNA transcriptional activity. RNA sequencing projects are carried out as follows figure 3.6.

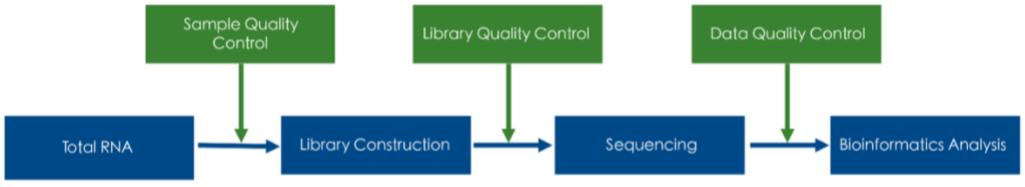

Figure 3.6 Workflow of RNA sequencing projects (Novogene, 2022)

3.4.2.1 RNA quantification and qualification

RNA integrity of total RNA of HS and LS was assessed using the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, CA, USA).

3.4.2.2 Library preparation for RNA sequencing

Total RNA of HS and LS were used as input material for the RNA sample preparations. Briefly, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. The fragment was carried out using divalent cations under high temperature in First Strand Synthesis Reaction Buffer (5X). First strand cDNA was synthesized using random hexamer primer and M-MuLV Reverse Transcriptase (RNase H-). Second strand cDNA synthesis was subsequently performed using DNA Polymerase I and RNase H. Remaining overhangs were converted into blunt ends via exonuclease/polymerase activities. After adenylation of 3’ ends of DNA fragments, adaptor with hairpin loop structure were ligated to prepare for hybridization. In order to select cDNA fragments of preferentially 370~420 bp in length, the library fragments were purified with AMPure XP system (Beckman Coulter, Beverly, USA). Then PCR was performed with Phusion High-Fidelity DNA polymerase, Universal PCR primers and Index (X) Primer. At last, PCR products were purified (AMPure XP system) and library quality was assessed on the Agilent Bioanalyzer 2100 system.

3.4.2.3 Clustering and sequencing (Novogene Experimental Department)

The clustering of the index-coded of HS and LS samples was performed on a cBot Cluster Generation System using TruSeq PE Cluster Kit v3-cBot-HS (Illumia) according to the manufacturer’s instructions. After cluster generation, the library preparations were sequenced on an Illumina Novaseq platform and 150 bp paired-end reads were generated.

3.4.3 Data Analysis

This is the workflow for mRNA sequencing data of standard bioinformatic analysis with a well-annotated reference genome, as follows figure 3.7.

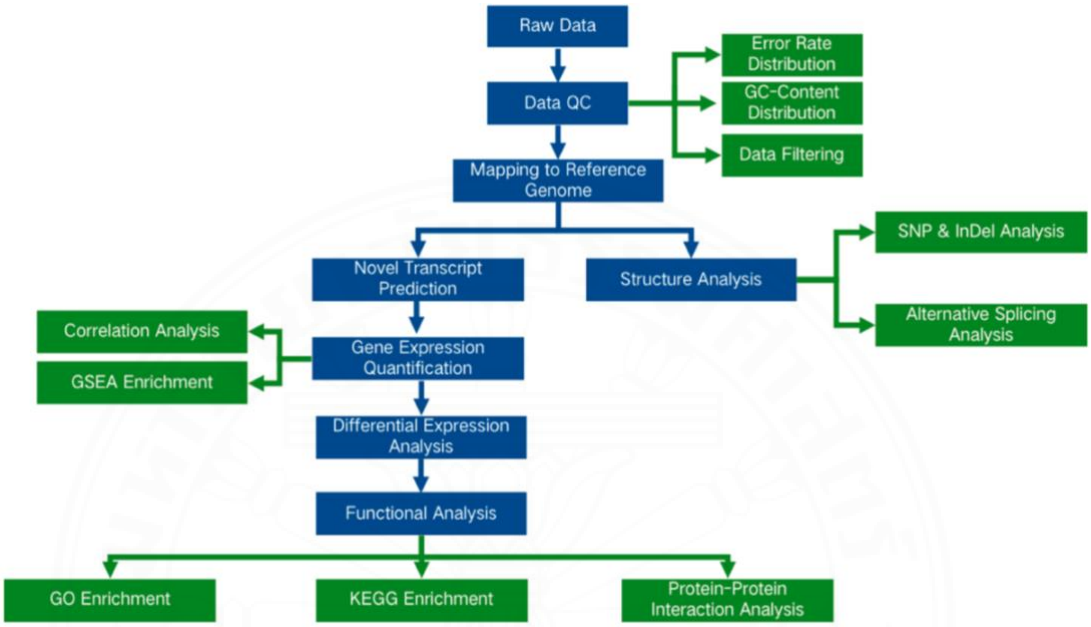

Figure 3.7 RNA-seq information analysis technology flow (Novogene, 2022)

3.4.3.1 Quality control

Raw data (raw reads) of fastq format of HS and LS were firstly processed through in-house perl scripts. In this step, clean data (clean reads) were obtained by removing reads containing adapter, reads containing ploy-N and low quality reads from raw data. At the same time, Q20, Q30 and GC content the clean data were calculated. All the downstream analyses were based on the clean data with high quality.

3.4.3.2 Reads mapping to the reference genome

Reference genome and gene model annotation files were downloaded from genome website 1 directly. In this study, we used *Oryza sativa Japonica* as the reference genome. Index of the reference genome was built using Hisat2 v2.0.5 and paired-end clean reads were aligned to the reference genome using Hisat2 v2.0.5. We selected Hisat2 as the mapping tool for that Hisat2 can generate a database of splice

junctions based on the gene model annotation file and thus a better mapping result than other non-splice mapping tools.

3.4.3.3 Quantification of gene expression level

In RNA-seq experiment, we estimate gene expression level by the abundance of transcripts (count of sequencing) that mapped to exon. Reads count is proportional gene expression level, gene length and sequencing depth. FPKM (short for the expected number of Fragments Per Kilobase of transcript sequence per Millions base pairs sequenced) is the most common method of estimating gene expression levels, which takes the effects into consideration of both sequencing depth and gene length on counting of fragments. Feature Counts v1.5.0-p3 was used to count the reads numbers mapped to each gene. And then FPKM of each gene was calculated based on the length of the gene and reads count mapped to this gene.

3.4.3.4 Differential expression analysis

The input data for differential gene expression analysis is read counts from gene expression level analysis. The differential gene expression analysis commonly contains three steps: 1) Read counts Normalization; 2) Model dependent p-value estimation and 3) FDR value estimation based on multiple hypothesis testing. The read counts of HS and LS were adjusted by edgeR program package through one scaling normalized factor. Differential expression analysis were analyzed into specific for HS and specific for LS. All sets were performed using the edgeR R package (3.22.5) in which R package analysis with parameter was seted as follows: TMM Normalized Method, the negative binomial distribution, filtering DEGs with a cut-off of log2 (FoldChange) ≥ 1, and padj ≤ 0.05 (Robinson et al., 2010). The P values were adjusted using the Benjamini & Hochberg method (Benjamini & Hochberg, 1995). Corrected P-value of 0.05 and absolute foldchange of 2 were set as the threshold for significantly differential expression. Log2 (FoldChange) was calculate using the read counts of each sample (Parnell, 2011), as follows the example:

Suppose 2 gene expression values A, B (treatment):

The read count A=10

The read count B=15

For log2-foldchange, its formula is

$$\log2\ FC = \log2\ (B) - \log2\ (A)$$

$$FC = 2^{\log2FC}$$

3.4.3.5 Functional analysis of differentially expressed genes

Gene Ontology (GO) enrichment analysis of differentially expressed genes was implemented by the clusterProfiler R package, in which gene length bias was corrected. GO terms with corrected Pvalue less than 0.05 were considered significantly enriched by differential expressed genes. KEGG is a database resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism and the ecosystem, from molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-through put experimental technologies (http://www.genome.jp/kegg/). We used clusterProfiler R package to test the statistical enrichment of differential expression genes in KEGG pathways.

3.5 Expression analysis

3.5.1 Removal of genomic DNA from RNA preparations

Prior first strand cDNA synthesis, genomic DNA was removed from RNA using DNase1 (Thermo scientific). Total RNA of ovaries of HS and LS groups cultivar were removed of genomic DNA. The reagents were mixed well and centrifuged briefly before pipetting. The mixture was prepared in a 0.2 mL microcentrifuge tube containing 1 μg of mRNA, 1 μl (1U) of 10x reaction buffer with $MgCl_2$ mix and top up to 10 μl of DEPC-treated. The mixture was incubated at 37°C for 30 min, and chilled on ice for 2 min. Then, the mixture was added with 1 μl of 50 mM EDTA and incubated at 65°C for 10 min, and ready to use the prepared RNA as a template for reverse transcriptase.

3.5.2 Complementary DNA (cDNA) synthesis

For first strand cDNA synthesis, the RNAs of HS and LS that is removed of genomic DNA, were used to be the template. Prior cDNA synthesis, the reagents were mixed well and centrifuged briefly before pipetting. RNA-primer mixture was prepared in a 0.2 mL microcentrifuge tube containing 0.1 to 5 µg of mRNA, 1 µl of primer: oligo $d(T)_{18}$ and top up to 12 µl of water, nuclease-free. The mixture was incubated at 65°C for 5 min, chilled on ice for 2 min. cDNA synthesis mix was added into each RNA - primer mixture including 4 µl of 5x Reaction Buffer, 1 µl of RiboLock RNase Inhibitor (20 U/ µl), 2 µl of dNTP Mix (10 mM) and 1 RevertAid M-MuLV RT (200 U/ µl), then mix and centrifuge briefly. The mixture was incubate for 60 min at 42°C. Finally, the reaction was terminated by heating at 70°C for 5 min. The synthesized cDNA was directly used in PCR by adding of 1 to 2 µl of the cDNA reaction mixture to a 10 µl PCR reaction.

3.5.3 RNA-Seq validation by qRT-PCR

For validation of RNA-Seq data, based on DEGs at P-value ≤ 0.05, used as the criteria for the significance of gene expression. Twenty up-regulated and Twenty down-regulated of DEGs were randomly selected for validation using RNA isolated from ovaries after pollination at 4-5 hr. of HS and LS (Table 3.1 and 3.2.). Quantitative real-time RT-PCR (qRT-PCR) was conducted using a Bio Rad CFX96 model C1000 Touch (Bio-Rad laboratories, Inc. U.S.) in a final volume of 10 µl containing 5 µl 2x qPCRBIO SyGreen Mix, 0.4 µl of the forward (10 µM) and reverse primer, 2 µl of cDNA and 2.2 µl of ddH_2O (qPCRBIO SyGreen Mix Lo-ROX, Oxford, UK). The cycling conditions were followed by PCR with 95 °C, 2 min initial denaturation; 95 °C, 5 seconds (denaturation), 58°C to 69 °C, 30 seconds (annealing) and 72 °C, 30 seconds (extension) for either 40 cycles; followed by melting curve analysis: 65 °C for 1 min. The relative gene expression was analyte using the data of comparative $2^{-\Delta\Delta}$CT method with the rice Ubiquitin 5 gene (*OsUBQ5*) used as a reference, and all of the results had two biological replicates (Schmittgen & Livak, 2008). Real-time PCR data was analyzed using the method of presenting qPCR data.

The data of comparative CT or CP method (also known as the $2^{\Delta\Delta}$CT method) was analyzed by the relative gene expression between reference gene and target gene (Schmittgen and Livak 2008). The data was analyzed by one-way analysis of variance in SPSS software. The Duncan's multiple range was tested at 0.05 level of significance (Bryman and Cramer, 2009).

$$R = 2^{-[\Delta CP\ sample - \Delta CP\ control]}$$

$$R = 2^{-\Delta\Delta}\ CP$$

Forty of DEGs-specific primers were designed using the Nipponbare reference genome obtained from GRAMENE database (http://www.gramene.org, accessed on 15 July 2022) and NCBI/Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/, accessed on 15 July 2022). All primer sequence lists used for qRT-PCR have been provided in Table 3.2.

Table. 3.2 qRT-PCR primer sequence of specific for HS and specific for LS

Gene Locus	Gene Product Name	Forward Primer	Reverse Primer
Os09g0467700	Conserved hypothetical protein	GCGTCATCTGCTGAGGGTG	CAGAAGTTCACTCCCCTGCC
Os09g0436500	Conserved hypothetical protein	CATTCTGCCCGAGCTTGGAA	GCTAATGAGATTTATGCATGAGCTT
Os09g0428900	Similar to Conserved protein	TCCACACCGGCCTAATCCAA	CCACAGCTAATGCACCGTCA
Os02g0467600	Similar to Cinnamate 4-hydroxylase CYP73	GTGGGCAACGACCTGAACC	GGCGGTGAAGATGTCGAAGA
Os03g0743650	Conserved hypothetical protein	CAGCAGCGTTTGGAGGAAG	ATGCTCCACTGCCAATCAAC
Os08g0122800	Conserved hypothetical protein	CCACTCATGGACAGCGCAAC	GCATAGTGAGCAACCTCATGCTG
Os06g0583900 (*OsPLL9*)	Similar to Pectate lyase homolog (EC 4.2.2.2) (Fragment)	CTCCTTCAGCGAGGACCAGA	TTGCTCCCTCCGATGGCATA
Os12g0563700 (*OsPME35*)	Pectin lyase fold domain containing protein	TGCACGATTGAGTCGGTGAC	AGGAGTAAACCACCCGGGAC
Os10g0204300	Similar to Phosphoenolpyruvate carboxykinase	CAAAATTCTGGCGAGAACAAGTTGA	AGCCAGATCTCTAACACCACCT
Os10g0163340	Similar to Cytosolic aldolase	TCGATGGCCCTCACGACATC	CACCTTCTTGGCGTCGGAAC

3.6 Statistical analysis

The experiment of F_4 to F_6 population design was a randomized complete block design with three replicates. Data of length of panicles, tiller number, panicle number, number of filled grains and unfilled grains per plant, number of filled grains and unfilled grains per panicle, filled grains and unfilled grains weight per panicle and 100 grains weight per plant were subjected to one-way analysis in SPSS software. The Duncan's multiple range was tested at 0.05 level of significance.

3.7 Co-expression analysis

The Rice Functionally Related Gene Expression Network Database (RiceFREND) version 2.0 was used to analyze a gene coexpression network (http://ricefrend.dna.affrc.go.jp, accessed on 30 April 2022). RiceFREND is a platform for identifying functionally related genes in various biological pathways and metabolic processes based on an extensive collection of microarray data derived from various tissues/organs at different stages (Khlaimongkhon et al., 2021; Sato et al., 2010). All data sets were combined into one gene expression matrix data for coexpression analysis among genes. The weighted Pearson's correlation coefficients were calculated to reduce any unsuitable effects. The mutual rank (MR) was used as an index for coexpression as described in ATTED-II (Obayashi & Kinoshita, 2009). In addition, the expression of gene at vary state was confirm by qRT-PCR. All primer sequence lists used for qRT-PCR have been provided in Table 3.2.

3.8 Chromosomal location of candidate genes

A total of twenty-one QTLs and six functional genes for seed setting rate, spikelet fertility and yield component (i.e., filled grain number, filled grain percentage, grain number per panicle, grain yield and grain yield per panicle) have been identified, according to the Gramene QTL database (http://archive.gramene.org/). We located significantly 8 and 3 DEGs specific HS and specific LS, respectively, from previous KEGG analysis on these QTL intervals. MG2C (http://mg2c.iask.in/mg2c_v2.1/) was used for mapping the candidate genes according to their relative distances and chromosomal positions with the base pair (bp) unit (Chao et al., 2021; Jiangtao et al., 2015). Two inputs for chromosomal location setting were as follows: input1; gene location including five fields; 1) gene id, 2) gene start position, 3) gene end position, 4) chromosome id, and 5) gene color label, and input 2; chromosome length including two fields; 1) chromosome id and 2) chromosome length.

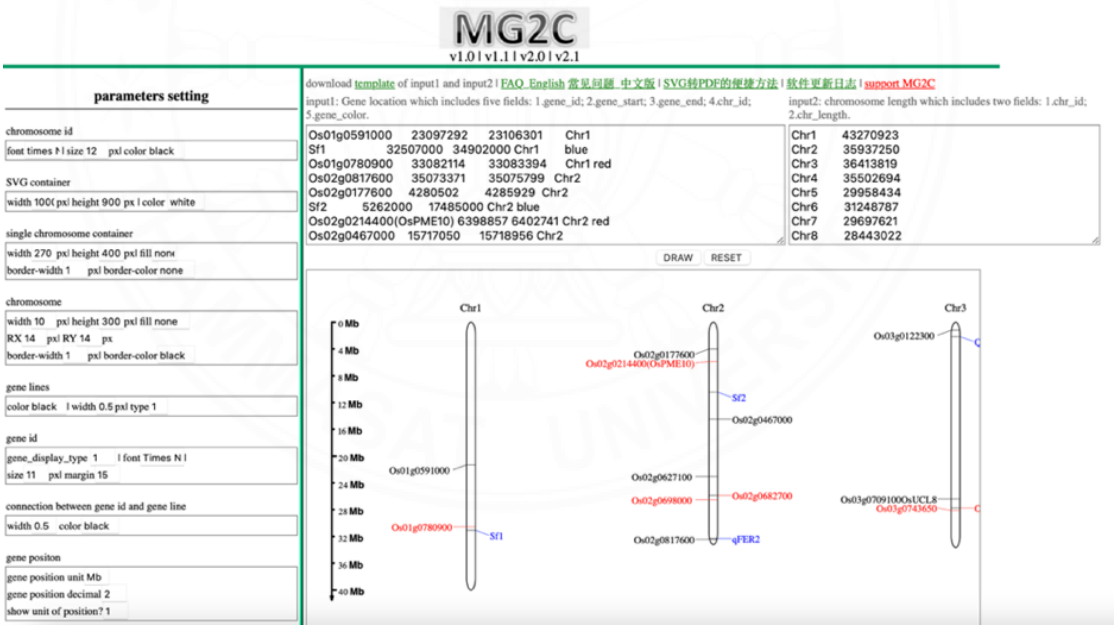

Figure 3.8 Parameters setting for chromosomal location of candidate genes

CHAPTER 4
RESULTS AND DISCUSSION

4.1 Construction of recombinant inbred lines and fertility evaluation of parents and F_{2-5} population

Two extreme groups of the F_2 to F_5 population having the low (≤ 25%) and high (≥ 70%) fertility were planted to construction of RILs (NB x KDML105) for identification of gene involved in spikelet fertility in rice using RNA-Seq. The result obtained by F_2 to F_5 plants having low (≤ 25%) fertility were 30, 43, 23 and 10 plants, respectively. In addition, the extreme groups of the high fertility (≥ 70%) in F_2 to F_5 population were 57, 20, 26 and 11 plants, respectively (Figure 4.1 to 4.4). Thus, in this study, we have ten and eleven plants of F_5 having low and high fertility, respectively, for planting to F_6 generation, which were used for RNA-Seq. In addition, the different significance of phenotypic characterization of HS and LS are shown in the figure 4.5.

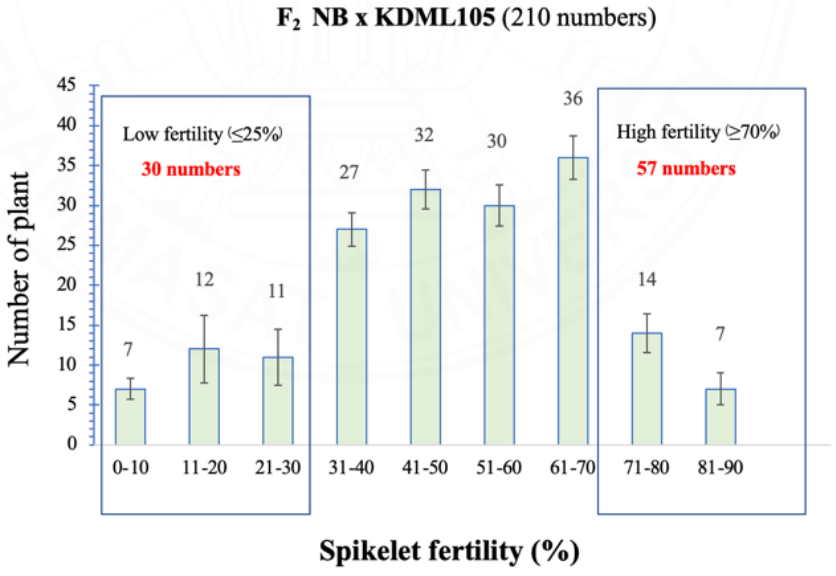

Figure 4.1 Number of plants in extreme groups of the F_2 population having the low and the high fertility

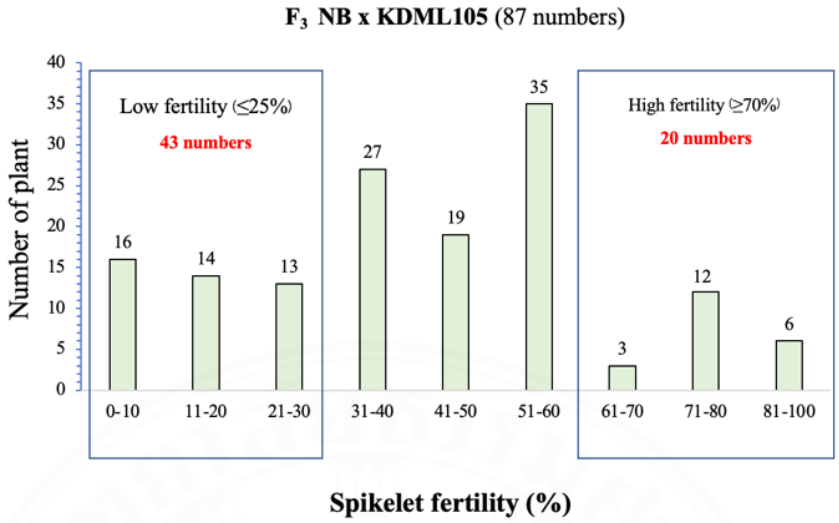

Figure 4.2 Number of plants in extreme groups of the F_3 population having the low and the high fertility

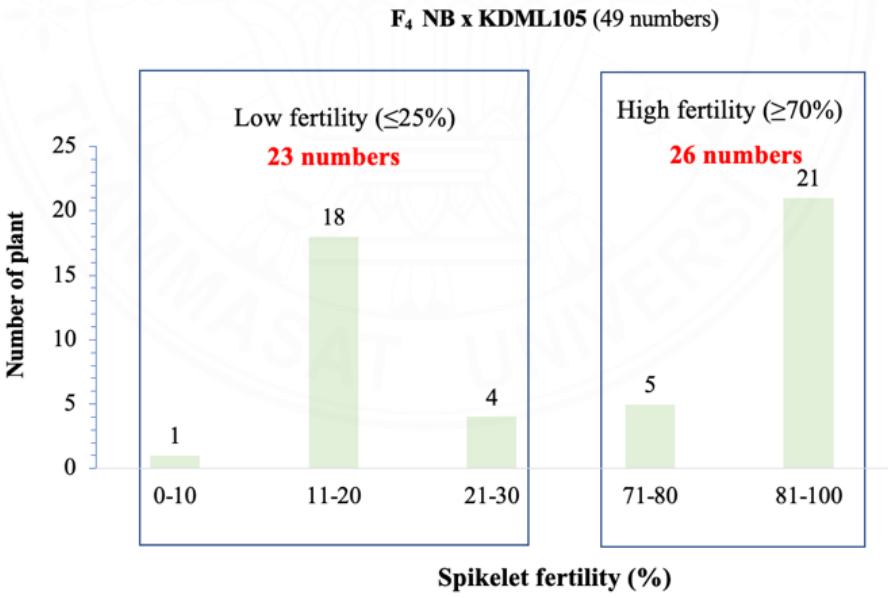

Figure 4.3 Number of plants in extreme groups of the F_4 population having the low and the high fertility

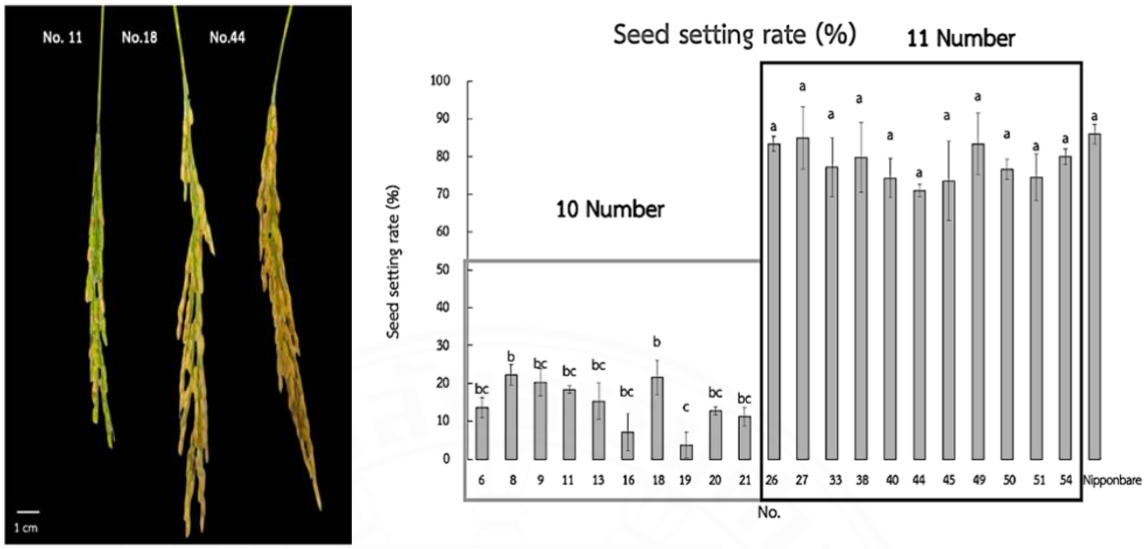

Figure 4.4 Seed setting rate in extreme groups of the F_5 population having the low and the high fertility

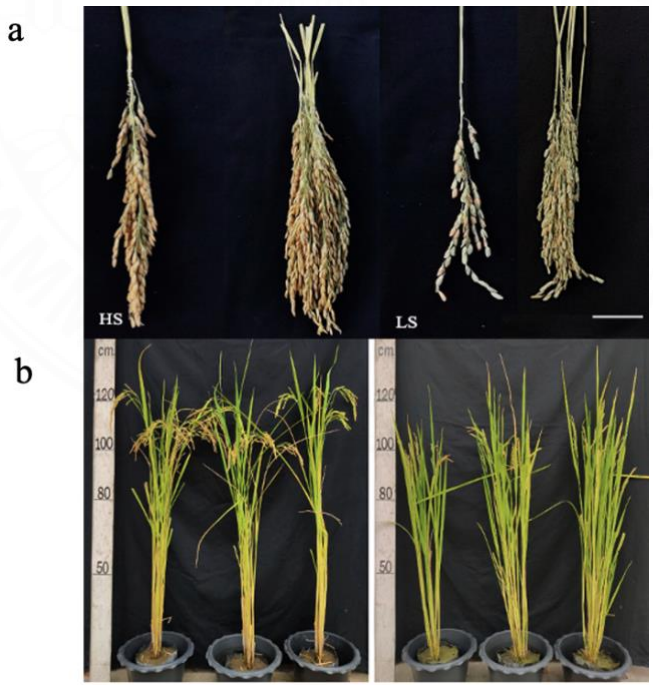

Figure 4.5 Phenotypic characterization of F_5 HS and LS lines. Mature panicles and panicle per plant (a) and plant morphology of HS and LS lines at mature stage (b).

4.2 Measurements of agronomic important traits

The average of agronomic important traits such as high plant (cm), days to heading (days), panicle number per plant, grain number per panicle and seed setting rate of 27 F_4 plants having high and 13 plants and the low (≤ 25%) fertility were shown in table 4.1 and 4.2, respectively. Cross NB x KDML105, average of plant height of HS ranged between 85 to 125 cm, while mean of plant height of LS was 80 to 113 cm. The parents NB and KDML105 recorded with 92 and 102 cm, respectively. The 70% day to heading values varied from 123 to 142 days with a mean of 131 days in HS. However, there are significant differences for 70% day to heading some in some line number when compare to the parents. The line number 27 had significant differences for 70% day to heading with NB. In addition, the line number 25 and 26 had significant differences for 70% day to heading with KDML105. In LS group, the 70% day to heading values from 119 to 134 days in LS. Only line number 12 had significant differences for 70% day to heading when compare to the parents. The 70% day to heading values had not different between HS and LS. The parental lines exhibited significant differences for 70% day to heading, which was 142 days and 113 days in NB and KDML105, respectively. The HS, LS and the parental lines had not significant difference of the number of panicles per plant. In addition, the number of panicles per plant HS ranged between 6 to 15 panicles, while LS ranged between 7 to 15 panicles. The parents NB and KDML105 recorded with 5 and 11 cm, respectively. The average of number of grains per panicle was significantly highest in line number 22 and 23 under HS group when compare to LS and the parental lines. Their number of grains per panicle were 174 and 169 grains. However, the LS had not significant difference both of each line number and the parental lines. The parental lines also had not significant difference for number of grains per panicle between NB and KDML105. The comparation of the percentage of seed setting rate between HS and LS was shown that there are significantly highest in HS (75-94%), while LS had low seed setting rate (6-22%). In addition, the parental lines had not significant difference in both cultivars, which was 90% in NB and 81% in KDML105.

Table 4.1 Phenotyping of agronomic important traits of extreme groups of F_4 plants having high fertility (≥ 70%) and its parent

No.	Plant height (cm)	70% Day to heading (day)	No. of panicle/plant (panicle)	No. grain/panicle (grain)	Seed setting rate (%)
1	89 ± 0.67 ^{abc}	134 ± 0.00 $^{a-d}$	9 ± 0.50 ^{ab}	111 ± 1.33 $^{c-l}$	88 ± 1.23 ^{ab}
2	117 ± 8.82 ^{ab}	134 ± 0.00 $^{a-d}$	8 ± 0.33 ^{ab}	125 ± 0.00 $^{a-i}$	80 ± 3.24 ^{ab}
3	100 ± 0.00 ^{abc}	130 ± 0.00 ^{bcd}	11 ± 2.91 ^{ab}	109 ± 5.36 $^{c-l}$	76 ± 2.68 ^{b}
4	110 ± 2.91 ^{abc}	132 ± 0.00 $^{a-d}$	6 ± 0.58 ^{ab}	92 ± 2.89 $^{f-m}$	77 ± 2.19 ^{ab}
5	118 ± 1.67 ^{bc}	128 ± 0.00 ^{cde}	12 ± 1.86 ^{ab}	90 ± 6.00 $^{g-m}$	91 ± 1.55 ^{ab}
6	118 ± 4.41 ^{bc}	132 ± 0.00 $^{a-d}$	13 ± 1.20 ^{ab}	99 ± 8.00 $^{d-m}$	80 ± 2.46 ^{ab}
7	112 ± 1.67 ^{abc}	139 ± 0.00 ^{ab}	11 ± 6.00 ^{ab}	68 ± 14.19 ^{klm}	88 ± 0.93 ^{ab}
8	125 ± 7.64 ^{a}	132 ± 0.00 ^{ab}	8 ± 0.33 ^{ab}	63 ± 3.00 ^{lm}	78 ± 2.91 ^{ab}
9	105 ± 10.00 ^{abc}	139 ± 0.00 $^{a-d}$	13 ± 2.65 ^{ab}	82 ± 0.50 $^{h-m}$	80 ± 2.20 ^{ab}
10	118 ± 3.33 ^{ab}	129 ± 0.00 $^{b-d}$	8 ± 0.50 ^{ab}	94 ± 4.00 $^{e-m}$	81 ± 2.70 ^{ab}
11	120 ± 10.41 ^{ab}	129 ± 0.00 $^{b-d}$	9 ± 0.50 ^{ab}	116 ± 8.00 $^{c-k}$	78 ± 4.88 ^{ab}
12	95 ± 8.66 ^{abc}	125 ± 2.00 ^{de}	15 ± 2.00 ^{a}	102 ± 9.50 $^{c-l}$	85 ± 3.56 ^{ab}
13	95 ± 12.58 ^{abc}	129 ± 3.00 $^{b-d}$	15 ± 4.00 ^{a}	49 ± 1.50 ^{m}	94 ± 3.57 ^{a}
14	107 ± 7.00 ^{abc}	132 ± 0.00 $^{a-d}$	10 ± 2.40 ^{ab}	87 ± 7.00 $^{h-m}$	82 ± 4.65 ^{ab}
15	102 ± 7.26 ^{abc}	130 ± 1.00 $^{a-d}$	10 ± 2.19 ^{ab}	93 ± 0.00 $^{e-m}$	81 ± 2.21 ^{ab}
16	105 ± 2.89 ^{abc}	131 ± 1.00 ^{bcd}	11 ± 2.67 ^{ab}	128 ± 16.15 $^{a-h}$	80 ± 4.92 ^{ab}
17	85 ± 10.00 ^{bc}	123 ± 4.33 ^{bcd}	12 ± 0.67 ^{ab}	103 ± 3.50 $^{c-l}$	82 ± 1.42 ^{ab}
18	95 ± 11.55 ^{abc}	126 ± 3.33 ^{cde}	12 ± 3.18 ^{ab}	111 ± 11.00 $^{c-l}$	83 ± 0.68 ^{ab}

Different lowercase letters within the same column show significant differences at the 0.05 level. N = 3

Table 4.1 Phenotyping of agronomic important traits of extreme groups of F_4 plants having high fertility (≥ 70%) and its parent

No.	Plant height (cm)	70% Day to heading (day)	No. of panicle/plant (panicle)	No. grain/panicle (grain)	Seed setting rate (%)
19	117 ± 9.28 ^{ab}	130 ± 0.33 ^{bcd}	11 ± 0.88 ^{ab}	151 ± 12.00 ^{abc}	82 ± 4.94 ^{ab}
20	118 ± 10.11 ^{ab}	136 ± 3.00 ^{abc}	11 ± 1.20 ^{ab}	149 ± 16.00 $^{a-d}$	79 ± 2.30 ^{ab}
21	115 ± 11.55 ^{abc}	130 ± 4.67 ^{bcd}	8 ± 1.00 ^{ab}	78 ± 17.00 $^{h-m}$	75 ± 3.95 ^{b}
22	98 ± 3.33 ^{abc}	125 ± 0.00 ^{de}	8 ± 1.45 ^{ab}	174 ± 23.50 ^{a}	78 ± 3.66 ^{ab}
23	93 ± 1.67 ^{abc}	125 ± 0.00 ^{de}	10 ± 0.33 ^{ab}	169 ± 8.00 ^{ab}	77 ± 3.15 ^{ab}
24	103 ± 4.41 ^{abc}	136 ± 5.67 ^{abc}	12 ± 0.00 ^{ab}	91 ± 0.50 $^{g-m}$	84 ± 6.57 ^{ab}
25	102 ± 1.67 ^{abc}	142 ± 0.00 ^{a}	11 ± 0.00 ^{ab}	104 ± 8.00 $^{c-l}$	85 ± 2.40 ^{ab}
26	97 ± 3.33 ^{abc}	142 ± 0.00 ^{a}	5 ± 0.33 ^{a}	104 ± 8.00 $^{c-l}$	90 ± 2.67 ^{ab}
27	102 ± 3.33 ^{abc}	113 ± 0.67 ^{f}	11 ± 1.45 ^{ab}	115 ± 5.61 $^{c-l}$	81 ± 4.54 ^{ab}
Nipponbare	97 ± 3.33 ^{abc}	142 ± 0.00 ^{a}	5 ± 0.33 ^{b}	118 ± 5.50 $^{c-l}$	90 ± 2.67 ^{ab}
KDML105	102 ± 3.33 ^{abc}	113 ± 0.67 ^{f}	11 ± 1.45 ^{ab}	111 ± 6.50 $^{c-l}$	81 ± 4.54 ^{ab}

Different lowercase letters within the same column show significant differences at the 0.05 level. N = 3

Table 4.2 Phenotyping of agronomic important traits of extreme groups of F_4 plants having the low fertility (≤ 25%) and its parent

No.	Plant height (cm)	70% Day to heading (day)	No. of panicle/plant (panicle)	No. grain/panicle (grain)	Seed setting rate (%)
1	93 ± 3.33 ^{abc}	134 ± 0.00 $^{a-d}$	11 ± 1.45 ^{ab}	97 ± 6.00 $^{e-m}$	18 ± 0.15 ^{c}
2	98 ± 2.50 ^{abc}	134 ± 0.00 $^{a-d}$	14 ± 1.00 ^{ab}	71 ± 1.50 $^{j-m}$	22 ± 3.65 ^{c}
3	93 ± 6.01 ^{abc}	131 ± 1.33 ^{bcd}	10 ± 1.20 ^{ab}	145 ± 7.21 $^{a-e}$	8 ± 2.18 ^{c}
4	95 ± 5.00 ^{abc}	131 ± 1.00 ^{bcd}	9 ± 0.50 ^{ab}	126 ± 29.00 $^{a-h}$	22 ± 3.55 ^{c}
5	80 ± 0.00 ^{c}	132 ± 0.00 $^{a-d}$	12 ± 0.67 ^{ab}	108 ± 2.50 $^{c-l}$	20 ± 0.72 ^{c}
6	90 ± 0.00 ^{abc}	132 ± 0.00 $^{a-d}$	9 ± 0.00 ^{ab}	73 ± 0.00 $^{i-m}$	15 ± 0.23 ^{c}
7	100 ± 0.00 ^{abc}	132 ± 0.00 $^{a-d}$	7 ± 1.00 ^{ab}	77 ± 3.00 $^{h-m}$	7 ± 0.23 ^{c}
8	93 ± 2.50 ^{abc}	134 ± 5.00 $^{a-d}$	15 ± 2.00 ^{a}	142 ± 12.50 $^{a-g}$	14 ± 2.38 ^{c}
9	113 ± 2.50 ^{abc}	129 ± 0.00 $^{b-d}$	9 ± 2.00 ^{ab}	143 ± 11.00 $^{a-f}$	18 ± 2.41 ^{c}
10	93 ± 1.67 ^{abc}	131 ± 1.00 ^{bcd}	9 ± 0.00 ^{ab}	128 ± 1.00 $^{a-h}$	18 ± 2.11 ^{c}
11	87 ± 1.67 ^{bc}	128 ± 4.33 ^{cde}	12 ± 0.67 ^{ab}	120 ± 1.76 $^{b-j}$	17 ± 0.74 ^{c}
12	105 ± 2.89 ^{abc}	119 ± 0.00 ^{ef}	12 ± 1.20 ^{ab}	99 ± 4.50 $^{d-m}$	21 ± 1.91 ^{c}
13	95 ± 2.89 ^{abc}	129 ± 0.00 $^{b-d}$	7 ± 0.33 ^{ab}	82 ± 0.00 $^{h-m}$	6 ± 0.81 ^{c}
Nipponbare	97 ± 3.33 ^{abc}	142 ± 0.00 ^{a}	5 ± 0.33 ^{b}	118 ± 5.50 $^{c-l}$	90 ± 2.67 ^{ab}
KDML105	102 ± 3.33 ^{abc}	113 ± 0.67 $^{b-d}$	11 ± 1.45 ^{ab}	111 ± 6.50 $^{c-l}$	81 ± 4.54 ^{ab}

Different lowercase letters within the same column show significant differences at the 0.05 level. N = 3

4.3 Strategy to identify tissue-specific involved in spikelet fertility

4.3.1 Fertility evaluations pistils, pollen and seed setting rate (%)

The pistils of HS and LS grow normally in the spikelet (Figure 4.6a). To identify defects in male reproductive development in LS, spikelets and florets were collected from panicles of the F_5 plants and studied in the pollen fertility and pollen tubes germination compared with the HS. Stain I_2-KI revealed that pollen grains both HS and LS were fertile with among typical and spherical forms (Figure 4.6b). The F_5 HS plants including line number 21, 40 and 50 were collected to stain with I_2-KI revealed that pollen grains showed normal fertile with the average of pollen fertility 95.79, 96.93 and 97.62%. Also, LS plants including line number 11, 18, 19 and 20 grow normally with the average of pollen fertility 92.98, 98.89, 95.46 and 95.84 (Figure 4.6c). Although pollen and the pistils fertility were observed in the two groups, there is a significant difference in seeding rate between F_5 HS and LS. The seed setting rates of F_5 HS varied from 71.60% to 84.89%, with average of 77.73%, while 4.14% to 22.26%, with average of 14.66 % were explored in LS. However, the seed setting rates of the parents were 85.93 and 80.95%, respectively (Table 4.3). This result confirms that the difference in the seed-setting rate between HS and LS may occur in the next process.

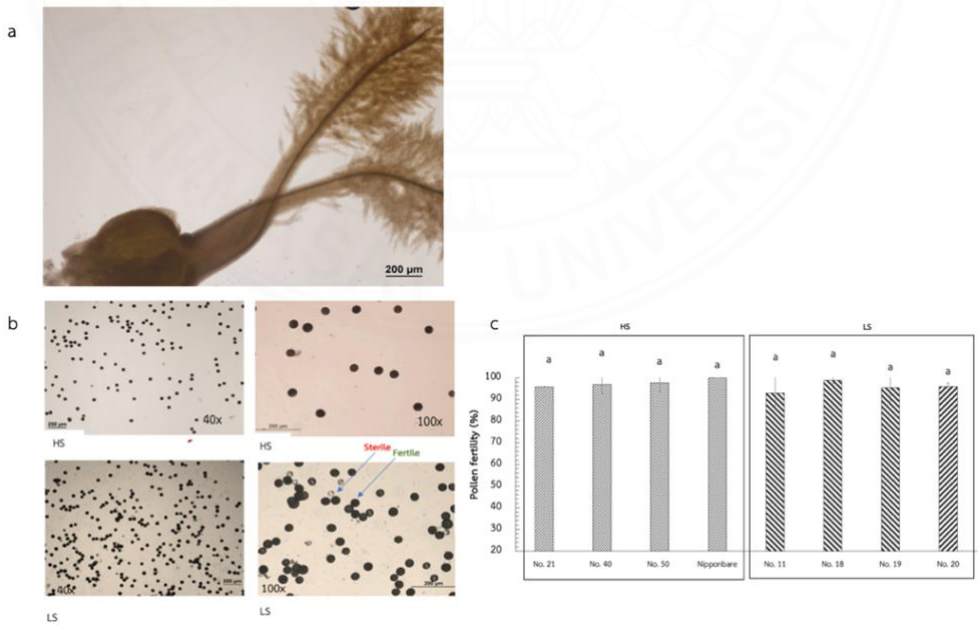

Figure 4.6 Pollen stained with I_2-KI solution (a), scale bar = 1 and pollen fertility F_5 HS and LS (b).

Table 4.3 Seed setting rate (%) of HS and LS (F_5 NB x KDML105)

No.	Seed setting rate (%)	
	HS	**LS**
1	83.28 ± 1.98 [a]	13.67 ± 2.64 [bc]
2	84.89 ± 8.25 [a]	22.26 ± 2.74 [b]
3	71.60 ± 9.07 [a]	20.34 ± 3.67 [bc]
4	79.69 ± 9.24 [a]	15.36 ± 4.75 [bc]
5	74.35 ± 5.18 [a]	7.14 ± 4.94 [bc]
6	73.60 ± 10.51[a]	21.58 ± 4.57 [b]
7	74.42 ± 6.13 [a]	4.14 ± 2.82 [c]
8	80.01 ± 2.06 [a]	12.76 ± 1.08 [bc]
Average	**77.73 ± 4.90 [a]**	**14.66 ± 6.66 [bc]**
Parents	Nipponbare (NB)	85.93 ± 2.69 [a]
	KDML105	80.95 ± 7.37 [a]

4.3.2 Pollen tube germination or pollination state

To monitor the pollen tube germination or pollination state, the germination of pollen grains and pollen tube growth were observed in *vivo* using the aniline blue stain at 8, 9, 10, 11, 12 am in R4 growth stages (Figure 4.7). Pollen germination and growth of HS and LS in *vivo* shown that pollens were not adhesive stigma, and pollen tubes were not reached the ovules at 8, 9, 10, and 11 am (Figure 4.7 a-b). However, there are little pollen-stigma adhesion, and pollen tubes were reached the ovules in HS and LS at 12 p.m. There are abundant pollen grains, multiple pollen-stigma adhesion and multiple pollen tubes reached the ovules of the HS pistils at 4-5 hr. after pollination was initiated at 12 am (Figure 4.7c). In contrast, little pollen grains, no pollen-stigma adhesion and no pollen tubes reached the ovules of the LS (Figure 4.7d).

In addition, the result showed little pollen grains, no germinated pollen grains and no pollen tubes reached the ovules in the HS and LS pistils at 1-2 hr. before flowering (BF) (Fig. 4.8a, d). At 1-2 hr. after flowering (AF) (Fig. 4.8b, e), the result showed little pollen grains, no germinated pollen grains and no pollen tubes reached the ovules in the HS and LS pistils. However, there are abundant pollen grains, multiple pollen-stigma adhesion and multiple pollen tubes reached the ovules of the HS pistils (Fig. 4.8c), at 4-5 hr. after pollination. In contrast, little pollen grains, no pollen-stigma adhesion and no pollen tubes reached the ovules of the LS at 4-5 hr. after pollination (Fig. 4.8f).

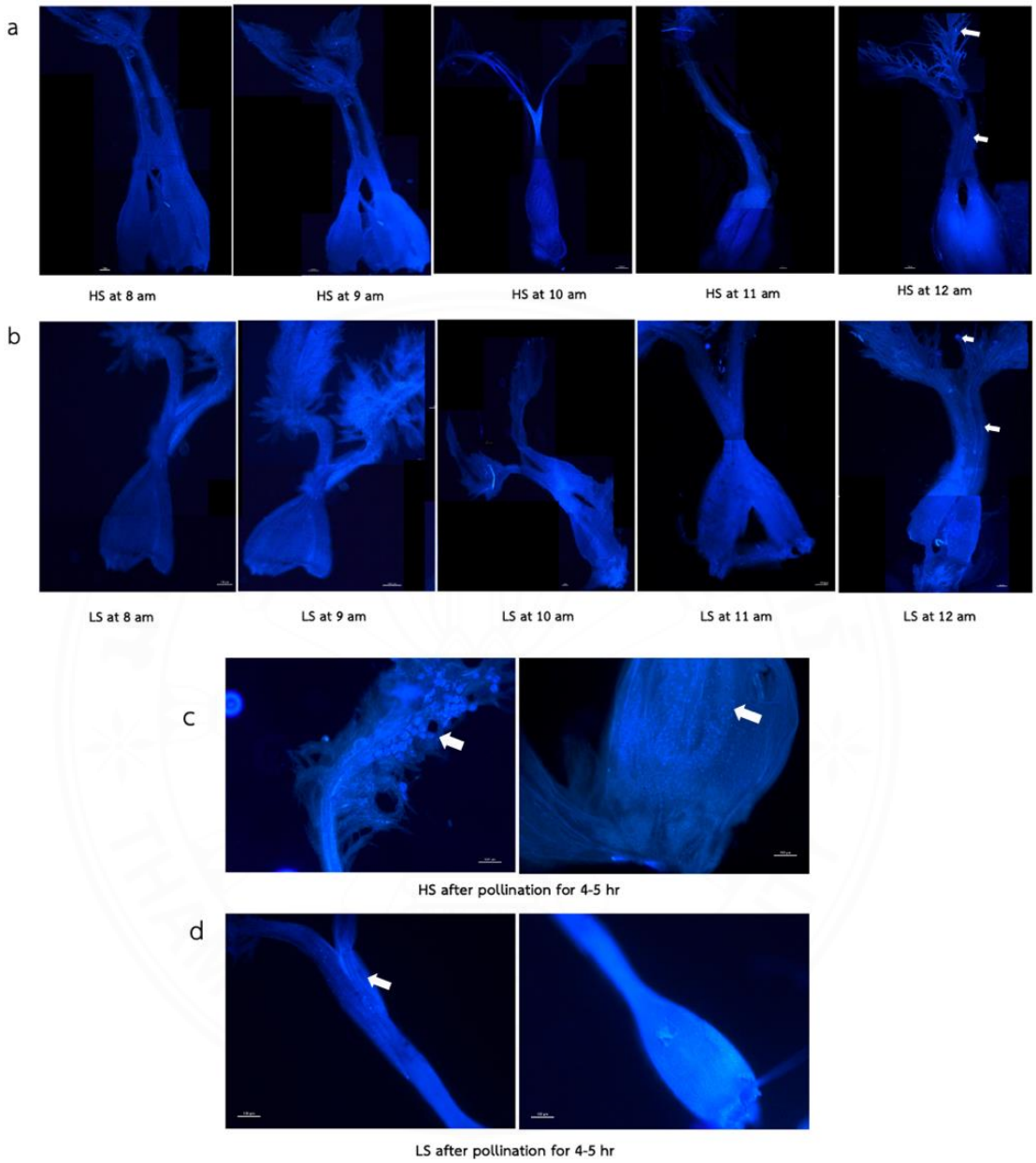

Figure 4.7 Pollen germination and growth of F_5 plant in HS and LS lines in *vivo*. Aniline blue staining of the HS and LS pistils at 8, 9, 10, 11, 12 am in R4 growth stages before flowering (a and b), the HS and LS pistils after flowering for 4-5 hr. (b and d), Scale bar = 100 μm

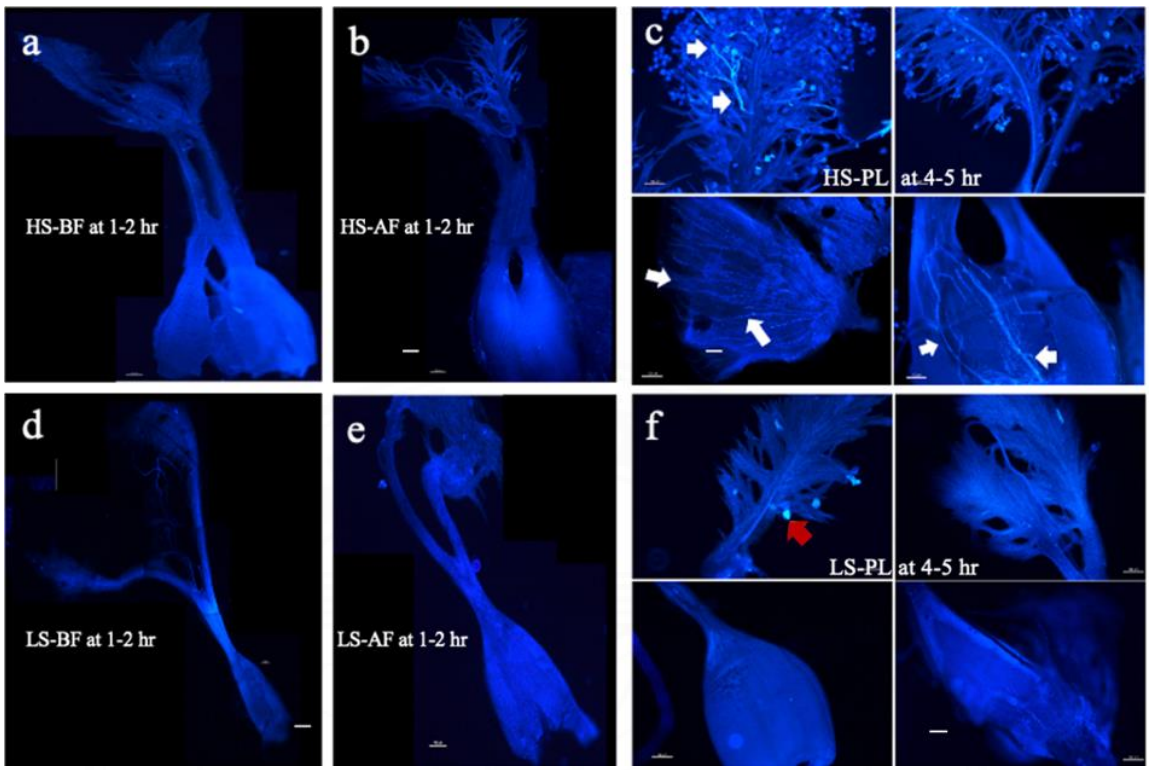

Figure 4.8 Pollen grain, pollen germination and growth of HS **(a-c)** and LS lines **(d-f)** in *vivo*. Aniline blue staining of the HS pistils at 1-2 hr. BF (**a)**, HS pistils at 1-2 hr. AF **(b)**, HS pistils at 4-5 hr. after PL. **(c)**, LS pistils at 1-2 hr. BF. **(d)**, LS pistils at 1-2 hr. AF. **(e)**, LS pistils at 4-5 hr. after PL **(f)**. Scale bar = 100 μm. White arrow indicate the germinated pollen tube and red arrow indicate the pollen-stigma adhesion

4.4 RNA-Seq

4.4.1 RNA quantification and qualification

Determining the integrity of RNA starting materials is a critical step in gene expression analysis. Profiles generated on the Agilent 2100 bioanalyzer yield information on concentration, allow a visual inspection of RNA integrity, and generate ribosomal ratios (Mueller et al., 2004). There is intensity in the 18S to 28S ribosomal band ratio in all sample (Figure 4.9). In HS-pooled tube had 18S to 28S ribosomal bant with 1774 nt and 3373 nt, respectively, which was not contaminated from test run (Figure 4.9 left). In LS-pooled tube had 18S to 28S ribosomal bant with 1765 nt and 3396 nt, respectively, which was not contaminated from test run (Figure 4.9 left). All sample is qualified and the library preparation can be prepared regularly.

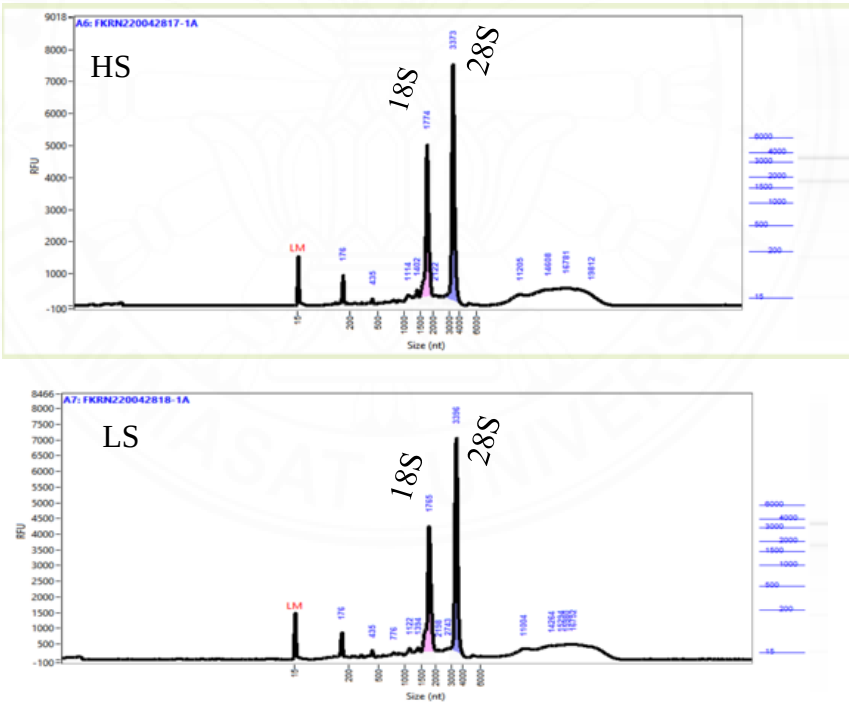

Figure 4.9 Agilent 2100 (right), and electrophoresis (left) and results for an acceptable total RNA sample.

4.4.2 Data Analysis

4.4.2.1 RNA data and read assembly

Our results showed that HS had the normal pollen tube growth in ovaries at 4-5 hr. after pollination, while no pollen tubes reached the ovules of the LS, indicating that the pollens tube growth is critical for spikelet fertility. Therefore, we used the pollinated pistils at 4-5 hr. from HS and LS for transcriptome profiling using RNA-seq analysis. A total of 79.15 million and 61.90 million raw reads were obtained from the HS and LS transcriptome libraries, respectively (Table 4.4). The results revealed that more than 80% of clean reads were high-quality. The error rate was low for both samples (0.03%). The average Q20 (error rate less than 1%) and Q30 (error rate less than 0.01%) were 97.69% and 93.68%, respectively. Both samples have average GC content approximately 52.36%. After aligning the clean reads of HS and LS to referent genome *japonica* variety Nipponbare, the results showed that 94.47% and 95.19% of clean reads were mapped, and 91.66% and 92.41% of mapped reads from HS and LS, respectively, were unique, while 2.82 and 2.79 reads of HS and LS were mapped to multiple loci, respectively (Table 4.4).

Table 4.4 Statistics of RNA sequencing results

Sample	HS	LS
Raw reads	79,151,432	61,897,814
Clean reads	77,430,622 (98.90%)	60,534,000 (98.89%)
Error rate	0.03	0.03
Q20 (%)	97.58	97.81
Q30 (%)	93.51	93.85
GC content (%)	52.29	52.44
Total mapped	73,150,117 (94.47%)	57,625,014 (95.19%)
Uniquely mapped	70,970,054 (91.66%)	55,938,804 (92.41%)
Multiple mapped	2,180,063 (2.82%)	1,686,210 (2.79%)

HS, High seed setting rates; LS, Low seed setting rates
Raw_reads: Reads count from the raw data, four rows as a unit, with statistics of reads count for every sequencing.
Clean_reads: Base number of raw data after filtering. (Number of clean reads) * (sequence length), converting unit to G.
Error_rate: Average sequencing error rate, which is calculated by Qphred=-10log10(e).
Q20: The percentage of the bases whose Q Phred values is greater than 20. (Number of bases with Q Phred value > 20) / (Number of total bases) *100. Q30: The percentage of the bases whose Q Phred values is greater than 30. (Number of bases with Q Phred value > 30) / (Number of total bases) *100.
GC_pct: The percentage of G&C base numbers of total bases. (G&C base number) / (Total base number) *100.

4.4.2.2 Reads mapping to the reference genome

Mapping the clean reads to the NB (reference genome) or the transcriptome is the basis for the next following analysis. We use HISAT2 to accomplish the mapping. Only the differential expression significant genes are needed, which we map reads to the transcriptome directly. After aligning clean reads to *japonica* varieties NB, the results showed that the total 94.47% and 95.19% of clean reads were mapped in HS and LS, respectively, 91.66% and 92.41% of mapped reads were unique in HS and LS, respectively and 2.79-3.07% of reads were mapped to multiple loci with NB (referent genome). Furthermore, mapped regions can be classified as exons, introns, or intergenic regions. Exon-mapped reads should be the most abundant type of reads when the NB (reference genome) is well-annotated (Table 4.5). Intron-reads may be

derived from pre-mRNA contamination or intron-retention from alternative splicing. Reads mapped to intergenic regions are mainly attributed to weak annotations of the reference genome. The distribution of sequencing reads of all samples in the genomic region is shown in the Figure 4.10. The percentage of number of reads aligned to exon regions of the genome was 90.12 and 89.69 in HS and LS, respectively. The percentage of reads aligned to intron regions of the genome was 2.56 and 2.78 in HS and LS, respectively. In addition, the intergenic regions of the genome were aligned around 7.32 and 7.53 % in HS and LS, respectively.

Table 4.5 Comparison of sample and reference genomes

Sample	HS	LS
Total_reads	77,430,622	60,534,000
Total_map	73,150,117 (94.47%)	57,625,014 (95.19%)
Unique_map	70,970,054 (91.66%)	55,938,804 (92.41%)
Multi_map	2,180,063 (2.82%)	1,686,210 (2.79%)

HS, High seed setting rates; LS, Low seed setting rates

Total_reads: Total clean reads used for analysis.

Total_map: Number and percentage of reads aligned to the genome, the ratio should higher than 70%, total mapping rate: (mapped reads)/ (total reads) *100.

Unique_map: Number and percentage of reads aligned to the unique position of the reference genome (for subsequent quantitative data analysis), unique mapping rate: (uniquely mapped reads)/(total reads)*100.

Multi_map: number and percentage of reads aligned to multiple locations in the reference genome, multiple mapping rate: (multiple mapped reads)/ (total reads) * 100.

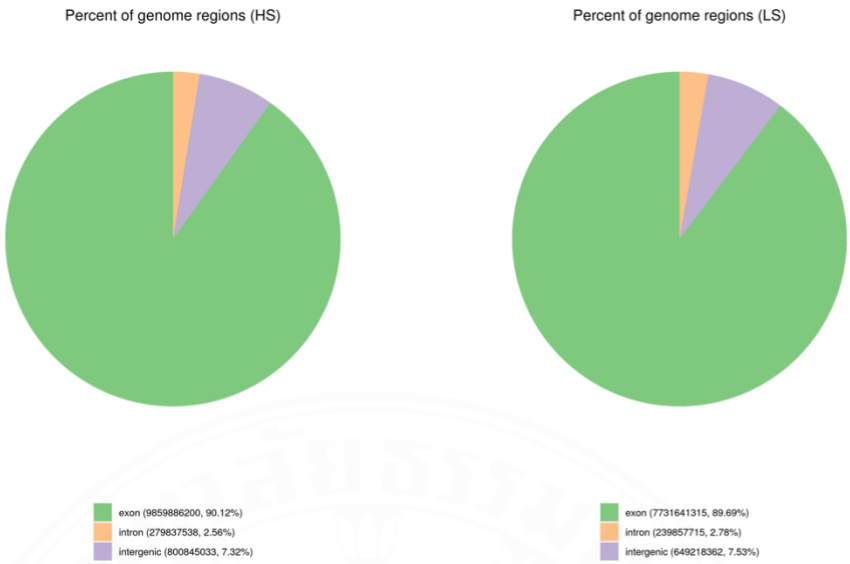

Figure 4.10 Sequencing reads in the genomic region in HS and LS, the ratios of the different colors in the figure represent the ratio of reads to different regions

4.4.2.3 Quantification of gene expression level

In RNA-seq experiments, gene expression level is estimated by the abundance of transcripts (count of sequencing) that mapped to genome or exon. Read counts is proportional to gene expression level, gene length and sequencing depth. Read counts were calculated in into FPKM. The FPKM not display because it is abundant of genes in HS and LS.

4.4.2.3.1 Verifying reliability and sample selection

Correlation of the gene expression levels between samples plays an important role in verifying reliability and sample selection, which can not only demonstrate the repeatability of the experiment but estimate the differential gene expression analysis. The closer the correlation coefficient is to 1, the higher similarity the samples have. Encode suggests that the square of the Pearson correlation coefficient should be greater than 0.92 (under ideal experiment conditions) and the R^2 should be greater than 0.8. The result shown the R^2 was accepted to be greater than 0.8 in all sample (Figure 4.11). It is intuitive to show sample differences and repeat cases between groups. The higher the correlation coefficient of the sample is, the closer the

expression pattern is. According to all gene expression levels (RPKM or FPKM) of each sample, the correlation coefficient of samples between groups is calculated and drawn as heat maps (Figure 4.11).

Principal component analysis (PCA) is also commonly used to evaluate intergroup differences and intragroup sample duplication. We performed PCA analysis on the gene expression value (FPKM) of all samples, as shown in the figure 4.11b. The result shown the samples between groups were dispersed and the samples within groups were gathered together.

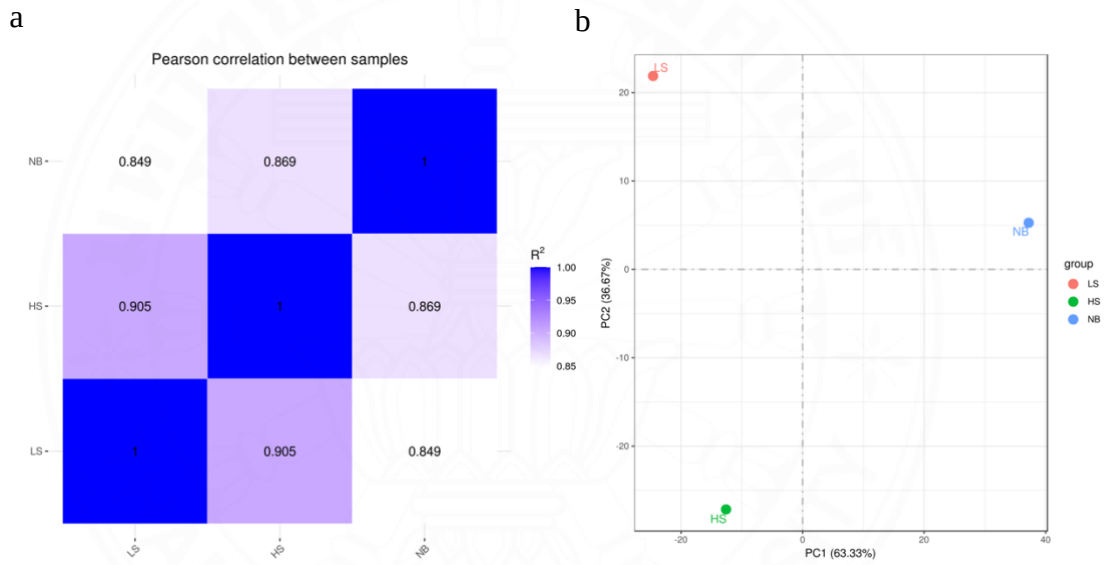

Figure 4.11 Inter-sample correlation heat map R^2 = Square of Pearson correlation coefficient (R) (a) and principal component analysis result (b)

4.4.2.3.2 Identification of total genes

To evaluate the gene expression in different high and low seed setting rate, we estimated gene expression level by the abundance of transcripts (count of sequencing) that mapped to genome or exon. The FPKM (short for the expected number of Fragments Per Kilobase of transcript sequence per Millions base pairs sequenced) of each gene were calculated as the normalized read counts. The genes with FPKM values > 0.5 were considered to be expressed genes. The results showed that a total of 21,803 genes were found to be commonly expressed in the two samples, while 901 and 1168 were specifically expressed for HS and LS respectively (Figure 4.12a). As shown in figure 4.12 b, there were variations in the expressions between the specific HS and the specific LS group.

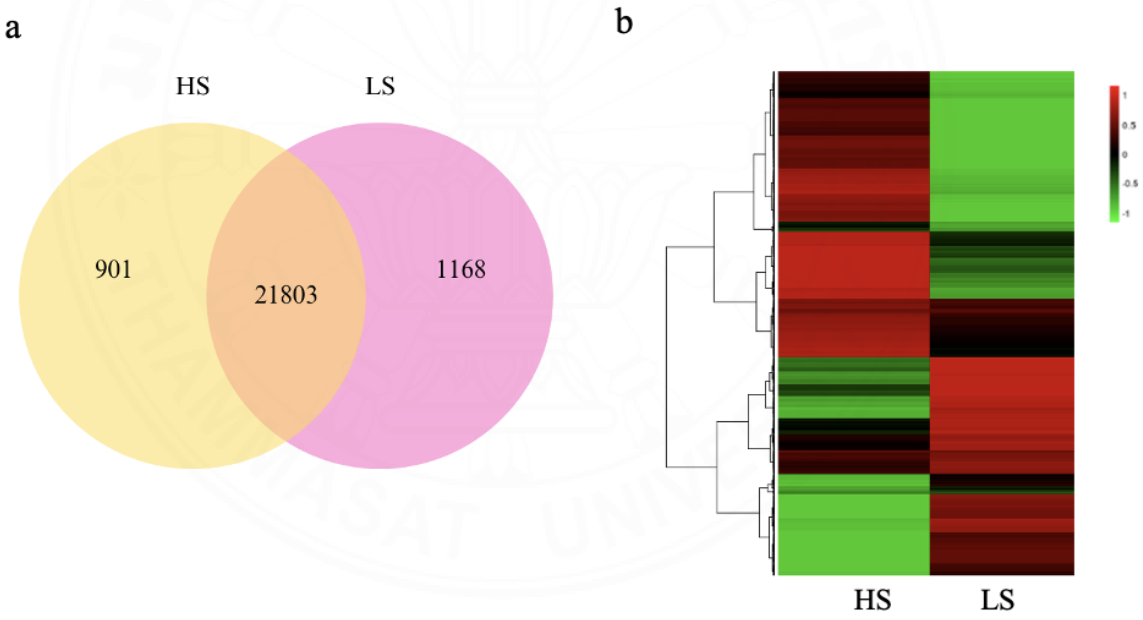

Figure. 4.12 Co-expression and the hierarchical cluster analysis (a and b). Venn diagram presents the number of genes that are uniquely expressed within each sample, with the overlapping regions showing the number of genes that are expressed in both samples base on the 0.5 FPKM as the threshold **(a)**. The hierarchical clustering of specific gene that are expressed in HS and LS base on the FPKM value. Green represents lower expression, red represents high expression, column represent individual experiments, and rows represent a gene **(b)**.

4.4.2.4 Identification of Differential Gene Expressions (DEGs)

To identify genes that exhibited significant changes in expression specific for high seed-setting (HS) and low seed-setting (LS) groups, differentially expression genes (DEGs) between the two groups were identified using fold change values in Fig. 4.12a. The significant DEGs were filtered based on criteria: |log2(Fold Change)| > 1 (up-regulated), |log2(Fold Change)| ≤ 1 (down-regulated), and a corrected p-value (padj) ≤ 0.05. The results showed that 249 DEGs were specific for HS including 118 up-regulated genes and 131 down-regulated genes. In addition, 473 DEGS including 214 up-regulated and 259 down-regulated genes were specific for LS, indicating that there were more DEGs in LS than those in HS (Figure 4.13).

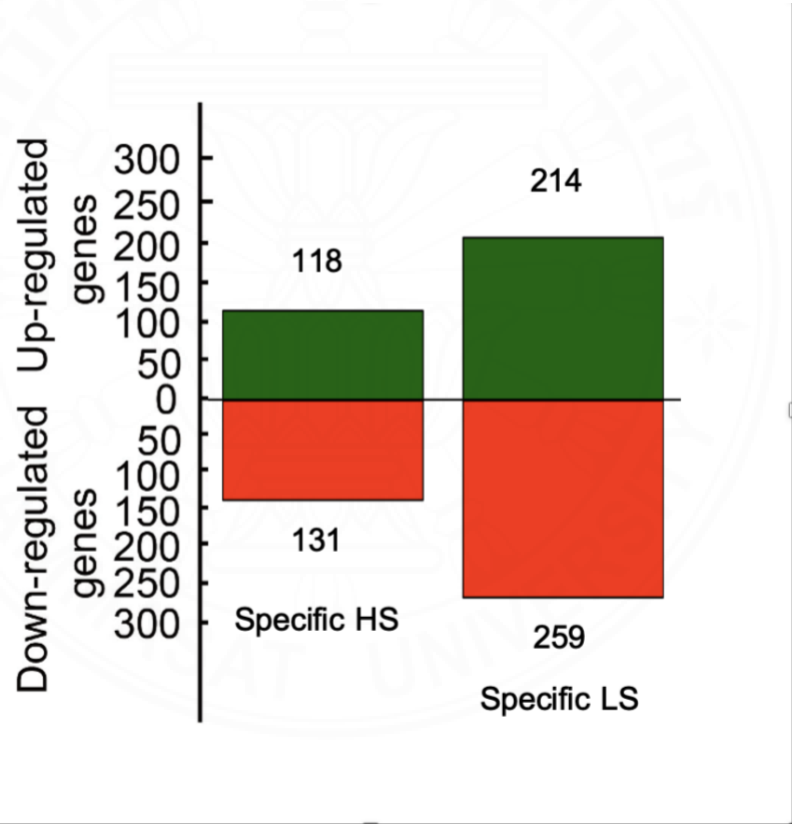

Figure. 4.13 The significant differentially expressed genes (DEGs) specific for HS and LS. Green and red represent up-regulated and down-regulated DEGs, respectively, and the numbers on the columns indicate the number of DEGs

4.4.2.5 Validation of RNA-seq results by quantitative real-time PCR

To validate expression results of RNA-seq, 10 DEGs specific for HS and LS were randomly selected for expression analysis using qRT-PCR (Table 3.2) The Log2fold change comparisons were performed between RNA-seq and qRT-PCR. The qRT-PCR results for all DEGs showed the same trend as those observed in RNA-seq (Figure 4.14).

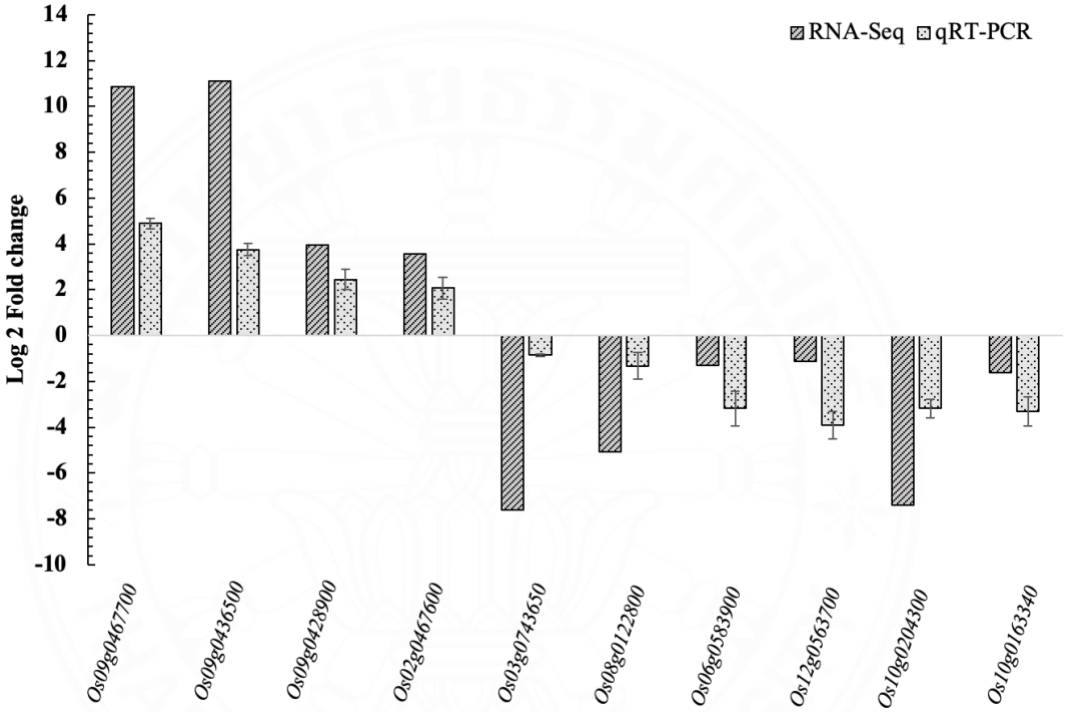

Figure. 4.14 Comparison of gene expression levels by RNA-seq and qRT-PCR. Ten DEGs specific for HS and LS were randomly selected. The relative expression values were normalized to the rice *UBQ5* gene (*OsUBQ5)*. The results were obtained from two biological replicates. Error bar indicated standard deviation.

4.4.2.6 Functional categorization, pathway enrichment, and identification of candidate genes associated with spikelet fertility in the HS group

To identify the functional significant of 249 DEGs that exhibited specific for HS, these DEGs was subjected for annotation through Gene Ontology (GO) and KEGG analyses. The significant GO terms and KEGG enrichments were filtered based on criteria: p-value (padj) ≤ 0.05 and FDR ≤ 0.05. The significant enrichment of 5 out of 56 GO terms were up-regulated in the HS group (Figure 4.15). These terms were related to cellular components and include vesicle cytoplasmic, membrane-bounded vesicle, cytoplasmic vesicle, intracellular vesicle, and membrane-bounded vesicle. However, 18 out of the 77 GO terms showed significant enrichment among down-regulated genes in the HS group (Figure 4.16). These terms include organelle, intracellular organelle, cell, cell part, and intracellular part, all pertaining to cellular components. This indicated that these functional categorizations were responsible to spikelet fertility in the HS group.

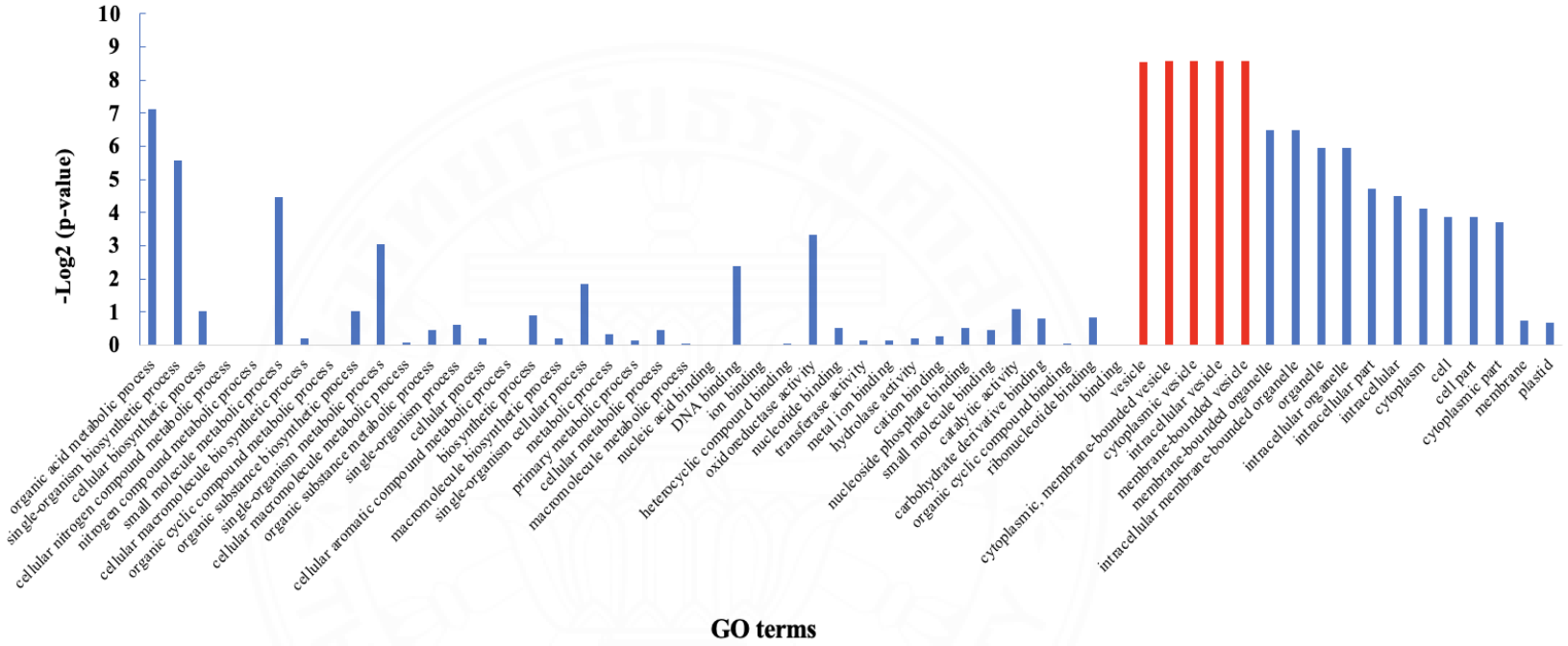

Figure. 4.15 The significant enrichment of 5 out of 56 GO terms of up-regulated specific for HS group

Blue bar represented non-significant GO term and red bar represented significant GO term

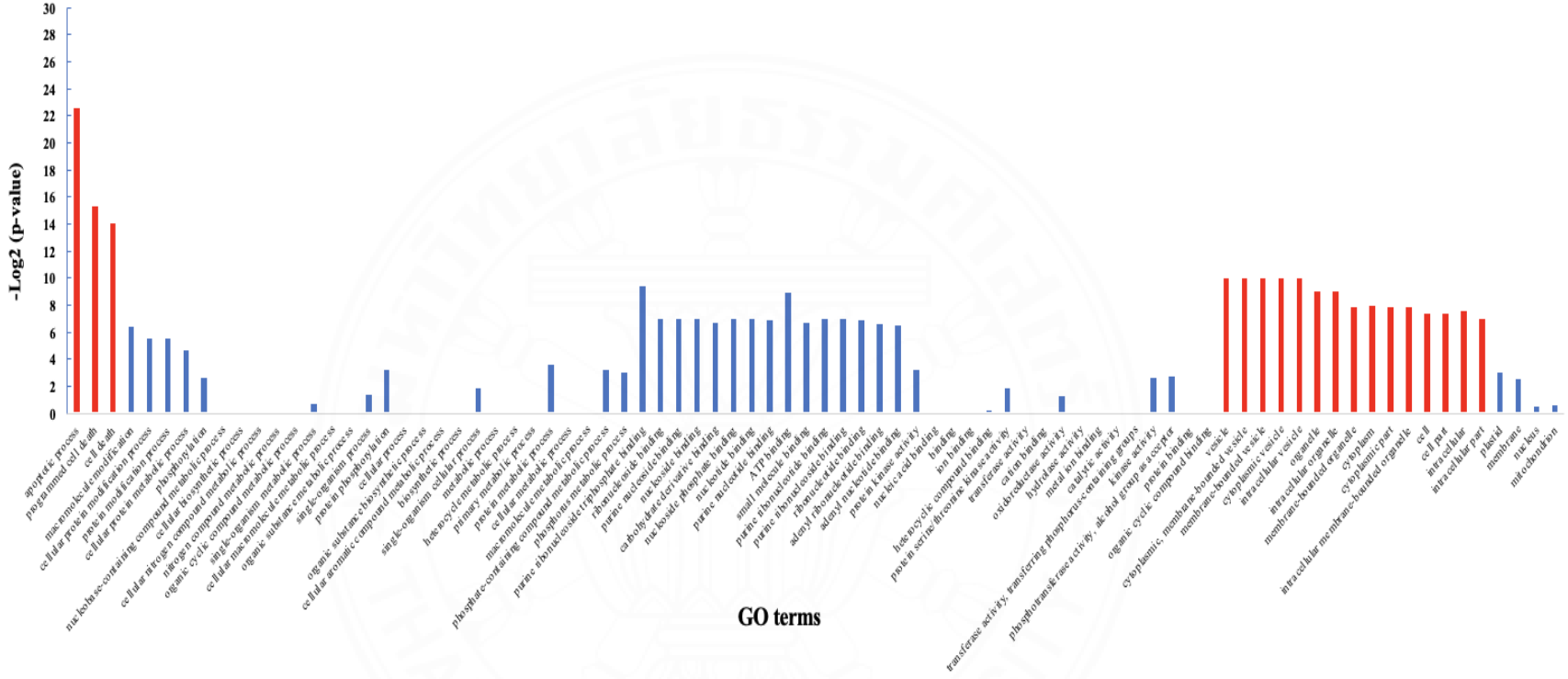

Figure. 4.16 The significant enrichment of 18 out of the 77 GO terms of down-regulated specific for HS group

Blue bar represented non-significant GO term and red bar represented significant GO term

To determine the pathway in which the 249 DEGs were implicated and enriched, a pathway-based analysis was conducted utilizing the KEGG pathway database using ShinyGO v.0.77 (Ge et al., 2020). Only the metabolic pathway was significantly associated with 11 DEGs (Table 4.6). These DEGs were composed with 6 up-regulated DEGs such as *Os06g0133000* (*Wx*), *Os09g0457400* (*Amy3A*), *Os06g0185100*, *Os11g0701100*, *Os02g0467600*, and *Os03g038210*, while 5 DEGs, namely *Os02g0134400*, *Os03g0670200*, *Os03g0368900*, *Os04g0573100* (*HTH1*), and *Os03g0321800*, were found to be down-regulated (Table 4.6).

Table 4.6 HS specific DEGs assigned to metabolic pathways using KEGG pathway database

Genes Id	Gene annotation	KO	LogFC
Os06g0133000	Granule-bound starch synthase, Synthesis of amylose in endosperm	K13679	2.32
Os09g0457400	Alpha-amylase isozyme 3A precursor (EC 3.2.1.1) (1,4-alpha-D-glucan glucanohydrolase)	K01176	2.07
Os06g0185100	Similar to estradiol 17-beta dehydrogenase 8	K00059	1.94
Os11g0701100	Similar to Class III chitinase homologu	K01183	1.83
Os02g0467600	Similar to Cinnamate 4-hydroxylase CYP73	K00487	1.82
Os03g0382100	Thiolase-like, subgroup domain containing protein	K15397	1.45
Os02g0134400	Fumarate reductase/succinate dehydrogenase flavoprotein, C-terminal domain containing protein	K00278	-1.07
Os03g0670200	Similar to YJR013Wp (Fragment)	K05284	-1.20
Os03g0368900	Haem peroxidase family protein	K00430	-1.20
Os04g0573100	Putative glucose-methanol-choline (GMC) oxidoreductase, Cutin biosynthesis, anther development, Pollen fertility	-	-1.25
Os03g0321800	Similar to WRKY transcription factor 55	K23735	-1.46

4.4.2.7 Functional categorization, pathway enrichment, and identification of candidate genes associated with the irregular regulation of pollen development in the LS group

In order to identify functional significance of 473 DEGs related to abnormalities in pollen tube growth in the LS, we subjected these DEGs to annotation using Gene Ontology (GO) and KEGG analyses. The significant GO terms and KEGG enrichments were filtered based on the following criteria: p-value (padj) ≤ 0.05 and FDR ≤ 0.05. For the up-regulated genes specific for the LS group, a total of 16 out of 55 GO terms were significantly enriched, particularly in cellular components, including membrane-bounded vesicles, vesicles, cytoplasmic regions, and intracellular vesicles (Figure 4.17). However, among the down-regulated genes specific for the LS group, a total 23 out of 131 GO terms showed significant enrichment in apoptotic process (biological process), receptor activity electron carrier activity, molecular transducer activity and monooxygenase activity (molecular function), vesicle, cytoplasmic, membrane-bounded vesicle, intracellular vesicle, membrane-bounded vesicle, cytoplasmic vesicle (cellular component), suggesting played a role in spikelet sterility (Figure 4.18).

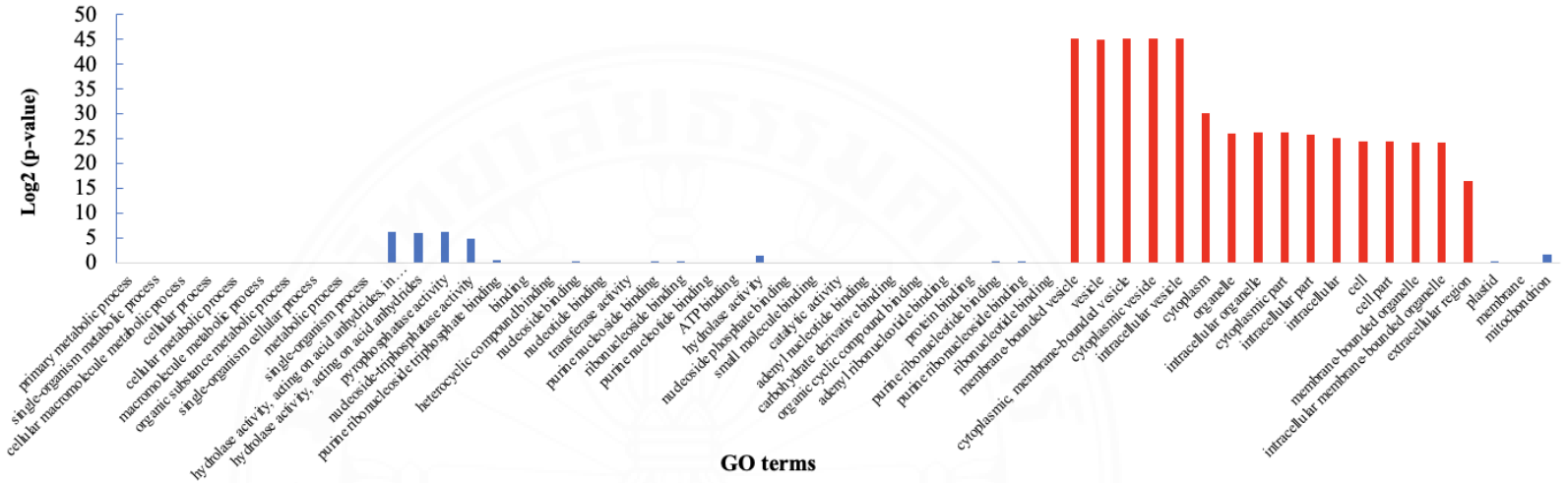

Figure. 4.17 The significant enrichment of 16 out of the 55 GO terms of up-regulated specific for LS group

Blue bar represented non-significant GO term and red bar represented significant GO term

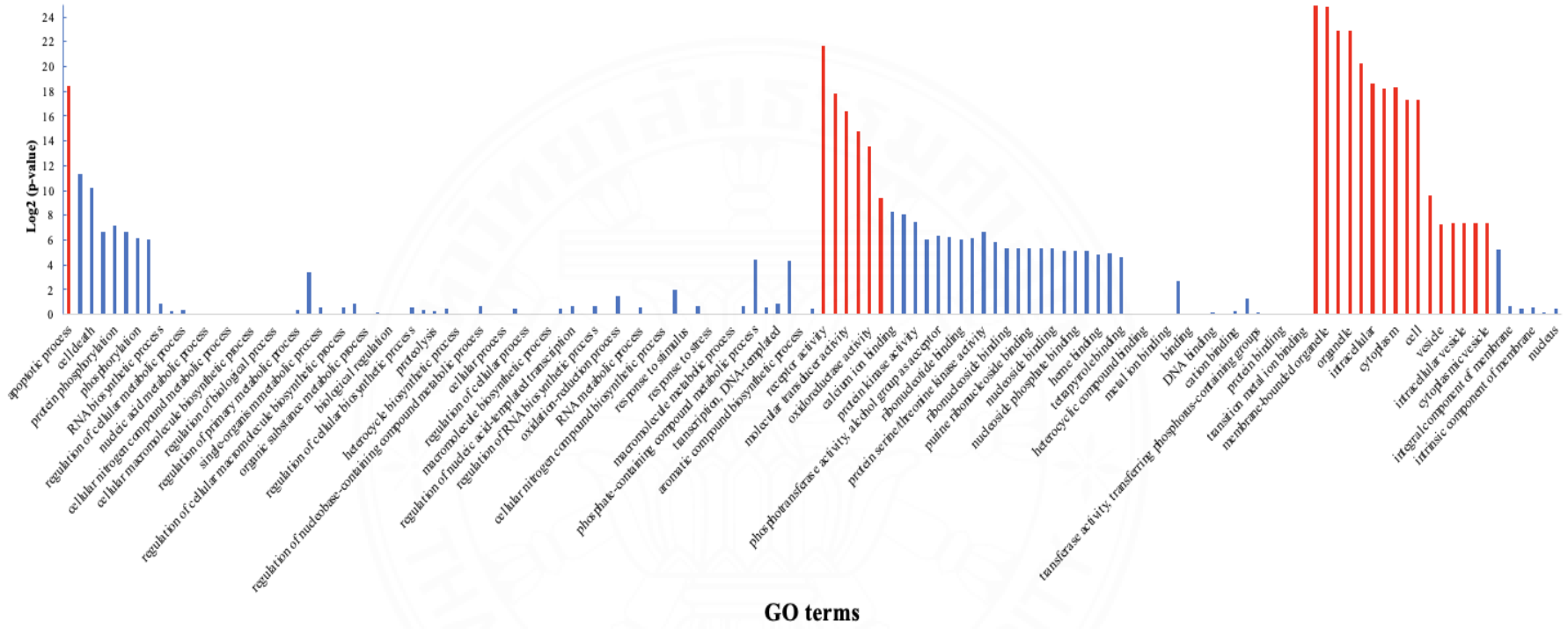

Figure. 4.18 The significant enrichment of 23 out of the 131 GO terms of down-regulated specific for LS group

Blue bar represented non-significant GO term and red bar represented significant GO term

For the KEGG pathway analysis, twenty-four significant DEGs were identified. Among these, ten DEGs exhibited up-regulation in pentose and glucuronate interconversions, biosynthesis of various plant secondary metabolites and metabolic pathways. Conversely, fourteen DEGs showed down-regulation in flavonoid biosynthesis, biosynthesis of secondary metabolites and metabolic pathways. The pentose and glucuronate interconversions pathway revealed the highest fold enrichment (40%) under up-regulation, while the flavonoid biosynthesis pathway exhibited the highest fold enrichment (27.78%) under down-regulation. The pentose and glucuronate interconversions pathway exhibited significant enrichment of four DEGs that encoded for pectin methylesterase (PME) and pectate lyase (PLL). These DEGs were identified as pectin methylesterase 6 (*Os03g0300500*; *OsPME10*), pectin lyase fold domain-containing protein (*Os12g0563700*; *OsPME35*), pectate lyase-like protein (*Os02g0214400*; *OsPLL3*), and pectate lyase (*Os06g0583900*; *OsPLL9*) (Table 4.7). Contrastingly, in the context of flavonoid biosynthesis, the genes *Os02g0767300* and *Os04g0101400* exhibited down-regulation. These genes are responsible for encoding flavonol synthase/flavanone 3-hydroxylase and flavone synthase II (FNSII), respectively (Table 4.7). Exploring the functionality of these LS-associated genes is anticipated to provide deeper insights into the complex regulatory network responsible for disrupting pollen tube growth in rice.

Table. 4.7 The highest fold enrichment of DEGs specific for LS in pentose and glucuronate interconversions pathway, and flavonoid biosynthesis pathways

Genes Id	Gene annotation	KO	LogFC
Pentose and glucuronate interconversions pathway			
OsPME10	Pectin methylesterase 6	K01051	1.43
OsPME35	Pectin lyase fold domain containing protein	K01051	1.13
OsPLL3	Pectate lyase-like protein	K01728	1.21
OsPLL9	Pectate lyase	K01728	1.30
Flavonoid biosynthesis pathway			
Os02g0767300	Flavonol synthase/flavanone 3-hydroxylase	K05278	-1.34
Os04g0101400	Flavone synthase II (FNSII), Biosynthesis of tricin O-linked conjugate	K23180	-1.73

4.5 Co-expression analysis

4.5.1 Co-expression network of 11 specific DEGs for HS

In order to explore the co-expression network among the 11 specific DEGs for HS within the metabolic pathways, we conducted an analysis of their interactions using 4 genes of pollen development available on the rice database (http://ricefrend.dna.affrc.go.jp). The gene network was generated through a multi-guide gene search (MR = 5, hierarchy = 1) (Figure 4.19a). The results showed a network comprising 60 nodes and 71 edges. Three out of the eleven DEGs, namely *Amy3A*, *HTH1*, and *Os03g0382100*, showed co-expression with pollen development genes. These genes included *Os01g0725900*, which encodes the pollen Ole e 1 allergen and extension domain-containing protein, *Os08g0302000*, which encodes a peroxidase 40 precursor, and *Os03g0251000*, which encodes a seed storage/trypsin-alpha amylase inhibitor domain-containing protein, respectively. Furthermore, *Os02g0467600* (cinnamate 4-hydroxylase CYP73) is co-expressed with AP2-EREBP transcription factors (TFs), which play a crucial role in regulating gene expression associated with spikelet fertility. To study the potential function of *Os02g0467600* related to spikelet fertility, we investigated the expression patterns of this gene before and after pollination using qRT-PCR (Figure 4.19b). The findings revealed that *Os02g0467600* was preferentially expressed in pistils of both HS and LS before and after pollination (Figure 4.13b). *Os02g0467600* had lower expression in the pistils of both HS and LS before pollination. However, after pollination for 4-5 hours, *Os02g0467600* was highly expressed in the HS pistil but low in the LS pistil, suggesting that it might be induced by the germination and growth of pollen tubes.

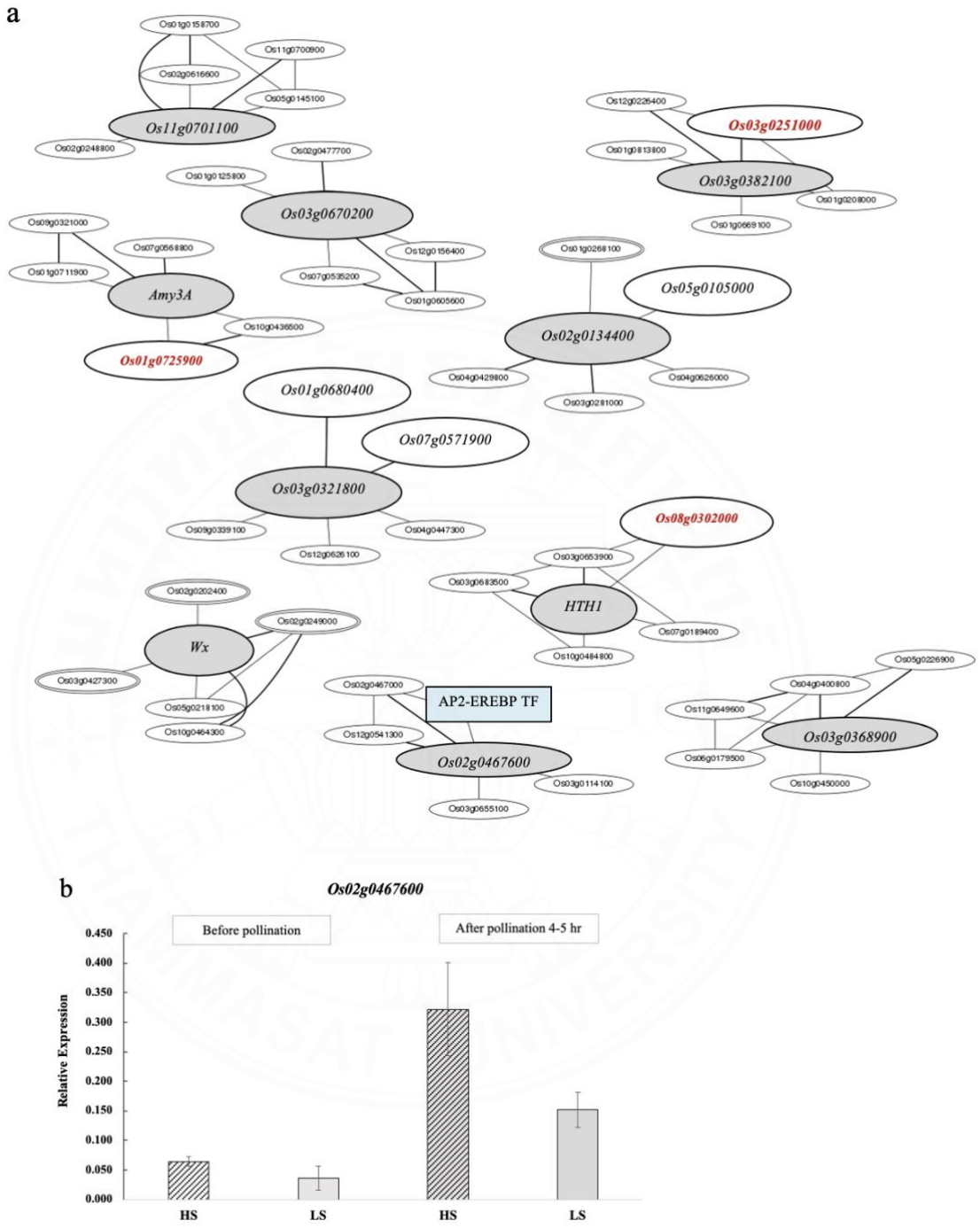

Figure. 4.19 The gene co-expression network and expression analysis by qRT-PCR of *Os02g0467600* specific for HS **(a-b)**. The co-expression of 11 specific DEGs for HS within the metabolic pathways **(a)**, expression analysis of *Os02g0467600* specific HS involved in spikelet fertility by qRT-PCR at before and after pollination **(b)**.

4.5.2 Co-expression network of 6 specific DEGs for LS

To develop a more comprehensive understanding of the intricate gene network of the 6 LS-specific DEGs with the highest fold enrichment, we analyzed co-expression network of these DEGs using RiceFRIEND database. The application of the multiple guide gene search (MR = 5, hierarchy = 1) resulted in the generation of a gene network containing 35 nodes and 54 edges (Figure 4.20a). The most extensive cluster comprises *OsPME10*, *OsPME35*, and *OsPLL3*, which are prominently enriched in the pentose and glucuronate interconversions pathway. Additionally, *OsPME10*, *OsPME35*, and *OsPLL3* were co-expressed with another set of 14 genes. Nevertheless, *OsPLL9*, which is one of the genes enriched in the pentose and glucuronate interconversions pathway, was separated from a cluster of *OsPME10, OsPME35,* and *OsPLL3.* Thus, we investigated expression patterns of *OsPME35* and *OsPLL9* specific for LS before and after pollination using qRT-PCR. The results showed that *OsPME35* and *OsPLL9* exhibited a similar expression pattern (Figure 4.20b). *OsPME35* and *OsPLL9* had higher expression in the pistils of the HS group compared to the LS group before pollination. Additionally, their expression decreased in both pistils, especially in HS pistils, after 4-5 hours of pollination, possibly due to the germination and growth of pollen tubes. Among the DEGs within flavonoid biosynthesis, we identified that *Os02g0767300* and *Os04g0101400* were assigned to a separate cluster. Especially, *Os02g0767300* was predicted to interact with *Os04g0137100* (*OsPLL4*) encoding pectate lyase. These genes could potentially associate with the disruption of pollen tube growth in the LS.

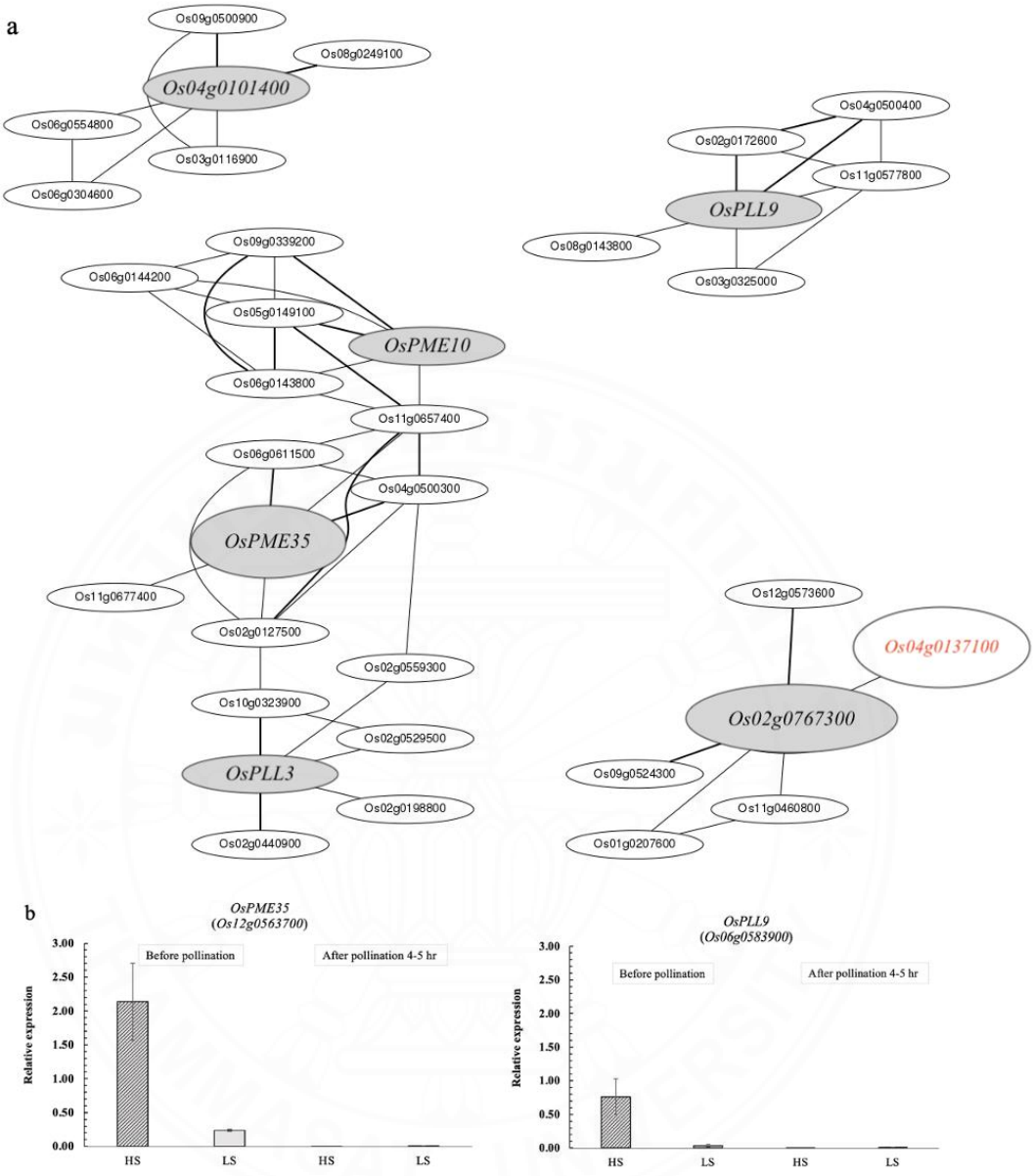

Figure. 4.20 The gene co-expression network and expression analysis by qRT-PCR of *OsPME35* and *OsPLL9* specific for LS **(a-b)**. The co-expression of 6 specific DEGs for LS within the pentose and glucuronate interconversions and flavonoid biosynthesis pathway **(a)**, expression analysis of *OsPME35* and *OsPLL9* specific LS involved in spikelet sterility by qRT-PCR at before and after pollination **(b)**.

4.6 The dominant enrichment DEGs mapped to the previously identified genes involved in seed setting, pollen germination, spikelet fertility and yield components in rice

A total of 21 QTLs and 6 functional genes associated with seed setting rate, spikelet fertility yield components and pollen tube growth in rice have been identified according to the Grammene QTLs database (https://www.gramene.org). A total of 11 DEGs including 8 DEGs specific for HS (Fig. 4.21, black label), 3 DEGs specific for LS (Fig. 4.21, red label), were located near QTL or in QTL intervals in 7 chromosomes. Among them, both of DEGs specific for HS, *Os02g0467600*, and LS, *OsPLL3*, were located to 2 QTLs associated with seed setting rate and spikelet fertility including *sf2* and *OsPLL3* on chromosome 2. Furthermore, *OsPME35*, a gene specific for LS, was found to be closely associated with gpp12.1, a locus that impacts the number of grains per panicle, located on chromosome 12.

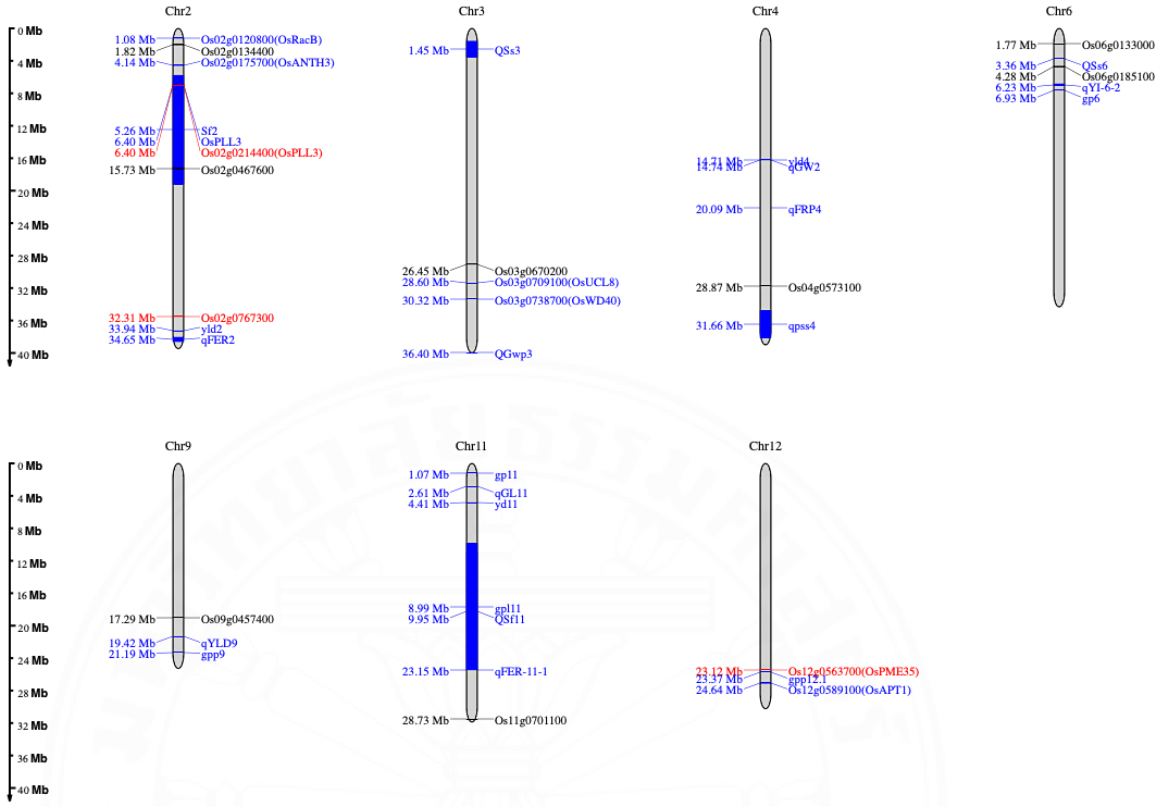

Figure. 4.21 Physical map showing the dominant enrichment genes specific for HS and co-localized with previous identified QTL/genes involved in spikelet fertility. The vertical bars represent individual chromosomes with numbers on left indicating physical position (in Mb). The previously identified QTL and functional genes involved in spikelet fertility, pollen growth and yield components are labeled in blue, the dominant enrichment DEGs specific for HS are labeled in black, the dominant enrichment DEGs specific for LS are labeled in red.

CHAPTER 5
DISCUSSION AND CONCLUSIONS

5.1 Discussion

Rice is one of the most important staple foods, and a model plant for the molecular genetic studies. Spikelet sterility is a crucial significant grain yield and quality in rice. Hybrid rice generated from intersubspecific crosses between *indica* and *japonica* rice was applied to construct the segregation of seed-setting rate in F_2 plants in our study. The maximum segregation and recombination for imposing agronomic traits can be selected to develop high-yield cultivars (Nevhudzholi et al., 2020). In this study, significant differences in seed-setting rates between the HS (high seed-setting) and LS (low seed-setting) were observed in the F_5 population. The application of Bulk Segregant Analysis (BSA) of HS and LS in this study played a pivotal role in selecting materials with distinct genetic backgrounds. This approach was instrumental in our finding to identify genes involved in spikelet fertility using RNA-Seq.

5.1.1 The importance of pollen tube growth for spikelet fertility

The pre/post-zygotic reproductive isolation, are important involving spikelet fertility in rice. Our results showed that the pre-zygotic plays an important role of spikelet fertility. Especially, a striking irregularity in the pollen tube growth is critical steps that lead to low seed setting rate of the LS, in agreement with previous studies by C. Zhang et al. (2018) and Yang Xu et al. (2017) reported that many pollen tubes appeared to lose the ability to grow directly into the style affecting the seed setting rate in rice. However, rapid pollen tube growth in the style is required for high seed-setting rate. Many previous studies showed that several genes have been identified to be important for pollen tube growth, such as mutant phenotype of *PTB1, OsCNGC13, OsPLL3, OsUCL8, OsWD40, OsRacB, OsANTH3, OsAPT1 and RopGEF* which are defective in reproductive development in rice causing sterility (E.-J. Kim et al., 2020; S. K. Lee et al., 2021; Li et al., 2013; Liu et al., 2021; Ouyang et al., 2012; Xu et al., 2022; Yang Xu et al., 2017; Zheng et al., 2018). Although numerous genes have been

reported, there has been no previous transcriptome analysis on a large scale for controlling late pollen development in inter-subspecific crosses of rice.

5.1.2 Transcriptome analysis

RNA-Seq provides benefits over traditional expression profiling methods and enables a comprehensive assessment of global gene expression patterns (Marioni et al., 2008; Nagalakshmi et al., 2008). In our study, RNA-Seq revealed 79.15 million and 61.90 million raw reads from the HS and LS transcriptome libraries, respectively In a previous study, Illumina technology was used to sequence the transcriptomes of eight related developmental stages of flower and seed of three Poaceae species including Brachypodium, rice (*Oryza sativa*), and sorghum (*Sorghum bicolor*), were sequenced using Illumina technology (Davidson et al., 2012). They obtained 25,532, 40,331, and 34,497 genes in Brachypodium, rice and sorghum, respectively (Initiative, 2010; Paterson et al., 2004; Project, 2005). In this study, RNA-seq revealed a common 21,803 expressed genes in pistil pollinated of HS and LS group, while 901 and 1168 were specifically expressed for HS and LS respectively, which is consistent with previous the transcriptome in three Poaceae species. Several transcriptome analyses have been conducted to identify DEGs during pollen tube development across various plant species such as Arabidopsis, maize, papaveraceae and tobacco (Becker et al., 2014; Conze et al., 2017; Mazuecos-Aguilera & Suárez-Santiago, 2023; Warman et al., 2020). For instance, Wu et al. (2014) conducted comparative study involving cytological and transcriptomic analyses of pollen development in autotetraploid and diploid rice. Their research uncovered a total of 1,251 DEGs within the pollen transcriptome of autotetraploid rice, comprising 1,011 up-regulated genes and 240 down-regulated genes. Similarly, Conze et al. (2017) conducted RNA sequencing of pollen and pollen tubes in tobacco (*Nicotiana tabacum*), leading to the identification of 8,979 DEGs, with 1,577 up-regulated DEGs and 7,402 down-regulated DEGs. A total of 249 and 473 DEGs out of 23,872 expressed genes were specific for HS and LS respectively, suggesting that a small number of specific genes involved in seed setting rate in F_5 groups derived from the inter-subspecific crosses of rice. These suggesting that the number of expressed genes and DEGs identified in our study were different

than those in other researches, possibly due to the plant species, conditions and analytical methods.

5.1.3 The role of the dominant enrichment of candidate genes to spikelet fertility

The functional classification of the dominant enrichment for spikelet fertility genes in this study was the metabolic pathways. The metabolic pathways associated with spikelet fertility in rice are complex and involve a variety of biochemical processes that collectively contribute to the successful development of spikelets into fertile seeds (Alqudah et al., 2021). The metabolic pathways are crucial for providing energy, nutrients, and regulatory signals necessary for proper seed development (GournasMaarouf-Bouteau, 2022). There are several the metabolic pathways related to spikelet fertility in rice such as photosynthesis, carbohydrate metabolism, amino acid metabolism, hormone signaling, secondary metabolites, nutrient uptake and transport, redox and Reactive Oxygen Species (ROS) signaling (Chen et al., 2019; Chen et al., 2022; Gournas et al., 2008; Kovaleva & Zakharova, 2004; Scholz et al., 2020; Selinski & Scheibe, 2014; Wang et al., 2018). The specific genes and enzymes involved in these pathways contribute to the genetic regulation of spikelet fertility (Zhang et al., 2023). Understanding the metabolic pathways associated with spikelet fertility is essential for improving crop yield and quality through targeted breeding and biotechnological interventions. In this study, 11 DEGs specific for HS including *Os02g0134400*, *Os02g0467600*, *Os03g0321800*, *Os03g0368900*, *Os03g0382100*, *Os03g0670200*, *HTH1*, *Wx*, *Os06g0185100*, *Amy3A* and *Os11g0701100* were in the metabolic pathways. Some of these genes were previously reported to be involved in anthers fertility, pollen fertility, grain and yield production in rice such as *Amy3A, HTH1, Wx,* and *Os06g0185100* (Chen et al., 2018; Sera et al., 2019; Wang et al., 1995; Y. Xu et al., 2017). Furthermore, *Amy3A, HTH1*, and *Os03g0382100* appear to be co-expressed with genes related to various cellular processes, including pollen development (*Os01g0725900*), allergen production (*Os08g0302000*), and seed development (*Os03g0251000*) (Benitez et al., 2013; Hu et al., 2011). This co-expression with pollen development genes suggests their potential involvement in these processes.

Transcription factors (TFs) play an important role in regulating gene expression in spikelet fertility. In this study, the co-expression network showed *Os02g0467600* (similar to cinnamate 4-hydroxylase CYP73) co-expression with AP2-EREBP TF. AP2-EREBP TF had been reported to be the part of gene regulatory networks and integrate metabolic, hormonal and environmental signals in stress (Dietz et al., 2010). Pande et al. (2018) reported that AP2/EREBP family also regulates diverse processes of plant development and metabolism such as vegetative and reproductive development, cell proliferation, secondary metabolism and responses to different plant hormones. Cinnamate 4-hydroxylase CYP73 are required for downstream phenylpropanoid production including of lignin and flavonoid biosynthesis (Ehlting et al., 2006; J. I. Kim et al., 2021; Lucheta et al., 2007). Lignin, a complex polymer located in the cell walls of plants, is essential for reinforcing and fortifying these walls. It serves a versatile function in protecting plants against a range of threats and environmental adversities. Flavanoids serve a multitude of functions, including the regulation of cellular growth, the attraction of pollinating insects, and the provision of protection against biotic and abiotic stresses (Dias et al., 2021). Muhlemann et al. (2018) reported that flavanols play a role in preserving the growth and structural stability of pollen tubes through the regulation of reactive oxygen species (ROS) balance when exposed to elevated temperature stress in tomato. However, *Os02g0467600* (similar to cinnamate 4-hydroxylase CYP73) had not been previously reported to be involved in spikelet fertility in rice by controlling pollen tube growth in rice pistils. Our results provided valuable information for understanding the molecular mechanism involving spikelet fertility in rice.

The qRT-PCR analysis of *Os02g0467600* before pollination showed low levels of expression, at a similar level, in both HS and LS pistils. However, following pollination for 4-5 hours, the expressions of *Os02g0467600* were significantly higher in HS pistils than that in the LS pistils, suggesting the role of this gene after pollination such as, interaction for germination and growth of pollen tubes. Furthermore, we found that *Os02g0467600* were colocalized with 2 QTLs associated with seed setting rate and spikelet fertility including *sf2* and *OsPLL3* on chromosome 2, confirming that this gene is important for spikelet fertility.

5.1.4 The complex network of candidate genes to irregulate in pollen tube growth state in rice

The complex network of genes involved in the irregularity of pollen tube growth in rice consists of a diverse set of genes that affect various aspects of pollen tube development and growth (Scholz et al., 2020). In this study, we identified six DEGs that were significantly enriched in the pentose and glucuronate interconversions pathway and the flavonoid biosynthesis pathway. *OsPME10*, *OsPME35*, *OsPLL3* and *OsPLL9* were significantly enriched in the pentose and glucuronate interconversions pathway, demonstrating the highest fold enrichment under up-regulation. Contrastingly, in the context of flavonoid biosynthesis, the genes *Os02g0767300* encoding flavonol synthase/flavanone 3-hydroxylase and *Os04g0101400* encoding flavone synthase II (FNSII) exhibited down-regulation.

In this study, we identified *OsPLL3* as the functional DEGs specific to LS. Recent studies have similarly shown that *OsPLL3* and *OsPLL9* play crucial roles in rice panicle development and are involved in male sterility (Zheng et al., 2018). Several studies have reported the importance of PME and PMEI in coordinating wall loosening and wall stiffening, affecting both the growth pulse and growth retardation of pollen tubes (Bosch & Hepler, 2005). In Arabidopsis, reduced PME activities have been shown to result in severe defects in pollen tube growth and morphology (Y.-J. Kim et al., 2020). Mutation of the pollen-specific Arabidopsis PME genes has been reported. The mutation *vgd1* resulted in an unstable pollen tube burst, while the lack of *AtPPME1* resulted in an altered shape of the pollen tube (Jiang et al., 2005; Tian et al., 2006). *AtPME48* was shown to be important for pollen intine layer during pollen maturation (Leroux et al., 2015). In maize (*Zea mays* L.), three PME proteins (ZmPME3, ZMGa1P, and ZmPME-10-1) have also been reported to exhibit self-incompatibility (Moran Lauter et al., 2017). The exogenous treatment of PME or pectinase on a pollen germination medium resulted in abnormal growth of the pollen tube in potato (*Solanum chacoense*) (Parre & Geitmann, 2005). In angiosperms, the enzymes responsible for transition are PMEs, that cleave the ester bonds in the secreted galacturonan monomers, enabling Ca^{2+}-crosslinking of the pectin chains. It was regulated by the complex process of pollen tip growth including cell organization, cell wall integrity, secretion, and signaling networks (Scholz et al., 2020). In rice (*Oryza sativa* L.), although Nguyen et

al. (2016) reported that the rice genome contains 43 PMEs and 49 PMEIs. Only one pollen-specific gene, *OsPME1*, has been reported to be regulated by various abiotic stresses and hormones in rice (Z. Zhang et al., 2018). In this study, we found that two DEGs (*OsPME 10* and *OsPME 35*) had not been previously reported to be associated with the pollen tube growth state in rice.

OsPME10, OsPME35, OsPLL3, OsPLL9, Os02g0767300, and *Os04g0101400* were LS-specific DEGs involving irregulated pollen tube growth with low seed setting rate in this study. The co-expression network revealed that *OsPME10, OsPME35,* and *OsPLL3* were co-expressed with a distinct set of 14 genes. Nevertheless, *OsPLL9* was separated from a cluster of *OsPME10, OsPME35,* and *OsPLL3*. The qRT-PCR analysis of *OsPME35* and *OsPLL9* before and after pollination using qRT-PCR showed that *OsPME35* and *OsPLL9* exhibited a similar expression pattern, by showing much higher expression in the HS pistils than that in LS before pollination, but at 4-5 hours after pollination their expression were very low in the both pistils, suggesting the role of this gene beginning before pollination. These results uncovered deeper mechanisms wherein *OsPME35* and *OsPLL9* are required for the style cell wall, potentially modifying the cell wall during pollen-style interaction. Among the DEGs within flavonoid biosynthesis, *Os02g0767300*, specific for LS, is shown to share the same pathway with *Os02g0467600* (similar to cinnamate 4-hydroxylase CYP73) specific for HS (*Os02g0467600*). It was also predicted to interact with *Os04g0137100* (*OsPLL4*) encoding pectate lyase in the co-expression study. Previously, Zheng et al. (2018) reported that knockdowns of *OsPLL3* and *OsPLL4* by artificial microRNA (amiRNA) disrupted normal pollen development and resulted in partial male sterility.

By utilizing the previously identified QTLs involving seed setting, spikelet fertility, and genes related to pollen tube growth, we were able to co-localize the DEGs identified in this RNA-Seq dataset. F. Zhang et al. (2018) searching the functions of DEGs by co-localization with previous identified QTL, the result showed that chromosome 2 had the greatest number co-localized DEGs with 11 QTLs responsible for panicle traits. Similarly, our results showed that chromosome 2 had the greatest number of DEGs with QTL and functional genes associated with spikelet fertility. One such gene was *Os02g0467600* (similar to cinnamate 4-hydroxylase CYP73), which co-localized with sf2 and *OsPLL3*. Furthermore, on chromosome 12, *OsPME35* was found

to be closely associated with gpp12.1, a locus that impacts the number of grains per panicle. These results suggest the roles of these genes in spikelet fertility in rice.

In summary, genes related to spikelet fertility in rice play a crucial role in the successful formation of grains within the spikelet. In our study, we discovered that *Os02g0467600* (similar to cinnamate 4-hydroxylase CYP73) had not previously been reported to be involved in rice spikelet fertility, specifically by controlling pollen tube growth in rice pistils. *OsPME10* and *OsPME35* were also found to be associated with the pollen tube growth state in rice, a novel discovery. Furthermore, we unveiled a new mechanism involving *OsPME35* and *OsPLL9*, which was identified before pollination, suggesting that the style potentially modifies the cell wall during pollen-style interaction, while *Os02g0467600* (similar to cinnamate 4-hydroxylase CYP73) appears to control later processes, such as fertilization and seed development. Furthermore, *Os02g0467600* (similar to cinnamate 4-hydroxylase CYP73) and *OsPME35* co-localized with QTLs and functional genes associated with spikelet fertility and the number of grains per panicle, respectively. Manipulating these genes through breeding or biotechnology approaches can potentially lead to the development of rice varieties with higher yields and improved agricultural productivity.

5.2 Conclusions

A small number of spikelet fertility differentially expressed genes were observed in specific HS that generated from inter-subspecific crosses as compared to specific LS. The transcriptome analysis of pollinated pistils presented in this study expands our understanding of the potential pathway and candidate genes involving spikelet fertility by identifying differentially expressed genes in the metabolic process. Furthermore, only eleven DEGs of specific HS were significantly enriched in metabolic pathways. These may be the candidate genes and potential pathway involving spikelet fertility. However, six DEGs of specific LS having the highest fold enriched to carbohydrate metabolism pathway that included pentose and glucuronate interconversions and flavonoid biosynthesis were also significantly enriched in our data, implying that these processes may play important roles in a striking irregularity in pollen development caused to spikelet sterility in rice. The co-expression network may help to understanding of the molecular mechanisms involved in spikelet fertility and provides an essential knowledge base from which to design further studies. By previously identified seed setting rate, spikelet fertility and the number of grains per panicle involving pollen tube growth and fine-mapping QTLs, the co-localized DEGs detected in this RNA-Seq data. This knowledge empowers researchers to develop innovative strategies for crop improvement, conservation, and sustainable agriculture.

REFERENCES

Books and Book Articles

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. & Walter, P. (2002). The shape and structure of proteins. In *Molecular Biology of the Cell. 4th edition*. Garland Science.

Brar, D. S., Dalmacio, R., Elloran, R., Aggarwal, R., Angeles, R., & Khush, G. S. (1996). Gene transfer and molecular characterization of introgression from wild *Oryza* species into rice. In *Rice Genetics III: (In 2 Parts)* (pp. 477-486). Bryman, A. & Cramer, D. (2009). *Quantitative data analysis with SPSS 14, 15 & 16: A guide for social scientists*. Routledge/Taylor & Francis Group.

Chang, T. T., & Bardenas, E. A. (1965). *The morphology and varietal characteristics of the rice plant* (Vol. 4). Int. Rice Res. Inst.

Institute, I. P. G. R. & Association, W. A. R. D. (2007). *Descriptors for Wild and Cultivated Rice (Oryza Spp.)*. Biodiversity International.

Oka, H. I. (2012). *Origin of cultivated rice*. Elsevier.

Chandler Jr, R. F. (1975). Case history of IRRI's research management during the period 1960 to 1972. *Asian Vegetable Research and Development Center. Republic of China*.

Vijay, D., & Roy, B. (2013). Chapter-4 Rice (*Oryza Sativa* L.). *Breeding, Biotechnology and Seed Production of Field Crops (December)*, 71-122.

Wang, H.-J., Huang, J.-C. & Jauh, G.-Y. (2010). Pollen germination and tube growth. In *Advances in Botanical Research* (Vol. 54, pp. 1-52). Elsevier.

Articles

Alqudah, A. M., Sharma, R. & Börner, A. (2021). Insight into the genetic contribution of maximum yield potential, spikelet development and abortion in barley. *Plants. People. Planet., 3*(6), 721-736.

Arunachalam, V. & Bandyopadhyay, A. (1984). Limits to genetic divergence for occurrence of heterosis-experimental evidence from crop plants. *Indian. J. Genet. Plant. Breed., 54*, 548-554.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S. & Eppig, J. T. (2000). Gene ontology: tool for the unification of biology. *Nat. Genet., 25*(1), 25-29.

Ayub, M., Hardwick, S. W., Luisi, B. F. & Bayley, H. (2013). Nanopore-based identification of individual nucleotides for direct RNA sequencing. *Nano. Lett., 13*(12), 6144-6150.

Beche, E., da Silva, C. L., Pagliosa, E. S., Capelin, M. A., Franke, J., Matei, G. & Benin, G. (2013). Hybrid performance and heterosis in early segregant populations of Brazilian spring wheat. *Aust. J. Crop. Sci., 7*(1), 51-57.

Becker, J. D., Takeda, S., Borges, F., Dolan, L. & Feijó, J. A. (2014). Transcriptional profiling of *Arabidopsis* root hairs and pollen defines an apical cell growth signature. *BMC Plant. Biol., 197*, 1-14.

Bedinger, P. (1992). The remarkable biology of pollen. *Plant. Cell., 4*(8), 879.

Benitez, L. C., da Maia, L. C., Ribeiro, M. V., Pegoraro, C., Peters, J. A., de Oliveira, A. C. & Braga, E. J. (2013). Salt induced change of gene expression in salt sensitive and tolerant rice species. *J. Agric. Sci., 5*(10), 251.

Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc., 57*(1), 289-300.

Birchler, J. A. (2015). Heterosis: The genetic basis of hybrid vigour. *Nat. Plants., 1*(3), 1-2.

Bosch, M. & Hepler, P. K. (2005). Pectin methylesterases and pectin dynamics in pollen tubes. *Plant. Cell., 17*(12), 3219-3226.

Branton, D., Deamer, D. W., Marziali, A., Bayley, H., Benner, S. A., Butler, T., Di Ventra, M., Garaj, S., Hibbs, A. & Huang, X. (2008). The potential and challenges of nanopore sequencing. *Nat. Biotechnol., 26*(10), 1146-1153.

Cazzola, F., Bermejo, C. J. & Cointry, E. (2020). Transgressive segregations in two pea F_2 populations and their respective $F_{2:3}$ families. *Pesqui. Agropecu. Bras., 55*, e01623.

Chan, P. P. & Lowe, T. M. (2016). GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic. Acids. Res., 44*(1), 184-189.

Chao, J., Li, Z., Sun, Y., Aluko, O. O., Wu, X., Wang, Q. & Liu, G. (2021). MG2C: a user-friendly online tool for drawing genetic maps. *Hortic. Res., 1*(1), 16.

Chelong, I.-a. & Sdoodee, S. (2012). Pollen viability, pollen germination and pollen tube growth of shogun (*Citrus reticulate Blanco*) under climate variability in southern Thailand. *J. Agric. Technol., 8*, 2297-2307.

Chen, H., Zhao, Z., Liu, L., Kong, W., Lin, Y., You, S., Bai, W., Xiao, Y., Zheng, H. & Jiang, L. (2017). Genetic analysis of a hybrid sterility gene that causes both pollen and embryo sac sterility in hybrids between *Oryza sativa* L. and *Oryza longistaminata*. *Heredity, 119*(3), 166-173.

Chen, L., Bian, J., Shi, S., Yu, J., Khanzada, H., Wassan, G. M., Zhu, C., Luo, X., Tong, S., Yang, X., Peng, X., Yong, S., Yu, Q., He, X., Fu, J., Chen, X., Hu, L., Ouyang, L. & He, H. (2018). Genetic analysis for the grain number heterosis of a super-hybrid rice *WFYT025* combination using RNA-Seq. *Rice 11*(1), 37.

Chen, L., Yuan, Y., Wu, J., Chen, Z., Wang, L., Shahid, M. Q. & Liu, X. (2019). Carbohydrate metabolism and fertility related genes high expression levels promote heterosis in autotetraploid rice harboring double neutral genes. *Rice, 12*(1), 34.

Chen, L., Zhang, Y.-H., Wang, S., Zhang, Y., Huang, T. & Cai, Y.-D. (2017). Prediction and analysis of essential genes using the enrichments of gene ontology and KEGG pathways. *PLOS ONE, 12*(9), e0184129.

Chen, S.-q., Zhong, W., Liu, M.-x., Xie, Z.-w. & Wang, H.-h. (2008). Pollen grain germination and pollen tube growth in pistil of rice. *Rice. Sci., 15*(2), 125-130.

Chen, T., Ma, J., Xu, C., Jiang, N., Li, G., Fu, W., Feng, B., Wang, D., Wu, Z., Tao, L. & Fu, G. (2022). Increased ATPase activity promotes heat-resistance, high-yield, and high-quality traits in rice by improving energy status. *Front. Plant. Sci., 13*, 1035027.

Chutimanukul, P., Saputro, T. B., Mahaprom, P., Plaimas, K., Comai, L., Buaboocha, T., Siangliw, M., Toojinda, T. & Chadchawan, S. (2021). Combining genome and gene co-expression network analyses for the identification of genes potentially regulating salt tolerance in rice. *Front. Plant. Sci., 12,* 704549.

Clancy, S. & Brown, W. (2008). Translation: DNA up mRNA to Protein. *Nat. Sci. Educ., 1*(1), 101.

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M. W., Gaffney, D. J., Elo, L. L. & Zhang, X. (2016). A survey of best practices for RNA-seq data analysis. *Genome. Biol., 17*(1), 1-19.

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

Crick, F. (1970). Central dogma of molecular biology. *Nature, 227*(5258), 561-563.

Dametto, A., Sperotto, R. A., Adamski, J. M., Blasi, É. A., Cargnelutti, D., de Oliveira, L. F., Ricachenevsky, F. K., Fregonezi, J. N., Mariath, J. E. & da Cruz, R. P. (2015). Cold tolerance in rice germinating seeds revealed by deep RNAseq analysis of contrasting *indica* genotypes. *Plant. Sci., 238,* 1-12.

Davidson, R. M., Gowda, M., Moghe, G., Lin, H., Vaillancourt, B., Shiu, S. H., Jiang, N. & Robin Buell, C. (2012, Aug). Comparative transcriptomics of three Poaceae species reveals patterns of gene expression evolution. *Plant. J., 71*(3), 492-502.

Dias, M. C., Pinto, D. C. & Silva, A. M. (2021). Plant flavonoids: Chemical characteristics and biological activity. *Molecules, 26*(17), 5377.

Dietz, K. J., Vogel, M. O. & Viehhauser, A. (2010). AP2/EREBP transcription factors are part of gene regulatory networks and integrate metabolic, hormonal and environmental signals in stress acclimation and retrograde signalling. *Protoplasma, 245*(1-4), 3-14.

Dittmar, K. A., Goodenbour, J. M. & Pan, T. (2006). Tissue-specific differences in human transfer RNA expression. *PLoS. Genet., 2*(12), e221.

Ehlting, J., Hamberger, B., Million-Rousseau, R. & Werck-Reichhart, D. (2006). Cytochromes P450 in phenolic metabolism. *Phytochem. Rev., 5,* 239-270.

Fahad, S., Hussain, S., Saud, S., Tanveer, M., Bajwa, A. A., Hassan, S., Shah, A. N., Ullah, A., Wu, C. & Khan, F. A. (2015). A biochar application protects rice pollen from high-temperature stress. *Plant. Physiol. Biochem., 96,* 281-287.

Fiebig, A., Mayfield, J. A., Miley, N. L., Chau, S., Fischer, R. L. & Preuss, D. (2000). Alterations in *CER6,* a gene identical to *CUT1*, differentially affect long-chain lipid content on the surface of pollen and stems. *Plant. Cell., 12*(10), 2-47.

Fu, D., Xiao, M., Hayward, A., Fu, Y., Liu, G., Jiang, G. & Zhang, H. (2014). Utilization of crop heterosis: a review. *Euphytica, 197*(2), 161-173.

Gao, Y., Du, L., Ma, Q., Yuan, Y., Liu, J., Song, H. & Feng, B. (2022). Conjunctive analyses of bulk segregant analysis sequencing and bulk segregant RNA sequencing to identify candidate genes controlling spikelet sterility of Foxtail millet. *Front. Plant. Sci., 13*, 842336.

Ge, S. X., Jung, D. & Yao, R. (2020). ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics, 36*(8), 2628-2629.

Gournas, C., Papageorgiou, I. & Diallinas, G. (2008). The nucleobase–ascorbate transporter (NAT) family: genomics, evolution, structure–function relationships and physiological role. *Mol. Biosyst., 4*(5), 404-416.

GournasMaarouf-Bouteau, H. (2022, Jan 20). The seed and the metabolism regulation. *Biology (Basel), 11*(2), 168.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R. & Zeng, Q. (2011). Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol., 29*(7), 644.

Guo, H., Mendrikahy, J. N., Xie, L., Deng, J., Lu, Z., Wu, J., Li, X., Shahid, M. Q. & Liu, X. (2017). Transcriptome analysis of neo-tetraploid rice reveals specific differential gene expressions associated with fertility and heterosis. *Sci. Rep., 7*(1), 1-11.

Guo, J., Xu, X., Li, W., Zhu, W., Zhu, H., Liu, Z., Luan, X., Dai, Z., Liu, G. & Zhang, Z. (2016). Overcoming inter-subspecific hybrid sterility in rice by developing *indica*-compatible *japonica* lines. *Sci. Rep., 6*(1), 1-9.

Guo, J.-X. & Liu, Y.-G. (2012). Molecular control of male reproductive development and pollen fertility in rice. *J. Integr. Plant. Biol., 54*(12), 967-978.

Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. (2012). CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell. Rep., 2*(3), 666-673.

Hernandez-Alias, X., Katanski, C. D., Zhang, W., Assari, M., Watkins, C. P., Schaefer, M. H., Serrano, L. & Pan, T. (2023). Single-read tRNA-seq analysis reveals coordination of tRNA modification and aminoacylation and fragmentation. *Nucleic. Acids. Res., 51*(3), e17-e17.

Heslop-Harrison, J. (1979). Supplement 1: Aspects of the Structure, Cytochemistry and Germination of the Pollen of Rye. *Ann. Bot, 44*, 2-47.

Hong, W. J., Lee, S. K., Kim, S. H., Kim, Y. J., Moon, S., Kim, E. J., Silva, J. & Jung, K. H. (2022). Comparative transcriptome analysis of pollen and anther wall reveals novel insights into the regulatory mechanisms underlying anther wall development and its dehiscence in rice. *Plant. Cell. Rep., 41*(5), 1229-1242.

Hrdlickova, R., Toloue, M. & Tian, B. (2017). RNA-Seq methods for transcriptome analysis. *Wiley Interdiscip. Rev. RNA., 8*(1), e1364.

Hu, L., Liang, W., Yin, C., Cui, X., Zong, J., Wang, X., Hu, J. & Zhang, D. (2011). Rice MADS3 regulates ROS homeostasis during late anther development. *Plant. Cell., 23*(2), 515-533.

Hu, Q., Wang, W., Lu, Q., Huang, J., Peng, S. & Cui, K. (2021). Abnormal anther development leads to lower spikelet fertility in rice (*Oryza sativa* L.) under high temperature during the panicle initiation stage. *BMC. Plant. Biol., 21*(1), 1-17.

Huang, X., Peng, X. & Sun, M. X. (2017). *OsGCD1* is essential for rice fertility and required for embryo dorsal-ventral pattern formation and endosperm development. *New. Phytol., 215*(3), 1039-1058.

Hwang, B., Lee, J. H. & Bang, D. (2018). Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp. Mol. Med., 50*(8), 1-14.

Initiative, I. B. (2010). Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature, 463*(7282), 763-768.

Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., Lönnerberg, P. & Linnarsson, S. (2014). Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods., 11*(2), 163-166.

Itoh, J.-I., Nonomura, K.-I., Ikeda, K., Yamaki, S., Inukai, Y., Yamagishi, H., Kitano, H. & Nagato, Y. (2005). Rice plant development: from zygote to spikelet. *Plant. Cell. Physiol., 46*(1), 23-47.

Ji, Q., Lu, J., Chao, Q., Zhang, Y., Zhang, M., Gu, M. & Xu, M. (2010). Two sequence alterations, a 136 bp InDel and an A/C polymorphic site, in the *S5* locus are associated with spikelet fertility of *indica-japonica* hybrid in rice. *J. Genet. Genom., 37*(1), 57-68.

Jiang, C., Wang, Y., Zhou, J., Rashid, M. A. R., Li, Y., Peng, Y., Xie, L., Zhou, G., He, Y., Sun, W., Zheng, C. & Xie, X. (2023). Genome-Wide Scan for Genetic Signatures Based on the Whole-Genome Resequencing of Salt- and Drought-Tolerant Rice Varieties. *Agronomy, 13*(7), 1936.

Jiang, L., Yang, S.-L., Xie, L.-F., Puah, C. S., Zhang, X.-Q., Yang, W.-C., Sundaresan, V. & Ye, D. (2005). VANGUARD1 encodes a pectin methylesterase that enhances pollen tube growth in the *Arabidopsis* style and transmitting tract. *Plant. Cell., 17*(2), 584-596.

Jiangtao, C., Yingzhen, K., Qian, W., Yuhe, S., Daping, G., Jing, L. & Guanshan, L. (2015). MapGene2Chrom, a tool to draw gene physical map based on Perl and SVG languages. *Yi. Chuan., 37*(1), 91-97.

Julius, M., Masuda, T. & Herzenberg, L. (1972). Demonstration that antigen-binding cells are precursors of antibody-producing cells after purification with a fluorescence-activated cell sorter. *Proc. Natl. Acad. Sci., 69*(7), 1934-1938.

Kamara, N., Jiao, Y., Lu, Z., Aloryi, K. D., Wu, J., Liu, X. & Shahid, M. Q. (2021). Cytological observations and bulked-segregant analysis coupled global genome sequencing reveal two genes associated with pollen fertility in tetraploid rice. *Int. J. Mol. Sci., 22*(2), 841.

Kanehisa, M. & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic. Acids. Res., 28*(1), 27-30.

Kashima, M., Sakamoto, R. L., Saito, H., Ohkubo, S., Tezuka, A., Deguchi, A., Hashida, Y., Kurita, Y., Iwayama, K. & Adachi, S. (2021). Genomic basis of transcriptome dynamics in rice under field conditions. *Plant. Cell. Physiol., 62*(9), 1436-1445.

Katara, J. L., Verma, R. L., Parida, M., Ngangkham, U., Molla, K. A., Barbadikar, K. M., Mukherjee, M., Samantaray, S., Ravi, N. R. & Singh, O. N. (2020). Differential expression of genes at panicle initiation and grain filling stages implied in heterosis of rice hybrids. *Int. J. Mol. Sci, 21*(3), 1080.

Kennedy, G., Burlingame, B. & Nguyen, N. (2002). Nutrient impact assessment of rice in major rice consuming countries. *Korea, 165*(23.3), 12.15.

Khlaimongkhon, S., Chakhonkaen, S., Tongmark, K., Sangarwut, N., Panyawut, N., Wasinanon, T., Sikaewtung, K., Wanchana, S., Mongkolsiriwatana, C. & Chunwonges, J. (2021). RNA sequencing reveals rice genes involved in male reproductive development under temperature alteration. *Plants, 10*(4), 663.

Khush, G. S. (1997). Origin, dispersal, cultivation and variation of rice. *Plant. Mol. Biol., 35*(1), 25-34.

Khush, G. S. (2001). Green revolution: the way forward. *Nat. Rev. Genet., 2*(10), 815-822.

Khush, G. S. (2005). What it will take to feed 5.0 billion rice consumers in 2030. *Plant. Mol. Biol., 59*(1), 1-6.

Kim, E.-J., Kim, J.-H., Hong, W.-J., Kim, E. Y., Kim, M.-H., Lee, S. K., Min, C. W., Kim, S. T., Park, S. K., Jung, K.-H. & Kim, Y.-J. (2023). Rice pollen-specific *OsRALF17* and *OsRALF19* are essential for pollen tube growth. *J. Integr. Plant. Biol., 65*(9), 2218-2236.

Kim, E.-J., Park, S.-W., Hong, W.-J., Silva, J., Liang, W., Zhang, D., Jung, K.-H. & Kim, Y.-J. (2020). Genome-wide analysis of *RopGEF* gene family to identify genes contributing to pollen tube growth in rice (*Oryza sativa*). *BMC Plant. Biol., 20*(1), 1-15.

Kim, J. I., Hidalgo-Shrestha, C., Bonawitz, N. D., Franke, R. B. & Chapple, C. (2021). Spatio-temporal control of phenylpropanoid biosynthesis by inducible complementation of a cinnamate 4-hydroxylase mutant. *J. Exp. Bot, 72*(8), 3061-3073.

Kim, S., Cho, K., Lim, S.-H., Goo, T.-W. & Lee, J.-Y. (2021). Transcriptome profiling of transgenic rice seeds lacking seed storage proteins (globulin, prolamin, and glutelin) by RNA-Seq analysis. *Plant. Biotechnol. Rep, 15*(1), 77-93.

Kim, Y.-J., Jeong, H. Y., Kang, S.-Y., Silva, J., Kim, E.-J., Park, S. K., Jung, K.-H. & Lee, C. (2020). Physiological importance of pectin modifying genes during rice pollen development. *Int. J. Mol. Sci., 21*(14), 4840.

Kim, Y.-J., Zhang, D. & Jung, K. H. (2019, Dec). Molecular basis of pollen germination in cereals. *Trends. Plant. Sci., 24*(12), 1126-1136.

Koch, C. M., Chiu, S. F., Akbarpour, M., Bharat, A., Ridge, K. M., Bartom, E. T. & Winter, D. R. (2018). A beginner's guide to analysis of RNA sequencing data. *Am. J. Respir. Cell. Mol. Biol., 59*(2), 145-157.

Kovaleva, L. & Zakharova, E. (2004). Gametophyte–sporophyte interactions in the pollen–pistil system: 4. the hormonal status and the mechanism of self-incompatibility. *Russ. J. Plant. Physiol., 51*(3), 402-406.

Kukurba, K. R. & Montgomery, S. B. (2015). RNA sequencing and analysis. *Cold. Spring. Harb. Protoc., 2015*(11), 951-969.

Kumar, R., Ichihashi, Y., Kimura, S., Chitwood, D. H., Headland, L. R., Peng, J., Maloof, J. N. & Sinha, N. R. (2012). A high-throughput method for Illumina RNA-Seq library preparation. *Front. Plant. Sci., 3*, 202.

Kuroha, T., Nagai, K., Kurokawa, Y., Nagamura, Y., Kusano, M., Yasui, H., Ashikari, M. & Fukushima, A. (2017). eQTLs regulating transcript variations associated with rapid internode elongation in deepwater rice. *Front. Plant. Sci., 8*, 1753.

Kwon, C.-T., Kim, S.-H., Kim, D. & Paek, N.-C. (2015). The rice floral repressor Early flowering1 affects spikelet fertility by modulating gibberellin signaling. *Rice, 8*(1), 1-11.

Lee, C.-M., Suh, J.-P., Park, H.-S., Baek, M.-K., Jeong, O. Y., Yun, S.-J., Cho, Y.-C. & Kim, S.-M. (2021). Identification of QTL combinations that cause spikelet sterility in rice derived from interspecific crosses. *Rice, 14*(1), 99.

Lee, S. K., Hong, W.-J., Silva, J., Kim, E.-J., Park, S. K., Jung, K.-H. & Kim, Y.-J. (2021). Global identification of *ANTH* genes involved in rice pollen germination and functional characterization of a key member, *OsANTH3*. *Front. Plant. Sci., 12*, 609473.

Lenka, S. K., Katiyar, A., Chinnusamy, V. & Bansal, K. C. (2011). Comparative analysis of drought-responsive transcriptome in Indica rice genotypes with contrasting drought tolerance. *Plant. Biotechnol. J., 9*(3), 315-327.

Leroux, C., Bouton, S., Kiefer-Meyer, M.-C., Fabrice, T. N., Mareck, A., Guénin, S., Fournet, F., Ringli, C., Pelloux, J. & Driouich, A. (2015). PECTIN METHYLESTERASE48 is involved in *Arabidopsis* pollen grain germination. *Plant. Physiol., 167*(2), 367-380.

Li, S., Li, W., Huang, B., Cao, X., Zhou, X., Ye, S., Li, C., Gao, F., Zou, T. & Xie, K. (2013). Natural variation in *PTB1* regulates rice seed setting rate by controlling pollen tube growth. *Nat. Commun., 4*(1), 1-13.

Li, W., Lin, Z. & Zhang, X. (2007). A novel segregation distortion in intraspecific population of Asian cotton (*Gossypium arboretum* L.) detected by molecular markers. *J. Genet. Genomics., 34*(7), 634-640.

Li, Z. (1980). A preliminary discussion about the classification of male sterile lines of rice in China. *Acta. Agron. Sin., 6*, 17-26.

Liang, T., XU, Z.-j. & CHEN, W.-f. (2017). Advances and prospects of super rice breeding in China. *J. Integr. Agric., 16*(5), 984-991.

Liu, K., Chen, Y., Huang, J., Qiu, Y., Li, S., Zhuo, X., Yu, F., Gao, J., Li, G. & Zhang, W. (2022). Spikelet differentiation and degeneration in rice varieties with different panicle sizes. *Food. Energy. Secur., 11*(1), e320.

Liu, S., Zhong, J., Ling, S., Liu, Y., Xu, Y. & Yao, J. (2021). *OsAPT1* a pollen preferentially expressed gene is essential for pollen tube germination and elongation in rice. *Plant. Mol. Biol. Rep., 39*(1), 87-97.

Liu, X., Qu, X., Jiang, Y., Chang, M., Zhang, R., Wu, Y., Fu, Y. & Huang, S. (2015). Profilin regulates apical actin polymerization to control polarized pollen tube growth. *Mol. Plant., 8*(12), 1694-1709.

Liu, X. C. & Wu, J. L. (1998). SSR heterogenic patterns of parents for marking and predicting heterosis in rice breeding. *Mol. Breed., 4*, 263-268.

Lord, E. (2001). Adhesion molecules in lily pollination. *Sex. Plant. Reprod, 14*, 57-62.

Lucheta, A. R., Silva-Pinhati, A. C. O., Basílio-Palmieri, A. C., Berger, I. J., Freitas-Astúa, J. & Cristofani, M. (2007). An in silico analysis of the key genes involved in flavonoid biosynthesis in *Citrus sinensis*. *Genet. Mol. Biol, 30*, 819-831.

Marguerat, S. & Bähler, J. (2010). RNA-seq: from technology to biology. *Cell. Mol. Life Sci, 67*(4), 569-579.

Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M. & Gilad, Y. (2008). RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome. Res., 18*(9), 1509-1517.

Matsui, T. (2005). Function of long basal dehiscence of the thecae in rice (*Oryza sativa* L.) pollination under hot and humid condition. *Phyton, 45*, 401-407.

Matsui, T., Omasa, K. & Horie, T. (1999). Mechanism of anther dehiscence in rice (*Oryza sativa* L.). *Ann. Bot., 84*(4), 501-506.

Matsuo, T. (1952). Genecological studies on cultivated rice. II. Varietial differences in damages by the rice stem borer. *Bull. natn. Inst, agric. Sei. D., 3*, 30-39.

Mazuecos-Aguilera, I. & Suárez-Santiago, V. N. (2023). Identification of candidate genes involved in the determinism of pollen grain aperture morphology by comparative transcriptome analysis in Papaveraceae. *Plants, 12*(7), 1570.

Montgomery, S. B., Sammeth, M., Gutierrez-Arcelus, M., Lach, R. P., Ingle, C., Nisbett, J., Guigo, R. & Dermitzakis, E. T. (2010). Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature, 464*(7289), 773-777.

Moran Lauter, A. N., Muszynski, M. G., Huffman, R. D. & Scott, M. P. (2017). A pectin methylesterase *ZmPme3* is expressed in gametophyte factor1-s (Ga1-s) silks and maps to that locus in maize (*Zea mays* L.). *Front. Plant. Sci., 8*, 1926.

Morin, R. D., Aksay, G., Dolgosheina, E., Ebhardt, H. A., Magrini, V., Mardis, E. R., Sahinalp, S. C. & Unrau, P. J. (2008). Comparative analysis of the small RNA transcriptomes of Pinus contorta and *Oryza sativa*. *Genome. Res., 18*(4), 571-584.

Morozova, O. & Marra, M. A. (2008). Applications of next-generation sequencing technologies in functional genomics. *Genomics*, *92*(5), 255-264.

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods., 5*(7), 621-628.

Muhlemann, J. K., Younts, T. L. & Muday, G. K. (2018). Flavonols control pollen tube growth and integrity by regulating ROS homeostasis during high-temperature stress. *Proc. Natl. Acad. Sci, 115*(47), E11188-E11197.

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M. & Snyder, M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*, *320*(5881), 1344-1349.

Nevhudzholi, K. M., Gwata, E. T. & McHau, G. R. A. (2020). A genetic study of pod shattering resistance in F2 soybean (*Glycine max*) germplasm. *S. Afr. J. Plant. Soil., 37*(2), 174-176.

Nguyen, H. P., Jeong, H. Y., Kim, H., Kim, Y. C. & Lee, C. (2016). Molecular and biochemical characterization of rice pectin methylesterase inhibitors (*OsPMEIs*). *Plant. Physiol. Biochem., 101*, 105-112.

Nivala, J., Marks, D. B. & Akeson, M. (2013). Unfoldase-mediated protein translocation through an α-hemolysin nanopore. *Nat. Biotechnol., 31*(3), 247-250.

Obayashi, T., Aoki, Y., Tadaka, S., Kagaya, Y. & Kinoshita, K. (2018). ATTED-II in 2018: a plant coexpression database based on investigation of the statistical property of the mutual rank index. *Plant. Cell. Physiol., 59*(1), e3-e3.

Obayashi, T. & Kinoshita, K. (2009). Rank of correlation coefficient as a comparable measure for biological significance of gene coexpression. *DNA research, 16*(5), 249-260.

Oka, H.-I. & Morishima, H. (1967). Variations in the Breeding Systems of a Wild Rice, *Oryza perennis*. *Evolution, 21*(2), 249-258.

Ouyang, Y., Huang, X., Lu, Z. & Yao, J. (2012). Genomic survey, expression profile and co-expression network analysis of *OsWD40* family in rice. *BMC Genom., 13*, 100.

Pachter, L. (2011). Models for transcript quantification from RNA-Seq. *arXiv preprint arXiv:1104.3889*.

Pande, A., Saxena, S., Thapliyal, M., Guru, S., Kumar, A. & Arora, S. (2018). Role of *AP2/EREBP* transcription factor family in environmental stress tolerance. *Life Sci. J., 3*, 1-9.

Parkhomchuk, D., Borodina, T., Amstislavskiy, V., Banaru, M., Hallen, L., Krobitsch, S., Lehrach, H. & Soldatov, A. (2009). Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic. Acids. Res., 37*(18), e123-e123.

Parnell, J. A. (2011). Strategic capabilities, competitive strategy, and performance among retailers in Argentina, Peru and the United States. *Manag. Decis., 49*(1), 139-155.

Parre, E. & Geitmann, A. (2005). Pectin and the role of the physical properties of the cell wall in pollen tube growth of *Solanum chacoense*. *Planta, 220*(4), 582-592.

Pastinen, T. (2010). Genome-wide allele-specific analysis: insights into regulatory variation. *Nat. Rev. Genet., 11*(8), 533-538.

Paterson, A. H., Bowers, J. E. & Chapman, B. A. (2004). Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proc. Natl. Acad. Sci., 101*(26), 9903-9908.

Pickrell, J. K., Marioni, J. C., Pai, A. A., Degner, J. F., Engelhardt, B. E., Nkadori, E., Veyrieras, J.-B., Stephens, M., Gilad, Y. & Pritchard, J. K. (2010). Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature, 464*(7289), 768-772.

Pinkard, O., McFarland, S., Sweet, T. & Coller, J. (2020). Quantitative tRNA-sequencing uncovers metazoan tissue-specific tRNA regulation. *Nat. Commun., 11*(1), 4104.

Portères, R. (1956). Taxonomie Agrobotanique des Riz cultivés O. *sativa* L. et O. *glaberrima Steudel*. *JATBA, 3*(7), 341-384.

Pranathi, K., Viraktamath, B. C., Neeraja, C. N., Balachandran, S. M. & et al. (2016). Development and validation of candidate gene-specific markers for the major fertility restorer genes, *Rf4* and *Rf3* in rice. *Mol. Plant. Breed., 36*(10), 145.

Project, I. (2005). The map-based sequence of the rice genome. *Nature, 436*(7052), 793-800.

Qi, B. & Wu, C. (2022). Potential roles of stigma exsertion on spikelet fertility in rice (*Oryza sativa* L.) under heat stress. *Front. Plant. Sci., 13*, 983070.

Reddyyamini, B., Reddy, K. H., Reddy, V. L. N., Babu, P. R. & Sudhakar, P. (2019). Transgressive segregation for yield and its component traits in rice (*Oryza sativa* L.). *Int. J. Curr. Microbiol. App. Sci., 8*(6), 2450-2455.

Robinson, M. D., McCarthy, D. J. & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics, 26*(1), 139-140.

Rosselli, R., Romoli, O., Vitulo, N., Vezzi, A., Campanaro, S., de Pascale, F., Schiavon, R., Tiarca, M., Poletto, F. & Concheri, G. (2016). Direct 16S rRNA-seq from bacterial communities: a PCR-independent approach to simultaneously assess microbial diversity and functional activity potential of each taxon. *Sci. Rep., 6*(1), 32165.

Saliba, A.-E., Westermann, A. J., Gorski, S. A. & Vogel, J. (2014). Single-cell RNA-seq: advances and future challenges. *Nucleic. Acids. Res., 42*(14), 8845-8860.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci., 74*(12), 5463-5467.

Santiago, J. P. & Sharkey, T. D. (2019). Pollen development at high temperature and role of carbon and nitrogen metabolites. *Plant. Cell. Environ., 42*(10), 2759-2775.

Sato, Y., Antonio, B. A., Namiki, N., Takehisa, H., Minami, H., Kamatsuki, K., Sugimoto, K., Shimizu, Y., Hirochika, H. & Nagamura, Y. (2010). RiceXPro: a platform for monitoring gene expression in *japonica* rice grown under natural field conditions. *Nucleic. Acids. Res., 39*(1), 1141-1148.

Schmittgen, T. D. & Livak, K. J. (2008). Analyzing real-time PCR data by the comparative CT method. *Nat. Protoc., 3*(6), 1101-1108.

Scholz, P., Anstatt, J., Krawczyk, H. E. & Ischebeck, T. (2020). Signalling pinpointed to the tip: the complex regulatory network that allows pollen tube growth. *Plants, 9*(9), 1098.

Schulz, M. H., Zerbino, D. R., Vingron, M. & Birney, E. (2012). Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics, 28*(8), 1086-1092.

Schuster, S. C. (2008). Next-generation sequencing transforms today's biology. *Nat. Methods., 5*(1), 16-18.

Sekhar, S., Kumar, J., Mohanty, S., Mohanty, N., Panda, R. S., Das, S., Shaw, B. P. & Behera, L. (2021). Identification of novel QTLs for grain fertility and associated traits to decipher poor grain filling of basal spikelets in dense panicle rice. *Sci. Rep, 11*(1), 13617.

Selinski, J. & Scheibe, R. (2014). Pollen tube growth: where does the energy come from?. *Plant. Signal. Behav., 9*(12), e977200.

Sera, Y., Hanamata, S., Sakamoto, S., Ono, S., Kaneko, K., Mitsui, Y., Koyano, T., Fujita, N., Sasou, A., Masumura, T., Saji, H., Nonomura, K.-I., Mitsuda, N., Mitsui, T., Kurusu, T. & Kuchitsu, K. (2019). Essential roles of autophagy in metabolic regulation in endosperm development during rice seed maturation. *Sci. Rep., 9*(1), 18544.

Shabalin, A. A. (2012). Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics, 28*(10), 1353-1358.

Shankar, R., Bhattacharjee, A. & Jain, M. (2016). Transcriptome analysis in different rice cultivars provides novel insights into desiccation and salinity stress responses. *Sci. Rep., 6*(1), 1-15.

Shi, Z., Ren, W., Zhao, Y., Wang, X., Zhang, R., Su, A., Wang, S., Li, C., Wang, J., Wang, S., Zhang, Y., Ji, Y., Song, W. & Zhao, J. (2021). Identification of a locus associated with genic male sterility in maize via EMS mutagenesis and bulked-segregant RNA-seq. *Crop. J., 9*(6), 1263-1269.

Shindo, C., Tsujimoto, H. & Sasakuma, T. (2003). Segregation analysis of heading traits in hexaploid wheat utilizing recombinant inbred lines. *Heredity, 90*(1), 56-63.

Sims, D., Sudbery, I., Ilott, N. E., Heger, A. & Ponting, C. P. (2014). Sequencing depth and coverage: key considerations in genomic analyses. *Nat. Rev. Genet., 15*(2), 121-132.

Singh, D., Singh, P. K., Chaudhary, S., Mehla, K. & Kumar, S. (2012). Exome sequencing and advances in crop improvement. *Adv. Genet, 79*, 87-121.

Stark, R., Grzelak, M. & Hadfield, J. (2019). RNA sequencing: the teenage years. *Nat. Rev. Genet., 20*(11), 631-656.

Sweeney, M. & McCouch, S. (2007). The complex history of the domestication of rice. *Ann. Bot., 100*(5), 951-957.

Tang, C., Zhu, X., Qiao, X., Gao, H., Li, Q., Wang, P., Wu, J. & Zhang, S. (2020). Characterization of the pectin methyl-esterase gene family and its function in controlling pollen tube growth in pear (*Pyrus bretschneideri*). *Genomics, 112*(3), 2467-2477.

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B. B. & Siddiqui, A. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods., 6*(5), 377-382.

Tao, D., Xu, P., Li, J., Hu, F., Yang, Y., Zhou, J., Tan, X. & Jones, M. P. (2004). Inheritance and mapping of male sterility restoration gene in upland *japonica* restorer lines. *Euphytica, 138*(3), 247-254.

Taylor, L. P. & Hepler, P. K. (1997). Pollen germination and tube growth. *Annu. Rev. Plant. Biol, 48*(1), 461-491.

Thirugnanakumar, S., Narasimman, R., Anandan, A. & Kumar, N. S. (2011). Studies of genetics of yield and yield component characters in F_2 and F_3 generations of rice (*Oryza sativa* L.). *Afr. J. Biotechnol., 10*(41), 7987-7997.

Tian, G.-W., Chen, M.-H., Zaltsman, A. & Citovsky, V. (2006). Pollen-specific pectin methylesterase involved in pollen tube growth. *Dev. Biol., 294*(1), 83-91.

Tran, L. T., Sugimoto, K., Kasozi, M., Mitalo, O. W. & Ezura, H. (2023). Pollination, pollen tube growth, and fertilization independently contribute to fruit set and development in tomato. *Front. Plant. Sci., 14*, 1205816.

Wan, J., Yamaguchi, Y., Kato, H. & Ikehashi, H. (1996). Two new loci for hybrid sterility in cultivated rice (*Oryza sativa* L.). *Theor. Appl. Genet., 92*, 183-190.

Wang, H.-c. & Hickey, D. A. (2002). Evidence for strong selective constraint acting on the nucleotide composition of 16S ribosomal RNA genes. *Nucleic. Acids. Res., 30*(11), 2501-2507.

Wang, W., Li, Y., Dang, P., Zhao, S., Lai, D. & Zhou, L. (2018). Rice secondary metabolites: structures, roles, biosynthesis, and metabolic regulation. *Molecules, 23*(12), 3098.

Wang, Z., Gerstein, M. & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet., 10*(1), 57-63.

Wang, Z. Y., Zheng, F. Q., Shen, G. Z., Gao, J. P., Snustad, D. P., Li, M. G., Zhang, J. L. & Hong, M. M. (1995). The amylose content in rice endosperm is related to the post-transcriptional regulation of the waxy gene. *Plant. J., 7*(4), 613-622.

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

Wongsurawat, T., Jenjaroenpun, P., Wanchai, V. & Nookaew, I. (2022). Native RNA or cDNA Sequencing for Transcriptomic Analysis: A Case Study on Saccharomyces cerevisiae. *Front. Bioeng. Biotechnol., 10*, 842299.

Wu, J., Shahid, M. Q., Guo, H., Yin, W., Chen, Z., Wang, L., Liu, X. & Lu, Y. (2014). Comparative cytological and transcriptomic analysis of pollen development in autotetraploid and diploid rice. *Plant. Reprod., 27*(4), 181-196.

Xiang, X., Zhang, P., Yu, P., Zhang, Y., Yang, Z., Sun, L., Wu, W., Khan, R. M., Abbas, A., Cheng, S. & Cao, L. (2019). *LSSR*1 facilitates seed setting rate by promoting fertilization in rice. *Rice, 12*(1), 31.

Xu, Y., Cai, W., Chen, X., Chen, M. & Liang, W. (2022). A small Rho GTPase *OsRacB* is required for pollen germination in rice. *Dev. Growth. Differ., 64*(2), 88-97.

Xu, Y., Liu, S., Liu, Y., Ling, S., Chen, C. & Yao, J. (2017). HOTHEAD-Like *HTH1* is involved in anther cutin biosynthesis and is required for pollen fertility in rice. *Plant. Cell. Physiol., 58*(7), 1238-1248.

Xu, Y., Yang, J., Wang, Y., Wang, J., Yu, Y., Long, Y., Wang, Y., Zhang, H., Ren, Y., Chen, J., Wang, Y., Zhang, X., Guo, X., Wu, F., Zhu, S., Lin, Q., Jiang, L., Wu, C., Wang, H. & Wan, J. (2017). *OsCNGC13* promotes seed-setting rate by facilitating pollen tube growth in stylar tissues. *PLoS. Genet., 13*(7), e1006906.

Yu, L., Ma, T., Zhang, Y., Hu, Y., Yu, K., Chen, Y., Ma, H. & Zhao, J. (2017). Identification and analysis of the stigma and embryo sac-preferential/specific genes in rice pistils. *BMC. Plant. Biol., 17*(1), 60.

Yuan, L. & Virmani, S. (1988). Status of hybrid rice research and development. *Hybrid. rice.*, 7-24.

Zeng, Y., Zhang, Y., Xiang, J., Uphoff, N. T., Pan, X. & Zhu, D. (2017). Effects of low temperature stress on spikelet-related parameters during anthesis in *indica–japonica* hybrid rice. *Front. Plant. Sci., 8*, 1350.

Zenoni, S., Ferrarini, A., Giacomelli, E., Xumerle, L., Fasoli, M., Malerba, G., Bellin, D., Pezzotti, M. & Delledonne, M. (2010). Characterization of transcriptional complexity during berry development in Vitis vinifera using RNA-Seq. *Plant. Physiol. Biochem., 152*(4), 1787-1795.

Zhang, C., Li, G., Chen, T., Feng, B., Fu, W., Yan, J., Islam, M. R., Jin, Q., Tao, L. & Fu, G. (2018). Heat stress induces spikelet sterility in rice at anthesis through inhibition of pollen tube elongation interfering with auxin homeostasis in pollinated pistils. *Rice, 11*(1), 1-12.

Zhang, F., Zhang, Y.-C., Zhang, J.-P., Yu, Y., Zhou, Y.-F., Feng, Y.-Z., Yang, Y.-W., Lei, M.-Q., He, H. & Lian, J.-P. (2018). Rice *UCL8*, a plantacyanin gene targeted by miR408, regulates fertility by controlling pollen tube germination and growth. *Rice, 11*(1), 1-6.

Zhang, G.-q. (2020). Prospects of utilization of inter-subspecific heterosis between *indica* and *japonica* rice. *J. Integr. Agric., 19*(1), 1-10.

Zhang, H., Kjemtrup-Lovelace, S., Li, C., Luo, Y., Chen, L. P. & Song, B.-H. (2017). Comparative RNA-seq analysis uncovers a complex regulatory network for soybean cyst nematode resistance in wild soybean (*Glycine soja*). *Sci. Rep., 7*(1), 1-14.

Zhang, H., Zhu, J., Gong, Z. & Zhu, J.-K. (2022). Abiotic stress responses in plants. *Nat. Rev. Genet., 23*(2), 104-119.

Zhang, M. J., Zhang, X. S. & Gao, X.-Q. (2020). ROS in the male–female interactions during pollination: Function and regulation. *Front. Plant. Sci., 11*, 177.

Zhang, Z., Sun, W., Wen, L., Liu, Y., Guo, X., Liu, Y., Yao, C., Xue, Q., Sun, Z. & Wang, Z. (2023). Dynamic gene regulatory networks improving spike fertility through regulation of floret primordia fate in wheat. *Plant. Cell. Environ., 1*, 1-16.

Zhang, Z., Zhang, B., Chen, Z., Zhang, D., Zhang, H., Wang, H., Zhang, Y. e., Cai, D., Liu, J. & Xiao, S. (2018). A PECTIN METHYLESTERASE gene at the maize Ga1 locus confers male function in unilateral cross-incompatibility. *Nat. Commun., 9*(1), 3678.

Zheng, Y., Yan, J., Wang, S., Xu, M., Huang, K., Chen, G. & Ding, Y. (2018). Genome-wide identification of the *pectate lyase-like* (*PLL*) gene family and functional analysis of two *PLL* genes in rice. *Mol. Genet. Genom., 293*(6), 1317-1331.

Zhou, L., Wang, Y., Xu, X., Yan, D., Yu, W., Miao, Y. & Xu, B. (2022). Conjunctive analyses of BSA-Seq and BSR-Seq unveil the *Msβ-GAL* and *MsJMT* as Key candidate genes for cytoplasmic male sterility in alfalfa (*Medicago sativa* L.). *Int. J. Mol. Sci., 23*(13).

Zhou, L.-Z., Juranić, M. & Dresselhaus, T. (2017). Germline development and fertilization mechanisms in maize. *Mol. Plant., 10*(3), 389-401.

Zong, Y., Chen, Z., Innes, J. B., Chen, C., Wang, Z. & Wang, H. (2007). Fire and flood management of coastal swamp enabled first rice paddy cultivation in east China. *Nature, 449*(7161), 459-462.

BIOGRAPHY

Name Miss Kanokwan Kaewmungkun

Educational Attainment 2013-2016: Master Degree of Science (Plant Physiology), Khon Kaen University, Khon Kaen, THAILAND.

2009-2012: Bachelor Degree of Science (Biology), Khon Kaen University, Khon Kaen, THAILAND.

Activities and awards

- Rice Green Production Technology in Lancang-Mekong Region Organized by Guangxi Academy of Agricultural Sciences on September 19th, 2022 in Nanning, Guangxi, P. R. China (Online). Topic "Development of new aromatic rice lines with high eating and cooking quality"
- The National and International Graduate Research Conference 2016, 15 January 2016, Khon Kaen University, Thailand. Topic "Effect of exogenous polyamines on some physiological characteristic and carbohydrate metabolism in rice (*Oryza sativa* l.) under drought stress"
- The 9th Botanical Conference of Thailand: BCT9 -Stepping toward Global Forum 2015, 3-5 June 2015, Thailand. Topic "Effects of calcium chloride on growth, carbohydrate metabolism and antioxidant enzyme activities in rice (O*ryza sativa* l.) under drought stress"

Publications

Kaewmungkun, K., Tongmark, K., Chakhonkaen, S., Sangarwut, N. & Muangprom, A. Bulked segregant RNA-Seq (BSR-Seq) analysis of pollinated pistils reveals genes influencing spikelet fertility in rice. *Rice Sci.* (Accepted: 18 April 2024)

Kaewmungkun, K., Tongmark, K., Chakhonkaen, S., Sangarwut, N., Wasinanon, T., Panyawut,N., Ditthab, K., Situng, K., Qi, Y.-b., Dapha, S., Panya, A., Phonsatta, N. & Muangprom, A. (2023). Development of new aromatic rice lines with high eating and cooking qualities. *J. Integr. Agric.*, *22*(3), 679-690.

Kaewmungkun, K. & Limpisophon, K. (2023). Characteristics of coconut protein-enriched lipid and the effects of pH, NaCl, and xanthan gum on its dispersibility. *J. Saudi Soc. Agric. Sci.*, *22*(6), 351-360.

Publications (Cont.)

Yongbin, Q., Summat, P., Panyawut, N., Sikaewtung, K., Ditthab, K., Tongmark, K., Kaewmungkun & Muangprom, A. (2023). Identification of rice accessions having cold tolerance at the seedling stage and development of novel genotypic assays for predicting cold tolerance. *Plants*, *12*(1), 215.

Kaewmungkun, K., Pattanagul, W., Monthatong, M., & Theerakulpisut, P. Effect of Exogenous Polyamines on Some Physiological Characteristic and Carbohydrate Metabolism in Rice (*Oryza sativa* L.) Under Drought Stress. (2016). The National and International Graduate Research Conference, Khon Kaen University, Thailand.