MONASH University

Population Genomics Study on Sabah Weedy Rice Using Genotyping-by-Sequencing (GBS)

Rupini Yesudasan
Bachelor of Biotechnology (Hons)

A thesis submitted for the degree of Master of Science at
Monash University in 2019
School of Science

Abstract

Weedy rice (*Oryza sativa*) is a conspecific form of cultivated rice that infests rice fields worldwide and poses a significant threat to sustainable rice production. Despite being a serious problem in the rice industry, little is known about the genetics of weedy rice in Sabah, an eastern Malaysian state on Borneo Island. Given that the occurrence of Sabah weedy rice is chronologically later than the spread of the weedy rice in Peninsular Malaysia and the fact that some weedy rice forms were found to share phenotypic similarities with their weedy counterparts from the Peninsular, the emergence of these weeds in Sabah may have an exotic origin outside of the state. Using two complementary approaches: genotyping-by-sequencing (GBS) and candidate gene analysis on *TAC1* domestication gene, the genetic diversity, population structure and potential origins of Sabah weedy were assessed with 182 samples of weedy, cultivated and wild rice accessions collected from Sabah, Peninsular Malaysia and neighboring South Asian countries. The results showed that the local cultivars are the main contributors to weedy rice genomes with considerable evidence for relative contributions of Peninsular Malaysian weedy ecotypes in the establishment of weedy rice in Sabah. The rapid spread of weedy rice with evolutionary origin outside of the state highlights the need for management of current weedy populations and measures to prevent further introductions into Sabah.

Keywords: population genetic, genotyping-by-sequencing (GBS), domestication gene, exotic origin, weedy rice

Declaration

This thesis is an original work of my research and contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Print Name: Rupini Yesudasan

Date: 20 July 2019

Acknowledgements

With boundless love and appreciation, I would like to extend my heartfelt gratitude and appreciation to the people who helped me bring this study to reality.

I would like to acknowledge my indebtedness and render my warmest thanks to my supervisor, Dr. Song Beng Kah who made this work possible. His friendly guidance and expert advice have been invaluable throughout all stages of the work. His consistence guidance, ample time spent and consistent advices that helped me bring this study into success. His dedicated supervision and constant encouragement towards the completion of this thesis encouraged me to do my best.

I would also wish to express my gratitude to my Co-Supervisor, Prof. Dr. Sadequr Rahman for extended discussions and valuable suggestions which have contributed greatly to the improvement of the thesis. His guidance helped me in all the time of research and writing of this thesis. I am greatly thankful to Associate Prof. Qasim Ayub for his insightful comments and encouragement, and for the hard question which incented me to widen my research from various perspectives. His wide knowledge and his logical way of thinking have been of great value for me.

I am also grateful to my lab mates for their patience and support in overcoming numerous obstacles that I had been facing through my research. Their fully devoted into this study inspire me to work together. Their optimistic attitudes especially encouraged me not to give up even facing any obstacles or difficulties.

Nobody has been more important to me in the pursuit of this project than the members of my family and friends. I would like to thank my parents; whose love and guidance are with me in whatever I pursue. I must express my profound gratitude to my best friends for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVATION

®	Registered trademark
μ	Micro
GBS	Genotyping-by-sequencing
BERNAS	Padiberas Nasional Berhad (Malaysian national rice corporation)
ha	Hectare
IRRI	International Rice Research Institute
MARDI	Malaysian Agriculture Research and Development Institute
MT	Metric tons
NGS	Next generation sequencing
O.	*Oryza*
PCR	Polymerase chain reaction
PCoA	Principle Coordinate Analysis
RAPD	Random amplified polymorphic DNA
Sdn.Bhd	Sendirian Berhad
Sp.	Species
SSR	Simple sequence repeat
SNP	Single nucleotide polymorphism

CHAPTER 1
INTRODUCTION

1.0 INTRODUCTION

Since the advent of agriculture, humans have encountered weeds that have thwarted their goal to manage crop cultivation. Weeds are plants that grow where they are not desired, and known as 'plants in the wrong place' (Monaco et al., 2002; Vigueira et al., 2013).These weeds have evolved to survive in agricultural settings, mainly through unintentional human selection processes (Warwick and Stewart, 2005). Unlike crops, which are characterized by domestication traits due to human cultivation, weeds possess weedy traits such as prolific seed production with high shattering, prolonged seed dormancy, vigorous vegetative growth and crop mimicry that allow them to thrive under harsh environmental conditions, giving them survival advantages (Rodenburg et al., 2008; Qi et al., 2015). Particularly, agricultural weeds lower farmers' income through yield loss, resulting in 10% worldwide reduction in crop productivity and approximately $33 billion annual cost to the United States alone (Pimentel et al., 2005). Although billions of dollars have been spent on weed management, yield loss due to weed problems still continues (Kane and Rieseberg, 2008). When new and improved weed management techniques are developed, new resistant weeds appear. Where do these weeds come from? Are they pre-adapted to be weeds as soon as they first encounter agricultural settings, or do they appear due to specific genetic changes? Answers to these questions are clearly important to establish a long-term weed management plan and for protection of the global food supply.

Weeds that are closely related to crops provide best study systems for examining weed evolution especially when the genetic resources available for an economically important crop can be leveraged to unravel the genetic basis and evolutionary history of weed-adaptive traits (Gross and Olsen, 2010). One such example is weedy rice (*Oryza sativa*), which infests rice fields globally. Weedy rice (or locally known as "Padi Angin") belongs to the same biological species as cultivated rice and is a conspecific taxon of the AA genome complex of rice that shares traits of both cultivated and wild rice types (Chen et al., 2004). Also known as red rice, it is typically taller, has rapid growth, early shattering, higher tillering capacity, flowers earlier than cultivated rice, has a pigmented caryopsis, pigmented hulls with or without awns , longer seed dormancy and competes with

cultivated rice for nutrients, light and space (Olsen et al., 2007; Shivrain et al., 2009; Gressel and Valverde, 2009). As most weedy rice types share similar morphological, physiological and biochemical characteristics with cultivated rice, eradication of weedy rice infestation without affecting cultivated rice is difficult (Shivrain et al., 2007; Saha et al., 2014). The rapid proliferation of weedy rice appears to be a direct consequence of changes in rice establishment methods, from traditional transplanting to industrialized direct-seeded method (Chauhan, 2013; Song et al., 2014). The changes in rice establishment methods were mainly due to water and labor scarcity and increased production cost (Shivrain et al., 2009; Chauhan, 2013). Additionally, farmers' limited knowledge on weedy rice and poor weed management, such as use of contaminated seeds and shared use of machinery have exacerbated weedy rice problems (Azmi and Baki, 2007; Song et al., 2014; Sudianto et al., 2016).

The evolutionary origin of weedy rice has been a controversial topic. There are multiple efforts underway to comprehend its worldwide origin (Thurber et al., 2010). Following these studies, several hypotheses that explain the origin of weedy rice have been put forward: (i) through de-domestication process of cultivated rice; (ii) weedy forms developed from ongoing selection and adaptation of wild rice to disturbed habitats; and (iii) hybridization between cultivated rice and its wild ancestor *Oryza rufipogon* (Londo and Schaal, 2007; Chauhan, 2013). Genetic studies of weedy rice population worldwide using molecular markers indicate that multiple genetically distinct weedy rice strains exist and the origin of weedy rice populations may vary across sites (Azmi and Baki, 2007; Londo and Schaal, 2007; Huang et al., 2012; Song et al., 2014, Li et al., 2017). Studies based on neutral markers, simple sequence repeat (SSR) and single nucleotide polymorphism (SNP) showed the two major weedy rice populations found in the United States, strawhull awnless (SH) and blackhull awned (BHA) are more closely related to the Asian cultivar groups *indica* and *aus*, respectively (Londo and Schaal, 2007; Gealy et al., 2009; Reagon et al., 2010). Studies using isozyme and SSR markers revealed that weedy rice populations in Bhutan and China are genetically related with *japonica* cultivated rice varieties (Ishikawa et al., 2005; Cao et al., 2006). Another recent study in Thailand, where rice is cultivated in close proximity to wild rice populations, suggested

that hybridization with wild population have contributed to the genetic makeup of weedy rice populations (Pusadee et al., 2013).

In Malaysia, the first weedy rice incident was reported in southern part of the Muda area, Kedah in 1988 and it became a major problem over the last two decades (Wahab and Suhaimi, 1991; Azmi and Baki, 2007). Only limited surveys and few restricted and localized studies have been performed, highlighting the detection, diversity and degree of weedy rice infestations (Karim et al., 2004; Sudianto et al., 2016). A recent molecular study (Song et al., 2014) based on 24 SSR markers and sequence variations at two domestication genes (*sh4*, controls seed shattering; *Bh4*, controls hull colour) revealed a complex origin where both wild and cultivated rice likely to have contributed to the composition of Malaysian weedy rice populations. Another study on sequence variations at the *Rc* gene, controlling pericarp colour (Cui et al., 2016), showed weedy rice populations in Peninsular Malaysia appear to have diverse origins with evidence of gene flow from *indica* cultivars and *O. rufipogon,* which is consistent with the previous analysis by Song et al. (2014).

Sabah is an East Malaysian state located in the northern part of Borneo Island which accounts for 20% of rice production in Malaysia (Islam et al., 2007). The rice industry in Sabah has been given a special priority based on its importance as a strategic crop and staple food commodity of the state. Like rice planting in Peninsular Malaysia, weedy rice infestation poses a serious threat to Sabah rice production. The first incident of Sabah weedy rice was recorded approximately a decade after the initial reports of weedy rice infestation in Peninsular Malaysia (Wahab and Suhaimi, 1991). The occurrence of weedy rice in the local fields was first observed in Kota Belud and Kota Marudu districts in late 2000s (S.S. Teo, unpublished observations), since then it has become a widespread problem in other rice production areas as well. In 2009, so far 1,100ha of paddy fields in Kota Marudu and 500ha in Kota Belud were affected by the weedy rice (The Star, 2009).

In temperate regions where no wild rice relatives occur, a fundamentally different set of population dynamics has likely governed the weed’s emergence and aggressive spread in rice fields (Olsen et al., 2007). However, its origin and evolutionary mechanism have remained poorly understood despite its significant threat. Past studies have shown that

cultivars could have contributed to evolution of weedy rice, with natural hybridization between *indica* and *japonica* varieties (Cao et al., 2006), or accidental introduction of weedy rice seeds that evolved from cultivated varieties elsewhere (Reagon et al., 2010) as possible mechanisms of the processes. In tropical countries such as Malaysia, although *O. rufipogon* is not native to Sabah state, extensive morphological diversity in Sabah weedy rice types such as wild-like vegetative and reproductive traits (e.g. open panicles, presence of awns, seed dormancy) were observed (BK Song, unpublished observations). Various distinct morphotypes of strawhull awnless (SH), strawhull awned (SHA), brown-striped-hull awnless (BR), brown-striped-hull awned (BRA) and their associated wild traits detected in fields implicated a dynamic evolutionary origin and complexity of weedy rice populations in Sabah state. Given that Sabah weedy rice occurrence was observed about a decade later than the initial reports of the weed in Peninsular Malaysia, and the fact that some weedy rice forms were found to share phenotypic similarities for some wild traits with their weedy counterparts from the Peninsular (e.g. BR and BRA with open panicles, high plant stature, grain shape and hull colour), it was hypothesized that emergence of these weeds in Sabah may have an exotic origin outside of the state (BK Song, Unpublished data). Some weedy rice types are also morphologically similar to local cultivars in Sabah, hence one would expect that these weedy rice types are genetically close to local cultivated rice.

For a better understanding of the occurrence of weedy rice, knowledge of two aspects on weedy rice evolution is needed. First, which wild and/or domesticated ancestors are they derived from? Second, what genetic mechanisms account for the traits that have enabled weedy rice to successfully invade the rice field? (Vigueira et al., 2013). Based on these two aspects, the origin and evolution of Sabah weedy rice would be examined in comparison with Peninsular weedy, wild and cultivated rice, by using two complementary approaches; Genotyping-by-sequencing (GBS) and candidate gene analyses. Firstly, the genetic diversity and population structure will be assessed the using neutral and possibly selected SNP markers across the rice genome using GBS, which has recently emerged as a promising genomic approach using next generation sequencing (NGS) technologies to explore plant genetic diversity on a large scale (Elshire et al., 2011; Peterson et al., 2014). The relatively straightforward, robust, and cost-effective GBS protocol is currently

being applied in numerous species including weedy rice by a large number of researchers (Glaubitz et al., 2014).

Secondly, given key weedy traits appear to have corresponding and well-characterized domestication genes, genetic basis of weediness can also be examined using candidate gene markers (Song et al., 2014). Domestication traits are those traits selected by humans that allow the plants to be cultivated in the agricultural settings such as seeds that have reduced shattering, awnless seed, and close, compact architecture. In this study, sequence variation of a domestication gene known as *TAC1* (controlling tiller angle development) was assessed. Tiller angle is the angle between the outermost tillers and main stem of rice plant. As one of the important factors in plant architecture, it determines the planting density per area and contributes greatly to grain yield. An optimum tiller angle significantly affects the ability to trap light efficiently, resist diseases, and maximizes yield (Xu et al., 1998; Li et al., 2007a). Rice plants with extreme spreading morphology occupy too much space and increase shading and lodging, thus decreasing the photosynthetic efficiency. On the other hand, the minimal tiller angle characteristic of extremely compact varieties restricts the penetration of light and air (Jiang et al., 2012a). *O. rufipogon*, the wild progenitor shows a spread-out growth however in contrast, cultivated rice usually exhibits a better plant architecture, with a smaller tiller angle that leads to high potential yields (Dong et al., 2016).*Tiller Angle Control 1 (TAC1)* is a major QTL located on chromosome 9 that controls tiller angle during heading stage in rice and encodes an expressed protein without homologous genes in rice (Jiang et al., 2012a; Dong et al., 2016).

1.2 Objectives

As evolutionary mechanisms associated with the emergence of the Sabah weedy rice populations have not been explored, it is important to examine the genetic variation and relationships among different weedy and cultivated rice strains in the local planting areas. More importantly, knowledge of weedy rice genetic diversity and population structure in this region would be able to facilitate the establishment of effective measures to manage the infestation. This present study sought to assess the genetic diversity within Sabah weedy rice accessions and examine the origin of Sabah weedy rice in comparison with local cultivar and weed strains from Peninsular Malaysia. Since the agronomic traits often differ between cultivated rice and wild / weedy relatives, candidate genes have opened up new sources of potential information about the evolution of weediness-enhancing traits. Combining with information about genetic diversity obtained from genotyping-by-sequencing, a candidate genes approach would help provide a more complete picture of the evolutionary origin of weedy rice groups. Here, the nucleotide variations in *TAC1* gene and its tiller angle traits in wild, weedy and cultivated rice were analysed. Since the tiller angle phenotypes have been selected at different stages of the domestication process, this *TAC1* analysis may tell us complementary accounts about the origin of Sabah weedy rice. To my knowledge, this is the first evolutionary study on weedy rice populations in the East Malaysian state.

The objectives of this project are as follows:

1. To assess genetic diversity and population structure of Sabah weedy rice populations.
2. To determine the relationship of Sabah weedy rice to local cultivars, and to weedy rice strains in the Peninsular Malaysia, hence, elucidate the origin of weedy rice in Sabah.
3. To investigate the role of *TAC1* gene in contemporary evolution of its corresponding traits in Sabah weedy rice populations.

CHAPTER 2

LITERATURE REVIEW

2.0 LITERATURE REVIEW

2.1 Rise of 'RICE'

More than 10,000 years ago, ancient people slowly shifted from hunter-gatherer to farmer where they began to gather wild rice species that grew in the swamps and marshes and started cultivating them in their surroundings (Kovach et al., 2007). The domestication of rice has fundamentally altered the course of human civilization. Indeed, the largest Asian civilization was once built on rice (Fuller et al., 2010; Callaway, 2014). Through a course of continuous selection on desired and favourable traits, the ancient farmers slowly transformed the unruly wild rice (*Oryza rufipogon* Griff.) into domesticated rice crop (*Oryza sativa* L.) (Figure 2.1), which is now the primary food source for more than a third of the world's populations (Kovach et al., 2007; Sang and Ge, 2007). Compared to the wild progenitor, the cultivated rice exhibit non shattering spikelet, reduced seed dormancy, absence of awns, loss of pigmented pericarp and hull, compact plant architecture and close panicles with densely packed grains (Kovach et al., 2007; Sweeney and McCouch, 2007, Vaughan et al., 2008). There are 21 wild species in the genus *Oryza* and 6 of them, *O. rufipogon, O. nivara, O. glumaepatula, O. meridionalis, O. breviligulata, O. longistaminata* constitute the primary diploid gene pool and share the AA genome that facilitates hybridization among them (Brondani et al., 2005). There are two species of rice that man has domesticated: *O. sativa*, the Asian rice and *O. glaberrima*, the African rice. Domestication of Asian rice believed to have started in Yangtze River Valley region of China and spread to other parts of Asia (Gross and Zhao, 2014). *O. sativa* evolved into two main types; *japonica* and *indica* and these two groups further differentiated into five major sub-varieties which referred as *indica*, *aus*, *temperate japonica*, *tropical japonica* and aromatic (Figure 2.2) (Garris et al., 2005; Vaughan et al., 2008). *O.glaberrimma*, the other cultivated species is thought to be domesticated by people lived near the bend of Niger River, West Africa and is still confined to its native land (Linares, 2002; Delouche and Labrada, 2007).

Today, rice is grown in all continents of the world, between 55º N in China and 36º S in Chile with exception of Antarctica and more than 3.5 billion people depend on rice for obtaining 20% of their daily calorie intake (Muthayya et al., 2014; Fahad et al., 2019). Cultivation of rice under various geographical region and climates has resulted in countless rice varieties. Khush (1997) reported about 120,000 distinct rice varieties existed and approximately 80,000 are preserved in International Rice Research Institute (IRRI).

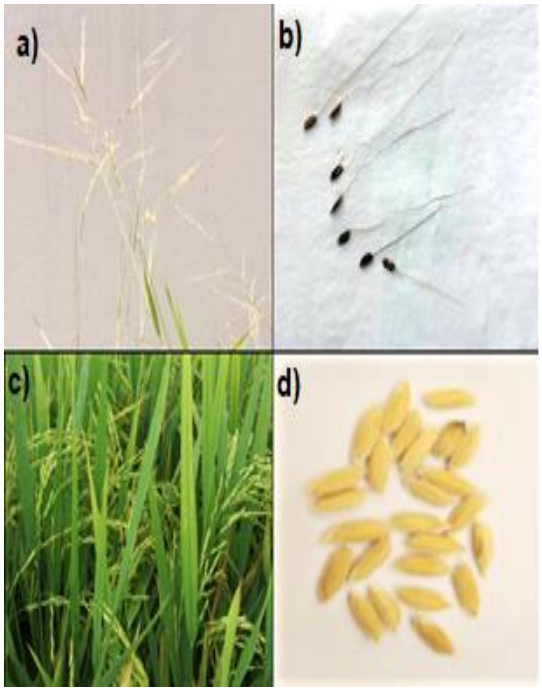


Figure 2.1: Transformation from *O. rufipogon* to *O. sativa*. a) panicles from *O. rufipogon*, b) seeds from *O. rufipogon*, c) panicles from *O. sativa*, d) seeds from *O. sativa*

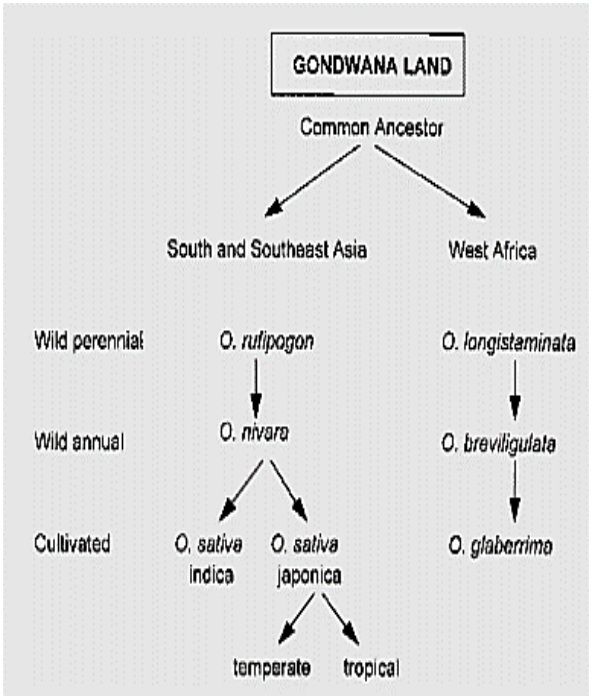


Figure 2.2: Evolutionary of two cultivated species .Extracted from Khush (1997).

2.1.1 Classification of *Oryza*

Oryza is a genus under the tribe *Oryzeae* in the family of *Poaceae* or *Gramineae*, commonly known as the grass family (the fourth largest flowering plant family) (Tzvelev, 1989; Peterson, 2013). There are 12 genera within the *Oryzeae* tribe (Vaughan, 1994). The genus *Oryza* contains approximately 23 species of which 21 are wild species and two are cultivated varieties (Vaughan, 1994). The genus *Oryza* has also been divided into four complexes, including the *Oryza sativa*, *Oryza officinalis*, *Oryza ridelyi* and *Oryza granulate* (Sweeney and McCouch, 2007). The genus contains both diploid (2n = 24) as well as tetraploid (2n = 48) species. Based on genome analysis all the species were grouped into nine distinct genomes, viz. A, B, C, D, E, F, G, H and J (Vaughan et al., 2003). These groups, along with the geographic locations of each species, are shown in Table 2.1.

Table 2.1 Species complexes of different Oryza species and their geographical distribution

SL NO	Species complex	Chromosome Number (2n)	Genome	Geographical distribution
	Sativa Complex			
1	*O. sativa* L.	24	AA	South & Southeast Asia
2	*O. nivara* Sharma et Shastry	24	AA	Tropical Asia
3	*O. rufipogon* Griff	24	AA	Tropical Asia
4	*O. meridionalis* Ng	24	AA	Tropical Asia
5	*O. glumaepetula I.*	24	AA	Tropical Australia
6	*O. glaberrima* Steud.	24	AA	Tropical America
7	*O. barthii* A.Chev.et Roehr	24	AA	West Africa
8	*O. longistaminata* A.Chev. et Roehr	24	AA	Africa
	Officinalis / Latifolia Complex			
9	*O. punctate Kotschy* ex Steud.	24	BB	Africa
10	*O. rhizomatis* Vaughan	24	CC	Sri Lanka
11	*O. minuta* J.S.Pesl. ex C.B. Presl	48	BBCC	Philippines, New Guinea
12	*O. mlamphuzensis* Krishn. Et Chandr.	48	BBCC	South India (Kerala)
13	*O. officinalis* Wall.ex Watt	24	CC	Asia, New Guinea
14	*O. eichingeri* A. Peter	24	CC	East Africa & Sri Lanka
15	*O. latifolia* Desv.	48	CCDD	Central & South America
16	*O. alta* Swallen	48	CCDD	Central & South America
17	*O. grandiglumis* (Doell) Prod.	48	CCDD	South America
18	*O. austaliensis* Domin	24	EE	Northern Australia
19	*O. schweinfruthiana* Prod	48	BBCC	Tropical Africa
	Meyeriana complex			
20	*O. granulate* Nees et Arn.ex Watt	24	GG	South & Southeast Asia
21	*O. meyeriana* (Zoll et Mor.ex Steud) Baill.	24	GG	Southeast Asia

Source : Extracted from Dunna and Roy (2013)

2.1.2 The wild rice, *Oryza rufipogon*

The common wild rice, *O. rufipogon*, known as the ancestor of Asian cultivated rice (*O. sativa* L.), is known as the most important germplasm for rice improvement (Song et al., 2003). *O. rufipogon* is a perennial, tufted grass that lives in relatively a-seasonal habitats and has large, indehiscent and pendant anthers (Ngu et al., 2010). It is a short –day plant that flowers near the end of the monsoon season (October to March) (Ngu et al., 2010). The vital differences from the crop are the tendency for the seed to shatter once mature and their prolonged seed dormancy. *O. rufipogon* populations harbor significantly higher genetic diversity than does the cultivated rice (Zhou et al., 2003). Based on 44 RFLP markers in Sun et al. (2001), the genetic diversity of *O. rufipogon* (A = 4.02, P = 97.7%, H_O = 0.033) was much higher than that of cultivated rice (A = 2.34, P = 75.0%, H_O = 0.014), and found that a great number of genes that occurred in *O. rufipogon* could not be found in cultivated rice. It is also reported that *O. rufipogon* outcrosses at a much higher rate (~7–56%) compared to *O. sativa* (~1–2%) (Gao et al., 2007; Ngu et al., 2010).

O. rufipogon is widely distributed in the tropics and subtropics of Asia including China, Nepal, India, Sri Lanka, Bangladesh, Myanmar, Laos, Thailand, Cambodia, Vietnam, Malaysia, Indonesia, Philippines, Papua New Guinea, and Australia (Vaughan, 1994; Ngu et al., 2010). In Malaysia, this species can be found only in the northern part of Peninsular Malaysia (Seberang Perai of Penang, Kedah, Kelantan and Terengganu), growing in swamps, along river banks, irrigation canals and in or at the margins of rice fields (Abdullah, 1991; Ngu et al., 2010).

O. rufipogon has been proven to be a valuable germplasm for rice improvement (Zhou et al., 2003). The first male-sterility (MS) gene was found in *O. rufipogon* and introduced to cultivated rice, which has led to the development of high yielding hybrid rice varieties (Song et al., 2003; Li et al., 2007b). Other agronomically beneficial traits, such as rice *tungro* virus resistance, elongation ability, tolerance to acid sulfate soil found in the wild rice and strong deep water tolerance, are of great importance for rice breeding (Song et al., 2003). Nevertheless, there is an urgent need for its conservation since its natural populations are recently declining in their natural habitats and their inherent genetic diversity is rapidly decreasing (Ngu et al., 2010).

2.1.3 Cultivated rice, *Oryza sativa*

Oryza sativa is known to have been originally domesticated in the southeastern part of Asia about 5000 years ago. Since then, several types of rice cultivars have been introduced to adapt to prevailing local culture techniques and environmental conditions (Sasaki, 2001). There are two major ecotypes of *O. sativa*, namely *indica*, adapted to the tropics, and *japonica*, adapted to the temperate regions and tropical uplands (Sang and Ge, 2007). The basic differences between these ecotypes can be clearly recognized according to the distinct shape of the seeds (Sasaki, 2001). *Indica* is characteristically long and slim whereas *japonica* appears short and round with the *tropical japonica* similar to *temperate japonica* but larger in size (Dunna and Roy, 2013). The *indica-japonica* division was clearly distinguished by restriction fragment length polymorphism (*RFLP*) studies (Wang and Tanksley, 1989; Sweeney and McCouch, 2007) and additional population structure consisting of *aus*, *aromatic*, *rayada*, and *ashina* was discerned using 15 isozyme loci (Glaszmann, 1987). The *aus*, *rayada* and *ashina* are minor groups that have generally been considered to be subspecies of *indica* ecotypes and all have a comparatively small geographic distribution along the Himalayan foothills (Garris et al., 2005). The drought-tolerant, early maturing *aus* rices are grown during the summer season while *rayada* and *ashina* are known to be floating rices, characterized by their vigorous growth with rising water level and have mainly been cultivated in areas where a flood always occurs during the wet season (Garris et al., 2005). *Aromatic* rices such as *basmati* from Pakistan and India and *sadri* from Iran have a distinctive aroma and are highly prized for their qualities (Garris et al., 2005).

In Peninsular Malaysia, three paddy varieties were frequently planted i.e., MR 220, MR 219, MR 263 (Department of Agriculture, 2015). Other varieties were also planted by paddy farmers but were in small percentages. For the state of Sabah, a total of 21 new varieties were introduced to farmers by the Department of Agriculture to date (Lamdin et al., 2015). The popular paddy varieties planted in Sabah were Tempatan, TR 8 and Sarawak (Table 2.2) (Department of Agriculture, 2015). Nevertheless, traditional rice varieties are still widely cultivated in Sabah especially by small land holders (Souki, 2015). Average yield of local rice varieties were lower compared to modern varieties, however,

majority of farmers in Sabah prefer to plant local rice varieties since they have good grain characteristics and good taste (Souki, 2015).

Table 2.2 Paddy varieties and average yield of wet paddy in Sabah during main season (August to February) 2013/2014

Variety	% Planted	Average yield of wet paddy (Kg/ha)
Tempatan	44.8	4194
TR 8	32.4	4213
MR 159	6.0	3829
TQR 2	2.5	4781
Padi Wangi	4.8	5387
Sarawak	6.7	4098
TR 7	1.0	4050
MR 269	1.0	2333
Bario	0.3	5950
Cileung	0.6	4050

Source : Extracted from Department of Agriculture (2015)

2.2 Rice Industry in Malaysia

Rice is the third most important crop in Malaysia after oil palm and rubber which were the first and second respectively in terms of production (Akinbile et al., 2011). Rice sector is contributing less than 1% to the gross domestic product (GDP) (Khairun Nisaa et al., 2018). Despite having a small contribution towards the nation's GDP, rice industry has garnered much interest from policymakers given its complex relationship with food security, culture and socio-economic factors (Khazanah Research Institute, 2019). Paddy planted area throughout Malaysia is estimated to be 672,000 hectares the average national paddy production is 3,660 metric ton per hectare (Khairun Nisaa et al., 2018). Paddy cultivation in Malaysia comprises the designated granary areas and non-granary areas and Malaysia relies primarily on ten key granary areas for its domestic rice production (Najim et al., 2007). The average yields in the granary areas are higher than in the non-granary areas. In 2016, the nation produced a total of 2.7 m MT of rice and out of this, 2.0m MT or 74.1% of the total paddy produced was from the granary areas (Khazanah Research Institute, 2019). The yield also varies between granary areas, with MADA, IADA Penang, IADA Ketara and IADA Barat Laut Selangor being the top performers with yields above 5.0 MT/Ha (Figure 2.3) (Khazanah Research Institute, 2019). These differences can be attributed to many combinatorial factors, including soil condition, weather, farm management, irrigation, pests and diseases and use of technology (Khazanah Research Institute, 2019). The main purpose of meeting a high-level rice production is to achieve a total rice self-sufficiency; the ability of a country to be able to supply rice one's own or its own needs without importing from other countries. Over the last 30 years, the total rice production in Malaysia has increased, allowing the self-sufficiency level (SSL) to reach 75% (Chamhuri et al., 2014). While the balance percentage of rice is sourced mainly from Thailand, Vietnam and Pakistan with the net rice imports grow by 2.2% per year by 2021 (Chamhuri et al., 2014). Malaysia envisioned to achieve 100% rice production SSL by year 2020 (Chamhuri et al., 2014).

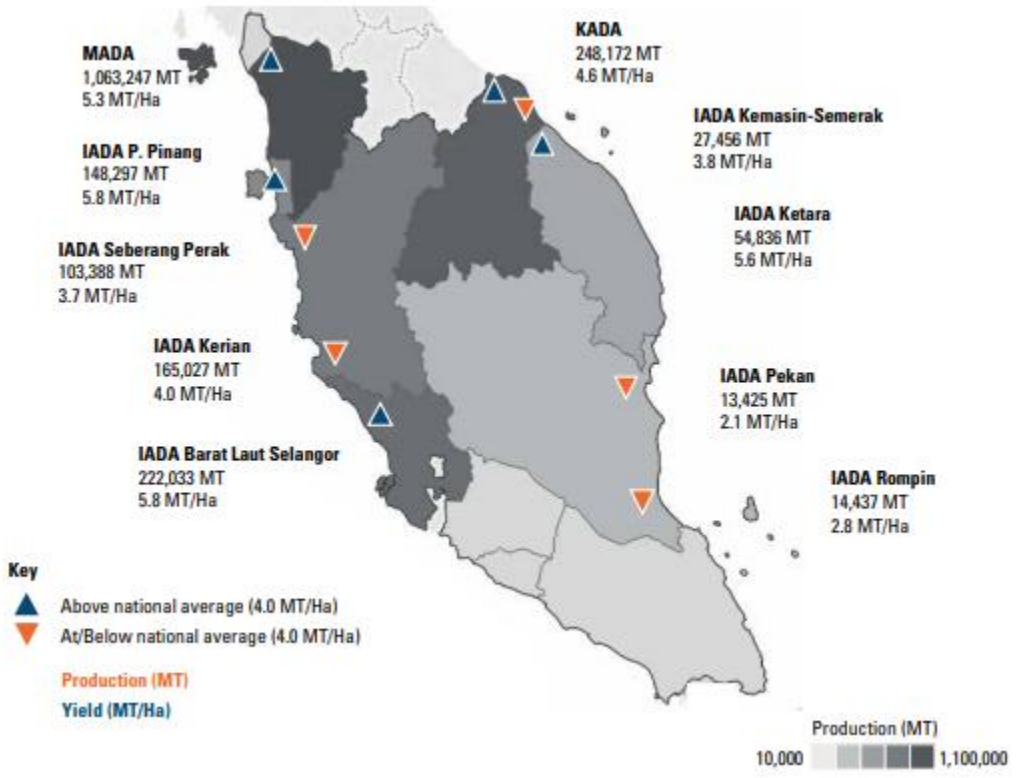
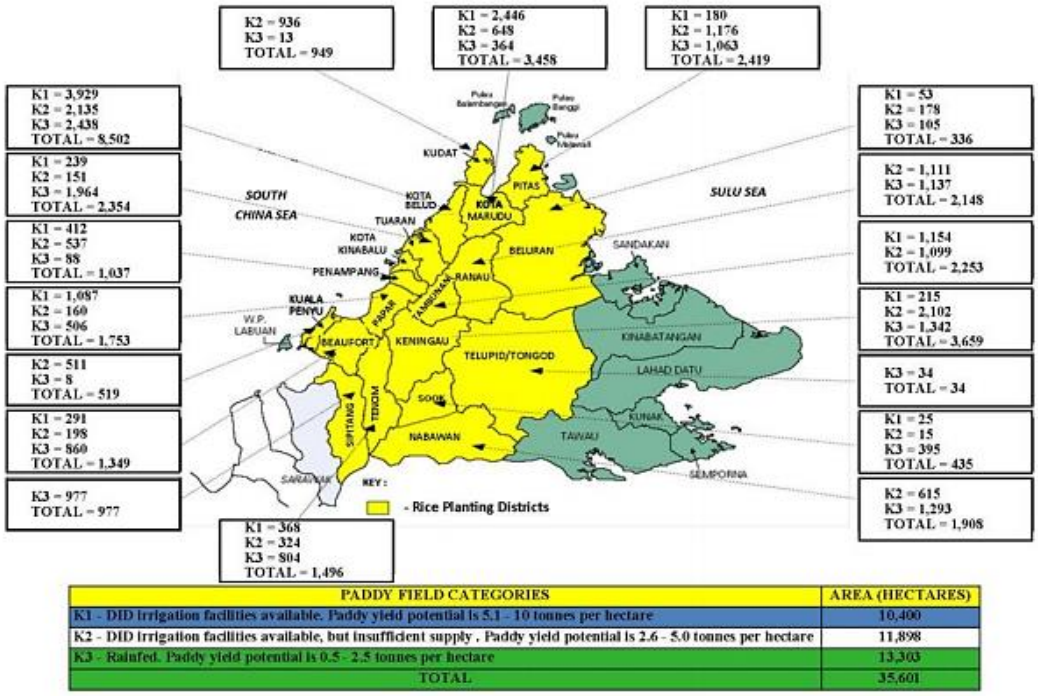


Figure 2.3. Rice production and yield in the granary areas, Malaysia, 2016.
Source : Extracted from Khazanah Research Institute (2019)

2.2.1 Rice production in Sabah

Rice industry in Sabah has been given a special priority based on its importance as a strategic crop and a staple food commodity. Rice also significantly shapes the cultural systems, beliefs, and traditions of people particularly the rural population in Sabah (Hashmi and Siau, 2016). The rice industry in Sabah is still under-developed compared to Peninsular Malaysia in which the rice production is only able to accommodate approximately 30-40% of the Sabah population needs (Hashmi and Siau, 2016). The main districts contributing to the production of rice in Sabah are Kota Belud (27%), Kota

Marudu (15%), Keningau (8%) and 16 other districts (Lamdin et al., 2015). Kota Belud, which is known as the rice bowl of Sabah is the biggest rice producing area with approximately 10,000 ha and produces more than 33,000 tonnes of paddy (Lamdin et al., 2015; Souki, 2015). Rice cultivated areas in Sabah are comprises of wetland and hill/dryland paddy areas. In recent years wetland paddy remains the most important paddy type with small volume of dryland paddy (Idris, 2018). For the wetland rice cultivated areas are made up of three categories, namely K1, K2 and K3 with a total area of approximately 35,600 ha (Figure 2.4); K1 are fully irrigated, where irrigation facilities and water are always available, K2 areas are equipped with irrigation facilities as in K1 but water is not always available and K3 areas are mainly dependent on rainfall and no irrigation facilities (Lamdin et al., 2015). Sabah's SSL is only about 25% in 2016 whereas Sabah's imports of rice are quite large (Idris, 2018). In 2012, rice import is reported to be worth RM330 over million and in 2013 total local paddy production is only 35 % of the value of import the same year hence indicates the increased level of rice dependency (Rajamoorthy et al.; 2015, Idris, 2018). The state government has always increased the effort to decrease the dependency and measures to boost domestic production through certain project, policy or program such as the rice bowl project and State Agricultural Policy 3, aimed at raising the SSL of rice in the state (Lamdin et al., 2015; Idris, 2018).



PADDY FIELD CATEGORIES	AREA (HECTARES)
K1 - DID irrigation facilities available. Paddy yield potential is 5.1 - 10 tonnes per hectare	10,400
K2 - DID irrigation facilities available, but insufficient supply . Paddy yield potential is 2.6 - 5.0 tonnes per hectare	11,898
K3 - Rainfed. Paddy yield potential is 0.5 - 2.5 tonnes per hectare	13,303
TOTAL	35,601

Figure 2.4 Rice producing districts in Sabah (Hectares)
Source : extracted from Lamdin et al. (2015).

2.3 Agricultural weeds

"Weeds evolved and are still evolving within the man-made habitat three principal ways: (a) from colonizers through selection towards adaptation to continuous habitat disturbance, (b) as derivatives of hybridization between wild and cultivated races of domestic species and (c) through selection towards re-establishing natural seed dispersal mechanisms in abandoned domesticates." De Wet and Harlan (1975).

A weed might, be defined in various ways but, in broad sense, weeds are any plants species that thrives in man-made habitats and interfering with human activities (Radosevich et al., 1997). The relationships of crops and weeds are ancient, dating back to the beginning of agriculture or previous proto-agricultural period when human started domesticating plants through artificial selection (Harlan, 1982). Domestication leads to a

conscious selection of favorable traits including the ability to grow in densely planted environments, easily controlled mating, reduced seed dormancy and seed shattering, good nutritive or fiber quality, and adaptation to agricultural settings (Meyer and Purugganan, 2013; Olsen and Wendel, 2013). However, alongside with domestication process, some cultivated plants may escape cultivation and evolve in unintended ways which referred as de-domestication. Plant de-domestication is an evolutionary process in which plants develop genetically and phenotypically distinct "feral" populations through hybridization and/or adaptation (Gressel, 2005; Ellstrand et al., 2010). During de-domestication, the weedy populations reacquire those wild traits that were lost in the process of domestication such as seed shattering yet retaining some crop-specific traits especially the ability to thrive in planting field (Table 2.3) (Gressel, 2005; Ellstrand et al., 2010). Unfortunately compared to crop domestication, the de-domestication process is far less studied particularly in terms of genomic evolution (Ellstrand et al., 2010; Wedger and Olsen, 2018).

The result of the process of de-domestication is often referred to as "ferality" (Gressel 2005). As indicated in the introductory quote on mode of weed evolution by De Wet and Harlan, two models for ferality are proposed for the evolution of crop weed species: endoferality and exoferality (Gressel, 2005; Ellstrand et al., 2010) (Figure 2.5). Endoferality refers to weeds that directly descended from crop through ancestral standing variation or new mutation. One best example of crop origins of weedy plants would be weedy rye *(Secale cereale)*. Weedy rye, which is problematic in wheat fields was originally thought to be a hybrid derivative of cultivated rye and the wild perennial mountain rye *[S. strictum (C. Presl) C. Presl.]*. However, the genetic analysis with 14 allozyme and three microsatellite loci of several populations of North American weedy rye failed to detect any ancestry from *S. strictum* or any other wild *Secale*. Apparently, the weedy populations are more similar to cultivar and evolved directly from one or more cultivars of cereal rye as described in Burger et al. (2007). On the other hand, exoferality is a process which weedy forms arising from continued introgression of wild species. One such scenario can be found in the evolution of weedy sunflowers (*Helianthus annuus*) in Spain and France. The genetic analysis of 16 microsatellite markers revealed the European weedy sunflower

strains most probably originated from the unintentional pollination by wild plants growing nearby the seed production field (Muller et al., 2011).

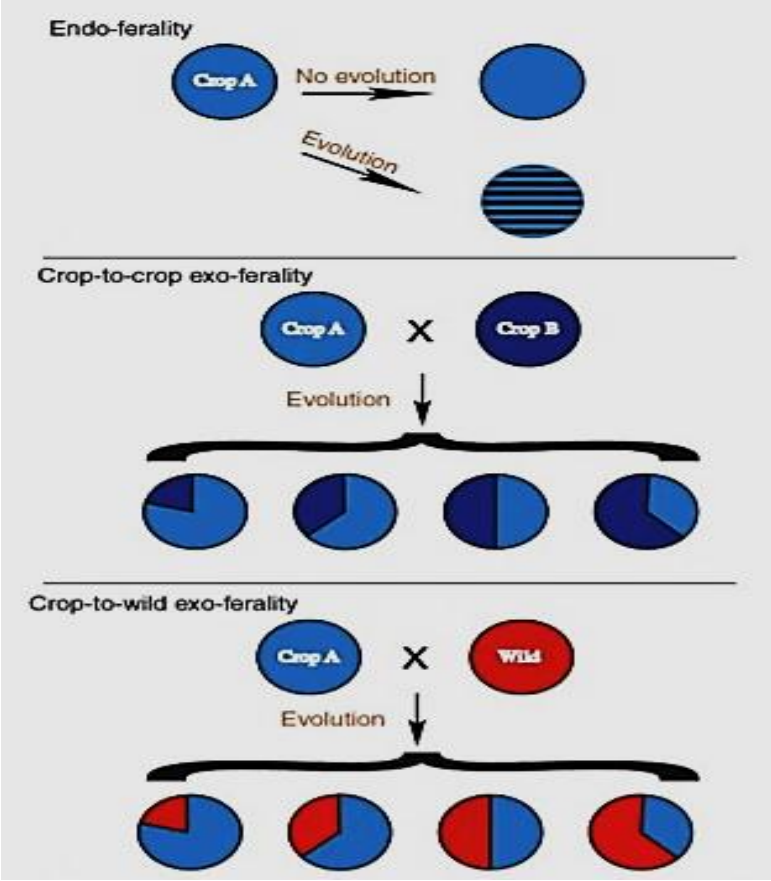


Figure 2.5 Models for the de-domestication of crops to weeds by endoferality and exoferality.
Source: adapted from Ellstrand et al. (2010)

Table 2.3 Components of Domestication, Weediness, and Wild Syndromes

Domestication Traits	Weedy or Invasive Traits	Wild Traits
Retention of the seed or fruit on the plant at maturity	Propagules that are adapted to long-distance dispersal and easily distributed	Propagules that are not adapted to long-distance dispersal
Loss of germination inhibitors	Seed dormancy	Seed dormancy
Synchrony in germination (loss of secondary dormancy)	Discontinuous germination (secondary dormancy)	Discontinuous germination (secondary dormancy)
Narrow germination requirements	Broad germination requirements	Special germination requirements
Short-lived seeds (no seed bank)	Long-lived seeds (seed bank)	Long-lived seeds (seed bank)
Synchrony of flowering and fruit development	Rapid growth to flowering, annual	Slow growth to flowering, perennial
More determinate growth	Continuous seed production for as long as growing conditions permit	More determinate growth
Reduction in seed dispersal (shattering)	Propagule (seed) shattering	Propagule (seed) shattering
Increase in vegetative vigor	Vigorous vegetative reproduction, if perennial	Not vigorous vegetative reproduction, if perennial
Adaptation to disturbed habitats	Adaptation to disturbed habitats	Not adapted to disturbed habitats

Source: adapted from Gressel (2005).

2.3.1 Weed problem in rice

Rice weeds appear as complex ecological entities. There are about 350 species have been reported as weeds of rice, of which grasses are ranked as first followed by sedges and broadleaf weeds (Holm et al., 1977). The composition of weed communities in rice fields is influenced by cultural, mechanical, chemical, and environmental factors (Matloob et al., 2014). The types of rice culture (irrigated, rainfed lowland, upland, deep water or tidal wetland), crop establishment method (transplanted or direct-seeded), moisture regime (irrigated and rainfed), land preparation (lowland or upland) and cultural practices (flooding, fertilizer application, cultivar types) are some of the principal determinants of weed spectrum and the degree of their infestation in rice field (Moody, 1993; De Datta and Baltazar, 1996). A list of major weeds found in rice fields in Asia has been presented in Table 2.4.

Table 2.4 Major weeds in rice fields in Asia

Weed category	Scientific Name	Family name
Grass	*Digitaria setigera*	Poaceae
	D. ciliaris	Poaceae
	Echinochloa colona	Poaceae
	E. glabrescents	Poaceae
	Eleusine indica	Poaceae
	Ischaemum rugosum	Poaceae
	Leptochloa chinensis	Poaceae
	Oryza sativa (weedy rice)	Poaceae
	Paspalum distichum	Poaceae
Sledge	*Cyperus iria*	Cyperaceae
	C.difformis	Cyperaceae
	C.rotundus	Cyperaceae
	Fimbristylis miliacea	Cyperaceae
Broad leaf	*Commelina benghalensis*	Commelinaceae
	Eclipta prostrata	Asteraceae
	Ipomoea aquatica	Convolvulaceae
	Ludwigia octovalvis	Onagraceae
	L. adscendents	Onagraceae
	Monochoria vaginalis	Pontederiaceae
	Sphenoclea zeylanica	Sphenocleaceae

Source: extracted from Juraimi et al. (2013)

On average, rice yield loss due to weed ranges from 15 to 20%, but in severe cases the yield loss may exceed 50% or even 100% (Mishra and Singh, 2007, Hasanuzzaman et al., 2009). This loss depends mainly on the season of crop sowing, weed species, weed density, rice cultivars, growth rate and density of weed and rice (Juraimi et al., 2013). In China, rice yield reduction caused by weeds is estimated at 10-20% as reported in Zhang (2003), while Savary et al. (1997) stated yield losses rice due to weeds ranged from 32-83% in India. It was estimated in Malaysia that the yield loss by grasses (mainly *Echinochloa crus-galli*), broad-leaved weeds and sedges was 41, 28 and 10%, respectively (Azmi, 1992, Karim et al., 2004). In Perak, Malaysia, rice weeds caused a yield reduction of 5–18% in the main season of 1989, which was equivalent to RM 70000.00 (US$17 500) (Mislamah, 1990). Yield reduction due to weeds is more critical in direct-seeded rice (DSR) than in transplanted rice (Karim et al., 2004). Jaya Suria et al. (2011) from their field trials in Penang, Malaysia, concluded that grasses were the most dominant category of weeds in DSR, constituting approximately 80% of the weed community. The increased incidence of weedy rice (*O. sativa f. spontanea*) in Asian rice fields has also been attributed to the adoption of DSR (Ferrero, 2003).

2.4 Weedy rice: A troublesome weed

Weedy rice is a noxious weed of rice and it belongs to the same species of cultivated rice (*O. sativa* L.). Weedy rice is a major threat to paddy fields worldwide since it has the typical properties of invasive weeds, such as strong reproductive ability, invasiveness, ability to compete for resources, and high phenotypic plasticity (Sun et al., 2019). Weedy rice known by diverse names in different countries for instance, ‘La Lon’ in Vietnam, ‘Luta’ in China, ‘Akamai’ in Japan, ‘Sharea’ in Korea, ‘Khai’ in Thailand, ‘ Jhor Dhan’ in Bangladesh, ‘Varinellu’ in India and ‘Padi angin’ in Malaysia (Watanabe, 1996). Morphological and physiological characteristics of weedy rice are almost the same as those of cultivated rice varieties (Azmi and Karim, 2008; Tewari, 2008). However, weedy

rice possesses considerable morphological diversity; these include the presence of red pericarps, greater dormancy, and high level of seed shattering (Thurber et al., 2010a; Vaughan et al., 2001). Several of these traits are also found in the wild ancestor of cultivated rice, *O. rufipogon*, and other wild *Oryza* relatives, but weedy rice differs from truly wild species in its adaptation to the agroecosystem and presence of some traits characterizing cultivated (Delouche et al., 2007). It is difficult to control weedy rice due to similar morphological, biochemical as well physiological features of cultivated rice (Tang and Morishima, 1997). Physical weed management is difficult as weedy rice are difficult to distinguish from the crop in early stages, and chemical weed management is limited as herbicides controlling weedy rice also kill rice crop plant (Ferrero, 2003; Nadir et al., 2017). Weedy rice has also been reported to adapt and survive in a habitat for many years, and this trait could play an important role in evolutionary studies on cultivated rice (Nadir et al., 2017).

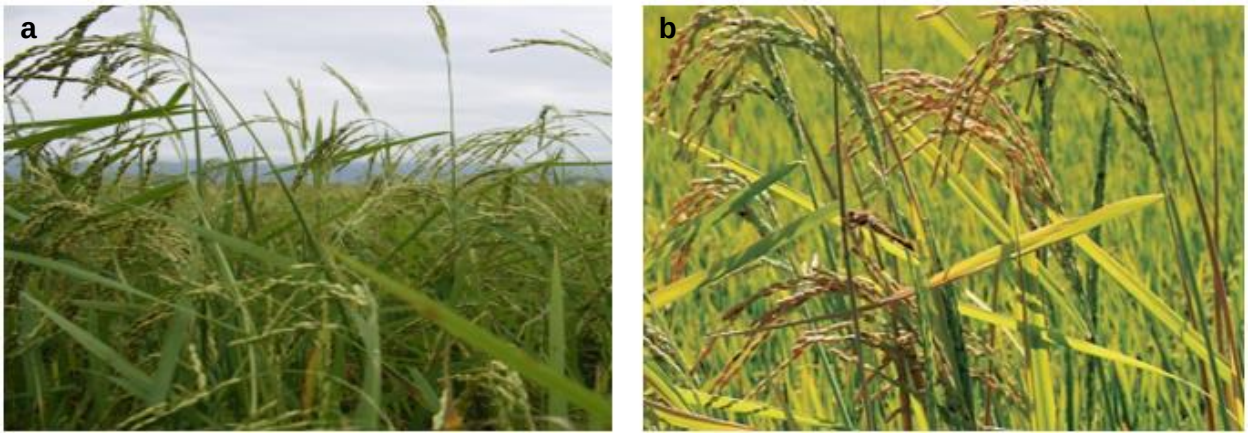


Figure 2.6 a) Weedy rice infestation in Kota Belud, Sabah. b) Weedy rice infestation in Sekinchan, Peninsular Malaysia.
Picture by Dr.Song Beng Kah

Weedy rice infestations were first reported from the Americas when red rice infestations were known to have occurred as early as 1846 as contaminants in imported seed rice (Allston, 1846). Weedy rice infestations in rice fields have been reported from European countries since the 1970s after the cultivation of weak, semi-dwarf *indica*-type rice varieties (Ferrero, 2003). The occurrence and infestation rate of weedy rice exploded around the globe due to and after the adoption of direct-seeded rice culturing (Fogliatto et al., 2010; Nadir et al., 2017). This contrasts with the traditional practice of transplanting seedlings from seedbeds to the paddy fields where weedy rice infestation was relatively in low level (Chauhan, 2013). In Malaysia, the phenomenon is said to have started in the 1980s (Karim et al., 2004; Azmi and Karim, 2008). In Thailand, weedy rice has become a serious problem over the whole country, after having been first observed in 2001 in Kanchanaburi province, where direct-seeded high yielding varieties were double cropped (Jindalouang et al., 2018). Currently, weedy rice has been reported in most of the rice-growing countries like China, India, the USA, Bangladesh, Bhutan, Brazil, Nepal, Thailand, Japan, the Philippines, Korea, Thailand, Sri Lanka, Vietnam, and Malaysia (Shrestha et al., 2018)

Yield losses due to weedy rice infestation have been reported from many parts of the world. Weedy rice infestation rates in European and US rice fields are estimated to be 30 and 70%, respectively (Catala et al., 2002; Gealy, 2005). In Arkansas, the highest rice-producing state in the USA, economic loss due to weedy rice has been estimated to be $274/hectare (Burgos et al., 2008). It is also reported that around 3 million hectares of rice fields in China have been infested with weedy rice, leading to a yield loss of about 3.4 million tonnes per year (Nadir et al., 2017; Rathore et al., 2016). According to Azmi et al. (2000), weedy rice caused yield losses up to 74% in Malaysia. If weedy rice-induced losses cannot not be controlled, they could adversely affect global food security and economy.

2.4.1 Spread of weedy rice

Transplanting is the main method of crop establishment in South East Asian countries. Transplanting of rice seedlings from nurseries into 'well-puddled' or flooded soils gives an advantage to rice over weeds due to seedling size and the flooded soil in which weed species must initially germinate and establish (Rao et al., 2007). In many areas, transplanting of rice and subsequent manual hand weeding have provide good weedy rice management, but it is labour intensive and requires considerable water for land preparation (Ismail et al., 2012). Most of the South East Asian countries are facing the problem of water scarcity and Tuong and Bouman (2003) reported that, in Asia, 39 million ha of irrigated rice may suffer from "physical water scarcity" or "economic water scarcity" by 2025. Besides, the farmers are also facing the problem of labour shortage and hike in wage rate due to migration of rural labour to urban areas. Ojha and Kwatra (2014) reported that transplanting requires 240 to 250 man hour per hector, which is 25 % of the total labour requirement of the rice crop. Due to these reasons, rice farmers are adopting the direct-seeded rice (DSR) method (Figure 2.7) (Rao et al., 2007; Rathore et al., 2016). DSR involves rice stand establishment directly by sowing seeds in the fields and uses less water and labours (Chauhan, 2013). However, direct rice seeding does not offer the weed suppression advantages that transplanted rice does, which leads to high weedy rice infestation and potential yield losses (Durand-Morat et al., 2018)

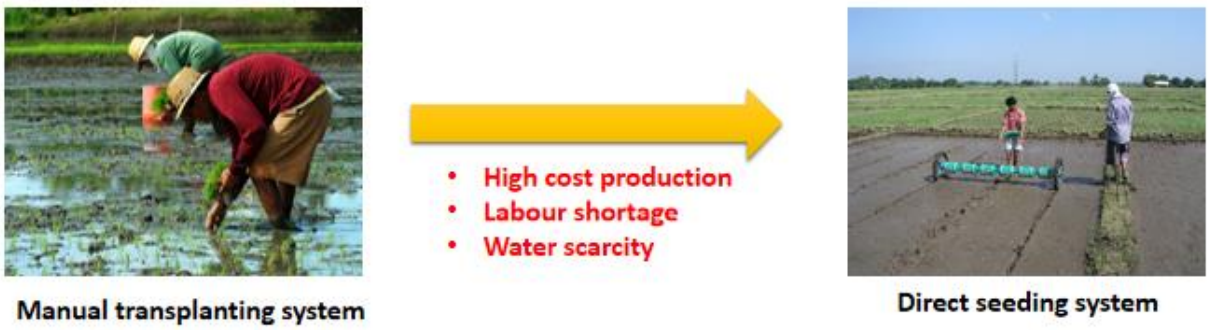


Figure 2.7. Shift of rice establishment method.

Weedy rice spreads rapidly from infested fields to new non-infested areas. Knowledge of the sources for the dispersal of weedy rice can help in preventing its spread to non-infested areas. Use of weedy rice contaminated seed stock, either purchased or exchanged, is the most important source of its spread to new areas (Saha et al., 2014). The use of weedy rice contaminated agricultural equipment/machinery, e.g. harvester also plays a vital role in its dispersal. It is therefore important to use weedy rice–free certified seeds and shared machinery should be cleaned before moving it to new areas to prevent weedy rice spread (Singh et al., 2013). In addition, weedy rice can also be dispersed from one field to another through irrigation channels or irrigation water, flooding and heavy winds or storms. The major dispersal mechanisms of weedy rice are summarized in Figure 2.8.

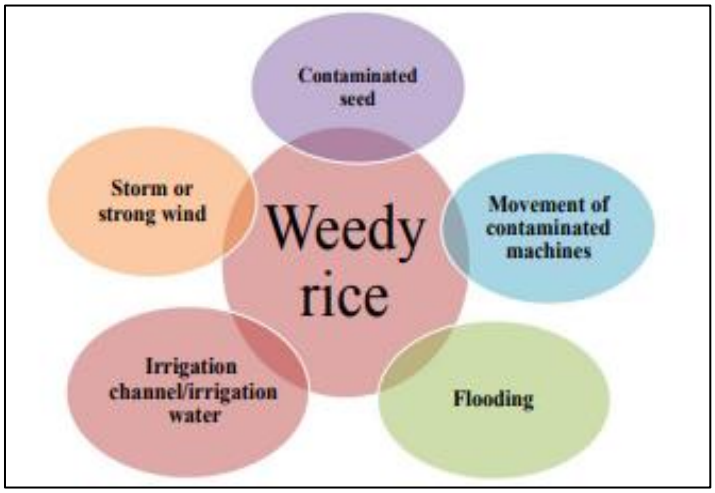


Figure 2.8 Sources of weedy rice spread or movement from infested areas to new areas Extracted from Singh et al. (2013)

2.4.2 Characteristic of weedy rice

Weedy rice has competitive advantage over cultivated rice as it grows taller and faster, tillers profusely and competes with cultivated rice for nutrients, light and space (Rathore et al., 2013). It flowers much earlier than cultivated rice and produces grains that shatter easily, thus enhancing the weed seed bank. Weedy rice has comparatively better resistance to adverse dry conditions and possesses long and varied seed dormancy in soil. Besides, weedy rice are diverse and exhibit more rapid seedling growth when compared to the cultivated rice to which they infest (Delouche et al., 2007). Characteristics and interference of weedy rice have been elaborated by Watanabe (1996) (Table 2.5). The three critical weedy traits; strong seed dormancy, greater seed viability and high seed shattering are crucial for the establishment and continuation of weedy rice as a serious weed problem in rice production.

Table 2.5: Characteristics and interference of weedy rice

Type of characteristics	Traits	Characteristics	Interference
Morphological	Culm	Long	Cause lodging, competitive edge over rice
	Grain	Short, pigmented grain, coloured pericarp, awned grain	Reduce rice quality
Ecological	Plant mimicry		Difficult to identify, escape weeding
	Thresh ability	Spontaneous shattering	Reduce rice yield, increase seed bank in soil
	Dormancy	Seed dormancy	Difficult to control
	Seed longevity	Viable in soil for several months to year	Difficult to reduce size of seed bank
	Germination	Variable	Adapt to wet-seeded fields
Physiological	Herbicide	Tolerant	Less effective with chemical control

Source: adapted from Watanabe (1996)

Seed shattering

Seed shattering is an undesirable agronomic trait in rice cultivation. The persistence of weedy rice has been partly attributed to its ability to shatter seeds (Thurber et al., 2011). Seed shattering trait, which distinguishes cultivated rice from its wild forms, is variable in weedy rice (Ferrero, 2003). In weedy rice, the abscission layer degrades earlier as compared to cultivated rice, leading to earlier shattering and increasing its fitness for survival in the environment (Thurber et al., 2011). Ferrero and Vidotto (1998) found that seed shattering in weedy rice started nine days after flowering and increased gradually for 30 days. A study from Vietnam reported that seed shattering time and percentage are not correlated and seed shattering in weedy rice varies from 20% to 95% in different seasons and biotypes (Chin et al., 2000). To date, 4 shattering genes with large effects have been cloned, *sh4* (Li et al., 2006), *qSH1* (Konishi et al., 2006), *sh-h* (Ji et al., 2010), and *SHAT1* (Zhou et al., 2012). Thurber et al. (2010b) concluded that the single nucleotide mutation in the *sh4* gene alone may not be sufficient to reduce shattering. Similarly, Zhu et al. (2012) reported the presence of the non-shattering *sh4* allele in all weedy rice varieties and in wild species with heavy shattering phenotype in high frequency. Hence, there are still unidentified shattering loci, which may have played important role shattering phenotype.

Seed dormancy

The success of weedy rice as weeds can be attributed to its high seed dormancy. Seed dormancy is a state of temporary developmental arrest and is an adaptive trait used by many plants (Nadir et al., 2017). Weedy rice seeds show dormancies that range from a few days to years depending on the biotype and the storage conditions of the seeds after shattering (Vidotto and Ferrero, 2000; Tseng et al., 2013). Variable seed dormancy is important in establishing weed populations as it allows seeds to germinate non-synchronously, thus avoiding adverse environmental conditions that many prevent seedling establishment at different times of growing season. Viable weedy rice seeds with red pericarp remained dormant up to three years in Brazil (Leitão et al., 1972) and up to ten years in the United States (Goss and Brown, 1939). From the inheritance study of

seed dormancy in weedy rice, Gu et al. (2003) concluded that weedy rice populations maintained a higher level of heritability for seed dormancy during a longer period after harvest. It was observed by Veasey et al. (2004) that the seeds of the dry region developed longer periods of seed dormancy, waiting for the wet season when environmental conditions again became favourable for germination and seedling survival.

Seed viability

The longevity of weedy rice seeds in the soil varies with ecotype and also affected by burial depth, soil type and moisture, cultivation practices, the magnitude of seed production (fecundity) and dormancy intensity (Noldín, 1995). A study conducted at two locations in Texas, USA, found that weedy rice seed longevity was greater with deeper seed burial (Noldin, 2000). The study found that seeds buried at 5 cm or kept at the soil surface had almost no viability after one year but, when buried at 12 and 25 cm, seeds remained viable longer than two years. Noldin et al. (2006) found that the differential level of dormancy among weedy rice ecotypes buried under the soil at different depths, and all of them were more viable than commercial rice cultivar. Five weedy ecotypes had viable seeds even after 36 months of burial in the soil whereas the commercial rice seeds were nonviable after 5 months of burial in the soil. Weedy rice therefore has greater viability than cultivated rice under certain environmental conditions and can emerge from deeper soil surface, thus forming persistent seedbank in the soil (Shrestha et al., 2018). Therefore, rice farmers should avoid deep tillage practices after the rice harvest in weedy rice infested fields because deep tillage places seeds deeper in the soil where the environment is less suitable for germination and seeds can remain viable longer (Singh et al., 2013)

2.4.3 Evolution of weedy rice

The genetic background of weedy rice varies considerably between countries, and arguments regarding its origin still debatable among scientific disciplines. However, the various hypotheses pertaining to the evolution of weedy rice have been supported by numerous studies. Three general routes to the origin of weedy rice have been proposed; (a) de-domestication of cultivated rice giving rise to weedy rice populations; (b) hybridization between wild rice species and cultivated groups; or, (c) direct colonization of wild rice species of rice agricultural fields (Londo and Schaal, 2007). Several genetic and phenotypic studies have supported de-domestication model in determining weedy rice evolution in different regions. For instance, in regions where the *japonica* subspecies is predominantly cultivated, weedy rice is phenotypically '*japonica*-like' (e.g. shorter and round grains); whereas in areas where the *indica* subspecies is cultivated, weedy rice is depicting '*indica* like' characters (e.g., longer grains, taller) (Ishikawa et al., 2005; Cao et al., 2006; Zhang et al., 2012). Similarly, genetic analysis using microsatellite and single nucleotide polymorphism (SNP) markers have implicated local crops of *japonica* or *indica* in the origin of weedy rice. For example, in Zhang et al. (2012), microsatellite technique with 21 pairs of SSR markers was utilized to estimate the genetic structure of two biotypes of weedy rice with *japonica* and *indica* rice characteristics, collected from Liaoning and Guangdong provinces, respectively. The result showed that weedy rice populations in the two provinces had very close genetic relationships with cultivated rice from the same field. Likewise, weedy rice in the Mediterranean region (where *japonica* rice was traditionally grown) was most closely related to *japonica* rice whereas in Brazil, where *indica* rice is typically grown, the weedy rice was also closely related to *indica* varieties (Ferrero, 2003).

De-domestication origins of weedy rice have even been demonstrated for cases where the ancestral cultivar is not local or endemic to the continent, such as in the U.S., where weedy populations descend from *aus* and *indica* cultivated varieties (Londo and Schaal, 2007; Gealy et al., 2009; Reagon et al., 2010). The U.S. weedy rice are genetically distinct from local crop varieties and likely evolved through de-domestication in Asia, with subsequent unintentional introductions into their present range (Reagon et al., 2010; Li et al., 2017).

The second hypothesis suggested that weedy rice originated from hybridization between cultivated rice and its progenitor type. In areas of tropical Asia where wild rice is present, weedy rice strains have typically been found to show some evidence of introgression from wild populations, although they are still primarily descended from domesticated rice (Cao et al., 2006; Huang et al., 2017). Studies in Thailand and Malaysia have documented hybridization between cultivated rice and wild *O. rufipogon* as a likely source of weedy rice origins (Prathepha, 2009; Pusadee et al., 2013; Song et al., 2014; Vigueira et al., 2019). In addition, weedy rice may also originate from ongoing and multidirectional hybridization between weedy rice and cultivated types as well as hybridization among weedy types (Londo and Schaal, 2007). For example, Xiong et al. (2012) demonstrated that weedy-type progenies could emerge after inter-subspecies (*japonica* x *japonica*) and intervarietal (*japonica* x *indica*) hybridization in rice (Qiu et al., 2014).

Some studies supported the notion that weedy rice was formed from the ongoing selection and adaptation of wild rice. In this case, the weedy rice is more likely to be “crop wild relative”; wild rice species that are genetically related to cultivated rice. These species are found growing around rice cultivation area, benefiting from the conditions in agricultural settings and invade the rice fields when time is favourable. In environment where no native wild rice is present, these weedy rices can be introduced through seed movement. For example, the weedy rice in Brazil is hypothesized to originated from Africa *via* trading during the colonial era (Carney, 2004).

2.5 Malaysian weedy rice scenario

Among rice growing regions of Southeast Asia, Malaysia presents a dynamic evolutionary history of weedy rice emergence (Song et al., 2014; Neik et al., 2019). Malaysia comprises of two distinct geographical regions (West and East Malaysia) separated by the South China Sea and these two regions have unique weedy rice history at two distant timelines. In Peninsular Malaysia (the west part of Malaysia; also known as the mainland), weedy rice was first reported in 1988 in Sekinchan, Selangor (Wahab and Suhaimi, 1991). Since then, the infestation was fast spread to other rice granaries in Peninsular Malaysia areas including Perak, Terengganu and Kedah with wide phenotypic variations (Baki, 2004; Anuar et al., 2014; Sudianto et al., 2016). More than 50 per cent of rice granaries in Selangor and Terengganu were reported to be infested with weedy rice (Azmi et al., 2000; Anuar et al., 2014; Mispan et al., 2019), while in weedy infestation level ranging from less than 10 per cent to more than 20 per cent coverage (Azmi et al., 2005). As with other Southeast Asian countries, one of the major factors contributing to the spread of weedy rice in Malaysia has been the change in rice establishment methods from transplanting of seedlings, which provides opportunities for hand-weeding of fields, to direct seeding (Sudianto et al., 2016). The emergence and fast spread of weedy rice were also resulted from poor land preparation which increased the survival fate of weedy rice in the seedbank (Azmi and Karim, 2008). It was reported that the national rice yield loss can be projected to 64,880 tons by only 5 per cent weedy rice infestation and can lead to monitory loss exceeding $20 million (Baki, 2004; Anuar et al., 2014). Previous morphological and genetic studies of Peninsular Malaysian weedy rice revealed that that both local wild rice populations and modern elite cultivars, especially MR series, have likely contributed to the composition of contemporary Malaysian weedy rice populations (Song et al., 2014; Cui et al., 2016; Sudianto et al., 2016; Vigueira et al., 2019).

The occurrence of weedy rice in the local fields was first observed in Kota Belud and Kota Marudu districts in late 2000s (Teo SS, unpublished observations), and has become a widespread problem in other rice production areas since the 2000's. Depending on the severity of infestation, it could cause reduced rice grain quality and up to 90% yield loss

in the local rice fields, and eventually lower farmer income (The Star, 2009). In 2009, it is reported about 1,100 ha of paddy fields in Kota Marudu and 500 ha in Kota Belud were affected by the weedy rice (The Star, 2009). Despite no wild rice (*O. rufipogon*) is native to Sabah state, extensive morphological diversity in vegetative and reproductive traits were observed (B. K. Song, unpublished observations). This implicated a dynamic evolutionary and complexity of weedy crop relatives in Sabah state. Given that Sabah weedy rice occurrence was observed a decade later than the first reports of the weed in Peninsular Malaysia, the existence of planting seed sharing norm among farmers in both Peninsular Malaysia and Sabah, and the fact that some weedy rice forms were found to share phenotypic similarities for some wild traits with their weedy counterparts from the peninsular (e.g. awned seed with open panicles, high plant stature, open tillers; B. K. Song, unpublished observations), suggest that emergence of these weedy populations in Sabah may likely originated through accidental introduction from the western part of the country (Neik et al., 2019). Little is known about origin and genetic background of weedy rice in Sabah. A very recent study by Neik et al. (2019) on Sabah weedy rice evolution using SSR markers showed considerable contributions of Peninsular Malaysian weedy rice ecotypes to Sabah populations.

2.6 Plant Architecture and tiller angle

Rice plant architecture is an important agronomic trait that influence global rice production (Peng et al., 1999). Through the course of rice domestication, several important morphological traits have been artificially selected by farmers to have an ideal plant architecture (Jin et al., 2008). A significant step in the domestication of rice was the shift from the prostate growth habit of wild rice species to more compact growth in modern cultivated rice. The selection of rice plants with ideal tiller angle is one the most crucial event in rice domestication for effective high yield grain production. The recent advancement of cloning and characterization have shed light on the molecular mechanisms that control rice tiller angle. The first insight of tiller angle mechanism came

from the analysis of *LAZY1 (LA1),* which was isolated from spreading mutant *la(k)* and controls tiller angle by regulating shoot gravitropism through negative regulation of basipetal polar auxin transport (PAT) (Figure 2.9) (Li et al., 2007a; Yoshihara and Iino, 2007). Further understanding on the role of auxin transport in regulating rice tiller angle shown in the studies of *OsPIN1 and OsPIN2* where over-expression of these auxin efflux transporters lead to changes in tiller angle formation (Chen et al., 2012; Xu et al., 2016). Besides, PROSTATE GROWTH 1 (*Prog1*) is also an important regulator of more vertical tiller growth. *Prog1* gene encodes a single Cys-His zinc-finger protein suggesting the protein may functions as transcriptional factors (Jin et al., 2008, Tan et al., 2008). In the work presented by Jin et al. (2008), molecular evidence provided that *Prog1* was subjected to selection during domestication with strong signature of selection in the coding region. An orthologue of *PROG1* gene, *PROG7* was recently identified in African domesticated rice, *O. glaberrima,* and found that the variations in the promoter regions reduces expression level in the tiller base, resulting in the erect stature of African cultivar (Hu et al., 2018). *TAC1*, another protein encoded gene related to *LAZY1* is a major quantitative trait locus controlling rice tillering. A single nucleotide change from 'AGGA' to 'GGGA' in the 3' splicing site of fourth intron resulted in low expression of *TAC1* ultimately leading to erect tillers (Yu et al., 2007). Recently, the QTL for *TAC3* and *qTAC8* loci controlling tiller angle have been identified (Dong et al., 2016; He et al., 2017).

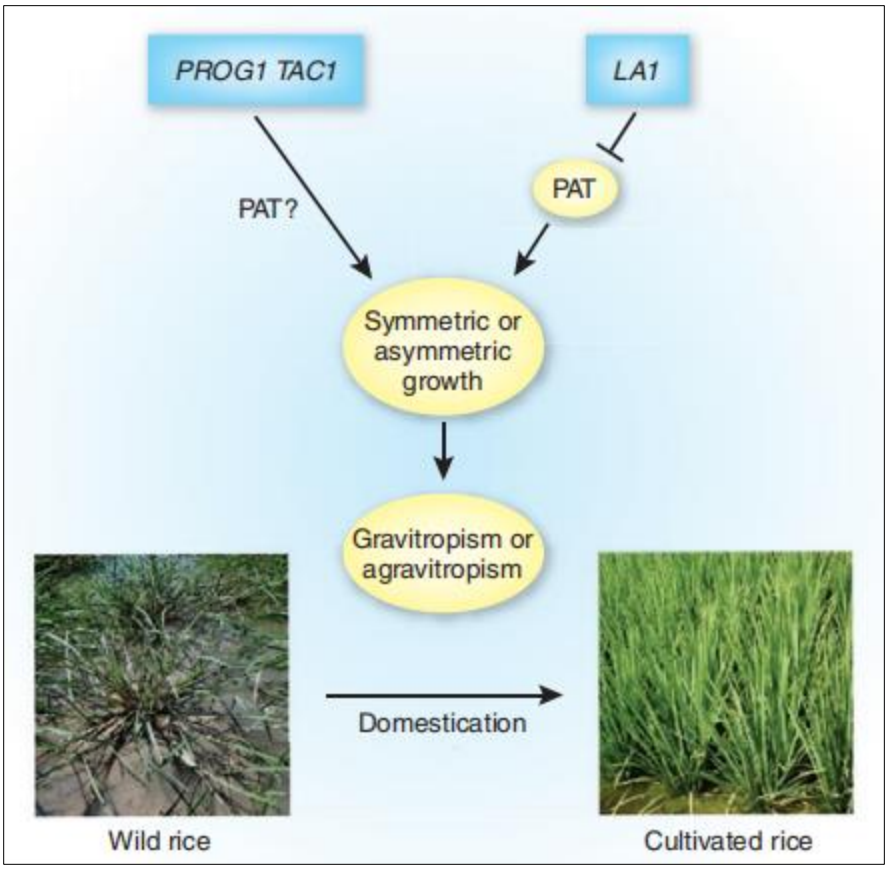


Figure 2.9 Genes and their proposed functions in controlling rice tiller angle. *LA1* negatively regulates rice tiller angle through the polar auxin transport (PAT)-mediated pathway. *PROG1* and *TAC1* are positive regulators that determine rice tiller angle through an unknown pathway.
Source : extracted from Wang and Li (2008)

2.7 Genotyping-by-sequencing (GBS)

The advent of next-generation sequencing (NGS) technologies holds the potential to dramatically impact the crop breeding improvement process. NGS enables whole-genome sequencing (WGS) and re-sequencing, metagenomics, transcriptome sequencing as well as high-throughput genotyping, which can be applied for genome selection (Gedil et al., 2016). It can also be applied to diversity analysis, genetic and epigenetic characterization of germplasm (Guo et al., 2014). Recently, NGS has enabled rice genetic improvements by identifying hundreds of key functional genes or by establishing associations between SNPs and agronomic traits. Various next-generation-based reduced representation protocols have been developed for genome-wide SNP data that have been applied to linkage mapping, quantitative trait locus (QTL) analysis, genome selection, and population genetics analysis. The most widely used assays for complexity reduction genotyping are restriction-site-associated DNA (RAD) (Baird et al., 2008) and genotyping-by-sequencing (GBS) (Elshire et al., 2011), and diversity array technology (DArT)-seq (Kilian et al., 2012). Researchers around the world presently focusing the reduced representation platforms such as GBS for genotyping rice, maize, cassava and banana for diversity analysis and molecular breeding (Huang et al., 2014, Gedil et al., 2016).

The GBS method incorporates a multiplex sequencing strategy (96 to 384 samples in parallel on a single run) for constructing reduced representative libraries for the Illumina NGS platform that uses an inexpensive barcoding system for increased efficiency at a lower cost compared to other genotyping methods (Elshire et al., 2011). Methylation-sensitive restriction enzymes are often employed to reduce the complexity of the genome and specifically to avoid sequencing through repetitive (methylated) DNA (Chung et al., 2017). Additionally, by choosing appropriate restriction enzyme(s), GBS can provide high SNP coverage in gene-rich regions of the genome in a highly cost-effective manner (Gore et al., 2007; Gore et al., 2009). Generally, GBS is simple, specific, highly reproducible, and rapid due to the simultaneous detection of SNPs and genotyping (He et al., 2014). The key components of this system have a lower cost, reduced sample handling, fewer

PCR and purification steps, no size fractionation, no reference sequence limits, and efficient barcoding, and the system is easy to scale up (Davey et al., 2011; Glaubitz et al., 2014). These features undoubtedly make GBS as a powerful tool for plant genomic studies. For instance GBS has been used in development of high density map of 20000 SNPs in wheat and 34000 SNPs in barley (Poland et al., 2012a) and to map QTLs for spike architecture and reduced plant height in barley (Liu et al., 2014). Besides, Moumouni et al. (2015) reported a genetic map of pearl millet using GBS method. They generated high-quality SNPs to construct a genetic map with an average interval of 2.1 (± 0.6) cM between the SNP markers. Their study demonstrated that GBS can quickly produce a denser and more uniform genetic map than previously published maps (Poland et al., 2012b).

CHAPTER 3

MATERIALS AND METHODS

3.0 MATERIALS AND METHODS

3.1 Population analysis of Sabah weedy rice

3.1.1 Plant materials and sequence data

A total of 182 sequence data used in this study (Appendix 1), including both newly generated DNA sequences data and previously sequenced accessions. A total of 89 rice accessions comprising 61 Peninsular Malaysia and 20 Sabah weedy rice populations were collected in 2011 and 2012 by Dr. Song Beng Kah (Monash University Malaysia). The Peninsular Malaysia samples were collected across three major rice planting areas (northwestern, northeastern and central-western) and Sabah accessions were collected from two main rice producing districts of Sabah state, Kota Marudu and Kota Belud across 10 rice planting sites (Kampung Balagaton, Kampung Masalog, Kampung Sangkir ,Kampung Longob, Kampung Mangkulua, Kampung Nolatan, Kampung Tamau, Kampung Taun Usik, Kampung Tawadakan and Kampung Telangtang) (Appendix 2). These accessions were then classified into six major morphotype groups which include straw hull (SH), straw hull awned (SHA), brown hull (BR), brown hull awned (BRA), black hull (BH) and black hull awned (BHA) (Appendix 3). These rice samples were then grown in the plant house of Monash University Malaysia. In addition to these samples, analyses also included GBS dataset derived from Vigueira et al. (2019) consisting of 44 accessions including domesticated rice (2 Malaysian elite cultivars), 22 Southeast Asia weedy rice (5 Indonesia, 2 Vietnam, 8 Thailand, and 7 Cambodia) and wild species (5 *O. rufipogon* and 6 *O. officinallis*). Besides, an additional 46 sequenced data (41 Sabah weedy rice and 5 Sabah cultivars) were provided by Dr. Song Beng Kah's group (Monash University, Malaysia).

Figure 3.1. Sampling locations of Malaysian *Oryza* accessions.
Sampling location were shaded in colour. purple: Perlis; green: Kedah; brown: Penang, yellow: Perak; orange: Selangor; pink: Kelantan; blue: Terenganu; red: Sabah

3.1.2 Rice planting and DNA extraction

Before planting the rice seeds were surface strelized with distilled water, then soaked in water at 70 ˚C and left for 3 hours at room temperature. The seeds, five replicates per accession were then germinated at 27˚C for 3 to 4 days in petri dishes lined with moist filter papers. Germinated seeds were transplanted to 22-litter pots filled with mixture of soil. The pots were flooded with water to mimic the paddy field condition and fertilizer and herbicide were applied at 2 weeks in a timely manner. Seedling were thinned to one per pot when seedlings reached 4 to 5 leaves growth stage. Due to limited space in greenhouse, the growth experiment was carried out in four stages from March 2017 to May 2018. Young leaves were obtained from 4-week old seedlings for DNA extraction using DNeasy Plant Mini Kit (Qiagen, Germany) following manufacture's instruction. Only high quality DNA samples, a concentration higher than 25 ng/ml and purity ~1.8 ratios for A260/ A280, were used for the following NGS based GBS experiments. The concentration of DNA was quantified by Qubit® 3.0 Fluorometer while the quality of the extracted DNA

was assessed through a 1.0% agarose gel electrophoresis and Biodrop µLITE UV-Visible Spectrophotometer (Biodrop®, Cambridge,UK).

3.1.3 GBS Library Preparation

Genotyping-by-sequencing (GBS) libraries were prepared using NEBNext® Ultra DNA Library Prep Kit for Illumina, (New England Biolabs, USA), with slight modification to the Elshire et al.,2011 protocol. Briefly 1500ng of genomic DNA from each of the samples were digested by 5 Units of ApeKI enzyme (New England Biolabs, USA) with an incubation at 75˚C for 3 hours. The quality of digested DNA was assessed by 1.5% of agarose gel electrophoresis. The digested DNA was then purified by using 1.5 x digested DNA volume of Agencourt AMPure XP (Beckman Coulter, USA). End repair and adenylation on sticky ends was performed along with adapter ligation. Another round of purification was carried out with 1.2 x DNA volume of Agencourt AMPure XP. The purified genomic DNA was then be subjected to polymerase chain reaction, PCR and amplified using barcode index/i7 primer and universal PCR primer/ i5 primer with initial denaturation at 98˚C, for 30 seconds, followed by eight cycles of denaturation and annealing/ extension at 98˚C for 10 seconds and at 65˚C for 75 seconds and a final cycle of extension at 65˚C for 5 minutes. The PCR with indexing Primers creates a tag for each library. NEBNext® Multiplex Oilgos for Illumina® (Index Primer Set 2) contains 24 primers (8*i5 and 12*i7 primers). Each primer contains the motif for the primers used in the sequencing later on and a variable sequence for the tag. A library tag is created of the combination of the i5 and i7 variable sequence. Thus, 96 tags can be created by using 24 primers combination. The final purification was performed using 1.0 x DNA volume of Agencourt AMPure XP. 2 µl of each library were quantified on the Qubit® 3.0 Fluorometer (Invitrogen, USA) using the high sensitivity kit (dsDNA HS Assay, Invitrogen, USA) and 5 µl were run on a 1.5% agarose gel to visualize the fragment distribution. Libraries which centred around the 200 bp - 300 bp range band in the marker were considered as successfully constructed and taken in for the further analysis. The libraries were then futher qualified through Agilent 2200 Tapestation (Agilent Technologies, USA). The barcoded libraries were sequenced

on Illumina Hiseq2000 sequencer (Macrogene Inc, Korea) with paired end 100 bp chemistry.

.

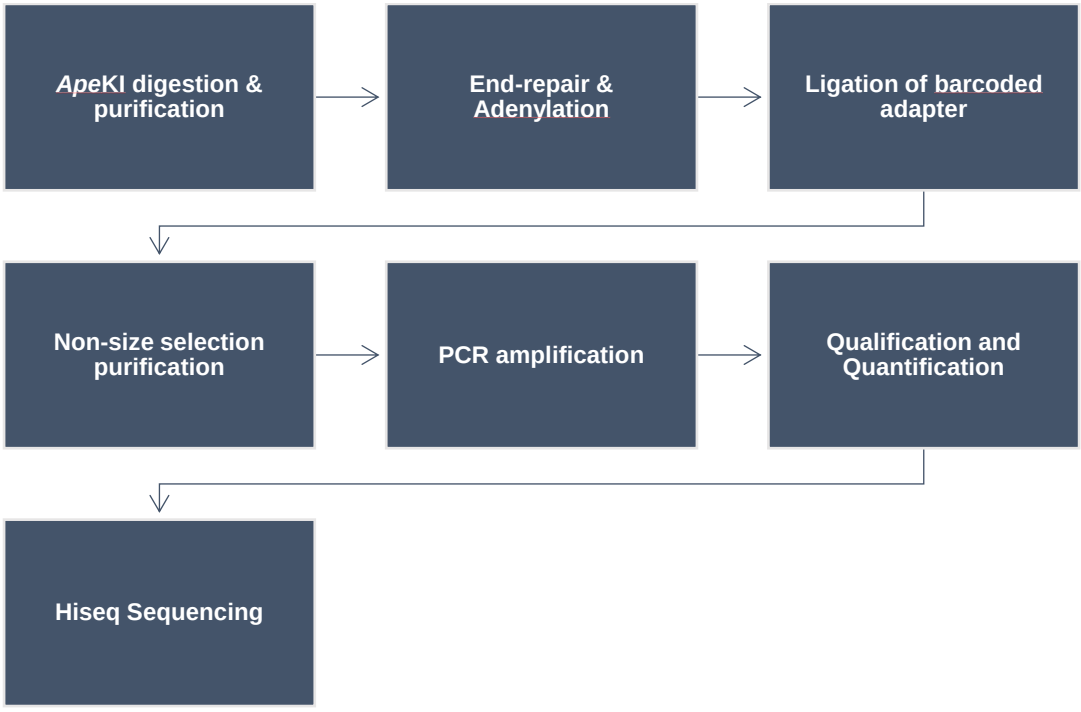


Figure 3.2.Schematic diagram of GBS library preparation

3.1.4 GBS-Raw Sequence Processing, Quality Filtering through TASSEL

Raw sequence reads in FASTQ format obtained from Hiseq sequencing will be processed using a standard TASSEL-GBS 5.0 pipeline (http://www.maizegenetics.net/) (Bradbury et al., 2007). The first step in discovery pipeline will filter out raw reads with N in the first 64 bases as well as reads without barcode and the expected 4-bp remnant of ApeKI cut site (CWGC). Tags comprising fewer than two reads of identical sequence and with less than minimum quality score of 20 were also discarded. All filtered tags were then aligned to the rice genome MSU 6.0 assembly using Burrow-Wheeler alignment (BWA) tool. SAM

tool was employed to read SAM files and covert to TagsOnPhysicalMap (TOPM). Tags positioned at the same physical location are aligned against one another and SNPs will be called from each tag alignment based on the following parameters: minimum minor allele frequency (mnMAF) of 0.01 and minimum locus coverage (mnLCov) of 0.2. The options -mnMAF can be used to filter out SNPs with rare minor alleles that possibly result from sequencing errors. While the -mnLCov option can be used to filter out SNPs with very high amounts of missing data from the output (Glaubitz et al., 2014). The SNP genotypes will be recorded in a HapMap (HMP) format genotype file per chromosome. After these filtering steps, a total of 441973 SNPs were yielded. A 'Distance filter' step was used to discard close proximity SNPs with less than 1.5 kb spacer and resulted in a total of 11325 SNPs. With further filtering using Tassel 5.0 Graphical User Interphase based on minimum taxa coverage (mnTCov) of 0.2 and minimum site coverage (mnScov) of 80% parameter, a total of 5379 high quality SNPs were retained for population genetic analysis.

3.1.5 Data Analysis

3.1.5.1 Genetic and nucleotide diversity analysis

The genotypic data matrix based on the 5379 loci of 182 *Oryza* samples representing 15 populations were analyzed to estimate their genetic diversity. The following parameters were calculated: (i) percentage of polymorphic loci (*P*), (ii) observed heterozygosity (H_o), (iii) expected heterozygosity (H_e), (iv) Shannon's information index (*I*), (v) number of private alleles (M_A), and (vi) fixation index (*F*). The statistical analyses were undertaken using the software GenAlEx version 6.5 (Peakall and Smouse, 2012).

A sliding window analysis for each identified rice population was performed to estimate relative nucleotide diversity between each rice groups across the genome. SNPs were converted from HAPMAP format to VCF format using TASSEL (Bradbury et al., 2007), and the genome-wide nucleotide diversity (π) were calculated using VCFtools (Danecek

et al., 2011) for 300-kb window size with 100-kb step size across the genome. The mean nucleotide diversity and variation were calculated and visualized in R.

3.1.5.2 Population Structure Analysis

The population structure for Malaysian weedy rice was determined using STRUCTURE 2.3.4 (Pritchard et al., 2000). The analysis was conducted with correlated allele frequencies and an admixture model for K values ranging from 2 to 10. Each run was conducted with burn-in period of 50,000 steps and 50,000 Monte Carlo Markov Chain (MCMC) replicates. Five independent runs were done for each value of K to generate our estimate of the true number of sub populations (Pritchard et al., 2000). Software program Structure Harvester v6.0 (Earl and vonHoldt, 2012) were used to identify the optimal number of populations (K), with the delta K statistic (Evanno et al., 2005) used as the selection criterion for optimal K (Appendix 4). The graphical representation of the population structure was generated ClumpaK (Kopelman et al., 2015), a web server that provides a full pipeline for clustering, summarizing and visualizing Structure results.

3.1.5.3 Genetic relationship analysis

A principal coordinate analysis (PCoA) was performed, using the EIGEN procedure in GenAlEx version 6.5 (Peakall and Smouse, 2012). The PCoA examines overall genetic similarity among individuals by using EIGEN value analysis to reduce the genetic variation between populations into two dimensions scatter plot. The differences between individuals (groups or clusters) are plotted on the dimension between the two first axes.

To estimate the relationships of weedy rice populations, Maximum Likelihood tree was constructed with the rate variation among sites was modeled with a gamma distribution (shape parameter = 1) and with 1,000 bootstrap iterations in Molecular Evolutionary Genetics Analysis version (MEGA) version 5 (Kumar et al., 2016)

3.2 Candidate gene analysis

3.2.1 Plant materials and sequence data

Samples used in this study included both newly generated DNA sequences and previously sequenced accessions that are publicly available. For newly generated DNA sequences, the panel set consisted of 52 *Oryza* samples, a subset of those accessions used in previous genotyping-by-sequencing (GBS) analysis: 21 Peninsular Malaysian weedy rice samples, 23 Sabah weedy rice samples, 2 Peninsular elite cultivars, 2 Sabah traditional cultivars and 4 Malaysian *O. rufipogon* samples. Accessions were chosen to represent the degree of tiller erectness covering all major groups of weedy population (SH, SHA, BR, BRA and BH). Sequence data of an additional 5 Malaysian *japonica* cultivars from Rice SNP-Seek database were included in this analysis as potential sources of weedy rice alleles. For phylogenetic analyses, a total of 34 geographically diverse *Oryza* samples were downloaded from Rice SNP-Seek database; these were 32 cultivated rice varieties (10 *tropical japonica*, 17 *indica* and 5 *aus* cultivars) and 2 wild ancestors (*O. rufipogon*). In addition, *O. nivara*, a wild rice from South America, *O. glaberrima*, a cultivated African rice, *O. barthii*, the wild ancestor of domesticated African rice and *O. meridionalis*, a species native to Oceania were included as potential progenitors of the weedy forms and as outgroups.

3.2.2 Scoring of tiller morphology

Thirty days after seeding, the plants were culled to leave one healthiest plant in each pot; five seedlings were planted in each pot initially. Upon heading, typically three months after germination, the tiller angle of each plant was measured. Tiller angles were estimated between the main tiller and the vertical culm where the angles were manually calculated by measuring the angle of the tangent of each lateral tiller point or tip, with respect to the upper main stem. We divided tiller angle into two categories, < 40° (erect tillers), and 40°-90° (open tillers). The angle was designated as ‘0’ for plants with erect tillers and compact architecture whilst ‘1’ was designated as open tillers with prostrate plant architecture

(Appendix 7). The degree of tiller erectness evaluated in this study and the measurement methods were based on rice descriptors published by the International Rice Research Institute.

3.2.3 DNA extraction and sequencing

Freshly excised young leaves were ground in a mortar and pestle containing liquid Nitrogen and genomic DNA was extracted using the Qiagen DNeasy Plant Mini Kit according to the manufacturer's instructions (Qiagen, Germany). DNA purity was checked on 1% agarose gel electrophoresis, and concentration and quality were measured by using Biodrop®. The targeted gene was the 3.14-kb genomic fragment harboring *TAC1* (LOC_Os09g35980), obtained from Yu et al. (2007). A total of 6 pairs of primers were designed using Primer3 (Rozen and Skaletsky, 2000) based on the *O. sativa spp. japonica cv. Nipponbare* genome (MSU 6.0 assembly), taking into account previously described polymorphisms in Yu et al. (2007) and Jiang et al. (2012a) (Appendix 8). PCR amplifications were carried out in 25 ul reactions containing the following: 1x Promega GoTaq Flexi Buffer, 2.0 mM $MgCl_2$, 200 µM dNTPs, 0.2 µM of forward and reverse primers, 1.25 unit of GoTaq polymerase and 25 ng of genomic DNA. The reactions were performed on a 96 well plate thermocycler, MyClycler (BIO RAD, USA) with 95 °C initial denaturation for 2 minutes, followed by 35 cycles of 30 seconds denaturation at 95 °C, 45 seconds annealing at 52°C and 1 min elongation at 72 °C. The reaction was terminated with a final elongation step of 72°C for 5 minutes and then cooled down to 4 °C. Direct sequencing in forward direction was carried out by First BASE Laboratories Sdn Bhd (Malaysia).

3.2.4 Genetic analysis

The obtained sequences were aligned by Sequencher® version 4.10.1 software (Codes, 2011) with further manual refinement and localization of relevant polymorphisms. Information about un-translated regions, exons and introns were obtained from the EnsemblPlants database (http://plants.ensembl.org/index.html) and Rice Genome Annotation Project database (http://rice.plantbiology.msu.edu/). Genetic variation parameters including nucleotide diversity (π), singleton site number (S), indel number (I), the proportion of segregating sites (θ), haplotype diversity (hd) and average number of nucleotide differences (K) were calculated using DnaSP v6.00 (http://www.ub.es/dnasp) (Rozas et al., 2017). Significance of Tajima's D value was also tested to detect the depature from neutrality.

3.2.5 Phylogenetic analysis

A haplotype network was constructed for comparison of phylogenetical relationship among *TAC1* haplotype using PopART (Leigh and Bryant, 2015) using TCS mode. Genealogical relationships among *TAC1* alleles were observed in Neighbour Joining (NJ) analyses as implemented in Mega 7 (Kumar et al., 2016). Distance were calculated using the Kimura 2-parameter modal along with gamma nucleotide substitution model. Pairwise deletion was considered with branch bootstrap values calculated *via* 1000 replicates.

CHAPTER 4

RESULTS

4.0 RESULTS

4.1 Population analysis of Sabah weedy rice

4.1.1 Genetic diversity analysis

To analyze the genetic diversity of Malaysian weedy rice, basic statistics were calculated to determine the mean of observed heterozygosity (H_o) and expected heterozygosity (H_e), Shannon diversity Index (*I*), fixation index (*F*), the number of private alleles (*MA*) at each locus and percentage of polymorphic alleles for each population. More than half of the total accessions had at least 50% of polymorphic loci *(P).* The H_o values in this analysis ranged from 0 to 0.3 with an average of 0.146 and the fixation index also varied between 0 and 1.00 with an average of 0.315 (Table 4.1). The *He* values computed in this analysis varied from 0.002 to 0.37 with an average of 0.212. Gene diversity among Malaysian weedy accessions found in this study (H_e = 0.219) was similar to the gene diversity reported by Song et al. (2014) (H_e = 0.368). Comparison of genetic diversity patterns among weedy rice indicate that South Asian weedy rice (Vietnam, Thailand, Cambodia, Indonesia; VTCI) is the most diverse group with the highest heterozygosity (H_e = 0.373) and high percentage of polymorphic loci (82.71%). This high genetic diversity could be driven by multiple sources of genetic input in VTCI weedy category (samples from 4 different geographical region).

Genetic diversity were comparable for both geographical regions, Peninsular Malaysia and Sabah, with the highest value detected in Sabah weedy rice population (H_e = 0.22, *I* = 0.357). These results complement previous observations on weedy populations from United States (H_e = 0.270, Gealy et al., 2009) where no native wild rice was found. When comparing the Malaysian weedy rice based on their morphotypes, gene diversity was found to be highest in the peninsular brownhull-awned (PM-BRA) population (H_e = 0.294, *I* = 0.484) which is consistent with Song et al. (2014), where highest diversity was observed in BRA population with H_e = 0.384. In Peninsular Malaysia, a high level of gene

Table 4.1. Genetic diversity of weedy rice, cultivated and wild rice samples based on 5379 SNP loci.

Morphotype	*N*	*%P*	H_o	H_e	*I*	M_A	*F*
Weedy rice							
Peninsular Malaysia							
Peninsular-Strawhull (PM-SH)	15	49.38	0.045	0.117	0.192	0	0.458
Peninsular-Strawhull Awned (PM-SHA)	8	78.03	0.166	0.272	0.438	4	0.299
Peninsular-Brownhull (PM-BR)	16	34.91	0.019	0.112	0.170	2	0.787
Peninsular-Brownhull Awned (PM-BRA)	12	83.44	0.262	0.294	0.484	12	0.058
Peninsular-Blackhull (PM-BH)	4	73.82	0.218	0.248	0.399	4	0.032
Peninsular-Blackhull Awned (PM-BHA)	6	73.02	0.292	0.276	0.438	9	0.000
Pooled	**61**	**65.43**	**0.167**	**0.219**	**0.354**	**5.16**	**0.272**
Sabah							
Sabah-Strawhull (SB-SH)	30	72.58	0.126	0.265	0.423	9	0.465
Sabah-Strawhull Awned (SB-SHA)	6	38.91	0.081	0.145	0.222	1	0.351
Sabah-Brownhull (SB-BR)	26	85.83	0.116	0.250	0.426	15	0.438
Pooled	**62**	**65.77**	**0.107**	**0.220**	**0.357**	**8.33**	**0.418**
All Malaysian weedy	**123**	**66.10**	**0.147**	**0.219**	**0.354**	**6.22**	**0.328**
South-Asian							
Vietnam-Thailand-Cambodia-Indonesia (VTCI)	22	82.71	0.177	0.373	0.602	13	0.506
All weedy	**145**	**67.56**	**0.150**	**0.235**	**0.379**	**6.9**	**0.335**
Cultivars							
Peninsular-Cultivar (PM-CV)	8	14.58	0.021	0.051	0.077	1	0.506
SB-Cultivar (SB-CV)	6	64.21	0.264	0.249	0.387	1	0.000
All cultivar	**14**	**39.40**	**0.143**	**0.150**	**0.232**	**1**	**0.253**
Wild species							
Malaysian *O.rufipogon*	7	69.06	0.305	0.318	0.493	41	0.028
VTCI *O.rufipogon*	11	52.59	0.096	0.203	0.319	20	0.474
Malaysian *O.officinalis*	5	0.69	0.002	0.002	0.004	21	0.334
All wild	**23**	**40.78**	**0.134**	**0.174**	**0.272**	**27.33**	**0.372**
Total	182	58.25	0.146	0.212	0.338	10.2	0.315

Abbreviations of parameters: *N* – number of accessions in population; *P (%)* – percentage of polymorphic loci; H_o – observed heterozygosity; H_e – expected heterozygosity and *I* – Shannon's diversity index; M_A – number of morphotype-specific allele/private allele; *F* – fixation index

diversity was found among wild-like weedy groups (PM-BRA, PM-BH, PM-BHA) (H_e = 0.294, 0.248, 0.276 respectively) compared with the crop-like weedy rice (PM-SH, PM-BR) (H_e = 0.117, 0.112 respectively). In general, wild-like weedy rice shows higher phenotypic and genetic variation compared to the crop-like strawhull weedy groups, hence it is not surprising that the awned or black hulled weedy populations are more genetically diverse than the non-awned strawhull types (Shivrain et al., 2010). However, In Sabah weedy rice, those with grain morphology most closely resembling cultivated rice (Strawhull, SB-SH) exhibited the highest level of genetic diversity (H_e = 0.265, I = 0.425), a similar observation made by Neik et al. (2019) on Sabah weedy rice using SSR makers.

Wild relatives are expected to have higher gene pool variability due to lack of human selection that will reduce gene diversity (Shivrain et al., 2009). The wild species in this study showed higher diversity (H_e = 0.22) compared with weedy and cultivars. However, the seven Malaysian *O. rufipogon* accession showcasing higher diversity (H_e = 0.318) than the eleven *O. rufipogon* collected from Vietnam-Thailand-Cambodia-Indonesia region (H_e = 0.203). The peninsular elite cultivated rice showed low heterozygosity H_e = 0.041), which is consistent with their narrow genetic background due to domestication bottleneck. In contrast, Sabah local cultivars exhibited considerably high level of genetic diversity than that of peninsular elite cultivars reflecting the diverse genetic background that characterizes Sabah cultivars (Ministry of Agriculture Malaysia 1996), as compared to the more homogenous elite 'MR' series cultivars planted in the peninsular mainland (Song et al. 2014). Low observed heterozygosity (H_o) in Sabah weedy rice (H_o = 0.107) suggesting high selfing rate for weeds which is consistent with previously documented observation in other region (Cao et al., 2006; Song et al., 2014). Furthermore, the mean F, fixation index, was 0.418, confirming the predominance of inbreeding in Sabah weedy rice. Some populations, such as PM-BHA and SB-CV had a very low F-value (F = 0.00), indicating the occurrence of cross-fertilization. This corroborates with the high value of observed heterozygosity found in these populations (PM-BHA: H_o = 0.292, H_e = 0.276 ; SB-CV: H_o = 0.264, H_e = 0.249). The overall results showed that Sabah weedy populations have relatively high genetic diversity, although there is considerable variation among

them. In fact, in natural populations of autogamous species, a low genetic diversity is expected due to the high degree of homozygosity. In the case weedy rice, however, the diversity within populations may be higher (Xia et al., 2011, dos Reis Goulart et al., 2012). This is further confirmed through the partitioning of genetic variability in weedy rice populations, which was mainly related to the diversity within each population (AMOVA within population; 24%), as compared with the differences between populations (AMOVA among population; 20%) (Appendix 5). The level of genetic diversity detected in the Sabah weedy rice populations (H_e = 0.22) was lower than that of Sabah weedy rice population analysed with SSR markers (H_e = 0.456; Neik et al. 2019). Similarly, the overall genetic diversity of weedy rice in this study, including both Sabah and peninsular (H_e = 0.219), was lower than Malaysia weedy rice (H_e = 0.368; SSR data of Song et al., 2014).

To further evaluate the genetic diversity, genome wide nucleotide diversity was surveyed for all *Oryza* population (Figure 4.1). The overall mean nucleotide diversity for all accessions was 2.3×10^{-4}. Nucleotide diversity is highest in Peninsular weedy rice samples as well as Malaysian cultivated rice (4.8×10^{-4} and 4.7×10^{-4} respectively) and lowest in Sabah weedy rice population (2.6×10^{-4}). Weedy rice from VTCI and wild species were having a similar range of nucleotide diversity (VTCI weedy = 4.1×10^{-4}, wild = 3.9×10^{-4}).

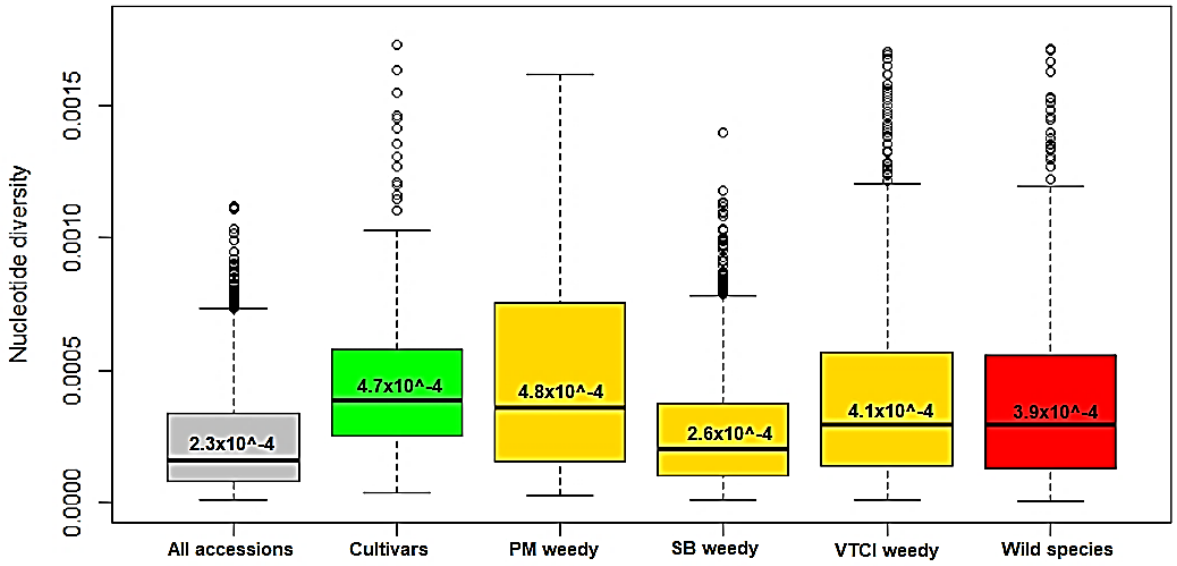


Figure 4.1. Box plots of nucleotide diversity from sliding windows across the rice genome for each *Oryza* group. Boxes represent 25th to 75th percentile and mean values are indicated inside each box. Groups include cultivated rice varieties of Sabah and Peninsular (green box) and weedy rice populations (yellow boxes) and wild rice populations (red box). PM: peninsular, SB: Sabah, VTCI: Vietnam-Thailand-Cambodia-Indonesia.

4.1.2 Population structure analysis

In order to gain insight about the origin of Sabah weedy rice, a Bayesian clustering based on MCMC method was used to estimate the admixture coefficient value (Q). The maximized marginal likelihood values of STRUCTURE outputs indicate K = 7 as the optimal grouping (Figure 4.2), where K indicates individual ancestries or populations discernable among the samples analysed. Results from K = 4 and K = 6 are also included for comparison.

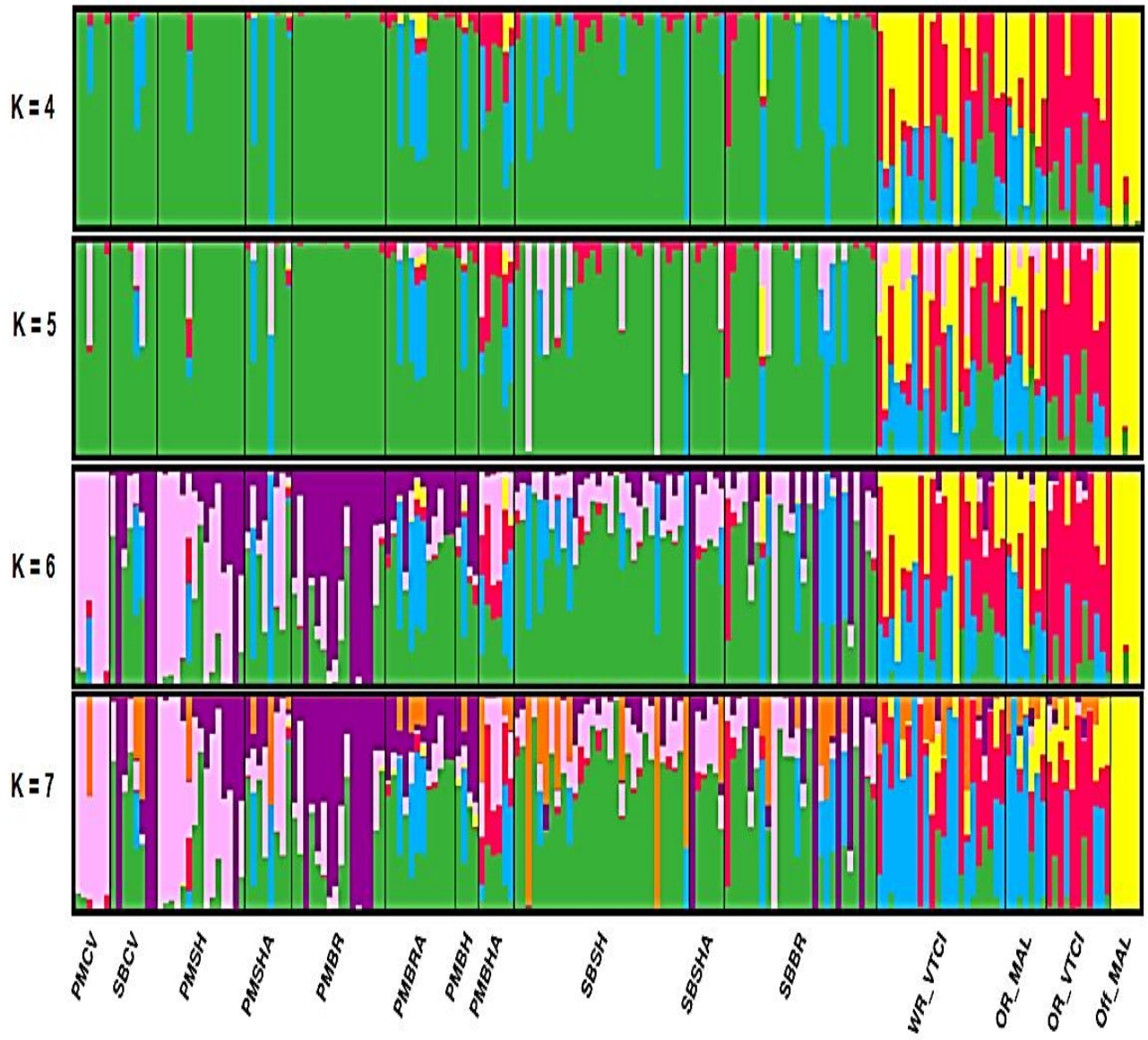


Figure 4.2. Population structure of 182 *Oryza* accessions based on 5379 SNPs. STRUCTURE results are shown from K = 5 - 7. The maximized marginal likelihood indicates K = 7 as the best fit grouping. Abbreviations of rice groups: PM, Peninsular; SB, Sabah; MAL, Malaysia; PMCV, peninsular elite cultivars; SBCV, high yielding Sabah cultivars; PMSH, strawhull awnless PM weeds; PMSHA, strawhull awned PM weeds; PMBR, brown-striped hull, awnless PM weeds; PMBRA, brown-striped hull, awned PM weeds; PMBH, blackhull awnless PM weeds; PMBHA, blackhull awned PM weeds; SBSH, Sabah strawhull awnless weeds; SBSHA, Sabah strawhull awned weeds; SBBR, Sabah brown-striped hull, awnless weeds; VTCI, Vietnam,Thailand,Cambodia,Indonesia WR, weedy rice; OR, *O. rufipogon*; Off*, O. officinalis*.

The majority of weedy rice individuals exhibited multiple cluster assignments, implying an admixed genetic constitution. At K = 4, wild relative *O. officinalis* (Off_MAL) was differentiated from other rice varieties which represented by yellow segment and with increasing K, the *O. officinalis* individuals remained in single cluster. At K = 4 and 5, the Malaysian weedy (both Peninsular and Sabah) rice samples are genetically distinct from the other Southeast Asian (VTCI) weeds. Majority of Malaysian weedy and cultivars assigned to a single cluster (green segment) while the VTCI weedy and wild rice individuals assigned to multiple clusters indicating higher background genetic diversity or admixture between clusters. This pattern is congruous with earlier observation by Viguiera et al. (2019) using FASTSTRUCTURE where the Malaysian weedy strains primarily with *indica* crop varieties. Hence the green segment in the STRUCTURE plot likely to reflect the *indica* crop genotype since all the cultivated and weedy accessions used in this study were *indica* varieties.

With increasing K value, more subgrouping was visible among Malaysian weedy rice. At K = 7 the Peninsular cultivars (PMCV) and Peninsular brown-hulled weedy types (PMBR) emerged into distinct clusters, represented by light-pink and purple clusters respectively (Figure 4.2). The distinct cluster of PMBR (purple) hypotheses that this weedy rice might have independently evolved from de-domestication of peninsular cultivar (light-pink) through standing variation. Both peninsular strawhull (PMSH) and strawhulled awned (PMSHA) groups contained subset of diversity found in the PMCV (light-pink) and PMBR (purple) groups in structure analysis suggests these two weedy rice populations were potentially be the results of hybridization between peninsular cultivar and BR weedy group. Similarly, peninsular weedy rice population in brown hull awned (PMBRA), black hull (PMBH) and black hull awned (PMBHA) were identified as admixture of peninsular cultivars (PMCV), brown-hulled (PMBR), South Asian (VTCI) weedy and wild species. The genetic gradation of weedy rice correlated with grain morphology particularly hull colour and awn type, from predominantly crop-like to predominantly wild like, as mentioned in Song et al. (2014). Though dissimilar gradation pattern in comparison with Song et al. (2014) and Vigueira et al. (2019) structure analyses on peninsular Malaysian weedy rice, a general trend is apparent where the lighter hulled weeds (crop-like, SH and SHA) clustered with peninsular elite *indica* rice, while darker hulled weedy rice (wild-like,

BH and BHA) has the pattern of genetic admixture resembling wild species of *O. rufipogon.* Together, these findings support a prominent role of Malaysian elite *indica* cultivated rice and wild rice species in the origin of peninsular weedy rice.

The Sabah cultivars used in this study showed admixture pattern of four genotypes; SBSH like genotype (green segment), PMBr like genotype (purple), PMCV like (light-pink) and VTCI weedy like genotype (blue). This genetic containment is consistent with our genetic diversity analysis where Sabah cultivars exhibited high diversity (provide some values of the genetic diversity parameters) and polymorphic percentage of around 64% (Table 4.1), explaining the broader genetic composition of Sabah cultivars than homogenous elite cultivars of peninsular Malaysia (pink cluster). Similarly, the Sabah weedy populations were also classified as admix of the abovementioned genotypes with predominantly showing SBSH-like genotype background. Additionally, the substantial contribution of PMBR genotype, ~60% membership coefficient (Figure 4.2), found in SBBR population gave rise to the question whether the Sabah weedy rice could be descended directly from Peninsular weedy rice. Generally, most of the Sabah weedy rice populations have close genetic resemblance to Sabah cultivars despite their phenotypic difference which support the notion that gene flow and genetic introgression have occurred between weedy rice and cultivars in Sabah for many years and acting as ongoing evolutionary mechanism that shaped the population structure of Sabah weedy rice.

Approximately five clusters were observed within the wild progenitor, *O. rufipogon,* all individuals appeared to be admixtures, particularly sharing similar genotype with VTCI weedy rice. Vigueira et al. (2019) described the VTCI weedy rice shared high genetic similarity with wild *Oryza* samples and also contained genetic composition of *indica*- and *aus* cultivars which explained the high heterogeneity of VTCI weedy population in this study.

Taken this, the smallest proportion of VTCI weedy-like genetic background (blue segments) found in Sabah weedy rice population more likely to reflect the genetic component of *indica*, *tropical japonica* or *aus* landraces rather than wild rice genotypes considering the fact that Sabah rice are growing outside sympatry of wild rice.

4.1.3 Genetic relationship among weedy rice

The first three axes of the PCoA account for 27.52%, 7.72% and 4.94% of the genetic variation present (Figure 4.3). The scatter plot based showed a clear genetic distribution pattern among weedy rice, cultivated rice and wild species. The weedy rice individuals were scattered over a very large area in the plot, suggesting their high level of genetic diversity, and weedy rice individuals from each region clustered together in the first coordinate (x-axis). The geographical barriers can be expected to have constrained gene flow between the different rice groups, and hence are likely to have contributed to their genetic differentiation. The weedy rice from peninsular were clustered in one cohesive group along with peninsular elite cultivar (Cluster 1) while the weedy rice individuals from Sabah were clustered into fragmented groups with samples dispersed across the positive and negative loads of the first principal component. The majority of Sabah weedy rice grouped closely with co-occurring local cultivars (Cluster 2) whereas some weedy rice that share higher percentage of peninsular cultivar and weedy genetic background fell into Cluster 1. Some Sabah and peninsular strawhulled, brownhulled and brownhull awned weedy rice which appeared to be admixture of peninsular and Sabah weeds are arrayed between Cluster 1 and 2. On the contrary, all South Asian (VTCI) weedy rice grouped together with common wild, *O. rufipogon* which corroborated by structure analysis (Cluster 3). Noticeably, the five *O. officinalis* samples were clearly differentiated into single cluster. Meanwhile, the second coordinate (y-axis) showed the gradation across hull morphotypes from lighter hulled to darker hulled weed. As observed in the structure analysis, the strawhull (SH) population occupied space closest to the cultivated rice whereas the blackhull (BH) and blackhull-awned (BHA) occupied space close to *O.rufipogon* as well as cultivated rice.

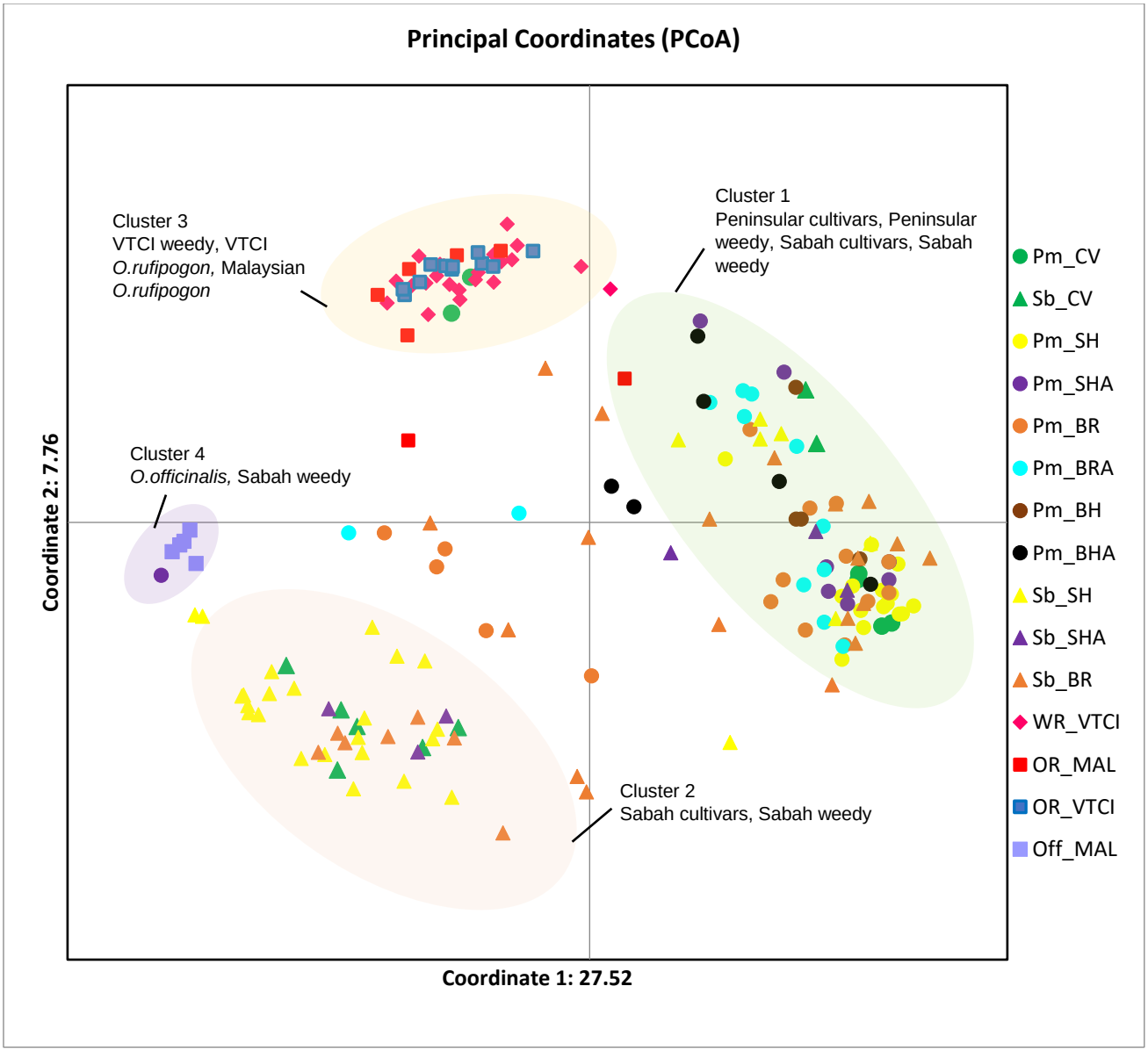


Figure 4.3. Principle Coordinate Analysis (PCoA) plot of weedy and cultivated rice with other *Oryza* species. Symbols represent different regions and colors represent different hull morphotypes. Abbreviations of rice groups: Pm, Peninsular; Sb, Sabah; MAL, Malaysia; Pm_CV, peninsular elite cultivars; Sb_CV, high yielding Sabah cultivars; Pm_SH, strawhull awnless Pm weeds; Pm_SHA, strawhull awned Pm weeds; Pm_BR, brown-striped hull, awnless Pm weeds; Pm_BRA, brown-striped hull, awned Pm weeds; Pm_BH, blackhull awnless Pm weeds; Pm_BHA, blackhull awned Pm weeds; Sb_SH, Sabah strawhull awnless weeds; Sb_SHA, Sabah strawhull awned weeds; Sb_BR, Sabah brown-striped hull, awnless weeds; VTCI, Vietnam,Thailand,Cambodia,Indonesia WR, weedy rice; OR, *O. rufipogon*; Off, *O. officinalis*

A maximum likelihood tree was constructed as a supplementary to PCoA analyses. (Appendix 6). In general, the clustering topology of the coordinate analysis is preserved although occasional transitions can be observed. The dendrogram showed that all of the samples were clustered into two main groups. The first group contained the weedy rice and cultivated rice population from Malaysia. The Malaysian weedy accessions formed a monophyletic group, with one containing peninsular weedy rice and the other containing Sabah weedy rice. As expected, the South Asian (VTCI) weeds in second group were phylogenetically distinct from Malaysian weeds, clustering tightly with *O. rufipogon* individuals. Coinciding PCoA observation, the weedy rice populations from same regions showed a closer genetic relationship than those from different regions suggesting that the weedy rice populations followed the isolation by-distance model. A few sub-groups and varieties of Malaysian weedy rice specifically those with admixed genotypes, however, did not cluster with respective to their geographic origin. Some of the Sabah cultivars and weedy rice form clades with peninsular and VTCI weeds. Similarly, few peninsular weedy rice lines found to lie within South Asian weedy clades. Although the Malaysian weedy rice samples were found to be dispersed throughout the tree, the phenotypic gradation were still evident. The weedy rice accessions with lighter hulled (crop-like) morphology, (SH and SHA) are nested in the clades that includes cultivars which suggesting the de-domestication origin for these weeds. However, some admixed weedy accessions with high proportion of wild like genotype (BH and BHA) appeared together in wild clade, suggesting origins from wild ancestors or hydridization with wild rice that gave rice to weedy rice. The wild relative, *O. officinalis* formed a relatively independent clade from the rest of rice individuals at the tree basal with a large distance from Malaysian weedy rice, showing the limited contribution to the evolution of weedy rice in Malaysia.

4.2 Candidate gene analysis

4.2.1 Phenotyping tiller angle and agronomic traits

The maximum angle of tillering was scored for a subset of 52 accessions representing multiple population groups of weedy, wild and cultivated *Oryza* (Table 4.2.). The measurements revealed variation among the *Oryza* groups skewing towards narrow tiller angles (< 40º). A larger number of weedy rice accessions, around 64% (Table 4.2), fell within smaller range of tiller angles, hence showing that Malaysian *indica* variety weedy rice groups have relatively compact architecture which are similar to cultivated rice. A similar distribution was observed by Huang et al. (2018), where the South Asian weedy rice subgroup have relatively compact plant architecture with smaller tiller angles. However, there is a significant difference in tiller angle between Sabah and Peninsular weedy rice. In Sabah, nearly half of the samples exhibited wider range of tiller angles compared with its Peninsular counterparts. Dong et al. (2016) postulates that the tiller angles are mainly controlled by genetic factors, and to some extent, the interactions with environment also play a role. Different environmental conditions and climate patterns of West (Peninsular) and East (Sabah, Borneo) Malaysia, for example different irrigation pattern and soil structure (peninsular cultivation is highly irrigated while Sabah is more towards dryland cultivation (Herman et al., 2015), might govern the variation in tiller angles. The Malaysian wild rice (*O. rufipogon*), as theorized has the most wide spread tillers, with majority (75%) having angles between 40-90 ˚. However, the other accession of *O. rufipogon*, (IRGC105491) believed to be a hybrid of cultivated and wild rice (McCouch et al., 2007, Song BK whole genome sequence data., pers. comm) which explained the compactness of tiller angle observed (<40 ˚) (Table 4.2)

Table 4.2 Phenotypic traits of Malaysian *Oryza* groups

Group	N	Hull					Tiller angle	
		SH	SHA	BR	BRA	BHA	< 40°	40°-90°
Weedy rice (44)								
Sabah	23	8	2	11	2	0	13 (56%)	10 (44%)
Peninsular	21	6	2	5	5	3	15 (71%)	6 (29%)
Mean							28 (64%)	16 (36%)
***indica* cultivars (4)**								
Sabah	2	2	0	0	0	0	2 (100%)	0 (0%)
Peninsular	2	2	0	0	0	0	2 (100%)	0 (0%)
Mean							4 (100%)	0 (0%)
Wild rice *O. rufipogon* (4)	4	0	0	0	0	4	1 (25%)	3 (75%)
Total mean	52						33 (63%)	19 (37%)

Abbreviations of rice groups: SH, strawhull awnless weeds; SHA, strawhull awned weeds; BR, brown-striped hull, awnless weeds; BRA, brown-striped hull, awned weeds; BHA, blackhull awned weeds.

4.2.2 Genealogy of *TAC1*

A 3.0 kb region was sequenced for *TAC1* encompassing 1277 bp of the genic region and 1691 bp of the 3' flanking region containing *TAC1* functional nucleotide polymorphism (FNP) at splicing site (coordinates 20731589 – 20734728 on chromosome 9 according to MSU 6.0 Nipponbare). 52 individuals were genotyped at this locus, including SH, SHA, BR, BRA, BHA weedy rice (43 accessions, 23 Sabah weedy and 21 peninsular weedy), cultivated *indica* rice (2 Peninsular elite and 2 Sabah traditional cultivars) and wild (4 accessions of *O. rufipogon*) (Table 4.3). Since *tropical japonica* group was absent in our sample panel, we further added 5 representatives of Malaysian *tropical japonica* landraces with compact tiller architecture and the sequence data was obtained from 3000 rice genome project, totalling up our sample panel to 57 accessions. Sequence variation

including single nucleotide polymorphisms (SNPs) and insertions/deletion (indels) occurred with different nucleotide frequency in different regions. Occurrence of indels (31) were more prevalent than SNPs (17). three polymorphic sites were identified and four indels in weedy rice groups in the gene region (from start to stop codon) (Table 4.3). Four indels with no polymorphic sites were detected in the *indica* cultivar group while no variation was detected in the genic region within the *japonica* cultivar population. One polymorphic singleton site was observed in the genic region of wild rice along with four indels. Analysis of the 3' flanking untranslated region (UTR) containing the *TAC1* FNP site of weedy rice showed nine polymorphic sites and 27 indels, with 3 of the nine polymorphic sites being parsimony sites. The cultivar group had three singleton variable sites and seven parsimony sites with 27 indels. All *japonica* varieties had the *TAC1* genotype, GGGA (compact plant architecture) at the 3' splicing site in the fourth intron whereas no *TAC1* allele was found in any weedy (except one peninsular weedy rice, MU026), wild and *indica* cultivated rice. Five polymorphic sites all of which were singleton variable sites were identified in the 3' flanking region of the wild rice. This study showed that nucleotide variation in 3' flanking sites occurred more frequently than in the gene region in all sampled accessions. Jiang et al. (2012a) had a similar observation in cultivated and wild rice with more variations found in both flanking (3' and 5') region. Furthermore, this study also showed weedy rice groups harboured more nucleotide polymorphisms than cultivated and wild rice.

Variation in the coding region, CDS, often leads to functional changes in a gene. All the accessions analysed had 780bp *TAC1* coding sequence encoding 259 amino acids. A comparison of the coding region from these 57 varieties of weedy, cultivated and wild accession with *OsTAC1* CDS of *Japonica Nipponbare* as reference revealed two SNPs, one synonymous and one nonsynonymous substitution, in the third exonic region. The synonymous substitution, G to A, at position +684 was observed in the *tropical japonica* group but the nonsynonymous substitution G to T, at position +486, was only detected in wild group. This mutation resulted in the change of amino acid from arginine (CGA) to leucine (CTA). The nonsynonymous substitution in the wild population was only detected in three *O. rufipogon* accessions (PLWr02, PLWr08 and PLWr16) from Kepala Batas, Penang, Malaysia, exhibiting prostrate growth habit. The intron region harboured more

SNPs than the exons (Table 4.4), with 11 SNPs spotted in introns of weedy and cultivated rice varieties and five SNPs detected in the intron of wild species. These results implied that all the analysed *Oryza* population (with exception of three abovementioned wild rice species) in this study had identical *TAC1* amino acid sequences hence suggesting the genetic makeup of *TAC1* is highly conserved between cultivated, weedy and its wild progenitor (Jiang et al., 2012a).

Table 4.3 Summary of DNA variation of the *TAC1* region of Malaysian *Oryza*

Region	*Species*	*N*	*L*	*I*	*PS*	*P*	*S*	*h*	π	*hd*	*K*	*Tajima's D*
GENE REGION (FROM START TO STOP CODON)	**All Weedy Rice**	44	1277	4	3	0	3	2	0.00011	0.045	0.136	-1.708 ^{NS}
	Sb weedy	23	1277	4	0	0	0	1	0	0	0	n
	Pm weedy	21	1277	4	3	0	3	2	0.00022	0.095	0.286	-1.727 ^{NS}
	All Cultivars	9	1277	4	3	3	0	2	0.00131	0.556	1.667	1.948 ^{NS}
	indica	4	1277	4	0	0	0	1	0	0	0	n
	trp.japonica	5	1277	0	0	0	0	1	0	0	0	n
	Wild rice	4	1277	4	1	0	1	2	0.00039	0.500	0.500	-0.612 ^{NS}
3' FLANKING UTR	**All Weedy Rice**	44	1673	27	9	3	6	3	0.00107	0.532	1.798	-0.375 ^{NS}
	Sb weedy	23	1673	27	3	3	0	2	0.00089	0.498	0.813	2.136*
	Pm weedy	21	1673	27	9	3	6	3	0.00128	0.567	2.502	-0.499 ^{NS}
	All Cultivars	9	1673	27	9	7	2	3	0.00256	0.639	3.311	1.343 ^{NS}
	indica	4	1673	27	3	0	3	2	0.00090	0.500	1.500	-0.754 ^{NS}
	trp.japonica	5	1673	6	0	0	0	1	0	0	0	n
	Wild rice	4	1673	7	5	0	5	2	0.00148	0.500	2.500	-0.797 ^{NS}
FULL LENGTH *TAC1*	**All Weedy Rice**	44	2944	31	12	3	9	3	0.00066	0.532	1.932	-0.904 ^{NS}
	Sb weedy	23	2944	31	3	3	0	2	0.00051	0.498	1.494	2.136*
	Pm weedy	21	2944	31	12	3	9	3	0.00082	0.567	2.419	-0.973 ^{NS}
	All Cultivar	9	2944	31	12	10	2	3	0.00202	0.639	5.944	1.640 ^{NS}
	indica	4	2944	31	3	0	3	2	0.00051	0.500	1.500	-0.754 ^{NS}
	trp.japonica	5	2969	6	0	0	0	1	0	0	0	n
	Wild rice	4	2964	11	6	0	6	2	0.00101	0.500	3.000	-0.808 ^{NS}

Abbreviations: *N:* number of samples; *L*: length of the alignment in which all sequence contain bases, gaps excluded; I: indels; PS : polymorphic sites; *P*: parsimony informative sites; *S*: singleton sites; *h*: number of haplotypes; π: nucleotide diversity; *hd*: haplotype diversity; *K*: average number of nucleotide difference; * : $P < 0.05$; NS: not significant; n: no tajima's D values, polymorphism not detected; Sb: Sabah; Pm: Peninsular; trp: tropical

Table 4.4. Summary of DNA variation in *TAC1* coding region and introns

REGIONS	SPECIES	*N*	*L*	*PS*	π	*h*	*SS*	*NS*
CDS	All Weedy Rice	44	780	1	0.0006	2		
	Sb Weedy	23	780	0	0	1		
	Pm Weedy	21	780	1	0.00012	2		
	All Cultivars	9	780	1	0.00072	2		
	indica	4	780	0	0	1		
	trp.japonica	5	780	0	0	1	$G_{684} \rightarrow A$	
	Wild rice	4	780	1	0.00065	2		$G_{486} \rightarrow T$
INTRONS	All Weedy Rice	44	2170	11	0.00087	3		
	Sb Weedy	23	2170	3	0.00069	2		
	Pm Weedy	21	2170	11	0.00107	3		
	All Cultivars	9	2170	11	0.00248	3		
	indica	4	2170	3	0.00069	2		
	trp japonica	5	2170	0	0	1		
	Wild rice	4	2170	5	0.00115	2		

Abbrevations: CDS: coding region; *N:* number of samples; *L*: length of sequence (gaps included if polymorphism exists); *PS*: polymorphic sites; *h*: number of haplotypes; π: nucleotide diversity; SS: synonymous substitution; NS: nonsynonymous substitution

4.2.3 Nucleotide diversity of *TAC1*

Nucleotide diversity (π) of the targeted regions of the *TAC1* locus for different subgroups of Sabah and Peninsular weedy rice, *indica* and *japonica* cultivars and wild rice was computed (Table 4.3). Analysis of the gene region (from start to stop codon) indicated very low diversity for weedy rice group with a value of 0.11×10^{-3} (Table 4.3). Looking into subgroups of weedy rice, the Peninsular weedy rice showed 0.22×10^{-3} diversity value, but no nucleotide diversity was detected in weedy rice collected from Sabah. However, this elevated diversity is attributable to a single weed accession, MU026 which carries gene sequence that differs from the other peninsular weedy rice by three singleton variable sites. If this diverged accession is removed, the diversity estimates will be

reduced to zero for Peninsular weedy rice varieties which is consistent with a nearly fixed *TAC1* CDS in both cultivar groups (refer Table 4.4). Nucleotide diversity was not observed in gene region within *indica* and *japonica* respectively owing to the failure to identify any polymorphic sites. Nucleotide diversity was not observed in gene region within *indica* and *japonica* respectively due to absence of polymorphic site. Nevertheless, the variation becomes more pronounced when both cultivar types are grouped together where nucleotide variation is (1.31×10^{-3}). The increase in diversity could be the inherent differences in the genome between *indica* and *japonica*. The wild rice, *O. rufipogon*, also showed a reduced nucleotide diversity, 0.39×10^{-3} with a nonsynonymous mutation at the third exon (Table 4.3). The 3' flanking region recorded exceptionally high nucleotide diversity due to the presence of highly divergent haplogroups (Table 4.3). Cultivars exhibited greater nucleotide diversity value of 2.56×10^{-3} across the 3' flanking region, followed by wild group and weedy rice with values at 1.48×10^{-3} and 1.07×10^{-3} respectively. Unlike the flanking regions, analysis of the entire *TAC1* locus indicated reduced nucleotide diversity in Sabah weedy rice (0.51×10^{-3}), Peninsular weedy rice (0.89×10^{-3}) and *indica* cultivar (0.51×10^{-3}) and moderately low respective pi value of 1.01×10^{-3} for their wild progenitor.

The genetic diversity of the *TAC1* was further investigated in the intronic region of all samples. As expected the coding region indicated very low diversity across the groups (Table 4.4). Significantly lower nucleotide diversity was seen in the weedy rice population (0.06×10^{-3}) compared to the cultivated and wild rice groups. Conversely, the nucleotide diversity in the intronic region was an order of magnitude higher than the CDS of *TAC1* gene for all groups.

The distribution *TAC1* gene variation was well represented in the sliding window analysis of π at every 100bp (Figure 4.4). High level of nucleotide polymorphism of Malaysian *Oryza* was apparent across the entire 3' flanking region. The diversity peaks were present mostly in the intronic region of the gene with the exception of the third exon which carries both synonymous and nonsynonymous mutation in the cultivars and wild rice accessions, respectively. A plateau was noticed in all exonic regions (except exon 3 of cultivars and wild), explaining fewer to no nucleotide polymorphism were observed. These findings

reaffirmed the high degree of conservation of *TAC1* CDS in weedy, wild and cultivated rice populations.

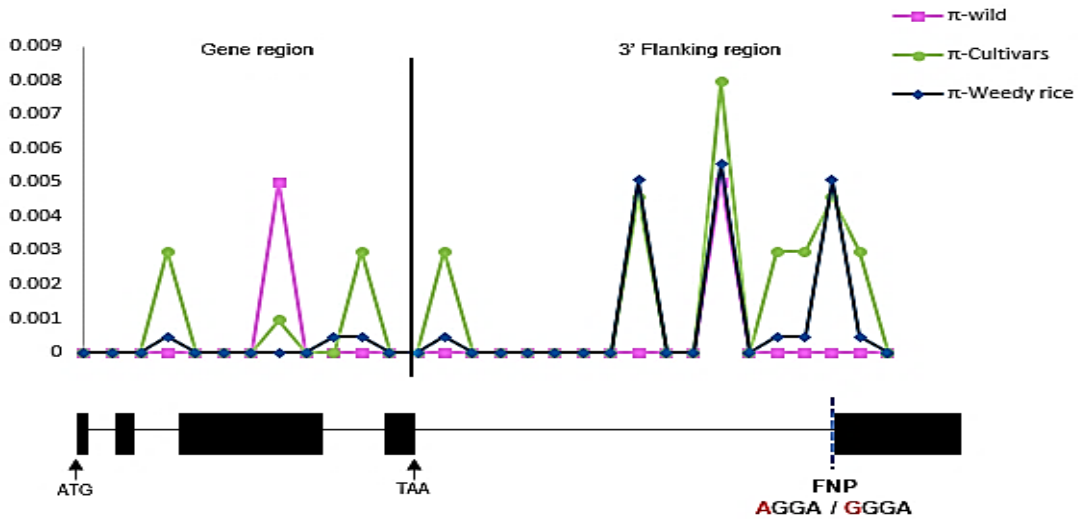


Figure 4.4. Sliding window analysis of nucleotide diversity for *TAC1* gene. Window length : 100bp; black box: exon; thin line; intron; ATG : start codon; TAA : stop codon.

4.2.4 Neutrality test

Recent studies on the evolutionary study of rice have shown that compact plant architecture with favorable tiller angles has been selected in the long history of domestication. To test this hypothesis, neutrality test on the *TAC1* gene has been performed with Tajima's D parameter in each group (Table 4.3). Statistically significant positive Tajima's D values (*D* = 2.136) in the 3' flanking and entire region of *TAC1* were detected in Sabah weedy rice, suggesting that potential recent balancing selection might be acting on these genomic regions of different subpopulation of Sabah weedy rice during de-domestication. Since the increased polymorphism in Peninsular weedy rice was ascribed to the MU026, the Tajima's D parameter in peninsular weedy rice as well as in weedy rice pooled populations excluding MU026 has been re-calculated. Surprisingly, the

Tajima's D become significantly positive at both 3' flanking region and whole length of *TAC1* (peninsular : 2.266; whole weedy population: 2.578, *P*<0.05) (refer Appendix 9). Taken together, it is believed that Malaysian weedy rice experienced strong balancing selection since their de-domestication from cultivated rice populations. Consistent with Jiang et al. (2012), the four *indica* cultivars and wild rice showed no significant deviation from neutrality at *TAC1*. However, the signature of selection could not be calculated in *japonica* group due to lack of polymorphism within them; this is in contrast to the expectations of selection on this gene in all Asian *japonica* varities from Jiang at el. (2012) where the *TAC1* gene in the *japonica* group was strongly selected (*D* = -1.8050, *P* < 0.05) during rice domestication. We presume the Malaysian *tropical japonica* crop varietes have likely experienced a very recent bottleneck event or a selective sweep at this gene, resulting in the fixation of a single haplotype.

4.2.5 Haplodiversity of *TAC1*

To understand the genetic mechanisms underlying differences in the rice tiller architecture among weed groups and their crop and wild progenitors, haplotype analysis of the *TAC1* gene were performed. In particular, it is of paramount importance to know whether Malaysian weedy rice shared similar *TAC1* alleles as their putative cultivated ancesters and whether Sabah weedy rice exhibit the same *TAC1* gene pattern as their peninsular counterparts. The analysed 57 Malaysian *Oryza* accessions were characterised by five haplotypes (H1 – H5) based on 13 mutations (Figure 4.5a). In the cultivated *indica* rice populations, two main haplotypes were identified, H1 and H2, where they are separated by four substitutions in the fourth intron. Sabah cultivars were classified into H1 and H2, each accounted for 50% while the peninsular cultivars were only shared by a single haplotype, H1 (Figure 4.5b).

a)

H	330	730	945	1032	1358	2029	2331	2336	2602	2635	2717	2760	2831	*N*	Tiller angle	
															<40	40-90
H1	C	G	G	A	G	T	T	G	G	T	-	G	A	27	74%	26%
H2	C	G	G	A	G	C	C	A	G	T	-	T	A	21	57%	43%
H3	T	G	A	T	A	T	T	A	A	C	T	G	G	5	100%	0%
H4	T	G	A	T	A	T	T	A	A	C	-	G	G	1	100%	0%
H5	C	T	G	A	G	T	T	G	G	T	-	G	A	3	0%	100%

b)

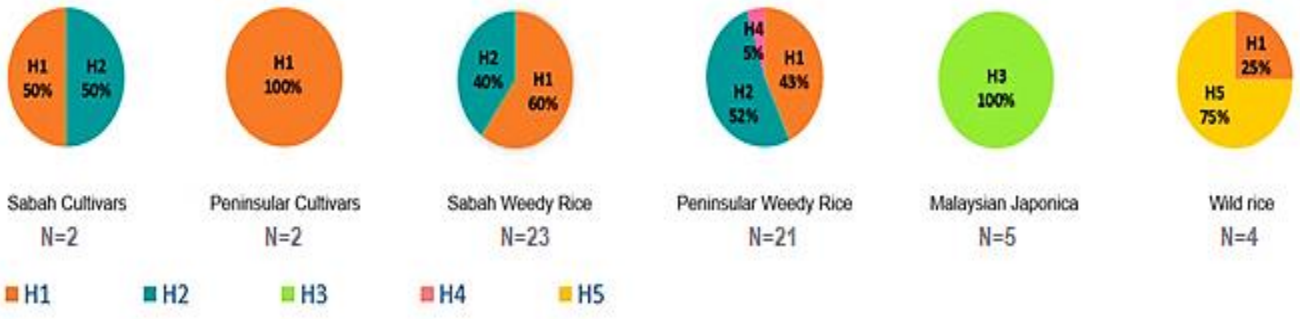


Figure 4.5. Haplotype diversity analysis of *TAC1*. a) Haplotype sequence analysis, green shaded box: NS substitution, orange shaded box: SS substitution. Red letters: FNP. b) Haplotype frequencies shown in percentage. H: haplotypes, N: number of samples.

As expected, all Sabah weedy haplotypes were identical to those detected in their cultivated rice counterpart, since cultivar type alleles were found in weedy rice accessions. About 60% of the total Sabah weedy rice accessions carried haplotype 1 while the remaining 40% were defined as haplotype 2. However not all weedy rice shared the similar haplotypes with their cultivars. While the peninsular cultivars were fixed with the H1 haplotype, three distinct haplotypes, two major haplotypes H1 (43%) and H2 (52%), and a minor haplotype, H4 (5%) were detected in the peninsular weedy rice populations. The minor haplotype 4 in peninsular weedy rice with single accession, MU026, was derived from H3 haplotype by one mutation at SNP position 2717bp. As observed, the haplotypes 1 and 2 appeared to be highly diverged alleles for *TAC1* within weed groups. This configuration of polymorphism, two divergent haplotypes at similar

frequency in our weedy rice together with the allele frequency spectrum that was strongly skewed towards intermediate frequency alleles (Tajima's *D* =*2*.578, P<0.05, all weedy rice excluding MU026 accession, refer appendix 9), is suggesting that both Sabah and peninsular weedy rice individuals had undergone long-term maintenance by balacing selection for the *TAC1* locus. Meanwhile, the H3 haplotype was composed exclusively of all the Malaysian *japonica* cultivars. This implied that the *TAC1* genomic region had lost diversity in *japonica* cultivars and might have experienced *japonica*-specific selective sweeps that led to single allele fixation. Of the two haplotypes found, about 75% (three out of four) wild rice accessions were unique to haplotype 5 while haplotype 1 was represented by a single *O. rufipogon* accession (MU201). Each of the haplotypes differing from one another by single non-synonymous substituition (Arg to Leu) at the third exon. Interestingly our data did not yield an obvious association between *TAC1* haplotypes and tiller angles (Table 4.2.1). Although previous studies reported that the *TAC1* mutation 'GGGA' (type G allele) was associated with compact plant tiller angle and the wild type *TAC1* 'AGGA' (type A allele) was associated with spread out tillers (Yu et al., 2007, Jiang et al., 2012a), all *indica* cultivated rice and weedy rice in our panel with tiller angles that are lesser than 40° carried the wild type 'A' allele at functional nucleotide polymorphism (FNP) site. In addition, a single wild accession, MU201 which carried 'A' allele had compact plant architecture. The mutational type 'G' allele was only found in *japonica* population, yet one *indica* weedy rice accession in haplotype 4 with narrow tiller angle also contained the similar allele. Thus, our results evince that rice tillering likely to have other genetic and environment determinants.

To resolve the evolutionary history of *TAC1* , the scope of haplotype analysis was broadened by comparing the Malaysian *TAC1* alleles with a panel of 37 *O. sativa* accessions (including *tropica japonica*, *indica* and *aus*), five wild rice varieties (*O. rufipogon*, *O. barthii*, *O. meridionalis*, *O. nivara*) and one cultivar *O. glaberrima* accession collected from 14 rice growing regions across the world. Eleven haplotypes were constructed and the network was dominated by three major haplogroups; Hap_1, Hap_2 and Hap_3 (Figure 4.6).

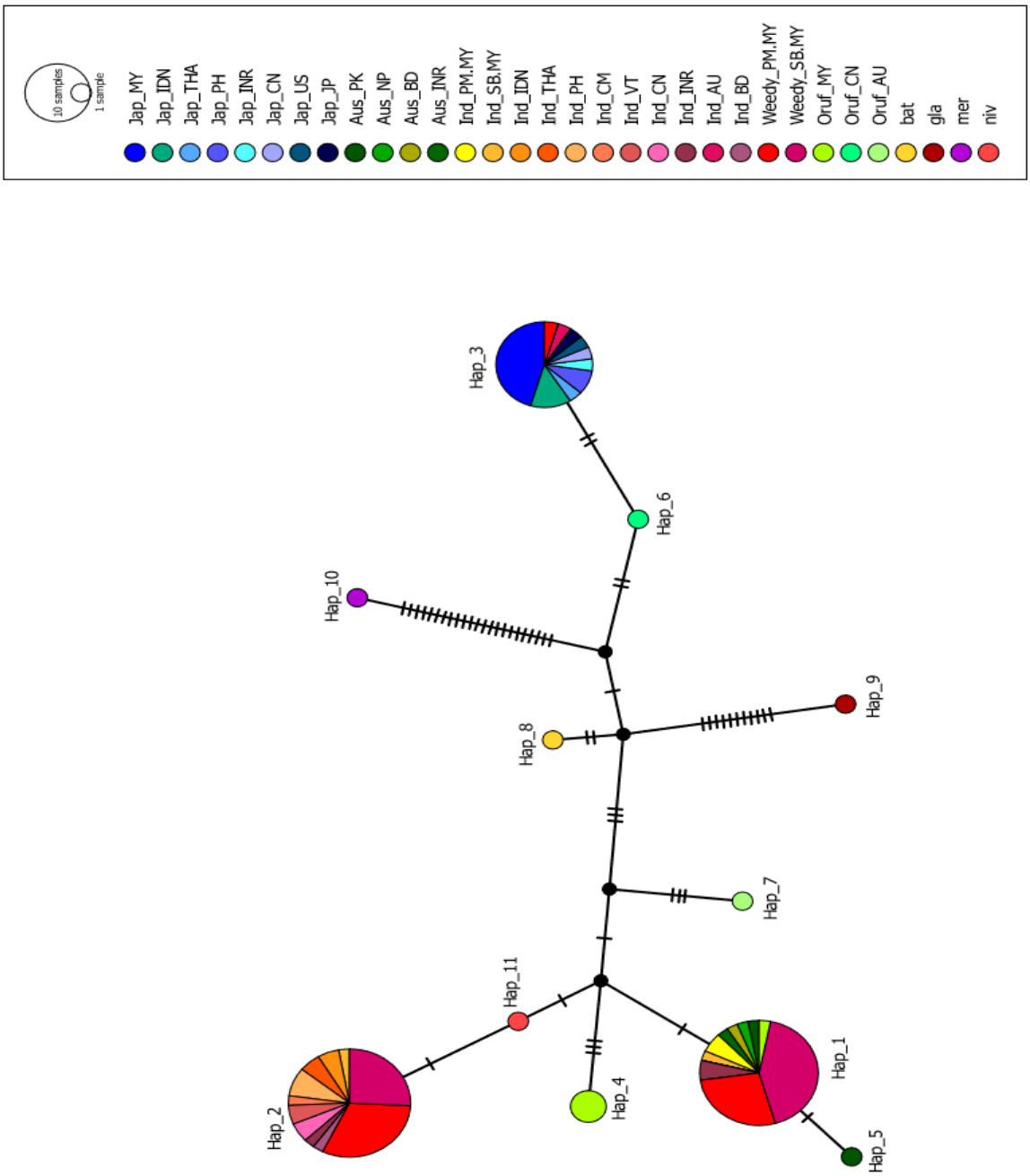


Figure 4.6. Haplotype TCS network of *TAC1* gene. Each nodes represents a haplotype, its size being proportional to its frequency. Abbreviations: Hap, haplotypes; Jap, *japonica*; Ind, *indica*, MY, Malaysia; IDN, Indonesia; THA, Thailand; PH, Philippines; INR, India; CN, China; JP, Japan; PK, Pakistan; NP, Nepal; BD, Bangladesh; CM, Cambodia; VT, Vietnam; AU, Australia; US, United States; PM, peninsular; SB, Sabah; Oruf, *O. rufipogon*; bat, *O. barthii*; gla, *O. glaberrima*; mer, *O. meridionalis*; niv, *O. nivara*.

The first cluster, Hap_1 was observed in *aus* (minor variety group related to *indica*) and *indica* cultivars and we referred this haplogroup as "*aus*-like" since recent research has revealed that *aus* is genetically complex and its origin is distinct (Kim et al., 2016, Civáň et al., 2015). The Hap_2 haplotype which occured primarily in *indica* cultivars grown within Southeast Asia was denoted as " *indica*-like". All *japonica* accessions across the globe formed an isolated cluster, Hap_3, thus we labelled it as "*japonica*-like" haplogroup. The haplotype classifications describe the geographical distribution of genotypes including "*aus*-like", "*indica*-like" or "*japonica*-like". All *japonica* or *indica* samples clustered together within its own clade respectively and the '*japonica*-like' clade was distantly located from the '*indica*-like' clade, showing clear segregation between them. Similar to what we observed in nucleotide diversity, all *japonica* lines in our panel originating from eight different countries underwent stronger selective sweep or severe genetic bottleneck to fixation of single allele for *TAC1* gene while *indica* individuals spread gradually across all major haplotypes, suggesting the two major rice varities may have arisen from different ancestral gene pools. Supporting this statement, we found that the "*indica*-like" individuals were closely related to *O. nivara* (Hap_11), an annual AA genome wild species commonly distributed in Southeast Asia (Vaughan et al., 2008) whereas the "japanica-like" accessions were close to *O. rufipogon* (Hap_6). Since *indica* and *aus* showed closer relationship in the network, it is possible that they shared common origin but *aus* likely has originated earlier or had evolved from a genetically distinct *O. rufipogon* or *O. nivara* ancesetral progenitor (Choi et al., 2017).

It is noteworthy that the wild rice used in this study were just a single representation of each wild rice species across the region and they might not truly represent the direct wild progenitor for domesticated population. Most Malaysian weedy individuals entered either *aus*-like or *indica*-like clades concordant with previously described balancing selection. Combining the haplotype and network analyses, convergence is observed among these two weedy rice groups (*aus*-like and *indica*-like) for plant architecture where cultivars (*aus* and *indica*) tend to have compact architecture yet some of the cultivars-derived weedy rice groups tend to have spreaded architecture.

We further mapped the geographical distribution of *TAC1* haplotypes (Figure 4.7). The "*aus*-like" genotype (Hap_1; shaded in blue) predominantly distributed in Indian subcontinent (from Pakistan to Bangladesh) which was identified as the source of *aus* gene pool. As mentioned by (Vigueira et al., 2019), unlike *indica*, *aus* cultivation is restricted to nothern Indian subcontinent and not being cultivated in Southeast Asia, hence the "*aus*-like" genotype in Malaysian weedy rice and cultivars could most probably reflect the introgression of *aus*-like wild rice rather than from cultivated *aus* ancestors. However, in Sabah where no wild *Oryza* occur, the "*aus*-like" genotype contribution in weedy rice could best be explained by the accidental introduction of *aus*-like weedy rice from peninsular Malaysia in contaminated rice grain stock. The "*indica*-like" genotype (Hap_2; shaded in orange) primarily concentrated in Indochina and Indian valley while "*japonica*-like" genotype (Hap_3; shaded in green) was found in high propotion in Southern Asia where upland rice cultivation is still prevalent. In conjuction with the three major *TAC1* alleles with consistenly distinct geographic origin, rice domestication in Malaysia was found to be a multiregional process independently producing the *japonica*, *indica* and *aus* rice varieties.

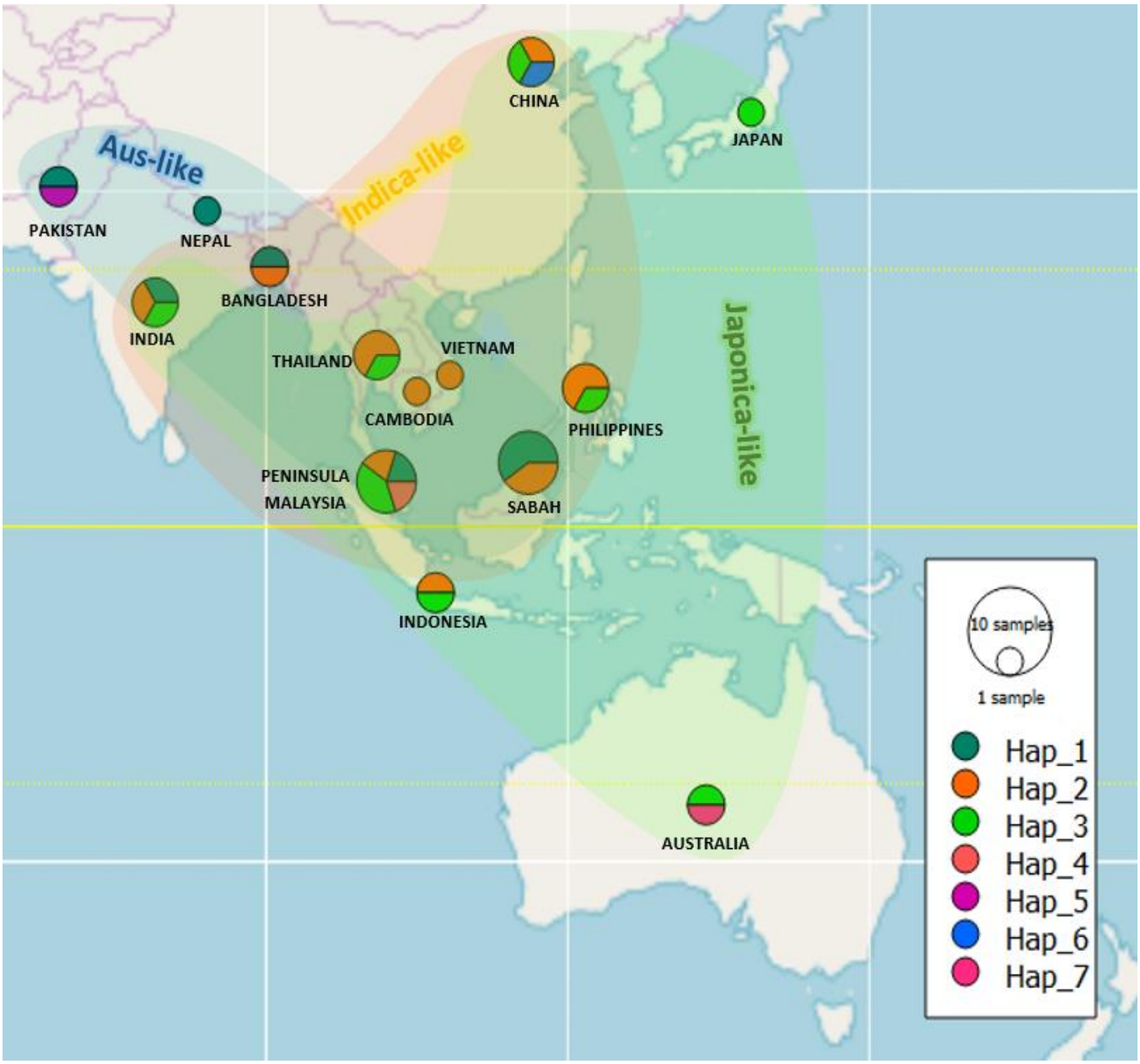


Figure 4.7. Geographical distribution of *TAC1* haplotypes. Blue shaded region: *aus*-like genotype, Orange shaded region: *indica*-like genotype, Green shaded region: *japonica*-like genotype

4.2.6 Phylogenetic pattern in *TAC1*

The neighbour-joining distance analysis of *TAC1* sequence revealed three major well supported and highly diverged clades; *indica*-like, *aus*-like and *japonica*-like (Figure 4.8), which matched our haplotype analysis. The distribution of wild rice accessions having sequences similar to the regions selected during domestication can provide a clue of the geographical origins of rice cultivation. In the phylo-tree, the *japonica* branches separately and forms a single clade and closely allied with wild accession (*O. rufipogon*) from China, in agreement with previous genetic analysis by Huang et al. (2012), where the common wild rice (Or-III) located in southern China was likely the direct ancestor of *japonica*. On the other hand, the *indica* accessions in "*indica*-like" clade formed a monophyletic group with *O. nivara*, an annual wild rice mainly distributed in South Asia, as proposed by a previous genome-wide population relationship analysis (Garris et al., 2005; Xu et al., 2011; Choi et al., 2017). Most of the morphological and physiological traits of *O. nivara* are similar to cultivated rice, hence it is possible that *O. nivara* could serve as *indica* rice progenitor (Ge and Sang, 2011). This results are also concordant with Jiang et al. (2012a), who analysed *TAC1* sequences of 113 varieties of cultivated rice and 48 wild progenitors and showed that all the *japonica* varieties were clustered with Hainan *O. rufipogon* while the other cluster was solely composed of *indica* varieties, *O. rufipogon* and *O. nivara*. The three Malaysian wild rice (*O. ruf* 2, *O. ruf* 3 and *O. ruf* 4) formed a minor clade branching out from "*aus*-like" cluster which was exclusively composed of *aus* cultivars originated along the Indian continent and weedy rice from Malaysia. Surprisingly, two Malaysian elite cultivars (MR219 and MR220) shared *aus*-like genotype at the *TAC1* locus. Considering that *aus* cultivars were grown in a comparatively smaller geographical region along the Himalayan foothills and not cultivated in Malaysia, the *aus* genotypes in Malaysia *Oryza* were likely to arise from introgression of *aus*-like wild rice that are growing within the vicinity of rice cultivation area. Taken together, the results of this study support the notion of multiple independent origins for *japonica*, *indica* and *aus*.

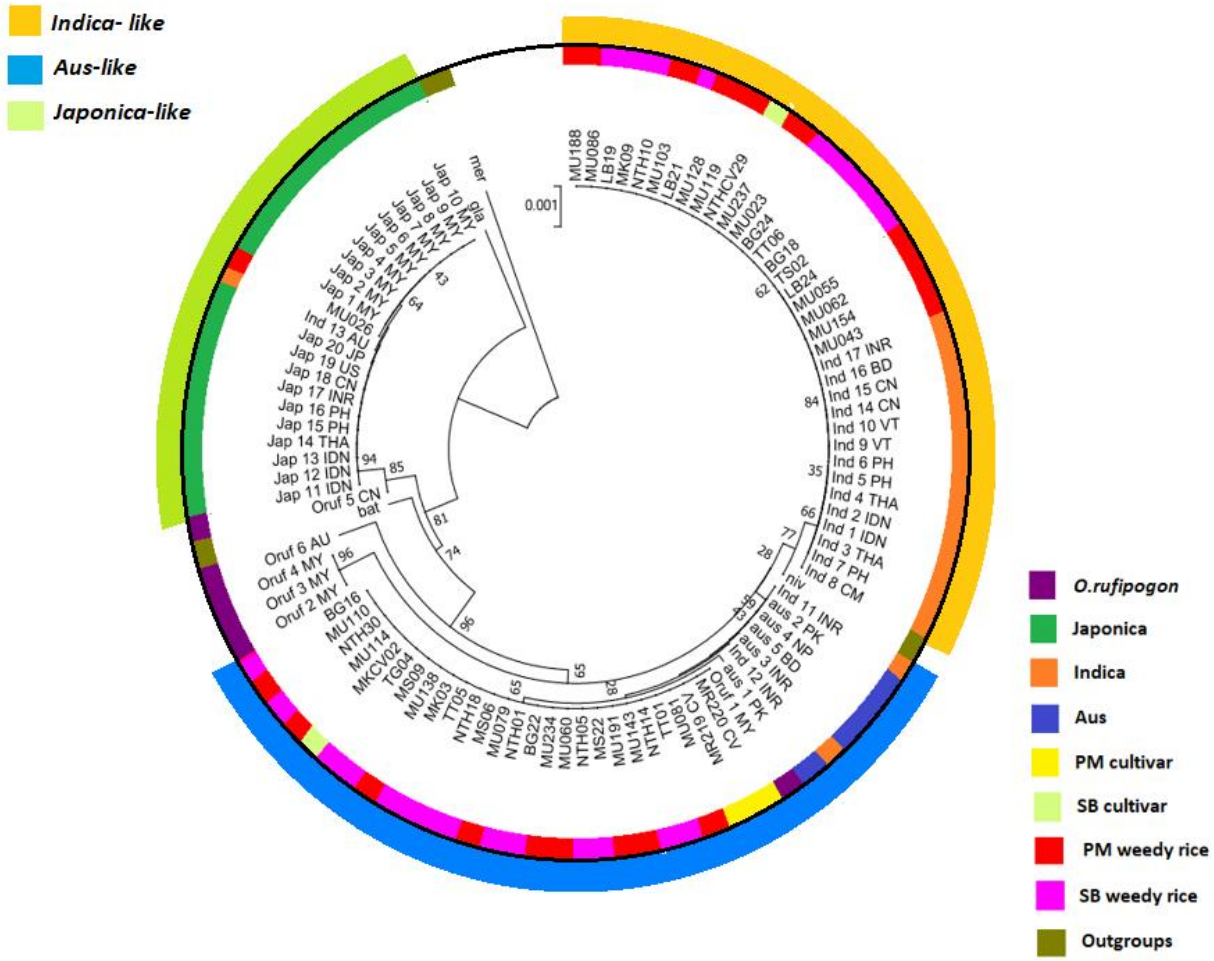


Figure 4.8: Neighbour-Joining tree constructed based on the sequence data of entire *TAC1* region. The colour range indicates the *Oryza* groups as labelled in the key. The outer ring indicates the *TAC1* genotype. The numbers on the branches indicates the bootstraps values. Abbrevation: Oruf: *O. rufipogon*, niv: *O. nivara*, Jap: *japonica*, Ind: *indica*, *aus*: *aus* type, gla*: O. glaberrima*, bat: *O. barthii*, mer: *O. meridonolis*
AU: Australia, BD: Bangladesh, CM: Cambodia, CN: China, IND: Indonesia, INR: India, JP: Japan, MY: Malaysia, NP: Nepal, PK: Pakistan, PH: Philippines, THA: Thailand, VT: Vietnam, US: United States of America, PM: Peninsular Malaysia, SB: Sabah

Two haplotypes, "*indica*-like" and "*aus*-like" coexisted in almost equal proportion of Malaysian weedy rice, creating an intermediate frequency of these alleles. Given the deep divergence between the two clades, these distributions suggest a long-term selective maintenance of both the *indica* and *aus*-like *TAC1* alleles in the Malaysian weedy rice population. Hence, it is believed that this is an ancient selected polymorphism which could be the result of introgression after speciation or retention of some alleles since the species split. Whether this non-neutral pattern of balancing polymorphism is a residual effect of the selection acting on *TAC1* or whether it is an indicative of other independent factors will require additional data for analysis.

CHAPTER 5

DISCUSSION

5.0 DISSCUSSION

Most rice farmers in Asia have shifted from transplanting to direct seeding owing to the rising scarcity of labour and water; however, due to physical and physiological similarities of weedy rice to cultivated rice and the absence of standing water at the time of crop emergence, adoption of direct-seeded rice systems makes weedy rice infestation one of the major threats to rice production (Chauhan, 2013; Rathore et al., 2013). Various studies have been conducted aiming to reveal the origin of weedy rice in different regions of the world (Chen et al., 2004; Cao et al., 2006; Londo et al., 2006; Pusadee et al., 2013; Song et al., 2014; Zhang et al., 2014). The great genetic variability in weedy rice, coupled with the long history of rice cultivation in Asia and the lack of historical records, have made it challenging to reconstruct its evolutionary history (Grimm et al., 2013).

History of weedy rice in Sabah is rather an interesting story. Being isolated by South China Sea from the Peninsular Malaysia, the occurrence of Sabah weedy rice was chronologically later than the spread of the weed in Peninsular Malaysia. Some weedy rice forms were found to share phenotypic similarities for some wild traits with their weedy counterparts from the mainland (Wahab and Suhaimi, 1991). These preconditions make Sabah weedy rice population a good model to study the emergence and evolution of weedy rice. The results of this study revealed Sabah local high yielding cultivars is the major contributor to Sabah weedy rice evolution. Besides, the peninsular weedy biotypes showed close genetic similarity in both genome wide GBS markers and *TAC1* gene variation, suggesting peninsular weedy rice, particularly the BR types, have shaped the evolution of Sabah weedy rice populations over recent time. In addition, presence of Peninsular Malaysia CV genetic constitution across Sabah weedy populations suggesting peninsular elite cultivars have played a role in Sabah weedy rice evolution.

5.1 Evolutionary origin of Sabah weedy rice

5.1.1 The role of local cultivars in Sabah weedy rice evolution

In Peninsular Malaysia, where rice is grown in close proximity to its wild ancestor, *O. rufipogon*, hybridization with wild populations appears to have contributed to the genetic composition of weedy rice types (Song et al., 2014; Cui et al., 2016; Vigueira et al., 2019). In regions where there is no wild rice population exists, weedy rice would likely be descended from co-occurring cultivated rice (Ellstrand et al., 2010), most likely through introgression from cultivated rice as described in Xia et al. (2011) and Jiang et al. (2012b) or de-domestication as suggested by Kanapeckas et al. (2016) and Qiu et al. (2017). Given there are no wild *O. rufipogon* present in northern Borneo, it ruled out the possibilities that Sabah weedy rice directly originated from the wild rice or from its hybrids. This study showed that weedy rice populations from Sabah state had very close genetic relationship with its very own high yielding cultivars. The results of STRUCTURE analysis (Figure 4.2) showed that most Sabah cultivars and weedy rice shared the genetic background of *indica* crop-like (green segment), indicating weedy rice may arise independently from cultivated rice. Similar observation was recorded by Tong et al. (2017) in which weedy rice were more likely from cultivated rice rather from wild rice in mitochondrial genome level. On the other hand, the considerable high affinity of weedy rice towards the Sabah local cultivars in coordinate and phylogeny analysis (Figure 4.3) strongly suggested the on-going crop-weed gene flow which in turn proves Sabah weedy rice are not derived from the wild rice but from the cultivated rice itself. Zhang et al. (2012) and Kanapeckas et al. (2016) also showed that weedy rice lineage is more closely related to rice cultivar grown in the sample, supporting the notion of gene introgression from cultivated rice as presented in this study. In addition, haplotype analyses of *TAC1* diversity equally suggest that coexisting cultivars could be the source of weeds evolution. For loci sampled across the entire 3 kb gene *TAC1* region, two distinct haplotypes shared between Sabah local cultivars and weedy rice accessions (H1 and H2; Figure 4.5) with each haplogroups tend to contain one Sabah cultivar. Judging from the abovementioned analyses, contribution of crop rice into weedy rice plays a significant role in the patterning of genetic diversity and population structure of Sabah weedy rice.

5.1.2 The exotic origin of Sabah weedy rice.

Results of this study reveal an important role for Peninsular Malaysia weeds in the recent establishment of Sabah weedy rice populations. Pooling the three Sabah weedy ecotypes (SB-SH, SB-SHA, SB-BR), 11% of weedy rice plants examined (7 out of 62) had peninsular brownhull (PM-BR) characterized by purple fragment with inferred ancestry membership more than 50 % in STRUCTURE output (Figure 4.2). In particular, four Sabah brownhull accessions from four different planting sites (NTH18, TA16, TT19 and TW05) exhibited PMBR genotype with greater than 75% of ancestry membership and it is well evident in the PCoA clustering. This suggesting the introduction of exotic weedy rice germplasm closely related to Sabah weedy rice varieties can facilitate the easiness of hybridization between the local weedy strains and the exotic germplasm due to limited reproductive barriers thus facilitating more successful hybridization (Cai et al., 2013). Strikingly, three Sabah cultivars; 2 from Kampung Telangtang (TT) and one from Kampung Longob (LB) planting areas contained genotype composition of PM-BR with more than 80% of membership and it is clearly depicted in phylogenetic tree where these three cultivars nested within the PM-BR clade. One possible reasoning for this scenario would be the peninsular brown hull (PM-BR) weedy rice were accidently brought into Sabah as contaminants of seed stocks. Lack of stringent enforcement at entry points in Sabah would facilitate frequent informal seeds sharing among farmers that lead to accidental introduction of weedy rice via contaminated seed stock. Weedy groups were introduced as stock seed contaminants had been reported previously across the world (Kane and Baack, 2007; Prathepha, 2009; Kanapeckas et al., 2016; Merotto Jr et al., 2016). Besides, frequent exchange of rice cultivar seeds or through other media, such as the machinery used for rice harvesting and water canals widespread weedy infestation to neighboring paddy fields. It is believed that weedy rice remained a marginal problem until the shift to mechanised direct seeding in Sabah in early 2000 (Chauhan, 2013, Grimm et al., 2013; Rathore et al., 2013).

In addition, the modern elite MR cultivar-derived genotype (light pink component) were found nearly in 90% of Sabah weedy rice and cultivars with approximately 20 - 30% ancestry membership. Given that peninsular elite MR cultivars have not been widely adopted by Sabah traditional farmers, the PM-CV genotype might have come indirectly from peninsular weedy rice that have been accidently introduced as seed stock contaminants. Besides, the *TAC1* analyses unravels a clear separation of '*aus*-like' haplotype in network analysis (Figure 4.6). Since Malaysia has no *aus* rice varieties, recent geneflow between *aus* cultivars and weedy rice in Sabah is unlikely. Transition from cultivating the indigenous traditional landraces to high yielding cultivars over the past several decades could have passed down several *aus*-like traits and their allelic variations including tiller architecture (conferred in part by *TAC1*). These traits might be expected to undergo adaptive introgression into present day high yielding cultivar populations in Sabah. Some of the weedy populations in U.S. are closely related to *aus* varieties, which is not cultivated in the USA, indicating US weeds are also of exotic origin just like Sabah weedy populations (Reagon et al., 2010). It will be interesting to investigate any other potential genetic sources contributing to Sabah weedy rice evolution particularly from neighboring regions namely Sarawak, Kalimantan and Philippines.

5.2 Multiple origins of Malaysian weedy rice

The results of this study show that the genetic diversity of Malaysian weedy rice population was not evenly distributed across geographical regions with the highest diversity in east Malaysia (H_e = 0.22) and the lowest in West Malaysia (H_e = 0.19). This might be attributed to different cultivation environment, different farming practices, seed source (e.g. BERNAS is the sole seed producer for Peninsular Malaysia whereas for Sabah, certified seeds produced by Department of Agriculture Sabah) and number of rice varieties used in different region. Multiple origins for weedy rice in Malaysia involved de-domestication of cultivars (endoferality) and genetic introgression from the wild rice *O. rufipogon* (exoferality). The wild rice strains appear to have contributed to genetic makeup of Malaysian weedy rice populations, particularly the "wild-like" Peninsular Malaysia BH

and BHA weedy populations, as shown in STRUCTURE result (Figure 4.2). Vigueira et al. (2019) explained the introgression of wild rice alleles into weedy rice populations could occur by two mechanisms: (i) hybridization of crop and wild populations; and (ii) de-domestication followed by introgression from wild rice. Wild rice contributions have been also been detected in weedy rice populations in Thailand (Prathepha, 2009; Pusadee et al., 2013). In the regions where there are no wild rice present (e.g., Japan, North America, Korea) (Cho et al., 1995; Cao et al., 2006; Londo and Schaal, 2007), weedy rice has evolved multiple times independently from different cultivated rice varieties. Weedy rice PM-SH and PM-SHA were suspected to be hybrid between the Peninsular Malaysia elite MR cultivars and PM-BR weedy rice, as supported by combined analysis of STRUCTURE and PCoA. Both PM-SH and PM-SHA exhibited a heavy admixed background containing PM-CV (light pink component) and PM-BR (purple component background) as well found staggered between elite cultivars and peninsular brown hull accessions in PCoA analysis. Besides, STRUCTURE analysis also suggested few weedy strains in Sabah show admixture pattern between the SB-SH and SB-BR gene pool. These results indicate that there is not only ongoing hybridization between cultivar and weedy rice, but also between weedy types. These observation suggest that hybridization contributes to formation of new intermediate weedy population that could be more competitive than the original types (Shivrain et al., 2009). Recently, Mispan et al. (2019), reported the emergence of new biotypes of weedy rice in Peninsular Malaysia which morphologically *vis-à-vis* the elite cultivar MR series (e.g. MR219, MR220) particularly in height, making the weedy rice unidentifiable and caused hand weeding impossible. The contribution of *aus* gene in Malaysian weedy rice was apparent in *TAC1* analysis. *Aus* cultivars does not occur in Malaysia suggesting *aus*-like landraces probably play a limited role in contemporary evolution of Malaysian weed population and these indigenous landraces, if any, should be investigated more carefully to understand the development of weedy populations from domesticated crops (Cui et al., 2016; Huang et al., 2017; Vigueira et al., 2019). The results of this study show that formation and evolution of weedy rice is an ongoing process, hence the future weed management therefore cannot only be the efficient treatment of already existing weedy rice populations but also the avoidance of the formation of new weedy rice

populations by adapting suitable agriculture practices, for example integrated management system (Grimm et al., 2013).

5.3 Weediness of tiller angle

One potential benefit in focusing on well characterized domestication is that the allelic variation at these genes can provide complementary insights into evolution of weediness traits. In weedy rice, compact architecture have been proposed as an adaptive trait (Huang et al., 2018). The *TAC1* gene has been implicated in this transition (Jiang et al., 2012a, Yu et al., 2007). In the present study, considering the small sampling area in Sabah (approximately 250 km^2), the tiller angle exhibited by Sabah weedy rice were varies greater than those of peninsular which are largely homogenous with narrow angle (Table 4.2). Since the tillering dynamic greatly influence by environmental factors including soil moisture (Mohapatra et al., 2011), this variation could be due the different farming environment between the Peninsular and Borneo, where rice planting areas in Sabah comprise both wet and drylands while the peninsular mainly consist of wetlands. Besides, the *indica* population in Malaysia presented wider variation in tiller angle compared with the *tropical japonica* population (Table 4.5a). Although genetic effects could explain the variation in tiller angle formation, the environment interactions with genotype also significantly affected tiller angle (Dong et al., 2016).

5.3.1 Signature selection of *TAC1*

In this study, the nucleotide diversity of *TAC1* gene region for all Malaysian *Oryza* (wild, weedy and cultivated rice) populations is relatively low and was the lowest among the three genic regions examined (gene region, from start to stop codon, 3' UTR and full length of *TAC1* regions). An identical 780bp long CDS with only two polymorphic sites that existed in *tropical japonica* and wild rice accessions has implicated the *TAC1* is highly conserved within *O. sativa* and its wild progenitor groups. The neutrality test of Jiang et al. (2012) demonstrated that when each of three regions of *TAC1* tested, only the 3' flanking region in the *Japonica* group showed significant difference (-2.0032, $P<0.05$),

suggested that *TAC1* of the *japonica* group was strongly selected during rice domestication. Contrary to this finding, no signature of selection was detected in any of the Malaysian weedy, wild and cultivated rice populations in this study. A similar finding was observed in *MsTAC1* of *Miscantus*, no significant selection was recorded in the genomic sequence or CDS (Zhao et al., 2014). The structural evolution of *TAC1* is important for plant architecture and haplotypes differences were expected between upland and lowland rice accessions where the upland accessions adapted to compact architecture while the lowland accession has wider plant habit (Yu and Nguyen, 1994). On the basis of previous studies, it is widely accepted that *indica* rice generally grows in tropical and subtropical humid environments, that is, it tends to be distributed in lower latitude areas and temperate low altitude areas, where it is adapted to the hot and humid climate (Xiong et al., 2010; 2011). In contrast, *Japonica* rice is generally distributed in temperate, subtropical cold environments (i.e., in high latitudes or altitudes of tropical areas, where it is adapted to the cold climate) (Zhao, 1991). So, the upland adapted japonicas to share the same haplotype H3 (*japonica*-like genotype) while the lowland adapted *indica* comprise two haplotypes; H1 (*aus*-like genotype) and H2 (*indica*-like genotype). A similar topology has been found in domesticated genes in Civáň et al. (2015). In this study, the *Japonica* is fixed to a single haplotype for *TAC1* and such fixation subjected to human selection pressure. Such scenario shown for the waxy locus where strong selective sweep detected in Olsen et al. (2006). Loss of diversity in *japonica* indicates a severe bottleneck during the selection of *TAC1* gene which might have been selected in ancestral progenitor and was subsequently passed to *Japonica* that facilitate the upland cultivation. The "*aus*-like" cluster is an interesting evolutionary observation. Though *aus* cultivars is not cultivated in Malaysia yet we observed substantial contribution of *aus*-like *TAC1* allele in Malaysian rice. The *aus*-like genotypes are most likely derived from introgression of *aus*-like wild types which exist around cultivation areas. In areas with no native wild species, Sabah in this case, this genotype could have been indirectly derived from *aus*-like wild rice through wild-like weedy rice from Peninsular Malaysia. This is in line with STRUCTURE results (Figure 4.2) where genetic background of peninsular weedy rice (PM-BR, purple segment) can be seen in Sabah weedy populations. In accordance with our observation, Wang et al. (2017) has reported a considerable

proportion of wild accession from Malaysia carrying substantial *aus* ancestry. Similar observation was made by Londo et al. (2006) for *Rc* gene, most U.S. red rice strains show evidence of past hybridisation between *aus*-like weeds.

5.3.2 Balancing selection of *TAC1*

Evolution of domestication crop primarily driven by strong directional selection imposed by humans. On the contrary, de-domesticated or feral plants are subjected to various pressure for survival in the agricultural environment (Qiu et al., 2017). In this candidate gene analysis, *TAC1* gene found has undergone potential balancing selection. It might due to the local environmental adaptation during which *TAC1* gene was target of balancing selection in response to the diverse nature selective pressure where the weedy rice were thriving. This could be explained by higher heterozygosity found in weedy rice compared to cultivated rice; the *indica* cultivars contain 3 polymorphic sites with nucleotide diversity (π) of 0.00066, however weedy rice had 9 polymorphic sites with 0.00107 nucleotide value (Table 4.3). Moreover, the Tajima's D values of these *TAC1* region in *indica* cultivars were very low (-0.754) depicting alleles with very low frequency in cultivated rice evolved to an intermediate frequency in weedy rice during de-domestication. Extensive phenotypic variation in tiller angle was also observed in weedy rice population (Table 4.2), indicating the better plasticity of tillers in different ecotypes. Balancing selection in plants often involved in pathogenic resistance and self-incompatibility (Delph and Kelly, 2014). One such example would be the abiotic stress related gene *PSTOL1,* where both functional and non-functional alleles were distinctly maintained in Asian rice (Vigueira et al., 2016). Similarly, Qiu et al. (2017) reported salt tolerance gene *OsEXPA3* has very high Tajima's D in weedy rice but low in cultivated rice (Tajima's D: weedy rice = 2.519; cultivated rice = -1.675), suggesting a strong balancing selection at this locus. As genetic introgression from wild and cultivar aided in weedy rice adaptability in agricultural settings, balancing selection might be expected to maintain high polymorphism of these adaptive alleles which leads to weedy rice evolution (Qiu et al., 2017).

CHAPTER 6

CONCLUSIONS

6.0 CONCLUSIONS

The recent changes of farming practices and cultivation methods with application of direct seeding with less weed management may have promoted the rapid proliferation and genetic diversification of weedy rice in Sabah. Though this study has not completely uncovered the enigma about the origin of weedy rice in East Malaysia, this combined analysis of GBS-SNP markers and *TAC1* suggest multiple origins for weedy rice in Sabah. A small fraction of Sabah weedy accessions derived from a genetically identical genetic make-up of Peninsular Malaysia weedy rice, supports the notion that weedy rice in Sabah may have an exotic origin. This is consistent with a recent finding revealing contribution of accidentally-introduced Peninsular Malaysia weedy rice to Sabah weedy counterparts using 24 SSR markers and *An-1* domestication gene (Neik et al., 2019). Besides, this study also has provided clear evidence that co-occurring local cultivars are major contributors to weedy rice genomes in Sabah predominantly through crop-weed introgression processes. The high level of gene flow and introgression of weedy rice in Sabah indicates that genes from cultivars can be easily incorporated into the gene pool of weedy rice. Therefore, adoption of herbicide resistance rice cultivation in Sabah might not be a suitable long-term option to eradicate weedy rice infestation as the high level of reproductive compatibility might promote rapid dissemination of the resistance trait to weedy rice populations (Song et al., 2014). If weedy strains were to receive herbicide-resistant alleles, perhaps in addition to other adaptive traits from cultivated counterparts which can significantly enhance the ecological fitness of the weeds, it may result in much complicated weedy rice problems (Chen et al., 2004). Future works could investigate other weedy populations from neighboring countries that could uncover additional aspects in the evolution of weedy rice not yet discovered in this study. Additional samples of landraces from Sabah should be included to understand the development of weedy populations from domesticated crops. It will be interesting to examine domestication genes controlling other traits such as pericarp colour, seed dormancy or seed shattering in Sabah weedy population. Practical implications such as strict enforcement at international boundaries to prevent further introductions of weedy rice into Sabah and

practicing integrated weed management strategies would help to curtail evolution of new weedy types in Sabah.

REFERENCES

ABDULLAH, M. Z. (1991) Wild relatives of rice in Malaysia: Their characteristics, distribution, ecology and potential in rice breeding. *MARDI Report,* **145**.

AKINBILE, C., et al. (2011) Rice Production and Water use Efficiency for Self-Sufficiency in Malaysia: A Review. *Trends in Applied Sciences Research,* **6** (10), 1127-1140.

ALLSTON, R. F. W. (1846) The rice plant. *DeBow's Review,* (1), 320-356.

ANUAR, N. H. S., et al. (2014) A comparative study of vegetative and reproductive growth of local weedy and clearfield® rice varieties in Malaysia. *Journal of the International Society for Southeast Asian Agricultural Sciences,* **20** (1), 41-51.

AZMI, M. (1992) Competitive ability of barnyard grass in direct-seeded rice. *Teknol. Padi,* **8**, 19-25.

AZMI, M., et al. (2000) *Management of weedy rice (Oryza sativa L.): The Malaysian experience.*

AZMI, M., et al. (2005) Weedy rice (Padi angin): a real threat to rice industry and farmers. In: *Proceedings of 4th National Seed Symposium, Malaysian Association of Seed Technologists, Putrajaya, Malaysia.*

AZMI, M. and BAKI, B. (2007) Weed flora landscapes of the Muda rice granary in the new millennium: a descriptive analysis. *J Trop Agric Food Sci,* **35**, 319-331.

AZMI, M. and KARIM, S. M. R. (2008) *Weedy rice : biology, ecology and management.* Kuala Lumpur, Malaysia: Malaysian Agricultural Research and Development Institute.

BAIRD, N. A., et al. (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLOS ONE,* **3** (10), e3376.

BAKI, B. B. (2004) Invasive Weed Species in Malaysian Agro-Ecosystems : Species, Impacts and Management. *Malaysian Journal of Science,* **23** (1), 1-42.

BRADBURY, P. J., et al. (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics,* **23** (19), 2633-2635.

BRONDANI, R. P. V., et al. (2005) Genetic Structure of Wild Rice *Oryza* Glumaepatula Populations in Three Brazilian Biomes Using Microsatellite Markers. *Genetica,* **125** (2), 115-123.

BURGER, J. C., et al. (2007) Rapid Phenotypic Divergence of Feral Rye from Domesticated Cereal Rye. *Weed Science,* **55** (3), 204-211.

BURGOS, N. R., et al. (2008) Red Rice (*Oryza sativa*) Status after 5 Years of Imidazolinone-Resistant Rice Technology in Arkansas. *Weed Technology,* **22** (1), 200-208.

CAI, X., et al. (2013) The Puzzle of Italian Rice Origin and Evolution: Determining Genetic Divergence and Affinity of Rice Germplasm from Italy and Asia. *PLOS ONE,* **8** (11), e80351.

CALLAWAY, E. (2014) Domestication: The birth of rice. *Nature,* **514**, S58.

CAO, Q., et al. (2006) Genetic diversity and origin of weedy rice (*Oryza sativa f. spontanea*) populations found in north-eastern China revealed by simple sequence repeat (SSR) markers. *Annals of botany,* **98** (6), 1241-1252.

CARNEY, J. A. (2004) 'With grains in her hair': rice in colonial Brazil. *Slavery & Abolition,* **25** (1), 1-27.

CATALA, M. M., et al. (2002) Red rice control in Southern Europe with pre- and postemergence applied herbicides.), International Rice Research Institute.

CHAMHURI, S., et al. (2014) *Issues and Challenges Facing Rice Production and Food Security in the Granary Areas in the East Coast Economic Region (ECER), Malaysia.* Vol. 7.

CHAUHAN, B. S. (2013) Strategies to manage weedy rice in Asia. *Crop Protection,* **48**, 51-56.

CHEN, L. J., et al. (2004) Gene flow from cultivated rice (*Oryza sativa*) to its weedy and wild relatives. *Annals of botany,* **93** (1), 67-73.

CHEN, Y., et al. (2012) Over-expression of *OsPIN2* leads to increased tiller numbers, angle and shorter plant height through suppression of *OsLAZY1*. *Plant Biotechnology Journal,* **10** (2), 139-149.

CHIN, D. V., et al. (2000) Weedy rice situation in Vietnam.) *FAO Report on Global Workshop on Red Rice Control.* Varadero, Cuba, FAO.

CHO, Y.-C., et al. (1995) Genetic characteristics of Korean weedy rice (*Oryza sativa L*.) by RFLP analysis. *Euphytica,* **86** (2), 103-110.

CHOI, J. Y., et al. (2017) The Rice Paradox: Multiple Origins but Single Domestication in Asian Rice. *Molecular biology and evolution,* **34** (4), 969-979.

CHUNG, Y. S., et al. (2017) Genotyping-by-sequencing: a promising tool for plant genetics research and breeding. *Horticulture, Environment, and Biotechnology,* **58** (5), 425-431.

CIVÁŇ, P., et al. (2015) Three geographically separate domestications of Asian rice. *Nature plants,* **1**, 15164-15164.

CODES, G. (2011) Sequencher version 4.10. 1 sequence analysis software. Gene Codes Corporation. *Ann Arbor, MI. http://genecodes. com (accessed 6 Dec. 2011)*.

CUI, Y., et al. (2016) Little white lies: pericarp color provides insights into the origins and evolution of southeast Asian weedy rice. *G3: Genes, Genomes, Genetics,* **6** (12), 4105-4114.

DANECEK, P., et al. (2011) The variant call format and VCFtools. *Bioinformatics (Oxford, England),* **27** (15), 2156-2158.

DAVEY, J. W., et al. (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics,* **12** (7), 499.

DE DATTA, S. K. and BALTAZAR, A. M. (1996) Weed control technology as a component of rice production systems. In: AULD, B. A. and KIM, K. U. (Eds.) *In Weed management in rice.* Vol. 139. Rome, Italy: FAO. Plant Production and Protection Paper, pp. 27-52.

DE WET, J. M. J. and HARLAN, J. R. (1975) Weeds and Domesticates: Evolution in the man-made habitat. *Economic Botany,* **29** (2), 99-108.

DELOUCHE, J. C., et al. (2007) *Weedy rices - origin, biology, ecology and control / by James C. Delouche ... [et al.] ; with the collaboration of Michael Larinde and Cadmo Rosell.* (FAO plant production and protection papers, 0259-2517 ; 188.) Rome: Food and Agriculture Organization of the United Nations.

DELOUCHE, J. C. and LABRADA, R. (2007) *Weedy rices: origin, biology, ecology and control.* Vol. 188. Food & Agriculture Org., 200.

DELPH, L. F. and KELLY, J. K. (2014) On the importance of balancing selection in plants. *The New phytologist,* **201** (1), 45-56.

DEPARTMENT OF AGRICULTURE, M. (2015) *Paddy Production Survey Report Malaysia.* Department of Agriculture Peninsular Malaysia.

DONG, H., et al. (2016) A Novel Tiller Angle Gene, *TAC3*, together with *TAC1* and *D2* Largely Determine the Natural Variation of Tiller Angle in Rice Cultivars. *PLOS Genetics,* **12** (11), e1006412.

DOS REIS GOULART, I. C. G., et al. (2012) Identification of origin and analysis of population structure of field-selected imidazolinone-herbicide resistant red rice (*Oryza sativa*). *Euphytica,* **187** (3), 437-447.

DUNNA, V. and ROY, B. (2013) Rice (*Oryza sativa L.*). In) p. 52.

DURAND-MORAT, A., et al. (2018) The implications of red rice on food security. *Global Food Security,* **18**, 62-75.

EARL, D. A. and VONHOLDT, B. M. (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources,* **4** (2), 359-361.

ELLSTRAND, N. C., et al. (2010) Crops gone wild: evolution of weeds and invasives from domesticated ancestors. *Evolutionary Applications,* **3** (5-6), 494-504.

ELSHIRE, R. J., et al. (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLOS ONE,* **6** (5), e19379.

EVANNO, G., et al. (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology,* **14** (8), 2611-2620.

FAHAD, S., et al. (2019) Chapter 1 - Major Constraints for Global Rice Production. In: HASANUZZAMAN, M., et al. (Eds.) *Advances in Rice Research for Abiotic Stress Tolerance.* Woodhead Publishing, pp. 1-22.

FERRERO, A. (2003) *Weedy rice, biological features and control.*

FERRERO, A. and VIDOTTO, F. (1998) Red rice control in rice pre- and post-planting. In: FAO (Ed.) *FAO Report of the Global workshop on red rice control.* Varadero ,Cuba.

FOGLIATTO, S., et al. (2010) Effects of winter flooding on weedy rice (*Oryza sativa* L.). *Crop Protection,* **29** (11), 1232-1240.

FULLER, D. Q., et al. (2010) Consilience of genetics and archaeobotany in the entangled history of rice. *Archaeological and Anthropological Sciences,* **2** (2), 115-131.

GAO, H., et al. (2007) A Markov Chain Monte Carlo Approach for Joint Inference of Population Structure and Inbreeding Rates From Multilocus Genotype Data. *Genetics,* **176** (3), 1635.

GARRIS, A. J., et al. (2005) Genetic Structure and Diversity in *Oryza sativa* L. *Genetics,* **169** (3), 1631.

GE, S. and SANG, T. (2011) Inappropriate model rejects independent domestications of *indica* and *japonica* rice. *Proceedings of the National Academy of Sciences,* **108** (39), E755.

GEALY, D. (2005) Gene Movement between Rice (*Oryza sativa*) and Weedy Rice (*Oryza sativa*) — a U.S. Temperate Rice Perspective. In: GRESSEL, J. (Ed.) *Crop ferality and volunteerism.* Vol. 99. Boca Raton: CRC Press, pp. 323-354.

GEALY, D. R., et al. (2009) Exploring genetic and spatial structure of US weedy red rice (*Oryza sativa*) in relation to rice relatives worldwide. *Weed Science,* **57** (6), 627-643.

GEDIL, M., et al. (2016) Perspectives on the application of next-generation sequencing to the improvement of Africa's staple food crops. In) *Next Generation Sequencing–Advances, Applications and Challenges.* InTechOpen, pp. 287-321.

GLASZMANN, J. C. (1987) Isozymes and classification of Asian rice varieties. *Theoretical and Applied Genetics,* **74** (1), 21-30.

GLAUBITZ, J. C., et al. (2014) TASSEL-GBS: a high capacity genotyping-by-sequencing analysis pipeline. *PLOS ONE,* **9** (2), e90346.

GOSS, W. L. and BROWN, E. (1939) Buried red rice seed. *Journal of the American Society of Agronomy,* **31** (7), 633-637.

GRESSEL, J. (2005) *Crop ferality and volunteerism.* Boca Raton, FL: Boca Raton, FL : Taylor & Francis.

GRESSEL, J. and VALVERDE, B. E. (2009) A strategy to provide long-term control of weedy rice while mitigating herbicide resistance transgene flow, and its potential use for other crops with related weeds. *Pest Management Science,* **65** (7), 723-731.

GRIMM, A., et al. (2013) Microsatellite markers reveal multiple origins for Italian weedy rice. *Ecology and Evolution,* **3** (14), 4786-4798.

GROSS, B. L. and ZHAO, Z. (2014) Archaeological and genetic insights into the origins of domesticated rice. *Proceedings of the National Academy of Sciences,* **111** (17), 6190.

GU, X.-Y., et al. (2003) Inheritance of Seed Dormancy in Weedy Rice. *Crop Science,* **43** (3), 835-843.

GUO, L., et al. (2014) Application of resequencing to rice genomics, functional genomics and evolutionary analysis. *Rice (New York, N.Y.),* **7** (1), 4-4.

HARLAN, J. R. (1982) Relationships between weeds and crops. In: HOLZNER, W. and NUMATA, M. (Eds.) *Biology and ecology of weeds.* Dordrecht: Springer Netherlands, pp. 91-96.

HASANUZZAMAN, M., et al. (2009) Evaluation of Preemergence Herbicide and Hand Weeding on the Weed Control Efficiency and Performance of Transplanted Aus Rice. *Am Eurasian J Agron. ,* **2(3)**, 138-143.

HASHMI, M. and SIAU, T., JOANNA. (2016) *Mineral contents of some indigenous rice varities of Sabah Malaysia.* Vol. 2.

HE, J., et al. (2017) Fine mapping and candidate gene analysis of qTAC8, a major quantitative trait locus controlling tiller angle in rice (*Oryza sativa* L.). *PLOS ONE,* **12** (5), e0178177-e0178177.

HE, J., et al. (2014) Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Frontiers in plant science,* **5**, 484-484.

HERMAN, T., et al. (2015) Rice Production and Climate Change: A Case Study of Malaysian Rice. *Pertanika Journal of Tropical Agricultural Science,* **38** (3).

HOLM, L. G., et al. (1977) *The world's worst weeds.* Honolulu, Hawaii: University Press.

HU, M., et al. (2018) The domestication of plant architecture in African rice. *The Plant Journal,* **94** (4), 661-669.

HUANG, P. U., et al. (2012) Phylogeography of Asian wild rice, *Oryza rufipogon*: a genome-wide view. *Molecular Ecology,* **21** (18), 4593-4604.

HUANG, Y.-F., et al. (2014) Using Genotyping-By-Sequencing (GBS) for Genomic Discovery in Cultivated Oat. *PLOS ONE,* **9** (7), e102448.

HUANG, Z., et al. (2018) The Role of Standing Variation in the Evolution of Weedines Traits in South Asian Weedy Rice (*Oryza* spp.). *G3: Genes|Genomes|Genetics,* **8** (11), 3679.

HUANG, Z., et al. (2017) All roads lead to weediness: Patterns of genomic divergence reveal extensive recurrent weedy rice origins from South Asian *Oryza*. *Molecular Ecology,* **26** (12), 3151-3167.

IDRIS, R. (2018) *Food Security in Sabah: An Overview for Rice Self Sufficient Target.* Vol. 5.

ISHIKAWA, R., et al. (2005) Origin of weedy rice grown in Bhutan and the force of genetic diversity. *Genetic Resources and Crop Evolution,* **52** (4), 395-403.

ISLAM, Z., et al. (2007) Influence of mechanization on rice pests in Peninsular Malaysia and their management. *Outlooks on Pest Management,* **18** (2), 87.

ISMAIL, A. M., et al. (2012) Adaptation to flooding during emergence and seedling growth in rice and weeds, and implications for crop establishment. *AoB PLANTS,* **2012**, pls019-pls019.

JAYA SURIA, A. S. M., et al. (2011) Efficacy and economics of different herbicides in aerobic rice system. *African Journal of Biotechnology,* **10** (41), 8007-8022.

JI, H., et al. (2010) Inactivation of the CTD phosphatase-like gene *OsCPL1* enhances the development of the abscission layer and seed shattering in rice. *The Plant Journal,* **61** (1), 96-106.

JIANG, J., et al. (2012a) Molecular Evolution of the *TAC1* Gene from Rice (*Oryza sativa* L.). *Journal of Genetics and Genomics,* **39** (10), 551-560.

JIANG, Z., et al. (2012b) Introgression from cultivated rice influences genetic differentiation of weedy rice populations at a local spatial scale. *Theoretical and Applied Genetics,* **124** (2), 309-322.

JIN, J., et al. (2008) Genetic control of rice plant architecture under domestication. *Nat Genet,* **40** (11), 1365-1369.

JINDALOUANG, R., et al. (2018) Growth and nutrients competition between weedy rice and crop rice in a replacement series study. *International Journal of Agriculture and Biology,* **20** (4), 784-790.

JURAIMI, A., et al. (2013) Sustainable weed management in direct-seeded rice culture: A review. *Australian Journal of Crop Science,* **7** (7), 989-1002.

KANAPECKAS, K. L., et al. (2016) Escape to Ferality: The Endoferal Origin of Weedy Rice from Crop Rice through De-Domestication. *PLOS ONE,* **11** (9), e0162676.

KANE, N. C. and BAACK, E. J. (2007) The origins of weedy rice. *Molecular Ecology,* **16** (21), 4423-4425.

KANE, N. C. and RIESEBERG, L. H. (2008) Genetics and evolution of weedy Helianthus annuus populations: adaptation of an agricultural weed. *Molecular Ecology,* **17** (1), 384-394.

KARIM, R. S., et al. (2004) Weed problems and their management in rice fields of Malaysia: An overview. *Weed Biology and Management,* **4** (4), 177-186.

KHAIRUN NISAA, M. N., et al. (2018) Assessing economic sustainability of Malaysian rice farming (1972-2014). *Journal of Academia Universiti Teknologi MARA Negeri Sembilan,* **6** (1), 6.

KHAZANAH RESEARCH INSTITUTE (2019) *The Status of the Paddy and Rice Industry in Malaysia.* Kuala Lumpur: Khazanah Research Institute.

KHUSH, G. S. (1997) Origin, dispersal, cultivation and variation of rice. *Plant Molecular Biology,* **35** (1), 25-34.

KILIAN, A., et al. (2012) Diversity arrays technology: a generic genome profiling technology on open platforms. In) *Data production and analysis in population genomics.* Springer, pp. 67-89.

KIM, H., et al. (2016) Population Dynamics Among six Major Groups of the *Oryza rufipogon* Species Complex, Wild Relative of Cultivated Asian Rice. *Rice,* **9** (1), 56.

KONISHI, S., et al. (2006) An SNP Caused Loss of Seed Shattering During Rice Domestication. *Science,* **312** (5778), 1392.

KOPELMAN, N. M., et al. (2015) Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular ecology resources,* **15** (5), 1179-1191.

KOVACH, M. J., et al. (2007) New insights into the history of rice domestication. *Trends in Genetics,* **23** (11), 578-587.

KUMAR, S., et al. (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular biology and evolution,* **33** (7), 1870-1874.

LAMDIN, J., et al. (2015) Sabah's roles and contributions towards fulfilling the nation's strategies in achieving 100% SSL in rice production.) Kota Kinabalu, Department of Agriculture Sabah.

LEIGH, J. and BRYANT, D. (2015) *PopART: Full-Feature Software for Haplotype Network Construction.* Vol. 6.

LEITÃO, F. E. D. F., et al. (1972) Estudos de competição entre o arroz vermelho e o arroz cultivado. *Bragantia,* **31**, 249-258.

LI, C., et al. (2006) Rice Domestication by Reducing Shattering. *Science,* **311** (5769), 1936.

LI, L.-F., et al. (2017) Signatures of adaptation in the weedy rice genome. *Nature Genetics,* **49** (5), 811-814.

LI, P., et al. (2007a) *LAZY1* controls rice shoot gravitropism through regulating polar auxin transport. *Cell Research,* **17**, 402.

LI, S., et al. (2007b) Characterization and Use of Male Sterility in Hybrid Rice Breeding. *Journal of Integrative Plant Biology,* **49** (6), 791-804.

LINARES, O. (2002) African rice (*Oryza glaberrima*): History and future potential. *Proc. Natl. Acad. Sci. U. S. A.,* **99** (25), 16360-16365.

LIU, H., et al. (2014) An evaluation of genotyping-by-sequencing (GBS) to map the Breviaristatum-e (ari-e) locus in cultivated barley. *BMC genomics,* **15** (1), 104.

LONDO, J. and SCHAAL, B. (2007) Origins and population genetics of weedy red rice in the USA. *Molecular Ecology,* **16** (21), 4523-4535.

LONDO, J. P., et al. (2006) Phylogeography of Asian wild rice, *Oryza rufipogon* reveals multiple independent domestications of cultivated rice, *Oryza sativa*. *Proceedings of the National Academy of Sciences,* **103** (25), 9578.

MATLOOB, A., et al. (2014) *Weeds of rice in Asia: problems and opportunities.*

MCCOUCH, S. R., et al. (2007) Through the genetic bottleneck: *O. rufipogon* as a source of trait-enhancing alleles for O. *sativa*. *Euphytica,* **154** (3), 317-339.

MEROTTO JR, A., et al. (2016) Evolutionary and social consequences of introgression of nontransgenic herbicide resistance from rice to weedy rice in Brazil. *Evolutionary Applications,* **9** (7), 837-846.

MEYER, R. S. and PURUGGANAN, M. D. (2013) Evolution of crop species: genetics of domestication and diversification. *Nature Reviews Genetics,* **14**, 840.

MISHRA, J. S. and SINGH, V. P. (2007) Integrated weed management in zero-till direct-seeded rice (*Oryza sativa*) - Wheat (*Triticum aestivum*) cropping system. *Indian J Agron,* **52**, 198-203.

MISLAMAH, A. B. (1990) Rice weeds and problems in weed control in Seberang Perak, Malaysia. *Proceedings of the Third Tropical Weed Science Conference (Kuala Lumpur, Malaysia, 4-6 December 1990)*.

MISPAN, M. S., et al. (2019) Managing weedy rice (*Oryza sativa L*.) in Malaysia: challenges and ways forward. *Journal of Research in Weed Science,* **2** (Issue 2. pp. 90-179), 149-167.

MONACO, T. J., et al. (2002) *Weed science: principles and practices.* John Wiley & Sons.

MOODY, K. (1993) Weed Control in Wet-Seeded Rice. *Experimental Agriculture,* **29** (4), 393-403.

MOUMOUNI, K., et al. (2015) Construction of a genetic map for pearl millet, Pennisetum glaucum (L.) R. Br., using a genotyping-by-sequencing (GBS) approach. *Molecular breeding,* **35** (1), 5.

MULLER, M.H., et al. (2011) The origin and evolution of a recent agricultural weed: population genetic diversity of weedy populations of sunflower (*Helianthus annuus L*.) in Spain and France. *Evolutionary applications,* **4** (3), 499-514.

MUTHAYYA, S., et al. (2014) An overview of global rice production, supply, trade, and consumption. *Annals of the New York Academy of Sciences,* **1324** (1), 7-14.

NADIR, S., et al. (2017) Weedy rice in sustainable rice production. A review. *Agronomy for Sustainable Development,* **37** (5), 46.

NAJIM, M. M. M., et al. (2007) *Sustainability of rice production: A Malaysian perspective.* Vol. 3.

NEIK, T. X., et al. (2019) When West Meets East: The Origins and Spread of Weedy Rice Between Continental and Island Southeast Asia. *G3: Genes|Genomes|Genetics*, g3.400021.2019.

NGU, M. S., et al. (2010) Genetic Structure of *Oryza rufipogon Griff*. Natural Populations in Malaysia: Implications for Conservation and Genetic Introgression of Cultivated Rice. *Tropical Plant Biology,* **3** (4), 227-239.

NOLDIN, J. A. (2000) Red Rice Status and Management in the Americas. *Wild and Weedy Rice in Rice Ecosystems in Asia-A Review*, 21-24.

NOLDÍN, J. A. (1995) Characterization, seed longevity, and herbicide sensitivity of red rice (*Oryza sativa L*.) ecotypes, and red rice control in soybeans (*Glycine max* (*L.*) *Merr.*).) College Station, Texas, Texas A&M University

NOLDIN, J. A., et al. (2006) Seed longevity of red rice ecotypes buried in soil. *Planta Daninha,* **24**, 611-620.

OJHA, P. and KWATRA, S. (2014) Analysis of different paddy transplanting methods in northern India: Ergo - economical study. *Journal of Applied and Natural Science,* **6** (2).

OLSEN, K. and WENDEL, J. (2013) Crop plants as models for understanding plant adaptation and diversification. *Frontiers in Plant Science,* **4** (290).

OLSEN, K. M., et al. (2007) Evolutionary genomics of weedy rice in the USA. *Journal of Integrative Plant Biology,* **49** (6), 811-816.

OLSEN, K. M., et al. (2006) Selection Under Domestication: Evidence for a Sweep in the Rice Genomic Region. *Genetics,* **173** (2), 975.

PEAKALL, R. and SMOUSE, P. E. (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics,* **28** (19), 2537-2539.

PENG, J., et al. (1999) 'Green revolution' genes encode mutant gibberellin response modulators. *Nature,* **400** (6741), 256-261.

PETERSON, G. W., et al. (2014) Genotyping-by-sequencing for plant genetic diversity analysis: a lab guide for SNP genotyping. *Diversity,* **6** (4), 665-680.

PETERSON, P. M. (2013) Poaceae (Gramineae). *eLS*.

POLAND, J., et al. (2012a) Genomic selection in wheat breeding using genotyping-by-sequencing. *The Plant Genome,* **5** (3), 103-113.

POLAND, J. A., et al. (2012b) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLOS ONE,* **7** (2), e32253.

PRATHEPHA, P. (2009) Seed morphological traits and genotypic diversity of weedy rice (*Oryza sativa f. spontanea*) populations found in the Thai Hom Mali rice fields of north-eastern Thailand. *Weed Biology and Management,* **9** (1), 1-9.

PRAVAT K. MOHAPATRA, et al. (2011) Plasticity of Tiller Dynamics in Wild Rice *Oryza rufipogon* Griff.: A Strategy for Resilience in Suboptimal Environments. *International Journal of Agronomy,* **2011**.

PRITCHARD, J. K., et al. (2000) Inference of population structure using multilocus genotype data. *Genetics,* **155** (2), 945-959.

PUSADEE, T., et al. (2013) Population structure of the primary gene pool of *Oryza sativa* in Thailand. *Genetic resources and crop evolution,* **60** (1), 335-353.

QI, X., et al. (2015) More than one way to evolve a weed: parallel evolution of US weedy rice through independent genetic mechanisms. *Mol. Ecol.,* **24** (13), 3329-3344.

QIU, J., et al. (2017) Genomic variation associated with local adaptation of weedy rice during de-domestication. *Nature Communications,* **8**, 15323.

QIU, J., et al. (2014) Genome re-sequencing suggested a weedy rice origin from domesticated *indica-japonica* hybridization: a case study from southern China. *Planta,* **240** (6), 1353-1363.

RAJAMOORTHY, Y., et al. (2015) Rice Industry in Malaysia: Challenges, Policies and Implications. *Procedia Economics and Finance,* **31**, 861-867.

RAO, A. N., et al. (2007) Weed Management in Direct-Seeded Rice. In: SPARKS, D. L. (Ed.) *Advances in Agronomy.* Vol. 93. Academic Press, pp. 153-255.

RATHORE, M., et al. (2013) Weedy rice: an emerging threat to rice cultivation and options for its management. *Current Science,* **105** (8), 1067-1072.

RATHORE, M., et al. (2016) Characterization of functional trait diversity among Indian cultivated and weedy rice populations. *Scientific reports,* **6**, 24176-24176.

REAGON, M., et al. (2010) Genomic patterns of nucleotide diversity in divergent populations of US weedy rice. *BMC Evolutionary Biology,* **10** (1), 180.

RODENBURG, J., et al. (2008) Weedy Rices - Origin, Biology, Ecology and Control. FAO Plant Production and Protection Paper 188. *Experimental Agriculture,* **44** (2), 276-276.

ROZAS, J., et al. (2017) DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular biology and evolution,* **34** (12), 3299-3302.

ROZEN, S. and SKALETSKY, H. (2000) Primer3 on the WWW for general users and for biologist programmers. In) *Bioinformatics methods and protocols.* Springer, pp. 365-386.

SAHA, S., et al. (2014) Weedy rice: problems and its management. *Indian Journal of Weed Science,* **46** (1), 14-22.

SANG, T. and GE, S. (2007) The Puzzle of Rice Domestication. *Journal of Integrative Plant Biology,* **49** (6), 760-768.

SASAKI, T. (2001) *Oryza sativa* (Rice). In: BRENNER, S. and MILLER, J. H. (Eds.) *Encyclopedia of Genetics.* New York: Academic Press, pp. 1394-1395.

SAVARY, S., et al. (1997) A characterisation of rice pests and quantification of yield losses in the rice-wheat system of India. *Crop Protection,* **16** (4), 387-398.

SHIVRAIN, V. K., et al. (2007) Gene flow between Clearfield™ rice and red rice. *Crop Protection,* **26** (3), 349-356.

SHIVRAIN, V. K., et al. (2009) Gene flow from weedy red rice (*Oryza sativa* L.) to cultivated rice and fitness of hybrids. *Pest management science,* **65** (10), 1124-1129.

SHIVRAIN, V. K., et al. (2010) Diversity of weedy red rice (*Oryza sativa* L.) in Arkansas, USA in relation to weed management. *Crop Protection,* **29** (7), 721-730.

SHRESTHA, S., et al. (2018) Weedy Rice: Competitive Ability, Evolution, and Diversity. In) *Integrated View of Population Genetics.*

SINGH, K., et al. (2013) *Weedy Rice: An Emerging Threat for Direct-seeded Rice Production Systems in India.* Vol. 01.

SONG, B.-K., et al. (2014) Malaysian weedy rice shows its true stripes: wild *Oryza* and elite rice cultivars shape agricultural weed evolution in Southeast Asia. *Molecular Ecology,* **23** (20), 5003-5017.

SONG, Z. P., et al. (2003) Genetic diversity in the northernmost *Oryza rufipogon* populations estimated by SSR markers. *Theoretical and Applied Genetics,* **107** (8), 1492-1499.

SOUKI, H. (2015) Prestasi dan ciri-ciri biji padi varieti tempatan yang ditanam di Kota Belud, Sabah. *Buletin Teknologi MARDI,* **8**, 52-62.

SUDIANTO, E., et al. (2016) Morphology of Malaysian weedy rice (*Oryza sativa*): diversity, origin and implications for weed management. *Weed Science,* **64** (3), 501-512.

SUN, C. Q., et al. (2001) Comparison of the genetic diversity of common wild rice (*Oryza rufipogon* Griff.) and cultivated rice (O. *sativa* L.) using RFLP markers. *Theoretical and Applied Genetics,* **102** (1), 157-162.

SUN, J., et al. (2019) Population Genomic Analysis and *De Novo* Assembly Reveal the Origin of Weedy Rice as an Evolutionary Game. *Molecular Plant,* **12** (5), 632-647.

SWEENEY, M. and MCCOUCH, S. (2007) The Complex History of the Domestication of Rice. *Annals of Botany,* **100** (5), 951-957.

TAN, L., et al. (2008) Control of a key transition from prostrate to erect growth in rice domestication. *Nature Genetics,* **40**, 1360.

TANG, L. H. and MORISHIMA, H. (1997) Genetic Characterization of Weedy Rices and the Inference on their Origins. *Breed. Sci,* **47**, 153-160.

TEWARI, A. N. (2008) Weedy rice and its management. In: Y. SING, et al. (Eds.) *Direct seeding or rice and weed management in the irrigated rice-wheat cropping system of the Indo-Gangetic plains.* pp. 205-211.

THE STAR. (2009) Wild padi species a threat to Sabah farmers.) *The Star Online.* Malaysia. Available at https://www.thestar.com.my/news/community/2009/11/10/wild-padi-species-a-threat-to-sabah-farmers/ [Viewed 15 June 2019]

THURBER, C. S., et al. (2011) Timing is everything: early degradation of abscission layer is associated with increased seed shattering in U.S. weedy rice. *BMC Plant Biology,* **11** (1), 14.

THURBER, C. S., et al. (2010) Molecular evolution of shattering loci in U.S. weedy rice. *Molecular ecology,* **19** (16), 3271-3284.

TONG, W., et al. (2017) Genetic variation architecture of mitochondrial genome reveals the differentiation in Korean landrace and weedy rice. *Scientific Reports,* **7**, 43327.

TSENG, T. M., et al. (2013) Inter- and intrapopulation variation in dormancy of *Oryza sativa* (weedy red rice) and allelic variation in dormancy-linked loci. *Weed Research,* **53** (6), 440-451.

TUONG, T. P. and BOUMAN, B. A. M. (2003) *Rice production in water-scarce environments.* (Water Productivity in Agriculture: Limits and Opportunities for Improvement) Wallingford: CABI Publishing.

TZVELEV, N. N. (1989) The system of grasses (Poaceae) and their evolution. *The Botanical Review,* **55** (3), 141-203.

VAUGHAN, D. A. (1994) *The Wild Relatives of Rice: A Genetic Resources Handbook.* IRRI, International Rice Research Institute.

VAUGHAN, D. A., et al. (2008) The evolving story of rice evolution. *Plant Science,* **174** (4), 394-408.

VAUGHAN, D. A., et al. (2003) Diversity in the *Oryza* genus. *Current Opinion in Plant Biology,* **6** (2), 139-146.

VAUGHAN, L., et al. (2001) Is all red rice found in commercial rice really *Oryza sativa*? *Weed Science,* **49**, 468-476.

VEASEY, E. A., et al. (2004) Variation in the loss of seed dormancy during after-ripening of wild and cultivated rice species. *Annals of botany,* **94** (6), 875-882.

VIDOTTO, F. and FERRERO, A. (2000) Germination behaviour of red rice (*Oryza sativa* L.) seeds in field and laboratory conditions. *Agronomie,* **20** (4), 375-382.

VIGUEIRA, C., et al. (2013) The red queen in the corn: agricultural weeds as models of rapid adaptive evolution. *Heredity,* **110** (4), 303.

VIGUEIRA, C. C., et al. (2019) Call of the wild rice: *Oryza rufipogon* shapes weedy rice evolution in Southeast Asia. *Evolutionary Applications,* **12** (1), 93-104.

VIGUEIRA, C. C., et al. (2016) Long-term balancing selection at the Phosphorus Starvation Tolerance 1 (*PSTOL1*) locus in wild, domesticated and weedy rice (*Oryza*). *BMC Plant Biology,* **16** (1), 101.

WAHAB, A. and SUHAIMI, O. (1991) Padi angin characteristics, adverse effects and methods of its eradication. *Teknologi Padi,* **7**, 21-31.

WANG, H., et al. (2017) Asian wild rice is a hybrid swarm with extensive gene flow and feralization from domesticated rice. *Genome research,* **27** (6), 1029-1038.

WANG, Y. and LI, J. (2008) Rice, rising. *Nature Genetics,* **40**, 1273.

WANG, Z. Y. and TANKSLEY, S. D. (1989) Restriction fragment length polymorphism in *Oryza sativa* L. *Genome,* **32** (6), 1113-1118.

WATANABE, H. (1996) Weedy rice problems in southeast Asia and control strategy. *JIRCAS International Symposium Series (Japan),* (4), 27-37.

WEDGER, M. J. and OLSEN, K. M. (2018) Evolving insights on weedy rice. *Ecological Genetics and Genomics,* **7-8**, 23-26.

XIA, H.-B., et al. (2011) Conspecific Crop-Weed Introgression Influences Evolution of Weedy Rice (*Oryza sativa* f. spontanea) across a Geographical Range. *PLOS ONE,* **6** (1), e16189.

XIONG, H. B., et al. (2012) Origin and evolution of weedy rice revealed by inter-subspecific and inter-varietal hybridizations in rice. *Molecular Plant Breed,* **10**, 131-139.

XU, K., et al. (2016) The *PIN1* family gene *PvPIN1* is involved in auxin-dependent root emergence and tillering in switchgrass. *Genetics and molecular biology,* **39** (1), 62-72.

XU, X., et al. (2011) Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nature Biotechnology,* **30**, 105.

XU, Y., et al. (1998) Transgressive segregation of tiller angle in rice caused by complementary gene action. *Crop Science,* **38** (1), 12.

YOSHIHARA, T. and IINO, M. (2007) Identification of the Gravitropism-Related Rice Gene *LAZY1* and Elucidation of *LAZY1*-Dependent and -Independent Gravity Signaling Pathways. *Plant and Cell Physiology,* **48** (5), 678-688.

YU, B., et al. (2007) *TAC1*, a major quantitative trait locus controlling tiller angle in rice. *The Plant Journal,* **52** (5), 891-898.

YU, L. X. and NGUYEN, H. T. (1994) Genetic variation detected with RAPD markers among upland and lowland rice cultivars (*Oryza sativa L*.). *Theoretical and Applied Genetics,* **87** (6), 668-672.

ZHANG, L., et al. (2012) Genetic diversity and origin of *japonica*- and *indica*-like rice biotypes of weedy rice in the Guangdong and Liaoning provinces of China. *Genetic Resources and Crop Evolution,* **59** (3), 399-410.

ZHANG, S., et al. (2014) Genetic Differentiation of Asian Weedy Rice Revealed with InDel Markers. *Crop Science,* **54**, 2499-2508.

ZHANG, Z. P. (2003) Development of chemical weed control and integrated weed management in China. *Weed Biology and Management,* **3** (4), 197-203.

ZHAO, H., et al. (2014) Natural variation and genetic analysis of the tiller angle gene *MsTAC1* in Miscanthussinensis. *Planta,* **240** (1), 161-175.

ZHOU, H.-F., et al. (2003) Microsatellite analysis of genetic diversity and population genetic structure of a wild rice (*Oryza rufipogon* Griff.) in China. *Theoretical and Applied Genetics,* **107** (2), 332-339.

ZHOU, Y., et al. (2012) Genetic Control of Seed Shattering in Rice by the *APETALA2* Transcription Factor *SHATTERING ABORTION1*. *The Plant Cell,* **24** (3), 1034.

ZHU, Y., et al. (2012) Sequence polymorphisms in wild, weedy, and cultivated rice suggest seed-shattering locus *sh4* played a minor role in Asian rice domestication. *Ecology and evolution,* **2** (9), 2106-2113.

APPENDICES

Appendix 1

Table 1. Summary of 182 rice accessions used in this study including country of collection, type and hull phenotype.

Accession ID	*Oryza* Species	Type	Hull morphology	Country	Source
CL01	*sativa (indica)*	Cultivar (elite)	Strawhull	Malaysia,Peninsular	
CL02	*sativa (indica)*	Cultivar (elite)	Strawhull	Malaysia, Peninsular	
MR218	*sativa (indica)*	Cultivar (elite)	Strawhull	Malaysia,Peninsular	
MR219	*sativa (indica)*	Cultivar (elite)	Strawhull	Malaysia,Peninsular	
MR220	*sativa (indica)*	Cultivar (elite)	Strawhull	Malaysia,Peninsular	
MR221	*sativa (indica)*	Cultivar (elite)	Strawhull	Malaysia, Peninsular	
LBCV04	*sativa (indica)*	Cultivar (high yielding)	Strawhull	Malaysia, Sabah	Dr. Song BK group
LBCV25	*sativa (indica)*	Cultivar (high yielding)	Strawhull	Malaysia, Sabah	Dr. Song BK group
MKCV01	*sativa (indica)*	Cultivar (high yielding)	Strawhull	Malaysia, Sabah	Dr. Song BK group
MKCV04	*sativa (indica)*	Cultivar (high yielding)	Strawhull	Malaysia, Sabah	Dr. Song BK group
NTHCV29	*sativa (indica)*	Cultivar (high yielding)	Strawhull	Malaysia, Sabah	
TGCV01	*sativa (indica)*	Cultivar (high yielding)	Strawhull	Malaysia, Sabah	
TTCV01	*sativa (indica)*	Cultivar (high yielding)	Strawhull	Malaysia, Sabah	Dr. Song BK group
TTCV03	*sativa (indica)*	Cultivar (high yielding)	Strawhull	Malaysia, Sabah	Dr. Song BK group

MUSC001	*sativa (indica)*	weedy	Strawhull	Malaysia, Peninsular	
MUSC032	*sativa (indica)*	weedy	Strawhull	Malaysia, Peninsular	
MUSC043	*sativa (indica)*	weedy	Strawhull	Malaysia, Peninsular	
MUSC069	*sativa (indica)*	weedy	Strawhull	Malaysia, Peninsular	
MUSC086	*sativa (indica)*	weedy	Strawhull	Malaysia, Peninsular	
MUSC089	*sativa (indica)*	weedy	Strawhull	Malaysia, Peninsular	
MUSC103	*sativa (indica)*	weedy	Strawhull	Malaysia, Peninsular	
MUSC106	*sativa (indica)*	weedy	Strawhull	Malaysia, Peninsular	
MUSC110	*sativa (indica)*	weedy	Strawhull	Malaysia, Peninsular	
MUSC117	*sativa (indica)*	weedy	Strawhull	Malaysia, Peninsular	
MUSC138	*sativa (indica)*	weedy	Strawhull	Malaysia, Peninsular	
MUSC143	*sativa (indica)*	weedy	Strawhull	Malaysia, Peninsular	
MUSC150	*sativa (indica)*	weedy	Strawhull	Malaysia, Peninsular	
MUSC161	*sativa (indica)*	weedy	Strawhull	Malaysia, Peninsular	
MUSC225	*sativa (indica)*	weedy	Strawhull	Malaysia, Peninsular	
MUSC023	*sativa (indica)*	weedy	Strawhull awned	Malaysia, Peninsular	
MUSC067	*sativa (indica)*	weedy	Strawhull awned	Malaysia, Peninsular	
MUSC098	*sativa (indica)*	weedy	Strawhull awned	Malaysia, Peninsular	
MUSC100	*sativa (indica)*	weedy	Strawhull awned	Malaysia, Peninsular	
MUSC114	*sativa (indica)*	weedy	Strawhull awned	Malaysia, Peninsular	
MUSC122	*sativa (indica)*	weedy	Strawhull awned	Malaysia, Peninsular	
MUSC170	*sativa (indica)*	weedy	Strawhull awned	Malaysia, Peninsular	
MUSC187	*sativa (indica)*	weedy	Strawhull awned	Malaysia, Peninsular	

MUSC007	*sativa (indica)*	weedy	Brownhull	Malaysia, Peninsular	
MUSC021	*sativa (indica)*	weedy	Brownhull	Malaysia, Peninsular	
MUSC038	*sativa (indica)*	weedy	Brownhull	Malaysia, Peninsular	
MUSC052	*sativa (indica)*	weedy	Brownhull	Malaysia, Peninsular	
MUSC055	*sativa (indica)*	weedy	Brownhull	Malaysia, Peninsular	
MUSC062	*sativa (indica)*	weedy	Brownhull	Malaysia, Peninsular	
MUSC065	*sativa (indica)*	weedy	Brownhull	Malaysia, Peninsular	
MUSC079	*sativa (indica)*	weedy	Brownhull	Malaysia, Peninsular	
MUSC119	*sativa (indica)*	weedy	Brownhull	Malaysia, Peninsular	
MUSC128	*sativa (indica)*	weedy	Brownhull	Malaysia, Peninsular	
MUSC132	*sativa (indica)*	weedy	Brownhull	Malaysia, Peninsular	
MUSC134	*sativa (indica)*	weedy	Brownhull	Malaysia, Peninsular	
MUSC142	*sativa (indica)*	weedy	Brownhull	Malaysia, Peninsular	
MUSC146	*sativa (indica)*	weedy	Brownhull	Malaysia, Peninsular	
MUSC157	*sativa (indica)*	weedy	Brownhull	Malaysia, Peninsular	
MUSC141	*sativa (indica)*	weedy	Brownhull	Malaysia, Peninsular	
MUSC008	*sativa (indica)*	weedy	Brownhull-awned	Malaysia, Peninsular	
MUSC014	*sativa (indica)*	weedy	Brownhull-awned	Malaysia, Peninsular	
MUSC022	*sativa (indica)*	weedy	Brownhull-awned	Malaysia, Peninsular	
MUSC060	*sativa (indica)*	weedy	Brownhull-awned	Malaysia, Peninsular	
MUSC072	*sativa (indica)*	weedy	Brownhull-awned	Malaysia, Peninsular	
MUSC081	*sativa (indica)*	weedy	Brownhull-awned	Malaysia, Peninsular	
MUSC145	*sativa (indica)*	weedy	Brownhull-awned	Malaysia, Peninsular	

MUSC154	*sativa (indica)*	weedy	Brownhull-awned	Malaysia, Peninsular	
MUSC188	*sativa (indica)*	weedy	Brownhull-awned	Malaysia, Peninsular	
MUSC232	*sativa (indica)*	weedy	Brownhull-awned	Malaysia, Peninsular	
MUSC233	*sativa (indica)*	weedy	Brownhull-awned	Malaysia, Peninsular	
MUSC237	*sativa (indica)*	weedy	Brownhull-awned	Malaysia, Peninsular	
MUSC005	*sativa (indica)*	weedy	Blackhull	Malaysia, Peninsular	
MUSC026	*sativa (indica)*	weedy	Blackhull	Malaysia, Peninsular	
MUSC066	*sativa (indica)*	weedy	Blackhull	Malaysia, Peninsular	
MUSC147	*sativa (indica)*	weedy	Blackhull	Malaysia, Peninsular	
MUSC027	*sativa (indica)*	weedy	Blackhull awned	Malaysia, Peninsular	
MUSC137	*sativa (indica)*	weedy	Blackhull awned	Malaysia, Peninsular	
MUSC171	*sativa (indica)*	weedy	Blackhull awned	Malaysia, Peninsular	
MUSC178	*sativa (indica)*	weedy	Blackhull awned	Malaysia, Peninsular	
MUSC234	*sativa (indica)*	weedy	Blackhull awned	Malaysia, Peninsular	
MUSC235	*sativa (indica)*	weedy	Blackhull awned	Malaysia, Peninsular	
KS05	*sativa (indica)*	weedy	Strawhull	Malaysia,Sabah	Dr. Song BK group
KS08	*sativa (indica)*	weedy	Strawhull	Malaysia,Sabah	Dr. Song BK group
LB021	*sativa (indica)*	weedy	Strawhull	Malaysia,Sabah	Dr. Song BK group
LB08	*sativa (indica)*	weedy	Strawhull	Malaysia,Sabah	Dr. Song BK group
MK03	*sativa (indica)*	weedy	Strawhull	Malaysia,Sabah	
MK04	*sativa (indica)*	weedy	Strawhull	Malaysia,Sabah	Dr. Song BK group
MK05	*sativa (indica)*	weedy	Strawhull	Malaysia,Sabah	Dr. Song BK group
MK12	*sativa (indica)*	weedy	Strawhull	Malaysia,Sabah	Dr. Song BK group

MS13	*sativa (indica)*	weedy	Strawhull	Malaysia,Sabah	Dr.Song BK group
MS22	*sativa (indica)*	weedy	Strawhull	Malaysia,Sabah	
NTH08	*sativa (indica)*	weedy	Strawhull	Malaysia,Sabah	Dr. Song BK group
NTH11	*sativa (indica)*	weedy	Strawhull	Malaysia,Sabah	Dr. Song BK group
NTH13	*sativa (indica)*	weedy	Strawhull	Malaysia,Sabah	
NTH19	*sativa (indica)*	weedy	Strawhull	Malaysia,Sabah	Dr. Song BK group
NTH32	*sativa (indica)*	weedy	Strawhull	Malaysia,Sabah	Dr. Song BK group
TA08	*sativa (indica)*	weedy	Strawhull	Malaysia,Sabah	Dr. Song BK group
TA12	*sativa (indica)*	weedy	Strawhull	Malaysia,Sabah	Dr. Song BK group
TA14	*sativa (indica)*	weedy	Strawhull	Malaysia,Sabah	Dr. Song BK group
TT01	*sativa (indica)*	weedy	Strawhull	Malaysia,Sabah	Dr. Song BK group
TT02	*sativa (indica)*	weedy	Strawhull	Malaysia,Sabah	Dr. Song BK group
TT04	*sativa (indica)*	weedy	Strawhull	Malaysia,Sabah	Dr. Song BK group
TT05	*sativa (indica)*	weedy	Strawhull	Malaysia,Sabah	
TT07	*sativa (indica)*	weedy	Strawhull	Malaysia,Sabah	Dr. Song BK group
TT08	*sativa (indica)*	weedy	Strawhull	Malaysia,Sabah	Dr. Song BK group
TT10	*sativa (indica)*	weedy	Strawhull	Malaysia,Sabah	Dr. Song BK group
TT14	*sativa (indica)*	weedy	Strawhull	Malaysia,Sabah	Dr. Song BK group
TT15	*sativa (indica)*	weedy	Strawhull	Malaysia,Sabah	Dr. Song BK group
TT16	*sativa (indica)*	weedy	Strawhull	Malaysia,Sabah	Dr. Song BK group
TT20	*sativa (indica)*	weedy	Strawhull	Malaysia,Sabah	Dr. Song BK group
TT06	*sativa (indica)*	weedy	Strawhull	Malaysia,Sabah	
BG43	*sativa (indica)*	weedy	Strawhull awned	Malaysia,Sabah	Dr. Song BK group

KS04	*sativa (indica)*	weedy	Strawhull awned	Malaysia,Sabah	Dr. Song BK group
MS10	*sativa (indica)*	weedy	Strawhull awned	Malaysia,Sabah	Dr. Song BK group
MS17	*sativa (indica)*	weedy	Strawhull awned	Malaysia,Sabah	Dr. Song BK group
NTH14	*sativa (indica)*	weedy	Strawhull awned	Malaysia,Sabah	Dr. Song BK group
TS02	*sativa (indica)*	weedy	Strawhull awned	Malaysia,Sabah	
BG03	*sativa (indica)*	weedy	Brownhull	Malaysia,Sabah	Dr. Song BK group
BG09	*sativa (indica)*	weedy	Brownhull	Malaysia,Sabah	Dr. Song BK group
BG12	*sativa (indica)*	weedy	Brownhull	Malaysia,Sabah	Dr. Song BK group
BG18	*sativa (indica)*	weedy	Brownhull	Malaysia,Sabah	
BG22	*sativa (indica)*	weedy	Brownhull	Malaysia,Sabah	
LB02	*sativa (indica)*	weedy	Brownhull	Malaysia,Sabah	
LB16	*sativa (indica)*	weedy	Brownhull	Malaysia,Sabah	Dr. Song BK group
LB19	*sativa (indica)*	weedy	Brownhull	Malaysia,Sabah	Dr. Song BK group
MK06	*sativa (indica)*	weedy	Brownhull	Malaysia,Sabah	Dr. Song BK group
MK09	*sativa (indica)*	weedy	Brownhull	Malaysia,Sabah	
MS09	*sativa (indica)*	weedy	Brownhull	Malaysia,Sabah	
MS18	*sativa (indica)*	weedy	Brownhull	Malaysia,Sabah	Dr. Song BK group
NTH01	*sativa (indica)*	weedy	Brownhull	Malaysia,Sabah	
NTH05	*sativa (indica)*	weedy	Brownhull	Malaysia,Sabah	
NTH10	*sativa (indica)*	weedy	Brownhull	Malaysia,Sabah	
NTH18	*sativa (indica)*	weedy	Brownhull	Malaysia,Sabah	
NTH30	*sativa (indica)*	weedy	Brownhull	Malaysia,Sabah	
TA01	*sativa (indica)*	weedy	Brownhull	Malaysia,Sabah	Dr. Song BK group

TA03	*sativa (indica)*	weedy	Brownhull	Malaysia,Sabah	Dr. Song BK group
TA16	*sativa (indica)*	weedy	Brownhull	Malaysia,Sabah	
TA21	*sativa (indica)*	weedy	Brownhull	Malaysia,Sabah	Dr. Song BK group
TT09	*sativa (indica)*	weedy	Brownhull	Malaysia,Sabah	Dr. Song BK group
TW02	*sativa (indica)*	weedy	Brownhull	Malaysia,Sabah	Dr. Song BK group
TW05	*sativa (indica)*	weedy	Brownhull	Malaysia,Sabah	Dr. Song BK group
TW07	*sativa (indica)*	weedy	Brownhull	Malaysia,Sabah	Dr. Song BK group
TS17	*sativa (indica)*	weedy	Brownhull	Malaysia,Sabah	Dr. Song BK group
CM105728	*Sativa*	weedy		Cambodia	Vigueira et al. 2019
CM106316	*Sativa*	weedy		Cambodia	Vigueira et al. 2019
CM106338	*Sativa*	weedy		Cambodia	Vigueira et al. 2019
CM89191	*Sativa*	weedy		Cambodia	Vigueira et al. 2019
CM89235	*Sativa*	weedy		Cambodia	Vigueira et al. 2019
CM89240	*Sativa*	weedy		Cambodia	Vigueira et al. 2019
IDN100206	*Sativa*	weedy		Indonesia	Vigueira et al. 2019
IDN106097	*Sativa*	weedy		Indonesia	Vigueira et al. 2019
IDN99601	*Sativa*	weedy		Indonesia	Vigueira et al. 2019
IDN99599	*Sativa*	weedy		Indonesia	Vigueira et al. 2019
TH100218	*Sativa*	weedy		Thailand	Vigueira et al. 2019
TH104406	*Sativa*	weedy		Thailand	Vigueira et al. 2019
TH105353	*Sativa*	weedy		Thailand	Vigueira et al. 2019
TH105362	*Sativa*	weedy		Thailand	Vigueira et al. 2019
TH105798	*Sativa*	weedy		Thailand	Vigueira et al. 2019

TH105851	*Sativa*	weedy		Thailand	Vigueira et al. 2019
TH82037	*Sativa*	weedy		Thailand	Vigueira et al. 2019
TH86449	*Sativa*	weedy		Thailand	Vigueira et al. 2019
VT106431	*Sativa*	weedy		Vietnam	Vigueira et al. 2019
VT102157	*Sativa*	weedy		Vietnam	Vigueira et al. 2019
VT106332	*Sativa*	weedy		Vietnam	Vigueira et al. 2019
VT86494	*Sativa*	weedy		Vietnam	Vigueira et al. 2019
MUSC201	*rufipogon*	wild		Malaysia	
MUSC202	*rufipogon*	wild		Malaysia	Vigueira et al. 2019
MUSC203	*rufipogon*	wild		Malaysia	Vigueira et al. 2019
MUSC207	*rufipogon*	wild		Malaysia	Vigueira et al. 2019
MUSC208	*rufipogon*	wild		Malaysia	Vigueira et al. 2019
MUSC209	*rufipogon*	wild		Malaysia	Vigueira et al. 2019
PL16	*rufipogon*	wild		Malaysia	
CM105720	*rufipogon*	wild		Cambodia	Vigueira et al. 2019
CM106321	*rufipogon*	wild		Cambodia	Vigueira et al. 2019
IDN106453	*rufipogon*	wild		Indonesia	Vigueira et al. 2019
IDN86478	*rufipogon*	wild		Indonesia	Vigueira et al. 2019
TH104812	*rufipogon*	wild		Thailand	Vigueira et al. 2019
TH104764	*rufipogon*	wild		Thailand	Vigueira et al. 2019
TH104875	*rufipogon*	wild		Thailand	Vigueira et al. 2019
TH105855	*rufipogon*	wild		Thailand	Vigueira et al. 2019
TH105942	*rufipogon*	wild		Thailand	Vigueira et al. 2019
VT106518	*rufipogon*	wild		Vietnam	Vigueira et al. 2019

VT106166	*rufipogon*	wild		Vietnam	Vigueira et al. 2019
MUSC212	*officinalis*	wild		Malaysia	Vigueira et al. 2019
MUSC213	*officinalis*	wild		Malaysia	Vigueira et al. 2019
MUSC214	*officinalis*	wild		Malaysia	
MUSC215	*officinalis*	wild		Malaysia	Vigueira et al. 2019
MUSC216	*officinalis*	wild		Malaysia	Vigueira et al. 2019
VT106332	*Sativa*	weedy		Vietnam	Vigueira et al. 2019
VT86494	*Sativa*	weedy		Vietnam	Vigueira et al. 2019
MUSC201	*rufipogon*	wild		Malaysia	
MUSC202	*rufipogon*	wild		Malaysia	Vigueira et al. 2019
MUSC203	*rufipogon*	wild		Malaysia	Vigueira et al. 2019
MUSC207	*rufipogon*	wild		Malaysia	Vigueira et al. 2019
MUSC208	*rufipogon*	wild		Malaysia	Vigueira et al. 2019
MUSC209	*rufipogon*	wild		Malaysia	Vigueira et al. 2019
PL16	*rufipogon*	wild		Malaysia	
CM105720	*rufipogon*	wild		Cambodia	Vigueira et al. 2019
CM106321	*rufipogon*	wild		Cambodia	Vigueira et al. 2019
IDN106453	*rufipogon*	wild		Indonesia	Vigueira et al. 2019
IDN86478	*rufipogon*	wild		Indonesia	Vigueira et al. 2019
TH104812	*rufipogon*	wild		Thailand	Vigueira et al. 2019
TH104764	*rufipogon*	wild		Thailand	Vigueira et al. 2019
TH104875	*rufipogon*	wild		Thailand	Vigueira et al. 2019
TH105855	*rufipogon*	wild		Thailand	Vigueira et al. 2019
TH105942	*rufipogon*	wild		Thailand	Vigueira et al. 2019

VT106518	*rufipogon*	wild		Vietnam	Vigueira et al. 2019
VT106166	*rufipogon*	wild		Vietnam	Vigueira et al. 2019
MUSC212	*officinalis*	wild		Malaysia	Vigueira et al. 2019
MUSC213	*officinalis*	wild		Malaysia	Vigueira et al. 2019
MUSC214	*officinalis*	wild		Malaysia	
MUSC215	*officinalis*	wild		Malaysia	Vigueira et al. 2019
MUSC216	*officinalis*	wild		Malaysia	Vigueira et al. 2019

Appendix 2

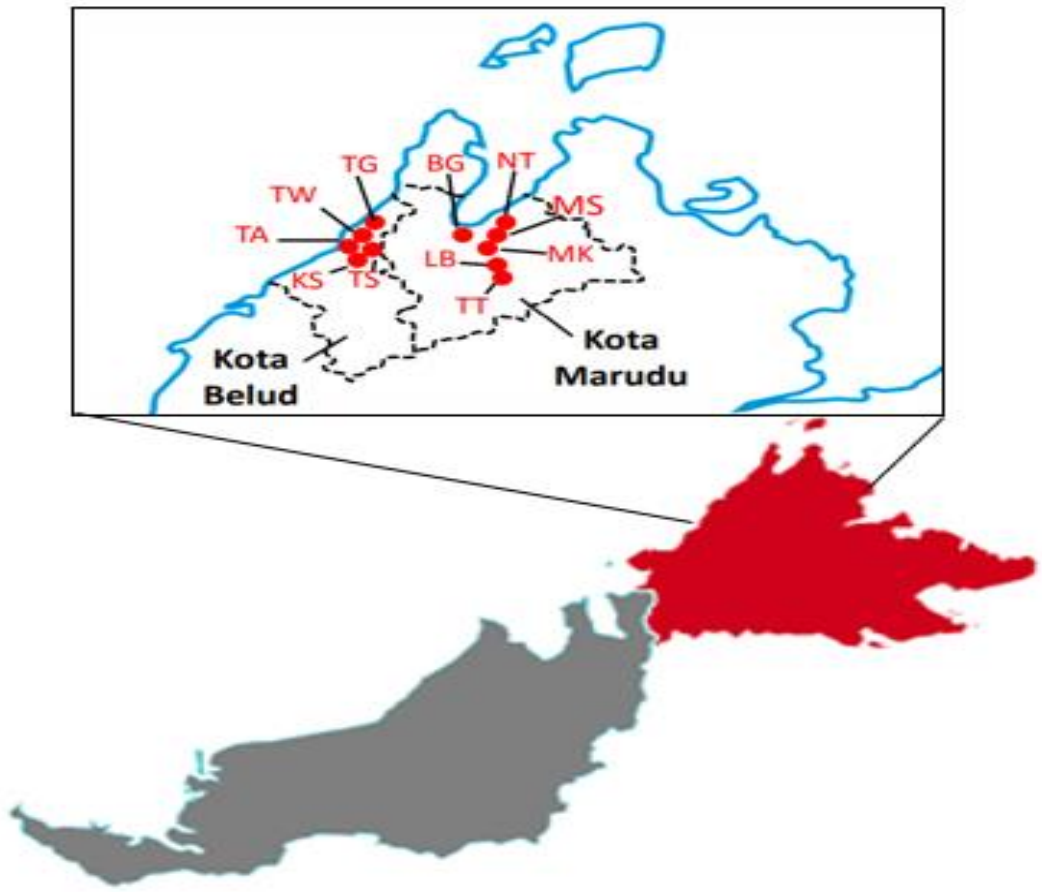


Figure 1. Sampling locations in Sabah. Abbrevations: BG- Kampung Balagaton; MS- Kampung Masalog; KS- Kampung Sangkir; LB- Kampung Longob; MK- Kampung Mangkalua; NT- Kampung Nolotan; TA- Kampung Tamau; TS- Kampung Taun Usik; TW- Kampung Tawadakan; TT- Kampung Telangtang. Extracted from Neik et al. (2019).

Appendix 3

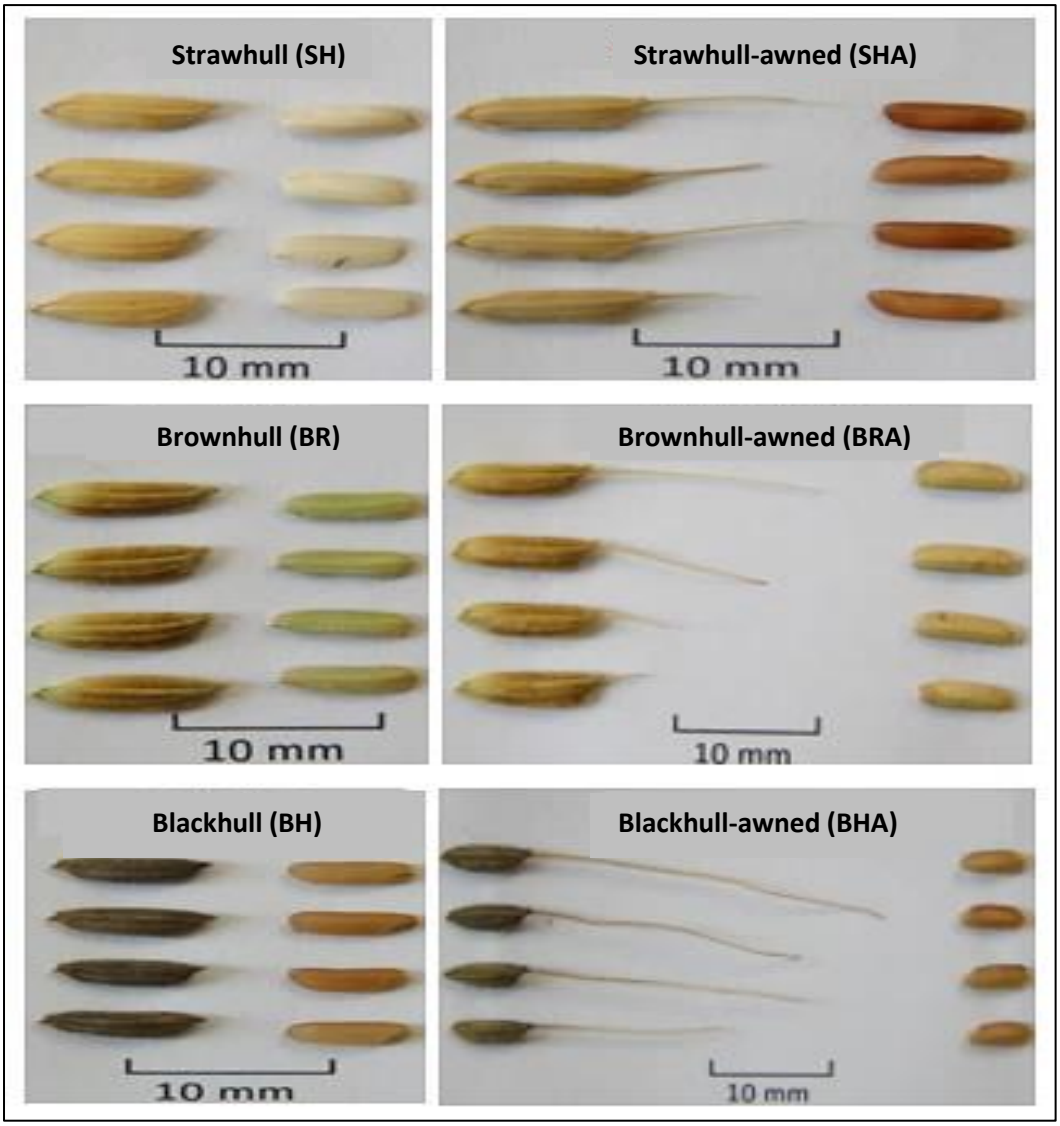


Figure 2. Hull morphology of Malaysian weedy rice

Appendix 4

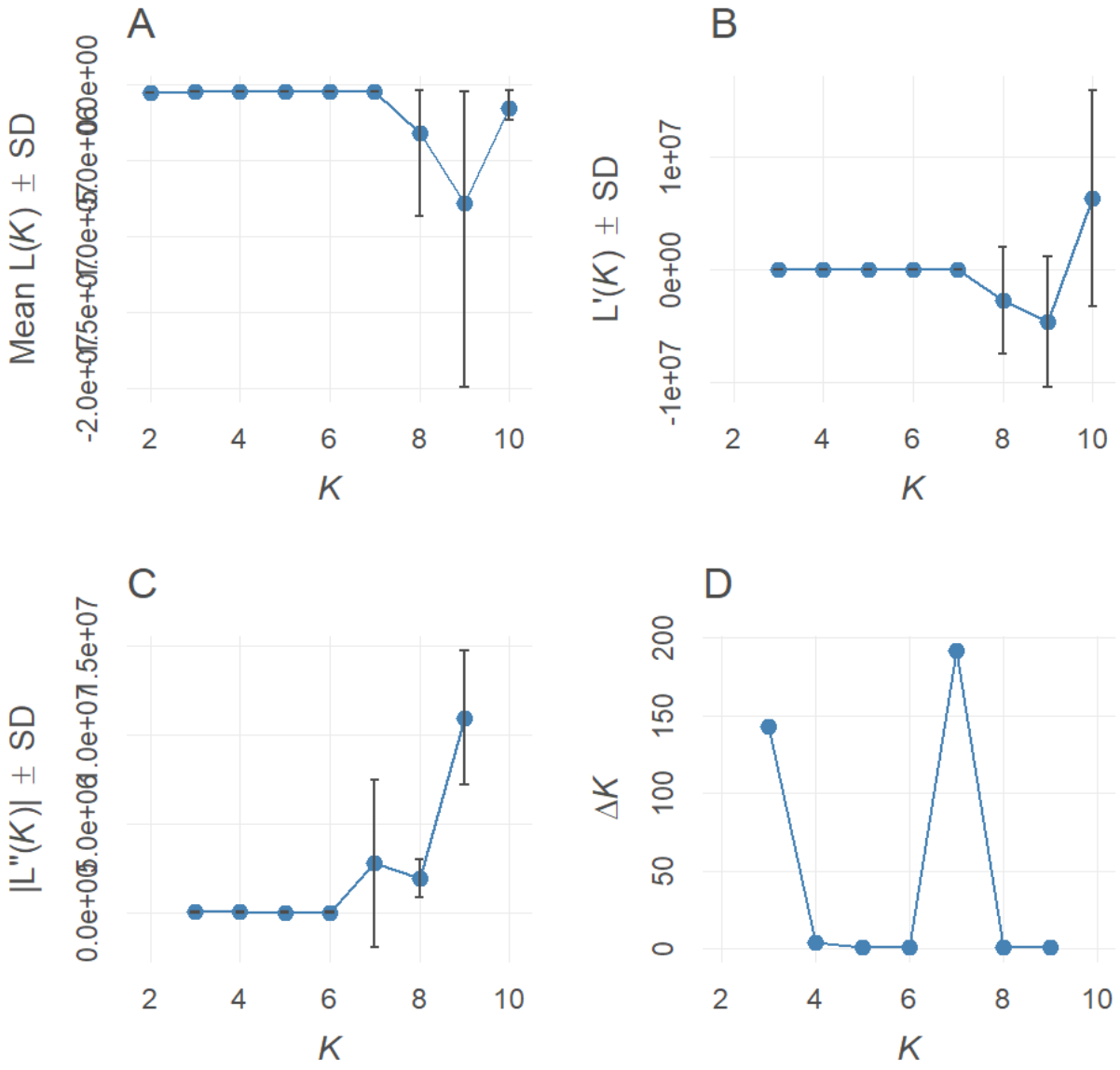


Figure 3. STRUCTURE analysis by Evanno method. A) Plot of L(*K*) vs *K* B) Plot of L'(*K*) vs *K* C) Plot of |L''(*K*)| vs *K* D) Plot of Delta *K* vs *K.*

Appendix 5

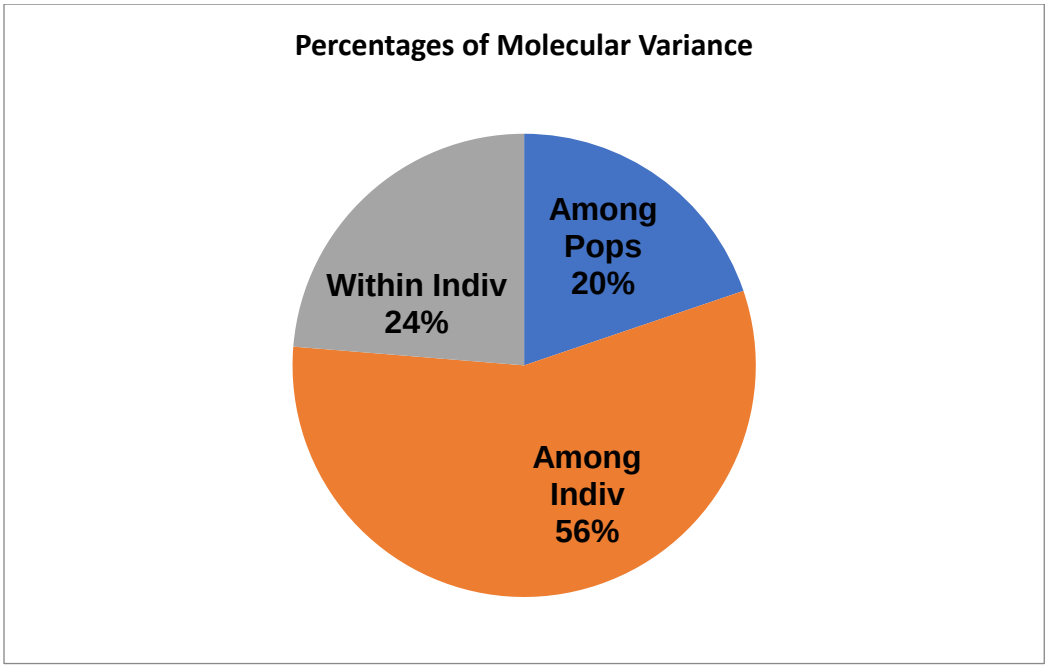


Figure 4. Analysis of Molecular variance (AMOVA) of 182 rice varieties based on GBS-SNP data. Abbrevations: indiv – individuals; pops – populations

Appendix 6

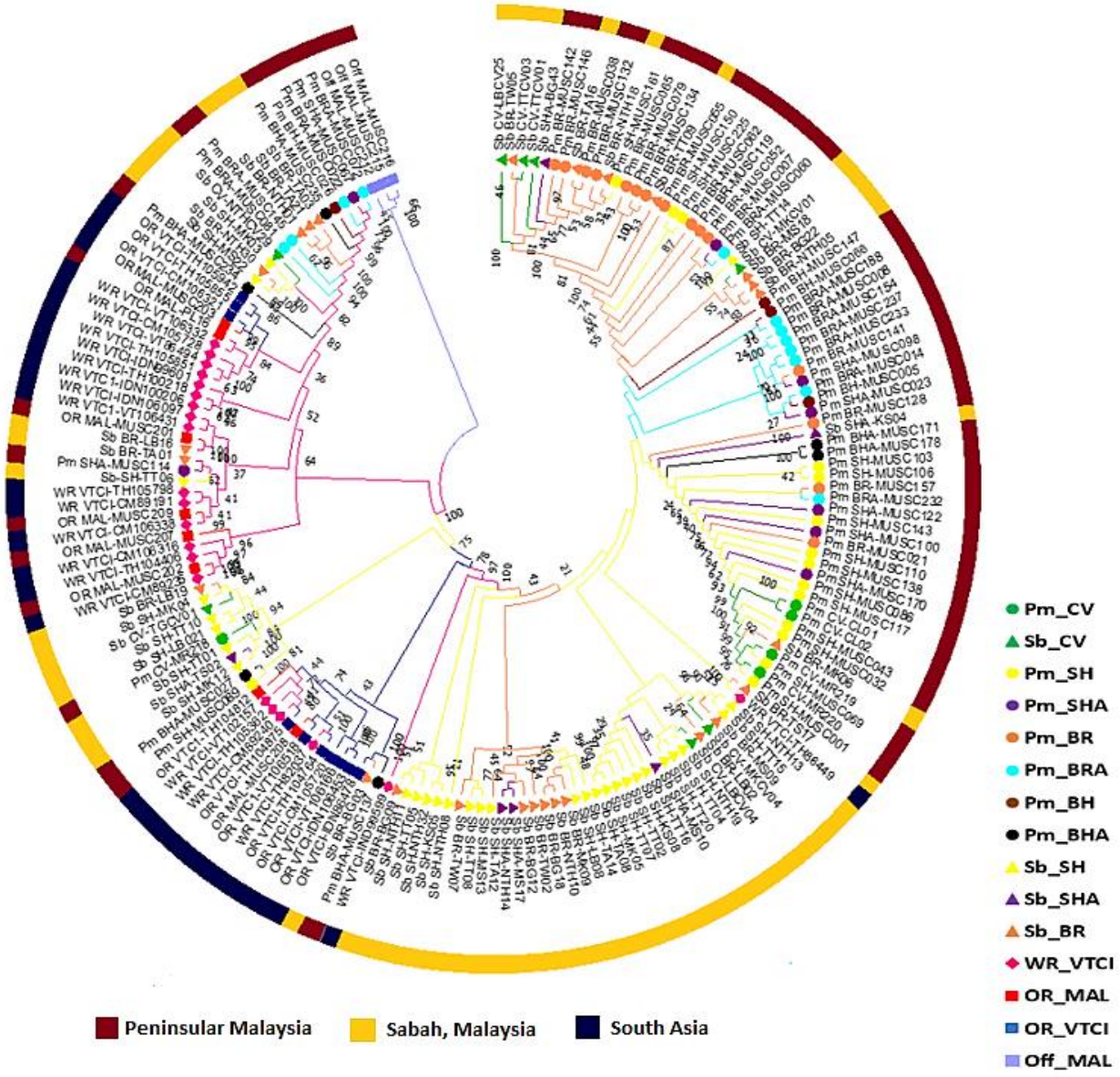


Figure 5. Evolutionary relationships of weedy rice by Maximum Likelihood (ML). The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analyzed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. Colours and taxa markers corresponds to PCoA analysis.

Appendix 7

Table 2. Tiller angle scoring and associated haplotypes of 52 Malaysian *Oryza* samples.

Accession ID	Hull morphology	Type	Region	Haplotypes	Tiller angle scoring (0= <40°, 1= 40°- 90°)
MR219	SH	Cultivar	Peninsular	1	0
MR220	SHA	Cultivar	Peninsular	1	0
NTHCV29	BR	Cultivar	Sabah	2	0
MKCV02	BRA	Cultivar	Sabah	1	0
MS06	SH	weedy	Sabah	1	1
TT01	SH	weedy	Sabah	1	0
TT05	SH	weedy	Sabah	1	0
MS22	SH	weedy	Sabah	1	0
MK03	SH	weedy	Sabah	1	0
NTH14	SHA	weedy	Sabah	1	0
NTH30	BR	weedy	Sabah	1	0
TG04	SH	weedy	Sabah	1	1
NTH18	BR	weedy	Sabah	1	0
BG22	BR	weedy	Sabah	1	0
MS09	BR	weedy	Sabah	1	1
NTH05	BR	weedy	Sabah	1	0
NT01	BR	weedy	Sabah	1	0
BG16	BRA	weedy	Sabah	1	1
LB21	SH	weedy	Sabah	2	1
TT06	SH	weedy	Sabah	2	0
TS02	SHA	weedy	Sabah	2	0
LB19	BR	weedy	Sabah	2	1
NTH10	BR	weedy	Sabah	2	1

BG24	BR	weedy	Sabah	2	1
BG18	BR	weedy	Sabah	2	1
MK09	BR	weedy	Sabah	2	0
LB24	BRA	weedy	Sabah	2	1
MUSC143	SH	weedy	Peninsular	1	0
MUSC138	SH	weedy	Peninsular	1	0
MUSC110	SH	weedy	Peninsular	1	0
MUSC079	BR	weedy	Peninsular	1	1
MUSC081	BRA	weedy	Peninsular	1	1
MUSC060	BRA	weedy	Peninsular	1	1
MUSC114	SHA	weedy	Peninsular	1	0
MUSC191	BHA	weedy	Peninsular	1	0
MUSC234	BHA	weedy	Peninsular	1	0
MUSC103	SH	weedy	Peninsular	2	0
MUSC086	SH	weedy	Peninsular	2	0
MUSC023	SHA	weedy	Peninsular	2	1
MUSC043	SH	weedy	Peninsular	2	0
MUSC055	BR	weedy	Peninsular	2	1
MUSC062	BR	weedy	Peninsular	2	0
MUSC128	BR	weedy	Peninsular	2	0
MUSC119	BR	weedy	Peninsular	2	0
MUSC154	BRA	weedy	Peninsular	2	0
MUSC188	BRA	weedy	Peninsular	2	1
MUSC237	BRA	weedy	Peninsular	2	0
MUSC026	BHA	weedy	Peninsular	3	0
Or201/IRGC105491		wild	Peninsular	1	0
PLWr02		wild	Peninsular	5	1
PLWr08		wild	Peninsular	5	1
PLWr16		wild	Peninsular	5	1

Appendix 8

Table 3. Primer pairs used for *TAC1* analysis in this study. All primers were taken from Yu et al. (2007).

Name	Forward 5' to 3'	Reverse 3' to 5'
***TAC1*_a**	GTACTGTCTGGCTTTCTCTTCTGGT	CCTCAGCCTTTTCTTCTTCG
***TAC1*_b**	TGCTCCGTGATGTGCTTATT	GTGTAAGATGTGGCGCACTG
***TAC1*_c**	AAGCCAACAAAACCAACGTC	GTGATTCGTGGGTCATGCTA
***TAC1*_d**	GGCCCTTGTGTTTTGCATAC	TTCCATTATCAGAGGCCAGAA
***TAC1*_e**	ATCTGAGCTGCCCCCTTATT	TTTGGTCCATCATAGCTCCA
***TAC1*_f**	AGTGGACACTAGACATGCAG	CAAGACACGTCAACTGTAGC

Appendix 9

Table 4. DNA variation of the *TAC1* region of weedy rice. The values below recorded after removal of peninsular weedy rice MU026.

$Region$	$Species$	N	L	h	π	hd	K	$Tajima's\ D$
Gene region (from start to stop codon)	Pm weedy	20	1277	0	0.00	0.00	0.00	n
	All weedy	43	1277	0	0.00	0.00	0.00	n
3' flanking UTR	Pm weedy	20	1673	2	0.00093	0.521	1.563	2.266*
	All weedy	43	1673	2	0.00091	0.529	1.528	2.578*
Full length *TAC1*	Pm weedy	20	2944	2	0.00053	0.521	1.563	2.266*
	All weedy	43	2944	2	0.00052	0.509	1.528	2.578*

Abbreviations: N: number of samples; L: length of the alignment in which all sequence contain bases, gaps excluded; h: number of haplotypes; π: nucleotide diversity; hd: haplotype diversity; K: average number of nucleotide difference; * : $P < 0.05$; n: no tajima's D values.