Genetic studies on resistance to *Rice tungro bacilliform virus* and evolution of its related endogenous viruses in the genus *Oryza*

(イネのツングロ病ウイルス RTBV 抵抗性およびオリザ属における
その内在性ウイルスの進化に関わる遺伝学的研究)

北海道大学　大学院農学院

生物資源科学専攻　博士後期課程

齋藤　希

TABLE OF CONTENTS

ABBREVIATIONS

AGO	Argonaute
CP	coat protein
DCL	Dicer-like protein
dsRNA	double-strand RNA
EPRV	endogenous pararetrovirus
eRTBVL	endogenous Rice tungro bacilliform virus like
EVE	endogenous virus element
IGR	intergenic region
miRNA	micro RNA
MP	movement protein
ORF	open reading frame
PR	protease
RTBV	Rice tungro bacilliform virus
RTD	Rice tungro disease
RT/RH	reverse transcriptase/RNase H
RTSV	Rice tungro spherical virus
siRNA	small interfering RNA
vsRNA	viral small RNA

LIST OF FIGURES

LIST OF TABLES

Chapter 1 General introduction

1.1 Rice tungro disease caused by two rice viruses

Rice, *Oryza sativa* L., is one of the important cereal crops that provide a staple food for half of the world's population. Rice viral diseases constitute a major obstacle to rice production and rice is known to be infected by more than 30 viruses, but 25 viruses do not directly have economic effects in terms of rice production (Abo and Sy, 1998). Rice tungro disease (RTD) negatively affects rice production in South Asia and Southeast Asia (Azzam and Chancellor, 2002). Dwarfing and leaf yellowing are the major symptoms of RTD (Azzam and Chancellor, 2002). *Rice tungro bacilliform virus* (RTBV) primarily causes RTD with the assistance of *Rice tungro spherical virus* (RTSV) (Hibino *et al*., 1987) (Figure 1.1). RTBV is one of the pararetroviruses belonging to the genus *Tungrovirus* in the *Caulimoviridae* family (Hull *et al*., 2005). RTBV has a circular double-stranded DNA genome of about 8 kb encapsidated in a bacilliform particle (Hay *et al*., 1991; Hull, 1996). RTSV is a member of the genus *Waikavirus* in the *Secoviridae* family (Sanfaçon *et al*., 2009; Thompson *et al*., 2017). RTSV has a single-stranded plus-sense RNA genome of about 12 kb encapsidated in a polyhedral particle (Shen *et al*., 1993; Hull 1996; Hull *et al*., 2005). The two viruses are transmitted by green leafhoppers (GLH) such as *Nephotettix virescens* Distant (Cabauatan *et al*., 1995; Hull 1996). Transmission of RTBV by GLH requires the presence of RTSV or a factor encoded in the RTSV genome (Hibino 1983; Hibino *et al*., 1987).

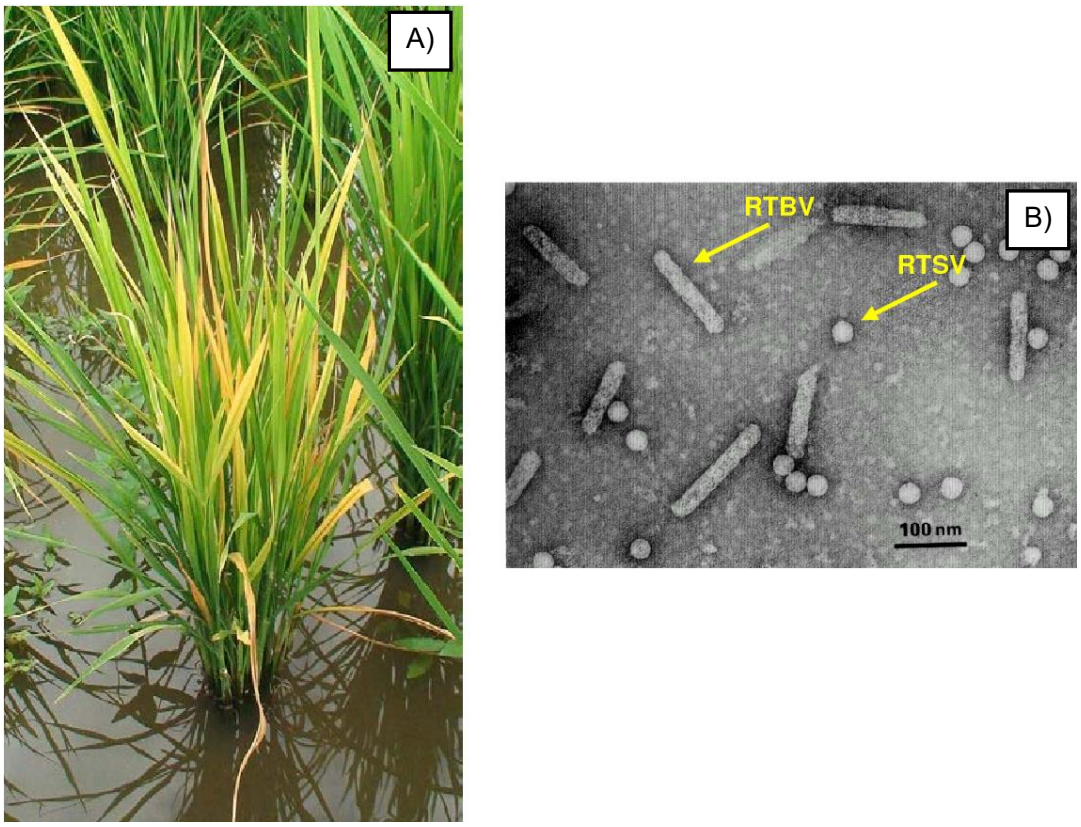


Figure 1.1 A) Rice tungro disease symptoms and B) *Rice tungro bacilliform virus* (RTBV) and *Rice tungro spherical virus* (RTSV) (adopted from Chen, 2015).

Different levels of resistance to RTD have been observed among rice cultivars (Encabo *et al*., 2009; Hibino *et al*., 1990; Sta Cruz *et al*., 2003). Utri Merah, which is a cultivar of the Asian rice species *Oryza sativa*, was reported to be one of the most resistant rice cultivars to RTD (Encabo *et al*., 2009). Majority of *O. sativa* plants show dwarfing and leaf yellowing symptoms when affected by RTD, however, the disease usually does not develop into necrosis or death (Hibino *et al*., 1990; Encabo *et al*., 2009). Contrastingly, RTD causes lethal symptoms such as necrosis in African rice *Oryza glaberrima* (Kobayashi and Ikeda 1992). It was also reported that *O. glaberrima* plants were more stunted than *O. sativa* plants when they were infected with RTSV alone (Budot *et al*., 2014).

1.2 Overview of anti-viral system in rice

Plants have developed multilayered defense mechanisms against microbial pathogens. Natural resistance against virus was mainly provided by two machineries, physical and physiological barriers and natural resistance (*R*) genes. For engineered resistance, gene-mediated resistance and RNA silencing-based defenses have been studied for crop protections (Chisholm *et al*., 2006; Jones and Dangl, 2006; Dong and Ronald, 2019; Kang *et al*., 2005; Wang and Krishnaswamy, 2012). RNA silencing, one of the major anti-viral strategies, can be exploited to confer viral immunity by targeting viral RNA for degradation (Rosa *et al*., 2018). This effect appeared to occur in post-transcriptional manners (post-transcriptional gene silencing; PTGS). The main actors of RNA silencing are small RNAs, which are classified into three main categories: short interfering RNAs (siRNAs), microRNAs (miRNAs) and Piwi-interacting RNAs (piRNAs). siRNAs and miRNAs are widely spread in animals, plants and insects and are characterized by the double-strand (ds) nature of their precursors (Carthew and Sontheimer, 2009). Dicer-like proteins (DCLs) process dsRNAs

with hairpin structures, producing mature small RNAs, which are loaded into Argonaute (AGO)-containing RNA induced silencing complexes (RISCs) to degrade their complementary targets (Figure 1.2). In antiviral silencing, viral small RNAs (vsRNAs) production drives degradation of viral RNAs in a sequence-specific manner (Hamilton and Baulcombe, 1999; Parent *et al*., 2015). AGO1 and AGO2 have been known to work for viral RNA degradation, while AGO4 is involved in viral DNA methylation (Mallory *et al*., 2004; Várallyay *et al*., 2010; Harvey *et al*., 2011; Wang *et al*., 2012; Carbonell and Carrington, 2015) (Figure 1.2A). In rice, a total of 19 members of *AGO* gene families have been identified (Zheng *et al*., 2019). For *Rice stripe virus* (RSV) and *Rice dwarf virus* (RDV), *AGO1* and *AGO18* are the most important for antiviral defense against virus disease (Wu *et al*., 2015). It is because plant virus infection induces the accumulation of AGO18, which sequester miR168 away from AGO1, to promote the accumulation of AGO1 for antiviral defense (Figure 1.2B) (Carbonell and Carrington, 2015). There are few reports about natural anti-viral mechanisms against RTBV infection, but engineered resistance against RTBV infection have not been successful to date (Rajeswaran *et al*., 2014; Zarreen *et al*., 2018).

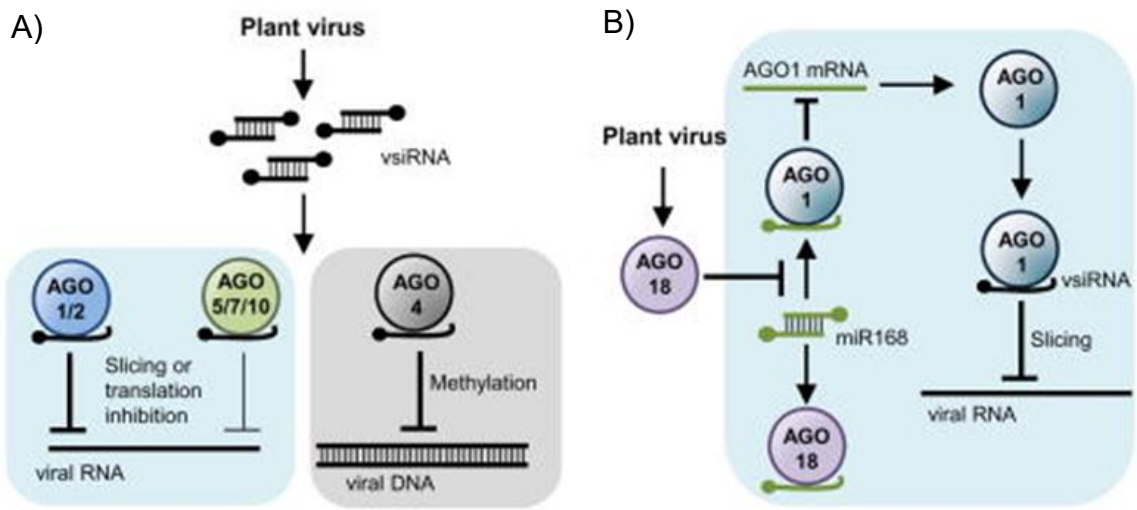


Figure 1.2 ARGONAUTE functions in plant antiviral RNA silencing (Carbonell and Carrington, 2015).

A) AGOs programmed with vsiRNAs repress viral RNA through slicing or translation inhibition in post-transcriptional gene silencing (PTGS), or through hypermethylation of viral DNA in transcriptional gene silencing (TGS). AGO1 and AGO2 are the main antiviral AGOs against RNA viruses, with AGO5, AGO7 and AGO10 having minor roles in some cases. AGO4 is the main antiviral AGO against DNA viruses. B) AGOs programmed with plant small RNAs promote antiviral defense. In rice, plant virus infection induces the accumulation of AGO18 which sequesters miR168 away from AGO1, and consequently induces AGO1 accumulation at elevated levels necessary for antiviral defense.

1.3 Endogenous viral element and eRTBVL sequence

The integration of viral sequences into the host germline genomes can accidentally occur during virus infection (Liu *et al*., 2012). Viral sequences that were integrated into the host genome are known as endogenous viral elements (EVEs), called "genomic fossil" (Feschotte and Gilbert, 2012). Most of the identified EVEs are derived from retroviruses since chromosomal integration were required for their replication (Feschotte and Gilbert, 2012). However, EVEs derived from other viruses than retroviruses were recently identified even though they do not have integration ability. One of the well-researched non-retrovirus EVEs is endogenous pararetrovirus (EPRV), which have been found in the genomes of various plants, such as tobacco, banana, petunia, rice, potato, and tomato (Jakowitsch *et al*., 1999; Ndowora *et al*., 1999; Richert-Pöggeler *et al*., 2003; Kunii *et al*., 2004; Hansen *et al*., 2005; Staginnus *et al*., 2007). In the rice genome, endogenous RTBV like (eRTBVL) sequences were identified (Kunii *et al*., 2004; Liu *et al* .2012). Based on the homology with RTBV at amino acid level, the viral genome structure of the reassembled eRTBVL consists of three ORFs (ORFx, y and z) and one intergenic region (IGR), but the reassembled eRTBVL lacks an ORF corresponding to ORF II in the RTBV genome (Figure 1.3). The longest ORFy encodes proteins that are predicted to be movement protein (MP), coat protein (CP), protease (PR) and reverse transcriptase/RNaseH (RT/RH) (Kunii *et al*., 2004) (Figure 1.3). eRTBVL were classified into five families (eRTBVL-A, B, C, D, and X) based on the amino acid sequence of RT/RH. eRTBVL-D family is known as an ancient family among them since the virus, which originated from eRTBVL-D, was integrated into rice genome before the speciation of *Oryza* AA-genome species (Chen *et al*., 2014) (Figure 1.4). eRTBVL-A family is divided into A1 and A2 subfamilies. The A1 subfamily was thought to be the initial genotype of this lineage (Chen *et al*., 2014). The recombination of the eRTBVL-B family

with an unknown lineage resulted in the viral genome of eRTBVL-A and -C. The virus sequences of eRTBVL-A2, which were generated from a recombination event between the viral genomes of eRTBVL-A1 and -C (Chen *et al*., 2014) (Figure 1.4).

Each of the genomes of *O. sativa* and *Oryza rufipogon* carry about 100 copies of the eRTBVL families, though their eRTBVL segments were of different proportions (Liu *et al* .2012; Chen *et al*., 2014). In contrast, only a few copies of eRTBVL have been found in the genome of African rice including *O. glaberrima* and *Oryza barthii* (Kunii *et al*., 2004) (Figure 1.5). The differences in eRTBVL copy number among *Oryza* AA-genome species have been considered as one of the reasons for disease response of the AA-genome species based on the previous analyses (Kunii *et al*., 2004; Saito *et al*., 2019). These eRTBVL families are considered to have arisen when *O. sativa* and *O. rufipogon* differentiated during the *Oryza*-AA genome speciation (Chen, 2015; Chen *et al*., 2014, 2018). Chen *et al*. (2018) discovered the oldest eRTBVL family, namely eRTBVL-D, from the genomes of all *Oryza* AA-genome species. The family of eRTBVL-D might allow us to trace the evolutionary processes of the local chromosomal regions with eRTBVL-D during the *Oryza*-AA genome speciation.

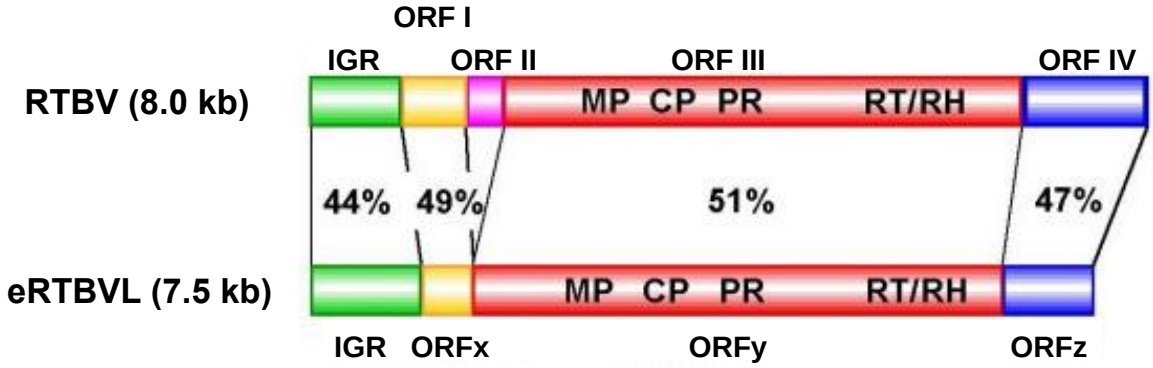


Figure 1.3 Genomic structure of RTBV and virus of eRTBVLs (modified from Kunii *et al.*, 2004).

Percentages indicate the nucleotide similarity of each of the corresponding regions or ORFs between RTBV and the virus of eRTBVLs. IGR, intergenic region; ORF, open reading frame; MP, movement protein; CP, coat protein; PR, protease; RT, reverse transcriptase; RH, RNase H.

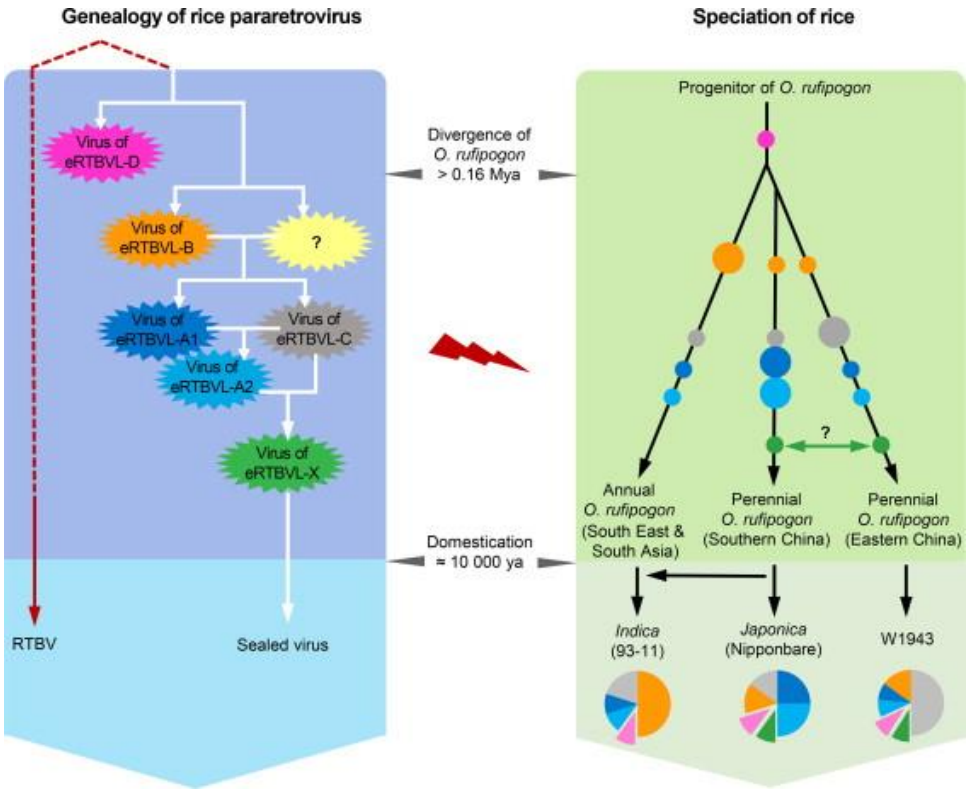


Figure 1.4 Schematic representation of possible integration events of eRTBVLs during rice and pararetrovirus evolution (adopted from Chen *et al.*, 2014).

The genealogy of virus lineages of eRTBVLs is summarized in the left panel using white lines. The oblong circles with jagged edge mark the occurrences of the various virus lineages. Among them, the yellow one represents the proposed unknown ancient virus lineage. The red dotted line indicates the unknown evolution of RTBV after its divergence from a common ancestor. The main processes of the rice speciation inferred from previous studies are summarized in the right panel (Huang *et al.*, 2012; Lu *et al.*, 2008; Zheng and Ge, 2010). The black lines indicate the time flow from the progenitor of *O. rufipogon* to the currently recognized accessions. Colored circles on the lines indicate the possible onset of endogenization; the colors correspond to the colors that indicate the virus lineages in the left panel. The size of the circle indicates the number of virus copies from these viral lineages that were integrated into the three rice genomes, reflecting the different geographic origin and prevalence of the virus lineages in different rice habitats. The abundances of the eRTBVL families shown in the pie charts at the bottom of the panel are not absolute. The timescales between the panels indicate two major branches in the rice speciation; Mya, million years ago.

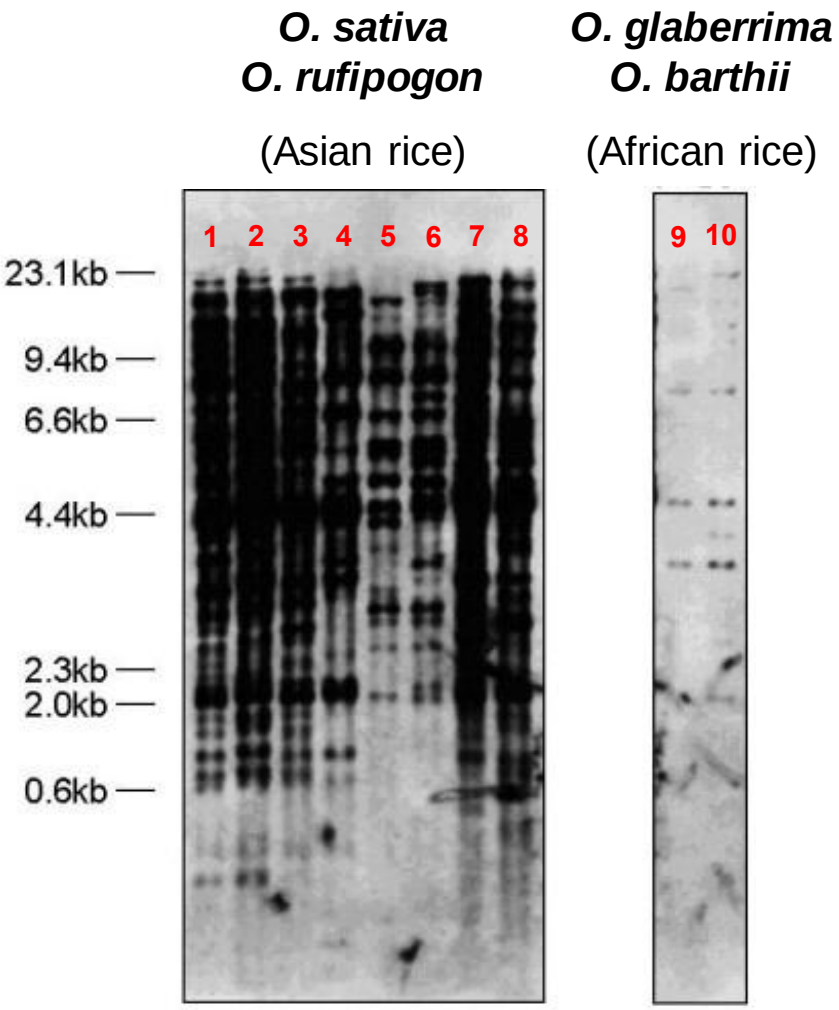


Figure 1.5 Detection of eRTBVL segments in the genomes of *Oryza* species by Southern blotting (modified from Kunii *et al.*, 2004).

Lanes 1-8 represent accessions or cultivars of *Oryza sativa* and *O. rufipogon*. Lanes 9 and 10 indicate accessions of *O. glaberrima* and *O. barthii*. Some accessions/lines of *O. sativa* and *O. rufipogon* are being used as the sources of resistance against RTD.

1.4 Objectives

In Chapter 2, I examined the vulnerability of *O. glaberrima* carrying a few eRTBVL copies and examined the relationship between the reaction to RTBV and the copy number of eRTBVL sequences by comparing between *O. sativa* and *O. glaberrima*. In addition, small RNA sequencing was performed to see whether the expression level of small RNA derived from eRTBVL were varied in the two species, which carry different number of eRTBVL copies.

In Chapter 3, I investigated how RTBV-resistant and -susceptible rice plants responded to RTBV through the phenotypic analysis and small RNA sequencing. Several candidate miRNAs and vsRNAs that might be associated with RTBV resistance were identified. The possible function of eRTBVL for anti-viral defense was also discussed in this chapter.

In Chapter 4, I analyzed for the evolution of eRTBVL-D family that is the oldest eRTBVL family that arose before *Oryza*-AA genome speciation. The eRTBVL-D sequences were available for the evolutionary study of the local-chromosomal regions. I described the functional implications of eRTBVL-D loci by evolutionary analyses for natural selection.

Chapter 2 Functional analysis of eRTBVL in the *O. sativa* and *O. glaberrima* genomes in relation to RTBV infection

2.1 Introduction

Different levels of resistance to RTD have been observed among rice cultivars (Encabo *et al*., 2009; Hibino *et al*., 1990; Sta Cruz *et al*., 2003). Utri Merah, which is a cultivar of the Asian rice species *O. sativa*, was reported to be one of the rice cultivars resistant to RTD (Encabo *et al*., 2009). The simultaneous infection with RTBV and RTSV caused whole plant necrosis in *O. glaberrima* but it did not induce any obvious necrosis in majority of *O. sativa* (Hibino *et al*., 1990; Kobayashi and Ikeda, 1992; Encabo *et al*., 2009). It was also reported that *O. glaberrima* plants were more stunted than *O. sativa* plants when they were infected with RTSV alone (Budot *et al*., 2014). Several *O. glaberrima* accessions were examined for reaction against infection with RTBV alone, but their vulnerability was not fully understood.

O. glaberrima has been adapted to the African environment where RTBV has not been reported. It was reported that *O. glaberrima* has gone through the domestication bottleneck, and shows significantly less genetic diversity than its progenitor, *O. barthii* (Li *et al*., 2011, Nabholz *et al*., 2014). These suggest that plants of *O. glaberrima* would be uniformly susceptible to the virus because they have not been exposed to the same selection pressures as *O. sativa* cultivars have been. To confirm this hypothesis, we examined the reactions of *O. glaberrima* plants to RTBV and compared their reactions with those of RTBV-susceptible *O. sativa* plants.

The genomic elements underlying the different reactions of *O. glaberrima* and *O. sativa* to RTBV are of interest as they could be associated with a common genetic background

in *Oryza* species tolerant to RTBV. Both *O. sativa* and *O. glaberrima* are AA-genome species and their genomic structures are generally similar (Wang *et al*., 2014). There are, however, many genomic differences between these two species (Wang *et al*., 2014). One genomic difference, which might be associated with their different reactions to RTBV, is the occurrence of eRTBVLs (Kunii *et al*., 2004; Liu *et al*., 2012; Chen and Kishima. 2016). The distribution of eRTBVLs differs between the two genomes; the *O. sativa* genome contains more than 100 copies of eRTBVLs, while the *O. glaberrima* genome has only a few copies (Kunii *et al*., 2004). Differences in the eRTBVL copy number in these genomes and the presence or lack of particular eRTBVLs might be associated with the different reactions of these two species to RTBV (Chen and Kishima, 2016). In this study, we performed small RNA-sequencing for mock-inoculated and RTBV-infected plants of *O. sativa* and *O. glaberrima* to examine the expression of the small RNAs derived from eRTBVL in the two species with and without RTBV infection.

2.2 Materials and Methods

2.2.1 Plant materials

The seeds of 18 accessions of *O. glaberrima* originating from four countries in Africa were obtained from the T. T. Chang Genetic Resources Center, International Rice Research Institute (IRRI) (Table 2.1). Two *O. sativa* ssp. *indica* varieties, Taichung Native 1 (TN1) and IR 64 were included as susceptible control materials (Encabo *et al*., 2009; Shibata *et al*., 2007). TN1 is known to be one of the most susceptible Asian rice plants to RTBV (Hibino *et al*., 1990, Shahjahan *et al*., 1990). All seeds were incubated at 28 ºC for 2 to 3 days in dark and wet conditions after removing the hull. After germination, the seeds were sown in pots and grown at 27–29 ºC at the containment facility of IRRI.

Table 2.1. Reactions of *Oryza glaberrima* and *O. sativa* plants to *Rice tungro bacilliform virus* (RTBV).

Species	Accession number[1] / variety	Country / region of origin	Number of plants examined	RTBV accumulation[2]	Leaf discoloration[3]	Height reduction rate (%)[4]
Oryza glaberrima	IRGC 86741	Nigeria	4	1.44 ± 0.12	+	49.8 ± 4.7
	IRGC 96718	Senegal	5	1.38 ± 0.23	nd / +	45.5 ± 6.6
	IRGC 96790	Nigeria	7	1.48 ± 0.11	+	31.5 ± 2.6
	IRGC 96793	Nigeria	7	1.07 ± 0.33	+	40.0 ± 4.9
	IRGC 96864	Liberia	3	1.16 ± 0.02	+	48.8 ± 4.8
	IRGC 96868	Liberia	3	1.45 ± 0.31	+	42.1 ± 2.3
	IRGC 100139	Guinea	4	1.33 ± 0.29	+	46.3 ± 6.4
	IRGC 100153	Guinea	3	1.21 ± 0.17	+	41.5 ± 5.9
	IRGC 102569	Liberia	4	0.37 ± 0.10	+	47.3 ± 14.1
	IRGC 102489	Liberia	3	0.72 ± 0.32	+	36.0 ± 8.8
	IRGC 102510	Liberia	3	0.95 ± 0.40	+	46.2 ± 6.7
	IRGC 102556	Liberia	3	1.28 ± 0.25	+	44.0 ± 6.1
	IRGC 103437	Senegal	3	0.86 ± 0.17	+	45.3 ± 4.5
	IRGC 103477	Senegal	3	0.67 ± 0.63	+	43.5 ± 3.1
	IRGC 104038	Senegal	7	1.29 ± 0.17	nd / +	38.5 ± 4.8
	IRGC 104545	Nigeria	8	1.28 ± 0.20	+	30.0 ± 4.1
	IRGC 104914	Nigeria	4	1.07 ± 0.28	+	34.3 ± 5.2
	IRGC 115633	Liberia	3	1.02 ± 0.17	+	49.0 ± 2.7
Oryza sativa	Taichung Native 1	Taiwan	7	1.31 ± 0.07	nd	32.8 ± 2.6
	IR64	Philippines	9	0.72 ± 0.13	nd	31.1 ± 2.7

[1]Accession number according to the International Rice Germplasm Collection (IRGC).

[2]Average ± standard deviation of the optical density at 405nm for RTBV in plant extracts determined by enzyme-linked immunosorbent assay at 21 days post-inoculation.

[3]Leaf yellowing was evaluated as either nd: no distinct leaf discoloration or +: yellow to yellow-orange leaf discoloration observed.

[4]Average ± standard deviation of height reduction rate (100 × [(height of mock-inoculated plant − height RTBV infected plant)/height of mock-inoculated plant]%).

2.2.2 Agroinoculation of RTBV

The plants were inoculated with RTBV via the agroinoculation method that has been used to evaluate the reactions of rice plants to RTBV (Sta Cruz *et al*., 1999, 2003; Dasgupta *et al*., 1991). Fourteen-day-old plants were inoculated with *Agrobacterium tumefaciens* GV3850 containing the RTBV infectious clone (qRTRB1162) (Dasgupta *et al*., 1991). Mock-treated plants were inoculated with *A. tumefaciens* carrying the empty vector (pBin19) (Sta Cruz *et al*., 2003). *Agrobacterium* carrying the RTBV infectious clone were cultured on Luria-Bertani agar (LBA) plates supplemented with kanamycin (50 µg/mL) and rifampicin (25 µg/mL) at 28ºC for 48 h; *Agrobacterium* used for mock inoculation was cultured on LBA plates with rifampicin. Bacterial cells were scraped from the plates and were resuspended in 50 mL LB broth with kanamycin (50 µg/mL) and rifampicin (25 µg/mL). The bacterial cells were cultured overnight in an incubator at 28ºC with shaking. Three mL of the overnight culture of bacterial cells was transferred into 250 mL LB broth with kanamycin (50 µg/mL) and rifampicin (25 µg/mL) that was then grown overnight in the same conditions. After overnight incubation, bacterial cells were harvested by centrifugation in a Beckman J-14 rotor at 5000 rpm for 15 minutes. The bacterial pellet was resuspended in 1.5 mL distilled water per 250 mL culture. Using a 1.0 mL disposable syringe with a 25-gauge (0.5 × 16 mm) needle, the bacterial suspension (200 µL) was injected into each plant at the base of the stem. After inoculation, plants were maintained under natural lighting conditions and temperatures ranging from 27 to 29ºC at the containment facility of IRRI.

2.2.3 Evaluation for RTBV accumulation

Enzyme-linked immunosorbent assay (ELISA). The presence of RTBV in plants was determined by ELISA with an RTBV-specific antibody (Bajet *et al*., 1986). The second

youngest leaf from each plant was collected at 14 and 21 days-post inoculation (dpi). Collected leaves were homogenized in phosphate buffered saline (0.02 M, pH 7.4)-Tween 20 (PBST, composition: 0.015 M NaCl, 0.735 mM KH_2PO_4, 0.01 M Na_2HPO_4, 0.0015 mM NaN_3 and 0.5 mL/L Tween 20) in 1:10 dilutions.

Polystyrene 96-well plates were sensitized with IgG for RTBV. A volume of 90 µL of coating buffer diluted with IgG (1.5 µL/mL) was applied to each well. The plates were incubated for four hours at 37°C or overnight at 4°C. In case of incubation at 37°C, the plates must be covered with damp paper towel and put into plastic bag to avoid the evaporation of liquid. After incubation, contents of the plates were discarded and plates were washed five times with by filling up each well with 1× PBST for five minutes each time. A volume of 90 µL of the leaf extracts were loaded into each well after wash. Plates containing the samples were incubated for four hours at 37°C or overnight at 4°C. After incubation, samples were removed, and the plates were washed as above. Afterwards, IgG-alkaline phosphatase in PBST was prepared and 90 µL of the solution was loaded into each well. Plates were incubated for three to four hours at 37°C or overnight at 4°C and were washed as above. Para-nitrophenol tablets (Sigma, USA cat. no. S0942) were dissolved in diethanolamine (10%, pH 9.8) to a final concentration of 1 mg/mL. A volume of 90 µL solution was loaded into each well. Plates were measured for optical density at 405 nm (OD_{405}) by an ELISA reader after incubation for 30 minutes at room temperature. A plant was considered to be infected with RTBV if the OD_{405} value obtained with the ELISA using the plant extract was higher than 0.1 (Shibata *et al*., 2007).

Quantitative PCR. The second or third youngest leaf was collected from the individual plants at 7, 14 and 21 dpi and quickly frozen in liquid nitrogen and kept at -70 to -80°C until use. Infection of RTBV was individually confirmed by ELISA prior to quantitative

PCR (qPCR). DNA was extracted using the cetyl trimethyl ammonium bromide (CTAB) method (Murray and Thompson, 1980). The qPCR was carried out using SYBR Master mix (Invitrogen, California, USA) and primers targeted at a region of the *ORF III* (1030-1200 bp) in the RTBV genome (forward primer: 5'-ATAACCCAAGATTCAACCTCTGA-3', reverse primer: 5'-AACTTTTTCTTGCCAATTTTTCTCA-3'). *Ubiquitin* (Gene ID: Os01g0328400/ LOC_Os01g22490.1) was used as the internal reference for normalization of each sample. Reactions were run on Mx3000P (Agilent Technologies, Wilmington, DE, USA) with the following profile: 95°C for 10 min, and 40 cycles of 95°C for 15 sec and 60°C for 60 sec. A total of 10 µL reaction mix containing 1 ng of DNA, forward and reverse primers to a final concentration of 200 nM, and 5 µL of 2× master mix were loaded into a 96-well plate. The amount of RTBV in each sample was analyzed by using the RTBV plasmid DNA (qRTRB1162 extracted from *Agrobacterium* containing the RTBV infectious clone) as a standard to determine the copy number of RTBV genomic sequence in each sample (Applied Biosystems, 2003).

2.2.4 Phenotyping

Percent reduction in plant height was calculated as 100 × [(height of mock-inoculated plant − height RTBV infected plant) / height of mock-inoculated plant]. Symptoms and plant heights were recorded at 7 to 21 days post-inoculation (dpi). Data were examined by an analysis of variance and the significance of differences among the means was examined by the least significant difference (LSD) test ($\alpha = 0.05$).

2.2.5 Small RNA sequencing and statistical analysis

An accession of *O. glaberrima*, CG 14 (IRGC accession # 96717), and a cultivar of *O. sativa* ssp. *indica*, IR 64, were selected for small RNA sequencing. IR 64 was included as susceptible control material (Shibata *et al*., 2007), and CG 14 was found out to be extremely susceptible to RTBV infection based on the result of phenotyping in this study. For RNA samples of RTBV-inoculated plants for small RNA sequencing, we used only those confirmed to be infected with RTBV. One RTBV-infected and one mock-inoculated plants of CG 14 and IR 64 were used as a sample for small-RNA sequencing.

The total RNA was extracted from the second youngest leaf of individual plants collected at 21 dpi using the TRIzol regent (Invitrogen, California, USA). After RNA isolation, individual sample of total RNA was treated with RQ1 RNase-Free DNase (Promega, Wisconsin, USA) following the manufacturer's instructions. The concentration and quality of total RNA were analyzed by NanoDrop2000 (Agilent Technologies, Wilmington, DE, USA) and gel electrophoresis, respectively. For sequencing, a cDNA library was constructed for each total RNA sample and then sequenced on the Illumina HiSeq2000 platform (Illumina, San Diago, CA) with 4 Gb output of sequence in 100bp paired ends by Hokkaido System Science Co., Ltd. (Sapporo, Japan).

The Illumina adapter of raw paired-end reads in fastq format was trimmed with *cutadapt*. The quality of all the raw reads without adapter sequences were evaluated and filtered by FaQCs (Lo and Chain, 2014) and PRINSEQ (Schmieder and Edwards, 2011) using the following condition: the average of quality score of all bases in read >20, length of read = 19-27 bp. All high-quality reads were aligned to the rice genome (IRGSP 1.0, ftp://ftp.gramene.org/pub/gramene/release-62/fasta/*Oryza_sativa*/dna/*Oryza_sativa*.IRGSP-1.0.dna.toplevel.fa.gz), the RTBV genomic sequence (NCBI accession: NC_001914.1) and

eRTBVL sequence (NCBI accession: BR000029.1, BR000030.1 and BR000031.1) using Bowtie (version 1.1.2, http://bowtie-bio.sourceforge.net/index.shtml/) with the following parameters: -S -a --best --strata -v 0-2. For the alignment of the clean reads against known rice miRNA, miRDeep-P was performed for the alignment of the clean reads against known rice miRNAs obtained from miRBase (Release 22, http://www.mirbase.org/) with following parameters: mapper.pl = -e -h -j -l 17 -m -p -s -t -v; quantifier.pl = -p -m -r (Yang and Li 2011; Yang *et al*., 2016).

2.3 Results

2.3.1 Comparison of phenotypes caused by RTBV infection between *O. sativa* and *O. glaberrima*

Based on the narrow genetic diversity of *O. glaberrima* (Li *et al*., 2011; Nabholz *et al*., 2014), the observation that most *O. glaberrima* accessions show necrosis upon infection with RTSV and RTBV (Kobayashi and Ikeda, 1992), and the fact that *O. glaberrima* plants originated from Africa where RTBV has not been reported, I hypothesized that *O. glaberrima* plants would be more vulnerable to RTBV than *O. sativa* plants. To determine whether *O. glaberrima* plants show common vulnerability to RTBV, 18 accessions of *O. glaberrima* and two *O. sativa* varieties, TN1 and IR 64, were examined for disease symptoms after RTBV inoculation (Table 2.1). TN1 was included because it is one of the *O. sativa* varieties that show the most severe symptoms with RTD (Cabauatan *et al*., 1995; Encabo *et al*., 2009). IR 64 was included because it is susceptible to RTD but still remains one of the most popular varieties in tropical Asia (Shibata *et al*., 2007). Three to eight infected plants per *O. glaberrima* accessions were examined for height reduction, leaf discoloration, and the accumulation of RTBV by ELISA at 21 dpi (Table 2.1). The average height reduction rates of

O. glaberrima accessions ranged from approximately 30% to 50% (Table 2.1). The height reduction rates for 15 of 18 accessions were more than 35%. Contrastingly, the average height reduction rates for TN1 and IR 64 were about 33% and 31%, respectively. All accessions of *O. glaberrima* showed yellow discoloration of leaves by RTBV infection, although the degree of discoloration varied among the accessions (Table 2.1). Contrastingly, evident discoloration of leaves was not observed in TN1 and IR 64 plants infected with RTBV. Therefore, the 18 *O. glaberrima* accessions examined appeared to be more sensitive to RTBV infection than *O. sativa* plants. The accumulation of RTBV was also examined in the 18 *O. glaberrima* accessions and the two varieties of *O. sativa* by ELISA. The average OD_{405} values for the extracts from the 18 *O. glaberrima* accessions at 21 dpi varied, ranged from 0.37 to 1.48 (Table 2.1). The average OD_{405} values for the extracts from 13 of 18 accessions were higher than 1.0 while those for the extracts from TN1 and IR 64 were 1.31 and 0.72, respectively. Therefore, it appeared that the levels of RTBV accumulation in majority of the *O. glaberrima* accessions were similar to the levels in TN1 that is the most susceptible *O. sativa* plant (Hibino *et al*., 1990; Shahjahan *et al*., 1990), and higher than the level in IR 64.

Based on the symptoms and RTBV accumulation levels observed in the 18 accessions of *O. glaberrima* (Table 2.1), we selected three accessions of *O. glaberrima* (IRGC accession numbers 96790, 96864, and 115633) for which we assessed more plants for height reduction and RTBV accumulation to confirm whether their reactions to RTBV were significantly different from those of the *O. sativa* plants. *O. glaberrima* IRGC 96790 was selected because it was the accession least stunted by RTBV infection (average height reduction rate of 31.5%) despite showing the highest level of RTBV accumulation (average OD_{405} value of 1.48). *O. glaberrima* IRGC 96864 and IRGC 115633 were selected as they were two of the most stunted accessions (average height reduction rates of 48.8% and 49.0%

for IRGC 96864 and 115633, respectively). Seventeen to twenty RTBV-infected plants for each of the three accessions of *O. glaberrima* were examined for their symptoms (Figure 2.1 and 2.2) and RTBV accumulation (Table 2.1). For comparison, two varieties of *O. sativa*, TN1 and IR 64 were also included in the evaluation. Consistent with the initial evaluation (Table 2.1), the three accessions of *O. glaberrima* were more stunted by RTBV infection than the *O. sativa* varieties (Figure 2.1), although the differences in height reduction rate were significant only between the *O. glaberrima* accessions (the height reduction rates of approximately 33 to 38%) and IR 64 (16.7%). Yellow to orange color discoloration of leaves was also observed in the three accessions of *O. glaberrima* by 21 dpi, but not in the two varieties of *O. sativa* (Figure 2.2). The levels of RTBV accumulation varied among the three accessions of *O. glaberrima*, and also among the two varieties of *O. sativa* (Table 2.2) and the difference in the level of RTBV accumulation was not evident between the two species. The levels of RTBV accumulation in *O. glaberrima* IRGC 96790 and 115633 were significantly higher than those in *O. glaberrima* IRGC 96864 and in *O. sativa* TN1 and IR 64 (Table 2.2). Collation of the data for height reduction (Figure 2.1) and RTBV accumulation (Table 2.2) indicated that the level of RTBV accumulation in the plants did not correlate with the severity of symptoms. The lack of correlation between the level of virus accumulation and the severity of symptoms was also observed when the reactions of *O. glaberrima* and *O. sativa* plants to RTSV were compared (Budot *et al*., 2014).

Due to the practical difficulty in inoculating a large number of plants by the RTBV agroinoculation method in a contained facility, the number of *O. glaberrima* accessions examined for the reactions to RTBV in this study was very limited. Therefore, to confirm whether the high susceptibility to RTBV is a typical reaction of *O. glaberrima* plants, it is desirable to examine more accessions of *O. glaberrima*. However, considering the previous

observations that most *O. glaberrima* accessions were hypersensitive to RTD (Kobayashi and Ikeda, 1992) and that *O. glaberrima* has a very narrow genetic diversity (Li *et al*., 2011; Nabholz *et al*., 2014), the results presented in this study presumably indicated that the *O. glaberrima* plants were generally more sensitive to RTBV infection than *O. sativa* plants.

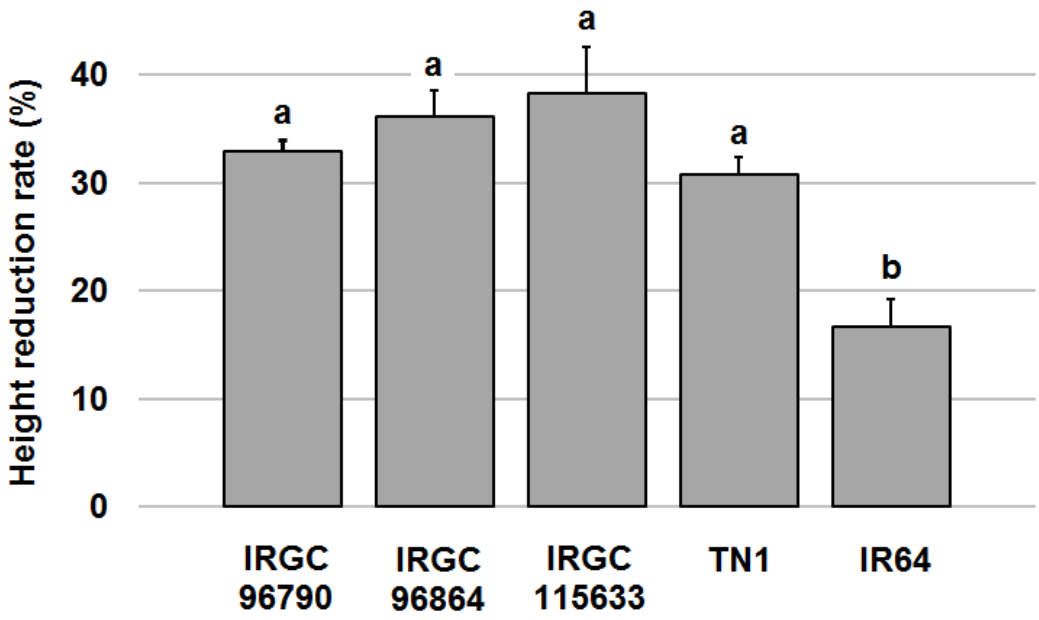


Figure 2.1 Height reduction rate for *Oryza glaberrima* and *O. sativa* plants infected with Rice tungro bacilliform virus (RTBV).

Bars with different letters are significantly different as determined by an LSD test ($\alpha = 0.05$). Vertical lines above the bars represent the standard deviation. TN1: Taichung Native 1.

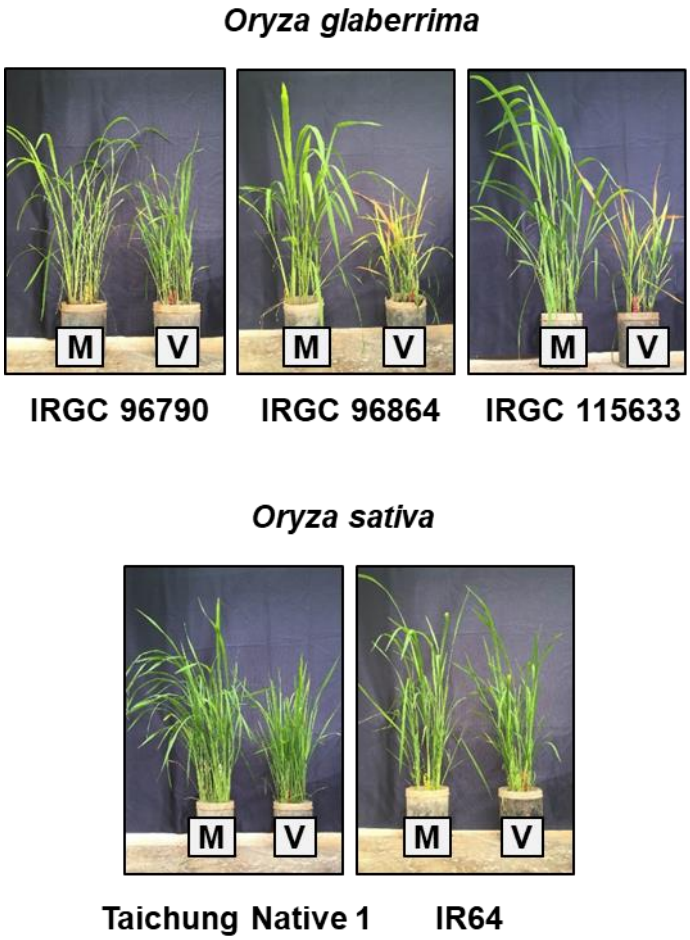


Figure 2.2 Representative symptoms in *Oryza glaberrima* and *O. sativa* plants infected with *Rice tungro bacilliform virus* (RTBV) at 21 days post-inoculation (dpi).

M: Mock-inoculated plant, V: Plant inoculated with RTBV.

Table 2.2 ***Rice tungro bacilliform virus*** **(RTBV) accumulation in** ***Oryza glaberrima*** **and** ***O. sativa*** **plants.**

Species	Accession number[1]/variety	Number of plants examined	Accumulation of RTBV at 21 days post-inoculaion[2]
Oryza glaberrima	IRGC 96790	20	1.12 ± 0.07^{c}
	IRGC96864	16	0.47 ± 0.04^{a}
	IRGC 115633	18	0.99 ± 0.06^{c}
Oryza sativa	Taichung Native 1	18	0.68 ± 0.06^{b}
	IR64	17	0.42 ± 0.05^{a}

[1]Accession number according to the International Rice Germplasm Collection (IRGC).

[2]Average ± standard deviation of OD_{405} for RTBV in plant extracts determined by an enzyme-linked immunosorbent assay at 21 days post-inoculation. Values followed by different letters are significantly different as determined by an LSD test ($\alpha = 0.05$).

2.3.2 Expression analysis of small RNAs

Four libraries of small RNAs derived from leaves of RTBV-infected or mock-inoculated plants of IR 64 or CG 14 were examined by sequencing. After removing adaptor sequences and low-quality reads (Q < 30) and selecting small RNAs or 19-27 nt in length, about 22-31 million clean reads were obtained for each library, representing 4,500,390 and 4,636,846 unique reads for RTBV- and mock-inoculated IR 64 respectively, and 6,577,620 and 5,733,538 unique reads for RTBV- and mock-inoculated CG 14, respectively (Table 2.3). The size classes of clean small RNA reads that were the most abundant in the four libraries were 21 and 24 nt (Figure 2.3A and B). The percentages of 21 and 24-nt reads in RTBV-infected plants decreased compared with mock-inoculated plants in IR 64, whereas the percentages of 21 and 24-nt reads in RTBV-infected CG 14 were higher than those in mock-inoculated CG 14 probably due to the RTBV infection (Figure 2.3C). The abundance of 21 and 24-nt small RNA in IR 64 and CG 14 were common in most plants (Axtell, 2013) (Figure 2.3 A, B). The prevalent sizes of RTBV vsRNA in the infected samples were 21 nt and 22 nt. The three abundant classes of eRTBVL small RNA were consistent with that of vsRNA, which were 21 nt, followed by 22 and 24 nt, and were not affected by the RTBV infection (Figure 2.3C). The frequencies of small RNAs of different lengths (21, 22 and 24 nt) derived from RTBV and eRTBVL were almost common in IR 64 and CG 14 (Figure 2.3C, D).

All the clean reads from IR 64 and CG 14 were mapped to the *O. sativa* genome (IRGSP 1.0) and the *O. glaberrima* genome (v1.0), respectively. The reads from each sample were mapped onto the genomes of *O. sativa* and *O. glaberrima*; 72% of the total reads from the library of RTBV-infected IR 64, 60% of the total reads from the library of RTBV-infected CG 14 reads, and 76% of the reads from each library of mock-inoculated samples (Table 2.3). The reads of endogenous small RNAs were the most abundant in all four samples, and in the

infected samples. The vsRNA reads from the RTBV genome accounted for approximately 10 and 13% of the total reads of RTBV-infected IR 64 and CG 14, respectively (Table 2.3). I also aligned all the clean reads with eRTBVL-A, B and C reconstructed sequences without allowing any mismatches (Table 2.3). The mapped reads with eRTBVL sequences were mapped to RTBV genome to confirm if the reads were not derived from RTBV genome. I removed the common small RNA reads, which were perfectly mapped to both eRTBVL and RTBV sequences, from the small RNA analyses. The accumulation of eRTBVL small RNA was observed in IR64, which carries approximately 100 copies of eRTBVL segments in its genome. The small RNA from eRTBVL in IR 64 accounted for 26,228 reads (0.11%) and 21,335 reads (0.08%) in the RTBV-infected sample and the mock-inoculated IR 64 plants, respectively. The small numbers of reads from the CG 14 samples that were mapped on eRTBVL were only 91 for the RTBV-infected sample and 16 from the mock-inoculated sample. These results suggested that the accumulation of eRTBVL small RNAs was dependent on the copy number of eRTBVL in the genomes. Considering the mapping percentage, the expression of small RNAs derived from eRTBVL was not affected by RTBV infection because the expression level of eRTBVL-derived small RNAs was similar between mock- and RTBV-inoculated IR 64 plants and much lower than that of vsRNAs.

Table 2.3 Summary of small RNAs from RTBV-infected and mock-inoculated IR 64 and CG 14.

	RTBV-infected				Mock-inoculated			
	IR 64		CG14		IR 64		CG14	
	Reads	Uniques	Reads	Uniques	Reads	Uniques	Reads	Uniques
Clean reads	22,903,918	4,500,390	31,174,326	6,577,620	25,206,525	4,636,846	24,483,068	5,733,538
Mapped to rice genome[1]	16,385,569 (71.54%)	2,274,677 (50.54%)	18,570,285 (59.57%)	4,191,585 (63.72%)	19,253,518 (76.38%)	3,027,742 (65.30%)	18,521,095 (75.65%)	3,138,426 (54.74%)
Mapped to virus geome	2,352,644 (10.27%)	58,147 (1.29%)	4,034,418 (12.94%)	67,530 (1.03%)	206 (0.00%)	188 (0.00%)	221 (0.00%)	184 (0.00%)
Mapped to eRTBVL	26,228 (0.11%)	4,067 (0.09%)	91 (0.00%)	55 (0.00%)	21,335 (0.08%)	3,706 (0.08%)	16 (0.00%)	7 (0.00%)

[1]The versions of rice genomic database used for mapping were IRGSP 1.0 for IR 64 and *Oryza_glaberrima*_V1 for CG 14.*Oryzaglaberrima*

[2]Percentage of mapping reads = Mapped reads (reads or uniques)/Clean reads (reads or uniques) × 100.

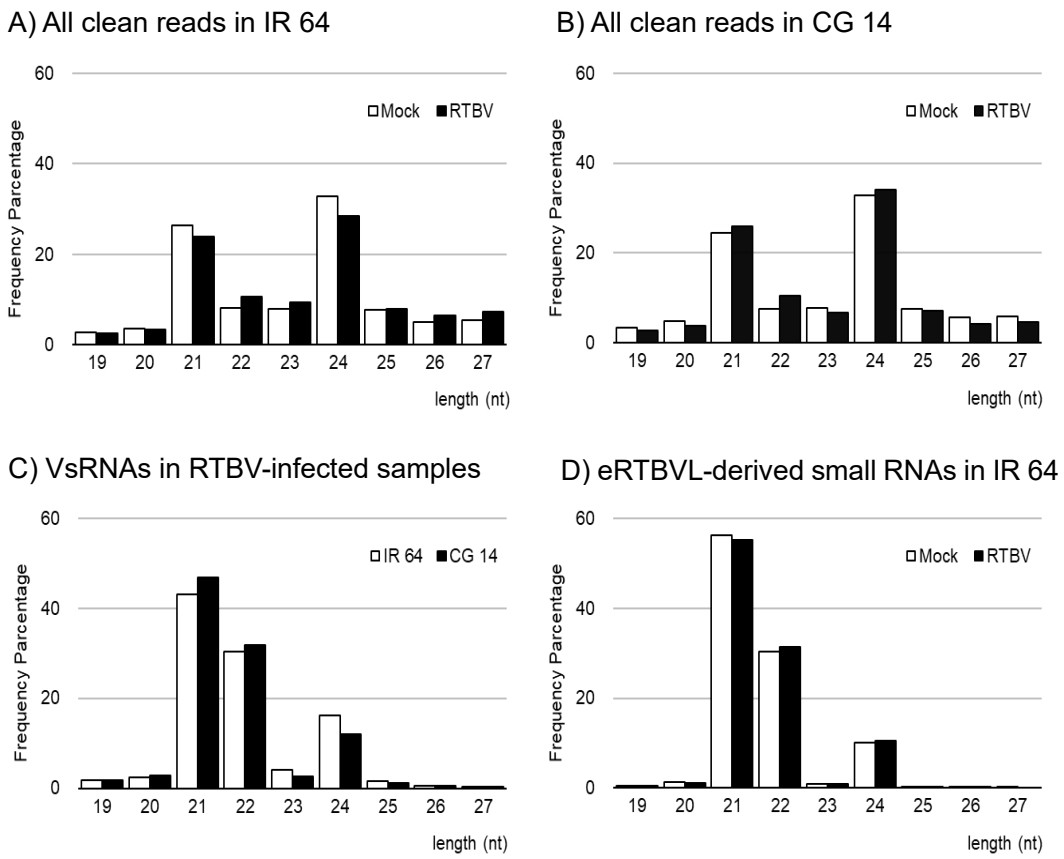


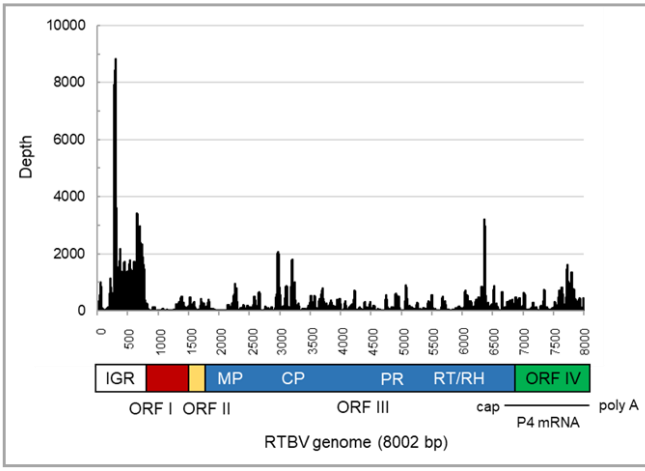
Figure 2.3 The nucleotide length distribution of small RNAs from RTBV-infected and mick-inoculated IR 64 and CG 14.

Bars indicate the frequency percentage for the number of reads in respective length. A) all clean reads of IR 64, B) all clean reads of CG 14, C) small RNAs mapped to the RTBV genome in infected samples, D) eRTBVL-derived small RNAs in IR 64.

2.3.3 Expression patterns of RTBV- and eRTBVL-derived small RNAs

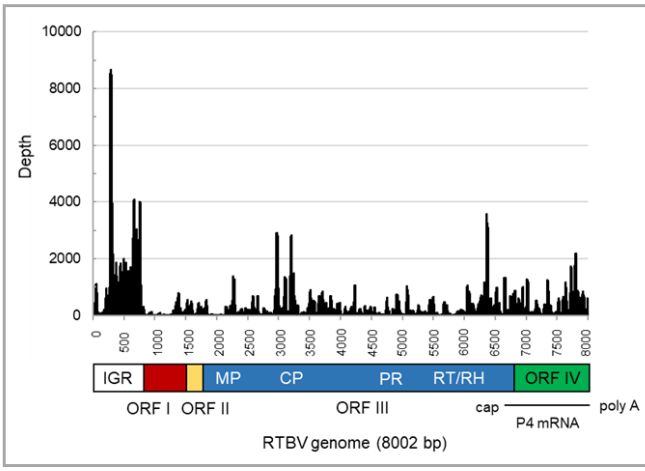
The distribution of vsRNAs along the RTBV genome is shown in Figure 2.4A and B. vsRNAs from the two plants covered over 99.5% of the viral genome with the largest hotspot located at IGR. This result was similar to the previous reports that the vsRNA accumulation was restricted only in the non-coding leader region, suggesting a strategy of the virus, namely "decoy" to maintain the virus defense machinery such as TGS and PTGS far away from coding region (Rajeswaran *et al*., 2014; Zarreen *et al*., 2018). The expression level of vsRNA was relatively higher around IGR, CP and RT/RH regions of ORF III, and ORF IV region in the RTBV genome (Figure 2.4). Depth for vsRNA, which represented the mapped number of nucleotide on each region in RTBV genome, was the highest at IGR (929,709 in IR 64 and 1,041,662 in CG 14), followed by ORF IV (313,276 in IR 64 and 475,838 in CG 14), RT/RH (275,625 in IR 64 and 347,986 in CG 14) and ORF III (935,910 in IR 64 and 347,986 in CG 14) (Figure 2.4). These patterns seem to be similar between IR 64 and CG 14 (Figure 2.4A, B). The expression level of eRTBVL-derived small RNAs in RTBV-infected IR 64 was a little higher than those in mock-inoculated IR64 (Figure 2.5). However, the read numbers of small RNA from eRTBVL was about 1/60 of that of vsRNA (Figure 2.4, 2.5). Similar to RTBV-derived small RNAs, the small RNAs of eRTBVL were also intensively localized at IGR and observed around ORFz in both libraries of IR64, indicating that the hotspots of the mapped small RNA were common to each other (Figure 2.4, 2.5). According to the distribution of small RNAs mapped to eRTBVL sequences, the locations and the expression levels of small RNAs derived from eRTBVL were similar between the RTBV-infected and the mock-inoculated IR 64.

A) VsRNA in RTBV-infected IR 64



	Depth	Average of depth
IGR	929,709	1,082
ORF I	48,057	80
ORF II	46,834	136
ORF III	935,910	186
MP	61,095	126
CP	174,562	185
PR	46,501	129
RT/RH	275,625	227
ORF IV	313,276	247

B) VsRNA in RTBV-infected CG 14



	Depth	Average of depth
IGR	1,041,662	1,213
ORF I	53,466	89
ORF II	54,938	159
ORF III	1,234,057	245
MP	76,010	157
CP	230,947	244
PR	54,835	152
RT/RH	347,986	287
ORF IV	475,838	376

Figure 2.4 Distribution of small RNA abundance along the RTBV genomes in A) IR 64 and B) CG 14.

A) vsRNA in IR 64 infected with RTBV, B) vsRNA in CG 14 infected with RTBV. Black bars indicated the depth counts which were mapped to the position on RTBV genome. All the depth counts were normalized by using read per million (RPM) for each sample following the formula: [depth counts/the total number of clean reads × 1,000,000]. X-axis represents RTBV genomic sequence (8002 bp, NCBI accession number NC_001914.1) and the color box below the bar graph indicates the structures and ORFs of RTBV genome. The table at the right side represents the actual number of depth counts which was normalized using RPM on each ORFs and motif. Average of depth was calculated by (Depth of the region / the length of the region).

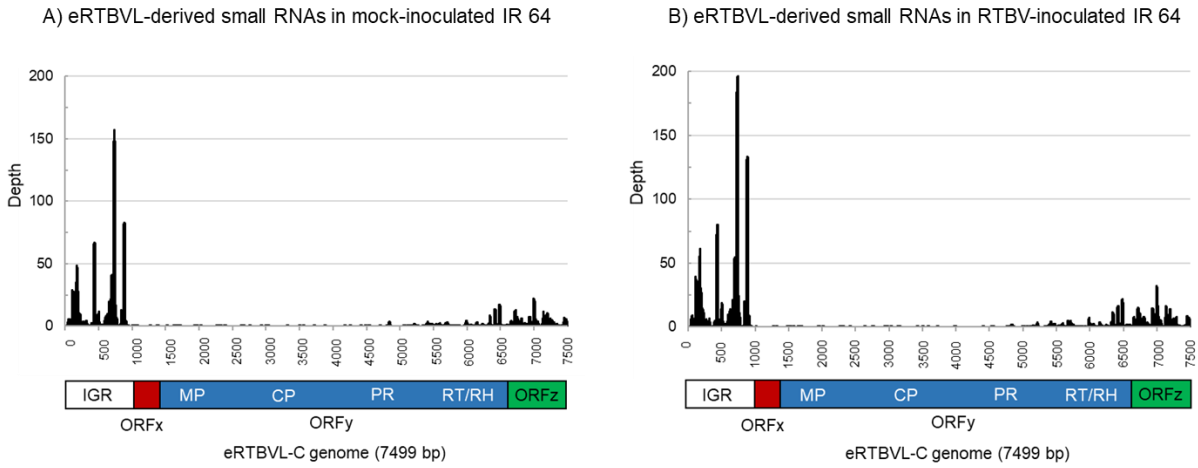


Figure 2.5 Distribution of small RNA abundance along eRTBVL genomes A) mock- and B) RTBV- inoculated IR 64.

A) eRTBVL-derived small RNAs in IR 64 inoculated with mock, B) eRTBVL-derived small RNAs in IR 64 infected with RTBV. Black bars indicate the depth counts which were mapped to the position on eRTBVL sequence. All the depth counts were normalized using RPM for each sample following the formula: [depth counts/the total number of clean reads × 1,000,000]. X-axis represents eRTBVL genomic sequence (7499 bp, NCBI accession number BR000031.1) and the color box below the bar graph indicates the structures and ORFs of RTBV genome.

2.4 Discussion

The symptom and the expression of small RNAs during RTBV infection were compared between *O. sativa* and *O. glaberrima* to predict the relationship between the reaction to RTBV and the copy number of eRTBVL. I selected IR 64 from *O. sativa* and CG 14 from *O. glaberrima* and examined the expressions of small RNAs derived from IR 64 and CG 14 which were inoculated with mock and RTBV. In terms of length distribution of small RNAs, the accumulation of 21 nt and 22 nt vsRNA of the RTBV-infected IR 64 plants was higher than that of the RTBV-infected CG 14 plants. This characteristic of vsRNA was closely similar to that of *Pagoda yellow mosaic associated virus* (PYMAV) and *Banana streak virus* (BSV), and the accumulation of 21- and 22-nt vsRNA was reported to be caused by the role of homologues of Dicer-like protein 4 (DCL4) and DCL2 in mounting defense response against PYMAV and BSV (Wang *et al*., 2014; Rajeswaran *et al*., 2014; Zarreen *et al*., 2018). The result suggested that the RTBV-infected IR 64 plants might produce more 21- 22-nt vsRNA for defense response against RTBV. I confirmed the constitutive expression of small RNA from eRTBVL sequences in IR 64, whereas a little expression of small RNA from eRTBVL was detected from CG14. The expression level of vsRNA seems to be similar between the RTBV-infected plants of IR 64 and CG 14. According to these results, the expression of small RNAs from eRTBVL might be associated with the phenotypic differences of height reduction rate and RTBV copy number between IR64 and CG14 after RTBV infection. However, the expression of small RNAs derived from eRTBVL was not dramatically up-regulated due to RTBV infection. Therefore, eRTBVLs might have an important role to mitigate the more severe symptom of RTBV by constant expressions of eRTBVL-derived small RNAs, although they might not contribute to strong resistance against RTBV.

Chapter 3 Occurrence of small RNAs related to RTBV resistance in *O. sativa*

3.1 Introduction

Virus infection affects the expression of host gene. Gene expression changes upon various virus infection as shown by using microarrays based on studies on responses at the gene expression level against viruses (Dardick, 2007; Senthil *et al*., 2005; Havelda *et al*., 2008; Babu *et al*., 2008; Agudelo-Romero *et al*., 2008; Catoni, *et al*., 2009; Shimizu *et al*., 2007; Satoh *et al*., 2010). According to these results, there was a similarity in different plant-virus interactions in terms of gene responses to virus infections, which were associated with development and morphogenesis of host plants (Dardick *et al*., 2007; Senthil *et al*., 2005). Disease symptoms caused by virus infection showed that the normal growth of plants was suppressed. Host genes affected by virus infection vary depending on plant species and tissues, and virus species (Senthil *et al*., 2005; Havelda *et al*., 2008; Babu *et al*., 2008; Agudelo-Romero *et al*., 2008; Catoni *et al*., 2009).

Recently, it has been reported that pathogen infection altered not only the expression of host genes, but also the expression of the endogenous microRNAs (miRNAs) (Wu *et al*., 2014; Zhao *et al*., 2015). For instance, a total of 69 differentially expressed miRNAs (56 up-regulated and 13 down-regulated) were found to be differentially expressed in rice due to the infection with Rice stripe virus (Yang *et al*., 2016). miRNAs play important roles in negative gene regulation in plants and animals via mRNA cleavage or translational inhibition of their targets (Hannon, 2002; Baulcombe, 2004; Vaucheret, 2006). In this study, I characterized vsRNA and rice-derived miRNA by Illumina deep sequencing and compared the expression of small RNAs between resistant plants and susceptible plants in order to examine the RNAi-based defense machinery in rice during RTBV infection. RTBV has been known to arm the

machinery for anti-silencing against rice plants, while the host defense mechanism on RTBV has not been revealed yet (Rajeswaran *et al*., 2014). This study revealed several responses of the host plant to RTBV infection by examining small RNAs which might be associated with RNAi-based defense.

3.2 Materials and Methods

3.2.1 Plant materials

TW 16 (BC_5F_8) was developed by backcrossing between the RTBV-resistant donor rice cultivar Utri Merah (IRGC accession number 16682) and the RTBV-susceptible recurrent parent TN1 (Encabo *et al*., 2009) (Figure 3.1).

3.2.2 Agroinoculation analysis

Rice plants were inoculated with RTBV by the Agroinoculation method (Dasgupta *et al*., 1991; Sta Cruz *et al*., 1999, 2003). The details of method were described in 2.2.2.

3.2.3 Evaluation for RTBV accumulation

The presence of RTBV in individual plants was examined by ELISA (see in 2.2.3). DNA was extracted from the second youngest leaf from TN1 and TW16 using the CTAB method as described in 2.2.3. After the DNA extraction from individual plants, DNA samples from three individual plants were combined into one. Three pooled samples were prepared for each of mock- and RTBV-inoculated TN1 and TW16 at each sampling time point (5, 10, 15, and 20 dpi). The copy number of RTBV in the plants were estimated by qPCR as described in 2.2.3.

3.2.4 Phenotyping

Ten to fifteen infected plants and three to five mock-inoculated plants per each genotype were examined at a replicate for phenotyping. Plant height reduction (%) was calculated as 100 × [(height of mock-inoculated plant - height RTBV infected plant)/height of mock-inoculated plant]. Plant heights were measured at 5, 10, 15 and 20 dpi. Differences in plant height reduction were examined by the least significant difference (LSD) test (α = 0.05). Evaluation for plant height reduction was repeated three times with 2 to 5 days intervals between the repeated experiments.

Plant weight (aboveground parts, PW, g), grain weight (GW, g) and stem and leaf weight (aboveground parts, SLW, g) were evaluated as the biomass traits. For each measurement, five mature plants were bulked, dried for about a week at 27-29ºC at the containment facility of IRRI and the samples including seeds were transferred to the drying rack and store there for two weeks, and then dried weight was measured. The seeds and vegetative parts above the ground of matured plants were evaluated for: harvest index (HI, %), panicle number (PN), spikelet number per panicle (SN), 1000-grain weight (1000GW, g), spikelet fertility (SF, %). HI was calculated as GW/PW × 100. Hundred-grain weight was measured with five replicates from each genotype and then the mean was converted to 1000-grain weight. For PN, SN, and SF, the values from five plants per genotype were used to compute means for statistical analysis. Differences in all the traits between mock- and RTBV-inoculated plants of each genotype were examined by t-test at the 95% confidence level.

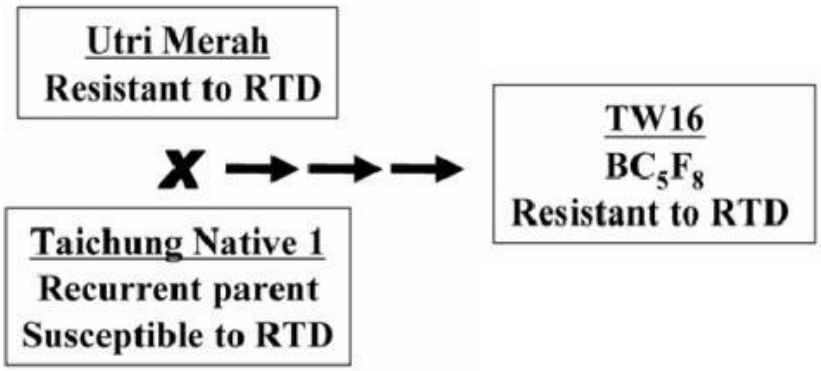


Figure 3.1 TW16 and their parent cultivars used in this study (modified from Encabo *et al*., 2009).

3.2.5 Small RNA sequencing and statistical analysis

Total RNA samples, which were extracted from mock-inoculated and RTBV-infected TN1 and TW16 were sequenced. The second youngest leaves were collected at 10 dpi from individual plants and RNA was extracted by the TRIzol reagent (Invitrogen, California, USA). The method of RNA extraction and quality check were mentioned in 2.2.5. Total RNA samples from three individual plants of the same genotype and treatment were bulked and submitted to Macrogen Japan (Kyoto, Japan) for small RNA sequencing. After performing quality control, the sequencing library was prepared by random fragmentation of cDNA sample, followed by 5' and 3' adapter ligation. Adapter-ligated fragments were PCR amplified and gel purified, and then sequencing was performed. Small RNA sequencing was done by Illumina HiSeq 2500 with 4 Gb output of sequence and 50 bp single-end reads were piled up as a fastq format. The steps from library construction to small RNA sequencing were done by Macrogen Japan (Kyoto, Japan).

In silico cleaning sequence reads was done as described in 2.2.5. All the clean reads were mapped to the RTBV genomic sequence and the rice genome (IRGSP 1.0) without allowing any mismatches by using Bowtie (http://bowtie-bio.sourceforge.net/index.shtml/). For identification of miRNA, the reads were mapped on the rice genome, and were aligned with noncoding RNA sequences, which were collected from the Rfam database (http://www.sanger.ac.uk/Software/Rfam/) and Gramene (http://www.gramene.org/), without allowing any mismatches by Bowtie. The alignments of all clean reads with the known rice mature miRNA and precursor miRNA (miRbase, version 22, http://www.mirbase.org/) were achieved by miRDeep-P with the parameters: mapper.pl = -e -h -j -l 17 -m -p -s -t -v; quantifier.pl = -p -m -r (Yang and Li, 2011; Yang *et al*., 2016).

Analysis for differentially expressed miRNAs. To compare the expressions of small RNAs in the RTBV-infected plants and the mock-inoculated plants, R (version 3.6.1) and the R package edgeR (https://bioconductor.org/packages/release/bioc/html/edgeR.html) were employed. I considered DE miRNA if its edgeR FDR-adjusted P value was <0.05.

3.2.6 RT-qPCR for expression analysis

Gene expression analysis. The expressions of *AGO* genes were quantitatively analyzed by RT-qPCR method as follows. Synthesis of cDNA from total RNA was performed with RT primer mix using the Quantitect reverse transcription kit (Qiagen, Hilden, Germany). A total of 1 µg of total RNA was mixed with 2 µL of 7× gDNA Wipeout buffer and nuclease-free water was added up to 14.5 mL. Then the solution containing total RNA was incubated at 42ºC for 2 min. After incubation, 4 µL of 5× Quantiscript RT buffer, 0.5 µL of the Quantiscript reverse transcriptase and 1 µL of RT primer mix was added into RNA solution. The PCR profile for cDNA synthesis followed the manufacturer's instruction. RT-qPCR was conducted according to the following procedures; each 10 µL volume of reaction contained 5 µL of 2× PowerUP SYBR Green master mix (Applied Biosystems, California, USA), 0.2 µL each of 10 µM forward and reverse primers (Supplementary Table S1), 1 µL of RT product, and 3.6 µL of nuclease-free water. qPCR was performed on the CFX-Connect Real-time System (BIO-RAD, California, USA) with the program as shown: 95ºC for 10 min, followed by 40 cycles of 95ºC for 15 sec, 60ºC for 1 min. The melting curves (60 to 95ºC) were analyzed to confirm the specificity of PCR products. The primer used in this study were indicated in Supplementary Table S1 and showed 90-110% of PCR efficiency in the analysis of their standard curve.

Expression analysis of small RNA from eRTBVL. The expression of eRTBVL-derived small RNA was compared among three genotypes. One of the small RNAs highly expressed from IGR was selected for Real-time quantitation. Total RNA was extracted with TRIzol as mentioned in 3.2.5. Total RNA from three individual plants were bulked and a biological replicate consisted of three bulked RNA samples, which contained 9 individual plants. All the RNA samples used for the quantification were treated with DNase (see in 2.2.5). According to the modified method described by Chen (2005), RNA reverse transcription was performed with specific stem-loop primer using SuperScript III first strand synthesis system for RT-qPCR (Invitrogen, Carlsbad, CA, USA). The stem-loop reverse transcription contained 4μg of total RNA, 2 μM small RNA-specific stem-loop primer, 1 × RT buffer, 0.25 mM dNTP, 40 U/μL RNaseOUT and 200U/μL Superscript III RT (Invitrogen, Carlsbad, CA, USA). The 20 μL reaction volume was incubated for 30 min at 16ºC, followed by pulsed RT of 60 pulsed cycles of 30ºC for 30 sec, 42ºC for 30 sec and 50ºC for 1 sec in G-STORM (Genetic Technologies, Australia). The reverse transcriptase was inactivated by incubation at 85ºC for 5 min. qPCR was performed with 10 μL volume of reaction containing 5 μL of 2× Power SYBR Green master mix, 0.2 μL each of 10 μM forward and universal reverse primers, 1 μL of RT product, and 3.6 μL of nuclease-free water. qPCR was conducted with the StepOne Plus Real-time PCR system (Applied Biosystems, California, USA) using the following program: 95ºC for 10 min, followed by 45 cycles of 95ºC for 15 sec, 60ºC for 30 min and 72ºC for 45 sec. All the primers used in this study were indicated in Supplementary Table S1.

3.3 Results

3.3.1 Phenotypes caused by RTBV infection

Due to RTBV infection, the three genotypes showed phenotypically different reactions, TN1 susceptible to RTBV, TW16 and Utri Merah resistant to RTBV and RTD. The height reduction was monitored until 20 dpi to characterize the stunting phenotypes of the infected plants (Table 3.1, Figure 3.2). All the RTBV-inoculated plants in this study were confirmed to be infected with RTBV by ELISA. The reduction rates of TN1, TW16 and Utri Merah were approximately 33%, 22% and 11% at 15 dpi, respectively, indicating that stunting of TN1 was more severe at 15 dpi than those of TW16 and Utri Merah. The height reduction rate of TW16 decreased from about 22% at 15 dpi to about 20% at 20 dpi, indicating that stunting of TW16 became milder (Figure 3.2). The height reduction rates for Utri Merah were approximately between 3 and 13% during observation from 5 to 20 dpi (Figure 3.2), but the differences in height between the mock-inoculated plants and the corresponding RTBV-infected plants of Utri Merah were not statistically significant (Table 3.1). In TW16, the heights of mock- and RTBV-inoculated plants were significantly different at 10 dpi and onwards (Table 3.1). However, the reduction rates of TN1 were significantly higher than those of Utri Merah at 15 and 20 dpi, and TW16 at 20 dpi (Figure 3.2). Stunting of TN1 plants at 20 dpi was obviously more severe than the other two genotypes (Figure 3.3). These results suggested that the stunting significantly varied between the RTBV-susceptible cultivar, TN1 and the RTBV-resistant plants, TW16 and Utri Merah.

Table 3.1 Heights of mock and RTBV-inoculated plants of TN1, TW16 and Utri Merah

Plant	Treatment	Plant height[1] (cm) at			
		5 dpi	10 dpi	15 dpi	20 dpi
TN1	Mock	38.5 ± 1.7	$50.1 \pm 2.2_a$	$68.1 \pm 1.4_a$	$74.4 \pm 3.6_a$
	RTBV	38.3 ± 2.2	$43.4 \pm 3.5_b$	$45.3 \pm 3.1_b$	$47.1 \pm 2.5_b$
TW16	Mock	39.5 ± 2.1	$49.7 \pm 1.9_a$	$61.8 \pm 2.1_a$	$66.6 \pm 2.3_a$
	RTBV	37.5 ± 2.0	$44.6 \pm 2.0_b$	$48.0 \pm 2.4_b$	$53.3 \pm 2.2_b$
Utri Merah	Mock	50.6 ± 3.3	58.1 ± 2.8	67.9 ± 4.6	$89.5 \pm 3.7_a$
	RTBV	48.8 ± 2.3	59.4 ± 1.7	67.5 ± 1.2	$77.0 \pm 3.5_b$

[1]Mean ± standard error of mean based on three independent experiments. Values for mock- and RTBV-inoculated plants of the same rice genotype at the same days post-inoculation indicated different subscript letters are significantly different by a t-test at the 95% confidence level.

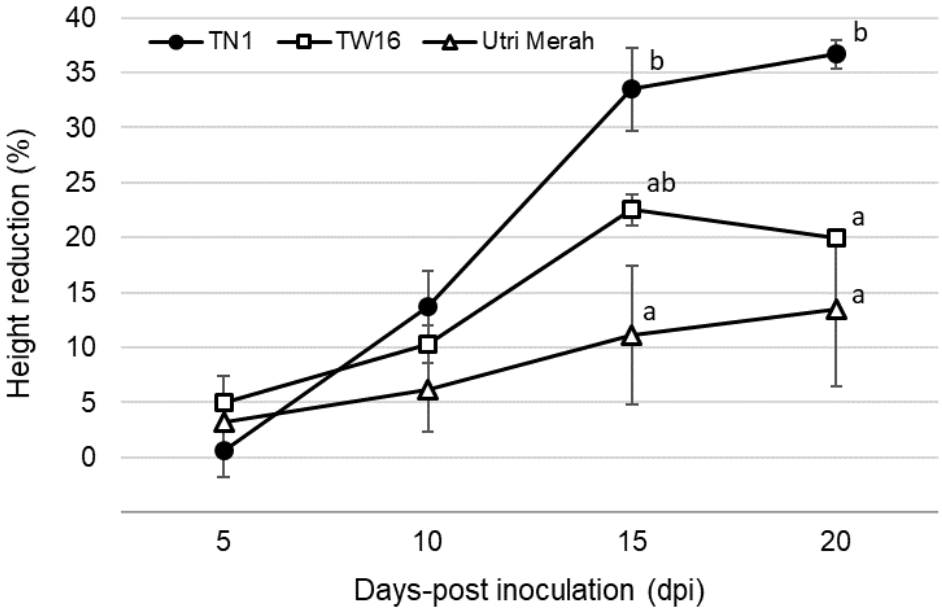


Figure 3.2 Height reduction rate in the RTBV-infected plants.

Height reduction in TN1, TW16 and Utri Merah infected with RTBV. Temporal changes in height reduction rates were shown in figure. The markers and the vertical lines indicate mean and standard error of mean, respectively, for the height reduction rate at the respective time points. Means and standard error of mean were based on three independent experiments (n = 5 per treatment and time point in one experiment). The markers at the same days post-inoculation (dpi) indicated by different letters are significantly different by the least significant difference test at the 95% confidence level.

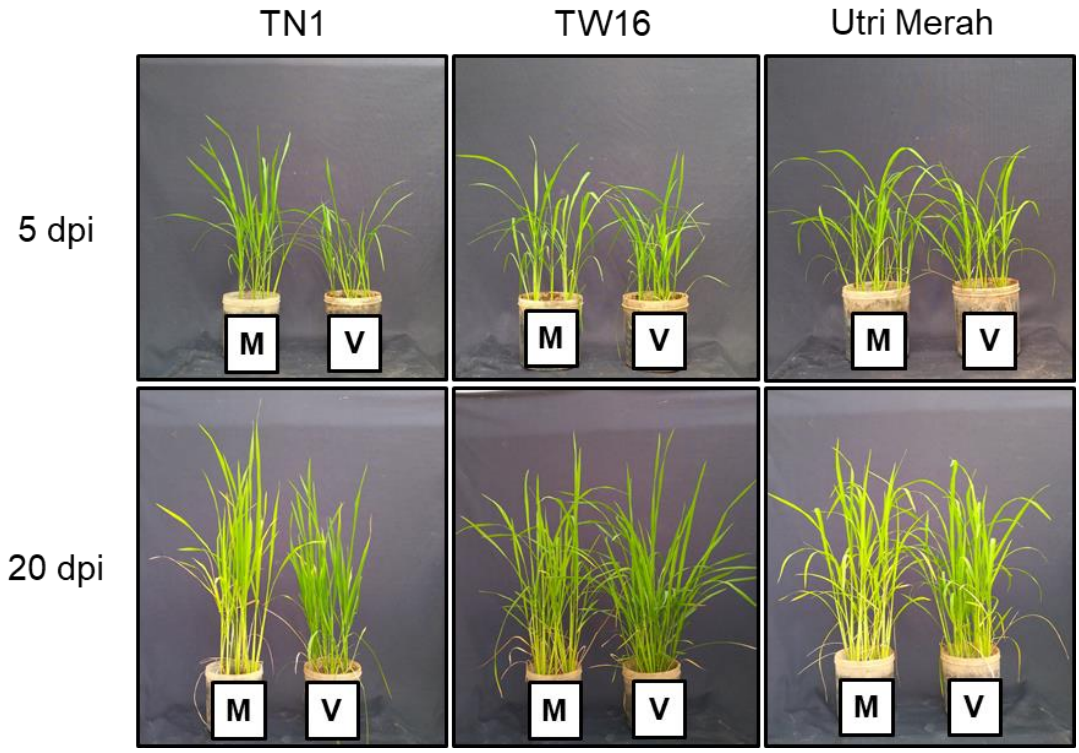


Figure 3.3 Plants of TN1, TW16 and Utri Merah at 5 and 20 dpi with RTBV.

M, mock-inoculated plant; V, RTBV-infected plant.

I carried out qPCR to examine whether the accumulation of RTBV is correlated with the height reduction rates among the three genotypes. The amount of RTBV in TN1 and TW16 increased until 15 dpi and then decreased, whereas in Utri Merah it started to decrease after 10 dpi (Figure 3.4). The copy number of RTBV in TN1 (17,093,803 copies) was significantly higher than that in TW16 (4,396,713 copies) and Utri Merah (3,282,182 copies) at 10 dpi and onwards. The significant differences of virus accumulation between TN1 and the resistant cultivars were observed at earlier time point compared with the height reduction rates between them, implying that the earlier RTBV accumulation led to the severity of stunting.

All the infected plants flowered and produced some seeds. TN1 also set seeds from the stunted plants that were stunted due to RTBV infection. I collected the vegetative parts and seeds from all the mature plants, and five biological replicates were bulked for the biomass analysis. TN1 showed the largest reduction of five agronomic traits (GW, 1000GW, PW, SF and HI) between mock- and RTBV-inoculated plants. Especially, a large reduction of approximately 78% in GW was observed in TN1 (Table 3.2, Figure 3.5). The 1000GW and HI in TW16 and Utri Merah, which were the important factors for the yield of rice, were not affected by the RTBV infection, whereas the decrease in PW observed in TW16 and Utri Merah impaired the values of GW and SN (Table 3.2, Figure 3.5).

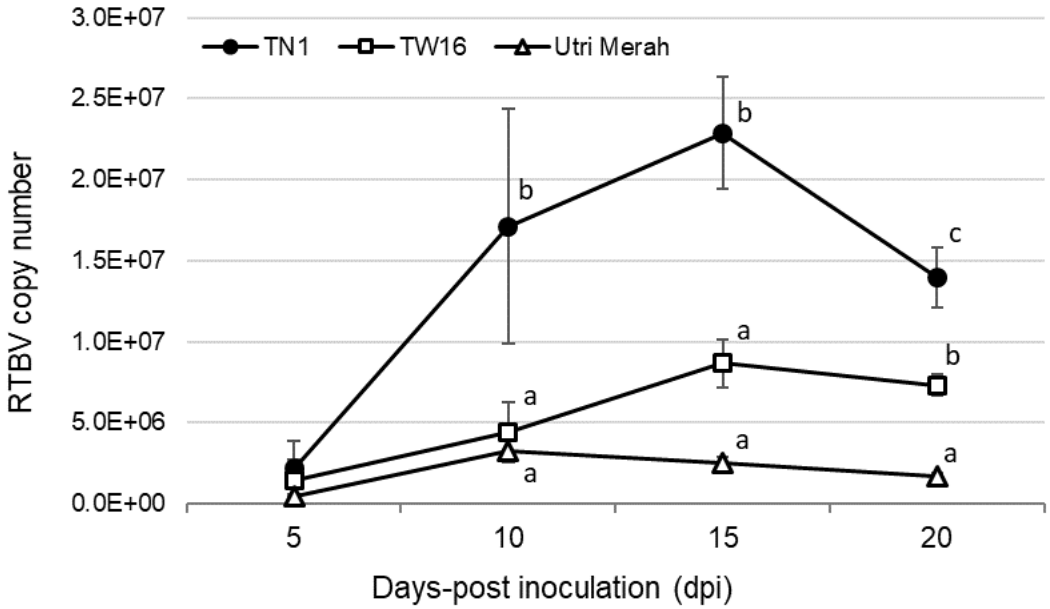


Figure 3.4 Temporal changes in the copy number of RTBV genome in three rice genotypes.

The copy number of RTBV genome in TN1, TW16 and Utri Merah infected with RTBV were indicated. The markers and the vertical lines indicate mean and standard error of mean, respectively, for the copy numbers of RTBV at the respective time points. Means and standard error of mean were based on three independent experiments (n = 5 per treatment and time point in one experiment). The markers at the same days post-inoculation (dpi) indicated by different letters are significantly different by the least significant difference test at the 95% confidence level.

Table 3.2 Effect of RTBV inoculation on some agronomic traits in TN1, TW16 and Utri Merah

Agronomic traits	TN1			TW16			Utri Merah		
	Mock	RTBV	% of Mock [1]	Mock	RTBV	% of Mock	Mock	RTBV	% of Mock
PW (g)	$30.6 \pm 1.6_a$	$24.8 \pm 1.2_b$	81	25.8 ± 2.0	22.7 ± 1.8	88.1	62.8 ± 3.2	61.9 ± 2.3	98.6
GW (g)	$23.7 \pm 1.5_a$	$5.2 \pm 0.4_b$	22.1	19.4 ± 1.5	18.6 ± 1.0	95.8	23.4 ± 1.5	20.7 ± 1.2	88.5
1000GW (g)	$21.6 \pm 0.3_a$	$18.7 \pm 0.2_b$	86.3	22.6 ± 0.2	22.3 ± 0.3	98.7	24.6 ± 0.2	24.1 ± 0.2	97.8
PN	3.6 ± 0.2	3.6 ± 0.1	101.0	3.2 ± 0.2	3.5 ± 0.2	110.1	3.5 ± 0.3	3.4 ± 0.3	99.0
SN	$85.7 \pm 5.0_a$	$41.5 \pm 2.0_b$	48.4	66.9 ± 0.4	66.1 ± 0.4	99.0	$72.9 \pm 7.0_a$	$57.6 \pm 4.8_b$	79
SF (%)	$89.4 \pm 1.0_a$	$65.5 \pm 2.8_b$	73.3	93.3 ± 0.8	92.3 ± 1.5	99.0	92.4 ± 1.3	92.5 ± 2.2	100.0
HI (%)	$77.7 \pm 7.6_a$	$24.1 \pm 2.4_b$	31.1	75.7 ± 2.3	79.0 ± 2.3	104.4	37.5 ± 2.7	34.0 ± 1.4	90.6

[1] % of Mock was computed as [RTBV/Mock × 100] for each trait and represented the changes due to RTBV infection.

Mean ± standard error of mean based on ten to fifteen 5 individual plants-pooled samples. Values for mock- and RTBV-inoculated plants of the same rice genotype indicated different subscript letters are significantly different by a t-test at the 95% confidence level.

PW, plant weight; GW, grain weight; PN, panicle number; SN, spikelet number; SF, spikelet fertility; HI, harvest index.

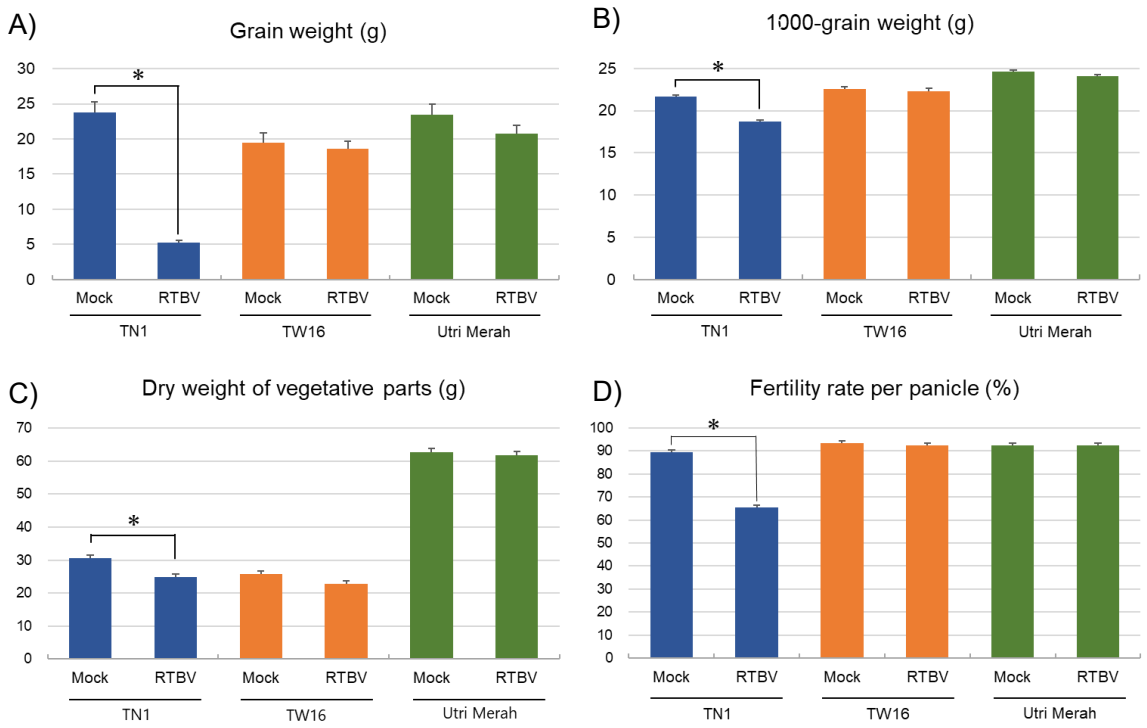


Figure 3.5 Effect of RTBV infection on four agronomic traits.

Four agronomic traits were analyzed between mock- and RTBV-inoculated plants among TN1, TW16 and Utri Merah. A) grain weight, B) 1000-grain weight, C) dry weight of vegetative parts, D) fertility rate per panicle. Asterisk represents significantly difference between mock- and RTBV-inoculated plants by the t-test at the 95% confidence level.

3.3.2 Overview of small RNA sequencing

TN1 and TW16 were examined to identify small RNAs, which might be associated with their difference in the reaction to RTBV. Total RNA was extracted from individual plants collected at 10 dpi, and four small RNA libraries for RTBV-infected and mock-inoculated plants were constructed and sequenced. After trimming adaptor and removing low-quality reads, 18,260,272 to 22,546,201 high-quality reads were obtained from the respective libraries (Table 3.3). In terms of the length distribution of small RNA of 19 to 27 nt of small RNAs, the pattern was very close between mock-inoculated and RTBV-infected TN1 (Figure 3.6A). On the other hand, for the RTBV-infected TW16, the frequency of 21 nt-long small RNAs was lower and that of 24 nt-long small RNAs was higher than the mock-inoculated plants (Figure 3.6B). The majority of 21 nt-long small RNAs were endogenous small RNA derived from the rice genomic sequences (Axtell and Merchant, 2013). Indeed, the frequency of the rice-derived small RNA in TW16 infected with RTBV was lower than that in mock-inoculated plants (Table 3.4). The major sizes of vsRNA were 21 nt, 22 nt, and 24 nt in both TN1 and TW16, and the range in length of vsRNA detected in TN1 was broader than that in TW16 (Figure 3.6C). According to previous reports, DCL1, DCL2, and DCL3 involved in silencing machineries of host plant against DNA virus generated 21 nt, 22 nt, and 24 nt of vsRNAs, respectively (Akbergenov *et al*., 2006; Blevins *et al*., 2006, 2011). According to Figure 3.6B, the frequency of 21- and 22-nt small RNA in the RTBV-infected TW16 was lower than those in the mock-inoculated TW16. This result might be because the expression level of rice miRNA in the RTBV-infected TW16 was approximately 26.5% lower than that in the mock-infected one (Table 3.4). Indeed, more vsRNAs with 21 nt and 22 nt in length were observed in TW16 rather than TN1, therefore, TW16 might produce those vsRNA for silencing against RTBV.

Using the Bowtie alignment program, the reads of small RNAs from RTBV-infected and mock-inoculated plants of TN1 and TW16 were sorted out into three classes based on origin: the rice genome, the RTBV genome, and known miRNAs (Table 3.4). Of the total reads, unique reads (uniques) accounted for approximately 10% in the respective samples, and the small RNAs mapped to the rice genome, were found to be 45 to 62%. In the samples of the RTBV-infected plants, 1.5 to 2.9% of the small RNAs were mapped to the RTBV genome, while almost no small RNAs mapped to the RTBV genome were found in the samples from mock-inoculated plants (Table 3.4). In fact, 330,884 reads from TN1 and 590,933 reads from TW16 in the libraries of RTBV-infected plants were perfectly matched with the RTBV genome, whereas a negligible number of reads (less than 50 reads) in the libraries from mock-inoculated plants were mapped to the RTBV genome (Table 3.4). All the clean reads were also attempted to be aligned with known rice miRNAs, which were collected from miRBase (version 22). In all the samples, the reads aligned with miRNAs accounted for 66 to 88% of reads mapped to the rice genome, whereas the ratios of unique reads were less than 1% of total unique reads (Table 3.4). The differences between total and unique reads mapped with rice known miRNA implied that specific rice miRNAs were highly expressed but various kind of miRNA were not found in all the libraries (Table 3.4). I paid attention to the RTBV-infected samples from TW16, which had a significantly lower amount of miRNA than the other samples. I speculated that the lower ratio of miRNA present in TW16 might be related to RTBV resistance which could be mediated by the generation of vsRNA.

Table 3.3 The overview of quality check of raw reads derived from sequencing of small RNAs from mock- and RTBV-inoculated TN1 and TW16

		Total raw reads	After trimming (Q>30, 17≦length≧29)
RTBV-infected	TN1	34,320,830	18,260,272 (53.20%)
	TW16	32,398,657	20,718,758 (63.95%)
Mock-inoculated	TN1	29,790,910	22,546,201 (75.68%)
	TW16	28,642,966	21,331,387 (74.47%)

The raw reads were filtered by following condition: the average of base quality in read is more than 30, the range of the read length is from 17 to 29 nt. Percentage represent that the ratio of the number of reads after trimming to the total number of raw reads.

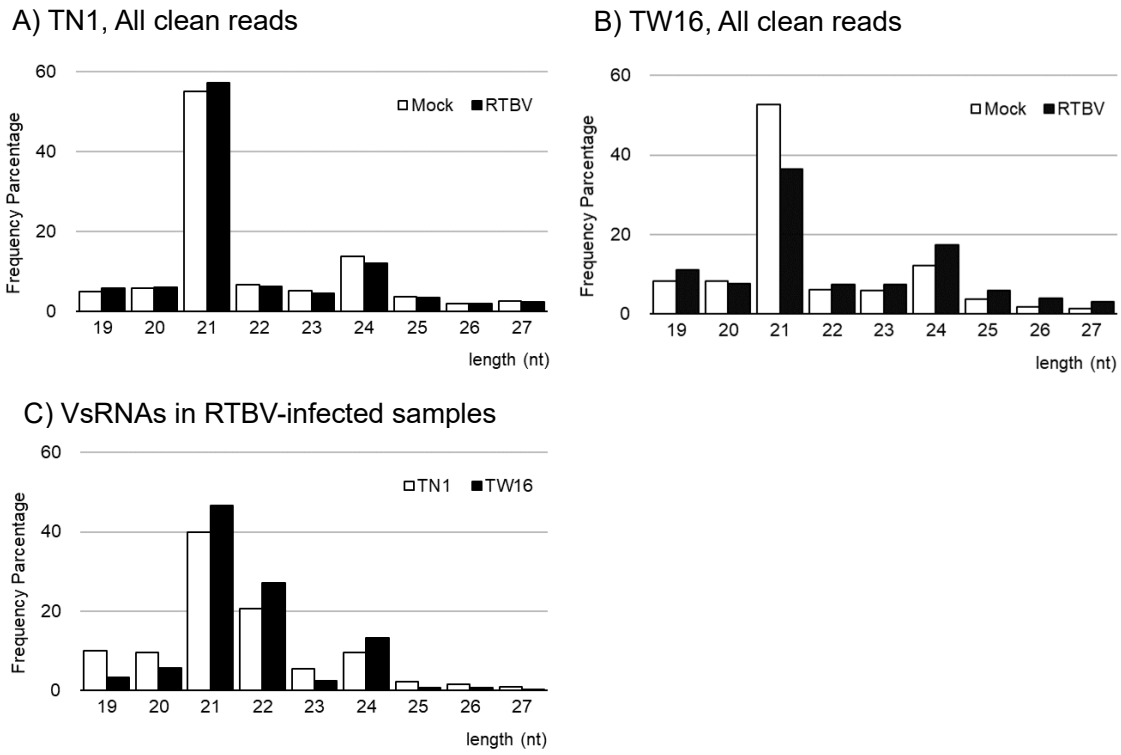


Figure 3.6 The nucleotide length distributions of small RNAs in TN1 and TW16.

Bars indicate the frequency percentage for the number of reads in respective length. A) all clean reads of TN1, B) all clean reads of TW16, C) small RNAs mapped to the RTBV genome in infected samples of TN1 and TW16.

Table 3.4 Summary of small sequencing results: small RNAs from virus-infected and mock-inoculated rice small RNA libraries

	RTBV-infected				mock-inoculated			
	TN1		TW16		TN1		TW16	
	Reads	Uniques	Reads	Uniques	Reads	Uniques	Reads	Uniques
Total number	18,260,272	1,972,933	20,718,758	2,059,099	22,546,201	1,955,876	21,331,387	1,540,567
Mapped to rice genome	11,262,898 (61.68%)	963,278 (48.83%)	9,407,143 (45.40%)	1,042,662 (49.36%)	13,767,400 (61.06%)	991,538 (49.31%)	13,217,711 (61.96%)	769,875 (50.64%)
Mapped to virus geome	330,884 (1.81%)	29,369 (1.49%)	590,933 (2.85%)	21,943 (1.07%)	32 (0.00%)	29 (0.00%)	35 (0.00%)	29 (0.00%)
Mapped to known miRNA	8,929,360 (48.90%)	17,002 (0.86%)	6,201,798 (29.93%)	15,654 (0.76%)	11,858,393 (52.60%)	14,406 (0.74%)	11,669,239 (54.70%)	13,210 (0.86%)

Percentage was calculated by the following formula: [Mapped reads (Reads or Uniques)/Clean reads (Reads or Uniques) × 100].

3.3.3 The comparison of accumulation pattern of vsRNAs with the expression of silencing-related gene between TN1 and TW16

The number of vsRNA reads were obtained from each library of infected samples at 10 dpi and the number of nucleotides mapped in the virus genome was shown as depth (Figure 3.7). The hotspot of vsRNA, which comprised three peaks with more than 30000 reads, appeared at IGR of the virus genome in TW16 (Figure 3.7). Fewer accumulations of vsRNAs from TN1 were obviously throughout the RTBV genome (Figure 3.7). The two largest peaks (sense and antisense reads) with nearly 90000 reads appeared in the leader region were apparently in contrast to the pattern of TN1 (Figure 3.7). The previous studies also detected such peaks, but the reads of vsRNA was distributed across the whole IGR (Rajeswaran *et al*., 2014; Zarreen *et al*., 2018). The details of vsRNA hotspots at IGR were analyzed to predict their role in resistant plants. One of the two peaks of vsRNA hotspot which appeared in TW16 was located around the start codon of the short upstream ORF I (uORF I), and the other corresponded to the stem region of the basal helix structure (Figure 3.8). The two positions with the abundant reads of vsRNA in TW16 were structurally committed with the virus function of RTBV. That is, uORF I containing start codon (AUG) is essential to initiate shunt-mediated translation at the AUU start codon of ORF I, and the highly-structured leader region is shunted over during translation initiation (Pooggin *et al*., 1999; 2006 and 2012) (Figure 3.8). The top of the large stem-loop structure in the leader region interacts with viral CP (Guerra-Peraza *et al*., 2000) (Figure 3.8).

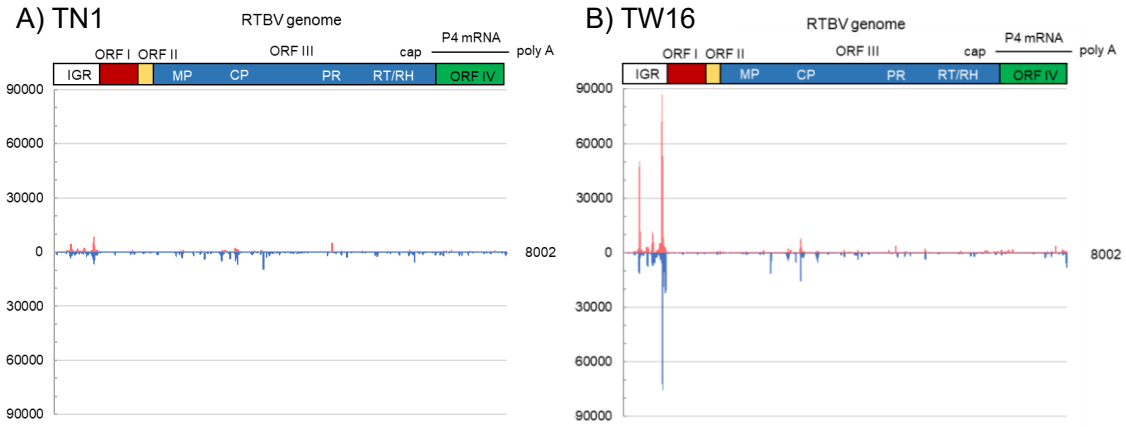


Figure 3.7 Distributions A) for TN1 and B) for TW16 of small RNA abundance along RTBV genome in infected plants.

Bars represent the depth for each nucleotide of RTBV genome. Red bars are for sense reads mapped to the RTBV genome and blue bars are for antisense reads. X-axis represents RTBV genomic sequence (8002 bp, NCBI accession number NC_001914.1) and the color box above the bar graph indicates the structures and ORFs of RTBV genome.

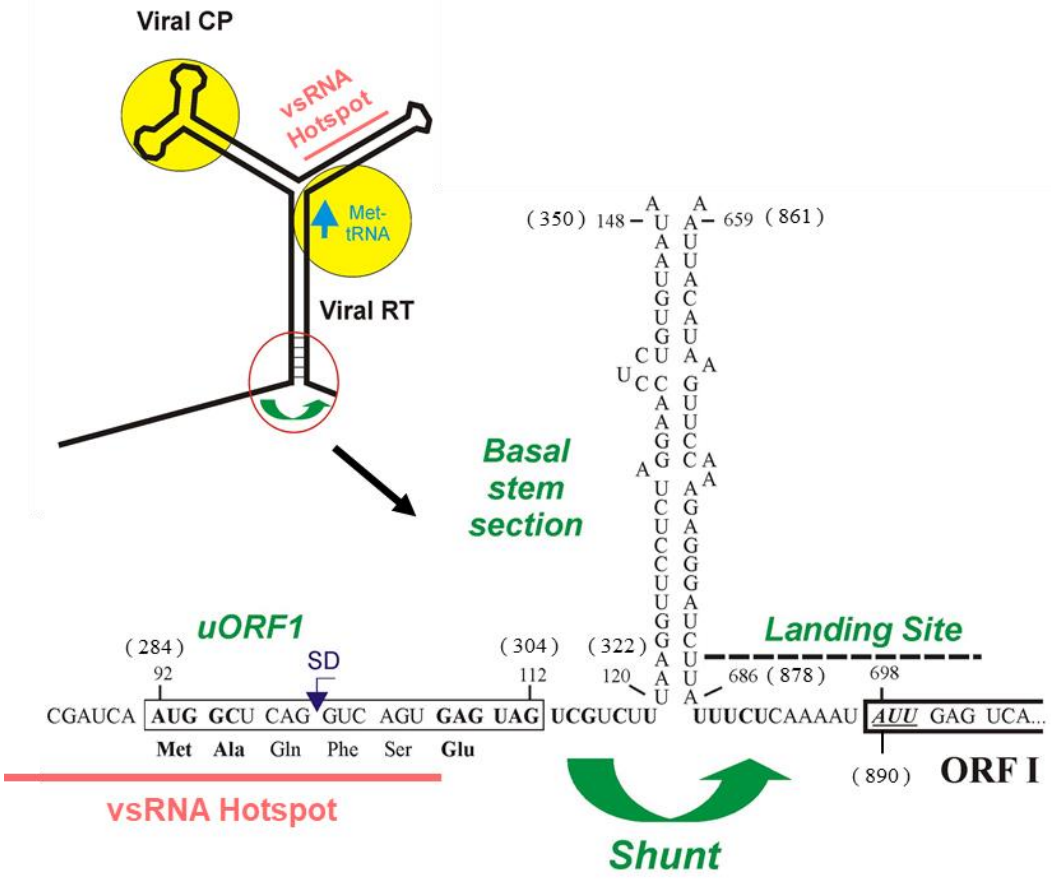


Figure 3.8 RTBV leader primary containing secondary structure and vsRNA hotspot (modified from Pooggin and Ryabova. 2018).

The numbers in parentheses represent the positions of nucleotide corresponding to that in Figure 3.7. The number is the position of the transcript. The nucleotide numbering is from the RNA 59-end. The 59-proximal short upstream ORF (uORF1) is boxed. The uORF1 AUG and the non-AUG start codons in the shunt landing site are underlined. The identical nucleotide stretches/motifs in the shunt take-off and landing sites are highlighted in bold.

3.3.4 Comparison of host silencing-related genes between TN1 and TW16

The expressions of host gene silencing-related genes were quantified by RT-qPCR (Figure 3.9). I tested the expressions of three *AGO* genes, *AGO1b*, *AGO2*, and *AGO18*, related to production of vsRNAs in the processes of RNA silencing. *AGO2* and *AGO1b* are known as effectors of vsRNA to directly slice virus mRNA in rice (Carbonell and Carrington, 2015) (Figure 1.2). *AGO18* targets osa-miR168, which inhibits functions of transcripts of the *AGO1* family (Wu *et al*., 2015); i.e. *AGO18* can sequester osa-miR168 from *AGO1* family. To make a relative standard index of the expression level, I used a value of the mock-inoculated plants of TN1 at 5 dpi as 1.0. At 5 dpi, both the expressions of *AGO1b* and *AGO2* genes in the RTBV-infected TW16 were the highest level among all the different dpi samples. The *AGO1b* and *AGO2* genes were up-regulated immediately after RTBV infection in TW16 (Figure 3.9A, 3.9B). Then, the expression levels of both genes decreased to less than one-sixth. These results may imply a correlation between RTBV resistance in TW16 and the expressions of *AGO1b* and *AGO2*. The accumulation of RTBV might be quickly suppressed by an anti-virus silencing machinery associated with the *AGO1b* and *AGO2* expressions in the resistant plants. The expression of *AGO18* was up-regulated due to RTBV infection in both cultivars. Particularly the expression of *AGO18* in TN1 was significantly higher than that in TW16 at 5 and 10 dpi (Figure 3.9C). These results suggested that the *AGO18* expression in TN1 and TW16 is not directly connected with RTBV resistance. In the resistant plants, the expressions of the three *AGO* genes were commonly induced by RTBV infection immediately after 5 dpi. Contrastingly, the low level of *AGO1b* and *AGO2* gene expressions were observed in the plants susceptible to RTBV. Taking together these results, the different degrees of the resistance to RTBV in TW16 and TN1 might be reflected by the differences of anti-viral machinery associated with *AGO1b* and *AGO2*.

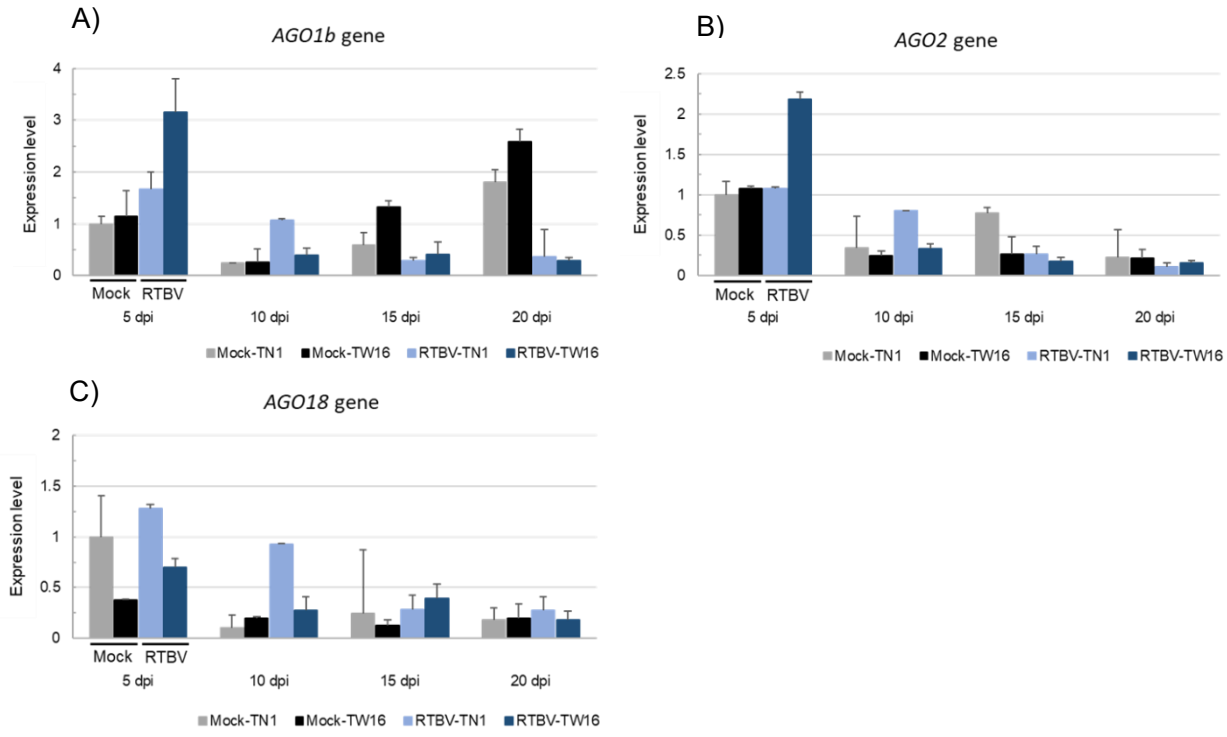


Figure 3.9 Temporal changes in the expression of three *AGO* genes in mock- and RTBV-inoculated TN1 and TW16.

Temporary changes in the expression of A) *AGO1b* gene, B) *AGO2* gene, and C) *AGO18* gene. Expression level was calculated by ΔΔCt method. The expression level for the mock-inoculated TN1at 5 dpi is converted to 1.0 in each graph and expression level of other samples was relative value. Vertical lines indicate standard error. Primers used for the expression analysis are described in Supplementary Table S1.

3.3.5 Possible association of rice known miRNAs to RTBV resistance

The most recent version of miRBase (version 22, http://www.mirbase.org/), which included a catalogue for the identified rice miRNAs (mature miRNAs derived from 604 precursors) was used for the identification of miRNAs in our deep sequencing data. A total of 6,201,798-11,858,393 reads (13,210-17,002 uniques) perfectly matched with known rice miRNA from the four small RNA libraries (Table 3.4). In RTBV-infected TW16, the percentage of reads derived from known rice miRNA was 29.93%, which was almost a half of the percentage of the known miRNAs (54.70%) in the mock-inoculated plants. On the other hand, the TN1 miRNAs mapped to the known miRNAs showed nearly similar percentages between the RTBV-infected (48.90%) and mock-inoculated (52.60%) plants. These results demonstrated that the expression of the known miRNAs in the resistant cultivar TW16 might be down-regulated due to the RTBV infection, but not in the sensitive cultivar TN1. To compare the expression level in each of the known miRNAs among the four miRNA libraries, I quantified differential expression (DE) levels of miRNAs by edgeR after normalization of the raw read counts (Table 3.5, 3.6). A stringent parameter ($p<0.05$, LogFC>2 and more than 10 RPM in at least one library after normalization) was employed to filter the outcomes of edgeR. The logarithm to base 2 was calculated for LogFC. Then, I identified 13 known miRNAs (11 DE miRNAs in TN1, and 2 in TW16), of which DE was compared between mock-inoculated samples and RTBV-infected samples (Table 3.5, Figure 3.10). The RTBV infection greatly down-regulated the expressions of 13 miRNAs (Figure 3.10). There were no common DE miRNAs due to RTBV infection between TN1 and TW16, indicating that the expression patterns of rice miRNAs against RTBV infection varied in the genotypes. These identified DE miRNAs might be associated with RTBV resistance if the resistance is constitutive.

Table 3.5 Differentially expressed (DE) rice miRNAs between mock-inoculated and RTBV-infected plants in two genotypes

	DE miRNA[1]	Mock-inoculated sample (RPM)	RTBV-infected sample (RPM)	LogFC[2]	p-value
TN1	osa-miR166j-3p	148021.6	9879.6	-4.0	0.0117
	osa-miR166c-3p	148460.3	9925.5	-4.0	0.0118
	osa-miR399d	924.2	88.7	-3.5	0.0256
	osa-miR166b-3p	113499.8	11914.8	-3.4	0.0300
	osa-miR166d-3p osa-miR166f	113679.9	11968.1	-3.4	0.0301
	osa-miR166a-3p	113679.7	11968.1	-3.4	0.0301
	osa-miR166i-3p	1221.1	139.4	-3.2	0.0361
	osa-miR166m	2391.9	277.9	-3.2	0.0372
	osa-miR166e-3p	45.9	3.9	-3.6	0.0419
	osa-miR166g-3p	54009.7	6830.2	-3.1	0.0434
TW16	osa-miR156g-3p	194.1	3.9	-5.3	0.0453
	osa-miR156c-3p	193.2	3.9	-5.3	0.0456

[1] DE miRNAs were selected by following option: $|logFC| > 2$, $p<0.05$. One of the samples' read count in comparison should have more than 10.0 reads per million (RPM).

[2] LogFC was calculated by the following formula: [(log base 2 converted from RPM of reads mapped with miRNA in RTBV-infected sample) – (log base 2 converted from RPM of reads mapped with each miRNA in mock-inoculated sample)]

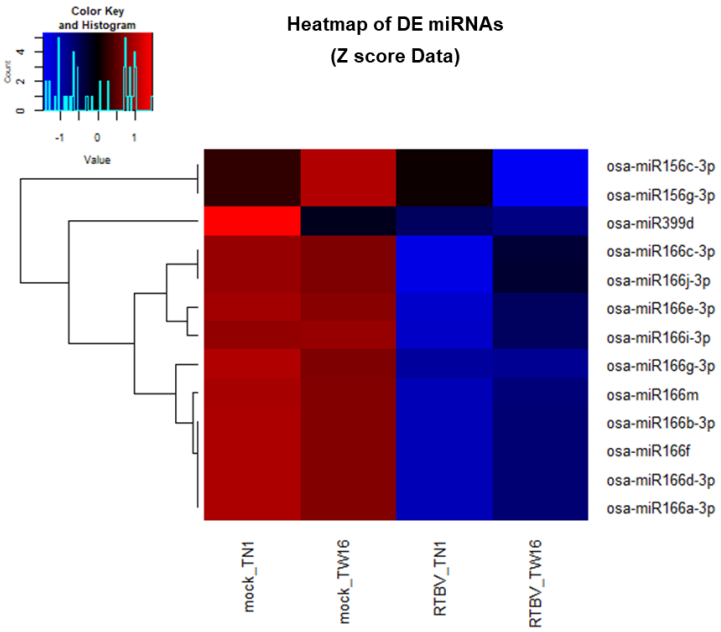


Figure 3.10 Heatmap of DE rice miRNAs between mock-inoculated and RTBV-infected plants in two genotypes.

The heatmap was computed using a function of heatmap.2 in gplots by R platform. The red indicated relatively higher miRNA expression level and the blue showed lower miRNA expression level. The actual expression level of each miRNA in the right of the heatmap was indicated in Table 3.5. Four libraries are represented at the bottom. miRNAs are arranged by the clustering of the expression patterns.

The expressions of the rice miRNAs were compared between the two genotypes of the same treatment (RTBV-infected or mock-inoculated). Seven and five were identified as DE miRNAs between the mock-inoculated TN1 and TW16 and between RTBV-infected plants, respectively (Table 3.6, Figure 3.11). Because of the differences between the RTBV-infected samples, the five DE miRNAs were the candidates likely to be associated with the RTBV resistance. The third comparison was conducted to estimate miRNAs by the integration of the two factors: the evaluation was based on the fold changes of the expressions between mock-inoculation and RTBV-infection, then the fold changes of the same miRNA was compared between the susceptible and resistant genotypes; [(LogFC due to RTBV infection in TW16) – (LogFC due to RTBV infection in TN1)]. To select candidate miRNAs, I set the filter conditions where the difference of expression changes was more than or less than 1.5 differences, and each expression in RPM (reads per million) was more than 100. After filtering, 15 miRNAs were selected (Figure 3.12, Table 3.7). Among those, osa-miR399d, osa-miR166j-3p, osa-miR166c-3p and osa-miR530-5p were also found in the previous analyses (Table 3.5, 3.6, 3.7). Considering the differences of the expression level and LogFC due to RTBV infection between the two genotypes by the three computational analyses, a total of 11 miRNAs were selected as candidates, which were expected to be involved in the RTBV infection or resistance (Table 3.8). The expression of all the candidate miRNAs decreased at 10 dpi in TW16 (Figure 3.12). The target of the selected miRNAs were predicted by psRNATarget (http://plantgrn.noble.org/psRNATarget/), using the following option: expectation<3, and non-annotated genes were removed (Table 3.8). The target prediction in this study was mostly consistent with known target gene for rice miRNA. Of the 11 candidate miRNAs, only osa-miR399d was characterized by Zarreen *et al*., (2018). According to the report, RTBV infection impaired osa-miR399d expression (Figure 3.12). This study also

showed that the expression of osa-miR399 in TN1 was more down-regulated than that in TW16 after RTBV infection (Table 3.7, Figure 3.12). The other miRNAs such as osa-miR530-5p, osa-miR444b.2/c.2, and osa-miR159a.1/b were predicted to target the same zinc finger proteins, which contained C3HC4 type RING finger domain, and these target genes were located at different chromosomes (Table 3.8). A MADS-box family gene was predicted as a target of osa-miR444, which was known as regulator of silencing machinery via vsRNA (Wang *et al*., 2016) (Table 3.8). The target genes of these identified miRNAs were previously reported, but the association with RTBV infection have not been reported except for osa-miR399d (Zarreen *et al*., 2018) (Table 3.7, 3.8).

Table 3.6 DE rice miRNAs between the two genotypes of the same treatment (RTBV-infected or mock-inoculated).

	DE miRNA[1]	TN1 (RPM)	TW16 (RPM)	LogFC [2]	p-value
Mock-inoculated sample	osa-miR530-5p	393.3	10343.2	4.8	9.96E-15
	osa-miR399d	924.2	138.8	-2.7	2.47E-06
	osa-miR390-3p	2.0	37.8	4.2	1.87E-05
	osa-miR399j	226.3	50.8	-2.1	3.48E-04
	osa-miR169n	6.1	29.2	2.3	0.006
	osa-miR169o	5.9	29.0	2.3	0.007
	osa-miR530-3p	4.3	21.0	2.3	0.012
RTBV-infected sample	osa-miR159a.1 osa-miR159b	2751.2	42.2	-6.0	6.88E-04
	osa-miR159f	121.7	2.7	-5.4	0.022
	osa-miR1423-3p	57.1	0.9	-5.8	0.025
	osa-miR1876	177.8	7.5	-4.5	0.038

[1] DE miRNAs were selected by following option: |logFC| > 2, $p<0.05$. One of the samples' read count in comparison should have more than 10.0 RPM.

[2] LogFC was calculated by the following formula: [(log base 2 converted from RPM of reads mapped with miRNA in RTBV-infected sample) – log base 2 converted from RPM of reads mapped with each miRNA in mock-inoculated sample)]

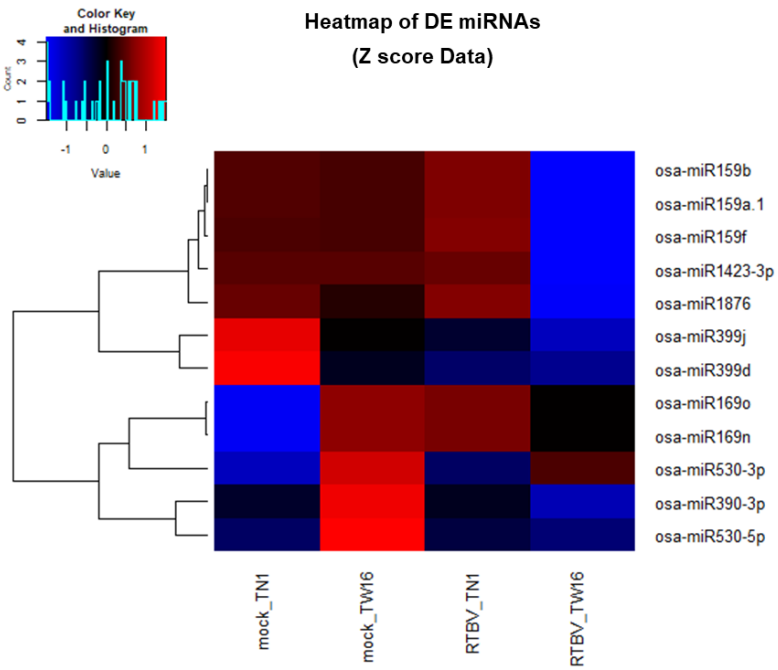


Figure 3.11 Heatmap of DE rice miRNAs between the two genotypes of the same treatment (RTBV-infected or mock-inoculated).

The heatmap was computed using a function of heatmap.2 in gplots by R platform. The red indicated relatively higher miRNA expression level and the blue showed lower miRNA expression level. The actual expression level of each miRNA in the right of the heatmap was indicated in Table 3.6. Four libraries are represented at the bottom. miRNAs are arranged by the clustering of the expression patterns.

Table 3.7 The list of miRNAs which have great differences of expression change due to RTBV infection between TN1 and TW16

miRNA ID	TN1 (LogFC)	TW16 (LogFC)	Difference of [1] expression change
osa-miR399d	-3.4	-1.0	2.4
osa-miR166j-5p	-0.6	1.7	2.4
osa-miR166j-3p	-3.9	-1.8	2.1
osa-miR166c-3p	-3.9	-1.8	2.1
osa-miR167b	-1.6	0.0	1.7
osa-miR167a-5p/c-5p	-1.6	0.0	1.7
osa-miR166h-3p	-1.7	-3.2	-1.5
osa-miR1425-5p	0.7	-1.4	-2.0
osa-miR444a-3p.2	0.4	-1.8	-2.2
osa-miR444e	0.4	-1.8	-2.2
osa-miR444b.2/c.2	-0.4	-2.7	-2.3
osa-miR444d.2	0.4	-1.9	-2.3
osa-miR444b.1/c.1	-0.4	-2.7	-2.3
osa-miR530-5p	0.4	-4.9	-5.3

[1] The differences of more than 2.0 or less than -2.0 were selected.

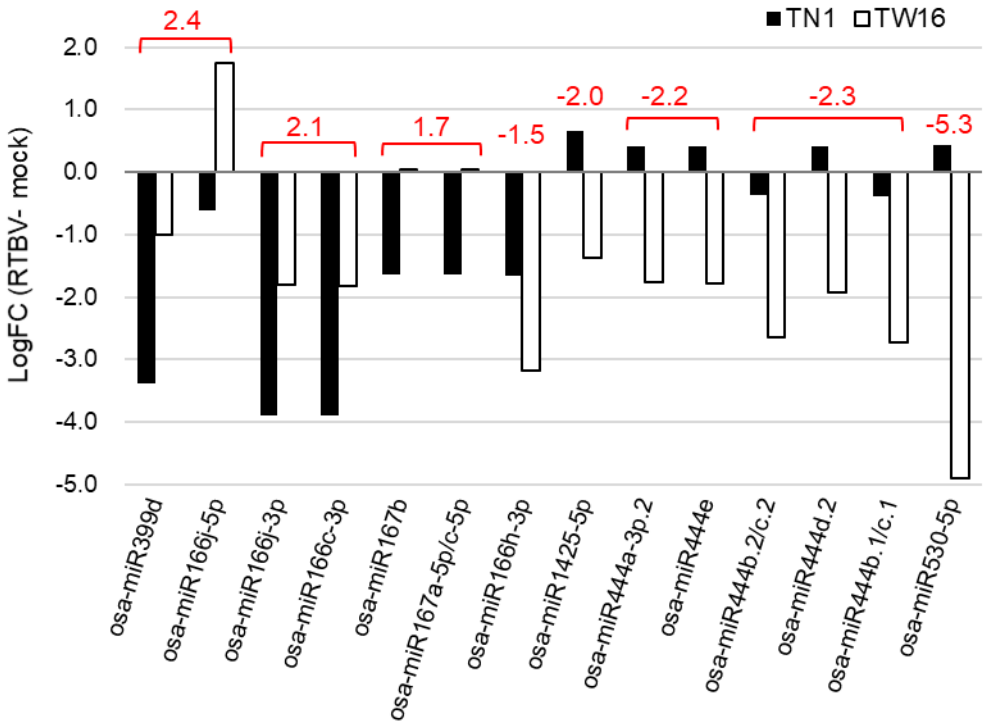


Figure 3.12 The comparison of the expression change due to RTBV infection between TN1 and TW16.

High-expressed miRNAs, which were used in this figure, were filtered by logarithm to base 2 > 6. The number in red color above the bar represents the difference of expression change due to RTBV infection between TN1 and TW16, and over 1.5 or less than -1.5 were showed in this graph. The formula is [TN1 {LogFC (RTBV-mock)}-TW16 {LogFC (RTBV-mock)}]. The black bar is logFC (RTBV-mock) of TN1, and the white bar is for TW16.

Table 3.8 The list of predicted genes which were targeted by known miRNA in rice

miRNA ID	Gene ID	Target discription
osa-miR530-5p	LOC_Os02g14990.1 LOC_Os04g51400	zinc finger, C3HC4 type domain containing protein
	LOC_Os05g44740.1	transposon protein, unclassified
	LOC_Os08g32100.1	transposon protein, CACTA, En/Spm sub-class
	LOC_Os04g56980.1	SWIB/MDM2 domain containing protein
	LOC_Os03g18790.1	SAG20
	LOC_Os06g16600.1	RNA-directed DNA polymerase
	LOC_Os05g44720.1	retrotransposon protein, unclassified
	LOC_Os11g04520	PP2A regulatory subunit TAP46
	LOC_Os01g56780.1	plus-3 domain containing protein
	LOC_Os03g63950	plastid-specific 30S ribosomal protein 1, chloroplast precursor
	LOC_Os01g01120.1	enolase-phosphatase E1
	LOC_Os12g44340.1	ATMAP70 protein
osa-miR444b.2	LOC_Os08g06510.1	zinc finger, C3HC4 type domain containing protein
osa-miR444c.2	LOC_Os04g38780.1	transcription factor
	LOC_Os07g14514	retrotransposon protein, unclassified
	LOC_Os02g49840 LOC_Os02g36924.1 LOC_Os08g33488.1	OsMADS57 - MADS-box family gene with MIKCc type-box
	LOC_Os01g44990.1	ercc6 protein
	LOC_Os10g02920.1	cytochrome b561
	LOC_Os01g67410.1	AP2/EREBP transcription factor BABY BOOM
osa-miR399d	LOC_Os08g45000.1	inorganic phosphate transporter
(Archak et al. 2007)	LOC_Os05g48390.1	ubiquitin conjugating enzyme protein
	LOC_Os05g45350.2	dnaJ domain containing protein
	LOC_Os04g55230.1	tetratricopeptide repeat domain containing protein
osa-miR166c-3p	LOC_Os02g49670.1	zinc knuckle family protein
osa-miR166j-3p	LOC_Os12g41860.1 LOC_Os10g33960 LOC_Os03g43930 LOC_Os03g01890	START domain containing protein
	LOC_Os03g35380.1 LOC_Os01g33740.1	retrotransposon protein, unclassified
	LOC_Os12g43900.1	retrotransposon protein, LINE subclass

osa-miR159a.1	LOC_Os10g05230.2	zinc finger, C3HC4 type domain containing protein
osa-miR159b	LOC_Os03g21380.1	OsCML27 - Calmodulin-related calcium sensor protein
	LOC_Os01g59660 LOC_Os03g38210.1 LOC_Os04g46384.1 LOC_Os05g41166.1 LOC_Os06g40330.1 LOC_Os01g12700.1	MYB family transcription factor
osa-miR156c-3p	LOC_Os06g37085.1	transposon protein, unclassified
osa-miR156g-3p	LOC_Os10g37620.1	retrotransposon protein, unclassified
	LOC_Os12g04800.1 LOC_Os04g08776.1	retrotransposon protein, Ty3-gypsy subclass
	LOC_Os03g44540.1	nuclear transcription factor Y subunit
	LOC_Os08g42960.1	ankyrin repeat-containing protein
	LOC_Os01g25450.1	AIG1 family protein
osa-miR1425-5p	LOC_Os05g40700.1	transmembrane protein 56
	LOC_Os10g35640.1 LOC_Os10g35436.1 LOC_Os10g35240 LOC_Os10g35230.1 LOC_Os08g01650.1 LOC_Os08g01640.1	Rf1, mitochondrial precursor
	LOC_Os05g34540.1 LOC_Os05g23860.1	rab GDP dissociation inhibitor alpha
	LOC_Os01g49614.1 LOC_Os01g49580.1	Protein kinase domain containing protein

3.3.6 Function of eRTBVL sequence regarding RTBV resistance

The high-quality reads were mapped to the sequences of eRTBVL to see whether the expression of eRTBVL-small RNA varied between susceptible and resistant samples. Using Bowtie, 19,968 reads and 3,434 uniques for TN1, and 31,742 reads and 2,910 uniques for TW16 were well aligned to at least one eRTBVL family in infected samples (Figure 3.13A). In mock-inoculated samples, 17,967 reads and 1,948 uniques for TN1, and 20,116 reads and

1,833 uniques for TW16 were mapped to eRTBVL sequence (Figure 3.13A). The accumulation of eRTBVL-derived small RNAs were examined to see where the small RNAs were distributed among eRTBVL sequence (Figure 3.13B-E). The eRTBVL-derived small RNAs were mostly accumulated at IGR, especially in the minus strands (Figure 3.13B-E). The position of small RNA hotspot was consistent with the results described in Chapter 2 using *O. sativa* cv. IR 64 (Figure 2.4). The small peaks of small RNAs in both the RTBV-infected cultivars broadly distributed in RT/RH, ORFz and IGR regions of the eRTBVL-C genome (Figure 3.13A, B, C), whereas the lower accumulation levels of the peaks in IGR region were observed in the RTBV-infected samples relative to mock-inoculation samples. These results indicated that RTBV infection did not induce the massive changes in the accumulation patterns of the small RNAs. In infected sample of TW16, the small peak of the accumulation for eRTBVL-small RNA were observed at RT/RH but not in TN1 infected with RTBV, indicating that the accumulation of these specific small RNAs in TW16 were induced by RTBV infection.

The expression of IGR-specific small RNA derived from eRTBVL was monitored in RTBV-infected plants, which were collected at earlier time points (Figure 3.14). The most accumulated eRTBVL-derived small RNA, which was identified by small RNA sequencing analysis for IR 64 (Figure 2.5), were used for expression analysis (Figure 3.14A). This IGR-specific eRTBVL-derived small RNA was located at position of 716 to 738 (Figure 3.14A). The significant differences were not observed for the expression of IGR-specific small RNA between TN1 and TW16. At the early time points such as 3 and 5 dpi, the expression levels for both genotypes were higher than the late time points 7 and 10 dpi. The results suggest that the rice plants rapidly responded to RTBV infection and induced the high expression of IGR-

specific small RNA. Thus, eRTBVL derived small RNA from IGR appeared to be not directly associated with the resistance between TN1 and TW16 against RTBV.

A)

	RTBV-infected				mock-inoculated			
	TN1		TW16		TN1		TW16	
	Reads	Uniques	Reads	Uniques	Reads	Uniques	Reads	Uniques
Mapped to eRTBVL	19,968 (0.11%)	3,434 (0.17%)	31,742 (0.15%)	2,910 (0.14%)	17,967 (0.08%)	1,948 (0.10%)	20,116 (0.09%)	1,833 (0.12%)

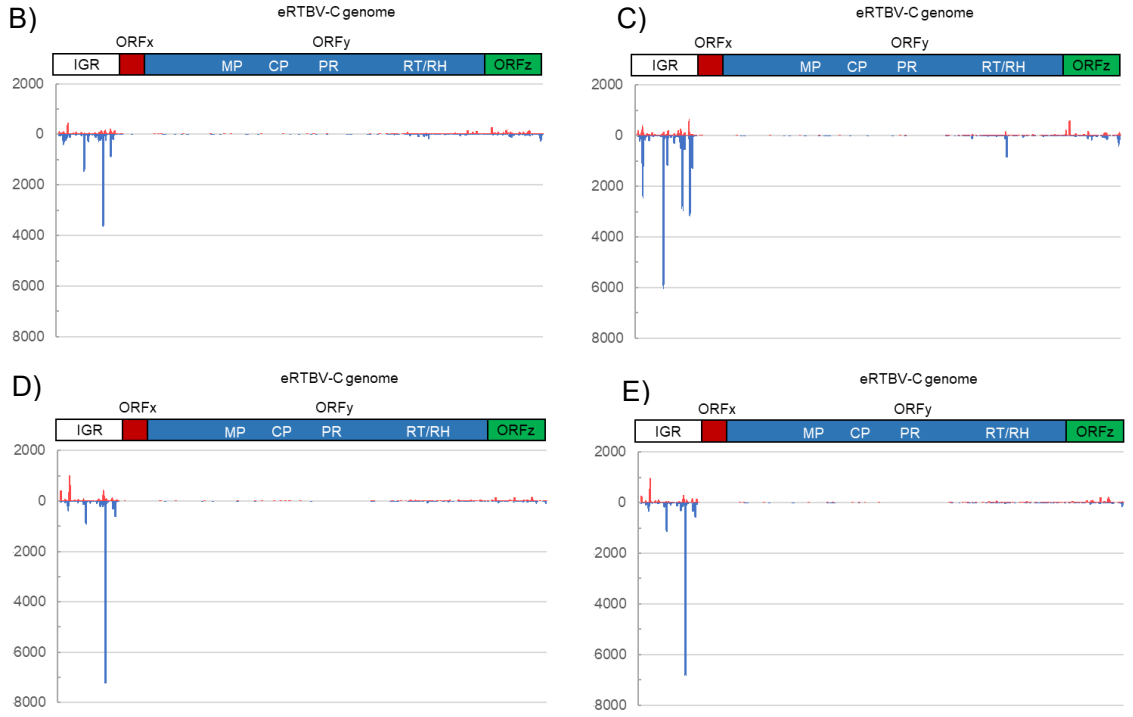


Figure 3.13 Mapping of small RNAs derived from eRTBVL in TN1 and TW16.

A) The number of mapped reads to eRTBVL. Distribution for B) TN1 and for C) TW16 of small RNA abundance along eRTBVL genome in RTBV-infected plants. Distribution for D) TN1 and for E) TW16 of small RNA abundance along eRTBVL genome in mock-inoculated plants.

A)

eRTBVL-derived small RNA sequence	Length (nt)	Position in eRTBVL sequence		
		Start	End	Region
UUCCCUUCGGGUGGAACUGCCU	22	716	738	IGR

B)

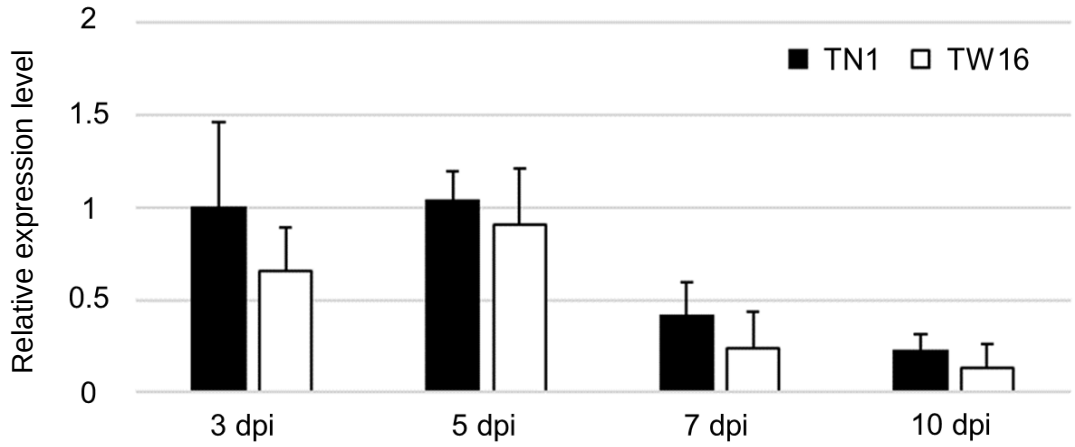


Figure 3.14 The characteristics of IGR-specific small RNA derived from eRTBVL and its expression in infected plants.

A) The ribonucleotide sequence and the position of IGR-specific small RNA derived from eRTBVL. B) Temporary changes of the expression of eRTBVL-derived small RNA described in A. Expression level was calculated by ΔΔCt method. The expression level for mock-inoculated TN1 at 3 dpi is converted to 1.0 in each graph and expression level of other samples was relative value. Vertical lines indicate standard error. Primers used for the expression analysis are presented in Supplementary Table S1.

3.4 Discussion

In this study, the reactions to RTBV such as phenotype, the expression of small RNA and silencing-related genes were compared between the susceptible and resistant plants. This would be the first report about the level of resistance for Utri Merah and TW16 in terms of agronomic traits such as grain quality and yield. Based on the phenotyping, most of the agronomic traits of Utri Merah and TW16 analyzed in this study were not affected by RTBV infection (Table 3.2, Figure 3.5). Harvest index (HI) was not affected by RTBV infection for Utri Merah, of which HI was lower than that of the healthy plants of TN1 (Table 3.2). When TW16, which has genetic background of TN1, was infected by RTBV, TW16 showed a high HI similar to mock-inoculated TN1 plants (Table 3.2). Taking account of RTBV disease resistance, TW16 was the cultivar that holds the high yield performance even when infected by RTBV.

Distribution of vsRNA in the RTBV genome was examined for the RTBV-infected samples. Surprisingly, the very low level of the accumulation of IGR-specific vsRNAs at 10 dpi occurred in TN1 (Figure 3.7A). In contrast, vsRNAs mapped in IGR were abundantly accumulated in TW16 (Figure 3.7B). However, previous studies described that the accumulation of IGR-specific vsRNA play an important role as “decoy” dsRNA to protect the coding sequence (CDS) in the RTBV genome from an anti-viral silencing machinery (Rajeswaran *et al*., 2014; Zarreen *et al*., 2018). These results seemed to be inconsistent with those in this study based on the comparison of the vsRNA accumulation between TW16 and TN1 (Figure 3.7). The previous reports observed the expression of vsRNAs from the RTBV-infected sample collected at 25 dpi, whereas I used 10-dpi plants for the expression analyses. In addition, the expression of vsRNAs covered the entire IGR of the RTBV genome in their reports (Rajeswaran *et al*., 2014; Zarreen *et al*., 2018). Rajeswaran *et al*. (2014) concluded

that RTBV could engage the host defense machinery away from the coding region of the virus genome by restricting the accumulation of vsRNA to the non-coding region in IGR. In this study, the two peaks of the accumulated vsRNAs were observed at IGR. One of the peaks observed at IGR of the RTBV genome was located at the translation initiation site for uORF I, which is essential to translate virus transcripts (Figure 3.8), suggesting an anti-viral defense mechanism against RTBV infection in TW16.

The up-regulation of *AGO1b* and *AGO2* genes occurred in TW16 (RTBV-resistant cultivar) immediately after the RTBV infection (Figure 3.9). The expression of *AGO1b* and *AGO2* genes might lead to the production of huge amount of IGR-specific small RNAs, and then *AGO1b* and *AGO2* may suppress the translation of the virus transcripts, which is essential for RTBV proliferation. The *AGO* genes-mediated silencing might have occurred at the beginning of RTBV infection. In contrast, TN1 (RTBV-sensitive cultivar) showed low accumulation of vsRNAs derived from IGR and the low-expression of *AGO1b* and *AGO2* genes, which are gene silencing-related genes. These results implied that an anti-silencing machinery of TN1 were not functionally active at the early timepoint in RTBV-infected TN1 because there are no reports except this study about the expression of vsRNA from infected rice plants observed at early time point.

I identified 11 miRNAs which might be associated with RTBV infection and resistance (Table 3.8). In rice, many miRNAs that target genes related to development were reported to be regulated by plant pathogen or environmental stress (Archak *et al*., 2007; Lim *et al*., 2009; Zarreen *et al*., 2018; Wang *et al*., 2016). Of 11 miRNAs, only osa-miR399d have been previously characterized that RTBV infection affected to the expression of osa-miR399d. This study demonstrated that osa-miR399d was significantly down-regulated by RTBV single infection, and similar results were obtained upon Tungro (RTSV + RTBV) infection by

Zarreen *et al*., (2018). *PHO2* (LOC_Os05g48390.1, Table 3.8), which was involved in the phosphate starvation response, was targeted by osa-miR399d (Zarreen *et al*., 2018; Archak *et al*., 2007; Bari *et al*., 2006; Hackenberg *et al*., 2013). The molecular function of osa-miR399d is to guide the RISC complex to the *PHO2* mRNA. PHO2 is annotated as a ubiquitin-conjugating E2 enzyme (Bari *et al*., 2006; Hackenberg *et al*., 2013) (Table 3.8).

For the remaining ten miRNAs, no association has been reported with RTBV or tungro disease resistance in previous studies. However, the relationships between these miRNAs and the other virus infections have been described. In this study, the expression of osa-miR444 decreased in the 10-dpi plants by RTBV infection. *Rice stripe virus* infection elevated osa-miR444b.1 and osa-miR444c.2 expressions that repressed the expression of *OsMADS57* and promoted the silencing machinery by increasing the expression of RNA-dependent RNA Polymerase 1 (*OsRDR 1*) (Wang *et al*., 2016). Another miRNA in rice, osa-miR166 (Table 3.8), was also down-regulated in TW16. osa-miR166 targets LOC_Os03g43930 involved in the auxin signaling associated with leaf and shoot development (Itoh *et al*., 2008). Most of the miRNAs identified in this study was down-regulated due to RTBV infection in TW16 rather than in TN1 (Table 3.5, 3.7, Figure 3.10, 3.12). These results implied that the down-regulation of osa-miRNA in TW16 activated an anti-viral system to protect rice plants from RTBV.

The expression of eRTBVL-derived small RNA was examined between TW16 and TN1 to see whether eRTBVL sequences in rice genome are involved in the RTBV resistance. Unlike the comparison between Asian rice and African rice, I could not find any correlation between the expression level and the level of RTBV resistance because the significant differences in the expression of the eRTBVL-derived small RNA were not observed between the two genotypes from 3 to 15 dpi (Figure 3.13B-E, 3.14). The two genotypes were expected

to have the same number of eRTBVL copies, suggesting that the expression level might be dependent on the copy number and not related to RTBV resistance.

In summary, the impact of RTBV infection upon agronomic traits and the in expression of small RNA were compared between rice genotypes resistant and susceptible to RTBV. I proposed the usefulness of TW16 for breeding because the major agronomic traits of TW16 were not affected by RTBV infection. A total of 11 crucial miRNAs related to RTBV resistance were identified in this study. The identified miRNAs might be valuable candidates to study novel insight into the defense mechanisms of genotypes resistant against RTBV.

Chapter 4 Origin of eRTBVL and *Oryza*-AA speciation

4.1 Introduction

eRTBVLs are the traces of viral integration, which has similar nucleotide sequence to RTBV belonging to *Caulimoviridae*, a pararetrovirus family and have been found in rice genome (Kunii *et al*., 2004) (Figure 1.3). They were present in the chromosomes of certain origin of AA-genome species and widely distributed over the species. A total of about 100 copies have been found in the genome of *O. sativa* ssp. *japonica* and approximately 80 copies in *indica*. eRTBVL-D, which was known as an ancient family among the five main eRTBVL families, was integrated into a common ancestor of AA-genome species and spread over the species during speciation (Kunii *et al*., 2004; Chen *et al*., 2014; Liu *et al*., 2012).

A total of 15 eRTBVL-D loci have been found in Nipponbare (*japonica* species) (Figure 4.1) and shared by AA-genome species (Figure 4.2A, Table 4.1) (Chen *et al*., 2014, 2018). Relative to the other eRTBVL families, eRTBVL-D is the oldest family, and the sequences had inserted into the chromosomes of the ancestor species of the *Oryza*-AA genome species. Considering that the speciation of *Oryza*-AA genome species occurred after the integrations of eRTBVL-D sequences, eRTBVL-D sequences are useful in exploring the evolutionary events of local chromosomal regions based on comparisons with general phylogenetic tree of *Oryza* AA-genome species (GPT) (Figure 4.2). This chapter describes evolutionary patterns of the local chromosomal regions during the *Oryza* AA-genome speciation through the comparisons of eRTBVL-D sequences among the *Oryza* AA-genome species.

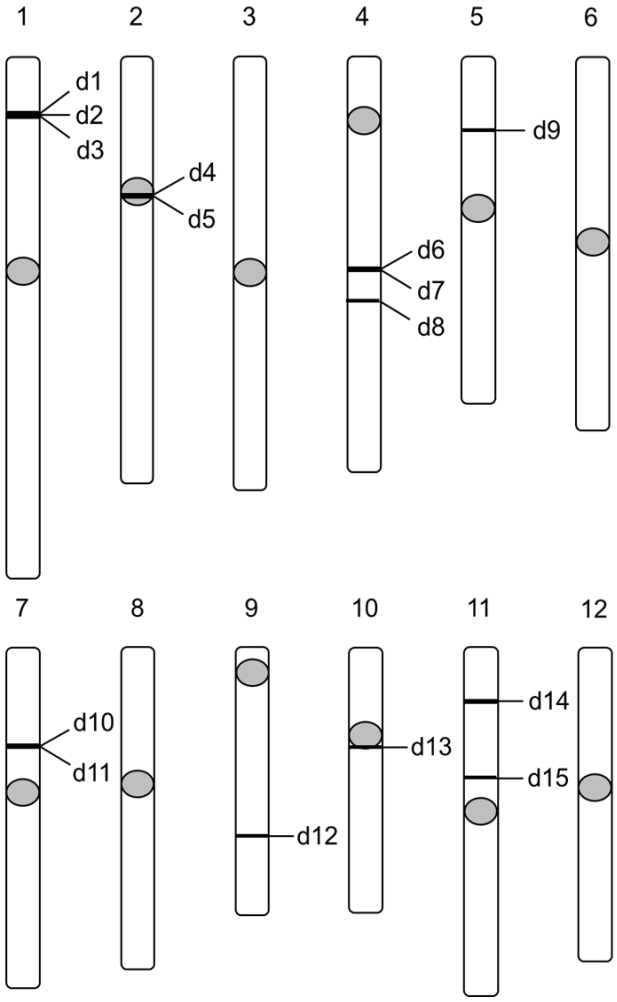


Figure 4.1 Genomic distribution of eRTBVL-D locus on *Oryza sativa* ssp. *japonica*.

Chromosomal numbers are shown at the top of the bar. Horizontal lines indicate the positions of eRTBVL-D loci in *japonica*. Circles indicate the positions of centromere.

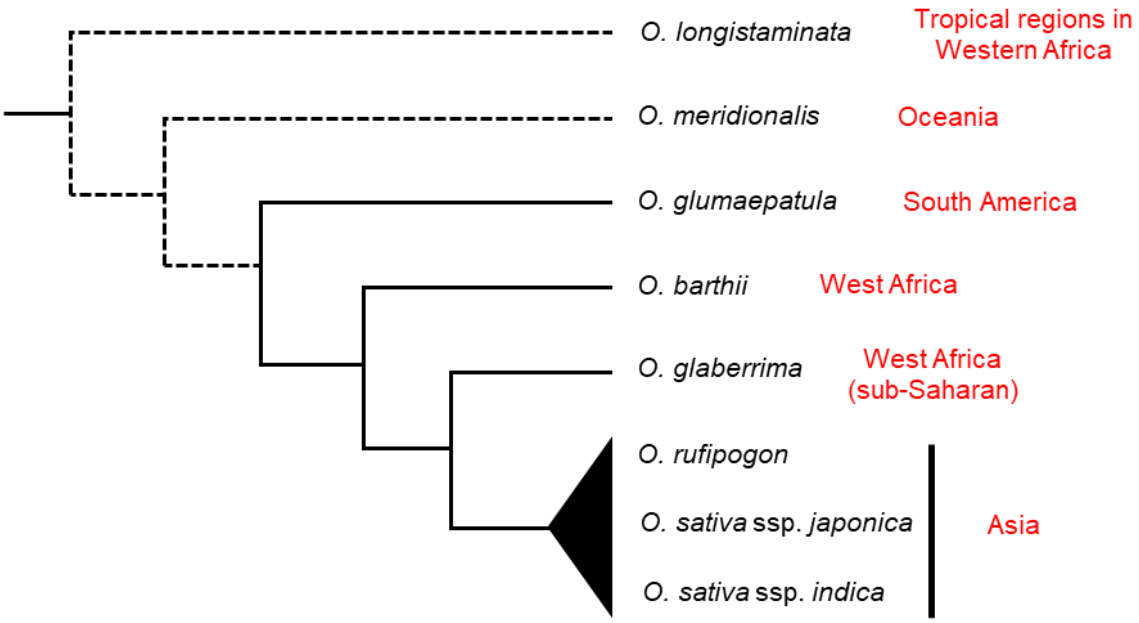


Figure 4.2 General phylogenetic tree (GPT) based on eight *Oryza* AA-species and their origin.

The phylogeny of *Oryza* AA-genome species is displayed according to previous references (Chen *et al*., 2013; Zhu *et al*., 2014). Dotted lines mean that the phylogenetic relationships between *O. longistaminata* and *O. meridionalis* were not fully understood. The original areas of AA-genome species are described at the right in red color.

Table 4.1 The position of eRTBVL-D segments in *Oryza* AA-genome species

ID	Previous name	*O. sativa ssp.* japonica					*O. sativa ssp.* indica					*O. rufipogon*					*O. glaberrima*				
		Chr	Start	End	Length	Strand	Chr	Start	End	Length	Sence	Chr	Start	End	Length	Strand	Chr	Start	End	Length	Strand
d1	JaE1-4a	1	5193920	5194407	488	+	1	5290026	5290512	487	+	1	4757193	4757679	487	+	1	4154711	4155109	399	+
d2	JaE1-4b	1	5194744	5195008	265	+	1	5290849	5291113	265	+	1	4758016	4758280	265	+	1	deleted			+
d3	JaE1-4c	1	5195124	5198055	2932	+	1	5291228	5294159	2932	+	1	4758395	4761331	2937	+	1	4155410	4158083	2674	+
d4	JaE2-3.1a	2	12120385	12120601	217	+	2	12319652	12319868	217	+	2	11229018	11229234	217	+		by PCR			
d5	JaE2-3.1b	2	12120644	12121245	602	−	2	12319911	12320512	602	−	2	11229277	11229878	602	−		by PCR			
d6	JaE4-2.1a	4	18488307	18488450	144	+	4	17481205	17481346	142	+	4	14730131	14730273	143	+	4	11186433	11186575	143	+
d7	JaE4-2.1b	4	18488544	18488674	131	−	4	17481442	17481572	131	−	4	14730369	14730499	131	−	4	11186668	11186799	132	−
d8	JaE4-3.1	4	21097520	21097778	259	−	4	20464869	20465127	259	−	4	17280374	17280631	258	−	4	13690391	13690644	254	−
d9	JaE5-0.1	5	6506720	6507050	331	−	5	6544418	6544748	331	−	5	5838059	5838389	331	−		by PCR			
d10	JaE7-4	7	8906209	8907980	1772	+	7	8659380	8664699	5320	+	7	8096430	8101787	5358	+	7	7642657	7648227	5571	+
d11	JaE7-5&JaE7-6	7	8909007	8920257	11251	−	7	8665267	8676319	11053	−	7	8102844	8113944	11101	−	7	7662915	7673393	10479	−
d12	JaE9-5.1	9	16726920	16727025	106	+	9	15848884	15848989	106	+	9	14423721	14423826	106	+	9	12218973	12219078	106	+
d13	JaE10-0.1	10	9099571	9099688	118	−	10	8894585	8894702	118	−	10	7953593	7953710	118	−		by PCR			
d14	JaE11-2	11	5069912	5072964	3053	+	11	5391598	5392179	582	+	11	4617423	4618004	582	+	Oglab11_unplaced045	205	3284	3080	+
d15	JaE11-5	11	11654495	11657114	2620	−	11	11879939	11882560	2622	−	11	11220479	11223106	2628	−	11	9061408	9063961	2554	−

ID	Previous name	*O. barthii*					*O. glumaepatula*					*O. meridionalis*					*O. longistaminata*				
		Chr	Start	End	Length	Strand	Chr	Start	End	Length	Strand	Chr	Start	End	Length	Strand	Chr	Start	End	Length	Strand
d1	JaE1-4a	1	4387743	4388227	485	+	1	5279656	5280128	473	+	1	5261177	5261700	524	+	1	5041899	5042383	485	+
d2	JaE1-4b			deleted					deleted			1	5262036	5262300	265			Deleted*			
d3	JaE1-4c	1	4385375	4388048	2674	+	1	5280432	5282642	2211	+	1	5262415	5264151	1737	+	1	5039533	5042204	2672	+
d4	JaE2-3.1a	2	11025428	11025644	217	+	2	12220660	12220876	217	+		by PCR					Deleted*			
d5	JaE2-3.1b	2	11025687	11026288	602	−	2	12220919	12221520	602	−		by PCR					Deleted*			
d6	JaE4-2.1a	4	12208515	12208658	144	+	4	15674968	15675110	143	+	4	14452390	14452532	143	+	4	12251597	12251739	143	+
d7	JaE4-2.1b	4	12208751	12208882	132	−	4	15675205	15675336	132	−	4	14452628	14452758	131	−	4	12251833	12251963	131	−
d8	JaE4-3.1	4	14694693	14694946	254	−			by PCR			4	17044210	17044517	308	−	4	14877184	14877441	258	−
d9	JaE5-0.1	5	5433872	5434188	317	−	5	6588049	6588341	293	−		Unknwon				5*	5522035	5522328	294	−
d10	JaE7-4	7	7798798	7804120	5323	+	1	39940422	39945744	5323	−	LONC01000297	254106	259439	5334	−		Unknwon			
d11	JaE7-5&JaE7-6	7	7818783	7829197	10415	−	1	39912399	39939365	26967	+	PCR&LONC01000297	228171	253049	24879	+		Unknwon			
d12	JaE9-5.1	9	13332228	13332333	106	+	9	16644418	16644523	106	+	9	13756069	13756174	106	+	9*	14346341	14346446	106	+
d13	JaE10-0.1	10	7045450	7045567	118	−	10	8287131	8287247	117	−	10	5812893	5813010	118	−	10	13820912	13821017	106	−
d14	JaE11-2	11	4551560	4554632	3073	+	11	4748694	4749289	596	+	11	4423885	4423992	108	+	11*	4400316	4400872	557	+
d15	JaE11-5	11	9891208	9893829	2622	−	11	11473849	11476492	2644	−		Unknwon				11*	9393575	9395464	1890	−

4.2 Materials and Methods

4.2.1 Plant materials and taxonomy

Whole genome sequences of *Oryza* species used in this study were listed in Table 4.2. Eight *Oryza* AA-genome species, *O. sativa* ssp. *japonica* and *indica*, *O. rufipogon*, *O. glaberrima*, *O. barthii*, *O. glumaepatula*, *O. meridionalis* and *O. longistaminata*, and then *O. punctata* (BB-genome), of which genome contains a few eRTBVL-sequences, were also used in this study (Table 4.1). Rice tungro bacilliform virus (RTBV) sequence (NCBI accession: NC_001914.1) was used as an outgroup for several analyses. For eRTBVL family, the consensus sequences of eRTBVL-A, B, C were obtained from NCBI (NCBI accession: BR000029.1, BR000030.1, and BR000031.1) and the consensus sequence of eRTBVL-X used in this study was mentioned by Chen *et al*., (2018).

Table 4.2 Information regarding the *Oryza* AA-genome assemblies examined in this study

Species	Cultivar/Accession/Sample name	Data source	Data version
O. sativa ssp. japonica	Nipponbare	Gramene	IRGSP-1.0
O. sativa ssp. indica	93-11	Molecular Plant	v2Plus
O. sativa ssp. indica	93-11	Gramene	ASM465v1
O. rufipogon	W1943	Gramene	OR_W1943
O. glaberrima	IRGC 96717	Gramene	AGI1.1
O. barthii	IRGC 105608	Gramene	v1
O. glumaepatula	GEN1233_2	Gramene	v1.5
O. meridionalis	W2112	Gramene	v1.3
O. meridionalis	Taxon B	NCBI	LONC00000000.1
O. longistainata	RD23	NCBI	LQBC00000000.1
O. longistainata	SAMN03247844	Gramene	v1.0
O. longistainata	IRGC110404	http://olinfres.nig.ac.jp	v2.0
O. punctata	IRGC105690	Gramene	v1.2

4.2.2 Collection of genomic data

The chromosomal locations of eRTBVL-D sequences in *Oryza* AA-genome species except for *O. longistaminata* have been published by Chen *et al*., (2018) (Table 4.1). The eRTBVL-D loci in *japonica* genome were also mapped to these references, and the highly reliable hit for each locus (e-values < 1 x 10^{-3} and lengths > 100 bp) was extracted as an eRTBVL-D ortholog when the flanking sequence was properly contiguous to eRTBVL-D sequence. The two reference genomes of *O. longistaminata* were obtained from BLASTN (NCBI-blast+ 2.9.0). Both sides flanking 5000 bp sequences of eRTBVL-D sequences were extracted from the *japonica* genome. When eRTBVL-D locus was not well mapped to either reference genome of *O. longistaminata*, I determined the presence or absence of this D locus in *O. longistaminata* by confirming whether both sides of the flanking sequences corresponded to those of the *japonica* eRTBVL-D locus.

To collect the sequence of flanking genes, I firstly found the closest gene to eRTBVL-D locus in *O. sativa* ssp. *japonica* and then the orthologous genes in *Oryza*-AA genome species were searched with Gramene (http://gramene.org/). The second flanking gene was used for analysis when the closest one has less than five orthologs in the AA-genome species.

4.2.4 Phylogenetic analysis

A set of the orthologous sequences of *Oryza* AA-genome species such as the eRTBVL-D locus, flanking sequence, and gene were applied to phylogenetic analysis. PR, RT/RH and ORFz region of five eRTBVL families were also used to refer to the phylogenetic tree. The five RTBVL-D sequences at d2, d8, d9, d12 and d13 loci were removed from data sets due to insufficient length for phylogenetic analysis. The two pairs of RTBVL-D

sequences at the d3/d4 and d6/d7 loci with intervals shorter than 200 bp were combined into the respective single sequence for the analysis. The sequence alignment of each data set was generated in MUSCLE (Edgar, 2004) and all gaps were removed. Maximum likelihood (ML) phylogenetic analysis was performed and the best-fitting substitution models for each locus were determined by model selection analysis. Support for the ML trees was evaluated by 1000 bootstrap replicates. All analyses related to phylogenetic analysis were completed by MEGA version 7.0 (Kumar, 2016).

4.2.5 Gene structure analysis

The structural rearrangements such as insertion, deletion, transformation and inversion were examined based on the nucleotide sequence. Bowtie and BLASTN (NCBI-blast+ 2.9.0) were used for the comparison of the chromosomal structure around eRTBVL-D loci including the flanking sequences among AA-genome species.

4.2.6 Tajima's D analysis

A total of 240 accessions each of *japonica* and *indica* (80 each from subtropical, temperate, and tropical *japonica*, and then 60 each from ind1a, ind1b, ind2 and ind3) were randomly selected from SNP-seek database (https://snp-seek.irri.org/) (Supplementary Table S2, S3). SNP dataset of eRTBVL-D loci, the flanking sequence and gene was collected from the database and Tajima's D (Tajima, 1989) statistical analysis was performed by TASSEL 5 (Bradbury, 2007) following the default option. A sliding window analysis (slide size = 20 SNPs, step size = 1 SNP each) based on the number of polymorphic sites across the region of both franking gene and eRTBVL-D locus was completed with TASSEL 5. The standard

neutral model and confidence limit for Tajima's D value were predicted based on the coalescent simulation by DNasp v5.0 and *ms*.

4.3 Results

4.3.1 Phylogenetic relationships of eRTBVL families

A consensus sequence of eRTBVL-D has not been assembled, because the 15 eRTBVL-D sequences in *japonica* genome does not fully cover a whole sequence of the eRTBVL genomes (Chen *et al*., 2018). Thus, 15 eRTBVL-D sequences of *japonica* genome were available as the segments of the eRTBVL-D family. The 15 eRTBVL-D sequences corresponded to the functional regions for eRTBVL as follows: d14 and d15 to PR, d3, d11, d14 and d15 to RT/RH, and d3, d10, d11, d14 and d15 to ORFz/ORF IV. The datasets of five eRTBVL families were aligned to the PR, RT/RH, and ORFz/ORF IV regions. The phylogenetic trees were constructed to analyze the characteristics of each family (Figure 4.3 and Supplementary Figure S1-S3). The three phylogenetic trees displayed that eRTBVL-D was more distantly related to the other eRTBVL families, and eRTBVL-A and B, X and C were topologically closer (Figure 4.3). My phylogenetic trees of PR, RT/RH, and ORFz/ORF IV supported the previous report that eRTBVL-D was integrated into rice genome before the integration of other families (Chen *et al*., 2017, 2018).

4.3.2 Distribution of eRTBVL-D sequences in *O. longistaminata*

The accurate positions of eRTBVL-D loci in *O. longistaminata* were unclear. According to orthologous analysis based on the eRTBVL-D sequences of *japonica* species, the presence of 10 loci and the absence of three loci were identified in the *O. longistaminata* genome (Figure 4.4, Table 4.1). The d4 and d5 loci in *O. longistaminata* were not found in

the reference nor detected by PCR using primers designed from *japonica* sequence, while the d4 and d5 in *O. meridionalis* and *O. glaberrima* were detected by PCR only using primers designed from *japonica* sequence (Chen *et al*., 2018) (Table 4.1). Thus, I judged the eRTBVL-D sequences from d4 and d5.

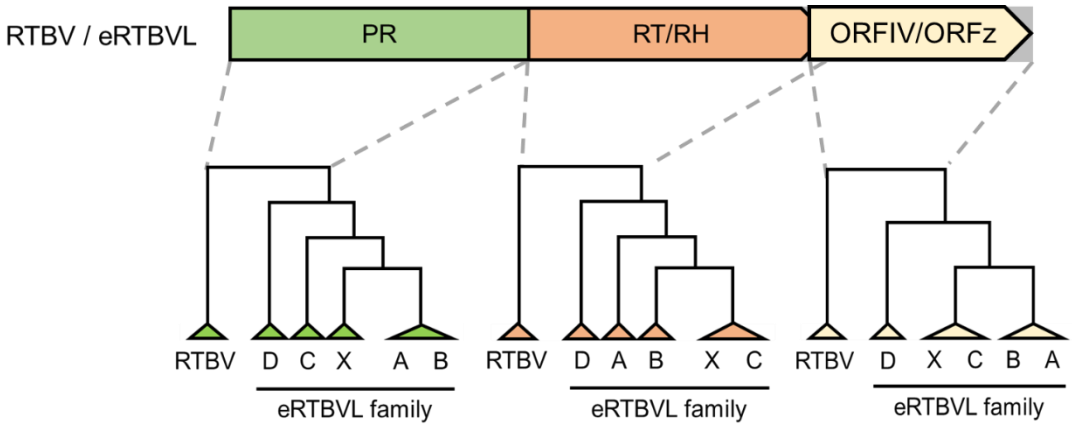
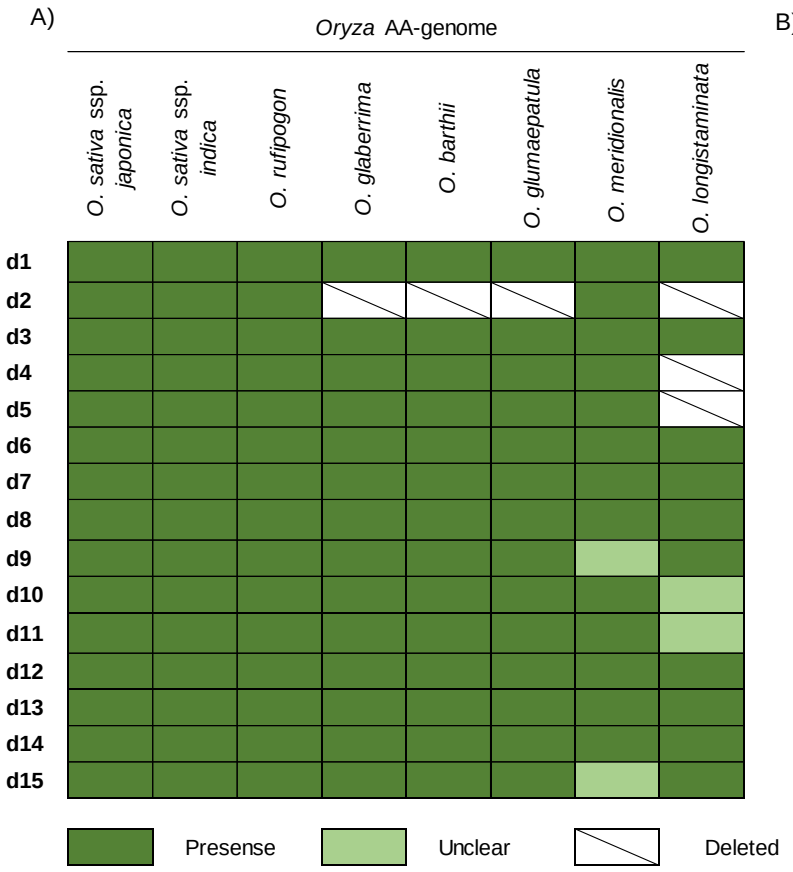


Figure 4.3 Phylogenetic relationships between RTBV and eRTBVL based on the nucleotide sequence of three regions of the RTBV/eRTBVL sequence.

The eRTBVL structure is shown schematically with the summarized phylogenetic trees (midpoint rooted for display purpose) for PR, RT/RH and ORF IV/ORFz. The details of alignment are presented in Supplementary Figure S1-S3.



B)

Phylogenetic Analysis	
The number of available sequence	Type of phylogenetic tree
8	B
4	n.d.
8	B
7	A
7	
8	B
8	
8	n.d.
7	n.d.
7	A
7	B
8	n.d.
8	n.d.
8	B
7	A

Figure 4.4 The information about A) sequence set for phylogenetic analysis of each eRTBVL-D locus and B) the classification of phylogeny.

The number of available sequences indicates the number of AA-genome species which have number of the eRTBVL-D locus in the genomes of the AA-genome species. n.d. represents that a set of eRTBVL-D loci reported in the left was not applied for the phylogenetic analysis. Phylogenetic trees based on eRTBVL-D loci were divided into type A and B. Phylogenetic trees classified into type A are almost topologically consistent with GPT. Type B represents phylogenetic trees which are different in shape from GPT.

4.3.3 Phylogenetic tree analysis based on eRTBVL-D sequences

The phylogenetic trees of eight eRTBVL-D loci were classified into two types based on GPT of *Oryza* AA-genome species (Figure 4.4), although the sequences for d2, d8, d9, d12, and d13 were not used in the phylogenetic analysis because of the insufficient length for the construction of a reliable tree. The trees classified into Type A (d4-5, d10, and d15), which were almost topologically consistent with GPT. The trees (d1, d3, d6-7, d11, and d14) in Type B were different in shape from GPT (Figure 4.5, 4.6). The trees of Type B exhibited the structures varied among the five RTBVL-D loci. In Type B, the four phylogenetic trees (Figure 4.6A, B, C, D) showed that the close relationship was observed between *O. meridionalis*/*O. glumaepatula* and Asian rice species, which have more distant relationships in GPT. The phylogenetic relationships of the AA-genome species in d11 were according to their geographical clusters (Figure 4.6D). In the tree for d6-7, *O. longistaminata* was a sister group of *japonica* with high value of bootstrap support (Figure 4.6C). The d14 tree presented *indica* located in the same clade with *O. longistaminata*, which should be distantly related to *O. sativa* (Figure 4.6E). The eRTBVL-D loci gave rise to various patterns of phylogenetic tree. Five out of eight trees were inconsistent with GPT of *Oryza* AA-genome speciation even though eRTBVL-D expected not to have certain functions for the host plants. In particular, *O. sativa*/*O. rufipogon* and *O. glaberrima*/*O. barthii* always belong in similar phylogenetic groups, while their relationships with *O. glumaepatula*, *O. meridionalis*, and *O. longistaminata* were inconsistent with that of GPT.

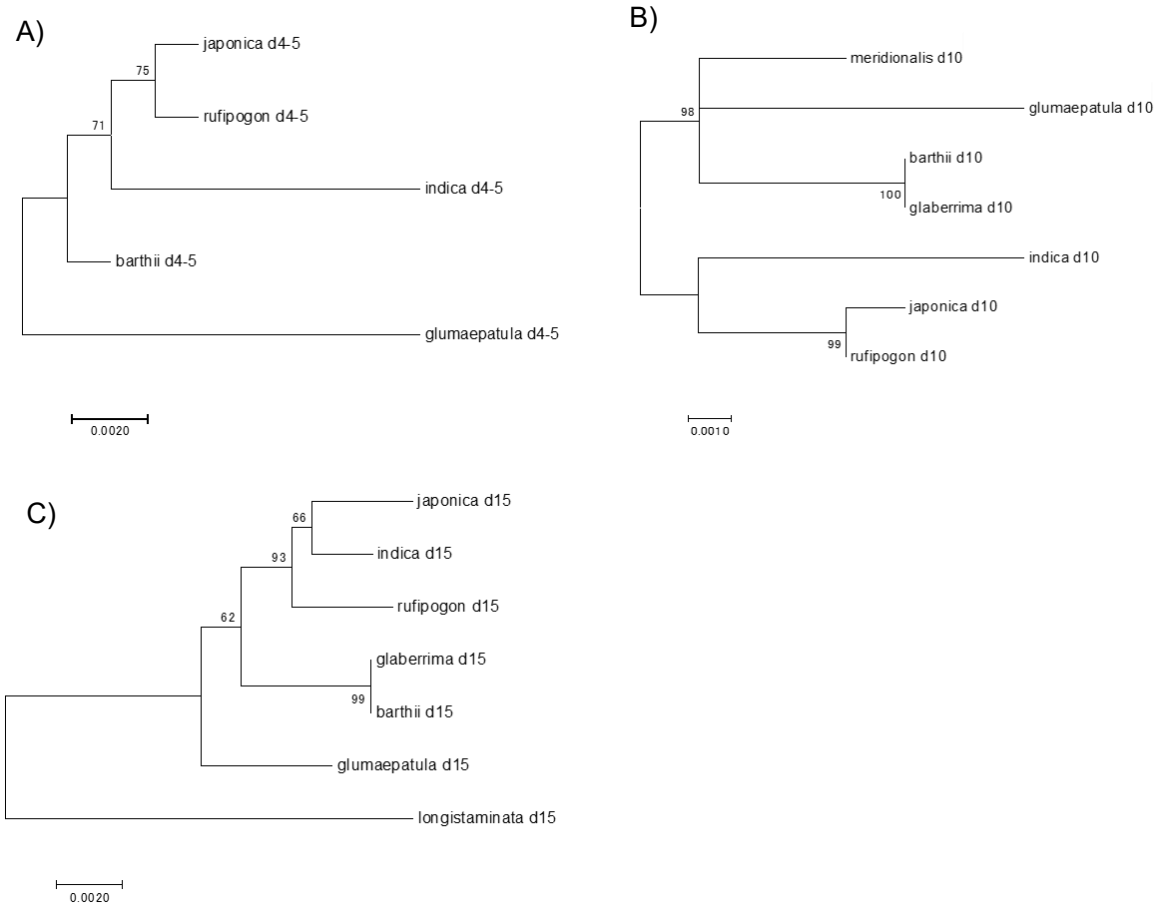


Figure 4.5 The phylogenetic tree of eRTBVL-D locus classified into Type A in *Oryza* AA-genome species.

Phylogenetic trees are constructed based on A) d4-5 loci, B) d10 loci, C) d15 loci. The numbers above the branch is the value of bootstrap. Bootstrap values more than 60 were indicated in this figure. The tree was drawn to scale, with branch lengths measured in the number of substitutions per site.

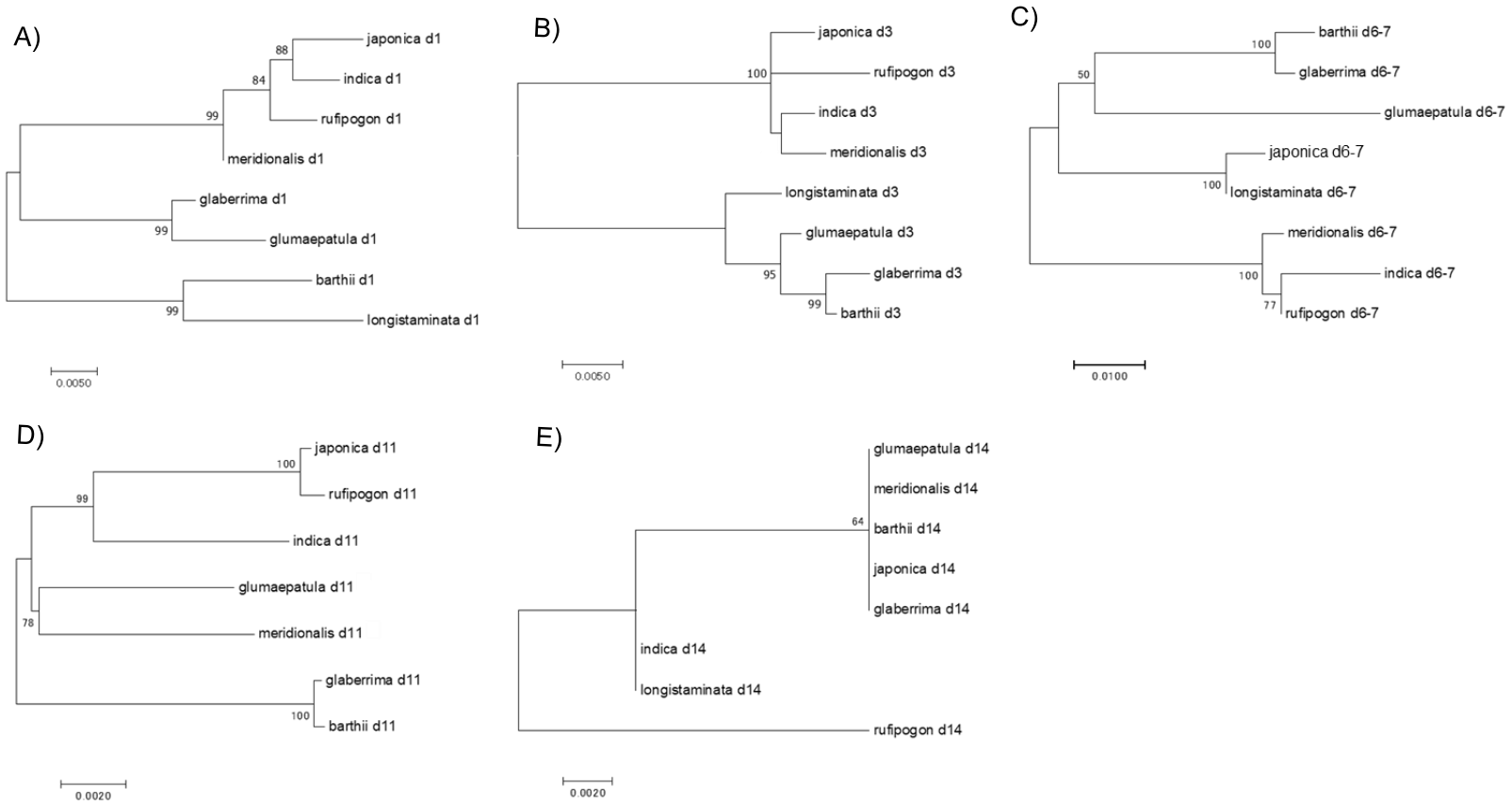


Figure 4.6 Phylogenetic tree of eRTBVL-D locus classified into Type B in *Oryza* AA-genome species.

Phylogenetic trees are constructed based on A) d1 locus, B) d3 locus, C) d6-7 locus, D) d11 locus, E) d14 locus. Bootstrap values greater than 60 was shown in this figure. The tree was drawn to scale, with branch lengths measured in the number of substitutions per site.

4.3.4 Phylogenetic relationships of the flanking genes of eRTBVL-D loci

To test whether the adjacent genes of eRTBVL-D sequences exhibited the same evolutionary patterns, phylogenetic trees were built using the upstream and downstream of flanking gene of each eRTBVL-D locus. Since the eRTBVL loci of d1, d2 and d3 are clustered in chromosome 1, I combined these three eRTBVL-D loci. The franking genes were obtained from upstream of d1 and downstream of d3. By the same reason, d10 and d11 located in chromosome 7 were considered as a single region for phylogenetic analysis of the flanking genes. The details of flanking genes were shown in Table 4.3. The phylogenetic trees for the flanking gene of d4-5 and d10 classified into Type A showed different phylogenetic relationship with GPT (Figure 4.7). The phylogenetic trees for the flanking gene of Type B were compared with that for eRTBVL-D sequences in Type B (Figure 4.6, 4.8). Either or both of the trees based on two flanking genes for all the eRTBVL-D loci in Type B showed similar phylogeny with that of corresponding eRTBVL-D (Figure 4.6, 4.8). This result revealed that the chromosomal regions containing the eRTBVL Type B sequences were unlikely to have evolved synchronizing with GPT.

Table 4.3 The nucleotide position of the flanking gene of eRTBVL-D locus and its ortholog in AA-genome species

		O. sativa ssp. *japonica*			*O. sativa* ssp. *indica*			*O. rufipogon*		*O. glaberrima*	
		Flanking gene of eRTBVL loci	gramene	The distance from eRTBV	gene_ID	gramene	v2Plus	gene_ID	gramene	gene_ID	gramene
d1-3	Upstream	Os01g0195801	1: 5,149,335-5,150,341:+	43579, 44403, 44783	BGIOSGA002940	1:5,542,296-5,543,724:+	1:5239359-5240787:+	ND	ND	ORGLA01G0397100	Oglab01_unplaced002:4,132-5,579:+
	Downstream	Os01g0196300	1: 5,201,862-5,203,996:+	7455, 6854, 3807	BGIOSGA002943	1:5,598,865-5,600,573:+	1:5298194-5299902:+	ORUFI01G06710	1:4,765,164-4,767,170:+	ORGLA01G0058200	1:4,161,901-4,163,606:+
d4-5	Upstream	Os02g0308400	2: 12,113,907-12,114,966:-	5419, 5678	BGIOSGA006614	2:12,920,636-12,921,439:-	2:12312483-12313286:-	ORUFI02G13950	2:11,222,821-11,223,624:-	-	-
	Downstream	Os02g0308800	2: 12,127,913-12,129,233:-	7312, 6668	ND	ND	ND	ORUFI02G13970	2:11,236,928-11,237,734:-	ORGLA02G0121500	2:10,596,792-10,597,598:-
d6-7	Upstream	Os04g0378200	4: 18,468,243-18,473,626:+	14681	BGIOSGA027847	8:415,571-417,553:+	4:406941-408923:+	ORUFI04G10880	4:14,701,139-14,706,241:+	ORGLA04G0069700	4:11,134,896-11,139,127:+
	Downstream	Os04g0378575	4: 18,490,353-18,494,159:+	1903, 1679	BGIOSGA016228	4:15,948,469-15,952,275:+	4:17483257-17487063:+	ORUFI04G10900	4:14,732,181-14,735,987:+	ORGLA02G0344000	ADWL01005695.1:624-2,750:+
d8	Upstream	Os04g0425100	4: 21,001,722-21,010,737:+	86783	BGIOSGA016418	4: 18,846,334-18,853,763	4:20364069-20371498	ORUFI04G13880	4:17,198,789-17,205,317:+	ORGLA04G0096900	4:13,631,890-13,637,638:+
	Downstream	Os04g0428300	4:21,207,698-21,208,285:+	110178	BGIOSGA016422	4: 19,112,498-19,113,757	4:20585589-20586848	ND	ND	ND	ND
d9	Upstream	Os05g0203900	5: 6,468,141-6,476,455:-	8332	BGIOSGA018620	5:6,730,218-6,733,145:-	5:6506327-6509254:-	ORUFI05G07110	5:5,789,582-5,799,014:-	ND	ND
	Downstream	Os05g0204600	5: 6,514,746-6,517,280:+	7696	BGIOSGA019342	5:6,777,434-6,778,939:+	5:6552673-6554178:+	ORUFI05G07170	5:5,846,088-5,848,621:+	ND	ND
d10-11	Upstream	Os07g0257200	7: 8,871,643-8,878,905:-	27304, 30102	BGIOSGA024510	7:8,883,713-8,890,474:-	7:8627398-8634159:-	ORUFI07G09370	7:8,065,847-8,072,635:-	ORGLA07G0078500	7:7,601,244-7,607,887:-
	Downstream	Os07g0258000	7: 8,926,134-8,928,872:+	18154, 5877	ND	ND	ND	ORUFI07G09390	7:8,119,821-8,122,749:+	ORGLA07G0078800	7:7,679,419-7,681,845:+
d12	Upstream	Os09g0447500	9: 16,718,633-16,720,923:-	5997	BGIOSGA029663	9:15,266,334-15,268,285:-	9:15836065-15838016:-	ORUFI09G13070	9:14,415,795-14,417,744:-	ORGLA09G0091600	9:12,211,029-12,212,977:-
	Downstream	Os09g0447600	9: 16,728,178-16,729,625:-	1153	BGIOSGA029662	9:15,280,743-15,281,858:-	12:15850474-15851589:-	ORUFI09G13080	9:14,425,325-14,426,441:-	ND	ND
d13	Upstream	Os10g0326900	10: 9,088,337-9,090,418:-	11234	BGIOSGA032140	10: 7,466,884-7,472,982:-	10:8883321-8889418:-	ORUFI10G06190	10:7,942,307-7,958,996:-	ORGLA10G0048700	10:6,997,079-7,003,180:-
	Downstream	Os10g0327000	10: 9,101,768-9,104,973:-	2080	BGIOSGA032139	10:7,480,575-7,483,786:-	10:8896783-8899994:-	ND	ND	ORGLA10G0048800	10:7,007,870-7,011,075:-
d14	Upstream	Os11g0200600	11: 5,059,892-5,063,060:+	10020	BGIOSGA030692	9:11,823,605-11,825,537:+	11:5376682-5378614:-	ORUFI11G05930 ORUFI11G05940	11:4,611,928-4,615,055:+ 11:4,612,166-4,615,055:+	ORGLA11G0054600	11:4,157,150-4,158,259:+
	Downstream	Os11g0202000	11: 5,153,098-5,156,354:+	80134	BGIOSGA003295	1:11,319,223-11,326,804:+	11:5471217-5478798:-	ND	ND	ND	ND
d15	Upstream	Os11g0307300	11: 11,650,220-11,651,887:-	2608	BGIOSGA034101	11:10,453,094-10,454,377:-	11:11875976-11877259:-	ORUFI11G11350	11:11,216,370-11,217,653:-	ORGLA11G0092500	11:9,057,475-9,058,758:-
	Downstream	Os11g0307500	11: 11,661,330-11,662,253:+	4216	BGIOSGA035175	11:10,474,431-10,480,033:+	11:11893686-11898084:+	ORUFI11G11360	11:11,227,126-11,228,517:+	ORGLA11G0092700	11:9,068,311-9,069,041:+

ND, No identified orthologous gene

Table 4.3 (continued) The nucleotide position of the flanking gene of eRTBVL-D locus and its ortholog in AA-genome species

		O. sativa ssp. *japonica*			*O. sativa* ssp. *indica*			*O. rufipogon*		*O. glaberrima*	
		Flanking gene of eRTBVL loci	gramene	The distance from eRTBV	gene_ID	gramene	v2Plus	gene_ID	gramene	gene_ID	gramene
d1-3	Upstream	Os01g0195801	1: 5,149,335-5,150,341:+	43579, 44403, 44783	BGIOSGA002940	1:5,542,296-5,543,724:+	1:5239359-5240787:+	ND	ND	ORGLA01G0397100	Oglab01_unplaced002:4,132-5,579:+
	Downstream	Os01g0196300	1: 5,201,862-5,203,996:+	7455, 6854, 3807	BGIOSGA002943	1:5,598,865-5,600,573:+	1:5298194-5299902:+	ORUFI01G06710	1:4,765,164-4,767,170:+	ORGLA01G0058200	1:4,161,901-4,163,606:+
d4-5	Upstream	Os02g0308400	2: 12,113,907-12,114,966:-	5419, 5678	BGIOSGA006614	2:12,920,636-12,921,439:-	2:12312483-12313286:-	ORUFI02G13950	2:11,222,821-11,223,624:-	-	-
	Downstream	Os02g0308800	2: 12,127,913-12,129,233:-	7312, 6668	ND	ND	ND	ORUFI02G13970	2:11,236,928-11,237,734:-	ORGLA02G0121500	2:10,596,792-10,597,598:-
d6-7	Upstream	Os04g0378200	4: 18,468,243-18,473,626:+	14681	BGIOSGA027847	8:415,571-417,553:+	4:406941-408923:+	ORUFI04G10880	4:14,701,139-14,706,241:+	ORGLA04G0069700	4:11,134,896-11,139,127:+
	Downstream	Os04g0378575	4: 18,490,353-18,494,159:+	1903, 1679	BGIOSGA016228	4:15,948,469-15,952,275:+	4:17483257-17487063:+	ORUFI04G10900	4:14,732,181-14,735,987:+	ORGLA02G0344000	ADWL01005695.1:624-2,750:+
d8	Upstream	Os04g0425100	4: 21,001,722-21,010,737:+	86783	BGIOSGA016418	4: 18,846,334-18,853,763	4:20364069-20371498	ORUFI04G13880	4:17,198,789-17,205,317:+	ORGLA04G0096900	4:13,631,890-13,637,638:+
	Downstream	Os04g0428300	4:21,207,698-21,208,285:+	110178	BGIOSGA016422	4: 19,112,498-19,113,757	4:20585589-20586848	ND	ND	ND	ND
d9	Upstream	Os05g0203900	5: 6,468,141-6,476,455:-	8332	BGIOSGA018620	5:6,730,218-6,733,145:-	5:6506327-6509254:-	ORUFI05G07110	5:5,789,582-5,799,014:-	ND	ND
	Downstream	Os05g0204600	5: 6,514,746-6,517,280:+	7696	BGIOSGA019342	5:6,777,434-6,778,939:+	5:6552673-6554178:+	ORUFI05G07170	5:5,846,088-5,848,621:+	ND	ND
d10-11	Upstream	Os07g0257200	7: 8,871,643-8,878,905:-	27304, 30102	BGIOSGA024510	7:8,883,713-8,890,474:-	7:8627398-8634159:-	ORUFI07G09370	7:8,065,847-8,072,635:-	ORGLA07G0078500	7:7,601,244-7,607,887:-
	Downstream	Os07g0258000	7: 8,926,134-8,928,872:+	18154, 5877	ND	ND	ND	ORUFI07G09390	7:8,119,821-8,122,749:+	ORGLA07G0078800	7:7,679,419-7,681,845:+
d12	Upstream	Os09g0447500	9: 16,718,633-16,720,923:-	5997	BGIOSGA029663	9:15,266,334-15,268,285:-	9:15836065-15838016:-	ORUFI09G13070	9:14,415,795-14,417,744:-	ORGLA09G0091600	9:12,211,029-12,212,977:-
	Downstream	Os09g0447600	9: 16,728,178-16,729,625:-	1153	BGIOSGA029662	9:15,280,743-15,281,858:-	12:15850474-15851589:-	ORUFI09G13080	9:14,425,325-14,426,441:-	ND	ND
d13	Upstream	Os10g0326900	10: 9,088,337-9,090,418:-	11234	BGIOSGA032140	10: 7,466,884-7,472,982:-	10:8883321-8889418:-	ORUFI10G06190	10:7,942,307-7,958,996:-	ORGLA10G0048700	10:6,997,079-7,003,180:-
	Downstream	Os10g0327000	10: 9,101,768-9,104,973:-	2080	BGIOSGA032139	10:7,480,575-7,483,786:-	10:8896783-8899994:-	ND	ND	ORGLA10G0048800	10:7,007,870-7,011,075:-
d14	Upstream	Os11g0200600	11: 5,059,892-5,063,060:+	10020	BGIOSGA030692	9:11,823,605-11,825,537:+	11:5376682-5378614:-	ORUFI11G05930 ORUFI11G05940	11:4,611,928-4,615,055:+ 11:4,612,166-4,615,055:+	ORGLA11G0054600	11:4,157,150-4,158,259:+
	Downstream	Os11g0202000	11: 5,153,098-5,156,354:+	80134	BGIOSGA003295	1:11,319,223-11,326,804:+	11:5471217-5478798:-	ND	ND	ND	ND
d15	Upstream	Os11g0307300	11: 11,650,220-11,651,887:-	2608	BGIOSGA034101	11:10,453,094-10,454,377:-	11:11875976-11877259:-	ORUFI11G11350	11:11,216,370-11,217,653:-	ORGLA11G0092500	11:9,057,475-9,058,758:-
	Downstream	Os11g0307500	11: 11,661,330-11,662,253:+	4216	BGIOSGA035175	11:10,474,431-10,480,033:+	11:11893686-11898084:+	ORUFI11G11360	11:11,227,126-11,228,517:+	ORGLA11G0092700	11:9,068,311-9,069,041:+

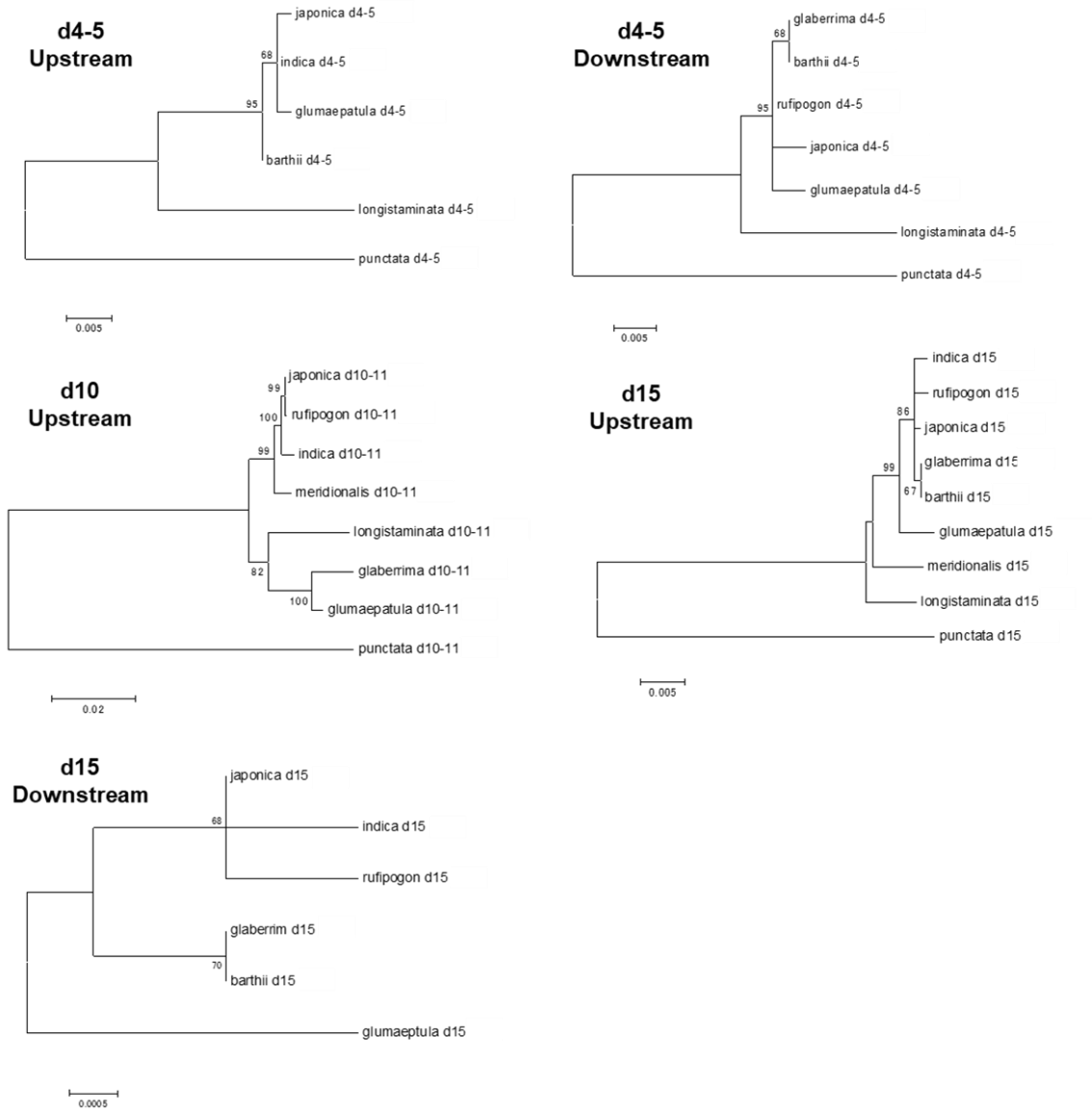


Figure 4.7 Phylogenetic tree of the flanking gene for eRTBVL-D locus classified into Type A in *Oryza* AA-genome species.

Phylogenetic trees are constructed based on the upstream and downstream flanking genes for d4-5, d10, and d15 locus. The bootstrap values greater than 60 are shown above the branches. The tree was drawn to scale, with branch lengths measured in the number of substitutions per site.

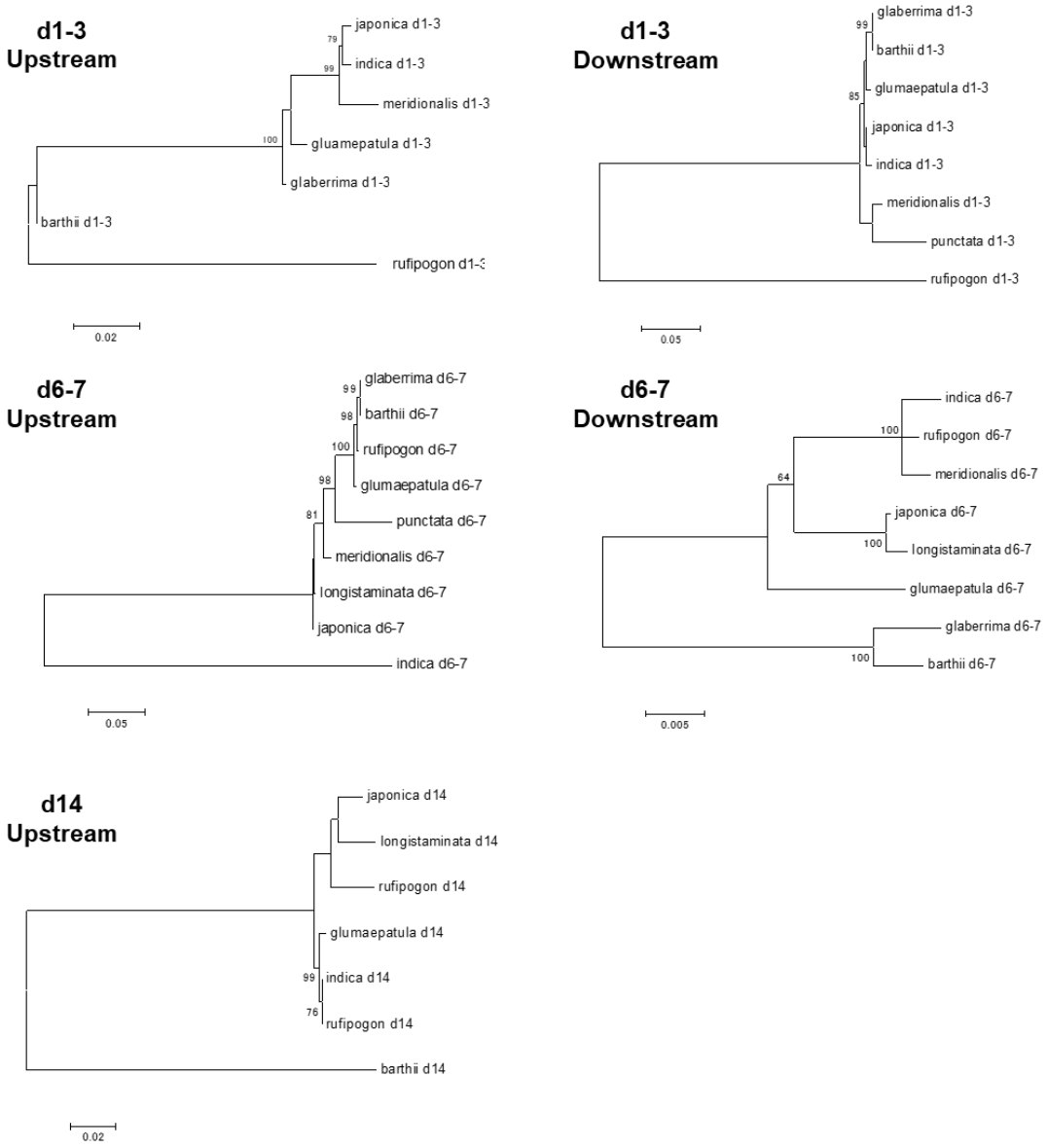


Figure 4.8 The phylogenetic tree of the flanking gene for eRTBVL-D locus classified into Type B in *Oryza* AA-genome species.

Phylogenetic trees are constructed based on the upstream and downstream flanking genes for d1-3, d6-7, and d14 locus. The bootstrap values greater than 60 are shown above the branches. The tree was drawn to scale, with branch lengths measured in the number of substitutions per site.

4.3.5 Chromosomal rearrangements of the region containing eRTBVL-D sequences and flanking genes

The chromosomal rearrangements such as translation, inversion and deletion occurred around eRTBVL-D locus in AA-genome species. The chromosomal structure around eRTBVL-D of *japonica* genome was compared with that of the other species. Chromosome 1 retained three eRTBVL-sequences, d1, d2, and d3, which have different alignments among the AA-genome species. The d2 locus was present in Asian rice and *O. meridionalis*, while the genomes of *O. glumaepatula* and *O. glaberrima* have deleted d2 between d1 and d3 present in *O. sativa* (Figure 4.9). In *O. barthii* and *O. longistaminata*, d2 was also missing, and d1 and d3 sequences overlapped with 300 bp in tandem (Figure 4.9).

I discovered inter- and intra-chromosomal rearrangements regarding the array of d10 and d11 located on chromosome 7 of *O. sativa* (Figure 4.10). Most of the AA-genome species had a similar structure of the region around d10-11 represented by *O. sativa*, but not *O. glumaepatula* and *O. meridionalis*. *O. glumaepatula* exhibited translocation of the region containing d10-11 and a vicinity gene from chromosome 7 to chromosome 1 (Figure 4.10). *O. meridionalis* has an inversion of d10-11 in chromosome 7 (Figure 4.10). These small chromosomal rearrangements observed in the eRTBVL-D locus might have also occurred in other many chromosomal segments in the AA-genome species.

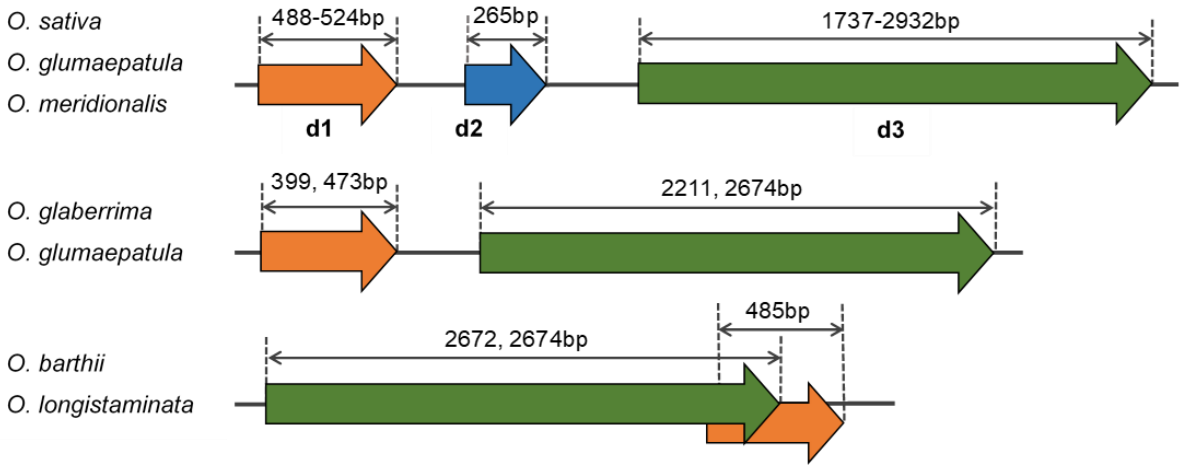


Figure 4.9 The rearrangements occurred around d1, d2 and d3 loci in *Oryza* AA-genome species.

Arrows represent a gene or eRTBVL-D locus. The numbers above each arrow indicate the length of the region of gene or eRTBVL-D in the AA-genome species in the left.

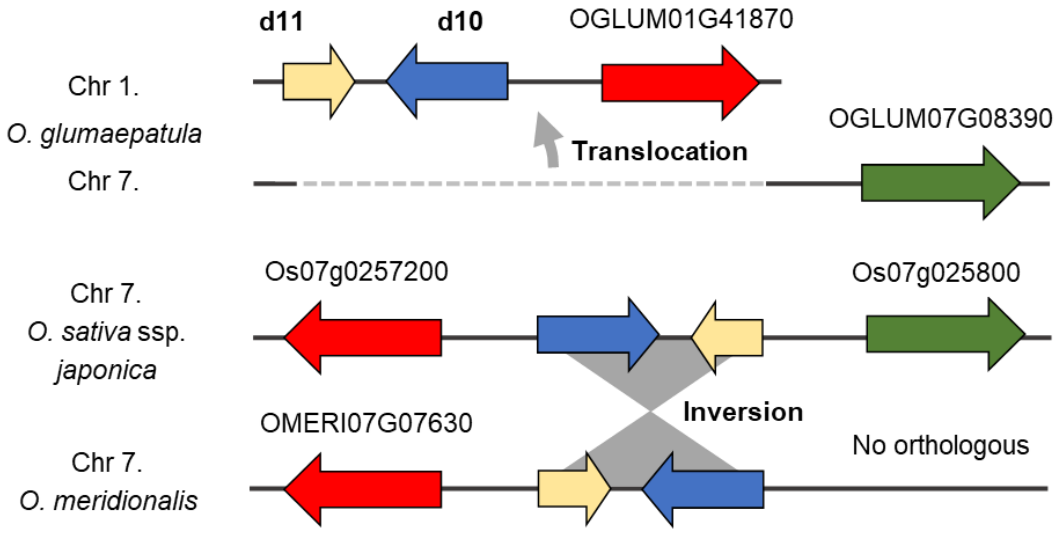


Figure 4.10 The translocation and inversion observed around d10 and d11 loci in *Oryza* AA-genome species.

Arrows indicate a gene or eRTBVL-D loci. The observed inversion and translocation in *O. glumaepatula* and *O. meridionalis* were based on the comparison with *O. sativa* ssp. *japonica* (the center).

4.3.6 Neutrality test of the regions around eRTBVL-D in the *O. sativa* populations

Tajima's D (Tajima, 1989) could test signatures of natural selection based on the number of polymorphic sites in the regions around eRTBVL-D between *japonica* and *indica* varieties. When Tajima's D is >0, natural selection that actively maintains genetic variation (equilibrium selection) is working, whereas if Tajima's D is <0, it indicates that natural selection (purification selection) works to eliminate harmful mutations. A total of 240 accessions from each of *japonica* and *indica* were used for the sliding window analysis of Tajima's D (Supplementary Table S2, S3). To investigate the degree of natural selection for eRTBVL-D loci in the speciation in Asian cultivated rice varieties, I estimated Tajima's D with the regions of eRTBVL-D loci and their flanking genes. Significant positive values for Tajima's D were observed for the d1-3 and d14 loci of the *japonica* population and d6-7 of the *indica* population (Table 4.4). The significant polymorphic spectrum of the eRTBVL sequence from d1-3 did not accompany the corresponding vicinity genes with significant value. So, the eRTBVL sequence from d1-3 itself was responsible for the positive Tajima's D value (Table 4.4). The other positive loci, d14 in *japonica* and d6-7 in *indica*, were linked with the adjacent regions, which possessed the positive Tajima's D values (Table 4.4). The regions containing the d4-5 locus in *japonica* and the d1-3 and d4-5 loci in *indica* were indicated at significantly negative Tajima's D values (Table 4.4). These negative loci were associated with both vicinity regions (Table 4.4).

Figure 4.11 represents the sliding-window plots of the observed Tajima's D values in the regions containing eRTBVL-D sequences for the *japonica* and *indica* populations. The Tajima's D value of the d1-3 locus was significantly positive, though the sliding window analysis showed that the level of the d1-3 was even through the region of the locus (Table 4.4, Figure 4.11). In another case, the D value of d4-5 locus was significantly negative-biased,

and the flanking regions also showed a similar negative-biased D value (Table 4.4, Figure 4.11). These results indicated that the flanking region including eRTBVL-D loci have evolved under the same evolutionary selection. The plotted D values for the regions containing eRTBVL-D and its flanking sequences such as d1-3, d6-7, d9, and d14 were greatly different between *japonica* and *indica* populations (Figure 4.11). The negative values were detected in majority of the regions around d6-7, d9 and d14 in *japonica* population, whereas positive value was detected for these regions in *indica* population. In contrast, the D values for the regions around d1-3 and d9 were positive in the *japonica* population and negative in *indica* population (Figure 4.11). These results indicated that these regions containing eRTBVL-D sequences differed in the signatures of natural selection between *japonica* and *indica*.

Table 4.4 Tajima's D test for the region of eRTBVL-D loci and the flanking gene within the population of *japonica* and *indica*

	japonica			*indica*		
	upstream gene	eRTBVL	downstream gene	upstream gene	eRTBVL	downstream gene
d1-3	0.55061	2.26868*	-1.41503	-0.38941	-1.84824*	-1.57525*
d4-5	-1.29868	-1.90216*	-1.47562	-1.23096	-1.93885*	-1.57779*
d6-7	-2.31392*	-0.49646	0.59173	3.38549*	2.38403*	2.73045*
d9	3.34977*	0.35427	3.91622*	-0.6118	-1.25132	1.39187
d10-11	-1.34619	-1.19016	-1.08147	0.50616	0.75278	0.14155
d14	4.59152*	3.14968*	-0.21657	-0.4058	-0.09516	0.14967
d15	-1.05987	-1.54036	n.d. [1]	-1.04944	0.88986	0.9449

[1] The region with no D value because of no segregation site.

Asterisks represent Tajima's D value with a significant difference with the standard neutral model ($p<0.05$).

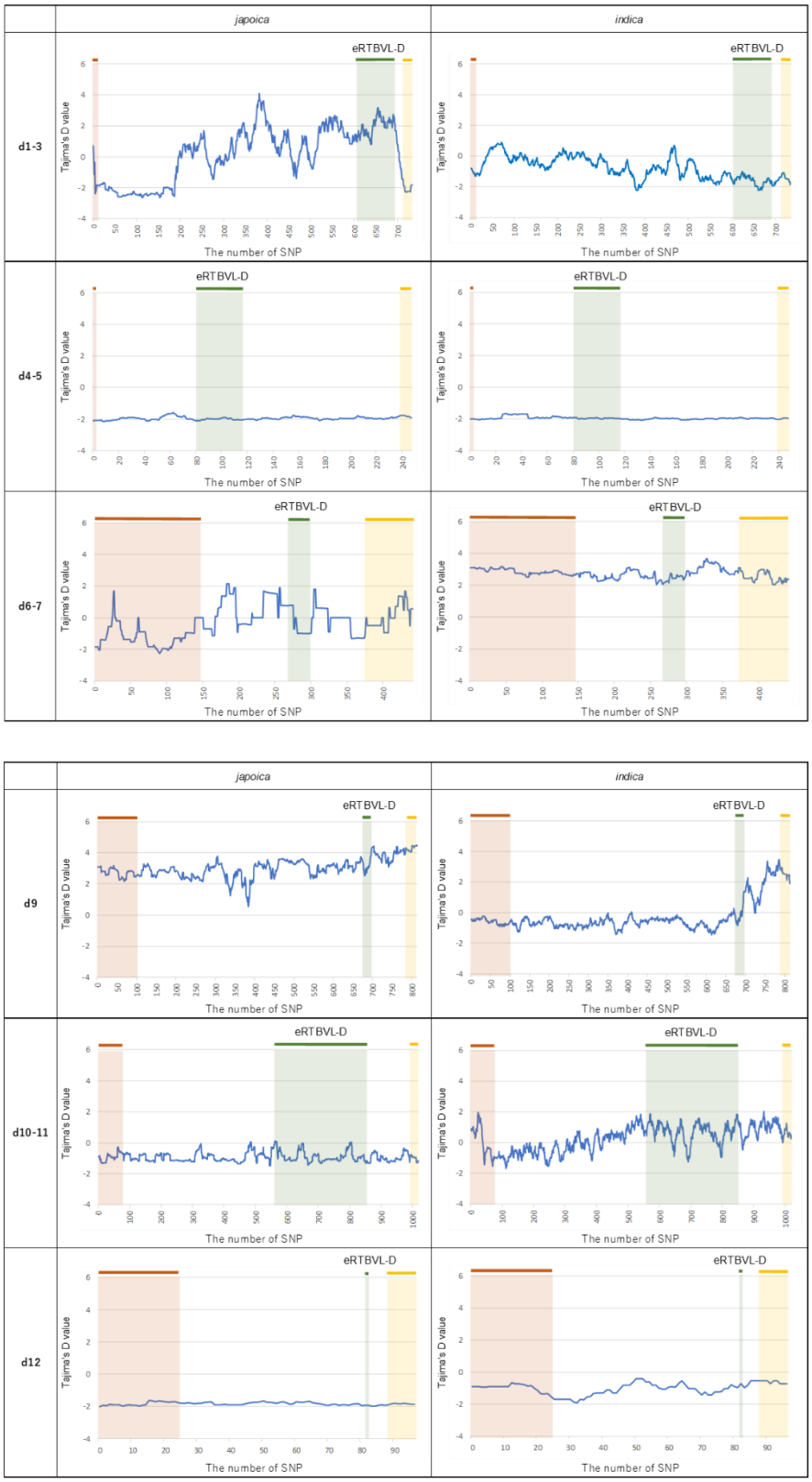


Figure 4.11 Sliding window analysis of Tajima's D across the franking gene and eRTBVL-D locus in *japonica* and *indica*.

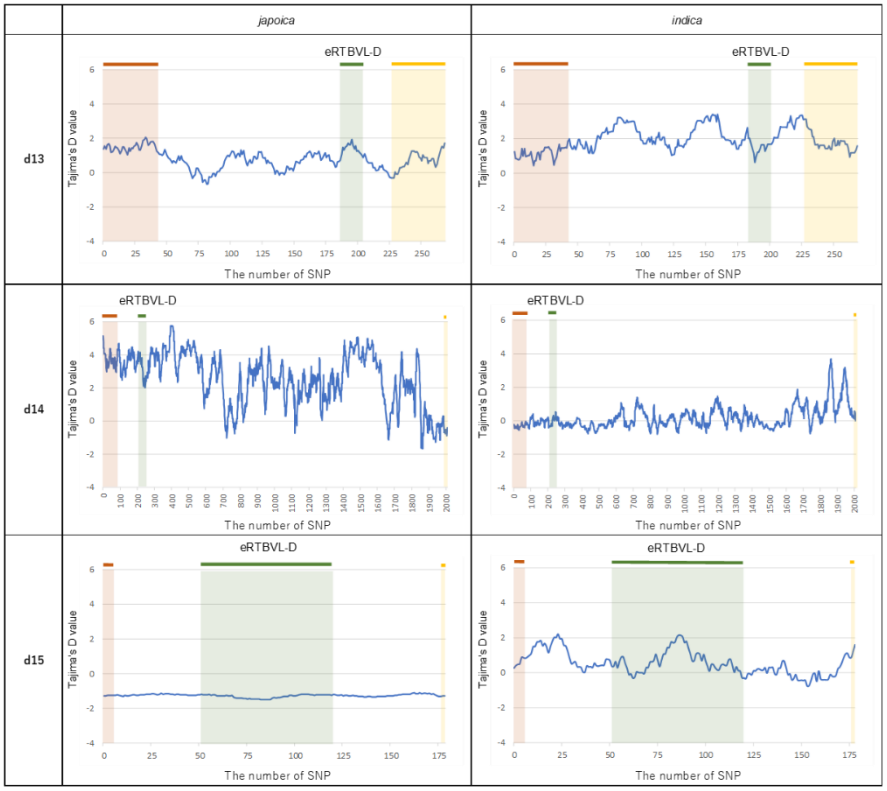


Figure 4.11 (Continued) Sliding window analysis of Tajima's D across the franking gene and eRTBVL-D locus in *japonica* and *indica*.

Sliding window plot of Tajima's D was constructed by a window size of 20 SNPs with step of a SNP. Color bars indicate the regions of eRTBVL-D loci or its flanking gene. The region of upstream flanking gene (red), eRTBVL-D region (green), and the region of downstream flanking gene (yellow). The range of color region represents a part of eRTBVL-D or gene sequence.

4.4 Discussion

In this study, I analyzed the evolution and the rearrangement of chromosomal regions with eRTBVL-D loci among the *Oryza* AA-genome species. In the phylogenetic relationships of the eRTBVL families, eRTBVL-D was distantly related to the other eRTBVL families, which have been found in Asian rice only. Therefore, the origin of eRTBVL-A, B, C and X virus could be integrated with the rice genome in Asia where the insect vector for RTD inhabits.

The phylogenetic analysis of eight eRTBVL-D loci were performed in this study. Only three of eRTBVL-D loci showed the same phylogenetic relationships as GPT (Figure 4.5), although the other five loci represented different phylogenetic relationships compared with GPT. Either flanking gene of the five loci also possessed inconsistent phylogenetic relationships (Figure 4.8). Currently, I do not define how such unusual phylogenetic relationships established during the speciation, but many of these relationships swerved from the known phylogeny might be present behind the chromosome structures.

Out of the 10 genomic regions including eRTBVL-D segment, seven were conserved in AA-genome species, the remaining three exhibited translocation, inversion and deletion during the AA-genome speciation (Figure 4.9 and Figure 4.10). It is not clear whether the chromosomal rearrangements observed around eRTBVL-D sequences triggered by itself or its flanking sequences. Further evolutionary evidence is necessary to resolve the issue too.

According to Tajima's D test, D values in d4-5 and d1-3 within *japonica* were significantly biased to the positive and negative states, respectively, while the corresponding flanking genes were rather at neutral state (Table 4.4). However, the outcomes for Tajima's D on the sliding window analysis for the same locus showed nearly even level (Figure 4.11). These results suggested that the eRTBVL-D loci have synchronously evolved with the

surrounding regions (Figure 4.11). The natural selections might not have been acted on eRTBVL-D as the central factor. Hence, eRTBVL could exist as “genomic fossils”.

I proposed that eRTBVL-D loci could be a useful resource as a neutral marker, because they rarely showed selective effect. The analyses of eRTBVL-D provided novel evolutionary insight for local chromosomal regions during the AA-genome speciation.

Chapter 5 General discussion

Plant viruses cause various types of disease and are responsible for huge losses in crop production and quality in the world (Nicaise, 2014). Plants have developed anti-viral defense machinery during their co-evolution with viruses. Regarding the anti-viral defense against RTBV, I considered three factors namely, endogenous virus sequence, vsRNA and rice miRNA, in this study.

Previous studies have pointed out that eRTBVL sequences found in the *O. sativa* genome were the possible elements to be involved in RTBV resistance (Kunii *et al*., 2004; Staginnus and Richert-Pöggeler, 2006; Liu *et al*., 2012). Agroinoculation analysis using 18 accession from *O. glaberrima* indicated that *O. glaberrima* plants were generally more susceptible to RTBV infection than *O. sativa* plants (Table 2.1, Table 2.2, Figure 2.1, Figure 2.2). That is, species with low eRTBVL copy numbers such as *O. glaberrima* tended to be vulnerable to RTBV infection. A possible mechanism of the RTBV-resistance due to EPRV sequences is suggested to be attributable to homology-dependent gene silencing (Chabannes and Iskra-Caruana, 2013). Based on small RNA sequencing of the RTBV-infected plants, the expression of eRTBVL-derived small RNAs was observed in *O. sativa*, but not observed in *O. glaberrima* (Figure 2.4C). Stable expression of eRTBVL-derived small RNAs in *O. glaberrima* was observed regardless of the RTBV-infected or mock-inoculated samples. In the comparison between RTBV-resistant and -susceptible cultivars in *O. sativa*, there was no correlation between the expression of eRTBVL-derived small RNAs and the genotypes (Figure 3.13, Figure 3.14). Unfortunately, I did not perform further analyses for the genes related to the silencing machineries of eRTBVL-derived small RNAs. However, I would say that the strong resistance against RTBV does not rely on the expression of small RNAs from

eRTBVL sequence in rice. Based on these agroinoculation analyses, it is likely that the presence of eRTBVL sequence, which was constantly expressed at low level, might function to mitigate symptoms and avoid serious damages by the virus.

According to the evolutionary analysis of the oldest eRTBVL-D family in *Oryza* AA-genome species, I supposed that eRTBVL-D, which is shared with all the AA-genome species, became genomic fossils without strong evolutionary selections. That is, the evolutionary history of the flanking region and vicinity genes influenced the evolution of nearby eRTBVL-D locus. On the other hand, Asian rice cultivars had frequent opportunities of infection with the other eRTBVL-related viruses during the speciation in Asia, resulting in the presence of eRTBVL families in the host genomes. Unlike eRTBVL-D family, eRTBVL-A, B, X, and C families comprise the large copy numbers in *O. sativa* and *O. rufipogon* (Liu *et al*., 2012; Chen *et al*., 2018). Together with the result of agroinoculation analysis, the four eRTBVL families, which were integrated into Asian rice after speciation, might contribute to mitigate the disease symptoms.

The regulation of small RNAs expression upon RTBV infection varied between resistant and susceptible plants (Figure 3.7, 3.11, 3.12, Table 3.6, 3.7). Two factors, which were the expressions of vsRNAs and rice miRNAs, might be interacted each other, resulting in the symptoms of RTBV infection. Virus silencing due to vsRNA was well known as an anti-viral defense machinery in plants (Calil and Fontes, 2017). The anti-silencing machinery of RTBV has been previously reported, but the anti-RTBV defense system of host plants was not fully understood. According to several reports about the function of RTBV-derived vsRNA (Rajeswaran *et al*., 2014; Zarreen *et al*., 2018), IGR-specific vsRNA was known to accumulate at around 25 dpi due to RTBV infection to disturb the silencing of viral RNA by *AGO* genes. In this study, the accumulation of vsRNA at IGR was observed in only the

resistant plants at 10 dpi or at earlier time points. I also detected the up-regulated expressions of *AGO1b* and *AGO2*, immediately after RTBV infection (Figure 3.7B, 3.9). These results indicated that the time from virus infection to produce vsRNAs might be important for RTBV resistance. Another point of view, vsRNAs derived IGR might have different function between the early and late infection stages. In the early stage of RTBV infection, the resistant plants activated the anti-viral system mediated by vsRNA-silencing and the abundant vsRNA derived from IGR repressed the translation of viral ORFs, minimizing the accumulation of virus in the plants. In the late stage of infection, vsRNAs derived from various region of IGR interfere with the host silencing machinery. My experimental outcomes resulted in the new insight for the host RTBV-defense system to produce the RTBV resistant plants by vsRNAs.

Over the past decade, a large number of miRNAs have been identified, and it has been revealed that miRNAs play critical roles on many biological processes or defense mechanisms at the post-transcription stage in plants (Jones-Rhoades *et al*., 2006; Gan *et al*., 2008). Several rice miRNAs, which were related to RTBV infection have been reported (Wang *et al*., 2016; Zarreen *et al*., 2018; Rajeswaran *et al*., 2014), but RTBV resistance-associated miRNAs have not been revealed yet. This is the first report of the analysis of miRNAs that are probably connected to RTBV-resistant plants. It was reported that RSV infection might suppress genes associated with rice growth and development by up-regulated miRNA for interrupting host-defense mechanisms against viruses, resulting in viral disease symptoms. Contrastingly, the resistant plants against RTBV might interfere in the cleavage of target RNAs by down-regulation of miRNAs and protect rice growth from viral disease symptom. This scenario was based on the DE analysis of rice miRNAs (Figure 3.12, Table 3.5, 3.6, 3.7). Most of the identified miRNAs related to RTBV were newly identified in this study (Table 3.8).

Taken together, this study proposed the possible anti-RTBV mechanism in the resistant plants and the evolutionary history of related endogenous viruses. I expect that my study gives an insight into how to develop resistant rice cultivars against Tungro disease.

Chapter 6 Summary

Rice, *Oryza sativa* L., is one of the important cereal crops that provides a staple food for half of the world's population. Rice viral diseases constitute a major obstacle to rice production. Rice is known to be infected by more than 30 viruses, but 25 viruses do not have direct economic effects in terms of rice production. Rice tungro disease (RTD) caused by *Rice tungro bacilliform virus* (RTBV) negatively affects rice production in South Asia and Southeast Asia. Plants have developed multilayered defense mechanisms against microbial pathogens such as gene silencing of viral small RNA (vsRNA) or host micro RNA (miRNA). Regarding the anti-viral defense against (RTBV), I considered three factors namely, endogenous RTBV like (eRTBVL) sequence, vsRNA and rice miRNA, in this study.

In Chapter 2, I focused on eRTBVL sequences, which have been identified in the rice genome. The distribution of eRTBVL sequences differs between the two genomes; the *O. sativa* genome contains more than 100 copies of eRTBVLs, while the *O. glaberrima* genome has only a few copies. Differences in the eRTBVL copy number in these genomes might be associated with the different reactions of these two species to RTBV. Firstly, I confirmed innate vulnerability to RTBV of *O. glaberrima* carrying a few eRTBVL copies, then the possible function of eRTBVL-derived small RNAs were examined by the comparison between *O. sativa* and *O. glaberrima*. Small RNA sequencing analysis indicated that the expression of small RNAs from eRTBVL might be associated with the phenotypic differences between the two genotypes after RTBV infection. I could suggest two possible ideas to explain the interaction between the susceptibility to RTBV and the presence of eRTBVL. One is that the constitutive expression of small RNAs from eRTBVL promotes the silencing machinery mediated by vsRNA in RTBV-infected plants. The other possibility is that the

small RNAs from eRTBVL respond immediately after RTBV infection to inhibit the replication of the virus at an initial stage of the infection.

In Chapter 3, the resistant (TW16) and susceptible (TN1) genotypes against RTBV were used for RTBV infection test and small RNA sequencing. To examine the RNAi-based defense machinery in rice during RTBV infection, I characterized vsRNA and rice-derived miRNA by small RNA sequencing and compared the expression of small RNAs between TN1 and TW16. A total of 11 crucial miRNAs related to RTBV resistance were identified in this study. Based on the expression analysis of *AGO* genes, vsRNA and rice miRNA indicated that the resistant plants against RTBV might interfere with the cleavage of target RNAs by down-regulation of miRNAs and protect the rice growth from viral disease symptom. The identified miRNAs might be valuable candidates to study novel insight into the defense mechanisms of resistant genotypes against RTBV.

Chapter 4 describes evolutionary patterns of the local chromosomal regions during the *Oryza* AA-genome speciation through the comparisons of eRTBVL-D sequences among the *Oryza* AA-genome species. eRTBVL-D is known as the oldest eRTBVL family that was integrated into the *Oryza* AA-genome before speciation. The eRTBVL-D sequences were available for the evolutionary study of the local-chromosomal regions. I found the swerved phylogenetic relationships from the known *Oryza* speciation and chromosomal rearrangements at the eRTBVL-D loci. I described the functional implications of eRTBVL-D loci by the evolutionary analyses for natural selection. The results of Tajima's D test indicated that natural selection might not have acted on eRTBVL-D as the central factor. Hence, eRTBVL could exist as "genomic fossils". In this chapter, I proposed that eRTBVL-D loci could be a useful resource as genetic neutral markers, because they have rarely selective effect. The analysis of eRTBVL-D provided novel evolutionary insight for local chromosomal

regions during the rice AA-genome speciation.

Taken together, this study proposed the possible RTBV-viral mechanism in the resistant plants and the evolutionary history of the related endogenous viruses. I expect that my study gives an insight into how to develop resistant rice cultivars against Tungro disease.

Supplementary Figures

Figure S1 Informative SNP patterns in the PR sequences of eRTBVL families and RTBV genome in the alignment.

Orange, A; green, T; blue, C; purple, G. Missing data are indicated by hyphens.

Figure S2 Informative SNP patterns in the ORF IV/ORFz sequences of eRTBVL families and RTBV genome in the alignment.

Orange, A; green, T; blue, C; purple, G. Missing data are indicated by hyphens.

Figure S3 Informative SNP patterns in the RT/RH sequences of eRTBVL families and RTBV genome in the alignment.

Orange, A; green, T; blue, C; purple, G. Missing data are indicated by hyphens.

Supplementary Tables

Table S1 Primers used in this study

1. Primers for the expression analysis of AGO genes				
Target	**Primer name**	**Primers**	**Primer sequence (from 5' to 3')**	**Length (mer)**
OsAGO1b	OsAGO1b_F	Forward	GGTTGCTGGAAATGTAGCCGTC	22
	OsAGO1b_R	Reverse	ACGATACAAGTACCAGACTTGCGT	24
OsAGO2	OsAGO2_F	Forward	CAGGCTCTACTACGAGGGCAT	21
	OsAGO2_R	Reverse	CCATGCGACGAAAGGCTGTC	20
OsAGO18	OsAGO18_F	Forward	GCTAAGTTCCAGTGGGTACTTCCA	24
	OsAGO18_R	Reverse	CACATGGAAACGAATGGCAACCTG	24
2. Primers for the expression analysis of eRTBVL-derived small RNAs				
Target	**Primer name**	**Primers**	**Primer sequence (from 5' to 3')**	**Length (mer)**
The region of IGR	SL1_siRNA5	For cDNA synthesis	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACAGGCAG	50
	siRNA5_F1	Forward	ACTGATTCCCTTCGGGTGGA	20
	siRNA5_R1	Reverse	GTGCAGGGTCCGAGG	15
U6 snRNA	U6_F	For cDNA synthesis, forward	TACAGATAAGATTAGCATGGCCCC	24
	U6_R	Reverse	GGACCATTTCTCGATTTGTACGTG	24

Table S2 The list of the accessions used for Tajima's D analysis in *O. sativa* ssp. *japonica* population

	japonica population		
	Subtrop	**temp**	**trop**
1	BUENG MONG LENG WE::IRGC 71088-1	GUANTUIBAIHE 1	JAERAERYUKDO::IRGC 73053-1
2	YAHNG LEU::IRGC 78279-1	GONGCHENGXIANG	R 27::IRGC 9498-1
3	ARC 13257::IRGC 22608-1	K 78-13::IRGC 36794-2	MALAGKIT (PINELIPE)::IRGC 67444-1
4	ARC 11768::IRGC 21630-1	BULI INIA::IRGC 116961-1	K 2 C 45::IRGC 68552-1
5	HO MOK::IRGC 73171-1	SR 113::GERVEX 553-C1	ZANTON::IRGC 31248-1
6	AE WAWNG POO DOO::IRGC 71007-1	HEIBIAO	NIAW KHAMIN::IRGC 40222-1
7	KHAO KHOT::IRGC 93892-1	TOPAZIO::GERVEX 1332-C1	MALAGKIT (ITIM)::IRGC 96124-1
8	A MOUK::IRGC 94474-1	MUGA::GERVEX 1099-C1	BALIK SEMAH::IRGC 17201-2
9	MAK KHEUA DENG::IRGC 107709-1	AMARELO::IRGC 9389-1	BULAK::IRGC 24884-1
10	NIA::IRGC 82728-1	QINGJINZAOSHENG	IR 71525-19-1-1::C1
11	CIRAD 358::C1	WIR 1951::IRGC 51643-1	IRAT109
12	DOK HIEN NOI::IRGC 107021-1	TEXAS PATNA 49::IRGC 6077-1	SALAZANA::IRGC 97219-1
13	R 258::IRGC 23702-2	CUNGUNUO	KINANDANGPATONG
14	KAUK PAHLING::IRGC 95823-2	MAGUZI	D 3-130::IRGC 30986-2
15	YARSABA::IRGC 97567-1	YUAN JING 7	GANDAMANA::IRGC 17593-2
16	TA SENG::IRGC 99176-1	YUEFU	LEKATAN::IRGC 71554-1
17	KHAO' NAN::IRGC 74981-1	K 113::IRGC 34107-1	ALEXANDROS::GERVEX 1678-C1
18	KHAU TRA BONG::IRGC 78342-1	KARA SHALI::IRGC 9293-1	INAPORAONAN::IRGC 19418-1
19	THATENG::IRGC 86377-1	CHIPKA::GERVEX 837-C1	IR 68704-145-1-1-B::C1
20	ARC 11538::IRGC 21463-1	LITCHIKIANG::IRGC 7287-1	GOGO::IRGC 43390-C1
21	LEUANG GLIANG::IRGC 71271-1	SHINCHIKU IKU 97::IRGC 10429-1	AMERICAN HUANGKEDAO
22	MANDASHERPO::IRGC 64913-1	CT 45::IRGC 34027-2	TOBOHUN::IRGC 63433-1
23	DON::IRGC 95117-1	C 722323::IRGC 73147-1	LIMBAYAN::IRGC 71556-1
24	THANGONE::IRGC 11816-1	M401	MINATANDA::IRGC 74613-1
25	VIETNAM 1::C1	BLAU BLA::IRGC 78460-1	REXARK ROGUE::IRGC 1972-1
26	CIRAD 358::C1	CUNGUNUO	MAMON::IRGC 30792-1
27	VIETNAM 1::C1	CX578	NERICA 8
28	ARC 11538::IRGC 21463-1	PIEMONTE::GERVEX 81-C1	LOHAMBITRO::IRGC 69857-1
29	ARC 12493::IRGC 22098-1	S 102/2::GERVEX 1251-C1	KOSAGI::IRGC 57692-1
30	KON SIM::IRGC 95626-1	91-382::IRGC 63464-1	KETAN DONGGO::IRGC 54201-1
31	POU::IRGC 98968-1	SETTANTUNO::GERVEX 1279-C1	LENKIBE-N::IRGC 11216-1
32	ARC 11768::IRGC 21630-1	TEXAS PATNA 49::IRGC 6077-1	REA::IRGC 73031-1
33	BER YA::IRGC 99970-1	YANDAO 9	HAOGELAO
34	DOK HIEN NOI::IRGC 107021-1	KINUGASAWASE::IRGC 2609-1	NIAW KHAMIN::IRGC 40222-1
35	KHAO LOUANG::IRGC 107770-1	WUYUGENG 20	CHATAO::IRGC 50456-1
36	YOI::IRGC 82855-1	GUANGKEXIANGNUO	NONGLIMNA 1::IRGC 58347-1
37	ARC 11946::IRGC 21768-2	YANDAO 8	KOSAGI::IRGC 57692-1
38	NAM ROO::IRGC 66996-1	NORIN 21::IRGC 493-1	D 4-193::IRGC 31083-2
39	A MOUK::IRGC 94474-1	MIARA::GERVEX 129-C1	KASIGNAYAN::IRGC 82788-1
40	NABESHI	SAFARI::GERVEX 1256-C1	IRAT 234::C1
41	YARSABA::IRGC 97567-1	TAICHUNG 65::IRGC 79-1	ANDEL::IRGC 17153-2
42	HAI NA::IRGC 107117-1	POLIZESTI 28::GERVEX 1170-C1	GAWICHAL::IRGC 77653-1
43	ARC 11424::IRGC 21380-2	WUYUGENG 3	A 2-257::IRGC 68342-1
44	PHIT CHZK::IRGC 73317-1	QINGNUO KYOHATAMOCHI	GRAAL::GERVEX 1682-C1
45	LEP HED::IRGC 11917-1	FENGZA	TJOKRON::IRGC 16653-2
46	KON SIM::IRGC 95626-1	HONGMISANDAN	32 UPLA::GERVEX 176-C1
47	KAMNAM::IRGC 64902-1	LIGEN 2::IRGC 82398-1	CALIFORNIA BELLE::IRGC 66836-1
48	SET::IRGC 92200-1	LIAOGENG 287	MANONGAZATO::IRGC 77876-1
49	MAK KHEUA KANG::IRGC 107712-1	SRAU SLAP::IRGC 23267-2	GUO DI HEI::IRGC 72697-1
50	MAK KHEUA DENG::IRGC 107709-1	WEIGUO	SAL BUI BAO
51	PHIT CHZK::IRGC 73317-1	BULI INIA::IRGC 116961-1	PADI ADONG DUMARAT::IRGC 14356-1
52	MAK KHEUA DENG::IRGC 107709-1	HONGKEZHENUO	PULUT MAS::IRGC 82676-1
53	AE WAWNG POO DOO::IRGC 71007-1	M 7::IRGC 34281-1	HYB 5-2-1-1-4::IRGC 34283-2
54	ARC 11538::IRGC 21463-1	KINUGASAWASE::IRGC 2609-1	PULUT CENRANA::IRGC 27400-2
55	VIENG::IRGC 107375-2	SELN 244 A 6-20::GERVEX 1273-C1	ASE PULU BOLONG::IRGC 25171-1
56	MUANG TAY::IRGC 98382-1	PRECOZ 2 F A::GERVEX 1182-C1	MANYALOJOPOIHUN::IRGC 63350-1
57	KHAO' NAN::IRGC 74981-1	C 722323::IRGC 73147-1	SEMENDANG::IRGC 93633-1
58	MUANG TAY::IRGC 98382-1	BERGREIS::IRGC 3150-1	BUSIYETAN::IRGC 11205-1
59	YAHNG LEU::IRGC 78279-1	053A-3	TEBONNET::IRGC 66760-1
60	PANG::IRGC 91987-1	GIZA159	BODA 148-3::GERVEX 8258-C1
61	KON SIM::IRGC 95626-1	RUBI::GERVEX 1247-C1	PAKWALAN::IRGC 24156-1
62	DAM::IRGC 23710-1	HWANGJO::IRGC 55547-1	RANGKAI KEPEL::IRGC 25640-1
63	HOKURIKU 52::IRGC 72491-1	WUDADAOZHONG	K 2 C 45::IRGC 68552-1
64	LAMUJIA	CAPATAZ::GERVEX 521-C1	KAYELI KUNING::IRGC 31974-1
65	LAI HANG HMA::IRGC 29608-2	CHALBYEO::IRGC 77639-1	GBRA::IRGC 32474-2
66	PHIT CHZK::IRGC 73317-1	RODINA::GERVEX 1234-C1	CERE KAWAT::IRGC 17401-2
67	ARC 6188::IRGC 20370-1	BULI INIA::IRGC 116961-1	BOND::IRGC 66755-1
68	NYON::IRGC 107099-1	SANT ANDREA::IRGC 65732-1	PADI BURIH::IRGC 44081-1
69	XISHI 15	S 102/2::GERVEX 1251-C1	SLOBOK::IRGC 16571-1
70	DAGPA BARA::IRGC 64887-1	GIZA159	LAYANDABU::IRGC 25214-1
71	MAK KHEUA::IRGC 11628-1	JINDAO 1	MANONGAZATO::IRGC 77876-1
72	KON SIM::IRGC 95626-1	HUA 24::IRGC 82127-1	DUNA AL::IRGC 71842-1
73	DAGPA BARA::IRGC 64887-1	XINGGUO	HYB 5-2-1-1-4::IRGC 34283-2
74	NAM ROO::IRGC 66996-1	YUEFU	AEN METAN::IRGC 17142-1
75	YOI::IRGC 82855-1	WA BANG::IRGC 19880-1	CT 13582-15-5-M::C1
76	MAK KHEUA KANG::IRGC 107712-1	GONG SHE 9::IRGC 62693-1	SLAMET::IRGC 18900-1
77	PLE LIA::IRGC 73562-1	S 201::IRGC 55230-1	PANGETAN::IRGC 26942-1
78	KAO SONG HAI::IRGC 61827-1	GYEONGSAN 1::IRGC 79404-1	AMERICAN HUANGKEDAO
79	THATENG::IRGC 86377-1	CHIMAO	GODAWEL::IRGC 15750-1
80	KHAO::IRGC 107359-1	H 305-84::IRGC 116988-1	AMERICAN HUANGKEDAO

Table S3 The list of the accessions used for Tajima's D analysis in *O. sativa* ssp. *indica* population

	indica population			
	ind1a	**ind1b**	**ind2**	**ind3**
1	MIN KE ZHAN::IRGC 72230-1	TIKAL 3::IRGC 50649-1	KUSHIARA::IRGC 34709-1	BATU::IRGC 71508-1
2	HONG DU BAI::IRGC 72098-1	KHAO GRADOOK CHAHNG::IRGC 17111-1	LARHA MUGAD::IRGC 52339-1	BONGOLON::IRGC 13094-1
3	CE IN TSAN::IRGC 4362-1	AMISTAD 82::IRGC 116954-1	MODDAI KARUPPAN::IRGC 15465-1	BUCAYAB::IRGC 44357-1
4	WUGU::IRGC 93479-1	RP 9-4::IRGC 39735-1	TSIPALA MENA 626::GERVEX 5392-C1	SIGUPAI::IRGC 24642-2
5	MR 136-1::IRGC 50707-1	CS94	CYLINDRICAL 30-605::IRGC 37845-2	DINGRAS::IRGC 44402-1
6	HSIN-T'AO YUAN CHING YU::IRGC 65074-1	WAS 203-B-B-2-4-1::C1	ROJOKELY::GERVEX 8410-C1	TERUNG DAUN::IRGC 25823-2
7	ACC 1286::IRGC 60116-1	KR200	PECALO::IRGC 30595-2	RADIN EBOS 35::IRGC 14204-1
8	HSIN-T'AO YUAN CHING YU::IRGC 65074-1	IR06G113	LAKHA KUAR::IRGC 74760-1	P 41::IRGC 68987-1
9	FANHAOPI	SEQ 93::C1	BR 116-3B-53::IRGC 39559-2	REMOL::IRGC 18624-1
10	PENGSHANTIEGANZHAN	MATATAG2	ARC 11322::IRGC 21315-2	LEUANG YAI 29-12-2::IRGC 881-1
11	G ZHENSHAN 97B	IR 28::IRGC 30411-C1	PURA BINNI::IRGC 26772-1	PAHK NOK GAEW::IRGC 64572-1
12	MEIHUANUO	AI ZI DAO	NCS 766::IRGC 62478-1	NEANG MAC::IRGC 23071-1
13	NANXIONGZAOYOUZHAN	NERICA-L-20	DUBRAJ::IRGC 34904-1	PADI TIMBUN::IRGC 13451-1
14	KAHAMU	PR106	LATA MONA::IRGC 49333-1	LIMA::IRGC 81487-1
15	QI GU::IRGC 77381-1	B 6136 E-3-TB-0-1-5::IRGC 117311-1	SUFALDHULA::IRGC 46698-2	KHAN THAI DO::IRGC 12990-1
16	GONGJU 73	XUANENCHANGTANQINGZHAN	PADI ARA::IRGC 54252-2	PIN GAEW 56::IRGC 7887-1
17	PAI CHUEH CHIU LIU::IRGC 34259-1	BG90-2	SHYAMJIRA::IRGC 77522-1	LINANGAN::IRGC 47255-2
18	QUN XUAN ZAO::IRGC 70371-1	CNA 4081::GERVEX 1494-C1	MARAGBE::IRGC 69263-1	JENDA ABANG::IRGC 25413-2
19	PENGSHANTIEGANZHAN	GUANG122	PERIYA VELLAI::IRGC 15475-1	B 12::IRGC 31293-1
20	XIA ZHI BAI::IRGC 53437-1	ITA 117::IRGC 75235-1	CHITRAKALI::IRGC 49524-2	BAHN DAENG::IRGC 65771-1
21	ZALE	CAOZHAO-2	CHAKOL::IRGC 77226-1	EA NOUAN::IRGC 95147-1
22	MAN TIAN QING::IRGC 72776-1	SADA RUPA::IRGC 77299-1	BANIKAT::IRGC 67720-1	KACIK TARAM::IRGC 35310-1
23	ESINIU	CX426	MENALAVA::IRGC 97122-1	BA TUC RAN::IRGC 73239-1
24	XIANLUOSICHI	G341	ARC 12180::IRGC 21965-2	MANUS::IRGC 71570-1
25	XIA HONG GU::IRGC 76803-1	IR6	19::IRGC 70786-1	POTO BALA::IRGC 15986-2
26	DA GU AI 7::IRGC 70075-1	IR42	VELLAI KOLOMBAN::IRGC 15517-1	EX EBOKOZURU::IRGC 79912-1
27	VIETNAM ZAODAO	PSBRC82	NOLOS::IRGC 77285-1	PIN GAEW 56::IRGC 7887-1
28	I KUNG PAO::IRGC 114-1	CICA 8::C1	MUTA GANJE::IRGC 26744-1	BIDAPIK::IRGC 43354-1
29	TAISHANNUO	AUS 177::IRGC 29009-1	DUDH KADAR::IRGC 67707-1	DICULA::IRGC 36729-2
30	MAN TIAN QING::IRGC 72776-1	IRBB7	VADAI::IRGC 37341-1	EMAHTAKYAN::IRGC 97789-1
31	TN1	IR2061-522-6-9	NCS 477::IRGC 62299-1	BANDANG BUNGKUAKLAN::IRGC 24697-2
32	WANLIXIAN	AUS 177::IRGC 29009-1	NCS 830::IRGC 62518-1	ALAMINOS::IRGC 44268-2
33	CHIH SHEN LI::IRGC 1306-1	TYPE3	GUTTI AKKULLU::IRGC 26850-2	SOK SOV::IRGC 83966-1
34	QITOUGU	TOS 9795::IRGC 80142-1	CO 2::IRGC 6186-2	NKUMBAN NDINGO::IRGC 77718-1
35	PENGSHANTIEGANZHAN	CX416	LUDI GOCHYA::IRGC 26504-2	ES 21::IRGC 56171-1
36	LIUYENIAN	PR 106::IRGC 53418-1	STRAW 23-471::IRGC 38456-1	PUEY TAW::IRGC 71405-1
37	AI ZI HUNG::IRGC 51255-1	IR 19746-28-2-2::IRGC 78072-C1	VARY MALADY MENA::IRGC 51555-1	KHAO' PAWM::IRGC 64416-2
38	AC 74::IRGC 33967-1	ECIA 24-107-1::IRGC 116978-1	MADHUKAR	DAMM NOEB THNGORN'::IRGC 87428-1
39	SAN ZHAO QI::IRGC 72289-1	IRGA 370-38-1-1F-C4-2::IRGC 117342-1	ARC 10159::IRGC 20754-2	PADI BENOT::IRGC 24757-2
40	BEIZINUO	CHENGNONGSHUIJING	PUILLIPINA KATARI::IRGC 77293-1	SANPATONG NYAY::IRGC 99050-1
41	SANBAILI	CHENGNONGSHUIJING	DISSI::IRGC 101346-1	PADI SIRANDAH KUNING::IRGC 73762-1
42	CHI GU::IRGC 71988-1	SUWEON 311::IRGC 61890-1	HERATH BANDA::IRGC 67630-1	SIMET 2::IRGC 25734-1
43	LAI YIP ZIM::IRGC 4955-1	IRGA 318-11-9-2A::IRGC 117340-1	PECALO::IRGC 30595-2	SRI RAJA
44	ZHENGSHAN97	CHATO::IRGC 55825-1	O. SATIVA::IRGC 10164-2	MOSHI::IRGC 70730-1
45	WANLIXIAN	BR IRGA 409::IRGC 55915-1	PANIKELASH::IRGC 46500-2	RADIN EBOS 35::IRGC 14204-1
46	YOUNIAN	XIANGAIZAO 10HAO	PERIYA VELLAI::IRGC 15475-1	PEUAK NAM 43::IRGC 17129-1
47	LING TA 150::IRGC 1106-1	ZHONGHUA 1	PATALASAFED SUNGHAWADO::IRGC 61133-1	SI HAO NDURIA::IRGC 66649-1
48	CHAOYANGYIHAO B	MOUI SAMBA::IRGC 75751-1	VARY LAVA DE MAROVATO::GERVEX 8396-C1	THIERNO BANDE::IRGC 16009-1
49	LAOZAOGU::IRGC 93444-1	SEBERANG MR 77::C1	GANDHISAIL::IRGC 43826-2	ZEINGYI::IRGC 33923-2
50	HENG YE ZAO::IRGC 72092-1	PSBRC 18::IRGC 99639-1	NONA BOKRA::IRGC 22710-C1	GAEW DAENG::IRGC 66707-1
51	SHAN GU::IRGC 59870-1	IR 72967-12-2-3::C1	IB 11::IRGC 74420-2	DAMNOEUB KHSE SAUT::IRGC 22819-2
52	NAN JING XIAO DAO::IRGC 72236-2	FENGAIZHAN	KOLONGI BAO::IRGC 24135-2	ZERO::IRGC 70707-2
53	SAN BAO GU::IRGC 76732-1	B6136-3-TB-0-1-5	JALMAGNA	E PLUAK::IRGC 64297-1
54	PILIT (7480) SELN (CI 12007)::IRGC 3758-1	IR 9560-2-6-3::IRGC 40451-1	NCS 477::IRGC 62299-1	KOUPENEDOU KOUBOURY::IRGC 16011-1
55	SAN BAO GU::IRGC 76732-1	CAMPECHE A 79::IRGC 51102-1	FOFFI::IRGC 68438-1	ARITH::IRGC 84347-2
56	HONG YANG ZAO 3::IRGC 67177-1	WAS 203-B-B-2-4-1::C1	ARC 14060::IRGC 41374-1	LONG KHI KRAI::IRGC 64490-1
57	YA NONG ZAO 4::IRGC 63908-1	CX417	STRAW 23-438::IRGC 38426-2	NELI::IRGC 66603-1
58	CHAOYANGYIHAO B	SABRAN::IRGC 34630-1	ROJOFOTSY::IRGC 69402-1	KHAO THI RATE::IRGC 58041-1
59	JIE CAO ZHAN::IRGC 70305-1	B 6136 E-3-TB-0-1-5::IRGC 117311-1	ARC 11322::IRGC 21315-2	PENDEK LAYUNG::IRGC 43542-2
60	MI GU::IRGC 76687-1	UPR 1201-1-20-1::IRGC 117357-1	MOCHICA::IRGC 11070-1	TAMBAS JANGKOK::IRGC 78724-1

Acknowledgement

Mr. Kazuhiro Nanba, Mr. Hiroshi Miyauchi and the stuffs, who were related to the conclusion of *Cotutelle* agreement, grew out of my doctoral dissertation for the *Cotutelle* program at Hokkaido university and International Rice Research Institute (IRRI) in 2017. I benefited from the assistance and support of numerous individuals and institutions during the development of the dissertation.

My deepest gratitude goes to Dr. Yuji Kishima and Dr. Il-Ryong Choi, Dr. Maria Genaleen Q. Diaz and Dr. Chikara Masuta, who are my thesis supervisors. They helped me with considering my future plan and how my research should be advanced by plenty of discussions. Their comments and constructive criticism of them supported pave the way to write this doctoral thesis. I am deeply grateful to Dr. Sunlu Chen, Dr. Itsuro Takamure, Dr. Yohei Koide, Dr. Maria Stefanie Dwiyanti and Dr. Nguyen Thanh Tam for giving me their ideas and useful methods of the experiments such as bioinformatics analysis and growing the rice plants. I also thank to the Ph.D. students, Mr. Kanaoka, Mr. Ohta, Mr. Kuniyoshi and Mr. Yamamori for insightful comments and suggestions. Ms. Yamada, Mr. Utsu, Ms. Yoshizumi and Mr. Kitajima, who are the member of virus team, have been tolerant and supportive of me. I also would like to thank to the staffs and students of Plant breeding Laboratory, Faculty of Agriculture, Hokkaido University for their support and cooperation and especially for their warm friendship during the course. Without all members, it would not be possible to enjoy my life and research in the laboratory.

I would like to express the deepest appreciation to all members of the virology laboratory at IRRI, Ma'am Jaymee, Ma'am Ida, Jerlie, IA, Kuya Celso, Kuya Onie, Kuya Jun, Kuya Moises and Kuya Paul. They helped me with my stay in the Philippines and the

experiments which have been done at IRRI. I have had nice time at IRRI because of everyone I met in the Philippines.

Computations were partially performed on the NIG supercomputer at ROIS National Institute of Genetics. This work was supported by Grant-in-Aid for JSPS Fellows of the Japan Society for the Promotion of Science (JSPS), Japan.

References

Abo, M. & A. Sy (1998) Rice virus diseases: Epidemiology and management strategies. *Journal of Sustainable Agriculture*, 11, 113-134.

Agudelo-Romero, P., P. Carbonell, F. de la Iglesia, J. Carrera, G. Rodrigo, A. Jaramillo, M. Perez-Amador & S. Elena (2008) Changes in the gene expression profile of Arabidopsis thaliana after infection with Tobacco etch virus. *Virology Journal*, 5, 92

Akbergenov, R., A. Si-Ammour, T. Blevins, I. Amin, C. Kutter, H. Vanderschuren, P. Zhang, W. Gruissem, F. Meins, T. Hohn & M. Pooggin (2006) Molecular characterization of geminivirus-derived small RNAs in different plant species. *Nucleic Acids Research*, 34, 462-471.

Archak, S., V. Lakshminarayanareddy & J. Nagaraju (2007) High-throughput multiplex microsatellite marker assay for detection and quantification of adulteration in Basmati rice (*Oryza sativa*). *Electrophoresis*, 28, 2396-2405.

Axtell, M. & S. Merchant (2013) Classification and Comparison of Small RNAs from Plants. *Annual Review of Plant Biology, Vol 64*, 64, 137-159.

Azzam, O. & T. Chancellor (2002) The biology, epidemiology, and management of rice tungro disease in Asia. *Plant Disease*, 86, 88-100.

Babu, M., J. Griffiths, T. Huang & A. Wang (2008) Altered gene expression changes in Arabidopsis leaf tissues and protoplasts in response to Plum pox virus infection. *Bmc Genomics*, 9, 325

Bajet, N., V. Aguiero, R. Daquioag, G. Jonson, R. Cabunagan, E. Mesina & H. Hibino (1986) Occurrence and spread of Rice Tungro Spherical Virus in the Philippines. *Plant Disease*, 70, 971-973.

Bari, R., B. Pant, M. Stitt & W. Scheible (2006) PHO2, microRNA399, and PHR1 define a phosphate-signaling pathway in plants. *Plant Physiology*, 141, 988-999.

Baulcombe, D. (2004) RNA silencing in plants. *Nature*, 431, 356-363.

Blevins, T., R. Rajeswaran, M. Aregger, B. Borah, M. Schepetilnikov, L. Baerlocher, L. Farinelli, F. Meins, T. Hohn & M. Pooggin (2011) Massive production of small RNAs from a non-coding region of Cauliflower mosaic virus in plant defense and viral counter-defense. *Nucleic Acids Research*, 39, 5003-5014.

Blevins, T., R. Rajeswaran, P. Shivaprasad, D. Beknazariants, A. Si-Ammour, H. Park, F. Vazquez, D. Robertson, F. Meins, T. Hohn & M. Pooggin (2006) Four plant Dicers mediate viral small RNA biogenesis and DNA virus induced silencing. *Nucleic Acids Research*, 34, 6233-6246.

Bradbury, P., Z. Zhang, D. Kroon, T. Casstevens, Y. Ramdoss & E. Buckler (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23, 2633-2635.

Budot, B., J. Encabo, I. Ambita, G. Atienza-Grande, K. Satoh, H. Kondoh, V. Ulat, R. Mauleon, S. Kikuchi & I. Choi (2014) Suppression of cell wall-related genes associated with stunting of *Oryza glaberrima* infected with Rice tungro spherical virus. *Frontiers in Microbiology*, 5, 26

Cabauatan, P., R. Cabunagan & H. Koganezawa (1995) Biological variants of Rice Tungro Viruses in the Philippines. *Phytopathology*, 85, 77-81.

Calil, I. & E. Fontes (2017) Plant immunity against viruses: antiviral immune receptors in focus. *Annals of Botany*, 119, 711-723.

Carbonell, A. & J. Carrington (2015) Antiviral roles of plant ARGONAUTES. *Current Opinion in Plant Biology*, 27, 111-117.

Carthew, R. & E. Sontheimer (2009) Origins and Mechanisms of miRNAs and siRNAs. *Cell*, 136, 642-655.

Catoni, M., L. Miozzi, V. Fiorilli, L. Lanfranco & G. Accotto (2009) Comparative Analysis of Expression Profiles in Shoots and Roots of Tomato Systemically Infected by Tomato spotted wilt virus Reveals Organ-Specific Transcriptional Responses. *Molecular Plant-Microbe Interactions*, 22, 1504-1513.

Chabannes, M., F. Baurens, P. Duroy, S. Bocs, M. Vernerey, M. Rodier-Goud, V. Barbe, P. Gayral & M. Iskra-Caruana (2013) Three Infectious Viral Species Lying in Wait in the Banana Genome. *Journal of Virology*, 87, 8624-8637.

Chen, S (2015) The evolution of endogenous pararetrovirus sequences in the genomes of rice and other grass species. Doctoral thesis, Hokkaido University, Sapporo

Chen, S. & Y. Kishima (2016) Endogenous pararetroviruses in rice genomes as a fossil record useful for the emerging field of palaeovirology. *Molecular Plant Pathology*, 17, 1317-1320.

Chen, S., R. Liu, K. Koyanagi & Y. Kishima (2014) Rice genomes recorded ancient pararetrovirus activities: Virus genealogy and multiple origins of endogenization during rice speciation. *Virology*, 471, 141-152.

Chen, S., N. Saito, J. Encabo, K. Yamada, I. Choi & Y. Kishima (2018) Ancient Endogenous Pararetroviruses in *Oryza* Genomes Provide Insights into the Heterogeneity of Viral Gene Macroevolution. *Genome Biology and Evolution*, 10, 2686-2696.

Chen, S., H. Zheng & Y. Kishima (2017) Genomic fossils reveal adaptation of non-autonomous pararetroviruses driven by concerted evolution of noncoding regulatory sequences. *Plos Pathogens*, 13, 6.

Chisholm, S., G. Coaker, B. Day & B. Staskawicz (2006) Host-microbe interactions: Shaping the evolution of the plant immune response. *Cell*, 124, 803-814.

Cruz, F., R. Hull & O. Azzam (2003) Changes in level of virus accumulation and incidence of infection are critical in the characterization of Rice tungro bacilliform virus (RTBV) resistance in rice. *Archives of Virology*, 148, 1465-1483.

Cruz, F., H. Koganezawa & H. Hibino (1993) Comparative cytology of Rice Tungro Viruses in selected rice cultivars. *Journal of Phytopathology*, 138, 274-282.

Dardick, C. (2007) Comparative expression profiling of Nicotiana benthamiana leaves systemically infected with three fruit tree viruses. *Molecular Plant-Microbe Interactions*, 20, 1004-1017.

Dasgupta, I., R. Hull, S. Eastop, C. Phoggipollini, M. Blackebrough, M. Boulton & J. Davies (1991) Rice Tungro Bacilliform Vius-DNA Independently Infects Rice After Agrobacterium-mediated Transfer. *Journal of General Virology*, 72, 1215-1221.

Dong, O. & P. Ronald (2019) Genetic Engineering for Disease Resistance in Plants: Recent Progress and Future Perspectives. *Plant Physiology*, 180, 26-38.

Edgar, R. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32, 1792-1797.

Encabo, J., P. Cabauatan, R. Cabunagan, K. Satoh, J. Lee, D. Kwak, T. De Leon, R. Macalalad, H. Kondoh, S. Kikuchi & I. Choi (2009) Suppression of Two Tungro Viruses in Rice by Separable Traits Originating from Cultivar Utri Merah. *Molecular Plant-Microbe Interactions*, 22, 1268-1281.

Feschotte, C. & C. Gilbert (2012) Endogenous viruses: insights into viral evolution and impact on host biology. *Nature Reviews Genetics*, 13, 283-U88.

Guerra-Peraza, O., M. de Tapia, T. Hohn & M. Hemmings-Mieszczak (2000) Interaction of the Cauliflower mosaic virus coat protein with the pregenomic RNA leader. *Journal of Virology*, 74, 2067-2072.

Hackenberg, M., B. Shi, P. Gustafson & P. Langridge (2013) Characterization of phosphorus-regulated miR399 and miR827 and their isomirs in barley under phosphorus-sufficient and phosphorus-deficient conditions. *Bmc Plant Biology*, 13, 214.

Hamilton, A. & D. Baulcombe (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science*, 286, 950-952.

Hannon, G. (2002) RNA interference. *Nature*, 418, 244-251.

Hansen, C., G. Harper & J. Heslop-Harrison (2005) Characterisation of pararetrovirus-like sequences in the genome of potato (Solanum tuberosum). *Cytogenetic and Genome Research*, 110, 559-565.

Harvey, J., M. Lewsey, K. Patel, J. Westwood, S. Heimstadt, J. Carr & D. Baulcombe (2011) An Antiviral Defense Role of AGO2 in Plants. *Plos One*, 6, 1.

Havelda, Z., E. Várallyay, A. Valoczi & J. Burgyan (2008) Plant virus infection-induced persistent host gene downregulation in systemically infected leaves. *Plant Journal*, 55, 278-288.

Hay, J., M. Jones, M. Blackbrough, I. Dasgupta, J. Davis & R. Hull (1991) An analysis of the sequence of an infectious clone of Rice tungro bacilliform virus, A plant pararetrovirus. *Nucleic Acids Research*, 19, 2615-2621.

Hibino, H. (1983) Transmission of 2 Rice Tungro-associates Viruses and Rice Waika Virus from Doubly or Singly Infected Source Plants by Leafhopper Vectors. *Plant Disease*, 67, 774-777.

Hibino, H. & P. Cabauatan (1987) Infectivity Neutralization of Rice Tungro-associated Viruses Acquired by Vector Leafhoppers. *Phytopathology*, 77, 473-476.

Hibino, H., R. Daquioag, E. Mesina & V. Aguiero (1990) Resistances in Rice to Tungro-associated Viruses. *Plant Disease*, 74, 923-926.

Huang, P., J. Molina, J. Flowers, S. Rubinstein, S. Jackson, M. Purugganan & B. Schaal (2012) Phylogeography of Asian wild rice, *Oryza* rufipogon: a genome-wide view. Molecular Ecology, 21, 4593-4604.

Hull, R. (1996) Molecular biology of rice tungro viruses. *Annual Review of Phytopathology*, 34, 275-297.

Hull, R. *et al*., (2005) Family Caulimoviridae. In Fauquet, C. *et al*., ed, Virus taxonomy: Classification and nomenclature of viruses: Eighth report of the international committee on taxonomy of viruses. *Elsevier/Academic Press, San Diego, California*, pp. 385-437.

Itoh, J., K. Hibara, Y. Sato & Y. Nagato (2008) Developmental role and auxin responsiveness of class III homeodomain leucine zipper gene family members in rice. *Plant Physiology*, 147, 1960-1975.

Jakowitsch, J., M. Mette, J. van der Winden, M. Matzke & A. Matzke (1999) Integrated pararetroviral sequences define a unique class of dispersed repetitive DNA in plants. *Proceedings of the National Academy of Sciences of the United States of America*, 96, 13241-13246.

Jones, J. & J. Dangl (2006) The plant immune system. *Nature*, 444, 323-329.

Jones-Rhoades, M., D. Bartel & B. Bartel (2006) MicroRNAs and their regulatory roles in plants. *Annual Review of Plant Biology*, 57, 19-53.

Kang, B., I. Yeam & M. Jahn (2005) Genetics of plant virus resistance. *Annual Review of Phytopathology*, 43, 581-621.

Kobayashi, N. & R. Ikeda (1992) Necrosis Caused by Rice Tungro Viruses in *Oryza*. *glaberrima* and O. barthii. *Japanese Journal of Breeding*, 42, 885-890.

Kumar, S., G. Stecher & K. Tamura (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, 33, 1870-1874.

Kunii, M., M. Kanda, H. Nagano, I. Uyeda, Y. Kishima & Y. Sano (2004) Reconstruction of putative DNA virus from endogenous rice tungro bacilliform virus-like sequences in the rice genome: implications for integration and evolution. *Bmc Genomics*, 5, 80.

Li, Z., X. Zheng & S. Ge (2011) Genetic diversity and domestication history of African rice (*Oryza glaberrima*) as inferred from multiple gene sequences. *Theoretical and Applied Genetics*, 123, 21-31.

Liu, R., K. Koyanagi, S. Chen & Y. Kishima (2012) Evolutionary force of AT-rich repeats to trap genomic and episomal DNAs into the rice genome: lessons from endogenous pararetrovirus. *Plant Journal*, 72, 817-828.

Lo, C. & P. Chain (2014) Rapid evaluation and quality control of next generation sequencing data with FaQCs. *Bmc Bioinformatics*, 15, 366.

Lu, J., X. Zhang, H. Wang, X. Yuan, Q. Xu, Y. Wang, H. Yu, S. Tang & X. Wei (2008) SSR Analysis on Diversity of AA Genome *Oryza* Species in the Southeast and South Asia. Rice Science, 15, 289-294.

Mallory, A., B. Reinhart, M. Jones-Rhoades, G. Tang, P. Zamore, M. Barton & D. Bartel (2004) MicroRNA control of PHABULOSA in leaf development: importance of pairing to the microRNA 5 ' region. *Embo Journal*, 23, 3356-3364.

Murray, M. & W. THOMPSON (1980) Rapid isolation of high molecular-weight plant DNA. *Nucleic Acids Research*, 8, 4321-4325.

Nabholz, B., G. Sarah, F. Sabot, M. Ruiz, H. Adam, S. Nidelet, A. Ghesquiere, S. Santoni, J. David & S. Glemin (2014) Transcriptome population genomics reveals severe bottleneck and domestication cost in the African rice (*Oryza glaberrima*). *Molecular Ecology*, 23, 2210-2227.

Ndowora, T., G. Dahal, D. LaFleur, G. Harper, R. Hull, N. Olszewski & B. Lockhart (1999) Evidence that badnavirus infection in Musa can originate from integrated pararetroviral sequences. *Virology*, 255, 214-220.

Nicaise, V. (2014) Crop immunity against viruses: outcomes and future challenges. *Frontiers in Plant Science*, 5, 660.

Parent, J., N. Bouteiller, T. Elmayan & H. Vaucheret (2015) Respective contributions of Arabidopsis DCL2 and DCL4 to RNA silencing. *Plant Journal*, 81, 223-232.

Pooggin, M., J. Futterer, K. Skryabin & T. Hohn (1999) A short open reading frame terminating in front of a stable hairpin is the conserved feature in pregenomic RNA leaders of plant pararetroviruses. *Journal of General Virology*, 80, 2217-2228.

Pooggin, M., R. Rajeswaran, M. Schepetilnikov & L. Ryabova (2012) Short ORF-Dependent Ribosome Shunting Operates in an RNA Picorna-Like Virus and a DNA Pararetrovirus that Cause Rice Tungro Disease. *Plos Pathogens*, 8, 3.

Pooggin, M. M., & Ryabova, L. A. (2018). Ribosome Shunting, Polycistronic Translation, and Evasion of Antiviral Defenses in Plant Pararetroviruses and Beyond. *Frontiers in microbiology*, *9*, 644.

Pooggin, M., L. Ryabova, X. He, J. Futterer & T. Hohn (2006) Mechanism of ribosome shunting in Rice tungro bacilliform pararetrovirus. *Rna*, 12, 841-850.

Rajeswaran, R., V. Golyaev, J. Seguin, A. Zvereva, L. Farinelli & M. Pooggin (2014) Interactions of Rice Tungro Bacilliform Pararetrovirus and Its Protein P4 with Plant RNA-Silencing Machinery. *Molecular Plant-Microbe Interactions*, 27, 1370-1378.

Richert-Poggeler, K., F. Noreen, T. Schwarzacher, G. Harper & T. Hohn (2003) Induction of infectious petunia vein clearing (pararetro) virus from endogenous provirus in petunia. *Embo Journal*, 22, 4836-4845.

Rosa, J., I. Suzuki, M. Kravicz, A. Caron, A. Pupo, F. Praca & M. Bentley (2018) Current Non-viral siRNA Delivery Systems as a Promising Treatment of Skin Diseases. *Current Pharmaceutical Design*, 24, 2644-2663.

Saito, N., J. Encabo, S. Chen, G. Jonson, Y. Kishima & I. Choi (2019) Innate Vulnerability of *Oryza glaberrima* to Rice tungro bacilliform virus. *Jarq-Japan Agricultural Research Quarterly*, 53, 1-6.

Sanfaçon, H., J. Wellink, O. Le Gall, A. Karasev, R. van der Vlugt & T. Wetzel (2009) Secoviridae: a proposed family of plant viruses within the order Picornavirales that combines the families Sequiviridae and Comoviridae, the unassigned genera Cheravirus and Sadwavirus, and the proposed genus Torradovirus. *Archives of Virology*, 154, 899-907.

Satoh, K., H. Kondoh, T. Sasaya, T. Shimizu, I. Choi, T. Omura & S. Kikuchi (2010) Selective modification of rice (*Oryza sativa*) gene expression by rice stripe virus infection. *Journal of General Virology*, 91, 294-305.

Schmieder, R. & R. Edwards (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, 27, 863-864.

Senthil, G., H. Liu, V. Puram, A. Clark, A. Stromberg & M. Goodin (2005) Specific and common changes in Nicotiana benthamiana gene expression in response to infection by enveloped viruses. *Journal of General Virology*, 86, 2615-2625.

Shahjahan, M., B. Jalani, A. Zakri, T. Imbe & O. Othman (1990) Inheritance of Tolerance to Rice Tungro Bacilliform Virus (RTBV) in Rice (*Oryza*. *Sativa* L). *Theoretical and Applied Genetics*, 80, 513-517.

Shen, P., M. Kaniewska, C. Smith & R. Beachy (1993) Nucleotide Sequence and Genomic Organization of Rice Tungro Spherical Virus. *Virology*, 193, 621-630.

Shibata, Y., R. Cabunagan, P. Cabauatan & I. Choi (2007) Characterization of *Oryza* rufipogon-derived resistance to tungro disease in rice. *Plant Disease*, 91, 1386-1391.

Shimizu, T., K. Satoh, S. Kikuchi & T. Omura (2007) The repression of cell wall- and plastid-related genes and the induction of defense-related genes in rice plants infected with Rice dwarf virus. *Molecular Plant-Microbe Interactions*, 20, 247-254.

Staginnus, C., W. Gregor, M. Mette, C. Teo, E. Borroto-Fernandez, M. Machado, M. Matzke & T. Schwarzacher (2007) Endogenous pararetroviral sequences in tomato (Solanum lycopersicum) and related species. *Bmc Plant Biology*, 7, 24.

Staginnus, C. & K. Richert-Poggeler (2006) Endogenous pararetroviruses: two-faced travelers in the plant genome. *Trends in Plant Science*, 11, 485-491.

Tajima, F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123, 585-595.

Várallyay, E., A. Valoczi, A. Agyi, J. Burgyan & Z. Havelda (2010) Plant virus-mediated induction of miR168 is associated with repression of ARGONAUTE1 accumulation. *Embo Journal*, 29, 3507-3519.

Vaucheret, H. (2006) Post-transcriptional small RNA pathways in plants: mechanisms and regulations. *Genes & Development*, 20, 759-771.

Wang, A. & S. Krishnaswamy (2012) Eukaryotic translation initiation factor 4E-mediated recessive resistance to plant viruses and its utility in crop improvement. *Molecular Plant Pathology*, 13, 795-803.

Wang, L., R. Liang & C. Chim (2012) Methylation of tumor suppressor microRNAs: lessons from lymphoid malignancies. *Expert Review of Molecular Diagnostics*, 12, 755-765.

Wang, M., Y. Yu, G. Haberer, P. Marri, C. Fan, J. Goicoechea, A. Zuccolo, X. Song, D. Kudrna, J. Ammiraju, R. Cossu, C. Maldonado, J. Chen, S. Lee, N. Sisneros, K. de Baynast, W. Golser, M. Wissotski, W. Kim, P. Sanchez, M. Ndjiondjop, K. Sanni, M. Long, J. Carney, O. Panaud, T. Wicker, C. Machado, M. Chen, K. Mayer, S. Rounsley & R. Wing (2014) The genome sequence of African rice (*Oryza glaberrima*) and evidence for independent domestication. *Nature Genetics*, 46, 982-988.

Wu, J., D. Wang, Y. Liu, L. Wang, X. Qiao & S. Zhang (2014) Identification of miRNAs involved in pear fruit development and quality. *Bmc Genomics*, 15, 953.

Wu, J., Z. Yang, Y. Wang, L. Zheng, R. Ye, Y. Ji, S. Zhao, S. Ji, R. Liu, L. Xu, H. Zheng, Y. Zhou, X. Zhang, X. Cao, L. Xie, Z. Wu, Y. Qi & Y. Li (2015) Viral-Inducible Argonaute18 Confers Broad-Spectrum Virus Resistance in Rice by Sequestering A Host MicroRNA. *Elife*, 4.

Yang, J., F. Zhang, J. Li, J. Chen & H. Zhang (2016) Integrative Analysis of the microRNAome and Transcriptome Illuminates the Response of Susceptible Rice Plants to Rice Stripe Virus. *Plos One*, 11, 1.

Yang, X. & L. Li (2011) miRDeep-P: a computational tool for analyzing the microRNA transcriptome in plants. *Bioinformatics*, 27, 2614-2615.

Zarreen, F., G. Kumar, A. Johnson & I. Dasgupta (2018) Small RNA-based interactions between rice and the viruses which cause the tungro disease. *Virology*, 523, 64-73.

Zheng, S., J. Li, L. Ma, H. Wang, H. Zhou, E. Ni, D. Jiang, Z. Liu & C. Zhuang (2019) OsAGO2 controls ROS production and the initiation of tapetal PCD by epigenetically regulating OsHXK1 expression in rice anthers. *Proceedings of the National Academy of Sciences of the United States of America*, 116, 7549-7558.

Zheng, X. & S. Ge (2010) Ecological divergence in the presence of gene flow in two closely related *Oryza* species (*Oryza* rufipogon and O. nivara). Molecular Ecology, 19, 2439-2454.

Zhao, Y., M. Wang, Z. Wang, R. Fang, X. Wang & Y. Jia (2015) Dynamic and Coordinated Expression Changes of Rice Small RNAs in Response to Xanthomonas *Oryza*e pv. *Oryza*e. *Journal of Genetics and Genomics*, 42, 625-637.