5.0 DISCUSSION

In this study, two receptor-like kinase genes, putative *Receptor-like Protein Kinase 1* (putative *RPK1*) and putative *CLAVATA1 Receptor Kinase* (putative *CLV1*), were selected for study as potential yield-related genes in rice. The putative *RPK1* is one of fourteen predicted genes from a 172 kb genomic DNA region of wild rice *Oryza rufipogon* (accession number IRGC 105491) around the RM 5 locus associated with the yield-enhancing QTL *yld1.1* (Song *et al*., 2009). The putative RPK1 was suspected to have a similar function to the Leucine-rich Repeat receptor-like Kinase1 (LRK1) protein because both proteins belong to the LRR-like type transmembrane receptor kinase group, and were derived from yield-enhancing QTLs. The LRK1 protein was identified experimentally to contribute to yield increase in rice, where over-expression study of *LRK1* showed that it regulates rice branch number (He *et al*., 2006; Zha *et al*., 2009). The putative *CLV1* was derived from a patented sequence (WO/2000/004761). However, the patented document did not name any specific species of rice. The patented sequence has been claimed to increase growth and yield ability in rice (Zhong *et al*., 2000). Putative CLV1 was also categorised into the transmembrane receptor kinase group. Although a large number of LRR-like type transmembrane receptor kinase groups are known in plants, only a small fraction of members have identified biological roles, especially rice yield traits. According to He *et al*. (2006), the LRR-like type receptor-like protein kinase gene cluster has a potential to contribute to increased rice yield. Therefore, putative *RPK1* and putative *CLV1* were selected as candidate yield-related genes.

Plant receptor-like kinases are grouped into the RLK/*Pelle* family. In general, the function of plant RLK/*Pelle* family member can be categorised into defence or developmental systems. Therefore, putative *RPK1* and *CLV1* genes may either be involved in developmental or plant defence system. According to Afzal *et al*. (2008), the function of plant RLK/*Pelle* gene family has become more complicated as a result of the multiple ligand-receptor recognition or the cross talk between disease and developmental pathways. For example, two well studied proteins, ERECTA and BRASSINOSTEROID-INSENSITIVE 1-associated receptor kinase1, are identified experimentally to be involved in both developmental and defence systems in *Arabidopsis thaliana* (Eckardt, 2005; Godiard *et al*., 2003; Chinchilla *et al*., 2007; Pillitteri *et al*., 2007). Thus, putative *RPK1* and *CLV1* genes could be involved in both plant defence and developmental pathways.

5.1 Gene Structure of Putative *RPK1* and Putative *CLV1*

From the previously reported coding region and gene structure of the *Oruf_RPK1* (*Oryza rufipogon* putative *RPK1*) by Song *et al*. (2009), an 855 bp open reading frame (ORF) sequence and 284 amino acid residues of *Oruf_RPK1* were predicted by a combination of *ab initio* and homology search approaches. However, 2,055 bp ORF sequence of *Oruf_RPK1* and *OsI_RPK1* (*Oryza sativa* ssp. *indica* cv. MR219 putative *RPK1*) were experimentally identified (Table 4.1). The ORF sequence of *Oruf_RPK1* and *OsI_RPK1* were predicted to encode a polypeptide of 684 amino acids (Table 4.1). The ORF sequence length and amino acid residues of *Oruf_RPK1* and *OsI_RPK1*

were different compared with those previously reported in *Oryza rufipogon* by Song *et al*. (2009). This is because the GC content of *Oruf_RPK1* and *OsI_RPK1* were shown to be greater than 70 % (Table 4.1). The unusually high GC content of putative *RPK1* might be a major contributor to inaccurate rice genome annotation by in silico prediction (Raghuvansh *et al*., 2009). This same situation has occurred during initial stage of rice genome annotation, where more than 40,000 over-estimated protein coding genes were annotated by *ab initio* gene prediction methods (Itoh *et al*., 2007; Bennetzen *et al*., 2004). Interestingly, the coding region and structure of Oruf_RPK1, OsI_RPK1 and OsJ_RPK1 (*Oryza sativa* ssp. *japonica* putative RPK1; gi:18677097) were shown to be 99 % identical (Figure 4.8). Therefore, it is suggested that putative RPK1 plays an important role in evolution selection because the Oruf_RPK1, OsI_RPK1 and OsJ_RPK1 (gi:18677097) were shown to be highly conserved in their coding regions (Song *et al*., 2009). Southern hybridization analysis showed that one to three copies of putative *RPK1* probably may be present in the genomes of *Oryza rufipogon* and *Oryza sativa* ssp. *indica* cv. MR219 genomes (Figure 4.22).

A total of five non-synonymous substitutions SNPs were identified among Oruf_RPK1, OsI_RPK1 and OsJ_RPK1 (gi:18677097) (Figur 4.8). Three of these are located at positions 139, 194 and 213, within a non-conversed region. These three SNPs result in amino acid changes with no influence to the charge. However, two other SNPs were identified within the highly conversed superfamily kinase domain, at positions 478 and 491. This result suggested that the positively charged Arg^{478} of

OsJ_RPK1 and Oruf_RPK1 were replaced by a neutral Gln^{478} in OsI_RPK1. Another amino acid substitution at position 491 involves a change from His^{491} (OsI_RPK1 and OsJ_RPK1) to Arg^{491} (Oruf_RPK1). Single nucleotide substitutions and amino acid substitutions may be due to rice domestication. The mutations presumably occurred prior to the divergence between *japonica* and *indica* subspecies and they might be involved in complex patterns of subpopulation isolation and convergence (Sweeney *et al*., 2007). Nevertheless, the amino acid substitutions would lead to a minor change in the secondary protein structure in these putative RPK1s. They are also suspected to be associated with the differences of putative *RPK1* gene expression levels among the rice studied (Figure 4.20), or could impact upon their functional significance.

The patented sequence (WO/2000/004761) was searched using Basic Local Alignment Search Tool (BLAST) database (Karlin and Altschul, 1993). OsJ_CLV1 (*Oryza sativa* ssp. *japonica* putative CLV1; gi:125602183) was one of the highest score protein in the database, which showed 90 % identity to the patented sequence (WO/2000/004761). The primers were aligned with regions showing 100 % identity between patented sequence (WO/2000/004761) and *OsJ_CLV1* (gi:125602183). Amplified partial 402 bp long gene sequences of *Oruf_CLV1* (*Oryza rufipogon* putative *CLV1*) and *OsI_CLV1* (*Oryza sativa* ssp. *indica* cv. MR219 putative *CLV1*) showed 99 % and 100 % identity respectively to *OsJ_CLV1* (gi:125602183); whereas, amplified partial 402 bp long gene sequences of *Oruf_CLV1* and *OsI_CLV1* only showed 92 % and 91 % identity respectively to the patented sequence

(WO/2000/004761). According to Buell (2002), simple sequencing errors (incorrect bases and low quality regions) and misassemblies are the main disadvantages for drafting the rice sequence. With potential errors like this, there could be a posibility that patented sequence (WO/2000/004761) and *OsJ_CLV1* (gi:125602183) may be the same gene. Southern hybridization analysis suggests that three to five copies of putative *CLV1* may be present in the genomes of *Oryza rufipogon* and *Oryza sativa* ssp. *indica* cv. MR219 (Figure 4.23). At the moment, at least six orthologous *CLV1* gene sequences have been identified in the rice genome (Miwa *et al*., 2009). Most of them may be involved in plant development and environmental responses (Miwa *et al*., 2009).

Based on sequence analysis, Oruf_RPK1, OsI_RPK1, OsJ_RPK1 (gi:18677097) and OsJ_CLV1 (gi:125602183) were identified and categorised into the LRR-like type transmembrane receptor kinase group. This group represents the largest group within the plant RLK/*Pelle* family (Diévart and Clark, 2004; Afzal *et al*., 2008). Its proteins contain an extracellular domain connected to an intracellular cytoplasmic protein kinase via a single pass transmembrane helix (Shiu and Bleecker, 2001a; Diévart and Clark, 2004). Several characteristic motifs of the extracellular domain of Oruf_RPK1, OsI_RPK1, OsJ_RPK1 (gi:18677097) and OsJ_CLV1 (gi:125602183) were identified, including signal peptide and LRR motif. Signal peptide regions of transmembrane receptor kinases are responsible for targeting cellular compartmentalisation (Eisenhaber *et al*., 2003). According to Afzal *et al*. (2008), it is

hypothesized that the LRR motif structure of transmembrane receptor kinases might be correlated to plant RLK/*Pelle* family functional categories, i.e. developmental or defence systems. However, the LRR-like type ERECTA of *Arabidopsis thaliana* was identified to be involved both in ovule development and in resistance to bacterial wilt (Godiard *et al*., 2003; Pillitteri *et al*., 2007). Although the function of Oruf_RPK1, OsI_RPK1, OsJ_RPK1 (gi:18677097) and OsJ_CLV1 (gi:125602183) have not yet been identified, they might be involved in either or both developmental and defence systems.

The putative RPK1 and putative CLV1 genes possesses a cytoplasmic protein kinase. The protein kinase activity of the RLK family plays a vital role in the regulatory mechanisms for protein activity and cellular signaling, through specific phosphorylation events (Johnson *et al*., 1996; Dardick and Ronald, 2006). OsJ_CLV1 (gi:125602183) was categorised into the RD kinase class in the plant RLK/*Pelle* family based on the presence of a conserved lysine (K) kinase subdomain II, conserved arginine (R) and aspartic acid (D) in kinase subdomain VIb, and aspartic acid (D) in kinase subdomain VII (Figure 4.14). Subdomain IVb of the RD kinase class contains an activation loop involved in regulation of phosphorylation and dephosphorylation (Nolen *et al*., 2004). According to Dardick and Ronald (2006), RD kinase class in the plant RLK/*Pelle* family members may be involved in development. The well studied RD kinase class members, *Arabidopsis thaliana* CLV1 and rice FLORAL NUMBER 1 (FON1), are also involved in cell growth and development

(Torii *et al*., 1996; Clark *et al*., 1997; Suzaki *et al*., 2004; Dardick and Ronald, 2006; Afzal *et al*., 2008).

Oruf_RPK1, OsI_RPK1 and OsJ_RPK1 (gi:18677097) were categorised into the RD minus kinase class in the plant RLK/*Pelle* family based on the absence of conserved arginine (R) and aspartic acid (D) in kinase subdomain VIb (Figure 4.14). Mutation of conserved residues of the RD minus class kinase subdomain IVb impairs autophosphorylation (Nolen *et al*., 2004). However, function or mechanism of RD minus class kinases is not fully understood (Afzal *et al*., 2008). Interestingly, conserved glycine (G) and asparagine (N) residues in subdomain IVb of the RD minus kinase class Oruf_RPK1 (382-643), OsI_RPK1 and OsJ_RPK1 (gi:18677097; 382-643) are also present in the RD minus kinase class *Glycine max* receptor-like kinase protein Gm_RHG1 (gi:206584433; 578-834) shown in Figure 4.14. Gm_RHG1 (gi:206584433) is reported to be involved in plant defence systems (Afzal and Lightfoot, 2007). Although only conserved arginine (R) and aspartic acid (D) residues in subdomain IVb have been determined, it could be suggested that the conserved glycine (G) and asparagine (N) residues in subdomain IVb of RD minus kinase class Oruf_RPK1, OsI_RPK1 and OsJ_RPK1 (gi:18677097) may imply a similar function in plant defence.

5.2 Phylogenetic Analysis

Based on the phylogentic analysis, the putative RPK1 and putative CLV1 were grouped into LRR-III and LRR-XI subfamilies respectively in the RLK/*Pelle* family (Figure 4.15 and Figure 4.16). According to Shiu *et al*. (2004), RLKs have been duplicated at a higher level in *Arabidopsis thaliana* and *Oryza sativa*. This could be due to involvement in the plant defence system. If they are involved in plant growth and development, RLKs have rarely been duplicated after the *Arabidopsis thaliana*-*Oryza sativa* split (Shiu *et al*., 2004). The LRR-III RLKs and LRR-XI RLKs subfamilies were catogerised into a highly duplicated subfamily (Shiu *et al*., 2004). Therefore, both putative RPK1 and putative CLV1 are suggested to be involved in plant defence.

According to Castells and Casacuberta (2007), most members of the LRR-III subfamily in the RLK/*Pelle* family were identified as atypical RLK because of an aspartate mutation to arginine in subdomain VIb. This is also observed in Oruf_RPK1, OsI_RPK1 and OsJ_RPK1 (gi:18677097; Figure 4.14). Orthologous RPK1 amino acid sequences, *Zea mays* atypical receptor kinase (MARK) and *Arabidopsis thaliana* Transmembrane Kinase-Like 1 (TMKL1), were reported to contain a kinase-defective atypical receptor kinase, which is involved in transducing signals by a phosphorylation independent mechanism (Valon *et al*., 1993; Castells and Casacuberta, 2007). Kinase-defective atypical receptor kinase domains are able to retain the power to interact with downstream effectors (Chevalier *et al*., 2005).

Therefore, Oruf_RPK1, OsI_RPK1 and OsJ_RPK1 (gi:18677097) might be involved in phosphorylation-independent mechanisms for signal transduction.

5.3 Potential Effects of Putative *RPK1* and Putative *CLV1* Alleles on Yield

Two candidate yield-related genes, putative *RPK1* and putative *CLV1*, were up-regulated during transition from the vegetative to flowering phases. During rice development, the transition from vegetative to flowering phase is the most critical phase, as it can directly or indirectly influence the potential yield in rice production (Wang and Li, 2005; Dwivedi *et al*., 2008). The transition from vegetative to flowering phase could be regulated by inducers or repressors (Tang *et al*., 2005). During the transition period, rice inflorescences are generated. The rice inflorescences consist of spikelets on each panicle branch. When spikelets are formed at the booting stage, the gametes are generated by meiosis (Nonomura *et al*., 2004; Martinez-Perez *et al*., 2008). Gene expression of putative *RPK1* of backcross progenies and its parental lines were up-regulated at the booting stage (Figure 4.20). *Arabidopsis thaliana* LRR1 (At5g16590), an ortholog of RPK1, interacted with three other receptor-like kinases (At3g17840, At1g48480, At3g02880) that were reported to be involved in gametophyte development (Gray-Mitsumune *et al*., 2008). Therefore, it is suggested that putative *RPK1* may be involved in gametophyte development.

After spikelet meristems and branches form, floral meristems are generated (Wang and Li, 2005). Gene expression of putative *CLV1* of backcross progenies and

its parental lines were up-regulated at booting and heading stages (Figure 4.21). Two well-studied orthologous of CLV1, rice FON1 and rice leucine-rich repeat receptor-like protein kinase1 (OsLRK1), are involved in inflorescence architecture and rice floral meristem development (Kim *et al*., 2000; Suzaki *et al*., 2004; Dwivedi *et al*., 2008). Therefore, it is suggested that the putative *CLV1* sequences in this study may be involved and possibly correlated with floral meristem development.

Although putative *RPK1* from *Oryza rufipogon* was suspected to contribute to the yield-enhancing QTL *yld1.1* (Song *et al*., 2009), BC_2F_7 line 23 relatively low yield progeny was identified to be homozygous for putative *RPK1* and putative *CLV1* alleles from *Oryza rufipogon* (Figure 4.4 and Figure 4.12). However, the expression profiles of putative *RPK1* and putative *CLV1* did not correlate between *Oryza rufipogon* and BC_2F_7 line 23 (Figure 4.20; Figure 4.21). Expression levels of putative *RPK1* and putative *CLV1* in BC_2F_7 line 23 were severely reduced relative to two housekeeping genes, *eEF-1α* and *UBQ5*, particularly during transition phase from vegetative to reproductive phase (Figure 4.20; Figure 4.21). The variation of gene expression can be the result of natural allelic variation or transcript abundance between different genotypes (Benfey and Mitchell-Olds, 2008).

BC_2F_7 line 7 relatively high yield progeny was identified to have inherited homozygous putative *RPK1* and putative *CLV1* alleles from *Oryza sativa* ssp. *indica* cv. MR219 (Figure 4.4 and Figure 4.12). However, the correlation of putative *RPK1*

and putative *CLV1* expression between *Oryza sativa* ssp. *indica* cv. MR219 and BC_2F_7 line 7 were not observed (Figure 4.20; Figure 4.21). The BC_2F_7 line 7's field trial performance for yield component was better than that of the parental lines (Appendix H). Therefore, optimum expression level of putative *RPK1* in BC_2F_7 line 7 might lead to better performance in rice yield related components at the booting stage (Figure 4.20). Meanwhile, expression level of putative *CLV1* in BC_2F_7 line 7 was highly up-regulated relative to the *eEF-1α* and *UBQ5*, especially at booting and heading stages (Figure 4.21). Thus, both of the putative *RPK1* and putative *CLV1* alleles from *Oryza sativa* ssp. *indica* cv. MR219 may correlate with enhanced yield performance in BC_2F_7 line 7. However, real time qRT-PCR profiles of putative *RPK1* and putative *CLV1* are not enough to directly determine rice yield performance. Yield is complex trait, possibly due to the effect of polygenic inheritance. Both genes may be involved in pathways contributing to yield performance in rice.