Haplotype Diversity Analysis of Bacterial Leaf Blight Resistance Gene *Xa21* in Rice

N.H.L.D.L.D. Nanayakkara, V. Herath[1] and D.V. Jayatilake[1*]

Postgraduate Institute of Agriculture
University of Peradeniya
Sri Lanka

ABSTRACT: *Bacterial leaf blight (BLB) caused by Xanthomonasoryzaepv. oryzae (Xoo) is a devastative rice disease. Resistance gene Xa21 is known to convey a stable resistance against a broad spectrum of Xoo races. Assessment of Xa21 genetic diversity and defining haplotypes is the initial step towards identification of haplotypes that associate with desired phenotypes. In this study, an alignment of a 2,778 bp coding sequence (CDS) from 1,618 rice accessions revealed 18 single nucleotide polymorphisms and six insertion/deletions (InDel), each with ≥ 5% occurrence among the accessions. A total of 13 confirmed-haplotypes (≥ 1% occurrence) and 113 putative haplotypes (≤ 1 % occurrence) were identified. The 13 confirmed haplotypes were represented by 1,341 rice accessions that deferred at 22 sites belonging to 12 rice subpopulations from 70 countries. The Xa21 CDS has a haplotype diversity of 0.8203, and a nucleotide diversity of 0.15448. Further, the haplotype network shows a high substitution rate between the haplotype pairs of Xa21, with two distinctive clusters created as a result of a 19-bp InDel in CDS. Based on 14 non-synonymous sites identified within the CDS, nine protein-types were identified. Of these three protein-types are truncated owing to a premature stop codon at 180th position of the polypeptide chain due to the said 19-bp InDel in the ectodomain, and represents 45% of the accessions. The ectodomain of XA21 protein is crucial for immune responses, and hence, these haplotypes with truncated proteins could be important in eliciting the function of XA21.*

Keywords: *Bacterial Leaf blight, Haplotype analysis, Xanthomonasoryzae pv. oryzae, Xa21*

INTRODUCTION

Rice (*Oryza sativa* L.) was the first cereal to get a fully assembled, well-curated 389 Mbp genome (International Rice Genome Sequencing Project 2005). The availability of such a genome has made the understanding of the rice genetics, and detection of genotypic variations a much easier task (Jackson, 2016). The first map-based sequencing of rice was done using the *japonica* rice cultivar "Nipponbare", and since nearly 3000 accessions including both *indica* and *japonica* types from all over the world have been sequenced, and are publicly available through the 3K Rice Genomes Project (Li *et al*., 2014): a wealth of information important for the assessment of existing genetic diversity, and for the discovery of novel alleles.

[1] Department of Agricultural Biology, Faculty of Agriculture, University of Peradeniya, Sri Lanka
* Corresponding author:djayatilake@yahoo.com

Bacterial leaf blight (BLB) caused by *Xanthomonas oryzae* pv. *oryzae* (*Xoo*) is one of the most devastative diseases in rice with the potential of causing ~50 - 80% crop loss under severe epidemics (Mew, 1993; Huang *et al.*, 1997; Deshmukh *et al.*, 2017). In the rice genome more than 40 genes conferring resistance against BLB has been identified (Dilla-Ermita *et al.*, 2017). Among them theresistance allele of *Xa21* is unique as it is known to deliver the highest degree of stable and broad-spectrum resistance against a diverse range of *Xoo* strains (Ikeda *et al*.,1990; Khush *et al*., 1991). The *Xa21* was introgressed in to the modern rice varieties from the rice wild relative *Oryza longistaminata*, and is physically mapped to the rice chromosome 11 (Ronald *et al.*, 1992). The *Xa21* was cloned and characterized as a gene coding for a receptor kinase-like protein (Song *et al.*, 1995), and is involved in plant pathogen recognition, and activation of defense responses (Century *et al.*, 1999; Gnanamanickam, 1999). Initially it was proposed that the nuclear localization signal (NLS) of the cytoplasmic domain of XA21 was the key to *Xa21*-mediated immune responses (Park and Ronald, 2012). However, through transgenic lines containing an alanine substitution at the NLS of XA21 kinase domain developed through site-directed mutagenesis the earlier theory on the significance of NLS in expression of the resistant phenotypes were dismissed (Wei *et al.*, 2016). Recently, it was proposed that even in the absence of a functional kinase domain, the ectodomain of XA21,and its recognition of RaxX (*Xa21*-mediated immunity X protein produced by *Xoo*) can mediate the expression of a robust immune response (Thomas *et al.*, 2018).

Domestication of rice has made the genetic structure of cultivated rice (both *indica* and *japonica* sub species) undergo genetic enhancements through complex introgressions (Garris *et al.*, 2005; Zhao *et al.*, 2010). Over the years the evolutionary process has added to the genetic diversity through mutations, and natural selection. Haplotype analysis is a platform to understand and estimate the genetic diversity. Haplotypes are genomic sequence variants, which co-occur along a single chromosome/ a single parent (Snyder *et al.*, 2015). The inherited genetic variations have created diversity among these haplotypes, and the assessment of the existing diversity into haplotype blocks generates knowledge for effective utilization of the existing rice germplasm. A haplotype analysis will illustrate the genetic patterns in a deeper context with higher precision, and it will assist in crop improvement programs. Further, it is the first step towards grouping accessions in a meaningful manner to workout functional relationships, and design molecular markers to specifically identify the desired haplotypes.

So far, the haplotype diversity has been analyzed for BLB resistance genes *Xa4, Xa26, xa25* and *xa13* (Bimolata *et al.*, 2015; Zhang *et al.*, 2017; Yu *et al.*, 2018), however, haplotypes have not been defined for *Xa21*, the gene conferring the highest stable BLB resistance against a broad–spectrum of *Xoo* races. Here we report, the defining of genomic haplotypes and protein-types for *Xa21* using a diverse rice germplasm collection consisting of 1,341 rice accessions representing 70 countries, and 12 rice subpopulations.

METHODOLOGY

Multiple alignment of *Xa21* genomic sequences and haplotype analysis

To identify the genomic variations in the coding sequence (CDS) of *Xa21*, and to define genomic haplotypes, a 2,748-bp CDS was retrieved from the 3000 Rice Genome Project sequence repositoryat the Rice SNP-Seek Database (Mansueto *et al*., 2016; http://snp-

seek.irri.org) for a panel of 2,896 rice accessions, using the position of *Xa21* in the *Oryza sativa* subsp. *japonica* genomic reference line "Nipponbare" (*Os11g0569733;* synonym *LOC_Os11g36180*; at position chr11: 21274696.21277443 in IRGSP-1.0 genome assembly). The sequences corresponding to the same region was retrieved from *O. sativa* subsp. *indica* genomic reference "Cultivar 93-11" (*BGIOSGA033042*) from Gramene (Tello-Ruiz *et al*., 2018; http://www.gramene.org). The retrieved 2,898 sequences were aligned using multiple alignment feature in ClustalOmega v1.2.0 (Goujon *et al.,* 2010) using UGENE v1.28.1 (Okonechnikov *et al.*, 2012) with manual editing.

In the alignment variable sites (single nucleotide polymorphism (SNP) and insertion/deletion (InDel) sites) with more than 5% occurrence in 2,898 accessions were identified using Geneious v7.1 (Drummond *et al.,* 2011). Sequences carrying ambiguous nucleotide calls at these identified SNP/ InDel sites were excluded from further analysis. Based on the above selected sites haplotypes were identified using DnaSP v6(Rozas *et al.*, 2017). Of the identified haplotypes only the haplotypes that are representing at least1%of the accessions were considered as confirmed haplotypes, and the others were considered as putative haplotypes. For the selected population the haplotype diversity and the nucleotide diversity was calculated using DnaSP v6.

Haplotype networks

In order to understand the relationships of the different haplotypes a haplotype network (of confirmed haplotypes) was constructed using Network v5.0.0.3 (www.fluxus-engineering.com) with median-joining calculation method *(*Bandelt *et al*.,1999), an epsilon of zero, an equal weight of 10 and a connection cost criterion (other analysis parameters set as default). The pie-slices of the network diagram were colour coded according to the representative subpopulations based on the information retrieved from the accession passport entries in Rice SNP-Seek Database (Rozas *et al*., 2017; http://snp-seek.irri.org).

Protein-type analysis

As an initial step towards identifying the haplotypes that could have an association to the desired phenotypes, protein-types were defined based on the SNP/InDel sites that occurred at the exon. Here, only the SNP/InDel sites leading to non-synonymous mutations were considered after translating the CDS to the respective polypeptide sequence. The translation frame was decided based on the polypeptide sequence of XA21 *japonica* reference "Nipponbare" (*Os11t0569733*) and Cultivar 93-11 (*BGIOSGA033042*) retrieved from Gramene database (Tello-Ruiz *et al*., 2018; http://www.gramene.org).The domains and repeats of the polypeptide chain were identified via UniProt database (https://www.uniprot .org).

RESULTS AND DISCUSSION

The *Xa21* plays a vital role in conveying resistance against a wide range of *Xoo* races, and it is the first BLB resistance gene to get cloned, and characterized (Song *et al.,* 1995). Even though diversity assessments has been carried out for the *Xa4, Xa26, xa25* and *xa13* resistance loci (Bimolata *et al.,* 2015; Zhang *et al.,* 2017; Yu *et al.,* 2018), the genetic diversity of *Xa21* remains unknown. To get a well-represented estimation on the genetic diversity, it is important to make assessments based on a larger panel of accessions that represents a wider genetic background. The 3000 Rice Genomes Database presents genomic

sequences of rice accessions coming from 86 countries representing 12 subpopulations (Rozas *et al*., 2017; http://snp-seek.irri.org). In the current study, we report the assessment of haplotype and protein-type diversity of the BLB resistance gene *Xa21*, and the haplotype network illustrating their interrelationships using 1,341 rice accessions.

Thirteen confirmed *Xa21* haplotypes and nine protein-types

The *Xa21* resistance region was originally introgressed in to IR24 variety at International Rice Research Institute(IRRI), Philippines from rice wild relative *O. longistaminata*, and then subsequently to IRBB21 to be used as a donor of resistance in rice breeding programs (Khush *et al*.,1989; Ikeda *et al*.,1990; Khush *et al*.,1991). Through map-based cloning, *Xa21* resistance gene, originally transferred from *O. longistaminata* Genebank accession U37133, was identified as undergoing single gene inheritance conferring resistance to BLB(Song *et al.*, 1995; Wang *et al.*, 1996). In an expression study Peng *et al.*, (2015), identified *Os11g0569733* (synonym *LOC_Os11g36180*; at position chr11: 21274696..21277443 in IRGSP-1.0 genome assembly) as one of the highly expressed *Xa21* genes in resistant lines. The *Os11g0569733* comprises of one exon, and codes for a protein with 915 amino acids. So far 13 orthologs and 28 paralogs has been identified for *Os11g0569733*, with two one: oneorthologs in wild relatives *O. longistaminata* and *Oryza nivara* (Monaco *et al.*, 2014).

In the current study, 24 polymorphic sites (18 SNP and six InDel sites) were identified in the coding sequence of *Xa21*. These sites were selected on the basis that their occurrence frequency was more than 5%from an original diverse panel of 2, 898 rice accessions. Based on 24 selected polymorphic sites,only 1,618 rice accessions carrying no ambiguous nucleotide calls at the said 24 sites were selected for further analysis.

Given the availability of the high-throughput sequencing techniques, the diversity assessments have now focused on the genes, considering sequences of multiple individuals (Buckler and Thornsberry, 2002). Previously such diversity assessments has been conducted for several rice genes and the alleleic haplotypes have been defined,and diversity assessments has been carried out (Garris*et al*., 2003; de Jong *et al.*, 2011; Lu *et al.*, 2012; Ogiso-Tanaka *et al.*, 2013). For the BLB resistance locus *Xa21* haplotypic analysis, and genetic diversity assessments have not been reported. Here, we report defining of a 126 haplotypes, including 13 confirmed haplotypes (with $\geq$ 1% occurrence) and 113 putative haplotypes (with $\leq$ 1% occurrence; 60% singletons) in a panel of 1,618 rice accessions. The confirmed 13 haplotypes were represented in 1,341 rice accessions. The identified confirmed haplotypes are denoted as H1 to H13 (Figure 1).The highest occurrence frequency of a 33%was reported for H5, followed by H11 with an 18% occurrence frequency. Lowest occurrence was of haplotypes H8 and H9, with an occurrence frequency of 1.5% (Figure 1).

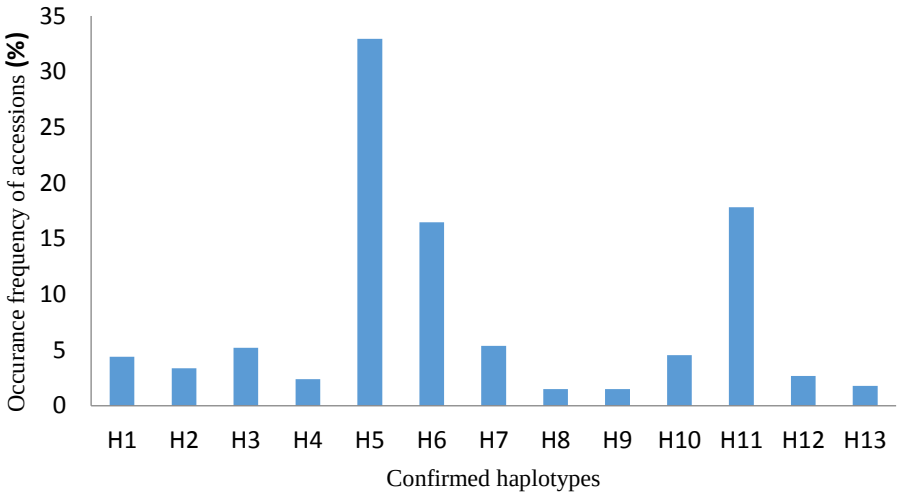
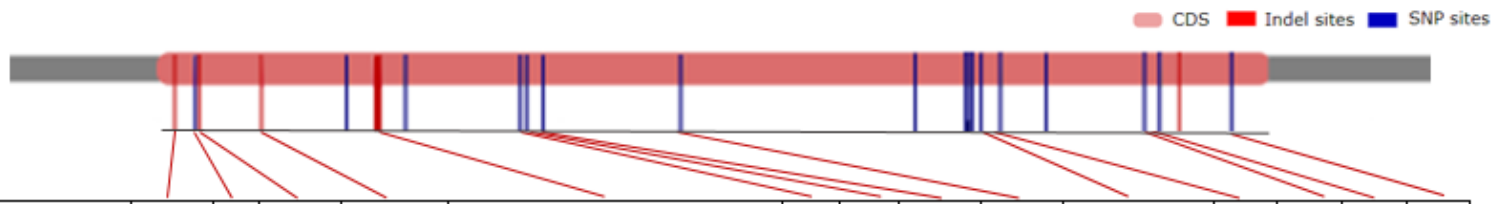


Figure 1. Occurrence frequency of accessions represented in the 13 confirmed *Xa21* haplotypes identified in a panel of 1,341 rice accessions

The non-synonymous mutations that occur in the CDS will lead to creating diversity at the protein level (Lu *et al.*, 2012). Based on these non-synonymous mutations it is possible to identify functional protein-types owing to the fact that mutations at functional domains could lead to a conformational change of the protein resulting altered functionality (Lu *et al.*, 2012; Wei *et al*., 2014; Li *et al*., 2015). In the current study, based on 14 such sites with non-synonymous mutations (10 SNPs and 4 InDel sites) within the CDS, nine protein-types XA21-1 to XA21-9) were identified (Figure 2). Among are three protein-types that were truncated as a result of premature stop codons at position 180 of the polypeptide sequence, created due to a 19-bp InDel in the genomic sequence (position 521-539 on the CDS).The said 19-bp InDel is within the defined exon region of *Xa21* genomic sequence of Nipponbare (*Os11g0569733*), but in the intron 1 of Cultivar 93-11 (*BGIOSGA033042*).

As a result, the protein gets truncated at the 180th amino acid that fall over the extracellular domain/ ectodomain. Recently it was reported that the ectodomain of XA21 protein (1-650), and its recognition of RaxX is critical for the immune responses (Pruitt *et al.*, 2015; Thomas *et al.*, 2018). In the truncated proteins only the first 178 amino acids remain. Therefore, it is very likely that these threeprotein-types will result null/pseudo proteins which may not be functional. Of the other six protein-types several had mutations on leucine rich repeats (LRR); LRR4, LRR9, LRR10 and LRR22, and on the NLS. The LRR repeats are previously reported to be associated with disease resistance conveyed *via Xa* genes (Caddell *et al.*, 2017), and hence, a mutation in these repeats could alter the function of the gene. Therefore, further studies to verify the identified protein-types through phenotypic assays are recommended.



Position from transcription start site	40..42	83	90..92	259..261	535..553	895	909	955	1869	2011	2082	2433	2475	2648
Domains		Extracellular domain										Kinase domain		
	Ectodomain									Helical domain	Cytoplasmic domain			
Repeats					LRR4	LRR9		LRR10	LRR22		NLS			
Nucleotide change	CTG	C	TGC	CGC	CGAGGTATGATCCCACGTG	G	C	G	G	C	A	G	A	T
	---	A	---	---	-------------------	A	T	A	A	T	C	A	C	C
Amino acid change	L	A	A	R	RGMIPRE	G	H	R	V	A	H	S	Q	F
	-	D	-	-	-------	D	Y	Q	M	V	N	N	P	S
XA21-1	-	D	-	R	-------	Truncated								
XA21-2	L	D	-	R	-------									
XA21-3	-	D	A	R	-------									
XA21-4	-	D	-	R	RGMIPRE	G	H	Q	M	V	H	S	P	F
XA21-5	-	D	A	R	RGMIPRE	D	H	R	M	V	H	S	P	F
XA21-6	-	D	A	R	RGMIPRE	D	H	R	V	A	N	N	Q	S
XA21-7	-	A	A	R	RGMIPRE	G	Y	Q	M	V	H	S	Q	F
XA21-8	-	A	A	R	RGMIPRE	G	Y	Q	M	V	H	S	P	F
XA21-9	L	D	-	-	RGMIPRE	D	H	R	V	A	N	N	Q	F

Figure 2. Position of the 14 non-synonymous sites on the gene and the identified nine protein-types for *Xa21*

A region with high genetic diversity

Based on the passport entries of the accessions in the confirmed haplotpes, country information was available only for 987 accessions, and they represented 70 countries and 12 *O. sativa* subpopulations (*admixture* (*admix*), *aromatic* (*aro*), *aus*, *indica* (*indx*, *ind2*, *ind3*, *ind1A* and *ind1B*), *japonica* (*japx*), *subtropical* (*subtrop*), *temperate* (*temp)*, and *tropical* (*trop*); Supplementary File 1 available at https://tar.sljol.info).The *aus, indica,* and *admix* subpopulations represents *indica* subspecies, while *temperate, tropical, sub-tropical, aromatic,* and *admix* represents *japonica* subspecies (Huang *et al.*, 2012; Mccouch *et al.*, 2016). Highest representation was from the subpopulations *indx* and *trop*, each accounting to 17%, followed by *temp* with a 16% representation. The subpopulations *aro* and *admix* was the least represented accounting to 2% of the total (Figure 3; Supplementary File 1 available at https://tar.sljol.info). Therefore, the selected panel was a good representative cohort to study the diversity of rice germplasm.

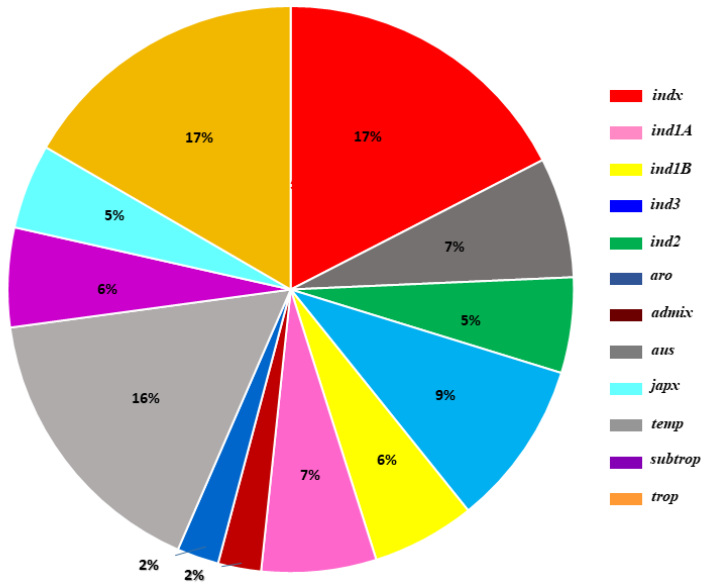


Figure 3. Occurrence frequency of the 12 subpopulations representing 1,341 rice accessions

High throughput sequencing techniques have resulted information that can be efficiently used to understand the underlying genetic diversity. This is an initial step towards assessing the evolutionary relationships, and sorting of haplotypes that can be used for the benefit of crop improvement programs (Buckler and Thornsberry, 2002). The haplotype network revealed the relationships of the identified 13 haplotypes (Figure 4). The subpopulations *indx* and *admix* are the most widely represented in the haplotypes given that they are only absent in H8 and H2, respectively. The least occurring subpopulation is *aro*, and it is only found in haplotypes H1, H2, H9 and H12.In general, each haplotype is represented by at least 5 subpopulations to 10 subpopulations. The finding indicates that the haplotypes are not specific to a narrow genetic background, but is present in a wider and divergent population. The occurrence of several polymorphic sites between haplotype pairs indicates that the considered region has a high substitution rate. The highest substitution rate was between the

H5 and H7, dividing the haplotypes in to two clusters based on the 19-bp InDel in the CDS. These findings indicate that the occurrence of the 19-bp InDel in the CDS pre-dates the structuring of rice in to subpopulations, and no geographic genetic structuring is observed with respect to *Xa21*.

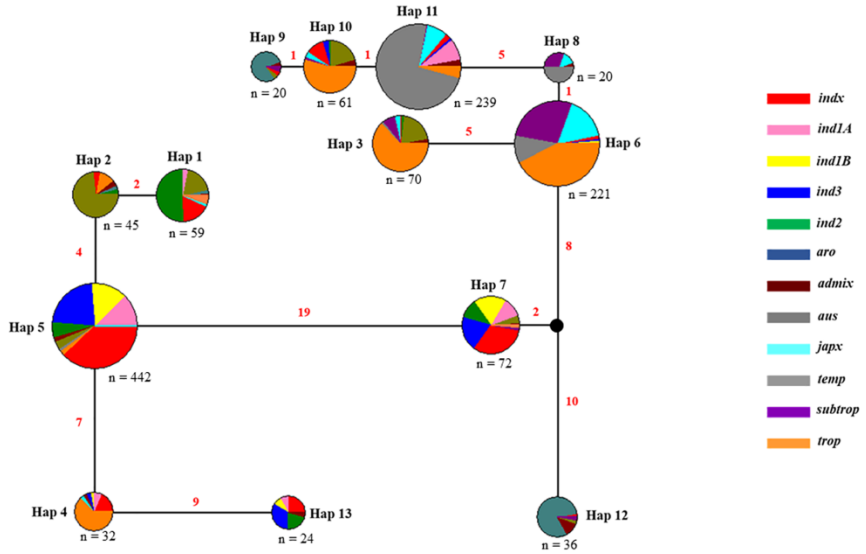


Figure 4. Haplotype network and subpopulation diversity of the 13 confirmed *Xa21* haplotypes (The colored pie slices correspond to subpopulations, and are drawn proportional to the occurrence frequency in each haplotype. The lines connecting the haplotypes illustrate the number of polymorphic sites between the haplotype pairs and n refers to the number of individuals represented in each haplotype. The black dot represents a hypothetical haplotype)

The advancements in the sequencing technologies has made it easier to assess the genetic variations in a DNA fragment representing a locus of interest (Genomics and Lee, 2011). Predictions made based on several SNPs/ InDel regions that are in linkage disequilibrium called haplotypes, are more effective in reflecting the actual phenotypic groupings than when considering single polymorphic sites (as reviewed by Rafalski, 2002). The genetic diversity within a population can be assessed based on the haplotype diversity (H) or based on the nucleotide diversity (π). The H (also referred to as the gene diversity) represents the probability of two randomly selected alleles to be different from each other (de Jong *et al.*, 2011). The π reflects the nucleotide divergence of two individuals with respect to one locus (Buckler and Thornsberry, 2002). In the current study, for the considered region among 1,341 rice accessions the H was 0.8203 and π was 0.15448. This indicates a higher genetic diversity in the population considered, and thus, will be a better representation of the existing rice germplasm for defining haplotypes. The reported high H for the CDS is in the range reported by previous research, and the reported π ismoderately high than in previous studies (H = 0.828, π = 0.292 for *RSUS3* in rice (Lestari *et al.*, 2011); H = 0.6822, π = 0.0188 for *RFT1* and H = 0.6292, π = 0.0048 for *Hd3a* (Ogiso-Tanaka *et al*., 2013).

This is the first reporting of haplotype and protein-types for the BLB resistance gene *Xa21*. The analysis reported here reflects the high genetic diversity of *Xa21* leading to the identification of confirmed haplotypes. The genetic divergence and the relationships between these haplotypes were established, and based on non-synonymous mutations protein-types were defined. Mutations at important repeats/ domains, and/or creation of truncated proteins were observed in the XA21 polypeptide. Some of these protein-types could be associated with desired resistance phenotypes. Hence, it is recommended to assess the association of resistance response of each protein-type identified in the current study, and develop molecular markers that can effectively screen the desirable haplotypes through marker-assisted selection.

CONCLUSIONS

The genetic diversity of the BLB resistance gene *Xa21* was assessed for the first time and based on the CDS 13 confirmed haplotypes, 113 putative haplotypes, and nine protein-types were defined based on 24 polymorphic sites identified from 1,618 rice accessions. The CDS of *Xa21* carried a relatively high H of 0.8203 and a moderate π of 0.15448, indicating high genetic diversity. The haplotype network indicates that the *Xa21* have a high substitution rate. Accessions were categorized in to two clusters based on a 19-bp InDel in the CDS.

ACKNOWLEDGEMENT

The authors wish to acknowledge the University Research grant (URG/2016/91/Ag) awarded by the University of Peradeniya, Sri Lanka and Kanishka Ukuwela of the Faculty of Applied Sciences of the Rajarata University of Sri Lanka for his advice on analysis.

REFERENCES

Bandelt, H. J., Forster, P. and Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. *16(1)*, 37–48.

Bimolata, W., Kumar, A., M, S. K. R., Sundaram, R. M., Laha, G. S., Qureshi, I. A. and Ghazi, I. A. (2015). Nucleotide diversity analysis of three major bacterial blight resistance genes in rice. PLoS One. *10(3)*, e0120186.

Buckler, E. S. and Thornsberry, J. M. (2002). Plant molecular diversity and applications to genomics. Curr. Opin. Plant. Biol. *5(2)*, 107–111.

Caddell, D. F., Park, C. J., Thomas, N. C., Canlas, P. E. and Ronald, P. C. (2017). Silencing of the rice gene *LRR1* compromises rice *Xa21* transcript accumulation and XA21-mediated immunity. Rice. *10(1)*, 1–11.

Century, K. S., Lagman, R. A., Adkisson, M., Morlan, J., Tobias, R., Schwartz, K., Smith, A., Love, J., Ronald, P. C. and Whalen, M. C. (1999). Developmental control of *Xa21*-mediated disease resistance in rice. Plant J. *20(2)*, 231–236.

Deshmukh, U. C., Verma, R. K., Saxena, R. R., Mohan, P. and Verulkar, S. B. (2017). Marker assisted selection for bacterial leaf blight resistance in segregating populations of

Karma Mahsuri'.Vegetos. *30(1).* doi: 10.5958/2229-4473.2017.00010.6.

Dilla-Ermita, C. J., Tandayu, E., Juanillas, V. M., Detras, J., Lozada, D. N., Dwiyanti, M. S., Vera Cruz, C., Mbanjo, E. G. N., Ardales, E., Diaz, M. G., Mendioro, M., Thomson, M. J. and Kretzschmar, T. (2017). Genome-wide association analysis tracks bacterial leaf blight resistance loci in rice diverse germplasm. Rice. *10(1).* doi: 10.1186/s12284-017-0147-4

Drummond, A. J., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Duran, C., Field, M., Heled, J., Kearse, M., Markowitz, S., Moir, R., Stones-Havas, S., Sturrock, S., Thierer, T. and Wilson, A. (2011). Geneious v5.4. doi: http://www.geneious.com.

Garris, A. J., McCouch, S. R. and Kresovich, S. (2003). Population structure and its effect on haplotype diversity and linkage disequilibrium surrounding the *xa5* locus of rice (*Oryza sativa* L.). Genetics, *165(2)*, 759-769.

Garris, A. J., Tai, T. H., Coburn, J., Kresovich, S. and Mccouch, S. (2005). Genetic structure and diversity in *Oryza sativa* L . Genetics.1638*(March)*, 1631–1638.

Genomics, P. and Lee, G. (2011). Single nucleotide polymorphisms and haplotype diversity in rice sucrose. J. Hered. *102(6)*, 735–746.

Gnanamanickam, S. S. (1999). An overview of bacterial blight disease of rice and srategies for its manangement. Curr. Sci. *77(11)*,1135–1144.

Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J. and Lopez, R. (2010). A new bioinformatics analysis tools framework at EMBL-EBI. Nucleic Acids Res. *38*: W695-9. doi: 10.1093/nar/gkq313

Huang, C. L., Hung, C. Y., Chiang, Y. C., Hwang, C. C., Hsu, T. W., Huang, C. C., Hung, K. H., Tsai, K. C., Wang, K. H., Osada, N., Schaal, B. A. and Chiang, T. Y. (2012). Footprints of natural and artificial selection for photoperiod pathway genes in *Oryza*, Plant J.*70(5)*, 769–782.

Huang, N., Angeles, E. R., Domingo, J., Magpantay, G., Singh, S., Zhang, G., Kumaravadivel, N., Bennett, J. and Khush, G. S. (1997). Pyramiding of bacterial blight resistance genes in rice: marker-assisted selection using RFLP and PCR. Theor Appl Genet. *95(3)*, 313–320.

Ikeda, R., Khush, G. and Tabien, R. (1990). A new resistance gene breed, to bacterial blight derived from *O. longistaminata*. Jpn J Breed. *(40)*, 280–281.

International Rice Genome Sequencing Project (2005). The map-based sequence of the rice genome. Nature. 436,793.

Jackson, S. A. (2016). Rice: The First Crop Genome. Rice. *9*. doi: 10.1186/s12284-016-0087-4.

de Jong, M. A., Wahlberg, N., van Eijk, M., Brakefield, P. M. and Zwaan, B. J. (2011). Mitochondrial DNA signature for range-wide populations of *Bicyclus anynana* suggests a rapid expansion from recent refugia. PLoS One. *6(6)*. doi: 10.1371/journal.pone.0021385.

Khush, G., Bacalangco, E. and Ogawa, T. (1991). A new gene for resistance to bacterial blight from *O. longistaminata*. Rice Genet. News Lett. *7*, 121–122.

Khush, G. S., Mackill, D. J. and Sidhu, G. S. (1989). Breeding rice for resistance to bacterial blight. Bacterial blight of rice. IRRI, Manila, Philippines, 207–217.

Li, J. B., Sun, Y. D., Liu, H., Wang, Y. Y., Jia, Y. L. and Xu, M. H. (2015). Natural variation of rice blast resistance gene *Pi-d2*. Genet. Mol. Res. *14*, 1235–1249.

Li, J. Y., Wang, J. and Zeigler, R. S. (2014). The 3,000 rice genomes project: new opportunities and challenges for future rice research. GigaScience. *3(1)*,8.

Lu, L., Yan, W., Xue, W., Shao, D. and Xing, Y. (2012). Evolution and association analysis of *Ghd7* in rice. PLoS One. *7(5)*. doi: 10.1371/journal.pone.003402.

Mansueto, L., Fuentes, R. R., Chebotarov, D., Borja, F. N., Detras, J., Abriol-Santos, J. M., Palis, K., Poliakov, A., Dubchak, I., Solovyev, V., Hamilton, R. S., McNally, K. L., Alexandrov, N. and Mauleon, R. (2016). SNP-Seek II: A resource for allele mining and analysis of big genomic data in *Oryza sativa*. Current Plant Biology. *6628(16).* doi: 10.1016/j.cpb.2016.12.003.

Mccouch, S. R., Wright, M. H., Tung, C., Maron, L. G., Mcnally, K. L., Fitzgerald, M., Singh, N., Declerck, G., Agosto-perez, F., Korniliev, P., Greenberg, A. J., Naredo, M. E. B., Mercado, S. M. Q., Harrington, S. E., Shi, Y., Leung, H., Ebana, K., Yano, M., Eizenga, G., Branchini, D. A., Kuser-falca, P. R., Mcclung, A. and Mezey, J. (2016). Open access resources for genome-wide association mapping in rice. Nat Commun. *7*, 10532.

Mew, T. W. (1993). Focus on bacterial blight of rice. Plant Dis.*77(1)*, 5.

Monaco, M. K., Stein, J., Naithani, S., Wei, S., Dharmawardhana, P., Kumari, S., Amarasinghe, V., Youens-Clark, K., Thomason, J., Preece, J., Pasternak, S., Olson, A., Jiao, Y., Lu, Z., Bolser, D., Kerhornou, A., Staines, D., Walts, B., Wu, G., D'Eustachio, P., Haw, R., Croft, D., Kersey, P. J., Stein, L., Jaiswal, P. and Ware, D. (2014). Gramene 2013: Comparative plant genomics resources. Nucleic Acids Res. *42(D1)*, 1–7.

Ogiso-Tanaka, E., Matsubara, K., Yamamoto, S., Nonoue, Y., Wu, J., Fujisawa, H., Ishikubo, H., Tanaka, T., Ando, T., Matsumoto, T. and Yano, M. (2013). Natural variation of the *Rice flowering locus T1* contributes to flowering time divergence in rice. PLOS One. *8(10)*.

Okonechnikov, K., Golosova, O., Fursov, M., Varlamov, A., Vaskin, Y., Efremov, I., German Grehov, O. G., Kandrov, D., Rasputin, K., Syabro, M. and Tleukenov, T. (2012). Unipro UGENE: A unified bioinformatics toolkit. Bioinformatics. *28(8)*, 1166–1167.

Park, C. J. and Ronald, P. C. (2012). Cleavage and nuclear localization of the rice XA21 immune receptor. Nat Commun. 3*(May)*, 1–6.

Peng, H., Chen, Z., Fang, Z., Zhou, J., Xia, Z. and Gao, L. (2015). Rice *Xa21* primed genes and pathways that are critical for combating bacterial blight infection. Sci. Rep. *5*, 20165.

Pruitt, R. N., Schwessinger, B., Joe, A., Thomas, N., Liu, F., Albert, M., Robinson, M. R.,

Chan, L. J. G., Luu, D. D., Chen, H., Bahar, O., Daudi, A., De Vleesschauwer, D., Caddell, D., Zhang, W., Zhao, X., Li, X., Heazlewood, J. L., Ruan, D., Majumder, D., Chern, M., Kalbacher, H., Midha, S., Patil, P. B., Sonti, R. V, Petzold, C. J., Liu, C. C., Brodbelt, J. S., Felix, G. and Ronald, P. C. (2015). The rice immune receptor XA21 recognizes a tyrosine-sulfated protein from a Gram-negative bacterium. Sci. Adv. *1(6)*, 1–13.

Rafalski, A. (2002). Applications of single nucleotide polymorphisms in crop genetics. Curr. Opin. Plant. Biol. *5(2)*, 94–100.

Ronald, P. C., Albano, B., Tabien, R., Abenes, L., Wu, K. S., McCouch, S. and Tanksley, S. D. (1992). Genetic and physical analysis of the rice bacterial blight disease resistance locus, *Xa21*. Mol Gen Genet. *236(1)*,113–120.

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E. and Sánchez-Gracia, A. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol. Biol. Evol. *34(12)*, 3299–3302.

Snyder, M. W., Adey, A., Kitzman, J. O. and Shendure, J. (2015). Haplotype-resolved genome sequencing : experimental methods and applications. Net. Rev. Genet. *16(6)*, 344-358.

Song, W.Y., Wang, G.L., Chen, L.L., Kim, H.S., Pi, L.Y., Holsten, T., Gardner, J., Wang, B., Zhai, W.X. and Zhu, L.H. (1995). A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. Science. *270(5243)*,1804–1806.

Tello-Ruiz, M. K., Naithani, S., Stein, J. C., Gupta, P., Campbell, M., Olson, A., Wei, S., Preece, J., Geniza, M. J., Jiao, Y., Lee, Y. K., Wang, B., Mulvaney, J., Chougule, K., Elser, J., Al-Bader, N., Kumari, S., Thomason, J., Kumar, V., Bolser, D. M., Naamati, G., Tapanari, E., Fonseca, N., Huerta, L., Iqbal, H., Keays, M., Munoz-Pomer Fuentes, A., Tang, A., Fabregat, A., D'Eustachio, P., Weiser, J., Stein, L. D., Petryszak, R., Papatheodorou, I., Kersey, P. J., Lockhart, P., Taylor, C., Jaiswal, P. and Ware, D. (2018). Gramene 2018: Unifying comparative genomics and pathway resources for plant research. Nucleic Acids Res. *46(D1)*, 1181–1189.

Thomas, N. C., Oksenberg, N., Liu, F., Caddell, D., Nalyvayko, A. and Nguyen, Y. (2018). The rice XA21 ectodomain fused to the Arabidopsis EFR cytoplasmic domain confers resistance to *Xanthomonas oryzae* pv. *oryzae*. PeerJ. *6*, e4456.

Wang, G.L., Song, W.-Y., Ruan, D.L., Sideris, S. and Ronald, P. C. (1996). The cloned gene, *Xa21*, confers resistance to multiple *Xanthomonas oryzae* pv. *oryzae* isolates in transgenic plants. *9(9)*, 850-855.

Wei, T., Chen, T., Ho, Y. T. and Ronald, P. C. (2016). Mutation of the rice XA21 predicted nuclear localization sequence does not affect resistance to *Xanthomonas oryzae* pv. *oryzae*. PeerJ. *4*, e2507.

Wei, X., Qiao, W., Yuan, N., Chen, Y., Wang, R., Cao, L., Zhang, W., Yang, Q. and Zeng, H. (2014). Domestication and association analysis of *Hd1* in Chinese mini-core collections of rice. Genet Resour. Crop Evol. *61(1)*, 121–142.

Yu, A. P., Wang, X. M., Yuan, X. P., Wang, C. H., Xu, Q., Feng, Y., Yu, H. Y., Wang.

(2016). Sequence variations and haplotypes of the bacterial blight resistance gene *xa13* in rice. J. Plant Pathol. *98(1)*, 167–169.

Zhang, F., Wu, Z. C., Wang, M. M., Dingkuhn, M., Xu, J. L., Zhou, Y. L. and Li, Z. K. (2017). Genome-wide association analysis identifies resistance loci for bacterial blight in a diverse collection of indica rice germplasm. PLoS One. *12(3)*, e0174598.

Zhao, K., Wright, M., Kimball, J., Eizenga, G., McClung, A., Kovach, M., Tyagi, W., Ali, M. L., Tung, C. W., Reynolds, A., Bustamante, C. D. and McCouch, S. R. (2010). Genomic diversity and introgression in *O. sativa* reveal the impact of domestication and breeding on the rice genome. PLoS One. *5(5)*, e10780.