Rakesh Bhowmick
1) ICAR-Vivekananda Parvatiya Krishi Anusandhan Sansthan, Almora, Uttarakhand, India
2) ICAR-National Research Centre on Plant Biotechnology, New Delhi, India

Soham Choudhury
ICAR-National Research Centre on Plant Biotechnology, New Delhi, India

ARNS Subbanna
ICAR-Vivekananda Parvatiya Krishi Anusandhan Sansthan, Almora, Uttarakhand, India

Suman Roy
ICAR-Central Research Institute for Jute and Allied Fibres, Barrackpore, West Bengal, India

Laxmi Sharma
ICAR-Central Research Institute for Jute and Allied Fibres, Barrackpore, West Bengal, India

Genome wide identification of genic and non-genic Microsatellites in *Nilaparvata lugens* Stål

Rakesh Bhowmick, Soham Choudhury, ARNS Subbanna, Suman Roy and Laxmi Sharma

Abstract
The brown plant hopper (BPH), *Nilaparvata lugens* Stål (Homopterous: Delphacidae) is one of the lethal pests in rice. In the present study, genome wide distributed Simple Sequence Repeats (SSR) were identified and analysed. The screening of 1.2 Gb of BPH genome resulted in identification of 271785 SSRs. Out of these, 3231 SSRs were from genic region. Microsatellite distribution in genome was found to be 238.2 SSRs per Mb. Altogether the microsatellite motifs contribute only 7.45Mb of BPH genome. Trinucleotide repeats were most abundant representing 50.8% of total genomic microsatellites. Longer repeat motifs were found to be least abundant in the genome representing 1.1% and 0.5% for pentanucleotide and hexanucleotide, respectively. SSR markers identified from this study can be useful for diversity analysis and genetic map development of BPH.

Keywords: brown plant hopper, microsatellite, simple sequence repeats

Introduction

The brown planthopper (BPH), *Nilaparvata lugens* Stål (Homoptera: Delphacidae), is one of the major insect pests of rice (*Oryza sativa*). Widespread damage of BPH cause significant yield reduction in rice. Moreover, higher dose of nitrogenous fertilizer application and indiscriminate use of pesticide often lead to outbreak of BPH (1). More than 50 million hectare of rice field has been affected by this insect during the past decade in Asia (2-4).

BPH has 30 chromosomes and genome size is estimated to be 1.2 Gb (5). Several attempts have been made to develop microsatellite markers in BPH, however no genome wide comprehensive analysis of microsatellite has been reported till date. More than 350 expressed sequence tag (EST)-derived simple sequence repeat (e-SSR) markers have been developed after creation of large EST database (6). Four hundred and seventy four simple sequence repeat markers were also developed and used to generate linkage map of BPH (7). In addition, 136 SSRs were isolated from BPH genomic DNA and were used to assess genetic diversity of BPH population (8).The availability of draft genome sequence of BPH (9) has been a useful resource in the present study to comprehensively analyse distribution of SSRs in BPH genome. Here, we have identified genome wide distributed genic and non-genic SSRs in BPH genome. Detailed analysis and frequency distribution of different microsatellite motifs identified in this study were made to visualize composition of microsatellite motifs in genome. Further, SSR markers developed in this study can be used to map virulence genes and other traits in BPH.

Material Methods
Data Retrieval
Draft sequence assembly of *Nilaparvata lugens* was downloaded from NCBI database (https://www.ncbi.nlm.nih.gov), under accession number AOSB00000000 (BioProject PRJNA177647). Draft assembly of BPH genome has total of 1,140,803,929 nucleotides and 46,559 scaffolds. For identification of genic microsatellite in BPH genome, CDS sequences were retrieved from NCBI database. As the sequence file was already filtered for low quality sequences, it was directly used for further analysis.

SSR identification
For identification of microsatellite motifs in the genome, the perl script MISA (downloaded from http://pgrc.ipk-gatersleben.de/misa) and software package GMATA (10) were used.

Correspondence
Rakesh Bhowmick
1) ICAR-Vivekananda Parvatiya Krishi Anusandhan Sansthan, Almora, Uttarakhand, India
2) ICAR-National Research Centre on Plant Biotechnology, New Delhi, India

In both the programs, microsatellite motif length of 2 to 6 nucleotides were considered. The minimum number of repeat for each microsatellite was defined as six for di-, and five for tri-, tetra-, penta- and hexa-nucleotides. Compound SSRs were defined when two SSRs in the same sequence scaffold was interrupted by less than 100 nucleotide bases. For designing the primer flanking microsatellite motifs, primer3 software integrated within GMATA package was used. Optimal parameter for primer designing was set as, 60ºC for annealing temperature and amplicon length from 120 to 400bp.

Result and Discussion
Genomic SSRs
Perfect SSRs with ≥5 repeats and a minimum length of 12bp were mined out from 1.14 Gb draft genome as well as predicted CDS of *Nilaparvata lugens*. It was found that 12,578 out of 46,559 scaffolds and 2,371 out of 25,643 CDS contain SSRs. A total of 271,785 SSRs were identified in BPH genome. On an average 238.2 SSRs were found in single Mb of genome and one SSR was found in 4.1 Kb of genome. Overall the microsatellite repeats together contribute to 7.45 Mb of genome in BPH.

Among the five different types of microsatellite studied, di (106,225) and tri (138,120) nucleotide repeats were found to be most abundant in BPH genome, representing 39% and 50.8% of total SSRs, respectively (Table-1). On the other hand, hexa-nucleotide repeats were found to be least abundant, representing only 0.51% of total microsatellite (Figure 1). From the overall analysis of SSR motifs, it was observed that occurrence of SSR decreases with increase in length of the repeat motif. A total of 918 different microsatellite variants were found in BPH genome (Figure 2). A thorough analysis was performed for each type of SSRs detected in BPH genome. The results indicate a significant difference in occurrence of repeat motifs in genome. A total of 918 different microsatellite motifs were detected. Twelve different types of dinucleotide repeats were found. Among these dinucleotide repeats, TG, AC, AG, CT, TC and GA motifs were most abundant representing 72.9% of total dinucleotide repeats. On the other hand, GC and CG was least abundant dinucleotide repeat in BPH genome, representing 0.12% and 0.1% respectively. Analysis of different trinucleotide repeats showed that there are sixty different trinucleotide repeats present in BPH genome. Repeats of AGA (11.3%) were found to be most predominant type of repeats, followed by TTC (10.1%), TCT (9.57%), CTT

(9.18%), AAG (8.9%), and GAA (8.4%). Together these six trinucleotide repeats account for 57.7% of total trinucleotide repeats. Rest fifty four trinucleotide repeats contribute 42.3% of total trinucleotide repeats (figure 3).

Further, we also found 176 tetranucleotide repeats in BPH genome. Among them, the motif TCAA had the highest frequency, representing 10% of the sampled sequences. Interestingly out of 176 tetranucleotide repeat variants, 22 repeats contribute 81% of total tetranucleotide repeats.

Though the abundance of longer repeats of pentanucleotide and hexanucleotide were least abundant throughout the genome, representing 1.1% and 0.5% respectively. 324 diverse pentanucleotide and 346 diverse hexanucleotide repeat variants were found in BPH genome. GGTTA (4%) and TTCTCC (6.5%) were most abundant pentanucleotide and hexanucletide repeat, respectively. In addition, most of the penta- and hexa-nucleotide repeat variants were found only once in genome. Altogether, twenty most abundant microsatellite motifs BPH genome in decreasing order were AGA, TG, AC, TTC, TCT, AG, CTT, CT, AAG, GAA, TC, GA, TA, AT, CA, GT,CTC, GAG, TAT and AAT. It was also interesting to note that, out of 918 different SSR motif variants identified in genome, 336 SSRs were found only once in genome. Further analysis of the frequencies of SSR motifs clearly indicated AT-rich motifs were far more abundant than AT = GC and GC-rich motifs. A total of AT-rich, AT = GC, and GC-rich motifs represent 60.7%, 18.9%, and 20.2% of the total SSRs, respectively. Flanking primer pairs were designed for PCR amplification of the microsatellite. A total of 252368 primer pairs were designed to amplify the identified microsatellites. Details of primer sequence, annealing temperature and amplicon length is provided in supplementary file.

Genic SSRs
A total of 3,231 SSRs were identified after screening 25,643 gene sequences. The most frequent repeats found within the gene sequences of BPH were trinucleotide repeats (82.2 %) followed by dinucleotide (16%). Longer repeat motifs (tetranucleotide, pentanucleotide, hexanucleotide) together contributing only 1.73% of total genic SSRs. The observed frequencies of different repeats in genic region of BPH are mentioned in table 2. Taken together all the microsatellite repeat motifs in gene, top ten most abundant motifs are CAG, GAG, GAT, CAA, AAC, GGA, AAG, AGG, ACA and AT, in decreasing order.

Table 1: Distribution of perfect microsatellite in BPH genome

SSR type	Count	Relative frequency (%)	Mean repeat number	Density/Mb	Cumulative Sequence length (kb)
Dinucleotide	106225	39.08	11.26	93.1	2393
Trinucleotide	138120	50.81	9.65	121	3999
Tetranucleotide	22872	1.16	7.39	20	676
Pentanucleotide	3169	1.16	13.45	2.7	213
Hexanucleotide	1399	0.51	9.5	1.2	172

Table 2: Description of genic SSRs in BPH genome

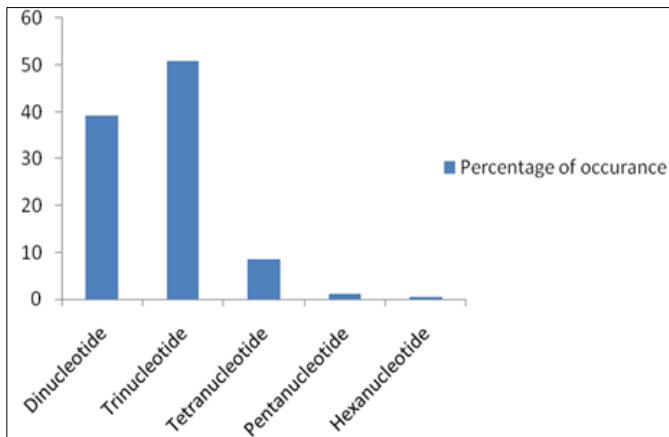| Content	Dinucleotide	Trinucleotide	Tetranucleotide	Pentanucleotide	Hexanucleotide
No. of motifs	519	2656	35	11	10
Percentage of total motifs	16	82.2	1.08	0.34	0.3
Motif variants	12	58	17	8	8
Most abundant motifs	AT,GA	AAC, AAG	AGCA,TATT	AATTG ATAAT	ACCACG, CAGCAC
Least abundant motifs	CG, GC	TTC, TTG	TCAT,TCAT	GCGGT GTGCT	TCATCT, TCTTCA

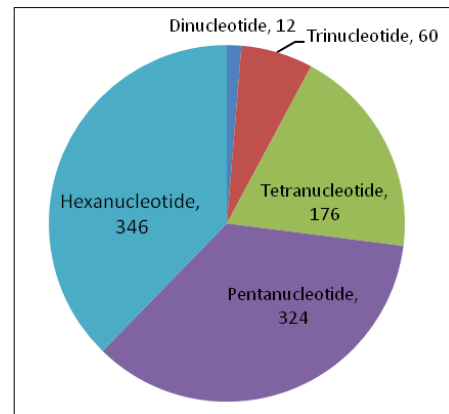**Fig 1:** Distribution with respect to the motif repeats of di to hexa-nucleotide microsatellites

Fig 2: Distribution of microsatellite motif length variants. Pie diagram represents motif length of 918 different microsatellite variants
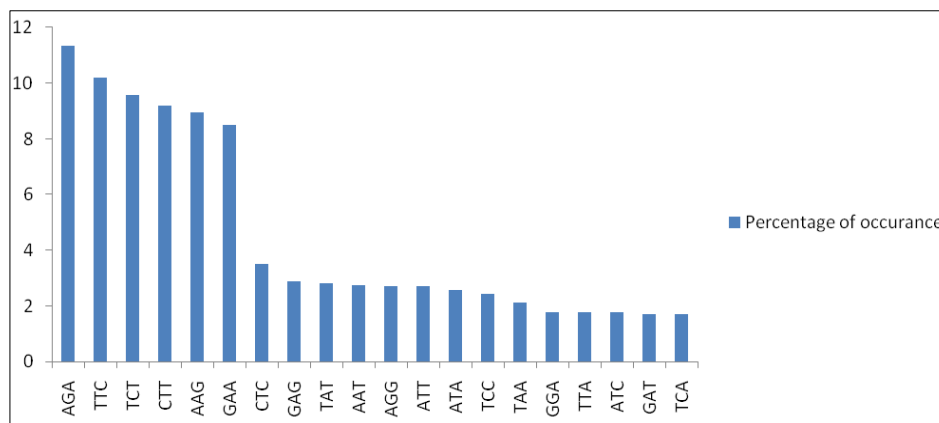

Fig 3: Distribution of trinucleotide repeat motifs in genome. Y axis represents percentage of occurrence in genome

Conclusions
Advancement of genome sequencing technology enabled us to develop novel molecular markers. In this study, genome-wide identification of microsatellites was conducted based on the assembled genomic sequences of brown plant hopper. Comprehensive analysis of BPH genome revealed that 7.4 Mb of
BPH genome is contributed by microsatellites. Trinucleotide repeats were most predominant than any other type of motifs. A high density of microsatellites was identified BPH, which in turn may be useful for analysis of genetic diversity of this species.

References
1. Bottrell DG, Schoenly KG. Resurrecting the ghost of green revolutions past: the brown planthopper as a recurring threat to high-yielding rice production in tropical Asia. J Asia-Pac Entomol. 2012; 15(1):122-140.
2. Catindig JLA, Arida GS, Baehaki SE, *et al*. In: Heong, K.L. and Hardy, B. (eds), Plant hoppers: New Threats to the Sustainability of Intensive Rice Production Systems in Asia. IRRI: Los Banos, Philippines, 2009, 191-220.
3. Heong KL, Are plant hopper problems caused bya breakdown in ecosystem services? In: Heong, K.L. and Hardy, B. (eds), Plant hoppers: New Threats to the Sustainability of Intensive Rice Production Systems in Asia. IRRI: Los Banos, Philippines, 2009, 221-31.
4. Gallagher KD, Kenmore PE, Sogawa K. Judicial use of insecticides deter plant hopper out breaks and extend the life of resistant varieties in South west Asian rice. In: Denno, R.F. and Perfect, T.J. (eds), Planthoppers: Their Ecology and Management. Chapman and Hall: New York, 1994, 599-614.
5. Wang Y, Fan HW, Huang HJ, Xue J, Wu WJ, Bao YY, *et al*. Chitin synthase 1 gene and its two alternative splicing variants from two sap-sucking insects, *Nilaparvata lugens* and *Laodelphax striatellus* (Hemiptera: Delphacidae). Insect Biochem. Mol. Biol. 2012; 42(9):637-646.
6. Jing S, Liu B, Peng L, Peng X, Zhu L, Fu Q, He G. Development and use of EST-SSR markers for assessing genetic diversity in the brown planthopper (*Nilaparvata lugens* Stål). Bull. Ent. Res. 2012; 102(1):113-22.
7. Jairin JI, Kobayashi TE, Yamagata YO, Sanada-Morimura SA, Mori KA, Tashiro KO, *et al*. A simple sequence repeat-and single-nucleotide polymorphism-based genetic linkage map of the brown plant hopper, *Nilaparvata lugens*. DNA Res. 2012; 20(1):17-30.
8. Jing S, Zhou X, Yu H, Liu B, Zhang C, Wang S, *et al*. Isolation and characterization of microsatellite markers in brown planthopper (*Nilaparvata lugens* Stål). Int. J Mol. Sci. 2012; 13(8):9527-9533.
9. Xue J, Zhou X, Zhang CX, Yu LL, Fan HW, Wang Z, *et al*. Genomes of the rice pest brown plant hopper and its endosymbionts reveal complex complementary contributions for host adaptation. Genom. Biol. 2014; 15(12):521.
10. Wang X, Wang L. GMATA: An Integrated Software Package for Genome-Scale SSR Mining, Marker Development and Viewing. Front Plant Sci. 2016; 7:1350.