Genetic Diversity and Population structure analysis of *Puccinia striiformis* f. sp. *tritici* in Yunnan, China

Md. Ashraful Alam[1, 2,*] and Mingju Li [1]

[1]Yunnan Key Laboratory of Green Prevention and Control of Agricultural Transboundary Pests, Agricultural Environment and Resource Research Institute, Yunnan Academy of Agricultural Sciences, Kunming 650205, China; [2] Bangladesh Agricultural Research Council (BARC), Farmgate, Dhaka 1215, Bangladesh

*Correspondence: e-mail: ashrafulw@yahoo.com; limingju1996@hotmail.com

Abstract

The isolates of wheat stripe rust disease were collected in the year of 2004, 2012, 2014 and 2015 during the main wheat growing season February to May from ten counties of Yunnan province in China. The results from single nucleotide polymorphisms (SNPs) of 168 isolates showed that two housekeeping genes namely Citrate Synthase (CS) and heat shock protein 90kDa (HSP) were found to contain a total of 11 SNP sites. Thirty-four haplotypes were inferred from the concatenated sequences, with three haplotypes (H23, H32 and H16) comprising of over 40.5% of population and were shared haplotypes the population of Dehong, Yuxi, Lincang, Qujing and other counties. The haplotype diversity, nucleotide diversity, mutation rate and recombination events were 0.930, 4.46E-03, 4.46E-03 and 4 respectively, which revealed virulent diversity of *Pst* populations among all locations. All the three grouping methods (UPGMA-tree, PCA and Structure) used in the current study to classify the *Pst* populations according to their races and geographical locations, were found meaningful and most of the groups were co-linear in all the grouping methods. Using analysis of molecular variance (AMOVA), we found 6.13% of the total variation was among populations and 86.63% within populations. This revealed a relatively high genetic diversity came from inside population but low genetic divergence came from among populations. Moreover, the molecular data on gene flow (*Nm*=6.09) confirmed the migration of pathogen populations occurred among all locations in Yunnan Province. The ancestral haplotype (H25) detected in Zhaotong, Yuxi, Lincang, Chuxiong and Kunming and the other results also indicated that the frequent pathogen exchange within the locations. We analyzed the population structure in different locations of Yunnan intensively. Based on trajectories of upper airflow and genetic diversity of *Pst* populations in different locations, we suggested that Lincang, Dehong and Dali are probably source of *Pst* in Yunnan.

Introduction

Around the globe, stripe rust disease of wheat is considered the most devastative disease, which is caused by fungus *Puccinia striiformis* f. sp. *tritici (Pst)*. In terms of the area that can be affected by the disease, China is the largest epidemic region for stripe rust disease of wheat in the world. The Yunnan Province is one of the wheat-producing Provinces in China, which is situated in the south-western part of the country. From west to east the Yunnan-Guizhou Plateau of China is crossed to this Province. Single nucleotide polymorphisms (SNPs) are relatively different types of molecular markers and AFLP, RAPD (Random Amplification of Polymorphic DNA) and SSR (Simple Sequence Repeats) are replacing by SNPs with the development of sequencing technology. Polymorphisms can be detected in coding and non-coding regions by SNPs marker of an organism that can cover a large part of genome. A SNP is the mutation of a single base pair at a specific locus position and SNPs can conserve during evolution. In recent years, SNP markers have been used to study of population structure in plant pathogens. The current study aimed to investigate the genomic assortment and population structure of *Pst* in Yunnan Province from the year 2004, 2012, 2014 and 2015 through using housekeeping genes SNP primers.

Materials and Methods

Sampling and multiplication

The samples of wheat stripe rust were collected in the year of 2004, 2012, 2014 and 2015 during the main wheat growing season February to May from ten counties of Yunnan province in China. A total of 168 isolates were used in this study.

Primer design

The sequences of *Pst* housekeeping genes were searched in Gen-Bank. Two protein-coding housekeeping genes were identified for developing SNP primers, namely Citrate Synthase (CS) and heat shock protein 90kDa (HSP). The SNP primers were designed using Premier 5.0 software (https://en.freedownloadmanager.org/users-choice/Primer_Premier_5_64.html).

DNA extraction

DNA was extracted directly from urediniospores using a modified cetyltrimethylammonium bromide (CTAB) procedure.

PCR and sequencing

PCR was carried out in 30µl volume and all primers amplified under the same thermal cycling conditions and reagent concentrations except the annealing temperatures. Purification and sequencing was carried out at Tsingke Biological Technology Co. Kunming, China

Analysis of the recorded data

Recorded data were analyzed by the multi-evolutionary analysis software. The arrangements were aligned and split for a single gene and all of the samples by using MEGA 4.0. The haplotypes, counting the records of SNP loci and the category, collapse sequence, ancestral and isolate number of haplotypes, were examined by Map tariff options for collapsing sequences and removing indels into haplotypes and eliminating infinite-sites desecrations, using SNAP Workbench 2.0. The diversity of haplotype (*Hd*) and nucleotide (*Pi*), neutrality tests (Tajima's *D* and Fu's *Fs* values), re-amalgamation incident (*Rm*), coefficient of genetic differentiation (*Gst*) based on haplotypes and gene flow (*Nm*) was calculated in DnaSP v.5.10. The mutation rate of the populations (Θ) was computed in MEGA 4.0. Analysis of molecular variance (AMOVA) was carried out using Arlequin 3.1. To assess the degree of isolates concentration changes (heat maps) and PCA clustering were performed using the Metabo-Analyst 2.0 software. The phylogenetic-tree was constructed using MEGA 5.0 software and illustrated by FigTree v1.4.2. The STRUCTURE 2.3.4 was used for inferring population structure.

Results and Discussion

Genetic diversity in the Yunnan Pst isolates

A total of 34 haplotypes were detected from 168 samples using two SNP primers collected from 10 counties of Yunnan province in the year of 2004, 2012. 2014 and 2015. There were 11 SNP loci, collected samples from all locations, where 8 were phylogenetically informative. By using 2 primers, a total of 77 polymorphic alleles were found across all populations. Haplotypes H23, H32 and H16 had the maximum incidence among the haplotypes, which was added up to 40.5% and were shared haplotypes the population of Dehong, Yuxi, Lincang, Qujing and other counties (Figure 1). Among them, H23 was comparatively widespread haplotypes and shared eight counties. The outcomes of the diversity of haplotype designated that the maximum *Hd* value was in Wenshan, 0.967 and the lower most was in Lijiang, 0.811. The diversity of nucleotides (*Pi*) fluctuated from 3.05E-03 to 5.59E-03 in the diversified populations. The mutation rate was the highest in the Zaothong population (5.14E-03) and lowest in Yuxi (3.11E-03) population. The recombination tests revealed that Wenshan had the maximum recombination, with *Rm* = 4, and lowest in Dali, Qujing and Lincang *Rm*=1. The overall Tajima's *D* was positive and not significant (*D* = 0.47601, *P* =0.15636) indicating low levels with low-frequency polymorphisms within locations. The Fu's Fs, was highly significant and negative (*Fs* = -25.654, *P* = 0.0000) indicating an excess number of alleles, as would be expected from a recent population expansion or from genetic hitchhiking. The coefficient of genetic differentiation (*Gst*) among all populations of Yunnan was 0.03941 while it was 0.03308, 0.01054, 0.02503, 0.00896, 0.01963, 0.001245, 0.00230, 0.00646 and 0.00083 between Kunming and Lijiang (LJ), Dehong (DH), Dali (DL), Qujing (QJ), Zhaotong (ZT), Yuxi (YX), Lincang (LC), Wenshan (WS), Chuxiong (CX) indicating a low differentiation among the nine counties except Lijiang and Dali. The *Gst* was low among all populations, indicating lower heterogeneity. The results of AMOVA signposted that modification largely originated from within populations, accounted for 86.63%, while accounted for 6.13% among populations within assemblies and accounted for 7.24% among clusters.

The dendrogram (Figure 2) was prepared from the genetic variation matrix derivatives from 11 SNP loci for 34 haplotypes. In UPGMA (unweighted pair-group method using arithmetic averages) dendrogram, the haplotypes assembled into three groups; however, one of them contained only one haplotype (H 31). The other 2 groups were characterized as key groups comprising more than 5 haplotypes. Groups 1 contained 24 haplotypes with 144 isolates, where 86.11% isolates from Lijiang (19), Zhaotong (12), Yuxi (24), Lincang (12), Wenshan (13), Chuxiong (19) and Kunming (25). Group 2 contained 9 haplotypes with 23 isolates, where 69.6% isolates from Dali (5), Zhaotong (5), Wenshan (3) and Chuxiong (3).

Principal component analysis (PCA) assists as a podium to deliver a three-dimensional graphical image of the proportional genetic detachments within the populations. It also measures the strength of the diversity between the groups categorized by a dendrogram. The haplotypes grouped by PCA were carefully arranged with a UPGMA-based tree. In PCA scatterplots, the first two principal components explained 38.5 and 25.1 % (Figure 3) of the entire dissimilarity, respectively. In agreement with the UPGMA-tree, haplotypes were obviously detached by PC1 and found 3 distinct groups however group 1 and group 2 were very close.

Population structure of the Yunnan Pst isolates

For population genetic structure analysis, Bayesian clustering modelling was performed in the STRUCTURE software using 34 haplotypes, where data were generated by SNP marker. As the clustering model assumes the fundamental reality of K clusters, and Evano test was done and generated K = 2 as the maximum log-likelihood (Figure 4). In STRUCTURE-software analysis, concurrences were further characterized as unadulterated or admixture, concurrences with >0.80 score were measured as pure and <0.80 as an admixture. The population I comprised 44.12% of haplotypes (15 haplotypes), where 11 haplotypes were pure and 4 were admixed. The total of 64 isolates consisted in population I, 11 from Kunming, 10 isolates from Lijiang, 9 from Lincang, 9 from Wenshan, 7 from Wenshan and 7 from Zhaotong, which covered 83% of the population I isolate. Population II comprised of total of 19 haplotypes with 104 isolates, out of those 1 haplotypes were found admixed. Out of 104 isolates, 18 isolates from Yuxi, 17 isolates from Kunming, 15 isolates from Chuxiong, 10 isolates from Zhaotong, 10 isolates from Lijiang and 9 isolates from Dali, covered 76% of population II isolates.

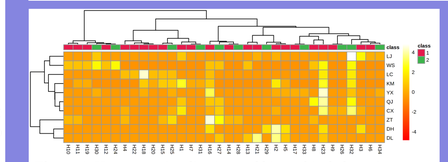

Figure 1. Heat-map visualization and hierarchical clustering analysis with Metabo-Analyst's data annotation tools constructed based on the different haplotypes for 10 locations. Rows: locations; Columns: haplotypes; Color key indicates haplotypes value, red: Lowest, white: highest.

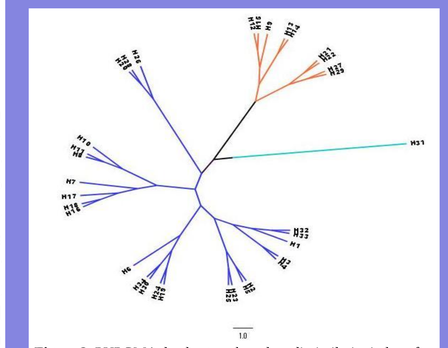

Figure 2. PUPGMA dendrogram based on dissimilarity index of 11 SNP loci for 34 haplotypes.

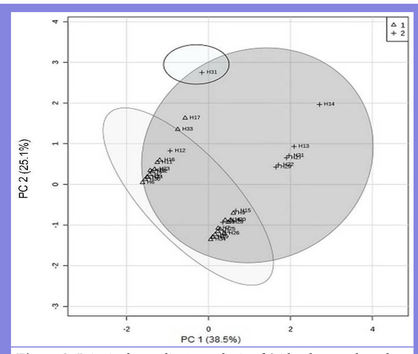

Figure 3. Principal coordinate analysis of 34 haplotypes based on SNP markers' data

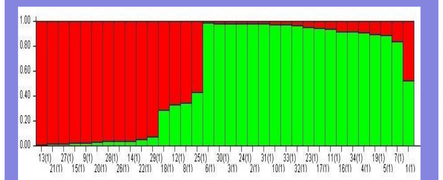

Figure 4. Model-based population structure plot for each isolates with K=2, using Structure with SNP markers data. Note: Colour codes are as follows: Population I red, Population II green, population III blue

Conclusion

In the study, results from SNPs of 168 segregates exhibited that two housekeeping genes were established to comprise a total of 11 SNP positions A total of 34 haplotypes were inferred from the concatenated sequences, with three haplotypes (H23, H32 and H16) comprising of over 40.5% of population and were shared haplotypes the population of Dehong, Yuxi, Lincang, Qujing and other counties. The haplotype diversity, nucleotide diversity, mutation rate and recombination events were 0.930, 4.46E-03, 4.46E-03 and 4 respectively, which revealed virulent diversity of *Pst* populations among all locations. All the three grouping methods (UPGMA-tree, PCA and Structure) used in the current study to classify the *Pst* populations according to their races and geographical locations, were found meaningful and most of the groups were co-linear in all the grouping methods. By using AMOVA, the study recognized about 6.13% of the total dissimilarity and 86.63% within populations. The findings of the study also exposed a comparatively the maximum hereditary assortment came from inside the population, but lower genetic discrepancy came from among populations. Furthermore, the genomic data on gene flow (*Nm*=6.09) established that the movement of pathogens occurred among all locations in Yunnan Province. Based on trajectories of upper airflow and genetic diversity of *Pst* populations in different locations, we suggested that Lincang, Dehong and Dali are probably source of *Pst* in Yunnan.