Viruses Without Borders: Global Analysis of the Population Structure, Haplotype Distribution, and Evolutionary Pattern of Iris Yellow Spot Orthotospovirus (Family Tospoviridae, Genus Orthotospovirus)

Afsha Tabassum¹, S. V. Ramesh², Ying Zhai¹, Romana Iftikhar¹, Cristian Olaya¹ and Hanu R. Pappu*¹*

¹ Department of Plant Pathology, Washington State University, Pullman, WA, United States, ² Indian Council of Agricultural Research-Central Plantation Crops Research Institute, Kasaragod, India

Iris yellow spot, caused by Iris yellow spot orthotospovirus (IYSV) (Genus: Orthotospovirus, Family: Tospoviridae), is an important disease of Allium spp. The complete N gene sequences of 142 IYSV isolates of curated sequence data from GenBank were used to determine the genetic diversity and evolutionary pattern. In silico restriction fragment length polymorphism (RFLP) analysis, codon-based maximum likelihood studies, genetic differentiation and gene flow within the populations of IYSV genotypes were investigated. Bayesian phylogenetic analysis was carried out to estimate the evolutionary rate. In silico RFLP analysis of N gene sequences categorized IYSV isolates into two major genotypes viz., IYSV Netherlands (IYSVNL; 55.63%), IYSV Brazil (IYSVB; 38.73%) and the rest fell in neither group [IYSV other (IYSVother; 5.63%)]. Phylogenetic tree largely corroborated the results of RFLP analysis and the IYSV genotypes clustered into IYSVNL and IYSVB genotypes. Genetic diversity test revealed IYSVother to be more diverse than IYSVNL and IYSVB. IYSVNL and IYSVB genotypes are under purifying selection and population expansion, whereas IYSVother showed decreasing population size and hence appear to be under balancing selection. IYSVB is least differentiated from IYSVother compared to IYSVNL genotype based on nucleotide diversity. Three putative recombinant events were found in the N gene of IYSV isolates based on RDP analysis, however, RAT substantiated two among them. The marginal likelihood mean substitution rate was 5.08 × 10⁻⁵ subs/site/year and 95% highest posterior density (HPD) substitution rate between 5.11 × 10⁻⁵ and 5.06 × 10⁻⁵. Findings suggest that IYSV continues to evolve using population expansion strategies. The substitution rates identified are similar to other plant RNA viruses.

Keywords: BEAST, evolutionary genomics, gene flow, genetic differentiation, genetic recombination, iris yellow spot orthotospovirus, in silico RFLP, phylogenetics

Abbreviations: IYSV, Iris yellow spot orthotospovirus; N gene, Nucleocapsid gene; RFLP, Restriction Fragment Length Polymorphism; RDP, Recombination Detection Program; RAT, Recombination Analysis Tool; BEAST, Bayesian Evolutionary Analysis by Sampling Trees.
INTRODUCTION

Tospoviruses continue to be a major production constraint for a wide range of agronomic and horticultural crops worldwide (Gent et al., 2006; Pappu et al., 2009; Mandal et al., 2012; Mandal et al., 2012; Bag et al., 2015; Oliver and Whitfield, 2016; Turina et al., 2016; Resende et al., 2020). Iris yellow spot orthotospovirus (IYSV; genus: Orthotospovirus, family: Tospoviridae) (Resende et al., 2020) primarily infect Allium spp., which includes onion (Allium cepa), green onion (Allium fistulosum), garlic (Allium tuberosum), leek (Allium porrum) (Gent et al., 2006; Cordoba-Selles et al., 2007; Bag et al., 2015; Karavina et al., 2016; Tabassum et al., 2016. The virus was first described in southern Brazil in 1981 on infected onion (Allium cepa; family: Amaryllidaceae) inflorescence stalks (scapes). The disease was referred to as “Sapeca.” In the US, the disease was first described in the Treasure Valley of southwestern Idaho and southeastern Oregon in 1989 (Gent et al., 2006). In 2003, the disease epidemic in Colorado (United States) caused a crop loss estimated at US $ 2.5–5 million (Gent et al., 2006). The disease has spread to most of the onion-growing areas in Africa, Asia, Europe, the Americas, and the Oceanía (Centre for Agriculture and Bioscience International (CABI) - Invasive Species Compendium, 2019).

The disease caused by IYSV is characterized by chlorotic or necrotic, straw-colored to white, dry, elongated or spindle shaped lesions along the scape (Figure 1A). Lesions are frequently in middle to lower portions of the scape. The diamond-shaped lesions tend to be less defined on leaves (Pappu et al., 2008; Bag et al., 2015). The photosynthetic activity is affected in the infected plants leading to reduced bulb size. As the disease progresses, the lesions girdle the scape causing the seed head to collapse leading to severe crop losses (Figure 1B; Gent et al., 2006).

IYSV, as other tospoviruses, consists of a tripartite genome: Small (S) and Medium (M) RNAs encode proteins in both sense and antisense orientations (ambisense) while the Large (L) RNA encodes protein from negative sense strand. The L RNA codes for RNA dependent RNA polymerase (RdRp), M RNA codes for glycoprotein precursors (G\textsubscript{N} and G\textsubscript{C}) and the non-structural movement protein (NSm), and the S RNA codes for nucleocapsid (N) and non-structural silencing suppressor protein (NSs) (reviewed in Bag et al., 2015; Pappu et al., 2020; Resende et al., 2020).

Genetic evolution of viruses directly impacts the host-virus interactions and as such is important to ascertain genetic diversity within a viral species (Sacristan and Garcia-Arenal, 2008; Gibbs and Ohshima, 2010). Genetic drift, migration, mutation, natural selection, segment reassortment, and recombination are the major sources of evolutionary changes in the genetic architecture of viral populations (Moya et al., 2004; Butkovic et al., 2021). Phylo-geographical analysis is a powerful tool to determine the geographical distribution pattern of virus, assessing their genetic variation, and historical events that are shaping the genetic architecture of the viral populations (Hewitt, 2004; Chen et al., 2012). Comprehensive genetic architecture and evolutionary genomic analysis of viral populations have become a subject of increasing attention in a number of viruses.

While IYSV is widely distributed in the world, the complete genome of only a few isolates are sequenced. Since the N gene is considered as one of the descriptors for tospovirus identification and classification, the N gene of a large number of isolates was sequenced and the genetic diversity was determined (de Avila et al., 1993; Pappu et al., 2006; Nischwitz et al., 2007; Iftikhar et al., 2014; Bag et al., 2015). The number of N gene sequences of IYSV isolates reported since the last study (Iftikhar et al., 2014) has been on the rise. Building on the earlier findings, we carried out a more detailed and a global analysis of the extent of genetic recombination, genetic diversity, genetic differentiation, and gene flow among different genotypes of IYSV isolates reported from different parts of the world. Further, Bayesian model-based coalescent approaches were used to gain insights into the molecular evolutionary pattern of IYSV population.

MATERIALS AND METHODS

Data Source of Nucleocapsid (N) Gene Sequences

Complete nucleocapsid (N) gene sequences of 142 IYSV isolates reported from across the globe were obtained the nucleotide sequence repository, GenBank. IYSV isolates analyzed were from 19 countries spread over six continents—Africa, Asia, Australia and New Zealand, Europe, North America (Canada, Mexico and the United States) and South America, infecting 10 different hosts including Allium cepa (the most commonly reported host), Allium porrum, Eustoma russellianum, Allium tuberosum, Allium chinense, Wild onion, Alstroemeria sp., Allium sativum, and Allium fistulosum. The N gene sequence (HQ267713) derived from tomato spotted wilt orthotospovirus (TSWV) infecting pepper crop in South Korea was used as an outgroup (Supplementary Table 1). Only complete IYSV N gene sequences...
(822 nt-long open reading frame coding for a 273-amino acid protein) were considered for analysis.

**In silico Restriction Fragment Length Polymorphism (RFLP) Analysis**

N gene sequences were analyzed for sequence variations by performing *in silico* RFLP analysis using Restriction Mapper (Restriction Mapper, 2009). The complete N gene sequence was virtually digested, and sites were mapped as recognized by restriction enzyme *Hinfl* (Zen et al., 2005). Based on *Hinfl* digestion, IYSV isolates can be grouped into IYSV Netherlands (IYSV<sub>NL</sub>) or IYSV Brazil (IYSV<sub>BR</sub>) types. The size of the largest fragment generated by digestion is considered for differentiating the given isolate into two groups. The genotypes *viz.*, IYSV<sub>NL</sub> and IYSV<sub>BR</sub> are differentiated based on the resultant 308 and 468 bp fragments, respectively. Those isolates that yielded any different fragment size upon restriction digestion were grouped into IYSV<sub>other</sub>.

**Phylogeny Construction**

Multiple sequence alignment (MSA) was performed using MUSCLE algorithm available in MEGA7 (Edgar, 2004). Best-fit model of nucleotide substitution was determined using MODEL TEST in MEGA7. Aligned sequence relatedness was evaluated using the Maximum Likelihood (default parameters with 1,000 bootstrap replicates) method based on Tamura parameter 3 model (T92) with Gamma distributed (G) available in MEGA7 (Kumar et al., 2016). The phylogenetic tree was rooted using TSWV N gene reported from South Korea as an outgroup.

**Population Selection Studies and Neutrality Test**

Mean rates of non-synonymous (dN) and synonymous substitutions (dS) were calculated using codon-based maximum likelihood methods, i.e., SLAC (single like ancestor counting), FEL (fixed effects likelihood), and REL (random effects likelihood). DATAMONKEY server (Weaver et al., 2018) was used to calculate dN/dS ratio. To test the theory of neutral evolution, the test statistics such as Tajima’s D (Tajima, 1989), Fu and Li’s D, and Fu and Li’s F (Fu and Li, 1993; Fu, 1997) were computed in DnaSP software.

**Genetic Differentiation and Gene Flow Estimates**

DnaSP was used to compute nucleotide test statistics such as Ks, Kst (Kst value close to zero indicate no differentiation), Snn (Snn value close to one indicates differentiation) (Hudson, 2000) and haplotype statistics Hs, Hst (Hudson et al., 1992a,b). These tests estimate genetic differentiation within the populations of IYSV genotypes. Fst statistics was used to estimate the extent of the gene flow (panmixia or free gene flow has values close to zero whereas infrequent gene flow attains values close to one) (Hudson et al., 1992b).

**Recombination Detection Analysis (RDA)**

Unaligned sequences were loaded in SDT v1.2 program, pairwise scan was performed with the MUSCLE, and the sequence data was saved with minimum identity of 70% and maximum of 100% to ensure sequences were properly aligned. The aligned IYSV N sequences were then used as an input query and analyzed for recombination events using Recombination Detection Program (RDP) v 4.0 (Martin and Rybicki, 2000), BOOTSCAN (Salminen et al., 1995), 3SEQ, GENECONV (Sawyer, 1999), MAXCHI (Maynard, 1992), CHIMAERA (Posada and Crandall, 2001) and SISCAN (Gibbs et al., 2000) available in RDP 4 Beta 4.88. Default settings for the different recombination detection methods and a Bonferroni corrected P-value cut-off of 0.05 were used for analysis.

**Recombination Analysis Tool (RAT)**

Recombination analysis tool (RAT) was used for the analysis of aligned nucleotide sequences (Etherington et al., 2005). RAT algorithm uses pairwise comparisons between sequences based on the distance method to identify recombinants in nucleotide sequence alignment. Percentage of nucleotide similarities were compared using a sliding window size of 10% of the sequence length and an increment size being half of the window size.

**Bayesian Evolutionary Analysis by Sampling Trees (BEAST)**

Bayesian phylogenetic analysis was performed in BEAST v2.4.6 (Bouckaert et al., 2014) to estimate evolutionary rate. Strict, relaxed (exponential, lognormal) and random local clocks were utilized for comparison (Bouckaert et al., 2014). Demographic models—coalescent constant population, coalescent exponential population, coalescent Bayesian skyline and coalescent extended Bayesian skyline were used to infer demographic history. “Temporal signal” (i.e., genetic changes between sampling times are sufficient and there is statistical relationship between genetic divergence and time) in the dataset was assessed using TempEst program (Rambaut et al., 2016). Using Markov Chain Monte Carlo (MCMC) method Bayesian phylogenies were constructed in BEAST v2.4.6. First 10% of the samples were discarded as burn-in. Convergence of the chain to stationary distribution and adequate sampling were analyzed using Tracer v1.6 (Tracer, 2018). Tracer was used to analyze the Effective Sample Size (ESS) and other prior parameter values. Tree Annotator was used for generating Maximum Clade Credibility (MCC) phylogenetic trees with common heights node. FigTree (2018) was used to generate the dendrograms.

**RESULTS**

**In silico Restriction Fragment Length Polymorphism Analysis**

Computational RFLP-based analysis of N gene sequences recognizing *Hinfl* restriction site divided the population into two major groups [79 NL (55.63%), 55 BR (38.73%)]. Thus, the genotype IYSV<sub>NL</sub> was found to be predominant...
over IYSV\textsubscript{BR} while the rest (5.63\%) fell in neither category (IYSV\textsubscript{other}) (Figure 2).

**Molecular Phylogeny of IYSV N Gene**

Phylogenetic tree of the N gene of IYSV constructed based on the aligned nucleotide sequences (Figure 3) using Maximum Likelihood method broadly clustered IYSV genotypes into two major clades (NL and BR types). Four IYSV\textsubscript{other} isolates (EU750697, KT27882, KT272883, and EU586203) and one IYSV\textsubscript{BR} isolate (FJ713700) clustered with the NL group. Similarly, four more IYSV\textsubscript{other} isolates (HQ148173, HQ148174, EU310290, and EU310291) and two IYSV\textsubscript{NL} isolates (FJ785835 and AF001387) clustered with the BR group. One IYSV\textsubscript{BR} isolate from Brazil (AF067070) formed a separate monophyletic clade along with a TSWV N gene isolate HQ267713 from South Korea (Outgroup). The clades also followed a geographical pattern as majority of IYSV\textsubscript{NL} genotypes are from North America and IYSV\textsubscript{BR} are from the Asian countries. Only one IYSV isolate has been reported from Brazil which formed a separate monophyletic clade even though in silico RFLP characterized it as an IYSV\textsubscript{BR} type.

**Population Selection Studies, Neutrality Test, and Genetic Diversity Test**

Gene codons that are in positive or negative selection pressure provide knowledge regarding the molecular evolution pattern of the N gene. The mean dN/dS (dN—rate of non-synonymous substitutions and dS—rate of synonymous substitutions) for N gene accessions belonging to IYSV\textsubscript{NL} group were found to be 0.192 with no positively selected codon site. SLAC methodology identified 20 negatively selected codons in the N gene of IYSV\textsubscript{NL} type. The data set when analyzed by FEL methodology revealed...
one positively selected codon site (codon no. 139) against 62 negatively selected codons. The dN-dS (mean difference between dN and dS) was –0.803 based on REL analysis denoting that the codon sites are under purifying selection acting against deleterious non-synonymous substitutions (Table 1).

For N gene accessions derived from IYSV
BR, the mean dN/dS was found to be 0.191 with no positively selected codon site. SLAC methodology identified 27 negatively selected codons. The data set when analyzed by FEL revealed one positively selected codon site (codon no. 139) against 63 negatively selected codons. The dN-dS was –0.813 based on REL analysis suggesting that the codon sites are under purifying selection acting against deleterious non-synonymous substitutions. Six positively selected codon (codon nos. 30, 40, 109, 139, 210, and 225) and seven deleterious non-synonymous substitutions. Six positively selected codon (codon no. 270) and zero negatively selected codons were identified by REL analysis, respectively (Table 1).

For N gene accessions of IYSV
other group, the mean dN/dS was found to be 0.172 with no positively selected codon site. SLAC methodology identified three negatively selected codons. FEL methodology revealed no positively selected codon site against 38 negatively selected codons. The dN-dS was –0.805 based on REL analysis and it denotes codon sites are under purifying selection acting against deleterious non-synonymous substitutions. One positively selected codon (codon no. 270) and zero negatively selected codons were identified by REL analysis, respectively (Table 1).

Nucleotide diversity (\(\pi\)) of IYSV
BR was about two folds higher than that for IYSV
NL (0.04513 and 0.01990; respectively, Table 2).

However, the highest nucleotide diversity among the IYSV isolates was found in IYSV
other (0.08042) indicating IYSV
other is more genetically diverse than the IYSV
NL and IYSV
BR (Table 2). Number of polymorphic sites (S) was 136 from the N gene sequences of eight isolates of IYSV
other, IYSV
NL showed 189 polymorphic sites among the N gene sequences obtained from the 79 isolates, whereas IYSV
BR showed 230 polymorphic sites obtained from the 55 isolates (Table 2). IYSV
other is more diverse than IYSV NL and BR based on number of polymorphic sites (S).

### Neutrality Test

The test of neutral evolution analyzed based on the total number of mutations and segregating sites, revealed statistically significant and non-significant negative values of test statistic Tajimas’s D for IYSV
NL and IYSV
BR, respectively (Tables 3, 4). It indicates the operation of purifying selection and population expansion in major IYSV genotypes (NL and BR). Similarly, negative values of other test statistics such as Fu and Li’s D and Fu and Li’s F also corroborate the above findings with regard to IYSV
BR and IYSV
NL genotypes. However, positive values of all the test statistics such as Tajima’s D, Fu and Li’s D, and Fu and Li’s F with respect to the genotype IYSV
other indicate the decrease in population size and act of balancing selection.

### TABLE 1 | Codon substitution in the nucleocapsid gene of Iris yellow spot orthotospovirus genotypes.

| Genotype      | Positively selected codon positions | No. of negatively selected codons | \(\omega\) | dN-dS |
|---------------|------------------------------------|----------------------------------|--------|------|
| IYSV
NL         | 139\(^a\)                           | 20\(^b\)                         | 0.192444 | –0.803 |
|              |                                    | 62\(^b\)                         |         |      |
|              |                                    | 73\(^b\)                         |         |      |
| IYSV
BR         | 139\(^a\)                           | 27\(^a\)                         | 0.191119 | –0.813 |
|              |                                    | 30\(^a\)                         |         |      |
|              |                                    | 63\(^b\)                         |         |      |
|              |                                    | 40\(^b\)                         |         |      |
|              |                                    | 109\(^c\)                        |         |      |
|              |                                    | 210\(^c\)                        |         |      |
|              |                                    | 225\(^c\)                        |         |      |
| IYSV
other     | 270\(^d\)                           | 3\(^a\)                          | 0.172079 | –0.805 |
|              |                                    | 36\(^b\)                         |         |      |
| IYSV
NL         | 109\(^d\)                           | 54\(^a\)                         | 0.205279 |      |
|              |                                    | 90\(^b\)                         |         |      |

\(dN\), the number of non-synonymous substitutions per non-synonymous site; \(dS\), the number of synonymous substitutions per synonymous site \(\omega\)—Ratio of dN/dS from SLAC (single like ancestor counting) methodology, dN-dS obtained from REL (random effects likelihood).

\(^a\) Codons identified by SLAC at a cut-off p-value 0.1.

\(^b\) Codons identified by FEL at a cut-off p-value 0.1.

\(^c\) Codons identified by REL at a cut-off Bayes factor value 50.

\(^d\) Codons identified by SLAC at a cut-off Bayes factor value 50.

\(\omega\) = dN/dS, IYSV
NL, IYSV
BR and IYSV
other.

### TABLE 2 | Genetic diversity test of Iris yellow spot orthotospovirus genotypes.

| Genotype       | N | S   | \(\pi\) | Hd |
|----------------|---|-----|--------|----|
| IYSV
NL         | 79 | 189 | 0.01990 | 0.982 |
| IYSV
BR         | 55 | 230 | 0.04513 | 0.999 |
| IYSV
other     | 08 | 136 | 0.08042 | 0.964 |
| IYSV
All        | 142| 317 | 0.07220 | 0.994 |

\(N\), Number of isolates; \(S\), Number of polymorphic (segregating) sites; \(\pi\), nucleotide diversity within species; IYSV
NL = IYSV
BR, IYSV
other.

### TABLE 3 | Neutrality test of Iris yellow spot orthotospovirus genotypes based on total number of mutations.

| Genotypes | Tajima’s D | Fu and Li’s D | Fu and Li’s F |
|-----------|------------|---------------|---------------|
| IYSV
NL     | –2.14425\(^*\) | –1.43877 | –2.07919 |
| IYSV
BR     | –1.24392 | –1.92209 | –1.98807 |
| IYSV
other | 1.08486 | 0.71725 | 0.89745 |

Calculated using total number of mutations. \(^*\) statistically significant at \(P < 0.01\).

### TABLE 4 | Neutrality tests of Iris yellow spot orthotospovirus genotypes based on total number of segregating sites.

| Genotypes | Tajima’s D | Fu and Li’s D | Fu and Li’s F |
|-----------|------------|---------------|---------------|
| IYSV NL    | –1.96208\(^*\) | –1.83392 | –2.26501 |
| IYSV BR    | –0.93734 | –2.14943 | –2.01338 |
| IYSV Other | 1.42119 | 0.69566 | 0.96703 |

Calculated using total number of segregating sites. \(^*\) statistically significant at \(P < 0.01\).
Genetic Differentiation and Gene Flow

Haplotype-based statistics (Hs and Hst) and nucleotide-based statistics (Ks, Kst, Snn) were estimated to evaluate genetic differentiation between the IYSV genotypes (Table 5). The statistically significant test values of Ks, Kst and Z reveal strong genetic differentiation among the IYSV genotypes studied. Snn value close to one indicates genetic differentiation even though insignificant test statistical values were obtained. IYSV_BR is more differentiated from IYSV_other (Kst value of 0.03724*) compared to IYSV_NL genotypes (0.21377*) based on the Kst values. Fst values show that the extent of gene flow between major genotypes, IYSV_BR and IYSV_NL, is relatively high than the gene flow between individual BR and NL genotypes with IYSV_other. Among the major genotypes, IYSV_NL shows greater gene flow with IYSV_other than IYSV_BR.

Recombination Detection Analysis

Three potential recombination events were detected among the IYSV N genes analyzed (Table 6 and Figure 4). AF067070 (BR type) IYSV isolate from Brazil is a potential recombinant of isolates: JX861126 Bosnia (major parent) and AY345226 Australia (minor parent). This recombinant was detected by GeneConv, 3Seq algorithms in RDP. The recombination breakpoint begins at 789 in alignment (789 without gaps) with breakpoint clustering at 99% confidence interval ranging from 730 to 3 in alignment (731–3 without gaps) and breakpoint ends at 820 in alignment (820 without gaps) with breakpoint clustering at 99% confidence interval ranging from 731 to 3 in alignment (731–3 without gaps) when the major and minor parent isolates intersect at two points in the graph (Figure 5).

Recombination Analysis Tool

Recombination analysis tool (RAT) was used to substantiate the findings of the RDP. An isolate was considered recombinant when the major and minor parent isolates intersect at two points in the graph (Figure 5). Based on this criterion, HQ148174_Iran and AB180918_Japan were considered potential recombinants.
even though only MaxChi in the RDP4 program detected HQ148174_Iran as a recombinant.

**Bayesian Evolutionary Analysis by Sampling Trees (BEAST)**

The rates of nucleotide substitution in tospovirus genomes have not been reported. Therefore, the global repository of IYSV N gene sequences was used to estimate the rates of nucleotide substitution and discern the rapidity with which molecular evolution might occur in the tospoviruses. Genetic recombinants were removed for BEAST analysis since their inclusion violates the assumption of coalescent-based analyses and thus could result in incorrect estimates of the rate of evolution. For nucleotide models, Hasegawa-Kishino-Yano (HKY)-based analysis was performed and it converged satisfactorily. While comparing two models if the marginal posterior distributions of the log-likelihoods do not overlap then the model with the higher posterior distribution of log-likelihood was preferred. Estimate is a better approximation of the true posterior distribution when larger Effective Sample Size (ESS) is available (ESS > 200 are desirable). Based on the above criteria, General Time Reversible (GTR) relaxed exponential growth clock model with coalescent
constant population was found to be the best fit with a marginal likelihood mean substitution rate of $5.08 \times 10^{-5}$ subs/site/year, $95\%$ highest posterior density (HPD) substitution rate between $5.11 \times 10^{-5}$ and $5.06 \times 10^{-5}$ and ESS was 305 (Supplementary Table 2). Bayesian phylogenetic tree separated the IYSV isolates into two distinct clades, clade I comprising of IYSV$_{BR}$ isolates and clade II comprising IYSV$_{NL}$ isolates. The isolates that belonged to the same geographic region (or same country) clustered together (Figure 6).

DISCUSSION

The global population structure and temporal dynamics of the IYSV conducted previously (Iftikhar et al., 2014) based on N gene sequences delineated that the viral isolates could be categorized into two major genotypes (IYSV$_{BR}$ and IYSV$_{NL}$). Further, temporal dynamics of IYSV showed greater incidence of IYSV$_{BR}$ post-2005 compared to IYSV$_{NL}$. Since the last publication, the number of N gene sequences added to the public repository has increased significantly. To gain a better understanding of the evolutionary genomics and to further gain deeper insights into the evolution rate of IYSV, we analyzed 142 complete N gene sequences using a wide range of computational tools to infer molecular evolutionary genomics.

In silico RFLP analysis to categorize the genotype of IYSV isolates showed that the majority of the isolates belonged to IYSV$_{NL}$ category (55.63%), whereas 38.73% of IYSV$_{BR}$ isolates were observed. There was an increment in IYSV$_{NL}$ genotype incidence or characterization compared to IYSV$_{BR}$ since the last report (Iftikhar et al., 2014). Interestingly, gene flow estimates showed greater gene flow between NL and BR genotypes, rather than between the individual major genotypes and the “other” category. Even between the major genotypes, IYSV$_{NL}$ exhibited a greater gene flow with IYSV$_{other}$. Also, a greater genetic diversity was observed in IYSV$_{other}$, compared to NL and BR. However, codon substitution analysis of N gene showed little change since the last study (Iftikhar et al., 2014). In fact, the positively selected
codons (codon positions 139 in BR and NL and 270 in IYSV other) remained intact despite the substantial increase in the number of isolates examined, suggesting the importance of these codon positions in improving the fitness of nucleocapsid protein. The negative selections in the other codon positions imply that the deleterious mutations in those positions are effectively removed in the IYSV population as a whole. Most of the codons of the N gene are neither under positive nor negative selection suggesting the neutral evolution of these codons.

Recombination is a common phenomenon in RNA viruses but the implications of recombination for evolution is not well studied (Sztuba-Solinska et al., 2011). There is a serious limitation in understanding the contribution of recombination to evolution of IYSV due to lack of full-length genome sequences. The potential recombinants identified in this study belonged to BR type which seems to be evolving using population expansion strategies. The recombination breakpoints were at 5' and 3' ends suggesting that these are potential hot spots for recombination (Gawande et al., 2015).

In the BEAST analysis, General Time Reversible (GTR) relaxed exponential growth clock model with coalescent constant population was found to be the best fit model explaining the genetic architecture of IYSV population. In a similar analysis of PVY genomic sequences, it was found that the relaxed uncorrelated log normal clock was the best fit with a population of constant size (Gibbs et al., 2017). Further, similar topology of the phylogenetic tree was obtained by both ML method and Bayesian MCC based phylogeny for IYSV.

PVY dating was reported by comparing the estimated phylogenetic dates with historical events in the worldwide adoption of potato and other PVY hosts (Gibbs et al., 2017). While the potato-PVY analysis was based on the sample collection dates over several decades, onion-IYSV interactions are relatively new and hence predicting the phylodynamic patterns and demographic history of IYSV require more such data on temporal scale. The PVY demographic history and population expansion was deduced and compared with that of geographic distribution of host (potato) suggesting direct influence of potato cultivation area on the population size of the virus (Mao et al., 2019). In this context, further studies how expansion of onion cultivation area influences the population expansion of IYSV will be interesting.

Bayesian coalescent estimates of evolutionary dynamics of citrus tristeza virus, based on the p25 gene, showed that the rate of substitution was at $1.19 \times 10^{-3}$ subs/site/year (Benitez-Galeano et al., 2017). Similarly, Bayesian phylogenetic reconstruction-based nucleotide substitution rates of CP gene derived from four species of viruses in Secoviridae family estimated it to range $9.29 \times 10^{-3}$ to $2.74 \times 10^{-3}$ (subs/site/year) (Thompson et al., 2014) while for tobamovirus the estimate ranged from $1 \times 10^{-5}$ and $1.3 \times 10^{-3}$ substitutions per site, per year (Pagan et al., 2010). Further, Bayesian analysis of VPg gene of PVY reveals that it has been evolving at a rate of $5.60 \times 10^{-4}$ subs/site/year (Mao et al., 2019). Thus, the mean substitution rates identified for the IYSV N gene are comparable to those found in other plant-infecting RNA viruses. Substitution rates tend to be higher in RNA viruses as they are shown to mutate at faster rate. These mutations help in viral emergence on novel hosts but are not adaptive (Sacristan and Garcia-Arenal, 2008). Furthermore, the time of divergence of PVY clades, clade N and clade O, was found to be the year 1861.
CONCLUSION

IYSV_{NL} was found to be the predominant genotype on a global scale. Interestingly, the IYSV_{other} genotype is genetically more diverse than IYSV_{BR} and IYSV_{NL} genotypes. Population structure analysis revealed that it is under purifying selection and the phenomenon of population expansion is occurring. BEAST-based molecular clock analysis showed that the rates of molecular evolution of IYSV N gene are similar to other plant RNA viruses. This study is a step forward in identifying molecular factors that contribute to the evolution of IYSV, and serves as a foundation for further evolutionary genomic studies on one of the economically important plant virus groups.

REFERENCES

Bag, S., Schwartz, H. F., Cramer, C. S., Havey, M. J., and Pappu, H. R. (2015). Iris yellow spot virus (Tospovirus: Bunyaviridae): from obscurity to research priority. *Mol. Plant Pathol.* 16, 224–237. doi: 10.1111/mpp.12177

Benítez-Galeano, M. J., Castells, M., and Colina, R. (2017). The further evolutionary genomic studies on one of the economically important plant virus species, next generation sequencing-based sequencing followed by *de novo* assembly would provide a near complete genomic sequences that could be used to generate a more comprehensive picture of the genetic diversity of the virus populations (Zarghani et al., 2018).

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

AT and HP conceived and designed the experiments. AT, SR, YZ, and HP analyzed the data. AT, CO, SR, YZ, RI, and HP contributed reagents, materials, and analysis tools. AT, SR, and HP wrote the manuscript, proofread and finalized the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported in part by the Specialty Crop Block Grant Program from the Washington State Department of Agriculture (Grant # K2527), USDA National Institute of Food and Agriculture (NIFA), Specialty Crop Research Initiative (Grant No. 2018-5118128435), and USDA-NIFA Hatch project Accession #1016563 “Reducing the Impact of Pests and Diseases Affecting Washington Agriculture.” The funders have no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ACKNOWLEDGMENTS

AT would like to thank the Indian Council of Agricultural Research and Washington State University for their fellowship and financial support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.633710/full#supplementary-material
Annu. Rev. Phytopathol. 48, 205–223. doi: 10.1146/annurev-phyto-073009-114404

Gibbs, A., Ohshima, K., Yasaka, R., Mohammadi, M., Gibbs, M. J., and Jones, R. A. C. (2017). The phylogenetics of the global population of potato virus Y and its necrogenic recombinants. Virus Evol. 3: vex002. doi: 10.1093/ve/vex002

Gibbs, M. J., Armstrong, J. S., and Gibbs, A. J. (2000). Sister – scanning: a monte carlo procedure for assessing signals in recombinant sequences. Bioinformatics 16, 573–582. doi: 10.1093/bioinformatics/16.7.573

He, Z., Chen, W., Yasaka, R., Chen, C., and Chen, X. (2019). Temporal analysis and adaptive evolution of the global population of potato virus M. Infect. Genet. Evol. 73, 167–174. doi: 10.1016/j.meegid.2019.04.034

Hewitt, G. M. (2004). Genetic consequences of climatic oscillations in the quaternary. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 359, 183–195. doi: 10.1098/rstb.2003.1388

Hudson, R. R. (2000). A new statistic for detecting genetic differentiation. Genetics 155, 2011–2014.

Hudson, R. R., Boes, D. D., and Kaplan, N. L. (1992a). A statistical test for detecting recombination amongst DNA polymorphism. Genetics 132, 583–589.

Iftikhar, R., Ramesh, S. V., Bag, S., Ashfaq, M., and Pappu, H. R. (2014). Global analysis of population structure, spatial and temporal dynamics of genetic diversity, and evolutionary lineages of Iris yellow spot virus (Tospovirus:Bunyaviridae). Gene 547, 111–118. doi: 10.1016/j.gene.2014.06.036

Karcavica, C., Ibagà, J. D., Gubbá, A., and Pappu, H. R. (2016). First report of Iris yellow spot virus infecting garlic and leek in Zimbabwe. Plant Dis. 100:657. doi: 10.1094/PDIS-09-15-1022-PDN

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874. doi: 10.1093/molbev/msw054

Mandal, B., Jain, R. K., Krishnareddy, M., KrishnaKumar, N. K., Ravi, K. S., and Pappu, H. R. (2012). Emerging problems of Tospoviruses (Bunyaviridae) and their management in the Indian subcontinent. Plant Dis. 96, 468–479. doi: 10.1094/PDPS-06-11-0520

Mao, Y., Bai, Y., Shen, J., Gao, F., Sun, X., Qu, G., et al. (2019). Molecular evolutionary analysis of potato virus Y infecting potato based on the VPg gene. Front. Microbiol. 10:1708. doi: 10.3389/fmicb.2019.01708

Martin, D., and Rybicki, E. (2000). RDP: detection of recombination amongst aligned sequences. Bioinformatics 16, 562–563. doi: 10.1093/bioinformatics/16.6.562

Maynard, S. J. (1992). Analyzing the mosaic structure of genes. J. Mol. Evol. 34, 126–129. doi: 10.1007/BF00182389

Moya, A., Holmes, E. C., and Gonzalez-Candelas, F. (2004). The population genetics and evolutionary epidemiology of RNA viruses. Nat. Rev. Microbiol. 2, 279–288. doi: 10.1038/nrmicro8683

Nischwitz, C., Pappu, H. R., Mullis, S. W., Sparks, A. N., Langston, D., Cainos, A. S., et al. (2007). Phylogenetic analysis of Iris yellow spot virus isolates from onion (Allium cepa) in Georgia (USA) and Peru. J. Phytopathol. 155, 531–535.

Olver, J. E., and Whitfield, A. E. (2016). The genus Tospovirus: emerging Bunyaviruses that threaten food security. Annu. Rev. Virol. 3, 101–124. doi: 10.1146/annurev-virology-100114-055036

Pagan, L., Firth, C., and Holmes, E. C. (2010). Phylogenetic analysis reveals rapid evolutionary dynamics in the plant RNA virus genus Tobamovirus. J. Mol. Evol. 71, 298–307. doi: 10.1007/s00705-010-1985-4

Pappu, H. R., du Toit, L. J., Schwartz, H. F., and Mohan, K. (2006). Sequence diversity of the nucleoprotein gene of Iris yellow spot virus (genus Tospovirus family Bunyaviridae) isolates from the western region of the United States. Arch. Virol. 151, 1015–1023. doi: 10.1007/s00705-005-0681-2

Pappu, H. R., Jones, R. A. C., and Jain, R. K. (2009). Global status of tospovirus epidemics in diverse cropping systems: successes gained and challenges ahead. Virus Res. 141, 219–236. doi: 10.1016/j.virusres.2009.01.009

Pappu, H. R., Rosales, I. M., and Druffel, K. L. (2008). Serological and molecular assays for rapid and sensitive detection of Iris yellow spot virus infection of bulb and seed onion crops. Plant Dis. 92, 588–594. doi: 10.1094/pdis-92-4-0588

Pappu, H. R., Whitfield, A. E., and Oliveira, A. (2020). Tomato Spotted Wilt Virus. Encyclopedia of Virology, Reference Module in Life Sciences, 4th Edn, Cambridge, MA: Elsevier Press. doi: 10.1016/B978-0-12-809633-8.21329-0

Posada, D., and Crandall, K. A. (2001). Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proc. Natl. Acad. Sci. U.S.A. 98, 13757–13762. doi: 10.1073/pnas.241370698

Rambaut, A., Lam, T. T., Carvalho, L. M., and Pybus, O. G. (2016). Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol. 2:vew007. doi: 10.1093/viruses/vew007

Resende, R., Whitfield, A. E., and Pappu, H. R. (2020). “Orthotospoviruses (Tospoviridae),” in Encyclopedia of Virology, 4th Edn., eds D. H. Bamford and M. Zuckerman (Cambridge, MA: Academic Press), 507–515. doi: 10.1016/B978-0-12-809633-8.21337-X

Restriction Mapper (2009). Available online at: www.restrictionmapper.org (accessed March 29, 2021).

Sacristán, S., and Garcia-Arenal, F. (2008). The evolution of virulence and pathogenicity in plant. Mol. Plant Pathol. 9, 369–384. doi: 10.1111/j.1364-3703.2007.00460.x

Salminen, M. O., Carr, J. K., Burke, D. S., and McCutchan, F. E. (1995). Identification of breakpoints in intergenotypic recombinants of HIV type 1 by Bootscanning. AIDS Res. Hum. Retrovir. 11, 1423–1425. doi: 10.1007/aip.1995.11423

Sawyer, S. A. (1999). Geneconv: A Computer Package for the Statistical Detection of Gene Conversion. St. Louis, MI: Department of Mathematics, Washington University.

Sztuba-Solinska, J., Urbanowicz, A., Figlerowicz, M., and Bujarski, J. J. (2018). The role of gene conversion in the evolution of virulence and pathogenicity in plant. Frontiers in Microbiology | www.frontiersin.org 11 September 2021 | Volume 12 | Article 633710

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

Thompson, J. R., Kamath, N., and Perry, K. L. (2014). An evolutionary analysis of the Secoviridae family of viruses. PLoS One 10:e0119267. doi: 10.1371/journal.pone.0119267

Tracer (2018). Available online at: http://tree.bio.ed.ac.uk/software/tracer/ (accessed March 29, 2021).

Turina, M., Kormelink, R., and Resende, R. O. (2016). Resistance to Tospoviruses in vegetable crops: epidemiological and molecular aspects. Annu. Rev. Phytopathol. 54, 347–371. doi: 10.1146/annurev-phyto-080615-095843

Weisser, S., Shank, S. D., Spielman, S. J., Li, M., Muse, S. V., and Pond, S. L. K. (2018). Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. Mol. Biol. Evol. 35, 773–777. doi: 10.1093/molbev/msx335
Tabassum et al. IYSV Evolutionary Genomics

Zarghani, S. N., Hily, J. M., Glasa, M., Marais, A., Wetzel, T., Faure, C., et al. (2018). Grapevine virus T diversity as revealed by full-length genome sequences assembled through high-throughput sequence data. *PLoS One* 13:e0206010. doi: 10.1371/journal.pone.0206010

Zen, S., Okuda, M., Ebihara, K., Uematsu, S., Hanada, K., Iwanami, T., et al. (2005). Genetic differentiation of *Iris* yellow spot virus on onion (*Allium cepa*) and pathogenicity of two IYSV strains on onion and leaf onion (*A. schoenoprasum*). *Jpn. J. Phytopathol.* 71, 123–126. doi: 10.3186/jjphytopath.71.123

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Tabassum, Ramesh, Zhai, Iftikhar, Olaya and Pappu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.