Phylogenetic and population genetic analysis of *Thrips tabaci* Lindeman (Thysanoptera: Thripidae) on *Allium* host in India (#90576)

First submission

Guidance from your Editor

Please submit by **1 Jan 2024** for the benefit of the authors (and your token reward) .

Structure and Criteria
Please read the 'Structure and Criteria' page for general guidance.

Custom checks
Make sure you include the custom checks shown below, in your review.

Raw data check
Review the raw data.

Image check
Check that figures and images have not been inappropriately manipulated.

If this article is published your review will be made public. You can choose whether to sign your review. If uploading a PDF please remove any identifiable information (if you want to remain anonymous).

Files

Download and review all files from the materials page.

6 Figure file(s)
5 Table file(s)
19 Other file(s)

Custom checks

DNA data checks

- Have you checked the authors data deposition statement?
- Can you access the deposited data?
- Has the data been deposited correctly?
- Is the deposition information noted in the manuscript?

Structure and Criteria

Structure your review

The review form is divided into 5 sections. Please consider these when composing your review:

1. **BASIC REPORTING**
2. **EXPERIMENTAL DESIGN**
3. **VALIDITY OF THE FINDINGS**
4. General comments
5. Confidential notes to the editor

You can also annotate this PDF and upload it as part of your review

When ready submit online.

Editorial Criteria

Use these criteria points to structure your review. The full detailed editorial criteria is on your guidance page.

BASIC REPORTING

- Clear, unambiguous, professional English language used throughout.
- Intro & background to show context. Literature well referenced & relevant.
- Structure conforms to PeerJ standards, discipline norm, or improved for clarity.
- Figures are relevant, high quality, well labelled & described.
- Raw data supplied (see PeerJ policy).

EXPERIMENTAL DESIGN

- Original primary research within Scope of the journal.
- Research question well defined, relevant & meaningful. It is stated how the research fills an identified knowledge gap.
- Rigorous investigation performed to a high technical & ethical standard.
- Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

- Impact and novelty not assessed. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
- All underlying data have been provided; they are robust, statistically sound, & controlled.
- Conclusions are well stated, linked to original research question & limited to supporting results.

Standout reviewing tips

The best reviewers use these techniques

Tip	*Example*
Support criticisms with evidence from the text or from other sources	*Smith et al (J of Methodology, 2005, V3, pp 123) have shown that the analysis you use in Lines 241-250 is not the most appropriate for this situation. Please explain why you used this method.*
Give specific suggestions on how to improve the manuscript	*Your introduction needs more detail. I suggest that you improve the description at lines 57- 86 to provide more justification for your study (specifically, you should expand upon the knowledge gap being filled).*
Comment on language and grammar issues	*The English language should be improved to ensure that an international audience can clearly understand your text. Some examples where the language could be improved include lines 23, 77, 121, 128 – the current phrasing makes comprehension difficult. I suggest you have a colleague who is proficient in English and familiar with the subject matter review your manuscript, or contact a professional editing service.*
Organize by importance of the issues, and number your points	*1. Your most important issue* *2. The next most important item* *3. …* *4. The least important points*
Please provide constructive criticism, and avoid personal opinions	*I thank you for providing the raw data, however your supplemental files need more descriptive metadata identifiers to be useful to future readers. Although your results are compelling, the data analysis should be improved in the following ways: AA, BB, CC*
Comment on strengths (as well as weaknesses) of the manuscript	*I commend the authors for their extensive data set, compiled over many years of detailed fieldwork. In addition, the manuscript is clearly written in professional, unambiguous language. If there is a weakness, it is in the statistical analysis (as I have noted above) which should be improved upon before Acceptance.*

Phylogenetic and population genetic analysis of *Thrips tabaci* Lindeman (Thysanoptera: Thripidae) on *Allium* host in India

Tushar Gawai [1, 2], **Sharwari Sadawarte** [3], **Kiran Khandagale** [1], **Anusha R** [1], **Abhijeet Kulkarni** [3], **Durgesh Kumar Jaiswal** [1, 2], **Avinash Ade** [2], **Suresh Gawande** [Corresp. 1]

[1] ICAR-Directorate of Onion and Garlic Research, Pune, India

[2] Department of Botany, Savitribai Phule Pune University, Pune, India

[3] Department of Bioinformatics,, Savitribai Phule Pune University, Pune, India

Corresponding Author: Suresh Gawande
Email address: Suresh.Gawande@icar.gov.in

Background. Understanding population-level genetic variations can aid in explaining variance in insect vector populations' contributions to the spread of particular viruses. *Thrips tabaci* populations' varying propensity to transmit the viruses has been linked to the existence of cryptic species that display various ways of transmission ranges for reproduction and hosts. Thus, precise characterization of Thrips is foremost and fundamental step in effective disease management. **Methods.** This study aims to explore intra-species genetic diversity and molecular evolutionary relationships of mitochondrial cytochrome oxidase gene subunit I (mtCOI) in Indian *Thrips tabaci*. Total 48 sequences were retrieved for mtCOI gene from NCBI-Nucleotide database for this analysis. Amplicon sequencing of thrips mtCOI gene from eight diverse localities in India was performed. **Results.** The frequency and distribution of SNPs polymorphisms at different locations in India are studied using amplicon sequencing data of mtCOI genes. Numerous Insertions and deletions of varying lengths were found in the genomic positions of different localities. At the genome position 300-400 for each locality, highest variation was noticed. Molecular diversity analysis revealed that the population contains 30 haplotypes. Higher *N*m values indicate higher gene flow is observed between western: southern sub-populations, north-eastern: eastern and north-eastern: western sub-populations of group A and northern-southern subpopulations of group B. Fst values of northern-western subpopulation indicate high levels of genetic variability which can be observed in haplotype networks as well with majority of haplotypes coming from northern region (Delhi). Although most population shows stable (haplotypes from Maharashtra and Delhi) and long evolutionary history, subpopulations from northern western and north-eastern Indian thrips population indicates rapid growth.

Phylogenetic and population genetic analysis of *Thrips tabaci* Lindeman (Thysanoptera: Thripidae) on *Allium* host in India

Tushar B Gawai[1,2], Sharwari S Sadawarte[3], Kiran Khandagale[1], Anusha R, Abhijeet Kulkarni[3], Durgesh Kumar Jaiswal[2], Avinash B Ade[2], Suresh J Gawande[1]*

[1]Department of Plant Pathology, ICAR- Directorate of Onion and Garlic Research, Rajgurunagar, Pune-410505

[2]Department of Botany, Savitribai Phule Pune University, Pune-411007

[3]Department of Bioinformatics, Savitribai Phule Pune University, Pune-411007

Corresponding Author:
Suresh J. Gawande
[1]ICAR-Directorate of Onion and Garlic Research, Pune, India
Email address: suresh.gawande@icar.gov.in

Abstract

Background. Understanding population-level genetic variations can aid in explaining variance in insect vector populations' contributions to the spread of particular viruses. *Thrips tabaci* populations' varying propensity to transmit the viruses has been linked to the existence of cryptic species that display various ways of transmission ranges for reproduction and hosts. Thus, precise characterization of Thrips is foremost and fundamental step in effective disease management.

Methods. This study aims to explore intra-species genetic diversity and molecular evolutionary relationships of mitochondrial cytochrome oxidase gene subunit I (mtCOI) in Indian *Thrips tabaci*. Total 48 sequences were retrieved for mtCOI gene from NCBI-Nucleotide database for this analysis. Amplicon sequencing of thrips mtCOI gene from eight diverse localities in India was performed.

Results. The frequency and distribution of SNPs polymorphisms at different locations in India were studied using amplicon sequencing data of mtCOI genes. Numerous Insertions and deletions of varying lengths were found in the genomic positions of different localities. At the genome position 300-400 for each locality, highest variation was noticed. Molecular diversity analysis revealed that the population contains 30 haplotypes. Higher *N*m values indicate higher gene flow is observed between western: southern sub-populations, north-eastern: eastern and north-eastern: western sub-populations of group A and northern-southern subpopulations of group B. Fst values of northern-western subpopulation indicate high levels of genetic variability which can be observed in haplotype networks as well with majority of haplotypes coming from northern region (Delhi). Although most population shows stable (haplotypes from Maharashtra and Delhi) and long evolutionary history, subpopulations from northern western and north-eastern Indian thrips population indicates rapid growth.

Keywords: mtCOI, Amplicon sequencing, *Thrips tabaci*, population genetics, phylogenetic analysis

Introduction

Thrips is a widespread polyphagous insect pest from the order Thysanoptera. Several species of thrips cause serious damage to crops worldwide (Mouden et al., 2017; Stuart et al., 2011). Among them, the onion thrips, *Thrips tabaci* Lind., is an economically important pest that costs significant monetary losses to various crops each year (De Kogel et al., 2015). In addition to direct damage to the crop, this insect is also a vector of two devastating viruses, Iris Yellow Spot Virus (IYSV), and Tomato Spotted Wilt Virus (TSWV). *T. tabaci* is considered as a complex of genetically differentiated and widespread insect pests adapted to diverse climatic conditions (Brunner et al., 2004). The genetic variation within this species has a positive impact on their adaptability under diverse climatic conditions. This genetic variation is also linked to their

reproductive modes, such as thelytoky, arrhenotoky, and deuterotoky (Kobayashi and Hasegawa 2003; Nault et al., 2006; Gawande et al., 2017), as well as vector competence (Jacobson et al., 2013).

To study the genetic variation in thrips, various genes were targeted and analyzed for phylogenetic study and species discrimination (Buckman et al., 2013; Rebijith et al., 2013). DNA barcoding is a novel system of species identification using the cytochrome c oxidase subunit 1 mitochondrial gene (cox1 or COI) as a standardized single molecular marker for the identification, comparison and categorization for animal species (Hebert et al., 2003). One prerequisite of the DNA barcoding method is that species identification be linked to a voucher from a curated biological collection. This allows for follow-up and the development of a method for verifying species identification.

In Thysanopetra, targeting mitochondrial genome is most suitable approach for molecular identification as well as species discrimination irrespective of life stages, sex and polymorphism (Asokan et al., 2007), discriminating cryptic species (Glover et al., 2010; Rebijith et al., 2013), biotypes (Shufran et al., 2000), haplotypes and host and geographic associated genetic differences (Rebijith et al., 2013; Brunner et al., 2004). Three distinct lineages (T-Tobacco associated, L1 and L2-Leek associated) in *Thrips tabaci* based on mitochondrial data were proposed by (Brunner et al., 2004; Kobayashi et al., 2013; Jacobson et al., 2016). Using mtCOI sequences, microsatellite markers and vector competence in accordance with transmission efficiency of the viruses in *thrips* population was documented (Jacobson and Kennedy 2013). DNA barcoding using mitochondrial cytochrome c oxidase I gene partial fragments (mtCOI) for species-level identification has received widespread support as an additional technique to clear up taxonomic ambiguities. According to the International Barcode of Life consortium, the

genetic variety below the level of a species has also been estimated using mitochondrial genes. Its value as a quick and reliable instrument for species identification is widely acknowledged in a huge range of animal taxa around the world (International Barcode of Life Consortium 2023). This method has also been applied to thrips identification, phylogenetic analysis, population structure and invasive genetics (Tyagi et al., 2017).

In India, phylogenetic analyses based on the mtCOI gene were carried out, revealing different biotypes of thrips on different hosts and exhibiting differences in their mtCOI gene (Asokan et al., 2007). The existence of heteroplasmy identified in *T. tabaci* by analyzing mtCOI haplotypes from different geographic regions in India. An exhaustive analysis of genetic diversity and population genetic variation of *T. tabaci* in India is yet to be done. Additionally, its correlation to other countries needs to be studied. Up until now, no population genetics analysis on *T. tabaci* has been reported in India (Gawande et al., 2017). Though much work has been done on thrips genotyping using mitochondrial mtCOI markers, the extent of crypticness within *T. tabaci* is still unknown. With the advent of next-generation sequencing (NGS), characterizing the level of variation through amplicon sequencing of the mtCOI gene in the population has become much easier. This technique allows for the processing and characterization of a large amount of sequence data, making it simple to capture sequence variations within a single gene fragment. Therefore, we aimed to unravel the genetic variation and structure of *T. tabaci* on Allium hosts from different geographic locations across India using amplicon sequencing of the mtCOI gene. In this study, we characterized the extent of genetic crypticness within the onion-infesting *T. tabaci* population in India using advanced bioinformatics tools.

Materials & Methods

Ethics Statement: *Thrips tabaci* sampling from *Allium* host did not involve any endangered species.

Specimen Collection

T. tabaci adults were collected from *Allium* hosts at 10 locations among different climatic regions across India between 2017 and 2019, as mentioned below in table 1. Collected specimens were preserved in 95% ethanol and stored at −20°C prior to DNA extraction.

PCR Amplification of a COI Fragment and amplicon sequencing

Thrips were removed from ethanol stock using fine brush. Preserved thrips were dried for 5 minutes and total genomic DNA was extracted using DNeasy Blood & Tissue Kit (Qiagen GmbH, Germany). A 525 bp fragment of the mitochondrial cytochrome oxidase I was amplified using primer pair COI-1F TGTTACGGCTCACGCTTTTG and COI-3R-TAAAACAGGGTCCCCTCCCC (Gawande et al., 2017). PCR was performed in 20 µl reaction volume, with the following cycling conditions; 95°C (3 min), 35 cycles of 95°C (40 s), 48°C (1 min), 72°C (2 min); and final extension of 72°C (5 min). The PCR reaction yielded a 525 bp amplicon. The PCR products from each locality were examined on 1.5% agarose gel electrophoresis. PCR products were purified using PCR purification kit (Qiagen) and equal concentration of each sample was sent for amplicon sequencing to the commercial firm. Amplicon gene sequencing was performed at AgriGenome Labs Private Limited (Kerala, India) on an Illumina MiSeq 250 x 2 platform (Illumina, Inc., San Diego, CA, USA). Read quality check parameters such as base quality score distributions, average base content per read and GC distribution in the reads, check for over-represented sequences, Adapter trimming were used for quality analysis.

Preparation of data set for genetic diversity analyses

MtCOI gene sequences were retrieved from NCBI-GenBank which were collected from India. Due to significant sequence length variation, two groups were formed and processed separately for computational analysis. Group A consists sequences having maximum length of 516 nucleotides whereas group B consists 675 nucleotide long sequences. Group A and group B have 25 and 22 sequences respectively (Table 2 and 3). Multiple sequence alignment was done using CLUSTALW in MEGA-X. Mean genetic distance was calculated using uncorrected p-distance model in MEGA-X. Analysis was carried out in MEGA X using maximum likelihood method on the Jukes-Cantore model (Tamura et al., 2013). Genetic diversity analysis was done using DNASP v5 and Arlequin 3.5 tools.

SNP analysis

The raw amplicon sequence data was checked with respect to quality using FASTQC (https://github.com/s-andrews/FastQC/releases/tag/v0.12.1). The sequences were aligned taking isolate DOGR_1 (KF724977.1) as reference sequence using Bowtie2 (https://bowtie-bio.sourceforge.net/bowtie2/index.shtml). SNP calling and filtering was done using BCFtoolsmpileup (https://samtools.github.io/bcftools/bcftools.html).

Results

Sequencing Data

Data generated from amplicon sequencing used for phylogenetic analysis. The sequence produced by Illumina sequencing was used to find out the homologs of DOGR isolate using NCBI BLAST. The result obtained from BLAST given closest homologs of DOGR new isolate. These homologs were used for phylogenetic analysis. A raw amplicon sequencing statistics is given in Table 4. The highest number of reads were generated in

sample from Sikkim (247,362) and lowest reads were from Tamilnadu (204,698). GC content was ranged between 33.6 to 34.5 %. More than 84 % reads were passed the 30 Phred score which indicated good quality of data. Phred quality score numerically expresses the accuracy of each nucleotide. Higher Q number signifies higher accuracy of data. Raw sequencing data was submitted NCBI SRA database (PRJNA917098).

Genomic distribution of SNPs

We obtained partial (420 bp) mitochondrial DNA (mtDNA) Cytochrome Oxidase-I (COI) sequences for *T. tabaci* sampled from *Allium* hosts for 8 different climatic regions of India. Single nucleotide polymorphisms (SNPs) within the partial COI gene differentiated *T. tabaci* populations into 30 mtDNA haplotypes in total, with 17 haplotypes in group A and 13 haplotypes in group B.

Locality wise percent variation was calculated from total reads recorded. Among all localities, 307_Palampur represents highest nucleotide polymorphism followed by 306_Tamilnadu, 305_Maharashtra, 303_Haryana and 302_Gujarat. Compared to other localities, low level of mitochondrial genetic variation was observed in 304_Delhi isolates. We examined the frequency and distribution of SNPs polymorphisms at different locations in India. Numerous insertions and deletions of varying lengths were found in the genomic positions of different localities. At the genome position 300-400 for each locality, highest variation was noticed. Besides, lowest variation was observed at the genomic position 150-250. We concluded that, the range of diversity at nucleotide level reported within mtCOI gene is from (1-420 bp) which was used in this study using high throughput sequencing (Fig. 1, 2).

Molecular phylogenetic analysis

169 Phylogenetic trees inferred that molecular diversity in same geographical region is
170 possible if the environmental conditions are favourable. Favourable environmental condition led
171 to higher molecular diversity. Molecular evolution process might not get affected by
172 geographical proximity in various subpopulations as sequences have clustered significantly
173 different (not clustered as per their geographical locations). Both phylogenetic trees most
174 sequences form individual nodes suggesting highly diverse population. In Fig.3, northern and
175 north-eastern (Sikkim strain) sub-populations have similarities as they form one clade which is
176 most distant from isolate of central region of India (Madhya Pradesh) whereas southern sub-
177 population (from Karnataka) forms one clade indicating presence of uniformly evolving
178 population. Diversity in north-eastern sub-population can be seen as two nodes grouped
179 differently. Similarly strains from eastern India (Odisha and West Bengal as well as Bihar).
180 In Fig.4, newly sequenced isolates were closely related irrespective of geographical
181 locations indicating stabilizing of *thrips* population (302_Gujrat, 305_Maharashtra,
182 309_Tripura, 306_Tamilnadu). Strains from West Bengal have uniformity in composition as
183 they group with different geographical locations where as majority of strains isolated from
184 Karnataka have more diverse compositions. Most population forms individual nodes indicating
185 increasing molecular diversity in group B as well.
186
187 ***Genetic diversity analysis***
188 Genetic diversity parameters such as number of polymorphic sites (S), number of
189 haplotypes (H), haplotype diversity (Hd), nucleotide diversity (Pi), and average number of
190 nucleotide differences (K) were determined in DnaSP 5.12.01 (Librado & Rozas 2009). All
191 sequence analysis of group A suggests that the population is stable with long evolutionary

history with high Hd and Pi values (Hd > 0.5 and Pi > 0.005) (Table 5). Eastern population of group A and group B as well as northern population of group A indicate the same. Southern population of group A have high Hd values with low Pi values (Hd > 0.5 and Pi < 0.005) representing population bottleneck followed by rapid growth in population (Supplementary File 1). Western population of group A shows recent population bottleneck or very few mtDNA lineage as they have low Hd and Pi values (Hd < 0.5 and Pi < 0.005) (Grant & Bowen 1998).

High Nm values indicating high gene flow were observed across population in Group A as compared to Group B. North-eastern and eastern Indian sub-populations with respect to northern and southern sub-populations from group A and western population with northern subpopulation from group B show Nm values less than 1 indicating low gene flow which show there might be higher genetic differentiation occurring in these sub-population within specific geographical territories. Fst values are observed to be below 0.5 which implies the overall similarity between various subpopulations of group A and group B. The low Fst value of between western, southern and northern subpopulations with north-eastern sub-populations in group A indicates low degree of differentiation amongst both subpopulations.

Haplotype networks using TCS algorithm were constructed in PopART (Fig. 5, 6). The networks show that group A consists of 8 haplotypes with 307_Himachal Pradesh, 303_Haryana as most distant haplotypes. In group B, with presence of 7 haplotypes; most distant haplotype is from Tripura. Some states represent more haplotypes like Maharashtra (KF724977.1_Maharashtra), Delhi (MN594551.1_Delhi) and West Bengal that may be indication of stabilizing population or successful niche formation by the species. Study with larger dataset is required to study the diversity and molecular evolution especially in these geographical regions.

Discussion

Identification of Thrips species (Insecta, Thysanoptera) using conventional methods is difficult. In the current study, DNA barcoding was utilized to help characterize thrips species from a large samples collected in onion fields. This study presents results from the informative region of the COI partial gene in *Thrips tabaci* sampled from 8 different geographical regions in India. Based on the survey data, it was found that throughout the selected localities, there is significant differences among them. The phylogenetic tree generated in the study shows high degree of diversity with more and new gene sequences which suggests that Indian thrips isolates continue to be distinct and more diverse. Diverse study of grouping thrips species with higher intraspecific genetic variation inferred from mtCOI gene predominantly in *T. tabaci* reported from different countries on important vegetables and crops host (Kadirvel et al., 2013; Li et al., 2020).

In the present study, nucleotide sequences of mtCOI gene were used for population genetics analysis that led to characterisation of within species diversity, geographical distribution of a species and tracing the phylogeographic origin (Lombaert et al., 2014). In the literature, up till now only 28 haplotypes of mtCOI gene of *Thrips tabaci* were reported worldwide. These haplotypes were retrieved from nearly 282 nucleotide sequences. It includes Asia, Europe, Australia and America and other 23 countries (Iftikhar et al., 2016). This study investigated the Single nucleotide polymorphisms (SNPs) within the partial COI gene differentiated *T. tabaci* populations into 30 mtDNA haplotypes in total, with 17 haplotypes in group A and 13 haplotypes in group B. It represents higher genetic variation or diversity of haplotypes in Indian population of *T. tabaci.* Among all localities, the mtCOI region of thrips from Palampur

represents highest nucleotide polymorphism followed by Tamilnadu, Maharashtra, Haryana and Gujarat. Compared to other localities, low level of mitochondrial genetic variation was observed in Delhi isolates.

Our study represents a comprehensive sequence data analysis with respect to molecular phylogeny and genetic diversity aspects. In order to achieve the computational based results of genetic diversity, two groups i.e. A & B were formed with respect to the variation in length of mtCOI gene of *thrips*. High Nm values indicating high gene flow were observed across population in Group A as compared to Group B. This analysis revealed that these groups showed recent population bottleneck with varying Hd and Pi values (Grant and Bowen 1998). Small Hd and Pi values (Hd 0.5 and Pi 0.005) indicate a recent population bottleneck or founder event by a single or a few mtDNA lineages; high Hd and low Pi values (Hd > 0.5 and Pi 0.005) indicate a population bottleneck followed by rapid population increase and mutation accumulation; low Hd and high Pi (Hd 0.5 and Pi > 0.005) represent geographically subdivided populations; High Hd and Pi values (Hd > 0.5 and Pi > 0.005) indicate a large stable population with a long evolutionary history or secondary contact between distinct lineages (Grant & Bowen 1998). This helps us to note that, genetic drift might confer in the change of diversity and number of haplotypes (Balloux et al., 2003). *T. tabaci* populations from central and southern region of India exhibited high levels of genetic differentiation and diversity compared to other populations, while the north-eastern region showed low genetic flow. The presence of isolation by distance and absence of recent gene flow suggests that genetic drift plays an important role in genetic differentiation. The data from this study extended our knowledge on diversity of *T. tabaci* in India, and afford idea about the management practices against it.

260 The pairwise *FST* and gene flow (*N*m) among different zones i.e. West, South, East and
261 North of Indian subcontinent were calculated. Significant difference was observed in Group A
262 and Group B, where Group A compromises southern and western part of India shows high *N*m.
263 For Group B also high *N*m was recorded which compromises southern part. Presence of *N*m < 1,
264 reflects low level of gene flow and ultimately higher genetic variation such as in north-eastern
265 and northern part (Tiroesele et al., 2014). Varying level of gene flow might due to the
266 geographical and genetic distance. On the other hand, Fst values were found to be under 0.5 for
267 majority of the samples which is relative measure of genetic variation that revealed overall
268 similarity. But for Group A compromising northern, north-eastern and western subpopulation,
269 Fst value < 0.5 suggesting low level of genetic variability.

270 Haplotype networks generated in this study suggest that Delhi and Maharashtra states
271 represent more haplotypes. Also, these result support molecular phylogeny result which states
272 that, geographical proximity cannot be used to measure molecular evolutionary patterns in
273 mtCOI. The haplotype networks show that group A consists of 8 haplotypes with 307_Himachal
274 Pradesh, 303_Haryana as most distant haplotypes. In group B, with presence of 7 haplotypes;
275 most distant haplotype is from Tripura. Some states represent more haplotypes like Maharashtra
276 (KF724977.1_Maharashtra), Delhi (MN594551.1_Delhi) and West Bengal that may be
277 indication of stabilizing population or successful niche formation by the species.

278 There have been reports on link between traits, reproductive modes and distribution of
279 genetic variation from different countries where particularly mode of reproduction of thrips
280 population affects the diversity among them (Westmore et al., 2013; Gawande et al., 2017;
281 Jacobson et al., 2013, Nault et al., 2014, Jacobson et al., 2016). In this study, we could not have

highlighted the presence of arrhenotokous and thelytokous populations which helps to rule out low or high genetic diversity as this part need to explore for Indian thrips.

Conclusions

Our investigation showed agreement as per previous reports, that both IYSV and *Thrips tabaci* isolates are most diverse in Indian subcontinent. This is the first study to examine the population genetic structure of *T. tabaci* populations using amplicon sequencing of mtCOI gene. MtCOI gene targeted here revealed genetic structure among the populations that corresponded to both the geographic locations where the populations were collected, and the grouping of individuals observed in the phylogenetic analysis conducted with mtCOI sequences.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

Suresh Gawande conceived the idea and designed the experiment, Tushar Gawai and Sharwari Sadawarte performed experiment and analysed data, Kiran Khandagale, Anusha R prepared Manuscript, Abhijeet Kulkarni and Avinash Ade supervised the experiment and reviewed the manuscript. Suresh Gawande and Durgesh Kumar Jaiswal helped in improving MS. All authors reviewed manuscript and approved the final draft.

References

Asokan, R., Krishna Kumar, N. K., Kumar, V., & Ranganath, H. R. (2007). Molecular differences in the mitochondrial cytochrome oxidase I (mtCOI) gene and development of a species-specific marker for onion *thrips*, *Thrips* tabaci Lindeman, and melon *thrips*, T. palmi Karny (Thysanoptera: Thripidae), vectors of tospoviruses (Bunyaviridae). Bulletin of Entomological Research, 97(5), 461-470. doi:10.1017/S0007485307005147

Balloux, F., Lehmann, L., & de Meeûs, T. (2003). The population genetics of clonal and partially clonal diploids. *Genetics*, *164*(4), 1635-1644.

Brunner, P. C., Chatzivassiliou, E. K., Katis, N. I., & Frey, J. E. (2004). Host-associated genetic differentiation in Thrips tabaci (Insecta; Thysanoptera), as determined from mtDNA sequence data. Heredity, 93(4), 364-370. doi:10.1038/sj.hdy.6800512

Buckman, R. S., Mound, L. A., & Whiting, M. F. (2013). Phylogeny of thrips (Insecta: Thysanoptera) based on five molecular loci. Systematic Entomology, 38(1), 123-133. doi:https://doi.org/10.1111/j.1365-3113.2012.00650.x

De Kogel, W. J., Koschier, E. H., Broughton, S., Castañé, C., Davidson, M. M., Hamilton, J. G. C., & Teulon, D. A. J. (2015). Semiochemicals for sustainable thrips management. *Bodenkultur*, *66*(3-4), 17-25.

Diaz-Montano, J., Fuchs, M., Nault, B. A., Fail, J., & Shelton, A. M. (2011). Onion Thrips (Thysanoptera: Thripidae): A Global Pest of Increasing Concern in Onion. *Journal of Economic Entomology, 104*(1), 1-13. doi:10.1603/EC10269

Gawande, S. J., Anandhan, S., Ingle, A. A., Jacobson, A., & Asokan, R. (2017). Heteroplasmy due to coexistence of mtCOI haplotypes from different lineages of the Thrips tabaci cryptic species group. *Bulletin of entomological research*, *107*(4), 534. DOI:10.1017/S0007485317000025

Glover, R. H., Collins, D. W., Walsh, K., & Boonham, N. (2010). Assessment of loci for DNA barcoding in the genus Thrips (Thysanoptera:Thripidae). *Molecular Ecology Resources, 10*(1), 51-59. doi:https://doi.org/10.1111/j.1755-0998.2009.02723.x

Grant, W. A. S., & Bowen, B. W. (1998). Shallow population histories in deep evolutionary lineages of marine fishes: insights from sardines and anchovies and lessons for conservation. *Journal of heredity*, *89*(5), 415-426.

Hebert, P. D. N., Cywinska, A., Ball, S. L., & deWaard, J. R. (2003). Biological identifications through DNA barcodes. Proceedings of the Royal Society of London. Series B: Biological Sciences, 270(1512), 313-321. doi:10.1098/rspb.2002.2218

Iftikhar, R., Ashfaq, M., Rasool, A., & Hebert, P. D. (2016). DNA barcode analysis of thrips (Thysanoptera) diversity in Pakistan reveals cryptic species complexes. *PLoS One*, *11*(1), e0146014.

International Barcode of Life Consortium, n.d., Centre for Biodiversity Genomics in Guelph, Canada, accessed 23 March 2023, <http://www.ibol.org/resources>

Jacobson, A. L., & Kennedy, G. G. (2013). Specific insect-virus interactions are responsible for variation in competency of different Thrips tabaci isolines to transmit different Tomato spotted wilt virus isolates. *PLoS One*, *8*(1), e54567.

Jacobson, A. L., Nault, B. A., Vargo, E. L., & Kennedy, G. G. (2016). Restricted gene flow among lineages of Thrips tabaci supports genetic divergence amongag cryptic species groups. *PloS one*, *11*(9), e0163882.

Jenser, G., & Szénási, Á. (2004). Review of the biology and vector capability of Thrips tabaci Lindeman (Thysanoptera: Thripidae). Acta Phytopathologica et Entomologica Hungarica, 39(1-3), 137-155. doi:10.1556/aphyt.39.2004.1-3.14

Jenser, G., Alamasi, A., Fail, J. & Tobias, I. (2011). Differences in the vector efficiency of Thrips tabaci in Europe and North America. Acta Phytopathologica et Entomologica Hungarica 46: 311–317.

Kadirvel, P., Srinivasan, R., Hsu, Y.C., Su, F.C. and De La Pena, R., 2013. Application of cytochrome oxidase I sequences for phylogenetic analysis and identification of thrips species occurring on vegetable crops. *Journal of Economic Entomology*, *106*(1), pp.408-418.

Kobayashi, K., & Hasegawa, E. (2012). Discrimination of reproductive forms of Thrips tabaci (Thysanoptera: Thripidae) by PCR with sequence specific primers. *Journal of Economic Entomology*, *105*(2), 555-559.

Kobayashi, K., Yoshimura, J., & Hasegawa, E. (2013). Coexistence Librado, P., & Rozas of sexual individuals and genetically isolated asexual counterparts in a thrips. *Scientific reports*, *3*(1), 1-4.

Librado, P., & Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics, 25(11), 1451-1452. doi:10.1093/bioinformatics/btp187

Lombaert, E., Estoup, A., Facon, B., Joubard, B., Grégoire, J. C., Jannin, A., . . . Guillemaud, T. (2014). Rapid increase in dispersal during range expansion in the invasive ladybird Harmonia axyridis. Journal of Evolutionary Biology, 27(3), 508-517. doi:https://doi.org/10.1111/jeb.12316

372 Mouden, S., Sarmiento, K. F., Klinkhamer, P. G., & Leiss, K. A. (2017). Integrated pest
373 management in western flower thrips: past, present and future. *Pest management*
374 *science*, *73*(5), 813-822.
375 Nault, B. A., A. M. Shelton, J. L. Gangloff-kaufmann, M. E. Clark, J. L. Werren, J. C. Cabrera-
376 la Rosa, and G. G. Kennedy. 2006. Reproductive modes in onion thrips (Thysanoptera:
377 Thripidae) populations from New York onion fields. Environ. Entomol. 35: 1264–1271.
378 Nault, B. A., Kain, W. C., & Wang, P. (2014). Seasonal Changes in Thrips tabaci Population
379 Structure in Two Cultivated Hosts. PLOS ONE, 9(7), e101791.
380 doi:10.1371/journal.pone.0101791
381 Rebijith, K. B., Asokan, R., Krishna Kumar, N. K., Krishna, V., Chaitanya, B. N., And
382 Ramamurthy, V. V. 2013. DNA barcoding and elucidation of cryptic aphid species
383 (Hemiptera: Aphididae) in India. Bull. Entomol. Res. 103: 601-610.
384 Shufran, K.A., Burd, J.D., Anstead, J.A. and Lushai, G.U.G.S., 2000. Mitochondrial DNA
385 sequence divergence among greenbug (Homoptera: Aphididae) biotypes: evidence for
386 host-adapted races. *Insect molecular biology*, *9*(2), pp.179-184.
387 Stuart, R.R., Gao, Y.L. and LEI, Z.R., 2011. Thrips: pests of concern to China and the United
388 States. *Agricultural Sciences in China*, *10*(6), pp.867-892.
389 Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: molecular
390 evolutionary genetics analysis version 6.0. *Molecular biology and evolution*, *30*(12),
391 2725-2729.
392 Tiroesele, B., Skoda, S. R., Hunt, T. E., Lee, D. J., Molina-Ochoa, J., & Foster, J. E. (2014).
393 Population structure, genetic variability, and gene flow of the bean leaf beetle, Cerotoma
394 trifurcata, in the Midwestern United States. J Insect Sci, 14, 62. doi:10.1093/jis/14.1.62
395 Tyagi, K., Kumar, V., Singha, D., Chandra, K., Laskar, B. A., Kundu, S., ... & Chatterjee, S.
396 (2017). DNA barcoding studies on thrips in India: cryptic species and species
397 complexes. *Scientific reports*, *7*(1), 1-14.
398 Westmore, G. C., Poke, F. S., Allen, G. R., & Wilson, C. R. (2013). Genetic and host-associated
399 differentiation within Thrips tabaci Lindeman (Thysanoptera: Thripidae) and its links to
400 Tomato spotted wilt virus-vector competence. *Heredity*, *111*(3), 210-215.
401 Li, X., Zhang, Z., Zhang, J., Huang, J., Wang, L., Li, Y., Hafeez, M. and Lu, Y., 2020.
402 Population genetic diversity and structure of Thrips tabaci (Thysanoptera: Thripidae) on

Allium hosts in China, inferred from mitochondrial COI gene sequences. *Journal of Economic Entomology*, *113*(3), pp.1426-1435.

Figure Captions

Fig 1. SNPs distribution vs Genome position (The x-axis corresponds to the genomic position of each SNP represented by a dot on the graph. The y-axis is distribution of each SNP. Each colour corresponds to a sample collected from different localities.)

Fig 2. Locality wise percent reads with variations

Fig. 3: Phylogenetic analysis of Group A using Maximum likelihood method. Violet represents sequences from northern India, blue from southern India, red from western India, green from eastern India, pink from north-eastern India, and yellow from central part of India.

Fig. 4: Phylogenetic analysis of Group B using Maximum likelihood method. Blue represents sequences from southern India, green from eastern India, red from Northern India, and pink from western India.

Fig. 5: Haplotype network visualized in POPART for group A Numbers in bracket on the edges signify number of mutations and KF724977.1-Maharashtra represents predominant haplotype. Network has 8 haplotypes

Fig. 6: Haplotype network visualized in POPART for group B Numbers in bracket on the edges signify number of mutations and MN594551.1-Delhi represents predominant haplotype. Network has 7 haplotypes.

Figure 1

SNPs distribution vs Genome position (The x-axis corresponds to the genomic position of each SNP represented by a dot on the graph. The y-axis is distribution of each SNP. Each colour corresponds to a sample collected from different localities.)

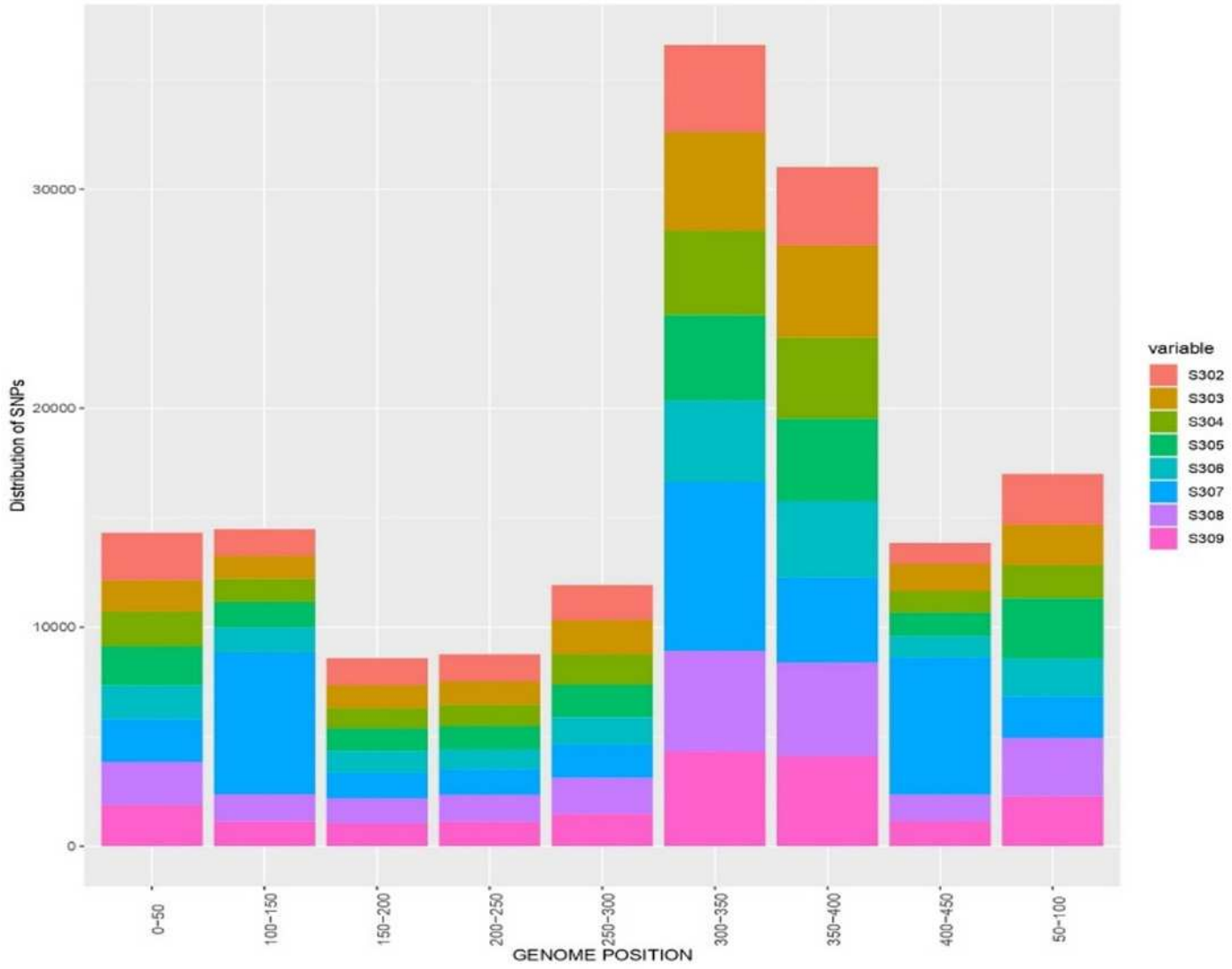

Figure 2

Locality wise percent reads with variations

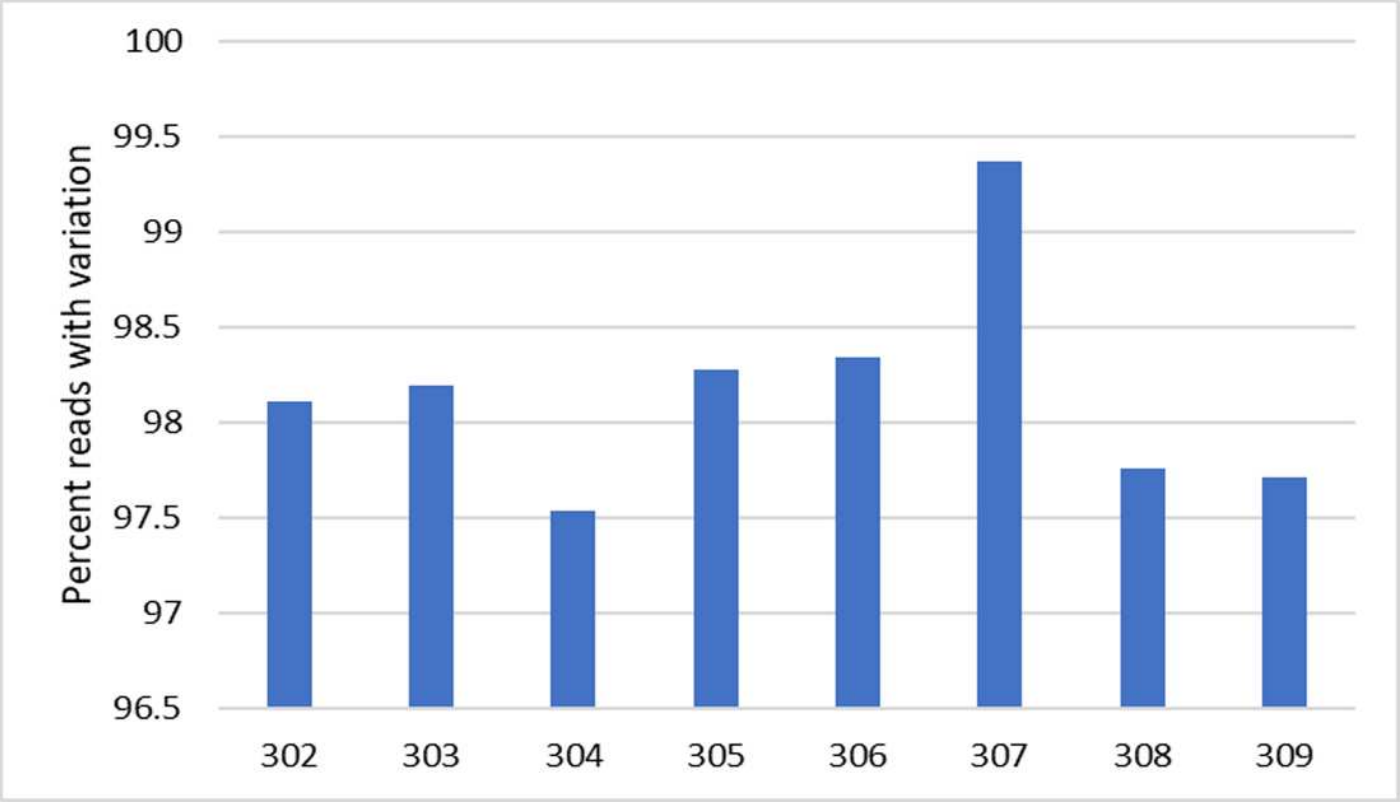

Figure 3

Phylogenetic analysis of Group A using Maximum likelihood method.

Phylogenetic analysis of Group A using Maximum likelihood method. Violet represents sequences from northern India, blue from southern India, red from western India, green from eastern India, pink from north-eastern India, and yellow from central part of India.

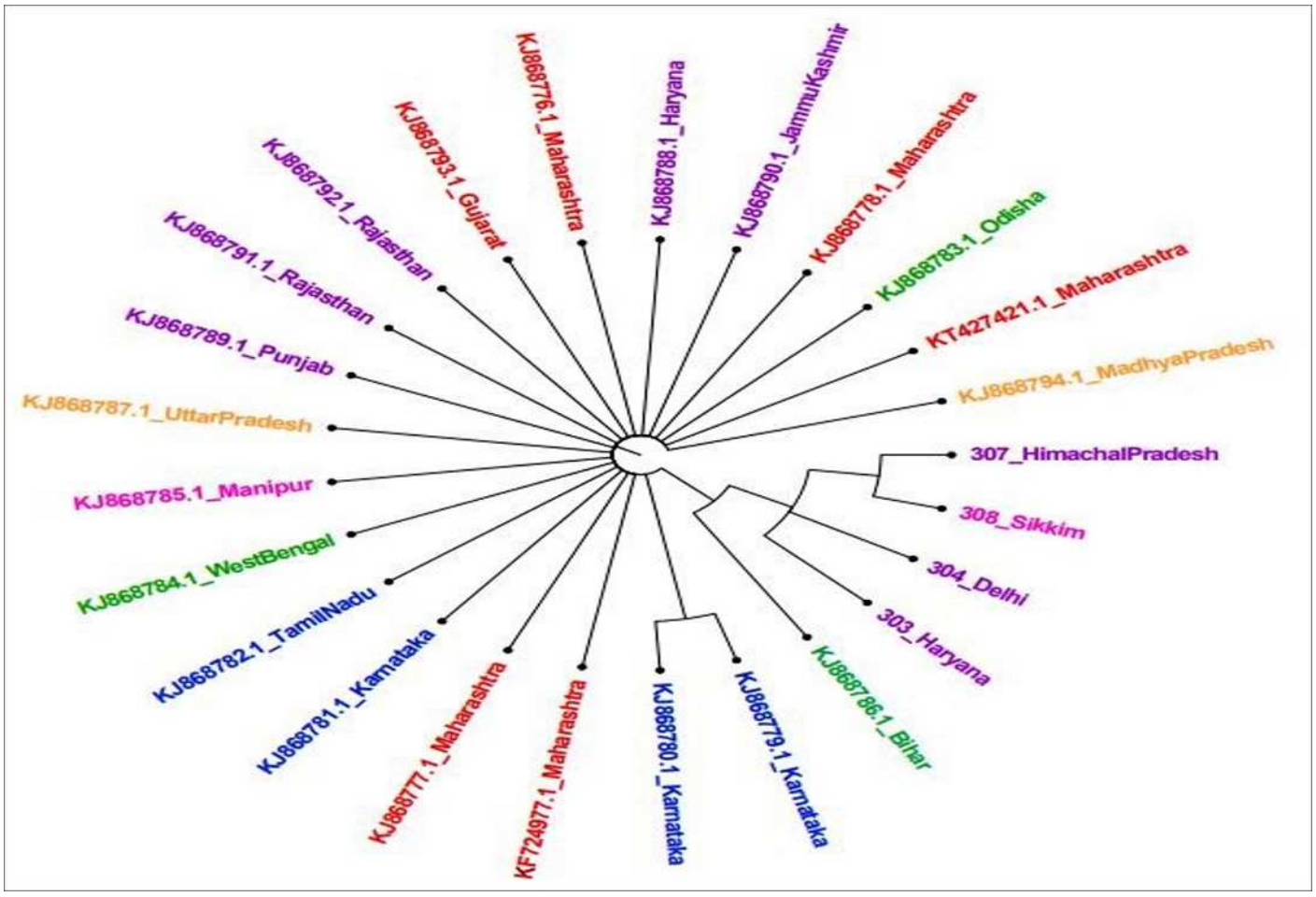

Figure 4

Phylogenetic analysis of Group B using Maximum likelihood method. Blue represents sequences from southern India, green from eastern India, red from Northern India, and pink from western India.

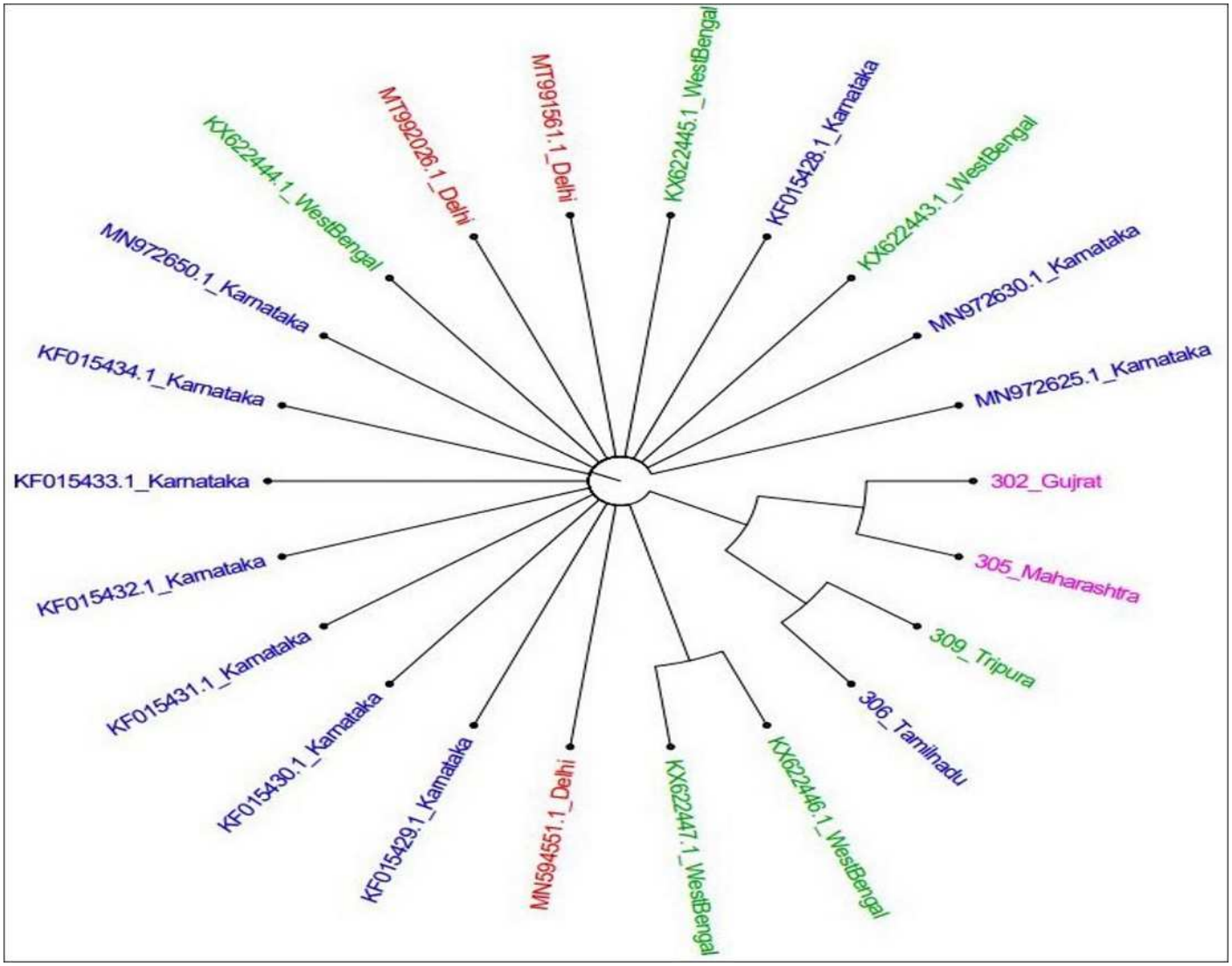

Figure 5

Haplotype network visualized in POPART for group A Numbers in bracket on the edges signify number of mutations and KF724977.1-Maharashtra represents predominant haplotype. Network has 8 haplotypes

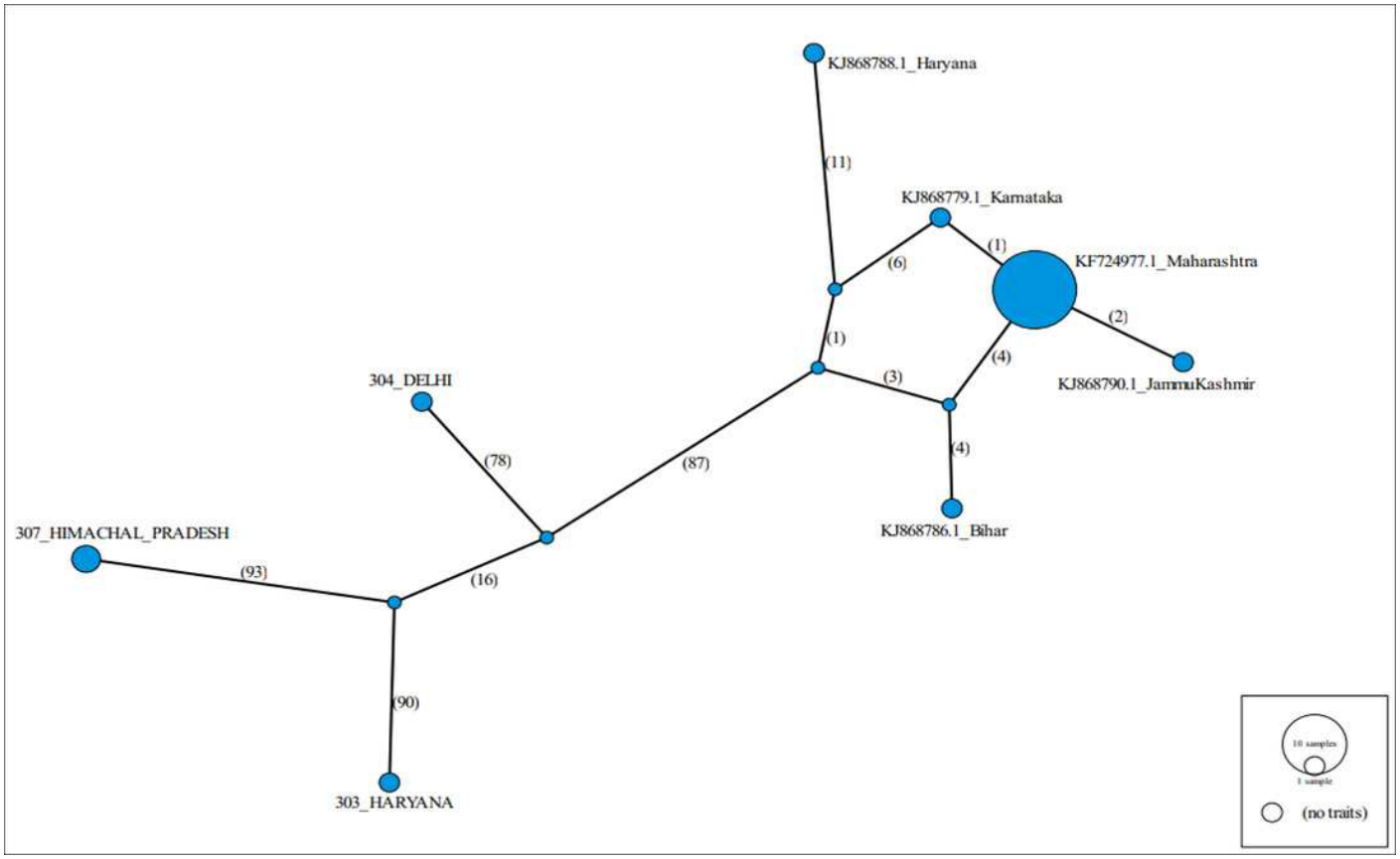

Figure 6

Haplotype network visualized in POPART for group B Numbers in bracket on the edges signify number of mutations and MN594551.1-Delhi represents predominant haplotype. Network has 7 haplotypes.

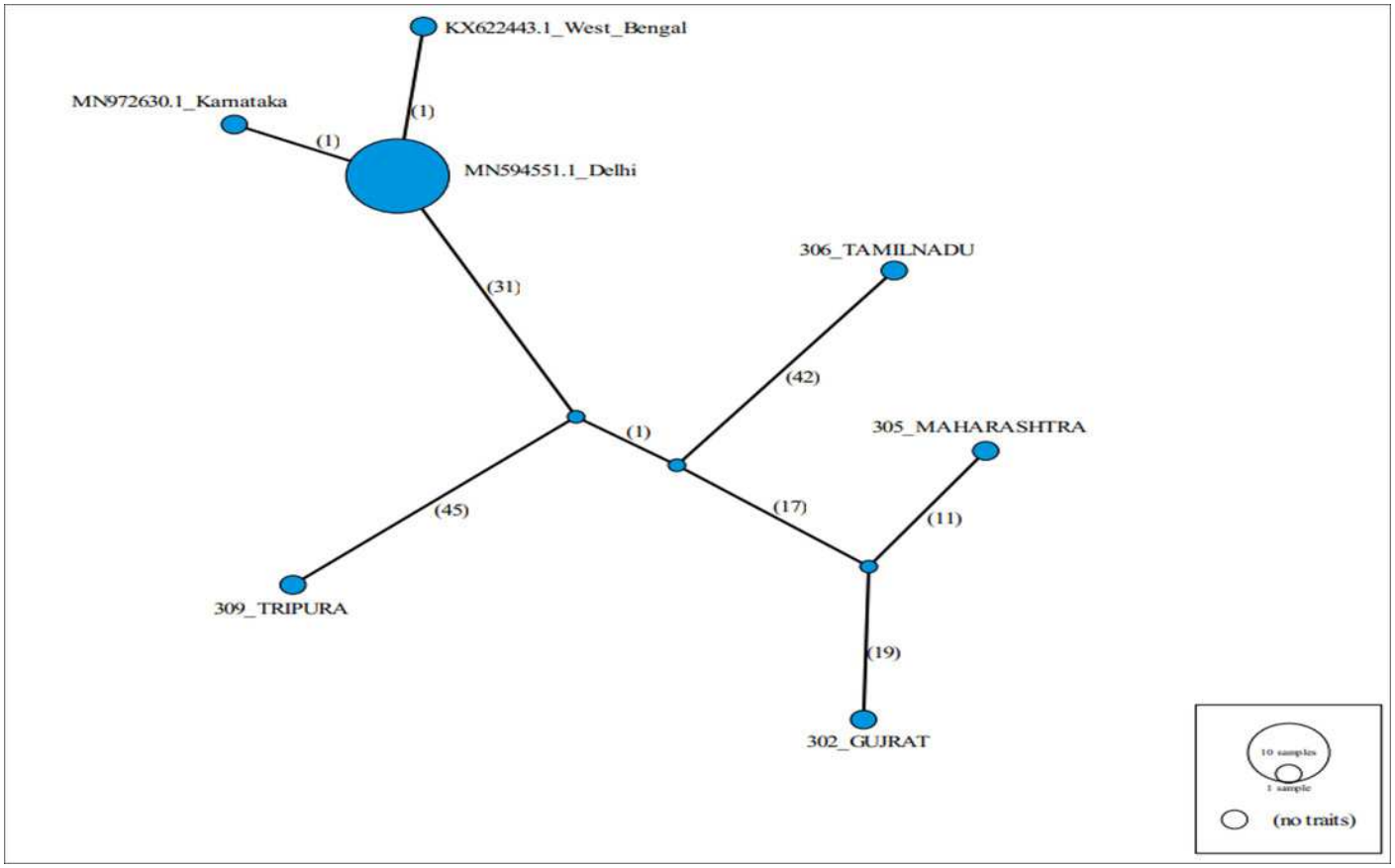

Table 1(on next page)

Detail information of collection of *Thrips tabaci* from India

Table 1: Detail information of collection of *Thrips tabaci* from India

Agroclimatic Zone	State	Collection site	Longitude (E) Latitude (N)
I	Himachal Pradesh	Palampur	32°06'02.0"N 76°32'47.6"E
III	Sikkim	Gangtok	27°19'11.1"N 88°36'07.4"E
	Tripura	Lembucherra	23°54'24.0"N 91°18'47.6"E
VI	Delhi	IARI	28°38'10.6"N 77°9'27.5"E
	Haryana	Karnal	29°44'55.4"N 76°59'48.5"E
	Gujarat	Junagadh	21°30'22.7"N 70°26'57.8"E
VII	Maharashtra	Rajgurunagar	18°50'34.6"N 73°53'04.8"E
VIII	Tamilnadu	Coimbatore	11°0'32.2"N 73°56'4.33"E

Table 2(on next page)

Distribution of Group A

1 **Table 2: Distribution of Group A**

Accession ID	Geographical location	Sub-population group
KF724977.1	Maharashtra	Sub-population 1: WEST
KJ868776.1	Maharashtra	
KJ868777.1	Maharashtra	
KJ868778.1	Maharashtra	
KT427421.1	Maharashtra	
KJ868793.1	Gujarat	
KJ868792.1	Rajasthan	
KJ868791.1	Rajasthan	
KJ868779.1	Karnataka	Sub-population 2: SOUTH
KJ868780.1	Karnataka	
KJ868781.1	Karnataka	
KJ868782.1	Tamilnadu	
KJ868783.1	Odisha	Sub-population 3: EAST
KJ868784.1	West Bengal	
KJ868786.1	Bihar	
KJ868788.1	Haryana	Sub-population 4: NORTH
KJ868789.1	Punjab	
KJ868790.1	Jammu Kashmir	
303	Haryana	
304	Delhi	
307	Himachal Pradesh	
KJ868794.1	Madhya Pradesh	Sub-population 5: CENTRAL

KJ868787.1	Uttar Pradesh	
KJ868785.1	Manipur	Sub-population 6: NORTH-EAST
308	Sikkim	

2

Table 3(on next page)

Distribution of Group B

Table 3: Distribution of Group B

NCBI GenBank Accession ID	Geographical Location	Sub-population
MN594551.1	Delhi	Sub-population 1: NORTH
MT991561.1	Delhi	
MT992026.1	Delhi	
MN972625.1	Karnataka	Sub-population 2: SOUTH
MN972630.1	Karnataka	
KF015428.1	Karnataka	
KF015429.1	Karnataka	
KF015430.1	Karnataka	
KF015431.1	Karnataka	
KF015432.1	Karnataka	
KF015433.1	Karnataka	
KF015434.1	Karnataka	
306	Tamilnadu	
KX622444.1	West Bengal	Sub-population 3: EAST
KX622445.1	West Bengal	
KX622446.1	West Bengal	
KX622447.1	West Bengal	
KX622443.1	West Bengal	
309	Tripura	
302	Gujrat	Sub-population 3: WEST
305	Maharashtra	

Table 4(on next page)

Raw amplicon sequencing data

Table 4: Raw amplicon sequencing data

Localities	Total read bases (bp)	Total reads	GC(%)	AT(%)	Q20(%)	Q30(%)
Junagadh	71,282,820	236,820	34.04	65.96	92.73	86.52
Karnal	70,529,116	234,316	34.53	65.47	90.5	84.24
Delhi	66,422,874	220,674	34.05	65.95	93.34	87.39
Maharashtra	64,141,294	213,094	33.87	66.13	94.22	88.41
Tamilnadu	61,614,098	204,698	34.37	65.63	91.73	85.59
Palampur	66,067,694	219,494	33.61	66.39	93.62	87.75
Sikkim	74,455,962	247,362	33.7	66.3	95.25	89.83
Tripura	67,244,604	223,404	33.73	66.27	95.04	89.25

Table 5(on next page)

Genetic diversity analysis of group A and B

1 **Table 5: Genetic diversity analysis of group A and B**

Population	sequences analyzed	Number of polymorphic sites (S)	Haplotype diversity (Hd)	Nucleotide diversity (Pi)	Average number of nucleotide differences (K)	Number of haplotypes	Group
All sequences	25	640	0.69	0.46	74.94	17	A
WEST	8	3	0.25	0.0014	0.75	2	
SOUTH	4	2	0.83	0.0022	1.16	3	
EAST	3	11	1.00	0.014	7.33	3	
NORTH	6	640	1.00	1.65	195.4	6	
CENTRAL	2	0	0.00	0	0.00	1	
NORTH EAST	2	499	1.00	1.16	247.00	2	
All sequences	22	1035	0.87	0.20	62.72857	13	B
NORTH	3	78	1.00	0.25	0.33	3	
SOUTH	11	584	0.49	0.12	59.09	4	
EAST	5	20	9.0	0.00683	4.40	4	
WEST	2	666	1.0	0.39333	118.00	2	

2

3