Mitochondrial COI based genetic diversity and phylogeographic structure of whitefly Bemisia tabaci *(Gennadius) on cotton in India*

Prabhulinga T, Sandhya Kranthi, Raghavendra K P, Rishi Kumar, Ruchika Suke, Shilpa Chawla & Keshav Raj Kranthi

International Journal of Tropical Insect Science

e-ISSN 1742-7592

Int J Trop Insect Sci
DOI 10.1007/s42690-020-00354-x

Springer

International Journal of Tropical Insect Science
https://doi.org/10.1007/s42690-020-00354-x

ORIGINAL RESEARCH ARTICLE

Mitochondrial COI based genetic diversity and phylogeographic structure of whitefly *Bemisia tabaci* (Gennadius) on cotton in India

Prabhulinga T[1] · Sandhya Kranthi[1] · Raghavendra K P[1] · Rishi Kumar[2] · Ruchika Suke[1] · Shilpa Chawla[1] · Keshav Raj Kranthi[3]

Received: 28 April 2020 / Accepted: 23 October 2020

Abstract

Whitefly *Bemisia tabaci* (Gennadius) is an important pest of cotton in India. The study of taxonomic diversity and its distribution on cotton is lacking. Such studies are necessary to identify the genetic groups of *B. tabaci* and its distribution on cotton in India. The proper identification of genetic groups and their distribution, which, ultimately, can lead to the timely development and utilization of management practices. The current study was undertaken to explore the genetic diversity and phylogeographic structure of *B. tabaci* from all the three major cotton growing zones representing different agro-climatic conditions of India. 290 partial mitochondrial Cytochrome Oxidase I (mt COI) sequences of whitefly population of six major cotton-growing states covering 22 districts were used in the analysis. Phylogenetic analysis revealed the presence of two monophyletic clades, Asia I and Asia II1 genetic groups represents South-Central India and North India respectively. Asia II1 is found more predominately distributed in the cotton leaf curl virus (CLCuV) disease-prone north Indian cotton-growing states. Higher genetic divergence (16.8–19.2%) was observed between the populations of Asia I and Asia II1 genetic group. Genetic differentiation analysis confirmed the phylogeographic structure of *B. tabaci* as isolated by distance. Our results in mapping the distribution of genetic groups in cotton ecosystems paved the way for the further studies and formulation of area-wide management practices for cotton whitely in India.

Keywords Cotton · Genetic diversity · Mitochondrial cytochrome oxidase (COI) · Phylogenetic analysis · Whitefly

Introduction

Cotton whitefly, *B. tabaci* (Gennadius) (Hemiptera: Aleyrodidae) is one of the most devastating pest worldwide (Byrne and Bellows Jr. 1991). The pest is reported to feed on 900 different hosts (Cahill et al. 1996; Hsieh et al. 2006). Causes economic damages to the crop in many cotton-growing courtiers viz., India, China, Pakistan, Brazil, USA (Kranthi et al. 2002; Naveen et al. 2017; Ashfaq et al. 2014) etc. Among sucking pests of cotton in India, whitefly causes significant damages to the crop directly and indirectly by feeding on phloem sap and by transmitting Cotton leaf curl virus (Geminivirus) respectively (Jones 2003; Briddon 2003; Naveed et al. 2007). Recently, northern cotton growing zone (Punjab, Haryana and Rajasthan) of India witnessed severe pest outbreak accompanied with cotton leaf curl virus disease (CLCuD) (Prabhulinga et al. 2019) caused loss to the tune of 0.4b USD (Kranthi 2015). Though whitefly incidence was first reported in 1905 on cotton in India (Misra and Lambda 1929), varying pest population dynamics and showed severe outbreaks were reported only after1984. The population dynamics of whiteflies are known to be regulated by several biotic and abiotic factors. In the recent past one of the major triggers for outbreaks of whiteflies were linked to hot humid climatic conditions of the cotton-growing regions (Prabhulinga et al. 2017) and indiscriminate use of pesticides, especially, synthetic pyrethroids (Kranthi et al. 2002; Sethi

T. Prabhulinga, Sandhya Kranthi and K. P. Raghavendra contributed equally to this work.

✉ Prabhulinga T
prabhut07@gmail.com

Keshav Raj Kranthi
http://www.icac.org

[1] ICAR-Central Institute for Cotton Research (CICR), Postbag No 2, Shankar Nagar PO, Nagpur, Maharashtra 440010, India

[2] ICAR-Central Institute for Cotton Research (CICR), Regional Station, Sirsa, Haryana, India

[3] International Cotton Advisory Committee (ICAC), 1629, K Street, NW, Suite 702, Washington, DC 20006, USA

Published online: 14 November 2020

and Dilawari 2008; Naveen et al. 2017). Whitefly was first described by Reaumur in 1736, however incorrectly placed in the order Lepidoptera. Later in 1795 whiteflies replaced in the suborder Homoptera of Hemiptera (Douglas 1877–78). Presently, the *B. tabaci* is a species complex; having several cryptic species (Ashfaq et al. 2014, Kumar et al. 2016a, 2016b, Khatun et al. 2018). Classification of cryptic species of *B. tabaci* became an overwhelming task for the classical taxonomists of its morphological indistinctness (Gill and Brown 2010; Chaubey et al. 2015). Because of these problems, researchers relied on parameters such as genetics, host range, insecticide resistance, behaviour, virus transmission and interactions with viruses and host plants for classification of whiteflies (Xu et al. 2011; Naveen et al. 2017). Dinsdale et al. (2010) used 3.5% genetic divergence as a limit for grouping genetic groups and there are more than 36 genetic groups are identified so far based on mitochondrial encoded Cytochrome C Oxidase I genetic divergence. Among 36 global genetic groups, India have 11 of them (Asia I, Asia I-India, Asia II-1, Asia II-5, Asia II-7, Asia II-8, and Asia II-11, Middle East Asia Minor (MEAM) -1 and Middle East Asia Minor (MEAM) - K genetic groups on different alternate hosts including cotton (Chowda-Reddy et al. 2012; Ellango et al. 2015; Roopa et al. 2015). India is the largest cotton growing country in the world with an area of 10.5 mha (AICCIP 2016–17). Though cotton is grown more than 10 major states of India, comprehensive information on whiteflies, their distribution and diversity (genetic groups or putative species) on cotton are scanty. Therefore, the study comprising genetic diversity and phylogeographic structure of *B. tabaci* is carried out.

Materials and method

Sampling

Adult whiteflies were collected from 58 geographical locations of major cotton-growing regions representing south, central and northern cotton growing zone of India (Fig. 1 Geographic distribution of *Bemisia tabaci* on cotton in India). The collected whiteflies adults were kept in 70% ethanol (Bedford et al. 1994) and stored at 4 °C until further use. Five whitefly samples were used from each location

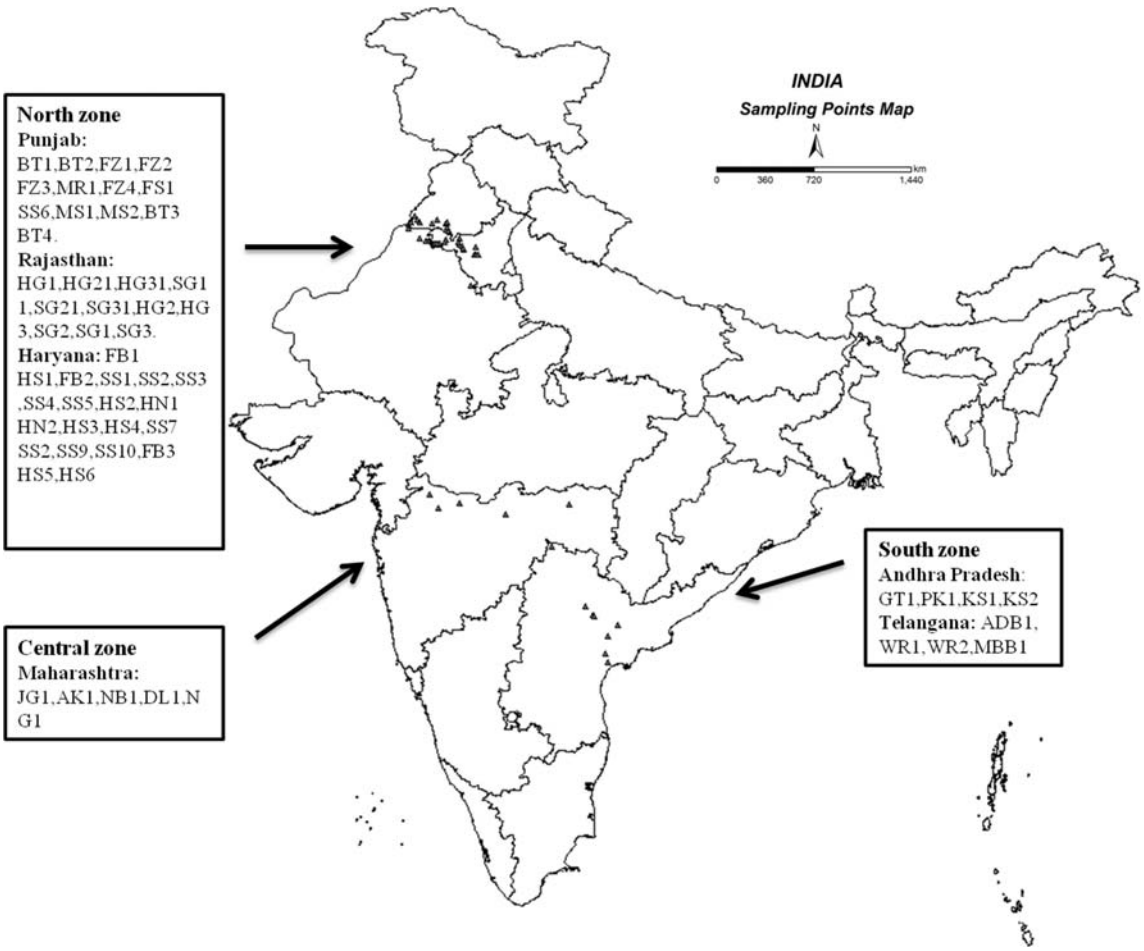

Fig. 1 Geographic distribution of *B. tabaci* on cotton in India

amounting 290 samples. The survey covered 22 districts of 6 major cotton-growing states of India. Geographic details of cotton whiteflies collection in India are enlisted in Table 1.

Genomic DNA extraction

DNA was extracted from individual whiteflies by following the protocol of Frohlich et al. 1999. The whiteflies were placed on a strip of parafilm and ground with 40 μL of ice-cold lysis buffer from the blunt end of a 0.2 ml PCR tubes. The lysis buffer comprised of 5 mM of Tris-HCl, pH 8.0, 0.5 mM of EDTA, 0.5% Nonidet P-40 and 1 mg/mL of proteinase K. The Extracts obtained after crushing were incubated at 65 °C for 15 min followed by 95 °C for 10 min before centrifugation (12,000 rpm for 10 min at 4 °C) to pellet debris. The supernatant collected was used as the template for PCR.

Polymerase chain reaction (PCR) and sequencing

The synthesized PCR primers used for amplification of mitochondrial partial COI gene region are, C1-J-2195 (5′-TTGA TTTTTTGGTCATCCAGAAGT-3′) and L2-N-3014 (5′-TCCAATGCACT- AATCTGCCATATTA-3′) (Simon et al. 1994), which were outsourced from Eurofins Genomics India Private Limited, Bangalore. PCR assays were performed by using 2 μL of genomic DNA as a template in a 25 μL reaction volume. The PCR reaction mixture comprised of 16.7 μL of nuclease-free water, 2.5 μL of 10X buffer, 0.5μLof 10 mM of dinucleotide triphosphates (dNTPs), 2 μL 25 mM of MgCl2, 0.3 μL of 3 units/ μL of Taq DNA polymerase and 0.5 μL of each 10 μM primer. All the PCR reagents were purchased from Merck India (Bangalore GeNei Private Limited) Bengaluru, India. The PCR performed with the specification of initial denaturation at 95 °C for 2 min, denaturation at 95 °C for 1 min, annealing at 52 °C for 1 min and extension at 72 °C for 1 min for 35 cycles and a final extension at 72 °C for 5 min. PCR products were separated on 1.2% agarose gel electrophoresis and later were purified by ExoSAP-IT™ PCR product cleanup reagent by following manufacturer's instructions before subjected to DNA sequencing (Eurofins genomics India Pvt. Ltd).

Genetic diversity and phylogenetic analysis

The 290 partial Mt COI sequences generated from the sequencing of 650 bp PCR product were trimmed in Bio-edit and obtained 425 bp clean and good sequences. Finally, the 58 consensus sequences representing each selected geographical locations were identified for further analysis. Eleven reference sequences covering 9 different genetic groups (Liu et al. 2012; Ellango et al. 2015; Kumar et al. 2016a, 2016b; Khatun et al. 2018) downloaded from the NCBI database were used along with the 58 consensus sequences generated from the current study. *Bemisia afer* and *Trialeurodes vaporarorium* species was used as an outgroup in the alignment (Ashfaq et al. 2014; Khatun et al. 2018). The aligned sequences were utilized to construct a Maximum likelihood tree via default neighbour-joining method based on the Jukes-Cantor model of evolution (Jukes and Cantor 1969) with the bootstrap resampling of 1000 times(Gueguen et al. 2010; Kumar et al. 2016a, 2016b). Automatic Barcode Gap Discovery (ABGD) was performed to delimit genetic clusters by detecting the significant gap in the pairwise distance distribution (Puillandre et al. 2012). Used web version of ABGD (https://bioinfo.mnhn.fr/abi/public/abgd/abgdweb.html) with default settings and K2P, JC69, and P distance model.

Population genetic analysis

The number of haplotypes, haplotype group, haplotype diversity (Hd), nucleotide diversity (pi), polymorphic sites and genetic differentiation (FST) analysis was performed by the software package DnaSP v6.10.01 (Rozas et al. 2017). Tajima's D (Tajima 1989), Fu and Li's F and Fu and Li's D neutrality tests were performed to detect the deviation from the neutral model of evolution and mismatch distribution analyses to infer the long term demographic history of populations using DnaSP v6.10.01.

The Analysis of Molecular Variance (AMOVA) test appraising the genetic variations among the populations were performed with Arlequin version 3.5 (Excoffier and Lischer 2010). The levels of genetic differentiation was categorized as great differentiation (FST = 0.25), moderate differentiation (FST = 0.15–0.25) and negligible differentiation (FST = 0.05) (Wright 1978). The TCS haplotype network constructed in PopART version 1.7 (Clement et al. 2002) package. The correlation between the genetic distance (FST) and geographic distance (km) obtained by Mantel test in Arlequin v3.5.

Results

Phylogenetic analysis

The phylogenetic tree for 58 Mt COI sequences of *B. tabaci* containing 425 nucleotides each was constructed by using *Bemisia afer* and *Trialeurodes vaporarorium* as an outgroup in MEGA 7 software. The maximum likelihood phylogenetic tree resulted in two monophyletic clades represents Asia I belongs to the south and central India and Asia II1 represents north Indian whitefly populations. The tree constructs supported moderately to strong with bootstrap value range from 50 to 93% (Fig. 2 Maximum-likelihood phylogenetic tree of mt COI gene sequences of *B. tabaci.* Various colours in phylogenetic tree indicate different sequences obtained in this study. Brown-Asia II1 genetic group, Blue-Asia I genetic group, Green-reference sequences and Red-outgroups).

Table 1 Geographic details of cotton Whiteflies collection in India

State	Locations	Districts	Sample code	Latitude	Longitude	NCBI Accession number
North India						
Haryana (HR)	Dhangar	Fatehabad	FB1	29° 37′ 58.6”	75° 30′ 49.4”	MG448545
	Chikanwas	Hisar	HS1	29° 16′ 40.4”	75° 38′ 17.6”	MG448546
	Badopal	Fatehabad	FB2	29° 25′ 57.7”	75° 32′ 14.3”	MG448547
	Mallekan	Sirsa	SS1	29° 25′ 57.1”	74° 53′ 36”	MG448548
	Keshupur	Sirsa	SS2	29° 26′ 44.3”	74° 51′ 42.6”	MG448549
	Choti mameran	Sirsa	SS3	29° 27′ 4.3”	74° 42′ 33.2”	MG448550
	Nezadella	Sirsa	SS4	29° 22′ 1.6”	75° 35′ 3.9”	MG448551
	Farwain	Sirsa	SS5	29° 34′ 25.4”	75° 4′ 33.4”	MG448552
	Fatehabad	Fatehabad	HS2	29° 29′ 37.8”	75° 28′ 50.8”	MG448554
	Richpura	Hansi	HN1	29° 5′ 39.5”	75° 59′ 59.1”	MG448555
	Dhana Kalan	Hansi	HN2	29° 4′ 52.4”	76° 1′ 41.1”	MG448556
	Rakhi Shahpur	Hisar	HS3	29° 7′ 6.7”	76° 6′ 21.5”	MG448557
	Panihari	Hisar	HS4	29° 20′ 39.2”	76° 2′ 58.3”	MG448558
	Moju Khera	Sirsa	SS7	29° 27′ 5.1”	74° 43′ 2.4”	MG448559
	Keshupura	Sirsa	SS2	29° 26′ 41.3”	74° 51′ 48.6”	MG448560
	Kuthabad	Sirsa	SS9	29° 27′ 15.1”	74° 47′ 51”	MG448561
	Surtia	Sirsa	SS10	28° 48′ 37.6”	75° 11′ 19.8”	MG448562
	Dariyapur	Fatehabad	FB3	29° 31′ 11.1”	75° 28′ 50.8”	MG448563
	Agroha	Hisar	HS5	29° 17′ 52.1”	75° 38′ 46.4”	MG448564
	Mayar	Hisar	HS6	28° 6′ 9.7”	75° 52′ 13.1”	MG448565
	Kotfatta	Bathinda	BT1	30° 6′ 25.5”	75° 3′ 45.6”	MG448566
Punjab (PB)	Balluana	Bathinda	BT2	30° 13′ 53.4”	74° 47′ 41.7”	MG448567
	Abohar	Fazilka	FZ1	30° 9′ 49.1”	74° 15′ 12”	MG448568
	Dangar khera	Fazilka	FZ2	30° 12′ 47.3”	74° 9′ 10.6”	MG448569
	Banawala	Fazilka	FZ3	30° 20′ 24.7”	74° 4′ 11.2”	MG448570
	Mann	Muktsar	MR1	30° 5′ 18.2”	74° 39′ 7.4”	MG448571
	Baluana Thakar Singh	Fazilka	FZ4	30° 13′ 53.4”	74° 47′ 41.7”	MG448572
	Chanartal Kalan	Fatehgarh Sahib	FS1	30° 7′ 44.6”	75° 8′ 1.8”	MG448573
	Sardulgarh	Mansa	SS6	29° 54′ 25.4”	75° 4′ 33.3”	MG448574
	Tibbi Hari Singh	Mansa	MS1	29° 59′ 12.8”	75° 7′ 31.6”	MG448575
	Sardulewala	Mansa	MS2	29° 29′ 1.8”	75° 1′ 8.5”	MG448576
	Lahri	Bathinda	BT3	29° 51′ 52.6”	75° 9′ 17.1”	MG448577
	Talawandi Sabo	Bathinda	BT4	29° 59′ 12.8”	75° 7′ 31.6”	MG448578
Rajasthan (RJ)	Rathikhera	Hanumangarh	HG1	29° 32′ 37.7”	74° 32′ 45.9”	MG448579
	Tibbi	Hanumangarh	HG21	29° 32′ 41.7”	74° 30′ 26.7”	MG448580
	Zandawali	Hanumangarh	HG31	29° 38′ 3”	74° 14′ 45.6”	MG448581
	Hindumalkot	Sri Ganganagar	SG11	30° 8′ 36.1”	73° 54′ 35”	MG448582
	Khanka	Sri Ganganagar	SG21	30° 2′ 48.9”	73° 52′ 22”	MG448583
	Sri Ganganagar	Sri Ganganagar	SG31	29° 56′ 3.9”	73° 53′ 22”	MG448584
	Tibbi	Hanumangarh	HG2	29° 33′ 28.7”	74° 25′ 38.1”	MG448585
	Zandawali	Hanumangarh	HG3	29° 38′ 3”	74° 14′ 45.6”	MG448586
	Khanka	Sri Ganganagar	SG2	30° 2′ 48.9”	73° 52′ 22”	MG448587
	Hindumalkot	Sri Ganganagar	SG1	30° 8′ 34.9”	73° 57′ 35.3”	MG448588
	Sri Ganganagar	Sri Ganganagar	SG3	29° 56′ 3.9”	73° 53′ 22”	MG448589
Central India						
Maharashtra (MH)	Jalgaon	Jalgaon	JG1	21° 2′ 58.7”	75° 31′ 56.4”	MG448590
	Akola	Akola	AK1	20° 42′ 4.4”	77° 1′ 11.6”	MG448591

Table 1 (continued)

State	Locations	Districts	Sample code	Latitude	Longitude	NCBI Accession number
	Nandurbar	Nandurbar	NB1	21° 20′ 56.1”	74° 32′ 51.3”	MG448592
	Dhule	Dhule	DL1	20° 53′ 31.7”	74° 51′ 19.9”	MG448593
	Nagpur	Nagpur	NG1	21° 2′ 22.7”	79° 3′ 36.2”	MG448594
South India						
Telangana (TA)	Adilabad	Adilabad	ADB1	19° 40′ 30”	78° 28′ 19”	MG448595
	Warangal	Warangal	WR1	17° 44′ 16”	79° 35′ 42”	MG448596
	Mahabubabad	Mahabubabad	WR2	17° 27′ 11”	79° 48′ 50”	MG448597
	Mahabubabad	Mahabubabad	MBB1	17° 24′ 41”	79° 52′ 13”	MG448598
Andhra Pradesh (AP)	Guntur	Guntur	GT1	16° 12′ 24.4”	80° 12′ 45.8”	MG448599
	Prakasam	Prakasam	PK1	15° 56′ 26”	80° 17′ 1.82”	MG448600
	Krishna,	Krishna,	KS1	16° 69′ 9.68”	80° 36′ 1.82”	MG448601
	Krishna	Krishna	KS2	16° 45′ 20”	80° 17′ 44”	MG448602

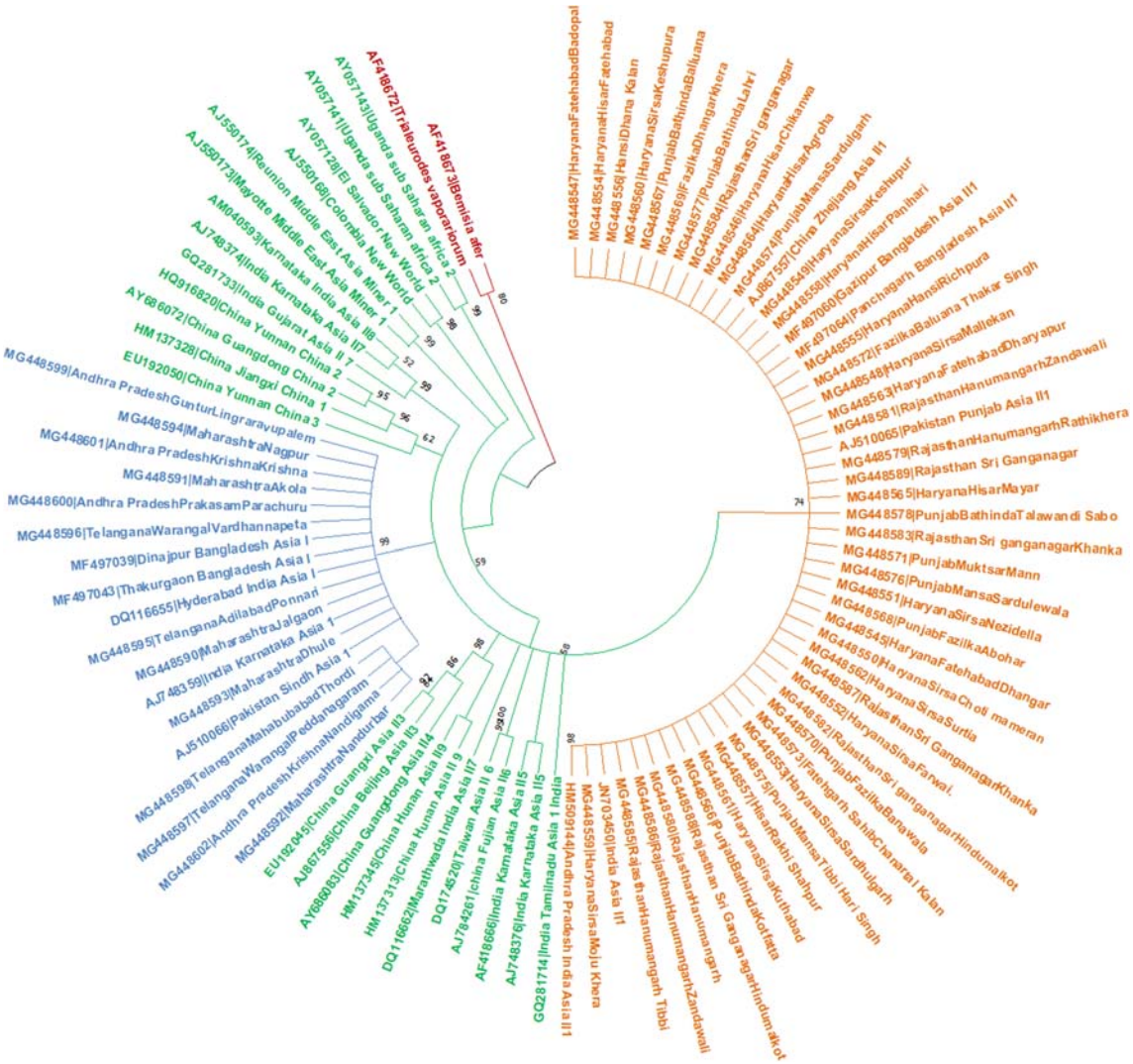

Fig. 2 Maximum-likelihood phylogenetic tree of mt COI gene sequences of *B. tabaci*. Various colours in phy-logenetic tree indicate different sequences obtained in this study. Brown-Asia II1 genetic group, Blue-Asia I genetic group, Green-reference sequences and Red-outgroups

Automatic barcode gap discovery (ABGD) analysis

The ABGD analysis results showed clear gab between the intraspecific and interspecific K2P distances, with a majority of the intraspecific distances falling well below 3% (Fig. 3 Automatic Barcode Gap Discovery (ABGD) analysis for CO I sequences. a) Distribution of pairwise differences, b) Ranked pairwise differences, c) Automatic partition results for

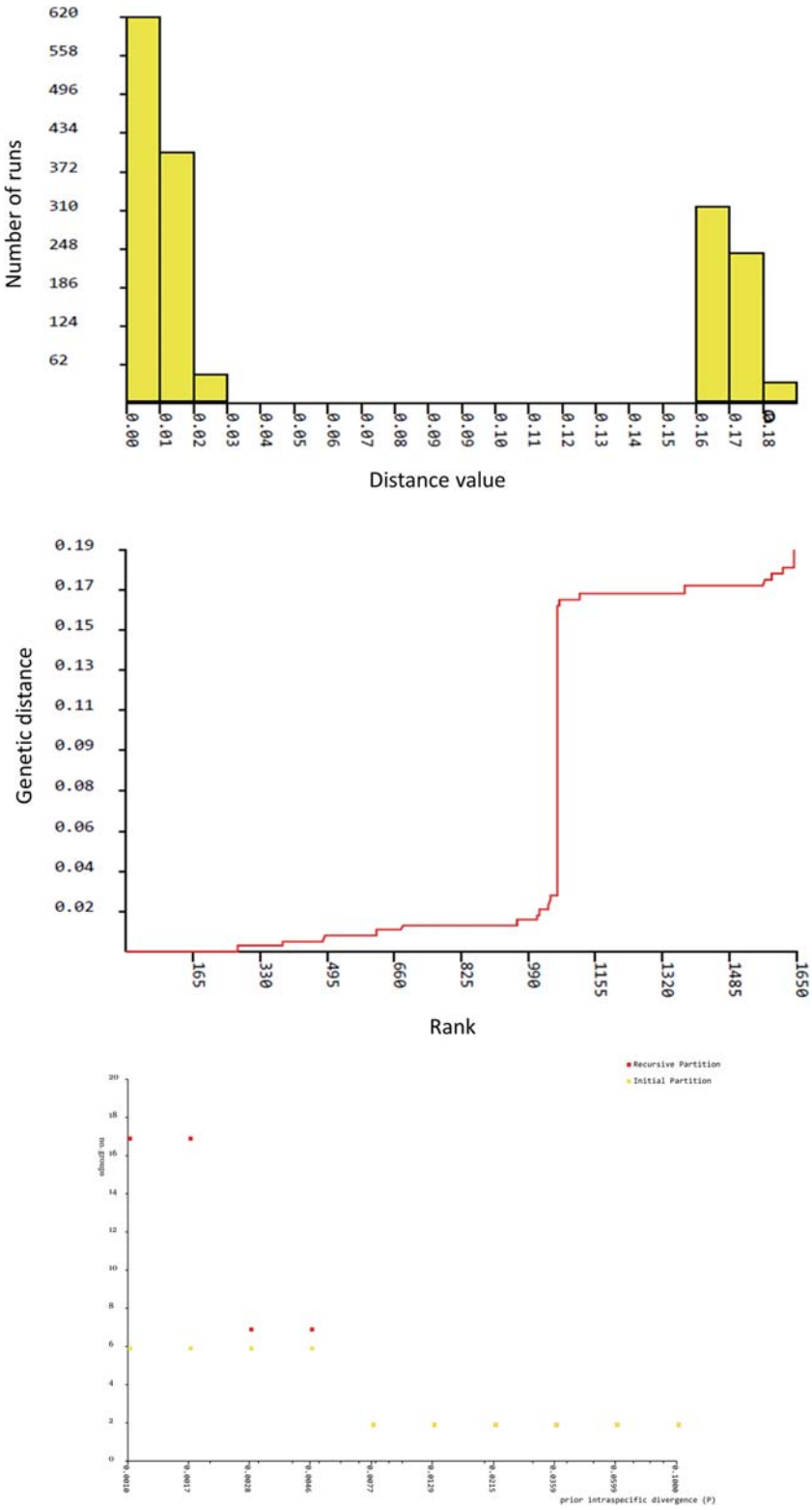

Fig. 3 Automatic Barcode Gap Discovery (ABGD) analysis for CO I sequences. **a**) Distribution of pairwise differences, **b**) Ranked pairwise differences, **c**) Automatic partition results for *B. tabaci*

B. tabaci). Both initial partition and recursive partition were employed to partition the dataset and it partitioned the data from 17 (prior maximal distance $P = 0.001000$) groups to 2 groups (prior maximal distance $P = 0.007743$), which also showed that the initial partition is more stable and divided the data set into 2 partitions at the value of $P = 0.0077$ of prior intraspecific divergence. The same number of groups or genetic groups delimited by MEGA phylogenetic analysis.

Haplotype diversity and distribution

The alignment of good quality sequences (425 bp) revealed 72 nucleotide polymorphic sites. Among the different population of *B. tabaci* haplotype diversity ranged from 0.5 to 0.9 and nucleotide diversity ranged from 0.00118 to 0.00808 (Table 2). Fourteen haplotypes resulted from 58 partial mt COI sequences of whitefly. Two dominant haplotypes namely haplotype 1 and haplotype 2 were grouped by 20 & 15 sequences majorly represented by whiteflies locations of Punjab, Haryana and Rajasthan states of North India. Haplotype 11 and 12 shared by 6 and 4 sequences respectively and represented Central and South India. Haplotype 6 and 8 shared by 3 and 2 sequences respectively and represented by north India.

Haplotype network

PopART analysis revealed two distinct haplogroups in TCS haplotype network (Clement et al. 2002) with the sufficient number of mutations in north and south-central group. In north India two dominant haplotype groups forming two distinctive circles sharing, PB, HR and RJ *B. tabaci* populations. North and south-central haplogroups are separated into two with 57 mutation steps (Fig. 4 The TCS haplotype network tree for the mitochondrial COI region of *B.tabaci* Circles represents the haplotypes identified and the size of each circle are proportional to the frequency of the haplotypes. The lines between each haplotype represent the mutations, each line represents a single mutation. The black colour of circles in a network represents imaginary haplotypes).

Genetic divergence

Nucleotide identity shared between the cotton whitefly populations of north India (PB, HR and RJ) were 97.9–100% with 0–2.1% genetic divergence. Populations from the south and central India together shared 98.8–100% identity with a minimal genetic divergence of 0–1.2%. The two major clades separate north and south-central Indian whitefly population exhibit divergence of 16.8–19.2% with the 84.7–85.2% identity between them. The higher genetic divergence revealed between the whitefly populations of northern and south-central states of India might be attributed to greater geographic distance and cropping pattern prevailing in that area.

Genetic differentiation and geographic distance

Genetic differentiation (FST) ranged from 0.018 to 0.976 between the geographically distinct populations of *B. tabaci.* Out of 15 populations, 9 populations found significant pairwise FST values (Table 3). The highest pairwise FST values (0.976) were found between the two geographically divergent populations of northern and south-central India. Comparison between geographically farthest locations resulted in highest pairwise FST value viz., 0.976, 0.975, 0.970 and 0.959 between AP and RJ, AP and PB, MH and PB and TA and RJ, respectively. Lowest FST value (≃0) obtained for the populations of nearest geographical locations depicted the least genetic differentiation (Table 3). The significant positive correlation found between the genetic and geographic distances of *B. tabaci* populations in Mantel test analysis.

Analysis of molecular variance (AMOVA) was also performed to distinguish variations between and among the population groups by using Arlequin 3.5 (Excoffier et al. 2005). Results revealed no significant molecular variations (4.62%, $p = 0.001$) within populations, though

Table 2 Haplotype and nucleotide diversity of cotton whitefly population in India

Location	No. of sequences	K	Hd	Pi	Haplotype	S
North India	45	2.502	0.698	0.00593	9	13
Punjab (PB)	13	2.615	0.731	0.00615	4	5
Haryana (HR)	21	3.433	0.695	0.00808	6	12
Rajasthan (RJ)	11	2.473	0.709	0.00586	5	7
Central India	5	1.2	0.9	0.00282	4	3
Maharashtra (MH)	5	1.2	0.9	0.00282	4	3
South India	8	1.536	0.679	0.00363	3	5
Andhra Pradesh (AP)	4	0.5	0.5	0.00118	2	1
Telangana (TA)	4	2.667	0.833	0.00636	3	5
Total	58	23.564	0.806	0.05611	14	72

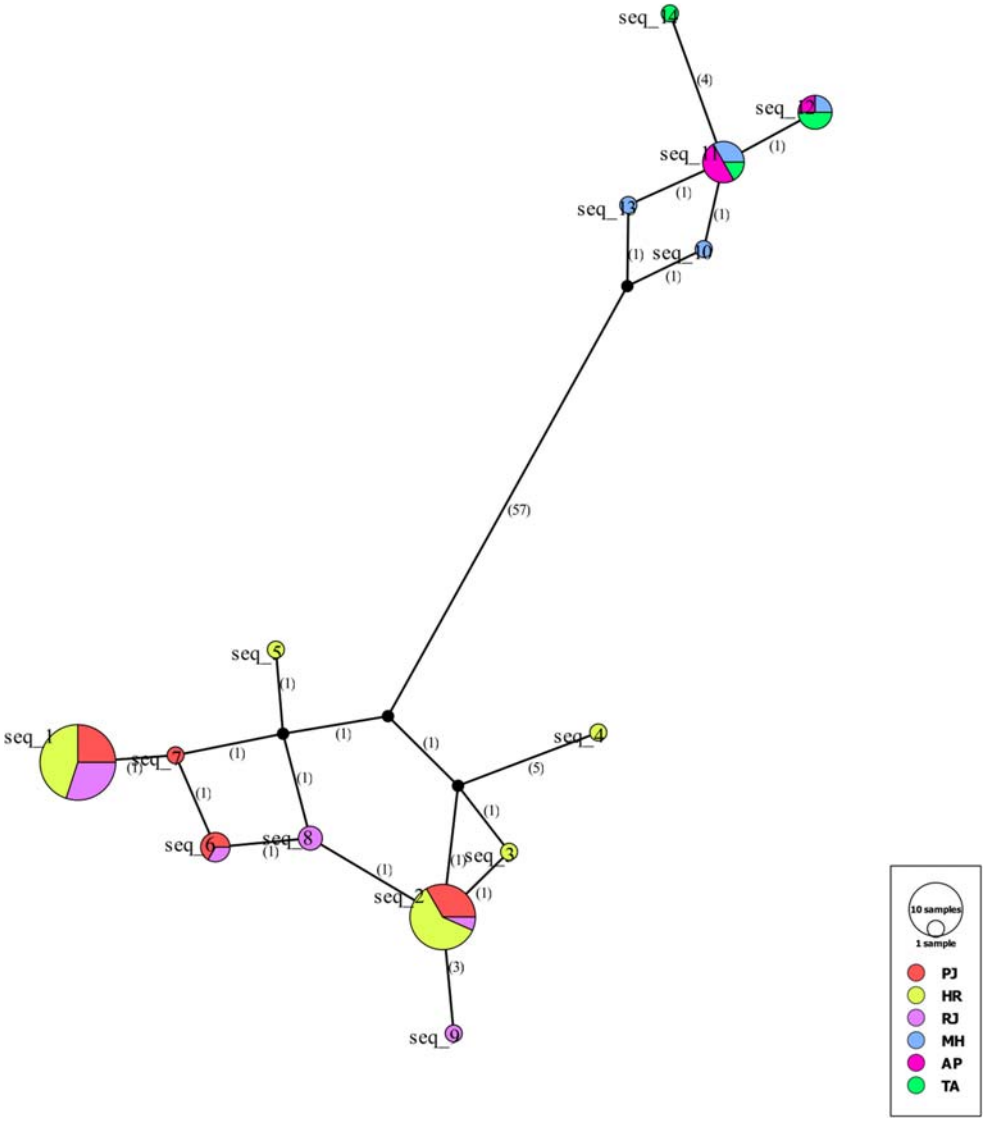

Fig. 4 The TCS haplotype network tree for the mitochondrial COI region of *B. tabaci* Circles represents the haplotypes identified and the size of the each circle are proportional to the frequency of the haplotypes. The lines between each haplotype represents the mutations, each line represents single mutation. Black colour of circles in a network represents imaginary haplotypes

significant variations exhibited among the groups of population (95.38%; $p = 0.001$) (Table 4).

Demographic history

Neutrality tests viz., Tajima's D, Fu and Li's D and Fu and Li's F were performed to analyse the neutral evolution theory. These neutrality test parameters provide information about demographic histories. All the three neutrality tests resulted in non-significant negative values for Haryana, Rajasthan (Except Tajima D test), Maharashtra, Telangana and Andhra Pradesh states. All neutrality tests resulted in non-significant negative values for central and south India population except north India, where, only Fu and Li's test showed significant negative value. In North India, among Punjab, Haryana and Rajasthan only for the Punjab whitefly populations, all the neutrality tests showed significantly positive values in Tajima's D test and Fu and Li's F (Table 5).

Discussion

Whitefly *B. tabaci* is one of the most important sucking pests of cotton in India, its species composition and distribution on cotton in India is unexplored. The current study is the first of its kind to explore the species composition and diversity within *B. tabaci* species complex on cotton in India. Such studies are critical for the development of sustainable management strategies for whitefly and its transmitted viral disease. The mt COI sequence-based phylogenetic analysis, pairwise sequence divergence, haplotype network analysis, ABGD analysis confirmed the presence of Asia I and Asia II 1 genetic group on cotton. Cotton being a prominent economical fibre crop in India growing in more than 12 mha area, supporting livelihoods of various sections of the country. Since the crop is being cultivated in wider geographic and climatic conditions, the existence of genetic groups/putative species of *B. tabaci* on different host species is obvious, owing to the

Table 3 AMOVA haplotype F_{ST} results for pairwise population comparison, matrix of the geographic distances (Crow flying distance in km; above diagonal) and the genetic differentiation (F_{ST}; below diagonal)

TA		AP	MH	PB	RJ	HR
TA	0	580	518	2186	2210	2065
AP	−0.05556	0	929	2197	2222	2077
MH	−0.01754	−0.13333	0	1412	1437	1291
PB	0.95945*	0.97581*	0.97019*	0	167	123
RJ	0.95917*	0.97610*	0.97028*	−0.04263	0	148
HR	0.95327*	0.96952*	0.96387*	−0.03335	0.01806	0

*Significance $P < 0.05$

acclimatization and evolution to the prevailing condition. Whitefly sampling done in all the three cotton growing regions, however, the sampling was comparatively more in Northern India than in South and Central India because of lower/no incidence of whitefly on cotton in central and south India as compared to north India. The populations within the northern cotton-growing region and south-central regions are genetically identical with minimum divergence but there is 16.8–19.2% genetic divergence between the two major clades representing north, and south-central region. The results of phylogeny in delimit the genetic group has been successfully validated by ABGD analysis (Fig. 3). Our results are in congruence with the reports on delimiting the cryptic species of sucking insects (Ashfaq et al. 2014; Tyagi et al. 2017; Li et al. 2020). In supplement to this, the phylogeographic structure of the *B. tabaci* population is isolated by distance (R2 = 0.852, $n = 15$) and which is confirmed by strong positive correlations of FST values with geographic distance in Mantel test. These results of the study are in support with similar studies elsewhere (Frohlich et al. 1999; Brown 2000; Legg et al. 2002). Old-world whitefly and new world whiteflies clade separated with 15–22% divergence were reported by Berry et al. (2004). The occurrence of such genetic variations in populations might be attributed to geographical location, the existence of the climatic condition, migratory behaviour, human interventions etc. (Hampe and Petit 2005; Bridle and Vines 2007; Kellermann et al. 2009).

B. tabaci is found to be polyphagous and reported all over the country on different host crops and perpetuates throughout the year. The force of evolution might have acted on different populations of *B. tabaci* differently in different geographical regions which might be the reason for the existence of more than 11 genetic groups/putative species (Asia I, Asia I-India, Asia II-1, Asia II-5, Asia II-7, Asia II-8, and Asia II-11, Middle East Asia Minor (MEAM) -1 and Middle East Asia Minor (MEAM) - K genetic groups) in India (Banks et al. 2001; Chowda-Reddy et al. 2012; Ellango et al. 2015; Kumar et al. 2016a, 2016b). Maximum likelihood phylogenetic analysis revealed the presence of Asia II1 genetic group in a northern cotton-growing region and Asia I in a south-central cotton-growing region, these results supported by the previous reports on different host crops of whiteflies in India (Chowda-Reddy et al. 2012; Ellango et al. 2015; Kumar et al. 2016a, 2016b). Fortunately, in our studied populations, one of the most dangerous globally invasive *B. tabaci* genetic group MEAM-1 (Díaz et al. 2015; and Hu et al., 2011, b) is not observed on cotton though it is present on tomato in India (Banks et al. 2001). However, Asia II1 is also equally dangerous because in neighbouring country Pakistan the Asia II 1 was previously restricted to Punjab, but now it is a dominant genetic group and also spreading to the southern part of the country (Ashfaq et al. 2014). Also, there are reports of high association of Asia II1 genetic group with the CLCuD (Ahmed et al. 2011; Rana et al. 2012), our study also recorded the same results where the presence of Asia II1 genetic group with the most dreaded viral disease CLCuD on cotton was found only in northern India populations. It is hypothesised that the geographical distribution of virus genotypes and vector genotypes is highly correlated (Brown, 2000) but this aspect needs to be explored experimentally. Apart from the association of Asia II1 genetic group with CLCuD it is dominating on other genetic groups of whitefly and expanding to southern Pakistan (Ashfaq et al. 2014), similarly in India the reports reveal that the Asia II1 is also the dominant genetic group in soybean (Prasanna et al. 2015) and is noticed in most of the soybean growing areas in India. This dominating nature of Asia II1 genetic group is worrisome, as it can increase its distribution by expanding its population on cotton by replacing previously existed genetic group like its already been witnessed in the neighbouring country Pakistan (Ashfaq

Table 4 AMOVA for Mt CO I sequences of *B. tabaci* in India

Source of variation	df	Sum of squares	Variance components	Percent variation	*P* value
Among groups	2	631.585	29.17529	95.38	0.06256
Among populations within groups	3	4.184	−0.00181	−0.01	0.37537
Within populations	52	73.541	1.41426	4.62	0.00000 + −0.0000

Table 5 Results of Tajima's *D* test, Fu and Li's *D* test and Fu and Li's *F* for COI gene in populations of *Bemisia tabaci*

Location	Tajima's D	Significance	Fu and Li's D	Significance	Fu and Li's F	Significance
North India	−0.67521	NS, $P > 0.10$	−2.69144	*, $P < 0.05$	−2.38644	NS, $0.10 > P > 0.05$
Central India	−1.04849	NS, $P > 0.10$	−1.04849	NS, $P > 0.10$	−1.05189	NS, $P > 0.10$
South India	−0.92337	NS, $P > 0.10$	−1.11466	NS, $P > 0.10$	−1.17916	NS, $P > 0.10$
Punjab (PB)	2.2219	*, $P < 0.05$	1.24617	NS, $P > 0.10$	1.70404	*, $P < 0.05$
Haryana (HR)	−0.17852	NS, $P > 0.10$	−1.25372	NS, $P > 0.10$	−1.09083	NS, $P > 0.10$
Rajasthan (RJ)	0.1406	NS, $P > 0.10$	−0.19261	NS, $P > 0.10$	−0.1228	NS, $P > 0.10$
Maharashtra (MH)	−1.04849	NS, $P > 0.10$	−1.04849	NS, $P > 0.10$	−1.05189	NS, $P > 0.10$
Andhra Pradesh (AP)	−0.61237	NS, $P > 0.10$	−0.61237	NS, $P > 0.10$	−0.47871	NS, $P > 0.10$
Telangana (TA)	−0.21249	NS, $P > 0.10$	−0.21249	NS, $P > 0.10$	−0.2008	NS, $P > 0.10$
Over all	1.44078	NS, $P > 0.10$	0.74487	NS, $P > 0.10$	1.20834	NS, $P > 0.10$

*, $P < 0.05$, **, $P > 0.10$

et al. 2014). The results of the neutrality tests revealed non-significant negative values of neutrality tests viz., Tajima's D, Fu and Li's F and Fu and Li's D tests which indicates probable recent population expansion in all the studied states except Punjab state (Table 5). However, in our study, the Asia II1 genetic group is not reported on cotton in the south and central region and also there are no reports of CLCuD in the respective regions. The population expansion of Asia II1 through the displacement of Asia I genetic group on cotton in south and central regions is needed to be monitored, to avoid the similar consequences happened in tomato where Asia I and Asia II7 genetic group replaced by dangerous MEAM-1 genetic group (Banks et al. 2001). Hence, monitoring and surveillance of specific genetic groups of whiteflies for the factor such as fecundity, survival ability, insecticide resistance and susceptibility to the biological control agents for its population build-up and the possibility of displacement of the genetic group should be studied (Legg et al. 2002; McKenzie et al. 2012, Chowda-Reddy et al. 2012).

Conclusion

Population structure in many insect species is affected by various factors viz., environmental and ecological factors, alternative host plants, migration, natural barriers, and human. The diversity and distribution of species are also affected by these factors, thus, the study of diversity, distribution and changes in the population structure is very important. Such studies can be done through population genetics and phylogenetic analysis and which is executed in our study for the most important pest *B. tabaci* of the important commercial fibre crop in India. The present study revealed the diversity and the distribution of genetic groups on cotton in India. The identification and mapping the distribution of genetic groups in cotton ecosystems paved the way for the further studies on understanding the factors contributing dominance of genetic groups in a particular region and its probable dispersion for timely development and utilization of management strategies directed towards reducing the economic damage caused by *B. tabaci* and the virus (CLCuD) it transmits.

Funding Indian Council of Agricultural Research (ICAR)-Central Institute for Cotton Research (CICR), Nagpur, and National Initiative on Climate Resilient Agriculture (NICRA), India.

Data availability The authors declare that all the data used in the present study are original and generated from the study.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethics approval Not applicable, as the research didn't involve human/animals.

Consent to participate The authors are equally contributed to the present study.

Consent for publication All the authors of the manuscript declare their consent for publication.

Code availability MEGA 7, PopART, DnaSP v6.10.01 and Arlequin v3.5 software were used for the analysis.

References

Ahmed MZ, De Barro PJ, Greeff JM, Ren S-X, Naveed M et al (2011) Genetic identity of the *Bemisia tabaci* species complex and association with high cotton leaf curl disease (CLCuD) incidence in Pakistan. Pest Manag Sci 67:307–317

AICCIP (2016–17) all India coordinated cotton improvement project report http://www.aiccip.cicr.org.in/CD_16_17/3_PC_Report.pdf

Ashfaq M, Hebert PDN, Mirza MS, Khan AM, Mansoor S, Shah GS, Zafar Y (2014) DNA barcoding of *Bemisia tabaci* Complex (Hemiptera: Aleyrodidae) reveals southerly expansion of the dominant whitefly species on cotton in Pakistan. PLoS One 9(8): e104485. https://doi.org/10.1371/journal.pone.0104485

Banks GK, Colvin J, Chowda RV, Maruthi MN, Muniyappa V, Venkatesh HM, Kiran Kumar M, Padma AS, Beitia FJ, Seal SE (2001) First report of the *Bemisia tabaci* B biotype in India and an associated tomato leaf curl virus disease epidemic. Plant Dis 85:231

Berry SD, Fondong VN, Rey C, Rogan D, Fauquet CM, Brown JK (2004) Molecular evidence for five distinct *Bemisia tabaci* (Homoptera: Aleyrodidae) geographic haplotypes associated with cassava plants in sub-Saharan Africa. Ann Entomol Soc Am 97: 852–859

Briddon WR (2003) Cotton leaf curl disease, a multi-component begomovirus complex. Mol Plant Pathol 4:427–434

Bridle JR, Vines TH (2007) Limits to evolution at range margins: when and why does adaptation fail? Trends Ecol Evol 22:140–147

Brown JK (2000) Molecular markers for the identification and global tracking of whitefly vector-begomovirus complexes. Virus Res 71: 233–260

Byrne DN, Bellows TS Jr (1991) Whitefly biology. Annu Rev Entomol 36:431–457

Cahill M, Denholm I, Ross G, Gorman K, Johnston D (1996) Relationship between bioassay data and the simulated field performance of insecticides against susceptible and resistant *Bemisia tabaci*. Bull Entomol Res 86:109–116

Chaubey R, Andrew RJ, Naveen NC, Rajagopal R, Ahmad B, Ramamurthy VV (2015) Morphometric analysis of three putative species of *Bemisia tabaci* (Hemiptera: Aleyrodidae) species complex from India. Ann Entomol Soc Am 108(4):600–612

Chowda-Reddy RV, Kirankumar M, Seal SE, Muniyappa V, Valand GB, Govindappa MR, Colvin J (2012) *Bemisia tabaci* phylogenetic groups in India and the relative transmission efficacy of tomato leaf curl Bangalore virus by an indigenous and an exotic population. J Integr Agric 11:235–248

Clement M, Snell Q, Walker P, Posada D, Crandall K (2002) TCS: Estimating gene genealogies. Parallel and Distributed Processing Symposium. Int Proc 2:84

Díaz F, Endersby NM, Hoffmann AA (2015) Genetic structure of the whitefly *Bemisia tabaci* populations in Colombia following a recent invasion. Insect Sci 22(4):483–494

Dinsdale A, Cook L, Riginos C, Buckley YM, De Barro PJ (2010) Refined global analysis of *Bemisia tabaci* (Gennadius) (Hemiptera: Sternorrhyncha: Aleyrodoidea) mitochondrial CO1 to identify species level genetic boundaries. Ann Entomol Soc Am 103:196–208

Ellango R, Singh ST, Rana VS, Priya G, Raina H, Chaubey R, Naveen NC, Mahmood R, Ramamurthy VV, Asokan R, Rajagopal R (2015) Distribution of *Bemisia tabaci* genetic groups in Indian. Environ Entomol 44(4):1258–1264

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and windows. Mol Ecol Resour 10:564–567

Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evol Bioinformatics Online 1:47–50

Frohlich DH, Torres-Jerez I, Bedford ID, Markham G, Brown JK (1999) A phylogeographical analysis of the *Bemisia tabaci* species complex based on mitochondrial DNA markers. Mol Ecol 8:1683–1691

Gill RJ, Brown JK (2010) Systematics of Bemisia and Bemisia relatives: can molecular techniques solve the *Bemisia tabaci* complex conundrum-a taxonomist's viewpoint. In: Stansly PA, Naranjo SE (eds) Bemisia: bionomics and management of a global pest. Springer, New York, pp 5–29

Gueguen G, Vavre F, Gnankine O, Peterschmitt M, Charif D, Chiel E, Gottlieb Y, Ghanim M, Fein EZ, Fleury F (2010) Endosymbiontmetacommunities, mtDNA diversity and the evolution of the Bemisiatabaci (Hemiptera: Aleyrodidae) species complex. Mol Ecol 19:4365–4376. https://doi.org/10.1111/j.1365-294X.2010.04775.x

Hampe A, Petit RJ (2005) Conserving biodiversity under climate change: the rear edge matters. Ecol Lett 8:461–467

Hsieh CH, Wang CH, Ko CC (2006) Analysis of *Bemisia tabaci* (Hemiptera: Aleyrodidae) species complex and distribution in eastern Asia based on mitochondrial DNA markers. Ann Entomol Soc Am 99:768–775

Hu J, De Barro P, Zhao H, Wang J, Nardi F, Liu SS (2011) An extensive field survey combined with a phylogenetic analysis reveals rapid and widespread invasion of two alien whiteflies in China. PLoS One 6:e16061

Jones RD (2003) Plant viruses transmitted by whiteflies. Eur J Plant Pathol 109:195–219

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian protein metabolism. Academic Press, New York, pp 21–132

Kellermann V, van Heerwaarden B, Sgrò CM, Hoffmann AA (2009) Fundamental evolutionary limits in ecological traits drive Drosophila species distributions. Science 325:1244–1246

Khatun MF, Hemayet Jahan SM, Lee S, Lee K-Y (2018) Genetic diversity and geographic distribution of the *Bemisia tabaci* species complex in Bangladesh. Acta Trop 187:28–36

Kranthi KR (2015) Whitefly –the black story. Cotton Stat News 23:1–3

Kranthi KR, Jadhav DR, Kranthi S, Wanjari RR, Ali SS, Russell DA (2002) Insecticide resistance in five major insect pests of cotton in India. Crop Prot 21:449–460

Kumar S, Stecher G, Tamura K (2016a) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33(7):1870–1874

Kumar NR, Chang JC, Narayanan MB, Ramasamy S (2016b) Phylogeographical structure in mitochondrial DNA of whitefly, *Bemisia tabaci* Gennadius (Hemiptera: Aleyrodidae) in southern India and Southeast Asia. Mitochondrial DNA Part A 28:621–631. https://doi.org/10.3109/24701394.2016.1160075

Legg J, French R, Rogan D, Okao-Okuja G, Brown JK (2002) A distinct *Bemisia tabaci* (Gennadius) (Hemiptera: Sternorrhyncha: Aleyrodidae) genotype cluster is associated with the epidemic of severe cassava mosaic virus disease in Uganda. Mol Ecol 11: 1219–1229

Li Q, Deng J, Chen C, Zeng L, Lin X, Cheng Z, Qiao G, Huang X (2020) DNA barcoding subtropical aphids and implications for population differentiation. Insects 11:11

Liu S-s, Colvin J, De Barro PJ (2012) Species Concepts as Applied to the Whitefly *Bemisia tabaci* Systematics: How Many Species Are There? J Integr Agric 11(2):176–186

McKenzie CL, Bethke JA, Byrne FJ, Chamberlin JR, Dennehy TJ, Dickey AM, Gilrein D, Hall PM, Ludwig S, Oetting RD, Osborne LS, Schmale L, Shatters RG Jr (2012) Distribution of *Bemisia tabaci* (Hemiptera: Aleyrodidae) biotypes in North America after the Q invasion. J Econ Entomol 105:753–766

Misra CS, Lambda SK (1929) The cotton whitefly (*Bemisia gossypiperda sp.*). Bulletin of agricultural research, Inst. Pusa 196:1–7

Naveed M, Salam A, Saleem MA (2007) Contribution of cultivated crops, vegetable weeds and ornamental plants in harbouring of *Bemisia tabaci* (Homoptera: Aleyrodidae) and associated parasitoids (Hymenoptera: Aphelinidae). J Pest Sci 80:191–197

Naveen NC, Chaubey R, Kumar D, Rebijith KB, Rajagopal R, Subrahmanyam B, Subramaniana S (2017) Insecticide resistance status in the whitefly, *Bemisia tabaci* genetic groups Asia-I, Asia-II-1 and Asia-II-7 on the Indian subcontinent. Sci Rep 7:40634

Prabhulinga T, Rameash K, Madhu TN, Vivek S, Suke R (2017) Maximum entropy modelling for predicting the potential distribution of cotton whitefly *Bemisia tabaci* (Gennadius) in North India. J Entomol Zool Stud 5(4):1002–1006

Prabhulinga T, Gawande SP, Kranthi S, Monga D, Kumar R, Kranthi KR (2019) Status of cotton leaf curl virus disease and its vector whitefly in North India: a survey report. J Entomol Zool Stud 7(3):183–186

Prasanna HC, Kanakala S, Archana K, Jyothsna P, Varma RK, Malathi VG (2015) Cryptic species composition and genetic diversity within *Bemisia tabaci* complex in soybean in India revealed by *mtCOI* DNA sequence. J Integr Agric 14(9):1786–1795

Puillandre N, Lambert A, Brouillet S, Achaz G (2012) ABGD, automatic barcode gap discovery for primary species delimitation. Mol Ecol 21(8):1864–1877

Rana VS, Singh ST, Priya NG, Kumar J, Rajagopal R (2012) Arsenophonus GroEL interacts with CLCuV and is localized in the midgut and salivary gland of whitefly *B. tabaci*. PLoS One 7: e42168

Roopa HK, Asokan R, Rebijith KB, Ranjitha H, Hande RM, Krishna Kumar NK (2015) Prevalence of a new genetic group, MEAM-K, of the whitefly *Bemisia tabaci* (Hemiptera: Aleyrodidae) in Karnataka, India, as evident from mt COI sequences. Fla Entomol 98(4):1062–1071

Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A (2017) DnaSP 6: DNA sequence polymorphism analysis of large datasets. Mol Biol Evol 34(12):3299–3302. https://doi.org/10.1093/molbev/msx248

Sethi A, Dilawari VK (2008) Spectrum of insecticide resistance in whitefly from upland cotton in the Indian subcontinent. J Entomol 5:138–147

Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Ann Entomol Soc Am 87:651–701

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123(3):585–595

Tyagi K, Kumar V, Singha D, Chandra K, Laskar BA, Kundu S, Chakraborty R, Chatterjee S (2017) DNA barcoding studies on Thrips in India: cryptic species and species complexes. Sci Rep 7: 4898

Wright S (1978) Evolution and the genetics of populations, Vol. 4. Variability Within and Among Natural Populations. University of Chicago Press, Chicago

Xu J, Lin KK, Liu SS (2011) Performance on different host plants of an alien and an indigenous *Bemisia tabaci* from China. J Appl Entomol 135:771–779

Bedford ID, Briddon RW, Brown JK, Rosell RC, Markham PG (1994) Geminivirus transmission and biological characterisation of *Bemisia tabaci* (Genn.) biotypes from different geographic regions. Ann Appl Biol 125:311–325