Whole genome sequencing of Asia II 1 species of whitefly reveals that genes involved in virus transmission and insecticide resistance have genetic variances between Asia II 1 and MEAM1 species

Sonia Hussain1,2†, Muhammad Farooq1†, Hassan Jamil Malik1,2, Imran Amin1, Brian E. Scheffler3, Jodi A. Scheffler4, Shu-Sheng Liu5 and Shahid Mansoor1*

Abstract

Background: Whiteflies (Bemisia tabaci) are phloem sap-sucking pests that because of their broad host range and ability to transmit viruses damage crop plants worldwide. B. tabaci are now known to be a complex of cryptic species that differ from each other in many characteristics such as mode of interaction with viruses, invasiveness, and resistance to insecticides. Asia II 1 is an indigenous species found on the Indian sub-continent and south-east Asia while the species named as Middle East Asia Minor 1 (MEAM1), likely originated from the Middle-East and has spread worldwide in recent decades. The purpose of this study is to find genomic differences between these two species.

Results: Sequencing of the nuclear genome of Asia II 1 with Illumina HiSeq and MiSeq generated 198.90 million reads that covers 88% of the reference genome. The sequence comparison with MEAM1 identified 2,327,972 SNPs and 202,479 INDELs. In total, 1294 genes were detected with high impact variants. The functional analysis revealed that some of the genes are involved in virus transmission including 4 genes in Tomato yellow leaf curl virus (TYLCV) transmission, 96 in Tomato crinivirus (ToCV) transmission, and 14 genes in insecticide resistance.

Conclusions: These genetic differences between Asia II 1 and MEAM1 may underlie the major biological differences between the two species such as virus transmission, insecticide resistance, and range of host plants. The present study provides new genomic data and information resources for Asia II 1 that will not only contribute to the species delimitation of whitefly, but also help in conceiving future research studies to develop more targeted management strategies against whitefly.

Keywords: Whitefly, Asia II 1, MEAM1, Sequencing, Virus, Insecticide
Introduction

_Bemisia tabaci_ (Hemiptera: Aleyrodidae), commonly known as ‘whiteflies’ are phloem sap sucking pests some of whose differentiation has become major constraint to important food, fiber and ornamental crops worldwide. The whiteflies can infest as many as 1000 plant species [1] and they damage host plants by infestation, but more importantly by transmitting plant viruses. These whiteflies can potentially vector over 300 plant viruses, mostly viruses in the genus _Begomovirus_ [2]. Major crops affected by _B. tabaci_-transmitted viruses on a global scale include cotton, cassava, tomato, sweet potato, cucurbits and other crop plant species.

Whiteflies (_B. tabaci_) are now known to be as a cryptic species complex, based on recent molecular phylogenetic analyses and evidence of reproductive incompatibility [3, 4]. These putative whitefly species differ in many biological aspects such as host range [1], resistance to insecticides [5, 6], specificity and capacity of virus transmission [7, 8] and composition of harbored symbionts [9]. Although the use of ≥3.5% mtCOI divergence as the criterion for species delimitation has been occasionally shown to be inadequate [10], it has been widely used to differentiate species. Based on sequence divergence of mtCOI (≥3.5% divergence), _B. tabaci_ has been deduced to include more than 39 cryptic species that are morphologically indistinguishable but genetically distinct [11–13].

The long-term association between begomoviruses and whitefly has brought some co-evolved adaptations [14] that allow them to live in equilibrium. Begomoviruses are single-stranded (ss) DNA viruses that are transmitted mostly in a persistent circulative manner. Once ingested through the stylet, these plant viruses move across the mid gut membrane and then via hemolymph translocate to salivary glands and from there these are egested while feeding [15]. In circulation of viruses, mid gut and salivary glands are the main barriers to overcome [16, 17]. Some mid gut proteins and proteins produced by endosymbionts in hemolymph are associated with circulation of viruses in whitefly. These interacting proteins are the main points which lead to the differentiation of cryptic species on the basis of specificity and capacity of virus transmission. The heat shock protein HSP70 is co-localized with _Tomato yellow leaf curl virus_ (TYLCV) coat protein within midgut epithelial cells and inhibits virus transmission [18]. Knottin-1 restricts the virus (TYLCV) amount in whitefly and thus shields the whitefly against its deleterious effects [19]. While cyclophilin B enhances the translocation of virus from mid gut to hemolymph [20]. Another protein peptidoglycan recognition protein (BtPRPG) is involved in whitefly immunity and has a potential binding site for TYLCV. Its co-localization with TYLCV is also reported within the midgut [21]. Endosymbionts which have been living in whitefly for millions of years [22] are also involved in virus transmission. Different cryptic species harbor different endosymbionts. Endosymbionts reside in bacteriocytes and some of them (e.g. _Hamiltonella_) produce GroEL homologue in the hemolymph which helps in virus circulation in whitefly.

Middle East-Asia Minor 1 (MEAM1, formerly known as “biotype B”) and Mediterranean (MED, formerly “Q biotype”) are globally important cryptic species of whitefly [23, 24] because of their invasiveness and broad host range. The two species originated in the Middle East regions, but are now reported from many regions of the world, and its presence has also been well reported in the southern Sindh region of Pakistan [25, 26]. Asia 1 and Asia II 1 are two species of whitefly indigenous to Pakistan, with Asia II 1 being the most prevalent whitefly in the central region of the country [26]. The different species of whitefly recorded from Pakistan have been shown to differ in many aspects including virus transmission, insecticide resistance, and host range. For example, MEAM1 is more efficient than Asia II 1 in transmitting _Tomato yellow leaf curl virus_ (TYLCV) [27]. In a study in Vietnam where Asia II 1 is indigenous, Asia II 1 is reported to be more efficient in transmitting _Tomato leaf curl Hainan virus_ (ToLCHnV) than that of TYLCV, while MEAM1 is more efficient in transmitting TYLCV than ToLCHnV [28]. Asia II 1 has been reported to be the most abundant species of whitefly in areas of high incidence of cotton leaf curl disease (CLCuD) in Pakistan and the western region of India. Two recent studies in China [17, 29] directly compared the transmission efficiency of begomoviruses by MEAM1, Asia II 1 and two more species, and showed that among these species Asia II 1 is the most efficient in transmitting both _Cotton leaf curl Multan virus_ (CLCuMuV) and _Tobacco curly shoot virus_ (TbCSV). Apart from differences in transmission efficiency of viruses, these species of whiteflies also differ in insecticide resistance [30] and host plant preference [31]. However, the physiological and molecular mechanisms underlying the differences between species of whitefly are yet poorly known.

Over the past several years, next generation sequencing (NGS) technology has emerged as an innovative approach to high-throughput sequencing [32], and the rapid development of this modern technology provides us an unprecedented opportunity to understand and explore numerous genetic findings, which can help to improve our research on the physiology and molecular biology of the whiteflies. These results can also provide new knowledge and concepts for the development of novel strategies and technology to manage whitefly pests and the viral disease agents they vector. In this study, our aim is to unravel some genetic information from Asia II 1 and MEAM1, the two major whitefly pests in Pakistan. First, with access to the data of MEAM1 [4], we performed high throughput sequencing of Asia II 1...
and aligned with that of MEAM1, to identify major genomic differences between the two species. We detected some high impact variants in genes (which were previously reported as differentially expressed genes) that have been predicted to be associated with virus transmission and insecticide resistance.

**Results**

**Mapping summary of nuclear genome**

Genome sequencing of Asia II 1 with Illumina HiSeq and MiSeq generated a total of 31.15 Gb of data comprising 198.90 million reads with read size 100 and 300 bp (the summary of raw data generated from each of seven libraries is given in Additional file 1). Approximately 91% of the reads passed the quality control criteria and 82 to 86% of these reads were mapped correctly to the reference genome. The available sequence from the reference genome is 615 Mb of the assessed total genome size of ~680–690 Mb as estimated in a previous study [33] using both flow cytometry and Kmer analysis. These reads covered 88% of the reference genome. The mean read length was 159 bp. The summary of the sequencing and mapping is shown in Table 1. The average depth of coverage of genome after filtration was 34X. Total length of the coding region of the reference genome is 44.43 Mb, 51% of which was covered with more than 5X depth of coverage, and 53% of the number of coding regions with 100% of length have at least 5X depth of coverage. The mean coverage of the coding region is 32X. Figure 1 displays the different number of coding regions with different lengths having at least 5X depth of coverage. Approximately 8366 coding regions have at least 5X coverage with full length genes.

**Variant statistics**

After variant calling and two times filtration with Genome Analysis Tool Kit (GATK), total number of 2,530, 451 high quality variants were discovered. Variant annotations and effect prediction through SnpEff resulted in 3,504,011 effects. Effects are greater in number than number of variants as one variant could have more than one effect. For example, one variant could be non-synonymous for one gene while being downstream to another. A variant statistics summary is given in Table 2 (raw variant calling data and the data after each filtration is provided in Additional file 2). Approximately 2,327, 972 SNPs and 202,479 INDELs were detected. In eight amplified regions ranging in size from 500 to 600 bp, there are 96 SNPs which were all validated through Sanger sequencing. The primers pairs list and validated SNPs positions are given in Additional file 4. The initial average variant rate was 1/20 bp, but that was decreased to 1/235 bp after filtration (when depth of read coverage at a variant point was increased to 30X in variant calling criterion). Variant rate also varied in different regions, the maximum variant rate recorded was 1/27 bp and minimum variant rate was 1/32,808 bp. Transition to transversion ratio is 1.71 and heterozygous to homozygous variant ratio is 0.05. In this study, insertions and deletions ranging from 1 to 100 bp were considered as INDELs. The maximum number of INDELs were 1 bp in length while lowest number of INDELs were of 14, 15, 20, 21, 23, 28, 33, 69, 89 or 100 bp in length. The distribution and types of variant effects in the whole genome are given in Table 3. According to functional effects of variants, these were distributed into three classes; silent (69.94%), missense (29.77%) and nonsense (0.29%).

Among the total estimated genes in whitefly MEAM1 (15,664), 1294 genes were found to have high impact variants in this data. These genes were selected for further analysis of ontology. The distribution and number of variants and their effects in different genic regions are given in Fig. 2. The number of genes in each class of high impact variants are also provided in Table 4.

**Gene ontology**

Coding regions that have high impact variants (1294 genes) were selected for gene ontology analysis. The Blast2GO results are shown in Fig. 3. The functions of these were classified into three groups: biological process (BP), molecular function (MF), and cellular components (CC). The greatest number of genes were associated with the BP category. IDs of genes associated with each sub

| Table 1 Mapping Summary |
|-------------------------|
| Total NGS Library       | 7 |
| Total Insert Size        | 550 |
| Sequencer               | IlluminaHiSeq2500 &MiSeq |
| Total Raw Data Generate | (HiSeq: 14GB) (MiSeq: 16 Gb) |
|                         | Total: 31.15 Gb |
| Average Coverage        | 47.34 X |
| Average Coverage After Filtration | 34.52 X |
| Total No of Reads Generate | HiSeq: 142605246 MiSeq: 56300942 |
|                         | Total: 198906188 |
| Total No of Reads Quality Passed | 181,434,767 |
| Total No of Reads Mapped | 156,293,812 (86%) |
| Total No of Reads Mapped Properly | 149,439,368 (82%) |
| Reference Genome Covered | 88% |
| Mean Read Length         | 159 bp |
category of these three functional classes are given in Additional file 3. Additional file 7 shows the associated pathways for the genes (with high impact variants), which were predicted by Blast2GO.

**Genes involved in virus transmission and insecticide resistance**

Fourteen genes of MEAM1 that were reported for potential involvement in insecticide resistance [4] and 96 genes which were reported to be associated with virus transmission [6] were expected to be high impact variants between Asia II 1 and MEAM1. In the present study, there were 15 high impact variants found in 14 genes which could potentially be involved in insecticide resistance. High impact variants include frame shift, start loss, stop gain, splice acceptor and splice donor. These lead to truncated or modified proteins with partial or complete loss of function. There is also a chance that because of these mutations some of the proteins may gain more efficiency rather than to be dis-functional. These 14 genes belong to 4 gene families: acetylcholinesterase

**Table 2** Variant Statistics

|                | Count | Percent |
|----------------|-------|---------|
| Number of variants | 2,530,451 |         |
| Number of effects   | 3,504,011 |         |
| Variant rate        | 1 /235 bases |       |
| SNP               | 2,327,972 |         |
| INS               | 103,960  |         |
| DEL               | 98,519   |         |
| Missense / Silent  | 0.4257   |         |
| Ts/Tv ratio        | 1.7147   |         |
| Heterozygous       | 122,045  |         |
| Homozygous         | 2,349,906 |        |
| Heterozygous/Homozygous | 0.05193612 |     |

**Table 3** Classification of effects and their number in the whole genome

| Type of Effects                  | No of Effects | Count | Percent |
|----------------------------------|---------------|-------|---------|
| High Effect                      | Total         | 1821  | 0.052   |
| Splice acceptor variant          | 96            | 0.003 |
| Splice donor variant             | 135           | 0.004 |
| Start loss                       | 56            | 0.002 |
| Stop gain                        | 371           | 0.003 |
| Stop lost                        | 96            | 0.003 |
| Frame shift                      | 1102          | 0.031 |
| Moderate Effect                  | Total         | 35,583| 2.724  |
| Conservative inframe deletion    | 49            | 0.001 |
| Conservative inframe insertion   | 83            | 0.002 |
| Disruptive inframe deletion      | 98            | 0.003 |
| Disruptive inframe insertion     | 74            | 0.002 |
| Missense variant                 | 35,285        | 1.004 |
| Low Effect                       | Total         | 95,439| 1.015  |
| 5′ UTR premature start gain      | 3020          | 0.086 |
| Splice region variant            | 10,980        | 0.312 |
| Stop retained                    | 106           | 0.003 |
| Synonymous variant               | 83,150        | 2.366 |
| Initiator codon/ non syn start   | 15            | 0      |
| Modifier Effects                 | Total         | 3,371,168 | 96.209 |
| 3′ UTR                           | 174,811       | 4.974 |
| 5′ UTR                           | 23,577        | 0.671 |
| Downstream                       | 485,837       | 13.823|
| Upstream                         | 421,908       | 12.041|
| Non-coding transcript            | 470           | 0.013 |
| Intron variant                   | 1,479,087     | 42.082|
| Intergenic regions               | 794,375       | 22.67 |
like protein, cathepsin (B, F, cathepsin L like), Cytochrome P450, and phosphatidylethanolamine-binding protein 1. A list of these insecticide resistance gene IDs is shown in Table 5 and those for virus transmission in Table 6 (TYLCV) and 7 (ToCV). All the genes described in Table 5 are reported for the potential involvement in insecticide resistance by Chen et al., [4], and those for virus transmission in Table 6 and Table 7 are reported by Hasegawa et al., [34] and Kaur et al., [6] respectively.

Structural variants

Structural variants were predicted through CNVnator in which the method of detection of structural variants is based on assessing the read of depth of the mapping genome. With CNVnator, among all the structural variants (duplications, deletions, insertions, inversions and translocations), some duplications were detected in the present study. Duplications with more than 1.5 cnv value are enlisted in Table 8 with their positions on the scaffolds and included genes in them. Functional annotations of these genes are presented in Additional file 6. Copy number variations were detected by CNVkit, which are described in Additional file 5. The structural variants in this study is not a comprehensive data and it is necessary to mention that reference genome is a draft genome that is about 90% of total estimated genome (~ 680–690 Mb) and in present study, 88% of this draft genome was covered with mapping reads. When the complete reference genome would be used to detect the structural variants, the results may include some more structural variants.

Discussion

Whitefly divergence into different distinct genotypes initiates the question whether the divergence results in a complex of different biotypes or it is a complex of different species! In order to resolve the divergence of whitefly question, it would be helpful to set criterion for sorting the different biotypes of whitefly and set a limit above which the difference is sufficient to declare new species status. Biological features e.g. virus transmission capacity, gut microbe diversity, host range, capacity to induce physiological changes in host plants, intermating capabilities, and capacity to spread widely have been used to differentiate cryptic species. Some of the genetic groups share common biological characters and some of the characters also show within group variability. Thus, most of the differences are uninformative or unable to resolve the cryptic species of whitefly. Molecular markers (such as AFLP, RAPD, 16S, CAPS, SCAR and mtCOI) have been used to show genetic differences between genotypes. The 3.5% genetic difference in terms of mtCOI sequences, differentiates almost all reproductively isolated groups according to available biological

---

**Table 4** Number of variant genes in each sub-class of high effects. One gene may have more than one effect and same gene may count in more than one category of high effects.

| Type of High Effects       | No of Genes |
|----------------------------|-------------|
| Splice acceptor variant    | 92          |
| Splice donor variant       | 129         |
| Start Loss                 | 55          |
| Stop gain                  | 346         |
| Stop lost                  | 91          |
| Frame shift                | 765         |
| Total                      | 1294        |

---

**Fig. 2** Distribution of variants in different genic regions
data. But some reports show disagreements with the outcomes using partial sequences of mitogenome. For example, a recent study [35], using genome wide analysis, suggested that MEAM2 might not be a separate genetic group but fall entirely into the MEAM1 group, whereas previously it was considered as a separate genetic group using mitochondrial genes. A similar phenomenon was observed in a recent study [36], where a combined analysis of experimental biological data with mitogenome sequences proposed that the African silverleafing (ASL)

Table 5 Genes potentially involved in insecticide resistance with variants between Asia II 1 and MEAM1

| Gene ID   | Annotation                              | Type of Variant            | Variant Position                  |
|-----------|-----------------------------------------|----------------------------|-----------------------------------|
| Bta08717  | Acetylcholinesterase-like protein        | Frame Shift                | Scaffold325:2419087               |
| Bta12286  | Cathepsin B                             | start lost                 | Scaffold562:2252138               |
| Bta06690  | Cathepsin F                             | stop gain                  | Scaffold2605:1316025              |
| Bta07152  | Cathepsin L-like protease               | Frame Shift                | Scaffold2737:56518                |
| Bta02560  | Cathepsin L-like protease               | Frame Shift                | Scaffold132:3684567               |
| Bta04696  | Cytochrome P450                          | Splice acceptor            | Scaffold1685:811440               |
| Bta06044  | Cytochrome P450                          | Stop lost                  | Scaffold231:1494714               |
| Bta01556  | Phosphatidylethanolamine-binding protein| Frame Shift                | Scaffold1224:613594               |
| Bta01355  | Phosphatidylethanolamine-binding protein1| start lost, splice        | Scaffold1195:116803, Scaffold1195:118926|
| Bta15207  | Phosphatidylethanolamine-binding protein1| Start lost                 | Scaffold2923:587527               |
| Bta07891  | Phosphatidylethanolamine-binding protein1| splice donor               | Scaffold300:6735496               |
| Bta12136  | Phosphatidylethanolamine-binding protein1| Frame Shift                | Scaffold545:18333                 |
| Bta13188  | Phosphatidylethanolamine-binding protein1| Splice acceptor            | Scaffold637:1563358               |
| Bta02907  | Phosphatidylethanolamine-binding protein, putative| Frame Shift                | Scaffold142449776               |
genotype, formerly treated as MED, may form a separate cryptic species. Thus, in view of these reports, species delimitation across the *B. tabaci* species complex requires data in addition to sequence divergence of mtCOI. Many recent reports show that species delimitation of a cryptic species complex requires a multi-method approach that integrates genetic differentiation, biological character, DNA barcoding, molecular phylogenetic analysis and possibly other biological features. In this regard, our study provides whole genome nuclear variants data, which will be useful to improve species delimitation of the *B. tabaci* species complex. It is also necessary to mention that although we detected all these variants between the two species Asia II 1 and MEAM1, but it may also possible that some of the variants may segregate within the same species.

In this study, we have sequenced the genome of the Asia II 1 species of whitefly and have used published transcriptomic data to infer biological differences between Asia II 1 and MEAM1. The sequencing of Asia II 1 not only provided new genomic resources for Asia II 1, but its comparative study with MEAM1 also provided insight into the comprehensive genetic differences between them.

With Blast2GO analysis, high impact variant genes were analyzed to identify the involvement of these genes in molecular pathways. The goal was to find out how genetic variances may alter or affect pathways which may then help in understanding the biological differences between the two species. Signal transduction pathways were considered as one of main points where gene alterations might help the whitefly to deal with any changes in the environment or inside the whitefly cells. Phosphatase and kinases are well-known enzymes in signal transduction pathways [37] as they activate or deactivate the functional proteins by either phosphorylation or dephosphorylation. Kinase and phosphatase functions in antagonistic ways as kinase initiates the phosphorylation and phosphatase removes the phosphate group from its substrate protein. In Additional file 7, it is noticeable that most of the genes are encoding phosphatase and kinase in different pathways e.g. phosphatase in T cell receptor signaling pathways, purine and thymine metabolism, and kinase in drug metabolism (important for pesticide resistance) and phosphatidylinositol signaling pathways. Genetic variants of these genes may alter their systematic regulatory role in biological functions.

Another prominent group of genes comprised “oxidase, dehydrogenase and reductase” enzymes performing functions in oxidative phosphorylation, amino acid (glycine, serine, threonine, valine, isoleucine, arginine and proline) metabolism, steroid degradation and biosynthesis, and biosynthesis of antibiotics. The robustness of a phloem sap sucking pest depends on the amino acid and carbohydrate contents of phloem sap of their host [38] as well as on their processing power of amino acids. For example, a Florida strain of whitefly processes more phloem sap that allows it to have more expanded host range [39]. Phloem sap lacks some essential amino acids and vitamins, so phloem sap sucking pests rely heavily on endosymbionts for some essential amino acids. There are number of genes which are present in more than one pathway for example Bta13274 encodes an oxidase involved in biosynthesis of antibiotics as well as arginine and proline metabolism, indirectly contributes to environmental fitness. A previous study reported that MEAM1 performed better than Asia II 1 on many commonly cultivated crops in China [40], and in another study MEAM1 showed the ability to adapt to unsuitable hosts [41]. Genetic variants in these genes may provide clues to the differential capacity of Asia II 1 and MEAM1 to adapt to changing environments.

Some recent studies report genes showing differential expression upon treatment of insecticide or virus infection. In our data, we identified high impact variants in 14 genes associated with insecticide resistance, 4 genes involved in TYLCV transmission, and 96 genes involved in ToCV transmission. The cathepsin gene family is involved in both insecticide resistance and ToCV transmission. Our results identified high impact variants in cathepsin B (3 genes), cathepsin F (1 gene) and cathepsin L-like genes (3 genes) that are involved in insecticide resistance and ToCV transmission. Cathepsins are proteases involved in many biological functions such as protein degradation, apoptosis, and signaling, and their activity in lysosomes has been broadly connected to virus transmission. The cathepsin B family is expanded in *B. tabaci* and also a novel clade of cathepsin L-like genes is identified in comparison to 15 other arthropods [4] which lead to the prediction of a possible contribution of cathepsin in virus acquisition or other responses that are involved in whitefly-virus interactions. Another important family in which genetic variants were found, associated with insecticide resistance is cytochrome P450 [42]. Two high impact variants were identified in two CYP 450 genes (Bta04696 and Bta06044). Chen et al., [4] inferred the involvement of these genes in insecticide resistance in MEAM1 on the basis of their

### Table 6 List of gene IDs which are potentially involved in TYLCV virus transmission and have genetic variants between Asia II 1 and MEAM1

| Gene ID       | Annotation                  | Type of Variant | Variant Position          |
|---------------|-----------------------------|-----------------|---------------------------|
| Bta10341      | Aldo-keto reductase         | Frame Shift     | Scaffold 403:3624744      |
| Bta04072      | Elicitin-like protein 6     | Frame Shift     | Scaffold161:5952976       |
| Bta02276      | Ubiquitin carboxyl-terminal hydrolase | Frame Shift | Scaffold130:858376        |
| Bta14634      | Unknown protein             | Frame Shift, Splice Donor Variant | scaffold811: 176696, scaffold811:176710 |

*Hussain et al. BMC Genomics* (2019) 20:507
| Gene ID   | Gene Name                                                                 | Type of Variants     | Scaffold:Snp Position |
|----------|---------------------------------------------------------------------------|----------------------|-----------------------|
| Bta08892 | 70 kDa heat shock protein                                                 | Frame shift          | Scaffold3328264318    |
| Bta01665 | AAA-ATPase-like domain-containing protein                                 | frame shift          | Scaffold12245022892   |
| Bta12603 | AAA-ATPase-like domain-containing protein                                 | stop gain            | Scaffold5972708628    |
| Bta05346 | Afadin, putative                                                          | stop lost            | Scaffold1991272506    |
| Bta11978 | Alpha-glucosidase                                                         | Frame shift          | Scaffold521859550     |
| Bta01804 | Ankyrin repeat and LEM domain-containing protein                         | Frame shift          | Scaffold1234405622    |
| Bta01772 | Cathepsin B                                                               | Frame shift          | Scaffold1232832350    |
| Bta07402 | Cathepsin B                                                               | Frame shift          | Scaffold21621342943   |
| Bta02120 | Cathepsin L-like protease                                                 | Frame shift          | Scaffold1261554552    |
| Bta02560 | Cathepsin L-like protease                                                 | Frame shift          | Scaffold123684567     |
| Bta07152 | Cathepsin L-like protease                                                 | Frame shift          | Scaffold273756518     |
| Bta06739 | Cation transport regulator-like protein 1                                 | Frame shift          | Scaffold26052569958   |
| Bta08022 | CG13675, isoform D                                                        | Frame shift          | Scaffold30403085831    |
| Bta04412 | CG14375                                                                   | Frame shift          | Scaffold165195426     |
| Bta03710 | CG17612, isoform A                                                         | splice acceptor variant | Scaffold155194033   |
| Bta11746 | CG7120, isoform F                                                          | splice donor, frame shift | Scaffold524009995    |
| Bta10928 | Chromodomain Y-like protein 2                                             | Frame shift          | Scaffold4771214758    |
| Bta12891 | Citron Rho-interacting kinase                                             | Frame shift          | Scaffold6132332964    |
| Bta02184 | Cystatin                                                                  | Frame shift          | Scaffold1281309680    |
| Bta07162 | DDB1-and CUL4-associated factor                                           | start loss, stop gain, | Scaffold2737410077,   |
| Bta07434 | DNA-directed RNA polymerase, omega subunit family protein                 | stop gain, frame shift | Scaffold2890209245    |
| Bta14689 | Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3B | splice acceptor variant | Scaffold8112521172   |
| Bta15680 | E3 ubiquitin-protein ligase TTC3                                          | Frame shift          | Scaffold9883252798    |
| Bta03681 | Eukaryotic translation initiation factor 3 subunit A                      | Frame shift          | Scaffold15121481746   |
| Bta14560 | Galectin                                                                  | Frame shift          | Scaffold8093964391    |
| Bta10009 | General transcription factor 3C polypeptide 2                             | Frame shift          | Scaffold3822610001    |
| Bta04387 | GH16255p                                                                  | splice acceptor variant | Scaffold16472597569 |
| Bta00770 | GK11989                                                                   | stop lost, frame shift | Scaffold1103753116   |
| Bta01833 | Klarshicht, isoform E                                                     | stop gain, frame shift | Scaffold1235600056   |
| Bta09051 | Laminin subunit beta-1                                                    | stop gain            | Scaffold3381218247    |
| Bta01704 | Loquacious                                                                | Frame shift          | Scaffold123255863     |
| Bta03800 | Lyososomal-trafficking regulator                                          | Frame shift          | Scaffold1554253425    |
| Bta05467 | Major royal jelly-related protein                                         | stop lost            | Scaffold1996871677    |
| Bta05773 | NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 12             | stop gain            | Scaffold2229151223    |
| Bta15454 | Neuroendocrine convertase 1                                               | Frame shift          | Scaffold9593096344    |
| Bta10191 | Nidogen-2                                                                 | stop gain            | Scaffold39782723     |
| Bta13257 | Protein patched                                                           | Frame shift          | Scaffold641259689     |
| Bta15368 | Protein phosphatase 1 L                                                   | Frame shift          | Scaffold95915812      |
| Bta13589 | Protein unc-45-like protein A                                             | Frame shift          | Scaffold66512187902   |
Table 7 List of gene IDs which are potentially involved in ToCV virus transmission and have genetic variants between Asia II 1 and MEAM 1 (Continued)

| Gene ID   | Gene Name                                           | Type of Variants | Scaffold:Snp Position |
|-----------|-----------------------------------------------------|------------------|-----------------------|
| Bta02051  | Regucalcin                                          | start loss       | Scaffold1240:111531   |
| Bta10926  | Replication factor-a protein 1                     | Frame shift      | Scaffold477:1149858   |
| Bta12190  | Sortilin-related receptor                           | Frame shift      | Scaffold545:2426274   |
| Bta02847  | Sulfotransferase                                    | stop gain        | Scaffold1467127       |
| Bta08229  | Symplekin                                           | splice acceptor variant | Scaffold317:609334 |
| Bta07946  | Terribly reduced optic lobes, isoform AN            | splice acceptor  | Scaffold3040:56936    |
| Bta05242  | Transcriptional protein SWT1                        | Frame shift      | Scaffold1898:321532   |
| Bta09856  | Trehalase                                           | stop gain        | Scaffold374:3016858   |
| Bta03928  | Trypsin-like serine protease                        | stop lost        | Scaffold147:7182519   |
| Bta09090  | Tudor domain protein                                | stop lost        | Scaffold338:1990664   |
| Bta08596  | Tudor domain-containing protein 1                   | Frame shift      | Scaffold3224722919    |
| Bta03892  | Ubiquitin carboxyl-terminal hydrolase               | start lost       | Scaffold1580568946    |
| Bta01518  | Unknown protein                                     | stop gain        | Scaffold147234936     |
| Bta01571  | Unknown protein                                     | Frame shift      | Scaffold1424139169    |
| Bta01615  | Unknown protein                                     | Frame shift      | Scaffold12243224397   |
| Bta02665  | Unknown protein                                     | Frame shift      | Scaffold1339520464    |
| Bta02767  | Unknown protein                                     | Frame shift      | Scaffold1371379435    |
| Bta02836  | Unknown protein                                     | Frame shift      | Scaffold1391098948    |
| Bta02920  | Unknown protein                                     | Frame shift      | Scaffold143202002     |
| Bta03301  | Unknown protein                                     | stop gain, frame shift | Scaffold147:7328434 |
| Bta03426  | Unknown protein                                     | stop gain        | Scaffold1496690294    |
| Bta03435  | Unknown protein                                     | Frame shift      | Scaffold14961047497   |
| Bta04551  | Unknown protein                                     | Frame shift      | Scaffold1655163918    |
| Bta04829  | Unknown protein                                     | Frame shift      | Scaffold17652047      |
| Bta04921  | Unknown protein                                     | Frame shift      | Scaffold17652047      |
| Bta05143  | Unknown protein                                     | stop gain        | Scaffold18461072084   |
| Bta05268  | Unknown protein                                     | stop gain, frame shift | Scaffold197132055    |
| Bta05546  | Unknown protein                                     | Frame shift      | Scaffold2013237841    |
| Bta05683  | Unknown protein                                     | Frame shift      | Scaffold2124427571    |
| Bta05758  | Unknown protein                                     | Frame shift      | Scaffold22251204179   |
| Bta05761  | Unknown protein                                     | stop gain        | Scaffold22251258041   |
| Bta05893  | Unknown protein                                     | stop gain        | Scaffold22261397519   |
| Bta06123  | Unknown protein                                     | splice donor     | Scaffold2313876649    |
| Bta07727  | Unknown protein                                     | Frame shift      | Scaffold300708005     |
| Bta07839  | Unknown protein                                     | stop gain        | Scaffold3004527825    |
| Bta08000  | Unknown protein                                     | stop gain, frame shift | Scaffold30402567504  |
| Bta08242  | Unknown protein                                     | Frame shift      | Scaffold3171074159    |
| Bta08287  | Unknown protein                                     | stop gain        | Scaffold320265593     |
| Bta08375  | Unknown protein                                     | stop gain        | Scaffold3203813827    |
| Bta08462  | Unknown protein                                     | Frame shift      | Scaffold322385722     |
| Bta08745  | Unknown protein                                     | Frame shift      | Scaffold3253471439    |
differential expression upon treatment with insecticides. Another gene that encodes a heat shock protein known to be involved in virus transmission [18] has frame shift variant in Asia II 1. Three genes which have high impact variants and are linked to ToCV transmission are associated with three KEGG pathways: oxidative phosphorylation (Bta05773), T cell receptor signaling pathway (Bta15368), and sucrose and starch metabolism (Bta09856). Bta09856 encodes trehalase a glycosidase which convert trehalose (major sugar reserve in insects play a vital role as an instant source of energy and in dealing with abiotic stresses) into glucose in sucrose and starch metabolism. The inhibition of trehalase causes abnormal growth and unsuccessful stress recovery [43]. Inhibition of trehalase provides promising area towards formulating strategy for insect control. There are also some genes with unknown functions, associated with transmission of ToCV [6]. We reported the genetic variants between Asia II 1 and MEAM1 for these genes, and future annotation of these unknown genes may provide further clues about the mechanism through which whitefly interact with a virus. This comprehensive data set of variations between indigenous and invasive species provide insights into the variations in mechanisms which give different attributes to whitefly species. Based on all these results we conclude that the MEAM1 species is more invasive due to its genetic variations.

**Conclusion**

In present study, whole genome wide variants between Asia II 1 (indigenous to the Indian sub-continent and south-east Asia) and MEAM1 (originated in the Middle East but has spread worldwide in recent decades) are presented with their detailed annotations and impact. Variants detection in some important genes such as genes associated with virus transmission and insecticide resistance will help in conceiving future research towards targeted management strategies against whitefly. Furthermore, this study provides a genomic resource of

| Table 7 | List of gene IDs which are potentially involved in ToCV virus transmission and have genetic variants between Asia II 1 and MEAM 1 (Continued) |
|---------|----------------------------------------------------------------------------------------------------------------------------------|
| Gene ID | Gene Name | Type of Variants | Scaffold:Snv Position |
| Bta10862 | Unknown protein | splice acceptor variant | Scaffold:471:791307 |
| Bta11840 | Unknown protein | Frame shift | Scaffold:52:7764853 |
| Bta12278 | Unknown protein | stop gain | Scaffold:562:2009445 |
| Bta12668 | Unknown protein | start lost | Scaffold:607:1307735 |
| Bta12727 | Unknown protein | Frame shift | Scaffold:607:2833985 |
| Bta13235 | Unknown protein | Frame shift | Scaffold:646:3239 |
| Bta13327 | Unknown protein | splice donor | Scaffold:6641:3718364 |
| Bta13745 | Unknown protein | Frame shift | Scaffold:657:1097200 |
| Bta13859 | Unknown protein | stop gain | Scaffold:667:1393372 |
| Bta13954 | Unknown protein | Frame shift | Scaffold:699:810303 |
| Bta15302 | Unknown protein | Frame shift | Scaffold:942:1732675 |
| Bta15415 | Unknown protein | Frame shift | Scaffold:959:1270849 |
| Bta07758 | Zinc finger protein | Frame shift | Scaffold:300:1778386 |
| Bta06175 | Zinc finger protein | stop gain | Scaffold:232:1822927 |
| Bta08766 | Zinc finger protein | Frame shift | Scaffold:325:3972542 |
| Bta11305 | Zinc finger protein | Frame shift | Scaffold:493:2884873 |

| Table 8 | Structural Variants |
|---------|---------------------|
| Type    | scaffold | start | end | length | CNV  | Genes* |
| duplication | Scaffold112 | 2,190,001 | 2,470,000 | 280,000 | 1.59861 | Bta02314 Bta02317 Bta02318 Bta02321 Bta02311 Bta02319 Bta02313 Bta02315 Bta02320 Bta02322 Bta02312 Bta02316 |
| duplication | Scaffold130 | 2,120,001 | 2,590,000 | 470,000 | 1.50412 | Bta08154 Bta08157 Bta08159 Bta08153 Bta08161 Bta08158 Bta08160 Bta08155 Bta08156 |
| duplication | Scaffold310 | 2,080,001 | 2,870,000 | 790,000 | 1.51561 | Bta09326 |
| duplication | Scaffold343 | 3,950,001 | 4,160,000 | 210,000 | 2.19297 | Bta10316 Bta10315 Bta10317 Bta10318 Bta10319 |

*Annotation of genes are described in Additional file 6
Asia II 1 that will contribute to resolving species delimitation of whitefly.

Methods

Colony maintenance and confirmation of cryptic species

The source of whitefly (Asia II-1) population collected from NIBGE, Faisalabad in 2016. An isogenic population was established and maintained in aired glass confinement on cotton (Gossypium hirsutum) plants at 32°C. The universal mtCOI primers C1-J-2195 (5′-TTGATTTTTTGCTACGAGAAGAGC-3′) and TL2-N-3014 (5′-TCCAATGCAC-TAATCTGCCCATTATA-3′) were used to confirm the cryptic species (Asia II 1) [44]. PCR amplifications were performed in 20 μL reactions using DreamTaq Green PCR Master Mix (Thermo Fisher Scientific). The polymerase chain reaction (PCR) cycling parameters were one denaturation cycle of 94°C for 5 min, followed by 35 cycles of 94°C for 1 min, 45°C for 1 min, and 72°C for 1 min, followed by a final extension of 72°C for 7 min. PCR products were visualized on a 1% agarose gel. Sanger sequencing [45] confirmed the Asia II 1 culture.

Genomic DNA extraction and library preparation

DNA extraction was done with “ISOLATE II Genomic DNA Kit” (Bioline Cat No. BIO-52066). Eight libraries with 550 bp insert size were prepared by the Illumina NeoPrep automation system with the library kit, Illumina #NP-101-1001, “TruSeq Nano DNA Library Kit for Neo-Prep”, which includes the adapter set “TruSeq LT” (adapter sequences: adapter read1 AGATCGGAAGAGC-CAACGTCTGAACTTCAGTCA, adapter read2 AGAT CGGAAGACCCGTCTGGTACAGGAAGATGTG). The target insert size selection was performed by the “Illumina NeoPrep Library Prep System”. Actual insert size ranges were calculated by CLC Genomics Workbench (v. 8.5.1).

Sequencing and mapping with reference genome

Sequencing was performed on the Illumina MiSeq and HiSeq2500 with Rapid v2 chemistry, 2x100bp, across 2 flow cell lanes. The Illumina bcl2fastq v2.16 software was used to convert raw basecalls (.bcl) to fastq.gz, and demultiplex the sequenced pool of libraries by the TruSeq LT indices in the NeoPrep process. The bcl2fastq script was set to automatically trim the adapters, if present. All duplicated reads, low quality regions (phred score less than threshold value) and reads containing N were trimmed. Cleaned reads were mapped onto the total reference genome of whitefly. Reference genome was accessed through ftp://www.whiteflygenomics.org/pub/MEAM1/MEAM1/ [4]. Mapping was done using BWA V0.7.12 with MEM algorithm using CLC Genomics Workbench 7.5. Raw data was visualized and analyzed to pass through quality control steps. Variant calling was performed by Haplotype caller GATK (using ‘ERC GVCF-variant_index_type LINEAR -variant_index_parameter 128,000). Variant filtration was performed two times using parameters (filtration1: DP > 20 & QD > 25.0 & FS < 5.00, filtration 2: DP > 30 & QD > 30.0 & FS = 0.00).

Analysis of variants

SnpEff [46] was used to annotate variants and effect prediction, and to classify the effects of variants by ‘functional classes’ (missense, nonsense and silent), by ‘impact’ (high, moderate, low and modifier), and by ‘type and region’ (downstream, exon, intergenic, intron, splice site acceptor, splice site donor, splice site region, transcript, upstream, UTR 3′, and UTR 5′). Then all genes that had “high impact variants” were analyzed with “Blast2GO Pro” (trial version) software [47] for gene ontology and to categorize gene functions into three classes: biological process, cellular components and molecular function. With Blast2GO Pro, KEGG pathways of these genes were also developed to analyze their function. All the mapped reads were evaluated to find structural variants. CNVnator [48] was used in the present study to find structural variants. CNVnator analyzes the “read of depth” from alignment to predict the structural variants. Copy number variations were detected by CNVkit [49].

SNPs validation

Some SNPs were randomly selected for the validation. Eight primer pairs were designed to amplify the regions (each with 500-600 bp length) which have a total of 96 SNPs. DNA was extracted from single whiteflies by the CTAB method [50]. Each region was amplified using DNA extracted from a single whitefly. PCR were performed in 50 μL reactions using DreamTaq Green PCR Master Mix (Thermo Fisher Scientific). PCR cycling parameters were one denaturation cycle of 94°C for 5 min, followed by 35 cycles of 94°C for 1 min, 54°C for 30s, and 72°C for 40 s, followed by a final extension of 72°C for 7 min. PCR products were visualized on a 1% agarose gel. Amplified products were purified by “AxyPrep PCR Clean-up Kit” and then these were sequenced by the Sanger method [45]. The sequenced reads were aligned with MEAM1 sequences by DNAStar software to validate the predicted SNPs.

We also analyzed the previously published transcriptomic data of MEAM1 [4, 6, 34]. They reported some genes that were associated with virus transmission (TYLCV and ToCV) and insecticide resistance. In our data we identified genes which had high impact variants and as well as genes previously reported as differentially expressed under virus or insecticide treatment.
Additional files

Additional file 1: Summary of raw data generated in each library. (XLSX 9 kb)

Additional file 2: Variant statistics. Variants statistics describe number of variants and effects and their types. (XLSX 12 kb)

Additional file 3: BLAST2GO results. Annotations of genes with high impact variants are classified in three main classes (1-biological function, 2-molecular function, 3-cellular components). GO IDs and number of variants in each subclass are also given in this file. (XLSX 22 kb)

Additional file 4: SNP validation. This file describes the positions of amplified regions in scaffolds and position of validated variants in each amplified region. (XLSX 13 kb)

Additional file 5: CNVkit results. (XLSX 19 kb)

Additional file 6: Gene ontology of genes having structural variants. Annotations and InterPro IDs of genes are given. (XLSX 10 kb)

Additional file 7: Gene IDs of genes contributing in different pathways. (XLSX 12 kb)

Abbreviations
BP: Biological Process; CC: Cellular Component; CTAB: Cetyl Trimethyl Ammonium Bromide; GATK: Genome Analysis Tool Kit; INDELS: Insertions Deletions; MEA11: Middle East Asia Minor 1; MF: Molecular Function; mtCOI: Mitochondrial Cytochrome Oxidase Subunit I; SNP: Single Nucleotide Polymorphism; ToCV: Tomato curl virus; ToLCNv: Tomato leaf curl Hainan virus; TYLCV: Tomato yellow leaf curl virus

Acknowledgements
We thank Atiq Ur Rehman, Imran Rauf, Muhammad Hamza Rana, Maryam Masood and Nasim Ahmad for technical assistance.

Authors’ contributions
SH and MF have equal contribution in this study. SH performed functional analysis (BLAST2G0), SNPs validation and contributed in detection of structural variants. SH also drafted the manuscript. MF performed the genome mapping, variant calling, variant analysis (SnipEff) and detection of structural variants. HJM maintained the whitefly isogenic culture and performed the mapping, variant calling, variant analysis (SnpEff) and detection of structural variants. SH also drafted the manuscript. BS designed the study and contributed in writing the manuscript. All authors read and approved the final manuscript.

Funding
This work is supported by the Pak-US cotton productivity enhancement program of the International Center for Agricultural Research in the Dry Areas (ICARDA) funded by United States Department of Agriculture (USDA), Agricultural Research Service (ARS), under agreement 58–6402–0-178F. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the USDA or ICARDA.

Availability of data and materials
The raw data of sequencing reads is submitted to SRA database of NCBI. The data can be accessed with these accession numbers: SRR8656460, SRR8656459, SRR8656466, SRR8656463, SRR8656456, SRR8656455, SRR8656458, SRR8656457, SRR8656461, SRR8656464, SRR8656465.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
1National Institute for Biotechnology and Genetic Engineering, Faisalabad, Pakistan. 2Department of Biotechnology, Pakistan Institute of Engineering & Applied Sciences (PIEAS), Nillore, Islamabad, Pakistan. 3USDA-ARS, Genomics and Bioinformatics Research Unit, 141 Experiment Station Rd., Stoneville, MS 38776, USA. 4USDA-ARS, Crop Genetics Research Unit, 141 Experiment Station Rd, Stoneville, MS 38776, USA. 5Institute of Insect Sciences, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou 310058, China.

Received: 3 December 2018 Accepted: 31 May 2019

Published online: 18 June 2019

References
1. Abd-Rabou S, Simmons AM. Survey of reproductive host plants of Bemisia tabaci (Hemiptera: Aleyrodidae) in Egypt, including new host records. Entomol News. 2010;121(3):456–65.
2. Gilbertson RL, Batuman O, Webster CG, Adkins D. Role of the insect supervector Bemisia tabaci and Frankliniella occidentalis in the emergence and global spread of plant viruses. Annu Rev Virol. 2015;2(1):67–93.
3. Geng L, Qian L-X, Shao R-X, Liu Y-Q, Liu S-S, Wang X-W. Transcriptome profiling of whitefly guts in response to tomato yellow leaf curl virus infection. Virol J. 2018;15(1).
4. Chen W, Hasegawa DK, Kaur N, Klott A, Pinheiro PV, Luan J, Stensmyr MC, Zheng Y, Liu W, Sun H, et al. The draft genome of whitefly Bemisia tabaci MEAM1, a global crop pest, provides novel insights into virus transmission, host adaptation, and insecticide resistance. BMC Biol. 2016;14(1):11.
5. Yao F-L, Zheng Y, Huang X-H, Ding X-L, Zhao J-W, Desneux N, He Y-X, Wang Q-Y. Dynamics of Bemisia tabaci biotypes and insecticide resistance in Fujian province in China during 2005–2014. Sci Rep. 2017;7:40823.
6. Kaur N, Chen W, Zheng Y, Hasegawa DK, Ling K-S, Fei Z, Wintermantel WM. Transcriptome analysis of the whitefly, Bemisia tabaci MEAM1 during feeding on tomato infected with the crinivirus, tomato chlorosis virus, identifies a temporal shift in gene expression and differential regulation of novel orphan genes. BMC Genomics. 2017;18(1):370.
7. Huipeng P, PE L, Dong C, Shao J, Qiu W, Chen Z. Insecticides promote viral outbreaks by altering herbivore competition. Ecol Appl. 2015;25(6):1585–95.
8. Ning W, Shi X, Liu B, Pan H, Wei W, Zeng Y, Sun X, Xie W, Wang S, Wu Q, et al. Transmission of tomato yellow leaf curl virus by Bemisia tabaci as affected by whitefly sex and biotype. Sci Rep. 2015;5:10744.
9. Marubayashi JM, Klott A, Yuki VA, Rezende JAM, Krause-Sakate R, Pavan MA, Ghanim M. Diversity and localization of bacterial endosymbionts from whitefly species collected in Brazil. PLoS One. 2014;9:9.
10. Qin L, Pan LL, Liu SS. Further insight into reproductive incompatibility between putative cryptic species of the Bemisia tabaci whitefly complex. Insect Sci. 2016;23(2):215–24.
11. Alemandri V, Vagg Mediina CG, DumOn AD, Arguello Caro EB, Mattio MF, Garcia Medina S, Lopez Lambertini PM, Truel G. Three members of the Bemisia tabaci (Hemiptera: Aleyrodidae) cryptic species complex occur sympatrically in argentine horticultural crops. J Econ Entomol. 2015;108(2):405–13.
12. Dinisdale A, Cook L, Riginos C, Buckley YM, De Barro P. Refined global analysis of Bemisia tabaci (Hemiptera: Sternorrhyncha: Aleyrodidae: Aleyrodidae) mitochondrial cytochrome oxidase 1 to identify species level identity and genetic boundaries. Ann Entomol Soc Am. 2010;103(2):196–208.
13. Firdaus S, Vosman B, Hidayati N, Daya Supena ED, Visser RG, van Heusden AW. The Bemisia tabaci species complex: additions from different parts of the world. Insect Sci. 2013;20(6):723–33.
14. Luan J-B, Li J-M, Varella N, Wang Y-L, Li F-F, Bao Y-Y, Zhang C-X, Liu S-S, Wang X-W. Global analysis of the transcriptional response of whitefly to tomato yellow leaf curl China virus reveals the relationship of coevolved adaptations. J Virol. 2011;85(7):3330–40.
15. Czosnek H, Hariton-Shalay A, Sobol I, Gorovits R, Ghanim M. The incredible journey of begomoviruses in their whitefly vector. Viruses. 2017;9(10):273.
16. Hofgenhout SA, Ammar el D, Whitfield AE, Redinbaugh MG. Insect vector interactions with persistently transmitted viruses. Annu Rev Phytopathol. 2008;46:327–59.
17. Pan L, Chen Q, Guo T, Wang X, Li P, Wang X, Liu S. Differential efficiency of a begomovirus to cross the midgut of different species of whiteflies results in variation of virus transmission by the vectors. Sci China Life Sci. 2018;61(10):1254–65. doi: https://doi.org/10.1007/s11427-017-9283-4.
18. Götz M, Popovski S, Kollenberg M, Gorovits R, Brown JK, Cicero JM, Czosnek H, Winter S, Ghanim M. Implication of *Bemisia tabaci* whitefly stress factor protein 70 in begomovirus-whitefly interactions. J Virol. 2012;86(24):13241–52.
19. Haritun Shalev A, Sobol I, Ghanim M, Liu S-S, Czosnek H. The whitefly *Bemisia tabaci* KNOTTIN-1 gene is implicated in regulating the quantity of tomato yellow leaf curl virus infected and transmitted by the insect. Viruses. 2016;8(7):205.
20. Kanakala S, Ghanim M. Implication of the whitefly *Bemisia tabaci* cytoplasmic protein 9 in the transmission of tomato yellow leaf curl virus. Front Plant Sci. 2016;7:1702.
21. Wang Z-Z, Shi M, Huang Y-C, Wang X-W, Stanley D, Chen X-X. A peptidoglycan recognition protein acts in whitefly (*Bemisia tabaci*) immunity and involves in begomovirus acquisition. Sci Rep. 2016;6:37806.
22. Baumann P, Munson MA, Lai C-Y, Clark MA, Baumann L, Moran N, Campbell BC. Origin and properties of bacterial endosymbionts of aphids, whiteflies, and mealybugs. ASM News. 1993;59:921–4.
23. De Barro PJ, Liu S-S, Boykin LM, Dinsdale AB. *Bemisia tabaci*: a species status. Annu Rev Entomol. 2010;56(1):1–19.
24. Liu S-S, Colvin J, De Barro PJ. Species concepts as applied to the whitefly *Bemisia tabaci* systematics: how many species are there? J Integr Agric. 2012;11(2):176–86.
25. Ashfaq M, Hebert PDN, Pirzai MA, Khan AM, Mansoor S, Shah GS, Zafar Y. DNA barcoding of *Bemisia tabaci* complex (Hemiptera: Aleyrodidae) reveals southerly expansion of the dominant whitefly species on cotton in Pakistan. PLoS One. 2014;9(8):104485.
26. Masood MA, Amin I, Hassan I, Mansoor S, Brown JK, Briddon RW. Diversity and distribution of cryptic species of the *Bemisia tabaci* (Hemiptera: Aleyrodidae) complex in Pakistan. J Econ Entomol. 2017;110:2295–300.
27. Li M, Hu J, Xu F-C, Liu S-S. Transmission of tomato yellow leaf curl virus by two invasive biotypes and a Chinese indigenous biotype of the whitefly *Bemisia tabaci*. Int J Pest Manag. 2010;56(3):275–80.
28. Götz M, Winter S. Diversity of *Bemisia tabaci* in Thailand and Vietnam and indications of species replacement. J Asia Pac Entomol. 2016;19(2):537–43.
29. Pan L-L, Cui X-Y, Chen Q-F, Wang X-W, Liu S-S. Cotton leaf curl disease: which whitefly is the vector? Phytopathology. 2018;108(10):1172–83.
30. Naveen NC, Chauhary R, Kumar D, Rejibith KB, Rajagopal R, Subrahmanyan B, Subramanian S. 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. 2017;7:40694.
31. Ahmed MZ, Naveed M, Noor ul Ane M, Ren SX, De Barro P, Qiu BL. Host suitability comparison between the MEAM1 and MEA1 cryptic species of *Bemisia tabaci* in cotton-growing zones of Pakistan. Pest Manag Sci. 2014;70(10):1531–7.
32. Schuster SC. Next-generation sequencing transforms today’s biology. Nat Methods. 2007;5(1):16.
33. Chen W, Hasegawa D, Arumuganathan K, Simsmons A, Wintermantel W, Fei Z, Ling K-S. Estimation of the whitefly *Bemisia tabaci* genome size based on k-mer and flow cytometric analyses. Insects. 2015;6(31):704.
34. Hasegawa DK, Chen W, Zheng Y, Kaur N, Wintermantel WM, Simmons AM, Fei Z, Ling K-S. Comparative transcriptome analysis reveals networks of genes activated in the whitefly *Bemisia tabaci* when fed on tomato plants infected with tomato yellow leaf curl virus. Virology. 2018;533:52–64.
35. Ellefkh S, Etter P, Tay WT, Fumagalli M, Gordon K, Johnson E, De Barro P. Genome-wide analyses of the *Bemisia tabaci* species complex reveal contrasting patterns of admixture and complex demographic histories. PLoS One. 2018;13(1):0190555.
36. Vyskočilová S, Tay WT, van Brunschot S, Seal S, Colvin J. An integrative approach to discovering cryptic species within the *Bemisia tabaci* whitefly species complex. Sci Rep. 2018;8(11):10886.
37. Seger R, Krebs EG. The MAPK signaling cascade. FASEB J. 1995;9(9):726–35.
38. Avvikack CS, Leather SR. Host plant quality and fecundity in herbivorous insects. Annu Rev Entomol. 2002;47:817–44.
39. Byrne DN, Miller WB. Carbohydrate and amino acid composition of phloem sap and honeydew produced by *Bemisia tabaci*. J Insect Physiol. 1990;36(6):433–9.
40. J. X, K. K, L. S. Performance on different host plants of an alien and an indigenous *Bemisia tabaci* from China. J Appl Entomol. 2011;135(10):771–9.
41. Carabali A, Bellotti AC, Montoya-Lerma J, Cuellar ME. Adaptation of *Bemisia tabaci* biotype B (Gennadius) to cassava, *Manihot esculenta* (Crantz). Crop Prot. 2005;24(7):643–9.
42. Yang X, Xie W, Wang S-L, Wu Q-J, Pan H-P, Li R-M, Yang N-N, Liu B-M, Xu B-Y, Zhou X, et al. Two cytochrome P450 genes are involved in imidacloprid resistance in field populations of the whitefly, *Bemisia tabaci*, in China. Pest Biochem Physiol. 2013;107(3):343–50.
43. Shukla E, Thorrat LI, Nath BB, Galkowd SM. Insect trehalase: physiological significance and potential applications. Gycobiology. 2013;23(4):357–67.
44. Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Annt Enmol Soc Am. 1994;87(6):651–701.
45. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977;74(12):5463–7.
46. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms. Nven. Flie. 2012;2(8):20–92.
47. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M, Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. J Bioinform. 2005;21(18):3674–6.
48. Abyzov A, Urban AE, Snyder M, Gerstein M. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res. 2011;21(6):974–84.
49. Televisch E, Shain AH, Botton T, Bastian BC. CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. PLoS Comput Biol. 2016;12(4):1004873.
50. Clarke JD. Cetyltrimethyl ammonium bromide (CTAB) DNA miniprep for plant DNA isolation. Cold Spring Harb Protoc. 2009;2009(3):pdb.prot5177. doi: https://doi.org/10.1101/pdb.prot5177.

**Publisher’s Note**
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

---

**Ready to submit your research?** Choose BMC and benefit from:
- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more biomedcentral.com/submissions