Comparison of Phenacoccus solenopsis specimens from different regions of Pakistan using COI molecular barcoding (Hemiptera: Pseudococcidae)

Article (Accepted Version)

Comparison of *Phenacoccus solenopsis* specimens from different regions of Pakistan using COI molecular barcoding.

Hayat Badshah[1], Farman Ullah[2], Hayat Zada[2], Qamar Zeb[1], Ruidar A. Shah[1], Paul-André Calatayud[3, 4] & Neil Crickmore[5*]

[1]Entomology Section, Agricultural research Institute Tarnab (24330), Peshawar, Pakistan, [2]Department of Plant Protection, The University of Agriculture, Peshawar, Pakistan, [3]UMR EGCE (Evolution, Génome, Comportement, Ecologie), CNRS-IRD-Univ. Paris-Sud, IDEEV, Université Paris-Saclay, Gif-sur-Yvette Cedex, France, [4]IRD c/o icipe P.O. Box 30772-00100, Nairobi, Kenya, [5]School of Life Sciences, University of Sussex, Brighton, UK.

***Correspondence:** Neil Crickmore, School of Life Sciences, University of Sussex, Brighton, UK, BN1 9QG. E-mail n.crickmore@sussex.ac.uk

Keywords: Cotton pest, pest management, mealybug, Pseudococcidae, biological control

Summary. Because correct identification of insects is crucial for pest management involving chemical or biological control agents, we have used a molecular approach to identify and characterize specimens of the cotton pest *Phenacoccus solenopsis* Tinsley (Sternorrhyncha: Pseudococcidae) present in different regions of Pakistan. The specimens were analyzed through DNA sequence analysis of their mitochondrial COI (mtCOI) gene using an improved procedure that could distinguish between the pest and its associated parasitoid. Our analysis showed no variation among the mealybug specimens from different geographical locations of Pakistan and confirmed that this is the same species and haplotype that is infesting cotton plants in other parts of Asia. This information will assist in the development of biological control programs against *P. solenopsis* in Pakistan and other Asian countries.

Résumé. Comparaison des spécimens de *Phenacoccus solenopsis* de différentes régions du Pakistan utilisant le barcoding moléculaire COI. Parce que l'identification correcte des insectes est cruciale pour la gestion des ravageurs par des agents chimiques ou biologiques, nous avons utilisé une approche moléculaire pour identifier et caractériser des spécimens du ravageur du coton *Phenacoccus solenopsis* Tinsley (Sternorrhyncha: Pseudococcidae) présents dans différentes régions du Pakistan. Les échantillons ont été analysés par analyse de la séquence d'ADN de leur gène mitochondrial COI (mtCOI) en utilisant une procédure améliorée qui permet la distinction du ravageur et de son parasitoïde associé. Notre analyse n'a montré aucune variation entre les spécimens de cochenilles de différentes localisations géographiques du Pakistan et confirme qu'il s'agit de la même espèce et de l'haplotype qui infeste des plants de coton dans d'autres pays d'Asie. Ces informations vont aider à l'élaboration d'un programme de lutte biologique contre *P. solenopsis* au Pakistan et dans d'autres pays d'Asie.

INTRODUCTION

The mealybug *Phenacoccus solenopsis* Tinsley (Sternorrhyncha: Pseudococcidae) is a highly polyphagous pest, feeding on more than 154 plant species including fruits, vegetables, crops and ornamentals in Pakistan (Arif et al. 2009). It was first described by Tinsley in 1898, while Wang et al. (2009) reviewed its distribution and suggested that it is native of the US. Its first record from outside of the US mainland was from Hawaii in 1966 (Kumashiro et al. 2001) and then in Asia from 2000 onward (Hodgson et al. 2008). In Pakistan, it was detected for the first time in 2005 (Abbas et al. 2008; Muhammad 2007). It is a major risk to cotton,

vegetables and fruits causing enormous losses worldwide (779.43 US$/ha) and has reduced average seed cotton yield by 44% (Dhawan et al. 2007). During 2005, in the cotton producing belt of Pakistan (Punjab and Sindh provinces), 60,700 ha from a 3,237,485 ha crop were seriously damaged by *P. solenopsis* and a 14% loss of the crop was reported. In 2007, almost 40% damage, equivalent to 3.1 million bales of cotton in the Punjab, was reported (Mahmood et al. 2011; Kakakhel 2007). In India, *P. solenopsis* damage in 2007 was estimated at US$400,000–500,000 and an even more severe attack was reported in 2008-2009 (Dutt 2007; Nagrare et al. 2009). In addition to the direct losses that the insects can cause by sucking the phloem sap, its feeding secretions (honeydew) cause additional losses to the plants by disturbing their photosynthetic activity and inducing fungal contaminations (Arif et al. 2012; Babar et al. 2013).

This species was observed to have considerable morphological variation which, along with the presence of abundant mealy waxes, led to misidentifications (Thomas & Ramamurthy 2008). Hodgson et al. (2008) reported that this morphological variation might be induced by environmental factors. But Singh et al. (2012) suggested that the wide host range of this pest might facilitate the occurrence of new biotypes, which may render it difficult to control the pest. Being able to identify different genetic variants of *P. solenopsis* could therefore potentially facilitate bespoke control procedures.

Various molecular markers have been used as tools to differentiate insect species and to study the evolutionary relationships among multiple species (Mowry & Barbour 2004; Gariepy et al. 2007). The most common genes used for insect species/subspecies differentiation have been cytochrome oxidase I and II (COI and COII) of mitochondrial DNA (Ashfaq et al. 2010; Li et al. 2005). More recently, the COI sequence has been established as a reference gene for species identification and classification through DNA barcoding (Miller 2007; Linares et al. 2009). Saccaggi et al. (2008) differentiated three mealybug species on the basis of COI gene sequence.

Ashfaq et al. (2010) identified *P. solenopsis* in Pakistan through DNA sequencing using partial nucleotide sequences of nuclear (elongation factor-1α, ribosomal DNA subunits 18S and 28S) and mitochondrial (COI) genes. PCR-RFLP analysis using three different endonucleases did not show any polymorphisms in individuals collected from various host plants and geographical locations of the Punjab and Sindh districts of Pakistan, indicating that the use of one gene sequence, such as COI, could be enough to identify other mealybug specimens in Pakistan.

In this context, since there is little information on the genetic variability of *P. solenopsis* in other regions of Pakistan (Ashfaq et al. 2010), in this study we have augmented the existing molecular dataset of mitochondrial COI gene sequences with those of *P. solenopsis* from four different, and previously unsampled, regions of Pakistan.

MATERIALS AND METHODS

Specimen collection and DNA extraction/amplification

Phenacoccus solenopsis females were collected from four different regions of Pakistan (Table 1). From each region, about 20 cm of infested shoots from five different plants of the species listed in the table were collected and brought to the laboratory. The mealybugs were removed from the shoots using a camel hair brush, and a total of one hundred mealybug females from each location were separated, identified and confirmed to be *Phenacoccus solenopsis* Tinsley using the taxonomic keys of Hodgson *et al*. (2008). For each locality, fifty mealybug females were preserved in 95% alcohol for future use. For the molecular analyses, eight females from a given preserved sample were put in a 5ml glass pestle and crushed with a mortar (Wheaton-USA). DNA was extracted using the Qiagen DNeasy Blood & Tissue kit. Total genomic DNA was preserved at –20 °C. We first used primers previously published from other Pseudococcids to amplify by PCR an 877 fragment of the COI gene (C1J2195: TTGATTYTTTGGTCATCCAGAAGT and TL2N3014: TCCAATGCACTAATCTGCCATATTA) (Ashfaq et al. 2010). The PCR conditions used were: initial denaturation at 94°C for 120 sec, 35 cycles of 94°C for 30 seconds, annealing 48°C for 90 seconds, extension at 72°C for 120 seconds and then a final extension at 72°C for 5 minutes. The Promega GoTaq Green Master Mix was used according to the manufacturer's protocol. PCR amplicons were purified from an agarose gel using the Qiaquick DNA extraction kit from Qiagen. As sequencing results showed that amplicons from two species were then obtained (see results), the corresponding PCR products were cloned into a plasmid vector and introduced into *E. coli*. A ligation reaction solution of 10 µl was made up of 5 µl of 2X Rapid ligation buffer, 1 µl of pGEM-T Easy vector (50ng), 123 3µl purified PCR product and 1 µl of T4 DNA Ligase (3Weiss units/µl) in an Eppendorf tubes. The reaction solution was mixed by pipetting and then incubated for 1 hour at room temperature and then stored at 4°C. The mixture was used to transform *E. coli* DH5α cells via electroporation. We then used the sequences obtained following the cloning procedure to design new primers that specifically amplified the COI gene of *P. solenopsis*. These primers were SolenopsisF (GGAAAAATTGAAATTTTTAGAAAAATTA) and SolenopsisR (CATTAAATGAATGATTTAAATTTGGAG). The corresponding amplicons were directly sequenced (MWG Eurofins).

Ligation into pGEM-T Easy Vector

In order to separate individual amplicons from a PCR reaction the products were cloned into a plasmid vector and introduced into *E. coli*. A ligation reaction solution of 10 µl was made up of 5 µl of 2X Rapid ligation buffer, 1 µl of pGEM-T Easy vector (50ng), 3µl purified PCR product and 1 µl of T4 DNA Ligase (3Weiss units/µl) in an Eppendorf tube. The reaction solution was mixed by pipetting and then incubated for 1 hour at room temperature and then stored at 4^{o}C. The mixture was used to transform *E. coli* DH5α cells via electroporation.

Phylogenetic Analysis

The sequences were compared to other published sequences using MEGA7 (Kumar et al. 2016). The Clustal alignment option was chosen to generate the multiple sequence alignment. The distance matrix was calculated from within MEGA7 using the pairwise Kimura 2-parameter (uniform rate) model. The evolutionary history was inferred using the Neighbor-joining method (Saitou & Nei 1987). The evolutionary distances were computed using the Maximum Composite Likelihood method of Tamura et al. (2004) and are in the units of the number of base substitutions per site. The analysis involved 13 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 140 positions in the final dataset.

RESULTS

Specimens of mealybug collected from four regions of Pakistan (Table 1) were identified as *Phenacoccus solenopsis* through morphological examination. Specifically adult females had two dark stripes either side of a middle body ridge, short waxy filaments around the body and quarter-length anal filaments (Arif et al. 2009). When amplified with the previously published primers, the sequence results suggested a mixture of species. Several positive clones were thus sent for sequencing and returned sequences that had a 100% match to a GenBank entry (AB521131) for the COI gene from *Aenasius* sp. (deposited in GenBank as

KJ995915-17). Thus it appeared that these primers amplified both *P. solenopsis* and one of its parasitoids, possibly *Aenasius bambawalei* (Hymenoptera: Encyrtidae). To address this problem an alignment was performed between the COI sequences from host and parasitoid, the new primers were designed and direct sequencing was then possible. We finally obtained sequences that exactly matched published sequences for *P. solenopsis*. (Table 2). Sequences were obtained from four samples from each geographical region but no variation was found, that is all our sequences, regardless of where they were sampled, presented the same haplotype. A representative sequence from each region was deposited in GenBank as KJ995911 (Peshawar), KJ995912 (Mohmand Agency), KJ995913 (Dera Ismail Khan) and KJ995914 (Naseer Abad). A recent study (Ahmed et al. 2015) used COI gene sequences to characterize various species of *P. solenopsis* and suggested that the species could be split into two major groups, one found in the United States and another in Asia. Within the Asian group nine different haplotypes could be identified, five of which (H1-5) appear more closely related than the other four (H6-9). The distance matrix (Table 2) and phylogenetic tree (Figure 1) comparing the DNA sequences from our specimens with these nine haplotypes both show that our samples cluster with the main H1-5 clade.

DISCUSSION

For effective pest control there is a need to identify species of mealybug and to differentiate between populations of the same species that could have different insecticide tolerances. The identification of mealybugs is generally difficult because of their similar size and morphological features and abundance of mealy waxes. Classical taxonomic keys require a high level of skill and can be time consuming especially as young crawlers often have to be reared up to the adult, or at least the preovipositing, stage (Daane et al. 2011; Hardy et al. 2008; Thomas & Ramamurthy 2008). We have found that sequencing of the COI gene provides a reliable identification but was compounded by the fact that published primers are capable of also amplifying the gene from associated parasitoids. The design of new primers overcame this limitation.

We found no sequence differences between specimens from the four geographical regions which is in agreement with Ashfaq et al. (2010) who found no molecular differences between mealybugs collected from various host plants and geographical locations of the Punjab and Sindh provinces of Pakistan. In contrast, Singh et al. (2012) reported only 68 to 78% similarity between different populations analyzed by four different RAPD markers and claimed that there is a significant possibility of having different biotypes of *P. solenopsis*

which may differ in resistance to insecticides. Our study indicates that there is no evidence for such different biotypes since this species entered Pakistan. This basic molecular barcode however will not detect more subtle differences in genetic structure that could affect their susceptibility to control agents. Should significant differences in susceptibility be found then more detailed analyses will be needed to characterize the different biotypes.

Detailed analyses of *P. solenopsis* populations using genetic markers have indicated two main clades, one Asian and one North American (Ahmed et al. 2015; Dong et al. 2009; Thomas & Ramamurthy 2014). The former study further divided the Asian clade into separate haplotypes. Figure 1 shows that the haplotype that we identified most closely matches, and is placed in the same monophyletic clade as, the H2 and H5 haplotypes which were derived from specimens found in Pakistan, China, India and Vietnam. The lack of any variation in the COI gene between specimens collected from different regions of Pakistan suggests that there is a low degree of genetic diversity between populations which will hopefully be reflected in a uniform response to pest control measures.

ACKNOWLEDGEMENTS

This research/visit was sponsored by Higher Education Commission of Pakistan under IRSIP (International research support initiative program) for PhD students. The cooperation of Alicia Elhigazi and Barbara Domanska, PhD students at the School of Life Sciences, University of Sussex, UK throughout the course of this study is thankfully acknowledged.

REFERENCES

Abbas G, Arif MJ, Saeed S, Kava H. 2008. Increasing menace of a new mealybug, *Phenacoccus gossypiphilous*, to the economic crops of Southern Asia. Proc. XI Int. Symp. on Scale Ins. Studies, Oeiras, Portugal, 24-27th Sept. 2007.

Ahmed AM, Jun MA, Qiu BL, Ri-Rong HE, Mu-Tao WU, Liang F, Zhao JP, Lin L, Hu XN, Lv LH, Jesse W, Breinhol T, Yong YL. 2015. Genetic record for a recent invasion of *Phenacoccus solenopsis* in Asia. Environmental Entomology. 44:907–918.

Arif MI, Rafiq M, Ghaffar A. 2009. Host plants of cotton mealybug (*Phenacoccus solenopsis*): a new menace to cotton agroecosystem of Punjab. International Journal of Agriculture Biology. 11:163–167.

Arif MI, Rafiq M, Wazir S, Mehmood N, Ghaffar A. 2012. Studies on cotton mealybug, *Phenacoccus solenopsis* (Pseudococcidae: Homoptera), and its natural enemies in Punjab, Pakistan. International Journal of Agriculture Biology. 14:557–562.

Ashfaq M, Noor AR, Mansoor S. 2010. DNA-based characterization of an invasive mealybug (Hemiptera: Pseudococcidae) species damaging cotton in Pakistan. Applied Entomology and Zoology. 45:395-404.

Babar TK, Karar H, Hasnain M, Shahazad MF, Saleem M, Ali A. 2013. Performance of some transgenic cotton cultivars against insect pest complex, virus incidence and yield. Pakistan Journal of Agricultural Sciences. 50:367-372.

Daane KM, Middleton MC, Sforza R, Cooper ML, Walton VM, Walsh DB, Zaviezo T, Almeida RP. 2011. Development of a multiplex PCR for identification of vineyard mealybugs. Journal of Environmental Entomology. 40:1595-1603.

Dhawan AK, Singh K, Saini S, Mohindru B, Kaur A, Singh G, Singh S. 2007. Incidence and damage potential of mealy bug, *Phenacoccus solenopsis* Tinsley, on cotton in Punjab. Indian Journal of Ecology. 34:166-172.

Dong C, Guo-Xia L, Hai-Bin F, Wei X. 2009. Phylogenetic analysis of mtCO1 reveals the cryptic lineages in *Phenacoccus solenopsis* complex (Hemiptera: Pseudococcidae). Acta Entomologica Sinica**.** 52:1261-1265.

Dutt U. 2007. Mealybug infestation in Punjab: Bt. cotton falls flat. Available from: http://www.countercurrents.org/dutt2108 07.htm

Gariepy TD, Kuhlmann U, Gillott C, Erlandson M. 2007. Parasitoids, predators and PCR: The use of diagnostic molecular markers in biological control of arthropods. Journal of Applied Entomology. 131:225-240.

240 Hardy NB, Gullan PJ, Hodgson CJ. 2008. A classification of mealybugs (Hemiptera:
241 Pseudococcidae) based on integrated molecular and morphological data. Systematic
242 Entomology**.** 33:51-71.
243 Hodgson C, Abbas G, Arif MJ, Saeed S, Karkar H. 2008. *Phenacoccus solenopsis* Tinsley
244 (Sternorrhyncha: Coccoidea: Pseudococcidae), an invasive mealybug damaging cotton
245 in Pakistan and India, with a discussion on seasonal morphological variation. Zootaxa**.**
246 1913:1-35.
247 Kakakhel I. 2007. Mealybug attack affects cotton crop on 150,000 acres. Available from:
248 http://www.dailytimes.com.pk/default.asp pg 54.
249 Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular evolutionary genetics analysis
250 version 7.0 for bigger datasets. Molecular Biology and Evolution**.** 33:1870-1874.
251 Kumashiro BR, Heu RA, Nishida GM, Beardsley JW. 2001. New state records of immigrant
252 insects in the Hawaiian Islands for the year 1999. Proceedings of Hawaiian
253 Entomological Society. 35:170-184.
254 Li HM, Deng RQ, Wang JW, Chen ZY, Jia FL, Wang XZ. 2005. DNA-based
255 characterization of cotton mealybug a preliminary phylogeny of the pentatomomorpha
256 (Hemiptera: Heteroptera) based on nuclear 18S rDNA and mitochondrial DNA
257 sequences. Molecular Phylogenetics and Evolution. 37:313–326.
258 Linares MC, Soto-Calderon ID, Lees DC, Anthony NM. 2009. High mitochondrial diversity
259 in geographically widespread butterflies of Madagascar: a test of the DNA barcoding
260 approach. Molecular Phylogenetics and Evolution. 50:485–495.
261 Mahmood R, Aslam MN, Solangi GS, Samad A. 2011. Historical perspective and
262 achievements in biological management of cotton mealybug *Phenacoccus solenopsis*
263 Tinsley in Pakistan. Proc. 5th meeting Asian Cotton Res. Dev. network, Conf. held at
264 Lohore Pak. pp. 17.
265 Miller SE. 2007. DNA barcoding and the renaissance of taxonomy. Proceedings of the
266 National Academy of Science of the United States of America**.** 104:4775–4776.
267 Mowry TM, Barbour JD. 2004. Distinguishing the parasitic wasp, *Peristenus howardi*, from
268 some of its congeners using polymerase chain reaction and restriction endonuclease
269 digestion. Journal of Insect Science**.** 4:25–31.
270 Muhammad A. 2007. Mealybug: cotton crop's worst catastrophe. Centre for Agro-
271 Informatics Research (CAIR), Pakistan. Available from:
272 http://agroict.org/pdf_news/Mealybug.pdf

Nagrare VS, Kranthi S, Biradar VK, Zade NN, Sangode V, Kakde G, Shukla RM, Shivare D, Khadi BM, Kranthi KR. 2009. Widespread infestation of the exotic mealybug species, *Phenacoccus solenopsis* (Tinsley) (Hemiptera: Pseudococcidae), on cotton in India. Bulletin of Entomological Research**.** 99:537-541.

Saccaggi DL, Kruger K, Pietersen G. 2008. A multiplex PCR assay for the simultaneous identification of three mealybug species (Hemiptera: Pseudococcidae). Bulletin of Entomological Research. 98:27–33.

Saitou N, Nei M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution. 4:406-425.

Singh S, Sharma R, Kumar R, Gupta VK, Dilawari VK. 2012. Molecular typing of mealybug *Phenacoccus solenopsis* populations from different hosts and locations in Punjab, India. Journal of Environmental Biology**.** 33:539-543.

Tamura K, Nei M, Kumar S. 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proceedings of the National Academy of Science of the United States of America**.** 101:11030–11035.

Tinsley JD. 1898. An ants'-nest coccid from New Mexico. The Canadian Entomologist**.** 30:47-48.

Thomas A, Ramamurthy VV. 2014. Morphological and molecular studies on the intra specific variations between populations of the cotton mealybug *Phenacoccus solenopsis* Tinsley (Hemiptera: Pseudococcidae). Entomological News**.** 123:339-347.

Thomas A, Ramamurthy VV. 2008. On the problems in the diagnostics of cotton mealybug *Phenacoccus solenopsis*. Entomon**.** 33:277-282.

Wang YP, Wu SA, Zhang RZ. 2009. Pest risk analysis of a new invasive pest, *Phenacoccus solenopsis*, to China. Chinese Bulletin of Entomology. 46:101-106.

Table 1. Locations where the mealybugs were collected

Sample	Host plant	Location
Mb1	China rose (*Hibiscus rosa-sinensis* Linn)	Peshawar (Latitude 34° 1′ 0″ North, Longitude, 71° 35′ 0″ East, Elevation, 359 m above the sea level)
Mb2	Tomato (*Lycopersicon esculentum* Mill)	Mohmand Agency (Latitude 34°19'16″ North, Longitude 71°24'0″, East, Elevation, 651 m above sea level).
Mb3	China rose (*Hibiscus rosa-sinensis* Linn)	Dera Ismail Khan (Latitude 31054'33" North, Longitude 070053'47" East, Elevation,166 m above sea level)
Mb4	Cotton (*Gossipium hirsutum* L.)	Naseer Abad (Baluchistan) (Latitude 28° 37' 35" North, Longitude 68° 7' 45" East, Elevation, 63 m above sea level)

Table 2. Distance matrix for the Asian *P. solenopsis* specimens

Peshawar												
Mohmand Agency	0.00											
Dera Ismail Khan	0.00	0.00										
Naseer Abad	0.00	0.00	0.00									
AsiaH5	0.00	0.00	0.00	0.00								
AsiaH1	0.00	0.00	0.00	0.00	0.00							
AsiaH4	0.00	0.00	0.00	0.00	0.00	0.00						
AsiaH2	0.00	0.00	0.00	0.00	0.00	0.00	0.00					
AsiaH3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00				
AsiaH7	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49			
AsiaH6	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.01		
AsiaH8	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.01	0.01	
AsiaH9	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.03	0.04	0.02

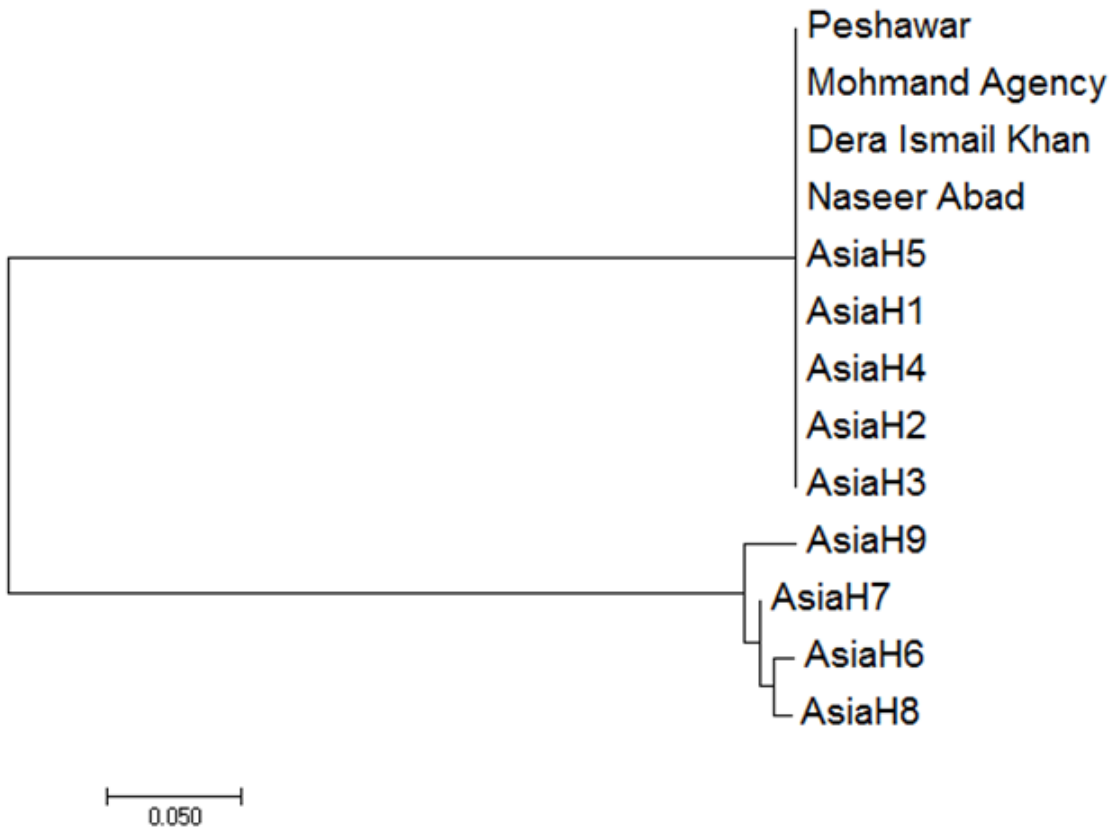

Figure 1. Evolutionary relationships of Asian *P. solenopsis* specimens. The optimal tree with the sum of branch length = 0.60796449 is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree.