Homology difference analysis of invasive mealybug species *Phenacoccus solenopsis* Tinsley in Southern China with COI gene sequence variability

F.Z. Wu¹,²*, J. Ma³, X.N. Hu³ and L. Zeng¹*

¹College of Resources and Environment, South China Agricultural University, Guangzhou, China: ²Huizhou Entry-Exit Inspection and Quarantine Bureau, Huizhou, China: ³Guangdong Inspection and Quarantine Technology Center, Guangzhou, China

Abstract

The mealybug species *Phenacoccus solenopsis* (*P. solenopsis*) has caused much agricultural damage since its recent invasion in China. However, the source of this invasion remains unclear. This study uses molecular methods to clarify the relationships among different population of *P. solenopsis* from China, USA, Pakistan, India, and Vietnam to determine the geographic origin of the introduction of this species into China. *P. solenopsis* samples were collected from 25 different locations in three provinces of Southern China. Samples from the USA, Pakistan, and Vietnam were also obtained. Parts of the mitochondrial genes for cytochrome oxidase I (COI) were sequenced for each sample. Homologous DNA sequences of the samples from the USA and India were downloaded from Gen Bank. Two haplotypes were found in China. The first was from most samples from the Guangdong, Guangxi, and Hainan populations in the China and Pakistan groups, and the second from a few samples from the Guangdong, Guangxi, Hainan populations in the China, Pakistan, India, and Vietnam groups. As shown in the maximum likelihood of trees constructed using the COI sequences, these samples belonged to two clades. Phylogenetic analysis suggested that most *P. solenopsis* mealybugs in Southern China are probably closely related to populations in Pakistan. The variation, relationship, expansion, and probable geographic origin of *P. solenopsis* mealybugs in Southern China are also discussed.

Keywords: *Phenacoccus solenopsis* Tinsley, COI gene, geographic populations, genetic differentiation, phylogenetic analysis

(Introduced 21 August 2014; First published online 29 October 2014)

Introduction

*Phenacoccus solenopsis* (*P. solenopsis*) Tinsley (Hemiptera: Sternorrhyncha: Coccoidea: Pseudococcidae) was first reported as a cotton pest in Texas, USA (Fuchs *et al.* 1991). Given the high variation of this species in terms of its morphological characteristics, biological adaptability, and ecological adjustability, it has reportedly settled in over 35 localities around the world, including Central America, the Caribbean, Ecuador (Ben-Dov, 1994), Chile (Larrain, 2002), Argentina (Granara de Willink, 2003), Brazil (Culik & Gullan, 2005), Pakistan, India (Karar, 2008), Nigeria, Sri Lanka, and recently, China (Wang *et al.*, 2009).
To adapt to a new environment, an invasive species may undergo a series of genetic and physiological changes. Several studies found that an invasive species can adapt through addictive genetic variation, epistasis, hybridization, genetic tradeoff, and a small number of gene and genomic rearrangements (Rhymer & Simberloff, 1996; Lee, 1999, 2002; Tsutsui et al., 2000). In genetic variation, invasive species typically exhibit a broad range of behavior adaptability. For example, *Capsella bursa-pastoris* blooms earlier in the desert than on the beach or in snow forests (Neuffer & Hurka, 2002), and almost all invasive weeds have smaller seeds and shorter flowering times (Maillet & Lopez-Garcia, 2000).

Thus, the genetic or behavioral variation of a newly imported species somehow reflects its relationship with the population of the surrounding area. Genetic data provide information on the source of the invading population as well as hypotheses on the environmental and evolutionary factors that ensure the success of biological invasions (Sax et al., 2005).

*P. solenopsis*, a notorious pest in cotton and a wide array of host plants, was first discovered in Guangzhou, China and has spread to several other provinces since then (Ma et al., 2009). However, whether this insect has its own genetic and behavioral differences from native species remains unclear. Data illustrating whether relationships exist among the same-species populations from different areas after invasion are also unavailable. Thus, we obtained 31 *P. solenopsis* samples from 25 different locations in Southern China (fig. 1) and three samples from Pakistan to identify the relationship between the Chinese and Pakistani species. Maconellicoccus hirsutus from Southern China was included as an outgroup. These samples were initially classified based on their morphology. Figure 1 gives an overview of the geographical distribution of the samples. Details regarding the sampling locations are provided in table 1. Live insects were carefully removed from the host plants. Once collected, the insects were immersed in 95% ethanol. Voucher samples were preserved at the Plant Quarantine Laboratory, Guangdong Inspection and Quarantine Technology Center.

Materials and methods

Sampling collection

The *P. solenopsis* samples used in the present investigation were obtained from three provinces in Southern China, at the port, Pakistan (three samples), Vietnam (one sample), and the USA (two samples). *Maconellicoccus hirsutus* from Southern China was included as an outgroup. These samples were initially classified based on their morphology. Figure 1 gives an overview of the geographical distribution of the samples. Details regarding the sampling locations are provided in table 1. Live insects were carefully removed from the host plants. Once collected, the insects were immersed in 95% ethanol. Voucher samples were preserved at the Plant Quarantine Laboratory, Guangdong Inspection and Quarantine Technology Center.

**DNA extraction, amplification, and sequencing**

All specimens were examined for the presence of parasitoids using a microscope. Total genomic DNA was extracted from the entire body of the *P. solenopsis* sample using the DNeasy Blood and Tissue Kit (QIAGEN) according to the manufacturer’s instructions. DNA concentration and purity were estimated by means of absorbance at 260–280nm in a NanoDrop ND-1000 (NanoDrop Technologies, Willminton, USA) spectrophotometer. Primers CIJ2195 TTGATTYYTTG-GTCATCCAGAAGT and TL2N3014 TCCAATGCATAAT-CTGCCCATATTA (Simon et al., 1994) were initially used to amplify 840bp of the mitochondrial COI gene. All samples were successfully amplified.
We amplified a partial sequence of the COI gene for phylogenetic analysis. The reaction was conducted in a thermal cycler in a 30 μl system composed of 3 μl of PCR buffer, 2.4 μl of dNTP mixture, 0.15 μl of Taq DNA polymerase, 1 μl of each primer, 1.8 μl of the DNA template, and ddH2O added to a final reaction volume of 30 μl. The PCRs were conducted in Eppendorf Mastercycler Thermal Cyclers using the following profile: an initial step of 4 min at 94°C, followed by 35 cycles of 30 s at 95°C, 45 s at 48°C, and 1 min at 72°C, which was followed by a final extension of 5 min at 72°C. The PCR products were evaluated by placing 3 μl of the product on 1.2% agarose gel. DNA identity was then confirmed via sequencing using an ABI 3730xl sequencer (BGI Life Tech Co. Ltd., Shenzhen, China). The sequences of all haplotypes of the mealybug species were deposited in Gen Bank under accession Nos. KF878037 to KF878073 (COI) (table 1).

### Data analysis

We performed a BLAST search on NCBI (http://www.ncbi.nlm.nih.gov/) after obtaining the sequence of the PCR products. Once the high identity score (above 99%) of the *P. solenopsis* COI gene was obtained, we aligned the sequence data using ClustalW2 (Thompson et al., 1994) at the default parameter settings. Small changes were made to the alignment. The hyper-variable regions were excluded from further analysis because of the ambiguity of the alignment (Swofford et al., 1996).

Unique haplotypes were identified using ARLEQUIN ver. 3.5. Descriptive statistics (number of variable sites, number of haplotypes, haplotype diversity, nucleotide diversity, and average number of nucleotide differences between haplotypes) were calculated using DNASP ver. 5.0.

An AMOVA hierarchical analysis of variance was performed using ARLEQUIN to partition the total variance in its components among groups, among populations, and within
populations based on the groups inferred by the AMOVA analysis. The median joining (MJ) networks of each of the COI genes was constructed using NETWORK ver. 4.6 to examine the evolutionary relationships among haplotypes.

A phylogenetic analysis of the COI based on the maximum likelihood (ML) method was performed separately using MEGA4 (Tamura et al., 2007) under the Kimura 2-parameter model and with the nearest-neighbor-interchange by means of the ML Heuristic method. The reliability of the branches was estimated using 1000 bootstraps.

**Results**

**Amplification result and sequence information**

We successfully amplified the COI in the samples and obtained bands of approximately 800bp on the gel. We aligned the sequence once the PCR products were sequenced. Several base pairs were removed because of ambiguous alignment, which resulted in a final count of 852 COI base pairs. We determined 23 variable sites at the 84, 114, 145, 147, 168, 174, 189, 324, 354, 381, 399, 400, 402, 417, 442, 450, 598, 603, 636, 648, 660, 733, and 735 sites (table 2).

**Population and haplotype division**

The samples from Southern China were found to include two COI haplotypes characterized by 23 variable sites of COI. Two other haplotypes were observed among the samples from the USA, one of which was from California and another in Arizona. Most *P. solenopsis* populations in Southern China possessed haplotype 2. Several samples possessed haplotypes 2 and 3, including those from the Guangdong, Guangxi, and Hainan populations in Southern China, as well as from the Pakistan population, Hanoi population in Vietnam, and Delhi population in India (table 3). To determine the source of the introduction of *P. solenopsis* into China, the COI sequences we analyzed included not only all of the *P. solenopsis* samples we had obtained but also those from the USA and India, which were obtained by downloading homologous mtDNA sequences from Gen Bank.

**Genetic diversity and distances**

Haplotype 2 (populations from Guangdong, Guangxi, Hainan, and Pakistan) was the predominant haplotype in Southern China with a frequency of 0.535. Haplotypes 3 (populations in Guangdong, Guangxi, Hainan, Pakistan, Hanoi, and Delhi), 1 (population in California), and 4 (population in Arizona) had frequencies of 0.326, 0.070, and 0.070, respectively (table 3). The Pakistan population had the highest haplotype diversity (0.66667) and nucleotide diversity (0.00090) among all populations. By contrast, the haplotype and nucleotide diversities had no significant differences among all population from India, Vietnam, and the USA. In Southern China, the haplotype and nucleotide diversities were 0.39916 and 0.00054, respectively. The highest haplotype diversity was found in the Hainan population (0.533571) followed by the Guangdong population (0.43956), and the Guangxi population had the lowest haplotype diversity (0.22222). The average number of nucleotide differences was 3.885 for the species.
The genetic distances among the various populations of *P. solenopsis* varied from 0 to 0.0290 (table 4). No significant genetic differentiation was observed among the populations in Southern China (genetic distances ranged from 0 to 0.0095) as well as among the populations in India, Vietnam, and Pakistan. Conversely, significant genetic differentiation was detected from the population in Arizona.

### Genetic structure

The eight populations were clustered in terms of geographic location as follows: USA (California, Arizona); China (Guangdong, Guangxi, Hainan); South Asia (Pakistan, India); and Southeast Asia (Vietnam). These four geographically well-defined regions are discussed in the following presentation.

The AMOVA revealed that a portion of genetic differentiation was partitioned among groups (59.56% based on COI sequences) and within populations (5.67%), whereas excessive genetic differentiation was observed between populations within each of the four groups identified (34.76%) (table 5). Accordingly, differentiation was highly significant within populations (*F*ST) and among populations between groups (*F*SC).

### MJ networks of haplotypes

MJ networks reconstructed from the haplotypes of the COI gene are shown in fig 2. Except for the USA samples, all samples from Southern China, Pakistan, Vietnam, and India had Hap-2 and Hap-3. Most samples from the Guangdong, Guangxi, and Hainan populations in the China group and two Pakistan samples in the South Asia group shared Hap-2. By contrast, a few samples from the Guangdong, Guangxi, and Hainan populations, one sample from the Vietnam-Hanoi population, one sample from the Pakistan population, and a few samples from the India-Delhi population shared Hap-3. The MJ networks of the COI haplotypes revealed that *P. solenopsis* has spread in Southern China.

### Phylogenetic analysis and the potential distribution patterns of *P. solenopsis* in Southern China

We generated the ML trees using the aligned COI sequence. The tree included two distinct clades: one clade...
consisted of samples from the USA, and another clade consisted of samples from China, Vietnam, India, and Pakistan. However, slight differences among the Guangdong, Guangxi, and Hainan populations in the China group were observed (fig. 3). These data suggest that the invasion of *P. solenopsis* into China may have originated from Pakistan rather than from America. After their invasion into China, the *P. solenopsis* population may have undergone specific genetic-level variations to adapt to the new environment.

**Discussion**

**Variation of *P. solenopsis* in Southern China**

Based on our knowledge, this study is the first to report on the molecular identification of *P. solenopsis* in Southern China based on the mtDNA COI gene. The *P. solenopsis* samples from the Guangdong, Guangxi, and Hainan populations in Southern China were found to include two COI haplotypes characterized by 23 variable sites of COI, which show limited genetic variation in the mealybugs in Southern China.

Although a number of structures were identified using AMOVA, which led to the partitioning of the eight populations being studied into four groups, namely, China, USA, South Asia, and Southeast Asia, the overall level of differentiation was low. Based on AMOVA results, a variability of more than 5% was observed within populations.

Based on the partial sequence variation of COI within the geographical *P. solenopsis* population in Southern China, we can infer that this species has undergone a series of genetic-level variations since it invaded China. To our knowledge, once an invasive species invades a new place, the main selective pressure originates from the change in environment (Sakai et al., 2001). To adapt to this new environment, the invasive species must also change. Moreover, the invasive species can inherit this adaptation. *P. solenopsis* has spread to several other countries since 1991, and a previous research has reported that *P. solenopsis* underwent changes in a new environment (Karar, 2008). Genetic variations in *P. solenopsis* may occur in a new habitat, and this variation may accelerate the invasion of *P. solenopsis*. Thus, questions of whether genetic-level variations in *P. solenopsis* in China and whether this variation affects the expansion of *P. solenopsis* still remain.

**Possible source of *P. solenopsis* in Southern China**

The phylogenetic tree showed that two divergent clades exist in this mealybug species. The major clade contained sequences of *P. solenopsis* from Guangdong, Guangxi, and Hainan in Southern China, Pakistan, Delhi in India, and Hanoi in Vietnam, whereas the other clade-contained sequences of *P. solenopsis* from the USA. The low genetic distance between the Guangdong and Guangxi populations from the China group and the Pakistan population in the South Asia group was 0, whereas the genetic distance between the Guangdong and Guangxi populations from the China group and the Arizona population from the USA group was 0.029. These data show that the *P. solenopsis* found in China may have originated from Pakistan rather than from America. These outcomes are in agreement with previous studies, wherein *P. solenopsis* was analyzed using the mitochondrial COI gene (Chu et al., 2009; Chen et al., 2012).

The genetic distance of *P. solenopsis* ranged from 0 to 0.029 when samples from the China, Pakistan, India, Vietnam,
and USA populations were compared, and the genetic differentiation within populations in Southern China was low. Although *P. solenopsis* invaded China within 5 years, we can still infer that the *P. solenopsis* that invaded China were still in their primary stages.

The genetic data on which the source of the invading populations is based confirm the hypotheses concerning the environmental and evolutionary factors that ensure the success of biological invasions (Sax et al., 2005). A recent study identified the global invasion of imported red fire ant using genetic data (Ascunce et al., 2011). We aim to identify the potential source of *P. solenopsis* in China. Therefore, we compared the genetic data of *P. solenopsis* from Southern China, America, India, Vietnam, and Pakistan. We determined that all of the COI data samples from China had a closer relationship with those from Pakistan rather than those from America. These results suggest that the *P. solenopsis* in China may not have originated from the USA. More data are required to determine whether the *P. solenopsis* in China originated from Pakistan.

**Spread of *P. solenopsis* in Southern China**

Given the rapid development of international trade and tourism, *P. solenopsis* has been widely distributed worldwide. This species has caused severe economic losses over the last two decades in many countries. At present, many countries and regions where *P. solenopsis* are found include Mexico, USA, Cuba, Jamaica, Guatemala, Dominica, Ecuador (Ben-Do, 1994), Panama, Brazil, Chile (Larrain, 2002), Argentina (Granara de Willink, 2003), Nigeria, Benin, Cameroon, New Caledonia, Pakistan, India (Karar, 2008), and Thailand. In China, *P. solenopsis* was first found in 2008 in Guangzhou City (Ma et al., 2009). Since its discovery, *P. solenopsis* has also been observed in Guangxi, Hainan, and other provinces (Wang et al., 2009). The MJ networks of COI haplotypes revealed that *P. solenopsis* has expanded in Southern China.

Based on the COI data, we determined that most of the *P. solenopsis* in Southern China, including those found within the same provinces and those in different provinces, belong to the same clade in the Pakistan tree of *P. solenopsis*. However, the samples obtained from the same provinces displayed slight differences with regard to COI sequence. This finding suggests that since *P. solenopsis* invaded China, this species may have spread from one place to another, which resulted in variations.

The data obtained in this study show that the source of the populations of *P. solenopsis* in Southern China might have originated from Pakistan instead of the USA. In future studies, we should collect more samples from all regions in China where *P. solenopsis* exist and use more types of markers to analyze the genetic differentiation and genetic structure of the species. These studies will help us understand the invasion history, invasion pattern, and source of this invasive population in Southern China. This molecular ecological evolution information on mealybugs may help us eliminate the invasion pathway or introduce biological control agents as well as learn from the measures taken by other countries regarding this matter.

**Acknowledgements**

We express our heartfelt thanks to the editor and two anonymous reviewers, for critical reading, insightful comments, and suggestions on this manuscript. This study was supported by the project of Science & Technology of General Administration of Quality Supervision, Inspection and Quarantine of the People’s Republic of China (grant no. 2012IK271) and Science & Technology of Huizhou City, Guangdong Province, China (grant no. 2013B040090004) and National ‘Twelfth Five-Year’ Plan program for Science & Technology Support (grant no. 2012BAK11B01). The English writing of the manuscript was polished by EnPapers.

**References**

Ascunce, M.S., Yang, C.C., Oakley, J., Calcata, L., Wu, W.J., Shih, C.J., Goudet, J., Ross, K.G. & Shoemaker, D.W. (2011) Global invasion history of the fire ant *Solenopsis invicta*. Science 331, 1066–1068.

Bellemain, E., Carlsen, T., Brochmann, C., Coissac, E., Taberlet, P. & Kauserud, H. (2010) ITS as an environmental DNA barcode for fungi: an in silico approach reveals potential PCR biases. BMC Microbiology 10, 189.

Ben-Dov, Y. (1994) *A Systematic Catalogue of the Mealybugs of the World* (Insecta: Homoptera: Coccoidea: Pseudococcidae and Ploiodei) with Data on Geographical Distribution, Host plants, Biology and Economic Importance. London, Intercept Limited.

Chen, Z., Zhang, J., Fu, H.F., Xu, Z.Z., Deng, K.Z. & Zhang, J.Y. (2012) On the validity of the species *Phenacoccus solenopsis* based on morphological and mitochondrial COI data, with the description of a new body color variety. Biodiversity Science 20, 443–450.

Chu, D., Liu, G.X. & Fu, W. (2009) Phylogenetic analysis of mtCOI reveals the cryptic lineages in *Phenacoccus solenopsis* complex (Hemiptera:Pseudococcidae). Acta Entomologica Sinica 52, 1261–1265.

Culik, M.P. & Gullan, P.J. (2005) A new pest of tomato and other records of mealybugs (Hemiptera: Pseudococcidae) from Espirito Santo, Brazil. Zootaxa 964, 1–8.

Fuchs, T.W., Stewart, J.W., Minzennayer, R. & Rose, M. (1991) First record of *Phenacoccus solenopsis* Tinsley in cultivated cotton in the United States. Southwestern Entomologist 16, 215–221.

Granara de Willink, M.C. (2003) Nuevas citas y huéspedes de *Phenacoccus* para la Argentina (Hemiptera: Pseudococcidae). Revista de la Sociedad Entomológica Argentina 62(3–4), 80–82.

Havill, N.P., Foottit, R.G. & von Dohlen, C.D. (2007) Evolution of host specialization in the Adelgidae (Insecta: Hemiptera) inferred from molecular phylogenetics. *Molecular Phylogenetics and Evolution* 44, 357–370.

Hebert, P.D.N., Cywinska, A. & Ball, S.L. (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London Series B: Biological Sciences* 270, 313–321.

Karar, H. (2008) *Phenacoccus solenopsis* Tinsley ( Sternorrhyncha: Coccoidea: Pseudococcidae), an invasive mealybug damaging cotton in Pakistan and India, with a discussion on seasonal morphological variation. *Zootaxa* 1913, 1–35.

Kress, W.J., Wurdack, K.J., Zimmer, E.A., Weigt, L.A. & Janzen, D.H. (2005) Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences of the United States of America* 102, 8369–8374.

Larrain, S. (2002) Insect and mite pest incidence on sweet peppino (*Solanum muricatum* Ait) cultivated in the fourth Region, Chile. *Agricultura Técnica* 62, 15–26.
Lee, C.E. (1999) Rapid and repeated invasions of fresh water by the copepod *Eurytemora affinis*. *Evolution* 53, 1423–1434.
Lee, C.E. (2002) Evolutionary genetics of invasive species. *Trends in Ecology and Evolution* 17, 386–391.
Ma, J., Hu, X.N., Liu, H.T. & Peng, Z.Q. (2009) *Phenacoccus solenopsis* Tinsley was found on *Hibiscus rosa-sinensis* in Guangzhou. *Plant Quarantine* 23, 35–36.
Maillet, J. & Lopez-Garcia, C. (2000) What criteria are relevant for predicting the invasive capacity of a new agricultural weed? The case of invasive American species in France. *Weed Research – Oxford* 40, 11–26.
Neuffer, B. & Hurka, H. (2002) Colonization history and introduction dynamics of *Capsella bursa-pastoris* (Brassicaceae) in North America: isozymes and quantitative traits. *Molecular Ecology* 8, 1667–1681.
Pons, J., Barraclough, T.G., Theodorides, K., Cardoso, A. & Vogler, A.P. (2004) Using exon and intron sequences of the gene *Mp20* to resolve basal relationships in *Cicindela* (Coleoptera: Cicindelidae). *Systems Biology* 53, 554–570.
Rhymer, J.M. & Simberloff, D. (1996) Extinction by hybridization and introgression. *Annual Review of Ecology and Systematics* 27, 83–109.
Sakai, A.K., Allendorf, F.W., Holt, J.S., Lodge, D.M., Molofsky, J., With, K.A., Baughman, S., Cabin, R.J., Cohen, J.E., Ellstrand, N.C. & McCauley, D.E. (2001). The population biology of invasive species. *Annual Review Ecology Systematics* 32, 305–332.
Sax, D.F., Stachowicz, J.J. & Gaines, S.D. (2005) *Species Invasions: Insights into Ecology, Evolution and Biogeography*. Sunderland, MA, Sinauer Associates.
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. *Annals of the Entomological Society America* 87, 651–701.
Swofford, D.L., Olsen, G.J., Waddell, P.J. & Hillis, D.M. (1996) Phylogenetic inference. pp. 407–514 in Hillis, D.M., Moritz, C. & Mable, B.K. (Eds) *Molecular Systematics*. Sunderland, MA, Sinauer Associates.
Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software, version 4.0. *Molecular Biology and Evolution* 24, 1596–1599.
Thompson, J.D., Higgins, D.G. & Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 22, 4673–4680.
Tsutsui, N.D., Suarez, A.V., Holway, D.A. & Case, T.J. (2000) Reduced genetic variation and the success of an invasive species. *Proceedings of the National Academy of Sciences of the United States of America* 97, 5948–5953.
Wang, Y.P., Wu, S.A. & Zhang, R.Z. (2009) Pest risk analysis of a new invasive pest, *Phenacoccus solenopsis*, to China. *Chinese Bulletin of Entomology* 46, 101–106.