

# Global phylogenetic relationships, population structure and gene flow estimation of *Trialeurodes vaporariorum* (Greenhouse whitefly)

J.M. Wainaina<sup>1\*</sup>, P. De Barro<sup>2</sup>, L. Kubatko<sup>3</sup>, M. A. Kehoe<sup>4</sup>, J. Harvey<sup>5,6</sup>, D. Karanja<sup>7</sup> and L. M. Boykin<sup>1</sup>

<sup>1</sup>The University of Western Australia, Australian Research Council Centre of Excellence in Plant Energy Biology and School of Molecular Sciences, Crawley, Perth 6009, Western Australia, Australia: <sup>2</sup>CSIRO, GPO Box 2583, Brisbane QLD 4001, Australia: <sup>3</sup>The Ohio State University 12th Avenue Columbus, Ohio, USA: <sup>4</sup>Departments of Agriculture and Food Western Australia, South Perth WA 6151, Australia: <sup>5</sup>Feed the Future Innovation Lab for the Reduction of Post-Harvest Loss, Kansas State University, Manhattan, Kansas, USA: <sup>6</sup>Biosciences eastern and central Africa (BeCA-ILRI Hub)–International Livestock Research Institute (ILRI), Nairobi, Kenya: <sup>7</sup>Kenya Agriculture and Livestock Research Organization (KARLO) Box 340-90100, Machakos, Kenya

## Abstract

*Trialeurodes vaporariorum* (Westwood, 1856) (Greenhouse whitefly) is an agricultural pest of global importance. It is associated with damage to plants during feeding and subsequent virus transmission. Yet, global phylogenetic relationships, population structure, and estimation of the rates of gene flow within this whitefly species remain largely unexplored. In this study, we obtained and filtered 227 GenBank records of mitochondrial cytochrome c oxidase I (mtCOI) sequences of *T. vaporariorum*, across various global locations to obtain a final set of 217 GenBank records. We further amplified and sequenced a ~750 bp fragment of mtCOI from an additional 31 samples collected from Kenya in 2014. Based on a total of 248 mtCOI sequences, we identified 16 haplotypes, with extensive overlap across all countries. Population structure analysis did not suggest population differentiation. Phylogenetic analysis indicated the 2014 Kenyan collection of samples clustered with a single sequence from the Netherlands to form a well-supported clade (denoted clade 1a) nested within the total set of sequences (denoted clade 1). Pairwise distances between sequences show greater sequence divergence between clades than within clades. In addition, analysis using migrate-n gave evidence for recent gene flow between the two groups. Overall, we find that *T. vaporariorum* forms a single large group, with evidence of further diversification consisting primarily of Kenyan sequences and one sequence from the Netherlands forming a well-supported clade.

---

\*Author for correspondence

Phone: +61469733615

Fax: +61864884401

E-mail: [james.wainaina@research.uwa.edu.au](mailto:james.wainaina@research.uwa.edu.au);

[jwmbora@yahoo.com](mailto:jwmbora@yahoo.com)

**Keywords:** *Trialeurodes vaporariorum*, population structure, phylogenetics gene flow, Kenya

(Accepted 23 March 2017; First published online 23 May 2017)

Whiteflies are important agricultural pests with a global distribution (Anderson *et al.*, 2004; Lapidot *et al.*, 2014). Their agricultural importance is associated with damage caused during feeding on the plant phloem and in the production of honeydew (Colvin *et al.*, 2006; Prijović *et al.*, 2013). Honeydew serves both to reduce transpiration and as an inoculation point for saprophytic fungi, which grow over the leaf surface reducing photosynthesis (Colvin *et al.*, 2006). Among the most economically important whiteflies is the greenhouse whitefly, *Trialeurodes vaporariorum* (Westwood, 1856), which transmits a number of plant viruses of the genera *Crinivirus* and *Torradovirus* (Navas-Castillo *et al.*, 2011, 2014). Of these, the *Criniviruses*, Tomato chlorosis virus (ToCV) and Tomato infectious chlorosis virus (TICV), are of major economic importance in tomato production globally (Wisler, 1998; Wintermantel *et al.*, 2009). In Africa, ToCV and TICV have been reported from Morocco and South Africa (EPPO, 2005) and more recently from greenhouses in Sudan (Fiallo-Olivé *et al.*, 2011). The viruses are able to infect a wide host range beyond tomato (Fortes *et al.*, 2012). Production losses attributed to *T. vaporariorum* transmitted viruses such as the sweet potato chlorotic stunt virus are a major constraint on sweet potato production in Kenya, with uninfected sweet potatoes yielding 50% more production than infected plants (Ateka *et al.*, 2004; Miano *et al.*, 2008).

#### *Phylogeographical structure of T. vaporariorum in Sub-Saharan Africa (SSA)*

Though *T. vaporariorum* is globally distributed, few studies have explored the phylogeographical structuring and genetic diversity of *T. vaporariorum* populations from sub-Saharan Africa (SSA) and in particular Kenya (Mound & Halsey, 1978; Njaramb, 2000; Kapantaidaki *et al.*, 2015). The majority of phylogeographical studies of *T. vaporariorum* have been confined to temperate regions (Roopa *et al.*, 2012; Prijović *et al.*, 2014; Kapantaidaki *et al.*, 2015). Most of these studies showed a lack of distinct phylogeographical structuring coupled with low genetic diversity.

Roopa *et al.* (2012) used both mitochondrial cytochrome c oxidase I (mtCOI) and a nuclear marker (ITS-1) to assess the genetic diversity and phylogeographical structuring of *T. vaporariorum* within populations in India. They reported no significant differences among *T. vaporariorum* populations. Kapantaidaki *et al.* (2015) supported these findings using a combination of mtCOI and secondary endosymbionts. In addition, Prijović *et al.* (2013) reported a similar lack of phylogeographical structuring across populations in Serbia and surrounding countries. However, some studies have reported population structuring in different regions, with several drivers implicated as influencing population structure. In particular, Ovčarenko *et al.* (2014) reported population structuring in Finland and attributed this to habitat, specifically to field vs. greenhouses populations. Gao *et al.* (2014) reported population structuring across the different regions of China and suggested the primary driver was multiple introduction points into China. There have been no similar studies in SSA.

#### *Assessment of global population structure*

On the basis of these studies we posed the following question: what is the phylogenetic relationship of *T. vaporariorum* samples from SSA, and in particular Kenya, compared with global populations?

To help answer this question, we carried out a field survey in 2014 across smallholder farms located in the Kenyan highlands, which produce primarily common bean (*Phaseolus vulgaris*, var *humilis Alef*). Using *T. vaporariorum* mtCOI sequences of both the Kenyan *T. vaporariorum* sequences and global GenBank records, we assessed the following: haplotype distribution and phylogenetic and population structuring within *T. vaporariorum* sequences. In addition, we estimated the rate of gene flow using Bayesian estimation. We report for the first time global *T. vaporariorum* population, phylogenetic and haplotype distributions and gene flow estimates in this invasive pest.

## Materials and methods

### *Sample collection*

Sample collection of whiteflies was carried out on private property, following informed consent from the landowners. Adult whiteflies were collected on common beans (*Phaseolus vulgaris*) growing within smallholder farms in Kenya. Sampling was done in the short rain season of November 2014, across three counties in Katumani/Machakos, Kiambu and Kajiado (table 1). In each farm, common bean plants were examined for the presence of whiteflies across a 'z' transect. Whiteflies were collected using a mouth aspirator and aspirated directly into a falcon tube (BD, Biosciences, USA) with RNAlater (Sigma, USA; cat no. R0901). Whiteflies were then shipped to the University of Western Australia and stored at room temperature.

### *Genomic DNA isolation*

Genomic DNA was extracted from individual whiteflies collected from each of the three sites using the Zymo Insect and Tissue extraction kit (Zymo, USA) as described by the manufacturer. Individual whiteflies were added into a ZR Bashing bead lysis tube containing 750 µl lysis solution. Bashing beads were then mounted on a Fast Prep<sup>®</sup>-24 MP bead beater and ground at 6 M/S for 40 s. Bashing beads were centrifuged at 10,000 × g for 1 min. Subsequent steps were as described by the manufacturer, with the only modification being the elution of genomic DNA with 30 µl of nuclease-free water followed by a centrifugation step at 10,000 × g for 30 s. Genomic DNA was stored at -20°C until analysis was carried out.


### *Polymerase chain reaction (PCR) and sequencing*

Mitochondria COI *T. vaporariorum*-specific primers COI-F: 5'-GCCTGGTTTGGCATT-3' and the reverse primer COI-R:



Fig. 1. Global map showing source of *Trialeurodes vaporariorum* sequences from GenBank and collected in this study.

Source: [www.premiumslides.com/powerpoint/world-map-kit.html](http://www.premiumslides.com/powerpoint/world-map-kit.html)

Key:  *Trialeurodes vaporariorum* samples included in this study.

5'-GCTTATTTAGCACCCACTCTA-3' were used to amplify a ~750 bp of mtCOI gene (Gao, *et al.*, 2014). PCR was carried out in a total reaction mix of 20  $\mu$ l with 5  $\mu$ l of whitefly DNA as template, 10  $\mu$ l Bulk AccuPower<sup>(R)</sup> PCR Premix containing 1 U of Top DNA polymerase, 250  $\mu$ M of dNTP (dATP, dCTP, dGTP, dTTP), 10 mM Tris-HCl (pH9.0), 30 mM KCl, 1.5 mM MgCl<sub>2</sub>, 1  $\mu$ l of 5  $\mu$ M forward and 5  $\mu$ M reverse primers and 3  $\mu$ l of nuclease free water. The PCR reaction was run on an eppendorf thermocycler (Eppendorf, Hamburg, Germany) under the following conditions; initial denaturation for 2 min at 94°C followed by 35 cycles of 94°C for 20 s, annealing for 30 s at 52°C, elongation for 1 min at 72°C, final elongation for 30 min at 72°C. PCR amplicons were separated in a 2% agarose gel pre-stained with gel red (Biotium, CA, USA) and run in 0.5 $\times$  Tris-Boric EDTA (TBE) electrophoresis buffer for 35 min at 100 V and detected under UV light (Biorad, CA, USA). PCR products were purified using Qiaquick PCR purification kit (Qiagen Inc.) as described by the manufacturer. Purified PCR products were sent for bi-directional Sanger sequencing at Macrogen Korea.

#### Sequence alignment and GenBank data retrieval

A total of 31 forward and reverse sequences of the 3' end of mtCOI were obtained. Raw chromatograms were visually inspected using Geneious version 8 (<http://www.geneious.com>, Kearse *et al.*, 2012) across all the bases, with ambiguities corrected. Consensus sequences were generated for each whitefly, and used for subsequent analysis. Basic local alignment search (BLAST) was carried out to confirm that all sequences were *T. vaporariorum*. In addition, 227

*T. vaporariorum* mtCOI sequences were retrieved from GenBank and assigned to geographical location based on GenBank metadata file records and supporting information in publication. We filtered these sequences and excluded sequences where geographical location was missing and those that were representative haplotypes rather than individual sequences leaving a total of ( $n = 217$ ) sequences for analysis. *T. vaporariorum* from each country were mapped on a global map (fig. 1). mtCOI sequences retrieved from GenBank were combined with sequences from this study (Supplementary table 1). The combined sequences were aligned using MAFFT in Geneious version 8 with default parameters (Kearse *et al.*, 2012). Manual inspection, adjustment of gaps and trimming of overhangs was carried out.

#### Haplotypes networks based on mtCOI sequences of *T. vaporariorum*

Associations between global haplotypes were determined using median joining algorithms implemented in NETWORK version 5.0 (Bandelt *et al.*, 1999) with two outgroups, *T. ricin* and *T. lauri*, included under default parameters.

#### Phylogenetic analysis

Selection of the optimal model for phylogenetic analysis was determined using jModeltest 2 (Darriba *et al.*, 2012), using 11 substitution schemes. The optimal model selected based on the Akaike information criterion was TPM2uf + G, while the optimal model using the Bayesian information criterion was HKY + I. The HKY + I model was used for subsequent analysis

Table 1. *Trialeurodes vaporariorum* location and host plants collected in this study.

*Location	Coordinates	Host plant	No of samples sequenced	GenBank Accession numbers
Kajiado	1°32.454S 36°44.573E	<i>Phaseolus vulgaris</i> (Common Bean)	2	KX058204, KX058211
Katumani	1°34.95S 37°14.503E	<i>P. vulgaris</i> (Common Bean)	4	KX058202 KX058205 KX058207 KX058214
Kiambu	1°7.781S 36°31.198E	<i>P. vulgaris</i> (Common Bean)	25	KX058201-KX058203 KX058206 KX058208-KX058209 KX058210 KX058212 KX058213 KX058215-KX058231

incorporates variable base frequencies and separate transition and transversion rates. MrBayes was run on the Magnus Supercomputer at the Pawsey Supercomputer Centre at the University of Western Australia using 76 core/h (274230s). Bayesian analysis was carried out for two independent runs using the command 'lset nst = 2 rates = gamma' for 50 million generations, with trees sampled every 1000 generations. In each of the runs, the first 25% (2500) trees were discarded as burn-in. Convergence and mixing of the chains were evaluated using Tracer v1.6 (Rambaut *et al.*, 2014). LnL had an effective sample size (ESS) of greater than 2000 a mean of -1509.4812, and stdev 11.0176. Maximum-likelihood (ML) phylogenetic trees were estimated under the HKY+I model in PAUP\* 4.0a147 (Swofford, 2003), with bootstrap support based on 100 replicates.

#### Estimates of evolutionary divergence between sequences across clades

To assess the extent of divergence between the main group (1) and a nested group (1a), we computed the proportion of base difference between representative sequences in the two groups using *p* distances implemented in MEGA 5 (Tamura *et al.*, 2011) (fig. 4).

#### Evaluation of *T. vaporariorum* population structure using Bayesian analysis

To determine the population structure of *T. vaporariorum* mtCOI sequences across different geographical locations we used Structurama 2.0 (Huelsenbeck *et al.*, 2011), which implements a Dirichlet process prior for the number of populations. The default parameters of Gamma (0.100, 10.000) were run, and the model without admixture was used. MCMC was run for one million generations and samples were collected every 1000 generations. Two independent runs were carried out to assess whether the MCMC algorithm converged to the same values (table 2).

#### Bayesian estimation of the rate gene flow between phylogenetic groups

To evaluate gene flow between *T. vaporariorum* found in clades 1 and 1a, we carried out Bayesian estimation of migration rates using migrate-n v 3.6.11 (Beerli & Palczewski, 2010). We evaluated four models: (a) model one assumed *T. vaporariorum* sequences were from a single population (b) model two

Table 2. Posterior probabilities on the number of populations (K) within *Trialeurodes vaporariorum* for the COI sequences across all geographical locations using Structurama 2.0.

Number of population (K)	Posterior probability
1	0.67
2	0.28
3	0.05
4	0.00

assumed *T. vaporariorum* migration from clades 1a to 1 (c) model three assumed *T. vaporariorum* migration from clade 1 to the nested clade 1a, and (d) model four assumed migration from clades 1 to 1a and the reverse. To test our models, we used the following search parameters: short chains 10, (10,000 genealogies) long chains = 4, genealogies (100,000) with a burn-in of 10,000. Evaluation of the most suitable models was based on the highest model probabilities and the natural log of Bayes factor (LBF).

## Results

#### Distribution of *T. vaporariorum* haplotypes across global collections

In this study, 31 *T. vaporariorum* sequences were generated and deposited in GenBank under accession numbers KX058201–KX058231 (table 1). An additional 217 sequences from GenBank (after the filtering step) were combined with our sequences for subsequent analysis. All sequences and accession numbers from GenBank and generated from this study are provided in Supplementary table 1. Analysis across the combined global dataset based on the 248 sequences, including two outgroups revealed 18 haplotypes, with 16 unique haplotypes within *T. vaporariorum* and two haplotypes from the outgroups *T. lauri* and *T. ricini* (fig. 2).

#### Median joining networks based on mtCOI sequences of *T. vaporariorum*

Assessments of genetic relationships among the 18 haplotypes identified were based on a median joining network of the mtCOI sequences (fig. 2), showing two outgroup haplotypes (red circles), while the *T. vaporariorum* samples are shown by green and yellow circles. Haplotypes 2 (HAP\_2) and 3 (HAP\_3) were the dominant haplotypes across all geographical

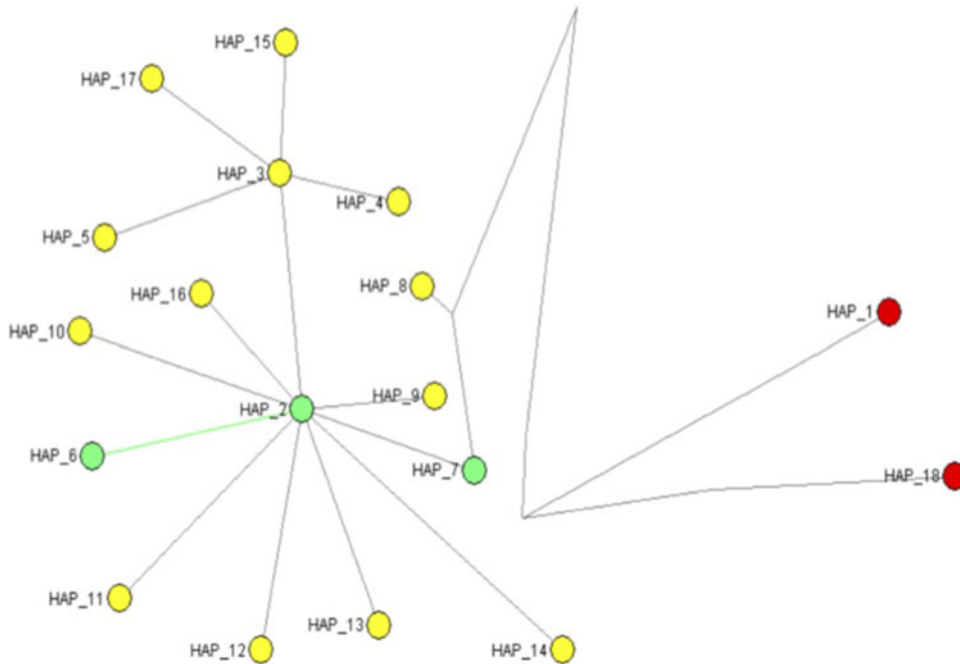


Fig. 2. Median joining network generated using the program NETWORK version 5.0 for 16 mtCOI haplotypes of *Trialeurodes* including two outgroup sequences (*Trialeurodes ricin*, *Trialeurodes lauri*). \*Haplotypes in circulation across global location are highlighted in yellow, those present in Kenya are highlighted in green and the outgroups are in red. \*Number of individual sequences within the two dominant haplotypes were Haplotype 2 ( $n = 26$ ), Haplotype 3 ( $n = 201$ ).

locations, while within Kenya haplotype HAP\_2, HAP\_7 and HAP\_8 were present (fig. 2).

#### Phylogenetic analysis of *T. vaporariorum* and assessment of sequence divergence

Phylogenetic relationships of *T. vaporariorum* across geographical locations were assessed using Bayesian phylogenetic inference as implemented in MrBayes 3.2.2 (Huelsenbeck *et al.*, 2011) and maximum-likelihood implemented in PAUP\* (Swofford, 2003). Phylogenetic trees from both methods (fig. 3 and Supplementary figs 1 and 2) yielded one main clade (denoted clade 1) and a nested clade (denoted clade 1a) containing a subset of Kenyan and one sequence from the Netherlands that was supported with >0.9 posterior probability and 50% bootstrap support, respectively. *T. ricin* and *T. lauri* were used as the outgroup for rooting the tree (Kinene *et al.*, 2016). The two outgroups selected were the closest recent relatives of *T. vaporariorum* based on the NCBI taxonomic classification database. To evaluate evolutionary divergence of sequences between clades 1 and 1a, we assessed pairwise nucleotide sequence comparisons using  $p$  distance (fig. 4). Pairwise comparison showed differences ranging between 0.023 and 0.0068 between clades, while distances within clades were <0.00001 (fig. 4).

#### Bayesian analysis of *T. vaporariorum* population structure

Population structure of *T. vaporariorum* was evaluated using Structurama 2.0, with default parameter settings and no admixture. All of the *T. vaporariorum* sequences clustered as one population ( $K = 1$ ), supported by a posterior probability of 0.67 (table 2), with posterior probability of 0.28 for  $K = 2$

populations and 0.05 for  $K = 3$  populations. All other numbers of populations received posterior probability of <0.01.

#### Estimation of the rate of gene flow

Estimation of the rate of gene flow between *T. vaporariorum* from the two groups on the phylogenetic tree was assessed using four hypotheses. Assessment of the most probable models was based on model probability and the natural LBF (table 3). The optimal model was three that had the highest model probability of 0.879622055 and log marginal likelihood -10842.01932. It hypothesised gene flow from individuals found in clade 1 to those found within the nested clade 1a of the phylogenetic tree (table 3).

## Discussion

In this study, we assessed the phylogeographical structuring and genetic diversity of *T. vaporariorum* globally. Phylogeographical structuring resulted in one main clade (1) with global sequences and a nested clade (1a) containing a subset of Kenyan and one sequence from the Netherlands. Population structure analysis identified a single population ( $K = 1$ ) across all *T. vaporariorum* sequences. Sequence pairwise distance (fig. 4) shows signs of sequence divergence between groups and minimal divergence within groups. This was further assessed through the estimation of gene flow between *T. vaporariorum* sequences clustering within the two groups on the phylogenetic tree. Bayesian estimation using migration showed evidence for gene flow from clade 1 to the nested clade 1a on the phylogenetic tree.

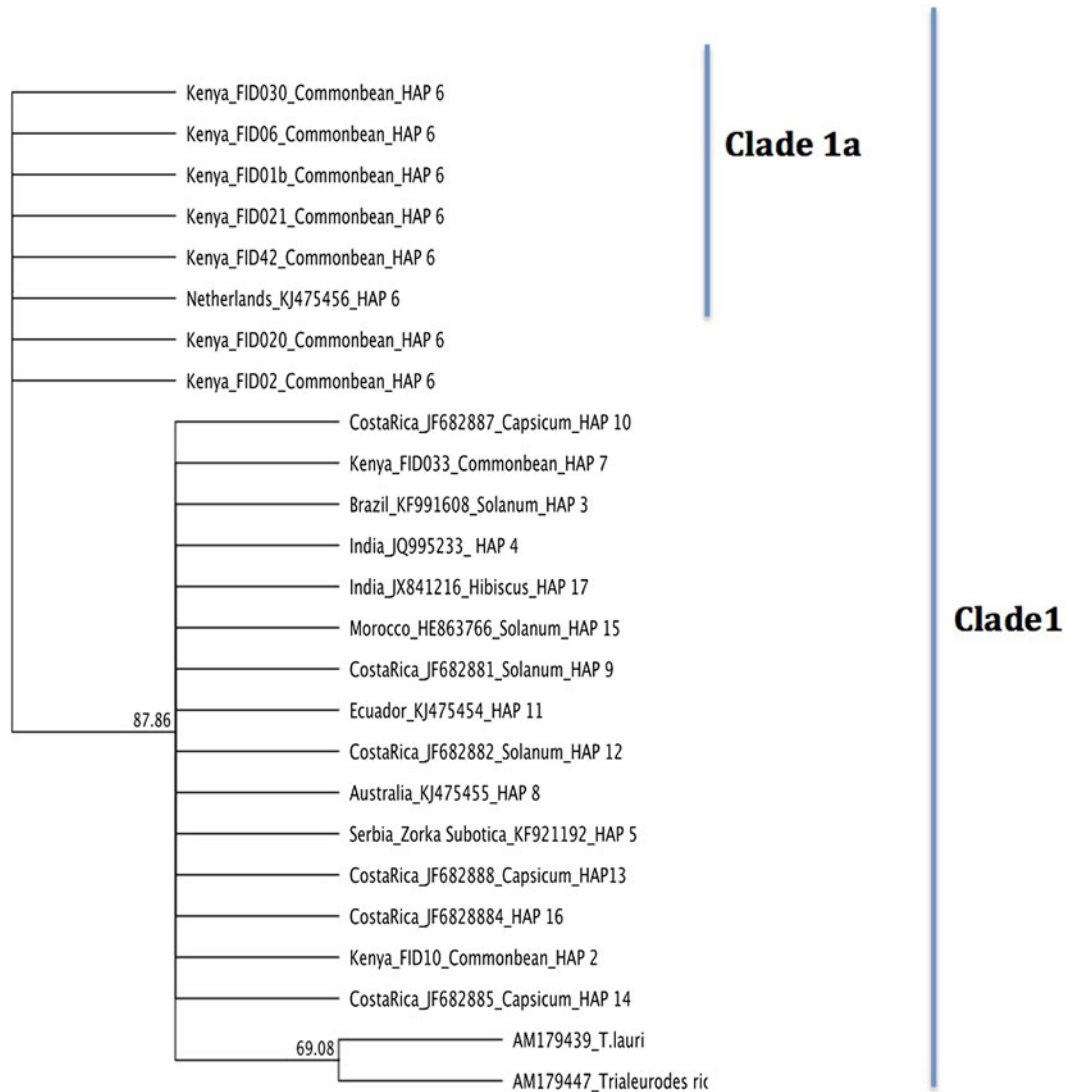


Fig. 3. Bayesian phylogenetic relationships of representative haplotype sequences of *T. vaporariorum* and two outgroups (*T. ricin*, *T. lauri*) generated with MrBayes 3.2.2. HAP: Haplotype numbers assigned for each tip.

Key: 1: The main group. 1a: Nested group within the main group.

### Haplotype networks

The relationship among the *T. vaporariorum* haplotypes (fig. 2) across different global geographical location was based on the median joining network (fig. 2). The two main nodes (HAP\_2, HAP\_3) (fig. 2) were present across all geographical locations. Within Kenyan *T. vaporariorum* sequences three haplotypes were identified in circulation (HAP\_2, HAP\_7, HAP\_8) (fig. 2), while Haplotype\_8 was found as a unique haplotype within Kenya. Kenyan and Netherlands sequences found within the nested clade 1a were from haplotype HAP\_7.

Our findings report fewer haplotypes than previously reported among *T. vaporariorum* (Prijošević *et al.*, 2014). This is likely due to the selection criteria used in this study. However a majority of the haplotypes identified overlap across

Table 3. Bayesian estimation of migration rates across all *Trialeurodes vaporariorum* sequences using Migrate 3.6.

Model	Bezier ImL (log marginal likelihood)	Model probability	LBF
Hypothesis 1	-10887.62978	<0.00001	-91.220914
Hypothesis 2	-10844.00818	0.120377945	-3.977714
Hypothesis 3	-10842.01932	0.879622055	-
Hypothesis 4	-110032.70967	<0.00001	-323.380696

geographical locations and could indicate a single source of a *T. vaporariorum* lineage that is easily adapted to different niche habitats (Kapantaidaki *et al.*, 2015). It is probable that continuous expansion of *T. vaporariorum* haplotypes across global locations and specifically from this study could be two-

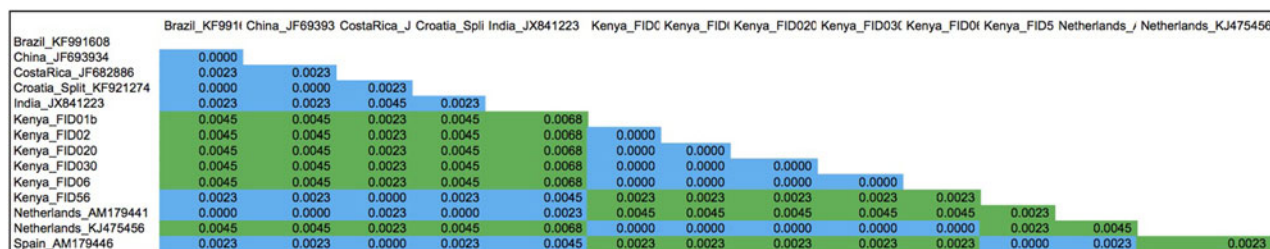


Fig. 4. *P* distances calculated between *Trialeurodes vaporariorum* nucleotide sequences clustering on clade 1 and nested clade 1a on the phylogenetic tree using MEGA 5.1. The proportion of base differences per site between sequences is shown. The analysis involved 14 nucleotide sequences. All ambiguous positions were removed for each sequence pair. There were a total of 443 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (Tamura *et al.*, 2011).

Key: (i) Comparison within group clade 1 and clade 1a ■■■. (ii) Comparison between groups clades 1 and 1a ■■■.

fold. First, there has been an increase in global trade on horticultural products, such as cut flowers from Kenya (Ulrich, 2014). This has resulted in increased importation of flower cuttings as propagation material from temperate ecoregions. Flower cuttings are potential carriers of *T. vaporariorum* because the eggs are difficult to detect, especially when handling large consignments, but readily hatch under ideal climatic conditions. This mode of *T. vaporariorum* movement has been implicated as one of the main modes of distribution within invasive pests globally (Grapputo *et al.*, 2005). Secondly, proliferation of polyhouses for horticulture production potentially provides appropriate microclimatic conditions sufficient for the emergence of *T. vaporariorum* and subsequent release into the open field. The importance of the polyhouse in building up *T. vaporariorum* populations and controlling gene flow has previously been reported (Roopa *et al.*, 2012). However, distinct haplotypes present within Kenya (HAP\_7), Costa Rica (HAP\_12, HAP\_13, HAP\_14), and India (HAP\_17) could be indications of local adaptations of *T. vaporariorum* within these regions.

#### Phylogeographical structuring and evolutionary divergence of *T. vaporariorum* sequences

The evolutionary relationships among the *T. vaporariorum* sequences across geographical locations were reconstructed using both maximum-likelihood and Bayesian phylogenetic analysis. Both phylogenetic trees resulted in one main clade 1 and a nested clade 1a (fig. 3 and Supplementary fig. 1). Clade 1 was composed of *T. vaporariorum* samples from all the geographical locations, while the nested clade 1a was composed of a subset of *T. vaporariorum* sequences from Kenya and one Netherlands sample. This is contrary to previous phylogenetic relationships of *T. vaporariorum*, where a single group was observed across *T. vaporariorum* sequences collected (Roopa *et al.*, 2012; Prijović *et al.*, 2014; Kapantaidaki *et al.*, 2015). In this study, we report the existence of one main clade (1) and a well-supported nested clade 1a within *T. vaporariorum*. We further assessed the pairwise differences between representative sequences from clade 1 and nested clade 1a, with nucleotide pairwise differences ranging between 0.0023 and 0.0068 between clades, while distances within clades were <0.00001 (fig. 4).

We postulate, based on the nucleotide pairwise differences between clade 1 and the nested clade 1a within the phylogenetic tree, that this divergence could be an early indicator of

segregation leading to an early differentiating species within the nested clade. Further information may be obtained through the additional sequencing of nuclear genes within *T. vaporariorum* from SSA. Thus over time, with more sampling, and an increased number of genetic markers, this might result in the identification of distinct *T. vaporariorum* populations.

#### Population structure based on Bayesian analysis

Population structure analysis of *T. vaporariorum* using Structurama showed the strongest support for one population (posterior probability 0.67; table 3). Population structure results of one population differed with the phylogenetic tree of one main group (clade 1), and subclade 1a within *T. vaporariorum* (fig. 3 and Supplementary fig. 1). Our findings concur with (Gao *et al.*, 2014) where they report lack of geographical structuring within the *T. vaporariorum* populations though several populations were reported. This further supports our hypothesis of a single source of *T. vaporariorum* introduced into SSA and other non-native *T. vaporariorum* range areas.

#### Gene flow estimation and population subdivision

To further evaluate phylogenetic and population structure results of *T. vaporariorum* globally, we used isolation by migration models. We used the Bayesian estimation method implemented in migrate-n to estimate migration rates and evaluated the models based on model probabilities and LBF. Our findings indicate there was gene flow (table 3) between sequences in clade 1 and those in clade 1a (Kenyan and Netherlands samples). Unrestricted gene flow between clades is an indication that sequences on clades 1 and 1a were not sufficiently distinct to be considered as distinct species, but also rejected the hypothesis that they all belonged to the same panmictic population. However, LBF comparisons between models were low. These findings were in agreement with the phylogenetic analysis but contrary to Structurama output ( $K = 1$ ) that considered all sequences to belong to one panmictic population. The difficulty in resolving the actual population by Structurama could be due to the few loci used in the study coupled with low sequence variation across mtCOI sequences, as has been previously observed in other studies (Wilkinson *et al.*, 2011).

Our findings highlight differences in gene flow between populations in *T. vaporariorum* vs. *B. tabaci*. *B. tabaci* is recognized as a species complex of 34 species that are morphologically indistinguishable (Boykin *et al.*, 2007; De Barro & Ahmed, 2011;

Boykin & De Barro, 2014). *B. tabaci* putative species exhibit differences in host range, phylogeographical clustering and potential to transmit viruses (Abdullahi *et al.*, 2003; Sseruwagi *et al.*, 2006, De Barro & Ahmed, 2011; Ashfaq *et al.*, 2014; Legg *et al.*, 2014). These differences are yet to be reported within *T. vaporariorum*, and our findings indicate the need for further investigation of *T. vaporariorum* within new geographic locations.

Future studies should include more expansive sampling outside native *T. vaporariorum* range areas of Europe and into new geographical locations coupled with a multilocus approach to improve our understanding of this highly invasive pest. This will provide for more robust and stringent phylogenetic, population and gene flow estimates to examine the possible existence of species divergence within *T. vaporariorum*. However, it is probable that the existence of a differentiating nested clade 1a is an early indicator of divergence within *T. vaporariorum* potentially driven by phylogeographical factors as observed within *B. tabaci* and *T. vaporariorum* in Europe (Boykin *et al.*, 2013; Ovčarenko *et al.*, 2014). In addition, differences in the bacterial endosymbionts among *T. vaporariorum* haplotypes might induce genetic sweeps within the mitochondria and thus drive the diversification of *T. vaporariorum* (Kapantaidaki *et al.*, 2015).

### Conclusion

This study provides the first comprehensive analysis of the genetic diversity, phylogenetic and population structure of *T. vaporariorum* coupled with gene flow estimation globally. Our findings provide some evidence of *T. vaporariorum* genetic divergence across geographical regions, but this is not sufficient to suggest species complex status and does not demonstrate sufficient isolation to indicate distinct populations. Our findings indicate the need for future research on understanding the drivers of genetic divergences within *T. vaporariorum* across different geographical regions.

### Supplementary Material

The supplementary material for this article can be found at <https://doi.org/10.1017/S0007485317000360>

### Acknowledgements

J.M.W is supported by an Australian Award scholarship by the Department of Foreign Affairs and Trade (DFAT), this work forms part of his PhD research. Fieldwork was supported by the kind consideration of Dr John Carr through a BBSRC GRANT: BBSRC GRANT: BB/J011762/1 and the Bioscience eastern and central Africa (BecA-ILRI Hub). Laboratory work was performed at the Australian Research Council Centre of Excellence Plant Energy Biology (PEB), The University of Western Australia. We thank Dr Charles Kariuki Director Kenya Agricultural and Livestock Research Organization (Katumani) for facilitating sample collection in Eastern Kenya and insightful discussions. Supercomputer analysis was supported by resources provided by the Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia. We acknowledge the contribution of the anonymous reviewers who greatly improved this manuscript.

### References

- Abdullahi, I., Winter, S., Atiri, G.I. & Thottappilly, G. (2003) Molecular characterization of whitefly, *Bemisia tabaci* (Hemiptera: Aleyrodidae) populations infesting cassava. *Bulletin of Entomological Research* **93**, 97–106. doi: 10.1079/BER2003223.
- Anderson, P.K., Cunningham, A.A., Patel, N.G., Morales, F.J., Epstein, P.R. & Daszak, P. (2004) Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends in Ecology and Evolution* **19** (10), 535–544. doi:10.1016/j.tree.2004.07.021
- Ateka, E.M., Njeru, R.W., Kibaru, A.G., Kimenju, J.W., Barg, E., Gibson, R.W. & Vetten, H.J. (2004) Identification and distribution of viruses infecting sweet potato in Kenya. *Annals of Applied Biology* **144**, 371–379. doi: 10.1111/j.1744-7348.2004.tb00353.x.
- Ashfaq, M., Hebert, P.D., Mirza, M.S., Khan, A.M., Mansoor, S., Shah, G.S. & Zafar, Y. (2014) DNA barcoding of *Bemisia tabaci* complex (Hemiptera: Aleyrodidae) reveals southerly expansion of the dominant whitefly species on cotton in Pakistan. *PLoS ONE* **9**(8), e104485.
- Bandelt, H.J., Forster, P. & Rohlf, A. (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology* **16**(1), 37–48. doi:10.1093/oxfordjournals.molbev.a026036
- Beerli, P. & Palczewski, M. (2010) Unified framework to evaluate panmixia and migration direction among multiple sampling locations. *Genetics* **185**(1), 313–326. <http://doi.org/10.1534/genetics.109.112532>
- Boykin, L.M. & De Barro, P.J. (2014) A practical guide to identifying members of the *Bemisia tabaci* species complex: and other morphologically identical species. *Frontiers in Ecology and Evolution* **2**(August), 1–5. doi:10.3389/fevo.2014.00045
- Boykin, L.M., Shatters, R.G., Rosell, R.C., McKenzie, C.L., Bagnall, R.A., De Barro, P. & Frohlich, D.R. (2007) Global relationships of *Bemisia tabaci* (Hemiptera: Aleyrodidae) revealed using Bayesian analysis of mitochondrial COI DNA sequences. *Molecular Phylogenetics and Evolution* **44**, 1306–1319. doi:10.1016/j.ympev.2007.04.020
- Boykin, L.M., Bell, C.D., Evans, G., Small, I. & De Barro, P.J. (2013) Is agriculture driving the diversification of the *Bemisia tabaci* species complex (Hemiptera: Sternorrhyncha: Aleyrodidae)? dating, diversification and biogeographic evidence revealed. *BMC Evolutionary Biology* **13**(1), 228. doi:10.1186/1471-2148-13-228
- Colvin, J., Omongo, C.A., Govindappa, M.R., Stevenson, P.C., Maruthi, M.N., Gibson, G., Seal, S.E. & Muniyappa, V. (2006) Host-plant viral infection effects on arthropod-vector population growth, development and behaviour: management and epidemiological implications. *Advances in Virus Research* **67**, 419–452. doi:10.1016/S0065-3527(06)67011-5
- Darriba, D., Taboada, G.L., Doallo, R. & Posada, D. (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9**(8), 772–772. doi:10.1038/nmeth.2109
- De Barro, P. & Ahmed, M.Z. (2011) Genetic networking of the *Bemisia tabaci* cryptic species complex reveals pattern of biological invasions. *PLoS ONE* **6**(10), e25579. doi:10.1371/journal.pone.0025579
- EPPO (2005) Tomato chlorosis crinivirus. *OEPP/EPPO Bulletin* **35**, 439–441. doi:10.1111/j.1365-2338.2005.00888.x
- Fiallo-Olivé, E., Hamed, A.A., Moriones, E. & Navas-Castillo, J. (2011) First report of tomato chlorosis virus infecting tomato in Sudan. *Plant Disease* **95**(12), 1592. doi:10.1094/PDIS-08-11-0631



- Fortes, I.M., Moriones, E. & Navas-Castillo, J. (2012) Tomato chlorosis virus in pepper: prevalence in commercial crops in southeastern Spain and symptomatology under experimental conditions. *Plant Pathology* **61**(5), 994–1001. doi:10.1111/j.1365-3059.2011.02584.x
- Gao, R.R., Zhang, W.-P., Wu, H.-T., Zhang, R.-M., Zhou, H.-X., Pan, H.-P., Zhang, Y.-J., Brown, J.K. & Chu, D. (2014) Population structure of the greenhouse whitefly, *Trialeurodes vaporariorum* (Westwood), an invasive species from the Americas, 60 years after invading China. *International Journal of Molecular Sciences* **15**(8), 13388–13400. doi:10.3390/ijms150813388
- Grapputo, A., Boman, S., Lindström, L., Lyytinen, A. & Mappes, J. (2005) The voyage of an invasive species across continents: genetic diversity of North American and European Colorado potato beetle populations. *Molecular Ecology* **14**(14), 4207–4219. doi:10.1111/j.1365-294X.2005.02740.x
- Huelsenbeck, J.P., Andolfatto, P. & Huelsenbeck, E.T. (2011) Structurama: bayesian inference of population structure. *Evolutionary Bioinformatics* **2011**(7), 55–59. doi:10.4137/EBO.S6761
- Kapantaidaki, D.E., Ovcarenko, I., Fytrou, N., Knott, K.E., Bourtzis, K. & Tsagkarakou, A. (2015) Low levels of mitochondrial DNA and symbiont diversity in the Worldwide agricultural pest, the greenhouse whitefly *Trialeurodes vaporariorum* (Hemiptera: Aleyrodidae). *Journal of Heredity* **106** (1), 80–92. doi:10.1093/jhered/esu061
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Simon Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. & Drummond, A. (2012) Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**(12), 1647–1649. doi:10.1093/bioinformatics/bts199
- Kinera, T., Wainaina, J., Maina, S. & Boykin, L.M. (2016) Rooting trees, methods for. *Encyclopedia of Evolutionary Biology* **3**, 489–493. doi:10.1016/B978-0-12-800049-6.00215-8
- Lapidot, M., Legg, J.P., Wintermantel, W.M. & Polston, J.E. (2014) Management of whitefly-transmitted viruses in open-field production systems. *Control of Plant Virus Diseases. Advances in Virus Research* (1st ed., Vol. 90). Elsevier Inc., doi:10.1016/B978-0-12-801246-8.00003-2
- Legg, J.P., Sseruwagi, P., Boniface, S., Okao-Okuja, G., Shirima, R., Bigirimana, S., Gashakae, G., Herrmannf, H.-W., Jeremiahg, S., Obieroh, H., Ndyetabulai, I., Tata-Hangy, W., Masembe, C. & Brown, J.K. (2014) Spatio-temporal patterns of genetic change amongst populations of cassava *Bemisia tabaci* whiteflies driving virus pandemics in East and Central Africa. *Virus Research* **186**, 61–75. doi:10.1016/j.virusres.2013.11.018
- Miano, D.W., LaBonte, D.R. & Clark, C.A. (2008) Identification of molecular markers associated with sweet potato resistance to sweet potato virus disease in Kenya. *Euphytica* **160**(1), 15–24. doi:10.1007/s10681-007-9495-2
- Mound, L.A. & Halsey, S.H. (1978) *Whitefly of the World. A Systematic Catalogue of the Aleyrodidae (Homoptera) with Host Plant and Natural Enemy Data*. British Museum (Natural History) & John Wiley & Sons, Chichester, Great Britain.
- Navas-Castillo, J., Fiallo-Olivé, E. & Sánchez-Campos, S. (2011) Emerging virus diseases transmitted by whiteflies. *Annual Review of Phytopathology* **49**, 219–248. doi:10.1146/annurev-phyto-072910-095235
- Navas-Castillo, J., Lopez, J.J. & Aranda, M.A. (2014) Whitefly-transmitted RNA viruses that affect intensive vegetable production. *Annals of Applied Biology* **165**, 155–171. doi:10.1111/aab.12147
- Njaramb, N.T. (2000) Assessment of the level of resistance to bifenthrin cypermethrin and methomyl insecticides in *Bemisia tabaci* (Gennadius) and *Trialeurodes vaporariorum* (Westwood) whitefly populations from selected sites in Kenya. Msc. dissertation, Kenyatta University.
- Ovcarenko, I., Kapantaidaki, D., Lindström, L., Gauthier, N., Tsagkarakou, A., Knott, K. & Vänninen, I. (2014) Agroecosystems shape population genetic structure of the greenhouse whitefly in Northern and Southern Europe. *BMC Evolutionary Biology* **14**, 165. doi:10.1186/s12862-014-0165-4
- Prijović, M., Marčić, D., Drobnjaković, T., Međo, I. & Perić, P. (2013) Life history traits and population growth of greenhouse whitefly on different tomato genotypes. *Pesticid Phytomed* **28**(4), 239–245. doi:10.2298/PIF1304239P
- Prijović, M., Skaljic, M., Drobnjaković, T., Zanić, K., Perić, P., Marčić, D., & Puizina, J. (2014) Genetic variation of the greenhouse whitefly, *Trialeurodes vaporariorum* (Hemiptera: Aleyrodidae), among populations from Serbia and neighbouring countries, as inferred from COI sequence variability. *Bulletin of Entomological Research* **1**, 1–10. doi:10.1017/S0007485314000169
- Rambaut, A., Suchard, M.A., Xie, D. & Drummond, A.J. (2014) Tracer v1.6. Available online at <http://beast.bio.ed.ac.uk/Tracer>
- Roopa, H.K., Kumar, N.K.K., Asokan, R., Rebijith, K.B., Mahmood, R. & Verghese, A. (2012) Phylogenetic analysis of *Trialeurodes Spp.* (Hemiptera: Aleyrodidae) from India based on differences in mitochondrial and nuclear DNA. *Florida Entomologist* **95**, 1086–1094. doi:10.1653/024.095.0438
- Sseruwagi, P., Maruthi, M.N., Colvin, J., Rey, M.E.C., Brown, J.K. & Legg, J.P. (2006) Colonization of non-cassava plant species by cassava whiteflies (*Bemisia tabaci*) in Uganda. *Entomologia Experimentalis et Applicata* **119**(Cmd), 145–153. doi:10.1111/j.1570-7458.2006.00402.x
- Swofford, D.L. (2003) *PAUP\*. Phylogenetic Analysis Using Parsimony (\*and Other Methods)*. Version 4. Sinauer Associates, Sunderland, Massachusetts.
- Tamura, K., Peterson, D., Stecher, G., Nei, M. & Kumar, S. (2011) MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* **28**(10), 2731–2739. <http://doi.org/10.1093/molbev/msr121>
- Ulrich, A. (2014) Export-oriented horticultural production in Laikipia, Kenya: assessing the implications for rural livelihoods. *Sustainability* **6**, 336–347. doi:10.3390/su6010336
- Wilkinson, S., Haley, C., Alderson, L. & Wiener, P. (2011) An empirical assessment of individual-based population genetic statistical techniques: application to British pig breeds. *Hereditas* **106**(2), 261–269. <http://doi.org/10.1038/hdy.2010.80>
- Wintermantel, W.M., Hladky, L.L., Gulati-Sakhuja, A., Li, R., Liu, H.Y. & Tzanetakis, I.E. (2009) The complete nucleotide sequence and genome organization of tomato infectious chlorosis virus: a distinct *crinivirus* most closely related to lettuce infectious yellows virus. *Archives of Virology* **154**(8), 1335–1341. doi:10.1007/s00705-009-0432-7
- Wisler, G.C., Li, R.H., Liu, H.Y., Lowry, D.S. & Duffus, J.E. (1998) Tomato chlorosis virus: a new whitefly-transmitted, Phloem-limited, bipartite closterovirus of tomato. *Phytopathology* **88**, 402–409.