Evolutionary Bioinformatics

ORIGINAL RESEARCH

Species Delimitation and Global Biosecurity

Laura M. Boykin[1], Karen F. Armstrong[1], Laura Kubatko[2] and Paul De Barro[3]

[1]Bio-Protection Research Centre, PO Box 84, Lincoln University, Lincoln 7647, New Zealand. [2]Departments of Statistics and Evolution, Ecology, and Organismal Biology, The Ohio State University, Columbus, Ohio 43210, USA. [3]CSIRO Ecosystem Sciences, GPO Box 2583, Brisbane QLD 4001. Corresponding author email: lboykin@mac.com

Abstract: Species delimitation directly impacts on global biosecurity. It is a critical element in the decisions made by national governments in regard to the flow of trade and to the biosecurity measures imposed to protect countries from the threat of invasive species. Here we outline a novel approach to species delimitation, "tip to root", for two highly invasive insect pests, *Bemisia tabaci* (sweetpotato whitefly) and *Lymantria dispar* (Asian gypsy moth). Both species are of concern to biosecurity, but illustrate the extremes of phylogenetic resolution that present the most complex delimitation issues for biosecurity; *B. tabaci* having extremely high intra-specific genetic variability and *L. dispar* composed of relatively indistinct subspecies. This study tests a series of analytical options to determine their applicability as tools to provide more rigorous species delimitation measures and consequently more defensible species assignments and identification of unknowns for biosecurity. Data from established DNA barcode datasets (COI), which are becoming increasingly considered for adoption in biosecurity, were used here as an example. The analytical approaches included the commonly used Kimura two-parameter (K2P) inter-species distance plus four more stringent measures of taxon distinctiveness, (1) Rosenberg's reciprocal monophyly, (P(AB)),[1] (2) Rodrigo's (P(randomly distinct)),[2] (3) genealogical sorting index, (*gsi*),[3] and (4) General mixed Yule- coalescent (GMYC).[4,5] For both insect datasets, a comparative analysis of the methods revealed that the K2P distance method does not capture the same level of species distinctiveness revealed by the other three measures; in *B. tabaci* there are more distinct groups than previously identified using the K2P distances and for *L. dipsar* far less variation is apparent within the predefined subspecies. A consensus for the results from P(AB), P(randomly distinct) and *gsi* offers greater statistical confidence as to where genetic limits might be drawn. In the species cases here, the results clearly indicate that there is a need for more gene sampling to substantiate either the new cohort of species indicated for *B. tabaci* or to detect the established subspecies taxonomy of *L. dispar*. Given the ease of use through the Geneious species delimitation plugins, similar analysis of such multi-gene datasets would be easily accommodated. Overall, the tip to root approach described here is recommended where careful consideration of species delimitation is required to support crucial biosecurity decisions based on accurate species identification.

Keywords: reciprocal monophyly, genealogical sorting index (*gsi*), randomly distinct, taxonomic distinctiveness, GMYC, *Bemisia tabaci*, *Lymantria dispar*, species identification, invasive species

Video Abstract Available from http://la-press.com/t.php?i=va001

Evolutionary Bioinformatics 2012:8 1–37

doi: 10.4137/EBO.S8532

This article is available from http://www.la-press.com.

Introduction

Species delimitation and assigning individuals to species should not be confused with the species concept debate.[6–15] Species delimitation is the methodological problem of inferring boundaries and numbers of species; it is an essential prerequisite for accurate species identification with implications at multiple levels, from founding taxonomy to systematic biology, organismal research and measuring biodiversity. This is distinct from the concept of a species, which is a theoretical matter of defining categories based on any one of many contemporary and often conflicting views.[9] The latter has its own critical influence on subsequent decisions and actions, but for most practical purposes the taxonomic unit of a species is frequently delimited through the use of discriminating morphological characters. Where these are lacking, as in the case of cryptic species complexes, species boundaries may be fuzzy, leading to the practical problem of then assigning identity.

In such situations, where the morphological species boundaries are fuzzy, molecular genetic information is often relied upon as an additional means by which to delimit and identify species. Methods based on the use of DNA sequencing and large reference datasets for both research and operational application have become more popular over the last 20 years. Of these, data for genes such as mitochondrial cytochrome oxidase one (mtCOI) DNA barcodes for vertebrates and invertebrates,[16–18] mitochondrial cytochrome b for fish,[19,20] 18S and 28S rDNA for nematodes,[21,22] ITS rDNA for fungi[23,24] and 16S for bacteria species[25,26] are all potential assets for critical applications such as biosecurity diagnoses.

Analytical methods for species delimitation are frequently based on exclusivity and have typically relied upon genetic distance, gene tree monophyly or statistical parsimony networks. These measures also require subjective decisions regarding thresholds for species boundaries[23,27–30] and are vulnerable to producing false negative and false positive assignments.[31] The use of lineage divergence based on the Kimura 2 parameter (K2P) model of molecular evolution[32] to produce generic rules such as the 3% and 10× rule used in DNA barcoding[17,33] is a case in point.[34–36] Improvements on this include methods for utilizing multiple genes to infer species trees that can then be used to aid in species delimitation[3,37–46] and avoid the potential pitfalls of single gene phylogenies.[47,48] Such progress in phylogenetic theory has lead to the development of ideas and software to generate a species tree from several gene trees with statistical support, hence obviating the need to rely on use of a single gene tree and subjective assignment decisions.

The benefit of using multilocus phylogenies to confidently delimit species is clear (eg[49]). However, in contrast to research-driven queries, there is not the time to develop the ideal dataset in response to often unpredictable biosecurity events where rapid decisions are necessary eg, a day to decide whether to reject a shipment at a port of entry. As such, for the foreseeable future, the appeal of using data that largely exists across a useful taxonomic range for only single genes will remain. Unfortunately species identity based on sequence similarity usually relies on the convenience of rudimentary analyses and subjective interpretation of species limits. Further, in cases without a well-resolved phylogeny, it is difficult to identify unknowns; therefore species delimitation of a well resolved phylogeny must precede attempts at identification. This study therefore tests a series of analytical options to determine their applicability as tools to provide more rigorous species delimitation measures and consequently more defensible species assignments for biosecurity.

Biosecurity as defined by the Food and Agriculture Organization of the United Nations (FAO)[50] is *"A strategic and integrated approach that encompasses the policy and regulatory frameworks (including instruments and activities) for analysing and managing relevant risks to human, animal and plant life and health, and associated risks to the environment."*[50] In particular, this covers areas such as food safety, zoonoses, the introduction of animal and plant pathogens and plant, vertebrate and invertebrate pests, the introduction and release of living modified organisms and deliberate introduction and management of alien species.[51] A key role in the delivery of biosecurity is the regulation of trade and market access and underpinning these are the international standards, guidelines and recommendations that exist under the International Plant Protection Convention (IPPC), the World Organization for Animal Health and the Codex Alimentarius Commission.[50,52] In terms of plant biosecurity, the central role of the IPPC is to coordinate work to prevent the spread and

introduction of pests of plants and plant products, and to promote appropriate measures for their control, with minimal disruption to trade (https://www.ippc.int/). A key element here is the international standards for phytosanitary measures which are administered through member countries National Plant Protection Organisations (NPPO). The NPPO plays the lead role in ensuring regulatory compliance so as to reduce the likelihood that pests of plants and plant products are spread via trade. Central to this is the capacity to identify species of concern either to the NPPO's country or to its trading partners accurately and in a timely manner. It is therefore crucial that NPPOs have the capacity to accurately assign organisms of concern to the correct species. In other words, NPPOs must incorporate an ability to delimit species accurately.

These present NPPOs with a particular challenge as they are tasked with making species assignments as part of their role as regulators and imposers of standards. The consequences of inaccurate species identification as a consequence of poor species delimitation are highlighted by the following examples. In 2004 (http://www.worldtradereview.com/news.asp?pType=N&iType=A&iID=79&siD=26&nID=14049) Pakistan rejected a shipment of wheat from Australia worth AUD$18 million due to contamination with the fungus fungal pathogen karnal bunt, *Tilletia horrida* Takah. It was subsequently found that, rather than *T. horrida*, the shipment contained the recently discovered close relative, *T. walkeri*, which is of no biosecurity significance and therefore did not warrant the economically significant quarantine intervention that had ensued. Central to this was the taxonomic confusion over the identity of the species used as the positive control in the diagnostic analysis; these initial species were assigned as positive controls. This, which in turn undermined the reliability of the diagnostic test, led to a false positive result.[53] In response, Rossman (2008) observed "This situation illustrates the dire need for accurate phylogenetic information upon which to base molecular diagnostic tests. Such tests are not accurate without the essential underpinning of systematic knowledge"; a key element here is the capacity to delineate reliably between species.

Another example is the case of myrtle rust, *Uredo rangelii* and guava rust, *Uredo psidii* or *Puccinia psidii* (the former is the name assigned to the asexual stage and the latter to the sexual stage). These pathogens are of biosecurity concern to regulators in Australia and overseas as they have the potential to infect and cause serious disease in many species of Myrtaceae, a family of many significant Australian native plant species. The identification of the specific pathogen responsible is critical as it has different and important implications in terms of international quarantine and market access as the quarantine and market access measures imposed with the aim of restricting spread from Australia will be based on the identity of the pathogen and will vary considerably depending on the species. To date, debate continues as to whether they are the same or different species and indeed how many species there are in the guava rust complex.[54–57] Furthermore, it has been suggested that *Puccinia psidii* sensu lato is neither a species of *Puccinia* nor a member of the Pucciniaceae.[58] The solution to the identity of the pathogen is considered to lie in molecular phylogenetic analysis and the challenge will be to generate a new molecular phylogenetic analysis to determine species boundaries has been called for.[57]

Similarly, in the true fruit flies (Tephritidae), which contain a number of economically significant pests, there are several cases where species boundaries are uncertain eg, the species complexes of *Anastrepha fraterculus*,[59] *Ceratitis* FAR (*fasciventris, anonae, rosa*)[60] and *Bactrocera dorsalis*.[61] These present considerable challenges for NPPO's as species identity is a key element, from demonstrating area freedom as part of market access arrangements, to mobilizing implementing effective eradication or management strategies that rely on species-specific methods such as the sterile insect technique (mating of releasing sterile male flies with to mate with wild female flies). A case in point is the incursion of *B. papayae*, which eventually cost AUD$35 million to eradicate from Australia between 1995–1999. Here, difficulty the failure of amongst taxonomists to agree on the specifics identification within the complex resulted in resistance by growers to implement the crucial implementation of initial quarantine restrictions during the initial phase of the eradication campaign.[62]

To explore the issue of species delimitation based on DNA data, the recently developed phylogenies for two invasive species of both systematic biology and biosecurity interest, *Bemisia tabaci* (sweetpotato whitefly)[63] and *Lymantria dispar* (Asian gypsy moth),[64] are considered here. Both cause millions of dollars

of damage globally[65,66] and are regarded as regulated species by a number of countries or regions, eg, Australia, the EU and New Zealand. *Bemisia tabaci* is globally distributed and capable of causing extensive damage to major vegetable, grain, legume and fiber crops.[67] It is currently described as a single species, but this has been subject to ongoing debate with the most recent publications arguing that it is a species complex (see[63] for review). In Europe, for example (EPPO data sheets on quarantine pests, http://www.eppo.org/QUARANTINE/insects/Bemisia_tabaci/BEMITA_ds.pdf), phytosanitary certificates for plants and parts of plants for propagation, some cut flowers and fresh foliage are required to declare freedom from non-European populations of *B. tabaci*[68]—yet what defines non-European is unresolved. *Bemisia tabaci* is genetically complex[63] with at least 28 distinct genetic groups identified based on mtCOI[63,69–71] and these are regarded in Dinsdale et al[70] and De Barro et al[63] as putative species. The species level delimitation proposed by Dinsdale et al[70] and De Barro et al[63] is supported by all available mating compatibility studies which show either no copulation between putative species or significant declines in fitness in the resultant F1 and F2 generations, but at present no morphological characters have been found to distinguish the different putative species.[72–77] The two putative species of considerable biosecurity concern are Middle East Asia Minor 1 (includes what is commonly referred to as biotype B, hereon MEAM1) and Mediterranean (includes what is currently referred to as biotype Q, hereon MED).

The second species, *Lymantria dispar* s.l., is one of the most destructive pests of forest, shade, fruit and ornamental trees throughout the northern hemisphere.[78,79] In contrast to *B. tabaci*, it is already recognized as being composed of subspecies; originally *L. dispar dispar* and *L. dispar japonica*,[80] but more recently including a third *L. dispar asiatica*.[81] Each has different implications for biosecurity. Primarily Asian females of *L. dispar* have larger wings and are capable of sustained flight (therefore capable of greater geographic spread)[82–86] whereas those of European and North American "populations" are not.[87,88] Unfortunately, female flight ability is not the only a trait associated with the Asian strain,[85] and delimitation is further confounded by overlapping geographic ranges, variation in behaviour (attraction to light) and host preferences within a broad host range; all of which are important indicators of their relative invasive capability and biosecurity importance. Molecular data have so far failed to clarify taxonomically assigned subspecies boundaries. MtCOI restriction haplotypes have suggested broadly three groups, North America, Europea/Siberia, and Asia,[85] while mtCOI barcode data and unsupported clades in a neighbor-joining (NJ) phylogeny indicate groupings of *L. dispar dispar* from North America and France (2), *L. dispar dispar* from Europe and Western Asia, and *L. dispar asiatica*/*japonica*.[64]

This study utilizes a "tip to root" approach for assessing taxonomic distinctiveness as a novel means of removing the subjectivity of species delimitation when considering phylogenetic relationships and levels of divergence. Several statistical measures were used to characterize the phylogenies of these two invasive insects as a basis for defining their species limits, ultimately to improve the level of confidence with which unknown individuals can be placed for identification purposes. The measures used are (1) P(AB), a test for taxonomic distinctiveness as determined by the null hypothesis that monophyly is a chance outcome under a model of random coalescence in a single group,[1] (2) The genealogical sorting index (*gsi*), which quantifies the degree of exclusive ancestry of a particular group on a rooted phylogeny[3] (3) P(RD) which is the probability that a clade has the observed degree of distinctiveness due to a random coalescent process[2] and (4) GMYC.[4,5] These measures were used to identify, where possible, potential points that represent taxonomic distinctiveness within existing phylogenies for *B. tabaci* and *L. dispar* and then compare these findings with the K2P inter-species distances and posterior probabilities that are currently used to support "species" cut-off points.

Materials and Methods

Samples

Bemisia tabaci dataset—The *B. tabaci* dataset consisted of a global sampling of 370 individuals and 657 base pairs of the mtCOI 3′ region (as described in[63,70]). The 24 low-level groups described in Dinsdale et al.[70] were identified using a Bayesian analyses and K2P distances and these groups were used as basis for the subsequent analyses of taxonomic distinctiveness. The Bayesian phylogeny was also used to test for

taxonomic distinctiveness with no a priori bias of the latter (see description below).

Lymantria dispar dataset—Data from de Waard et al,[64] consisting of 658 base pairs of the DNA barcode mtCOI 5′ for 319 individuals, was re-analysed to produce a Bayesian phylogeny with node support (PP) (lacking in the DeWaard[64] phylogeny) using MrBayes 3.1.2[89] on the BlueFern® supercomputer at the University of Canterbury, Christchurch, New Zealand. Two independent runs of 8 million generations utilizing eight processors were used, with every 100th tree retained, resulting in a sample of 80,000 trees for each run. The sumt command was used with 25% of the trees discarded as burn-in to produce a consensus tree. Convergence of the Bayesian runs was assessed by the potential scale reduction factor.[90] In addition, the average standard deviation of split frequencies was consistently close to 0.05 for the last 1 million generations of the runs. There was no indication of a lack of convergence of the MCMC. Clades identified within this phylogeny with high posterior probability were used to test for species distinctiveness. K2P distances were calculated in Geneious.

Species distinctiveness measures

The species delimitation plugin[91] for Geneious[92] was used to calculate Rosenberg's reciprocal monophyly, P(AB)[1] and Rodrigo's (P(RD)[2] measures. The genealogical sorting index (*gsi*)[3] statistic was calculated in R based on the estimated tree and the assignment file that contains user specified groups (see http://www.genealogicalsorting.org/). Two different assignment files were generated for the *gsi* for each dataset: one based on previously-defined taxonomic groups, and the other containing groups within those as determined a priori here. Each of the assignment files was run with the known phylogeny and an R script that specifies the number of permutations (200,000 permutations across 4 processors). All of the *gsi* analyses were run using R on the BlueFern® cluster at The University of Canterbury. To assess the significance of the *gsi P*-values the Bonferroni correction was used as follows. For the *B. tabaci* previously defined groups, excluding those with only one representative, resulted in 17 tests therefore the *P*-value cut-off used is 0.05/17 = 0.002; for the non-predefined groups there were 227 tests (some of the clades were nested and had to be specified differently in various *gsi* assignment files and required several runs to get all the configurations and associated *gsi*/*P*-values) therefore the *P*-value cut-off is 0.05/227 = 0.00002. The *L. dispar gsi* run involved 14 tests and the cut-off is 0.05/14 = 0.004.

Species boundaries were also assessed using the GMYC approach,[4,5] which requires a fully resolved phylogeny with branch length estimates. A Bayesian analyses using BEAST 1.6.1[93] was run on BeSTGRID and the Bluefern Supercomputer at The University of Canterbury, New Zealand. The analysis performed used a relaxed lognormal clock and branch lengths were estimated using a coalescent prior and a GTR + I + γ model of evolution. The GMYC employs a coalescent as the null model to explain branching patterns and so the coalescent prior gives more conservative results as it is more likely to ignore a coalescent-speciation transition.[4,5] Two independent runs were completed for each insect data set. The BEAST runs consisted of 50 million generations with trees sampled every 5,000th generation. Convergence of the runs was checked using Tracer v1.5.4[94] and the ESS values were well above 200 for each run. Log Combiner v1.5.4[95] was used to combine the trees from each of the runs and the burnin (1001 trees) were removed. TreeAnnotator v1.5.4[96] summarized the trees (maximum clade credibility and median height specified) and produced one singe maximum clade credibility tree that was then used for the input into the SPLITS package for the R statistical environment (https://r-forge.r-project.org/projects/splits).

In addition to species delimitation measures, the Geneious plugin[91] also generates values for species identification based on the groups being tested. For the groups of interest P ID(Strict) and P ID Liberal) were also calculated (Tables 2 and 3). This was done following the methods described in Ross et al.[31] In brief, the plugin facilitates the calculation of the probability (and the 95% confidence interval) of a hypothetical unknown taxon being positively identified in the group of interest.

Analyses using previously defined groups

Previously published phylogenies[63,64,70] defining species or taxonomic groups were used to test predefined groups with current measures of species distinctness

(P(AB), P(RD), *gsi* and GMYC). Groups for *B. tabaci* were based on percentage divergence (K2P distances).[70] For *L. dispar*, subspecies were defined based on geographic distribution limits and also on Bayesian assignment tests.[97] These groups were then tested using the species distinctiveness measures (as described above).

Analyses without a priori groups defined

Fixed phylogenies were imported into Geneious and each group (two or more individuals) or clade was tested against its sister group to assess whether it was distinct according to the P(AB) and the P(RD). The *B. tabaci* analyses consisted of 231 pairwise comparisons across the fixed phylogeny. Each major clade was assigned a number 1–9 (Fig. 2), essentially partitioning a naked phylogeny, with no preconceived bias, by systematically working from the tips to the root of the tree analyzing taxonomic distinctiveness. Within each of these clades, additional groups were also assigned a number, for example clade 2, group 1 is given the number 2–1. This process starts at the tips of the tree and works along the branches asking the question: Where on this phylogeny is there enough "distinctiveness" according to the measure to call the groups in question a "species"? For *L. dispar*, the subspecies previously described[64,81] were not monophyletic in the consensus phylogeny estimated using MrBayes, estimated here, therefore the P(AB) was not calculated. The iterative tip to root process of assessing taxonomic distinctiveness described above was carried out, but the lack of resolution meant that only 14 pair-wise clade comparisons could be included in the analysis.

The species distinctiveness measures K2P distance/P(RD)/posterior probability/P(AB)/*gsi* were evaluated and their significance (+ or −) assigned. Significance (+) was determined as >1% K2P difference/>0.05/>0.70/Bonferonni correction values (*gsi P*-values, see above) respectively; non-significance was coded as "−". For example, with a clade assigned +/+/+/+/+ indicates significant species distinctiveness for all five measures. The criteria to identify portions of the tree that were taxonomically distinct was to have posterior probabilities above 0.70. The other four measures were then evaluated for these groups to determine taxonomic distinctiveness.

Results

Testing *B. tabaci* predefined groups

Seventeen of the 24 previously defined genetic groups[63,70] (Fig. 1) contained multiple haplotypes and were tested using the various species distinctiveness measures (Table 1). All of these groups were monophyletic and the inter-species distance between them ranged from 1.114 to 3.34% divergence (Table 1). Using the "strict" criterion described by Ross et al,[31] where the reference data set contains similar sequences in a monospecific clade, the probabilities of correctly identifying a hypothetical unknown (query sequences) ranged from 0.54 (0.39, 0.69 with a 95% confidence interval) for Asia II 6 to 0.96 (0.91, 1.0) for the Australia/Indonesia clade. Utilizing a more liberal criterion, where the query sequence falls within a monospecific clade or a sister clade, the probabilities were considerably higher for all of the clades (excluding Asia II 5) having probabilities above 0.90 (Table 1).

Rodrigo et al (2008) defines distinctive clades as those that have P(RD) values <0.05. All of the previously defined genetic groups have P(RD) values <0.05 (Table 1). Clade support for the 24 genetic groups ranged from 0.69 for SubSaharan Africa 1 to 1.00 in 10 of the other genetic groups. Thirteen of the 24 groups tested had significant P(AB) values ($P < 10^{-5}$) (Table 1). In this analysis, 13 groups have *gsi* values of 1.00, with 12 of these having significant *gsi P*-values of 5.00E-06, while Asia II 5 had a non-significant *P*-value after Bonferroni correction (2.10E-03). The two groups with the lowest *gsi* values were sub-Saharan Africa 1 (0.56, *P*-value = 5.00E-06) and Asia I (0.49, *P*-value = 5.00E-06). Australia/Indonesia and sub-Saharan Africa 2 deviated from 1.0 (0.75, *P*-value = 5.00E-06 and 0.85, *P*-value = 5.00E-06, respectively).

Testing all *B. tabaci* groups (no a priori groups defined)

Every group/clade on the estimated phylogeny that contained two or more individuals was tested to assess P(AB) and P(RD) and other various measures plus the *gsi* (Table 2), using R (http://www.r-project.org/) on the University of Canterbury computer cluster. Figure 2 and the supplemental data show the clustering strategy, with reference to the global

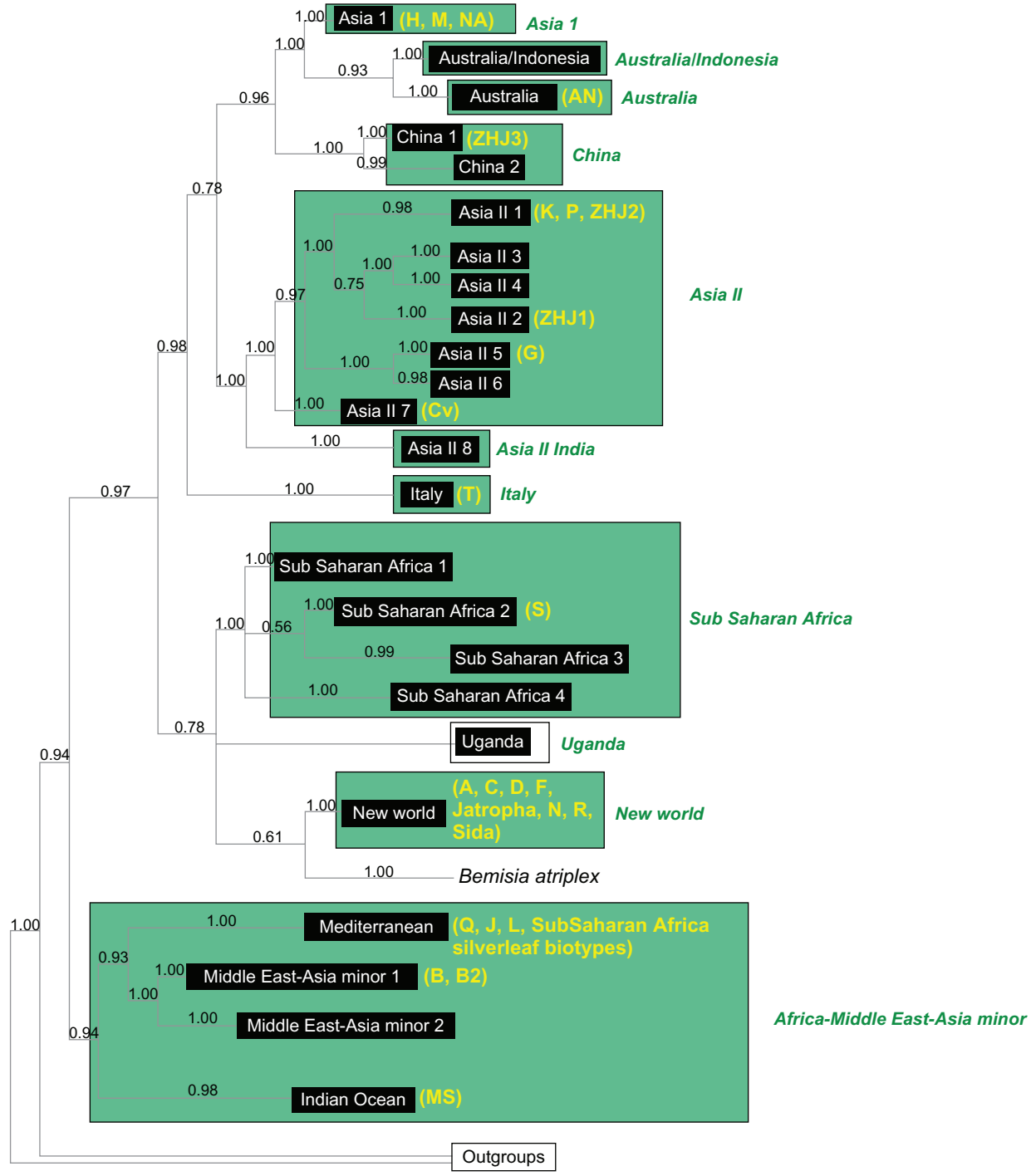

Figure 1. Duplicate phylogeny as seen in De Barro et al[63] to show the previously defined groups being tested for species distinctiveness (results shown in Table 1).

phylogeny (Figs. 1 and 2). Table 2 shows the pairwise combinations and corresponding species distinctiveness measures. All groups defined in the supplemental data and Table 2 were tested using *gsi* (200,000 permutations) and resulted in a greater number of distinctive groups (Table 2) than were previously described as predefined groups in Figure 1 and Table 1. This included all of the 24 genetic groups described in Dinsdale et al,[70] which were supported by posterior probabilities of at least 0.70 (Table 1) and which were defined based on the K2P distances. In addition, results of the clustering strategy outlined in Figure 3 and the supplemental data reveal that there were more clades that were taxonomically distinct based on the criteria outlined in the

Table 1. The clades described by Dinsdale et al.[70] (Column 1 in parenthesis are the number of individuals in each clade) were tested for species distinctiveness as measured by the Geneious species delimitation plugin[91] and the genealogical sorting index (*gsi*).

	Intra dist	K2P	Intra/ inter	P ID(Strict)	P ID(Liberal)	Av (MRCA-tips)	P(RD)	PP	P(AB)	*gsi*	*P*-value
Dinsdale putative species											
SubSaharan Africa 1 (19)	0.344	1.867	0.18	0.93 (0.88, 0.98)	0.98 (0.95, 1.0)	0.2356	0.05	0.69	1.10E-16	0.559	5.00E-06
SubSaharan Africa 2 (44)	0.266	1.867	0.14	0.93 (0.86, 1.00)	0.98 (0.93, 1.0)	0.2541	0.05	1	1.00E-02	0.847	5.00E-06
SubSaharan Africa 3 (1)											
SubSaharan Africa 4 (8)	0.182	1.817	0.1	0.90 (0.79, 1.0)	0.97 (0.91, 1.0)	0.1126	0.95	0.98	1.80E-12	1	5.00E-06
New World (19)	0.158	3.34	0.05	0.92 (0.82, 1.0)	0.99 (0.93, 1.0)	0.126	0.05	1	1.20E-11	1	5.00E-06
India Ocean (12)	0.09	1.222	0.07	0.74 (0.57, 0.92)	0.97 (0.82, 1.0)	0.0598	0.05	1	2.00E-07	1	5.00E-06
Mediterranean (88)	0.616	1.538	0.4	0.87 (0.82, 0.92)	0.96 (0.94, 0.99)	0.5439	0.05	1	2.80E-05	1	5.00E-06
Middle East—Asia Minor 2 (1)											
Middle East—Asia Minor 1 (75)	0.41	1.538	0.27	0.91 (0.85, 0.96)	0.97 (0.94, 1.00)	0.693	0.05	0.78	2.70E-04	1	5.00E-06
Italy (7)	0.158	3.119	0.05	0.92 (0.82, 1.0)	0.99 (0.93, 1.0)	0.126	0.05	1	1.20E-11	1	5.00E-06
Asia II 1 (15)	0.389	2.11	0.18	0.93 (0.88, 0.98)	0.98 (0.95, 1.0)	0.8325	0.05	1	4.90E-06	1	5.00E-06
Asia II 2 (1)											
Asia II 3 (4)	0.379	2.11	0.18	0.81 (0.69, 0.94)	0.96 (0.86, 1.0)	0.4519	0.05	1	4.90E-06	1	5.00E-06
Asia II 4 (1)											
Asia II 5 (2)	0.367	1.114	0.33	0.57 (0.39, 0.75)	0.82 (0.68, 0.97)	0.2575	0.05	1	0.05	1	2.10E-03
Asia II 6 (3)	0.119	1.114	0.11	0.54 (0.39, 0.69)	0.92 (0.77, 1.0)	0.0596	0.05	1	0.05	1	5.00E-06
Asia II 7 (7)	0.37	2.519	0.15	0.88 (0.77, 0.98)	0.96 (0.89, 1.0)	0.3512	0.05	1	1.40E-08	1	5.00E-06
Asia II 8 (4)	1.836	2.826	0.65	0.79 (0.74, 0.85)	0.95 (0.92, 0.97)	1.4719	0.05	1	8.40E-07	1	5.00E-06
Ex China EU 192051 (1)											
China 1 (5)	0.853	3.101	0.28	0.81 (0.71, 0.92)	0.93 (0.87, 0.99)	1.0981	1	0.98	1.90E-08	1	5.00E-06
China 2 (1)											
Australia/Indonesia (4)	0.276	2.818	0.1	0.96 (0.91, 1.0)	0.99 (0.96, 1.0)	0.2005	0.05	0.93	4.70E-07	0.748	5.00E-06
Asia I (19)	1.003	2.818	0.36	0.69 (0.57, 0.82)	0.92 (0.81, 1.0)	1.2642	0.05	0.87	1.50E-06	0.489	5.00E-06
Australia (1)											

Notes: The species delimitation plugin generates: Intra Dist: average pairwise tree distance among members of a predefined clade, Inter Dist: average pairwise tree distance between members of the group of interest and its sister taxa (K2Pdistance), Intra/Inter: The ratio of Intra Dist to Inter Dist, P ID(Liberal): mean probability, with a 95% confidence interval (CI) for a prediction of making a correct identification of an unknown specimen being sister to or within the group of interest, P ID(Strict): mean probability, with a 95% confidence interval (CI) for a prediction of making a correct identification of an unknown specimen being found only in the group of interest,[31] Av(MRCA): mean distance between the most recent common ancestor of the species and its members, P(Randomly Distinct): probability that a clade has the observed degree of distinctiveness,[2] Clade Support: Bayesian posterior probability (PP), and Rosenberg's P_{AB}: Reciprocal monophyly and lastly, the *gsi* statistic and associated *P*-value are included.[3] Shaded numbers indicate significance, see figure legends for details and also visual representation as indicated by "+" for significance or "–" for non-significance.

methods section and with higher than 0.70 posterior probabilities. Specifically, clade 2 (sub-Saharan Africa) had 14 additional groups, clade 3 (New World) 8 additional groups, clade 5 (Asia I) 6 additional groups, clade 6 (Asia II) 12 additional groups, clade 8 (Middle East Asia Minor 1) 4 additional groups and clade 9 (Mediterranean) 7 additional groups. These groups were the basis of further investigation while clade 4 (Italy) and clade 7 (Indian Ocean) were omitted as they contained no other well supported groups (PP $<$ 0.70) (Table 2).

Of the original clades showing additional groups, clade 2, the SubSaharan African clade with 14 well supported clades defined by PP $>$ 0.70, had only seven clades with significant *gsi P*-values, four of which also had significant P(AB) values. Further analyses showed that clades 2-27 and 2-39 were distinct based on measures excluding the K2P distance (Fig. 3). For the New World clade (clade 3) with eight well supported clades (PP $>$ 0.70), none were significant by the P(AB) measure, but three had significant *gsi P*-values (3-1, 3-4, and 3-5). Clade 5 (Asia 1) had six well supported clades (PP $>$ 0.70) all of which had significant *gsi P*-values and two with significant P(AB) values. In addition, three, clades 5-15, 5-16, and 5-19, had significant K2P distance measures. There was significantly more taxonomic distinctiveness in clade 6 than previously described[70] with 12 additional clades (PP $>$ 0.70), seven of which had significant P(AB) and eight with a significant *gsi P*-value. Clade 8, Middle East Asia Minor 1, had four groups with significant posterior probabilities, but only 8-6 had significant P(AB) and *gsi P*-values. Clade 9, Mediterranean, had seven well supported (PP $>$ 0.70) clades, of these six had significant *gsi P*-values and five significant P(AB) values. Table 4 shows a summary of the clades that were taxonomically distinct. The GMYC analysis supported all the groups identified with the above measures (data not shown).

L. dispar phylogeny and species distinctiveness

The phylogeny for *L. dispar* (Fig. 4) was not as well resolved as that for *B. tabaci* (Figs. 1 and 2) and far less complex. In total, 14 clades, based on PP values, (boxed in Fig. 4) were apparent amongst the three taxonomically defined sub-species. The *L. dispar dispar* group (clade 12) was well supported (PP = 0.85), with eight groups supported as indicated by posterior probabilities above 0.70 (boxed in Fig. 4). In contrast, the *L. dispar asiatica*/*L. dispar japonica* group did not form a monophyletic clade, but contained three well-supported groups (13, 13a, and 14, Fig. 4). These 11 well-resolved groups were used in the assignment file for *gsi*. With the exception of clade 13, the *gsi* values were 1.0, with *P*-values ranging from 5.00E-05 to 1.25E-04, (Table 3) indicating taxonomic distinctiveness. The intra/inter distance ratios were low (Table 3) with even the divergence between clade 12 (*L. d. dispar*) and clade 13 (part of *L.d. asiatica*/*japonica*) being well below the typical species split threshold of 2% for Lepidoptera.[34] In a DNA barcoding context, the ability to accurately assign an unknown to one of these clades as measured by P ID(strict) or P ID(liberal) which ranges from 0.39 to 0.89 and 0.74 to 0.97, respectively, is questionable. Clades 3, 12, 13 and 14 had high values ($>$0.70) for both measures indicating a high probability of arriving at the correct assignment for an unknown as seen in Table 3. The measure for P(RD) ranged from 0.05–1.0, indicating a range from complete distinctiveness to no distinctiveness and did not appear to correlate with any of the other measures. All of the *P*-values for the *gsi* statistic were significant for all 14 clades (Table 3). The GMYC results did not identify any additional groups representing of taxonomic distinctiveness (data not shown).

Discussion

Species delimitation, the process by which species boundaries are determined and new species are described, is not just a matter for theoretical consideration. Operational issues also fuel the debate as eloquently summarized by Sites and Marshall.[14] For areas such as the delivery of national biosecurity obligations, immediate practical implications exist when delimitation is unclear. This is becoming increasingly apparent with the adoption of DNA sequence analysis and the generation of large single-gene datasets for both taxonomic research[98] and diagnosis of high risk pest species.[99] The two species analysed here represent extremes of the taxonomic problems faced; one with a well-resolved phylogeny supporting large intra-specific variation (*B. tabaci*) and another with a poorly resolved phylogeny and limited genetic

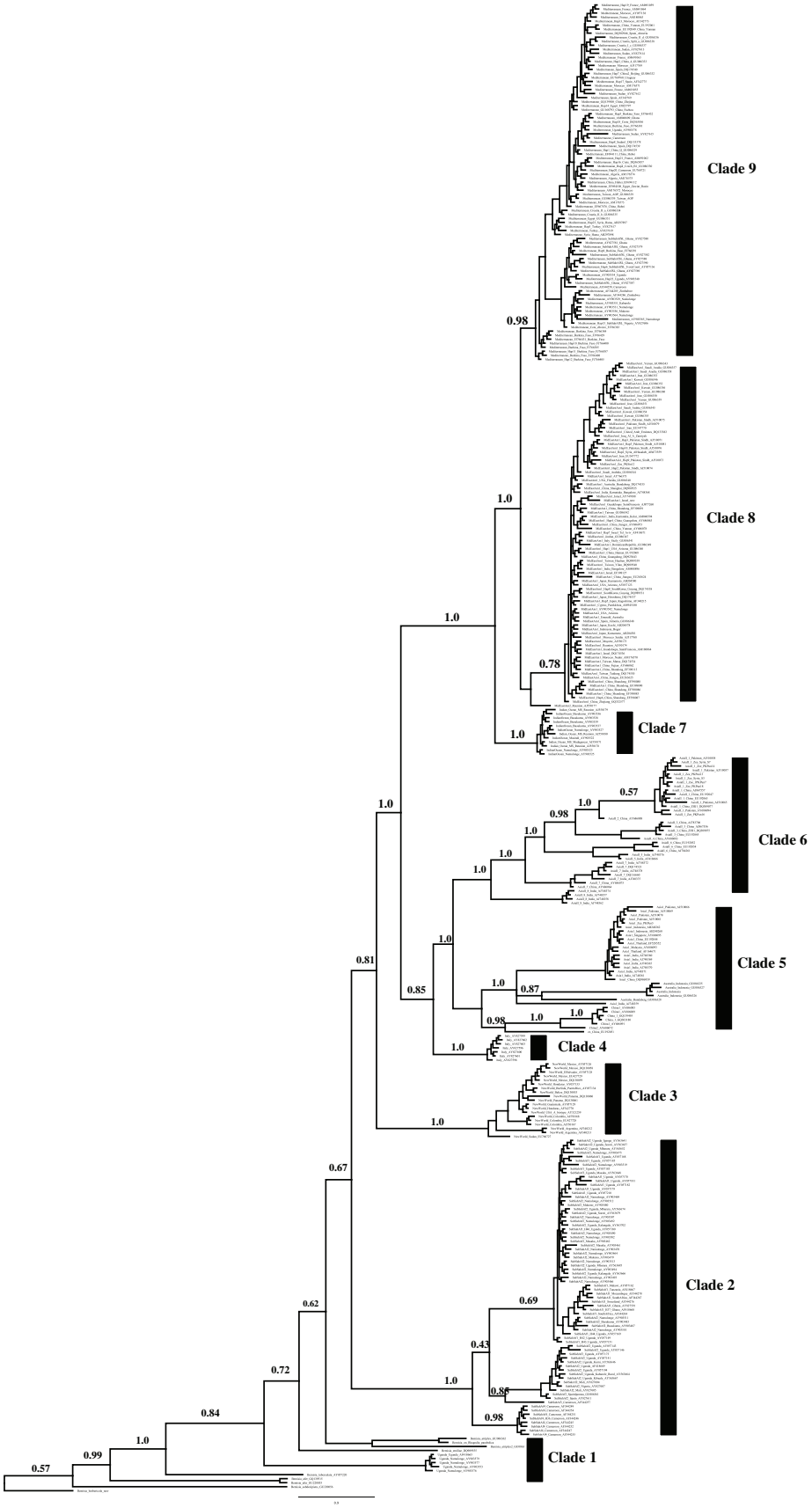

Figure 2. Global phylogeny for *B. tabaci* generated using MrBayes and sequences from De Barro et al.[63]
Note: Major clades are numbered and these numbers correspond to the first number in Tables 1 and 2.

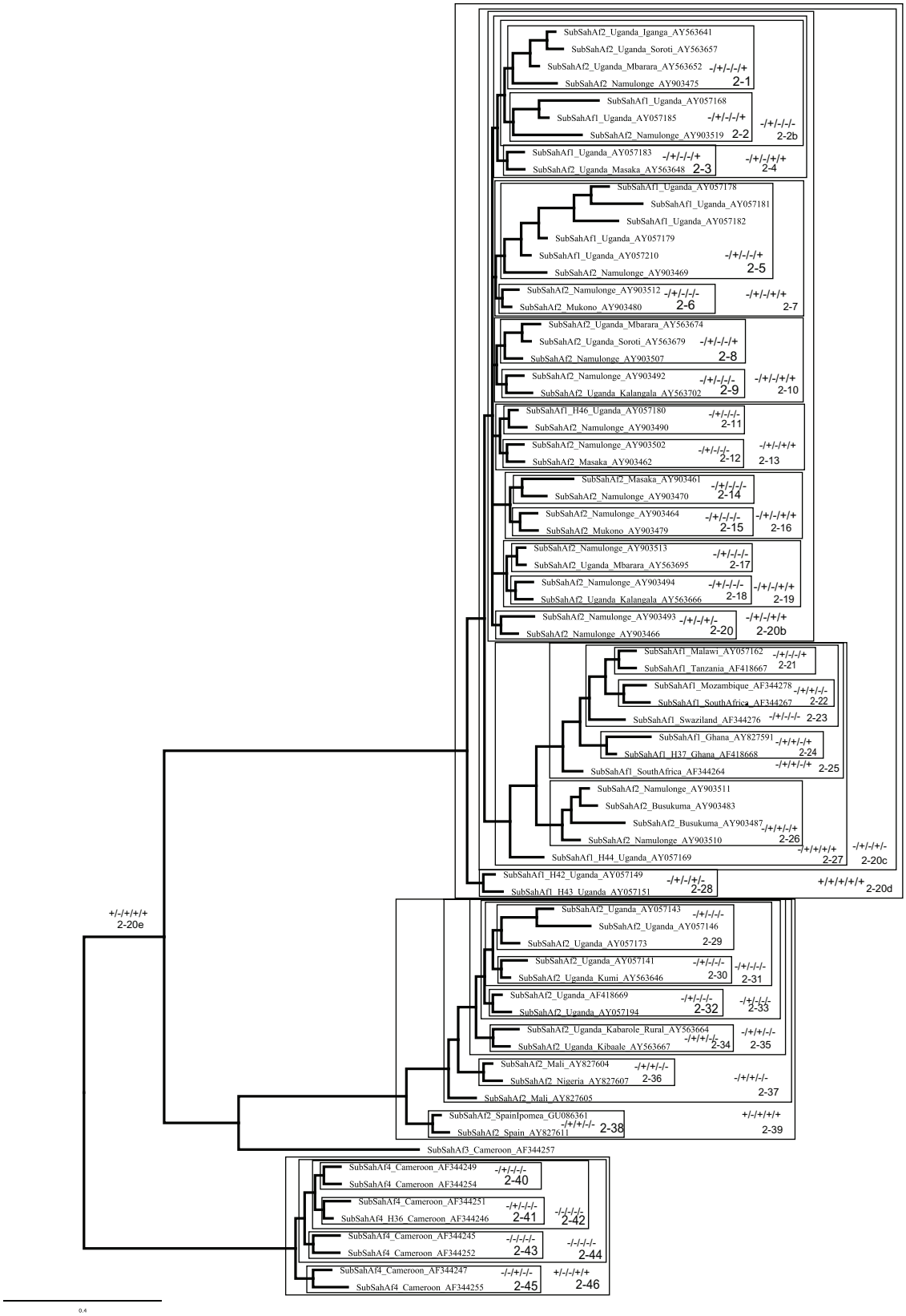

Figure 3. Clade 2 extracted from Figure 2 with boxes around groups being tested using the species delimitation software implemented in the Geneious plugin[91] and the *gsi* statistic.[3]

Notes: The decoder for the "+" and "–" is as follows: K2P distance/P(RD)/posterior probability/P(AB)/*gsi*. An example of all measures indicating species distinctiveness would look like +/+/+/+/+ on the figure. Significance was determined by: >1% difference/>0.05/>0.70/ the Bonferroni correction described in the methods section.

Table 2. Tip to root approach for *B. tabaci*. Clade numbers refer to boxed individuals found in Figures 2 and 3 and supplemental data, the Table 1 legend contains information about each of the columns.

Clade 1	Clade 2	Intra dist	K2P	Intra/Inter	P ID(Strict)
Clade 1—Outgroups					
1-1 (2)	1-2 (2)	0.166	4.858	0.03	0.57 (0.42, 0.72)
1-2 (2)	1-1 (2)	4.625	4.858	0.95	0.10 (0.00E+00, 0.26)
1-4 (2)	1-5 (2)	0.064	0.135	0.47	0.35 (0.20, 0.50)
1-5 (2)	1-4 (2)	0.128	0.135	0.95	0.11 (0.00E+00, 0.26)
1-3 (5)	1-6 (5)	4.309	5.995	0.72	0.45 (0.32, 0.58)
1-6 (5)	1-3 (5)	0.151	5.995	0.03	0.92 (0.79, 1.0)
Clade 2—SubSaharan Africa					
2-1 (4)	2-2 (3)	0.147	0.304	0.48	0.54 (0.40, 0.69)
2-2 (3)	2-1 (4)	0.274	0.304	0.9	0.19 (3.38E-03, 0.37)
2-2b (7)	2-3 (2)	0.255	0.233	1.09	0.26 (0.15, 0.37)
2-3 (2)	2-2b (7)	0.085	0.233	0.36	0.40 (0.25, 0.56)
2-5 (6)	2-6 (2)	0.258	0.266	0.97	0.28 (0.15, 0.41)
2-6 (2)	2-5 (6)	0.062	0.266	0.23	0.47 (0.32, 0.62)
2-7 (8)	2-4 (9)	0.255	0.341	0.75	0.53 (0.42, 0.63)
2-8 (3)	2-9 (2)	0.102	0.164	0.62	0.38 (0.19, 0.56)
2-9 (2)	2-8 (3)	0.108	0.164	0.66	0.26 (0.10, 0.41)
2-4 (9)	2-10 (5)	0.242	0.251	0.96	0.46 (0.37, 0.55)
2-10 (5)	2-4 (9)	0.14	0.251	0.56	0.56 (0.43, 0.69)
2-11 (2)	2-12 (2)	0.066	0.115	0.57	0.30 (0.15, 0.45)
2-12 (2)	2-11 (2)	0.084	0.115	0.73	0.22 (0.06, 0.37)
2-13 (4)	2-4 (9)	0.102	0.234	0.43	0.57 (0.43, 0.72)
2-14 (2)	2-15 (2)	0.19	0.194	0.98	0.09 (0.00E+00, 0.25)
2-15 (2)	2-14 (2)	0.087	0.194	0.45	0.36 (0.21, 0.52)
2-16 (4)	2-19 (4)	0.175	0.234	0.75	0.36 (0.21, 0.51)
2-19 (4)	2-16 (4)	0.096	0.234	0.41	0.59 (0.45, 0.74)
2-17 (2)	2-18 (2)	0.043	0.112	0.38	0.40 (0.24, 0.55)
2-18 (2)	2-17 (2)	0.086	0.112	0.77	0.20 (0.04, 0.36)
2-20 (2)	20-19 (4)	0.125	0.177	0.71	0.23 (0.07, 0.39)
2-21 (2)	2-22 (2)	0.085	0.183	0.46	0.35 (0.20, 0.51)
2-22 (2)	2-21 (2)	0.12	0.183	0.66	0.25 (0.10, 0.41)
2-23 (5)	2-24 (2)	0.182	0.281	0.65	0.50 (0.37, 0.63)
2-24 (2)	2-23 (5)	0.134	0.281	0.48	0.35 (0.19, 0.50)
2-25 (8)	2-26 (4)	0.229	0.391	0.59	0.63 (0.53, 0.74)
2-26 (4)	2-25 (8)	0.147	0.391	0.38	0.61 (0.47, 0.76)
2-20b (36)	2-27 (13)	0.253	0.478	0.53	0.84 (0.78, 0.89)
2-27 (13)	2-20b (36)	0.309	0.478	0.65	0.75 (0.68, 0.81)
2-20c (49)	2-28 (2)	0.346	0.321	1.08	0.50 (0.44, 0.55)
2-28 (2)	2-20c (49)	0.077	0.321	0.24	0.47 (0.32, 0.62)
2-29 (3)	2-30 (2)	0.206	0.219	0.94	0.16 (0.00E+00, 0.35)
2-30 (2)	2-29 (3)	0.086	0.219	0.39	0.39 (0.24, 0.54)
2-31 (5)	2-32 (2)	0.202	0.194	1.04	0.23 (0.11, 0.36)
2-32 (2)	2-31 (5)	0.067	0.194	0.35	0.41 (0.26, 0.57)
2-33 (7)	2-34 (2)	0.192	0.233	0.82	0.47 (0.36, 0.58)
2-34 (2)	2-33 (7)	0.105	0.233	0.45	0.36 (0.21, 0.51)
2-35 (9)	2-36 (2)	0.205	0.27	0.76	0.60 (0.52, 0.69)
2-36 (2)	2-35 (9)	0.067	0.27	0.25	0.46 (0.31, 0.62)
2-37 (12)	2-38 (2)	0.229	0.377	0.61	0.76 (0.70, 0.83)
2-38 (2)	2-37 (12)	0.061	0.377	0.16	0.51 (0.36, 0.66)
2-20d (51)	2-39 (15)	0.344	1.852	0.19	0.93 (0.88, 0.98)
2-39 (15)	2-20d (51)	0.382	1.852	0.21	0.91 (0.84, 0.98)
2-40 (2)	2-41 (2)	0.084	0.131	0.65	0.26 (0.11, 0.42)
2-41 (2)	2-40 (2)	0.087	0.131	0.67	0.25 (0.09, 0.40)
2-42 (4)	2-43 (2)	0.116	0.179	0.64	0.43 (0.29, 0.58)

P ID(Liberal)	Av(MRCA-tips)	P(RD)	PP	P(AB)	*gsi*	*P*-value
0.96 (0.81, 1.0)	0.0831	0.05	1	1.10E-01	1.0	1.82E-03
0.40 (0.24, 0.55)	2.3124	0.05	0.57	1.10E-01	1.0	1.63E-03
0.69 (0.54, 0.85)	0.0318	0.05	0.17	0.11	1.0	1.80E-03
0.40 (0.24, 0.56)	0.0641	0.05	0.17	1.10E-01	1.0	1.97E-03
0.76 (0.66, 0.86)	2.7978	0.99	1	8.80E-04	1.0	1.97E-03
0.98 (0.87, 1.0)	0.1088	0.05	1	8.80E-04	1.0	1.82E-03
0.83 (0.72, 0.94)	0.1023	0.05	0.02	1.00E-02	1.0	5.00E-06
0.50 (0.35, 0.65)	0.1588	0.05	0.02	1.00E-02	1.0	2.49E-05
0.60 (0.53, 0.67)	0.1479	0.05	0	1.00E-02	1.0	1.97E-03
0.76 (0.60, 0.91)	0.0425	0.05	0.02	1.00E-02	1.0	5.00E-06
0.60 (0.50, 0.71)	0.1958	0.05	0.02	1.00E-02	1.0	5.00E-06
0.84 (0.69, 0.99)	0.0312	0.05	0.02	1.00E-02	1.0	1.97E-03
0.83 (0.76, 0.89)	0.1755	0.05	0	1.80E-09	1.0	5.00E-06
0.66 (0.51, 0.80)	0.0657	0.05	0.02	5.00E-02	1.0	2.49E-05
0.58 (0.42, 0.73)	0.0539	0.05	0.02	5.00E-02	1.0	1.97E-03
0.76 (0.71, 0.82)	0.1439	0.05	0	6.00E-10	1.0	5.00E-06
0.84 (0.74, 0.94)	0.083	0.05	0	1.50E-07	0.75	5.00E-06
0.63 (0.48, 0.79)	0.0328	0.05	0.03	1.10E-01	1.0	1.82E-03
0.53 (0.37, 0.69)	0.0422	0.05	0.02	1.10E-01	1.0	1.82E-03
0.85 (0.74, 0.96)	0.0575	0.05	0	0.00000097	1.0	5.00E-06
0.38 (0.22, 0.54)	0.095	0.05	0.04	1.10E-01	1.0	1.97E-03
0.71 (0.55, 0.86)	0.0435	0.05	0.02	1.10E-01	1.0	1.82E-03
0.68 (0.57, 0.79)	0.0969	0.05	0	9.70E-07	1.0	5.00E-06
0.86 (0.75, 0.97)	0.0561	0.05	0	9.70E-07	1.0	5.00E-06
0.75 (0.59, 0.90)	0.0213	0.05	0.01	1.10E-01	1.0	1.63E-03
0.51 (0.35, 0.67)	0.043	0.05	0.02	1.10E-01	1.0	1.80E-03
0.55 (0.39, 0.70)	0.0626	0.05	0.02	9.00E-05	1.0	1.80E-03
0.70 (0.54, 0.85)	0.0424	0.05	0.43	1.10E-01	1.0	1.82E-03
0.58 (0.42, 0.73)	0.0602	0.05	0.72	1.10E-01	1.0	1.82E-03
0.80 (0.70, 0.90)	0.1234	0.05	0.23	2.00E-02	1.0	1.82E-03
0.69 (0.53, 0.84)	0.0668	0.05	0.91	2.00E-02	0.83	5.00E-06
0.87 (0.81, 0.94)	0.1693	0.05	0.72	3.60E-04	1.0	5.00E-06
0.87 (0.76, 0.98)	0.0886	0.05	1	3.60E-04	1.0	5.00E-06
0.96 (0.93, 0.99)	0.1325	0.05	0	1.50E-13	1.0	5.00E-06
0.93 (0.88, 0.97)	0.2605	0.05	0.71	1.50E-13	1.0	5.00E-06
0.78 (0.75, 0.81)	0.1982	0.05	0.09	3.10E-05	0.18	3.35E-03
0.83 (0.68, 0.99)	0.0385	0.05	0.63	3.10E-05	0.18	3.35E-03
0.47 (0.32, 0.62)	0.133	0.05	0.09	5.00E-02	1.0	5.00E-06
0.74 (0.59, 0.90)	0.0428	0.05	0.1	5.00E-02	1.0	1.82E-03
0.55 (0.45, 0.66)	0.1184	0.05	0.03	2.00E-02	1.0	1.82E-03
0.77 (0.61, 0.92)	0.0337	0.05	0.1	2.00E-02	0.75	5.00E-06
0.80 (0.73, 0.86)	0.1153	0.05	0.09	1.00E-02	1.0	1.97E-03
0.70 (0.55, 0.86)	0.0527	0.05	0.95	1.00E-02	1.0	1.97E-03
0.86 (0.81, 0.91)	0.1277	0.05	0.98	3.64E-03	1.0	1.80E-03
0.83 (0.68, 0.98)	0.0337	0.05	1	3.64E-03	1.0	1.80E-03
0.93 (0.89, 0.98)	0.1738	0.05	1	1.69E-03	1.0	1.82E-03
0.88 (0.73, 1.0)	0.0304	0.05	1	1.69E-03	1.0	1.97E-03
0.98 (0.95, 1.0)	0.2356	0.05	0.69	1.10E-16	1.0	5.00E-06
0.97 (0.93, 1.0)	0.6632	1	0.85	1.10E-16	1.0	5.00E-06
0.58 (0.43, 0.74)	0.0421	0.05	0.1	1.10E-01	1.0	1.97E-03
0.57 (0.41, 0.73)	0.0437	0.05	0.1	1.10E-01	1.0	1.97E-03
0.74 (0.63, 0.85)	0.0653	0.96	0.03	3.00E-02	1.0	1.82E-03

(*Continued*)

Table 2. (*Continued*)

Clade 1	Clade 2	Intra dist	K2P	Intra/Inter	P ID(Strict)
2-43 (2)	2-42 (4)	0.132	0.179	0.74	0.21 (0.06, 0.37)
2-44 (6)	2-45 (2)	0.151	0.224	0.67	0.48 (0.35, 0.61)
2-45 (2)	2-44 (6)	0.133	0.224	0.59	0.29 (0.13, 0.44)
2-46 (8)	2-20e (66)	0.884	1.817	0.49	0.85 (0.80, 0.90)
2-20e (66)	2-46 (8)	0.182	1.817	0.1	0.90 (0.79, 1.0)
Clade 3—New World					
3-1 (6)	3-2 (2)	0.15	0.308	0.49	0.61 (0.48, 0.73)
3-2 (2)	3-1 (6)	0.064	0.308	0.21	0.49 (0.33, 0.64)
3-3 (8)	3-4 (2)	0.215	0.434	0.49	0.69 (0.58, 0.80)
3-4 (2)	3-3 (8)	0.325	0.434	0.75	0.21 (0.05, 0.37)
3-6 (3)	3-5 (10)	0.295	0.359	0.82	0.56 (0.48, 0.65)
3-5 (10)	3-6 (3)	0.084	0.359	0.23	0.64 (0.46, 0.81)
3-7 (13)	3-8 (3)	0.312	0.637	0.49	0.81 (0.74, 0.88)
3-8 (3)	3-7 (13)	0.099	0.637	0.16	0.69 (0.51, 0.86)
3-9 (16)	3-10 (2)	0.412	1.07	0.39	0.88 (0.82, 0.93)
3-10 (2)	3-9 (16)	0.082	1.07	0.08	0.55 (0.40, 0.70)
Clade 4—Italy					
4-1 (4)	4-2 (2)	0.123	0.167	0.73	0.37 (0.22, 0.52)
4-2 (2)	4-1 (4)	0.069	0.167	0.41	0.38 (0.23, 0.53)
Clade 5—Asia I, China, Australia					
5-1 (3)	5-2 (2)	0.301	0.251	1.2	0.00E+00 (0.00E+00, 0.18)
5-2 (2)	5-1 (3)	0.044	0.251	0.17	0.50 (0.35, 0.65)
5-4a (2)	5-4b (2)	0.066	0.12	0.55	0.31 (0.15, 0.46)
5-4b (2)	5-4a (2)	0.088	0.12	0.74	0.21 (0.06, 0.37)
5-3 (6)	5-4c (9)	0.236	0.299	0.79	0.40 (0.28, 0.53)
5-4c (9)	5-3 (6)	0.106	0.299	0.35	0.63 (0.49, 0.77)
5-5 (10)	5-6 (2)	0.252	0.284	0.89	0.52 (0.43, 0.61)
5-6 (2)	5-5 (10)	0.126	0.284	0.44	0.36 (0.21, 0.52)
5-8 (2)	5-9 (2)	0.126	0.169	0.74	0.21 (0.05, 0.37)
5-9 (2)	5-8 (2)	0.127	0.169	0.75	0.21 (0.05, 0.36)
5-7 (12)	5-10 (4)	0.26	0.318	0.82	0.65 (0.58, 0.72)
5-10 (4)	5-7 (12)	0.155	0.318	0.49	0.54 (0.39, 0.68)
5-11 (16)	5-12 (2)	0.278	0.254	1.09	0.48 (0.42, 0.53)
5-12 (2)	5-11 (16)	0.069	0.254	0.27	0.45 (0.30, 0.60)
5-13 (19)	5-15 (5)	0.276	2.818	0.1	0.96 (0.91, 1.0)
5-15 (5)	5-13 (19)	1.003	2.818	0.36	0.69 (0.57, 0.82)
5-17 (2)	5-18 (2)	0.088	0.167	0.52	0.32 (0.17, 0.48)
5-18 (2)	5-17 (2)	0.066	0.167	0.4	0.39 (0.23, 0.54)
5-16 (25)	5-19 (7)	1.273	3.101	0.41	0.87 (0.82, 0.92)
5-19 (7)	5-16 (25)	0.853	3.101	0.28	0.81 (0.71, 0.92)
Clade 6—Asia II					
6-1 (4)	6-2 (2)	0.245	0.226	1.08	0.13 (0.00E+00, 0.28)
6-2 (2)	6-1 (4)	0.063	0.226	0.28	0.45 (0.30, 0.60)
6-3 (6)	6-4 (2)	0.223	0.238	0.94	0.30 (0.18, 0.43)
6-4 (2)	6-3 (6)	0.079	0.238	0.33	0.42 (0.27, 0.57)
6-6 (2)	6-7 (2)	0.13	0.251	0.52	0.33 (0.17, 0.48)
6-7 (2)	6-6 (2)	0.289	0.251	1.15	3.89E-03 (0.00E+00, 0.16)
6-5 (8)	6-8 (5)	0.224	0.332	0.67	0.58 (0.47, 0.68)
6-8 (5)	6-5 (8)	0.219	0.332	0.66	0.49 (0.36, 0.62)
6-9 (13)	6-10 (2)	0.279	0.409	0.68	0.73 (0.66, 0.80)
6-10 (2)	6-9 (13)	0.189	0.409	0.46	0.36 (0.20, 0.51)
6-12 (4)	6-11 (16)	0.379	2.11	0.18	0.81 (0.69, 0.94)
6-11 (16)	6-12 (4)	0.389	2.11	0.18	0.93 (0.88, 0.98)

P ID(Liberal)	Av(MRCA-tips)	P(RD)	PP	P(AB)	*gsi*	*P*-value
0.53 (0.37, 0.68)	0.0662	0.89	0.1	3.00E-02	1.0	1.82E-03
0.79 (0.69, 0.89)	0.0903	0.61	0.09	1.00E-02	1.0	1.63E-03
0.62 (0.46, 0.77)	0.0665	0.73	0.94	1.00E-02	1.0	1.63E-03
0.96 (0.93, 0.99)	0.967	0.98	0.43	1.80E-12	1.0	5.00E-06
0.97 (0.91, 1.0)	0.1126	0.74	0.98	1.80E-12	1.0	5.00E-06
0.87 (0.77, 0.97)	0.1256	0.05	0.99	1.00E-02	0.86	5.00E-06
0.86 (0.70, 1.0)	0.0318	0.05	1	1.00E-02	0.14	0.323
0.89 (0.83, 0.96)	0.1727	0.05	0.43	4.94E-03	1.0	1.97E-03
0.52 (0.36, 0.68)	0.1624	0.05	0.85	4.94E-03	1.0	5.00E-06
0.84 (0.78, 0.89)	0.2139	1	0.43	5.80E-04	1.0	5.00E-06
0.87 (0.73, 1.0)	0.0555	1	0.75	5.80E-04	1.0	5.00E-06
0.95 (0.90, 0.99)	0.2223	0.38	0.95	2.30E-04	0.50	3.65E-03
0.92 (0.77, 1.0)	0.0569	0.05	1	2.30E-04	0.50	3.65E-03
0.96 (0.94, 0.99)	0.3412	0.84	0.86	7.60E-04	1.0	1.82E-03
0.94 (0.78, 1.0)	0.041	0.05	1	7.60E-04	1.0	1.82E-03
0.69 (0.58, 0.80)	0.0864	0.1	0.16	3.00E-02	1.0	5.00E-06
0.73 (0.57, 0.88)	0.0346	0.05	0.17	3.00E-02	1.0	1.97E-03
0.33 (0.18, 0.48)	0.1647	0.05	0.91	5.00E-02	1.0	2.49E-05
0.88 (0.72, 1.0)	0.0218	0.05	0.23	5.00E-02	1.0	1.82E-03
0.64 (0.49, 0.80)	0.033	0.05	0.17	1.10E-01	1.0	1.97E-03
0.53 (0.37, 0.69)	0.0442	0.05	0.18	1.10E-01	1.0	1.97E-03
0.72 (0.62, 0.83)	0.1548	0.05	0.88	1.06E-03	1.0	5.00E-06
0.88 (0.77, 0.99)	0.0599	0.05	0.66	1.06E-03	1.0	5.00E-06
0.81 (0.75, 0.86)	0.1592	0.05	0.16	2.75E-03	1.0	1.63E-03
0.71 (0.56, 0.86)	0.0628	0.05	0.05	2.75E-03	1.0	1.63E-03
0.52 (0.37, 0.68)	0.0629	0.05	0.06	1.10E-01	1.0	1.80E-03
0.52 (0.36, 0.68)	0.0635	0.05	0.05	1.10E-01	1.0	1.97E-03
0.88 (0.84, 0.93)	0.1792	0.05	0	7.30E-05	1.0	5.00E-06
0.82 (0.72, 0.93)	0.0846	0.05	0	7.30E-05	1.0	5.00E-06
0.77 (0.74, 0.80)	0.179	0.05	0	7.60E-04	1.0	1.97E-03
0.81 (0.66, 0.97)	0.0347	0.05	0.06	7.60E-04	1.0	1.97E-03
0.99 (0.96, 1.0)	0.2005	0.05	0.17	4.70E-07	1.0	5.00E-06
0.92 (0.81, 1.0)	1.2642	0.05	0.87	1.50E-06	1.0	5.00E-06
0.66 (0.50, 0.81)	0.0438	0.05	1	1.10E-01	1.0	1.97E-03
0.74 (0.58, 0.89)	0.0332	0.05	0.34	1.10E-01	1.0	1.97E-03
0.96 (0.94, 0.99)	1.309	1	1	1.90E-08	0.69	5.00E-06
0.93 (0.87, 0.99)	1.0981	1	0.98	1.90E-08	0.69	5.00E-06
0.44 (0.32, 0.56)	0.1474	0.05	0.17	3.00E-02	1.0	1.97E-03
0.81 (0.66, 0.96)	0.0317	0.05	0.14	3.00E-02	1.0	5.00E-06
0.63 (0.53, 0.73)	0.133	0.05	0.13	1.00E-02	1.0	5.00E-06
0.78 (0.62, 0.93)	0.0397	0.05	0.73	1.00E-02	1.0	1.97E-03
0.66 (0.51, 0.82)	0.0649	0.05	0.11	1.10E-01	1.0	1.80E-03
0.27 (0.11, 0.43)	0.1445	0.05	0.11	1.10E-01	1.0	1.97E-03
0.85 (0.79, 0.92)	0.1368	0.05	0.74	1.20E-04	1.0	5.00E-06
0.79 (0.69, 0.90)	0.1272	0.05	0.09	1.20E-04	1.0	5.00E-06
0.92 (0.88, 0.96)	0.1694	0.05	0.31	1.36E-03	1.0	1.82E-03
0.70 (0.54, 0.85)	0.0945	0.05	0.85	1.36E-03	1.0	1.82E-03
0.96 (0.86, 1.0)	0.4519	0.05	1	4.90E-06	1.0	5.00E-06
0.98 (0.95, 1.0)	0.8325	0.05	0.51	4.90E-06	0.93	5.00E-06

(*Continued*)

Table 2. (*Continued*)

Clade 1	Clade 2	Intra dist	K2P	Intra/Inter	P ID(Strict)
6-14 (3)	6-15 (2)	0.367	1.114	0.33	0.57 (0.39, 0.75)
6-15 (2)	6-14 (3)	0.119	1.114	0.11	0.54 (0.39, 0.69)
6-13 (5)	6-16 (5)	1.044	2.628	0.4	0.87 (0.82, 0.93)
6-16 (5)	6-13 (5)	0.79	2.628	0.3	0.73 (0.61, 0.86)
6-16b (26)	6-17 (7)	1.548	2.519	0.61	0.81 (0.75, 0.86)
6-17 (7)	6-16b (26)	0.37	2.519	0.15	0.88 (0.77, 0.98)
6-17b (26)	6-18 (4)	1.836	2.826	0.65	0.79 (0.74, 0.85)
6-18 (4)	6-17b (26)	0.138	2.826	0.05	0.84 (0.70, 0.98)
Clade 7					
7-1 (2)	7-2 (2)	0.17	0.179	0.95	0.11 (0.00E+00, 0.26)
7-2 (2)	7-1 (2)	0.085	0.179	0.47	0.35 (0.20, 0.50)
7-3 (4)	7-4 (4)	0.162	0.219	0.74	0.37 (0.22, 0.51)
7-4 (4)	7-3 (4)	0.132	0.219	0.6	0.46 (0.31, 0.61)
7-5 (8)	7-6 (2)	0.188	0.184	1.02	0.32 (0.21, 0.43)
7-6 (2)	7-5 (8)	0.068	0.184	0.37	0.40 (0.25, 0.55)
7-7 (10)	7-8 (2)	0.184	0.226	0.81	0.57 (0.48, 0.66)
7-8 (2)	7-7 (10)	0.069	0.226	0.3	0.44 (0.28, 0.59)
Clade 8—Middle East/Asia Minor					
8-2 (3)	8-3 (2)	0.105	0.164	0.64	0.37 (0.18, 0.55)
8-3 (2)	8-2 (3)	0.059	0.164	0.36	0.41 (0.25, 0.56)
8-1 (5)	8-3b (10)	0.125	0.267	0.47	0.62 (0.49, 0.75)
8-3b (10)	8-1 (5)	0.136	0.267	0.51	0.59 (0.46, 0.72)
8-4 (12)	8-5 (2)	0.209	0.309	0.68	0.73 (0.66, 0.80)
8-5 (2)	8-4 (12)	0.064	0.309	0.21	0.48 (0.33, 0.64)
8-6 (14)	8-7 (5)	0.234	0.452	0.52	0.80 (0.73, 0.87)
8-7 (5)	8-6 (14)	0.138	0.452	0.31	0.73 (0.60, 0.85)
8-8 (3)	8-9 (2)	0.122	0.152	0.8	0.26 (0.07, 0.44)
8-9 (2)	8-8 (3)	0.064	0.152	0.42	0.38 (0.22, 0.53)
8-9b (5)	8-10 (2)	0.134	0.204	0.66	0.49 (0.36, 0.62)
8-10 (2)	8-9b (5)	0.153	0.204	0.75	0.21 (0.05, 0.37)
8-7b (19)	8-11 (8)	0.317	0.48	0.66	0.79 (0.74, 0.84)
8-11 (8)	8-7b (19)	0.163	0.48	0.34	0.78 (0.67, 0.89)
8-12 (28)	8-13 (2)	0.376	0.46	0.82	0.71 (0.66, 0.76)
8-13 (2)	8-12 (28)	0.11	0.46	0.24	0.47 (0.32, 0.62)
8-14 (30)	8-15 (3)	0.386	0.489	0.79	0.73 (0.67, 0.78)
8-15 (3)	8-14 (30)	0.122	0.489	0.25	0.63 (0.45, 0.80)
8-16 (3)	8-17 (2)	0.205	0.258	0.79	0.26 (0.08, 0.44)
8-17 (2)	8-16 (3)	0.097	0.258	0.38	0.40 (0.25, 0.55)
8-18 (5)	8-19 (4)	0.226	0.343	0.66	0.49 (0.36, 0.62)
8-19 (4)	8-18 (5)	0.193	0.343	0.56	0.49 (0.34, 0.63)
8-15b (30)	8-20 (9)	0.402	0.573	0.7	0.77 (0.72, 0.82)
8-20 (9)	8-15b (30)	0.286	0.573	0.5	0.74 (0.65, 0.82)
8-21 (2)	8-22 (2)	0.067	0.199	0.34	0.42 (0.27, 0.57)
8-22 (2)	8-21 (2)	0.154	0.199	0.78	0.19 (0.04, 0.35)
8-23 (4)	8-24 (3)	0.17	0.224	0.76	0.35 (0.21, 0.50)
8-24 (3)	8-23 (4)	0.109	0.224	0.49	0.47 (0.29, 0.65)
8-25 (7)	8-20 (9)	0.192	0.341	0.56	0.65 (0.54, 0.75)
8-26 (3)	8-27 (2)	0.081	0.255	0.32	0.58 (0.40, 0.76)
8-27 (2)	8-26 (3)	0.23	0.255	0.9	0.13 (0.00E+00, 0.29)
8-27b (5)	8-28 (2)	0.2	0.219	0.91	0.32 (0.19, 0.45)
8-28 (2)	8-27b (5)	0.091	0.219	0.42	0.38 (0.23, 0.53)
8-29 (7)	8-32 (5)	0.204	0.23	0.89	0.42 (0.32, 0.53)
8-32 (5)	8-29 (7)	0.142	0.23	0.62	0.52 (0.39, 0.64)
8-30 (3)	8-31 (2)	0.105	0.17	0.61	0.38 (0.20, 0.56)

P ID(Liberal)	Av(MRCA-tips)	P(RD)	PP	P(AB)	*gsi*	*P*-value
0.82 (0.68, 0.97)	0.2575	0.05	1	5.00E-02	1.0	5.00E-06
0.92 (0.77, 1.0)	0.0596	0.05	1	5.00E-02	1.0	1.80E-03
0.96 (0.94, 0.99)	1.0862	0.05	1	1.20E-06	1.0	5.00E-06
0.93 (0.83, 1.0)	0.5949	0.05	1	1.20E-06	1.0	5.00E-06
0.95 (0.92, 0.98)	1.4011	0.05	0.98	1.40E-08	1.0	5.00E-06
0.96 (0.89, 1.0)	0.3512	0.05	1	1.40E-08	1.0	5.00E-06
0.95 (0.92, 0.97)	1.4719	0.05	1	8.40E-07	1.0	5.00E-06
0.97 (0.86, 1.0)	0.0957	0.05	1	8.40E-07	1.0	5.00E-06
0.40 (0.24, 0.56)	0.0849	0.05	0.09	1.10E-01	1.0	1.80E-03
0.69 (0.54, 0.85)	0.0424	0.05	0.07	1.10E-01	1.0	1.82E-03
0.69 (0.58, 0.80)	0.0895	0.05	0.01	4.08E-03	1.0	5.00E-06
0.77 (0.66, 0.88)	0.0909	0.05	0.07	4.08E-03	1.0	5.00E-06
0.67 (0.60, 0.73)	0.1093	0.05	0.02	4.94E-03	1.0	1.82E-03
0.75 (0.60, 0.91)	0.0342	0.05	0.09	4.94E-03	1.0	1.82E-03
0.84 (0.79, 0.89)	0.1141	0.05	0.53	2.75E-03	1.0	1.80E-03
0.79 (0.64, 0.95)	0.0345	0.05	0.54	2.75E-03	1.0	1.80E-03
0.65 (0.50, 0.79)	0.0646	0.05	0.34	5.00E-02	1.0	5.00E-06
0.76 (0.61, 0.91)	0.0295	0.05	0.19	5.00E-02	1.0	1.63E-03
0.88 (0.78, 0.98)	0.1016	0.05	0.1	8.80E-04	1.0	5.00E-06
0.86 (0.76, 0.96)	0.0874	0.05	0.35	8.80E-04	1.0	5.00E-06
0.92 (0.88, 0.96)	0.1917	0.05	0.42	1.69E-03	1.0	1.63E-03
0.85 (0.70, 1.0)	0.0322	0.05	0.6	1.69E-03	1.0	1.63E-03
0.94 (0.90, 0.99)	0.213	0.05	1	9.50E-06	0.92	5.00E-06
0.93 (0.83, 1.0)	0.1054	0.05	0.41	9.50E-06	1.0	5.00E-06
0.55 (0.40, 0.70)	0.0753	0.05	0.08	5.00E-02	1.0	5.00E-06
0.72 (0.57, 0.88)	0.0321	0.05	0.09	5.00E-02	1.0	1.82E-03
0.80 (0.69, 0.90)	0.0805	0.05	0.02	2.00E-02	1.0	1.82E-03
0.52 (0.36, 0.68)	0.0764	0.05	0.09	2.00E-02	1.0	5.00E-06
0.94 (0.92, 0.97)	0.2531	0.05	0.5	3.40E-08	1.0	5.00E-06
0.92 (0.85, 0.98)	0.1132	0.05	0.29	3.40E-08	1.0	5.00E-06
0.91 (0.88, 0.94)	0.3377	0.05	0.07	1.50E-04	1.0	1.97E-03
0.83 (0.68, 0.99)	0.0552	0.05	0.02	1.50E-04	1.0	1.97E-03
0.92 (0.89, 0.95)	0.3638	0.05	0	1.10E-05	1.0	5.00E-06
0.87 (0.72, 1.0)	0.0757	0.05	0.01	1.10E-05	1.0	2.00E-03
0.56 (0.41, 0.71)	0.1299	0.05	0.42	5.00E-02	0.19	7.41E-05
0.75 (0.60, 0.90)	0.0487	0.05	0.3	5.00E-02	1.0	1.82E-03
0.79 (0.69, 0.90)	0.1382	0.05	0.09	1.98E-03	1.0	5.00E-06
0.79 (0.68, 0.90)	0.1174	0.05	0.57	1.98E-03	1.0	5.00E-06
0.94 (0.91, 0.97)	0.361	0.05	0	1.30E-24	1.0	5.00E-06
0.92 (0.86, 0.97)	0.1726	0.05	0	1.20E-13	1.0	5.00E-06
0.77 (0.62, 0.93)	0.0337	0.05	0.95	1.10E-01	1.0	1.97E-03
0.50 (0.35, 0.66)	0.0772	0.05	0.58	1.10E-01	1.0	1.97E-03
0.68 (0.56, 0.79)	0.0995	0.05	0.01	1.00E-02	1.0	5.00E-06
0.73 (0.58, 0.88)	0.0695	0.05	0.02	1.00E-02	1.0	5.00E-06
0.88 (0.82, 0.94)	0.1146	0.05	0	8.80E-12	1.0	5.00E-06
0.83 (0.68, 0.97)	0.0501	0.05	0.75	5.00E-02	1.0	5.00E-06
0.43 (0.27, 0.58)	0.1151	0.05	0.79	5.00E-02	1.0	1.82E-03
0.64 (0.54, 0.75)	0.121	0.05	0.01	2.00E-02	1.0	1.82E-03
0.73 (0.57, 0.88)	0.0456	0.05	0.02	2.00E-02	1.0	5.00E-06
0.76 (0.70, 0.83)	0.1268	0.05	0	8.80E-12	0.83	5.00E-06
0.81 (0.71, 0.91)	0.0884	0.05	0	1.10E-09	1.0	5.00E-06
0.66 (0.51, 0.81)	0.0746	0.05	0.02	5.00E-02	0.50	0.0041

(*Continued*)

Table 2. (*Continued*)

Clade 1	Clade 2	Intra dist	K2P	Intra/Inter	P ID(Strict)
8-31 (2)	8-30 (3)	0.089	0.17	0.52	0.32 (0.17, 0.48)
8-33 (2)	8-34 (2)	0.066	0.12	0.55	0.31 (0.16, 0.47)
8-34 (2)	8-33 (2)	0.087	0.12	0.73	0.22 (0.06, 0.38)
8-35 (4)	8-38 (4)	0.106	0.167	0.63	0.44 (0.29, 0.58)
8-38 (4)	8-35 (4)	0.142	0.167	0.85	0.29 (0.14, 0.44)
8-36 (2)	8-37 (2)	0.088	0.159	0.56	0.31 (0.15, 0.46)
8-37 (2)	8-36 (2)	0.129	0.159	0.81	0.18 (0.02, 0.33)
8-39 (2)	8-40 (2)	0.09	0.133	0.67	0.25 (0.09, 0.40)
8-40 (2)	8-39 (2)	0.09	0.133	0.67	0.25 (0.09, 0.40)
8-41 (4)	8-44 (4)	0.119	0.223	0.53	0.51 (0.36, 0.65)
8-44 (4)	8-41 (4)	0.123	0.223	0.55	0.50 (0.35, 0.64)
8-42 (2)	8-43 (2)	0.089	0.139	0.64	0.26 (0.11, 0.42)
8-43 (2)	8-42 (2)	0.092	0.139	0.66	0.25 (0.10, 0.41)
8-45 (2)	8-44 (4)	0.11	0.187	0.59	0.29 (0.13, 0.44)
8-45b (69)	8-46 (5)	0.397	0.415	0.96	0.61 (0.56, 0.66)
8-46 (5)	8-45 (69)	0.143	0.415	0.34	0.70 (0.58, 0.83)
Clade 9—Mediterranean					
9-1 (5)	9-2 (3)	0.134	0.254	0.53	0.58 (0.45, 0.70)
9-2 (3)	9-1 (5)	0.141	0.254	0.56	0.42 (0.24, 0.60)
9-4 (3)	9-5 (2)	0.156	0.282	0.55	0.42 (0.24, 0.60)
9-5 (2)	9-4 (3)	0.192	0.282	0.68	0.24 (0.09, 0.40)
9-3 (8)	9-6 (5)	0.199	0.298	0.67	0.58 (0.47, 0.69)
9-6 (5)	9-3 (8)	0.235	0.298	0.79	0.40 (0.27, 0.53)
9-7 (2)	9-8 (2)	0.087	0.142	0.62	0.28 (0.12, 0.43)
9-8 (2)	9-7 (2)	0.107	0.142	0.75	0.21 (0.05, 0.36)
9-9 (4)	9-12 (4)	0.127	0.221	0.57	0.48 (0.33, 0.62)
9-12 (4)	9-9 (4)	0.137	0.221	0.62	0.45 (0.30, 0.59)
9-10 (2)	9-11 (2)	0.068	0.151	0.45	0.36 (0.21, 0.51)
9-11 (2)	9-10 (2)	0.149	0.151	0.99	0.09 (0.00E+00, 0.25)
9-13 (21)	9-14 (2)	0.246	0.312	0.79	0.73 (0.68, 0.78)
9-14 (2)	9-13 (21)	0.197	0.312	0.63	0.27 (0.11, 0.42)
9-16 (25)	9-17 (2)	0.264	0.319	0.83	0.71 (0.65, 0.76)
9-17 (2)	9-16 (25)	0.086	0.319	0.27	0.45 (0.30, 0.60)
9-19 (3)	9-20 (2)	0.102	0.153	0.67	0.34 (0.16, 0.53)
9-20 (2)	9-19 (3)	0.087	0.153	0.57	0.30 (0.15, 0.46)
9-21 (5)	9-22 (3)	0.131	0.247	0.53	0.58 (0.45, 0.70)
9-22 (3)	9-21 (5)	0.22	0.247	0.89	0.20 (0.01, 0.38)
9-23 (3)	9-24 (5)	0.237	0.248	0.96	0.47 (0.38, 0.55)
9-24 (5)	9-23 (3)	0.067	0.248	0.27	0.45 (0.30, 0.60)
9-25 (11)	9-30 (9)	0.238	0.342	0.69	0.64 (0.55, 0.73)
9-30 (9)	9-25 (11)	0.215	0.342	0.63	0.68 (0.59, 0.76)
9-26 (4)	9-27 (2)	0.147	0.246	0.6	0.46 (0.32, 0.61)
9-27 (2)	9-26 (4)	0.115	0.246	0.47	0.35 (0.20, 0.51)
9-28 (6)	9-29 (3)	0.198	0.245	0.81	0.39 (0.26, 0.52)
9-29 (3)	9-28 (6)	0.116	0.245	0.47	0.47 (0.29, 0.65)
9-31 (20)	9-18 (27)	0.271	0.488	0.56	0.83 (0.77, 0.88)
9-18 (27)	9-31 (20)	0.288	0.488	0.59	0.82 (0.76, 0.87)
9-33 (48)	9-34 (2)	0.376	0.421	0.89	0.66 (0.61, 0.71)
9-34 (2)	9-33 (48)	0.084	0.421	0.2	0.49 (0.34, 0.64)
9-32 (3)	9-31b (47)	0.382	0.35	1.09	0.48 (0.42, 0.53)
9-31b (47)	9-32 (3)	0.104	0.35	0.3	0.59 (0.42, 0.77)
9-36 (2)	9-37 (2)	0.13	0.167	0.78	0.19 (0.04, 0.35)
9-37 (2)	9-36 (2)	0.109	0.167	0.66	0.26 (0.10, 0.41)
9-35 (50)	9-38 (5)	0.379	0.553	0.68	0.78 (0.73, 0.83)

P ID(Liberal)	Av(MRCA-tips)	P(RD)	PP	P(AB)	*gsi*	*P*-value
0.66 (0.50, 0.82)	0.0445	0.05	0.02	5.00E-02	1.0	1.82E-03
0.64 (0.49, 0.80)	0.033	0.05	0.02	1.10E-01	1.0	1.82E-03
0.53 (0.38, 0.69)	0.0437	0.05	0.02	1.10E-01	1.0	1.82E-03
0.75 (0.64, 0.86)	0.0602	0.05	0	1.70E-08	1.0	5.00E-06
0.61 (0.50, 0.73)	0.0794	0.05	0	1.70E-08	1.0	5.00E-06
0.64 (0.48, 0.79)	0.0442	0.05	0.02	1.10E-01	1.0	1.97E-03
0.48 (0.32, 0.64)	0.0645	0.05	0.02	1.10E-01	1.0	1.63E-03
0.57 (0.41, 0.72)	0.045	0.05	0.02	1.10E-01	0.07	0.186
0.57 (0.41, 0.72)	0.0449	0.05	0.02	1.10E-01	0.33	0.0078
0.80 (0.69, 0.91)	0.0667	0.05	0	1.70E-08	1.0	5.00E-06
0.79 (0.68, 0.90)	0.0695	0.05	0	1.70E-08	1.0	5.00E-06
0.59 (0.43, 0.74)	0.0445	0.05	0.02	1.10E-01	1.0	1.82E-03
0.57 (0.42, 0.73)	0.046	0.05	0.02	1.10E-01	1.0	1.63E-03
0.62 (0.46, 0.77)	0.0551	0.05	0.01	8.30E-06	1.0	1.97E-03
0.85 (0.83, 0.88)	0.229	0.05	0	7.80E-10	1.0	5.00E-06
0.92 (0.82, 1.0)	0.1141	0.05	0.51	7.80E-10	1.0	5.00E-06
0.85 (0.75, 0.95)	0.1027	0.05	0.03	1.00E-02	1.0	5.00E-06
0.69 (0.54, 0.84)	0.1047	0.05	0.03	1.00E-02	1.0	5.00E-06
0.69 (0.55, 0.84)	0.099	0.05	0.28	5.00E-02	1.0	2.49E-05
0.56 (0.41, 0.72)	0.096	0.05	0.55	5.00E-02	1.0	1.97E-03
0.85 (0.79, 0.92)	0.1265	0.05	0	5.10E-06	0.72	5.00E-06
0.72 (0.62, 0.83)	0.1412	0.05	0.01	2.00E-05	0.77	5.00E-06
0.60 (0.45, 0.76)	0.0437	0.05	0.04	1.10E-01	1.0	1.97E-03
0.52 (0.36, 0.68)	0.0533	0.05	0.04	1.10E-01	1.0	1.97E-03
0.78 (0.67, 0.89)	0.0709	0.05	0	5.20E-05	1.0	5.00E-06
0.76 (0.65, 0.87)	0.0755	0.05	0	1.60E-05	1.0	5.00E-06
0.70 (0.55, 0.86)	0.0341	0.05	0.04	1.10E-01	1.0	5.00E-06
0.37 (0.21, 0.53)	0.0746	0.05	0.03	1.10E-01	1.0	1.97E-03
0.92 (0.89, 0.95)	0.1537	0.05	0	3.50E-04	1.0	1.82E-03
0.59 (0.44, 0.75)	0.0986	0.05	0.52	3.50E-04	1.0	5.00E-06
0.91 (0.88, 0.94)	0.2377	0.05	0.03	2.10E-04	1.0	1.97E-03
0.82 (0.66, 0.97)	0.0431	0.05	0.04	2.10E-04	1.0	1.97E-03
0.63 (0.48, 0.78)	0.0658	0.05	0.09	5.00E-02	1.0	2.49E-05
0.63 (0.48, 0.79)	0.0433	0.05	0.11	5.00E-02	1.0	1.97E-03
0.85 (0.75, 0.95)	0.079	0.05	0.04	1.00E-02	1.0	2.49E-05
0.50 (0.35, 0.65)	0.125	0.05	0.11	1.00E-02	1.0	5.00E-06
0.77 (0.71, 0.82)	0.1638	0.05	0.05	3.64E-03	1.0	1.97E-03
0.82 (0.66, 0.97)	0.0335	0.05	0.03	3.64E-03	1.0	1.97E-03
0.88 (0.82, 0.93)	0.1672	0.05	0	6.20E-07	1.0	5.00E-06
0.89 (0.84, 0.95)	0.1321	0.05	0	6.20E-07	1.0	5.00E-06
0.77 (0.66, 0.88)	0.1038	0.05	0.67	3.00E-02	1.0	5.00E-06
0.69 (0.54, 0.85)	0.0577	0.05	0.76	3.00E-02	1.0	2.49E-05
0.71 (0.61, 0.81)	0.1311	0.05	0.02	2.98E-03	1.0	5.00E-06
0.74 (0.59, 0.88)	0.0726	0.05	0.04	2.98E-03	1.0	5.00E-06
0.96 (0.93, 0.98)	0.2421	0.05	0	4.40E-15	1.0	5.00E-06
0.95 (0.93, 0.98)	0.1728	0.05	0	4.40E-15	1.0	5.00E-06
0.88 (0.86, 0.91)	0.2679	0.05	1	2.70E-05	1.0	5.00E-06
0.86 (0.71, 1.0)	0.0422	0.05	0.34	2.70E-05	1.0	1.82E-03
0.77 (0.74, 0.80)	0.2474	0.05	0	2.00E-06	0.75	5.00E-06
0.84 (0.69, 0.98)	0.0736	0.05	0.04	2.00E-06	1.0	5.00E-06
0.50 (0.34, 0.66)	0.0651	0.05	0.12	1.10E-01	1.0	2.49E-05
0.58 (0.42, 0.73)	0.0547	0.05	0.12	1.10E-01	1.0	2.49E-05
0.94 (0.91, 0.97)	0.3436	0.05	1	7.60E-09	1.0	5.00E-06

(*Continued*)

Table 2. (*Continued*)

Clade 1	Clade 2	Intra dist	K2P	Intra/Inter	P ID(Strict)
9-38 (5)	9-35 (50)	0.159	0.553	0.29	0.74 (0.62, 0.87)
9-40 (3)	9-41 (2)	0.159	0.236	0.67	0.34 (0.16, 0.52)
9-41 (2)	9-40 (3)	0.177	0.236	0.75	0.21 (0.05, 0.36)
9-42 (5)	9-43 (4)	0.207	0.274	0.76	0.42 (0.30, 0.55)
9-43 (4)	9-42 (5)	0.154	0.274	0.56	0.49 (0.34, 0.63)
9-44 (9)	9-45 (3)	0.235	0.283	0.83	0.56 (0.47, 0.64)
9-45 (3)	9-44 (9)	0.136	0.283	0.48	0.47 (0.29, 0.65)
9-46 (13)	9-53 (25)	0.259	0.468	0.55	0.79 (0.72, 0.86)
9-53 (25)	9-46 (13)	0.245	0.468	0.52	0.72 (0.64, 0.81)
9-47 (2)	9-48 (2)	0.146	0.195	0.75	0.21 (0.05, 0.36)
9-48 (2)	9-47 (2)	0.111	0.195	0.57	0.30 (0.15, 0.46)
9-50 (2)	9-51 (2)	0.087	0.288	0.3	0.44 (0.28, 0.59)
9-51 (2)	9-50 (2)	0.403	0.288	1.4	0.00E+00 (0.00E+00, 0.04)
9-49 (4)	9-52 (5)	0.173	0.279	0.62	0.45 (0.30, 0.59)
9-52 (5)	9-49 (4)	0.273	0.279	0.98	0.20 (0.05, 0.35)
9-39 (58)	9-54 (23)	0.405	0.91	0.45	0.86 (0.81, 0.91)
9-54 (23)	9-39 (58)	0.364	0.91	0.4	0.87 (0.82, 0.92)
9-55 (81)	9-56 (6)	0.61	0.651	0.94	0.63 (0.58, 0.68)
9-56 (6)	9-55 (81)	0.117	0.651	0.18	0.81 (0.69, 0.94)
9-57 (87)	9-58 (3)	0.613	0.657	0.93	0.63 (0.58, 0.68)
9-58 (3)	9-57 (87)	0.09	0.657	0.14	0.70 (0.52, 0.88)

diversity that does not support the current sub-species divisions (*L. dispar*). By examining how different species delimitation measures affect the characterization of species boundaries within each of these datasets, the analyses revealed levels of distinctiveness within them that have implications for interpretation of the data in the context of the currently accepted taxonomy of these species. Much of this was not apparent from the commonly used K2P inter-species distance or PP.

Species delimitation of *B. tabaci*

The results suggest that for *B. tabaci* there is more taxonomic distinctiveness than previously reported.[63,69,70] Of particular interest from a biosecurity context are clades 8 and 9 (Fig. 2 and supplemental data) which relate to the putative species MEAM1 and MED,[70] respectively. These are both globally invasive, resistant to a wide range of insecticides[100] and cause severe economic losses.[67] Within these, clade 8-6, (haplotypes from Yemen, Saudi Arabia, Iran and Kuwait) and clade 9-33 (haplotypes from France, China, Spain, Croatia, Sudan, Morocco, Uruguay, Egypt, Ghana, Cameroon, Crete, Algeria, Taiwan) are significant for five of the six measures of distinctiveness. Clade 8-6 is of particular biosecurity interest as none of the haplotypes that belong to this clade have so far been detected beyond what is regarded as part of the home range of MEAM1. Given the economic damage caused by other members of clade 8, measures need to be considered to prevent incursions by clade 8-6. In contrast, clade 9-33 is made up of haplotypes from the presumed home range, Algeria, Croatia, Egypt, Morocco, Spain and possibly France and Sudan and those which have already spread over the past 10 years to countries outside the Mediterranean Basin home range to Cameroon, China, Ghana, Taiwan and Uruguay.[101] Therefore, despite the concern that mtCOI DNA is not necessarily adequate as a sole source of species-defining data,[102,103] the level and consistency of distinctiveness here indicates these groups should be investigated further for possible "species" status using an integrative taxonomy approach.[104] Importantly, neither clade 8-6 nor 9-33 have significant K2P distances and would be overlooked if relying on standard DNA barcoding species delimitation practices.

However, elevation of these new groups to "species" status would have major impacts on the regulation of these new species from growers, governmental agencies, chemical companies, etc, in this large range

P ID(Liberal)	Av(MRCA-tips)	P(RD)	PP	P(AB)	*gsi*	*P*-value
0.93 (0.83, 1.0)	0.0973	0.05	0.16	7.60E-09	1.0	5.00E-06
0.63 (0.48, 0.77)	0.0938	0.05	0.09	5.00E-02	1.0	2.49E-05
0.52 (0.36, 0.68)	0.0886	0.05	0.06	5.00E-02	1.0	1.97E-03
0.74 (0.64, 0.84)	0.1196	0.05	0.01	1.97E-03	1.0	5.00E-06
0.79 (0.68, 0.90)	0.1085	0.05	0.03	1.98E-03	1.0	5.00E-06
0.83 (0.78, 0.89)	0.1373	0.05	0.01	8.20E-04	1.0	5.00E-06
0.73 (0.59, 0.88)	0.0973	0.05	0.06	8.20E-04	1.0	5.00E-06
0.94 (0.90, 0.98)	0.1731	0.05	0.36	7.90E-08	1.0	5.00E-06
0.91 (0.86, 0.97)	0.2168	0.05	0.3	7.90E-08	1.0	5.00E-06
0.52 (0.36, 0.68)	0.0732	0.05	0.7	1.10E-01	1.0	1.97E-03
0.63 (0.48, 0.79)	0.0554	0.05	0.08	1.10E-01	1.0	5.00E-06
0.79 (0.64, 0.95)	0.0437	0.05	0.08	1.10E-01	1.0	2.49E-05
0.12 (0.00E+00, 0.28)	0.2014	0.05	0.08	1.10E-01	1.0	2.49E-05
0.76 (0.65, 0.87)	0.0977	0.05	0.02	4.08E-03	1.0	5.00E-06
0.52 (0.41, 0.64)	0.1438	0.05	0.01	4.08E-03	1.0	5.00E-06
0.96 (0.93, 0.99)	0.4022	0.05	0.93	2.60E-22	1.0	5.00E-06
0.96 (0.94, 0.99)	0.2316	0.05	1	2.60E-22	1.0	5.00E-06
0.86 (0.84, 0.89)	0.4694	0.05	0.85	6.70E-10	1.0	5.00E-06
0.96 (0.86, 1.0)	0.0868	0.05	0.18	6.70E-10	1.0	5.00E-06
0.87 (0.84, 0.89)	0.5146	0.05	0.58	2.00E-07	1.0	9.99E-06
0.93 (0.79, 1.0)	0.0598	0.05	0.53	2.00E-07	1.0	9.99E-06

of countries as well as their trading partners. Every country has protocols in place for handling *B. tabaci* (s.l.). Thus creating more species names for this group would force the generation of additional protocols and strategies for handling the "new" species and for trade compliance documentation and measures to be modified. There may also be flow on effects to chemical companies that manufacture and market pesticides for use against *B.tabaci* (s.l.). Here, there may be the additional need to confirm efficacy of their products against every species. While it is premature to recommend the description of additional species based on the analysis of a single locus,[105–107] in the case of *B. tabaci*, the literature surrounding the biology and taxonomy all point to the presence of multiple species (see review[63,73]) and this study adds further weight to the argument that *B. tabaci* is most likely a complex of numerous cryptic species and that any other explanation is less parsimonious.

Species delimitation of *L. dispar*

The mtCOI data is useful, based on standard PP for identifying *L. dispar dispar* (Clade 12-Fig. 4 and Table 3), but not for delineating the taxonomically recognized subspecies *L. dispar asiactica* and *L. dispar japonica*.[81] Similarly, utilizing the P ID(strict) and P ID(liberal) measures described in Ross et al[31] provides relatively high confidence, with 0.86 and 0.96 probability respectively, of correct identification if the unknown is either in this group or sister to it. In contrast, clades for the other two subspecies *L. dispar asiatica* and *L. dispar japonica* are not recovered with these more rigorous statistical tests (eg, Mr Bayes). This has implications for interpretation of past reports which provided no clade support[64] or rigorous forms of assessing branch support,[108] therefore possibly misleading users of the K2P distance and DNA barcoding for these subspecies. Here the lack of variation and resolution in the *L. dispar* phylogeny demonstrates the drawbacks of utilizing mtCOI alone for delimiting the three subspecies and other genes with more phylogenetic signal need to be used. On the other hand, there are other clades that exhibit substantial taxonomic distinctiveness (1 and 2 and 14, Table 3), and might form a better basis from which to guide future taxonomic revisions. Delimiting *L. dispar* s.l. within the genus is, however, accomplished with this gene region.[16,109] Therefore it remains a useful first approach in biosecurity for distinguishing it from other pest tussock moth species when immature life stages are intercepted, with

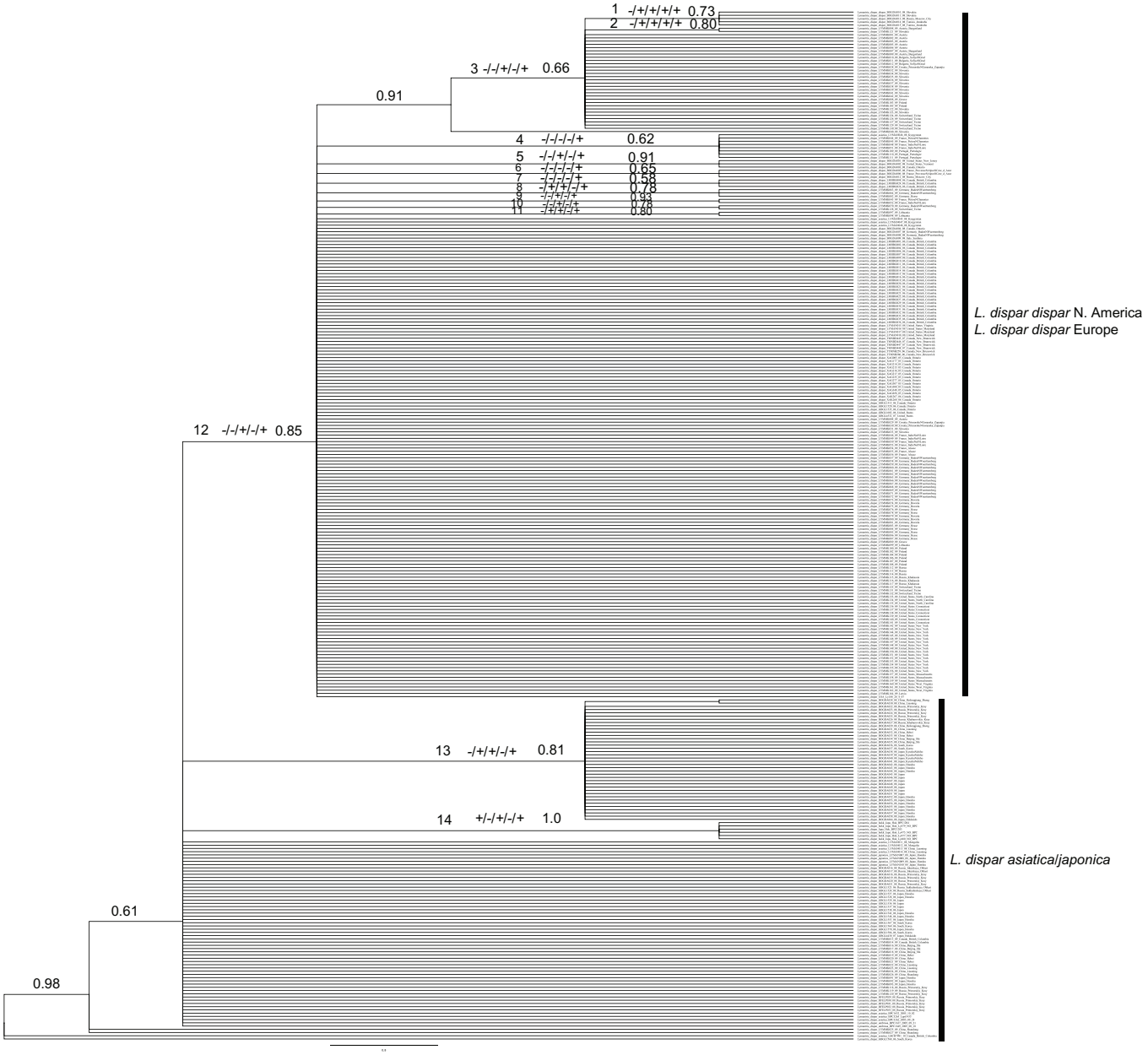

Figure 4. MrBayes phylogeny for *Lymantria dispar* generated using 8 million generations, trees sampled every 100 generations and 25% discarded as burnin.
Notes: Boxed individual and assigned numbers correspond to Table 3. The decoder for the "+" and "–" is as follows: K2P distance/P(RD)/posterior probability/P(AB)/*gsi*. An example of all measures indicating species distinctiveness would look like +/+/+/+/+ on the figure. Significance was determined by: >1% difference/>0.05/>0.70/ the Bonferroni correction described in the methods section.

no clues as to subsequent adult morphology, plant host preference or geographic origin.

Discordance in species distinctiveness measures

So where do you draw the "species" line on a phylogeny? This question has been addressed with multiple methods,[4,5,17,28,29,97,110–112] all with varying degrees of success, including four of the measures used in this study, P(AB), P(RD), *gsi* and GMYC. In reality, many systematists also rely heavily on either bootstrap support or posterior probability or both as clade support to delineate species. A direct comparison of all these methods with data for the two cases analysed here revealed no consistency between any combination of interspecies distance (based on the K2P), PP, Rosenberg's P(AB), Rodrigo's P(RD) and the *gsi* statistic and associated *P*-values (Tables 1–3).

Table 3. *L. dispar* species distinctiveness measures generated from the Geneious plugin[91] and the *gsi* software.[3]

Clade 1	Clade 2	Intra dist	K2P	Intra/ inter	P ID(Strict)	P ID(Liberal)	Av(MRCA-tips)	P(RD)	PP	P(AB)	*gsi*	*P*-value
1 (5)	2 (2)	0.188	0.534	0.35	0.70 (0.57, 0.82)	0.92 (0.82, 1.0)	0.094	0.05	0.73	1.20E-07	1	5.00E-06
2 (2)	1 (5)	0.192	0.534	0.36	0.41 (0.25, 0.56)	0.76 (0.61, 0.91)	0.096	0.05	0.8	8.30E-05	1	6.50E-05
3 (37)	8 (3)	0.263	0.758	0.35	0.89 (0.83, 0.94)	0.97 (0.94, 0.99)	0.2972	1	0.91	NAN	1	5.00E-06
4 (8)	8 (3)	0.195	0.539	0.36	0.77 (0.66, 0.87)	0.91 (0.85, 0.98)	0.0977	0.54	0.62	NAN	1	5.00E-06
5 (3)	8 (3)	0.183	0.538	0.34	0.56 (0.39, 0.74)	0.81 (0.67, 0.96)	0.0914	0.61	0.91	NAN	1	5.00E-06
6 (3)	8 (3)	0.221	0.565	0.39	0.53 (0.35, 0.71)	0.79 (0.64, 0.93)	0.1105	0.83	0.65	NAN	1	5.00E-06
7 (3)	8 (3)	0.211	0.544	0.39	0.53 (0.35, 0.71)	0.79 (0.64, 0.93)	0.1054	0.16	0.58	NAN	1	5.00E-06
8 (3)	9 (2)	0.183	0.531	0.35	0.56 (0.38, 0.74)	0.81 (0.67, 0.96)	0.0917	0.05	0.78	NAN	1	5.00E-06
9 (2)	8 (3)	0.183	0.531	0.34	0.41 (0.26, 0.57)	0.77 (0.62, 0.92)	0.0914	0.21	0.93	NAN	1	9.00E-05
10 (2)	8 (3)	0.216	0.549	0.39	0.39 (0.24, 0.54)	0.74 (0.59, 0.89)	0.1078	0.46	0.78	NAN	1	1.35E-04
11 (2)	8 (3)	0.21	0.54	0.39	0.39 (0.24, 0.55)	0.74 (0.59, 0.90)	0.105	0.05	0.8	NAN	1	9.50E-05
12 (212)	13 (35)	0.36	0.839	0.43	0.86 (0.81, 0.92)	0.96 (0.94, 0.99)	0.1917	0.09	0.85	NAN	1	5.00E-06
13 (35)	12 (212)	0.238	0.839	0.28	0.90 (0.85, 0.95)	0.97 (0.94, 1.00)	0.1193	0.05	0.81	NAN	0.9687868	5.00E-06
14 (6)	13 (35)	0.292	1.364	0.21	0.79 (0.66, 0.92)	0.95 (0.85, 1.0)	0.146	0.84	1	NAN	1	5.00E-06

Notes: Clade numbers refer to boxed individuals seen in Figure 4. See the legend for Table 2 describing the measures included.

Table 4. Summary of supported clades from previously suggested groupings and results presented here. *B. tabaci* were based on 4+ out of 5 statistical measures; *L. dispar* based on 3 out of 5 due to P(AB) being unable to be calculated for the majority of clades due to the unresolved phylogeny.

Before	Revised
Bemisia tabaci	
(Dinsdale et al[70])	(This study)
SubSaharan Africa 1**	SubSaharan Africa 1/2
SubSaharan Africa 2**	Clade 2-27 (4+)
	Clade 2-39 (4+)
SubSaharan Africa 3	Untested as only 1 haplotype
SubSaharan Africa 4	SubSaharan Africa 4
New World	New World
India Ocean	Indian Ocean
Mediterranean**	Mediterranean
	(Clades 9-56 + 9-57)
	Clade 9-39 (4+)
	Clade 9-54 (4+)
Middle East—Asia Minor 2	Untested as only 1 haplotype
Middle East—Asia Minor 1**	Middle East—Asia Minor 1
	Clade 8-6 (4+)
Italy	Italy
Asia I	Asia I
Asia II-1	Asia II-1
Asia II-2	Untested as only 1 haplotype
Asia II-3	Asia II 3
Asia II-4	Untested as only 1 haplotype
Asia II-5	Asia II-5
Asia II-6	Asia II-6
Asia II-7	Asia II-7
Asia II-8	Asia II-8
Ex China EU 192051	Untested as only 1 haplotype
China 1	China 1
China 2	Untested as only 1 haplotype
Australia/Indonesia	Australia/Indonesia
Australia	Untested as only 1 haplotype
Lymantria dispar	
(De Waard et al[64])	(This study)
L. d. dispar (N. America & France)**	Clade 1 (4+)
L. d. dispar (Europe)**	Clade 2 (4+)
	Clade 8 (3+)
	Clade 11 (3+)
*L. d. asiatica/japonica***	Clade 13 (3+)
	Clade 14 (3+)

Note: **Changes between studies.

To understand this, it is important to consider the assumptions and different questions being asked of these measures. Clade support (PP) is data-driven and is a measure of how strongly the data support the particular clade. In contrast, the P(AB) and *gsi*

measures are dependent on the estimated tree topology and on the data only through the estimated tree. More specifically, the null hypothesis for both P(AB) and P(RD) is based on panmixis. As a test for cryptic species identification or species distinctiveness, they are based on the coalescent[113] and can be applied to genetic data from one locus. P(RD) is defined as the probability of an observed degree of distinctiveness. The first step is calculating *M*, which is the sum of the intervals spanning the node to the tips (the "species defining node") and the sum of lengths of the intervals between the node and the root. Here a *P*-value for this measure of less than 0.05 indicates that the focal group has branching significantly different to what would be expected under the coalescent process, ie, the lineage is not conforming to the Wright-Fisher model and therefore a cryptic species is present. For *B. tabaci* the majority of clades tested for distinctiveness had P(RD) *P*-values lower than 0.05, indicating the rejection of panmixis and the possibility of mating restrictions indicative of separate species. For *L. dispar*, there were also a few clades with a P(RD) *P*-value of less than 0.05 indicating the possibility that cryptic species are present, although these did not correlate with the taxonomically defined subspecies units. However, generally, the measure P(RD) was overly sensitive to detecting taxonomic distinctiveness for our data compared to the other measures. This over-estimation of statistical significance, which had already been predicted by Rodrigo et al,[2] misleadingly indicates the presence of cryptic species due to the fact that only one of many possible coalescent models is utilized in the calculations. Rodrigo et al[2] suggest an *a posteriori* correction of the level of significance is needed and an uncorrected *P*-value will be too liberal, and this is exactly what is seen in Tables 1–3.

In contrast, the *gsi* can track divergence before complete monophyly and is therefore good for looking at distinctiveness amongst "young" species.[3] The *gsi* is an estimate of the degree of exclusive ancestry of individuals in predefined groups on rooted trees. Both of the datasets included in this study had large sample sizes lending to the utility of the *gsi* and rejection of the null hypothesis indicating species distinctiveness. The *gsi* was the only measure in this study that supported distinctiveness of all of the clades in the *L. dispar* phylogeny. In contrast, these recent divergences were not detected with the more conservative measure of P(AB), where only two clades were significant, these included samples from Slovakia, Russia, Tunisia, and Austria, and P(RD), where five clades were significant. Of these, the most conservative approach to assessing species distinctiveness is the reciprocal monophyly described by P(AB) in Rosenberg.[1] Here the null hypothesis is that monophyly is a chance outcome of random branching and can be rejected at $P < 10^{-5}$. The *B. tabaci* dataset presented here clearly has enough variation supporting the recognition of several groups being reciprocally monophyletic and therefore distinct compared to what has been previously proposed (Table 4), whereas the *L. dispar* dataset had only two groups with a significant P(AB), clades 1 and 2 (Table 3).

One further set of assumptions that need to be considered in cases such as these relate to the idiosyncrasies of mitochondrial DNA evolution. Characteristics such as its susceptibility to selective sweeps, hybrid introgression and ancestral polymorphism contribute to the questionable use of mtDNA as an ideal genetic marker.[114] Selective sweeps effectively reduce intraspecific variation, thereby making the 'gap' between intra- and inter-specific diversity more pronounced, enhancing the appearance of discrete taxa. Bacterial endosymbiont-induced selective sweeps have been suggested as a mechanism for speciation in the *B. tabaci*[115] this hypothesis seems less plausible as available evidence shows that most studies exploring mating between different members of the complex are unable to copulate[77] making it highly unlikely that selective sweeps are involved in *B. tabaci* evolution. Conversely, homogenization through hybrid introgression and ancestral polymorphism effectively reduces mtDNA inter-specific variation resulting in an apparent lack of distinction among otherwise biologically distinct sub-species groups, possibly in the case here of *L. dispar*. Introgression has been reported for other Lepidoptera (examples given in[116]), however, as it is the female Lepidoptera that are the heterogametic sex and would exhibit reduced viability as hybrids, and assuming mtDNA to be largely maternally inherited, introgression may not be as important as shared ancestral polymorphisms producing the same effect. Each of these senario's could be elucidated by comparison to data from multiple nuclear markers (eg,[117]) and emphasises the need ultimately for a mulitlocus, integrated taxonomic approach. Barcoding combined

with the 'tip-to-root' analytical approach does have the potential to reveal evolutionary units and potential new species at a faster rate and can be followed up by taxonomists. Therefore, incorporating the approaches developed here into an integrative workflow, such as that proposed by Padial et al[104] for assessing "candidate species", is ideally needed to underpin subsequent use of such datasets in a biosecurity setting.

So to answer the question of where to draw the line, perhaps, where five of the six taxonomic distinctiveness measures are significant the nominated clades could be the focus of further studies to verify their "species" status. For the test cases here, a guide based on comparison of previously proposed taxonomic groups with the tip to root approach described in this study is presented in Table 4.

Practical suggestions

There is an increasing body of literature that supports the need to include multiple genes to generate a species tree. However, the reality for biosecurity applications is that for the near future at least, DNA barcode-type approaches based on use of a single gene are becoming a practical option for distinguishing species of biosecurity concern. Although less vulnerable to blind false negative and false positive results than alternative species 'specific' PCR and RFLP-based methods of identification, it is equally susceptible to incomplete taxonomic knowledge and/or gene regions that poorly reflect the species delimitations. There is clearly a need to bridge this gap by taking advantage of the recent advances in species tree estimation[38,39,41,118] from multiple gene trees. This would serve the main goals of NPPOs (National Plant Protection Organisations) to protect the commodities within their borders from unknown non-native invasive pests[119] much more effectively. Unfortunately, as Figure 5 outlines, the decision making process would require diagnosticians to be skilled in the current taxonomy of individual groups, species concepts, DNA barcoding, phylogenetic methods, gene tree versus species tree, etc.; a daunting prospect and an impractical resource for all potential species threats. Also, the time to collect data from multiple genes, carry out an analysis (eg, BEST[41] can take months for the chains to converge) and then perform the additional step of assessing species delimitation measures is inappropriately time consuming; relying on availability and

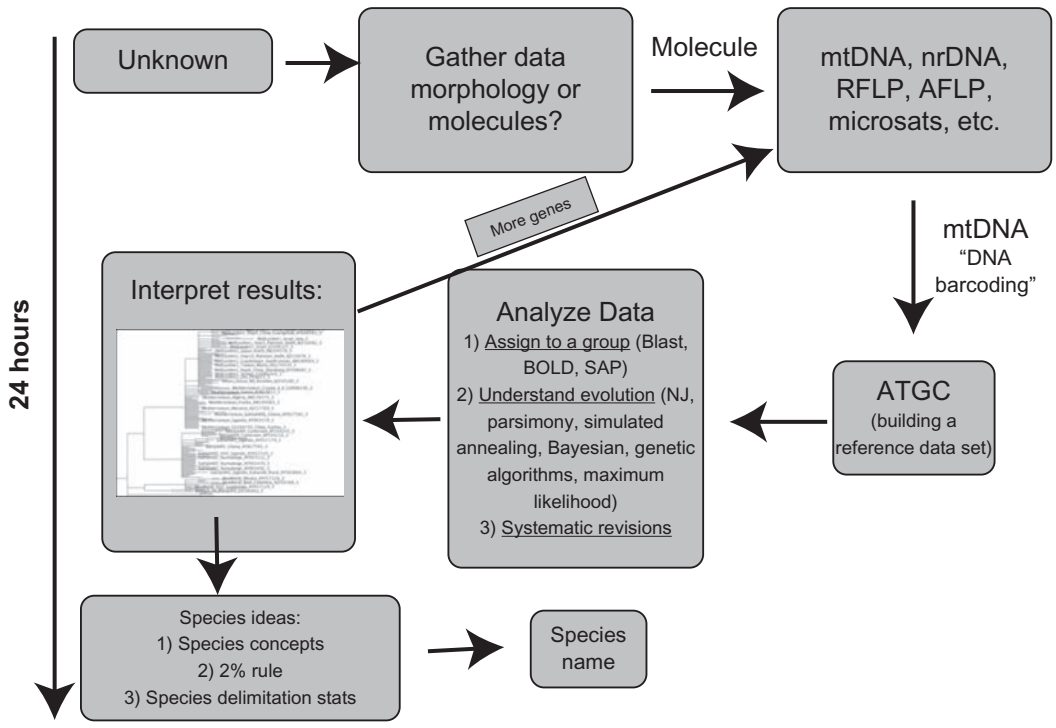

Figure 5. Outline of the decision making process for identifying an unknown intercept at the border illustrating the incompatibility between undertaking rigorous species delimitation process and requiring species identification within 24 hours.

use of multiple genes is simply not a feasible option in most biosecurity circumstances.

Where do we go from here? The six species distinctiveness measures described here are all dependent on the phylogeny. Having a reliable robust phylogeny that has been rigorously tested using all phylogenetic techniques available is ideally the first step in testing for species distinctiveness and ultimately identification. In cases of significant importance to biosecurity, such as cryptic species where the taxonomy is not clear, relying solely on clade support may be severely misleading as to the taxa that are actually distinct. The mtCOI region is a viable option for accurately identifying *B. tabaci* s.l., though clearly the taxonomy has to catch-up with the genetic diversity that exists within this "species". Conversely, in cases like *L. dispar* where the taxonomy based on morphology and behavioral characters appears to be describing more diversity than is apparent from the mtCOI region, additional gene regions need to be analyzed to confirm if multiple species are present or not. It is important that users and developers of the methodology for estimating both gene tree and species tree realize the impacts that such methods could have on global trade. Similarly, the regulators need to recognize that utilizing one gene may not provide the power of resolution needed to robustly delimit species. Future directions for biosecurity focused research might include scrutinizing similar phylogenies of other pests besides *B. tabaci* and *L. dispar* that are listed amongst 100 of the most invasive species (http://www.issg.org/database/species/search.asp?st=100ss). Where DNA barcoding fails to provide useful information for species level identification these could easily, through implementation of the tools in Geneious, be targeted for multiple gene analyses and more sophisticated species tree generation and analysis tools.

While DNA methods of delimitation are more amenable to quantitative analysis than morphological data, the defensibility of decisions based on sequence-similarity of an unknown within a gene tree is very much dependent on how well the species are delimited by that data in the first place. Ideally, a framework of statistically supported clades (as we have described here) against which assignments of unknown specimens could be more confidently made would assist regulatory agencies in decision making. In terms of species delimitation, it is recommended to consider several species delimitation measures before making decisions on species boundaries. In our case we implemented the strategy of five out of six (*B. tabaci*) and four out of six (*L. dispar)* of the measures being significant for taxonomic distinctiveness analyses to question current descriptions. Keeping in mind that the order of which the measures detect taxonomic distinctiveness is as follow P(RD) > GMYC > *gsi* > P(AB) > PP > K2P (Table 2), meaning that P(RD) was the most liberal measure and K2P interspecies distance the most conservative. Solely relying on the K2P interspecies distance or PP is going to underestimate taxonomic distinctiveness. Once species delimitation is investigated (as described above) assignment of an unknown can proceed. It is recommended by Ross et al,[31] based on a simulation study, that for correct identification of an unknown where the species has not been included in the reference data set, relying on the strict P(ID) described above coupled with a distance threshold decreases the chance for false positive identification. Most of this (species delimitation and identification) is easily implemented in a user friendly Geneious plugin,[91] *gsi* and GMYC being the exceptions.

Acknowledgements

Vladimir Mencl at the Bluefern supercomputer facilities at the University of Canterbury, New Zealand was instrumental in implementing the *gsi* software on the cluster. Rupert Collins (Lincoln University, New Zealand) provided assistance with the GMYC analyses. Mikael Hirsch (Australian Government Department of Agriculture, Fisheries and Forestry (DAFF)) advised on the myrtle/guava rust example. Ron DeBry (University of Cincinnati) for useful comments on earlier drafts of the manuscript. LMB and KF were funded by the Tertiary Education Council of New Zealand, L.K. by the U.S. National Science Foundation grant DEB-0842219.

Highlights

- A novel approach ("tip to root") to delineate species objectively is described herein; this approach can be effectively applied to any phylogeny.
- K2P genetic distances alone do not identify all taxonomic distinctiveness present in a given phylogeny;

other measures such as P(AB), P(RD), *gsi* and GMYC are more useful in identifying taxonomic distinctiveness.

- A consensus of such analyses statistically supported five more distinct clades within the *Bemisia tabaci* mtDNA phylogeny than are taxonomically described to date, while delineation of *Lymantria dispar* subspecies remains problematic due to lack of phylogenetic resolution.
- As part of the regulation of trade under international biosecurity arrangements, the capacity to delimit species using DNA data, will have a direct bearing on whether trade takes place and will represent a significant departure from the current processes which are mostly reliant of morphological separation.

Disclosures

Author(s) have provided signed confirmations to the publisher of their compliance with all applicable legal and ethical obligations in respect to declaration of conflicts of interest, funding, authorship and contributorship, and compliance with ethical requirements in respect to treatment of human and animal test subjects. If this article contains identifiable human subject(s) author(s) were required to supply signed patient consent prior to publication. Author(s) have confirmed that the published article is unique and not under consideration nor published by any other publication and that they have consent to reproduce any copyrighted material. The peer reviewers declared no conflicts of interest.

References

1. Rosenberg NA. Statistical tests for taxonomic distinctiveness from observations of monophyly. *Evolution*. 2007;61(2):317–23.
2. Rodrigo A, Bertels F, Heled J, Noder R, Shearman H, Tsai P. The perils of plenty: what are we going to do with all these genes? *Philos Trans R Soc Lond B Biol Sci*. 2008;363(1512):3893–902.
3. Cummings MP, Neel MC, Shaw KL. A genealogical approach to quantifying lineage divergence. *Evolution*. 2008;62(9):2411–22.
4. Pons J, Barraclough TG, Gomez-Zurita J, et al. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Syst Biol*. 2006; 55(4):595–609.
5. Monaghan MT, Wild R, Elliot M, et al. Accelerated species inventory on Madagascar using coalescent-based models of species delineation. *Syst Biol*. 2009;58(3):298–311.
6. Mayr E. Systematics and the origin of species. Columbia: University Press; 1942.
7. De Queiroz K. Ernst Mayr and the modern concept of species. *Proc Natl Acad Sci U S A*. 2005;102 Suppl 1:6600–7.
8. Hey J. On the failure of modern species concepts. *Trends Ecol Evol*. 2006;21(8):447–50.
9. De Queiroz K. Species concepts and species delimitation. *Syst Biol*. 2007; 56(6):879–86.
10. Wiens JJ, Servedio MR. Species delimitation in systematics: inferring diagnostic differences between species. *Proc Biol Sci*. 2000;267(1444): 631–6.
11. De Queiroz K. The general lineage concept of species, species criteria and the process of speciation. In: Howard DJ, Berlocher SH, editors. *Endless forms Species and Speciation*. New York: Oxford University Press; 1998:57–75.
12. Frost DR, Kluge AR. A consideration of epistemology in systematic biology, with special reference to species. *Cladistics*. 1994;10:259–94.
13. Nixon KC, Wheeler WC. An amplification of the phylogenetic species concept. *Cladistics*. 1990;6:211–23.
14. Sites JW, Marshall JC. Delimiting species: a Renaissance issue in systematic biology. *Trends Ecol Evol*. 2003;18(9):462–70.
15. Sites JW, Marshall JC. Operational criteria for delimiting species. *Annu Rev Ecol Evol Syst*. 2004;35:199–227.
16. Armstrong KF, Ball SL. DNA barcodes for biosecurity: invasive species identification. *Philos Trans R Soc Lond B Biol Sci*. 2005;360(1462):1813–23.
17. Hebert PD, Cywinska A, Ball SL, deWaard JR. Biological identifications through DNA barcodes. *Proc Biol Sci*. 2003;270(1512):313–21.
18. Floyd R, Lima J, deWaard JR, Humble LM, Hanner RH. Common goals: policy implications of DNA barcoding as a protocol for identification of arthropod pests. *Biol Invasions*. 2010;12(9):2947–54.
19. Jérôme M, Lemaire C, Verrez-Bagnis V, Etienne M. Direct sequencing method for species identification of canned sardine and sardine-type products. *J Agric Food Chem*. 2003;51(25):7326–32.
20. Sullivan JP, Lavoue S, Hopkins CD. Discovery and phylogenetic analysis of a riverine species flock of African electric fishes (Mormyridae: Teleostei). *Evolution*. 2002;56(3):597–616.
21. Abebe E, Mekete T, Thomas WK. A critique of current methods in nematode taxonomy. *Afr J Biotechnol*. 2011;10(3):312–23.
22. Floyd R, Abebe E, Papert A, Blaxter M. Molecular barcodes for soil nematode identification. *Mol Ecol*. 2002;11(4):839–50.
23. Goker M, Garcia-Blazquez G, Voglmayr H, Telleria MT, Martin MP. Molecular taxonomy of phytopathogenic fungi: a case study in Peronospora. *PLoS One*. 2009;4(7):e6319.
24. Grube M, Kroken S. Molecular approaches and the concept of species and species complexes in lichenized fungi. *Mycol Res*. 2000;104(11): 1284–94.
25. Barraclough TG, Hughes M, Ashford-Hodges N, Fujisawa T. Inferring evolutionarily significant units of bacterial diversity from broad environmental surveys of single-locus data. *Biol Lett*. 2009;5(3):425–8.
26. Janda JM, Abbott SL. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *J Clin Microbiol*. 2007;45(9):2761–4.
27. Carstens BC, Dewey TA. Species delimitation using a combined coalescent and information-theoretic approach: an example from North American Myotis bats. *Syst Biol*. 2010;59(4):400–14.
28. O'Meara BC. New heuristic methods for joint species delimitation and species tree inference. *Syst Biol*. 2010;59(1):59–73.
29. Yang Z, Rannala B. Bayesian species delimitation using multilocus sequence data. *Proc Natl Acad Sci U S A*. 2010;107(20):9264–9.
30. Goker M, Grimm GW, Auch AF, Aurahs R, Kucera M. A Clustering Optimization Strategy for Molecular Taxonomy Applied to Planktonic Foraminifera SSU rDNA. *Evol Bioinformatics*. 2010;6:97–112.
31. Ross HA, Murugan S, Li WL. Testing the reliability of genetic methods of species identification via simulation. *Syst Biol*. 2008;57(2):216–30.
32. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol*. 1980;16(2):111–20.
33. Cywinska A, Hunter FF, Hebert PD. Identifying Canadian mosquito species through DNA barcodes. *Med Vet Entomol*. 2006;20(4):413–24.
34. Burns JM, Janzen DH, Hajibabaei M, Hallwachs W, Hebert PD. Dna barcodes of closely related (but morpholigcally and ecologically distinct) species of skipper butterflies (Hesperiidae) can differ by only one to three nucleotides. *J of Lep Soc*. 2007;61(3):138–53.

35. Meyer CP, Paulay G. DNA barcoding: error rates based on comprehensive sampling. *PLoS Biol*. 2005;3(12):e422.
36. Nielsen R, Matz M. Statistical approaches for DNA barcoding. *Syst Biol*. 2006;55(1):162–9.
37. Knowles LL, Kubatko LS, editors. Estimating Species Trees. Practical and Theoretical Aspects: Wiley-Blackwell; 2010.
38. Kubatko LS. Identifying hybridization events in the presence of coalescence via model selection. *Syst Biol*. 2009;58(5):478–88.
39. Kubatko LS, Carstens BC, Knowles LL. STEM: species tree estimation using maximum likelihood for gene trees under coalescence. *Bioinformatics*. 2009;25(7):971–3.
40. Huang H, He Q, Kubatko LS, Knowles LL. Sources of error inherent in species-tree estimation: impact of mutational and coalescent effects on accuracy and implications for choosing among different methods. *Syst Biol*. 2010;59(5):573–83.
41. Liu L. BEST: Bayesian estimation of species trees under the coalescent model. *Bioinformatics*. 2008;24(21):2542–3.
42. Liu L, Yu L, Pearl DK, Edwards SV. Estimating species phylogenies using coalescence times among sequences. *Syst Biol*. 2009;58(5):468–77.
43. Ane C, Larget B, Baum DA, Smith SD, Rokas A. Bayesian estimation of concordance among gene trees. *Mol Biol Evol*. 2007;24(2):412–26.
44. Than C, Nakhleh L. Species tree inference by minimizing deep coalescences. *PLoS Comput Biol*. 2009;5(9):e1000501.
45. Than C, Ruths D, Nakhleh L. PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinformatics*. 2008;9:322.
46. Than C, Sugino R, Innan H, Nakhleh L. Efficient inference of bacterial strain trees from genome-scale multilocus data. *Bioinformatics*. 2008;24(13): i123–31.
47. Funk DJ, Omland KE. Species-level paraphyly and polyphyly frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Ann Rev Ecol Evol Syst*. 2003;34:397–423.
48. Knowles LL, Carstens BC. Delimiting species without monophyletic gene trees. *Syst Biol*. 2007;56(6):887–95.
49. Belfiore NM, Liu L, Moritz C. Multilocus phylogenetics of a rapid radiation in the genus Thomomys (Rodentia: Geomyidae). *Syst Biol*. 2008;57(2): 294–310.
50. FAO. Food and Agriculture Organization of the United Nations. Secretariat of the International Plant Protection Convention, Rome; 2007:341–52.
51. Goldson SL. Biosecurity, risk and policy: a New Zealand perspective. *J of Consum Protect and Food Safety*. 2011;6:41–7.
52. FAO. Food and Agriculture Organization of the United Nations. International standards for phytosanitary measures. In: *Diagnostic Protocols for Regulated Pests*. 2006:27.
53. Rossman AY. The impact of invasive fungi on agricultural ecosystems in the United States. In: Langor DW, Sweeney J, editors. *Ecological impacts of Non-Native Invertebrates and Fungi on Terrestrial Ecosystems*. U.S. Government; 2008:97–107.
54. Carnegie AJ, Lidbetter JR, Walker J, et al. *Uredo rangelii*, a taxon in the guava rust complex, newly recorded on Myrtaceae in Australia. *Australas Plant Path*. 2010;39:463–6.
55. Simpson JA, Thomas K, Grgurinovic CA. Uredinales species pathogenic on species of Myrtaceae. *Australas Plant Path*. 2006;35:549–62.
56. Glen M, Alfenas AC, Zauza V, Wingfield MJ, Mohammed C. Puccinia psidii: a threat to the Australian environment and economy—a review. *Australas Plant Path*. 2007;36:1–16.
57. Ramsfield T, Dick M, Bulman L, Ganley R. Briefing document on myrtle rust, a member of the guava rust complex, and the risk to New Zealand. In: *Scion*, *Next Generation Biomaterials*. www.scionresearch.com; 2010.
58. Van der Merwe MM, Walker J, Ericson L, Burdon JJ. Coevolution with higher taxonomic host groups within the *Puccinia*/*Uromyces* rust lineage obscured by host jumps. *Mycol Res*. 2008;112:1387–408.
59. Smith-Caldas MRB, McPheron BA, Silva JG, Zucchi R. Phylogenetic relationships among species of the *fracterculus* group (*Anastrepha*: Diptera: Tephritidae) inferred from DNA sequences of mitochondrial cytochrome oxidase I. *Neotrop Entomol*. 2001;30(4):565–73.
60. Virgilio M, Backeljau T, Barr N, Meyer MD. Molecular evaluation of nominal species in the *Ceratitis fasciventris*, *C. anonae*, *C. rosa* complex (Diptera: Tephritidae). *Mol Phylogenet Evol*. 2008;48(1):270–80.
61. Clarke AR, Armstrong KF, Carmichael AE, et al. Invasive phytophagous pests arising through a recent tropical evolutionary radiation: the *Bactrocera dorsalis* complex of fruit flies. *Annu Rev Entomol*. 2005;50: 293–319.
62. Cantrell BK, Chadwick B, Cahill A. Fruit fly fighters: eriadication of papaya fruit fly. In.: CSIRO Publishing, Collingwood, Victoria, Australia; 2002.
63. De Barro PJ, Liu SS, Boykin LM, Dinsdale AB. Bemisia tabaci: A Statement of Species Status. *Ann Rev Entomol*. 2011;56:1–19.
64. de Waard JR, Mitchell A, Keena MA, et al. Towards a global barcode library for Lymantria (Lepidoptera: Lymantriinae) tussock moths of biosecurity concert. *PLoS Biol*. 2010;5(12):e14280.
65. Pimental D, Zuniga R, Morrison D. Update of the environmental and economics costs associated with alien-invasive species in the United States. *Ecol Econ*. 2005;52:273–88.
66. Byrne DN, Bellows TS Jr. Whitefly biology. *Ann Rev Entomol*. 1991;36: 431–57.
67. Oliveira MRV, Henneberry TJ, Anderson PK. History, current status, and collaborative research projects for *Bemisia tabaci*. *Crop Protection*. 2001; 20:709–23.
68. Drogue S, Gozlan E. Are SPS-TBT regulations in the European union discriminating agricultural trade from Africa? The case of live plants and cut flowers. *Int J Business Res*. 2009;9(3):70–91.
69. Boykin LM, Shatters RG Jr, Rosell RC, et al. Global relationships of *Bemisia tabaci* (Hemiptera: Aleyrodidae) revealed using Bayesian analysis of mitochondrial COI DNA sequences. *Mol Phylogenet Evol*. 2007;44(3): 1306–19.
70. Dinsdale A, Cook L, Riginos C, Buckley YM, De Barro P. Refined global analysis of *Bemisia tabaci* (Gennadius) (Hemiptera: Sternorrhyncha: Aleyroidea) mitochondrial CO1 to identify species level genetic boundries. *Ann Entomol Soc Am*. 2010;103(2):196–208.
71. Hu J, De Barro P, Zhao H, Wang J, Nardi F, Liu SS. An extensive field survey combined with a phylogenetic analysis reveals rapid and widespread invasion of two alien whiteflies in China. *PLoS One*. 2011;6(1):e16061.
72. Elbaz M, Lahav N, Morin S. Evidence for pre-zygotic reproductive barrier between the B and Q biotypes of *Bemisia tabaci* (Hemiptera: Aleyrodidae). *Bull Entomol Res*. 2010;100(5):581–90.
73. Liu SS, Colvin J, De Barro P. Species concepts as applied to the whitefly *Bemisia tabaci* systematics: how many species are there? *Agricul Sci in China*. 2011;10.
74. Sun DB, Xu J, Luan JB, Liu SS. Reproductive incompatibility between the B and Q biotypes of the whitefly *Bemisia tabaci* in China: genetic and behavioural evidence. *Bull Entomol Res*. 2010:1–10.
75. Wang P, Sun DB, Qiu BL, Lui SS. The presence of six cryptic species of the whitefly *Bemisia tabaci* complex in China as revealed by crossing experiments. *Insect Sci*. 2011:18(67–77).
76. Wang Z, Yan H, Yang Y, Wu Y. Biotype and insecticide resistance status of the whitefly Bemisia tabaci from China. *Pest Manag Sci*. 2010;66(12): 1360–6.
77. Xu J, De Barro PJ, Liu SS. Reproductive incompatibility among genetic groups of *Bemisia tabaci* supports the proposition that the whitefly is a cryptic species complex. *Bull Entomol Res*. 2010;100(3):359–66.
78. Liebhold AM, Elkinton JS. Gypsy moth dynamics. *Trends Ecol Evol*. 1991; 6(8):263–4.
79. Liebhold AM, Tobin PC. Population ecology of insect invasions and their management. *Annu Rev Entomol*. 2008;53:387–408.
80. Schintlmeister A. The Taxonomy of the genus *Lymantria* HÜBNER, [1819] (Lepidoptera: Lymantriidae). *Quadrifina*. 2004;Band 7:1–248.
81. Pogue MG, Schaefer PW. A review of selected species of *Lymantria* (Hubner [1819]) (Lepidoptera: Noctuidae: Lymantriinae) from subtropical and temperate regions of Asia including the description of three new speces, some potentially invasive to North America. Publication FHTET-2006-7, United States Department of Agriculture, Forest Service. Fort Collins, Colorado: Forest Health Technology Enterprise Team; 2007.

82. Baranchikov YN. Ecological basis of the evolution of host relationships in Eurasian gypsy moth populations. In: Wallner WE, McManus KA, editors. *Proceedings*, *Lymantriidae: A Comparison of Features of New and Old World Tussock Moths*. U.S. Department of Agriculture; 1989.
83. Koshio C. Pre-ovipositional behaviour of female gypsy moth, *Lymantria dispar* L. (Lepidoptera, Lymantriidae). *Appl Entomol Z*. 1996;31:1–10.
84. Schaefer PW, Weseloh RM, Sun X, Wallner WE, Yan J. Gypsy moth *Lymantria* (=*Ocneria*) *dispar* L. (Lepidoptera: Lymantriidae), in the People's Republic of China. *Env Entomol*. 1984;13:1535–41.
85. Keena MA, Cote MJ, Grinberg PS, Wallner WE. World distribution of female flight and genetic variation in *Lymantria dispar* (Lepidoptera: Lymantriidae). *Environ Entomol*. 2008;37(3):636–49.
86. Keena MA, Grinberg PS, Wallner WE. Inheritance of female flight in Lymantria dispar (Lepidoptera: Lymantriidae). *Environ Entomol*. 2007; 36(2):484–94.
87. Carde RT, Charlton RE, Wallner WE, Baranchikov YN. Pheromone-mediated diel activity rhythms of male Asian gypsy moths (Lepidoptera: Lymantriidae) in relations to female eclosion and temperature. *Ann Entomol Soc Am*. 1996; 89:745–53.
88. Wallner WE, Humble LM, Levin RE, Baranchikov YN, Carde RT. Response of adult lymantriid moths to illuminiation devices in the Russian Far East. *J Econ Entomol*. 1995;88:337–42.
89. Huelsenbeck JP, Ronquist F. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 2003;19:1572–4.
90. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Stat Sci*. 1992;7:457–511.
91. Masters BC, Fan V, Ross HA. Species delimitation—a geneious plugin for the exploration of species boundaries. *Mol Ecol Resourc*. 2010;11(1):154–7.
92. Drummond AJ, Ashton B, Buxton S, et al. Geneious v5.1, Available from: http://www.geneious.com; 2010.
93. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol*. 2007;7:214.
94. Rambaut A, Drummond AJ. Tracer v1.5. [Online]. Available from: http://beastbioedacuk/Tracer; 2007.
95. Drummond AJ, Rambaut A, Suchard M. LogCombiner v1.5.4 [Online]. Available from: http://beastbioedacuk/LogCombiner; 2010.
96. Drummond AJ, Rambaut A, Suchard M. TreeAnnotator v1.5.4 [Online]. Available from: http://beastbioedacuk/TreeAnnotator; 2010.
97. Abdo Z, Golding GB. A step toward barcoding life: a model-based, decision-theoretic method to assign genes to preexisting species groups. *Syst Biol*. 2007;56(1):44–56.
98. Teletchea F. After 7 years and 1000 citations: comparative assessment of the DNA barcoding and the DNA taxonomy proposals for taxonomists and non-taxonomists. *Mitochondrial DNA*. 2010;21(6):206–26.
99. Armstrong KF. DNA barcoding for plant biosecurity: a new module in New Zealand's diagnostic toolbox. *EPPO Bulletin*. 2010;40(1):91–100.
100. Prabhaker N, Courdriet DL, Toscano NC. Effect of synergists on organophosphate and permethrin resistance in sweetpotato whitefly (Homoptera: Aleyrodidae). *J Econ Entomol*. 1998:81(34–9).
101. De Barro P, Ahmed M. Genetic Networking of the *Bemisia tabaci* Cryptic Species Complex Reveals Pattern of Biological Invasions. *PLoS Biol*. 2011;6(10):e25579.
102. Cognato AI. Standard percent DNA sequence difference for insects does not predict species boundaries. *J Econ Entomol*. 2006;99(4):1037–45.
103. Dasmahapatra KK, Elias M, Hill RI, Hoffman JI, Mallet J. Mitochondrial DNA barcoding detects some species that are real, and some that are not. *Mol Ecol Resour*. 2010;10(2):264–73.
104. Padial JM, Miralles A, De la Riva I, Vences M. The integrative future of taxonomy. *Front Zool*. 2010;7:16.
105. Hudson RR, Coyne JA. Mathematical consequences of the genealogical species concept. *Evolution*. 2002;56(8):1557–65.
106. Linnen CR, Farrell BD. Mitonuclear discordance is caused by rampant mitochondrial introgression in Neodiprion (Hymenoptera: Diprionidae) sawflies. *Evolution*. 2007;61(6):1417–38.
107. Shaw KL. Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: what mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. *Proc Natl Acad Sci U S A*. 2002; 99(25):16122–7.
108. Bogdanowicz SM, Schaefer PW, Harrison RG. Mitochondrial DNA variation among worldwide populations of gypsy moths, *Lymantria dispar*. *Mol Phylogenet Evol*. 2000;15(3):487–95.
109. Ball SL, Armstrong KF. DNA barcodes for insect pest identification: a test case with tussock moths (Lepidoptera: Lymantriidae). *Can J Forest Res*. 2006;36:337–50.
110. Avise JC, Walker D. Species realities and numbers in sexual vertebrates: perspectives from an asexually transmitted genome. *Proc Natl Acad Sci U S A*. 1999;96(3):992–5.
111. Ence DD, Carstens BC. SpedeSTEM: a rapid and accurate method for species delimitation. *Mol Ecol Resourc*. 2010:1–8. DOI: 10.1111/j.1755-0998.2010.02947.x.
112. Kubatko LS, Gibbs HL, Bloomquist EW. Inferring Species-Level Phylogenies and Taxonomic Distinctiveness Using Multilocus Data In Sistrurus Rattlesnakes. *Syst Biol*. 2011;60(4):393–409.
113. Kingman JFC. The coalescent. *Stochastic Processes and Their Applications*. 1982;13:235–48.
114. Galtier N, Nabholz B, Glemin S, Hurst GD. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Mol Ecol*. 2009;18(22): 4541–50.
115. Gueguen G, Vavre F, Gnankine O, et al. Endosymbiont metacommunities, mtDNA diversity and the evolution of the Bemisia tabaci (Hemiptera: Aleyrodidae) species complex. *Mol Ecol*. 2010;(19):4365–76.
116. Forister ML, Nice CC, Fordyce JA, Compert Z, Shapiro AM. Considering evolutionary processes in the use of single-locus genetic data for conservation, with examples from the Lepidoptera. *J of Insect Conserv*. 2008;12:37–51.
117. Gompert Z, Nice CC, Fordyce JA, Forister ML, Shapiro AM. Identifying units for conservation using molecular systematics: the cautionary tale of the Karner blue butterfly. *Mol Ecol*. 2006;15(7):1759–68.
118. Heled J, Drummond AJ. Bayesian inference of species trees from multilocus data. *Mol Biol Evol*. 2010;27(3):570–80.
119. Waage JK, Mumford JD. Agricultural biosecurity. *Philos Trans R Soc Lond B Biol Sci*. 2008;363(1492):863–76.

Supplemental Data

Bemisia tabaci clades extracted from Figure 2. Boxes indicate groups of species that were tested using the species distinctiveness measures described in the text and the results are shown in Table 2.

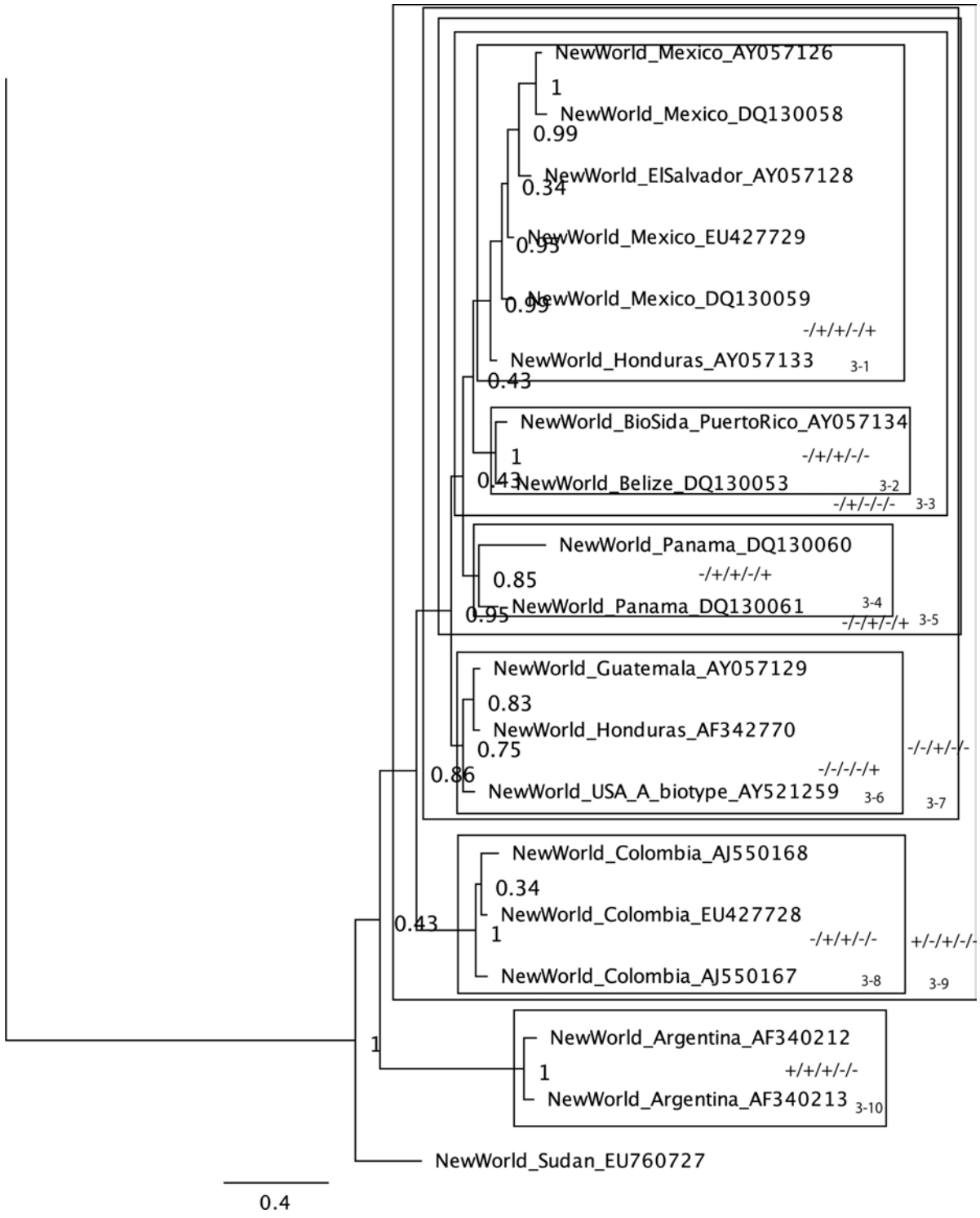

Figure S1.

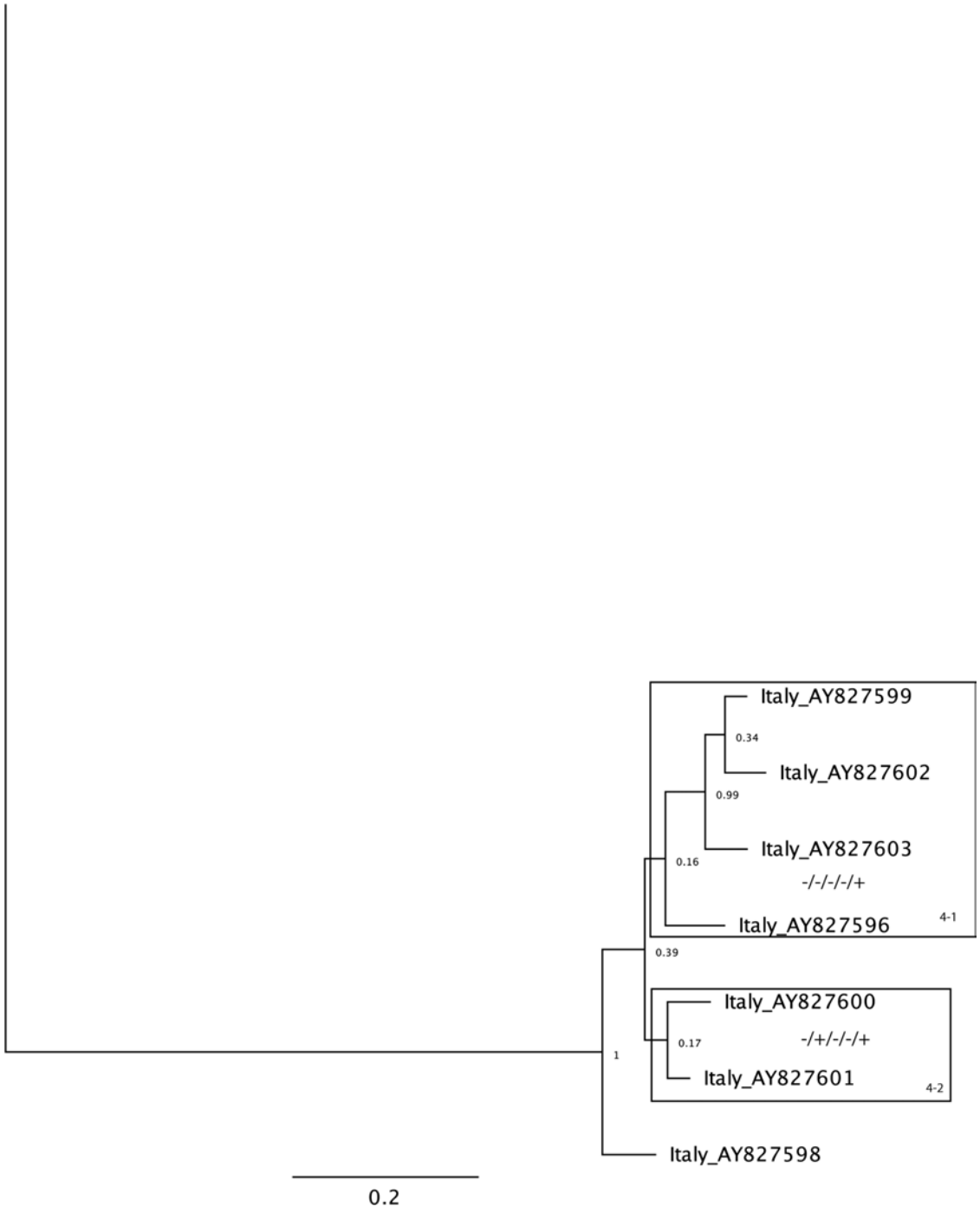

Figure S1. (*continued*)

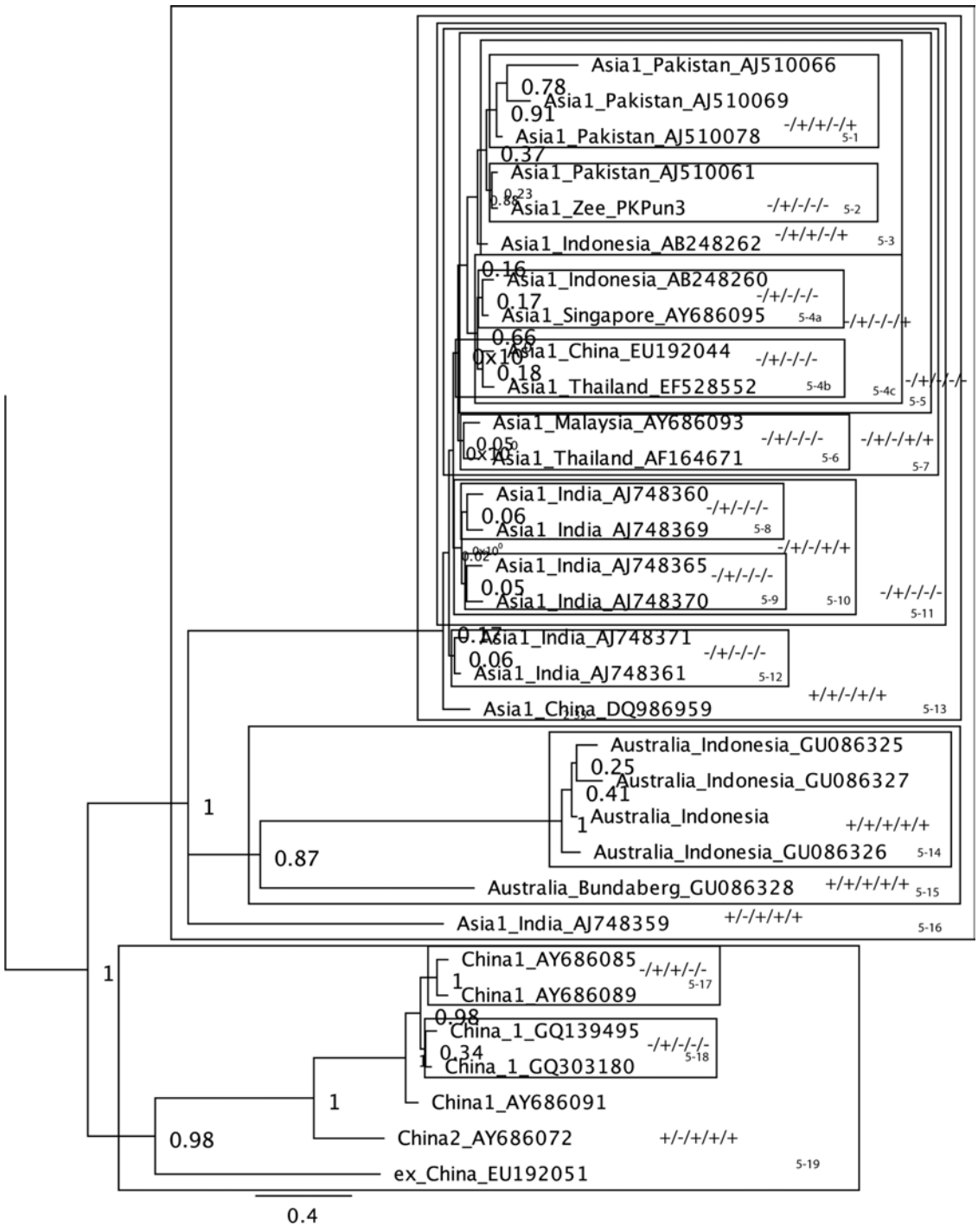

Figure S1. (*continued*)

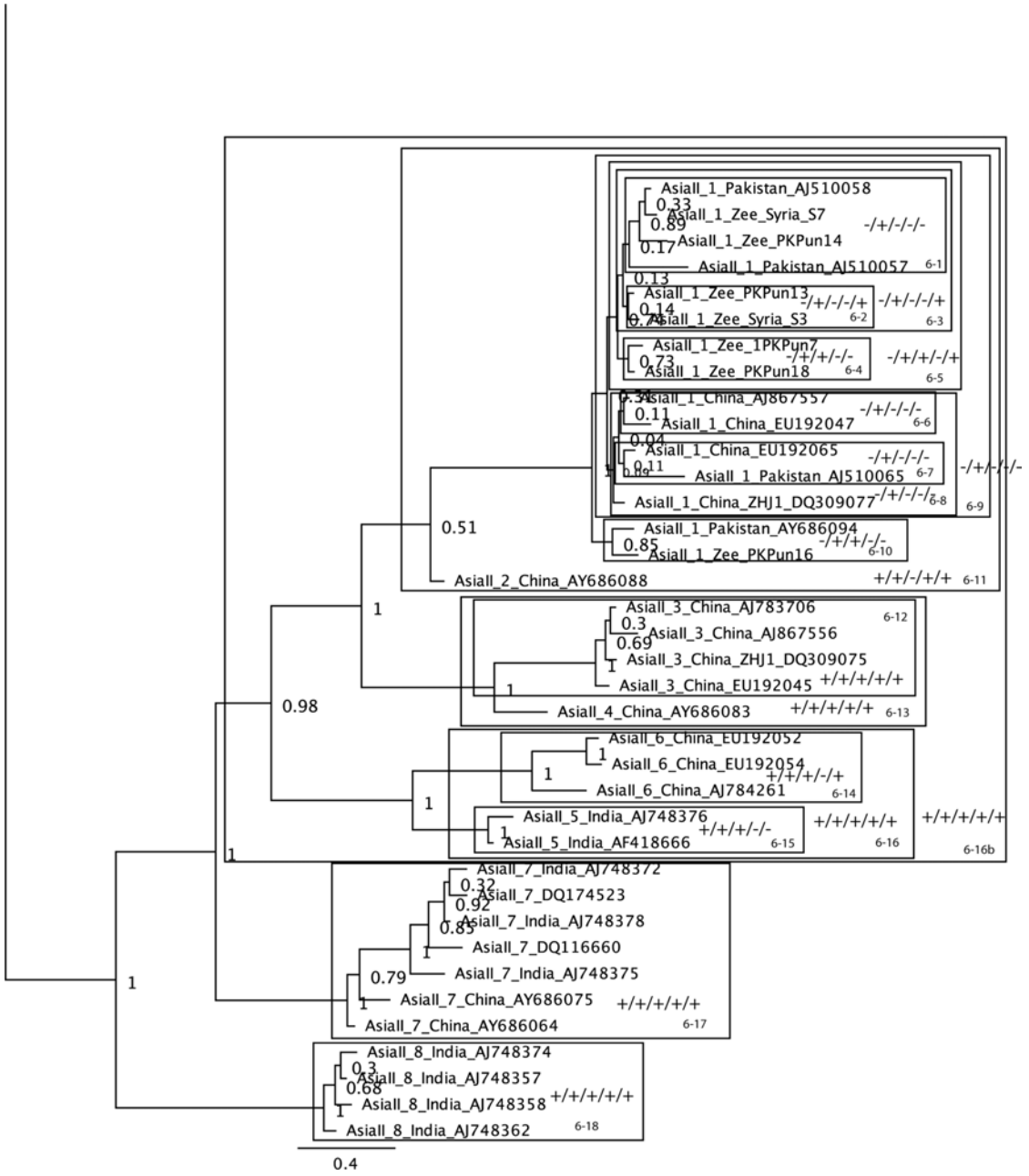

Figure S1. (*continued*)

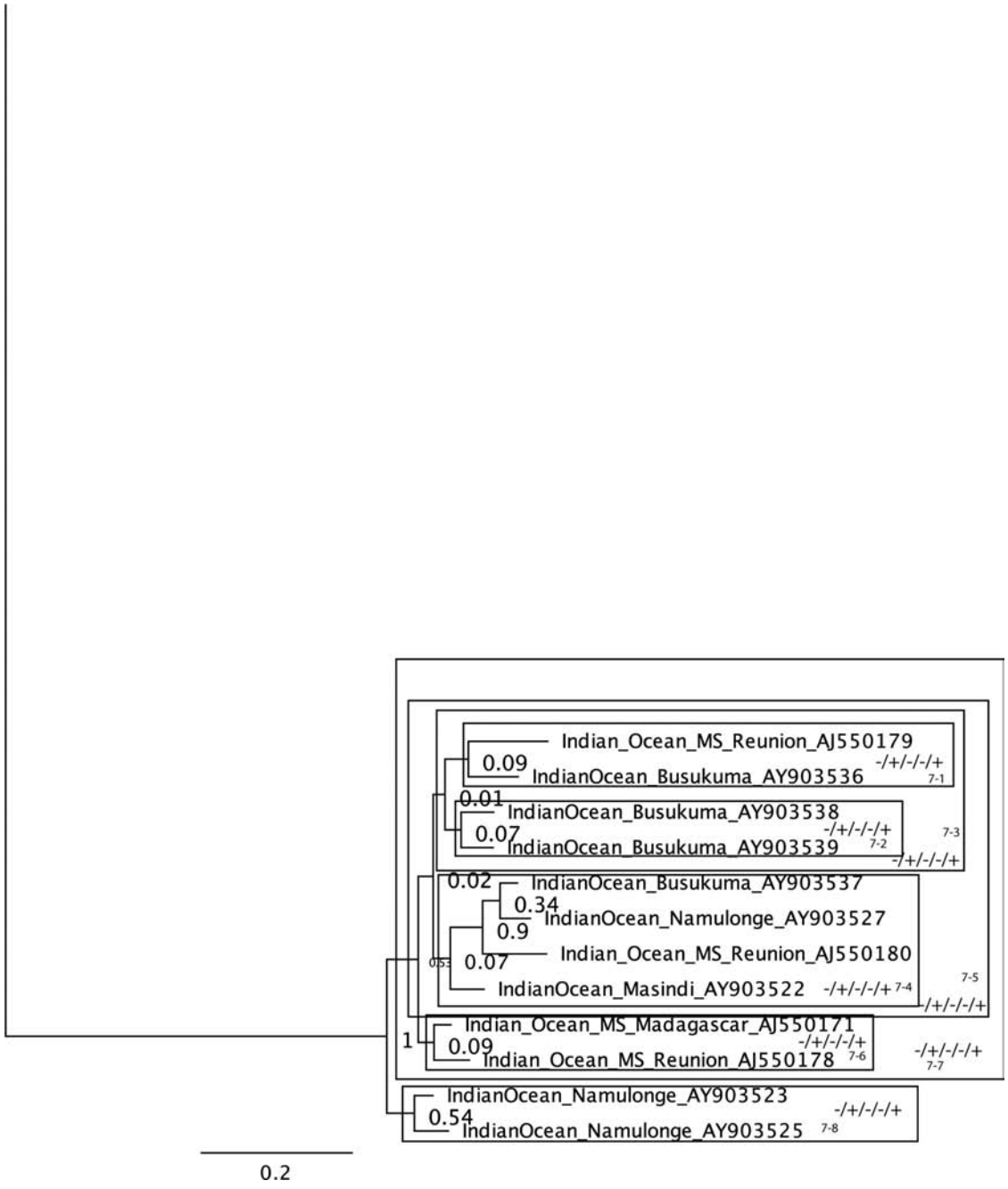

Figure S1. (*continued*)

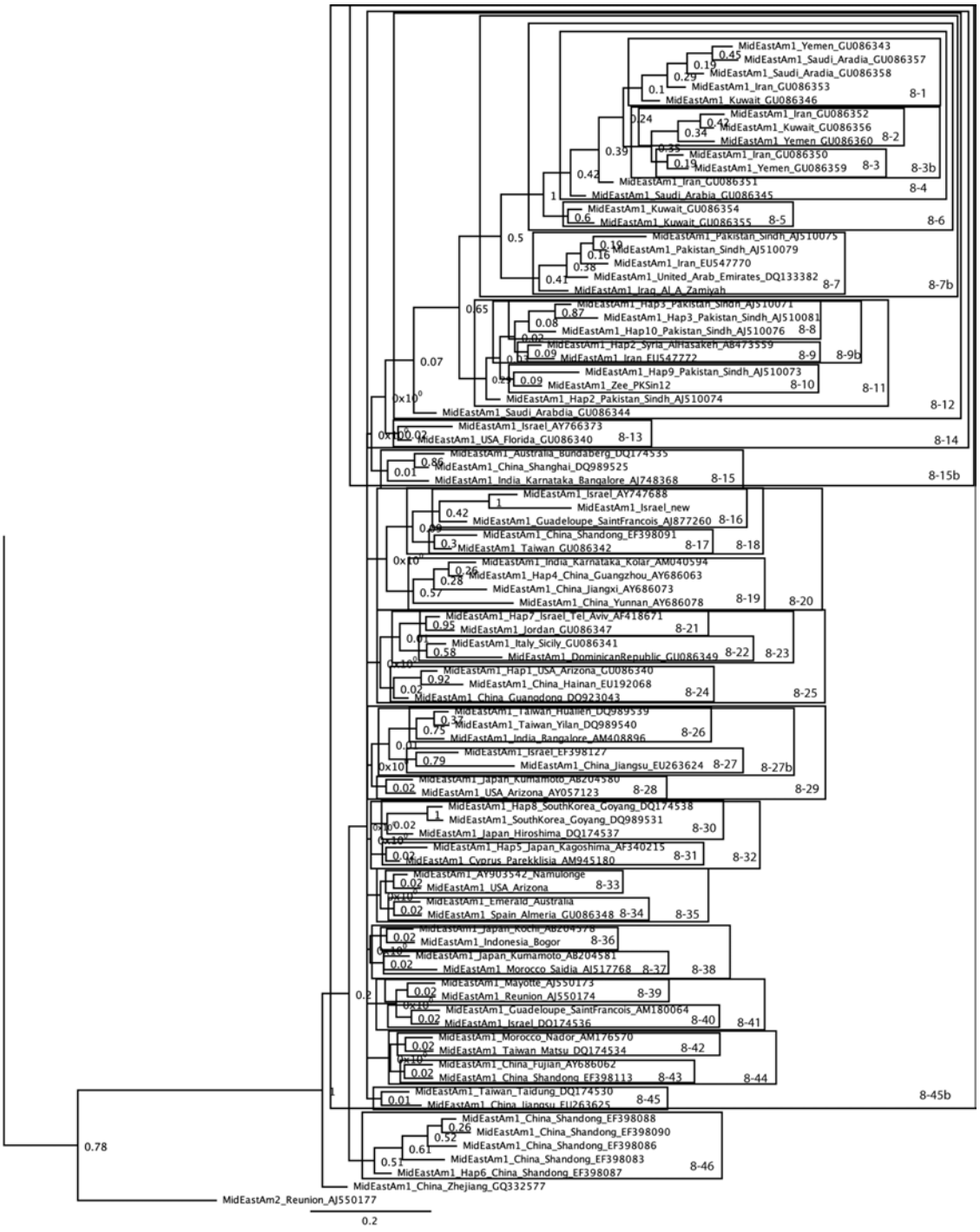

Figure S1. (*continued*)

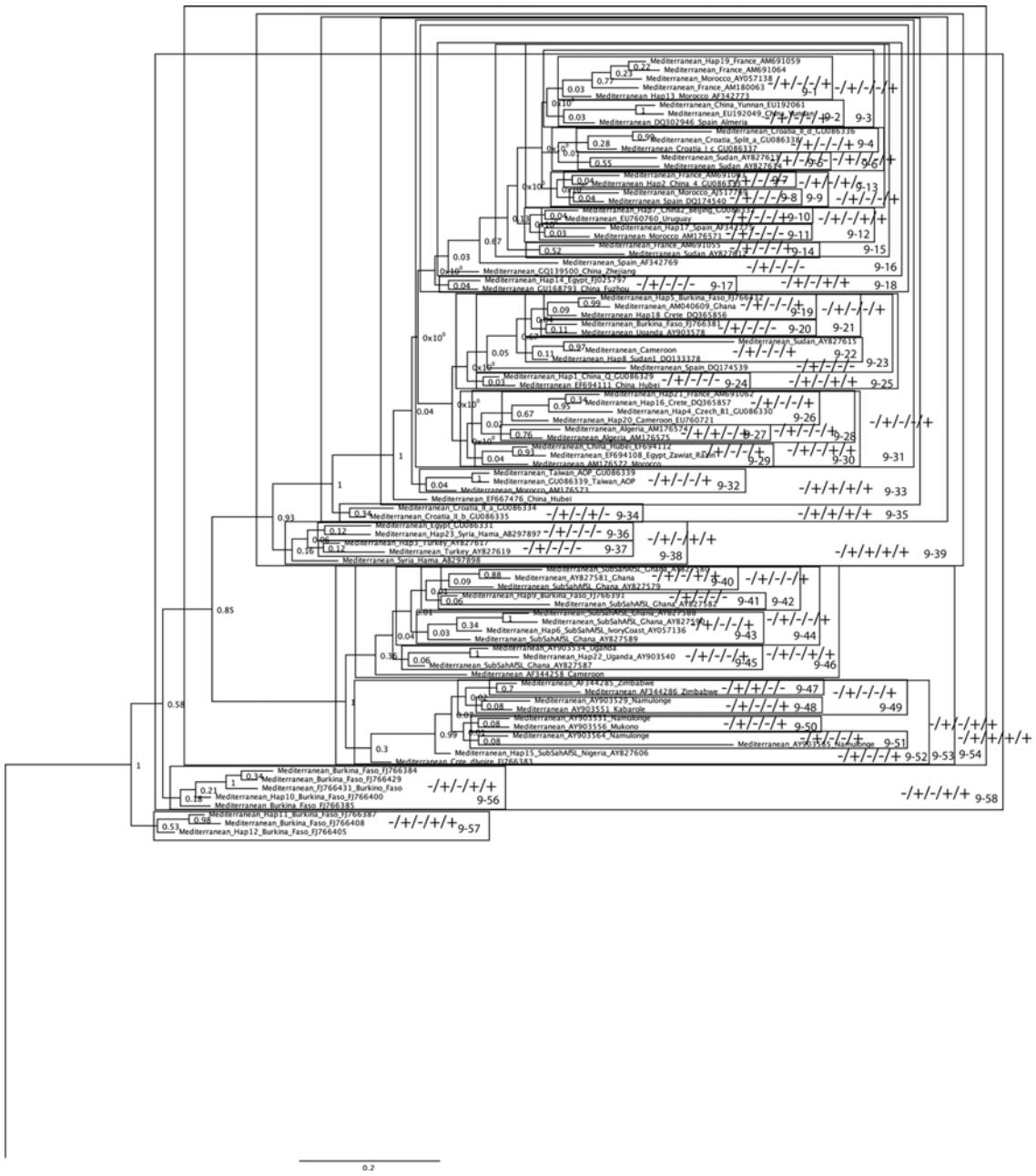

Figure S1.

A video abstract by the authors of this paper is available. video-abstract001.m4v.