

Global invasion history of the tropical fire ant: a stowaway on the first global trade routes

DIETRICH GOTZEK,*†¹ HEATHER J. AXEN,‡¹ ANDREW V. SUAREZ,* SARA HELMS CAHAN‡ and DEWAYNE SHOEMAKER§

*Department of Animal Biology and Department of Entomology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA, †Department of Entomology, National Museum of Natural History, Smithsonian Institution, Washington, DC 20013, USA, ‡Department of Biology, University of Vermont, Burlington, VT 05405, USA, §CMAVE, USDA-ARS, Gainesville, FL 32608, USA

Abstract

Biological invasions are largely thought to be contemporary, having recently increased sharply in the wake of globalization. However, human commerce had already become global by the mid-16th century when the Spanish connected the New World with Europe and Asia via their Manila galleon and West Indies trade routes. We use genetic data to trace the global invasion of one of the world's most widespread and invasive pest ants, the tropical fire ant, *Solenopsis geminata*. Our results reveal a pattern of introduction of Old World populations that is highly consistent with historical trading routes suggesting that Spanish trade introduced the tropical fire ant to Asia in the 16th century. We identify southwestern Mexico as the most likely source for the invasive populations, which is consistent with the use of Acapulco as the major Spanish port on the Pacific Ocean. From there, the Spanish galleons brought silver to Manila, which served as a hub for trade with China. The genetic data document a corresponding spread of *S. geminata* from Mexico via Manila to Taiwan and from there, throughout the Old World. Our descriptions of the worldwide spread of *S. geminata* represent a rare documented case of a biological invasion of a highly invasive and globally distributed pest species due to the earliest stages of global commerce.

Keywords: biogeography, fire ant, global trade, invasion biology, *Solenopsis geminata*

Received 20 September 2014; revision received 24 November 2014; accepted 6 December 2014

Introduction

Human-mediated changes to the environment and globalization of trade have brought biological invasions into the public, political and scientific limelight (Vitousek *et al.* 1997; Bright 1999; Pyšek & Richardson 2010; Lowry *et al.* 2013). Several recently introduced species have become notorious invaders and highly invasive and damaging pests, costing billions to control and taking a toll on human health and the environment (Pimentel *et al.* 2000, 2005). Human migration and trade have always aided dispersal of other organisms as com-

mensals. Indeed, Darwin (1859) had already recognized the transformation of ecosystems by invasions during his 19th century travels. Some of the earliest documented cases of dispersal via human migration are the spread of a skink and gecko during the colonization of Polynesia starting from 1500 BCE (Fisher 1997; Austin 1999) and the movement of house mice (Rajabi-Maham *et al.* 2008) and domestic animals following the Neolithic expansion (Larson *et al.* 2007; Sacks *et al.* 2013). However, historical records or archaeological evidence for species invasions is often lacking or difficult to acquire, and most evidence for commerce-mediated invasions is recent. This bias is undoubtedly augmented by human-mediated introductions having increased sharply with the rise in global transportation and trade (Meyerson & Mooney 2007; Westphal *et al.* 2007; Hulme

Correspondence: Dietrich Gotzek, Fax: +1-217-244-1224; E-mail: dgotzek@uga.edu

¹These authors contributed equally to the work.

2009). Although such global transport is largely assumed to be contemporary, Europe, Africa and Asia were already well connected through a vibrant commercial network, and trade had become truly global in the 16th century when the Spanish established trade routes across the Pacific and Atlantic Oceans (Flynn & Giráldez 2004). The Manila galleons and West Indies fleet connected Spain to the Asian markets via the New World. The extent to which early global trade may have been important for setting up many current biogeographic patterns is unknown, as we lack studies that examine range expansions at this time (but see Aplin *et al.* 2011).

Ants are ideal models to test whether the emergence of global trade had a far-reaching biogeographic impact, as they are readily moved via ship transport, making them highly susceptible to human dispersal. There are a number of widespread ant species with global or near global distributions (McGlynn 1999). Their native ranges are often unknown, as is accurate and detailed historical documentation of their spread. Many of these species are known to have reached a near global distribution by the 19th century (Wetterer 2005, 2008, 2010, 2011, 2012; Wetterer *et al.* 2009), raising the possibility that they were distributed along the same routes and during the same times. If so, comparative analyses of invasion histories should recover the predominant trade routes during the time of invasion and highlight the importance of human commerce as a main vector for human-mediated dispersal.

One such globally distributed ant species is the tropical fire ant (TFA), *Solenopsis geminata* (Fabricius 1804). While TFA is not as well studied as its congener, the Red Imported Fire Ant (*S. invicta*), TFA is often the most common, abundant and conspicuous ant in disturbed areas and has a much wider distribution. Having colonized virtually all tropical regions in the world (Wetterer 2011), it is one of the most widespread tropical and subtropical ants. Like many other widely distributed early invasives, there has been uncertainty regarding the native range of TFA. It has long been known outside the New World (Wetterer 2011); a red variant (subspecies *S. geminata rufa*, now a junior synonym of the nominal form (Ettershank 1966)) was initially considered to be indigenous to Asia and distinct from the New World forms (Creighton 1930). However, this red form was later found to occur from Florida to Panama, which suggests a New World origin for *S. g. rufa* (Creighton 1930). Trager (1991) hypothesized that the Old World *S. g. rufa* derived from a single introduction event based on the high uniformity of the Asian and Pacific populations, but noted that the dark West African specimens of the TFA more resembled forms from the southeastern U.S.A. and the Caribbean. Clearly, without exact knowledge of an invasive species'

population and invasion history, it is difficult to identify native ranges.

Tropical fire ant is a pioneer species (Perfecto 1991), a generalist keystone predator (Risch & Carroll 1982), and readily occupies urban and agro-ecosystems (Perfecto 1991; Holway *et al.* 2002; Perfecto & Vandermeer 2011). This ant also is well known as an agricultural pest (Lakshminantha *et al.* 1996; Holway *et al.* 2002) and is thought to be one of the infamous Hispaniolan plague ants of the early 16th century (Wilson 2005). Damage to crops is mostly indirect, by tending honeydew-producing aphids and other Hemiptera (Risch & Carroll 1982; Carroll & Risch 1984), but foragers also are known to girdle citrus trunks (Wolcott 1933), deter specialist pollinators (Carroll & Risch 1984) and damage irrigation tubing (Chang & Ota 1976). Aside from its negative impacts on agriculture, TFA can impact both vertebrate (Travis 1938; Kroll *et al.* 1973; Moulis 1996; Plentovich *et al.* 2009) and invertebrate (Lake & O'Dowd 1991; Way *et al.* 1998; Geetha *et al.* 2000) faunas. The latter has led it to be described as an important biocontrol agent of invertebrate pest species (Way *et al.* 2002; Way & Heong 2009).

Research on invasive populations of TFA has lagged far behind that of the other invasive ant species, particularly its congener *S. invicta* (Tschinkel 2006). Systematic study and understanding of biological invasions are not only important to better control, manage and prevent establishment of invasive species, but also represent ideal model systems for the study of important questions in evolution and ecology (e.g. rapid evolution, genetics of adaptation to new environments) (Sax *et al.* 2007; Suarez & Tsutsui 2008; Simberloff *et al.* 2013). Molecular genetic methods have become an important tool for the study of biological invasions (Estoup & Guillemaud 2010; Fitzpatrick *et al.* 2012; Kirk *et al.* 2013), allowing inference of parameters otherwise difficult or impossible to obtain by other means. Here, we present the first population genetic study using microsatellites and mtDNA sequences of the TFA on a global scale. We describe the population structure of this ant, identify the most likely source population, estimate times and infer routes of invasion, and document recent, human-mediated long distance dispersal. Our data indicate that TFA spread in the wake of the first global trade network in the 16th century.

Materials and methods

Data generation

Sample collection. A total of 192 TFA colonies were sampled from across the current geographic distribution of *Solenopsis geminata* (Fig. 1, Table S1, Supporting

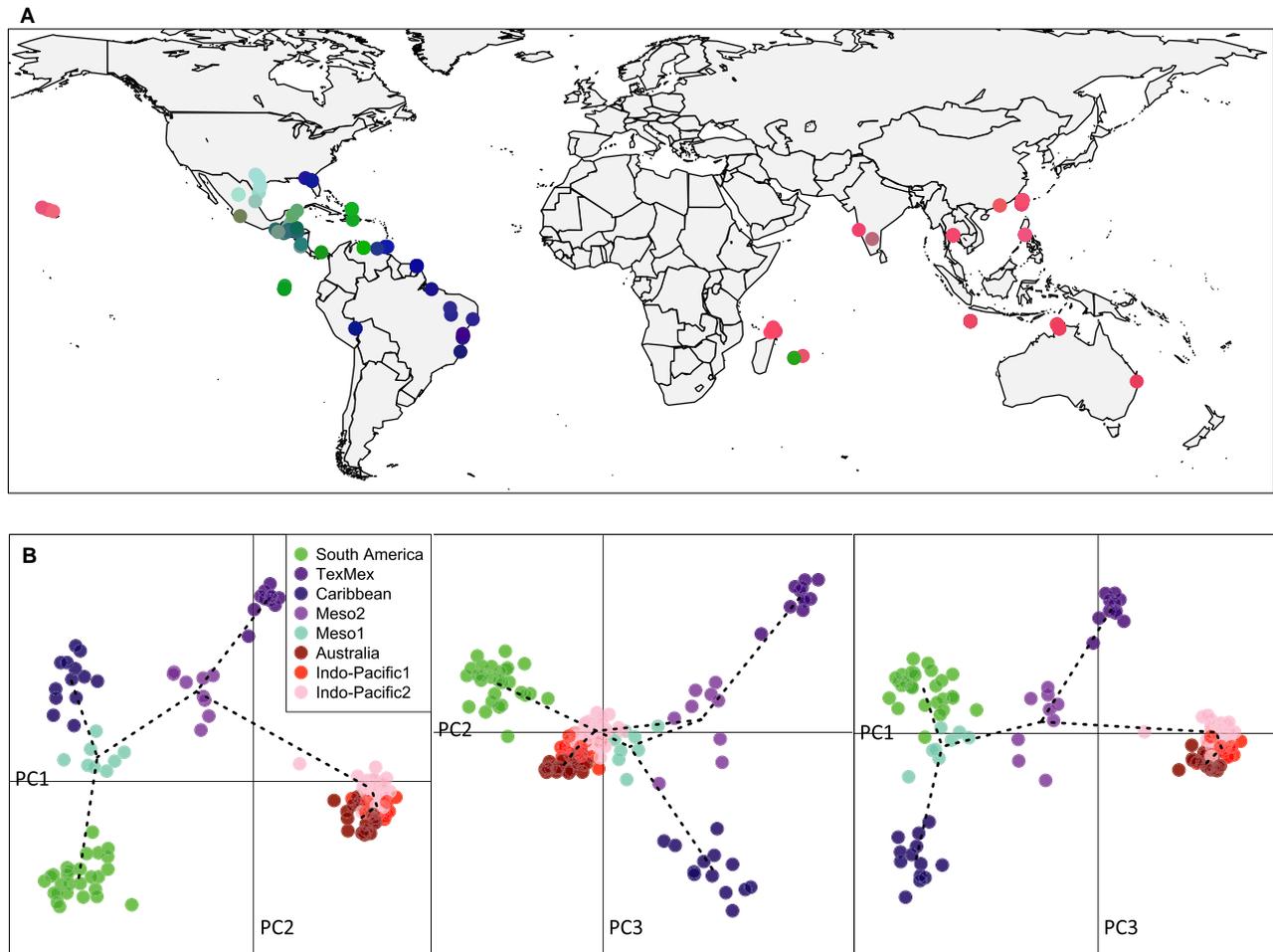


Fig. 1 Group memberships and projection in geographic and discriminant space. A) Geographic localities of samples. Samples are RGB colour coded according to the first three principal components of the DAPC (using four clusters which clearly distinguish an Old World (red), South American (dark blue), Caribbean (green) and North American (light blue) cluster. The Mesoamerican samples are intermediate between the New World clusters. B) Scatterplot of eight clusters recognized by DAPC. The first three principal components are shown (PC1: 26.1%, PC2: 7%, PC3: 4.8%). Cluster centroids are connected by a minimum-spanning tree.

information). To generate statistically independent samples, only a single worker ant was used from each colony. We roughly distinguish between the New World (i.e. the Americas including the Galápagos) and Old World (Africa, Asia and Australia). Samples of the three most closely related species to *S. geminata* (*S. xyloni*, *S. amblychila*, and *S. aurea*) were used as outgroups for the phylogenetic analyses (Trager 1991; Pitts *et al.* 2005).

Mitochondrial DNA sequencing. We amplified 646 bp of the mitochondrial gene *cytochrome c oxidase 1* (CO1) from 182 specimens of *S. geminata* and five outgroup specimens following published procedures (Ross *et al.* 2003). Resulting sequences were checked for the presence of premature stop codons and indels and compared to *S. geminata* mtDNA sequences from GenBank; amplified fragments were consistent with mtDNA and

not nuclear homologs. Sequences were readily aligned by hand and are deposited in GenBank (Table S1, Supporting information).

Microsatellite genotyping. Forty-five microsatellite markers were amplified following methods outlined in previous studies (Ascunce *et al.* 2009, 2011). Briefly, genotypes of a single female from 151 colonies were determined at 45 nuclear microsatellite loci. We only included individuals with genotypic data for >30 microsatellite markers in our analyses, resulting in 68 native and 77 invasive samples retained for subsequent study. Following best practice procedures (Bonin *et al.* 2004; Hoffman & Amos 2005), we estimated microsatellite genotyping error rate by independent allele scoring and double-checking of random individuals, which also helped to identify and eliminate errors that had

occurred during scoring of alleles by hand. Loci were screened for null alleles, Hardy–Weinberg equilibrium and linkage disequilibrium.

Data analysis

Generally, we analysed the mtDNA sequence data and nuclear microsatellite data separately to be able to compare and contrast potential differences between maternal and biparental gene flow and divergence. To examine population genetic structure at a finer resolution in the nuclear data set, we also analysed the New World and Old World/Australian samples separately.

Multivariate analyses of microsatellite data. As part of data exploration, we performed several multivariate analyses, which do not make strong assumptions about the underlying genetic model (Jombart *et al.* 2009). We conducted discriminant analysis of principal components (DAPC), which finds principal components best summarizing the differences between these clusters while minimizing within-cluster variation (Jombart *et al.* 2010). DAPC was carried out with the *adegenet* 1.3-8 package (Jombart & Ahmed 2011) implemented in R 3.0.1 (R Development Core Team 2013). As we wanted a naïve comparison between this method and the Bayesian clustering methods, we assumed no prior group membership and used sequential *K*-means clustering (up to $K = 15$) and model selection to infer genetic groups. The Bayesian information criterion (BIC) (Schwarz 1978) was used to assess the support for the model (i.e. the number of clusters and assignment of individuals to them), which is an efficient measure of support (Lee *et al.* 2009). Although *K*-means clustering is performed on transformed data using PCA, we used all 145 principal components, thus retaining all variation of the original data. We used 10^9 iterations and 10^3 random starting centroids for each run, which both aid in the convergence of the algorithm. To not overfit the discriminant function, we chose the optimal number of principal components for the DAPC using the *optim.a.score* function. The *a*-score captures the trade-off between the power of discrimination and overfitting using too many principal components in the analysis by measuring the proportion of successful reassignments of the DAPC analysis compared to *K*-means clustering (observed discrimination) and random clustering (random discrimination). Subsequent DAPCs were conducted with three principal components (explaining 37.8% of variance) and three retained discriminant functions. The use of principal components also ensures that information provided to the discriminant analysis is uncorrelated, thus removing potential effects of linkage disequilibrium. Plotting discriminant functions with

minimum-spanning trees connecting the cluster centroids allows a visual representation of affinities between clusters.

To validate our decision to recognize substructure in our data set, we compared stability of group membership probabilities of the eight inferred clusters to random clustering, giving an indication of how well supported the clusters are. Low group membership probabilities suggest that clusters are not supported by the data. As retention of too many principal components can result in overinflated membership probabilities and hence perfect discrimination, we conservatively used only three principal components and discriminant functions.

We estimated how well supported the group membership was relative to collection locality. Posterior group memberships can be used to indicate admixture or misclassification when prior groups are used to conduct the DAPC. We repeated the DAPC, grouping samples by region or country of origin (Fig. 2B). Membership assignment probabilities to each region based on retained discriminant functions were compared to the groups identified by *K*-means. Posterior assignment probabilities to a priori selected groups indicate the validity of a given group.

Bayesian clustering of microsatellite data. As the multivariate methods inferred cluster sizes varying up to 10-fold, we used the Bayesian clustering software STRUCTURAMA 2.0 (Huelsenbeck *et al.* 2011) to assign samples to clusters, which has been shown to outperform other methods (Fogelqvist *et al.* 2010; Hausdorf & Hennig 2010). STRUCTURAMA uses a Dirichlet process prior (DPP) to calculate the posterior probability of numbers of populations in the data set (Pella & Masuda 2006), which has been shown to be an efficient and powerful method to infer population structure (Huelsenbeck & Andolfatto 2007; Onogi *et al.* 2011; Shringarpure *et al.* 2011). STRUCTURAMA implements a Hierarchical Dirichlet Process model (Teh *et al.* 2006) to accommodate admixture of infinite ancestral populations while treating the number of populations as a random variable.

As this model is computationally demanding, we ran short exploratory analyses to assess the impact of the model (admixture vs. no admixture), concentration prior and hyperpriors on the analysis, with five or ten replicates for each model to test for consistency between analyses. Following the recommendations of François & Durand (2010), we employed various models using Bayes Factor to guide model choice. While we recognize that the harmonic mean reported by *Structurama* is a poor estimator of the marginal likelihood as its variance is often unreasonably large, we found that $K = 4$ was

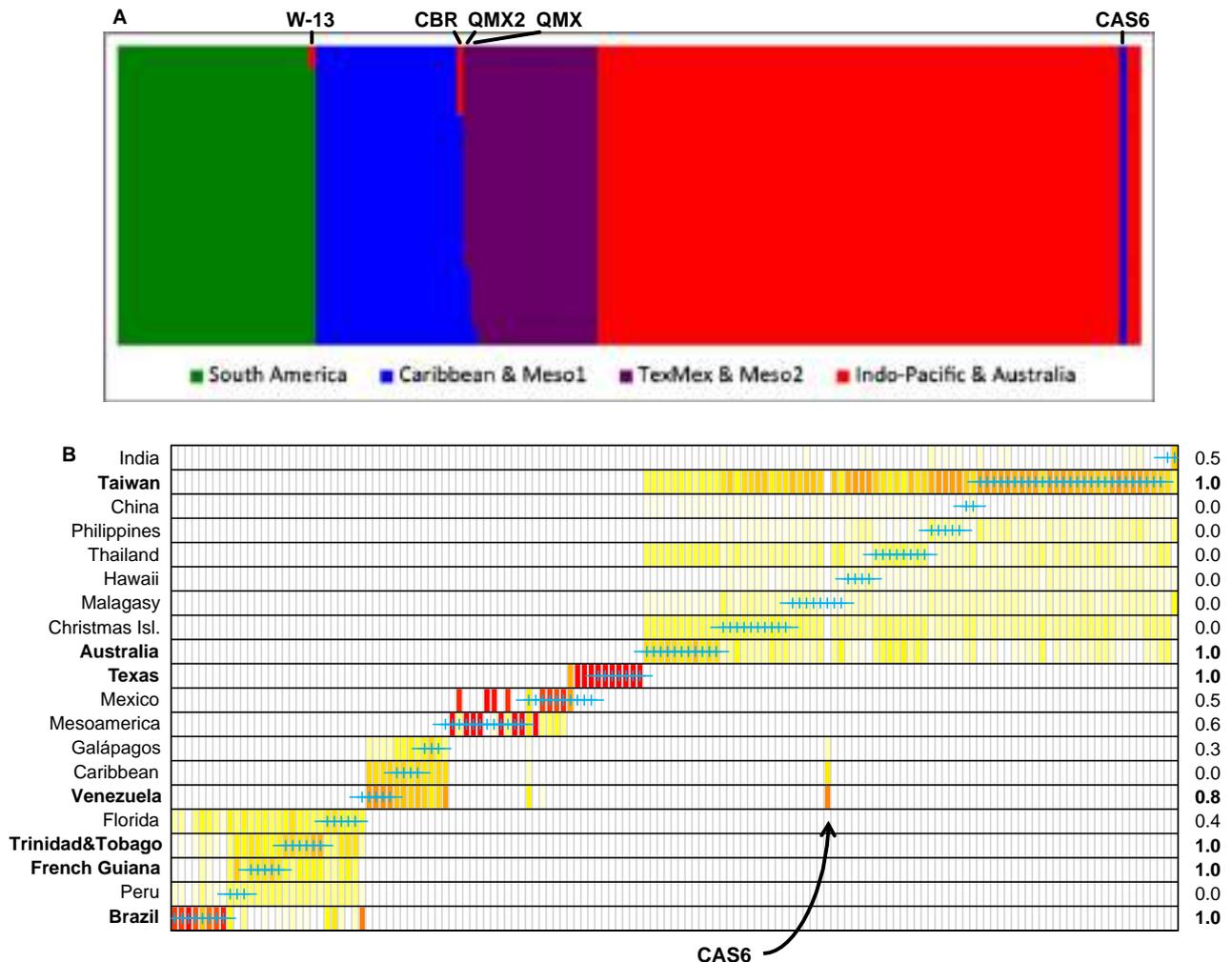


Fig. 2 Cluster assignment and admixture proportions as inferred by A) *Structure* for $K = 4$ and B) DAPC using collection sites as prior clusters. The heatmap (red = 1, white = 0) shows proportions of successful reassignment of individuals to their original clusters. Sites and their successful reassignments proportions are in rows, and individuals and their prior cluster (i.e. collection sites; blue crosses) are in columns. Large values indicate clear-cut clusters, and low values suggest admixed or poorly supported groups. Well-supported sites (≥ 0.8 reassignment proportions) are indicated in bold.

most often chosen, regardless of model used. In addition to the marginal likelihood (which were often very similar between analyses), we also used low sum of squares score for mean partition (i.e. the partition distance; Huelssenbeck & Andolfatto 2007) to identify the best analysis. This partition distance, which is the minimum number of individuals that must be moved between populations in one of the partitions to make it identical to the other partition (Gusfield 2002), measures the stability the partitioning scheme of a given analysis.

For the final model, we fixed α to give the desired mean of the prior for the expected number of populations [$E(K) = 5$] and modelled admixture using a gamma distribution shape and scale parameters (hyperpriors) set to 1 each. All analyses were conducted with the program default of sampling from a

single chain run for 100 000 generations with a burnin of 100.

General description of genetic variation. Measures of genetic diversity and population differentiation were separately estimated for the mtDNA and microsatellite data sets using the groups identified by the clustering methods. We analysed the mtDNA using MEGA 5.2.2 (Tamura *et al.* 2011) and DNASP 5.10.01 (Librado & Rozas 2009). All sequences were included in these analyses, not only unique haplotypes. For the nuclear data, we calculated indices of genetic variation with GENODIVE 2.0 (Meirmans & Van Tienderen 2004) and GENEALX 6.5 (Peakall & Smouse 2012). Following Meirmans & Hedrick (2011), we report pairwise population F_{ST} , G'_{ST} and D values (Hedrick 2005; Jost 2008;

Gerlach *et al.* 2010; Meirmans & Hedrick 2011) as none is an ideal summary statistic. We used nearly unbiased estimates of heterozygosity (H_S and H_T) (Nei & Chesser 1983) to reduce the bias of D and G''_{ST} values due to small sample size of some of the inferred populations. Statistical significance of pairwise F_{ST} values was tested using an analysis of molecular variance with 10 000 permutations (Excoffier *et al.* 1992; Michalakis & Excoffier 1996) with Bonferroni correction (Rice 1989).

Phylogenetic analyses. To obtain a more explicitly phylogenetic perspective of the relationships between clusters using the microsatellite data, we constructed a neighbour-joining tree (Saitou & Nei 1987) of interindividual Nei's chord distances (D_A ; Nei *et al.* 1983) using the *neighbor* program in the PHYLIP 3.69 package (Felsenstein 2005). This is expected to reflect genealogical relationships when a large number of informative markers are used (Chakraborty & Jin 1993; Bowcock *et al.* 1994). We used Nei's chord distance as it has been shown to outperform other distance measures for reconstructing phylogenetic trees using microsatellite data (Takezaki & Nei 1996, 2008). One thousand bootstrap replicates were used to estimate branch support.

Bayesian inference of the mtDNA genealogy was performed using BEAST 1.7.5 (Drummond *et al.* 2012). We estimated the best fitting partition and model of nucleotide substitution using the greedy heuristic search algorithm in PARTITIONFINDER 1.1. (Lanfear *et al.* 2012), which selected the HKY+I, F81 and GTR+G model for the first, second and third codon position, respectively. Clock and tree models were linked across partitions, but substitution models remained unlinked. We applied an exponential size coalescent model (Griffiths & Tavaré 1994) with a lognormal prior on the coalescent size parameter to estimate the tree. All other priors were kept at default. As we did not extensively sample the outgroup species and this violates the assumption of random sampling of OTUs in coalescent analyses (Wakeley 2008), we conducted the phylogenetic analyses with and without outgroups. No significant differences were found. Ten million generations were run, sampled every 1000th generation, of which the first 10% were discarded as burn-in. Stationarity of the runs was assessed in TRACER 1.5 (Rambaut & Drummond 2007) by plateauing of log-likelihoods and effective sample sizes (ESS) >200. Four independent runs were combined, all of which had similar mean log-likelihoods. To test for undue influence of the priors on the posterior parameter estimates, we compared posteriors from analyses estimated by sampling with and without (i.e. sampling only from the prior distribution) data.

Assignment and exclusion tests of microsatellite data. Assignment and exclusion tests were carried out with GENECLASS2 (Piry *et al.* 2004) using Bayesian approaches, which generally outperform distance and frequency-based approaches (Paetkau *et al.* 2004). We used the four native clusters as determined by the clustering methods as reference populations. Assignment tests were used to assign individuals collected in non-native areas to reference clusters from the native range. Assignment probabilities of $\geq 95\%$ to a given source population were considered to be significant support for the native population to be the source of invasives. In the absence of statistically significant assignment probabilities, the native reference population showing the highest average likelihood value was considered the most likely source population. Results of the Bayesian assignment tests were independent of the prior used (Rannala & Mountain 1997; Baudouin & Lebrun 2000). However, assignment methods assume that the actual source population is represented among the reference populations and can thus erroneously assign individuals to one or another reference population with high probability if the true source population has not been sampled (Paetkau *et al.* 2004). Exclusion tests are not prone to such an error, as they can exclude all reference populations as putative sources of introduced populations. To perform exclusion tests, we used the resampling algorithm of Paetkau *et al.* (2004), as other Monte Carlo resampling methods (Rannala & Mountain 1997; Cornuet *et al.* 1999) exclude an excess of resident individuals. All simulations were conducted with 100 000 simulated individuals and an alpha level of 0.01.

Once individuals from the invasive range were assigned to reference source populations, we tested whether any of these were first-generation migrants. As we could not be confident that we have sampled every native population, we used both the likelihood of an individual's genotype within the population where the individual has been sampled (L_{home}) and the ratio of L_{home} to the highest likelihood value among all populations excluding the population where the individual was sampled (L_{home}/L_{max}) as statistical criteria for the detection of first-generation migrants (Paetkau *et al.* 2004). While the likelihood ratio L_{home}/L_{max} has more power than the L_{home} statistic, it is only appropriate if all source populations for immigrants have been sampled. The L_{home} statistic is more appropriate when some source populations are clearly missing (Paetkau *et al.* 2004; Piry *et al.* 2004).

Testing invasion scenarios using approximate Bayesian computation. We used approximate Bayesian computation (ABC; Beaumont 2010; Bertorelle *et al.* 2010) to compare invasion scenarios and infer the invasion history of the

tropical fire ant. ABC is a Bayesian inference approach that does not require the specification of a likelihood function and can hence be used to efficiently carry out complex model-based inferences using large numbers of simulated data sets which are compared to the observed data set using summary statistics. All steps of the analyses were conducted with DIYABC 2.0 (Cornuet *et al.* 2014) using only microsatellite data. To limit the number of scenarios to test and for lack of a robust and reliable evolutionary history of the native populations due to rooting problems of both the microsatellite and mtDNA trees, we first sought to determine the source population(s) of the introduced clusters. To this end, we tested three competing scenarios (Fig. S3, Supporting information), with successively more narrowly defined putative source populations. For these analyses, we considered the invasives to belong to one population and successively more narrowly defined two competing putative source populations, loosely following the clusters recovered with increasing K : (A) the South American and TexMex/Meso1/Meso2/Caribbean clusters; (B) the TexMex/Meso2 and Meso1/Caribbean clusters; and (C) the TexMex and Meso2 clusters. The three competing scenarios allowed the introduced population to derive from one or the other source population or to be admixed from both.

Second, we tested more complex and specific invasion scenarios in analyses D from which we also derived parameter estimates of invasion times, bottleneck sizes and duration (Fig. S4, Supporting information). For this analysis, we recognized the Australian, Indo-Pacific and Meso2 clusters and we sought to distinguish between four invasion scenarios: an independent invasion, a serial invasion, an independent invasion from an unsampled ghost population and a serial invasion from an unsampled ghost population (Fig. S4A, Supporting information).

Following Cornuet *et al.* (2008), we considered only the simple generalized stepwise-mutation model (Estoup *et al.* 2002) to reduce the number of parameters. We left the mutation model at default settings. We also implemented a 5:6 female to male sex ratio (Travis 1941) and haplo-diploid locus model for all analyses, although these settings did not substantially influence the results (not shown). We only considered uniform priors and the following constraints on parameters: $db < t1 < t2 < ta$, $Nb < N$. For analyses A–C, we kept all parameter priors at their default. We set $N = [10-10\ 000]$, $t_{1,2} = [150-1000]$, $ta = [10-10\ 000]$ and $db = [1-1000]$ for analysis D. We used mean size variance, mean number of alleles, mean Garza-William's M index across loci, $(\delta\mu)^2$ distance between samples, mean size variance across loci, mean number of alleles across loci and F_{ST} as summary statistics. We produced reference

tables with 10^5 simulated data sets per scenario for analyses A–C and 10^6 data sets per scenario for analysis D for parameter estimation. To lighten the computational burden for analyses D, we used LDA-transformed summary statistics (Estoup *et al.* 2012). For analyses (A–C), we used raw summary statistics.

To reveal model (scenario) and/or prior misspecification prior to full analyses, we pre-evaluated scenario and prior distributions using both PCA and locating observed within simulated summary statistics (Cornuet *et al.* 2010) to verify that at least one prior–scenario combination can produce simulated data sets that are sufficiently close to the observed data set.

We estimated posterior probabilities of competing scenarios using polychotomous logistic regression of the 1% simulated data sets closest to the observed data set (which is generally more discriminant than the direct estimates). We assessed confidence in scenario choice by computing 95% confidence intervals and type I and II errors for the most probable scenario of analysis D. We validated the choice of thresholds by repeating posterior probability calculations with fewer (0.1% and 0.0025%) and more (10% and 0.1%) simulated data sets (Cornuet *et al.* 2008; Guillemaud *et al.* 2010), which produced similar results.

We estimated parameters after applying a *logit* transformation to the parameter values of the 1% simulated data sets closest to the observed data set. Use of other transformations (*log* or *log-tangent* transformation; Estoup *et al.* 2004; Hamilton *et al.* 2005) produced similar results (not shown). We additionally measured the performance of parameter estimation by calculating the median of the absolute error divided by the true parameter value of the 500 pseudo-observed data sets simulated using the median and mode of the posterior distribution as point estimates (relative median absolute errors, RMAE). Finally, we performed model checking with all summary statistics not used for the primary analysis using both PCA and ranking of summary statistics (Cornuet *et al.* 2010) to assess the goodness-of-fit of our model/parameter/posterior combination.

Results

Population structure

Four to nine clusters were supported by K -means clustering and Bayesian methods (Fig. 1, Fig. S1A, Supporting information). Four clusters represent the simplest summary of the data at the highest hierarchical level (Evanno *et al.* 2005), but the substantial substructure, in both geographic and discriminant space (Fig. 1), suggests eight clusters among which genetic differentiation was pronounced and always statistically significant

after Bonferroni correction ($P < 0.0005$; Table S3, Supporting information). Stability of group membership probabilities, derived from proportions of successful reassignments based on retained discriminant functions of the DAPC based on maximal substructuring, was high (100% for the New World clusters and $>82\%$ for the Old World clusters compared to 0–50% of random clustering (Fig. S2, Supporting information)).

Of the eight clusters, five were in the New World (Fig. 1B) and were consistently recovered in both global and New World-only analyses. The South American cluster contains all specimens from the Amazon, Guiana Shield, Brazilian coast and Florida. The Caribbean cluster contains samples from Panama, central Venezuela, the Dominican Republic, the Grand Turks, all samples from the Galápagos and one individual from Reunion (CAS6). The TexMex cluster contains samples from Texas and northern Mexico. The first Mesoamerican cluster (Meso1; containing samples from Costa Rica, Nicaragua, Honduras, Guatemala and a singleton from Mexico) partially overlaps the more northern second Mesoamerican cluster (Meso2; with samples from Mexico, Guatemala, Belize and a singleton from Costa Rica).

Finally, three clusters were identified in the Old World (Fig. 1B). Minimum-spanning trees of cluster centroids and distance in discriminant space of New to Old World samples suggest a close affinity between Mesoamerica and the Old World. Global analyses and Bayesian analyses of Old World samples only differentiate between the Australian cluster (containing all samples from Australia, Christmas Island and a singleton from Reunion (CAS3)) and an Indo-Pacific cluster (containing all other Old World samples). *K*-means clustering of Old World samples supports further substructure. At $K = 3$, the Indo-Pacific cluster is split into the Indo-Pacific1 (containing samples from Hawaii, Taiwan, India and all samples from China, Thailand and Madagascar) and Indo-Pacific2 clusters (containing the remaining samples from Hawaii, Taiwan and India, all samples from the Philippines and one sample from Mauritius). Due to the relatively poor separation in discriminant space and decreasing stability of group membership relative to the New World clusters in the global analysis, we chose to ignore finer substructure (Fig. 2B, Fig. S2, Supporting information).

Estimates of diversity

Estimates of genetic diversity show a reduction in the Old World clusters in both the nuclear and cytoplasmic genomes, pointing to an invasive origin of these samples (Tables S3 and S4, Supporting information). There were a total of 64 unique haplotypes in 186 mtDNA sequences generated. The putative native

populations ($n = 106$) contain the vast majority of mtDNA diversity (57 haplotypes; Table S2, Supporting information), whereas the putative invasive ants ($n = 75$) shared a total of 4 haplotypes. One of these (CAS6 from Reunion Island) is shared with two native specimens collected in the Caribbean islands of Turks and Caicos. Two specimens from Hawaii and a singleton from Christmas Island share the second haplotype. The third haplotype is unique to a sample from the Philippines. All remaining Old World samples (93%; $n = 70$) and a single specimen from Brazil (BrGem) share the fourth and most common haplotype. This haplotype is thus found throughout the Indo-Pacific region (Madagascar, Reunion Island, Mauritius, Australia, India, the Philippines, Thailand, China and Taiwan). The low number of haplotypes is reflected in the very low estimate of mitochondrial diversity (Table S2, Supporting information). The nuclear genome of the Old World clusters showed similar patterns of reduced genetic diversity relative to the native populations, but to a lesser degree (Table S2, Supporting information).

Admixture

Bayesian posterior assignment probabilities of individuals to a single cluster were generally very high ($>99\%$; Fig. 2A), indicating that there is little evidence for admixture between clusters, with few exceptions. Two Meso2 samples from the Yucatan (QMX and QMX2) contained a minority proportion (6.7% and 25.6%, respectively) of their genome from the Meso1 cluster. A Brazilian sample, W-13, was also admixed, with a small proportion (7.8%) of its genome derived from the Old World clusters. Finally, CBR from Jalisco, Mexico was estimated to derive 23.3% of its genome from the Old World cluster.

Reassignment of individuals to areas of origin based on the discriminant function of the DAPC (Fig. 2B) resulted in seven sites having high ($\geq 80\%$) reassignment proportions, indicating clear-cut groups. Six strongly differentiated clusters can be visually recognized mirroring the results of the clustering methods, validating our recognition of substructure in the data set, especially in the New World. Assignment proportions for sites and clusters outside the cluster inferred by the clustering methods were generally zero, indicating no admixture between clusters. Assignment proportions were often distributed across sites within a given cluster, indicating lack of within-cluster structure. Substructure is also visible in the Old World, suggesting that recognition of an Australian cluster is valid. CAS6 from Reunion is again not reassigned to its sampled site.

Phylogenetic analyses

Both the nuclear NJ tree and mtDNA coalescent generally recover the same groups as the clustering methods (Fig. 3). Importantly, in both data sets, CBR from southern Jalisco, Mexico is always the sister group to the invasive clade (with high support) and CAS6 from Reunion falls within a clade of Caribbean cluster samples (the Dominican Republic and Venezuela). In the NJ tree (Fig. 3A), the Philippine and most Taiwanese samples are sister to all samples within the Indo-Pacific clade, and the Australian cluster is situated well within the invasive samples. Despite uncertainty in rooting and lack of support for the deeper nodes in the mtDNA tree (Fig. 3B), the placement of haplotypes from the invasive range is consistent and well supported. The three Old World haplotypes form a well-supported monophyletic clade (PP = 1.0), which is always placed within a clade of North American samples.

Assignment and exclusion tests

We attempted to assign the Old World specimens to a native population (as identified by the clustering methods; Paetkau *et al.* 2004). Virtually all samples of the invasive range were assigned to the native Meso2 cluster with >95% probability (Table S4, Supporting information), with five exceptions. CAS6 from Reunion had a 100% assignment probability to the Caribbean cluster and four samples had a majority assignment to the TexMex cluster, but only one of these was significant at 95% assignment probability. Assignment tests can

incorrectly assign individuals to a reference population with high probability if the true source population is not sampled. Exclusion tests do not suffer from this problem. Exclusion probabilities for our data set were generally very low ($P < 0.05$), suggesting that we have not sampled the true source population. However, Meso2 could not be excluded as source population for most samples (at $P \geq 0.01$). Meso2 was excluded as source population only for CAS6, for which the Caribbean cluster could not be excluded ($P = 0.164$). Both the direct likelihood (L_{home}) and likelihood ratio ($L_{\text{home}}/L_{\text{max}}$) methods identified CAS6 from Reunion as a first-generation migrant with high probability (<0.0001%). No other F0 migrants were identified.

Inference of invasion scenarios using ABC

Progressive exclusion of native TFA populations as the source for the invasive Old World populations identified the Meso2 cluster as the most probable source. All other scenarios where the invasives are derived from other native populations (even through admixture) were rejected (Fig. S3, Supporting information). Of the more complex models using the Meso2, Australian and Indo-Pacific clusters, a serial invasion scenario (where the Australian cluster is derived from the Indo-Pacific cluster) was always preferred over an independent invasion scenario, indicating one initial invasion event took place with subsequent dispersal within the Old World (Fig. S4A, Supporting information). Scenarios including an unsampled ghost population were also always preferred over scenarios in which the source population had been

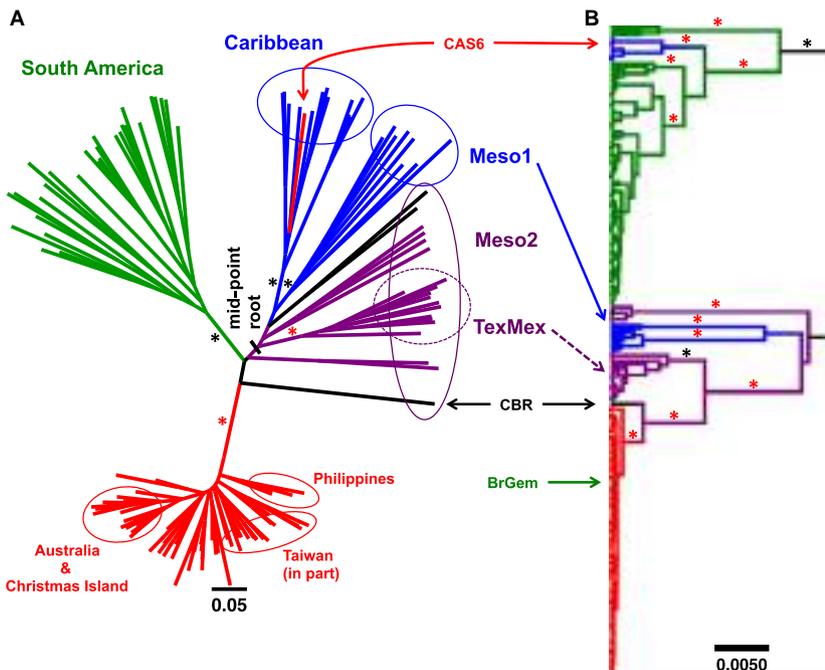


Fig. 3 Phylogenetic hypotheses based on nuclear (A) and cytoplasmic (B) genomes. Cluster membership is indicated by colour (red: Indo-Pacific & Australia; green: South America; blue: TexMex & Meso2; purple: Caribbean & Meso1). The three Mexican samples which are inconsistently placed between the Meso1 and Meso2 clusters are in black. Branch support is indicated by red (1.0) and black (0.95–0.99) stars; only deeper nodes are labelled. A) Bootstrapped (1000 replicates) neighbour-joining tree using microsatellite Nei's chord distances. B) Bayesian maximum clade credibility tree based on mtDNA sequences.

sampled. Scenario 3 (serial invasion from a ghost population) had the highest posterior probability (PP = 0.6575), and its 95% CI (0.5802, 0.7348) did not overlap with the 95% CI of the next best scenario (scenario 1; Fig. S4A, Supporting information). Type I and II errors were estimated at 0.31 and 0.124, respectively.

Posterior distributions of parameters are well estimated with peaked posteriors and clear differences to prior distributions and generally low RMAE values (Table 1, Fig. S4B, Supporting information). Estimates of population size indicate large effective population sizes for both native and invasive populations. Founding propagule size estimates are also rather large, which explains the modest loss of nuclear diversity in the introduced populations. The rather old divergence time of the ghost population suggests it is a native and not an invasive population. Divergence estimates of the invasive populations using the mode indicate the founding event occurred approximately 241 generations ago (95% CI 178, 938). This places the initial invasion event in the early 16th century, based on colony reproductive ages of 2 years. The Australian cluster diverged approximately in the mid-19th century (~163 generations). The bottlenecks were estimated to be quite long (~100 generations).

Discussion

We use nuclear and cytoplasmic genotype data for samples collected throughout the known global range of the

Table 1 Demographic parameter estimates, their relative median absolute errors (RMAE) and their 95% confidence interval of scenario 3 using ABC based on simulated 1% of simulated data sets closest to the observed values. All population sizes (N) are effective population sizes. N_{xb} : bottlenecked founding populations, db : bottleneck duration, t_j : divergence time. Time (t and db) is given in generations

| Parameter | Median | RMAE (median) | Mode | RMAE (mode) | q [2.5] | q [97.5] |
|---------------|--------|---------------|------|-------------|---------|----------|
| N1 (Meso2) | 9230 | 0.141 | 9430 | 0.150 | 7110 | 9960 |
| N2 (Austral.) | 4290 | 0.184 | 1800 | 0.253 | 990 | 9590 |
| N2b | 589 | 0.306 | 381 | 0.325 | 116 | 2770 |
| N3 (Indo-P.) | 1530 | 0.147 | 1010 | 0.172 | 354 | 7790 |
| N3b | 357 | 0.371 | 156 | 0.452 | 52 | 2450 |
| Nu (ghost) | 3920 | 0.245 | 2350 | 0.277 | 801 | 9530 |
| db | 123 | 0.452 | 77 | 0.610 | 11 | 427 |
| t_1 | 188 | 0.199 | 163 | 0.208 | 153 | 357 |
| t_2 | 382 | 0.149 | 241 | 0.176 | 178 | 938 |
| t_a | 3280 | 0.188 | 2980 | 0.220 | 850 | 8660 |

TFA, one of the most widely distributed pest ants, and describe its invasion history in great detail. We show that the Old World samples are introduced and derived from a single New World source population. Multiple lines of evidence support the Meso2 cluster and more specifically, southwestern Mexico, as the likely source, including results from clustering methods, admixture proportions, assignment and exclusion tests, measures of genetic diversity, tree-based methods and ABC. Creighton (1930) was the first to recognize the connection between the Old World and Mexican populations of the TFA, pointing out that the Old World *S. g. rufa* form was especially common in Texas and Mexico. Our analyses further suggest the invasive populations originated during one main introduction period followed by more recent long distance dispersal events from other source populations.

Divergence time estimates of the invasive populations from the native source population suggest that the main founding event(s) occurred in the early 16th century, a time of burgeoning colonialism and trade. These results are consistent with historical data. Since the early 15th century, Europeans explored and traded with Africa, Asia, the Americas and Oceania. This Age of Exploration culminated in the first truly global trade network by way of the Spanish Manila galleon trade, which connected the Asian and American markets (Flynn & Giráldez 2004). For 250 years (from 1565 to 1815 CE), one or two Spanish galleons set sail annually to trade New World silver for Chinese silk, porcelain and spices in Manila (Spate 2004). Importantly, the Spanish-Mexican colonists not only brought maize, sweet potatoes and other crops to Asia, but additionally carried rock, sand and soil for ballast (Carlton 1992), giving ample opportunity for entire TFA colonies to have been inadvertently transported during these long trans-Pacific voyages. Manila served as a hub for Spanish trade with the Chinese Ming and Qing Dynasties via the southern Chinese province of Fujian and to a lesser degree with Canton (Guangdong) province and the Molucca Spice Islands. Thus, the Spanish Philippines were well connected with the main trade centres of Asia, not only via the Portuguese and Dutch trade networks, but perhaps more importantly through an already existing vast and sophisticated Asian network (Spate 2004; Bjork 2005). The expansive and dynamic 16th–18th century trade system would have allowed the TFA to rapidly spread to the major economic and agricultural centres of coastal Asia, Africa, and the Pacific and Indian Oceans.

The Manila galleons sailed from the Pacific port of Acapulco, Oaxaca, in southwestern Mexico. Even though we lack samples from this area, our sample from Jalisco (CBR) is both geographically closest to

Acapulco (~450 km) and genetically closest to the introduced populations. This lack of sampling of populations around Acapulco and southwestern Mexico could explain divergence of the introduced populations from the native samples (e.g. the lack of shared mitochondrial haplotypes, the inference of an unsampled native ghost population and low power of the exclusion tests). However, low mitochondrial diversity stands in contrast to the relatively high nuclear genetic diversity. Although the low mtDNA haplotype variation in the invasive ants suggests a severe bottleneck during the founding event or selective sweeps associated with adaptation to the new environment, the nuclear genetic diversity of invasives is comparable to the diversity estimates of native populations and estimates of founding propagule size based on nuclear data number in the hundreds. Several factors may explain this observation. First, genetic bottlenecks are expected to be more severe for the cytoplasmic than the nuclear genome due to its smaller effective population size (Moritz *et al.* 1987). Second, founding propagules may have contained multiple reproductive queens (polygyny; Banks *et al.* 1973; Mackay *et al.* 1990). Queens in polygynous fire ant colonies are often derived from the same matriline (Ross *et al.* 1996), resulting in a shared cytoplasmic genome, while the nuclear genome is not similarly impacted. While the social form of most introduced populations has not been conclusively determined, field observations of nest structure and spacing suggest that polygyny is common. Third, selective sweeps due to adaptation to the newly invaded environment or endosymbiont-driven reductions can result in reduced mitochondrial diversity (Hurst & Jiggins 2005; Suarez & Tsutsui 2008). On the other hand, loss of genetic diversity can be modest or nonexistent if the founding propagule is large, the bottleneck brief and population growth rapid (Suarez & Tsutsui 2008). Moreover, historical processes operating in the source population(s) shape the level of genetic diversity introduced during the invasion process (Taylor & Keller 2007).

Despite the relative genetic homogeneity of the invasive clusters, several inferences on the spread of the TFA within the invasive range can be made. First, the DAPC, minimum-spanning tree and measures of population differentiation indicate that the Indo-Pacific2 cluster is closest to the source population, suggesting that the other non-native clusters stem from there. The Indo-Pacific2 cluster, containing all Philippine and most Taiwanese samples, also has the highest mtDNA haplotype diversity of the invasive clusters, as the two rare haplotypes are found within this cluster. The NJ tree places the Philippine samples sister to the remaining invasives, followed by most Taiwanese samples. This is congruent with direction of Spanish trade from Acapulco via Manila to

southern China and Formosa (Taiwan), where the Spaniards briefly settled in the 17th century. Finally, all analyses suggest that the Australian cluster is derived from the Indo-Pacific cluster and does not represent an independent invasion event. The ABC analysis estimates the invasion of Australia and the Christmas Island occurred more recently (188–376 years ago), which is consistent with the relatively late (ca. 1869 CE) European settlement of northern Australia.

Human-mediated movement of the TFA is still ongoing. We identified both samples from Reunion as recent or first-generation immigrants. While CAS3 may not be an F0 migrant, CAS6 (belonging to the Caribbean cluster) clearly is. Such long distance dispersals are a well-known feature of invasions (Suarez *et al.* 2001). There also is evidence for reintroduction of the TFA into its native range. Two samples from Bahia, Brazil (BrGem and W-13), have admixed genomes derived from the invasive and South American populations. A likely explanation for the genetic footprint of the invasive cluster in Brazil is due to a secondary invasion of TFA from the invasive range. Given that both samples were collected close to the city of Salvador, the economic and cultural hub of northeastern Brazil, it is perhaps not surprising to find evidence of a secondary invasion here.

We demonstrate that the TFA achieved its current pantropical distribution through Spanish trade with Asia during the 16th–18th century. One major introduction event from a single source population first brought the TFA from southwestern Mexico to the Philippines and then China, and from there, it was dispersed throughout the Old World tropics. However, movement of this highly invasive ant is still ongoing as indicated by the finding of long distance, first-generation migrants from the Caribbean to Reunion, or reintroduction from the invasive populations to the native range in Brazil. To the best of our knowledge, this is the first documented evidence of human-mediated introduction of a highly invasive pest species during the first truly global trade network. Our results are in contrast to many other unintentionally introduced social insects where data largely supports patterns of establishment much later, ranging from the late 19th century to mid-20th century (Tsutsui *et al.* 2001; Krushelnycky *et al.* 2005; Ascunce *et al.* 2011; Beggs *et al.* 2011; Evans *et al.* 2013). However, some invasive ant distributions look surprisingly similar to that of TFA and already appeared to be well established by the 19th century (Wetterer 2005, 2008, 2011, 2012; Wetterer *et al.* 2009). Thus, it is likely that other ants were spread along the same maritime trade routes, a conclusion supported by the fact that soil often was used as ballast (which was common well into the early 20th century; Carlton 1992), a habit that would suggest the transport of soil nesting ant colonies is highly likely. Further study of

the TFA and other ant species in their native and invasive ranges (e.g. Suarez *et al.* 2001; Ascunce *et al.* 2011) will shed light on the causes, processes and consequences of biological invasions and allow a richer assessment of the impact of human history on contemporary biogeographic patterns.

Acknowledgements

We thank the following scientists and institutions for generously contributing specimens: B. Hoffman, K.L. Heong and the International Rice Research Institute, S. Hasin, V. Framenau and the Western Australian Museum, W. Tschinkel, J. Longino, S.C.-C. Yang, L. van Aesch, D. Cherix, L. Davis, J. Orivel, B. Fisher, E. LeBrun, H. Herrerra, R. Arauco, K. Ross, and the National Museum of Natural History for sharing samples. E. Carroll kindly provided assistance in the laboratory. M. Ciolek pointed out important literature. K. Ross, J. Wetterer and three anonymous reviewers improved the manuscript with helpful comments and discussions. A.V.S. and D.G. gratefully acknowledge financial support from NSF (DEB 1020979) and USDA APHIS (292 AG 11-8130-0068-CA).

References

- Aplin KP, Suzuki H, Chinen AA *et al.* (2011) Multiple geographic origins of commensalism and complex dispersal history of black rats. *PLoS ONE*, **6**, e26357.
- Ascunce MS, Bouwma AM, Shoemaker DD (2009) Characterization of 24 microsatellite markers in 11 species of fire ants in the Genus *Solenopsis* (Hymenoptera: Formicidae). *Molecular Ecology Resources*, **9**, 1476–1479.
- Ascunce MS, Yang C-C, Oakey J *et al.* (2011) Global invasion history of the fire ant *Solenopsis invicta*. *Science*, **331**, 1066–1068.
- Austin CC (1999) Lizards took express train to Polynesia. *Nature*, **397**, 113–114.
- Banks WA, Plumley JK, Hicks DM (1973) Polygyny in a colony of the fire ant *Solenopsis geminata*. *Annals of the Entomological Society of America*, **66**, 234–235.
- Baudouin L, Lebrun P (2000) An operational Bayesian approach for the identification of sexually reproduced cross-fertilized populations using molecular markers. Proceedings of the International Symposium on Molecular Markers for Characterizing Genotypes and Identifying Cultivars in Horticulture, March 6–9, 2000, Montpellier, France. Leuven: International Society Horticultural Science, 81–93.
- Beaumont MA (2010) Approximate Bayesian computation in evolution and ecology. *Annual Review of Ecology, Evolution, and Systematics*, **41**, 379–406.
- Beggs JR, Brockerhoff EG, Corley JC *et al.* (2011) Ecological effects and management of invasive alien Vespidae. *BioControl*, **56**, 505–526.
- Bertorelle G, Benazzo A, Mona S (2010) ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Molecular Ecology*, **19**, 2609–2625.
- Bjork K (2005) The link that kept the Philippines Spanish: Mexican merchant interests and the Manila Trade, 1571–1815. *Journal of World History*, **9**, 25–50.
- Bonin A, Bellemain E, Bronken Eidesen P, Pompanon F, Brochmann C, Taberlet P (2004) How to track and assess genotyping errors in population genetics studies. *Molecular Ecology*, **13**, 3261–3273.
- Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL (1994) High resolution of human evolutionary trees with polymorphic microsatellites. *Nature*, **368**, 455–457.
- Bright C (1999) Invasive species: pathogens of globalization. *Foreign Policy*, **1999**, 50–64.
- Carlton JT (1992) Blue immigrants: the marine biology of maritime history. *Log*, **44**, 499–509.
- Carroll CR, Risch SJ (1984) The dynamics of seed harvesting in early successional communities by a tropical ant, *Solenopsis geminata*. *Oecologia*, **61**, 388–392.
- Chakraborty R, Jin L (1993) Determination of relatedness between individuals using DNA fingerprinting. *Human Biology*, **65**, 875–895.
- Chang VCS, Ota AK (1976) Fire ant damage to polyethylene tubing used in drip irrigation systems. *Journal of Economic Entomology*, **69**, 447–450.
- Cornuet J, Piry S, Luikart G, Estoup A, Solignac M (1999) New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics*, **153**, 1989–2000.
- Cornuet J-M, Santos F, Beaumont MA *et al.* (2008) Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics*, **24**, 2713–2719.
- Cornuet J-M, Ravigné V, Estoup A (2010) Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC Bioinformatics*, **11**, 401.
- Cornuet J-M, Pudlo P, Veysier J *et al.* (2014) DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism. DNA sequence and microsatellite data. *Bioinformatics*, **30**, 1187–1189.
- Creighton WS (1930) The new world species of the genus *Solenopsis* (Hymenop. Formicidae). *Proceedings of the American Academy of Arts and Sciences*, **66**, 39–151.
- Darwin C (1859) *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. John Murray, London.
- Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian Phylogenetics with BEAUTI and the BEAST 1.7. *Molecular Biology and Evolution*, **29**, 1969–1973.
- Estoup A, Guillemaud T (2010) Reconstructing routes of invasion using genetic data: why, how and so what? *Molecular Ecology*, **19**, 4113–4130.
- Estoup A, Jarne P, Cornuet J-M (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology*, **11**, 1591–1604.
- Estoup A, Beaumont M, Sennedot F, Moritz C, Cornuet J-M (2004) Genetic analysis of complex demographic scenarios: spatially expanding populations of the cane toad, *Bufo marinus*. *Evolution*, **58**, 2021–2036.
- Estoup A, Lombaert E, Marin J-M *et al.* (2012) Estimation of demographic model probabilities with Approximate Bayesian Computation using linear discriminant analysis on summary statistics. *Molecular Ecology Resources*, **12**, 846–855.
- Ettershank G (1966) A generic revision of the world Myrmicinae related to *Solenopsis* and *Pheidologeton* (Hymenoptera: Formicidae). *Australian Journal of Zoology*, **14**, 73–171.

- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.
- Evans TA, Forschler BT, Grace JK (2013) Biology of invasive termites: a worldwide review. *Annual Review of Entomology*, **58**, 455–474.
- Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes - application to human mitochondrial-DNA restriction data. *Genetics*, **131**, 479–491.
- Felsenstein J (2005) *PHYLIP v.3.6*. Distributed by the author. Dept. Genome Sci, Univ. Washington, Seattle.
- Fisher RN (1997) Dispersal and evolution of the Pacific Basin Gekkonid Lizards *Gehyra oceanica* and *Gehyra mutilata*. *Evolution*, **51**, 906–921.
- Fitzpatrick BM, Fordyce JA, Niemiller ML, Reynolds RG (2012) What can DNA tell us about biological invasions? *Biological Invasions*, **14**, 245–253.
- Flynn DO, Giráldez A (2004) Born with a “silver spoon”: the origin of world trade in 1571. *Journal of World History*, **6**, 201–221.
- Fogelqvist J, Niittyvuopio A, Ågren JA, Savolainen O, Lascoux M (2010) Cryptic population genetic structure: the number of inferred clusters depends on sample size. *Molecular Ecology Resources*, **10**, 314–323.
- François O, Durand E (2010) Spatially explicit Bayesian clustering models in population genetics. *Molecular Ecology Resources*, **10**, 773–784.
- Geetha V, Ajay N, Viswanathan G, Narendra A (2000) The effect of urbanisation on the biodiversity of ant fauna in and around Bangalore. *Journal of Ecobiology*, **12**, 115–122.
- Gerlach G, Jueterbock A, Kraemer P, Deppermann J, Harmand P (2010) Calculations of population differentiation based on Gst and D: forget Gst but not all of statistics!. *Molecular Ecology*, **19**, 3845–3852.
- Griffiths RC, Tavaré S (1994) Sampling theory for neutral alleles in a varying environment. *Philosophical Transactions of the Royal Society of London B*, **344**, 403–410.
- Guillemaud T, Beaumont MA, Ciosi M, Cornuet J-M, Estoup A (2010) Inferring introduction routes of invasive species using approximate Bayesian computation on microsatellite data. *Heredity*, **104**, 88–99.
- Gusfield D (2002) Partition-distance: a problem and class of perfect graphs arising in clustering. *Information Processing Letters*, **82**, 159–164.
- Hamilton G, Currat M, Ray N, Heckel G, Beaumont M, Excoffier L (2005) Bayesian estimation of recent migration rates after a spatial expansion. *Genetics*, **170**, 409–417.
- Hausdorf B, Hennig C (2010) Species delimitation using dominant and codominant multilocus markers. *Systematic Biology*, **59**, 491–503.
- Hedrick PW (2005) A standardized genetic differentiation measure. *Evolution*, **59**, 1633–1638.
- Hoffman JL, Amos W (2005) Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. *Molecular Ecology*, **14**, 599–612.
- Holway DA, Lach L, Suarez AV, Tsutsui ND, Case TJ (2002) The causes and consequences of ant invasions. *Annual Review of Ecology and Systematics*, **33**, 181–233.
- Huelsenbeck JP, Andolfatto P (2007) Inference of population structure under a dirichlet process model. *Genetics*, **175**, 1787–1802.
- Huelsenbeck JP, Andolfatto P, Huelsenbeck ET (2011) STRUCTURAMA: Bayesian inference of population structure. *Evolutionary Bioinformatics*, **7**, 55–59.
- Hulme PE (2009) Trade, transport and trouble: managing invasive species pathways in an era of globalization. *Journal of Applied Ecology*, **46**, 10–18.
- Hurst GDD, Jiggins FM (2005) Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceedings of the Royal Society B*, **272**, 1525–1534.
- Jombart T, Ahmed I (2011) ADEGENET 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, **27**, 3070–3071.
- Jombart T, Pontier D, Dufour A-B (2009) Genetic markers in the playground of multivariate analysis. *Heredity*, **102**, 330–341.
- Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, **11**, 94.
- Jost L (2008) Gst and its relatives do not measure differentiation. *Molecular Ecology*, **17**, 4015–4026.
- Kirk H, Dorn S, Mazzi D (2013) Molecular genetics and genomics generate new insights into invertebrate pest invasions. *Evolutionary Applications*, **6**, 842–856.
- Kroll JC, Arnold KA, Gotic RF (1973) An observation of predation by native fire ants on nestling barn swallows. *Wilson Bulletin*, **85**, 478–479.
- Krushelnicky PD, Loope LL, Reimer NJ (2005) The ecology, policy and management of ants in Hawaii. *Proceedings of the Hawaiian Entomological Society*, **37**, 1–25.
- Lake PS, O’Dowd DJ (1991) Red crabs in rain forest, Christmas Island: biotic resistance to invasion by an exotic snail. *Oikos*, **62**, 25–29.
- Lakshminantha BP, Lakshminarayan NG, Ali TMM, Veeresh GK (1996) Fire-ant damage to potato in Bangalore. *Journal of the Indian Potato Association*, **23**, 75–76.
- Lanfear R, Calcott B, Ho SYW, Guindon S (2012) PARTITIONFINDER: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, **29**, 1695–1701.
- Larson G, Albarella U, Dobney K *et al.* (2007) Ancient DNA, pig domestication, and the spread of the neolithic into Europe. *Proceedings of the National Academy of Sciences USA*, **104**, 15276–15281.
- Lee C, Abdool A, Huang C-H (2009) PCA-based population structure inference with generic clustering algorithms. *BMC Bioinformatics*, **10**, S73.
- Librado P, Rozas J (2009) DNASP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.
- Lowry E, Rollinson EJ, Laybourn AJ *et al.* (2013) Biological invasions: a field synopsis, systematic review, and database of the literature. *Ecology and Evolution*, **3**, 182–196.
- Mackay WP, Porter S, Gonzalez D *et al.* (1990) A comparison of monogyne and polygyne populations of the tropical fire ant, *Solenopsis geminata* (Hymenoptera: Formicidae), in Mexico. *Journal of the Kansas Entomological Society*, **63**, 611–615.
- McGlynn TP (1999) The Worldwide transfer of ants: geographical distribution and ecological invasions. *Journal of Biogeography*, **26**, 535–548.
- Meirmans PG, Hedrick PW (2011) Assessing population structure: FST and related measures. *Molecular Ecology Resources*, **11**, 5–18.

- Meirmans PG, Van Tienderen PH (2004) GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, **4**, 792–794.
- Meyerson LA, Mooney HA (2007) Invasive alien species in an era of globalization. *Frontiers in Ecology and Evolution*, **5**, 199–208.
- Michalakis Y, Excoffier L (1996) A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. *Genetics*, **142**, 1061–1064.
- Moritz C, Dowling TE, Brown WM (1987) Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annual Review of Ecology Evolution and Systematics*, **18**, 269–292.
- Moulis RA (1996) Predation by the imported fire ant (*Solenopsis invicta*) on loggerhead sea turtle (*Caretta caretta*) nests on Wassaw National Wildlife Refuge, Georgia. *Celonian Conservation Biology*, **36**, 439–472.
- Nei M, Chesser R (1983) Estimation of fixation indices and gene diversities. *Annals of Human Genetics*, **47**, 253–259.
- Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. *Journal of Molecular Evolution*, **19**, 153–170.
- Onogi A, Nurimoto M, Morita M (2011) Characterization of a Bayesian genetic clustering algorithm based on a Dirichlet process prior and comparison among Bayesian clustering methods. *BMC Bioinformatics*, **12**, 263.
- Paetkau D, Slade R, Burden M, Estoup A (2004) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology*, **13**, 55–65.
- Peakall R, Smouse PE (2012) GENALEX 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. *Bioinformatics*, **28**, 2537–2539.
- Pella J, Masuda M (2006) The Gibbs and split-merge sampler for population mixture analysis from genetic data with incomplete baselines. *Canadian Journal of Fisheries and Aquatic Sciences*, **63**, 576–596.
- Perfecto I (1991) Dynamics of *Solenopsis geminata* in a tropical fallow field after ploughing. *Oikos*, **62**, 139–144.
- Perfecto I, Vandermeer J (2011) Discovery dominance tradeoff: the case of *Pheidole subarmata* and *Solenopsis geminata* (Hymenoptera: Formicidae) in neotropical pastures. *Environmental Entomology*, **40**, 999–1006.
- Pimentel D, Lach L, Zuniga R, Morrison D (2000) Environmental and economic costs of nonindigenous species in the United States. *BioScience*, **50**, 53.
- Pimentel D, Zuniga R, Morrison D (2005) Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological Economics*, **52**, 273–288.
- Piry S, Alapetite A, Cornuet J-M, Paetkau D, Baudouin L, Estoup A (2004) GENECLASS2: a software for genetic assignment and first-generation migrant detection. *Journal of Heredity*, **95**, 536–539.
- Pitts JP, McHugh J, Ross KG (2005) Cladistic analysis of the fire ants of the *Solenopsis saevissima* species-group (Hymenoptera: Formicidae). *Zoologica Scripta*, **34**, 493–505.
- Plentovich S, Hebshi A, Conant S (2009) Detrimental effects of two widespread invasive ant species on weight and survival of colonial nesting seabirds in the Hawaiian Islands. *Biological Invasions*, **11**, 289–298.
- Pyšek P, Richardson DM (2010) Invasive species, environmental change and management, and health. *Annual Review of Environment and Resources*, **35**, 25–55.
- R Development Core Team (2013) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rajabi-Maham H, Orth A, Bonhomme F (2008) Phylogeography and postglacial expansion of *Mus musculus domesticus* inferred from mitochondrial DNA coalescent, from Iran to Europe. *Molecular Ecology*, **17**, 627–641.
- Rambaut A, Drummond AJ (2007) TRACER v1.4, Available from <http://beast.bio.ed.ac.uk/Tracer>
- Rannala B, Mountain J (1997) Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences USA*, **94**, 9197–9201.
- Rice WR (1989) Analyzing tables of statistical tests. *Evolution*, **43**, 223–225.
- Risch SJ, Carroll CR (1982) Effect of a keystone predaceous ant, *Solenopsis geminata*, on arthropods in a tropical agroecosystem. *Ecology*, **63**, 1979–1983.
- Ross KG, Vargo EL, Keller L (1996) Social evolution in a new environment: the case of introduced fire ants. *Proceedings of the National Academy of Sciences USA*, **93**, 3021–3025.
- Ross KG, Krieger MJ, Shoemaker DD (2003) Alternative genetic foundations for a key social polymorphism in fire ants. *Genetics*, **165**, 1853–1867.
- Sacks BN, Brown SK, Stephens D, Pedersen NC, Wu J-T, Berry O (2013) Y chromosome analysis of dingoes and Southeast Asian Village dogs suggests a neolithic continental expansion from Southeast Asia followed by multiple Austronesian dispersals. *Molecular Biology and Evolution*, **30**, 1103–1118.
- Saitou N, Nei M (1987) The Neighbor-Joining Method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, **4**, 406–425.
- Sax DF, Stachowicz JJ, Brown JH *et al.* (2007) Ecological and evolutionary insights from species invasions. *Trends in Ecology and Evolution*, **22**, 465–471.
- Schwarz GE (1978) Estimating the dimension of a model. *Annals of Statistics*, **6**, 461–464.
- Shringarpure S, Won D, Xing EP (2011) STRUCTHDP: automatic inference of number of clusters and population structure from admixed genotype data. *Bioinformatics*, **27**, i324–i332.
- Simberloff D, Martin J-L, Genovesi P *et al.* (2013) Impacts of biological invasions: what's what and the way forward. *Trends in Ecology and Evolution*, **28**, 58–66.
- Spate OHK (2004) *The Spanish Lake*. ANU E Press, Canberra.
- Suarez AV, Tsutsui ND (2008) The evolutionary consequences of biological invasions. *Molecular Ecology*, **17**, 351–360.
- Suarez AV, Holway DA, Case TJ (2001) Patterns of spread in biological invasions dominated by long-distance jump dispersal: insights from Argentine ants. *Proceedings of the National Academy of Sciences USA*, **98**, 1095–1100.
- Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics*, **144**, 389–390.
- Takezaki N, Nei M (2008) Empirical tests of the reliability of phylogenetic trees constructed with microsatellite DNA. *Genetics*, **178**, 385–392.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using

- maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.
- Taylor DR, Keller SR (2007) Historical range expansion determines the phylogenetic diversity introduced during contemporary species invasion. *Evolution*, **61**, 334–345.
- Teh YW, Jordan MI, Beal MJ, Blei DM (2006) Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, **101**, 1566–1581.
- Trager JC (1991) A revision of the fire ants, *Solenopsis geminata* group (Hymenoptera: Formicidae: Myrmicinae). *Journal of the New York Entomological Society*, **99**, 141–198.
- Travis BV (1938) The fire ant (*Solenopsis* spp.) as a pest of quail. *Journal of Economic Entomology*, **31**, 649–652.
- Travis BV (1941) Notes on the biology of the fire ant *Solenopsis geminata* (F.) in Florida and Georgia. *Florida Entomologist*, **24**, 15–22.
- Tschinkel WR (2006) *The Fire Ants*. Belknap Press, Cambridge.
- Tsutsui ND, Suarez AV, Holway DA, Case TJ (2001) Relationships among native and introduced populations of the Argentine ant (*Linepithema humile*) and the source of introduced populations. *Molecular Ecology*, **10**, 2151–2161.
- Vitousek PM, D'Antonio CMD, Loope LL, Rejmánek M, Westbrooks R (1997) Introduced species: a significant component of human-caused global change. *New Zealand Journal of Ecology*, **21**, 1–16.
- Wakeley J (2008) *Coalescent Theory: An Introduction*. Roberts & Company Publishers, Greenwood Village, Colorado.
- Way M, Heong K (2009) Significance of the tropical fire ant *Solenopsis geminata* (Hymenoptera: Formicidae) as part of the natural enemy complex responsible for successful biological control of many tropical irrigated rice pests. *Bulletin of Entomological Research*, **99**, 503–512.
- Way MJ, Islam Z, Heong KL, Joshi RC (1998) Ants in tropical irrigated rice: distribution and abundance, especially of *Solenopsis geminata* (Hymenoptera: Formicidae). *Bulletin of Entomological Research*, **88**, 467–476.
- Way MJ, Javier G, Heong KL (2002) The role of ants, especially the fire ant, *Solenopsis geminata* (Hymenoptera: Formicidae), in the biological control of tropical upland rice pests. *Bulletin of Entomological Research*, **92**, 431–437.
- Westphal MI, Browne M, MacKinnon K, Noble I (2007) The link between international trade and the global distribution of invasive alien species. *Biological Invasions*, **10**, 391–398.
- Wetterer JK (2005) Worldwide distribution and potential spread of the long-legged ant, *Anoplolepis gracilipes* (Hymenoptera: Formicidae). *Sociobiology*, **45**, 77–97.
- Wetterer JK (2008) Worldwide spread of the longhorn crazy ant, *Paratrechina longicornis* (Hymenoptera: Formicidae). *Myrmecological News*, **11**, 137–149.
- Wetterer JK (2010) Worldwide spread of the Pharaoh ant, *Monomorium pharaonis* (Hymenoptera: Formicidae). *Myrmecological News*, **13**, 115–129.
- Wetterer JK (2011) Worldwide spread of the tropical fire ant, *Solenopsis geminata* (Hymenoptera: Formicidae). *Myrmecological News*, **14**, 21–35.
- Wetterer JK (2012) Worldwide spread of the african big-headed ant, *Pheidole megacephala* (Hymenoptera: Formicidae). *Myrmecological News*, **17**, 51–62.
- Wetterer JK, Wild AL, Suarez AV, Roura-Pascual N, Espadaler X (2009) Worldwide spread of the Argentine ant, *Linepithema humile* (Hymenoptera: Formicidae). *Myrmecological News*, **12**, 187–194.
- Wilson EO (2005) Early ant plagues in the New World. *Nature*, **433**, 32.
- Wolcott GN (1933) Recent experiments in the control of two Puerto Rican ants. *Journal of the Department of Agriculture Puerto Rico*, **17**, 223–239.

D.G. and H.J.A. conceived the study; D.G., H.J.A. and D.D.S. performed research; and D.G., H.J.A., A.V.S., S.H.C. and D.D.S. wrote the manuscript.

Data accessibility

Sampling information can be found in Table S1, Supporting information. All data, scripts and input files are deposited in DataDryad (doi:10.5061/dryad.256kh) and DNA sequences are additionally deposited in GenBank (Accession nos: KP145683 – KP145854).

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1. A) Number of clusters inferred for the complete dataset. Left: BIC scores for $K = 1-15$ using K -means clustering. Right: Posterior probabilities using the DPP implemented in *Structurama*. B) Number of clusters for inferred for Old World samples only.

Fig. S2. Stability of group membership probabilities for maximal clustering ($K = 8$) of the complete dataset.

Fig. S3. (A) Three scenarios compared in ABC analyses A – C to identify the source population. (B) Posterior probabilities of competing scenarios (see SI Figure S3A) of ABC.

Fig. S4. (A) Scenarios and posterior probabilities (with 95% CI) used in analyses D based on logistic regression estimates. (B) Prior and posterior density distributions of demographic parameters for scenario 3 (serial invasion from unsampled ghost population) of analysis D.

Table S1. List of samples, cluster assignment, collection locality and coordinates, and data type.

Table S2. Estimates of genetic diversity (and standard error for microsatellites and standard deviation for mtDNA in brackets) within clusters for nuclear and cytoplasmic genomes.

Table S3. Degree of genetic differentiation between clusters (F_{ST} , Jost's D_{est} , Hedrick's G'_{ST}).

Table S4. Assignment and Exclusion tests of the invasive Old World samples to the New World clusters.