Extensive human-mediated jump dispersal within and across the native and introduced ranges of the invasive termite Reticulitermes flavipes


To cite this version:

Extensive human-mediated jump dispersal within and across the native and introduced ranges of the invasive termite *Reticulitermes flavipes*

Pierre-André Eyer†, Alexander J. Blumenfeld†, Laura N. L. Johnson1,2, Elfie Perdereau3, Phillip Shults1, Shichen Wang4, Franck Dedeine3, Simon Dupont3, Anne-Geneviève Bagnères3,5#, Edward L. Vargo1#

1Department of Entomology, 2143 TAMU, Texas A&M University, College Station, Texas, 77843-2143, USA
2Department of Veterinary Sciences, University of Wyoming, Laramie, Wyoming, 82070, USA
3IRBI, UMR 7261 CNRS-Université de Tours. Avenue Monge, Parc Grandmont, 37200 Tours, France
4Texas A&M Agrilife Genomics and Bioinformatics Service, College Station, USA
5CEFE, CNRS, Univ Montpellier, Univ Paul Valéry Montpellier, EPHE, IRS, 34000 Montpellier, France

†Pierre-André Eyer and Alexander J. Blumenfeld should be considered joint first author
#
Edward L. Vargo and Anne-Geneviève Bagnères should be considered joint last authors

*Corresponding author:
Pierre-André Eyer
Department of Entomology,
Texas A&M University,
College Station, 77843, Texas, USA
e-mail: pieyer@live.fr

Keywords: Population genetics, Phylogeography, Social insects, Invasive species, ABC
Abstract

As native ranges are often geographically structured, invasive species originating from a single source population only carry a fraction of the genetic diversity present in their native range. The invasion process is thus often associated with a drastic loss of genetic diversity resulting from a founder event. However, the fraction of diversity brought to the invasive range may vary under different invasion histories, increasing with the size of the propagule, the number of re-introduction events, and/or the total genetic diversity represented by the various source populations in a multiple-introduction scenario. In this study, we generated a SNP dataset for the invasive termite *Reticulitermes flavipes* from 23 native populations in the eastern United States and six introduced populations throughout the world. Using population genetic analyses and approximate Bayesian computation Random Forest, we investigated its worldwide invasion history. We found a complex invasion pathway with multiple events out of the native range and bridgehead introductions from the introduced population in France. Our data suggest that extensive long-distance jump dispersal appears common in both the native and introduced ranges of this species, likely through human transportation. Overall, our results show that similar to multiple introduction events into the invasive range, admixture in the native range prior to invasion can potentially favor invasion success by increasing the genetic diversity that is later transferred to the introduced range.
INTRODUCTION

The transport of species beyond their native ranges by human activity is breaking down biogeographical barriers and causing global reorganization of biota (Capinha et al. 2015; van Kleunen et al. 2015), with the ensuing invasions posing a serious threat to biodiversity, agriculture and human health (Simberloff et al. 2013). Successful invaders must disperse into a geographically distant area, establish a viable and fertile population, and spread throughout this new environment, where the biotic and abiotic pressures may differ from those they faced in their native range (Kolar and Lodge 2001). Biological invasions have long been seen as paradoxical, as the invasion process was thought to occur in spite of the reduction of genetic diversity that typically follows introductions of invasive species (Sax and Brown 2000). However, data from a growing number of studies suggest that biological invasions are not always associated with a loss of genetic diversity, and that a loss of genetic diversity is not always accompanied with inbreeding costs and a loss of adaptive potential (Facon et al. 2006, Roman & Darling 2007, Estoup et al. 2016, Eyer et al. 2018a, Blumenfeld et al. 2021). In addition, the ecological dominance of invaders in their novel environments is not necessarily the result of superior competitive ability compared to native species, but may simply involve the filling of vacant niches (Dlugosh and Parker 2008, Dlugosch et al. 2015, Bates et al. 2020).

Several life-history traits may enhance the invasive success of some species (Eyer & Vargo 2021). Specific breeding systems, modes of dispersal or physiological characteristics may influence the ability of species to spread and to become established. Investigating the mechanisms underlying the invasion process requires determining whether these traits differ between introduced and native populations. Such differences may arise after the introduction due to new ecological pressures occurring in the invaded area (Wares et al. 2005; Keller and Taylor 2008), or they may already be present within native populations, thereby pre-adapting the source population for invasion success. Therefore, determining the source population of invasive species is critical to conduct comparative studies of life-history traits between introduced and native
ranges to understand how they evolved under distinct biotic and abiotic pressures (Barker et al. 2017).

Investigating invasion mechanisms also requires knowledge of the invasion history, in which a series of demographic events may influence the invasion process and patterns of genetic diversity. The introduced range may consist of a single invasive population. This introduced population may have originated from a single introduction out of the native range, or from multiple introductions out of the native range, either from the same or different source populations. In contrast, the introduced range may comprise multiple invasive populations, which may originate from separate introduction events from one source population, or from different source populations out of the native range (Oficialdegui et al. 2019; Acevedo-Limón et al. 2020). Finally, an established invasive population itself may become a source for subsequent invasions, a phenomenon coined the ‘bridgehead effect’ (Lombaert et al. 2010; Bertelsmeier and Keller 2018).

Therefore, reconstructing invasion histories is important for explaining the global distribution of genetic diversity and understanding adaptive evolution in new environments (Cristescu 2015, van Boheemen and Hodgins 2020).

A bottleneck event following an introduction usually results in a loss of genetic diversity in the introduced population (Dlugosch and Parker 2008), but the amount of genetic diversity lost may vary under different invasion scenarios. The degree to which genetic diversity is reduced may be limited when the initial colonizing force is large, when the introduced population is subsequently re-invaded by additional individuals during multiple introduction events, and/or when the introduced population is invaded by individuals from several genetically distinct source populations (Facon et al. 2006). Sometimes, when there are several introductions from different source populations and these interbreed within an invasive population, genetic diversity may even be higher within this population than its native source population(s) (Facon et al. 2008). In contrast, the bridgehead effect may result in a severe loss of diversity, as subsequent introductions arise from an already depauperate introduced population. The bridgehead effect
has been suggested to promote the spread of phenotypic traits enhancing invasion success in secondary invasive populations, as these traits are already selected for and widespread in the initial introduced population, although there is limited support for such a phenomenon (Bertelsmeier and Keller 2018). Investigating patterns of genetic diversity in native and introduced populations can therefore provide insights into the introduction history of invasive species (e.g., Winkler et al. 2019, Geburzi et al. 2020, Hirsch et al. 2021, Resh et al. 2021, Wesse et al. 2021).

*Reticulitermes flavipes* is a subterranean termite species native in the eastern USA, where it ranges from Texas to Massachusetts. The termite has become invasive in localities both near to and distant from the eastern USA. In both its native and introduced ranges, this termite species is responsible for large amounts of damage to human structures (Evans et al. 2013, Shults et al. 2021). This includes the western USA (Austin et al. 2005; McKern et al. 2006), the Province of Ontario in Canada (Kirby 1965), the Bahamas (Scheffrahn et al. 1999), Chile (Clément et al. 2001) and Uruguay in South America (Austin et al. 2005; Su et al. 2006) and France, Germany, Austria and Italy in Western Europe, where it was first reported in 1837 (Kollar 1837; Weidner 1937; Clément et al. 2001; Ghesini et al. 2010). This species has also been reported (GBIF) from Mexico and the outermost regions of Spain (Canary Island; Hernández-Teixidor et al. 2019) and Portugal (Azores; Austin et al. 2012).

The native and invasive populations of *R. flavipes* have been the focus of numerous studies investigating its breeding system. In the French invasive range, colonies are large, readily fuse together and contain several hundred neotenics (worker or nymph-derived reproductives that replace the primary or alate-derived reproductives who found new colonies) (Dronnet et al. 2005; Vargo and Husseneder 2009; Perdereau et al. 2010a). Although substantial variability in breeding structure is present among the native USA populations of *R. flavipes*, colonies from most native populations are spatially less expansive, fuse only occasionally and are headed by a monogamous pair of primary reproductives or a few neotenics (Vargo and Husseneder 2009; Vargo et al. 2013; Aguero et al. 2020). Interestingly, colonies in a population from Louisiana share
some of the same traits as those in France (Perdereau et al. 2010a; Perdereau et al. 2010c; Perdereau et al. 2015).

Previous genetic analyses based on microsatellite markers and mtDNA haplotypes have shown that the introduced French population of *R. flavipes* exhibits an average decrease in genetic diversity of 60-80% compared to native USA populations (Perdereau et al. 2013). The analysis also revealed the occurrence of three main genetic clusters within the native USA range – the ‘Eastern cluster’ (West Virginia, Virginia, Delaware, North and South Carolina), the ‘Gulf Coast cluster’ (Florida and Eastern Mississippi–Louisiana) and the ‘Southern Louisiana cluster’ (the New Orleans and Baton Rouge regions in Louisiana) (Perdereau et al. 2013). Notably, some microsatellite and mtDNA haplotypes found in France were unique to the Southern Louisiana cluster (Perdereau et al. 2013). This finding, together with similarities in chemical profiles and breeding structures found between France and Louisiana (Perdereau et al. 2010b, Perdereau et al. 2015), suggested that the French population of *R. flavipes* was introduced from Louisiana, most likely during the 17th and 18th centuries via wood and plant trade between New Orleans and the major French ports on the Atlantic coast (Dronnet et al. 2005, Perdereau et al. 2010a, Perdereau et al. 2013).

Although the Louisiana origin of the invasive French population appears well supported, several points remain unclear. First, Perdereau et al. (2019) recently identified a French haplotype more closely related to the ‘Eastern cluster’ than the ‘Southern Louisiana cluster’, suggesting multiple native populations from the USA may have invaded France. Additionally, the source(s) of the Canadian and Chilean invasions remain unidentified. Although several populations of *R. flavipes* occur in the Northeastern and Midwestern USA (*i.e.*, adjacent to Ontario), the only haplotype found in Canada was shared with populations in Louisiana and France (Perdereau et al. 2013). Therefore, it is unclear whether the Canadian population arose from a primary introduction from Louisiana or from a secondary introduction through France (*i.e.*, bridgehead introduction), as eastern Canada and France share a close historical bond. Similarly, Chile's
unique haplotype was closest to one shared between Louisiana and France (Perdereau et al. 2013), raising the same question regarding primary versus secondary introduction. Overall, these findings suggest a complex invasion history for *R. flavipes* and raise the question of how many native populations may have served as sources for the introduced populations and what the role of bridgeheads might be in the global distribution of this species.

Here, we used population genetic analyses and approximate Bayesian computation Random Forest (ABC-RF) to investigate the invasion history of *R. flavipes*. Using ddRadSeq, we first generated a SNP dataset sequencing 23 native populations in the USA and six introduced populations in France, Germany, Chile, Uruguay, the Bahamas and Canada. We then assessed patterns of genetic structure within the entire native range of the species, and within each of the introduced populations. Finally, in order to elucidate the invasion history of *R. flavipes*, we compared support for different invasion scenarios modeling the number, size and origin of each introduction event using ABC-RF.
2 | MATERIALS AND METHODS

2.1 | Population sampling and sequencing

A total of 257 individuals of *R. flavipes* were collected from 29 populations spanning both native (USA) and different introduced populations in Europe (*i.e.*, France, Germany), North America (Canada and Bahamas) and South America (Chile and Uruguay) (Figure 1; Detailed sampling is provided in Table S1). In addition, 19 individuals of the sister species *R. virginicus* were collected to serve as an outgroup for the phylogenetic analysis. Samples were stored in 96% ethanol at 4°C until DNA extraction. Total genomic DNA was extracted from each individual using a modified Gentra Puregene extraction method (Gentra Systems, Inc. Minneapolis, MN, USA). DNA quality was assessed by agarose gel electrophoresis and DNA concentration was measured with Qubit® 2.0 Fluorometer (Invitrogen, USA). Non-degraded genomic DNA (100-300ng) was used to construct ddRAD libraries. Libraries were prepared and sequenced at the Texas A&M AgriLife Genomics and Bioinformatics Service facility using SphI and EcoRI restriction enzymes following the protocol of Peterson et al. (2012). Each sample was identified using unique combinatorial barcodes of 6 and 8 base pairs. Samples were amplified through PCR with iProof™ High-Fidelity DNA Polymerase (Bio-Rad). PCR products were purified using AMPure XP beads (Beckman Coulter Inc.). Libraries were size-selected to a range of 300–500 bp using the BluePippin system (Sage Science Inc.). Libraries were sequenced on six flowcell lanes using an Illumina HiSeq 2500 (Illumina Inc., USA) to generate 150 bp paired-end reads.

The paired-end reads were checked for quality control using FastQC v0.11.8 (Andrews 2010). Forward and reverse reads were demultiplexed from their barcodes, assigned to each sample and assembled using Stacks v.2.41 (Rochette et al. 2019). Reads were first aligned to the *R. flavipes* reference genome (Zhou et al. unpublished data, Table S2) using the Burrows-Wheeler Aligner (Li and Durbin 2009). Aligned reads were then run through the reference-based pipeline of Stacks, which built and genotyped the paired-end data, as well as called SNPs using the population-wide data per locus. Only SNPs present in at least 70% of individuals in half of the
populations were kept for downstream analyses. Furthermore, SNPs with mean coverage lower than 5x and higher than 200x were removed using VCFtools v.0.1.15 (Danecek et al. 2011), to prevent unlikely SNPs and highly repetitive regions. Low frequency alleles (< 0.05) and highly heterozygous loci (> 0.7) were sorted out, as they likely represent sequencing errors and paralogs (Benestan et al. 2016). A single random SNP was kept for each locus, to prevent linkage disequilibrium that may potentially affect population structure and phylogenetic analyses. The dataset was formatted for downstream software programs using PGDSpider v.2.1.1.5 (Lischer and Excoffier 2011).

### 2.2 Population structure and phylogenetic relationship

Expected ($H_E$) and observed ($H_O$) heterozygosity, inbreeding coefficient ($F_{is}$), and population differentiation values ($F_{ST}$) were calculated using Stacks (Rochette et al. 2019). Population structure among the 23 native and six introduced populations was analyzed using three complementary approaches.

First, the most likely number of genetic clusters (i.e., $K$) in the dataset was estimated, and individuals were assigned into each of them using fastSTRUCTURE v1.040 (Raj et al. 2014). The algorithm ran following an admixture model with allele frequencies correlated and did not use a priori information on localities. The algorithm was parallelized and automated using Structure_threader (Pina-Martins et al. 2017), and ran for $K$ ranging from one to 29. The chooseK.py function was used to select the most likely number of genetic clusters. Plots were created by Distruct v2.3 (Chhatre 2019) (available at [http://distruct2.popgen.org](http://distruct2.popgen.org)).

Second, genetic clustering was estimated using a principal component analysis (PCA) and a discriminant analysis of principal components (DAPC). DAPC uses discriminant functions that maximize variance among groups while minimizing variance within groups (Jombart et al. 2010). The most likely number of genetic groups was first inferred by the find.clusters algorithm
on the principal component analysis (PCA) outputs, with the Bayesian information criterion utilized
to select the number of genetic groups. The optimal number of principal components to inform
the DAPC (i.e., maximizing discriminatory power between groups, while preventing overfitting)
was then defined using the function optim.a.score. Both the PCA and DAPC were performed in
R (R Core Team 2020) using the adegenet package (Jombart 2008).

Third, population structure was visualized using the relatedness matrix produced by the
RADpainter and fineRADstructure software (Malinsky et al. 2018). This method calculates co-
ancestry between samples as an independent assessment of population structure. Analyses ran
using default parameters of 100,000 burn-in and 100,000 MCMC iterations, and results were
visualized in R through scripts provided with the program (available at
http://cichlid.gurdon.cam.ac.uk/fineRAD structure.html).

Phylogenetic relationships among R. flavipes individuals were inferred using maximum
likelihood (ML) analysis implemented in RAxML v8.2.12 (Stamatakis 2014). Phylogenetic
relationships were also estimated using a Bayesian analysis (Figure S5). After filtering, only 16
out of the 19 individuals of R. virginicus were used as an outgroup; these R. virginicus samples
were not used in any other analyses. An acquisition bias correction was applied to the likelihood
calculations, removing invariant sites from the alignment through the Phrynomics R script
(available at https://github.com/bbanbury/phrynomics/). The rapid bootstrap analysis and search
for the best-scoring maximum likelihood tree was performed using the extended majority rule
(MRE)-based bootstopping criterion (Pattengale et al. 2010) under the GTR+G nucleotide
substitution model.

2.3 | Assessing the invasion history

The global invasion history of R. flavipes was inferred through ABC analyses by comparing
support for different invasion scenarios. The scenarios varied according to the origin(s) of
introduced populations, the founding population size, the bottleneck duration and the admixture rate if multiple sources were detected. To reduce computational effort, model selection and parameter estimation were performed using the recently developed random forests (RF) machine learning method (ABC-RF) available in the abcrf R package (Pudlo et al. 2015, Raynal et al. 2018). This method requires a reduced number of simulated datasets while still providing robust posterior estimates. To reduce computational effort, we also only tested scenarios relevant to biological and historical data; for example, we did not consider that the Chilean and Canadian introduced populations could be the source of the French population. A step-by-step approach (9 different steps divided into 4 parts; fully explained in Supplementary Information 1) was used to infer the different episodes of the invasion history of *R. flavipes*, as this type of approach is commonly performed in ABC studies to distribute the computational effort (Fraimout et al. 2017, Javal et al. 2019, Ryan et al. 2019). The introduced populations in Germany, Uruguay and the Bahamas were not used in ABC computations as they were represented by too few individuals. Briefly, the first part estimated whether each introduced population (*i.e.*, France, Canada and Chile) arose from independent or bridgehead introduction events (Part A). As this first part indicated that the French population may have played a role in the introductions to Canada and Chile, we first sought to decipher the source(s) of the introductions to France alone (Part B). Next, we attempted to identify the sources of the Canadian (Part C) and Chilean (Part D) populations using France as a potential source. For all scenarios tested, introduction events were followed by a decrease in effective population sizes that varied from one to 100 migrants for a duration of zero to 50 years. Divergence time is given in generations, with a generation length of one year. Posterior distributions of preliminary simulated data sets were used to adjust the range of other priors as wide as possible while retaining biological meaning. For each step, 10,000 simulated datasets, including all of the summary statistics implemented in DIYABC v.2.1.0 (Cornuet et al. 2014), were generated per scenario from 2,000 randomly sampled SNPs. Priors were set uniform for all model parameters and selected based on historical records. Simulated datasets were first
generated by DIYABC, and later exported for model selection and parameter estimation in ABC-RF. The different scenarios tested within each step are provided in the Supplementary Information.
3 | RESULTS

The 257 *R. flavipes* samples yielded an average of 7.0 million paired reads per individual (range: 0.03 – 23.5). Twenty-eight individuals were removed due to a significant amount of missing data (≥ 60%) or low coverage (≤ 9.5×). After filtering, the final dataset contained 229 individuals of *R. flavipes* from 29 populations and included 51,116 SNPs, with an average coverage of 27× and 32% missing data. Weak inbreeding was found within *R. flavipes* populations ($F_{IS} \pm SE = -0.053 \pm 0.031$). Consequently, values of observed heterozygosity ($H_o \pm SE = 0.196 \pm 0.031$) were higher than values of expected heterozygosity ($H_e \pm SE = 0.135 \pm 0.020$; values for all populations are provided in Table S3).

3.1 | Population structure

Strong genetic structure was uncovered among the *R. flavipes* individuals from fastSTRUCTURE, with $K = 4$ best explaining the structure in the data (Figure 1). At this value of $K$, more than half of the individuals in the dataset (57.2%) were clearly assigned to one of the four clusters (assignment probability higher than 99%; 73.3% of individuals were assigned to a unique cluster probability higher than 80%). However, the strong genetic structure uncovered among individuals in the native range was inconsistent with their geographic origin, as neighboring samples often exhibited completely different assignment profiles (Figure 1). This pattern was also found when populations from the native range were analyzed separately (Figure S1). In the French introduced range, most samples could be assigned to the same cluster, although some samples from the Paris region had a mixed assignment; a similar mixed assignment was found for the lone German sample. A comparable pattern was observed in the Chilean introduced range, with most samples displaying fixed assignments and only a few with mixed assignments. Although a single genetic group was mostly found within each introduced population (France, Chile and Canada), the three introduced
populations were separately assigned to three different genetic groups and did not segregate into a single ‘introduced’ cluster; a finding also uncovered at lower values of K (Figure S2). Because the genetic clustering of the native range did not consistently align with geographic origin, inferring a source population for each introduced population becomes difficult. For example, most samples from Chile were assigned to the same cluster as samples from New York, Wisconsin and Texas. Similarly, although the introduced population in France shared its strongest tie to the native range with Arkansas, France also had ties with Louisiana, Missouri and even one sample in South Carolina. The origin of the samples in Canada was even more complicated, as the genetic cluster present in this population was spread across most native localities. Similar findings were uncovered for different values of K (Figure S2).

Similar results to that of fastSTRUCTURE were uncovered using the PCA and DAPC approach (Figure 2). The PCA indicated strong differentiation across the *R. flavipes* samples, as they broadly segregated along the two axes. For most localities, genetic clustering was not associated with geography, as samples from a given locality did not always cluster together. Likewise, low genetic similarity was observed between geographically neighboring localities. Interestingly, such a pattern was also found to a lesser extent in the introduced populations of France and Chile (only a single sample was available from Germany and Uruguay, and just two from the Bahamas). In France, most of the samples segregated together, except for six individuals from the Paris region, which clustered separately from the rest of the main population and had mixed assignments. A similar pattern was observed for the samples from Chile, with three samples clustering apart from the main Chilean population. Interestingly, fastSTRUCTURE found the occurrence of two and three genetic clusters in the Chilean and French populations respectively, when those populations were analyzed separately (Figure S3). The *find.clusters* algorithm found the best support for four genetic clusters in the dataset (Figure 2). Notably, the introduced localities of *R. flavipes* did not cluster together; instead, the different introduced
populations were spread across the four different DAPC clusters, with some even split between two clusters (Chile and France). Remarkably, a similar pattern was observed from localities within the native range, with samples from a given locality clustering into two (e.g., Texas, Mississippi, Wisconsin) or even three (Louisiana) distinct DAPC clusters.

The co-ancestry matrix highlighted similar patterns when clustering individuals based on their level of relatedness (Figure 3). Using fineRADstructure, all samples from a given locality were no more related to one another than they were to samples from another locality (Figure 3). This result is indicative of a weak geographic structure in the native range, as most localities were disjunct in the co-ancestry matrix. Notably, the same pattern was observed for the introduced populations, with clustering observed in two (Canada) or three (France and Chile) distinct co-ancestry groups. Accordingly, although significant, the mean genetic differentiation between populations was rather low (mean $F_{ST} \pm SE = 0.091 \pm 0.054$; pairwise $F_{ST}$ values between each pair of populations is provided in Figure S4).

### 3.2 | Phylogenetic relationship

The ML phylogeny was constructed on 29,875 SNPs after filtering out invariant sites, using 650 bootstrap replicates, as suggested by the MRE-based bootstopping-criterion. Overall, the tree was consistent with results from the clustering analyses, despite weak bootstrap support throughout the topology (Figure 4). Interestingly, the entire introduced range did not fall out as a single clade; instead, introduced populations arose throughout different branches of the tree. Furthermore, all invasive populations fall out as at least two (Canada and Bahamas) or more different clades (France and Chile). This result also suggests that different introduced populations arose from separate introduction events out of the native range, and that there were several
introduction events in most invasive populations (similar findings were found when Bayesian
inferences were used to build the tree; Figure S5).

### 3.3 | Invasion history

Part A of the ABC analysis found that introduced populations in Canada and Chile most likely
originated, at least partially, from bridgehead introductions from the previously introduced
population in France (Figure 5), rather than directly from the native range (Supplementary
Information 1). The RF votes were mostly split between three scenarios describing a bridgehead
introduction from France to either Canada (220 RF votes), Chile then Canada (221 RF votes) or
both countries (215 RF votes).

When analyzing the introduced French population alone in part B, the first step found that
this introduced population could not be unambiguously assigned to a single origin, as all three
regions of the native range received a substantial amount of support (*Louisiana/Mississippi*: 417
RF votes, *East*: 414 RF votes and *Central*: 169 RF votes). The ‘least bad’ single introduction event
scenario (151 RF votes) was outvoted when compared against a two-population admixture
scenario (319 RF votes, second step); and this two-population admixture scenario (271 RF votes)
was itself outvoted by scenarios simulating the contemporary French population arising through
admixture of all three native regions (394 RF votes, third step). When groups of scenarios were
compared, the group of scenarios with admixture outvoted the group without admixture in the
second step (660 against 340 RF votes); and the group of scenarios with three-population
admixture outvoted the group with two-population admixture in the third step (612 against 388 RF
votes). The fourth step of part B (further dividing the native range) found that Massachusetts,
Maryland and New York (222 RF votes) obtained the highest support as the origin for the French
population. However, several other source populations obtained a significant number of RF votes,
casting doubt on the ability to undeniably assign the introduced population of France to a unique source. This ambiguity is further emphasized when the scenarios were divided into groups, as both the Southeastern region (504 RF votes) and the rest of the native range (496 RF votes) obtained an almost identical number of RF votes. Overall, these findings suggest the occurrence of multiple introduction events out of the native range. However, at both large (step1) and finer scales (step4), no scenario received a majority vote, preventing a definitive determination of the source for the introduced population in France and calling for caution in the appraisal of the estimated parameters.

Part C aimed at analyzing the origins of the Canadian introduced population, using the French introduced population as a potential source. ABC-RF analyses revealed that the most probable scenario for the origin of the Canadian population was an introduction from a French bridgehead and its admixture with a separate introduction event from the native range (405 RF votes), rather than originating entirely from the native range (227 RF votes) or a French bridgehead (368 RF votes). The presence of a French bridgehead is also supported, as the group of scenarios including a bridgehead event (623 RF votes) outvoted the group without a bridgehead event (377 RF votes). When the native range was further divided, ABC-RF analyses also failed to confidently link the origin of the Canadian introduced population to a unique geographic region, as several source populations obtained a significant number of RF votes. This doubt is also emphasized when groups of scenarios were compared, as both the Southeastern region (520 RF votes) and the rest of the native range (480 RF votes) obtained a similar number of RF votes.

A similar invasion history was identified for Chile in part D, as a bridgehead from France combined with an additional introduction event from the native range was found most likely (506 RF votes), rather than entirely from the native range (339 RF votes) or a French bridgehead (155 RF votes). Similar to the origin of the introduced populations of France and Canada, ABC-RF did
not confidently infer the source of the Chilean population when the native range was divided.

Several source populations obtained a similar number of RF votes when each scenario was analyzed separately, and the Southeastern group of scenarios (539 RF votes) obtained a similar number of RF votes to the group that included the rest of the native range (461 RF votes). Overall, the ABC results cast doubt on the ability to connect each introduced population to one or a few specific source populations, as most simulated scenarios poorly fit the observed dataset with no scenario receiving a clear majority of the votes. This finding is also suggested by the divergence between the simulated and observed datasets present in the LDA graphs, potentially highlighting that more sophisticated scenarios are needed to better explain the data. Although all of the posterior probabilities and posterior parameter estimates are provided in the Supplementary Information for the 'least bad' scenario in each step, we call for caution in interpreting those values given the ambiguous results obtained in most steps.
Our study provides insights into the invasion history of the termite *Reticulitermes flavipes*, highlighting frequent and recent human-mediated jump dispersal in both the native and introduced range of this species. We first revealed strong genetic structure among individuals within the native range of this species with individuals grouping into four distinct clusters. Yet, these clusters were not strictly associated with geography, as highly different individuals were found in the same locality and highly similar ones in localities separated by several thousand kilometers. This finding indicates extensive movement of colonies throughout the native range, likely through human transportation of wood. We also highlight a complex invasion history with multiple introduction events out of the native range and bridgehead spread from the introduced population in France. The apparent genetic shuffling within the native range limits our ability to assign an exact source population(s) for the different introduced ranges. However, similar to the effect of multiple introductions into the invasive range, admixture in the native range prior to invasion can potentially favor invasion success by increasing the genetic diversity later conveyed to the introduced ranges.

Our findings revealed the occurrence of multiple introductions from different native localities serving as sources for the invasive ranges of France, Chile and Canada. Additionally, Canada and Chile received secondary invasions from the introduced population in France, which acted as a bridgehead. Some previous results indicated that there may have been several introductions into France (Perdereau et al. 2019). *Reticulitermes flavipes* was first reported in Europe (Austria) in 1837 and was first reported in France as *R. santonensis* in 1924 (Feytaud 1924), where it was widespread and therefore probably introduced much earlier (Bagnères et al. 1990). Despite being unable to definitively link its source population(s) to the New Orleans region as previously suggested (Perdereau et al. 2013, 2015), our data, based on a larger sample size and more informative markers, do not rule out this possibility. Our data instead suggest that
individuals genetically similar to this invasive population were found across the entire native range, from Louisiana to Maryland. However, it is possible that the French population originated from colonies originally from the New Orleans region that had been transported elsewhere within the native range, such as South Carolina or Arkansas. Such long-distance jump dispersal within the native range can therefore hamper the clear identification of the source population(s). Likewise, although our results suggest that the Canadian and Chilean introduced populations originated from admixture between the introduced population of France and native localities in the northern range of *R. flavipes*, these results suffer from low confidence, potentially due to genetic mixing between native localities. Although the French connections with Louisiana and eastern Canada are well-established, France also has historical ties with Chile. Notably, most of the human immigrants to Chile between the 18th and 20th centuries come from the Basque region of Southern France (Fernandez-Domingo 2006), where *R. flavipes* occurs. During the 18th century, Chile experienced massive immigration from this region, reaching 27% of the total Chilean colonial population. Overall, these findings indicate that jump dispersal may not be restricted to a single region within the native range of this species. Instead, such dispersal appears common in *R. flavipes* in both its native and invasive ranges, suggesting that this species possesses traits that promote its spread and have contributed to its global distribution.

The genetic patterns observed in the native range of *R. flavipes* may be explained by numerous and recent jump dispersal events across the native range, likely mediated via human trade and transportation. This finding exemplifies species spread by *stratified dispersal*, whereby individuals disperse at different spatial scales, from local to long-distance movement (Shigesada et al. 1995). Local scale dispersal relies on the biological dispersal ability of the species, ranging from short-range (*i.e.*, budding) to moderate dispersal (*i.e.*, nuptial flights). In contrast, long-distance dispersal is often human-mediated and therefore considered stochastic and difficult to identify. Notably, our study revealed both genetically distinct individuals inhabiting the same
locality and genetically similar individuals separated by several thousand kilometers. The geographic distance separating highly similar individuals far exceeds the biological dispersal ability of this species, which suggests that these individuals were artificially transported to a different locality. Additionally, the finding of genetically distinct individuals within the same or adjacent localities indicates a low level of mixing between those individuals. This may stem from reduced local dispersal, whereby transported individuals inbreed and do not disperse far from their landing point. A high proportion of new reproductives of *R. flavipes* do in fact couple with their nestmates during mating flights (25%); however, the proportion of inbred founders is significantly reduced among established colonies (DeHeer and Vargo 2006). Therefore, this inbreeding depression may select against the interbreeding of artificially transported colonies. Also, *R. flavipes* usually disperses through nuptial flights, which should enhance gene flow over large scales (Vargo 2003). Consequently, a scenario where transported individuals interbreed and do not disperse far from their landing point may not alone explain the pattern observed in this study. The finding of highly genetically different individuals within the same locality therefore suggests that some of the long-distance jump dispersal events are probably too recent to allow transported individuals to admix with local colonies and homogenize the gene pool within populations.

The global spread of invasive species is strongly influenced by long-distance jump dispersal events, even once established within an introduced range (Suarez et al. 2001). Jump dispersal events are more effective, and often required, for species to rapidly reach a widespread distribution (Gippet et al. 2019, Bertelsmeier 2021). For example, the worldwide distribution of the Argentine ant has been shown to primarily stem from human-mediated jump dispersal, rather than from its classical spread through colony budding, as the latter would have to be three orders of magnitude higher to explain its actual distribution (Suarez et al. 2001). This finding is also typified in the global distribution of the red imported fire ant *Solenopsis invicta*, which utilized long-range
jump dispersal to first invade the southeastern USA, and subsequently Asia and Australia from this USA bridgehead (Ascunce et al. 2011). In general, human-mediated jump dispersal appears common in eusocial invaders with a global distribution, like ants (Bertelsmeier et al. 2017, 2018) and termites (Buczkowski and Bertelsmeier 2017, Blumenfeld and Vargo 2020). These multiple long-distance movements are also observed among regions within invasive ranges, across a wide variety of taxa, such as the aforementioned *S. invicta* throughout the southern USA (Lofgren 1986) and China (Ascunce et al. 2011), the western mosquitofish *Gambusia affinis* in New Zealand (Purcell and Stockwell 2015), and plants in China (Horvitz et al. 2017). Many studies have demonstrated the role of human-mediated jump dispersal in shaping invasive distributions and genetic diversity. However, it often remains unclear whether long-distance dispersal pre-exists in the native range of invasive species, and whether it plays a role in determining the pattern of genetic diversity observed at the global scale of these species.

Native ranges of many invasive species often remain geographically structured (Voisin et al. 2005, Beck et al. 2008, Leinonen et al. 2008, Verhoeven et al. 2011). For example, native populations of *S. invicta* are strongly geographically differentiated (Ross et al. 2007). Though rare long-distance dispersal of *S. invicta* has been reported (Ahrens et al. 2005), these events occurred far in the past and have been attributed to strong winds during nuptial flights or the rafting of entire colonies during flooding events (Hölldobler and Wilson 1990), rather than from human-mediated transport (Ahrens et al. 2005). Native populations of another termite invader *Coptotermes formosanus* in China are highly structured, with distinct native populations representing different genetic clusters (Blumenfeld et al. 2021). This structuring suggests reduced gene flow across populations, and therefore a limited number of human-mediated dispersal events within the native range of this species. Our results stand in sharp contrast with the strong population structure commonly uncovered within the native ranges of invasive species, as frequent jump dispersal appears to have occurred in the native range of *R. flavipes*. 
Understanding the factors driving the differences between *C. formosanus* and *R. flavipes* may shed light on key evolutionary mechanisms underlying their invasion success. Furthermore, while most studies focus on unraveling invasion pathways out of a native range, our results stress the need to consider evolutionary processes and human-mediated dispersal that may already be present within the native range of an invasive species, as these can affect the level and distribution of genetic diversity in both the native and invasive ranges.

Extensive human-mediated jump dispersal has been reported in the native range of a few species. For example, in the invasive tree *Acacia pycnantha*, extensive transport and replanting throughout its native Australian range prior to its introduction to South Africa resulted in highly admixed genotypes already present in the native range. This feature has consequently prevented an accurate identification of the native source population(s), as highly admixed genotypes and comparable genetic diversity were present in both ranges of the species (Le Roux et al. 2013). A similar pattern has been found in the North American rangeland weed, *Centaurea diffusa*, where an extremely low level of population structure in the native range hindered the assignment of its introduced population to its likely native source location (Marrs et al. 2008). However, the genetic patterns observed in the *Acacia* and *Centaurea* plants are slightly different than the one observed in *R. flavipes*, as the inability to pinpoint the origins of invasive populations of these plants stems from the near-panmixia found across the native range (Marrs et al. 2008, Le Roux et al. 2013). Therefore, the patterns in these other species most likely stem from an ancient and continuous genetic shuffling throughout the native range. In contrast, the lack of geographic structure despite highly genetically different individuals indicates recent and stochastic long-distance dispersal in *R. flavipes*. Consequently, the genetic structure of *R. flavipes* may have been different (with less admixture) in both the native and invasive range(s) a few centuries ago, at the beginning of the French, Canadian and Chilean invasions. The complex genetic structure currently observed,
together with multiple introduction events, makes it difficult to accurately reconstruct the invasion history of this species.

The invasion success of termites is tightly linked with their ability to eat wood, nest in wood and cultivated plants, and readily generate secondary reproductives, as all 28 species of invasive termites share these three traits (Evans et al. 2013, Eyer and Vargo 2021). These traits may enhance the frequency of human-mediated dispersal because any piece of wood serving as a nest or foraging site has the potential to become a viable propagule (Evans et al. 2010, Evans et al. 2013). However, these traits are common in lower termites like R. flavipes and C. formosanus, therefore their occurrence in both species cannot explain why R. flavipes has experienced a greater frequency of long-distance dispersal than C. formosanus. In R. flavipes, repeated human-mediated dispersal could reflect a higher degree of propagule pressure from different USA regions, representing multiple hubs of intense human activity and timber production. Forests and timber production are unequally distributed across the eastern USA (Brown et al. 1999, Howard and Liang 2019), and may therefore require significant wood transportation throughout this part of the country from high to low timber-producing regions. Similarly, the frequency of human-mediated dispersal may reflect the connectivity between native regions. In the introduced population of R. flavipes in France, the distribution of genetic diversity is associated with the construction of the railway network and stations, highlighting its possible role in spreading termites over long distances (Andrieu et al. 2017, Suppo et al. 2018, Perdereau et al. 2019). In the USA, about 14,000km of track were active by 1850, mainly in the eastern USA (141,000km in 1880 and over 400,000km in 1916) (United States Census Bureau 1890, Chandler 1965). In contrast, the first 10km railway was built in China in 1881, but less than 13,000km were in use by 1948 for the whole country. This difference in connectivity may explain the numerous long-distance dispersal events in the native range of R. flavipes and their absence in the Chinese native range of C. formosanus. Interestingly, the USA railroad network has been suggested to represent a major
dispersal mode for the invasive population of *C. formosanus* (Austin et al. 2008). Overall, many invasive social insect species originate from South America or East Asia (Tsutsui et al. 2000, Heller 2004, Ross et al. 2007, Eyer et al. 2018a, Eyer et al. 2018b, Eyer et al. 2020). The population structure observed in most native populations of invasive termites may simply reflect the reduced connectivity between native regions in these areas, potentially resulting from a lack of internal trade among regions or difficulty in reaching isolated geographic areas. Our findings in *R. flavipes* indicate that frequent long-distance dispersal may already be present within the native ranges of some invasive species, especially those originating from regions with a long history of dense transport networks.

Native populations of many invasive species often remain geographically isolated and locally adapted (Verhoeven et al. 2011). In the introduced range, a temporary loss of local adaptation through admixture has been suggested to alter the fitness consequences of admixture in recent invaders (Verhoeven et al. 2011). In our study, the levels of admixture observed in the introduced populations of France and Chile may be explained by numerous introductions from distinct source populations and their interbreeding within the invasive range. However, we cannot rule out the possibility that populations were already admixed before propagules were transported worldwide. Similarly, it is possible that admixed introduced populations re-invaded the native range of *R. flavipes*. In the native range of species, long-distance dispersal enhances gene flow between distant populations that are otherwise isolated. Similar to post-introduction increases of genetic diversity through multiple introduction events (Kolbe et al. 2004, Stenoien et al. 2005, García et al. 2017), admixture between native populations prior to an introduction event may enhance the amount of genetic diversity brought to the invasive range. Admixture may improve invasion success through recombination of distinct genotypes, potentially creating novel combinations of traits, and/or increasing the level of genetic diversity upon which natural selection can act. Pre- or post-introduction admixture may also relax the inbreeding load by reducing the
expression of recessive deleterious alleles or lead to heterosis effects, potentially improving the establishment and early success of invasive species (Ellstrand and Schierenbeck 2000, Drake 2006, Keller and Taylor 2008, Hahn and Rieseberg 2016). Overall, increased genetic diversity via admixture may favor subsequent introductions given the novel selection pressures invasive species face in their new environments (Verhoeven et al. 2011).

Conclusion

In this study, we infer the occurrence of long-distance jump dispersal in the native range of the termite *R. flavipes*. This long-distance dispersal may facilitate admixture between populations that are otherwise isolated. Admixture in native populations prior to introduction may favor invasion success by increasing the amount of genetic diversity brought to the introduced range, achieving an effect similar to that produced by multiple introductions from the native range. However, pre-introduction admixture may not be as common as multiple introduction scenarios (*i.e.*, post-introduction admixture), because the benefits of admixture in the novel environment of the invasive range are probably higher, and the costs smaller (Rius and Darling 2014). As native populations are locally adapted, long-distance dispersal and admixture may disturb this local adaptation, thereby reducing population fitness (Verhoeven et al. 2011, Palacio-Lopez et al. 2017). In contrast, populations in invaded ranges are generally too recent to be locally adapted (but see Batz et al. 2020). This lack of local adaptation may release introduced populations from maintaining specific locale-selected allelic combinations, and thereby fully benefit from admixture in early stages of the invasion. The relative roles of pre- and post-introduction admixture in biological invasions should be fertile ground for future studies.
ACKNOWLEDGEMENTS

We wish to thank Renato Ripa (Instituto de Investigaciones Agropecuarias, Chile), Ariel Camousseight (National History Museum of Science from Santiago, Chile), James Austin (BASF, Research Triangle Park, NC, USA), Uwe Noldt (Germany) and Guillaume Baudouin (Paris City Hall, France) for giving us samples. This work was supported by USDA-NRICGP grants to ELV (00-35302-9377 and 2002-35352-12490), the Urban Entomology Endowment at Texas A&M University and CNRS grants and contracts to AGB.

DATA ACCESSIBILITY

The data reported in this study will be deposited in the Open Science Framework database upon acceptance, https://osf.io. Raw sequence files are deposited at the National Center for Biotechnology Information under BioProject accession number PRJNA667438.

AUTHOR CONTRIBUTIONS

ELV and AGB designed the study. EP, SD, FD, AGB and ELV collected the data. LNLJ performed the molecular analyses. AJB, PAE, LNLJ and PTS analyzed the data. PAE and AJB wrote the paper with contributions of ELV.

SUPPLEMENTARY MATERIAL

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

Table S1: Sampling information for every individual used in this study.
Table S2: Number of samples, observed and expected heterozygosity, FIS and nucleotide diversity for each population.

Table S3: Number of samples, observed and expected heterozygosity, FIS and nucleotide diversity for each population.

Figure S1: fastSTRUCTURE assignment for individuals from native populations of *R. flavipes* for K ranging from 2 to 4.

Figure S2: fastSTRUCTURE assignment for each individual of *R. flavipes* for K ranging from 2 to 5.

Figure S3: fastSTRUCTURE assignment for individuals from invasive populations of France and Chile.

Figure S4: Pairwise FST matrix between each pair of populations of *R. flavipes*.

Figure S5: Bayesian inferences tree of *Reticulitermes flavipes*. Native populations are highlighted in grey; introduced populations are highlighted in reddish colors.

Supplementary Information 1: A detailed description of the step-by-step ABC RF analysis is covered, including: 1) priors used for the analyses, 2) graphical representation and random forest votes for each scenario within each step, 3) an overall PCA of the simulated datasets for every scenario for each step and 4) parameter estimates for the final invasion model.
FIGURE LEGENDS

Figure 1: Sampling map and fastSTRUCTURE assignment for each individual of *R. flavipes* for K = 4. Each vertical bar represents an individual and each color represents a distinct genetic cluster. Individual fastSTRUCTURE assignments are geographically located in the native and introduced ranges of *R. flavipes*.

Figure 2: Principal Component Analysis (PCA) of *Reticulitermes flavipes* individuals. Each circle represents an individual. Each individual is colored according to its population of origin; introduced populations are depicted in reddish colors, native populations are colored in grey. Individuals are grouped according the Discriminant Analysis of Principal Components (DAPC) with best support for K = 4 genetic clusters.

Figure 3: Co-ancestry matrix between each pair of individuals inferred using fineRADstructure. Each pixel represents a pair of individuals. Co-ancestry coefficients between two individuals are designated on a color spectrum. Low values are shown in yellow; higher values are shown in darker colors.

Figure 4: Maximum likelihood phylogenetic tree of *Reticulitermes flavipes* individuals from RAxML. Individuals are colored according to their fastSTRUCTURE assignments (K = 4). Samples from the introduced ranges are highlighted with an emphasized tip. The phylogenetic tree is rooted with 16 *R. virginicus* samples.

Figure 5: Graphical representation of the invasion pathway of *Reticulitermes flavipes* out of the eastern USA inferred through ABC RF in France, Canada and Chile. The estimated time of introduction and rate of admixture is provided for each introduction event. The large 95%CI however calls for caution in interpreting those values. All scenarios tested and results for each ABC step, as well as all of the posterior parameter estimates, are provided in the Supplementary Information.
REFERENCES


Chhatre, V. E. 2019. Distruct v2.3, a modified cluster membership plotting script.


United States Census Bureau. 1890. Report on transportation business in the United States at the eleventh census


* bootstrap value above 0.95; * value below 0.95; all other values are 1.00
Canada

Chile

Native range

~210 years ago
60 - 365
45%
6 - 90%

~280 years ago
53 - 389
56%
7 - 93%

~180 years ago
91 - 398
7 - 92%

France