High-Resolution Population Genetic Structure of Tawny Crazy Ant (Nylanderia fulva Mayr: Hymenoptera: Formicidae) from the Origin in South America and Introduced Regions of the United States

Jocelyn R. Holt (Jocelyn.Holt@rice.edu)
Rice University

James Montoya Lerma
Universidad del Valle

Luis A. Calcaterra
CONICET and Fundación para el Estudio de Especies Invasivas

Tyler J. Raszick
Texas A&M University

Raul F. Medina
Texas A&M University

Research Article

Keywords: genetic differentiation, high throughput sequencing, SNPs, invasive insect

Posted Date: January 4th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2399319/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Background

The tawny crazy ant (*Nylanderia fulva* Mayr) is native to South America and was first reported in the continental United States (US) in 1938. It was not until the 1990s in Florida and 2000s in Texas that this ant was considered a serious pest in the US. Tawny crazy ant (TCA) is currently considered an invasive pest in six US states and this ant’s invasion success is attributed in part to a unicolonial nature, multiple queens per nest, natural enemies release in the invasive range, and ability to detoxify venom from other competitor ant species. A limited number of low-density molecular markers have previously shown little genetic differentiation among TCA populations across their geographic distribution in the US.

Results

Using High Throughput Sequencing (HTS) we obtained high-density molecular markers (i.e., SNPs) for TCA samples. With 26,657 SNPs we identified genetic variation among TCA populations in different states across the US (i.e., Texas, Louisiana, Alabama, Mississippi, Georgia, and Florida) and in South America (i.e., Argentina, Colombia, and Peru).

Conclusion

Our results underscore that for recently introduced invasive species, increasing the number of molecular markers used in population genetic studies can provide greater resolution. High-resolution information on regional genetic differences can help inform pest management strategies.

Background

Genetic differentiation among insects that are recently introduced may be difficult to detect due to a single introduction event that results in a genetic bottleneck. In contrast, multiple introductions of genetically differentiated propagules can facilitate detection of population structure. For instance, genetic variability among ambrosia beetle (*Xylosandrus crassiusculus* Motschulsky) populations was due to multiple introduction events and the identification of cryptic species [1]. Similarly, invasive populations, particularly those that are small, are likely to experience genetic drift over the course of an invasion event. Microevolutionary forces, such as selection and genetic drift, can result in genetic differentiation in as quickly as 10 to 50 years [2, 3]. Geographically widespread and relatively recent introductions of invasive insect species (i.e., 50 years or less) provide an opportunity to characterize the degree of genetic differentiation among different geographic locations. This information can then be used to increase monitoring at potential points of entry, determine centers of origin, and determine whether genetically differentiated invasive pest populations vary in traits relevant to their control [4]. In this study, we characterized the population genetic structure of tawny crazy ant (*Nylanderia fulva* Mayr) in the invaded region of the US and the native origin in South America.

Invasive tawny crazy ant (*Nylanderia fulva* Mayr; hereafter referred to as TCA) was first identified in Texas in 1938 and later in Florida in 1953 [5, 6]. This species has a native distribution along the Rio de La Plata basin and the southern portion of the Atlantic Coastal Forest biome [7]. TCA is considered a pest species in urban, agricultural, and natural/wildland areas in introduced ranges of South America and the US [8–10]. After the initial reports and identification of TCA in the US, little to no information was recorded about further range expansion or its impacts on urban and natural settings, suggesting that these first reported propagules may have failed to establish. TCA was later reported in large numbers in Florida hospitals in the 1990s, and subsequently during a 2002 outbreak in Houston at NASA, and the greater Houston metro area [11, 12]. After the 1990s, TCA expanded their geographic distribution and, in addition to Texas and Florida, are now reported in Alabama, Georgia, Louisiana, and Mississippi [12]. TCA range expansion was likely facilitated by this ant’s ability to detoxify fire ant venom and displace other ant species from habitats [10, 13], in addition to dietary flexibility [14]. Because this ant is a relatively recent invader, microevolutionary forces may not have had enough time to leave a strong signature of genetic differentiation.

Previous population genetic studies of TCA used low-density molecular markers (sometimes called diagnostic markers; e.g., COI, EF1α-F1, EF1α-F2, CAD, argK, and microsatellites) and were unable to detect genetic differentiation among TCA populations in the US. The most recent such study used COI and 13 microsatellite markers that characterized TCA in the US as unicolonial [15], and failed to detect differences in population genetic structure from five different states (i.e., Texas, Louisiana, Mississippi, Georgia, and Florida). In contrast, within their native range in South America, significant population genetic structure was detected [15, 16].

The choice of molecular markers used to detect genetic variation in invasive insects should be informed by the degree to which populations in the invaded regions are reproductively isolated from the center of origin and by the duration of time since isolation occurred. The combination of reduced gene flow and decreased effective population sizes, needs to be taken into consideration when characterizing species’ population genetic structure [17]. The use of fewer (i.e., low-density) molecular markers that are designed for barcoding or diagnostic identification (e.g., COI, ITS) is sufficient for detecting greater genetic differences among species or populations that have a long history of reproductive isolation [18–20]. In instances where there is moderate reproductive isolation among populations, the use of tens to hundreds of molecular markers (i.e., SSRs, ISSRs) is often sufficient to detect genetic variation. In instances with ongoing gene flow or recent reproductive isolation a higher number of molecular markers is needed [21]. Thus, population genetic studies of recently invaded species should use high-density molecular markers (e.g., AFLPs, SNPs) to detect genetic variation.

High-density markers are able to detect genetic structure among populations that may have low genetic diversity due to a recent invasion or ongoing gene flow [22–24]. There are many instances where the use of high-density molecular markers detected genetic differences among populations, where low-density molecular markers failed to detect variation [24–29]. For example, in native populations of geographically isolated American lobsters (*Homarus americanus* H.
Milne-Edwards), the use of microsatellite markers failed to detect population genetic structure between the northern and southern range, while the use of 8,144 SNPs (single nucleotide polymorphisms) allowed for the detection of significant genetic differences between these populations [30]. Similarly, analysis of 18,147 SNPs showed clear delineation of yellow fever mosquito (Aedes aegypti L.) populations in Asia and Australia when eight microsatellites failed to detect these genetic differences [31]. Using high-density molecular markers allows for the identification of genetic structure among populations with recent gene flow, and identification of biologically relevant genetic differences can be used to inform management practices as was the case for the American lobster [32].

In this study, we resolved the population genetic structure of TCA in the US and South America. We hypothesized that high-density molecular markers would allow for the detection of genetic differentiation among ants from different geographic locations. In addition, we hypothesized that analysis of high-density molecular markers might inform the potential population(s) of origin from the native range in Argentina.

Results

Data Filtering

Approximately 745 million reads were retained after using the process radtags filtering step in STACKS [33, 34]. There were between 4.4 to 12 million reads per sample. To minimize potential bias due to linkage (i.e., linked alleles), only one SNP per locus was used for data analysis. Most loci met Hardy Weinberg Equilibrium (HWE), with less than 5% not meeting HWE. This resulted in 96 individuals with 26,657 SNPs among nine geographic locations.

Population Genetic Analysis

When all TCA from South America and the US were analyzed, the number of putative populations estimated by \( \Delta K \) peaked at five with a second peak at 11; suggesting population structure (Fig. 1). Ants from the US and South America clustered in seven genetically differentiated populations according to their geographic origin (Fig. 2). When only samples from the US were analyzed, the \( \Delta K \) peaked at six (Supplemental Fig. 1) while the optimal BIC values were three and five (Supplemental Fig. 2). When samples from only Texas were analyzed \( \Delta K \) of five and seven was detected (Supplemental Fig. 3), however clustering was not strongly associated with geographic location (Supplemental Fig. 4). When only samples from the native range (i.e., Argentina) were analyzed, the \( \Delta K \) peaked at three (Supplemental Fig. 5).

Population genetic statistics for each collection location are provided in Table 1. The number of private alleles was highest for ants sampled from Argentina (2120) and Colombia (2294) while ants from Alabama in the US had the lowest number (50) of private alleles or alleles exclusive to a particular population (Table 1). There were several populations that had moderate levels of genetic differentiation (Table 2). Both Colombia and Peru had high \( F_{ST} \) values when compared with other populations (Table 2). The inbreeding coefficient was low for the sampled populations (Table 1). When only samples from Argentina were analyzed, they were grouped into three populations associated with geographic region (Table 4).

A DAPC showed that TCA in both South America and the US had genetically differentiated clusters based mainly upon geographic location (Fig. 3). Ants collected from Argentina clustered separately from those collected in Colombia and Peru. Within the US, ants collected from each state clustered separately; although ants collected from Louisiana and Mississippi clustered closest to ants from Texas, which was also supported by low \( F_{ST} \) values among these locations indicating little genetic differentiation among these locations (Fig. 3).

- An AMOVA on all collection locations for TCA showed that there was no significant variation among populations despite some moderate to high \( F_{ST} \) values among collection locations. The components of variance broken into variation among locations at 0.87% (DF = 7, phi = 0.16, \( P < 0.001 \)) of the variability, while variation within collection locations accounted for 15.78% (DF = 88, phi = 0.15, \( P > 0.01 \)) of the variability, and variation among all individual samples was 84.7% (DF = 96, phi = -0.0048, \( P > 0.5 \)). When only samples from the US were evaluated with an AMOVA, there was a lack of significant variation detected among locations (Supplemental Table 1). Similarly with samples from Argentina, although three clusters were identified using an AMOVA, these did not have significant variation detected among locations (Supplemental Table 2). This lack of significance with an AMOVA is potentially due to the low sample size and a low number of distinct populations [35]. In contrast, pairwise \( F_{ST} \) values revealed geographic structure for populations in the US (Table 3).
- Calculation of IBD with Genepop did not detect geographic distance as a significant driver of genetic differentiation when all samples were analyzed (\( P = 0.09400 \)), when only samples from the US were analyzed (\( P = 0.96500 \)), or when only samples from South America were analyzed (\( P = 0.83410 \)). These results suggest other factors are driving genetic differentiation. A linear regression of the IBD pairwise comparisons indicated \( F_{ST} \) values were not significantly correlated with distance for all samples as this explained only 0.005% variation. When samples were analyzed separately for North and South America samples, distance explained 36.01% variation for US samples, and 70.76% variation for South American samples, as indicated by the \( R^2 \) value. When samples from Colombia and Peru were removed from the analysis, IBD was still not significant (\( P = 0.14500 \)).
Table 1

Number of TCA samples from US and International collection locations along with summary statistics. The inbreeding coefficient ($F_{IS}$) was negative or low except for all ants sampled. Among all positions (variant – SNP variation in a gene among individuals in a population resulting in different alleles and fixed – same SNP in a gene for all individuals in a population resulting in no allele variation): n = number of individuals, V = variant sites that are polymorphic in at least one collection location, % poly = percent polymorphic sites among variant positions: SNPs = polymorphic sites within a collection location, P = average frequency of the most common allele, Ho = observed heterozygosity, He = expected heterozygosity, $\pi$ = average nucleotide diversity, $F_{IS}$ = inbreeding coefficient.

<table>
<thead>
<tr>
<th>Collection Region</th>
<th>n</th>
<th>V</th>
<th>% poly</th>
<th>SNPs/Poly-morphic Sites</th>
<th>Private Alleles</th>
<th>P</th>
<th>Ho</th>
<th>He</th>
<th>$\pi$</th>
<th>$F_{IS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>US</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alabama</td>
<td>2</td>
<td>17301</td>
<td>4.1</td>
<td>5424</td>
<td>50</td>
<td>0.9164</td>
<td>0.1469</td>
<td>0.8861</td>
<td>0.1367</td>
<td>-0.0180</td>
</tr>
<tr>
<td>Florida</td>
<td>6</td>
<td>20240</td>
<td>7.6</td>
<td>9751</td>
<td>84</td>
<td>0.905</td>
<td>0.8383</td>
<td>0.1348</td>
<td>0.1443</td>
<td>-0.0379</td>
</tr>
<tr>
<td>Georgia</td>
<td>6</td>
<td>22935</td>
<td>7.4</td>
<td>10225</td>
<td>186</td>
<td>0.9018</td>
<td>0.1667</td>
<td>0.1384</td>
<td>0.1518</td>
<td>-0.0316</td>
</tr>
<tr>
<td>Louisiana</td>
<td>6</td>
<td>21836</td>
<td>7.9</td>
<td>10468</td>
<td>189</td>
<td>0.9021</td>
<td>0.1614</td>
<td>0.1392</td>
<td>0.1505</td>
<td>-0.0238</td>
</tr>
<tr>
<td>Mississippi</td>
<td>6</td>
<td>23137</td>
<td>7.7</td>
<td>10038</td>
<td>231</td>
<td>0.9032</td>
<td>0.1615</td>
<td>0.1350</td>
<td>0.1480</td>
<td>-0.0259</td>
</tr>
<tr>
<td>Texas</td>
<td>57</td>
<td>21872</td>
<td>11.3</td>
<td>14931</td>
<td>344</td>
<td>0.9016</td>
<td>0.1604</td>
<td>0.1439</td>
<td>0.1454</td>
<td>-0.0122</td>
</tr>
<tr>
<td>South America</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Argentina</td>
<td>10</td>
<td>21348</td>
<td>10.8</td>
<td>14123</td>
<td>2120</td>
<td>0.8563</td>
<td>0.1908</td>
<td>0.2010</td>
<td>0.2125</td>
<td>0.0549</td>
</tr>
<tr>
<td>Colombia</td>
<td>2</td>
<td>22169</td>
<td>2.65</td>
<td>3841</td>
<td>2294</td>
<td>0.9454</td>
<td>0.1023</td>
<td>0.0706</td>
<td>0.0942</td>
<td>-0.0122</td>
</tr>
<tr>
<td>Peru</td>
<td>1</td>
<td>17145</td>
<td>2.0</td>
<td>3024</td>
<td>209</td>
<td>0.9118</td>
<td>0.1764</td>
<td>0.1453</td>
<td>0.1764</td>
<td>0</td>
</tr>
<tr>
<td>Total Individuals</td>
<td>96</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2

Pairwise $F_{ST}$ values among each collection location for TCAs. Samples from South America show genetic differentiation from US samples. *** = very great ($\geq$ 0.26), ** = great (0.15 - 0.25), * = moderate (0.15 - 0.05), and + = little genetic differentiation ($\leq$ 0.05) (Hartl & Clark 1997).

<table>
<thead>
<tr>
<th>Collection Region</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Argentina</td>
<td>Colombia</td>
<td>Peru</td>
<td>Texas</td>
<td>Louisiana</td>
<td>Mississippi</td>
<td>Alabama</td>
<td>Georgia</td>
<td>Florida</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Argentina</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colombia</td>
<td>0.3176***</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peru</td>
<td>0.1441*</td>
<td>0.5723***</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Texas</td>
<td>0.0904*</td>
<td>0.3044***</td>
<td>0.1454*</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Louisiana</td>
<td>0.0983*</td>
<td>0.4328***</td>
<td>0.2482**</td>
<td>0.0083+</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mississippi</td>
<td>0.1011*</td>
<td>0.4576***</td>
<td>0.2697***</td>
<td>0.0098+</td>
<td>0.0441+</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alabama</td>
<td>0.0909*</td>
<td>0.5501***</td>
<td>0.3415***</td>
<td>0.0100+</td>
<td>0.0503+</td>
<td>0.0633+</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Georgia</td>
<td>0.0980*</td>
<td>0.4482***</td>
<td>0.2617***</td>
<td>0.0100+</td>
<td>0.0396+</td>
<td>0.0446+</td>
<td>0.0591**</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Florida</td>
<td>0.1012*</td>
<td>0.4291***</td>
<td>0.2376**</td>
<td>0.0102+</td>
<td>0.0348+</td>
<td>0.0446+</td>
<td>0.0470+</td>
<td>0.0412+</td>
<td>-</td>
</tr>
</tbody>
</table>
Table 3
Pairwise $F_{ST}$ values among TCAs in the US. Samples show genetic. *** = very great ($≥ 0.26$), ** = great ($0.15 - 0.25$), * = moderate ($0.15 - 0.05$), and + = little genetic differentiation ($≤ 0.05$) (Hartl & Clark 1997).

<table>
<thead>
<tr>
<th>Collection Region</th>
<th>1 Texas</th>
<th>2 Louisiana</th>
<th>3 Mississippi</th>
<th>4 Alabama</th>
<th>5 Georgia</th>
<th>6 Florida</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Texas</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Louisiana</td>
<td>0.1789**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 Mississippi</td>
<td>0.0611*</td>
<td>0.1785**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Alabama</td>
<td>0.0979*</td>
<td>0.1429*</td>
<td>0.0746*</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Georgia</td>
<td>0.0897*</td>
<td>0.1800**</td>
<td>0.0831*</td>
<td>0.0947*</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>6 Florida</td>
<td>0.0403+</td>
<td>0.1172*</td>
<td>0.0393*</td>
<td>0.0362+</td>
<td>0.0556*</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 4
Pairwise $F_{ST}$ values among TCAs in Argentina. Samples show genetic. *** = very great ($≥ 0.26$), ** = great ($0.15 - 0.25$), * = moderate ($0.15 - 0.05$), and + = little genetic differentiation ($≤ 0.05$) (Hartl & Clark 1997).

<table>
<thead>
<tr>
<th>Collection Region</th>
<th>1 Misiones</th>
<th>2 Corrientes &amp; Entre Ríos</th>
<th>3 Buenos Aires</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Misiones</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Corrientes &amp; Entre Ríos</td>
<td>0.1571**</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>3 Buenos Aires</td>
<td>0.1923**</td>
<td>0.2087**</td>
<td>-</td>
</tr>
</tbody>
</table>

Discussion

Our results show that invasive TCAs within the US belong to at least four genetically differentiated populations, and when substructure is considered, our results suggest six differentiated populations based upon geographic location. The high-resolution genetic variation among introduced TCAs was possible due to the increased power of using thousands of randomly generated SNPs.

Our findings with high-density molecular markers (i.e., 26,657 SNPs) are in contrast to recent research that identified TCAs in the US as a relatively undifferentiated superclone using microsatellite markers [15]. Our results show that the use of high-density molecular markers can provide greater resolution among differentiating populations than the use of low-density molecular markers. Several studies have shown the emergence of population genetic structure when using high-density molecular markers. For example, the genetic characterization of yellow fever mosquito (Aedes aegypti L.) populations in California, using SNPs generated from HTS, identified previously undetected distinct Northern and Southern populations [36]. The two genetically differentiated mosquito populations suggested independent introduction events and limited admixture [36]. Similarly, greater resolution of the population genetic structure of a coral reef fish (Elacatinus lorl D.S. Jordan) along the Belize barrier reef were identified with the use of 2,418 SNPs compared with the use of 89 microsatellite loci [37]. Thus, when some species are reported as lacking genetic variation, it is important to determine how much of this reporting may be due to the selection of a limited quantity of diagnostic molecular markers (i.e., low-density molecular markers).

The population genetic variation of TCAs detected by our study, could be the result of several different potential factors. For instance, there could have been multiple independent propagule introductions into the US, which is supported by a study conducted in a wider geographic region within the native range of the TCA using COI (Fernández, MB; pers. com.). The differences in population genetic structure detected among TCA in different states aligns with this scenario (Fig. 2). These findings are also supported by a study that found differences among ultraconserved elements (UCEs) of TCAs from different states [38]. While transportation of TCA propagules can happen via contaminated soil, wood (i.e., logs or pallets), or vegetation [8], the population structure revealed by our study suggests that while this happens within a state it is infrequent among states (Fig. 2). The great amount of genetic differentiation among ants in the US and South America suggests historic introduction events of propagules without ongoing gene flow. This inference is based upon our limited sampling of TCAs from Argentina, which did not capture all the representative genotypes present in the native range.

In addition, differentiation post introduction could also have contributed to the current population structure. TCAs have been in the US for over 20 years, with propagules reported in 1990s for Florida and the 2000s for Texas. Under the right conditions, microevolutionary forces can generate genetic and phenotypic differentiation relatively rapidly. For instance, in less than 50 years, soapberry bugs feeding on invasive plants with smaller seeds than their native host plants, evolved decreased beak lengths in north America [39]. Furthermore, soapberry bugs in Texas and New Mexico were reported as undergoing host-associated differentiation from Western soapberry to Mexican buckeye in the past approximately 72 years or less (i.e., 1950s to 1960s) [3]. Similarly, invasive spotted-wing drosophila (Drosophila suzuki) Matsumura, estimated to have colonized the Hawaiian archipelago in the 1980s, underwent morphological and population genetic differentiation within approximately 30 years. Flies at higher elevation have larger wing sizes and are genetically different than flies
occurring at lower elevations across several islands [2]. These examples show that relatively rapid genetic and phenotypic differences among insect populations can be detected, may be biologically relevant, and should not be dismissed.

Limited morphological differentiation among crazy ant workers makes visual identification of closely related species difficult [16, 38]. Based upon the high levels of genetic differentiation of ants from Colombia and Peru, it is likely that these are two cryptic species of Nylanderia (Table 2). In a study of red wood ants, $F_{ST}$ values of 0.2 and above were reported among different species [40]. Similarly, different species of North American Fire Ants were found to have $F_{ST}$ values between 0.2 and 0.442 [41]. Our findings of great genetic differentiation among ants from Colombia and Peru further supports findings by Williams et al. (2022) that Nylanderia fulva is a species complex. Additional morphological analysis of male ants from these regions could further support our molecular findings and aid in determining whether these species have been previously described or are a new addition in the fulva/pubens species complex.

Insects of different genotypes may interact (i.e., conspecific or interspecific) in invaded regions in ways dissimilar to their native habitats [42, 43], which can result in admixture or hybridization events [44–46]. In particular, this process might be facilitated by different invasive genotypes (or even species) arriving to the same novel locations and mating, when these genotypes might not mate in their native region. Two species of invasive fire ant (i.e., Solenopsis invicta and Solenopsis richteri) were reported to admix in the introduced range of North America [42]. Admixture events could promote genotypes that have enhanced insecticide resistance or greater tolerance towards biological control agents. For instance, admixture between diamond back moths that are genetically adapted to consume rice and those genetically adapted to consume corn occurred in conjunction with insect dispersal, which maintained alleles for insecticide resistance in the surrounding geographic areas [47]. Additionally, hybridization events among different species could result in novel phenotypes. When the native corn earworm (Helicoverpa zea Boddie) mated with invasive cotton bollworms (Helicoverpa armigera Hübner) in Brazil, this resulted in hybrid offspring that were more resistant to pesticide applications than the native corn earworns [48]. Thus, admixture in invaded regions could result in novel adaptations, which might influence invasive pest behaviors and subsequent management practices.

Incorporating information on variability of genetic composition can be valuable for managing populations that have different responses to control practices. In addition, the use of high-density molecular markers can allow for the correlation of pest traits to different populations. For instance, genetically differentiated populations of green peach aphids (Myzus persicae Sulzer), Mediterranean fruit flies (Ceratitis capitata Wiedemann), and mountain pine beetles (Dendroctonus ponderosae Hopkins) have all been shown to vary in their susceptibility to pesticides and stress tolerance [49–51]. Recently, the application of a fungus-like microsporidian pathogen was found effective at decreasing and in some instances eliminating TCA populations [52–54]. Further investigation is required into the effectiveness of this biopesticide on other TCA populations.

The correlation of genomic information with pest traits can be used to tailor management approaches towards genetically differentiated populations. The next steps are to correlate these genetic differences with to potential pest traits, which would provide information that can be incorporated into management practices. Further studies should characterize how these invasive TCA populations differ in traits relevant to pest control. We think that assessing potential variation among populations in traits such as vector competency, insecticide susceptibility, symbiotic interactions, behavior, etc., are areas relevant to pest management [4, 36, 55–57].

**Conclusion**

Our study shows that high-density molecular markers (i.e., thousands of SNPs) revealed population genetic structure among recently invaded TCA propagules. In addition, we identified two different species from Colombia and Peru within the fulva/pubens species complex. Identifying genetic differences both in the introduced and native range is the beginning to understanding the evolutionary ecology of these pests. Continued monitoring of different TCA populations is recommended, as well as periodic genetic characterization that might identify ongoing changes in pest traits influenced by microevolutionary forces.

**Material And Methods**

**Sample Collection**

TCA worker ants (female and diploid) were collected from across the US (i.e., Alabama, Florida, Georgia, Louisiana, Mississippi, and Texas) (Table 1). In addition, specimens from native (i.e., Argentina) and invasive (i.e., Colombia and Peru) ranges in South America were used to determine potential points of origin (Table 1). To maximize detection of potential genetic variation within each state and country, when possible, worker ants were collected from nests at least 1 km apart from each other (Supplemental Table 3). In addition, when possible, workers were collected from at least three different locations within each state or country. Ants were preserved in 95% ethanol and labeled with collection site information, collection date and the collector’s name.

**DNA Extraction**

DNA was extracted from individual worker ants using a Gentra PureGene Kit (QIAGEN, Valencia, CA, USA). The quantity and purity of DNA was checked using a NanoDrop spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA, USA) and a PicoGreen dye assay on a NanoDrop Fluorospectrometer (NanoDrop Technologies, Inc., DE). DNA samples were submitted to Texas A&M University AgriLife Genomics and Bioinformatics Service (TXGen, College Station, TX). Samples with good quantities and quality of DNA (i.e. 20 + ng/μL DNA concentrations on Nanodrop, when possible samples had a Genomic Quantity Number Technologies, Inc., DE). DNA samples were submitted to Texas A&M University AgriLife Genomics and Bioinformatics Service (TXGen, College Station, TX).

**Data Analysis and Single Nucleotide Polymorphism Identification of Ants**
An average of 3.8 million reads with phred scores ≥ 24 (an average sequence phred score of 35) were obtained for each individual TCA worker, for a total of 96 samples, with known geographic locations. Sequence reads were filtered for quality using a FastQC version 0.72 tool to summarize all Phred scores, which met a minimum score of 20 for each sample [60]. Reads were demultiplexed and further filtered using the process radtags program in STACKS version 2.53, providing both restriction enzymes (−index-index -e ecoRI –renz_2 nlalll) and quality filtering (size of sliding window, -w 0.15; Phred score, -s 20). Reads were mapped onto the tawny crazy ant (Nylanderia fulva) reference genome (NCBI BioProject PRJNA517949 by Kranti Konganti and Aaron Tarone) following the STACKS protocol [33, 61]. The demultiplexed reads were run first through the reference map pipeline (comprised of gstacks and populations). The Marukilow model was used to call variants and genotypes in gstacks. Then the populations program was run with TCA split into nine geographic locations (Argentina, Colombia, Peru, Texas, Louisiana, Florida, Georgia, Alabama, and Mississippi) and only the first SNP in a locus was kept (−write-single-snp) [62] and a single representative for overlapping sites for reference aligned reads was enabled (−ordered-export). Settings for the populations program included filtering loci to keep those shared by 80% or more of samples within a population (−min-samples-per-pop; -r 0.8), at or above a minimum minor allele frequency of 2% (−min-maf 0.02), and shared by a minimum of 1 population (−min-populations; -p 1), with a maximum observed heterozygosity of 70% (−max-obs-het 0.7); no required value was set for the metapopulation, meaning the minimum percentage of individuals across populations required was 0% (−min-samples-overall; -R) [61, 63–66]. The populations program in STACKS was used to generate population genetic summary statistics (Table 1) [33].

### Evaluating Genetic Relationships

The data file generated by the populations program in STACKS was used to assess population genetic variation in STRUCTURE version 2.3.4 [67]. STRUCTURE runs (K = 1 – 12) were done for the nine ant collection locations from the introduced and native ranges, plus an additional three potential populations. Each run had a 10,000 burnin with 10,000 iterations for MCMC (Markov Chain Monte Carlo) and was replicated 10 times for each value of K [36, 68]; taking into consideration the lower sample size of Argentina, these were run for 25000 burnin with 25000 iterations. The number of putative populations in the data was analyzed with two different programs for all values of K, with Structure Harvester Web with five runs per value of K from 1 to 10, and with the packages poppr 2.9.3 [69], ape [70], and magrittr [71] to run a BIC (Bayesian Information Criterion) in R [72–74]. The estimated number of populations was determined by assessing peak ΔK values [73, 75].

An Analysis of MOlecular Variance (AMOVA) in R was used to analyze population genetic statistics [76]. The following packages in R were loaded to run 999 permutations: adegenet [77, 78], poppr [69], genepop [79], and ggplot2 [80]. Sample collections were analyzed following the pipeline from the GitHub repository: Population genetics in R [81, 82]. The missing loci parameter was set to ignore (missing = "ignore").

Sample clusters based on similarities of shared SNPs was visualized using a DAPC (Discriminant Analysis of Principal Components) with the adegenet program following the recommended protocols [77, 83].

The software Genepop on the Web version 4.7.5 [79, 84] was used to run a Mantel test that analyzed Isolation by Distance (IBD) [85, 86] among population pairs. A matrix of F\textsubscript{ST} values generated from STACKS and a matrix of estimated geographic distances in kilometers among populations was constructed in a .txt file (Supplemental Table 4a, 4b, 4c). Option 6. F\textsubscript{ST} and other correlations was selected, with the following parameter settings (Allele identify (F-statistics): Estimation Ploidy: Diploid, Isolation by distance: 9. Isolation by distance (using Isolde), Isolation by distance parameters: Linear geographic distances, Convert F-statistics to F/(1-F) statistics: Yes, Minimum distance between samples to be taken in account for regression 0.0001, Number of permutations for Mantel test: 10000, Please enter 4 random number generator seeds: 5, 13, 37, 75, Output format & Delivery: HTML – Plain Text). The .txt file containing the matrixes was upload to Genepop on the Web for data analysis. The pairwise comparison output from Isolde for F\textsubscript{ST} test: 10000, Please enter 4 random number generator seeds: 5, 13, 37, 75, Output format & Delivery: HTML – Plain Text). The .txt file containing the matrixes was upload to Genepop on the Web for data analysis. The pairwise comparison output from Isolde for F\textsubscript{ST} values and linear geographic distances was analyzed in Excel with a linear regression to obtain the slope, intercept, and R\textsuperscript{2} values of the dataset.

### Abbreviations

- AFLPs = amplified fragment length polymorphisms
- COI = cytochrome oxidase I
- DAPC = discriminant analysis of principal components
- HTS = high throughput sequencing
- IBD = isolation by distance
- ITS = internal transcribed spacer
- SNPs = single nucleotide polymorphisms
- TCA = tawny crazy ant
- UCEs = ultraconserved elements
- US = United States

### Declarations
Ethics declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable.

Availability of data

Raw data can be found on NCBI under accession number PRJNA892980. Scripts for data analysis are available through GitHub at: https://github.com/holtjocelyn/tca_pop_gen.git

Competing interests

All the authors declare no competing interests.

Funding

This study was supported by the Texas Invasive Ants Exceptional Item Grant. JRH was supported by an AFRI Competitive Grant from USDA National Institute of Food and Agriculture (Education and Workforce Development project, accession number: TEX09837) during the writing of this manuscript.

Author Contributions

JRH and RFM wrote the main manuscript text. JRH produced figures and analyzed data. LAC edited the manuscript and provided content. JML edited the manuscript and provided content. TJR provided bioinformatics assistance, reviewed scripts, and edited the manuscript. All authors reviewed the manuscript.

Acknowledgements

Thank you to the Texas Ecological Research Laboratory for providing access to private land associated with sample collection. We also thank Dr. Spencer Johnston for assisting with initial genome size assessments, which suggested there was genetic differentiation among TCA from Texas and Florida. We are thankful to the many individuals who contributed ant samples for analysis. Thank you to Mackenzie Tietjen and Kyle Harrison for advice on data analysis. Thank you to Michael Dickens for troubleshooting advice for the TAMU Cluster. We also thank our undergraduate researchers Robert Chapa and Jose Torres for their assistance in maintaining ant colonies and making collection vouchers. Thank you to our colleagues who provided location information for collection of TCA or TCA specimens: Ed LeBrun, Robert Puckett, and Bryant McDowell. We also thank our colleagues who contributed samples: Robert Puckett, Ed Vargo, J.A. MacGown, Jessica Warren, Nathalie Baena Andrés Posso-Terranova, Fernando Díaz, Rachel Streccker, Ben Gochnoun & Dan Suter, David Oi, Bryant McDowell, Crys Wright, J. McMullan, John Gordy, Universidad del Valle - Colombia, J. Vasquez, Gregg Henderson, Wade Sharp, Odair C. Bueno, David Cross, Melissa Jones, Mackenzie Kjeldgaard, John Warner, Fudd Graham, Dan Devenport. Thank you to our colleagues who provided location information for collection of TCA or TCA specimens: Ed LeBrun, Robert Puckett, and Bryant McDowell. We also thank our colleagues who contributed samples: Robert Puckett, Ed Vargo, J.A. MacGown, Jessica Warren, Nathalie Baena Andrés Posso-Terranova, Fernando Díaz, Rachel Streccker, Ben Gochnoun & Dan Suter, David Oi, Bryant McDowell, Crys Wright, J. McMullan, John Gordy, Universidad del Valle - Colombia, J. Vasquez, Gregg Henderson, Wade Sharp, Odair C. Bueno, David Cross, Melissa Jones, Mackenzie Kjeldgaard, John Warner, Fudd Graham, Dan Devenport. Thank you to Mariana Mateos and Aaron Tarone for their invaluable insight and feedback. Thank you to María Belén Fernández and Fernando Díaz for valuable observations and comments to the manuscript.

References


74. Population genetics in R: Population differentiation for sequence data [https://popgen.nescent.org/PopDiffSequenceData.html]

Figures

DeltaK = mean(|L"'(K)|) / sd(L(K))

![Figure 1](image-url)
Structure Harvester results for all TCA from South America and the US showing a peak of DK at five and another peak at 11 putative populations.

Figure 2

Structure plot showing geographically differentiated populations of TCAs. A) The colors show seven genetically differentiated clusters based upon geographic location. B) Further analysis of only US samples support six genetically differentiated clusters based upon geographic location, for a total of nine population from South America and the US. The geographic collection location is provided at the top of the structure plot. C) When only samples from Argentina were analyzed, this revealed three genetically differentiated clusters, which when added with the six from the US, one from Colombia, and one from Peru equals eleven total clusters that were identified by Structure Harvester.
Figure 3

DAPC analysis in R. The colored squares represent each geographic region from which samples were collected in South America and the US.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementalmaterialallfiguresandtablescapopgen27Dec2022pdf.pdf