

# SCIENTIFIC REPORTS



OPEN

## The reunion of two lineages of the Neotropical brown stink bug on soybean lands in the heart of Brazil

Patricia L. Soares, Erick M. G. Cordeiro, Frederico N. S. Santos, Celso Omoto & Alberto S. Correa

The rapid pace of conversion of natural areas to agricultural systems is highly concerning, and the consequences for conservation and pest management are not yet fully understood. We examined mitochondrial (COI and *Cytb*) and nuclear (ITS1) gene regions of 21 populations of the stink bug *Euschistus heros*, to investigate the genetic diversity, genetic structure, and demographic history of this emerging soybean pest in South America. Two deep lineages that diverged in the Pliocene (4.5 Myr) occur over wide areas of Brazil. Historical changes during the Plio-Pleistocene led to significant genetic differences between *E. heros* populations, which differentiated further in several biomes. The northern lineage is older, more diverse, and prevalent in the Amazon and Caatinga, while the southern lineage is younger, less diverse, and prevalent in the Atlantic Forest and Chaco biomes. *Euschistus heros* populations are expanding in size and range but at different rates, strongly affected by environmental variables. Secondary contact between the main lineages is now occurring, mainly in areas of intensive farming and particularly in the Cerrado, an important agricultural frontier. Individuals adapted to different environmental conditions and to large monocultures might currently be combining into a panmictic and hard-to-control pest population.

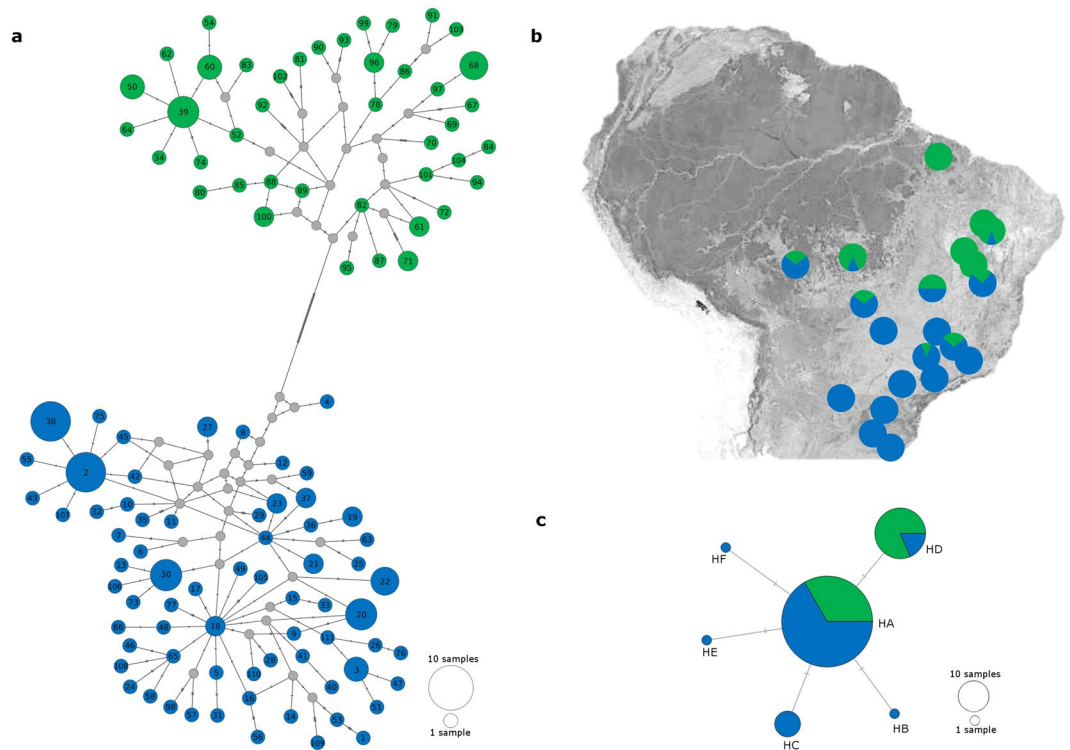
The drastic climate changes during the Plio-Pleistocene have been considered the main cause of high levels of diversification in many areas in Brazil<sup>1–3</sup>. The four major Brazilian biomes, the Atlantic Forest (AF), Cerrado (central savannas), Amazon and Caatinga (seasonally dry tropical forest) have undergone profound changes during glaciation and interglacial cycles<sup>4</sup>. Due to this complexity and landscape composition, the processes of diversification by vicariance and habitat refugia are frequently invoked to explain the high levels of species endemism and diversity in this part of the planet<sup>5–8</sup>. Recent findings suggest that forest expansion and contraction and ‘historically stable areas’ may also have played a major role in the differentiation of lineages<sup>9</sup>.

Today, the Amazon and Atlantic forests are separated by a unique mosaic composed of savannas and woodlands that covers a large area between the two forest biomes, from Argentina and Paraguay (the Argentinean and Paraguayan Chaco) and continues through the central Brazilian Cerrado to the Caatinga in northeastern Brazil. This belt of mostly sparse dry vegetation is known as the ‘major South American disjunction’<sup>3,10,11</sup> and forms a natural barrier preventing the movement of organisms between the northern Brazilian biomes and the Atlantic Forest<sup>12–14</sup>. Recent human impacts on these biomes, particularly the expansion of agricultural areas, have drastically changed the landscape and likely the connection between ecosystems, rearranging patterns that began to be formed millions of years ago.

The expansion of agricultural frontiers necessitates conversion of the native habitat to agriculture<sup>15</sup>. Fragments of native vegetation become embedded in a matrix of cropland and pasture that will eventually affect the species and ecosystem dynamics<sup>16</sup>. In some instances, habitat loss can limit the geographic range of endemic species, although certain species may thrive in their new surroundings. Cropland areas can provide exploitable resources, e.g. the expansion of soybean cultivation in Brazil in recent decades<sup>17</sup>. Soybean crops were formerly limited to southern Brazil (Atlantic Forest); in the early 1970s, advances in farming methods and new varieties allowed soybean farmers to expand into a new and important frontier, the Cerrado<sup>18–20</sup>.

Farming in the Cerrado has had both negative and positive impacts over the last 40 years. The expansion of soybean crops caused extensive environmental impacts such as fragmentation and the loss of natural areas, a

Department of Entomology and Acarology, University of Sao Paulo, Luiz de Queiroz College of Agriculture (USP/ESALQ), Piracicaba, SP, 13418-900, Brazil. Correspondence and requests for materials should be addressed to A.S.C. (email: [ascorrea@usp.br](mailto:ascorrea@usp.br))



**Figure 1.** Median-joining network and geographic distribution of *Euschistus heros* haplotypes in South America. **(a)** Network of 159 concatenated mitochondrial COI-Cytb sequences. Size of haplotype circles reflects sample size, and gray nodes represent missing haplotypes. Colors indicate the two mitochondrial lineages of haplotypes: lineage N (northern) in green, and lineage S (southern) in blue. Number of mutation steps is shown as hatch marks. **(b)** Geographic distributions of the mitochondrial haplotypes. Circles represent the proportion of each lineage. The map was obtained from Google Maps, Map ©2017 Google, INEGI. Pie charts were drawn with Microsoft Excel. Maps were modified with GIMP 2.8.22 **(c)** Network of 124 nuclear ITS1 region sequences. The haplotype color refers to the mitochondrial lineages of the individual from which the ITS1 sequence was obtained.

matter of great concern for ecologists and conservationists<sup>2,21,22</sup>. On the other hand, Brazil is the second largest soybean producer after the USA, and soybeans account for an important share of its GDP<sup>23</sup>. For these reasons, it is important to understand both the impacts of soybean expansion on connecting natural populations, and the influx of insect pests from natural areas into soybean croplands.

The Neotropical brown stink bug, *Euschistus heros* (Fabr. 1798) (Hemiptera: Pentatomidae), is one of the most important pests of soybeans<sup>24</sup>. This Neotropical native is widespread in South America, living in markedly different environments including the Amazon Forest, Cerrado, Caatinga and Atlantic Forest<sup>25</sup>. The dispersal ability of *E. heros* is not well known, but is considered to be poor, perhaps because of its limited flight activity and diapause behavior<sup>26,27</sup>. This polyphagous pest feeds on members of Fabaceae, Solanaceae, Brassicaceae, Compositae, and Malvaceae, and often reaches high population densities on soybeans<sup>28–31</sup>. Rarely reported before the 1970s<sup>24,32,33</sup>, *E. heros* has since increased in abundance and is now found in all major soybean-producing regions<sup>24</sup>. Recently, this pest was recorded in Argentina<sup>34</sup> and Paraguay<sup>24</sup>, raising concerns regarding a possible range expansion to other locations in South, Central and North America<sup>24</sup>.

Here we present a genealogy of mitochondrial and nuclear DNA sequences from *E. heros*. We addressed several questions regarding the genetic diversity, population structure, and demographic history of *E. heros* populations in Brazil and Paraguay, examining the potential role of past events in the differentiation of lineages, and the recent events (i.e. soybean expansion) promoting the admixture of ancient lineages. Our first objective was to determine the genetic distribution, studying population divergence and population structure. Second, we investigated the demographic history of *E. heros* in different biomes in South America. Finally, we used a modelling approach to explore how environmental variables and soybean expansion can explain the genetic pattern of current populations of *E. heros* in Brazil.

## Results

**Genealogical inferences.** Genealogical relationships among 111 mitochondrial haplotypes indicate two well-supported *E. heros* lineages separated by 52 mutational steps, and an estimated genetic distance of  $D = 0.042$  (Fig. 1a). The southern lineage (S) haplotypes occur mainly in southern regions of South America (Fig. 1b). Small percentages of lineage S haplotypes were also found in central and northeastern Brazil (Fig. 1b), in a wide range of habitats (Table 1). The northern lineage (N) occurs mainly in northern and northeastern South America and

| ID | Locations (City, State)      | Code | Biome           | mtDNA haplotypes (n)*   | ITS1 haplotypes (n) | Latitude (S) | Longitude (W) |
|----|------------------------------|------|-----------------|---|---------------------|--------------|---------------|
| 1  | Teutônia, RS                 | RST  | Atlantic Forest | <i>H1; H2(2); H3; H4; H5; H6; H7; H8</i>                      | HA(4); HB           | 29°26'48.83" | 51°48'50.44"  |
| 2  | Santa Bárbara do Sul, RS     | RSSB | Atlantic Forest | <i>H9; H10; H11; H12; H13; H14; H15; H16; H17</i>             | HA(5)               | 28°22'01.95" | 53°15'06.23"  |
| 3  | Chopininho, PR               | PRC  | Atlantic Forest | <i>H18; H19(2); H20(3); H21; H22(4); H23; H24; H25</i>        | HA(6); HC(3)        | 25°51'23.28" | 52°32'14.01"  |
| 4  | Cornélio Procópio, PR        | PRCP | Atlantic Forest | <i>H2(3); H20; H26; H27; H28; H29</i>                         | HA(3)               | 23°10'57.89" | 50°38'44.37"  |
| 5  | Anhembi, SP                  | SPA  | Atlantic Forest | <i>H23; H30(2); H31; H32; H33</i>                             | HA(3)               | 22°47'17.09" | 48°07'52.29"  |
| 6  | Lavras, MG                   | MGL  | Atlantic Forest | <i>H27; H43; H44; H45</i>                                     | HA(2); HC           | 21°14'54.56" | 45°00'04.95"  |
| 7  | General Higinio Morinigo, PY | PY   | Chaco           | <i>H2; H20; H38; H105; H106; H107; H108; H109; H110; H111</i> | HA(3); HC(2)        | 25°09'19.55" | 55°29'59.24"  |
| 8  | Costa Rica, MS               | MSCR | Cerrado         | <i>H3; H30; H37; H48; H49</i>                                 | HA(2); HC           | 18°32'37.15" | 53°07'45.17"  |
| 9  | Jaboticabal, SP              | SPJ  | Cerrado         | <i>H2; H30; H34; H35; H36; H37; H38(2)</i>                    | HA(8)               | 21°15'09.05" | 48°19'32.43"  |
| 10 | Capitólio, MG                | MGC  | Cerrado         | <i>H21; H30; H39(2); H40; H41; H42</i>                        | HA(6); HD           | 20°36'50.88" | 46°02'52.35"  |
| 11 | Santa Juliana, MG            | MGSJ | Cerrado         | <i>H38(2); H46; H47</i>                                       | —                   | 19°18'40.47" | 47°31'57.69"  |
| 12 | Padre Bernardo, GO           | GOBP | Cerrado         | <i>H18; H38; H39; H50; H51; H52</i>                           | HA(6); HD(3)        | 15°09'39.38" | 48°17'01.46"  |
| 13 | Rondonópolis, MT             | MTR  | Cerrado         | <i>H39; H53; H54; H55; H56; H57; H58</i>                      | HA(7); HD; HE       | 16°27'55.71" | 54°38'19.04"  |
| 14 | Sorriso, MT                  | MTS  | Amazon Forest   | <i>H39; H50(2); H59; H60; H61; H62</i>                        | HA(7); HD(2); HF    | 12°32'34.61" | 55°43'17.53"  |
| 15 | Cerejeiras, RO               | ROC  | Amazon Forest   | <i>H2; H3; H60; H63; H64; H65; H66</i>                        | HA(5)               | 13°11'14.64" | 60°49'02.48"  |
| 16 | Paragominas, PA              | PAP  | Amazon Forest   | <i>H67; H68(4); H69; H70; H71(2); H72</i>                     | HA(4); HD(4)        | 03°00'09.95" | 47°21'11.19"  |
| 17 | Correntina, BA               | BAC  | Caatinga        | <i>H38(2); H60; H73; H74; H75; H76; H77</i>                   | HA(8); HD(2)        | 13°20'33.19" | 44°38'08.06"  |
| 18 | São Desidério, BA            | BASD | Caatinga        | <i>H78; H79; H80; H81; H82</i>                                | HA(3); HD(5)        | 12°21'27.47" | 44°58'38.23"  |
| 19 | Luis Eduardo Magalhães, BA   | BALE | Caatinga        | <i>H61; H83; H84; H85; H86; H87; H88; H89; H90</i>            | HA(3); HD(6)        | 12°05'25.63" | 45°46'49.94"  |
| 20 | Bom Jesus, PI                | PIB  | Caatinga        | <i>H91; H92; H93; H94; H95; H96; H97; H98; H99</i>            | —                   | 09°04'17.95" | 44°21'33.65"  |
| 21 | Bom Jesus, PI                | BJPI | Caatinga        | <i>H96; H100(2); H101; H102; H103; H104</i>                   | HA(2); HD(3)        | 09°04'17.95" | 44°21'33.65"  |

**Table 1.** Sampling localities of *Euschistus heros*, with code, biomes, mitochondrial haplotype from two concatenated genes (COI-Cytb), haplotype nuclear ITS1 region, and geographic coordinates. \*Italics haplotype = Southern lineage; Bold haplotype = Northern lineage.

was not present in Paraguay or southern Brazil (Table 1). A total of 91 haplotypes were identified as private haplotypes; the most frequent variants were H2 and H38 ( $n = 8$ ), both from lineage S (Table 1).

Analyses of the ITS1 region revealed a single nucleotide polymorphism variation separating the haplotypes. The six haplotypes were separated by a one-step mutation (Fig. 1c and Supplementary material S2). We created two alternative ITS1 sequences for all individuals that showed ambiguity in the polymorphic site, to recreate the heterozygotes. The ITS1 network had lower haplotype diversity than the mitochondrial network. Haplotype HA was the most frequent (70.16%) and was widely distributed across all regions (Table 1). Haplotype HC was found in southern and central South America, and associated individuals previously identified as mitochondrial lineage S (Table 1 and Supplementary Fig S2). Haplotype HD (21.77%) was found only in northern South America, and associated individuals previously identified as mitochondrial lineages N or S (Table 1). The single haplotypes HB, HE and HF were found in populations RS1, MT1, and MT2, respectively (Table 1). The sharing of nuclear haplotypes by specimens from both mitochondrial lineages may indicate that these lineages can interbreed (Fig. 1c).

**Diversity statistics.** These South American *E. heros* populations showed extensive mitochondrial diversity. Of 159 concatenated mitochondrial sequences of COI and Cytb analyzed, 111 haplotypes were found; most (82%) of these haplotypes were private and only 18% were shared among individuals. The overall haplotype diversity, nucleotide diversity and mean number of nucleotide differences were  $h = 0.991$ ,  $\pi = 0.03312$  and  $K = 32.892$ , respectively (Table 2). The haplotype diversity was similar in lineage N ( $h = 0.984$ ) and lineage S ( $h = 0.982$ ); however, lineage N ( $\pi = 0.0090$ ) had higher nucleotide diversity than lineage S ( $\pi = 0.0062$ ). For the different biomes, the haplotype diversity among biomes (groups) was low, ranging from 0.967 in the Amazon to 1.000 in the Chaco. The nucleotide diversity varied more widely among biomes, from 0.00627 in the Atlantic Forest to 0.03147 in the Amazon Forest (Table 2). Locations where both lineages were present had the highest levels of nucleotide diversity and mean numbers of nucleotide differences. Sequence analysis of the ITS1 region identified six haplotypes with a haplotype diversity of 0.461, nucleotide diversity of 0.0008 and mean number of nucleotide differences of 0.499 (Table 2). The haplotype diversity and nucleotide diversity were higher in lineage N ( $h = 0.503$ ;  $\pi = 0.0008$ ) than lineage S ( $h = 0.355$ ;  $\pi = 0.0006$ ), previously defined by the mitochondrial network. Among biomes, the highest diversity was observed in the Chaco ( $h = 0.600$ ;  $\pi = 0.0009$ ) and the lowest diversity in the Atlantic Forest ( $h = 0.315$ ;  $\pi = 0.0005$ ) (Table 2). The Amazon Forest, Caatinga and Cerrado have higher nucleotide diversity due to mixing of the two lineages in these areas.

**Mitochondrial divergence dating.** The estimated age of origin of the lineage S (southern lineage) clade was 4.5 Myr (95% C.I 2.801–6.453 Myr), during the Pliocene, with intense diversification in the Pleistocene and Holocene (Fig. 2b).

**Population Structure.** Lineage N haplotypes were associated mainly with the Amazon Forest and Caatinga, with one, more recent clade (CW) associated with central Brazil, a transitional region among the Cerrado,

| Geographic regions | Sample size (N) | Haplotype number (H) | Haplotype diversity (h) | Nucleotide diversity ( $\pi$ ) | Average # of nucleotide difference (K) |
|--------------------|-----------------|----------------------|-------------------------|--------------------------------|--|
| <b>COI-Cytb</b>    |                 |                      |                         |                                |  |
| Pooled             | 159             | 111                  | 0.991                   | 0.0331                         | 32.892                                 |
| Atlantic Forest    | 50              | 36                   | 0.979                   | 0.0063                         | 6.224                                  |
| Chaco              | 10              | 10                   | 1.000                   | 0.0066                         | 6.511                                  |
| Cerrado            | 37              | 27                   | 0.970                   | 0.0256                         | 25.396                                 |
| Amazon Forest      | 24              | 18                   | 0.967                   | 0.0315                         | 31.246                                 |
| Caatinga           | 38              | 35                   | 0.996                   | 0.0252                         | 25.007                                 |
| Lineage N*         | 57              | 42                   | 0.984                   | 0.0090                         | 8.921                                  |
| Lineage S*         | 102             | 69                   | 0.982                   | 0.0062                         | 6.198                                  |
| <b>ITS1 region</b> |                 |                      |                         |                                |  |
| Pooled             | 124             | 6                    | 0.461                   | 0.0008                         | 0.499                                  |
| Atlantic Forest    | 28              | 3                    | 0.315                   | 0.0005                         | 0.325                                  |
| Chaco              | 5               | 2                    | 0.600                   | 0.0009                         | 0.600                                  |
| Cerrado            | 36              | 4                    | 0.340                   | 0.0006                         | 0.357                                  |
| Amazon Forest      | 23              | 3                    | 0.466                   | 0.0008                         | 0.490                                  |
| Caatinga           | 32              | 2                    | 0.516                   | 0.0008                         | 0.516                                  |
| Lineage N*         | 50              | 2                    | 0.503                   | 0.0008                         | 0.503                                  |
| Lineage S*         | 74              | 6                    | 0.355                   | 0.0006                         | 0.382                                  |

**Table 2.** Measures of genetic diversity for *Euschistus heros* based on two concatenated mitochondrial genes (COI-Cytb) and ITS1 region. \*Lineages were previously defined by mitochondrial network.

Caatinga and Amazon Forest (Fig. 2a,c). Lineage S predominates in the Atlantic Forest and Chaco, with a lower frequency in the Cerrado, Caatinga and Amazon Forest (Fig. 2a,c).

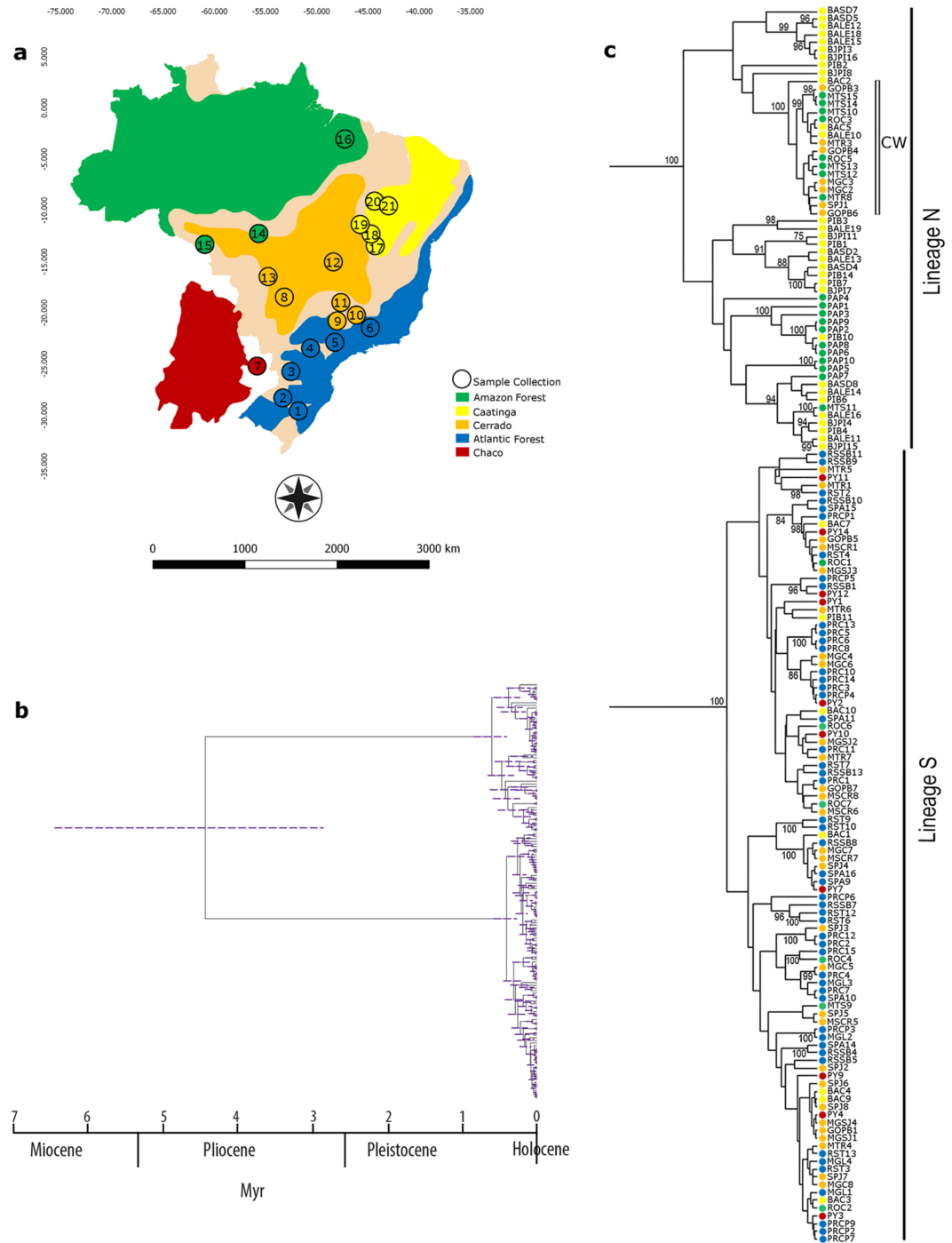
At the regional scale, the *E. heros* populations showed high genetic structure, as assessed by the Analysis of Molecular Variance (AMOVA). Differences among populations accounted for most of the genetic variances in mtDNA (56.57%,  $P < 0.001$ ) and a high and significant value in ITS1 regions (20.18%,  $P < 0.001$ ) (Table 3). A test of the hypothesis that the genetic variation is structured by biomes showed that 45.07% of the mtDNA total variance was distributed among biomes ( $\Phi_{CT} = 0.450$ ,  $P < 0.001$ ). Furthermore, the larger portion of genetic variation within populations (39.61%,  $\Phi_{ST} = 0.603$ ) indicates overall genetic differentiation in these populations (Table 3). Analysis of the ITS1 region supported the mitochondrial data, showing a significant structuring by biome ( $\Phi_{CT} = 0.173$ ,  $P < 0.001$ ), in which most of the genetic variation was within populations (Table 3).

**Demographic statistics inferred for mitochondrial genes.** Considering the two lineages, significant negative values were found in both the Tajima's D and Fu's  $F_s$  neutrality tests, indicating population expansion or purifying selection. Considering the biomes, the neutrality test statistics did not fully agree with one another. Fu's  $F_s$  statistic was significantly negative for all biomes, but only the Atlantic Forest biome had a significant negative Tajima's D value (Table 4).

For the lineages, the mismatch distribution analysis resulted in a nonsignificant SSD ( $P > 0.05$ ), indicating a recent demographic expansion of lineage S but not lineage N ( $P = 0.04$ ). For the biomes, a nonsignificant SSD ( $P > 0.05$ ) was also found for the *E. heros* populations in all biomes but the Caatinga ( $P = 0.03$ ) (Table 4). The non-significant raggedness index ( $P > 0.17$ ) supports the spatial-expansion model of populations of lineages, biomes, and the entire group (all populations combined) (Table 4). The  $\tau$  values were higher in the Cerrado and the Amazon Forest,  $\tau = 57.5$  and  $\tau = 58.7$ , respectively, compared to the other three biomes, Atlantic Forest ( $\tau = 6.4$ ), Chaco ( $\tau = 6.0$ ) and Caatinga ( $\tau = 7.3$ ) (Table 4).

The expansion of populations in the Amazon Forest, Chaco, Caatinga, Cerrado and Atlantic Forest occurred within the last 500 years, corresponding to a recent expansion during the Quaternary according to the Bayesian skyline plot analysis (Fig. 3). The Chaco and Atlantic Forest populations remained stable during the past 100 years, while the Caatinga and Cerrado populations are still expanding. According to the effective population size ( $N_e$ ), the Atlantic Forest population is the largest, followed by the Caatinga. The Cerrado, Amazon Forest and Chaco populations have similar sizes, but the Cerrado population is still expanding very rapidly, while the Chaco population is expanding slowly and the Amazon population is now contracting (Fig. 3).

**Environmental features and soybean expansion modelling the current mitochondrial lineage distribution.** Three models passed the cutoff (i.e. models that were less than 2 units away from the "best" model) to explain the presence (%) of the southern lineage at a given location (i.e. the probability of finding an individual from the southern lineage). The best predictors were the 'max temperature of the warmest month', 'latitude', and 'annual mean temperature' (Fig. 4 and Table 5). The most important variables were the 'max temperature of the warmest month' and 'latitude', which received the highest score in all top models. The model performance improved when 'annual mean temperature' was excluded, i.e.  $AIC_c$  (36.22 and 35.35) and  $w_i$  (0.09 and 0.14) (Table 5). The best model selected ( $AIC_c = 35.35$ ) was 22.29 units away from the null model ( $AIC_c = 57.64$ ).



**Figure 2.** Bayesian coalescent tree for *Euschistus heros*. (a) Geographic and biome distribution of each population of *E. heros* sampled (see Table 1) (map based on Ab’Saber<sup>77</sup> and modified with GIMP 2.8.22); (b) Bayesian phylogeny tree of 159 concatenated mitochondrial sequences (COI-Cytb). Gray bars at nodes indicate 95% highest probability density intervals (HPD) confidence intervals for nodal age. (c) Bayesian phylogeny tree showing posterior probability values (>75) and biome where individuals were collected (taxon names provided in Table 1).

The variable ‘latitude’ (0.50) was more important than ‘max temperature of the warmest month’ (0.43) in the best model according to the z-scored beta (null deviance = 18 on 18 *d.f.*, residual deviance = 4.04 on 16 *d.f.*) (Table 5). Latitude (0.9) was also the most important variable in the second-best model compared to the ‘max temperature of the warmest month’ (0.7) and ‘annual mean temperature’ (0.65) (null deviance 18 on 18 *d.f.*, residual deviance = 3.469 on 15 *d.f.*) and in the third-best model (0.66) compared to ‘mean temperature of wettest quarter’ (0.35) and ‘max temperature of the warmest month’ (0.56) (null deviance 18 on 18 *d.f.*, residual deviance = 3.531 on 15 *d.f.*). None of the three best models included soybean variables. The two soybean variables, ‘time since first



| Source of variation             | d.f. | Variance components | Percentage variance | Fixation indices (P-value)         |
|---------------------------------|------|---------------------|---------------------|------------------------------------|
| <b>(a) COI-CytB</b>             |      |                     |                     |                                    |
| Among populations               | 20   | 9.550               | 56.57               | $\Phi_{ST} = 0.566$ ( $P = 0.00$ ) |
| Within populations              | 138  | 7.333               | 43.43               |                                    |
| Total                           | 158  | 16.883              |                     |                                    |
| Among biomes                    | 4    | 8.342               | 45.07               | $\Phi_{CT} = 0.450$ ( $P = 0.00$ ) |
| Among populations within biomes | 16   | 2.836               | 15.32               | $\Phi_{SC} = 0.278$ ( $P = 0.00$ ) |
| Within populations              | 138  | 7.333               | 39.61               | $\Phi_{ST} = 0.603$ ( $P = 0.00$ ) |
| Total                           | 158  | 18.511              |                     |                                    |
| <b>(b) ITS1 region</b>          |      |                     |                     |                                    |
| Among populations               | 18   | 0.050               | 20.18               | $\Phi_{ST} = 0.201$ ( $P = 0.00$ ) |
| Within populations              | 105  | 0.201               | 79.82               |                                    |
| Total                           | 123  | 0.252               |                     |                                    |
| Among biomes                    | 4    | 0.045               | 17.30               | $\Phi_{CT} = 0.173$ ( $P = 0.00$ ) |
| Among populations within biomes | 14   | 0.014               | 5.49                | $\Phi_{SC} = 0.066$ ( $P = 0.07$ ) |
| Within populations              | 105  | 0.201               | 77.21               | $\Phi_{ST} = 0.227$ ( $P = 0.00$ ) |
| Total                           | 123  | 0.260               |                     |                                    |

**Table 3.** Analysis of molecular variance (AMOVA) for genetic structure of *Euschistus heros* based on two concatenated mitochondrial genes (COI-Cytb) and ITS1 region.

| Geographic regions | Sample size (n) | Tajima's D | Fs de Fu  | $\tau$ (SD <sub>95%</sub> ) | SSD (P-value)      | r (P-value)       |
|--------------------|-----------------|------------|-----------|-----------------------------|--------------------|-------------------|
| <b>COI-Cytb</b>    |                 |            |           |                             |                    |                   |
| Pooled             | 159             | 0.445      | -23.799*  | 56.2 (5.30 - 89.24)         | 0.0268 (P = 0.08)  | 0.0041 (P = 0.98) |
| Atlantic Forest    | 50              | -1.920*    | -25.224*  | 6.4 (4.25 - 7.48)           | 0.0016 (P = 0.66)  | 0.0136 (P = 0.55) |
| Chaco              | 10              | -1.253     | -4.883*   | 6.0 (2.81 - 7.01)           | 0.0325 (P = 0.12)  | 0.0563 (P = 0.34) |
| Cerrado            | 37              | 0.858      | -18.632*  | 57.5 (3.66 - 79.53)         | 0.0394 (P = 0.09)  | 0.0160 (P = 0.80) |
| Amazon Forest      | 24              | 0.805      | -7.198*   | 58.7 (3.98 - 196.99)        | 0.0363 (P = 0.10)  | 0.0419 (P = 0.26) |
| Caatinga           | 38              | -0.067     | -19.748*  | 7.3 (5.36 - 11.13)          | 0.0199 (P = 0.03*) | 0.0071 (P = 0.84) |
| Lineage N          | 57              | -1.779**   | -24.741** | 10.0 (7.25 - 11.03)         | 0.0112 (P = 0.04*) | 0.0145 (P = 0.17) |
| Lineage S          | 102             | -2.271**   | -25.120** | 6.5 (4.50 - 7.27)           | 0.0033 (P = 0.11)  | 0.0144 (P = 0.20) |

**Table 4.** Neutrality test statistics and mismatch distribution analysis for *Euschistus heros* based on two concatenated mitochondrial genes (COI-Cytb). Lineage were previously defined by mitochondrial network.  $\tau$  = Expansion parameter; SSD = Sum of Squared Deviation; r = Harpending's Ruggedness Index.

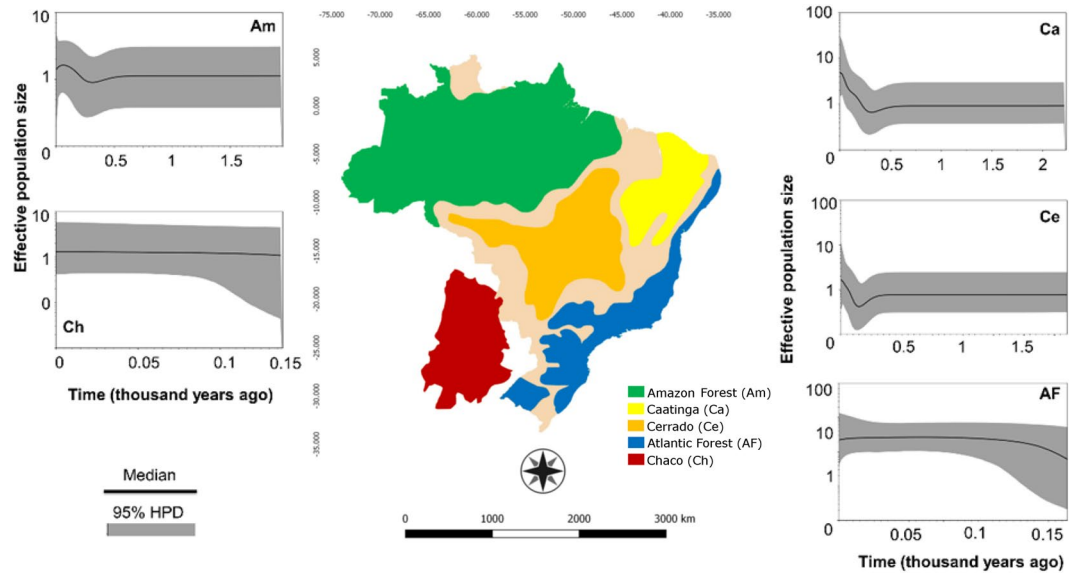
soybean harvest' and 'soybean expansion rate', ranked 7<sup>th</sup> and 16<sup>th</sup> in overall importance. The 'time since first soybean harvest' was strongly correlated with 'latitude' ( $r = 0.90$ ,  $d.f. = 17$ ,  $P = 0.000$ ), 'max temperature of the warmest month' ( $r = 0.61$ ,  $d.f. = 17$ ,  $P = 0.005$ ), and 'annual mean temperature' ( $r = 0.76$ ,  $d.f. = 17$ ,  $P = 0.000$ ).

## Discussion

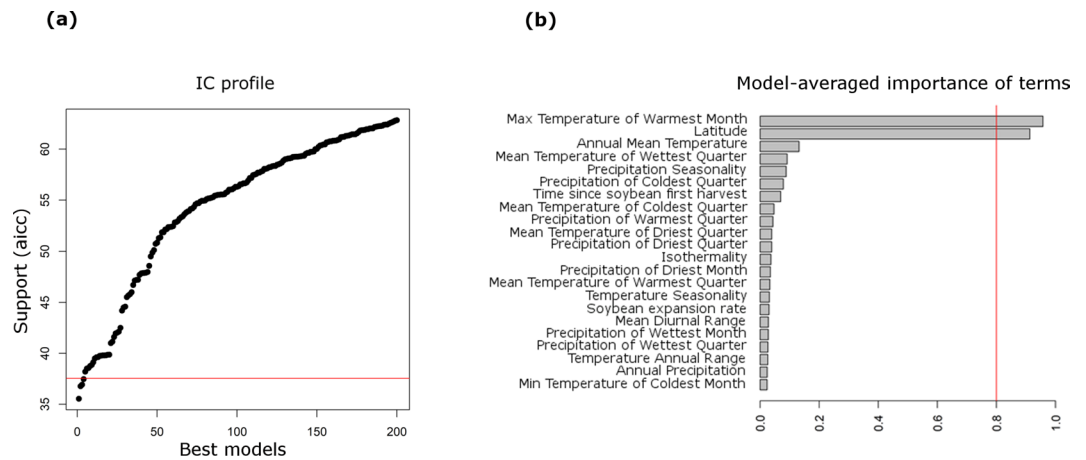
Our results revealed two divergent deep lineages of *E. heros* in South America. The two COI-Cytb haplotype groups are separated by 52 mutational steps and have an estimated genetic distance of 4.2% (K2P). The number of mutation steps separating the two *E. heros* lineages is exceptionally high, raising the question of the possible presence of cryptic species. Rolston<sup>35</sup> described a species in Middle America that is morphologically similar to *E. heros*, *Euschistus atrox* (Westwood) 1837, distributed in Colombia, Guiana, and Panama. However, examination of the external morphology, the sharing of ITS1 alleles by both lineages, and the admixture of lineages in the laboratory support the hypothesis of a single species that encompasses the two divergent lineages. Our results reveal the need for a thorough systematic review of the genus *Euschistus*.

The two *E. heros* lineages are geographically separated from one another, with one clade clustering the northernmost populations (i.e. northern and northeastern Brazil), and a second clade clustering the southernmost populations (i.e. southern and southeastern regions). Both mitochondrial lineages expanded to form a mixed zone upon secondary contact in the central and southwestern regions. It is not clear when the reunion occurred, but the formerly isolated populations seem to have come into contact before reproductive isolation was complete<sup>36,37</sup>. A related point to consider is that all but one of the northern haplotypes found in the Cerrado (CW) were phylogenetically grouped together in one clade, indicating a subgroup differentiation. The central-western (CW) subgroup likely occupied the region much earlier than the southern lineage arrived and before the first soybean fields were established. This is strong evidence against the hypothesis that the *E. heros* expansion was purely associated with the expansion of soybean cultivation during the 1970s.

The divergence time of the two main clades is estimated as occurring during the Pliocene (i.e. 4.5 Myr). This divergence seems to be associated with a cooling and drying of the global environment, which caused the



**Figure 3.** Bayesian skyline plot (BSP) showing population size dynamics for *Euschistus heros* in five biomes. The y-axis indicates effective population size ( $N_e$ ) scaled by mutation rate ( $\mu$ ) as a function of time. Black horizontal line shows median BSP estimate, and gray area shows upper and lower 95% highest posterior density limits. The map was based on Ab'Saber<sup>77</sup> and modified with GIMP 2.8.22.



**Figure 4.** Model selection and variable importance. (a)  $AIC_c$  values for the 200 top models for percentage of the southern lineage. Horizontal red line separates models with  $AIC_c$  less than 2 units away from the “best” model. (b) Relative importance of predictors, considering all models. Relative importance is the sum of weights for the models in which the variable appears. Vertical red line indicates the 0.8 cutoff used to separate important variables.

separation of the Amazon Forest from the southern part of the Atlantic Forest and the consequent expansion of grasslands and savannas<sup>38</sup>. Temperature cycles were also associated with more recent diversification events during the Pleistocene (i.e. differentiation of the CW group). Deep sequence divergence dating to the Pliocene is also reported for other organisms in the Neotropics<sup>39,40</sup>, and phylogeographic structure has been found in amphibians in the Atlantic Forest<sup>3</sup>, reptiles in the Cerrado<sup>41</sup>, and plants<sup>4,42</sup>.

Spatial genetic structuring by biomes was also found among subpopulations of *E. heros*. Separation into the Amazon Forest, Caatinga, Cerrado, Atlantic Forest and Chaco biomes seems to be the best way to explain the genetic variance hierarchically. Thus, separating insects by biomes can help us to understand the pattern of lineage mixing, diversity and demographic history. The haplotype diversity of *E. heros* was high and similar among biomes and lineages. This pattern is the result of the high number of private haplotypes found in *E. heros* populations in all biomes. The higher nucleotide diversity of lineage N compared to lineage S can be explained under the ‘historical climate’ stability models, where a stable environment such as the Amazon Forest can offer conditions for a population to persist, resulting in elevated intraspecific genetic diversity<sup>43,44</sup>. Unstable regions, on the other hand, would be associated with recent or multiple-event colonization, resulting in lower intraspecific genetic

| Competing models                        |   | AIC <sub>c</sub> | w <sub>i</sub> (AIC) |
|---|---|------------------|----------------------|
| <b>Presence of southern lineage (%)</b> |   |                  |                      |
| (1)                                     | +0.50 Latitude<br>−0.43 Max Temperature of Warmest Month  | 35.35            | 0.14                 |
| (2)                                     | +0.90 Latitude<br>−0.70 Max Temperature of Warmest Month<br>+0.65 Annual Mean Temperature             | 36.22            | 0.09                 |
| (3)                                     | +0.66 Latitude<br>+0.35 Mean Temperature of Wettest Quarter −0.56<br>Max Temperature of Warmest Month | 36.56            | 0.07                 |

**Table 5.** Model-selection results showing the top three models for the response variable ‘presence of the southern lineage’. Selected models are those with lowest AIC<sub>c</sub> values within two units from that of the best model. Best models are those with lower AIC<sub>c</sub> values. The Akaike weights ( $w_i$ ) represent the relative likelihood of a model and can be used to compare the strength of evidence of alternative models.

diversity and signatures of expansion<sup>4</sup>. Therefore, the northern biomes (Amazon Forest and Caatinga) were the most stable environment, while the Atlantic Forest was the least stable environment. Another consideration is that lineage S is associated with areas that have undergone intense transformation due to agricultural practices, and has experienced population dynamics linked with farming cycles and control tactics.

Although *E. heros*’ limited dispersal capacity likely helped to preserve the pattern formed during the late Tertiary and Quaternary as an outcome of the climate changes, the last 100 years were an important turning point for *E. heros* populations (i.e. soybean introduction and expansion of farming starting at the end of the 19<sup>th</sup> century). It is plausible that farming and trade routes have increased the admixture process in certain areas, especially the Cerrado and connecting areas, even though there are still large areas where the two lineages have not yet encountered each other, showing that the pattern is still well preserved.

Recent signals of expansion were detected for *E. heros* lineages and in all biomes sampled. The inferences regarding population growth were supported by the neutrality tests, the unimodal mismatch distribution and the demographic expansion parameters ( $\tau$ ). Spatial expansion is also occurring, given that no significant Raggedness values were found. Apart from differences in test sensitivity, the lack of full agreement between tests for *E. heros* in each biome might indicate a more complex scenario. Multiple processes affecting local diversity and the noise from human intervention causing population reduction, population subdivision, bottlenecks, and facilitation of dispersal resulting in the secondary contact might affect the precise demographic estimates for a species<sup>45–47</sup>.

We also conducted a Bayesian Skyline analysis to test the hypothesis of recent expansion in all biomes and to determine how the effective sizes of the populations behaved over time. The period of *E. heros* population growth in all areas overlaps with the period of intense changes caused by the increase of urban occupation and agricultural area in South America<sup>48</sup>. It may be that the resulting habitat loss not only did not affect *E. heros* populations negatively, but has even been advantageous. One possible hypothesis to explain the success of *E. heros* is shifting hosts from natural areas to agricultural fields, especially soybeans but also cotton and bean fields<sup>31</sup>. A second hypothesis is that one or more traits occur in a latitudinal cline<sup>49–51</sup>. The species’ association with environmental gradients should also be considered, given the possibility of differences in traits and adaptations such as reproductive diapause<sup>27,52,53</sup>.

We used environmental and soybean variables to make phylogeographic inferences to predict the predominance of lineage S over lineage N at a given location. Selected models had similar AIC<sub>c</sub> scores and considerably reduced the number of variables, down to 4 for the percentage of lineage S models. The two most important variables were the maximum temperature of the warmest month and the latitude. Temperature and photoperiod both affect this species, and might induce quiescence behavior and other possible differences in physiological responses. Latitude, on the other hand, can be correlated with geographic distance, environmental gradients, and agricultural gradients, as in the case of the soybean expansion. Our data support the predictions of the latitudinal-gradient hypothesis, even though distinct demographic scenarios can be expected at different times of *E. heros*’ evolutionary history. The time since the first soybean harvest correlates with latitude and other bioclimate variables, which likely decreased the importance of this variable in the model.

The reunion of the two long-separated lineages might have unforeseen consequences for one of the largest soybean-producing regions in the world. The two lineages are united again in central Brazil, where an agricultural revolution started in the 1970s and continues today, pushing soybean fields northward<sup>54</sup>. It is possible that the northern and southern populations of *E. heros* are exchanging adaptations in admixture zones. However, knowledge of the differences between the two lineages is limited, because their presence was unknown until this point<sup>55,56</sup>. The changing status of *E. heros* from a secondary to a primary pest in soybean crops and the reasons for this are poorly understood. In recent years, the increase of population densities in soybean fields, the shorter quiescence period, larger host range (i.e. damage in cotton crops) and pesticide tolerance/resistance have been frequently reported in *E. heros* populations<sup>30,31,57</sup>. These concerns increase in a scenario of GM soybean introduction, no-till management, and expansion to diversity-hotspot areas.

## Material and Methods

**Sample collection and DNA extraction.** One hundred fifty-nine specimens of *E. heros* were collected between 12/2015 and 07/2016 from 21 different localities across five South American biomes. Twenty sampling sites were in Brazil and one site in Paraguay. Specimens were collected as adults, from the canopy of soybeans,



using a beating cloth under the plants. Individuals were preserved in ethanol (>95%) at  $-20^{\circ}\text{C}$  until laboratory manipulation, after which the remaining tissue from all specimens was stored at  $-80^{\circ}\text{C}$ . DNA was extracted from the head tissue of an adult specimen, using the modified CTAB protocol<sup>58</sup>.

**PCR amplification and DNA sequencing.** Fragments of two mitochondrial and one nuclear region were amplified by polymerase chain reaction (PCR), using specific mitochondrial primers developed for this project and previously developed ITS1 primers<sup>59</sup>. The Cytochrome c Oxidase Subunit 1 (COI) fragment was amplified using the forward primer (5'-ACCGCACATGCATTTGTAATAA-3') and the reverse primer (5'-GTGGCTGATGTGAAGTATGCTC-3'), and the Cytochrome b (Cytb) fragment was amplified using the forward primer (5'-GGATATGTTTTACCTTGAGGACA-3') and the reverse primer (5'-GGAATGATCGTAAGATTGCGTA-3'). To amplify the ITS1 rDNA region (18S partial – ITS1 complete – 5.8S partial) we used the forward primer CAS18SF1 (5'-TACACACCGCCCGTCTACTA-3') and the reverse primer CAS5p8sB1d (5'-ATGTGCGTTCRAAATGTCGATGTTCA-3'). The PCR reactions were performed in a total volume of 25  $\mu\text{L}$  containing 80 ng total DNA, 1.5 mM/ $\mu\text{L}$   $\text{MgCl}_2$ , 0.1 mM/ $\mu\text{L}$  dNTPs, 0.4 pmol/ $\mu\text{L}$  of each primer, 1 U of Taq DNA Polymerase (Synapse Inc.) and Buffer (10  $\times$  Taq DNA Buffer). PCR cycles consisted of denaturation at  $95^{\circ}\text{C}$  for 3 min, followed by 35 cycles with denaturation at  $95^{\circ}\text{C}$  for 30 s, annealing at  $54^{\circ}\text{C}$  for 40 s, polymerization at  $72^{\circ}\text{C}$  for 1.5 min and final extension at  $72^{\circ}\text{C}$  for 10 min. Subsequently, the PCR products were separated on agarose gel (1.5% w/v) and observed under ultraviolet light. The amplicons were purified using 0.33  $\mu\text{L}$  EXO I, 0.33  $\mu\text{L}$  FastAp and 0.34  $\mu\text{L}$  of ultra-pure water together with 10  $\mu\text{L}$  of each PCR product, held at  $37^{\circ}\text{C}$  for 30 min, then at  $80^{\circ}\text{C}$  for 15 min. The PCR product Sanger sequencing was performed by the Animal Biotechnology Laboratory at ESALQ, University of São Paulo.

**Assembly of sequence datasets.** All sequences were aligned and edited manually using the software Sequencher 4.0.1 (Gene Codes Corp., Ann Harbor, MI, USA). To eliminate missing data, sequences were interrupted at 607 bp for the COI gene, 386 bp for the Cytb gene and 638 bp (18S partial – 52 bp; ITS1 complete – 416 bp; 5.8S partial – 170 bp) for the ITS1 region. There were no insertions or deletions in the sequences obtained. All sequences (datasets) obtained in this study were deposited in NCBI-GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) with the accession numbers MG651970 - MG652128 (COI), MG652129 - MG652287 (Cytb), and MG654513 - MG654636 (ITS1).

The presence of nuclear paralogs of mitochondrial origin (termed numts)<sup>60</sup> was inspected in the mitochondrial gene fragments, using the software MEGA v.5.2<sup>61</sup>. Three signatures of numts were searched: (i) indels that introduce frameshifts, (ii) out-of-place inframe stop codons that lead to premature termination of protein translation, and (iii) lack of codon position substitution bias toward the 3rd position, that lead to a higher rate of non-synonymous mutations. The presence of signatures (i) and (ii) is enough to consider a given sequence a numt. No numts were detected in the COI or Cytb sequences; therefore, we included all mitochondrial sequences in our analysis. The posterior analyses were performed using concatenated mitochondrial genes (COI-Cytb).

**Genealogical inferences.** The genealogical relationships between haplotypes of the mitochondrial and ITS1 regions were reconstructed by a network of median-joining haplotypes, using the PopArt software<sup>62</sup>. Preliminary analysis revealed two putative mitochondrial lineages associated with *E. heros* populations in Brazil and Paraguay. The genetic distance ( $D$ ) between the two mitochondrial lineages was inferred by dividing the haplotypes in two groups and calculating the 2-parameter Kimura method (K2P) in MEGA v.5.2 software<sup>61</sup>.

**Diversity statistics.** The diversity analysis was performed by dividing individuals into two groups according to the mitochondrial lineages, or into five groups according to the biome to which the individuals belonged: Amazon Forest, Cerrado, Caatinga, Atlantic Forest or Chaco. Number of haplotypes, haplotype diversity ( $h$ ), nucleotide diversity ( $\pi$ ) and mean number of nucleotide differences ( $S$ ) were estimated using the DNAsp v.5<sup>63</sup>.

**Divergence dating.** We estimated the relative age of divergence between the two mitochondrial lineages using the Bayesian relaxed phylogenetic approach implemented in BEAST v.1.8.4<sup>64</sup>, based on the combined mitochondrial genes. The substitution model was determined using the software PARTITIONFINDER version 1.1.1.1<sup>65</sup> that selected the GTR + G + I model. A strict molecular-clock model to estimate the substitution rate and coalescent tree priors set to the constant size model were implemented. We used the insect molecular clock (mean = 0.0177, SD = 0.001)<sup>66</sup> that corresponds to 3.54% pairwise divergence per Myr. Three independent runs were performed for 150 million generations, sampling every 1000 steps and discarding 20% as burn-in. TRACER v.1.6 was used to determine convergence, measure the effective sample size (ESS), and calculate the mean and 95% highest posterior density interval (HPD) for divergence times. Effective sample size (ESS) for all parameters exceeded 200, and the three runs converged to similar distributions. Runs were then combined with LogCombiner v.1.4.7<sup>64</sup>.

**Population Structure.** Variance Analysis (AMOVA) was performed in Arlequin with parametric bootstrap (1000 replicates) using a 5% significance level<sup>67</sup>. The analyses were conducted to examine the presence of genetic structure among individuals, considering all sites sampled (non-hierarchical), among populations according to the sampling location (populations) and among biomes in three hierarchical levels.

**Demographic statistics inferred for mitochondrial data.** Tajima's  $D$  and Fu's  $F_s$  neutrality tests were calculated using Arlequin v.3.5<sup>67</sup>. Both tests used 1,000 permutations using coalescing simulations. Fu's  $F_s$  statistic was considered significant at the 95% confidence level when the  $P$ -value was less than 0.02. For each biome, we also estimated tau ( $\tau$ ) with its 95% confidence intervals, using a generalized least-squares approach and 1,000 coalescent simulations in Arlequin v.3.5. The parameter  $\tau$  denotes the age of the expansion ( $t$ ), so that  $t = \tau/2 \mu$ ;

$u = \mu Lg^{68}$ . The parameter  $\mu$  represents the estimated mutation rate,  $L$  is the length of the sequence, and  $g$  is the generation time. For *E. heros*, we did not estimate  $t$  directly, because the number of generations per year cannot be estimated straightforwardly. Thus, if we assume that the estimated mutation rate has not changed in *E. heros* (substitution rate = 1.345%), then  $u$  will be constant and we can consider that a smaller  $\tau$  value indicates a newly established population, and a larger  $\tau$  value an older one.

We conducted a mismatch distribution analysis using a spatial expansion model. The sum of square deviations (SSD), raggedness index ( $r$ ) statistics, and their associated P-value were calculated using Arlequin v.3.5. A nonsignificant SSD value means that the hypothesis of population expansion cannot be rejected, and a nonsignificant raggedness index indicates a good fit of the data to the spatial expansion model. We also used a Bayesian Skyline Plot (BSP) in Beast to reconstruct the demographic history, using TRACER v.1.6, based on the COI-Cytb data using 10 groups. We used the same substitution model and molecular-clock model that were used to estimate the divergence time.

### Environmental features and soybean expansion modelling the current mitochondrial lineage distribution.

We used a model selection approach to identify and select variables that could be influencing the presence of a lineage at a given location<sup>69,70</sup>. Therefore, our response variable was the proportion of the southern lineage S calculated at each location as a percentage of the total composition. As predictor variables, we used 'latitude', 19 WorldClim variables based on all pixels of a CFR at 30 arc-second image<sup>71</sup>, and two soybean variables. The soybean variables consisted of the estimated time since the first harvest and the rate of increase of soybean production, given the cultivated area. We used linear regression to compile data from different sources and to estimate the two soybean variables, using the regression slope and the predicted year when the cultivated area was 100 hectares (Supplementary Table S1 and Fig. S1)<sup>54,72,73</sup>.

We evaluated the fit and plausibility of possible candidate models using glmulti<sup>74</sup>. We used a selection considering only the main effect, keeping the 200 best models. The criterion for selection was the corrected Akaike Information Criterion (AIC<sub>c</sub>)<sup>75</sup>. We selected models with AIC<sub>c</sub> less than two units away from the best model. We also evaluated the Akaike weight of the best models, to assess the probability that a model is the best<sup>76</sup>. All variables were standardized by z-score, and the significance of each predictor was assessed by a GLM. We also assessed the importance of each variable by summing the Akaike weight for the models in which the variable appeared. Variables that appear many times in the top models, tend to be more important. We used the cutoff of 0.8 to separate the most important variables under the weight criterion<sup>69,70</sup>.

### References

- Mittermeier, R. A. *et al.* Hotspots Revisited: Earth's biologically richest and most endangered terrestrial ecoregions. *Washington, Cemex* **1**, 392 (2005).
- Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. & Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858 (2000).
- Thomé, M. T. C. *et al.* Phylogeography of endemic toads and post-Pliocene persistence of the Brazilian Atlantic Forest. *Molecular Phylogenetics and Evolution* **55**, 1018–1031 (2010).
- Werneck, F. P. The diversification of eastern South American open vegetation biomes: Historical biogeography and perspectives. *Quaternary Science Reviews* **30**, 1630–1648 (2011).
- Haffer, J. Speciation in Amazonian forest birds. *Science* **165**, 131–137 (1969).
- Bennett, K. D. & Provan, J. What do we mean by 'refugia'? *Quaternary Science Reviews* **27**, 2449–2455 (2008).
- Hewitt, G. M. The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907–913 (2000).
- Carnaval, A. C. & Moritz, C. Historical climate modeling predicts patterns of current biodiversity in the Brazilian Atlantic forest. *Journal of Biogeography* **35**, 1187–1201 (2008).
- Leite, Y. L. R. *et al.* Neotropical forest expansion during the last glacial period challenges refuge hypothesis. *PNAS* **113**, 1008–1013 (2016).
- Costa, L. The historical bridge between the Amazon and the Atlantic Forest of Brazil: a study of molecular phylogeography with small mammals. *Journal of Biogeography* **30**, 71–86 (2003).
- Brieger, F. G. Contribuição à fitogeografia do Brasil com referência especial as orquídeas. *Anais do XX Congresso Nacional de Botânica*, 41–44 (1969).
- Moojen, J. Speciation in the Brazilian Spiny Rats (Genus *Proechimys*, Family Echimyidae). *University of Kansas Publications, Museum of Natural History* **1**, 301–406 (1948).
- Mori, S. A., Boom, B. M. & Prance, G. T. Distributional patterns and conservation of eastern Brazilian coastal forest tree species. *Brittonia* **33**, 233–245 (1981).
- Por, D. F. S. The Atlantic rain forest of Brazil. SPB Academic Publishing (1992).
- Schiesari, L., Waichman, A., Brock, T., Adams, C. & Grillitsch, B. Pesticide use and biodiversity conservation in the Amazonian agricultural frontier. *Philosophical Transactions of the Royal Society B* **368**, 20120378 (2013).
- Tabarelli, M., Da Silva, J. M. C. & Gascon, C. Forest fragmentation, synergisms and the impoverishment of neotropical forests. *Biodiversity & Conservation* **13**, 1419–1425 (2004).
- Paterniani, E. & Malavolta, E. La conquista del 'cerrado' en el Brasil. Victoria de la investigación científica. *Interciencia* **24**, 173–181 (1999).
- Klink, C. A. De Grão em Grão: O Cerrado Perde Espaço. Brasília, Brazil: *World Wide Fund for Nature (WWF-Brazil)*: 66 (1995).
- Klink, C. A., Macedo, R. H. & Mueller, C. C. Cerrado: Processo de ocupação e implicações para a conservação e utilização da sua diversidade biológica. Brasília, Brazil: *World Wide Fund for Nature (WWF-Brazil)*: 104 (1994).
- Spehar, C. R. Impact of strategic genes in soybean on agricultural development in the Brazilian tropical savannahs. *Field Crops Research* **1**, 141–146 (1995).
- Klink, C. A., Moreira, A. G. & Solbrig, O. T. Ecological impacts of agricultural development in the Brazilian Cerrados. The World's savannahs: Economic driving forces, ecological constraints and policy options for sustainable land in the Biosphere **12**, (eds Young, M. D. & Solbrig O. T.) 259–283 (Parthenon Publishing, London, 1993).
- Fearnside, P. M. Soybean cultivation as a threat to the environment in Brazil. *Environmental Conservation* **28**, 23–38 (2001).
- Brazilian Institute of Geography and Statistics – IBGE. <http://www.ibge.gov.br/home/> (2017).
- Panizzi, A. R. Growing problems with stink bugs (Hemiptera: Heteroptera: Pentatomidae): species invasive to the US and potential Neotropical invaders. *American Entomologist* **61**, 223–233 (2015).

25. Panizzi, A. R. & Slansky, F. Jr. Review of phytophagous pentatomids (Hemiptera: Pentatomidae) associated with soybean in the Americas. *Florida Entomologist* **68**, 184–214 (1985).
26. Aldrich, J. R. Dispersal of the southern green stink bug, *Nezara viridula* (L.) (Heteroptera: Pentatomidae), by hurricane Hugo. *Proceedings of the Entomological Society of Washington* **92**, 757–759 (1990).
27. Mourão, A. P. M. & Panizzi, A. R. Photophase influence on the reproductive diapause, seasonal morphs, and feeding activity of *Euschistus heros* (Fabr., 1798) (Hemiptera: Pentatomidae). *Brazilian Journal of Biology* **62**, 231–238 (2002).
28. Link, D. & Grazia, J. Pentatomídeos da região central do Rio Grande do Sul (Heteroptera). *Anais da Sociedade Entomológica do Brasil* **16**, 115–129 (1987).
29. Panizzi, A. R. Stink bugs on soybean in northeastern Brazil and a new record on the southern green stink bug, *Nezara viridula* (L.) (Heteroptera: Pentatomidae). *Neotropical Entomology* **31**, 331–332 (2002).
30. Soria, M. F., Degrande, P. E. & Panizzi, A. R. Algodoeiro invadido. *Revista Cultivar* **131**, 18–20 (2010).
31. Smaniotta, L. F. & Panizzi, A. R. Interactions of selected species of stink bugs (Hemiptera: Heteroptera: Pentatomidae) from leguminous crops with plants in the Neotropics. *Florida Entomologist* **98**, 7–17 (2015).
32. Rolston, L. H. Revision of the genus *Euschistus* in Middle America (Hemiptera, Pentatomidae, Pentatomini). Revisión del género *Euschistus* en Centroamérica (Hemiptera, Pentatomidae, Pentatomini). *Entomologica Americana* **48**, 1–102 (1974).
33. Panizzi, A. R. *et al.* Insetos da soja no Brasil. *EMBRAPA Soja – Boletim de Pesquisa e Desenvolvimento (INFOTECA-E)* (1977).
34. Saluso, A., Xavier, L., Silva, F. A. C. & Panizzi, A. R. An invasive pentatomid pest in Argentina: Neotropical brown stink bug, *Euschistus heros* (F.) (Hemiptera, Pentatomidae, Pentatomini). *Neotropical Entomology* **40**, 704–705 (2011).
35. Rolston, L. H. Revision of the genus *Euschistus* (Hemiptera: Pentatomidae) in Middle America. *Entomologica Americana* **48**, 1–102 (1974).
36. Coyne, J. A. & Orr, H. A. Speciation. *Sinauer*, (Sunderland, MA, 2004).
37. Sedghifar, A., Brandvain, Y. & Ralph, P. Beyond clines: lineages and haplotype blocks in hybrid zones. *Molecular Ecology* **25**, 2559–2576 (2016).
38. Haywood, A. M. & Valdes, P. J. Vegetation cover in a warmer world simulated using a dynamic global vegetation model for the Mid-Pliocene. *Palaeogeography, Palaeoclimatology, Palaeoecology* **237**, 412–427 (2006).
39. Bermingham, E. & Martin, A. P. Comparative mtDNA phylogeography of neotropical freshwater fishes: testing shared history to infer the evolutionary landscape of lower Central America. *Molecular Ecology* **7**, 499–517 (1998).
40. Da Silva, M. N. F. & Patton, J. L. Molecular phylogeography and the evolution and conservation of Amazonian mammals. *Molecular Ecology* **7**, 475–486 (1998).
41. Santos, M. G., Nogueira, C., Giugliano, L. G. & Colli, G. R. Landscape evolution and phylogeography of *Micrablepharus atticolus* (Squamata, Gymnophthalmidae), an endemic lizard of the Brazilian Cerrado. *Journal of Biogeography* **41**, 1506–1519 (2014).
42. Garcia, M. G. *et al.* Molecular evidence of cryptic speciation, historical range expansion, and recent intraspecific hybridization in the Neotropical seasonal forest tree *Cedrela fissilis* (Meliaceae). *Molecular Phylogenetics and Evolution* **61**, 639–649 (2011).
43. Carnaval, A. C. & Moritz, C. Historical climate modelling predicts patterns of current biodiversity in the Brazilian Atlantic forest. *Journal of Biogeography* **35**, 1187–1201 (2008).
44. Carnaval, A. C., Hickerson, M. J., Haddad, C. F., Rodrigues, M. T. & Moritz, C. Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. *Science* **323**, 785–789 (2009).
45. Bremer, J. R. A., Viñas, J., Mejuto, J., Ely, B. & Pla, C. Comparative phylogeography of Atlantic bluefin tuna and swordfish: the combined effects of vicariance, secondary contact, introgression, and population expansion on the regional phylogenies of two highly migratory pelagic fishes. *Molecular Phylogenetics and Evolution* **36**, 169–187 (2005).
46. Mari-Mena, N., Lopez-Vaamonde, C., Naveira, H., Auger-Rozenberg, M. A. & Vila, M. Phylogeography of the Spanish moon moth *Graellsia isabellae* (Lepidoptera: Saturniidae). *BMC Evolutionary Biology* **16**, 139 (2016).
47. Corrêa, A. S., Vinson, C. C., Braga, L. S., Guedes, R. N. C. & de Oliveira, L. O. Ancient origin and recent range expansion of the maize weevil *Sitophilus zeamais*, and its genealogical relationship to the rice weevil *S. oryzae*. *Bulletin of Entomological Research* **107**, 9–20 (2017).
48. Laurance, W. F., Sayer, J. & Cassman, K. G. Agricultural expansion and its impacts on tropical nature. *Trends in Ecology & Evolution* **29**, 107–116 (2014).
49. Tyukmaeva, V. I., Salminen, T. S., Kankare, M., Knott, K. E. & Hoikkala, A. Adaptation to a seasonally varying environment: a strong latitudinal cline in reproductive diapause combined with high gene flow in *Drosophila montana*. *Ecology and Evolution* **1**, 160–168 (2011).
50. Hut, R. A., Paolucci, S., Dor, R., Kyriacou, C. P. & Daan, S. Latitudinal clines: an evolutionary view on biological rhythms. *Proceedings of the Royal Society of London B: Biological Sciences* **280**, 20130433 (2013).
51. Paolucci, S., Zande, L. & Beukeboom, L. W. Adaptive latitudinal cline of photoperiodic diapause induction in the parasitoid *Nasonia vitripennis* in Europe. *Journal of Evolutionary Biology* **26**, 705–718 (2013).
52. Ruberson, J. R., Yeagan, K. V. & Newton, B. L. Variation in diapause responses between geographic populations of the predator *Geocoris punctipes* (Heteroptera: Geocoridae). *Annals of the Entomological Society of America* **94**, 116–122 (2001).
53. Pegoraro, M. *et al.* Geographical analysis of diapause inducibility in European *Drosophila melanogaster* populations. *Journal of Insect Physiology* **98**, 238–244 (2017).
54. C.N de Abastecimento – CONAB. S.H de Á Plantada, Produtividade e Produção, Relativas às Safras 1976/77 a 2015/16 de soja <http://www.conab.gov.br/conteudos.php?a=1252> (2017).
55. Lombaert, E. *et al.* Bridgehead effect in the worldwide invasion of the biocontrol harlequin ladybird. *PLoS one* **5**, e9743 (2010).
56. Rius, M. & Darling, J. A. How important is intraspecific genetic admixture to the success of colonising populations? *Trends in Ecology & Evolution* **29**, 233–242 (2014).
57. Sosa-Gómez, D. R. *et al.* Insecticide susceptibility of *Euschistus heros* (Heteroptera: Pentatomidae) in Brazil. *Journal of Economic Entomology* **102**, 1209–1216 (2009).
58. Clark, T. L., Meinke, L. J. & Foster, J. E. Molecular phylogeny of *Diabrotica* beetles (Coleoptera: Chrysomelidae) inferred from analysis of combined mitochondrial and nuclear DNA sequences. *Insect Molecular Biology* **10**, 303–314 (2001).
59. Ji, Y. J., Zhang, D. 'X. & He, L. J. Evolutionary conservation and versatility of a new set of primers for amplifying the ribosomal internal transcribed spacer regions in insects and other invertebrates. *Molecular Ecology Resources* **3**, 581–585 (2003).
60. Lopez, J. V., Yuhki, N., Masuda, R., Modi, W. & O'Brien, S. J. Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. *Journal of Molecular Evolution* **39**, 174–190 (1994).
61. Tamura, K. *et al.* MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance and maximum parsimony methods. *Molecular Biology and Evolution* **28**, 2731–2739 (2011).
62. Leigh, J. W. & Bryant, D. POPART: full-feature software for haplotype network construction. *Methods in Ecology and Evolution* **6**, 1110–1116 (2015).
63. Librado, P. & Rozas, J. DnaSPv5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452 (2009).
64. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **7**, 214 (2007).
65. Lanfear, R., Calcott, B., Ho, S. Y. & Guindon, S. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution* **29**, 1695–1701 (2012).

66. Papadopoulou, A., Anastasiou, I. & Vogler, A. P. Revisiting the insect mitochondrial molecular clock: the mid-Aegean trench calibration. *Molecular Biology and Evolution* **27**, 1659–1672 (2010).
67. Excoffier, L. & Lischer, H. E. L. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564–567 (2010).
68. Rogers, A. R. & Harpending, H. Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology & Evolution* **9**, 552–569 (1992).
69. Anderson, D. R. Information Theory and Entropy. *Model Based Inference in the Life Sciences: A Primer on Evidence*, 51–82 (2008).
70. Burnham, K. P. & Anderson, D. R. Model selection and multimodel inference: a practical information-theoretic approach. Springer Science & Business Media (2003).
71. Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**, 1965–1978 (2005).
72. Conceição, A. O. A expansão da soja no Rio Grande do Sul 1975-75. (Dissertation). Porto Alegre. *Universidade Federal do Rio Grande do Sul*. FEE (1986).
73. Pissinato, B. A cultura de cana-de-açúcar no Estado de São Paulo entre 1950 e 2010: evolução histórica da área e da produtividade. (Dissertation) *Universidade de São Paulo* (2013).
74. Calcagno, V. & de Mazancourt, C. glmulti: an R package for easy automated model selection with (generalized) linear models. *Journal of Statistical Software* **34**, 1–29 (2010).
75. Johnson, J. B. & Omland, K. S. Model selection in ecology and evolution. *Trends in Ecology & Evolution* **19**, 101–108 (2004).
76. Wagenmakers, E. J. & Farrell, S. AIC model selection using Akaike weights. *Psychonomic Bulletin & Review* **11**, 192–196 (2004).
77. Ab'Saber, A. N. Domínios morfoclimáticos e províncias fitogeográficas do Brasil. *Orientação* **3**, 45–48 (1967).

## Acknowledgements

We thank all collaborators who helped with insect sampling (Arias, O.; Azevedo, T.; Baptistella, M.G.; Bordignon, A.; Brogin, R.L.; Farinha, J.F.; Ferreira, V.E.N.; Gomes, J.; Gontijo, P.; Leite, N.A.; Lerner, M.A.; Machado, R.; Netto, J.; Omoto, C.; Pavan, G.B.; Rondon, M.; Rossi, G.D.; Silva, E.H.; Silva, L.B.; Tonelli, M.; Trebien, E.; Vidotti, M.S.; Vivan, L.M.; Zachrisson, B.; Zenati, L.). Our thanks to Julia Barbosa de Paiva for valuable assistance with the maps. We also thank the following entities for financial support: Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP (grant #2014/11495-3 and fellowship #2017/02393-0), Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq (grant #423040/2016-2 and fellowship #309167/2015-9) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES for fellowships.

## Author Contributions

A.S.C., C.O. and P.L.S. conceived and designed the study. E.M.G.C., F.N.S. and P.L.S. collected the data. A.S.C. and C.O. provided reagents and analytical tools. A.S.C., E.M.C.G. and PLS analyzed the data. A.S.C., E.M.C.G. and P.L.S. wrote the manuscript. All authors read, corrected, and approved the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-20187-6>.

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018