Whole genome comparisons reveal panmixia among fall armyworm (*Spodoptera frugiperda*) from diverse locations

Katrina A. Schlum[1], Kurt Lamour[2], Caroline Placidi de Bortoli[2], Rahul Banerjee[2], Scott J. Emrich[1,3], Robert Meagher[4], Eliseu Pereira[5], Maria Gabriela Murua[6], Gregory A. Sword[7], Ashley E. Tessnow[7], Diego Viteri Dillon[8], Angela M. Linares Ramirez[9], Komivi S. Akutse[10], Rebecca Schmidt-Jeffris[11], Fangneng Huang[12], Dominic Reisig[13], and Juan Luis Jurat-Fuentes[2]

[1]Genome Science and Technology Graduate Program, University of Tennessee, Knoxville, TN 37996, USA

[2]Department of Entomology and Plant Pathology, University of Tennessee, Knoxville, TN 37996, USA

[3]Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN, USA

[4]USDA-ARS Center for Medical, Agricultural and Veterinary Entomology (CMAVE), Insect Behavior and Biocontrol Research Unit, Gainesville, FL 32608, USA

[5]Departamento de Entomologia, Universidade Federal de Viçosa, Viçosa, MG 36570, Brazil

[6]Estación Experimental Agroindustrial Obispo Colombres, Las Talitas, Tucumán T4101XAC, Argentina

[7]Department of Entomology, Texas A&M University, College Station, TX 77843, USA

[8]Isabel Research Substation, Department of Agro-Environmental Sciences, University of Puerto Rico, Isabela, PR 00662, USA

23 [9]Lajas Research Substation, Department of Agro-Environmental Sciences, University of Puerto
24 Rico, Lajas, PR 00667, USA
25 [10]International Centre of Insect Physiology and Ecology, Nairobi, Kenya
26 [11]USDA-ARS Temperate Fruit & Vegetable Research Unit, Wapato, WA 98951, USA
27 [12]Department of Entomology, Louisiana State University Agricultural Center, Baton Rouge, LA
28 70803, USA
29 [13]Department of Entomology and Plant Pathology, North Carolina State University, Raleigh, NC
30 27695, USA
31

32 **Abstract**

33 The fall armyworm (*Spodoptera frugiperda* (J.E. Smith)) is a highly polyphagous
34 agricultural pest with long-distance migratory behavior threatening food security worldwide.
35 This pest has a host range of >80 plant species, but two host strains are recognized based on their
36 association with corn (C-strain) or rice and smaller grasses (R-strain). In this study, the
37 population structure and genetic diversity in 55 *S. frugiperda* samples from Argentina, Brazil,
38 Kenya, Puerto Rico and the United States (USA) were surveyed to further our understanding of
39 whole genome nuclear diversity. Comparisons at the genomic level suggest panmixia in this
40 population, other than a minor reduction in gene flow between the two overwintering populations
41 in the continental USA that also corresponded to genetically distinct host strains. Two maternal
42 lines were detected from analysis of mitochondrial genomes. We found members from the
43 Eastern Hemisphere interspersed within both continental USA overwintering subpopulations,
44 suggesting multiple individuals were likely introduced to Africa. Comparisons between
45 laboratory-reared and field collected *S. frugiperda* support similar genomic diversity, validating

46 the experimental use of laboratory strains. Our research is the largest diverse collection of United
47 States *S. frugiperda* whole genome sequences characterized to date, covering eight continental
48 states and a USA territory (Puerto Rico). The genomic resources presented provide foundational
49 information to understand gene flow at the whole genome level among *S. frugiperda*
50 populations.
51

52 **Introduction**

53 Larvae of the fall armyworm, *Spodoptera frugiperda* (J.E. Smith, 1797) (Lepidoptera:
54 Noctuidae), are highly polyphagous agricultural pest affecting key food and fiber staples such as
55 corn (*Zea mays* L.), cotton (*Gossypium spp.* L.), sorghum (*Sorghum bicolor* L.), rice (*Oryza*
56 *sativa* L.), and vegetable crops [1]. Damage by *S. frugiperda* in its native subtropical range in the
57 Americas resulted in 15-100% yield loss, depending on the level of infestation [2]. In most of the
58 continental USA this pest does not diapause or survive cold winters. Overwintering populations
59 in southern Texas and Florida migrate yearly northward over several generations to populate
60 northern regions in USA and Canada [3].

61 For more than a decade, effective control of *S. frugiperda* in the Western hemisphere has
62 been provided mainly by genetically modified corn and cotton producing insecticidal Cry and
63 Vip3A proteins from the bacterium *Bacillus thuringiensis* (*Bt*). However, resistance to Cry1F,
64 Cry1Ab and Cry1A.105 insecticidal proteins quickly developed in Puerto Rico and the
65 continental USA (Florida and North Carolina) [4, 5], as well as Brazil [6] and Argentina [7].
66 More recently, the economic importance of this pest has further increased with its introduction in
67 sub-Saharan Africa [8], subsequent spread to India and Southeastern Asia [9], and more recently
68 Australia [10]. An estimate using data from twelve African countries indicates that yield losses

69 resulting from *S. frugiperda* injury could be 21-53% of their annual corn production, which
70 equals to a loss of US$2.5-$6.2 billion in losses [11]. While yet to be detected in molecular
71 screens [12], resistance alleles to *Bt* toxins could have been carried by invasive *S. frugiperda* to
72 the Eastern hemisphere.

73 Populations of *S. frugiperda* are composed of sympatric mixtures of two genetically
74 differentiated strains based on host preference, a "rice" (R) strain feeding on rice, millet and
75 smaller grasses, and a "corn" (C) strain feeding preferentially on corn and sorghum [13, 14].
76 There is evidence supporting that this differentiation involves reproductive incompatibility [15]
77 and differential susceptibility to xenobiotics [16]. Signatures of positive selection for genes
78 involved in chemoreception, detoxification, and digestion were also detected in whole-genome
79 comparisons of nine C- and R-strain individuals [17]. Being morphologically indistinguishable,
80 C and R individuals are discriminated using genetic markers located on the mitochondrial *COI*
81 (cytochrome oxidase subunit I) and sex-linked (Z chromosome) *Tpi* (*triosephosphate isomerase*)
82 genes [18-21]. However, host preference is not absolute and discrepancies among haplotype
83 markers have been reported. For example, the *Tpi* marker agrees with assortative mating and host
84 assessments [22] in describing *S. frugiperda* samples from corn in the Eastern Hemisphere as C-
85 strain, yet a predominant *COI*-R strain marker is obtained in these collections [9]. These
86 conflicting results may reflect interstrain hybridization [22] or be driven by maternally inherited
87 symbionts that skew the distribution and diversity of certain haplotypes [23]. Additional factors
88 influencing these conflicting haplotyping results may include incomplete lineage sorting and
89 plant host behavioral plasticity.

90 Nucleotides found at *COI* host strain marker sites have been commonly grouped into
91 "haplotypes" with relative proportions associated to geographic origin [24], enabling

92 differentiation of "Texas" and "Florida" overwintering *S. frugiperda* populations in the USA.
93 Migratory studies with these haplotypes determined that the "Texas" population is distributed
94 throughout South, Central and North America [25, 26], while the "Florida" population locates to
95 Florida and the Caribbean but migrates through the eastern USA seaboard to reach Canada [27,
96 28]. Haplotype profiling of *S. frugiperda* from Africa, India and Southeastern Asia supports
97 introductions in Africa of individuals from the "Florida" population and subsequent spread to
98 Asia [9, 12].

99 In this work, we collected and sequenced 55 genomic DNA (gDNA) samples of *S.*
100 *frugiperda* from three continents, with an emphasis on C-strain individuals from North and South
101 America. Genome-wide comparisons allowed testing for gene flow between geographically
102 distant *S. frugiperda* populations and comparing genetic diversity between field and laboratory-
103 reared *S. frugiperda.* We find that based on F_{st} values there is low genetic differentiation between
104 Texas vs. Florida and that samples from these locations display similar genetic nucleotide
105 diversity. Host strain differentiation was supported by nuclear and mitochondrial genome
106 differences. Comparison of mitochondrial genomes detected two clusters where the clusters
107 correlate by and large with host strain and resistance to the Cry1F toxin. The exceptions to this
108 correlation suggest that C-strain individuals based on nuclear genome have likely mated with R-
109 strain individuals. We also observed no detectable reduction in genetic diversity in well-
110 established *S. frugiperda* laboratory-reared colonies, which supports using laboratory strains of
111 this pest as a model for field populations.

112

113 **Methods**

114 **Samples and strain typing**

115 Details of the 55 laboratory and field-collected *S. frugiperda* samples analyzed are found
116 in Supplementary Table 1. Adults (moths) were captured using sex pheromone baited traps [29],
117 with most collection sites near corn plantings to optimize trap capture efficiency of C-strain
118 males. Laboratory-reared samples were obtained from rearing facilities. Susceptibility to Cry1F
119 toxin from *B. thuringiensis* for these laboratory samples was tested in bioassays presented
120 elsewhere [5, 30-32]. Larval samples were collected at field and laboratory rearing locations and
121 used when available. The collected specimens were identified as *S. frugiperda* by morphology
122 features [33] and stored at -20°C until required for analysis.

123 We followed a sample naming protocol that included the first three letters representing
124 the country of origin (Bra for Brazil, Arg for Argentina, USA for United States of America, Ken
125 for Kenya, and Pue for Puerto Rico), the next two letters representing the first two letters of the
126 state/province of origin (SP for Sao Paulo, TX for Texas, etc. XX when unknown), and the third
127 letter representing the Cry1Fa susceptibility phenotype when known ("r" for resistant, "s" for
128 susceptible and "u" for unknown). A number was used when necessary to differentiate samples
129 with the same geographic origin and Cry1F susceptibility phenotype.

130 Host strain (C versus R) was determined based on sequence identity at specific marker
131 positions in reference mt*CoI*1164 and *Tpi*183 sequences, as described elsewhere [12]. The
132 mitochondrial *COI1287* marker identified all samples as C-strain, so the nuclear *Tpi*183 marker
133 was used as a more reliable marker of host strain [34]. The mitochondrial *COI*1164 marker was
134 not used because it did not fall under any of the pre-defined *COI* haplotypes [34] for any of the
135 samples.

136 **DNA extraction, library preparation, and sequencing**

137 Genomic DNA was isolated from individual carcasses of fifth instar *S. frugiperda* larvae
138 after dissecting the gut tissue or from legs or heads of adults using the Pure Link Genomic DNA
139 mini kit (Invitrogen), following manufacturer's protocols, and then quantified using a Nanodrop
140 spectrophotometer (Thermo Scientific). Extracted DNA was then sheared randomly to between
141 250-500 bp using a Covaris M220 focused ultrasonicator (Woburn, MA), according to
142 manufacturer's instructions. The fragmented DNA was then ligated with dual-indices using a
143 KAPA Hyper prep PCR-free library kit (Roche) according to the manufacturer's directions. The
144 ligated fragments were quantified using quantitative PCR and a KAPA Library quantification kit
145 (Roche), and then submitted for sequencing on an Illumina HiSeqX device running a 2 x 150 bp
146 paired-end configuration (Admera Health, NJ). Raw paired-end sequence reads for each sample
147 (55) are available for download in the NCBI Sequence Read Archive (SRA) under
148 SRR12044614-SRR12044668 with associated metadata available under NCBI BioProject id
149 PRJNA640063. The raw data was processed to remove low quality reads using the CLC
150 Genomics Workbench v9.5.2 (Qiagen) trim function with default parameters.
151
152 **Filtering and mapping**
153 Raw reads were quality-trimmed at both ends, filtered for adapter sequences and error
154 corrected for known Illumina artifacts and PhiX sequencing control using BBDuk
155 (https://jgi.doe.gov/data-and-tools/bbtools/) with options trimq = 15 and filter = 23. Further
156 quality control and confirmation that adapter and non-relevant (contaminants, primer artifacts)
157 sequences were removed from filtered reads were performed using FastQC [35]. Given the
158 prevalence of C-strain among our samples and similar number of variants called when using both
159 *S. frugiperda* corn and rice genomes [36], we mapped the remaining reads to the *S. frugiperda*

160 corn reference genome (v3.1) downloaded from

161 https://bipaa.genouest.org/sp/spodoptera_frugiperda_pub/ (assembled length of 312 Mb across

162 29,949 scaffolds [17]) using a Burrows-Wheeler aligner algorithm (bwa-mem) [37]. Variants

163 were then called using the SAMtools mpileup utility [38], resulting in detection of 126,977,977

164 single nucleotide polymorphisms (SNPs) and indels. The unfiltered dataset contained

165 120,398,863 SNPs. Three samples (USATXu1, BraSPr2 and BraSPr1) were removed from SNP

166 analysis because they had missing variant call rate greater than 50%, resulting in 52 samples

167 being used for SNP PCA-based analyses. For nucleotide diversity analyses, SNPs were filtered

168 using VCFtools [39] based on both minor allele frequency being less than or equal to 0.05 and

169 presence in at least 50% of the 52 samples surveyed. This filtering further reduced SNP number

170 to 204,784.

171

172 **Phylogenomics**

173 The Mash distance method was used to measure pairwise dissimilarity of genomic

174 sequences [40]. One key advantage of sketching-based approaches is that they neither require *de*

175 *novo* genome assembly nor a pre-existing reference genome to identify related individuals. The

176 Mash method estimates sequence similarity via the Jaccard similarity coefficient (the ratio of

177 shared k-mers to total k-mers) for a "random" subset of k-mer pairs for each genome surveyed.

178 The Jaccard similarity (J) is then used to calculate dissimilarity between two genomes (*D*) as D =

179 1-J. All *S. frugiperda* individuals were sketched with Mash [40] using a 10,000 sketch size, k-

180 mer size of 21 and minimum copy of each k-mer equal to 2, per previous recommendations of

181 representing genomes with minimal computational costs [40]. All sample sketches were screened

182 against the Mash Refseq database and other *Spodoptera* species (*Spodoptera exigua*,
183 GCA_011316535.1; and *Spodoptera litura*, GCA_002706865.2) as outgroups.
184
185 **SNP diversity**
186 Weighted F_{st} values as indicators of total genetic variance in a subpopulation were
187 calculated using VCFtools [39] by the Weir and Cockerham estimator on 1 kbp, 3 kbp and 5 kbp
188 windows. The variation in window size was used to confirm that there was little to no change in
189 F_{st} when the window was varied. For F_{st} calculations using the mitochondrial genome, each base
190 pair position was compared. Nuclear genome (pi) values were calculated using 1 kbp windows
191 when comparing corn vs. rice host strain, and 3 kbp when comparing F_{st} values for overwintering
192 FL/TX vs. USA laboratory samples. We also used 3 kbp windows for F_{st} comparisons across the
193 nuclear genome with populations defined by mitochondria haplotype network groupings, Mash
194 genome defined "haplotypes", and host strain haplotypes. For all nucleotide diversity analyses, a
195 total of 3,558,854 SNPs were used to assess diversity using VCFtools [39].
196
197 **Mitochondrial genome assembly**
198 Mitochondrial sequences were extracted from the whole genome sequence reads using
199 NOVOplasty [41] with the mitochondrial partial *COI* sequence MH932092.1 as the seed. A
200 continuous complete mitochondrial chromosome of around 15 kbp was generated for all 55
201 samples, except for BraBAr5, which was excluded from further analysis. The whole
202 mitochondrial FASTA sequences were assembled for each sample and then aligned via bwa-
203 mem (https://github.com/lh3/bwa) with default options. Mitochondrial variants were called using
204 SAMtools mpileup based on the *S. frugiperda* corn reference genome at LepidoDB

205 (https://bipaa.genouest.org/sp/spodoptera_frugiperda_pub/). The vcf files were compressed using
206 SAMtools bgzip, indexed using SAMtools tabix and merged using SAMtools BCFtools merge
207 [38]. Missing genotypes were replaced with sequence at the mitochondrial reference genome
208 from the *S. frugiperda* corn genome at LepidoDB. The vcf files were then filtered by removing
209 16 multi-allelic sites and 92 uninformative sites due to being found in 2 samples or less via the
210 informloci function in the adegenet R package [42], resulting in 298 sites remaining. These
211 remaining mitochondrial sites were used to build a haplotype network using the R package pegas
212 haploNet function [43] with Euclidean distance and an infinite site model used for building the
213 network.
214
215 **Structure**
216 PLINK –make-bed [44] was used to generate the input file to FastStructure [45], a
217 variation of Structure made for larger SNP datasets. FastStructure is a generative model-based
218 approach based on Hardy-Weinberg equilibrium assumptions between alleles and linkage
219 disequilibrium between genotyped loci. The FastStructure script structure.py was used to
220 determine k, the number of assumed populations or genetic groups that share a subset of allele
221 frequencies from k = 2 to 10. Then k =2 was chosen by the FastStructure's choosek.py script for
222 model complexity that maximizes marginal likelihood, and k = 1 for model components to
223 explain structure. Additionally, PCA eigenvalues and eigenvectors were generated using PLINK
224 2.0 [44, 46] and visualized in R using ggplot2 [47].
225
226 **Results**
227 **Host strain typing**

228 The majority (26, 47.3%) of the 55 *S. frugiperda* samples used in this study originated
229 from field collections representing eight states in the USA. The average mapping rate for all
230 samples was 88.2% with a range of between 60.62% to 96.82% range per SAMtools flagstat
231 [38]. Identification of the C- and R- host strains for each individual was performed using nuclear
232 *Tpi* and mitochondrial *COI* genetic markers described elsewhere [26]. Given that sample
233 collections were from or around cornfields, we expected to observe a majority of C-strain
234 samples. The *COI*1164 marker identified all samples as C-strain. In contrast, typing based on the
235 nucleotides at the *TPi*183 site resulted in 42 samples identified as C-strain individuals (C/C), 8 as
236 R-strain individuals (T/T), 3 as interstrain hybrids (C/T) and 2 that were undetermined due to
237 low sequencing coverage at this locus. All eight identified R-strain samples were from the USA
238 (3 from Texas, 2 from Maryland and Tennessee each, and 1 from Puerto Rico).
239
240 **Whole genome diversity, nucleotide diversity and differentiation**
241 The FastStructure tool was used first to determine possible population structure among
242 the samples. The number of groups (k) was analyzed for k = 1 to k = 10, but all samples
243 clustered as one group even when the number of groups was increased. This lack of clear
244 population structure was further confirmed using principal component analysis (PCA) on a
245 filtered set of 204,784 nuclear SNPs (see Methods). PC1 through PC10 was plotted and no
246 structure was found based on country of origin, host strain or susceptibility to the Cry1F
247 insecticidal protein from *Bt* (data not shown for PC2 to PC10). We noted that PCA using the first
248 two principal components uncovered weak substructure within the USA and Brazil samples,
249 where the majority of samples clustered in a larger cluster and four samples (USAFLu9 from
250 Florida, and BraMGr2, BraBAr2, and BraMGr5 from Minas Gerais in Brazil) grouped together in a

251 smaller cluster with 1 Florida sample (USAFLr2) the most differentiable from the two clusters
252 (Fig. 1). All these small cluster samples except USAFLu9 are laboratory samples, and therefore
253 this grouping may be a result of the side effect of genetic drift associated with the colony
254 formation process.

255 Given the overall lack of clear population structure observed based on origin, we next
256 applied genome sketching [40] and traditional distance-based hierarchical clustering as a third
257 strategy to analyze these samples. The hierarchical cluster dendrogram defined by the genome-
258 wide distance estimates based on the fraction of shared k-mers between each pair of genomes
259 was cut at height 0.05 as it resulted in at least two primary clusters (Texas vs. non-Texas) among
260 the samples analyzed (Fig. 2). This cut height generated 8 clusters, 6 containing less than 3
261 samples and were merged to the major cluster, while the remaining 10 samples were left in the
262 remaining 2 minor cluster. The first cluster was comprised of samples from the USA
263 (USATXu2-5 from Texas, USATNu1-2 from Tennessee, USAMDu1-2 from Maryland,
264 USAFlu2 from Florida and PueSlu1 from Puerto Rico), while the second cluster included
265 individuals from all countries sampled with the majority of samples from USA representing
266 Florida/eastern seaboard. The three samples from Puerto Rico were interspersed across both
267 clusters and always clustered with Florida samples, yet genomic distances between clusters were
268 small (< 0.05 Mash distance). Similar to the PCA clusters, these two clusters did not correlate
269 with host strain designation or geographic origin, with the first cluster including both R- and C-
270 strain samples and the second cluster containing both R- and C-strain samples as well as three
271 interstrain hybrids (Supplemental Fig. S2).

272 Mash distance is a proxy for genome pairwise distance, with 0 representing identical
273 genomes (all k-mers that comprise the sketches are present in both sequences). The Mash

274 distances across all countries surveyed were averaged to be <0.05 (Table S2, range of average
275 distance was 0.043 to 0.045), probably reflecting recent population expansion. Moreover, these
276 Mash distances support that R and C-strains are members of the same species, as the estimated
277 genetic variation is comparable to at least 95% average nucleotide identity (ANI), which is the
278 ANI cutoff for eukaryotic species [40].
279
280 **Gene flow between overwintering populations in the USA**
281 Previous work based on the ratios of four pre-determined *COI* haplotypes suggested that
282 the overwintering Florida *S. frugiperda* population is mostly reproductively isolated from the
283 overwintering population in Texas, with genetic exchange mostly occurring at the north and
284 south ends of the Appalachian mountain chains [27]. As mentioned above, we observed that C-
285 strain samples from Texas and Florida/eastern seaboard did not separate based on hierarchical
286 clustering of genome-wide distances, and Florida samples were found across both groups (Fig.
287 2). Further, when comparing the two R-strain samples from Texas and the fourteen C-strain
288 samples from Florida/eastern seaboard, we detected a low pairwise weighted F_{st} value (0.0111)
289 based on 3,558,854 nuclear sites. This very low F_{st} value suggests no population separation at the
290 nuclear level, supporting recent population separation or ongoing gene flow between Texas and
291 Florida overwintering populations, independently of host strain.
292 Based on the same Texas R-strain and Florida C-strain groupings, mean diversity was
293 determined. Mean diversity for the field samples from Texas R-strain and Florida C-strain
294 samples was estimated at 0.0631 and 0.0556, respectively (Fig. 3). Next, we determined the F_{st}
295 value based on 406 mitochondrial SNPs found across all samples in these groups. The
296 mitochondrial F_{st} for Texas R-strain and Florida C-strain samples was estimated at 0.7961. This

297 F_{st} estimate being close to 1 suggests genetic differentiation between the R and C strains is the
298 strongest at the mitochondrial level, possibly due to two main ancestral lineages. Although we
299 see a stronger difference between geographic regions than between R and C strains in this study,
300 the estimated mitochondrial to nuclear F_{st} ratio of 46.8 is consistent with previously reported sex-
301 based differences in dispersal [48] and/or mating preference differences between strains [49].
302 Even so, our data suggest hybridization occurs and some C-strain nuclear genomes have a clearly
303 R-strain associated mitochondrial genome (see Fig. 5).
304
305 **Laboratory vs. field diversity**
306 Laboratory-reared strains of *S. frugiperda* are commonly used for diverse physiological,
307 toxicological and genetic studies, yet their genetic resemblance to field populations has not been
308 investigated. To investigate the differentiation level between two well established laboratory
309 strains (>20 years) originating from the same locations in the USA (Mississippi, MS) and all
310 field USA populations surveyed, we estimated a weighted F_{st} (fixation index). As shown in Table
311 1, there was moderate differentiation between the MS laboratory strains and field USA field
312 populations, with an F_{st} range of 0.0056 to 0.1109. Moderate differentiation was also observed
313 when comparing the MS laboratory strains with other more recently established (<10 years)
314 laboratory populations from North Carolina (NC) and Florida (FL), with an F_{st} range of 0.0024 to
315 0.0874. The samples least differentiated (F_{st} = 0.0631) from the MS laboratory populations were
316 collected from TX/MN locations, while the most distinct (F_{st} = 0.1263) were the samples from
317 Maryland (MD). Overall, these moderate F_{st} values [50] suggest small differences at the allelic
318 level between laboratory and field-collected groups. Genetic diversity in the two laboratory MS

319 colonies was slightly different but comparable to diversity in field USA populations (F_{st} = 0.0981
320 on average) and to either of the two overwintering populations (F_{st} = 0.0594).

321 Additionally, we surveyed overall genomic diversity using 3 kbp windows and estimated
322 mean diversity for the same two MS laboratory strains compared to all field and laboratory *S.*
323 *frugiperda* samples surveyed. Mean nuclear diversity was 0.0011 for Benzon (USAMSs1) and
324 SIMRU (USAMSs2) laboratory strains (n=2), compared to 0.0015 for field Texas samples (n=7),
325 and 0.0015 for field Florida/eastern USA samples (n=14).

326

327 **Diversity across the mitochondrial genome**

328 We investigated the maternal lineage among all but one of the *S. frugiperda* samples (due
329 to lack of a complete mitochondrial genome in sample BraBAr4) by comparing genetic diversity
330 of whole mitochondrial (mt) genomes. Using a total of 390 SNPs, we generated a haplotype
331 network as described in Materials and Methods (Fig. 4). Consistent with a large amount of
332 standing variation in the species, even the mitochondrial sequences appeared unique with all but
333 eight samples forming a unique haplotype. These exceptions included three haplogroups of only
334 Brazil samples, a second group of Kenya and USA samples, and a third haplogroup of samples
335 from Kenya and two haplogroups of Florida, USA and Puerto Rico samples (Fig. 4). There were
336 two main "haplotype" clusters that separated in the network, with average genetic difference
337 between the two clusters of 29.53, supporting evidence for two main maternal lineages within the
338 samples surveyed. Using these two haplotype clusters as labels, we mapped the mt haplotype
339 clusters to the two whole genome sequence (WGS) "haplotype" clusters found in Mash distance-
340 based tree against a mitochondrial variant UPGMA distance-based tree (Fig. 5). In accordance
341 with the previously estimated F_{st}, the mt and WGS clusters did not correlate completely.

342 However, all samples in one of the mt haplotypes clustered in one of the WGS haplotypes ("one"
343 in purple in Fig. 5), while the other mt haplotype contained a mix of WGS haplotypes.
344 Interestingly, and while not all samples were phenotyped, all samples classified as resistant to the
345 Cry1F toxin clustered in one of the mt haplotype groups.

346

347 **Discussion**

348 We report on the genome-level comparison of 55 *S. frugiperda* samples from diverse
349 locations, mostly from the native range in the Americas but also including samples from an
350 African location (Kenya) where the insect is a devastating exotic pest. While using the C-strain
351 *S. frugiperda* genome for alignments could have underestimated diversity in R-strain individuals,
352 this approach is supported by previous surveys using both corn and rice genomes with minimal
353 differences found based on the number of variants called and F_{st} values [36].

354 The population structure of the 55 genomic *S. frugiperda* samples was surveyed using
355 parametric and non-parametric methods. Based on FastStructure (parametric), no population
356 structure was found either within samples from the Eastern Hemisphere or between the Eastern
357 and Western Hemisphere samples. Since methods based on parametric estimation of allele
358 frequencies methods are sensitive to sample size, one possibility is that the lack of structure
359 detected is a reflection of the relatively small number of R-strain samples (only 8 individuals and
360 potentially 2 hybrids) analyzed. A more even distribution would address potential limited
361 sampling of some locations surveyed, especially samples representing the Eastern Hemisphere
362 (e.g. Africa).

363 Although PCA based on the nuclear genome agreed with the little to no population
364 structure as found by FastStructure, the Mash distance-based analysis on the nuclear genome and

365 SNP Hamming-based distance method based on the mitochondrial genome detected two
366 subpopulations within the samples surveyed, which associate with the previously speculated two
367 subpopulations in the continental U.S. based on four mitochondrial *COI* haplotypes [51]. The
368 two clusters clearly indicate that the corn and rice strain individuals are more similar locally than
369 between geographic areas, suggesting some level of hybridization between host strains in the
370 USA. Further, we have analyzed individuals that appear to have a C-strain nuclear genome with
371 an R-strain associated mitochondria (maternal lineage). This is consistent with the conflicting
372 host strain assignments using specific genetic markers between R-strain and interstrain hybrids
373 that was previously reported [22]. We still observe host differentiation at the nuclear genome
374 level, though, and may suggest current single locus markers are not sufficient for surveying all *S*.
375 *frugiperda*.

376 Our analysis included samples collected from locations (Maryland & Tennessee) close to
377 predicted hybridization zones for TX and FL overwintering *S. frugiperda* populations [27], and
378 as expected these associated with both geographic regions. There were also three exceptions to
379 the geographic separation: two Texas samples were found in the predominately Florida
380 represented group and one Florida sample was found in the predominately Texas population.
381 This observation suggests that the reproductive isolation between Texas and Florida/Eastern
382 seaboard is not absolute and mating may be occurring in overlapping locations, as previously
383 suggested [27]. Future work should include increased sampling from potential hybrid zones
384 (Georgia/New York) and Texas. The Mash distance-based clusters suggested that Puerto Rico
385 samples are closer to Florida, since they were found in sub-clusters with Florida-based samples.
386 However, due to low representation and small genomic distances between each cluster, our data

387 is not able to show conclusive support for migration models suggesting significant exchanges
388 between *S. frugiperda* from Puerto Rico and Florida [28].
389 The low amount of genetic variation across the whole genome between the C- and R-
390 strains is in line with previous reports, [17] and suggest that the genome differentiated after the
391 ecological differentiation of host strains or that barriers are isolating C- and R-strains (i.e.
392 prezygotic reproductive isolation or mating preference). In testing this hypothesis, future
393 research should involve analyses of a larger sample size of C- and R-strain individuals from
394 Texas and Florida.
395 Mitochondrial F_{st} estimates of 0.7961 obtained between C- and R- host strains were
396 highly similar to the F_{st} value (0.938) from a prior analysis of 9 corn and 9 rice sympatric
397 samples from Mississippi (USA) [17], and the nuclear DNA F_{st} found in our study (0.0173) is
398 also similar to the previous study's estimate (0.019). The two mitochondrial haplotypes found in
399 our analysis could be explained by selective mating or some form of reproductive
400 incompatibility. Although susceptibility to Cry1F was not known for all samples, we observed
401 that all Cry1F-resistant strains clustered to one of the mitochondrial haplotypes, despite their
402 geographic origin. While speculative given the comparatively low number of confirmed Cry1F-
403 resistant samples, this grouping may reflect partitioning of resistance with the haplotype group
404 only containing C-strain samples. Further work including Cry1F-resistant R-strain samples
405 would be necessary to test for associations between mitochondrial haplotype and resistance to
406 Cry1F.
407 One possibility to explain the mitochondrial haplotypes detected could be that they are
408 based on reproductive incompatibility from infection by parasites such as *Wolbachia* [52].
409 However, we only detected short fragments (<100 bp) matching to the RefSeq *Wolbachia*

410 genomes on NCBI as of May 9, 2020 using BLAST (data not shown). This observation is in
411 agreement with leg and head tissues being mostly used for isolating genetic material. In testing
412 for *Wolbachia* infection, we extracted genomic DNA from a limited number of available
413 abdominal samples and performed sequencing of PCR amplicons using *Wolbachia*-specific
414 primers (Fig. S3). *Wolbachia* amplicons suggestive of infection were detected in all samples
415 tested, independently of mitochondrial haplotype, supporting no influence of *Wolbachia*
416 infection on reproductive isolation of clustered populations. Further, lack of clear
417 correspondence between genome-based and mitochondria-based haplotypes suggest on-going
418 hybridization between C- and R-strains, consistent with detection of two potential hybrids among
419 our 54 samples. Consistent lack of concordance between the nuclear and mitochondrial F_{st} values
420 suggests that biological mitochondria-nuclear interactions are not maintained, further supporting
421 that the samples analyzed are part of a panmictic population with little to no population structure.
422 Previous reports comparing genetic diversity in laboratory-reared insects to wild
423 populations mostly reported loss of allele number and heterozygosity during adaptation to mass
424 rearing [53, 54], although in some examples high genetic variability remained [55], even in long-
425 established lines [56]. We compared diversity in two long-established (>20 years) *S. frugiperda*
426 laboratory lines. Both laboratory lines, Benzon (USAMSs1) and SIMRU (USAMSs2), originated
427 from field collections in Mississippi and they were the least differentiated from the TX/MN field
428 collections. Interestingly, more recently laboratory populations from NC and FL [5] presented
429 lower F_{st} values compared to field samples. This moderate differentiation between laboratory and
430 field samples could be indicative of good culture management involving maintenance of large
431 populations and/or regular introgression of field individuals. One possibility could also be that
432 some of the field samples compared display low diversity, yet this is unlikely given samples

433 from Central and South America (for Texas population) and the Caribbean (for Florida
434 populations) based on current *S. frugiperda* migratory models [25]. However, these observations
435 of moderate differentiation in laboratory and field populations are in contrast to previous
436 research on locusts where genetic differentiation was decreased in laboratory colonies compared
437 to field samples [57]. It is important to consider that despite the lack of overall differentiation
438 between laboratory and field *S. frugiperda*, differentiation probably occurs during adaptation to
439 mass rearing, resulting in selection of specific alleles that may not be representative of field
440 populations. Thus, while our observations suggest laboratory strains contain genetic diversity
441 similar to field populations, laboratory strains may have specific allele frequencies that may not
442 be representative of field populations.

443 Overall, this study provides a diverse characterization of *S. frugiperda* whole and
444 mitochondrial genome diversity. Findings from the study support lack of clear *S. frugiperda*
445 population structure among the locations sampled, with only mitochondrial genomes indicating
446 two different maternal lines mostly separating host strains. The genomic resources generated
447 allow further exploration of gene flow and how it may affect management of *S. frugiperda* as an
448 expanding global superpest using laboratory-reared and field-collected individuals.

449

450 **Acknowledgements**

451 This project was funded by Agriculture and Food Research Initiative Foundational
452 Program competitive grant 2018-67013-27820 and Hatch Multistate NC-246 from the US
453 Department of Agriculture National Institute of Food and Agriculture. We would like to thank
454 Rodney Nagoshi for review and helpful comments.

455

References

1. Montezano DG, Specht A, Sosa-Gómez D, Roque-Specht V, Sousa-Silva J, Paula-Moraes SV, Peterson J, Hunt T: **Host plants of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) in the Americas**. *African Entomology* 2018, **26**(2):286-300.
2. Hruska AJ, Gould F: **Fall armyworm (Lepidoptera: Noctuidae) and *Diatraea lineolata* (Lepidoptera: Pyralidae): Impact of larval population level and temporal occurrence on maize yield in Nicaragua**. *Journal of Economic Entomology* 1997, **90**(2):611-622.
3. Westbrook J, Fleischer S, Jairam S, Meagher R, Nagoshi R: **Multigenerational migration of fall armyworm, a pest insect**. *Ecosphere (Washington, DC)* 2019, **10**(11):n/a.
4. Storer NP, Babcock JM, Schlenz M, Meade T, Thompson GD, Bing JW, Huckaba RM: **Discovery and characterization of field resistance to Bt maize: *Spodoptera frugiperda* (Lepidoptera: Noctuidae) in Puerto Rico**. *Journal of Economic Entomology* 2010, **103**(4):1031-1038.
5. Huang F, Qureshi JA, Meagher RL, Jr., Reisig DD, Head GP, Andow DA, Ni X, Kerns D, Buntin GD, Niu Y *et al*: **Cry1F resistance in fall armyworm *Spodoptera frugiperda*: single gene versus pyramided Bt maize**. *PLoS One* 2014, **9**(11):e112958.
6. Farias JR, Andow DA, Horikoshi RJ, Sorgatto RJ, Fresia P, dos Santos AC, Omoto C: **Field-evolved resistance to Cry1F maize by *Spodoptera frugiperda* (Lepidoptera: Noctuidae) in Brazil**. *Crop Protection* 2014, **64**:150-158.
7. Chandrasena DI, Signorini AM, Abratti G, Storer NP, Olaciregui ML, Alves AP, Pilcher CD: **Characterization of field-evolved resistance to *Bacillus thuringiensis*-derived Cry1F delta-endotoxin in *Spodoptera frugiperda* populations from Argentina**. *Pest Management Science* 2018, **74**(3):746-754.
8. Goergen G, Kumar PL, Sankung SB, Togola A, Tamò M: **First report of outbreaks of the fall armyworm *Spodoptera frugiperda* (J E Smith) (Lepidoptera, Noctuidae), a new alien invasive pest in West and Central Africa**. *PLoS ONE* 2016, **11**(10).
9. Nagoshi RN, Htain NN, Boughton D, Zhang L, Xiao Y, Nagoshi BY, Mota-Sanchez D: **Southeastern Asia fall armyworms are closely related to populations in Africa and India, consistent with common origin and recent migration**. *Scientific Reports* 2020, **10**(1).
10. Department of Argiculture and Fisheries. **First mainland detection of fall armyworm.** [https://www.daf.qld.gov.au/news-media/media-centre/biosecurity/news/first-mainland-detection-of-fall-armyworm]
11. Day R, Abrahams P, Bateman M, Beale T, Clottey V, Cock M, Colmenarez Y, Corniani N, Early R, Godwin J *et al*: **Fall armyworm: Impacts and implications for Africa**. *Outlooks on Pest Management* 2017, **28**(5):196-201.
12. Nagoshi RN, Koffi D, Agboka K, Tounou KA, Banerjee R, Jurat-Fuentes JL, Meagher RL: **Comparative molecular analyses of invasive fall armyworm in Togo reveal**

strong similarities to populations from the eastern United States and the Greater Antilles. *PLoS One* 2017, **12**(7):e0181982.

13. Pashley DP: **Host-associated genetic differentiation in fall armyworm (Lepidoptera: Noctuidae): A sibling species complex?** *Annals of the Entomological Society of America* 1986, **79**:898-904.
14. Pashley DP: **Quantitative genetics, development, and physiological adaptation in host strains of fall armyworm**. *Evolution* 1988, **42**(1):93-102.
15. Pashley DP, Martin JA: **Reproductive incompatibility between host strains of the fall armyworm (Lepidoptera: Noctuidae)**. *Annals of the Entomological Society of America* 1987, **80**:731-733.
16. Hay-Roe MM, Meagher RL, Nagoshi RN: **Effects of cyanogenic plants on fitness in two host strains of the fall armyworm (*Spodoptera frugiperda*)**. *Journal of Chemical Ecology* 2011, **37**(12):1314-1322.
17. Gouin A, Bretaudeau A, Nam K, Gimenez S, Aury J-M, Duvic B, Hilliou F, Durand N, Montagné N, Darboux I *et al*: **Two genomes of highly polyphagous lepidopteran pests (*Spodoptera frugiperda*, Noctuidae) with different host-plant ranges**. *Sci Rep* 2017, **7**(1):11816-11816.
18. Levy H, Garcia-Maruniak A, Maruniak J: **Strain identification of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) insects and cell line; PCR-RFLP of cytochrome oxidase C subunit 1 gene**. *Florida Entomologist* 2002, **85**(1):186.
19. Nagoshi RN, Meagher RL, Adamczyk JJ, Jr., Braman SK, Brandenburg RL, Nuessly G: **New restriction fragment length polymorphisms in the cytochrome oxidase I gene facilitate host strain identification of fall armyworm (Lepidoptera: Noctuidae) populations in the southeastern United States**. *Journal of Economic Entomology* 2006, **99**(3):671-677.
20. Nagoshi RN, Meagher RL: **Using intron sequence comparisons in the triose-phosphate isomerase gene to study the divergence of the fall armyworm host strains**. *Insect Mol Biol* 2016, **25**(3):324-337.
21. Nagoshi RN: **The fall armyworm triose phosphate isomerase (Tpi) gene as a marker of strain identity and interstrain mating**. *Annals of the Entomological Society of America* 2010, **103**(2):283-292.
22. Nagoshi RN: **Evidence that a major subpopulation of fall armyworm found in the Western Hemisphere is rare or absent in Africa, which may limit the range of crops at risk of infestation**. *PLoS ONE* 2019, **14**(4):e0208966.
23. Cook James M, Delgado Ana M: **Effects of a sex-ratio distorting endosymbiont on mtDNA variation in a global insect pest**. *BMC evolutionary biology* 2009, **9**(1):49.
24. Nagoshi RN, Silvie P, Meagher RL: **Comparison of haplotype frequencies differentiate fall armyworm (Lepidoptera: Noctuidae) corn-strain populations from Florida and Brazil**. *Journal of Economic Entomology* 2007, **100**(3):954-961.
25. Nagoshi RN, Fleischer S, Meagher RL, Hay-Roe M, Khan A, Murua MG, Silvie P, Vergara C, Westbrook J: **Fall armyworm migration across the Lesser Antilles and the potential for genetic exchanges between North and South American populations**. *PLoS One* 2017, **12**(2):e0171743.
26. Nagoshi RN, Nagoshi BY, Canarte E, Navarrete B, Solorzano R, Garces-Carrera S: **Genetic characterization of fall armyworm (*Spodoptera frugiperda*) in Ecuador and**

comparisons with regional populations identify likely migratory relationships. *PLoS ONE* 2019, **14**(9):e0222332.

27. Nagoshi RN, Meagher RL, Hay-Roe M: **Inferring the annual migration patterns of fall armyworm (Lepidoptera: Noctuidae) in the United States from mitochondrial haplotypes**. *Ecol Evol* 2012, **2**(7):1458-1467.
28. Nagoshi RN, Rosas-Garcia NM, Meagher RL, Fleischer SJ, Westbrook JK, Sappington TW, Hay-Roe M, Thomas JM, Murua GM: **Haplotype profile comparisons between *Spodoptera frugiperda* (Lepidoptera: Noctuidae) populations from Mexico with those from Puerto Rico, South America, and the United States and their implications to migratory behavior**. *J Econ Entomol* 2015, **108**(1):135-144.
29. Meagher RL: **Trapping fall armyworm (Lepidoptera: Noctuidae) adults in traps baited with pheromone and a synthetic floral volatile compound**. *Florida Entomologist* 2001, **84**(2):288-292.
30. Banerjee R, Hasler J, Meagher R, Nagoshi R, Hietala L, Huang F, Narva K, Jurat-Fuentes JL: **Mechanism and DNA-based detection of field-evolved resistance to transgenic Bt corn in fall armyworm (*Spodoptera frugiperda*)**. *Scientific Reports* 2017, **7**(1):10877.
31. Oscar FS-A, João VCR, Thadeu CS, Clébson ST, Silverio OC, Raul NCG, Eliseu JGP: **Resistance to dual-gene Bt maize in *Spodoptera frugiperda*: selection, inheritance, and cross-resistance to other transgenic events**. *Scientific Reports* 2015, **5**(1).
32. Santos-Amaya OF, Tavares CS, Rodrigues JVC, Campos SO, Guedes RNC, Alves AP, Pereira EJG: **Fitness costs and stability of Cry1Fa resistance in Brazilian populations of *Spodoptera frugiperda***. *Pest Management Science* 2017, **73**(1):35-43.
33. **Fall armyworm. University of Florida Publication Number EENY-98** [http://entnemdept.ufl.edu/creatures/field/fall_armyworm.htm#host]
34. Nagoshi RN, Silvie P, Meagher RL, Lopez J, Machado V: **Identification and comparison of fall armyworm (Lepidoptera: Noctuidae) host strains in Brazil, Texas, and Florida**. *Annals of the Entomological Society of America* 2007, **100**(3):394-402.
35. **FastQC: A quality control tool for high throughput sequence data** [https://qubeshub.org/resources/fastqc]
36. Nam K, Nhim S, Robin S, Bretaudeau A, Nègre N, d'Alençon E: **Divergent selection causes whole genome differentiation without physical linkage among the targets in *Spodoptera frugiperda* (Noctuidae)**. *bioRxiv* 2018:452870.
37. Li H: **Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM**. 2013.
38. Li H: **A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data**. *Bioinformatics* 2011, **27**(21):2987-2993.
39. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST *et al*: **The variant call format and VCFtools**. *Bioinformatics* 2011, **27**(15):2156-2158.
40. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, Phillippy AM: **Mash: fast genome and metagenome distance estimation using MinHash.(Report)**. *Genome Biology (Online Edition)* 2016, **17**(1).
41. Dierckxsens N, Mardulyn P, Smits G: **NOVOPlasty: de novo assembly of organelle genomes from whole genome data**. *Nucleic acids research* 2017, **45**(4):e18.

42. Jombart T, Ahmed I: **Adegenet 1.3-1: new tools for the analysis of genome-wide SNP data**. *Bioinformatics (Oxford, England)* 2011, **27**(21):3070.
43. Paradis E: **pegas: an R package for population genetics with an integrated–modular approach**. *Bioinformatics* 2010, **26**(3):419-420.
44. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ: **Second-generation PLINK: rising to the challenge of larger and richer datasets**. *GigaScience* 2015, **4**(1).
45. Pritchard JK, Stephens M, Raj A: **fastSTRUCTURE: Variational inference of population structure in large SNP data sets.** *Genetics* 2014, **197**(2):573-589.
46. **PLINK 2.0.** [www.cog-genomics.org/plink/2.0/]
47. Wickham H: **ggplot2 Elegant Graphics for Data Analysis**. 2016.
48. Rašić G, Schama R, Powell R, Maciel-de Freitas R, Endersby-Harshman NM, Filipović I, Sylvestre G, Máspero RC, Hoffmann AA: **Contrasting genetic structure between mitochondrial and nuclear markers in the dengue fever mosquito from Rio de Janeiro: implications for vector control**. *Evolutionary Applications* 2015, **8**(9):901-915.
49. Finley LWS, Haigis MC: **The coordination of nuclear and mitochondrial communication during aging and calorie restriction**. *Ageing Research Reviews* 2009, **8**(3):173-188.
50. Stephan W: **Principles of Population Genetics. Fourth Edition.** *The Quarterly review of biology* 2007, **82**(4):416-416.
51. Nagoshi RN, Meagher RL, Flanders K, Gore J, Jackson R, Lopez J, Armstrong JS, Buntin GD, Sansone C, Leonard BR: **Using haplotypes to monitor the migration of fall armyworm (Lepidoptera: Noctuidae) corn-strain populations from Texas and Florida**. *Journal of Economic Entomology* 2008, **101**(3):742-749.
52. Correa CC, Ballard JWO: **Wolbachia associations with insects: winning or losing against a master manipulator**. *Frontiers in Ecology and Evolution* 2016, **3**.
53. Zygouridis NE, Argov Y, Nemny-Lavy E, Augustinos AA, Nestel D, Mathiopoulos KD: **Genetic changes during laboratory domestication of an olive fly SIT strain**. *Journal of applied entomology (1986)* 2014, **138**(6):423-432.
54. Gilchrist AS, Cameron EC, Sved JA, Meats AW: **Genetic consequences of domestication and mass rearing of pest fruit fly *Bactrocera tryoni* (Diptera: Tephritidae)**. *Journal of Economic Entomology* 2012, **105**(3):1051-1056.
55. Simões P, Pascual M, Coelho MM, Matos M: **Divergent evolution of molecular markers during laboratory adaptation in *Drosophila subobscura***. *Genetica* 2010, **138**(9-10):999-1009.
56. Parreño MA, Scannapieco AC, Remis MI, Juri M, Vera MT, Segura DF, Cladera JL, Lanzavecchia SB: **Dynamics of genetic variability in *Anastrepha fraterculus* (Diptera: Tephritidae) during adaptation to laboratory rearing conditions**. *BMC genetics* 2014, **15 Suppl 2**(Suppl 2):S14.
57. Berthier K, Chapuis MP, Simpson SJ, Ferenz HJ, Habib Kane CM, Kang L, Lange A, Ott SR, Babah Ebbe MA, Rodenburg KW *et al*: **Laboratory populations as a resource for understanding the relationship between genotypes and phenotypes: A global case study in locusts**. *Advances in insect physiology* 2010, **39**:1-37.

634

635 **Figure legends**

636 **Figure 1:** Principal component analysis (PCA) plot of 52 *Spodoptera frugiperda* samples (3
637 samples removed due to missing rate ≥ 50%) based on 3,558,854 SNPs. Country of origin and
638 host strain as determined by the nuclear *Tpi* marker are visualized by color and shape,
639 respectively, as shown in the legend.

640

641 **Figure 2**: Dissimilarity heatmap of Mash distances across all 55 *Spodoptera frugiperda* samples,
642 colored by country of origin as indicated. Two outgroups (*Spodoptera litura* and *Spodoptera*
643 *exigua*) were added. Samples were named as described in Materials and Methods.

644

645 **Figure 3:** Nucleotide diversity box plots between C-strain Florida (FL) vs. R-strain Texas (TX)
646 samples.

647

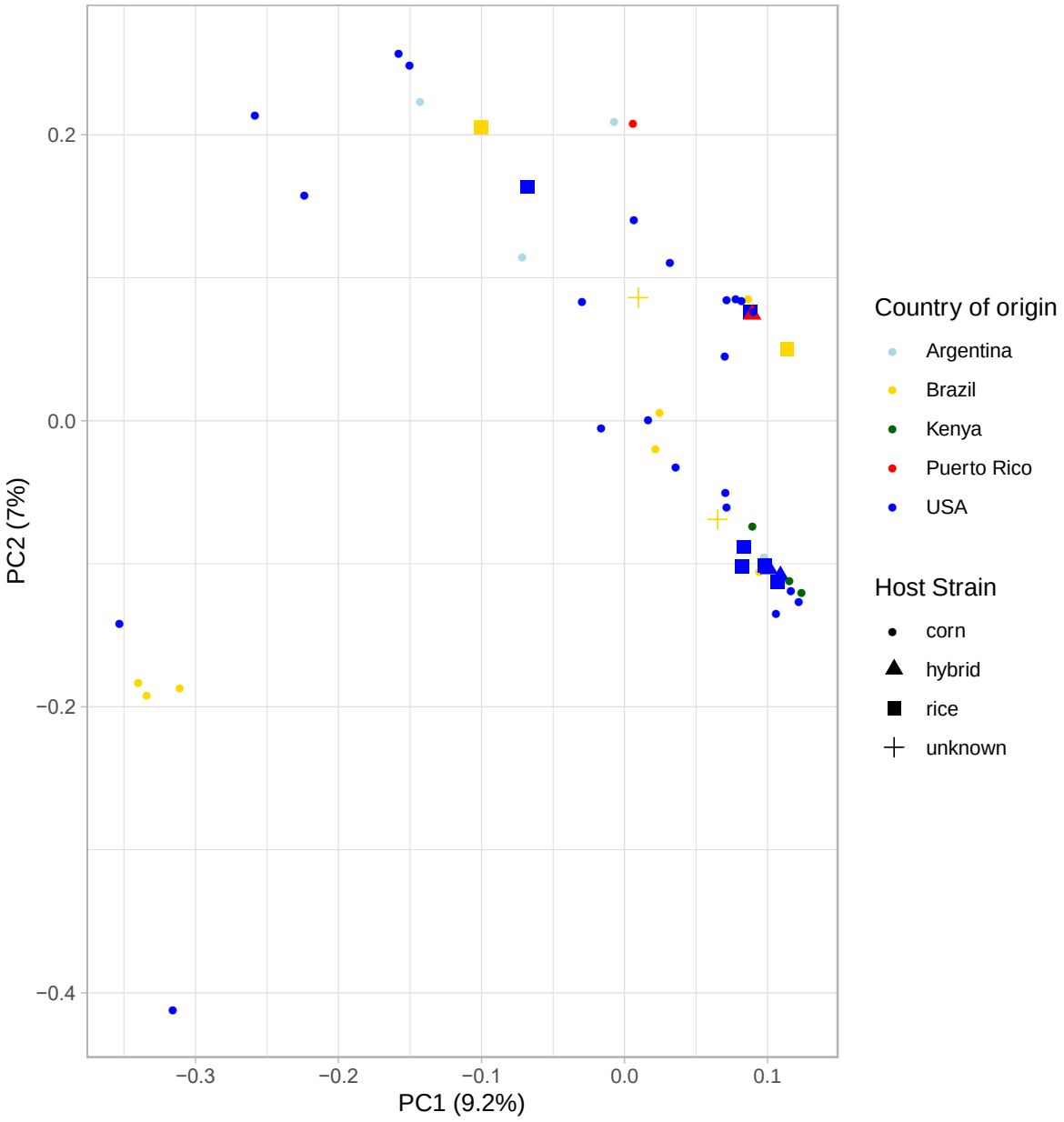
648 **Figure 4**: Mitochondrial haplotype network of all 54 *S. frugiperda* mitochondrial chromosome
649 level assembled individuals, colored by country of origin.

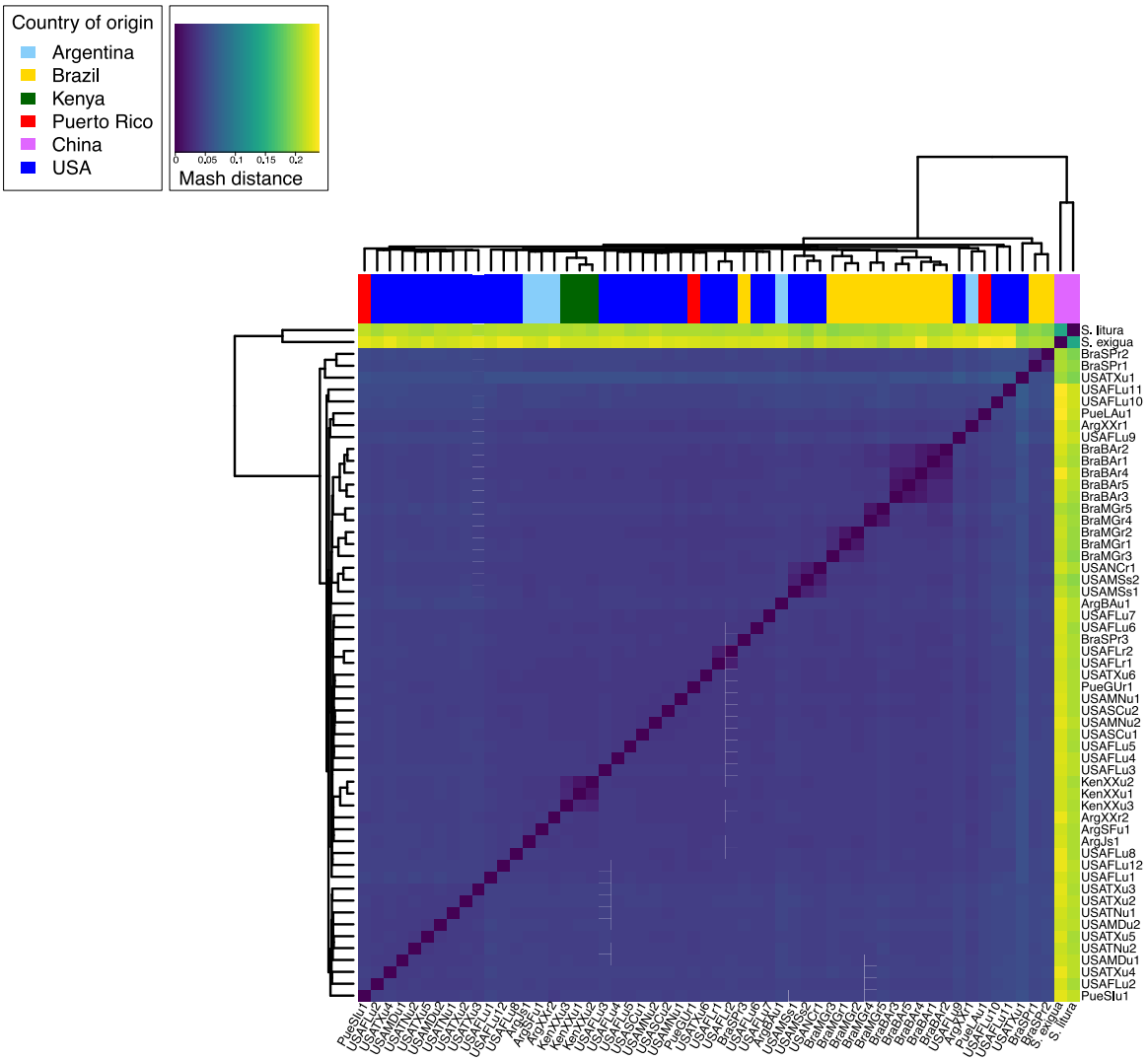
650

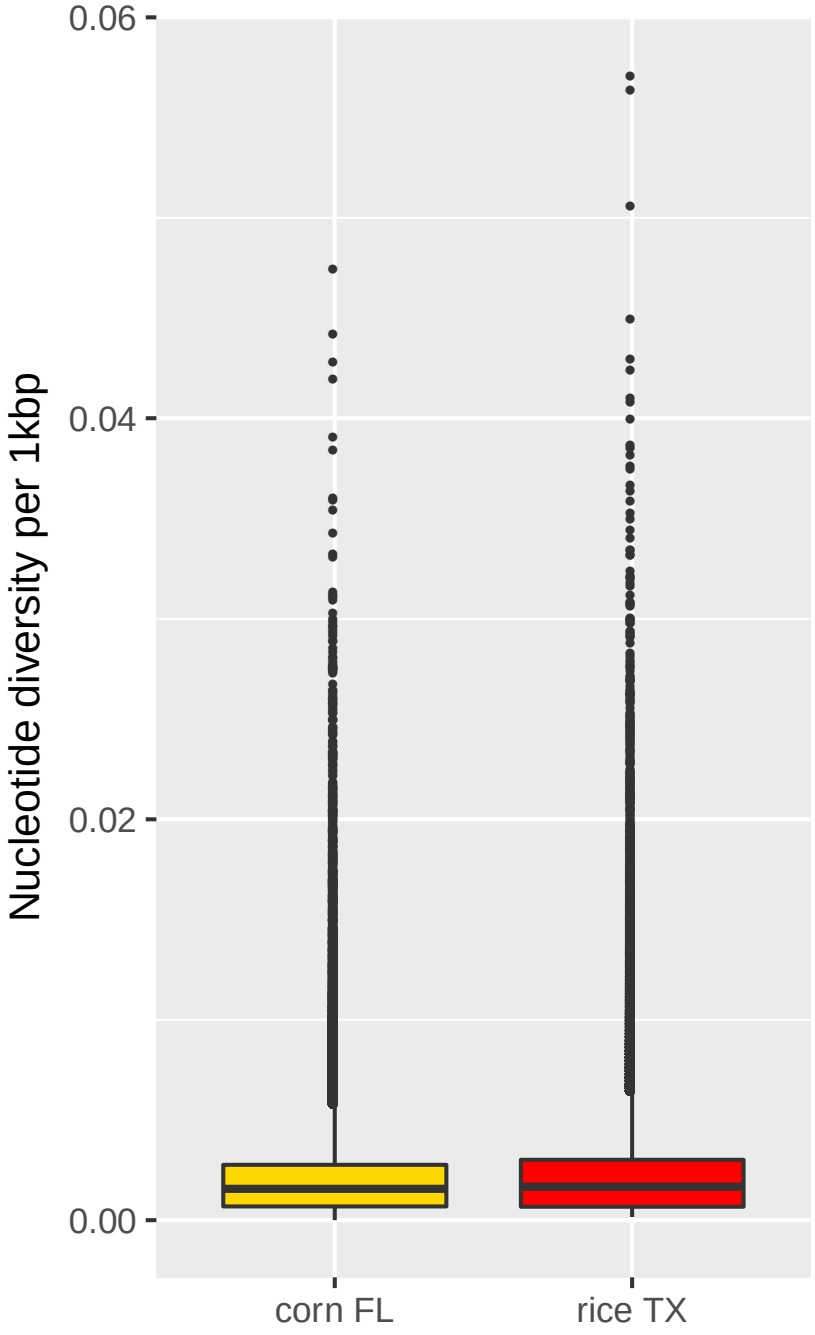
651 **Figure 5:** UPGMA distance-based tree on biallelic variant SNPs in 54 mitochondrial
652 chromosome level assembled individuals rooted with *Spodoptera litura* as an outgroup.

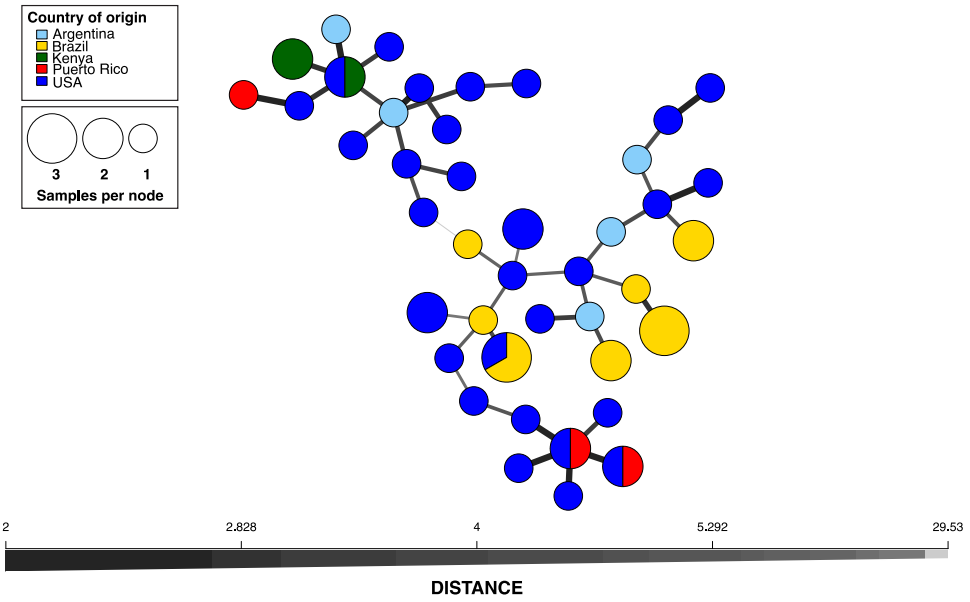
653

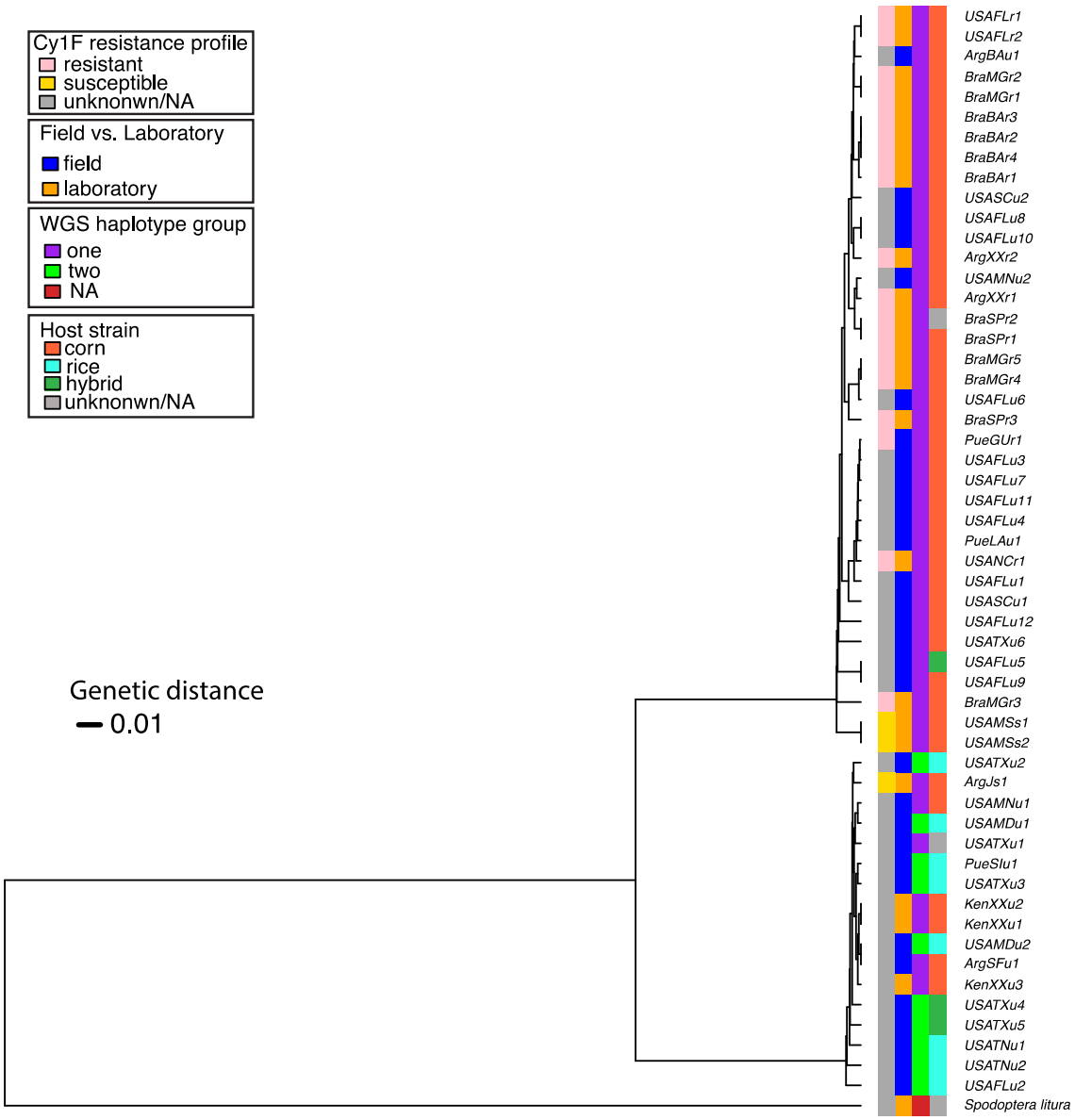
654 **Table 1:** Pairwise F_{st} values across all populations surveyed. NA was used when the samples
655 overlap in the defined populations.


0.2
0.0
−0.2
−0.4
PC2 (7%)
−0.3
−0.2
−0.1
0.0
0.1
PC1 (9.2%)
Country of origin
Argentina
Brazil
Kenya
Puerto Rico
USA
Host Strain
corn
hybrid
rice
unknown


Country of origin
Argentina
Brazil
Kenya
Puerto Rico
China
USA
0
0.05
0.1
0.15
0.2
Mash distance
S. litura
S. exigua
BraSPr2
BraSPr1
USATXu1
USAFLu11
USAFLu10
PueLAu1
ArgXXr1
USAFLu9
BraBAr2
BraBAr1
BraBAr4
BraBAr5
BraBAr3
BraMGr5
BraMGr4
BraMGr2
BraMGr1
BraMGr3
USANCr1
USAMSs2
USAMSs1
ArgBAu1
USAFLu7
USAFLu6
BraSPr3
USAFLr2
USAFLr1
USATXu6
PueGUr1
USAMNu1
USASCu2
USAMNu2
USASCu1
USAFLu5
USAFLu4
USAFLu3
KenXXu2
KenXXu1
KenXXu3
ArgXXr2
ArgSFu1
ArgJs1
USAFLu8
USAFLu12
USAFLu1
USATXu3
USATXu2
USATNu1
USAMDu2
USATXu5
USATNu2
USAMDu1
USATXu4
USAFLu2
PueSIu1


0.06
0.04
0.02
0.00
Nucleotide diversity per 1kbp
corn FL
rice TX


Country of origin
Argentina
Brazil
Kenya
Puerto Rico
USA
3
2
1
Samples per node
2
2.828
4
5.292
29.53
DISTANCE


Cy1F resistance profile
resistant
susceptible
unknonwn/NA
Field vs. Laboratory
field
laboratory
WGS haplotype group
one
two
NA
Host strain
corn
rice
hybrid
unknonwn/NA
Genetic distance
0.01
USAFLr1
USAFLr2
ArgBAu1
BraMGr2
BraMGr1
BraBAr3
BraBAr2
BraBAr4
BraBAr1
USASCu2
USAFLu8
USAFLu10
ArgXXr2
USAMNu2
ArgXXr1
BraSPr2
BraSPr1
BraMGr5
BraMGr4
USAFLu6
BraSPr3
PueGUr1
USAFLu3
USAFLu7
USAFLu11
USAFLu4
PueLAu1
USANCr1
USAFLu1
USASCu1
USAFLu12
USATXu6
USAFLu5
USAFLu9
BraMGr3
USAMSs1
USAMSs2
USATXu2
ArgJs1
USAMNu1
USAMDu1
USATXu1
PueSIu1
USATXu3
KenXXu2
KenXXu1
USAMDu2
ArgSFu1
KenXXu3
USATXu4
USATXu5
USATNu1
USATNu2
USAFLu2
Spodoptera litura