Genomic balancing selection is key to the invasive success of the fall armyworm

Sudeeptha Yainna[1], Wee Tek Tay[2], Estelle Fiteni[1,3], Fabrice Legeai[4,5], Anne-Laure Clamens[6], Sylvie Gimenez[1], Marie Frayssinet[1], R Asokan[7], CM Kalleshwaraswamy[8], Sharanabasappa Deshmukh[8], Robert L. Meagher, Jr.[9], Carlos A. Blanco[10], Pierre Silvie[11,12,13], Thierry Brévault[11,12], Anicet Dassou[14], Gael J. Kergoat[6], Thomas Walsh[2], Karl Gordon[2] , Nicolas Nègre[1], Emmanuelle d'Alençon[1], Kiwoong Nam[1*]

[1]DGIMI, Univ Montpellier, INRAE, Montpellier, France

[2]CSIRO, Black Mountain Laboratories, Canberra, Australia

[3]Université Claude Bernard Lyon 1, Lyon, France

[4]INRA, UMR-IGEPP, BioInformatics Platform for Agroecosystems Arthropods, Campus Beaulieu, Rennes, 35042, France

[5]INRIA, IRISA, GenOuest Core Facility, Campus de Beaulieu, Rennes, France

[6]CBGP, INRAE, CIRAD, IRD, Institut Agro, Univ Montpellier, Montpellier, France

[7]Division of Biotechnology, ICAR - Indian Institute of Horticultural Research, Bengaluru, India

[8]Department of Entomology, College of Agriculture, University of Agricultural and Horticultural Sciences, Shivamogga, India

[9]United States Department of Agriculture, Agricultural Research Service, Gainesville, Florida, U.S.A

[10]United States Department of Agriculture, Animal and Plant Health Inspection Service, Maryland, U.S.A.

[11]CIRAD, UPR AIDA, Montpellier, France

[12]AIDA, Univ. Montpellier, CIRAD, Montpellier, France

[13]IRD, UMR IPME, 34000, Montpellier, France

[14]ENSBBA, UNSTIM, Dassa, Benin

*Correspondence: ki-woong.nam@inrae.fr

Abstract

A successful biological invasion involves survival in a newly occupied environment. If a population bottleneck occurs during an invasion, the resulting depletion of genetic variants could cause increased inbreeding depression and decreased adaptive potential, which may result in a fitness reduction. How invasive populations survive in the newly occupied environment despite reduced heterozygosity and how, in many cases, they maintain moderate levels of heterozygosity are still contentious issues[1]. The Fall armyworm (FAW; Lepidoptera: *Spodoptera frugiperda*), a polyphagous pest, is native to the Western hemisphere. Its invasion in the Old World was first reported from West Africa in early 2016, and in less than four years, it swept sub-Saharan Africa and Asia, finally reaching Australia. We used population genomics approaches to investigate the factors that may explain the invasive success of the FAW. Here we show that genomic balancing selection played a key role in invasive success by restoring heterozygosity before the global invasion. We observe a drastic loss of mitochondrial polymorphism in invasive populations, whereas nuclear heterozygosity exhibits a mild reduction. The population from Benin in West Africa has the lowest length of linkage disequilibrium amongst all invasive and native populations despite its reduced population size. This result indicates that balancing selection increased heterozygosity by facilitating the admixture of invasive populations from distinct origins and that, once heterozygosity was sufficiently high, FAW started spreading globally in the Old World. As comparable heterozygosity levels between invasive and native populations are commonly observed[1], we postulate that the restoration of heterozygosity through balancing selection could be widespread among successful cases of biological invasions.

Keywords: adaptive evolution, Fall armyworm, invasive pests, population genomics, *Spodoptera frugiperda*

Text

A successful biological invasion involves the survival of an introduced population, which is typically associated with rapid adaptation processes in the newly occupied environment[2,3]. If a bottleneck occurs during an invasion as a result of the introduction of a small number of individuals, the invasive population may have a decreased fitness due to inbreeding depression because the level of heterozygosity is decreased[4]. Moreover, small populations may have a lower adaptive potential than large populations because of a lower population-scaled rate of mutation[5–7] or a lower number of existing genetic variants[8], of which a proportion provides a beneficial effect for the survival in a new environment. The expectation that invasive populations have a reduced fitness appears to be contradictory with ample cases of invasive success, which has been often coined up as the 'genetic paradox of biological invasion'[9].

The occurrence of multiple introduction events has been proposed to be the solution to this paradox because genetic admixture among heterogeneous populations results in an increase in heterozygosity, which may decrease inbreeding depression and increase adaptive potential (reviewed in Estoup et al.[1]). However, the co-existence of allopatrically-originated individuals does not necessarily cause an increase in the level of heterozygosity because of the following two reasons. First, admixed individuals may have reduced fitness due to genetic incompatibilities between two haplotypes or strains. An established population is expected to have an optimal allelic combination through natural selection. Thus, admixed individuals between two established populations may have a substantial number of incompatible alleles, which decreases fitness. Indeed, genetic incompatibilities between populations are common in *Drosophila* fruit flies[10]. In addition, during the entire process of admixture, the stochastic effect of genetic drift may cause a substantial loss of variants if the initial number of invading individuals is small. In other words, a large effective population size is required to maintain variants from heterogeneous populations by overcoming genetic drift.

If selective advantages of admixed genotypes are sufficiently high to overcome the potential genetic incompatibilities in admixed individuals or to overcome genetic drift at the initial phase of admixture, then balancing selection may act in the way of facilitating admixture between different sets of genotypes from the different invasive origins. Therefore, it is tempting to hypothesize that invasive populations experience balancing selection, which restores heterozygosity during the lagging time between initial introductions to rapid range expansion.

The fall armyworm, *Spodoptera frugiperda* (J.E. Smith) (Lepidoptera: Noctuidae: Noctuinae), is one of the most infamous insect pests due to an extremely high-level of polyphagy (more than 353 host-plants belonging to 76 plant families are reported[11]), high dispersal capacity and migratory behavior[12], the rapid development of insecticide resistance[13,14], including resistance to Bt proteins[15–18], and occasional outbreaks[19]. The FAW is native to North and South America, and its presence in West Africa was first reported in 2016[20]. In the following years, the FAW spread across sub-Saharan Africa, followed by global detection in India, South East Asia, East Asia, Egypt, and Australia (https://www.cabi.org/isc/fallarmyworm). Invasive FAW larvae cause significant economic losses, especially on corn, with yield loss of corn production averaging 21%-53% in Africa[21]. The FAW consists of two strains, corn strain (sfC) and rice strain (sfR) (named after their supposedly preferred host-plants), which are observed sympatrically in all their native range[22–24]. Both strains are observed in invasive populations, while the relative proportion of the identified strains depends on their geographic location[25–27] . Tay et al., reported genomic signature of multiple introductions of FAW from Mississippi and South America to the Old World based on 870 unlinked single nucleotide variants (SNV)[28]. Potential multiple introductions and the recent explosive global invasion of the FAW makes this species an ideal model to test the potential effect of balancing selection in invasion success.

In this paper, we aim at testing the potential role of balancing selection in the global invasion of the FAW using population genomics. First, we identified genomic SNV (Single Nucleotide Variants)

from 177 samples in both native and invasive populations. Then, we inferred multiple origins of invasion, and tested balancing selection in the invasive population. Lastly, we identified adaptive evolution specific to invasive populations. We generated a new reference genome assembly from sfC using 30X PacBio Reads and Hi-C data[29]. The assembly size and N50 are 385 Mbp and 10.6 Mbp, respectively. L90 is 26, which is close to the chromosome number in FAW (31), implying nearly chromosome-sized scaffolds in this assembly. BUSCO analysis[30] demonstrates that this assembly has the highest correctness among all published FAW genome assemblies (Table S1).

The origin of invasion

We performed whole genome sequencing from FAW samples collected in Benin (39 individuals), India (14), Mexico (26), Florida (24), French Guiana (3), and Guadeloupe (4) using novaseq 6000 with 20X coverage on average (Fig. S1). This dataset was combined with resequencing data from populations collected in Mississippi (17) and Puerto Rico (15), which were used in our previous studies[31,32]. In addition, we had the opportunity of using resequencing data of Brazil (10), Malawi (16), and Uganda (7) from CSIRO[28], Australia. Lastly, two individuals from China were also added to the dataset[33]. The resulting total number of individuals used in this study is 177 (99 from native populations and 78 from invasive populations). The mapping of genomes was performed against the reference assembly (Fig. S2), followed by variant calling using GATK[34]. After filtering, 27,117,672 SNPs remained (see methods for more detail). We identified the strain from a maximum likelihood phylogenetic tree reconstructed from the full sequences of mitochondrial Cytochrome C Oxidase subunit I (COX1) gene, which is the universal barcode gene and also commonly used for FAW strain identification[35]. The COX1 phylogenetic tree shows high bootstrapping confidence scores for both sfC and sfR (bootstrap supporting value > 92%) (Fig. S3), with 99 and 78 and individuals being assigned to sfC and sfR, respectively. The invasive populations have 29 and 49 sfC and sfR individuals, respectively, and native populations have 70 and 29 sfC and sfR individuals, respectively.

A principal component analysis was performed from nuclear genome sequences to identify the origin of invasive populations. The first principal component shows three groups of individuals (Fig. 1A). The first group (sfR group) consists of sfR from the Caribbean, including Florida, Guadeloupe, and French Guiana, but also one individual from Mississippi. The second group (sfC group) consists of sfC from Mexico only. The third group (hybrid group) is found between the first and second groups along the first principal component, suggesting that this group was probably generated through intraspecific hybridization between sfC and sfR. The second principal component separates the hybrid group into native (Mississippi, Puerto Rico, Brazil, and Florida) and invasive (Benin, Malawi, Uganda, India, and China) populations. This result shows that hybrids were first generated in native populations and that these hybrids further invaded the Old World. This result is in line with previous studies, indicating that the vast majority of individuals of invasive populations are hybrids[25–27]. We also observed that both native hybrid populations and invasive populations exhibit reproductive barriers between sfC and sfR from genetic differentiation (F_{ST}) with the statistical significance (Fig. S4).

The ancestry coefficient analysis[36] shows that invasive populations have homogeneous genomic sequences in a range of K values, while native populations show the heterogeneity except for sfC in Mexico (Fig. 1B). The BIO-NJ phylogenetic tree reconstructed from whole genome sequences exhibits 100% bootstrapping supports for sfC and sfR groups (Fig. 1C), in like with the PCA results (Fig. 1A). In addition, the tree also demonstrates that all invasive individuals belong to a single clade with bootstrap support of 100%, further highlighting the homogeneity of the invasive genomic background in all the locations of collections.

Reduction in genetic diversity during the invasion

We then compared the genetic diversity between native hybrid populations and invasive populations. We assembled whole mitochondrial genomes, and we observed that we were able to extract high-quality full-length ND5 and COX1 sequences from all 177 individuals. In ND5, the

longest gene in the mitochondrial genome, sfC and sfR of the native hybrid populations have 26 and six polymorphic sites, respectively (Fig. 1D, left). However, sfC of the invasive populations has only one polymorphic site from 29 individuals (96.6% reduction), and sfR of the invasive populations has no polymorphic site from 43 individuals (100% reduction). We also compared π (nucleotide diversity) between invasive populations and native hybrid populations from whole mitochondrial genomes. The nucleotide diversity of sfC and sfR was reduced during the invasion by 78.32% (6.100×10^{-4} and 1.323×10^{-4} for native hybrid populations and invasive populations, respectively) and by 78.45% (3.156×10^{-4} and 6.801×10^{-5} for native hybrid populations and invasive populations, respectively), respectively. We identified eight and nine mitochondrial SNVs from sfC and sfR, respectively, but none of them was identified from native hybrid groups. This result implies that the observed SNVs in invasive populations were generated after the invasion, although we cannot exclude the possibility that these SNVs were derived from native hybrid populations that are not included in this study. The dramatic reduction in the mitochondrial genetic diversity, which was already shown in a previous study[37], implies a severe genetic bottleneck during the invasion.

We further compared the number of nuclear biallelic heterozygous sites counted from each individual between native hybrid populations and invasive populations. We considered sites only if the genotype is determined from all 177 individuals to avoid potential statistical artifacts from missing data. Invasive populations have significantly lower numbers of heterozygous positions (Wilcoxon rank-sum test, $p = 1.2 \times 10^{-14}$), while the average difference is only 12.71% (15,854.18 and 18,162.53 for invasive populations and native hybrid populations, respectively, among 412,404bp) (Fig. 1D, right). Interestingly, two individuals from India show almost the complete depletion of heterozygosity (B4 and B9), and one individual from Puerto Rico (PR19) has particularly high heterozygosity. The dramatic difference in the reduction of genetic diversity between mitochondrial and nuclear genomes suggests that the evolutionary forces reshaping polymorphism patterns is different between these two genomes.

Multiple introductions have been suggested to contribute to an increase in the heterozygosity of invasive populations. Thus, multiple origins of FAW might explain the moderate level of heterozygosity in invasive populations. However, this explanation alone cannot explain the difference between nuclear and mitochondrial patterns shown in Fig. 1D, because it is not realistic that the admixture increased only nuclear genetic diversity (which is heterozygosity in the case of diploid nuclear genomes) while mitochondrial genetic diversity remained unchanged.

Instead, we postulate that genomic balancing selection increased nuclear heterozygosity in invasive FAW populations. In this scenario, (i) a severe bottleneck of an initially invasive population depleted heterozygosity, which caused inbreeding depression[4] (for example, reduced egg viability, increased mortality, and reduced life span as shown in inbred monarch butterflies[38]), (ii) this population had a lagging period where the nuclear heterozygosity gradually increased through genomic balancing selection, which facilitated admixture among populations with different invasive origins, while mitochondrial genetic diversity remained low, and (iii) when the heterozygosity has sufficiently increased to generate a stable population of the initially invasive population, the FAW was able to start its large scale invasion of the Old World.

Genomic balancing selection

To test the possibility of genomic balancing selection, we analyzed copy number variations (CNVs) to identify the origin of the invasive population with a higher resolution. As CNVs are much rarer than SNVs, we expected that CNVs have fewer noise signals from shared ancestral polymorphisms among multiple native populations to detect the invasive origin. We used CNVs only if minor allele frequency is higher than 0.2 to minimize false positives. The number of identified CNVs is 22,915. Ancestry coefficient analysis shows that, from a range of K values, invasive populations are divided into two groups (Fig. 2A). The first group includes Benin and India, and the second group includes Uganda, Malawi, and China. The first and the second groups have the same ancestry pattern from

sfC in Florida (Florida-sfC) and Brazil, respectively. This result demonstrates the occurrence of multiple introductions from Florida-sfC and Brazil.

The heterogeneous distribution of Florida-sfC-specific or Brazil-specific SNV among invasive individuals was tested. We counted the numbers of SNVs that are found only from Florida-sfC or Brazil for each individual in invasive populations, and these numbers were compared between the two invasive groups (Benin-India and Malawi-Uganda-China). Fig. 2B shows a nearly uniform distribution of SNV numbers specific to Florida-sfC-specific SNV across the entire invasive populations, and the SNV numbers were not significantly different between these two groups (p = 0.3502; 22,746.69 and 22,493.56 for Benin-India and Malawi-Uganda-China). The Malawi-Uganda-China group has a significantly higher number of Brazil-specific SNV than the Benin-India group ($p = 6.519 \times 10^{-7}$; 11,934.20 and 12,484.84 for Benin-India and Malawi-Uganda-China, respectively), but with only a 4.61% difference between the two. These results show an almost uniform distribution of the numbers of Florida-sfC-specific or Brazil-specific SNVs among individuals in invasive populations, unlike what is found with CNVs.

Subsequently, we estimated to what extent the heterozygosity can be increased by admixture from SNVs that are absent in Brazil for each individual from Florida-sfC, assuming that these SNVs may increase the genetic diversity compared with a case that only Brazil is the only invading population. The numbers of these SNVs range from 656,760bp to 695,100bp (Fig. 2C). We also identified SNVs that are absent in Florida-sfC for each individual from Brazil. The numbers of these SNVs range from 378,299bp to 520,133bp. This result shows that the admixture between Florida-sfC and Brazil populations may increase the number of SNPs from 378kb to 695kb. The number of heterozygous positions in the invasive population is 1,629,133bp on average. Thus, the mixture might contribute to the heterozygous positions up to 42.67% of total invasive SNPs (695,100bp / 1,629,133bp).

Then, we tested whether genomic balancing selection increases the level of heterozygosity by mixing genes between sfC-Florida and Brazil. In the presence of balancing selection, the length of the linkage disequilibrium is decreased because balancing selection has the same effect on the linkage disequilibrium with recombination hotspot[39]. Therefore, if invasive populations experienced genomic balancing selection, then these populations are expected to have shorter linkage disequilibrium than native populations. If balancing selection does not exist, invasive populations will have longer lengths of linkage disequilibrium than native populations because of smaller effective population sizes (i.e., smaller heterozygosity as shown in Fig. 1D). To test these alternative hypotheses, we compared the decay curve of linkage disequilibrium according to the distance from one locus to another for each strain of each population. We observed that the sfC and sfR from Benin had a faster decay of linkage disequilibrium than the other invasive populations as well as sfC-Florida or Brazil populations (Fig. 3A). When the decay of linkage disequilibrium was compared across all the native and invasive populations, sfC and sfR from Benin exhibit the fastest rate of decay (Fig S5). This result shows that the invasive population in Benin has a shorter linkage disequilibrium than native populations despite the smaller effective population size. This pattern is best explained by balancing selection that increases the genomic heterozygosity level of the population in Benin. Figure 3B shows a correlation of nucleotide diversity calculated from 100kb windows between invasive and native hybrid populations. The Pearson's correlation coefficient is very high ($r = 0.992$, $p < 2.2 \times 10^{-16}$), and outliers of this correlation are not observed. This pattern is in line with genomic balancing selection, rather than balancing selection affecting only a few loci. The shorter length of linkage disequilibrium in Benin is of particular interest because the FAW invasion was first reported on the Western coast of Africa, including Benin, Togo, Nigeria, and São Tomé and Príncipe[20]. Thus, we concluded that FAWs had increased heterozygosity by balancing selection in Benin (or other neighboring regions) and were able to spread eastward once their heterozygosity was sufficiently high.

Testing alternate hypotheses

An alternative explanation is that Florida-sfC-originated individuals co-existed with Brazil-originated individuals in Benin, while the admixture was incomplete compared with the other invasive populations. In this case, the heterogeneous genomic sequences among individuals in Benin may cause an underestimation of the length of linkage disequilibrium. We tested the heterogeneity in the population from Benin from CV (coefficient of variance) of Florida-sfC or Brazil derived variants (Fig. 2B) among invasive populations assuming that this heterogeneity among individuals increases the variance of Florida-sfC-specific or Brazil-specific SNV numbers. For the variants from Florida-sfC, CV was lowest in Benin (0.0194), followed by Malawi (0.0231), India (0.0241), and then Uganda (0.0535). For the variants from Brazil, CV was lowest in India (0.0184), followed by Benin (0.0202), Malawi (0.0332), and then Uganda (0.0527). This result shows that the population from Benin does not have a particularly high CV. Therefore, the heterogeneity of genomic sequences in Benin is not supported.

We then tested another alternative hypothesis that the level of heterozygosity in invasive populations is increased by interspecific hybridization with non-FAW species belonging to the same genus as there are several other *Spodoptera* species which are found in Africa and Asia including *S. littoralis* (Boisduval) in Africa, *S. mauritia* (Boisduval) and *S. litura* (Fabricius) in Asia, and *S. cilium* (Guenée) and *S. exigua* (Hübner) in Africa and Asia. In this case, the distribution of genetic differentiation is expected to show a bimodal distribution[40], in which each mode represents the FAW and non-FAW species, respectively. The histogram of F_{ST} calculated from 100kb windows shows a unimodal distribution, in which 99.0% of windows have F_{ST} greater than zero (Fig. 4A). This distribution does not support inter-specific hybridization. We also tested the interspecific hybridization from the numbers of homozygous variant positions, which are expected to be increased by interspecific hybridization because, in this case, the non-FAW species have a longer phylogenetic distance from organism used to generate the reference genomes than the FAW in the native populations. In order to remove statistical artifacts, we considered positions only if genotypes are determined from all individuals. We observed that invasive populations have lower numbers of

278 homozygous variant positions than native populations (2954.295bp and 3170.527bp in total
279 412,404bp for invasive and native populations, respectively; $p = 0.005319$ Wilcoxon rank-sum test)
280 (Fig. 4B), further showing that the interspecific hybridization between *Spodoptera* species is not
281 supported.

282 **Identification of adaptive evolution in the invasive population**

283 We calculated the composite likelihood of selective sweeps[41] from invasive populations to identify
284 positively selected genes that may contribute to adaptation in a new environment. The median value
285 of the composite likelihood is 0.4350, and a locus is considered to be targeted by selective sweep if
286 the composite likelihood is higher than 100, which was arbitrarily chosen. In total, we identified
287 seven loci on three chromosomes as potential targets of selective sweeps (Fig . 5A). As the high
288 composite likelihood of these loci might be generated by selective sweeps not specific to invasive
289 populations or by background selection[42], we calculated the composite likelihood from native
290 hybrid populations as well. Four out of the seven loci do not exhibit outliers of the composite
291 likelihood in native hybrid populations (Fig. S6). Therefore, we considered these four loci
292 potentially targeted by selective sweeps specific to invasive populations. These four loci contain 36
293 predicted protein-coding genes (Table S3), including 12 genes with unknown gene functions. We
294 carefully underwent a manual curation of these genes to determine the function. The locus on
295 chromosome 14 has CYP9A, which belongs to Cytochrome P450 gene family. This gene family
296 plays a key role in detoxifying xenobiotics[43], and CYP9A genes are overexpressed by plant
297 allelochemicals and pesticides in FAW[44]. Therefore, positive selection on this gene might contribute
298 to the adaptation to plants or pesticides in an invasive area. This locus also includes three copies of
299 tubulin genes, implying that the cytoskeleton could be under positive selection as well.

300 A locus on chromosome 29 includes a carboxylesterase gene, which may involve insecticides
301 resistance[45], and an ABC transporter homolog to mdr49, which protects organisms from cytotoxic
302 compounds in *Drosophila melanogaster* Meigen[46]. Therefore, positive selection of these three genes

might mitigate environmental stresses in an invasive area. This locus also includes a kunitz-type serine protease inhibitor gene, which plays a role in the digestion of plants[47]. The gene encoding odorant receptor 13, which could be important for the selection of foraging or oviposition sites[48], is also found from this loci. Invasive populations have reduced host plant ranges compared with native populations[27,49]. One of the possible explanations of this reduction is the genetic differentiation of the serine protease inhibitor gene or the odorant receptor gene by genetic linkage to selectively targeted carboxylesterase and mdr49 genes. In this explanation, the reduction of host-plant ranges is a by-product of the process of adaptive evolution to reduce environmental stress. However, the possibility of divergent selection on the host plant should be considered as well. Interestingly, this locus includes clk, a key circadian clock gene[50]. African populations of FAWs have an earlier mating time than American populations by three hours[51]. The genetic differentiation of clk could also be caused by genetic linkage to positively selected environmental stress genes or host-plant genes, while divergent selection on the circadian clock is also possible.

CNV exhibits two groups in the invasive population (Fig. 2A), unlike SNV. The first group includes Benin and India, and the second group includes China, Malawi, and Uganda. We tested the presence of positive selection by CNV that is specific to one or both groups in invasive populations. F_{ST} calculated from CNV between Benin-India and China-Malawi-Uganda is 0.0397 (Fig. 5B). F_{ST} calculated from SNV between these two groups is 0.00973, which represents only 24.5% of F_{ST} from F_{ST} from CNV (0.00973/0.0397). In order to test if CNV having much higher F_{ST} than SNV is a general phenomenon, we also calculated F_{ST} between pairs among native hybrid populations, sfC group, and sfR group. The ratio of F_{ST} between these pairs from SNV to CNV ranges from 0.607 to 2.36 (Fig. 5C), which is higher than the ratio of F_{ST} between Benin-India and China-Malawi-Uganda (0.245). Thus, we concluded that the F_{ST} calculated from CNV between Benin-India and China-Malawi-Uganda could be affected by positive selection on CNV. In total, six loci with CNV have almost complete genetic differentiation between the two groups ($F_{ST} > 0.8$).

We identified only one gene, Decaprenyl-diphosphate synthase subunit 2 (DDSS2), from these loci.

Most individuals in the China-Malawi-Uganda group have this gene as single-copy, while the Benin-India group lacks this gene in most individuals. In FAW, the DDSS gene is down-regulated by bat-induced stress[52], and a region near Benin exhibits a hotspot for bat-species diversity[53]. Thus, the CNV of DDSS gene could possibly be a consequence of adaptation to local bat communities in West Africa (or India). More ecological studies are required to test the differential stress from predators across multiple invasive populations.

In this study, we showed that the restoration of the level of heterozygosity by genomic balancing selection is key to invasive success in FAW and that it likely enables its rapid global invasion of the Old World. We do not argue that invasive FAW in Western Africa obtained a new trait by adaptive evolution that increased invasiveness (e.g., Bridgehead Effects[54,55]). FAWs in native populations exhibit high migratory behavior, and invasive populations have probably equally high mobility as native populations. Instead, we argue here that the generation of a stable population in West Africa by genomic balancing selection played a key role in invasive success in FAW.

In addition, we do not argue that West Africa is the only initially invaded area. It is possible that the initial introduction of FAW might occur elsewhere in the Eastern Hemisphere[28], while invasive FAW remained undetected due to their small population size. We argue here that genomic balancing selection is one of the causal evolutionary forces responsible for explosive population growth in West Africa by facilitating admixtures and that this population migrated eastward, as shown from the chronological order of detection of invasive FAW. If populations of FAW existed in the Eastern Hemisphere before the first detection in West Africa, potential gene flow among invasive populations could explain the different patterns of ancestry coefficients between CNV (Fig. 2A) and SNV (Fig. 1B) among invasive populations.

The majority of reported cases show that the reduction in heterozygosity is mild (e.g., < 20%) in a wide range of taxa[1]. Therefore, it could be postulated that balancing selection may play a key role in the invasive success of a large range of organisms. Future studies should involve population genomics analysis in other invasive taxa to test this possibility. This study also highlights the

355 importance of rapid and vigorous pest control during the early phase of the invasion, as emphasized
356 by many researchers, before heterozygosity is sufficiently increased to generate a stable population
357 by genomic balancing selection. For an early eradication, early monitoring of pest species is
358 mandatory, and a small number of individuals should not be overlooked, like the case of the Asian
359 hornet (*Vespa velutina* Lepeletier) that started from a small invasive population which then went on
360 to rapidly colonize large areas of Western Europe[56].

361 **Methods**

362 ***Genome assembly***

363 We performed the mapping of Illumina reads (~80X)[57] against an assembly, which was generated
364 from 30X PacBio Reads in our previous study[32], using SMALT[58], and potential errors in the
365 assemblies were identified using reapr[59]. If an error is found over a gap, the scaffold was broken
366 into two using the same software to remove potential structural errors in the assembly. The broken
367 assemblies were concatenated using SALSA2[60] or 3D-DNA[61], followed by gap filling with the 80X
368 Illumina reads using SOAP-denovo2 Gap-Closer[62] and with the PacBio reads using LR_GapCloser
369 v1.1[63]. We observed that 3D-DNA generated a slightly more correct assembly than SALSA2 from
370 BUSCO analysis (Table S1). Thus, the assembly from 3D-DNA was used in this study. Gene
371 annotation was transferred from the previously generated assemblies to current assembly using
372 RATT[64].

373

374 ***Resequencing Data***

375 FAW larvae were collected from Wagou and Gando Villages in Benin (2017), from Citra and
376 Jacksonville in Florida (2015), from Texcoco in Mexico (2009), from French Guiana (1992), and
377 from Petit-Bourg and Port-Louis in Guadeloupe (2013). We obtained gDNA from India, which was

used by Sharanabasappa et al[65]. Genomic DNA was extracted using the Wizard Genomic DNA kit or the Qiagen Dneasy blood and tissues kit. Libraries for whole genome resequencing were constructed from 1.0µg DNA per sample using NEBNext DNA Library Prep Kit. Novaseq 6000 with ~20X coverage was used to perform whole genome resequencing with 150bp paired-end and 300bp insert length. Then, we combined the resequencing data from Puerto Rico and Mississippi, which were generated for our previous studies (Hiseq 2500, Hiseq 4000, and Novaseq 6000)[31,32], as well as the resequencing data of Brazil, Malawi, and Uganda from CSIRO (Novaseq 6000, 150bp paired-end sequencing)[28]. Lastly, resequencing data from China[33] was also combined with the dataset. Adapter sequences were removed using adapterremoval[66]. Then, we performed mapping of reads against the reference genome using bowtie2[67]. Then, we performed a variant calling using GATK[34] . Filtering was performed if QD is lower than 2.0, or FS is higher than 60.0, or MQ is lower than 40.0, or MQRankSum is lower than -12.5, or ReadPosRankSum is lower than -8.0. CNVs were identified using CNVCaller[68]. We discarded all CNVs unless minor allele frequency is higher than 0.2 to reduce false positives.

Phylogenetic analysis

To identify strains, we mapped the Illumina reads against mitochondrial genomes (NCBI: KM362176) using bowtie2[67], followed by extracting mitochondrial reads using samtools[69]. Mitochondrial genomes were assembled using MitoZ[70], and COX1 sequences were identified. These COX1 sequences were aligned together with a COX1 sequence from a specimen of another *Spodoptera* species, *S. exigua* (NCBI ID, JX316220), using MUSCLE [71], and a maximum likelihood phylogenetic tree was reconstructed using PhyML[72]. The phylogenetic tree was visualized using iTOL [73].

We calculated the nuclear genetic distance between each pair of individuals from the difference in allele frequency at biallelic sites in which genotypes are determined from all individuals using VCFphylo (https://github.com/kiwoong-nam/VCFPhylo). Transversional variants were weighted to

two. Then, a bootstrapping distance matrix was generated with 1,000 replications, and we generated BIO-NJ trees for each matrix using FastME[74]. Then, a consensus tree was made using consense in Phylip package[75], and the tree was visualized using iTOL[73].

Population genomics analysis

The principal component analysis was performed using plink[76]. We used admixture[36] for the ancestry coefficient analysis. Weir and Cockerham's F_{ST}[77] was calculated using VCFtools[78]. Potential targets of selective sweeps were identified using SweeD[41]. The number of the grid is 1,000 per chromosome. If a locus has the composite likelihood of selective sweeps higher than 100, we considered that this locus was targeted by a selective sweep. The decay curves of linkage disequilibrium were generated using PopLDdecay[79]. To identify mitochondrial SNVs, a mitochondrial VCF was generated from the bam files, which was made to identify strains (see above), using GATK[34].

415 # References

1. Estoup, A. *et al.* Is there a genetic paradox of biological invasion? *Annu. Rev. Ecol. Evol. Syst.* **47**, 51–72 (2016).
2. Lee, C. E. Evolutionary genetics of invasive species. *Trends Ecol. Evol.* **17**, 386–391 (2002).
3. Whitney, K. D. & Gabler, C. A. Rapid evolution in introduced species, 'invasive traits' and recipient communities: challenges for predicting invasive potential. *Divers. Distrib.* **14**, 569–580 (2008).
4. Charlesworth, D. & Willis, J. H. The genetics of inbreeding depression. *Nat. Rev. Genet.* **10**, 783–796 (2009).
5. Lanfear, R., Kokko, H. & Eyre-Walker, A. Population size and the rate of evolution. *Trends Ecol. Evol.* **29**, 33–41 (2014).
6. Grossman, S. R. *et al.* Identifying recent adaptations in large-scale genomic data. *Cell* **152**, 703–713 (2013).
7. Nam, K. *et al.* Evidence that the rate of strong selective sweeps increases with population size in the great apes. *Proc. Natl. Acad. Sci.* **114**, 1613–1618 (2017).
8. Hermisson, J. & Pennings, P. S. Soft Sweeps: Molecular population genetics of adaptation from standing genetic variation. *Genetics* **169**, 2335–2352 (2005).
9. Allendorf, F. W. & Lundquist, L. L. Introduction: population biology, evolution, and control of invasive Species. *Conserv. Biol.* **17**, 24–30 (2003).
10. Corbett-Detig, R. B., Zhou, J., Clark, A. G., Hartl, D. L. & Ayroles, J. F. Genetic incompatibilities are widespread within species. *Nature* **504**, 135–137 (2013).
11. Montezano, D. G. *et al.* Host plants of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) in the Americas. *Afr. Entomol.* **26**, 286–300 (2018).
12. Westbrook, J. K., Nagoshi, R. N., Meagher, R. L., Fleischer, S. J. & Jairam, S. Modeling seasonal migration of fall armyworm moths. *Int. J. Biometeorol.* **60**, 255–267 (2016).
13. Gutiérrez-Moreno, R. *et al.* Field-Evolved Resistance of the Fall Armyworm (Lepidoptera: Noctuidae) to Synthetic Insecticides in Puerto Rico and Mexico. *J. Econ. Entomol.* **112**, 792–802 (2019).
14. Mota-Sanchez, D. & John C., W. Arthropod Pesticide Resistance Database. https://www.pesticideresistance.org/index.php.
15. Storer, N. P. *et al.* Discovery and Characterization of Field Resistance to Bt Maize: *Spodoptera*

frugiperda (Lepidoptera: Noctuidae) in Puerto Rico. *J. Econ. Entomol.* **103**, 1031–1038 (2010).

16. Jakka, S. R. K. *et al.* Field-Evolved Mode 1 Resistance of the Fall Armyworm to Transgenic Cry1Fa-Expressing Corn Associated with Reduced Cry1Fa Toxin Binding and Midgut Alkaline Phosphatase Expression. *Appl. Environ. Microbiol.* **82**, 1023–1034 (2016).
17. Omoto, C. *et al.* Field-evolved resistance to Cry1Ab maize by *Spodoptera frugiperda* in Brazil. *Pest Manag. Sci.* **72**, 1727–1736 (2016).
18. Chandrasena, D. I. *et al.* Characterization of field-evolved resistance to Bacillus thuringiensis-derived Cry1F δ-endotoxin in *Spodoptera frugiperda* populations from Argentina. *Pest Manag. Sci.* **74**, 746–754 (2018).
19. Sparks, A. N. A review of the biology of the fall armyworm. *Fla. Entomol.* 82–87 (1979).
20. Goergen, G., Kumar, P. L., Sankung, S. B., 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* **11**, e0165632 (2016).
21. Day, R. *et al.* Fall Armyworm: impacts and implications for Africa. *Outlooks Pest Manag.* **28**, 196–201 (2017).
22. Pashley, D. P. Host-associated genetic differentiation in fall armyworm (Lepidoptera: Noctuidae): a sibling species complex? *Ann. Entomol. Soc. Am.* **79**, 898–904 (1986).
23. Pashley, D. P. & Martin, J. A. Reproductive incompatibility between host strains of the fall armyworm (Lepidoptera: Noctuidae). *Ann. Entomol. Soc. Am.* **80**, 731–733 (1987).
24. Dumas, P. *et al. Spodoptera frugiperda* (Lepidoptera: Noctuidae) host-plant variants: two host strains or two distinct species? *Genetica* **143**, 305–316 (2015).
25. Zhang, L. *et al.* High-depth resequencing reveals hybrid population and insecticide resistance characteristics of fall armyworm (*Spodoptera frugiperda*) invading China. *bioRxiv* 813154 (2019) doi:10.1101/813154.
26. Nagoshi, R. N. *et al.* 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* **12**, (2017).
27. Nagoshi, R. N., Goergen, G., Plessis, H. D., van den Berg, J. & Meagher, R. Genetic comparisons of fall armyworm populations from 11 countries spanning sub-Saharan Africa provide insights into strain composition and migratory behaviors. *Sci. Rep.* **9**, 8311 (2019).

28. Tay, W. T. *et al.* Whole genome sequencing of global *Spodoptera frugiperda* populations: evidence for complex, multiple introductions across the Old World. *bioRxiv* 2020.06.12.147660 (2020) doi:10.1101/2020.06.12.147660.

29. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).

30. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).

31. Nam, K. *et al.* Adaptation by copy number variation increases insecticide resistance in fall armyworms. *bioRxiv* 812958 (2019) doi:10.1101/812958.

32. Nam, K. *et al.* Divergent selection causes whole genome differentiation without physical linkage among the targets in *Spodoptera frugiperda* (Noctuidae). *bioRxiv* 452870 (2018) doi:10.1101/452870.

33. Liu, H. *et al.* Chromosome level draft genomes of the fall armyworm, *Spodoptera frugiperda* (Lepidoptera: Noctuidae), an alien invasive pest in China. *bioRxiv* 671560 (2019) doi:10.1101/671560.

34. McKenna, A. *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

35. Lu, Y. & Adang, M. J. Distinguishing fall armyworm (Lepidoptera: Noctuidae) strains using a diagnostic mitochondrial DNA marker. *Fla. Entomol.* **79**, 48–55 (1996).

36. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).

37. Nagoshi, R. N. *et al.* Analysis of strain distribution, migratory potential, and invasion history of fall armyworm populations in northern Sub-Saharan Africa. *Sci. Rep.* **8**, 3710 (2018).

38. Mongue, A. J., Tsai, M. V., Wayne, M. L. & de Roode, J. C. Inbreeding depression in monarch butterflies. *J. Insect Conserv.* **20**, 477–483 (2016).

39. DeGiorgio, M., Lohmueller, K. E. & Nielsen, R. A model-based approach for identifying signatures of ancient balancing selection in genetic data. *PLoS Genet.* **10**, e1004561 (2014).

40. Anderson, C. J. *et al.* Hybridization and gene flow in the mega-pest lineage of moth, *Helicoverpa*. *Proc. Natl. Acad. Sci.* **115**, 5034–5039 (2018).

41. Pavlidis, P., Živković, D., Stamatakis, A. & Alachiotis, N. SweeD: likelihood-based detection of selective sweeps in thousands of genomes. *Mol. Biol. Evol.* **30**, 2224–2234 (2013).

42. Charlesworth, B., Morgan, M. T. & Charlesworth, D. The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289–1303 (1993).

43. McDonnell, A. M. & Dang, C. H. Basic review of the cytochrome P450 system. *J. Adv. Pract. Oncol.* **4**, 263–268 (2013).

44. Giraudo, M. *et al.* Cytochrome P450s from the fall armyworm (*Spodoptera frugiperda*): responses to plant allelochemicals and pesticides. *Insect Mol. Biol.* **24**, 115–128 (2015).

45. Cui, F. *et al.* Carboxylesterase-mediated insecticide resistance: Quantitative increase induces broader metabolic resistance than qualitative change. *Pestic. Biochem. Physiol.* **121**, 88–96 (2015).

46. Tapadia, M. G. & Lakhotia, S. C. Expression of mdr49 and mdr65 multidrug resistance genes in larval tissues of *Drosophila melanogaster* under normal and stress conditions. *Cell Stress Chaperones* **10**, 7–11 (2005).

47. Lin, H. *et al.* Characterization and expression profiling of serine protease inhibitors in the diamondback moth, *Plutella xylostella* (Lepidoptera: Plutellidae). *BMC Genomics* **18**, 162 (2017).

48. de Fouchier, A. *et al.* Functional evolution of Lepidoptera olfactory receptors revealed by deorphanization of a moth repertoire. *Nat. Commun.* **8**, 15709 (2017).

49. Goergen, G., Kumar, P. L., Sankung, S. B., 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* **11**, e0165632 (2016).

50. Tataroglu, O. & Emery, P. The molecular ticks of the *Drosophila* circadian clock. *Curr. Opin. Insect Sci.* **7**, 51–57 (2015).

51. Haenniger, S. *et al.* Sexual communication of *Spodoptera frugiperda* from West Africa: Adaptation of an invasive species and implications for pest management. *Sci. Rep.* **10**, 2892 (2020).

52. Cinel, S. D. & Taylor, S. J. Prolonged bat call exposure induces a broad transcriptional response in the male fall armyworm (*Spodoptera frugiperda*; Lepidoptera: Noctuidae) brain. *Front. Behav. Neurosci.* **13**, (2019).

53. Herkt, K. M. B., Barnikel, G., Skidmore, A. K. & Fahr, J. A high-resolution model of bat diversity and endemism for continental Africa. *Ecol. Model.* **320**, 9–28 (2016).

54. Lombaert, E. *et al.* Bridgehead effect in the worldwide invasion of the biocontrol Harlequin ladybird. *PLoS ONE* **5**, e9743 (2010).

55. Bertelsmeier, C. & Keller, L. Bridgehead effects and role of adaptive evolution in invasive populations.

Trends Ecol. Evol. **33**, 527–534 (2018).

56 Rortais, A. *et al.* A new enemy of honeybees in Europe: the Asian Hornet, Vespa velutina. (2010).

57. Gouin, A. *et al.* Two genomes of highly polyphagous lepidopteran pests (*Spodoptera frugiperda*, Noctuidae) with different host-plant ranges. *Sci. Rep.* **7**, 11816 (2017).

58. Sanger Institute. *SMALT (https://www.sanger.ac.uk/tool/smalt-0/)*.

59. Hunt, M. *et al.* REAPR: a universal tool for genome assembly evaluation. *Genome Biol.* **14**, R47 (2013).

60. Ghurye, J. *et al.* Integrating Hi-C links with assembly graphs for chromosome-scale assembly. *PLOS Comput.Biol.* **15**, e1007273 (2019).

61. Dudchenko, O. *et al.* De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).

62. Luo, R. *et al.* Erratum: SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* **4**, 30 (2015).

63. Xu, G.-C. *et al.* LR_Gapcloser: a tiling path-based gap closer that uses long reads to complete genome assembly. *GigaScience* **8**, (2018).

64. Otto, T. D., Dillon, G. P., Degrave, W. S. & Berriman, M. RATT: Rapid Annotation Transfer Tool. *Nucleic Acids Res.* **39**, e57 (2011).

65. Sharanabasappa, et al. First report of the fall armyworm, *Spodoptera frugiperda* (J E Smith) (Lepidoptera: Noctuidae), an alien invasive pest on maize in India. *Pest Manag. Hortic. Ecosyst.* **24**, 23–29 (2018).

66. Schubert, M., Lindgreen, S. & Orlando, L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **9**, 88 (2016).

67. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

68. Wang, X. *et al.* CNVcaller: highly efficient and widely applicable software for detecting copy number variations in large populations. *GigaScience* **6**, (2017).

69. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

70. Meng, G., Li, Y., Yang, C. & Liu, S. MitoZ: a toolkit for animal mitochondrial genome assembly, annotation and visualization. *Nucleic Acids Res.* **47**, e63–e63 (2019).

71. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

72. Guindon, S. *et al*. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).

73. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).

74. Lefort, V., Desper, R. & Gascuel, O. FastME 2.0: A comprehensive, accurate, and fast distance-based phylogeny inference program. *Mol. Biol. Evol.* **32**, 2798–2800 (2015).

75. Plotree, D. & Plotgram, D. PHYLIP-phylogeny inference package (version 3.2). *cladistics* **5**, 163–166 (1989).

76. Rentería, M. E., Cortes, A. & Medland, S. E. Using PLINK for genome-wide association studies (GWAS) and data analysis. in *Genome-Wide Association Studies and Genomic Prediction* (eds. Gondro, C., van der Werf, J. & Hayes, B.) 193–213 (Humana Press, 2013). doi:10.1007/978-1-62703-447-0_8.

77. Weir, B. S. & Cockerham, C. C. Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370 (1984).

78. Danecek, P. *et al*. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

79. Zhang, C., Dong, S.-S., Xu, J.-Y., He, W.-M. & Yang, T.-L. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* **35**, 1786–1788 (2019).

416

417

418 # End notes

419 ## Acknowledgements

420 This work (ID 1702-018, given to KN) was publicly funded through ANR (the French National
421 Research Agency) under the "Investissements d'avenir" programme with the reference ANR-10-
422 LABX-001-01 Labex Agro and coordinated by Agropolis Fondation under the frame of I-SITE
423 MUSE (ANR-16-IDEX-0006). In addition, a grant from the department of Santé des Plantes et
424 Environnement at Institut national de la recherche agronomique for KN (adaptivesv). This work
425 was also financially supported by EUPHRESCO (FAW-spedcom, given to Anne-Nathalie Volkoff)
426 and by CSIRO Health & biosecurity (given to WTT, TW, and KG). SY was supported by a
427 CIRAD-INRAE PhD fellowship.

428

429 ## Author Contributions

430 FL generated reference genome assembly. WTT, MF, SD, RA, CMK, RLMJ, CAB, PS, TB, AD,
431 TW, KG, and NN provided samples for whole genome resequencing. EF, ANC, SG, and GJK
432 prepared samples. EF performed variant calling. SY and KN performed analysis. NN and EA
433 performed gene annotation. SY and KN wrote manuscript. KN involved in planning and supervised
434 the work.

435

436 ## Competing interests

437 The authors declare no competing interests.

438

439 ## Additional Information

440 The raw reads of these samples are available from NCBI SRA (PRJNA639296 for samples from
441 Florida and PRJNA639295 for the rest of the samples). The reference genome assembly used in this
442 study is available at BIPAA (https://bipaa.genouest.org/sp/spodoptera_frugiperda).

443 Supplementary Information is available for this paper. We declare a full code availability upon

444 request.

Figure legends

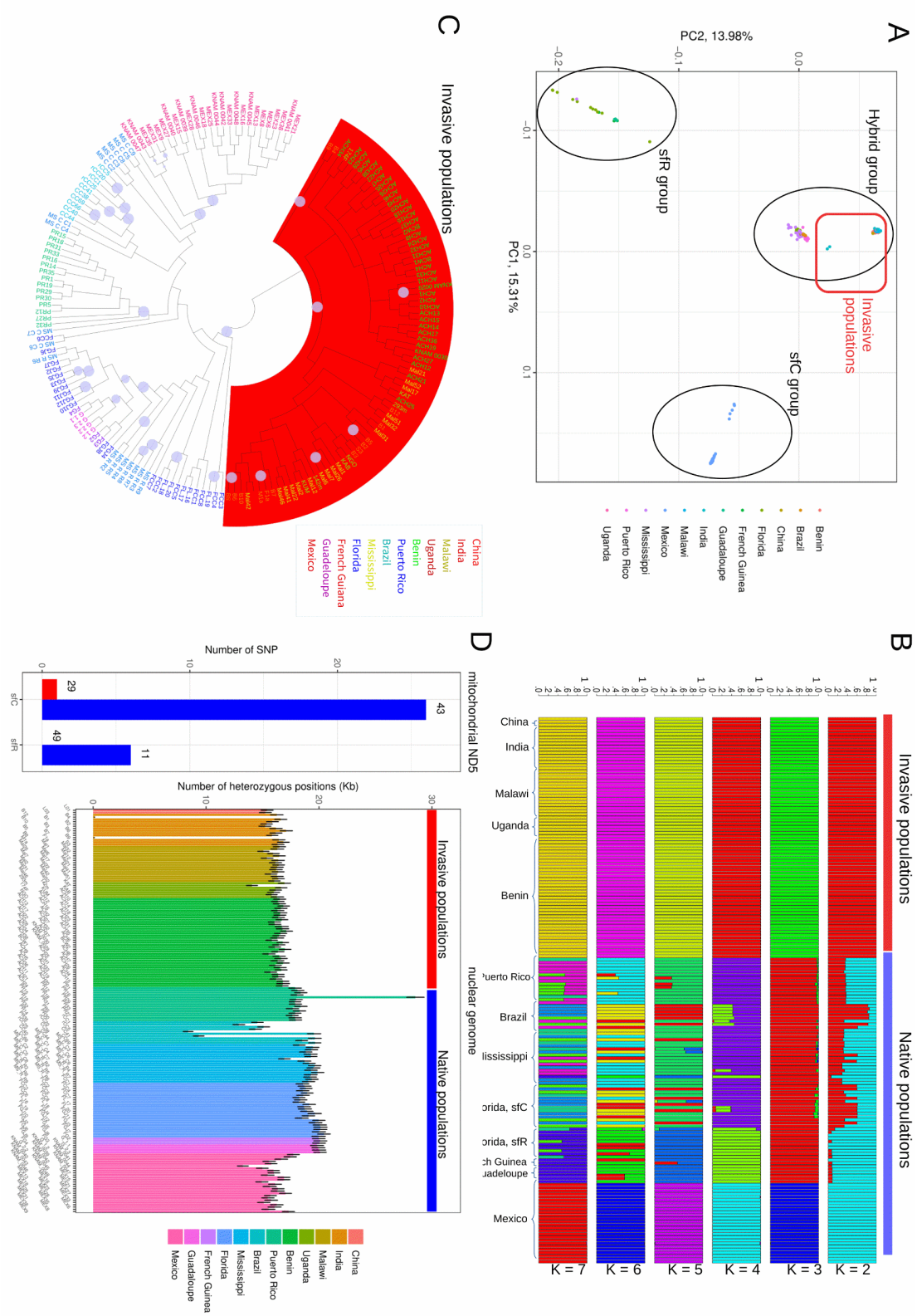
Figure 1. **Population structure of fall armyworms.** A. Principal component analysis. B. Ancestry coefficient analysis with varying K values. C. BIO-NJ phylogenetic tree was reconstructed from the allelic differentiation between a pair of individuals with 1,000 replication of bootstrapping. The circles on the branches show bootstrapping support higher than 90%. D. (left) The numbers of SNPs on the mitochondrial ND5 gene in sfC and sfR. The numbers above the bars indicate the number of sequences. (right) The number of heterozygous positions counted from positions of which genotypes are determined from all individuals. The error bars indicate 95% confidence intervals calculated from 1,000 times of bootstrapping replications in the way of resampling from 100kb windows.

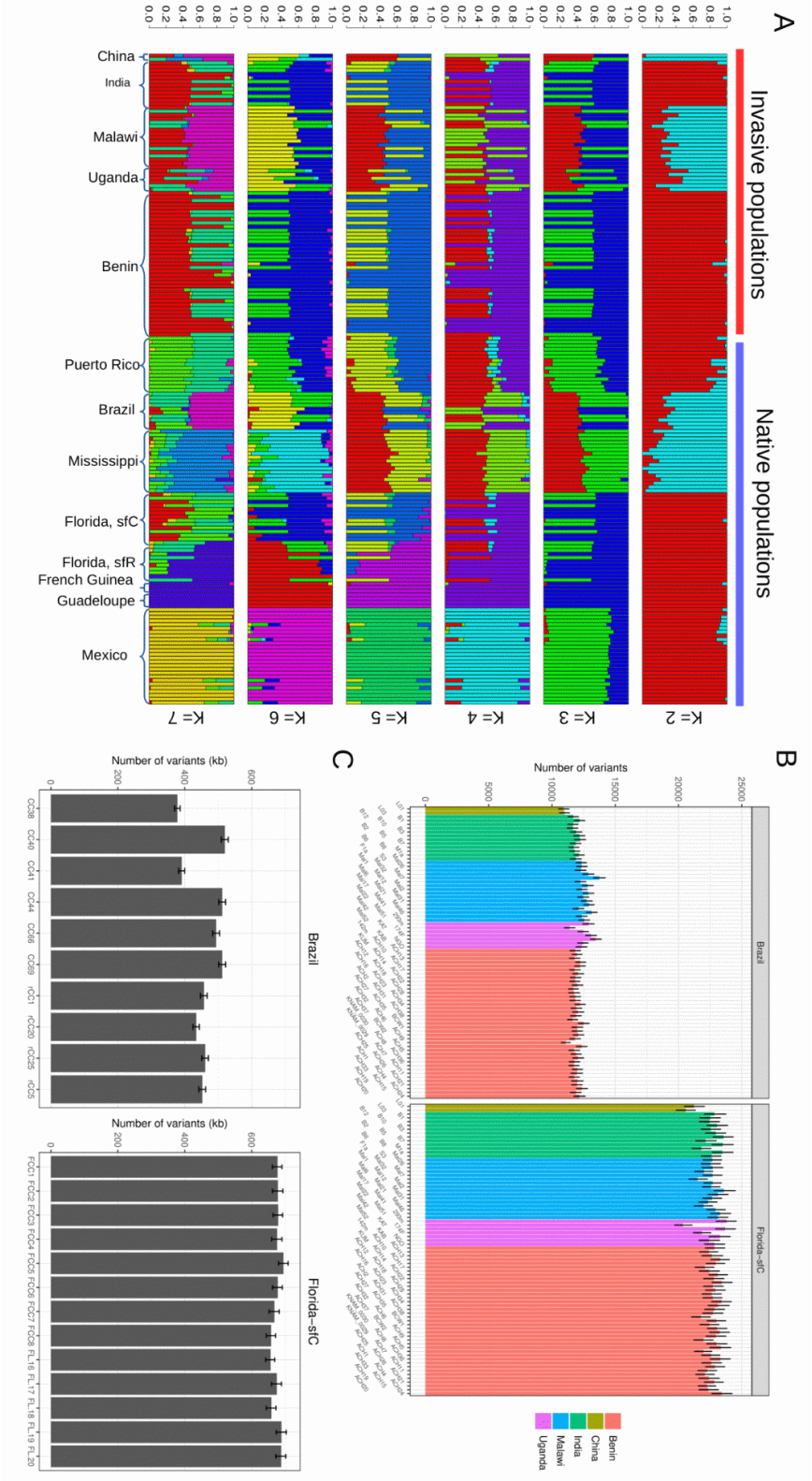
Figure 2. **Multiple introduction of invasive fall armyworm** A. Ancestry coefficient analysis of CNV with varying K values. B. (left) The number of SNVs specifically found from the population in Brazil and absent from all the other populations, counted from each individual in the invasive populations. (right) The number of SNVs specifically found from sfC-Florida and absent from all the other populations, counted from each individual from invasive populations. C. (left) The number of SNVs in each of individuals from Brazil that are not found from sfC-Florida. (right) The number of SNVs in each of individuals from sfC-Florida that are not found from Brazil. The error bars indicate 95% confidence intervals calculated from 1,000 times of bootstrapping replication in the way of resampling from 100kb windows.

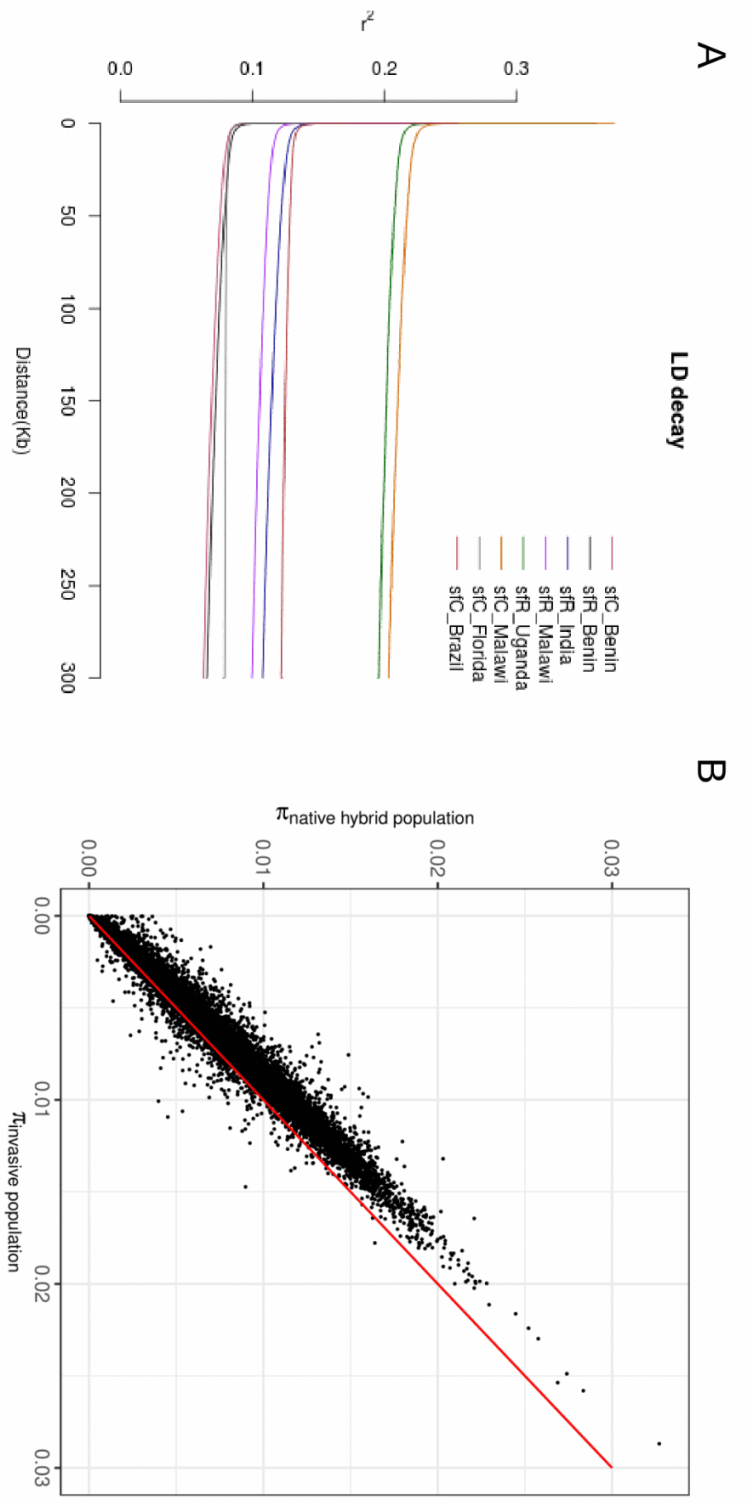
Figure 3. **Genomic balancing selection.** A. The LD decay curves calculated from each strain in each invasive population and their origins (sfC_Brazil and sfC_Florida). B. Correlation of nucleotide diversity between invasive populations and native hybrid population.

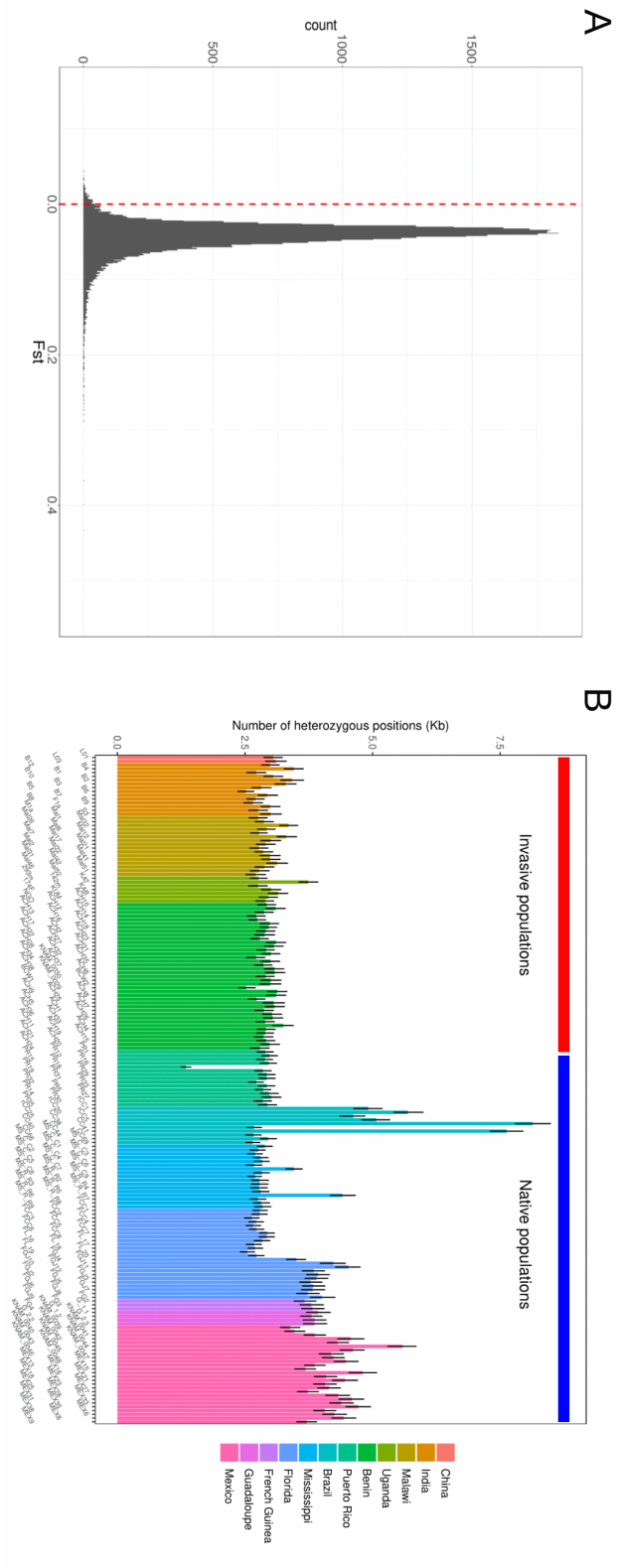
Figure 4. **Testing interspecific hybridization** A. Histogram of F_{ST} calculated from 100kb windows between invasive populations and native hybrid groups. The red vertical bar indicates F_{ST} equals to zero. B. Homozygous variant positions were counted for each individual. The error bars indicate 95% confidence intervals calculated from 1,000 times of bootstrapping replication in the way of resampling from 100kb windows.

Figure 5. **Loci under positive selection** A. The composite likelihood of being targeted by selective sweeps in invasive populations. The red asterisks indicate invasive population-specific outliers of the composite likelihood (>100), potentially targeted by selective sweeps. B. F_{ST} calculated from pairs of groups in CNV and SNV. The error bars indicate 95% confidence intervals calculated from 1,000 times of bootstrapping replication in the way of resampling from 100kb windows. C. Allele frequency of the CNV locus containing the DDSS gene. CH0, CH1, and CH2 indicate zero, one, and two copies in a haploid genome, respectively.









488