



# OPEN The evolutionary process of invasion in the fall armyworm (*Spodoptera frugiperda*)

Sudeeptha Yainna<sup>1,2</sup>, Wee Tek Tay<sup>3</sup>, Karine Durand<sup>1</sup>, Estelle Fiteni<sup>1</sup>, Frédérique Hilliou<sup>4</sup>, Fabrice Legeai<sup>5,6</sup>, Anne-Laure Clamens<sup>7</sup>, Sylvie Gimenez<sup>1</sup>, R. Asokan<sup>8</sup>, C. M. Kalleshwaraswamy<sup>9</sup>, Sharanabasappa S. Deshmukh<sup>9</sup>, Robert L. Meagher Jr.<sup>10</sup>, Carlos A. Blanco<sup>11</sup>, Pierre Silvie<sup>2,12,13</sup>, Thierry Brévault<sup>2,12</sup>, Anicet Dassou<sup>14</sup>, Gael J. Kergoat<sup>6</sup>, Thomas Walsh<sup>3</sup>, Karl Gordon<sup>3</sup>, Nicolas Nègre<sup>1</sup>, Emmanuelle d'Alençon<sup>1</sup> & Kiwoong Nam<sup>1</sup>✉

The fall armyworm (FAW; *Spodoptera frugiperda*) is one of the major agricultural pest insects. FAW is native to the Americas, and its invasion was first reported in West Africa in 2016. Then it quickly spread through Africa, Asia, and Oceania, becoming one of the main threats to corn production. We analyzed whole genome sequences of 177 FAW individuals from 12 locations on four continents to infer evolutionary processes of invasion. Principal component analysis from the TPI gene and whole genome sequences shows that invasive FAW populations originated from the corn strain. Ancestry coefficient and phylogenetic analyses from the nuclear genome indicate that invasive populations are derived from a single ancestry, distinct from native populations, while the mitochondrial phylogenetic tree supports the hypothesis of multiple introductions. Adaptive evolution specific to invasive populations was observed in detoxification, chemosensory, and digestion genes. We concluded that extant invasive FAW populations originated from the corn strain with potential contributions of adaptive evolution.

Biological invasion by pest species is one of the main causes of economic losses in agriculture<sup>1–4</sup>. As human trade is constantly increasing, the number of reported cases of invasive species has consequently surged up<sup>5,6</sup>, particularly in insect species<sup>7</sup>. Indeed, insects comprise 88.24% of reported invertebrate invasive cases in the Global Invasive Species Database<sup>8</sup> (accessed on 27th July 2021). Consequently, invasive pest insects cause serious losses in major crop production<sup>9</sup>, especially in sub-Saharan African countries, where the national economy is heavily dependent on agriculture, as up to 35% of national gross domestic product (GDP) is lost due to invasive pests<sup>10</sup>.

The process of an invasion typically involves three key steps<sup>11</sup>. The first step is the introduction of an alien species outside its native range. Invasive species are often introduced through anthropogenic transport, such as trade, urbanization, irrigation, and roads and railways<sup>12</sup>. The second step is the establishment of stable populations, thanks to a higher intrinsic growth rate<sup>13</sup> or a higher competitive ability<sup>14</sup> than native species. The third step is range expansion from the established population. Adaptive evolution may contribute to all of these steps by (i) fixation of beneficial de novo mutations or standing genetic variations, (ii) hybridization between introduced and native species, or (iii) genome doubling (reviewed in<sup>11</sup>). A time gap between introduction and range expansion has been often observed, and this gap period was coined the ‘lag phase’<sup>15</sup>. Since the eradication of an introduced pest species is only realistic during its lag phase, it is important to identify alien species before they become invasive for pest management.

<sup>1</sup>DGIMI, Univ Montpellier, INRAE, Montpellier, France. <sup>2</sup>CIRAD, UPR AIDA, Montpellier, France. <sup>3</sup>Black Mountain Laboratories, CSIRO, Canberra, Australia. <sup>4</sup>INRAE, Institut Sophia Agrobiotech, Université Côte D’Azur, CNRS, Sophia Antipolis, France. <sup>5</sup>INRAE, UMR-IGEPP, Bioinformatics Platform for Agroecosystems Arthropods, Campus Beaulieu, 35042 Rennes, France. <sup>6</sup>INRIA, IRISA, GenOuest Core Facility, Campus de Beaulieu, Rennes, France. <sup>7</sup>CBGP, INRAE, CIRAD, IRD, Institut Agro, Univ Montpellier, Montpellier, France. <sup>8</sup>Division of Biotechnology, ICAR - Indian Institute of Horticultural Research, Bengaluru, India. <sup>9</sup>Department of Entomology, College of Agriculture, University of Agricultural and Horticultural Sciences, Shivamogga, India. <sup>10</sup>United States Department of Agriculture, Agricultural Research Service, Gainesville, FL, USA. <sup>11</sup>United States Department of Agriculture, Animal and Plant Health Inspection Service, Maryland, USA. <sup>12</sup>AIDA, Univ Montpellier, CIRAD, Montpellier, France. <sup>13</sup>PHIM, Univ Montpellier, IRD, CIRAD, INRAE, Institut Agro, Montpellier, France. <sup>14</sup>ENSBBA, UNSTIM, Dassa, Benin. ✉email: ki-woong.nam@inrae.fr

The fall armyworm, *Spodoptera frugiperda* (J.E. Smith) (FAW, Lepidoptera: Noctuidae: Noctuinae), is one of the most infamous insect pests due to its severe impact on major crops including corn, cotton, rice, sorghum, and soybean<sup>16,17</sup>. The FAW exhibits high dispersal ability combined with a marked migratory behavior<sup>18</sup>, the ability to rapidly develop resistance against synthetic insecticides<sup>19</sup> and *Bacillus thuringiensis* proteins<sup>20–22</sup>, alongside the potential for occasional outbreaks. The FAW consists of two host plant strains with differentiated ranges of host plants, the corn strain (sfC) and the rice strain (sfR) (named after their preferred host plants). These two strains are observed in sympatry throughout their native range<sup>23–25</sup>. Since sfC and sfR are morphologically indistinguishable, the two strains can only be identified with molecular markers<sup>26,27</sup>. The nuclear triosephosphate isomerase (TPI) gene on Z chromosome<sup>28</sup> and the mitochondrial cytochrome *c* oxidase subunit 1 (COX1) gene<sup>27,29</sup> are commonly used for this purpose.

The FAW is native to the Americas, and its invasion of the Old World was first reported in West Africa in 2016<sup>30</sup>. In the years following its first report, FAWs were detected throughout sub-Saharan Africa, followed by their widespread detection in India, South East Asia, East Asia, Egypt, and Oceania (<https://www.cabi.org/isc/fallarmyworm>). Invasive FAW larvae cause significant economic losses, especially on corn with a production yield loss of 21–53% in Africa<sup>31</sup>, where corn is one of the most important staple crops (more than 30% of total caloric intake comes from corn<sup>32</sup>). Chemical insecticides are commonly used to control invasive FAW populations, but field-evolved insecticide resistance against chemical insecticides has been reported from Chinese populations<sup>33,34</sup>.

Since the global invasion of FAWs is a pressing issue due to the severe economic impacts, there is an urgent need to obtain information on the population structure of invasive and native FAW populations in the context of geography and host plants. Gui et al.<sup>35</sup> reported that the Chinese populations most likely originated from African populations using population genomics analyses. Schlum et al.<sup>36</sup> also performed population genomics analyses of one invasive (Kenya) and four native (Argentina, Brazil, Puerto Rico, and mainland USA) populations, but they observed no clear pattern of population structure. Since these studies are based on the FAW samples collected from corn or sorghum, which are presumably the preferred plants of sfC<sup>25</sup>, the genetic relatedness among native sfC, native sfR and invasive FAW populations has yet to be determined in these studies.

Even though invasive FAW populations are found almost exclusively in sfC-preferred plants<sup>37–39</sup>, such as corns and sorghums, several genetics studies argued that invasive FAWs are hybrids between sfC and sfR. Marker-based studies showed that most invasive FAWs are sfC according to the mitochondrial COX1 genes and sfR according to the TPI genes, arguing that invasive FAWs were generated through interstrain hybridization<sup>40,41</sup>. But it is still unclear whether the same pattern can be supported if whole genome sequences are analyzed. Zhang et al. reported through genomics analysis that Chinese FAW populations are hybrids between sfC and sfR<sup>33</sup>. However, since they included only one sfC sample and one sfR sample from native populations, it is still unclear whether the same pattern will be observed when population data from native population are included. In short, the argument that invasive FAWs are hybrids between sfC and sfR needs to be revisited using population genomics data. If invasive FAWs are indeed hybrids, it is not easy to explain why they are found almost exclusively in sfC-preferred host plants.

In this study, we aim at inferring the population structure of FAW populations, including native sfC, sfR, and invasive populations using whole genome sequences from samples collected globally from North and South America (native range), East and West Africa, and South and East Asia (invasive ranges). Native FAW samples were collected from both corn (sfC preferred plants), as well as grass or rice (sfR preferred plants). We also identified loci with signatures of adaptive evolution that may contribute to the FAW invasive success.

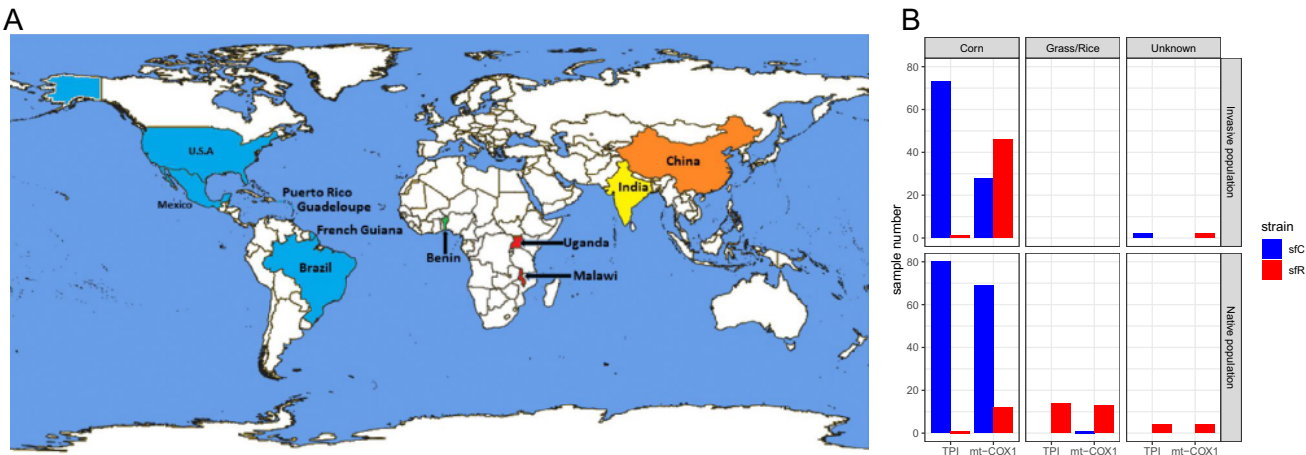
## Results

**Invasive origins.** The total number of individuals used in this study was 177 (99 from native populations and 78 from invasive populations) from 12 geographic populations (Fig. 1A and Table 1), making this dataset the largest and most comprehensive to date. The identified strain had an almost perfect correlation with host plants (Fig. 1B) according to the TPI marker. When the mitochondrial COX1 marker was used, almost all samples from rice or grasses were sfR, but sfR was also often observed from corn plants.

The mitochondrial COX1 phylogenetic tree recovered two clades corresponding to sfC and sfR mitochondrial strains with high bootstrap confidence scores (bootstrap support value of 97.3%) (Fig. 2A). The existence of two mitochondrial strains in invasive populations suggests a possibility of multiple introductions from the Americas to the Old World, as shown by Tay et al.<sup>49</sup>.

A principal component analysis was performed using 27,117,672 nuclear SNVs to infer FAW population structure. The first principal component explained 15.31% of the variance in allele frequency (Fig. S1A), of which the eigenvectors displayed three groups of individuals (Figs. 2B, S1B). The first group (referred to as ‘Rice group’ in this paper) consisted of samples collected from rice or grasses in the Caribbean (including Florida, Guadeloupe) and French Guiana, but also an individual from corn in the Mississippi. The second group (referred to as ‘Corn.Mex group’) consisted of samples from corn in Mexico only. The third group was found between the first and second groups along the first principal component. This group was composed of samples from corns only (referred to as ‘Corn group’).

The first principal component in Fig. 2B did not separate invasive populations (Benin, Malawi, Uganda, India, and China) from any native populations. But invasive populations were separated from the native populations in the Corn group along the second principal component, which explained 13.98% of the variance in allele frequency.  $F_{ST}$  statistics show significant genetic differentiation between the native and invasive populations in the Corn group ( $F_{ST} = 0.0432$ ;  $p < 0.01$ ). This result shows that genetic variations can be explained primarily by three groupings (e.g., Rice group, Corn.Mex group, and Corn group) and that the invasive populations were derived from the native population(s) in the Corn group.



**Figure 1. Samples used in this study** (A) Map indicating the countries from which samples of *Spodoptera frugiperda* were sequenced. The blue color indicates the native countries of *S. frugiperda*. The green, red, yellow, and orange colors indicate that the invasion was reported in 2016, 2017, 2018, and 2019 respectively. The map was generated using MapChart<sup>42</sup>. (B) The numbers of sFC and sFR samples according to the host plants using different diagnostic markers (TPI or mitochondrial COX1).

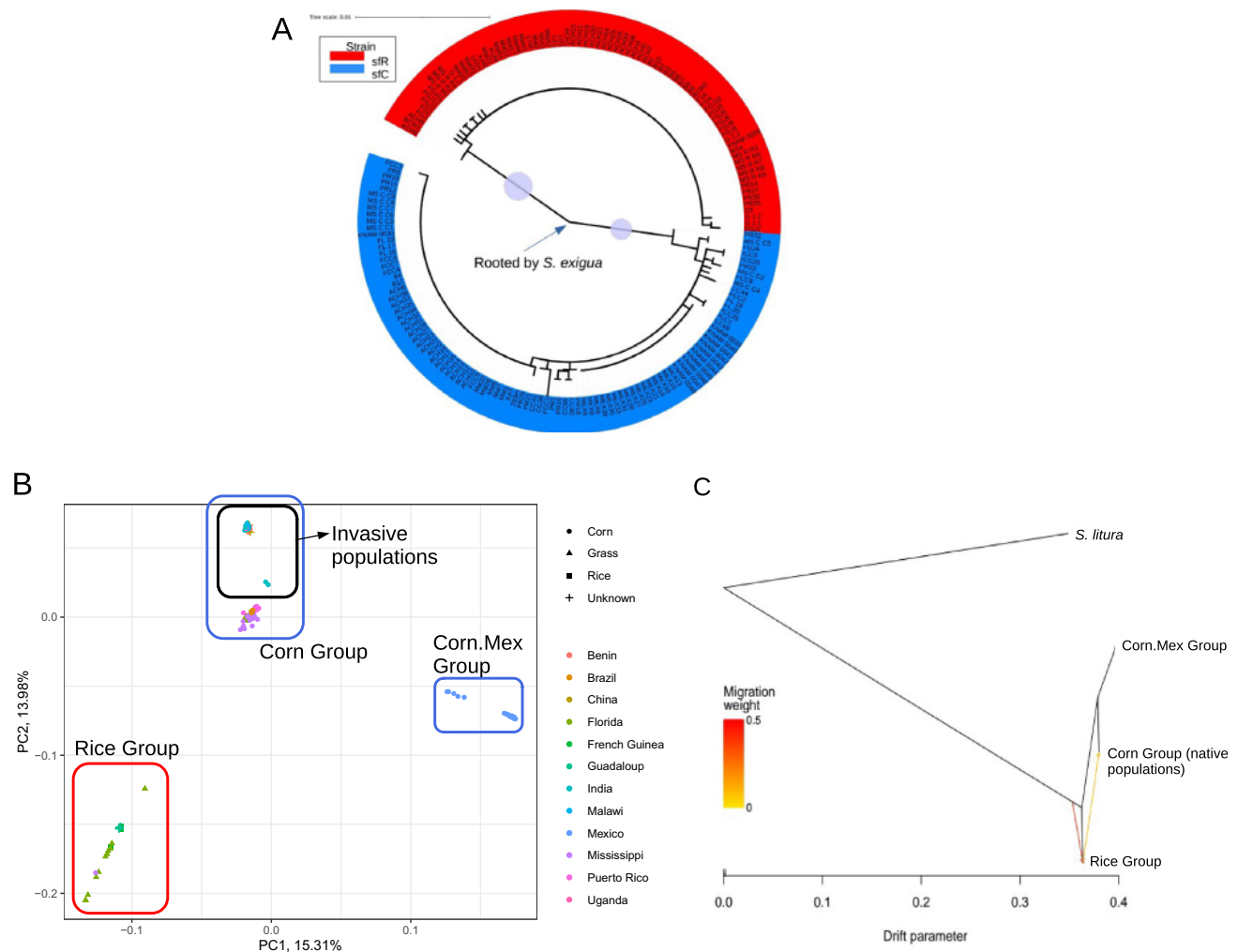
Country	District/village	The number of samples	Source
India	Karnataka	14	43
China	Kunming	2	44
Mexico	Texcoco	26	Current study
French Guiana		3	Current study
Guadeloupe	Petit-Bourg, Port-Louis	4	Current study
Puerto Rico	Santa Isabel	15	45
USA	Mississippi—Stoneville	17	46,47
	Florida—Citra and Jacksonville	24	48
Brazil	State of Goiás	10	49,50
Malawi	Blantyre, Chiradzulu, Machinga, Mulanje, Thyolo, Zomba, Salima, Karonga, Mzimba, Nkhata Bay	16	49,50
Uganda	Amolatar, Katakwi, Kumi, Ngora, Pallisa, Soroti, Mbarara, Wakiso	7	49,50
Benin	Wagou, Gando	10	Current study

**Table 1.** Information of resequencing data used in the study.

A whole nuclear genome phylogenetic tree reconstructed from native FAWs showed that ancestral FAWs split between the Rice group and (the Corn.Mex + the Corn groups) in the area of origin, and a unidirectional gene flow from Rice group to Corn group was observed when the number of migration edges was six (Fig. 2C). This phylogenetic relationship was consistently observed when different numbers of migration edges were assumed (Fig. S2). This result shows that the Corn group was derived from the ancestry of the Corn and Corn.Mex groups.

**A single ancestry of invasive FAWs.** We performed the ancestry coefficient analysis<sup>51</sup> to infer the ancestry of each individual. Invasive populations appeared to share a single common ancestor, distinct from all native populations, in a range of K values (Figs. 3A and S3). The single ancestry of invasive populations was further tested by reconstructing a BIO-NJ phylogenetic tree inferred from all individuals. If all the invasive individuals from populations originated from a single ancestry, these individuals will show monophyly with the assumption that invasive populations originated from native populations. The resulting tree exhibited a clade composed of invasive individuals with 100% of the bootstrapping support, and this clade does not contain any individuals from native populations (Fig. 3B). This result shows the monophyly of invasive populations.

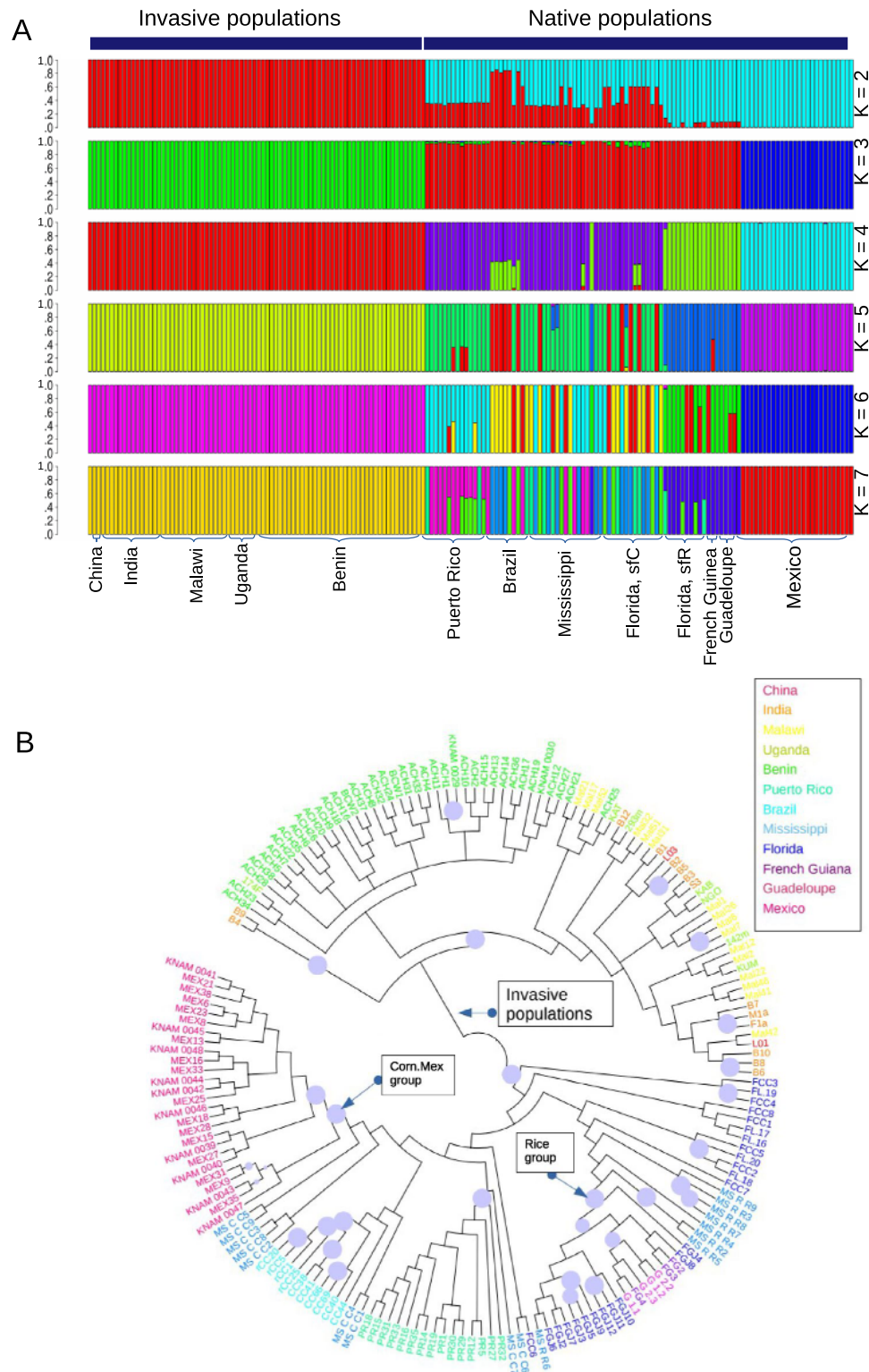
If invasive populations experienced genetic admixture between two sources with different mitochondrial strains (Fig. 2A), then invasive populations are expected to have lower  $D_{XY}$  between the mitochondrial strains than native populations due to the gene flow<sup>53</sup>.  $F_{ST}$  statistics showed nuclear genomic differentiation between the mitochondrial strains both in the invasive populations ( $F_{ST}=0.00281$ ;  $p<0.01$ ) and in the native populations ( $F_{ST}=0.0124$ ;  $p<0.01$ ) (Fig. S4A). Invasive populations had lower  $D_{XY}$  than native populations across whole genomes (Fig. S4B), suggesting that invasive populations experienced an increased rate of genetic exchanges between mitochondrial strains.



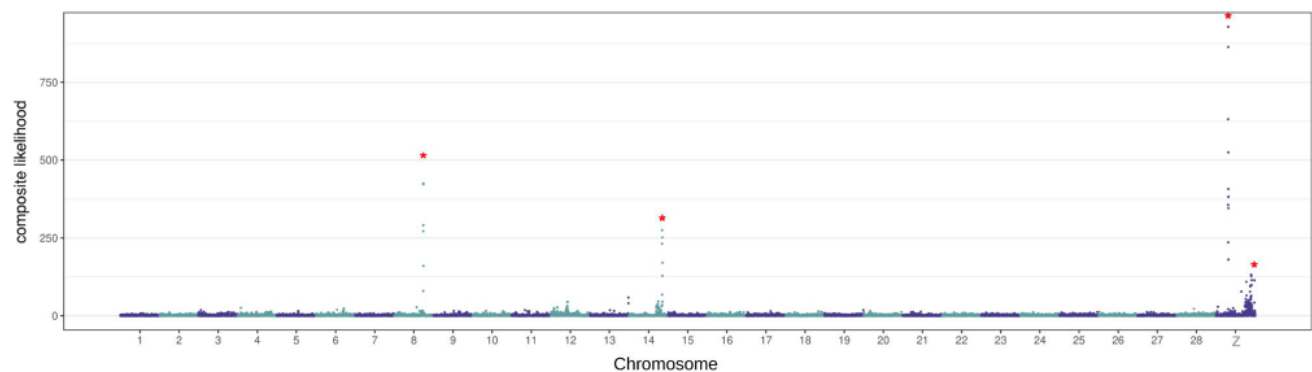
**Figure 2. Invasive origins inferred from mitochondrial or nuclear sequences** (A) Maximum-likelihood phylogenetic tree was reconstructed from the full-length mitochondrial COX1 gene (1536 bp). The red and blue clades indicate sFR and sFC, respectively. The circles on the branches show bootstrap support values higher than 90%. A branch leading to *Spodoptera exigua* was deleted while the position was indicated for a visualization purpose. (B) The result of the principal component analysis showed that FAW populations are composed of Rice, Corn, and Corn.Mex groups and that invasive populations belong to the Corn groups. (C) TreeMIX analysis shows that the Corn group was derived from the ancestry of the Corn.Mex and the Corn groups. The arrows indicate the direction of gene flow, which was detected from the Rice group to the Corn group. The color gradient of the arrows indicate the migration weight.

**Adaptively evolved genes in the invasive population.** We tested if adaptive evolution specific to invasive populations is supported using the composite likelihood of selective sweeps<sup>54</sup>. The median value of the composite likelihood was 0.4350. A locus was considered to be targeted by selective sweep if the composite likelihood was higher than 100, a threshold that was arbitrarily chosen. We were able to pinpoint the apparent outlier of the composite likelihood through eyeballing as well. In total, 26 grids had a composite likelihood higher than 100. These grids represented 0.090% of the total 29,000, meaning that a very stringent criterion was applied here to infer loci under selective sweeps. Neighboring grids constituted seven loci on three chromosomes. We considered that these loci are potential targets of selective sweeps (Fig. 4). As the high composite likelihood of these loci could be generated by selective sweeps not specific to invasive populations or by background selection<sup>55</sup>, we also calculated the composite likelihood from native individuals from the Corn group. Four out of the seven loci did not exhibit outliers of the composite likelihood in the native population in the Corn group (Fig. S5). We, therefore, considered that these four loci were potentially targeted by selective sweeps specific to invasive populations. These four loci contained 36 predicted protein-coding genes (Table S1), including 12 genes with unknown functions. We carefully underwent a manual curation of these genes by NCBI blasting to determine their functions. The locus on chromosome 14 encompasses CYP9A, which belongs to the Cytochrome P450 gene family. A locus on the Z chromosome includes a carboxylesterase gene, an ABC transporter homolog to *mdr49*, a kunitz-type serine protease inhibitor gene, odorant receptor 13, and the *clk* gene.





**Figure 3. A single origin of invasive populations** (A) Ancestry coefficient analysis from SNVs showed homogeneous genomic sequences of invasive populations. (B) A nuclear BIO-NJ phylogenetic tree was reconstructed from whole nuclear genome sequences. The circles on the branches show bootstrap support values higher than 90%. The Corn.Mex group and the Rice group are indicated to the corresponding nodes. The remaining individuals belong to the Corn Group. Invasive individuals were also indicated to the corresponding node. The phylogenetic tree was visualized using iTOL v4<sup>52</sup>.



**Figure 4. Loci under adaptive evolution** The composite likelihood of being targeted by selective sweeps in invasive populations. The x-axis represents the coordinates of grids in the reference genomes, and the y-axis is the composite likelihood of each grid. The red asterisks indicate potential targets of invasive population-specific selective sweeps.

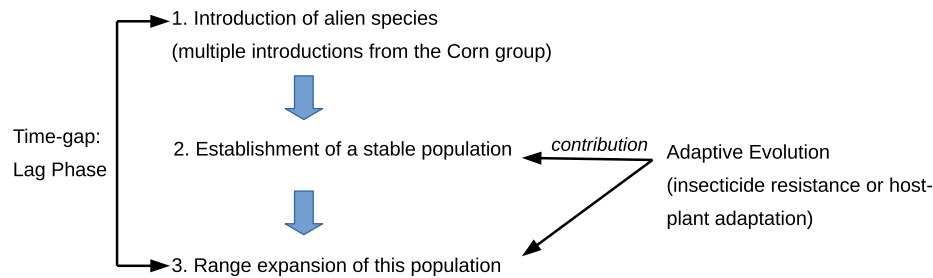
## Discussion

Invasive FAW insects have been one of the main threats against corn production across the Old World. In native areas, FAW is composed of two sympatric strains with differentiated host plants, such that sfC prefers corns and sfR prefers rice and grass<sup>23</sup>. Population structure among invasive FAWs, native sfC, and native sfR remained elusive. In this study, we inferred the population structure of FAW using whole genome sequences from 177 individuals collected from 12 geographic locations in native and invasive areas. Principal component analysis from the whole nuclear genome demonstrates that invasive FAWs originated from the corn-eating group (i.e., the Corn group) (Fig. 2B). The phylogenetic tree in Fig. 2C shows that the Corn group was derived from the ancestry of the Corn + Corn.Mex groups. We argue that the Corn group in Fig. 2B should be considered to be sfC, even though a proportion of this group was classified as sfR according to the mitochondrial COX1 marker because these samples were sampled only from corns. We reported that the TPI marker can be reliably used to identify host plant strains (sfC or sfR) whereas the COX1 marker should be used to identify two phylogenetic entities within sfC<sup>48</sup>. In this study, according to the TPI marker, almost all invasive FAWs were classified as sfC (Fig. 1B), in line with the conclusion that invasive FAWs should be considered to be sfC. The reason why invasive FAWs are found almost exclusively in sfC-preferred host plants is probably due to the fact that they are sfC. Therefore, we argue that invasive FAWs should not be considered to be hybrids anymore<sup>33,40,41</sup>.

Interestingly, the analyzed invasive FAWs appear to originate from a single population (Fig. 3) even though the mitochondrial phylogenetic tree shows that invasive FAWs are introduced from at least two genetic entities (e.g., sfC and sfR according to the mitochondrial COX1 marker) (Fig. 2A), as previously shown by Tay et al.<sup>49</sup>. Invasive populations appear to have experienced genetic admixture between these genetic entities with different mitochondrial markers (Fig. S4). We postulate that if FAWs were introduced from multiple origins a long time before 2016, when FAW invasion was first reported in West Africa<sup>30</sup>, the introduced FAWs might have had enough time to experience genetic admixtures. According to this hypothesis, introduced FAWs had genetically admixed nuclear genomes through recombination while non-recombining mitochondrial genomes were still distinctly separated. During this time of the lag phase, the introduced FAWs might have been undetected, possibly due to small population sizes before experiencing a population explosion in 2016. The presence of lag phases has been reported mostly in plants<sup>56</sup>, but also in insects<sup>57</sup>, birds<sup>58</sup>, and fishes<sup>59</sup>. In the FAWs, the existence of the lag phase should be formally tested in future studies.

**Adaptive evolution contributed to invasive success.** The identified genes under invasive population-specific selective sweeps (Fig. 4) suggest the possibility that insecticide resistance has contributed to the FAW invasive success. For example, the CYP9A gene on chromosome 14 may play an important role in host plant adaptation or insecticide resistance. CYP9A belongs to the cytochrome P450 gene family, a key player in the detoxification of xenobiotics<sup>45,60</sup>. CYP9A genes are overexpressed by plant allelochemicals and pesticides in FAW<sup>61</sup>, and CYP9A gene amplification has been observed in Puerto Rico<sup>45</sup>, where extensive field-evolved resistance against various insecticides has been reported in FAW<sup>19</sup>. A selectively targeted locus on the Z chromosome includes the carboxylesterase gene, which may be also involved in insecticide resistance<sup>62</sup>. Therefore, insecticide resistance could be one of the key selective pressures for invasion success. In our previous study, we observed that invasive FAW populations have higher gene copy numbers of CYP genes than native populations and that invasive populations have higher allele frequencies causing insecticide resistance<sup>63</sup>, consistent with the argument that insecticide resistance may have contributed to the invasive success. Indeed, invasive FAW populations have shown to have resistance against insecticides<sup>33,34</sup>.

The locus on the Z chromosome also includes an ABC transporter homolog to *mdr49*, which protects organisms from cytotoxic compounds in *Drosophila melanogaster*<sup>64</sup>. This locus also includes a Kunitz-type serine protease inhibitor gene, which plays a role in plant digestion<sup>65</sup>, and a gene encoding odorant receptor 13, which may be important for the selection of foraging or oviposition sites<sup>66</sup>. Since detoxification genes, digestion genes, and chemosensory genes are the key genetic elements determining the range of host plants in the FAW<sup>46</sup>, selective



**Figure 5. Models of the invasive process in FAW** First, FAWs were introduced from the Corn group between sfC and sfR, possibly multiple times. Second, a stable population was established in a location in the Old World. This population experienced growth in size, and FAW invasion was first reported. Third, range expansion occurred from this population. The first step occurred before 2016, implying the existence of a lag phase. Adaptive evolution contributed to the second and third steps, potentially through insecticide resistance or host plant adaptation.

sweeps on the locus including these genes may suggest the possibility that invasive FAWs experienced adaptation to host plants.

Another locus on the Z chromosome under selective sweeps includes *clk*, a key circadian clock gene<sup>67</sup>. Native populations have different mating times between sfC and sfR<sup>68,69</sup>, while invasive African populations of FAWs do not exhibit such different mating times while having earlier mating times than American populations by three hours<sup>70</sup>. The genetic differentiation of *clk* from native populations could promote hybridization between sfC and sfR through changes in mating time in African populations. Further genomics and behavior studies are required to test the association between genotypes at the *clk* gene and mating behavior.

**Process of FAW invasion.** Taken together, we propose the following evolutionary invasive scenario (summarized in Fig. 5). First, FAW has been introduced in the Old World multiple times. The introduced FAWs experienced a lag phase<sup>15</sup> and genetically admixed FAWs were generated. Second, a stable population at one location in the Old World was established, and this population experienced explosive growth in population size. At this time, the FAW invasion was first reported<sup>30</sup>. Third, invasive FAWs exhibited very rapid range expansion across sub-Saharan Africa, Asia, Oceania, and Egypt. We also argue that this proposed evolutionary history can be promoted by adaptive evolution, especially in the second and third stages.

The observed genetic differentiation between the analyzed invasive and native populations in the Corn group could be caused by multiple evolutionary forces. First, newly generated mutations during a lag phase might have contributed to differentiation. Second, multiple introductions of FAWs themselves may generate the observed genetic differentiation since genetically admixed individuals have differentiated genotypes from each of the origins. Third, mild positive selection targeting many genes may have accelerated the genomic differentiation between invasive and native populations through synergistic effects among linkage disequilibrium<sup>47,71</sup>. Since these explanations are not mutually exclusive, it is possible that these three evolutionary forces simultaneously increased genetic differentiation between invasive and native populations.

**Potential issues of sampling bias.** Since our sampling does not cover the entire range of FAW (especially in its native range in the Americas), the patterns observed in this study could be affected by temporal or geographical sampling bias, even though the analyzed FAWs were collected from a wide range of geographic locations in four continents. Since geography does not have a major effect on the population structure within the Corn group (Fig. 3A), we do not believe that an unknown geographic population in the Corn group has different genomic sequences from all samples analyzed in the native Corn group while being highly similar to the invasive Corn groups. However, it may be possible to pinpoint the exact invasive origins in native areas if additional geographic populations are included in future studies.

## Conclusion

In this study, we showed that invasive FAWs originated from sfC in the native area. The invasive FAWs appear to be derived from a single origin despite multiple introductions, suggesting the possibility of genetic admixture during a lag phase. We also observed footprints of adaptive evolution specific to invasive populations. Adaptive evolution specific to the identified invasive populations may have caused insecticide resistance or host plant adaptation, possibly contributing to invasive success. We argue that comparative genomics and functional studies on host plant adaptation and insecticide resistance are urgently needed to identify the main evolutionary selective forces responsible for invasive success, as this information may be useful to better control FAW in the Old World.

## Materials and methods

**Sampling and sequencing.** FAW larvae were collected from Wagou and Gando villages in Benin (2017), from Texcoco in Mexico (2009), from French Guiana (1992), and from Petit-Bourg and Port-Louis in Guadeloupe (2013). We obtained gDNA from India, which was used by Sharanabasappa et al.<sup>43</sup>. Genomic DNA was extracted using the Promega Wizard Genomic DNA Kit or the Qiagen DNeasy<sup>®</sup> blood and Tissue Kit. Libraries

for whole genome resequencing were constructed from 1.0 µg of DNA per sample using the NEBNext DNA Library Prep Kit. 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 the NovaSeq 6000 System with 20X coverage on average with 150 bp paired-end and 300 bp insert length. Then, we combined resequencing data from Santa Isabel in Puerto Rico (15) (NCBI SRA: PRJNA577869), Stoneville in Mississippi (17) (PRJNA494340), and from Citra and Jacksonville in Florida (USA) (24) (PRJNA639296), which were generated for our previous studies<sup>45,47,48</sup>. The resequencing data from Kunming in China (2)<sup>44</sup> was also included in the analysis. Finally, we took advantage of using resequencing data from the State of Goiás in Brazil (10), from nine districts including Blantyre, Chiradzulu, Machinga, Mulanje, Thyolo, Zomba, Salima, and Karonga, Mzimba, and Nkhata Bay in Malawi (16), and from eight districts including Amolatar, Katakwi, Kumi, Ngora, Pallisa, Soroti, Mbarara and Wakiso in Uganda (7), which were generated by Commonwealth Scientific and Industrial Research Organisation for other studies<sup>49,50</sup>. The total number of geographic locations is 12, which covers a large proportion of the FAW distribution area. The total number of individuals used in this study is 177 (99 from native populations and 78 from invasive populations) (Fig. 1 and Table 1). Since the geographic locations include both North and South America (native populations), West and East Africa, and South and East Asia (invasive populations), the resequencing data represents a large proportion of distribution in FAW.

**Variant calling.** Nucleotides with a Phred score less than 20 and adapter sequences were removed from the reads using AdapterRemoval v2.1.7<sup>72</sup>. Reads were mapped against the ver7 reference genome ([https://bipaa.genouest.org/sp/spodoptera\\_frugiperda\\_pub/download](https://bipaa.genouest.org/sp/spodoptera_frugiperda_pub/download))<sup>48</sup>, which had chromosome-sized scaffolds, using bowtie2 v2.3.4.1 with the -very-sensitive-local preset<sup>73</sup>. The read depth of mapping of resequencing data was 17.72X in the median. Then, we performed local re-assembly of haplotypes using GATK v4.1.2.0<sup>74</sup>. The resulting gvcf files were merged into one gvcf file with CombineGVCFs command and variant calling was performed with GenotypeGVCF command in the same software. Filtering was performed from the resulting VCF (variant calling format) file if QD is lower than 2.0, FS is higher than 60.0, MQ is lower than 40.0, or MQRankSum is lower than -12.5, or ReadPosRankSum is lower than -8.0. The total number of genetic variance is 158,277,530, which includes 89,387,997 SNVs (single nucleotide variations). After filtering, 27,117,672 SNVs remained.

**Strain identification.** FAW strains were identified using a phylogenetic tree reconstructed from the mitochondrial COX1 gene, which is the universal barcode gene and is also commonly used for FAW strain identification<sup>75</sup>. We first mapped the Illumina reads against mitochondrial genomes (NCBI: KM362176) using bowtie2 v2.3.4.1 with -very-sensitive-local preset<sup>73</sup>, followed by extracting mitochondrial reads using samtools v1.9<sup>76</sup>. Mitochondrial genomes were assembled from these reads, and mitochondrial genes were annotated using MitoZ using the default options<sup>77</sup>. Mitochondrial COX1 sequences were aligned together with a COX1 sequence from a specimen of another *Spodoptera* species, *S. exigua* (Hübner) (NCBI: JX316220), using MUSCLE v3.8.31<sup>78</sup>, and a maximum likelihood phylogenetic tree was reconstructed using MEGA X<sup>79</sup>. We used the Kimura 2-parameter model, and a uniform substitution rate across the alignment was assumed to avoid potential statistical artifacts caused by parameter richness. Non-parametric bootstrapping was performed with 1,000 replications to calculate statistical support at each node. The phylogenetic tree was visualized using iTOL v6<sup>52</sup>. We determined the mitochondrial strains from sFC and sFR clades with the individuals from Mississippi and Puerto Rico, in which strains were already determined for each individual in our previous studies<sup>45,47</sup>.

FAW strains were also identified using the TPI marker. First, we extracted the TPI locus from the vcf file using tabix v1.10.2-3<sup>80</sup>. Principal component analysis was performed from the TPI locus using plink v1.9<sup>81</sup>. Then, the strain of each sample was assigned according to the grouping.

**The inference of origin of the introduction.** We identified a native population that was genetically most closely related to the invasive populations from the population structure. The principal component analysis was performed to identify the main groups in FAW using plink v1.9<sup>81</sup>. Ten principal components were extracted, and the main groups in FAW were identified from two major principal components. Significant genetic differentiation between groups was tested using  $F_{ST}$ <sup>82</sup> calculated from VCFtools v0.1.15<sup>83</sup>.  $F_{ST}$  was calculated from 100 kb windows. The genomic average  $F_{ST}$  was calculated from Weir and Cockerham's weighted mean  $F_{ST}$ . Significant genetic differentiation between two groups was tested with a permutation test<sup>84</sup>. More specifically, (i)  $F_{ST}$  was calculated from two groups, (ii)  $F_{ST}$  was calculated from randomly generated groups with 100 replications, and (iii) the proportion of random groups that have higher  $F_{ST}$  than real groups was calculated. Since this proportion represents type I statistical error, the proportion was interpreted as a p-value. Ancestry coefficient analysis was performed using admixture v1.3.0 in a wide range of K values<sup>51</sup>.

Phylogenetic relationships with gene flow among the main groups of native populations were estimated to infer evolutionary history using TreeMIX v1.13<sup>85</sup>. First, a new VCF file with individuals from native populations was generated by subsetting the original VCF file with 177 individuals. Second, *S. litura* (Fabricius), which is one of the most closely related species to FAW in the *Spodoptera* genus<sup>86</sup>, was used as an outgroup, and a VCF file was generated by mapping *S. litura* Illumina reads (NCBI SRA, SRX2446353) against the FAW reference genome using bowtie2 with -very-sensitive-local preset<sup>73</sup> and by variant calling using GATK v4.1.2.0<sup>74</sup>. Third, these two VCF files were merged using VCFtools v0.1.15<sup>83</sup>, while missing data of *S. litura* was not allowed. The number of used SNVs is 697,498. Fourth, the newly merged VCF file was converted to TreeMIX format using vcf2treemix.sh (<https://github.com/speciationgenomics/scripts/blob/master/vcf2treemix.sh>). Fifth, we ran a TreeMIX binary with the number of migration edges ranging from three to six while the resulting tree was rooted by *S. litura*. The number of SNVs per block was 500 to estimate covariance matrices. Finally, the resulting file was visualized using the internal R library (plotting\_funcs.R) in the TreeMix software.



We also used the VCFphylo approach<sup>47</sup> to infer phylogenetic relationships at the level of individuals out of whole genome sequences. First, the euclidean distance of allele frequency between each pair of individuals was calculated from the biallelic variant positions in which genotypes are determined from all individuals using VCFphylo<sup>47</sup>. Transversional variants were weighted to two. Then, bootstrapping distance matrices were generated with 1,000 replications, and we generated BIO-NJ trees for each matrix using FastME v2.1.5<sup>87</sup>. Then, a consensus tree was made using consense in the Phylip package v3.697<sup>88</sup>. The tree was visualized using iTOL v6<sup>52</sup>.

**The inference of loci under selective sweeps.** Potential targets of selective sweeps were identified from the composite likelihood of being targeted by selected sweeps from the site frequency spectrum using SweepD v3.2.1<sup>54</sup> from the largest 29 scaffolds, which are believed to be chromosome-sized. The proportion of the assembly covered by these 29 scaffolds is 93.52%. The number of grids in which composite likelihood was calculated was 1,000 per scaffold. Based on the assumption that most genomic sequences were not targeted by selective sweeps, outliers of composite likelihood were identified as loci targeted by selective sweeps. The annotation of Cytochrome P450 genes was manually curated, as described in Gouin et al.<sup>46</sup>. The name of the CYP gene was validated by the P450 nomenclature committee<sup>89</sup>.

## Data availability

The resequencing dataset generated during the current study is available in the NCBI SRA (Accession Number: PRJNA639295).

## Code availability

Computer programming scripts used in this study are available on GitHub ([https://github.com/kiwoong-nam/FAW\\_invasion](https://github.com/kiwoong-nam/FAW_invasion)).

Received: 8 August 2022; Accepted: 30 November 2022

Published online: 06 December 2022

## References

- Pimentel, D. *et al.* Economic and environmental threats of alien plant, animal, and microbe invasions. *Agric. Ecosyst. Environ.* **84**, 1–20 (2001).
- Hulme, P. E. Trade, transport and trouble: managing invasive species pathways in an era of globalization. *J. Appl. Ecol.* **46**, 10–18 (2009).
- Bebber, D. P., Holmes, T. & Gurr, S. J. The global spread of crop pests and pathogens. *Glob. Ecol. Biogeogr.* **23**, 1398–1407 (2014).
- Diagne, C. *et al.* High and rising economic costs of biological invasions worldwide. *Nature* **592**, 571–576 (2021).
- McNeely, J. A. As the world gets smaller, the chances of invasion grow. *Euphytica* **148**, 5–15 (2006).
- Seebens, H. *et al.* Global rise in emerging alien species results from increased accessibility of new source pools. *Proc. Natl. Acad. Sci. USA* **115**, E2264–E2273 (2018).
- Roques, A. *et al.* Temporal and interspecific variation in rates of spread for insect species invading Europe during the last 200 years. *Biol. Invasions* **18**, 907–920 (2016).
- de Poorter, M. & Browne, M. The Global Invasive Species Database (GISD) and international information exchange: Using global expertise to help in the fight against invasive alien species. *Plant Prot. Plant Health Eur.* **9–11**, 49–54 (2005).
- Tay, W. T. & Gordon, K. H. J. Going global: Genomic insights into insect invasions. *Curr. Opin. Insect Sci.* **31**, 123–130 (2019).
- Paini, D. R. *et al.* Global threat to agriculture from invasive species. *Proc. Natl. Acad. Sci. USA* **113**, 7575–7579 (2016).
- Prentis, P. J., Wilson, J. R. U., Dormontt, E. E., Richardson, D. M. & Lowe, A. J. Adaptive evolution in invasive species. *Trends Plant Sci.* **13**, 288–294 (2008).
- Bertelsmeier, C. Globalization and the anthropogenic spread of invasive social insects. *Curr. Opin. Insect Sci.* **46**, 16–23 (2021).
- Crawley, M. J. *et al.* The population biology of invaders. *Philos. Trans. R. Soc. Lond. B* **314**, 711–731 (1986).
- Petren, K. & Case, T. J. An experimental demonstration of exploitation competition in an ongoing invasion. *Ecology* **77**, 118–132 (1996).
- Kowarik, I. Time lags in biological invasions with regard to the success and failure of alien species. *Plant Invasions Gen. Asp. Spec. Probl.* **1**, 15–38 (1995).
- Andrews, K. L. The whorlworm, *Spodoptera frugiperda*. *Cent. Am. Neighb. Areas Fla. Entomol.* **63**, 456–467 (1980).
- Sparks, A. N. A review of the biology of the fall armyworm. *Fla. Entomol.* **1**, 82–87 (1979).
- 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).
- 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).
- Blanco, C. A. *et al.* Susceptibility of isofamilies of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) to Cry1Ac and Cry1Fa proteins of *Bacillus thuringiensis*. *Southw. Entomol.* **35**, 409–416 (2010).
- 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).
- Chandrasena, D. I. *et al.* Characterization of field-evolved resistance to *Bacillus thuringiensis*-derived Cry1F  $\delta$ -endotoxin in *Spodoptera frugiperda* populations from Argentina. *Pest Manag. Sci.* **74**, 746–754 (2018).
- Pashley, D. P. Host-associated genetic differentiation in fall armyworm (Lepidoptera: Noctuidae): A sibling species complex?. *Ann. Entomol. Soc. Am.* **79**, 898–904 (1986).
- 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).
- Dumas, P. *et al.* *Spodoptera frugiperda* (Lepidoptera: Noctuidae) host-plant variants: Two host strains or two distinct species?. *Genetica* **143**, 305–316 (2015).
- Lu, Y. J., Kochert, G. D., Isenhour, D. J. & Adang, M. J. Molecular characterization of a strain-specific repeated DNA sequence in the fall armyworm *Spodoptera frugiperda* (Lepidoptera: Noctuidae). *Insect Mol. Biol.* **3**, 123–130 (1994).
- Pashley, D. P. Host-associated differentiation in armyworms (Lepidoptera: Noctuidae): An allozymic and mtDNA perspective. in *Electrophoretic Studies on Agricultural Pests*, vol. 39, 103–114 (Clarendon Press, 1989).
- Nagoshi, R. N. The fall armyworm *Triosephosphate Isomerase* (Tpi) gene as a marker of strain identity and interstrain mating. *Ann. Entomol. Soc. Am.* **103**, 283–292 (2010).

29. Dumas, P. *et al.* Phylogenetic molecular species delimitations unravel potential new species in the pest genus *Spodoptera* Guenée, 1852 (Lepidoptera, Noctuidae). *PLoS ONE* **10**, e0122407 (2015).
30. 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).
31. Day, R. *et al.* Fall armyworm: impacts and implications for Africa. *Outlooks Pest Manag.* **28**, 196–201 (2017).
32. Nuss, E. T. & Tanumihardjo, S. A. Maize: A paramount staple crop in the context of global nutrition. *Compr. Rev. Food Sci. Food Saf.* **9**, 417–436 (2010).
33. Zhang, L. *et al.* Genetic structure and insecticide resistance characteristics of fall armyworm populations invading China. *Mol. Ecol. Resour.* **20**, 1682–1696 (2020).
34. Zhang, D. *et al.* Insecticide resistance monitoring for the invasive populations of fall armyworm, *Spodoptera frugiperda* in China. *J. Integr. Agric.* **20**, 783–791 (2021).
35. Gui, F. *et al.* Genomic and transcriptomic analysis unveils population evolution and development of pesticide resistance in fall armyworm *Spodoptera frugiperda*. *Protein Cell* **1**, 1–19 (2020).
36. Schlum, K. A. *et al.* Whole genome comparisons reveal panmixia among fall armyworm (*Spodoptera frugiperda*) from diverse locations. *BMC Genomics* **22**, 179 (2021).
37. Stokstad, E. New crop pest takes Africa at lightning speed. *Science* **356**, 473–474 (2017).
38. Nagoshi, R. N. *et al.* The fall armyworm strain associated with most rice, millet, and pasture infestations in the Western Hemisphere is rare or absent in Ghana and Togo. *PLoS ONE* **16**, e0253528 (2021).
39. Caniço, A., Mexia, A. & Santos, L. Farmers' knowledge, perception and management practices of fall armyworm (*Spodoptera frugiperda* Smith) in Manica province, Mozambique. *NeoBiota* **68**, 127 (2021).
40. 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**, e0181982 (2017).
41. 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).
42. Create your own Custom Map. *MapChart* <https://mapchart.net/index.html>.
43. Sharanabasappa, S. *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).
44. Liu, H. *et al.* Chromosome level draft genomes of the fall armyworm, *Spodoptera frugiperda* (Lepidoptera: Noctuidae), an alien invasive pest in China. *BioRxiv* <https://doi.org/10.1101/671560> (2019).
45. Gimenez, S. *et al.* Adaptation by copy number variation increases insecticide resistance in the fall armyworm. *Commun. Biol.* **3**, 664 (2020).
46. Gouin, A. *et al.* Two genomes of highly polyphagous lepidopteran pests (*Spodoptera frugiperda*, Noctuidae) with different host-plant ranges. *Sci. Rep.* **7**, 11816 (2017).
47. Nam, K. *et al.* Positive selection alone is sufficient for whole genome differentiation at the early stage of speciation process in the fall armyworm. *BMC Evol. Biol.* **20**, 152 (2020).
48. Fiteni, E. *et al.* Host-plant adaptation as a driver of incipient speciation in the fall armyworm (*Spodoptera frugiperda*). *BMC Ecol. Evol.* **22**, 133 (2022).
49. Tay, W. T. *et al.* Global population genomic signature of *Spodoptera frugiperda* (fall armyworm) supports complex introduction events across the Old World. *Commun. Biol.* **5**, 1–15 (2022).
50. Guan, F. *et al.* Whole-genome sequencing to detect mutations associated with resistance to insecticides and Bt proteins in *Spodoptera frugiperda*. *Insect Sci.* <https://doi.org/10.1111/1744-7917.12838> (2020).
51. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
52. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).
53. Cruickshank, T. E. & Hahn, M. W. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol. Ecol.* **23**, 3133–3157 (2014).
54. Pavlidis, P., Živković, D., Stamatakis, A. & Alachiotis, N. SweepD: likelihood-based detection of selective sweeps in thousands of genomes. *Mol. Biol. Evol.* **30**, 2224–2234 (2013).
55. Charlesworth, B., Morgan, M. T. & Charlesworth, D. The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289–1303 (1993).
56. Aikio, S., Duncan, R. P. & Hulme, P. E. Lag-phases in alien plant invasions: separating the facts from the artefacts. *Oikos* **119**, 370–378 (2010).
57. Morimoto, N., Kiritani, K., Yamamura, K. & Yamanaka, T. Finding indications of lag time, saturation and trading inflow in the emergence record of exotic agricultural insect pests in Japan. *Appl. Entomol. Zool.* **54**, 437–450 (2019).
58. Aagaard, K. & Lockwood, J. Exotic birds show lags in population growth. *Divers. Distrib.* **20**, 547–554 (2014).
59. Azzurro, E., Maynou, F., Belmaker, J., Golani, D. & Crooks, J. A. Lag times in Lessepsian fish invasion. *Biol. Invasions* **18**, 2761–2772 (2016).
60. McDonnell, A. M. & Dang, C. H. Basic review of the cytochrome P450 system. *J. Adv. Pract. Oncol.* **4**, 263–268 (2013).
61. 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).
62. Cao, W. *et al.* Multi-faceted epigenetic dysregulation of gene expression promotes esophageal squamous cell carcinoma. *Nat. Commun.* **11**, 3675 (2020).
63. Yainna, S. *et al.* Geographic monitoring of insecticide resistance mutations in native and invasive populations of the Fall Armyworm. *Insects* **12**, 468 (2021).
64. Tapadia, M. G. & Lakhota, 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).
65. 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).
66. de Fouchier, A. *et al.* Functional evolution of Lepidoptera olfactory receptors revealed by deorphanization of a moth repertoire. *Nat. Commun.* **8**, 15709 (2017).
67. Tataroglu, O. & Emery, P. The molecular ticks of the *Drosophila* circadian clock. *Curr. Opin. Insect Sci.* **7**, 51–57 (2015).
68. Hänniger, S. *et al.* Genetic basis of allochronic differentiation in the fall armyworm. *BMC Evol. Biol.* **17**, 68 (2017).
69. Schöfl, G., Heckel, D. G. & Groot, A. T. Time-shifted reproductive behaviours among fall armyworm (Noctuidae: *Spodoptera frugiperda*) host strains: Evidence for differing modes of inheritance. *J. Evol. Biol.* **22**, 1447–1459 (2009).
70. 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).
71. Feder, J. L. *et al.* Genome-wide congealing and rapid transitions across the speciation continuum during speciation with gene flow. *J. Hered.* **105**, 810–820 (2014).

72. Schubert, M., Lindgreen, S. & Orlando, L. AdapterRemoval v2: Rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **9**, 88 (2016).
73. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
74. McKenna, A. *et al.* The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
75. Lu, Y. & Adang, M. J. Distinguishing fall armyworm (Lepidoptera: Noctuidae) strains using a diagnostic mitochondrial DNA marker. *Fla. Entomol.* **1**, 48–55 (1996).
76. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
77. 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).
78. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
79. Tamura, K. *et al.* MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739 (2011).
80. Li, H. Tabix: fast retrieval of sequence features from generic TAB-delimited files. *Bioinformatics* **27**, 718–719 (2011).
81. Renteria, M. E., Cortes, A. & Medland, S. E. Using PLINK for genome-wide association studies (GWAS) and data analysis. *Methods Mol. Biol.* **1019**, 193–213 (2013).
82. Weir, B. S. & Cockerham, C. C. Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370 (1984).
83. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
84. Ernst, M. D. Permutation methods: a basis for exact inference. *Stat. Sci.* **4**, 676–685 (2004).
85. Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**, e1002967 (2012).
86. Kergoat, G. J. *et al.* A novel reference dated phylogeny for the genus *Spodoptera* Guenée (Lepidoptera: Noctuidae: Noctuidae): New insights into the evolution of a pest-rich genus. *Mol. Phylogenet. Evol.* **161**, 107161 (2021).
87. 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).
88. Plotree, D. & Plotgram, D. PHYLIP-phylogeny inference package (version 3.2). *Cladistics* **5**, 163–166 (1989).
89. Nelson, D. R. The cytochrome p450 homepage. *Hum. Genomics* **4**, 1–7 (2009).

## Acknowledgements

This work (ID 1702-018, given to KN) was publicly funded through ANR (the French National Research Agency) under the "Investissements d'avenir" programme with the reference ANR-10-LABX-001-01 Labex Agro and coordinated by Agropolis Fondation under the frame of I-SITE MUSE (ANR-16-IDEX-0006). In addition, a grant from the department of Santé des Plantes et Environnement at Institut national de la recherche agronomique for KN (adaptivesv). This work was also financially supported by EUPHRESKO (FAW-spedcom, given to Anne-Nathalie Volkoff) and by CSIRO Health & biosecurity (given to WTT, TW, and KG). SY was supported by a CIRAD-INRAE PhD fellowship. We are grateful to the genotoul bioinformatics platform Toulouse Occitanie (Bioinfo Genotoul, <https://doi.org/10.15454/1.5572369328961167E12>) for providing help and/or computing and/or storage resources.

## Author contributions

W.T.T., M.F., S.S.D., R.A., C.M.K., R.L.M.J., C.A.B., P.S., T.B., A.D., T.W., K.G., and N.N. provided samples for whole genome resequencing. E.F., A.N.C., S.G., G.J.K., and E.A. prepared samples. E.F. and F.L. performed variant calling. S.Y., K.D., and K.N. performed analysis. F.H. performed the CYP gene annotation. S.Y. and K.N. wrote the manuscript with input from all co-authors.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-25529-z>.

**Correspondence** and requests for materials should be addressed to K.N.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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



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

© The Author(s) 2022