

<https://doi.org/10.1038/s42003-022-03230-1>

OPEN

Global population genomic signature of *Spodoptera frugiperda* (fall armyworm) supports complex introduction events across the Old World

Wee Tek Tay¹✉, Rahul V. Rane¹, Amanda Padovan¹, Tom K. Walsh¹, Samia Elfekih², Sharon Downes³, Kiwong Nam⁴, Emmanuelle d'Alençon⁴, Jianpeng Zhang⁵, Yidong Wu⁵, Nicolas Nègre⁴, Daniele Kunz⁶, Darren J. Kriticos¹, Cecilia Czepak⁷, Michael H. Otim⁸ & Karl H. J. Gordon¹

Native to the Americas, the invasive *Spodoptera frugiperda* (fall armyworm; FAW) was reported in West Africa in 2016, followed by its chronological detection across the Old World and the hypothesis of an eastward Asia expansion. We explored population genomic signatures of American and Old World FAW and identified 12 maternal mitochondrial DNA genome lineages across the invasive range. 870 high-quality nuclear single nucleotide polymorphic DNA markers identified five distinct New World population clusters, broadly reflecting FAW native geographical ranges and the absence of host-plant preferences. We identified unique admixed Old World populations, and admixed and non-admixed Asian FAW individuals, all of which suggested multiple introductions underpinning the pest's global spread. Directional gene flow from the East into eastern Africa was also detected, in contrast to the west-to-east spread hypothesis. Our study demonstrated the potential of population genomic approaches via international partnership to address global emerging pest threats and biosecurity challenges.

¹ CSIRO Black Mountain Laboratories, Clunies Ross Street, Canberra ACT 2602, Australia. ² CSIRO Australian Centre for Disease Preparedness, Geelong, VIC, Australia. ³ CSIRO FD McMaster Laboratories, New England Highway, Armidale NSW2350, Australia. ⁴ DGIMI, Université Montpellier, INRAE, Montpellier, France. ⁵ College of Plant Protection Nanjing Agricultural University, Nanjing, China. ⁶ Gordon Institute, University of Cambridge, Cambridge CB2 1QN, UK. ⁷ Universidade Federal de Goiáss, Escola de Agronomia, Goiânia, GO, Brazil. ⁸ National Crops Resources Research Institute, Namulonge, Kampala, Uganda. ✉email: weetek.tay@csiro.au

Long an important pest of agriculture in its native New World range, the fall armyworm (FAW) *Spodoptera frugiperda* was first reported in West Africa (Nigeria, São Tomé and Príncipe) in early 2016¹, followed by confirmation across central (Congo²; Togo³, Southern⁴ and Eastern⁵) sub-Saharan Africa between 2017/2018⁶; the Middle East (Yemen⁷), India^{8,9} and surrounding nations, Myanmar¹⁰ and Thailand¹¹ (August and December 2018), followed by Southern China (Yunnan Province) in early January 2019^{12–14}. Detections of FAW since January 2019 have gathered speed: south-ward to Malaysia (March 2019) and Indonesia (Sumatra, April 2019; Java, July 2019; Kalimantan July 2019); Hong Kong (April 2019), Taiwan (May/June 2019); Laos and Vietnam (April 2019¹⁵), the Philippines (June 2019^{16,17}), South Korea (June 2019), and Japan (June 2019)¹⁸. Within China, the FAW has been reported in a northward expansion pattern from Yunnan to 18 provinces by July 2019^{19–21}. As of September 2021, over 70 African and Asian nations have reported FAW²². In January 2020, FAW was trapped in Australia's special biosecurity zone in the Torres Strait islands of Saibai and Erub, and confirmed on 3 February 2020, and on mainland Australia in Bamaga on 18 February 2020^{23,24}.

This chronologically ordered eastward spread of detections led to a widely adopted assumption²⁵ that the FAW was actually spreading west-to-east across and then from Africa. Based on the detection timeline, predictive simulations that assumed human-assisted spread, in particular, eggs on aircraft carrier surfaces²⁶ and from agricultural trade (e.g., associated with various fresh agricultural commodities including asparagus, capsicum, *Solanum melongena*, *S. macrocarpon*; with cut flowers, and fresh cuttings²⁷) associated with egg and larval stages, have modelled this very vagile pest's movement from the east coast of America/the Greater Antilles to West Africa³ between Central and Southern America and Africa, and between Africa and Asia²⁸). Movements of soil from countries known to have FAW into the EU are prohibited and were put in place to limit accidental introductions of pupae²⁷. The human-assisted spread model²⁸ was also used to warn China and South East Asian nations of imminent impact by FAW following confirmation of the pest in India²⁹. This model further forms the basis of international research efforts to track the movement, including using molecular tools to examine invasion biology (e.g.,^{3,30,31}), and simulations to model long-distance dispersal (e.g.,^{28,32,33}). Indeed a meteorological data-based simulation study concluded the Yunnan FAW populations originated from Myanmar, consistent with FAW being officially reported earlier in Myanmar^{34,35} than in China¹⁹. Other work has examined the impact and implications for global plant health and agricultural biosecurity (e.g.,^{12,36}), integrated pest management (IPM) and bioeconomics^{37–39}, and insecticide resistance^{21,31,40}.

Genetic studies on the spread of FAW have focussed on single genes on the mitochondrial genome, occasionally supported by a single partial nuclear gene marker. These markers have been widely used because, throughout much of the native range, FAW populations consist of two morphologically identical host races, the rice-preferred and corn-preferred *S. frugiperda* ('Sfr' and 'Sfc', respectively), that have also been considered as potential sister species^{5,41,42} that exhibited low but significant genomic variation (~1.9%)³⁰. These two host races are supported by phylogenetic analyses based on nuclear and mitochondrial DNA genomes³⁰, and partial mitochondrial DNA genes (e.g.,^{1,5,9,13,41}). The distribution of these Sfr and Sfc populations in their New World range has only recently been investigated based on partial mitochondrial and nuclear genes⁴³, while at the whole genome level they are less well-understood. Genotypes from both host races/sister species are present in the invasive populations (e.g.,^{3,44–46}). Since 2010^{47,48} and especially in recent times during the FAW

range expansion^{13,31,45,49}, the partial *Triose Phosphate Isomerase* (*Tpi*) gene on the Z-chromosome has been adopted to clarify the *Sfc/Sfr* host race status. The *Tpi* marker relies on the presence of a single nucleotide polymorphic (SNP) marker at position 183^{48,49} to differentiate between corn- or rice-preferred FAW. Similarly, inconclusive host preferences based on the mtCOI gene marker also detected *Sfc* and *Sfr* on corn-host plants (e.g.,⁵). Contrary to the introduction patterns of the noctuid *H. armigera* in South America⁵⁰ which showed high genetic diversity^{51–53} similar to that reported for global *H. armigera* populations^{54–56}, the current global partial mtCOI signatures of both *Sfc* and *Sfr* have each been consistent with a single 'bridgehead effect'⁵⁷ introduction, which, when considered together with the *Tpi* locus, was suggested to likely have a Florida/Greater Antilles source population³.

What is missing from current research into the spread of FAW is an analysis of broader genomic evidence. Genome-wide SNP markers aligned to well-annotated genomes can provide powerful genomic evidence for understanding introduction pathways⁵⁸ and eliminate candidate populations⁵⁹ as well as to elucidate hybrid signatures⁶⁰. Furthermore, under the current assumption (i.e., a west-to-east spread involving a single western African invasive bridgehead population), subsequent invasive populations would share the founding population's genomic signatures, since as the more evolutionarily parsimonious scenario, only a single genetic change to match between the environment and the bridgehead invasive individuals would be needed⁵⁷.

In this study, we provide an assessment of global FAW movement history based on genomic data that incorporates populations from Northern, Central, and Southern Americas, and the Caribbean (i.e., representing the original population range), Western and Eastern Africa, and Western and Eastern Asia, representing the pest's Old World expansion. Here we reveal a multi-locus invasion that is likely independent of the reported detection patterns and their timelines, and provide genomic-based evidence to support multiple introductions of the FAW into the Old World, with movements of FAW detected between Asia and Africa. We also re-evaluated the pest's global spread directionality to highlight implications in the future management of FAW, and the need for ongoing global agricultural biosecurity research and cooperation to improve preparedness for emerging invasive agricultural pest issues.

Results

Across the native and invasive ranges, FAW individuals have been classified into rice- or corn-preferred strains, either based on the partial mtCOI gene, or through the *TPI* partial gene from the z-chromosome. Due to the non-recombinant and maternally inherited mode of the mitochondrial DNA genome (cf. biparental and recombinant nuclear genome), it is also possible to infer the minimum number of unique female founders responsible for establishing the invasive populations, as well as their likely native population origin/s through matching between mitochondrial genomes. We first examined the mitochondrial genome diversity in the invasive range to determine minimum maternal lineages, and an overview of the strain composition in both native and invasive populations. Of the 197 FAW individuals sequenced (Supplementary Data 1), 102 were from the native New World range and 95 from the invasive Old World range (Fig. 1). From the pest's native range, we detected 25 'rice' mitochondrial genome (i.e., mitogenome) haplotypes, and 51 'corn' mitogenome haplotypes. All FAW from Mexico and Peru had the 'corn' mitogenome while FAW from Guadeloupe and French Guiana were all 'rice' mitogenomes. Of the FAW from the invasive range nine 'corn' and 'rice' mitogenome haplotypes were identified; one

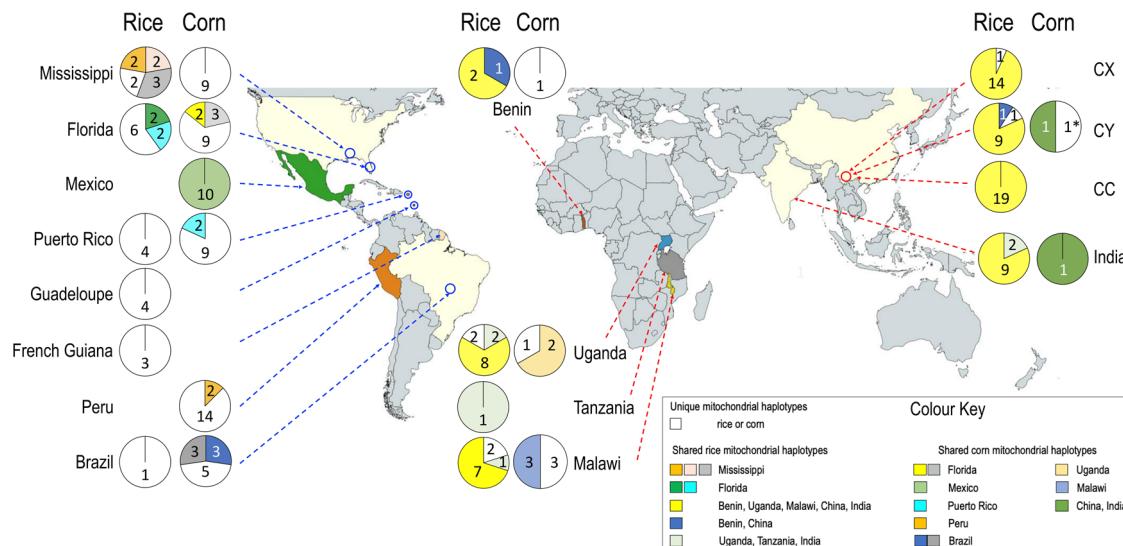


Fig. 1 Native New World and invasive Old World FAW populations, and proportions of rice-strain and corn-strain mitochondrial DNA haplotypes from 15,059 bp of the mitochondrial DNA genomes. New and Old Worlds' FAW populations and proportions of mitochondrial DNA haplotypes based on 15,059 bp of the mitochondrial DNA genomes and excluding four regions of low complexity. For the New World 'rice' FAW, a total of 20 unique mitogenome haplotypes (represented by the white colour proportion of each pie chart), and 11 non-unique mitogenome haplotypes were detected (i.e., a total of 25 mitochondrial haplotypes in rice FAW in the New World). For the 'corn' mitogenomes, 46 unique haplotypes were detected from the native range, while 25 corn FAW individuals shared a total of seven haplotypes (i.e., a total of $46 + 7 = 53$ mitochondrial haplotypes). In the invasive range, six unique 'rice' mitogenomes (i.e., white portion of the pie charts, representing two individuals from Uganda, two individuals from Malawi, and two individuals from China (CY, $n=1$; CX, $n=1$) and three shared mitogenomes (i.e., dark blue, yellow, pale green) were detected from 76 individuals from Africa ($n=22$), India ($n=11$) and China ($n=43$). For the 'corn' FAW from the invasive range, six unique mitogenome haplotypes (i.e., white portions of pie charts) and three non-unique mitogenome haplotypes (pale orange, pale blue and dark green) were detected, although only one individual each from China and India shared a common mitogenome (represented by dark green). With the exception of white colour representing unique mitogenomes, colour schemes are otherwise independent between 'corn' and 'rice' mitogenome haplotypes. China FAW populations from Yunnan Province of Cangyuan (CC), Yuanjiang (CY), and Xinping (CX) are indicated. One pre-border FAW intercepted in December 2016 from cut flowers that originated from Yunnan China (CH06) with a unique corn mitogenome is indicated with ** (placed together with the CY corn pie-chart). Numbers within pie-charts indicate individuals for each mitogenome haplotype.

of the 'corn' mitogenome haplotypes (represented by green colour, Fig. 1) was shared between CY and Indian individuals. No African corn mitogenome haplotypes were shared with Asian FAW populations. In contrast, 83% (i.e., 68/82) of African and Asian FAW with 'rice' mitogenomes shared a common haplotype (represented by the yellow colour, Fig. 1). FAW individuals from China and Benin shared a rare rice mitogenome haplotype (blue colour haplotype), and individuals from Uganda, Tanzania, Malawi and India shared one (i.e., light green colour) haplotype (Fig. 1). In general, the high diversity of haplotypes in both 'rice' and 'corn' in the native range and the lack of diversity in the invasive range is consistent with patterns observed in invasive populations that have a relatively small number of founders.

Mitochondrial DNA genome phylogeny. The trimmed (15,059 bp) mitochondrial DNA genomes of all individuals identified two distinct clades that corresponded to the 'rice-preferred' and 'corn-preferred' clusters (Fig. 2). Based on the near-complete mitogenome phylogeny, a minimum of four and five introduction events were associated with the 'rice' and 'corn' maternal lineages, respectively (Fig. 2). Except for the 'corn' specimen (CH06) from Yunnan that clustered strongly with an individual from Mississippi (UM04) within a clade (node support: 80%) that included also North and South Americas individuals, all 'corn' individuals from the invasive range (i.e., MW26, BE30, MW01, MW06, IN12, MW16, UG03, UG06) clustered weakly with individuals from Florida. Similarly, apart from the Benin individual (i.e., BE01), all remaining 'rice' FAW from the invasive range also clustered weakly with individuals from

Florida. Therefore, the likely origins of the Old World invasive 'corn' and 'rice' FAW remained inconclusive, while divergent mitochondrial genomes nevertheless supported multiple introductions to underpin the current invasive Old World FAW populations.

Nuclear SNP phylogeny. The ML phylogeny based on 870 unlinked and neutral SNPs revealed four distinct clades (clades I, II, III, IV; Fig. 3) across the sampled populations. Native and invasive individuals were a component of each clade which enabled a side-by-side comparison of population structure. Members within each clade were grouped with high (90–96%) bootstrap branch node support values. Clade I included the majority of the invasive FAW individuals from China (CX, CY, CC), India (IN), Uganda (UG), and Benin (BE) as well as individuals from Brazil. Overall, subclades within Clade I indicated unique genomic signatures between the CC and CY/CX populations. Indian and African populations (i.e., Uganda, Benin) were scattered among the CC and CY/CX populations. This interspersed clustering of subclades from Chinese, African and Indian populations suggest a complex FAW spread across the Old World, with some of the China CY individuals potentially sharing a New World origin similar to the Brazil rCC (i.e., 'BR' code, Fig. 3 Clade I) individuals.

Clade II, which is phylogenetically most closely related to Clade I, is dominated by individuals from Mississippi. Within this clade, individuals from China (i.e., CX), Uganda, Benin and India are also present, indicative of likely separate introductions of FAW from the population(s) with genetic similarity to the Mississippi

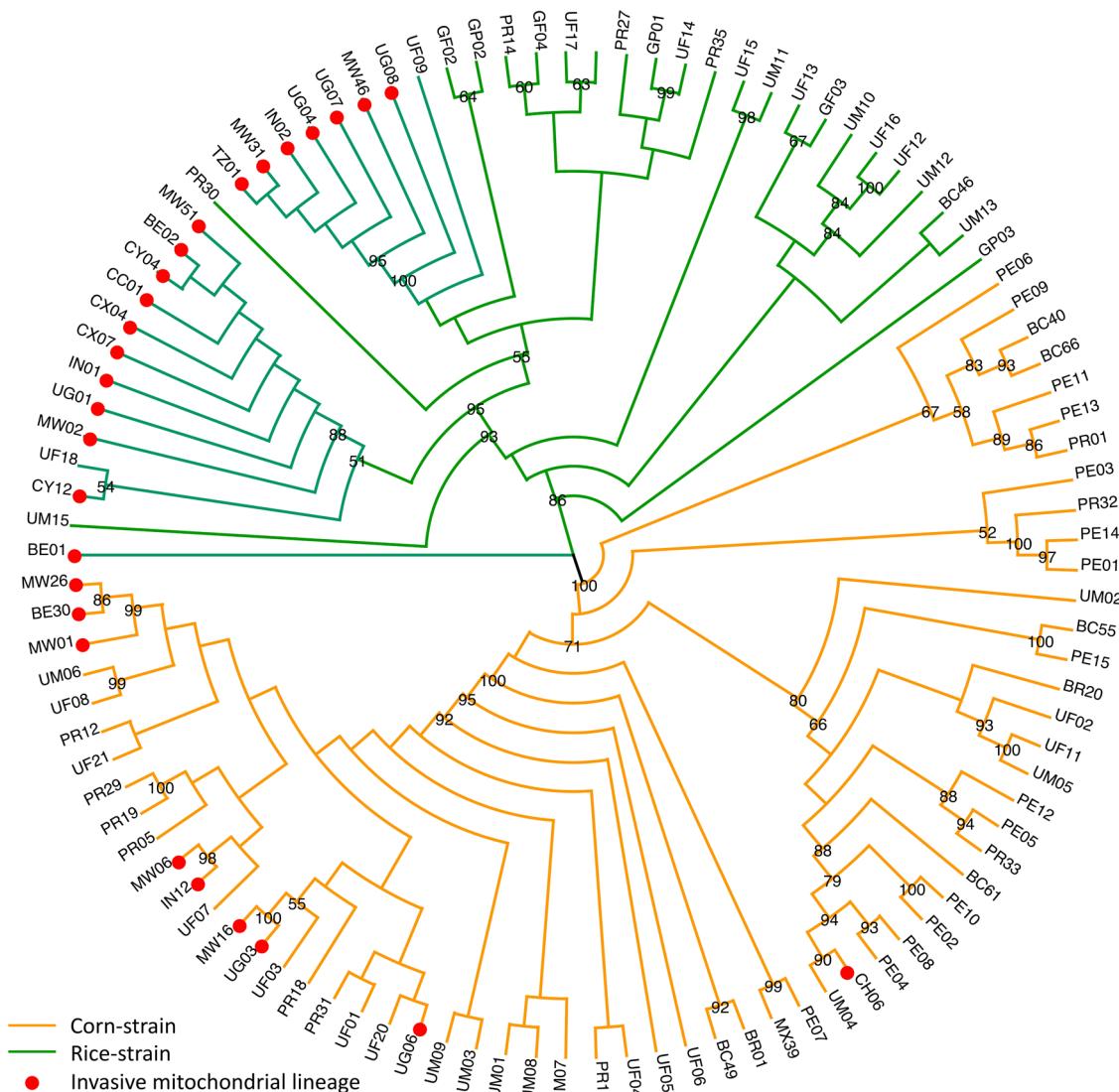


Fig. 2 Partial (15,059 bp) mitochondrial genome Maximum Likelihood phylogeny, showing a clear dichotomy between the rice-strain (green branches) and corn-strain (orange branches) *Spodoptera frugiperda*. Invasive haplotypes from multiple maternal lineages are indicated by red dots. FAW maximum likelihood phylogeny was constructed using IQ-Tree based on 15,059 bp partial mitochondrial genome with edge-linked partition for the 13 protein-coding genes and excluding four regions of low complexity. Node support is estimated from 1000 bootstrap replications, node support values are shown for $\geq 50\%$. ‘Rice’ clade is indicated by branches in green (native range) and ‘Corn’ clade is indicated by branches in orange (native). Unique haplotypes from all populations are included. Country codes are UF (USA-Florida), UM (USA-Mississippi), PR (Puerto Rico), GP (Guadeloupe), GF (French Guiana), PE (Peru), MX (Mexico), BC (Brazil-CC), BR (Brzil-rCC), BE (Benin), UG (Uganda), TZ (Tanzania), MW (Malawi), IN (India), and four populations from China Yunnan Province (Australia pre-border interception (CH06); Cangyuan (CC), Yuanjing (CY), and Xinping (CX)). Invasive mitochondrial lineages are indicated by red dots.

population into the Old World. Clade III is represented by a separate Brazilian (i.e., ‘BC’) FAW population and the Peru FAW individuals. Invasive populations clustered within clade III were the Malawi FAW population, a single Tanzania and three Ugandan individuals, suggesting that these African FAW shared a similar origin that is different from other African (e.g., Benin, rest of Uganda) and Asian populations. The Ugandan population, in particular, appears genetically highly heterogeneous, indicating it too has mixed introduction backgrounds.

Clade IV is dominated by the Florida population and other Caribbean islands/Greater Antilles (e.g., Puerto Rico)/Lesser Antilles (e.g., Guadeloupe)/ Central American (e.g., Mexico), and parts of the northern region of South America (e.g., French Guiana) FAW populations. Clade IV contained a single invasive Chinese FAW (i.e., CH06). Taken as a whole, the nuclear SNP

phylogeny provides clear evidence for multiple introductions of FAW to the Old World, while identifying populations associated with the Mississippi and the Brazilian ‘BR’ populations as likely sources of invasive populations into the Old World. The source population for Malawi’s FAW was likely population(s) from South America, currently represented by Peru/Brazil (BC) populations. Based on interception data, with the exception of a single unique FAW, Florida and the Greater Antilles do not appear to be likely sources for the current invasive populations in the Old World.

Our nuclear SNP phylogeny therefore clearly showed that the native range FAW populations could be classified based on their geographic origins. The unexpected direct phylogenetic relationship between the US Mississippi and Brazil ‘BR’ population, suggested potential movements of populations within North

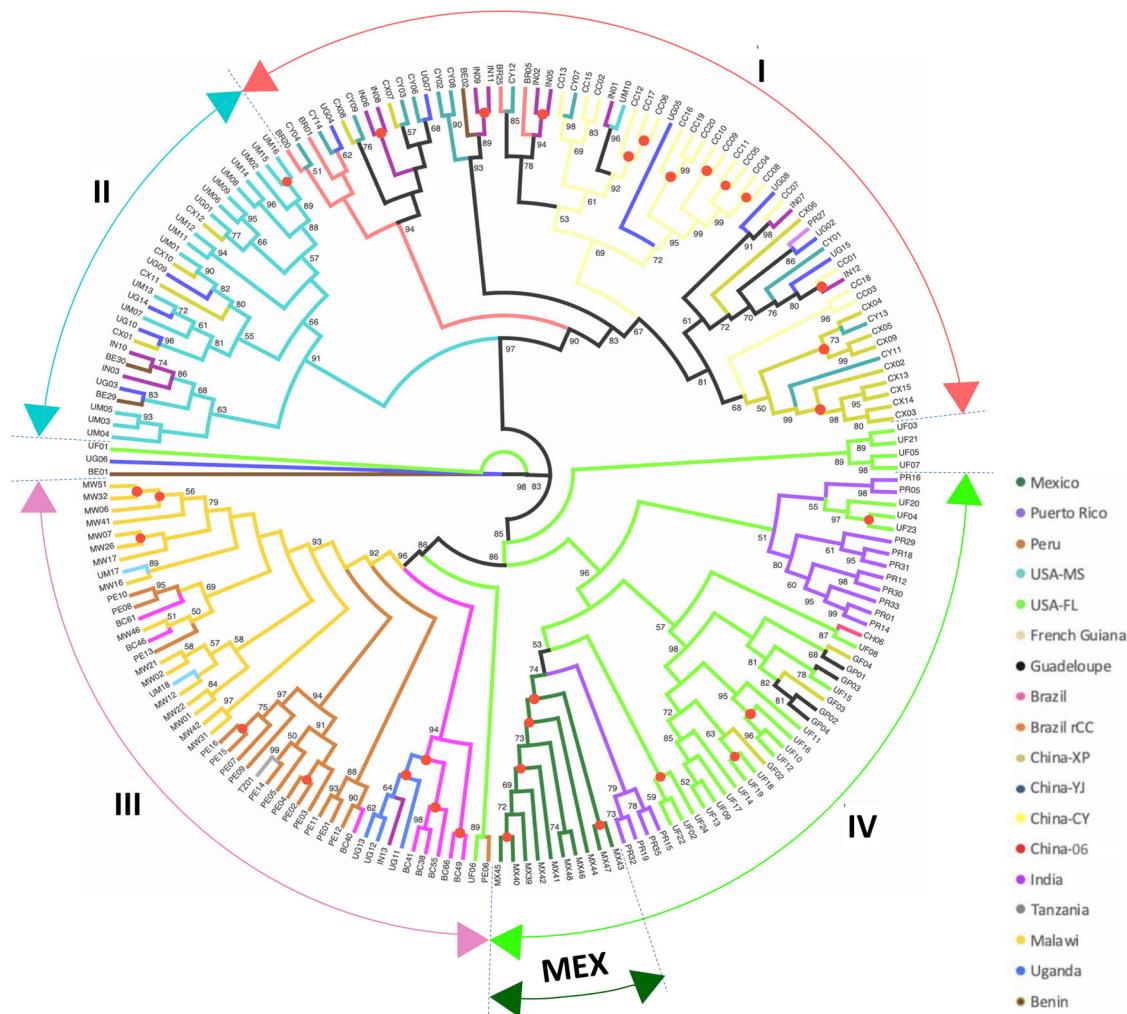


Fig. 3 Maximum likelihood phylogeny of *Spodoptera frugiperda* populations from the native range of Northern America, Caribbean, South America, and the Old World invasive populations from Africa and Asia as inferred from 870 genome-wide SNP loci. IQ-Tree with 1000 bootstraps replications to estimate node support for *Spodoptera frugiperda* populations from Northern America (Mississippi, Florida), Caribbean (Puerto Rico, Guadeloupe, French Guiana), and South America (Peru, Brazil), as well as *S. frugiperda* populations representing the Old World invasive range from Western Africa (Benin), Eastern Africa (Uganda, Tanzania, Malawi), and Asia (India, China). A total of 870 independent SNPs (i.e., unlinked) from non-coding regions distributed across the genome with no missing data were used. Populations are represented by unique colour schemes as indicated. Three populations of *S. frugiperda* from China Yunnan Province are Cangyuan (CC), Yuanjiang (CY), and Xinping (CX), and two populations of *S. frugiperda* from Brazil are Brazil-CC (BC) and Brazil-rCC (BR). Branch nodes with 100% bootstrap support are indicated by red dots. Bootstrap values of <50% are not shown. The legend shows branch colours of sampling countries.

America (i.e., Mississippi is not the wintering ground for FAW and represents the melting-pot for summer migrants from Texas and Florida⁶¹ and between North and South America. Finally, an important overall finding was that our panel of neutral SNPs selected from whole-genome sequencing did not separate individuals based on ‘corn’ or ‘rice’ mitochondrial DNA genome signatures, nor did they support the host strain characterisation based on the *Tpi* partial gene marker.

Genetic diversity. Basic population diversity statistics for each population are listed in Table 1. Nucleotide diversity (π) varied across a narrow range (0.287–0.329), for the 870 variable and independent SNPs analysed, that included no invariant loci. No significant overall difference was observed between the native and invasive range populations. All populations showed higher average observed heterozygosities (H_{obs}) than the average expected heterozygosities (H_{exp}), both in the native and invasive ranges, with the highest H_{obs} seen in the Malawi population. Negative F_{IS} values for all populations were consistent with H_{obs} being higher

than H_{exp} , and suggested systematic avoidance of consanguineous mating⁶² within FAW subpopulations as a whole. The lower expected heterozygosity in all these populations (i.e., $H_{obs} > H_{exp}$; see⁶³) is most likely indicative of the recent mixing of previously distinct populations and does not support that these invasive populations originated from a single introduction (e.g.,^{2,49,64,65}) or had undergone a recent bottleneck from the widely suggested recent western Africa arrival. It is more likely that they represent the result of multiple introductions into the invasive range, as already suggested by mitochondrial and nuclear SNP phylogenies (Figs. 2, 3) and PCA (Fig. 4). The observed heterozygosity excess detected for the native range populations may similarly be due to factors such as structure between these populations and the breaking of isolation through periodic migration among native populations. Consistent with these observations, a number of the populations including most from the invasive range also contained significant numbers of loci not in Hardy-Weinberg equilibrium (HWE). This was especially the case for the two largest Chinese populations (i.e., CY, CX), Malawi and Uganda, as well

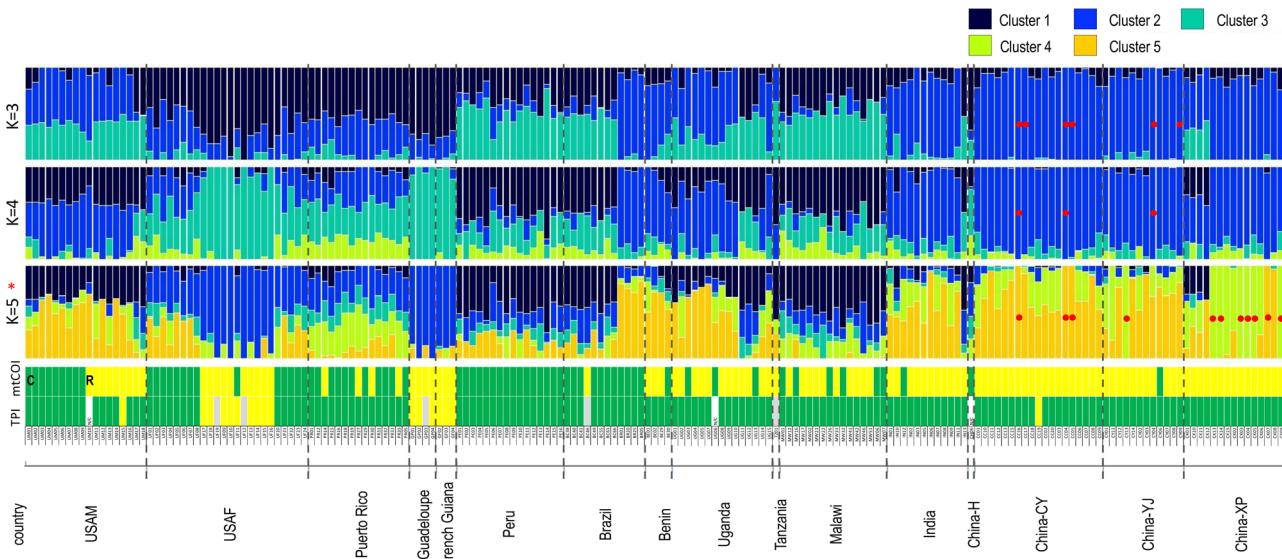


Fig. 5 Admixture analysis of invasive and native *Spodoptera frugiperda* populations based on 870 neutral SNP loci, and individual host strain identity based on the *Tpi* and the *mtCOI* markers. Admixture analysis based on $K = 3$ (Cross-Validation Error (CVE): 0.375), $K = 4$ (CVE: 0.377), and $K = 5$ (CVE: 0.380)). Populations 'USAM' and 'USAf' are from Mississippi and Florida, respectively. Populations from China were from Cangyuan (CY), Yuanjiang (YJ), Xinping (XP) in Yunnan Province. Individuals in the invasive ranges that lacked the signature of admixture are indicated by red dots. Corn- ('C') or rice- ('R') preferred plant hosts are identified based on *mtCOI* as per Dumas et al.⁴¹, and by *Tpi* SNP approach as per Nagoshi⁴⁸ are indicated by green or yellow bars, respectively. Specimen ID's and sampling countries are as labelled. 'N/C' for *Tpi* indicates no coverage. Grey bars indicate individuals with heterozygous *Tpi* SNPs.

and $K = 5$. Interestingly, these individuals with non-admixed genomic signatures (at either $K = 3$, 4 and/or 5) also possessed the rice mitogenome haplotypes (Fig. 5). This observation is similar to that observed for the non-admixed Chinese individuals that have mitogenomes which also exhibited the rice haplotypes. Admixture analysis also revealed most Mexican individuals as having non-admixed genome patterns and with the corn mitogenome haplotypes (Supplementary Fig. 1a, b). As with the SNP phylogeny above, a comparison of the admixture patterns to mitogenomes and the *Tpi* locus of native and invasive FAW populations failed to find evidence to support FAW host-strain characterisation. The genome admixture signatures of FAW across its African and Asian invasive range supported a complex pattern for multiple introductions. For example, given the highly admixed genomic patterns detected in African and Indian individuals, it is unlikely that matings between these admixed populations would lead to individuals with non-admixed genomic signatures in China, unless there was some very strong selection pressure acting across specific genomic regions of these selected CY, CC, and CX individuals.

Admixture networks. To explore the population substructure revealed by the admixture analysis in relation to the ML clusters obtained from phylogeny and PCA, we performed network analysis using the *plotAdmixture* function in the NetView R package. The ML network of individuals belonging to each of the specified populations is shown in Fig. 6. The four major clusters, I–IV, correspond to those shown in the ML tree (Fig. 3). Individuals from some populations were shown to be spread across multiple clades, e.g., PR, UF and UM from the native range and IN, BE and CX from the invasive populations. Of the populations in the invasive range, those from China were found predominantly in cluster I, with some CX individuals in cluster II and the single CH06 individual in cluster IV.

Plotting admixture proportions at $K = 5$ on this network showed the different populations from China that predominantly comprise Cluster I each have distinct admixture profiles that are

shared with those of individuals from Uganda and India. In cluster II, China-XP (CX), India (IN), Benin (BE), and Uganda (UG) formed networks with USA-Mississippi (UM) individuals. In Cluster III, all Malawi (MW) individuals and various Ugandan (UG) individuals and the single Tanzanian (TZ) individual formed a network cluster with Peru (PE), Brazil-CC (BC), and some USA Florida (UF) individuals. In cluster IV, only one Chinese FAW (CH) was found to group to this predominantly Caribbean/Central America FAW group (consisting of UF, Puerto Rico (PR), French Guiana (GF), Guadeloupe (GP), and Mexico (MX) FAW individuals).

Directionality of gene flow analysis using divMigrate. Analysis of the directionality of gene flow (i.e., relative directional migration) between populations using divMigrate enabled investigation of possible introduction pathways leading to the complex population substructure patterns seen in the above analyses. The most significant directional gene flow signatures seen were from all three Chinese populations (i.e., CX, CY, CC) into Malawi and from the Cangyuan (CC) population into Uganda (Fig. 7). Significant gene flow from Florida (UF) and from Puerto Rico (PR) into the Mississippi (UM) FAW population, which the above (e.g., Figs. 3, 4a, 5, 6; Supplementary Fig. 2) had shown to be distinct was also detected. No evidence was found for directional gene flow from any of the populations studied into China, nor any from or into India. Together with the Admixture results (Fig. 5), these results indicate the East African FAW populations likely originated from China, with some independent 'non-China' introductions also detected in Malawi. The Admixture signatures within the Ugandan FAW population suggested the presence of two genetically distinct FAW populations (Figs. 5, 6), one of which originated from Asia and involved genetic contribution from the Yunnan Cangyuan (CC) population (Fig. 7), as well as gene flow from Malawi (Fig. 5). While the Malawi population overall showed admixture patterns similar to Peru (Fig. 5) with the PCA showing the Malawi, Peru and Brazil-CC (BC) populations clustered together (Fig. 4b, c), directionality analysis

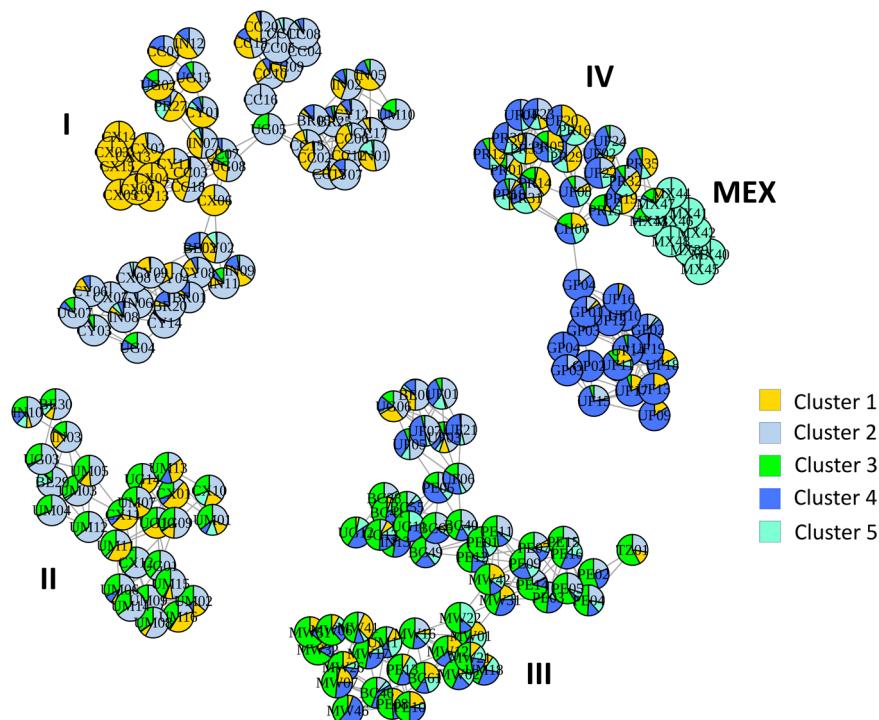


Fig. 6 Maximum Likelihood distance network with admixture analysis of invasive and native *Spodoptera frugiperda* populations. Maximum likelihood (ML) distance network with admixture analysis inferred from five genetic clusters ($K = 5$) presented as pie charts for each individual analysed. The network was drawn using the plotAdmixture function in the R package NetView^{120, 121}, and is based on a ML distance matrix calculated from the IQ-Tree shown in Fig. 3, using the R package ape¹²². The four major clusters, I–IV, correspond to those shown in the phylogenetic tree (Fig. 3). Individuals are identified by country codes as follows: China XP (CX), China YJ (CY), China CY (CC), China CH06 (CH), India (IN), Uganda (UG), Tanzania (TZ), Malawi (MW), Benin (BE), Brazil CC (BC), Brazil rCC (BR), Peru (PE), French Guiana (GF), Mexico (MX), Guadeloupe (GP), Puerto Rico (PR), USA-Florida (UF), and USA-Mississippi (UM). See Supplementary Data 1 for complete information about the individuals. Cluster I comprises predominantly different Chinese populations each with distinct admixture profiles but included also genetic cluster profiles of individuals from Uganda, India, Brazil-rCC (BR) and Puerto Rico. In cluster II, China-XP (CX), India, Benin, and Uganda formed networks with USA-Mississippi individuals. In Cluster III, all Malawi individuals and various Tanzania and Uganda individuals were grouped with Peru, Brazil-CC (BC), and selected USA-FL individuals. In cluster IV, only one Chinese FAW (CH) was found to a group to this predominantly Caribbean/Central America FAW group (consisting of USA-FL, Puerto Rico, French Guiana, Guadeloupe, and Mexico FAW individuals). Note that individuals sharing the same colour schemes do not necessarily have the same genetic content, and that the MEX group consisted only of individuals from Mexico showing little admixture with any other population.

indicated genetic contributions from all three Chinese FAW populations (Fig. 7).

Discussion

The genomic analysis of FAW from native and invasive ranges contradicts recently published theories on the pathway, origin, and direction of the spread of this pest across the Old World. Neutral and unlinked genome-wide SNPs obtained from early stages of the FAW invasion showed, through population admixture analysis, ML distance network, and gene flow directionality analyses, that there were likely multiple introductions to both Africa and Asia. Studies to date have relied on analyses of limited partial mitochondrial DNA (e.g., partial *COI* and *CYTB*^{5,46}); and the nuclear *Tpi* partial gene (e.g.,⁴⁵) of various African, Asian and South East Asian invasive FAW populations, with comparisons to native New World FAW populations. These studies inferred the directionality of spread from the timing of official reporting to the FAO/IPPC, and described a single introduction of FAW to the Old World from an eastern American/Greater Antilles population, that spread rapidly across the sub-Saharan African nations, before moving to the Indian sub-continent via the Middle East, and then to South East Asia, and China⁴⁵.

Under the bridgehead effect of invasion scenario where subsequent successful invasive populations (e.g., FAW populations

from sub-Saharan Africa (south, central, east Africa); Asia (e.g., India, China)) originated via an intermediate and successful population (i.e., western African FAW population¹), one should expect all populations to share the same genomic signatures as the bridgehead (i.e., western African FAW) populations⁵⁷. Genome-wide SNP analyses in this present study, however, showed the populations in China and Africa to be genetically diverse and demonstrates strong evidence for a complex spread pattern across the Old World, including a substantial proportion of east-to-west movement, with populations from Asia as a potential source of invasive FAW populations in Africa (e.g., Malawi, Uganda), although our study lacked other populations (e.g., from Southeast Asia) that may be alternative sources of invasive FAW populations to Africa. The confirmation of FAW after reports by farmers of crop damage, i.e., in Nigeria and São Tomé and Príncipe in early 2016¹ and in northern and eastern Uganda since 2014²², suggested that *S. frugiperda* was present in the African continent earlier, and given the genomic evidence reported here would suggest that the FAW was present in Asia and/or Southeast Asia prior to 2016.

Incidences of FAW attacking stem/leaf parts were reported from a farm producing turf grass for parks in Hanoi, Vietnam, between March and June (Spring/Summer seasons) of 2008, with heavy infestations reported between the months of April and May in 2008^{67–69}, as well as being reported as an agriculture insect

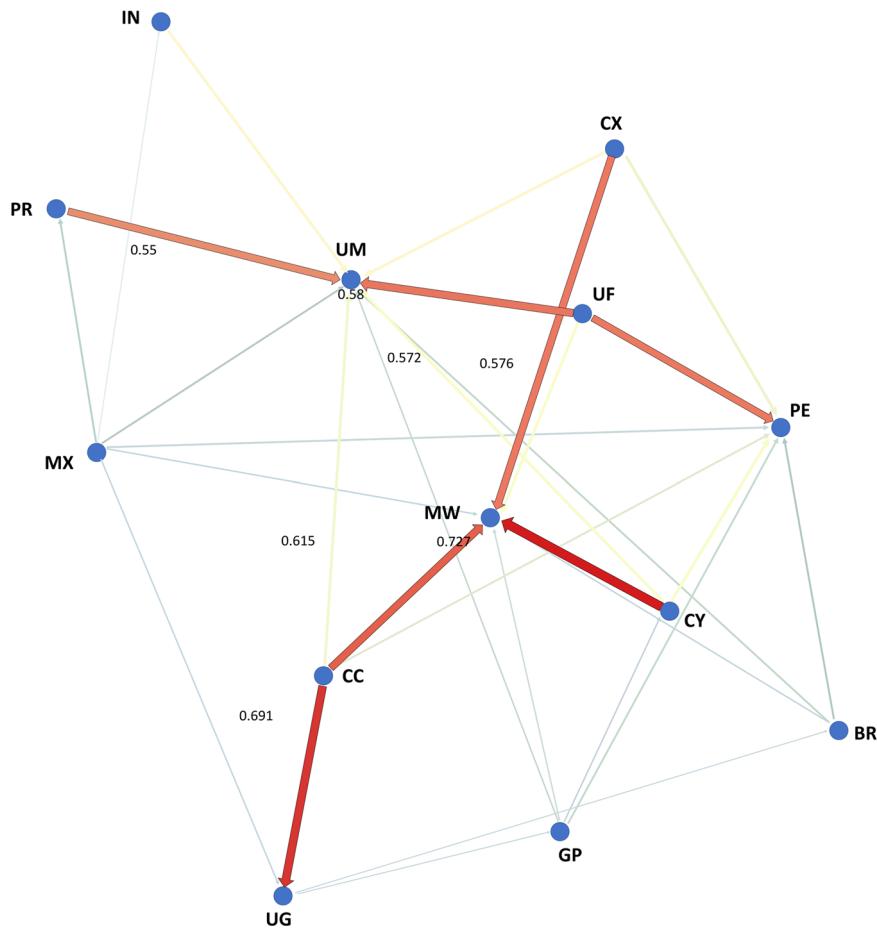


Fig. 7 divMigrate Analysis inferred directionality of gene flow between native and invasive *Spodoptera frugiperda* populations. Analysis using divMigrate to infer directionality of gene flow (i.e., relative directional migration) between New World native and Old World invasive *Spodoptera frugiperda* populations. The divMigrate analysis was run using the online server <<https://popgen.shinyapps.io/divMigrate-online/>>¹²². The analysis was performed with the G_{ST} migration statistic of¹²⁷ and¹²⁸ at filter threshold = 3.0 and 1000 bootstrap replications to assess confidence with alpha value set at 0.05 (i.e., 95% confidence). Weighted values above 0.50 are indicated. Population codes are IN (India), PR (Puerto Rico), MX (Mexico), UG (Uganda), CC (China Cangyuan), CY (China Yuanjiang), CX (China Xiping), MW (Malawi), PE (Peru), GP (Guadeloupe), BR (Brazil-rCC), UM (USA Mississippi), and UF (USA Florida). High migration (i.e., gene flow; sensu Sundqvist et al.¹²⁴) is seen from all three Chinese populations into Malawi and from Cangyuan (CC) to Uganda (UG). High migration from Florida and from Puerto Rico into the Mississippi FAW population is also detected.

pest in the production areas around Hanoi since 2008⁷⁰. We also provided clear evidence for multiple introductions of this agricultural pest into Africa, demonstrating conclusively that the Malawian FAW population has a distinct genomic signature different from Chinese populations. The pre-border interceptions of FAW larvae (identified via morphological characters) that originated from countries outside of the Americas since prior to 2014⁷¹ (although molecular diagnostics of these suspect FAW larvae will be required to provide definitive confirmation of such non-native range interceptions), the early detections and report of FAW in Asia/S.E. Asia (e.g., CH06¹² (GenBank MT897262); 2008 Hanoi outbreaks⁶⁹), and the complex pattern of multiple introductions including potential North American origins for various Chinese FAW populations (e.g., this study for individual CH06; the Yunnan ‘NJ05’ Individual⁷²), are consistent with the perceived rapid spread experienced across the African⁷³ and Asian continents⁷⁴.

Despite being one of the worst agricultural pests in the New World, there has been limited population genomic work on the FAW in their native ranges. Through our genome-wide SNP analyses, we have identified unexpected complexity in the FAW population structure in the New World. While the mitochondrial genome analysis confirmed the two canonical clades that have

long been suggested to define two strains with different host preferences, i.e., corn (*Sfc*) and rice (*Sfr*), the neutral nuclear SNP analyses showed a more complex population and genomic structure. FAW populations in the New World could be differentiated into at least five distinct groups that broadly followed the species’ geographic distributions, and with no obvious pattern related to host race determination by mitochondrial or *Tpi* markers, providing the first genome-wide support for suggestions that these mitochondrial genomes (and the often associated *Tpi* marker) do not define any real population structure across the native range of FAW^{43,75}, while a lack of consistent correlation between host plant and mitochondrial genome in native range populations were observed⁷⁶. Frequent hybridisation has been known to occur in the field (e.g.,⁷⁷), and would also account for the observed pattern. Furthermore, African populations contained hybrids of F2 or even later generations, and mating time differences within the African populations were likely related to the differences in circadian gene expression previously identified in *Sfc* or *Sfr* populations in their native range⁷⁸. Differences in mitochondria function could be directly related to host preferences⁷⁹, which could explain the absence of a correlation between the mitochondrial and nuclear genotypes, but this lack of genomic correlation that was also in part due to the non-

recombination nature for mitochondrial genome *cf.* nuclear genome, points clearly to the need of genome-wide studies in field populations, and that persistence to classify the invasive FAW populations into either 'rice' or 'corn' strains would contribute little in predicting crop damage, and may even hinder the management of these invasive FAW populations.

We detected directional migration from Florida and the Puerto Rican populations to the genetically distinct Mississippi one. This is consistent with findings based on mtCOI sequences that the Mississippi populations were established through seasonal migration from Texas and Florida⁶¹. There also seems to be evidence for a wider Caribbean population including Florida, Puerto Rico, Mexico, the Lesser Antilles (e.g., Guadeloupe) and the north-eastern region of South America (e.g., French Guiana). Mexican FAW formed a separate sub-clade within the Florida/Greater Antilles/Lesser Antilles FAW group. Significant pairwise F_{ST} estimates between Mexico and all native and invasive FAW populations and suggested population genetic differentiation that indirectly indicated limited gene flow. Northern Mexican populations have been shown to be similar to the Southern Texas overwintering population⁶¹, and this is reflected by our finding that the Mexican population sits within the broader Caribbean clade that includes Florida. Across the native range, evidence of population substructure has been reported (e.g., ^{66,80,81}), and a population genomic approach could help identify biological/ecological factors that underpinned patterns of population differentiation, gene flow directionality, and prevalence of admixture to further assist with their management.

Our PCA on genome-wide SNPs identified the Brazilian FAW as two genetically distinct populations, with one population ('BC') being phylogenetically more closely related to the Peruvian FAW population, and the BR population which is phylogenetically more closely related to the Mississippi population. The Brazilian 'BR' population included individuals that had been found to have a novel 12 bp deletion mutation in the *ABCC2* gene⁴⁰. The implications of the close phylogenetic relationship between the BR and Mississippi populations are great given that FAW is regarded as a major agricultural pest in Brazil⁴⁰, and the possible movements of alleles that could potentially underpin resistance, especially to Cry1F and Cry1A toxins, would add to the challenge of managing this pest in the Americas.

Genomic analyses in the present study support multiple introductions of FAW from different sources into Africa, rather than via a single western Africa introduction. Phylogenetic inference and PCA clearly identified the South American FAW population, as represented by the Peru/Brazil (BC) samples, as the likely source for the Malawi population, although this could also represent the 'bridgehead effect'⁵⁷ from other invasive FAW populations not yet included in the current analysis, such as other Asian/East Asian/South East Asian populations. Global movements of invasive pests, exemplified by the spread of FAW, and other agriculturally important pests (e.g., *H. armigera*^{51,82}; the Harlequin ladybird *Harmonia axyridis*⁸³; the whitefly *Bemisia tabaci* species complex⁵⁸; the tomato leaf miner *Tuta abosulta*⁸⁴) are timely reminders of the need for global coordination of enhanced biosecurity preparedness strategies that build on advancement in genomic research. The potential negative impacts of introductions of alien species include the introgression of genetic traits to local species through hybridisation^{12,59,60,85,86}. Development of new trans-continental trade routes to increase economic growth between trading partners must therefore recognise these potential risks and take into consideration the biosecurity implications associated with the rapid spreading of highly invasive pests and pathogens of plants, animals and humans⁸⁷ that could instead undermine the aim to grow the global economy.

Methods

Spodoptera frugiperda populations sampled and analysed in this study were sourced from Florida ($n = 24$)⁷⁹, Mississippi (Stoneville; $n = 18$)³⁰, Puerto Rico (Ayala; $n = 15$)⁸⁸, Peru ($n = 16$), Brazil ($n = 12$; IBAMA Permit number: 18BR028445/DF), Mexico (Texcoco, Estado de Mexico, sampling date: 2009; $n = 10$), Guadeloupe ($n = 4$), French Guiana ($n = 3$), Benin ($n = 4$), India ($n = 12$)⁸⁹; Tanzania ($n = 1$), Uganda ($n = 15$), Malawi ($n = 16$), and three populations from Yunnan Province, China (CC = 19; CY = 12; CX = 15)⁴⁰, and one individual (CH06) from Australia's pre-border interception program overseen by the Department of Agriculture, Water and the Environment (DAWE), also from Yunnan, China (Supplementary Data 1). Sampling of FAW did not require ethics approval as this was an invertebrate/insect pest widely found occurring and attacking agricultural crops. The initial differentiation of these individuals as 'corn-preferred' or 'rice-preferred' was based on the partial mtCOI gene region⁴¹ and a polymorphism within the Triose Phosphate Isomerase (*Tpi*) gene⁴⁸.

The genomes of both *Sfr* and *Sfc* have been sequenced and annotated³⁰, allowing higher resolution analysis of genetic structure, migration patterns and sub-species status based on a high number of genome-wide SNPs to enable identification of the potential New World origins, and the species and admixture status of the invasive *Sfc* and *Sfr* populations. Extraction of total genomic DNA was carried out at the CSIRO Black Mountain Laboratories site in Canberra Australia for the Brazil, Tanzania, Malawi and Uganda populations, as well as the pre-border intercepted FAW sample from Peru and China, using the Qiagen Blood and Tissue DNA extraction kit following instructions as provided, with genomic DNA eluted in 200 μ L EB. Total genomic DNA for the other three Chinese populations were extracted at Nanjing Agricultural University as detailed in Guan et al.⁴⁰. Total genomic DNA from Mississippi, Florida, Puerto Rico, Guadeloupe, Mexico, and French Guiana, and Indian populations was carried out at INRAE DGIMI (Univ. Montpellier, INRAE, France) as reported in Yainna et al.⁸⁹.

Genomic libraries prepared by CSIRO were constructed using an Illumina Nextera Flex DNA Library Prep Kit following manufacturer's instructions and sequenced by Illumina NovaSeq6000 S4 300 sequencing system at the Australian Genome Research Facility (AGRF). Sequencing efforts were shared between three research institutions: 61 samples were prepared at CSIRO (populations from Brazil, Peru, Uganda, Tanzania, and Malawi), 46 samples were prepared by NJAU for populations from China Yunnan Province (CC, CY and CX counties), and 89 samples were prepared by DGIMI, France (populations from Florida, Mississippi, Puerto Rico, Guadeloupe, French Guiana, Mexico, Benin and India). The Peru FAW samples and the single FAW sample CH06 from Yunnan China were intercepted at Australia's pre-border inspections of imported agricultural and horticultural commodities by the Department of Agriculture, Water and the Environment (DAWE) on fresh vegetables and cut flowers, respectively. The FAW CH06 was sequenced using the Illumina MiSeq high throughput sequencing (HTS) platform following the methods of Tay et al.⁹⁰. Sequencing coverage ranged from 2–56x with a mean coverage of 19x.

Mitochondrial genomes assembly and haplotypes characterisation. The mitochondrial DNA genome for all samples were assembled using Geneious 11.1.5 based on strategies previously used for assembly of *Helicoverpa* species as outlined in Walsh et al.⁹¹. Assembled mitogenomes were annotated using MITOS⁹² selecting invertebrate mitochondrial genetic code. All annotated protein-coding genes/coding sequences (PCGs/CDS) were re-annotated visually to identify putative stop codons and to align start codon positions. Four regions of low complexity (corresponding to BC55 nt6065–6092; nt9544–9580; nt12807–12838; nt15047–15276) were trimmed due to alignment difficulties and low genome assembly confidence associated with simple repeat units, resulting in all samples having final mitochondrial DNA genome length of 15,059 bp. We identified unique mitogenomic haplotypes using the DNACollapser in FaBox (1.5) <<https://users-birc.au.dk/~palle/php/fabox/dnacollapser.php>>⁹³ after alignment using MAFFT Alignment v7.450^{94,95} within Geneious 11.1.5 and selecting the Auto option for Algorithm and 200PAM / K = 2 for Scoring matrix, Gap open penalty of 1.53, and offset value of 0.123. GenBank accession numbers for full mitochondrial genomes from all individuals are listed in Supplementary Data 1.

Nuclear SNPs selection. In this study, we used the originally assembled genome³⁰ for our raw data processing. While the nuclear genomes of the two strains were found to be ~1.9% different³⁰, however, invasive populations analysed to-date have consisted predominantly of hybrids^{31,89}. We used the native rice reference genome *Sfr* from Florida (see also^{88,89} for high-quality assemblies of native population genomes for *Sfr*, and⁹⁶ for high-quality genome assemblies of native *Sfc*^{31,97,98}) to map as it was found to be superior to the corn genome in terms of assembly statistics (e.g., Busco score for the corn strain indicated more missing genes than the rice strain; N50 contig size is greater for *Sfr*; see³⁰). Genomic raw data was cleaned and trimmed using Trimmomatic (v0.39)⁹⁹ and aligned to the *S. frugiperda*³⁰ (rice v1) genome using BWA-MEM (v2.1)¹⁰⁰. Variants were predicted using BBMap (v38.81)¹⁰¹ using the following parameters: bgzip = t maxcov = 300 ploidy = 2 multisample = t; followed by indel normalisation using BCFtools (v1.10)¹⁰² to obtain a whole-genome SNP panel. Variants were filtered to remove SNPs with minimum allele frequency of 0.01, any missing data and linkage disequilibrium (LD) pruned with stringent parameters (-indep-pairwise 50 kb

Author contributions

W.T.T., T.K.W., K.N., E.d'A., N.N., S.D., D.J.K., C.C., and K.H.J.G. designed the study. W.T.T., K.H.J.G., T.W., R.V.R., A.P., S.E., D.K., Z.J., and Y.W. analysed the data. W.T.T., T.W., S.D., Y.W., K.N., E.d'A., N.N., C.C., and M.H.O. sourced the samples. W.T.T., T.K.W., K.H.J.G., R.V.R., and D.J.K. drafted the manuscript, all authors contributed on improving the manuscript. W.T.T., R.V.R., K.H.J.G., K.N., T.K.W., N.N., and M.H.O. revised the manuscript.

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/s42003-022-03230-1>.

Correspondence and requests for materials should be addressed to Wee Tek Tay.

Peer review information *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Luciano Matzkin and Caitlin Karniski. Peer reviewer reports are available.

Reprints and permission information is available at <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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022