ARTICLE

Check for updates

Variation among 532 genomes unveils the origin and evolutionary history of a global insect herbivore

Minsheng You [1,2,20✉], Fushi Ke[1,2,20], Shijun You [1,2,3,20], Zhangyan Wu[4,20], Qingfeng Liu [4,20], Weiyi He [1,2,20], Simon W. Baxter[1,2,5], Zhiguang Yuchi [1,6], Liette Vasseur [1,2,7✉], Geoff M. Gurr [1,2,8✉], Christopher M. Ward [9], Hugo Cerda[1,2,10], Guang Yang[1,2], Lu Peng[1,2], Yuanchun Jin[4], Miao Xie[1,2], Lijun Cai[1,2], Carl J. Douglas[1,2,3,21], Murray B. Isman [11], Mark S. Goettel[1,2,12], Qisheng Song [1,2,13], Qinghai Fan[1,2,14], Gefu Wang-Pruski[1,2,15], David C. Lees[16], Zhen Yue [4✉], Jianlin Bai[1,2], Tiansheng Liu[1,2], Lianyun Lin[1,5], Yunkai Zheng[1,2], Zhaohua Zeng[1,17], Sheng Lin[1,2], Yue Wang[1,2], Qian Zhao[1,2], Xiaofeng Xia[1,2], Wenbin Chen[1,2], Lilin Chen[1,2], Mingmin Zou[1,2], Jinying Liao[1,2], Qiang Gao[4], Xiaodong Fang[4], Ye Yin[4], Huanming Yang[4,17,19], Jian Wang[4,18,19], Liwei Han[1,2], Yingjun Lin[1,2], Yanping Lu[1,2] & Mousheng Zhuang[1,2]

The diamondback moth, *Plutella xylostella* is a cosmopolitan pest that has evolved resistance to all classes of insecticide, and costs the world economy an estimated US $4-5 billion annually. We analyse patterns of variation among 532 *P. xylostella* genomes, representing a worldwide sample of 114 populations. We find evidence that suggests South America is the geographical area of origin of this species, challenging earlier hypotheses of an Old-World origin. Our analysis indicates that *Plutella xylostella* has experienced three major expansions across the world, mainly facilitated by European colonization and global trade. We identify genomic signatures of selection in genes related to metabolic and signaling pathways that could be evidence of environmental adaptation. This evolutionary history of *P. xylostella* provides insights into transoceanic movements that have enabled it to become a worldwide pest.

[1] State Key Laboratory of Ecological Pest Control for Fujian and Taiwan Crops, Institute of Applied Ecology, Fujian Agriculture and Forestry University, Fuzhou 350002, China. [2] Joint International Research Laboratory of Ecological Pest Control, Ministry of Education, Fuzhou 350002, China. [3] Department of Botany, University of British Columbia, Vancouver, BC V6T 1Z4, Canada. [4] BGI Genomics, BGI-Shenzhen, Shenzhen 518083, China. [5] School of BioSciences, The University of Melbourne, Melbourne, VIC 3010, Australia. [6] Tianjin Key Laboratory for Modern Drug Delivery & High-Efficiency, Collaborative Innovation Center of Chemical Science and Engineering, School of Pharmaceutical Science and Technology, Tianjin University, Tianjin 300072, China. [7] Department of Biological Sciences, Brock University, 1812 Sir Isaac Brock Way, St. Catharines, ON L2S 3A1, Canada. [8] Graham Centre, Charles Sturt University, Orange, NSW 2800, Australia. [9] School of Biological Sciences, University of Adelaide, Adelaide, Australia. [10] Instituto Superior de Formación Docente Salomé Ureña (ISFODOSU), Recinto Lus Napoleón Núñez Molina, Carretera Duarte, Km 10 1/2, Municipio de Licey Al Medio, Provincia de Santiago, República Dominicana. [11] Faculty of Land and Food Systems, University of British Columbia, Vancouver, BC V6T 1Z4, Canada. [12] Agriculture and Agri-Food Canada, Lethbridge Research Centre, Lethbridge, AB, Canada. [13] Division of Plant Sciences, University of Missouri, Columbia, MO 65211, USA. [14] Plant Health & Environment Laboratory, Ministry for Primary Industries, Auckland, New Zealand. [15] Department of Plant, Food, and Environmental Sciences, Faculty of Agriculture, Dalhousie University, PO Box 550, Truro, NS B2N 5E3, Canada. [16] Natural History Museum, Cromwell Road, South Kensington, SW7 5BD London, UK. [17] Institute of Plant Protection, Fujian Academy of Agricultural Sciences, Fuzhou 350003, China. [18] BGI-Shenzhen, Shenzhen 518083, China. [19] James D. Watson Institute of Genome Sciences, Hangzhou 310058, China. [20] These authors contributed equally: Minsheng You, Fushi Ke, Shijun You, Zhangyan Wu, Qingfeng Liu, Weiyi He. [21] Deceased: Carl J. Douglas. ✉email: msyou@fafu.edu.cn; lvasseur@brocku.ca; ggurr@csu.edu.au; yuezhen@bgi.com

The diamondback moth, *Plutella xylostella* (L.) (Lepidoptera: Plutellidae), is the most widely distributed species among all butterflies and moths on Earth[1]. This major pest is an oligophagous herbivore of cultivated and wild cruciferous plants (Brassicaceae), including many economically important food crops such as cabbage, cauliflower, and rapeseed[1,2]. The total economic cost of its damage and management worldwide is estimated at US $4–5 billion per year[1,3]. This is the first species for which its field populations have been documented to have evolved resistance to DDT, the iconic chemical insecticide of the 1950s[4] and to toxins of *Bacillus thuringiensis* (Bt) developed as an insecticide for control of Lepidoptera in the 1990s[5]. *P. xylostella* has now developed resistance to all major classes of insecticide and is increasingly difficult to control in the field[1].

Although intensive research has been done on the biology, ecology and management of *P. xylostella* during recent decades[1,2,6], our knowledge of its geographical origin and how it has become such a highly successful pest in all continents except Antarctica remains surprisingly incomplete and highly controversial[7–10]. To date, little is known about the global patterns of genomic variation in this species, which is essential for understanding the evolutionary history of *P. xylostella* together with the genetic basis of its rapid adaptation to insecticides.

In this study, we identify the origin of *P. xylostella* as South America using a global sample collection and nuclear/mitochondrial genome sequencing of all individuals, along with *COI* sequences for these and other specimens from BOLD. Further, we analyze the nuclear genomes of our specimens combined with geographical and historical information to reveal its dispersal routes and the progressive timing of global expansion. Based on the sequenced SNPs, we investigate the genomic signatures of selection to address the underlying mechanism associated with the local adaptation of this pest species.

Results

Global pattern of genomic variation. We first characterized the global pattern of variation among 532 genomes of *P. xylostella* using a worldwide sample of specimens collected from different locations (sites) in a stratified fashion reflecting a diverse range of biogeographical regions (Fig. 1a, Supplementary Fig. 1 and Supplementary Table 1) and covering an extensive scope of the eco-climatic index (Supplementary Fig. 2). Each of the individual genomes was sequenced with the Illumina sequencing system (HiSeq 2000) to produce 90-bp paired-end raw reads (Supplementary Fig. 1). A total of 1,797 Gb quality filtered reads (Supplementary Table 2) were mapped to the *P. xylostella* reference genome[11] using Stampy v1.0.27[12]. Individuals with low mapping rate or coverage (<60%) were excluded, and a total of 532 individuals were retained for variant discovery (Supplementary Table 2). After calibrating and filtering of the low-quality variants (Supplementary Fig. 1), we generated a genomic dataset containing 40,107,925 SNPs and 22,736,441 indels (Supplementary Tables 3 and 4), representing one variant on average in every six bp of the reference genome[11]. This is the densest variant map for any organism, including the recently released data of human[13] and *Arabidopsis thaliana*[14]. The global pattern of genomic variation (Fig. 1b), regional diversity of individual-based SNPs (Fig. 1c), and low ratio of shared SNPs (7.20%) among different geographical populations (Supplementary Table 5) revealed a high level of polymorphism that provided the capacity for *P. xylostella* to readily expand and adapt to different habitats worldwide.

Geographical origin. An earlier study proposed that *P. xylostella* originated from Mediterranean Europe[8] as many Brassicaceae crops were first domesticated in the this region. Other studies

predicted a South African[9] or Chinese origin[7] based on the regional diversity of indigenous Brassicaceae hosts or parasitoids of *P. xylostella*. An mtDNA-based analysis had supported claims of Africa as the possible area of origin of the species but used as few as 13 sampling sites worldwide without samples from South America[10]. A much larger and more representative collection of samples was required for accurate identification of the pest's geographical origin and better understanding of the evolutionary history of *P. xylostella*.

We extended these efforts to conduct a genomic study with high-quality nuclear SNPs of the worldwide sample (Fig. 1a, Supplementary Fig. 1 and Supplementary Table 1). Using the nuclear SNP data, a neighbor-joining (NJ) tree was constructed for global *P. xylostella* populations with the congeneric *P. australiana* as an outgroup according to our *COI*-based phylogenetic analysis (see Methods). Results revealed that multiple individuals collected from South America (SA) formed a distinct and basal clade (Fig. 2a and Supplementary Fig. 3). This was further confirmed by mitochondrial genomic data using the same specimens (Supplementary Fig. 4) and *COI* gene data from our specimens combined with additional published data (Supplementary Fig. 5). We also generated summary trees of different groups (see population genetic structure analysis below) based on 3,256 genome-wide local trees and revealed that the most prevalent topologies across the genome support *P. xylostella* populations of South America as the basal node closest to *P. australiana* (Supplementary Table 6). These results strongly suggest that *P. xylostella* originated in South America, where other endemic *Plutella* species are known[15]. Our study convincingly repolarizes the evolutionary history of *P. xylostella* from hypotheses of Old-World origin[7–10] to the New World.

The Brassicaceae family contains >3,700 species found on all continents but Antarctica[2,16]. In South America, with the richest cruciferous flora of the Southern Hemisphere[16,17], *P. xylostella* would have evolved on native Brassicales. Following European colonization of South America in the late 15th to early 16th century, the introduction and widespread use of domesticated Brassicaceae crops[18–20] would have expanded host plant resources used by *P. xylostella* on this continent. Initially, the species would have been confined to South America, isolated by oceans and limited by habitat/eco-climatic constraints until human interference[21].

Evolutionary and expansion history. Based on our phylogenetic analysis, in addition to the basal clade of South America (SA), we found four additional clades of North America (NA), Afro-Eurasia (A-E), South East Asia (SEA), and Oceania (OC) (Fig. 2a). These geographically clustered groups were supported by genetic structure analysis (Fig. 2b). A principal component analysis (PCA) (Fig. 2c) provided further evidence of the population structure of *P. xylostella* worldwide, with gene flow across the continent of Afro-Eurasia. The d_{XY}-based analysis[22] showed the lowest genetic differentiation between SA *P. xylostella* and *P. australiana*, followed by NA *P. xylostella* and *P. australiana*, A-E *P. xylostella* and *P. australiana*, SE Asian *P. xylostella* and *P. australiana*, and OC *P. xylostella* and *P. australiana* (Fig. 2d), which outlined the global colonization process of *P. xylostella* populations.

To further investigate the demographic history of *P. xylostella*, we estimated the population sizes and divergence times using a pairwise sequentially Markovian coalescent (PSMC) model[23]. It revealed a strikingly concordant history among geographical groups with a sharp decline following the last glacial maximum (LGM) in the early phase of evolution and a pattern of divergence in the recent past (with low resolution) among
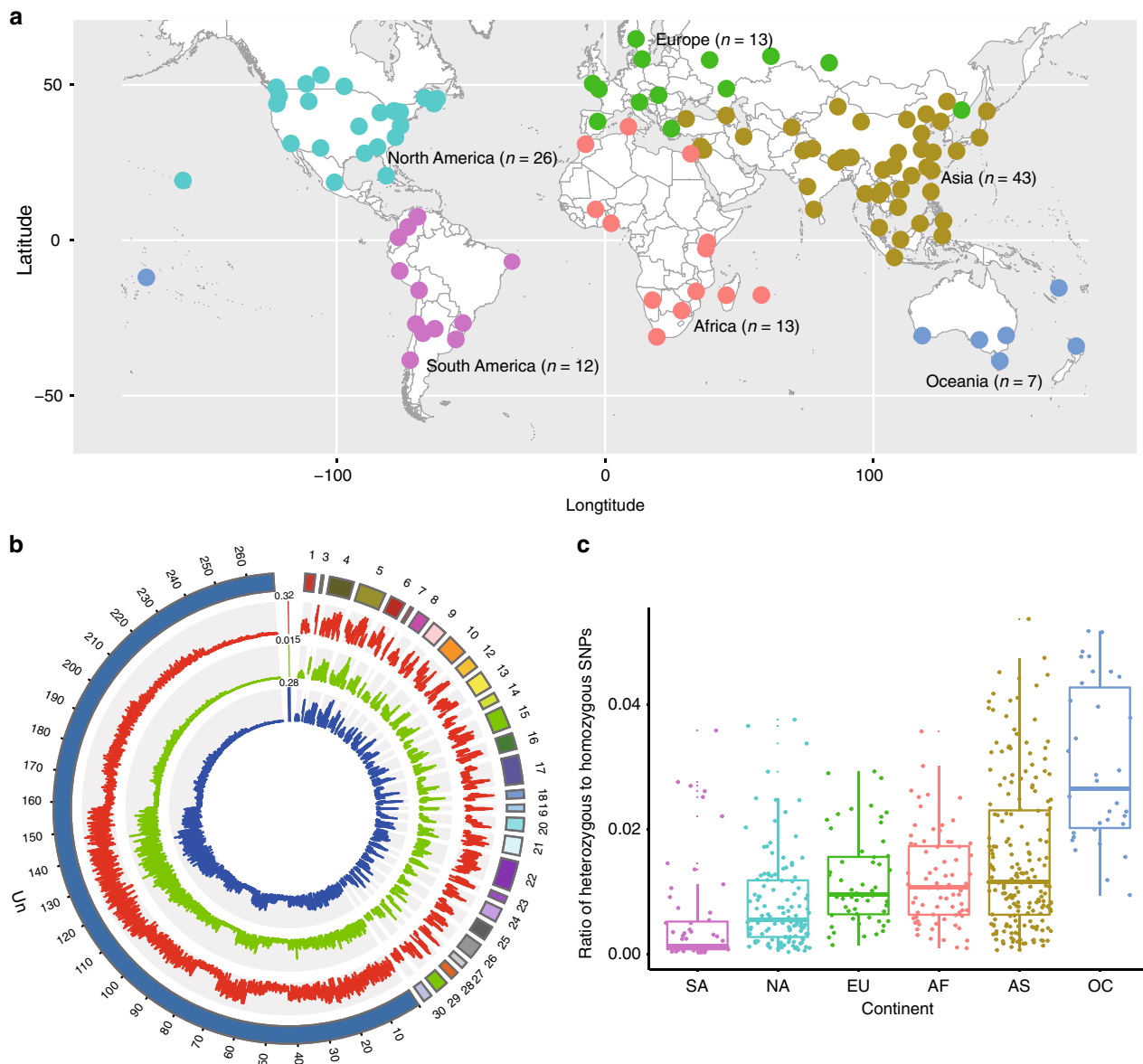
Fig. 1 Sampling locations and genomic variation of 532 *Plutella xylostella* individuals. a Illustration of 114 sampling locations showing that the *P. xylostella* specimens were collected from 55 countries across six of the seven continents in the world, i.e. excluding Antarctica. **b** Global patterns of the genomic variation. Circles from the outermost to innermost represent the reference genome of *P. xylostella* (including the partial sequences of 28 chromosomes and the scaffolds that were unable to be assigned (Un)), SNPs, nucleotide diversity and indels, respectively. **c** Ratio of individual-based heterozygous to homozygous SNPs in different continents of South America (SA), North America (NA), Europe (EU), Africa (AF), Asia (AS), and Oceania (OC). Boxes show the first and third quartile range (IQR) while whiskers extend to a maximum of 1.5 * IQR. Values for each of the individuals are shown as points surrounding boxplots. Source data are provided in the Source Data file. The map was generated with the rworldmap package v1.3-6[73].

different geographical groups (Supplementary Fig. 6). A recently published approach, SMC++ [24], which estimates the historical population sizes with higher resolution in the recent past compared to other methods such as PSMC[25], was used to predict the historical population sizes and divergence times of different groups. We found that *P. xylostella* experienced three major expansions across the world, with both North American and Afro-Eurasian lineages splitting from the ancestral lineage approximately 500 years ago, followed by South East Asia and then Oceania (Fig. 3a).

P. xylostella has remarkable genetic plasticity[26] and high level of genomic variation[11] that enable it to rapidly adapt to local environments, potentially leading to the change in the levels of genetic diversity among geographical populations. After a new

founder event of *P. xylostella*, the sizes of derived populations tend to rapidly grow based on the SMC++ analysis (Fig. 3a). This may have led to accumulation of mutations that were not present in the ancestral population, especially reflecting the likelihood that the species would have been subject to new and diverse selection pressures from novel plant hosts, novel agonists (e.g., pathogens and parasitoids), habitats, and climates[27,28], as well as becoming established in extensive, heterogeneous geographical regions that limited mixing. Genetic admixture, which may generate novel genotypes[27,29], among the *P. xylostella* populations in both the Old-World and North America was frequent according to our phylogenic reconstruction (Supplementary Fig. 3) and population structure analyses (Fig. 2b and c). Together, these effects help explain the pattern of increasing
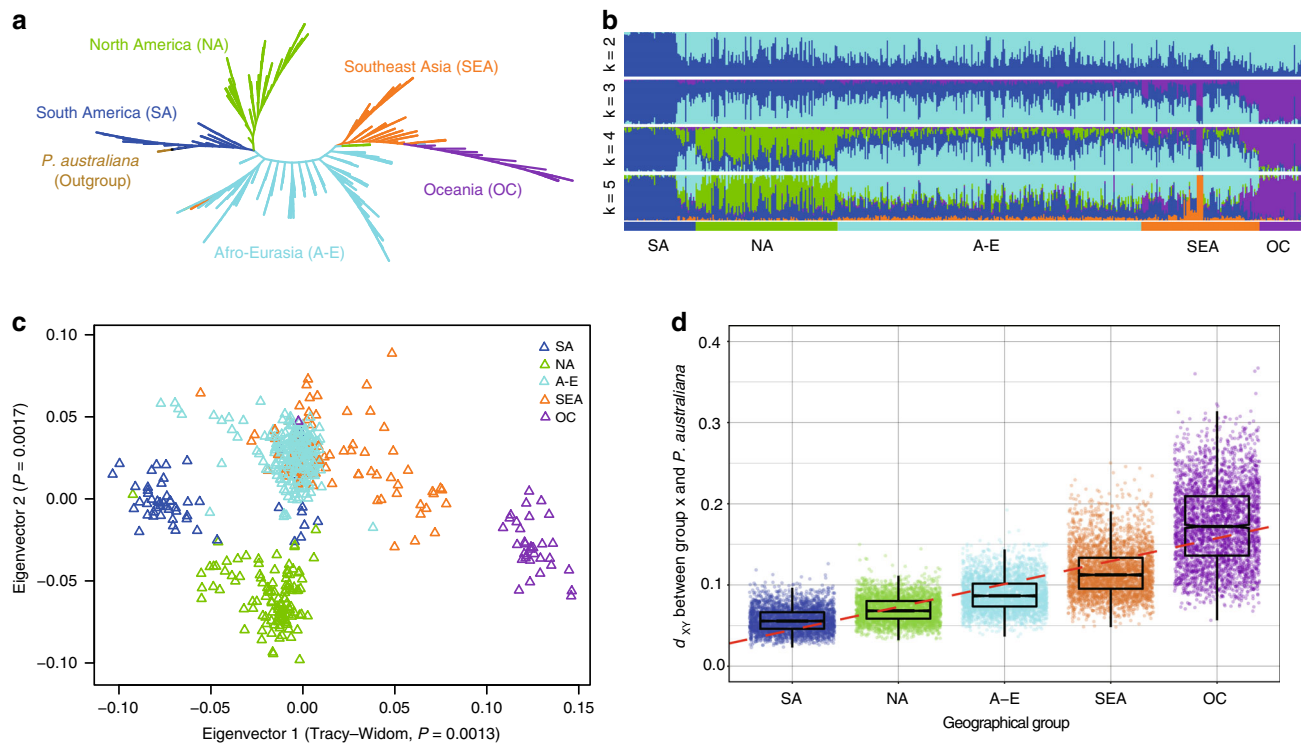
Fig. 2 Origin and global dispersal of *P. xylostella*. a Neighbor-joining phylogeny based on the nuclear genomes (SNPs) of 532 *P. xylostella* individuals collected worldwide, with two congeneric *P. australiana* individuals being used as an outgroup. The branch lengths are not scaled. See also Supplementary Fig. 3. **b** sNMF-based genetic structure and individual ancestry with colors in each column representing ancestry proportion over range of population sizes ($K = 2$–5, with an optimal $K = 5$). **c** PCA plot of the first two components generated by 2,839 SNPs. **d** d_{XY} calculated between each *P. xylostella* group and *P. australiana* based on 3,256 non-overlapping genome-wide windows. Boxes show the first and third quartile range (IQR) while whiskers extend to a maximum of 1.5 * IQR. Values for each window are shown as individual points surrounding boxplots. A simple linear model was fitted to the data (red dashed line), the line of best fit has an estimated R^2 value of 0.6149 and a slope of 0.02821. Source data are provided in the Source Data file.

genetic diversity in the range expansion process of *P. xylostella* populations (Fig. 1c).

Our phylogenetic and genetic analyses, PCA plot of the first two components, and d_{xy}-based analysis of differentiation (Fig. 2) supported the demographic evidence that the SA lineage was basal with the NA, A-E and SEA lineages diverging at progressively later stages and the OC lineage the most recent (Fig. 3a). These results were integrated with historical information[18–20], allowing us to propose a scenario of dispersal events for *P. xylostella* (Fig. 3b). We found that the major expansion events of *P. xylostella* were associated with human activities of agricultural production and trade. With European colonization, particularly the domestication of cruciferous crops with reduced glucosinolates and the introduction of Brassicaceae crops by European colonizers to South America[18,30], the original populations of *P. xylostella* in South America appear to have dispersed to and colonized various regions of the world (Fig. 3a and b). After colonizing the Mediterranean region, founder populations of *P. xylostella* likely dispersed across Europe, Western Asia, and Africa (Figs. 2 and 3)[31,32]. Like the spreading trend of *A. thaliana*, the diversity of *P. xylostella* populations in Europe and Eurasia exhibits a latitudinal pattern along the east-west axis (Fig. 3b), which has been facilitated by the rapid expansion of agriculture[14,33]. Around 200 years ago, independent dispersal events led the founder populations expanding eastwards into Asian countries first and then proceeding to Oceania (Figs. 2 and 3)[34,35], which corresponds to the most recent major region of colonization by Europeans. Records of Brassicaceae date from the "First Fleet" arrival in Australia in 1788 that carried produce and seeds of several *Brassica* species[35], and this was followed by

introduction and widespread cultivation of other brassicas by Chinese Australians[34]. Relative to the predicted earliest time when a Plutellidae ancestor may have become a cruciferous specialist (~54–90 million years ago)[11,36], the recent expansion events of *P. xylostella* (~200–500 years ago) further indicate that it could have survived on the indigenous Brassicaceae plants in South America for a long time, possibly in the timeframe of millions of years, after its putative divergence from an ancestor shared with its closest known relative, *P. australiana*[37].

Genomic signatures of local adaptation. Based on our globally sampled genomic data, we found that *P. xylostella* populations across the world exhibited a dense map of variants (Supplementary Tables 3 and 4), high level of polymorphism (Fig. 1b, c and Supplementary Fig. 7), and rapid decay of linkage disequilibrium (LD; Supplementary Fig. 8). These findings suggest a large effective population size for this species. Considering its genetic heterozygosity and rapid insecticide-resistance evolution, this species is well suited for a study of evolutionary adaptation under strong environmental selection pressure[38,39]. The intensive use of insecticides against *P. xylostella* has led to increased selection pressure for development of insecticide resistance[1,2,4,40–45]. We identified a global pattern of adaptive variation shown by the frequency distribution of three reported SNPs associated with insecticide resistance[46,47] (Supplementary Fig. 9). Such a global genotype distribution of three insecticide-resistance-related point mutation loci revealed that selection pressure resulting from insecticide applications had strong geographical dependence.
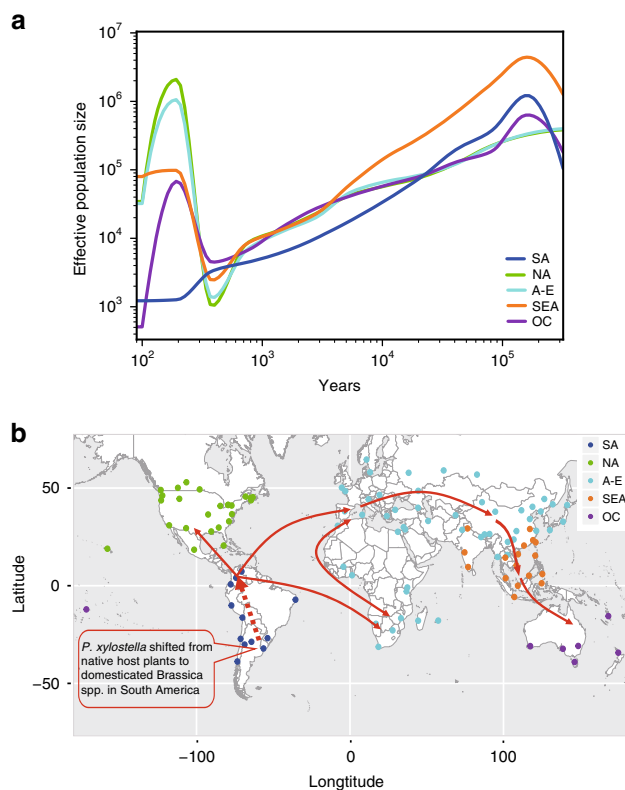
Fig. 3 Global expansion and demographic history of the *P. xylostella* populations. a Demographic history of *P. xylostella* illustrating the effective population sizes and divergence times based on multiple unphased individuals from different geographical groups and predicted by SMC++[24]. Coordinates are logarithmically scaled. **b** A proposed scenario of global colonization of the *P. xylostella* populations. The red arrows denote the proposed dispersal events of *P. xylostella* from South America towards other continents based on phylogenetic result (Fig. 2a), population genetic analyses (Fig. 2b–d), and demographic history (Fig. 3a). Source data are provided in the Source Data file. The map was generated with the rworldmap package v1.3-6[73].

To identify the genomic signatures of evolutionary adaptation for *P. xylostella*, we ran a genome-wide association study (GWAS) using the eigenvector1 of PCA as a "phenotype" (EigenGWAS)[48] to isolate a group of 75 individuals from Southeast Asia and Oceania (Supplementary Fig. 10). This reflects the fact that in these tropical and subtropical regions, cruciferous crops are massively and continuously grown year-round in a variety of cropping systems from backyard gardens to large-scale farms, resulting in favorable conditions for *P. xylostella* to develop and frequently outbreak throughout the year[1,2].

We identified 3,827 significantly differentiated SNPs ($P \leq 1e^{-8}$) (Supplementary Fig. 11a, right) with high level of genetic differentiation (F_{ST}) from 64,960 filtered SNPs (Supplementary Fig. 11a, left), which indicated that numerical distribution the significantly differentiated SNPs was proportionally similar to that of the filtered SNPs in each of the four genomic regions (Supplementary Fig. 11a). These outliers contained 1,179 candidate genes, being the most highly represented in metabolic and molecular signaling-related pathways according to the GO and KEGG analysis (shown with top 20 GO terms and KEGG pathways; Figs. 11b and 12). Among the 1,179 candidate genes under divergent selection we found 93 that were annotated in the published *P. xylostella* genome with known functions of detoxification of plant defense compounds and insecticide resistance[11]. We then identified six genes with non-synonymous

SNPs in coding regions. Three of them, including carboxypeptidase A (Px005867), P450-CYP2 (Px002515), and juvenile hormone esterase (JHE, Px003448) have reliable structural templates available in the Protein Data Bank, which allowed us to create homology models for these three enzymes using Schrödinger software[49].

Signals of divergent selection for the three non-synonymous mutations were identified according to their global distribution of genotype frequency (Supplementary Fig. 14). Comparison of the predicted structures between wild-type (WT) and mutant (Mut) enzymes revealed the potential impacts of these three mutations on the structural changes of these enzymes (Supplementary Fig. 15), which provides a cue for further experimentally-based research to establish a functional relationship between mutations and insecticide resistance[50,51].

Discussion

The present study improves our understanding of the origin, evolution, and genetic bases of adaptation in *P. xylostella*, a species with worldwide importance for pest management and food safety. Using a global sample collection (532 individuals) covering all six continents where the species occurs and nuclear and mitochondrial genomes as well as *COI* sequencing of all individuals, we have identified the area of origin of *P. xylostella* as South America. The result contrasts with previous hypotheses that suggested the Mediterranean region[8], South Africa[9,10] or China[7] as possible areas of origin of the species. Further, the phylogeographical profiling reveals that *P. xylostella* expansion events and timing have been facilitated by human socioeconomic activities. Genes in metabolic and molecular signaling-related pathways are putative candidates involved in evolutionary adaptation under the strong selective pressure of insecticides. Our results illustrate the utility of emerging genomic approaches to understand historical patterns of species expansion, and further address the underlying mechanisms associated with the worldwide dispersal of this notorious pest species.

Methods

Sample collection and DNA extraction. Based on the globally-distributed nature of *P. xylostella*, we developed a sampling plan with broad geographic scope (Fig. 1a and Supplementary Fig. 1). The global samples of *P. xylostella* were collected in 2012–2014 from 114 locations that cover broad regions throughout the world, with 13 samples from Africa including Madagascar, 43 samples from Asia, 13 samples from Europe, 26 samples from North America including Hawaii, 12 samples from South America, and 7 samples from Oceania (Fig. 1a and Supplementary Table 1). Our collection covers an extensive range of the eco-climatic index and areas that support differing numbers of annual generations, including those regions with year-round persistence of *P. xylostella* to others that are only seasonally suitable for growth and development of the species (Supplementary Fig. 2). Within each location, larvae, pupae, or adults were collected from cruciferous vegetable fields. Field-collected samples were morphologically inspected and genetically checked with *COI* sequences to confirm their identity.

The samples were preserved in 95% alcohol at −80 °C prior to DNA extraction. At least five individuals from each sampling location were used for DNA extraction. For quality control, each individual was washed twice using double-distilled water, and then dissected to remove the midgut including its microbiome and parasitoids to eliminate potential DNA contamination (Supplementary Fig. 1). To avoid unintentional biases, the individuals were each allocated a code number (Supplementary Table 2) in a double-blind fashion to obscure the origin of the insect to all handlers and analysts who identified the insect, its DNA or any associated genomic data.

DNA was extracted from each individual using DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) following the manufacturer's protocol. DNA was eluted from the DNeasy Mini spin column in 200 µl TE buffer. Concentration and integrity of the total DNA for each individual was measured with a Qubit Fluorometer (Invitrogen, Carlsbad, CA, USA) and agarose gel electrophoresis.

DNA barcoding and sequencing. A *Cytochrome Oxidase I* (*COI*) mitochondrial gene fragment of up to 658 bp was amplified and sequenced using Sanger sequencing, and then queried to the BOLD system[52] to confirm the species identity for each individual. In the 468 sequences of this dataset [https://doi.org/10.5883/DS-PLUT1][37,52],

399 sequences belong to *P. xylostella* (BOLD:AAA1513 [http://www.boldsystems.org/index.php/Public_BarcodeCluster?clusteruri=BOLD:AAA1513]) and 58 to *P. australiana* (BOLD:AAC6876 [http://www.boldsystems.org/index.php/Public_BarcodeCluster?clusteruri=BOLD:AAC6876]) while the rest belongs to other potential outgroup taxa in the family Plutellidae.

Genomic sequencing was performed with Illumina HiSeq 2000 at BGI, Shenzhen, China, to produce 90 bp paired-end reads for every individual. Considering the cosmopolitan distribution of *P. xylostella*, we aimed to sequence a large number of individual genomes across various geographical locations with a 5–10× coverage for each individual, which is a strategy previously used for the 1000 human genomes project[53] and *Apis mellifera*[54]. Two *P. australiana*[37] individuals were also sequenced with a 30× coverage and used as an outgroup for comparative analysis of genetic differentiation with the *P. xylostella* populations and construction of the phylogenetic tree. Sequencing libraries for each of the *P. xylostella* individuals were constructed according to the manufacturer's protocol. The quality and yield of the library were tested using the Agilent 2100 Bioanalyzer and ABI StepOnePlus Real-Time PCR System.

Data filtering, mapping and SNP calling. Raw reads were processed to obtain clean reads using custom scripts. Poor reads with 10 ambiguous "N" bases, >40% low-quality bases, or identical sequences at the two ends were filtered out.

We artificially allocated scaffolds of the *P. xylostella* genome into 20 synthetic chromosomes, and the SNP calling, and subsequent analyses were all performed on these 20 "chromosomes". Stampy (v1.0.27)[12] was employed to map the clean reads onto our *P. xylostella* reference genome (v2)[11] using default parameters. Subsequently, alignments for each of the individual samples were sorted with SortSam of Picard-tools (v-1.117, https://sourceforge.net/projects/picard/), and processed by removing duplicate reads with MarkDuplicates of Picard-tools. Reads with indels were realigned using RealignerTargetCreator and IndelRealigner in the Genome Analysis Toolkit (GATK v-3.2.2)[55] to avoid misalignment around indels. After realignment, base-quality scores were recalibrated using BaseRecalibrator based on the reference SNP set, which was generated using UnifiedGenotyper and SAMtools[56] from the 532 individuals. The sequencing and mapping statistics are summarized in Supplementary Table 2.

SNP calling was then performed using the GATK HaplotypeCaller with parameters --emitRefConfidence GVCF --variant_index_type LINEAR --variant_index_parameter 128,000. VariantRecalibrator was first used to create a Gaussian mixture model to examine the annotation values over a high-quality subset previously generated by UnifiedGenotyper and SAMtools and evaluate all input variants. ApplyRecalibration was used to designate the best parameters for each of the variants. Finally, VariantFiltration was used to filter the SNPs. The filtering settings were as follows: QD < 2.0 || MQ < 40.0 || ReadPosRankSum < −8.0 || FS > 60.0 || HaplotypeScore > 13.0 || MQRankSum < −12.5.

To resolve the origin of *P. xylostella*, two congeneric *P. australiana* individuals were used as an outgroup species. The sequencing reads of *P. australiana* were aligned to the *P. xylostella* genome using Stampy (v1.0.27)[12] with default parameters and reordered and sorted by Picard (https://broadinstitute.github.io/picard/). SOAPsnp (http://soap.genomics.org.cn/soapsnp.html) was then used to detect SNPs in each *P. australiana* individual with at least three supporting reads, and to assemble the consensus sequence for the *P. australiana* individuals based on the alignment of the sequencing reads with the *P. xylostella* genome. The genomic dataset of two *P. australiana* individuals and 532 *P. xylostella* individuals was used for phylogenetic tree construction.

To identify mitochondrial variants of *P. xylostella*, we also called SNPs using the mitochondrial genome of *P. xylostella* (GenBank KM023645 [https://www.ncbi.nlm.nih.gov/search/all/?term=KM023645]) as a reference. The same SNP calling procedure as done for nuclear SNP calling was employed, while a haplotype setting was used. The mitochondrial genome of *P. australiana* was reconstructed using MITObim[57] with a *P. australiana* COI barcode sequence as the seed and the *P. xylostella* mitochondrial genome as the reference.

Construction of the phylogenetic trees. Phylogenetic relationships of nuclear and mitochondrial genomes were analyzed among 532 individual samples of *P. xylostella* with two samples of *P. australiana* used as an outgroup. A phylogenetic tree based on the nuclear genomes was constructed (Fig. 2a and Supplementary Fig. 3), using the neighbor-joining (NJ) method[58], based on a genetic distance matrix (https://github.com/BGI-shenzhen/VCF2Dis), and calculated by the software PHYLIP v3.695 (http://evolution.genetics.washington.edu/phylip.html). Mitochondrial genomes were also used for phylogenetic tree construction using the NJ method implemented in PHYLIP, and a frequency tree (Supplementary Fig. 4) was generated using the consensus module with 1000 bootstraps.

To further confirm the origin and evolutionary relationships of *P. xylostella* populations based on nuclear and mitochondrial SNPs, a *COI*-based phylogenetic tree (Supplementary Fig. 5) was constructed based on NJ method with 1000 bootstraps using MEGA5[59]. This tree included the sequences of 532 *P. xylostella* individuals collected worldwide and two *P. australiana* individuals collected in Australia, as well as individual sequences of five non-Australian *Plutella* species (with two individual sequences for each of *P. armoraciae*, *P. porrectella*, *P. geniatella*, *P. hyperboreella* and one of *P. notabilis*), *Eidophasia vanella*, *P. australiana*, and *P. xylostella* downloaded from BOLD [https://doi.org/10.5883/DS-PLUT1][37,52], and

two undescribed taxa ('kaloko' and 'napoopoo') from Hawaii[60,61] downloaded from GenBank, with accession codes AF019041 [https://www.ncbi.nlm.nih.gov/nuccore/AF019041] and AF019042 [https://www.ncbi.nlm.nih.gov/nuccore/AF019042].

Population genetic pattern analysis. Bi-allelic SNPs presenting >95% individuals with a minor allele frequency of over 0.2 in the dataset were kept using vcftools[62] and used for population genetic structure analysis. We sampled one SNP from a 25 bp DNA window to generate loci independent of linkage disequilibrium. A total of 2,839 SNPs was retained for further analysis. The population genetic structure was analyzed using sNMF[63] with the pre-defined genetic clusters increased from $K = 2$ to $K = 8$ and illustrated with POPHELPER[64]. Principal component analysis (PCA) was also conducted using PLINK[65] with the same dataset, and a Tracy-Widom test was used to determine the significant level of the eigenvectors. The results (Fig. 2b) further confirmed and supported the five geographically clustered groups of *P. xylostella* populations worldwide based on previous nuclear phylogenetic analysis.

We generated genome-wide summary trees of different groups based on local trees. Variants with a maximum missing rate of 70% were filtered, and then converted to Genomic Data Structure (GDS) using SeqArray[66]. Local distance matrix was calculated using R scripts (https://github.com/CMWbio) with a bin of 5,000 SNPs in a maximum interval of 100 kb. A total of 3,256 local trees were generated across the genome. TWISST[67] was then used to calculate topology weighting for each local tree with 1,000 iterations (Supplementary Table 6). We also calculated d_{XY} values[22] in these 3,256 windows between five identified groups and the outgroup population to investigate genetic differentiation pattern during the global colonization of *P. xylostella* (Fig. 2d).

We presented the global genotype distribution of three previously reported SNPs associated with insecticide resistance[46,47] to show the geographical dependence of these point mutation loci (G4946E, L1014F and T929I; Supplementary Fig. 9). G4946E in ryanodine receptor was involved in resistance to diamide[46], and L1014F and T929I in sodium channel were associated with resistance to pyrethroid[47].

Demographic history. We selected one individual with high sequencing depth from each group to estimate the demographic history of *P. xylostella* using Pairwise Sequentially Markovian Coalescence (PSMC)[23], with a generation time of 0.1 years and a mutation rate[68] of 8.4×10^{-9} (Supplementary Fig. 6). A recently published approach with higher resolution in the recent past compared to PSMC accuracy, SMC++[24,25], was used to predict the demographic history (or population sizes and divergence times) of *P. xylostella* based on multiple unphased individuals (Fig. 3a). Five previously defined groups (or clades) were used for the analysis. We used a mutation rate of 8.4×10^{-9} from *Drosophila*[68], and 10 generations of *P. xylostella* per year estimated with global observations and records[1–3]. The short generation time of *P. xylostella* makes possible the reliable and precise estimation of effective population sizes in the recent past using the method of SMC++[25].

Identification of the loci under selection. An approach of genome-wide association study with the first eigenvector from the PCA as a "phenotype" (Eigen-GWAS)[48] was recently developed to identify single SNPs that contribute to the genetic differentiation (eigenvector) of two populations based on regression analysis. By using individual-level eigenvectors as phenotypes (Y in regression analysis) and single SNPs (X in regression analysis) in a linear regression, the resulting regression coefficients are equivalent to singular value decomposition (SVD) SNP effects and used to identify loci under selection along gradients of ancestry[48]. EigenGWAS also used a correction parameter to filter out signals of population stratification (i.e. caused by geography/drift), which allows the loci under selection to be identified. This approach has been successfully used to identify the loci under divergent selection between the UK and Dutch populations of great tit (*Parus major*) for better understanding of how genetic signatures of selection translate into variation in fitness and phenotypes[38].

To identify the genomic signatures of selection for *P. xylostella*, 64,960 SNPs (Supplementary Fig. 11a) were obtained after filtering with a missing rate ≤5% and a minor allele frequency ≥0.05 by vcftools[62] from the genomic dataset of our global samples. Based on the filtered SNPs, we ran the approach of EigenGWAS using a stringent level of genome-wide significance threshold ($P \le 1e^{-8}$)[69]. A total of 3,827 loci (or outlier) under selection (Supplementary Fig. 11a) were identified for further functional annotation.

Based on the genomic database of *P. xylostella* (http://iae.fafu.edu.cn/DBM/index.php), we searched for candidate genes with the outliers. GO annotation and classification of the candidate genes were conducted using Blast2GO (version 2.5.0)[70] and WEGO[71]. Pathways of the candidate genes were identified (Supplementary Fig. 11b) using the KEGG database (http://www.genome.jp/kegg/pathway.html). The genes enriched in the first 20 GO terms and KEGG pathways, and the value of fixation index (F_{ST}) was calculated for each of the identified loci using vcftools[58] (Supplementary Figs. 11b; 12 and 13).

Homology modeling. The structural models for the wild-type carboxypeptidase A, P450-CYP2 and juvenile hormone esterase were created by Prime module of Schrödinger software using human carboxypeptidase structure (PDB ID: 1PCA), fish cytochrome P450 structure (PDB ID: 4R1Z) and human acetylcholinesterase

structure (PDB ID: 4BDT) as templates, respectively (Supplementary Fig. 15). The resistant-mutant models were developed by introducing mutations to the wild-type structural models, followed by further energy minimization using Chimera[72]. All of the structural figures were also generated by Chimera[69].

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Raw reads of all 532 sequenced individuals have been deposited in the CNSA (https://db. cngb.org/cnsa/) of CNGBdb with the accession code CNP0000018, and been synchronously deposited in the EMBL Nucleotide Sequence Database (ENA) (https:// www.ebi.ac.uk/ena) with the accession code PRJEB24034. Sequences of five non-Australian *Plutella* species (with two individual sequences for each of *P. armoraciae*, *P. porrectella*, *P. geniatella*, *P. hyperboreella* and one of *P. notabilis*), *Eidophasia vanella*, *P. australiana*, and *P. xylostella* were downloaded from BOLD [https://doi.org/10.5883/DS-PLUT1][37,52]. Sequences of two undescribed taxa ('kaloko' and 'napoopoo') from Hawaii[60,61] were downloaded from the GenBank, with accession codes AF019041 [https://www.ncbi.nlm.nih.gov/nuccore/AF019041] and AF019042 [https://www.ncbi. nlm.nih.gov/nuccore/AF019042]. The source data underlying Figs. 1b, c, 2b–d, 3a as well as Supplementary Figs. 2, 6, 7, 8, 9, 10, 11, 12, 13, 14 and15 are provided as a Source Data file.

References

1. Furlong, M. J., Wright, D. J. & Dosdall, L. M. Diamondback moth ecology and management: problems, progress, and prospects. *Annu. Rev. Entomol.* **58**, 517–541 (2013).
2. Talekar, N. S. & Shelton, A. M. Biology, ecology, and management of the diamondback Moth. *Annu. Rev. Entomol.* **38**, 275–301 (1993).
3. Zalucki, M. P. & Furlong, M. J. Predicting outbreaks of a migratory pest: an analysis of DBM distribution and abundance revisited. In *Proceedings of The Sixth International Workshop on Management of Diamondback Moth and Other Crucifer Pests* (eds Srinivasan, R. Shelton, A. M. & Collins, H. L.) 8–14 (AVRDC, 2011).
4. Ankersmit, G. W. DDT-resistance in *Plutella maculipennis* (Curt.) (Lep.) in Java. *Bull. Entomol. Res.* **44**, 421–425 (1953).
5. Tabashnik, B. E., Cushing, N. L., Finson, N. & Johnson, M. W. Field development of resistance to *Bacillus thuringiensis* in diamondback moth (Lepidoptera: Plutellidae). *J. Econ. Entomol.* **83**, 1671–1676 (1990).
6. Li, Z., Feng, X., Liu, S., You, M. & Furlong, M. J. Biology, ecology, and management of the diamondback moth in China. *Annu. Rev. Entomol.* **61**, 277–296 (2016).
7. Liu, S., Wang, X., Guo, S., He, J. & Shi, Z. Seasonal abundance of the parasitoid complex associated with the diamondback moth, *Plutella xylostella* (Lepidoptera: Plutellidae) in Hangzhou, China. *Bull. Entomol. Res.* **90**, 221–231 (2000).
8. Hardy, J. E. *Plutella maculipennis*, Curt., its natural and biological control in England. *Bull. Entomol. Res.* **29**, 343–372 (2009).
9. Kfir, R. Origin of the diamondback moth (Lepidoptera: Plutellidae). *Ann. Entomol. Soc. Am.* **91**, 164–167 (1998).
10. Juric, I., Salzburger, W. & Balmer, O. Spread and global population structure of the diamondback moth *Plutella xylostella* (Lepidoptera: Plutellidae) and its larval parasitoids *Diadegma semiclausum* and *Diadegma fenestrale* (Hymenoptera: Ichneumonidae) based on mtDNA. *Bull. Entomol. Res.* **107**, 155–164 (2017).
11. You, M. S. et al. A heterozygous moth genome provides insights into herbivory and detoxification. *Nat. Genet.* **45**, 220–225 (2013).
12. Lunter, G. & Goodson, M. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res.* **21**, 936–939 (2011).
13. 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
14. Alonso-Blanco, C. et al. 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell* **166**, 481–491 (2016).
15. Meyrick, E. *Exotic Microlepidoptera* (Taylor & Francis Group, 1931).
16. Appel, O. & Al-Shehbaz, I. A. Cruciferae. In *The Families and Genera of Vascular Plants vol.5 Flowering Plants · Dicotyledons* (eds Kubitzki, K. Rohwer, J. G. & Bittrich, V.) 75–174 (Springer, 2003).
17. Al-Shehbaz, I. A., Cano, A., Trinidad, H. & Navarro, E. New species of *Brayopsis*, *Descurainia*, *Draba*, *Neuontobotrys* and *Weberbauera* (Brassicaceae) from Peru. *Kew Bull.* **68**, 219–231 (2013).
18. Super, J. C. *Food, Conquest, and Colonization in the Sixteenth-century Spanish America* 192 (University of New Mexico Press, Albuquerque, 1988).
19. de Oviedo y Valdés, G. F. *La Historia General de Las Indias* (1535).
20. Gallagher, D. American plants in Sub-Saharan Africa: a review of the archaeological evidence. *Azania.: Archaeol. Res. Afr.* **51**, 24–61 (2016).
21. Frenot, Y. et al. Biological invasions in the Antarctic: extent, impacts and implications. *Biol. Rev.* **80**, 45–72 (2005).
22. Nei, M. *Molecular Evolutionary Genetics* (Columbia University Press, 1987).
23. Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011).
24. Terhorst, J., Kamm, J. A. & Song, Y. S. Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat. Genet.* **49**, 303–309 (2017).
25. Patton, A. et al. Contemporary demographic reconstruction methods are robust to genome assembly quality: a case study in Tasmanian devils. *Mol. Biol. Evol.* **36**, 2906–2921 (2019).
26. Henniges-Janssen, K. et al. Complex inheritance of larval adaptation in *Plutella xylostella* to a novel host plant. *Heredity* **107**, 421–432 (2011).
27. Estoup, A. et al. Is there a genetic paradox of biological invasion? *Annu. Rev. Ecol. Syst.* **47**, 51–72 (2016).
28. Lavergne, S. & Molofsky, J. Increased genetic variation and evolutionary potential drive the success of an invasive grass. *Proc. Natl Acad. Sci. USA* **104**, 3883–3888 (2007).
29. Facon, B., Pointier, J. P., Jarne, P., Sarda, V. & David, P. High genetic variance in life-history strategies within invasive populations by way of multiple introductions. *Curr. Biol.* **18**, 363–67 (2008).
30. Fahey, J. W., Zalcmann, A. T. & Talalay, P. The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. *Phytochemistry* **59**, 5–51 (2002).
31. Dixon, G. R. *Vegetable Brassicas and Related Crucifers* (CABI, Wallingford, 2006).
32. Franzke, A., Lysak, M. A., Al-Shehbaz, I. A., Koch, M. A. & Mummenhoff, K. Cabbage family affairs: the evolutionary history of Brassicaceae. *Trends Plant Sci.* **16**, 108–116 (2011).
33. François, O., Blum, M. G., Jakobsson, M. & Rosenberg, N. A. Demographic history of European populations of *Arabidopsis thaliana*. *PLoS Genet.* **4**, e1000075 (2008).
34. Wahlqvist, M. L. Asian migration to Australia: food and health consequences. *Asia Pac. J. Clin. Nutr.* **11**, S562–S568 (2002).
35. Frost, A. *The First Fleet: The Real Story*. (Black Inc., 2012).
36. Wang, X. et al. The genome of the mesopolyploid crop species *Brassica rapa*. *Nat. Genet.* **43**, 1035 (2011).
37. Landry, J. F. & Hebert, P. D. N. *Plutella australiana* (Lepidoptera, Plutellidae), an overlooked diamondback moth revealed by DNA barcodes. *Zookeys* **327**, 43–63 (2013).
38. Bosse, M. et al. Recent natural selection causes adaptive evolution of an avian polygenic trait. *Science* **358**, 365–368 (2017).
39. Hawkins, N. J., Bass, C., Dixon, A. & Neve, P. The evolutionary origins of pesticide resistance. *Biol. Rev.* **94**, 135–155 (2018).
40. Atumurirava, F. & Furlong M. J. Diamondback moth resistance to commonly used insecticides in Fiji. In *Proceedings of The Sixth International Workshop on Management of Diamondback Moth and Other Crucifer Pests* (eds Srinivasan, R. Shelton, A. M. & Collins, H. L.) 216–221 (AVRDC, 2011).
41. Heisswolf, S. Houlding, B. J. & Deuter, P. L. A decade of integrated pest management (IPM) in Brassica vegetable crops-the role of farmer participation in its development in southern Queensland, Australia. In *Proceedings of The Third International Workshop on Management of Diamondback Moth and Other Crucifer Pests* (eds Sivapragasam, A. Loke, W. H. Hussan, A. K. & Lin, G. S.) 228–232 (MARDI, 1997).
42. Walker, G. P. Cameron, P. J. & Berry, N. A. Implementing an IPM programme for vegetable brassicas in New Zealand. In *Proceedings of The Fourth International Workshop on Management of Diamondback Moth and Other Crucifer Pests* (eds Endersby, N. M. & Ridland, P. M.) 365–370 (The Regional Institute Ltd, 2001).
43. Furlong, M. J. et al. Ecology of diamondback moth in Australian canola: landscape perspectives and the implications for management. *Aust. J. Exp. Agr.* **48**, 1494–1505 (2008).
44. Sonoda, S., Tsukahara, Y. M. & Tsumuki, H. Genomic organization of the para-sodium channel alpha-subunit genes from the pyrethroid-resistant and -susceptible strains of the diamondback moth. *Arch. Insect Biochem. Physiol.* **69**, 1–12 (2008).
45. He, W. et al. Developmental and insecticide-resistant insights from the de novo assembled transcriptome of the diamondback moth, *Plutella xylostella*. *Genomics* **99**, 169–177 (2012).
46. Troczka, B. et al. Resistance to diamide insecticides in diamondback moth, *Plutella xylostella* (Lepidoptera: Plutellidae) is associated with a mutation in the membrane-spanning domain of the ryanodine receptor. *Insect Biochem. Mol. Biol.* **42**, 873–880 (2012).

47. Endersby, N. M. et al. Widespread pyrethroid resistance in Australian diamondback moth, *Plutella xylostella* (L.), is related to multiple mutations in the para sodium channel gene. *Bull. Entomol. Res.* **101**, 393–405 (2011).

48. Chen, G., Lee, S. H., Zhu, Z., Benyamin, B. & Robinson, M. R. EigenGWAS: finding loci under selection through genome-wide association studies of eigenvectors in structured populations. *Heredity* **117**, 51–61 (2016).

49. Schrödinger Release 2017-1: Prime, Schrödinger, LLC, New York, NY. (2017).

50. Jackson, C. J. et al. Structure and function of an insect α-carboxylesterase (αEsterase7) associated with insecticide resistance. *Proc. Natl Acad. Sci. USA* **110**, 10177–10182 (2013).

51. Amichot, M. et al. Point mutations associated with insecticide resistance in the *Drosophila* cytochrome P450 Cyp6a2 enable DDT metabolism. *Eur. J. Biochem.* **271**, 1250–1257 (2004).

52. Ratnasingham, S. & Hebert, P. D. N. BOLD: the barcode of life data system (http://www.barcodinglife.org). *Mol. Ecol. Notes* **7**, 355–364 (2007).

53. The 1000 Genomes Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).

54. Wallberg, A. et al. A worldwide survey of genome sequence variation provides insight into the evolutionary history of the honeybee *Apis mellifera*. *Nat. Genet.* **46**, 1081–1088 (2014).

55. DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).

56. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).

57. Hahn, C., Bachmann, L. & Chevreux, B. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Res.* **41**, e129 (2013).

58. Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987).

59. 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).

60. Chang, W. X. Z. et al. Mitochondrial DNA sequence variation among geographic strains of diamondback moth (Lepidoptera: Plutellidae). *Ann. Entomol. Soc. Am.* **90**, 590–595 (1997).

61. Robinson, G. S. & Sattler, K. *Plutella* in the Hawaiian Islands: relatives and host-races of the diamondback moth (Lepidoptera: Plutellidae). *Bish. Mus. Occas. Pap.* **67**, 1–27 (2001).

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

63. Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G. & Francois, O. Fast and efficient estimation of individual ancestry coefficients. *Genetics* **196**, 973–983 (2014).

64. Francis, R. M. POPHELPER: an R package and web app to analyze and visualize population structure. *Mol. Ecol. Resour.* **17**, 27–32 (2017).

65. Purcell, S. et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

66. Zheng, X. et al. SeqArray—a storage-efficient high-performance data format for WGS variant calls. *Bioinformatics* **33**, 2251–2257 (2017).

67. Simon, M. & Van Belleghem, S. Exploring evolutionary relationships across the genome using topology weighting. *Genetics* **206**, 429–438 (2017).

68. Haag-Liautard, C. et al. Direct estimation of per nucleotide and genomic deleterious mutation rates in. *Drosop. Nat.* **445**, 82–85 (2007).

69. Kanai, M., Tanaka, T. & Okada, Y. Empirical estimation of genome-wide significance thresholds based on the 1000 Genomes Project data set. *J. Hum. Genet.* **61**, 861–866 (2016).

70. Conesa, A., Götz, S., Garcíagómez, J. M., Terol, J. & Robles, M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676 (2005).

71. Ye, J. et al. WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res.* **34**, W293–W297 (2006).

72. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

73. South, A. rworldmap: a new R package for mapping global data. *R J.* **3**, 35–43 (2011).

Author contributions

M.Y., F.K., S.Y., Z.W., Q.L., W.H., S.W.B., and Z. Yuchi contributed equally to this work. M.Y., L.V., G.M.G., C.J.D., Z. Yue, H.Y., and J.W. conceived, designed and/or managed the project. H.C., M.S.G., L.V., G.M.G., S.W.B., Q.S., Q.F., G.W.-P., D.C.L., J.B., T.L., L.P., M.X., L. Cai, Y.Z., Z.Z., S.L., Y.W., Q.Z., X.X., W.C., L. Chen, M. Zou, J.L., L.H., Y. Lin, Y. Lu, and M. Zhuang collected insects and/or prepared DNA samples for sequencing. F.K., S.Y., Z.W., Q.L., W.H., Z. Yuchi, Y.J., C. M.W., L.L., T.L., J.B., M.Z., Q.G., X.F., and Y.Y. performed experiments and/or data analyses. M.Y., F.K., S.Y., Z.W., W.H., Z. Yuchi, L.V., and G.M.G. co-wrote the manuscript. M.Y., F.K., S.Y., Z.W., W.H., Z. Yuchi, L.V., G.M.G., D.C.L., S.W.B., C.M.W., H.C., G.Y., M.B.I., M.S.G., Q.S., Q.F., G.W.-P. interpreted results and/or revised the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-16178-9.

Correspondence and requests for materials should be addressed to M.Y., L.V., G.M.G. or Z.Y.

Peer review information *Nature Communications* thanks Nicola Nadeau, and the other, anonymous, reviewers for their contribution to the peer review of this work. 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.