



OPEN

## Genome-wide analysis reveals distinct global populations of pink bollworm (*Pectinophora gossypiella*)

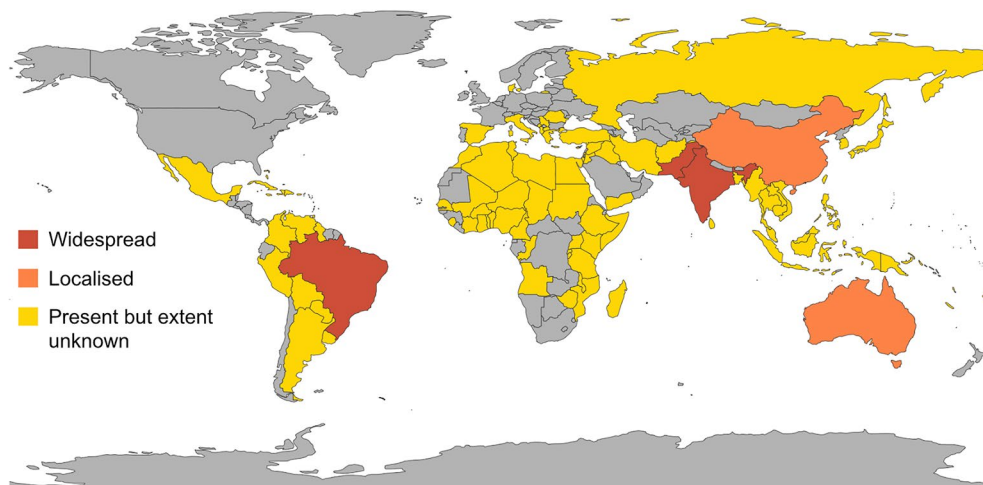
Paige Matheson<sup>1✉</sup>, Elahe Parvizi<sup>1</sup>, Jeffrey A. Fabrick<sup>2</sup>, Hamid Anees Siddiqui<sup>3</sup>, Bruce E. Tabashnik<sup>4</sup>, Tom Walsh<sup>5</sup> & Angela McGaughan<sup>1</sup>

The pink bollworm (*Pectinophora gossypiella*) is one of the world's most destructive pests of cotton. This invasive lepidopteran occurs in nearly all cotton-growing countries. Its presence in the Ord Valley of North West Australia poses a potential threat to the expanding cotton industry there. To assess this threat and better understand population structure of pink bollworm, we analysed genomic data from individuals collected in the field from North West Australia, India, and Pakistan, as well as from four laboratory colonies that originated in the United States. We identified single nucleotide polymorphisms (SNPs) using a reduced-representation, genotyping-by-sequencing technique (DArTseq). The final filtered dataset included 6355 SNPs and 88 individual genomes that clustered into five groups: Australia, India-Pakistan, and three groups from the United States. We also analysed sequences from Genbank for mitochondrial DNA (mtDNA) locus cytochrome c oxidase I (COI) for pink bollworm from six countries. We found low genetic diversity within populations and high differentiation between populations from different continents. The high genetic differentiation between Australia and the other populations and colonies sampled in this study reduces concerns about gene flow to North West Australia, particularly from populations in India and Pakistan that have evolved resistance to transgenic insecticidal cotton. We attribute the observed population structure to pink bollworm's narrow host plant range and limited dispersal between continents.

Invasive species represent a significant threat to agriculture, due to economic costs associated with management and reduced crop yields<sup>1</sup>. Pink bollworm (*Pectinophora gossypiella*) is a major pest of cotton that has colonised more than 150 countries worldwide, including much of tropical America, Africa, Asia, and Australasia (See Fig. 1)<sup>2,3</sup>. The origin of this invasive lepidopteran pest is not known, but a leading candidate is India<sup>4,5</sup> where it was first discovered damaging cotton in 1843<sup>6</sup>. It has also been hypothesised to have originated in Australia or South East Asia<sup>7</sup>.

In Australia, pink bollworm was first reported on cotton in Queensland in 1924<sup>8</sup>. Currently, it is thought to occur primarily in Western Australia and the Northern Territory<sup>8–11</sup>, which are over 1000 km from Australia's primary cotton production areas of New South Wales and southern Queensland. However, pink bollworm's presence in the Ord Valley in the northern part of Western Australia is a potential threat to the expanding cotton industry there, where over 1000 hectares of cotton were planted in 2020 for the first time since 2011<sup>12</sup>. Cotton genetically engineered to produce insecticidal proteins from the bacterium *Bacillus thuringiensis* (Bt) has been grown in Australia since 1996 and has been effective against pink bollworm in the United States and China<sup>13,14</sup>. However, pink bollworm populations in India and Pakistan have evolved resistance to Bt cotton, with serious practical consequences<sup>15</sup>. In principle, connectivity among populations could result in the introduction of Bt-resistant pink bollworm into North West Australia. Thus, a greater understanding of pink bollworm population structure could be useful for assessing the threat to cotton in Australia, as well for improving surveillance and management elsewhere.

<sup>1</sup>Te Aka Mātuatua - School of Science, University of Waikato, Hamilton, New Zealand. <sup>2</sup>United States Department of Agriculture Agricultural Research Service, United States Arid Land Agricultural Research Center, Maricopa, AZ 85138, USA. <sup>3</sup>Department of Biology, University of York, York, UK. <sup>4</sup>Department of Entomology, University of Arizona, Tucson, AZ 85721, USA. <sup>5</sup>Commonwealth Scientific Industrial Research Organisation Environment, Clunies Ross St, Acton, ACT 2601, Australia. ✉email: paige.matheson14@gmail.com



**Figure 1.** Map depicting where pink bollworm is widespread, localised, and present but extent unknown. Distribution data was extracted from distribution table at <https://www.cabidigitallibrary.org/doi/10.1079/cabicompndium.39417> and visualised on a map created in R version 4.1.2<sup>27</sup> using packages ‘rworldmap’ version 1.3.6<sup>37</sup> and ‘ggplot2’ version 3.4.2<sup>38</sup>.

Five previous studies of population structure in pink bollworm have analysed genetic variation based on DNA sequences from one to 13 loci<sup>5–7,16,17</sup>. Two of these studies evaluated only the mitochondrial DNA (mtDNA) cytochrome *c* oxidase subunit I (COI) locus in samples from India<sup>6,17</sup>, two focused primarily on China and used two mtDNA loci<sup>16</sup> or 13 microsatellites<sup>5</sup> respectively, and one was based on a *piggyBac*-like transposon insertion and its flanking sequences<sup>7</sup>. Of these, only the latter included populations from Australia. The conclusions from these studies were mixed, including apparently conflicting results from mtDNA and microsatellite DNA analysis of the same set of populations<sup>5,16</sup>. Population structure of pink bollworm has not been explored previously using genome-wide data that includes Australian samples and a large number of genetic markers. Such analysis may provide increased accuracy and resolution to address these discrepancies.

Here, we aimed to understand pink bollworm population structure by using a genotyping-by-sequencing technique (DARtseq)<sup>18</sup> to produce genome-wide data in the form of 6355 single nucleotide polymorphisms (SNPs) identified from 88 individual pink bollworm from seven wild populations in Australia, India, and Pakistan; and four laboratory colonies in the U.S. We used these data to assess the population structure of pink bollworm, with a particular focus on the Australian population. We also analysed data from Genbank for mtDNA COI for pink bollworm from six countries. Our results show minimal gene flow between Australia and the other populations that were sampled in this study, which reduces concerns about the threat of introductions of Bt-resistant pink bollworm from India and Pakistan.

## Methods

**Pink bollworm samples.** We analysed the genomes of 88 pink bollworm from Australia, India, Pakistan, and the U.S. (Table S1). Male moths were collected from near Kununurra, Western Australia (–15.65, 128.70) in March and April 2017 using pheromone traps (Agrisense Recharge Lures for *Pectinophora gossypiella*; Entosol (Australia) Pty Ltd., NSW Australia). Fourth instar larvae were collected from fields of Bt cotton producing Cry1Ac + Cry2Ab (Bt toxins) in 2010 from the states of Telangana, Maharashtra, Karnataka, and Andhra Pradesh in India<sup>19</sup>, and from fields of Bt cotton producing Cry1Ac in 2016–2017 from the provinces of Punjab and Sindh in Pakistan.

For pink bollworm from the U.S., we used individuals from four laboratory colonies in Arizona: APHIS-S\_1, APHIS-S\_2, Bt4R, and Bt4-R2, where this pest was declared eradicated in 2018<sup>20</sup>. Both APHIS-S colonies were obtained from the United States Department of Agriculture (USDA) Animal and Plant Health Inspection Service (APHIS) laboratory in Phoenix, Arizona, where they had been maintained for over 40 years without exposure to Bt toxins or other insecticides<sup>21,22</sup>. Subsets of the APHIS-S colony were provided to the USDA Agricultural Research Service in Maricopa, Arizona in 2006 (APHIS-S\_2) and 2018 (APHIS-S\_1). Bt4R and Bt4-R2 are laboratory-selected resistant colonies: Bt4R was derived from the Bt-susceptible Western Cotton Research Laboratory (WCRL) colony from 2007 to 2008 by laboratory selection for Cry1Ac resistance<sup>23</sup>. Bt4-R2 was obtained from Bt4R in 2010 and has 28-fold resistance to Cry1Ac and > 10,000-fold resistance to Cry2Ab<sup>24</sup>. Larvae from all four colonies were reared in the laboratory on wheat germ diet<sup>25</sup> at 26 °C and a photoperiod of 14 h light:10 h dark.

**DARtseq genotyping.** We sent 90 pink bollworm samples to Diversity Arrays Technology (DARt Pty Ltd, Canberra, Australia) for DNA extraction, quantification, and genotyping (see Supplementary Information Table S1 for further sample details). DARtseq™ is a restriction enzyme-based, complexity reduction method that employs a next generation sequencing platform, as detailed by Kilian et al. and Georges et al.<sup>18,26</sup>. Short-read sequences were then processed following the DARtseq bioinformatic pipeline to yield genotyping data in the form of SNPs<sup>26</sup>.

**SNP filtering.** We received genotyping outputs in the DarT ‘2 row’ format, where alleles are scored “0” for homozygous reference state, “1” for heterozygous, and “2” for homozygous alternate (or SNP) state<sup>26</sup>. This initial dataset contained 59,262 SNPs across 90 individuals. Using R version 4.1.2<sup>27</sup>, we converted this data into a genlight object using ‘adegenet’ version 2.1.7<sup>28</sup>, and then used ‘dartR’ version 2.0.4<sup>29</sup> and ‘radiator’ version 1.2.2<sup>30</sup> for data manipulation and filtering. We filtered SNPs by reproducibility (threshold: 0.98), call rate (threshold: 0.95), and minor allele frequency (threshold: 0.02). The final filtered dataset contained 88 individual genotypes, 6355 SNP markers, and 1.41% missing data.

**Heterozygosity and inbreeding.** In R, we used ‘hierfstat’ version 0.5.11<sup>31</sup> to calculate observed ( $H_o$ ) and expected ( $H_e$ ) heterozygosity, as well as inbreeding coefficients ( $F_{IS}$ ; a metric that ranges from  $-1$  to  $1$ , where values close to  $0$  meet the neutral expectation, values that approach  $1$  indicate a deficit of heterozygotes indicating inbreeding, and values approaching  $-1$  indicate an excess of heterozygotes) for each population from the filtered dataset.

**Population structure.** We used R to analyse genetic variation between and within populations by performing an Analysis of Molecular Variance (AMOVA) with the ‘poppr’ package version 2.9.3<sup>32</sup>. We randomly permuted the AMOVA output 1000 times to test if populations and colonies were significantly different using the function ‘randtest’ from the package ‘ade4’ version 1.7.19<sup>33</sup>. We then estimated pairwise genetic differentiation ( $F_{ST}$ ) between populations and colonies using the function ‘genet.dist’ and method ‘WC84’ from the ‘hierfstat’ package, version 0.5.11<sup>34,35</sup>.

We conducted a principal component analysis based on Euclidean genetic distances, using the function ‘glPCA’ implemented in the ‘adegenet’ package in R. We performed two analyses, one with only the 11 populations and colonies of pink bollworm, and another that also included an ‘outgroup’ population of spotted pink bollworm (*Pectinophora scutigera*; see Fig. S1).

We calculated individual admixture coefficients by first converting SNP data into ‘STRUCTURE’ format using the ‘gl2faststructure’ function implemented in the R package ‘dartR’, and then to ‘geno’ format using the ‘struct2geno’ function from ‘LEA’ version 3.6.0<sup>36</sup>. We ran sparse non-negative matrix factorisation on individuals using the ‘sNMF’ function also implemented in ‘LEA’ and analysed  $K$  values of  $1$ – $10$ , with 100 replications for each  $K$  value. We identified the  $K$  value that best explained our results using the cross-entropy criterion (Fig. S2). Admixture results were presented on a map created using the R packages ‘rworldmap’ version 1.3.6<sup>37</sup> and ‘ggplot2’ version 3.4.2<sup>38</sup>.

To examine relationships among samples in a phylogenetic context, we converted our SNP data to variant call format (VCF) using ‘radiator’ version 1.2.2<sup>30</sup> and then to phylip format using the command ‘vcf2phylip’ script (available at: <https://github.com/edgardomortiz/vcf2phylip>). A maximum likelihood (ML) phylogeny was constructed in Iqtree version 1.2.1<sup>39</sup> using the best-fit substitution model automatically selected by the software, with 10,000 bootstrap iterations to assess clade support. The resulting output was read into R using ‘ape’ version 5.6.2<sup>40</sup> for data visualisation.

Finally, to further investigate genetic diversity and population structure of global pink bollworm populations, we obtained 31 mitochondrial COI sequences in FASTA format from GenBank (Table S2) for populations from Australia, India, Pakistan, the U.S., Israel, and Kenya. COI sequences were aligned using MEGA version 11<sup>41</sup>. The nucleotide diversity of each population was then calculated using the ‘nuc.div’ function in R from the ‘pegas’ package version 1.1<sup>42</sup>. COI sequences were converted to haplotypes using the same package, and the function ‘haploNet’ was used to construct a haplotype network.

## Results

**Heterozygosity and inbreeding.** Observed heterozygosity ( $H_o$ ) was low within each of the 11 populations and colonies studied (range =  $0.060$ – $0.207$ , Table 1). As expected, mean  $H_o$  was lower for the four U.S. laboratory colonies ( $0.09$ ) than the seven field populations from Australia, India, and Pakistan ( $0.19$ ) ( $t_9 = 4.87$ ;  $P < 0.001$ ). Mean  $H_o$  was significantly higher for the field populations from India and Pakistan ( $0.20$ ) than for Australia ( $0.14$ ) ( $t_5 = 11.5$ ,  $P < 0.001$ ).  $H_o$  was not correlated with the number of field sites sampled per population across the seven field populations ( $r_5 = 0.34$ ;  $P = 0.46$ ) nor the number of individuals sampled per population across all 11 populations and colonies ( $r_9 = 0.12$ ;  $P = 0.72$ ).

Overall, observed heterozygosity (mean =  $0.15$ ) was significantly lower than expected heterozygosity ( $H_e$ ) (mean =  $0.18$ ) (paired  $t$ -test,  $t_{10} = 4.72$ ;  $P < 0.001$ , Table 1). The mean inbreeding coefficient ( $F_{IS}$ ) was higher for field populations ( $0.13$ ) than laboratory colonies ( $0.03$ ) ( $t_9 = 2.63$ ,  $P = 0.03$ ) (Table 1).

**Population structure.** AMOVA indicated that 25.6% of genetic variation was partitioned between populations and colonies, while 12.3 and 62.1% of variation was partitioned between and within individuals, respectively ( $P < 0.001$  in all cases; Table 2).

We observed clear genetic structuring of populations and colonies based on geography, with pairwise  $F_{ST}$  values greater than  $0.185$  for populations pertaining to Australia, U.S., and India-Pakistan. U.S. colonies were strongly genetically differentiated (pairwise  $F_{ST} = 0.273$ – $0.583$ ; Table 3), whereas Indian and Pakistani populations were genetically similar to one another (pairwise  $F_{ST} < 0.022$ ; Table 3). These findings were reinforced by the principal component analysis (PCA; Fig. 2) and maximum likelihood (ML) phylogeny (Fig. 3). The first two principal components in the PCA explained 26.6% of the total genetic variance in our dataset, and clearly demonstrated the distinction between individuals assigned to U.S. (blue shades), Australian (green), and India-Pakistan (red/orange/yellow shades) population and colonies (Fig. 2), with a bootstrap confidence value of 100 supporting the divergence of the three main groups in the ML phylogeny (Fig. 3). The PCA assigned U.S. individuals to three

Population	Country	State or province	Field sites	n	Ho	He	F <sub>IS</sub>
Aus-K	Australia	Western Australia	1	12	0.141	0.209	0.280
Ind-A	India	Andhra Pradesh	3	10	0.197	0.231	0.117
Ind-K	India	Karnataka	1	4	0.194	0.233	0.110
Ind-M	India	Maharashtra	3	9	0.195	0.230	0.121
Ind-T	India	Telangana	2	5	0.198	0.232	0.101
Pak-P	Pakistan	Punjab	5	14	0.199	0.236	0.129
Pak-S	Pakistan	Sindh	1	4	0.207	0.235	0.068
APHIS-S_1	U.S.	Arizona	NA	12	0.160	0.178	0.083
APHIS-S_2	U.S.	Arizona	NA	11	0.060	0.060	0.000
Bt4R	U.S.	Arizona	NA	4	0.080	0.089	0.055
Bt4-R2	U.S.	Arizona	NA	3	0.072	0.075	-0.015

**Table 1.** Location, number of field sites and individuals (n) sampled, observed heterozygosity (Ho), expected heterozygosity (He), and inbreeding coefficient (F<sub>IS</sub>) for 11 populations of pink bollworm.

Source	df	SS	MS	Est.Var	PV	P value
Between populations	10	79,799.7	7979.9	421.4	25.6	<0.001
Between individuals	77	107,396.5	1394.8	193.2	12.3	<0.001
Within individuals	88	88,741.7	1008.4	62.1	62.1	<0.001
Total	175	275,937.9	1576.8	1623.0	100.0	–

**Table 2.** Analysis of molecular variance (AMOVA) to assess variation between populations, and between and within individuals, for 11 populations of pink bollworm. *df* degrees of freedom, *SS* sum of squares, *MS* mean square, *Est. Var* estimated variance, *PV* percentage of variance.

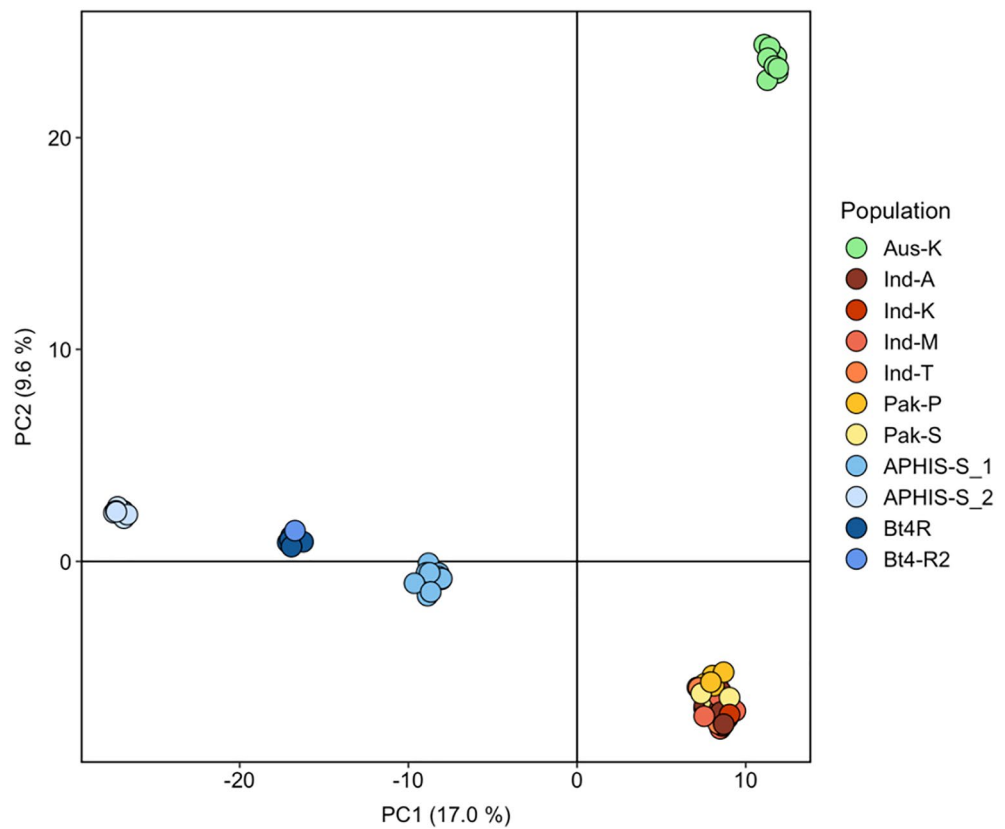
	Aus-K	Ind-A	Ind-K	Ind-M	Ind-T	Pak-P	Pak-S	APHIS-S_1	APHIS-S_2	Bt4R
Ind-A	0.260	–	–	–	–	–	–	–	–	–
Ind-K	0.259	0	–	–	–	–	–	–	–	–
Ind-M	0.259	0.020	0.014	–	–	–	–	–	–	–
Ind-T	0.255	0.001	0.003	0.007	–	–	–	–	–	–
Pak-P	0.243	0.016	0.011	0.012	0.006	–	–	–	–	–
Pak-S	0.253	0.022	0.021	0.018	0.016	0.003	–	–	–	–
APHIS-S_1	0.347	0.204	0.224	0.206	0.212	0.185	0.211	–	–	–
APHIS-S_2	0.550	0.450	0.549	0.458	0.518	0.408	0.541	0.393	–	–
Bt4R	0.431	0.312	0.369	0.318	0.351	0.287	0.365	0.273	0.558	–
Bt4-R2	0.423	0.310	0.366	0.313	0.337	0.282	0.356	0.327	0.583	0.327

**Table 3.** Pairwise F<sub>ST</sub> values for the 11 populations of pink bollworm listed in Table 1.

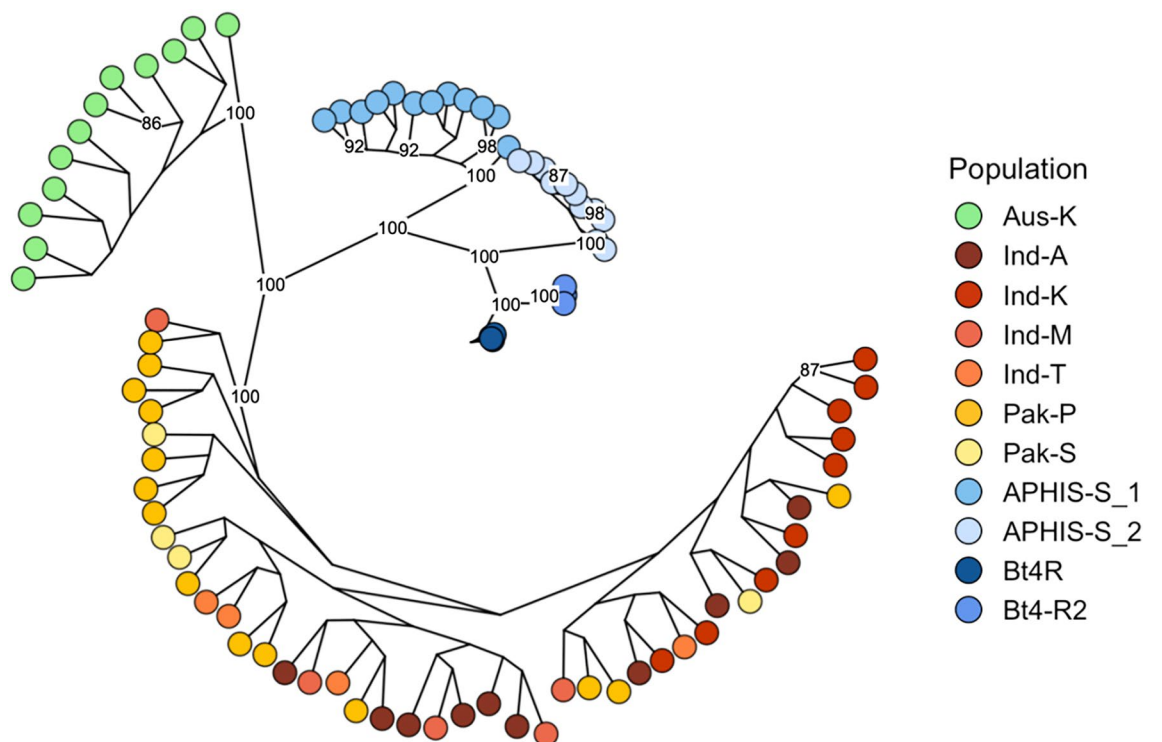
main clusters: (1) APHIS-S\_1, (2) APHIS-S\_2, and (3) Bt4R and Bt4-R2 (Fig. 2). ML further distinguished these groupings, placing each of the four U.S. colonies as genetically distinct with a bootstrap confidence value of 100 (Fig. 3). Consistent with pairwise F<sub>ST</sub> values, individuals from India and Pakistan lacked population structure, forming one genetic aggregate in the PCA plot and showing limited divergence in the ML phylogeny (Figs. 2, 3).

Sparse Non-Negative Matrix Factorisation (sNMF) analysis of the filtered dataset (6355 SNPs) indicated an optimal K value of five clusters (Fig. S2; see Fig. S3 for admixture proportions at K = 3, K = 4, and K = 6). These corresponded to Australia, India-Pakistan, and three groups for the U.S. colonies: (1) APHIS-S\_2 (light blue), (2) APHIS-S\_1 (medium blue), and (3) Bt4R and Bt4-R2 (dark blue) (Fig. 4). We found limited admixture, with each individual corresponding primarily to one genetic group (Fig. 4). The India-Pakistan populations showed the highest absolute admixture, with very small proportions (< ~5%) of shared ancestry with U.S. and Australian clusters. Meanwhile, Australia was highly isolated, sharing limited genetic ancestry or migration pathways with other populations (Fig. 4).

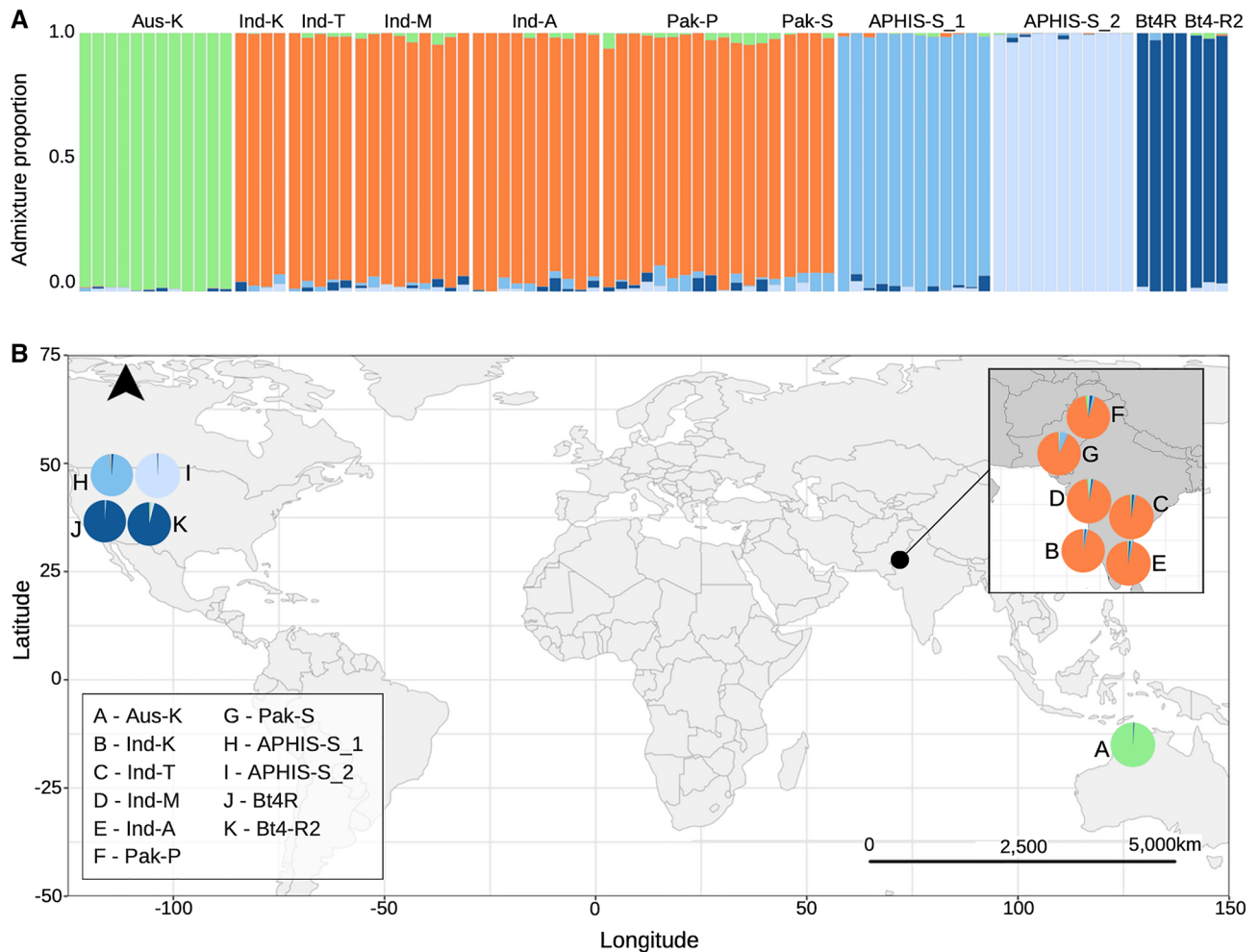
We identified 13 haplotypes from 31 COI sequences from GenBank, obtained from pink bollworm individuals in six countries: Australia (n = 3 individuals), India (n = 12), Pakistan (n = 5), the U.S. (n = 3), Israel (n = 4), and Kenya (n = 4). Nucleotide diversity was 0.002 in India (10 haplotypes), 0.0007 in Israel (two haplotypes), and 0 in the other four countries (one haplotype each) (Fig. 5). Haplotype A occurred in 16 individuals, including



**Figure 2.** Principal component analysis (PCA) of the filtered dataset of 6355 SNP loci for 11 populations of pink bollworm from Table 1.



**Figure 3.** Maximum likelihood tree for individuals from 11 populations of pink bollworm with bootstrap confidence values > 85 attached to nodes.



**Figure 4.** Admixture proportions for 11 populations of pink bollworm from Table 1. (A) Each bar represents the admixture coefficient for one individual. (B) Geographic distribution of populations with the admixture coefficient for each population represented by a pie chart. Analyses in (A) and (B) are based on five genetic clusters ( $K=5$ ).

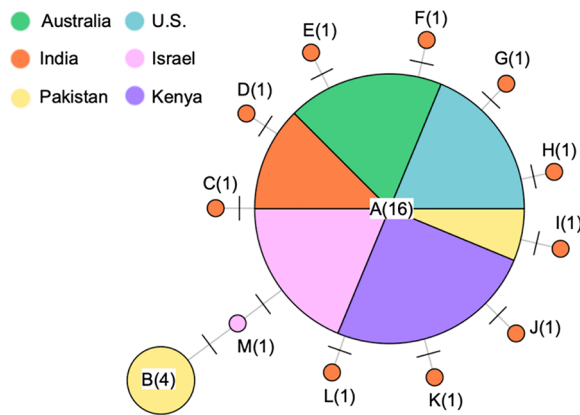
at least one individual from each country (Fig. 5). The second most frequent haplotype (B) occurred in four individuals from Pakistan, but not in any other country (Fig. 5). The other 11 haplotypes (C–M) occurred once each, with 10 from India (C–L) and one from Israel (M). Haplotype B differed by two mutational steps from A, whereas each of the less frequent haplotypes differed by one mutational step from A (Fig. 5).

## Discussion

We used population genomic analyses to assess population structure of seven wild pink bollworm populations from Australia, India, and Pakistan, and four laboratory colonies from the U.S. Overall, our results showed strong population differentiation between continents, including between a population in North West Australia and the other populations and colonies that were sampled in this study. This isolation reduces concerns about the threat of introduction of Bt-resistant insects from India and Pakistan into North West Australia, where the cotton industry is currently expanding.

Conflicting results were obtained in five previous studies using genetic markers to investigate pink bollworm population structure and genetic diversity. For example, using microsatellites<sup>5</sup>, relatively high genetic diversity and clear population structure were found among pink bollworm from China, Pakistan, and the U.S. However, low genetic variation and weak differentiation was later found for the same populations using mtDNA<sup>16</sup>. Among the other studies, high haplotype diversity within populations from Australia, China, India, Israel, Mexico, and the U.S. was found by analysing transposable element sequence data<sup>7</sup>, but low diversity was found in two studies that analysed mtDNA sequences from India<sup>6,17</sup>. These discrepancies may reflect differences between studies in the genetic markers used. Namely, mutation rate likely differs between the marker regions studied, and relative to nuclear DNA, mtDNA is more prone to genetic drift because of its maternal inheritance<sup>16,43,44</sup>. Whereas the previous studies of pink bollworm used one to 13 loci, we analysed genome-wide data for SNPs, which is expected to increase resolution and accuracy for understanding genetic processes<sup>45</sup>.

Our results showed that genetic diversity based on heterozygosity at 6355 SNP markers and nucleotide diversity at mtDNA COI was low within all populations and colonies. As expected, heterozygosity was significantly



**Figure 5.** Network of 13 haplotypes (A–M) based on 31 COI sequences from 31 pink bollworm from Australia, India, Pakistan, the U.S., Israel, and Kenya. Circles represent haplotypes, with the circle size proportional to the haplotype frequency and numbers in parentheses indicating the frequency of each haplotype. Lines connecting haplotypes represent mutations, with each dash across a line representing one mutational step between haplotypes.

lower within laboratory colonies from the U.S. compared to within field populations from Australia, India, and Pakistan. However, putative inbreeding ( $F_{IS}$ ) was higher in the field populations. This was surprising as laboratory colonies are expected to show greater  $F_{IS}$  because their smaller population sizes increase mating between related individuals<sup>46,47</sup>. However, a meta-analysis of self-incompatible plant species also found that  $F_{IS}$  was positively correlated with population size<sup>48</sup>, which is similar to our findings. Estimates of diversity and  $F_{IS}$  using SNP data may be influenced by missing data, rare alleles, sample size, and population structure<sup>49</sup>. However, in our study  $H_o$  was not correlated with the number of individuals nor the number of field sites sampled per population, and our SNP matrix had < 1.5% missing data. As noted by Schmidt et al.,<sup>49</sup> more work is needed to evaluate the implications of differences between observed and expected heterozygosity<sup>50</sup>.

Consistent with the microsatellite-based findings of Liu et al.<sup>5</sup>, we found strong genetic differentiation between populations and colonies of pink bollworm from different continents. We also found extremely low admixture between populations and colonies, indicative of low realised gene flow. In the previous studies, strong genetic differentiation, limited gene flow, and low heterozygosity in wild pink bollworm populations was attributed to potential genetic bottlenecks caused by larval mortality from Bt cotton, post-Pleistocene range expansion with limited founders, limited flight activity, and/or narrow host specificity<sup>5–7,16,17</sup>. Pink bollworm moths have been found at altitudes up to 1000 m and their dispersal up to 100 km is documented in the U.S.<sup>51,52</sup>. However, our results imply that inter-continental gene flow is negligible for the populations we studied.

Although pink bollworm is present in over 150 countries worldwide (Fig. 1), genomic data is unavailable for most of these populations. We analysed genomic data from samples obtained at least five to ten years ago from field populations in four to six countries, and four laboratory colonies from the U.S. (see “Methods”). Thus, more recent data from global populations would be valuable. In particular, it would be useful to investigate populations from Indonesia, Philippines, and Malaysia, as these may represent possible bridgehead intermediaries between India and Australia and may therefore be a potential avenue of spread for Bt-resistant alleles in the future<sup>53</sup>. Future studies might also reconstruct the demographic history of pink bollworm to advance understanding of invasion pathways, and apply selection-based analyses to investigate the evolutionary factors that underpin invasion success (e.g., the specific alleles involved in Bt-resistance)<sup>54</sup>. In such studies, the newly available reference genome for pink bollworm<sup>55</sup> and wider geographic sampling could yield insights facilitating better monitoring and management of this cosmopolitan pest.

### Data availability

The data generated and/or analysed during the current study are available in the Figshare digital repository: <https://doi.org/10.6084/m9.figshare.22871558.v1>.

Received: 20 January 2023; Accepted: 11 July 2023

Published online: 20 July 2023

### References

- Gippet, J. M. W., Liebhold, A. M., Fenn-Moltu, G. & Bertelsmeier, C. Human-mediated dispersal in insects. *Curr. Opin. Insect Sci.* **35**, 96–102. <https://doi.org/10.1016/j.cois.2019.07.005> (2002).
- Henneberry, T. & Naranjo, S. Integrated management approaches for pink bollworm in the southwestern United States. *Integr. Pest Manag. Rev.* **3**, 31–52. <https://doi.org/10.1023/A:1009673622862> (1998).
- CABI Compendium. *Pectinophora gossypiella* (pink bollworm). <https://doi.org/10.1079/cabicompendium.39417> (2021)
- Ingram, W. *Pectinophora* (Lepidoptera: Gelechiidae), insect pests of cotton (CAB International Publisher, 1994).
- Liu, Y. D., Wu, K.-M. & Guo, Y. Y. Population structure and introduction history of the pink bollworm, *Pectinophora gossypiella*, in China. *Entomol. Exp. Appl.* **130**, 160–172. <https://doi.org/10.1111/j.1570-7458.2008.00800.x> (2009).

6. Sridhar, J. *et al.* Population genetic structure of cotton pink bollworm, *Pectinophora gossypiella* (Saunders) (Lepidoptera: Gelechiidae) using mitochondrial cytochrome oxidase I (*COI*) gene sequences from India. *Mitochondrial DNA Part A* **28**, 941–948. <https://doi.org/10.1080/24701394.2016.1214727> (2017).
7. Wang, J., Simmons, G. S., Miller, T. A., Tabashnik, B. E. & Park, Y. Global variation in the *piggyBac*-like element of pink bollworm *Pectinophora gossypiella*. *J. Asia Pac. Entomol.* **14**, 131–135. <https://doi.org/10.1016/j.aspen.2010.09.006> (2011).
8. Holdaway, F. G. The pink bollworm of Queensland. *Bull. Entomol. Res.* **17**, 67–83. <https://doi.org/10.1017/S000748530001909X> (1926).
9. Common, I. A revision of the pink bollworms of cotton (*Pectinophora busck* (Lepidoptera: Gelechiidae) and related genera in Australia. *Aust. J. Zool.* **6**, 268–306. <https://doi.org/10.1071/ZO9580268> (1951).
10. Naumann, I. D. & Sands, D. P. A. Two Australian *Elasmus* spp. (Hymenoptera: Elasmidae), parasitoids of *Pectinophora gossypiella* (Saunders) (Lepidoptera: Gelechiidae): Their taxonomy and biology. *Aust. J. Entomol.* **23**, 25–32. <https://doi.org/10.1111/j.1440-6055.1984.tb01901.x> (1984).
11. Richards, K. Insect pests of cotton in the Ord river irrigation area. *J. Dep. Agric. West. Aust. Ser.* **4**(5), 79–86 (1964).
12. Fowler, C. Ord valley cotton industry hits new highs in WA's far north, as national production plummets. <https://www.abc.net.au/news/rural/2020-08-31/ord-cotton-industry-hits-new-highs-in-wa-far-north/12610730> (2020).
13. Tabashnik, B. E., Liesner, L. R., Ellsworth, P. C. & Carrière, Y. Transgenic cotton and sterile insect releases synergize eradication of pink bollworm a century after it invaded the United States. *Proc. Natl. Acad. Sci. USA* **118**, e2019115118. <https://doi.org/10.1073/pnas.2019115118> (2020).
14. Wan, P. *et al.* Hybridizing transgenic Bt cotton with non-Bt cotton counters resistance in pink bollworm. *Proc. Natl. Acad. Sci. USA* **114**, 5413–5418. <https://doi.org/10.1073/pnas.1700396114> (2017).
15. Tabashnik, B. E. & Carrière, Y. Global patterns of resistance to Bt crops highlighting pink bollworm in the United States, China, and India. *J. Econ. Entomol.* **112**, 2513–2523. <https://doi.org/10.1093/jee/toz173> (2019).
16. Liu, Y., Hou, M. & Wu, K. Genetic variation of mitochondrial DNA in Chinese populations of *Pectinophora gossypiella* (Lepidoptera: Gelechiidae). *Environ. Entomol.* **39**, 1344–1351. <https://doi.org/10.1603/EN09355> (2010).
17. Naik, V. C. B. *et al.* Evidence for population expansion of cotton pink bollworm *Pectinophora gossypiella* (Saunders) (Lepidoptera: Gelechiidae) in India. *Sci. Rep.* **10**, 4740. <https://doi.org/10.1038/s41598-020-61389-1> (2020).
18. Kilian, A. *et al.* Diversity arrays technology: A generic genome profiling technology on open platforms. *Methods Mol. Biol.* **888**, 67–89. [https://doi.org/10.1007/978-1-61779-870-2\\_5](https://doi.org/10.1007/978-1-61779-870-2_5) (2012).
19. Mathew, L. G. *et al.* ABC transporter mis-splicing associated with resistance to Bt toxin Cry2Ab in laboratory- and field-selected pink bollworm. *Sci. Rep.* **8**, 13531. <https://doi.org/10.1038/s41598-018-31840-5> (2018).
20. USDA. Eradication of pink bollworm. In *A proclamation by the secretary of agriculture of the United States of America*. USDA. <https://www.usda.gov/sites/default/files/documents/usda-pink-bollworm-proclamation.pdf> (2018).
21. Liu, Y. *et al.* Effects of Bt cotton and Cry1Ac toxin on survival and development of pink bollworm (Lepidoptera: Gelechiidae). *J. Econ. Entomol.* **94**, 1237–1242. <https://doi.org/10.1603/0022-0493-94.5.1237> (2001).
22. Bartlett, A. C. Resistance of the pink bollworm to Bt transgenic cotton. Beltwide Cotton Conferences (USA). In *Beltwide Cotton Conferences (USA)* (1995).
23. Fabrick, J. A., Forlow Jech, L. & Henneberry, T. J. Novel pink bollworm resistance to the Bt toxin Cry1Ac: Effects on mating, oviposition, larval development and survival. *J. Insect Sci.* **9**, 24. <https://doi.org/10.1673/031.009.2401> (2009).
24. Fabrick, J. A. *et al.* Multi-toxin resistance enables pink bollworm survival on pyramided Bt cotton. *Sci. Rep.* **5**, 16554. <https://doi.org/10.1038/srep16554> (2015).
25. Bartlett, A. C. & Wolf, W. W. *Pectinophora gossypiella*. In *Handbook of insect rearing* (eds Singh, P. & Moore, R. F.) 415–430 (Elsevier, 1985).
26. Georges, A. *et al.* Genomewide SNP markers breathe new life into phylogeography and delimitation for the problematic short-necked turtles (Chelidae: *Emydura*) of eastern Australia. *Mol. Ecol.* **27**, 5195–5213. <https://doi.org/10.1111/mec.14925> (2018).
27. R Core Team. R: A language and environment for statistical computing. R Foundation for statistical computing, Vienna, Austria. <https://www.R-project.org/> (2021).
28. Jombart, T. adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405. <https://doi.org/10.1093/bioinformatics/btn129> (2008).
29. Gruber, B., Unmack, P. J., Berry, O. F. & Georges, A. dartr: An R package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Mol. Ecol. Resour.* **18**, 691–699. <https://doi.org/10.1111/1755-0998.12745> (2018).
30. Gosselin, T. Radiator: RADseq data exploration, manipulation and visualization using R. R package version 1.1.9. <https://thierrygosselin.github.io/radiator/> (2020).
31. Goudet, J. hierfstat, a package for r to compute and test hierarchical F-statistics. *Mol. Ecol. Notes* **5**, 184–186. <https://doi.org/10.1111/j.1471-8286.2004.00828.x> (2005).
32. Kamvar, Z. N., Tabima, J. F. & Grünwald, N. J. Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* **2**, e281. <https://doi.org/10.7717/peerj.281> (2014).
33. Dray, S. & Dufour, A. B. The ade4 package: Implementing the duality diagram for ecologists. *J. Stat. Softw.* **22**, 1–20. <https://doi.org/10.18637/jss.v022.i04> (2007).
34. Weir, B. S. & Cockerham, C. C. Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370. <https://doi.org/10.2307/2408641> (1984).
35. Weir, B. S. & Goudet, J. A unified characterization of population structure and relatedness. *Genetics* **206**, 2085–2103. <https://doi.org/10.1534/genetics.116.198424> (2017).
36. Fricot, E. & François, O. LEA: An R package for landscape and ecological association studies. *Methods Ecol. Evol.* **6**, 925–929. <https://doi.org/10.1111/2041-210X.12382> (2015).
37. South, A. rworldmap: A new R package for mapping global data. *R J.* **3**, 35–43 (2011).
38. Wickham, H. *ggplot2: Elegant graphics for data analysis* (Springer, 2016).
39. Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274. <https://doi.org/10.1093/molbev/msu300> (2015).
40. Paradis, E. & Schliep, K. ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528. <https://doi.org/10.1093/bioinformatics/bty633> (2018).
41. Tamura, K., Stecher, G. & Kumar, S. MEGA11: Molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* **38**, 3022–3027. <https://doi.org/10.1093/molbev/msab120> (2021).
42. Paradis, E. pegas: An R package for population genetics with an integrated–modular approach. *Bioinformatics* **26**, 419–420. <https://doi.org/10.1093/bioinformatics/btp696> (2010).
43. Song, W. H., Ballard, J. W., Yi, Y. J. & Sutovsky, P. Regulation of mitochondrial genome inheritance by autophagy and ubiquitin-proteasome system: Implications for health, fitness, and fertility. *Biomed Res. Int.* **2014**, 981867. <https://doi.org/10.1155/2014/981867> (2014).
44. Xu, S. *et al.* High mutation rates in the mitochondrial genomes of *Daphnia pulex*. *Mol. Biol. Evol.* **29**, 763–769. <https://doi.org/10.1093/molbev/msr243> (2012).
45. North, H. L., McGaughan, A. & Jiggins, C. D. Insights into invasive species from whole-genome resequencing. *Mol. Ecol.* **30**, 6289–6308. <https://doi.org/10.1111/mec.15999> (2021).



46. Briscoe, D. A. *et al.* Rapid loss of genetic variation in large captive populations of *Drosophila* flies: Implications for the genetic management of captive populations. *Biology* **6**, 416–425. <https://doi.org/10.1046/j.1523-1739.1992.06030416.x> (1992).
47. Ross, P. A., Endersby-Harshman, N. M. & Hoffmann, A. A. A comprehensive assessment of inbreeding and laboratory adaptation in *Aedes aegypti* mosquitoes. *Evol. Appl.* **12**, 572–586. <https://doi.org/10.1111/eva.12740> (2019).
48. Leimu, R., Mutikainen, P. I. A., Koricheva, J. & Fischer, M. How general are positive relationships between plant population size, fitness and genetic variation?. *J. Ecol.* **94**, 942–952. <https://doi.org/10.1111/j.1365-2745.2006.01150.x> (2006).
49. Schmidt, T. L., Jasper, M.-E., Weeks, A. R. & Hoffmann, A. A. Unbiased population heterozygosity estimates from genome-wide sequence data. *Methods Ecol. Evol.* **12**, 1888–1898. <https://doi.org/10.1111/2041-210X.13659> (2021).
50. McDonald, R. E. & Loftin, U. C. Dispersal of the pink bollworm by flight or wind carriage of the moths. *J. Econ. Entomol.* **28**, 745–755. <https://doi.org/10.1093/jee/28.5.745> (1935).
51. Bariola, L. A., Keller, J. C., Turley, D. L. & Farris, J. R. Migration and population studies of the pink bollworm 1 in the Arid West. *Environ. Entomol.* **2**, 205–208. <https://doi.org/10.1093/ee/2.2.205> (1973).
52. Tabashnik, B. E. *et al.* Dispersal of pink bollworm (Lepidoptera: Gelechiidae) males in transgenic cotton that produces a *Bacillus thuringiensis* toxin. *J. Econ. Entomol.* **92**, 772–780. <https://doi.org/10.1093/jee/92.4.772> (1999).
53. Bertelsmeier, C., Ollier, S., Liebhold, A. M. & Keller, L. Recurrent bridgehead effects accelerate global alien ant spread. *Proc. Natl. Acad. Sci. USA* **115**, 5486–5491. <https://doi.org/10.1073/pnas.1801990115> (2018).
54. Estoup, A. & Guillemaud, T. Reconstructing routes of invasion using genetic data: Why, how, and so what?. *Mol. Ecol.* **19**, 4113–4130. <https://doi.org/10.1111/j.1365-294X.2010.04773.x> (2010).
55. Stahlke, A. R. *et al.* Chromosome-scale genome assembly of the pink bollworm, *Pectinophora gossypiella*, a global pest of cotton. *G3 Genes Genomics Genet.* **13**, jkad040. <https://doi.org/10.1093/g3journal/jkad040> (2023).

## Acknowledgements

We thank Jeyakumar Ponnuraj (Indian Council of Agricultural Research, Indian Institute of Rice Research, Hyderabad, India), Bheemanna Mallappa (University of Agricultural Sciences, Karnataka, India), and Lingutla R. Chowdary (Ranga Agricultural University, Andhra Pradesh, India), and Shahid Mansoor and Muhammad Asif (both from Agricultural Biotechnology Division, National Institute for Biotechnology and Genetic Engineering, College Pakistan Institute of Engineering and Applied Sciences, Faisalabad, Pakistan) for providing pink bollworm samples from India and Pakistan, respectively. Rebecca Clarke from Raitech collected pink bollworm samples Kununurra, Australia, and Tim Hogarty extracted DNA from the samples at Commonwealth Scientific Industrial Research Organisation, Canberra. This research was supported by the Centre for Biodiversity Analysis at the Australian National University (Ignition Grant to A.M. and T.W.), and in part by Commonwealth Scientific Industrial Research Organisation and the U.S. Department of Agriculture (USDA) Agricultural Research Service, National Program 304—Crop Protection & Quarantine (CRIS Project #2020-22620-022-00D). Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal opportunity provider and employer.

## Author contributions

A.M. and T.W. wrote the original grant application and A.M. designed the eventual analysis plan, with J.F., B.T., H.A.S., and T.W. collecting or facilitating sample collection. P.M. analysed the data with assistance as required from A.M. and E.P., and created all figures. P.M. wrote the first draft of the manuscript and all authors provided feedback, with P.M. leading subsequent revision of the final 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/s41598-023-38504-z>.

**Correspondence** and requests for materials should be addressed to P.M.

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

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



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

© The Author(s) 2023