Genetic hitch-hiking and resistance evolution to transgenic *Bt* toxins: insights from the African stalk borer *Busseola fusca* (Noctuidae).

AUTHORS: Pascal Campagne[1,2,3], Claire Capdevielle-Dulac[1], Rémy Pasquet[1], Stephen J. Cornell[3], Marlene Kruger[4], Jean-François Silvain[1], Bruno Le Rü[1,2], Johnnie Van den Berg[4]

Affiliations:

[1]Unité de Recherche IRD 072, CNRS UPR9034, Laboratoire Évolution, Génome et Spéciation, Gif-sur-Yvette, France – Université Paris-Sud 11, Orsay, France.

[2]Noctuid Stem Borers Biodiversity in Africa Project, Environmental Health Division, International Centre for Insect Physiology & Ecology, PO Box 30772-00100 Nairobi, Kenya

[3]Institute of Integrative Biology, University of Liverpool, Liverpool, L69 7ZB, UK.

[4]Unit of Environmental Sciences and Management, North-West University, Potchefstroom, 2520, South Africa

Corresponding author: Pascal Campagne

Institute of Integrative Biology,
Biosciences Building,
University of Liverpool, Liverpool L69 7ZB, Liverpool, UK.
Email : pascal.campagne@liv.ac.uk

Running title: Population genetics of a *Bt* resistant stem borer

Word count: 6123

Abstract

Since transgenic crops expressing *Bacillus thuringiensis* (*Bt*) toxins were first released, resistance evolution leading to failure in control of pests populations has been observed in a number of species. Field resistance of the moth *Busseola fusca* was acknowledged eight years after *Bt* maize was introduced in South Africa. Since then, field resistance of this corn borer has been observed at several locations, raising questions about the nature, distribution and dynamics of the resistance trait.

Using genetic markers, our study identified four outlier loci clearly associated with resistance. In addition, genetic structure at neutral loci reflected extensive gene flow among populations. A realistically parameterised model suggests that resistance could travel in space at a speed of several kilometres a year.

Markers at outlier loci delineated a geographic region associated with resistance spread. This was an area of approximately 100 km radius, including the location where resistance was first reported. Controlled crosses corroborated these findings and showed significant differences of progeny survival on *Bt* plants depending on the origin of the resistant parent.

Last, our study suggests diverse resistance mutations, which would explain the widespread occurrence of resistant larvae in *Bt* fields across the main area of maize production in South Africa.

Keywords: AFLP markers, genome scan, *Bt* maize, spatial distribution, resistance spread, controlled crosses.

INTRODUCTION

Adaptation to human driven environmental pressures present insightful illustrations of the various facets of contemporary evolution and their implications. The evolution of resistance, the process by which a population of organisms acquires a genetically-based decrease of susceptibility to a toxic compound (Tabashnik *et al*., 2009), may result in spectacularly fast responses to strong selective pressures (e.g., Mallet, 1989). Genetically modified crops, expressing *Bacillus thuringiensis* (*Bt*) toxins, are expected to result in elevated and continuous selection whenever *Bt* crops dominate the landscape (Tabashnik *et al*., 2013).

Evolution of resistance to *Bt* toxins expressed by crops is a direct threat to sustained control of pest populations, which may seriously impact agricultural yields. Even if a history of 10−20 years has revealed the first empirical trends (Tabashnik *et al*., 2013), the questions of how fast resistance to *Bt* crops toxins is likely to evolve and how quickly it may spread remain challenging. While *Bt* crops have been deployed for almost two decades, loss of control in pest populations, due to field-evolved resistance, arose in a few pest species only. However, decrease of susceptibility to transgenic *Bt* toxins (field-evolved resistance), which has not led to control failure, has been observed in many species (Tabashnik *et al*., 2013). Empirical knowledge about the spatial dynamics of resistance spread is scarce. Extensive gene flow is commonly observed in pest populations (e.g., Daly, 1985; Korman et al. 1993; Bourguet *et al*., 2000; Endersby *et al*., 2006; Kim *et al*., 2009) and may promote rapid diffusion of resistance alleles across space (e.g., Peck *et al*., 1999, Bourguet *et al*., 2000).

The main option for delaying resistance evolution is the refuge strategy, which amounts to planting non-*Bt* plants within or surrounding *Bt*-fields, allowing the survival of some

susceptible larvae in a *Bt*-dominated environment. The proportion of refuge crop required to effectively delay resistance evolution is strongly dependent on the recessiveness of the resistance trait and, more generally, on the survival of heterozygote-resistant individuals at the dose of toxin expressed in plant tissues (e.g., see Vacher *et al*., 2003; Vacher *et al*., 2004).

The nature and uniformity of the genetic bases of resistance to *Bt* toxins are thus crucial aspects influencing resistance evolution and management (e.g., Tabashnik *et al*., 1998). Multiple recessive mutations that cause resistance to *Cry*1A toxins in Lepidoptera have been reported (see Heckel *et al*., 2007) in genes encoding Cadherin-like proteins (e.g., Morin *et al*., 2003) and ABC transporters (e.g., Baxter *et al*., 2011). In the field, diverse alleles conferring resistance were identified within Chinese populations of *Helicoverpa armigera* (Zhang *et al*., 2012), including recessive resistance alleles at the Cadherin locus as well as two non-recessive mutations, one of which was not linked to this locus.

Strong selection acting at a particular resistance locus is expected to affect not only the frequency of the selected mutation but also flanking polymorphisms. Alleles present on the ancestral chromosome on which a favourable mutation arose will tend to increase in frequency as the mutation invades the population, a consequence which is known as the genetic hitch-hiking effect (Kojima and Schaffer, 1967; Maynard Smith and Haigh, 1974). Selective shifts may therefore give rise to strong linkage disequilibrium, i.e, some non-random allelic associations at the population level, between loci of the genomic region under selection.

Various population genetics approaches have aimed at taking advantage of such phenomena to capture signatures from positive selection (Jensen, 2014), notably in non-model organisms (e.g., de Villemereuil and Gaggiotti, 2015). The size of a hitch-hiking effect depends on several factors, including: age and frequency of the mutation under selection; the intensity of

100 selection; and recombination distance to the locus under selection. In particular, the initial
101 frequency of the selected allele in the population prior to a selective shift, as well as the age of
102 a mutation, play a crucial role (Maynard Smith and Haigh, 1974; Jensen, 2014). As a
103 consequence, the effectiveness of genome scans is lowered when for example: populations
104 have undergone a particular demographic history (e.g., bottleneck); a recessive allele is
105 involved; or selection acts on an ancient and previously neutral allele (Teshima *et al*., 2006;
106 Jensen, 2014).

107 In South Africa, the pest *Busseola fusca* (Lepidoptera, Noctuidae) evolved resistance to the *Bt*
108 toxin *Cry*1Ab expressed by transgenic maize within 8 years of the technology being released
109 (Van Rensburg, 2007). While this pest species is seldom hosted by wild grasses (LeRü *et al*.,
110 2006), it is responsible for substantial yield losses in maize, sorghum (Kfir *et al*., 2002), and
111 more marginally sugarcane crops (Assefa *et al*., 2015). Being endemic to Sub-Saharan Africa
112 and widely distributed throughout the region (Dupas *et al*., 2014), *B*. *fusca* is considered a
113 pest of economic importance. To date, field resistance in *B*. *fusca* remains one of the few
114 cases in the world where *Bt* crop failure has clearly been documented (see Tabashnik *et al*.,
115 2013). Resistance evolution in this species was initially ascribed to an insufficient
116 implementation of the refuge strategy (Kruger *et al*., 2012, Van den Berg *et al*., 2013).
117 Nonetheless, the existence of a dominant resistance trait, such as found in this pest
118 (Campagne *et al*., 2013), is expected to have greatly accelerated the spread of resistance and
119 to have drastically reduced refuge efficiency. Furthermore, the presumed low survival of
120 resistant individuals on non-*Bt* plants is unlikely to have mitigated the fitness benefits
121 conferred by the resistance trait since no fitness cost associated with resistance was found in
122 *B*. *fusca* (Kruger *et al*., 2014).

123 Four years after being first recorded in the field in 2006, resistance in *B*. *fusca* was observed
124 at a distance of ~40 km from the location reported initially (Kruger *et al*., 2011). By 2011, *Bt*

maize fields infested by *B. fusca* were observed in various places up to ~400 km distance from the initial location (see van den Berg *et al*., 2013). The latter sequence of observations raises a number of puzzling questions regarding both the initial distribution and spatial dynamics of resistance traits to *Bt*-maize in this species. While gene flow may either swamp local resistance evolution or contribute to the spread of resistance alleles (Bourguet *et al*., 2000), little is known about the levels of migration and gene flow taking place in populations of *B. fusca* at sub-regional scales.

Based on empirical genetic data collected in the field, controlled crosses, and theoretical considerations, the purpose of this study was to investigate dynamics and diversity of resistance evolution in field populations of *B. fusca*. Our aims were to: (i) to evaluate the extent of gene flow in this pest within and between sites distributed across the major region of maize cultivation in South Africa; (ii) to identify markers associated with resistance and further investigate spatial distribution of resistance traits; (iii) to explore the geographic variability of resistance inheritance in *B. fusca* within the study region.

MATERIAL AND METHODS

Study context and sampling

The first official observation of field resistance in *B. fusca* was recorded in the Christiana area (Van Rensburg, 2007; Kruger *et al*., 2011) which is located at the western extremity of the main maize production area (Van den Berg *et al*., 2013). Available data indicate, between 1998/1999 and 2010/2011, an average proportion of 60% *Bt* maize in the area where

148 resistance was first reported (Kruger et al. 2012). The landscape is composed of 20% crops
149 (the main one being maize) while most of the matrix consists of semi-natural or natural
150 habitats (source: FAO GeoNetwork, www.fao.org). In this area, populations of *B. fusca* are
151 bivoltine with a diapause stage between cropping seasons. Sampling was performed during
152 the growing season 2010/2011, i.e., approximately $k \approx 26$ generations after *Bt* maize was
153 released, across the main region of maize production in South Africa (stretching across more
154 than 600 km East-West).

155 In order to carry out population genetics analyses, 20 to 25 larvae per field were sampled in a
156 total of 28 fields (14 non-*Bt* fields and 14 *Bt* fields, Table 1). Sampling was performed within
157 maize fields at advanced stage of development (late vegetative stage or higher), and consisted
158 in 5 collection points distributed along a line, separated by approximately 50 m. At each
159 point, we collected 5 larvae originating from 5 different plants separated by a few metres from
160 each other. Care was taken to sample larvae in their 4^{th} instar at least, in order to minimize the
161 number of susceptible larvae collected in *Bt* fields. Larvae originating from different plants
162 were conserved separately in 90% ethanol. Finally, a crude assessment of field damage level
163 was performed in most sampled fields by recording the number of plants with leaf damage in
164 3 separate rows chosen at random (at a distance > 20 m from each other, depending on an
165 arbitrary number of steps) of 100 plants each (see Table 1). In order to perform controlled
166 crosses, an additional number of larvae was collected in *Bt* fields at sites where resistance had
167 been reported.

168

169 **Genetic markers**

170 Total DNA isolation of 20 mg of larvae thoraces was done using the DNeasy Tissue Kit
171 (Qiagen). DNA concentration was determined using a Nanodrop (Thermo Scientific). The

AFLP reaction was performed as described by Vos *et al*. (1995) with slight modifications. A quantity of 300 ng of DNA was digested using the following restriction enzymes: EcoRI and PstI. Digestion was made overnight at 37°C with 5 units of EcoRI (New England Biolabs), and 5 units of PstI (New England Biolabs) in a total volume of 25 µL. Ligation of double-stranded EcoRI and PstI adaptors to the ends of the restriction fragments was performed (for 8 h at 15°C) by adding 10 pmol of EcoRI and PstI adaptors, 5 units T4 DNA ligase (Promega). A 5 µL volume of 10X diluted ligation products were used as a template for preamplification, using 10 pmol of EcoRI and MseI primers, 0.4 mM dNTPs, 1.5 mM $MgCl_2$ and 1 unit of GoTaq Flexi DNA Polymerase (Promega) in a final volume of 25 µL. The preamplification thermocycle profile was: 94°C for 5 min, followed by 20 cycles at 94°C for 30 s, 56°C for 1 min, 72°C for 1 min and 72°C for 5 min. Selective amplification steps were performed using 5 pmol EcoRI (+NN) and PstI (+NN) primers with 2.5 µL of 20X diluted preamplification product in a final volume of 12.5 µL. Each selective amplification reaction mixture contained 0.4 mM dNTPs, 1.5 mM MgCl and 0.5 unit of GoTaq Flexi DNA Polymerase (Promega). The selective amplification thermocycler profile was: 94°C for 2 min, followed by 12 cycles at 94°C for 30 s, 65°C for 30 s, 72°C for 1 min, followed by 23 cycles at 94°C for 30 s, 56°C for 30 s, 72°C for 1 min, and 72°C for 5 min. The different primers combinations used were : EcoRI+AA/PstI+AG, EcoRI+AG/PstI+AG, EcoRI+GA/PstI+AT, EcoRI+AG/PstI+AT, EcoRI+GA/PstI+AG, EcoRI+CA/PstI+AA, EcoRI+CT/PstI+AT, EcoRI+TG/PstI+GA, and EcoRI+GG/PstI+AC. AFLP markers were scored according to the absence/presence of peaks using Genemapper V4.1 (Life Technologies). In order to optimise the reliability of the markers, 65 individuals were processed twice. Loci characterized by more than 3 discrepancies (5%) were considered prone to genotyping errors and were discarded. A total of 629 polymorphic markers were retained for further statistical analysis. The overall genotyping

196 error rate per locus (see Bonin et al. 2004), calculated as the ratio between observed number
197 of differences and total number of comparisons, was 1.3%.

198

199 **Data analysis**

200 The inbreeding coefficient F_{IS} was estimated at both population and individual levels, for each
201 sampled field using the software I4A which is designed to deal with dominant genetic
202 markers (Chybicki *et al*., 2011). After testing the convergence of different run lengths, data
203 was analyzed using 10,000 burn-in and 50,000 sampling iterations considering flat priors for
204 the inbreeding coefficients. Estimates were provided with 95%-credibility intervals.

205 Rapid resistance evolution is expected to have left specific signatures in the genome of *B*.
206 *fusca* as directional selection imposed by *Bt* toxins may result in increased differentiation
207 between populations at hitch-hiking loci. Using Bayescan (see Foll and Gaggiotti, 2008), we
208 scanned the genome of this pest for loci at which genetic differentiation between populations
209 (F_{ST}) is significantly higher than the genome baseline (outlier loci). Two sets of markers were
210 defined and analysed separately: non-outlier loci, expected to reflect neutral processes such as
211 genetic drift and gene flow between populations; and outlier loci, further tested for being
212 associated with resistance.

213 In order to analyse the set of non-outlier loci, Euclidean genetic distances between individuals
214 were computed (Peakall and Smouse, 2012) and genetic differentiation (Φ_{ST}) among fields
215 was estimated with an Analysis of Molecular Variance (AMOVA, Excoffier *et al*., 1992;
216 Peakall and Smouse, 2012). Statistical significance was tested using 999 permutations. A
217 between-field analysis was performed on Principal Coordinate Analysis (PCO) axes (ade4
218 package, R Core Team, 2012). In addition, pairwise kinship coefficients (F_{ij}) were estimated
219 using the software SPAGeDi (Hardy et and Vekemans, 2002). Estimation of kinship

220 coefficients based on dominant markers requires an externally supplied estimate of the
221 inbreeding coefficient. As the vast majority of individuals did not exhibit evidence of
222 inbreeding (see Results section) we used $F = 0$. Average kinship was then calculated within
223 and between fields. The genetic neighbourhood size (N_b*), a measure of the local extent of
224 breeding, was estimated following Hardy's (2003) procedure: $N_b^* \approx (F_0 - 1) / b$, where F_0 is
225 the average kinship for adjacent individuals (within a field) and b is the slope of the linear
226 decay between kinship and the natural logarithm of distance. Pairs of geographic distances
227 higher than 100 km were excluded (see Hardy, 2003) and a confidence interval was calculated
228 using 999 bootstrap iterations. N_b* is an empirical estimate of the theoretical demographic
229 parameter $N_b = 4\pi\sigma\,\rho_e$, where ρ_e denotes the effective population density (e.g., see Rousset
230 1997; Hardy, 2003) and σ (km per generation) is the axial variance of an isotropic Gaussian
231 dispersal distribution (2σ being the mean squared parent-offspring-displacement).

232 Association between field resistance and the genetic structure at outlier loci was then
233 investigated. Analyses of population clustering were performed exclusively on individuals
234 which were sampled in *Bt* fields. Ward's method for hierarchical clustering based on pairwise
235 genetic distances (Peakall and Smouse, 2012) and the software STRUCTURE 2.3.4 (Pritchard et
236 al. 2000) were used to identify genetic clusters. The latter analysis was executed under the
237 admixture model and correlated allele frequencies, with 10,000 burn-in iterations and 50,000
238 main iterations. The optimal number of genetic clusters (K) was determined following the
239 procedure described in Evanno *et al*. (2005), with 5 replicates for each K-value. For the sake
240 of comparison, the same analysis was performed using the set of non-outlier loci.

241 The difference in marker frequency at outlier loci between *Bt* and non-*Bt* fields was tested
242 using Generalised Linear mixed Models (Venables and Ripley, 2002) with spatially correlated
243 random effects, assuming a binomial distribution. Crop type (i.e., *Bt* vs. non-*Bt*) was treated
244 as a fixed effect and the geographic coordinates of the fields were considered as random

components of the analysis. Correlations between marker frequencies at outlier loci were evaluated with Pearson's correlation coefficient (R Core Team, 2012). Linkage disequilibrium between outlier loci was tested following Hill's procedure (1974), which provides a parametrically explicit Maximum-Likelihood computation method for dominant markers. Both a standardized estimate of linkage disequilibrium (r) and its associated $\log_{10}$-likelihood ratio (LR) were calculated.

Controlled crosses

Heritability of resistance traits provides insights into the genetic bases of resistance. In order to explore geographic variations in larvae survival on *Bt* maize, we performed single-pair crosses between resistant parents and a strain of *a priori* susceptible larvae originating from Tanzania (Tz), where no *Bt* crops are grown. Three sites, where resistance had been reported, were selected: Valhaarts (Va), an irrigation scheme close (~50 km) to the location where resistance was first reported and in which resistance has been previously studied (Kruger *et al*., 2011); Lichtenburg (Li) at 180–200 km distance from the initial location; and Bethleem (Be), at ~300 km from the initial location (see also Fig. 1). Controlled crosses were performed between adults originating from *Bt* fields and susceptible individuals: Va × Tz; Li × Tz; and Be × Tz. Progeny survival was assessed in a greenhouse on *Bt* and non-*Bt* maize planted in pots. Maize plants were inoculated with 20 to 30 neonates, which were gently transferred with a paintbrush within the whorl of young maize plants (i.e., from the 5th leaf collar stage). Plants were contained in fine-meshed nets to avoid larvae escaping. The number of individuals surviving on a plant was recorded after a period of 21 days. Preliminary observations indicate that survival after 2 weeks on *Bt* maize of individuals originating from Tanzania, is typically ≤ 2%. A fraction of the offspring were reared on non-*Bt* maize as a control. About 1700 larvae

were used in the experiment. Survival of the progeny on maize was compared using Fisher's exact tests (R Core Team, 2012). Contingent upon emergence of F1 adults and various experimental constraints, sib-crosses were undertaken in order to test whether homozygosity could reinforce resistance in F2 offspring of selected parents. One sib-cross was successfully performed among individuals originating from the Li × Tz cross and which survived on *Bt* maize.

Theoretical resistance spread

One approach for understanding the geographic distribution of resistance is to model the spatial spread of resistance alleles. In our survey, larvae were found to survive on *Bt* maize at various locations which also correspond to sites where resistance was reported (Van den Berg *et al*., 2014), translating into a vast expanse of approximately 200 km × 300 km. Assuming relatively simple genetic bases of resistance, the most elementary scenarios consistent with these observations could include: (i) a rapid propagation of a unique mutation conferring resistance from an initial location; or (ii) ancient and widespread resistance mutation(s), prior to the selective shift. We explored the feasibility of a dominant mutation spreading across space, from an initially confined distribution, under the selective pressure of *Bt* maize. Resistance was considered to involve a single locus, with a resistance allele (R) of frequency q, and a susceptible allele (S) of frequency $1 - q$. For modelling purposes, the spatial distribution of *Bt* and non-*Bt* plants is considered continuous and random. The proportion of *Bt* crop in the landscape (ω) is assumed to determine the survival probability (e.g., see Tabashnik *et al*., 2008; Jin *et al*., 2015) of the three genotypes (RR, RS and SS): $W_{SS} = 1 - \omega$, $W_{RS} = (1 - \omega) + h \times \omega$ and $W_{RR} = 1$, where h denotes the dominance of the resistance allele. Spatial spread of a resistance allele under selection was considered under a demo-genetic

system of deterministic reaction-diffusion equations proposed by Tyutyunov *et al*. (2008) and slightly modified to our purpose (see Appendix S1). The reaction term of the model accounts for local increase of resistance allele frequency, in relation to ω, h, W_{SS}, W_{RR}, q, and, the birth rate β per individual and per generation. The diffusion term captures the effect of individual dispersal under diffusive Brownian motion (diffusion constant D), scaling to $2\sigma = 4D$ after one generation.

RESULTS

Genetic structure at non-outlier loci

The five primer combinations yielded 629 polymorphic and reproducible fragments in *B. fusca*, for the entire set of 693 individuals. All individuals had different AFLP profiles, differentiated by at least 8 markers. Estimates of population inbreeding coefficient were low, ranging between 0.007 and 0.058 in the 28 sampled fields (Table 1), indicating no strong heterozygote deficiency. Similarly, estimates of median individual inbreeding were < 0.01 in approximately 97% of individuals. The outlier analysis carried out with all 28 fields detected 26 loci characterized by $F_{ST} > 0.07$ (logarithm of the posterior odds, $\log_{10} PO > 1$) while the non-outliers were characterised by a lower differentiation $F_{ST} = 0.027$ (standard deviation = 0.008). On this basis, loci were split into two groups: "outlier" and "non-outlier" loci and were further analyzed separately.

Evidence for extensive gene flow between populations was found when investigating the genetic structure of non-outlier loci. Overall genetic differentiation among fields was low but

316 non-null (Φ_{ST} = 0.03, P < 0.001). Even at long distances (> 400 km), average pairwise
317 genetic differentiation between sampled fields was weak ($\Phi_{ST} \approx 0.04$). In spite of a lack of
318 genetic differentiation at broad spatial scales, a smooth pattern of isolation by distance was
319 found. Firstly, pairwise Φ_{ST} were associated with geographic distances (R = 0.47; P < 0.001).
320 Similarly, kinship coefficients were significantly correlated with log-transformed distances (R
321 = 0.52; P < 0.001; Fig. 2a). Parameters of the latter regression yielded the estimate for the
322 neighbourhood size, $N_b^* \approx (F_0 - 1) / b = 706$ with a 95% bootstrap interval [526, 1010],
323 which translates into $\rho_e.\sigma \approx 55$ (i.e., with $N_b^* \approx 4\pi\sigma\rho_e$), which is well above the theoretical
324 threshold of $\rho_e.\sigma > 1$ at which a population distributed in a continuous space behaves nearly
325 as a panmictic population (see Maruyama 1972). Secondly, a between-class (i.e., fields) PCO
326 (Fig. 2b) revealed congruent mapping between B-PCO axes and the geographical origin of the
327 field-collected populations (see Fig. 1 for the locations); e.g., clusters of fields {5,6,7},
328 {2,3,4} and {54,64,68,71,75}. The small amount of variation captured by the analysis (6.7%
329 of the total inertia; 1.7% onto axes 1 and 2) consistently reflected low levels of genetic
330 structure.

331

332 **Outlier loci and resistance**

333 The genetic structure of *Bt* fields at outlier loci revealed two spatially segregated groups of *Bt*
334 fields, henceforth A_1 and A_2 (Fig. 3). The optimal number of clusters obtained using a
335 Bayesian clustering method (STRUCTURE) was K = 2 in both sets of loci; higher values of K
336 did not result in better fits of the model (see Appendix S2). While essentially no genetic
337 structure appeared in non-outliers, a clustering based on outlier loci brought to light a clearer
338 pattern which was further supported by the Ward's clustering method (Fig. 3). One cluster was
339 composed of five *Bt* fields (3, 5, 7, 43, 65) which were spatially aggregated (area A_1, Fig. 1)

340 in the surroundings of the location where resistance was first reported. While the extent of A_1
341 could be sketched as an area of approximately 100 km radius, A_2 was defined as an outgroup
342 which included the remaining nine *Bt* fields.

343 A second round of outlier analysis (using Bayescan) was performed within each group A_1 and
344 A_2, including the 4 closest non-*Bt* fields (1, 2, 4, 6) in A_1 and the remaining ones in A_2. These
345 additional analyses did not yield outliers which were not detected previously (Table 2).
346 However, some outliers returned in the first round were not detected as significant ($\log_{10}$ *PO*
347 < 1) in the second round, presumably due to a reduction in statistical power.

348 The association of marker frequency at outlier loci with *Bt* and non-*Bt* fields was then
349 examined. Mixed GLM with spatially-correlated random effects showed significant
350 differences in frequency at four outlier markers (L192, L204, L299 and L369) in A_1 ($P <$
351 0.005, Table 2) whereas no significant association was found in A_2.

352 These four loci were therefore considered strong candidates associated with resistance to *Bt*
353 toxins. Spatial distribution of these putative hitch-hikers within A_1 was highly consistent (Fig.
354 4): marker frequency in the *Bt* fields was either systematically lower (L192, L204, 299) or
355 systematically higher (L369) than in the non-*Bt* fields. It is worth noticing that these contrasts
356 were stronger in the two most Southern *Bt* fields of A_1 (5 and 7), suggesting a higher
357 resistance allele frequency in these fields. Overall, allelic frequencies at the four outlier
358 markers (in the 28 fields) were significantly inter-correlated (see also Appendix S3). Pearson's
359 correlation coefficients were: $0.6 < R < 0.7$ ($P < 0.001$) in the pairs {L192-L204}, {L192-
360 299} and {L204-299}; $R = -0.51$ ($P = 0.005$) in the pair {L204-L369}; $R = -0.44$ ($P = 0.017$)
361 in the pair {L192-L369}; and not significant in one pair of loci only, $R = -0.30$ ($P = 0.115$).

362 In addition, linkage disequilibrium was detected among putative hitch-hikers within the
363 smaller and more homogeneous area A_1. Highly significant pairwise linkage disequilibrium

364 values $0.42 < | r | < 0.47$ ($\log_{10}$-likelihood ratios $LR \geq 3$) were obtained for the pairs of loci
365 {L192-L369},{L204-L369} and {L204-L299}, while the remaining three pairs were
366 characterised by *r*-values > 0.25: $r = 0.38$ for the pair { L192-L369} ($LR = 2.4$), $0.25 < | r | <$
367 0.32 ($LR > 1.13$) for {L299-L369} and {L192-L299}. Consistently, genetic differentiation at
368 these four outlier loci was elevated between *Bt* and non-*Bt* fields of A_1 ($\Phi_{ST} = 0.24$; $P <$
369 0.001) as well as between *Bt* fields of A_1 and A_2 ($\Phi_{ST} = 0.21$; $P < 0.001$). By contrast, it was
370 low ($\Phi_{ST} < 0.05$) within either type of field (*Bt* or non-*Bt*) and within each area (A_1 and A_2).

371

372 **Controlled crosses**

373 The site (Va) was located in the area A_1 in which genetic contrasts were observed, while (Li)
374 and (Be) were located farther apart in A_2 (see Fig. 1). Overall survival was 0.67 when
375 offspring was reared on non-*Bt* maize, significantly higher ($P < 0.001$) than in offspring
376 reared on *Bt* maize (Fig. 5). When reared on *Bt* maize, survival of Va × Tz offspring (0.32)
377 was clearly higher than both Li × Tz and Be × Tz offspring ($P < 0.001$). Given substantial
378 heterogeneity within the two latter crosses, the highest survival in each type of offspring
379 (0.11, 0.13) was compared to that of Va × Tz offspring reared on *Bt*; it was unambiguously
380 confirmed that progeny obtained with individuals from Valhaarts outperformed progeny of
381 the other crosses on *Bt* maize ($P < 0.001$; Fig. 5). While survival in a Li × Tz progeny reared
382 on *Bt* maize was particularly low (6/383), a F2 sib-cross between 2 of the 6 survivors resulted
383 in significantly increased survival on *Bt* maize (51/251, $P < 0.001$), suggesting the existence
384 of a recessive resistance trait in (Li).

385

386 **Theoretical resistance spread**

Based on a demo-genetic system of deterministic reaction-diffusion equations, we show that the lower propagation speed (V) of resistance is a constant which takes the form (Appendix S1, see also Tyutyunov *et al*. 2008):

$$V = \sqrt{2\sigma^2 \cdot \beta \cdot h \cdot \omega} \qquad [1]$$

The time required for a concentric wave of resistance to travel a distance d (km) may simply be approximated by $T \approx d / V$ generations.

From controlled crosses performed in the laboratory we estimated that a female of *B. fusca* produces on average approximately 250 viable neonates (see also Khadioli *et al*., 2014). In laboratory experiments, survival (egg to adult stage) of *B. fusca* has been estimated to be, respectively, of order 40% and 15-20% in optimal and in sub-optimal temperature conditions (Khadioli *et al*., 2014). In a different species of Noctuidae (*Helicoverpa zea*) the estimate of survival to adulthood in the field was close to 5% (Vargas and Nishida, 1980). Assuming a 1:1 sex ratio, we determined a conservative estimate of $\beta = 0.05 \times 250/2 \approx 6$.

A resistance allele being propagated under constant selective pressure typically results in fast rapid spread. Based on Eq. [1], we explored the scale at which a dominant resistance allele may spread within ~26 generations (i.e., 13 years, and assuming $h = 1$, $\omega = 0.6$, $\beta = 6$) from an initially local distribution. Selection due to a moderate fraction of *Bt* crop in the landscape ($\omega \approx 0.6$) and operating on a fully dominant resistance allele ($h = 1$) yields speeds of several kilometres per year, e.g., 4.6 km.year^{-1} $< V <$ 17 km.year^{-1} for $1 < \sigma < 10$. Over 26 generations, the spread of resistance would attain a radius of several tens of kilometres (i.e., approximately 60 km to 220 km) (Fig. 6). Under the same parameter values, a high axial dispersal variance ($\sigma^2 \approx 25$ km per generation) would be required for the resistance wave to travel from the location where resistance was first reported to the farthest *Bt*-fields in which individuals were found to survive (Fig. 6).

411

412

413 **DISCUSSION**

414

415 Being one of the few instances of field evolved resistance leading to *Bt* crop failure
416 (Tabashnik *et al*., 2009; Tabashnik *et al*., 2013), *Bt*-resistance in *B*. *fusca* may be considered
417 an insightful study case in the context of *Bt* crop deployment worldwide. Our results
418 confirmed that a sharp selective shift, as indicated by the extensive uptake of *Bt* maize
419 (Kruger *et al*., 2012), resulted in rapid evolution of resistance. Beyond the detection of outlier
420 loci, our results provide further understanding of both gene flow and field resistance in this
421 pest, notably in terms of spatial distribution, dynamics and diversity of resistance in the field.

422

423 **Gene flow and resistance spread**

424 In line with what was shown in several other moth species (e.g., Korman *et al*., 1993;
425 Nibouche *et al*., 1998; Bourguet *et al*., 2000; Han and Caprio, 2002), our empirical results
426 reflect extensive gene flow and low levels of population inbreeding in *B*. *fusca*. To a first
427 approximation, the whole population seems close to panmixia at coarse scales (700 km × 400
428 km). It has been suggested that *B*. *fusca* could be characterized by limited dispersal capacities
429 (Sezonlin *et al*., 2006). However, consistent with Dupas *et al*. (2014), our study did not reveal
430 strong barriers to gene flow. Estimate of neighbourhood size in *B*. *fusca* (N_b* ≈ 700
431 individuals) had the same order of magnitude as in the European corn borer *O*. *nubilalis* (N_b*
432 ≈ 500 on maize, see Martel *et al*., 2003). Classical methods for estimating gene flow (see
433 Rousset, 1997; Hardy, 2003) cannot yield a separate estimation of the parameters σ^2 and ρ_e,

such that the product of these has to be considered, i.e., $\rho_e.\sigma^2 \approx 55$ in our study. Assuming an effective density of $14 < \rho_e < 55$ individuals.km^{-2} therefore amounts to assuming a dispersal parameter $1 < \sigma^2 < 4$ km . Similar effective population densities, i.e., $5 < \rho_e < 50$ individuals.km^{-} , have been proposed in other species of wild moths (Saccheri *et al*., 2008) and moth pests (Martel *et al*., 2003). Mean flight distance of several kilometres was observed in males of the moth pest *Spodoptera litura*, using a release-recapture experiment. In *B*. *fusca*, observations of dispersing females crossing an inter-field of 1.6 km have been reported (see Harris and Nwanze, 1992). Although dispersal distances and timing may differ between males and females, these observations are roughly compatible with a dispersal parameter σ^2 scaling to a few km per generation.

Extensive gene flow in pest populations may affect the pace of resistance evolution by mitigating local increase of resistance with susceptible alleles originating from non-*Bt* fields, or by contributing to an accelerated spatial spread of resistance pockets. Besides gene dispersal (σ), the speed of resistance propagation crucially depends on the inheritance of resistance (i.e., h, the dominance parameter in our model) and the population ecology of a pest, notably its birth rate (β). These parameters would clearly deserve further consideration in order to better comprehend the spatial dynamics of resistance. Nonetheless, realistic numerical applications suggested a travelling speed of a few kilometres per generations. It seems plausible that a resistance pocket, initially small, may have quickly expanded into a bigger area of radius several tens of kilometers or more. Within 13 years ($k = 26$ generations), one might expect the radius of such a pocket to increase by approximately 100 km (provided $\omega \approx 0.6$, $h = 1$, $\beta = 6$, $\sigma^2 \approx 2.1$ km , $\rho_e \approx 26.8$), an order of distance which is commensurate with the size of A_1. However, our model suggested that only high dispersal variance could result in a propagation radius that would encompass all apparently resistant populations, which might lie at the margin of what can be reasonably assumed.

459

460 **Outlier loci and resistance**

461 While larvae surviving in *Bt* fields were observed across a vast area, a genetic signature
462 associated with field-evolved resistance could be attributed to a genetic hitch-hiking process
463 within a smaller region only. The existence of such a pattern is also suggestive of a recent
464 geographical spread of resistance. Indeed, due to genetic recombination, associations between
465 resistance allele and alleles at linked loci are not expected to persist in time, unless
466 recombination distance is very low (see Maynard Smith and Haigh, 1974). Correlated allele
467 frequencies and linkage disequilibrium are not necessarily indicative of physical linkage as
468 they may also result from various stochastic processes at population level. Nevertheless, given
469 an observed lack of neutral genetic structure, even at coarse spatial scale, we expect neither
470 population differentiation, demographic history nor genetic drift to strongly affect linkage
471 disequilibrium patterns in the wild.

472 In line with genetic hitch-hiking arising due to resistance evolution, consistent differences in
473 outlier marker frequencies were observed between *Bt* and non-*Bt* fields. The size of these
474 differences could reflect an intermediate frequency of resistance alleles in A_1 (see also
475 Appendix S4). On the one hand, a review study (Tabashnik *et al*., 2013) evaluated the
476 proportion of resistant *B*. *fusca* individuals to be $\approx q^2 + 2hq(1 - q) \geq 50\%$, in the core
477 resistance region, which translates into a resistance allele frequency $q > 0.25$ assuming a
478 strictly dominant ($h = 1$) monogenic trait. On the other hand further assuming a resistance
479 allele exclusively associated with null alleles (a_0) at either outlier loci L192, L204 or L299,
480 one would expect the frequency of homozygotes a_0/a_0 (corresponding to an absence of
481 marker) to be close to $P(a_0/a_0) = q^2 / (q^2 + 2hpq)$ in *Bt* fields, i.e., leading to $q = 2 \times P(a_0/a_0) /$
482 $[1 + P(a_0/a_0)]$. Given $P(a_0/a_0) \approx 0.3$ at these loci in *Bt* fields (see Table 2), we attain a crude

483 estimate $q \approx 0.46$. Finally, relaxing the stringent assumptions of full linkage and strict
484 dominance, this exercise suggests a resistance allele frequency $0.25 < q < 0.5$ (see also
485 Appendix S4).

486 Note that some outlier loci could not be associated with resistance, which suggests that these
487 loci may not be related to resistance at all, or reflect a lack of power in some analyses, notably
488 due to the use of dominant genetic markers. However, it cannot be excluded that a fraction of
489 these outliers, loosely linked to the resistance trait, underwent quick hitch-hiking but did not
490 shift significantly during the time for which associations persisted, contributing to some levels
491 of genetic differentiation between areas (see also Appendix S5).

492

493 **Uniformity of resistance**

494 Since the first observations of field resistance in *B. fusca* were reported, in the Western part of
495 the maize cutivation region, most conspicuous failures to control populations of *B. fusca* in *Bt*
496 fields (e.g., see Van Rensburg *et al*., 2007; Kruger *et al*., 2011) have been observed in the
497 surroundings of this location. Moreover, the resistance trait associated with these failures
498 appeared to be dominant (Campagne *et al*., 2013). Our results showing differential progeny
499 survival on *Bt* plants and genetic contrasts (between A_1 and A_2) support the idea of a
500 dominant resistance trait being confined within A_1. Furthermore, it is coherent with the
501 chronology of resistance development in South Africa (Van den Berg *et al*., 2013) and with
502 the levels of infestation in *Bt* fields perceived by farmers at different locations: perceived field
503 infestations > 10% mostly occurred within 100−150 km from the site where resistance was
504 initially reported (Kruger *et al*., 2012; Van den Berg and Campagne, 2014).

505 By contrast, information about resistance found within A_2 is rather scarce (Van den Berg *et*
506 *al*., 2013). Even if survival of some susceptible individuals on plant tissues expressing a low

dose of toxins might partly explain the development of larvae on *Bt* fields within A_2, our results would advocate for independent evolution of resistance. Indeed, increased survival on *Bt* maize obtained in a F2 sib-cross clearly suggested the existence of at least one recessive resistance trait in A_2. Similarly to what has already been reported in another pest (e.g., Zhang *et al*., 2012, Jin *et al*., 2013), diverse resistance mutations characterised by different levels of recessiveness might therefore be responsible for the observed pattern of widespread field resistance to *Bt* maize in South Africa.

Finally, our results draw a valuable picture of resistance distribution and dynamics for one of the few instances of resistance evolution leading to control failure in the field. Evidence suggests resistance being a non-uniform trait across the extended region of maize production. The distribution of the main dominant resistance trait seems confined to a smaller region (100 km radius). The scale at which the distribution of this trait takes place could correspond to a recent geographical spread as confirmed by theoretical considerations showing a fast propagation of a dominant resistance trait across space. This study illustrates both potential challenges and insights of a strong evolutionary response scaling, in space and time, with the selective pressure of drastic environmental change.

Acknowledgements: The authors thank four anonymous reviewers for their constructive comments about earlier versions of the mansucript as well as Peter E Smouse and Stéphane Dupas for helpful discussions. PC was supported by IRD and Natural Environment Research Council grant NE/J022993/1; RP, JFS and BLR were funded by IRD; JVdB was supported by Biosafety South Africa (Grant 08-001).

Conflict of Interest:

The authors declare no conflict of interest.

Data Archiving:

AFLP data are available from Dryad doi:10.5061/dryad.n3r35

538 **References:**

539 Assefa Y, Conlong DE, Van den Berg J, Martin LA (2015). Ecological Genetics and Host
540 Range Expansion by *Busseola fusca* (Lepidoptera: Noctuidae). *Environ Entomol* **44**:1265–
541 1274.

542 Baxter SW, Badenes-Pérez FR, Morrison A, Vogel H, Crickmore N, Kain W, Wang P,
543 Heckel DG, Jiggins CD (2011). Parallel evolution of *Bacillus thuringiensis* toxin
544 resistance in Lepidoptera. *Genetics* **189**:675–679.

545 Bonin A, Bellemain E, Bronken Eidesen P, Pompanon F, Brochmann C, Taberlet P. (2004).
546 How to track and assess genotyping errors in population genetics studies. *Mol Ecol*
547 **13**:3261–3273.

548 Bourguet D, Bethenod MT, Pasteur N, Viard F (2000) Gene flow in the European corn borer
549 *Ostrinia nubilalis*: implications for the sustainability of transgenic insecticidal maize.
550 *Proc Roy Soc B* **267**:117–122.

551 Campagne P, Kruger M, Pasquet R, Le Ru B, Van den Berg J (2013) Dominant inheritance of
552 field-evolved resistance to *Bt* corn in *Busseola fusca*. *PLoS ONE* **8**(7): e69675.

553 Charlesworth B, Charlesworth D, Barton NH (2003). The effects of genetic and geographic
554 structure on neutral variation. *Annu Rev Ecol Evol Syst* **34**:99–125..

555 Chybicki IJ, Oleksa A, Burczyk J (2011). Increased inbreeding and strong kinship structure in
556 *Taxus baccata* estimated from both AFLP and SSR data. *Heredity* **107**:589–600.

557 Daly JC, Gregg P (1985). Genetic variation in *Heliothis* in Australia: species identification
558 and gene flow in the two pest species *H. armigera* (Hübner) and *H. punctigera*
559 Wallengren (Lepidoptera: Noctuidae). *Bull Entomol Res* **75**:169–184.

560 de Villemereuil P, Gaggiotti OE (2015). A new F_{ST}-based method to uncover local adaptation
561 using environmental variables. *Meth Ecol Evol* **6**:1248–1258.

562 Dupas S, Le Rü B, Branca A, Faure N, Gigot G, Campagne P, Sezonlin M, Ndemah R,
563 Ong'amo G, Calatayud PA, Silvain JF (2014). Phylogeography in continuous space:
564 coupling species distribution models and circuit theory to assess the effect of contiguous
565 migration at different climatic periods on genetic differentiation in *Busseola fusca*
566 (Lepidoptera: Noctuidae). *Mol Ecol* **23**:2313–2325.

567 Endersby NM, Hoffman AA, McKechnie SW, Weeks AR (2007) Is there genetic structure in
568 populations of *Helicoverpa armigera* from Australia? *Entomol Exp Appl* **122**: 253–263.

569 Excoffier L, Smouse PE, Quattro JM (1992). Analysis of molecular variance inferred from
570 metric distances among DNA haplotypes: application to human mitochondrial DNA
571 restriction data. *Genetics* **131**:479–491.

572 Evanno G, Regnaut S, Goudet J (2005). Detecting the number of clusters of individuals using
573 the software STRUCTURE: a simulation study. *Mol Ecol* **14**:2611–2620.

574 Foll M, Gaggiotti OE (2008). A genome scan method to identify selected loci appropriate for
575 both dominant and codominant markers: a Bayesian perspective. *Genetics* **180**:977-993

576 Han Q, Caprio MA (2002) Temporal and spatial patterns of allelic frequencies in Cotton
577 Bolloworm (Lepidoptera: Noctuidae). *Environ Entomol* **31**:462–468.

578 Hardy OJ, Vekemans X (2002). SPAGeDi: a versatile computer program to analyse spatial
579 genetic structure at the individual or population levels. *Mol Ecol Notes* **2**:618–620.

580 Hardy OJ (2003). Estimation of pairwise relatedness between individuals and characterization
581 of isolation-by-distance processes using dominant genetic markers. *Mol Ecol* **12**:1577–
582 1588.

583 Harris, K M and Nwanze, K E (1992) *Busseola fusca* (Fuller), the African maize stalk borer: a
584 handbook of information. Manual. International Crops Research Institute for the Semi-
585 Arid Tropics.

586 Heckel DG, Gahan LJ, Baxter SW, Zhao JZ, Shelton AM, Gould F, Tabashnik BE (2007).
587 The diversity of Bt resistance genes in species of Lepidoptera. *J Invert Pathol* **95**:192–
588 197.

589 Hill WG (1974). Estimation of linkage disequilibrium in randomly mating populations.
590 *Heredity* **33**:229–239.

591 Jensen JD (2014). On the unfounded enthusiasm for soft selective sweeps. *Nat Commun*
592 **5**:5281.

593 Jin L., Wei Y., Zhang L., Yang Y., Tabashnik BE and Wu Y. (2013), Dominant resistance to
594 Bt cotton and minor cross-resistance to Bt toxin Cry2Ab in cotton bollworm from China.
595 *Evol Appl* **6**:1222–1235.

596 Jin L, Zhang H, Lu Y, Yang Y, Wu K, Tabashnik BE, Wu Y (2015). Large-scale test of the
597 natural refuge strategy for delaying insect resistance to transgenic *Bt* crops. *Nat*
598 *Biotechnol* **33**:169–174.

599 Kfir R, Overholt WA, Khan ZR, Polaszek A (2002). Biology and management of
600 economically important lepidopteran cereal stem borers in Africa. *Ann Rev Entomol*
601 **47**:701–731.

602 Khadioli N, Tonnang ZEH, Ong'amo G, Achia T, Kipchirchir I, Kroschel J, Le Rü B (2014),
603 Effect of temperature on the life history parameters of noctuid lepidopteran stem borers,
604 Busseola fusca and Sesamia calamistis. *Ann Appl Biol* **165**: 373–386.

605 Kim KS, Bagley MJ, Coates BS, Hellmich RL, Sappington TW (2009) Spatial and temporal
606 genetic analyses show high gene flow among European corn borer (Lepidoptera:
607 Crambidae) populations across the central US corn belt. *Environ Entomol* **38**:1312–1323.

608 Kojima KI, Schaffer HE (1967). Survival process of linked mutant gene. *Evolution* **21**:518–
609 531.

610 Korman AK, Mallet J, Goodenough JL, Graves JB, Hayes JL, Hendricks DE, Wall M (1993).
611 Population structure in *Heliothis virescens* (Lepidoptera: Noctuidae): An estimate of gene
612 flow. *Ann Entomol Soc Am* **86**:182–188.

613 Kruger M, Van Rensburg JBJ, Van den Berg J (2011). Resistance to *Bt* maize in *Busseola*
614 *fusca* (Lepidoptera: Noctuidae) from Vaalharts, South Africa. *Environ Entomol* **40**:477–
615 483.

616 Kruger M, Van Rensburg JBJ, Van den Berg J (2012) Transgenic *Bt* maize: farmers'
617 perceptions, refuge compliance and reports of stem borer resistance in South Africa. *J*
618 *Appl Entomol* **136**: 38–50.

619 Kruger M, Van Rensburg JBJ, Van den Berg J (2014). No fitness costs associated with
620 resistance of *Busseola fusca* (Lepidoptera: Noctuidae) to genetically modified *Bt* maize.
621 *Crop Prot* **55**:1–6.

622 Le Rü BP, Ong'amo GO, Moyal P, Muchugu E, Ngala L et al. (2006). Geographic distribution
623 and host plant ranges of East African noctuid stem borers. *Ann Soc Entomol Fr* **42**:353–
624 361.

625 Mallet J (1989). The evolution of insecticide resistance: have insects won? *Trends Ecol Evol*
626 **4**: 336–340.

627 Martel C, Réjasse A, Rousset F, Bethenod MT, Bourguet D (2003) Host-plant-associated
628 genetic differentiation in Northern French populations of the European corn borer.
629 *Heredity* **90**:141–149.

630 Maruyama T (1972) Rate of decrease of genetic variability in a two-dimensional continuous
631 population of finite size. *Genetics* **70**:639–651.

632 Maynard-Smith J, Haigh J (1974). The hitch-hiking effect of a favourable gene. *Genet Res*
633 **23**:23–35.

634 Morin S, Biggs RW, Sisterson MS, Shriver L, Ellers-Kirk C, Higginson D, Holley D, Gahan
635 LJ, Heckel DG, Carrière Y, Dennehy TJ, Brown JK, Tabashnik BR (2003). Three
636 cadherin alleles associated with resistance to *Bacillus thuringiensis* in pink bollworm.
637 *Proc Natl Acad Sci USA* **100**:5004–5009.

638 Peakall R, Smouse PE (2006). Genalex 6: genetic analysis in Excel. Population genetic
639 software for teaching and research. *Mol Ecol Notes* **6**:288–295.

640 Peck SL, Gould F, Ellner SP (1999). Spread of Resistance in Spatially Extended Regions of
641 Transgenic Cotton: Implications for Management of Heliothis virescens (Lepidoptera:
642 Noctuidae). *J Econ Entomol* **92**:1–16.

643 Pritchard JK, Stephens M, Donelly P (2000). Inference of population structure using
644 multilocus genotype data. *Genetics* **155**:945–959.

R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Rousset F (1997). Genetic Differentiation and Estimation of Gene Flow from F-Statistics Under Isolation by Distance. *Genetics* **145**:1219-1228.

Saccheri IJ, Rousset F, Watts PC, Brakefield PM, Cook LM (2008) Selection and gene flow on a diminishing cline of melanic peppered moths. *Proc Natl Acad Sci USA* **105**:16212–16217.

Sezonlin M, Dupas S, Le Rü B, Le Gall P, Moyal P, Calatayud PA, Silvain JF (2006) Phylogeography and population genetics of the maize stalk borer *Busseola fusca* (Lepidoptera, Noctuidae) in sub-Saharan Africa. *Mol Ecol* **15**:407–420.

Tabashnik BE, Liu YB, Malvar T, Heckel DG, Masson L et al. (1998). Insect resistance to *Bacillus thuringiensis*: uniform or diverse? *Philos Trans R Soc Lond B* **353**: 1751–1756.

Tabashnik BE, Gassmann AJ, Crowder DW, Carriere Y (2008). Insect resistance to *Bt* crops: evidence versus theory. *Nat Biotechnol* **26**:199–202.

Tabashnik BE, Van Rensburg JBJ, Carriere Y (2009) Field-evolved insect resistance to *Bt* crops: definition, theory, and data. *J Econ Entomol* **102**:2011–2025.

Tabashnik BE, Brévault, Carrière Y (2013). Insect resistance to Bt crops: lessons from the first billion acres. *Nature Biotechnol* **31**:510–521.

Tay WT, Mahon RJ, Heckel DG, Walsh TK, Downes S, James WJ, et al. (2015). Insect resistance to *Bacillus thuringiensis* toxin *Cry*2 Ab is conferred by mutations in an ABC transporter subfamily A protein. *Plos Genetics* **11**:e1005534.

667 Teshima KM, Coop G, Przeworski M (2006). How reliable are empirical genomic scans for
668 selective sweeps? *Genome Res* **16**:702–712.

669 Tyutyunov Y, Zhadanovskaya E, Bourguet D, Arditi R (2008). Landscape refuges delay
670 resistance of the European corn borer to Bt-maize: a demo-genetic dynamic model. Theor
671 *Popul Biol* **74**:138–46.

672 Vacher C, Bourguet D, Rousset F, Chevillon C, Hochberg ME (2003). Modelling the spatial
673 configuration of refuges for a sustainable control of pests: a case study of *Bt* cotton. *J Evol*
674 *Biol* **16**:378–387.

675 Vacher C, Bourguet D, Rousset F, Chevillon C, Hochberg ME (2004). High dose refuge
676 strategies and genetically modified crops – reply to Tabashnik et al. *J Evol Biol* **17**:913–
677 918.

678 Van Rensburg JBJ (2007) First report of field resistance by the stem borer, *Busseola fusca*
679 (Fuller) to *Bt*-transgenic maize. S. Afr. *J Plant Soil* **24**:147–151.

680 Van den Berg J, Hilbeck A, Bhøn T (2013). Pest resistance to *Cry*1Ab *Bt* maize: field
681 resistance, contributing factors and lessons from South Africa. *Crop Prot* **54**:154–160.

682 Van den Berg J, Campagne P (2014). Resistance of *Busseola fusca* to Cry1Ab *Bt* Maize
683 Plants in South Africa and Challenges to Insect Resistance Management in Africa. In:
684 Soberón M, Gao Y, Bravo A (eds). *Bt resistance: characterization and strategies for GM*
685 *crops expressing Bacillus thuringiensis toxins*. CAB International, pp 36–48.

686 Vargas R, Nishida T (1980). Life table of the corn earworm, *Heliothis zea* (Boddie), in sweet
687 corn in Hawaii. *Proc Hawaiian Entomol Soc* **23**:301–307.

688 Venables WN, Ripley BD (2002). Modern Applied Statistics with S. Fourth Edition. Springer,
689 New York. ISBN 0-387-95457-0.

690 Vos P, Hogers R, Bleeker M, Reijans M, van De Lee T, Hornes M, Zabeau M (1995). AFLP:
691 a new technique for DNA fingerprinting. *Nucleic Acids Res* **23**:4407–4414.

692 Zhang H, Tian W, Zhao J, Jin L, Yang J, Wu S, Wu K, Cui J, Tabashnik BE, Wu Y (2012).
693 Diverse genetic basis of field-evolved resistance to *Bt* cotton in cotton bollworm from
694 China. *Proc Natl Acad Sci USA* **109**: 10275–10280.

695

696

697 Table 1: General characteristics of the 28 fields sampled: IL is the infestation level (%); *n*, the
698 sample size; *F* the inbreeding coefficient in a given field and its 95% estimation interval
699 ($IC_{95\%}$).

Field ID	Longitude	Latitude	Altitude	Crop	IL (%)	*n*	*F*	$IC_{95\%}$
3	24.76	-27.75	1096 m	corn*	15	25	0.011	[0.000−0.053]
5	24.70	-29.10	1130 m	corn*	5	25	0.007	[0.000−0.037]
7	24.72	-29.29	1100 m	corn*	5	25	0.010	[0.000−0.047]
43	25.96	-28.04	1136 m	corn*	10	25	0.011	[0.000−0.049]
65	24.67	-28.05	1149 m	corn*	> 30	25	0.026	[0.002−0.081]
2	24.85	-27.93	1150 m	corn	15	25	0.011	[0.001−0.082]
4	24.79	-27.71	1130 m	corn	5	25	0.009	[0.000−0.042]
6	24.74	-29.17	1100 m	corn	*na*	25	0.009	[0.000−0.041]
11	26.79	-27.41	1136 m	corn	5	25	0.013	[0.005−0.111]
89	28.82	-25.72	1473 m	corn	5	25	0.009	[0.000−0.043]
1	25.47	-27.68	1213 m	sorghum	> 30	24	0.031	[0.002−0.098]
8	26.23	-28.11	1290 m	corn*	5	20	0.010	[0.000−0.047]
10	27.09	-26.69	1356 m	corn	*na*	25	0.044	[0.007−0.115]
12	28.55	-27.94	1754 m	corn*	> 30	25	0.058	[0.012−0.136]
13	28.24	-27.79	1615 m	corn*	15	25	0.028	[0.001−0.093]
14	28.60	-27.75	1670 m	corn*	5	25	0.030	[0.001−0.093]
15	28.08	-27.46	1638 m	corn*	5	25	0.024	[0.001−0.081]
17	27.70	-26.70	1461 m	corn	5	25	0.008	[0.000−0.038]
24	29.66	-26.79	1660 m	corn	5	25	0.019	[0.000−0.073]
40	26.12	-26.22	1454 m	corn	10	23	0.023	[0.001−0.087]
41	25.51	-27.23	1363 m	corn*	> 30	25	0.008	[0.000−0.038]
42	26.54	-27.77	1288 m	corn	15	24	0.009	[0.000−0.043]
51	28.08	-28.10	1623 m	corn*	> 30	25	0.037	[0.002−0.106]
54	26.77	-26.22	1520 m	corn*	5	25	0.021	[0.001−0.070]
57	26.59	-25.96	1551 m	corn	25	25	0.011	[0.000−0.052]
64	26.94	-26.14	1529 m	corn	25	24	0.012	[0.001−0.057]
71	26.01	-26.39	1427 m	corn	5	24	0.034	[0.003−0.099]
75	25.63	-26.80	1387 m	corn*	15	24	0.010	[0.000−0.049]

700

701 *, *Bt* field

702 *na*, fields for which information is not available.

703 Table 2: Overview of the outlier scan. The three first columns display outlier loci which were
704 detected when performing an overall analysis. In each geographical region, A_1 and A_2,
705 comparisons of marker frequency between *Bt* fields (f_{Bt}) and non-*Bt* fields (f_{nBt}) were carried
706 at various outlier loci. The *P*-values refers to a mixed GLM (binomial family) with spatially
707 correlated random effects.

Area			A_1				A_2		
Locus	F_{ST}	log(PO)	f_{Bt}	f_{nBt}	*P*		f_{Bt}	f_{nBt}	*P*
73	0.10	> 1000	–	–	–		–	–	–
81	0.14	> 1000	0.91	0.66	0.254		0.72	0.75	0.974
110	0.20	> 1000	0.56	0.29	0.597		0.11	0.18	0.147
184	0.08	1.6	–	–	–		–	–	–
187	0.07	1	0.98	0.87	0.231		–	–	–
192	0.12	> 1000	0.71	0.99	0.005	**	–	–	–
204	0.14	> 1000	0.67	0.97	0.002	**	–	–	–
208	0.07	1	–	–	–		–	–	–
231	0.08	1.3	–	–	–		–	–	–
299	0.08	1.3	0.66	0.93	0.004	**	–	–	–
306	0.08	1.2	–	–	–		–	–	–
331	0.14	> 1000	–	–	–		0.80	0.83	0.639
369	0.15	> 1000	0.98	0.68	0.002	**	0.81	0.82	0.677
390	0.10	2.5	–	–	–		0.63	0.69	0.602
392	0.08	1.4	–	–	–		–	–	–
492	0.13	> 1000	0.01	0.02	0.675		0.09	0.19	0.057
572	0.10	> 1000	–	–	–		–	–	–
596	0.12	> 1000	–	–	–		0.96	0.92	0.110
615	0.15	> 1000	–	–	–		0.78	0.80	0.844
617	0.16	> 1000	0.75	0.48	0.319		–	–	–
631	0.08	1.6	–	–	–		–	–	–
738	0.16	> 10	0.97	0.81	0.700		0.88	0.95	0.176
759	0.11	2.4	–	–	–		–	–	–
782	0.12	> 1000	–	–	–		0.90	0.82	0.421
818	0.11	> 1000	–	–	–		–	–	–
821	0.09	1.3	–	–	–		0.77	0.82	0.534

708

709 −, locus detected as an outlier in the overall analysis but no longer significant when analyzing
710 A_1 and A_2 separately. Note that the absence of contrasts between *Bt* and non-*Bt* fields was
711 verified at all outlier loci within each zone but are not displayed here for the sake of
712 simplicity. Significance level: ** $P < 0.01$; * $0.05 < P < 0.01$.

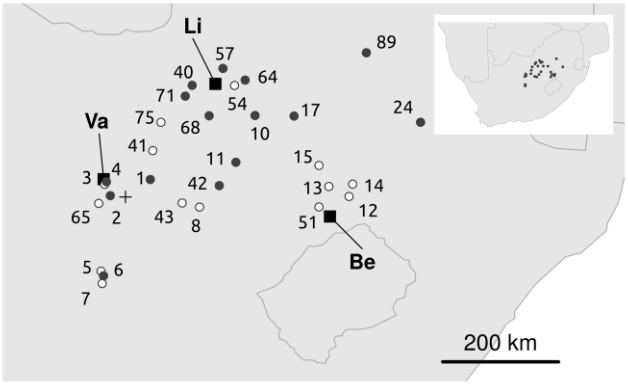

Figure 1: Map of the South African localities where larvae of *Busseola fusca* were collected. Empty dots represent *Bt* fields; dark-grey dots, non-*Bt* field; square dots, locations sampled for the controlled crosses (Va, Be, Li). Location where resistance was initially reported is represented by a cross. Numbers are the field IDs as displayed in Table 1.

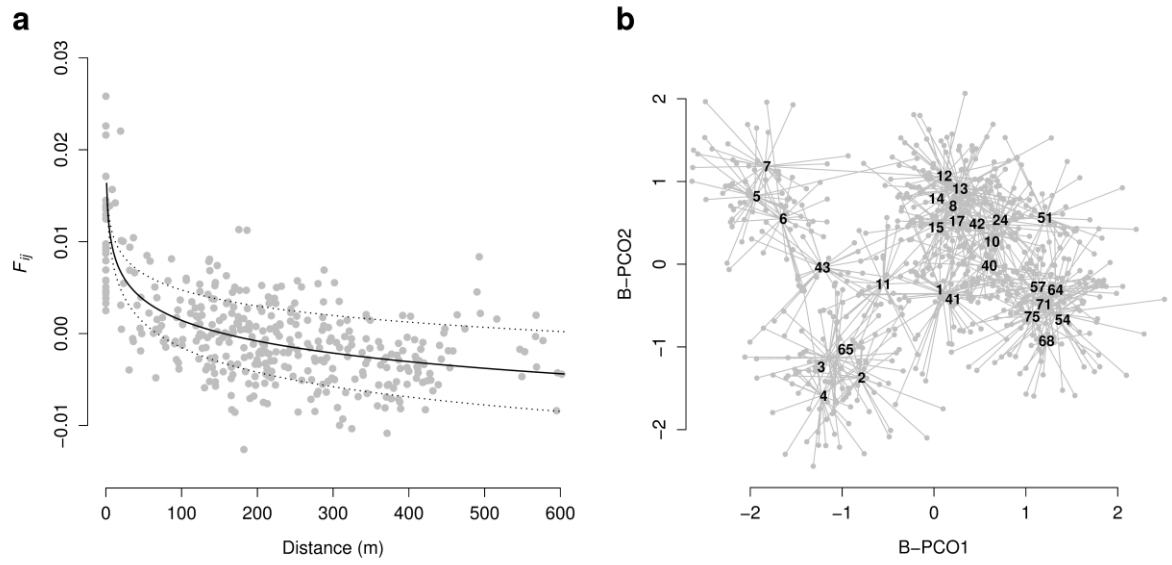

Figure 2: Isolation by distance in *B. fusca* populations (non-outlier loci). (A) Pairwise kinship coefficients (F_{ij}) as a function of distance (km). The curve corresponds to the regression of kinship as a function of log-transformed distance; dotted lines represent the 95%-Confidence Interval of the regression slope. (B) Between-field analysis performed on Principal Coordinate (PCO) axes. Barycentre of each field sampled is denoted by its ID (number); individuals collected in the different fields correspond to grey dots; lines relate each individual to its field of origin.

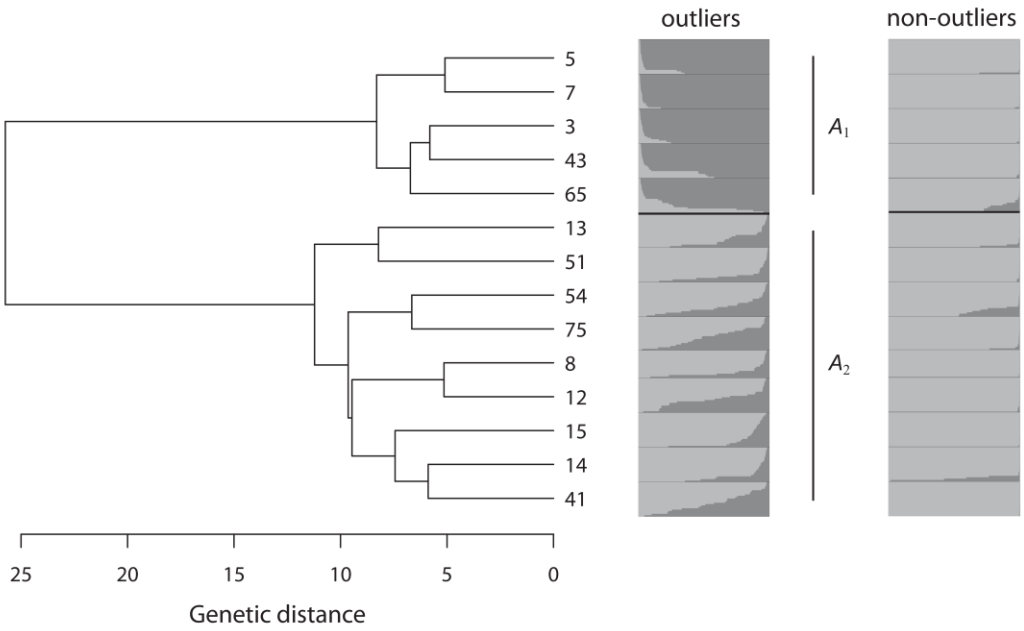

Figure 3: Comparison of population structures between both the outliers and non-outlier loci, for individuals collected in *Bt*-fields. The dendrogram (left) was obtained based on Euclidean genetic distances obtained with the dataset of outlier loci. Two genetic groups are distinguished: A_1 and A_2. Comparative STRUCTURE plots (right) of the 14 *Bt* fields using outlier loci and non-outlier loci. The number of clusters in each analysis was set at $K = 2$. The length of individual bars corresponds to the probability of membership to a cluster. Numbers identify the *Bt* field sampled.

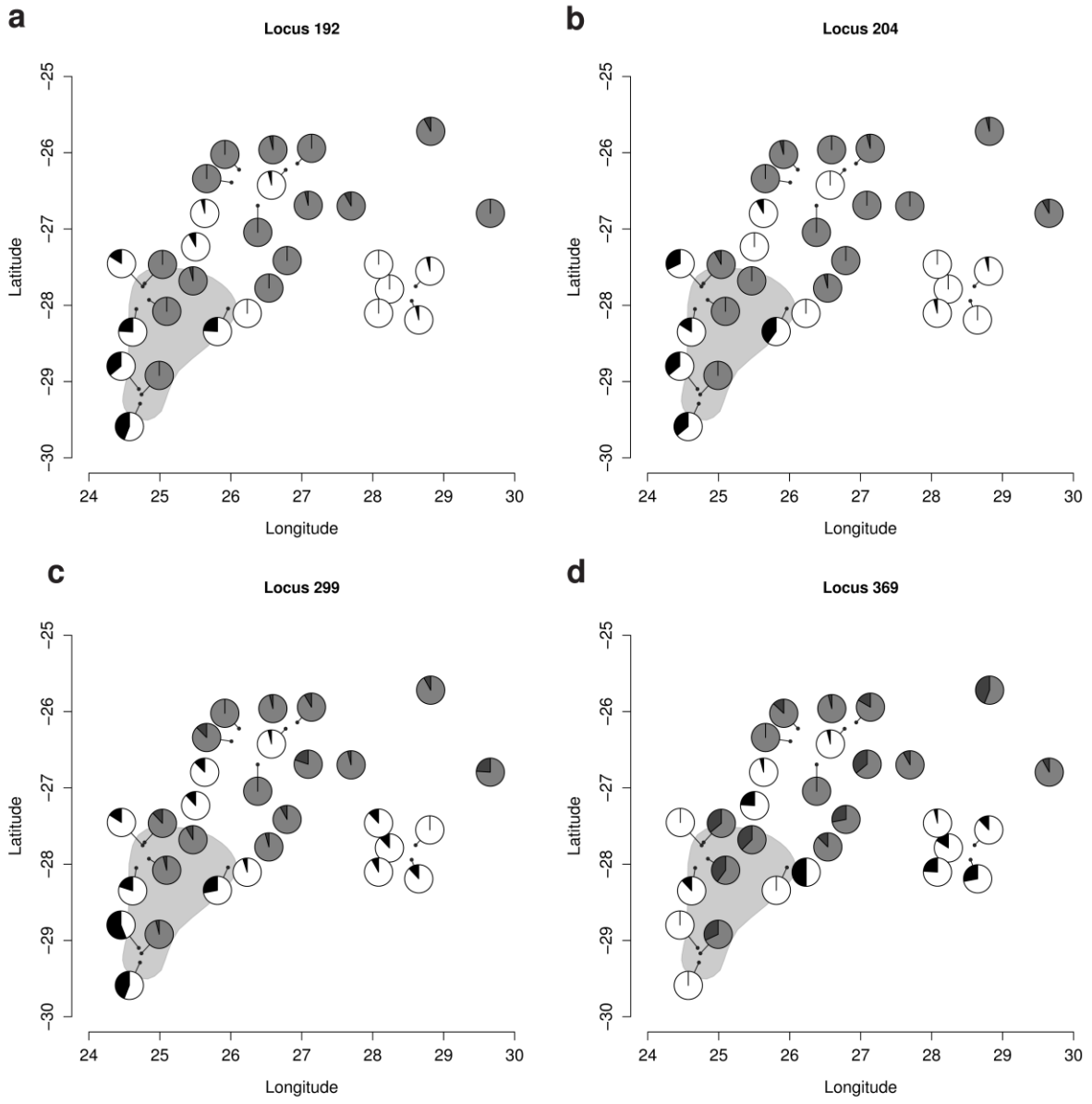

Figure 4: Marker frequency at four outliers loci for the different locations sampled. Black-and-white pies represent *Bt* fields; shaded-grey pies, non-*Bt* fields. The darker fraction of the pies corresponds to the absence of marker (i.e., homozygote for the null allele). The shaded area delineates the region A_1, while by default A_2 is composed of any fields outside this region.

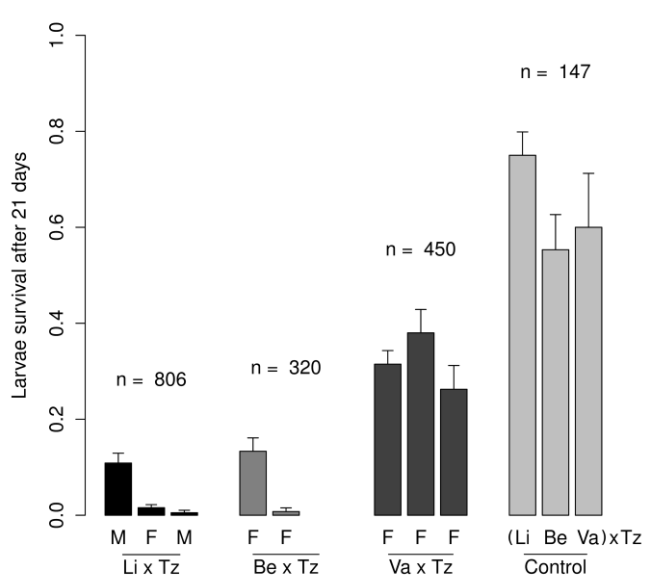

Figure 5: Survival of progeny on *Bt* maize after 21 days. Crosses were performed between individuals originating from a susceptible Tanzanian strain (Tz) and individuals originating from *Bt* fields in three different locations: Lichtenburg (Li × Tz), Bethleem (Be × Tz) and Valhaarts (Va × Tz), the latter being in the area A_1. The control consisted of the three same types of cross for which progeny was reared on non-*Bt* maize. M/F indicate whether the resistant parent was a male (M) or a female (F); error bars represent standard errors of the different proportions.

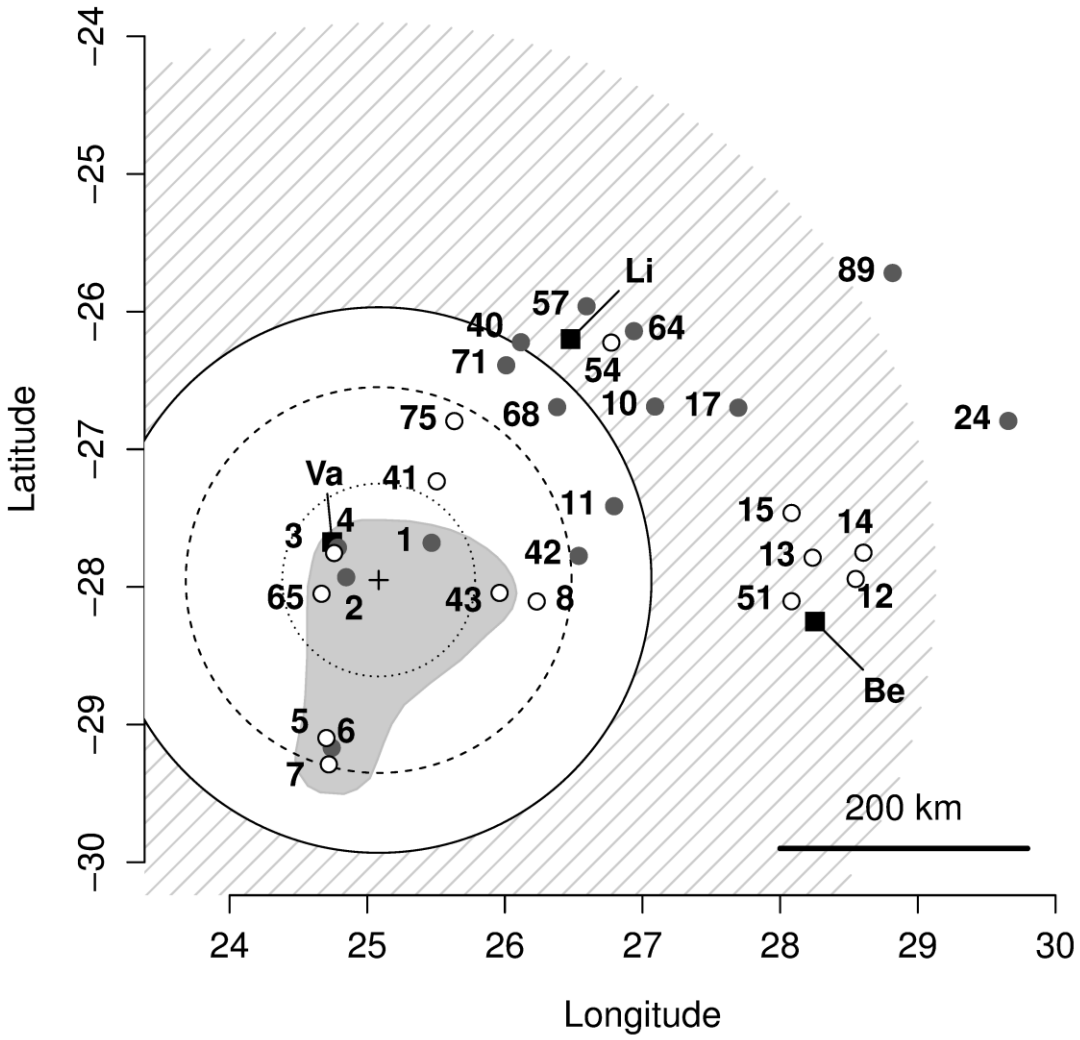

Figure 6: Expected resistance spread due to selection operated by *Bt* toxins ($\omega = 0.6$) after $k =$ 26 generations. Resistance was assumed initially confined at the location where it was first reported (centre of the circles). Dotted, dashed and solid circles correspond to distances obtained with increased axial dispersal variance (σ = 0.75, 3 and 6 km per generation). External border of the hatched ring represents the dispersal parameter required $\sigma^2 \approx 25$ km to encompass all resistant populations collected in this study. Empty dots correspond to sampled *Bt* fields; dark-grey dots, non-*Bt* field; square dots, locations sampled for the controlled crosses (Va, Be, Li). Numbers are the field IDs as displayed in Table 1.

Online Appendix S1

Invasion by a fitter 1-locus diploid genotype with logistic population dynamics

Resistance is assumed to involve a single locus with two alleles: a *Bt*-resistance and a *Bt*-susceptibility allele. Let *S*, *G* and *R* denote the densities of the susceptible homozygote, heterozygote, and resistant homozygote, which have respective birth rates (defined as the number of zygotes surviving to adulthood and assuming local random mating) $s = \beta \times (1 - \omega)$, $g = \beta \times [(1 - \omega) + h\omega]$, $r = 1$. Population is regulated by density dependent mortality which is neutral with respect to genotype. Individual dispersal is a Brownian motion characterised by a diffusion constant *D*. The model is defined by the following partial differential equations (PDEs):

$$\frac{\partial R(x,y,t)}{\partial t} = rNq^2 - mR - kRN + D\nabla^2 R$$
$$\frac{\partial G(x,y,t)}{\partial t} = 2gNq(1-q) - mG - kGN + D\nabla^2 G$$
$$\frac{\partial S(x,y,t)}{\partial t} = sN(1-q)^2 - mS - kSN + D\nabla^2 S$$

where $q = \frac{R + G/2}{N}$ is the frequency of the resistant allele; $N = R + G + S$ is the total population density; (x, y) and t represent space and time; the parameters m and k are account for density-independent and density-dependent mortality, respectively (see also Tyutyunov et al. 2008).

Because of the neutrality of the interactions and motion, the model may be simplified to the following two PDEs:

$$\frac{\partial q}{\partial t} = q\left(rq + g\left(1-q\right) - rq^2 - 2gq\left(1-q\right) - s\left(1-q\right)^2\right) + D\nabla^2 q + 2\nabla q \cdot \nabla \ln N$$
$$\frac{\partial N}{\partial t} = N\left(rq^2 + 2gq\left(1-q\right) + s\left(1-q\right)^2 - m - kN\right)$$

Regarding spatially independent solutions, the equation relative to the variations of q becomes independent of N and yields the solutions $q = 0$, $q = 1$, and $q = \frac{s-g}{r-2g-s}$. The third solution falls within the range $0 \le q \le 1$ if $g > \max(r,s)$ or $g < \min(r,s)$. The solution $q = 0$ is unstable (with eigenvalue $g - s$) and $q = 1$ is stable (with eigenvalue $g - r$). $N = 0$ is always an equilibrium value, irrespective of p, and is linearly unstable provided $m < \min(r; s)$. Assuming $s < g \le r$, the combined equilibria are $\left(q,N\right) = \left(0, \frac{s-m}{k}\right)$ (saddle point) and $\left(q,N\right) = \left(1, \frac{r-m}{k}\right)$ (stable).

We are interested in the spread of a small inocculum of the resistance allele, within a background population of susceptible alleles. We expect a travelling wave solution, where the stable $\left(1, \frac{r-m}{k}\right)$ invades the saddle-point solution $\left(0, \frac{s-m}{k}\right)$. The van Saarlos (2003, doi:10.1016/j.physrep.2003.08.001) method yields the "linear" wave speed under such

circumstances, which generally provides a lower bound estimate of (or, in many simple cases, is equal to) the true wave speed. This method firstly consists in linearising the equation about the unstable solution. If we write $(q,N)=\left(q,\frac{s-m}{k}+n\right)$, then the linearised equation for v becomes:

$$\begin{aligned}\frac{\partial q}{\partial t} &= q(h-s)+D\nabla^2 q+\frac{2k}{s-m}\nabla q\cdot\nabla n+O\left(q^2\right)\\ &= q(h-s)+D\nabla^2 q+O\left(q^2,qn,n^2\right)\end{aligned}$$

q being decoupled from the dynamics of n, which simplifies the calculation. Moreover, the linear form of the equation is the same as the Fisher-Kolmogorov wave equation, for which the wave speed is known. Therefore, without further calculation the linear wave speed of the system is given by:

$$V=2\sqrt{D(g-s)}$$

The correspondence between the axial dispersal variance and the diffusion coefficient D is, in a 2-dimensional space $2\sigma^2 = 4Dt = 4D$, per generation ($t = 1$). We therefore get the following wave speed equation:

$$V_\omega=\sqrt{4D\cdot\beta\cdot h\cdot\omega}=\sqrt{2\sigma^2\cdot\beta\cdot h\cdot\omega}\,,$$

i.e., in km per generation, using the manuscript notations.

Online Appendix S2

Determining the optimal number of genetic cluster with STRUCTURE

Table S2.1: Detection of *K* genetic clusters in both the outlier and non-outlier loci datasets.

	Non-outlier loci			Outlier loci		
	ln(*K*)	sd	Δ ln(*K*)	ln(*K*)	sd	Δ ln(*K*)
K = 2	-115678	115.0	—	-9205	15.6	—
K = 3	-116085	371.9	-406.2	-9214	41.1	-9.2
K = 4	-116987	956.7	-902.9	-9635	209.1	-421.2
K = 5	-117923	1906.4	-935.2	-10064	451.9	-429.0

ln(*K*): posterior log-probability of the data given *K*

sd: standard deviation

$\Delta \ln(K) = \ln(K) - \ln(K - 1)$

Online Appendix S3

Coordinated changes in frequency at four outlier loci

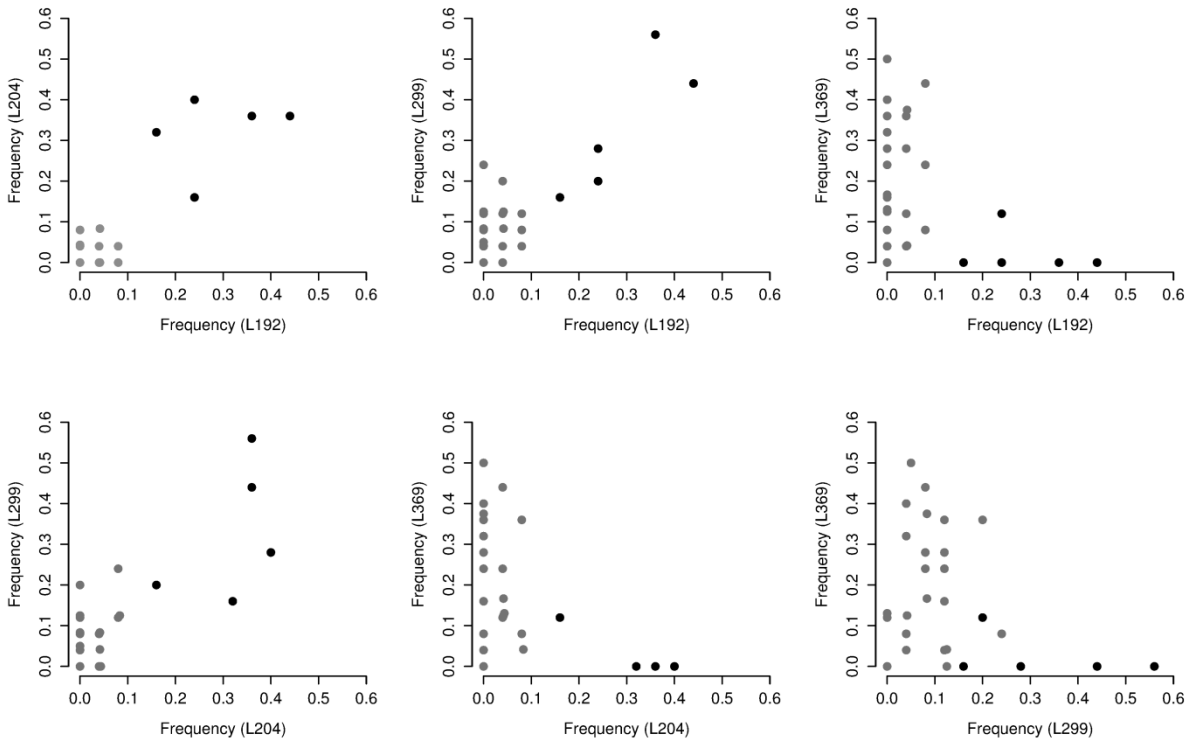

Figure S3.1: Frequency of the absence of marker at the four outlier loci L192, L204, L299 and L369 in the 28 fields sampled. Plots are displayed for the six pairs of loci. Black dots denote Bt-fields of A_1.

Online Appendix S4

Bayesian projection of resistance allele frequency assuming a monogenic resistance trait

We propose here an estimation of the resistance allele frequency assuming that the 4 most convincing outlier loci are physically linked to a monogenic resistance trait in A_1. On this basis, we infer the likely allelic phases, i.e., the allelic fractions, at neutral loci, linked to both the resistance and susceptible alleles. Negligible deviations from Hardy-Weinberg equilibrium are assumed.

The resistance allele (R) has a uniform frequency q in A_1 while the frequency of the susceptible allele (S) has a frequency $p = 1 - q$. We define u and v as the conditional null allele frequencies at a neutral locus, linked to the respective alleles, R and S. The model is thus defined by the parameters q (or p), u, v and h, exclusively. Provided the frequency of the null allele is $qu + pv$ in the population, the frequency of the null AFLP type "00" (absence of marker) is $\alpha = (qu + pv)$, which may be further expanded (see below). In a *Bt* field, a fraction $p + 2pq \times (1 - h)$ of individuals is expected to be killed by the *Bt* toxin quickly after eggs have hatched, leading to the frequency of homozygotes "00": $\beta = (q\,u + 2h \times qupv) / [1 - p - 2pq \times (1 - h)]$. Under this model, the distribution of homozygote null allele "00", in a *Bt* field i and a non-*Bt* field j follows a binomial distribution which leads to the likelihood function:

$$P\left(D|q,s,u,h\right) = \prod_{i=1}^{B} C_{n_i}^{k_{0i}} \cdot \left(\beta\right)^{k_{0i}} \cdot \left(1-\beta\right)^{n_i - k_{0i}} \times \prod_{j=1}^{\bar{B}} C_{m_j}^{l_{0j}} \cdot \left(\alpha\right)^{l_{0i}} \cdot \left(1-\alpha\right)^{m_i - l_{0i}}$$

where C is the binomial coefficient, k_0 and l_0 are the "null phenotype" counts in *Bt* and non-*Bt* fields, respectively; n_i and m_j, the number of individuals sampled in *Bt* field i and non-*Bt* field j; B and $\bar{B}$ are the respective number of *Bt* and non-*Bt* fields considered. Given a prior distribution for the different parameters, computation of the posterior probability density of q, u and v at each locus is achieved.

Uniform prior distributions were used for the parameters u and v. The prior distribution of h was a Beta distribution of parameters (5,1) which corresponds to a probability $P(h < 0.5) < 0.032$ and whose PDF monotonically increases as $h \rightarrow 1$. The prior for the proportion of resistant individuals $q + h \times 2q(1 - q)$ was modelled symmetric Beta distribution of parameters (1.3; 1.3) which mostly penalizes extremes values for the resistance allele: $q < 0.05$, $q < 0.95$.

Results: The mean of the posterior distributions for the resistance allele frequency at the time of the survey (Fig. S4.1) was similar for all four loci ($0.25 < q_t < 0.35$) suggesting a predominance of resistant heterozygote individuals in the population. Higher values dominance h lowers tends to be associated with lower q_t while lower h yields somewhat higher expectations for q_t. Irrespective of assumptions about the value of h, strong contrasts appear between allele frequencies linked to the resistance allele (v_t) and the ones related to the susceptible allele (u_t). The difference $|u_t - v_t|$ were of order ≈ 0.7-0.8 for all loci reflecting a strong hitch-hiking effect (see Maynard Smith and Haigh, 1974).

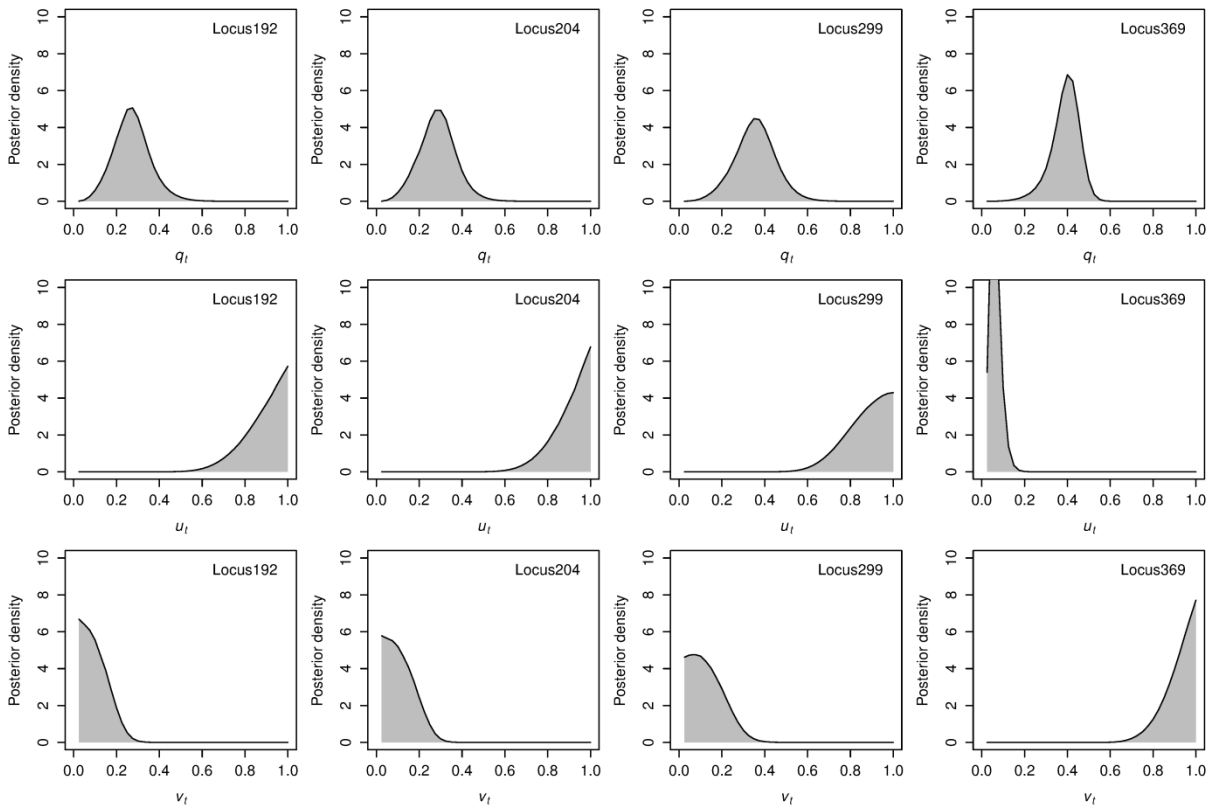

Figure S4.1: Inference of allelic phases at the four most convincing outlier loci. Posterior distributions associated with the parameters describing the three allele frequencies q_t, u_t, v_t: (q_t) denotes the resistance allele frequency; (u_t), the fraction null allele linked to the resistance allele; (v_t), the fraction null allele linked to the susceptible allele.

Online Appendix S5

Genetic structure of *Bt* and non-*Bt* fields at all outlier loci

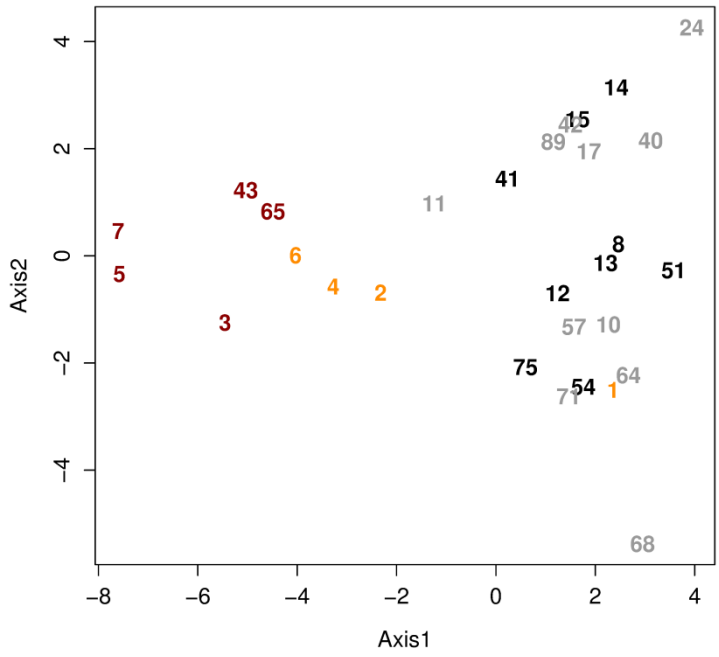

Figure S5.1: Map of a Principal Component Analysis based on the frequency markers at outlier loci (see Table 2), in both *Bt* and non-*Bt* fields. The first axis accounts for 31% of the total inertia; axis 2, 11%. Numbers correspond to the field ID (see Table 1). Darkred labels represent the *Bt* fields of A_1; orange, non-*Bt* fields of A_1; black, *Bt* fields of A_2; grey, non-*Bt* fields of A_2.