**Supplementary material and methods**

**Rearing protocol to obtain 3rd instar WCR larvae for midgut dissection**

The corn seeds used in the rearing of 3rd instar larvae did not express a Cry toxin and were purchased from Hancock Seed Company (Dade City, FL) as Reid’s Yellow Dent Field Corn. Seeds were planted before egg hatching to ensure food availability for neonate larvae. Seeds were equilibrated to room temperature, then soaked in water for 24 hrs. They were then mixed in a large plastic box, 34.6 cm L x 21.0 cm W x 12.4 cm H (Sterilite Corporation, Townsend, MA), containing humidified soil and kept at 25°C for 5-7 days to grow enough roots to feed larvae for 24 hours.

On the day eggs were received, they were placed at 25°C in a growth chamber and incubated for 10-15 days before hatching. On the first day of hatching, one or two germinated seeds were placed in a petri dish containing the eggs in soil. Petri dishes were closed with parafilm to avoid neonate escapes. The hatched neonates fed on the seedlings for 24 hours before being transferred to a small plastic box with lid, 18.4 cm W, 16.3 cm L x 6.7 cm H (Genpack, LLC; Glens Falls, NY), containing germinated seeds. On the same day, the small plastic boxes were prepared to receive 24 hour-old larvae. These boxes contained humidified soil (half of the box) and germinated seeds. Every day, the seedling on which neonates fed, was transferred in the small plastic box and the number of larvae added was recorded to avoid overabundance of larvae. The maximum number of larvae transferred in a box was 150 individuals. The seedling once placed in the small box was replaced by a new seed in the petri dish to feed the new neonates over the following 24 hours. After 10 to 15 days, depending on when the first larvae were transferred to the small plastic box, the 2nd instar larvae were transferred into large plastic boxes (Sterilite boxes described above), containing humidified soil and germinated seeds. The seedling mat (from the small plastic box) containing the larvae, was placed upside down into the large plastic box. The material of the large box (seedling mat and germinated seeds) was mixed with soil. The date of transfer in the small plastic box was kept for each large plastic box, as well as the date of transfer in the large box to keep track of the age of the larvae. Larvae transferred in the large plastic boxes were 2 weeks old and were thus in the last days of the 2nd instar. All larvae were recovered after 3 days spent in the large plastic box to make sure they were old enough for dissection and all the same age range.

### Polymorphism detection and analysis

#### Establishment of a list of polymorphisms with Freebayes

The program Freebayes (Garrison and Marth 2012), v1.0.2-dirty, was used to computationally detect small polymorphisms between the susceptible Standard strain (Branson 1976) and resistant Hopkinton colony (Gassmann et al. 2011) based on RNA sequencing data. Freebayes is a Bayesian genetic variant detector able to model multiallelic loci from individuals with non-uniform copy numbers, across samples. It produces short haplotypes detected from the alignment data. The underlying process associates the probability of a specific combination of genotypes, genotype likelihoods, to both the quality of sequencing observations (reads mapped for each potential locus) and *a priori* expectations about the distribution of alleles within a set of individuals sampled from the same population (priors). The calculated quality of the mapping corresponds to the probability that a read is mis-mapped against the reference. This program was chosen because the samples obtained for RNA sequencing consisted of pools of individuals; thus, virtual polyploidy was considered for the current analysis. In other programs commonly used, such as Samtools or GATK, the ploidy is set for diploid organisms by default and cannot be modified. However, Freebayes allows more flexibility, and because each sample contained 30 individuals, the ploidy of the dataset was then set to 60 (2n times 30). Setting a polyploidy avoided the false assumption of a unique diploid individual per sample.

The transcriptome alignments, per sample, were edited beforehand to add read groups to delineate the samples with Picard tool “AddOrReplaceReadGroups” (Broad Institute 2017) and remove the PCR duplicates to get a realistic count of the reads supporting a polymorphism, using the Samtools command “rmdup” (Li et al. 2009). Because the analysis with Freebayes required a reference sequence, the transcriptome sequence used for the alignments was the reference chosen for this study (Eyun et al. 2014).

Only variations (SNPs and indels) with more than 5 supporting reads, resulting in 120 reads over the 24 samples, and with a minimal frequency of 0.2 and minimal quality of 20 were conserved. The output was obtained in a Variant Call Format (vcf file), which was then modified for the statistical analysis with a custom Python script.

#### Statistical analysis of the detected polymorphisms: hierfstat

The list of polymorphisms was analyzed to determine which loci contributed to population structure and provided a distinction between the two WCR strains Standard and Hopkinton. F-statistics, introduced by Wright (1951) and redefined by Weir and Cockerham (1984) are population structure estimators. The F-statistics estimators and frequency of an allele are relative, the measures of variance of this frequency between populations and between individuals within populations can thus be estimated using the F-statistics.

Measure of the variance of allele frequencies between populations and between individuals within populations was performed with an ANOVA-type analysis, implemented in the package hierfstat (Goudet 2005) for R (R Development Core Team 2013). This package allows the estimate of (hierarchical) F-statistics from hierarchical factors, using Yang’s algorithm (Yang 1998) allowing to set any number of levels, to test for population differentiation, by a generalized likelihood-ratio test. In the present study, the package was used to extract loci of interest from the Freebayes output, with outlier F-statistic values between susceptible and Hopkinton input samples, hence with high variances in allele frequencies between populations.

The input data for the analysis was obtained from the vcf file with a custom Python script, which also conserved only biallelic loci, since SNPs are mostly biallelic due to the low frequency of single nucleotide substitutions (Vignal et al. 2002). Each read was considered as an individual in a population of reads in the analysis, carrying either the reference allele (“1”) or the alternative allele (“2”), belonging to a given sample (factor “sample”, levels 1 to 24), either from the susceptible population (factor “population status”, level 1) or the resistant population (level 2), creating a hierarchy of factor levels with a coding specific to each type of read. Each potential polymorphic locus was analyzed separately and as haploid, for consistency with parameters set in Freebayes assuming non-diploid samples due to larval pooling in each sample, and since each read (“individual”) carries only one allele at a time.

The analysis was then performed to determine the influence on the strain of allele frequency variations, per locus, measuring the variance component (variance of the allele frequency) for each polymorphism, to calculate the hierarchical F-statistics for each locus. The output was represented in a boxplot. The outliers, with FST values outside the main distribution, were conserved and considered as the loci with the most different frequencies between the two tested strains, or populations of reads. A BLAST2GO (Conesa et al. 2005) analysis was performed to obtain annotations of the outlier transcripts where polymorphisms were identified.

#### Confirmation of the polymorphisms: MassArray

##### Biological material

Loci detected by Freebayes and exhibiting high (outlier) FST values in the hierfstat analysis, and thus potentially representing alleles with variable frequencies between susceptible and resistant populations, were confirmed in the two strains used for their discovery. They were also tested to assess their suitability as markers of the resistance to Cry3Bb1 in other strains of WCR.

The resistant Hopkinton strain, used in RNA sequencing and polymorphism discovery, described above, exhibits a strong linkage disequilibrium between the region potentially including the gene(s) of resistance and nearby loci (Flagel et al. 2015). This strain was thus a good candidate for SNP discovery since markers linked with the locus (or loci) responsible for resistance might be used as markers of the resistance in other populations. To confirm the difference in allele frequencies with a susceptible population, the non-diapausing Standard colony (Branson 1976) was chosen as susceptible counterpart. Standard was also used in the RNA-seq and thus in the polymorphism discovery. It was also the strain used for the introgression of the non-diapausing trait in Hopkinton, to facilitate laboratory rearing (Gassmann et al. 2011).

Two other colonies were provided by Dr. Lance Meinke at the University of Nebraska Lincoln. They were collected from two fields in Nebraska, “Jackson” from eastern NE, and “Keith” from western NE, showing greater than expected injury in Cry3Bb1-expressing corn fields (Wangila 2016), with an average root injury rating >1.00, as defined by the EPA (EPA 2011). Their resistance status was confirmed in Wangila and Meinke (2016) and Wangila 2016 by single plant bioassay method, described in Gassmann et al. (2011, 2014). They were then crossed with a non-diapausing susceptible colony for rearing, resulting in a non-diapausing offspring (G1). A subsequent generation of WCR was obtained for each colony to increase population sizes. These two first laboratory-reared generations were maintained on isoline corn. The following generation, each of the two colonies was split int two sub-colonies, one reared continuously on a *Bt*-corn hybrid expressing the full length Cry3Bb1 protein, Stone 6021VT3 corn (“*Vt3*”), and the other reared on the near isoline hybrid Stone 6021RR2, for 4 subsequent generations. Both seed types were provided by Monsanto Company (Saint Louis, MO). Adults all fed on non-transgenic corn. An additional resistant population from Minnesota called “Hills”, described as resistant in Zukoff et al. (2016), was provided by Dr. Bruce Hibbard from the University of Missouri / USDA-ARS. A similar rearing procedure was applied for Hills and the Nebraskan populations, starting with a cross of the resistant individuals from the field with a susceptible non-diapausing strain. The non-diapausing offspring was reared for one generation on isoline to increase population size and the following generation was split on either *Vt3* (reselected) or isoline (non-selected colony) corn as larvae, and exclusively isoline corn as adults, for 8 additional generations. Thirty adults, 15 females and 15 males, from each of the three Cry3Bb1-reselected colonies and their unselected counterparts, were stored at 4°C in 95% ethanol for genomic DNA extraction.

##### DNA Extraction method

Genomic DNA extraction was performed using DNAzol® (Molecular Research Center, Inc., Cincinnati, OH). DNAzol® is an extraction buffer solubilizing cellular components and allowing genomic DNA precipitation in presence of ethanol (Chomczynski et al. 1997). The protocol was adapted from Tabashnik et al. (2005). The specimens contained in 95% ethanol were rinsed 3 times in a solution of NaCl at 0.065% to eliminate the residual ethanol from the tissues. The abdomen of each individual was removed to avoid contamination by DNA from bacteria contained in the gut, but also from fertilized eggs or stored sperm in female’s reproductive tract, and allowing sex confirmation to keep a consistent sex ratio for each population. The head and thoracic regions were then crushed with a pestle in 980 µL of DNAzol® and 10 µL of PolyAcryl Carrier (Molecular Research Center, Inc., Cincinnati, OH) to increase the recovery of DNA, then incubated for 1 hr with 10 µL of RNase A (AkronBiotech, Boca Raton, FL) at 37°C. The mixture was then treated with 8 µL of proteinase K (Promega, Madison, WI) at 20 ng/ µL at 55°C during 1 hr before incubation at 4°C overnight. After several centrifugations and rinses with 100% and 70% ethanol, the DNA pellets were re-suspended in 100 µL of low EDTA TE buffer at pH 8. Low EDTA buffer was used to avoid chelation of Mg2+ ions, which are co-factors of the *Taq* polymerase, used for DNA amplification, by EDTA. The concentration of the resulting DNA samples was measured with a NanoDrop 2000 UV spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA). Each sample was then diluted to reach 20ng/µL for the experiment, using low EDTA buffer at pH 8.

##### Primer design and selection of loci

The detection of potential polymorphisms relied on a transcriptome. However, genomic data was needed to design primers since the template for the MassArray experiment is gDNA. Transcript sequences were aligned to the draft genome of WCR (H.M. Robertson, personal communication) using the software Splign (Kapustin et al. 2008) to perform a cDNA to genome spliced alignment of the contigs containing the loci of interest, against the draft genome. This alignment was tolerant of sequencing errors and polymorphisms. About 100-200 bases around the polymorphic locus were provided for the primer design based on the gDNA sequences. Some contigs did not have a corresponding sequence in the genome and the transcript sequence was therefore used for primer design. Assay Design Suite (Agena Bioscience Inc., San Diego, CA) was used for primer and multiplex designs based on assay compatibility within a multiplex.

##### MassArray experiment and statistical analysis

The experiment was performed on the Agena Bioscience MassArray System at the UIC Core Genomics Facility. Two multiplexes were used, one including primers for 27 loci and another amplifying 24 loci. Each primer pair generated a short amplicon, between 80 and 120 bp. Samples were placed in a 384-well plate. The first step of the experiment was a PCR amplification, based on the designed assays, followed by a SAP (shrimp alkaline phosphatase) reaction cleanup that removes unincorporated nucleotides and primers. The third step was a single-base extension using mass-modified dideoxynucleotide terminators in order to identify the locus-specific alleles by their mass. Finally, the extension products were submitted to a mass spectrometry analysis, differentiating the mass of the two different alleles by matrix-assisted laser desorption ionization - time-of-flight (MALDI-TOF) method (Gabriel et al. 2001; Agena Bioscience 2015).

The genotypes resulting from the MassArray experiment were analyzed to detect, and remove from the analysis, loci deviating from the Hardy-Weinberg equilibrium. A test of departure from the H-W equilibrium is often used in genotyping experiments as a quality control method, since it is sensitive to genotyping errors, which are a source of overestimation of the population structure (Morin et al. 2009; Anderson et al. 2010). The statistical test answers the question: “Are genotype frequencies observed in a sample compatible with those expected under H-W equilibrium?” (Waples 2015). Exact tests for Hardy-Weinberg equilibrium were computed with GENEPOP version 4.4 (Raymond and Rousset 1995; Rousset 2008). The results of the Hardy-Weinberg equilibrium tests were corrected for multiple testing, by locus, for each population, with a Bonferroni correction.

A graphical representation was chosen to visualize separation of the assayed populations based on their genotypes. Discriminant Analysis of Principal Components (DAPC) is a multivariate statistical method providing a visual identification of clusters of genetically similar individuals by maximizing the difference between clusters while minimizing the difference within the clusters (Jombart et al. 2010). The method transforms data using a Principal Component Analysis and the clusters generated are identified using a discriminant analysis, based on the clusters known *a priori* (populations). The loci which passed the H-W equilibrium quality control were analyzed with the package adegenet (Jombart 2008), offering a graphical representation of population differentiation by DAPC. This package also allows a reassignment of each individual to its likely population of origin based on the conserved principal components of the DAPC that are considered as discriminant. Principal components are the association of alleles allowing to differentiate the populations by maximizing the variance between the clusters, also called discriminant functions, and are represented by the axes.

**References cited**

Agena Bioscience, 2015. iPLEX-Reagents-Brochure. Agena Biosci.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ, 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT, 2010. Data quality control in genetic case-control association studies. Nat. Protoc. 5, 1564–1573.

Benjamini Y, Hochberg Y, 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B Methodol. 57, 289–300.

Branson TF, 1976. The selection of a non-diapause strain of *Diabrotica virgifera* (Coleoptera: Chrysomelidae). Entomol. Exp. Appl. 19, 148–154.

Broad Institute, 2017. Picard [WWW Document]. URL http://broadinstitute.github.io/picard/

Chomczynski P, Mackey K, Wilfinger W, 1997. DNAzol: a reagent for the rapid isolation of genomic DNA. BioTechniques 22, 550–553.

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M, 2005. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21, 3674–3676.

Eyun S-I, Wang H, Benson AK, 2014. Molecular evolution of glycoside hydrolase genes in the western corn rootworm (*Diabrotica virgifera virgifera*). PLOS ONE 9.

Flagel LE, Swarup S, Chen M, Bauer C, Wanjugi H, Carroll M, Hill P, Tuscan M, Bansal R, Flannagan R, Clark TL, Michel AP, Head GP, Goldman BS, 2015. Genetic markers for western corn rootworm resistance to *Bt* toxin. G3 Genes Genomes Genet. 5, 399–405.

Gabriel S, Ziaugra L, Tabbaa D, 2001. SNP genotyping using the Sequenom MassARRAY iPLEX platform, in: Current Protocols in Human Genetics. John Wiley & Sons, Inc.

Garrison E, Marth G, 2012. Haplotype-based variant detection from short-read sequencing. ArXiv12073907 Q-Bio.

Gassmann AJ, Petzold-Maxwell JL, Clifton EH, Dunbar MW, Hoffmann AM, Ingber DA, Keweshan RS, 2014. Field-evolved resistance by western corn rootworm to multiple *Bacillus thuringiensis* toxins in transgenic maize. Proc. Natl. Acad. Sci. U. S. A. 111, 5141–5146.

Gassmann AJ, Petzold-Maxwell JL, Keweshan RS, Dunbar MW, 2011. Field-evolved resistance to *Bt* maize by western corn rootworm. PloS One 6, e22629–e22629.

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A, 2003. ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 31, 3784–3788.

Goudet J, 2005. hierfstat, a package for R to compute and test hierarchical F-statistics. Mol. Ecol. Notes 5, 184–186.

Jombart T, 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24, 1403–1405.

Jombart T, Devillard S, Balloux F, 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. 11, 94.

Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-Y, Lopez R, Hunter S, 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30, 1236–1240.

Kapustin Y, Souvorov A, Tatusova T, Lipman D, 2008. Splign: algorithms for computing spliced alignments with identification of paralogs. Biol. Direct 3, 20.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079.

Li W, Cowley A, Uludag M, Gur T, McWilliam H, Squizzato S, Park Y, Buso N, Lopez R, 2015. The EMBL-EBI bioinformatics web and programmatic tools framework. Nucleic Acids Res. 43, W580-4.

McWilliam H, Li W, Uludag M, Squizzato S, Park Y, Buso N, Cowley A, Lopez R, 2013. Analysis Tool Web Services from the EMBL-EBI. Nucleic Acids Res. 41, W597-600.

Morin PA, Leduc RG, Archer FI, Martien KK, Huebinger R, Bickham JW, Taylor BL, 2009. Significant deviations from Hardy–Weinberg equilibrium caused by low levels of microsatellite genotyping errors. Mol. Ecol. Resour. 9, 498–504.

R Development Core Team, n.d. R Development Core Team (2013). 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.

Raymond M, Rousset F, 1995. GENEPOP (Version 1.2): Population genetics software for exact tests and ecumenicism. J. Hered. 86, 248–249.

Rousset F, 2008. genepop’007: a complete re-implementation of the genepop software for Windows and Linux. Mol. Ecol. Resour. 8, 103–106.

Tabashnik BE, Biggs RW, Higginson DM, Henderson S, Unnithan DC, Unnithan GC, Ellers-Kirk C, Sisterson MS, Dennehy TJ, Carrière Y, Morin S, 2005. Association between resistance to *Bt* cotton and cadherin genotype in pink bollworm. J. Econ. Entomol. 98, 635–644.

Vignal A, Milan D, SanCristobal M, Eggen A, 2002. A review on SNP and other types of molecular markers and their use in animal genetics. Genet. Sel. Evol. GSE 34, 275–305.

Wangila DS, 2016. Resistance management of western corn rootworm, *Diabrotica virgifera virgifera* LeConte to *Bt* corn traits in Nebraska. PhD Entomology Dissertation. Univ. Neb. Lincoln, Nebraska.

Wangila DS, Meinke LJ, 2016. Effects of adult emergence timing on susceptibility and fitness of Cry3Bb1-resistant western corn rootworms. J. Appl. Entomol. n/a-n/a.

Waples RS, 2015. Testing for Hardy–Weinberg proportions: Have we lost the plot? J. Hered. 106, 1–19.

Weir BS, Cockerham CC, 1984. Estimating F-statistics for the analysis of population structure. Evolution 38, 1358–1370.

Wright S, 1951. The genetical structure of populations. Ann. Eugen. 15, 323–354.

Yang R-C, 1998. Estimating hierarchical F-statistics. Evolution 52, 950–956.

Zukoff SN, Ostlie KR, Potter B, Meihls LN, Zukoff AL, French L, Ellersieck MR, French BW, Hibbard BE, 2016. Multiple assays indicate varying levels of cross resistance in Cry3Bb1-selected field populations of the western corn rootworm to mCry3A, eCry3.1Ab, and Cry34/35Ab1. J. Econ. Entomol. 109, 1387–1398.