



**HAL**  
open science

## Genetic basis of codling moth (*Cydia pomonella*) resistance to the original isolate of *C. pomonella* Granulovirus (CpGV-M)

Jérôme Olivares, Myriam Siegwart, Mathieu Gautier, Sandrine Maugin, Léa Gingueneau, Bertrand Gauffre

### ► To cite this version:

Jérôme Olivares, Myriam Siegwart, Mathieu Gautier, Sandrine Maugin, Léa Gingueneau, et al.. Genetic basis of codling moth (*Cydia pomonella*) resistance to the original isolate of *C. pomonella* Granulovirus (CpGV-M). *Entomologia Generalis*, 2023, 43 (3), pp.649-658. 10.1127/entomologia/2023/2052 . hal-04224150

**HAL Id: hal-04224150**

**<https://hal.science/hal-04224150>**

Submitted on 10 Oct 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License



# Genetic basis of codling moth (*Cydia pomonella*) resistance to the original isolate of *C. pomonella Granulovirus* (CpGV-M)

Jérôme Olivares<sup>1</sup>, Myriam Siegwart<sup>1,\*</sup>, Mathieu Gautier<sup>2</sup>, Sandrine Maugin<sup>1</sup>, Léa Gingueneau<sup>1</sup>, and Bertrand Gauffre<sup>1</sup>

<sup>1</sup> INRAE, Plants and Cropping Systems in Horticulture research unit (PSH), 228 route de l'aérodrome, 84914 Avignon, France

<sup>2</sup> CBGP, INRAE, CIRAD, IRD, Montpellier SupAgro, Université de Montpellier, Montpellier, France

\* Corresponding author: myriam.siegwart@inrae.fr

With 2 figures and 2 tables

**Abstract:** The utilization of insect pest pathogens as biocontrol agents can serve as an effective alternative to chemical pesticides. However, pests can develop resistance to these pathogens, similar to resistance observed with chemicals. Identifying the genetic bases of resistance is a crucial step in comprehending the co-evolution mechanisms between insect hosts and their pathogens, thereby enhancing the durability of these biocontrol methods. *Cydia pomonella*, a major pest of apple and pear trees found across all continents, can be effectively controlled in commercial orchards through the use of the granulovirus, CpGV. However, since 2005, resistance to the CpGV-M isolate has been reported in Europe, leading to significant challenges in managing *C. pomonella* in organic orchards. Studies conducted on CPGV-M resistance have revealed that viral replication is blocked in resistant phenotypes, and resistance is linked to a major gene located on the Z sex chromosome. Nevertheless, this major gene appears to be insufficient in explaining resistance to CPGV-M, as other forms of resistance involving autosomal genes have been identified subsequently. The objective of this study was to gain insights into the genetic architecture of *C. pomonella* resistance to CpGV-M by conducting a genome-wide association study using newly generated whole-genome resequencing data from pools of individuals (Pool-Seq) sampled from 13 susceptible and resistant populations across four regions in France and Italy. Through this analysis, eight candidate regions were identified, including one located on the Z sex chromosome. These eight regions encompass a total of 27 lncRNAs (long non-coding RNAs) and 40 genes, of which three are known to be involved in insects immune pathways.

**Keywords:** *Cydia pomonella granulovirus*, Genome-Wide Association Studies (GWAS), insect immunity, host-virus interaction, resistance

## 1 Introduction

While chemical pesticides have been successfully employed for years, the current societal and political demand for a reduction in their use is strong. In this context, the use of insect pathogens as biocontrol agents could be a valuable alternative as illustrated by some bacteria [mainly various strains of *Bacillus thuringiensis* (Bt)], entomopathogenic fungi (e.g., *Beauveria bassiana* and *Metarhizium* species) or viruses such as baculoviruses. However, strong directional selection imposed by repeated use of a microbial control agent might lead to a rapid evolution and increase in frequency of resistance alleles in pest populations (Tabashnik et al. 2009). It is thus particularly important to identify the natural variation in resistance of insect populations and the

mechanisms of resistance as pathogens evolution could be managed to counter resistance mechanisms developed by the host (Cory & Franklin 2012).

The advance of evolutionary population genomics facilitated by high-throughput sequencing technologies has expanded the genomic scale of studies investigating the evolution of resistance and the co-evolutionary arms race of pathogen evasion and host resistance in non-model species (Anderson et al. 2010). In particular, genome scans represent an efficient approach to target genomic regions involved in adaptation, as they analyze genome-wide variations in order to detect locus-specific signatures of positive directional selection (Storz 2005). Pool-Seq experiments that consist of sequencing pools of individual DNAs provide cost-effective alternatives to facilitate and even improve allele frequency

estimation at genome-wide markers (Gautier 2015). In addition, methods that take into account population structure may allow to study genomic adaptation by investigating a large range of populations from different areas (Gautier 2015). In this study, we applied these population genomics approaches to investigate virus resistance in an insect pest.

The baculovirus *Cydia pomonella granulovirus* (CpGV) is an important and commercially successful biocontrol agent, which is used against larvae of the codling moth, *Cydia pomonella*, a major global pest of pome fruits and walnuts (Lacey et al. 2008). The CpGV occlusion bodies (OBs) can be sprayed in their natural form, and cause larval death within 4–6 days (Lacey et al. 2008). It is harmless to non-target animals and has no detrimental impact on the environment (Lacey et al. 2008). The first CpGV products were commercialized in the early 1990s from the isolate CpGV-M discovered in Mexico (Huber & others 1990). Since then, CpGV has become a cornerstone of *C. pomonella* control in both integrated pest management and organic orchards (Lacey et al. 2008). However, in 2005, field resistant *C. pomonella* populations to CpGV-M were detected in Germany and France (Sauphanor et al. 2006). CpGV-M resistance rapidly spread to all apple production basins in France (Siegwart et al. 2017) and all over Europe (Zichová, et al. 2013). Some of these field populations could be 1,000 to 100,000 times less susceptible to CpGV-M compared to susceptible strains (Sauphanor et al. 2006). Following this discovery, resistance-breaking CpGV isolates were successfully tested and commercialized (Berling et al. 2009).

Resistance to CpGV-M in *C. pomonella* was determined to be associated with a major dominant gene linked to the Z chromosome, resulting in an early inhibition of viral DNA replication within cells (Asser-Kaiser et al. 2011). However, the contribution of other genes was suggested to fully explain CpGV-M resistance in a French *C. pomonella* resistant laboratory strain (Berling et al. 2013). These studies suggest a complex picture of CpGV-M resistance, which could potentially be linked to complicated molecular mechanisms and most likely acts on one of its immune pathways.

The aim of this study was to provide insights into the genetic architecture of the acquired resistance of *C. pomonella* genomic regions to CpGV-M infection. To grasp a large range of variation associated with CpGV-M resistance, we sampled 13 populations in four areas of France and Italy, including at least one resistant and one susceptible population in each area. For each population, pools of individual DNAs were sequenced. We investigated their population genetic structure and used a model-based approach to test for adaptive differentiation and association with the estimated level of resistance of the corresponding populations (Gautier 2015). We took advantage of the recently published chromosome-level and annotated assembly of the *C. pomonella* genome to identify candidate genes (Wan et al. 2019).

## 2 Material and methods

### 2.1 Wild populations and laboratory strains of *C. pomonella*

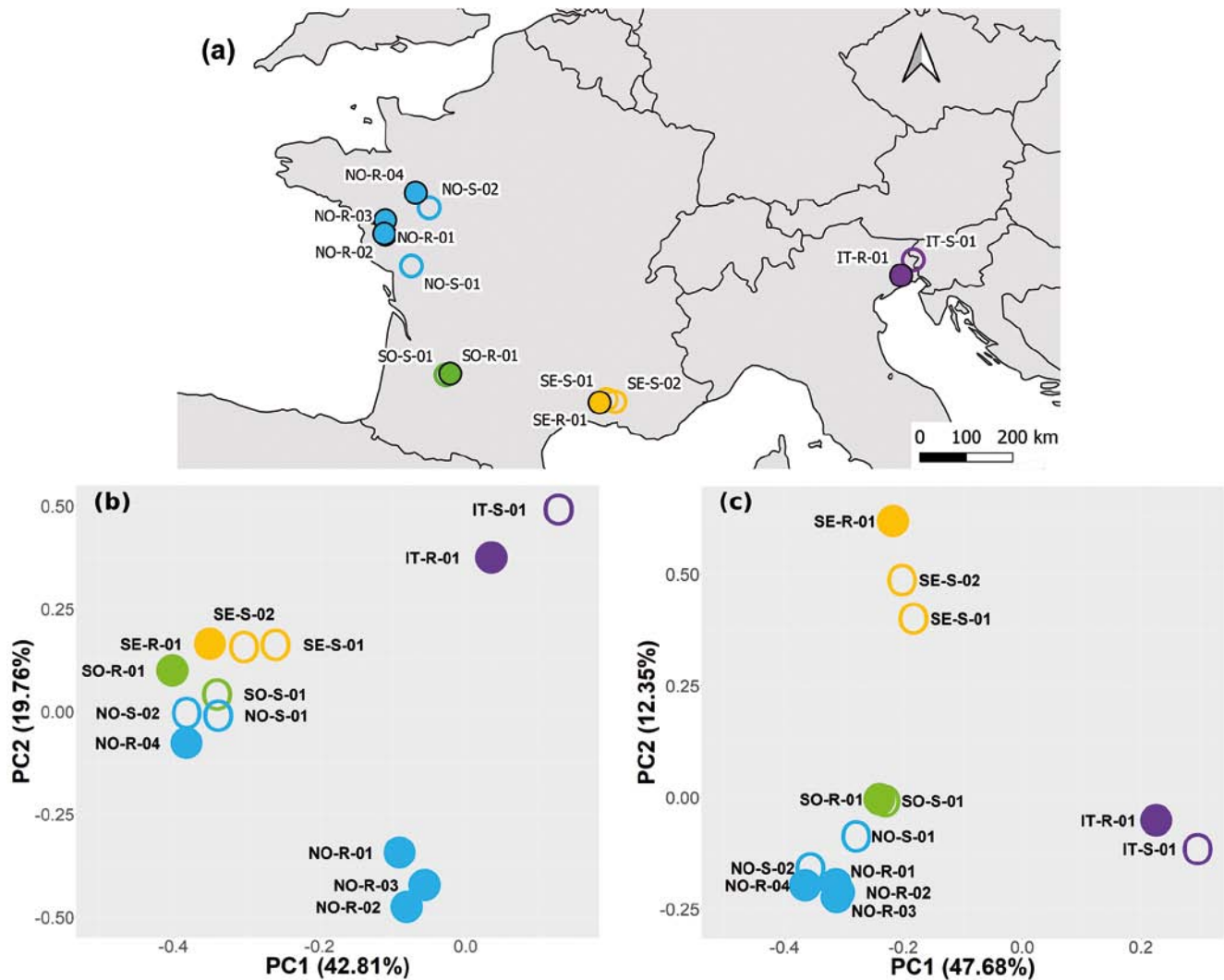
To grasp a large range of genomic variation associated with CpGV-M resistance, wild populations were sampled in 4 distant geographical areas in France and Italy (>300 km, Fig. 1 and Table S1). Four populations were sampled in 2017, 1 in 2018 and 8 in 2019 (Table S1). Fifth instar larvae were trapped in commercial orchards with cardboard bands that were wrapped around trunks of apple trees in early summer and collected in autumn. After 4 months at 4°C to break diapause, larvae were placed in a rearing room at 23°C with a photoperiod of 16 hours of light and 8 hours of night to induce pupation and moth emergence. Adults were sexed at emergence and placed in plastic containers for reproduction. Finally, adults were placed in 90° alcohol for genomic analysis while eggs were collected to perform bioassay by virus ingestion on neonatal larvae.

### 2.2 Discriminating dose bioassays

Neonatal larvae aged between 0 and 12h were used for bioassays since it is the stage targeted by CpGV. They were individually deposited in 96 microwell plates, each microwell containing approximately 150 µL of artificial diet (Stonefly diet Industries Ltd, Rochester, NY) and treated on the diet surface with 6 µl of a solution containing 6,250 viral particles (GV)/µl of CpGV-M diluted in osmosed water, or water for control. This dose is considered discriminating as it kills 100% of susceptible individuals. Plates were incubated in climatic chambers at 23°C and 16D:8N photoperiod and mortality was measured after 7 days. A larva was considered dead if it did not respond to a probe with dissecting forceps. Missing larvae were removed from the data (< 1%). Each test consisted of a minimum 24 individuals per modality (control or treated). Corrected Mortality (CM) was calculated using Abbott formula (Abbott 1925). Populations with CM < 60% were considered resistant. CM value was used as a fine population-specific covariable of resistance level to CpGV-M infection in the association testing analysis.

### 2.3 Total genomic DNA extraction and sequencing

All DNA extraction was conducted using the Qiagen Blood & Cell Culture DNA Mini Kit. To account for average difference in size, males and females total genomic DNA was extracted separately: thorax were dissected from each adult moth, cut in half and pooled up to 50 to be digested overnight at 56°C in 10 ml final volume of G2 buffer, all subsequent steps were made following the manufacturer's instructions. DNA integrity and purity were assessed on 1% agarose gel electrophoresis and by the A260/280 ratio, using a Nanodrop 1000 spectrophotometer (Thermo Scientific). Accurate DNA quantity was assessed using the PicoGreen dsDNA Assay Kit (Molecular Probes). For each sample, males and females



**Fig. 1.** Location and genetic structure of the 13 wild populations sampled (a) and representation in a two-dimension plot of genetic structure based on a eigen-decomposition of the scaled covariance matrix of the population allele frequencies ( $\Omega$ ) for (b) the Z chromosome and (c) autosomes. Filled dots represent resistant populations and empty dots represent susceptible populations (different colors represent different regions).

pools were mixed to reach an equimolar concentration of each individual in the final DNA pool.

The 5 DNA pools from populations sampled in 2017 and 2018 were paired-end sequenced by the Macrogen company (Korea) on an Illumina Hi-Seq TenX instrument providing  $2 \times 150$  bp reads. The remaining 8 DNA pools from populations sampled in 2019 were paired-end sequenced ( $2 \times 150$  bp) by BGI (Hong Kong) using whole genome DNBseq sequencing technologies. The Pool-Seq data were deposited in the Sequence Read Archive repository under the BioProject accession number PRJNA942582.

## 2.4 SNP calling and filtering

Raw reads (SRA accession number SAMN33695182 to SAMN33695194) were trimmed by quality using TrimGalore

v.0.6.5 (Krueger 2015) with a Q30 quality cutoff and the following DNBseq adaptators sequences

```
:-a AAGTCGGAGGCCAAGCGGTCTTAGGAAGACAA
-a2 AAGTCGGATCGTAGCCATGTCGTTCTGTGAG
CCAAGGAGTTG.
```

The software BWA mem v.0.7.17 (Li & Durbin 2009) was used to align the filtered paired reads to the *C. pomonella* reference genome V1 (<http://v2.insect-genome.com/Organism/224>). The resulting bam files were sorted using SAMtools v. 1.14 (Li et al. 2009) and duplicates were removed using the *RemoveDuplicates* function from Picard's Tools v.2.20.7 (<https://broadinstitute.github.io/picard/>). The 13 pools filtered BAM files were finally combined using the mpileup program in SAMtools to perform variant calling with Varscan *mpileup2cns* v2.3.6 (options: --min-coverage

10 (--min-var-freq 0.05 --min-avg-qual 20 --p-value 0.8) (Koboldt et al. 2012)

The resulting VCF file was subsequently filtered with the `vcf2pooldata` function from the R package *poolfstat* v1.1 (Gautier et al. 2022) retaining only biallelic variants (including insertions and deletions but hereafter referred to as SNPs), covered by > 4 reads, < 99.9<sup>th</sup> overall coverage percentile in each pool and with an overall MAF > 0.05. The *poolfstat* function `pooldata2genobypass` was then used to generate the input files for association analyses.

## 2.5 Genome scan for association with resistance level

Genome-wide scans for association with the estimated resistance level to CpGV-M infection of the population represented by the pools were performed under the standard covariate model implemented in BayPass version 2.3 (Gautier 2015). Following the manual recommendations, we capitalized on the large number of available SNPs by using a subsampling strategy consisting of dividing each data set into sub-datasets containing only one SNP every 4 SNPs for Z chromosome and one every 100 for autosomes representing approximately 80K SNP per subset. These sub-datasets were further analyzed in parallel under the BayPass standard covariate model using default options for the Markov Chain Monte Carlo (MCMC) algorithm (including `-scalecov` options to ensure proper scaling of the population covariable). Three independent runs (using the option `-seed`) were performed for each dataset. The estimated model hyperparameters were highly consistent across both runs and datasets, including in particular the scaled covariance matrix of population allele frequencies ( $\Omega$ ) that quantifies the genetic relationship among each pair of populations. Support for association of each SNP with the population resistance level was then assessed using the median Bayes Factor (BF) computed over the three independent runs. BFs were reported in deciban units (db) with values between 15 db and 20 db providing “very strong evidence” and values >20 db providing “decisive evidence” in favor of association according to Jeffrey’s rule (Jeffreys 1998). We also scan the genome for overall adaptive differentiation (that may be related to other selective pressures than the one imposed by CpGV-M) based on the XtX statistics which may be viewed as a SNP-specific Fst-like that account for the confounding effect of population structure (Gautier 2015). SNP-specific XtX were also estimated as the median estimated values over the three independent runs of the program. We finally deemed SNPs as candidate for association if they both displayed a BF  $\geq$  15 db and an estimated XtX > 99% quantile of the empirical XtX distribution. Visualization plots and results manipulations were computed using standard packages in R (R Core 2013).

## 2.6 Population genetic structure

Genetic differentiation between all populations was estimated with SNP markers using pairwise  $F_{ST}$  which were

estimated using the estimator for PoolSeq data implemented in *poolfstat* function `computePairwiseFSTmatrix`. Pairwise  $F_{ST}$  between each population pair were estimated separately with autosomal SNPs and SNPs on the Z chromosome. We drew up dendrograms using the R function *hclust* from the *stats* package and isolation by distance was tested by regressing pairwise estimates of  $F_{ST}$  against the logarithm of the geographical distances using a Mantel test (10,000 permutations). The spectral decomposition of the scaled covariance matrix of population allele frequencies ( $\Omega$ ) estimated with Baypass was also used to describe population genetic structure.

## 2.7 Identification and annotation of candidate genomic regions

To identify genomic regions containing the highest concentrations of SNPs potentially related to resistance level to CpGV-M infection we used overlapping sliding windows of 50 kb with an offset of 5 kb and focused our analysis on genes located within windows that contained at least 5 highly associated SNPs (Fig. 2). Gene annotation, ontology and LncRNA list were obtained from the first version of the *C. pomonella* chromosome-level annotated genome (CpomGenome-1), which can be accessed at <http://v2.insect-genome.com/Organism/224>. Genes predictions were individually checked using Blast analysis. GO annotations were carried out on the QuickGO platform (QuickGO 2023) and the gene names were determined by ortholog search on the UniProt platform (UniProt Consortium 2023). The localization (e.g., intronic, exonic, intergenic) and effect (e.g., synonymous vs. non-synonymous) of the variants were predicted using a local database and SnpEff 5.0c (Cingolani et al. 2012), the putative localization of LncRNAs along the *C. pomonella* reference genome was deduced by BLAST analyses.

# 3 Results

## 3.1 Discriminating dose bioassays

Bioassays performed at the discriminating dose (6,250 GV/ $\mu$ l) showed a gradient of susceptibility to CpGV-M infection among the sampled populations, with corrected mortality (CM) ranging from 8.35% to 100% (Table S1). Based on a threshold of 60% on CM, we considered our dataset to be composed of 6 susceptible and 7 resistant populations, including at least one resistant population in each geographical area studied (Fig. 1 and Table S1).

## 3.2 Sequencing, mapping and variant calling

Sequencing generated an average of 275,437,098 reads per library with an average depth of coverage of 45X ranging from 18 to 70X. The low depth of coverage in some samples was due to a lower mapping ratio. Blast analysis of the unmapped reads revealed that these reads were primarily

from bacteria, likely due to partial degradation of some samples (data not shown). All libraries exhibited almost the same breadth of coverage around 84% of the 763Mb *C. pomonella* reference genome (see details in Table S2).

Among the 29,706,673 variants detected across the 13 pools, 8,511,535 were biallelic, present in all pools (no missing data), and were selected using the R package *poolfstat* for subsequent analyses.

### 3.3 Population genetic structure

The pairwise  $F_{ST}$  estimated with autosomal SNPs ranged from 0.002 to 0.141.  $F_{ST}$  between pairs of French populations and between the two Italian populations were relatively low ( $<0.055$ , average  $F_{ST} = 0.03$ ) while  $F_{ST}$  involving French and Italian population pairs were important ( $>0.055$ , average  $F_{ST} = 0.106$ ). Autosomal SNPs supported an isolation by distance pattern with a positive relationship between  $F_{ST}$  and the logarithm of geographic distance (slope = 0.022 Mantel test  $p < 0.001$ , Fig. S1). A consistent picture, reflecting geography, was obtained from the spectral decomposition of the scaled covariance matrix of population allele frequencies ( $\Omega$ ) estimated with BayPass on autosomal SNPs (Fig. 1c). Interestingly, SNPs on the Z chromosome showed slightly lower  $F_{ST}$  among populations (range 0.013 – 0.123) and the dendrogram constructed from the  $F_{ST}$  (Fig. S2) showed 4 clusters, one with the two Italian populations, one with 3 resistant populations from NO-France, one with 4 susceptible populations from all 3 French areas and the last one with 3 resistant populations from NO and SE-France plus one susceptible population from SO-France (Fig. S2). Isolation by distance pattern also differed from those based on autosomal SNPs (slope = 0.003, Mantel test  $p < 0.001$ , Fig. S1). Consistently,  $\Omega$  estimated for the chromosome Z revealed a complex pattern separating populations according to both their geographical origin and their resistance status (Fig. 1b). In particular, French populations formed two clusters, one including 3 resistant populations from NO-France (out of 4), and the other including the other French populations (including both susceptible and resistant populations).

### 3.4 Genome scan for association with resistance level

Based on median values computed over the three replicated BayPass analyses, the top 1% posterior mean of XtX was 20.01 and included 85,115 SNPs. Among them, 488 had a BF  $\geq 15$  dB (max = 49.05) and were selected as candidates. Our sliding window approach reduced to 8 the number of genomic regions strongly associated with resistance level to CpGV-M infection (Fig. 2). One was located on the Z chromosome and 7 were located on autosomes 2, 3, 4, 15, 16, 19 and 21 (Fig. 2).

### 3.5 Functional annotation analysis

Out of the 488 candidate SNPs (Table S4), i) 16 were annotated as exonic; ii) 33 and 30 were located within 5kb upstream

the 5' end or downstream the 3' end of a gene respectively; iii) 97 were intronic; and iv), 312 were intergenic (Table 1a). The 8 genomic regions strongly associated with resistance level to CpGV-M infection accounted, in total, for 25.8% of the 488 candidate SNPs and included from 6 to 30 SNPs (Table 1b). Interestingly, they were enriched in exonic variants (8.7% of the candidate SNPs within the 8 genomic regions compared to 3.3% over all candidate SNPs were exonic, Table 1a).

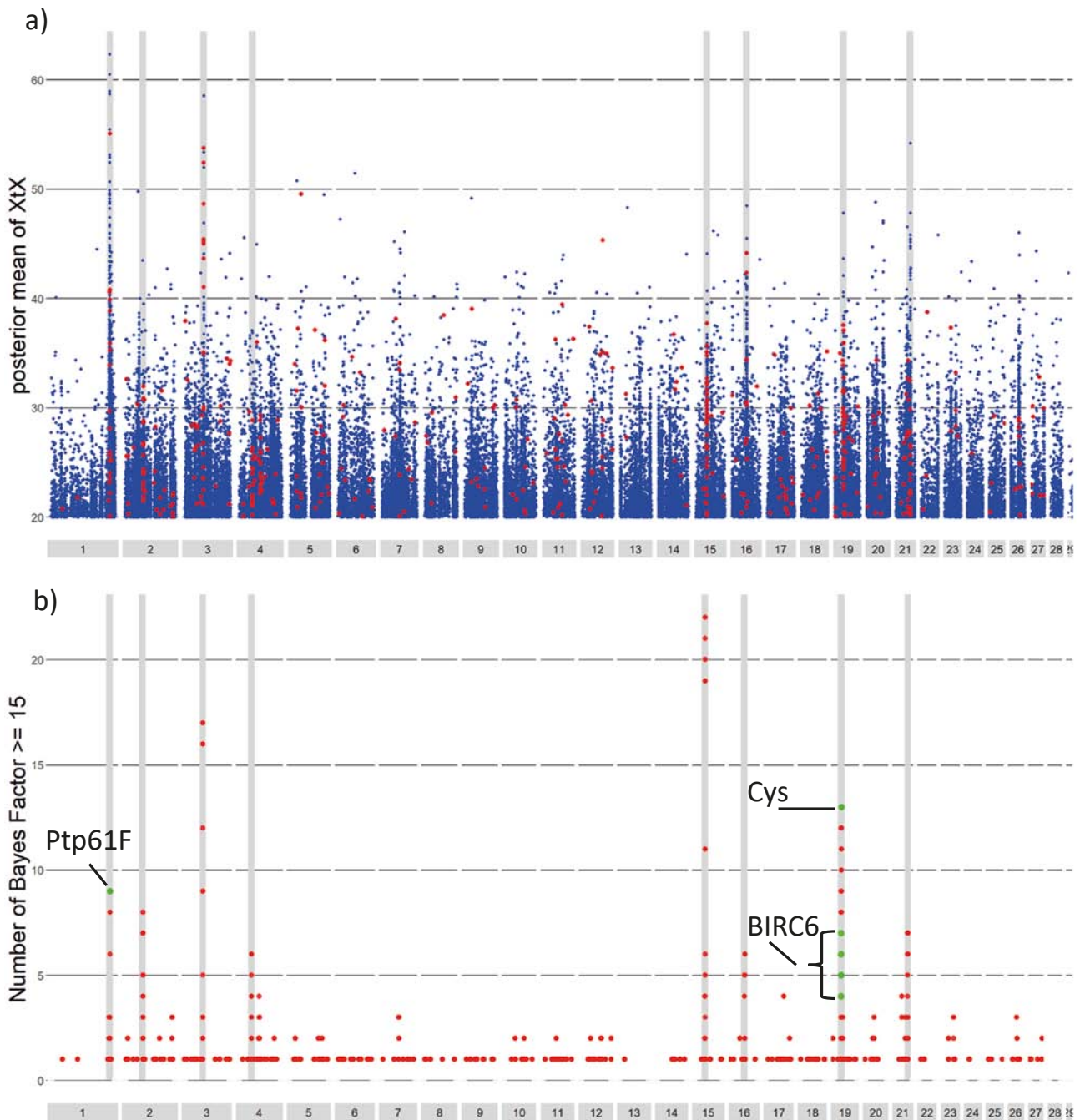
According to published *C. pomonella* annotation, we found 40 coding proteins (11 of them are “uncharacterized protein”) and 27 putatives lncRNA within the 8 genomic regions strongly associated with resistance (Table S3). Of these, 3 have been described as having a role in immune pathways (Table S3). Two were found on chromosome 19: *BIRC6* (spread over 3 predicted ORF: cpo034700 + cpo034720 + cpo034750), a baculoviral IAP repeat-containing protein, and *Cys* (cpo034660), a cystatin-like protein (Table S3). One was located on the Z-chromosome: a tyrosine-protein phosphatase non-receptor type 61F gene (*Ptp61F*, cpo122180). A lncRNA (lnc004846) is present 15 to 20 kb from this gene and contain 9 candidates SNPs.

## 4 Discussion

Our genome-wide association study, conducted on samples from wild populations collected in four regions of France and Italy, has identified eight genomic regions in *C. pomonella* that are strongly associated with resistance to CpGV-M. This includes a region on the Z chromosome, as expected based on previous studies (Asser-Kaiser et al. 2011). Furthermore, analysis of the population genetic structure among populations revealed contrasting patterns between autosomes and the Z chromosome. SNPs on the autosomes exhibited a hierarchical population genetic structure, indicating a clear signal of isolation by distance in France and a relative isolation of the two Italian populations. On the other hand, SNPs on the Z chromosome showed a more complex genetic structure, combining geography and phenotype, and suggesting a stronger selective pressure on this chromosome.

Since the sampled populations were geographically distant, it is plausible that the genetic events leading to the resistance phenotypes occurred independently through parallel or convergent evolution. Therefore, despite the limited number of tested populations, the identification of 8 candidate regions supports the existence of an oligogenic architecture of resistance to CpGV-M, which includes autosomal genes.

Although previous knowledge suggested that CpGV-M resistance was primarily attributed to a single gene on the Z chromosome, our study demonstrates the involvement of other genes, some of which exhibit stronger support. However, the approach employed in our study does not allow us to evaluate the relative contribution of different genes to the resistance phenotypes.



**Fig. 2.** Representation of SNPs associated with CpGV-M resistance along the genome of *Cydia pomonella*. **(a)** Manhattan plot showing the 99<sup>th</sup> percentile of the posterior mean of XtX. SNPs being highly associated with the resistance phenotypes (with a Bayes Factor  $\geq 15$  dB) are shown in red. **(b)** Represent the number of SNP (with BF  $\geq 15$  dB) in 50 kb sliding windows along the genome. The 3 candidate genes (*Ptp61F*, *BIRC6* and *Cys*) known to be involved in immunity mechanisms are represented in green. Numbers on the x-axes correspond to the chromosomes. The 8 candidate regions are represented by the vertical gray bars in both graphs.

The functional annotation analysis of the genes identified in the 8 genomic regions strongly associated with resistance to CpGV-M pointed 3 genes involved in insects immune pathways. The first one, *CpPtp61F*, was found on the Z chromosome and is coding for a tyrosine phosphatase (*Ptp61F*). In insects, *Ptp61Fs* are known to be involved in immunity

(Müller et al. 2005). *Ptp61F* is acting in Janus kinase/signal transducer and activator of transcription (JAK/STAT), one of the 6 immunity pathways against viruses identified in *B. mori* and in *Drosophila melanogaster* (Fig. S3, Jiang et al. 2021, Müller et al. 2005). In *Bombyx mori*, an ortholog of this gene, *BmPtp61F*, is involved in viral infection of the

**Table 1.** a) Location in ORFs of the 488 candidate SNPs and the 126 candidate SNPs within genomic regions strongly associated with resistance. b) position, size and number of candidate SNP in the 8 genomic regions strongly associated with resistance to CpGV-M.

a)	Candidate SNPs in all genome		Candidate SNPs in genomic regions strongly associated with resistance		
exonic	16 (3.2%)		11 (8.8%)		
upstream gene variant	33 (6.6%)		10 (7.9%)		
downstream gene variant	30 (6.1%)		14 (11.1%)		
Intron	97 (19.9%)		20 (15.9%)		
intergenic	312 (63.9%)		71 (56.3%)		
total	488		126		

b)	Range		size	candidate SNP	Coding protein	non coding RNA
chr1	53 560 000	53 655 000	95 000	9	3	4
chr15	10 350 000	10 480 000	130 000	25	2	3
chr16	13 255 000	13 340 000	85 000	6	5	2
chr19	8 040 000	8 265 000	225 000	30	16	7
chr2	16 585 000	16 685 000	100 000	10	2	1
chr21	13 050 000	13 290 000	240 000	15	5	4
chr3	16 945 000	17 050 000	105 000	18	5	3
chr4	12 145 000	12 215 000	70 000	6	2	3

baculovirus BmNPV, and its over-expression enables virus replication in host cells (Wang et al. 2016). This is particularly interesting since previous studies showed that CpGV-M replication is blocked in *C. pomonella* resistant individuals (Asser-Kaiser et al. 2011; Berling et al. 2009). We thus can formulate the hypothesis that over-expression of *CpPtp61F* by CpGV-M might not be triggered in resistant *C. pomonella* individuals. Interestingly, our results pointed to a lncRNA (lnc004846) located in close proximity to *CpPtp61F*. Gene cis-activation by lncRNA have been previously described (Gil & Ulitsky 2020), and a recent study has demonstrated the importance of lncRNA in the infection cycle of a baculovirus in *B. mori* (Z. Zhang et al. 2021). New prediction providing links between lncRNAs and their partner genes (expression regulation) have been published on the website <http://v2.insect-genome.com/lncRNA>, based on analysis conducted by the FEELnc software. These data indicate that lnc004846 has only one partner gene, which is *Ptp61F*. Thus, it would be interesting to deeply investigate whether lnc004846 is involved in *CpPtp61F* expression cis-regulation. Other lncRNAs found associated with CpGV-M resistance in our study may also influence the expression of more distantly located genes through TRANS activation.

The two other candidate genes involved in insects' immune pathways, *BIRC6* and *Cys*, were both in chromosome 19 genomic region strongly associated with resistance to CpGV-M. The inhibitor of apoptosis (*BIRC6*) protein has been described as interacting in the insect B immunity

pathway (Imd) (Fig. S2) (Kleino et al. 2005). *BIRC6* (also referred to as *BRUCE* or *Apollon*) is carrying an inhibitor of apoptosis (*iap*) domain known to act on caspase inhibition (O'Riordan et al. 2008). Insect *iaps* act as sentinels of viral infection as they contain degrons that make them unstable and easily degradable when viral infection occurs, causing apoptosis (Vandergaast, et al. 2015). It is worth noting that *BIRC6* was spread over 3 ORF predictions, which illustrate that the *in silico* annotation of the *C. pomonella* genome need to be improved. The second autosomal candidate gene is the *Cys* gene which has been described as involved in the autophagic cell death in *Drosophila melanogaster* (Gorski et al. 2003). It is also known to be involved in immunity deregulation by inhibiting host target proteases in the infection of the particular parasitoid/bracovirus/caterpillar trio (Espagne et al. 2005). In this gene, we detected a non-synonymous mutation replacing an asparagine to aspartic acid replacing a non-polar amino acid with a negatively charged one. To conclude on the implications of the three candidate genes identified in our study, further investigations involving functional validation methods, such as RNA interference (RNAi) or CRISPR-Cas9, should be undertaken in future research, similar to studies conducted by Zhang et al. (2022) in related research domains.

We have highlighted regions of the genome for which we do not yet have information on their role in immunity but which may prove interesting as progress is made in *C. pomonella* genome annotation and in knowledge on insect-



**Table 2.** List of genes carrying candidate SNPs identified as involved in type 1 resistance to CpGV-M and their annotation. Only already known genes (identified in Wan et al. 2019) have been considered (gene number refers to the annotation available on insect-base.com), and only proteins related to immunity or involved in responses to baculovirus infections are indicated. Gene name and description were obtained in QuickGO and UniProt platforms. The full list of genes including those not known to be involved in immune pathways is available in supplementary material S2.

Chromosom	gene number	Gene abbreviation and synonyms	Gene name and description	Annotation
chr 1	cpo122180	Ptp61F	Tyrosine-protein phosphatase non-receptor type 61F: Inhibitor of 3 immunes pathways, tyrosine phosphatase, inhibitor cell proliferation, inhibitor insuline pathway, inhibitor of stress responds, inhibitor apoptosis, actor in cell proliferation	cytoplasm; cytosol; endomembrane system; membrane; mitochondrial envelope; nucleus; protein tyrosine phosphatase activity; axon guidance; cellular response to hypoxia; germ-band extension; mitotic cell cycle; negative regulation of cell population proliferation; negative regulation of epidermal growth factor receptor signaling pathway EGFR, immune pathway F in lepidoptera; negative regulation of ERK1 and ERK2 cascade, immune pathway F in lepidoptera; negative regulation of insulin receptor signaling pathway; negative regulation of JNK cascade (apoptosis mediated stress pathway); negative regulation of phosphatidylinositol 3-kinase signaling (PI3K), immune pathway E in lepidoptera; negative regulation of receptor signaling pathway via JAK-STAT, immune pathway C in lepidoptera, negative regulation of vascular endothelial growth factor receptor signaling pathway; peptidyl-tyrosine dephosphorylation; peptidyl-tyrosine phosphorylation; receptor signaling pathway via JAK-STAT; regulation of protein localization; regulation of protein stability
chr 19	cpo034720 cpo034700 cpo034750	BIRC6, Apollon, BRUCE	Baculoviral IAP repeat-containing protein 6: Anti-apoptotic protein which can regulate cell death by controlling caspases and by acting as an E3 ubiquitin-protein ligase cell division, microtubule organizing, placenta development	endosome; Flemming body; membrane; microtubule organizing center; midbody; nucleus; spindle pole; trans-Golgi network; cysteine-type endopeptidase inhibitor activity (Binds to and stops, prevents or reduces the activity of a cysteine-type endopeptidase); cell cycle; cell division; labyrinthine layer development (placenta development); negative regulation of apoptotic process; negative regulation of extrinsic apoptotic signaling pathway; placenta development; positive regulation of cell population proliferation; protein ubiquitination; regulation of cell population proliferation; regulation of cytokinesis; spongiotrophoblast layer development (placenta); ubiquitin-dependent protein catabolic process
	cpo034660	Cys	Cystatin-like protein: protease inhibitor in the extracellular space	multicellular organism reproduction; cysteine-type endopeptidase inhibitor activity; extracellular space; apoptosis regulator, immune pathway F

virus interactions. For example, the 4 genomic regions strongly associated with resistance on chromosomes 2, 3, 10 and 19 contained non-synonymous mutations in uncharacterized proteins.

This study confirmed that pool-seq, combined with paired population sampling design and methods that consider population structure is an efficient approach for detecting convergent adaptation in wild populations and identify candidate genes (Carrier et al. 2021). We provide genomic resources that would help future studies on insect-virus interactions and, in particular, to understand how *C. pomonella* resists to CpGV-M, which is a unique case of an insect pest adaptation to a microorganism used as a biocontrol agent (Siegwart

et al. 2015). Future works needs to focus on functional validation of the candidate genes found in this study and on *C. pomonella* resistances to other CpGV isolates

**Acknowledgements:** This project was funded and supported by ANSES (grant agreement no. PPV/2020) via the tax on sales of plant protection products. The proceeds of this tax are assigned to ANSES to finance the establishment of the system for monitoring the adverse effects of plant protection products, called ‘phytopharmacovigilance’ (PPV), established by the French Act of the future of agriculture of 13 October 2014. We are grateful to the NPP-UPL, France for providing some wild populations. We are grateful to

the genotoul bioinformatics platform Toulouse Occitanie (Bioinfo Genotoul, <https://doi.org/10.15454/1.5572369328961167E12>) for providing help and/or computing and/or storage resources.

## References

- Abbott, W. S. (1925). A method of computing the effectiveness of an insecticide. *Journal of Economic Entomology*, *18*(2), 265–277. <https://doi.org/10.1093/jee/18.2.265a>
- Anderson, J. P., Gleason, C. A., Foley, R. C., Thrall, P. H., Burdon, J. B., & Singh, K. B. (2010). Plants versus pathogens: An evolutionary arms race. *Functional Plant Biology*, *37*(6), 499–512. <https://doi.org/10.1071/FP09304>
- Asser-Kaiser, S., Radtke, P., El-Salamouny, S., Winstanley, D., & Jehle, J. A. (2011). Baculovirus resistance in codling moth (*Cydia pomonella* L.) caused by early block of virus replication. *Virology*, *410*(2), 360–367. <https://doi.org/10.1016/j.virol.2010.11.021>
- Berling, M., Blachere-Lopez, C., Soubabere, O., Lery, X., Bonhomme, A., Sauphanor, B., & Lopez-Ferber, M. (2009). *Cydia pomonella granulovirus* Genotypes Overcome Virus Resistance in the Codling Moth and Improve Virus Efficiency by Selection against Resistant Hosts. *Applied and Environmental Microbiology*, *75*(4), 925–930. <https://doi.org/10.1128/AEM.01998-08>
- Berling, M., Sauphanor, B., Bonhomme, A., Siegwart, M., & Lopez-Ferber, M. (2013). A single sex-linked dominant gene does not fully explain the codling moth's resistance to granulovirus. *Pest Management Science*, *69*(11), 1261–1266. <https://doi.org/10.1002/ps.3493>
- Carlier, J., Bonnot, F., Roussel, V., Ravel, S., Martinez, R. T., Perez-Vicente, L., ... Wright, S. (2021). Convergent adaptation to quantitative host resistance in a major plant pathogen. *mBio*, *12*(1), e03129–e20. <https://doi.org/10.1128/mBio.03129-20>
- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., ... Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, *6*(2), 80–92. <https://doi.org/10.4161/fly.19695>
- Cory, J. S., & Franklin, M. T. (2012). Evolution and the microbial control of insects. *Evolutionary Applications*, *5*(5), 455–469. <https://doi.org/10.1111/j.1752-4571.2012.00269.x>
- Espagne, E., Douris, V., Lalmanach, G., Provost, B., Cattolico, L., Lesobre, J., ... Huguët, E. (2005). A virus essential for insect host-parasite interactions encodes cystatins. *Journal of Virology*, *79*(15), 9765–9776. <https://doi.org/10.1128/JVI.79.15.9765-9776.2005>
- Gautier, M. (2015). Genome-wide scan for adaptive divergence and association with population-specific covariates. *Genetics*, *201*(4), 1555–1579. <https://doi.org/10.1534/genetics.115.181453>
- Gautier, M., Vitalis, R., Flori, L., & Estoup, A. (2022). f-Statistics estimation and admixture graph construction with Pool-Seq or allele count data using the R package poolfstat. *Molecular Ecology Resources*, *22*(4), 1394–1416. <https://doi.org/10.1111/1755-0998.13557>
- Gil, N., & Ulitsky, I. (2020). Regulation of gene expression by cis-acting long non-coding RNAs. *Nature Reviews. Genetics*, *21*(2), 102–117. <https://doi.org/10.1038/s41576-019-0184-5>
- Gorski, S. M., Chittaranjan, S., Pleasance, E. D., Freeman, J., Anderson, C. L., Varhol, R. J., ... Marra, M. A. (2003). A SAGE approach to discovery of genes involved in autophagic cell death. *Current Biology*, *13*(4), 358–363. [https://doi.org/10.1016/S0960-9822\(03\)00082-4](https://doi.org/10.1016/S0960-9822(03)00082-4)
- Huber, J. & others. (1990). History of the CPGV as a biological control agent-its long way to a commercial viral pesticide. In *Proceedings and abstracts, Vth International Colloquium on Invertebrate Pathology and Microbial Control, Adelaide, Australia*, 20–24 August 1990. (pp. 424–427). Department of Entomology, University of Adelaide.
- Jeffreys, H. (1998). *The theory of probability*. OuP Oxford.
- Kleino, A., Valanne, S., Ulvila, J., Kallio, J., Myllymäki, H., Enwald, H., ... Rämetsä, M. (2005). Inhibitor of apoptosis 2 and TAK1-binding protein are components of the *Drosophila* Imd pathway. *The EMBO Journal*, *24*(19), 3423–3434. <https://doi.org/10.1038/sj.emboj.7600807>
- Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., McLellan, M. D., Lin, L., ... Wilson, R. K. (2012). VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research*, *22*(3), 568–576. <https://doi.org/10.1101/gr.129684.111>
- Krueger, F. (2015). *TrimGalore: A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files*. Babraham Bioinformatics.
- Lacey, L. A., Thomson, D., Vincent, C., & Arthurs, S. P. (2008). Codling moth granulovirus: A comprehensive review. *Bio-control Science and Technology*, *18*(7), 639–663. <https://doi.org/10.1080/09583150802267046>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics (Oxford, England)*, *25*(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics (Oxford, England)*, *25*(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Müller, P., Kutenkeuler, D., Gesellchen, V., Zeidler, M. P., & Boutros, M. (2005). Identification of JAK/STAT signalling components by genome-wide RNA interference. *Nature*, *436*(7052), 871–875. <https://doi.org/10.1038/nature03869>
- O’Riordan, M. X. D., Bauler, L. D., Scott, F. L., & Duckett, C. S. (2008). Inhibitor of Apoptosis Proteins in Eukaryotic Evolution and Development: A Model of Thematic Conservation. *Developmental Cell*, *15*(4), 497–508. <https://doi.org/10.1016/j.devcel.2008.09.012>
- QuickGO (2023). GO version 2023-02-08 Annotation set created on 2023-01-31 09:05. Retrieved from <https://www.ebi.ac.uk/QuickGO/>
- R Core (2013). R: A language and environment for statistical computing.
- Sauphanor, B., Berling, M., Toubon, J. F., Reyes, M., Delnatte, J., & Allemoz, P. (2006). Cases of resistance to granulosis virus in the codling moth. *Phytoma*, *590*, 24–27.
- Siegwart, M., Graillot, B., Blachère-Lopez, C., Besse, S., Bardin, M., Nicot, P., & Lopez-Ferber, M. (2015). Resistance to bio-insecticides or how to enhance their sustainability: A review. *Frontiers in Plant Science*, *6*. <https://doi.org/10.3389/fpls.2015.00381>
- Siegwart, M., Graillot, B., Crabos, L., Besse, S., & Lopez-Ferber, M. (2017). Résistance du carpocapse au virus de la granulo. *Phytoma*, *708*, 4.

- Storz, J. F. (2005). Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology*, *14*(3), 671–688. <https://doi.org/10.1111/j.1365-294X.2005.02437.x>
- Tabashnik, B. E., Van Rensburg, J. B. J., & Carrière, Y. (2009). Field-evolved insect resistance to Bt crops: Definition, theory, and data. *Journal of Economic Entomology*, *102*(6), 2011–2025. <https://doi.org/10.1603/029.102.0601>
- UniProt Consortium. (2023). UniProt: The Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, *51*, D523–D531.
- Vandergaast, R., Mitchell, J. K., Byers, N. M., & Friesen, P. D. (2015). Insect inhibitor-of-apoptosis (IAP) proteins are negatively regulated by signal-induced N-terminal degrons absent within viral IAP proteins. *Journal of Virology*, *89*(8), 4481–4493. <https://doi.org/10.1128/JVI.03659-14>
- Wan, F., Yin, C., Tang, R., Chen, M., Wu, Q., Huang, C., ... Li, F. (2019). A chromosome-level genome assembly of *Cydia pomonella* provides insights into chemical ecology and insecticide resistance. *Nature Communications*, *10*(1), 4237. <https://doi.org/10.1038/s41467-019-12175-9>
- Wang, F., Xue, R., Li, X., Hu, C., & Xia, Q. (2016). Characterization of a protein tyrosine phosphatase as a host factor promoting baculovirus replication in silkworm, *Bombyx mori*. *Developmental and Comparative Immunology*, *57*, 31–37. <https://doi.org/10.1016/j.dci.2015.12.002>
- Zhang, Y., Liu, S., De Meyer, M., Liao, Z., Zhao, Y., Virgilio, M., ... Wee, S. L. (2022). Genomes of the cosmopolitan fruit pest *Bactrocera dorsalis* (Diptera: Tephritidae) reveal its global invasion history and thermal adaptation. *Journal of Advanced Research*. <https://doi.org/10.1016/j.jare.2022.12.012>
- Zhang, Z., Zhao, Z., Lin, S., Wu, W., Tang, W., Dong, Y., ... Guo, X. (2021). Identification of long noncoding RNAs in silkworm larvae infected with *Bombyx mori cypovirus*. *Archives of Insect Biochemistry and Physiology*, *106*(3), 1–12. <https://doi.org/10.1002/arch.21777>
- Zichová, T., Stará, J., Kundu, J. K., Eberle, K. E., & Jehle, J. A. (2013). Resistance to *Cydia pomonella* granulovirus follows a geographically widely distributed inheritance type within Europe. *BioControl*, *58*(4), 525–534. <https://doi.org/10.1007/s10526-013-9507-1>

Manuscript received: 17 March 2023

Revisions requested: 3 June 2023

Modified version received: 21 June 2023

Accepted: 28 June 2022

The pdf version (Adobe JavaScript must be enabled) of this paper includes an electronic supplement: Fig. S1–S3, Table S1–S4