Correlations Between Read Depth, Copy Number Variation, and Bt Resistance in *Helicoverpa zea*

Bhargav Srinivasan, Megan L. Fritz

Department of Entomology, University of Maryland, College Park, MD, U.S.A.

Introduction

The corn earworm, also known as *Helicoverpa zea*, is a major pest of vegetable crops across the Americas (Fig. 1). They recently evolved high levels of resistance to transgenic Bt-expressing crops meant for their control. Resistance resulted, in part, from a copy number variant (CNV; Fig. 2). This class of mutations is often associated with rapid adaptation [1], yet their detection can be challenging in non-model organisms. In this study, we explored three CNV identification tools to determine the geographic extent of this Bt pesticide resistance mutation in *H. zea*.

Figure 1. *Helicoverpa zea*

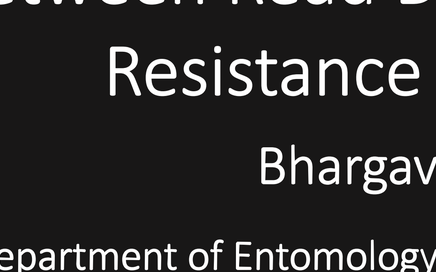
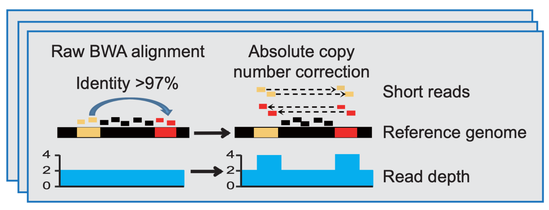


Figure 2. Potential CNV events

Objectives

- Evaluate the accuracy of 3 different CNV identification methods (Figs. 3 & 4) on a whole genome short read dataset from a non-model organism.
- Determine the geographic extent of the Bt resistance CNV across 11 North American populations.
- Detect additional high frequency (>50% per pop) CNVs across the *H. zea* genome.

Read-Depth (RD)



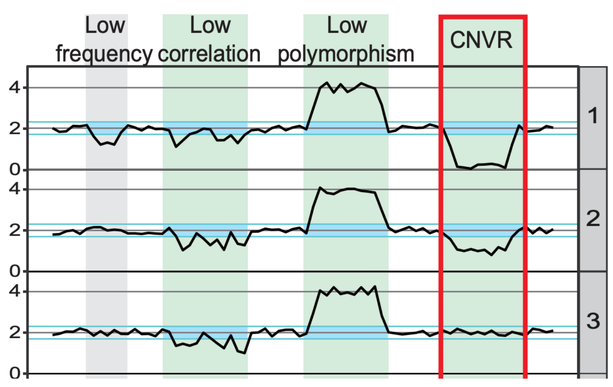


Figure 3. Adapted from [2]. CNVCaller's methods for CNV identification. Sliding windows are normalized for their read depth, then analyzed for their polymorphism within the population.

Integrated Paired-End (PE) & Split-Read (SR)

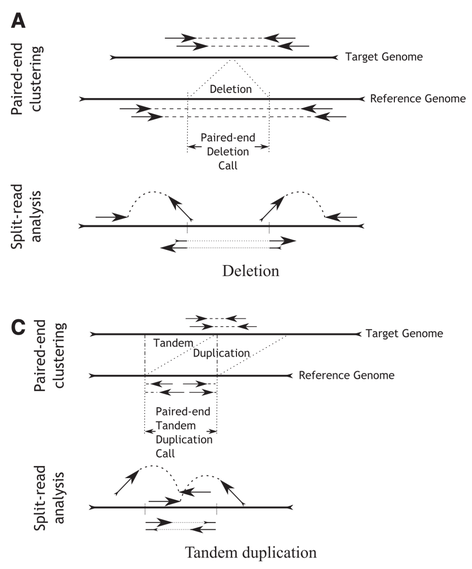


Figure 4. Adapted from [3]. Delly's methods for tandem duplication and deletion identification. Split reads and paired end reads are analyzed for their cutoff points among potential structural variants based on unusual orientation and gaps.

Methods

We used an archived WGS dataset of 237 wild *H. zea* samples collected by North et al. [4] from 10 North American locations. These samples are from populations with varying resistance levels and thus, likely have varying frequencies of the Bt resistance-related CNV and other potentially adaptive CNVs. We also used a positive control dataset of 22 wild *H. zea* collected by Taylor et al. [5], which were previously validated to contain the Bt resistance CNV. We tested both datasets with the following methods:

1) Identify statistically significant adjacent windows of higher than expected read depth (RD) across the population, using custom scripts that we developed.
2) Use CNVCaller [2] to identify population wide CNVs with read depth normalized across 1600 base pair sliding windows across the genome.
3) Use Delly [3] to identify population wide CNVS with paired-end and split-read methods.

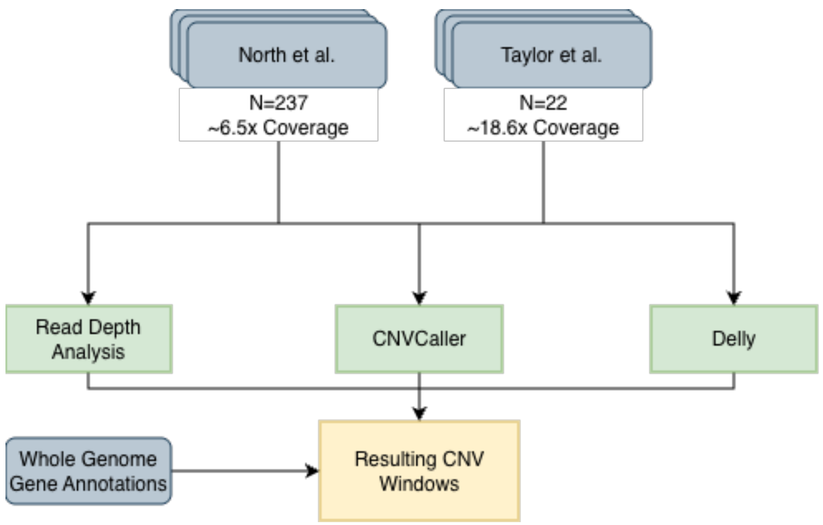


Figure 5. Bioinformatics pipeline to find duplicated genes. North et al. files are pooled by each of 10 locations during analysis. Each method produces windows of duplication that are paired with annotated genes for output.

Results

Our custom RD analysis and CNVCaller identified the Bt resistance CNV in our positive control dataset, while Delly 2 did not. Moreover, Delly 2 did not identify any duplicated genes in the Taylor et al. files (Fig. 6). Of the 3 tools, our RD analysis identified the most unique CNVs, followed by Delly 2 and CNVCaller (Fig. 6).

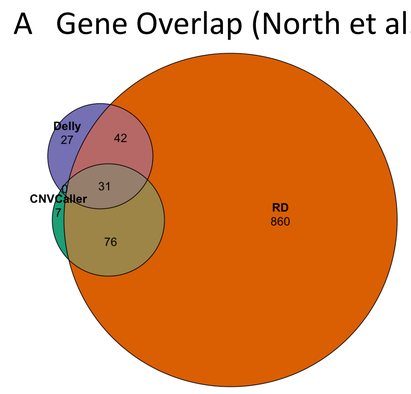


Figure 6. Distribution of duplicated genes among all North et al. populations (A) and Taylor et al. samples (B). Genes analyzed in plot A are taken as the union of output genes between all 10 locations.

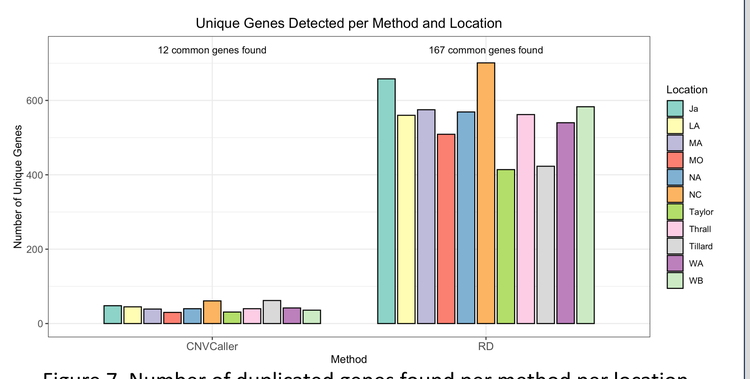


Figure 7. Number of duplicated genes found per method per location. When finding the intersection between all locations per method, we find that 12 common genes were found for CNVCaller and 167 common genes were found for RD analysis. Delly 2 is not shown because it failed to detect any CNVs in our positive control data.

After running each method on our 11 groups (10 North et al. locations + Taylor et al. positive control), we found a list of common (frequency > 50%) duplicated genes for each method in each location (Fig. 7). The Bt resistance-related CNV could be found in all populations, but at varying frequencies. (Fig. 8)

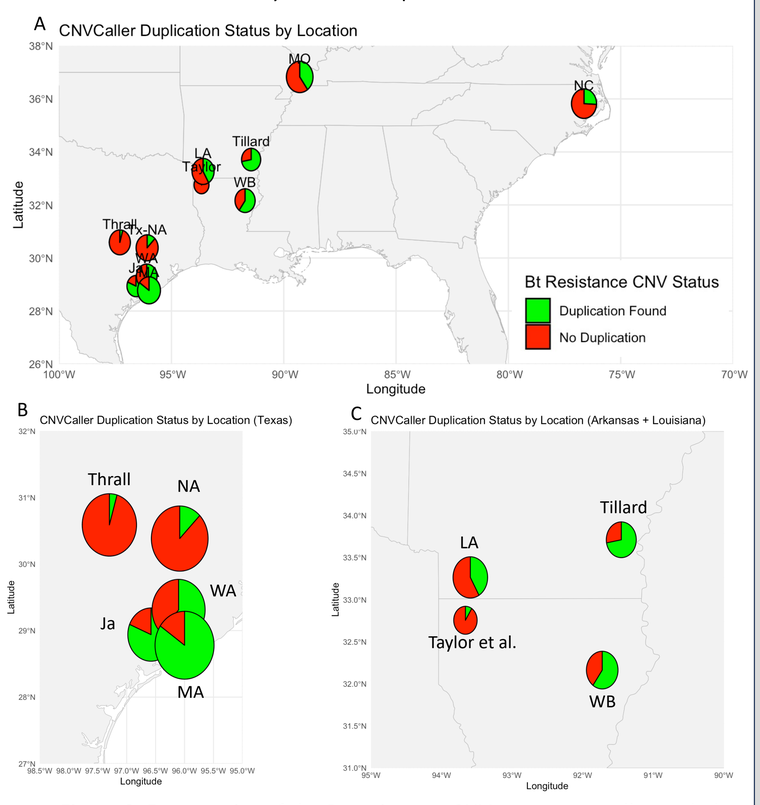


Figure 8. Frequencies of the Bt resistance CNV gene across all sampled populations. Plots B and C show zoomed in replicas of plot A for clarity.

To find potential genes of adaptive importance, we inspected common duplicated genes found amongst many locations (Fig. 9).

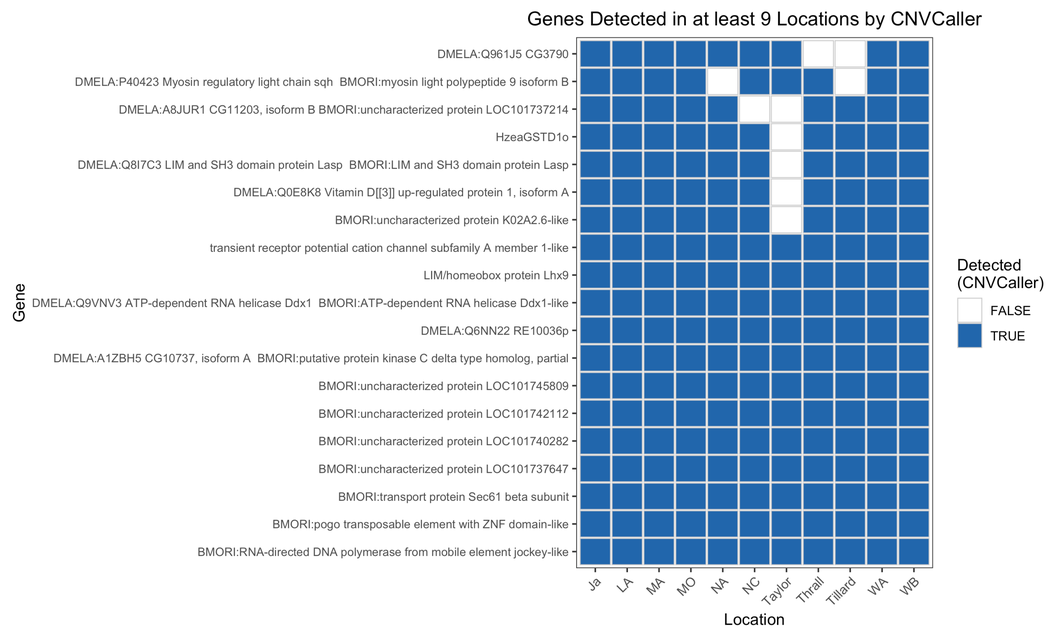


Figure 9. Most common genes found by CNVCaller per location (10 from North et al. and 1 from Taylor et al.).

Conclusions

- One CNV detection tool (Delly2), did not identify the known Bt resistance allele in our positive control data (Taylor et al.). This highlights a challenge for CNV detection in non-model organisms with currently available tools.
- The Bt resistance CNV is found in multiple sites across North America, highlighting its widespread distribution (Fig. 8).
- Future work could examine the adaptive importance of common *H. zea* CNVs (Fig. 9).

References and Acknowledgements

[1] Heckel DG., Pest Manag Sci. 2022 Jan;78(1):12-18
[2] Wang et al., GigaScience, 0, 2017, 1–12
[3] Rausch et al., Vol. 28 ECCB 2012, pages i333–i339
[4] North et al., *Molecular Biology and Evolution*, Volume 41, Issue 7, July 2024
[5] Taylor et al., BMC Genomics, Volume 25, article number 1247, (2024)

This work was supported by USDA NIFA Biotechnology Risk Assessment Grant 2024-33522-43706 to MLF.