

Article

Viral Prevalence and Genomic Xenology in the Coevolution of HzNV-2 (Nudiviridae) with Host *Helicoverpa zea* (Lepidoptera: Noctuidae)

Luke R. Tembrock ^{1,*} , Frida A. Zink ¹  and Todd M. Gilligan ² ¹ Department of Agricultural Biology, Colorado State University, Fort Collins, CO 80523, USA² USDA-APHIS-PPQ-Science & Technology, Identification Technology Program, Fort Collins, CO 80526, USA

* Correspondence: tembrock@colostate.edu

Simple Summary: Most insect species, like most other animal species, host a number of different viruses, and because the number of insect species are so great, the diversity of viruses in insects is expected to be larger than in any other animal lineage. Despite this expectation, few studies are conducted to characterize such viruses and fewer still are carried out to describe the evolutionary dynamics between host and virus. Here we studied a virus that infects an agriculturally important moth species corn earworm (*Helicoverpa zea*) by comparing host and viral genomes. From this comparison it was found that the virus has incorporated several host genes among which the gene cytosolic serine hydroxymethyltransferase is not known from any other virus and thus may provide a unique study system for better understanding the function of this gene. Additionally, undescribed genes were found in both the virus and host that possess sequences originating from bacteria. Lastly the prevalence of the virus was tested across a broad geographic range and found to be present in nearly all host populations tested. Given these findings, more studies should be initiated to characterize insect viruses for novel molecular interactions with possible uses in gene editing and insect biological control.



Citation: Tembrock, L.R.; Zink, F.A.; Gilligan, T.M. Viral Prevalence and Genomic Xenology in the Coevolution of HzNV-2 (Nudiviridae) with Host *Helicoverpa zea* (Lepidoptera: Noctuidae). *Insects* **2023**, *14*, 797. <https://doi.org/10.3390/insects14100797>

Academic Editor: Angela Smilanich

Received: 28 June 2023

Revised: 16 September 2023

Accepted: 24 September 2023

Published: 30 September 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Insect viruses have been described from numerous lineages, yet patterns of genetic exchange and viral prevalence, which are essential to understanding host–virus coevolution, are rarely studied. In *Helicoverpa zea*, the virus HzNV-2 can cause deformity of male and female genitalia, resulting in sterility. Using ddPCR, we found that male *H. zea* with malformed genitalia (agonadal) contained high levels of HzNV-2 DNA, confirming previous work. HzNV-2 was found to be prevalent throughout the United States, at more than twice the rate of the baculovirus HaSNPV, and that it contained several host-acquired DNA sequences. HzNV-2 possesses four recently endogenized lepidopteran genes and several more distantly related genes, including one gene with a bacteria-like sequence found in both host and virus. Among the recently acquired genes is cytosolic serine hydroxymethyltransferase (cSHMT). In nearly all tested *H. zea*, cSHMT contained a 200 bp transposable element (TE) that was not found in cSHMT of the sister species *H. armigera*. No other virus has been found with host cSHMT, and the study of this shared copy, including possible interactions, may yield new insights into the function of this gene with possible applications to insect biological control, and gene editing.

Keywords: evolution of insect genitalia; speciation; *Helicoverpa armigera*; *Helicoverpa zea*; dsDNA viruses; nudivirus; transposable elements; direct repeats

1. Introduction

Viruses are ubiquitous biological entities associated with nearly every form of cellular life. Viral genomes are diverse in structure and can consist of single or double-stranded RNA or single or double-stranded DNA (dsDNA), with sizes varying from 2 Kb to 2 Mb [1,2]. In prokaryotes DNA viruses are the most common, while in eukaryotes RNA

viruses are the most common [3,4]. However, diverse dsDNA viruses are also known from eukaryotes, with arthropods hosting a broad spectrum of these viruses [5]. Long-term persistence and high levels of prevalence have been noted among numerous dsDNA viruses in animals, such as lifelong infections from HSV (herpes simplex virus) and papillomaviruses in mammals [6,7]. The highest degree of persistence among dsDNA viruses is known from braconid parasitoid wasps, where entire viral genomes have been integrated into some wasp genomes and remain functionally active in blocking insect host immune responses [8,9]. Relatives of these fully endogenized viruses (Polydnviridae), the Nudiviridae, have been described from hosts in the arthropod orders Coleoptera, Decapoda, Diptera, Lepidoptera, and Orthoptera [10]. These nudiviruses are similar to dsDNA viruses like HSV regarding persistence, transmission, and prevalence [11]. While studies of Nudiviridae prevalence are few [12], quantification of prevalence is essential to understanding host–pathogen interactions and coevolution therein.

Numerous viruses are known to infect lepidopteran hosts including at least three in the moth *Helicoverpa zea* [13,14]. The “*Helicoverpa zea* Reproductive Virus,” hereafter referred to with the taxonomically correct HzNV-2 (*Helicoverpa zea* Nudivirus strain 2) naming convention (*sensu* [15,16]), is noted as causing gonadal atrophy among some infected individuals. Previous studies [13,17,18] have demonstrated that the deformed genitalia found in sterile forms of *H. zea* are a result of HzNV-2 infection, specifically, females were noted as having a greatly deformed bursa copulatrix [19]. Transmission of HzNV-2 can occur both horizontally, through individuals mating, and vertically, via the female laying infected eggs with some of the offspring being asymptomatic carriers and others symptomatically sterile [18,20]. Lupiani et al. [12] found that, in some populations, a majority of the wild *H. zea* females tested were positive for HzNV-2, but in most cases the infected individuals were fertile, asymptomatic carriers [20]. Furthermore, the occurrence of HzNV-2 varies in different geographic locations of North America, where prevalence ranges from 0 to 69% [12] due to the multiple modes of transmission and inability to cause mortality or immediate sterility among infected individuals [20]. However, the prevalence of HzNV-2 in *H. zea* has not been examined since the work by Lupiani et al., in 1999, and thus it is unknown if factors such as recent introgression with the sister species *H. armigera* in South America [21] and modern farming practices such as widespread use of Bt toxins have altered the prevalence of HzNV-2.

Historical inferences of HzNV-2 prevalence can be drawn from *Heliothis stombleri* Okumura and Bauer, 1969 which was described as a species separate from *Heliothis* (now *Helicoverpa*) *zea* Boddie, 1850. Male *H. stombleri* were distinguished from *H. zea* based on morphological differences in the valvae and absence of cornuti in the vesica of the genitalia. In females, sclerotization of the ostium and significant deformation of the bursa copulatrix and appendix bursae were used to separate the two species. These descriptions of genitalic malformation match those for symptomatic cases of HzNV-2. A year after the description of *H. stombleri*, Hardwick [22] reared a series of *H. zea* from a single female that consisted of a high percentage of sterile moths matching the description of *H. stombleri*. Because both typical and sterile moths were reared from the same female, Hardwick determined that *H. stombleri* was nothing more than an aberrant sterile form of *H. zea*. The two names were formally synonymized by Pogue [23]. Phylogenetic work by Balbi et al. [24] using DNA sequence data from CO1 and EF-1 α further confirmed that *H. stombleri* is synonymous with *H. zea*. To this day, individuals matching the *H. stombleri* description are routinely collected (e.g., [24,25]), but the mechanism by which these sterile forms of *H. zea* persist in the population has not been thoroughly linked to HzNV-2 symptoms in the taxonomic literature.

Symbiotic microbial interactions are known to mediate numerous host functions as well as the process of speciation [26–28]. While viral infections have not historically been viewed as microbial symbioses (e.g., [29]), it is becoming increasingly apparent that viruses play an important, even at times beneficial, role in host–genome evolution through horizontal gene or genome transfer (HGT), acquisition of novel function and/or form [30–33], and

an increase in nucleotide diversity via selection on immune system related genes [34–36]. Such viral alterations to host genomes are also observed to be important drivers in certain modes of speciation (e.g., [28,37,38]). In a similar manner, acquisition of host DNA by viruses is thought to be important in the evolution of viral genomes and often associated with novel adaptations to the host environment [39–41]. Viral mutation rates are higher than eukaryotic mutation rates [42,43], resulting in an increased substitution rate within acquired genes in the viral genome than within those same genes in the host genome. Because of this, determining the timing of gene integration into a viral genome using phylogenetic methods devised for eukaryotes can be difficult. The correct resolution of a rapidly evolving xenologous viral gene with the slowly evolving genes in the host lineage can produce artifactual results caused by long branch attraction [44]. However, recently acquired genes or those undergoing strong selection for a similar function in both host and virus can be resolved in a position close to the host from which they were acquired.

Here we screened individuals of *H. zea* with aberrant genitalia (consistent with *H. stombleri*) using a highly sensitive droplet digital PCR (ddPCR) assay designed to detect and quantify HzNV-2 viral DNA. Additionally, wild caught *H. zea* individuals with normal genitalia were assayed to quantify viral DNA in asymptomatic males from the same collection locations. Lastly, bulk samples from pheromone traps were assayed in order to provide estimates of viral abundance from a large number of individuals throughout the United States. To study the coevolution of host and virus, we identified highly similar DNA sequences shared between them by comparing HzNV-2 to *H. zea* and the closely related taxa *H. armigera* Hübner, 1809 and *Chloridea virescens* Fabricius, 1777, to establish what genes or DNA fragments may have been recently transferred between the host and HzNV-2. We then annotated the relevant sequences to provide insights into the origins and putative functions of these genes in HzNV-2 (e.g., [45,46]). Phylogenetic analyses were employed to infer the timing of gene transfer. Prevalence of HzNV-2 in *H. zea* populations and the rate, type, and timing of DNA transfers can provide insights into the coevolution of nudiviruses and insect hosts.

2. Materials and Methods

2.1. *Helicoverpa* Collection and Identification

Adult male *Helicoverpa* were obtained from pheromone trap surveys for *H. armigera* in Colorado and Puerto Rico, USA conducted from 2014–2015 and used for individual testing of agonadal individuals. The genitalia from these specimens, used for individual assays, were dissected following procedures in Brambila [47] and Pogue [23]. In summary, the abdomen was removed from the specimen, soaked in 10% KOH for approximately 60 min, and descaled using small brushes. The genitalia were removed from the abdomen using forceps and the phallus was separated from the valvae. Cornuti counts were used to verify specimens were *H. zea* (≥ 15 “sets” of cornuti), *H. armigera* (≤ 9 “sets” of cornuti), or sterile (no cornuti). In some cases, the vesica was everted to examine the number of basal lobes (3 in *H. zea*, 2 in *H. armigera*). Individual (dissected) and bulk (undissected) moths from pheromone traps were obtained from Colorado, Florida, Georgia, Illinois, Indiana, Louisiana, Maine, Mississippi, Missouri, New Mexico, Oregon, Tennessee, Texas, and Puerto Rico, and used for ddPCR testing. Hybrids were attained from the U.S. Department of Agriculture, Forest Pest Methods Laboratory in Buzzards Bay, Massachusetts through controlled crosses between male *H. zea* and female *H. armigera*, while *H. armigera* samples were collected by collaborators in Spain, Australia, and South Africa (Supplementary Tables S1 and S2). The species identity of all dissected agonadal individuals and undissected individuals was confirmed using the real-time PCR assay described in Gilligan et al. [48] or ddPCR for all bulk samples as described in Zink et al. [49]. All specimens were preserved in ethanol and stored at $-80\text{ }^{\circ}\text{C}$.

2.2. DNA Extraction

DNA was extracted from male individuals with aberrant or atypical genitalia matching the description for *H. stombleri*. Positively identified specimens of *H. armigera*, *H. zea*, and lab-reared hybrids between these two species were used as controls. Genomic DNA was extracted from the terminal segment of adult abdomens using a Qiagen DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA, USA) following the protocol described in Tembrock et al. [50]. DNA concentration and purity were measured on a NanoDrop 2000 Ver. 1.6 spectrophotometer (Thermo Scientific/NanoDrop, Wilmington, DE, USA). Bulk extractions were conducted following the procedures described in Gloor et al. [51] and modified in Perera et al. [52] and Zink et al. [49,53].

2.3. Droplet Digital PCR for Detection of HzNV-2 DNA

After DNA extraction, genomic DNA was digested using HindIII (New England Biolabs, Ipswich, MA, USA) following manufacturer's instructions, or fragmented using QIAshredder columns (Qiagen) to ensure DNA fragments were sufficiently small for efficient packaging into droplets. The ddPCR assay was then carried out using primers P4-I, P4-II, P13-I, and P13-II (Table 1) that were designed to amplify two separate loci in the HzNV-2 genome [12]. Due to a single nucleotide mismatch in P4-II, all data presented were generated using the P13 primer set.

Table 1. Primers used in this study.

Name	Description	Sequence	Tm (°C)	Source
P4-I	HzNV-2 detection	5'-GCACGATTCGTAATGTTC-3'	51.8 *	[12]
P4-II	HzNV-2 detection	5'-GCACACCTATCAATCACC-3'	52.8 *	[12]
P13-I	HzNV-2 detection	5'-TCGATGCCGTAATACC-3'	49.7 *	[12]
P13-II	HzNV-2 detection	5'-GTCGCTGAATCAAGTCTG-3'	53.2 *	[12]
Hz_SHMT_1738F	cSHMT TE insert intron 7	5'-CCAGCGCCTCTGCAAAGG-3'	61.4	This Study
Hza_SHMT_24R	cSHMT TE insert intron 7	5'-TAAATGTTAAGCTGTTTCCTCT-3'	53.2–55.9	This Study

* Tm values were recalculated for these primers using the same methods as for the PCR primers designed in this study.

The ddPCR analysis was performed on a QX200 Droplet Digital PCR system (Bio-Rad Laboratories Inc., Hercules, CA, USA) using EvaGreen intercalating DNA dye to detect positive droplets following the protocol described in Zink et al. [49]. The assay was optimized for use with the primers from Lupiani et al. [12] regarding primer concentration, annealing temperature, and PCR protocol. For individual specimens, 0.2 ng of total DNA (includes virus and host DNA) was used for each reaction. For bulk samples, DNA was not quantified and 2 µL of extract was used per reaction. The final primer concentration for each reaction was 175 nM for each forward and reverse primer. The PCR program used was as follows: (1) 5 min at 95 °C, (2) 1 min at 95 °C, (3) 1 min at 53.5 °C, (4) 1 min at 72 °C, (5) 34 repetitions of steps 2 through 4, (6) 5 min at 4 °C, (7) 5 min at 95 °C, and (8) an infinite hold at 4 °C. Lid temperature was maintained at 105 °C through all cycles with a ramp rate of 2 °C/s between each step. After reading, droplet data were analyzed using 'definetherain' [54] to establish the threshold cutoff above which droplets are considered positive.

2.4. Genomic Sequence Comparisons

Several search strategies were employed to compare viral and non-viral DNA. First, host and viral genomes were compared using the complete genome of HzNV-2 (GenBank accession JN418988.1) as a query to BLASTn [55] search against whole-genome contigs (employing the whole genome shotgun database for the three complete genome searches) of *H. zea*, *H. armigera*, and *Chloridea virescens* (taxid: 7113, 29058, and 7102, respectively) separately. The following BLASTn settings were employed: program = blastn; wordsize = 11; expect value 0.05; hitlist size = 200; match/mismatch scores = 2, −3; gapcosts = 5, 2; low

complexity filter = yes; filter string = L,m; genetic code = 1. All searches were performed using the NCBI webserver unless otherwise specified. Hits from each lepidopteran species were grouped based on HzNV-2 genomic location to avoid calling hits as unique that corresponded to the same regions of the HzNV-2 genome. Second, each of the best hits (cutoff e-value ≤ 0.01 ; Pertsemlidis and Fondon 2001 [56]) from the three lepidopteran genomes were then used as queries to BLASTn search (using above settings) the entire nt collection to crosscheck that hits of similar function were found across species. Third, complete ORF sequences (using [16] ORF boundaries) or intergenic DNA from HzNV-2 that were shared with at least one of the three lepidopteran species (e-value ≤ 0.01) were annotated by BLASTn searches to the nt database. Additionally, PSI-BLASTp [57] searches (for ORFs only) to the non-redundant protein database were employed with the following settings: program = blastp, word size = 3; expect value = 0.05; hitlist size = 500; gapcosts = 11, 1; matrix = BLOSUM62; filter string = F; genetic code = 1; window size = 40; threshold = 11; composition-based statistics = 2, and to the clustered nr database using default settings. Lastly, BLASTn searches (using the above parameters for nt searches) were conducted using custom genomic databases from HiRise scaffolded genomes of *H. armigera*, *H. zea*, and hybrid *H. zea* x *H. armigera* (described below). All search results and subsequent annotations were checked against annotations from Burand et al. [16].

2.5. Phylogenetic Analyses

Two complementary phylogenetic approaches were utilized, wherein one approach used a consistently taxon-constrained approach and the other employed a similarity-only criteria in making sequence comparisons. In the first approach, the best matches from the BLAST searches were further scrutinized using several phylogenetic methods to infer the origin of sequences in the HzNV-2 genome. Four HzNV-2 ORFs (Hz2V047, Hz2V066, Hz2V023, and Hz2V035) were aligned with genes (cds only, introns omitted) of the same predicted function from the lepidopteran species *Bombyx mandarina*, *Bombyx mori*, *H. armigera*, *Spodoptera litura*, and *Trichoplusia ni*; the dipteran species *Aedes albopictus*, *Anopheles gambiae*, and *Culex quinquefasciatus*; the hymenopteran species *Megachile rotundata*, *Neodiprion lecontei*, and *Orussus abietinus*; and the coleopteran species *Dendroctonus ponderosae* as an outgroup taxon. These species were chosen because they have high quality genomic resources including predicted gene boundaries, functional annotations, and/or represent species from lineages related to *H. zea*. Where needed, follow-up sequencing was conducted to confirm comparisons. Follow-up sequencing included two HiRise assembled draft genomes (*H. armigera*, and an *H. zea* x *H. armigera* hybrid; GenBank accessions to be added after acceptance), an *H. zea* reference genome ([58]; GCA_022581195.1), as well as PCR generated Sanger sequencing described below. From the predicted gene sequences downloaded from GenBank, ORFs were extracted using ORF finder [59] and then aligned using MUSCLE v 3.8.425 [60]. From the alignment, a NJ (neighbor-joining) tree was resolved with Geneious 9.1.8 using a Tamura-Nei genetic distance model, and branch support assessed with 1000 jackknife replicates. The NJ tree was used as a guide to perform AICc and BIC tests using MODELTEST v 3.7 [61] to determine the best substitution model and rate variation for the Bayesian analyses (Supplementary Table S3). A phylogenetic analysis using MrBayes v 3.2.6 [62] was run with the following parameters: substitution model and rate variation = GTR + I + G; gamma categories = 4; heated chains = 4; chain length = 1,100,000; subsampling frequency = 200; burn-in = 100,000; and an unconstrained branch length prior. Where AICc and BIC model test results differed, an additional run using the HKY85 substitution model was completed to assess differences in branch support and/or tree topology when using different substitution models. Lastly, phylogenetic analyses using the parsimony method were conducted in PAUP v 4.0a168 [63] in two separate runs using a heuristic tree search in one and a branch-and-bound tree search in the other. For both runs, 1000 bootstrap pseudo-replicates starting with a random seed were used to assess branch support.

In the second phylogenetic approach not constrained by taxon, each HzNV-2 ORF was searched in BLAST using the above settings and the top 100 sequences were downloaded and aligned with MAFFT v 7.450 [64,65]. The alignment was edited to remove short sequences and non-isoform duplicates and realigned with MAFFT. The alignment was used to generate a splits network in SplitsTree v 6.0 [66] using the default Hamming distance calculation with splits assessed using 1000 bootstrap replicates. The ORF Hz2V091 was also assessed in similar manner as above but because it is thought to have a diverse microbial origin the same taxon constrained approach was not applied. Instead, ORF Hz2V091 was searched against the BLAST nr/nt database, and the top 250 sequences were downloaded and aligned with MAFFT from which an unrooted NJ tree was generated. The most distant sequence on the NJ tree was used as the outgroup in a Bayesian analysis using MrBayes. All settings for these analyses were the same as outlined above.

2.6. PCR Amplification and Sequencing of cSHMT Segments from *H. armigera*, *H. zea*, and Hybrids

Because cSHMT has not been described as a host-acquired gene in other dsDNA viruses, we chose to characterize the host copies in greater detail with follow-up sequencing. Primers were designed using Primer 3 v 2.3.7 [67] with default settings and the SantaLucia [68] method for calculating T_m. Primers were designed in homologous regions of the alignment that would amplify both species and span regions where inserts of interest were inferred. Individuals of *H. armigera*, *H. zea*, and laboratory-reared hybrids between the two species were employed to generate additional sequence data to confirm cSHMT alignments from GenBank-attained sequences. DNA extraction and quantification were carried out as stated above. The primer set eventually employed after testing was Hz_SHMT_1738F and Hza_SHMT_24R (Table 1). PCR reactions were run in 50 µL volumes and consisted of 5 µL 10X Ex Taq buffer (Takara Bio, Shiga, Japan), 4.0 µL dNTPs, 200 nM of each forward and reverse primer, 0.25 µL Taq (TaKaRa Ex Taq HS polymerase), and 36.75 µL water with 1 µL of template DNA. Thermocycling for PCR amplification was: (1) 5 min at 95 °C, (2) 30 s at 95 °C, (3) 1 min at 53 °C, (4) 1 min 30 s at 72 °C, (5) 34 repetitions of steps 2 through 4, (6) 72 °C for 7 min, and (7) 11 °C infinite hold with a lid temperature of 105 °C maintained through all cycles. Amplicons were run on 1% agarose gels to check for multiple bands. In lanes with multiple bands consistent with those expected from the alignments, the bands were excised from the gel using gel cutting tips and purified using the QIAquick Gel Extraction kit (Qiagen, Valencia, CA, USA). Purified PCR amplicons were sent to the University of Chicago Cancer Research Center DNA Sequencing Facility for Sanger sequencing using an Applied Biosystems 3730XL DNA sequencer (Applied Biosystems, Foster City, CA, USA). Sequence electropherograms were edited by removal of terminal low-quality calls, assembled into contigs (forward and reverse) using Geneious, and aligned using MUSCLE for comparison to previous alignments from GenBank sequences. The Sanger sequencing-confirmed insert was BLAST searched against the reference *H. zea* genome (GCF_022581195.2) to find other locations in the genome where similar insertion events have occurred. Draft genomes of *H. armigera* and *H. zea* × *H. armigera* were also employed in the detection of insert activity and similar comparisons (under NCBI BioProject ID PRJNA1020878). These HiRise genomes were generated by Dovetail (Dovetail Genomics, Scotts Valley, CA, USA) from CHiCAGO and HiC proximity ligation libraries and sequenced on an Illumina HiSeq X (Illumina Inc., San Diego, CA, USA), after which the HiRise pipeline was employed for scaffolding [69–71].

3. Results

3.1. Specimen Identification, Confirmation of HzNV-2 in Agonadal Specimens, and Viral Prevalence Determined from Individual and Bulk Samples

In all instances where dissections were completed, the sterile agonadal individuals were visually aberrant, usually missing all cornuti in the vesica, and the vesica deformed or absent. Using the ddPCR assay, we were able to detect the presence of viral DNA in all agonadal samples collected in the USA with the P13 primer set (Table 1), indicating

high levels of viral DNA in the agonadal samples. Some *H. zea* samples with typical genitalia also showed amplification in the form of a small number of positive droplets that were not present in the No Tissue Control (NTC) assays run with water instead of DNA (Figure 1). No HzNV-2 DNA was detected in *H. armigera* or hybrids (Figure 2). A set of 26 undissected individuals from across the USA were tested for HzNV-2 with ddPCR (primer set P13) of which eight (31%) had detectable levels of HzNV-2 DNA (Supplementary Table S1). Similarly, 29 bulk samples containing between three and 200 individuals (1014 individuals in total) from across the USA were processed with ddPCR with 17 (59%) containing detectable amounts HzNV-2 DNA (Supplementary Table S2). Among the bulk samples, those from Indiana were the only set that did not contain detectable levels of HzNV-2 whereas all samples from Colorado, Florida, and Texas contained detectable levels of virus with the remaining sample sets having a mixture of positive and negative samples.

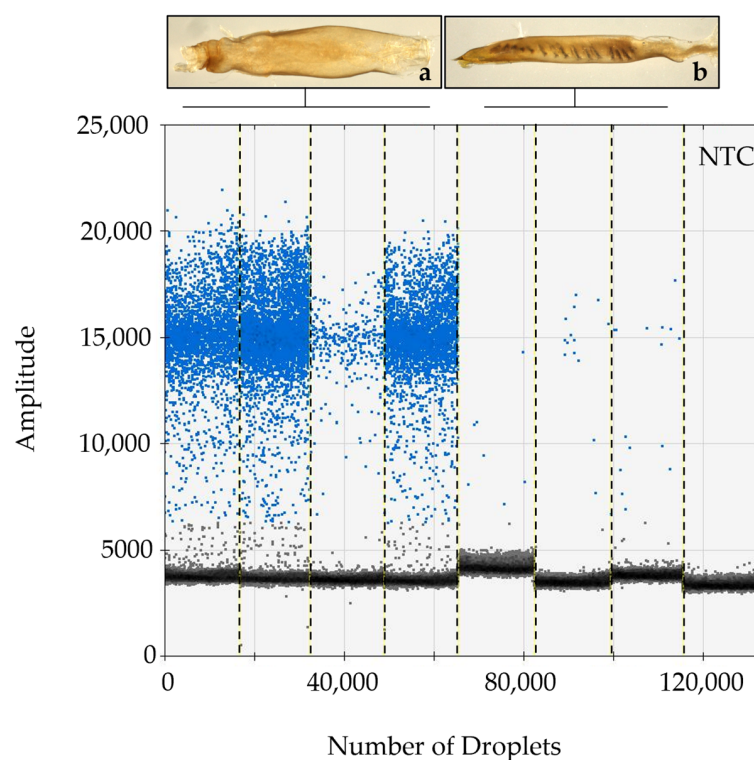


Figure 1. HzNV-2 is detectable at high levels in dissected sterile *H. zea* males compared to morphologically normal dissected *H. zea* males. Results from ddPCR using primer set P13 indicate the presence of HzNV-2 with positive droplets in blue and negative droplets in grey from agonadal *H. zea* males (a), typical *H. zea* males (b), and a no tissue control (NTC). Inset photographs from dissections made for this study illustrate the differences in male genital morphology between agonadal males (a) and typical males (b).

3.2. Genomic Comparisons between HzNV-2, Host Species *H. zea*, and Closely Related Species *H. armigera* and *C. virescens*

The complete HzNV-2 genome was compared to whole genome contigs of *H. zea*, *H. armigera*, and *C. virescens* separately using BLAST searches (Table 2). The searches were performed against each genome separately to optimize results for recently endogenized genome segments between the virus and each lepidopteran species, and to minimize the number of hits to unrelated species. When DNA and protein annotations from viral ORFs agreed, matches were considered high quality representing probable recent endogenizations with functions potentially similar to those in the host genome. These searches revealed four ORFs shared between the HzNV-2 genome and all three individual lepidopteran genomes that could be considered recently acquired based on gene structure and sequence similarity.

These results were found to have the lowest e-values across all searches. The four ORFs were annotated as ribonucleotide reductase (RNR), cytosolic serine hydroxymethyltransferase (cSHMT), proton-coupled folate transporter (PCFT), and thymidylate synthase (TS) based on both DNA and protein searches (Table 2; Supplementary Figure S1). These annotations were similar to those in Burand et al. [16] but additions to the GenBank database since 2012 have improved the specificity of annotations and the e-values of the results. Five other hits were found among the individual searches that had a 0.01 or lower e-value among at least one species. Of the remaining five hits associated with identifiable ORFs, annotations between DNA and protein searches were not as closely linked as those for the top four hits. The exception to this was ORF Hz2V091 (Table 2) that was inferred to be homologous with DNA and protein sequences from several bacterial species. ORF Hz2V091 was not annotated in Burand et al. [16].

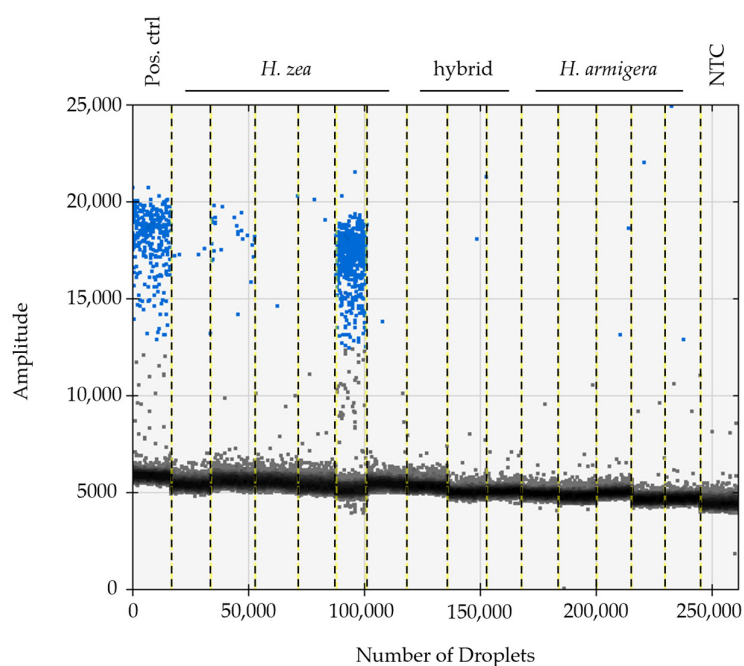


Figure 2. The ddPCR assay using the P13 primer set can detect varying concentrations of viral DNA present in specimens of *H. zea*. DNA from specimens of field-caught *H. zea* and *H. armigera*, and lab-reared hybrids used in the study were assayed using ddPCR to determine the presence or absence of viral DNA. The total DNA (host and virus) concentration is normalized to 0.1 ng/μL. Positive droplets are shown in blue, negative droplets are shown in grey.

3.3. Phylogenetic Analyses of HzNV-2 Genes of Recent Exogenous Origin

Because BLAST hits do not provide lineage-based gene comparisons, we also conducted several phylogenetic analyses using longer complete protein-coding sequences to assess gene origin and timing of acquisition in HzNV-2. When employing a taxon-constrained method, the genes cSHMT, PCFT, and TS all resolved with a lepidopteran clade across all three methods (BI, NJ, and parsimony). The RNR gene only resolved with the lepidopteran clade when using the Bayesian inference (BI) method (Supplementary Figure S2A). Regarding branching order, the ORF sequences from HzNV-2 resolved on an early diverging branch to the lepidopteran clade except for cSHMT which resolved late diverging and sister to *H. armigera* in the BI tree (PP = 0.88) and early diverging in the NJ tree (JK = 100) and parsimony tree (BS = 85; Supplementary Figure S2B). Resolution of the other insect lineages used here generally followed previous analyses (<http://tolweb.org/Endopterygota/8243> accessed on 24 February 2021; Supplementary Figure S2C,D) for insects except for the placement of the lepidopteran clade in the PCFT tree where dipteran and hymenopteran clades are sister in the BI (PP = 0.87) and NJ methods (JK = 98) but not the parsimony method.

Table 2. Annotations and e-values for genomic comparisons and database searches. Values in parenthesis after the best DNA hits are e-value/% query coverage/% identity from the best hit.

Location in HzNV2 Genome	Annotation from Similar Sequences	<i>H. zea</i>	<i>H. armigera</i>	<i>C. virescens</i>	Best DNA Hit	Best Protein Hit	Putative Viral Function	Burand et al. Annotation
ORF Hz2V047 (96009-98822)	Ribonucleotide reductase	1×10^{-30}	6×10^{-34}	8×10^{-21}	<i>Spodoptera littoralis</i> nucleopolyhedrovirus isolate SpliNPV-Tun2 (2×10^{-130} /66/67.1)	Ribonucleoside-diphosphate reductase large subunit, <i>Hypomocoma kahamanoa</i> (0.0)	Nucleotide anabolism, inhibit host cell signalling	Ribonuclease reductase
ORF Hz2V066 (127832-126510)	Serine hydroxymethyl-transferase	7×10^{-27}	5×10^{-22}	6×10^{-23}	Serine hydroxymethyltransferase, cytosolic transcript variant $\times 2$, <i>Helicoverpa armigera</i> (5×10^{-113} /88/67.2)	Serine hydroxymethyltransferase, cytosolic isoform $\times 1$, <i>Helicoverpa armigera</i> (0.0)	Nucleotide anabolism, one-carbon metabolism	Serine hydroxymethyl-transferase
ORF Hz2V023 (49772-51307)	Proton-coupled folate transporter	2×10^{-21}	2×10^{-20}	3×10^{-20}	Proton-coupled folate transporter, <i>Helicoverpa armigera</i> (3.0E-166/81/70.8)	Proton-coupled folate transporter, <i>Helicoverpa armigera</i> (0.0)	Folate transport	Membrane transporter
ORF Hz2V035 (69749-70621)	Thymidylate synthase	1×10^{-18}	3×10^{-18}	5×10^{-17}	Thymidylate synthase, <i>Pectinophora gossypiella</i> (3×10^{-90} /90/70.3)	Thymidylate synthase, <i>Manduca sexta</i> (7×10^{-170})	Thymidine biosynthesis	Thymidylate synthesis
ORF Hz2V002 (7183-13047) hit associated with nested ORF 12222-11431	unknown	2×10^{-14}	1×10^{-14}	no match	Chromosome 19, <i>Perizoma flavofasciatum</i> (2×10^{-34} /60/74.6)	transcriptional regulatory protein AlgP, partial <i>Biomphalaria glabrata</i> (0.036)	unknown	Unidentified
ORF Hz2V098 (195699-199133)	unknown	7×10^{-8}	1×10^{-9}	no match	Chromosome7, <i>Amphipyra tragopoginis</i> (5×10^{-81} /21/70)	Hypothetical protein Cantr_06524, <i>Candida viswanathii</i> (3×10^{-4})	unknown	Unidentified
ORF Hz2V091 (177136-178524)	Cell surface protein (bacterial origin)	1×10^{-5}	3×10^{-12}	no match	HD-771 plasmid p07, <i>Bacillus thuringiensis</i> (4×10^{-34} /21/75.9)	Hypothetical protein BTXL6_27630, <i>Bacillus thuringiensis</i> (5×10^{-23})	Cell surface interaction	Unidentified
Between ORF Hz2V083 (158080-158743) and ORF Hz2V084 (163557-159694)	unknown	1×10^{-6}	6×10^{-9}	no match	Chromosome 6, <i>Harpalus rufipes</i> (1×10^{-21} /12/68)	no match	unknown	Unidentified
ORF Hz2V062 (120105-118786)	Occlusion-derived virus envelop protein e56	6×10^{-4}	1×10^{-5}	no match	Chromosome 14, <i>Lumbricus terrestris</i> (1×10^{-7} /9/84.6)	PIF-5a, <i>Tipula oleracea</i> nudivirus (9×10^{-38})	Viral envelope synthesis	Unidentified

When using a sequence similarity-only approach, splits networks for PCFT and cSHMT (Figure 3A,C) resolved viral ORFs with other lepidopteran genes with the same annotations. The cSHMT viral ORFs resolved on a branch similar in length to branches separating lepidopteran genes, while in the PCFT network the viral ORFs resolved on a long branch relative to the difference between lepidopteran genes. In the RNR splits network (Figure 3B) the viral ORFs resolved between clades containing lepidopteran genes and genes from *Spodoptera litura* Nucleopolyhedrovirus (SpliNPV) with branch lengths similar across the network. In the TS splits network the viral ORFs resolved on a long branch with other arthropod TS genes with splits separating this clade from a vertebrate TS clade (Figure 3B).

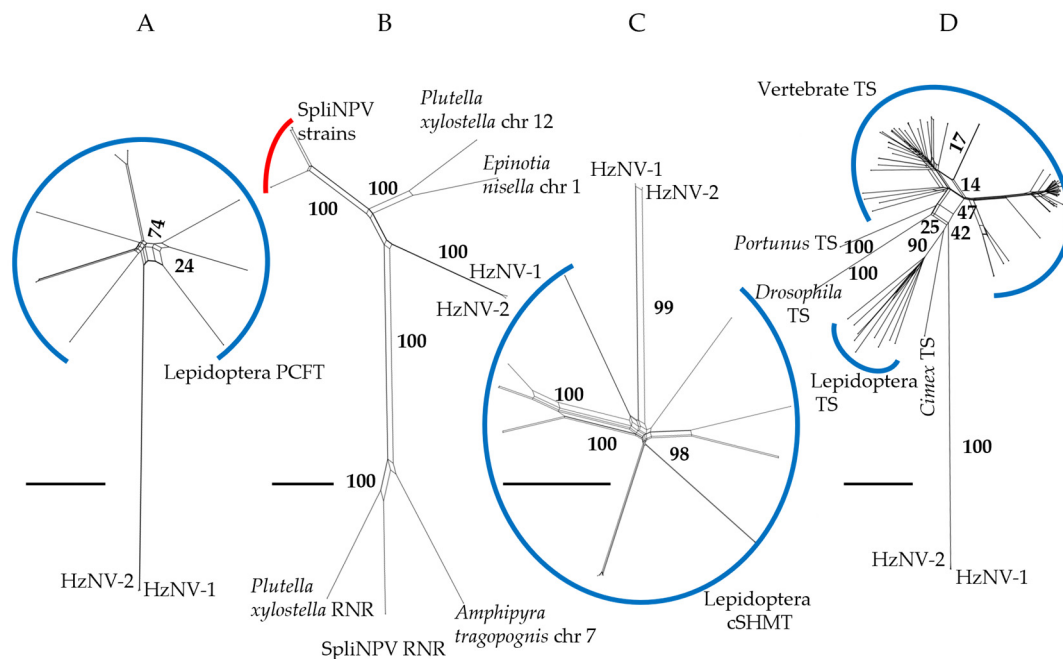


Figure 3. Splits networks from alignments of HzNV-2 ORFs with GenBank best hits using HzNV-2 ORFs (A) PCFT; (B) RNR; (C) cSHMT; and (D) TS. Numbers in networks are bootstrap values and a 0.1 scale bar is provided for each network for comparison at lower left. Blue arcs represent genes from a given metazoan lineage and the red arc indicates genes from a viral lineage. Star or branching patterns in the networks reflect a more even distribution of genetic distances between sequences whereas splits reflect clustering between groupings of sequences. See Figure S2 for the taxonomy constrained phylogenetic analyses.

The phylogenetic analyses of ORF Hz2V091 (Figure 4) resolved this sequence in a well-supported clade (PP = 1) with sequences from plasmids of *Bacillus thuringiensis*, *B. cereus*, and *Brevibacillus laterosporus* as well as the noctuid species *Amphipyra berbera*. This clade was nested within a well-supported clade (PP = 1) of similar sequences from *Streptococcus agalactiae* with functional annotations for cell surface proteins such as *adhesin* and *BibA*. The alignment of ORF Hz2V091 to the best BLAST hits (Supplementary Figure S3A) indicates that the region of similarity between the *A. berbera* and many bacterial matches is centered on direct repeat 4 (dr4) near the 3' end of the ORF. In addition, a region of similarity between a *Plasmodium* repeat motif and the 5' end of ORF Hz2V091 was also matched. An uncharacterized gene in *H. zea* and *H. armigera* (mRNA XM_049842877) with a region similar to ORF Hz2V091 dr4 and bacterial sequences was also found in several recently produced *Helicoverpa* genomes (Supplementary Figure S3B).

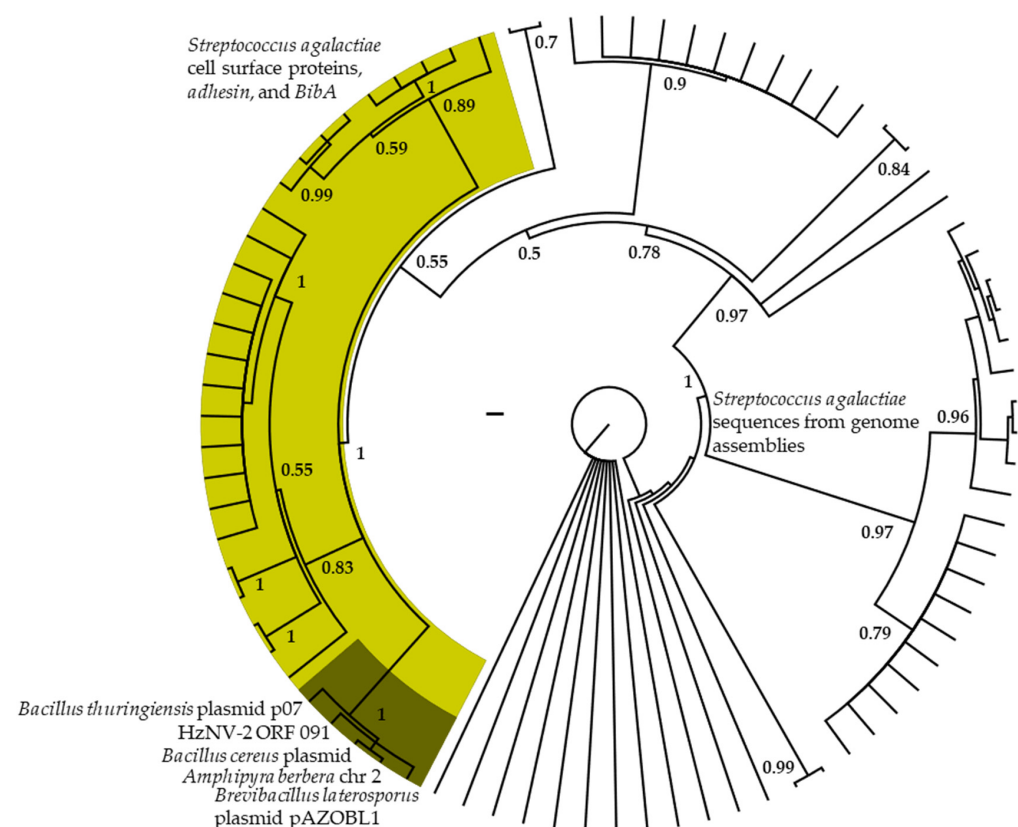


Figure 4. Phylogenetic analyses of the dr4 containing region of the HzNV-2 ORF Hz2V091 in an alignment with the best DNA matches from GenBank. The light-colored clade contains sequences annotated as cell surface proteins including *Adhesin* and *BibA*. The dark colored clade includes HzNV-2 ORF Hz2V091 and plasmid sequences from *Bacillus*, *Brevibacillus*, and the lepidoptera species *Amphipyra berbera*. All other terminals in the tree are from *Streptococcus agalactiae* genome assemblies. Posterior probability is shown for clade support and a 0.2 scale bar is shown near the center of the tree.

3.4. Differences in Length and Nucleotide Content of cSHMT in HzNV-2, *H. armigera*, *H. zea*, and Hybrids

PCR amplifications were conducted across multiple samples to confirm the presence or absence of a 200 bp insert inferred in intron 7 of the cSHMT gene in *H. zea* (Supplementary Figure S4). Both agarose gels and aligned Sanger sequence (GenBank accessions for sequences generated in this study OR609382–OR609386) results confirmed the presence of a 200 bp insert in intron 7 (Figures 5 and 6). All lab-reared hybrids and several field-collected *H. zea* were found to have two prominent bands separated by about 200 bp when the PCR products were resolved on agarose gel. Among the hybrid samples, a presumably *H. armigera*-specific band shorter than the *H. zea* short band was evident especially in samples a/z 3 and a/z 4 which contain the shorter bands from both species. A faint intermediate band was present in some double-banded samples but was not assessed further at this time. Because some field-collected *H. zea* samples had similar banding patterns to known lab-reared hybrids, a real-time PCR assay was run to confirm whether these field-collected individuals were hybrids. None of the double banded *H. zea* were found to be hybrids when assessed with real-time PCR (Supplementary Figure S5). As such, the double-banded *H. zea* were considered heterozygous for the insert. This was further confirmed with HiRise assembled genome data from a hybrid that did not contain the insert (Figure 5). When the insert sequence was BLAST-searched against the reference *H. zea* genome, homologous sequences were found in 10,963 separate locations in the genome although many were shorter than the query sequence (Supplementary Figure S6). The

contribution of this insert to the total length of each chromosome was proportionally higher among the shorter chromosomes (Supplementary Figure S7). Because cytochrome P450 genes are known to be central to mediating xenobiotic responses including viral infection in insects and have similar inserts as those found in cSHMT, we specifically examined these genes for such inserts. We found 13 cytochrome P450 genes with similar inserts from seven chromosomes mainly in the introns but with several spanning intron exon boundaries as well as inserts closely adjacent (less than 200 bp) to a gene (Table 3). Sequence similarity between the insert found in *H. zea* cSHMT and those described from *H. armigera* [72,73] suggests that this insert is a type of transposable element (TE) known as a short interspersed nuclear element (SINE). Specifically, the insert contains the imperfect direct repeats 5'-GGTAATGA-3' at the 5' and 5'-GGTAATGG-3' at the 3' ends of the insert and within each of the direct repeats are inverted repeats 5'-AATGAC-3' and 5'-GTCGTT-3' (Figure 6). Such a pattern of nucleotide repeat flanking is consistent with TE insertion [74,75].

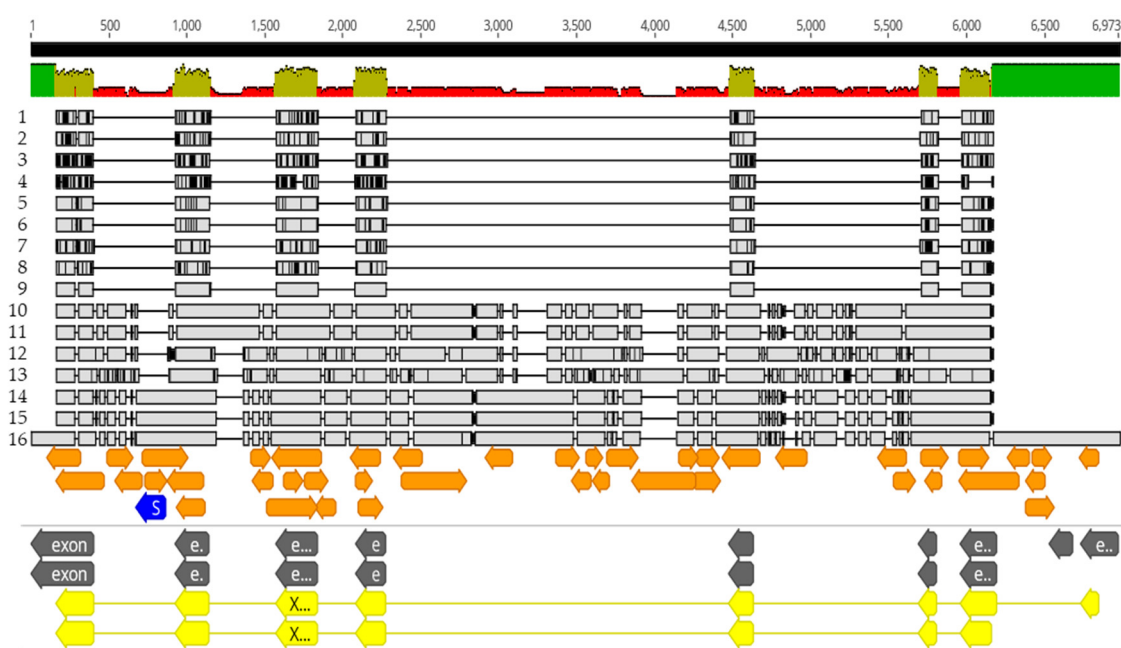


Figure 5. Alignment of cSHMT sequences. Below the alignment: orange tracks indicate alternative ORFs, gray tracks exons, yellow cds, and blue the SINE insert. The top gray and yellow tracks indicate isoform X2 and the bottom gray and yellow tracks indicate isoform X1. Sequences 1–9 are cds alignments from (1) *Culex quinquefasciatus*, (2) *Aedes albopictus*, (3) *Anopheles gambiae*, (4) HzNV-2, (5) *Bombyx mandarina*, (6) *B. mori*, (7) *Trichoplusia ni*, (8) *Spodoptera litura*, and (9) *Helicoverpa armigera*. Sequences 10–16 are complete scaffold sequences from (10) *H. armigera* NW_018395393.1, (11) a HiRise assembly for *H. armigera*, (12) a HiRise assembly for *H. zea* × *H. armigera*, (13) a HiRise assembly for *H. zea* × *H. armigera* (14) a HiRise assembly for *H. zea*, (15) *H. zea* NFMG01027888.1, and (16) *H. zea* reference genome NC_061460.1.

Table 3. All cytochrome p450 genes in the *H. zea* reference genome with detectable TE sequences similar to those found in intron 7 of cSHMT.

Chr	Length (bp)	% Identity	e-Value	5' Position	Annotation from Reference Genome	Genic Location	Gene #
1	63	89.71	2×10^{-17}	10,131,593	cytochrome p450 4C1-like	intron 7 of 7	1
5	183	73.30	4×10^{-32}	13,156,359	138 bp from cytochrome p450 4C1-like	UTR	2
5	62	91.05	5×10^{-18}	13,139,843	cytochrome p450 4C1-like isoform x1	intron 3 of 9	2
5	58	84.75	9×10^{-9}	13,139,943	cytochrome p450 4C1-like isoform x1	intron 3 of 9 into exon 4 of 10	2

Table 3. Cont.

Chr	Length (bp)	% Identity	e-Value	5' Position	Annotation from Reference Genome	Genic Location	Gene #
9	33	84.21	0.029	1,586,828	cytochrome p450 9E2-like	intron 2 of 9	3
9	83	94.05	7×10^{-29}	1,590,440	cytochrome p450 9E2-like	intron 6 of 9	3
9	31	87.50	0.029	1,590,561	cytochrome p450 9E2-like	intron 6 of 9	3
15	170	73.14	2×10^{-29}	11,280,289	cytochrome p450 6K1-like	intron 7 of 8	4
19	73	83.33	5×10^{-18}	4,236,290	cytochrome p450 4G15-like	intron 5 of 10	5
19	168	75.15	9×10^{-34}	4,240,146	119 bp from cytochrome p450 4G15-like	UTR	5
19	50	86.54	1×10^{-6}	4,250,534	cytochrome p450 4C1-like	intron 4 of 10	6
23	44	85.71	4×10^{-7}	7,407,358	cytochrome p450 4V2-like	intron 4 of 9	7
23	171	71.75	6×10^{-24}	7,426,917	cytochrome p450 4C1-like	intron 3 of 9 into exon 4 of 10	8
23	41	85.71	7×10^{-4}	7,447,063	cytochrome p450 4C3-like isoform x1	intron 2 of 10	9
23	63	86.77	9×10^{-15}	7,447,155	cytochrome p450 4C3-like isoform x1	intron 2 of 10	9
23	45	82.00	7×10^{-4}	9,722,754	cytochrome p450 4C1-like	intron 7 of 9	10
23	44	85.71	4×10^{-7}	9,722,850	cytochrome p450 4C1-like	intron 7 of 9	10
23	82	87.06	1×10^{-19}	9,783,243	cytochrome p450 4V2-like	intron 7 of 10	11
23	78	87.34	2×10^{-18}	9,820,391	cytochrome p450 4C1-like	intron 2 of 9	12
30	70	97.00	5×10^{-18}	2,071,964	cytochrome p450 307A1	intron 1 of 1	13

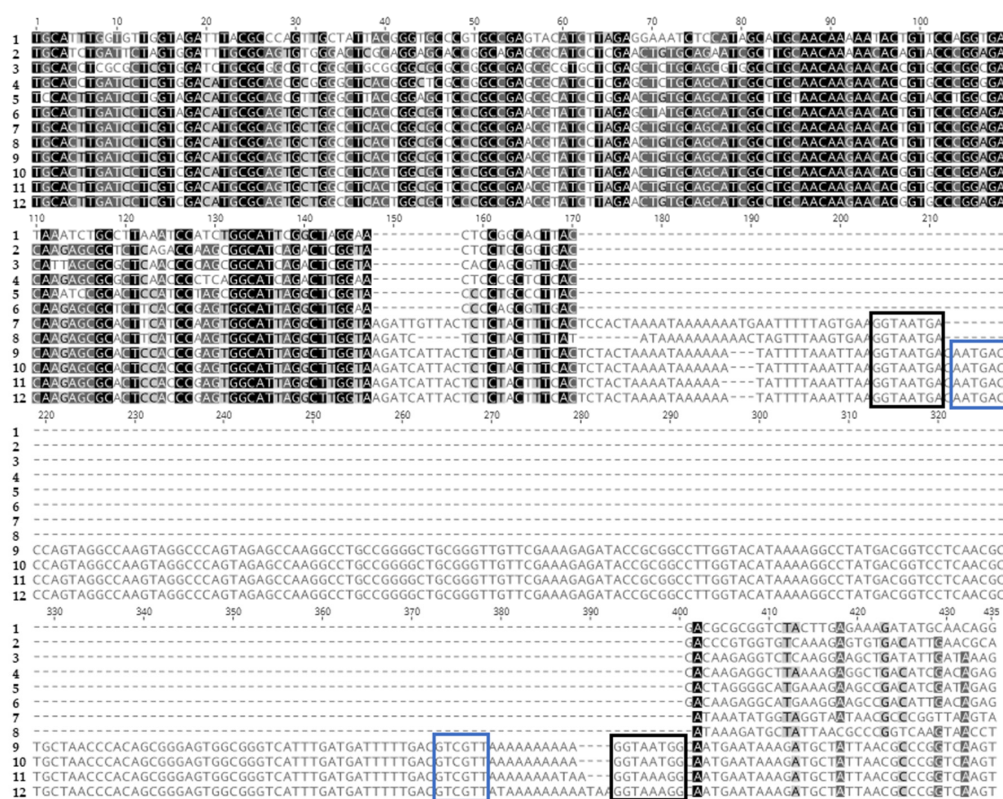


Figure 6. PCR-generated sequences from intron 7 of cSHMT aligned with sequences from GenBank. The sequences 7–12 were generated as part of this study except 11 from *Helicoverpa zea* NFMG01027888.1. The sequences are as follows: (1) *Anopheles gambiae*; (2) ORF Hz2V066; (3) *Bombyx mori*; (4) *Trichoplusia ni*; (5) *Spodoptera litura*; (6) *H. armigera*; (7) *H. zea* heterozygote light band; (8) *H. zea* × *H. armigera* lab reared light band; (9) *H. zea* heterozygote heavy band; (10) *H. zea* × *H. armigera* lab reared heavy band; (11) *H. zea*; and (12) *H. zea* field collected. Sequences from hybrids and heterozygotes were separated by gel electrophoresis, excised, and purified. Black boxes indicate TE flanking direct repeats and blue boxes indicate inverted repeats. *H. zea* sequence 11 was manually adjusted to align flanking direct repeats, the 5' and 3' ends of the alignment were trimmed to fit the figure.

4. Discussion

The dissected agonadal specimens examined in this study (consistent with the description of *H. stombleri*) produced many positive droplets from HzNV-2 DNA when amplified using ddPCR (Figure 1). The agonadal samples with a lower number of droplets still constitute a much higher number of positive droplets than the asymptomatic individuals. The lower number of droplets in this sample might be the result of several factors including differences in tissue collection, DNA extraction efficiency, sample preservation, and viral load. The three dissected individuals consistent with typical *H. zea*, which did not have malformed agonadal genitalia, had very few to no positive droplets. The presence of a small number of droplets sharing a consistent amplitude (RFU) with a positive result from asymptomatic samples has several possible explanations. Since this assay is not intended for diagnosis of the condition, we did not determine a false positive rate. A small number of false positives is considered standard in most ddPCR assays (e.g., [76,77]). Another explanation, which is likely due to the consistency of results for individuals between runs and the lack of false positives in NTC reactions, is that these are asymptomatic (or possibly latent) carriers of HzNV-2 with lower virus titers [20,78]. In any case, a clear difference in the number of positive droplets can be seen between the *H. stombleri*/agonadal forms (216–4580 copies of target per μL total DNA) and the typical *H. zea* samples (2.1–8.3 copies of target per μL total DNA; Figure 1). The lack of positive ddPCR results for *H. armigera* and hybrids indicates that HzNV-2 was not present in these specimens (Figure 2).

A set of undissected individuals and bulk samples were run using the ddPCR assay to estimate the prevalence of HzNV-2 in *H. zea* (Figure 2; Supplementary Tables S1 and S2). Among the individual samples no obvious geographic pattern of prevalence was observed; however, two samples (sample numbers 2330 and 3519) from Puerto Rico collected in different years were found to have very high copy numbers of HzNV-2 present (5520 and 2821 positive droplets). Like the individual sampling, our bulk sampling found the presence of HzNV-2 in nearly all locations except for Indiana which had small bulk sample sizes possibly explaining why no virus was detected there. Our sampling results included a broader geographic area and larger number of samples than Lupiani et al. [12], yet produced similar results, further supporting their assertion that HzNV-2 is an endemic virus to *H. zea* and not recently released from a small area of endemism in Mississippi as had been previously thought. The endemism of HzNV-2 is further confirmed when considering the frequent finds of sterile *H. zea* in South America and other parts the species range [24] are most likely the result of infection. Our sampling, together with past observations, make the case for the need to conduct intensive screening studies for HzNV-2 such that fine scale patterns associated with geographic location, ecological setting, and phenology can be related to prevalence and viral load in *H. zea* populations. Such studies could be applied to understanding how viral infection affects *H. zea* population dynamics and, in turn, crop health. Our results show that ddPCR is an excellent tool for conducting such studies. Pairing epidemiological ddPCR studies with high-throughput sequencing of viral genomes could allow for real-time monitoring of gene introductions from host and other sources, including transient TEs and short non-coding fragments.

Because of the extensive changes to genitalic morphology caused by HzNV-2 infection [17,19] and the importance of genitalia for species identification [79,80], it is not surprising that a species-level taxon designation (i.e., *H. stombleri*) was proposed to account for such observed changes. Given the fact that numerous other sexually transmitted viruses, and viruses in general, can alter the morphology of insect genitalia [14] and other morphological characters [81], it is possible that other insect taxonomic designations have been based on traits altered due to viral infection. This might be especially problematic in cases where the resulting morphological change is not as significant as that caused by HzNV-2 or, in the extreme case, has a selective advantage, as in the domesticated viruses in some ichneumonid wasp lineages [9,82]. Insect species descriptions based on genitalia characters that disagree with other data (e.g., gene trees) may warrant further review considering viral infection as a source of morphological change. Such studies might help to clarify taxonomic

inconsistencies among hosts as well as be valuable in discovering new species of viruses and novel host–pathogen interactions.

In our dataset, four genes were found to be very similar in genomic comparisons of HzNV-2 with *H. zea*, *H. armigera*, and another common heliothine pest *C. virescens* (Table 2). In these instances, timing of gene acquisition is reflected in the BLAST results and phylogenetic analyses of the complete ORF sequences. For the RNR BLASTn search, the best hit was to *Spodoptera littoralis* nucleopolyhedrovirus (SpliNPV) and the best PSI-BLASTp was to an RNR from *Hypomocoma kahamanoa* (Lepidoptera: Cosmopterigidae), while the taxonomically constrained phylogenetic analyses of the complete ORF sequences had HzNV-2 RNR resolving with the other lepidopteran species only in the Bayesian method and branching in a basal position from the outgroup in the other two methods (Supplementary Figure S2). The sequence similarity approach resolved the HzNV-2 RNR ORF with other viral and lepidopteran RNR genes (Figure 3B). This pattern of matches to both lepidopteran and viral sequences may suggest a more ancient acquisition of this gene by an ancestral nudivirus/baculovirus with subsequent sorting to extant viral lineages such as SpliNPV and HzNV-2. Burand et al. [16] identified 20 other ORFs (including a second copy of RNR) from HzNV-2 that were similar to ORFs from other nudiviruses and baculoviruses which may reflect a shared history of a single acquisition event for each ORF among related Nudiviridae species. Alternatively, the shared presence of RNR in separate viral lineages may be the result gene exchange between virus species during a simultaneous infection event which has been noted previously in other viruses [83,84]. The acquisition of RNR by other dsDNA viruses appears to have occurred multiple times as is evident from the presence of different RNR classes in distinct viral lineages similar to non-viral RNRs [85]. Given this frequent uptake of RNR by dsDNA viruses the shared copies in SpliNPV and HzNV-2 may also be explained by separate convergent acquisitions. The acquisition of RNR by viruses is often associated with increasing viral autonomy through the synthesis of deoxyribonucleotides needed for DNA replication of the viral genome independent of host cellular machinery [86,87]. Alternative functions for RNR subunits such as blocking the host inflammatory response and innate immune signaling have been described for viral RNRs, which corresponds with the loss of original enzymatic function of the acquired gene [88].

The PCFT and TS genes resolved in the same position across all three taxon-constrained phylogenetic methods with relatively high support values in all instances. The acquisition of TS in viral genomes is often associated with RNR as it produces a necessary enzyme for the synthesis of the DNA precursor 2'-deoxythymidine-5'-monophosphate [89,90]. The RNR and TS ORFs in HzNV-2, as well as others such as ORF Hz2V067 (Deoxynucleotide kinase) and ORF Hz2V069 (dUTPase), are likely involved with viral DNA replication but alternative functions could have evolved for these or similar genes (such as a second copy of RNR, ORF Hz2V065) as has occurred in herpes simplex virus [91]. The presence of PCFT in the HzNV-2 genome could be functionally associated with cSHMT (potentially both host and viral forms) as they both function in eukaryotic folate metabolism [92,93] and might similarly shuttle/regulate such molecules between virus and host in this instance as well.

The cSHMT ORF from HzNV-2 matches with very high confidence to *H. armigera* cSHMT sequences in both the BLASTn and PSI-BLASTp searches (Table 1). Similarly, the complete ORF phylogenetic analyses using a taxonomy-constrained alignment placed the HzNV-2 sequence with the lepidopteran species in an early diverging position by NJ and parsimony, or sister to *H. armigera* in the Bayesian method (Supplementary Figure S2). For both RNR and cSHMT the similarity-based methods placed the HzNV-2 sequence in a more distant relationship than the model-based Bayesian method, but the placements were consistent in the degree to which the HzNV-2 sequences were similar to other lepidopteran sequences. In the splits network, the HzNV-2 ORF resolved with other lepidopteran cSHMT genes on a relatively short branch with few internal splits suggesting minimal evolution between host and viral genes perhaps as a result of recent acquisition (Figure 3C). High levels of sequence similarity are also reflected in the BLAST hits to the nucleotide and

protein databases. From these results it is reasonable to infer that cSHMT may be the most recently acquired host gene. The recent acquisition of cSHMT from a lepidopteran host species, and the limited host specificity of HzNV-2, is consistent with gene acquisition as a speciation/adaptive event, as seen in similar viruses [40].

One-carbon units such as those produced by cSHMT are essential to many cellular pathways. For instance, cSHMT plays an important role in processes such as neurotransmitter synthesis, lipid and protein production, maintenance of cellular redox status, folate metabolism, and methylation [94,95]. In insects, cSHMT most likely functions similarly to that of other eukaryotes [96] but may also play a role in virus inhibition as has been shown for cytosine methyltransferases in *Aedes aegypti* controlled via *Wolbachia* [97]. Similarly, the cSHMT copy in the HzNV-2 genome (which is unique among xenologous viral genes) may have alternative functions beyond DNA replication including escape from host defense as with RNR in some herpes viruses [88]. The presence of multiple nested putative ORFs in viral and lepidopteran cSHMT (especially nested intronic ORFs in *H. zea* and *H. armigera*) may also point to additional uncharacterized functions and interactions for this gene (Figure 5; [98]).

From alignments of publicly available sequence data, HiRise assembled genomes, and PCR generated sequences from *H. zea*, and *H. zea* × *H. armigera* hybrids, the cSHMT gene of *H. zea* appears to have undergone several changes (mainly to intronic regions) compared to the sister species *H. armigera* (Figures 5 and 6). Given the presence of cSHMT in HzNV-2, the structural changes in *H. zea* cSHMT may affect host–virus interaction. We focused on the SINE found in intron 7, as homologous sequences were found throughout the *H. zea* genome in association with functional genes such as cytochrome P450. The cytochrome P450 family of enzymes are well known in the metabolism of xenobiotics [99] and have thus been much studied for their role in producing resistance to various insecticidal compounds [100,101]. In addition to metabolizing various xenobiotic compounds, cytochrome P450 enzymes have also been described in several types of pathogen defense responses [102,103]. The insertion of TEs in or near *H. armigera* cytochrome P450 genes has been linked to increased resistance to insecticidal compounds [72]. While TE insertion into genes can result in a loss of function, they have also been noted in providing functional advantages such as linking genes from distant regions of the genome through coordinated transcription in response to an external agent [104]. As such, could the potential linking of *H. zea* cSHMT to different cytochrome P450 genes (as well as many other genes) be related to the interaction of host and viral copies of cSHMT? Examples of cytochrome P450 genes responding to viral infection have been observed in *H. zea* to HzSNPV (*H. zea* single-nucleocapsid nucleopolyhedrovirus) [105], and *Trichoplusia ni* to *Autographa californica* multiple nucleopolyhedrovirus [106] as well as in more distant lineages such as *Aedes aegypti* to Dengue virus infection [103]. Viral infections have also been noted to increase the abundance of host TE transcripts in *Drosophila* [107]. That said, TE insertion may have some negative impacts on proper functioning of cSHMT, although mitochondrial copies of SHMT have been shown to be able to compensate in instances of loss of function in cSHMT in mammalian models [108,109]. To this point, it should be noted that all *H. zea* that were heterozygous for the TE in intron 7 of cSHMT were not found to contain HzNV-2 DNA (Table S1). As our sample sizes were small, more follow-up work should be conducted to understand how TE-bearing cSHMT copies and viral infection are correlated. Lastly our observation of hybrid gel banding patterns (Supplementary Figure S4) as well as sequence data (Figure 6) show that the TE in cSHMT is inherited in most hybrids (especially given the high rates of TE homozygosity in *H. zea*) which may provide novel adaptive links between genes which are not present in either of the parents, potentially exacerbating the spread of coadapted hybrids in the Americas [21]. The different arrangements of *H. zea* and *H. armigera* cSHMT (especially regarding intron sequences) will provide a useful variable locus for the development of species- and hybrid-level markers for use in screening projects. The presence of cSHMT in HzNV-2 is unique among viruses studied thus far and may provide an exceptional system (virus–host interaction and evolution) for understanding how this essential gene functions

and novel methods for blocking or altering this gene in the development of antifolate-like medications [90,110–112] or the development of folate-disruption-based sterility techniques for pest insect control [113].

When the Hz2V091 ORF was BLASTn searched, the region of greatest similarity was associated with the direct repeat (dr) 4 identified in Burand et al. [16] to numerous bacterial sequences (Supplementary Figures S1 and S3). Such repeats are known to be involved with creating binding recognition sites in translated proteins associated with bacterial cell surface interactions with host cells [114,115]. In addition, binding proteins are known to evolve via modular assemblage/recombination outside of the dr sites [116–118]. This is consistent with the bacterial dr-containing gene in *Helicoverpa* made up of a 5' region with matches to noctuid DNA, the dr region similar to HzNV-2 and bacterial motifs, and a region 3' of the dr matching to several noctuid species (Supplementary Figure S3). This too appears to be the case in Hz2V091, with a 5' region matching to repeats from *Plasmodium* and the dr4 region matching to bacterial sequences. Some of the outstanding questions related to this genomic region are: (1) How and when did this gene or motif (drs) transfer to or evolve in *H. zea* or HzNV-2 and what function does it have? (2) Did HzNV-2 act as an intermediary from bacteria to *H. zea*? (3) Was the dr-containing gene in *H. zea* the source of Hz2V091 with recombination in HzNV-2 thereafter to attain the current ORF sequence? (4) Did virus and host acquire the same gene motifs through separate unrelated events? More work is needed to answer these questions but a functional role for the dr-containing gene in *Helicoverpa* is reasonable based on similar genes from other insects [119] and the presence of the dr4-like sequence in the cds of gene mRNA XM_049842877. Additionally, the dr4 match in *Helicoverpa* genomes is not arrayed throughout the genome or paired with other similar drs and thus not convergent to viral or bacterial copies through chance paralogous duplication and mutation.

The acquisition of genes from bacteria that co-occur in the environment or coinfect host cells appears to be commonplace among large dsDNA viruses [84,120,121]. From analysis of several large dsDNA virus genomes, most genes acquired from bacteria are related to DNA replication and repair, or are membrane proteins [120]. Only two other HzNV-2 ORFs (Hz2V099; *Psychromonas ingrahamii*, esterase/lipase and Hz2V110; *Trichomonas vaginalis*, protein kinase) were found to contain sequences matching to other microbes in the Burand et al. [16] annotation. Most of the BLAST hits for ORF Hz2V091 in our comparisons were for cell surface proteins suggesting a possible cell-surface recognition/manipulation function for this gene [122].

The cooccurrence of HzNV-2 and bacteria such as *Brevibacillus laterosporus* and *Bacillus thuringiensis* in the insect host [123] provide suitable conditions for HzNV-2 to acquire bacterial genes for novel functions that may influence fitness. While HzNV-2 and similar viruses have been shown to acquire bacterial genes, the degree to which bacteria may be involved in the life cycle of these viruses, in roles such as secondary hosts, vectors, or agents of coinfection increasing pathogenicity (e.g., [124]), has not been thoroughly studied. The ORFs with motifs of bacterial origin in HzNV-2 are clustered closely in the genome, a pattern found in other large dsDNA viruses [120]. The genomic region where Hz2V091 and the 5' and 3' flanking ORFs (ORFs Hz2V090 and Hz2V092 with regions similar to *Acinetobacter soli*, e-value = 3×10^{-53} and *A. junii*, e-value = 1×10^{-42} , respectively, as attained from BLASTn searches) are located in the HzNV-2 genome is absent in the closely related HzNV-1 [16]. This suggests that this cluster of three ORFs may have been recently acquired by HzNV-2 or lost by HzNV-1 and by extension could be an important source of functional differences. Additionally, this region is enriched in direct repeats [16], with the Hz2V090 ORF possessing a protein motif from a DNA intra-strand crosslink recognition protein (e-value = 8×10^{-154} top hit when Hz2V090 is BLASTp searched against the nr database) which suggests that this area may be a recombination and exogenous DNA uptake hotspot [125–127]. The acquisition not only of host genetic material but of genetic material from bacteria and other microbes may be critical in gaining novel functions, adapting to new hosts, altering viral life cycles, broadening vector dynamics, and ultimately

resulting in functional differences between HzNV strains as well as in other large dsDNA viruses [84,120,121]. Because HzNV-2 appears able to acquire bacterial genes (or parts of bacterial genes), these types of viruses might also be involved with symbiont-induced insect speciation through the manipulation of microbial symbionts and/or host–microbe interactions [28].

Viral infection alone is unlikely to result in speciation of the host. However, long-term coexistence between virus and host can affect populations coevolving with viral pressure differently by providing an ‘environment’ more favorable to high mutation rates [128,129] versus populations without viral pressure, potentially increasing divergence rates. Rapid evolution of host response and immunity loci may not result in reproductive isolation of populations with viral pressure but could result in local extinction of infected populations without immunity and, in turn, accelerate allopatric speciation [130,131]. Viruses, in rare events, can also have a more direct effect on divergence by increasing the fitness of hosts when genetic rearrangements (including viral integration) in response to foreign DNA result in a gain of novel function [9,29,38,132]. This may allow an insect population to occupy a new niche and could help drive ecological speciation. A pattern of differentiation is evident when comparing *H. zea* and *H. armigera* where divergence in the cSHMT gene was noted between these two species. Given this pattern between the two *Helicoverpa* species, HzNV-2 may have evolved to infect *H. zea* after or as part of the divergence of *H. armigera* from *H. zea*. If this is the case, then HzNV-2 may have evolved to infect *H. zea* in the last 1.5 my based on whole genome divergence estimates between *H. zea* and *H. armigera* [133]. The closely related baculoviruses HaSNPV (*H. armigera* SNPV) and HzSNPV provide a useful comparison to HzNV-2 evolution given that speciation in the two baculovirus species followed host speciation [134]. If assuming a similar time since divergence between the two *Helicoverpa* SNPVs and HzNV-2 from a yet undetermined sister lineage, the rate of host-gene acquisition since divergence varies, with one acquisition in HzSNPV, to three in HaSNPV to four or more in HzNV-2 when using *H. armigera* wgs contigs as a basis for comparison (Supplementary Figure S8). Viral prevalence also seems to vary between viral lineages with rates of 31% in individual samples and 58% found in bulk samples for HzNV-2 from this study (Supplementary Tables S1 and S2), while infection rates of only 13% were noted from a field study of wildtype HaSNPV [135]. These differences could be related to the lifecycles of these viruses where latent infections have been noted in HzNV-2 [78] and high rates of mortality are known from *Helicoverpa* SNPVs. As with many questions in evolution, the determination of event order becomes an important part of linking pattern to process. For instance, did the ancestors of HzNV-2 acquire numerous essential host genes resulting in longer viral persistence in the host (and as a result, higher rates of prevalence), or did some other factor enable longer viral persistence in the host and increase the opportunity for host-gene acquisitions? Given the pronounced differences and shared ancestry between baculoviruses and nudiviruses, these lineages provide a superb comparative study system for understanding viral evolution.

Insect-virus infections should be characterized in greater detail, including HzNV-2 with *Helicoverpa*, to understand how these viruses might mediate speciation events and how viral, microbial, and host genes interact at the molecular level. Among animals, the insects (Hexapoda) contain the greatest species diversity [136], and if the number of viruses in each insect species is similar to that of other animal lineages [137,138], the number of insect viruses could greatly exceed 50 million (=1 million described insect species $\times$ 50 viruses/insect species). As such, the greatest number of animal viruses and by extension novel molecular interactions, some of which could have applications for biotechnology, are likely to be found among insects. Further studies of viruses should be encouraged to gain a better understanding of how they have influenced insect evolution and, in turn, the biosphere.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/insects14100797/s1>. Table S1, Individual samples tested for presence of HzNV-2 with ddPCR. Samples with GenBank accession numbers were Sanger sequenced; Table S2, ddPCR results from bulk samples tested for prevalence of HzNV-2; Table S3, Results of model tests for each gene cds alignment used in the DNA-based phylogenetic analysis; Figure S1, Translated HzNV-2 ORFs inferred to be recent host acquisitions BLASTP searched against the clustered nr database; Figure S2, Phylogenetic analyses of sequences from (A) RNR; (B) cSHMT; (C) PCFT; and (D) TS. Figure S3, Best sequence matches mapped to (A) HzNV-2 ORF Hz2V091 and (B) an *H. zea* gene of unknown function with a central region similar to HzNV-2 dr4 in ORF Hz2V091; Figure S4, Agarose gel (1%) of PCR products spanning the TE insert in intron 7 of cSHMT from *H. zea*, and lab-reared hybrids. HzNV-2 test results on these samples can be found in Supplementary Table S1. *Refers to samples that were Sanger sequenced; accessions given in Table S1; Figure S5, The cSHMT Intron 7 TE-inserted specimen with gel banding pattern similar to *H. zea* x *H. armigera* lab hybrid specimens but is identified as *H. zea* using the validated real-time PCR method from Gilligan et al. (2015); Figure S6, The size distribution of TEs with similar sequence identities (cutoff e-value < 0.01) to those found in P450 genes and cSHMT intron 7 across a subset of chromosomes; Figure S7, The proportion of each *H. zea* chromosome made up of TEs with similar sequence identities (cutoff e-value < 0.01) to those found in P450 genes and cSHMT intron 7. Chromosomes ordered largest (15.51 Mb chr 1) to smallest (6.32 Mb chr 30) with the Z chromosome at 18.81 Mb listed last; Figure S8, Inferred recent host DNA acquisitions mapped to baculoviruses HaSNPV (blue), HzSNPV (green), and nudivirus HzNV-2 (orange). Arrows indicate inferred host DNA acquisitions since *H. armigera* and *H. zea* divergence (for the baculoviruses only as the sister species of HzNV-2 has not been determined). All searches used complete viral genomes as queries to wgs *H. armigera* taxid 29058. Viral genomes not to scale.

Author Contributions: Conception and design of experiments, T.M.G., L.R.T. and F.A.Z.; ddPCR instrumentation and analyses, F.A.Z.; phylogenetic and network analyses, L.R.T.; sample acquisition, T.M.G., L.R.T. and F.A.Z.; specimen dissections, T.M.G.; comparative genomic analyses, L.R.T.; generation of sequence data, L.R.T.; writing—original draft preparation, L.R.T.; writing—review and editing, T.M.G., L.R.T. and F.A.Z.; funding acquisition, T.M.G. and L.R.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded, in part, through the Plant Protection Act (PPA) Section 7721 provided by the U.S. Department of Agriculture (USDA) Animal and Plant Health Inspection Service (APHIS) through Cooperative Agreement AP18PPQS&T00C074 awarded to Colorado State University. It may not necessarily express APHIS' views. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA; USDA is an equal opportunity provider and employer.

Data Availability Statement: All sequence data can be accessed through NCBI GenBank at the accession numbers OR609382–OR609386 for the Sanger data and under BioProject ID PRJNA1020878 for the genomic data. All other data are available by request, please contact Luke Tembrock (tembrock@colostate.edu).

Acknowledgments: Special thanks to John Burand and Mark Stenglein for their helpful insights on viral evolution while conceiving this study and to Laura Trozzo and Hannah Nadel for providing lab-reared *Helicoverpa* specimens for use in this study. Thanks to Wee Tek Tay and Tom Walsh of CSIRO for providing *H. armigera* samples from Australia and Joaquín Baixeras from the University of Valencia for providing *H. armigera* from Spain. Thanks to Janet Hardin from Colorado State University in Fort Collins, Colorado, USA for her ongoing efforts to trap and collect insects during survey work which we utilized in this and many other studies. Thanks to Thomas Swale, Shaune Hall, and Jasmine Haimovitz of Dovetail genomics for conducting HiC and CHiCAGO library preparation, Illumina sequencing, assembly, and scaffolding of the HiRise genomes used here. Thanks to the USDA-APHIS Cooperative Agricultural Pest Survey (CAPS) program and cooperators in all states that provided *H. zea* specimens for testing.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Krupovic, M.; Bamford, D.H. Order to the viral universe. *J. Virol.* **2010**, *84*, 12476–12479. [\[CrossRef\]](#)
- Dougan, T.J.; Quake, S.R. Viral taxonomy derived from evolutionary genome relationships. *PLoS ONE* **2019**, *14*, e0220440. [\[CrossRef\]](#) [\[PubMed\]](#)
- Koonin, E.V.; Dolja, V.V. Virus world as an evolutionary network of viruses and capsidless selfish elements. *Microbiol. Mol. Biol. R.* **2014**, *78*, 278–303. [\[CrossRef\]](#) [\[PubMed\]](#)
- Wolf, Y.I.; Kazlauskas, D.; Iranzo, J.; Lucia-Sanz, A.; Kuhn, J.H.; Krupovic, M.; Dolja, V.V.; Koonin, E.V. Origins and Evolution of the Global RNA Virome. *mBio* **2018**, *9*, e02329–18. [\[CrossRef\]](#) [\[PubMed\]](#)
- Koonin, E.V.; Krupovic, M.; Yutin, N. Evolution of double-stranded DNA viruses of eukaryotes: From bacteriophages to transposons to giant viruses. *Ann. N. Y. Acad. Sci.* **2015**, *1341*, 10–24. [\[CrossRef\]](#)
- Smith, J.S.; Robinson, N.J. Age-specific prevalence of infection with herpes simplex virus types 2 and 1: A global review. *J. Infect. Dis.* **2002**, *186* (Suppl. S1), S3–S28. [\[CrossRef\]](#)
- Doorbar, J. Latent papillomavirus infections and their regulation. *Curr. Opin. Virol.* **2013**, *3*, 416–421. [\[CrossRef\]](#)
- Bezier, A.; Annaheim, M.; Herbinier, J.; Wetterwald, C.; Gyapay, G.; Bernard-Samain, S.; Wincker, P.; Roditi, I.; Heller, M.; Belghazi, M.; et al. Polydnviruses of braconid wasps derive from an ancestral nudivirus. *Science* **2009**, *323*, 926–930. [\[CrossRef\]](#)
- Pichon, A. Recurrent DNA virus domestication leading to different parasite virulence strategies. *Sci. Adv.* **2015**, *1*, e1501150. [\[CrossRef\]](#)
- Petersen, J.M.; Bezier, A.; Drezen, J.M.; van Oers, M.M. The naked truth: An updated review on nudiviruses and their relationship to bracoviruses and baculoviruses. *J. Invertebr. Pathol.* **2022**, *189*, 107718. [\[CrossRef\]](#)
- James, C.; Harfouche, M.; Welton, N.J.; Turner, K.M.; Abu-Raddad, L.J.; Gottlieb, S.L.; Looker, K.J. Herpes simplex virus: Global infection prevalence and incidence estimates, 2016. *Bull. World Health Organ.* **2020**, *98*, 315–329. [\[CrossRef\]](#) [\[PubMed\]](#)
- Lupiani, B.; Raina, A.K.; Huber, C. Development and use of a PCR assay for detection of the reproductive virus in wild populations of *Helicoverpa zea* (Lepidoptera: Noctuidae). *J. Invertebr. Pathol.* **1999**, *73*, 107–112. [\[CrossRef\]](#) [\[PubMed\]](#)
- Herzog, G.A.; Phillips, J.R. Manifestation of an abnormal reproductive system in a laboratory strain of the bollworm *Heliothis zea* Lepidoptera, Noctuidae. *J. Georgia Entomol. Soc.* **1982**, *17*, 506–513.
- Knell, R.J.; Webberley, K.M. Sexually transmitted diseases of insects: Distribution, evolution, ecology and host behaviour. *Biol. Rev.* **2004**, *79*, 557–581. [\[CrossRef\]](#) [\[PubMed\]](#)
- Wang, Y.; Burand, J.P.; Jehle, J.A. Nudivirus genomics: Diversity and classification. *Virol. Sin.* **2008**, *22*, 128–136. [\[CrossRef\]](#)
- Burand, J.P.; Kim, W.; Afonso, C.L.; Tulman, E.R.; Kutish, G.F.; Lu, Z.; Rock, D.L. Analysis of the genome of the sexually transmitted insect virus *Helicoverpa zea* nudivirus 2. *Viruses* **2012**, *4*, 28–61. [\[CrossRef\]](#) [\[PubMed\]](#)
- Raina, A.K.; Adams, J.R. Gonad-specific virus of corn earworm. *Nature* **1998**, *374*, 770. [\[CrossRef\]](#)
- Hamm, J.J.; Carpenter, J.E.; Styer, E.L. Oviposition day effect on incidence of agonadal progeny of *Helicoverpa zea* (Lepidoptera: Noctuidae) infected with a virus. *Ann. Entomol. Soc. Am.* **1996**, *89*, 266–275. [\[CrossRef\]](#)
- Raina, A.K.; Adams, J.R.; Lupiani, B.; Lynn, D.E.; Kim, W.J.; Burand, J.P.; Dougherty, E.M. Further characterization of the gonad-specific virus of corn earworm, *Helicoverpa zea*. *J. Invertebr. Pathol.* **2000**, *76*, 6–12. [\[CrossRef\]](#)
- Burand, J.P. Pathology and replication of the sexually transmitted insect virus HzNV-2. In *Virology I: HIV and Related Issues*; Iconcept Press: Kowloon, Hong Kong, 2013.
- Valencia-Montoya, W.A.; Elfekih, S.; North, H.L.; Meier, J.I.; Warren, I.A.; Tay, W.T.; Gordon, K.H.J.; Specht, A.; Paula-Moraes, S.V.; Rane, R.; et al. Adaptive introgression across semipermeable species boundaries between local *Helicoverpa zea* and invasive *Helicoverpa armigera* moths. *Mol. Biol. Evol.* **2020**, *37*, 2568–2583. [\[CrossRef\]](#)
- Hardwick, D.F. The biological status of “*Heliothis stromboli*”. *Can. Entomol.* **1970**, *102*, 339–341. [\[CrossRef\]](#)
- Pogue, M.G. A new synonym of *Helicoverpa zea* (Bodie) and differentiation of adult males of *H. zea* and *H. armigera* (Hubner) (Lepidoptera: Noctuidae: Heliothinae). *Ann. Entomol. Soc. Am.* **2004**, *97*, 1222–1226. [\[CrossRef\]](#)
- Balbi, E.I.; Flores, F.M.; Tosto, D.S.; Arneodo, J.D. Further description of *Helicoverpa zea* (Lepidoptera: Noctuidae) male genitalia and new genetic evidence of synonymy with respect to the anomalous form, “*Heliothis stromboli*”. *J. Insect Sci.* **2017**, *17*, 1–6. [\[CrossRef\]](#) [\[PubMed\]](#)
- Nagoshi, R.N.; Gilligan, T.M.; Brambila, J. Combining Tpi and CO1 genetic markers to discriminate invasive *Helicoverpa armigera* from local *Helicoverpa zea* (Lepidoptera: Noctuidae) populations in the Southeastern United States. *J. Econ. Entomol.* **2016**, *109*, 2115–2124. [\[CrossRef\]](#)
- Wallin, I.E. *Symbioticism and the Origin of Species*; Williams and Wilkins Company: Philadelphia, PA, USA, 1927.
- Margulis, L.; Sagan, D. *Acquiring Genomes: A Theory of the Origins of Species*; Basic Books: New York, NY, USA, 2002.
- Brucker, R.M.; Bordenstein, S.R. Speciation by symbiosis. *Trends Ecol. Evol.* **2012**, *27*, 443–451. [\[CrossRef\]](#)
- Roossinck, M.J. The good viruses: Viral mutualistic symbioses. *Nat. Rev. Microbiol.* **2011**, *9*, 99–108. [\[CrossRef\]](#)
- Michel, G.; Tonon, T.; Scornet, D.; Cock, J.M.; Kloareg, B. Central and storage carbon metabolism of the brown alga *Ectocarpus siliculosus*: Insights into the origin and evolution of storage carbohydrates in Eukaryotes. *New Phytol.* **2010**, *188*, 67–81. [\[CrossRef\]](#)
- Aiewsakun, P.; Katzourakis, A. Endogenous viruses: Connecting recent and ancient viral evolution. *Virology* **2015**, *479–480*, 26–37. [\[CrossRef\]](#)
- Gilbert, C.; Cordaux, R. Viruses as vectors of horizontal transfer of genetic material in eukaryotes. *Curr. Opin. Virol.* **2017**, *25*, 16–22. [\[CrossRef\]](#)

33. Irwin, N.A.T.; Pittis, A.A.; Richards, T.A.; Keeling, P.J. Systematic evaluation of horizontal gene transfer between eukaryotes and viruses. *Nat. Microbiol.* **2022**, *7*, 327–336. [\[CrossRef\]](#)
34. Rodriguez-Valera, F.; Martin-Cuadrado, A.B.; Rodriguez-Brito, B.; Pasic, L.; Thingstad, T.F.; Rohwer, F.; Mira, A. OPINION Explaining microbial population genomics through phage predation. *Nat. Rev. Microbiol.* **2009**, *7*, 828–836. [\[CrossRef\]](#)
35. Manry, J.; Laval, G.; Patin, E.; Fornarino, S.; Itan, Y.; Fumagalli, M.; Sironi, M.; Tichit, M.; Bouchier, C.; Casanova, J.L.; et al. Evolutionary genetic dissection of human interferons. *J. Exp. Med.* **2011**, *208*, 2747–2759. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Obbard, D.J.; Dudas, G. The genetics of host-virus coevolution in invertebrates. *Curr. Opin. Virol.* **2014**, *8*, 73–78. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Maori, E.; Tanne, E.; Sela, I. Reciprocal sequence exchange between non-retro viruses and hosts leading to the appearance of new host phenotypes. *Virology* **2007**, *362*, 342–349. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Braga, L.P.P.; Soucy, S.M.; Amgarten, D.E.; da Silva, A.M.; Setubal, J.C. Bacterial diversification in the light of the interactions with phages: The genetic symbionts and their role in ecological speciation. *Front. Ecol. Evol.* **2018**, *6*, 1–12. [\[CrossRef\]](#)
39. Fridman, S.; Flores-Urbe, J.; Larom, S.; Alalouf, O.; Liran, O.; Yacoby, I.; Salama, F.; Bailleul, B.; Rappaport, F.; Ziv, T.; et al. A myovirus encoding both photosystem I and II proteins enhances cyclic electron flow in infected *Prochlorococcus* cells. *Nature Microbiol.* **2017**, *2*, 1350–1357. [\[CrossRef\]](#)
40. Schulz, F.; Yutin, N.; Ivanova, N.N.; Ortega, D.R.; Lee, T.K.; Vierheilig, J.; Daims, H.; Horn, M.; Wagner, M.; Jensen, G.J.; et al. Giant viruses with an expanded complement of translation system components. *Science* **2017**, *356*, 82–85. [\[CrossRef\]](#)
41. Krupovic, M.; Dolja, V.V.; Koonin, E.V. Origin of viruses: Primordial replicators recruiting capsids from hosts. *Nat. Rev. Microbiol.* **2019**, *17*, 449–458. [\[CrossRef\]](#)
42. Duffy, S.; Shackelton, L.A.; Holmes, E.C. Rates of evolutionary change in viruses: Patterns and determinants. *Nat. Rev. Genet.* **2008**, *9*, 267–276. [\[CrossRef\]](#)
43. Belshaw, R.; Sanjuan, R.; Pybus, O.G. Viral mutation and substitution: Units and levels. *Curr. Opin. Virol.* **2011**, *1*, 430–435. [\[CrossRef\]](#)
44. Moreira, D.; Lopez-Garcia, P. Evolution of viruses and cells: Do we need a fourth domain of life to explain the origin of eukaryotes? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2015**, *370*, 20140327. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Niikura, M.; Dodgson, J.; Cheng, H. Direct evidence of host genome acquisition by the alphaherpesvirus Marek's disease virus. *Arch. Virol.* **2006**, *151*, 537–549. [\[CrossRef\]](#) [\[PubMed\]](#)
46. Tang, S.; Patel, A.; Krause, P.R. Herpes simplex virus ICP27 regulates alternative pre-mRNA polyadenylation and splicing in a sequence-dependent manner. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 12256–12261. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Brambila, J. *Instructions for Dissecting Male Genitalia of Helicoverpa (Lepidoptera: Noctuidae) to Separate H. zea from H. armigera (Hubner)*; USDA-APHIS-PPQ: Gainesville, FL, USA, 2009.
48. Gilligan, T.M.; Tembrock, L.R.; Farris, R.E.; Barr, N.B.; van der Straten, M.J.; van de Vossenberg, B.T.; Metz-Verschure, E. A multiplex real-time PCR assay to diagnose and separate *Helicoverpa armigera* and *H. zea* (Lepidoptera: Noctuidae) in the New World. *PLoS ONE* **2015**, *10*, e0142912. [\[CrossRef\]](#)
49. Zink, F.A.; Tembrock, L.R.; Timm, A.E.; Farris, R.E.; Perera, O.P.; Gilligan, T.M. A droplet digital PCR (ddPCR) assay to detect *Helicoverpa armigera* (Lepidoptera: Noctuidae) in bulk trap samples. *PLoS ONE* **2017**, *12*, e0178704. [\[CrossRef\]](#)
50. Tembrock, L.R.; Farris, R.E.; Ledezma, L.; Barr, N.B.; Gilligan, T.M. A real-Time PCR assay for the separation of *Autographa gamma* (Noctuidae: Plusiinae) from morphologically similar species in North America. *J. Econ. Entomol.* **2017**, *110*, 2609–2617. [\[CrossRef\]](#)
51. Gloor, G.B.; Preston, C.R.; Johnson-Schlitz, D.M.; Nassif, N.A.; Phillis, R.W.; Benz, W.K.; Robertson, H.M.; Engels, W.R. Type I repressors of P element mobility. *Genetics* **1993**, *135*, 81–95. [\[CrossRef\]](#)
52. Perera, O.P.; Allen, K.C.; Jain, D.; Purcell, M.; Little, N.S.; Luttrell, R.G. Rapid identification of *Helicoverpa armigera* and *Helicoverpa zea* (Lepidoptera: Noctuidae) using ribosomal RNA internal transcribed spacer 1. *J. Insect Sci.* **2015**, *15*, 155. [\[CrossRef\]](#)
53. Zink, F.A.; Tembrock, L.R.; Timm, A.E.; Gilligan, T.M. A ddPCR assay for identification of *Autographa gamma* (Noctuidae: Plusiinae) in bulk trap samples. *J. Econ. Entomol.* **2018**, *111*, 1490–1495. [\[CrossRef\]](#)
54. Jones, M.; Williams, J.; Gartner, K.; Phillips, R.; Hurst, J.; Frater, J. Low copy target detection by droplet digital PCR through application of a novel open access bioinformatic pipeline, 'definetherain'. *J. Virol. Methods* **2014**, *202*, 46–53. [\[CrossRef\]](#)
55. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [\[CrossRef\]](#) [\[PubMed\]](#)
56. Pertsemlidis, A.; Fondon, J.W., 3rd. Having a BLAST with bioinformatics (and avoiding BLASTphemy). *Genome Biol.* **2001**, *2*, REVIEWS2002. [\[CrossRef\]](#)
57. Altschul, S.F.; Madden, T.L.; Schaffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402. [\[CrossRef\]](#) [\[PubMed\]](#)
58. Stahlke, A.R.; Chang, J.; Tembrock, L.R.; Sim, S.B.; Chudalayandi, S.; Geib, S.M.; Scheffler, B.E.; Perera, O.P.; Gilligan, T.M.; Childers, A.K.; et al. A chromosome-scale genome assembly of a *Helicoverpa zea* strain resistant to *Bacillus thuringiensis* Cry1Ac insecticidal protein. *Genome Biol. Evol.* **2023**, *15*, evac131. [\[CrossRef\]](#) [\[PubMed\]](#)
59. Wheeler, D.L.; Church, D.M.; Federhen, S.; Lash, A.E.; Madden, T.L.; Pontius, J.U.; Schuler, G.D.; Schriml, L.M.; Sequeira, E.; Tatusova, T.A.; et al. Database resources of the National Center for Biotechnology. *Nucleic Acids Res.* **2003**, *31*, 28–33. [\[CrossRef\]](#) [\[PubMed\]](#)

60. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [\[CrossRef\]](#)
61. Posada, D.; Crandall, K.A. MODELTEST: Testing the model of DNA substitution. *Bioinformatics* **1998**, *14*, 817–818. [\[CrossRef\]](#)
62. Huelsenbeck, J.P.; Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **2001**, *17*, 754–755. [\[CrossRef\]](#)
63. Swofford, D.L. *PAUP*. Phylogenetic Analysis Using Parsimony (and Other Methods)*; Sinauer Associates: Sunderland, UK, 2003.
64. Katoh, K.; Misawa, K.; Kuma, K.; Miyata, T. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **2002**, *30*, 3059–3066. [\[CrossRef\]](#)
65. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [\[CrossRef\]](#)
66. Huson, D.H.; Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **2006**, *23*, 254–267. [\[CrossRef\]](#) [\[PubMed\]](#)
67. Untergasser, A.; Cutcutache, I.; Koressaar, T.; Ye, J.; Faircloth, B.C.; Remm, M.; Rozen, S.G. Primer3—New capabilities and interfaces. *Nucleic Acids Res.* **2012**, *40*, e115. [\[CrossRef\]](#) [\[PubMed\]](#)
68. SantaLucia, J., Jr. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 1460–1465. [\[CrossRef\]](#)
69. Lieberman-Aiden, E.; van Berkum, N.L.; Williams, L.; Imakaev, M.; Ragoczy, T.; Telling, A.; Amit, I.; Lajoie, B.R.; Sabo, P.J.; Dorschner, M.O.; et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **2009**, *326*, 289–293. [\[CrossRef\]](#) [\[PubMed\]](#)
70. Cairns, J.; Freire-Pritchett, P.; Wingett, S.W.; Varnai, C.; Dimond, A.; Plagnol, V.; Zerbino, D.; Schoenfelder, S.; Javierre, B.M.; Osborne, C.; et al. CHiCAGO: Robust detection of DNA looping interactions in Capture Hi-C data. *Genome Biol.* **2016**, *17*, 127. [\[CrossRef\]](#) [\[PubMed\]](#)
71. Putnam, N.H.; O’Connell, B.L.; Stites, J.C.; Rice, B.J.; Blanchette, M.; Calef, R.; Troll, C.J.; Fields, A.; Hartley, P.D.; Sugnet, C.W.; et al. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Res.* **2016**, *26*, 342–350. [\[CrossRef\]](#)
72. Chen, S.; Li, X.C. Transposable elements are enriched within or in close proximity to xenobiotic-metabolizing cytochrome P450 genes. *BMC Evol. Biol.* **2007**, *7*, 46. [\[CrossRef\]](#)
73. Klai, K.; ChEnais, B.; Zidi, M.; Djebbi, S.; Caruso, A.; Denis, F.; Confais, J.; Badawi, M.; Casse, N.; Mezghani Khemakhem, M. Screening of *Helicoverpa armigera* mobilome revealed transposable element insertions in insecticide resistance genes. *Insects* **2020**, *11*, 879. [\[CrossRef\]](#)
74. Gray, Y.H. It takes two transposons to tango: Transposable-element-mediated chromosomal rearrangements. *Trends Genet.* **2000**, *16*, 461–468. [\[CrossRef\]](#)
75. Tu, Z. Insect transposable elements. In *Insect Molecular Biology and Biochemistry*; Gilbert, L.I., Ed.; Elsevier Academic Press: Amsterdam, The Netherlands, 2012.
76. Zink, F.A.; Tembrock, L.R.; Timm, A.E.; Gilligan, T.M. A duplex ddPCR assay for simultaneously detecting *Ips sexdentatus* and *Ips typographus* (Coleoptera: Curculionidae) in bulk trap samples. *Can. J. For. Res.* **2019**, *49*, 903–914. [\[CrossRef\]](#)
77. Franke, G.N.; Maier, J.; Wildenberger, K.; Cross, M.; Giles, F.J.; Muller, M.C.; Hochhaus, A.; Niederwieser, D.; Lange, T. Comparison of real-time quantitative PCR and digital droplet PCR for BCR-ABL1 monitoring in patients with chronic myeloid leukemia. *J. Mol. Diagn.* **2020**, *22*, 81–89. [\[CrossRef\]](#) [\[PubMed\]](#)
78. Wu, Y.L.; Wu, C.P.; Liu, C.Y.Y.; Hsu, P.W.C.; Wu, E.C.; Chao, Y.C. A non-coding RNA of insect HzNV-1 virus establishes latent viral infection through microRNA. *Sci. Rep.* **2011**, *1*, 60. [\[CrossRef\]](#)
79. Tuxen, S.L. *Taxonomists Glossary of Genitalia of Insects*; Munksgaard: Copenhagen, Denmark, 1970.
80. Masly, J.P. 170 years of “lock-and-key”: Genital morphology and reproductive isolation. *Int. J. Evol. Biol.* **2012**, *2012*, 247352. [\[CrossRef\]](#) [\[PubMed\]](#)
81. Williams, T. Natural invertebrate hosts of iridoviruses (Iridoviridae). *Neotrop. Entomol.* **2008**, *37*, 615–632. [\[CrossRef\]](#) [\[PubMed\]](#)
82. Herniou, E.A.; Huguet, E.; Theze, J.; Bezier, A.; Periquet, G.; Drezen, J.M. When parasitic wasps hijacked viruses: Genomic and functional evolution of polydnaviruses. *Philos. Trans. R. Soc. B* **2013**, *368*, 20130051. [\[CrossRef\]](#) [\[PubMed\]](#)
83. DaPalma, T.; Doonan, B.P.; Trager, N.M.; Kasman, L.M. A systematic approach to virus-virus interactions. *Virus Res.* **2010**, *149*, 1–9. [\[CrossRef\]](#)
84. Diesend, J.; Kruse, J.; Hagedorn, M.; Hammann, C. Amoebae, giant viruses, and virophages make up a complex, multilayered threesome. *Front. Cell Inf. Microbiol.* **2018**, *7*, 527. [\[CrossRef\]](#)
85. Lundin, D.; Gribaldo, S.; Torrents, E.; Sjöberg, B.M.; Poole, A.M. Ribonucleotide reduction—Horizontal transfer of a required function spans all three domains. *BMC Evol. Biol.* **2010**, *10*, 383. [\[CrossRef\]](#)
86. Shackelton, L.A.; Holmes, E.C. The evolution of large DNA viruses: Combining genomic information of viruses and their hosts. *Trends Microbiol.* **2004**, *12*, 458–465. [\[CrossRef\]](#)
87. Ball, A.L. Virus replication strategies. In *Fields Virology*; Knipe, D.M., Howley, P.M., Eds.; Lippincott Williams & Wilkins: Philadelphia, PA, USA, 2007.
88. Lembo, D.; Brune, W. Tinkering with a viral ribonucleotide reductase. *Trends Biochem. Sci.* **2009**, *34*, 25–32. [\[CrossRef\]](#)
89. Carreras, C.W.; Santi, D.V. The catalytic mechanism and structure of thymidylate synthase. *Annu. Rev. Biochem.* **1995**, *64*, 721–762. [\[CrossRef\]](#) [\[PubMed\]](#)

90. Arvizu-Flores, A.A.; Aispuro-Hernandez, E.; Garcia-Orozco, K.D.; Varela-Romero, A.; Valenzuela-Soto, E.; Velazquez-Contreras, E.F.; Rojo-Dominguez, A.; Yepiz-Plascencia, G.; Maley, F.; Sotelo-Mundo, R.R. Functional identity of the active sites of crustacean and viral thymidylate synthases. *Comp. Biochem. Physiol. C Toxicol. Pharmacol.* **2009**, *150*, 406–413. [\[CrossRef\]](#) [\[PubMed\]](#)
91. Chabaud, S.; Lambert, H.; Sasseville, A.M.J.; Lavoie, H.; Guilbault, C.; Massie, B.; Landry, J.; Langelier, Y. The R1 subunit of herpes simplex virus ribonucleotide reductase has chaperone-like activity similar to Hsp27. *FEBS Lett.* **2003**, *545*, 213–218. [\[CrossRef\]](#) [\[PubMed\]](#)
92. Stover, P.; Schrich, V. Serine hydroxymethyltransferase catalyzes the hydrolysis of 5,10-methyltetrahydrofolate to 5-formyltetrahydrofolate. *J. Biol. Chem.* **1990**, *265*, 14227–14233. [\[CrossRef\]](#) [\[PubMed\]](#)
93. Hou, Z.; Matherly, L.H. Biology of the major facilitative folate transporters SLC19A1 and SLC46A1. *Curr. Top. Membr.* **2014**, *73*, 175–204. [\[CrossRef\]](#)
94. Hebbbring, S.J.; Chai, Y.; Ji, Y.; Abo, R.P.; Jenkins, G.D.; Fridley, B.; Zhang, J.; Eckloff, B.W.; Wieben, E.D.; Weinshilboum, R.M. Serine hydroxymethyltransferase 1 and 2: Gene sequence variation and functional genomic characterization. *J. Neurochem.* **2012**, *120*, 881–890. [\[CrossRef\]](#)
95. Newman, A.C.; Maddocks, O.D.K. One-carbon metabolism in cancer. *Br. J. Cancer* **2017**, *116*, 1499–1504. [\[CrossRef\]](#)
96. Haque, M.R.; Hirowatari, A.; Nai, N.; Furuya, S.; Yamamoto, K. Serine hydroxymethyltransferase from the silkworm *Bombyx mori*: Identification, distribution, and biochemical characterization. *Arch. Insect Biochem. Physiol.* **2019**, *102*, e21594. [\[CrossRef\]](#)
97. Zhang, G.; Hussain, M.; O'Neill, S.L.; Asgari, S. *Wolbachia* uses a host microRNA to regulate transcripts of a methyltransferase, contributing to dengue virus inhibition in *Aedes aegypti*. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 10276–10281. [\[CrossRef\]](#)
98. Orr, M.W.; Mao, Y.; Storz, G.; Qian, S.B. Alternative ORFs and small ORFs: Shedding light on the dark proteome. *Nucleic Acids Res.* **2020**, *48*, 1029–1042. [\[CrossRef\]](#)
99. Gillam, E.M.J.; Hunter, D.J.B. Chemical defense and exploitation biotransformation of xenobiotics by cytochrome P450 enzymes. In *Metal Ions in Life Sciences*; Sigel, A., Sigel, H., Sigel, R.K.O., Eds.; John Wiley and Sons: West Sussex, UK, 2007; Volume 3, pp. 477–560.
100. Chen, S.; Yang, Y.H.; Wu, Y.D. Correlation between fenvalerate resistance and cytochrome P450-mediated O-demethylation activity in *Helicoverpa armigera* (Lepidoptera: Noctuidae). *J. Econ. Entomol.* **2005**, *98*, 943–946. [\[CrossRef\]](#)
101. Epelboin, Y.; Wang, L.; Gai Gianetto, Q.; Choumet, V.; Gaborit, P.; Issaly, J.; Guidez, A.; Douche, T.; Chaze, T.; Matondo, M.; et al. CYP450 core involvement in multiple resistance strains of *Aedes aegypti* from French Guiana highlighted by proteomics, molecular and biochemical studies. *PLoS ONE* **2021**, *16*, e0243992. [\[CrossRef\]](#) [\[PubMed\]](#)
102. Kim, Y.C.; Kim, S.Y.; Paek, K.H.; Choi, D.; Park, J.M. Suppression of CaCYP1, a novel cytochrome P450 gene, compromises the basal pathogen defense response of pepper plants. *Biochem. Biophys. Res. Commun.* **2006**, *345*, 638–645. [\[CrossRef\]](#) [\[PubMed\]](#)
103. Behura, S.K.; Gomez-Machorro, C.; Harker, B.W.; deBruyn, B.; Lovin, D.D.; Hemme, R.R.; Mori, A.; Romero-Severson, J.; Severson, D.W. Global cross-talk of genes of the mosquito *Aedes aegypti* in response to dengue virus infection. *PLoS Negl. Trop. Dis.* **2011**, *5*, e1385. [\[CrossRef\]](#) [\[PubMed\]](#)
104. Cowley, M.; Oakley, R.J. Transposable elements re-wire and fine-tune the transcriptome. *PLoS Genet.* **2013**, *9*, e1003234. [\[CrossRef\]](#)
105. Noland, J.E.; Breitenbach, J.E.; Popham, H.J.; Hum-Musser, S.M.; Vogel, H.; Musser, R.O. Gut transcription in *Helicoverpa zea* is dynamically altered in response to baculovirus infection. *Insects* **2013**, *4*, 506–520. [\[CrossRef\]](#)
106. Shrestha, A.; Bao, K.; Chen, W.; Wang, P.; Fei, Z.; Blissard, G.W. Transcriptional responses of the *Trichoplusia ni* midgut to oral infection by the baculovirus *Autographa californica* multiple nucleopolyhedrovirus. *J. Virol.* **2019**, *93*, e00353-19. [\[CrossRef\]](#)
107. Roy, M.; Viginier, B.; Saint-Michel, E.; Arnaud, F.; Ratiner, M.; Fablet, M. Viral infection impacts transposable element transcript amounts in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 12249–12257. [\[CrossRef\]](#)
108. MacFarlane, A.J.; Liu, X.; Perry, C.A.; Flodby, P.; Allen, R.H.; Stabler, S.P.; Stover, P.J. Cytoplasmic serine hydroxymethyltransferase regulates the metabolic partitioning of methylenetetrahydrofolate but is not essential in mice. *J. Biol. Chem.* **2008**, *283*, 25846–25853. [\[CrossRef\]](#)
109. Anderson, D.D.; Stover, P.J. SHMT1 and SHMT2 are functionally redundant in nuclear de novo thymidylate biosynthesis. *PLoS ONE* **2009**, *4*, e5839. [\[CrossRef\]](#)
110. Florio, R.; di Salvo, M.L.; Vivoli, M.; Contestabile, R. Serine hydroxymethyltransferase: A model enzyme for mechanistic, structural, and evolutionary studies. *Biochim. Biophys. Acta* **2011**, *1814*, 1489–1496. [\[CrossRef\]](#)
111. Anderson, D.D.; Woeller, C.F.; Chiang, E.P.; Shane, B.; Stover, P.J. Serine hydroxymethyltransferase anchors de novo thymidylate synthesis pathway to nuclear lamina for DNA synthesis. *J. Biol. Chem.* **2012**, *287*, 7051–7062. [\[CrossRef\]](#) [\[PubMed\]](#)
112. Visentin, M.; Zhao, R.; Goldman, I.D. The antifolates. *Hematol. Oncol. Clin. North. Am.* **2012**, *26*, 629–648. [\[CrossRef\]](#) [\[PubMed\]](#)
113. Sugai, E.; Ashoush, I. Male sterility in the silkworm, *Bombyx mori* L., induced by aminopterin (Lepidoptera: Bombycidae). *App Entomol. Zool.* **1970**, *5*, 202–207. [\[CrossRef\]](#)
114. Dramsi, S.; Bierne, H. Spatial organization of cell wall-anchored proteins at the surface of gram-positive bacteria. *Curr. Top. Microbiol.* **2017**, *404*, 177–201. [\[CrossRef\]](#)
115. Fischetti, V.A. Surface proteins on gram-positive bacteria. *Microbiol. Spectr.* **2019**, *7*, 1–27. [\[CrossRef\]](#)
116. Diaz, E.; Lopez, R.; Garcia, J.L. Chimeric phage-bacterial enzymes: A clue to the modular evolution of genes. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 8125–8129. [\[CrossRef\]](#)

117. Garcia, J.L.; Sanchez-Beato, A.R.; Medrano, F.J.; Lopez, R. Versatility of choline-binding domain. *Microb. Drug Resist.* **1998**, *4*, 25–36. [\[CrossRef\]](#)
118. Calatayud, S.; Garcia-Risco, M.; Capdevila, M.; Canestro, C.; Palacios, O.; Albalat, R. Modular evolution and population variability of *Oikopleura dioica* metallothioneins. *Front. Cell Dev. Biol.* **2021**, *9*, 702688. [\[CrossRef\]](#)
119. Wang, X.; Zhang, Y.; Zhang, R.; Zhang, J. The diversity of pattern recognition receptors (PRRs) involved with insect defense against pathogens. *Curr. Opin. Insect Sci.* **2019**, *33*, 105–110. [\[CrossRef\]](#)
120. Filee, J.; Siguier, P.; Chandler, M. I am what I eat and I eat what I am: Acquisition of bacterial genes by giant viruses. *Trends Genet.* **2007**, *23*, 10–15. [\[CrossRef\]](#) [\[PubMed\]](#)
121. Bertelli, C.; Greub, G. Lateral gene exchanges shape the genomes of amoeba-resisting microorganisms. *Front. Cell Infect. Microbiol.* **2012**, *2*, 110. [\[CrossRef\]](#) [\[PubMed\]](#)
122. Oliveira, G.; Silva, L.; Leao, T.; Mougari, S.; da Fonseca, F.G.; Kroon, E.G.; La Scola, B.; Abrahao, J.S. Tupanvirus-infected amoebas are induced to aggregate with uninfected cells promoting viral dissemination. *Sci. Rep.* **2019**, *9*, 183. [\[CrossRef\]](#) [\[PubMed\]](#)
123. Ruiu, L. *Brevibacillus laterosporus*, a pathogen of invertebrates and a broad-spectrum antimicrobial species. *Insects* **2013**, *4*, 476–492. [\[CrossRef\]](#)
124. McCullers, J.A. The co-pathogenesis of influenza viruses with bacteria in the lung. *Nat. Rev. Microbiol.* **2014**, *12*, 252–262. [\[CrossRef\]](#)
125. Klages-Mundt, N.L.; Li, L. Formation and repair of DNA-protein crosslink damage. *Sci. China Life Sci.* **2017**, *60*, 1065–1076. [\[CrossRef\]](#)
126. Novarina, D.; Desai, R.; Vaisica, J.A.; Ou, J.; Bellaoui, M.; Brown, G.W.; Chang, M. A genome-wide screen for genes affecting spontaneous direct-repeat recombination in *Saccharomyces cerevisiae*. *G3 (Bethesda)* **2020**, *10*, 1853–1867. [\[CrossRef\]](#)
127. Makarenkov, V.; Mazouze, B.; Rabusseau, G.; Legendre, P. Horizontal gene transfer and recombination analysis of SARS-CoV-2 genes helps discover its close relatives and shed light on its origin. *BMC Ecol. Evol.* **2021**, *21*, 5. [\[CrossRef\]](#)
128. Pal, C.; Macia, M.D.; Oliver, A.; Schachar, I.; Buckling, A. Coevolution with viruses drives the evolution of bacterial mutation rates. *Nature* **2007**, *450*, 1079–1081. [\[CrossRef\]](#)
129. M’Gonigle, L.K.; Shen, J.J.; Otto, S.P. Mutating away from your enemies: The evolution of mutation rate in a host-parasite system. *Theor. Popul. Biol.* **2009**, *75*, 301–311. [\[CrossRef\]](#)
130. Streicker, D.G.; Winternitz, J.C.; Satterfield, D.A.; Condori-Condori, R.E.; Broos, A.; Tello, C.; Recuenco, S.; Velasco-Villa, A.; Altizer, S.; Valderrama, W. Host-pathogen evolutionary signatures reveal dynamics and future invasions of vampire bat rabies. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 10926–10931. [\[CrossRef\]](#) [\[PubMed\]](#)
131. Dominguez-Andres, J.; Netea, M.G. Impact of historic migrations and evolutionary processes on human immunity. *Trends Immunol.* **2019**, *40*, 1105–1119. [\[CrossRef\]](#) [\[PubMed\]](#)
132. Goic, B.; Stapleford, K.A.; Frangeul, L.; Doucet, A.J.; Gausson, V.; Blanc, H.; Schemmel-Jofre, N.; Cristofari, G.; Lambrechts, L.; Vignuzzi, M.; et al. Virus-derived DNA drives mosquito vector tolerance to arboviral infection. *Nat. Commun.* **2016**, *7*, 12410. [\[CrossRef\]](#) [\[PubMed\]](#)
133. Pearce, S.L.; Clarke, D.F.; East, P.D.; Elfekih, S.; Gordon, K.H.J.; Jermini, L.S.; McGaughan, A.; Oakeshott, J.G.; Papanicolaou, A.; Perera, O.P.; et al. Genomic innovations, transcriptional plasticity and gene loss underlying the evolution and divergence of two highly polyphagous and invasive *Helicoverpa* pest species. *BMC Biol.* **2017**, *15*, 63. [\[CrossRef\]](#)
134. Ardisson-Araujo, D.M.P.; Sosa-Gómez, D.R.; Melo, F.L.; Bao, S.N. Characterization of the *Helicoverpa zea* single nucleopolyhedrovirus isolated in Brazil during the first old world bollworm (Noctuidae: *Helicoverpa armigera*) nationwide outbreak. *Virus Res.* **2015**, *20*, 1–4. [\[CrossRef\]](#)
135. Zhou, M.; Sun, X.; Sun, X.; Vlak, J.M.; Hu, Z.; van der Werf, W. Horizontal and vertical transmission of wild-type and recombinant *Helicoverpa armigera* single-nucleocapsid nucleopolyhedrovirus. *J. Invertebr. Pathol.* **2005**, *89*, 165–175. [\[CrossRef\]](#)
136. Stork, N.E. How many species of insects and other terrestrial arthropods are there on earth? *Annu. Rev. Entomol.* **2018**, *63*, 31–45. [\[CrossRef\]](#)
137. Anthony, S.J.; Epstein, J.H.; Murray, K.A.; Navarrete-Macias, I.; Zambrana-Torrel, C.M.; Solovyov, A.; Ojeda-Flores, R.; Arrigo, N.C.; Islam, A.; Ali Khan, S.; et al. A strategy to estimate unknown viral diversity in mammals. *mBio* **2013**, *4*, e00598-13. [\[CrossRef\]](#)
138. Junglen, S.; Drosten, C. Virus discovery and recent insights into virus diversity in arthropods. *Curr. Opin. Microbiol.* **2013**, *16*, 507–513. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.