## ABSTRACT

SCHUNKE, JAMES MICHAEL. Transcriptomic Profile of the Mexican Fruit Fly *Anastrepha ludens* under Desiccation Conditions and Structural Analysis of Geminiviral Proteins Rep and REn. (Under the direction of Dr. José Trinidad Ascencio-Ibáñez).

*Anastrepha ludens* is an invasive fly species that causes severe crop loss in climates without seasons to control the fly population. As climate change progresses and southerly latitudes begin to get warmer, the risk of *A. ludens* infestations increases. Some wild populations of *A. ludens* have developed the ability to resist desiccation through an undetermined mechanism; understanding this mechanism could result in treatments to eliminate wild *A. ludens* populations. In this first study we analyzed the whole tissue transcriptomic profiles of laboratory reared *A. ludens* (male and female) and compared them to biological replicates that were desiccation stressed using differential gene expression analysis and pathway analysis. Known and suspected genes important to the desiccation response in *Drosophila* were then selected for future qPCR analysis. Further, we analyzed the heritability of desiccation resistance of laboratory reared *A. ludens*. Our results show numerous similarities but also differences in the GO described for *Drosophila* when compared to *A. ludens*. We also found that the maintenance of desiccation resistance in *A. ludens* disappears after 20 generations and returns to the same susceptibility level of the non-resistant flies. Taken in its totality, the results would seem to suggest that the molecular and physiological profile *A. ludens* is closely related to that of *Drosophila* and provides a unique insight into the polygenic mechanism of desiccation resistance in the Mexican fruit fly.

The goal of the second project is twofold: first, to predictively model the structures of geminiviral proteins Rep and REn via various tools and second, to empirically validate these results using techniques such as Nuclear Magnetic Resonance (NMR) and X-Ray

Crystallography (XRC). Understanding the structures of Rep and REn will provide insight into their functions and therefore potential avenues of exploitation to inhibit geminiviral replication or repress symptoms. The utilization of various different mathematical, statistical, and criteria-based modelling approaches is used to cross-validate the results of each tool using both homology and *de novo*-based modelling techniques. Several versions of Rep (with different lengths from the 5' end) and full length REn were transformed into bacterial strains DH5-α and BL21-DE3 and then purified from BL21-DE3 using a 10X histidine tag, validated via immunoblotting and then analyzed using natural abundance $^{14}N$ NMR. The results of the modelling were marginally successful with each tool providing a unique insight that when viewed in its entirety was useful. The results of the experiments were generally positive with purification and NMR analysis of full length (pNSB1531) Rep achieved while analysis of REn with NMR was not attempted because of the laboratory shutdown due to the COVID-19 pandemic. These results have provided a set of general models and energetic information about Rep and REn that can be used in future molecular simulations as well as a framework for future experimental approaches to stabilizing Rep and REn and determining their structure. In addition, these models provide a basis to model peptide aptamers interaction with virus proteins for virus control.

# Transcriptomic Profile of the Mexican Fruit Fly *Anastrepha ludens* under Desiccation Conditions and Structural Analysis of Geminiviral Proteins Rep and REn.

by
James Michael Schunke

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Master of Science

Biochemistry

Raleigh, North Carolina
2020

APPROVED BY:

| | |
|---|---|
| _____ | _____ |
| Dr. José Trinidad Ascencio-Ibáñez<br>Committee Chair | Dr. Michael B. Goshe |
| _____ | |
| Dr. Robert B. Rose | |

## DEDICATION

To my wife, Dallas: you are amazing and make my life so enjoyable.

## BIOGRAPHY

James Schunke was born October 25th, 1995 in North Carolina. He grew up on a small farm in the foothills of Western North Carolina where they had a herd of about 30 registered Nubian dairy/show goats. He graduated with a B.S. in Biochemistry from NCSU in 2018 where he conducted undergraduate research in the lab of Dr. José Trinidad Ascencio-Ibáñez. After an invitation to complete his master's thesis under the direction of Dr. José Trinidad Ascencio-Ibáñez he began conducting research into the structure of viral proteins Rep and REn as well as a bioinformatic analysis of the desiccation specific transcriptome of the Mexican Fruit Fly (*Anastrepha ludens*).

James married his lovely wife Dallas in 2018 who also graduated from NCSU with a B.S. in Meteorology. They are currently expecting their first child, Henry James Schunke.

## ACKNOWLEDGMENTS

I would like to thank Dr. José Trinidad Ascencio-Ibáñez for mentoring and guiding me throughout my undergraduate career and giving me the opportunity to conduct my master's research under his tutelage. Your instruction has been invaluable to me personally and professionally. Finally, your openness and willingness to talk about literally any subject of interest to me has been a treasure to me intellectually and makes learning from you a delight.

I would also like to thank my fun undergraduate friends – Aanandi, Ahmad, Alex, and Ethan – for your stimulating conversations and willingness to help me around the lab. Alex, your help in completing my experiments in the lab was invaluable; you will go far in life.

Further, many thanks to my graduate student friends who provided friendship and input throughout our classes and research endeavors. Natalia, Mareca, and Monica – Thank you so much.

Natalia, you have been a great friend and support through graduate school, encouraging me to excel and to pursue my dreams; you will go far in life. I am so proud of you.

Many thanks to my mentor and friend Dr. Charles Cook for your guidance and instruction in pursuing my goals and dreams.

Finally, boundless Thanks to my wife, Dallas, for the constant love and support she has shown me throughout my pursuit of my degree. Your support allowed me to complete my master's on time, with continuity, and relatively stress free. Thank you, my beloved.

"*Faithful friends are a study shelter; whoever finds one finds a treasure. Faithful friends are beyond price, no sum can balance their worth*."

Ecclesiasticus 6:14-15 NAB

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

**CHAPTER 1: Transcriptomic Profile of *Anastrepha ludens* under desiccation conditions**

**To be submitted to the Journal of Molecular Ecology**

James M. Schunke[1], José Arredondo[2], Mario A. Arteaga-Vázquez[2], Marco T. Tejeda[2], José T. Ascencio-Ibáñez[1], and Francisco Díaz-Fleischer[3*].

[1]Department of Structural and Molecular Biochemistry, North Carolina State University.

[2]MOSCAFRUT, Subdirección de Desarrollo de Métodos, SADER-IICA, Metapa de Domínguez, Chiapas, México.

[3]INBIOTECA, Universidad Veracruzana, Av. De las Culturas Veracruzanas 101, CP 91090, Xalapa, Veracruz, México.

*Corresponding authors: fradiaz@uv.mx, jtascenc@ncsu.edu

## INTRODUCTION

Desiccation resistance is a phenomenon whereby certain organisms – especially insects – have the capability to tolerate extreme water deprivation (Hoffman & Parsons, 1989). While desiccation resistance at the transcriptomic level has been studied previously in various *Drosophila* species, it has not been characterized in the Mexican fruit fly, *Anastrepha ludens* (Matzkin & Markow, 2009; Sinclair et al., 2007). *A. ludens* is one of the three most invasive *Tetphritidae* whose distribution includes Central and South America into the southern United States (*USDA* 2019). They are unique from other members of *Anastrepha* as they are not only tropical, but subtropical as well (Stone, 1942). Their primary infestations include citrus and mango as well as dozens of other fruits through which they cause enormous financial damage each year (*USDA, 2019)*. Therefore, as climate change intensifies and the southern United States becomes increasingly warmer there is an elevated threat of potentially disastrous infestations by both desiccation resistant *A. ludens* and other invasive *Tetphritidae*'s (Aluja et al., 2014; A. A. Hoffmann, 2010). Thus, an understanding of the mechanisms of desiccation resistance in *A. ludens* could potentially lead to additional treatments to control *A. ludens* populations.

*A. ludens* is quite different from *D. melanogaster* morphologically, physiologically, and reproductively**.** *D. melanogaster* has an approximate lifespan of 50 days at optimal growth conditions, the females are approximately 2.5 mm in length with the males being slightly smaller, has four pairs of chromosomes, and females lay about 400 eggs in their lifetime (Linford et al., 2013). *A. ludens* can live up to one year in the wild, can lay up to 2000 eggs in their lifespan, females are 7-11 mm long with the males being slightly smaller, and have 12 chromosomes (Carey et al., 2005). Assessing the gene level of regulation during desiccation in *A. ludens* may identify additional protein targets.

There are three essential mechanisms that flies employ in response to desiccation stress: 1) retaining water, 2) increasing water intake, or 3) increasing their resistance to water loss (Telonis-Scott et al., 2006). Of these, water retention is the most predominant and most often utilized mechanism of desiccation resistance. There exist three methods of water loss: through the epicuticle, through the gut epithelial cells, and through respiration involving the spiracles (Ferveur, 2005). It seems that the greatest mechanism of water retention is related to an increased volume of water in desiccation resistant flies contained in the hemolymph. Correlatively, it seems that desiccation resistant flies are larger bodied and contain a greater volume of water. This is especially true when making the comparison between male and female desiccation resistant flies.

Previously, it has been noted that there are resistance differences in male vs female when exposed to the same desiccation conditions (Ary A. Hoffmann & Harshman, 1999). Part of this resistance has been attributed to the greater body volume of the females compared to the males. This increased body size leads to a greater volume of hemolymph that contains an overall greater quantity of water compared to males (Bazinet et al., 2010; Folk et al., 2001). In addition,

previous research has identified cuticular hydrocarbon impingement, cAMP dependent signaling protein *desi*, and carbohydrate metabolism regulation involving increased trehalose deposition (H. Chung et al., 2014; Kawano et al., 2010; Stinziano et al., 2015; Thorat et al., 2012). Interestingly, an aquaporin family has been implicated in the desiccation resistance in *D. melanogaster* larvae (Philip et al., 2008). Further, protein ubiquitination has been shown to play an important role in both cuticle regulation as well as the cytoskeleton of the fly (Kang et al., 2016).

Desiccation resistance has been shown in *D. melanogaster* to be highly heritable – up to 60% percent – and possesses the ability to rapidly evolve (Kellermann et al., 2009). However, this does not hold true when taking into consideration geographically constrained fly species. The rearing environment has also been shown to play an important role in the evolution and heritability of various *Drosophila* species. However, little is known concerning the heritability and evolution of the desiccation resistance in *A. ludens*.

In this study we analyzed the whole tissue transcriptomic profiles of laboratory reared *A. ludens* (male and female) and compared them to biological replicates that were desiccation stressed using differential gene expression analysis and pathway analysis. A flowchart of the complete methodological approach is found in **Figure 1.0**. Known and suspected genes important to the desiccation response in *Drosophila* were then selected for future qPCR analysis. Further, we analyzed the heritability and evolution of desiccation resistance of laboratory reared *A. ludens*. Our results show numerous similarities but also differences in the GO described for *Drosophila* when compared to *A. ludens*. We also found that the maintenance of desiccation resistance in *A. ludens* disappears after 20 generations and returns to the same susceptibility level of the non-resistant flies. Taken in its totality, the results would seem to suggest that the

molecular and physiological profile *A. ludens* is closely related to that of *Drosophila* and provides a unique insight into the polygenic mechanism of desiccation resistance in *A. ludens*.

## EXPERIMENTAL PROCEDURES

### Origin of the fly strains (Tejeda et al. 2016)

*Anastrepha ludens* flies were obtained as pupae from a mass reared strain produced at the MOSCAFRUT biofactory at Metapa de Domínguez, Chiapas, México. Approximately 300 million individuals of this fly strain are produced per week. This laboratory strain was reared for more than 100 generations in the biofactory. Thus, flies were well adapted to laboratory adult and larval crowding conditions. For example, females oviposit into an artificial medium and their larvae develop in an artificial diet made of corn cob fractions, corn flour, sodium benzoate, methylparaben, citric acid, guar gum, and purified water (Domínguez et al. 2010). From this original population, we derived 10 experimental populations of 400 flies each: five selected for desiccation resistance (D1–D5) and five unselected, control populations (C1–C5).

### Selection

Once emerged, five groups of four hundred individuals each were placed in Plexiglas cages of 25 × 25 × 25 cm. Males and females were placed in separate cages (200 females and 200 males per replicate) to prevent mating before selection. Desiccation conditions (relative humidity 20–30%) were achieved by supplementing each cage with three plastic containers, each with 50 g of silica gel (Sigma-Aldrich, PubChem Substance ID: 24899758) and covered with a nylon mesh to avoid direct contact of the flies with silica (Tejeda et al. 2014). A data-logger was placed inside a cage selected at random. All cages were then individually sealed with a single layer of self-adhesive plastic film. Relative humidity inside the cage stabilized at between 20–30% within the first 12 hours of exposure. Observations took place every eight hours or less until approximately 12% of the population remained alive. Survivors were transferred to wooden-framed, mesh cages of 30 × 30 × 30 cm where water and food (3:1 mixture of sugar-dry yeast

(ICN Biochemicals, Aurora, OH)) were provided *ad libitum*. Control populations (C1–C5) were handled in the same way but not subjected to desiccation stress and had access to water and food. At least 25 pairs of each of the 10 populations were used to reproduce the following generation. For selected populations (D), a second group was exposed if 25 pairs were not obtained in the first cohort (group of individuals that emerged on the same day). For control populations, pairs were sampled randomly.

**Rearing Conditions**

Selected and control populations were kept under the same rearing conditions. Adults were maintained in wooden framed mesh cages of 30 × 30 × 30 cm at 26 ± 2°C, 80 ± 10% relative humidity and a light: dark 12:12 (h/h) cycle. Eggs were collected with artificial water-filled oviposition devices made with a bottomless Petri dish covered with a black cloth whose surface was overlaid with a thin clear silicone layer (Sista, Henkel Mexicana S.A., Huixquilucan, Estado de México, México), and placed at the top of the wooden cage. Eggs were maintained wet in Petri dishes for four days until hatching. Larvae were transferred to plastic containers with the artificial diet (described above). After 13 days larvae were separated from the food and placed in moistened vermiculite (Strong-lite R, Products Corp. Seneca, Illinois, USA) to allow larvae to pupate. Once emerged and treated with the appropriate selection regime, adult flies were kept in groups of 65-80 pairs per cage with water and food. Total RNA was harvested from the flies by isolation of the alive resistant flies, crushed, and total RNA extracted.

**Fly Generational Resistance**

After emergence flies were separated by sex and by treatment (D, C y DNS), 300 individuals per cage of 25 x 25 x 25 cm. Desiccation conditions (relative humidity 20–30%) were achieved by supplementing each cage with three plastic containers, each with 50 g of silica

gel (Sigma-Aldrich, PubChem Substance ID: 24899758) and covered with a nylon mesh to avoid direct contact of the flies with silica (Tejeda et al. 2014). A data-logger was placed inside a cage selected at random. All cages were then individually sealed with a single layer of self-adhesive plastic film. Relative humidity inside the cage stabilized at between 20–30% within the first 12 h of exposure. Observations took place every 12 hours or less until approximately 25 % of the population remained alive. Survival flies were transferred to wooden-framed, mesh cages of 30 × 30 × 30cm where water and food (3:1 mixture of sugar-dry yeast (ICN Biochemicals, Aurora, OH)) were provided *ad libitum*. Control populations (C1–C5) and DNS (DNS1-DS5) were handled in the same way but not subjected to desiccation stress and had access to water and food. At least 75 pairs of each of the 15 populations were used to reproduce the following generation.

Following our experimental design, data were analyzed using population as the fundamental unit of replication (n = 5). The time, in hours, at which 50% of flies had died during desiccation (mean survival time, ST50) ± standard deviation was calculated for every sex (our data has been analyzed considering sex, population, and generation combination. Both parameters were estimated by the regression of probit-transformed percentage of surviving flies to exposure time. ST50 was averaged between cohorts of the same population and generation.

**RNA Purification and Library Preparation**

RNA-Seq experiments were performed from a cohort of flies of the five families both control and DR populations. Flies were collected after their emergence, over a four-hour period and then were sexed. The sexes were subsequently maintained in separate cages. Groups of 5 females and 5 males (one of each family of the two populations) were exposed to desiccation for 24 hours.

Total RNA from each pool, was extracted with TRI Reagent (Thermo Fisher Scientific, 15596026) according to manufacturer's protocol. Library construction (paired-end 2x100) and sequencing of ~106 - 58 million paired-end 2x100 reads per sample was performed at the CINVESTAV-IPN (Irapuato, Gto. Mex.) using an Illumina HiSeq 4000 platform.

**Differential Gene Expression Analysis**

The total RNA was sequenced using Illumina paired-end sequencing with 100 bp length. Two runs were made in the HiSeq sequencing machine. The sequences from each run for each sample were then combined to generate the total fragments for each sample. The output was saved as a Fastq file.

The Fastq files were then analyzed using the program FastQC v0.11.3 to subject the data to quality control (QC) checks (Andrews, 2010). The data passed the QC check with a Phred score greater than 28 with normal decay at the final bases of the sequences. It was not necessary to eliminate and remove adapter sequences as commercially available adapter sequences were used and the low-quality sequences were removed during the assembly process. The fragments were then assembled using Trinity v2.0.6 using the "—trimmomatic" parameter to remove the low-quality sequences (Grabherr et al., 2011). The output resulted in 16 fasta files. CD-HIT-EST was used to combine the sequences with more than 90% identity and more than 90% coverage.

The final transcriptome contains 261,536 sequences with a size average of 964.1 bp and a maximum sequence of 27,495 bp. With the assembled transcriptome, the table of counts generated when the fragments were aligned to the transcriptome. Alignment to the transcriptome was done using the program Bowtie v1.1.1 with the parameter "-a" to report all hits "-best –strata" to report only the best hits, and "-X 800" to make the maximum size of the insert 800 bp

(Langmead et al., 2009). The program Express v1.5.1 was used to adjust the table of counts generated in Trinity (Forster et al., 2013).

To identify the differentially expressed genes (DEG) Bioconductor packages EdgeR and Limma were employed using R and RStudio (Allaire, 2015; Robinson et al., 2009; Smyth, 2005). Comparisons were made between each of the 16 samples using a fold-change of 2X and a False Discovery Rate (FDR) of 0.01. Tables were made for each comparison with the number of genes UP and DOWN regulated. The output was saved as .fasta files. The DEG were then annotated by comparing to the *Drosophila melanogaster* sequences using "org.Dm.eg.db" where the best hit was used for identification as well as later use in the GO analysis.

Gene Ontology analysis was conducted using the ClusterProfiler R package in Bioconductor for each of the 12 generated comparisons (Gene & Consortium, 2000; Yu, 2018; Yu et al., 2012). FDR cutoffs of 0.05 were used after graphing FDR vs. Gene Name to find the inflection point that would glean the most data with the least amount of noise. Bonferroni corrected p-value cutoffs of 0.05 were used as is standard practice in transcriptomic data analysis. GO profiles for each comparison that were both UP and DOWN regulated were generated for Biological Process (BP), Cellular Component (CC), and Molecular Function (MF). In addition, Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis was conducted using ClusterProfiler for each comparison to validate the results obtained in the GO analysis. Further, iDEP.90 was utilized to identify the DEG using DESeq2 and Limma and both a GO and KEGG analysis was conducted to validate the results using EdgeR and Limma with the same cutoff criteria as used previously in the GO analysis (Ge et al., 2018; Gene & Consortium, 2000; Ogata et al., 1999).

**Pathway Analysis**

Of the three GO categories (Biological Process, Molecular Function, and Cellular Component) we choose the Molecular Function (MF) as it most directly relates to the functions of the proteins involved in the desiccation stress response. Three pathways upregulated in the GO term MF were identified as being the most significantly impinged upon by the desiccation response: serine hydrolase, lipase, and active transmembrane transporter. Three pathways were found to be downregulated in MF: chitin binding, cuticle structural construction, and actin binding. To find the gene overlap, gene lists from each sample that had a pathway of interest was compared using Venn diagrams using the program VennDiagram and the male vs female comparison and the stressed vs non-stressed comparisons made in previous analyses above (H. Chen & Boutros, 2011). Then, each area of overlap that was of interest was saved as a file and then compared to the Flybase for identification as to gene type and function. These genes were then curated based on p-value, FDR, and potential impact on the desiccation response to select candidates for further validation by qPCR.

**RESULTS**

The flies were successfully grown with no complications. Results indicated that *A. ludens* desiccation resistant stock lost their resistance after 20 generations. This loss of resistance occurred in a nearly linear fashion as shown in **Figure 1.1**. The control line represents *A. ludens* stock that are susceptible to desiccation. The topline represents *A. ludens* stock that are resistant to desiccation. Both controls remained consistent throughout the entire analysis indicating that they presented a good baseline from which to compare the self-selecting line.

The differential expression analysis (DEG) was successful with 12 comparisons being made and numerous genes in each comparison differentially expressed. **Table 1.1** has a complete list of the comparisons made and the number of differentially expressed transcripts UP or DOWN regulated for each comparison and the ID key is found in **Table 1.0.** The GO pathway analysis yielded numerous pathways that were overrepresented in each of the three categories: Biological Process, Molecular Function, and Cellular Component. The KEGG analysis for each comparison yielded similar but far fewer pathways - with many of the comparisons having no KEGG enriched pathway. Likely this is due to the functionally different purpose of the KEGG software and the methodology by which it assigns genes to pathways by attempting to link the genes to protein products. **Table 1.2** lists the three most overrepresented GO and KEGG pathways for each of the GO categories.

In addition, barplots, network plots, and dot plots for each GO term based on pathway p-value significance and number of genes in the pathway were generated to visualize pathway significance and potential importance for further analysis; to reproduce these figures see code stored in GitHub and can be accessed at https://github.com/jdmsncsu20/Anastrepha-ludens. To focus attention on and select those pathways that have the greatest potential, each of the bar and

dot plots were examined and the most common pathways were then selected in the Molecular Function category as it contains the most useful data in understanding the molecular response to desiccation stress in *A. ludens*. Attempting to make Venn Diagrams with the full set of DEG genes both UP and DOWN regulated presented an analytical nightmare as there were simply thousands of genes of overlap. Thus, by selecting the pathways of interest and then generating Venn Diagrams between each of the comparisons showing the pathways as overrepresented the number of genes become simply a few dozen to a hundred. These genes in each of the areas of overlap were then analyzed for those that may be unique to *A. ludens* or ubiquitous with *D. melanogaster*.

The selected MF pathways shown to be differentially impinged and the genes found to be significantly differentially regulated in our analysis were consistent with previous data identified in *Drosophila*. **Table 1.3** depicts ten genes identified previously as important in the desiccation response in *Drosophila* and those same genes we have identified in *A. ludens* as being differentially expressed in the stressed samples. **Table 1.4** contains a list of the most common pathways shown to be differentially expressed in the GO molecular function (MF) category.

Significant overlap was found between each of the Venn Diagrams made for males only, females only; stressed vs. non-stressed in both intrasex (e.g. Male vs. Male) and intersex (Female vs. Male) comparisons for each of the GO categories and their respective UP or DOWN regulation. Of special interest to us are the Venn Diagrams of the Molecular Function. **Figure 1.2** contains the Venn Diagram for the Male pathway 'active transmembrane transporter' from each of the male comparisons. This pathway is important as it plays an integral role in understanding and regulating the flow of nutrients and metabolites throughout the cell. Further, a pathway for 'neurotransmitter transporter activity' was found to be overrepresented only

between the stressed sensitive female and the stressed tolerant female and between the stressed sensitive male and the stressed tolerant male. No other comparison made yielded this pathway. This pathway is unique to *A. ludens* and not found in *D. melanogaster* and thus was selected as a potential source of desiccation resistance. For the remainder of the Venn Diagrams for each of the other pathways UP and DOWN for the males and for the female and intersex comparisons see code in GitHub at https://github.com/jdmsncsu20/Anastrepha-ludens. In some cases, identical terms were found to be UP and DOWN regulated though containing different genes. These disparities are further explored in the Discussion below.

Venn Diagrams were made to determine the gene overlap between each of the DEG comparisons. Comparisons were made between females only, males only, between non-stressed females and non-stressed males, and stressed females and stressed males. In addition, for the pathways of interest Venn Diagrams for each of the comparisons were generated. For example, **Figure 1.2** depicts the Venn Diagram for the genes shown to be upregulated and downregulated in the GO MF male pathway – active transmembrane transporter. Venn Diagrams like this were generated for each of the pathways of interest: active transmembrane transporter, cytoskeletal protein binding, and neurotransmitter transporter activity. Significant gene overlap was found in each Venn Diagram. Genes involved in each overlap were identified via FlyBase using CG numbers from the genes. Genes of interest were then selected for additional analysis using qPCR in the future and are found in **Table 1.3**.

## DISCUSSION

The purpose of this study was to identify the main players involved in desiccation resistance in the Mexican fly. To achieve this, we compared the *Drosophila* transcriptome to the *A. ludens* transcriptome with respect to desiccation resistance, to characterize the desiccation response in *A. ludens* at the transcriptome level, and to identify and validate, if applicable, novel genes differentially expressed via qPCR. Our results indicate that there is a high correlation of the *A. ludens* transcriptome to the *Drosophila* transcriptome – the first time this has been performed. In addition, the pathways identified as important in *Drosophila* literature were shown to be enriched in our analysis as well. Further, many of the genes shown to be key in the *Drosophila* desiccation response were shown to be present in *A. ludens* and a list is given in **Table 1.3**. These include *Desi* and *Frost*; two genes known to be involved specifically involved in the desiccation response in Drosophila. In addition, we have shown that *A. ludens* desiccation resistance is lost after 20 generations as shown in **Figure 1.1**; potential reasons are given below. Thus, our results provide a framework for understanding the desiccation stress response in the most invasive fruit fly in the world and an insight into further areas of research that could lead to mechanisms of control or eradication with species specificity.

To understand the desiccation resistance pattern of *A. ludens* we grew several biological replicates of *A. ludens* and then after exposing some to desiccation stress and others in their ideal environment we harvested the whole fly RNA. We then conducted RNA sequencing and DEG analysis to determine the genes impinged upon by the desiccation response in *A. ludens* and compared this response to that found in *Drosophila*. Our results indicated that many of the genes found to be significant in the desiccation response in *Drosophila* were also significant in *A. ludens* as shown in **Table 1.3**. In addition, we conducted a pathway analysis using GO and

KEGG and determined similar and overlapping results. This is true of both the EdgeR and Limma determined DEG and DESeq2 and Limma identified DEG.

The Gene Ontology Category Molecular Function (MF) yielded the most explicitly useful information concerning the *A. ludens* desiccation response. Three MF UP regulated pathways of interest that are consistent with previous literature are: serine hydrolase, lipase, and active transmembrane transporter. Three MF DOWN regulated pathways of interest are: chitin binding, cuticle structural construction, and actin binding. However, these data must be taken in context of other GO terms such as Biological Process (BP) and Cellular Component (CC). What we found interesting is that taken in sum these pathways provide a clearer insight into what is intuitive in the desiccation response and what is novel. For example, one would expect to see an upregulation in processes increasing access to stored reserves such as fat which we see in the MF pathways lipase and serine hydrolase, but we also see this theme reflected in the upregulated BP pathways: lipid metabolic process, cellular lipid metabolic process, and lipid catabolic process. Significantly, the most common MF downregulated pathways are related to the construction of fly structural features such as chitin binding and cuticle structural construction that are metabolically expensive to maintain (e.g. requiring phosphorylated energy sources such as ATP and UDP). This theme is reflected in the UP regulated GO category Cellular Component (CC) where the membrane coat is found to be enriched. This may suggest that the fly's metabolism is decreasing water loss through the membrane and that the upregulation of the lipid metabolism may serve both energetically AND to prevent water loss through the membranes and ultimately through the cuticle. This is important as the fly lipid metabolism is done to release the water molecules bound to the lipids.

Perhaps one of the more interesting outcomes of this analysis is that after 20 generations the desiccation resistant flies lost their resistance, a phenomenon not yet demonstrated in Drosophila. There could be several reasons for this: desiccation resistance is polygenic and that self-selecting and inbreeding the flies could lead to a dilution of the gene pool, continual stress on the flies may actually inhibit their resistance capabilities, there may potentially be bias in selecting the flies that survive the desiccation resistance for further breeding, hormonal regulation, and there may be an epigenetic component. While it could be a combination of many of these factors, epigenetics has been shown previously in *Drosophila* to play an important role in regulating the response to desiccation by the addition or removal of histone marks on chromatin in regions thought to be important in modulating the desiccation response (Sharma et al., 2017). Further, enrichment of genes thought to be putative targets of the Polycomb Group proteins were not affected whereas genes that were thought to be putative targets of the Trithorax group proteins had an expected impact on transcript levels (Sharma et al., 2017). It is possible that enhancer regions and enhancer marks and inheritance of histone marks play a key role in the desiccation resistance response. Enhancer marks serve as a recruitment nexus for transcription factors that can modulate and attenuate gene expression at the transcriptional level (Ong & Corces, 2011).

Further, research from our group – manuscript submitted – as well as others has indicated that hormonal regulation may play a key role in the regulation of the desiccation response in fruit flies (Kahsai et al., 2010; Terhzaz et al., 2018). Our group has unpublished data that suggests hormonal control using Methoprene vs. metformin in *A. ludens* plays a key role in the longevity and fecundity of the flies with significant differences noted between males and females. By modulating behavior via hormones and affecting fecundity, the resource availability to

ameliorate the effects of desiccation stress are attenuated and the longevity of the flies is shortened. Our results from this analysis indicate that neuropeptide pathway involvement is indicated as shown in **Table 1.5**. The extent to which hormones play cannot be determined directly from this analysis, but it provides a framework for understanding the systems level involvement in regulating the desiccation response in *A. ludens*.

Our results from the DEG pathway analysis indicate that the 'neurotransmitter transporter activity' pathway was present only in the comparisons between the stressed sensitive female and the stressed tolerant female and between the stressed sensitive male and the stressed tolerant male. No other comparison made yielded this pathway. Further, **Table 1.3** lists the genes found in each pathway and the significant overlap between comparisons. Interestingly, we found that both comparisons included the upregulation of gene SLC22A which codes for a Solute Carrier protein. SLC's are integral membrane proteins that transport carnitine into the cell via function as a symporter; this is borne out by the data which shows an upregulation of the CarT gene a protein that transports carcinine to photoreceptors and belongs to the SLC22A family of genes. **Table 1.5** also confirms that the gene that codes for tachykinin protein (Tk) are present and upregulated; Tk has been implicated previously in the regulation of intestinal lipid metabolism and when knocked out, mutants flies displayed increased susceptibility to desiccation conditions (Gai et al., 2016; Kahsai et al., 2010). The importance of the SLC22A family has not been fully elucidated in the desiccation response of *Drosophila* and certainly not *A. ludens*.

In conclusion, our results indicate that *A. ludens* though very dissimilar to *Drosophila* morphologically, reproductively, and geographically, shares a similar desiccation specific transcriptomic profile with *Drosophila*. Our data provides a more complete picture of the mechanism of desiccation resistance in *A. ludens* and provides a context for the epigenetic,

hormonal, and polygenic underpinnings of the desiccation resistance response. Our results further indicate that there is currently an unexplained mechanism whereby the desiccation resistant flies lose their resistance after 20 generations.

**Future Work**

Future research can explore a systems level understanding of the desiccation response in *A. ludens*, the epigenetic contributions to the desiccation response, the hormonal control of the desiccation response in *A. ludens*, and why the flies lose their resistance after several generations. Understanding the mechanisms of desiccation resistance of invasive fly species can allow for the development of targeted treatments that have marginal impact on the surrounding ecosystem. These targeted treatments could include hormone analogues, protein inhibitors specific to *A. ludens*, and a greater understanding of the influences of the behavior of *A. ludens* in response to desiccation stress that could be exploited.

A potential experimental approach could involve a time course experiment where total RNA is harvested from fly tissue for each of the selected lines: control desiccation resistant, control desiccation sensitive, and inbred line that loses its resistance after approximately 20 generations. Analysis could occur at both the gene regulation level and the epigenetic mark level. Specific genes that are known to be highly differentially expressed in response to desiccation stress such as those listed depicted in **Table 1.3** can be monitored for attenuation of expression that might account for the loss of resistance in continually desiccation stressed populations. Further, epigenetic markers such as silencing and activating histone marks H3K27me3 and H3K4me3, respectively, which are known to be involved in the desiccation response can be monitored for changes throughout the time course experiment. These marks are selected as of potential interest as they have been shown to be impinged during desiccation events in *Drosophila melanogaster*. This level of analysis would provide a key insight into the genetic and epigenetic landscape of *A. ludens* and how it changes in response to desiccation stress and thus mechanisms of regulation that could be exploited to control *A. ludens* invasive populations.

## *Anastrepha ludens* Prospectus Outline for Future Work

The goal of this prospectus is to examine the extent to which epigenetic influences impact desiccation resistance in *A. ludens* and correlate this analysis with the genomic profile of *Drosophila melanogaster*. As described in Chapter 1 there is a significant portion of the desiccation response in *D. melanogaster* that is regulated through mostly poorly understood epigenetic mechanisms. This project was initially undertaken in Dr. Hanley-Bowdoin's class at NCSU – BCH 761. This prospectus was initially written in NSF format also including a budget. It has been slightly updated from its original form for clarity and relevant research annotations. While the research I have articulated in Chapter 1 fulfills many of the purposes of Aim 1 in the following prospectus, there remains additional validation to *D. melanogaster* that could solidify the results from Aim 2. To this end, I have been collaborating with Dr. Francisco Diaz-Fleischer in Mexico who is an expert in fly behavioral ecology and Dr. Jose Trino Ascencio-Ibanez, my PI, who is an expert in molecular biology and transcriptome analysis.

**Limitations:**

Limitations of this analysis include the fact that RNA was total RNA and not tissue or functionally specific. Further, because the RNA from each comparison was pooled and not individualized this could lead to a dilution of the desiccation effect on transcript levels.

**Author Contributions:**

J. A. and M. T. selected the flies and then carried out the experiments. M. T. extracted the RNA and participated in the design of the experiment. J. S. conducted the bioinformatic analysis, pathway analysis, gene selection for future qPCR analysis, and participated in the experimental design. F. D-F. participated in the experimental design and conceived of the initial experiment. J. A-I participated in the experimental design

## REFERENCES


Ach, R. A., Durfee, T., Miller, A. B., Taranto, P., Hanley-Bowdoin, L., Zambryski, P. C., & Gruissem, W. (1997). RRB1 and RRB2 encode maize retinoblastoma-related proteins that interact with a plant D-type cyclin and geminivirus replication protein. *Molecular and Cellular Biology*. https://doi.org/10.1128/mcb.17.9.5077

Allaire, J. J. (2015). RStudio: Integrated development environment for R. *The Journal of Wildlife Management*. https://doi.org/10.1002/jwmg.232

Aluja, M., Birke, A., Ceymann, M., Guillén, L., Arrigoni, E., Baumgartner, D., Pascacio-Villafán, C., & Samietz, J. (2014). Agroecosystem resilience to an invasive insect species that could expand its geographical range in response to global climate change. *Agriculture, Ecosystems and Environment*. https://doi.org/10.1016/j.agee.2014.01.017

Andrews, S. (2010). FastQC. *Babraham Bioinformatics*. https://doi.org/citeulike-article-id:11583827

Argüello-Astorga, G. R., Guevara-González, R. G., Herrera-Estrella, L. R., & Rivera-Bustamante, R. F. (1994). Geminivirus Replication Origins Have a Group-Specific Organization of Iterative Elements: A Model for Replication. *Virology*. https://doi.org/10.1006/viro.1994.1458

Bazinet, A. L., Marshall, K. E., MacMillan, H. A., Williams, C. M., & Sinclair, B. J. (2010). Rapid changes in desiccation resistance in Drosophila melanogaster are facilitated by changes in cuticular permeability. *Journal of Insect Physiology*. https://doi.org/10.1016/j.jinsphys.2010.09.002

Berg, J., Tymoczko, J., & Stryer, L. (2002). Biochemistry, 5th edition. In *Biochemistry*.

Bhagavan, N. V., & Ha, C.-E. (2015). DNA Replication, Repair, and Mutagenesis. In *Essentials of Medical Biochemistry*. https://doi.org/10.1016/b978-0-12-416687-5.00022-1

Bhattacharyya, D., & Chakraborty, S. (2018). Chloroplast: the Trojan horse in plant–virus interaction. In *Molecular Plant Pathology*. https://doi.org/10.1111/mpp.12533

Bhattacharyya, D., Gnanasekaran, P., Kumar, R. K., Kushwaha, N. K., Sharma, V. K., Yusuf, M. A., & Chakraborty, S. (2015). A geminivirus betasatellite damages the structural and functional integrity of chloroplasts leading to symptom formation and inhibition of photosynthesis. *Journal of Experimental Botany*. https://doi.org/10.1093/jxb/erv299

Borah, B. K., Zarreen, F., Baruah, G., & Dasgupta, I. (2016). Insights into the control of geminiviral promoters. In *Virology*. https://doi.org/10.1016/j.virol.2016.04.033

Bottcher, B., Unseld, S., Ceulemans, H., Russell, R. B., & Jeske, H. (2004). Geminate Structures of African Cassava Mosaic Virus. *Journal of Virology*. https://doi.org/10.1128/jvi.78.13.6758-6765.2004

Boulton, M. I., Pallaghy, C. K., Chatani, M., MacFarlane, S., & Davies, J. W. (1993). Replication of maize streak virus mutants in maize protoplasts: Evidence for a movement protein. *Virology*. https://doi.org/10.1006/viro.1993.1010

Briddon, R. W., Bedford, I. D., Tsai, J. H., & Markham, P. G. (1996). Analysis of the nucleotide sequence of the treehopper-transmitted geminivirus, tomato pseudo-curly top virus, suggests a recombinant origin. *Virology*. https://doi.org/10.1006/viro.1996.0264

Briddon, R. W., Patil, B. L., Bagewadi, B., Nawaz-Ul-Rehman, M. S., & Fauquet, C. M. (2010). Distinct evolutionary histories of the DNA-A and DNA-B components of bipartite begomoviruses. *BMC Evolutionary Biology*. https://doi.org/10.1186/1471-2148-10-97

Buchan, D. W. A., & Jones, D. T. (2019). The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkz297

Cantú-Iris, M., Pastor-Palacios, G., Mauricio-Castillo, J. A., Bañuelos-Hernández, B., Avalos-Calleros, J. A., Juárez-Reyes, A., Rivera-Bustamante, R., & Argüello-Astorga, G. R. (2019). Analysis of a new begomovirus unveils a composite element conserved in the CP gene promoters of several Geminiviridae genera: Clues to comprehend the complex regulation of late genes. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0210485

Carey, J. R., Liedo, P., Müller, H. G., Wang, J. L., Senturk, D., & Harshman, L. (2005). Biodemography of a long-lived tephritid: Reproduction and longevity in a large cohort of female Mexican fruit flies, *Anastrepha ludens*. *Experimental Gerontology*. https://doi.org/10.1016/j.exger.2005.07.013

Castellano, M. M., Sanz-Burgos, A. P., & Gutiérrez, C. (1999). Initiation of DNA replication in a eukaryotic rolling-circle replicon: Identification of multiple DNA-protein complexes at the geminivirus origin. *Journal of Molecular Biology*. https://doi.org/10.1006/jmbi.1999.2916

Castillo, A. G., Collinet, D., Deret, S., Kashoggi, A., & Bejarano, E. R. (2003). Dual interaction of plant PCNA with geminivirus replication accessory protein (REn) and viral replication protein (Rep). *Virology*. https://doi.org/10.1016/S0042-6822(03)00234-4

Chen, H., & Boutros, P. C. (2011). VennDiagram: A package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics*. https://doi.org/10.1186/1471-2105-12-35

Chen, L. F., & Gilbertson, R. L. (2009). Curtovirus-cucurbit interaction: Acquisition host plays a role in leafhopper transmission in a host-dependent manner. *Phytopathology*. https://doi.org/10.1094/PHYTO-99-1-0101

Chowda-Reddy, R. V., Achenjang, F., Felton, C., Etarock, M. T., Anangfac, M. T., Nugent, P., & Fondong, V. N. (2008). Role of a geminivirus AV2 protein putative protein kinase C motif on subcellular localization and pathogenicity. *Virus Research*. https://doi.org/10.1016/j.virusres.2008.02.014

Chung, H., Loehlin, D. W., Dufour, H. D., Vaccarro, K., Millar, J. G., & Carroll, S. B. (2014). A single gene affects both ecological divergence and mate choice in Drosophila. *Science*. https://doi.org/10.1126/science.1249998

Chung, H. Y., & Sunter, G. (2014). Interaction between the transcription factor AtTIFY4B and begomovirus AL2 protein impacts pathogenicity. *Plant Molecular Biology*. https://doi.org/10.1007/s11103-014-0222-9

Clemson, A. S., Sgrò, C. M., & Telonis-Scott, M. (2018). Transcriptional profiles of plasticity

for desiccation stress in Drosophila. *Comparative Biochemistry and Physiology Part - B: Biochemistry and Molecular Biology*. https://doi.org/10.1016/j.cbpb.2017.11.003

Collin, S., Fernández-Lobato, M., Gooding, P. S., Mullineaux, P. M., & Fenoll, C. (1996). The two nonstructural proteins from wheat dwarf virus involved in viral gene expression and replication are retinoblastoma-binding proteins. *Virology*. https://doi.org/10.1006/viro.1996.0256

Davies, J. W., Stanley, J., Donson, J., Mullineaux, P. M., & Boulton, M. I. (1987). Structure and replication of geminivirus genomes. *Journal of Cell Science. Supplement*. https://doi.org/10.1242/jcs.1987.supplement_7.7

De Jager, S. M., & Murray, J. A. H. (1999). Retinoblastoma proteins in plants. *Plant Molecular Biology*. https://doi.org/10.1023/A:1006398232003

Desvoyes, B., De Mendoza, A., Ruiz-Trillo, I., & Gutierrez, C. (2014). Novel roles of plant RETINOBLASTOMA-RELATED (RBR) protein in cell proliferation and asymmetric cell division. In *Journal of Experimental Botany*. https://doi.org/10.1093/jxb/ert411

Dong, X., van Wezel, R., Stanley, J., & Hong, Y. (2003). Functional Characterization of the Nuclear Localization Signal for a Suppressor of Posttranscriptional Gene Silencing. *Journal of Virology*. https://doi.org/10.1128/jvi.77.12.7026-7033.2003

Donson, J., Morris-Krsinich, B. A., Mullineaux, P. M., Boulton, M. I., & Davies, J. W. (1984). A putative primer for second-strand DNA synthesis of maize streak virus is virion-associated. *The EMBO Journal*. https://doi.org/10.1002/j.1460-2075.1984.tb02260.x

Durmuş, S., & Ülgen, K. (2017). Comparative interactomics for virus–human protein–protein interactions: DNA viruses versus RNA viruses. *FEBS Open Bio*. https://doi.org/10.1002/2211-5463.12167

Elmer, J. S., Brand, L., Sunter, G., Gardiner, W. E., Bisaro, D. M., & Rogers, S. G. (1988). Genetic analysis of the tomato golden mosaic virus II. The product of the AL1 coding sequence is required for replication. *Nucleic Acids Research*. https://doi.org/10.1093/nar/16.14.7043

Fauquet, C. M., Briddon, R. W., Brown, J. K., Moriones, E., Stanley, J., Zerbini, M., & Zhou, X. (2008). Geminivirus strain demarcation and nomenclature. *Archives of Virology*. https://doi.org/10.1007/s00705-008-0037-6

Ferveur, J. F. (2005). Cuticular hydrocarbons: Their evolution and roles in Drosophila pheromonal communication. In *Behavior Genetics*. https://doi.org/10.1007/s10519-005-3220-5

Folk, D. G., Han, C., & Bradley, T. J. (2001). Water acquisition and partitioning in Drosophila melanogaster: Effects of selection for desiccation-resistance. *Journal of Experimental Biology*.

Fondong, V. N., Reddy, R. V. C., Lu, C., Hankoua, B., Felton, C., Czymmek, K., & Achenjang, F. (2007). The consensus N-myristoylation motif of a geminivirus AC4 protein is required for membrane binding and pathogenicity. *Molecular Plant-Microbe Interactions*. https://doi.org/10.1094/MPMI-20-4-0380

Fontes, E. P. B., Luckow, V. A., & Hanley-Bowdoin, L. (1992). A geminivirus replication protein is a sequence-specific DNA binding protein. *Plant Cell*. https://doi.org/10.1105/tpc.4.5.597

Formosa, T., & Alberts, B. M. (1986). DNA synthesis dependent on genetic recombination: Characterization of a reaction catalyzed by purified bacteriophage T4 proteins. *Cell*. https://doi.org/10.1016/0092-8674(86)90522-2

Forster, S. C., Finkel, A. M., Gould, J. A., & Hertzog, P. J. (2013). RNA-eXpress annotates novel transcript features in RNA-seq data. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btt034

*FRUIT FLY EXCLUSION AND DETECTION STRATEGIC PLAN FY 2019-2023*. (2019). https://www.aphis.usda.gov/plant_health/plant_pest_info/fruit_flies/downloads/feed-strategic-plan-en.pdf

Gai, Y., Liu, Z., Cervantes-Sandoval, I., & Davis, R. L. (2016). Drosophila SLC22A Transporter Is a Memory Suppressor Gene that Influences Cholinergic Neurotransmission to the Mushroom Bodies. *Neuron*. https://doi.org/10.1016/j.neuron.2016.03.017

Ge, S. X., Son, E. W., & Yao, R. (2018). iDEP: An integrated web application for differential expression and pathway analysis of RNA-Seq data. *BMC Bioinformatics*. https://doi.org/10.1186/s12859-018-2486-6

Gene, T., & Consortium, O. (2000). Gene Ontology : tool for the. *Gene Expression*. https://doi.org/10.1038/75556

Glick, E., Zrachya, A., Levy, Y., Mett, A., Gidoni, D., Belausov, E., Citovsky, V., & Gafni, Y. (2008). Interaction with host SGS3 is required for suppression of RNA silencing by tomato yellow leaf curl virus V2 protein. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.0709036105

Gnanasekaran, P., Ponnusamy, K., & Chakraborty, S. (2019). A geminivirus betasatellite encoded βC1 protein interacts with PsbP and subverts PsbP-mediated antiviral defence in plants. *Molecular Plant Pathology*. https://doi.org/10.1111/mpp.12804

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., Di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*. https://doi.org/10.1038/nbt.1883

Gronenborn, B. (2004). Nanoviruses: Genome organisation and protein function. *Veterinary Microbiology*. https://doi.org/10.1016/j.vetmic.2003.10.015

Gutierrez, C. (1999). Geminivirus DNA replication. In *Cellular and Molecular Life Sciences*. https://doi.org/10.1007/s000180050433

Gutierrez, Crisanto, Ramirez-Parra, E., Castellano, M. M., Sanz-Burgos, A. P., Luque, A., & Missich, R. (2004). Geminivirus DNA replication and cell cycle interactions. *Veterinary Microbiology*. https://doi.org/10.1016/j.vetmic.2003.10.012

Hallan, V., & Gafni, Y. (2001). Tomato yellow leaf curl virus (TYLCV) capsid protein (CP) subunit interactions: Implications for viral assembly. *Archives of Virology*. https://doi.org/10.1007/s007050170062

Hanley-Bowdoin, L., Settlage, S. B., Orozco, B. M., Nagar, S., & Robertson, D. (2000). Geminiviruses: Models for plant DNA replication, transcription, and cell cycle regulation. In *Critical Reviews in Biochemistry and Molecular Biology*. https://doi.org/10.1080/07352689991309162

Hanley-Bowdoin, L., Settlage, S. B., & Robertson, D. (2004). Reprogramming plant gene expression: A prerequisite to geminivirus DNA replication. In *Molecular Plant Pathology*. https://doi.org/10.1111/j.1364-3703.2004.00214.x

Hartitz, M. D., Sunter, G., & Bisaro, D. M. (1999). The tomato golden mosaic virus transactivator (TrAP) is a single- stranded DNA and zinc-binding phosphoprotein with an acidic activation domain. *Virology*. https://doi.org/10.1006/viro.1999.9925

Höfer, P., Bedford, I. D., Markham, P. G., Jeske, H., & Frischmuth, T. (1997). Coat protein gene replacement results in whitefly transmission of an insect nontransmissible geminivirus isolate. *Virology*. https://doi.org/10.1006/viro.1997.8751

Hoffmann, A. A. (2010). Physiological climatic limits in Drosophila: Patterns and implications. *Journal of Experimental Biology*. https://doi.org/10.1242/jeb.037630

HOFFMANN, A. A., & PARSONS, P. A. (1989). An integrated approach to environmental stress tolerance and life-history variation: desiccation tolerance in Drosophila. *Biological Journal of the Linnean Society*. https://doi.org/10.1111/j.1095-8312.1989.tb02098.x

Hoffmann, Ary A., & Harshman, L. G. (1999). Desiccation and starvation resistance in Drosophila: Patterns of variation at the species, population and intrapopulation levels. In *Heredity*. https://doi.org/10.1046/j.1365-2540.1999.00649.x

Hu, T., Huang, C., He, Y., Castillo-Gonzalez, C., Gui, X., Wang, Y., Zhang, X., & Zhou, X. (2019). βc1 protein encoded in geminivirus satellite concertedly targets MKK2 and MPK4 to counter host defense. *PLoS Pathogens*. https://doi.org/10.1371/journal.ppat.1007728

Jeske, H., Lütgemeier, M., & Preiß, W. (2001). DNA forms indicate rolling circle and recombination-dependent replication of Abutilon mosaic virus. *EMBO Journal*. https://doi.org/10.1093/emboj/20.21.6158

Joan L. Slonczewski, John W. Foster, K. M. (2017). Microbiology: An Evolving Science (2nd Edition). *IEEE Transactions on Wireless Communications*. https://doi.org/10.1109/twc.2017.2771181

Jose, J., & Usha, R. (2003). Bhendi yellow vein mosaic disease in india is caused by association of a DNA β satellite with a begomovirus. *Virology*. https://doi.org/10.1006/viro.2002.1768

Jupin, I., De Kouchkovsky, F., Jouanneau, F., & Gronenborn, B. (1994). Movement of tomato yellow leaf curl geminivirus (TYLCV): Involvement of the protein encoded by ORF C4. *Virology*. https://doi.org/10.1006/viro.1994.1512

Kahsai, L., Kapan, N., Dircksen, H., Winther, Å. M. E., & Nässel, D. R. (2010). Metabolic stress

responses in Drosophila are modulated by brain neurosecretory cells that produce multiple neuropeptides. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0011480

Kang, L., Aggarwal, D. D., Rashkovetsky, E., Korol, A. B., & Michalak, P. (2016). Rapid genomic changes in Drosophila melanogaster adapting to desiccation stress in an experimental evolution system. *BMC Genomics*. https://doi.org/10.1186/s12864-016-2556-y

Kawano, T., Shimoda, M., Matsumoto, H., Ryuda, M., Tsuzuki, S., & Hayakawa, Y. (2010). Identification of a gene, Desiccate, contributing to desiccation resistance in Drosophila melanogaster. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M110.168864

Kellermann, V., Van Heerwaarden, B., Sgrò, C. M., & Hoffmann, A. A. (2009). Fundamental evolutionary limits in ecological traits drive drosophila species distributions. *Science*. https://doi.org/10.1126/science.1175443

Kelman, Z. (1997). PCNA: Structure, functions and interactions. In *Oncogene*. https://doi.org/10.1038/sj.onc.1200886

Kong, L.-J. (2000). A geminivirus replication protein interacts with the retinoblastoma protein through a novel domain to determine symptoms and tissue specificity of infection in plants. *The EMBO Journal*. https://doi.org/10.1093/emboj/19.13.3485

Kong, L. J., & Hanley-Bowdoin, L. (2002). A geminivirus replication protein interacts with a protein kinase and a motor protein that display different expression patterns during plant development and infection. *Plant Cell*. https://doi.org/10.1105/tpc.003681

Krupovic, M., Ravantti, J. J., & Bamford, D. H. (2009). Geminiviruses: A tale of a plasmid becoming a virus. *BMC Evolutionary Biology*. https://doi.org/10.1186/1471-2148-9-112

Kunik, T., Palanichelvam, K., Czosnek, H., Citovsky, V., & Gafni, Y. (1998). Nuclear import of the capsid protein of tomato yellow leaf curl virus (TYLCV) in plant and insect cells. *Plant Journal*. https://doi.org/10.1046/j.1365-313X.1998.00037.x

Lacatus, G., & Sunter, G. (2009). The Arabidopsis PEAPOD2 transcription factor interacts with geminivirus AL2 protein and the coat protein promoter. *Virology*. https://doi.org/10.1016/j.virol.2009.07.004

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*. https://doi.org/10.1186/gb-2009-10-3-r25

Lazarowitz, S. G. (1992). Geminiviruses: Genome structure and gene function. *Critical Reviews in Plant Sciences*. https://doi.org/10.1080/07352689209382350

Legg, J. P., & Fauquet, C. M. (2004). Cassava mosaic geminiviruses in Africa. In *Plant Molecular Biology*. https://doi.org/10.1007/s11103-004-1651-7

Linford, N. J., Bilgir, C., Ro, J., & Pletcher, S. D. (2013). Measurement of lifespan in Drosophila melanogaster. *Journal of Visualized Experiments*. https://doi.org/10.3791/50068

Liu, L., Saunders, K., Thomas, C. L., Davies, J. W., & Stanley, J. (1999). Bean yellow dwarf virus RepA, but not Rep, binds to maize retinoblastoma protein and the virus tolerates

mutations in the consensus binding motif. *Virology*. https://doi.org/10.1006/viro.1999.9616

Lozano-Durán, R., Rosas-Díaz, T., Gusmaroli, G., Luna, A. P., Taconnat, L., Deng, X. W., & Bejarano, E. R. (2011). Geminiviruses subvert ubiquitination by altering CSN-mediated derubylation of SCF E3 ligase complexes and inhibit jasmonate signaling in Arabidopsis thaliana. *Plant Cell*. https://doi.org/10.1105/tpc.110.080267

Mansoor, S., Briddon, R. W., Zafar, Y., & Stanley, J. (2003). Geminivirus disease complexes: An emerging threat. In *Trends in Plant Science*. https://doi.org/10.1016/S1360-1385(03)00007-4

Mansoor, S., Zafar, Y., & Briddon, R. W. (2006). Geminivirus disease complexes: the threat is spreading. In *Trends in Plant Science*. https://doi.org/10.1016/j.tplants.2006.03.003

Markham, P. G., Bedford, I. D., Liu, S., & Pinner, M. S. (1994). The transmission of geminiviruses by Bemisia tabaci. *Pesticide Science*. https://doi.org/10.1002/ps.2780420209

Matzkin, L. M., & Markow, T. A. (2009). Transcriptional regulation of metabolism associated with the increased desiccation resistance of the cactophilic Drosophila mojavensis. *Genetics*. https://doi.org/10.1534/genetics.109.104927

Milward, J. H. (2018). Geminivirus: Rep as a Target for Conferring Resistance, the Host's Strigolactone Hormonal Response to Infection, and Host Resistance. *Master's Thesis*. https://repository.lib.ncsu.edu/bitstream/handle/1840.20/35671/etd.pdf?sequence=1&isAllowed=y

Nash, T. (2010). *Further Characterization of the Rep Protein and Using Peptide Aptamers as a Broad-based Resistance Strategy to Combat Geminivirus Disease.* [North Carolina State University]. https://repository.lib.ncsu.edu/bitstream/handle/1840.16/6423/etd.pdf?sequence=1&isAllowed=y

Navas-Castillo, J., Fiallo-Olivé, E., & Sánchez-Campos, S. (2011). Emerging Virus Diseases Transmitted by Whiteflies. *Annual Review of Phytopathology*. https://doi.org/10.1146/annurev-phyto-072910-095235

Núñez, E. D., & Aiello, A. (2013). Leafhoppers (Homoptera: Cicadellidae) that probe human skin: A review of the world literature and nineteen new records, from Panama. *Terrestrial Arthropod Reviews*. https://doi.org/10.1163/18749836-06001064

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., & Kanehisa, M. (1999). KEGG: Kyoto encyclopedia of genes and genomes. In *Nucleic Acids Research*. https://doi.org/10.1093/nar/27.1.29

Ong, C. T., & Corces, V. G. (2011). Enhancer function: New insights into the regulation of tissue-specific gene expression. In *Nature Reviews Genetics*. https://doi.org/10.1038/nrg2957

Orozco, B. M., & Hanley-Bowdoin, L. (1996). A DNA structure is required for geminivirus replication origin function. *Journal of Virology*. https://doi.org/10.1128/jvi.70.1.148-158.1996

Pasumarthy, K. K., Choudhury, N. R., & Mukherjee, S. K. (2010). Tomato leaf curl Kerala virus (ToLCKeV) AC3 protein forms a higher order oligomer and enhances ATPase activity of replication initiator protein (Rep/AC1). *Virology Journal*. https://doi.org/10.1186/1743-422X-7-128

Pavlov, Y. I., Maki, S., Maki, H., & Kunkel, T. A. (2004). Evidence for interplay among yeast replicative DNA polymerases alpha, delta and epsilon from studies of exonuclease and polymerase active site mutations. *BMC Biology*. https://doi.org/10.1186/1741-7007-2-11

Pedersen, T. J., & Hanley-Bowdoin, L. (1994). Molecular Characterization of the AL3 Protein Encoded by a Bipartite Geminivirus. *Virology*. https://doi.org/10.1006/viro.1994.1442

Petty, I. T. D., Carter, S. C., Morra, M. R., Jeffrey, J. L., & Olivey, H. E. (2000). Bipartite geminivirus host adaptation determined cooperatively by coding and noncoding sequences of the genome. *Virology*. https://doi.org/10.1006/viro.2000.0620

Philip, B. N., Yi, S. X., Elnitsky, M. A., & Lee, R. E. (2008). Aquaporins play a role in desiccation and freeze tolerance in larvae of the goldenrod gall fly, Eurosta solidaginis. *Journal of Experimental Biology*. https://doi.org/10.1242/jeb.016758

Pradhan, B., Tien, V. Van, & Dey, N. (2017). *Mol Biology of Virus Replication*.

Qin, S., Ward, B. M., & Lazarowitz, S. G. (1998). The Bipartite Geminivirus Coat Protein Aids BR1 Function in Viral Movement by Affecting the Accumulation of Viral Single-Stranded DNA. *Journal of Virology*. https://doi.org/10.1128/jvi.72.11.9247-9256.1998

Rizvi, I., Choudhury, N. R., & Tuteja, N. (2014). Insights into the functional characteristics of geminivirus rolling-circle replication initiator protein and its interaction with host factors affecting viral DNA replication. In *Archives of Virology*. https://doi.org/10.1007/s00705-014-2297-7

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2009). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btp616

Rojas, M. R., Jiang, H., Salati, R., Xoconostle-Cázares, B., Sudarshana, M. R., Lucas, W. J., & Gilbertson, R. L. (2001). Functional analysis of proteins involved in movement of the monopartite begomovirus, Tomato yellow leaf curl virus. *Virology*. https://doi.org/10.1006/viro.2001.1194

Roshan, P., Kulshreshtha, A., Kumar, S., Purohit, R., & Hallan, V. (2018). AV2 protein of tomato leaf curl Palampur virus promotes systemic necrosis in Nicotiana benthamiana and interacts with host Catalase2. *Scientific Reports*. https://doi.org/10.1038/s41598-018-19292-3

Ruhel, R., & Chakraborty, S. (2019). Multifunctional roles of geminivirus encoded replication initiator protein. *VirusDisease*. https://doi.org/10.1007/s13337-018-0458-0

Sanderfoot, A. A., Ingham, D. J., & Lazarowitz, S. C. (1992). A Viral Movement Protein as a Nuclear Shuttle. *Plant Physiology*.

Sanderfoot, A. A., Ingham, D. J., & Lazarowitz, S. G. (1996). A viral movement protein as a

nuclear shuttle: The geminivirus BR1 movement protein contains domains essential for interaction with BL1 and nuclear localization. *Plant Physiology*. https://doi.org/10.1104/pp.110.1.23

Saunders, K., Lucy, A., & Stanley, J. (1991). DNA forms of the geminivirus African cassava mosaic virus consistent with a rolling circle mechanism of replication. *Nucleic Acids Research*. https://doi.org/10.1093/nar/19.9.2325

Saunders, K., Lucy, A., & Stanley, J. (1992). RNA-primed complementary-sense DNA synthesis of the geminivirus African cassava mosaic virus. *Nucleic Acids Research*. https://doi.org/10.1093/nar/20.23.6311

Saunders, K., Norman, A., Gucciardo, S., & Stanley, J. (2004). The DNA β satellite component associated with ageratum yellow vein disease encodes an essential pathogenicity protein (βC1). *Virology*. https://doi.org/10.1016/j.virol.2004.03.018

Selth, L. A., Dogra, S. C., Rasheed, M. S., Healy, H., Randles, J. W., & Rezaian, M. A. (2005). A NAC domain protein interacts with Tomato leaf curl virus replication accessory protein and enhances viral replication. *Plant Cell*. https://doi.org/10.1105/tpc.104.027235

Settlage, S. B., See, R. G., & Hanley-Bowdoin, L. (2005). Geminivirus C3 Protein: Replication Enhancement and Protein Interactions. *Journal of Virology*. https://doi.org/10.1128/jvi.79.15.9885-9895.2005

Settlage, Sharon B., Miller, A. B., Gruissem, W., & Hanley-Bowdoin, L. (2001). Dual interaction of a geminivirus replication accessory factor with a viral replication protein and a plant cell cycle regulator. *Virology*. https://doi.org/10.1006/viro.2000.0719

Sharma, V., Kohli, S., & Brahmachari, V. (2017). Correlation between desiccation stress response and epigenetic modifications of genes in Drosophila melanogaster: An example of environment-epigenome interaction. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*. https://doi.org/10.1016/j.bbagrm.2017.08.001

Shen, W., Dallas, M. B., Goshe, M. B., & Hanley-Bowdoin, L. (2014). SnRK1 Phosphorylation of AL2 Delays Cabbage Leaf Curl Virus Infection in Arabidopsis. *Journal of Virology*. https://doi.org/10.1128/jvi.00761-14

Shen, Wei, & Hanley-Bowdoin, L. (2006). Geminivirus infection up-regulates the expression of two arabidopsis protein kinases related to yeast SNF1- and mammalian AMPK-activating kinases. *Plant Physiology*. https://doi.org/10.1104/pp.106.088476

Shepherd, D. N., Martin, D. P., Van Der Walt, E., Dent, K., Varsani, A., & Rybicki, E. P. (2010). Maize streak virus: An old and complex "emerging" pathogen. *Molecular Plant Pathology*. https://doi.org/10.1111/j.1364-3703.2009.00568.x

Sinclair, B. J., Gibbs, A. G., & Roberts, S. P. (2007). Gene transcription during exposure to, and recovery from, cold and desiccation stress in Drosophila melanogaster. *Insect Molecular Biology*. https://doi.org/10.1111/j.1365-2583.2007.00739.x

Singh, D. K., Islam, M. N., Choudhury, N. R., Karjee, S., & Mukherjee, S. K. (2007). The 32 kDa subunit of replication protein A (RPA) participates in the DNA replication of Mung bean yellow mosaic India virus (MYMIV) by interacting with the viral Rep protein. *Nucleic*

*Acids Research*. https://doi.org/10.1093/nar/gkl1088

Smyth, G. K. (2005). limma: Linear Models for Microarray Data. In *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. https://doi.org/10.1007/0-387-29362-0_23

Stanley, J. (1991). The molecular determinants of geminivirus pathogenesis. *Seminars in Virology*.

Stanley, John. (1993). Geminiviruses: plant viral vectors. *Current Opinion in Genetics and Development*. https://doi.org/10.1016/S0959-437X(05)80347-8

Stenger, D. C., Revington, G. N., Stevenson, M. C., & Bisaro, D. M. (1991). Replicational release of geminivirus genomes from tandemly repeated copies: Evidence for rolling-circle replication of a plant viral DNA. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.88.18.8029

Stinziano, J. R., Sové, R. J., Rundle, H. D., & Sinclair, B. J. (2015). Rapid desiccation hardening changes the cuticular hydrocarbon profile of drosophila melanogaster. *Comparative Biochemistry and Physiology -Part A : Molecular and Integrative Physiology*. https://doi.org/10.1016/j.cbpa.2014.11.004

Stone, A. (1942). Fruitflies of the genus anastrepha. In *Nature*.

Sun, M., Jiang, K., Li, C., Du, J., Li, M., Ghanem, H., Wu, G., & Qing, L. (2020). Tobacco curly shoot virus C3 protein enhances viral replication and gene expression in Nicotiana benthamiana plants. *Virus Research*. https://doi.org/10.1016/j.virusres.2020.197939

Sung, Y. K., & Coutts, R. H. A. (1995). Mutational analysis of potato yellow mosaic geminivirus. *Journal of General Virology*. https://doi.org/10.1099/0022-1317-76-7-1773

Sunter, G., & Bisaro, D. M. (1991). Transactivation in a geminivirus: AL2 gene product is needed for coat protein expression. *Virology*. https://doi.org/10.1016/0042-6822(91)90049-H

Sunter, G., & Bisaro, D. M. (1997). Regulation of a geminivirus coat protein promoter by AL2 protein (TrAP): Evidence for activation and derepression mechanisms. *Virology*. https://doi.org/10.1006/viro.1997.8549

Sunter, G., & Bisaro, D. M. (2003). Identification of a minimal sequence required for activation of the tomato golden mosaic virus coat protein promoter in protoplasts. *Virology*. https://doi.org/10.1006/viro.2002.1757

Sunter, G., Hartitz, M. D., Hormuzdi, S. G., Brough, C. L., & Bisaro, D. M. (1990). Genetic analysis of tomato golden mosaic virus: ORF AL2 is required for coat protein accumulation while ORF AL3 is necessary for efficient DNA replication. *Virology*. https://doi.org/10.1016/0042-6822(90)90275-V

Sunter, G., Sunter, J. L., & Bisaro, D. M. (2001). Plants expressing tomato golden mosaic virus AL2 or beet curly top virus L2 transgenes show enhanced susceptibility to infection by DNA and RNA viruses. *Virology*. https://doi.org/10.1006/viro.2001.0950

Swanson, M. M., & Harrison, B. D. (1993). Serological relationships and epitope profiles of isolates of okra leaf curl geminivirus from Africa and the Middle East. *Biochimie*. https://doi.org/10.1016/0300-9084(93)90101-W

Telonis-Scott, M., Guthridge, K. M., & Hoffmann, A. A. (2006). A new set of laboratory-selected Drosophila melanogaster lines for the analysis of desiccation resistance: Response to selection, physiology and correlated responses. *Journal of Experimental Biology*. https://doi.org/10.1242/jeb.02201

Terhzaz, S., Alford, L., Yeoh, J. G. C., Marley, R., Dornan, A. J., Dow, J. A. T., & Davies, S. A. (2018). Renal neuroendocrine control of desiccation and cold tolerance by Drosophila suzukii. *Pest Management Science*. https://doi.org/10.1002/ps.4663

Thorat, L. J., Gaikwad, S. M., & Nath, B. B. (2012). Trehalose as an indicator of desiccation stress in Drosophila melanogaster larvae: A potential marker of anhydrobiosis. *Biochemical and Biophysical Research Communications*. https://doi.org/10.1016/j.bbrc.2012.02.065

Unseld, S., Frischmuth, T., & Jeske, H. (2004). Short deletions in nuclear targeting sequences of African cassava mosaic virus coat protein prevent geminivirus twinned particle formation. *Virology*. https://doi.org/10.1016/j.virol.2003.09.003

Vinoth Kumar, R., Singh, A. K., Singh, A. K., Yadav, T., Singh, A. K., Kushwaha, N., Chattopadhyay, B., & Chakraborty, S. (2015). Complexity of begomovirus and betasatellite populations associated with chilli leaf curl disease in India. *Journal of General Virology*. https://doi.org/10.1099/jgv.0.000254

Wang, B., Li, F., Huang, C., Yang, X., Qian, Y., Xie, Y., & Zhou, X. (2014). V2 of tomato yellow leaf curl virus can suppress methylation-mediated transcriptional gene silencing in plants. *Journal of General Virology*. https://doi.org/10.1099/vir.0.055798-0

Ward, B. M., & Lazarowitz, S. G. (1999). Nuclear export in plants: Use of geminivirus movement proteins for a cell-based export assay. *Plant Cell*. https://doi.org/10.1105/tpc.11.7.1267

Yang, X., Baliji, S., Buchmann, R. C., Wang, H., Lindbo, J. A., Sunter, G., & Bisaro, D. M. (2007). Functional Modulation of the Geminivirus AL2 Transcription Factor and Silencing Suppressor by Self-Interaction. *Journal of Virology*. https://doi.org/10.1128/jvi.00617-07

Yu, G. (2018). clusterProfiler: universal enrichment tool for functional and comparative study. *BioRxiv*. https://doi.org/10.1101/256784

Yu, G., Wang, L. G., Han, Y., & He, Q. Y. (2012). ClusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS A Journal of Integrative Biology*. https://doi.org/10.1089/omi.2011.0118

Zacksenhaus, E., Jiang, Z., Phillips, R. A., & Gallie, B. L. (1996). Dual mechanisms of repression of E2F1 activity by the retinoblastoma gene product. *The EMBO Journal*. https://doi.org/10.1002/j.1460-2075.1996.tb00978.x

Zerbini, F. M., Briddon, R. W., Idris, A., Martin, D. P., Moriones, E., Navas-Castillo, J., Rivera-Bustamante, R., Roumagnac, P., & Varsani, A. (2017). ICTV virus taxonomy profile: Geminiviridae. *Journal of General Virology*. https://doi.org/10.1099/jgv.0.000738

Zhang, W., Olson, N. H., Baker, T. S., Faulkner, L., Agbandje-McKenna, M., Boulton, M. I., Davies, J. W., & McKenna, R. (2001). Structure of the maize streak virus geminate particle. *Virology*. https://doi.org/10.1006/viro.2000.0739

Zhou, Y., Rojas, M. R., Park, M.-R., Seo, Y.-S., Lucas, W. J., & Gilbertson, R. L. (2011). Histone H3 Interacts and Colocalizes with the Nuclear Shuttle Protein and the Movement Protein of a Geminivirus. *Journal of Virology*. https://doi.org/10.1128/jvi.00082-11

**Tables**

**Table 1.0:** ID Key for all of the analyses and figures.

| ID Code | Phenotype |
|---|---|
| HS | Sensitive Female |
| HD | Tolerant Female |
| HSE | Stressed Sensitive Female |
| HDE | Stressed Tolerant Female |
| MS | Sensitive Male |
| MD | Tolerant Male |
| MSE | Stressed Sensitive Male |

**Table 1.1**: List of all DEG for each comparison. DEG comparison made is in the first column and UP represents the number of upregulated genes for the comparison made and DOWN represents the number of downregulated genes for the comparison made. Comparison ID key found in **Table 1.0**.

| Comparison | UP | DOWN |
|---|---|---|
| HDE-MDE | 8721 | 1157 |
| HD-HDE | 1996 | 1183 |
| HD-MD | 4954 | 1828 |
| HSE-MSE | 1210 | 1407 |
| HSE-HDE | 1933 | 908 |
| HS-HD | 1391 | 1044 |
| HS-HSE | 1466 | 1196 |
| HS-MS | 2487 | 863 |
| MD-MDE | 7131 | 1831 |
| MSE-MDE | 1793 | 6445 |
| MS-MD | 1501 | 2139 |
| MS-MSE | 1272 | 951 |

**Table 1.2**: The three most differentially impinged pathways for each comparison in every GO category – Molecular Function (MF), Biological Process (BP), and Cellular Component (CC) – for both UP and DOWN regulated Differentially Expressed Genes.

| GO MF UP | GO MF DOWN | GO BP UP | GO BP DOWN | GO CC UP | GO CC DOWN |
|---|---|---|---|---|---|
| Membrane transporter activity | Transmembrane transporter | Lipid Metabolic Process | Cuticle Development | Neural Process | Axon Part |
| Lipase | Chitin binding | Lipid Catabolic Process | Carbon Metabolic Processes | Tight Junction | Sarcomere |
| Serine Hydrolase | cuticle structural construction | Cellular Lipid Metabolic Process | Import Into The Cell | Cell-to-Cell Junction | Synapse |

**Table 1.3:** Table depicting the genes ubiquitously expressed in both *Anastrepha* and *Drosophila*. The gene column contains the name of the genes found in *Drosophila* to be differentially regulated in response to desiccation stress and that were also found in our analysis of *A. ludens*. Also included are the accession numbers for the Flybase. All genes listed were found to be upregulated in every stressed sample.

| Gene | Accession Number | Function |
|---|---|---|
| Desi | CG 14686 | Transmembrane Helix - desiccation resistance |
| Frost | CG 9434 | Cold Tolerance / Thermal Adaptation |
| stmA | CG 8739 | Neurogenic development |
| HSP-23 | CG 4463 | Cold repsonse / thermal stress |
| Desat 2 | CG 5925 | Fatty acid desaturases |
| Gapdh 1 | CG 12055 | Cold Tolerance |
| Pepck 1 | CG 17725 | Gluconeogenesis |
| Pepck 2 | CG 10924 | Gluconeogenesis |
| Ino 80 | CG 31212 | Chromatin Remodeler |
| Mms 19 | CG 12005 | Mitotic Division Regulation |

**Table 1.4:** Most common pathways of the Gene Ontology Molecular Function category across all of the DEG comparisons.

| Molecular Function Pathways | |
|---|---|
| Histone Binding | Nucleosomal DNA Binding |
| Transmembrane Transporter | Membrane Transporter Activity |
| Hydrolase Activity | Serine Hydrolase |
| Chitin Binding | Cofactor Binding |
| Cuticle Structural Construction | Phosphotransferase |
| Actin Binding | Lipase |
| Nucleosome Binding | |

**Table 1.5:** Table 1.5 shows the DEG in the Neurotransmitter Transporter Activity pathway in the stress sensitive male vs stress tolerant male (MSE-MDE) and stress sensitive female vs stress tolerant female (HSE-HDE) comparisons, their respective Flybase ID, and the gene symbol for each gene. Overlap was found between the comparisons.

| Table 1.5 | | | |
|---|---|---|---|
| **MSE-MDE** | | **HSE-HDE** | |
| **Flybase ID** | **Gene Symbol** | **Flybase ID** | **Gene Symbol** |
| FBgn0033443 | CG1698 | FBgn0033443 | CG1698 |
| FBgn0033778 | Balat | FBgn0037976 | Tk |
| FBgn0033911 | VGAT | FBgn0032879 | CarT |
| FBgn0011603 | ine | FBgn0011603 | ine |
| FBgn0026438 | Eaat2 | FBgn0031939 | CG13796 |
| FBgn0037140 | SLC22A | FBgn0026439 | Eaat1 |
| FBgn0032879 | CarT | FBgn0037140 | SLC22A |
| FBgn0033708 | CG8850 | | |
| FBgn0010497 | dmGlut | | |

## Figures

**Figure 1.0:** Flowchart of the general process for the RNA-Seq analysis.

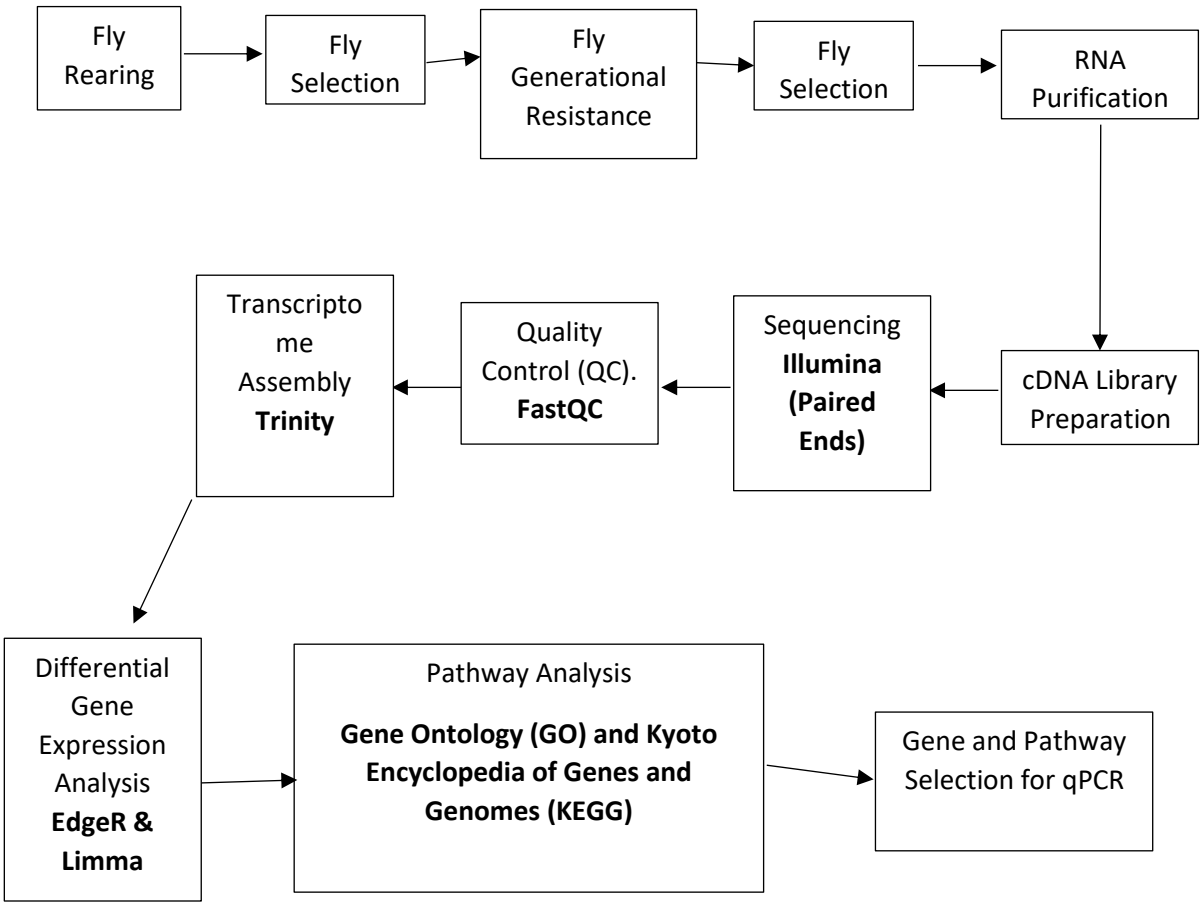
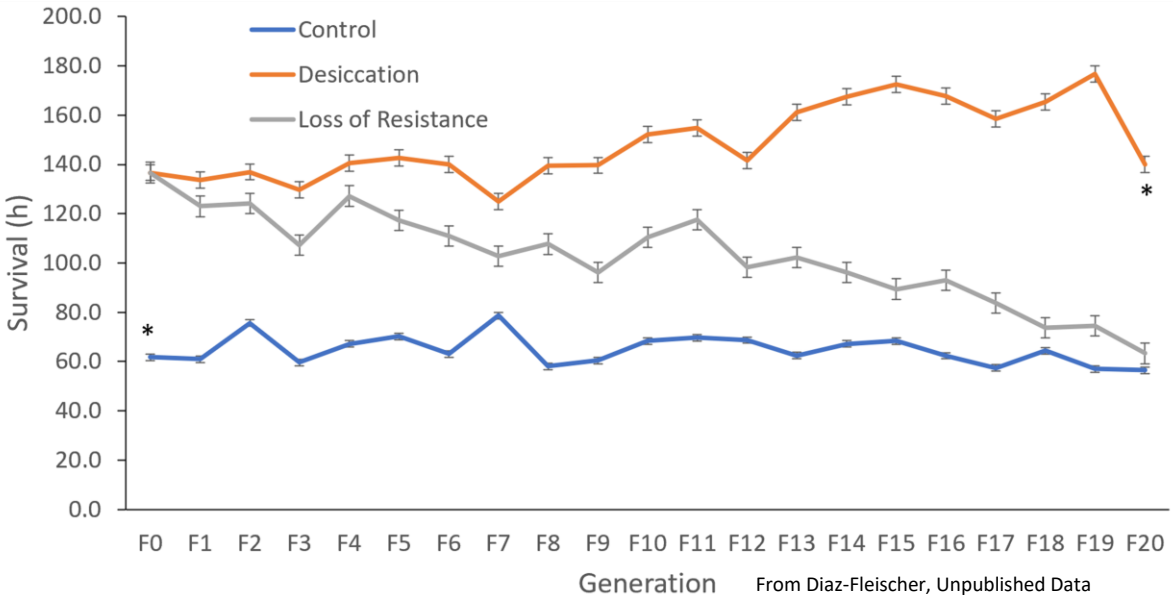


**Figure 1.1:** Graph showing a desiccation sensitive and desiccation resistant line compared to a self-selecting initially desiccation resistant line that lost its resistance after 20 generations. The blue control line represents the desiccation sensitive fly stock; the orange line represents the desiccation resistant flies; the gray line represents the inbred population of flies that lose resistance after 20 generations.




From Diaz-Fleischer, Unpublished Data

**Figure 1.2**: Depiction of all **male** GO MF UP regulated Active Transmembrane Transporter pathway on the left and the **female** GO MF UP regulated Active Transmembrane Transporter pathway on the right. Each of the large four ovals represents the number of genes in each of the comparisons (e.g. MS-MSE). The overlap between each of the four comparisons is segmented into different compartments for the genes that are shared between each of the comparisons. For example, the compartment with 14 represents 14 genes in the GO Molecular Function UP regulated Active Transmembrane Transporter pathway that are ubiquitous to the four comparisons. From the same pathway although in different sexes, the results are very similar.

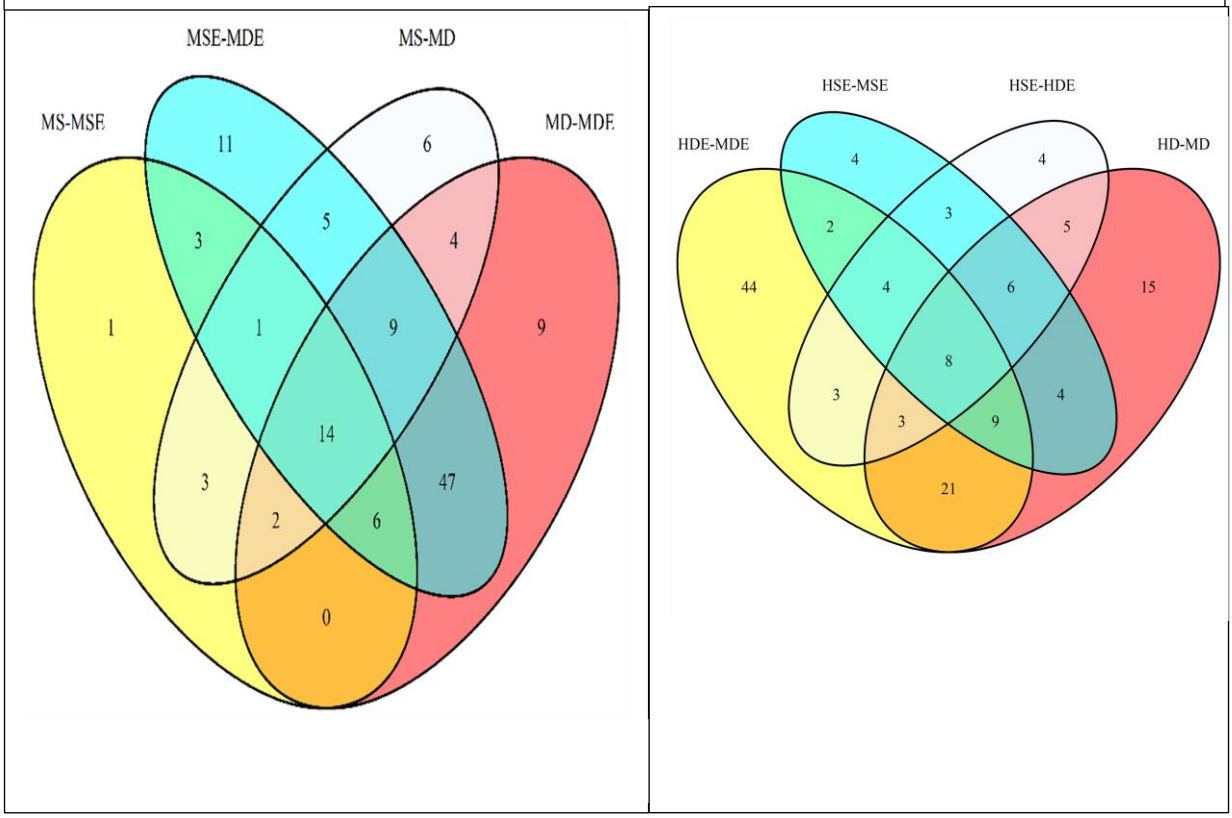
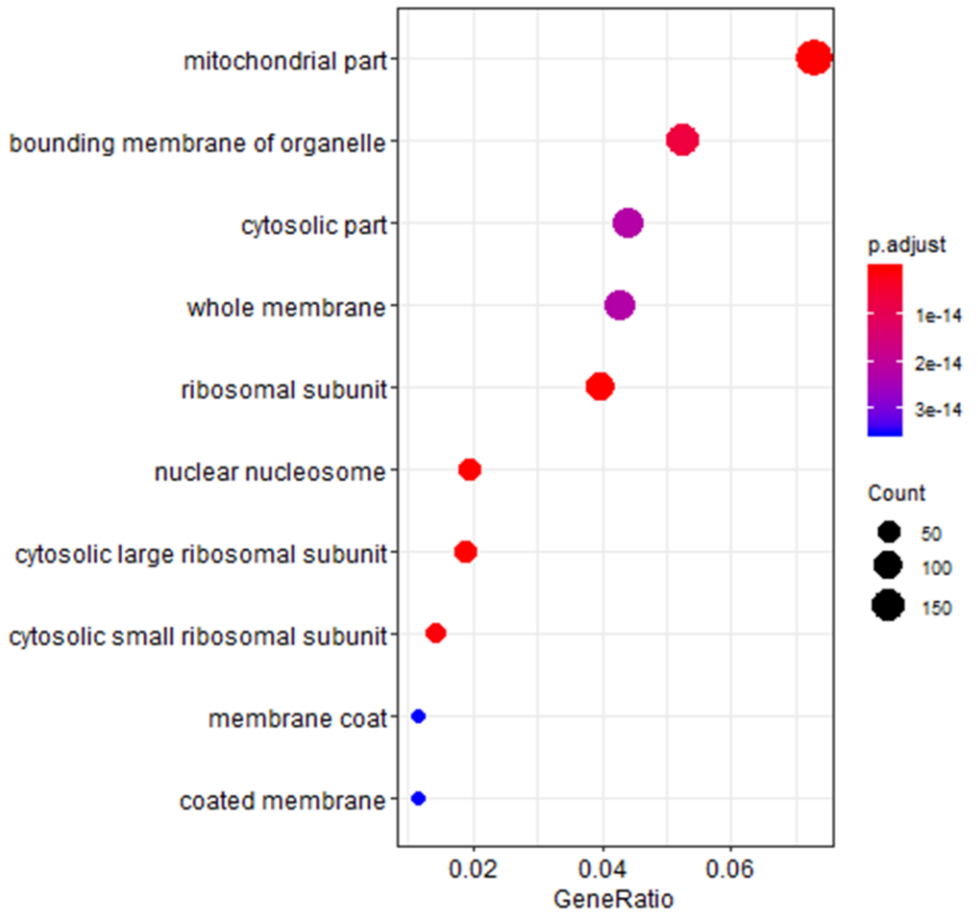


**Figure 1.3**: Dot plot showing the stressed tolerant female vs. stressed tolerant male (HDE-MDE) GO Biological Process (BP) UP pathways. Dot size is a function of the number of genes implicated in the pathway; color is a function of p-value significance – red is more significant than purple/blue; the x-axis is ranked according the gene ratio or the number of genes in the pathway compared to all the DEG in the comparison.



## CHAPTER 2: Structural Analysis of Geminiviral Proteins Rep and REn

### INTRODUCTION

Plant viruses are highly destructive financially causing billions of dollars a year in crop losses with concomitant staggering humanitarian impacts through food shortages and lost income (Mansoor et al., 2003, 2006). Disproportionately impacted are developing countries in equatorial climates with few or no cold seasons to eliminate plant virus vectors such as Kenya, Tanzania, Mexico, Guatemala, Honduras, and Brazil (Legg & Fauquet, 2004; Navas-Castillo et al., 2011; Shepherd et al., 2010). Climate, poor infrastructure, and lack of fundamental knowledge of plant viruses contribute to the inability to control and eradicate these plant viruses. A major family of highly destructive plant viruses and the focus of this chapter are Geminiviruses (*Geminiviridae*) which comprise nine genera: *Becurtovirus, Begomovirus, Curtovirus, Mastrevirus, Topocuvirus, Capulavirus, Eragrovirus, Grablovirus,* and *Turncurtovirus* (Zerbini et al., 2017). Understanding their structure and how they function is an essential tool in developing eradication and control strategies and is thus the focus of this chapter.

Perhaps the seminal supposition of biology is that the structure determines function. In protein biology, therefore, the structure of a protein is fundamentally related to and integral to the function of that protein (Berg et al., 2002). Viral proteins are highly specialized proteins that allow the virus to successfully propagate and are therefore both a product and a component of the virus (Joan L. Slonczewski, John W. Foster, 2017). Because of their inability to self-replicate, viruses rely on their ability to sequestrate host cell machinery – particularly those proteins concerned with viral replication (Berg et al., 2002). To facilitate recruitment of host cell machinery viral proteins must be relatively promiscuous and possess the ability to evade host cell surveillance factors. The ability to be resistant to host cell attack resides in the dynamic and

evolving nature of the viral protein sequence which disguises recognized pathogenic motifs. In addition, viral proteins are often highly multipurpose with numerous roles in the viral replication process (Berg et al., 2002).

It is important to recognize that that there are primarily two types of viruses: DNA based viruses and RNA based viruses (Durmuş & Ülgen, 2017). A DNA virus can be directly replicated in the host cell and transcribed whereas an RNA virus must utilize Reverse Transcriptase to convert the RNA to DNA for replication and then transcription in the host cell (Bhagavan & Ha, 2015). Viruses are currently classified into seven groups using the Baltimore and International Committee on Taxonomy of Viruses (ICTV) classification systems of which geminiviruses belong to Group II (ssDNA) (Zerbini et al., 2017). The criteria are based on strand polarity, double stranded or single stranded genomes, and DNA or RNA.

Geminiviruses are single stranded DNA (ssDNA) icosahedral shaped viruses with a monopartite or bipartite genome that infect plants (Davies et al., 1987; C. Gutierrez, 1999; Zerbini et al., 2017). Geminiviruses can be further subdivided into two subgroups based on host range and/or insect vector: dicotyledonous based geminiviruses that are transmitted by the whitefly and monocotyledonous based geminiviruses that are transmitted by different leafhopper and grasshopper species (John Stanley, 1993). A monopartite geminivirus consists of a single DNA component while a bipartite geminivirus consists of two DNA components (DNA A and DNA B) (Briddon et al., 2010; Fauquet et al., 2008; Petty et al., 2000). Bipartite geminiviruses have larger and each of the components plays a unique role in the viral lifecycle: the DNA A component codes for proteins that assist with viral DNA replication, encapsidation, viral suppressors of RNA silencing, and vector transmission; the DNA B component codes for proteins that support intercellular and intracellular movement of virus particles (Briddon et al.,

2010; Swanson & Harrison, 1993). Monopartite geminiviruses contain the equivalent of the DNA A component of the bipartite geminiviruses and thus the movement proteins normally supplied by the DNA B component are supplanted by the coat protein or the gene encoding Open Reading Frame (ORF) V2. Of the nine genera currently identified, only the *Begomovirus* genera contains members with a bipartite genome.

Geminiviruses have a small genome of approximately 2.5 kilobases (kb) (Hanley-Bowdoin et al., 2000; J Stanley, 1991). These genomes are packed into an icosahedral shaped capsid approximately 18-20 nm in diameter and a length of approximately 30 nm – see **Figure 2.0** for an image of the complete capsid (Gronenborn, 2004). The viral capsids are joined at the vertex of the icosahedra where there is a missing vertex to form a bipartite capsid (Bottcher et al., 2004; Krupovic et al., 2009; Zhang et al., 2001). Quite interestingly, bipartite viruses have a component distributed in separate capsids of the virion so as to facilitate an effective infective lifecycle both capsids must be transmitted to and their contents injected into the host cell. Accordingly, there are several steps to a geminivirus infection and successful replication and transcription that are reliant on the genes contained in the geminivirus genome.

Due to relatively small size of their genomes, geminiviruses only encode for a few genes and thus only a few protein products indicating that to effectively replicate they must sequester and utilize host cell machinery. All monopartite geminiviruses generally encode the following genes: Replication Associated Protein (Rep, AL1, and AC1), Transcriptional Activator Protein (TrAP, C2, and AC2), Replication Enhancer Protein (REn, C3, AC3) or RepA, C4/AC4, Pre-Coat V2 Protein V2/AV2, and Coat Protein (CP, V1, AV1). Bipartite geminiviruses however have the previously listed genes in Component A and in Component B they generally have two additional genes: Nuclear Shuttle Protein (NSP, BV1) and Movement Protein (MP, BC1)

(Zerbini et al., 2017). Further, geminiviruses sometimes have β-satellites that encode a gene entitled βC1 (Gnanasekaran et al., 2019; Jose & Usha, 2003; Saunders et al., 2004). Of note, geminiviruses do not encode for a polymerase from which the necessity of host cell polymerase arises (Hanley-Bowdoin et al., 2004). Each of these genes have different functions and locations throughout each of the various geminivirus classes and will further described below.

The most important gene in geminiviruses is Rep; it is critical that Rep be functional and available for replication or else the geminivirus will not replicate (Elmer et al., 1988; Saunders et al., 1991). Rep functions by recruiting and interacting with host cell factors as well as its own viral proteins. In fact, while Rep has been extensively studied and many of its domains have been mapped there remains a dearth of knowledge concerning its binding partners and structure – further analysis will follow below. Rep A is an alternately spliced version of Rep that in the four genera mastreviruses, becurtoviruses, capulaviruses, and grabloviruses is required for the geminiviruses to replicate in addition to the traditionally spliced Rep variant (Collin et al., 1996; Crisanto Gutierrez et al., 2004). As Rep A is a variant of Rep, it has many of the same functions as Rep and because the four aforementioned genera are lacking in REn, Rep A fulfills similar functions as REn (Sharon B. Settlage et al., 2001). C4 is present in all geminiviruses and has multiple functions that are highly varied across different geminiviruses species. Its functions can include activities as diverse as involvement in the cell-cycle regulator or involved in viral movement (Jupin et al., 1994; Rojas et al., 2001). Interestingly, C4 is the least well-conserved of the geminiviral proteins and its sequence is located within the Rep reading frame, but C4 has a different open reading frame (ORF) (Rojas et al., 2001). This could be the difference in the functions between different geminiviral species as each could have slightly different ORF's

leading to different structures and thus different functions. There are additional geminiviral proteins that are ubiquitous across all geminiviruses.

TrAP (AC2) has numerous functions which are essential to geminivirus replication and that are facilitated through its various domains. TrAP has been implicated in controlling replication by either enhancing or inhibiting replication, regulating basal cell metabolism, serving a role as a viral pathogenetic factor, and can suppress host cell-initiated gene silencing which serves to enhance viral gene expression and replication (H. Y. Chung & Sunter, 2014; Lozano-Durán et al., 2011). Despite the gene activation capabilities of TrAP, it does not bind directly to geminivirus gene promoter regions to enhance and facilitate their expression (Lacatus & Sunter, 2009; Sunter et al., 2001; Sunter & Bisaro, 1997; Yang et al., 2007). Instead, it can function by interacting with host cell proteins that then direct TrAP/shuttle complex to bind to the promoter regions of AV1 to facilitate its expression in the geminivirus infection lifecycle (Sunter & Bisaro, 1991, 1997). Interestingly, TrAP contains a functional zinc-finger domain, C-terminus acidic region, contains a nuclear localization signal (NLS), and is regulated by phosphorylation (Hartitz et al., 1999; W. Shen et al., 2014). Both the zinc-finger domain and the acidic region are required for TrAP to form oligomers with itself; in addition, phosphorylation is needed in order for complete functional utilization of TrAP in its oligomeric states (Hartitz et al., 1999). While TrAP contains an NLS, it has been found in either the cytoplasm or the nuclei or both depending on the geminivirus studied and the assay used to study its location and activity (Dong et al., 2003). However, to perform its primary function - gene activation – it must be localized to the nucleus where the DNA in located and thus the necessity of the NLS is observed.

In contrast to the Sunter and Bisaro perspective, the Rivera-Bustamante group has a different hypothesis concerning the binding of TrAP to the promoter region (Argüello-Astorga et

al., 1994). They propose that TrAP binds to Conserved Late Elements in the coat protein (CP) promoter region with the 8 bp sequence (GTGGTCCC) (Argüello-Astorga et al., 1994; Cantú-Iris et al., 2019). This is not inconsistent with the Sunter and Bisaro position as the 18 bp sequence (CGTCTAA*GTGGTCCC*GCA) that they identified as TrAP binding to non-specifically is within the CLE identified by Rivera-Bustamante's group (Sunter & Bisaro, 2003). Thus, since the 8 bp sequence is conserved and the flanking sequences on either side are not conserved, it is prescient to assert that this is the interaction site for TrAP.

AV2 is known as the precoat protein in geminiviruses. It has important roles in silencing the transcriptional gene silencing (TGS) surveillance mechanism in the plant host cell (Wang et al., 2014). TGS is the RNAi silencing mechanism utilized by plants to interfere with and attenuate pathogenic viral gene transcription. Thus, AV2 acts as a suppressor of TGS through interaction with host SGS3 that is known to be involved in the RNA-silencing pathway (Glick et al., 2008). Interestingly, via a naturally derived defense mechanism, TGS developed to suppress viral RNA transcript level and AV2 would seemingly be a natural case of virus-induced gene silencing (VIGS) (Chowda-Reddy et al., 2008; Fondong et al., 2007; Roshan et al., 2018). In contrast, AV1, is the only gene coding for a structural protein that contains several functions not intrinsically related to its primary function: protein packaging for virus components. In addition, AV1 can shuttle viral DNA between the nucleus and the cytoplasm which is critical for viral replication as well as packaging in the cytoplasm (Boulton et al., 1993; Liu et al., 1999). Finally, AV2 has been shown in monopartite Begomoviruses and monopartite Mastreviruses to facilitate movement of viral material cell-to-cell which is likely done in concert with the movement protein whose function is described below.

The remaining two proteins of interest that are unique to bipartite geminivirus genomes and are found in DNA component B are the nuclear shuttle protein and the movement protein (BV1 and BC1 respectively). BV1 binds to viral ssDNA to facilitate transport into the nucleus; in addition, this complex will bind to BC1 to facilitate cell-to-cell transmission (Hallan & Gafni, 2001; Kunik et al., 1998; Qin et al., 1998; Sanderfoot et al., 1992, 1996; Unseld et al., 2004; Ward & Lazarowitz, 1999). After new host cell infection with BC1, BC1 is released from the BV1-viral ssDNA-BC1 complex and BV1 will then shuttle the ssDNA into the host cell nucleus for replication. There are additional roles for BV1 such as interactions with Histone 3; there may be a synergistic role between AV2 which is also known to favorably alter methylation states for viral replication and BV1 (Wang et al., 2014; Zhou et al., 2011).

The final protein of interest unique to geminiviruses is the β-satellite protein βC1. βC1 is a novel protein that has been found to enhance viral pathogenicity via interactions with host cell protein to attenuate TGS (Vinoth Kumar et al., 2015). It has been recently discovered that βC1 can bind to *Nicotiana benthamiana* oxygen-evolving enhancer protein 2 (PsbP) which is involved in the Photosystem II (PSII) that has been shown to exhibit antiviral property involvement (Gnanasekaran et al., 2019). βC1 has been previously shown to induce symptoms *in planta* by interfering with and altering the structure of chloroplasts (Bhattacharyya et al., 2015; Bhattacharyya & Chakraborty, 2018). Further, it has been demonstrated that βC1 binds to and inhibits the activity of two proteins in the host mitogen-activated protein kinase (MAPK) cascade: MKK2 and MPK4 (Hu et al., 2019). The MAPK cascade is a long-term host-cell defense mechanism consisting of kinases activated to favorably influence host defense gene expression. Thus, by interfering and suppressing the ability of the MAPK cascade to proceed, βC1 effectively neuters a powerful element of the host cell defense response. In the end,

however, the most significant determinant of the combinatorial nature of geminivirus viral infectivity and symptomology appears to occur in the interaction between Rep and REn.

It has been shown that REn enhances the function of Rep by approximately one order of magnitude (Hanley-Bowdoin et al., 2000; S. B. Settlage et al., 2005; Sung & Coutts, 1995; Sunter et al., 1990). An exact explanation for this phenomenon has yet to be elucidated though good reasons have been hypothesized in the literature. The predominant understanding of these hypotheses hinge on understanding the binding partners of Rep and REn. It has been previously demonstrated that REn interacts much more favorably with the Proliferating Cell Nuclear Antigen (PCNA) and it is therefore thought that it can more readily recruit the PCNA to the viral replisome (S. B. Settlage et al., 2005). PCNA functions as a processivity factor for DNA pol δ and thus its quick recruitment to the replisome serves to enhance the overall viral replication levels (Castillo et al., 2003; Kelman, 1997). Interestingly, without REn, replication still occurs via Rep which may suggest that Rep contains the intrinsic ability to bind to and recruit PCNA to the replisome though not in as an efficient manner or that PCNA is recruited to the viral replisome after it is already assembled and mobilized. Perhaps another explanation is that it may bind just as readily to PCNA but not in the same ratio as it has greater numbers of binding partners than REn and thus stoichiometrically is unfavorable. A stochastic explanation of the interaction of Rep and REn with their binding partners would seem to support this stoichiometric explanation. Additionally, it could be that REn has an order of magnitude greater binding affinity for PCNA than Rep. To understand these potential differences, it is necessary to understand the mechanism of geminivirus transmission, infection lifecycle, and replication dynamics and the specific motifs contained in the geminiviral proteins that contribute to the replication process.

Geminiviruses are transmitted generally via two vector mechanisms: leafhoppers/treehoppers and the silver whitefly *Bemisia tabaci* (L. F. Chen & Gilbertson, 2009; Höfer et al., 1997). Leafhoppers are small insects that suck plant sap and belong to the family *Cicadellidae* and the order *Homoptera* (Núñez & Aiello, 2013). All of the genera of geminiviruses other than Begomoviruses are vectored by leafhoppers with the sole exception being Topocuviruses being transmitted by a treehopper that belongs to the same superfamily as the leafhoppers (Briddon et al., 1996). The whitefly *Bemisia tabaci* is the only known species of whitefly that transmits geminivirus and transmission occurs when the whitefly is feeding (Markham et al., 1994; Navas-Castillo et al., 2011). Geminiviruses are received and deposited through the following general mechanism: ingestion into the digestive tract, digestion and passage through the digestive tract to the hemolymph, entrance into the salivary glands, and then release with the saliva while feeding thereby introducing the virus particles into the new host. Once inside the host, a plethora of activity ensues that culminates in the replication and continued host transmission of the geminivirus leading to an infected plant, many times with concomitant symptomatology.

Viral infection consists of several steps that are generally consistent; geminivirus infection is simply a spin on the general steps herein outlined (Pradhan et al., 2017). The first step of viral infection is the attachment of the capsid to the cell membrane/wall; information on how this occurs in geminiviral infections is currently unknown and is a major gap in the geminivirus literature. The second step is the penetration of the virus or viral genetic contents into the cytoplasm of the cell. The third step of viral infection is replication of the viral genome after import into the host cell nucleus. The fourth step is the assembly/packaging of the replicated genome and the protein products into a capsid like structure. The fifth step is the release of the

encapsidated virus into the extracellular environment – either a facilitated mechanism or simply due to host cell rupture. Finally, the sixth step is transmission – the movement of the virus to a new host cell either via vector or proximity to infected cells, is important to point out that geminiviruses should reach the phloem to be transmitted. However, while geminiviruses follow these general steps, they utilize their own proteins to initiate and facilitate replication, encapsidation, and movement (Pradhan et al., 2017).

Geminiviruses mostly replicate via rolling circle replication (RCR) through dsDNA intermediates (Castellano et al., 1999; Lazarowitz, 1992; Saunders et al., 1991; Stenger et al., 1991). The other mechanism involved in geminivirus replication is referred to as Recombination-Dependent Replication (RDR) (Jeske et al., 2001). The only essential geminiviral protein required for this process is Rep and is shown graphically in **Figure 2.2** (Rizvi et al., 2014). RCR is initiated by the conversion of the geminiviral ssDNA to the dsDNA intermediate by extension of a short RNA primer for minus strand synthesis; although RCR is not fully understood in Begomoviruses (Donson et al., 1984; Saunders et al., 1992). Rep then binds to in a cooperative manner to the intergenic region (IR) highly conserved nonanucleotide sequence TAATATTAC and oligomerizes (Fontes et al., 1992). This presence of Rep results in the formation of a hairpin loop of the TAATATTAC sequence that is recognized by Rep – thus the hairpin loop formation is essential for replication (Orozco & Hanley-Bowdoin, 1996). This hairpin loop is then nicked by Rep at TAATATT//AC using the tyrosine residue in motif III of Rep; the 5’ end of the nick is then bound by Rep and elongation of the nick and therefore replication ensues at the 3’ end of the nick (Orozco & Hanley-Bowdoin, 1996). Rep then is essential in recruiting REn and various other host cell factors such as PCNA, minichromosome maintenance 2 (MCM2), replication protein A32 (RPA32), and replication factor C (REPC) to form the replisome and propagate

replication (L. J. Kong & Hanley-Bowdoin, 2002; Pavlov et al., 2004; Pradhan et al., 2017; Ruhel & Chakraborty, 2019; Wei Shen & Hanley-Bowdoin, 2006; Singh et al., 2007) . Upon completion of replication of the minus sense DNA strand, the same process is initiated on the parental geminiviral DNA strand.

Sometimes, however, RCR stalls and incomplete ssDNA exists and thus a further replication mechanism becomes necessary – in this case, Recombination Dependent Replication (RDR) (Formosa & Alberts, 1986; Jeske et al., 2001). RDR is merely an extension of the RCR process. Initially, viral ssDNA binds to circular dsDNA viral intermediate in a recombination event through a homologous sequence mediated through protein interactions. Next, loop migration and ssDNA elongation occur in much the same manner as RCR. Finally, complementary strand synthesis occurs, resulting in dsDNA – the same result as RCR (Jeske et al., 2001). To understand how Rep recruits and utilizes other viral protein and host cell factors, it is necessary to understand the structure and binding domains of Rep and REn.

Rep, as the only required viral protein for geminiviral replication, has a multifunctional capacity through several binding domains. As depicted in the **Figure 2.3**, the N-terminal region of Rep contains most of the binding domains: DNA cleavage and ligation, oligomerization, retinoblastoma and REn binding domain, and PCNA and GRIK binding domains. The C-terminal region contains the Helicase, ATP binding, and ATPase domain. The three conserved motifs I, II, and III are also contained in the N-terminal region. Rep also contains a Geminivirus Rep Sequence (GRS) in the N-terminal region that when mutated results in the lack of infectivity in the host organism.

Perhaps one of the most interesting binding domains in Rep is the retinoblastoma binding domain. In addition to Rep's capability to initiate replication, it also possesses the ability to

initiate the cell cycle and thus gene expression of replicative machinery. It accomplishes these functions through binding to plant retinoblastoma (pRBR) (Ach et al., 1997; L.-J. Kong, 2000). pBR is a protein that that interacts with the E2F family of transcription factors as well as Cyclin Dependent Kinases (CDK) (De Jager & Murray, 1999). The pBR-E2F interaction is perhaps of the most interest in the geminiviral replication lifecycle (Desvoyes et al., 2014). pRBR functions by binding to E2F transcription factors and inhibiting their function, which leads to a repression of the DNA replication (Zacksenhaus et al., 1996). Rep, however, can bind to pBR and remove it from the E2F complex, releasing the repression and thereby allowing the progression of the cell cycle to stimulate DNA replication. This is a positive benefit for the geminivirus as it increases the volume of the machinery/host cell replicative factors available for utilization in geminivirus replication. It is important to note that the plant DNA is also replicated at the same time as the viral DNA, so this is not a targeted DNA replication machinery that is de-repressed.

REn, also known as AC3, is the replication enhancing protein found in geminiviruses that increases the processivity of geminiviral replication. As stated previously, it has a high affinity for the processivity factor PCNA and thus serves to recruit PCNA to the replisome thereby increasing the overall viral load. However, despite similar functions to Rep, it is approximately one-third the length of Rep at 132 amino acids vs. 349 amino acids. REn has the capacity to bind to itself, to Rep, and to other host proteins such as PCNA and pRBR.

Functionally, REn is interesting as it has several binding domains at various locations throughout the protein. Perhaps one of the most fascinating elements of the binding interactions with REn is that the essential predictor of inhibitory functionality is charge reversal but not charge loss (S. B. Settlage et al., 2005). Charge reversal was found to affect binding to one of the primary protein interactors, PCNA, via yeast-two-hybrid assays in both Tomato Yellow Leaf

Curl Virus (TYLCV) and Tomato Golden Mosaic Virus (TGMV) (Castillo et al., 2003; S. B. Settlage et al., 2005; Sharon B. Settlage et al., 2001). Further, it has been demonstrated that the efficacy of REn in activating Rep is enhanced when it oligomerizes with itself to interact with its protein partners (Pasumarthy et al., 2010). For example, REn also enhances the ATPase activity of Rep at low concentrations by oligomerization and binding to already multimeric Rep (Pasumarthy et al., 2010).

REn also plays a regulatory role in the expression of the geminivirus genome and the transcript levels of Rep. In *N. clevelandii* protoplasts it has been shown that REn can bind to the Rep promoter and influence, via activation, Rep's transcription (Borah et al., 2016). In general, however, the exact mechanisms of autoregulation in geminiviruses are highly variable and can be quite different even amongst similar species of geminiviruses. For example, while REn may enhance Rep's function and thereby increase the overall viral load, the actual impact on the plant's symptomology may vary and thus not necessarily cause severe growth impingement as has been shown in *Nicotiana benthamiana* (Sun et al., 2020). Another example of the known regulatory capabilities of REn relates to its ability to interact with NAC-domain-containing transcription factor (SlNAC1) in *Solanum lycopersicum*. Through a yeast-two-hybrid experiment it has been shown that one of the potential mechanisms whereby REn helps in increasing the viral load is by inducing the gene expression of SlNAC1 (Selth et al., 2005). Then, REn and SlNAC1 colocalize to the nucleus where they have been shown to bind in concert to a yeast reporter gene promoter region. Clearly, REn plays an important although not indispensable role in regulating and modulating the geminiviral infection lifecycle both by directly recruiting host cell factors such as PCNA, inducing expression of host cell proteins, and then shuttling the expressed proteins it has induced.

The goal of this project is therefore twofold: to predictively model the structures of Rep and REn via various tools and to empirically validate these results using techniques such as NMR and XRC. A complete flowchart of the methodological approach is found in **Figure 2.1**. We attempted to use various different mathematical, statistical, and criteria-based modelling approaches to cross-validate the results of each tool using both homology and *de novo*-based modelling techniques to model the three-dimensional structure of Rep and REn. The results of the modelling were marginally successful with each tool providing a unique insight that when viewed in its entirety was useful. The results of the experiments were generally positive with purification and NMR analysis of full length (pNSB1531) Rep achieved while analysis of REn with NMR was not attempted due to constraints of the COVID-19 pandemic. These results have provided a set of general models and energetic information about Rep and REn that can be used in future molecular simulations as well as a framework for future experimental approaches to stabilizing Rep and REn and determining their structure. In addition, these models provide a basis to model peptide aptamers interactions with Rep and REn for virus control.

## EXPERIMENTAL PROCEDURES

### Modeling

To determine the structure of Rep and REn, we first decided to employ predictive modeling based on the sequence. Two primary structural modeling techniques exist and were utilized in this analysis: *de novo* modeling and homology modeling. *De novo* modeling takes an amino acid sequence and attempts to build a three-dimensional structure from the sequence using predicted secondary structural information and first principle physics concepts such as maximal and minimal angles. Homology modeling compares the amino acid sequence to similar sequences and then builds a three-dimensional structure based on similar motifs and physics concepts. The modeling program Rosetta (https://www.rosettacommons.org/, Version: 3.12) was the primary local software used. In addition, several online modeling programs were utilized that use less rigorous techniques but are more accessible to validate the results obtained via Rosetta.

Hardware used for all analyses: Ubuntu Version 20.04 LTS on local machine with 60 Gb RA, 2 Gb dedicated graphics card NVIDIA GeForce GT 710, and 500 Gb Samsung EVO Series Solid State Drive (SSD).

### Rosetta AbInitio

The first analysis used in Rosetta was *de novo* analysis. This requires four files: fasta sequence file, psipred_ss2, psipred_frag3 file, and psipred_frag9. The fasta file contains the amino acid sequence of the protein to be analyzed, psipred_ss2 is a file that contains the predicted secondary structure of the protein given the fasta input. The psipred_frag3 file contains a list of all possible 3-mers and the psipred_frag9 contains a list of all possible 9-mers from the fasta input. If the analysis were to do homology modeling, the input would contain a fifth file that is the .pdb file of the resolved structure or one very similar in sequence and function as the

input fasta. All psipred files were generated using an online tool from the UCL-CS workbench (Buchan & Jones, 2019).

To execute several structure resolution models at the same time, the desktop multiprocessor was overclocked and each "logical" processor was ascribed a structure and parameters to resolve using mpirun – oversubscribe -np X where X is the number of processors into which to partition the actual processor. The AbinitioRelax.mpi.linuxgcrelease script (V 3.12 and included in Rosetta local installation) was utilized and takes the – database command to read in the input files. -nstruct X was used to specify the number of structures where X is 20,000 structures. Because the lowest energy structure was desired, the option -abinitio:relax was given. Further, the option -abinitio::increase_cycle X, where X is 10 was passed. The following options: - abinitio::rg_reweight 0.5 -abinitio::rsd_wt_helix 0.5 -abinitio::rsd_wt_loop 0.5 -relax::fast, were passed based on the previously empirically determined best options as stated by Rosetta developers and help resolve structural clashes. Finally, the -out:file:silent ./XXX_silent.out was specified to save the output of the program iteratively to the output file for immediate viewing and to prevent data loss if the program was interrupted as well as to free up RAM space.

Because each processor output its structural data to a separate file (e.g. 8 logical processors = 8 files), it was necessary to combine them using the combine_silent.default.linuxgcrelease program built into Rosetta. This file is necessary as we want to extract the generated data for predicted structures into .pdb files. Rosetta assigns a score to the predicted structure based on the various parameters specified in the analysis. These scores must then be ordered sequentially if we want to view the highest confidence structures and ignore the low-quality structures. This was done by using the grep function in Linux to grad the scores from the automatically generated scores.fsc file and then the sort command was used to

sort them in descending order. It is now necessary to remove the X number of top highest scoring files from the sorted scores file where here X = 200 and save them to a new file "top_200" using the cat function in Linux.

From here we needed to extract those top 200 highest scoring .pdb's from the combined silent output file. We passed the xargs function in Linux and used the built in Rosetta extraction command extact_pdbs.default.linuxgcrelease. We then continued with an optional step to cluster the .pdb's based on shared parameter characteristics. We therefore made a directory into which all .pdb's were added. Then the native Rosetta clustering program cluster.default.linuxgccrelease was used to see if there were any clusters in the generated .pdb's. Finally, the .pdb's were extracted from the clusters if any were generated and saved is .pdb format for later analysis and viewing in a .pdb structure viewer. At this juncture, PyMol (https://pymol.org/2/, Version 2.4) was employed to visualize the generated structures.

**Robetta Server**

The online server corollary to Rosetta, Robetta (https://robetta.bakerlab.org/) was used to validate the results obtained from the Rosetta analysis. The Robetta server, maintained by the David Baker lab, takes an input the fasta sequence and the user then selects if the modeling to be done is comparative/homology, *ab initio*, and/or domain prediction. Further, the user can also manually input constraints to model only certain residues and then select the mathematical model to be initiated on the constrained model; e.g. gaussian, bounded, sigmoid, or harmonic. We simply allowed the default settings on Robetta – both homology and *ab initio* based modeling – to be the run on the Rep CbLCV and REn CbLCV amino acid sequences in fasta format as shown in **Table 2.0.** Sequence alignment for Rep CbLCV, Tomato Golden Mosaic Virus (TGMV), Tomato Yellow Leaf Curl Virus (TYLCV), and Wheat Dwarf Virus (WDV) are shown

in **Table 2.1** as are sequence alignments for REn CbLCV, TYLCV, and WDV. Seqeunce alignments were done in Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/). The psipred and variants to predict the secondary structure were automatically computed in Robetta and their input was not needed.

**Phyre2 Server**

To obtain an additional model of Rep and REn using a different analytical technique from Robetta, the online Phyre2 server (http://www.sbg.bio.ic.ac.uk/~phyre2/, Version 2.0) was utilized. Phyre2 conducts only homology-based modelling. The input was the amino acid sequence of Rep CbLCV and REn CbLCV; the intensive option was selected to obtain the most complete modelling information for each protein. The ProQ2 program native to Phyre2 is used to assess model quality from the models generated by Phyre2.

**I-TASSER**

I-TASSER (https://zhanglab.ccmb.med.umich.edu/I-TASSER/, Version 5.1) was used to obtain additional modelling information in the form of .pdb files as well as predicted function information through its homology based modeling. The available constraint options are much more significant than Robetta and Phyre2, but for this analysis they were not used as whole protein analysis was desired. Further, incomplete secondary structure is unknown and therefore in order not to bias the analysis all constraint information was omitted.

**SWISS-MODEL: Expasy**

The Swiss model using Expasy (https://swissmodel.expasy.org/) was also used to generate structural data for both Rep CbLCV and REn CbLCV through its homology based modeling techniques. The amino acid sequences for both proteins was inputted into the server

and after generation of the models, the server identified templates were examined for comparison and the overlap and Ramachandran plots were analyzed.

**PROTEIN EXPRESSION**

The proteins Rep CbLCV and REn CbLCV were selected as candidates for protein structural analysis using XRC and NMR. We used previously generated plasmids for our analysis; pNSB1529, pNSB1530 and pNSB1531 plus pNSB195 (Nash, 2010). Rep CbLCV plasmids generated are $pNSB1529_{1\text{-}121}$ (one-third length Rep), $pNSB1530_{1\text{-}177}$ (half-length Rep), and $pNSB1531_{1\text{-}349}$ (full length Rep) with the subscript representing the length of the plasmid generated. The REn plasmid has the identifier $pNSB195_{1\text{-}132}$ and is the full length REn (Pedersen & Hanley-Bowdoin, 1994). However, because they had not been used recently, it was necessary to first validate the plasmids as well as transform them into BL21-DE3 *E. coli* cells (ThermoFisher Cat: EC0114) for optimal expression.

**Plasmid Transformation**

BL21-DE3 chemically competent *E. coli* cells were transformed with the plasmids pNSB1529 (31 ng/µl), pNSB1530 (22 ng/µl), pNSB1531 (29 ng/µl), and pNSB195 (32 ng/µl) containing the antibiotic selection gene for Ampicillin / Carbenicillin. After following the manufacturer's protocol for transformation (electroporation), aliquots of the transformed cells were plated onto LB-Agar plates with 100 µg/ml of Ampicillin (Amp). Plates were left overnight at 37 °C. Single colonies were picked the next day and placed into a 3 ml solution of LB and 3 ml Amp (100 µg/ml) and agitated at 250 rpm overnight at 37 °C; the plates were placed at 4 °C for later use. 1 ml of each overnight culture was used to purify DNA plasmid using a Qiagen DNA plasmid miniprep kit (Cat No./ID: 27104) following the Qiagen protocol: new plasmid was obtained for later analysis as original stock was extremely low. Further, 1 ml of each overnight

culture was then mixed with glycerol to make a 50/50 mixture of cells to sterile glycerol as a stock and stored at -80 °C in glycerol stocks.

**Induction**

To establish in principle that our techniques would work, an initial working culture of 10 ml was prepared and analyzed as follows. Once validated, the protein production was scaled up to 1-liter cultures and the volumes of the reactants were adjusted accordingly.

A swipe through each glycerol stock with a pipette tip was then thoroughly eluted into 5 ml LB and 100 µg/ml Amp and agitated at ~ 250 rpm overnight at 37°C. 1 ml of this overnight culture was then added to 10 ml of fresh LB and 100 µg/ml Amp and incubated at 20 °C while shaking at ~ 250 rpm until the $OD_{600}$ was > 0.6. The culture was then induced with 1 mM IPTG (to induce protein production via the inducible promoter in front of Rep as a function of the vector design for BL21-DE3 cells) and was shaken at ~ 250 rpm overnight at 20 °C.

To determine the ideal length of induction, seven 100 µl aliquots were taken each hour after the addition of 1 mM IPTG while shaking at 250 rpm at 20 °C. These aliquots were stored at -20 °C for later Immunoblot analysis.

**Cell Lysis**

Cultures were centrifuged at 4360 x g for 10 minutes to pellet the cells, the supernatant was poured off, and then the pellet was resuspended in 1 ml lysis buffer (20 mM $NaH_2PO_4$, 300 mM NaCl, 1 mM $MgCl_2$), 1 mM Pefabloc, and 1 mM lysozyme on ice and sonicated at 20% amplitude in five 30 second pulses on ice. Then the suspended and lysed cells were centrifuged at 12,100 x g at 4 °C for 30 min to separate the cellular components from the total protein. The pellet was then stored at -80 °C for later use in Immunoblots. The supernatant was filtered

through a 20 µm filter to remove any protein aggregates and cell detritus that would block the purification column.

**Purification**

The supernatant was applied to a $Ni^{2+}$ pre-packed, equilibrated column (1 ml HisTrap from GE: 71502813-EI) at a flow rate of 1 ml per minute. The column was then washed with 10 column volumes (CV) of binding buffer (20 mM $NaH_2PO_4$, 500 mM NaCl, 100 mM imidazole). The bound protein was eluted from the column using 2 CV of elution buffer (20 mM $NaH_2PO_4$, 500 mM NaCl, 1 M imidazole). The eluted protein concentration was then determined on a NanoDrop machine at $A_{280}$. This process had to be repeated a few times until we worked out the ideal buffer conditions. The column has a theoretical binding capacity of 40 mg/ml and the desired target for NMR is 10-15 mg/ml at ~ 350 µl while XRC is about 1 ml of 1-10 mg/ml.

**Immunoblots**

The next step was to run Immunoblots to ensure that we had actually obtained Rep and not a spurious protein or incorrectly cloned plasmid. To determine peak protein production an induction was performed from T0 – T7 in hours; each hour 100 µl was saved for Immunoblot analysis. The Immunoblot was run with a protein ladder and each induction aliquot via the instructions from the preassembled gel from ThermoFisher (e.g. 20 µl 2X SDS buffer, 16 µl sample, 4 µl DTT and incubate at 82 ℃ for 2 minutes). Twenty µl of each sample was loaded onto the gel in addition to the protein molecular weight marker (MWM). The gel was allowed to run in buffer condition for 35 minutes at 225 volts. The gel was then stained with reversible Aquastain and images captured of the stained gel. The gel was then destained with water and the protein transferred to a PVDF membrane in buffer for 1 hour at 30 V. The membrane was then washed with a washing buffer (5% BSA w/v, 0.1% Tween 20, 1 X TBS) for 30 minutes while

gently rocking. A primary antibody to Rep produced in Rabbits (Rabbit α-Rep) was then allowed to rock at room temperature for 1 hour in washing buffer at 1:10,000 ratio. The membrane was then washed 3 times for 5 minutes each with the washing buffer. A secondary antibody to the rabbit antibody (IRDye 800 CW α Rabbit) was allowed to incubate for 1 hour at room temperature while rocking in washing buffer and then washed 3 times for 5 minutes each. The membrane was then visualized using a LiCor imaging machine scanning at 800 and 700 nm.

**Nuclear Magnetic Resonance**

The purified sample was then analyzed using natural abundance ($^{14}N$) NMR on the instrument Bruker Ascend$^{TM}$ 700 at the Molecular Education, Technology and Research Innovation Center (METRIC) at NCSU. This was done in our elution buffer that contains high salt and contains a high concentration of imidazole. The imaging took approximately 7 min after manually tuning the magnet to account for the unique magnetic frequency of the sample. Approximately six thousand different runs were made with the NMR machine to gather enough data to have a high degree of confidence in the results obtained. The sample was then collected and saved for later use, as necessary. All NMR acquisition and post-acquisition analysis was carried out by Dr. Peter Thompson in the Department of Structural and Molecular Biochemistry at NCSU.

**Mutation Analysis**

Rotameric analysis was done for TGMV at the conserved K145 residue for conversions to Alanine to determine if a cause for loss of function could be determined since it is known that a K145A mutation results in a diminution of function for Rep. Visualization was done in PyMol (https://pymol.org/2/, Version 2.4) and Rotameric analysis was done in UCSF Chimera (https://cgl.ucsf.edu/chimera/, Version 1.14 (build 42094)).

## RESULTS

### Modeling

The general consensus of each of the modeling tools employed seemed to indicate a relatively high degree of uncertainty in the predicted structure and function of both Rep and REn. With respect to Rep, there is relative consistency in predicting the secondary structures with marginal consensus as to the placement of those secondary structures in three-dimensional space. It should be mentioned, however, that the general location of these secondary structures when folded were in the same general location. In addition, many of the homologous structures independently chosen by the programs for homologous based modeling were often different with significantly varying sequence coverage. However, there were several similarities – aside from the predictive uncertainty – that addressed the unstable regions of Rep. At this point, there are no solved structures for full length Rep. Two structures – 1L2M.pdb and 6Q1M.pdb – have been determined using NMR and crystallography, respectively, to determine the first 120 amino acids of the N-terminal region of tomato yellow leaf curl virus and wheat dwarf virus, respectively, as shown in **Figure 2.8**. The important differentiation is that both viruses are monopartite and not bipartite Begomoviruses and thus the structure of Rep between each may be slightly different in important nuanced ways.

With respect to REn, the modelling accuracy was varied but generally quite poor. Template selection for each of the programs to develop a model for REn was the biggest differentiator between the various programs and the models generated. On average, the programs were successful in predicting the function of REn based on the understanding of current literature. The best model was generated via Robetta using *de novo* based analyses.

Summaries of the modeling techniques and programs used for both Rep and REn and the results obtained from the analyses are found in **Table 2.2**.

**Rosetta *Ab Initio* – Rep**

The results from the Rosetta analysis of Rep were varied and wide ranging. There was little consensus as to the lowest energy structure(s) despite running 20,000 model simulations. There was, however, consensus as to the numbers of secondary structures, type, and general placement in three-dimensional space. Analysis of the modeled structures generated from the modeling analysis indicated a plethora of steric clashes.

**Rosetta *Ab Initio* – REn**

The results from the *ab initio* Rosetta analysis of REn were more inconsistent than Rep with much greater variability in the placement of secondary structures in three-dimensional space. For example, the top 5 scoring .pdb models of REn were analyzed in PyMol and aligned to each other and they showed no consistent or significant structural overlap.

**Robetta Server – Rep**

The advantage of using the Robetta server is the computational power afforded to its users. The results from the analysis of Rep using the Robetta server were consistent with those obtained in other analyses with respect to the degree of confidence in the models obtained. Even modeling the individual domains detected automatically by Robetta and constrained by Gremlin and then modeling the entire structure the confidence levels in the generated structures were consistently about 22%. Oddly enough, this holds true for both *ab initio* and homology-based modeling with the Robetta server.

**Robetta Server - REn**

The results from the Robetta server concerning REn were actually quite positive as opposed to Rep. Robetta attempted both homology and *ab initio* based modelling. The homology modelling yielded poor results: specifically, a model with 0% confidence in the result. However, the *ab initio* based modelling yielded a fully folded and detailed structure unlike any other program herein attempted with a confidence level of 40% and is depicted in **Figure 2.9** This is impressive considering its confidence is much higher than all of the other programs attempted as well as the fact that the models produced were nicely folded with favorable interactions as evidenced by the Gremlin data. The average angstrom error estimate was less than 4 Å for each residue with a few outliers in either extreme.

**Phyre2 Server – Rep**

The results from the Phyre2 server were actually positive with a model predicted with high confidence – 100% in Normal mode and is depicted in **Figure 2.10**. The server only modelled the first 123 out 349 residues starting from the N-terminal region of the sequence due to limitations of the computational capacity of Phyre2 server allocation. Interestingly, out of all the various models attempted, Phyre2 was the only program capable of explicitly detecting and recognizing that the queried protein was an origin of replication-binding domain containing protein and that it belongs to the family of DNA binding Rep proteins. When the data was observed further, the Ramachandran analysis identified no energy anomalies in the predicted model. Further, when the alignment confidence was calculated there were no residues with completely disordered alignments. Finally, when looking at the theoretical conservation analysis of the aligned sequence (query sequence vs. modelled sequence) there was very little conservation needed throughout the sequence except for residues 220-227.

When the intensive modeling technique was utilized the results suggested that 91% of all the residues were modelled at greater than a 90% confidence threshold. When the ancillary analytical tools were employed, the crystal structure of the wheat dwarf virus Rep domain (WDVRD) was used as the structure/sequence with greatest coverage. Only the first 120 amino acids of the full-length Rep model were compared to the WDVRD as the WDVRD crystal structure consisted of only 120 amino acids. In addition, the alignment confidence in the comparison between the sequences was high; a stark contrast to the "Normal" modeling mode previously employed. The Ramachandran analysis indicated that only 2/120 residues had unfavorable energetic profiles. Again, the theoretical conservation analysis indicated that the only residues that are highly conserved are 220-227. Further, the intensive modeling detected approximately two binding pockets; 23-32 and 80-85 that are located approximately on the same quarter of the protein. Finally, the ProQ2 quality assessment of the Phyre2 server yielded an overall and local good score where the range is good to bad with a significant color gradient between them.

**Phyre2 Server – REn**

The results from the Phyre2 Server were similar to those obtained in other analyses. The models had a confidence level of ~ 20% with 11% coverage in normal mode. This is not surprising given how small REn is (132 amino acids) and how it can be challenging to model such a small protein on such relatively few molecular interactions as well as a lack of homologs for homologous based modelling. In addition, the domain analysis yielded few hits and thus had very low coverage at 11 %.

When the intensive modeling technique was utilized (full length REn) the results were even more dismal: 0% of the residues were modelled at a confidence level greater than 90%. The

closest structure (19.1%) to the predicted structure of REn by Phyre2 is c4pc4B.pdb and is known as a lipid binding protein; pdb name: bombyx mori lipoprotein 6. When viewing the Ramachandran plot for the overlay of the modelled REn onto the c4pc4B there were only 4/50 residues with unfavorable energetics. Further, only the terminal amino acids at the C and N-termini were considered relatively disordered. Interestingly, the alignment with c4pc4B aligned predominantly with the N-terminal region of the REn protein in a manner similar to the alignments noted in Rep. Finally, the ProQ2 quality assessment of the modeled REn vs. c4pc4B alignment at the local and global levels was overall of moderate quality.

**I-TASSER – Rep**

The models generated via I-TASSER for Rep were of poor quality and inconsistent in structure. The confidence score was -4.82 with -5 being the lowest confidence in the model and the estimated RMSD was 19.0±2.1Å. In addition, the estimated similarity value (TM) - also referred to as the Root-Mean-Square-Deviation (RMSD), but referred to by I-TASSER as TM - assigned to the generated models and their respective homolog template was 0.22±0.06. This is a very low number that reflects the inconsistent topology between the template and the generated model. A normal RMSD from other homology-based modelling programs would be in the range of 10-25 Å; a good crystal structure RMSD is in the range of 1.6-2.9 Å. The lower the RMSD the better the resolution of the model. Furthermore, the top 10 identified structural analogs in the Protein Data Bank (PDB) had little or no similar functions to Rep. I-TASSER also failed to identify the known partially solved structures of Rep in the PDB and utilize them as homologs. Also, I-TASSER has a function that identifies GO terms that are related to the homologs of the modelled protein. I-TASSER failed to identity most of Rep's critical functions as a replication protein. It successfully identified that Rep is a nucleic acid binding protein as well as having

helicase activity and is depicted in **Figure 2.11**. Overall, I-TASSER was not an ideal tool in the determination of the structure of Rep predictively as it seems its homology-based modelling for Rep was not ideal and failed to pick structures that had the most similar functions.

**I-TASSER – REn**

The results from the I-TASSER analysis were consistent with the other analyses we have obtained. The modelling generated several relatively low confidence models and one is depicted in **Figure 2.12**. While the program was able to successfully predict secondary structures with the same efficacy of the other programs, it was not able to generate a high confidence low RMSD model. While the model did have a TM-score of 0.29±0.09 which was above the 0.17 cutoff - signifying that the model and the template had a non-random level of similarity - the confidence score in the predicted model (C-score) was -3.9 with the range from -5 to 2. Thus, the overall quality of the model and the confidence in the accuracy of the model were on the low end and therefore relatively poor.

However, I-TASSER successfully predicted the functions of REn that were reported in GO terms. For example, ATP binding, nucleotide binding, helicase activity, DNA replication, and nucleic acid metabolic process were all GO terms that describe many of the functions of the top 10 homologous protein hits.

Interestingly, many of the top structural homolog hits to REn were proteins that have helicase and replication initiation functions. For example, 2z4rA.pdb and 5xdrA.pdb were the first and second highest ranking structural hits, respectively, and both are essential to replication initiation and helicase activity in *E. coli* and *H. sapiens*. Further, structurally similar proteins had functions compatible with helicase loading, DNA remodeling, transposase recruitment, and DNA replication regulation protein.

In addition, I-TASSER provides feedback on the predicted ligand binding domains and the ligands that could bind to those binding sites. I-TASSER identified several ligand binding sites throughout the fully modelled REn; however, most of the binding sites were between residues 80 – 110, 50-70, and 10-32. The ligands thought to bind in these binding domains include peptides and nucleic acids. Specifically, the program identified transhydrogenase as a potential peptide binding partner of REn which is currently not described in the literature.

**SWISS MODEL: Expasy – Rep**

Expasy generated five models for Rep. Two of the metrics used by Expasy to estimate the quality of the models generated are the Global Model Quality Estimation (GQME) and the Qualitative Model Energy Analysis (QMEAN). The GQME is a quality estimation for the alignment of the query sequence with the searched template and the quality of the resulting model based on the sequence coverage and thus the quality of the template alignment. The range of GQME values is from 0 – 1 with a value closer to 1 being desired. The QMEAN is a measure of the global and local quality estimates for all the residues in each individual model. The QMEAN values have a dynamic range with anything less than -4.0 representing a low-quality model and a value around zero representing a model with good agreement between the predicted structure and the homologous template.

For Rep, Expasy generated five potential models. The highest GMQE was 0.19 and the associated QMEAN value was -2.50 indicating a relatively decent model based on a poor alignment with the best fit template. It quite accurately determined the function of Rep as a replication initiation protein, helicase loader, and primosome component. The maximal sequence identity between any of the homologs as determined by Expasy was 51 %. All of the templates

for the modelling of Rep were .pdb structures from the PDB and whose function was either as a replication protein or replication associated protein: 1L2M.pdb, 6Q1M.pdb, and 1L5I.pdb.

**SWISS MODEL: Expasy – REn**

The two models generated in Expasy for REn were slightly better than those generated for Rep. The top GQME score was 0.12 while the top QMEAN score was -1.66. While these models are adequate and the QMEAN score is in the desired range, the overall quality of the alignment and the model built from it is not ideal. In fact, there is only an approximately 20% certainty in the generated models. In fact, Expasy did not detect any of the identified functions of REn through homologous structures. The identified structures included membrane associating proteins, cell adhesion proteins, and fatty acid multifunctional proteins. In fact, the max sequence identity for detected homologous proteins was 31%; a low number that makes analysis challenging.

**Plasmid Transformation**

Transformation was successful for all the plasmids transformed into BL21-DE3 *E. coli*: pNSB1529, pNSB1530, pNSB1531, and pNSB195 as shown in **Figure 2.4**. DH5-α *E. coli* cells were transformed only with pNSB195 and the transformation was successful.

**Induction**

The inductions were successful based on Immunoblots and gels showing increased band width as a function of growth as shown in **Figure 2.5** and **Figure 2.6**. The timed induction studies for each of the plasmids indicated that the greatest protein production occurred the longer the induction for up to 16 hours. The ideal induction temperature was 16 °C for 16 hours while shaking at ~ 250 rpm.

**Cell Lysis**

Cell lysis was attempted using just the lysis buffer and a vortex with manual agitation with a glass stirrer. However, it was determined that the best lysis technique was to add the lysis buffer and let the solution sit on ice for five minutes and then gently agitate with a glass stirring rod and then to sonicate at the 20% amplitude setting as described in the methods. When properly completed, cell lysis resulted in complete resuspension of the bacterial pellet with no visible lumps. It is essential for protein solubility and stability that the sample remain on ice from pelleting to resuspension or be stored at -80 °C.

**Purification**

Purification was successful when using a cleared, equilibrated, and freshly charged HisTrap. Purification from cleared lysates were successful for plasmids: pNSB1529, pNSB1530, and pNSB1531 with concentrations in the range of 8-14 mg/ml per liter of culture as shown in **Figure 2.13**. In addition, purification of pNSB1529 and pNSB1531 were successful when grown in minimal media with concentrations in the range of 8-14 mg/ml per liter of culture. No capacity issues were encountered when purifying with the HisTrap. To maximize flow-through on the HisTrap it was essential to first clear the lysates by running through a 20 µm filter. In addition, purification occurred after the addition of the second column volume of the elution buffer. For example, the first column volume of elution buffer displaced the binding buffer and filled the HisTrap, and the second column volume of elution buffer displaced and removed the first column volume of the elution buffer. Due to the extremely high concentration of imidazole in the elution buffer (1 M) it was not necessary to apply additional column volumes (CV's) to the HisTrap and elute any remaining bound protein.

Purification was unsuccessful for pNSB195 (REn). When BL21-DE3 and DH5-α cells were retransformed with new pNSB195 plasmid growth was exhibited. However, when cultured, induced, and purified no protein was observed on a gel or via Immunoblot. Further analysis will need to be conducted to determine the cause of this discrepancy as the control for the transformation grew with no problem.

**Immunoblots**

Immunoblots were done for various analyses for all the Rep plasmids: pNSB1529, pNSB1530, and pNSB1531. None were attempted on the REn plasmid pNSB195 due to failure to purify. To validate the PAGE gels, Immunoblots with primary antibodies specific for Rep were used from rabbits. In addition, antibodies for Rep from chicken and yolk were tested and used in conjunction with secondary HRP conjugated enzyme. The rabbit-anti-Rep primary was found to be the most effective and specific, thus it was the primary antibody used in the Immunoblot analyses. No Immunoblot was done on REn as we could not detect any purified protein via stained PAGE gels and $A_{280}$ after 1-liter cultures.

The Immunoblot for the seven hour/overnight induction was successful and indicated that the greatest production of Rep occurred overnight and that the cells were not killed by the toxic nature of Rep being produced in large quantities as shown in **Figure 2.5**. Further, the Immunoblots for the purification of LB media Rep plasmids pNSB1529 and pNSB1531 demonstrated that the purification technique was sufficient, and that the purification process was successful as shown in **Figure 2.13**. The Immunoblots for pNSB1529 and pNSB1531 plasmids that were grown in minimal media indicated slightly less growth than if grown in regular LB media, but still a sufficient quantity at approximately 8 mg/ml per liter of culture. In addition, immunoblots were done for each step of the purification process: initial washes, addition of the

total protein lysate, washes with binding buffer, and several additions of elution buffer. The results indicated that the equilibration of the column using ten CV of binding buffer was sufficient, that washing the column with no more than 8 CV of binding buffer after addition of the protein was optimal, and that the vast majority of the protein was eluted in the second CV of elution buffer. These results informed and influenced the protocol for optimal purification.

**Nuclear Magnetic Resonance**

Natural abundance $^{14}N$ NMR was done on purified pNSB1531 in the elution buffer. The results of the NMR analysis indicated that the protein was likely folded as shown in **Figure 2.7**. This is the first time that a natural abundance NMR has been done on the full-length Rep and its stability in its elution buffer. However, the signal was overwhelmed by the presence of the 1 M imidazole in the elution buffer so to obtain a clearer signal the buffer would need to be exchanged and the high salt and high imidazole buffer replaced with a lower salt no imidazole buffer. $^{14}N$ NMR was not done on purified pNSB1529 or pNSB195 as complications arising from Covid-19 prevented laboratory access to conduct the analysis.

**Mutation Analysis**

The results of the mutation analysis of K145A yielded interesting results as to potential causes of the inhibition of the functioning of Rep. K145 is a conserved lysine residue on the oligomerization domain. Because no currently modeled version of Rep exists that includes the oligomerization domain, the best model for Rep from Phyre2 was used for Rotameric analysis. Using the two highest probability positions post substitution seemed to indicate the N-terminus end of the oligomerization was destabilized by the removal of the carbon chains in going from a lysine to an alanine.

## DISCUSSION

The purpose of this project was twofold: to model the structures of Rep and REN and then to clone, purify, and analyze via NMR various lengths of folded Rep and full length REn. The modeling analysis had somewhat varied success though generally consistent results for Rep while REn had poor success though consistent results. The bench work results were mixed with Rep being successfully cloned and purified while REn was not successfully purified. Further, full length Rep was analyzed using natural abundance ($^{14}$N) NMR to assess its stability while REn was never analyzed with $^{14}$N NMR. Future work that both enhances the modeling techniques employed here and utilizes different structural analysis techniques will be key to finally solving the structures of CbLCV Rep and REn and thereby understanding their functions and binding interactions with each other and other binding partners.

The results from the modeling analysis were generally consistent with low confidence in the generated models though some programs were more accurate and/or confident in their models than others. Much of this confidence, however, is seemingly dependent on the mathematical analytical techniques employed and the rigor of the analysis. For example, Rosetta and Robetta are based on the premise that repeatedly modeling a structure while slightly modifying the inputted parameters each time will have a dramatic effect on the outcome and thus the lowest energy structure can eventually be realized. However, this approach requires extensive computational efforts that are best suited to a server grade cluster; this is especially true when modeling structures *ab initio* and not using a template. On the other hand, I-TASSER employs template/homology-based modelling that once a reasonable structure is determined then analyzes the three-dimensional structures for their potential functions. The computational requirements of I-TASSER are much less than those programs that utilize *ab initio* model construction.

The results for Rep from Rosetta were inconclusive and of low model confidence. From Rosetta, runs of 5,000, 10,000, 15,000, and 20,000 *ab initio* structures were generated. The top five scoring models (lowest energy) from each run were compared to each other in PyMol where they showed little alignment similarity. Oddly enough, the scores for each of the top five models from each run were very close to each other -- though it must be stated that the greater the number of structures generated the greater the number of low energy structures. On the other hand, when using the Robetta server the top five scoring models were also fairly poor with only about 20% confidence in the *ab initio* and comparative modelling derived structures. Thus, it is interesting to note that despite the capabilities of server structure generation the results are not too dissimilar. Further, while Robetta found structures similar in known function to Rep such as 1L2M.pdb and 6Q1M.pdb, the modelling based off of these structures was inconsistent. While it could be argued that this inconsistency points to the independent and relatively unbiased mathematical models of Robetta, it is quite challenging to use for predictive purposes. Despite the hype surrounding Rosetta and Robetta and the incredible flexibility in modelling afforded by both platforms, in the structural analysis of Rep and REn they generally underperformed in comparison to other programs such as Phyre2 and I-TASSER.

The results for REn in Rosetta and Robetta were not quite as dismal as those obtained for Rep. Several low confidence models were developed that had much difference in the three-dimensional overlap of the various structures. This held true for both homology and *ab initio* based modelling techniques in both platforms. Unfortunately, while Robetta was able to identify proteins with similar functions in the PDB such as 1T9I.pdb and 3MXB.pdb the models produced were not consistent structurally. Robetta performed a homology-based analysis of REn and the result was 0% confidence in any the resultant models. While the *ab initio* based Robetta

modeling resulted in a model with 35% confidence. This is interesting considering that Gremlin was the primary method of modeling constraint and detection of sequence conservation. Generally speaking, better modelling results are obtained when the modelling is biased with templates with similar functions and in the case of Robetta and REn the results are quite the opposite. These results stand in stark contrast to the results obtained for the structure of REn obtained via Phyre2.

Phyre2 had quite nice success at predicting the structure and function of Rep through its homology-based modelling. For example, its modeling based on 1L5I.pdb and 6Q1M.pdb (these are structures deposited in the Protein Data Bank (PDB) that have high sequence similarity as selected by Phyre2) had 100% confidence in the model produced. When analyzing the Ramachandran plots there were only two energetically unfavorable residues. This is especially notable given the energetically poor profile produced in Robetta through the Gremlin analysis. Further, when looking at the ProQ2 quality assessment rendered by Phyre2 for Rep, the models produced had overall no bad residue placements and most residues at the highest favorability rating. Finally, when looking at the clash analysis in Phyre2, only two residues in the entire model had a predicted bad placement in the model with all others being generally of the highest quality. It is important to remember, however, that these classifications might change based on the selection criteria of Robetta or I-TASSER. Despite this weakness, the ability of Phyre2 to provide detailed feedback as to the various elements of model quality in concert with the strict cutoff criteria, allow for confidence in the generated models.

With respect to the predicted functions of Rep, Phyre2 was the only program that correctly determined that Rep was an DNA origin binding protein. This is important considering

the fundamental importance of Rep is its ability to nick the stem loop at the origin of the geminivirus genome to initiate replication.

Phyre2 for REn, however, had much less confidence in the resultant model based on 4PC4.pdb at 20%. This was consistent with most analyses that had more confidence in the models predicting the structure of Rep rather than REn. When analyzing the details of the REn and 4PC4.pdb based model, the Ramachandran analysis had numerous unfavorable residue angles that is likely a function of the low sequence identity at 33%. It is likely that the poor modelling of REn is due to its relatively short sequence and the fact that Phyre2 is a homology-based modelling technique and thus relies heavily on sequence similarity to produce a high-quality model.

While I-TASSER was not very successful at predicting high confidence models of either Rep or REn, it did excel at determining the functions of each proteins and finding homologous proteins. These functions were then expressed as GO terms which allows for easy analysis and comparison between homologs. Moreover, I-TASSER provided a very interesting and useful tool that seeks to predictively determine potential ligand binding sites on the proteins. Surprisingly, the tool was quite effective at identifying known ligand binding sites on Rep as well as a few new sites that potentially warrant investigation.

Finally, the results of the analysis through Expasy resulted in low quality models for both Rep and REn. However, despite the low quality of the generated models, they were generally consistent in their confidence level and overall three-dimensional conformation. Interestingly, Expasy correctly identified many of the functions of Rep as a member of the primosome, replication protein, and helicase yet failed to identify any known functions in REn.

As can be seen from the dynamic and varied nature of the results it is crucial to leverage various different programmatic techniques to obtain as diverse and thorough structural information as possible. Some programs were able to generate higher confidence models of Rep and other of REn and without utilizing these various programs this information would not have been gleaned. In addition, by using the various platforms different anecdotal and analytical data could be obtained. For example, information on the theoretical energetic profile of the generated models, theoretical ligand binding domains, and even information on the predicted functions of the modelled proteins based on sequence homologs. Thus, from understanding the general structure of Rep and REn a deeper understanding of its functions may be gleaned and thus inform empirical structural analysis techniques.

Transforming the plasmids for Rep in BL21-DE3 *E. coli* was relatively straightforward and occurred without much problem. It was necessary to first establish desired transformation protocols and once it was determined that standard manufacture guidelines were sufficient that became the standard for which transformations proceeded. The more technically challenging element was to establish protocols for Immunoblots that were effective but did not waste precious primary antibody. The Immunoblots were critical in validating the information that could be visualized in the PAGE gels as well as information that might not be readily visible to the naked eye. For analyses we were able to establish that the optimal primary antibody ratio was 1:5,000 with the rabbit-□-Rep and a 1:10,000 ratio of LiCor 800 CW secondary □-rabbit antibody. For example, we were able to establish that inducing BL21-DE3 overnight did not result in appreciable toxicity to the bacteria and yielded the greatest production of Rep. Furthermore, our Immunoblots were able to ascertain that the optimal number of column washes in the purification process was eight and that more significantly diminished our protein yield

while less washes allowed for the presence of protein contaminants. While Immunoblotting was successful in validating and understanding our empirical data, the most challenging element of this project was to determine how to stabilize Rep for a significant period of time.

Rep is an apparently inherently unstable protein in its non-native environment. When concentrated over 1.5-2 mg/ml, it crashes out within 7 hours. This is highly undesirable for analysis via NMR as it makes analyzing the solution impossible. In addition, having the extremely high imidazole (1 M) in the buffer seems to render the purified Rep highly unstable over long periods of time. Work to analyze the effect of completing a buffer exchange has not yet been attempted though the results are promising. Potential mechanisms to affect the buffer exchange include protein concentrators, dialysis, diafiltration, and size exclusion chromatography. Seemingly the easiest method would be to use protein concentrators so as not to lose protein yield and accomplish buffer exchange in one process. It will need to be determined, however, which method is most amenable to the long-term stability of Rep.

REn has posed a unique challenge in purification. Transformations in both BL21-DE3 and DH5-α were successful with clear and defined colonies. However, when attempting to purify REn from 1 L induced cultures, no protein was eluted. At this point no Immunoblot has been done to determine the presence of REn due to extremely limited quantities of the primary □-REn antibody. It seems, however, that the gels showing the culture, cellular lysate, total protein lysate, and the purified flow-through indicate that REn is being produced but not purified. Likely this is due to REn having a different tag; e.g. GST rather than a 10X histidine tag. This would cause REn to immediately pass through the column and not bind to the $Ni^{2+}$ beads. Further analyses such as restriction enzyme digestion and Immunoblots will need to be done to exactly pinpoint the cause of the error.

Future work that both enhances the modeling techniques employed here and utilizes different structural analysis techniques will be key to finally solving the structures of Rep and REn and thereby understanding their functions and binding interactions with each other and other binding partners. The aims of this project were two-fold: to predictively determine the structures of Rep and REn and to empirically determine the structure of Rep and REn. The results of this project have met with generally positive data that have laid the groundwork for future theoretical and empirical analyses. To understand the structures and functions of Rep and REn will serve to develop capabilities to impinge on their functions and thus to control and eventually eradicate geminiviruses. The humanitarian effects will be countless as people and societies become more stable and the food supply remains available and strengthened.

**Future Work**

Future work will need to be done in terms of modeling Rep and REn as technology develops. Specifically, dynamic molecular simulations will be extremely useful in modeling protein-protein interactions of the binding partners of Rep and REn as well as understanding the binding interaction between Rep and REn. Perhaps understanding the molecular dynamics of the geminiviral proteins could result in an understanding of potential inhibitors that are exclusive to geminiviral proteins. By searching the vast molecular space for all possible configurations of molecules known and theorized and relating to known chemicals and the concomitant interaction with other biological organisms, novel techniques impinging on geminiviral function could be discovered. Testing such discoveries would be relatively easy in a laboratory scenario and could provide interesting mechanistic insights into viral replication and viral protein function.

With respect to actually determining the structures of Rep and REn experimentally, there are several options as it relates to NMR and XRC. To further proceed with NMR analysis, it will be necessary to supplement a buffer that is relatively low salt and very low imidazole with a sodium phosphate base that is conducive to NMR and thereby provides the clearest signal. In addition, centrifugal protein concentrators may be highly useful in sample of purified protein as it provides an easy mechanism of protein concentration and buffer exchange as well as eliminating potentially spurious proteins that are contaminating the protein sample. Another option to replace the buffer with a more ideal buffer includes dialysis, size exclusion chromatography, and diafiltration. Each method has advantages and disadvantages that would need to be fully explored in the context of each particular plasmid. However, likely the most effective mechanisms of buffer exchange and maintenance of protein concentration and therefore yield is to use centrifugal protein concentrators or dialysis.

While NMR has an advantage in that if proteins fail to crystallize then NMR is a good option for proteins less than ~ 35kDa, for larger proteins it is significantly more challenging as well as the complexity and time commitment required to analyze the resulting data, since the masses of Rep and Ren are 39 kDa and 16 kDa, respectively. Thus, XRC is a much more desirable option if at all possible. While crystallization was attempted previously on pNSB1529, pNSB1530, and pNSB1531 there was little success. However, since then, new technologies have emerged as well as potentially new techniques. Future work could attempt to co-crystallize Rep with protein partners such as REn or PCNA. Previous work has attempted to co-crystallize with DNA oligomers; however, different length oligomers could be attempted as well as adjusting the concentration of oligomers.

Finally, another novel approach to gleaning functional information about Rep and REn would be to perform crosslinking experiments and analysis with mass spectrometry. By initiating a pull-down assay of Rep and/or REn in their native environments (*in planta*), the binding partners could be determined. These binding partners could then provide information as to the functions of Rep and REn and thus as to potential uses of motifs. Further, by crosslinking with reversible linkers of varying length and then fractionating, information can be found concerning which amino acids are interacting based on the mass of the modification to reacted residues. Then, a model can be built explaining the amino acid interactions and thus a rudimentary, empirical structure can be determined for the proteins. In addition, by varying the length of the crosslinkers an overall population level understanding of the most common binding partners for Rep and REn can be determined. Finally, by adjusting the length of exposure of the protein to hydrogen exchange using $D_2O$, information can be subtly gathered as to interior and exterior

residues and thus cross-validate the information obtained in other experiments using mass spectrometry analysis.

## REFERENCES


Ach, R. A., Durfee, T., Miller, A. B., Taranto, P., Hanley-Bowdoin, L., Zambryski, P. C., & Gruissem, W. (1997). RRB1 and RRB2 encode maize retinoblastoma-related proteins that interact with a plant D-type cyclin and geminivirus replication protein. *Molecular and Cellular Biology*. https://doi.org/10.1128/mcb.17.9.5077

Allaire, J. J. (2015). RStudio: Integrated development environment for R. *The Journal of Wildlife Management*. https://doi.org/10.1002/jwmg.232

Aluja, M., Birke, A., Ceymann, M., Guillén, L., Arrigoni, E., Baumgartner, D., Pascacio-Villafán, C., & Samietz, J. (2014). Agroecosystem resilience to an invasive insect species that could expand its geographical range in response to global climate change. *Agriculture, Ecosystems and Environment*. https://doi.org/10.1016/j.agee.2014.01.017

Andrews, S. (2010). FastQC. *Babraham Bioinformatics*. https://doi.org/citeulike-article-id:11583827

Argüello-Astorga, G. R., Guevara-González, R. G., Herrera-Estrella, L. R., & Rivera-Bustamante, R. F. (1994). Geminivirus Replication Origins Have a Group-Specific Organization of Iterative Elements: A Model for Replication. *Virology*. https://doi.org/10.1006/viro.1994.1458

Bazinet, A. L., Marshall, K. E., MacMillan, H. A., Williams, C. M., & Sinclair, B. J. (2010). Rapid changes in desiccation resistance in Drosophila melanogaster are facilitated by changes in cuticular permeability. *Journal of Insect Physiology*. https://doi.org/10.1016/j.jinsphys.2010.09.002

Berg, J., Tymoczko, J., & Stryer, L. (2002). Biochemistry, 5th edition. In *Biochemistry*.

Bhagavan, N. V., & Ha, C.-E. (2015). DNA Replication, Repair, and Mutagenesis. In *Essentials of Medical Biochemistry*. https://doi.org/10.1016/b978-0-12-416687-5.00022-1

Bhattacharyya, D., & Chakraborty, S. (2018). Chloroplast: the Trojan horse in plant–virus interaction. In *Molecular Plant Pathology*. https://doi.org/10.1111/mpp.12533

Bhattacharyya, D., Gnanasekaran, P., Kumar, R. K., Kushwaha, N. K., Sharma, V. K., Yusuf, M. A., & Chakraborty, S. (2015). A geminivirus betasatellite damages the structural and functional integrity of chloroplasts leading to symptom formation and inhibition of photosynthesis. *Journal of Experimental Botany*. https://doi.org/10.1093/jxb/erv299

Borah, B. K., Zarreen, F., Baruah, G., & Dasgupta, I. (2016). Insights into the control of geminiviral promoters. In *Virology*. https://doi.org/10.1016/j.virol.2016.04.033

Bottcher, B., Unseld, S., Ceulemans, H., Russell, R. B., & Jeske, H. (2004). Geminate Structures of African Cassava Mosaic Virus. *Journal of Virology*. https://doi.org/10.1128/jvi.78.13.6758-6765.2004

Boulton, M. I., Pallaghy, C. K., Chatani, M., MacFarlane, S., & Davies, J. W. (1993). Replication of maize streak virus mutants in maize protoplasts: Evidence for a movement protein. *Virology*. https://doi.org/10.1006/viro.1993.1010

Briddon, R. W., Bedford, I. D., Tsai, J. H., & Markham, P. G. (1996). Analysis of the nucleotide sequence of the treehopper-transmitted geminivirus, tomato pseudo-curly top virus, suggests a recombinant origin. *Virology*. https://doi.org/10.1006/viro.1996.0264

Briddon, R. W., Patil, B. L., Bagewadi, B., Nawaz-Ul-Rehman, M. S., & Fauquet, C. M. (2010). Distinct evolutionary histories of the DNA-A and DNA-B components of bipartite begomoviruses. *BMC Evolutionary Biology*. https://doi.org/10.1186/1471-2148-10-97

Buchan, D. W. A., & Jones, D. T. (2019). The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkz297

Cantú-Iris, M., Pastor-Palacios, G., Mauricio-Castillo, J. A., Bañuelos-Hernández, B., Avalos-Calleros, J. A., Juárez-Reyes, A., Rivera-Bustamante, R., & Argüello-Astorga, G. R. (2019). Analysis of a new begomovirus unveils a composite element conserved in the CP gene promoters of several Geminiviridae genera: Clues to comprehend the complex regulation of late genes. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0210485

Carey, J. R., Liedo, P., Müller, H. G., Wang, J. L., Senturk, D., & Harshman, L. (2005). Biodemography of a long-lived tephritid: Reproduction and longevity in a large cohort of female Mexican fruit flies, Anastrepha ludens. *Experimental Gerontology*. https://doi.org/10.1016/j.exger.2005.07.013

Castellano, M. M., Sanz-Burgos, A. P., & Gutiérrez, C. (1999). Initiation of DNA replication in a eukaryotic rolling-circle replicon: Identification of multiple DNA-protein complexes at the geminivirus origin. *Journal of Molecular Biology*. https://doi.org/10.1006/jmbi.1999.2916

Castillo, A. G., Collinet, D., Deret, S., Kashoggi, A., & Bejarano, E. R. (2003). Dual interaction of plant PCNA with geminivirus replication accessory protein (REn) and viral replication protein (Rep). *Virology*. https://doi.org/10.1016/S0042-6822(03)00234-4

Chen, H., & Boutros, P. C. (2011). VennDiagram: A package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics*. https://doi.org/10.1186/1471-2105-12-35

Chen, L. F., & Gilbertson, R. L. (2009). Curtovirus-cucurbit interaction: Acquisition host plays a role in leafhopper transmission in a host-dependent manner. *Phytopathology*. https://doi.org/10.1094/PHYTO-99-1-0101

Chowda-Reddy, R. V., Achenjang, F., Felton, C., Etarock, M. T., Anangfac, M. T., Nugent, P., & Fondong, V. N. (2008). Role of a geminivirus AV2 protein putative protein kinase C motif on subcellular localization and pathogenicity. *Virus Research*. https://doi.org/10.1016/j.virusres.2008.02.014

Chung, H., Loehlin, D. W., Dufour, H. D., Vaccarro, K., Millar, J. G., & Carroll, S. B. (2014). A single gene affects both ecological divergence and mate choice in Drosophila. *Science*. https://doi.org/10.1126/science.1249998

Chung, H. Y., & Sunter, G. (2014). Interaction between the transcription factor AtTIFY4B and begomovirus AL2 protein impacts pathogenicity. *Plant Molecular Biology*. https://doi.org/10.1007/s11103-014-0222-9

Clemson, A. S., Sgrò, C. M., & Telonis-Scott, M. (2018). Transcriptional profiles of plasticity

for desiccation stress in Drosophila. *Comparative Biochemistry and Physiology Part - B: Biochemistry and Molecular Biology*. https://doi.org/10.1016/j.cbpb.2017.11.003

Collin, S., Fernández-Lobato, M., Gooding, P. S., Mullineaux, P. M., & Fenoll, C. (1996). The two nonstructural proteins from wheat dwarf virus involved in viral gene expression and replication are retinoblastoma-binding proteins. *Virology*. https://doi.org/10.1006/viro.1996.0256

Davies, J. W., Stanley, J., Donson, J., Mullineaux, P. M., & Boulton, M. I. (1987). Structure and replication of geminivirus genomes. *Journal of Cell Science. Supplement*. https://doi.org/10.1242/jcs.1987.supplement_7.7

De Jager, S. M., & Murray, J. A. H. (1999). Retinoblastoma proteins in plants. *Plant Molecular Biology*. https://doi.org/10.1023/A:1006398232003

Desvoyes, B., De Mendoza, A., Ruiz-Trillo, I., & Gutierrez, C. (2014). Novel roles of plant RETINOBLASTOMA-RELATED (RBR) protein in cell proliferation and asymmetric cell division. In *Journal of Experimental Botany*. https://doi.org/10.1093/jxb/ert411

Dong, X., van Wezel, R., Stanley, J., & Hong, Y. (2003). Functional Characterization of the Nuclear Localization Signal for a Suppressor of Posttranscriptional Gene Silencing. *Journal of Virology*. https://doi.org/10.1128/jvi.77.12.7026-7033.2003

Donson, J., Morris-Krsinich, B. A., Mullineaux, P. M., Boulton, M. I., & Davies, J. W. (1984). A putative primer for second-strand DNA synthesis of maize streak virus is virion-associated. *The EMBO Journal*. https://doi.org/10.1002/j.1460-2075.1984.tb02260.x

Durmuş, S., & Ülgen, K. (2017). Comparative interactomics for virus–human protein–protein interactions: DNA viruses versus RNA viruses. *FEBS Open Bio*. https://doi.org/10.1002/2211-5463.12167

Elmer, J. S., Brand, L., Sunter, G., Gardiner, W. E., Bisaro, D. M., & Rogers, S. G. (1988). Genetic analysis of the tomato golden mosaic virus II. The product of the AL1 coding sequence is required for replication. *Nucleic Acids Research*. https://doi.org/10.1093/nar/16.14.7043

Fauquet, C. M., Briddon, R. W., Brown, J. K., Moriones, E., Stanley, J., Zerbini, M., & Zhou, X. (2008). Geminivirus strain demarcation and nomenclature. *Archives of Virology*. https://doi.org/10.1007/s00705-008-0037-6

Ferveur, J. F. (2005). Cuticular hydrocarbons: Their evolution and roles in Drosophila pheromonal communication. In *Behavior Genetics*. https://doi.org/10.1007/s10519-005-3220-5

Folk, D. G., Han, C., & Bradley, T. J. (2001). Water acquisition and partitioning in Drosophila melanogaster: Effects of selection for desiccation-resistance. *Journal of Experimental Biology*.

Fondong, V. N., Reddy, R. V. C., Lu, C., Hankoua, B., Felton, C., Czymmek, K., & Achenjang, F. (2007). The consensus N-myristoylation motif of a geminivirus AC4 protein is required for membrane binding and pathogenicity. *Molecular Plant-Microbe Interactions*. https://doi.org/10.1094/MPMI-20-4-0380

Fontes, E. P. B., Luckow, V. A., & Hanley-Bowdoin, L. (1992). A geminivirus replication protein is a sequence-specific DNA binding protein. *Plant Cell*. https://doi.org/10.1105/tpc.4.5.597

Formosa, T., & Alberts, B. M. (1986). DNA synthesis dependent on genetic recombination: Characterization of a reaction catalyzed by purified bacteriophage T4 proteins. *Cell*. https://doi.org/10.1016/0092-8674(86)90522-2

Forster, S. C., Finkel, A. M., Gould, J. A., & Hertzog, P. J. (2013). RNA-eXpress annotates novel transcript features in RNA-seq data. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btt034

*FRUIT FLY EXCLUSION AND DETECTION STRATEGIC PLAN FY 2019-2023*. (2019). https://www.aphis.usda.gov/plant_health/plant_pest_info/fruit_flies/downloads/feed-strategic-plan-en.pdf

Gai, Y., Liu, Z., Cervantes-Sandoval, I., & Davis, R. L. (2016). Drosophila SLC22A Transporter Is a Memory Suppressor Gene that Influences Cholinergic Neurotransmission to the Mushroom Bodies. *Neuron*. https://doi.org/10.1016/j.neuron.2016.03.017

Ge, S. X., Son, E. W., & Yao, R. (2018). iDEP: An integrated web application for differential expression and pathway analysis of RNA-Seq data. *BMC Bioinformatics*. https://doi.org/10.1186/s12859-018-2486-6

Gene, T., & Consortium, O. (2000). Gene Ontology : tool for the. *Gene Expression*. https://doi.org/10.1038/75556

Glick, E., Zrachya, A., Levy, Y., Mett, A., Gidoni, D., Belausov, E., Citovsky, V., & Gafni, Y. (2008). Interaction with host SGS3 is required for suppression of RNA silencing by tomato yellow leaf curl virus V2 protein. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.0709036105

Gnanasekaran, P., Ponnusamy, K., & Chakraborty, S. (2019). A geminivirus betasatellite encoded βC1 protein interacts with PsbP and subverts PsbP-mediated antiviral defence in plants. *Molecular Plant Pathology*. https://doi.org/10.1111/mpp.12804

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., Di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*. https://doi.org/10.1038/nbt.1883

Gronenborn, B. (2004). Nanoviruses: Genome organisation and protein function. *Veterinary Microbiology*. https://doi.org/10.1016/j.vetmic.2003.10.015

Gutierrez, C. (1999). Geminivirus DNA replication. In *Cellular and Molecular Life Sciences*. https://doi.org/10.1007/s000180050433

Gutierrez, Crisanto, Ramirez-Parra, E., Castellano, M. M., Sanz-Burgos, A. P., Luque, A., & Missich, R. (2004). Geminivirus DNA replication and cell cycle interactions. *Veterinary Microbiology*. https://doi.org/10.1016/j.vetmic.2003.10.012

Hallan, V., & Gafni, Y. (2001). Tomato yellow leaf curl virus (TYLCV) capsid protein (CP) subunit interactions: Implications for viral assembly. *Archives of Virology*. https://doi.org/10.1007/s007050170062

Hanley-Bowdoin, L., Settlage, S. B., Orozco, B. M., Nagar, S., & Robertson, D. (2000). Geminiviruses: Models for plant DNA replication, transcription, and cell cycle regulation. In *Critical Reviews in Biochemistry and Molecular Biology*. https://doi.org/10.1080/07352689991309162

Hanley-Bowdoin, L., Settlage, S. B., & Robertson, D. (2004). Reprogramming plant gene expression: A prerequisite to geminivirus DNA replication. In *Molecular Plant Pathology*. https://doi.org/10.1111/j.1364-3703.2004.00214.x

Hartitz, M. D., Sunter, G., & Bisaro, D. M. (1999). The tomato golden mosaic virus transactivator (TrAP) is a single- stranded DNA and zinc-binding phosphoprotein with an acidic activation domain. *Virology*. https://doi.org/10.1006/viro.1999.9925

Höfer, P., Bedford, I. D., Markham, P. G., Jeske, H., & Frischmuth, T. (1997). Coat protein gene replacement results in whitefly transmission of an insect nontransmissible geminivirus isolate. *Virology*. https://doi.org/10.1006/viro.1997.8751

Hoffmann, A. A. (2010). Physiological climatic limits in Drosophila: Patterns and implications. *Journal of Experimental Biology*. https://doi.org/10.1242/jeb.037630

HOFFMANN, A. A., & PARSONS, P. A. (1989). An integrated approach to environmental stress tolerance and life-history variation: desiccation tolerance in Drosophila. *Biological Journal of the Linnean Society*. https://doi.org/10.1111/j.1095-8312.1989.tb02098.x

Hoffmann, Ary A., & Harshman, L. G. (1999). Desiccation and starvation resistance in Drosophila: Patterns of variation at the species, population and intrapopulation levels. In *Heredity*. https://doi.org/10.1046/j.1365-2540.1999.00649.x

Hu, T., Huang, C., He, Y., Castillo-Gonzalez, C., Gui, X., Wang, Y., Zhang, X., & Zhou, X. (2019). βc1 protein encoded in geminivirus satellite concertedly targets MKK2 and MPK4 to counter host defense. *PLoS Pathogens*. https://doi.org/10.1371/journal.ppat.1007728

Jeske, H., Lütgemeier, M., & Preiß, W. (2001). DNA forms indicate rolling circle and recombination-dependent replication of Abutilon mosaic virus. *EMBO Journal*. https://doi.org/10.1093/emboj/20.21.6158

Joan L. Slonczewski, John W. Foster, K. M. (2017). Microbiology: An Evolving Science (2nd Edition). *IEEE Transactions on Wireless Communications*. https://doi.org/10.1109/twc.2017.2771181

Jose, J., & Usha, R. (2003). Bhendi yellow vein mosaic disease in india is caused by association of a DNA β satellite with a begomovirus. *Virology*. https://doi.org/10.1006/viro.2002.1768

Jupin, I., De Kouchkovsky, F., Jouanneau, F., & Gronenborn, B. (1994). Movement of tomato yellow leaf curl geminivirus (TYLCV): Involvement of the protein encoded by ORF C4. *Virology*. https://doi.org/10.1006/viro.1994.1512

Kahsai, L., Kapan, N., Dircksen, H., Winther, Å. M. E., & Nässel, D. R. (2010). Metabolic stress

responses in Drosophila are modulated by brain neurosecretory cells that produce multiple neuropeptides. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0011480

Kang, L., Aggarwal, D. D., Rashkovetsky, E., Korol, A. B., & Michalak, P. (2016). Rapid genomic changes in Drosophila melanogaster adapting to desiccation stress in an experimental evolution system. *BMC Genomics*. https://doi.org/10.1186/s12864-016-2556-y

Kawano, T., Shimoda, M., Matsumoto, H., Ryuda, M., Tsuzuki, S., & Hayakawa, Y. (2010). Identification of a gene, Desiccate, contributing to desiccation resistance in Drosophila melanogaster. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M110.168864

Kellermann, V., Van Heerwaarden, B., Sgrò, C. M., & Hoffmann, A. A. (2009). Fundamental evolutionary limits in ecological traits drive drosophila species distributions. *Science*. https://doi.org/10.1126/science.1175443

Kelman, Z. (1997). PCNA: Structure, functions and interactions. In *Oncogene*. https://doi.org/10.1038/sj.onc.1200886

Kong, L.-J. (2000). A geminivirus replication protein interacts with the retinoblastoma protein through a novel domain to determine symptoms and tissue specificity of infection in plants. *The EMBO Journal*. https://doi.org/10.1093/emboj/19.13.3485

Kong, L. J., & Hanley-Bowdoin, L. (2002). A geminivirus replication protein interacts with a protein kinase and a motor protein that display different expression patterns during plant development and infection. *Plant Cell*. https://doi.org/10.1105/tpc.003681

Krupovic, M., Ravantti, J. J., & Bamford, D. H. (2009). Geminiviruses: A tale of a plasmid becoming a virus. *BMC Evolutionary Biology*. https://doi.org/10.1186/1471-2148-9-112

Kunik, T., Palanichelvam, K., Czosnek, H., Citovsky, V., & Gafni, Y. (1998). Nuclear import of the capsid protein of tomato yellow leaf curl virus (TYLCV) in plant and insect cells. *Plant Journal*. https://doi.org/10.1046/j.1365-313X.1998.00037.x

Lacatus, G., & Sunter, G. (2009). The Arabidopsis PEAPOD2 transcription factor interacts with geminivirus AL2 protein and the coat protein promoter. *Virology*. https://doi.org/10.1016/j.virol.2009.07.004

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*. https://doi.org/10.1186/gb-2009-10-3-r25

Lazarowitz, S. G. (1992). Geminiviruses: Genome structure and gene function. *Critical Reviews in Plant Sciences*. https://doi.org/10.1080/07352689209382350

Legg, J. P., & Fauquet, C. M. (2004). Cassava mosaic geminiviruses in Africa. In *Plant Molecular Biology*. https://doi.org/10.1007/s11103-004-1651-7

Linford, N. J., Bilgir, C., Ro, J., & Pletcher, S. D. (2013). Measurement of lifespan in Drosophila melanogaster. *Journal of Visualized Experiments*. https://doi.org/10.3791/50068

Liu, L., Saunders, K., Thomas, C. L., Davies, J. W., & Stanley, J. (1999). Bean yellow dwarf virus RepA, but not Rep, binds to maize retinoblastoma protein and the virus tolerates

mutations in the consensus binding motif. *Virology*. https://doi.org/10.1006/viro.1999.9616

Lozano-Durán, R., Rosas-Díaz, T., Gusmaroli, G., Luna, A. P., Taconnat, L., Deng, X. W., & Bejarano, E. R. (2011). Geminiviruses subvert ubiquitination by altering CSN-mediated derubylation of SCF E3 ligase complexes and inhibit jasmonate signaling in Arabidopsis thaliana. *Plant Cell*. https://doi.org/10.1105/tpc.110.080267

Mansoor, S., Briddon, R. W., Zafar, Y., & Stanley, J. (2003). Geminivirus disease complexes: An emerging threat. In *Trends in Plant Science*. https://doi.org/10.1016/S1360-1385(03)00007-4

Mansoor, S., Zafar, Y., & Briddon, R. W. (2006). Geminivirus disease complexes: the threat is spreading. In *Trends in Plant Science*. https://doi.org/10.1016/j.tplants.2006.03.003

Markham, P. G., Bedford, I. D., Liu, S., & Pinner, M. S. (1994). The transmission of geminiviruses by Bemisia tabaci. *Pesticide Science*. https://doi.org/10.1002/ps.2780420209

Matzkin, L. M., & Markow, T. A. (2009). Transcriptional regulation of metabolism associated with the increased desiccation resistance of the cactophilic Drosophila mojavensis. *Genetics*. https://doi.org/10.1534/genetics.109.104927

Milward, J. H. (2018). Geminivirus: Rep as a Target for Conferring Resistance, the Host's Strigolactone Hormonal Response to Infection, and Host Resistance. *Master's Thesis*. https://repository.lib.ncsu.edu/bitstream/handle/1840.20/35671/etd.pdf?sequence=1&isAllowed=y

Nash, T. (2010). *Further Characterization of the Rep Protein and Using Peptide Aptamers as a Broad-based Resistance Strategy to Combat Geminivirus Disease.* [North Carolina State University]. https://repository.lib.ncsu.edu/bitstream/handle/1840.16/6423/etd.pdf?sequence=1&isAllowed=y

Navas-Castillo, J., Fiallo-Olivé, E., & Sánchez-Campos, S. (2011). Emerging Virus Diseases Transmitted by Whiteflies. *Annual Review of Phytopathology*. https://doi.org/10.1146/annurev-phyto-072910-095235

Núñez, E. D., & Aiello, A. (2013). Leafhoppers (Homoptera: Cicadellidae) that probe human skin: A review of the world literature and nineteen new records, from Panama. *Terrestrial Arthropod Reviews*. https://doi.org/10.1163/18749836-06001064

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., & Kanehisa, M. (1999). KEGG: Kyoto encyclopedia of genes and genomes. In *Nucleic Acids Research*. https://doi.org/10.1093/nar/27.1.29

Ong, C. T., & Corces, V. G. (2011). Enhancer function: New insights into the regulation of tissue-specific gene expression. In *Nature Reviews Genetics*. https://doi.org/10.1038/nrg2957

Orozco, B. M., & Hanley-Bowdoin, L. (1996). A DNA structure is required for geminivirus replication origin function. *Journal of Virology*. https://doi.org/10.1128/jvi.70.1.148-158.1996

Pasumarthy, K. K., Choudhury, N. R., & Mukherjee, S. K. (2010). Tomato leaf curl Kerala virus (ToLCKeV) AC3 protein forms a higher order oligomer and enhances ATPase activity of replication initiator protein (Rep/AC1). *Virology Journal*. https://doi.org/10.1186/1743-422X-7-128

Pavlov, Y. I., Maki, S., Maki, H., & Kunkel, T. A. (2004). Evidence for interplay among yeast replicative DNA polymerases alpha, delta and epsilon from studies of exonuclease and polymerase active site mutations. *BMC Biology*. https://doi.org/10.1186/1741-7007-2-11

Pedersen, T. J., & Hanley-Bowdoin, L. (1994). Molecular Characterization of the AL3 Protein Encoded by a Bipartite Geminivirus. *Virology*. https://doi.org/10.1006/viro.1994.1442

Petty, I. T. D., Carter, S. C., Morra, M. R., Jeffrey, J. L., & Olivey, H. E. (2000). Bipartite geminivirus host adaptation determined cooperatively by coding and noncoding sequences of the genome. *Virology*. https://doi.org/10.1006/viro.2000.0620

Philip, B. N., Yi, S. X., Elnitsky, M. A., & Lee, R. E. (2008). Aquaporins play a role in desiccation and freeze tolerance in larvae of the goldenrod gall fly, Eurosta solidaginis. *Journal of Experimental Biology*. https://doi.org/10.1242/jeb.016758

Pradhan, B., Tien, V. Van, & Dey, N. (2017). *Mol Biology of Virus Replication*.

Qin, S., Ward, B. M., & Lazarowitz, S. G. (1998). The Bipartite Geminivirus Coat Protein Aids BR1 Function in Viral Movement by Affecting the Accumulation of Viral Single-Stranded DNA. *Journal of Virology*. https://doi.org/10.1128/jvi.72.11.9247-9256.1998

Rizvi, I., Choudhury, N. R., & Tuteja, N. (2014). Insights into the functional characteristics of geminivirus rolling-circle replication initiator protein and its interaction with host factors affecting viral DNA replication. In *Archives of Virology*. https://doi.org/10.1007/s00705-014-2297-7

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2009). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btp616

Rojas, M. R., Jiang, H., Salati, R., Xoconostle-Cázares, B., Sudarshana, M. R., Lucas, W. J., & Gilbertson, R. L. (2001). Functional analysis of proteins involved in movement of the monopartite begomovirus, Tomato yellow leaf curl virus. *Virology*. https://doi.org/10.1006/viro.2001.1194

Roshan, P., Kulshreshtha, A., Kumar, S., Purohit, R., & Hallan, V. (2018). AV2 protein of tomato leaf curl Palampur virus promotes systemic necrosis in Nicotiana benthamiana and interacts with host Catalase2. *Scientific Reports*. https://doi.org/10.1038/s41598-018-19292-3

Ruhel, R., & Chakraborty, S. (2019). Multifunctional roles of geminivirus encoded replication initiator protein. *VirusDisease*. https://doi.org/10.1007/s13337-018-0458-0

Sanderfoot, A. A., Ingham, D. J., & Lazarowitz, S. C. (1992). A Viral Movement Protein as a Nuclear Shuttle. *Plant Physiology*.

Sanderfoot, A. A., Ingham, D. J., & Lazarowitz, S. G. (1996). A viral movement protein as a

nuclear shuttle: The geminivirus BR1 movement protein contains domains essential for interaction with BL1 and nuclear localization. *Plant Physiology*. https://doi.org/10.1104/pp.110.1.23

Saunders, K., Lucy, A., & Stanley, J. (1991). DNA forms of the geminivirus African cassava mosaic virus consistent with a rolling circle mechanism of replication. *Nucleic Acids Research*. https://doi.org/10.1093/nar/19.9.2325

Saunders, K., Lucy, A., & Stanley, J. (1992). RNA-primed complementary-sense DNA synthesis of the geminivirus African cassava mosaic virus. *Nucleic Acids Research*. https://doi.org/10.1093/nar/20.23.6311

Saunders, K., Norman, A., Gucciardo, S., & Stanley, J. (2004). The DNA β satellite component associated with ageratum yellow vein disease encodes an essential pathogenicity protein (βC1). *Virology*. https://doi.org/10.1016/j.virol.2004.03.018

Selth, L. A., Dogra, S. C., Rasheed, M. S., Healy, H., Randles, J. W., & Rezaian, M. A. (2005). A NAC domain protein interacts with Tomato leaf curl virus replication accessory protein and enhances viral replication. *Plant Cell*. https://doi.org/10.1105/tpc.104.027235

Settlage, S. B., See, R. G., & Hanley-Bowdoin, L. (2005). Geminivirus C3 Protein: Replication Enhancement and Protein Interactions. *Journal of Virology*. https://doi.org/10.1128/jvi.79.15.9885-9895.2005

Settlage, Sharon B., Miller, A. B., Gruissem, W., & Hanley-Bowdoin, L. (2001). Dual interaction of a geminivirus replication accessory factor with a viral replication protein and a plant cell cycle regulator. *Virology*. https://doi.org/10.1006/viro.2000.0719

Sharma, V., Kohli, S., & Brahmachari, V. (2017). Correlation between desiccation stress response and epigenetic modifications of genes in Drosophila melanogaster: An example of environment-epigenome interaction. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*. https://doi.org/10.1016/j.bbagrm.2017.08.001

Shen, W., Dallas, M. B., Goshe, M. B., & Hanley-Bowdoin, L. (2014). SnRK1 Phosphorylation of AL2 Delays Cabbage Leaf Curl Virus Infection in Arabidopsis. *Journal of Virology*. https://doi.org/10.1128/jvi.00761-14

Shen, Wei, & Hanley-Bowdoin, L. (2006). Geminivirus infection up-regulates the expression of two arabidopsis protein kinases related to yeast SNF1- and mammalian AMPK-activating kinases. *Plant Physiology*. https://doi.org/10.1104/pp.106.088476

Shepherd, D. N., Martin, D. P., Van Der Walt, E., Dent, K., Varsani, A., & Rybicki, E. P. (2010). Maize streak virus: An old and complex "emerging" pathogen. *Molecular Plant Pathology*. https://doi.org/10.1111/j.1364-3703.2009.00568.x

Sinclair, B. J., Gibbs, A. G., & Roberts, S. P. (2007). Gene transcription during exposure to, and recovery from, cold and desiccation stress in Drosophila melanogaster. *Insect Molecular Biology*. https://doi.org/10.1111/j.1365-2583.2007.00739.x

Singh, D. K., Islam, M. N., Choudhury, N. R., Karjee, S., & Mukherjee, S. K. (2007). The 32 kDa subunit of replication protein A (RPA) participates in the DNA replication of Mung bean yellow mosaic India virus (MYMIV) by interacting with the viral Rep protein. *Nucleic*

*Acids Research*. https://doi.org/10.1093/nar/gkl1088

Smyth, G. K. (2005). limma: Linear Models for Microarray Data. In *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. https://doi.org/10.1007/0-387-29362-0_23

Stanley, J. (1991). The molecular determinants of geminivirus pathogenesis. *Seminars in Virology*.

Stanley, John. (1993). Geminiviruses: plant viral vectors. *Current Opinion in Genetics and Development*. https://doi.org/10.1016/S0959-437X(05)80347-8

Stenger, D. C., Revington, G. N., Stevenson, M. C., & Bisaro, D. M. (1991). Replicational release of geminivirus genomes from tandemly repeated copies: Evidence for rolling-circle replication of a plant viral DNA. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.88.18.8029

Stinziano, J. R., Sové, R. J., Rundle, H. D., & Sinclair, B. J. (2015). Rapid desiccation hardening changes the cuticular hydrocarbon profile of drosophila melanogaster. *Comparative Biochemistry and Physiology -Part A : Molecular and Integrative Physiology*. https://doi.org/10.1016/j.cbpa.2014.11.004

Stone, A. (1942). Fruitflies of the genus anastrepha. In *Nature*.

Sun, M., Jiang, K., Li, C., Du, J., Li, M., Ghanem, H., Wu, G., & Qing, L. (2020). Tobacco curly shoot virus C3 protein enhances viral replication and gene expression in Nicotiana benthamiana plants. *Virus Research*. https://doi.org/10.1016/j.virusres.2020.197939

Sung, Y. K., & Coutts, R. H. A. (1995). Mutational analysis of potato yellow mosaic geminivirus. *Journal of General Virology*. https://doi.org/10.1099/0022-1317-76-7-1773

Sunter, G., & Bisaro, D. M. (1991). Transactivation in a geminivirus: AL2 gene product is needed for coat protein expression. *Virology*. https://doi.org/10.1016/0042-6822(91)90049-H

Sunter, G., & Bisaro, D. M. (1997). Regulation of a geminivirus coat protein promoter by AL2 protein (TrAP): Evidence for activation and derepression mechanisms. *Virology*. https://doi.org/10.1006/viro.1997.8549

Sunter, G., & Bisaro, D. M. (2003). Identification of a minimal sequence required for activation of the tomato golden mosaic virus coat protein promoter in protoplasts. *Virology*. https://doi.org/10.1006/viro.2002.1757

Sunter, G., Hartitz, M. D., Hormuzdi, S. G., Brough, C. L., & Bisaro, D. M. (1990). Genetic analysis of tomato golden mosaic virus: ORF AL2 is required for coat protein accumulation while ORF AL3 is necessary for efficient DNA replication. *Virology*. https://doi.org/10.1016/0042-6822(90)90275-V

Sunter, G., Sunter, J. L., & Bisaro, D. M. (2001). Plants expressing tomato golden mosaic virus AL2 or beet curly top virus L2 transgenes show enhanced susceptibility to infection by DNA and RNA viruses. *Virology*. https://doi.org/10.1006/viro.2001.0950

Swanson, M. M., & Harrison, B. D. (1993). Serological relationships and epitope profiles of isolates of okra leaf curl geminivirus from Africa and the Middle East. *Biochimie*. https://doi.org/10.1016/0300-9084(93)90101-W

Telonis-Scott, M., Guthridge, K. M., & Hoffmann, A. A. (2006). A new set of laboratory-selected Drosophila melanogaster lines for the analysis of desiccation resistance: Response to selection, physiology and correlated responses. *Journal of Experimental Biology*. https://doi.org/10.1242/jeb.02201

Terhzaz, S., Alford, L., Yeoh, J. G. C., Marley, R., Dornan, A. J., Dow, J. A. T., & Davies, S. A. (2018). Renal neuroendocrine control of desiccation and cold tolerance by Drosophila suzukii. *Pest Management Science*. https://doi.org/10.1002/ps.4663

Thorat, L. J., Gaikwad, S. M., & Nath, B. B. (2012). Trehalose as an indicator of desiccation stress in Drosophila melanogaster larvae: A potential marker of anhydrobiosis. *Biochemical and Biophysical Research Communications*. https://doi.org/10.1016/j.bbrc.2012.02.065

Unseld, S., Frischmuth, T., & Jeske, H. (2004). Short deletions in nuclear targeting sequences of African cassava mosaic virus coat protein prevent geminivirus twinned particle formation. *Virology*. https://doi.org/10.1016/j.virol.2003.09.003

Vinoth Kumar, R., Singh, A. K., Singh, A. K., Yadav, T., Singh, A. K., Kushwaha, N., Chattopadhyay, B., & Chakraborty, S. (2015). Complexity of begomovirus and betasatellite populations associated with chilli leaf curl disease in India. *Journal of General Virology*. https://doi.org/10.1099/jgv.0.000254

Wang, B., Li, F., Huang, C., Yang, X., Qian, Y., Xie, Y., & Zhou, X. (2014). V2 of tomato yellow leaf curl virus can suppress methylation-mediated transcriptional gene silencing in plants. *Journal of General Virology*. https://doi.org/10.1099/vir.0.055798-0

Ward, B. M., & Lazarowitz, S. G. (1999). Nuclear export in plants: Use of geminivirus movement proteins for a cell-based export assay. *Plant Cell*. https://doi.org/10.1105/tpc.11.7.1267

Yang, X., Baliji, S., Buchmann, R. C., Wang, H., Lindbo, J. A., Sunter, G., & Bisaro, D. M. (2007). Functional Modulation of the Geminivirus AL2 Transcription Factor and Silencing Suppressor by Self-Interaction. *Journal of Virology*. https://doi.org/10.1128/jvi.00617-07

Yu, G. (2018). clusterProfiler: universal enrichment tool for functional and comparative study. *BioRxiv*. https://doi.org/10.1101/256784

Yu, G., Wang, L. G., Han, Y., & He, Q. Y. (2012). ClusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS A Journal of Integrative Biology*. https://doi.org/10.1089/omi.2011.0118

Zacksenhaus, E., Jiang, Z., Phillips, R. A., & Gallie, B. L. (1996). Dual mechanisms of repression of E2F1 activity by the retinoblastoma gene product. *The EMBO Journal*. https://doi.org/10.1002/j.1460-2075.1996.tb00978.x

Zerbini, F. M., Briddon, R. W., Idris, A., Martin, D. P., Moriones, E., Navas-Castillo, J., Rivera-Bustamante, R., Roumagnac, P., & Varsani, A. (2017). ICTV virus taxonomy profile: Geminiviridae. *Journal of General Virology*. https://doi.org/10.1099/jgv.0.000738

Zhang, W., Olson, N. H., Baker, T. S., Faulkner, L., Agbandje-McKenna, M., Boulton, M. I., Davies, J. W., & McKenna, R. (2001). Structure of the maize streak virus geminate particle. *Virology*. https://doi.org/10.1006/viro.2000.0739

Zhou, Y., Rojas, M. R., Park, M.-R., Seo, Y.-S., Lucas, W. J., & Gilbertson, R. L. (2011). Histone H3 Interacts and Colocalizes with the Nuclear Shuttle Protein and the Movement Protein of a Geminivirus. *Journal of Virology*. https://doi.org/10.1128/jvi.00082-11

## Tables

**Table 2.0:** Table depicts the amino acids sequences of CbLCV Rep and CbLCV REn.

| Rep | REn |
|---|---|
| MPRNPKSFRLAARNIFLTYPQCDIPKDE<br>ALQMLQTLSWSVVKPTYIRVAREEHSD<br>GFPHLHCLIQLSGKSNIKDARFFDITHPR<br>RSANFHPNIQAAKDTNAVKNYITKDGD<br>YCESGQYKVSGGTKANKDDVYHNAV<br>NAGCVEEALAIIRAGDPKTFIVSYHNVR<br>ANIERLFTKAPEPWAPPFQLSSFTNVPD<br>EMSSWADDYFGRSAAARAERPISIIVEG<br>DSRTGKTMWARALGPHNYLSGHLDFN<br>SKVFSNNAEYNVIDDIAPHYLKLKHWK<br>ELIGAQRDWQSNCKYGKPVQIKGGIPSI<br>VLCNPGEGSSYISFLNKEENASLRAWTT<br>KNAKFITLEAPLYQSTAQDC | MDSRTGESITVAQAENSVFIW<br>EVPNPLYFRIQRVEDPLYTRTR<br>IYHIQVRFNHNLRKALDLRKA<br>YFNFQVWTTSMRASGPTYLS<br>RFKCLVMSHLDNLGVIGINHV<br>IRAVRFATDRSYVTHVHENHV<br>INFKIY |

**Table 2.1:** Sequence overlap of various different geminiviral Rep and REn proteins, respectively, using Clustal Omega. In panel A, four different variants of Rep are aligned: TGMVY (UniProtKB - P03567), TYLCI (UniProtKB - P27259), CALCV (UniProtKB - Q3LTQ7), and WDVS (UniProtKB - Q67622). In panel B, three variants of REn are aligned: CALCV (UniProtKB - Q96702), TYLCI (UniProtKB - P27265), and TGMVY (UniProtKB - P03563).

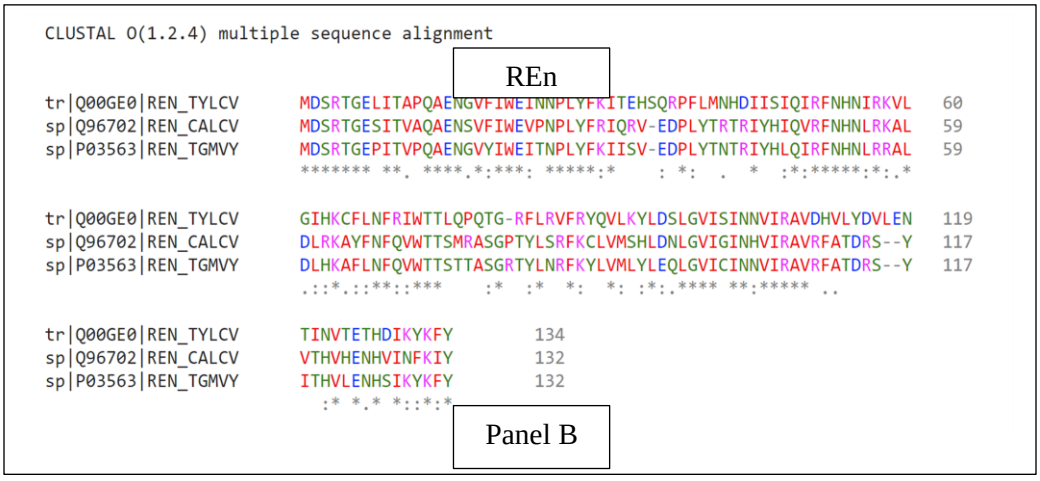



**Table 2.2:** Summary of modelling techniques for Rep and REn. If no functional information is given, then the program did not provide functional information.

**Rep**

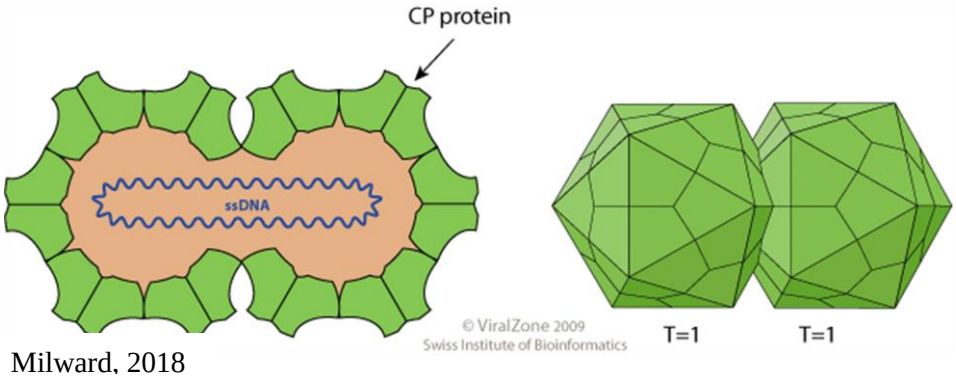
| Program | Modeling | Results |
|---|---|---|
| Rosetta | *de novo* | Poor Structural prediction |
| Robetta | *de novo* | Poor Structural prediction |
| Robetta | Homology | Poor structural prediction |
| Phyre2 | Homology | Best model with good functional prediction |
| I-TASSER | Homology | Worst structural prediction |
| Expasy | Homology | Good model with excellent functional prediction |

**REn**

| Program | Modeling | Results |
|---|---|---|
| Rosetta | *de novo* | Poor Structural prediction |
| Robetta | *de novo* | Best model with excellent functional prediction |
| Robetta | Homology | Poor structural prediction |
| Phyre2 | Homology | Worst structural prediction. |
| I-TASSER | Homology | Poor model with excellent functional prediction |
| Expasy | Homology | Subpar model with poor functional |

## Figures

**Figure 2.0:** Depiction of geminivirus icosahedral capsid with ssDNA distributed inside the capsid.




Milward, 2018

**Figure 2.1:** Flowchart of the methodological approach to solving the Structures of Rep and REn. The process applies to each plasmid.

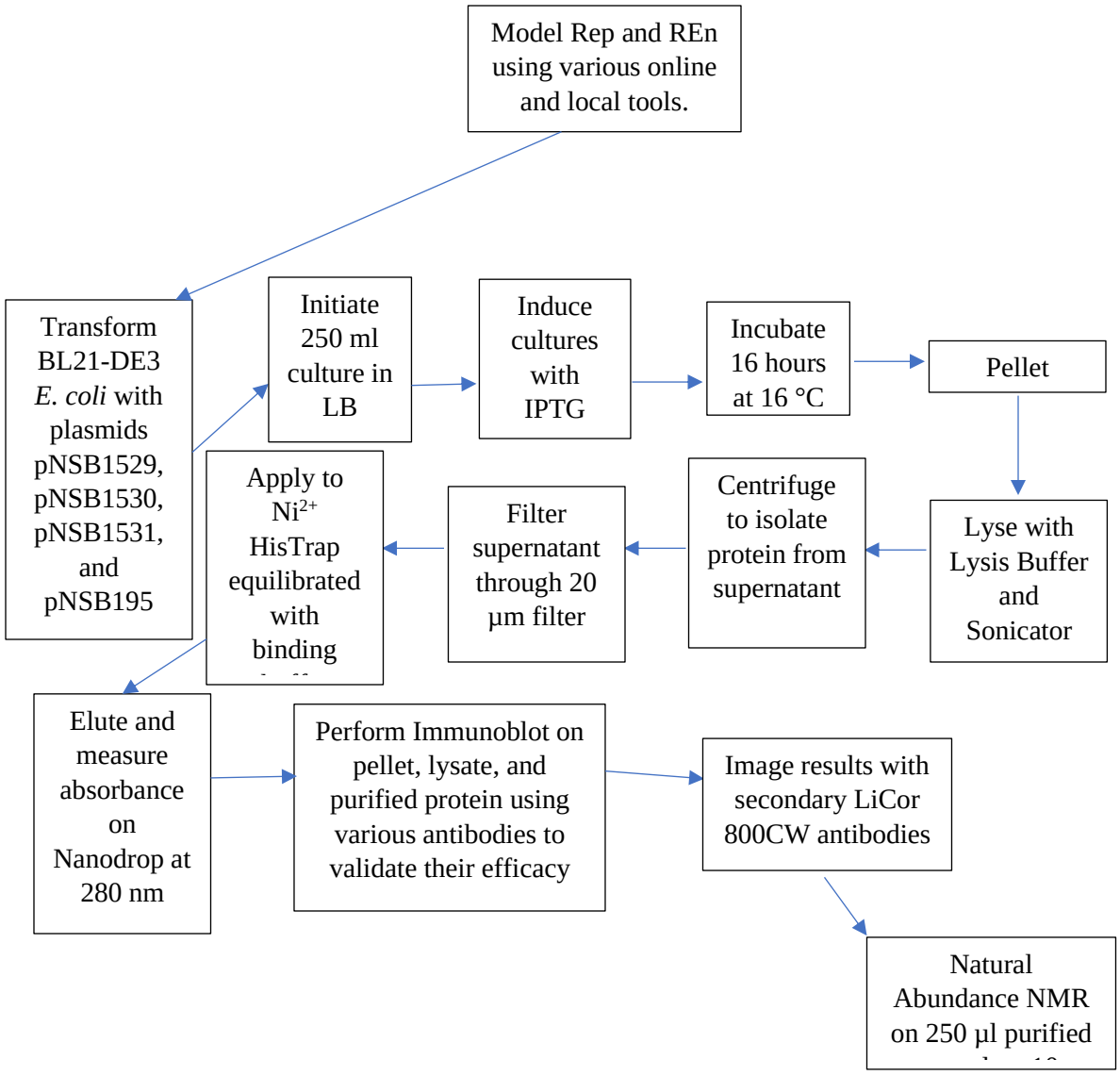


**Figure 2.2:** Artists representation of Rolling Circle Replication (RCR). ssDNA is converted to the dsDNA intermediate for replication. Rep is bound to the nonanucleotide consensus sequence. The stem loop is cut by Rep and viral DNA replisome is formed at the nick creating the replication fork.

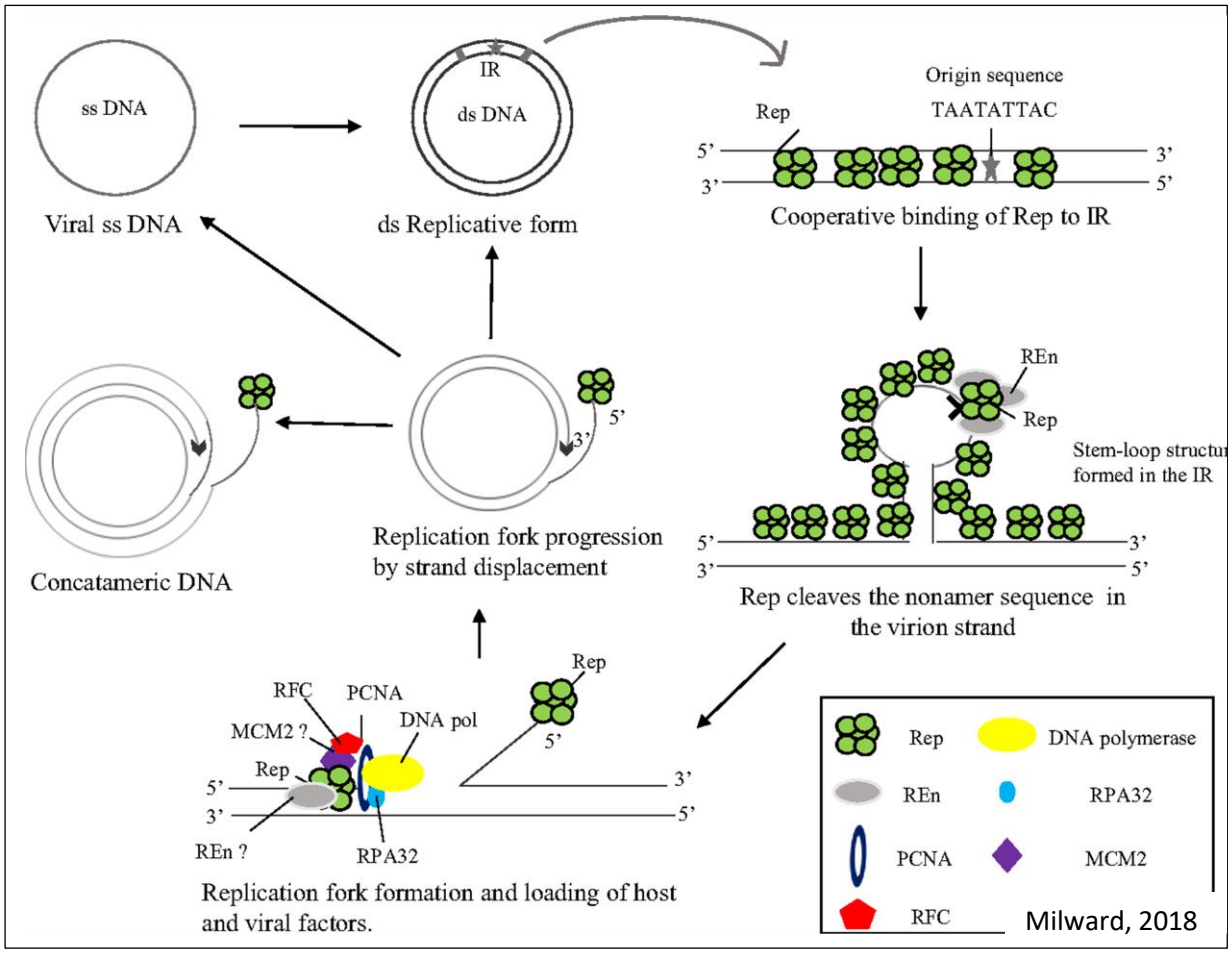


**Figure 2.3:** Graphic illustrating known Rep binding domains. Note that most of the known binding domains are on the N-terminus while the last approximately 45% of Rep is dedicated to the Helicase/ATPase domain.

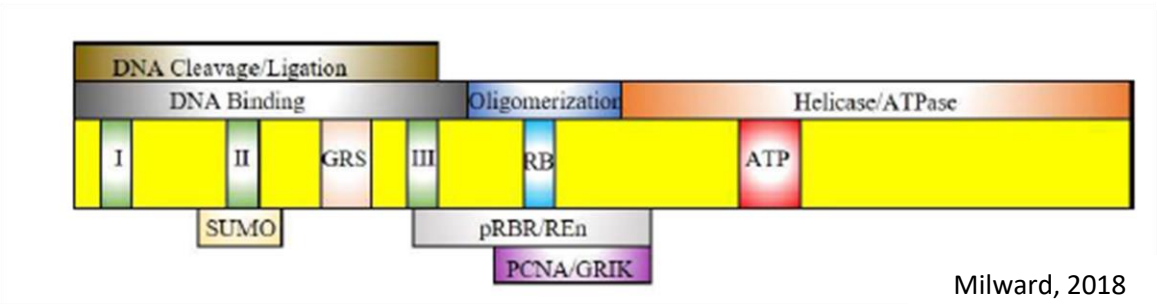


**Figure 2.4:** LB-agar selection plate of pNSB195 transformed BL21-DE3 *E. coli* in LB infused Carbenicillin.

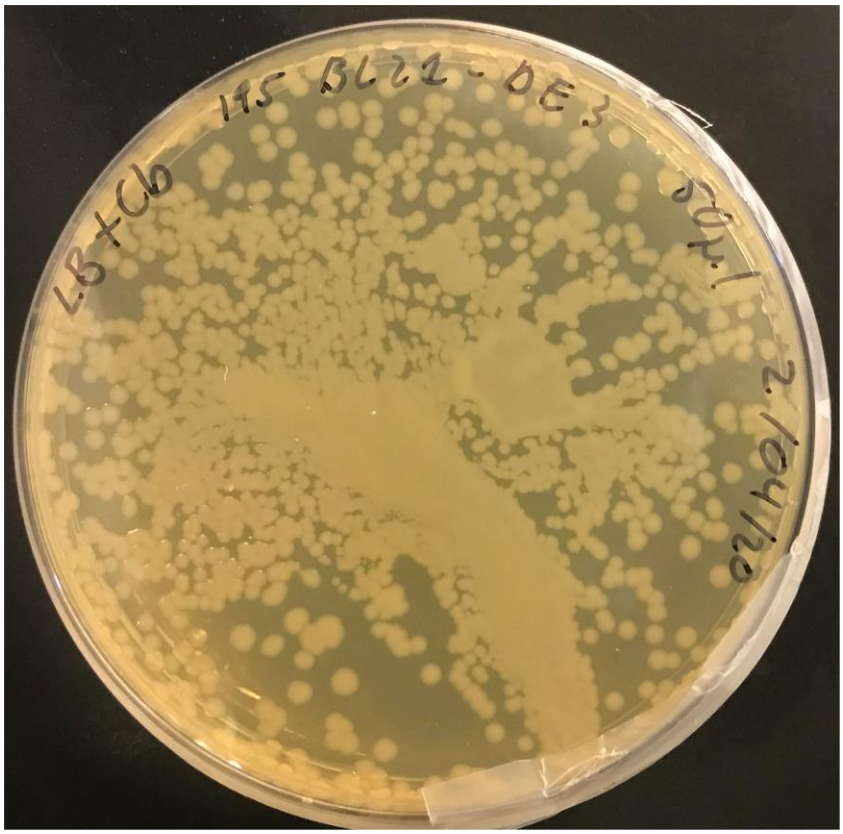


**Figure 2.5:** Immunoblot of pNSB1529 seven-hour induction using Rabbit-α-Rep primary antibodies and LiCor 800CW secondary antibodies. On the far left is the protein molecular weight marker (Amersham: RPN800E). Each lane represents a time point sample. The greatest accumulation of Rep occurs at T7. The double bands are non-specific proteins that interacted with the primary antibody.

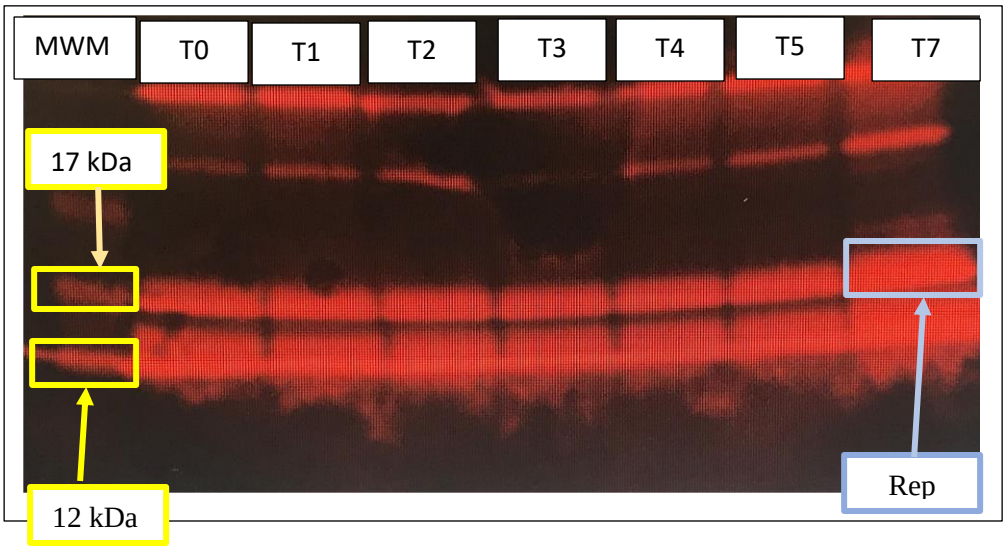


**Figure 2.6:** Aquastained Tris-Glycine-SDS gel of the overnight induction of pNSB1529. On the far left is the protein molecular weight marker. On the far right is the overnight induced pNSB1529. Each of the remaining lanes represents an aliquot from each time point listed: pre-induction, 1, 2, 3, 4, and 5 h post addition of IPTG as well as an overnight induced sample. pNSB1529 Rep is found at approximately the 17 kDa range.

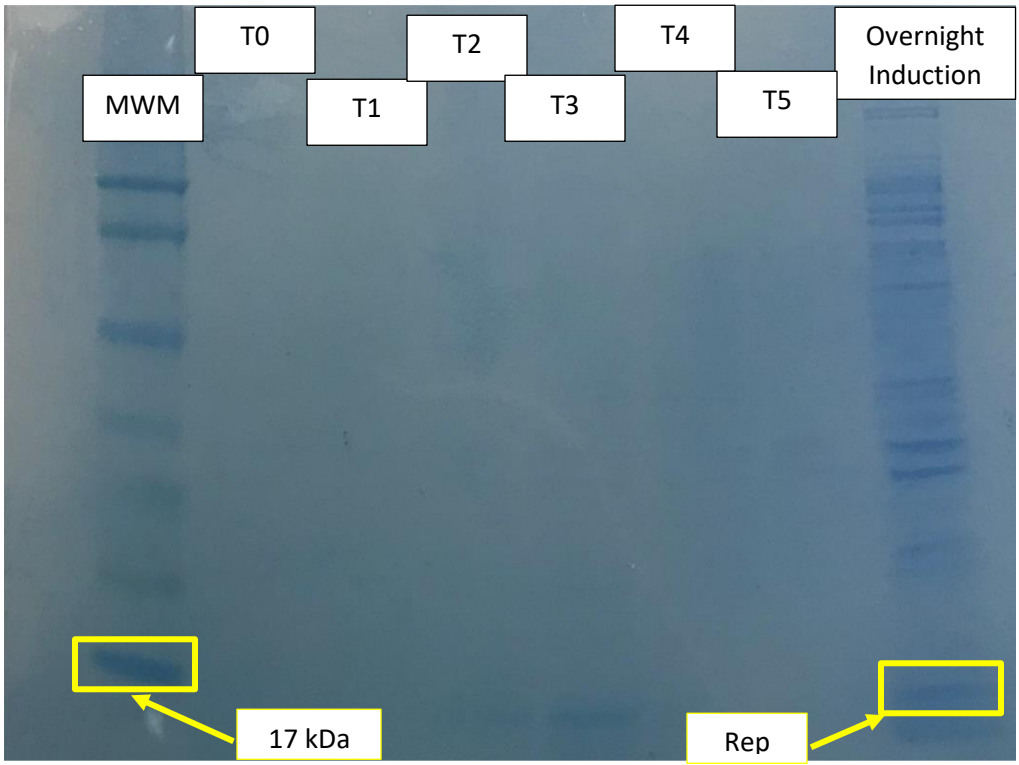


**Figure 2.7:** NMR profile of purified pNSB1529. It appears that Rep is marginally folded although it is difficult to distinguish from the overwhelming presence of Imidazole in the purified sample.

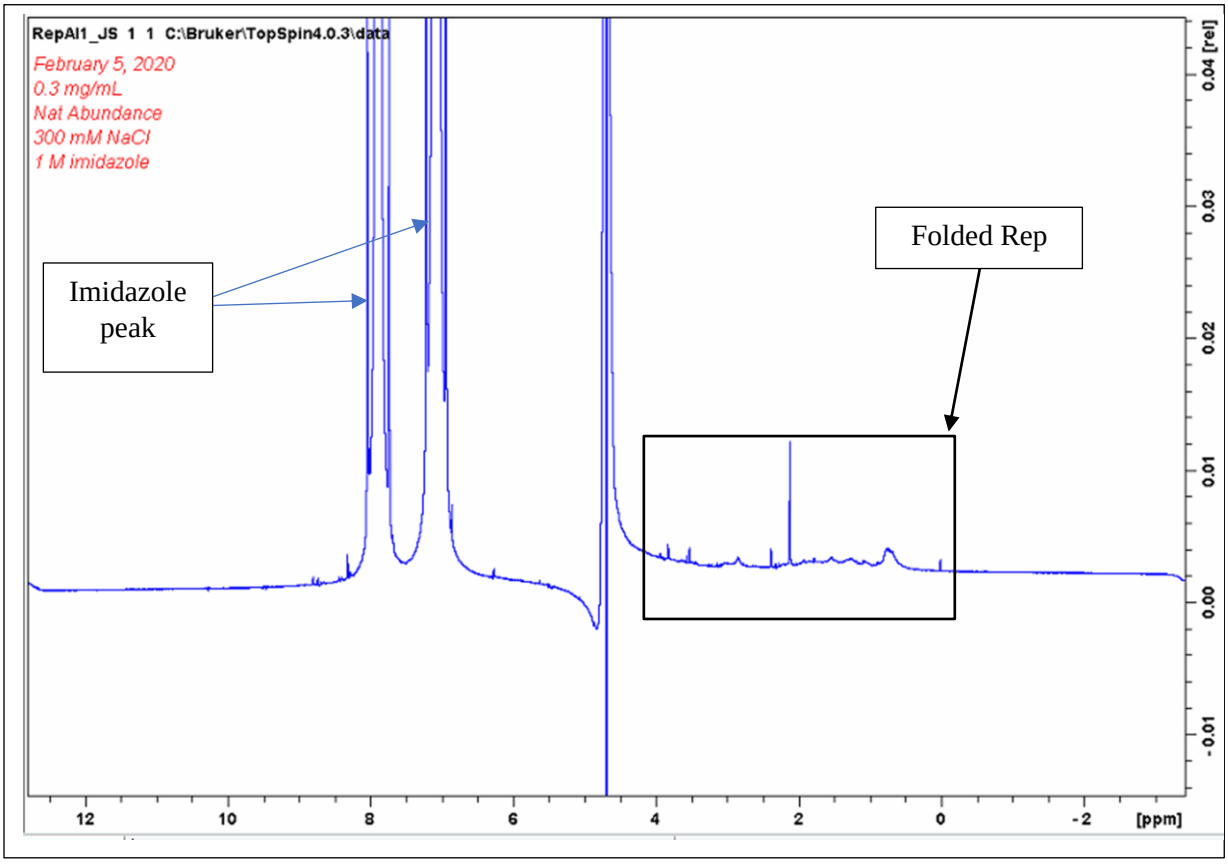


**Figure 2.8:** Partially solved structures of Rep. On the left is the NMR structure of the N-terminal region of TYLCV Rep for the first 120 amino acids. On the right is the Crystal structure of the N-terminal region of Wheat Dwarf Virus Rep for the first 120 amino acids. The bottom panel is the superimposition of 1L2M and 6Q1M.

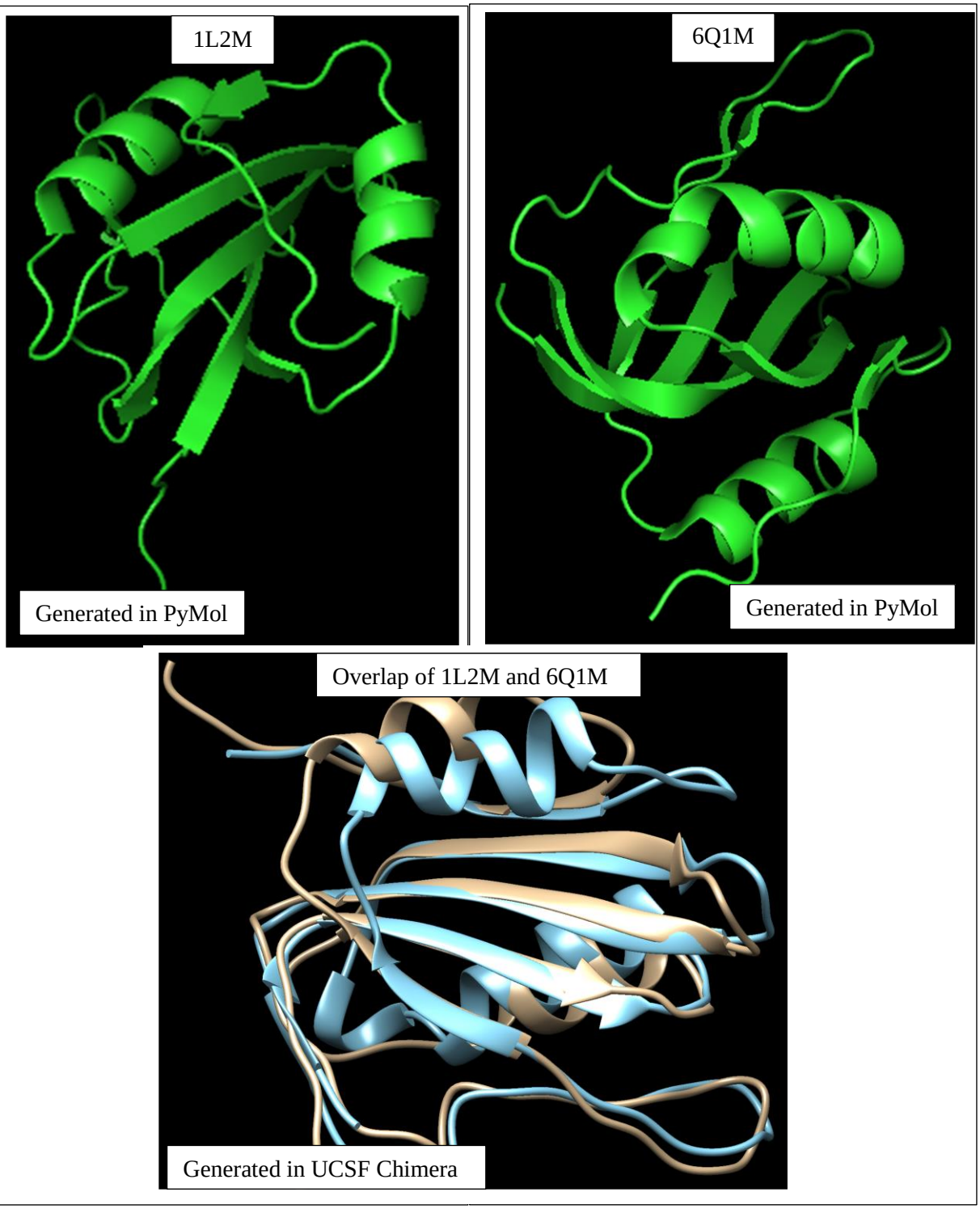


**Figure 2.9:** The *ab initio* modelled structure of REn from Robetta. This is the best *in silico* model from any of the attempted methods.

**Figure 2.10:** Phyre2 model of full-length Rep. This is the best model from all the attempted *in silico* methods. On the left is the N-Terminal region containing most of the binding domains; The long alpha-helix in the middle represents the Rep oligomerization domain; the domain on the right represents the ATPase domain. Note the generally poorly folded loops in the ATPase domain compared to the N-terminal domain.

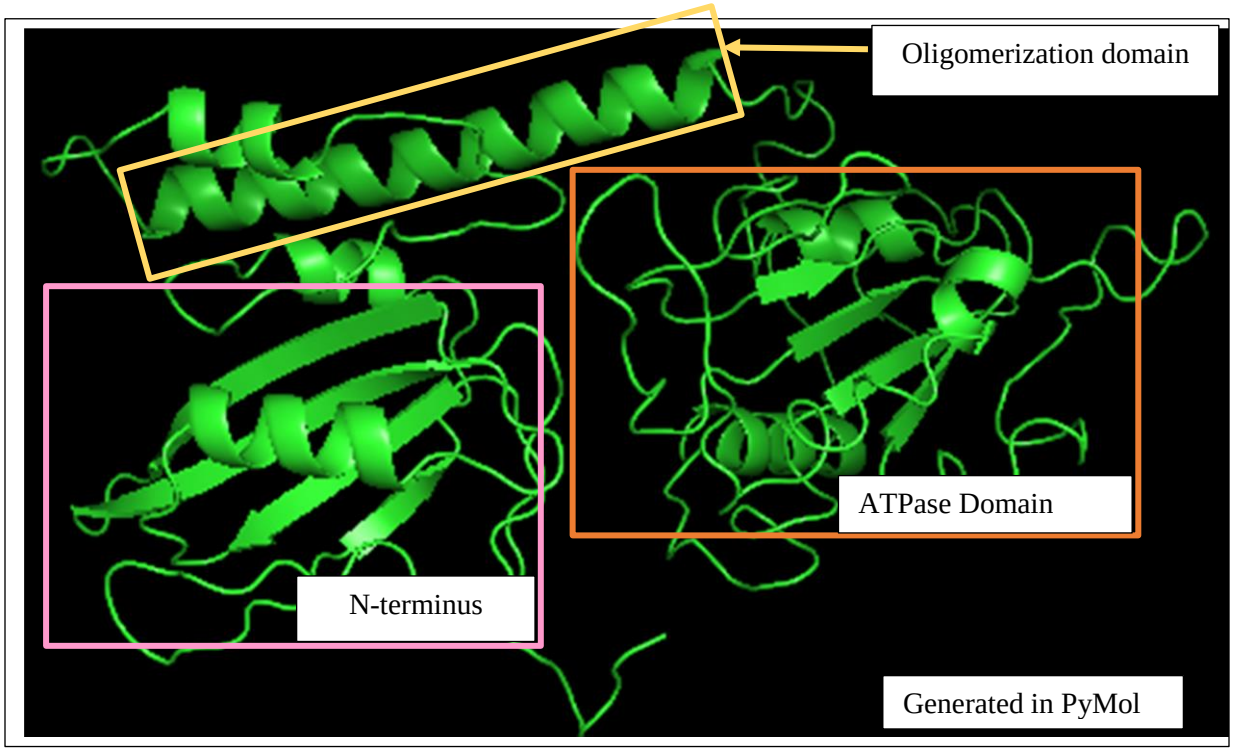


**Figure 2.11:** An example of a poorly folded Rep as modelled by I-TASSER. Note the almost non-existent folded ATPase domain and the almost incompletely folded oligomerization domain. The N-terminal region is folded quite well.

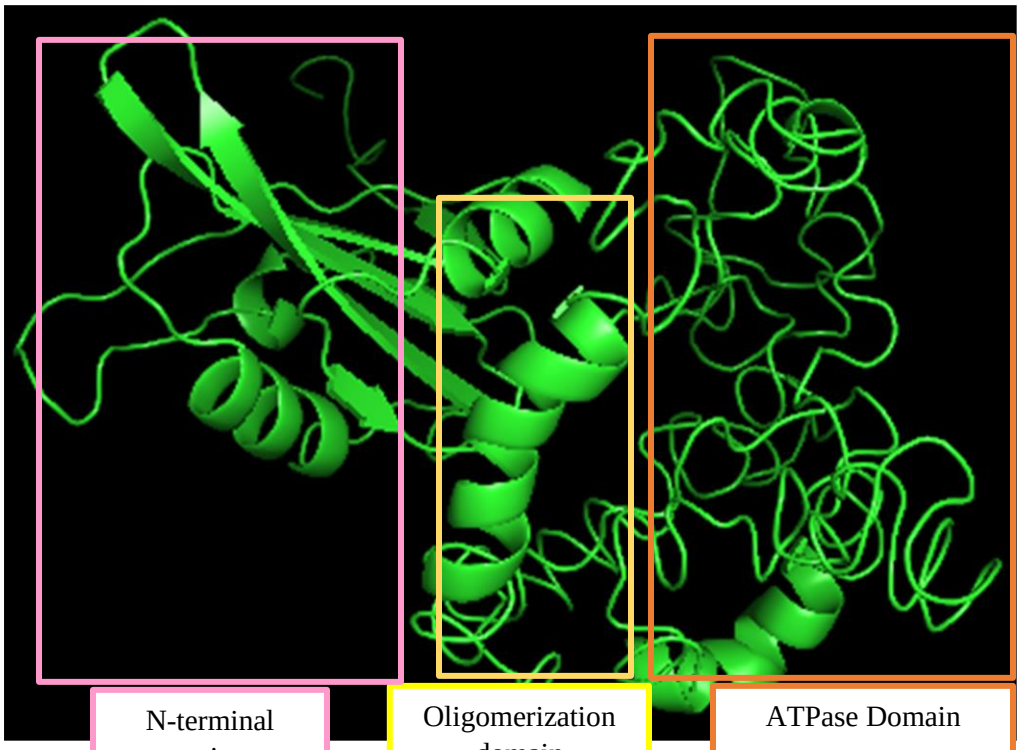


**Figure 2.12:** The comparative based model of REn from I-TASSER. It is an example of the generally poor results from the modelling methods.

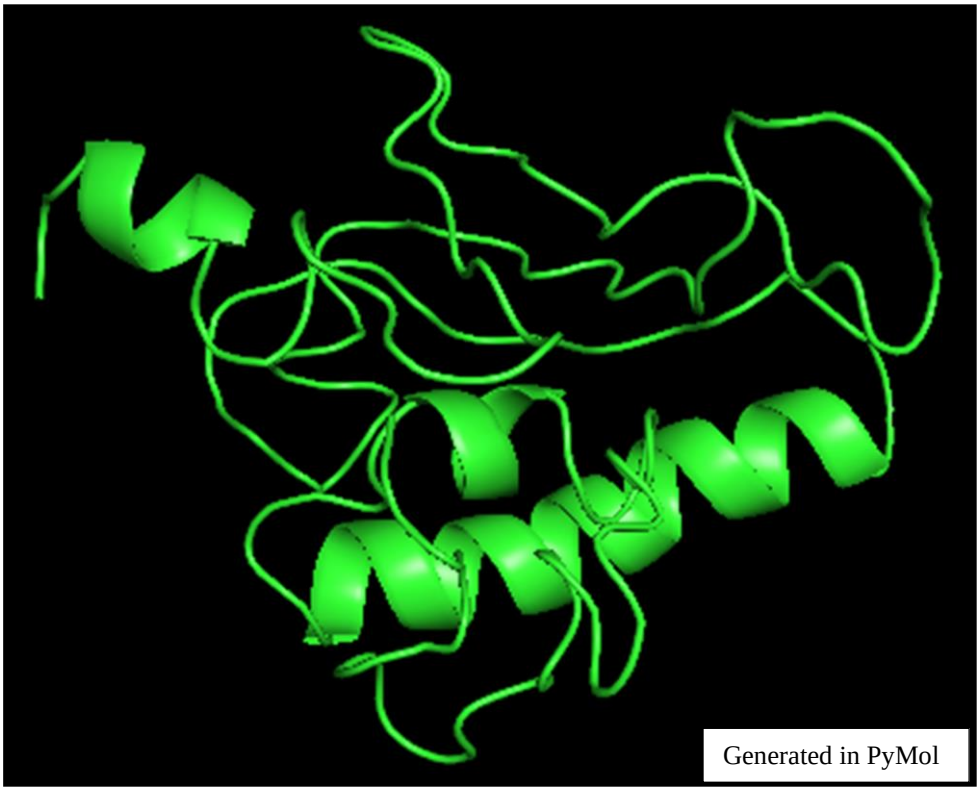


**Figure 2.13:** Immunoblot of purified pNSB1529 at 700 nm. In lane one is the protein molecular weight marker. Lane 2 represents sample from lysed pellet. Lane 3 represents sample from the lysis supernatant. Lane 4 represents the HisTrap with the addition of the supernatant and washed three times with binding buffer. Each subsequent lane was washed with two additional washes of binding buffer. Lane 8 contains the eluted Rep after 11 washes with binding buffer.