ABSTRACT

ENCAPSIDATION OF THE HOST RNAs BY BROAD BEAN MOTTLE VIRUS (BBMV) AND COWPEA CHLOROTIC MOTTLE VIRUS (CCMV)

Nipin Shrestha, M.S.
Department of Biological Sciences
Northern Illinois University, 2016
Dr. Jozef J. Bujarski, Advisor

Encapsidation of the viral genome into capsid is a highly specific process. Despite the viral RNA specificity, evidence of the host cellular RNA co-encapsidation have been shown in some insect viruses like Flock house virus (FHV) and plant viruses like Cucumber necrotic virus (CNV). This study is dedicated to finding the similar evidence of host RNA encapsidation by the two members, BBMV and CCMV of the Bromovirus family – model virus system. The Next Generation RNA sequencing (NGS RNA-Seq) can be implemented to sequence the RNAs encapsidated by the highly purified virions. We mapped and aligned the reads obtained from sequencing to the viral references to filter out the co-encapsidated non-viral reads. These co-encapsidated non-viral reads are then classified into different categories on the basis of their origin from the host plants. BBMV and CCMV co-encapsidated 1.465% and 1.323% of the host RNAs respectively. Majorities of the host RNAs encapsidated are nuclear including ribosomal RNAs (rRNAs) and messenger RNAs (mRNAs). Among the co-encapsidated plant RNAs, 21.228% and 35.220% reads are from transposable elements (TEs) in BBMV and CCMV respectively, which make these virions a potential carrier for the horizontal gene transfer (HGT) in plants.
ENCAPSIDATION OF THE HOST RNAs BY BROAD BEAN MOTTLE VIRUS (BBMV)
AND COWPEA CHLOROTIC MOTTLE VIRUS (CCMV)

BY

NIPIN SHRESTHA
©2016 Nipin Shrestha

A THESIS SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE
MASTER OF SCIENCE

DEPARTMENT OF BIOLOGICAL SCIENCES

Thesis Director:
Dr. Jozef J. Bujarski
ACKNOWLEDGEMENTS

I would like to thank Dr. Jozef J. Bujarski for providing me the opportunity to work in his lab and for his continuous support, guidance, and encouragement. I would also like to thank my committee members Dr. Melvin R. Duvall and Dr. Gabriel P. Holbrook for their time, support, and advice throughout my graduate education. I extend my special gratitude to Mr. Sean Burke and Dr. Duvall for the valuable guidance and assistance with bioinformatic analyses during the project. I would also like to thank Mrs. Alexandra Bujarska (Ola) for her assistance and guidance in the lab. I am grateful to all my current and former lab mates for their support and motivation. And last but not the least, I would like to thank my friends and family.
DEDICATION

This work is dedicated to my parents, Hari Krishna Shrestha and Ganga Shrestha for their faith, support and guidance throughout my life.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF TABLES</th>
<th>vii</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF FIGURES</td>
<td>viii</td>
</tr>
<tr>
<td>CHAPTER 1 LITERATURE REVIEW</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1.1 Discovery and the classification of plant viruses</td>
<td>1</td>
</tr>
<tr>
<td>1.1.2 Broad bean mottle virus (BBMV)</td>
<td>2</td>
</tr>
<tr>
<td>1.1.3 Cowpea chlorotic mottle virus (CCMV)</td>
<td>4</td>
</tr>
<tr>
<td>1.2 RNA synthesis in the Bromoviruses</td>
<td>6</td>
</tr>
<tr>
<td>1.2.1 De novo initiation of RNA synthesis</td>
<td>7</td>
</tr>
<tr>
<td>1.2.2 CPs in RNA synthesis</td>
<td>8</td>
</tr>
<tr>
<td>1.2.3 Spherules for the Bromovirus RNA replication</td>
<td>9</td>
</tr>
<tr>
<td>1.2.4 Roles of host factors in RNA replication</td>
<td>10</td>
</tr>
<tr>
<td>1.3 RNA encapsidation and assembly of virions</td>
<td>12</td>
</tr>
<tr>
<td>1.3.1 CP-RNA and CP-CP interaction in the encapsidation</td>
<td>14</td>
</tr>
<tr>
<td>1.3.2 Replicase-CP interaction and replication-coupled packaging</td>
<td>17</td>
</tr>
<tr>
<td>1.3.3 Host-dependent and -independent virion heterogeneity</td>
<td>18</td>
</tr>
<tr>
<td>1.4 Horizontal gene transfer in plants: brief evidence</td>
<td>19</td>
</tr>
</tbody>
</table>
1.5 An introduction to Next Generation RNA sequencing (Next Gen RNA-seq) ............... 20

1.6 Rationale and future prospects ................................................................................. 22

CHAPTER 2    MATERIALS AND METHODS ......................................................................... 26

2.1 Virus propagation ........................................................................................................ 26

2.2 Virus extraction and purification .............................................................................. 27
  2.2.1 Sucrose-cushioned virus purification .................................................................. 27
  2.2.2 Sucrose gradient ultracentrifugation .................................................................. 29

2.3 Electron microscopy and SDS-PAGE .......................................................................... 30

2.4 RNA purification ......................................................................................................... 31

2.5 Next Generation RNA sequencing (Next Gen RNA-seq) .......................................... 32

2.6 Sequence analysis ....................................................................................................... 32

CHAPTER 3    RESULTS ..................................................................................................... 34

3.1 Systemic infection on broad bean and cowpea plants .............................................. 34

3.2 Virus extraction and purification: cushioned pelleting and gradient purification .... 36

3.3 Quality analysis of the virus purification ................................................................... 39

3.4 NGS-RNA sequencing and analysis of the read data .............................................. 44
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.4.1 Summary of BBMV and CCMV RNA reads</td>
<td>44</td>
</tr>
<tr>
<td>3.4.2</td>
<td>46</td>
</tr>
<tr>
<td>Detailed characterization of reads originated from the host plant sources</td>
<td>46</td>
</tr>
<tr>
<td>3.4.3. The co-encapsidated transposable elements (TE)</td>
<td>49</td>
</tr>
<tr>
<td><strong>CHAPTER 4       DISCUSSION</strong></td>
<td>53</td>
</tr>
<tr>
<td>4.1 Host RNAs are encapsidated by BBMV and CCMV</td>
<td>53</td>
</tr>
<tr>
<td>4.1.1 Organellar RNA is co-encapsidated the least by two <em>Bromoviruses</em></td>
<td>53</td>
</tr>
<tr>
<td>4.1.2 Co-encapsidated transposable elements: a potential source of horizontal gene transfer (HGT)</td>
<td>55</td>
</tr>
<tr>
<td>4.2 Conclusion</td>
<td>57</td>
</tr>
<tr>
<td><strong>REFERENCES</strong></td>
<td>59</td>
</tr>
<tr>
<td><strong>APPENDIX      DETAILED PROTOCOLS AND TABLES</strong></td>
<td>69</td>
</tr>
<tr>
<td>Protocol A1. Virion extraction and purification from the infected leaves</td>
<td>70</td>
</tr>
<tr>
<td>Protocol A1.1 Sucrose-cushioned purification</td>
<td>70</td>
</tr>
<tr>
<td>Protocol A1.2 Sucrose gradient purification</td>
<td>71</td>
</tr>
<tr>
<td>Protocol A2: RNA extraction and purification from the purified virions</td>
<td>71</td>
</tr>
<tr>
<td>Protocol A3. Denaturing RNA gel electrophoresis</td>
<td>72</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1. Summary of the three Bromovirus genome. ................................................................. 6

Table 2. Summary of NGS RNA-seq reads obtained for the two Bromoviruses. ...................... 46

Table 3 Summary of plant-originated co-encapsidated reads in the two Bromoviruses........... 48

Table 4. Categories of transposable elements encapsidated by the two Bromoviruses.......... 50

Table A 1. Sample preparation for denaturing RNA gel electrophoresis................................. 73

Table A 2. Summary of the preparations for Tris-glycine SDS-PAGE.................................. 73
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Illustrative diagram of the Bromovirus genome</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>Schematic representation of roles of the host ESCRT proteins in the spherule formation by BMV.</td>
<td>11</td>
</tr>
<tr>
<td>3</td>
<td>Schematic illustration of the RNA3 and RNA4 co-encapsidation into a single virion particle</td>
<td>16</td>
</tr>
<tr>
<td>4</td>
<td>Symptoms on cowpea and broad bean seedlings by the CCMV and BBMV infection, respectively.</td>
<td>35</td>
</tr>
<tr>
<td>5</td>
<td>Virion purification by sucrose gradient ultracentrifugation.</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td><strong>Error! Bookmark not defined.</strong></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>The dark field transmission electron microscopy (TEM) analysis of the purified virion solutions</td>
<td>40</td>
</tr>
<tr>
<td>7</td>
<td>SDS-PAGE analysis of the virion preparations</td>
<td>42</td>
</tr>
<tr>
<td>8</td>
<td>Denaturing agarose gel electrophoresis of viral RNA preparations (1 % agarose gel).</td>
<td>43</td>
</tr>
</tbody>
</table>
CHAPTER 1

LITERATURE REVIEW

1.1 Introduction

1.1.1 Discovery and the classification of plant viruses

Viruses are among one of the most elusive infectious agents in the molecular pathology. The last universal cellular ancestor (LUCA) theory of the viral origin explains origin of viruses might have been parallel to or even before the origin of the first living cells (Holmes, 2011). The first virus to be discovered was Tobacco mosaic virus (TMV) in the year 1892 by Ivanoski. It was also the first virus to be visualized as a crystal through electron microscopy in the year 1935. Since then, it has been used as a prototype to study biology of the host plant, tobacco, and widely used as a model virus system in the study of host-virus interaction, cellular trafficking and many other sectors like health and pharmaceuticals (Scholthof, 2004).

The International Committee on Taxonomy of Viruses (ICTV) has been established to classify viruses from vertebrates, invertebrates, plants and bacteria. The plant virus classification system is based on morphology of virus particles (virions), their physical properties (size,
molecular weight, sedimentation coefficient and stability), chemical composition (proteins, carbohydrates and lipids), the nature and organization of their genome (RNA or DNA, strandedness – single or double, sense - positive or negative, linear or circular and number of segments), antigenic properties and biological properties (host range, disease pathogenicity, vectors, mode of transmission and geographical distribution) (Becker & Hadar, 1983; Hull, 2013). Among the various sixteen groups of plant virus families approved by ICTV, Bromoviridae falls under positive-sense single-stranded RNA (+ssRNA) virus with three classes, namely, *Cucumovirus, Bromovirus* and *Alfamovirus* (Harrison et al., 1971). *Bromovirus* includes the non-enveloped, icosahedral (T=3) viruses with tripartite ssRNA genome and a subgenomic RNA. They have been widely used as model viruses for the study of viral structure, plant virus replication and virus-host relationship. *Brome mosaic virus* (BMV), *Broad bean mottle virus* (BBMV), *Cowpea chlorotic mottle virus* (CCMV) and *Spring beauty latent virus* (SBLV) cumulatively form the *Bromovirus* group (Lane, 1974; Valverde, 1884).

1.1.2 Broad bean mottle virus (BBMV)

BBMV and CCMV are non-enveloped, spherical plant viruses made of 180 coat proteins forming an icosahedral (T=3) capsid that encapsulates three +ssRNA into three different viral particles separately. Virion particle of BBMV has shown a diameter of 260Å in X-ray crystallography and CCMV has a diameter of 280Å (Finch & Klug, 1967; Speir, Munshi, Wang, Baker, & Johnson, 1995).
BBMV infects broad bean (*Vicia faba*) as a natural host but it can also infect a broad range of other dicotyledonous plants like some *Chenopodium* spp, *Nicotiana benthamiana*, *Pisum sativum*, etc. BBMV infection shows systematic mottling symptoms on the leaves of the natural host. Various strains of BBMV on the basis of their origin have been reported, four of those strains serologically characterized are Morocco (Mo), Tunisia (Tu), Syria (Sy) and Sudan (Su) strains. BBMV is generally considered as a non-seed transmissible virus but incidences of seed transmission have been reported when co-infected with Bean yellow mosaic virus (BYMV) (Makkouk, Bos, Rizkallah, Azzam, & Katul, 1989). *Apion radiolus*, *Hypera variabilis*, *Pachytychius strumarius*, *Smicronyx cyaneus*, and *Sitona lineatus* are some of the insect vectors known to transmit the virus to healthy faba plants (Fortass & Diallo, 1993).

BBMV genome has been completely sequenced, which comprises three RNAs encapsulated into three different viral particles (Figure 1). RNA1 is a linear RNA with a length of 3158 base pairs (bp), which codes for a 109 kDa and 966 amino acid (a.a) non-structural protein. RNA2 is a 2,799 bp linear RNA that codes for another non-structural 2a protein of 815 a.a and 90.5 kDa. RNA3 is a di-cistronic linear RNA of 2,293 bp that codes for a 3a protein (32kDa) and a coat protein (CP). CP is coded from the subgenomic RNA4 (SgRNA4), which is coded from the intercistronic subgenomic RNA promoter on minus strand of RNA3 (Dzianott & Bujarski, 1991; Romero, Dzianott, & Bujarski, 1992).
1.1.3 Cowpea chlorotic mottle virus (CCMV)

Cowpea (*Vigna unguiculata*) plants infected with CCMV show systemic chlorotic mottling symptoms on the leaves. The virus can infect a broad range of other dicotyledonous plants like soybean (*Glycine max*), *Nicotiana benthamiana* and *Chenopodium quinoa*. On the basis of symptoms, CCMV can be a type strain (CCMV-T), mild strain (CCMV-M), soybean strain (CCMV-S) or resistant strain (CCMV-R) (Bijaisoradat & Kuhn, 1985; de Assis Filho, Paguio, Sherwood, & Deom, 2002; Wyatt & Kuhn, 1980). CCMV is a vector transmissible but non-seed transmissible virus. Insect vectors of CCMV include species of chrysomelid beetles including bean leaf beetle (*Ceratoma trifurcate*) and spotted cucumber beetle (*Diabrotica undecimpunctata howardi*). Transmission of the CCMV by beetle vectors has been found to be dependent on the CP of the virus (Mello, Clark, & Perry, 2010). The tripartite genome of CCMV has been sequenced completely. RNA1 (3,171 bp) codes for the 1a protein, RNA2 (2,775 bp) codes for 2a protein and RNA3 (2,173 bp) codes for 3a protein and CP (Table 1).

Both 1a and 2a polypeptides are required for RNA replication in *Bromoviruses*. N-terminal domain of the 1a protein has methyltransferase activity required for capping and the C-terminal domain is known to function as a helicase. The 2a protein has a polymerase-like core domain, the RNA dependent RNA polymerase (RdRp). A compatibility between 1a and 2a proteins is required for the efficient viral RNA replication. Specifically, RNA3 replication has been found to be altered to a great extent in the absence of 1a and 2a protein interaction. The 1a and 2a proteins are conserved along the members of *Bromovirus* spp. (BMV, CCMV and BBMV) but the 3a protein
has lesser sequence identity (~52% between CCMV and BMV), which indicates the role in host specificity. The 3a protein, responsible for the systemic infection by cell to cell movement of the viruses, is known as movement protein. Intercistronic region of RNA3 has a subgenomic promoter region that codes for the sgRNA4, which expresses the CP. RNA-CP interactions are known to have an important role in the efficient viral assembly (Allison, Janda, & Ahlquist, 1989; Smirnyagina, Lin, & Ahlquist, 1996).

---

**Figure 1. Illustrative diagram of the Bromovirus genome.** Genome of Bromoviruses has three genomic RNAs and a subgenomic RNA 4 (sgRNA4). 5'UTR in all three RNAs have hairpin structure known as B-box. 3'UTR has a tRNA like structure (TLS) required for replication and packaging. The intergenic region of RNA3 has replication enhancer overlapping the promoter for sgRNA4. Diagrams on the right represent three homogenous spherical Bromovirus particles measuring approximately 26 nm.
1.2 RNA synthesis in the Bromoviruses

The successful infection cycle of a virus includes multiple steps starting with disassembly of the viral particle upon the invasion to the host cell followed by genome replication, translation, packaging into a proper virion architecture and cell-to-cell movement. The infection cycle of viruses is diverse based on the type and nature of the genomic materials. All DNA viruses, double-stranded RNA (dsRNA) viruses and negatively ssRNA (-ssRNA) viruses need to synthesize the positive-sense RNA (mRNA sense RNA) before translation into proteins but the +ssRNA viruses can start the translation immediately by using the host components (Flint, Enquist, Racaniello, & Skalka, 2004). Genome replication and packaging among icosahedral viruses have resemblance to some extent though the mechanism might differ more extensively. Studies in spherical plant viruses like Tombusviruses, Tymoviruses, Sobemoviruses, Dianthoviruses, Bromoviruses, Cucumoviruses, and Alfamovirus have shown the interaction of viral RNA with host factors for replication and all positive-strand RNA viruses are known to replicate in a specialized intracellular compartments in the host cytoplasm known as spherules (Rao, 2006).
1.2.1 *De novo* initiation of RNA synthesis

*Bromoviruses* represent one of the most studied plant viruses with tripartite +ssRNA genome. Replication in BMV is initiated *de novo*, which requires the protein-protein interaction. The 1a protein directs the recruitment of 2a protein along with the viral RNA to induce the spherule formation in the host cytoplasm by the cellular membrane reorganization. Both the N-terminal capping activity by methyltransferase and C-terminal helicase activity are indispensable for the functioning of 1a protein in replication but for the enhancement of the 2a activity, only the helicase domain is sufficient. So, one of the roles of ATPase- and GTPase-bearing helicase domain might be to unwind secondary structures in the RNA template required for initiation. Also, helicase is known to separate template from the nascent RNA strand during elongation of replication by the RNA-dependent RNA polymerases (RdRp) (Kroner, Young, & Ahlquist, 1990; Subba-Reddy et al., 2012). The 5’ region of RNA1 and RNA 2 both have B-box region, which has the regulatory function in translation, and it is also known to be pivotal for RNA replication. The 1a protein binds to the region to form the membrane-associated replicase complex (Yi, Vaughan, Yarbrough, Dharmaiah, & Kao, 2009).

The tRNA-like structure at the 3’ untranslated region (UTR) of all three genomic RNAs and replication enhancer (RE) in the intergenic region of RNA3 are required for the efficient negative-strand RNA synthesis. For the synthesis of the positive-stranded RNAs, the *cis*-acting 5’ sequences are required. The subgenomic promoter overlapping the portion of RE is required for synthesis of the subgenomic mRNA (Ahola, den Boon, & Ahlquist, 2000). The 5’ UTR in RNA1 and 2 and the RE in RNA3 intergenic region are known to have conserved residues forming TΨC loop structure.
typical of t-RNA structure known as B-box. This region interacts with helicase domain of 1a protein for the viral RNA template recruitment. Mutations in the 1a domain have been found to be related to lowered or abolished viral RNA accumulation and stabilization. Mutations leading to disruption of the stem-loop structure are related to the lowered RNA replication. So, the viral RNA template specificity for the replication is directed by 1a interaction with the template at the stem-loop region followed by recruitment of the 2a protein (Chen, Noueiry, & Ahlquist, 2001).

1.2.2 CPs in RNA synthesis

The multifunctional, viral CPs along with the host proteins have vital roles in localization and functioning of the replicase complex. The CPs are responsive to the binding signal in the promoter region of the 3’UTR that directs negative strand synthesis. The region has a triloop of 5’-AUA-3’ forming the stem loop-C (SLC) hairpin structure. NMR study has shown stacking of A3 to the inside of the loop with U2 on the top forming clamped adenine motif (CAM). The similar hairpin structure is conserved in the intergenic region of the RNA3, subgenomic promoter of the RNA4. However, in the sgRNA4 promoter, an extra A residue has been found but a general loop structure is not altered, so the CP-induced recognition of the RNA template by the viral replicase complex seems to be structure conserved rather than the sequence, and the mechanism of the initiation of genomic negative RNA synthesis and sgRNA4 synthesis might be similar (Skov, Gaudin, Podbevšek, Olsthoorn, & Petersen, 2012). The accumulation of replicase complex has also been shown to be dependent on the CP abundance. The overexpression of BMV CP is known to reduce the viral RNA replication by depleting the translation of RNA1 and RNA2. Interestingly,
CP also binds to the 7-nucleotide stem loop of the B-box in the 5’ UTR, which is known to be one of the major components of the replicase complex, so CPs are also known to play some major, nonstructural roles in the fate of RNA templates, especially during replication, transcription or translation (Yi, Letteney, Kim, & Kao, 2009).

1.2.3 Spherules for the Bromovirus RNA replication

Subcellular localization of the replicase complex into a membrane-bound spherule is a well-known process for the replication of all +ssRNA viruses. The disease development in CCMV-infected cowpeas has been characterized by the formation of inclusion bodies similar to membranous vesicles in the host cell cytoplasm between the rough endoplasmic reticulum (RER) membrane and the nuclear envelope (Kim, 1977). Varieties of other organelle membranes are also known to contribute to a site of vesiculation for the spherules. Endoplasmic reticulum (ER) membranes are the most common sites; examples include Poliovirus and Bromovirus. However, other membranes like mitochondria, chloroplast and lysosomes also provide the site for vesiculation in many other viruses like Flock house virus (FHV) forms spherules from the mitochondrial membrane (Bamunusinghe, Seo, & Rao, 2011). In BMV, 1a protein localizes to the perinuclear ER membrane to induce 50-75 nm spherules. These individual structures serve as compartments for replication in case of relatively lower availability of 2a proteins, but in the presence of abundant 2a proteins, further membranes of spherules rearrange into the multilayered structures. The new structure also serves similar function to spherules by allowing replication with
the accumulation of 1a and 2a proteins and protection of template RNA from cellular nucleases (Schwartz, Chen, Lee, Janda, & Ahlquist, 2004).

1.2.4 Roles of host factors in RNA replication

The formation of spherule is induced by 1a proteins and is sufficient to induce the vesiculation in absence of other replicase components. However, 2a proteins can only be recruited by the interaction of its N-terminal region preceding the core polymerase domain with the C-terminal region of the 1a protein; 1a proteins then recruit RNA templates to the spherules by interacting with conserved stem loop structures in 3’ and 5’UTRs (den Boon, Chen, & Ahlquist, 2001; den Boon, Diaz, & Ahlquist, 2010). Along with these viral proteins, several host proteins and genes are also known to be involved in the spherule formation. As the structure of the spherules and their biogenesis are similar to the multivesicular bodies (MVB) in the host cell, most of the host factors involved are the components of cellular vesiculation pathways. Reticulon homology domain proteins (RHPs), which are a family of membrane-shaping proteins, conserved in yeast, human and plants, play a crucial role in spherule formation and replication. RHPs interact with 1a localized at the perinuclear membrane to initiate the spherule formation. Some investigations also hinted at the relation of the spherule formation to the lipid biosynthesis, as the removal of DOA4 and BRO1 genes caused depleted spherule formation and replication (Diaz, Wang, & Ahlquist, 2010; Wang et al., 2011).

Recently, Diaz et al (2015) has shown the effects of host endosomal sorting complexes required for transport (ESCRT) components on the spherule formation and RNA replication.
Different components of ESCRT had varied effects on the process, ranging from change in spherule size to spherule frequency and alteration of RNA replication up to 25 folds. *Snf7p* was detected as the most crucial ESCRTIII factor that abolished spherules completely on deletion even if other ESCRTs like *Vps23p* and *Vps20p* were functional, which signifies that the *Snf7p* recruits other ESCRTs to the spherule for the replication. However, deletion of later two genes resulted in depletion of replication in a spherule-independent manner, so host factors like ESCRTs are required for induction of spherules and might be part of the replication process as well (Diaz, Zhang, Ollwerther, Wang, & Ahlquist, 2015). The brief illustration is shown in the figure 2.

Figure 2. Schematic representation of roles of the host ESCRT proteins in the spherule formation by BMV (Diaz et al., 2015).

Among the other host factors, the deletion mutation analyses have shown altered BMV replication in yeast deleted with genes responsible for cell growth, protein homeostasis, protein
trafficking and translation. Deletion of several host genes (*DHR2, ECM16, NOP7, PWP1, RIO2, RPA43, UTP4, UTP18, YGR251W*) enhances the BMV replication, which indicates the possible competition of the cellular process against virus replication (Gancarz, Hao, He, Newton, & Ahlquist, 2011).

1.3 RNA encapsidation and assembly of virions

The successful infection cycle of viruses largely depends on the viral encapsidation and assembly, as it is known to be vital for the cell-to-cell and long-distance transport of the virion in the host system. The process of the virion assembly is a highly organized event requiring protein-protein interactions and protein-nucleic acid interactions. These interactions are known to increase the specificity of the encapsidation process. The overall mechanism of CP assembly and encapsidation of the viral genome varies among various virus families. In dsDNA viruses, procapsids are formed by assisted or self-assembly of the CPs and the genomic materials are directed into preformed capsid shells by packaging motors. In contrast to this, in ssRNA viruses the most common phenomenon is the co-assembly process, where CPs interacts with the specific region of the viral genome to serve as nucleation for the assembly of progeny virions. Also, the *in silico* analysis made by Dykeman et al. showed the possible kinetic traps in CPs assembly, and possibly the interaction of CPs with genome for nucleation helps evade the trap (Dykeman, Stockley, & Twarock, 2014). In concordance to this finding, encapsidation in BMV has been found to be dependent on RNA-CP interaction and CP-CP interaction. Specifically, BMV RNA3 has a
bipartite signal to direct the RNA encapsidation. The tRNA-like structure (TLS) in the 3’ UTR is the nucleation element (NE), which non-specifically interacts with CP to initiate the polymerization of the CP dimers into pentamers. The 187 nt element known as packaging element (PE) is another *cis*-acting signal for packaging found within the 3a-open reading frame (Choi & Rao, 2003).

In *Bromovirus*, all three genomic RNAs have structural similarities at their 3’ UTR, which form a tRNA-like structure (TLS). The structure along with the *cis*-acting packaging element (PE) has been shown to be important for the RNA3 virion particle assembly in BMV, so the rest of the two genomic RNAs might have the similar RNA-dependent CPs assemblies for the encapsidation (Choi & Rao, 2003). However, an *in vivo* packaging experiment in BMV has shown that RNA1 and RNA2 packaging is independent of the presence of 3’TLS, so there might also be possibilities that encapsidation requirement for the three different viral particles differ from each other (Annamalai & Rao, 2007). Also, in CCMV, the requirement of these bipartite signals and other regions of the 3a-open reading frame (ORF), CP-ORF or 3’ UTR is known to be dispensable for the viral RNA encapsidation (Annamalai & Rao, 2005a). So, the inference can be drawn that the mechanism and regulation of the encapsidation process in *Bromoviruses* might be different among different members of the family. Also, BMV has monocotyledons as their natural hosts, whereas CCMV and BBMV have dicotyledons as the natural host, so the regulation of RNA encapsidation might also have some roles in the host specificity in *Bromoviruses*. 
1.3.1 CP-RNA and CP-CP interaction in the encapsidation

Interaction of CPs with cis-acting elements in the viral RNA is through the highly basic N-terminal arginine-rich motif (ARM), which is conserved among a wide range of plant viruses (BMV, CCMV, BBMV, Cucumber mosaic virus [CMV], TMV) and animal viruses (Human immunodeficiency virus [HIV], Adeno associated virus [AAV], FHV) (Rao, 2006). The 25aa N-terminal motif has both specific and non-specific determinants in the case of BMV RNA. The presence of significant amount of positively charged basic residues justifies the non-specific binding to negatively charged phosphate backbone of BMV RNA. Duggal and Hall (1993) also showed that BMV RNA1 has specific domains that interact with ARM for the viral RNA selectivity. The PE in the MP ORF of RNA3 has three stem-loop structures, stem loops A, B and C (SLA, SLB and SLC). The interaction of CP ARM with PE thus is predicted to be structure conserved rather than the sequence. In absence or alteration of the nucleotides affecting the structure, the virion RNA failed to be encapsidated, whereas removal of 3’ TLS from the BMV RNA3 increased the proportion of non-specific RNA encapsidation, which was reduced after exogenous introduction of RNA3 with 3’TLS. This suggests the non-specific competitive interaction of CP with viral RNAs (Damayanti, Tsukaguchi, Mise, & Okuno, 2003). The N-terminal ARM and PE are also known to have roles in co-packaging of RNA4 along with the RNA3. Deformation of PE in MP ORF depleted the in vivo RNA4 packaging. Also, no known signal is present in RNA4 for the packaging so, RNA4 packaging might be directed by the signal present in the RNA3 PE. Choi and Rao (2003) hypothesized the model of co-packaging as shown in Fig.2. Initial interaction of NE and PE with ARM packages RNA3 into virion particle, which
exposes a portion of ARM towards the outside that specifically interacts with RNA4 to co-package it.

Differences between requirements of bipartite signal in BMV and CCMV are also accompanied by the differences in ARM requirements between the two viruses. Deletion of the amino acid from position 909 to 919 in CP ORF of CCMV has shown a detectable reduction in encapsidation of viral RNA along with formation of polymorphic virions. The effect was more prominent for encapsidation of RNA1 and 2, whereas RNA3 and sgRNA 4 were less affected. Interestingly, another mutant with disrupted α-helical structure of ARM by substitution of three arginine and one lysine with four proline had no effect on the competency of CCMV CPs for encapsidation, which is contrary to the requirement of the secondary structure for BMV RNA encapsidation (Annamalai, Apte, Wilkens, & Rao, 2005).

Amount of RNA encapsidated and virion morphology are other aspects of assembly impacted by the CP-CP and CP-RNA interaction. Generally, multi-partite virions like BMV maintain exclusive particle homogeneity and various investigations have shown possible roles of CP in the process. As described above, the N-terminal basic peptide tail in CP interacts with negatively charged genomic RNA to encapsidate it. The mutational analysis to alter number of positively charged residues in the N-terminal tail has shown significant variation in length and amount of RNA encapsidated without distinguishable virion size differences (Ni et al., 2012). However, in other instances the mutation has also been linked to the formation of polymorphic virion particles like in TMV, where the mixture of T=1 and T=3 virions was reported due to mutation in ARMs (Calhoun, Speir, & Rao, 2007). Cadena-Nava et al (2012) showed in an in
*vitro* assay with CCMV that as long as the ratio of CP:RNA is sufficiently high, CCMV encapsidates RNA ranging from 140 to 12,000 nt, provided that the total positive charge contributed by ARMS is comparable to the negative charge carried by the RNA. CCMV CP assembly accomplishes this encapsidation by incorporating either multiply truncated RNAs into single capsid or larger RNA into conjoined multiple capsids (Cadena-Nava et al., 2012). This explains the possible role of electrostatic interaction in the assembly pathways. Furthermore, from the perspective of electrostatics, a two-step pathway of CP assembly in which the initial nucleation of viral assembly starts by electrostatic interaction of CPs with RNA through ARMs, then the excess of CP dimers attached to RNA is transferred to developing capsid through electrostatic interaction between ARMs of displaced CP and outer negatively charged CPs attached to RNA (Garmann et al., 2014). The figure 3 illustrates the role of CP in co-encapsidation of the RNA 3 and RNA 4 into a single viral particle.

![Figure 3. Schematic illustration of the RNA3 and RNA4 co-encapsidation into a single virion particle. RNA3 is encapsidated into the virion, then the extended ARM in CP interacts with RNA4 to encapsidate it into the viral particle. Image rebuilt from Sztuba-Solinska & Bujarski (2008) with permission.](image)
1.3.2 Replicase-CP interaction and replication-coupled packaging

Viral replicase complex also impacts the efficiency of viral assembly. Particularly in +ssRNA viruses genome, packaging, is functionally coupled to the replication. In presence of the aberrant replicase, significant amount of cellular RNAs are encapsidated by BMV particles. Further, particle size homogeneity for the different genomic RNA content is altered in case of defective replicase complex. Thus, it is also functioning in the regulation of CP-CP interaction to maintain the size and the specificity of the virions. Possibly during the transport of the replicated RNA to the cytoplasm for the translation, the CPs residing on the neck of spherules interact with the replicase for the specificity of RNA to be encapsidated. So there might be some role of replicase in RNA selectivity and possibly replication is coupled to the encapsidation (Rao, 2006; Rao, Chaturvedi, & Garmann, 2014).

Another significance of the replicase-CP interaction is in the replication-coupled packaging, which is known to be inherent to the replication-dependent transcription and translation of CP subunits. The CPs transcribed from heterologous replication (BMV CP expressed in FHV RNA) product are reported to encapsidate higher amounts of non-viral RNA. Furthermore, exogenous sgRNA4 transcripts were not translated into CPs, but also addition of purified functional replicase failed to translate the transcripts into CP subunits (Annamalai, Rofail, DeMason, & Rao, 2008). Role of replicase in replication-independent manner has also been explicated by Annamalai and Rao by the transient expression of RNA3 in *N. benthamiana* (Annamalai & Rao, 2005b). Thus,
viral specificity in encapsidation is enhanced by the CP-viral replicase complex interaction both in the replication-dependent and -independent manners.

1.3.3 Host-dependent and -independent virion heterogeneity

Recently, encapsidation of the RNA has also been shown to be determined by the host. Relative abundance of the three BMV particles is found to be heterogeneous in barley, wheat and tobacco. The virion particles exhibited varied physiochemical properties like resistance to peptidases, buoyant densities, sizes, and the amount of RNA encapsidated. Further, incidence of recombination in the 3’UTR of RNA1 was found in virions from barley and wheat but not in tobacco. Interestingly, the degree of post-translational modification (PTM) for viruses from different hosts was also different. (Ni, Vaughan, Tragesser, Hoover, & Kao, 2014). Not only the virions from different hosts have heterogeneity, but also the three virion particles (RNA1, RNA2 and RNA3/4) in the same host exhibit physiochemical heterogeneity. This might be one of the reasons behind altered requirements for RNA encapsidation by different particles. Despite the similar changes in the conserved regions of CP or RNA, the influence on encapsidation efficiency and specificity remains varied, like mutation in ARM altering charge shows depletion of RNA1 but RNA2 abundance is not affected. These differences might be facilitating the timing of viral disassembly and gene expression in the infection process. Thus, these varied virion properties within and among hosts can be accounted for by the requirement of viruses to evade diverse host defense mechanisms for successful infection (Vaughan et al., 2014).
1.4 Horizontal gene transfer in plants: brief evidence

Horizontal gene transfer (HGT) is an asexual transfer of genes from one organism to another that is inheritable. HGT can be between individuals from the same species or different species. It is a well-documented process in bacteria and other prokaryotes. However, it is not yet a significantly established process in eukaryotes. Some incidences of horizontal transfer of transposons (HTT) have been documented in animals and plants (Diao, Freeling, & Lisch, 2005). HTT has been known to be linked to the host-parasite interactions influencing the genomic evolution (Gilbert, Schaack, Pace II, Brindley, & Feschotte, 2010). In plants, *Agrobacterium*-mediated transfer of plasmid genes is known naturally. In most of the other instances, majority of plant HGT has been contributed by bacteria. The most observed HGT among plants is the transfer of genetic elements from the mitochondrial genome of one plant to another. Even the transfer of an intron from angiosperm to gymnosperm has been reported (Richardson & Palmer, 2007).

In case of animals, lentiviral retrotransposons are known to be transferred to the genomes of many species. In humans, >50% of the genomic content are transposable elements (TE). Upon the infection of viral diseases, viral genomic contents are known to be integrated to the host genome for the expression of viral proteins (El Baidouri et al., 2014). A similar mechanism might be present in viruses to serve as a part of HGT in eukaryotes. In context of plants, no virus-mediated HGT of non-viral genetic element has been substantiated yet in land plants. However, in a green alga, *Prochlorococcus*, repeated exchange of photosynthetic genes with viruses has been observed (Lindell et al., 2004). Also, horizontal transfer of a DNA transposon, MULE, has been
documented between *Setaria* and rice lineages, which diverged from each other at least 50 million years ago. Mechanism for these HGT are not known, but vector-mediated transfers are speculated (Diao et al., 2005).

### 1.5 An introduction to Next Generation RNA sequencing (Next Gen RNA-seq)

A new dawn of genomics and transcriptomics started with the revolution in non-Sanger-based sequencing known as the Next Generation sequencing (NGS) or massively parallel sequencing. NGS employs the polymerase chain reaction (PCR)-based amplification of the fragmented DNA attached to adapters, cDNA in the case of RNA-seq, followed by the high-throughput parallel sequencing, which produces millions of reads in a single cycle. After the first commercially successful Roche454 NGS system, there has been a rise of many other commercially available sequencers like Illumina genome analyzer (GA), AB Solid system, Personal Genome Machine (PGM), Pacific Biosciences (PacBio) sequencers, etc. (Liu et al., 2012). This advent of NGS has made genome and transcriptome sequence easily accessible at an affordable cost and in quick time. Recently it has been widely used in whole genome sequencing (WGS), epigenetic studies, transcriptome profiling, identification of non-coding RNAs (ncRNA) and protein binding sites (Yang et al., 2009).

RNA-seq has proven to be a powerful tool for the recent transcriptomics supplanting the hybridization-based microarray techniques. It increased the horizon of analysis as microarray is feasible only for the known genes, but RNA-seq can be applied for identification of novel
transcripts including the exonic regions and ncRNA. Among the other benefits, RNA-seq is highly sensitive; 2-10ng of RNA is sufficient for the routine protocol. Although the construction of the directional library is laborious, determination of the polarity of the transcript provides better insights on the gene expression level and overlapping transcripts (Parkhomchuk et al., 2009). The reads produced from these sequencers are very short, ranging from 35bp -500bp, depending on the type of the sequencing system. Millions of these reads are required to be meticulously assembled into a full-length transcriptome with the help of different algorithms. For the sequences, which have reference genome/transcriptome with completely assembled or closely related genome/transcriptome, the reference-based genome/transcriptome assembly can be applied. Software for the assembly is available as web-based tools and stand-alone tools, whereas if the reference genome is not available, the assembly process can be done by de novo assembly. Most of the software and bioinformatics tools use the De Bruijn graph-based approach for the de novo assembly. The reference-based and de novo assembly strategies can also be combined as a hybrid assembly strategy to derive more comprehensive and insightful information (Martin & Wang, 2011).

The NGS RNA-seq has contributed to many biological findings, including novel viral pathogen discovery, miRNA and siRNA discoveries, and RNA profiling at the nucleotide level resolution (Kehoe, Coutts, Buirchell, & Jones, 2014; Kutnjak et al., 2015; Routh, Ordoukhanian, & Johnson, 2012; Roy, Shao, Hartung, Schneider, & Bransky, 2013; Yan et al., 2010). So the application of NGS in transcriptomics and genomics is not limited by the potentiality of
sequencing the samples. However, more powerful and robust storage and analytical computing systems can broaden the horizon further.

1.6 Rationale and future prospects

The project is dedicated to answering preliminary questions regarding possible encapsidation of the host RNA by two of the members of Bromoviruses. Encapsidation is a very tightly regulated process with unbelievable viral RNA specificity. As explained in the literature review above, viral RNAs have many regions that contribute to signaling from the initiation of capsid assembly to the progeny virion. In case of BMV RNA3 particle, 3’ TLS initiates the CP-RNA interaction for the assembly followed by the interaction of ARM with the PE in the coding region of MP. Similar PE sequences are also known to be present in RNA1 and RNA2, which direct the specificity of viral RNA encapsidation (Choi & Rao, 2003). Despite the CP-RNA and CP-CP interaction-directed specific encapsidation, there are instances that prove the significant amount of host RNA encapsidated by different animal viruses and plant viruses.

The FHV and Nudaurelia capensis omega virus (NωV), bipartite +ssRNA viruses, have been shown to encapsidate host RNAs up to 1% of the total viral RNA. These results were consistent for both the authentic viruses and the virus-like particles (VLPs), so viral CPs can be induced to encapsidate non-viral RNAs along with the viral RNAs (Routh, Domitrovic, & Johnson, 2012a). Cucumber necrosis virus (CNV) coat proteins are also known to encapsidate the host RNA. CNV encapsidated host RNA from 0.09% -0.7% of the total viral genome. Majorities of the
RNAs encapsidated were reported to be ribosomal RNA (rRNA) followed by chloroplast RNA. CNV replicates by compartmentalization on the chloroplast membrane, so chloroplast as the source of host RNA seems to be justifiable. Similar incidence has been reported in TMV, which encapsidated 2-2.5% of the host RNA, chloroplast being the major source (Ghoshal, Theilmann, Reade, Maghodia, & Rochon, 2015).

Brief review of molecular biology of the Bromoviruses clearly implicates the specific roles of various regions of the viral RNA and CP in encapsidation and the infection cycle as a whole. So it is predictable that possible lag in the co-ordination of interactions due to mutation or cellular environment might be among the causes of the misencapsulation. Further, CPs are shown to auto-assemble in vitro if the physiological conditions are satisfied, so hypotheses can be derived such that, after a successful RNA replication in spherules, the failure to transport replicated RNA sufficiently to the site of assembly can lead to assembly of viral CPs with the available cellular materials (Routh, Domitrovic, & Johnson, 2012b). This hypothesis can be boosted with the finding that CP assembly initiates by the co-operative and competitive binding of the ARM to the 3’TLS. So, in lack of viral RNAs, cellular RNAs with secondary structure (stem loops) probably act as substrate for CPs to interact non-specifically and initiate encapsidation (Damayanti et al., 2003; Garmann et al., 2014). Also, lack of functional replicase is known to depreciate the RNA specificity in the encapsidation by both the replication-dependent and -independent ways (Annamalai & Rao, 2005b).

The possible encapsidation of host RNAs can be identified by the NGS RNA-seq and the reads from RNA-seq can be classified by aligning them with available genome databases, which
have been completely sequenced (Routh, Domitrovic, et al., 2012a). The reads obtained from virally encapsidated RNA sequencing can be assembled by reference assembly. The aim of this project is to determine non-viral RNA fragments, so de novo assembly might not be necessary. So, the reads that are not aligned to viral genomic RNA might be the co-encapsidated host RNA. These unmapped reads are then aligned to plant genome database by BLASTn tool to identify the sources in the host.

The ultimate goal of this project is to extend the knowledge of host RNA encapsidated to the possible virus-mediated HGT in plants. HGT in eukaryotes is one of a little-studied process, but a few instances have already been evident. Moreover, virus-mediated HGT is thought to be an extremely rare process in plant evolution. But it has been shown that early terrestrial plants like mosses acquired some of their genes from prokaryotes, fungi and viruses (Yue, Hu, Sun, Yang, & Huang, 2012). There is a great possibility that the viruses in that case can transfer genes from one host to the other in the process of infection. However, genes delivered to new host might or might not perpetuate to the offspring. So, one of the other dimensions of this project would be important for VLPs mediated by genetically modified plants (Keese, 2008).

In the long term, the project has potential to explore the new dimensions of host-virus interrelationship. Ni et al. has shown that the BMV RNA encapsidation can be affected by the host harboring the virus. In addition, viral CP has multiple functions in the regulation of the viral infection cycle. Efficient encapsidation is also known to be related to the systemic transfer of the virus. The steps of viral infection cycle are interlinked; this project thus might prove a stepping stone to decode the functions of CP (Weber & Bujarski, 2015). Finally, CCMV CP has been widely
used as a nano-cargo for delivery of different components into live hosts like for efficient drug delivery at DNA level (Mikkilä et al., 2014). Encapsidation flexibility of CCMV virions has also been used in other fields, for example as a nano-reactor for enzyme activity (Minten et al., 2011) and for epitope expression in subunit vaccine development (Hassani-Mehraban, Creutzburg, Heereveld, & Kormelink, 2015). These findings hint at the possible capacity of viruses to deliver genetic materials from one host to another, and the knowledge on factors affecting encapsidation, both viral and host, can be helpful in generating VLPs as cargos for higher molecular weight components.
CHAPTER 2

MATERIALS AND METHODS

2.1 Virus propagation

The virus propagation was made by previously described methods for Bromovirus isolation (Bujarski, 1998). Broad bean (Vicia faba) and cowpea (Vigna unguiculata) plants were grown in 10 cm by 10 cm pots under the controlled conditions in a growth chamber with the day and the night temperatures of 24º C and 18 º C, respectively. The 10-day- old seedlings of broad bean were inoculated with BBMV - Tunisia strains (BBMV-TS) mechanically, by dusting the leaves with carborundum powder. Similarly, 12-day-seedlings (before the emergence of secondary leaves) of cowpea were inoculated with CCMV- Type strain. Both the virus inoculation solutions were diluted 10 times with the inoculation buffer (0.01M NaH2PO4 and 0.01M Mgcl2 at pH 6.0). The leaves were gently sprinkled with water after inoculation to remove the excess carborundum residues. Both plant species were grown in the growth chamber with 16 hours of daylight at 22º C. The infected leaves of broad bean were harvested 15 days post-inoculation (dpi) and the cowpea leaves with chlorotic symptoms were harvested after 12 dpi. Leaves were quickly frozen in liquid nitrogen and stored at -80 º C for the further steps.
2.2 Virus extraction and purification

Virus purification was performed under stringent conditions from the frozen leaves infected with the respective virus in two different phases. For BBMV and CCMV, the above preserved leaves were used, whereas for BMV the barley leaves that were stored in freezer (pre-infected) were used to purify the virions in two steps. First, the sucrose-cushioned purification was applied to purify the virus from the chloroform-clarified solution. Then, the virus solution from the cushioned purification was subjected to sucrose gradient ultracentrifugation. The virus purification was performed by the previously described procedure for purification of the BMV (Chaturvedi, Jung, Gupta, Anvari, & Rao, 2012).

2.2.1 Sucrose-cushioned virus purification

The infected leaves were ground in the presence of liquid nitrogen before homogenizing it in 1 ml of the virus extraction buffer (0.5M NaAc, 0.08M MgAc, pH 4.5) per gram of the leaf sample. Freshly prepared 0.1M ascorbic acid and 1/100 volume of β-mercaptoethanol were added to the extraction buffer just before the homogenization for BBMV. The homogenized extract was centrifuged at 5000 RCF for 10 minutes to remove the insoluble leaf components as pellets. The supernatant was mixed with equal volume of chloroform and centrifuged at 12000 RCF for 15 minutes. This step was repeated for two more times. Chloroform dissolves inorganic components of plant cells along with proteins and this centrifugation does not remove the virus with the bottom chloroform layer. Chloroform is also known to reduce the affinity of RNAs to the ribosomes, such
that the potentially embedded ribosomal RNA is liberated to the aqueous supernatant alongside with the virus.

Clarified extract was centrifuged with a 20% (W/V) sucrose cushion. The sucrose solution was prepared in the virus extraction buffer. Twenty milliliters (ml) of the sucrose solution was layered in the Beckmann Ultra Clear Ultracentrifuge tube (25mm X 89mm). On top of the sucrose layer a 15 ml phase of the clear extract was overlaid cautiously; the tube was centrifuged at 25,000 RPM for 5 hours in Beckman SW28 rotor. The resultant transparent glassy white pellet was re-dissolved in the virus suspension buffer (10 X diluted extraction buffer) and soaked to dissolve overnight. Depending on the amount of the pellet, 100-200 µl of the virus suspension buffer was used in each tube.

Virus solution was then treated with DNases and RNases to remove any RNA and DNA, respectively that co-purified along with virus particles. Ten folds DNase Buffer (NEB) (10 mM Tris pH 7.6, 2.5 mM MgCl2, and 0.5 mM CaCl2) was added to the virus solution to the final concentration of 1X. To the solution, 20 U of DNaseI (NEB) and 0.5µg of RNaseA (Roche) per 100ul virions were added (Routh, Domitrovic, et al., 2012a). The reaction mixture was allowed to stand at room temperature for 2 hours. At the end, the reaction tubes were transferred to ice to stop the nuclease activity. Virus was concentrated from the reaction mixture by using Amicon® Ultra Centrifuge Filters UltraceI® - 100K (UFC810008). The filter tube was filled up to 4mL with virus suspension buffer, then the treated virus solution was added to the tube (not more than 1 mL), followed by centrifugation at 6500 RPM for 15 minutes. This filter wash was performed at least three times to remove the possible residues of the enzymes. Finally, virus in the filter was dissolved into half the initial volume with the same virus suspension buffer.
2.2.2 Sucrose gradient ultracentrifugation

Purified virus solution from the sucrose cushion (not treated with DNase and RNase) was further purified by more robust sucrose gradient ultracentrifugation. Gradients of sucrose from 10% to 40% were prepared in the virus suspension buffer and the virus solution was layered cautiously on the topmost layer of sucrose. Eight milliliters of each sucrose solution was overlaid in a descending order, with the 40% (W/V) sucrose being at the bottom followed by 30%, 20% and 10% sucrose layers. On top of the 10% sucrose, 1 ml of the sucrose-cushioned virus solution was layered carefully. Then the tube was centrifuged at 25,000 RPM for 3 hours in the Beckman SW28 rotor. At the end of the centrifugation, the distinct opalescent band of the virus was observed approximately in the zone between 20% and 30% sucrose, under the white light illumination, and the virus band was then collected by puncturing underneath with a syringe and 22G1 needle. The virus suspension was diluted by at least two times then subjected to the ultracentrifugation at 26,500 RPM for 3 hours in Beckman SW41Ti rotor. Finally, pure form of the virus was collected as a glassy pellet, which was dissolved by soaking in 200-300 µL of the virus suspension buffer.

The final step of the purification involved treatment with DNase and RNase to remove any nucleic acids that co-purified and attached to the virion surfaces. Here, the method used was the same as the one used in the case of sucrose-cushioned virus solution (see above).
2.3 Electron microscopy and SDS-PAGE

Purified virions were visualized by the dark field transmission electron microscopy (TEM) in the core microscopy facility at Northern Illinois University, Department of Biological Sciences. Virions were negatively stained with 1% uranyl acetate, and the stained virions were visualized at the magnification of 60K. The virions were diluted at least 50 times prior of visualization.

The virion solutions were subjected to the SDS-PAGE analysis to detect the respective coat proteins for each virion preparation. The 12% resolving polyacrylamide gel was prepared and casted between two glass plates followed by layering with a thin layer of butanol. Once the gel was solidified, the butanol layer was removed, and after proper rinsing, the 5% polyacrylamide stacking gel was poured on top of the resolving gel and the comb was set to make sample loading wells. The reagent mixture was prepared as shown in the appendix table A2.

Two microliters (µl) of each virion purification was mixed to 6 µl of the 5X SDS gel-loading buffer. The final volume was maintained to 30 µl; 100mM of freshly prepared dithiothreitol (DTT) was added to each reaction mixture. The reaction mixture was heated up to 100°C for 3 minutes and 30 µl of the samples were loaded in each well. One of the wells was loaded with 20 µl of the Colorplus pre-stained protein ladder (New England Biolabs, Inc.), as size reference. The gel was then run in Tris-glycine buffer for 3 hours at 80 mV until the sample crossed the stacking gel; then the voltage was increased to 110mV for next 5 hours. The gel was then stained overnight with Coomassie Brilliant Blue followed by eight washings with the destaining solution. The staining solution was prepared by dissolving 250mg/100 ml of the methanol: H2O:
acetic acid (5:4:1) solution. The destaining solution was a mixture of methanol and acetic acid in the ratio of 3:1. Finally, the bands were visualized under the white light illumination.

2.4 RNA purification

RNA was extracted from the purified virions by sodium dodecyl sulfate (SDS) lysis and purified by sequential treatment of phenol and chloroform. One half percent (W/V) SDS was used to lyse the virions along with 50% phenol and 0.5% RNA extraction buffer (0.5M glycine, 0.5M sodium chloride, 0.1M EDTA at pH 9.5). Thorough mixing of the solution by vortexing followed by centrifugation at 14,000 RPM for 4 minutes at 4 °C separated the insoluble virion components from aqueous supernatant containing RNA. Equal volume of the phenol: chloroform (5:1, pH 4.3) was added to the supernatant. The mixture was vortexed to mix the solution properly and it was centrifuged at 14,000 RPM for 4 minutes at 4 °C. The final clarification was made by mixing the supernatant from the above step with equal volume of chloroform followed by centrifugation under the same parameters of temperature and speed. Sodium acetate (0.3 M) was added to the clarified aqueous layer containing RNA to facilitate the precipitation by 70% (V/V) ethanol. After addition of the ethanol, the solution was incubated at -20 °C overnight to give enough time to precipitate RNA. The next day, it was centrifuged at 14,000 RPM for 30 minutes at 4 °C to pellet the viral RNA. Traces of ethanol were allowed to evaporate in laminar hood for 5-10 minutes with the blower on. Finally, the semi-transparent white RNA pellet was dissolved in RNase-free water. The following denaturing gel electrophoresis showed the characteristic tripartite RNA genome of the virus was intact. The RNA was stored at -80 °C.
2.5 Next Generation RNA sequencing (Next Gen RNA-seq)

Sequencing of the three purified viral RNA preparations (that of BMV, CCMA and BBMV) were performed on the Illumina Next Generation Sequencing platform in the Core Genomic facility, Research Resource Center, at the University of Illinois at Chicago. Library preparation was performed by Wafergen PrepX RNA-Seq for Illumina, which uses RNAse III digestion prior to the library preparation. The process generated a directional library, which was sequenced on an Illumina HiSeq2000 instrument.

2.6 Sequence analysis

Reads downloaded from the sequencing facility were processed by trimming the low-quality reads in dynamic trim (Dynamictrim.pl) with the quality score of p<0.05. The trimmed reads were sorted to remove insignificantly short reads; the cutoff value for the length of reads was set to 25 bp using the Solexa QA, LengthSort program (Cox, Peterson, & Biggs, 2010). The sorted reads were mapped to reference (MTR) in Geneious Pro v. 8.1.2. (Biomatters, Ltd, Auckland NZ) for the respective viral genomes. The reference genomes were retrieved from the National Center for Biotechnology Information (NCBI) nucleotide database, NC_004006.1 - NC_004008.1 for BBMV and AF325739.1 - AF325741.1 for CCMV. The reads that were not MTR in Geneious were exported as FASTA files and aligned with the references using BLASTn tool. The reads that were not mapped or aligned to reference viral genome were allocated as a separate set, which will be referred to as total non-viral reads (TNR) henceforth in this disquisition.
The TNR datasets were aligned to different reference databases from respective hosts in a sequential order to filter out host sources of reads as organelle (mitochondrial and chloroplast), ribosomal (rRNA), mRNA, genomic and transposable elements (TE).

As the *Vicia faba* genome has not yet been sequenced to completion, mitochondrial and chloroplast genome databases for *Glycine max*, *Medicago trunculata*, and *Lotus japonicum* (members of *Fabaceae*) were compiled as the reference mitochondrial and chloroplast genomes to align the BBMV TNR. Further, BBMV TNR were queried against rRNA and transcriptome from *Glycine max*, as it is well annotated and closely related to *Vicia faba*. Similarly, CCMV TNR was queried against *Glycine max* database for mitochondria, rRNA, and transcriptome, whereas the *Vigna unguiculata* reference was used for the chloroplast data.
CHAPTER 3

RESULTS

3.1 Systemic infection on broad bean and cowpea plants

Local symptoms of the CCMV infection were visible on the inoculated primary leaves of cowpeas 5 days post-inoculation (dpi). The infected leaves showed the characteristic reddish necrotic lesions. The onset of systemic chlorotic symptoms on secondary leaves was visible starting from 8 dpi, which continued to propagate. The leaves from infected plants were collected after 12 dpi when about 5-6 leaves per plant were seen with chlorotic mottling symptoms (Fig. 4A.). Some plants were allowed to grow for longer time, 21 dpi, to see the severity of symptoms (Fig.4B). Severe, late-phase symptoms showed complete chlorosis of leaves followed by beginning of the reddish necrotic lesion at the central region of the leaves around midribs.

Similarly, systemic symptoms were visible on the broad bean leaves after 8 dpi with BBMV. The infected plants were harvested after 15 dpi, when about 5-6 leaves per plant were seen with chlorotic mottles. The light yellowish chlorotic spots distributed all along the leaves are the characteristic symptoms of the BBMV infection on the broad bean seedlings (Fig.4C.). Unlike CCMV in cowpea, BBMV did not show necrotic lesions even after 21dpi; instead, the new leaves emerging after about 25 dpi exhibited reduced chlorotic mottling.
Figure 4. Symptoms on cowpea and broad bean seedlings by the CCMV and BBMV infection, respectively. A) CCMV-infected cowpea after 12 dpi. Upper leaves are systematically infected while secondary leaves reveal chlorotic symptoms and the lower leaves are the locally infected primary leaves with necrotic lesions. The necrotic lesions were evident from the 5 dpi of CCMV inoculation. B) CCMV-infected secondary cowpea leaves after 21 dpi showing severe chlorotic symptoms and necrotic lesions. C) Broad bean plant infected with BBMV after 15 dpi showing mottling of the secondary leaves. “Continued on the following page.”
3.2 Virus extraction and purification: cushioned pelleting and gradient purification

BBMV and CCMV from the broad bean and cowpea leaves, respectively, were purified vigorously in three distinct phases. First phase involved the removal of debris and cellular components followed by chloroform clarification, which removes plant cellular materials and potentially weakens the ribosome-RNA interaction to release RNA to the supernatant layer along with the virions. Second phase was to pellet virions by sucrose-cushioned ultracentrifugation. The chloroform-clarified virions were layered on top of the 20% sucrose solution (Fig.5) and ultracentrifugation at the 25,000 rpm pelleted the glassy white virions. Pelleted virions were dissolved in virus buffer and processed further in two different fractions. One of the fractions of
virion solution was subjected to DNase/RNase treatment and another one was further purified by sucrose gradient ultracentrifugation at 25,000 rpm. Gradient of sucrose from 10% - 40% was layered in the Beckman centrifuge tubes, on top of which the cushion-purified virus solution was layered carefully and subjected to the ultracentrifugation. Virus formed a distinct blue band visible under the white light illumination in the region between 20% - 30% of the sucrose (Fig. 5.). Finally the band was collected and pelleted by further centrifugation at 26,500 rpm. The dissolved virion solution was treated with DNases and RNases, followed by virion concentration using Amicon® Ultra Centrifuge Filters Ultracel® - 100K (UFC810008).

![Image of virion purification by sucrose gradient ultracentrifugation]

**Figure 5. Virion purification by sucrose gradient ultracentrifugation.** A) Sucrose gradient (10%-40%) layered with the virus solution on top. B) BBMV illuminated in the white light and C) CCMV illuminated in white light after finishing the sucrose gradient ultracentrifugation. “Continued on the following page.”
Figure 5. Continued.
3.3 Quality analysis of the virus purification

Quality analysis of the virion purification was performed by using transmission electron microscopy (TEM) followed by SDS-PAGE. Virion particle integrity and purity are crucial for this project as the project is dedicated to analyzing RNA encapsidated inside the virion. The TEM analysis showed intact virions, which ensures that the RNAs extracted in further steps are from the encapsidated virions (Fig. 6). Co-precipitation of the host factors with the virion is strictly undesired as it falsifies the final result, so the virion purity was verified by the SDS-PAGE. As a reference, all the gradient-purified virion samples were run adjacent to the respective sucrose-cushion-purified virions. Cushioned virions had clear evidence of additional peptides to the characteristic coat proteins (CPs), whereas only the CP bands were visible for gradient-purified virions (Fig. 7). This reduces the possibilities of contamination with host genetic materials that have been complexed with cellular proteins. For the purpose of negative control, some wells were loaded only with the loading buffer. The SDS-PAGE gel shows the typical CP bands in the white light illuminance, which is 21 KDa for BBMV, 20.5 KDa for CCMV and 20.2 KDa for BMV. For CCMV that has been purified with the sucrose cushion, only some additional protein bands of sizes ranging from 30 KDa to 80 KDa were visible, which could be a potential source for the presence of host genetic materials to get embedded and evade RNases and DNases treatments. Similarly, for cushion-purified BBMV, a faint band of 50-60 KDa was visible. Both the cushioned and gradient-purified BMV from barley showed no other extra bands than the CP, so the proper purification of virions is crucial and the procedure might be varying for different viruses from different hosts.
The extractions of the encapsidated RNA from virion preparations were performed by the phenol-chloroform purification of the virions that were lysed with 10% SDS solution. The purified RNA was analyzed by electrophoresis in denaturing agarose gels (1% agarose). The gel analysis ensured the presence of intact RNA 1, RNA2, RNA3 and sgRNA4 after purification (Fig. 8). BMV RNA was used as a positive control in lane 3.

A.

Figure 6. The dark field transmission electron microscopy (TEM) analysis of the purified virion solutions. A) BBMV B) CCMV. The diameter of the virion particles averages to 28nm. The images were taken by Lori Bross, the manager of the Core Electron Microscopy at Northern Illinois University, Department of Biological Sciences. “Continued in the following page”
Figure 6. continued
Figure 7. SDS-PAGE analysis of the virion preparations. Lane 1, Lane 2, Lane 7, Lane 9 and Lane 11 were loaded only with the loading buffer as negative controls. Lane 3, Lane 5 and Lane 8 were loaded with sucrose-cushion-purified BMV, BBMV and CCMV, respectively. Lane 4, Lane 6 and Lane 10 were loaded with sucrose-gradient-purified BMV, BBMV and CCMV, respectively. BMV was used as positive control. As a size standard, the gel was also loaded with the Colorplus Pre-stained Protein Ladder, broad range (10-230KDa), New England Biolabs ® (lane M). All the sucrose-gradient-purified samples display only one distinct band at around 20KDa, in reference to the marker, whereas cushion-purified virions showed additional bands as well. The gel was stained with Coomassie-Blue protocol.
Figure 8. Denaturing agarose gel electrophoresis of viral RNA preparations (1% agarose gel). Lane 1 – CCMV RNA, Lane 2 – BBMV RNA and Lane 3 – BMV RNA (positive control). The largest bands represent RNA1 followed by RNA2 and RNA3. Farthest from the well, sgRNA4 is clearly visible in the gel as the smallest bands.
3.4 NGS-RNA sequencing and analysis of the read data

3.4.1 Summary of BBMV and CCMV RNA reads

The purified RNAs were sequenced on an Illumina HiSeq2000 instrument after the successful library preparation by using Wafergen PrepX RNASeq Library Kit at the DNA Services Facility University of Illinois, Chicago. The reads were 76 bp long or slightly shorter, with the mean length of 75.5 bp. The reads from both BBMV and CCMV viruses were mapped to their respective reference viral genome by using Geneious Pro v. 8.1.2. The unmapped reads were further run with the BLASTn program to find more viral reads that aligned to the reference genome. Then the reads that did not map in Geneious or align in BLASTn were separated as the total non-viral reads (TNR).

The total reads for BBMV after trimming and sorting the length were 121,512,620, out of which 101,847,536 originated from the viral genome. So the TNR for BBMV was 19,665,084, which was further queried against different host databases to classify the reads on the basis of their origin. Reference databases were derived from *Glycine max*, which belongs to the same taxonomic family as the *Vicia faba*, so they have a sufficient homology. Moreover, *Glycine max* is one of the most studied legumes and its genome has been well annotated; 103,361,910 reads were successfully assigned in this study for BBMV, out of which 101,847,536 (98.535%) did originate from the viral genome while 1,514,374 (1.465%) reads have been assigned as those of the host origin (Table 2.A).
Similarly, after trimming and sorting their length, total count of reads derived from CCMV were 108,911,947, out of which 104,959,847 reads originated from the viral genome. The TNR datasets value for CCMV was then 3,463,964, which was further queried against different reference databases in order to further categorize their origination. Overall, total reads assigned counted as 106,367,433, which included 104,959,847 (98.677%) of viral CCMV reads and 1,407,586 (1.323%) plant host-derived reads (Table 2.B).

**Table 2. Summary of NGS RNA-seq reads obtained for the two Bromoviruses. A) BBMV B) CCMV.**

*The percentages of viral reads and host reads were calculated based on the total assigned reads. TNR - total non-viral reads combined from both assigned and unassigned reads. “Continued on the following page”*

<table>
<thead>
<tr>
<th></th>
<th>BBMV</th>
<th>% Reads</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Reads</td>
<td>121,512,620</td>
<td>-</td>
</tr>
<tr>
<td>Assigned reads</td>
<td>103,361,910</td>
<td>100</td>
</tr>
<tr>
<td>Viral Reads</td>
<td>101,847,536</td>
<td>98.535</td>
</tr>
<tr>
<td>Plant reads</td>
<td>1,514,374</td>
<td>1.465</td>
</tr>
<tr>
<td>Unassigned</td>
<td>18,150,710</td>
<td>-</td>
</tr>
<tr>
<td>TNR</td>
<td>19,665,084</td>
<td>-</td>
</tr>
</tbody>
</table>
3.4.2 Detailed characterization of reads originated from the host plant sources

As explained above, the TNR datasets were queried against the respective host reference databases by using the program BLASTn to find the origination in the host genome. The first query was made against the mitochondrial genome followed by the chloroplast genome. For BBMV, the reference dataset was compiled from three closely related species, *Glycine max*, *Medicago trunculata*, and *Lotus japonicus*. The hits for mitochondrial origin were the lowest among all other categories with only 17 reads (0.001%). The rest of the TNR reads, not of a mitochondrial origin, were then run in BLASTn against the chloroplast genome database concatenated from the above three plant species; 218 reads (0.014%) were found to originate from the chloroplast genome. So, out of the 1,514,374 co-encapsidated host plant reads, 235 reads originated from the plant organelles. Further, the highest number of reads were found to be derived from the nuclear rRNA,

<table>
<thead>
<tr>
<th></th>
<th>CCMV</th>
<th>% Reads</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Reads</td>
<td>108,911,947</td>
<td></td>
</tr>
<tr>
<td>Assigned reads</td>
<td>106,367,433</td>
<td>100</td>
</tr>
<tr>
<td>Viral Reads</td>
<td>104,959,847</td>
<td>98.677</td>
</tr>
<tr>
<td>Plant reads</td>
<td>1,407,586</td>
<td>1.323</td>
</tr>
<tr>
<td>Unassigned</td>
<td>2,544,514</td>
<td></td>
</tr>
<tr>
<td>TNR</td>
<td>3,463,964</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Continued.

B)
which sums up to 959,785 reads (63.378%), followed by nuclear mRNA, 533,814 reads (35.250%). The remaining TNR was queried against the genome of the *Glycine max* to see if any other genomic RNA hits could be found, and consequently, there were 20,540 hits (1.356%) of this plant, which indicates that other types of genomic RNA could also be encapsidated by the BBMV virions.

Similarly, the analyses was made for CCMV to find different categories of reads originating from the host. The reference database for mitochondria, rRNA, mRNA and genomic RNA was derived from *Glycine max*, whereas the plastid genome was available for *Vigna unguiculata*. Twenty-five reads (0.002%) were found to originate from the plant mitochondria, which is the lowest number of reads among all assigned host categories; 1462 reads (0.104%) hit with the chloroplast genome. rRNA had hits with the highest number of reads, 755,476 (53.672%), followed by mRNA, 637,778 reads (45.310%) and 12,845 reads (0.913%) originated from other types of genomic RNA. The total hits related to the organellar RNAs sum up to reach 1,478 for CCMV.

The trend of co-encapsidation of host RNAs by BBMV and CCMV seems to be comparable. Both the virions encapsidated mitochondrial RNA least efficiently followed by chloroplast, genomic RNA, mRNA and rRNA. So the data signifies the common features of the RNA encapsidation specificity for the two viruses; also, since the fractions of different RNA categories co-encapsidated are comparable, the virion multiplication and packaging most likely take place at approximately the same subcellular location. The detailed comparative illustration for the co-encapsidated plant RNAs is shown in Table 3.
Table 3. Summary of plant-originated co-encapsidated reads in the two Bromoviruses. TNR datasets were classified into various plant source origins on the basis of the BLAST search against the respective reference database. The upper half of the table, from left to right, represents number of reads for BBMV and CCMV that had hits against the respective reference database. The lower half represents the percentage conversion of the reads among the total plant reads assigned.

<table>
<thead>
<tr>
<th>Plant source</th>
<th>No. of BBMV Reads</th>
<th>No. of CCMV Reads</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mitochondrial</td>
<td>17</td>
<td>25</td>
</tr>
<tr>
<td>Chloroplasts</td>
<td>218</td>
<td>1,462</td>
</tr>
<tr>
<td>Ribosomal RNA (rRNA)</td>
<td>959,785</td>
<td>755,476</td>
</tr>
<tr>
<td>mRNA</td>
<td>533,814</td>
<td>637,778</td>
</tr>
<tr>
<td>Genomic</td>
<td>20,540</td>
<td>12,845</td>
</tr>
<tr>
<td>Total plant reads</td>
<td>1,514,374</td>
<td>1,407,586</td>
</tr>
<tr>
<td>Total Organelle Reads</td>
<td>235</td>
<td>1,478</td>
</tr>
<tr>
<td>Putative transposable Elements (TE)</td>
<td>321,478</td>
<td>495,742</td>
</tr>
</tbody>
</table>

Plant reads in Percentage

<table>
<thead>
<tr>
<th>Plant source</th>
<th>BBMV %</th>
<th>CCMV %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mitochondrial</td>
<td>0.001</td>
<td>0.002</td>
</tr>
<tr>
<td>Chloroplasts</td>
<td>0.014</td>
<td>0.104</td>
</tr>
<tr>
<td>ribosomalRNA (rRNA)</td>
<td>63.378</td>
<td>53.672</td>
</tr>
<tr>
<td>mRNA</td>
<td>35.250</td>
<td>45.310</td>
</tr>
<tr>
<td>Genomic</td>
<td>1.356</td>
<td>0.913</td>
</tr>
<tr>
<td>Total</td>
<td>100.000</td>
<td>100.000</td>
</tr>
<tr>
<td>TE</td>
<td>21.228</td>
<td>35.22</td>
</tr>
</tbody>
</table>
3.4.3. The co-encapsidated transposable elements (TE)

One of the most interesting findings of this study is the encapsidation of TEs by both BBMV and CCMV. The mRNA references that had hits with TNR were pooled out by using bbmap program; then the pooled mRNA reads were queried against the TE database for *Glycine max*. So this approach gave the data for putative transposons among the host mRNA that were co-encapsidated by BBMV and CCMV: 321,478 reads (21.228%) were found to originate from the putative host TE in BBMV, whereas 495,742 reads (35.22%) were found for CCMV (Table 3.).

The TE sequences encapsidated by virions were further dissected to find their specific classes. Both the virions encapsidated mutators, class-II TE DNA transposons, most efficiently (495,665 reads for BBMV and 321022 reads for CCMV). Among other transposons found there were miniature inverted repeat transposable elements (MITEs) and CACTA superfamilies. The encapsidation of retrotransposons was less efficient in both the viruses; however, the comparison of two viruses showed BBMV to be more efficient with encapsidating retrotransposons. The detailed breakdown of the number of reads for TE is explained in Table 4A and Table 4B.
Table 4. Categories of transposable elements encapsidated by the two Bromoviruses. The table shows different types of transposable elements co-encapsidated by A) BBMV B) CCMV. TE are classified as transposons and retrotransposons. Transposons are further classified to superfamilies, CACTA, Mutators and Mites, whereas retrotransposons are classified to copia-like and LINE.

<table>
<thead>
<tr>
<th>Transposable elements</th>
<th>No. of reads</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retrotransposons, copia like</td>
<td>228</td>
</tr>
<tr>
<td>LINE</td>
<td>20</td>
</tr>
<tr>
<td>Mutators</td>
<td>321022</td>
</tr>
<tr>
<td>CACTA</td>
<td>131</td>
</tr>
<tr>
<td>Mites</td>
<td>77</td>
</tr>
<tr>
<td>Total Transposons</td>
<td>321250</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Transposable elements</th>
<th>No. of reads</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retrotransposons, copia like</td>
<td>7</td>
</tr>
<tr>
<td>Mutators</td>
<td>495665</td>
</tr>
<tr>
<td>CACTA</td>
<td>29</td>
</tr>
<tr>
<td>Mites</td>
<td>43</td>
</tr>
<tr>
<td>Total Transposons</td>
<td>495737</td>
</tr>
</tbody>
</table>

Based on the above results the current study shows that despite the stringency in the viral RNA selectivity during genome packaging, there are still possibilities for the encapsidation of significant amounts of the host RNA. This indicates that various viral and host factors might have roles in the encapsidation specificity. Moreover, encapsidation of the host transposable elements exhibits the immense promise for the requirement of further studies on this topic. Conclusively,
BBMV and CCMV are found to co-encapsidate host RNA in a manner comparable to each other in respect of types of RNA as well as the efficiency of encapsidation.
4.1 Host RNAs are encapsidated by BBMV and CCMV

These studies demonstrate that BBMV and CCMV encapsidated 1.465% and 1.323% of the host RNA, which is comparable to the previous finding that 1% of the host RNA is encapsidated by FHV (Routh, Domitrovic, et al., 2012a), whereas it is slightly higher than the data reported in CNV, which accounts for 0.07-0.9% of the encapsidated RNA to be the host RNAs (Ghoshal et al., 2015). This differences hints at the altered encapsidation efficiency among different virus families. The percentages of reads originating from plants and viruses were calculated out of the total reads assigned. For the reads originating from different plant sources, the percentage conversion was made out of the total plant-originated reads. Similarly, TEs were identified among the assigned mRNA-originated reads, and the percent conversion was made out of the total plant-originated reads. In this study, the efficiency of the host RNA encapsidation by both viruses is comparable. The Pearson correlation coefficient for the host RNAs encapsidated by BBMV and CCMV is 0.98 with the degree of freedom of 6 and the level of significance of p<0.001. The correlation coefficient shows that the categories of host RNAs encapsidated by the
two viruses are highly comparable. Inference can be drawn that these two *Bromoviruses* might share the mechanism and subcellular localization of the virion genome packaging.

### 4.1.1 Organellar RNA is co-encapsidated the least by two *Bromoviruses*

The least number of reads were found to have sequence homology with the mitochondrial RNAs followed by the chloroplast ones in both the viruses. Combined percentages of reads account for below 0.015%, which seems to be different from the findings from Ghoshal et al. (2015) in authentic CNV virions. The authors have found that chloroplast RNA was most efficiently encapsidated by the virions. The discrepancies in our findings can be accounted for by differences in sites of the virus multiplication for CNV versus *Bromoviruses*. Namely, for CNV in *N. benthamiana*, genome packaging and virion multiplication are known to take place near the chloroplast membrane, whereas for *Bromoviruses*, viral RNA replication takes place in spherules that are formed as distinct subcellular compartments in the cytoplasm near the endoplasmic reticulum (ER) membrane (Diaz et al., 2015).

Host ribosomal RNAs seem to be the most efficiently co-encapsidated RNA by both *Bromoviruses*, which account for 63.375% and 53.672% of the total host RNA co-encapsidated by BBMV and CCMV, respectively. Followed by rRNA, the host nuclear mRNAs have the highest chances of being co-encapsidated by the virions: 35.250% and 45.310% of the total host RNA co-encapsidated by BBMV and CCMV, respectively, involve mRNAs. This finding is in correlation with the findings from Routh et al. for FHV virions. The authors found mRNA and rRNA to be the most predominant RNAs encapsidated by the virions that accounted for 1.1 % of the total
assigned reads (Routh, Domitrovic, et al., 2012a), which is very similar to this study. Further, the result can be correlated with the site of the viral RNA replication and the host factors involved in viral RNA replication and in packaging.

Viral replication takes place in the cytoplasm, in the spherules. Spherules are known to be similar to multivesicular bodies (MVB) morphologically, and components of the endosomal sorting complex that is required for transport (ESCRT) have been shown to be required for the spherulation as well. Primary protein interaction to initiate the spherule formation is known to occur between the host reticulon (Rtn) and viral 1a protein. Further, other host components like snf7 and Vsp ATPase complex are required for initiation of replication. These ESCRT components also have some roles in translation of mRNA (Ghoshal et al., 2015), so this indicates that some host RNAs might also have access to the spherules along with viral RNAs. Since the encapsidation of viral RNAs is believed to take place in the close vicinity of the spherules and replication is coupled to the packaging, the findings from this study seem justifiable that nuclear rRNA and mRNA are the primary targets for co-encapsidation by the virions.

The above discussion can be further justified with the evidence that in the vicinity of endoplasmic reticulum (ER) membrane the host mRNA and rRNA are supposed to be abundant in light of the fact that endosomal sorting complex is known to be actively involved in cellular RNA localization. For example, transitional ER are known to play roles in transporting gurken mRNAs to the specific location in Drosophila oocytes (Herpers & Rabouille, 2004). So, the significant amount of host RNAs around the spherules during viral RNA replication might also be taken up by the viral coat proteins for encapsidation. This assumption is plausible since RNA encapsidation
is known to be directed by the structure-specific interaction of RNA with CP in addition to the sequences themselves. In BMV, depleted RNA encapsidation efficiency due to loss of 3’ TLS was shown to be compensated by supplementing with a cellular tRNA-like structure (Annamalai & Rao, 2007). Role of CP in maintaining the particle homogeneity could be another explanation that supports the co-encapsidation of host RNA. The N-terminal basic peptide tail in CP interacts with negatively charged genomic RNA to encapsidate it; alteration of the amount of positive charges in CP significantly altered both the amount and length of the RNAs encapsidated without any distinguishable viral particle polymorphism (Ni et al., 2012). So, if viral RNA synthesis is defective or not adequate viral RNA is not available, the CP might eventually pick up any amount of cellular RNAs to balance the electrostatic requirement for the capsid assembly.

4.1.2 Co-encapsidated transposable elements: a potential source of horizontal gene transfer (HGT)

One of the most interesting findings of this study is that the host transposable elements are encapsidated by the virions. In previous studies by Ghosal et al. and Routh et al., the significant numbers of transposable elements were reported to be encapsidated in CNV and FHV, respectively. In this study, 321,478 BBMV total non-viral reads (TNR) were found to be originated from the host mRNAs for TEs, whereas 495,742 reads from CCMV TNR had hits with the host mRNAs bearing TEs. The most abundant type of encapsidated TE is mutator, a DNA transposon, for both the virions. This seems obvious in light of the fact that mutators are among one of the most widely distributed transposons known in plants. Most importantly, mutators are also known to be autonomous elements and contribute at the elevated frequency of the spontaneous mutations
in the maize genes (Benito & Walbot, 1997). The horizontal transfer of transposon (HTT) of mutators has been reported in a wide range of the members in the grass family although the exact mediator is not known (Diao et al., 2005). Thus, the encapsidation of TEs by virions shows the potential as a mediator for the HGT. Also, genomic studies in the recent years have proven that TEs have a significant contribution to the eukaryotic genome structure, function and evolution. Moreover, a hint of retrotransposons among those TEs makes one of the hypotheses in this study about possible roles of virions in HGT more robust and plausible. Retrotransposons carry the coding regions for integrase, reverse transcriptase and polymerase activities so that these elements can integrate with the host genomic DNA. Whole genome sequencing has also revealed that eukaryotic genomes have an abundant number of reverse transcriptase coding regions, perhaps higher than other protein coding sequences. Moreover, other non-LTR retrotransposons and other transposons can then be also readily integrated and reverse transcribed into DNA (Finnegan, 2012).

The uptake of host cellular RNAs including transposons is also significant in terms of the viral genome evolution. Highly frequent spontaneous mutations are known to be major causes of the rapid RNA genome evolution (Holland et al., 1982). Based on this study and previous data on the host RNA encapsidation (in FHV and CNV), speculation can be made that these encapsidated host RNAs also might have some roles in viral genome evolution. The fragments of the encapsidated host RNAs can be incorporated into the viral RNA genome, giving rise to new or improved strains.

The expansion of this project in future can be substantial to the scientific community. VLPs have been widely used as tools in biomedicine for gene therapy and nano-studies (Verwegen &
Cornelissen, 2015). Especially, CCMV CPs have been used as vessels in nano-reactions for their flexibility in encapsidation (Hassani-Mehraban et al., 2015). The amount of particles encapsidated is contributed by the electrostatic interaction between positive charges on CP residues and negative charges on the encapsidated particles. So, if the factors influencing the viral encapsidation and detailed mechanism are known, then more efficient CP modifications can be introduced to generate VLPs with higher capacities (Cadena-Nava et al., 2012). Further, the results of this investigation show the comparable pattern and efficiency of the host and viral RNA encapsidation for BBMV and CCMV, so there are possibilities that BBMV has the similar potential as CCMV to be used as VLPs.

4.2 Conclusion

On the basis of findings in this study, we can draw inference that virions are capable of encapsidating host RNAs with differing efficiency for different cellular RNAs. The efficiency might be dependent on the intracellular localization of viral replication and packaging. Interestingly, virions also encapsidate TEs, which are potential mediators of HGT. Studies on HGT in plants are still not much advanced, but in light of the recent whole-genome studies there is adequate evidence of HTT in plants (El Baidouri et al., 2014), so it is plausible that viruses may serve as potential carriers for the HGT in plants through (or via) HTT.

Further studies on the project are definitely needed to jump into the above conclusions more robustly. The findings and results irrefutably support the discussions above. More
replications and studies of other members of *Bromoviridae* might be helpful in generating more firm conclusion


APPENDIX
DETAILED PROTOCOLS AND TABLES
Protocol A1. Virion extraction and purification from the infected leaves

Protocol A1.1 Sucrose-cushioned purification:

1) Collected 50 gm of the virus infected leaves and grinded in mortar pestle by the addition of 50 ml of the extraction buffer (0.5 M NaAc; 0.08 M MgAc, pH 4.5 ) For BBMV, 1/100 the volume of β-mercaptoethanol and 0.1M ascorbic acid was added to the extraction buffer just before grinding.

2) Centrifuged the mixture at 5000 RCF for 10 minutes to pellet insoluble cellular debris.

3) Collected the supernatant and added equal volume of the chloroform solution and vortexed 20 seconds to mix properly.

4) Centrifuged the mixture at 12,000 RCF for 15 minutes.

5) Collected the supernatant and repeated chloroform clarification.

6) The supernatant was stirred for 30 minutes in a magnetic stirrer to remove traces of chloroform from the supernatant.

7) Supernatant transferred to the ultracentrifuge tubes with 20% sucrose as a cushion. 15 ml of the supernatant overlaid to the 20 ml of the 20% sucrose prepared in the extraction buffer.

8) Centrifuged the tubes at 25,000 RPM for 5 hours using the Beckman SW32 rotor.

9) The glassy white pellets were dissolved in 500ul of the virus buffer (10X dilution of the extraction buffer).
Protocol A.1.2 Sucrose gradient purification:

1) Sucrose solutions from 10% sucrose to 40% sucrose was prepared in the virus buffer.
2) Layered carefully in an ultracentrifuge tube, 10% sucrose at the bottom, then 20%, 30% and 40% on the topmost layer. Incubated at 4°C overnight to homogenize the gradient.
3) Layered 1ml of the cushioned virion purification on top of the gradient.
4) Centrifuged at 25,000 RPM for 3 hours using Beckman SW32 rotor.
5) Virus separated as a bluish band in the white light illumination at the zone between 20-30% sucrose.
6) Ultracentrifuge tube was punctured just underneath the virus band and collected by using a 1ml syringe and 22 G1 needle.
7) Added equal volume of the virus buffer to the sucrose containing virus solution, then centrifuged at 25,000 RPM for 3 hours.
8) Glassy virus pellet was dissolved in 400 µl of the virus buffer.

Protocol A2: RNA extraction and purification from the purified virions

1) Transferred 200 µl of the purified virion solution to a clean Eppendorf tube, added 25 µl 10% SDS (w/v), 25 µl 10X RNA extraction buffer (0.5M glycine, 0.5M sodium chloride, 0.1M EDTA, pH 9.0), and 250 µl phenol.
2) Vortexed for 20 sec.
3) Centrifuged for 4 min at 14,000 RPM and transferred the supernatant to the new tube then added 250 µl of the phenol chloroform solution.

4) Vortexed for 20 sec, then centrifuged at 14,000 RPM for 4 min (4°C).

5) Transferred the clear supernatant to a new tube and added 250 µl of the chloroform.

6) Vortexed for 20 sec, followed by centrifugation at 14,000 RPM for 4 min (4°C).

7) Transferred supernatant to the new tube, added 1/10 th volume of 3M sodium acetate followed by addition of 2.5 volume of chilled ethanol.

8) Incubated the tube overnight at -20 ºC.

9) Centrifuged the tubes at 14,000 RPM for 25 min at 4 ºC. The RNA pellet was dissolved in 10 µl of RNase free water. The concentration of RNA was determined by the Nanodrop analysis.

**Protocol A3. Denaturing RNA gel electrophoresis**

1) Dissolved 0.5 g of agarose in 36 ml of the distilled water by boiling for 1 minutes.

2) Added 4ml of the 3-(N-morpholino) propane sulfonic acid (MOPS) (0.2M MOPS, 10 mm EDTA, 50 mm NaoAc, pH 7.0) and 3 ml of the formaldehyde.

3) Maintained final volume to 50 ml by adding distilled water. Heated for 15 seconds.

4) Casted in the gel tray with a comb.

5) RNA sample preparation was done as shown in the Table A1.
**Table A 1. Sample preparation for denaturing RNA gel electrophoresis.** 1200 ng of RNA was used for each sample. The gel was run at 110 mv for 180 minutes.

<table>
<thead>
<tr>
<th>Reagents</th>
<th>Volume/sample (µl)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formaldehyde</td>
<td>5.5</td>
</tr>
<tr>
<td>Formamide</td>
<td>15</td>
</tr>
<tr>
<td>MOPS</td>
<td>15</td>
</tr>
<tr>
<td>EtBr</td>
<td>1.5</td>
</tr>
<tr>
<td>H2O</td>
<td>4.6 ( 1200 ng RNA)</td>
</tr>
<tr>
<td>RNA</td>
<td></td>
</tr>
<tr>
<td>RNA dye</td>
<td>3</td>
</tr>
</tbody>
</table>

Incubate at 65°C for 15 min

**Table A 2. Summary of the preparations for Tris-glycine SDS-PAGE.** Upper half - 12% resolving gel and Lower half- 5% stacking gel. TEMED was added to the reagent mixture just before casting the gel.

<table>
<thead>
<tr>
<th>Resolving gel</th>
<th>Vol (20ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solutions</td>
<td></td>
</tr>
<tr>
<td>H2O</td>
<td>6.6</td>
</tr>
<tr>
<td>30% acrylamide mix</td>
<td>8</td>
</tr>
<tr>
<td>1.5M Tris (pH 8.8)</td>
<td>5</td>
</tr>
<tr>
<td>10% SDS</td>
<td>0.2</td>
</tr>
<tr>
<td>10% Ammonium persulfate</td>
<td>0.2</td>
</tr>
<tr>
<td>TEMED</td>
<td>0.008</td>
</tr>
</tbody>
</table>

Stacking gel | Vol (10ml) |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Solutions</td>
<td></td>
</tr>
<tr>
<td>H2O</td>
<td>6.84</td>
</tr>
<tr>
<td>30% acrylamide mix</td>
<td>1.7</td>
</tr>
<tr>
<td>1.0 M Tris (pH 6.8)</td>
<td>1.25</td>
</tr>
<tr>
<td>10% SDS</td>
<td>0.1</td>
</tr>
<tr>
<td>10% Ammonium persulfate</td>
<td>0.1</td>
</tr>
<tr>
<td>TEMED</td>
<td>0.01</td>
</tr>
</tbody>
</table>