



Sveriges lantbruksuniversitet  
Swedish University of Agricultural Sciences  
Faculty of Natural Resources  
and Agricultural Sciences

# **Complete genome characterization of Barley yellow dwarf virus-OYV**

**Kanitha Sathees**

**Independent project in Biology, 30 hp, EX0732**

**Examensarbete / Institutionen för växtbiologi, SLU**

**Uppsala, 2016**

**Nr: 153**

**Plant Biology-Master's programme**



# Complete genome characterization of Barley yellow dwarf virus-OYV

*KANITHA SATHEES*

Supervisor: Anders Kvarnheden  
Swedish University of Agricultural Sciences (SLU),  
Department of Plant Biology

Examiner: Johan Meijer  
Swedish University of Agricultural Sciences (SLU),  
Department of Plant Biology

Credits: 30 hp

Level: A1E

Course title: Independent project in Biology

Course code: EX0732

Programme: Plant Biology-Master's programme

Place of publication: Uppsala

Year of publication: 2016

Title of series: Examensarbete / Institutionen för växtbiologi, SLU

Nr: 153

Online publication: <http://stud.epsilon.slu.se>

**Keywords:** complete nucleotide sequence, genome organization, Barley yellow dwarf virus-OYV (BYDV-OYV), Luteovirus, recombination



## **POPULAR SCIENCE:**

Virus infections cause diseases in crop plants that result in yield and quality losses in crop production. Barley yellow dwarf (BYD) is one of the most important diseases of cereals worldwide and is caused by viruses of several closely related species. All cereals grown in Sweden are susceptible to BYD. In an earlier virus diversity study, a new tentative species, BYDV-oat yellowing virus (BYDV-OYV), was discovered in Latvia. This species has also been discovered in different parts of Sweden. This project was designed to investigate the origin of BYDV-OYV and its relationship with other species of BYD-associated viruses. Using various molecular techniques an almost complete genome sequence of BYDV-OYV was obtained. Further sequence analysis showed that BYDV-OYV differ significantly from other BYD-associated viruses according to standardized molecular diversity criteria. BYDV-OYV can thus be regarded as a distinct species and seem to have formed through a cross between two related BYD-associated viruses. The information can be used for better understanding of the biology and evolution of these viruses in different environments and ultimately develop strategies for improved plant protection.

## Abstract

Barley yellow dwarf-associated viruses (BYD-associated viruses) belong to the family *Luteoviridae* and are of high economical relevance for cultivation of cereals. BYD-associated viruses are phloem-restricted pathogens that cause yellowing and reddening of leaves, plant dwarfing and reduction in size and number of grains. The viruses are transmitted only by aphids and have a genome of single-stranded positive sense RNA (+ssRNA). BYDV-oat yellowing virus (BYDV-OYV) was discovered in Latvia in a study on molecular variability of the coat protein gene (CP gene) of BYD-associated viruses from Latvia and Sweden. The virus was proposed to be a new species within the family *Luteoviridae*. Recently, BYDV-OYV has also been found in Sweden. In my previous study, a partial genome sequence, which comprised 2,792 nucleotides (nt) of the 3' end of BYDV-OYV, was determined to investigate the relationship with other isolates of BYD-associated viruses. A phylogenetic analysis with the partial genome sequence of BYDV-OYV and other isolates of family *Luteoviridae* showed that BYDV-OYV is a new variant within the family *Luteoviridae*. BYDV-OYV shared higher nucleotide identity with the groups PAV-IIIa and PAV-IIIb than to the groups PAV-I and PAV-II. In the current study, a molecular investigation of BYDV-OYV was carried out by determining also the 5' part of the genome sequence to explore genome organization, relationship with other viruses of family *Luteoviridae*, recombination events and taxonomic classification. The total determined sequence contained 5,132 nt and covered an almost complete genome except the un-translated region 1 (UTR1), the 5' end of open reading frame 1 (ORF1) and the 3' end of UTR4. The genome organization of BYDV-OYV is similar to other members of the genus *Luteovirus*. BYDV-OYV differs enough in sequence from other isolates of the genus *Luteovirus* to be regarded as a separate species. Overall genome sequence identity was only 72 - 83 % between BYDV-OYV and other viruses of genus *Luteovirus*. BYDV-PAS-129 (GenBank accession no AY218796) belonging to group PAV-II, had the highest nucleotide identity (83 %). Nucleotide identities of individual ORFs and UTRs showed more complex relationship between BYDV-OYV and other viruses of genus *Luteovirus* and ranged from 50 - 97%. ORF1, ORF2 and UTR2 are more closely related to those of BYDV-PAS-129, while other regions are more closely related to those of BYDV-PAV-CN and BYDV-PAV isolate 05ZZ1 in group PAV-III. One putative recombination event was detected when the genome sequence of BYDV-OYV was analysed together with 12 other isolates of the genus *Luteovirus*. The recombinant seem to originate from BYDV-PAS-129 (minor) and BYDV-PAV-CN (major) with the recombinant region starting and ending at nt positions 772 and 1,013, respectively. The genome of BYDV-OYV has thus probably evolved through recombination between ancestors of BYDV-PAS-129 and BYDV-PAV-CN. BYDV-PAS-129 also seems to be a recombinant with a virus related to BYDV-OYV as minor parent. The recombination event in the genome of BYDV-OYV was detected within the ORF1 and ORF2 region, which is a hot spot in the evolution of the family *Luteoviridae*.



## Table of contents

1. Introduction .....	8
1.1. BYD-associated viruses .....	9
1.2. Current classification of BYD-associated viruses .....	9
1.3. A new tentative species of BYDV-PAV .....	10
1.4. Importance of a complete genome sequence .....	10
1.5. Genome organization and gene functions of family <i>Luteoviridae</i> .....	11
1.6. Relationship between family <i>Luteoviridae</i> genomes.....	12
2. Aims of the study .....	14
3. Materials and Methods .....	15
3.1. Plant material.....	15
3.2. Primer design.....	15
3.3. Immunocapture-reverse transcription-polymerase chain reaction (IC-RT-PCR), gel extraction and purification.....	17
3.4. Cloning of purified DNA fragments 1-3 .....	18
3.5. Overnight bacterial culture .....	18
3.6. Restriction analysis of purified plasmid DNA and sequencing.....	19
3.7. 5'/3' rapid amplification of cDNA end (5'/3' RACE) .....	19
3.8. Sequence analyses and phylogenetic analyses .....	22
3.9. Recombination detection .....	23
4. Results .....	25
4.1. PCR of fragments 1-3.....	25
4.2. Gel extraction and purification.....	27
4.3. Restriction analysis of purified recombinant plasmids for fragments 1-3.....	27
4.4. 5' RACE.....	28
4.5. Gel extraction and purification.....	29
4.6. Restriction analysis of recombinant plasmid for fragment 4.....	30
4.7. 3' RACE.....	31
4.8. Restriction analyses of recombinant plasmid of 3' end of BYDV-OYV genome .....	31
4.9. Genome organization of BYDV-OYV .....	32
4.10. Comparisons with other luteoviruses .....	34
4.11. Phylogenetic analysis .....	35



4.12. Evidence for recombination of BYDV-OYV .....	35
5. Discussion .....	38
References .....	41
Acknowledgements .....	47

## 1. Introduction

RNA viruses are common pathogens of all organisms including humans, animals and plants. Plant-infecting viruses may affect the fitness of infected plants by reducing their survival (Sacristan et al., 2005), yield or competitive ability with healthy plants. The viruses also show high evolutionary potential through a high mutation rate and high rate of recombination (Elena & Sanjuan, 2008). One of the best examples for plant RNA viruses, which have these features, is barley yellow dwarf-associated viruses (BYD-associated viruses). Oswald and Houston identified barley yellow dwarf virus (BYDV) in 1951 (Oswald & Houston, 1951). Since then, BYD-associated viruses have been recognized as being among the most economically important and damaging viruses of cereals and fodder grasses and are found in every cereal-growing area in the world (Irwin & Thresh, 1990; D'Arcy, 1995; Power & Remold, 1996; Nega, 2014). BYD-associated viruses are members of the genera *Luteovirus* and *Polerovirus* in the family *Luteoviridae* (Domier, 2012). The causal agents of barley yellow dwarf disease are only transmitted by aphids in a circulative and non-propagative manner (Gray & Gildow, 2003; Fiebig et al., 2004; Hogenhout et al., 2008). Initially, BYD-associated viruses were classified as five different strains by their aphid vector specificities (Rochow, 1969; Rochow & Muller, 1971). More recently the viruses have been classified into species within the family *Luteoviridae*. They are denoted as *Barley yellow dwarf virus-PAV* (BYDV-PAV), *Barley yellow dwarf virus-MAV* (BYDV-MAV), *Cereal yellow dwarf virus-RPV* (CYDV-RPV), *Barley yellow dwarf virus-RMV* (BYDV-RMV), *Barley yellow dwarf virus-SGV* (BYDV-SGV) and *Barley yellow dwarf virus-GPV* (BYDV-GAV) (Domier, 2012). Among these species, BYDV-PAV is the most economically important and widely distributed one. PAV isolates have been divided into two species, PAV and PAS, to reflect significant variation in the coat protein gene (CP gene) sequence (Mayo, 2002; Liu et al., 2007). A new tentative species, BYDV-OYV, has been discovered in Latvia and Sweden based on the CP gene analysis of BYD-associated viruses (Bisnieks et al., 2004; Eriksson et al., unpublished). In addition, a new species, BYDV-PAV-CN, has been identified in China (Liu et al., 2007). The new species were identified according to the recent agreement that viruses with a variation at the amino acid level exceeding 10 % for any viral gene product could be classified as belonging to separate species within the family *Luteoviridae* (Domier, 2012). New viruses of the family *Luteoviridae* have often evolved as a result of recombination

events. This is a major mechanism in virus evolution allowing viruses to evolve rapidly (Knierima et al., 2010; Sztuba-Solinska et al., 2011). Improved knowledge of recombination events within the family *Luteoviridae* enables a better understanding of how these viruses evolve and adapt to changes in natural ecosystems as well as in agricultural systems.

### **1.1. BYD-associated viruses**

Even though outbreaks of BYD were noted in the United States in 1907 and 1949, BYD in California was first shown to be caused by viruses in 1951 and then BYDV was subsequently found to have a worldwide distribution being able to infect nearly all members of the family *Poaceae* (Irwin & Thresh, 1990; D'Arcy, 1995; Power & Remold, 1996; Nega, 2014). BYD-associated viruses have a genome of single-stranded positive-sense RNA (+ssRNA). It is difficult to estimate the global yield losses due to the fact that these viruses induce symptoms that resemble the effects of other biotic and abiotic factors.

### **1.2. Current classification of BYD-associated viruses**

Recently, BYD-associated viruses have been classified as separate species in two genera based on the degree of identity between their nucleotide sequences (Fattouh et al., 1990; Domier, 1995). BYDV-PAV and BYDV-MAV are within the genus *Luteovirus*. The former designation as BYDV-RPV was changed to *Cereal yellow dwarf virus-RPV* (CYDV-RPV) and classified within the genus *Polerovirus* along with CYDV-RPS. BYDV-SGV, BYDV-GAV and BYDV-RMV have not yet been assigned to any genus (Domier, 2012). Recent analyses of the complete genome sequence of BYDV-RMV revealed it to be a new polerovirus, distantly related to other BYD-associated viruses, and with the suggested designation as *Maize yellow dwarf virus-RMV* (MYDV-RMV) (Kruger et al., 2013). In recent literature, an alternative naming of PAV-related isolates has also been proposed with the species PAV-I, PAV-II and PAV-III, where PAV-I corresponds to PAV, PAV-II to PAS and PAV-III to PAV-CN (Liu et al., 2007; Wu et al., 2011).

### **1.3. A new tentative species of BYDV-PAV**

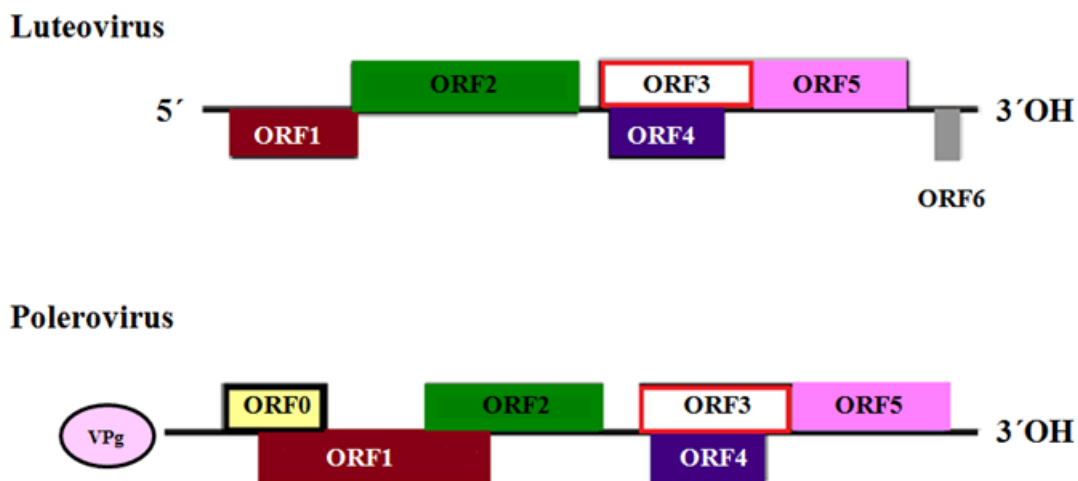
A new, distinct genotype of BYDV-PAV (PAV-Sal1) (GenBank accession number: AJ563410) was discovered in Latvia in a study on molecular diversity of the CP-encoding region of BYDV-PAV and BYDV-MAV isolates from Latvia and Sweden (Bisnieks et al., 2004). Initially, the isolate was classified as belonging to the species BYDV-PAV, since the isolate showed a positive reaction for BYDV-PAV based on detection by a PAV-specific antibody. However, it was revealed that PAV-Sal1 could represent a new species after CP gene sequence comparisons with other isolates of BYDV-PAV and BYDV-MAV, where the amino acid (aa) sequence of PAV-Sal1 differed by more than 10 % to any other isolate. Then, BYDV-OYV (OYV for oat yellowing virus) was proposed as a name for this new tentative virus. BYDV-OYV showed closest relationship with BYDV-PAV-CN, which represents a third distinct species within the PAV serotype (GenBank: AY855920). Since then, BYDV-OYV has also been identified in different parts of Sweden (Eriksson et al., unpublished). In a previous study (Sathees, 2015), a partial genome sequence of BYDV-OYV from Sweden was determined to investigate the relationship to other isolates of BYDV. The partial genome sequence of BYDV-OYV comprised 2,792 nucleotides (nt) of the 3' end. Organization of open reading frames (ORF) and un-translated regions (UTR) of the genome sequence was similar to those of viruses in the genus *Luteovirus*, but the nt identity of the partial genome sequence was less than 90 %. This study also revealed that BYDV-OYV isolates from Sweden and Latvia belong to the same species, since the CP gene of these isolates shared 95 % nt and 93 % aa identities.

### **1.4. Importance of a complete genome sequence**

Although the experiment revealed that BYDV-OYV is a new tentative species within the family *Luteoviridae*, a complete genome sequence of BYDV-OYV is necessary for a taxonomic classification and understanding of its evolutionary history. Some virus isolates have been reclassified according to the guidelines of the International Committee on Taxonomy of Viruses (ICTV) after determining their complete genome sequences. Complete genome sequences help to discover genetic recombination, which is the exchange of nucleotide segments among molecules. Recombination may play an important role in the emergence of new viral pathogens (Knierima et al., 2010; Sztuba-Solinska et al., 2011;

Bujarski, 2013). For instance, sequence analysis of individual ORFs and UTRs of PAV-CN and other PAV isolates showed that PAV-CN may have been formed through recombination events. The 5' UTR and ORF 1 of PAV-CN are most closely related to GAV, which is another Chinese species of BYDV, while PAV-CN is most closely related to BYDV-PAS for all other regions (Liu et al., 2007). Recombination may also lead to changes in the biological properties of a virus affecting host resistance and specificity (Sztuba-Solinska et al., 2011; Bujarski, 2013).

### 1.5. Genome organization and gene functions of family *Luteoviridae*



**Fig. 1.** Genome organization of viruses belonging to the genera *Luteovirus* and *Polerovirus*

The determination of complete genome sequences of virus isolates in family *Luteoviridae* has greatly enhanced our understanding of their genome organization and gene-expression strategies. The members of the genera *Luteovirus* and *Polerovirus* have six ORFs in their genome (Domier, 2012). The 5'-proximal ORFs are necessary for RNA replication. The 5' part of the genome of viruses in the genus *Polerovirus* contains ORF0, which encodes a suppressor of RNA silencing (VSR) (Pfeffer et al., 2002; Mangwende et al., 2009; Kozłowska-Makulska et al., 2010), but that is absent in BYDV-PAV and other viruses of the genus *Luteovirus*. ORF0 is also responsible for symptom induction and determination of viral host range, and it overlaps with ORF1. ORF1 and ORF2 overlap for both genera, but the length of nucleotide overlap varies. ORF2 encodes the RNA-dependent RNA polymerase (RdRp) but that must be fused to ORF1 to encode the RdRp. ORF1 and ORF2 are essential

for RNA replication in plant cells (Koev et al., 2002). ORF3 encodes the major CP. ORF5 is translated in fusion with ORF3 as a result of in-frame translational read-through of the ORF3 stop codon. ORF5 encodes a read-through domain (RTD) necessary for aphid transmission (Chay et al., 1996; Gray & Gildow, 2003) and it may also be involved in systemic virus movement within plants (Ziegler-Graff et al., 1996; Peter et al., 2009) and phloem limitation (Peter et al., 2009). ORF4 is entirely located within ORF3 and encodes a movement protein (MP) and permits infection of the phloem tissue of the host plant (Chay et al., 1996). BYD-associated viruses are phloem-limited (Peter et al., 2009) and interfere with translocation of photosynthesis products from cells in leaves to sieve tube elements through plasmodesmata (Rochow & Duffus, 1981; Gray, 1996). The genome of viruses in the genus *Luteovirus* contains ORF6, which encodes a silencing suppressor (Liu et al., 2012).

## **1.6. Relationship between family *Luteoviridae* genomes**

Family *Luteoviridae* falls into three genera: *Luteovirus*, *Polerovirus* and *Enamovirus* (Domier, 2012), differing in genome organization, replication strategies and expression mechanisms (Maia et al., 2000). Viruses of the genera *Luteovirus* and *Polerovirus* differ at the 5' end of their genomes (Domier, 2012) with viruses in the genus *Luteovirus* being most closely related to those of genus *Dianthovirus* in the family *Tombusviridae* (Miller et al., 2002). Thus, the replication proteins of genus *Luteovirus* most closely resemble those of viruses in the family *Tombusviridae*. ORF1 of poleroviruses is longer than ORF1 of luteoviruses. ORF1 of genus *Polerovirus* encodes different functions such as a proteinase motif and viral genome-linked protein (VPg) (van der Wilk et al., 1997). The VPg is absent in genus *Luteovirus*. In contrast, an important sequence similarity is detected between ORFs at the genome 3' end of genera *Luteovirus* and *Polerovirus*. This sequence similarity is detected between CP (ORF3), MP (ORF4) and RTD (ORF5) (Miller et al., 2002). A number of recombination events have been detected between luteoviruses and poleroviruses (Domier et al., 2002). These two genera diverged from a common ancestor through recombination (Miller et al., 2002; Pagan & Holmes, 2010). Since then, a number of recombination events have occurred between and among luteoviruses and poleroviruses (Domier et al., 2002). The RdRp genes of genera *Polerovirus* and *Enamovirus* show closest relationship to each other but they have a different origin to those of *Luteovirus* (Miller et al., 2002; Domier, 2012). In addition,

the genome of viruses in the genus *Enamovirus* has a genome-linked viral protein (VPg) attached to the 5' end of the genome like viruses of the genus *Polerovirus* (Pfeffer et al., 2002; Mangwende et al., 2009).

## **2. Aims of the study**

BYDV is a group of plant viruses of high economical relevance in cereal cultivation and that are transmitted by aphids. However, knowledge about the taxonomic status, evolutionary dynamics and genetic diversity of BYD-associated viruses in northern Europe is limited. Therefore, the aim of the study was to determine the complete genome sequence of a Swedish isolate of BYDV-OYV, which is a new tentative species within the family *Luteoviridae*, and to study its origin and relationship with other isolates within the family *Luteoviridae*.

### **Hypothesis**

BYDV-OYV is a new species within the family *Luteoviridae* and has evolved through a recombination event.



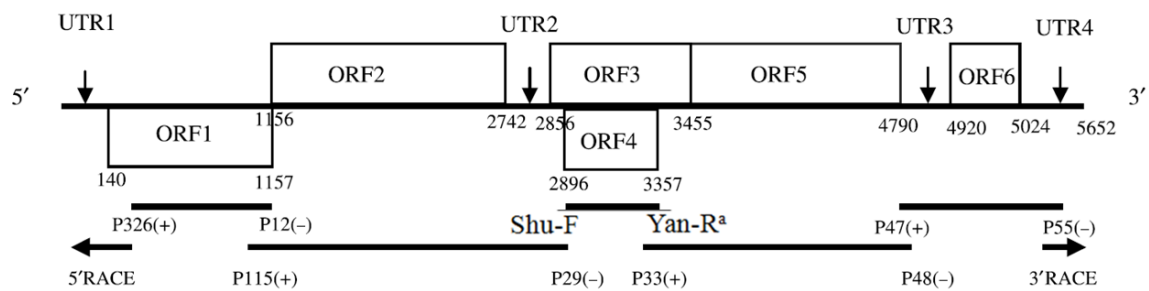
### 3. Materials and Methods

#### 3.1. Plant material

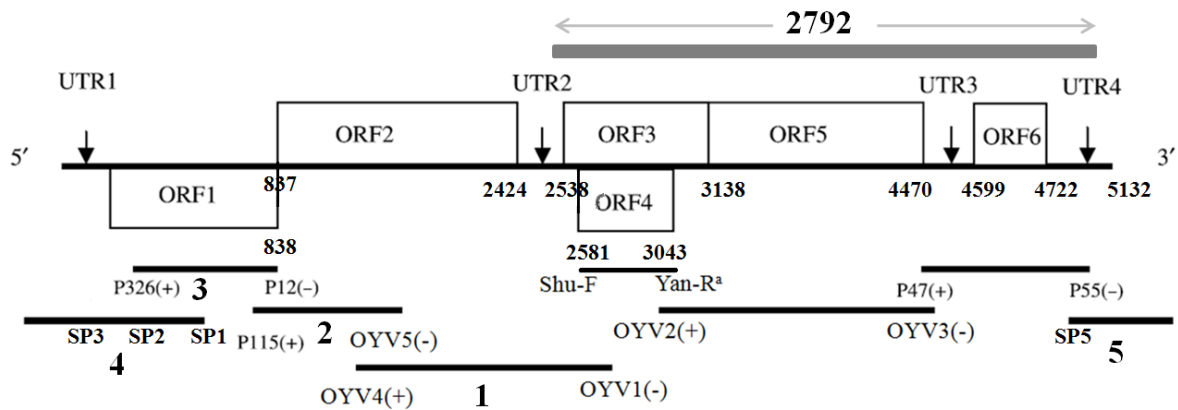
Plant material previously shown to be infected with BYDV-OYV (Sathees, 2015; Eriksson et al., unpublished) was used to obtain the complete viral genome sequence. Virus-infected plant extracts were obtained from two different grasses from the county of Jämtland, Sweden, collected in 2010: couch grass (*Elytrigia repens*; Kvickrot) and *Festuca pratensis* (Ängssvingel). In my previous study, two extracts (A2 and A30) of *F. pratensis* and two extracts (K15 and K19) of *E. repens*, which had previously been positive in ELISA tests for BYDV-PAV and -MAV, were tested by immunocapture reverse transcription polymerase chain reaction (IC-RT-PCR) for the CP gene (Bisnieks et al., 2004). Two extracts (A2 and A30) were then used to determine the partial genome sequence of BYDV-OYV. In the current study, A30 was used to obtain the complete viral genome sequence.

#### 3.2. Primer design

A.



B.



**Fig. 2. A.** Genomic map of members of BYDV-PAV-CN with localization of primers for amplification of different segments of the BYDV-OYV genome (adapted from Liu et al., 2007). **B.** Genomic map for the new tentative species BYDV-OYV with positions for newly designed primers. The highlighted part of the genome (2,792 nt) was already sequenced in a previous study (Sathees, 2015). For amplification of genome fragments 1, 2, 3, 4 and 5 by immunocapture RT-PCR, primers OYV1 (-), OYV4 (+), OYV5 (-), P115 (+), P12 (-), P326 (+), SP1, SP2, SP3 and SP5 were used. The boxes indicate ORFs and the block arrows indicate UTRs. The numbers indicate nucleotide positions. The bars under the genomic map indicate the overlapping products of RT-PCR.

Primers used in RT-PCR to amplify the 5' part of the BYDV-OYV genome were obtained from the published sequence of BYDV-PAV-CN (Liu et al., 2007) (Fig. 2). The 5' part of the genome, which includes part of ORF1, ORF3 and ORF4 as well as the complete ORF2 and UTR2, was attempted to be amplified as two overlapping fragments using two primer pairs: P29 (-) / P115 (+) and P12 (-) / P326 (+) (Fig. 2A). Although the fragment, which covered most of ORF1 could be amplified by RT-PCR using primers P12 (-) and P326 (+), primers P29 (-) and P115 (+) failed to amplify the other fragment. Thus, primer OYV1 (-) was designed instead of primer P29 (-) based on the determined nucleotide sequence of the BYDV-OYV CP gene (Sathees, 2015). After failing to clone the fragment, which was amplified using the primers OYV1 (-) and P115 (+), this region was divided into two overlapping fragments (1 and 2) (Fig. 2B). Two primers, OYV4 (+) and OYV5 (-), were

designed for highly conserved regions of other BYD-associated viruses (Table 1) to amplify the overlapping fragments 1 and 2.

**Table 1. Primers used for amplification of 5' part and 3' end of the BYDV-OYV genome**

<b>Primers</b>	<b>Nucleotide sequences (5' to 3')</b>	<b>Position</b>	<b>Reference</b>
OYV1 (-) <sup>a</sup>	AGTACGTGAGAGCTAATGTAC	2491-2471 <sup>b</sup>	This study
OYV4 (+)	ATGTTTCGTTGAGGATAAGATGC	1643-1665	This study
OYV5 (-)	TCACCATGTTGAAGCCGTATT	1890-1869	This study
SP1 (-)	CCTGCACCATTTTCGTGACC	490-471	This study
SP2 (-)	TCGTGACCTCCTTCATTAGACCA	478-456	This study
SP3 (-)	TGGATGTTGGTGGGTGCAGAGA	416-395	This study
SP5 (+)	ACCCACCACCGATCAAATAAGCT	4808-4831	This study
P12 (-)	GCTCCGTCTGTGACCGCAAT	895-875	Liu et al., 2007
P115 (+)	GGGTTTTTAGAGGGGCTCTGT	828-848	Liu et al., 2007
P326 (+)	GACTTCGAGGCNGANCTCGCT	1-21	Liu et al., 2007
P29 (-)	TCGATTGGGTTGGACCACAACC	2958-2937	Liu et al., 2007

<sup>a</sup>“+” indicates positive-sense sequence; “-” indicates complementary sequence.

<sup>b</sup> Position of the primers in the determined sequence of the BYDV-OYV genome.

### **3.3. Immunocapture-reverse transcription-polymerase chain reaction (IC-RT-PCR), gel extraction and purification**

The virus particles were captured using immunocapture techniques to amplify fragments 1, 2 and 3 (Fig. 2B). Then, cDNA for amplification of fragments 1- 3 was synthesized by RT using appropriate reverse primers (Fig. 2A and 2B; Table 1) using a cDNA synthesis kit (Superscript III Reverse Transcriptase, Invitrogen by Life Technologies) according to manufacturer’s instructions (Bisnieks et al., 2004; Sathees, 2015).

PCR amplification of fragments 1-3 was carried out using appropriate forward and reverse primers with Dream Taq DNA Polymerase (Thermo Scientific) (Malmstrom and Shu, 2004; Bisnieks et al., 2004; Liu et al., 2007; Sathees, 2015). As a negative control, PCR was run with water instead of cDNA to monitor any contamination. The fragments were amplified using a T100™ Thermal Cycler (Bio-Rad). PCR amplification of fragment 1 consisted of an initial denaturation step of 2 minutes and 30 seconds at 95°C followed by 35 cycles of

denaturation at 95°C for 30 seconds, annealing at 51°C for 2 minutes and primer extension at 72°C for 2 minutes. The final PCR cycle was followed by a final extension at 72°C for 10 minutes and indefinite hold at 12°C. In PCR amplification of fragments 2 and 3, the annealing temperatures were altered according to the melting temperatures of the primer pairs. The annealing temperature for amplification of fragment 2 was 63°C, while it was 55°C for fragment 3. However, the annealing temperatures were changed according to the results of PCR amplifications. For example, initially, the annealing temperature for amplification of fragment 1 was set to 56°C but the amplification of fragment 1 failed to give the expected result. Then the annealing temperature was changed and set to 51°C. The PCR products were analyzed in a 1 % agarose gel with 0.5x TBE and stained with 1 µg/ml Gel Red. Fragments 1 and 3 were purified from gels using a gel purification kit (Gene JET™ Gel Extraction Kit, Thermo Scientific) according to the manufacturer's instructions, since the PCR amplification resulted in more than one band. For confirmation, purified PCR fragments were analyzed by gel electrophoresis.

#### **3.4. Cloning of purified DNA fragments 1-3**

The purified DNA was cloned using CloneJET PCR Cloning Kit (Thermo Scientific) and Subcloning Efficiency *Escherichia coli* DH5α Competent Cells (Invitrogen) according to the manufacturers' instructions.

#### **3.5. Overnight bacterial culture**

The presence of clone inserts in the selected colonies was analyzed by gel electrophoresis after PCR amplification using 2 µl 10X Dream Taq buffer (Fermentas), 0.2 µl from 10 µM pJET Forward sequencing primer, 0.2 µl from 10 µM pJET Reverse sequencing primer, 0.4 µl from 2 mM dNTP, 0.1 µl Dream Taq (5 U/µl), 16 µl water and 1 µl bacterial suspension. Clones containing inserts of the expected size were inoculated into 4 ml LB containing ampicillin (4 µg/ml) and incubated overnight at 37°C with shaking. Plasmid DNA was isolated from the overnight cultures using the Gene JET Plasmid Miniprep Kit (Thermo

Scientific) according to the manufacturer's instructions. After purification, the yield and purity of plasmid DNA was measured using a nanodrop instrument (Thermo Scientific).

### **3.6. Restriction analysis of purified plasmid DNA and sequencing**

The purified plasmids were cut by the two restriction enzymes *XhoI* and *XbaI* (Sathees, 2015). The presence of inserts was analyzed and confirmed by agarose gel electrophoresis. Twenty  $\mu\text{l}$  of purified plasmid DNA (100 ng/ $\mu\text{l}$ ) with insert of the expected size was sent to Macrogen (Amsterdam, the Netherlands) for sequencing. The purified plasmid DNA was sequenced using primers pJET forward and pJET reverse (Thermo Scientific).

### **3.7. 5'/3' rapid amplification of cDNA end (5'/3' RACE)**

#### **3.7.1. Primer design for RACE**

Four primers were designed to amplify the 5' and 3' ends of the BYDV-OYV genome. For amplification of the 5' end of the genome, primers SP1, SP2 and SP3 were designed based on the sequence of fragment 3, which includes ORF1 and ORF2. For amplification of the 3' end of the genome, primer SP5 was designed from the previously determined sequence (Sathees, 2015) (Fig. 2B; Table 1).

#### **3.7.2. 5' RACE**

##### **3.7.2.1. First-strand cDNA synthesis**

To amplify the 5' end, virus particles were captured from plant extract using the immunocapture process as in step 3.3 (Sathees, 2015). cDNA synthesis was carried out using primer SP1 and cDNA synthesis kit (5'/3' RACE Kit, 2<sup>nd</sup> Generation, Roche) according to manufacturer's instructions. A control reaction was carried out using Control Primer neo1/rev and Control neo-RNA.

### 3.7.2.2. Purification of cDNA

The cDNA was purified using a PCR purification kit (GeneJET PCR purification kit, Thermo Scientific) according to the manufacturer's instructions. Concentration and purity of the purified cDNA was determined using a nanodrop instrument. One  $\mu$ l of the purified cDNA was taken for a later control by PCR. The purified cDNA was directly used for poly(A) tailing by recombinant Terminal Transferase and dATP (5'/3' RACE Kit, 2<sup>nd</sup> Generation, Roche) according to manufacturer's instructions.

### 3.7.2.3. PCR amplification of dA-tailed cDNA and control reaction

PCR amplification of dA-tailed cDNA was carried out using the Expand High Fidelity PCR System (Roche) according to manufacturer's instructions. Two PCR rounds were carried out in the amplification process. However, the second round PCR is optional. If the first round PCR results in sufficient amount of PCR product, the second round PCR is not necessary. Otherwise, the second round PCR is carried out. The first round of PCR was run using a T100<sup>TM</sup> Thermal Cycler with the conditions according to manufacturer's instructions (Roche). Control reaction was carried out using Oligo dT Anchor Primer and Control Primer neo2/rev to confirm that the dA-tailing was successful. A band of around 300 bp should be obtained if the dT-tailing reaction has been successful, according to the manufacturer's instructions for the 5'/3' RACE Kit (Roche). As a negative control, PCR was run without template to monitor for contamination.

The second PCR round was performed with diluted (product of the first round diluted 1:20 in double-distilled water) or undiluted product from the first PCR round as template using the Expand High Fidelity PCR System (Roche) and a T100<sup>TM</sup> Thermal Cycler according to manufacturer's instructions.

#### 3.7.2.4. Agarose gel electrophoresis

The products from the first and second rounds of PCR were analysed as described in section 3.3. DNA fragments of the first PCR round were purified from the gel (see section 3.3). For confirmation, purified PCR fragments were analyzed by gel electrophoresis.

The PCR fragments were cloned and analysed as described in sections 3.4 – 3.6, but using *E. coli* Library Efficiency DH5 $\alpha$  Competent Cells (Invitrogen).

### 3.7.3. 3'RACE

#### 3.7.3.1. Immunocapture for 3' end and poly(A) tailing

Virus particles were captured by immunocapture (section 3.3) to determine the sequence of the 3' end of the BYDV-OYV genome. A poly(A) tail was added to the captured viral RNA using a Poly(A) Tailing Kit (Invitrogen) according to manufacturer's instruction.

#### 3.7.3.2. First-strand cDNA synthesis and PCR amplification of cDNA

After poly(A) tailing, cDNA synthesis was carried out (5'/3' RACE Kit, 2<sup>nd</sup> Generation, Roche) according to manufacturer's instructions. The cDNA was used as template for PCR without prior purification. Then, PCR amplification using poly(A) tailed cDNA as template was carried out using the Expand High Fidelity PCR System (Roche) and a T100<sup>TM</sup> Thermal Cycler according to manufacturer's instruction (Roche). As a negative control, PCR was run without template to monitor for contamination.

#### 3.7.3.3. Agarose gel electrophoresis

The PCR product for the 3' end of the genome was analysed using gel electrophoresis and the DNA fragment was purified from the gel as described in section 3.3. The purified PCR

fragment was analyzed by gel electrophoresis and then cloned as described in sections 3.4 to 3.6.

### 3.8. Sequence analyses and phylogenetic analyses

At least two clones were sequenced for each fragment. Sequences of the five genome fragments (1, 2, 3, 4 and 5) of BYDV-OYV were analysed by nucleotide identity searches in the GenBank database with Blastn (<http://www.ncbi.nlm.nih.gov/BLAST/>). An almost complete genomic sequence of BYDV-OYV was assembled using MegAlign (DNASTAR Lasergene 11) from the four sequenced PCR fragments (1, 2, 3 and 4) and the 3' part of the BYDV-OYV genome (2,792 nt), which was determined in a previous study (Sathees, 2015).

The assembled sequence was analysed using ORF finder in NCBI (<http://www.ncbi.nlm.nih.gov/projects/gorf/>) to locate open reading frames (ORFs) and untranslated intergenic regions (UTRs). Nucleotide identities of BYDV-OYV ORFs and UTRs and deduced amino acid sequence identities of BYDV-OYV ORFs to other members of family *Luteoviridae* were calculated in ClustalW (<http://www.genome.jp/tools/clustalw/>). For phylogenetic analyses, a multiple alignment of the assembled genome sequence of BYDV-OYV and those of different BYD-associated viruses, which were obtained from GenBank (Table 2), was performed using ClustalW in MEGA 5.2.1 (Tamura et al., 2011) before phylogenetic analyses. The phylogenetic tree was constructed using the neighbour-joining algorithm in MEGA 5.2.1 with distances calculated using Tamura 3-parameter model. The robustness of the internal branches of the tree was estimated by bootstrap analysis using 500 replicates.

**Table 2. The virus sequences used for sequence comparisons and for phylogenetic analyses**

Name of virus	Accession number
PAV-Sal1	AJ563410
BYDV-PAS-129	AF218798
BYDV-PAV- III	AF235167



BYDV-GAV	AY220739
BYDV-PAV-CN	AY855920
BYDV-OYV	-
BYDV-MAV-PS1	D11028
BYDV-PAV isolate 016	EF521835
BYDV-PAV isolate 052	EF521841
BYDV-PAV isolate 068	EF521846
BYDV-PAV isolate 05GG2	EU332309
BYDV-PAV isolate 05GG5	EU332310
BYDV-PAV isolate 05ZZ4	EU332321
BYDV-PAV isolate 05ZZ1	EU332320
BYDV-PAV isolate 06GY5	EU332329
BYDV-PAV isolate 06KM14	EU332332
CYDV-RPV	L25299

---

### 3.9. Recombination detection

The sequences of 13 isolates of the genus *Lutovirus* (Table 2) were aligned using ClustalW in MEGA 5.2.1 (Tamura et al., 2011) and the output of the alignment was saved as a FASTA file for opening in the recombination detection program package (RDP v.4.46) (<http://en.bio-soft.net/tree/RDP.html>) (Martin et al., 2015) for detection of putative recombination events in BYDV-OYV. The advantage of the program is that the program simultaneously allows different recombination detection algorithms, which sequentially test every combination of three sequences of input data to identify recombination events based on the assumption of one recombinant and two non-recombinant sequences. RDP4 is also able to detect recombination

events without predefined non-recombinant reference sequences (Martin et al., 2011). The package contained a series of recombination detection algorithms, including GENECONV (Padidam et al., 1999), Bootscan (Martin et al., 2005), Chimaera (Posada & Crandall, 2001), SiScan (Gibbs et al., 2000), MaxChi (Maynard Smith, 1992), 3Seq (Boni et al., 2007) and RDP (Martin & Rybicki, 2010). All these methods were utilized and compared to identify and characterise individual recombination events, positions of beginning and ending breakpoints and sequences most closely related to the parental sequences (major and minor parents). The minor parent is the one contributing the smaller fraction of the recombinant, while the major parent is the one contributing the larger fraction of the recombinant (Martin et al., 2010). A putative recombination event that was identified by at least three of the seven recombination detection algorithms (Liu et al., 2010) was retained. Phylogenetic analysis with recombinant regions was performed to confirm results of recombination analysis. The trees were constructed using the neighbour-joining algorithm to show topological shift of specific sequences of minor and major parents in the recombinant (Liu et al., 2010).

## 4. Results

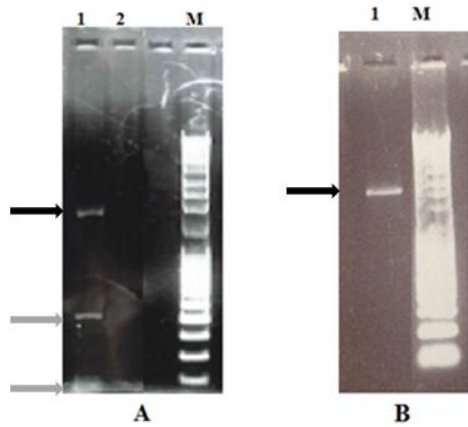
### 4.1. PCR of fragments 1-3

PCR primers were designed to amplify the 5' end of the BYDV-OYV genome as partially overlapping fragments (Fig. 2A; Table 1). PCR using primers P29 (-) and P115 (+) to amplify a fragment corresponding to ORF2 and UTR2 (Fig. 2) did not give any visible product (Fig. 3).



**Fig. 3.** Gel electrophoresis for analysis of attempt to amplify part of BYDV-OYV genome (ORF2 and UTR2) using the primers P29 (-) and P115 (+). Lane 1 was loaded with amplified PCR product (expected size ~1,765 bp). Lane 2 represents a negative control and lane M contains 80-10,000 bp Mass Ruler DNA.

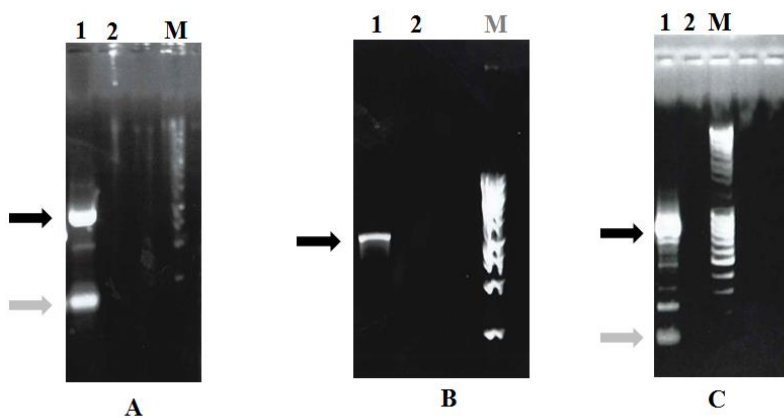
Instead, this genome segment was amplified using primer P115 (+) and the newly designed primer OYV1 (-) yielding an amplification product with the expected size of ~1,765 bp, (Fig. 4A). The gel electrophoresis also showed that this PCR amplification resulted in primer-dimer and an unexpected PCR product. However, purification of the expected band using a gel extraction kit was successful and the strength of the purified band in gel electrophoresis showed that the DNA amount was sufficient for cloning (Fig. 4B).



**Fig. 4. A.** Gel electrophoresis of PCR products, which were amplified using the primers OYV1 (-) and P115 (+). **B.** Gel electrophoresis of PCR product after gel purification. Black arrows in A and B indicate the expected PCR product (~1,765 bp) and gray arrows indicate primer-dimer and unwanted PCR product. Lane 2 represents a negative control and lane M contains 80-10,000 bp Mass Ruler DNA.

However, no colonies were obtained when attempting to clone the fragment. Hence, this genome region was divided into two overlapping fragments (1 and 2) and amplified. Fragment 1 mainly contained ORF2 and UTR2, while fragment 2 mainly contained the 5' end of ORF2. In addition, primers were designed for amplification of fragment 3 covering most of ORF1 (Fig. 2B).

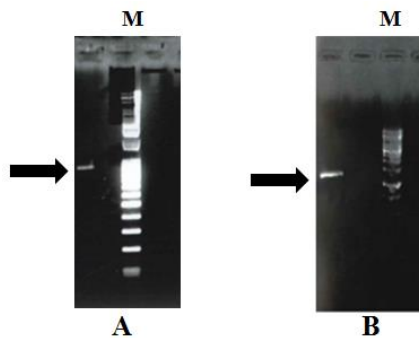
The obtained PCR products were of the expected sizes for the targeted genome regions of BYDV-OYV: ~805 bp (fragment 1), ~1,021 bp (fragment 2) and ~875 bp (fragment 3), respectively (Fig. 5).



**Fig. 5.** Gel electrophoresis of PCR products for the 5' end of the BYDV-OYV genome. Black arrows indicate the expected PCR product and gray arrows indicate primer-dimer. **A.** PCR product of fragment 1, which was amplified using the primers OYV1 and OYV4 (~805 bp). **B.** PCR product of fragment 2, which was amplified using primers OYV5 and P115 (~1,021 bp). **C.** PCR product of fragment 3, which was amplified using primers P12 and P326 (~875 bp). Lane 2 contains negative control. Lane M in A and C contain 80-10,000 bp Mass Ruler DNA Ladder, while lane M in B contains GeneRuler™ 1 kb DNA Ladder (250 to 10,000 bp).

#### 4.2. Gel extraction and purification

Gel extraction of fragments 1 and 3 resulted in pure fragments of the expected sizes (~805 bp and ~875 bp) and at a high concentration suitable for cloning (Fig. 6).

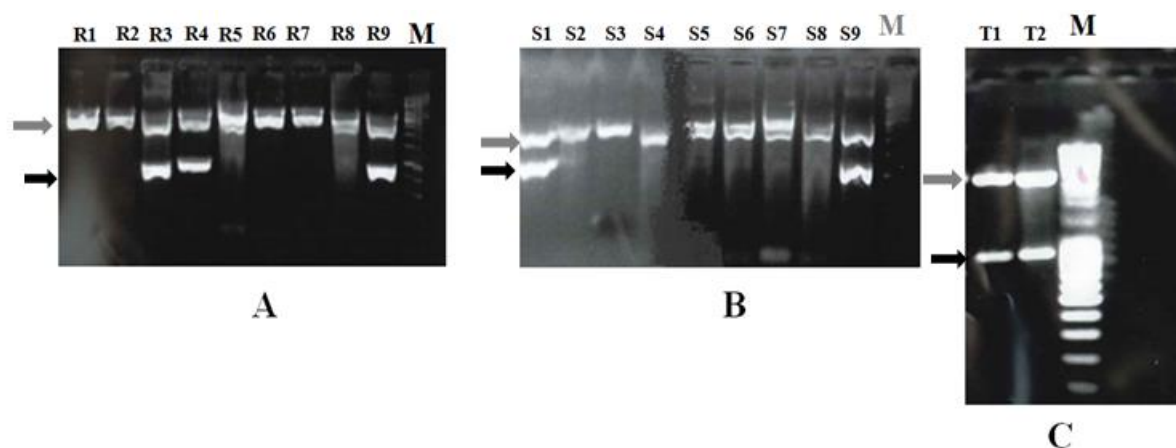


**Fig. 6.** Gel electrophoresis of PCR products for BYDV-OYV fragments 1 and 3 after gel purifications. **A.** Purified PCR product of fragment 1 (~805 bp). **B.** Purified PCR product of fragment 3 (~875 bp). Lane M in A is 80-10,000 bp Mass Ruler DNA Ladder and lane M in B contains GeneRuler™ 1 kb DNA Ladder (250 to 10,000 bp).

#### 4.3. Restriction analysis of purified recombinant plasmids for fragments 1-3

Restriction analysis of recombinant plasmids for the three fragments of the BYDV-OYV genome was performed to ensure that all the clones contained the expected DNA inserts. According to the gel electrophoresis of restriction products, out of nine clones, R3, R4 and R9

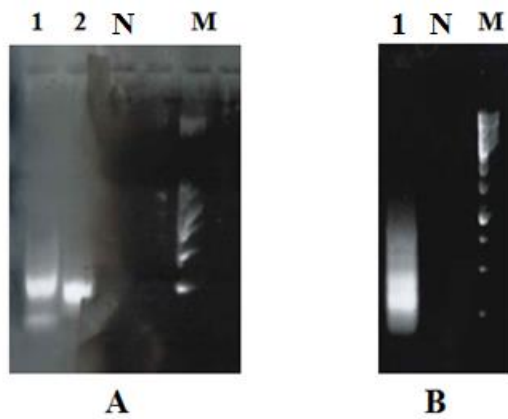
contained an insert of the expected size for fragment 1 (Fig. 7A). For fragment 2, two clones out of nine (S1 and S9) contained inserts of the expected size (Fig. 7B). For fragment 3, two clones (T1 and T2) contained inserts of the expected size (Fig. 7C). The insert sequences of these positive clones (R3, R4, R9, S1, S9, T1 and T2) showed high identity to other BYD-associated viruses in the GenBank database.



**Fig. 7.** Plasmid restriction analyses with *Xba*I and *Xho*I. **A.** Cloning of PCR products of fragment 1 (insert size ~805 bp). **B.** Cloning of PCR products of fragment 2 (insert size ~1021 bp). **C.** Cloning of PCR products of fragment 3 (~875 bp). Black arrows indicate insert and gray arrows indicate vector DNA. Lane M in A and C was loaded with 80-10,000 bp Mass Ruler DNA Ladder, while lane M in B contains GeneRuler™ 1 kb DNA Ladder (250 to 10,000 bp).

#### 4.4. 5' RACE

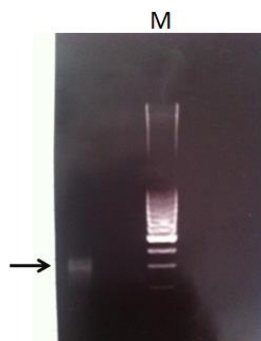
Two subsequent PCR rounds were performed to amplify the 5' end of the BYDV-OYV genome after poly(A)-tailing of first-strand cDNA (Fragment 4). In this experiment, two PCR rounds were carried out, even if the first PCR round was successful (Fig. 8A and B). The resulting band of around 500 bp in the control reaction confirmed the successful dA-tailing (Fig. 8A). No amplification was obtained in the negative control experiment indicating the absence of any contamination.



**Fig. 8.** Gel electrophoresis of PCR products from nested amplification to obtain the 5' end of the BYDV-OYV genome. **A.** Lane 1: First round PCR product amplified using specific primer SP2 and Oligo dT-Anchor Primer (~500 bp); Lane 2: positive control reaction. **B.** Lane 1: Second round PCR product amplified using specific primer SP3 and PCR Anchor Primer. M indicates GeneRuler™ 1 kb DNA Ladder (250 to 10,000 bp). N indicates negative control.

#### 4.5. Gel extraction and purification

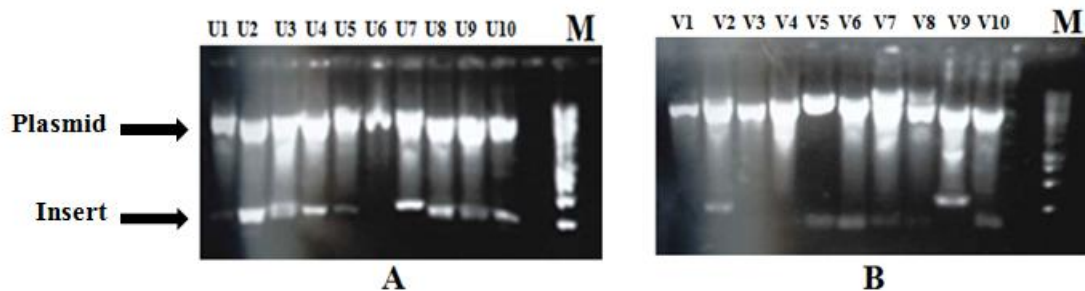
Gel extraction of first round PCR resulted in a pure fragment of the expected size (~500 bp) and at a high concentration suitable for cloning (Fig. 9).



**Fig. 9.** Gel electrophoresis of first round PCR product, which was obtained through amplification of the 5' end of the BYDV-OYV genome, after gel purification. Arrow indicates the PCR product. M indicates GeneRuler™ 1 kb DNA Ladder (250 to 10,000 bp).

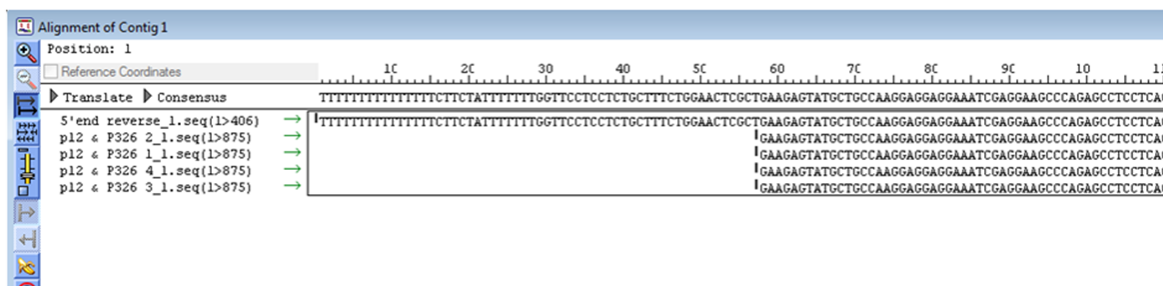
#### 4.6. Restriction analysis of recombinant plasmid for fragment 4

After cloning, restriction analysis was performed to confirm that clones contained inserts of the expected size (Fig. 10) of PCR product (~500 bp) from first or second round PCR. The gel showed that all of the ten clones for the product of the first PCR round had the expected size of DNA inserts (Fig. 10A), except U1 and U6 (Fig. 8A). For the second PCR round only two (V2 and V9) of ten clones showed the expected band size (Fig. 10B) from the second PCR round.



**Fig. 10.** Gel electrophoresis of digested plasmids with putative inserts for the 5' end of the BYDV-OYV genome. **A.** Cloning of first round PCR product. **B.** Cloning of second round PCR product. Lane M contains GeneRuler™ 1 kb DNA Ladder (250 to 10,000 bp).

Only one clone (U4) showed significant similarity to BYD-associated viruses in GenBank database searches with Blastn. The insert of this clone was 438 bp excluding primers and around 349 bp overlapped with fragment 3. However, the insert sequence of clone U4 had several T residues in the 5' end (Fig. 11) which did not correspond to the added dT-tail and the PCR Anchor primer.

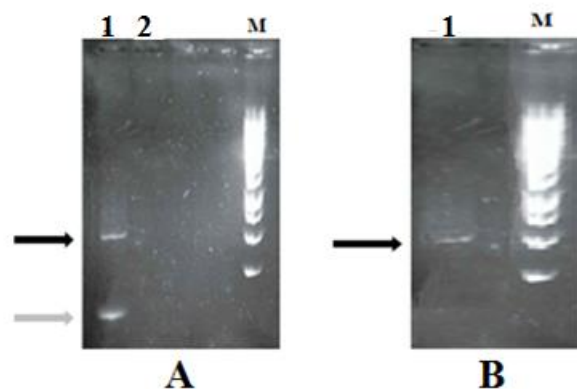


**Fig. 11.** Sequence alignment of fragments 3 and 4 of BYDV-OYV using MegAlign in DNASTAR.



#### 4.7. 3' RACE

Gel electrophoresis of the PCR product for the 3' end of the BYDV-OYV genome showed the expected DNA band of around 750 bp and a primer-dimer (Fig. 12A). Gel electrophoresis of extracted and purified PCR product showed that the amount of product was sufficient for cloning (Fig. 12B).

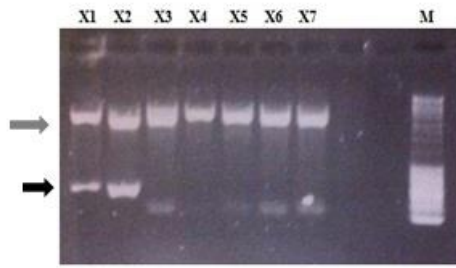


**Fig. 12.** **A.** PCR amplification of the 3' end of the BYDV-OYV genome. A black arrow marks the expected PCR product (~750 bp) and the gray arrow indicates primer-dimer in lane 1. Lane 2 indicates negative control. **B.** Gel electrophoresis of gel-purified PCR product of 3' end of BYDV-OYV genome. The black arrow marks the expected purified PCR product (~750 bp). M refers to the GeneRuler™ 1 kb DNA Ladder (250 to 10,000 bp).

#### 4.8. Restriction analyses of recombinant plasmid of 3' end of BYDV-OYV genome

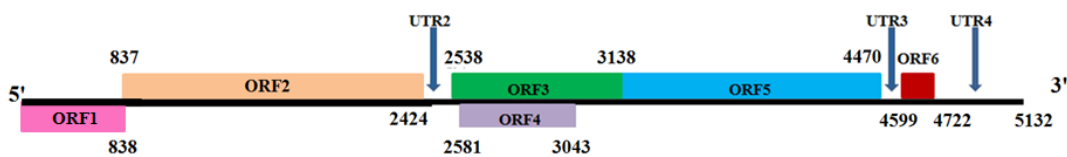
After cloning, restriction analysis was carried out to confirm that clones had the expected size of inserts corresponding to the PCR product of the 3' end of the BYDV-OYV genome. Clones X1 and X2 had the expected inserts of around 750 bp (Fig. 13).

Even though insert sequences of clones X1 and X2 contained forward (specific primer SP5) and reverse (PCR Anchor) primers in their sequences, the sequences between the primers did not show any significant identity to those of other members of family *Luteoviridae* in the GenBank database.



**Fig. 13.** Gel electrophoresis of digested plasmids with putative inserts for the 3' end of the BYDV-OYV genome. The black arrow indicates inserts and the gray arrow indicates vector. Lane M contains GeneRuler™ 1 kb DNA Ladder (250 to 10,000 bp).

#### 4.9. Genome organization of BYDV-OYV



**Fig. 14.** Genomic map of BYDV-OYV. The boxes indicate ORFs and the arrows indicate UTRs. The numbers indicate nucleotide positions.

The assembled BYDV-OYV genome sequence, with 5' and 3' ends missing, was 5,132 nt in length (Fig. 14). The genome sequence was assembled from sequences of overlapping fragments 1, 2 and 3 of the 5' end obtained in the current study (Fig. 3) and 3' end sequences from a previous study (Sathees, 2015). The assembled genome contained six ORFs and three UTRs, which were organized similarly as for other members of the genus *Luteovirus* (Liu et al., 2007; Domier, 2012). The partial sequence of ORF1 is 837 nt with the termination codon (TAG) at nt 838. ORF2 (nt 837-2424) partially overlapped with ORF1 and encoded a putative protein of 58.8 kDa. Similar to other luteoviruses, ORF2 is likely to be expressed by a frameshift regulated by the sequence GGGTTTT, which is immediately upstream of the

ORF1 stop codon (Mayo & Miller, 1999) (Table 4). UTR2 between ORF2 and ORF3 was 114 nt in length. ORF3, ORF4 and ORF5 showed similar structure to those of viruses in the genera *Luteovirus* and *Polerovirus* (Fig. 1). ORF3 (nt 2538-3138) encoded a coat protein of 22 kDa. ORF4 (462 nt) was contained completely within ORF3, but in a different reading frame, and encoded a movement protein of 17 kDa. ORF5 was immediately adjacent to the ORF3 stop codon and encoded a read-through protein of 49.4 kDa. UTR3 was 129 nt in length and located between ORF5 and ORF6. ORF6 (123 nt) was similar in length to ORF6 of BYDV-PAV isolate 05ZZ4, but longer than for BYDV-PAV-CN (Table 4).

**Table 4. Length in nucleotides of ORFs and UTRs of BYDY-OYV and other members of the genus *Luteovirus***

Region	BYDV- OYV	BYDV-PAV- CN	BYDV- PAS-129	BYDV-PAV isolate 05ZZ4	BYDV- MAV-PS1	BYDV- PAV isolate 016	BYDV- PAV isolate 06KM14
<b>UTR1</b>	-	141	138	144	105	141	138
<b>ORF1</b>	837 <sup>a</sup>	1020	1023	1020	1020	1020	1020
<b>ORF2</b>	1587	1581	1590	1587	1593	1593	1590
<b>UTR2</b>	114	114	114	114	114	114	114
<b>ORF3</b>	600	600	603	600	600	603	603
<b>ORF4</b>	462	462	462	462	465	462	462
<b>ORF5</b>	1332	1335	1335	1335	1374	1338	1350
<b>UTR3</b>	129	129	129	129	111	117	108
<b>ORF6</b>	123	105	120	123	132	126	129
<b>UTR4</b>	420 <sup>a</sup>	627	636	630	237	624	630

<sup>a</sup> Partial sequence

#### 4.10. Comparisons with other luteoviruses

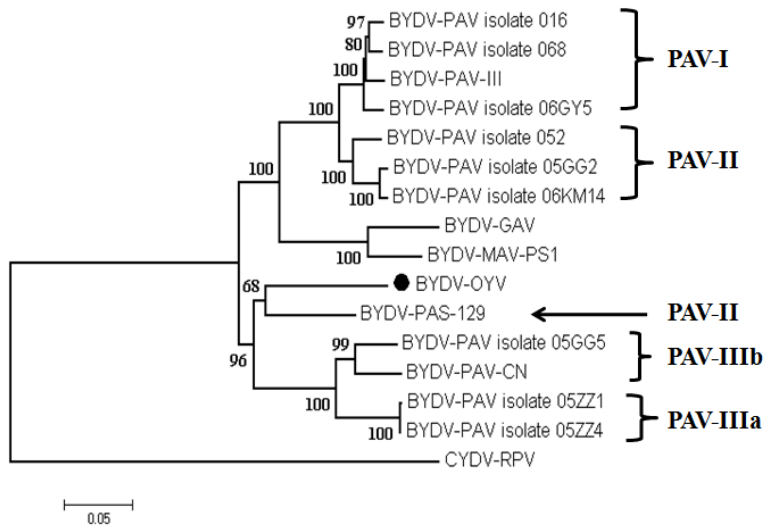
**Table 5. Nucleotide and deduced amino acid sequence identities of BYDV-OYV ORFs and UTRs with other members of the genus *Luteovirus***

Region	BYDV-PAV-CN	BYDV-PAS-129	BYDV-PAV-III	BYDV-PAV isolate 05ZZ1	BYDV-PAV isolate 06KM14	BYDV-GAV	BYDV-PAV isolate 016
ORF1	71.6 (71.3) <sup>a</sup>	<b>84.7 (87.8)</b>	74.0 (75.9)	73.0 (79.5)	72.9 (75.2)	72.4 (72.3)	75.5 (75.2)
ORF2	85.5 (84.9)	<b>90.0 (94.7)</b>	81.4 (89.2)	85.8 (91.3)	81.0 (87.9)	80.7 (87.5)	81.0 (88.8)
UTR2	95.6	<b>97.4</b>	92.9	95.6	93.8	88.5	92.9
ORF3	<b>88.2 (87.9)</b>	81.4 (71.5)	81.6 (74.0)	<b>88.5</b> (87.4)	81.8 (74.4)	75.2 (65.3)	81.9 (73.5)
ORF4	<b>91.1 (86.3)</b>	84.0 (79.1)	83.7 (78.4)	<b>91.1 (86.3)</b>	83.8 (80.4)	78.1 (71.9)	84.0 (80.4)
ORF5	74.7 (79.6)	<b>77.0</b> (62.4)	76.3 ( <b>82.8</b> )	73.4 (54.3)	76.0 (63.0)	58.7 (58.2)	<b>77.0</b> (60.0)
UTR3	86.1	<b>86.8</b>	79.3	84.4	82.1	59.4	79.3
ORF6	59.4 (55.0)	<b>62.2</b> (41.0)	50.4 (42.4)	61.8 (50.0)	58.7 (39.0)	53.5 (44.7)	52.9 (42.5)
UTR4	80.9	83.0	75.4	<b>85.4</b>	73.2	75.9	74.9
Genome	80.1	<b>83.3</b>	78.1	80.5	77.7	71.9	78.0

<sup>a</sup>The values within brackets are identities of deduced amino acid sequences. The values in bold show the highest nt and aa identities.

Nucleotide and deduced amino acid sequence identities were calculated for ORFs and UTRs of BYDV-OYV relative to other members of the genus *Luteovirus* available in GenBank. The results showed that the determined nucleotide sequence of BYDV-OYV (5,132 nt) shared 72 - 83 % identity with other luteoviruses. Sequence identities of each ORF and UTR of BYDV-OYV with other luteoviruses ranged from 50 to 97 %. The BYDV-OYV genome and BYDV-PAS-129 shared highest nucleotide identity (83.3 %). BYDV-OYV and BYDV-PAS-129 shared highest nucleotide and deduced amino acid identities for ORF1 (84.7 and 87.8 %, respectively), ORF2 (90.0 and 94.7 %, respectively) and highest nucleotide identity for UTR2 (97.4 %). Among individual ORFs and UTRs, UTR2 of all compared isolates shared highest nucleotide sequence identities at 88.5 - 97.4 % and ORF6 of all compared isolates shared lowest nucleotide sequence identities at 50.4 - 62.2 %. At the 3' end of the genome, UTR4 of BYDV-OYV showed highest identity to BYDV-PAV isolate 05ZZ1 (85.4 %) (Table 5).

#### 4.11. Phylogenetic analysis



**Fig. 15.** Neighbour-joining analysis of sequences of BYDV-OYV and members of the genera *Luteovirus* and *Polerovirus* based on alignment of the almost complete genome of BYDV-OYV (5,132 nt). The bootstrap values are indicated at the branch points. Horizontal lines are in proportion to the number of nucleotide differences between branch nodes. The scale bar corresponds to 0.05 estimated nucleotide substitutions per site. Accession numbers are listed in Table 3.

The phylogenetic analysis with the almost complete genome sequence (5,132 nt) indicated that BYDV-OYV showed closest relationship with BYDV-PAS-129 (PAV-II), but with a bootstrap value of only 68 % (Fig. 15). In the previous study (Sathees, 2015), the phylogenetic analysis of the 3' part (2,792 nt) suggested that BYDV-OYV is more closely related to PAV-III.

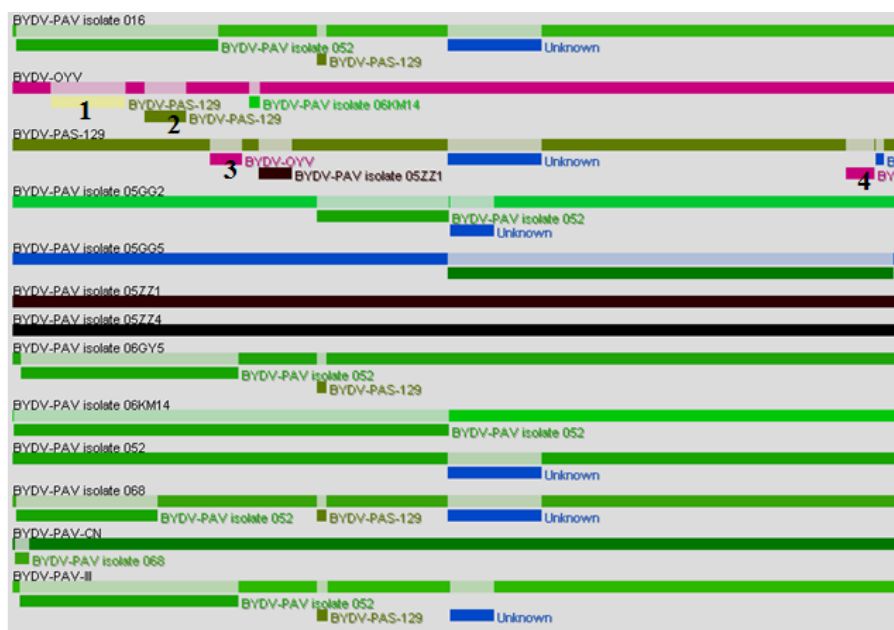
#### 4.12. Evidence for recombination of BYDV-OYV

Two putative recombination events were detected in the genome sequence of BYDV-OYV. The first recombination event was supported by only two of the seven detection methods used (Fig. 16) and it had starting and ending breakpoints at nt positions 112 and 660, respectively.

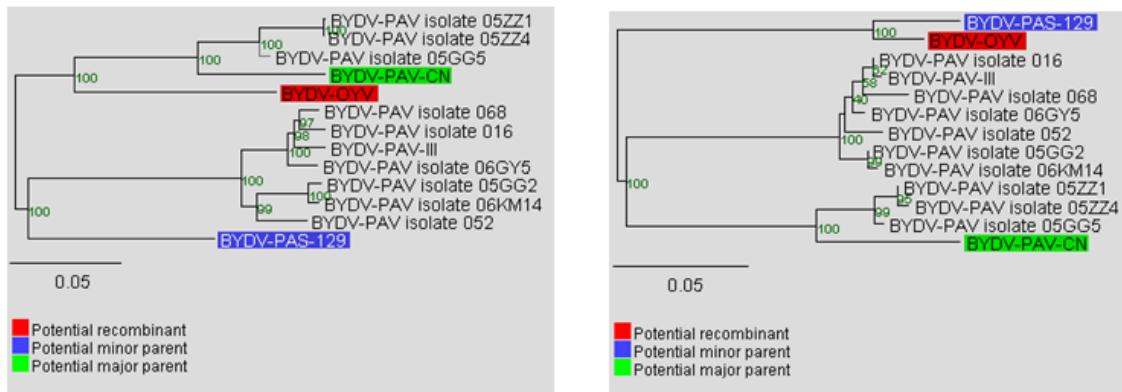
The second recombination event was supported by four methods (RDP -  $1.061 \times 10^{-06}$ , GENECONV -  $7.942 \times 10^{-04}$ , Chimaera -  $1.632 \times 10^{-03}$  and MaxChi -  $3.976 \times 10^{-04}$ ) (Fig. 16). The second recombination event with the parents BYDV-PAS-129 (GenBank AF218798) (minor) and BYDV-PAV-CN (GenBank AY855920) (major) had starting and ending breakpoints at nt positions 772 and 1013, respectively. The beginning and ending breakpoints were located within ORF1 and ORF2, respectively (Fig. 14 and Fig. 16).

Two putative recombination events (recombination events 3 and 4) were also detected with BYDV-OYV as the minor parent in the genome sequence of BYDV-PAS-129 (Fig. 16). The third event with BYDV-PAV isolate 016 as major parent had starting and ending breakpoints at nt positions 1155 and 1343, respectively. This recombination was identified in ORF2 of the genome of BYDV-PAS-129. The third recombination event was supported by three of the seven detection methods (RDP -  $2.072 \times 10^{-04}$ , GENECONV -  $1.821 \times 10^{-03}$  and Chimaera -  $2.483 \times 10^{-02}$ ). The fourth recombination event was supported by only two of the seven detection methods used (Fig. 16) and the recombination had starting and ending breakpoints at nt positions 4,849 and 5,017, respectively.

Phylogenetic analyses including the major and minor parental regions of recombination events 2 and 3 confirmed these two recombination events (2 and 3) with high bootstrap support (Fig. 17 and Fig. 18).



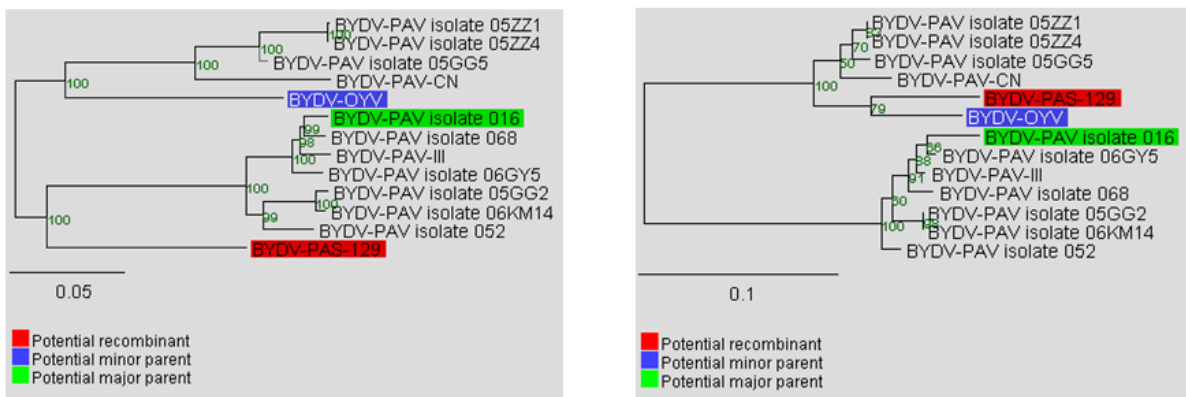
**Fig. 16.** Schematic sequence display of the recombinant regions.



A

B

**Fig. 17.** Phylogenetic trees constructed using neighbour-joining analysis to confirm recombination event 2 in the genome of BYDV-OYV. **A.** The tree was inferred from the major region. **B.** The tree was inferred from the minor region. The sequence marked with red is the recombinant (BYDV-OYV), blue indicates the minor parent (BYDV-PAS-129) and green represents the major parent (BYDV-PAV-CN). Numbers indicate bootstrap support values. All branch lengths are drawn to a scale of nucleotide substitutions per site.



A

B

**Fig. 18.** Phylogenetic trees constructed using neighbour-joining analysis to confirm recombination event 3 in the genome of BYDV-PAS-129. **A.** The tree was inferred from the major region. **B.** The tree was inferred from the minor region. The sequence marked with red is the recombinant (BYDV-PAS-129), blue indicates the minor parent (BYDV-OYV) and green represents the major parent (BYDV-PAV isolate 016). Numbers indicate bootstrap support values. All branch lengths are drawn to a scale of nucleotide substitutions per site.

## 5. Discussion

The study aimed to characterize a new virus, BYDV-OYV, by determination and analyses of the complete genome sequence. Initially, BYDV-OYV was discovered in Latvia and it was suggested as a new tentative species within the family *Luteoviridae* (Bisnieks et al., 2004). Subsequently, the same species was identified from different parts of Sweden, including the county of Jämtland, Sweden (Eriksson et al., unpublished). Determination of the partial genome sequence of BYDV-OYV (2,792 nt of the 3' end) gave further support for the classification of BYDV-OYV as a new species, because BYDV-OYV showed more than 10% nt difference to all other isolates of family *Luteoviridae* (Sathees, 2015). According to the guidelines of ICTV, a virus could be separated as a distinct species within the family *Luteoviridae*, if any viral product differs by more than 10% at amino acid level (Domier, 2012). Although other classification criteria, which include host range, symptomatology and vector specificity, could be used to classify the plant viruses, the degree of relatedness of the genome sequences of BYD-associated viruses is more suitable criterion (Sathees, 2015).

In the current study, the method of 5' and 3' RACE was used in attempt to determine the terminal sequences of the BYDV-OYV genome. However, the RACE attempts failed to provide the expected fragments. The missing sequences are in total around 520 nt (~323 nt at the 5' end and ~197 nt at the 3' end) for obtaining the complete genome sequence of BYDV-OYV. Mostly, the sequences obtained from 5' and 3' RACE did not correspond to known sequences of the BYDV-OYV genome or other members of the family *Luteoviridae*. According to the results, this could be due to mis-priming of the gene-specific primers designed. In this case it seems that the conditions for the first round PCR were not optimal. Poor quality RNA or low viral RNA concentration of the plant extract could also be the reason for non-specific products being amplified. The quality of the isolated RNA could have been assessed e.g. by high resolution denaturing gel electrophoresis using a Bioanalyzer. A new extract of a BYDV-infected plant could be used for PCR amplification. Other factors leading to long 5' T stretches in the 5' terminal region of the final PCR product may be wrong annealing conditions resulting in internal binding of the Oligo-d(T) Anchor Primer to the 5' poly(A) tail.

The almost complete genome sequence of BYDV-OYV showed highest nt identity with BYDV-PAS-129 (83.3 %). If the missing sequences (520 nt) had shown 100 % identity with



BYDV-PAS-129, the total nt identity would have been 84.8 %. Regarding the complex relationship among individual ORFs and UTRs between BYDV-OYV and other members of the genus *Luteovirus*, it is possible that BYDV-OYV may have evolved through recombination events similar to BYDV-PAV-CN. The 5' UTR and ORF1 of PAV-CN are most closely related to GAV, while PAV-CN is most closely related to BYDV-PAS-129 for all other regions (Liu et al., 2007). BYDV-OYV location in phylogenetic trees, which were constructed using different regions of the BYDV-OYV genome, also indicated recombination among isolates of the family *Luteoviridae*. The position of BYDV-OYV in a phylogenetic analysis with the 3' end of the genome of BYDV-OYV (2,792 nt) and other isolates of family *Luteoviridae* showed that BYDV-OYV is more closely related to the groups of PAV-IIIa and PAV-IIIb (Liu et al., 2007; Wu et al., 2011; Sathees, 2015), while the phylogenetic analysis with the almost complete genome sequence of BYDV-OYV showed that BYDV-OYV is more closely related to BYDV-PAS-129.

Sequences of thirteen virus isolates were analysed using RDP4 (Martin et al., 2015) to identify putative recombination events in the genome of BYDV-OYV. The analysis gave a good explanation for the complex relationship among isolates. BYDV-OYV seems to have formed through recombination between ancestors of BYDV-PAV-CN (major parent) and BYDV-PAS-129 (minor parent). Sequence comparisons showed that the 5' region of the BYDV-OYV genome (ORF1-84.7 % and ORF2-90.0 %) shared the highest nt identity with that of BYDV-PAS-129. Other parts of the genome of BYDV-OYV are most closely related to BYDV-PAV-CN.

Results of phylogenetic analyses with the recombinant genome region of BYDV-OYV supported that the genome of BYDV-OYV has undergone a recombination event during its evolution. The recombination event was detected within ORF1 and ORF2. In another study, a higher frequency of recombination was identified as a hot spot in the overlapping region between ORF1 and ORF2 for BYDV, CABYV, CYDV, ScYLV and TuYV, which belong to the family *Luteoviridae* (Pagan & Holmes, 2010). Similar results have been reported for viruses of the genus *Luteovirus* (Wu et al., 2011).

Overall, the current study and previous studies suggest that recombination is a noticeable and very frequent phenomenon within viruses of the family *Luteoviridae* (Moonan et al., 2000; Miller et al., 2002; Pagan & Holmes, 2010; Wu et al., 2011; Boulila, 2011). There is mounting evidence that recombination leads to rapid evolvement of new viruses (Chare &

Holmes, 2006; Codoner & Elena, 2008). Accordingly, a greater understanding of the role of recombination events in plant viruses is important for the development of appropriate control strategies. Recombination may result in host-switching and resistance breaking (Codoner & Elena, 2008; Sztuba-Solinska et al., 2011; Bujarski, 2013). For instance, virus isolates, which have undergone recombination events with rapid generation of new genotypes, may be able to overcome the immune system of new hosts (Chen et al., 2002; Miras et al., 2014).

In conclusion, molecular analysis conducted in this study shows that BYDV-OYV can be regarded as a new species within the family *Luteoviridae*. Further, BYDV-OYV seems to have evolved through recombination between ancestors of BYDV-PAS-129 and BYDV-PAV-CN. However, the 5' and 3' ends of the genome sequence of BYDV-OYV remain to be determined.

## References

- Bisnieks, M., Kvarnheden, A., Sigvald, R. & Valkonen, J.P.T. 2004. Molecular diversity of the coat protein-encoding region of *Barley yellow dwarf virus-PAV* and *Barley yellow dwarf virus-MAV* from Latvia and Sweden. *Arch. Virol.* 149: 843-853.
- Boni, M., Posada, D. & Feldman, M. 2007. An exact nonparametric method for inferring mosaic structure in sequence triplets. *Genetics* 176: 1035-1047.
- Boni, M., Zhou, Y., Taubenberger, J. & Holmes, E. 2008. Homologous recombination is very rare or absent in human influenza A virus. *J. Virol.* 82: 4807-4811.
- Boulila, M. 2011. Selective constraints, molecular recombination structure and phylogenetic reconstruction of isometric plant RNA viruses of the families *Luteoviridae* and *Tymoviridae*. *Biochimie* 93: 242-253.
- Bujarski, J.J. 2013. Genetic recombination in plant-infecting messenger-sense RNA viruses: overview and research perspectives. *Frontiers Plant Sci.* 4: 68.
- Chare, E.R. & Holmes, E.C. 2006. Phylogenetic survey of recombination frequency in plant RNA viruses. *Arch. Virol.* 151: 933-946.
- Chay, C.A., Gunasinge, U.B., Dinesh-Kumar, S.P., Miller, W.A. & Gray, S.M. 1996. Aphid transmission and systemic plant infection determinants of barley yellow dwarf virus luteovirus-PAV are contained in the coat protein, readthrough domain and 17 kDa protein, respectively. *Virology* 219: 57-65.
- Chen, Y.K., Goldbach, R. & Prins, M. 2002. Inter and intra molecular recombinations in the *Cucumber mosaic virus* genome related to adaptation to *Alstroemeria*. *J. Virol.* 76: 4119-4124.
- Codoner, F.M. & Elena, S.F. 2008. The promiscuous evolutionary history of the family *Bromoviridae*. *J. Gen. Virol.* 89: 1739-1747.
- D'Arcy, C.J. 1995. Symptomatology and host range of barley yellow dwarf. In: D'Arcy, C. J. & Burnett, P.A. (Eds). *Barley yellow dwarf: 40 years of progress*. APS Press. St. Paul, MN, pp. 9-28.

- Domier, L.L. 1995. Genome structure and function of barley yellow dwarf viruses. In: D'Arcy, C. J. & Burnett, P.A. (Eds). Barley yellow dwarf: 40 years of progress. APS Press. St. Paul, MN, pp. 181-201.
- Domier, L.L., McCoppin, N.K., Larsen, R.C. & D'Arcy, C.J. 2002. Nucleotide sequence shows that Bean leafroll virus has a luteovirus-like genome organization. J. Gen. Virol. 83: 1791-1798.
- Domier, L.L. 2012. Family *Luteoviridae*. In: King, A.M.Q., Adams, M.J., Carstens, E.B. & Lefkowitz, E.J. (Eds). Virus taxonomy. 9<sup>th</sup> report of the International Committee on Taxonomy of Viruses. Elsevier/Academic Press. Oxford, UK: pp. 1045-1053.
- Elena, S.F. & Sanjuan, R. 2008. Virus evolution, insights from an experimental approach. Annu. Rev. Ecol. Evol. Syst. 38: 27-52.
- Fattouh, F.A., Ueng, P.P., Kawata, E.E., Barbara, B.A., Larkins, B.A. & Lister, R.M. 1990. Luteovirus relationship assessed by cDNA clone from barley yellow dwarf viruses. Phytopathology 80: 913-920.
- Fiebig, M., Poehling, H.M. & Borgemeister, C. 2004. Barley yellow dwarf virus, wheat and *Sitobion avenae*: a case of trilateral interactions. Entomol. Exp. Appl. 110: 11-24.
- Gibbs, M., Armstrong, J. & Gibbs, A. 2000. Sister-Scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. Bioinformatics 16: 573-582.
- Gray, S.M. 1996. Plant virus proteins involved in natural vector transmission. Trends Microbiol. 4: 259-264.
- Gray, S. & Gildow, F. E. 2003. Luteovirus-aphid interactions. Annu. Rev. Phytopathol. 41: 539-566.
- Hogenhout, S.A., Ammar, D., Whitfield, A.E. & Redinbaugh, M.G. 2008. Insect vector interactions with persistently transmitted viruses. Annu. Rev. Phytopathol. 46: 327-359.
- Irwin, M.E. & Thresh, J.M. 1990. Epidemiology of barley yellow dwarf: a study in ecological complexity. Annu. Rev. Phytopathol. 28: 393-424.

- Knierim, D., Deng, T.C., Tsai, W.S., Green, S.K. & Kenyon, L. 2010. Molecular identification of three distinct polerovirus species and a recombinant cucurbit aphid-borne yellows virus strain infecting cucurbit crops in Taiwan. *Plant Pathol.* 59: 991-1002.
- Koev, G., Liu, S., Beckett, R. & Miller, W.A. 2002. The 3'-terminal structure required for replication of barley yellow dwarf virus RNA contains an embedded 3' end. *Virology* 292: 114-126.
- Kozłowska-Makulska, A., Guilley, H., Szyndel, M.S., Beuve, M., Lemaire, O., Herrbach, E. & Bouzoubaa, S. 2010. P0 proteins of European beet-infecting poleroviruses display variable RNA silencing suppression activity. *J. Gen. Virol.* 91: 1082–1091.
- Krueger, E.N., Beckett, R.J., Gray, S.M. & Miller, W.A. 2013. The complete nucleotide sequence of the genome of barley yellow dwarf virus-RMV reveals it to be a new polerovirus distantly related to other yellow dwarf viruses. *Frontiers Microbiol.* 4: 205.
- Lim, S., Yoo, R.H., Igori, D., Zhao, F., Kim, K.H. & Moon, J.S. 2015. Genome sequence of a recombination brassica yellows virus infecting Chinese cabbage. *Arch. Virol.* 160: 597-600.
- Liu, F., Wang, X., Liu, Y., Xie, J., Gray, S. M., Zhou, G. & Gao, B. 2007. A Chinese isolate of barley yellow dwarf virus-PAV represents a third distinct species within the PAV serotype. *Arch. Virol.* 152: 1365-1373.
- Liu, X., Wu, C. & Chen, A. 2010. Codon usage bias and recombination events for neuraminidase and hemagglutinin genes in Chinese isolates of influenza A virus subtype H9N2. *Arch. Virol.* 155: 685-693.
- Liu, Y., Hao, Z., Zhao, K., Wu, B. & Wang, X. 2012. Two suppressors of RNA silencing encoded by cereal-infecting members of the family *Luteoviridae*. *J. Gen. Virol.* 93: 1825-1839.
- Maia, I.G., Goncalves, M.C., Arruda, P. & Vega, J. 2000. Molecular evidence that sugarcane yellow leaf virus (ScYLV) is a member of the Luteoviridae family. *Arch. Virol.* 145: 1009-1019.
- Mangwende, T., Wang, M.L., Borth, W., Hu, J., Moore, P.H., Mirkov, E.T. & Albert, H.H. 2009. The P0 gene of *Sugarcane yellow leaf virus* encodes an RNA silencing suppressor with unique activities. *Virology* 384:38–50.

- Martin, D., Posada, D., Crandall, K. & Williamson, C. 2005. A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS RES. Hum. Retroviruses* 21: 98-102.
- Martin, D. & Rybicki, E. 2010. RDP: detection of recombination amongst aligned sequences. *Bioinformatics* 16: 562-563.
- Martin, D., Lemey, P. & Posada, D. 2011. Analysing recombination in nucleotide sequences. *Mol. Ecol. Resour.* 11: 943-955.
- Martin, D. P., Murrell, B., Michael, G., Khoosal, A. & Muhire, B. 2015. RDP4: detection and analysis of recombination patterns in virus genomes. *Virus Evol.* 1: 1-5.
- Maynard Smith, J. 1992. Analyzing the mosaic structure of genes. *J. Mol. Evol.* 34: 126-129.
- Mayo, M.A. 2002. ICTV at the Paris ICV: Results of the plenary session and the binomial ballot. *Arch. Virol.* 147: 2254-2260.
- Mayo, M.A., Barker, H., Robinson, D.J., Tamada, T. & Harrison, B.D. 1982. Evidence that potato leafroll virus RNA is positive-stranded, is linked to a small protein and does not contain polyadenylate. *J. Gen. Virol.* 59: 163-167.
- Mayo, M.A. & Miller, W.A. 1999. The structure and expression of luteovirus genome. In: Smith, H.G. & Barker, H. (Eds). *The Luteoviridae*. CABI publishing, Wallingford. pp. 23-42.
- Miller, W.A., Liu, S. & Beckett, R. 2002. Barley yellow dwarf virus: Luteoviridae or Tombusviridae? *Mol. Plant Pathol.* 3: 177-183.
- Miras, M., Sempere, R.N., Kraft, J.J., Miller, W.A., Aranda, M.A. & Truniger, V. 2014. Interfamilial recombination between viruses led to acquisition of a novel translation-enhancing RNA element that allows resistance breaking. *New Phytol.* 202: 233-246.
- Moonan, F., Molina, J. & Mirkov, T.E. 2000. *Sugargane yellow leaf virus*: an emerging virus that has evolved by recombination between luteoviral and poleroviral ancestors. *Virology* 269: 156-171.
- Nega, A. 2014. Review on barley yellow dwarf viruses. *J. Biol. Agric. Healthc.* 4: 64-90.
- Oswald, J.W. & Houston, B.R. 1951. A new virus disease of cereals, transmissible by aphids. *Plant Dis. Rep.* 35: 471-475.

- Padidam, M., Sawyer, S. and Fauquet, C. 1999. Possible emergence of new geminiviruses by frequent recombination. *Virology* 265: 218-225.
- Pagan, I. & Holmes, E.C. 2010. Long-term evolution of the *Luteoviridae*: time scale and mode of virus speciation. *J. Virol.* 84: 6177-6187.
- Peter, K.A., Gildow, F., Palukaitis, P. & Gray, S.M. 2009. The C terminus of the polerovirus P5 readthrough domain limits virus infection to the phloem. *J. Virol.* 83: 5419–5429.
- Pfeffer, S., Dunoyer, P., Heim, F., Richards, K.E., Jonard, G. & Ziegler-Graff, V. 2002. P0 of beet western yellows virus is a suppressor of posttranscriptional gene silencing. *J. Virol.* 70: 6815-6824.
- Posada, D. & Crandall, K.A. 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc. Natl. Acad. Sci. U.S.A.* 98: 13757-13762.
- Power, A.G. & Remold, S.K. 1996. Incidence of barley yellow dwarf virus in wild grass populations: implications for biotechnology risk assessment. Levin, M., Grim, C. & Angle, S. (Eds) *Proceedings of the Biotechnology Risk Assessment Symposium*, University of Maryland, Gaithersburg, MD, USA: pp. 58-65.
- Rochow, W.F. 1969. Biological properties of four isolates of barley yellow dwarf virus. *Phytopathology* 59: 1580–1589.
- Rochow, W.F. & Muller, I. 1971. A fifth variant of barley yellow dwarf virus in New York. *Plant Dis.* 55: 874-877.
- Rochow, W.F. & Duffus, J.E. 1981. Luteoviruses and yellows diseases. In: Kursiak, E. *Handbook of plant virus infections and comparative diagnosis* (Eds.). Elsevier Press, The Netherlands. pp. 147-170.
- Sacristan, S., Fraile, A., Malipica, J.M. & Garcia-Arenal, F., 2005. Analysis of host adaptation and its relationship with virulence in Cucumber mosaic virus. *Phytopathology* 95: 827-833.
- Sathees, K. 2015. Sequence characterization of Barley yellow dwarf virus-OYV. Independent project in Biology, 30 hp, EX0565. Examensarbete / Swedish University of Agricultural Sciences (SLU), Uppsala. 137.

Smith, J.M. 1992. Analyzing the mosaic structure of genes. *J. Mol. Evol.* 34: 126-129.

Sztuba-Solinska, J., Urbanowicz, A., Figlerowicz, M. & Bujarski, J.J. 2011. RNA-RNA recombination in plant virus replication and evolution. *Ann. Rev. Phytopathol.* 49: 415-443.

Van der Wilk, F., Verbeek, M., Dullemans, A.M. & van den Heuvel, J.F. 1997. The genome-linked protein of potato leafroll virus is located downstream of the putative protease domain of the ORF1 product. *Virology* 234: 300-303.

Wu, B., Blanchard-Letort, A., Liy, Y., Zhou, G., Wang, X. & Elena, S.F. 2011. Dynamics of molecular evolution and phylogeography of barley yellow dwarf virus-PAV. *PLoS One* 6: e16896.

Ziegler-Graff, V., Brault, V., Mutterer, D., Simonis, M.T., Herrbach, E., Guilley, H., Richards, K.E. & Jonard, G. 1996. The coat protein of beet western yellows luteovirus is essential for systemic infection but the viral gene products P29 and P19 are dispensable for systemic infection and aphid transmission. *Mol. Plant-Microbe Interact.* 9: 501-510.



## **Acknowledgements**

There are a number of people, I would like to thank, whose time and support made this thesis work possible. First, I would like to express my thanks and gratitude to my supervisor Professor Anders Kvarnheden for his research guidance and support and for sharing his knowledge and experience. I would also like to thank my examiner Professor Johan Meijer for his comments and advice. I am indebted to virology group members and Department of Plant Biology staff for their help during the lab work.

Finally, I owe much to my family, who indirectly enabled me to undertake and complete this research. I am thankful for my parents for their love. I am grateful to my husband for his love, support and constant source of encouragement. I am also grateful to my son, who makes me always happy and joy.





**Nr 153**  
**Uppsala 2016**

---

**Department of Plant Biology**  
**SLU**  
**Box 7080**  
**75007 Uppsala, Sweden**