Virus integration and tandem repeats in the genomes of Petunia

A thesis submitted to the University of Leicester for the degree of Doctor of Philosophy

By

Osamah Nadhim Kadhim Alisawi M.Sc.

January 2019

Department of Genetics and Genome Biology

Abstract

The integration of endogenous pararetroviruses (EPRVs) and tandemly repeated sequences were examined in whole genome raw reads, and two genome assemblies, in diploid *Petunia* species including hybrid-derivatives and their ancestors, using bioinformatics, molecular biology, cytogenetics and microscopy. Three types of EPRV clusters (petuvirus-, florendovirus- and caulimovirus-like sequences) were found. Chromosomal signals of PVCV (*Petunia vein clearing virus*) were seen by *in situ* hybridization in all *Petunia* species. Fragmented parts of four novel florendovirus-like sequences were found and the complete sequence was reconstructed, adding petunia to the 27 known host species. Chromosome III of *P. axillaris* and *P. hybrida* Rdc showed strong pericentromeric signal of PVCV and *Florendovirus* suggesting both EPRVs have similar positions, integration patterns and endogenization events (unlike *P. integrifolia* subsp *inflata* and *P. axillaris* subsp *parodii*). The caulimovirus-like sequence cluster was less abundant in genomes, with four novel members. RNA analysis from infected and healthy petunia samples revealed expression of endogenous PVCV and *Caulimovirus* sequences, unlike *Florendovirus* (not detected in RNA). The episomal form of vertically transmitted PVCV was integrated near the telomere of heterologous chromosomes. Transmission electron microscopy (TEM) showed differences in number and size of PVCV particles and inclusion bodies for both chlorotic spots and vein clearing symptoms, the latter correlated with PVCV particles in cytoplasm from vascular bundle cells. In plants with chlorotic symptoms, infected cells contained virions in parenchyma cells, while scattered virions were seen in chlorotic spots in *P. hybrida* W138 after heat induction of symptoms. Eight unique types of tandem repeat clusters were analysed within *Petunia* raw reads with variable genome proportions and different loci on mitotic chromosomes. Three were useful markers for chromosome identification. Taken together, the work shows the contribution of repetitive DNA to diversity and variation within petunia genomes, and has consequences for evolution, and both resistance and spread of some viruses.

Declaration

I hereby declare that no part of this thesis has been previously submitted to this or any other university as part of the requirements for a higher degree. The study described in this thesis, unless otherwise acknowledged in the text or by reference, was conducted by the undersigned who is fully responsible.

This work was achieved in the Department of Genetics and Genome Biology, University of Leicester and Julius Kühn-Institut (JKI), Braunschweig, Germany (see Appendix 4.1), during the period from January 2015 to December 2018.

Signed:……...

Date:……

Osamah Nadhim Kadhim Alisawi

Dedication

Thanks to ALLAH for blessing me much more than I deserve

This work is dedicated to the sacred two date palm trees grown on Euphrates river beach, covered me with their fronds, and taken care of me for my entire life, proud of you both my great father and mother

Nadhim Kadhim Alisawi and **Rabab Waheed Alisawi**

My lovely, amazing wife **Wasan Riyadh Alisawi**, who brilliantly supported me to build up this story.

Future men, my little boys, **Zaid, Yazan** and newborn baby **Taim**.

My beloved sisters (**Israa** and **Rawaa**) and brothers (**Ahmed** and **Yaseen**).

I would also send my fulfilment and sincerity to my great grandfather who passed away and left behind a massive heritage for the whole tribe (**Al-Isa**) and **Iraqi** community.

Al Sheikh Haj Waheed Abood Alisawi (1927-2012)

, my respected uncles

Dr. Abood Waheed Alisawi and **Dr. Riyadh Waheed Alisawi**

Finally, thanks must go to my homeland where the first ever civilization established

IRAQ

Acknowledgements

First and foremost, I would like to express my sincere gratitude and appreciation to my three supervisors Prof. Pat Heslop-Harrison, Dr. Trude Schwarzacher and Dr. Katja Richert-Pöggeler, for their support, guidance, kindness and generosity over the whole time of study.

I am indebted to Prof. David Twell and Dr. Richard Gornall for their positive comments and notes.

I would also like to acknowledge my thanks to Mr. Ramesh Patel and my colleagues in lab 201 Dr. Sarbast Ihsan Mustafa, Dr. Jotyar Jassim Muhammed, Dr. Rubar M. Salih, Dr. Nouf Fakieh Alsayied and Noorhariza Mohd Zaki.

My special thanks go to very gentle members of Dr. Katja Richert-Pöggeler's lab in Julius Kühn-Institut (JKI), Braunschweig, Germany; Mr. Dirk Schmalowski, Ms. Christina Maaß (Grille) and Ms. Sabine Schuhmann for their assistance and fruitful cooperation.

I highly appreciate the full cooperation of Assistant Prof. Aureliano Bombarely, Department of Horticulture, Virginia Tech, USA, who provided me with raw reads and assemblies of *Petunia* species.

I wish to thank my Iraqi friends Dr. Wadhah Mahbooba, Dr. Ghazwan Al Hassan, Dr. Pshtiwan Bebane, Dr. Tariq A. Kareem, Mr. Imad Younis, Mr. Muhammed Abdul Sattar, Mr. Wisam Aziz and Mr. Maitham Al-Jalali for their kind help.

I would like to thank my sponsorship; the Iraqi Ministry of Higher Education and Scientific research and my home department; Plant Protection Department, Faculty of Agriculture, University of Kufa for their funding and giving me this great opportunity.

Last, and by no means least, I wish to thank all people who have pushed me forward to be in progress all time.

Contents

Abbreviations

bp	Base pairs
kb	Kilo bases
Mbp	Mega bases
pg	Picogram
C	C-value
Gb	Giga bases
PVCV	*Petunia vein clearing virus*
ICTV	International Committee on Taxonomy of Viruses
TSD	Target site duplications
DNA	Deoxyribonucleic acid
rDNA	Ribosomal DNA
cDNA	Complementary deoxyribonucleic acid
ITS	Internal transcriped spacer
cpDNA	Chloroplast DNA
mtDNA	Mitochonderia DNA
RE	RepeatExplorer
RNA	Ribonucleic acid
RNAse	Ribonuclease
Ch.	Chromosome
ORF	Open reading frame
DAPI	4′, 6-diamidino-2-phenylindole
dNTPs	Deoxy nucleotide triphosphates
FISH	Fluorescent *in situ* hybridization
RT	Reverse transcriptase
RH1	Ribonuclase H1
BSA	Bovine Serum Albumin
Epon	Epoxy resin
DDSA	Dodecenyl succinic anhydride
MNA	Methyl nadic anhydride
DMP	Dimethoxy propane
CCD	Charge - coupled device
TEM	Transmission Electron Microscope
IEM	Immuno- Electron Microscopy
ELISA	Enzyme Linked Immuno-Sorbent Assay
IBs	Inclusion Bodies
PPT	Polypurine tract
PBS	Primer binding site
%	Percentage
g	Gravity
NOR	Nucleolar organising region
TEs	Transposable elements
UV	Ultra violet
MS	Murashige Skoog
v/v	Volume per volume
w/v	Weight per volume

µg	Microgram
CL	Cluster
Scf	Scaffold
F	Forward
R	Reverse
MYA	Million years ago
nm	Nanometer
µm	Micrometer
Btn	Biotin
DPV	Description of plant viruses
RFLP	Restriction fragment length polymorphism
TVCV	*Tobacco vein clearing virus*
CaMV	*Cauliflower mosaic virus*
BSV	*Banana streak virus*
BSOLV	*Banana streak Obino l'ewai virus*
BSGFV	*Banana streak Golden finger virus*
BSMYV	*Banana streak Mysore virus*
HHV-6	*Herpesvirus 6*
ICNV	International Committee on Nomenclature of Viruses
HTS	High throughput sequencing
NGS	Next generation sequencing
SNP	Single nucleotides polymorphism
AP	Aspartic proteinase
CP	Coat protein
MP	Movement protein

Chapter I. Introduction

1.1 The genus *Petunia*

The genus *Petunia* (Solanaceae) is a popular ornamental plant cultivated worldwide, but the taxonomy of the genus has not been clear throughout the years and has been changed a few times since the first description by Jussieu (1803). The name of *Petunia* derives from ‘’petum’’ or ‘’betum’’, an original name for the tobacco plant, *Nicotiana tabacum,* that is closely similar to *Petunia nyctaginiflora* (now = *P. axillaris*), that was one of the two *Petunia* species firstly described (Stehmann *et al.* 2009).

In 1803, based on collected material from Montevideo and Uruguay, Jussieu was the first person to describe *Petunia* (Jussieu 1803). As quite distinct species, *P. parviflora* Juss. and *P. nyctaginiflora* Juss. were described in the same paper. Earlier, Lamarck (1793) described *P. nyctaginiflora* Lam. as *Nicotiana axillaris* Lam, and then *Petunia* species were attributed to various genera of the Solanaceae such as *Nicotiana*, *Fabiana*, *Calibrachoa*, *Salpiglossis* and *Nierembergia*. The South American Solanaceae were revised by Miers (1846) who recognized ten species of *Petunia*, five of them new. In the *Flora of Brazil*, Sendtner (1850) revised Solanaceae with nine of the thirteen described *Petunia* species being new. In 1852, de Candolle’s *Prodromus* publication, presented sixteen *Petunia* species by Dunal (1852) and transferred three species to the genus *Fabiana*; a new genus *Leptophragma* was described but is now considered as a synonym of *Calibrachoa* (Stehmann *et al.* 2009).

The first monograph of *Petunia* was published by Fries (1911) which included 27 species with nine species described as new. The morphology, geographic distribution, circumscription, and relationships between *Petunia* and other genera within Solanaceae were discussed in detail in this monograph and it is still the latest available revision of the genus (Stehmann *et al.* 2009). Some species of *Petunia* and *Calibrachoa* share similar floral and vegetative morphology in addition to geographic distribution, presenting difficulties to differentiate these two genera. Nevertheless, Wijsman’s (1983) decision to separate the two genera (*Petunia* and *Calibrachoa*) has been confirmed by anatomical, reproductive, cytotaxonomic and chemical studies. Additionally, all *Petunia* species have seven pairs of chromosomes (Watanabe *et al.* 1996), whereas all examined *Calibrachoa* have nine pairs (Stehmann *et al.* 1996; Watanabe *et al.* 1997). There is a high level of cross-incompatibility between *Petunia* species that have

different numbers of chromosomes, indicating that they belong to genetically isolated groups (Wijsman 1983; Watanabe *et al.* 1996). Outcomes from recent molecular research have illuminated phylogenetic relationships using RFLP chloroplast DNA, ITS, cpDNA and mtDNA analyses and support the separation of the two genera, showing them as sister groups (Ando *et al.* 2005; Kulcheski *et al.* 2006). However, some commercial stocks are putative hybrids between *Petunia* and *Calibrachoa* and have been distributed in the floral markets with no clear explanation how they were hybridized. ×*Calitunia* and ×*Petchoa* Supercal were made by the Danziger and Sakata seed companies respectively. ×*Calitunia* is characterized by early blooming, resistance to high soil pH, lush foliage, and a wide colour range while, ×*Petchoa* is characterized by lush foliage, large vigour, sterile flowers and tolerance to hard weather conditions. These hybrids are triploid (3x=25) and two of the chromosome sets come from *Calibrachoa* (Van Meggelen 2005; Olschowski *et al.* 2012; Jędrzejuk *et al.* 2017).

P. inflata is considered as a synonym of *P. integrifolia* by Smith (1966), but Wijsman (1982) restored this plant taxon to be a subspecies under *P. integrifolia*, a common species in southern regions of South America that he divided into three regional sub-species *P. integrifolia* subsp *integrifolia, P. integrifolia* subsp *occidentalis* and *P. integrifolia* subsp *inflata*. Recently, Ando *et al.* (2005) resurrected *P. inflata* as a species separate from *P. integrifolia* according to a morphometric analysis and suggested using straight calyx lobes as a good diagnostic attribute.

Stehmann *et al*. (2009) reported that *Petunia* has 14 species (Table 1.1).

Table 1.1 List of wild species of petunia and their main descriptions (Stehmann *et al*. 2009).

	***Petunia* species**	**Morphology**	**Distribution and habitat**
1	*P. axillaris* (Lam.) Britton, Sterns and Poggenb	‘’Flowers white and tube slightly enlarged or cylindrical toward the top’’.	‘’It exhibits the largest geographic distribution in the genus and is known to occur in Brazil (Rio Grande do Sul), Argentina, Uruguay, Paraguay, and Bolivia. Three allopatric subspecies have been accepted based on corolla tube length and stamen arrangement. Individuals of *P. axillaris* are heliophilous and inhabit rocky sites, but can also be found along roadsides’’.
2	*P. integrifolia* (Hook.) Schinz and Thell	‘’Stems decumbent; flowers purple; capsule subglobose with peduncle deflexed’’.	‘’It inhabits the Pampas province and occurs in Argentina, Uruguay, and southern Brazil (from Rio Grande do Sul to the coast of Santa Catarina), growing on different kinds of substrata (latossols, sandsoils, and litosoils). It can also be found on disturbed areas such as roadsides or cultivated lands’’.
3	*P. interior* T. Ando & Hashim	‘’Flowers purple and anthers with channeled lobes at dehiscence’’.	‘’Its geographic distribution ranges from northwestern Rio Grande do Sul and western Santa Catarina (with some disjunct places) in Brazil to the province of Misiones, Argentina’’.
4	*P. bajeensis* T. Ando & Hashim	‘’Flowers purple and plant viscid; leaves with prominent	‘’ To date found only in the extreme southern region of Rio Grande do Sul, Brazil, in the municipalities

		venation. Vegetatively, the individuals of this species roughly resemble more robust plants of *P. bonjardinensis*, but the morphology of the flowers does not differ from that of *P. integrifolia* except for the larger size of the floral parts''.	of Baj'e, Cangucu, and Lavras do Sul, it can be found growing along roadside slopes''.
5	*P. secreta* Stehmann & Semir	''Flowers purple and plant erect; filaments adnate nearly to middle of corolla tube; pollen yellow''.	'' It is endemic to the place called "Pedra do Segredo" and adjacent areas around the municipality of Cac．apava do Sul, in Rio Grande do Sul, southern Brazil. It is clearly heliophilous, inhabiting the top of conglomerate sandstone towers at about 300–400 m elevation and visited by bees''.
6	*P. bonjardinensis* T. Ando & Hashim	''Flowers purple and stigma exserted above anthers of the longest stamens''.	'' It is endemic to a small area near to the border of the southern Brazilian plateau, in the municipality of Bom Jardim da Serra, Santa Catarina, where it is not difficult to find individuals growing on roadside slopes''.
7	*P. exerta* Stehmann	''Flowers red-orange and Sciophilous plants; anthers and stigma exserted from corolla tube''.	'' This strictly endemic species is known only from the "guaritas" and adjacent areas, at the municipality of Cac．apava do Sul, Rio Grande do Sul, Brazil, growing in shallow caves sculpted by the wind in sandstone towers''.
8	*P. mantiqueirensis* T. Ando & Hashim	''Flowers purple and plant procumbent; filaments adnate below the middle of tube; pollen violet or bluish''.	''It is restricted to the Serra da Mantiqueira, in Minas Gerais, southeastern Brazil, where few populations are known. Individuals of *P. mantiqueirensis* are shade tolerant and grow on the border of the *Araucaria* or montane forests, as well as on more open places, at altitudes ranging from 1000 to 1700 m above sea level''.
9	*P. reitzii* L. B. Sm. & Downs	''Flowers bright red and stigma located below the anthers of the longest pair of stamens''.	'' It is endemic to the oriental border of the southern Brazilian plateau in Santa Catarina and seems to be restricted to a small area between the municipalities of Bom Retiro and Urubici, at altitudes of about 1000 m and associated with *Araucaria* forest. It grows on the walls of small cliffs beside rivers, hanging freely in space, but can also be found along exposed roadside slopes''.
10	*P. saxicola* L. B. Sm. & Downs	''Flowers bright red and stigma slightly exserted above the anthers of longest pair of stamens''.	'' The saxicolous habit of this species is unique in the genus, and individuals are found growing on humid and rocky escarpments of a small area of the border of the southern Brazilian plateau, in the municipality of Otacilio Costa, Santa Catarina. Only one population of *P. saxicola* is known to exist''.
11	*P. scheideana* L. B. Sm. & Downs	''Flowers purple and stigma located at the same level to the anthers of the longest pair of stamens''.	'' The geographic distribution ranges from higher altitudes (800–1000 m) in Parana and Santa Catarina, Brazil, often associated with *Araucaria* forests, westward into the lowlands of extreme northern Misiones, Argentina (about 200–300 m)''.
12	*P. altiplana* T. Ando & Hashim	''Flowers purple and plant repent, rooting at the nodes; leaves widely obovate or orbicular''.	'' This species is distributed in the highlands of Santa Catarina and Rio Grande do Sul, Brazil, in altitudes from 800 to 1200 m, and grows in outcrops or exposed roadside slopes''.
13	*P. occidentalis* R. E. Fr	''Flowers purple and corolla limb 20–25 mm in diameter,	'' It is geographic distribution is restricted to the Sub-Andean mountains (from 650 to 2000 m of altitude) in northwestern Argentina (Jujuy, Salta) and southern Bolivia (Tarija), being separated from

		filaments adnated >7 mm to the corolla tube base''.	the other *Petunia* species by the Chaco, a large, flat region covered by a dry forest, in northern Argentina, Bolivia, and Paraguay''.
14	*P. inflata* R. E. Fr	''Flowers purple and corolla limb 25–40 mm in diameter, filaments adnated <5 mm to the corolla tube base''.	'' It is found in a hybrid zone in northwestern Rio Grande do Sul, Brazil''.

Recently, Reck-Kortmann *et al.* (2014) used nuclear and plastid DNA markers to study twenty species of petunia phylogenetically adding three wild species (*P. riograndensis, P. littoralis, P. guarapuavensis*) to the listed species above in Table 1.1 in addition to three subspecies (*P. axillaris* subsp *subandina, P. axillaris* subsp *parodii, P. integrifolia* subsp *depauperata*). The monophyly and the divergence based on the differentiation of corolla tube length of *Petunia* species were confirmed by this phylogenetic study, while *Petunia* species were geographically distributed as a result of divergence within main clades suggesting the Pampas region as the earliest area of petunia divergence (Figure 1.1).

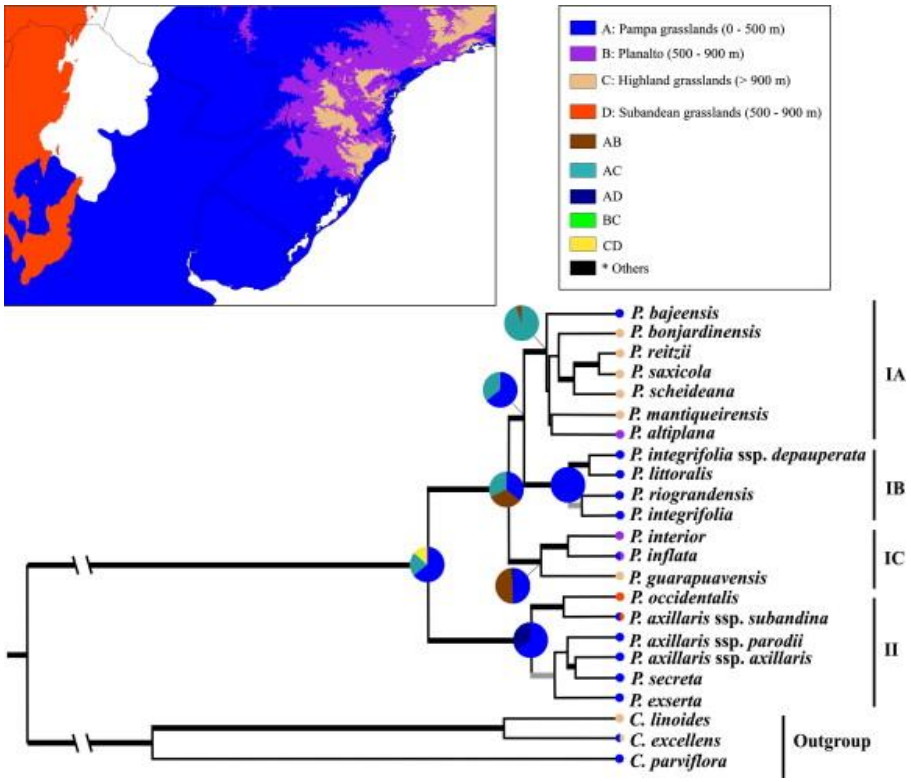


Figure 1.1 Phylogenetic tree of *Petunia* species based on nuclear and plastid sequences. Nodes with 1 or between 0.7 and 0.9 posterior probabilities appear as thick black and gray branches respectively. The principal clades identified in the analysis were appeared to the right as vertical bars. The most likely ancestral areas were shown as pie charts on the nodes, while other reconstructions indicated in black. This tree results were taken from four areas identified in the map at the top of this figure. Figure taken from Reck-Kortmann *et al*. (2014).

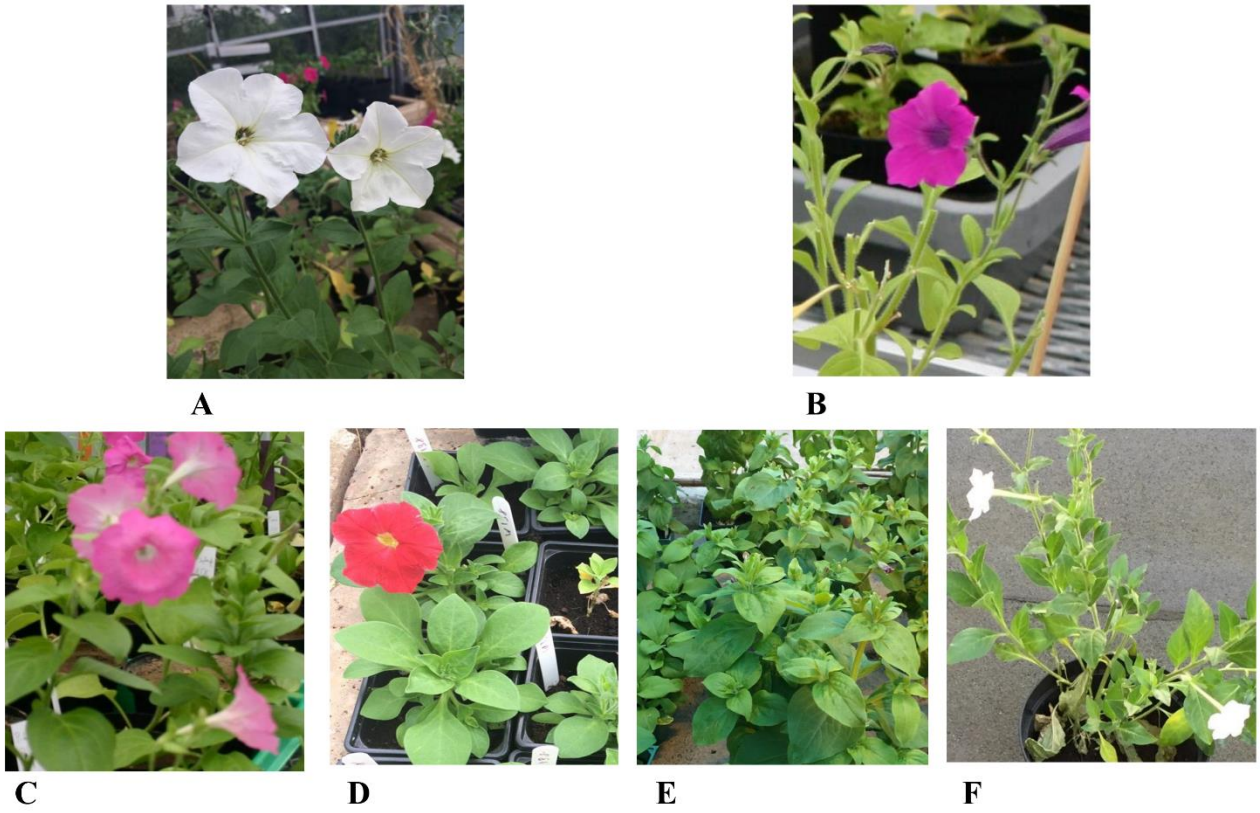


Figure 1.2 The examined *Petunia* species images: *Petunia axillaris* subsp *axillaris* N (A), *P. integrifolia* subsp *inflata* S6(B), *P. hybrida* Rdc (C), *P. hybrida* W138(D), *P. hybrida* V26 (E) and *P. axillaris* subsp *parodii* S7(F).

The garden petunia, *Petunia hybrida* has been considered as a decisive species for the genus taxonomy. It was obtained in (1834) by Atkins of Northampton, a British nurseryman, through hybridization and crossing, and rapidly spread to European nurseries (Sink 1984; Klemm *et al.* 2017). This hybrid described as *Nierembergia atkinsiana* by Sweet (1935), and later, Vilmorin (1863) represented the term *Petunia* x *hybrida* as the garden petunia (Ganga *et al.* 2011). Now, *P. hybrida* is cultivated around the world and is considered a very important Solanaceae ornamental plant. The origin of *P. hybrida* and genetic incompatibility mechanisms have been studied by many researchers (Ferguson & Ottley 1932; Mather & Edwardes 1943; Stout 1952; Van Der Donk 1974; Linskens 1975; Sink 1984). As parents of this hybrid, multiple species have been proposed: Wijsman (1982) reported that this plant was produced from breeding between two different species, *P. axillaris* that has white flowers pollinated by moths and *P. integrifolia* that has purple flowers pollinated by bees. Species of *Calibrachoa* such as *C. calycina* and *C. linearis* (2*n* = 18) have been crossed with the parents of the hybrid plant (2*n* = 14) by Wijsman (1903) without any success. Crosses between plants with different numbers of

chromosomes usually failed, unlike plants with similar chromosomes numbers (Wijsman 1983; Klemm *et al.* 2017).

Petunia is endemic to South America region and distributed between 22° and 39° S subtropical regions (Figure 1.3). Brazil has most Petunia species with thirteen species (except *P. occidentalis*), followed by Argentina with five species, and then Uruguay, Paraguay, and Bolivia with two species for each. Two principal areas in southern Brazil considered as centers of diversity of *Petunia* species: (a) Pampean region lowlands (Figure 1.3A) and (b) southern Brazilian plateau highlands (Figure 1.3B). The two regions are included in the Pampean and Paranense provinces, respectively (Cabrera and Willink 1980). Serra do Sudeste region is considered as the highest richness area at low altitudes in the Brazilian pampa, the Pampas also occupied a large areas in Argentina, Uruguay, and southernmost Brazil.

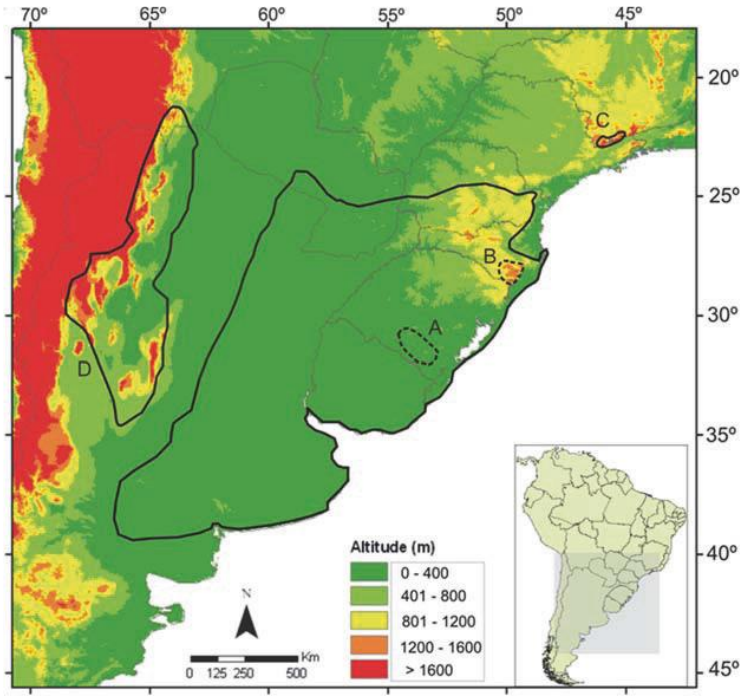


Figure 1.3 Map of geographic distribution of *Petunia* in solid lines. The two dotted lines are the centers of diversity: Serra do Sudeste in Rio Grande do Sul, Brazil (**A**), and Highlands of Serra Geral in Santa Catarina, Brazil (**B**), while the large disjunct regions: Serra da Mantiqueira, in Minas Gerais, Brazil (**C**), and the Sub-Andean region in Argentina and Bolivia (**D**). Figure taken from Gerats and Strommer (2009).

The Serra do Sudeste in southern Rio Grande do Sul has a low range of mountains with diverse edaphic conditions. *P. axillaris* and *Petunia integrifolia,* parental species of the hybrid petunia, are endemic in these areas (Ando *et al.* 2001). In Serra do Sudeste, three out of five *Petunia* species grow in this region are strict sympatric (*P. bajeensis, P. exserta,* and *P. secreta*).

The second region includes the Serra Geral borders in Santa Catarina state where *Petunia* species grow alongside with grasslands or with *Araucaria* moist forests. In this area, three out of four species are strict endemics (*P. bonjardinensis, P. reitzii and P. saxicola*) that restricted to the higher region of the Santa Catarina plateau. Two large disjunct regions of *Petunia* are reported: the first is Serra da Mantiqueira in Brazil, where only *P. mantiqueirensis* is registered as endemic species phylogenetically linked to the Brazilian highland group and the second is Sub-Andean region in Argentina and Bolivia (Figure 1.3C, 1.3D). The the Atlantic rainforest and savanna considered as geographical barrier covered most of the S˜ao Paulo state. The Sub-Andean area is separated by the Chaco, a drier region from the core *Petunia* distribution where two species are inhibited: *P. axillaris* subsp. *subandina,* and *P. occidentalis* (Fries 1911; Ando 1996; Tsukamoto & Kao 1998; Kokubun *et al.* 2006).

Fregonezi *et al.* (2012) reported that changes in climate, soil conditions and ecology probably played a significant role in *Petunia* speciation of lowland clade. Some species of *Petunia* are sympatric and associated with particular phytoecological areas while others, like *P. axillaris,* have widely distributed habitat within temperate South America. The role of ecological divergence is strongly accepted due to habitat changes that affected on population differentiation (Zheng & Ge 2010). Multiple ecological and environmental pressures, such as different pollinators and gene flux disruption among groups, might be affected in differentiation of subspecies morphology. The differences of floral characteristics and pollinator attraction are associated with each other, and probably drove speciation together with ecological effects in the Pampas area (Fregonezi *et al*., 2012).

In scientific research, the accessions/lines V26 and Mitchell of *P. hybrida* are very frequently used, having high transformation ability, in addition to *P. hybrida* W138 that has *dTPH1* transposon with high copy number and has been applied for transposon mutagenesis. These lines, as well as, genetic self-incompatibility, development, transposon activity, and integration with herbivores, pollinators and pathogens strongly support *Petunia* as a model plant (Bombarely *et al.* 2016; Vandenbussche *et al.* 2016).

Genomes of diploid *Petunia* species range between 1.30 to 1.57 pg 1C (corresponding to 1,300-1,570 Mb) (Mishiba *et al.* 2000). Compared with genomes of other Solanaceae, the genome of *Petunia* is larger than *Solanum tuberosum* (844 Mb) and *Solanum lycopersicum* (900 Mb), but

smaller than *Capsicum annuum* genome (3480 Mb) (Kim *et al.* 2014). Additionally, *Petunia* has important uses as an anti –microbial source, possesses slight anti-oxidation activity, and the leaves have insecticide properties (Gautam *et al.* 2012).

P. hybrida along with other species of *Chenopodium*, *Cucumis*, *Nicotiana*, *Phaseolus* and *Vigna* is a common host plant used in virology for virus amplification (Hull 2014). More than 150 plant viruses are reported to infect petunia using artificial inoculations (Engelmann & Hamacher 2008). The European and Mediterranean Plant Protection Organization (EPPO) requires testing of stock plants used in commercialized propagation schemes (OEPP/EPPO 2008) for nineteen viruses. *Petunia* viruses are known for their negative effect on its economic value such as *Tobacco mosaic virus, Tomato mosaic virus, Potato virus Y, Broad bean wilt I virus, Alfalfa mosaic virus, Cucumber mosaic virus, Petunia asteroid mosaic virus, Petunia ring spot virus*, *Petunia vein banding virus, Petunia flower mottle virus* and *Petunia vein clearing virus* (Lesemann 1996; Mavric *et al*. 1996; Cohen *et al*. 1999). Moreover, *Chilli leaf curl virus,* a new virus infecting *P. hybrida*, has been characterised in India for the first time by Nehra and Gaur (2015) while in Iran, Anabestani *et al.* (2017) found that *Beet curly top virus* could transmit through seeds of *P. hybrida*.

1.2 The karyotype of petunia chromosomes

Numbers and sizes of the somatic chromosomes as a complete set in each species are considered as a physical feature of the genome, presented as karyotypes (Stebbins & Dunn 1950). The karyotype includes some fundamental aspects like centromere positions, ratios of arms, as well as chromosome numbers and sizes, and presence of secondary constrictions at the Nucleolar Organization Region (NOR, defining the satellited chromosome), and sometimes supernumerary B chromosomes. Using paraffin-sectioned samples, many researchers studied the somatic metaphase karyotype of *P. hybrida* (Dermen 1931; Steere 1932; Malinowski 1935; Marthaler 1936; Wergin 1936; Levan 1937; Cooper 1946). From those names, however, Malinowski, Levan and Marthaler were the only successful workers to find out loci of kinetochore by camera-lucida drawings. Malinowski and Levan works show distinct nucleolar constrictions and positions of kinetochore except karyotypes that were appeared slightly different from Marthaler results. The employed species in the Malinowski paper was registered firstly as *P. violacea* and then later revealed that it was actually *P. hybrida* (Stout 1952). Takehisa (1964) reported that in all chromosomes, differentiation in chromosome thickness happens from late prophase to metaphase and this result was confirmed by comparing lengths

of chromosomes at both phases. Individual chromosomes were classified to metacentric (M) and submetacentric (SM) groups as well as given numbers from the biggest to the smallest as M1, SM1, SM2, M2, SM3, SM4 and M3. Malinowski (1935) and Marthaler (1936) pointed out that chromosome SM1 had a satellite at the short arm. Five of the seven pairs of petunia chromosomes have been recognized using staining protocols, but the only problem was with chromosomes V and VI that share arm ratio and relative length. Smith and Oud (1972) characterised differences between those two chromosomes (V and VI) based on fluorescence patterns using quinacrine fluorescence staining. Furthermore, chromosome I has been characterised based on its relative length, while chromosomes II and III appeared identical but the existence of the satellite in II was adequate to differentiate them from each other (Marthaler 1936; Bentzer *et al.* 1971) (Figure 1.4). Smith *et al.* (1973) applied an improved method using cellulase and pectinase with quinacrine staining to obtain chromosome sets of *P. hybrida* under fluorescence microscopy showing relative length, centromere index and fluorescence intensity. On the other hand, Dietrich *et al.* (1981) demonstrated that C-banding technique and pachytene analysis are not valuable methods for large scale karyotyping because of the dispersed pattern of the heterochromatin in petunia that is not restricted in particular positions unlike in tomato. Conia *et al.* (1987) applied flow cytometric analysis to produce a high metaphase index for *P. hybrida* chromosomes showing theoretical histograms and experimental flow karyotype. Fransz *et al.* (1996) applied fluorescent *in situ* hybridization (FISH) to characterise the whole chromosomes of some hybrid cultivars of petunia using probes of 18S rDNA and *chsA* gene.

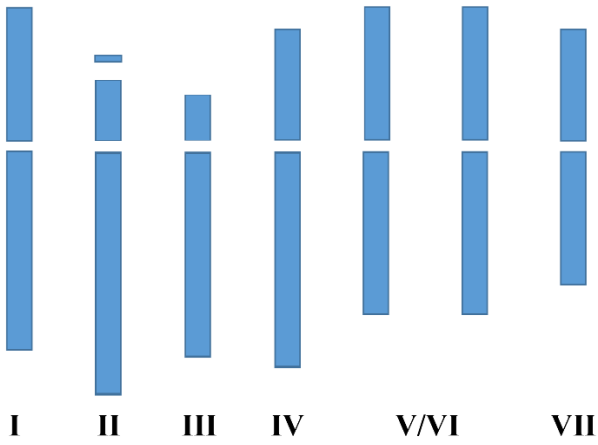


Figure 1.4 Idiogram of *P. hybrida* chromosomes according to (Marthaler 1936) showing the numbered seven chromosomes of petunia.

1.3 Repetitive DNA

Plant genomes contain large proportions of repetitive DNA that could reach up to 90-95%, while low copy sequences, coding regions, regulatory units are only a small portion of the genome (Heslop-Harrison 2000). The repetitive DNA sequences consist of highly heterogeneous sets of many thousands of super families, families and subfamilies with a variety of copy numbers, motif length and arrangements within the genome. It has been proven that changes in repetitive DNA parts happened rapidly in parallel with speciation in higher plants (Bennett & Leitch 2011). The existence of repetitive DNA sequences has different aspects from one to few sites inside the genome, whereas the widely dispersed motifs and tandem repeats throughout the DNA are making up 50-75% of the nucleus DNA (Flavell *et al.* 1974; Schmidt & Heslop-Harrison 1998; Heslop-Harrison & Schmidt 2001).

Repetitive sequences have often been referred to as 'junk DNA' (Schmidt & Heslop-Harrison 1998) as they are not transcriptionally active in general. Now, it is however clear that repetitive DNA sequences have impacts on the behaviour and structure of the genome and chromosomes, as well as chromosomes packaging and histone proteins modification in addition to gene expression, segregation and recombination (Martienssen 1998; Heslop-Harrison & Schwarzacher 2012).

It has been shown that the repetitive DNA has many benefits including the function (e.g. rDNA repeats), evolution of the genome and the structural roles of some sequences inside chromosomes like centromeres and telomeres (Heslop-Harrison & Schwarzacher 2012). The histone proteins modification is correlated with chromatin packaging or epigenetic fact, while some repetitive sequences might be transcribed to small RNAs which are included in chromatin or genome regulation and modification through evolution. This may lead to produce diversity, genome divergence and speciation as well as the modulation of gene expression can be controlled by losing and gaining of repeats (Schmidt & Heslop-Harrison 1998).

The reasons for maintaining and tolerating repetitive sequences within the genome, as well as the vast diversity and different types of repetitive DNA sequences (Figure 1.3) are not totally clear. However, there is an assumption that the importance of the repetitive DNA is correlated with the maintenance and stabilization of the chromosomes structure (Irick 1994; Vig 1994), or is linked with identification and chromosome segregation in meiosis and mitosis (Vershinin

et al. 1995; Kubis *et al.* 1998). For example, telomeres can be protected by telomere-associated repeats and involved with gene regulation in subtelomeric regions (Sýkorová *et al.* 2003).

Moreover, the repetitive DNA probably has a role in protecting coding DNA against shock during stress conditions (Pluhar *et al.* 2001), and it has been used for nuclear architecture study (Heslop-Harrison 2000) (see Figure 1.5).

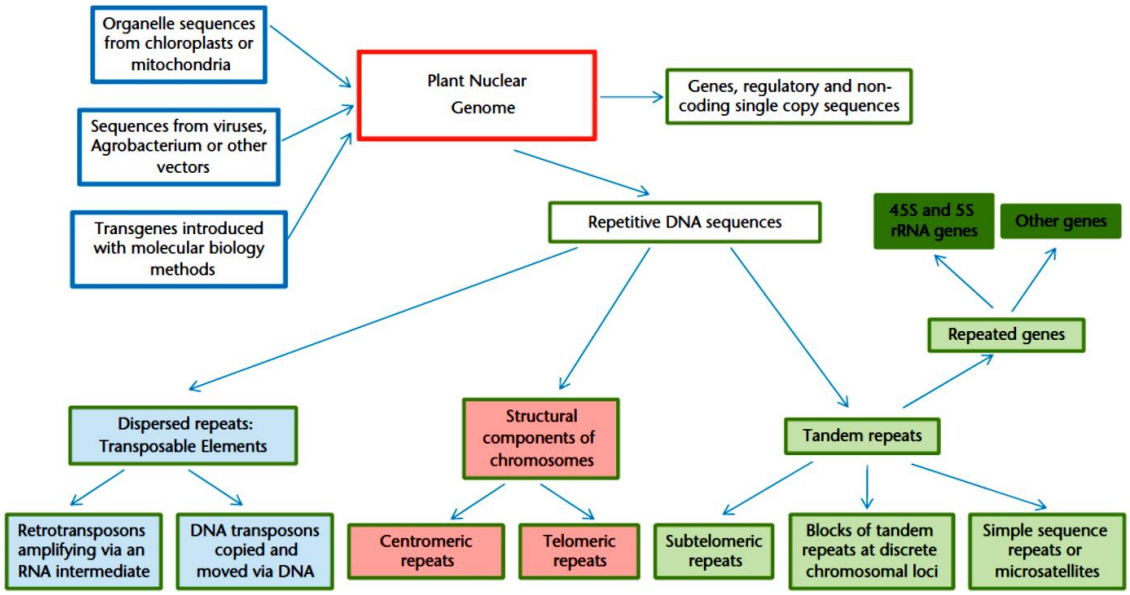


Figure 1.5 General image of plant DNA components in the nuclear genome of eukaryotes showing coding (Genes and regulatory sequences) and non-coding (repetitive DNA sequences include dispersed repeats, structural components, tandem repeats and repeated genes) divisions (Heslop-Harrison & Schmidt 2001).

1.3.1 Types of Repetitive DNA sequences

1.3.1.1 Tandemly repeated sequences (Microsatellite, Minisatellite and Satellite DNA)

Microsatellites consist of 2-6 bp nucleotide repeats within arrays up to 1 kb, and are also called simple sequence repeat (SSR). They are abundantly found in plants in coding and non-coding regions. Generally, tri-nucleotides are found in coding-regions and while they do not

cause frameshifts, large arrays are linked to malfunctioning genes and diseases (Morgante *et al.* 2002). In many plants, di- nucleotides are considered the main members of SSRs, and the common repeats are GA, CT, AT and TA (Tóth *et al.* 2000). These types of repeats evolve quickly and due to the variable length of arrays have been used successfully for DNA markers, fingerprinting species and accessions (Kubis *et al.* 1998).

Minisatellites are tandem repeats with a monomer size between 9 to about 40 bp, for example positioned in the pericentromeric regions of *Arabidopsis thaliana* depending on the *in situ* hybridization findings. In eukaryotic genomes, both micro/minisatellites have different distributions and possible functions (Vergnaud & Denoeud 2000). Interestingly, highly polymorphic minisatellites found hypervariable as well as too abundant within genomes as an efficient material to distinguish individuals (Weitzel *et al.* 1988).

Satellite sequences have larger monomer sizes of up to several kb and were initially discovered in Caesium density gradients for DNA isolation, to form distinct shoulders or satellites. They are the essential constituent of heterochromatin and found as large blocks of up to 1Mb arrays in pericentromeric and centromeric position, but also interstitially or in subtelomeric regions (telomere associated sequences, TAS) (Arney & Fisher 2004; Hall *et al.* 2004; Sharma & Raina 2005; Bloom 2007). Satellite DNAs are highly varying in nucleotide sequence, intricacy, genomic multiplicity. Monomers often are found to have high ratio of A+T nucleotides and while variable lengths in plant and animal genomes, the detected monomers have preferential lengths between 150 to 360 bp and significantly reflect requirements of wrapped DNA length around nucleosomes (Schmidt & Heslop-Harrison 1998; Henikoff *et al.* 2001). In eukaryotes genomes, satellite DNA content is considerably variable and can make up 50% of the whole DNA (Doolittle & Sapienza 1980; Cavalier-Smith 1985; Elder Jr & Turner 1995; Schmidt & Heslop-Harrison 1998; Gregory *et al.* 2006).

To detect and characterise satellite DNAs, digesting genomic DNA with restriction endonucleases and sequencing of short multimers or randomly cloned monomers is still the main technique despite the great progress of whole genome sequencing tools (Salih 2017). Tandemly repeated motifs assembly has been faced by serious restrictions: while the individual monomers might show some variation, these are generally too low for building contigs, and repeats are collapsed within an array, but also between different genome locations (Eichler *et al.* 2004; Rudd & Willard 2004). Most assemblies are therefore devoid of large satellite arrays. Although satellite DNAs are highly abundant within the heterochromatin, they are still

underrepresented in the processes of genome analysis, and the data available are not enough to conclude about their functional evidence and general organization (Nagaki *et al.* 2004; Hoskins *et al.* 2007). See more details of tandemly repeated sequences (Satellite DNAs) in chapter V.

1.4 Virus definition and taxonomy

Based on the International Committee on Taxonomy of Viruses (ICTV), viruses define as ‘‘an elementary biosystem that possesses some of the properties of living systems such as having a genome and being able to adapt to changing environments. However, viruses cannot capture and store free energy and they are not functionally active outside their host cells’’ (Hull, 2001; Hansen and Heslop-Harrison 2004). Also, the ICTV accepted the proposed definition of Van Regenmortel (1990) ‘’ A virus species is defined as a polythetic class of viruses that constitutes a replicating lineage and occupies a particular ecological niche’’(Murphy *et al.* 2012). These entities have genes with well identified properties of expression and replication in host tissues. Viruses were recognized till the 1990s based on their symptoms, physical and biological descriptions as well as host range, particle structure, replication and biochemical composition. Nowadays, viral genome nature, size and sequence have been accurately applied to categorize and classify viruses to different groups (Büchen-Osmond 2003; Hansen & Heslop-Harrison 2004a).

The earliest attempts to recognize viruses started to differentiate them from other detectable microbes through light microscope. Ivanowski (1892) and Beijerinck (1898) were the first scientists who discovered plant viruses (*Tobacco mosaic virus*) based on physiochemical properties using filterability to assess their small sizes (van Helvoort & Sankaran 2018). Most measurements made to classify viruses focused on the ability to do infections and diseases. The common features of pathogenic properties, organ tropisms, transmission and ecological characteristics were applied in virus classification at that early time such as viruses that share pathogenicity of causing mosaic symptoms (e.g., *Cauliflower mosaic virus*, *Alfalfa mosaic virus* and *Tobacco mosaic virus*). After 1930, the structure and composition of virus particles proposed by Bawden (1941, 1950) for the first time to identify viruses groups based on shared virion characteristics. On this basis, several groups of plant viruses constructed with filamentous or rod-shaped virions (Brandes & Wetter 1959). Massive numbers of new viruses discovered in 1950s and 1960s with rapidly growing data that prompted some committees and individuals to work independently in classification schemes. Therefore, the International Committee on Nomenclature of Viruses (ICNV) was established in 1966 at the International

Congress of Microbiology in Moscow as a result of this background that became later in 1973 the International Committee on Taxonomy of Viruses (ICTV). The ICTV developed universal scheme taking virion properties as a main criteria to divide families, subfamilies and genera (Murphy *et al*., 2012). In the 1990s, the progress of species taxon suggests that families and genera should be classified monothetically while species are best identified polythetically (Van Regenmortel 1990). Nowadays, existing viruses of economically important plants have been classified based on useful and usable taxonomic system to orders, families, subfamilies, genera and species. Virus families are named with suffix-*viridae* and viruses within same family have distinct morphology, genome structure and replication. Also, family members seem more stable and indicate phylogenetic independence or separation. Virus genera are named with suffix-*virus* and genus members share same characteristics and phylogenies, and have differences from members of other genera. The species taxon is considered the crucial hierarchical level and could be recognized by more than one property. In some cases, genomic, structural, physicochemical and serological properties could be used to differentiate species level, while in other examples, viruses have already named as species due to their very distinct properties (Murphy *et al*., 2012).

1.5 Transposable elements (TEs)

TEs are involved in the genome with high ability to transpose between different loci as well as to replicate themselves producing abundant copies within genomes (Craig 2002). They are widely distributed through all organisms so far analysed, and frequently make up a large portion of the genome (Feschotte & Pritham 2007; Pritham 2009). These elements can efficiently transmit between hosts through two styles, either by horizontal transfer between different species, or vertical inheritance over host generations (Schaack *et al.* 2010; Wallau *et al.* 2018). In genomes, TEs and endogenous viruses (see below) are considered main catalysts of innovation and variation with a more severe impact of their horizontal activity on eukaryotic evolution than expected before (Gilbert & Feschotte 2018). Most of interspersed repetitive sequences were created by activity of mobile genetic elements (MGEs) inside the nucleus and cells. These MGEs are classified into two classes:

Class I elements are retrotransposons or retroelements, several kb in length, that move via ‘copy-paste’ mechanism through an RNA intermediates.

Class II elements are DNA transposons that move as DNA copies via 'cut- paste' mechanism and often have in plants much lower copy numbers, and due to their much smaller size and genome proportion, than retrotransposons (Heslop-Harrison & Schmidt 2001).

MGEs represent DNA fragments that change their replication and movement within chromosomal positions of the same genome species, causing mutation and alteration in the genome (Jurka *et al.* 1992; Flavell *et al.* 1994). The elements show variable copy numbers between a few to millions per genome, and represent main and major component in the genome of eukaryotes (Schmidt 1999; Heslop-Harrison & Schmidt 2001). TEs have been considered as an ancient component of genomes because of their existence in most organisms (Kidwell 2002; Hua-Van *et al.* 2005). Many researchers suggested that these components in plants comprise >80-85% of the whole genomes sizes such as maize (Heslop-Harrison & Schmidt 2001; Schnable *et al.* 2009), wheat (Tenaillon *et al.* 2011) and Liliaceae (Vitte & Panaud 2005). Fungi, metazoans and yeasts do not have abundant TEs unlike plants (Daboussi & Capy 2003; Hua-Van *et al.* 2005; Kidwell 2005), while these elements are not found in 20% of prokaryotic genomes and several parasitic apicomplexa (Bringaud *et al.* 2006; Hua-Van *et al.* 2011).

1.5.1 Retrotransposons (Class I elements)

As a genomic element, the retrotransposons and retroviruses share a similar mode of propagation via transcription and translation, then package their transcripts to particles (Adams *et al.* 1987). 'Copy- paste' mechanism is used to transpose these elements through an RNA intermediate, and RNA polymerase II used to transcribe retrotransposons to mRNA, then convert to a complementary DNA (cDNA) and integrate into a new locus in the genome via an integrase (Wicker *et al.* 2007; López-Flores & Garrido-Ramos 2012; Lisch 2013). These elements were firstly characterised and identified based on *Drosophila* elements (Emori *et al.* 1985; Marlor *et al.* 1986). Retrotransposons are distributed with dispersal manner over chromosomes because of their life cycle (Heslop-Harrison *et al.* 1997). These elements go through duplicative transposition at the end of their cycle, increasing their numbers and expanding genome size (SanMiguel *et al.* 1996; Kumar & Bennetzen 1999; Slotkin & Martienssen 2007; López-Flores & Garrido-Ramos 2012). Retrotransposons can be divided according to the existence or absence of long terminal repeats (LTRs) into LTR and non- LTR-retrotransposons. In comparison to animal genomes, plants have higher amounts of LTR retrotransposons and considered the major retrotransposons order while animals have many non-LTR retrotransposons (Wicker *et al.* 2007; López-Flores & Garrido-Ramos 2012).

1.5.1.1 LTR- retrotransposons

LTR- retrotransposons with a full length of these elements up to 25kb, are present in plant genomes with varying proportion: in *Arabidopsis* (5%), rice (10%), sorghum (54,5%) and more than this percentage in maize (50-80%) (Sanmiguel & Bennetzen 1998; Kapitonov & Jurka 1999; Meyers *et al.* 2001; Neumann *et al.* 2003; Paterson *et al.* 2009). The LTRs have an internal domain that encodes the required proteins for retrotransposition (Schulman & Kalendar 2005). There are two major open reading frames (ORFs), the *gag* polyprotein ORF that encodes the essential proteins for virus like elements and genome integration, and secondly, the longer *pol* ORF that is more conserved than *gag*, is auto-operated and contains a polyprotein with aspartic proteinase (AP), reverse transcriptase (RT) that is considered as highly conserved domains in all retrotransposons types, RNase H (RH), and integrase (INT) (Suoniemi *et al.* 1998) (Figure 1.4). LTR- retrotransposons have been classified to five superfamilies according to their domain/ORF order, *Metaviridae* (Ty3-gypsy); *Pseudoviridae* (Ty1-copia); Retroviruses; Endogenous retroviruses (ERVs); and *Bel-Pao* (Kumar & Bennetzen 1999; Hansen & Heslop-Harrison 2004a; Wicker *et al.* 2007). These elements have been classified to three groups (*Gypsy*, *Copia* and *Bel-Pao*) by López-Flores and Garrido-Ramos (2012) based on the similarity of sequences and the encoded genes order. As reverse-transcribing viruses, The International Committee on Taxonomy of Viruses (ICTV) has classified these elements under new order *Ortervirales* to five families: *Caulimoviridae*, *Retroviridae*, *Hepadnaviridae*, *Metaviridae* and *Pseudoviridae*. Latterly, in 2018, the ICTV added a new family, *Belpaoviridae*, that was previously considered as a member of the *Metaviridae* family (Krupovic *et al.* 2018) (Figure 1.6).

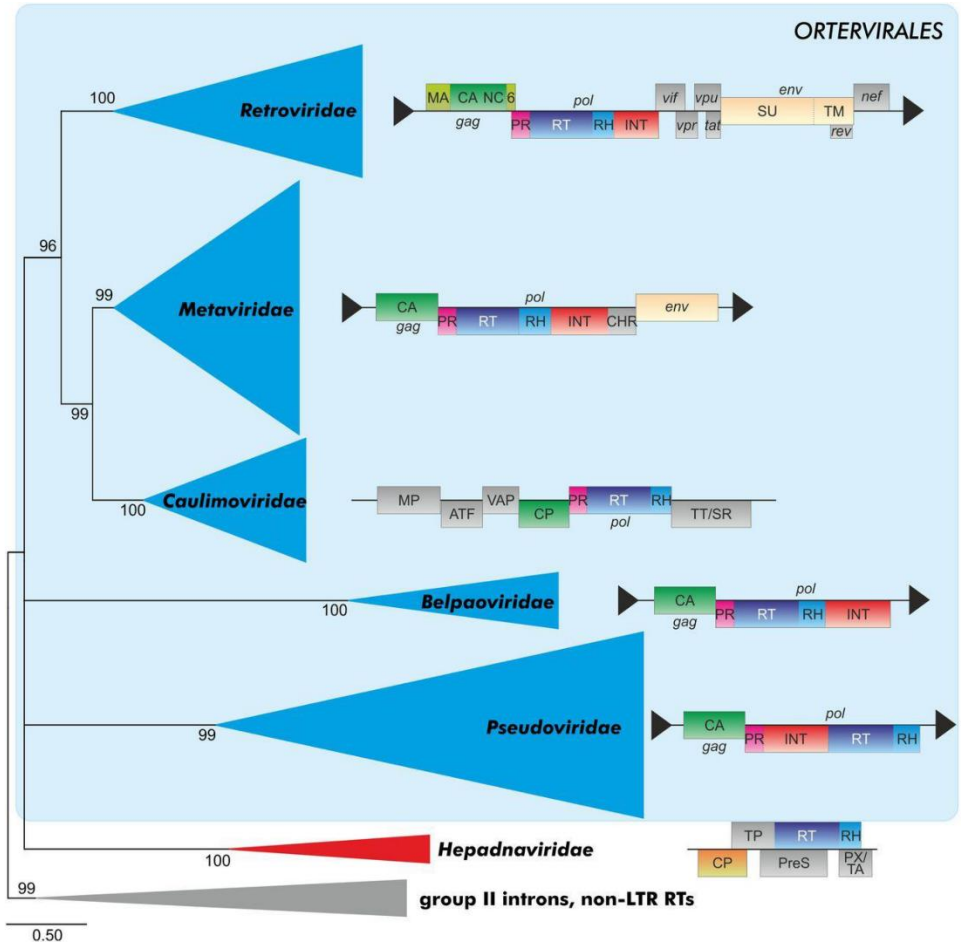


Figure 1.6 Phylogenetic tree of the six families of reverse-transcribing viruses (*Retroviridae*, *Metaviridae*, *Caulimoviridae*, *Belpaoviridae*, *Pseudoviridae* and *Hepadnaviridae*) showing genome arrangement of the protein domains and the close relationship between *Metaviridae* and *Caulimoviridae*. Figure taken from Krupovic *et al.* (2018).

Endogenous pararetroviruses are a relatively new discovery in plants. They include BSV (*Banana streak virus*), TVCV (*Tobacco vein clearing virus*) and PVCV (*Petunia vein clearing virus*). More detail about viruses and their nuclear integration are introduced in chapter III.

Ty3-gypsy (*Metaviridae*)

This superfamily is widely distributed in fungi, plants and animals, and generates 4-6 bp target site duplications (TSD) and is bound by LTRs of varying size. Protein domains PBS and PPT have encoded by the *gag-pol* genes across downstream and upstream of LTR. Ty3-gypsy has derived its name from the *Ty3* retrotransposons of *Saccharomyces cerevisiae* and *Drosophila*

melanogaster genomes. The integrase domain (INT) is localized downstream of RT and RH domains like in retroviruses that include the same ORF3 in the Ty3-gypsy (Marlor *et al.* 1986; Hansen *et al.* 1988; Krupovic *et al.* 2018).

Ty1-copia (*Pseudoviridae*)

The Ty1-copia has been found in plants as well as other living organisms (Manninen & Schulman 1993; White *et al.* 1994; Bennetzen 1996; Wicker *et al.* 2007), and named based on the *Ty1* retrotransposons of the genomes of *Saccharomyces cerevisiae* (Clare & Farabaugh 1985), and *Drosophila melanogaster* (Mount & Rubin 1985; Boeke & Corces 1989; Grandbastien *et al.* 1989). The RT domain is located downstream of the INT and both are flanked by LTRs while the PBS and PPT have been located towards downstream and upstream of LTRs. In plants, this group shows less divergence than fungi and insects, and have variable sequence heterogeneity among plant species (Flavell *et al.* 1992). Additionally, significant differences have been found between *Gypsy* and *Copia* with the three proteins domains (INT, RT, and RH), as *Gypsy* has the INT in the downstream of RT and RH while in *Copia* the INT is positioned at the upstream of RT and RH (Hansen & Heslop-Harrison 2004a; Krupovic *et al.* 2018).

1.5.1.2 Non LTR- retrotransposons

The non LTR-retrotransposons or retroposons have a very short LTR at the terminal region and transcribe from an internal promoter (Slotkin & Martienssen 2007). They have two sub-divisions, the first is the long interspersed nuclear elements (LINEs) and the second is the short interspersed nuclear elements (SINEs) based on the size and encoded domains. Plants have both types, but generally at lower levels (Kubis *et al.* 1998), while in animals, SINEs are abundantly found (Schmidt 1999; Jurka *et al.* 2007).

1.5.2 DNA transposons (Class II elements)

DNA transposons are transposons via a ‘cut-paste’ mechanism and do not involve an RNA intermediate. Also, these elements excise from some loci of chromosomes and then reintegrate within a new position of the genome by transposase enzyme. This class includes superfamilies of *CACTA*, *Mutator*-like element (MULE) and hAT (*hobo*, *Activator* and *Tam3*) (Lisch 2013). Another classification has put three subclasses based on DNA strand numbers that are excised through transposition, subclass I that works with ‘cut-paste’ mechanism, subclass II that works

with rolling circle approach (*Helitrons*), and subclass III with self-synthesizing DNA transposons (*Polintons*) (Feschotte & Pritham 2007; Kapitonov & Jurka 2008; Bao *et al.* 2009).

1.5.3 Transposable elements in petunia

Petunia genome has two major groups as viral (EPRVs), and non-viral (retrotransposons and DNA transposons) retroelements that play a crucial role in biodiversity of *Petunia* species (Richert-Pöggeler & Schwarzacher 2009). Long terminal repeat (LTR) retroelements, as a type of repetitive DNA, are found in the *Petunia* genome with comparatively lower ratio than in other Solanaceae with higher percentage of DNA transposons than retrotransposons (Gerats 2009; Bombarely *et al.* 2016). In *Petunia*, the identification of conserved domains of gag-*pol* regions in LTR-retrotransposons (Richert-Pöggeler *et al.* 2003) has revealed homology to similar components within different plant families (Richert-Pöggeler & Schwarzacher 2009). Although, reports of the *Copia* and *Gypsy* superfamilies (LTR retroelements) are still limited, Matsubara *et al.* (2005) studied the common features between *rTph1* the novel transposable element and *Copia* elements. Richert-Pöggeler and Schwarzacher (2009) had earlier reported relationships of *Gypsy* and *Copia* reverse transcriptase domains from some species of *Petunia*. Furthermore, Kriedt *et al.* (2014) discussed the relationship between the RNase H-3'LTR region of eight of *Copia* families in wild *Petunia* species.

1.6 Fluorescence *in situ* hybridization

The cytogenetic tool, is a technique in which widely used for detecting and proving the existence, abundance and location of genome sequences through hybridization between single stranded chromosomal DNA and a specific probe of labelled DNA(Schwarzacher & Heslop-Harrison 2000). Separately, the first description of *in situ* hybridization (ISH) was provided by Gall and Pardue (1969), and John *et al.* (1969), as initially, this technique used radioactive and colorimetric materials with some problems in radioactivity that could be released with low spatial resolution. Later, and since the mid-1980s, fluorophores have been used for detecting the hybridization sites instead of radioactive materials with more safely, saving time and high quality products. In addition, it is more flexible than before as we could mix more than one probe in multicolour tests (Salvo-Garrido *et al.* 2001; Schwarzacher 2003).

The metaphase and meiotic chromosomes, DNA fibres, nuclei, and tissues can be used as a target DNA (De Jong *et al.* 1999; Van Stedum & King 2002). Single and double stranded DNA,

RNA and oligonucleotides can be applied as a DNA probe (Nouri-Aria 2008). Suitable length of probe is between 100 and 300 bp (Salvo-Garrido *et al.* 2001), and different methods can be used for labelling probes like nick translation, PCR labelling and random primer labelling (Schwarzacher & Heslop-Harrison 2000). These methods can be used in two ways of labelling, the direct one in which the fluorophores are directly bound to the probe by using fluorophore conjugated nucleotides. The second way is indirect for detecting the hapten that has no fluorescence and integrated within the DNA probe, and then detected by a fluorophore-tagged antibody against the hapten (Volpi & Bridger 2008).

The biotin and digoxigenin are most widely used in indirect labelling methods by utilizing the high avidin (or streptavidin) affinity to biotin more than anti-biotin, and DAPI applied as a blue dye to counterstain the chromosomes (Sharma & Sharma 2001). The interaction of DNA: DNA in FISH is normally implemented in chromosomes preparations. In order to allow probe access and reduce image background, it is very important to make chromosomes free out the cells and spread on a glass slide. As well as, the inhibition of RNA and proteins should be done before the hybridization pre-treatments. After the hybridization process, the post-hybridization wash should be performed for removing free and weakly bound probes (Schwarzacher & Heslop-Harrison 2000).

Repetitive sequences, dispersed motifs, and (to some extent, although not routinely) single copy genes can be shown and localized on chromosomes with FISH (Bang *et al.* 1997). FISH is applied also to explore the phylogenetic relationships, chromosome identification, in addition to study evolutionary chromosome rearrangements, DNA mapping and genome organization (Heslop-Harrison 2000; Schwarzacher 2003; Contento *et al.* 2005).

1.7 Bioinformatic techniques

1.7.1 Whole genome sequencing data

NGS (Next Generation Sequencing) includes multiple high-throughput sequencing (HTS) techniques that generate huge numbers of reads from multiplexed specimens in a one run. A range of techniques have been developed extremely with time as this technology started with Sanger sequencing system by semi-automated tools as a first generation method for only a single strand DNA. Later, after twenty years, this approach has developed to sequence around

1000 bp per run (Shendure & Ji 2008). Thereafter, NGS platforms have been developed significantly with high capability to sequence the whole genome using cyclic-array technology (Hutchison III 2007; Mardis 2008; Metzker 2010; Van Dijk *et al.* 2014). NGS has two approaches, short reads sequencing such as Illumina and Qiagen GeneReader and long reads like PacBio and Oxford Nanopore (Goodwin *et al.* 2016).

Although, the whole genome sequencing is very efficient in genome coverage, only partial genome assembly can be produced from short read approach, due to the impact of different repetitive sequences classes. Single copy sequence and repetitive sequences can be assembled accurately only in the case of shorter repeats than the read length (Ricker *et al.* 2012). Bombarely *et al.* (2016) performed a mixed *de novo* assembly for *P. axillaris* subsp *axillaris* N by hybridizing short reads (including mate pairs, short reads separated by approximately known distances) from Illumina and long reads from PacBio, while only short reads have been *de novo* assembled exclusively for *P. integrifolia* subsp *inflata* S6. High-quality assemblies represent sizes of 1.26 Gb and 1.29 Gb for *P. axillaris* subsp *axillaris* N and *P. integrifolia* subsp *inflata* S6 respectively.
Illumina reads have been remapped to the assemblies, and the degree of heterozygosity has been estimated by single nucleotide polymorphism (SNP). The unassembled fractions of the genome have been rated with 140 Mb for *P. axillaris* subsp *axillaris* N and 110 Mb for *P. integrifolia subsp inflata* S6, because of the major impact of repetitive elements (Bombarely *et al.* 2016).

1.7.2 *k*-mer counting

Repetitive sequences contents in a genome can be analysed by using the *k*-mer frequency (Bergman & Quesneville 2007; Marçais & Kingsford 2011). *K-mer* means length *k* sequence included in the analysed dataset, for instance, the sequence AAGAG is a 5-mer and it is only one of the 5-mers positioned in the sequence AAGAGAAGAG. We can explore these sequences very repeatedly in the genome by counting all *k*-mers. Interestingly, *k*-mer counting can be useful tool for estimation of repeat libraries completeness and explore further sequences not found in the libraries (Krassovsky & Henikoff 2014). *K-mer* analysis has been used to count the frequency of DNA sequences of length *k* from raw reads data. It is an appropriate tool for measuring genome sizes and correcting sequence errors (Pevzner *et al.* 2001; Kelley *et al.* 2010), using some tools such as Jellyfish (Marçais & Kingsford 2011), and Tallymer

(Kurtz *et al.* 2008). This method is considered unbiased tool for counting repetitive sequences due to its independence of genome assembly (Bergman & Quesneville 2007; Marçais & Kingsford 2011).

K-mer was applied for identifying highly repeated structures from unassembled genome sequences and the correlation between these sequences and the centromeric regions of several mammalian genomes (Alkan *et al.* 2011), and Williams *et al.* (2013) counted the repeated DNA sequences in bacteria using this tool. In *Drosophila melanogaster*, *k*-mer frequencies were used for counting the repetitive sequences, identifying known transposons and short repeats (Krassovsky & Henikoff 2014). Recently, various lengths of *Taraxacum* microspecies motifs have been analysed using frequency analysis of all possible sequences, evaluating different lengths and complementing the graph-based outcomes (Salih *et al.* 2017). Using NGS data from the sheep genome, major classes of dispersed, tandemly repeated elements and endogenous retroviruses related repetitive sequences were identified by frequency analysis of short motifs (Mustafa *et al.* 2018).

1.7.3 Graph-based clustering of raw read sequences

1.7.3.1 RepeatExplorer

Sequences represented in multiple reads can be clustered using graph-based approaches. RepeatExplorer is a group of accessible software programs for identification of repetitive DNA sequences. *De novo* repeat characterisation could be achieved in the computational pipeline using the algorithm of a graph-based clustering without any demand for known reference databases. The main input of this pipeline is millions of short reads from next-generation sequencing that are random and non-selective. Phylogenetic relationships, comparative analysis and repeat classification of different types of retroelements result from the tools in this pipeline (Novák *et al.* 2010; Novák *et al.* 2013) (see Figure 1.7; Appendices 3.1 and 3.2).

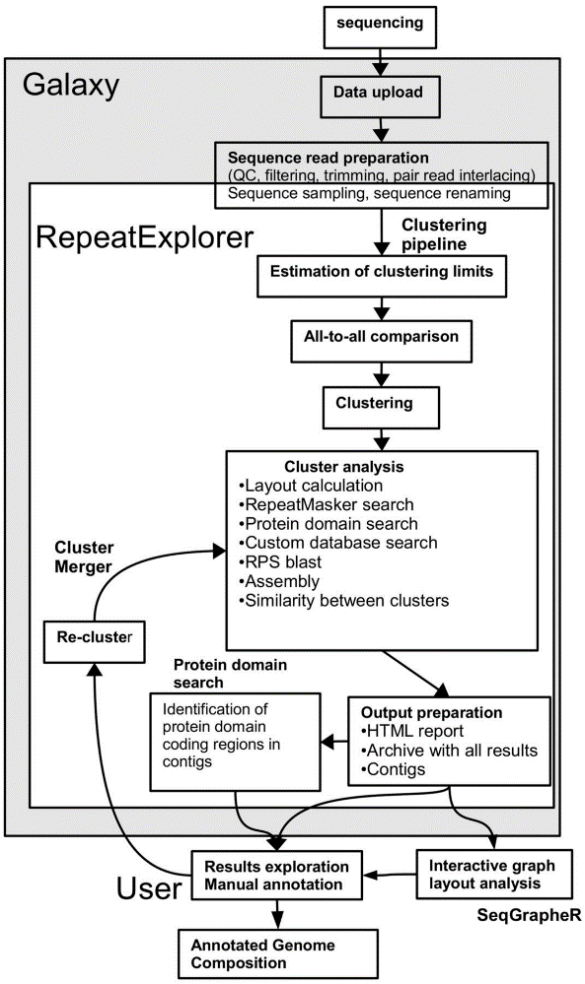


Figure 1.7 Graph-based clustering of repetitive raw reads using the RepeatExplorer pipeline. Figure taken from Novák *et al*. (2013). The pipeline runs on Galaxy, an open source, web-based platform.

In comparative study between two genera of the South American killifish (*Austrolebias charrua* and *Cynopoecilus melanotaenia*), RepeatExplorer has been used for providing the annotated repeated sequences and their repetitive sections in the two genomes (García *et al.* 2015). Further, genomes of *Ficus carica*, *Morus notabilis* and *Malus domestica* have comparatively analysed through RepeatExplorer using Illumina reads (Usai *et al.* 2017). In the Solanaceae family, transposable elements of twenty eight species of the *Physaleae* and *Solanaea* tribes have been studied by RepeatExplorer, identifying repetitive components in all plants (Mendieta 2015). As well as, RepeatExplorer algorithm has identified simple and low complexity repeats in petunia in addition to mixed repeat family clusters integrating retroelements (Bombarely *et al.* 2016).

1.7.3.2 TAREAN

Tandem repeat analyzer (TAREAN), is a novel pipeline that effectively detects satellite repeats in the unassembled short reads. The graph-based clustering has been employed to characterise reads types of repetitive elements. Satellite repeats could be putatively detected by the existent circular structures in the graph-based clusters. Repeat monomers from the most frequent *k*-mers are reconstructed through destructing read sequences from their clusters. TAREAN has been efficiently examined through low-pass genome reads of various plant species (see Figure 1.8; Appendix 5.1).

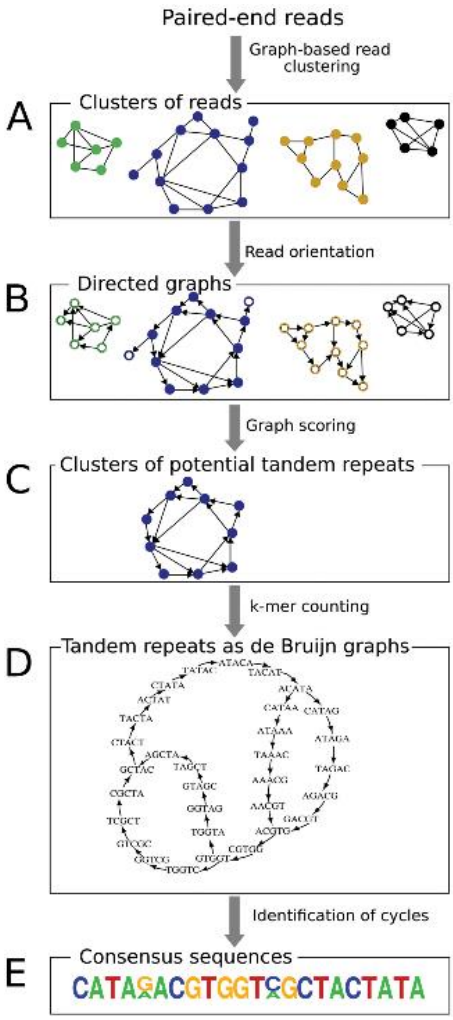


Figure 1.8 Identification of candidate tandem repeat sequences using TAREAN. Figure taken from Novák *et al*. (2017).

The results from graph-based or *k*-mer identification of satellite repeats reveal the presence and exact genomic abundance of repetitive motifs, but give no information about their chromosomal distribution. In some cases, the repeats can be identified in whole genome

sequence assemblies (despite the collapse in number of copies during assembly), but *in situ* hybridization (FISH) has probed essential to give detailed information about locations, number of sites and relative abundance between sites. An example of such characterisation was given in *Vicia faba* where three repeats were detected and their loci identified on chromosomes (Novák *et al.* 2017).

1.8 Aims and objectives

Aims: The main aim of this study is to determine the interaction, chromosomal location of tandemly repeated sequences, PVCV and other endogenous viruses in *Petunia* species, to find out more about episomal, *de novo* integrated and induced infections, study expression patterns, and any differences between vein clearing and spot symptoms of PVCV.

Objectives:

1. Organization of PVCV and other integrated viral sequences in *Petunia* species.

To explore sequences of PVCV, florendoviruses and other endogenous viruses in *Petunia* genomes, their location within petunia chromosomes, viral sequences activity and expression, the relationship and interaction of these sequences (chapter III).

2. Differences of PVCV symptoms and infections within petunia tissues.

To associate distinct symptom expression patterns as well as different modes of PVCV transmission (horizontal or vertical) with virus particle concentration as well as with changes in the cell ultrastructure of infected cells using transmission electron microscopy (TEM) together with immunogold labelling (chapter IV).

3. Identification of tandem repeats in petunia.

To reveal and characterise all tandem repeats that highly abundant motifs in petunia DNA, their organization within chromosome sequence assemblies, chromosomal location, and diversity among petunia species (chapter V).

Chapter II. Materials and methods

2.1 Plant material and cultivation

The parental species, *P. axillaris* subsp *axillaris* N and *P. integrifolia* subsp *inflata S6* in addition to three cultivars of *P. hybrida* (Rdc, V26 and W138) and *P. axillaris* subsp *parodii* S7 were used in this study. Seeds of the three wild *Petunia* species have been provided by Cris Kuhlemeier, University of Bern, Switzerland. *P. hybrida* seeds were obtained from Ronald Koes, University of Amsterdam (W138), NL Chrestensen, Erfurt, Germany (Rdc, aka "Himmelsröschen") and John Innes Institute, Norwich (V26) (Table 2.1; Figure 1.1). The seeds were placed in small pots and covered lightly with a peat moss (Levington advance), and irrigated regularly under greenhouse conditions (25 ˚C temperature and 16 hr of daylight) at the Department of Genetics and Genome Biology, University of Leicester (UK).

Table 2.1 List of suppliers for seeds from wild and hybrid petunia plants used in the studies.

No.	Species	Cultivar	Source
1.	*P. axillaris* subsp. *axillaris*	N	Cris Kuhlemeier, University of Bern, Switzerland
2.	*P. integrifolia* subsp *inflata*	S6	Cris Kuhlemeier, University of Bern, Switzerland
3.	*P. hybrida"Himmelsröschen"* or Rdc (*Rose du ciel*)	Rdc	NL Chtestensen, Erfurt, Germany
4.	*P. hybrida*	V26	John Innes Institute, Norwich
5.	*P. hybrida*	W138	Ronald Koes, University of Amsterdam

6.	*P. axillaris* subsp *parodii*	S7	Cris Kuhlemeier, University of Bern, Switzerland

2.2 Standard solutions and media

Table 2.2 Standard solutions used in experiments

Experiments	Solutions	Constitutions
DNA extraction	CTAB buffer	2% (w/v) cetyltrimethylammonium bromide, 100mM Tris–HCL, 1.4M NaCL, 20mM EDTA. (pH 7.5-8.0)
	10x TE buffer	100mM Tris (tris-hydroxymethylamino-methane)-HCl, 10mM EDTA (ethylene-diamine-tetra-acetic acid. (pH 8)
	DNA wash buffer	76 % ethanol, 10mM ammonium acetate.
	DNA extraction buffer	50 mM of Tris-HCl, pH 8.0, 25 mM EDTA, 400 mM NaCl.
Gel electrophoresis	6x Gel loading buffer	60% (v/v) glycerol; 0.25% (w/v) bromophenol blue; 0.25% (w/v) xylene cyanol FF; diluted to 1x in 50% (v/v) glycerol.
	50x TAE buffer (tris acetate-EDTA)	2M Tris-HCl; 50 mM EDTA (ethylenediaminetetraacetic acid; pH 8) 5.71% (v/v) glacial acetic acid, diluted to 1x in deionised H2O.
	Ethidium Bromide	10 mg/ml (1 g Ethidium bromide dissolved in 100 ml of sterile distilled water and stored at 4°C).
Cloning	Ampicillin	10 mg/ml (10 mg of ampicillin dissolved in 1 ml of distilled water and stored at -20°C).

	SOB medium (super optimal broth)	20g of tryptone, 5g yeast extract 0.5g NaCl, 10 ml 250 mM KCl. The final volume was 1000ml with sterile distilled water. (pH7.0)
	LB (lysogeny broth)	Agar plates: 2.5% LB broth (Melford), 1.5% agar (For Medium), 100 µg/ml ampicillin, 80 µg/ml x-gal, 0.5 mM IPTG (**i**so**p**ropyl β-D-1-**t**hio**g**alactopyranoside. (pH 7.2)
	LB medium (luria-bertani)	12.5g of LB broth (10g tryptone, 5g yeast extract, 10g NaCl). The final volume was 500ml with sterile distilled water. (pH 7.0)
	LB medium agar	25g of LB broth (10g tryptone, 5 g yeast extract, 10g NaCl). The final volume was 1000ml with sterile distilled water and 15 g of Bacto-Agar (DIFCO).
	IPTG	200 mM (476 mg/ml isopropyl-B-D-thiogalacto-pyronoside dissolved in 10 ml distilled water, filter sterilized and stored at -20°C).
	X-gal	40mg/ml (1g of 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside with 25ml of dimethylformamide, filter sterilized and stored at -20°C).
	2x rapid ligation buffer	60mM Tris-HCl (pH 7.8), 20mM MgCl2, 20mM DTT, 2mM ATP and 10% PEG.
Chromosome preparations	10x Enzyme buffer	100 mM citric acid, 100 mM tri-sodium citrate, stored at 4°C. (pH 4.6)
	1x Enzyme solution	3% (w/v) pectinase (Sigma), 0.2 % (w/v) cellulase [1.8% (w/v), Onozuka RS] cellulase (Calbiochem) in 1x enzyme buffer, stored at -20°C.
	8-hydroxyquinoline	0.002M (Dissolving 0.29 g of S-hydroxyquinoline in 1000 ml of ddH2O, stored in the dark at 4°C).

FISH	SDS (sodiumdodecyl sulfate)	20% (2g Sodium dodecyl sulfate (SDS) with 8ml water).
	20x SSC (Saline Sodium Citrate)	0.3M NaCl, 0.03M Sodium citrate. (pH 7.0)
	Paraformaldehyde	4% paraformaldehyde (Fisher Scientific) dissolved in distilled water. The final volume was 100 ml. (pH 7.0)
	50% Dextran sulphate	50 gm dextran sulfate with 100 ml distilled water, filter sterilized and stored at -20°C.
	Salmon sperm DNA	1mg/ml of sheared salmon sperm DNA.
	Blocking solution	5% BSA in 4xSSC with 0.2% Tween 20.
	Detection buffer	4xSSC, 0.2% (v/v) Tween 20.
	0.5 M EDTA	186.1g disodium ethylenediamine tetraacetate $_2$H**2**O into 800ml of distilled water. Adjust pH to 8.0 with NaOH. The final volume was 1 L.
	DAPI solution	5g of DAPI (4′, 6-diamidino-2-phenylindole) dissolved in molecular grade water. The final volume was 50ml (stock of 100µg/ml diluted in water. Final concentration of 4µg/ml was diluted with McIlvaine's buffer and stored at -20°C).
De novo integration	RLT buffer (RNeasy lysis buffer)	RLT buffer contains a high concentration of guanidine isothiocycanate that supports the binding of RNA to the silica membrane.

	RW1buffer (RNeasy washing buffer)	RW1 buffer contains a guanidine salt and ethanol. This buffer is used as a stringent washing buffer that efficiently removes biomolecules that are non-specifically bound to the silica membrane like carbohydrates, proteins and fatty acids.
	RDD buffer (RNase-free DNase)	RDD buffer provides efficient on column digestion of DNA and ensures that the RNA remains bound to the column.
	RPE buffer	RPE buffer is a mild washing buffer used to remove traces of salts on the column.
	PB buffer	PB buffer contains a high concentration of guanidine hydrochloride and isopropanol.
	PE buffer	PE buffer is a wash buffer in DNA cleanup procedures.
Electron microscopy	Elisa wash buffer	6.06 g Tris base, 8.2 g NaCl and 6.0 ml 6 M HCl dissolved in 1 L of distilled water with about 7.2 pH.
	Epon812	Glycerol based aliphatic epoxy resin that is relatively low viscosity: Epon 812 = 150-220 cps at 25 °C.
	DDSA	1-Dodecenylsuccinic anhydride is a hardener contains mixture of isomers.
	MNA	Methyl nadic anhydride (Methyl-5- norbornene-2,3-dicarboxylic Anhydride) is an epoxy curing harder.
	DMP30	2,4,6-Tris-(dimethylaminomethyl) phenol is a curing catalyst for epoxy resins.

2.3 Methods

2.3.1 Genomic DNA extraction

Genomic DNA was prepared from young leaves of petunia using the CTAB method (Doyle & Doyle 1990) with some modifications. One to two grams of asymptomatic young leaves were collected from each individual species, and washed with distilled water and then wrapped with aluminium foil followed by shock freezing in liquid nitrogen and immediately ground up using a pestle and mortar to avoid enzymatic degradation. About half a spatula of polyvinylpyrrolidone (Sigma) was added to the material before transferring it into 10 ml of preheated CTAB buffer (contains 50 µl of B-mercaptoethanol) in 50 ml tube and incubated at 60 °C for 30 min. Absolute chloroform: isoamyl alcohol (24:1) was added with equal volume to each tube and blended via inverting many times for 3 min, then centrifuged at 2040 g at room temperature for 10 min. The supernatant was transferred gently from the top of the tube to another new one using 1ml tip with cut end, the steps of washing and centrifugation were repeated another time. After that, pre-chilled isopropanol (0.6x volume) was added to the supernatant for DNA precipitation, then the material was mixed carefully by inverting several times, and kept for 10 min on ice. The precipitated DNA was spun down at 735 g for 3 min, and then air dried and washed with 5 ml of DNA wash buffer for 20 min. After this step, the DNA was re-suspended in 1ml of 1x TE buffer at RT overnight after air drying. It was incubated for 1 hr at 37 °C with 2 µl of 10 µg /µl RNase A (Bioline) to remove the RNA and re-precipitated with 1x sodium acetate 3M (pH 6.8) and 2x absolute ethanol. The DNA was spun down at 735 g for 3 min and then re-suspended in 500 µl of 1x TE buffer at RT overnight, the prepared DNA was stored at -20 °C.

2.3.2 DNA quantification

2.3.2.1 Gel electrophoresis

For conformation check and detection of denaturation, 5 µl of the extracted total genomic DNA was separated on agarose gel (1% w/v in 1xTAE buffer with ethidium bromide 1µg/µl) with a size and amount reference (5 µl of Bioline 1kb hyper ladder). The gel run was set up for 60 min at 75 V. Gel documentation system (Gen flash) was used to capture the gel image.

2.3.2.2 Spectrophotometry

Genomic DNA was assessed for purity, quality and amount using a Nanodrop ND-1000 spectrophotometer at 260/280 nm wavelength using 1 µl of each single DNA sample. Quality was considered as a high when the OD 260/280 nm value was about 1.8-2.2.

2.3.3 PCR amplification

Genomic DNA, dNTPs and primers were diluted with sterilized water, and the final concentration of the DNA was 30-50 ng/µl to use as a template. The master mix components (Table 2.3) were set up to amplify some tandem repeats (see chapter V) and EPRVs (Table 2.4). PCR reactions used a gradient PCR machine (Biometra) with a program that consisted of 3 min for initial denaturation at 95 °C and then 35 cycles of denaturation (95 °C for 30 sec), annealing (50.2-69.3 °C for 30 sec) and primer extension (72 °C for 1 min). The last cycle was 1 min at 72 °C as a final extension. After these steps, the temperature was set up at 16 °C as indefinite time (Table 2.5). PVCV primers (Table 3.1) were amplified as (Table 3.2), and PCR products were checked by gel electrophoresis (see below).

Table 2.3 Components and quantities of PCR Master Mix.

Components	**Amounts(µl)**
Molecular biology grade water	13.4
10x KAPA Taq Buffer	2
$MgCl_2$(25mM)	1.3
10mM dNTP Mix	1.2
10 µL Forward Primer	0.5
10 µL Reverse Primer	0.5
KAPA Taq DNA Polymerase	0.1
Template DNA	1
Total	20 µl

2.3.4 Purification of DNA fragments from agarose gel

Selected DNA fragments were cut from the gel by sterilized forceps and scalpels under a UV trans-illuminator (UVP). The gel pieces were placed into 1.5 ml Eppendorf tubes and then purified and washed from agarose parts according to the instructions in the clean-up kit (Nucleo Spin Extract II Clean up kit Macherey-Nagel Company).

2.3.5 Cloning of PCR products

The pGEM-T Easy Vector System I kit (Promega) was used for cloning of some purified PCR products.

2.3.5.1 Ligation

The ligation reactions were performed in a 300 µl Eppendorf tube. The reactions consisted of 2x rapid ligation buffer (7 µl), pGEM-T Easy Vector (0.9 µl), DNA sample (5.4 µl), T4 DNA ligase (1.2 µl), and sterile water (0.5 µl). The final volume was 15 µl as (Table 2.4), and mixed gently by flicking, then incubated at RT for 1 hr and then overnight at 4°C (see Table 2.2).

Table 2.4 Components of the ligation reaction.

Component	**Volume (µl)**	**Final concentration**
2x rapid ligation buffer	7	1x
pGEM-T Easy Vector	0.9	1-10 ng/µl
Purified PCR products	5.4	1-10 ng/µl
T4 DNA ligase	1.2	6U/µl
Molecular grade sterile water	0.5	NA
Total	15	

2.3.5.2 Transformation

Plasmid vectors containing the PCR product were transformed into transformation competent cells (*Escherichia coli*). After growth in selective media, culture was plated out on LB (Lysogeny Broth) agar plates with selective markers (pre-prepared media contains 100 µg/ml ampicillin, 500 µg/ml (IPTG) - isopropyl-β-A-thiogalactopyranoside and 40 µg/ml (X-Gal) - 5-bromo-4-chloro-3-indolyl-β-D-galactosidase) (see Table 2.2).

2.3.5.3 Screening and isolation of recombinant clones

Recombinant colonies with inserted sequences in the plasmid were white, while non-recombinant cells appeared as blue colonies. White colonies were picked by a sterile toothpick and transferred into 5ml LB medium tube and then incubated overnight at 37°C in a shaker (230 g) (see Table 2.2).

2.3.5.4 Colony PCR

Colony PCR was used to confirm recombinant clones and the size of insert using the universal M13 sequencing primers [forward (GTA AAA CGA CGG CCA GT) and reverse (GGA AAC AGC TAT GAC CAT G)]. The master mix components were prepared as (Table 2.5) in 1.5 ml Eppendorf tubes. 1-2 µl bacterial culture was used in addition to 15 µl of master mix components in each clean 0.2 ml PCR tube. After flicking and spinning down the mixture tube, the PCR machine was set up according to the PCR protocol in (Table 2.6). PCR products and fragment lengths were analysed in horizontal agarose gel electrophoresis.

Table 2.5 Materials for PCR reaction of cloned inserts products.

Components	**Amounts (µl)**
Molecular grade Water	9.65
10x KAPA Taq Buffer	1.5
MgCl2(25mM)	1
10Mm dNTP Mix	1.2
10 µL Forward Primer-M13	0.75
10 µL Reverse Primer-M13	0.75
KAPA Taq DNA Polymerase	0.15
Total	15

Table 2.6 Colony PCR protocol of cloned PCR products.

Steps		Temperature °C	Duration Minutes : seconds	Cycles
1	Initial denaturation	95	5:00	1
2	Denaturation	95	1:00	35
3	Annealing	55	00:30	
4	Extension	72	1:00	
5	Final extension	72	10:00	1
6	Hold	16	pause	1

2.3.5.5 Plasmid DNA isolation and purification

Plasmid DNA was purified using Machery-Nagel Kit following the Manufacturer's instructions. Recombinant colonies culture (1400 µl) was added into 1.5 ml Eppendorf tube and centrifuged directly at 9875 g for 30 sec. Then, the supernatant was gently removed and only pellets left at the bottom of the tube. The cell pellet was re-suspended totally by adding buffer A1 and pipetting up and down. In order to continue with cell lysis process, buffer A2 and buffer A3 were separately and respectively added and gently mixed by inverting the tube many times. The lysate was clarified by centrifugation at 9875 g for 5 min. The supernatant was loaded in a NucleoSpin Plasmid/Plasmid (no lid) column and centrifuged at 9875 g for 1 min. Next, the silica membrane was washed by adding buffer AW and buffer A4 separately and respectively with centrifugation at 9875 g for 1 min each. Finally, the silica membrane was dried by centrifugation and plasmid DNA was collected by adding 50 µl elution buffer and centrifuged at 9875 g for 1 min.

2.3.5.6 Sequencing of cloned PCR products

Cloned PCR products were sequenced at GATC biotech (Germany) by sending the gel extracted DNA including the universal M13 forward primer.

2.3.6 Probe labelling

Purified PCR products or amplified cloned inserts were used as template for labelling with biotin-11-dUTP or digoxigenin-11-dUTP (Roche Diagnostics). The BioPrime (Invitrogen) DNA Labelling System was used for biotin labelling following the manufacturer's instruction. Briefly, 25 µl of DNA sample (500-1000 ng) was added to 20 µl of 2.5X Random Primers Solution (final volume 45 µl) and then denatured at 95 °C in a water bath. After the denaturation step, 10x dNTP (5 µl) and Klenow Fragment of *E. coli* DNA polymerase I (40U) (1 µl) were added on ice and mixed carefully but thoroughly, and then incubated for 2 hr at 37 °C. The BioPrime (Invitrogen) Array CGH Genomic Labelling Module was applied for digoxigenin labelling. After denaturing 19 µl of DNA sample (500-1000 ng) and 16 µl of 2.5X Random Primers Solution (final volume 35 µl), 10x dUTP nucleotide (3 µl), digoxigenin-11-dUTP (1,8 µl) and Exo-Klenow Fragment (40U) (0,8 µl) were gathered and mixed carefully but thoroughly and incubated for 2 hr at 37 °C. After incubating each sample, stop buffer (5 µl) was added to end the reaction. Finally, a NucleoSpin Extract II Clean-up Kit (Machery-Nagel) was used for cleaning each labelled probe and stored at -20 °C. In addition, six probes of tandem repeats clusters (Table 5.2) have been labelled directly by synthesis with nucleotides linked with biotin.

2.3.7 *Petunia* chromosome preparations

Fresh root tips were firstly pretreated with 0.2 M 8-hydroxyquinoline for 4 hr and then fixed with freshly made ethanol:glacial acetic acid (3:1). Fixed roots were prepared and washed two times for 10 min each in 1x enzyme buffer diluted from 10x enzyme buffer to remove the fixative solution residues, and then digested in 1x enzyme solution at 37 ˚C in incubator for 15-20 min. After this period, root tips were soaked again in 1x enzyme buffer to stop the enzyme action. Clean glass slides (SuperFrost, Menzel-Glaser, Thermo Scientific) were used for preparation under a stereo microscope, with a fine needle and forceps, a single root tip was added to one drop of 60% acetic acid. Then, the root cap was removed keeping only the

meristematic tissues, and cells were separated with the needle and forceps. A glass coverslip (22 mm x 22 mm No. 1) was placed on the prepared cells and excess of acid removed by a filter paper. The coverslip was tapped gently to disperse the material and then squashed by force with the thumb. A phase contrast microscope was used for visualizing the slides to check cell quality, metaphase index, and to confirm if they were well squashed. After checking, the slides were transferred on to dry ice for 15 min and the cover slip was removed with a razor blade. The slides were air dried and stored at room temperature and checked again (see Table 2.2).

2.3.8 Fluorescent *in situ* hybridization

The method of Schwarzacher and Heslop-Harrison (2000) was applied for fluorescent *in situ* hybridization.

2.3.8.1 Pre-hybridization

Selected slides with free chromosomes and good metaphase distribution were re-fixed using immersion in new fixative absolute ethanol:glacial acetic acid (3:1) mixture for 10 min. Slides were soaked in absolute ethanol for two times 5 min each and air dried, and then incubated with 200 µl of RNase (100 µg/ml) solution each under a plastic cover slip at 37 °C for 1 hr in a humid chamber. The plastic coverslips were removed from slides by washing in 2xSSC for 2 and 10 min respectively, then slides were incubated in a pepsin solution (400 µl pepsin:80 ml 0,01M HCL) for 3 min at room temperature and washed with distilled water and 2xSSC for 1 and 5 min respectively. Next, slides were incubated at RT for 10 min with saline buffered paraformaldehyde 4 % (w/v) (Fisher scientific) under the fume hood. The 2xSSC solution was used again for washing slides, and then slides were dehydrated in 70%, 85% and 100% ethanol for 2 min each. Then, slides were air dried to prepare them for probe adding (see Table 2.2).

2.3.8.2 Hybridization

The total quantity of hybridization mixture including labelled probe was prepared in an Eppendorf tube (37 µl /slide) (see Table 2.7). The mixture was heated at 95°C for 10 min and cooled on ice for 10 min. The prepared mixture was added to the marked area of a slide and covered with a plastic cover slip. The slides were placed on a thermal cycler and heated to 70°C for 7 min denaturing both chromosomal and probe DNA. For probe hybridization to chromosomal DNA, temperature was decreased slowly to 37°C and kept for 16-20 hr.

Table 2.7 Amounts and concentrations of components in the hybridization mixtures used in FISH.

Components	**Concentrations**	**Amount for 1 slide (µl)**
100% formamide	50%	20
20x SSC	2x	4
50% dextran sulphate	10%	8
Salmon sperm DNA 1µg/µl	0.025µg	1
100mM EDTA	1.25 mM	0.5
10% SDS	0.125%	0.5
Probe DNA	(25-200 µg/µl)	3
Total		37

2.3.8.3 Post-hybridization washes

Post hybridization washes were performed using stringent conditions with 0.1x SSC at 42°C to decrease the background signals and to remove weakly bound probe (see Table 2.2).

2.3.8.4 Detection and DAPI staining

Washed slides were immersed in detection buffer for 5 min. Then, blocking solution was added to the slides and covered with a plastic cover slip and incubated for 30 min at 37 °C. The antibody solutions (40 µl) were added after removing coverslips. The plastic cover slips were added again on top to prevent dehydration and the whole assembly was incubated at 37 °C for 1 hr. Then, detection buffer was used to wash slides three times at 40°C for 2 min each. After that, DAPI and anti-fade (Citifluor AF) mixture (20 µl) were added to the marked area in each slide before adding a glass cover slip (No. 0.24x40 mm). Finally, the slides were placed in the

dark at 4°C overnight before they were checked under an epifluorescence microscope (see Table 2.2).

2.3.8.5 Microscopy and photography

The FISH slides were scanned and viewed by a Nikon Eclipse 80i epifluorescence microscope (B&W CCD camera, Nikon Digital Sight USB (H) EXT I/0 B2W camera controller and a Nikon Intenslight C-HGFI lamp). A drop of immersion oil was placed on the slide after pressing the cover slip carefully and firmly with a filter paper. Three types of band filters were used for checking FISH signals: DAPI (blue) was captured by UV-2E/C (wavelengths of excitation filter 340-380 nm), Digoxigenin (green) by B-2E/C (wavelengths of excitation filter 465-495 nm), and Alexa Flour 594 (red) by G-2E/C (wavelengths of excitation filter 528-553 nm). The red biotin/ Alexa594 streptavidin antibody signal was captured first, followed by the green digoxigenin/FITC-anti-digoxigenin signals, and later the DAPI signal. Each single image was captured using Nikon, NIS-Elements 4.0, false coloured, and overlaid for producing a composite image with the red and green colours for FISH probes and the blue for chromosomes. Adobe Photoshop CC2015.5 was used for preparing and overlaying images and hybridization signals, apart from cropping, using only functions affecting the whole image equally.

Chapter III. Integration of Endogenous Pararetroviruses (EPRVs)

3.1 Endogenous pararetroviruses (EPRVs)

For a long time, a principle in virology stated that there was no evidence for plant viruses integrating into the host genome unlike bacterial and animal viruses. This dogma has been overruled with the discovery of endogenous pararetroviruses in plants. Most of our knowledge about EPRVs biology and evolution was gained from two sources: i) the activation of integrated EPRVs that gave factual proof of their infectivity under particular conditions and ii) identification of positioning and diversity within sequenced plant genomes. Richert-Pöggeler and Shepherd (1997) reported that *Petunia vein clearing virus* transmits through seeds and integrates in the *Petunia hybrida* genome. The whole virus sequence is present in most of *P. hybrida* and the parental species *P. axillaris* subsp. *axillaris* and *P. integrifolia* subsp *inflata* , and capable of episomal virus replication in response to stress (Richert-Pöggeler *et al.* 2003). Ndowora *et al.* (1999) and Harper *et al.* (1999) showed that *Banana streak virus*-OL integrates at two loci in the *Musa* genome using FISH and could trigger virus infection in healthy plants. In *Musa balbisiana* genome, four species of BSV have recently been reported as pathogenic integrants: *Banana streak OL virus* (BSOLV), *Banana streak IM virus* (BSIMV), *Banana streak MY virus* (BSMYV), and *Banana streak GF virus* (BSGFV) (Gayral *et al.* 2008; Iskra-Caruana *et al.* 2010; Geering *et al.* 2014). Lockhart *et al.* (2000) revealed that *Tobacco vein clearing virus* could also integrate in the genome of *Nicotiana edwardsonii* using hybridization tests and genomic library screening. These viruses, *Petunia vein clearing virus* (PVCV), *Banana streak virus* (BSV), and *Tobacco vein clearing virus* (TVCV), show episomal diseases related with integrated motifs (Harper *et al.* 2002).

Nowadays, many sequences of plant viruses have been found as integrants in different plant genomes (Hohn *et al.* 2008; Geering *et al.* 2014).

According to tenth Report of the International Committee on Taxonomy of Viruses (ICTV), there are five DNA plant virus families with ssDNA genomes and the family of *Caulimoviridae* encapsidating dsDNA:

1- The family of *Geminiviridae* has ten genera.
2- The family of *Alphasatellitidae* has twelve genera.
3- The family of *Nanoviridae* has three genera.

4- The family of *Tolecusatellitidae* has two genera.
5- The family of *Caulimoviridae* has eight recognized genera.

Recently, one additional genus has been proposed to the family of *Caulimoviridae*. It was named *Florendovirus* which members are found abundantly in a variety of flowering plants (Geering *et al.* 2010; Geering *et al.* 2014).

3.1.1 *Caulimoviridae*

The *Caulimoviridae* family includes genera of *Caulimovirus, Soymovirus, Cavemovirus, Solendovirus, Petuvirus, Badnavirus, Rosadnavirus, Tungrovirus* and *Florendovirus* with viral genomes from 7.2 to 9.2 kb, which are essentially differentiated from each other by their genome organization (Geering *et al.* 2010; Geering & Hull 2012; Geering *et al.* 2014; Diop *et al.* 2018). Those viruses replicate uniquely by reverse transcription. Virus particles encapsidate a double-stranded DNA genome and virus integration step is not obligatory for viral replication. Their genome consists of open circular double-stranded DNA with single strand discontinuities at particular positions. Based on these biological properties which are distinct from retroviruses, they are also referred to as pararetoviruses (Temin 1985).

These gaps have resulted from DNA replication and are located at the priming positions for plus strand DNA synthesis as well as on the minus DNA strand (Gronenborn 1987; Harper *et al.* 2002). The family produce a more-than-genome-length RNA that takes two roles as a template for reverse transcription and as mRNA for some of gene products (Hull 2002).

Caulimovirus replication occurs in two cellular compartments: first, transcription of a genome length RNA from a viral minichromosomal form in the nucleus, followed by reverse transcription to give circular dsDNA with discontinuties in the cytoplasm (Gronenborn 1987; Harper *et al.* 2002). The genetic information can be either organized in up to eight ORFs comprising single gene products or like the ORF III of *Badnavirus*, encoding for the coat protein (CP), the movement protein (MP), an aspartate proteinase (AP), the reverse transcriptase (RT) and the ribonuclease H1 (RH1) that are involved in virus replication. The genera of *Badnavirus, Petuvirus, Solendovirus* and *Caulimovirus* could integrate within host nuclear genomes and transmit over generations like cellular genes (Geering *et al.* 2014). Recently, Diop *et al.* (2018) reported that both primitive and complex vascular plants contain endogenous caulimovirid-like sequences. Up to five different genera of *Caulimoviridae* have

been detected in angiosperms with florendoviruses being the most prominent representative followed by petuviruses.

3.1.1.1 Petunia vein clearing virus (PVCV)

The virus was detected and described for the first time in Germany by Lesemann and Casper (1973). In a study aimed to identify mycoplasma–like bodies and viruses in *Opuntia tuna* showing witches–broom disease, Lesemann and Casper (1973) reported that it was not possible to detect virus-like particles in ultrathin sections of *O. tuna* tissues and most transmission experiments from infected *Opuntia* were negative. *Petunia* plants that were not used in the transmission tests, but had been raised from the same seed source showed all of sudden vein clearing symptoms and leaf deformation in young shoots with stunted growth. In ultrathin sections of symptomatic leaf and stem material, virus-like particles and inclusion bodies were discovered. Horizontal transmission of the virus to other petunia plants was possible using grafting. Based on host plant and symptoms the infectious entity was named ‘`*Petunia vein clearing virus*’’ (Lesemann & Casper 1973). Later, it has been recorded in United States of America (Lockhart & Lesemann 1998), and in Israel (Gera *et al.* 2000). Particle measurements using purified virus suspensions showed that the average virion diameter with 40-50 nm is slightly smaller than other *Caulimoviridae* (Zeidan *et al*. 2000). Based on immune electron microscopy, no serological relationship to other members of caulimoviruses was found (Lesemann & Casper 1973). Phylogenetic analyses confirmed PVCV being distinct from so far known *Caulimoviridae* and it was classified as type member of the new genus *Petuvirus* (Richert-Pöggeler & Shepherd 1997; Fauquet *et al.* 2005).

The infected petunia plants appear stunted especially in young shoots, and display chlorosis along the leaf veins with either diffuse or sharp borders. Additionally symptomatic leaves show epinasty (Richert-Pöggeler *et al.* 2003). These symptoms (see Figure 3.1) are well expressed when the temperature is more than 25°C under nursery conditions (Zeidan *et al.* 2000). Furthermore, symptoms can arise from stressed plants that suffer from water and nutrient deficiency, in addition to wounding (Lockhart & Lesemann 1998; Harper *et al.* 2002) and tissue culture (Richert-Pöggeler *et al.* 2003).

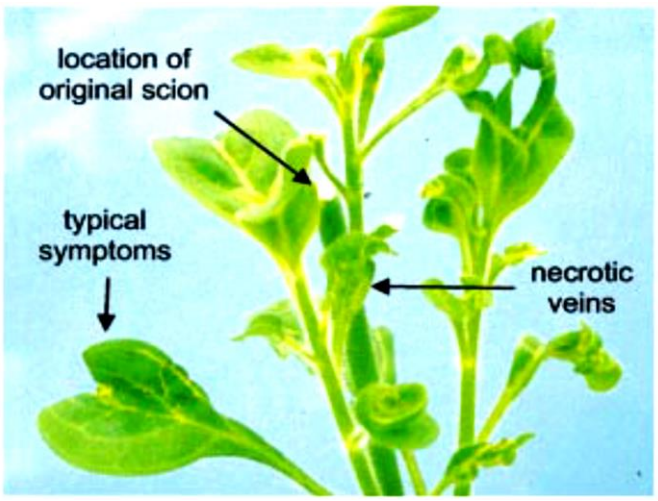


Figure 3.1 The symptoms on *P. axillaris* subsp *parodii* S7 resulted by grafting with a PVCV infected scion of *P. hybrida*. The main symptoms appear as chlorosis along the leaf veins with either diffuse or sharp borders and stunted young shoots (Richert-Pöggeler *et al*. 2003).

Horizontal transmission of episomal PVCV was successful to *Nicotiana glutinosa* and *P. axillaris* subsp *parodii* S7 using grafting. So far, the determined host range of PVCV is limited within Solanaceae, including species of *Petunia* like *P. integrifolia* subsp. *integrifolia, P. axillaris* subsp *axillaris, P. hybrida* and *N. glutinosa* but not *N. tabacum* (Richert-Pöggeler *et al.* 2003). Outbreak of vein clearing disease has been observed for some cultivars of *P. hybrida* such as "Himmelsröschen" (Lesemann & Casper 1973), "Fantasy Pink" (Lockhart & Lesemann 1998), and the "Cascadia" series (Gera *et al.* 2000). Although PVCV transmission by vectors has not been found yet, the viral DNA can be transmitted by biolistic inoculation of petunia plants (Richert-Pöggeler *et al.* 2003).

Screening a genomic library of *P. hybrida* cultivar W138 identified integrated PVCV sequences arranged in a tandem array from which direct full genome length transcript can be synthesized. Indeed these "provirus"-like structures were infectious using biolistic inoculation of provirus-free *P. axillaris* subsp *parodii* S7 (Richert-Pöggeler *et al.* 2003). Most likely release of epigenetic control leads to PVCV infections originating from such endogenous copies in *P. hybrida* W138 after wounding and tissue culture in 30% of the treated plants after 2-3 months (Noreen *et al.* 2007).

Detection of endogenous forms of PVCV in asymptomatic plants using FISH proved that there are homologous sequences in the genomes of *P. axillaris* subsp *axillaris, P. integrifolia* subsp *inflata* and *P. hybrida* (Richert-Pöggeler *et al.* 2003; Bombarely *et al.* 2016). No hybridization

signal was obtained in *P. axillaris* subsp *parodii* and *N. glutinosa* when stringent washing conditions were applied. Thus integrants are vertically transmitted via reproductive cells to the next generation (Richert-Pöggeler *et al.* 2003). Interestingly, FISH studies indicated the collocalisation of PVCV integrants with *Metaviridae* sequences that are phylogenetically the closest relative to *Caulimoviridae* (Bombarely *et al.* 2016; Krupovic *et al.* 2018).

The genome sequence of PVCV is 7206 bp in size (GenBank accession no. U95208). All genetic information is comprised in one large ORF consisting of 2179 amino acids which distinguishes PVCV from other members within *Caulimoviridae* having their genomic information distributed over 3 to 8 ORFs (Hohn & Rothnie 2013) . Within the polyprotein consensus sequences known from the *gag* and *pol* genes of retroviruses for a RNA binding domain (RB), protease (PR), reverse transcriptase (RT) and RNase H (RH) were identified (Richert-Pöggeler & Shepherd 1997; Richert-Pöggeler *et al.* 2003).

An inactivated integrase-like domain and quasi (long) terminal repeats were described for PVCV which are core elements for retrotransposon integration during replication supporting the ancestral link of both phylogenetic groups (Richert-Pöggeler & Shepherd 1997; Richert-Pöggeler *et al.* 2003; Krupovic *et al.* 2018).

The available data on genome assemblies of the two *Petunia* parental species *P. axillaris* subsp *axillaris* N (*PaxiN*) and *P. integrifolia* subsp *inflata* S6 (*PinfS6*) reveals numerous integrated sequences of PVCV (Bombarely *et al.* 2016). FISH was used to localize sequences on chromosomes. In cytological studies distinct patterns for PVCV sequence localization could be seen. In *P. axillaris* N they were mainly identified on chromosome III and less frequently on chromosome VI compared to *P. integrifolia* subsp *inflata* S6, in which only a weak signal on chromosome IV was observed in FISH. So far, almost the complete viral genome length as well as degenerated PVCV sequences have been found in arrays within both assemblies when screened for larger than 500 nt in length against PVCV (GenBank accession no. U95208). Different degrees of preservation of integrated PVCV sequence were found between the two *Petunia* species. Thirty sequences from *PaxiN* ranged between 542 and 2848 nt in length with 80-99% identity, while in *PinfS6*, only nine smaller fragments with 563-635 nt in size revealed 78-80% identity. The route of PVCV integration seems to be similar in both *Petunia* species resulting in single insertions as well as small tandem arrays (Richert-Pöggeler *et al.* 2003; Bombarely *et al.* 2016).

3.1.1.2 *Florendovirus*

The newly described genus *Florendovirus* (Geering *et al.* 2014), out of the family of *Caulimoviridae,* has colonized the genomes from a large variety of flowering plants comprising 34 distinct viral species. Endogenous viral sequences have been found in monocot as well as eudicot plants including potato, maize, papaya, soybean, apple, citrus, cacao, grape, tomato, cassava, rice and other plants. The name '*Florendovirus*' based on the existence in flowering plants referring to *Flora* the Roman goddess of flowers and to the fact of being *endo*genous. Further, depending on 80% nt identity in the reverse transcriptase (RT)-ribonuclease H (RH) domains and other sequence clusters, 34 representative distinct virus species have been identified and classified giving them names including the host species name. The sequence fragments have been assembled into a total of 76 entire or nearly full-length *Florendovirus* genomes sized between 7.2 and 8.5 kb. They can encode for one or two open reading frames (ORFs) and thus were divided into two categories. The majority of florendoviruses have two ORFs in different reading frames. Sequence analysis revealed potential open reading frames encoding various protein domains necessary for viral replication. Phylogenetic analyses have shown that the genus *Florendovirus* branches next to *Petuvirus* that comprises all genetic information for replication within one large ORF. Florendoviruses harbor conserved domains of *gag-pol* genes typical for retroelements, being clearly distinct from known reverse transcribing elements like monopartite *Caulimoviridae* and diploid retroviruses (Geering *et al.* 2014; Bombarely *et al.* 2016).

3.2 Aims and objectives

Aims: The main aim of this chapter is to determine the interaction, chromosomal location of PVCV and other endogenous pararetroviruses in *Petunia* species, to find out more about episomal and induced infections, study expression patterns and EPRVs sites within *Petunia* chromosomes.

Objectives:

- To identify sequences, copy numbers and genome proportions of PVCV, florendoviruses, and caulimoviruses in *Petunia* genomes.

- To define how EPRVs are organized and integrated in *Petunia* species and find their chromosomal location.

- To define relationships, pathways and mechanisms for evolution of integrated viral sequences and their episomal counterparts

- To investigate viral sequence activity or expression, involving nuclear and episomal copies, and sequences interactions.

3.3 Materials and methods

The range of approaches and workflow used in this chapter is presented in Figure 3.2.

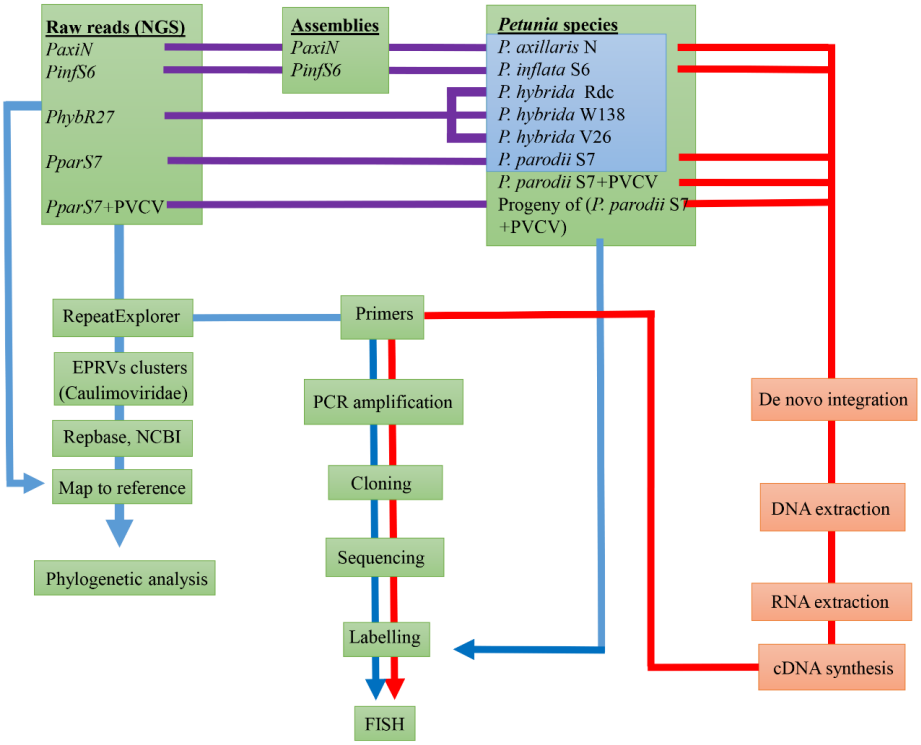


Figure 3.2 Flowchart shows the outline of the materials of five genome raw reads (NGS), two assemblies and three wild *Petunia* species, three hybrid species, plus two genomes as an episomal and progeny of infected plants. Use of bioinformatic tools was in parallel with molecular biology and cytology techniques (PCR amplification, cloning, sequencing and FISH) (blue line). On the other side, the line of *de novo* integration started with non infected three wild species as well as two infected species with horizontally (*P. axillaris* subsp *parodii* S7+PVCV) and vertically (progeny of *P. axillaris* subsp *parodii* S7+PVCV) transmitted infections. DNA and RNA extractions as well as cDNA synthesis have processed toward PCR amplification and FISH techniques (red line). Each species has particular raw reads in addition to two assemblies for the parents (*P. axillaris* subsp *axillaris* N and *P. integrifolia* subsp *inflata* S6), while one raw read set (*PhybR27*) was used as a representative for the three hybrid species (*P. hybrida* Rdc, V26 and W138)(purple line).

3.3.1 EPRV fragments in *Petunia* species

3.3.1.1 Next generation sequencing

The raw reads of *P. hybrida* R27 (*PhybR27*) as a representative of the hybrid genomes was obtained from Aureliano Bombarely, Department of Horticulture, Virginia Tech, USA in addition to the assemblies of *P. integrifolia* subsp *inflata* S6 (*PinfS6*) and *P. axillaris* subsp *axillaris* N (*PaxiN*) genomes (Bombarely *et al.* 2016). The genome of healthy, provirus-free *P. axillaris* subsp *parodii* S7 and PVCV infected *P. axillaris* subsp *parodii* S7 were sequenced commercially by Novogene Company Limited, Hong Kong, China using Illumina HiSeq-PE150 reads. The bi-directional reads have been paired by Geneious software before uploading through RepeatExplorer and doing 'Map to reference'. These data have been used in a basic search for EPRVs clusters within genomes sequences by RepeatExplorer (Novák *et al.* 2013), Repbase (Jurka *et al.* 2005; Bao *et al.* 2015)**,** and Ubuntu Linux 13.10, using Geneious software (http://www.geneious.com/) (Kearse *et al.* 2012).

3.3.1.2 Graph-based read clustering with RepeatExplorer

This pipeline (Novák *et al.* 2013) has been used to explore EPRV clusters in the whole raw reads. The program only has the capability to recognize EPRVs at family level as caulimoviruses, and clusters of EPRVs needed further characterisation to identify to species level. Total contigs were extracted and each cluster was submitted through Repbase (Jurka *et al.* 2005), Basic Local Alignment Search Tool (Altschul *et al.* 1990), and then aligned to known viral sequences from DPV web (Adams & Antoniw 2005) and Repbase dataset (Jurka *et al.* 2005) to identify virus sequences on the genus and species levels using the alignment tool in Geneious.

3.3.1.3 *De novo* assembly

Geneious software assembler was applied to *de novo* assembly of the whole raw reads of infected *P. axillaris* subsp *parodii* S7. The bi-directional reads have been paired by Geneious software before subjecting to the assembler, then 23 % of the reads were taken to generate more than 100 contigs with variable lengths. Every individual contig was constructed of multiple overlapping assembled reads with consensus sequence for each contig. These consensus

sequences were applied to ‘Map to reference’ tool using well identified references (PVCV sequence), and then aligned (Figure 3.3).

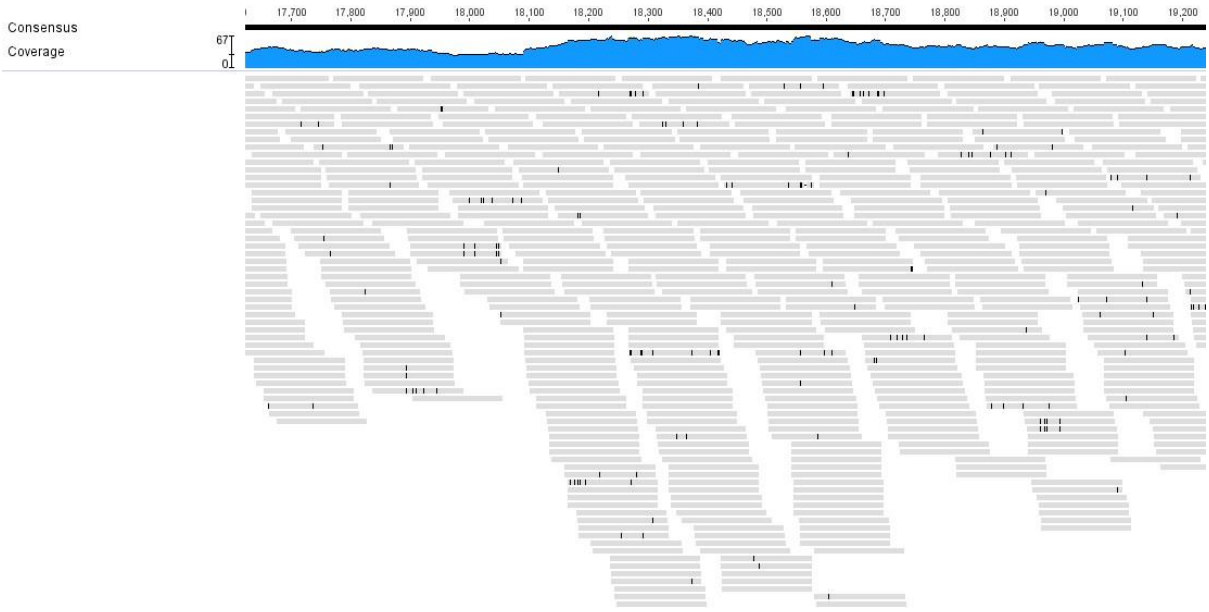


Figure 3.3 Image of *de novo* assembly tool showing assembled raw reads that overlapped to construct single contig with consensus sequence. The blue line indicates coverage with variable depths over the whole sequence. The scattered black dots within reads indicate single nucleotide polymorphisms (SNPs).

3.3.1.4 Map to reference

This bioinformatic tool was applied using Geneious and other software packages (Kearse *et al.* 2012), to assemble the raw reads of all examined NGS data for EPRV sequences (*Petuvirus*, *Florendovirus* and *Caulimovirus*). The result was included in the main report that has number of assembled reads and the whole used reads as well as highly frequent overlapped reads have been incorporated in one contig and consensus sequence. This report data was used to calculate genome proportions and copy numbers as below:

1- Genome proportion: number of assembled reads / number of total NGS reads x 100.
2- Copy number: number of assembled reads x read length/ reference sequence length. (Mustafa *et al.* 2018) (see Figure 3.4).

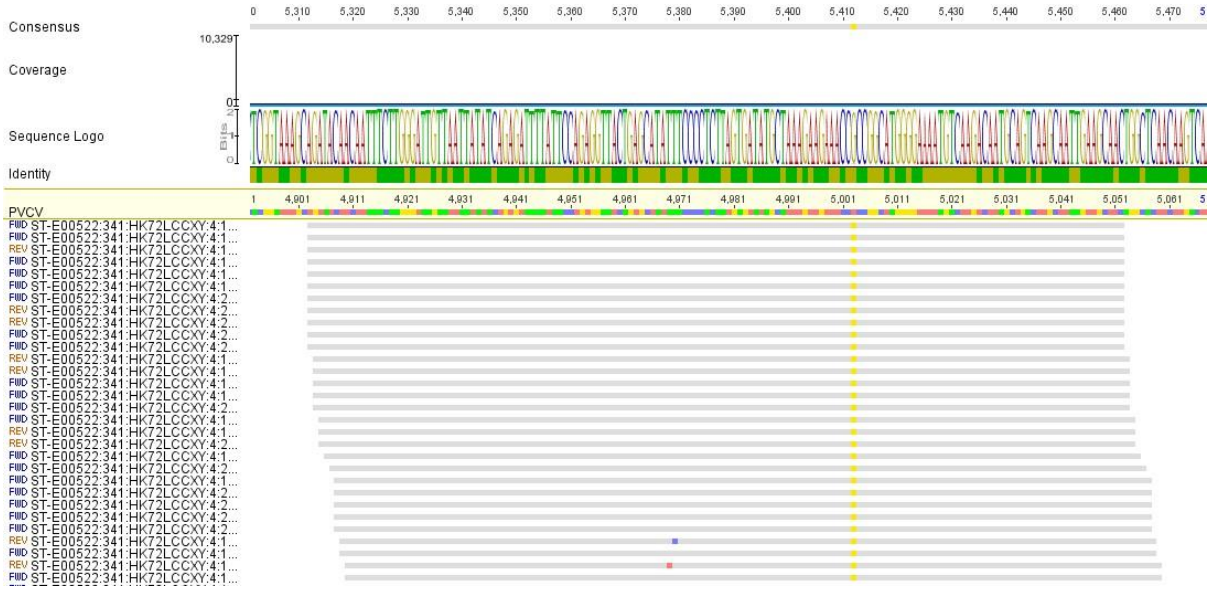


Figure 3.4 Reference mapping tool showing mapped raw reads to the reference sequence (PVCV genome) to construct consensus sequence. The yellow line within reads indicate recombinant single nucleotide polymorphism (SNP) observed in high numbers of reads, while the red and blue dots indicate only one SNP for each dot within individuale read.

3.3.1.5 Phylogenetic analysis

In order to choose a robust model for phylogeny, MEGA7 (Tamura *et al.* 2013) was used with maximum likelihood method. Firstly, the Geneious software (founded by Biomatters; available from http://www.geneious.com) was applied for alignment with default parameters and optimized manually. Then, ClustalW alignment was used for the extracted sequences (about 6500 bp for each). After that, a phylogenetic tree has been reconstructed by General Time Reversible (GTR) as a best substitution model. Bayesian phylogeny inference was used for analysis with Bayesian inference of phylogeny (MrBayes 3.2.6). (Huelsenbeck & Ronquist 2001).

3.3.1.6 PCR amplification

To confirm the "*in silico*" results obtained from the RepeatExplorer and 'Map to reference', primers based on sequences out of the identified EPRV-like clusters were designed as below:

1- Petuvirus-like sequences

Petuvirus cluster was found in the RepeatExplorer contigs (CL205), primers and cycling program listed in Tables 3.1 and 3.2 respectively. PCR amplification was applied to confirm presence or absence of missing parts of PVCV within petunias according to preliminary results

of ‘map to reference’ that show some parts of PVCV were missing from *P. integrifolia* subsp *inflata* S6 raw reads (Figure 3.5).

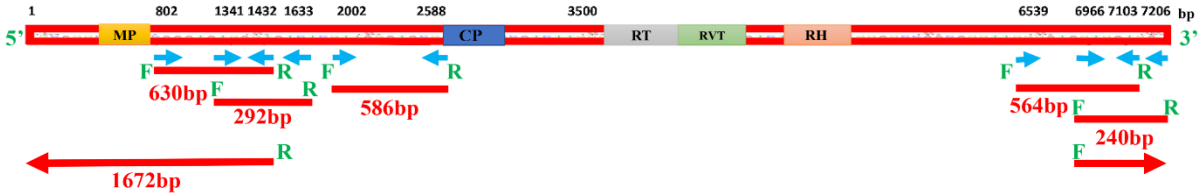


Figure 3.5 Distribution of primers and generated fragments within coding and non-coding regions. These parts have been selected as missing regions in *P. integrifolia* subsp *inflata* S6 raw reads (results from ‘Map to reference’) over the entire length of PVCV, while other parts have been found with variable pairwise identity and degeneration. The full length comprised five main protein domains, MP (movement protein), CP (coat protein), RT (reverse transcriptase), RVT (RNA dependent DNA polymerase), and RH (RNase H) respectively.

The main three fragments of endogenous PVCV (left, middle and right from accession no. AY228106) (Figure 3.6) were combined together by adding 220 ng of PVCV-L, 600 ng of PVCV-M, and 480 ng of PVCV-R, these concentrations have calibrated according to suitable volumes of the purified fragments. The mixture was applied as a probe in FISH (Table 3.4), and the cycling program mentioned in Table 3.2.

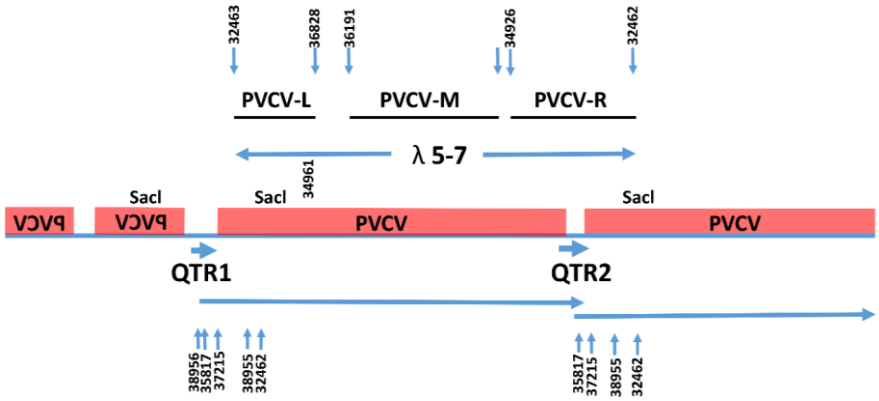


Figure 3.6 Organization of chromosomal PVCV in lambda clone 5 (Accession no. AY228106) of *P. hybrida* W138 library with PVCV-middle (-M), right (-R) and left (-L) fragments in addition to their primers with arrows above to cover each part. The reverse primer of the middle part (34961) located in real above pink box next to SacI position. The pink boxed regions are coding sequences interrupted with non-coding regions. PVCV sequences at the 5 end found in reverse orientation (mirrored PVCV). The two lines below QTRs indicate putative transcripts and numbers below those lines indicate primers used to analyse QTRs. Figure taken from (Richert-Pöggeler *et al.* 2003).

2- Florendovirus-like sequence

This cluster was extracted from the RepeatExplorer contigs (CL131) with about 3000 nt and then characterised by aligning the sequence with some well identified florendoviruses that were taken from Geering *et al.* (2014) in addition to further identification in Repbase and NCBI.

The *Florendovirus* primer was designed in two steps. First, scaffolds of the *Petunia* assemblies were searched for *Florendovirus* sequences and the shared scaffolds have been collected. Next, the identified scaffolds that have *Florendovirus* sequences were aligned and a consensus sequence determined in order to design the forward and reverse primers respectively (Figure 3.7). So far, designed primer represents about 63 of 76 of *Florendovirus* members, and sequence analysis revealed that the consensus sequence comprised conserved domains of RNaseH (RH), and two conserved domains out of the reverse transcriptase region of reverse transcriptase (RNA dependent DNA polymerase, RVT), and of reverse transcriptase (RT-LTR), with 697 bp length (Figure 3.8). Many primers were designed and the best set (FlorB) was chosen (see Tables 3.1 and 3.3).

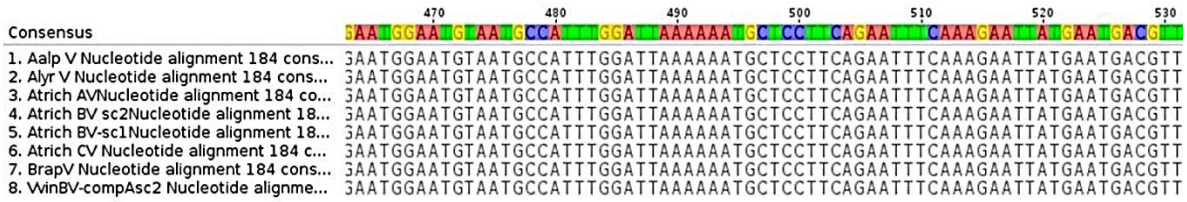


Figure 3.7 Alignment of multiple *Florendovirus* members that shared similar scaffolds in the assemblies; the consensus has been applied to design forward and reverse primers (Scf380F and Scf380R).

Figure 3.8 FlorB consensus sequence organization with conserved protein domains, the total size is 697 bp (Scf380F and Scf380R), the consensus sequence was inverted from 3' to 5', and comprised conserved domains of RNaseH (RH), of two conserved domains out of the reverse transcriptase (RT) out of ORF3 region of *Florendovirus* member in *P. axillaris* subsp *axillaris* N.

3- Caulimovirus-like sequence

Caulimovirus sequences were taken from the RepeatExplorer contigs (CL112) with about 1700 nt, and then one set of primers was designed from the consensus sequence with length of 1030 bp, comprising reverse transcriptase (RNA dependent DNA polymerase, RVT) as a part of ORF2 (Figure 3.9; Tables 3.1 and 3.3) .

Figure 3.9 Consensus sequence of caulimovirus-like sequence that shows forward and reverse primers of 1030 bp (RE-CL112F and RE-CL112R), comprising one conserved domain out of the reverse transcriptase domain as a part of ORF2.

Table 3.1 List of EPRVs primers.

No.	Primer name	Sequence	Tm °C	Product length (bp)
1	PVCV (802F-1432R)	(802F) CCC AAT GTA CCC AAG TCC C (1432R) GGC TCC ACT GTC AGA TGA GG	59.5	630
2	PVCV(1341F-1633R)	(1341F) GTG GGA TCA TTG AGC AGG CC (1633R) GGG TCT TTT GCA CCA GAT GG	62.8	292
3	PVCV(2002F-2588R)	(2002F) CAA ATC CGC CTA ACT CAC CG (2588R) CTT AGG TGG AGT GAT TTC AGG	58	586
4	PVCV(6539F-7103R)	(6539F) GCC ATC AAC CCC CTC ATG G (7103R) GGC ACC ATA CAG CCA TAA TAC C	61.9	564

5	PVCV(6966F -7206R)	(6966F) CTC TCT CTT GTT TCC AAA CTC (7206R) ATA CGA GAT ATG GAG GAA TT	50.6	240
6	*Florendovirus* (FlorB)	(Scf380F) GGT CAG CAC CAG AGT CTA G (Scf380R) GCG GAA CAA CTC CAA GGT GAC	57	697
7	*Caulimoivirus*	(RE-CL112F) CCT CAC CTG GAT CCG ATA TCT C (RE-CL112R) GCT CCG AAT AGT TTC AGC GG	60	1030

Table 3.2 PCR cycling program for PVCV primers.

Step		**Temperature °C**	**Duration** **Minuets :seconds**	**Cycles**
1	Initial denaturation	94	2:00	1
2	Denaturation	94	00:45	30
3	Annealing	58	00:45	
4	Extension	72	2:00	
5	Final extension	72	10:00	1
6	Hold	16	pause	1

Table 3.3 PCR cycling program for *Florendovirus* and *Caulimovirus* primers.

Step		**Temperature °C**	**Duration** **Minutes :seconds**	**Cycles**
1	Initial denaturation	95	3:00	1
2	Denaturation	95	00:30	35
3	Annealing	50.2-69.3	00:30	
4	Extension	72	1:00	
5	Final extension	72	1:00	1
6	Hold	16	pause	1

3.3.1.7 Cloning, sequencing and probes labelling

These methods were mentioned earlier in 2.3.5 and 2.3.6 respectively.

3.3.1.8 Chromosome preparations and fluorescent *in situ* hybridization

These protocols were explained earlier with details in 2.3.7 and 2.3.8 respectively. Three main fragments of chromosomal PVCV used in FISH are in Table 3.4.

Table 3.4 PVCV hybridization probes used in FISH.

No	Primer name	Nucleotide sequence (5'-3')	Tm °C	Length (bp)	Reference
1	PVCV-M (middle)	(36191F) CAG TCT AGC AGT CAC CTT GG (34961R) TGC TCT CAT GTC CAT TTC AAC C	58	3000	Richert-Pöggeler *et al*., 2003
2	PVCV-L (large)	(32463F) CAA GGA GCT CCC CTT ACA AAA GAC TCC (36828R) CGA GAA CTC TGA TAA GAC CTT G	60	1100	Richert-Pöggeler *et al*., 2003
3	PVCV-R (right)	(34926F) TTG CTG ATT TCC TAT CAA GGC C (32462R) AGG GGA GCT CCT TGG ATT TGG ACT TGG	66	2400	Richert-Pöggeler *et al*., 2003

3.3.2 *De novo* integration

P. axillaris subsp *parodii* S7 had been selected as a model plant to study PVCV *de novo* integration since it did not contain provirus-like sequences (Richert-Pöggeler *et al.* 2003). A model system to study *de novo* integration was set up by Richert-Pöggeler (unpublished). *P. axillaris* subsp *parodii* S7 had been infected using biolistic inoculation on plants grown under sterile conditions on Murashige Skoog (MS) agar in magenta boxes (100 x100 x120 mm). As inoculum, the infectious full-length clone p72-2, 3c of PVCV (Richert-Pöggeler *et al.* 2003) was used. Successful PVCV infection was proven using PCR verified by immune electron microscopy. PVCV infected plants were maintained using cuttings under sterile tissue culture conditions for more than 8 years. For seed production plants were transferred to soil and grown under greenhouse conditions and maintained using cuttings. In leaves with vein clearing symptoms presence of episomal virus replication was proven by immune electron microscopy. PVCV positive plants were registered under the number EM09-348 and manually pollinated. Harvested seeds were plated first on MS agar plates and maintained in tissue culture transferring cuttings to new media every 6-8 weeks. Episomal PVCV infection was proven using immune electron microscopy.

3.3.2.1 Plant material and cultivation

The parents *P. axillaris* subsp *axillaris* N and *P. integrifolia* subsp *inflata* S6 in addition to provirus free, PVCV-infected *P. axillaris* subsp *parodii* S7, and next generation of infected *P. axillaris* subsp *parodii* S7 maintained by tissue culture, were used in this study. Seeds of the three wild *Petunia* species have been provided by Cris Kuhlemeier, University of Bern, Switzerland. The seeds were placed after surface sterilization on Murashige Skoog (MS) (Appendix 4.3) agar plates (2.165 g of MS salts with macro and micronutrients added to 10 g of sucrose and dissolved in ultrapure water and then filled to 1 liter. After that, 8 g of bacto agar added to the mixture and autoclaved at 121°C for 30 min, 30 ml added for each plate). After three weeks, the young plants were transferred directly to soil or in peatmoss (Klassmann substrate 1 type) and cultivated at Julius Kühn-Institut (JKI) greenhouse, Braunschweig, Germany (see Appendix 4.1). Light (12-14 hr), temperature (20-25°C) and humidity (60-80 %) conditions (Table 3.5).

Table 3.5 The descriptions of the five DNA samples from different petunias.

Samples		Registration codes	Descriptions
i	*P. axillaris* subsp *axillaris* N	TEM16-124_1-6	Mixed sample of six plants TEM16-124_1 to _6, PVCV sequence insertions have been found in the genome assemblies.
ii	*P. integrifolia* subsp *inflata* S6	TEM15-652	PVCV sequence insertions have been found in the genome assemblies.
iii	*P. axillaris* subsp *parodii* S7+ PVCV	TEM16-35_1	Episomal PVCV genome, horizontally transmitted virus via grafting.
iv	*P. axillaris* subsp *parodii* S7	TEM16-315_1-3	Mixed sample of three plants TEM16-315_1to _3, no episomal PVCV genome detected.
v	progeny of *P. axillaris* subsp *parodii* S7 + PVCV	EM09-348	Parental plant registered as EM09-348 to study if episomal replicating virus can be vertically transmitted to the next generation (progeny) via *de novo* integration (Figure 3.35).

3.3.2.2 DNA extraction

Total DNA of *P. axillaris* subsp *axillaris* N, *P. integrifolia* subsp *inflata* S6, *P. axillaris* subsp *parodii* S7, PVCV-infected *P. axillaris* subsp *parodii* S7 (plants were grown in greenhouse), and progeny of PVCV-infected *P. axillaris* subsp *parodii* S7 (plant was grown in tissue culture) has been extracted and purified from leaves using Edwards's protocol (Edwards *et al.* 1991) with some modification. Small petunia leaves were harvested in a tube with one small grinding ball and then frozen in liquid nitrogen and ground well for 30 sec. After that, extraction buffer (800 µl) was added and mixed well, tubes have been applied for heat extraction in heat block at 60°C for 30 min followed by centrifugation for 15 min at 1306 g. Next, supernatant (600 µl) was transferred to a new tube and 600 µl iso-propanol (ice-cold) was added and then centrifuged for 5 min at 11752 g. After the supernatant was poured off, 1000 µl of 70% ethanol (ice-cold) was added and centrifuged for 5 min at 11752 g, and the supernatant was discarded and pellets left to dry at room temperature. Finally, pellets were dissolved in 100 µl water and measured for DNA concentration using a NanoDrop ND-1000.

3.3.2.3 RNA extraction

RNeasy Plant Mini Kit (QIAGEN) was used for RNA extraction and plant material was about 100 mg for each sample, RLT buffer (445.5µl) with mercaptoethanol (4.5µl) added to each sample and mixed well vigorously, and then transferred to QLAshredder spin column and centrifuged at 12752 g for 2 min. After that, the supernatant was relocated to a new Eppendorf tube without disturbing the cell debris pellet and 200µl of ethanol (96%) was added, and transferred then to the RNeasy Mini spin column for spinning at 12752 g for 15 sec. The flow through material was discarded, RW1 buffer (350µl) added and centrifuged for 15 sec at 12752 g. The DNase –RDD mixture was prepared earlier to be ready at this stage by adding 10µl of DNase with 70µl RDD buffer. Later, this 80µl samples were added to the center of RNeasy tube membrane and incubated for 15 min at room temperature. After this step, buffers [RW1 (350µl) and RPE (500µl)] were added to the membrane and then centrifuged for 15 sec at 12752 g. For collecting the purified RNA, the spin column was placed in a new 2 ml collection tube and centrifuged at 12752 g for 1 min and then placed in a new Eppendorf tube and 30µl of RNase free water added and centrifuged for 1 min at 12752 g. Finally, this sample was measured using a NanoDrop ND-1000 to check purity and concentration.

3.3.2.4 cDNA- synthesis

RNA samples were used for cDNA synthesis using oligo dT primer [M4-T, 5'-GTTTTCCCAGTCACGAC(T)15-3' described by Chen *et al.* (2002)] since a polyA-tail is present at the 3' end of the viral RNA. Generated cDNA is a prerequisite for PCR with virus specific primers to monitor transcription of viral sequences. Omniscript Reverse Transcription Kit (QIAGEN) was used for obtaining the complementary DNA of the RNA samples. In a PCR tube, 3µl of each RNA sample added to 9µl of RNase free water and then mixed well, the mixture denatured at 65°C for 5 min in a PCR machine to remove RNA secondary structures. Components of cDNA-synthesis (Table 3.6) mixed well in a PCR tube and then 8µl for each sample was added to the denatured sample, then incubated in a PCR machine at 37°C for 1 hr.

Table 3.6 Components and quantities of cDNA synthesis mixture.

cDNA synthesis	**µl/sample**
10x RT buffer	2
dNTP mix (5mM)	2
Primer M4-T (10 µM)	2
RNase inhibitor (4U/µl)	1
Omniscript Reverse Transcriptase	1
Total for each sample	8

3.3.2.5 Gel electrophoresis of RNA samples and cDNA-synthesis

3-4 µl of RNA sample mixed with 1-2µl of RNase free water then added to 5µl of 2x loading dye. 2µl of RNA marker (Thermo Scientific RiboRuler High Range RNA-Ladder, 6000 bp-200 bp), was mixed with 5µl 2x RNA loading dye (Thermo Scientific) and 3µl H_2O. Samples and markers were denatured in a PCR machine at 70°C for 10 min and then directly put on ice for 5 min before running on 1.5% agarose gel in 1x TAE-buffer.

For cDNA samples, 5 µl of each cDNA was added to 5µl of RNase free water then mixed with 2 µl of 6x loading dye. 1µl of DNA ladder added to 1µl of 6x DNA loading dye and then mixed with 4µl H_2O. Samples and markers were added on 1.5% agarose gel electrophoresis to run for 1 hr at 80 volts.

3.3.2.6 Reverse transcriptase-PCR amplification

cDNA of each sample was used as a template with specific primer pairs for PVCV named F 16a and R16b (Richert-Pöggeler *et al.* 2003). PVCV primers 16a and 16b produced a 483 bp fragment comprising part of the conserved RT domain within PVCV pol region (Figure 3.10):

F16a (3972 to 3988 bp): CGC ATT GGA GCA GAT GG
R16b (4454 to 4437 bp): GTG AGA GAA GAG TGT GAG

Figure 3.10 PVCV primer location within the whole sequence of PVCV, including RT domain and the total size is 483 bp.

For *Florendovirus* and *Caulimovirus*, primers were used as in Table 3.1.

Genomic DNA was diluted to 30-50 ng/µl with sterilized water and used as a template. The dNTPs and primers were also diluted with sterilized water. A PCR machine (PTC-100, Programmable Thermo Controller, MJ Research) was used for primer amplification. The master mix components (Table 3.7) were set up for PCR reactions using the specific program that consisted of 2 min for initial denaturation at 95°C and then 35 cycles for denaturation (95°C for 45 sec), annealing (54°C for 45 sec) and primer extension (72°C for 1 min). The last cycle was 10 min at 72°C as a final extension. After these steps, the temperature was set up at 10°C as indefinite hold time (Table 3.8). PCR products were checked using 1.5% agarose gel electrophoresis in 1x TAE-buffer with 75 volts for 1 hr.

Table 3.7 Components and quantities of PCR Master Mix.

PVCV-PCR	**Sample (µl)**
10X buffer	5
dNTP-Mix 10mM	1
Primer 16a forward	2
Primer 16 b reverse	2
$MgCl_2$ 25mM	4

Taq- Polymerase (5U/µl) (HotStart)	0.25
RNase-free water	35
Total for 1 sample	49.25

Table 3.8 PCR cycling program using PVCV primers F16a and R16b.

Step	**Temperature °C**	**Duration Min:Sec**	**Cycles**
Denature	95	2 :00	1
Denature	95	0:45	35
Annealing	54	0:45	
Extension	72	1 :00	
Final extension	72	10 :00	1
Hold	10	pause	1

3.3.2.7 Purification of PCR- product

QIAquick PCR Purification Kit was used for purifying PCR products following the manufacturer's instructions. A 5x volume of PB buffer were added to the sample in a new tube and mixed gently. The mixture was added on the lilac column and spun down for 1 min at 13793 g and then the flow-through was discarded. Next, PE buffer (750µl) was added and centrifuged for 1min at 13793 g, and then the flow-through was discarded. After that, the column was spun down for 1 min at 13793 g to remove residual wash buffer. Finally, the column was placed in a clean 1.5 ml Eppendorf tube and 50µl ultrapure-water added and left for 1min at room temperature before centrifugation for 1min at 13793 g, then the purified PCR product was measured using a NanoDrop ND-1000, and then sent for sequencing.

3.4 Results

3.4.1 EPRVs sequences

Three clusters of pararetroviruses as endogenous sequences within all examined raw reads were identified by the RepeatExplorer pipeline (see Appendix 3.2):

3.4.1.1 Petuvirus-like cluster

The typical member of this cluster is *Petunia vein clearing virus* (PVCV) which is found in CL205 with 34.8%. A very distinct shape of the virus cluster has produced as graph based sequence with interfered circles and the main protein domains (RT, RVT and RH) in the middle (Figure 3.11).

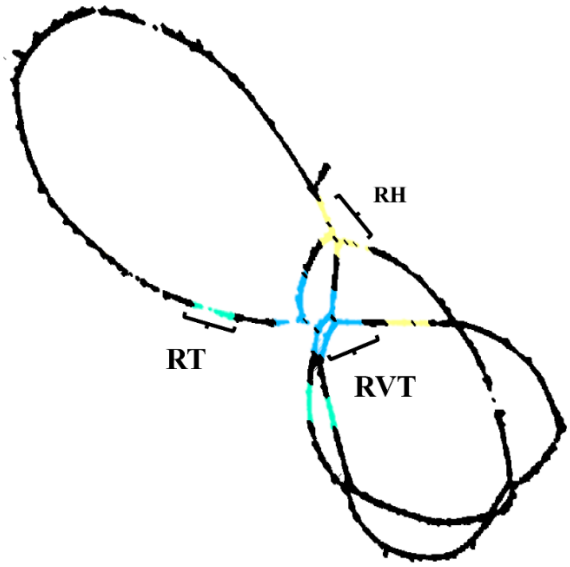


Figure 3.11 Graph based cluster of PVCV in CL205, showing interfered circular shapes and protein domains of RT (in green), RVT (in cyan) in the middle and RH-domain (in yellow). The graphs generated by RepeatExplorer show nodes (dots) representing groups of sequences, connected by edges to other nodes where there is similarity between domains of the sequences. Here, RVT domains are all quite similar, while there is more variability and a longer distance (base pairs) between the RH and RT domains (seen as three loops).

Genome proportions

Using the reference sequence of PVCV, 'map to reference' of raw reads showed variable proportions of PVCV between samples. *PhybR27* was the highest, (0.0278%); while *PaxiN* (0.0075 %) was more than the other parent *PinfS6* (0.0016%). In the healthy *P. axillaris* subsp *parodii* S7 (*PparS7*), the virus proportion (0.000273%) was at the lowest. DNA from the infected plant *PparS7*+PVCV, including episomal PVCV in the DNA used for sequencing, showed the highest proportion of reads mapping to the PVCV reference (0.0588%) (Figure 3.12).

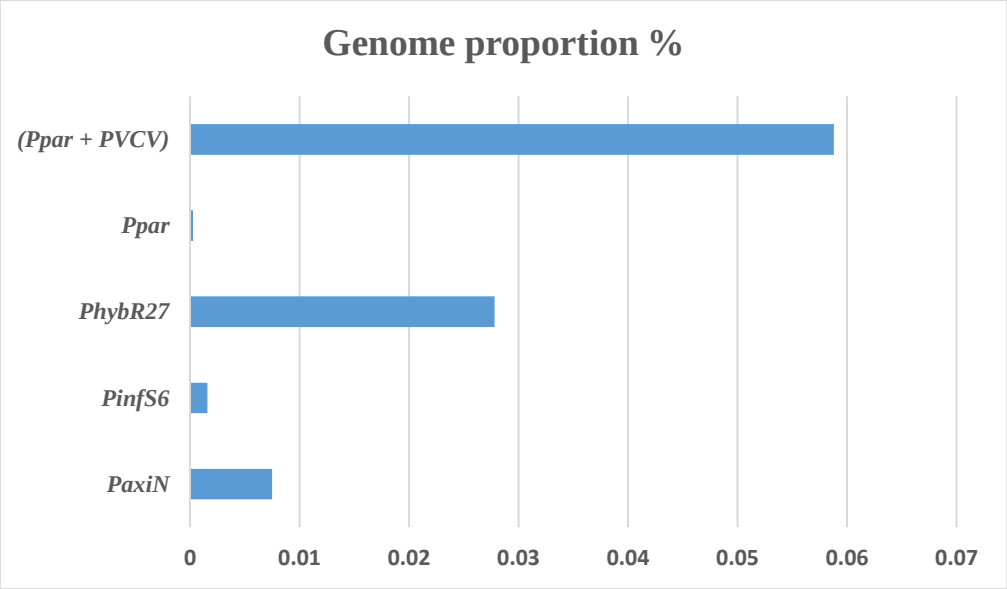


Figure 3.12 Genome proportions of PVCV within all examined raw reads, showing higher proportion in *PparS7*+PVCV, then *PhybR27*, *PaxiN*, *PinfS6* and the lower proportion was in *PparS7*.

Copy numbers

When converted to copy numbers per genome equivalent for integrated PVCV, *PhybR27* had 949 copies, *PaxiN* 172.7, and *PinfS6* 36.8 copies, with minimal numbers (4.5) in *PparS7*. The infected genome of *P. axillaris* subsp *parodii* S7 (*PparS7*+PVCV) has 1075 copies, much more that the integrated PVCV in *PhybR27* by 126 copies while differences were very high with other reads (Figure 3.13).

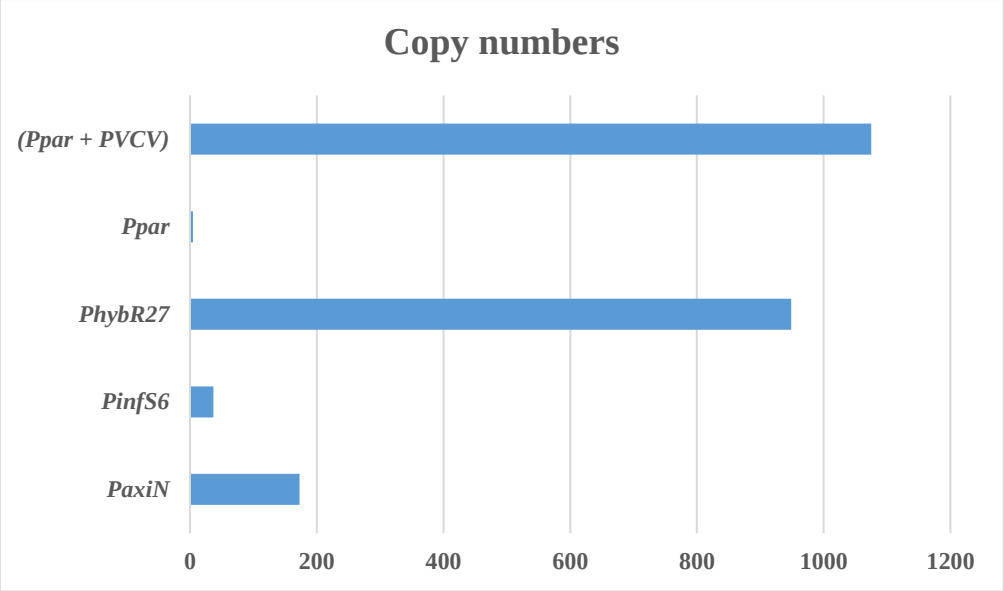


Figure 3.13 Copy numbers of PVCV in all examined raw reads, showing higher proportion in *PparS7*+PVCV and then *PhybR27*, *PaxiN*, *PinfS6* while the lower proportion was in *PparS7*.

PVCV fragments in all examined *Petunia* species

Using the reference sequence of PVCV against whole reads of *Petunia* genomes by map to reference tool revealed that some parts of PVCV were missing from raw reads of some species. To check that result, PCR amplification was applied for some specific regions over the whole sequence of the virus against all *Petunia* DNAs. The first primer set was located between 802 and 1432 bp and successfully amplified the expected size 630 bp in *P. axillaris* subsp *axillaris* N, *P. hybrida* Rdc, *P. hybrida* V26, *P. hybrida* W138, and *P. axillaris* subsp *parodii* S7 in addition to infected *P. axillaris* subsp *parodii* S7 while it is clearly missed in *P. integrifolia* subsp *inflata S6* (Figure 3.14).

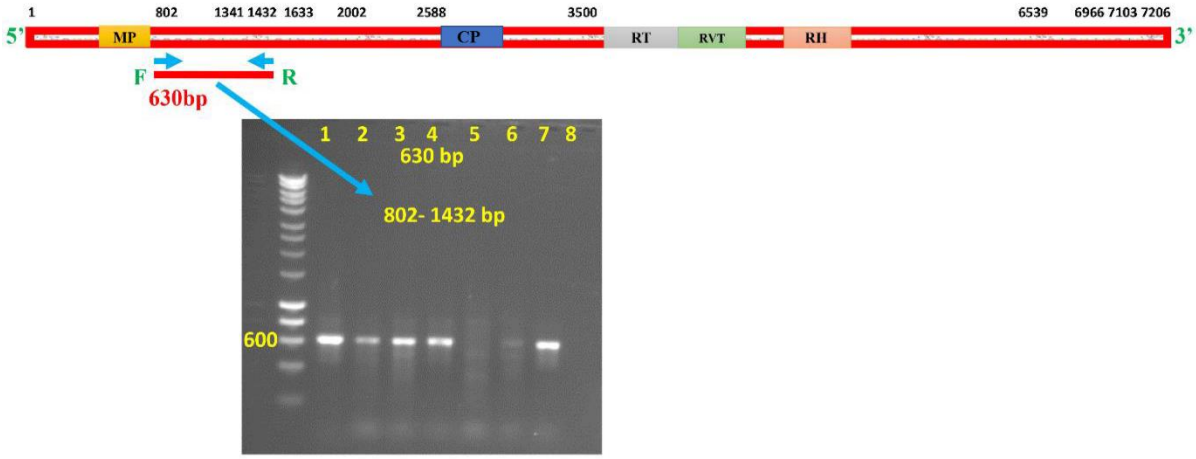


Figure 3.14 First set of primers to amplify 630 bp sequence size between 802 and 1432 bp over PVCV, and find out missing parts in all examined *Petunia* genomes: *P. axillaris* subsp *axillaris* N (1), *P. hybrida* Rdc (2), *P. hybrida* V26 (3), *P. hybrida* W138 (4), *P. integrifolia* subsp *inflata* S6 (5), *P. axillaris* subsp *parodii* S7 (6) *P. axillaris* subsp *parodii* S7 +PVCV(7), and negative control (8), showing that *P. integrifolia* subsp *inflata* S6 has missed this part of sequence unlike other species.

The second and third sets of primers (1341-1633 bp) and (2002-2588 bp) amplified 292bp and 586 bp bands sizes respectively in *P. axillaris* subsp *axillaris* N, *P. hybrida* Rdc, *P. hybrida* V26, *P. hybrida* W138, *P. axillaris* subsp *parodii* S7, and infected *P. axillaris* subsp *parodii* S7 while they failed in *P. integrifolia* subsp *inflata* S6 (Figure 3.15). The fourth set of primers (6539-7103 bp) failed to amplify the 564 bp product in both *P. integrifolia* subsp *inflata* S6 and *P. axillaris* subsp *parodii* S7, and existed in other *Petunia* genomes. The last primer set was designed to amplify 240 bp at the end of PVCV sequence that bioinformatically appeared as a polymorphic region; this failed by PCR in all genomes except the infected *P. axillaris* subsp *parodii* S7 that has full length episomal virus (Figure 3.15).

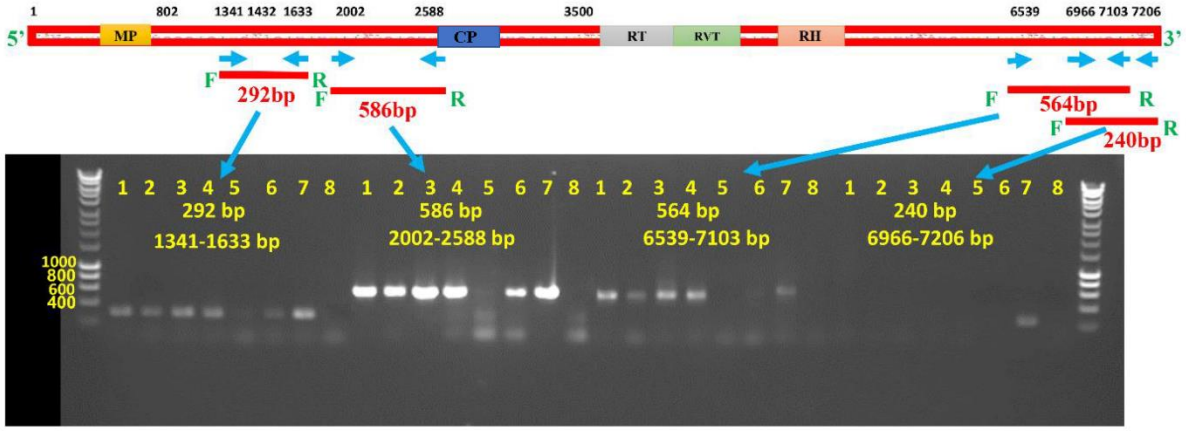


Figure 3.15 Four sets of primers to explore different sizes of sequences that expected to be missed in some species: *P. axillaris* subsp *axillaris* N (1), *P. hybrida* Rdc (2), *P. hybrida* V26 (3), *P. hybrida* W138 (4), *P. integrifolia* subsp *inflata* S6 (5), *P. axillaris* subsp *parodii* S7 (6), *P. axillaris* subsp *parodii* S7 +PVCV (7), and negative control (8). No fragment was amplified in *P. integrifolia* subsp *inflata* S6, while third set failed in *P. axillaris* subsp *parodii* S7. The last part 240 bp of PVCV did not amplify in all genomes except the episomal copies from infected *P. axillaris* subsp *parodii* S7.

In order to confirm whether the last region of integrated virus is really missed or not, another PCR was applied using primer set forward 6966 and reverse 1432 bp to cover the end and the start parts of the virus sequence taking into account that PVCV is tandemly arranged within genomes. This amplification was successful in all hybrids and infected *P. axillaris* subsp *parodii* S7 genomes with expected size 1672 bp, while in *P. axillaris* subsp *axillaris* N, *P. integrifolia* subsp *inflata* S6 and *P. axillaris* subsp *parodii* S7 genomes, the band was less than expected size 250, 900 and 900 bp respectively. Interestingly, in the case of *P. hybrida* Rdc and *P. hybrida* W138, two more bands were noticed with sizes of 900 and 700 bp, those bands were sharper in *P. hybrida* W138 than Rdc (Figure 3.16).

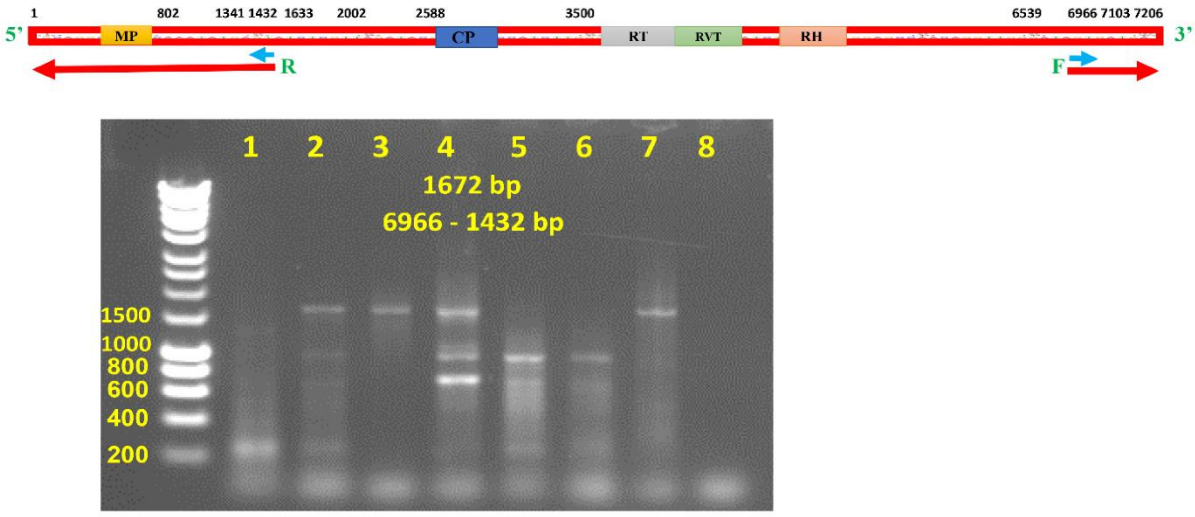


Figure 3.16 Amplified bands have covered the end and start regions of PVCV to confirm if there is missing part at the end of PVCV in genomes of *P. axillaris* subsp *axillaris* N (1), *P. hybrida* Rdc (2), *P. hybrida* V26 (3), *P. hybrida* W138 (4), *P. integrifolia* subsp *inflata* S6 (5), *P. axillaris* subsp *parodii* S7 (6), *P. axillaris* subsp *parodii* S7 +PVCV (7), and negative control (8).

According to the results of 'map to reference' and PCR, *P. axillaris* subsp *axillaris* N and all hybrid plants include the full sequence of chromosomal PVCV except the last part 240 bp that was polymorphic or degenerate. Virus sequence in *P. integrifolia* subsp *inflata* S6 has very degenerated regions that were not possible to amplify by PCR except for some shorter lengths with higher conservation. In *P. axillaris* subsp *parodii* S7 genome, 'map to reference' suggested the PVCV sequence has about nine missing regions and the last part is degenerated (Appendix 3.3), although regions spanning some of the 'missing parts' were amplified by PCR primers.

FISH

The amplified fragments PVCV-M (3000 bp), PVCV-L (1100 bp), and PVCV-R (2400 bp) (Figure 3.17) were combined and labelled using biotin-11-dUTP and digoxigenin-11- dUTP and then applied for *in situ* hybridization over all examined *Petunia* species metaphases.

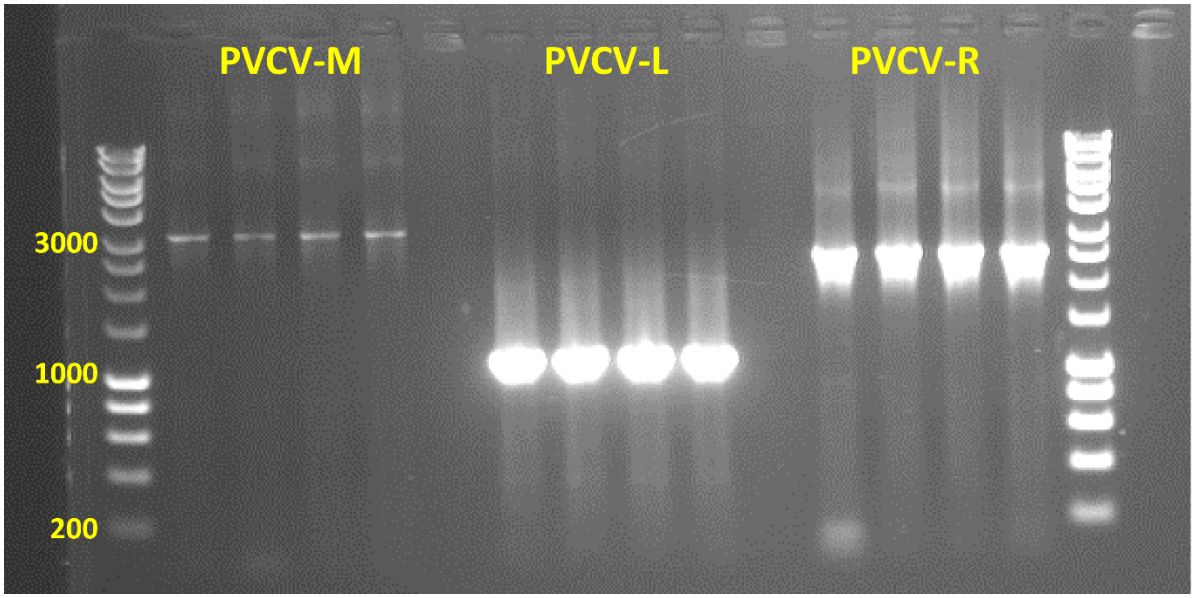
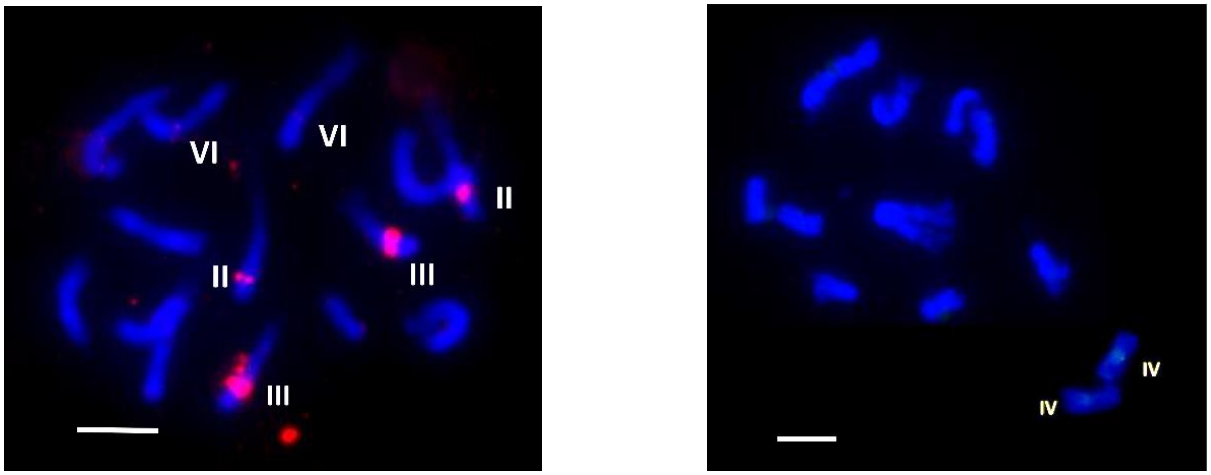
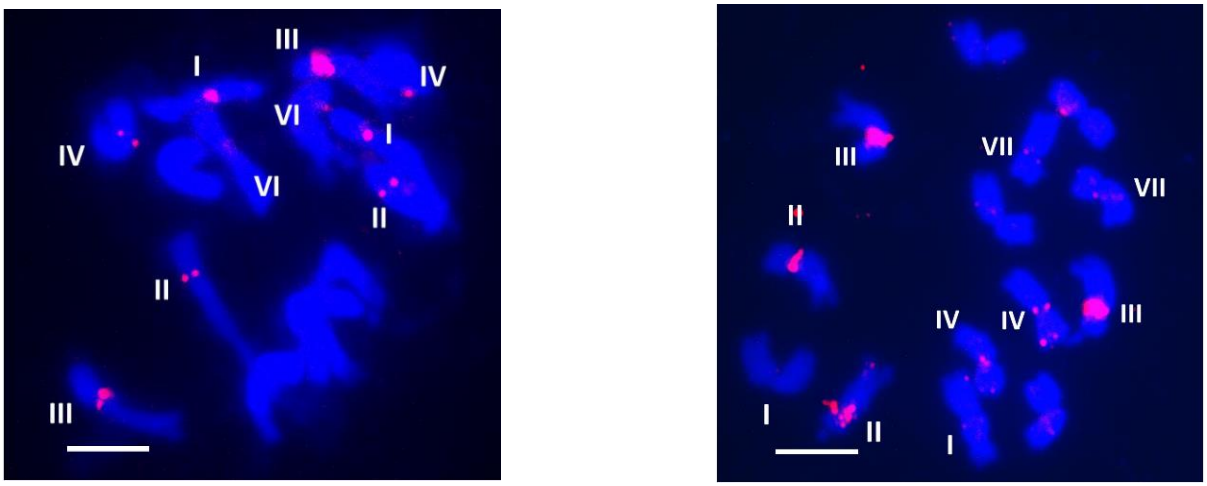


Figure 3.17 The main three fragments of PVCV, PVCV-M (3000 bp), PVCV-L (1100 bp), and PVCV-R (2400 bp) based on (Richert-Pöggeler *et al.* 2003), see Figure 3.6 and Table 3.4.

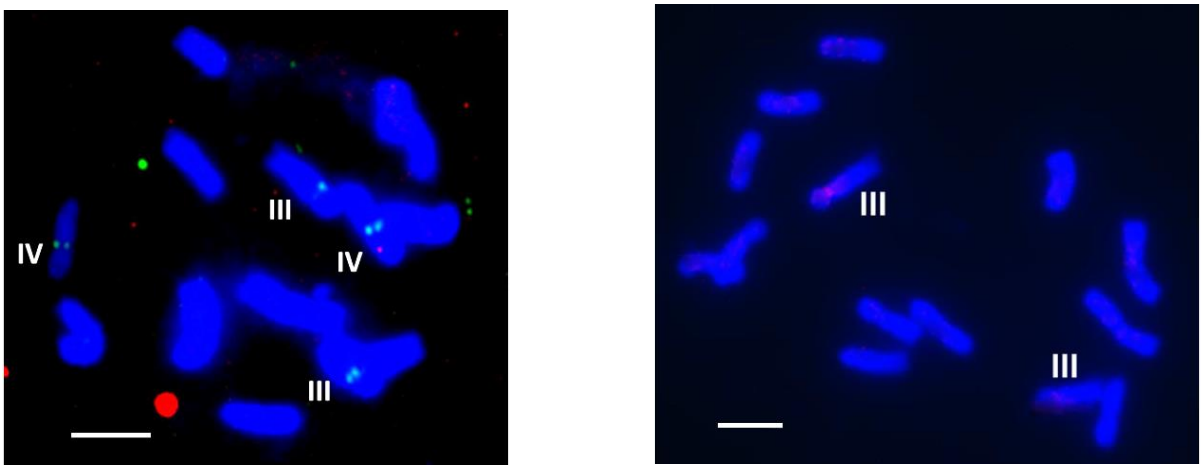
In *P. axillaris* subsp *axillaris* N, pericentromeric signals were seen on three pairs of chromosomes: a strong signal on Ch.III, lower strength signal on Ch.II, and a weak signal on Ch.VI. Hybrid plant chromosomes showed a distinct distribution for each species: *P. hybrida* Rdc shows nicely strong signals on the centromere of Ch. I, II, III, IV and VI with higher strength signal than others on Ch. III. In *P. hybrida* W138, five pairs of chromosomes clearly showed strong pericentromeric signals on Ch.III, lower strength on Ch.II, and weaker signals on Ch.IV, VII, and I, interestingly, however, Ch.IV shows telomeric signals on the short arm as well as the pericentromeric signal. Only two pairs of *P. hybrida* V26 chromosomes showed pericentromeric signals on Ch.III and IV. Very weak signals on the centromere of Ch.IV and III in *P. integrifolia* subsp *inflata* S6 and *P. axillaris* subsp *parodii* S7 respectively were sometimes detected. Ch.III in most species showed the hotspot of strong signals of chromosomal PVCV sequences (Figure 3.18).



A B



C D



E F

Figure 3.18 Signals of PVCV probe over all examined *Petunia* chromosomes, showing three pairs with signals (red) in *P. axillaris* subsp *axillaris* N (A), one pair (green) in *P. integrifolia* subsp *inflata* S6 (B), five pairs (red) in *P. hybrida* Rdc (C) and *P. hybrida* W138 (D), two pairs (green) in *P. hybrida* V26 (E), and one pair (red) in *P. axillaris* subsp *parodii* S7 (F). Chromosomes were stained with DAPI (blue) and probes were labelled with biotin 11-dUTP (detected in red) and digoxigenin-11-dUTP (detected in green). Bar = 10 μm.

3.4.1.2 Florendovirus-like cluster

This graph-based read cluster (CL131) includes 36.3% of reads with pararetrovirus homology (output from the Repeat Explorer). Further analysis of the reads showed a new genus of *Caulimoviridae* called *Florendovirus* (Geering *et al.* 2014) in this cluster at high copy numbers. The graph showed a circular pattern, with a butterfly half wing shape linked irregularly, having RT and RH domains in the middle of graph (Figure 3.19).

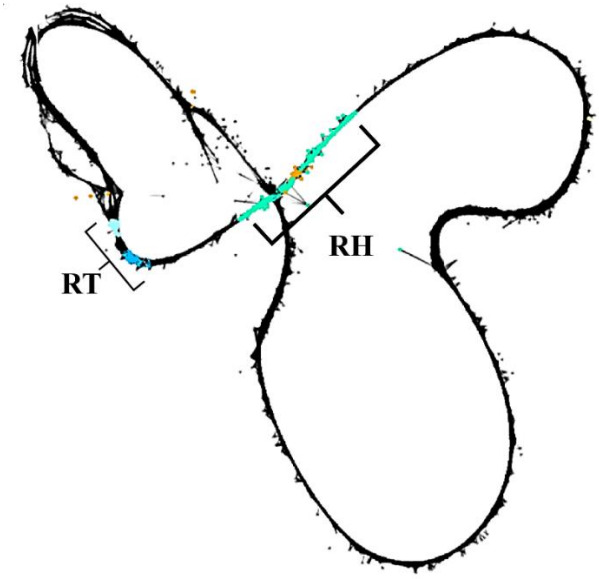


Figure 3.19 Florendovirus-like cluster of CL131, showing butterfly half wing shape linked with irregular circle, having RT (in blue) and RH (in green) domains in the middle of the graph.

Novel florendoviruses in *Petunia* species

Initially, all 76 entire sequences of *Florendovirus* members in (Geering *et al.* 2014) have been extracted and blasted against the parental genome assemblies; 74 of these sequences were found as fragmented parts with variable numbers of hits between *PaxiN* and *PinfS6*. Most of florendoviruses were registered in *PaxiN* with slightly higher numbers than *PinfS6*; two members were not found (OsatBV-compBsc1 and SbicV-compB) (Figure 3.20). From the RepeatExplorer, CL131 shared with most of florendoviruses in RT and RH domains with about 70% pairwise identity, although the entire sequence was different. To reveal the real identity of this cluster, the whole sequence was mapped to the whole raw reads of the four examined genomes, and the consensus sequence was extracted. After that, this sequence was blasted through NCBI and Repbase to find protein domains and open reading frames. Interestingly, in

PparS7 case, the extracted cluster from the RepeatExplorer was the entire sequence of *Florendovirus*. The new members of florendoviruses, one from each raw read set, were named following Geering *et al.* (2014). In *PaxiN*, the whole sequence length of its own *Florendovirus*, *PaxiV* is 7170 bp and has four open reading frames, and four conserved domains RH, RT, RVT and MP, also flanked by poly TA. The *PinfS6* member is *PinfV* with entire length 7142 bp and same protein domains as *PaxiV* with poly TA flanking regions and three open reading frames. The hybrid read has *PhybV* with 7519 bp and six open reading frames, while *PparV* in *PparS7* has 6519 bp and two open reading frames, and similar features as above (Figure 3.21; Appendices 3.4, 3.5, 3.6 and 3.7).

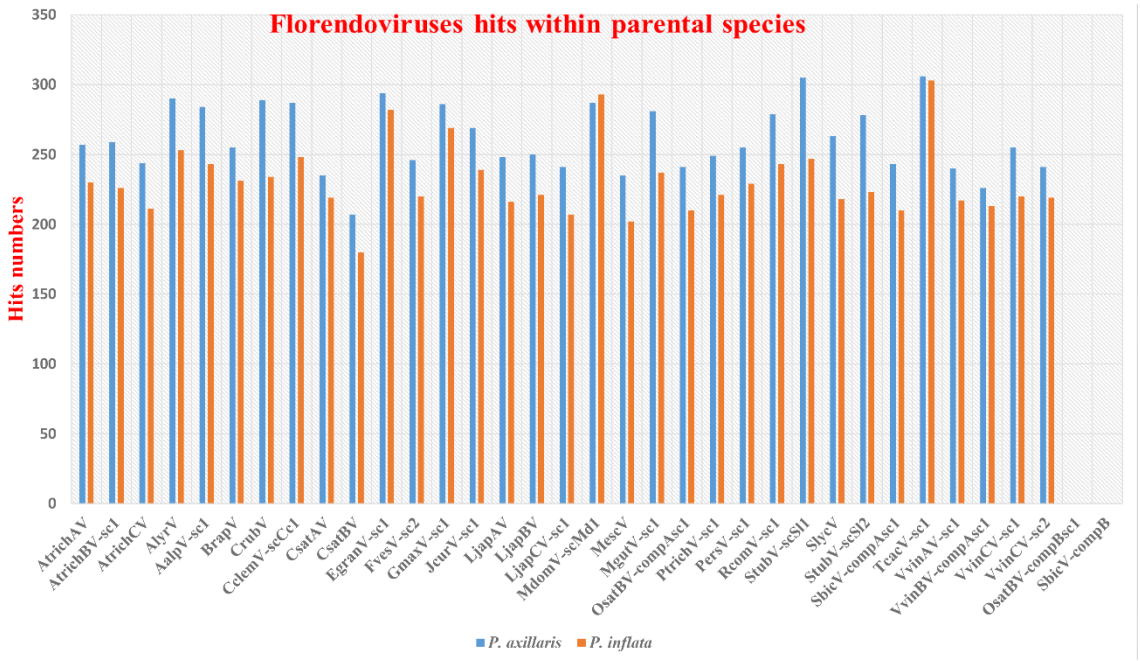


Figure 3.20 Hits of florendoviruses fragments over parental genomes *P. axillaris* N and *P. integrifolia* subsp *inflata* S6, showing *P. axillaris* subsp *axillaris* N has slightly higher numbers than *P. integrifolia* subsp *inflata* S6 except in MdomV-scMd1. The references *Florendovirus* types OsatBV-compBsc1 and SbicV-compB were not found in either genome.

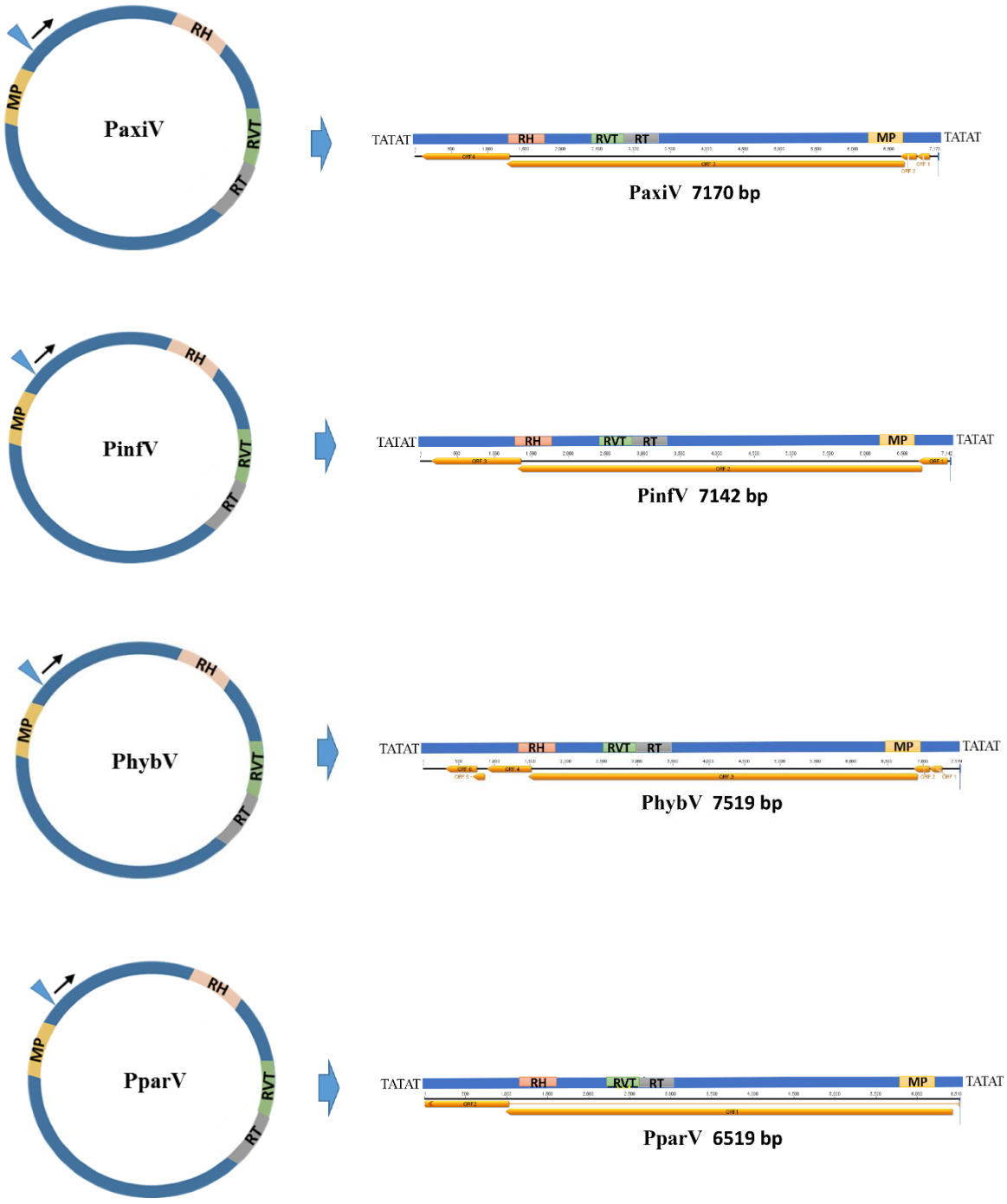


Figure 3.21 New sequences of florendoviruses in *Petunia* genome reads, *PaxiV*, *PinfV*, *PhybV* and *PparV* in *PaxiN*, *PinfS6*, *PhybR27* and *PparS7* respectively, having variable lengths and ORFs with inverted sequences. The circular forms on the left show putative exogenous forms with expected cutting side and the direction of integration (black arrow) before endogenisation event.

Genome proportion

Each *Florendovirus* member was used against the whole raw reads of its own genome to calculate genome proportion. Highly variable proportions within *Petunia* species have been registered, *PhybR27* shows most with 0.1331%, followed by *PaxiN* with 0.1012% and then *PparS7* 0.071% while the lowest one was *PinfS6* (0.0097%) (Figure 3.22).

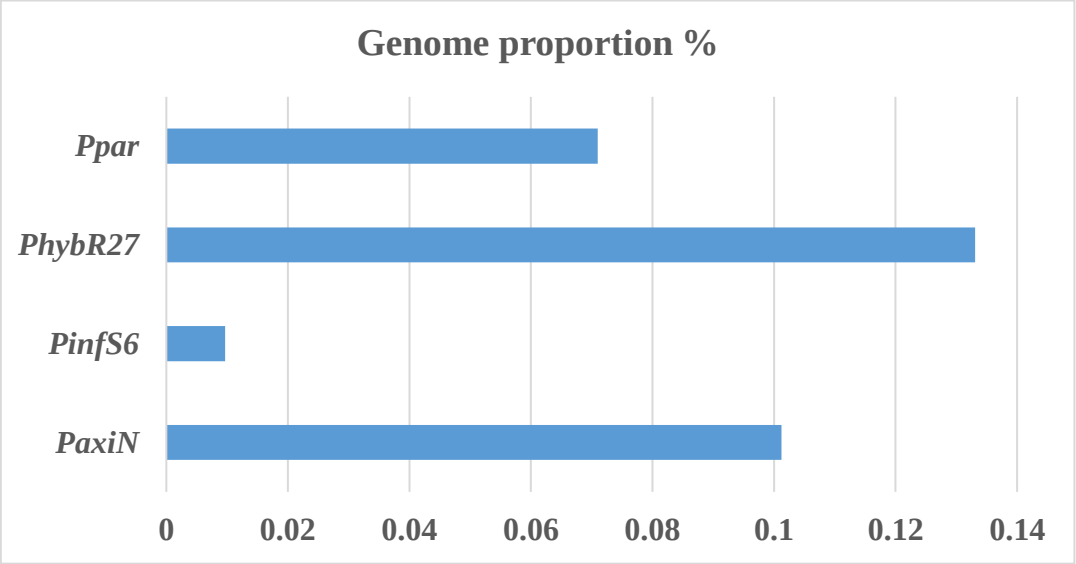


Figure 3.22 Genome proportions of florendovirus-like clusters over the examined *Petunia* species, showing higher ratio in *PhybR27*, then *PaxiN*, and *PparS7*, the lower proportion was in *PinfS6*.

Copy numbers

Florendovirus-like sequence integrated in *PhybR27* with highest number of copies (5018) that doubled *PaxiN* copies (2579), 1300 copies found in *PparS7* and only 176 copies were registered in *PinfS6* (Figure 3.23).

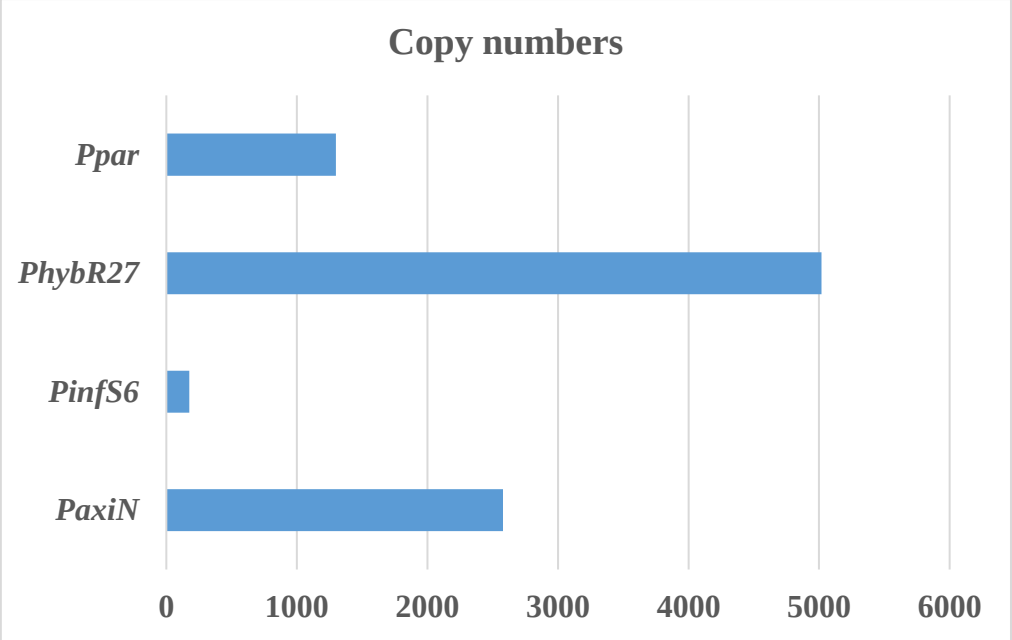


Figure 3.23 Copy numbers of florendovirus-like sequences in all used raw reads, showing higher copies in *PhybR27*, *PaxiN*, *PparS7*, and *PinfS6* respectively.

FISH

The amplified *Florendovirus* PCR product FlorB was cloned and then sequenced to confirm the sequence identity and purity (Figures 3.24 and 3.25). Next, the FlorB sequence was labelled and used as a probe for *in situ* hybridization. FlorB probe was not easy hybridizable over most of *Petunia* species that *in situ* was repeated many times to confirm results, and the most successful hybridizations were achieved for *P. axillaris* subsp *axillaris* N and *P. hybrida* Rdc. Clearly, pericentromeric signals have been identified at the centromere of five pairs of *P. axillaris* N chromosomes (Ch.II, III, IV, V and VI), and the strongest signal has spotted on Ch.III. In *P. hybrida* Rdc, pericentromeric signals have been spotted on five pairs (Ch.I, II, III, IV and VI), and like in *P. axillaris* N, Ch.III has shown high condensed signal on the centromere. Signals were not detected on *P. integrifolia* subsp *inflata* S6 and *P. axillaris* subsp *parodii* S7 chromosomes (Figure 3.26).

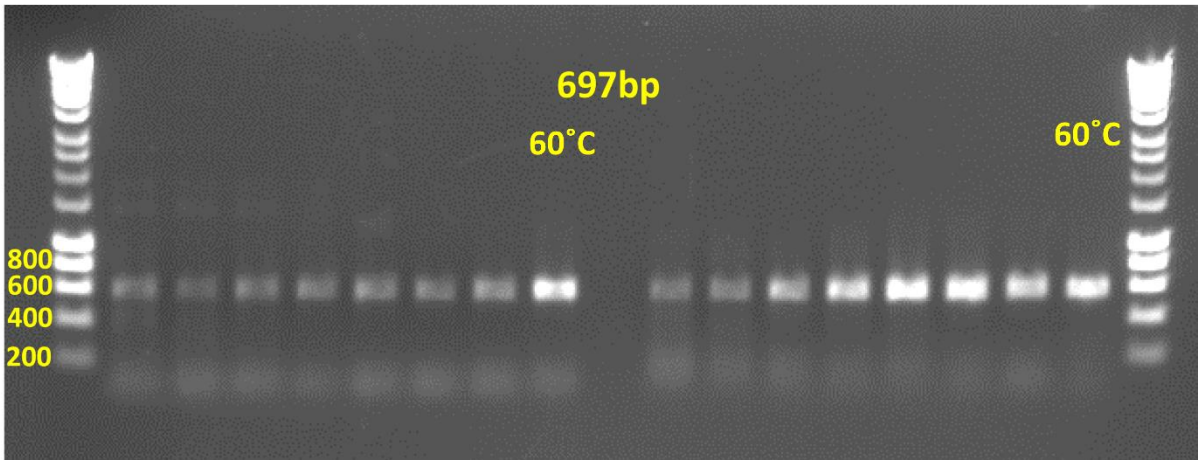


Figure 3.24 Amplified bands with multiple annealing temperatures using florB primer pair, a DNA template was for *P. axillaris* subsp *axillaris* N. Highest fragment concentration was gained at 60˚C, the length is 697 bp.

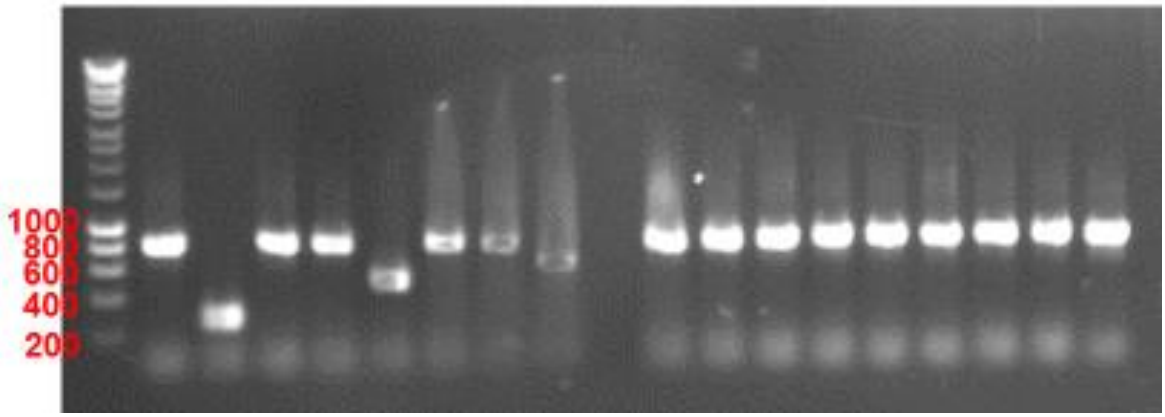
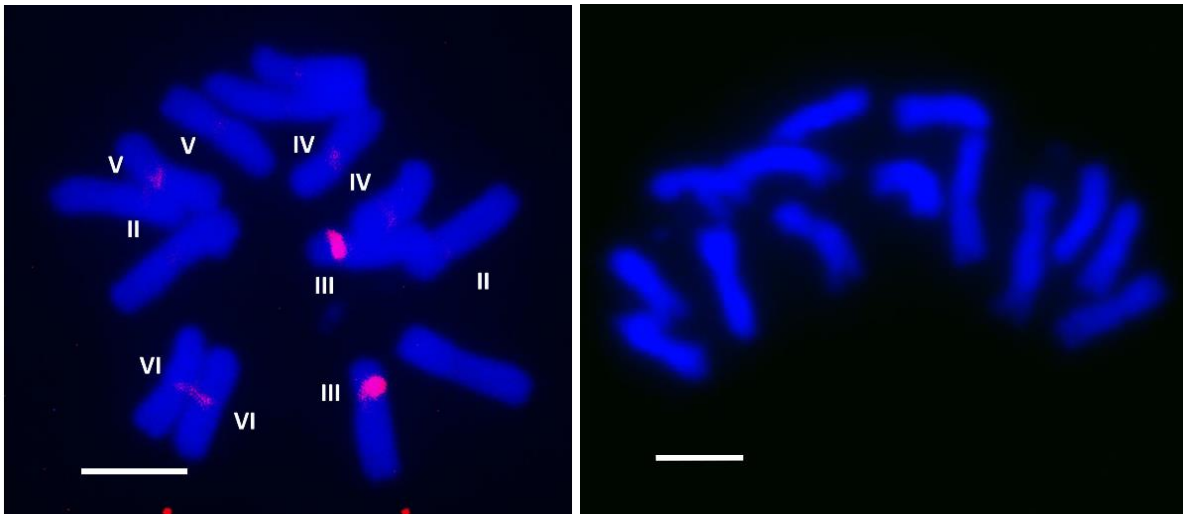
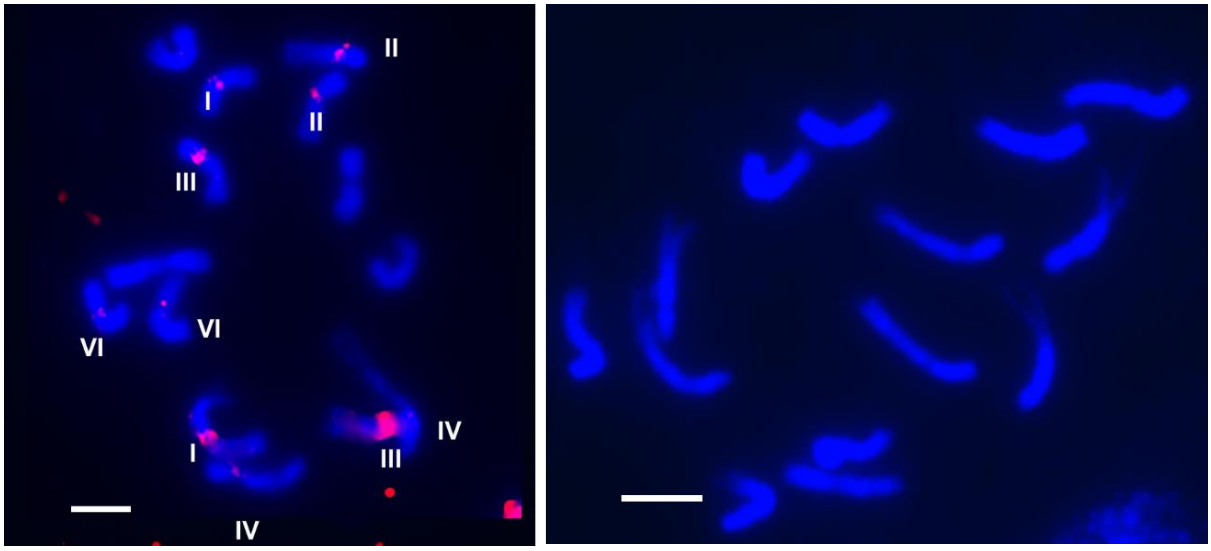


Figure 3.25 Amplification of cloned florB sequence at 55˚C. Product size is 903 bp, representing the insert plus M13 cloning site and primers (206 bp).



A B



C D

Figure 3.26 FISH of florB probe to metaphase of *P. axillaris* subsp *axillaris* N (A) and *P. hybrida* Rdc (C), showing signals at the centromere regions of five pairs of chromosomes, chromosome III has a stronger signal than others. No signals were seen in *P. integrifolia* subsp *inflata* S6 (B) and *P. axillaris* subsp *parodii* S7 (D). Chromosomes were stained with DAPI (blue) and the probe was labelled with biotin 11- dUTP (detected in red). Bar = 10 μm.

3.4.1.3 Caulimovirus-like cluster

Based on the RepeatExplorer pipeline, caulimovirus-like cluster (CL112) homology was identified in 27.2% of the reads. Graph based clustering showed a similar shape to the *Florendovirus* graph with opposite sizes of the graph parts, and RT and RH domains involved in the middle (Figure 3.27).

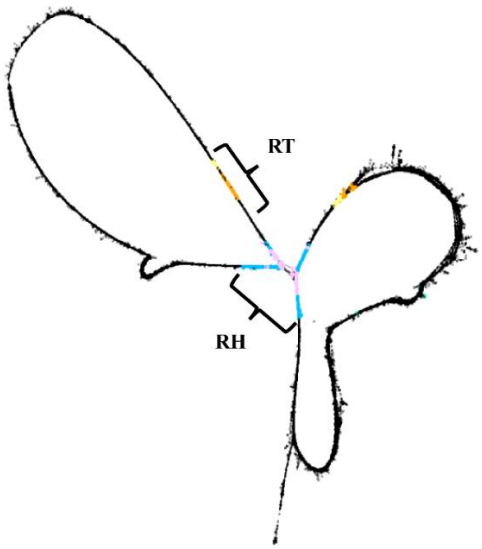


Figure 3.27 Caulimovirus-like cluster of CL112, showing butterfly half wing shape smaller than in florendovirus-like cluster linked with irregular circle, having RT (in yellow) and RH (in cyan) domains in the middle of graph.

Novel caulimoviruses in *Petunia* species

This cluster was preliminarily blasted through Repbase database after extraction from the RepeatExplorer (*PaxiN*), and a high similarity was found with integrated *Caulimovirus* members especially in RT and RH domains. After that, 'map to reference' tool was applied against the entire raw reads, and the consensus sequence extracted. As in florendoviruses, the whole sequence of *Caulimovirus* represented in the cluster of RepeatExplorer of *PparS7*. Obviously this consensus sequence has represented as new member of caulimoviruses in *P. axillaris* subsp *axillaris* N, and similar process has been achieved for other raw reads. Next, protein domains and ORFs were identified over the whole sequences. These sequences were named following Repbase dataset regulation that has big collection of integrated *Caulimovirus* members in a wide range of plant genomes. Accordingly, *Caulimovirus-PAx* represents caulimovirus-like sequence in *P. axillaris* subsp *axillaris* N (*PaxiN*), *Caulimovirus-PIn* in *P. integrifolia* subsp *inflata* S6 (*PinfS6*), *Caulimovirus-PHy* in *P. hybrida* (*PhybR27*), and *Caulimovirus-PPa* in *P. axillaris* subsp *parodii* S7 (*PparS7*). These virus-like sequences were found in *PaxiN*, *PinfS6*, *PhybR27* and *PparS7* with 7722, 8012, 8060 and 7999 bp lengths

respectively with a different series of protein domains. In *Caulimovirus-PAx* and *Caulimovirus-PPa*, protein domains start with MP, RT, RVT and RH with four and five ORFs respectively, while in *Caulimovirus-PIn*, begin with RH, MP, RT and RVT with eight open reading frames for all. Interestingly, these domains were inverted in *Caulimovirus-PHy* and start with RH, RT, RVT and MP, and has four open reading frames (Figure 3.28; Appendices 3.8, 3.9, 3.10 and 3.11).

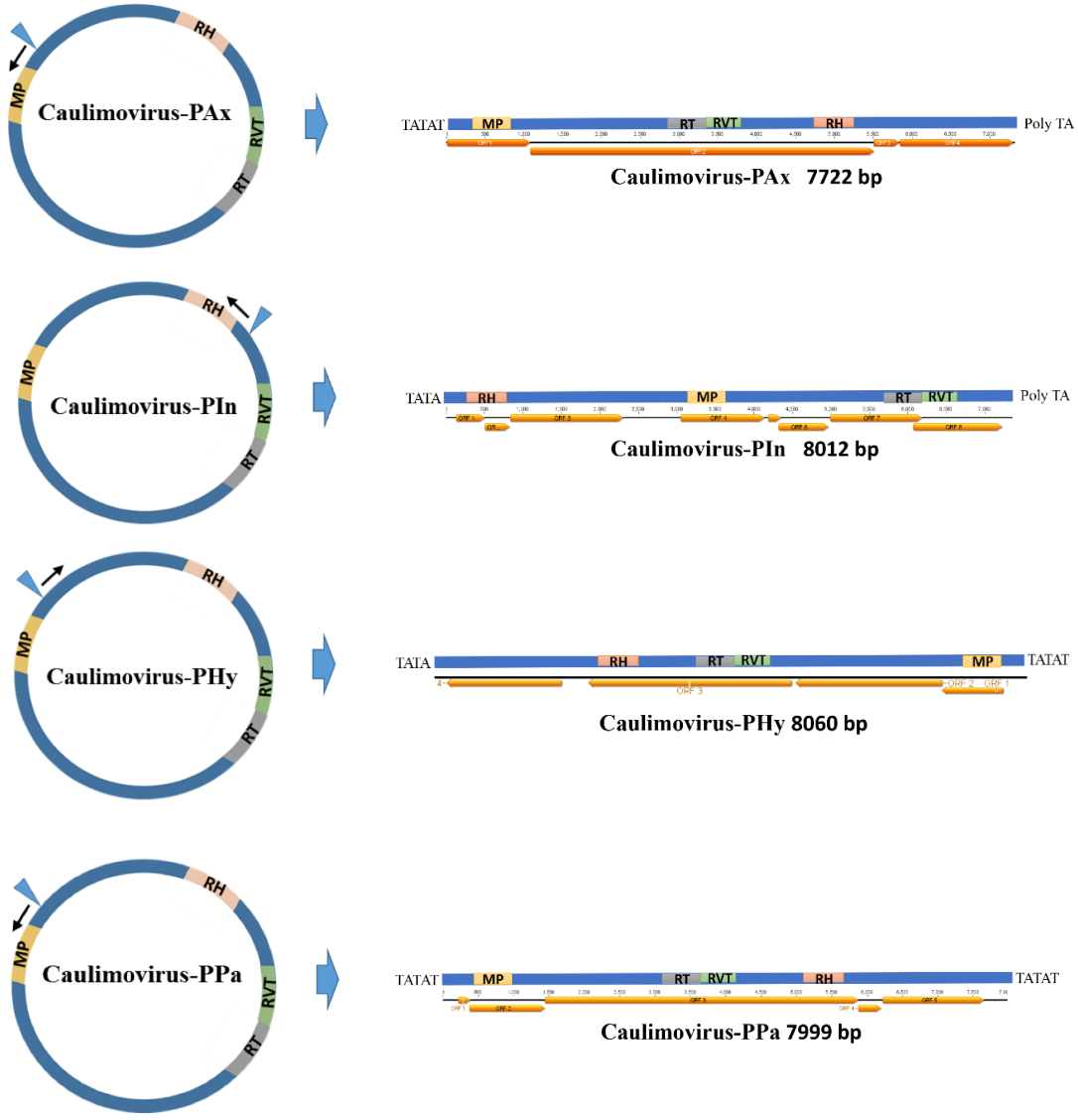


Figure 3.28 Novel members of caulimoviruses within *Petunia* genome reads, *Caulimovirus-PAx*, *Caulimovirus-PIn*, *Caulimovirus-PHy* and *Caulimovirus-PPa*, showing different lengths and domain distributions, and variable ORFs numbers. Interestingly, *Caulimovirus-PHy* has an inverted sequence unlike the others. The circular forms on the left show putative exogenous forms with suggested cutting side and the direction of integration (black arrow) before endogenisation event.

Genome proportion

Unlike the other two EPRV clusters, the caulimovirus–like cluster had a higher proportion (0.0279%) in *PparS7* than others, while *PaxiN* and *PhybR27* have approximately same ratio (0.015%), and the lowest proportion was in *PinfS6* (0.0062%) (Figure 3.29).

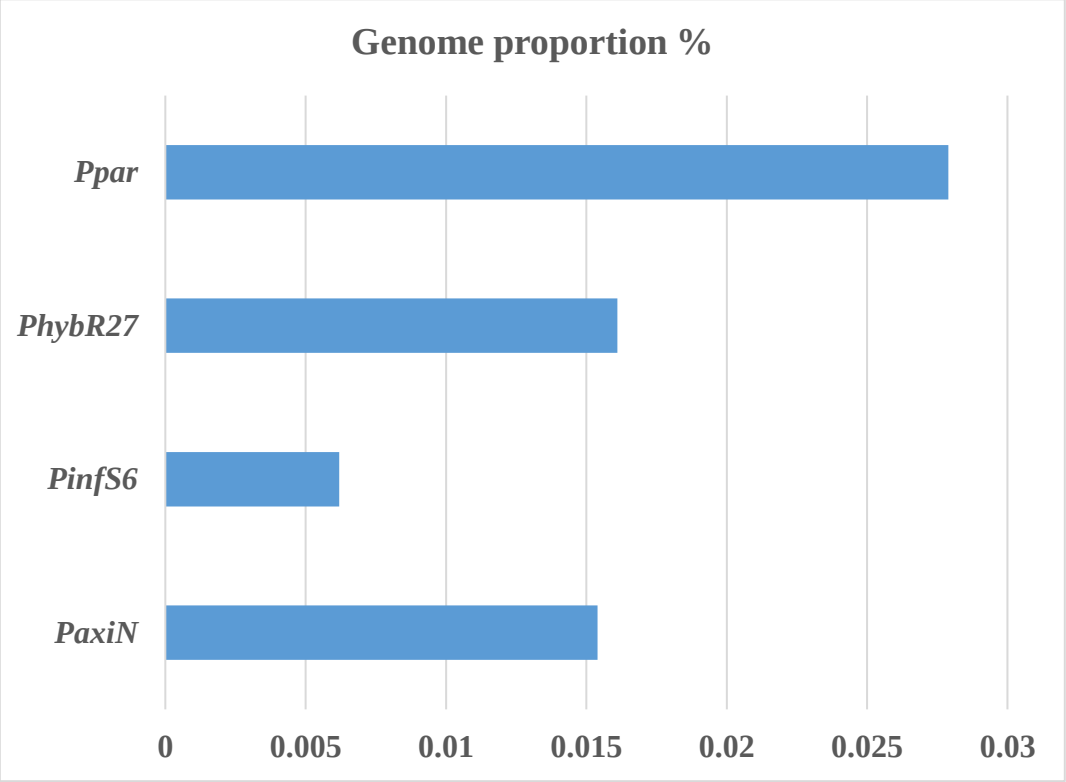


Figure 3.29 Genome proportions of caulimovirus-like clusters within the four petunia reads, showing higher ratio in *PparS7*, then *PaxiN* and *PhybR27* with similar ratio, the lower proportion was in *PinflS6*.

Copy numbers

Higher number of copies registered in *PhybR27* (3843) and then *PparS7* and *PaxiN* with 3237 and 2493 respectively, and like other EPRV, *PinfS6* has the lower copies (Figure 3.30).

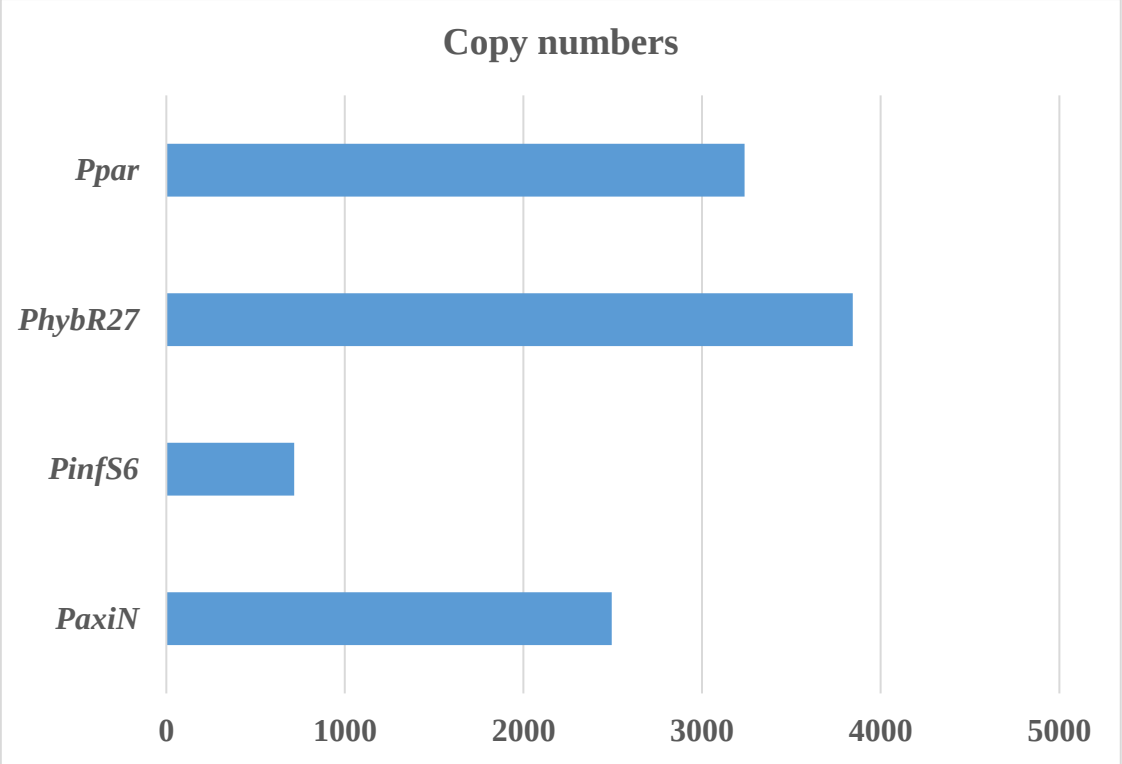


Figure 3.30 Copy numbers of caulimovirus-like clusters within the four petunia reads, showing higher number in *PhybR27*, then *PparS7* and *PaxiN*, the lower number was in *PinflS6*.

PCR amplification and *in situ* hybridization

The CL112 primers (1030 bp) amplified a product that was labelled for *in situ* hybridization to *Petunia* chromosomes (Figure 3.31). No hybridization signals were identified in any *Petunia* metaphases (Figure 3.32) using the protocol for localization of repetitive sequences at discrete loci.

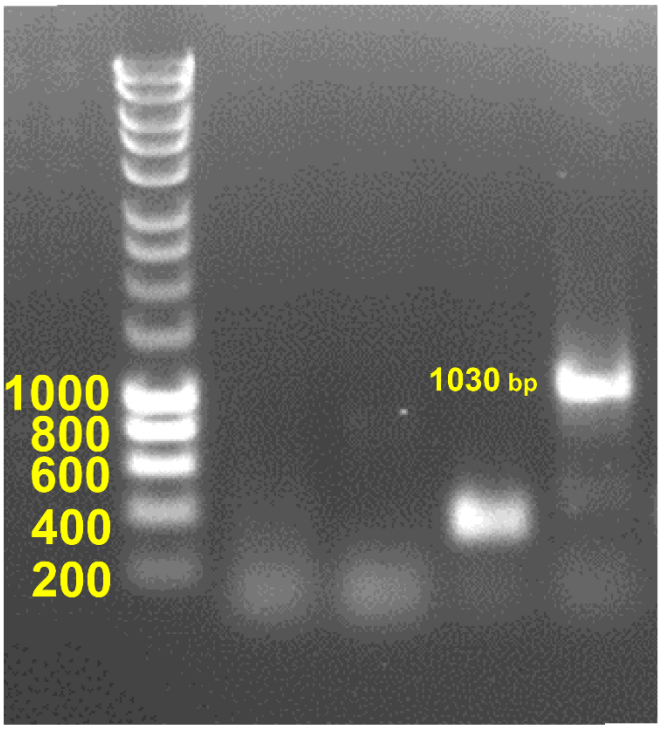


Figure 3.31 Amplified band of *Caulimovirus* primer (CL112) with 1030 bp at 60˚C.

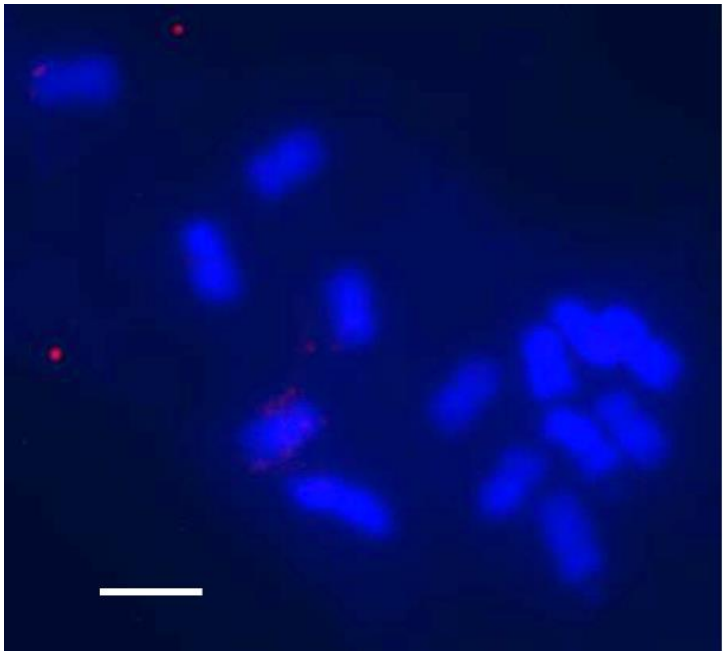


Figure 3.32 FISH image of caulimovirus-like cluster probe on *P. hybrida* Rdc metaphase, showing no spotted signals on chromosomes.

3.4.2 Phylogenetic relationships of EPRVs

The closely related EPRVs links have been confirmed through Bayesian phylogenetic tree of the whole EPRVs sequences (6500 bp each) that reconstructed after their alignments. The most closely related florendoviruses were *PhybV* and *PaxiV* (supported by posterior probability of 1), and both members have 0.61 support with *PparV*, then those three have 1 support with *PinfV*. A similar story was repeated with caulimoviruses: *Caulimovirus-PHy* shows high similarity with *Caulimovirus-PAx* (1 support), then the two linked with *Caulimovirus-PPa* (1 support), and all related to *Caulimovirus-PIn* with 1 support. The outgroup viruses were PVCV and CaMV, and the latter was closer to the EPRVs than PVCV especially in the region of RT and RH domains (Figure 3.33).

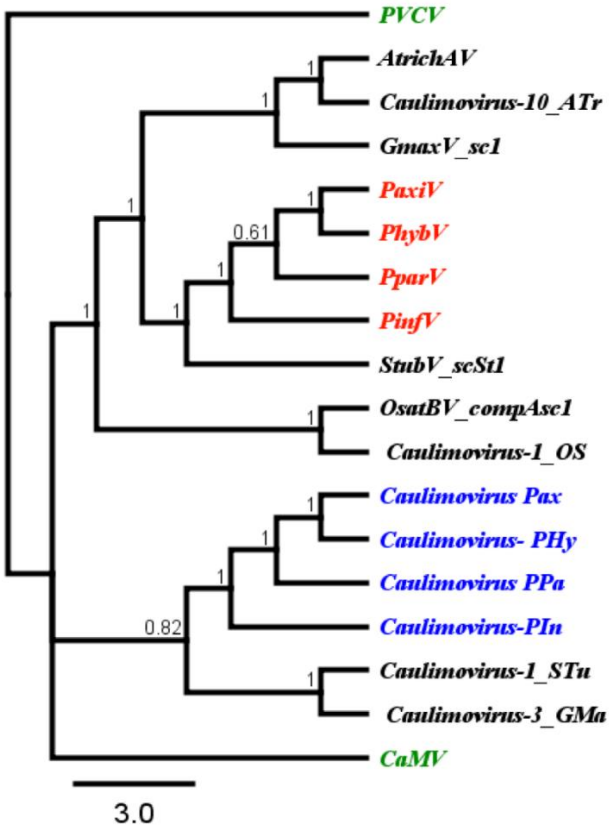


Figure 3.33 Bayesian phylogenetic tree of a group of caulimoviruses and florendoviruses in comparison with PVCV and CaMV using 18 entire sequences (about 6500 bp each), showing high relationship between *Florendovirus* members (in red), and *Caulimovirus* members (in blue) in *Petunia* genomes. The EPRVs of florendoviruses and caulimoviruses from *Amborella trichopoda* and *Oryza sativa* show closely related sequences due to host impact unlike EPRVs from *Solanum tuberosum* and *Glycine max*. The sequences were aligned by ClustalW alignment and then tree was reconstructed through MrBayes inference. The number next each branch is for posterior probability (> 0.5 support).

3.4.3 *De novo* integration

3.4.3.1 DNA extraction

Total DNA comprising genomic and potential episomal DNA of different *Petunia* species was extracted from leaf tissues. Quantity and quality of the samples were analysed and determined by spectrophotometry and gel electrophoresis. For all *Petunia* samples (see Table 3.5), high molecular weight DNAs (above 10 kb) were collected (Figure 3.34), genomic DNA from all samples was of good quality with a 260/280 ratio between 1.70 to 2.10 and the concentrations were about 30 to 800 ng/µl. The differences in the DNA quantity were due to different amount of leaves from each sample as well as plant age. All plants have been registered and submitted to TEM inspection for presence of PVCV virions or other viral contamination.

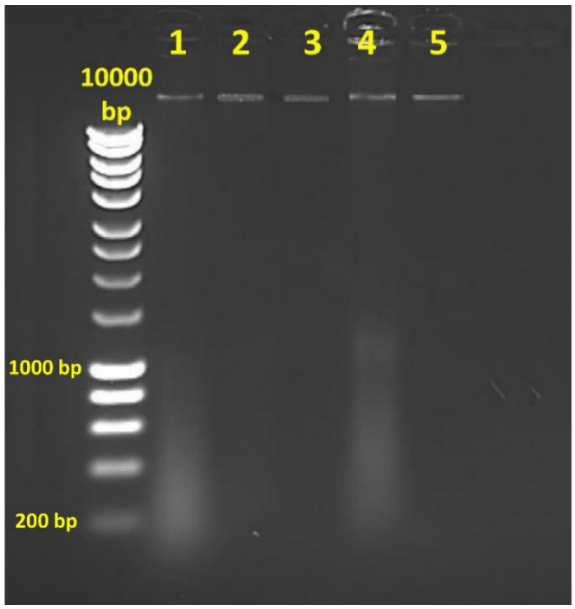


Figure 3.34 1% agarose gel of extracted DNA from *P. axillaris* subsp *axillaris* N (mixed sample from plants numbered TEM16-124_1 to _6, lane) (1), *P. integrifolia* subsp *inflata* S6 TEM15-652 (2) PVCV infected *P. axillaris* subsp *parodii* S7 TEM16-35_1 (3), *P. axillaris* subsp *parodii* S7 (mixed sample from plants TEM16-315_1to _3) (4) and progeny of PVCV infected *P. axillaris* subsp *parodii* S7 EM09-348 (5) by loading 10 µL purified DNA + 3 µL loading Buffer (5X). 10 µl of 1kb hyper ladder (Bioline), run on each side of the gel.

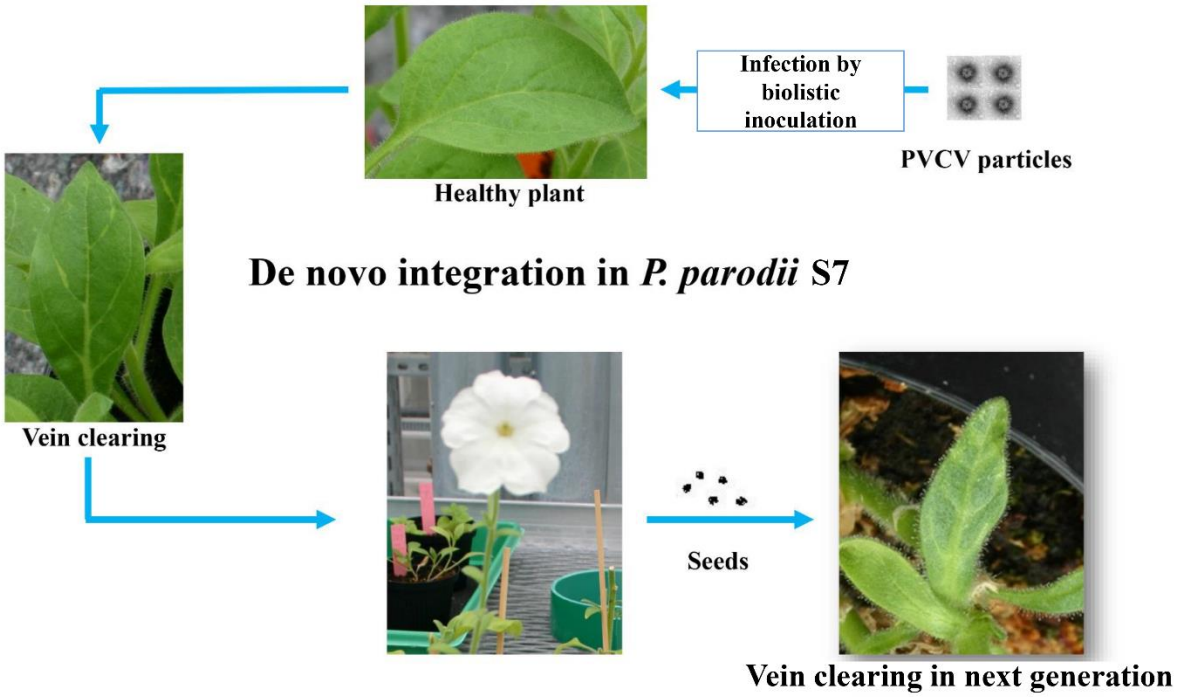


Figure 3.35 Model of horizontal followed by vertical transmission of PVCV in *P. axillaris* subsp *parodii* S7. Healthy plants can be infected with PVCV via biolistic inoculation (horizontal transmission). Vegetative propagation using cuttings probably promotes *de novo* integration of PVCV sequences into the *P. axillaris* subsp *parodii* S7 genome. Thus the virus gets vertically transmitted over generations by seeds.

3.4.3.2 RNA extraction

PCR for detection of episomal PVCV replication and activation of other pararetroviral-like sequences is challenging on DNA templates because of the presence of chromosomal viral copies which can also serve as template during PCR. RNA transcription is a first step in retroelement replication, therefore, total RNA extracts were prepared from equal weights of the selected five samples, including a DNase incubation to remove chromosomal DNA contamination in the final RNA sample. RNA concentration and quality were determined using a spectrophotometer and gel electrophoresis (Figure 3.36). The obtained RNA was variable, and samples with lower RNA content showed a higher degree of RNA degradation (see Table 3.9):

Table 3.9 RNA concentrations of the five petunias samples.

Samples		Registration codes	RNA (ng/µl)
i	*P. axillaris* subsp *axillaris* N	TEM16-124_1-6	141,66
ii	*P. integrifolia* subsp *inflata* S6	TEM15-652	314,36
iii	*P. axillaris* subsp *parodii* S7 + PVCV	TEM16-35_1	107,98
iv	*P. axillaris* subsp *parodii* S7	TEM16-315_1-3	416,50
v	Progeny of *P. axillaris* subsp *parodii* S7 + PVCV	EM09-348	252,72

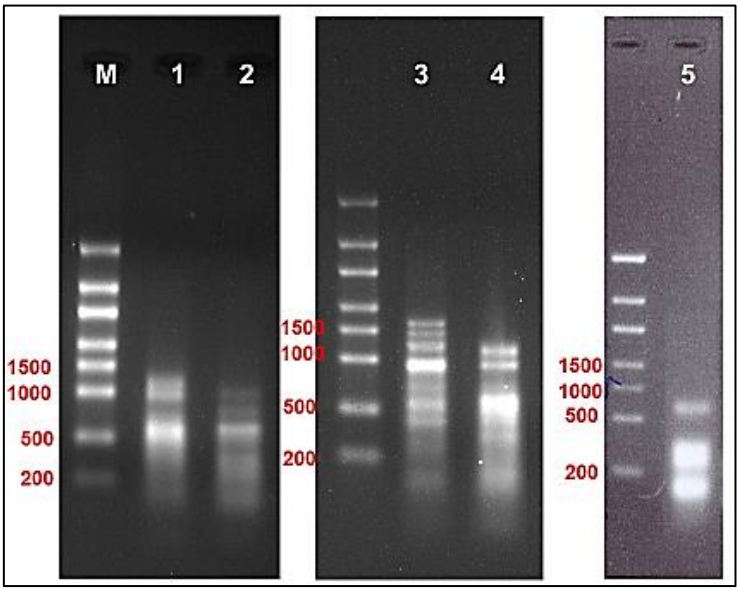


Figure 3.36 RNA extraction of *P. axillaris* subsp *axillaris* N (mixed sample from plants numbered TEM16-124_1 to _6, lane) (1), *P. integrifolia* subsp *inflata* S6 TEM15-652 (2) PVCV infected *P. axillaris* subsp *parodii* S7 TEM16-35_1 (3), *P. axillaris* subsp *parodii* S7 (mixed sample from plants TEM16-315_1to _3) (4) and progeny of PVCV infected *P. axillaris* subsp *parodii* S7 EM09-348 (5) by loading 5 µL purified RNA + 5 µL loading dye (2X). 10µl of RNA ladder (2µl RNA Ladder, 5µl RNA loading dye (2X) and 3µl H_20), run on the left lane (M) of each gel.

3.4.3.3 Reverse Transcriptase-PCR

Complementary DNAs were synthesized, and electrophoresed for the five RNA samples (Figure 3.37). All cDNA bands were between 250 and 1000 bp. cDNAs were used as templates for PCR using PVCV F16a-R16b (483 bp), *Florendovirus* florB (697 bp), *Caulimovirus* CL112 (1030 bp) primer pairs (Figure 3.38).

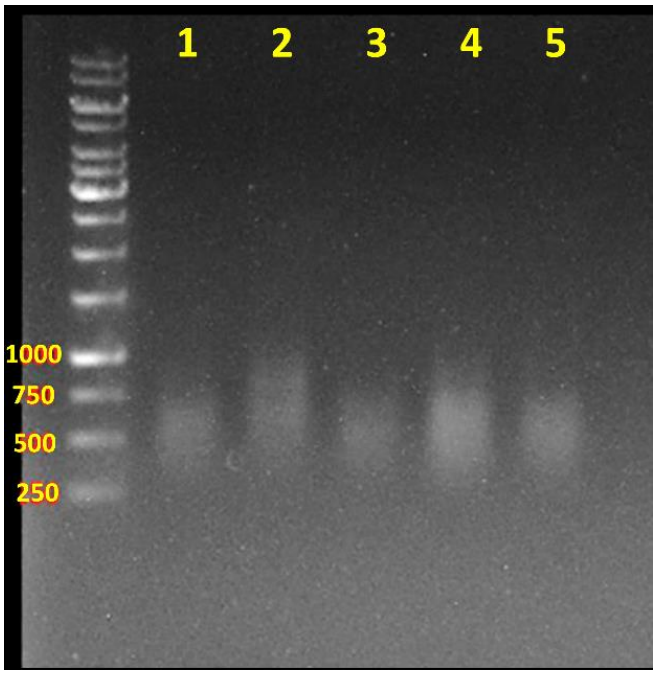


Figure 3.37 cDNA synthesis of *P. axillaris* subsp *axillaris* N (mixed sample from plants numbered TEM16-124_1 to _6, lane) (1), *P. integrifolia* subsp *inflata* S6 TEM15-652 (2) PVCV infected *P. axillaris* subsp *parodii* S7 TEM16-35_1 (3), *P. axillaris* subsp *parodii* S7 (mixed sample from plants TEM16-315_1to _3) (4) and progeny of PVCV infected *P. axillaris* subsp *parodii* S7 EM09-348 (5) with Omniscript Reverse Transcription Kit. Gel was run by loading 5 µl cDNA, 5µl H_2O and 2µl DNA loading dye (6X). 6µl of DNA Ladder (1µl DNA Ladder, 1µl 6X DNA loading dye and 4µl H_2O).

Although the PCR was not quantitative, there were suggestive differences in product amount between samples. For PVCV primer, the strongest band was produced with sample No.3 (*P. axillaris* subsp *parodii* S7 + PVCV), known to be episomally infected by PVCV. Weaker bands were found in samples No.5 (progeny of *P. axillaris* subsp *parodii* S7 + PVCV) and No.1 (*P. axillaris* N), very light bands were amplified for samples No.2 (*P. integrifolia* subsp *inflata* S6) and No.4 (*P. axillaris* subsp *parodii* S7). The faint PCR products in the wild *Petunia*

species may originate from residual DNA in the RNA sample. A control used directly extracted RNA as a template for PCR without the reverse transcription step. The PCR product obtained in *P. axillaris* subsp *parodii* S7 has confirmed earlier results in 3.4.1.1 as this wild *Petunia* was suggested to be free of any forms of PVCV before these experiments.

CL112 primer of *Caulimovirus* has amplified strongly with samples No.3 (*P. axillaris* subsp *parodii* S7 + PVCV) and No.4 (*P. axillaris* subsp *parodii* S7), and then sample No.5 (*P. axillaris* subsp *parodii* S7 + PVCV), light band for sample No.1 (*P. axillaris* subsp *axillaris* N) and very light band for sample No.2 (*P. integrifolia* subsp *inflata* S6).

Using the florB primer for *Florendovirus* sequence, the bands were often light for samples No.3, 4 and 5 even after amplifying them again using their PCR products as templates while very light bands were identified for samples No.1 and 2 (Figure 3.38).

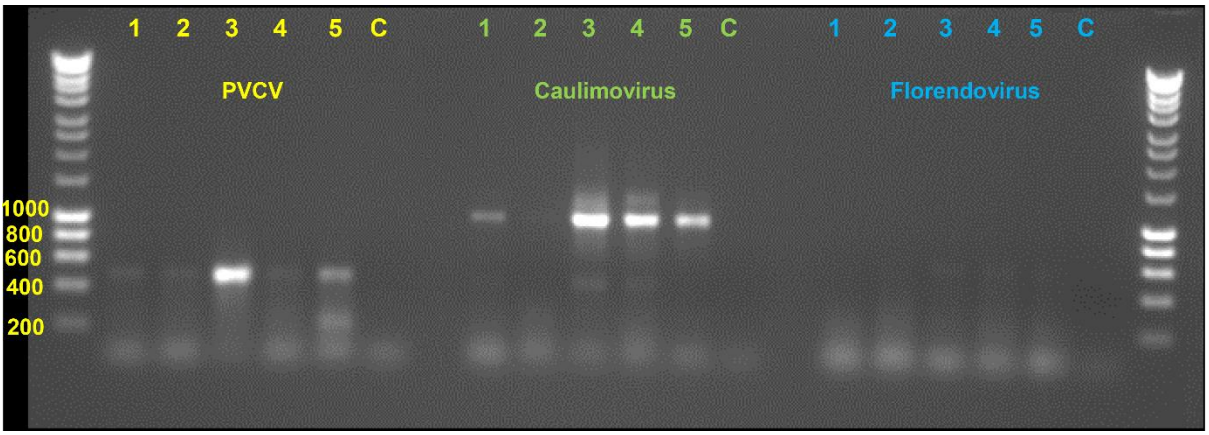


Figure 3.38 PCR of *P. axillaris* subsp *axillaris* N (mixed sample from plants numbered TEM16-124_1 to _6, lane) (1), *P. integrifolia* subsp *inflata* S6 TEM15-652 (2) PVCV infected *P. axillaris* subsp *parodii* S7 TEM16-35_1 (3), *P. axillaris* subsp *parodii* S7 (mixed sample from plants TEM16-315_1to _3) (4) and progeny of PVCV infected *P. axillaris* subsp *parodii* S7 EM09-348 (5) cDNAs as templates and primer pairs of PVCV(F16a+R16b), *Caulimovirus* (CL112) and *Florendovirus* (florB).

3.4.3.4 *In situ* hybridization

From tissue culture, *de novo* integrated PVCV sample with vein clearing symptom that was infected horizontally in 2001 by grafting and then maintained in a tissue culture condition was

checked and analysed. Further, produced seeds from this particular plant as vertically transmitted and *de novo* integrated virus have been examined with same protocol (Figure 3.35). In order to find out *de novo* integration sites of PVCV over host chromosomes, PVCV probe has been applied against progeny of PVCV infected *P. axillaris* subsp *parodii* S7 EM09-348 metaphases. Findings show telomeric signals of integrated virus on the telomere of heterologous chromosomes with variable strengths on short or long arms. In some cases, signal was shown on the telomere of one individual chromosome or two heterologous chromosomes, and signal strength was very concentrated with multiple dots or too weak with only two dots around the telomere. For instance, in Figure 3.39 fluorescent signals of PVCV were located on the short and long arms of Ch.III and IV chromosomes respectively, with a weak signal on the short arm of Ch.III and too condensed dots on the long arm of Ch.IV. Typical signal was registered as two dots on the telomere follow by a dotted ring around the telomere region (Figure 3.40). In contrast, signal features and loci of *de novo* integrated PVCV in host chromosomes are completely different from locations of the chromosomal PVCV in all *Petunia* chromosomes where it appeared as only two dots in the centromere or telomere regions over only homologous chromosomes (Figure 3.18).

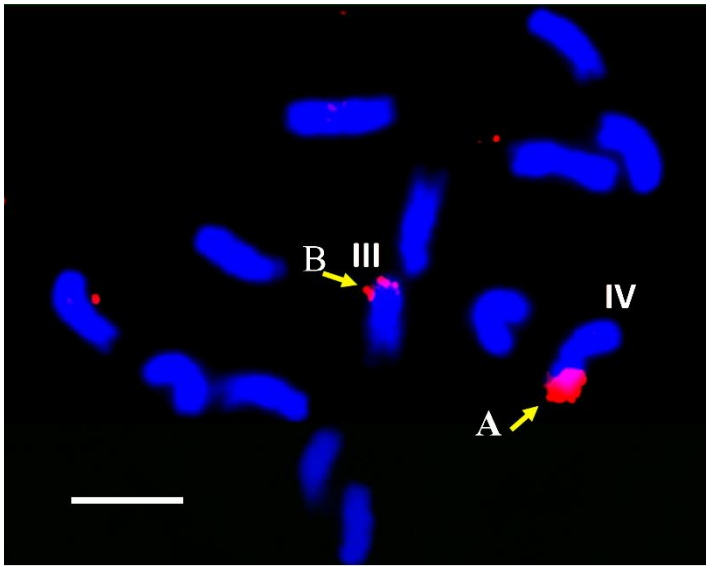


Figure 3.39 PVCV signals over next generation of PVCV infected *P. axillaris* subsp *parodii* S7 EM09-348 metaphases that show two signals with variable strength on the long arm (A) of Ch.IV and short arm (B) of Ch.III. Chromosomes were stained with DAPI (blue) and probe was labelled with biotin-11- dUTP (detected in red). Bar = 10 µm.

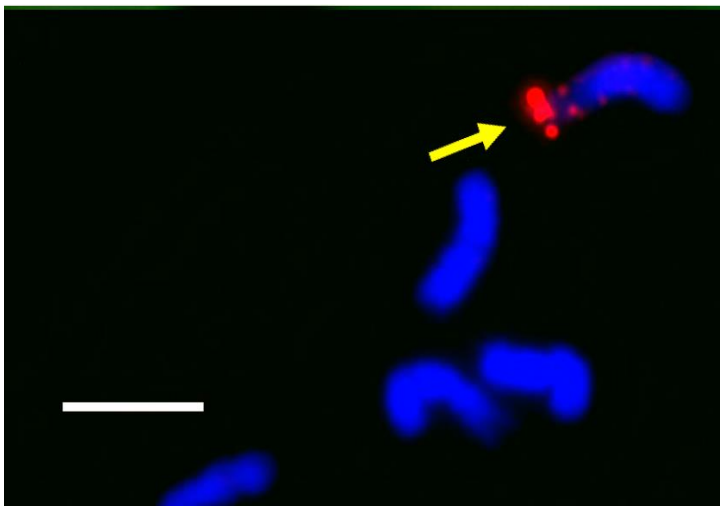


Figure 3.40 Typical signal of integrated PVCV on the long arm of the telomere region of progeny of PVCV infected *P. axillaris* subsp *parodii* S7 chromosome, showing an unusual hybridization pattern with two dots and a halo of dots looks like a ring around the telomere. Chromosomes were stained with DAPI (blue) and probe was labelled with biotin 11-dUTP (detected in red). Bar = 10 μm.

3.4.3.5 *De novo* assembly

In order to confirm the results of *in situ* hybridization above of *de novo* integrated PVCV, *de novo* assembly for the whole raw reads of infected *P. axillaris* subsp *parodii* S7 was applied. Then, the constructed consensus sequences were mapped to PVCV sequence to check if there is any telomeric sequence (TTTAGGG) near the virus sequence. After that, the mapped contig was extracted and then aligned against PVCV sequence. The result shows that some contigs have telomere sequence incorporated within a recombinant PVCV while other contigs have telomere sequences around the integrated virus sequence (Figure 3.41). More importantly, PVCV has two regions of telomeric repeat TTAGGG (mammalian-type) in at the start (590-595 bp) and end (6214-6219 bp) of the entire length.

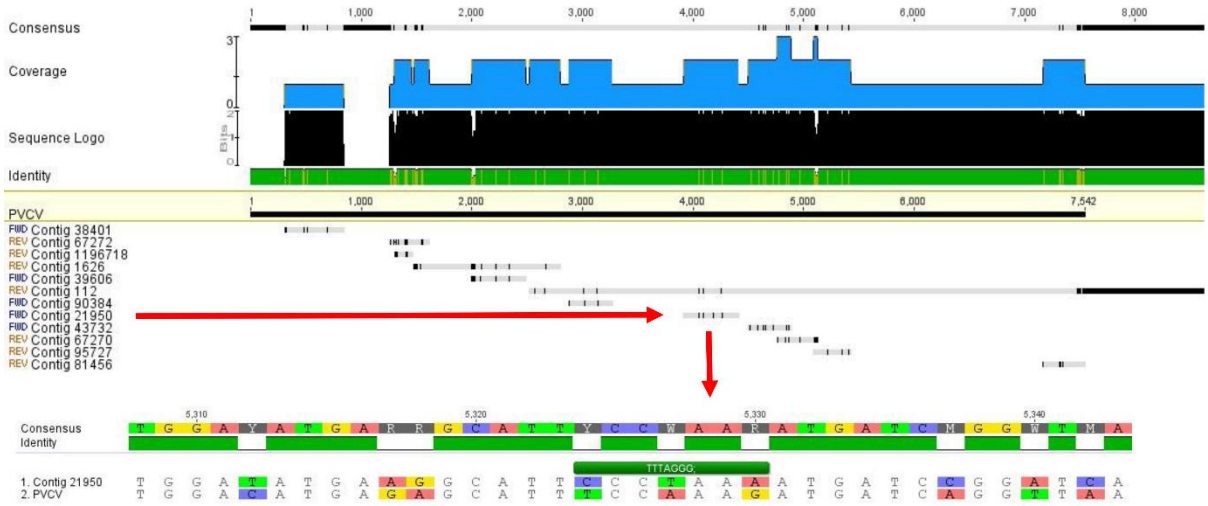


Figure 3.41 The mapped and aligned contig to PVCV shows contig 21950 extracted from *de novo* assembly has reverse complement telomere sequence (CCCTAAA) next to recombinant PVCV sequence.

3.5 Discussion

3.5.1 PVCV fragments in all examined *Petunia* species

Findings of this project strongly support the importance of endogenous viral sequences in plant genomes, and provided further evidence of the integration diversity between plant viruses and their hosts. Although the integrated sequences are inherited over generations and many are degenerated, some sequences are relatively intact, transcriptionally efficient and could be infectious under particular conditions (Geering 2001; Geering *et al.* 2010; Geering *et al.* 2014). The revolution of sequencing techniques confirmed that some plant genomes have been infected by some *Caulimoviridae* members especially in the Solanaceae that were reported as appropriate hosts for a wide number of plant viruses (Engelmann & Hamacher 2008; Geering *et al.* 2014; Kim *et al.* 2014). Endogenisation searches have been extended successfully within *Petunia* genomes, using bioinformatics tools and PCR amplification. According to the results, virus sequence has been found with full length in the hybrid genomes of *Petunia* with respect to some shorter lengths or degenerated copies. Although PVCV has been naturally fragmented in the genome of *Petunia*, necessarily conserved protein domains are still active and capable to make infection under stress conditions (Richert-Pöggeler *et al.* 2003; Bombarely *et al.* 2016). In the wild species, *P. axillaris* subsp *axillaris* N has similar length to PVCV in the hybrids

and only the last part was degenerated. While *P. axillaris* subsp *parodii* S7 was reported in Bombarely *et al.* (2016) as missing all PVCV fragments, this work has shown the existence of virus parts with some missing fragments. The virus in the other parent *P. integrifolia* subsp *inflata* S6 was found degenerated over the whole sequence (so primer homologies were probably the reason for failure of amplification of some regions). The degenerated nucleotides of the last part of chromosomal PVCV in most *Petunia* genomes enables primers from this region to act as markers to differentiate integrated virus copies from those copies of the exogenous form (Figures 3.14, 3.15 and 3.16). FISH results have confirmed variable existence of virus sequences over *Petunia* species: strong signals have been captured in the hybrids and *P. axillaris* subsp *axillaris* N, in contrast to very weak signal in one pair of *P. integrifolia* subsp *inflata* S6 and *P. axillaris* subsp *parodii* S7 chromosomes. In *P. axillaris* subsp *axillaris* N, one more pair of chromosomes (Ch.II) were captured with weak pericentromeric signals adding this chromosome to Bombarely *et al.* (2016) who showed only two pairs of *P. axillaris* N chromosomes with strong signals of PVCV (Figure 3.18). The existence of EPRVs may also play a role in genome diversification by providing coding or transcription regulatory elements as new genetic components.

3.5.2 PVCV as a cytogenetic marker

Chromosomal signals of PVCV in each species enabled recognition of *Petunia* species and chromosomes according to their signal positions, showing EPRV is a useful cytogenetic marker (Figure 3.18). For retrotransposons, their widespread distribution within chromosomes, high copy numbers, in addition to the ubiquitous nature, have made these components quite typical for developing of the DNA based markers (Flavell *et al.* 1992; Teo *et al.* 2005). As very closely related to retrotransposons, endogenous pararetroviruses (EPRVs) probably integrate within host genomes by hitchhiking on retrotransposons and the hybrid of the two types is eligible of real transposition might be made through this path (Hohn 1994; Richert-Pöggeler *et al.* 2003; Gregor *et al.* 2004; Froissart *et al.* 2005; Staginnus *et al.* 2007; Richert-Pöggeler & Schwarzacher 2009). Additionally, in the context of diversity among *Petunia* genomes, integration of PVCV has revealed evolution and phylogeny approaches over *Petunia* species, and clarified the real relationships between wild and hybrid hosts based on EPRV.

3.5.3 Florendovirus-like sequences

The florendovirus-like clusters in *Petunia* species show fragmented parts of the 74 members of florendoviruses (Geering *et al.* 2014) in the two parents assemblies (*PaxiN* and *PinfS6*) (Figure 3.20). Conversely to the PVCV case, these fragments registered in *P. axillaris* subsp *parodii* S7 with higher proportions and copies than in *P. integrifolia* subsp *inflata S6.* Interestingly, four novel florendoviruses were recorded as one virus-like sequence in raw reads from each genome, adding petunia to the list of 27 host species invaded by these viral units (Figure 3.21; Appendices 3.4-3.7). As Geering *et al.* (2014) reported about the phylogenetic link between the two sister groups of *Florendovirus* and PVCV, new members in petunia have shared the plesiomorphic trait particularly of RT and RH with PVCV. However, those two could be distinguished by the existence of more than one ORF in florendoviruses unlike PVCV that has one ORF. The results of FISH revealed condensed signal also in the centromere of Ch.III of *P. axillaris* subsp *axillaris* N and *P. hybrida* Rdc in addition to weaker signals in other chromosomes (Figure 3.26).

In this project, our findings have proved Geering *et al.* (2014) suggestion about integration loci of *Florendovirus* sequences in TE-rich regions of plant chromosomes flanked by TA dinucleotides repeats. Results did not particularly agree with Geering *et al.* (2014) reporting about 9% of *Florendovirus* positions within introns of plant genes based on data from *Vitis vinifera* genome suggesting biological effects of these elements on gene transcripts and expression levels. Here, inserted *Florendovirus* motifs accumulated in heterochromatin regions in fragile sites of poly-TA simple repeats, probably including secondary structures like hairpins and therefore avoiding chromosome fragility (Zlotorynski *et al.* 2003; Dillon *et al.* 2013). These elements could act as fillers to repair DNA breaks either through microhomology-mediated end or non-homologous end joining (Huertas 2010).

3.5.4 Hotspot of Chromosome III

Chromosome III in particular shows stronger pericentromeric signal in *P. axillaris* subsp *axillaris* N, the hybrid species and weaker signal in *P. axillaris* subsp *parodii* S7, and this might reflect a PVCV hotspot inside this chromosome. It probably happened due to high rates of recombination hotspots and frequencies within this chromosome that resulted from elevated DNA break formation. Croll *et al.* (2015) found that recombination rate frequency for each

chromosome was inversely linked to the chromosome length as it was higher in smallest core chromosome. In *Petunia* chromosomes, the centromere of Ch.III has the closest length to the telomere region among all chromosomes as eccentric type with relatively small size (Smith *et al.* 1973). More importantly, the existence of PVCV and florendovirus-like sequences in the same chromosome (Figure 3.42) might supported similarity of the two EPRVs sequence in phylogeny as they pass through the same ancient endogenisation events. It might be a source of virus sequences to other chromosomes by rearrangements, crossovers, frequency of recombination (Lively 2010; Morran *et al.* 2011) or transposition, all features that influence genetic variability of species and elimination of pathogenic sequences. On the virus side, in animals, a crucial role is played by pathogen recombination, leading to annual recurring outbreaks of recombined viral strains driving epidemic influenza (Nelson & Holmes 2007).

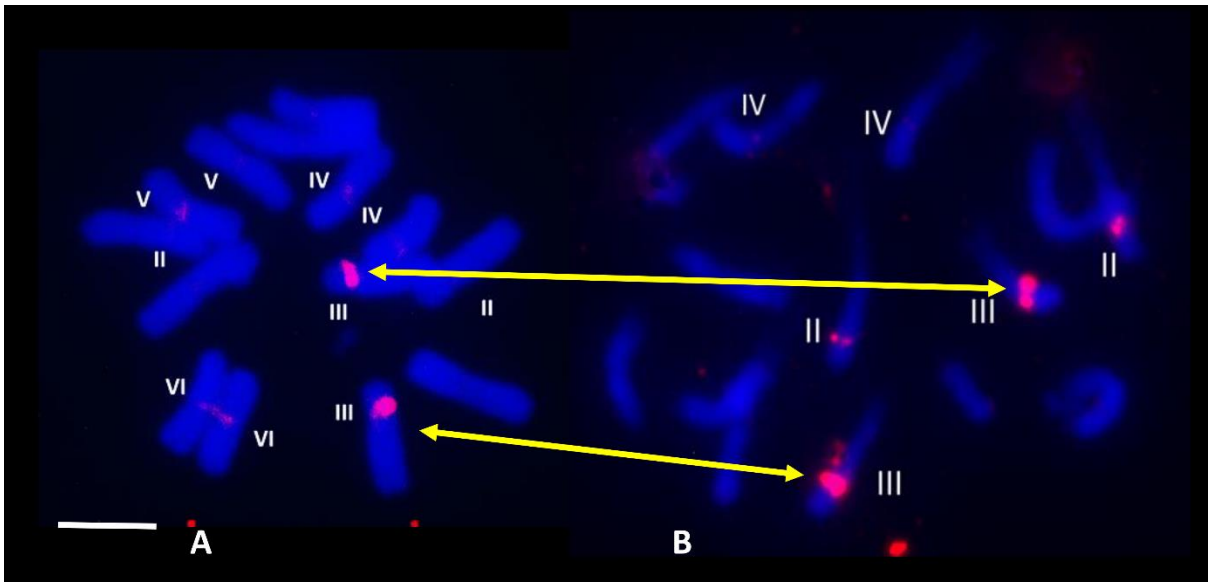


Figure 3.42 FISH image of florB (A) and PVCV (B) probes on *P. axillaris* subsp *axillaris* N chromosomes shows both probes hybridized at the same region of the centromere of Ch.III. The probe was labelled with biotin11-dUTP (red). Bar =10 µm.

3.5.5 Caulimovirus-like sequences

The caulimovirus-like sequence RepeatExplorer cluster was found in lower genome proportions than other two EPRVs clusters with varying proportions of copies in raw reads (Figures 3.29 and 3.30), and undetectable signals along chromosomes (Figure 3.32). Signal strength reflects copy numbers and abundance of applied sequence; although FISH is not a

quantitative tool, probes with low copy number will be hardly detectable and many researchers failed to quantify *in situ* experiments (Leitch & Leitch 2013). Here, four novel caulimovirus sequences have been identified in raw reads from each species, and they have multiple ORFs (more than four; Figure 3.28). In the assemblies, protein domains were surprisingly rearranged, and this might happen due to different cutting points and integration sites of their circular DNA at endogenisation (20-34 MYA), probably explaining why *Caulimovirus-PHy* has an inverted sequence with respect to other caulimovirus-like sequences. Recently, many endogenous viruses that belong to genus *Caulimovirus* have been characterised in plant genomes, and despite these integrants being found fragmented and degenerated, they could be reassembled through *in silico* (Geering *et al.* 2010; Geering *et al.* 2014), and data suggests *in vivo* too (Harper *et al.* 1999; Richert-Pöggeler *et al.* 2003). Sequences of endogenous caulimoviruses have been arranged and listed in a Repbase database (Appendices 3.8-3.11) that incorporates representative repetitive sequences of eukaryotes species. They have variable numbers of ORFs with normal sequence direction (5'-3') (Bao *et al.* 2015). Interestingly, this study has not confirmed that some caulimovirid sequences have bipartite genome structure like that mentioned in Geering *et al.* (2014) study about *Florendovirus* members depending on their gene organization as there is no such example of bipartite genomes in the members of the *Caulimoviridae*.

3.5.6 Expression of EPRVs

Exploring more details about the activity of all three EPRVs has further importance to investigate their ability for expression and being active under specific conditions. Synthesizing cDNAs for infected and healthy *Petunia* samples has revealed that PVCV and caulimovirus-like sequences have been strongly expressed, unlike *Florendovirus* that surprisingly have not been detected (Figure 3.38), suggesting that despite very abundant copy numbers (Figure 3.23), they are relatively inactive. So far, our results have disagreed with the comment of Geering *et al.* (2014) about good representation of *Florendovirus* members in expressed sequence tag (EST) databases, suggesting that *Florendovirus* sequences are transcribed. Interestingly, Geering *et al.* (2014) has reported findings above based on only three from 27 host genomes (*Citrus clementia*, *Oryza sativa* and *Prunus persica*) data with a range between 47-57% alignment identities. So far, the transcription of florendovirus-like sequence has dependent activity according to host effect and specificity, also extended search to explore more facts

from other host genomes is quite recommended. It is also noteworthy that early analysis of EST libraries from potato showed pararetrovirus-like sequences, particularly in those derived from callus (Hansen *et al.* 2005).

The chromosomal environment has some influences on the expression of genes that observed in some species such as in *Drosophila* where particular gene has been inactivated after it translocated into heterochromatic region (Reuter *et al.* 1990), while in tomato, the sulfurea locus instability was studied in closely located gene to heterochromatin region that suppressed its function (Hagemann & Snoad 1971). In *P. hybrida*, the foreign genes expression of lines 17 and 24 show stable and distinct methylation pattern at the integration regions in transgenic and wild plants. This pattern was imposed on the flanked regions of the integrated fragments and probably it's responsible for the intensity changes of gene expression (Pröls & Meyer 1992). Richert-Pöggeler *et al.* (2003) reported that integrated PVCV sequences of healthy petunias are likely to be methylated. While under induction factors, the virus could be released at short period of demethylation in one integration loci at single cell.

Phylogenetic relationships among florendoviruses have revealed that *P. axillaris* subsp *axillaris* N and *P. hybrida* were so far closely related genomes (Figure 3.33) and reflected that most of the hybrid characteristics came from the parent *P. axillaris* subsp *axillaris* N, the same story as for caulimoviruses. EPRV data from examined genomes has suggested that *P. integrifolia* subsp *inflata* S6 was not permissive with invasion of these elements, unlike the hybrid one, *P. axillaris* subsp *axillaris* N and *P. axillaris* subsp *parodii* S7 respectively (Figures 3.12, 3.13, 3.22, 3.23, 3.29 and 3.30). This outcome suggests that *P. integrifolia* subsp *inflata* S6 contributed less than *P. axillaris* subsp *axillaris* N in formation of the hybrid genome, a result in agreement with preliminary data that has been discussed within the Petunia genome sequencing consortium (Bombarely *et al.* 2016).

The role of endogenous pararetroviruses in inducing new virus infection from the host genome has opened the door widely to answer many questions about the benefits of EPRVs to their host. It was reported for long time that EPRVs could prevent new infections as one of defence role for EPRVs, but researches of inducible patterns of *Petunia vein clearing virus*, *Banana streak virus* and *Tobacco vein clearing virus* have contradicted the first statement of EPRVs role. In addition, these elements have been found with higher numbers of copies than silencing approaches need to prevent virus infection, as well as having different distributions in different plant families that have no infection from the exogenous related virus (Jakowitsch *et al.* 1999; Wesley *et al.* 2001; Kunii *et al.* 2004). The differences between the species in endogenous copy

numbers and in the domains represented, may suggest silencing mechanisms are more efficient in some genomes than others in parallel with cross-protection, and silencing of RNA-viruses (*Papaya ringspot virus*; Ming *et al.* 2008).

3.5.7 *De novo* integration

Vertical transmission has been studied here through *de novo* integration using our virus of interest PVCV that horizontally infected *P. axillaris* subsp *parodii* S7 by grafting, and then maintained vertically since 2001. Many plant viruses use vertical transmission through seeds (or other propagules such as seed potatoes) to cause severe infections on host plants by transmitting virus particles from parents to hybrids, as well as this method has a huge impact on pathogen virulence and plant adaptation. It seems most likely that vertical transmission rate has negative correlation with pathogen virulence that may have evolved with challenge to the hypothesis of trade-off and suggested models (Ewald 1983, 1987; Alizon *et al.* 2009).
PVCV is integrated into arbitrary loci on host plant chromosomes as virus signals were found on the telomere of the short or long arm of heterologous chromosomes and signal strength widely fluctuated within examined chromosomes (Figure 3.39 and 3.40). *De novo* assembly confirmed the telomere sequence (TTTAGGG) was associated around the recombinant virus sequence suggesting PVCV has integrated in the telomere region (Figure 3.41). In plant genomes, this study has been applied for the first time as only in recent years, few researches have started to explore this type of integration for herpesvirus 6 (HHV-6) within human chromosomes using fluorescent *in situ* hybridization (FISH). Herpesvirus 6 integrated in the telomere site of human chromosomes, and the examined individual's cells were 100% vertically infected, also virus signals found interestingly on short or long arm of heterologous chromosomes (Nacheva *et al.* 2008). Further studies have characterised that chromosomally integrated herpesvirus 6 is so far unstable and could be released to cause infection as a functioning virus (Huang *et al.* 2013; Wood & Royle 2017). Our data in this research has so far agreed with data from human genome, suggesting similar pattern of integration between two viruses (PVCV and HHV-6) and two hosts (*P. axillaris* subsp *parodii* S7 and *Homo sapiens*). However, PVCV sequence has TTAGGG (mammalian-type) telomeric repeat in two regions at the start (590-595 bp) and end (6214-6219 bp) of the whole length. PVCV integrates at the telomere of infected *Petunia* probably because of unstable proximal telomeric sites of these chromosomes and still prone to high recombination rate coupled together with abundant

virus genomes that integrated by such a mechanism contributing the telomere regions (McKnight *et al.* 2002; Wang *et al.* 2004). This mechanism perhaps starts with disturbing the t-loop site and allow emergence of telomere rapid deletion events as reported in yeast (Lustig 2003), and then truncated regions may arise by double strand break and end processing to produce single-strand overhang in presence of virus genome strands that is easily interfered with (Huang *et al.* 2013). However, this model was taken from results of other kingdoms, and the applicability of the model to plant genomes needs further investigation. Sequencing using technologies such as 10X Genomics or Nanopore Minion to identify long-range associations of sequences would be helpful.

Chapter IV. Transmission electron microscopy of PVCV

4.1 Introduction

4.1.1 Electron microscopy and detection of *Petunia vein clearing virus* (PVCV)

Transmission electron microscope (TEM) of ultrathin sections or sap from infected tissues is widely used for detection and identification of plant and animal viruses. In a study aimed to identify phytoplasma–like bodies and viruses in *Opuntia tuna* showing witches–broom disease-like symptoms, Lesemann and Casper (1973) reported that it was not possible to detect virus-like particles in ultrathin sections of embedded *O. tuna* tissues and that most transmission experiments from infected *Opuntia* were negative. *Petunia* test plants that were not used for the transmission experiments, showed all of sudden vein clearing symptoms and leaf deformation in young shoots with stunted growth. In ultrathin sections of embedded symptomatic leaf and stem material, virus-like particles and inclusion bodies were discovered. Only grafting from petunia to petunia transmitted the virus-like particles. Based on host plant and symptoms the infectious entity was named ‘`*Petunia vein clearing virus*’’ (PVCV). No serological relationships between PVCV and other caulimoviruses using immuno electron microscopy (IEM) was identified (Lesemann & Casper 1973). Based on phylogentic analysis and distinct genome organization PVCV was classified in a new genus amongst *Caulimoviridae* named *Petuvirus* (Richert-Pöggeler & Shepherd 1997; King *et al.* 2011). Lack of molecular tools and information about virus biology at that time made electron microscopy a useful diagnostic tool for PVCV detection in plant tissues. However, using electron microscopy for virus detection is hampered because of the low virus concentration and unequal virus distribution within plant tissues (Sikron *et al.* 1995). Consequently the risk of false negative results for PVCV existed which is critical when testing propagation material from commercial nurseries (Zeidan *et al.* 2000). A PCR based technique increased the sensitivity for plant virus detection (Rybicki & Hughes 1990; Navot *et al.* 1992; Zeidan *et al.* 2000). The virions can be detected in crude samples only in symptomatic tissues. Virion’s diameter was variable with the majority of particles being between 40 and 45 nm in size (Lesemann and Casper, 1973). Ultrathin sections of embedded symptomatic plant material revealed that many infected cells contained one or more densely stained inclusion bodies (IB) with a diameter up to 10 µm. The inclusion bodies were rounded, but in some cases appeared irregular, the matrix

appeared finely granular and within the bodies many lightly contrasted, rounded, vacuole-like regions were present. The IBs were not enveloped by a membrane. Virus particles have not been found within the IB matrix but sometimes close to the inclusion bodies and just within vacuole-like regions, particles have rarely occurred within a nucleus (Lesemann & Casper 1973). Here, the ultra-structure of variability in vein clearing symptom expression reaching from a more spot like appearance to extended chlorosis along the veins were analysed using transmission electron microscopy (TEM). Furthermore the ultra-structure of PVCV infected cells originating from horizontal or vertical transmission was compared.

4.2 Aims and objectives

Aims: One aim of this chapter is to determine the cellular ultra-structure and any differences between vein clearing and spot symptoms. The second aim addresses if viral integration is essential for virus replication. Therefore, episomal viral replication from induced chromosomal copies of PVCV which were preexisting or derived from *de novo* integration was compared with PVCV replication in petunia plants after horizontal virus transmission.

Objectives:

Using transmission electron microscopy (TEM) together with immunogold labelling technique to associate different modes of PVCV transmission (horizontal or vertical) with virus particle concentration as well as with changes in the ultra-cellular structure of infected cells in both of episomal, *de novo* integrated and induced infections of PVCV.

4.3 Materials and methods

4.3.1 Plant species

The examined *Petunia* species were: i) episomally infected *P. axillaris* subsp *parodii* S7, registered as TEM16-35_1, showing vein clearing and local lesion symptoms. ii) progeny of PVCV-infected *P. axillaris* subsp *parodii* S7 grown from seeds on Murashige-Skoog agar (Appendix 4.3) as tissue culture samples showing vein clearing registered as TEM16-483. iii) progeny of episomally infected *P. axillaris* subsp *parodii* S7 that showed vein clearing symptoms, registered as TEM12-584. iv) induced symptoms under heat application were taken

from *P. hybrida* W138, registered as 3645d [(examined plants at four weeks old were exposed to 16 h at 30°C and then 8 h at 25°C until viral symptoms were induced, while control plants were grown for 24 h at 25°C (Noreen *et al.* 2007)]. v) healthy leaves of *P. axillaris* subsp *parodii* S7 that do not carry infectious chromosomal copies of PVCV, registered as TEM 16-403_3. vi) healthy leaves of *P. hybrida* W138, that harbor inducible PVCV genomes within their chromosomes, registered as TEM 16-484 (Table 4.1).

Table 4.1 List of examined samples of petunia with symptoms, registration codes and descriptions.

Samples		Symptoms	Registration codes	Descriptions
i	*P. axillaris* subsp *parodii* S7	Vein clearing and chlorotic spot	TEM 16-35_1	PVCV infected *P. axillaris* subsp *parodii* S7 with episomal form that horizontally transmitted via grafting.
ii	*P. axillaris* subsp *parodii* S7	Vein clearing	TEM16-483	Progeny of PVCV-infected *P. axillaris* subsp *parodii S7* grown from seeds on Murashige-Skoog agar.
iii	*P. axillaris* subsp *parodii* S7	Vein clearing	TEM 12-584	Progeny of PVCV infected *P. axillaris* subsp *parodii S7* that vertically transmitted.
iv	*P. hybrida* W138	Chlorotic spot	3645d	Induced *P. hybrida* W138 by heat application.
v	*P. axillaris* subsp *parodii* S7	No symptoms	TEM 16-403_3	Healthy *P. axillaris* subsp *parodii* S7 leaves.
vi	*P. hybrida* W138	No symptoms	TEM 16-484	Healthy *P. hybrida* W138 leaves.

4.3.2 Electron microscopy

Dip preparations were done with symptomatic leaf tissues of petunia. TEM grids were mounted on the crude sap preparations for 5 min, washed with water and for contrast negative stained with 1% aqueous uranyl acetate. Samples were dried and screened for presence of virions using a transmission electron microscope, Tecnai G2 spirit, FEI. TEM examination was performed at acceleration voltage of 80 kV providing optimal contrast for biological material. Images were taken using a 2x2k digital camera (Veleta) which was attached at the side of the microscope column. A serological approach was used to enhance detection sensitivity using the electron microscope. For immuno electron microscopy (IEM), a polyclonal PVCV antibody

produced in rabbits (Richert 1992) was employed. IEM included immunosorbant electron microscopy (ISEM) for trapping of virus particles with PVCV specific antibody diluted 1:1000. For virus identification, samples were incubated with PVCV antibody diluted 1:50. For a more pronounced visualization of antibody-virion interaction, selected samples from dip preparations as well as ultrathin sections were immunogold labelled with a secondary gold-goat-anti-rabbit antibody conjugate diluted 1:50. For immunogold labelling, 5 or 10 nm gold particles were used, and both particle sizes obtained specific labelling of PVCV particles.

4.3.2.1 Embedding procedure

According to Lesemann (1991), symptomatic leaf tissue was sliced into 1x2 mm pieces using a sharp razor blade. The leaf sections were transferred carefully and directly with a preparation needle into 1 ml of the first fixation solution (2.5 % glutaraldehyde in 0.1 M (K, Na) P-buffer) in 3.5 ml glass tubes. Leaf tissues were submerged in the fixation solution and degassed for 15 min in a vacuum desiccator. Once all air had been removed from the sample tissue, fixation was done for at least 2 hr followed by a washing step to remove the first fixation solution. For the second (post) fixation served 1 ml 0.5% osmium tetroxide in 0.1 M (K, Na) P-buffer for at least 2 hr. Samples were washed using a Pasteur pipette filled with Millipore water four times for 5 min each. For contrast, samples were negative stained with 1% aqueous uranyl acetate.

4.3.2.2 Preparation of embedding forms

Freshly prepared embedding resin (Epon812+DDSA, MNA, catalysator DMP30) from stock solutions was added to half the volume of the embedding form that had been labelled with a small piece of paper by pencil at upper right corner of the form (see Appendix 4.2). A preparation needle was used to remove air bubbles and the resin was left for polymerizing overnight at 40°C in a drying oven (Memmert). On the second day, dehydration procedure was done by preparing 50% acetone (15ml 100% acetone + 15 ml Millipore water) and 70% acetone (35ml 100% acetone + 15ml Millipore water). Then, washing for two times (30 min) with 50% acetone, two times (30 min) with 70% acetone and four times (30 min) with 100% acetone. After the last 100% acetone step, solutions have been added into small glass tubes (5 ml) with flat bottom that were labelled accordingly. Solutions were transferred via fast pouring and remaining pieces were transferred using a preparation needle. 100% acetone were replaced with acetone/Epon 1:1 in order to remove acetone, glass tubes placed inside dry oven with open

door and rotated by rotator for 1 hr at 40°C. Solution replaced with Epon and acetone/Epon solution was removed as much as possible. Final infiltration step of the fixed sample was done before transfer to embedding forms two times for 1hr at 40°C with Epon**.** Embedding forms that have been prefilled with Epon (half volume, see above) were filled up with the sample in Epon solution and air bubbles removed. Polymerization occurred at 60°C for 48 hr in a drying oven (Memmert). For each sample 4 repetitions were done, the generated blocks have been numbered chronologically (registration codes) and each repetition is indicated by a letter (see Table 2.2).

4.3.2.3 Immunogold labelling using ultra-thin sections

Three to five ultra-thin sections of 60 nm thickness in average were mounted on a nickel grid (75 mesh). To avoid unspecific binding, the mounted sections were blocked for 15 min with 1% BSA in 0.1 M phosphate (P-) buffer, pH 7.0 followed by washing with 1.5ml ELISA wash-buffer and excess liquid was removed. Incubation with the primary antibody diluted 1:50 in ELISA wash-buffer+0.5 % BSA was done overnight followed by washing with 1.5ml ELISA wash-buffer and excess liquid was removed. After that, the grid was incubated with the secondary antibody (gold-goat-anti-rabbit, 10 nm) diluted 1:50 in ELISA wash-buffer+0.5 % BSA for 2h. Before the addition of 2% uranyl acetate for contrast, the sections were washed with ultra-pure H_2O. Finally, the grid was washed with few drops of ultra-pure H_2O and stored dry (see Table2.2).

4.4 Results

4.4.1 Horizontally transmitted PVCV

In order to study the cellular ultra-structure of observed symptomatic phenotypes comprising diffuse chlorosis or direct vein clearing (Figures 4.1and 4.2) in leaves after horizontal transmission, tissues were embedded for TEM analysis. In *P. axillaris* subsp *parodii* S7, no full-length PVCV sequences capable of triggering infection are present (Figure 3.18; Appendix 3.3). Therefore these plants are predestined to study the ultra-structure of cells for PVCV replication using episomal DNA templates only without interference from inducible chromosomal PVCV DNA copies.

Figure 4.1 Typical vein clearing symptoms on young shoot after vertical transmission of PVCV. The depicted plant on the left, registered as TEM12-584, is from the progeny of *P. axillaris* subsp *parodii* S7 that had been infected with PVCV using biolistic inoculation and maintained in tissue culture over several years. Prior to pollination and seed harvest plants were transferred from tissue culture to the green house. Healthy leaf of *P. axillaris* subsp *parodii* S7, registered as TEM16-403_3 on the right.

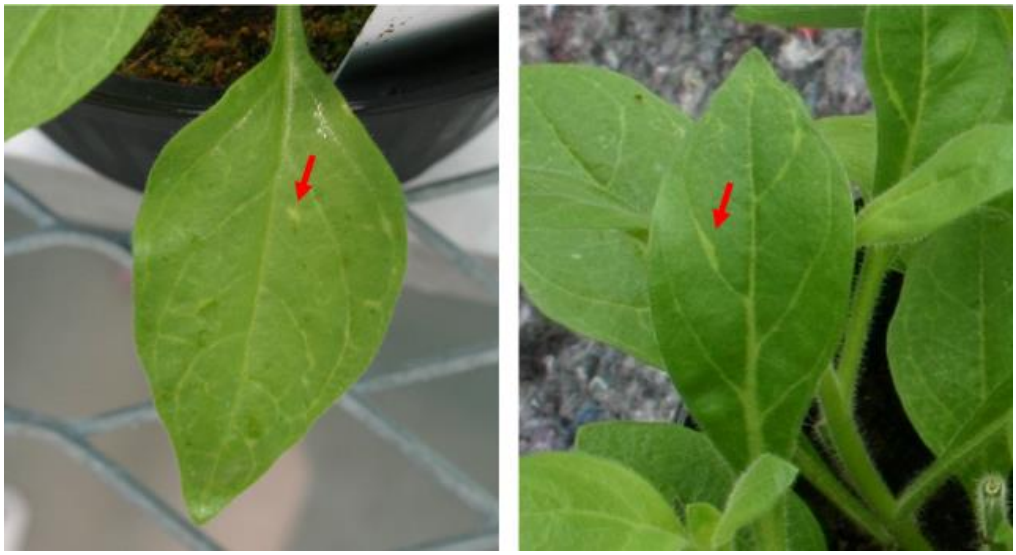


Figure 4.2 PVCV symptoms on infected *P. axillaris* subsp *parodii* S7 (TEM16-35_1) leaves start with yellow local lesion close to the leaf midrib (left), elongated spot along the side vein (right). *P. axillaris* subsp *parodii* S7 plants were infected with an infectious full-length clone of PVCV [72-2_3c, (Richert-Pöggeler *et al.* 2003)] using biolistic inoculation and propagated in tissue culture for several years and transferred to soil to increase growth under greenhouse conditions.

4.4.1.1 Cellular ultra-structure of diffuse chlorotic symptoms along the veins

In symptomatic leaves of *P. axillaris* subsp *parodii* S7 (TEM16-35_1) displaying elongated chlorotic spots (Figure 4.2 left panel), virus particles have been shown in mesophyll cells that are adjacent to the vasculature and concentrated in distinct cytoplasmic regions. Additionally, rough endoplasmic reticulum (ER) was noticed in these cytoplasmic regions adjacent to virus particles. The isometric particles had an average diameter between 36 and 46 nm. During temporal progression of infection these viroplasm seem to become more compact and appear as electron dense material in which virions are not distinguishable any more. In average 500 to 1000 virus particles were counted in infected cells. Various phases during maturation from viroplasm to inclusion bodies are displayed in Figures 4.3 and 4.4. Inclusion bodies (IB) have been noticed with a rounded shape in some cells and irregular shape in others with a diameter of 500 nm-2µm. IBs are without membranes. Vacuoles were observed in some infected cells and chloroplasts in the examined cells did not have any starch granules (Figures 4.3, 4.4 and 4.5). Immunogold labelling confirmed that the majority of virions can be found in the

cytoplasm and not within the inclusion body matrix where only few gold particles were located (Figure 4.6).

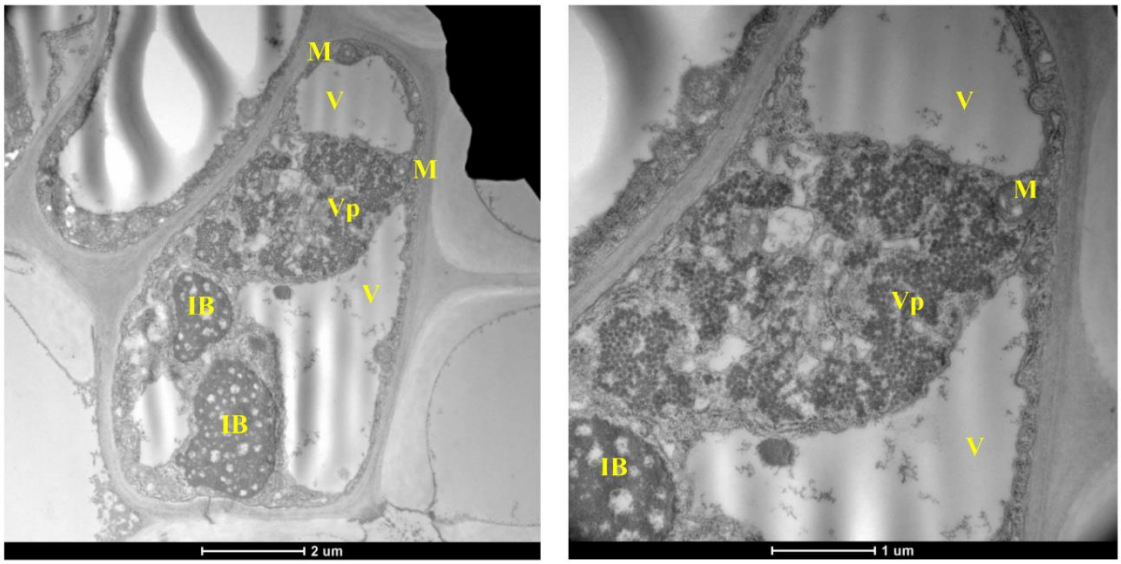


Figure 4.3 Ultrathin section of PVCV-infected *P. axillaris* subsp *parodii* S7 (TEM16-35_1) mesophyll leaf cell. High concentration of virions and two large inclusion bodies were identified in the cytoplasm. Vp= viroplasm, IB= inclusion body, M= mitochondria, V= vacuole. Bar= 2µm (left), and 1 µm (right).

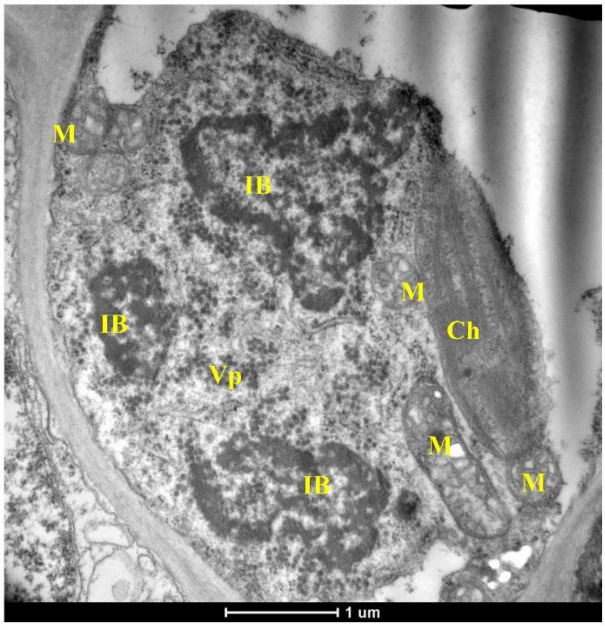


Figure 4.4 Transmission electron micrograph of ultrathin section of PVCV-infected *P. axillaris* subsp *parodii* S7 (TEM16-35_1) mesophyll leaf cell, showing distributed virions with less dense inclusion body matrix. Vp= viroplasm, IB= inclusion body, M=mitochondria, V=vacuole, Ch=chloroplasts without starch. Bar =1 µm.

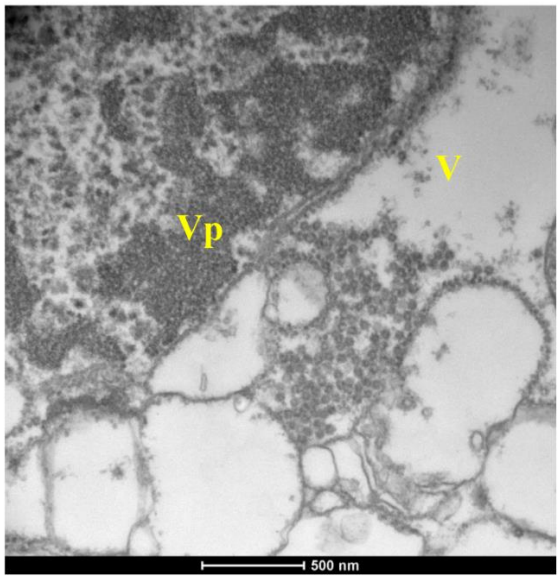


Figure 4.5 Details of PVCV virions within the cytoplasm in episomally infected *P. axillaris* subsp *parodii* S7 (TEM16-35_1) and interaction with vesicular membranes. Vp= viroplasm, V= vacuole. Bar=500 nm.

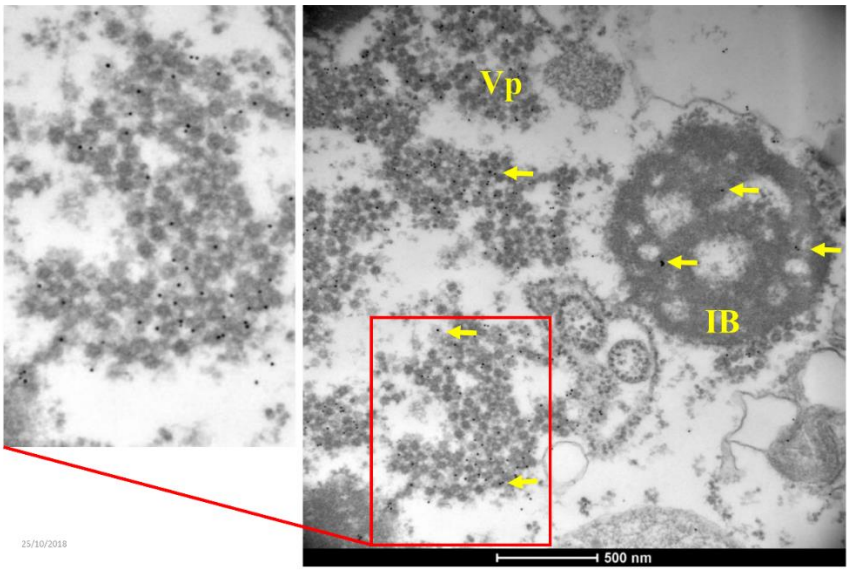


Figure 4.6 Immunogold labelling of PVCV particles in episomally infected *P. axillaris* subsp *parodii* S7 (TEM16-35_1) mesophyll leaf cell using 10 nm gold conjugated-anti-rabbit antiserum. The inclusion bodies have been noticed with a rounded shape. Some virions are found outside but in close contact with the IB (black arrow). Vp= viroplasm, IB= inclusion body, arrows indicate regions with gold label. Contrast and brightness have been adjusted to improve visibility of the gold particles. Bar=500 nm.

4.4.1.2 Cellular ultra-structure of vein restricted chlorosis symptoms

Tissue embedded from leaves of PVCV infected *P. axillaris* subsp *parodii S7* (TEM16-35_1) displaying vein clearing symptoms revealed a high concentration of PVCV particles in the cytoplasm of mesophyll cells adjacent to the vasculature. No inclusion bodies have been noticed in these cells indicating probably an early event in virus infection. Viroplasm has formed but has not yet reached a more condensed stage resulting in the formation of inclusion bodies. Vacuole like regions have been found with a big portion at the periphery of these cells. These results of vein clearing samples have been confirmed with another infected plant under the same greenhouse conditions (Figure 4.7).

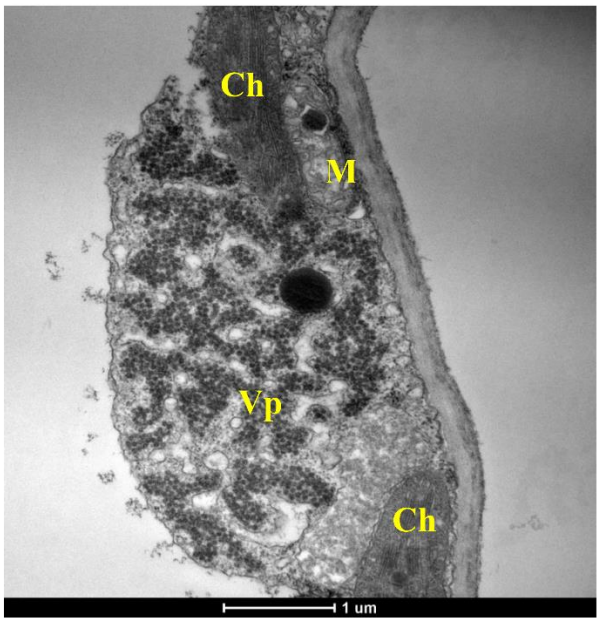


Figure 4.7 Transmission electron micrograph of ultrathin section of PVCV-infected *P. axillaris* subsp *parodii* S7 (TEM16_35_1) in the cytoplasm of a mesophyll cell adjacent to the vasculature, showing isometric virions. No inclusion bodies have been found in these cells. Vp= viroplasm, M=mitochondria, Ch=chloroplasts without starch. Bar =1 µm.

4.4.2 Vertical transmission of PVCV

4.4.2.1 TEM for progeny of PVCV-infected *P. axillaris* subsp *parodii* S7

PVCV infected *P. axillaris* subsp *parodii* S7 derived from biolistic inoculation using an infectious full-length clone have been vegetatively propagated for several years on Murashige Skoog medium by cuttings and transferred to the greenhouse for further growth and seed production. The seeds of such plants were grown under sterile as well as greenhouse conditions. Under both cultivation methods, vein clearing symptoms occurred and were tested for episomal PVCV infection which must have occurred during vertical transmission via seeds. PVCV particles have been shown in companion cells adjacent to the vasculature (Figure 4.8). The particles have been found in condensed regions within a cell and were spherical with an average diameter of 44 nm, immunogold labelled sections have confirmed this result as gold particles were attached to virions. Inclusion bodies have not been found in the infected cells (Figure 4.9).

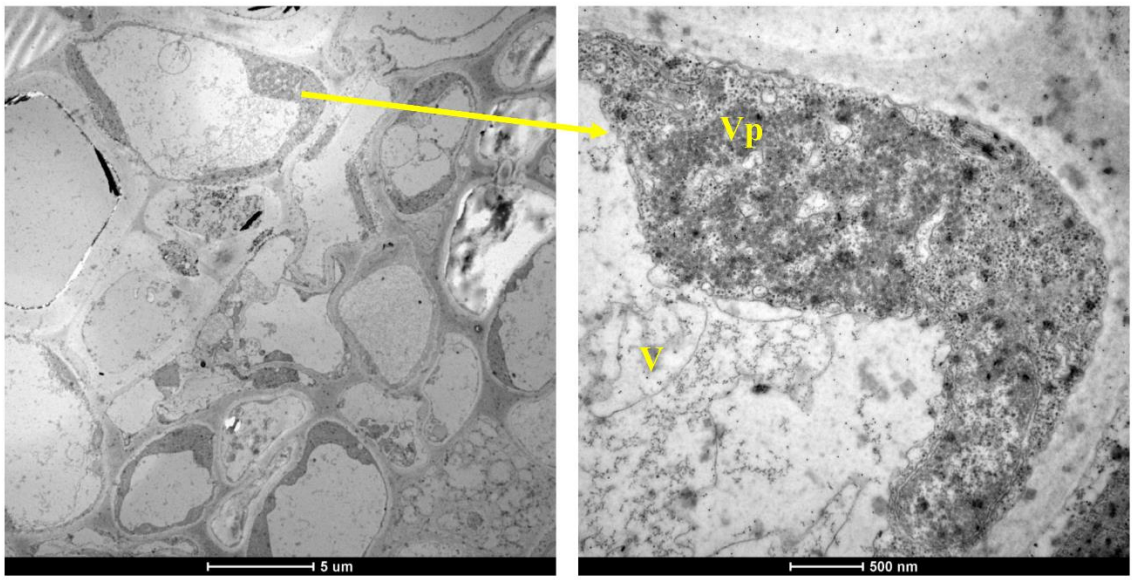


Figure 4.8 Ultrathin section of progeny of PVCV-infected *P. axillaris* subsp *parodii* S7 (TEM16-483) grown from seeds on Murashige-Skoog agar, showing virions in the cytoplasm of a bundle sheath cell. No inclusion bodies have been found in these cells. Vp= viroplasm, V= vacuole. Bar = 5 µm (left), and 500 nm (right).

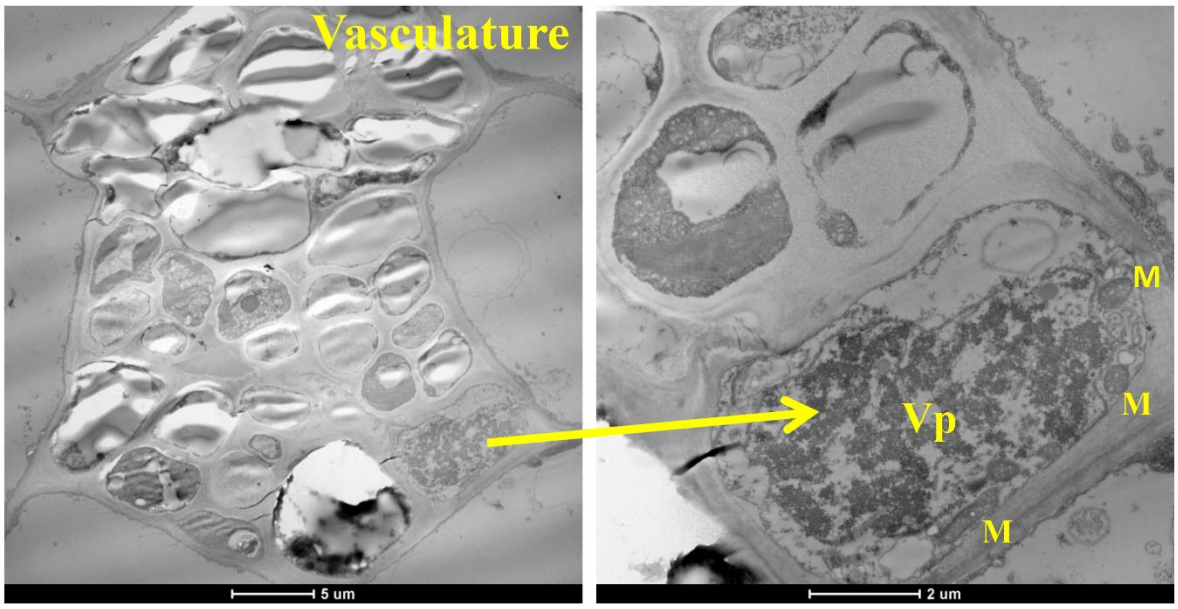


Figure 4.9 Overview and details of ultra-structure of progeny of PVCV-infected *P. axillaris* subsp *parodii* S7 (TEM12-584) leaves with vein clearing symptoms. Virions are distributed in the cytoplasm of a probable companion cell, showing condensed virions. No inclusion bodies have been found in these cells. Vp= viroplasm, M= mitochondria. Bar =5 µm (left), and 2 µm (right).

4.4.2.2 Ultrastructure of PVCV infection after induction of chromosomal PVCV forming a tandem array

In the *P. hybrida* cultivar W138 it has been shown that integrated PVCV genomes arranged in a tandem array are infectious (Richert-Pöggeler *et al.* 2003). Furthermore, it has been reported that such chromosomal copies can be activated after heat exposure (Noreen *et al.* 2007). Symptomatic leaf sections from induced *P. hybrida* W138 have been embedded and ultrathin sections were screened for PVCV infected cells originated from vertical transmission. Isometric virions as well as inclusion bodies resided in the cytoplasm (Figure 4.10). About 300-400 virions were counted in infected cells that have approximately the same dimensions in comparison to particles (about 40 to 52 nm) derived from horizontal transmission (Figures 4.10 and 4.11). Three different stages of inclusion bodies were present (Figures 4.10 and 4.11). They all displayed a rounded shape with an average diameter of 450-600 nm. The cell displayed in Figure 4.10 contained two IBs with vacuole-like inner core that have released generated particles in the cytoplasm. These particles reacted specifically with immunogold. The IB in the upper left corner is electron dense and surrounded by cellular (ER) membrane structures. In

Figure 4.11 an earlier stage of PVCV replication is depicted. Here the two IBs, flanking the virion containing region, are less electron dense and virions seem to egress from the viroplasm matrix into the cytoplasm. Mature particles in the cytoplasm are immunogold labelled.

When chromosomal copies of PVCV are present they can theoretically trigger PVCV infection in each cell. The host plant seems to employ epigenetic regulation to avoid this (Noreen *et al.* 2007). Instead the virus can move from the initially infected cell via plasmodesmata to the neighboring cell. An indication for such intercellular movement is depicted in Figure 4.11 on the right panel. A single virion is located adjacent to the cell wall with plasmodesmata exposed.

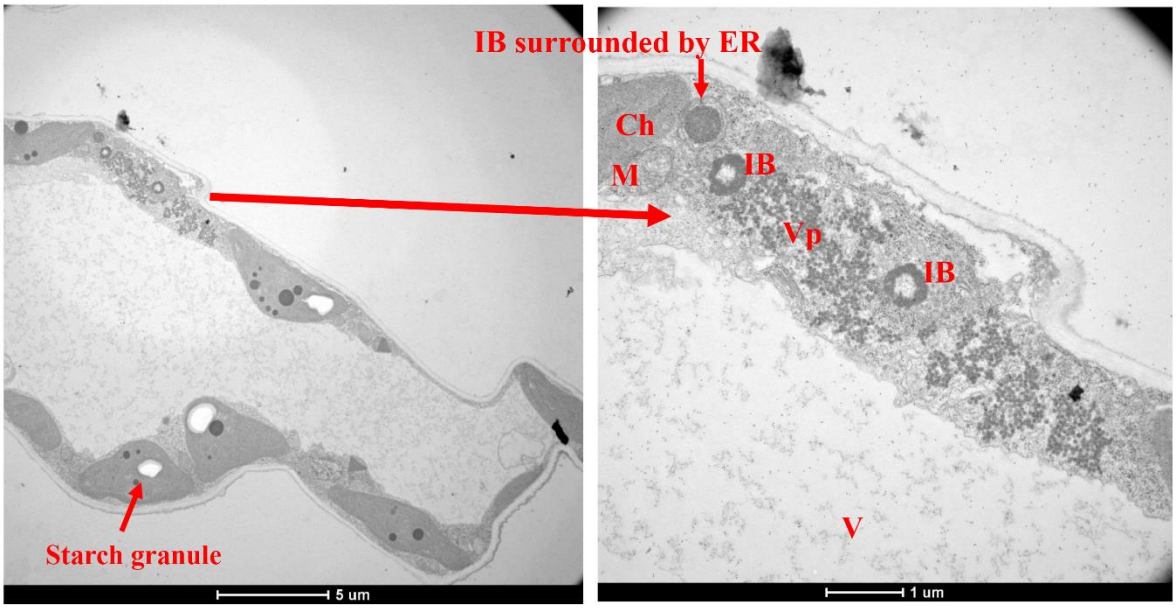


Figure 4.10 Transmission electron micrograph of an ultrathin section of induced *P. hybrida* W138 (3645d). In the cytoplasm of a mesophyll cell, scattered virions and three inclusion bodies are visible. Vp= viroplasm, M= mitochondria, IB= inclusion body, Ch= chloroplast with and without starch granules, V= vacuole. Bar = 5µm (left), and 1 µm (right).

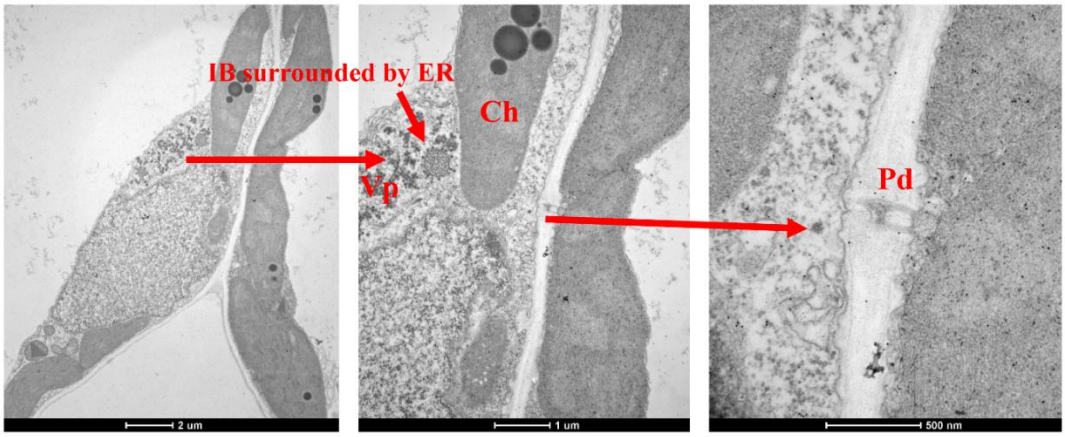


Figure 4.11 PVCV replication in induced *P. hybrida* W138 (3645d). In an early stage of replication PVCV particles seem to egress from the viroplasm matrix (VP) into the cytoplasm. Mature virions in the cytoplasm are immunogold labelled. Red arrow indicates single virion adjacent to the cell wall with plasmodesmata (PD) visible. Vp= viroplasm. Bar = 2µm (left), 1 µm (middle), and 500 nm (right).

4.4.2.3 TEM for healthy *P. axillaris* subsp *parodii* S7 and *P. hybrida* W138

Embedded tissue of healthy plants contained no abnormality regarding organelle sizes or shapes and vasculature. Neither PVCV virions nor inclusion bodies have been noticed in the respective cells. (Figures 4.12 and 4.13).

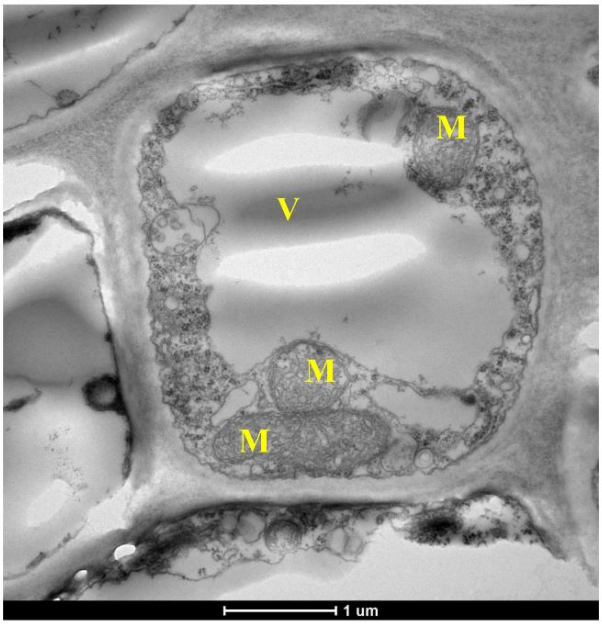


Figure 4.12 Healthy cell of *P. axillaris* subsp *parodii* S7, showing normal cell components in the vasculature. M= mitochondria, V=vacuole. Bar = 1µm.

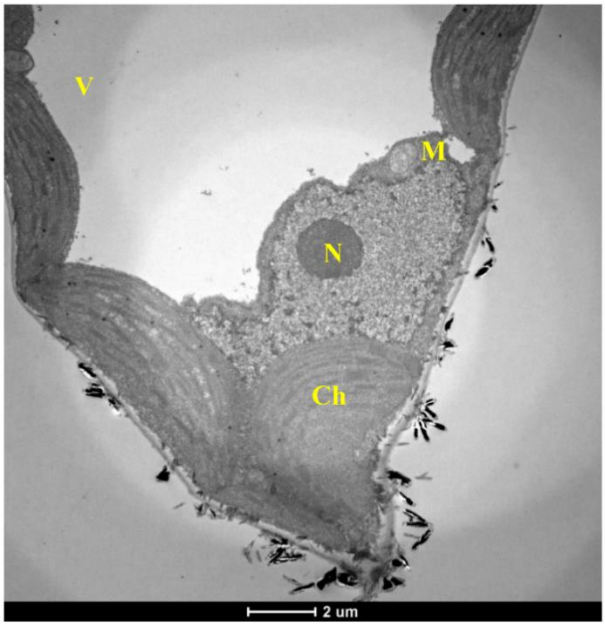


Figure 4.13 Ultrathin section of healthy *P. hybrid*a W138, showing normal cell components at parenchyma cells. N= nucleolus within the nucleus, M= mitochondria, Ch= chloroplast, V= vacuole. Bar = 2µm.

To sum up the results above see Table 4.2:

Table 4.2 Differences in virion and inclusion body numbers and sizes within applied samples contents.

	Type of symptom/origin of DNA	**Virion numbers**	**Virion sizes**	**Number of IBs**	**IB sizes**
1	Spreading chlorosis/episomal	500-1000	36-46 nm	2-3	0.5-2 µm
2	Vein clearing/episomal	500-750	40-44 nm	No	No
3	Vein clearing/chromosomal	300-750	44 nm	No	No
4	Spontaneous chlorosis/chromosomal	300-400	40-52 nm	1-3	450-600 nm

4.5 Discussion

Electron microscopy provides high resolution images of the viruses, and shows their relationship or interactions with the host cells. Together with microscopy techniques based on light (fluorescence microscopy, confocal laser scanning microscopy) and molecular biology tools (e.g. PCR, FISH, NGS) it enables comprehensive analysis of the virosphere (Hoang *et al.* 2011; Corbetta *et al.* 2017; Hesketh *et al.* 2017; Lei *et al.* 2017; Balke *et al.* 2018). In this chapter we employ electron microscopy to study PVCV in petunia cells (Figures 4.3- 4.13). PVCV can occur both in episomal as well as chromosomal forms depending on the host. Both templates can trigger viral infection. The aim was to compare PVCV replication after horizontal and vertical transmission respectively. In earlier studies it had been shown that the presence of PVCV particles is correlated with symptom expression (Richert 1992). Therefore, the impact of various phenotypes on viral replication also investigated in the distinct transmission modes.

4.5.1 Ultrastructural analysis of diffuse chlorosis and vein clearing symptoms after horizontal transmission of PVCV

Despite no vector for horizontal (from plant to plant) PVCV transmission being identified yet, plants can be infected artificially using grafting or biolistic inoculation (Richert-Pöggeler *et al.* 2003). The leaf symptoms on PVCV infected petunia are generally described as leaf yellowing and vein clearing in young shoots (Lockhart & Lesemann 1998; Zeidan *et al.* 2000; Richert-Pöggeler *et al.* 2003). In some cases, symptoms start with local chlorosis adjacent to one of small netted veins then developing either strictly along the vein or becoming a diffuse chlorosis spreading further into the leaf lamina (Figure 4.2). In horizontal PVCV transmission, the encapsidated circular double stranded DNA molecules are released from the capsid and are transcribed in the plant nucleus. The generated RNA molecule comprising the complete viral genome including promoter sequences is transported to the cytoplasm. The transcript serves as messenger RNA and as template for DNA synthesis using reverse transcription. To determine whether the differences in symptomatic phenotypes were also reflected on the cellular level of PVCV infected *P. axillaris* subsp *parodii* S7 plants, electron microscopy was applied, including use of immunolabelling of PVCV to confirm the identification of structures with viral capsids. So far, cells from tissue with vein clearing contained clustered virions but lacked viral inclusion bodies. The minimum diameter of the virion was 36nm and the maximum was

45nm; this is in accordance with the published value of 46nm which was obtained measuring purified virions directly (Richert-Pöggeler & Lesemann 2007). Many of these cells included vacuoles, and chloroplasts sometimes contain starch granules. In leaves displaying diffuse chlorosis, infected cells contained virions in parenchyma cells with minimum diameter 44nm and the maximum was 56nm. Conversely to cells derived from vein clearing tissue, there were condensed inclusion bodies of rounded or irregular shapes with a diameter of 500 nm – 2 µm similar to those described by Lesemann and Casper (1973) (Figures 4.3, 4.4, 4.5 and 4.6). The existence of rough endoplasmic reticulum (ER) in the cytoplasmic regions adjacent to virus particles may be an indication that these metabolic active sites are also used for translation of viral proteins (Schaad *et al.* 1997; Turner *et al.* 2004).

Comparing the two symptom types, differences in particle sizes and IB numbers have been found in infected cells for each phenotype (Table 4.2). These observations give first insights in stages during viral replication. Early on infected tissue in which chlorosis is limited to the vein region viral replication starts. Spreading of viral infection to the leaf lamina is accompanied by enhanced viral replication. As consequence more virus particles are present and inclusion bodies are formed. However, in virology, inclusion bodies are accumulated dead-end materials that contain unused viral proteins in the cytoplasm of infected cells at late times of infection, while viroplasms are electron-dense masses at early stages of infection include viral and cellular factors in viral replication sites within infected cells (Novoa *et al.* 2005). The variability in particle sizes may indicate different steps of particle interactions with viral proteins as well as with host components. For *Cauliflower mosaic virus* it has been shown that during virus movement interaction of coat protein with the virion associated protein occurs (Stavolone *et al.* 2005; Amari *et al.* 2010; Stavolone & Lionetti 2017).

4.5.2 *De novo* integration and vertical transmission of PVCV in *P. axillaris* subsp *parodii* S7

The *P. axillaris* subsp *parodii* S7 genome was assumed to be free of any integrated form of PVCV, since no signal was generated under stringent hybridization conditions using total genomic DNA as template and virus specific probes (Richert-Pöggeler *et al.* 2003). When artificially infected using grafting or biolistic inoculation, *P. axillaris* subsp *parodii* S7 proved to be a suitable host for PVCV allowing episomal replication (Figures 4.1 and 4.2). These experiments showed that unlike retroviruses, endogenous pararetroviruses do not depend on integration into the host genome to complete their replication cycle; indeed, before Harper *et*

al. (1999), integrated pararetroviruses were unknown in plants. *P. axillaris* subsp *parodii* S7 is an ideal model system to study *de novo* integration of PVCV sequences into the *Petunia* genome. Whereas it is estimated that integration occurs frequently in somatic cells, the manifestation of such an event in reproductive cells and thus in the progeny may be rare (Gayral *et al.* 2008). Therefore the progeny of PVCV infected *P. axillaris* subsp *parodii* S7 grown in tissue culture for prolonged times (>10 years) were screened. For infection of sterile plants, biolistic inoculation with an infectious full-length clone had been used (Richert-Pöggeler *et al.* 2003). Cuttings from PVCV infected plants were transferred to new MS medium every 4-8 weeks. The dedifferentiation of cells during the process of new root development most likely facilitates invasion of the host genome by linearized PVCV DNA molecules present in the nucleus. Furthermore it had been postulated that integrated copies get activated in such meristematic cells due to changes in epigenetic modifications (Noreen *et al.* 2007).

PVCV infection of symptomatic plants were confirmed by PCR and IEM. A first evidence for vertical transmission of PVCV was obtained when the progeny of these plants showed typical vein clearing symptoms and tested positive in PCR and IEM. FISH further confirmed *de novo* integration of PVCV in the telomeric regions of chromosomes III and IV (see Figure 3.39).

The fact that virions but no IBs were detectable (Figures 4.7, 4.8 and 4.9), may indicate either that an early step in viral replication is caught or that in this cell only intracellular replication occurs. The latter case would be similar to the replication of LTR retrotransposons (Richert-Pöggeler & Schwarzacher 2009).

4.5.3 PVCV symptom expression and replication after induction of chromosomal copies

Elevated temperatures above 25°C for a continuous time period of 7 days can enhance symptom expression in *P. hybrida* (Zeidan *et al.* 2000). Additionally, *P. hybrida* line W138 was shown to be more permissive than cultivar Rdc (Himmelsröschen) to induce symptoms after repeated heat treatments with a higher rate of infection (Noreen *et al.* 2007). Symptoms from induced infection are often starting as spot features more than vein clearing. In contrast to horizontally infected *P. axillaris* subsp *parodii* S7 cells, symptom expression resulted from activation of chromosomal copies. Particle concentration in W138 cells was at the maximum about 400 /cell (Figures 4.10 and 4.11), while in PVCV infected *P. axillaris* subsp *parodii* S7 it could reach 1000 /cell (Figures 4.3- 4.7). Virions were found loosely scattered in the cell

whereas in horizontally infected *P. axillaris* subsp *parodii* S7 additionally viroplasm and IBs were present.

The ultra-structural analyses provided first evidence that the efficiency in viral replication depends on the origin of PVCV infection. Virus activation from chromosomal copies seems to be less efficient. This may happen due to tighter epigenetic control by the host than from episomal PVCV copies. It may also take longer to build up high virus load because the number of integrated tandem array is less than the number of circular DNA molecules provided in artificial infection. The low efficiency of expression of EPRVs is also found in banana, where infection originating from endogenous copies is unusual, sometimes following stress such as tissue culture or cold nights (Harper *et al.* 1999; Harper *et al.* 2002).

The virulance of viruses could be significantly affected by transmission mode as its increase within horizontal transmission as a side effect of a trade off between infectivity and virulance. In vertical transmission, virulance decreases to allow virus transmission to wide host offsprings. Stewart *et al.* (2005) reported that *Barley stripe mosaic virus* (BSMV) horizontally transmitted in its host (*Hordeum vulgare*) with twofold increase in infectivity and tripled virulance, while in vertical transmission, infectivity slightly increased, while virulence highly reduced.

Interestingly, in the *P. hybrida* line W138 telomeric PVCV insertions as seen after *de novo* integration in *P. axillaris* subsp *parodii* S7 of this study were identified (Richert-Pöggeler and Schwarzacher, unpublished) (see Figure 3.18). The plasticity of telomeres harbouring various retroelements as shown for bdelloid rotifers and *Drosophila* (Arkhipova *et al.* 2017; Casacuberta 2017), and we can speculate similar plasticity and reorganization of sequences in telomeric regions may also be true for petunia. Notably, Vicient and Casacuberta (2017) discussed transposable elements and their insertion into heterochromatic regions of chromosomes (including centromeres and telomeres), potentially a model for the behaviour of EPRVs. Future studies should address if the invasion of the *Petunia* genome by PVCV starts at the telomere and then spreads to other loci in the pericentromeric regions of chromosomes. The identified telomeric copy in W138 may indicate the dynamics in virus host symbiosis and thus may represent a recent *de novo* integration by PVCV.

Chapter V. Tandem repeats

5.1 Introduction

Tandemly repeated or satellite (satDNA) DNA sequences are often present at pericentromeric and centromeric heterochromatin, and may related to chromosome pairing and separation of chromosomes in eukaryotic cell divisions (Arney & Fisher 2004; Hall *et al.* 2004; Bloom 2007; Sepsi *et al.* 2017). Within plant genomes, tandemly repeated DNA nature and content differs between species, and sequence may diverge within species. Tandem repeats typically represent 5% to 25% of all the DNA, with transposable elements being an even more abundant repetitive motif (Elder Jr & Turner 1995; Schmidt & Heslop-Harrison 1998; Melters *et al.* 2013), with the repetitive component of the genome responsible for much of the difference in DNA content between species (Doolittle & Sapienza 1980; Cavalier-Smith 1985; Gregory *et al.* 2006). With the introduction of large scale DNA sequencing, the nature, evolutionary mechanisms and functions of tandem repeats can be studied across the whole genome, complementing earlier studies of its abundance (Charlesworth *et al.* 1994; Elder Jr & Turner 1995; Schmidt & Heslop-Harrison 1998; Ugarković & Plohl 2002). Tandemly repeated DNA consists of monomers, often with high AT/GC nucleotide ratio, and variable lengths between few bp to more than 1kb within arrays consisting of tens to millions of the monomer. Monomers have preferential lengths around 175 and 360 bp, reflecting the DNA length wrapped around nucleosomes (Schmidt & Heslop-Harrison 1998; Henikoff *et al.* 2001). Tandemly repeated motifs assembly has proved impossible with current short-read technologies because of collapse of the sequence motifs within large contigs; while large-insert sequences, and long-molecule sequence reads are becoming helpful, there are still major challenges in measuring copy numbers and determining genomic locations of major arrays in both plants and animals (Kuhn *et al.* 2011).

Despite the abundance of tandem repeats, more data on their role in genome organization and function is required (Eichler *et al.* 2004; Nagaki *et al.* 2004; Rudd & Willard 2004; Hoskins *et al.* 2007; Heitkam & Schmidt 2009; Talbert *et al.* 2018). Large volumes of high-quality random sequence reads of 100 to 350 bp are suitable for identification of abundant tandemly repeated sequence motifs from genomes with appropriate analysis tools. *K-mer* analysis identifies the most abundant sequence motifs *k*-bases long in short raw read sequences. Graph-based clustering of raw reads, particularly using the RepeatExplorer software tools (Novák *et al.* 2013) is also allowing repeat sequence identification. The tandem repeat analyser (TAREAN)

is a computational pipeline that can be applied to explore tandem repeats using graph-based sequence clustering algorithms (Novák *et al.* 2013; Bilinski *et al.* 2017; Novák *et al.* 2017). Using *in situ* hybridization to chromosomes with repetitive motifs identified by assemblies of abundant motifs in *k*-mers or graph-based clusters shows the genomic organization and any clustering of the loci.

Like most plant genomes, repetitive DNA is abundant in *Petunia* genomes, but the existence of these sequences at 60% (see Appendix 3.1) is comparatively low with a larger space of genes and low copy sequences, taking into consideration the genome size. In the assembled genomes of the *P. axillaris* subsp *axillaris* N (*PaxiN*) and *P. integrifolia* subsp *inflata* S6 (*PinfS6*), abundant repeats of the typical 180 bp or 340 bp that wrap around nucleosomes have not been found, while repeat searches revealed the existence of short (60) bp and long repeats (500-1000 bp) (Bombarely *et al.* 2016). RepeatExplorer (Novák *et al.* 2013) has been used to identify simple and low complexity repeats, finding clusters including retroelements. Bombarely *et al.* (2016) revealed that tandem repeats are not very abundant in the *Petunia* genome assembly or raw reads, although some were extracted from smaller scaffolds, many composed of a 169 bp repeat. A strong *in situ* hybridization signal was found at the centromeres. Larger scaffolds (409-826 kb) had multiple tandem repeat copies as internal as well as terminal sequences, and dispersed short arrays have been found in other scaffolds. Interestingly, there was no evidence found for association with centromeres of repeat carrying scaffolds to give the distribution of a *Gypsy* retroelement-related fragment located near the centromeres of all chromosomes.

5.2 Aims and objectives

The main goal of this chapter is to:

- identify the tandemly repeated satellite DNA motifs in *Petunia axillaris* subsp *axillaris* N and *P. integrifolia* subsp *inflata* S6 with their hybrid cultivars *P. hybrida* Rdc, *P. hybrida* V26 and *P. hybrida* W138, and the sister *P. axillaris* subsp *parodii* S7.

- characterise their nature, organization, diversity and abundance to understand the evolutionary processes and potential consequences of the presence of this abundant component of the nuclear genome.

5.3 Materials and methods

5.3.1 Plant material and DNA extraction

Petunia species from the same sources in chapter II; 2.1 (see Table 2.1), were germinated and root-tips were used for somatic chromosome preparations. Genomic DNA was prepared from young leaves using the CTAB method (mentioned earlier in 2.3.1) (Doyle & Doyle 1990) with some modifications. Purified genomic DNA was used to amplify tandem repeats by PCR (see Table 2.5), cloned (see 2.3.5), and sequenced commercially (see 2.3.5.6).

5.3.2 Tandem repeat analysis

Illumina HiSeq PE126 raw reads of *P. hybrida* R27 (*PhybR27*) were obtained from Aureliano Bombarely, Department of Horticulture, Virginia Tech, USA, in addition to the assemblies and raw reads of *P. integrifolia* subsp *inflata* S6 (*PinfS6*) and *P. axillaris* subsp *axillaris* N (*PaxiN*) genomes (Bombarely *et al.* 2016). Genomic DNA of *P. axillaris* subsp *parodii* S7 was sequenced commercially by Novogene, Hong Kong, China using Illumina HiSeq-PE150 reads. These data have been used in a basic analysis of tandem repeat clusters within genome sequences using Geneious (http://www.geneious.com/) (Kearse *et al.* 2012). The programmes RepeatExplorer (Novák *et al.* 2013), and Tandem Repeat Analyser (TAREAN) (Novák *et al.* 2017), were used for graph-based clustering of repeated sequences in the raw reads. Additionally, Repbase (Jurka *et al.* 2005; Bao *et al.* 2015), Tandem repeats finder (Benson 1999), (BLAST); Basic Local Alignment Search Tool (Altschul *et al.* 1990), dot-plot tools and RepeatMasker (Smit & Hubley 2015) (http://www.repeatmasker.org) were used for checking, finding and characterising repeat sequences, and the database of conserved protein motifs of retroelements (Hansen & Heslop-Harrison 2004b). Copy numbers in raw reads were calculated by counting number of reads mapping to reference sequences (see 3.3.1.4).

5.3.3 Chromosomes preparation and *in situ* hybridization

Root tips from young plants of all examined *Petunia* species were pretreated with 0.2 M 8-hydroxyquinoline for 4 hr before fixation with freshly made ethanol:glacial acetic acid (3:1). Chromosome preparations were made on slides following proteolytic digestion of fixed and washed root tips with pectinase and cellulase (see 2.3.7). The method of Schwarzacher and Heslop-Harrison (2000) was applied for fluorescent *in situ* hybridization (mentioned earlier in

2.3.8). For making probes, PCR was used to amplify *Type I-CL43A* and *Type II-CL43B* (Tables 5.1and 5.2) and products were labelled with biotin-11-dUTP and digoxigenin-11-dUTP via BioPrime (Invitrogen) DNA Labelling System and Array CGH Genomic Labelling Module (see 2.3.6). Probes of the six other clusters (Table 5.3) were synthesized as oligonucleotides 5'-labelled directly with biotin-11-dUTP (Sigma Aldrich Company). Hybridization sites were detected by anti-avidin antibody conjugated to red fluorochomes, and anti-digoxigenin antibody conjugated with FITC. DAPI (4, 6-diamidino-2-phenylindole) and an antifade solution were used for counterstaining and mounting chromosomes. Slides were examined using a Zeiss Axioplan 2 or Nikon Eclipse 80i epifluorescence microscope and images were captured as mentioned earlier in 2.3.8.5. Adobe Photoshop CC2015.5 was used for preparing and overlaying images and hybridization signals, apart from cropping, using only functions affecting the whole image equally.

Table 5.1 The first two tandem repeat clusters primers.

No	**Primer name**	**Sequence**	**Tm °C**	**Product length (bp)**
1	CL43A	**(**Scf160-72F) CCG AAA GCG CAA AC TATC CC (Scf160-26R) AAA AAG AGG TAG GCG TTG A AG	64.5	157
2	CL43B	(Scf160-62F) ACC AGC AGA AAT AGA TCC AC (Scf160-64R) TGC AGA AGT ACA TCT ATG CG	53	1002

Table 5.2 PCR cycling program for the first tandem repeats primers.

Step		**Temperature °C**	**Duration (Minutes:seconds)**	**Cycles**
1	Initial denaturation	95	3:00	1
2	Denaturation	95	00:30	35
3	Annealing	50.2-69.3	00:30	
4	Extension	72	1:00	
5	Final extension	72	1:00	1
6	Hold	16	pause	1

Table 5.3 Direct probe sequences of six tandem repeat clusters.

No.	Probe name &type	Nucleotide sequence (5'-3')	Probe length (bp)
1	CL58[Btn]	TCACTAGAAATGACCAATTA TACTTGTTAGAGTGACAAAT GATGATCATTA	51
2	CL95[Btn]	CCTTTTTGGTATACTGTATAC TCTTTCGGTATACCTTGTTAT GTTTGGATCGAAG	55
3	CL102[Btn]	AACATACATAAATATTTGAT TGTAGAAAATATTTGAGCCG AAGCGGCCG	49
4	CL114[Btn]	AAACTGACTCGAAAAGGAA ATGATCGCTATCTTTTAGC	38
5	CL295[Btn]	TGATGATGATCATCACTAGA CATGACCAAATATACAAGTA AGAGTGATAAA	51
6	CL331[Btn]	GGCTACACCATGCGAAGTTC GGGGACGAACTTGCTTTAAA GAAAGGGGGGATGA	54

5.4 Results

5.4.1 Identification of tandem repeats by graph-based clustering

In the *P. hybrida* R27 reads, graph-based clustering with the program RepeatExplorer (Novák *et al.* 2013), and tandem repeat analysis (TAREAN) identified candidate tandem repeats by characteristic cluster graphs with star- or donut-like forms (Figures 5.2, 5.7, 5.9, 5.11, 5.13, 5.15, and 5.17; Appendix 5.1). In total, only eight candidate tandem repeats were identified in *P. hybrida* (*PhybR27*) (Table 5.1; two came from a division of a cluster), a number that is consistent with examination of sequence assembly data. Each tandem repeat had a characteristic monomer length and genome proportion.

Other bioinformatics approaches were also used to identify tandem-repeat motifs, including analysis of the most abundant *k*-mers (motifs *k* bases long, where $15<k<64$), and visualization of dot-plots covering the sequence assemblies to show tandem arrays by a series of lines of homology parallel to the major diagonal (Figure 5.4). These revealed the same tandem-repeat types with no additional motifs more than 50 bp long. Some of the identified tandem repeats marked the ends of assembled sequence scaffolds.

5.4.2 Genome organization

5.4.2.1 Assemblies and raw-reads

Analysis of sequence reads from the *Petunia* genomes (*P. axillaris* subsp *axillaris* N, *P. integrifolia* subsp *inflata* S6, the ancestors of the hybrid petunia; and the sister species *P. axillaris* subsp *parodii* S7) by mapping reads to reference motifs for each tandem repeat type identified in *P. hybrida* showed each was abundant, although there were substantial differences in relative copy numbers (Table 5.4; Figure 5.1). Copies of all tandem repeat motifs were found in the two whole genome assemblies, although collapsed by the assembly algorithms or being placed at ends of scaffolds (eg Figure 5.4); assembled lengths of tandem repeat arrays are also shown.

The genome sizes of the two parental and the hybrid-origin *Petunia* species are similar (1,400 Mb, all 2n=2x=14). The evolutionary differences were analysed in tandem repeat abundance.

The expectation would be stochastic processes would lead to differences in abundance in the parental species (separation 30 MYA), while *P. hybrida* would represent an average. Compared to *P. axillaris* subsp *axillaris* N, *P. integrifolia* subsp *inflata* S6 had a total repeat copy number of 30% less, while *P. hybrida* had 30% more; among individual tandem repeat types, some increased in copy number while some decreased. *P. axillaris* subsp *parodii* S7 (accurate genome size not known) had fewer tandem repeats, with a particular reduction (27%) in the most abundant Type I.

Table 5.4 Tandem repeat types with copy numbers and genome proportions within *Petunia* genomes reads plus array lengths in assemblies.

Types	**Consensus length (bp)**	**Proportion in *PhybR27* (%)**	**Copy numbers per genome**				**Location**	**Figures no.**	**PaxiN /PinfS6**	**PinfS6 + PaxiN /PhybR27**	**Copies in assembly & array size**	
			PaxiN	***PinfS6***	***PhybR27***	***PparS7***					***PaxiN***	***PinfS6***
TypeI-CL43A	157	0.21	314,586	316,813	389,959	111,816	Sub-telomeric	5.3	0.99	1.61	1,016	1,013
											5-10 kb	
TypeII-CL43B	1,002		147,144	318,384	401,170	89,908	Centromeric	5.5	0.46	1.16	2,486	1,702
TypeIII-CL58	51	0.150	140,814	40,308	98,868	42,894	Centromeric	5.8	3.49	1.83	4,252	4,316
											3 - 15 kb	
TypeIV-CL95	55	0.092	57,672	34,756	82,679	31,748	No spots	5.10	1.65	1.11	8,622	10,007
											1kb	
TypeV-CL102	49	0.080	45,909	22,652	87,498	22,131	Telomeric (TTTAGGG)	5.12	2	0.78	45	10
											1kb	
TypeVI-CL114	38	0.074	34,039	31,516	46,875	18,930	Centromeric	5.14	1	1.39	9,595	8,388
											1-2kb	
TypeVII-CL295	51	0.017	24,192	3,049	27,596	7,019	Centromeric and intercalary	5.16	7.93	0.98	9,082	5,411
											1-20kb	
TypeVIII-CL331	54	0.013	11,952	9,779	16,797	19,616	No spot	5.18	1.22	1.29	599	467
											0.1-1kb	
Total			776,308	777,257	1,151,442	344,062			0.99	1.34		

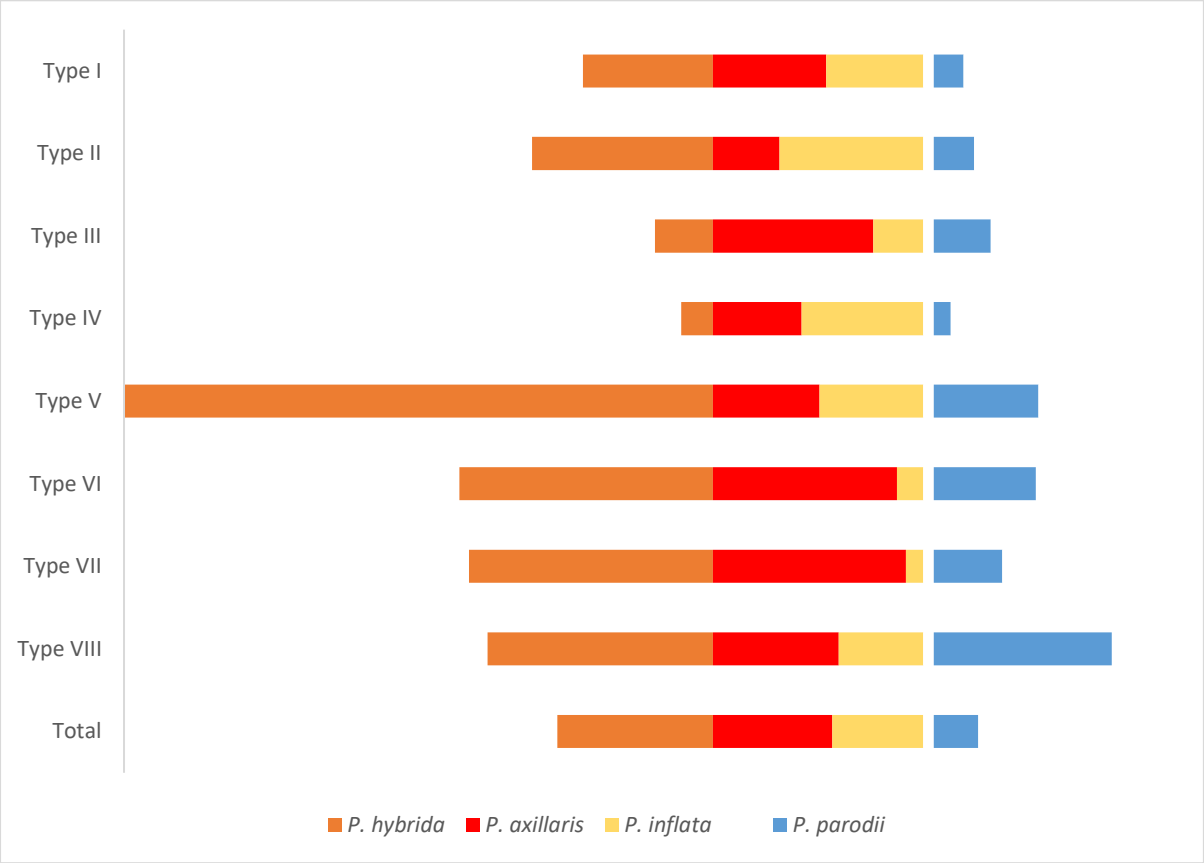


Figure 5.1 Relative copy number of each repeat type in *P. axillaris* subsp *axillaris* N (red) and *P. integrifolia* subsp *inflata* S6 (yellow), and the numbers in the hybrid species *P. hybrida* (orange) and sister species *P. axillaris* subsp *parodii* S7 (cyan). Each sequence shows a unique absolute (Table 5.4) and relative pattern of copy number variation.

Table 5.5 Ratios of mapped tandem repeat sequences to the four examined raw reads with 0% mismatch (absolute copy number) over 10% mismatch (relative copy number).

Types	***PaxiN***	**%**	***PinfS6***	**%**	***PhybR27***	**%**	***PparS7***	**%**
TypeI-CL43A	56,598/314,586	17.9	35,523/316,813	11.21	59,862/389,959	15.3	18,142/111,816	16.2
TypeII-CL43B	90/147,144	0.06	1/318,384	0.0003	1/401,170	0.000002	6/89,908	0.0066
TypeIII-CL58	12,688/140,814	9	3,022/40,308	7.4	6,958/98,868	7	3,170/42,894	7.39
TypeIV-CL95	285/57,672	0.49	465/34,756	1.3	2,318/82,679	2.8	757/31,748	2.38
TypeV-CL102	19,575/45,909	42	7,865/22,652	34.7	62,927/87,498	71.9	8,054/22,131	36.3
TypeVI-CL114	11,811/34,039	34.6	6,534/31,516	20.7	16,325/46,875	34.8	6,295/18,930	33.2
TypeVII-CL295	3,185/24,192	13	261/3,049	8.5	2,058/27,596	7.4	1,251/7,019	17.8
TypeVIII-CL331	3,613/11,952	30	3,006/9,779	30.7	4,573/16,797	27	3,733/19,616	19

In situ hybridization was used to find the location of each repeat on metaphase chromosomes. Except for the least abundant sequence *TypeVIII-CL331* (where no hybridization signal was observed), the chromosomal study showed the motifs were repeated and presented at multiple discrete loci.

5.4.3 Chromosomal organization of tandem repeats

5.4.3.1 *Type I- CL43A*

CL43A is a tandem repeat with consensus monomer length of 157 bp. It appears in the genome assembly 5-10 kb arrays, and marks the end of several scaffolds. There are 1,013 and 1,016 copies in the *PinfS6* and *PaxiN* assemblies respectively, but as expected, these represent a collapse of monomers from raw reads, with more than 300,000 copies per genome in *PhybR27*, *PaxiN* and *PinfS6*, and a lower content in *PparS7* (Table 5.4; Figure 5.1). The TAREAN and RepeatExplorer clustering graphs for *Type I-CL43A* and *Type II-CL43B* are condensed star-like shapes (Figure 5.2). *In situ* experiments showed that this sequence has strong (in agreement with the high copy number), subtelomeric signals, showing two dots on most chromosomes ends in the six *Petunia* species (Figure 5.3). Interestingly, this tandem repeat is sharing exactly similar features of the published repeat (169 bp) in Bombarely *et al.* (2016) that extracted by *k*-mer analysis.

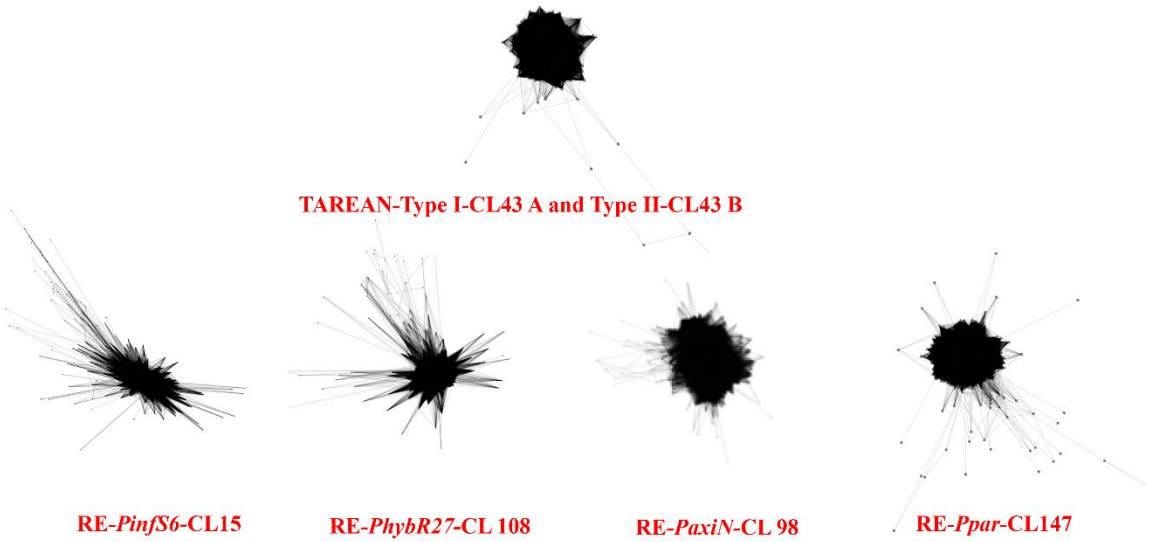
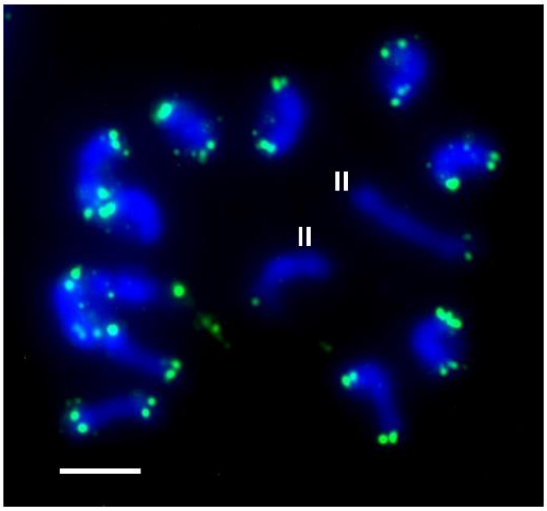
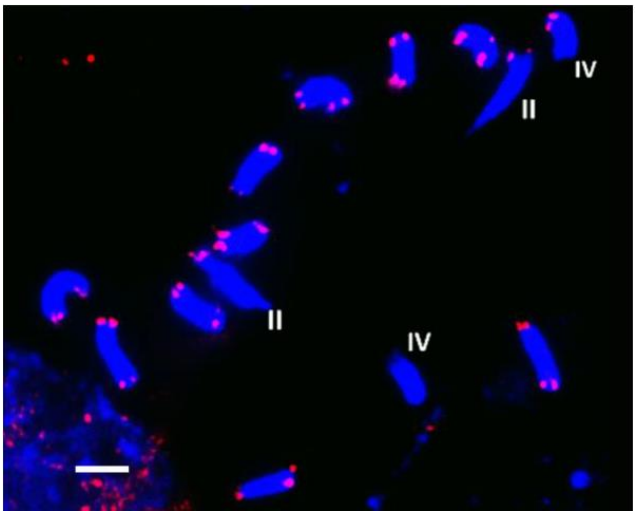


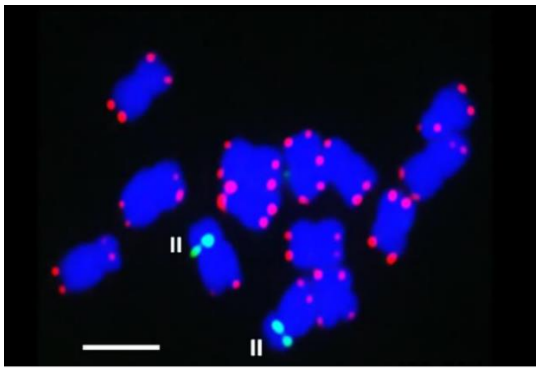
Figure 5.2 Graphs of tandem repeat clusters of *Type I-CL43A* and *Type II-CL43B*, showing similar condensed star like shapes with high similarity to other genome reads clusters (*PinfS6*-CL15, *PhybR27*-CL108, *PaxiN*-CL98, and *Ppar*-CL147) based on the results of TAREAN and RepeatExplorer pipelines.



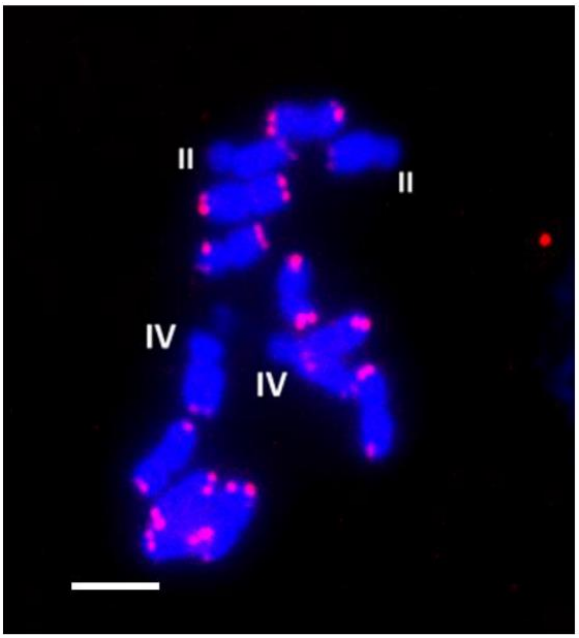
A



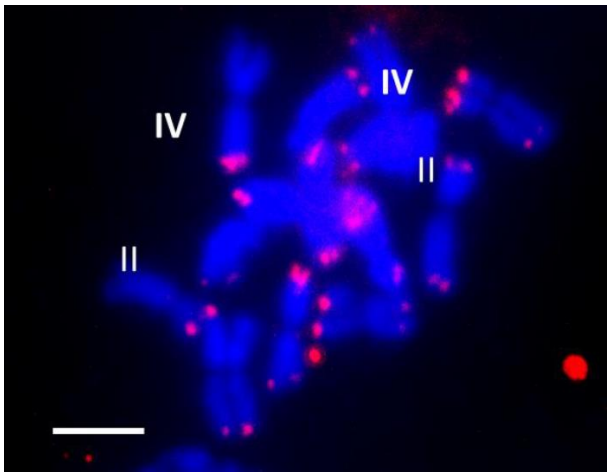
B



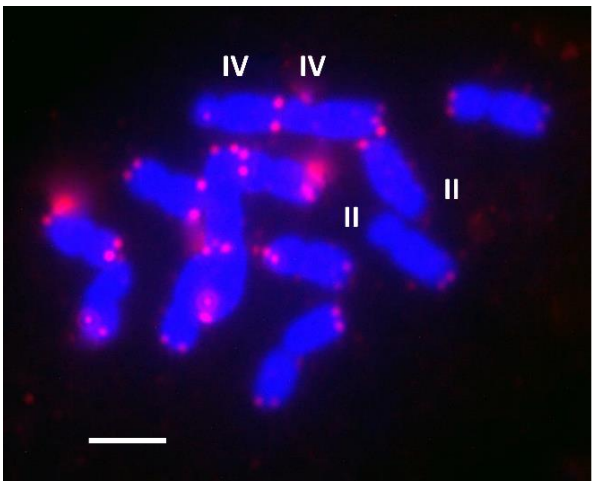
C



D



E



F

Figure 5.3 *In situ* hybridization of tandem repeat *Type I-CL43A* probe to metaphase chromosomes (2n=14) from six *Petunia* species, showing subtelomeric signals as pairs of dots on all telomeres, except one chromosome pair (Ch.II) on *P. axillaris* N (A) and *P. hybrida* Rdc (C), two pairs (Ch.II and Ch.IV) on *P. integrifolia* subsp *inflata* S6 (B), *P. hybrida* V26 (D), *P. hybrida* W138 (E) and *P. axillaris subsp parodii* S7 (F) with signals only at the end of the long arms, and no telomeric signal was detected on the short arms carrying the 5S rDNA arrays. Chromosomes were stained with DAPI (blue) and the probe was detected red for *P. integrifolia* subsp *inflata* S6 (B), *P. hybrida* Rdc (C), *P. hybrida* V26 (D), *P. hybrida* W138 (E) and *P. axillaris* subsp *parodii* S7 (F), while detected green was used for labelling this probe in *P. axillaris* subsp *axillaris* N (A) and to the 5S rDNA on one pair of *P. hybrida* Rdc (C). Bar = 10 μm for *P. axillaris* subsp *axillaris* N (A), *P. hybrida* W138 (E) and *P. hybrida* Rdc (C), and 5 μm for *P. hybrida* V26 (D), *P. axillaris* subsp *parodii* S7 (F), and 3 μm for *P. integrifolia* subsp *inflata* S6 (A).

5.4.3.2 *Type II- CL43B*

This repeat assembled with a long monomer of 1,002 bp, and was identified with 2,486 copies in *PaxiN* and 1,702 copies in *PinfS6* genome assemblies, while 147,144 and 318,384 copies (and 401,170 copies in *PhybR27*) per genome were identified in the reads. The assemblies showed the 1 kb tandem-repeat array, including interspersion with other repeat arrays of about 20 kb (Figure 5.4). The tandem repeat had strong, quite broad, signals at the centromere regions of all *Petunia* species chromosomes (Figure 5.5). Both Type I and Type II repeats shared the same graph based cluster feature as a condensed star like shape (Figure 5.2), had 79.8% pairwise identity, and were located together in some scaffolds (*PaxiN162Scf00160* scaffold) (Figure 5.6), *in situ* hybridization showed contrasting subtelomeric and pericentromeric locations for *Type I-CL43A* and *Type II-CL43B* respectively.

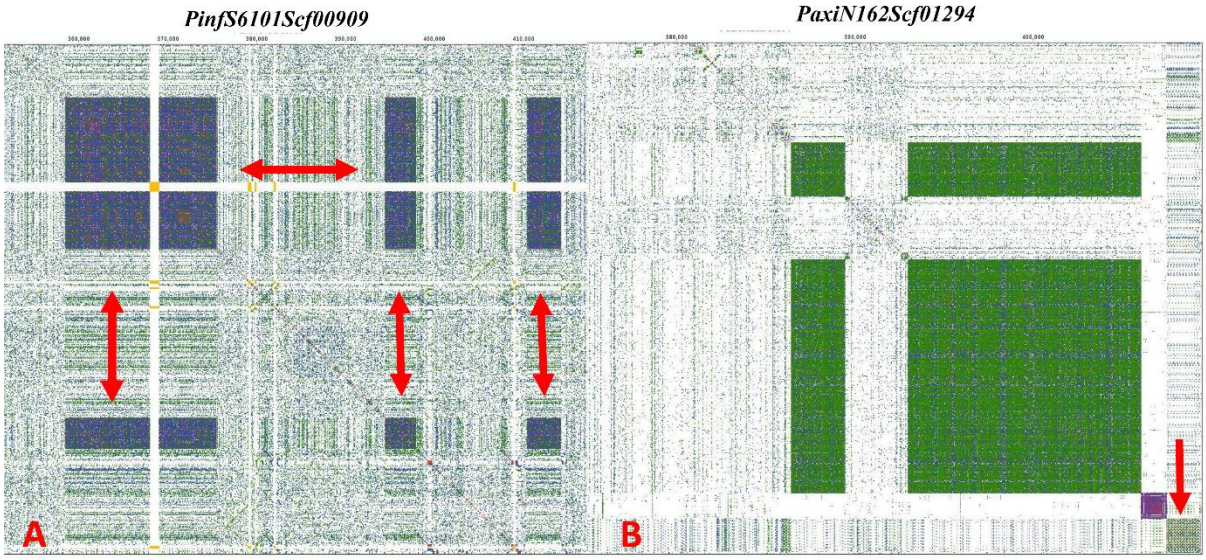
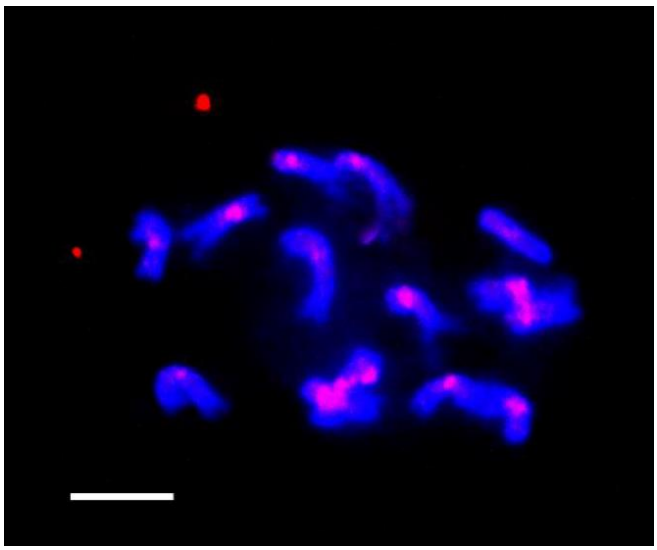
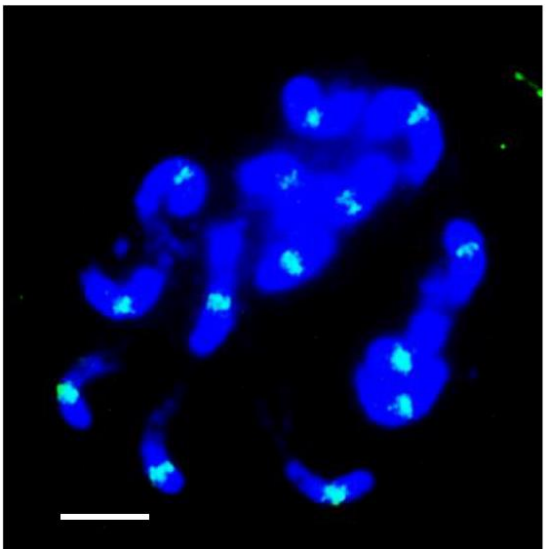


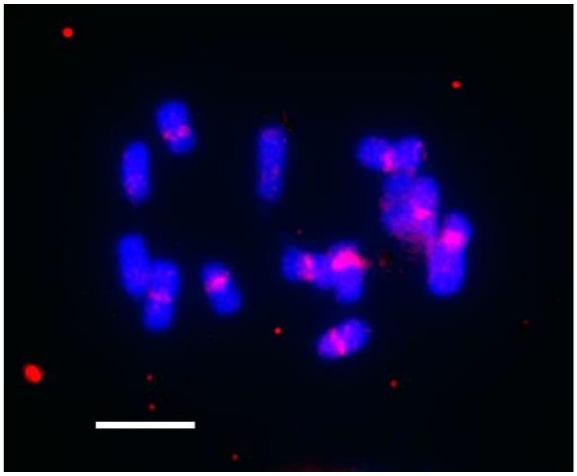
Figure 5.4 Self-dot plots of tandem repeat *Type II- CL43B* in *PinfS6101Scf00909* scaffold at *P. integrifolia* subsp *inflata* S6 assembly (A) showing tandem structure between 360,000 bp and 415,000 bp (total size of this scaffold is about 1,820,000 bp), and *PaxiN162Scf01294* scaffold at *P. axillaris* subsp *axillaris* N assembly (B) showing about 2 kb array at the end of this scaffold. Arrows in (A) indicate miss-assembled array of *Type II- CL43B* while in (B), single arrow indicates a small array at the end of the scaffold.



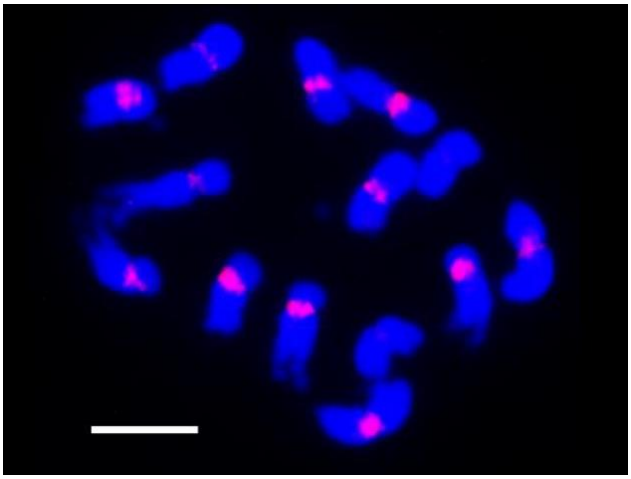
A



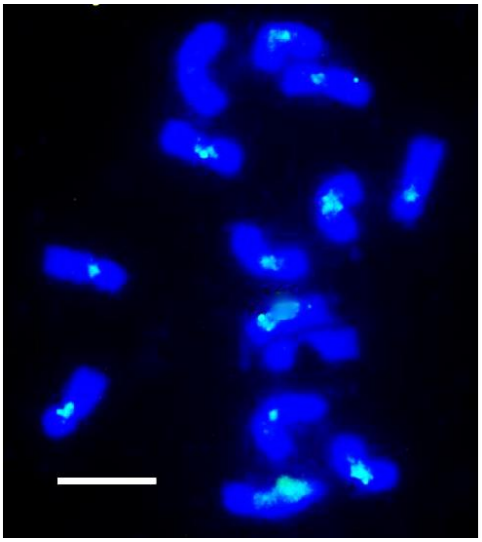
B



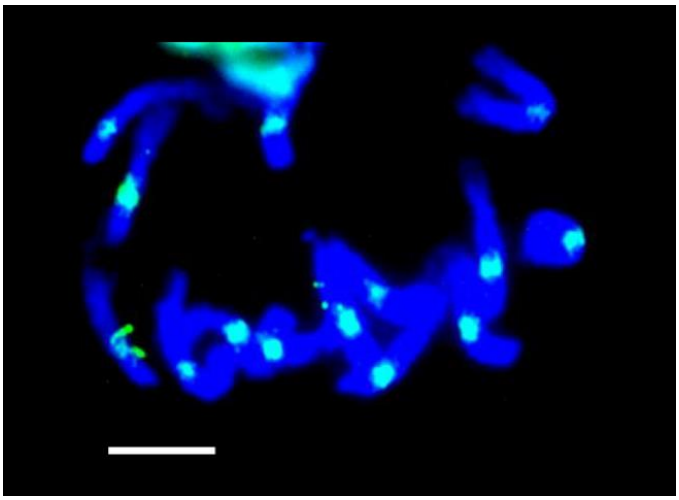
C



D



E



F

Figure 5.5 Metaphase chromosomes of *Petunia* (2n=14) species probed with tandem repeat *Type II-CL43B* showing dispersed and strong signals at the centromere regions of all *Petunia* species chromosomes. Chromosomes stained with DAPI (blue) and the probe detected red for *P. axillaris* subsp *axillaris* N (A), *P. hybrida* Rdc (C) and *P. hybrida* V26 (D), while in *P. integrifolia* subsp *inflata* S6 (B), *P. hybrida* W138 (E) and *P. axillaris* subsp *parodii* S7(F), the probe was detected green. Bar = 10 µm.

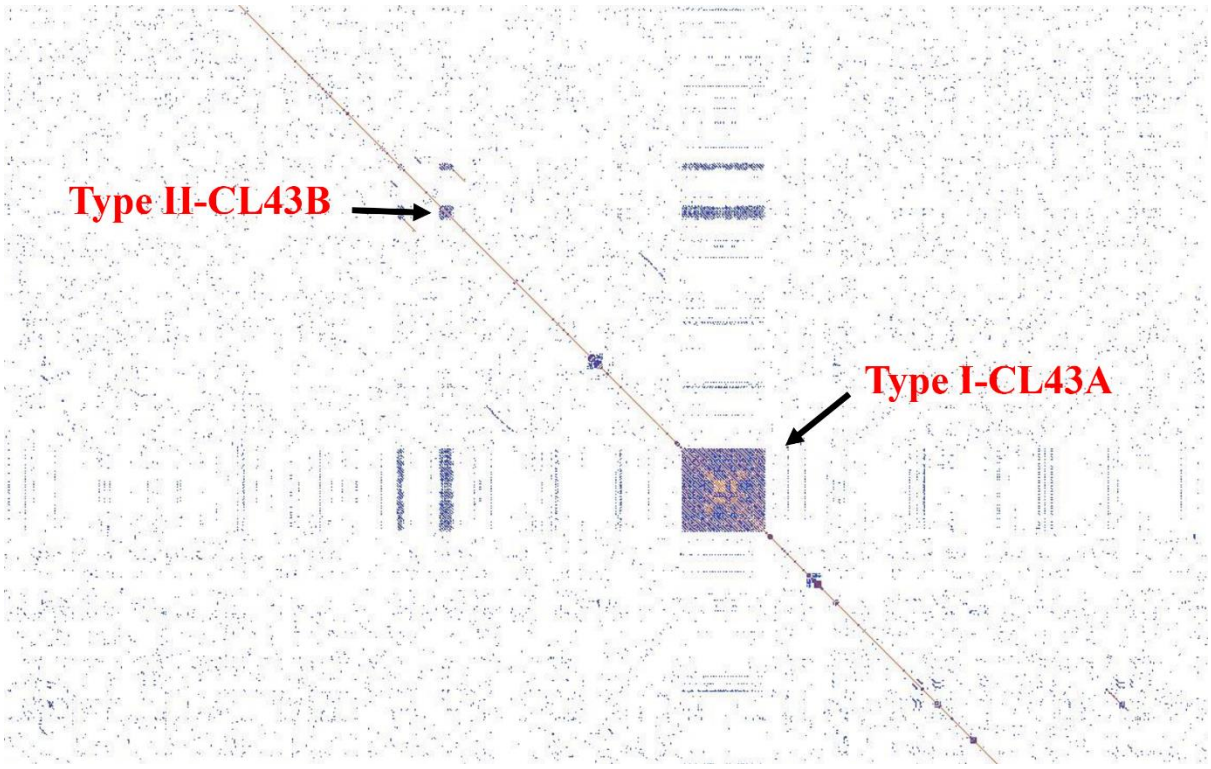


Figure 5.6 Self-dot plot of *PaxiN162Scf160* scaffold, showing tandem repeat *Type I-CL43 A* and *Type II-CL43 B* positions as 10 kb and 1 kb arrays respectively.

5.4.3.3 *Type III-CL58*

This tandem repeat is abundantly represented in the assemblies and raw reads with 0.150 % genome proportion. It showed variable sizes of arrays between 3 kb and 15 kb within small scaffolds, with 23% (*PinfS6*) and 7% (*PaxiN*) of the reads being represented in assemblies, presumably representing collapse of reads of the repeat in the assemblies, and different copy

numbers in each species (Table 5.4). In TAREAN, the repeat graph showed a solid star-like shape with multiple arms in comparison with shapes from RepeatExplorer that presented this star like feature in *PinfS6, PhybR27* and *PparS7* with long arm which branches with one loop (Figure 5.7). On chromosomes, a centromeric signal was found on four pairs of *P. axillaris* subsp *axillaris* N chromosomes (Ch. II, III, IV and V), with two pairs in *P. integrifolia* subsp *inflata* S6, *P. hybrida* Rdc (Ch. II and III) and *P. hybrida* V26 (Ch.II and IV), and three pairs on *P. hybrida* W138 (Ch.II, IV and V; hence showing intraspecific variation) and *P. axillaris* subsp *parodii* S7 (Ch.II, III and IV) (Figure 5.8).

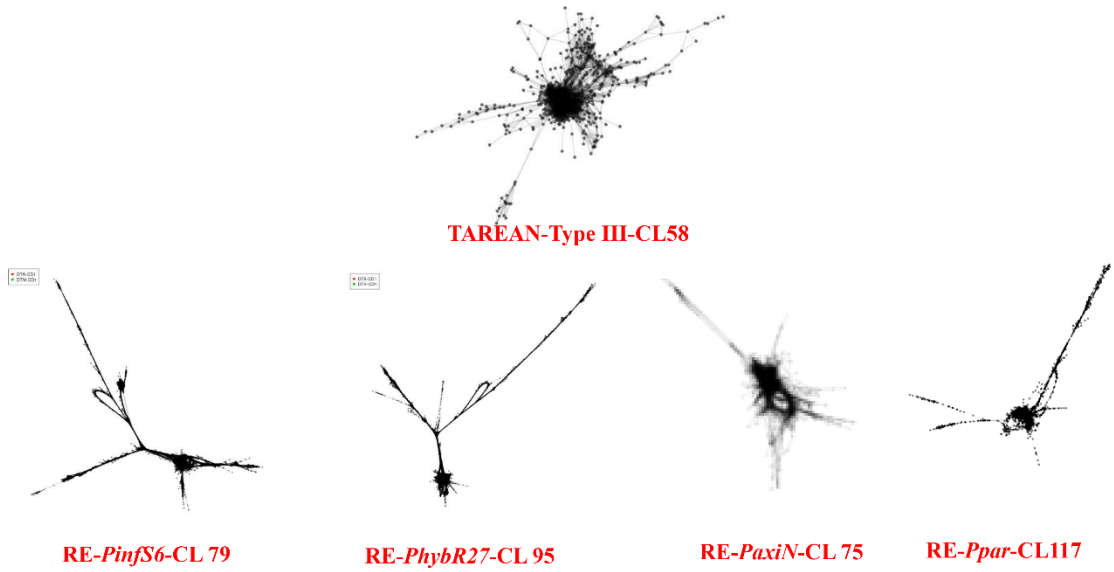
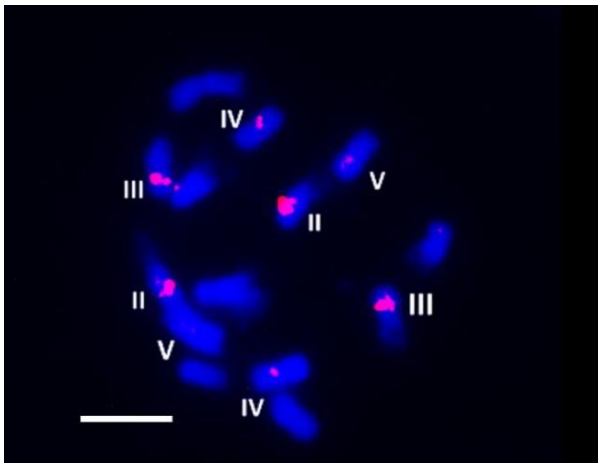
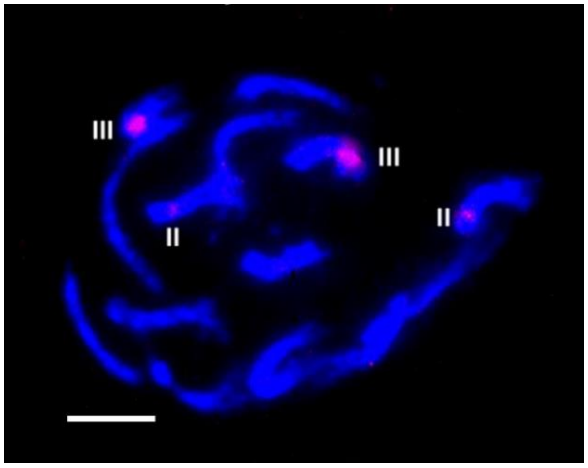


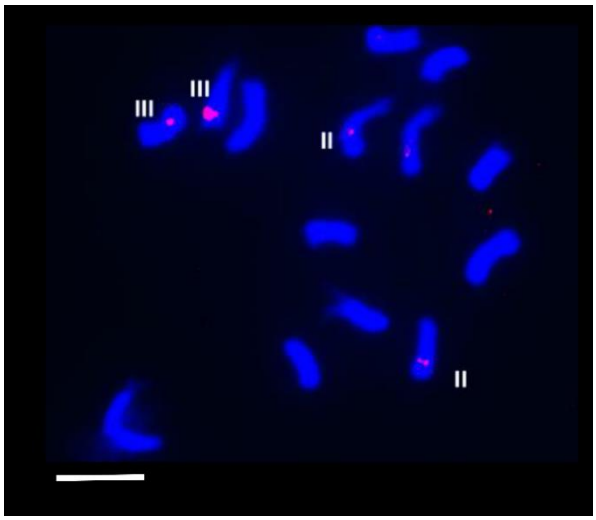
Figure 5.7 Tandem repeat *Type III-CL58* cluster graphs, showing high similarity in *PinfS6*-CL79, *PhybR27*-CL95 and *Ppar*-CL117 graphs, while *PaxiN*-CL75 is more likely to TAREAN graph.


IV
III
V
II
II
III
V
IV

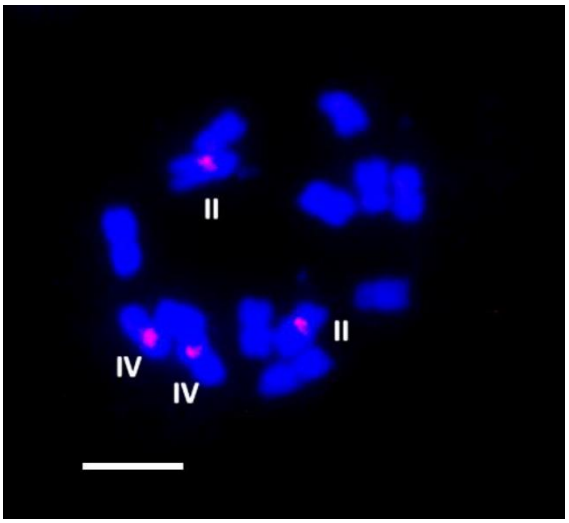
A


III
III
II
II

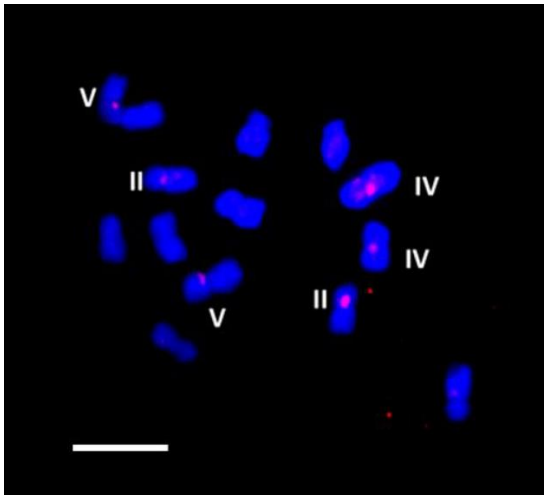
B


III
III
II
II

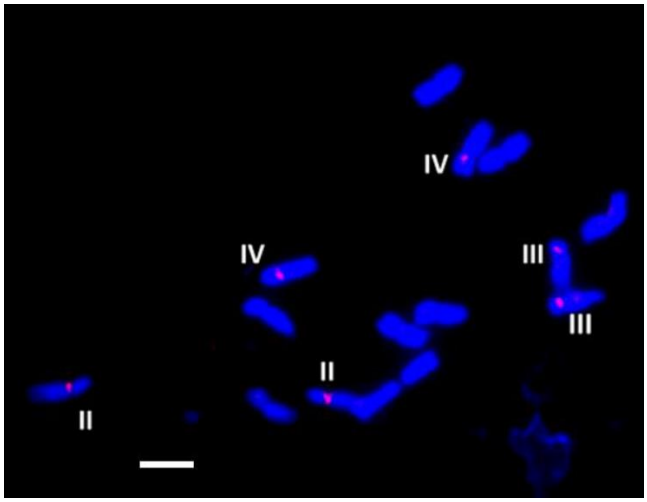
C


II
II
IV
IV

D


V
II
IV
IV
II
V

E


IV
IV
III
III
II
II

F

Figure 5.8 Metaphase chromosomes of *Petunia* (2n=14) species probed with tandem repeat *Type III-CL58* detected red shows that centromeric signals clearly found on four pairs of *P. axillaris* subsp *axillaris* N (A) (Ch.II, III, IV and V). Two pairs are seen in *P. integrifolia* subsp *inflata* S6 (B), *P. hybrida* Rdc (C) (Ch.II and III) and *P. hybrida* V26 (D) (Ch.II and IV) have centromeric signals; there are three pairs on *P. hybrida* W138 (E) (Ch.II, IV and V) and *P. axillaris* subsp *parodii* S7 (F) (Ch.II, III and IV). Bar = 10 µm for all species except *P. axillaris* subsp *parodii* S7 (F) that has 5 µm.

5.4.3.4 *Type IV-CL95*

Despite of 0.092% genome proportion and the abundant sequences of this repeat within *PinfS6* and *PaxiN* assemblies, it has low numbers of copies in *PhybR27*and *PparS7* (Figure 5.1). This repeat had small arrays with about 1 kb that were widely distributed over assemblies. Cluster graphs had a solid circle with few extensions especially in *PaxiN* and *PparS7* reads; *PinfS6* and *PhybR27* were similar, while *PaxiN* had a smaller central cluster and four branches (Figure 5.9). Consistent with the copy number and distribution in assemblies, and weak and dispersed signals were seen on chromosomes (Figure 5.10).

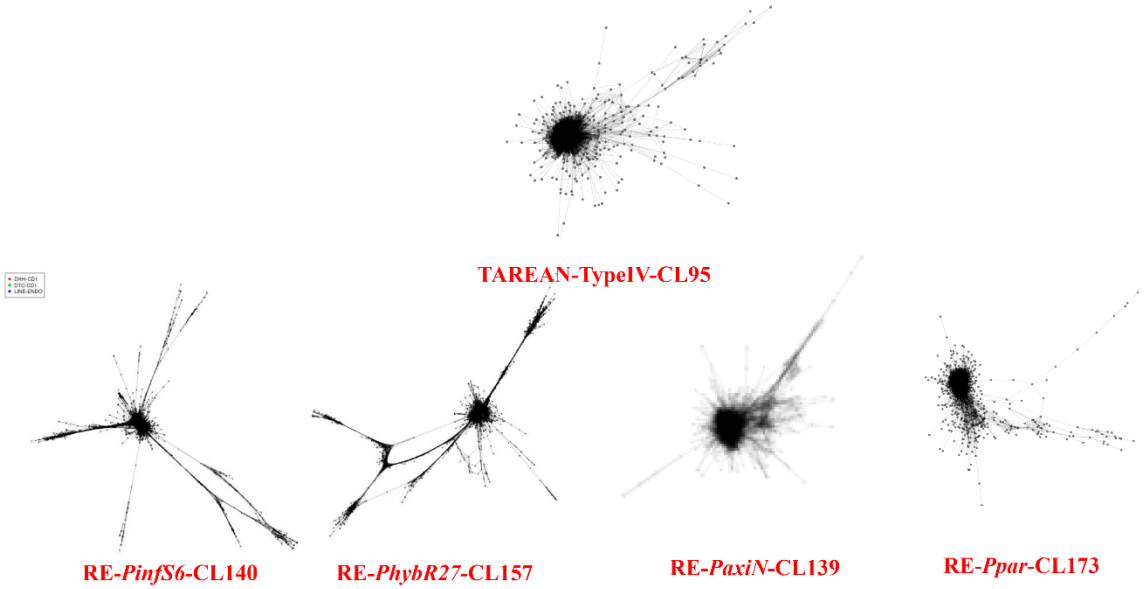
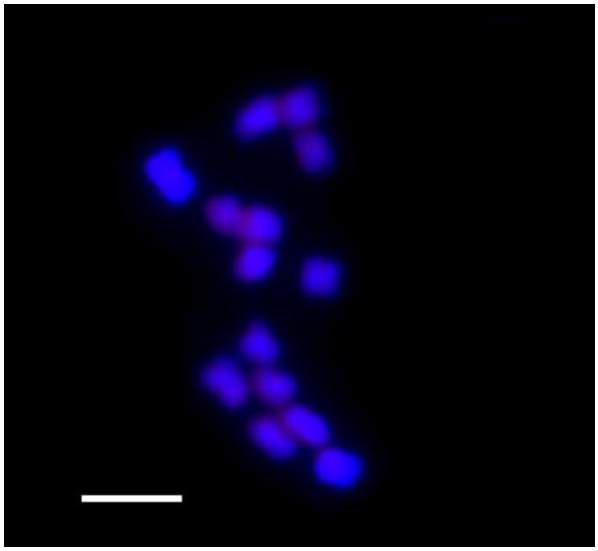
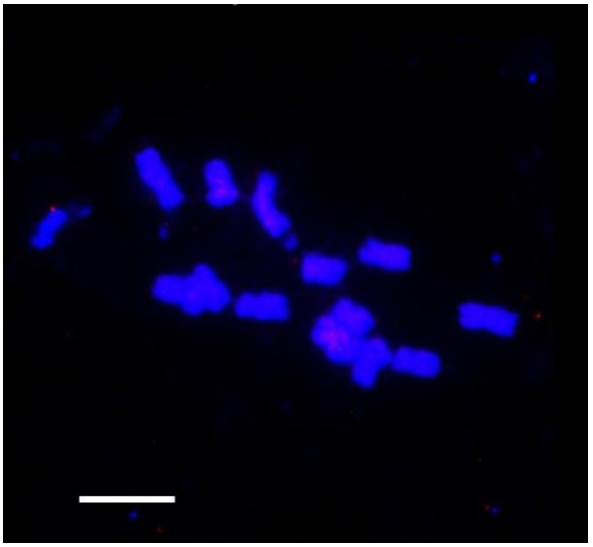


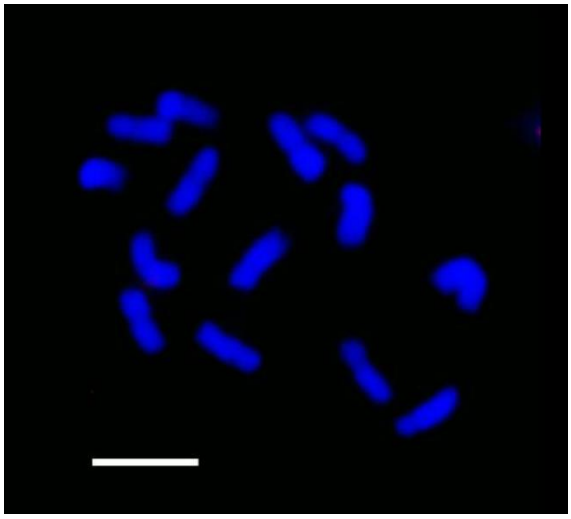
Figure 5.9 Graphs of four genomes clusters of tandem repeat *Type IV-CL95* that compared together with TAREAN shape, representing spitting images for *PinfS6*-CL140 and *PhybR27*-CL157 clusters with smaller central cluster and four branches, two branches are coming together at the middle, and between *PaxiN*-CL139, *Ppar*-CL173 and TAREAN with a solid circle and few extensions.



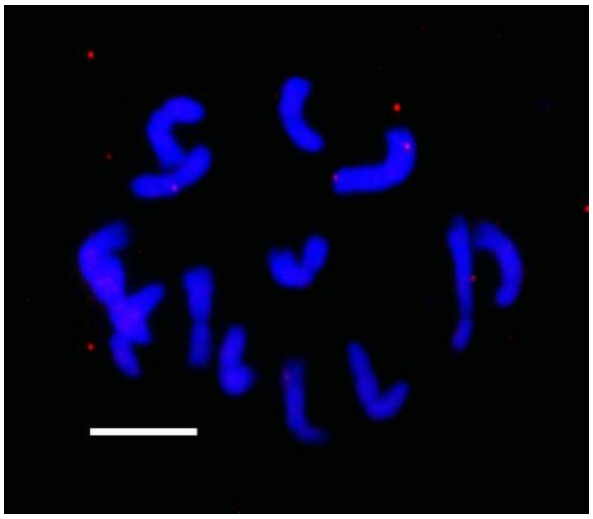
A



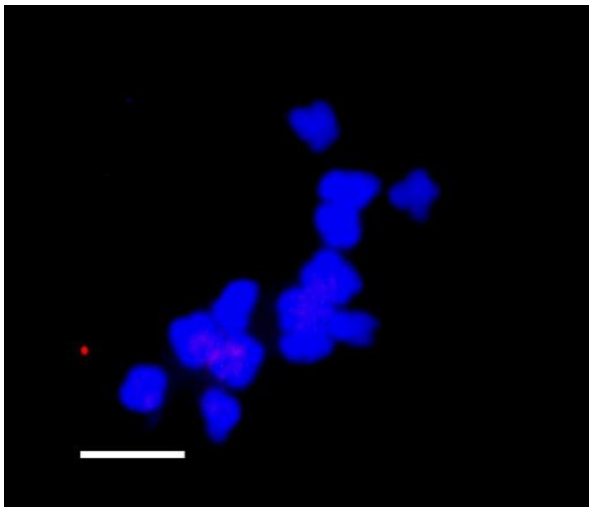
B



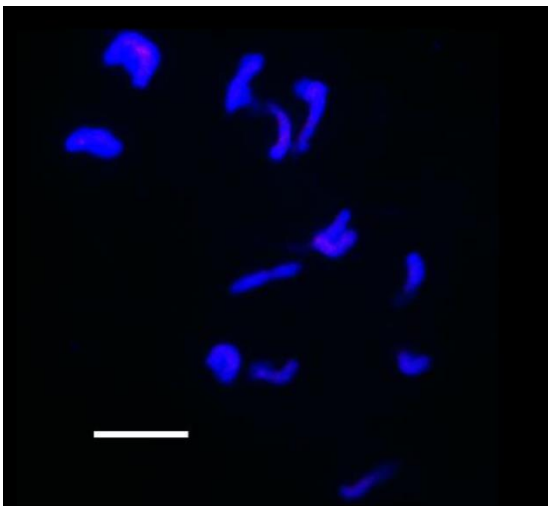
C



D



E



F

Figure 5.10 Metaphase chromosomes of *Petunia* (2n=14) species probed with repeat *Type IV-CL95* shows weak and dispersed signals (red) over all chromosomes (blue). *P. axillaris* subsp *axillaris* N (A), *P. integrifolia* subsp *inflata* S6 (B), *P. hybrida* Rdc (C), *P. hybrida* V26 (D), *P. hybrida* W138(E) and *P. axillaris* subsp *parodii* S7 (F). Bar = 10 µm for all species.

5.4.3.5 *Type V-CL102*

Type V was poorly represented within *PinfS6* and *PaxiN* assemblies (as small 1 kb arrays), although copies were abundant in all raw reads (0.08% genome proportion). The sequence included many copies of the telomeric repeat (TTTAGGG) and variants, and the graph had two connected solid circles or a circular pattern (Figure 5.11). *In situ* hybridization signals were strong on the telomere region of one pair of all *Petunia* species chromosomes, showing highly condensed signal on just the short arm of chromosome II (Figure 5.12).

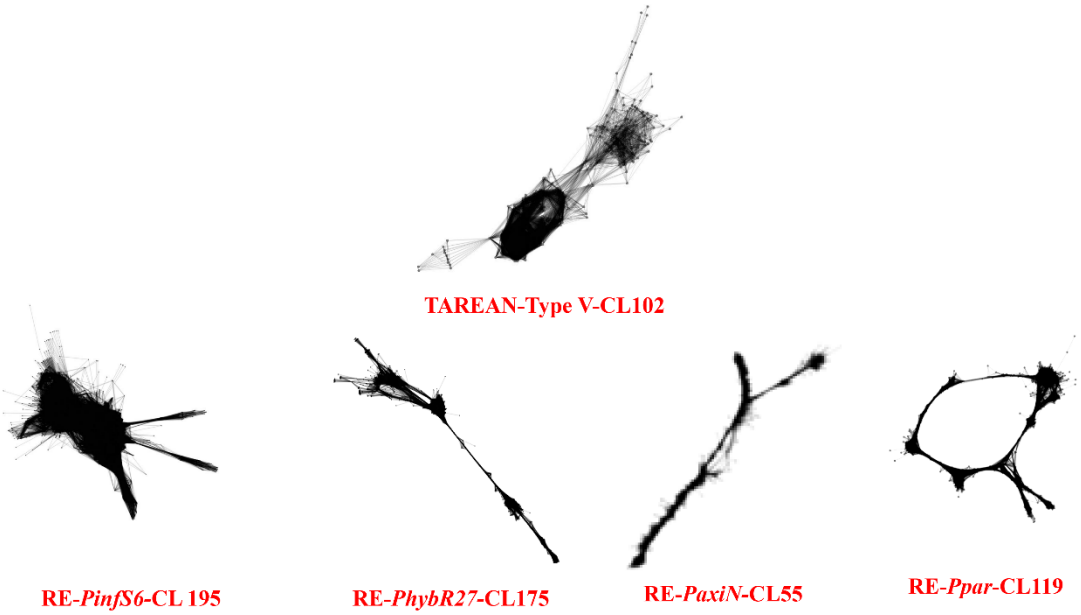
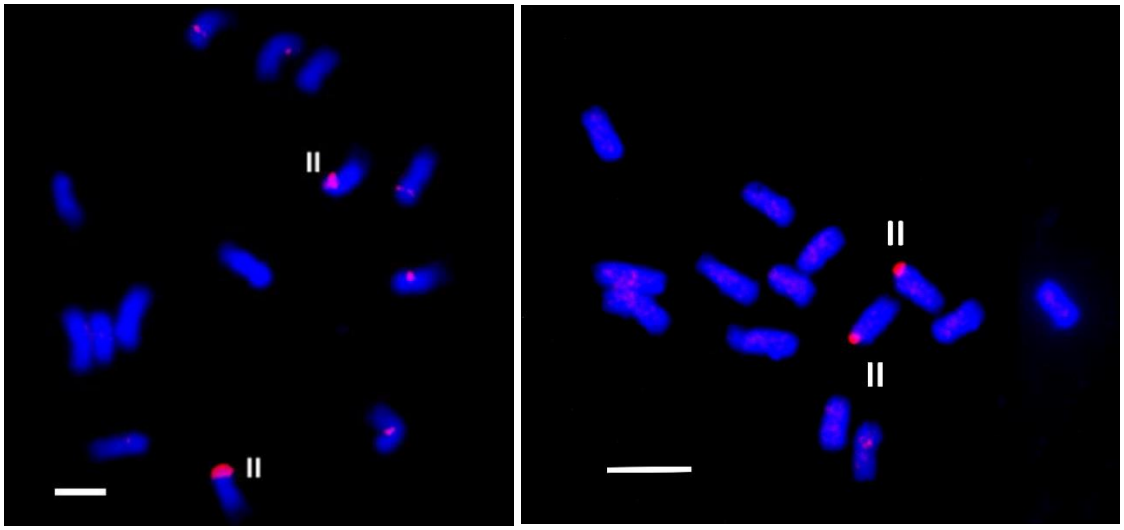
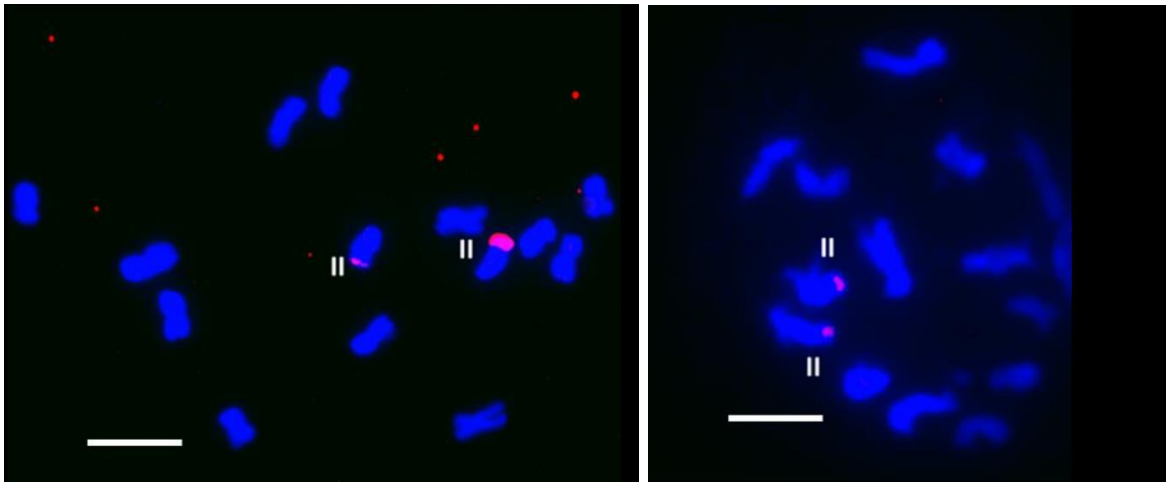


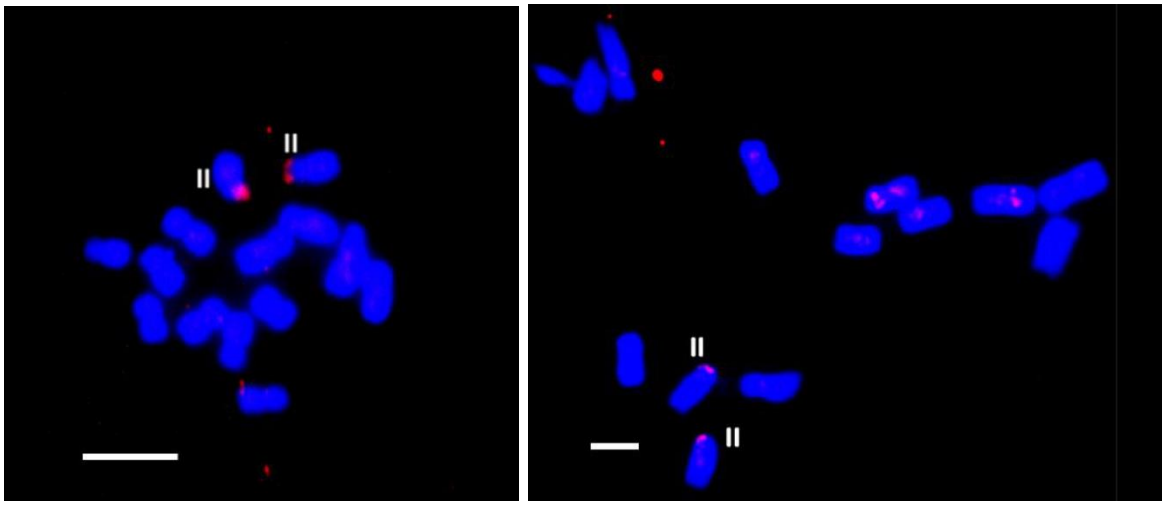
Figure 5.11 Tandem repeat *Type V-CL102* graphs, showing two connected solid circles for TAREAN and *PinfS6*-CL195, and a bit different from both *PaxiN*-CL55 and *PhybR27*-CL175 clusters shapes that look very close to each other, while *Ppar*-CL119 shows circular shape surrounded by condensed nodes.



A B



C D



E F

Figure 5.12 Metaphase chromosomes of *Petunia* (2n=14) species probed with tandem repeat *Type V-CL102*, showing strong signals (red) on the telomere region on the short arm of chromosome II. *P. axillaris* subsp *axillaris* N (A), *P. integrifolia* subsp *inflata* S6 (B), *P. hybrida* Rdc (C), *P. hybrida* V26 (D), *P. hybrida* W138 (E) and *P. axillaris* subsp *parodii* S7 (F). Bar =5 µm for A and F, and 10 µm for the rest.

5.4.3.6 *TypeVI-CL114*

Abundant sequences have been shown in the two assemblies with notably higher copy numbers in *PaxiN* and *PhybR27* than other two species. For this tandem repeat, arrays are between 1 to 2 kb with 0.074% genome proportion. Graphs showed that *PhybR27* has a solid circle with extensions linked and a ring shape, and *PparS7* shows similar graph without ring, while this ring shape looks bigger in *PaxiN*, and very complex feature of *PinfS6* (Figure 5.13). *P. axillaris* subsp *axillaris* N, *P. integrifolia* subsp *inflata* S6, *P. hybrida* V26, and *P. axillaris* subsp *parodii* S7 had centromeric signals on two chromosome pairs (Ch.II and IV) with variable site strengths, while chromosomes of *P. hybrida* Rdc and *P. hybrida* W138 only showed one pair (Ch.IV) (Figure 5.14).

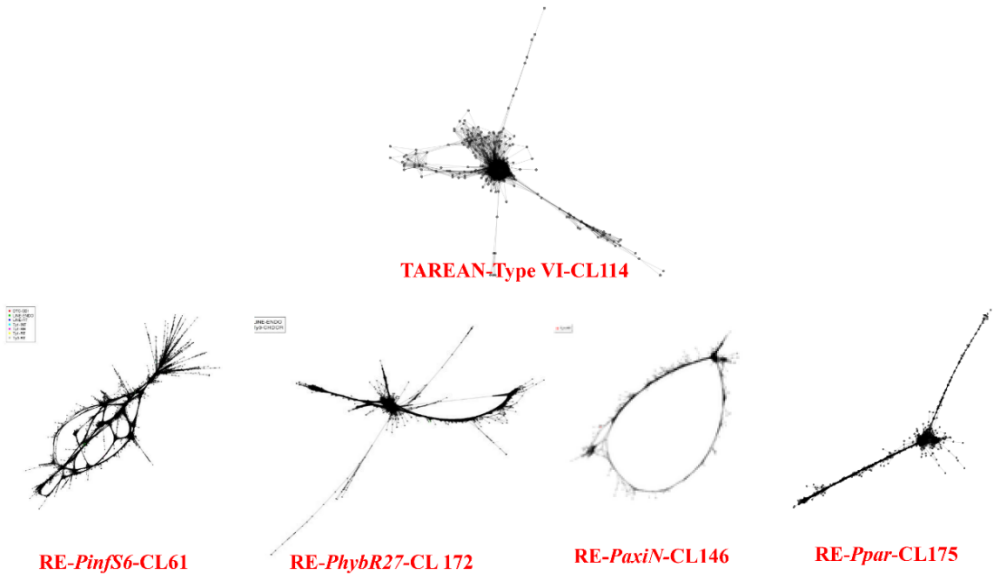
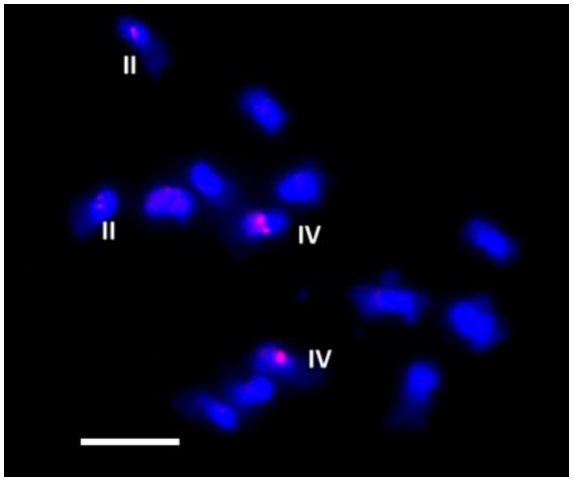
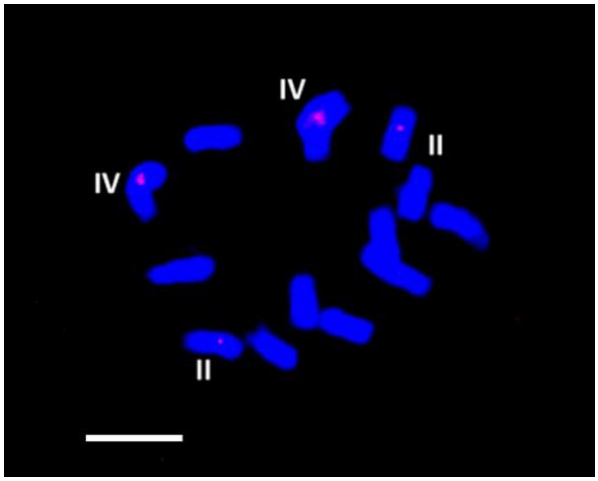


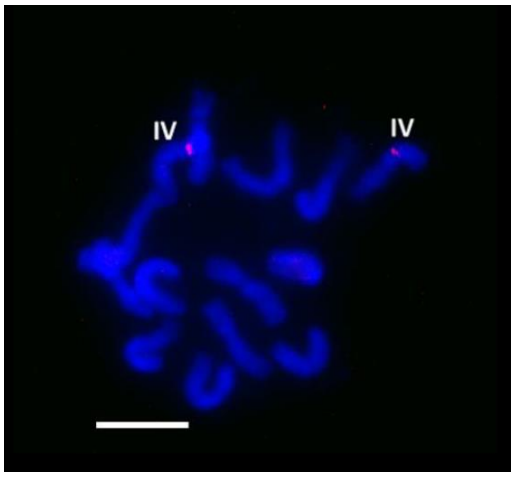
Figure 5.13 Graphs show tandem repeat Type VI-CL114 shapes that have a solid circle with extensions linked with a ring shape for PhybR27-CL172, and Ppar-CL175 shows similar graph without ring, while this ring shape looks bigger in PaxiN-CL146, and very complex feature of PinfS6-CL61.



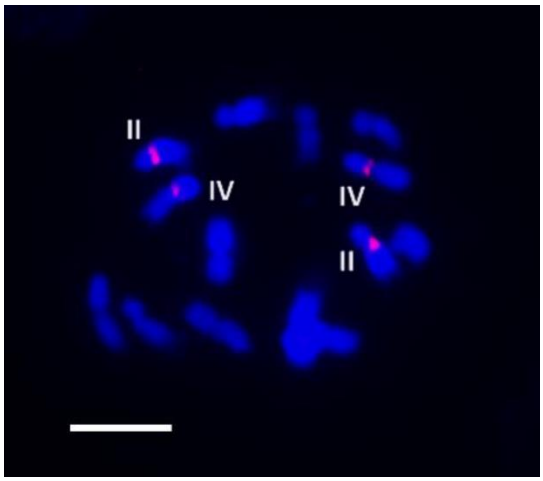
A



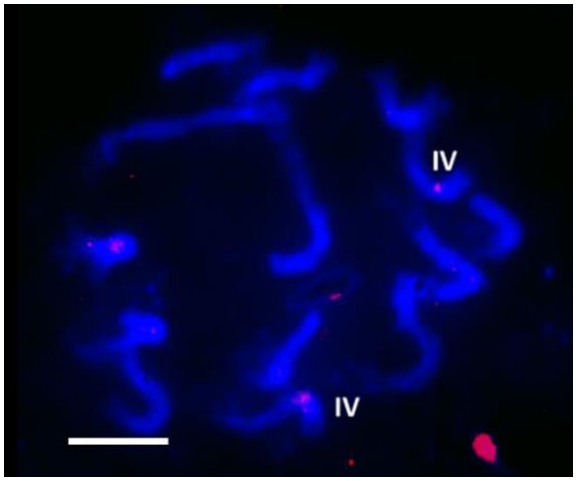
B



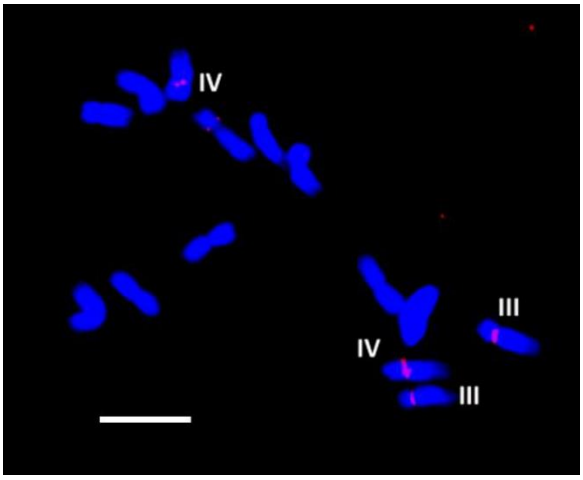
C



D



E



F

Figure 5.14 Metaphase chromosomes of Petunia (2n=14) species probed with tandem repeat Type VI-CL114 revealed that *P. axillaris* subsp *axillaris* N (A), *P. integrifolia* subsp *inflata* S6 (B), *P. hybrida* V26 (D), and *P. axillaris* subsp *parodii* S7 (F) chromosomes have centromeric signals (red) on two pairs (Ch.II and IV) with variable site strengths, while chromosomes of *P. hybrida* W138 (E) and *P. hybrida* Rdc (C) only has one pair (Ch.IV). Bar = 10 µm.

5.4.3.7 *Type VII-CL295*

This repeat was more abundant in *PhybR27* and *PaxiN* than *PinfS6* and *PparS7* with 0.017% genome proportion and variable arrays sizes between 1 to 20 kb. A star like graph shape of TAREAN is not as condensed as in the RepeatExplorer clusters, and graphs of *PhybR27* and *PaxiN* are similar (Figure 5.15). All *Petunia* species have intercalary strong sites on the short arm of Ch.I in addition to a centromeric signal on Ch.II in *P. axillaris* subsp *axillaris* N, *P. integrifolia* subsp *inflata* S6 and *P. hybrida* Rdc (Figure 5.16).

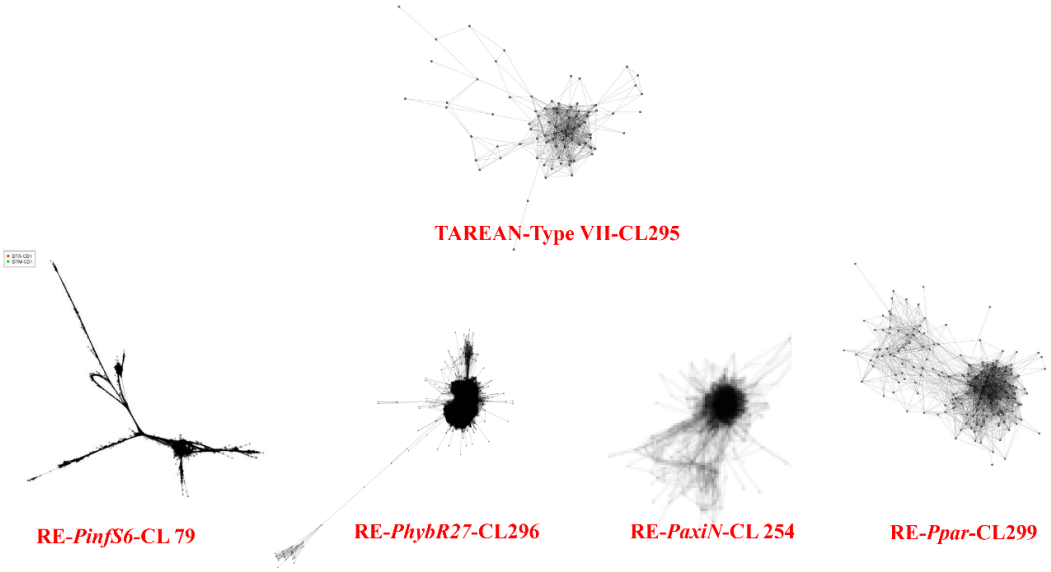
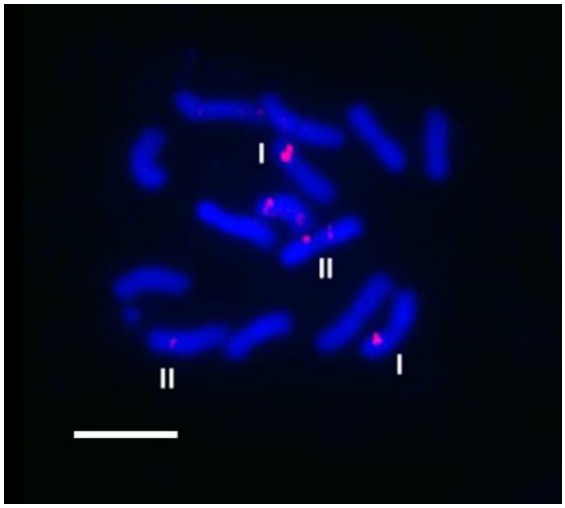
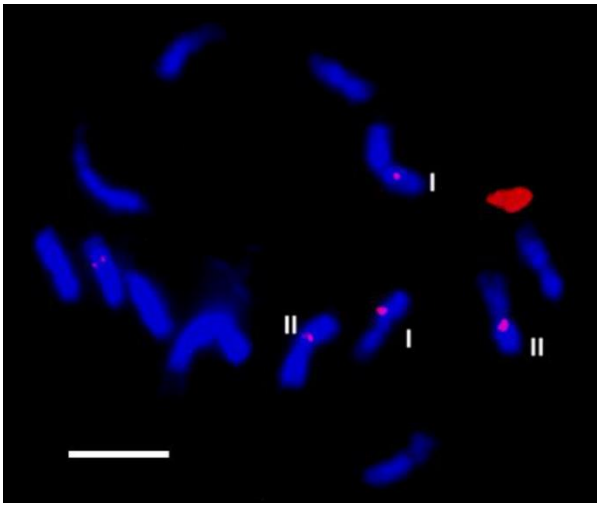


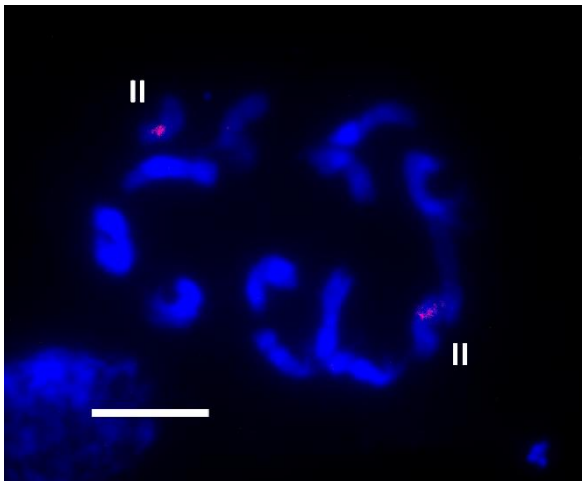
Figure 5.15 Graphs of tandem repeat *Type VII-CL295* show star like shape of TAREAN that not quite condensed like in the RepeatExplorer clusters except *Ppar*-CL299 that looks quite close, *PhybR27*-CL296 and *PaxiN*-CL254 shapes are very similar, while *PinfS6*-CL79 has linear branches linked with the condensed spot that interestingly same shape of *PinfS6*-CL79 in the tandem repeat *Type III-CL58* (Figure 5.7).



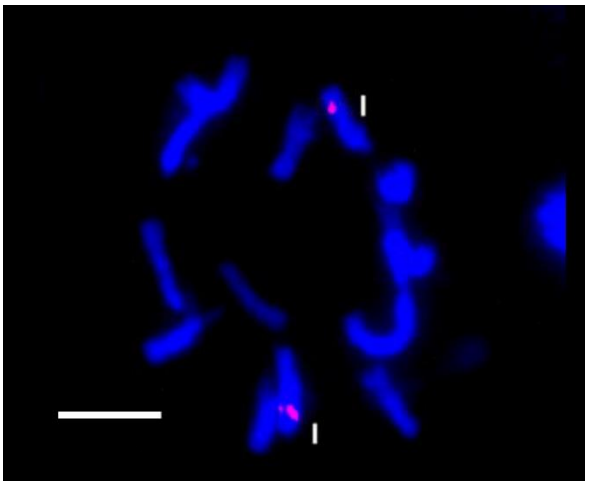
A



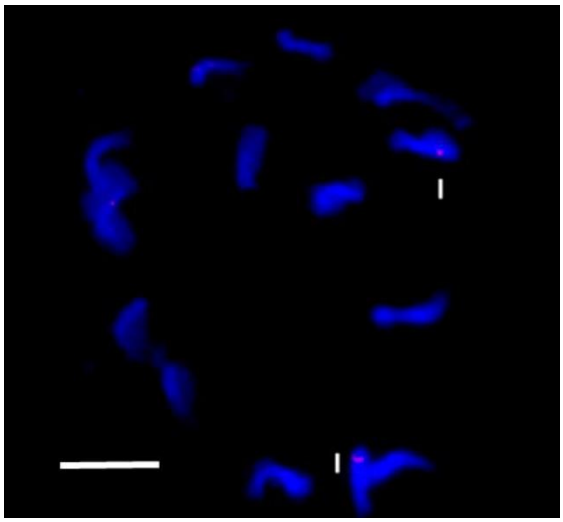
B



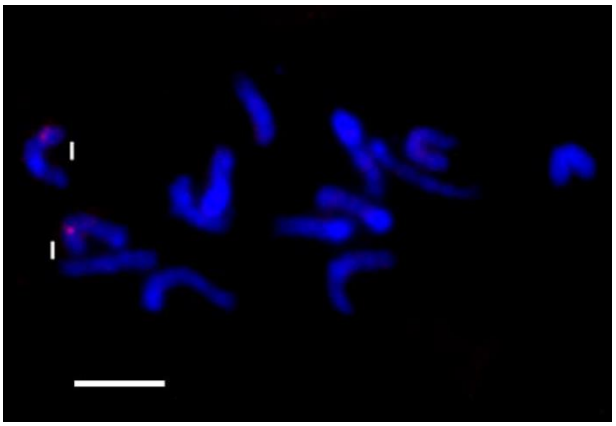
C



D



E



F

Figure 5.16 Metaphase chromosomes of *Petunia* (2n=14) species probed with tandem repeat *Type VII-CL295* shows intercalary strong sites on the short arm of Ch.I in *P. axillaris* subsp *axillaris* N (A), *P. integrifolia* subsp *inflata* S6 (B), *P. hybrida* Rdc (C), *P. hybrida* V26 (D), *P. hybrida* W138 (E), and *P. axillaris subsp parodii* S7 (F), in addition to a centromeric signal on Ch.II only in *P. axillaris* subsp *axillaris* N (A), *P. integrifolia* subsp *inflata* S6 (B), and *P. hybrida* Rdc (C). The probe was labelled with biotin 11- dUTP (red). Bar = 10 µm.

5.4.3.8 *Type VIII-CL331*

A small number of copies of this repeat were found in *PinfS6* and *PaxiN*, with only small arrays in the assembly. From TAREAN data, this repeat has donut graph shape, not as condensed as others in RepeatExplorer (Figure 5.17). No signal was detected by *in situ* hybridization on chromosomes over all petunias (Figure 5.18), confirming the low abundance. The result suggests that clusters with this abundance and lower can be discounted from analysis of major abundant tandemly repeat satDNA sequences.

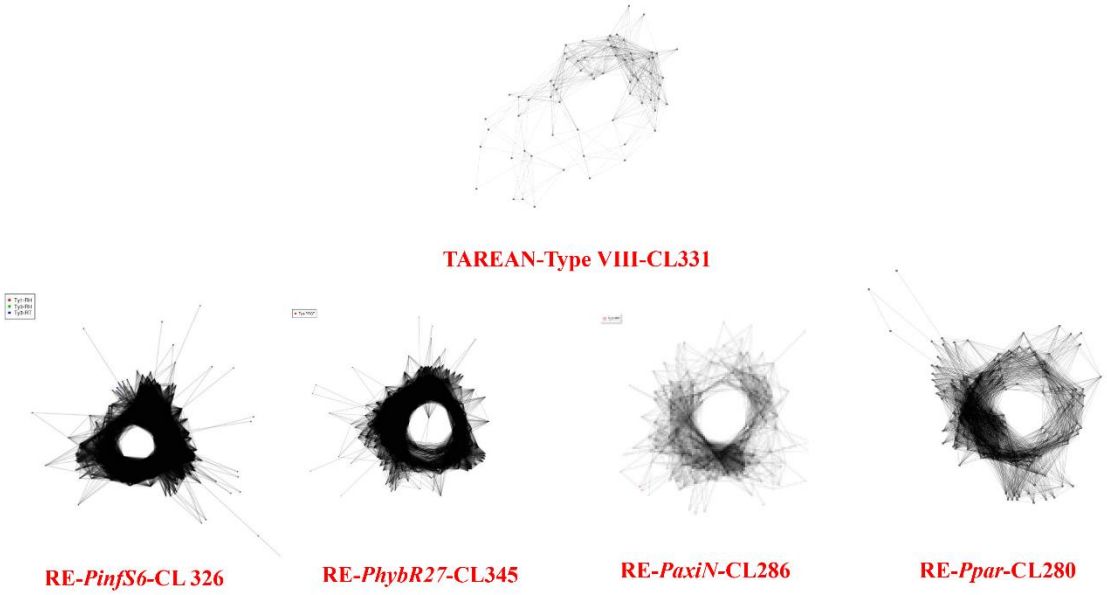
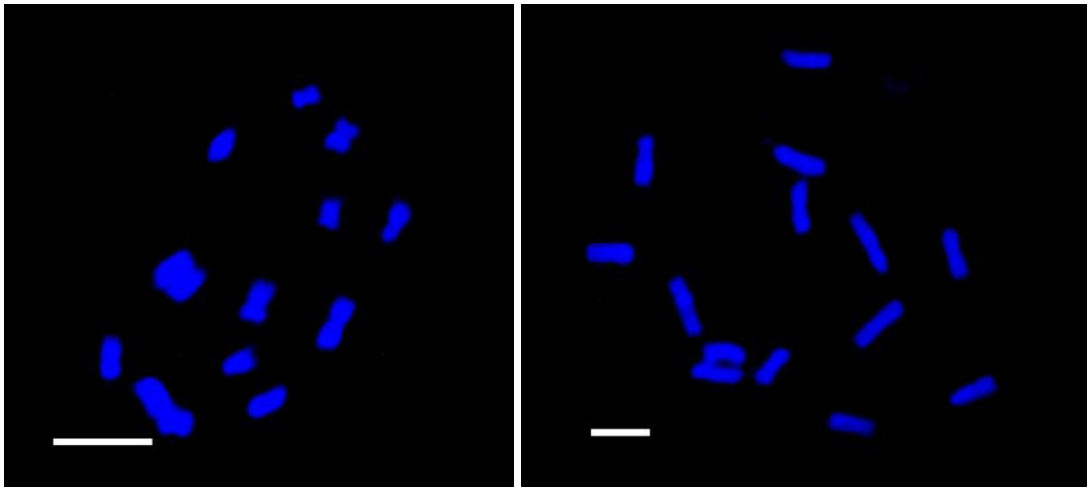
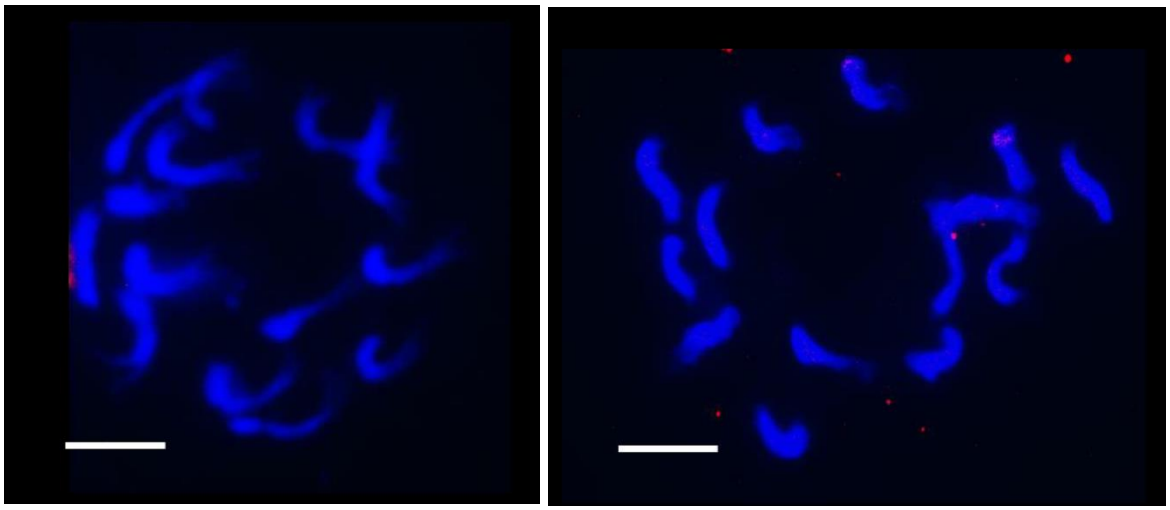


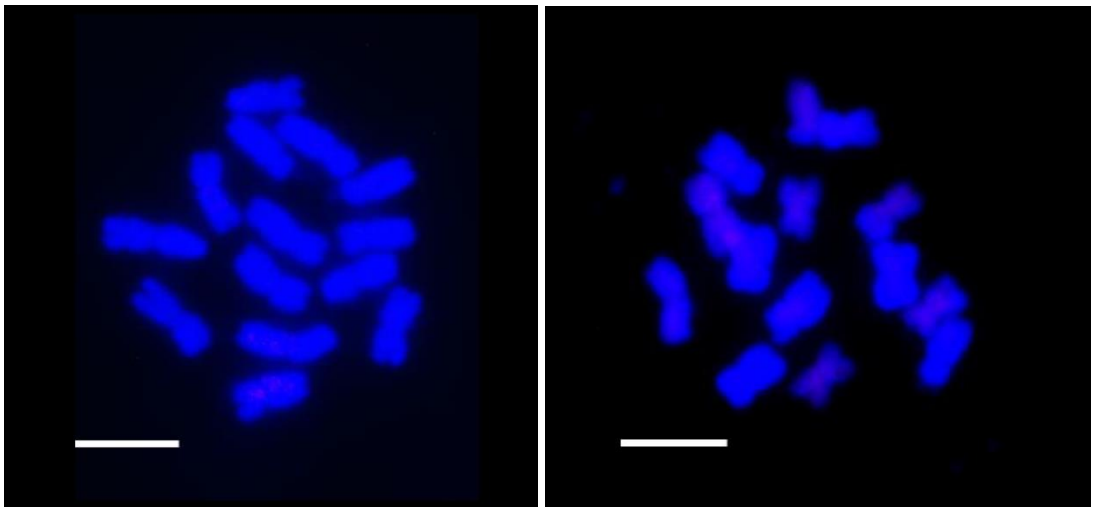
Figure 5.17 Donut shapes of *Type VIII-CL331*, showing TAREAN's graph that is not quite condensed in comparison with more concentrated shapes in the RepeatExplorer (*PinfS6*-CL326, *PhybR27*-CL345, *PaxiN*-CL286, and *Ppar*-CL280).



A B



C D



E F

Figure 5.18 Metaphase chromosomes of *Petunia* (2n=14) species probed with tandem repeat *Type VIII-CL331* shows no clear signals over all *Petunia* species chromosomes. Bar = 10 μm for all species except *P. integrifolia* subsp *inflata* S6 that has 5 μm.

5.5 Discussion

The detailed analysis of repetitive elements from genomes of four *Petunia* species used complementary tools (graph-based repeat clustering using unassembled raw reads, raw read mapping, sequence assemblies, *in situ* hybridization and chromosome studies).

5.5.1 Genome proportion of tandem repeat clusters

Repetitive elements represented 64-68 % of the *Petunia* genomes (Figure 1.3; Appendix 3.1), consistent with the measurements reported by Bombarely *et al.* (2016) in *P. axillaris* subsp *axillaris* N. The analysis showed that a relatively low proportion, 0.6%, of the genome (8.4 Mb of the 1,400 Mb genomes) was represented by a limited number – just eight– tandemly-repeated clusters of satellite DNA clusters (Table 5.4). This contrasts with a higher percentage of tandem repeats in other species with a range of genome sizes: 4.89 % in *Coix aquatica* (Cai *et al.* 2014; 2,335 Mb), 6.8% in *Beta vulgaris* (Zakrzewski *et al.* 2010; 714-758 Mb), 5% in *Zea mays* (Meyers *et al.* 2001; 2,300-2,700 Mb), 2.7% in *Vicia faba* (Robledillo *et al.* 2018; 1,300 Mb), 20% in *Citrus spp*. (380 Mb), 30% in *Cucumis sativus* genomes (367 Mb) (Hemleben *et al.* 2007). The *Petunia* figure is similar to that reported in *Coix lacryma-jobi* (0.60%, 1,600 Mb). Even *Arabidopsis thaliana* (176 Mb) includes some 8% of the genome as rDNA repeats, with 4% as the single *AtCon* tandemly repeated centromeric sequence (Heslop-Harrison *et al.* 1999). Except for the 45S rDNA sequences at the secondary constrictions (slightly darker in DAPI staining), the low proportion of tandemly repeated satellites in petunia is supported by the chromosome images where no positively- or negatively- staining bands are seen (Figures 5.3, 5.5, 5.8, 5.10, 5.12, 5.14, 5.16 and 5.18). In contrast, many other plant species have conspicuous bands, considered as heterochromatin, at centromeric, sub-telomeric or intercalary positions on chromosomes. It is important to study in detail the nature of the tandem repeats in a genome where they represent a small but significant proportion: is their number, distribution and even function the same as in genomes with a higher proportion? Do any features of the genome mitigate against amplification of tandem arrays?

5.5.2 Tandemly repeated sequences within assemblies

In assembled sequence scaffolds, various short satellite repeat arrays were found ranging in length from less than 1kb to some 20 kb (Table 5.4; Figures 5.4 and 5.6). Collapse of repeats, and mis-assembly of flanking regions, is often a consequence of the sequence assembly data and algorithms. Nevertheless, the assembly data in petunia, using both PacBio long-reads and mate-pairs (Bombarely *et al.* 2016) were supported the low proportion and diversity of tandem repeats. Depending on conformation, small repeat arrays (<10-50 kb) were not expected to be seen, by *in situ* hybridization, nor identify SSRs and microsatellites, typically represented by 10s of copies of 2-, 3- or 4- base pair motifs, in the repeat analysis here. However, the chromosomes do show some array loci that are absent from the *P. axillaris* N assembly: for example, tandem repeats *Type I-CL43A* (10 kb array length, with much higher copy number by *in situ* hybridization and read mapping than in the assembly) and *Type II-CL43B* (1 kb) both locate in scaffold *PaxiN162Scf160,* but *in situ* hybridization locates Type I at subtelomeric sites and Type II is pericentromeric (Figures 5.3, 5.5 and 5.6). Short arrays of telomeric sequences, their derivatives, or other subtelomeric repeats, are often also found near centromeres (Presting *et al.* 1996), as seen in the sub-centromeric scaffold *PaxiN162Scf160,* while the major subtelomeric Type I arrays, like the Type V telomere (Figure 5.12), are collapsed in the draft assemblies.

5.5.3 Graph based clusters pipelines

Graph-based clustering algorithms are widely applied to whole-genome sequence assembly (Zerbino & Birney 2008; Kingsford *et al.* 2010; Novák *et al.* 2010). Here, RepeatExplorer and TAREAN graphs are used to cluster repetitive elements, and the tandem repeats generally show a graph with a dense centre (star) or ring (donut) graph shape representing multiple similar reads with overlaps, with a lower number of peripheral reads with variants or representing ends of arrays (Figures 5.2, 5.7, 5.9, 5.11, 5.13, 5.15 and 5.17). While the graphs were helpful for identification of candidate tandem repeats, no link was evident between biological features of the repeat (abundance, location, monomer length and array size, or inter-monomer variation) and the detailed pattern of the graph.

5.5.4 The role of tandemly repeated sequences

Despite the relatively low proportion of repeats in petunia, tandem repeat classes were found with centromeric, peri-centromeric, sub-telomeric, telomeric and intercalary locations (Table 5.4; Figures 5.3, 5.5, 5.8, 5.10, 5.12, 5.14, 5.16 and 5.18). Their potential role in maintenance of nuclear of chromosome structure at metaphase and interphase has been considered, including centromeric function (Hemleben *et al.* 2007). Roles in large-scale genome organization (Biscotti *et al.* 2015), chromatin packaging, or maintenance of chromosome stability (Vershinin & Heslop-Harrison 1998), and effects on the environment of the expressed genes (Lamb *et al.* 2007), have also been proposed (Ohtsubo *et al.* 1991; Han *et al.* 2008; Torres *et al.* 2011) (Figure 5.1).

5.5.5 The evolution of tandem repeats

P. axillaris subsp *axillaris* N and *P. integrifolia* subsp *inflata* S6 separated from a common ancestor some 30 MYA (Wang *et al.* 2008), and came together in *P. hybrida* in the last 2 centuries (Gerats & Strommer 2008). Comparing *PaxiN* and *PinfS6*, of the eight tandem repeats, three have substantially more copies in *PaxiN* than *PinfS6*, and two have more in *PinfS6* than *PaxiN*. Notably, in the hybrid combination, one tandem repeat has a substantially greater copy number (similar copies in both ancestors, telomeric location) while two are diminished (one more *PaxiN* than *PinfS6*, one more *PinfS6* than *PaxiN*). Tandem repeats most likely evolve in copy number by unequal crossing-over at meiosis, or slippage during replication before mitosis, events that our data show have occurred in both recent and older evolutionary time with both amplification and loss. The abundance of all types I to VIII in the sister species *P. axillaris* subsp *parodii* S7 suggests that all types are ancient. Given the defined nature of the tandem repeats and availability of straightforward analysis pathway, it will be valuable to resynthesize *P. hybrida* from it ancestors and analyse the copy number changes occurring over a few generations, comparing both sexual and vegetative pathways to distinguish meiotic and mitotic events, and examining if the changes are salutatory, occurring in bursts soon after hybrid formation as with transposon activation, (Grandbastien 1998), and potentially related through chromatin remodelling and methylation changes. Genomic changes are regularly seen in hybrid species (Alix *et al.* 2017). Gaeta *et al.* (2007) have resynthesized tetraploid Brassicas and seen homoeologous chromosome exchanges, and it would be valuable to examine repetitive DNA changes over similar timescales in petunia. Ma and Gustafson (2008) examined changes in new triticale (wheat x rye) hybrids, and found more genomic changes in the rye chromosomes, occurring immediately after hybridization, also noting an

effect of cytoplasm; major rearrangements of the genome are also seen hybrids in the grass genus *Brachypodium* (Lusinska *et al.* 2018). In other Triticeae species, substantial differences in tandem repeat abundance between sister species combining in a polyploid hybrid are evident: the tandem repeat *dpTa1* is much more abundant in the N than D genomes in ancestral species and hybrid *Aegilops ventricosa* (Bardsley *et al.* 1999). Here, though, each type of repeat has a unique pattern of amplification in the three sister species *P. axillaris* subsp *axillaris* N , *P. integrifolia* subsp *inflata* S6 and *P. axillaris* subsp *parodii* S7 suggesting different evolution from that of the rest of the genome. Very rapid changes in copy number – from approximately 5x to 1/5x in the hybrid compared to ancestral species – were observed for the each type. No correlations between abundance nor copy number evolution and chromosomal location were evident (see Figure 5.10).

5.5.6 Tandem repeats as genetic markers

Most of tandem clusters involved in pericentromeric regions except *Type I-CL43A* and *Type V-CL102* that discovered in subtelomeric and telomeric loci respectively (Figures 5.3 and 5.12). This distribution is functionally important to strengthen chromosome structure of petunia as a typical genome. Findings of tandem repeat types III, VI and VII have been characterised as chromosomes markers, providing useful cytogenetic tool for easily identifying metaphase chromosomes, depending on repeat localization and fluorescence intensity (Figures 5.8, 5.14 and 5.16). Recently, many tools have been used for finding a specific approach to characterise genomes of different species, these methods are classified as non-PCR and PCR based techniques (Agarwal *et al.* 2008; Korir *et al.* 2013). So far, facts of the homology between *in situ* hybridization strength and the abundance of copy numbers in raw reads have been achieved for some tandem repeats but unlikely approved for others like Type IV and Type VIII (Figures 5.10 and 5.18). However, the two types have to some extent enough copy numbers and genome proportion (especially Type IV), with no clear signal has seen through FISH that is not a quantitative way for detecting unclear probes. This probably because of the existence of these repeats is much dispersed and in very small arrays. Interestingly, no correlation has been noticed between the copy numbers of some repeats and chromosome sizes.

Results obtained in the present chapter will consequently be used in integration of the chromosomal and genetic maps that are undertaken by the Petunia Genome Sequencing Consortium through molecular mapping and *in situ* hybridization, efficient chromosomes markers, providing useful cytogenetic tool for easily identifying metaphase chromosomes.

Chapter VI. General discussion: Repetitive elements in petunia

This work has characterised the repetitive DNA landscape in the *Petunia* genome (see Appendix 3.1), with DNA components that can be divided into two major divisions based on their organization and position within genome. One division contains semi-autonomous sequences such as DNA transposable elements, retroelements, and integrated retrovirus-related sequences. The retroviral-related sequences were the focus of more detailed studies of their consequences for the genome and plant. The second division includes tandemly repeated sequences like satellite DNAs.

PVCV and other integrated viral sequences in *Petunia* species

From the first group, EPRVs, as elements with major consequences for the host plant, were investigated here using bioinformatics tools, cytogenetic work (chapter III), and transmission electron microscopy (chapter IV) within *Petunia* genomes. Although these elements have extensively colonized flowering plant genomes since 20-34 MYA, EPRVs are still less well studied in plant genomes than in animals where they have been known for many years (Geering *et al.* 2014). Our results have highlighted further facts to understand the role of EPRVs elements (petu-like, florendo-like, and caulimo-like sequences) in formation of wild and hybrid host genomes, their possible functions, and relationship between different units. High variable proportions of EPRVs sequences have been counted within wild and hybrid genomes of petunia, and widely distinguished over their loci on chromosomes suggesting these elements (petu-like virus) as a novel cytogenetic marker alongside retrotransposons (Figure 3.18). Interestingly, some EPRVs (petu, and florendo-like sequences) share similar positions over host chromosomes (Figure 3.42) suggesting probably their closely related phylogenetic relationship with same endogenisation events. Endogenisation searches were extended to discover novel members of florendoviruses and caulimoviruses (four sequences of each type of both groups).

Also, the outcomes show higher levels of genome proportions and copy nymbers of EPRVs within hybrid petunia than wild parents, in addition to integration pattern of these elements inside hybrids as tandemly repeated sequences unlike their parents. As a consequence, hybrid petunias are considered as a reservoir of EPRVs and have more permissive and inducible genomes than wild parents (Richert-Pöggeler *et al.* 2003) (chapter III).

PVCV expression and infections within *Petunia* tissues

Expression protocols have revealed that florendoviruses were non-active integrants despite their abundance unlike petu- and caulimo-like sequences that were strongly expressed (Figure 3.38). For the first time in plants, the approach of *de novo* integration of chromosomally integrated PVCV has discovered to be integrated over next generations on the telomere of heterologous chromosomes (see Figure 3.39). Moreover, significant ultra-structure changes have been noticed within chlorotic spot, vein clearing symptoms of horizontally transmitted PVCV as well as progeny of vertically transmitted PVCV and induced symptoms of PVCV (chapter IV).

The wide existence of endogenous pararetroviruses in abundant copy numbers within flowering plant genomes has suggested these elements as a source of infection, but also have a protective function in silencing and suppression via RNA interference-induced resistance (Richert-Pöggeler *et al.* 2003). The number of copies in petunia (chapter III) is much higher than is known to induce RNAi-mediated resistance to RNA viruses (Ratcliff *et al*. 1997; Jakowitsch *et al*. 1999; Wesley *et al*. 2001; Kunii *et al*. 2004). With the close relationship to retroelements, EPRV insertion has many impacts on host genome, potentially including chromosome breakage and rearrangement, insertional mutation, sequence amplification and gene regulation. However, most retroelements such as the *Gypsy* and *Copia* families, or DNA elements including CACTA-EnSpm, have more dispersed distributions over broad chromosomal regions, compared to the EPRVs that occur a discrete chromosomal loci. Fragmented lengths of integrated EPRV units also could have activities of promoter/enhancer, ORFs, active splice sites and RT in addition to transposition and related amplification ability (Löwer 1999; Hansen & Heslop-Harrison 2004a). Additionally, they are playing a very clear role in their host evolution by adding novel genetic components at the levels of coding or transcriptional regulation taking into account that plants have higher plasticity in their genome structure than animals: the existence of EPRVs may play a role in genome diversification by providing coding or transcription regulatory elements as new genetic components. The spread of EPRVs is widley correlated with their host reproduction and distribution between different regions around the world. Although some EPRVs are located in geographically segregated regions, they probably spreaded to different areas of the world around 500 YA. Based on genetic population structure, these elements showed frequent spread among regions. Our findings confirmed the widespread distribution of EPRVs in the genomes of cultivated and wild plants

with extensive colonization patterns. These elements followed similar scale of high copy numbers families of transposable elements (Geering *et al.* 2014; Yasaka *et al.* 2014).

Tandemly repeats as major repetitive DNA sequences in petunia

Tandemly repeated sequences have been analysed in this work through bioinformatics analysis and cytogenetic tools to explore their diversity over host genomes (chapter V). These elements were found with relatively low abundance in *Petunia* genomes, and include eight types of clusters that involved in centromeric, peri-centromeric, sub-telomeric, telomeric and intercalary regions of *Petunia* chromosomes (Table 5.4). Tandemly repeated clusters are vary in their copy numbers and localizations over examined *Petunia* genomes. More importantly, three types of tandem repeats have been characterised as efficient chromosomes markers, providing useful cytogenetic tools for the identification of metaphase chromosomes based on repeat localization and fluorescence intensity (Figures 5.8, 5.14 and 5.16). This work is consistent with a functional role suggested for the repeats in genome evolution and structural roles within chromosomes regions like centromeres and telomeres (Schmidt & Heslop-Harrison 1998). Their maintenance role in chromosome structure at metaphase and interphase has been considered, including centromeric function (Hemleben *et al.* 2007). Roles in large-scale genome organization (Biscotti *et al*. 2015), chromatin packaging, or maintenance of chromosome stability (Vershinin & Heslop-Harrison 1998), and effects on the environment of the expressed genes (Lamb *et al.* 2007), have also been proposed (Ohtsubo *et al.* 1991; Han *et al.* 2008; Torres *et al.* 2011), as well as in the stability of chromatin packaging (Escudeiro, Adega, Robinson, Schwarzacher, Heslop-Harrison and Chaves, submitted). Moreover, as repetitive DNA, tandem repeats probably have a protective role for coding DNA from any shock through stress conditions due to its high stability and it has been used for nuclear architecture study, chromosomes and genomes identification, and phylogenetic analysis (Heslop-Harrison 2000; Pluhar *et al.* 2001).

Future extensions

Work should be continued to find out more integration features, localization, and transmission of different EPRVs within other plant species: given the well-defined presence and effects in petunia described here, and in banana (Harper *et al.* 1999), it is likely that they will be found more widely across plant species. Investigating new members of EPRVs should be extended

with assistance of the whole genome sequencing techniques. FISH work on caulimovirus-like sequences position in *Petunia* chromosomes should be extended using new techniques such as a massive single copy pools of oligonucleotides (Han *et al.* 2015) to cover all EPRVs types' locations. Expression patterns of EPRVs have been explored within and between wild *Petunia* species, and more data from hybrid varieties is strongly needed. In addition, *de novo* integration experiments have revealed novel facts about plant virus integration over plant generations, and exploring more details about integration sites within telomere regions are highly recommended. It would also be valuable to characterise further the mechanisms of vein clearing following PVCV infection, including the effects on chloroplasts. It will also be important to examine miRNA/siRNA species present in various infected and resistant or asymptomatic lines to futher characterise the endogenous effects of the integrated EPRV sequence motifs. For tandemly repeated sequences, it will be valuable to resynthesize *P. hybrida* from it ancestors and analyse the copy number changes occurring over a few generations, and examining chromosome elimination and diploidization, comparing both sexual and vegetative pathways to distinguish meiotic and mitotic events, and examining if the changes are salutatory, occurring in bursts soon after hybrid formation. For all the repetitive sequence classes, it will be interesting to further examine their role in epigenetics, including chromatin organization, their relationships with histone and DNA modifications, and potential consequences for the modulation of gene expression and chromosomes behaviour in adjacent regions.

Conclusion

In the context of diversity and evolution, using six petunia species, raw reads from four whole-genome survey sequencing experiments, and two assemblies with bioinformatics tools, cytogenetic techniques and electron microscopy allowed us to explore our collected results regarding EPRVs and tandem repeats. This project has achieved our main aims (see 1.7) in order to provide new insights and highlight further facts to understand the role of repetitive DNA elements in formation of wild and hybrid host genomes, their possible functions and relationships. These elements present fundamental diversity among the evolutionarily related species and individuals, suggesting an evolutionary impact associated with repeats on host genome variation.

References

Adams M.J. & Antoniw J.F. (2005) DPVweb: An open access internet resource on plant viruses and virus diseases. *Outlooks on Pest Management* **16**, 268-70.

Adams S.E., Mellor J., Gull K., Sim R.B., Tuite M.F., Kingsman S.M. & Kingsman A.J. (1987) The functions and relationships of Ty-VLP proteins in yeast reflect those of mammalian retroviral proteins. *Cell* **49**, 111-9.

Agarwal M., Shrivastava N. & Padh H. (2008) Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Reports* **27**, 617-31.

Alix K., Gérard P.R., Schwarzacher T. & Heslop-Harrison J. (2017) Polyploidy and interspecific hybridization: partners for adaptation, speciation and evolution in plants. *Annals of Botany* **120**, 183-94.

Alizon S., Hurford A., Mideo N. & Van Baalen M. (2009) Virulence evolution and the trade-off hypothesis: history, current state of affairs and the future. *Journal of Evolutionary Biology* **22**, 245-59.

Alkan C., Cardone M.F., Catacchio C.R., Antonacci F., O'Brien S.J., Ryder O.A., Purgato S., Zoli M., Della Valle G. & Eichler E.E. (2011) Genome-wide characterization of centromeric satellites from multiple mammalian genomes. *Genome Research* **21**, 137-45.

Altschul S.F., Gish W., Miller W., Myers E.W. & Lipman D.J. (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403-10.

Amari K., Boutant E., Hofmann C., Schmitt-Keichinger C., Fernandez-Calvino L., Didier P., Lerich A., Mutterer J., Thomas C.L. & Heinlein M. (2010) A family of plasmodesmal proteins with receptor-like properties for plant viral movement proteins. *PLoS Pathogens* **6**, e1001119.

Anabestani A., Behjatnia S.A.A., Izadpanah K., Tabein S. & Accotto G.P. (2017) Seed Transmission of *Beet Curly Top Virus* and *Beet Curly Top Iran Virus* in a Local Cultivar of *Petunia* in Iran. *Viruses* **9**, 299-310.

Ando T. (1996) Distribution of *Petunia axillaris* (Solanaceae) and its new subspecies in Argentina and Bolivia. *Acta Phytotax. Geobot.* **47**, 19-30.

Ando T., Kokubun H., Watanabe H., Tanaka N., Yukawa T., Hashimoto G., Marchesi E., Suárez E. & Basualdo I.L. (2005) Phylogenetic analysis of *Petunia* sensu Jussieu (Solanaceae) using chloroplast DNA RFLP. *Annals of Botany* **96**, 289-97.

Ando T., Nomura M., Tsukahara J., Watanabe H., Kokubun H., Tsukamoto T., Hashimoto G., Marchesi E. & Kitching I.J. (2001) Reproductive isolation in a native population of *Petunia* sensu Jussieu (Solanaceae). *Annals of Botany* **88**, 403-13.

Arkhipova I.R., Yushenova I.A. & Rodriguez F. (2017) Giant reverse transcriptase-encoding transposable elements at telomeres. *Molecular Biology and Evolution* **34**, 2245-57.

Arney K.L. & Fisher A.G. (2004) Epigenetic aspects of differentiation. *Journal of Cell Science* **117**, 4355-63.

Balke I., Reseviča G. & Zeltins A. (2018) Isolation and Characterization of Two Distinct Types of Unmodified Spherical Plant Sobemovirus-Like Particles for Diagnostic and Technical Uses. In: *Virus-Derived Nanoparticles for Advanced Technologies* (pp. 19-34). Springer.

Bang J.W., Kim J.S. & Lee M.K. (1997) Molecular cytogenetics of plant genome. *Journal of Plant Biology* **40**, 149-55.

Bao W., Jurka M.G., Kapitonov V.V. & Jurka J. (2009) New superfamilies of eukaryotic DNA transposons and their internal divisions. *Molecular Biology and Evolution* **26**, 983-93.

Bao W., Kojima K.K. & Kohany O. (2015) Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA* **6**, 11-16.

Bardsley D., Cuadrado A., Jack P., Harrison G., Castilho A. & Heslop-Harrison J. (1999) Chromosome markers in the tetraploid wheat *Aegilops ventricosa* analysed by *in situ* hybridization. *Theoretical and Applied Genetics* **99**, 300-4.

Bennett M.D. & Leitch I.J. (2011) Nuclear DNA amounts in angiosperms: targets, trends and tomorrow. *Annals of Botany* **107**, 467-590.

Bennetzen J.L. (1996) The contributions of retroelements to plant genome organization, function and evolution. *Trends in Microbiology* **4**, 347-53.

Benson G. (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research* **27**, 573-80.

Bentzer B., Bothmer R.v., Engstrand L., Gustafsson M. & Snogerup S. (1971) Some sources of error in the determination of arm rations of chromosomes. *Botaniska Notiser*.

Bergman C.M. & Quesneville H. (2007) Discovering and detecting transposable elements in genome sequences. *Briefings in Bioinformatics* **8**, 382-92.

Bilinski P., Han Y., Hufford M.B., Lorant A., Zhang P., Estep M.C., Jiang J. & Ross-Ibarra J. (2017) Genomic abundance is not predictive of tandem repeat localization in grass genomes. *PloS one* **12**, e0177896.

Biscotti M.A., Olmo E. & Heslop-Harrison J.P. (2015) Repetitive DNA in eukaryotic genomes. *Chromosome Res*. 23:415–20.

Bloom K. (2007) Centromere dynamics. *Current Opinion in Genetics & Development* **17**, 151-6.

Boeke J. & Corces V.G. (1989) Transcription and reverse transcription of retrotransposons. *Annual Reviews in Microbiology* **43**, 403-34.

Bombarely A., Moser M., Amrad A., Bao M., Bapaume L., Barry C.S., Bliek M., Boersma M.R., Borghi L. & Bruggmann R. (2016) Insight into the evolution of the Solanaceae from the parental genomes of *Petunia hybrida*. *Nature Plants* **2**, 16074-82.

Brandes J. & Wetter C. (1959) Classification of elongated plant viruses on the basis of particle morphology. *Virology* **8**, 99-115.

Bringaud F., Ghedin E., Blandin G., Bartholomeu D.C., Caler E., Levin M.J., Baltz T. & El-Sayed N. M. (2006) Evolution of non-LTR retrotransposons in the trypanosomatid genomes: *Leishmania major* has lost the active elements. *Molecular and Biochemical Parasitology* **145**, 158-70.

Büchen-Osmond C. (2003) Taxonomy and classification of viruses. *Manual of clinical microbiology* **2**, 1217-26.

Britton N.L. & Brown A. (1913) *An illustrated flora of the northern United States, Canada and the British possessions: from Newfoundland to the parallel of the southern boundary of Virginia, and from the Atlantic Ocean westward to the 102d meridian*. C. Scribner's sons.

Cabrera, A.L. andWillink, A. (1980) *Biogeografia de America Latina, 2nd Edn*. Secretaria Generalde la OEA, Washington.

Cai Z., Liu H., He Q., Pu M., Chen J., Lai J., Li X. & Jin W. (2014) Differential genome evolution and speciation of *Coix lacryma-jobi* L. and *Coix aquatica* Roxb. hybrid guangxi revealed by repetitive sequence analysis and fine karyotyping. *BMC Genomics* **15**, 1025-40.

Casacuberta E. (2017) *Drosophila*: retrotransposons making up telomeres. *Viruses* **9**, 192-207.

Cavalier-Smith T. (1985) Selfish DNA and the origin of introns. *Nature* **315**, 283-84.

Charlesworth B., Sniegowski P. & Stephan W. (1994) The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* **371**, 215-20.

Chen J., Chen J.P. & Adams M. (2002) Characterisation of some carla-and potyviruses from bulb crops in China. *Archives of Virology* **147**, 419-28.

Clare J. & Farabaugh P. (1985) Nucleotide sequence of a yeast Ty element: evidence for an unusual mechanism of gene expression. *Proceedings of the National Academy of Sciences* **82**, 2829-33.

Cohen J., Sikron N., Shuval S. & Gera A. (1999) Susceptibility of vegetatively propagated petunia to tobamovirus infection and its possible control. *Hortscience* **34**, 292-3.

Conia J., Bergounioux C., Perennes C., Muller P., Brown S. & Gadal P. (1987) Flow cytometric analysis and sorting of plant chromosomes from *Petunia hybrida* protoplasts. *Cytometry: The Journal of the International Society for Analytical Cytology* **8**, 500-8.

Contento A., Heslop-Harrison J. & Schwarzacher T. (2005) Diversity of a major repetitive DNA sequence in diploid and polyploid Triticeae. *Cytogenetic and Genome Research* **109**, 34-42.

Cooper D. (1946) Double fertilization in *Petunia*. *American Journal of Botany* **33**, 54-7.

Corbetta M., Kasdorf G., Engelbrecht D. & Wiid J. (2017) Detection of viral-like particles by electron microscopy of negatively stained extracts from grapevines. *South African Journal of Enology and Viticulture* **5**, 43-9.

Craig N.L. (2002) Mobile DNA: an introduction. In: *Mobile DNA II* (pp. 3-11). *American Society of Microbiology*.

Croll D., Lendenmann M.H., Stewart E. & McDonald B.A. (2015) The impact of recombination hotspots on genome evolution of a fungal plant pathogen. *Genetics* **201**, 1213-28.

Daboussi M.-J. & Capy P. (2003) Transposable elements in filamentous fungi. *Annual Reviews in Microbiology* **57**, 275-99.

De Jong J.H., Fransz P. & Zabel P. (1999) High resolution FISH in plants–techniques and applications. *Trends in Plant Science* **4**, 258-63.

Dermen H. (1931) Polyploidy in petunia. *American Journal of Botany* **18**, 250-61.

Dietrich A., De Jong J. & Mulder R. (1981) Location and variation of the constitutive heterochromatin in *Petunia hybrida*. *Genetica* **55**, 85-91.

Dillon L.W., Pierce L.C., Ng M.C. & Wang Y.-H. (2013) Role of DNA secondary structures in fragile site breakage along human chromosome 10. *Human Molecular Genetics* **22**, 1443-56.

Diop S.I., Geering A.D., Alfama-Depauw F., Loaec M., Teycheney P.-Y. & Maumus F. (2018) *Tracheophyte* genomes keep track of the deep evolution of the *Caulimoviridae*. *Scientific Reports* **8**, 572-80.

Doolittle W.F. & Sapienza C. (1980) Selfish genes, the phenotype paradigm and genome evolution. *Nature* **284**, 601-3.

Doyle J. & Doyle J. (1990) Isolation of DNA from small amounts of plant tissues. *BRL Focus* **12**, V15.

Edwards K., Johnstone C. & Thompson C. (1991) A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Research* **19**, 1349.

Eichler E.E., Clark R.A. & She X. (2004) An assessment of the sequence gaps: unfinished business in a finished human genome. *Nature Reviews Genetics* **5**, 345-54.

Elder Jr J.F. & Turner B.J. (1995) Concerted evolution of repetitive DNA sequences in eukaryotes. *The Quarterly Review of Biology* **70**, 297-320.

Emori Y., Shiba T., Kanaya S., Inouye S., Yuki S. & Saigo K. (1985) The nucleotide sequences of copia and copia-related RNA in *Drosophila* virus-like particles. *Nature* **315**, 773-76.

Engelmann J. & Hamacher J. (2008) Plant Virus Diseases: Ornamental Plants 207-229.

Ewald P.W. (1983) Host-parasite relations, vectors, and the evolution of disease severity. *Annual Review of Ecology and Systematics* **14**, 465-85.

Ewald P.W. (1987) Transmission Modes and Evolution of the Parasitism-Mutualism Continuum a. *Annals of the New York Academy of Sciences* **503**, 295-306.

Fauquet C.M., Mayo M.A., Maniloff J., Desselberger U. & Ball L.A. (2005) *Virus taxonomy: VIIIth report of the International Committee on Taxonomy of Viruses*. Academic Press.

Ferguson M.C. & Ottley A.M. (1932) Studies on *Petunia* III. A redescription and additional discussion of certain species of *Petunia*. *American Journal of Botany* **19**, 385-405.

Feschotte C. & Pritham E.J. (2007) DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* **41**, 331-68.

Flavell A.J., Dunbar E., Anderson R., Pearce S.R., Hartley R. & Kumar A. (1992) Ty1–copia group retrotransposons are ubiquitous and heterogeneous in higher plants. *Nucleic Acids Research* **20**, 3639-44.

Flavell A.J., Pearce S.R. & Kumar A. (1994) Plant transposable elements and the genome. *Current Opinion in Genetics & Development* **4**, 838-44.

Flavell R., Bennett M., Smith J. & Smith D. (1974) Genome size and the proportion of repeated nucleotide sequence DNA in plants. *Biochemical Genetics* **12**, 257-69.

Fransz P.F., Stam M., Montijn B., Hoopen R.T., Wiegant J., Kooter J.M., Oud O. & Nanninga N. (1996) Detection of single-copy genes and chromosome rearrangements in *Petunia hybrida* by fluorescence *in situ* hybridization. *The Plant Journal* **9**, 767-74.

Fregonezi J.N., Turchetto C., Bonatto S.L. & Freitas L.B. (2012) Biogeographical history and diversification of Petunia and Calibrachoa (Solanaceae) in the Neotropical Pampas grassland. *Botanical Journal of the Linnean Society* **171**, 140-53.

Fries R.E. (1911) Die Arten der Gattung *Petunia*. *Kungl. Svenska Vetenskapsakademiens Handlingar* **46**, 1-72.

Froissart R., Roze D., Uzest M., Galibert L., Blanc S. & Michalakis Y. (2005) Recombination every day: abundant recombination in a virus during a single multi-cellular host infection. *PLoS Biology* **3**, 389-95.

Gaeta R.T., Pires J.C., Iniguez-Luy F., Leon E. & Osborn T.C. (2007) Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. *The Plant Cell* **19**, 3403-17.

Gall J.G. & Pardue M.L. (1969) Formation and detection of RNA-DNA hybrid molecules in cytological preparations. *Proceedings of the National Academy of Sciences* **63**, 378-83.

Ganga M., Jayalakshmi S., Jegadeeswari V., Padmadevi K. & Jawaharlal M. (2011) *Petunia*. In: *Wild Crop Relatives: Genomic and Breeding Resources*. 209-42. Springer.

García G., Ríos N. & Gutiérrez V. (2015) Next-generation sequencing detects repetitive elements expansion in giant genomes of annual killifish genus *Austrolebias* (Cyprinodontiformes, Rivulidae). *Genetica* **143**, 353-60.

Gautam K.K., Kumar S. & Raj S.K. (2012) Molecular characterization of *Cucumber mosaic virus* strain causing severe mosaic disease on petunia in India. *Phytoparasitica* **40**, 425-31.

Gayral P., Noa-Carrazana J.-C., Lescot M., Lheureux F., Lockhart B.E., Matsumoto T., Piffanelli P. & Iskra-Caruana M.-L. (2008) A single *Banana streak virus* integration event in the banana genome as the origin of infectious endogenous pararetrovirus. *Journal of Virology* **82**, 6697-710.

Geering A. & Hull R. (2012) Family *Caulimoviridae*. *Virus taxonomy classification and nomenclature of viruses ninth Report of the International Committee on Taxonomy of Viruses*, 424-43.

Geering A.D. (2001) *Caulimoviridae* (Plant Pararetroviruses). *eLS*.

Geering A.D., Maumus F., Copetti D., Choisne N., Zwickl D.J., Zytnicki M., McTaggart A.R., Scalabrin S., Vezzulli S. & Wing R.A. (2014) Endogenous florendoviruses are major components of plant genomes and hallmarks of virus evolution. *Nature Communications* **5**, 5269-79.

Geering A.D., Scharaschkin T. & Teycheney P.-Y. (2010) The classification and nomenclature of endogenous viruses of the family *Caulimoviridae*. *Archives of Virology* **155**, 123-31.

Gera A., Sikron N., Cohen J. & Zeidan M. (2000) First report of *Petunia vein clearing virus* in Israel. *Plant Disease* **84**, 201.

Gerats T. (2009) Identification and exploitation of *Petunia* transposable elements: a brief history. In: *Petunia* (pp. 365-79). Springer.

Gerats T. & Strommer J. (2008) *Petunia: evolutionary, developmental and physiological genetics*. Springer Science & Business Media.

Gilbert C. & Feschotte C. (2018) Horizontal acquisition of transposable elements and viral sequences: patterns and consequences. *Current Opinion in Genetics & Development* **49**, 15-24.

Goodwin S., McPherson J.D. & McCombie W.R. (2016) Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics* **17**, 333-51.

Grandbastien M.-A. (1998) Activation of plant retrotransposons under stress conditions. *Trends in Plant Science* **3**, 181-7.

Grandbastien M.-A., Spielmann A. & Caboche M. (1989) Tnt1, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. *Nature* **337**, 376-80.

Gregor W., Mette M.F., Staginnus C., Matzke M.A. & Matzke A.J. (2004) A distinct endogenous pararetrovirus family in *Nicotiana tomentosiformis*, a diploid progenitor of polyploid tobacco. *Plant Physiology* **134**, 1191-9.

Gregory T.R., Nicol J.A., Tamm H., Kullman B., Kullman K., Leitch I.J., Murray B.G., Kapraun D.F., Greilhuber J. & Bennett M.D. (2006) Eukaryotic genome size databases. *Nucleic Acids Research* **35**, 332-8.

Greuter W., McNeill J., Barrie F.R., Burdet H., Demoulin V., Filgueiras T., Nicholson D., Silva P., Skog J. & Trehane P. (2000) International code of botanical nomenclature

(Saint Louis Code): Sixteenth International Botanical Congress, St Louis, Missouri, USA, July-August 1999. In: *International code of botanical nomenclature (Saint Louis Code): Sixteenth International Botanical Congress, St Louis, Missouri, USA, July-August 1999.* International Association for Plant Taxonomy.

Gronenborn B. (1987) The molecular biology of *Cauliflower mosaic virus* and its application as plant gene vector. In: *Plant DNA Infectious Agents* (pp. 1-29). Springer.

Hagemann R. & Snoad B. (1971) Paramutation (somatic conversion) at the Sulfurea locus of Lycopersicon esculentum V. The localisation of Sulf. *Heredity* **27**, 409.

Hall A.E., Keith K.C., Hall S.E., Copenhaver G.P. & Preuss D. (2004) The rapidly evolving field of plant centromeres. *Current Opinion in Plant Biology* **7**, 108-14.

Han Y., Zhang T., Thammapichai P., Weng Y. & Jiang J. (2015) Chromosome-specific painting in *Cucumis* species using bulked oligonucleotides. *Genetics* **200**, 771-9.

Han Y., Zhang Z., Liu J., Lu J., Huang S. & Jin W. (2008) Distribution of the tandem repeat sequences and karyotyping in cucumber (*Cucumis sativus* L.) by fluorescence *in situ* hybridization. *Cytogenetic and Genome Research* **122**, 80-8.

Hansen C., Harper G. & Heslop-Harrison J. (2005) Characterisation of pararetrovirus-like sequences in the genome of potato (*Solanum tuberosum*). *Cytogenetic and Genome Research* **110**, 559-65.

Hansen C. & Heslop-Harrison J.S. (2004b) Sequences and Phylogenies of Plant Pararetroviruses, Viruses, and Transposable Elements. *Advances in Botanical Research* **41**, 165-93.

Hansen L.J., Chalker D.L. & Sandmeyer S.B. (1988) Ty3, a yeast retrotransposon associated with tRNA genes, has homology to animal retroviruses. *Molecular and Cellular Biology* **8**, 5245-56.

Harper G., Hull R., Lockhart B. & Olszewski N. (2002) Viral sequences integrated into plant genomes. In: *Annual Review of Phytopathology*, pp. 119-36.

Harper G., Osuji J.O., Heslop-Harrison J.P. & Hull R. (1999) Integration of banana streak badnavirus into the *Musa* genome: molecular and cytogenetic evidence. *Virology* **255**, 207-13.

Heitkam T. & Schmidt T. (2009) BNR–a LINE family from Beta vulgaris–contains a RRM domain in open reading frame 1 and defines a L1 sub-clade present in diverse plant genomes. *The Plant Journal* **59**, 872-82.

Hemleben V., Kovarik A., Torres-Ruiz R.A., Volkov R.A. & Beridze T. (2007) Plant highly repeated satellite DNA: molecular evolution, distribution and use for identification of hybrids. *Systematics and Biodiversity* **5**, 277-89.

Henikoff S., Ahmad K. & Malik H.S. (2001) The centromere paradox: stable inheritance with rapidly evolving DNA. *Science* **293**, 1098-102.

Hesketh E.L., Meshcheriakova Y., Thompson R.F., Lomonossoff G.P. & Ranson N.A. (2017) The structures of a naturally empty *Cowpea mosaic virus* particle and its genome-containing counterpart by cryo-electron microscopy. *Scientific Reports* **7**, 539-48.

Heslop-Harrison J. (2000) Comparative genome organization in plants: from sequence and markers to chromatin and chromosomes. *The Plant Cell* **12**, 617-35.

Heslop-Harrison J.P., Brandes A., Taketa S., Schmidt T., Vershinin A.V., Alkhimova E.G., Kamm A., Doudrick R.L., Schwarzacher T. & Katsiotis A. (1997) The chromosomal distributions of Ty1-copia group retrotransposable elements in higher plants and their implications for genome evolution. *Genetica* **100**, 197-204.

Heslop-Harrison J.P. & Schwarzacher T. (2012) Genetics and genomics of crop domestication. In: *Plant Biotechnology and Agriculture* (pp. 3-18). Elsevier.

Heslop-Harrison J.S., Murata M., Ogura Y., Schwarzacher T. & Motoyoshi F. (1999) Polymorphisms and genomic organization of repetitive DNA from centromeric regions of *Arabidopsis* chromosomes. *The Plant Cell* **11**, 31-42.

Heslop-Harrison J. & Schmidt T. (2001) Plant nuclear genome composition. Chichester: John Wiley & Sons Ltd; 2012.

Hoang A.T., Zhang H.-m., Yang J., Chen J.-p., Hébrard E., Zhou G.-h., Vinh V.N. & Cheng J.-a. (2011) Identification, characterization, and distribution of *Southern rice black-streaked dwarf virus* in Vietnam. *Plant Disease* **95**, 1063-9.

Hohn T. (1994) Recombination of a plant pararetrovirus: *Cauliflower mosaic virus*. In: *Homologous Recombination and Gene Silencing in Plants* (pp. 25-38). Springer.

Hohn T., Richert-Pöggeler K.R., Staginnus C., Harper G., Schwarzacher T., Teo C.H., Teycheney P.-Y., Iskra-Caruana M.-L. & Hull R. (2008) Evolution of integrated plant viruses. In: *Plant Virus Evolution* (pp. 53-81). Springer.

Hohn T. & Rothnie H. (2013) Plant pararetroviruses: replication and expression. *Current Opinion in Virology* **3**, 621-8.

Hoskins R.A., Carlson J.W., Kennedy C., Acevedo D., Evans-Holm M., Frise E., Wan K.H., Park S., Mendez-Lago M. & Rossi F. (2007) Sequence finishing and mapping of *Drosophila melanogaster* heterochromatin. *Science* **316**, 1625-8.

Hua-Van A., Le Rouzic A., Boutin T.S., Filée J. & Capy P. (2011) The struggle for life of the genome's selfish architects. *Biology Direct* **6**, 19-47.

Hua-Van A., Le Rouzic A., Maisonhaute C. & Capy P. (2005) Abundance, distribution and dynamics of retrotransposable elements and transposons: similarities and differences. *Cytogenetic and Genome Research* **110**, 426-40.

Huang Y., Hidalgo-Bravo A., Zhang E., Cotton V.E., Mendez-Bermudez A., Wig G., Medina-Calzada Z., Neumann R., Jeffreys A.J. & Winney B. (2013) Human telomeres that carry an integrated copy of human herpesvirus 6 are often short and unstable, facilitating release of the viral genome from the chromosome. *Nucleic Acids Research* **42**, 315-27.

Huelsenbeck J.P. & Ronquist F. (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754-5.

Huertas P. (2010) DNA resection in eukaryotes: deciding how to fix the break. *Nature Structural and Molecular Biology* **17**, 11-16.

Hull R. (2002) Matthew's Plant Virology 4ª Ed.[e-book] Grã-Bretanha. Bath Press. Disponível em: http://pt. scribd. com/doc/51572781/Matthews-Plant-Virology-4th-Ed [Acedido em 2 de Setembro de 2013].

Hull R. (2014) Chapter 8–origins and evolution of plant viruses. *Plant Virology, 5th edn (ed. Hull R.)*, 423-76.

Hutchison III C.A. (2007) DNA sequencing: bench to bedside and beyond. *Nucleic Acids Research* **35**, 6227-37.

Irick H. (1994) A new function for heterochromatin. *Chromosoma* **103**, 1-3.

Iskra-Caruana M.-L., Baurens F.-C., Gayral P. & Chabannes M. (2010) A four-partner plant–virus interaction: enemies can also come from within. *Molecular Plant-Microbe Interactions* **23**, 1394-402.

Jakowitsch J., Mette M., van Der Winden J., Matzke M. & Matzke A. (1999) Integrated pararetroviral sequences define a unique class of dispersed repetitive DNA in plants. *Proceedings of the National Academy of Sciences* **96**, 13241-6.

Jędrzejuk A., Meyer L. & Serek M. (2017) Characterization of interspecific hybrids of *Petunia* and *Calibrachoa* by multiplex PCR, DNA content, and chromosome number. *The Journal of Horticultural Science and Biotechnology* **92**, 493-501.

John H., Birnstiel M. & Jones K. (1969) RNA-DNA hybrids at the cytological level. *Nature* **223**, 582-87.

Jurka J., Kapitonov V.V., Kohany O. & Jurka M.V. (2007) Repetitive sequences in complex genomes: structure and evolution. *Annual Review of Genomics and Human Genetics.* **8**, 241-59.

Jurka J., Kapitonov V.V., Pavlicek A., Klonowski P., Kohany O. & Walichiewicz J. (2005) Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research* **110**, 462-7.

Jurka J., Walichiewicz J. & Milosavljevic A. (1992) Prototypic sequences for human repetitive DNA. *Journal of Molecular Evolution* **35**, 286-91.

Jussieu A.d. (1803) Sur le *Petunia* genera nouveau de la famille des plants solanees. *Annales du Muséum d'histoire naturelle*. **2**, 214-6.

Kapitonov V.V. & Jurka J. (1999) The long terminal repeat of an endogenous retrovirus induces alternative splicing and encodes an additional carboxy-terminal sequence in the human leptin receptor. *Journal of Molecular Evolution* **48**, 248-51.

Kapitonov V.V. & Jurka J. (2008) A universal classification of eukaryotic transposable elements implemented in Repbase. *Nature Reviews Genetics* **9**, 411-12.

Kearse M., Moir R., Wilson A., Stones-Havas S., Cheung M., Sturrock S., Buxton S., Cooper A., Markowitz S. & Duran C. (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-9.

Kelley D.R., Schatz M.C. & Salzberg S.L. (2010) Quake: quality-aware detection and correction of sequencing errors. *Genome Biology* **11**, R116-28.

Kidwell M.G. (2002) Transposable elements and the evolution of genome size in eukaryotes. *Genetica* **115**, 49-63.

Kidwell M.G. (2005) Transposable elements. In: *The Evolution of the genome* (pp. 165-221). Elsevier.

Kim S., Park M., Yeom S.-I., Kim Y.-M., Lee J.M., Lee H.-A., Seo E., Choi J., Cheong K. & Kim K.-T. (2014) Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. *Nature genetics* **46**, 270-78.

King A.M., Lefkowitz E., Adams M.J. & Carstens E.B. (2011) *Virus taxonomy: ninth report of the International Committee on Taxonomy of Viruses*. Elsevier.

Kingsford C., Schatz M.C. & Pop M. (2010) Assembly complexity of prokaryotic genomes using short reads. *BMC bioinformatics* **11**, 21-31.

Klemm N., Dohm A., Sander U. & Capo A. (2017) Petunia plants having a unique flower color pattern. *Patent Application Publication* 461, 822.

Kokubun H., Nakano M., Tsukamoto T., Watanabe H., Hashimoto G., Marchesi E., Bullrich L., Basualdo I.L., Kao T.-h. & Ando T. (2006) Distribution of self-compatible and self-incompatible populations of *Petunia axillaris* (Solanaceae) outside Uruguay. *Journal of Plant Research* **119**, 419.

Korir N.K., Han J., Shangguan L., Wang C., Kayesh E., Zhang Y. & Fang J. (2013) Plant variety and cultivar identification: advances and prospects. *Critical Reviews in Biotechnology* **33**, 111-25.

Krassovsky K. & Henikoff S. (2014) Distinct chromatin features characterize different classes of repeat sequences in *Drosophila melanogaster*. *BMC Genomics* **15**, 105-18.

Kriedt R.A., Cruz G.M., Bonatto S.L. & Freitas L.B. (2014) Novel transposable elements in Solanaceae: evolutionary relationships among Tnt1-related sequences in wild *Petunia* species. *Plant Molecular Biology Reporter* **32**, 142-52.

Krupovic M., Blomberg J., Coffin J.M., Dasgupta I., Fan H., Geering A.D., Gifford R., Harrach B., Hull R. & Johnson W. (2018) *Ortervirales*: New Virus Order Unifying Five Families of Reverse-Transcribing Viruses. *Journal of Virology* **92**, e00515-18.

Kubis S., Schmidt T. & Heslop-Harrison J.S. (1998) Repetitive DNA elements as a major component of plant genomes. *Annals of Botany* **82**, 45-55.

Kuhn G.C., Küttler H., Moreira-Filho O. & Heslop-Harrison J.S. (2011) The 1.688 repetitive DNA of *Drosophila*: concerted evolution at different genomic scales and association with genes. *Molecular Biology and Evolution* **29**, 7-11.

Kulcheski F.R., Muschner V.C., Lorenz-Lemke A.P., Stehmann J.R., Bonatto S.L., Salzano F.M. & Freitas L.B. (2006) Molecular phylogenetic analysis of *Petunia* juss.(Solanaceae). *Genetica* **126**, 3-14.

Kumar A. & Bennetzen J.L. (1999) Plant retrotransposons. *Annual Review of Genetics* **33**, 479-532.

Kunii M., Kanda M., Nagano H., Uyeda I., Kishima Y. & Sano Y. (2004) Reconstruction of putative DNA virus from endogenous *Rice tungro bacilliform virus*-like sequences in the rice genome: implications for integration and evolution. *BMC Genomics* **5**, 80-94.

Kurtz S., Narechania A., Stein J.C. & Ware D. (2008) A new method to compute *K*-mer frequencies and its application to annotate large repetitive plant genomes. *BMC Genomics* **9**, 517-35.

Lamb J.C., Meyer J.M., Corcoran B., Kato A., Han F. & Birchler J.A. (2007) Distinct chromosomal distributions of highly repetitive sequences in maize. *Chromosome Research* **15**, 33-49.

Lei R., Jiang H., Hu F., Yan J. & Zhu S. (2017) Chlorophyll fluorescence lifetime imaging provides new insight into the chlorosis induced by plant virus infection. *Plant Cell Reports* **36**, 327-41.

Leitch I.J. & Leitch A.R. (2013) Genome size diversity and evolution in land plants. In: *Plant Genome Diversity Volume 2* (pp. 307-22). Springer.

Lesemann D.-E. (1991) Specific cytological alterations in virus-infected plant cells. In: *Electron Microscopy of Plant Pathogens* (pp. 147-59). Springer.

Lesemann D.-E. (1996) Viruses recently detected in vegetatively propagated *Petunia*. In: *IX International Symposium on Virus Diseases of Ornamental Plants 432*, pp. 88-95.

Lesemann D. & Casper R. (1973) Electron microscopy of *Petunia vein-clearing virus*, an isometric plant virus associated with specific inclusions in petunia cells. *Phytopathology* 63, 1118-1124.

Levan A. (1937) On chromosome fragments in *Petunia*. *Svensk Botanisk Tidskrift* 31**: 1-10.**

Linskens H. (1975) Incompatibility in *Petunia*. *Proceedings of the Royal Society of London* **188**, 299-311.

Lisch D. (2013) How important are transposons for plant evolution? *Nature Reviews Genetics* **14**, 49-61.

Lively C.M. (2010) A review of Red Queen models for the persistence of obligate sexual reproduction. *Journal of Heredity* **101**, S13-S20.

Lockhart B. & Lesemann D.-E. (1998) Occurrence of *Petunia vein-clearing virus* in the USA. *Plant Disease* **82**, 262.

Lockhart B.E., Menke J., Dahal G. & Olszewski N. (2000) Characterization and genomic analysis of *Tobacco vein clearing virus*, a plant pararetrovirus that is transmitted vertically and related to sequences integrated in the host genome. *Journal of General Virology* **81**, 1579-85.

López-Flores I. & Garrido-Ramos M. (2012) The repetitive DNA content of eukaryotic genomes. In: *Repetitive DNA* (pp. 1-28). Karger Publishers.

Löwer R. (1999) The pathogenic potential of endogenous retroviruses: facts and fantasies. *Trends in Microbiology* **7**, 350-6.

Lusinska J., Majka J., Betekhtin A., Susek K., Wolny E. & Hasterok R. (2018) Chromosome identification and reconstruction of evolutionary rearrangements in *Brachypodium distachyon*, *B. stacei* and *B. hybridum*. *Annals of Botany* 122, 445–59.

Lustig A.J. (2003) Clues to catastrophic telomere loss in mammals from yeast telomere rapid deletion. *Nature Reviews Genetics* **4**, 916-23.

Ma X.-F. & Gustafson J.P. (2008) Allopolyploidization-accommodated genomic sequence changes in triticale. *Annals of Botany* **101**, 825-32.

Malinowski E. (1935) Studies on unstable characters in *Petunia*. I. The extreme flower types of the unstable race with mosaic color patterns. *Genetics* **20**, 342-56.

Manninen I. & Schulman A.H. (1993) BARE-1, a copia-like retroelement in barley (*Hordeum vulgare* L.). *Plant Molecular Biology* **22**, 829-46.

Marçais G. & Kingsford C. (2011) A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. *Bioinformatics* **27**, 764-70.

Mardis E.R. (2008) The impact of next-generation sequencing technology on genetics. *Trends in Genetics* **24**, 133-41.

Marlor R.L., Parkhurst S. & Corces V.G. (1986) The *Drosophila melanogaster* gypsy transposable element encodes putative gene products homologous to retroviral proteins. *Molecular and Cellular Biology* **6**, 1129-34.

Marthaler H. (1936) Morphologie der Chromosomen des Zellkerns von *Petunia*. *Z. indukt. Ab- stamm* u. Vererb. 72: 258-266.

Martienssen R.A. (1998) Functional genomics: probing plant gene function and expression with transposons. *Proceedings of the National Academy of Sciences* **95**, 2021-6.

Mather K. & Edwardes P. (1943) Specific differences in *Petunia*. III. Flower colour and genetic isolation. *J. Genet* **45**, 243-60.

Matsubara K., Kodama H., Kokubun H., Watanabe H. & Ando T. (2005) Two novel transposable elements in a cytochrome P450 gene govern anthocyanin biosynthesis of commercial petunias. *Gene* **358**, 121-6.

Mavric I., Blatnik A. & Ravnikar M. (1996) Viruses infecting trailing petunias in Slovenia. In: *IX International Symposium on Virus Diseases of Ornamental Plants 432*, pp. 364-7.

McKnight T.D., Riha K. & Shippen D.E. (2002) Telomeres, telomerase, and stability of the plant genome. *Plant Molecular Biology* **48**, 331-7.

Melters D.P., Bradnam K.R., Young H.A., Telis N., May M.R., Ruby J.G., Sebra R., Peluso P., Eid J. & Rank D. (2013) Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biology* **14**, R10-30.

Mendieta J.P. (2015) Transposable Element Abundance and Variability in 28 Different Species in the Family Solanaceae. Undergraduate Honors Theses. 980.

Metzker M.L. (2010) Sequencing technologies—the next generation. *Nature Reviews Genetics* **11**, 31-46.

Meyers B.C., Tingey S.V. & Morgante M. (2001) Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. *Genome Research* **11**, 1660-76.

Ming R., Hou S., Feng Y., Yu Q., Dionne-Laporte A., Saw J.H., Senin P., Wang W., Ly B.V. & Lewis K.L. (2008) The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature* **452**, 991-96.

Mishiba K.-I., Ando T., Mii M., Watanabe H., Kokubun H., Hashimoto G. & Marchesi E. (2000) Nuclear DNA content as an index character discriminating taxa in the genus *Petunia* sensu Jussieu (Solanaceae). *Annals of Botany* **85**, 665-73.

Morgante M., Hanafey M. & Powell W. (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nature Genetics* **30**, 194.

Morran L.T., Schmidt O.G., Gelarden I.A., Parrish R.C. & Lively C.M. (2011) Running with the Red Queen: host-parasite coevolution selects for biparental sex. *Science* **333**, 216-8.

Mount S.M. & Rubin G.M. (1985) Complete nucleotide sequence of the *Drosophila* transposable element copia: homology between copia and retroviral proteins. *Molecular and Cellular Biology* **5**, 1630-8.

Murphy F.A., Fauquet C.M., Bishop D.H., Ghabrial S.A., Jarvis A.W., Martelli G.P., Mayo M.A. & Summers M.D. (2012) *Virus taxonomy: classification and nomenclature of viruses*. Springer Science & Business Media.

Mustafa S.I., Schwarzacher T. & Heslop-Harrison J. (2018) Complete mitogenomes from Kurdistani sheep: abundant centromeric nuclear copies representing diverse ancestors. *Mitochondrial DNA Part A*, 1-14.

Nacheva E.P., Ward K.N., Brazma D., Virgili A., Howard J., Leong H.N. & Clark D.A. (2008) Human herpesvirus 6 integrates within telomeric regions as evidenced by five different chromosomal sites. *Journal of Medical Virology* **80**, 1952-8.

Nagaki K., Cheng Z., Ouyang S., Talbert P.B., Kim M., Jones K.M., Henikoff S., Buell C.R. & Jiang J. (2004) Sequencing of a rice centromere uncovers active genes. *Nature Genetics* **36**, 138-45.

Navot N., Zeidan M., Pichersky E., Zamir D. & Czosnek H. (1992) Use of the polymerase chain reaction to amplify *Tomato yellow leaf curl virus* DNA from infected plants and viruliferous whiteflies. *Phytopathology* **82**, 1199-202.

Ndowora T., Dahal G., LaFleur D., Harper G., Hull R., Olszewski N.E. & Lockhart B. (1999) Evidence that badnavirus infection inmusacan originate from integrated pararetroviral sequences. *Virology* **255**, 214-20.

Nehra C. & Gaur R. (2015) Molecular characterization of *Chilli leaf curl virus* infecting new host plant *Petunia hybrida* in India. *Virus Genes* **50**, 58-62.

Nelson M.I. & Holmes E.C. (2007) The evolution of epidemic influenza. *Nature Reviews Genetics* **8**, 196-05.

Neumann P., Požárková D. & Macas J. (2003) Highly abundant pea LTR retrotransposon Ogre is constitutively transcribed and partially spliced. *Plant Molecular Biology* **53**, 399-410.

Noreen F., Akbergenov R., Hohn T. & Richert-Pöggeler K.R. (2007) Distinct expression of endogenous *Petunia vein clearing virus* and the DNA transposon *dTph1* in two *Petunia hybrida* lines is correlated with differences in histone modification and siRNA production. *Plant Journal* **50**, 219-29.

Nouri-Aria K.T. (2008) *In situ* hybridization. In: *Allergy Methods and Protocols* (pp. 331-47). Springer.

Novák P., Ávila Robledillo L., Koblížková A., Vrbová I., Neumann P. & Macas J. (2017) TAREAN: A computational tool for identification and characterization of satellite DNA from unassembled short reads. *Nucleic Acids Research* 45, 111-21.

Novák P., Neumann P. & Macas J. (2010) Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC bioinformatics* **11**, 378-90.

Novák P., Neumann P., Pech J., Steinhaisl J. & Macas J. (2013) RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics* **29**, 792-3.

Novoa R.R., Calderita G., Arranz R., Fontana J., Granzow H. & Risco C. (2005) Virus factories: associations of cell organelles for viral replication and morphogenesis. *Biology of the Cell* **97**, 147-72.

OEPP/EPPO (2008) Production of pathogen-tested herbaceous ornamentals. *Bulletin OEPP/EPPO Bulletin* 38, 31–52.

Ohtsubo H., Umeda M. & Ohtsubo E. (1991) Organization of DNA sequences highly repeated in tandem in rice genomes. *The Japanese Journal of Genetics* **66**, 241-54.

Olschowski S., Meyer L., Mußmann V. & Serek M. (2012) Investigation of crossing barriers between *Calibrachoa* and *Petunia* species. In: *II All Africa Horticulture Congress 1007*, pp. 765-9.

Paterson A.H., Bowers J.E., Bruggmann R., Dubchak I., Grimwood J., Gundlach H., Haberer G., Hellsten U., Mitros T. & Poliakov A. (2009) The *Sorghum bicolor* genome and the diversification of grasses. *Nature* **457**, 551-56.

Pevzner P.A., Tang H. & Waterman M.S. (2001) An Eulerian path approach to DNA fragment assembly. *Proceedings of the National Academy of Sciences* **98**, 9748-53.

Pluhar S., Erickson L. & Pauls K. (2001) Effects of tissue culture on a highly repetitive DNA sequence (E180 satellite) in *Medicago sativa*. *Plant Cell, Tissue and Organ Culture* **67**, 195-9.

Presting G.G., Frary A., Pillen K. & Tanksley S.D. (1996) Telomere-homologous sequences occur near the centromeres of many tomato chromosomes. *Molecular and General Genetics MGG* **251**, 526-31.

Pritham E.J. (2009) Transposable elements and factors influencing their success in eukaryotes. *Journal of Heredity* **100**, 648-55.

Pröls F. & Meyer P. (1992) The methylation patterns of chromosomal integration regions influence gene activity of transferred DNA in *Petunia hybrida*. *The Plant Journal* **2**, 465-75.

Ratcliff F., Harrison B.D. & Baulcombe D.C. (1997) A similarity between viral defense and gene silencing in plants. *Science* **276**, 1558-60.

Reck-Kortmann M., Silva-Arias G.A., Segatto A.L.A., Mäder G., Bonatto S.L. & de Freitas L.B. (2014) Multilocus phylogeny reconstruction: new insights into the evolutionary history of the genus Petunia. *Molecular phylogenetics and evolution* **81**, 19-28.

Reuter G., Giarre M., Farah J., Gausz J., Spierer A. & Spierer P. (1990) Dependence of position-effect variegation in Drosophila on dose of a gene encoding an unusual zinc-finger protein. *Nature* **344**, 219.

Richert-Pöggeler K. & Shepherd R. (1997) *Petunia vein-clearing virus*: a plant pararetrovirus with the core sequences for an integrase function. *Virology* **236**, 137-46.

Richert-Pöggeler K.R. & Lesemann D.-E. (2007) *Petunia vein clearing virus*. *AAB Descriptions of Plant Viruses*.

Richert-Pöggeler K.R. & Schwarzacher T. (2009) Impact of Retroelements in Shaping the *Petunia* Genome. In: *Petunia* (pp. 343-63). Springer.

Richert-Pöggeler K.R., Noreen F., Schwarzacher T., Harper G. & Hohn T. (2003) Induction of infectious *Petunia vein clearing* (pararetro) *virus* from endogenous provirus in petunia. *EMBO Journal* **22**, 4836-45.

Richert (1992) Untersuchungen zur Charakterisierung des *Petunia vein clearing virus* (PVCV), ein samenübertragbares Pararetrovirus. PhD thesis, , Göttingen, Germany. In: *PhD thesis*. University of Göttingen, Göttingen, Germany.

Ricker N., Qian H. & Fulthorpe R. (2012) The limitations of draft assemblies for understanding prokaryotic adaptation and evolution. *Genomics* **100**, 167-75.

Robledillo L.Á., Koblížková A., Novák P., Böttinger K., Vrbová I., Neumann P., Schubert I. & Macas J. (2018) Satellite DNA in *Vicia faba* is characterized by remarkable diversity in its sequence composition, association with centromeres, and replication timing. *Scientific Reports* **8**, 5838-49.

Rudd M.K. & Willard H.F. (2004) Analysis of the centromeric regions of the human genome assembly. *Trends in Genetics* **20**, 529-33.

Rybicki E.P. & Hughes F.L. (1990) Detection and typing of *Maize streak virus* and other distantly related geminiviruses of grasses by polymerase chain reaction amplification of a conserved viral sequence. *Journal of General Virology* **71**, 2519-26.

Salih M. (2017) Nuclear and chloroplast genome diversity in apomictic microspecies of *Taraxacum*. University of Leicester, Department of Genetics, PhD thesis.

Salih R.H.M., Majeský Ľ., Schwarzacher T., Gornall R. & Heslop-Harrison P. (2017) Complete chloroplast genomes from apomictic *Taraxacum* (Asteraceae): Identity and variation between three microspecies. *PloS one* **12**, e0168008.

Salvo-Garrido H., Travella S., Schwarzacher T., Harwood W. & Snape J. (2001) An efficient method for the physical mapping of transgenes in barley using *in situ* hybridization. *Genome* **44**, 104-10.

SanMiguel P. & Bennetzen J.L. (1998) Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. *Annals of Botany* **82**, 37-44.

SanMiguel P., Tikhonov A., Jin Y.-K., Motchoulskaia N., Zakharov D., Melake-Berhan A., Springer P.S., Edwards K.J., Lee M. & Avramova Z. (1996) Nested retrotransposons in the intergenic regions of the maize genome. *Science* **274**, 765-8.

Schaack S., Gilbert C. & Feschotte C. (2010) Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends in Ecology & Evolution* **25**, 537-46.

Schaad M.C., Jensen P.E. & Carrington J.C. (1997) Formation of plant RNA virus replication complexes on membranes: role of an endoplasmic reticulum-targeted viral protein. *The EMBO Journal* **16**, 4049-59.

Schmidt T. (1999) LINEs, SINEs and repetitive DNA: non-LTR retrotransposons in plant genomes. *Plant Molecular Biology* **40**, 903-10.

Schmidt T. & Heslop-Harrison J. (1998) Genomes, genes and junk: the large-scale organization of plant chromosomes. *Trends in Plant Science* **3**, 195-9.

Schnable P.S., Ware D., Fulton R.S., Stein J.C., Wei F., Pasternak S., Liang C., Zhang J., Fulton L. & Graves T.A. (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science* **326**, 1112-5.

Schulman A.H. & Kalendar R. (2005) A movable feast: diverse retrotransposons and their contribution to barley genome dynamics. *Cytogenetic and Genome Research* **110**, 598-605.

Schwarzacher T. (2003) Meiosis, recombination and chromosomes: a review of gene isolation and fluorescent *in situ* hybridization data in plants. *Journal of Experimental Botany* **54**, 11-23.

Schwarzacher T. & Heslop-Harrison P. (2000) Practical *in situ* hybridization. BIOS Scientific. Oxford: Bios Scientific Publishers Limited.

Sepsi A., Higgins J.D., Heslop-Harrison J.S. & Schwarzacher T. (2017) CENH 3 morphogenesis reveals dynamic centromere associations during synaptonemal complex formation and the progression through male meiosis in hexaploid wheat. *The Plant Journal* **89**, 235-49.

Sharma A.K. & Sharma A. (2001) Chromosome painting–principles, strategies and scope. *Methods in Cell Science* **23**, 1-5.

Sharma S. & Raina S. (2005) Organization and evolution of highly repeated satellite DNA sequences in plant chromosomes. *Cytogenetic and Genome Research* **109**, 15-26.

Shendure J. & Ji H. (2008) Next-generation DNA sequencing. *Nature Biotechnology* **26**, 1135-45.

Sikron N., Cohen J., Shoval S. & Gera A. (1995) Virus diseases in petunia. *Phytoparasitica* **23**, 273.

Sink K. (1984) Taxonomy. In: *Petunia* (pp. 3-9). Springer.

Slotkin R.K. & Martienssen R. (2007) Transposable elements and the epigenetic regulation of the genome. *Nature Reviews Genetics* **8**, 272-85.

S**mit A. & Hubley R.** (2015) GreenP. *RepeatMasker at http://repeatmasker. org Retrieved online*.

Smith F. & Oud J. (1972) The possibility to distinguish chromosomes of *Petunia hybrida* by quinacrine fluorescence. *Genetica* **43**, 589-96.

Smith F., Oud J. & De Jong J. (1973) A standard karyogram of *Petunia hybrida* Hort. *Genetica* **44**, 474-84.

Smith L. (1966) Flora Illustrada Catarinense. Solanaceas: Herbario. *Barbosa Rodrigues*, 261-91.

Staginnus C., Gregor W., Mette M.F., Teo C.H., Borroto-Fernandez E.G., Machado M.L.d.C., Matzke M. & Schwarzacher T. (2007) Endogenous pararetroviral sequences in tomato (*Solanum lycopersicum*) and related species.(Research article)(Clinical report). *BMC Plant Biology* **7**, 24-40.

Stavolone L. & Lionetti V. (2017) Extracellular matrix in plants and animals: Hooks and locks for viruses. *Frontiers in Microbiology* **8**, 1760-68.

Stavolone L., Villani M.E., Leclerc D. & Hohn T. (2005) A coiled-coil interaction mediates *Cauliflower mosaic virus* cell-to-cell movement. *Proceedings of the National Academy of Sciences* **102**, 6219-24.

Stebbins G. & Dunn L. (1950) Chapter XII: the karyotype. pp. 442-72. Columbia University Press New York.

Steere W.C. (1932) Chromosome behavior in triploid *Petunia hybrida*. *American Journal of Botany* **19**, 340-57.

Stehmann J. & Bohs L. (2007) Nuevas combinaciones en Solanaceae. *Darwiniana* **45**, 240-1.

Stehmann J., Semir J., Dutilh J. & Forni-Martins E. (1996) Solanaceae. IOPB chromosome data. *International Organization of Plant Biosystematics Newsletter*, 26-7.

Stehmann J.R., Lorenz-Lemke A.P., Freitas L.B. & Semir J. (2009) The genus *Petunia*. In: *Petunia* (pp. 1-28). Springer.

Stehmann J.R. & Semir J. (1997) A new species and new combinations in *Calibrachoa* (Solanaceae). *Novon*, 417-9.

Stewart A.D., Logsdon J.M. & Kelley S.E. (2005) An empirical study of the evolution of virulence under both horizontal and vertical transmission. *Evolution* **59**, 730-9.

Stout A.B. (1952) Reproduction in *Petunia*. *Memoirs of the Torrey Botanical Club* **20**, 1-202.

Suoniemi A., Tanskanen J. & Schulman A.H. (1998) Gypsy-like retrotransposons are widespread in the plant kingdom. *The Plant Journal* **13**, 699-705.

Sýkorová E., Lim K., Kunicka Z., Chase M., Bennett M., Fajkus J. & Leitch A. (2003) Telomere variability in the monocotyledonous plant order Asparagales. *Proceedings of the Royal Society of London B: Biological Sciences* **270**, 1893-904.

Takehisa S. (1964) The karyotype of Petunia hybrida and the differential chromosome condensation. *The Japanese Journal of Genetics* **38**, 237-43.

Talbert P., Kasinathan S. & Henikoff S. (2018) Simple and Complex Centromeric Satellites in *Drosophila* Sibling Species. *Genetics* 208, 977-990.

Tamura K., Stecher G., Peterson D., Filipski A. & Kumar S. (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* **30**, 2725-9.

Temin H. (1985) Reverse transcription in the eukaryotic genome: retroviruses, pararetroviruses, retrotransposons, and retrotranscripts. *Molecular Biology and Evolution* **2**, 455-68.

Tenaillon M.I., Hufford M.B., Gaut B.S. & Ross-Ibarra J. (2011) Genome size and transposable element content as determined by high-throughput sequencing in maize and *Zea luxurians*. *Genome Biology and Evolution* **3**, 219-29.

Teo C.H., Tan S.H., Ho C.L., Faridah Q.Z., Othman Y.R., Heslop-Harrison J.S., Kalendar R. & Schulman A.H. (2005) Genome constitution and classification using retrotransposon-based markers in the orphan crop banana. *Journal of Plant Biology* **48**, 96-105.

Torres G.A., Gong Z., Iovene M., Hirsch C.D., Buell C.R., Bryan G.J., Novák P., Macas J. & Jiang J. (2011) Organization and evolution of subtelomeric satellite repeats in the potato genome. *G3: Genes, Genomes, Genetics* **1**, 85-92.

Tóth G., Gáspári Z. & Jurka J. (2000) Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Research* **10**, 967-81.

Tsukamoto T., Ando, T., Kokubun, H., Watanabe, H., Tanaka, R., Hashimoto, G., Marchesi, E. & Kao T. (1998) Differentiation in the status of self-incompatibility among all natural taxa of *Petunia* (Solanaceae). *Acta Phytotaxonomica et Geobotanica* **49**, 115-33.

Turner K.A., Sit T.L., Callaway A.S., Allen N.S. & Lommel S.A. (2004) Red clover necrotic mosaic virus replication proteins accumulate at the endoplasmic reticulum. *Virology* **320**, 276-90.

Ugarković Đ. & Plohl M. (2002) Variation in satellite DNA profiles—causes and effects. *The EMBO Journal* **21**, 5955-9.

Usai G., Mascagni F., Zuccolo A., Giordani T., Natali L. & Cavallini A. (2017) Characterization of the repetitive component of the fig (*Ficus carica*) genome. In: *LX SIGA Annual Congress*.

Van Der Donk J. (1974) Gene activity and the incompatibility reaction in *Petunia*. In: *Proceedings International Symposium on Fertilization in Higher Plants*. North-Holland Publishing Co., Amsterdam, pp. 279–292.

Van Dijk E.L., Auger H., Jaszczyszyn Y. & Thermes C. (2014) Ten years of next-generation sequencing technology. *Trends in Genetics* **30**, 418-26.

van Helvoort T. & Sankaran N. (2018) How Seeing Became Knowing: The Role of the Electron Microscope in Shaping the Modern Definition of Viruses. *Journal of the History of Biology*, 1-36.

Van Meggelen I. (2005) Flower trails in den Niederlanden, Erste Kreuzung zwischen Petunia and Calibrachoa vorgestellt. *Dtsch Gartenbau* **25**, 10-6.

Van Regenmortel M.H. (1990) Virus species, a much overlooked but essential concept in virus classification. *Intervirology* **31**, 241-54.

Van Stedum S. & King W. (2002) Basic FISH techniques and troubleshooting. In: *Molecular Cytogenetics* (pp. 51-63). Springer.

Vandenbussche M., Chambrier P., Rodrigues Bento S. & Morel P. (2016) *Petunia*, your next supermodel? *Frontiers in Plant Science* **7**, 72-83.

Vergnaud G. & Denoeud F. (2000) Minisatellites: mutability and genome architecture. *Genome Research* **10**, 899-907.

Vershinin A.V. & Heslop-Harrison J.S. (1998) Comparative analysis of the nucleosomal structure of rye, wheat and their relatives. *Plant Molecular Biology* **36**, 149-61.

Vershinin A.V., Schwarzacher T. & Heslop-Harrison J.S. (1995) The large-scale genomic organization of repetitive DNA families at the telomeres of rye chromosomes. *The Plant Cell* **7**, 1823-33.

Vicient C.M. & Casacuberta J.M. (2017) Impact of transposable elements on polyploid plant genomes. *Annals of Botany* **120**, 195-207.

Vig B.K. (1994) Do specific nucleotide bases constitute the centromere? *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **309**, 1-10.

Vitte C. & Panaud O. (2005) LTR retrotransposons and flowering plant genome size: emergence of the increase/decrease model. *Cytogenetic and Genome Rresearch* **110**, 91-107.

Volpi E.V. & Bridger J.M. (2008) FISH glossary: an overview of the fluorescence *in situ* hybridization technique. *Biotechniques* **45**, 385-409.

Wallau G.L., Vieira C. & Loreto É.L.S. (2018) Genetic exchange in eukaryotes through horizontal transfer: connected by the mobilome. *Mobile DNA* **9**, 6-22.

Wang R.C., Smogorzewska A. & De Lange T. (2004) Homologous recombination generates T-loop-sized deletions at human telomeres. *Cell* **119**, 355-68.

Wang Y., Diehl A., Wu F., Vrebalov J., Giovannoni J., Siepel A. & Tanksley S.D. (2008) Sequencing and comparative analysis of a conserved syntenic segment in the Solanaceae. *Genetics* 180, 391–408.

Watanabe H., Ando T., Iida S.-i., Buto K.-i., Tsukamoto T., Kokubun H., Hashimoto G. & Marchesi E. (1997) Cross-compatibility of *Petunia pubescens* and *P. pygmaea* with native taxa of *Petunia*. *Journal of the Japanese Society for Horticultural Science* **66**, 607-12.

Watanabe H., Ando T., Iida S.-i., Suzuki A., Buto K.-i., Tsukamoto T., Hashimoto G. & Marchesi E. (1996) Cross compatibility of *Petunia* cultivars and *P. axillaris* with native taxa of *Petunia* in relation to their chromosome number. *Journal of the Japanese Society for Horticultural Science* **65**, 625-34.

Weitzel J.N., Hows J.M., Jeffreys A.J., Min G.L. & Goldman J.M. (1988) Use of a hypervariable minisatellite DNA probe (33.15) for evaluating engraftment two or more years after bone marrow transplantation for aplastic anaemia. *British Journal of Haematology* **70**, 91-7.

Wergin W. (1936) Zytologische Untersuchungen on *Petunia hybrida* hort. *Z, indukt. Abstamm* u. Vererb. 71: 120-155.

Wesley S.V., Helliwell C.A., Smith N.A., Wang M., Rouse D.T., Liu Q., Gooding P.S., Singh S.P., Abbott D. & Stoutjesdijk P.A. (2001) Construct design for efficient, effective and high-throughput gene silencing in plants. *The Plant Journal* **27**, 581-90.

White S.E., Habera L.F. & Wessler S.R. (1994) Retrotransposons in the flanking regions of normal plant genes: a role for copia-like elements in the evolution of gene structure and expression. *Proceedings of the National Academy of Sciences* **91**, 11792-6.

Wicker T., Sabot F., Hua-Van A., Bennetzen J.L., Capy P., Chalhoub B., Flavell A., Leroy P., Morgante M. & Panaud O. (2007) A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics* **8**, 973-82.

Wijnands D., Bos J., Wijsman H., Schneider F., Brickell C. & Zimmer K. (1986) Proposal to conserve 7436 petunia with *P. nyctaginiflora* as typs. cons.(Solanaceae). *Taxon* **35**, 748-9.

Wijsman H. (1982) On the interrelationships of certain species of Petunia : 1. Taxonomic notes on the parental species of *Petunia hybrida*. *Acta Botanica Neerlandica* **31**, 477-90.

Wijsman H. (1983) On the interrelationships of certain species of Petunia II. Experimental data: Crosses between different taxa. *Acta Botanica Neerlandica* **32**, 97-107.

Wijsman H. (1990) On the inter-relationships of certain species of *Petunia* VI. New names for the species of *Calibrachoa* formerly included into *Petunia* (Solanaceae). *Acta Botanica Neerlandica* **39**, 101-2.

Wijsman H. & De Jong J. (1985) On the interrelationships of certain species of *Petunia* IV. Hybridization between *P. linearis* and *P. calycina* and nomenclatorial consequences in the petunia group. *Acta Botanica Neerlandica* **34**, 337-49.

Williams D., Trimble W.L., Shilts M., Meyer F. & Ochman H. (2013) Rapid quantification of sequence repeats to resolve the size, structure and contents of bacterial genomes. *BMC Genomics* **14**, 537-48.

Wood M.L. & Royle N.J. (2017) Chromosomally integrated human herpesvirus 6: models of viral genome release from the telomere and impacts on human health. *Viruses* **9**, 184-98.

Yasaka R., Nguyen H.D., Ho S.Y., Duchêne S., Korkmaz S., Katis N., Takahashi H., Gibbs A.J. & Ohshima K. (2014) The temporal evolution and global spread of Cauliflower mosaic virus, a plant pararetrovirus. *PloS one* **9**, e85641.

Zakrzewski F., Wenke T., Holtgräwe D., Weisshaar B. & Schmidt T. (2010) Analysis of ac 0 t-1 library enables the targeted identification of minisatellite and satellite families in *Beta vulgaris*. *BMC Plant Biology* **10**, 8-22.

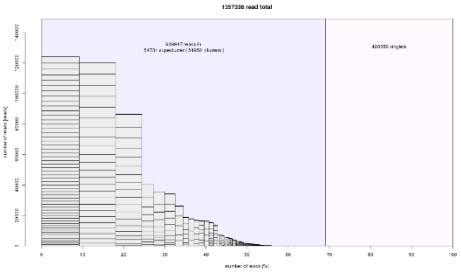
Zeidan M., Sikron N., Cohen J. & Gera A. (2000) Improved detection of *Petunia vein clearing* caulimovirus. *Hortscience* **35**, 1279-82.

Zerbino D.R. & Birney E. (2008) Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Research* **18**, 821-9.

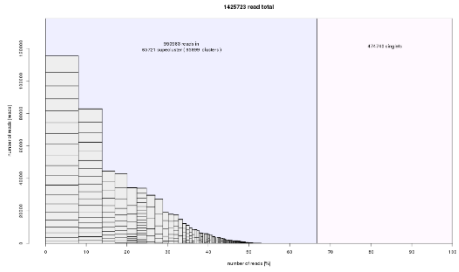
Zheng X.M. & Ge S. (2010) Ecological divergence in the presence of gene flow in two closely related Oryza species (Oryza rufipogon and O. nivara). *Molecular Ecology* **19**, 2439-54.

Zlotorynski E., Rahat A., Skaug J., Ben-Porat N., Ozeri E., Hershberg R., Levi A., Scherer S.W., Margalit H. & Kerem B. (2003) Molecular basis for expression of common and rare fragile sites. *Molecular and Cellular Biology* **23**, 7143-51.

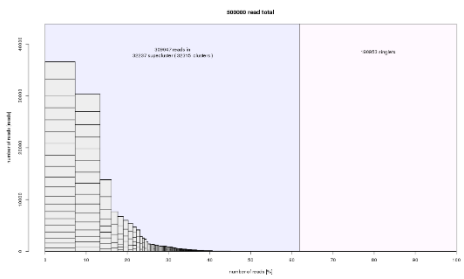
Appendices



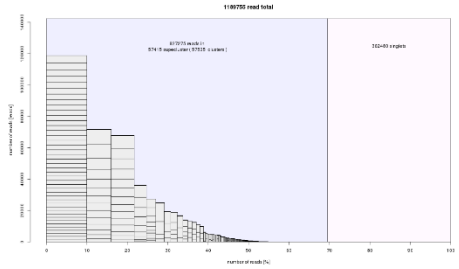
P. axillaris (*PaxiN*)



P. inflata (*PinfS6*)



P. hybrida (*PhybR27*)



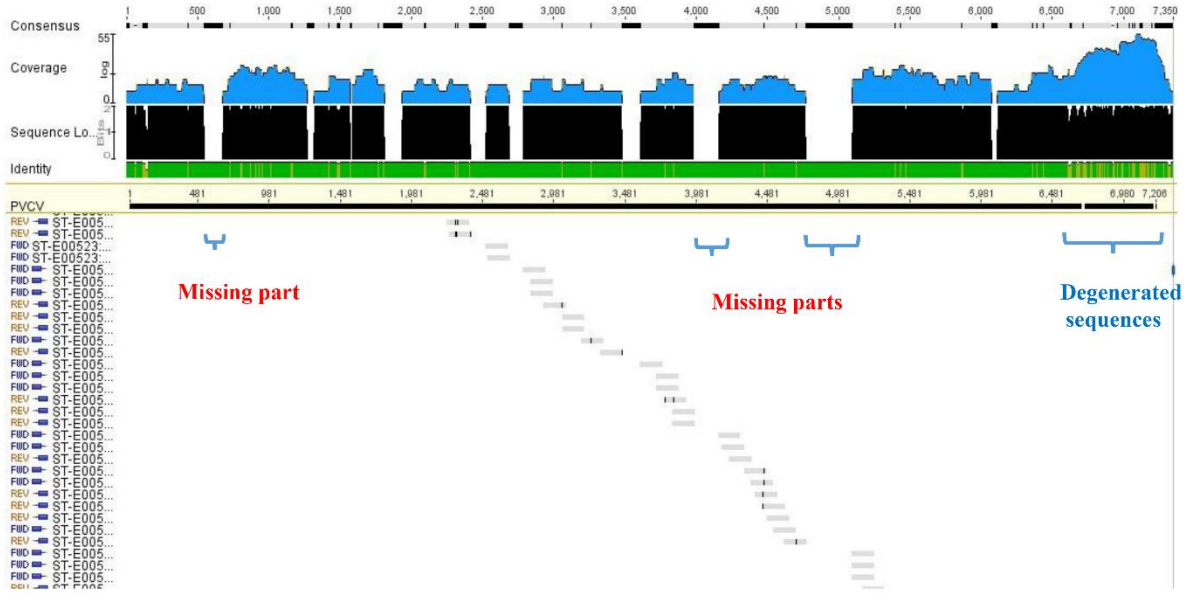
P. parodii (PparS7)

Appendix 3.1 Summary histogram of all used raw reads from RepeatExplorer report showing about 60% of the repetitive DNA components and ratios.

Show 1000 entries Search:

	Cluster	Super cluster	Genome Proportion [%]	Genome Proportion Adjusted [%]	Size real	Graph layout	Similarity based annotation	Pbs score	Ltr detection	Satellite probability	TAREAN classification	Consensus length
1	1	1	0.68	0.68	3376		0.06% Class_I:LTR:Ty3_gypsy/chromovirus/CRM:Ty3-INT 0.03% Class_I:LTR:Ty3_gypsy/chromovirus/CRM:Ty3-RT 0.03% Class_I:LTR:Ty1_copia/TAR:Ty1-RT		Met	2.64e-21	Other	
2	2	2	0.66	0.66	3325		0.03% organelle/mitochondria			2.34e-21	Other	
3	3	3	0.65	0.65	3237					2.7e-22	Other	
4	4	1	0.62	0.62	3108		17.57% Class_I:LTR:Ty3_gypsy/chromovirus/CRM:Ty3-PROT 12.29% Class_I:LTR:Ty3_gypsy/chromovirus/CRM:Ty3-GAG			9.81e-23	Other	
5	5	2	0.51	0.51	2537		0.04% organelle/mitochondria			1.65e-22	Other	
6	6	4	0.49	0.49	2459		24.16% Class_I:LTR:Ty1_copia/SIRE:Ty1-INT 6.02% Class_I:LTR:Ty1_copia/SIRE:Ty1-PROT 3.54% Class_I:LTR:Ty1_copia/SIRE:Ty1-RT 0.77% Class_I:LTR:Ty1_copia/Tork:Ty1-INT 0.73% Class_I:LTR:Ty1_copia/Ivana:Ty1-INT 0.73% Class_I:LTR:Ty1_copia/Ivana:Ty1-GAG 0.08% Class_I:LINE:LINE-RT 0.04% Class_I:LTR:Ty1_copia/SIRE:Ty1-GAG 0.04% Class_I:LTR:Ty3_gypsy/chromovirus/Tcn1:Ty3-RT 0.04% Class_I:LTR:Ty1_copia/Tork:Ty1-GAG			1.44e-23	Other	

Appendix 3.2 The first page of RepeatExplorer report showing repetitive DNA components with list starts with highest proportion clusters to the smallest ones.



Appendix 3.3 Mapped *P. axillaris* subsp *parodii* S7 reads (*PparS7*) to PVCV reference, showing nine missing parts within the entire PVCV sequence and degenerated part at the end.

```
;ID PaxiV DNA ; PLN ; 7170 BP
;XX
;DE Consensus Florendovirus in Petunia axillaris.
;XX
;AC .
;XX
;DT 04-Jun-2018 (Rel. -1, Created)
;DT 04-Jun-2018 (Rel. -1, Last updated, Version 1)
;XX
;KW Caulimoviridae; Integrated Virus; PaxiV.
;XX
;OS Petunia axillaris
;XX
;OC Petunia axillaris
;OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
;OC Spermatophyta; Magnoliophyta; eudicotyledons; Gunneridae;
;OC Pentapetalae; asterids; lamiids; Solanales; Solanaceae;
;OC Petunioideae; Petunia.
;XX
;RN [1] (bases 1 to 7170)
;RA Alisawi,O., Richert-Poggeler,K.,
;RA Heslop-Harrison,P. and Schwarzacher,T.
;RT New integrated Florendovirus in Petunia axillaris.
;RL Direct Submission to RR (04-Jun-2018)
;XX
;CC New integrated Florendovirus in Petunia axillaris.
```

```
;XX
;FT CDS 6704..1263
;FT /product="PaxiV_1p"
;FT /translation="LKVIIEKLVTVQEPIIKMEGSSQSVLNSINLSTEI
;FT DKISKIENNLLNWNIPKEKLEKIYEVGGFDFINRFNIKTCESTIAINQS
;FT VETIQLLSIKDIEKYRKSYKFLHIGLVQVAVKPLHRLGLDIPICLLLRD
;FT NRFLNFDDSLIAVLQSNMANGPVYFNCYPNFSVDINDPTILDTLTLNVK
;FT TKHLNSKHQAREIAIIYRVYYKLMKTTLAPKAKPSRAHKGVTMLMEANI
;FT KHHNIFVPRLLKWDEVLSKDEWHFDSITEPKVEDESIQIEQVIQHTDGS
;FT IDLKFLKSRSFNATSSSSRMSFSRPSTSCPLEEEDIISKNLKGVNFSKK
;FT IPQPEYKPVIINQPSSPTFSEINDSLGMITLEEPFEPDIKYLQQLWVKP
;FT ENNIKREWYRNNYTLEQRKSFRDKWMKDMKKWKCSLEFFKWFEYTGKLA
;FT NCESLNMVVKKWYTKSNKIIESTIPPLEEINIPVGDKVVIGNPFKRTSI
;FT DAQGEPKNKDINKLIEQSNYTNQILHVIANQVEEARSKSSFKPFPEKRI
;FT KFEDNIEENPIFRLPSMSKDKFPSLPKGGLEIKGDVLDEITKRLKSSLT
;FT INTIPKEKENSVKTLSELNNQLNKLGSNYHHKKNYYKKPTFPDIQYEEE
;FT HLVSFSSHSGDAITEWNIDGQAEHQILNRLHEMGIAATAYKIKGNTDKQ
;FT AATLLSYGFTGMLKNWWDNYLSDKDRETVLNATTIIKIVKQEGSSTNQI
;FT IEEQVREDAINTLLYAIIKHFIGEPKVFKDRSLEILSNLSCPKLHDFKW
;FT YKDIFLNKVMIREDCNHSYWKEKFISGLPHLFADKVRQKIKDRFNGEIP
;FT YDGLTYGDIVSFINITALELCTDMKLKNQLKRDYKITKKELGNFCSDFG
;FT YAKITAPSTEATKRSKRINRKDKSSNSNYKRKSSKRYRKKIPKNASQSK
;FT QPSFKERCYNCGKKGHKANDCRVKTKKKNKINSLELDEDIKNKIYAILD
;FT ENDNSEEESSSTESISDDEQINIAYESSDSYTSESECDCPRGLCTCGTG
;FT SIKVISNENKEFFFDLIDQIKDEEIKREYLLKLKDIVLSRESNKPNISN
;FT PFSMKNILQKFDKKEEKISISDLYAELATLKQEVNILKQKVNNIEEENL
;FT ALELIQKTQKELNSTIIGESSNTKQVEAQDDELGVLAISKIKSQKWYVP
;FT IKLVINSEYSIQTVALLDSGADQNCIVEGLIPTKYCDKTSAKLFSASGE
;FT KLKIKYKLSNAHICNNNICFKNTFILIKDLNEEVILGTPFLTQISPFTV
;FT NTTGITTNNLGKEIIFYFLNGPKQREIDTLKERSIYKIQMIKNQINYIK
;FT DEIKIVKIEDQLKDKLIQDKILQAQKEIESEVCSDLPNAFWERKRHIVR
;FT LPYIEGFEDGSIPTKARPIQMNQELMEYCKKEINELLEKKIIRPSKSPW
;FT SCSAFYVNNNAERERGAPRLVINYKPLNKVLRWIRYPIPNKRDLLKRTY
;FT KAKIYSKFDMKSGFWQIQIDERDKYKTAFNVPFGQYEWNVMPFGLKNAP
;FT SEFQNIMNNIFNPYSSFSIVYIDDVLIFSENIDQHFKHLKIFLNIIKKN
;FT GLVVSAKKIKLFQTSIRFLGHDLSQGMYKPIYRAIEFSTKFPDVLTDKT
;FT QLQRFLGSLNYVSDFIPNLRQICEPLYQRLRKNPIPWSQEQTQTVQKVK
;FT SIVQRLPCLGIPHPDAFMIVETDASEIGFGGILKQRLSPKESEQLVRYC
;FT SGIWNGTQKNYSTIKKEILSIVLCITKFQDDLINKEFLLRVDCKAAKDV
;FT LQKDVKNLVSKQIFARWQALLSSFDFQIEFIKGEDNSLPDFLTREFLQR
;FT RHEAIISSKTSEAIP"
;XX
;DR [1] (Consensus)
;XX
;SQ Sequence 7170 BP; 2014 A; 1105 C; 1016 G; 3035 T; 0 other;
PaxiV
gactaaataatcaaagtcaagaatcttatctttacaaaagtaaagtaagatctacattattgataccact
tggcacttatttgttgtctatcttttgctaaaaatccgtgagatatgcttgagcaatcccgtgtgggtcc
tccatatctccattatcttctgctgatgctgttgatgatgctgtatccaataattcagtaatttgattta
tttcactgccatttgcttggtctaaatttgctaaggcctgttgaagttttcttttcaacgcttcctttga
taatttgtctgttgatccgatcttcgtttgaacttgctgaacctgtttttctttgttcttaggggtccac
cccttaatttttatagttttaacaaggtacctgaccttgcataatacctctatattgaattcccaactaa
taatataggacactcttttagccatgaagaatttacataactgtatatgttctggtaatgctgctagttc
ttcgttttctttctggtaattggtgtgtaactgtttgaaagaaggtgacaacacccttttcgtctgctcca
aatgtattccaccattcataaaaccatcttgggatgggatccgttgataccttttcacagaacttgacaa
accatgaatggctgccaggtcttaagtatatgaagttaaaccacgcatttttgtagtcgtaaaaattata
acctgatattttaaaattcactgataatggaaattgaatcggcgtgtgtagatgatctactggccaatcc
gcagggcttaaaattcttttaatagtaaacttggagtacctaatactcctcggatcatttttatctgcta
```

actcattttcaatctctattgaattggttgctatcaatatagcttcataaaattctcttgttttataagg
gttgtctgtttccacataatttctgtctgtgtaacaagacttgataatatctgtaattcgtttttcaact
aattttctctcaatcgcaatgattggtacctgttgactgtttatccaatcgaatttctgttgacttgctg
ttgaagagcttgcttgttcttgtttaatgggaggttgaggttgtactgcctcagaataactgtttatggt
tcttttggcataactgtttaaggattgttttgaggttcctggggttgcttctctctttaccggtgtagga
ggaggtaattgaggaaactctgctaaagccgtatatctgtttatatgtgagttctggttatgatgctgtt
caaggaattgcctcggaggttttactggatataatggcctcatgtcttctctgtaaaaattctctagtta
aaaaatctggaagactattgtcttctccttttaataaattctatttgaaaatcaaagctagataataaagc
ttgccatctggcaaatatttgttttgaaactaaatttttaacatcttttgcaaaacatctttgctgct
ttacaatcaactcttaaaagaaattctttgtttatcaaatcatcttgaaatttagttatacataaaacta
ttgataaaatttctttttaatagtactataatttttttgagttccgttccatatgccagagcaatatcg
gactaattgttcagactcttttggagagagacgctgtttaagaattcctccgaaacctatttcagaggca
tcggtttcaacaatcataaaggcatccggatgtggtatgcctagacatggaagtctttgaacaatacttt
tgacctttggactgtttgtgtttgttcttggctccaaggaattggatttttcctaagacgctgataaag
gggttcacatatttgtctaagatttggaataaaatctgagacatagttgaggcttcctaaaaatctttgt
aactgagttttatctgttaaaacatctggaaattttgtagagaattctatggctctataaattggtttat
acatgccttgagataagtcatgtcctaaaaatctgatagaggtttgaaatagtttgatctttttggcact
cactaccagcccatttttcttaataatattaagaaatatctttaaatgcttaaaatgttgatctatattt
tctgaaaatattaatacgtcatctatataaactattgagaatgaactatatggattaaatatattattca
ttatattttgaaattctgatggggcatttttaacccaaagggcataacattccattcatattggccaaa
gggcacattaaatgctgttttatacttatccctttcatctatttggatttgccaaaatcctgatttcata
tcaaatttactatatattttagccttatacgttctttttaataaatctcttttatttggaattggatatc
gtatccatcttaatactttgtttaaaggtttataatttataactaatcttggagctcctctttctcgttc
agcattgttattaacataaaatgcggaacaactccaaggtgacttagagggacgaattattttcttttct
aataattcattaatttccttttttacaatattccattaattcttggttcatttgaattggtctagccttag
tggggatagatccatcttcaaatccttctatatatggtaatcgtactatatgccttttttctttcccaaaa
ggcattaggtagatctgaacaaacttcagattctatttcttttttgggcttgtagtattttatcttgaatt
aatttatctttaagctggtcttctattttaacaattttaatttcatcttttatgtaatttatttgatttt
ttatcatttgtattttatagatacttctttctttttagagtatctatttctctttgttttggaccatttag
gaaataaaatattatttctttaccaagattattggtagtaatacctgttgtattaactgtaaaaggagat
atctgagttaaaaagggagttcctaatataacttcctcatttaaatcttttattagaataaaagtatttt
taaagcatatgttattattacatatatgtgcattagataatttatattttatttttaagtttttctcctga
tgcagagaataatttttgctgatgttttatcacaatattttgttggaatcaatccttctactatacaattt
tggtcagcaccagagtctagtaaagctactgtctggatactgtattcactatttatgactaattttattg
gtacataccatttttgagacttaattttagagatagccaagactcctaattcgtcgtcttgtgcttctac
ttgtttagtatttgaactttctcctattattgtagaatttaattccttttgagtttttttgaattaattct
agagcaaggttttcttcctctatattattaacttttttgttttaaaatattaacttcttgttttaatgttg
ctaattctgcataaagatctgatatagatatttttttcttctttttttatcaaatttttgtagaatattttt
catactaaaaggattggatatgttaggtttatttgattctctacttaagactatatctttttaattttaaa
agatattctctttttaatttcttcatcctttatttgatctattagatcaaagaagaattctttatttttcat
ttgatatgacttttattgaacctgtcccgcaggtacataatcctctaggacaatcacactctgattcgga
agtgtaactatcactagactcatatgctatattaatttgttcatcatcacttatgctttccgtggatgaa
ctttcttcttcactattatcattttcgtctagtatagcatatatttttattcttaatgtcttcatctaatt
ctaaactattaatcttgtttttcttctttgtcttgactctacaatcatttgccttatggccttttttttcc
acaattgtagcatctttctttaaaagaaggttgtttactttgacttgcatttttttggaattttcttcctg
tatcttttggaagattttcttttataattactattgctacttttgtcttttctatttatccttttggatc
tttttgttgcctcggtagacggggcagttattttagcataaccaaaatcactacagaaatttcctaattc
cttcttagtgattttataatctctttttaactgatttttttaatttcatatcagtacatagttctaaagct
gttatattaataaagcttactatatcaccataggttaatccgtcataaggtatttcaccgttaaatctat
cttttatttttttgtctaactttatctgcaaataaatgtggtagtccacttataaatttttctttccaata
actatgattacaatcttctctaatcattacttttatttaaaaatatgtctttataccatttaaagtcatgt
aattttggacatgataaattagataatatttctaaacttctatctttaaatacttttggttctccaataa
agtgtttaattatggcatataataaagtgtttatggcatcttcacgtacttgttcttctattatttgatt
agtactactaccttcttgtttaactatttttatgattgtagtagcattaaggacagtttctctatcttta
tccgacagatagttgtcccaccaatttttttaacattcctgtaaatccgtaagaaagcagggttgctgctt
gcttatctgtattacctttttatttttataggctgttgctgctattcccatttcatgtaatctattaagtat
ttgatgttctgcttgtccatctatattccattcagttatagcatcacctgagtgagaactgaatgatact
agatgttcttcttcatactgtatatcaggaaatgtaggtttttttataataatttttcttatgatgataat

tagatcctaatttatttaattgattatttaattcagataaggtttttaccgaattttctttttctttggg
aattgtatttatcgttaaagatgattttaacctttttggtaatttcatctagaacatcacctttttatttcc
aatccacctttaggaagactagggaatttatctttttgacatactaggtagcctgaaaattggattttctt
caatattatcttcaaatttttattctttttctctgggaatggtttaaaagaagacttgcttcttgcttcttc
tacctgattggctatgacatgtagtatttgatttgtataattactttgttctattaatttattaatatct
ttatttttaggttctccttgggcatctattgatgttctttttaaaaggatttcctattactactttatcac
ctactggtatatttatttcttctaatgggggaatggtagattctattattttattacttttcgtatacca
ttttttaactaccatatttaaagattcacagttagctaattttcctgtatattcaaaccatttaaaaaat
tctaaactacacttccatttttttcatatctttcatccatttatctctaaatgatttcctctgttctaaag
tgtaattgtttctataccattctctttttaatattattctctggttttacccagagttgttggagatattt
tatatctggttcaaatggttcttctaatgttatcattcctaaggaatcatttatttcactaaaggttggg
cttgatggttgattgattatgacaggcttgtattctggttgtggaatttttttactaaagttaacacctt
ttaaatttttgcttattatgtcttcttcctccaggggacaagaagtagatggccttgaaaaggacatcct
ggaagatgaagatgtagcattaaaggatctagattttaggaattttaaatctattgatccatctgtatgt
tgtataacttgttctatttgtatagattcatcttccacttttggttccgtgatggaatcaaagtgccatt
catcttttgataagacttcatcccatttaagtaatcttgggacaaatatattatgatgtttaatatttgc
ttccattaacattgtaactccttttatgagctctacttggtttagctttgggagctaaagtagttttcatt
aatttataatatactctatatattatagctatttctctggcttgatgtttactatttaaatgttttgttt
taacatttaatgttagagtatctaatattgttggatcattaatatcaactgagaaatttgggtaacaatt
aaagtatactggcccatttgccatgttactttgtaatacggctattaatgaatcatcaaagtttaaaaat
ctattatctcttaataataaacatataggaatatctaatcctagtctatgaagtggttttacagccactt
gcactagtcctatatgtaggaacttataacttttttctatatttttctatatcttttatagacaatagttg
aattgtttcaactgactgatttatagcaattgttgattcacatgttttaatattaaatctatttatgaag
tcaaatcctccgacttcatatatctttttctaacttttctttgggaatattccaatttaacagattatttt
ctatttttgatattttatctatttctgttgataggtttatactgtttaatactgactgcgatgatccttc
cattttaataatgggctcctggaccgtcacaagcttttctataataactttcagctataggttttaatct
aagtatttctgaaatattattttgcatataatgtagaaaatcaatatttcctgttaagagatatctatgt
agattatccttttaaaagcttttcatggtcattaatttctttatcaatatttatccagagttctagataac
ctggatccatattcatttttaaaatattatatttttattttttattaaggaaatctcttctccaaagttta
taatatagcagaagagtttcttctacgtaaggattatcgttgtttattaaatcagaaaaacatatattca
taaaatagatataataatcattatcattatatttattaaaccagtgtttcatataatacgggagttaagt
gtttttttgttttcttatgggggggttatcctatagttaactctactcttttctattaaggtgttttcttt
attcagatttccttcgttcttatcctatgt

Appendix 3.4 Approved sequence of *PaxiV*, the *Florendovirus* member in *P. axillaris* subsp *axillaris* N by Repbase database.

```
;ID PinfV DNA ; PLN ; 7142 BP
;XX
;DE consensus sequence of new Florendovirus in Petunia inflata.
;XX
;AC .
;XX
;DT 04-Jun-2018 (Rel. -1, Created)
;DT 04-Jun-2018 (Rel. -1, Last updated, Version 1)
;XX
;KW Caulimoviridae; Integrated Virus; PinfV.
;XX
;OS Petunia integrifolia subsP. integrifolia subsp inflata
;XX
;OC Petunia integrifolia subsP. integrifolia subsp inflata
;OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
```

```
;OC Spermatophyta; Magnoliophyta; eudicotyledons; Gunneridae;
;OC Pentapetalae; asterids; lamiids; Solanales; Solanaceae;
;OC Petunioideae; Petunia.
;XX
;RN [1] (bases 1 to 7142)
;RA Alisawi,O., Richert-Poggeler,K.,
;RA Heslop-Harrison,P. and Schwarzacher,T.
;RT New integrated Florendovirus in Petunia inflata.
;RL Direct Submission to RR (04-Jun-2018)
;XX
;CC New integrated Florendovirus in Petunia inflata.
;XX
;FT CDS 6760..1319
;FT /product="PinfV_1p"
;FT /translation="LKVIIEKLVTVQELIIKMEGSSQSVLNSINLSTEV
;FT DKISKIENNLLNWNIPKEKIEKIYEVGGFDFINRFNIKTCESTIAINQS
;FT VETIQLLSIKDIEKYRKSYKFLHIGLVQVAVKPLHRLGLDIPICLLLRD
;FT NRFLNFDDSLIAVLQSNMANGPVYFNCYPNFSVDINDPTILDVLTLNVK
;FT TKHLNSKHQTREIAIIYRVYYKLMKTTLAPKAKPSRAHKGVTMLMEANI
;FT KHHNIFVPRLLKWDEVLSKNEWHFDSITEPKVEDESIQIEQVIQHTDGS
;FT IDLKFLKSRSFNATSSSSRMSFSRPSTSCPLEEEDIISKNLKGVNFSKR
;FT IPQPEYKPVIINQPSSPTFSEINDSLGMITLEEPFEPDIKYLQQLWVKP
;FT ENNIKREWYKNNYTLEQRKSFRDKWMKDMKKWKCSLEFFKWFEYTGKLA
;FT NCESLNMVVKKWYTKSNKIIESTIPPLEEINIPVGDKVVIGNPFKRTSI
;FT DAQGEPKNKDINKLIEQSNYTNQILHVIANQVEEARSKSSFKPFPEKRI
;FT KFEDNIEKNPIFRLPSMSKDKFPSLPKGGLEIKGDVLDEITKRLKSSLT
;FT INTIPIEKENSVKTLSELNNQLNKLGSNYHHKKNYYKKPTFPDIQYEEE
;FT HLVSFSSHTGDAITEWNIDGQAEHQILNRLHEMGIAATAYKIKGNTDKQ
;FT AATLLSYGFTGMLKNWWDNYLSDRDRETVLNATTIIKVVKQESNSTNQT
;FT IEEQVREDAINTLLYAIIKHFIGEPKVFKDRSLEILSNLSCPKLHDFKW
;FT YKDIFLNKVMIREDCNHSYWKEKFISGLPQLFADKVRQKIKDRFNGEIP
;FT YDGLTYGDIISFINITALELCTDMKLKNQLKRDYKITKKELGNFCSDFG
;FT YAKITAPSTEATKKSKRINRKDKSSNSYYKRKSSKRYRKKIPKNTSQNK
;FT QPSFEERCYNCGKKGHKANNCRVKAKKKNKINSLELDEDIKNKIYAILD
;FT ENDNSEEGSSSTEDISDDEQINIAYESSDSYTSESECDCPKGLCTCGTG
;FT SIKVISNENKEFFFDLIDQIKDEEIKREYLLKLKDIVLSRESNKPNISN
;FT PFSMKNILQKFDKKEEKISISDLYAELTTLKQEVNILKQKVNKIEEENL
;FT ALELIQKTQKELNSTIIGESSNTKQVEAQDDELGVLAISKIKSQKWYVP
;FT IKLVINSEYSIQTVALLDSGADQNCIVEGLIPTKYCDKTSAKLFSASGE
;FT KLKIKYKLSNAHICNNNICFKNTFILIKDLNEEVILGTPFLTQISPFTV
;FT NTTGITTNNLGKEIIFYFLNGPKQREIDTLKERSIYKIQLIKNQINYIK
;FT EEIKIVKIENQLKDKLIQDKILQTQKEIESEVCSDLPNAFWERKRHIVR
;FT LPYIEGFEEESIPTKARPIQMNQELMEYCKKEINELLEKKIIRPSKSPW
;FT SCSAFYVNNNAERERGAPRLVINYKPLNKVLRWIRYPIPNKRDLLKRTY
;FT KAKIYSKFDMKSGFWQIQIDERDKYKTAFNVPFGQYEWNVMPFGLKNAP
;FT SEFQNIMNNIFNPYSSFSIVYIDDVLIFSENIDQHFKHLKIFLNIIQKN
;FT GLVVSAKKIKLFQTSIRFLGHDLSQGMYKPICRAIEFSTKFPDVLTDKT
;FT QLQRFLGSLNYVSDFIPNLRQICEPLYQRLRKNPVPWSQEQTQTVQKVK
;FT SIVQRLPCLGIPHPDAFMIVETDASEIGFGGILKQRLSPKESEQLVRYC
;FT SGIWNGAQKNYSTIKKEILSIVLCITKFQDDLINKEFLLRVDCKAAKDV
;FT LQKDVKNLVSKQIFARWQALLSSFDFQIEFIKGEDNSLPDFLTREFLQR
;FT KHEAIISSKTSEAIP"
;XX
;DR [1] (Consensus)
;XX
;SQ Sequence 7142 BP; 2018 A; 1103 C; 1006 G; 3015 T; 0 other;
PinfV
tattaacggggaatcttatcttttataaagtaagataagatttacattattgatgctaatgggatcctat
```

tgtcaggaatcttttctttacaaaagtaaagtaagatttacattattgataccacttggcatctatttgt
tgtctatcttttgctaaaaatccgtgagatacgcttgagcaatcccgtgtgggtcctccatatctccatt
atctcctgctgatgctgttgatgacgctgtatctaataattcagtgatttgatttatttcactaccattt
gcttggtccaaatttgccaaggcctgttgaagttttcttttaacgcttcctttgataatttgtctgttg
atccgatctttgtttgaacttgttgaacctgtttttctttgtttttaggggtccacccccttaatttttat
agttttaacaaggtacctgaccttgcataatacttctatattgaattcccaactaataatataggacact
cttttagccatgaagaatttacataactgtatatgttctggtaatgctgctagttcttcattttctttct
ggtaattggtgtgtaactgtttgaaagaaggtgacaacacccttttcgtctgctccaaatgtattccacca
ttcataaaaccatcttgggatgggatccgttgatacctttttcacagaacttgacaaaccatgaatggctg
ccaggtcttaagtatatgaagttaaaccatgcatttttatagtcgtaaaaattataacctgatattttaa
aatccactgataatgggaattgaatcggcgtgtgtagatgatctactggccaatccgcagggcttaaaat
tcttttaatagtaaacttggagtacctaatactccttggatcatttttatctgctaactcattttcaatt
tctattgaatttgttgctatcaatatagcttcataaaattctcttgttttataagggttgtctgtttcca
cataatttctgtctgtgtaacaagacttgataatatctgtaattcgtttttcaactaattttctctcaat
tgcaatgattggtacctgttgactgtttatccaatcgaatttctgttgacttgctgttgaagagcttgct
tgttcttgtttaatgggaggttgaggttgtactgcctcagaataactgtttatggttcttttggcataac
tgtttaaggattgttttgaggttcctgaggttgcttctctctttaccggtgtaggaggaggtaattgagg
aaactctgctaaaactgtatatctgtttatatgtgagttctggttatgatgctgttcaaggaattgcctc
ggaggtcttactggatataatggcctcatgttttctctgtaaaaattctctagttaaaaaatccggaaga
ctattgtcttctcctttaataaattctatttgaaaatcaaagctagataataaagcttgccatctggcaa
atatttgttttgaaactaaattcttaacatctttttgcaaaacatcttttgctgctttacaatcaactct
taaaagaaattctttgtttatcaaatcatcttgaaatttagttatacataaaactattgataaaatttct
tttttaatagtactataatttttttgagctccgttccatattccagagcaatatcggactaattgttcag
actcttttggagagagacgctgtttaagaattcctccgaaacctatttcagaggcatcggtttcaacaat
cataaaggcatccggatgtggtatgcctagacaaggaagtctttgaacaatacttttgacttttttggact
gtttgtgtttgttcttggctccaaggaactggatttttcctaagacgctgataaaggggttcacatattt
gtctaaggtttggaataaaatctgaaacataattgaggcttcctaaaaacctttgtaactgagttttatc
tgttaaaacatctggaaattttgtagagaattctatggctctacaaattggtttatacatgccttgagac
aagtcatgtcctaaaaatctaatagaggtttgaaataatttaatcttcttggcactcactaccagcccat
ttttctgaataatattaagaaatatttttaaatgtttaaaatgttgatctatattttctgaaaatattaa
tacatcatcaatataaactattgagaatgaactatatggattaaatatattattcattatattttgaaat
tctgatggggcatttttaacccaaagggcataacattccattcatattggccaaagggcacattaaatg
ctgttttatacttatccctttcatctatttggatttgccaaaatcctgatttcatatcaaatttactata
tattttagccttatacgttctttttaataaatctcttttatttggaattggatatcgtatccatcttaat
actttgtttaaaggtttataatttataactaatcttggagctcctctttctcgttcagcattattattaa
cataaaatgcggaacaactccaaggtgacttagaaggacgaattatttttcttttctaataattcattaat
ttccttttttacaatattccattaattcttggttcatttgaattggtctagccttagtggggatagactct
tcttcaaatccttctatatacggtaatcgtactatatgccttttttctttcccaaaaggcattaggtagat
ctgagcagacttcagattctatttctttttgggtttgtaatattttatcttgaattagtttatctttaag
ctggttttctattttaacaattttaatttcttctttttatgtaatttatttgatttttttatcaattgtatt
ttatagatacttctttctttttagagtatctatttctctttgttttggaccatttaggaaataaaatatta
tttctttaccaagattattggtagtaatacctgttgtattaactgtaaaaggagatatctgagttaaaaa
gggagttcctaatataacttcttcatttaaatcctttatcagaataaaagtatttttaaagcatatgtta
ttattacatatatgtgcattagataatttatattttatttttaagtttttctcctgatgcagagaataatt
ttgctgatgttttatcacaatattttgttggaattaatccttctactatacaattttggtcagcaccaga
gtctaataaagctactgtctggatactgtattcactatttatgactaattttattggtacataccatttt
tgagacttaattttagagatagccaagactcccaattcgtcgtcttgtgcttctacttgtttagtatttg
aactttctcctattattgtagaatttaactccttttgagtttttttgaattaattctagagcaaggttttc
ttcctctatttttattaacttttttgttttaaaatattaacttcttgttttaatgttgttaattctgcataa
agatctgatatagatatttttttcttctttttttatcaaatttttgtagaatatttttcatactaaaaggat
tggatatgttaggtttatttgattctctacttaagactatgtcttttaatttttaaaagatattctctttt
aatttcttcatcctttatttgatctattagatcaaagaagaactctttatttttcatttgatatgacttttt
attgaacctgtcccgcaggtacataatcctttaggacaatcacactctgattcagaagtgtaactatcac
tagactcatatgctatattaatttgttcatcatcacttatgtcttccgtggatgaacttccttcttcact
attatcattttcgtctagtatagcatatattttattcttaatgtcttcatctagttctaaactattaatt
ttgtttttctttttttgcttttactctacaattatttgccttatggcctttctttccacaattgtagcatc
tttcttcaaaagaaggttgcttattttgacttgtatttttttggaattttcttcctgtatctttttggaaga
ttttctttttataataactattgctacttttgtctttttctatttatccttttggatttttttgttgcctcg

gtagacggggcagttatcttagcataaccaaaatcactacagaaatttcctaattctttcttagtgattt
tataatctctttttagctgatttttttaatttcatatcagtacaaagttctaaagctgttatgttaataaa
gcttattatatcaccataggttaaaccatcataaggtatttcaccgttaaatctatcttttatttttttgt
ctaactttatctgcaaataactgtggtaatccacttataaatttttctttccaataactatgattacaat
cttctctaatcattactttatttaggaatatgtctttataccatttaaagtcatgtaattttggacatga
taaattagataatatttctaaacttctatctttaaatacttttggttctccaataaagtgtttaattata
gcatataataaagtatttatagcatcttcacgtacttgttcttctattgtttgattagtactattacttt
cttgtttaactacctttatgattgtagtagcattaaggacagtttctctatctctatctgacagatagtt
gtcccaccaatttttttaacattcctgtaaatccgtaagaaagcagggttgctgcttgcttatctgtatta
ccttttattttataggctgttgctgctattcccatttcatgtaatctattaagtatttgatgttctgctt
gtccatctatattccactcagttatagcatcacctgtgtgcgaactgaatgatactagatgttcttcttc
atactgtatatcaggaaatgtaggtttttataataatttttcttatgatgataattagatcctaattta
tttaattgattatttaattcagataaggtttttactgaattttctttctctattggaattgtatttattg
ttaaagatgattttaacctttttagtaatttcatctagaacatcacctttttatttccaatccacctttagg
aagactagggaatttatcttttgacatactaggtagcctgaaaattggattcttttcaatattatcttca
aattttattcttttctctgggaatggtttaaaagaagacttgcttcttgcttcttctacctgattagcta
tgacatgtaatatttgattcgtataattactttgttctattaatttattaatatctttatttttaggttc
tccttgggcatctattgatgttcttttaaaaggatttcctattactactttatcacctactggtatattt
atttcttctaatgggggaatggtagattctattattttattacttttcgtataccatttttttaactacca
tatttaaagactcacagttagctaattttcctgtatattcaaaccatttaaaaaattctaaactacactt
ccatttttttcatatctttcatccatttatctctaaatgatttcctctgttctaaagtgtaattgttttta
taccattctctttttaatattattctctggttttacccagagttgttggagatattttatatctggttcaa
atggttcttctaacgttatcattcctaaggaatcatttatttcactaaaggttgggcttgatggttgatt
gattatgacaggcttgtattctggttgtggaattcttttactaaagttaacacctttttaaatttttgctt
attatgtcttcttcctccaggggacaagaagtagacggccttgaaaaggacatcctggaagatgaagatg
tagcgttaaaggatctagattttagaaatttttaaatctattgatccatctgtatgttgtataacttgttc
tatttgtatagattcatcttccacttttggttccgtgatggaatcaaagtgccattcatttttttgataag
acttcatcccatttaagtaatcttgggacaaatatattatgatgtttaatatttgcttccattaacattg
taactccttttatgagctctacttggtttagctttgggagctaaagtagttttcattaatttataatatac
tctatatattatagctatttctctggtttgatgtttactattcaaatgttttgttttaacatttaatgtt
agaacatctaatattgttggatcattaatatcaactgagaaattcggataacaattaaagtatactggcc
catttgccatgttactttgtaatacggctattaatgaatcatcaaagtttaagaatctattatctcttaa
taataaacatataggaatatctaatcctagtctatgaagtggttttacagccacttgcactagtcctata
tgtaggaacttataacttttttctatatttttctatatcttttatagacaatagttgaattgtttcaactg
actgatttatagcaattgttgattcacatgttttaatattaaatctatttatgaagtcaaatcctccgac
ttcatatatctttttctatcttttctttgggaatattccaatttaacagattattttcaatttttgatatt
ttatctacttctgttgataggtttatactgtttaatactgactgtgatgatccttccattttaataatga
gctcctggaccgtcacaagcttttctataataactttcagctataggttttaacctaagtatttctgaaa
tattattttgcatataatgtagaaaatcaatatttcctgttaaaagatatctatgtaaattatccttttaa
aagcttttcatggtcactaatttctttatcagtatttatccagagttctagataacctggatccatattc
atttttaaaatatttctatttcttctaaggaaatctcttctccaaagtctataatataactgaagagttt
cttttacgtagggattatcactgtttactaaatcagtgtaacatgtattcatatagtagatataatagtt
attattattatatttattaaaccagtgtttcattaagtacgggagtttaagtgttttagttttcttatgg
gg

Appendix 3.5 Approved sequence of *PinfV*, the *Florendovirus* member in *P. integrifolia* subsp *inflata S6* by Repbase database.

```
;ID PhybV DNA ; PLN ; 7519 BP
;XX
;DE Consensus sequence of Florendovirus.
;XX
;AC .
;XX
;DT 04-Jun-2018 (Rel. -1, Created)
;DT 04-Jun-2018 (Rel. -1, Last updated, Version 1)
;XX
;KW Caulimoviridae; Integrated Virus; PhybV.
;XX
;OS Petunia x hybrida
;XX
;OC Petunia x hybrida
;OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
;OC Spermatophyta; Magnoliophyta; eudicotyledons; Gunneridae;
;OC Pentapetalae; asterids; lamiids; Solanales; Solanaceae;
;OC Petunioideae; Petunia.
;XX
;RN [1] (bases 1 to 7519)
;RA Alisawi,O., Richert-Poggeler,K.,
;RA Heslop-Harrison,P. and Schwarzacher,T.
;RT New integrated Florendovirus in Petunia hybrida.
;RL Direct Submission to RR (04-Jun-2018)
;XX
;CC Integrated Florendovirus in Petunia hybrida.
;XX
;FT CDS 6922..1481
;FT /product="PhybV_1p"
;FT /translation="LKVIIEKLVTVQEFIIKMEGSSQSVLNSINLSTEV
;FT DKISKIENNLLNWNIPKEKLEKIYEVGGFDFINRFNIKTCESTIAINQS
;FT VETIQLLSIKDIEKYRKSYKFLHIGLVQVAVKPLHRLGLDIPICLLLRD
;FT NRFLNFDDSLIAVLQSNMANGPVYFNCYPNFSVDINDPTILDVLTLNVK
;FT TKHLNSKHQAREIAIIYRVYYKLMKTTLAPKAKPSRAHKGVTMLMEANI
;FT KHHNIFVPRLLKWDEVLSKDEWHFDSITEPKVEDESIQIEQVIQHTDGS
;FT IDLKFLKSRSFNATSSSSRMSFSRPSTSCPLEEEDIISKNLKGVNFSKR
;FT IPQPEYKPVIINQPSSPTFSEINDSLGMITLEEPFEPDIKYLQQLWVKP
;FT ENNIKREWYRNNYTLEQRKSFRDKWMKDMKKWKCSLEFFKWFEYTGKLA
;FT NCESLNMVVKKWYTKSNKIIESTIPPLEEINIPVGDKVVIGNPFKRTSI
;FT DAQGEPKNKDINKLIEQSNYTNQILHVIANQVEEARSKSSFKSFPEKRI
;FT KFEDNIEKNPIFRLPSMSKDKFPSLPKGGLEIKGDVLDEITKRLKSSLT
;FT INTIPKEKENSVKTLSELNNQLNKLGSNYHHKKNYYKKPTFPDIQYEEE
;FT HLVSFSSHSGDAITEWNIDGQAEHQILNRLHEMGIAATAYKIKGNTDKQ
;FT AATLLSYGFTGMLKNWWDNYLSDRDRETVLNATTIIKIVKQESSSTNQI
;FT IEEQVREDAINTLLYAIIKHFIGEPKVFKDRSLEILSNLSCPKLHDFKW
;FT YKDIFLNKVMIREDCNHSYWKEKFISGLPHLFADKVRQKIKDRFNGEIP
;FT YDGLTYGDIVSFINITALELCTDMKLKNQLKRDYKITKKELGNFCSDFG
;FT YAKITAPSTEATKRSKRINRKDKSSNSYYKRKSSKRYRKKIPKNTSQSK
;FT QPSFEERCYNCGKKGHKANDCRVKTKKKNKINSLELDEDIKNKIYAILD
;FT ENDNSEEESSSTESISDDEQINIAYESSDSYTSESECDCPRGLCTCGTG
;FT SIKVISNENKEFFFDLIDQIKDEEIKREYLLKLKDIVLSRESNKPNISN
;FT PFSMKNILQKFDKKEEKISISDLYAELTTLKQEVNILKQKVNKIEEENL
;FT ALELIQKTQKELNSTIIGESSNTKQVEAQDDELGVLAISKIKSQKWYVP
;FT IKLVINSEYSIQTVALLDSGADQNCIVEGLIPTKYCDKTSAKLFSASGE
;FT KLKIKYKLSNAHICNNNICFKNTFILIKDLNEEVILGTPFLTQISPFTV
;FT NTTGITTNNLGKEIIFYFLNGPKQREIDTLKERSIYKIQLIKNQINYIK
;FT DEIKIVKIENQLKDKLIQDKILQTQKEIESEVCSDLPNAFWERKRHIVR
;FT LPYIEGFEEGSIPTKARPIQMNQELMEYCKKEINELLEKKIIRPSKSPW
```

;FT SCSAFYVNNNAERERGAPRLVINYKPLNKVLRWIRYPIPNKRDLLKRTY
;FT KAKIYSKFDMKSGFWQIQIDERDKYKTAFNVPFGQYEWNVMPFGLKNAP
;FT SEFQNIMNNIFNPYSSFSIVYIDDVLIFSENIDQHFKHLKIFLNIIKKN
;FT GLVVSAKKIKLFQTSIRFLGHDLSQGMYKPICRAIEFSTKFPDVLTDKT
;FT QLQRFLGSLNYVSDFIPNLRQICEPLYQRLRKNPIPWSQEQTQTVQKVK
;FT SIVQRLPCLGIPHPDAFMIVETDASEIGFGGILKQRLSPKESEQLVRYC
;FT SGIWNGTQKNYSTIKKEILSIVLCITKFQDDLINKEFLLRVDCKAAKDV
;FT LQKDVKNLVSKQIFARWQALLSSFDFQIEFIKGEDNSLPDFLTREFLQR
;FT RHEAIISSKTSEAIP"
;XX
;DR [1] (Consensus)
;XX
;SQ Sequence 7519 BP; 2137 A; 1167 C; 1080 G; 3135 T; 0 other;
PhybV
aataaaattggaactgtttattgaaaaagaaggaacattgttatagacagagggaaaggagggaatattt
ctacaatgaggatttcgagtgatgccttacttctgatttctacttctatttatagctgctaatgacacta
ttctcctttacaggtggagagctattaacggggaatcttatcttttataaagtaagataagatttacatt
attgatgctaatgggatcctattgtcaagaatcttatctttacaaaagtaaagtaagatctacattattg
ataccacttggcacctatttgttgtctatcttttgctaaaaatccgtgagatatgcttgagcaatcccgt
gtgggtcctccatatctccattatctcctgctgatgctgttgatgatgctgtatctaataattcagtgat
ttgatttatttcactgccatttgcttggtctaaatttgccaaggcctgttgaagttttctttttaacgct
tcctttgataatttgtctgttgatccgatcttcgtttgaacttgttgaacctgtttttctttgttcttag
gggtccacccctttaatttttatagttttaacaaggtacctgaccttgcataatacttctatattgaattc
ccaactaataatataggacactcttttagccatgaagaatttacataactgtatatgttctggtaatgct
gctagttcttcgttttctttctggtaattggtgtgtaactgtttgaaagaaggtgacaacacccttttcgt
ctgctccaaatgtattccaccattcataaaaccatcttgggatgggatccgttgataccttttcacagaa
cttgacaaaccatgaatggctgccaggtcttaagtatatgaagttaaaccacgcatttttttagtcgtaa
aaattataacctgatattttaaaattcactgataatggaaattgaatcggcgtgtgtagatgatctactg
gccaatccgcagggcttaaaattcttttaatagtaaacttggagtacctaatactccttggatcatttttt
atctgctaactcattttcaatttctattgaattggttgctatcaatatagcttcataaaattctcttgtt
ttataagggttgtctgtttccacataatttctgtctgtgtaacaagacttgataatatctgtaattcgtt
tttcaactaattttctctcaatcgcaatgattggtacctgttgactgtttatccaatcgaatttctgttg
acttgctgttgaagagcttgcttgttcttgtttaatgggaggttgaggttgtactgcctcagaataactg
tttatggttcttttggcataactgtttaaggattgttttgaggttcctggggttgcttctctctttaccg
gtgtaggaggaggtaattgaggaaactctgctaaagctgtatatctgtttatatgtgagttctggttatg
atgctgttcaaggaattgcctcggaggtcttactggatataatggcctcatgtcttctctgtaaaaattc
tctagttaaaaaatctggaagactattgtcttctcctttaataaattctatttgaaaatcaaagctagat
aataaagcttgccatctggcaaatatttgttttgaaactaaattcttaacatctttttgcaaaacatctt
ttgctgctttacaatcaactcttaaaagaaattctttgtttatcaaatcatcttgaaatttagttataca
taaaactattgataaaatttctttttttaatagtactataatttttttgagttccgttccatatgccagag
caatatcggactaattgttcagactcttttggagagagacgctgtttaagaattcctccgaaacctattt
cagaggcatcggtttcaacaatcataaaggcatccggatgtggtatgcctagacaaggaagtctttgaac
aatacttttgacttttttggactgtttgtgtttgttcttggctccaaggaattggatttttcctaagacgc
tgataaaggggttcacatatttgtctaagatttggaataaaatctgaaacataattgaggcttcctaaaa
acctttgtaactgagttttatctgttaaaacatctggaaattttgtagagaattctatggctctacaaat
tggtttatacatgccttgagacaagtcatgtcctaaaaatctaatagaggtttgaaatagtttaatcttt
ttggcactcactaccagcccatttttcttaataatattaagaaatatctttaaatgcttaaaatgttgat
ctatattttctgaaaatattaatacatcatcaatataaactattgagaatgaactatatggattaaatat
attattcattatattttgaaattctgatggggcatttttaacccaaagggcataacattccattcatat
tggccaaagggcacattaaatgctgttttatacttatccctttcatctatttggatttgccaaaatcctg
atttcatatcaaatttactatatatttagccttatacgttctttttaataaatctcttttatttggaat
tggatatcgtatccatcttaatactttgtttaaaggtttataatttataactaatcttggagctcctctt
tctcgttcagcattattattaacataaaatgcggaacaactccaaggtgacttagaaggacgaattattt
tcttttctaataattcattaatttccttttacaatattccattaattcttggttcatttgaattggtct
agccttagtggggatagatccttcttcaaatccttctatatatggtaatcgtactatatgccttttttctt
tcccaaaaggcattaggtagatctgaacagacttcagattctatttctttttgggtttgtagtattttat
cttgaattagtttatctttaagctggttttctattttaacaattttaatttcatcttttatgtaatttat
ttgatttttatcaattgtattttatagatacttctttcttttagagtatctatttctctttgttttgga

ccatttaggaaataaaatattatttctttaccaagattattggtagtaatacctgttgtattaactgtaa
aaggagatatctgagttaaaaagggagttcctaatataacttcttcatttaaatcttttatcagaataaa
agtatttttaaagcatatgttattattacatatatgtgcattagataatttatattttatttttaagtttt
tctcctgatgcagagaataattttgctgatgttttatcacaatattttgttggaattaatccttctacta
tacaattttggtcagcaccagagtctaataaagctactgtctggatactgtattcactatttatgactaa
ttttattggtacataccatttttgagacttaattttagagatagccaagactcccaattcgtcgtcttgt
gcttctacttgtttagtatttgaactttctcctattattgtagaatttaactccttttgagtttttgaa
ttaattctagagcaaggttttcttcctctattttattaacttttgttttaaaatattaacttcttgttt
taatgttgttaattctgcataaagatctgatatagatatttttttcttctttttttatcaaatttttgtaga
atatttttcatactaaaaggattggatatgttaggtttatttgattctctacttaagactatatctttta
attttaaaagatattctcttttaatttcttcatcctttatttgatctattagatcaaagaagaattcttt
attttcatttgatatgacttttattgaacctgtcccgcaggtacataatcctctaggacaatcacactct
gattcggaagtgtaactatcactagactcatatgctatattaatttgttcatcatcacttatgctttccg
tggatgaacttttcttcttcactattatcattttcgtctagtatagcatatattttattcttaatgtcttc
atctaattctaaactattaattttgtttttctttttttgtcttgactctacaatcatttgccttatggcct
ttttttccacaattgtagcatctttcttcaaaagaaggttgtttactttgacttgtatttttttggaattt
tcttcctgtatcttttggaagattttcttttataataactattgctacttttgtcttttctatttatcct
tttggatctttttgttgcctcggtagacggggcagttatcttagcataaccaaaatcactacagaaattt
cctaattctttcttagtgatcttgtaatctctttttaactgatttttttaatttcatatcagtacatagtt
ctaaagctgttatgttaataaagcttactatatcaccataggttaaaccatcataaggtatttcaccgtt
aaatctatcttttatttttttgtctaactttatctgcaaataaatgtggtaatccacttataaatttttct
ttccaataactatgattacaatcttctctaatcattactttatttagaaatatgtctttataccatttaa
agtcatgtaattttggacatgataaattagataatatttctaaacttctatctttaaatacttttggttc
tccaataaagtgtttaattatagcatataataaagtatttatagcatcttcacgtacttgttcttctatt
atttgattagtactactactttcttgtttaactatttttatgattgtagtagcattaaggacagtttctc
tatctctatctgacagatagttgtcccaccaatttttttaacattcctgtaaatccgtaagaaagcagggt
tgctgcttgcttatctgtattacctttttatttttataggctgttgctgctattcccatttcatgtaatcta
ttaagtatttgatgttctgcttgtccatctatattccattcagttatagcatcacctgagtgcgaactga
atgatactagatgttcttcttcatactgtatatcaggaaatgtaggtttttttataataatttttcttatg
atgataattagatcctaatttatttaattgattatttaattcagataaggtttttactgaattttctttt
tcttttggaattgtatttatcgttaaagatgattttaacctttttagtaatttcatctagaacatcacctt
ttatttccaatccacctttaggaagactagggaatttatcttttgacatactaggtagcctgaaaattgg
atttttttcaatattatcttcaaattttattcttttctctgggaatgatttaaaagaagacttgcttctt
gcttcttctacttgattagctatgacatgtagtatttgatttgtataattactttgttctattaatttat
taatatctttatttttaggttctccttgggcatctattgatgttcttttaaaaggatttcctattactac
tttatcacctactggtatatttatttcttctaatgggggaatggtagattctattatttttattacttttc
gtataccatttttttaactaccatatttaaagattcacagttagctaattttcctgtatattcaaaccatt
taaaaaattctaaactacacttccatttttttcatatctttcatccatttatctctaaatgatttcctctg
ttctaaagtgtaattgtttctataccattctctttttaatattattctctggttttacccagagttgttgg
agatattttatatctggttcaaatggttcttctaacgttatcattcctaaggaatcatttatttcactaa
aggttgggcttgatggttgattgattatgacaggcttgtattctggttgtggaattctttttactaaagtt
aacacctttttaaatttttgcttattatgtcttcttcctccaggggacaagaagtagatggccttgaaaag
gacatcctggaagatgaagatgtagcgttaaaggatctagatttttaggaattttaaatctattgatccat
ctgtatgttgtataacttgttctatttgtatagattcatcttccacttttggttccgtgatggaatcaaa
gtgccattcatcttttgataagacttcatcccatttaagtaatcttgggacaaatatattatgatgttta
atatttgcttccattaacattgtaactccttttatgagctctacttggtttagctttgggagctaaagtag
ttttcattaatttataatatactctatatattatagctatttctctggcttgatgtttactatttaaatg
ttttgttttaacatttaatgttagaacatctaatattgttggatcattaatatcaactgagaaattcgga
taacaattaaagtatactggcccatttgccatgttactttgtaatacggctattaatgaatcatcaaagt
ttaagaatctattatctcttaataataaacatataggaatatctaatcctagtctatgaagtggttttac
agccacttgcactagtcctatatgtaggaacttataacttttctatatttttctatatcttttatagac
aatagttgaattgtttcaactgactgatttatagcaattgttgattcacatgttttaatattaaatctat
ttatgaagtcaaatcctccgacttcatatatttttttctaacttttctttgggaatattccaatttaacag
attattttctatttttgatattttatctacttctgttgataggtttatactgtttaatactgactgtgat
gatccttccatttttaataatgaactcctggaccgtcacaagcttttctataataactttcagctataggt
tttaacctaagtatttctgaaatattatttttgcatataatgtagaaaatcaatatttcctgttaaaagat
atctatgtaaattatccttttaaaagcttttcatggtcactaatttctttatcaatatttatccagagttc
tagataacctggatccatattcatttttaaaatattatatttttatttttattaaggaaatctcttctcc

```
aaagtttataatataacagaagagtttcttctacgtaaggattatcgttgtttattaaatcagagaaaca
tatattcataaaatagatataatattcattatcattatatttattaaaccagtgtttcataaaatacggg
agttaagtgttttttgttttcttatgggggggttatcctatagttaactctactcttttctattaaggtg
ttttctttattcagaatttccttcgttcttatcctatgtagctaagcgacgggaacaccaatccgggtcc
ggccctacgtcctagactcaacgggctaatccaacagaactttatggcttgccaacgtcgaaacgggggga
ttaatcggttacctcgacacaaggataag
```

Appendix 3.6 Approved sequence of *PhybV*, the *Florendovirus* member in *P. hybrida* by Repbase database.

```
;ID PparV DNA ; PLN ; 6519 BP
;XX
;DE Integrated Florendovirus in P. axillaris subsp parodii.
;XX
;AC .
;XX
;DT 05-Apr-2018 (Rel. -1, Created)
;DT 05-Apr-2018 (Rel. -1, Last updated, Version 1)
;XX
;KW Caulimoviridae; Integrated Virus; PparV.
;XX
;OS Petunia axillaris subsP. axillaris subsp parodii
;XX
;OC Petunia axillaris subsP. axillaris subsp parodii
;OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
;OC Spermatophyta; Magnoliophyta; eudicotyledons; Gunneridae;
;OC Pentapetalae; asterids; lamiids; Solanales; Solanaceae;
;OC Petunioideae; Petunia.
;XX
;RN [1] (bases 1 to 6519)
;RA Alisawi,O., Richert-Poggeler,K.,
;RA Schwarzacher,T. and Heslop-Harrison,P.
;RT Integrated Florendovirus within P. axillaris subsp parodii S7
genome.
;RL Direct Submission to RR (05-Apr-2018)
;XX
;CC Consensus sequence.
;XX
;FT CDS 6432..991
;FT /product="PparV_1p"
;FT /translation="LKVIIEKLVTVQEPIIKMEGSSQSVLNSINLSTEV
;FT DKISKIENNLLNWNIPKEKLEKIYEVGGFDFINRFNIKTCESTIAINQS
;FT VETIQLLSIKDIEKYRKSYKFLHIGLVQVAVKPLHRLGLDIPICLLLRD
;FT NRFLNFDDSLIAVLQSNMANGPVYFNCYPNFSVDINDPTILDVLTLNVK
;FT TKHLNSKHQAREIAIIYRVYYKLMKTTLAPKAKPSRAHKGVTMLMEANI
;FT RHHNIFVPRLLKWDEVLSKDEWHFDSITEPKVEDESIQIEQVIQHTDGS
;FT IDLKFLKSRSSNATSSSSRMSFSRPSTSCPLEEEDIISKNLKGVNFSKR
;FT IPQPEYKPVIINQPSSPTFSEINDSLGMITLEEPFEPDIKYLQQLWVKP
;FT ENNIKREWYRNNYTLEQRKSFRDKWMKDMKKWKCSLEFFKWFEYTGKLA
;FT NCESLNMVVKKWYTKSNKIIESTIPPLEEINIPVGDKVVIGNPFKRTSI
;FT DAQGEPKNRDINKLIEQSNYTNQILHVIANQVEEARSKSSFKPFPEKRI
;FT KFEDNIEENPIFRLPSMSKDKFPSLPKGGLEIKGDVLDEITKRLKSSLT
;FT INTIPKEKENSVKTLSELNNQLNKLGSNYHHKKNYYKKPTFPDIQYEEE
```

;FT HLVSFSSHSGDAITEWNIDGQAEHQILNRLHEMGIAATAYKIKGNTDKQ
;FT AATLLSYGFTGMLKNWWDNYLSDRDRETVLNATTIIKVVKQESNSTNQT
;FT IEEQVREDAINTLLYAIIKHFIGEPKVFKDRSLEILSNLSCPKLHDFKW
;FT YKDIFLNKVMIREDCNHSYWKEKFISGLPQLFADKVRQKIKDRFNGEIP
;FT YDGLTYGDIISFINITALELCTDMKLKNQLKRDYKITKKELGNFCSDFG
;FT YAKITAPSTEATKKSKRKNRKDKSSNSDYKRKYSKRYRKKIPKNTSQSK
;FT QPSFKERCYNCGKKGHKANDCRVKTKKKNKINSLELDEDIKNKIYAILD
;FT ENDNSEEESSSTESISDDEQINIAYESSDSYTSESECDCPRGLCTCGTG
;FT SIKVISNENKEFFFDLIDQIKDEEIKREYLLKLKEIVLSRESNKPNISN
;FT PFSMKKILQKFDKKEEKISISDLYAELTTLKQEVNILKQKVNNIEEENL
;FT ALELIQKTQKELNSTIIGESSNTKQVEAQDDELGVLAISKIKSQKWYVP
;FT IKLVINSEYSIQTVALLDSGADQNCIVEGLIPTKYCDKTSAKLFSASGE
;FT KLKIKYKLSNAHICNDNICFKNTFILIKDLNEEVILGTPFLTQISPFTV
;FT NTTGITTNNLGKEIIFYFLNGPKQREIDTLKERSIYKIQLIKNQINYIK
;FT DEIKIVKIENQFKDKLIQDKILQTQKEIESEVCSDLPNAFWERKRHIVR
;FT LPYIEGFEEGSIPTKARPIQMNQELMEYCKKEINELLEKKIIRPSKSPW
;FT SCSAFYVNNNAERERGAPRLVINYKPLNKVLRWIRYPIPNKRDLLKRTY
;FT KAKIYSKFDMKSGFWQIQIDERDKYKTAFNVPFGQYEWNVMPFGLKNAP
;FT SEFQNIMNNIFNPYSSFSIVYIDDVLIFSENIDQHFKHLKIFLNIIKKN
;FT GLVVSAKKIKLFQTSIRFLGHDLSQGMYKPICRAIEFSTKFPDVLTDKT
;FT QLQRFLGSLNYVSDFIPNLRQICEPLYQRLRKNPVPWSQEQTQTVQKVK
;FT SIVQRLPCLGIPHPDAFMIVETDASEIGFGGILKQRLSPKESEQLVRYC
;FT SGIWNGAQKNYSTIKKEILSIVLCITKFQDDLINKEFLLRVDCKAAKDV
;FT LQKDVKNLVSKQIFARWQALLSSFDFQIEFIKGEDNSLPDFLTREFLQR
;FT RHEAIISSKTSEAIP"
;XX
;DR [1] (Consensus)
;XX
;SQ Sequence 6519 BP; 1830 A; 1013 C; 921 G; 2755 T; 0 other;
PparV
tcctttgataatttgactgttgatccgatattcgtttgaacttgctgaacctgtttttctttgttcttag
gggtccatcccttaatttttatagttttaacaaggtacctgaccttgcataatacttctatattgaattc
ccaactaataatataggacactcttttagccatgaagaatttacataactgtatatgttctggtaatgct
gctagttcttcattttctttctggtaattggtgtgtagctgtttgaaagaaggtgacaacacccttttcgt
ctgctccaaatgtattccaccattcataaaaccatcttgggatgggatccgttgataccttttcacagaa
cttgacaaaccatgaatggctgccaggtcttaagtatatgaagttaaaccacgcatttttgtagtcgtag
aaattataacctgatattttaaaattcactgataatggaaattgaatcggcgtgtgtagatgatctactg
gccaatccgcagggcttaaaattcttttaatagtaaacttggagtacctaatactccttggatcatttttt
atctgctaactcattttcaatctctattgaattggttgctatcaatatagcttcataaaattctcttgtt
ttataagggttgtctgtttccacataatttctgtctgtgtaacaagacttgataatatctgtaattcgtt
tttcaactaattttctctcaatcgcaatgattggtacctgttgactgtttatccaatcgaatttctgttg
acttgctgttgaagagcttgcttgttcttgtttaatgggaggttgaggttgtactgcctcagaataactg
tttatggttcttttggcataactgtttaaggattgttttgaggttcctggggttgcttctctctttaccg
gtgtaggaggaggtaattgaggaaactctgctaaagccgtatatctgtttatatgtgagttctggttatg
atgctgttcaaggaattgcctcggaggttttactggatataatggcctcatgtcttctctgtaaaaattc
tctagttaaaaaatctggaagactattgtcttctcctttaataaattctatttgaaaatcaaagctagat
aataaagcttgccatctggcaaatatttgttttgaaactaaattcttaacatctttttgcaaaacatctt
ttgctgctttacaatcaactcttaaaagaaattctttgtttatcaaatcatcttgaaatttagttataca
taaaactattgataaaatttctttttttaatagtactataatttttttgagctccgttccatattccagag
caatatcggactaattgttcagactcttttggagagagacgctgtttaagaattcctccgaaacctattt
cagaggcatcggtttcaacaatcataaaggcatccggatgtggtatgcctagacaaggaagtctttgaac
aatacttttgacttttttggactgtttgtgtttgttcttggctccaaggaactggatttttcctaagacgc
tgataaaggggttcacatatttgtctaagatttggaataaaatctgaaacataattgaggcttcctagaa
acctttgtaactgagttttatctgttaaaacatctggaaatttttgtagagaattctatggctctacaaat
tggtttatacatgccttgagacaagtcatgtcctaaaaatctaatagaggtttgaaataatttaatcttc
ttggcactcactaccagcccatttttcttaataatattaagaaatatctttaaatgcttaaaatgttgat
ctatattttctgaaaatattaatacatcatcaatataaactattgagaatgaactatatggattaaatat
attattcattatattttgaaattctgatggggcattctttaacccaaagggcataacattccattcatat

tggccaaagggcacattaaatgctgttttatacttatccctttcatctatttggatttgccaaaatcctg
atttcatatcaaatttactatatattttagccttatacgttctttttaataaatctcttttatttggaat
tggatatcgtatccatcttaatactttgtttaaaggtttataatttataactaatcttggagctcctctt
tctcgttcagcattattattaacataaaatgcggaacaactccaaggtgacttagaaggacgaattattt
tcttttctaataattcattaatttccttttttacaatattccattaattcttggttcatttgaattggtct
agccttagtggggatagatccttcttcaaatccttctatatacggtaatcgtactatatgccttttttctt
tcccaaaaggcattaggtagatctgaacagacttcagattctatttctttttgggtttgtagtattttat
cttgaattagtttatctttaaactggttttctattttaacaattttaatttcgtcttttatgtaatttat
ttgatttttttatcaattgtattttatagatacttctttctttttagagtatctatttctctttgctttgga
ccatttaggaaataaaatattatttctttaccaagattattggtagtaatacctgttgtattaactgtaa
aaggagatatctgagttaaaaagggagttcctaatataacttcttcatttaaatccttttatcagaataaa
agtatttttaaagcatatgttatcattacatatatgtgcattagataatttatattttatttttaagtttt
tctcctgatgcagagaataattttgctgatgttttatcgcaatattttgttggaattaatccttctacta
tacaattttggtcagcaccagagtctaataaagctactgtctggatactgtattcactatttatgactaa
ttttattggtacataccatttttgagacttaattttagagatagccaagactcccaattcgtcgtcttgt
gcttctacttgtttagtatttgaactttctcctattattgtagaatttaactccttttgagtttttgaa
ttaattctagagcaaggttttcttcctctatgttattaacttttttgttttaaaatattaacttcttgttt
taatgttgttaattctgcataaagatctgatatagatatttttttcttctttttttatcaaatttttgtaga
atttttttcatactaaaaggattggatatgttaggtttatttgattctctacttaagactatttctttta
attttaaaagatattctcttttaatttcttcatcctttatttgatctattagatcaaagaagaattcttt
attttcatttgatatgacttttattgaacctgtcccgcaggtacataatcctctaggacaatcacactct
gattcggaagtgtaactatcactagactcatatgctatattaatttgttcatcatcacttatgctttccg
tggatgaacttttcttcttcactattatcattttcgtctagtatagcatatattttattcttaatgtcttc
atctaattctaaactattaattttgtttttcttctttgtcttgactctacaatcatttgccttatggcct
ttttttccacaattgtagcatctttctttaaaagaaggttgtttactttgacttgtattctttggaattt
tcttcctgtatcttttggaatattttcttttataatcactattgctacttttgtcttttctatttttcct
tttggatttttttgttgcctcggtagacggggcagttatcttagcataaccaaaatcactacagaaattt
cctaattctttcttagtgattttataatctctttttagctgatttttttaatttcatatcagtacatagtt
ctaaagctgttatgttaataaagcttattatatcaccataggttaaaccatcataaggtatttcaccgtt
aaatctatcttttatttttttgtctaactttatctgcaaataactgtggtaatccacttataaatttttct
ttccaataactatgattacaatcttctctaatcattactttatttaggaatatgtctttataccatttaa
agtcatgtaattttggacatgataaattagataatatttctaaacttctatctttaaatacttttggttc
tccaataaagtgtttaattatagcatataataaagtatttatagcatcttcacgtacttgttcttctatt
gtttgattagtactattactttcttgtttaactacctttatgattgtagtagcattaaggacagtttctc
tatctctatctgacagatagttgtcccaccaatttttttaacattcctgtaaatccgtaagaaagcagggt
tgctgcttgcttatctgtattacctttttatttttataggctgttgctgctattcccatttcatgtaatcta
tttagtatttgatgttctgcttgtccatctatattccactcagttatagcatcacctgagtgagaactga
atgatactagatgttcttcttcatactgtatatcaggaaatgtaggtttttttataataatttttcttatg
atgataattagatcctaatttatttaattgattatttaattcagataaggtttttactgaattttctttt
tcttttggaattgtatttatcgttaaagatgattttaacctttttagtaatttcatctagaacatcacctt
ttatttctaatccacctttaggaagactagggaatttatcttttgacatactaggtagcctgaaaattgg
attttcttcaatattatcttcaaattttattcttttctctgggaatggtttaaaagaagacttgcttctt
gcttcttctacctgattggctatgacatgtagtatttgatttgtataattactttgttctattaatttat
taatatctctatttttaggttctccttgggcatctattgatgttctttttaaaaggatttcctattactac
tttatcacctactggtatatttatttcttctaatgggggaatggtagattctattattttattacttttc
gtataccatttttttaactaccatatttaaagattcacagttagctaattttcctgtatattcaaaccatt
taaaaaattctaaactacacttccatttttttcatatctttcatccatttatctctaaatgatttcctctg
ttctaaagtgtaattgtttctataccattctctttttaatattattctctggttttacccagagttgttgg
agatattttatatctggttcaaatggttcttctaacgttatcattcctaaggaatcatttatttcactaa
aggttgggcttgatggttgattgattatgacaggcttgtattctggttgtggaattcttttactaaagtt
aacacctttttaaatttttgcttattatgtcttcttcctccaggggacaagaagtagatggccttgaaaag
gacatcctggaagatgaagatgtagcattagaggatctagattttaggaattttaaatctattgatccat
ctgtatgttgtataacttgttctatttgtatagattcatcttccacttttggttccgtgatggaatcaaa
gtgccattcatcttttgataagacttcatcccatttaagtaatcttgggacaaatatattatgatgtcta
atatttgcttccattaacattgtaactccttttatgagctctacttggtttagctttgggagctagagtag
ttttcattaatttataatatactctatatattatagctatttctctggcttgatgtttactatttagatg
ttttgttttaacatttaatgttagaacatctaatattgttggatcattaatatcaactgagaaattcgga
taacaattaaagtatactggcccatttgccatgttactttgtaatacggctattaatgaatcatcaaagt

```
ttaagaatctattatctcttaataataaacatataggaatatctaatcctagtctatgaagtggttttac
agccacttgcactagtcctatatgtaggaacttataacttttctatatttttctatatcttttatagac
aatagttgaattgtttcaactgactgatttatagcaattgttgattcacatgttttaatattaaatctat
ttatgaagtcaaatcctccgacttcatatatcttttctaacttttctttgggaatattccaatttaacag
attattttctatttttgatattttatctacttctgttgataggtttatactgtttaatactgactgtgat
gatccttccattttaataatgggctcctggaccgtcacaagcttttctataataactttcagctataggt
tttaatctaagtatttctgaaatattattttgcatataatgtaaaaaatcaatatttcctgttgaaagat
atctatgta
```

Appendix 3.7 Approved sequence of *PparV*, the *Florendovirus* member in *P. axillaris* subsp *parodii* S7 by the Repbase database.

```
;ID Caulimovirus-PAx DNA ; PLN ; 7722 BP
;XX
;DE Integrated caulimovirus sequence.
;XX
;AC .
;XX
;DT 07-Aug-2017 (Rel. -1, Created)
;DT 07-Aug-2017 (Rel. -1, Last updated, Version 1)
;XX
;KW Caulimoviridae; Integrated Virus; Caulimovirus-PAx.
;XX
;OS Petunia axillaris
;XX
;OC Petunia axillaris
;OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
;OC Spermatophyta; Magnoliophyta; eudicotyledons; Gunneridae;
;OC Pentapetalae; asterids; lamiids; Solanales; Solanaceae;
;OC Petunioideae; Petunia.
;XX
;RN [1] (bases 1 to 7722)
;RA Alisawi,O., Richert-Poggeler,K.,
;RA Heslop-Harrison,P. and Schwarzacher,T.
;RT New integrated viral sequence within Petunia axillaris genome.
;RL Direct Submission to RR (07-Aug-2017)
;XX
;CC 100% identical to consensus.
;CC ~ 61.1% identical to Caulimovirus-2_STu.
;CC ~ 55.8% identical to Caulimovirus-3_STu.
;CC ~ 61.4% identical to Caulimovirus-4_STu.
;XX
;FT CDS 1072..5505
;FT /product="Caulimovirus-PAx_1p"
;FT /translation="KYLMDFTSTIDFSRGSPARKKISEDFYKPRWELFR
;FT RWFFKSFNRTEQNDFQEEFYKDLEKVNKIISFIPWFMGKYVANYIAVLE
;FT RQFELLNGKIIDAVLPPQQPFKIKKEGKMMDFAAFSAMIEDNTLQVTVK
;FT HINIMLKQHNYTNIYVHILGEHIVSLHDKVDKLCSIISSNTASTSGKEK
;FT LEKPPRVAIPTIQPPPDVEDFKLKPDAYNIEKFLNEKFKELSLKPLINT
;FT EESDNAFSENEVFEDDFPKLTYDQINKLKSSYSSKFADKPRQRMFYYPR
;FT PTPQDVLFEEQDDMYAPNSFSGKQIYEWNIDGLTNRQMYVVMHRMMMYS
;FT TICKTNGNSDKTVANMITSGFTGQLKGWWDNYLTVDQRNEILNAVKSEP
```

;FT GTSTADNTNAVYFLVVNIIEHFSGRWSDNSETIRTLLQNMRCKTLTSFR
;FT LYKDTFLCRVMELPECNSTHWKSKFIDGLPNLFAERVKKVLRQEGVNIN
;FT YDDHTYGSLVGICIQEGLALCNDIKLNQELKRHRLNERQQLGEFCGQFG
;FT MDDPRTRTKSKSHKLHRSKTGKKHDYQEWRQRYSERRTEKRARRKKAFK
;FT ERKDDIKAKAKNPLNCWKCGRFGHRYKDCKVKERIKALNIDEDLKDSLL
;FT KILLNSEPEISDPGEENIPTKEELNKVFQDESDIVTSSEDECDPCKIGQ
;FT PCDNKENDEFYNLFSQFSLNTLDNDNLIELLKMVQDSKVRSKLIDQIKG
;FT NFEQKESEGLEKGFYAPMDNTPYTMKEVKRLLEQRKMSHSGSPTTSQDL
;FT EKEVYNLKKEIYNLKKENKILDQRISRIEGKGKEKDETTTPSNSENLDN
;FT IFSDMNEEFLTTLEVFTSQKFFIKITLLIDYNFKKEFIALVDSGADLNC
;FT VQEGLIPSRYFHKTTHSLRSANGEKMQIEYKLPTASICTNKVCIPQSFV
;FT LIKNMSNQVILGTPFLQKIFPITRIDKEGIIGTFQGKPIIFQFITEPIT
;FT RVLNQLKDMLAKKSQHINFLKQEIYTLNIDEILKNPKLREKIDFMINQF
;FT SLQICSEHPNAFWNRKKHVVTLPYEEEFSEANSIPTKARPCQMNSEYLE
;FT LCKKEIDSLLHKGLIRSSKSPWSCTAFYVNKHAEQERGVPRLVINYKPL
;FT NKVLKWIRYPIPNKKDLLDRLHNAIIFSKFDLKSGYWQIQIAETDRYKT
;FT AFNVPIGQFEWNVMPFGLKNAPSEFQKIMNDIFIPYSNFIIVYIDDILV
;FT FSNTIEMHFKHLELFKKIIIQNGLVISKPKMFLFQTKVRFLGHNIEKGK
;FT IIPINRSIEFASKFPDVITDKTQLQRFLGSLNYIAPFYKDLAKDTSILY
;FT DRLKKIPKPWTDAHTETVRKIKERVKNLPCLTLANPTWPKIIETDASDI
;FT GYGGILKQNAPGEKIEYLIQFHSGKWNNSQKNYSTVAKEILAIVKCVLK
;FT FQGDLYNQKFLIKTDCQSAKFMFNKDCKHDVSKQMFARWQALLAPFDFE
;FT IHYKKGEDNSLPDFLTREYLAS"
;XX
;DR [1] (Consensus)
;XX
;SQ Sequence 7722 BP; 2842 A; 1120 C; 1379 G; 2381 T; 0 other;
Caulimovirus-PAx
tttaatgattagaaacaagactactgttaaaaacagtagaaaggaggaatatgatattccacaacaccta
gatttacttaacaagtggactattcctaaagttcaacctaagttaatataccagttaggaacctttgaga
aatttggtttgaaacaagtagtgaaaactacagaagatactattactattgataatgatcatgctacctt
taatttattgactcagcatgatttattgccttatatgcacactcatagatttattcatattggtttaatt
caagttgcatttaaacctttaactttaagaggactgaatgaaacgtttatagctgctctcagggatggta
gaaaccataactggaagaaatcccttatagggattgtacaatcaagtttggcctatggcccggtatattt
taatgcttatccgaatttacaaatatcattatcggatgaaaatgctttaaaagccctagttttaaatgtt
aaactgcatggttatgactatatgcctaatacagaagtcatatgtatctgttataggatttattttaaac
cattatatactctaaatcctatgtgcaaaatcatggattttaaaaatgagactattttaatagaaactaa
ttttggtaaaacaaaggttactaccagaagaccaatcaagtgggaagaaatagatttccctaaagagtgg
gtcatggaggatgctgctcctccaaaagatttgaaaagtactgatttaactggaattgaacagactccag
agggcaatgtaaaaatcagatttgaatctgatgacgttgtaagtaggataaatatgggatcggatagatc
tagctctaaaaatttatctagatctaattctttgtattattctcatatctctccaattgattatattgtt
gaaacccccttctagagcttctacttctcaaataagagaagaaatgcctagatataatagtttgaaaattg
ataaagataatattgtcactcctgttgctcctataagagaacaacagattgacgatgcggatataactgc
atctgaaatggattttggtgaaaatatttaatggattttacatccacaattgattttagtcggggatcac
ccgcaagaaagaaaatatctgaggatttttataaacccagatgggaacttttagaagatggtttttcaa
aagttttaatagaactgaacaaaatgattttcaagaagaatttttataaagatcttgagaaagtcaacaaa
attatatcttttatcccatggtttatgggaaaatatgtggcaaattatattgctgttttggaaagacaat
tcgagttgttaaatggtaaaatcattgacgctgtattacctccacagcaacctttaaaataaagaagga
gggaaaaatgatggattttgcagcttttctgcgatgatagaggataataccttacaagttactgttaaa
catattaatattatgttaaaacagcataactatactaatatttatgttcatatattgggtgagcatattg
tatctttacatgataaggtagataaactatgttctataataagcagtaatactgctagcacatcaggaaa
agaaaaattggaaaagccacctcgagtagccattccaactattcagcctcctcctgatgttgaagatttt
aaattaaaacctgatgcatataatatcgaaaagtttttaaatgagaaatttaaagagttatcactaaagc
ctctaattaatacggaggaaagtgataatgcattttctgaaaatgaagtttttgaagatgattttccaaa
actaacttatgatcaaattaataaattaaagtcttcttatagttcgaaatttgctgataaaccaaggcaa
aggatgttttactatcctaggcctactcctcaggatgttctattcgaagagcaggatgatatgtatgctc
cgaatagtttcagcggtaagcagatatatgaatggaatattgatggccttaccaataggcaaatgtatgt
agttatgcatagaatgatgatgtatagcacgatttgcaagactaatggtaatagtgataaaactgttgcc

aatatgattacctcgggttttactggtcaattaaaagggtggtgggataattatctcactgttgaccaga
gaaatgaaattttgaatgcagttaaaagtgaaccaggaacttctacagctgataatactaatgctgttta
ttttttggttgttaatataattgaacacttttcaggaagatggtctgataatagtgagactattcgaacc
ctcttacagaatatgagatgtaagactttaacttcatttagactttataaagatacatttctttgtagag
tcatggagttaccagagtgtaatagcactcattggaaatctaaatttatagatggacttcctaacctctt
tgcagaaagagttaaaaaagttcttaggcaagaaggagtcaatattaactatgatgaccatacttatgga
agtttggttggaatatgtatccaagaaggattggcgttatgtaatgatatcaagctaaatcaggagctta
aacgccatcgtcttaatgaaagacaacagttgggtgaattttgtggtcaattcggcatggatgatccacg
aacccgaaccaaatcaaagtctcataagcttcatagaagcaagactggtaagaaacatgattaccaggaa
tggagacaaagatattcggagagacgcacagaaaagcgtgctagaagaaaaaaggcttttaaagagcgaa
aggatgacattaaagccaaagccaagaatccactaaattgttggaaatgtggaaggtttggtcatcgata
taaagactgcaaggtcaaagagagaatcaaagctcttaatatagatgaagatctgaaggattctttattg
aagatcttattaaattctgaaccagaaatatcggatccaggtgaggaaaatatccctaccaaggaagaac
taaataaagtcttccaggatgaaagtgatattgtcactagttctgaggatgaatgtgatccttgcaagat
tggacaaccttgtgataataaagaaaatgatgaattttataatttattttctcaatttagtcttaacact
ttggataatgataatttaatagaattattgaaaatggtacaggattccaaagtgagatccaaacttatcg
atcagatcaaaggaaatttcgaacagaaggaatccgaaggattggaaaagggtttctatgctcctatgga
taatactccatacaccatgaaagaagttaaaagacttctagaacaaagaaaaatgagtcatagtggttca
cctactacgtctcaggatttggaaaaagaggtttataacctcaaaaaagagatttataatctcaagaaag
aaaataagatattagatcaaaggatttctagaattgaaggaaaaggtaaagaaaaggatgaaaccactac
cccttctaattctgaaaatcttgataatatttttagcgatatgaatgaagagtttttaactaccctagaa
gtatttacttctcaaaagttttttataaaaattactcttcttattgattataatttcaaaaaagaattta
tagctcttgttgatagtggagcagatttaaattgtgttcaagaaggattaattccttcacgatattttca
taagactactcatagtttaaggtctgctaatggtgagaaaatgcagattgagtacaaattgccaacagct
tctatatgcacaaataaggtatgtatacctcaaagctttgttttaataaaaaatatgtctaaccaagtta
tactaggaacacccttttacaaaaaatatttcctattactaggattgataaagaagggataataggaac
ttttcaaggaaaacctattatttttcagtttataactgaacctattactagagttttaaaccagttgaaa
gatatgctagctaagaagagtcagcatatcaattttcttaaacaagaaatttatactttaaatatcgatg
agattttaaaaaatcctaaattgagagaaaaaattgattttatgataaatcagttttcattacaaatttg
tagtgaacatccaaatgctttttggaataggaaaaaacatgtagttacacttccatatgaggaagagttt
tctgaagctaatagtatccctactaaggcaagaccttgccagatgaattcagaatacttggaattatgca
aaaaggaaatagattctttattgcataaaggtctcatacgatcatcgaaatcaccatggtcctgtacagc
gttctatgttaataaacatgctgaacaggaaagaggagttccaaggttagtaataaactataaacctttg
aataaagttttgaaatggatcaggtatccaattcctaataaaaaggatttattagataggcttcataatg
ctattatatttttctaaatttgatttaaaatcagggtactggcagattcaaatagctgaaactgacaggta
caaaactgctttcaacgttccaataggacaatttgaatggaacgttatgccatttggtttaaaaaatgct
ccttctgaatttcagaaaattatgaatgacatattcatcccttatagtaattttattatcgtttatattg
acgatatcttagtattttctaatactatcgaaatgcattttaagcatttagaattatttaaaaaaattat
tatacaaaatggtttggttatatctaaacctaaaatgtttttgtttcaaactaaagtaagatttcttggc
cataatattgaaaaggggaaaattattcctattaatagaagtatagaatttgcttctaaatttccagatg
ttataactgacaaaactcaattacagagatttcttggaagtctaaattatattgctccattttataaaga
tctggctaaggatacatctattttatatgatagattgaaaaagattccaaagccatggacagatgctcat
actgagacagtccggaaaatcaaggaaagggtaaaaaatctcccttgtcttacattggctaatcctacat
ggccaaaaattatagagaccgatgcatcagacattggttatggaggtattttaaaacaaaatgctccagg
agagaaaattgaatatttgatacaattccattctggtaaatggaataacagtcagaaaaactattcaact
gtagcaaaagaaatacttgctatagtaaaatgtgttttaaaattccaaggtgatttatacaatcagaagt
ttcttataaagaccgattgccaatcagctaaattcatgtttaataaagattgtaaacatgatgtctctaa
gcaaatgtttgctagatggcaagcgctattagctccatttgattttgaaatacattataagaaaggagaa
gataatagccttcctgatttttgactagagaatatttggctagttaatgataataatcattatcttaat
gtgcaggttatttaatatgtgtagacgagataatatgctttattattgtccatactgccaaaatcaactc
ttacccaggcagatgtatgctatatattgggatttgcgtgcaacgcagagtagattgaaaaaactggtgt
atgaagctgaagatttattatcacgtgccaagttctttaataaagaagatgtagatcatgttatcagcat
tatatgcacagcaaaaactagaatgaaagatattacaaaggaaattataggaatccataaagtattggaa
atttatccaatatatctggcctgatgaatttatgatctattatatgcaggttatgcatccggcccgaggg
aacccaacctctaatagaggaaggggaaacagaggaagaggcactggtagaggtgctgtaatagcccaaa
tgggtaagcaaaggcttatagccgaagacttaacaggatcatcttcgaagagtaatacaagcaatgatac
aatgtcataccaacagtatcaagaattcctgaaatggaaaaattctcaagagaagcaaaagggaacagag
gataatatgcctccgtcatattcttctattatattagaagatgataacgagactatccagtcttatatta

agacagatcacaaggaaatagtattattccttgaagatagagaccttcaatggaaagatgagccatggaa
attaatggcaagatatttaaataatgcgtcatacgcagcctgtacttataagtacagaggtcattatgag
atgatacttaacaccactgaatcttgcgagatatcgcatttttacccagctaataccaaaaaagttttta
acttttccaaaatgataatcaaaaggattatatccgcagaagaatggggtttaagtaccctaactgaaag
aagttataatcatccagagatgaaaacaccggtaaaatataattactgggactatgtggatagcttttat
aaagcattgctatatgagaatgagaacaagaaacattcttggttcataaaaatatgctcaaatgtttata
aacagagcaatataccaaattggttttcacagtggtggatcgtttatggtcctacggtaaaattgctacc
ggagccattgctcagtttatatactgaatgggtggatgtatctcccaagattcatgaattacagggaaat
aacactcttatagagggaaattcttatatgaatttctttattgaatattcaatcccatggattatgaaat
ggagcgttgaagtggattataccccacataacattcctgccttatatagaacctttatacgaggttttg
gaaaaaaatgattcaaagaacagctgatggtaatctacactgtcaagatacaattgatttaataaaacaa
tcactcgagaaatataaacagattgatcatctggaacaagttaatacaccaagtccttatcaacacattc
tccataagttaaggatgaagaagggattaatgaccaaagaggaaaagttagccttctacttggaagaagt
caaaaaagatttaatcaaaaatcttgaggttgaaattcagtcagacacgtcaatggcttctgcaaataat
acagaagatgagggcgaatattgtctaccaggggaatcccaatcaattacagaagatgaagcaaatgatt
atatcaccaatataatgacacaggtgactgaacaagtgactaaagagagggaatcttccaaggcacaaaa
cgctaagggaaaggataaactctgacgaggacatattattcaagacccctcagcttttataaagttacct
ttaataaagtagcttcaataatgttgctctctctttaataaaagcagctttgataaagttgtgaaaataa
tgggccatctatcggccatttttctttctttaattttcaaccggccacttgtaaagtttgggcctaattt
gttgatttattttgtctataaatagcttaggagttagaacggagaagggcatcagtttacaaccctctca
aaattgtaaatctctcctattgtaaaaactttcaagcatataaaattctcttaattgcctgagcgcaaaa
ccaggtatttaacattgttcttttaattttgtttcttttaataaatttaagctttgatacttggcctaat
gataggctaaatcattaagtgt

Appendix 3.8 Approved sequence of *Caulimovirus-PAx*, the *Caulimovirus* member in *P. axillaris* subsp *axillaris* N by the Repbase database.

```
;ID Caulimovirus-PIn DNA ; PLN ; 8012 BP
;XX
;DE Integrated caulimovirus sequence.
;XX
;AC .
;XX
;DT 17-Aug-2017 (Rel. -1, Created)
;DT 17-Aug-2017 (Rel. -1, Last updated, Version 1)
;XX
;KW Caulimoviridae; Integrated Virus; Caulimovirus-PIn.
;XX
;OS Petunia integrifolia subsP. integrifolia subsp inflata
;XX
;OC Petunia integrifolia subsP. integrifolia subsp inflata
;OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
;OC Spermatophyta; Magnoliophyta; eudicotyledons; Gunneridae;
;OC Pentapetalae; asterids; lamiids; Solanales; Solanaceae;
;OC Petunioideae; Petunia.
;XX
;RN [1] (bases 1 to 8012)
;RA Alisawi,O., Richert-Poggeler,K.,
;RA Heslop-Harrison,P. and Schwarzacher,T.
;RT Integrated caulimovirus sequence within Petunia inflata genome.
;RL Direct Submission to RR (17-Aug-2017)
;XX
;CC 100% identical to consensus.
;XX
;FT CDS 787..2298
```

;FT /product="Caulimovirus-PIn_1p"
;FT /translation="ESIKYWKFIQYIWPDEFMIYYMQVMHPARGNPTSN
;FT RGRGNRGRGTGRGDVIAQMGKQRLIAEDLTGSSSKSNTSNDTMSYQQYQ
;FT EFLKWKNSQEKQKGTEDNMPPSYSSIILEDDNETIQSYIKTDHKEIILF
;FT LEDRDLQWKDEPWKLMARYLNNASYAACTYKYRGHYEMILNTTESCEIS
;FT HFYPANTKKVFNFSKMIIKRIISAEEWGLSTLTERSYNHPEMKTPVKYN
;FT YWDYVDSFYKALLYENENKKHSWFIKICSNVYKQSNIPNWFSQWWIVYG
;FT PTVKLLPEPLLSLYTEWVDVSPKIHELQGNNTLIEENSYMNFFIEYSIP
;FT WIMKWSVEVDYTPHNIPALYRTFYTRFWKKMIQRTADGNLHCQDTIDLI
;FT KQSLEKYKQIDHLEQVNTPSPYQHILHKLRMKKGLMTKEEKLAFYLEEV
;FT KKDLIKNLEVEIQSDTSMASANNTEDEGEYCLPGESQSITEDEANDYIT
;FT NIMTQVTEQVTKERESSKAQNAKGKDKL"
;XX
;DR [1] (Consensus)
;XX
;SQ Sequence 8012 BP; 2908 A; 1183 C; 1436 G; 2485 T; 0 other;
Caulimovirus-PIn
gcaagtatttcttttgctacagttgaatagtttttctgactgttattccatttaccagaatggaattgta
tcaaatattcaattttctctcctggagcattttgttttaaaatacctccataaccaatgtctgatgcatc
ggcctcctggagtattttaaaacaaaatgctccaggagagaaaattgaatatttgatacaattccatttt
ggtaaatggaataacagtcagaaaaactattcaactgtagcaaaagaaatacttgctatagtaaaatgtg
ttttaaaattccaaggtgatttatacaatcagaagtttcttataaagaccgattgccaatcagctaaatt
catgtttaataaagattgtaaacatgatgtctctaagcaaatgtttgctagatggcaggcgctattagcc
ccatttgattttgaaatacattataagaaaggagaagataatagccttcctgatttttgactagagaat
atttggctagttaatgataataatcattatcttaatgtgcaggttatttaatatgtgtagacgagataat
atgctttattattgtccatactgccaaaatcaactcttacccaagcagatgtatgctatatattgggatt
tgcgtgcaacgcagagtagattgaaaaaactggtgtatgaagctgaagatttattatcacgtgccaagtt
ctttaataaagaagatgtagatcatgttatcagcattatatgcacagcaaaaactagaatgaaagatatt
acaaagcaaattataggaatccataaagtattggaaatttatccaatatatctggcctgatgaatttatg
atctattatatgcaggttatgcatccggcccgagggaacccaacctctaatagaggaaggggaaacagag
gaagaggcactggtagaggtgatgtaatagcccaaatgggtaagcaaaggcttatagccgaagacttaac
aggatcatcttcgaagagtaatacaagcaatgatacaatgtcataccaacagtatcaagaattcctgaaa
tggaaaaattctcaagaaaagcaaaagggaacagaggataatatgcctccgtcatattcttctattatat
tagaagatgataacgagactatccagtcttatattaagacagatcacaaggaaataatattattccttga
agatagagacctccaatggaaagatgagccatggaaattaatggcaagatatttaaataatgcgtcatac
gcagcctgtacttataagtacagaggtcattatgagatgatacttaacaccactgaatcttgcgagatat
cgcatttttacccagctaataccaaaaaagtttttaacttttccaaaatgataatcaaaagaattatatc
cgcagaagaatggggtttaagtaccctaactgaaagaagttataatcatccagagatgaaaacaccggta
aaatataattactgggactatgtggatagctttataaagcattgctatatgagaatgagaacaagaaac
attcttggttcataaaaatatgctcaaatgtttataaacagagcaatataccaaattggttttcacagtg
gtggatcgtttatggtcctacggtaaaattgctaccagagccattgctcagtttatatactgaatgggtg
gatgtatctcccaagattcatgaattacagggaaataacactcttatagaggaaaattcttatatgaatt
tctttattgaatattcaatcccatggattatgaaatggagcgttgaagtggattataccccacataacat
tcctgccttatatagaacctttatacgaggttttggaaaaaaatgattcaaagaacagctgatggtaat
ctacactgtcaagatacaattgatttaataaaacaatcactcgagaaatataaacagattgatcatctgg
aacaagttaatacaccaagtccttatcaacacattctccataagttaaggatgaagaagggattaatgac
caaagaggaaaagttagccttctacttggaagaagtcaaaaaagatttaatcaaaaatcttgaggttgaa
attcagtcagacacgtcaatggcttctgcaaataatacagaagatgagggcgaatattgtctaccagggg
aatcccaatcaattacagaagatgaagcaaatgattatatcaccaatataatgacacaggtgactgaaca
agtgactaaagagagggaatcttccaaggcacaaaacgctaagggaaaggataaactctgacgaggacat
attattcaagacccctcagcttttataaagttacctttaataaagtagcttcaataatgttgctctctct
ttaataaaagcagctttgataaagttgtaatgggccatctatcggccatttttctttctttaatttcaa
ccggccacttgtaaagtttgggcctaatttgttgatttattttgtctataaatagcttaggagttagaac
ggagaagggcatcagtttacaaccctctcaaaattgtaaatctctcctattgtaaaaactttcaagcata
taaaattctcttaattgcctgagcgcaaaaccaggtatttaacattgttcttttaattttgtttctttta
ataaatttaagctttgatacttggcctaatgagaggctaaatcattaagtgttgagctctgtcacactag
gctactagggtaaaatgattctaataacagattatgattagaatctcagcaggtggtggtgttaaaggta
tggtttgatagacaacccaggcggaaggctgcatgccagaaccaaggttctgttgtttgcccgcttagcc

gaggatggcgcttagggctgttgttgaacctaattttgggttcagccactccttgccagggttgtgtcac
cagtaaaattttcctgttacccccatgctgttaagatccgtcataaatttggtatcagagccatggtgca
cagtagcaataggtgacgagttcaacacttaactttctgatgaatatcattagaaacaagactactgtta
aaaacagtagaaaggaggaatatgatattccacaacacctagatttacttaacaagtggactattcctaa
agttcaacctaagttaatataccagttaggaacctttgagaaattgggtttgaaacaagtagtgaaaact
acagaagatactattactattgataatgatcatgctacctttaatttattgactcagcatgatttattgc
cttatatgcacactcatagatttattcatattggtttgattcaagttgcatttaaacctttaactttaag
aggactaaatgaaacgtttatagctgctctcagggatggtagaaaccataactggaagaaatcccttata
gggattgtacaatcaagtttggcctatggcccggtatattataatgcttatccgaatttacagatatcat
tatcggatgaaaatgctttaaaagccctagttttaaatgttaaactgcatggttatgactatatgcctaa
tacagaagtcatatgtatctgttataggatttattttaaaccattatatactctaaatcctatgtgcaaa
atcatggattttaaaaatgagactattttaatagaaactaattttggtaaaacaaaggttactactagaa
gacctatcaagtgggaagaaatagattttcctaaagagtgggtcattgaggatgctgctcctccaaaaga
tttgaaaagtactgatttaactggaattgaacagactccagagggcaatgtaaaaatcagatttgaatct
gatgacgttgtaagtaggataaatatgggatcggatagatctagctctaaaaatttatctagatctaatt
ctttgtattattctcatatctctccaattgattatattgttgaaacccttctagagcttctacttctca
aataagagaagaaatgcttagatataatagtttgaaaattgataaagataatattgttactcctgttgct
ccaataaaggaacaacaaaggaatgatgttgatgcagatataactgagtcagaaatgggttttgaaattt
aatggattttaagcctaaaatagattttagtcggggatcacccgcaagaaaaaaaaattctgaagaattt
tataaacctagatgggcattatttcggaaatgatttttaaaagttttgataaaaccaagcaacatgcgt
ttcaagaagaattctatgaagatcttgaaaagcatgacaagcttatatcttttgtcttgtggtttatggg
aagatatgtagcaaataatattgctgtcatgggaaggcaatttgaaatggcaaatggtacaattgttgat
gctgtattacctccacagcaacctttaaaatgaaaaaacttgactttgcagcgtttgctacattatggg
atgagcaaactttgcaagttactgttaaacatattaatattatgttaaaacagcataactatactaatat
ttatgttcatgtattaggtgagcatattgtatctttacatgataaggtagataaaatatgttctataata
agcagtaatactgctagcacatcaggaaaagaaaaattggaaaagccacctcgagtagccattccaacta
ttcagcctcctcctgatgttgaagattttaaattaaaacctgatgcatataatatcgaaaagtttttaaa
tgagaaatttaaaaagttatcactaaagcctctaattaatacggaggaaagtgataatgcattatctgaa
aatgaagtttttgaaagtgattttccaaaactaacttatgatcaaattaataaattaaagtcttcttata
gttcgaaatttgctgataaaccaaggcaaaggatgttttactatcctaggcctactcctcaggatgttta
attcgaagagcaggatgatatgtatgctccgaatagtttcagcggtaagcagatatatgaatggaatatt
gatggccttaccagtaggcaaatgtatgtagttatgcatagaatgatgatgtatagcacgatttgcaaga
ctaatggtaatagtgataaaactgttgccaatatgattacctcgggttttactggtcaattaaaaggggtg
gtgggataattatctcactgttgaccagagaaatgaaattttgaatgcagttaaaagtgaaccaggaact
tctacagctgataatactaatgctgtttatttttttggttgttaatataattgaacacttttcaggaagat
ggtctgataatagtgagactattcgaaccctcttacagaatatgagatgtaagactttaacttcatttag
actttataaagatacatttctttgtagagtcatggagttaccagagtgtaatagcactcattggaaatct
aaatttatagatggacttcctaacctctttgcagaaagagttaaaaaagttcttaggcaagaaggagtca
atattaactatgatgaccatacttatggaagtttggttggaatatgtatccaagaaggattggcgttatg
taatgatatcaagctaaatcaggagcttaaacgccatcgtcttaatgaaagacaacagttgggtgaattt
tgtggtcaattcggcatggatgatccacgaacccgaaccaaatcaaagtctcataagcttcataaaagca
agactggtaagaaacatgattaccaagaatggagacaaaaatatttagaaagacgcacagaaaagtgtac
tagaagaaaaaaagctttaaagagcgaaaggatgacattaaagccaaagccaagaatccactaaattgt
tggaaatgtggaaggtttggtcatcgatataaagactgcaaggtcaaagagagaatcaaagctcttaata
tagatgaagatctgaaggattctttattgaagatcttattaaattctgaaccagaaatatcggatccagg
tgaggaaaatatccctaccaaggaagatctaaataaagtcttccaggatgaaagtgatattgtcactagt
tctgaggatgaatgtgatccttgcaagattggacaaccttgtgataatgaagaaaatgatgaatttttata
atttattttctcaatttagtctttaacactttggataatgataatttaatagaattattgaaaatggtac
aggattccaaagtgagatccaaacttatcgatcagatcaaaggaaatttcgaacagaaggaatccgaagg
attggaaaagggtttctatgctcctatggataatactccatacaccatgaaagaagttaaaagacttcta
gaacaaagaaaaatgagtcatagtggttcacctactacgtctcaggatttggaaaaagaggtttataacc
tcaaaaaagagatttataatctcaagaaagaaaataagatattagatcaaaggatttctagaattgaagg
aaaaggtaaagaaaaagatgaaaccactacccctctaattctgaaaatcttgataatatttttagcgat
atgaatgaagagtttttaactacactagaagtatttacttctcaaaagtttttttataaaaattactcttc
ttattgattataatttcaaaaaagaatttatagctcttgttgatagtggagcagatttaaattgtgttca
agaaggattaattccttcacgatattttcataagactactcatagtttaaggtctgctaatggtgagaaa
atgcagattgagtacaaattgccaacagcttctatatgcacaaataaggtatgtatacctcaaagctttg
ttttaataaaaaatatgtctaaccaagttatactaggaacacccttttacaaaaaatatttcctattac

```
taggattgataaagaagagataataggaacttttcaaggaaaacctattattttcagtttataactgaa
cctattactagagttttaaaccagttgaaagatatgctagctaagaagagtcagcatatcaattttctta
aacaagaaatttatactttaaatatcgatgagattttaaaaaatcctaaattgagagaaaaaattgattt
tatgataaatcagttttcattacaaatttgtagtgaacatccaaatgctttttggaataggaaaaaaaca
tgtagttacacttccatatgaggaagagttttctgaagctaatagtatccctactaagccaagaccttgc
cagatgaattcagaatacttggaattatgcaaaaaggaaatagattctttattgcataaaggtctcatac
gatcatcgaaatcaccatggtcctgtacagcgttctatgttaataaacatgctgaacaggaaagaggagt
tccaaggttagtaataaactataaacctttgaataaagttttgaaatggatcaggtatccaattcctaat
aaaaaggatttattagataggcttcataatgctattatattttctaaatttgatttaaaatcagggtact
ggcagattcaaatagctgaaactgacaggtacaaaactgctttcaacgttccaataggacaatttgaatg
gaacgttatgccatttggtttaaaaaatgctccttctgaatttcagaaaattatgaatgacatattcatc
ccttatagtaattttattatcgtttatattgacgatatcttagtattttctaatactatcgaaatgcatt
ttaagcatttagaattatttaaaaaaattattatacaaaatggtttggttatatctaaatctaaaatgtt
tttgtttcaaactaaagcaagatttcttggccataatattgaaaaggggaaaattattcctattaataga
agtatagaatttgcttctaaatttccagatgttataactgacaaaactcaattacagagatttcttggaa
gtctaaattatattgctccatttcataatcca
```

Appendix 3.9 Approved sequence of *Caulimovirus-PIn*, the *Caulimovirus* member in *P. integrifolia* subsp *inflata S6* by the Repbase database.

```
;ID Caulimovirus-PHy DNA ; PLN ; 8060 BP
;XX
;DE Integrated caulimovirus sequence.
;XX
;AC .
;XX
;DT 18-Aug-2017 (Rel. -1, Created)
;DT 18-Aug-2017 (Rel. -1, Last updated, Version 1)
;XX
;KW Caulimoviridae; Integrated Virus; Caulimovirus-PHy.
;XX
;OS Petunia x hybrida
;XX
;OC Petunia x hybrida
;OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
;OC Spermatophyta; Magnoliophyta; eudicotyledons; Gunneridae;
;OC Pentapetalae; asterids; lamiids; Solanales; Solanaceae;
;OC Petunioideae; Petunia.
;XX
;RN [1] (bases 1 to 8060)
;RA Alisawi,O., Richert-Poggeler, Heslop-Harrison,P. and
Schwarzacher,T.
;RT New integrated caulimovirus sequence within Petunia hybrida genome.
;RL Direct Submission to RR (18-Aug-2017)
;XX
;CC 95.9% identical to consensus.
;CC ~ 61% identical to Caulimovirus-2_STu.
;CC ~ 55.9% identical to Caulimovirus-3_STu.
;CC ~ 61.1% identical to Caulimovirus-4_STu.
;XX
;FT CDS 1549..722
;FT /product="Caulimovirus-PHy_1p"
;FT /translation="NIITGTMWIAFIKALLYENENKKHSWFIKICSNVY
```

;FT KQSNIPNWFSQWWIVYGPTVKLLPEPLLSLYTEWVDVSPKIHELQGNNT
;FT LIEGNSYMNFFIEYSIPWIMKWSVEVDYTPHNIPALYRTFYTRFWKKMI
;FT QRTADGNLHCQDTIDLIKQSLEKYKQIDHLEQVNTPSPYQHILHKLRMK
;FT KGLMTKEEKLAFYLEEVKKDLIKNLEVEIQSDTSMASANNTEDEGEYCL
;FT PGESQSITEDEANDYITNIMTQVTEQVTKERESSKAQNAKGKDKL"
;XX
;DR [1] (Consensus)
;XX
;SQ Sequence 8060 BP; 2520 A; 1418 C; 1146 G; 2976 T; 0 other;
Caulimovirus-PHy
gctcttgttgatagtggagcagatttaaattgtgtggaatctatttccttttgcataattccagacaca
accctggcaaggagtggctgaacccaaaattaggttcaacaacagccctaaacgccatcctcggctaagc
gggcaaacaacagaaccttggttctggcattcaaccttccgcctgggttgtctatcaaaccatacttta
acaccaccacctgctgagattctaatcttaatctgttatttagaatcattttacccccagtagcctagtg
tgacagagctcaacacttaatgaattttagcctatcatttaggccaagtatcaaagcttaaaatttatta
aaagaaacaaaattaaaagaacaatgttaaatacctggttttgcgctcaggcaattaagagaatttttat
atgcttgaaagtttttacaataggagagatttacaattttgagagggttgtgaaaactgatgcccttctc
cgttctaactcctaagctatttatagacaaagaataaatcaacaaattaggcccaaactttacaagtggc
cggttgaaaattaaagaaagaaaaatggccgatagatggcccattattttcacaactttatcaaagctgc
ttttattaaagagagagcaacattattgaagctactttattaaaggtaactttataaaagctgaggggtc
ttgaataatatgtcctcgtcagagtttatcctttcccttagcgttttgtgccttggaagattccctctct
ttagtcacttgttcagtcacctgtgtcattatattggtgatataatcatttgcttcatcttctgtaattg
attgggattcccctggtagacaatattcgccctcatcttctgtattatttgcagaagccattgacgtgtc
tgactgaatttcaacctcaagatttttgattaaatctttttttgacttcttccaagtagaaggctaacttt
tcctctttggtcattaatccctttcttcatccttaacttatggagaatgtgttgataaggacttggtgtat
taacttgttccagatgatcaatctgtttatatttctcgagtgattgttttattaaatcaattgtatcttg
acagtgtagattaccatcagctgttctttgaatcatttttttccaaaacctcgtataaaaggttctatat
aaggcaggaatgttatgtggggtataatccacttcaacgctccatttcataatccatgggattgaatatt
caataaagaaattcatataagaatttccctctataagagtgttatttccctgtaattcatgaatcttggg
agatacatccacccattcagtatataaactgagcaatggctccggtagcaattttaccgtaggaccataa
acgatccaccactgtgaaaaccaatttggtatattgctctgtttataaacatttgagcatatttttatga
accaagaatgtttcttgttctcattctcatatagcaatgcttttataaaagctatccacatagtcccagt
aattatattttaccggtgttttcatctctggatgattataacttctttcagttagggtacttaaacccca
ttcttctgcggatataatccttttgattatcattttggaaaaaagttaaaaacttttttggtattagctg
ggtaaaaatgcgatatctcgcaagattcagtggtgttaagtatcatctcataatgacctctgtacttata
agtacaggctgcgtatgacgcattatttaaatatcttgccattaatttccatggctcatctttccattgg
aggtctctatcttcaaggaataatattatttccttgtgatctgtcttaatataagactggatagtttcgt
tatcatcttctaatataatagaagaatatgacggaggcatattatcctctgttccctttgcttctcttg
agaatttttccatttcaggaattcttgatactgttggtatgacattgtatcattgcttgtattactcttc
gaagatgatcctgttaagatcttcggctataagcctttgcttacccatttgggctattacagcacctcta
ccagtgcctcttcctctgtttccccttcctctattagaggttgggtttcctcgggccggatgcataacct
gcatataatagatcataaattcatcaggccagatatattggataaatttccaatactttacggattccta
taatttcctttgtaatatctttcattctagtttttgctgtgcatataatgctgataacatgatctacatc
ttctttattaaagaacttggcacgtgataataaatcttcagcttcatacaccagtttttttcaatctact
ctgcgttgcacgcaaatcccaatatatagcatacatctgcctgggtaagagttgattttggcagtatgga
caataataaagcatattatctcgtctacacatattaaataacctgcacattcaaaataatgattctatta
tcattaactagccaaatattctctagttaaaaaaatcaggaagactattatcttctccttttttatattg
tatttcaaaatcaaatggggctaataatgcctgccatcttgcaaacatttgtttagaaacatcatgttta
caatctttattaaacataaatttagctgattggcaatcagtctttataataaacttttggttatataaat
caccttgaaatttaaaacacattttactatagcaagtatttcttttgctacagttgcataatttttttct
ggctattattccatttaccagaatggaatatgaaccaaatattcaattttattcttctggagagtattgc
tttaaaatacctccataaccaatgtcagaagcatcagtttctattatctttggccaagtaggattagcca
atgtaagacaaggaagagtttttactctttccttgatctttctggactgttttagtatgagcatcagtc
catggctttggaatctttttcaatctatcatataagatagctgtatctttagcaagattctcataaaat
ggagcaatataatttaaactccctagaaatctctgtaattgagttttatctgttataatatcaggaaatt
tagaagcaaattcaatacttctattaataggtataatctttccttttcaatattatgaccaaggaatct
tacctttagtttgaaataaaaacatttttggttttagatataaccaaaccattttgtataattattttct
taaataaatttcaaagatgcttaaaatgcatttcaatagtatttgaaaatactaatatatcgtcaatata

aacgataataaaattgctataaggaatgaatatatcattcataattttctgaaattcagaaggagcattt
ttttaatccaaatggcataacattccattcgaactgtcctattgggacattaaaagcagttttatacctg
tcagcttcagctatttgaatctgccagtatcctgatttcaaatcaaatttagaaaatataatagcattat
gaagcctatccaataaatccttttttattaggaataggatatctaatccatttcaaaactttattcaaagg
tttatagtttatgactaaccttggaactcctctttcttgttcagcatgtttattaacataaaaagctgta
caagaccatggggatttttttgaagatcgaataagacctttatgtaataaagaatctatttccttttttac
ataattccaaatattctgaattcatttggcaaggtcttgccttagtaggaatacttattagtttcagaaa
agtcttcctcatatggaagtgtaactacatgctttttcctattccaaaaaagcatttggatgttcactac
aaatttgtaatgaaaactgatttattataaaatcaatttttttctctcaatttgggatttttttataaat
ctcatcaatatttaaagtataaatttcttgtttaagaaaattgatatgctgacttttcttagctcagcat
atcttttaactggtttaaaactctagtaataggttcagttataaactgaaaaataataggtttttcttga
aaagttcctattatcccttctttatcaatcctagtaataggaaatatttttttgtaaaaagggtgttccta
gtataacttggttagacatatttttttattaaaacaaagctttgaggtatacataccttatttgtgcatat
agaagcttttggcaatttgtactcaatctgcattttctcaccattagcagaccttaaactatgagtagtt
ttatgaaaatatcgtgaaggaattaatccttcttgaacacaatttaaatctgctccactatcaacaagag
ctataaattctttttttgaaattataatcaataagaagagtaatttttataaaaaaactttgagaagtaa
atacttctagggtagttaaaaactctgcctcattcccgatatcgctaaaaatattatcaagattttcaga
attagaaggggtagtggtttcatccttttctttaccttttccttcaattctagaaatcctttgatctaat
atcctattttccttcttgagattataaatctccttttttatgaggttataaacctctttttttccaaatcct
gagacgtagtaggtgaactactatgactcatttttctttttgttctagaagtcttttaacttctttcatgg
tgtatggagtattatccataggggcatagaaatccctttttttcaatccttcggattccttctgttcgaaat
ttcttcctttgatctgatcgcgataagtttagatctaacttcggaatcctgtaccatcttcaataaattc
tattaaattatcattatccaaagtgttaaagactaaaattgagaaaataaattataaaattcatcatttt
cttcattatcacaaggttgtccaatcttgcaaggatcacattcatcctcagaactagtgacaatatcact
ttcatcctggaagactttaatttagttcttccttggtagggatattttcctcacctggatccgatatttc
tggttcagaatttaataagatctttaataaagaatcttacacaaagaatctttcagatcttcatctatat
taagagctttgatcttttcttttacttttgcaatctttataatgatgaccaaatcttccacacttccaaca
attttgtggtttctttgatttagtggattcttggctttggctttaatgtcatcctttcgctctttaaaag
ccttttttcttctagcacgcttttctgtgcgtctctccgaatatctttgtctccattcctggtaatcatg
tttcttaccagtcttgcttctatgaagcttatgagactttgatttggttcgggttcgtggatcatccatg
ccgaattgaccacaaaattcacccaactgttgtctttcattaagacgatggcgtttaagctcttgattta
gcttaatttcattacataacgccaatccttcttggatacatattccaactaaacttccataagtatggtc
atcataatcgattaatatgaactcctccttgtctaagagctttttaactctttctgcaaagaggttagg
aagtccatctataaatttggatttccaatgagtgctattacactctggtaactccataactctgcaaaga
aatgtatctttataaagtctaaatgaagttaaagtcttacatctcatattctgtaagagggttcgaatag
tctcactattatcagaccatcttcctgaaaagtgttcaattatattaacaaccaaagaataaacagcatt
agaattatcagctgtagaagatgttcctggttctgtagaagttcctggttcacttttaactgcattcaaa
atttcatttcgctggtcaacagtgagataattatcccaccacccttttaattgaccagtaaaacctgaga
taatcatattattggcaacagttttatcactattaccattagtcttgcaaatcgtgctatacatcatcat
tctatgtataactacatacatttgcctattggtaaaggccatcaatattccattcatatatctgcttacc
gctgaaactattcggagcatacagtatcatcctgctcttcgaatagaacatcctgaggagtaggcctagg
atagtaaaacatcctttgccttggtttatcagcaaatttcgaactataagaagactttaatttattaatt
tgatcataagtcaatttattaatttgatcataagttagttttggaaaatcactttcgaaaaacttcattt
tcagatgaaaatgcattatcactttcctccgtattaattagaggctttagtgataactctttaaatttct
catttaaaaacttttcgatattagatgcatcaggtcttttaatttaaaattttcaacatcaggaggaggt
tgaatagttggaatggctactcgtggtggctcttccaatttttctttttcctgatgtgctagcagtattac
tgcttattgatagaacataatttatctaccttatcatgtaaagatacaatatgctcacccaatatatgaa
cataaatattagtatagttatgctgttttaacataatattaatatgtttaacagtaacttgtaaagtatt
atcttcaattattgcagaaaaagctgcaaaatcaatcatttttccctccttctttatttaaaaggttgc
tgtggaggtaatacagcatcaatgattttaccatttgacaattcaaattgtctttccaaaacagcaatat
tatttgctacatatttttcccataaaccatgggataaaagatataatttttattaactttctcaagatcttt
ataaaattcttcttgaaaatcattttgttcagttttattaaaacttttaaaaaaccattttctaaatagt
tcccatctgggtttataaaaatcctcagatattttctttcttgcgggtgatccccgactaaaatcaattg
tggacttaaaatccattaaatatttatttcaaaactttcaaaatccatttcagatgcagttatgtctaca
tcatttatctgttgttctttttataggagcaacaggagtaacaatattatttttatcaattttcaaagta
tgatatctaggcacttcctctcttatctgagaagtagaagctctagaaggagtttgaacaatataatcaa
ttggagaaatatgagaatgatataaagaattagatctagataaaccttgaaaactagatctgtcagatct
catatccatcctacttataacatcttcagattcaaatctgatttttacatttccttctggagtttgttca

attccagtcaaatctgtatttttaaaatcttttggaggagctgcttcctccatgacccactctttagggg
aaatctatttcatcccacttgataggtcttctggtagtaactttagttttaccaaaattagtttctatta
aaatagtttcatttttaaaatccataattttgcacataggatttaaagtatataatggcttaaaataaat
cctataacagatacatattacttctgtattaggcatatatatcataaccatgtaatttaacatttaaaac
taaagctttttaaagcattttcatccgacaatgatatctgtaaattcggataagcattaaaatatactgg
gccataggccaaacttgattgtacaatccctataagggatttcttccagttatggtttctaccatccctg
agagcagctataaaaagtttcatttaatcctcttaaagttaaaggtttaaatgcaacttgaattaaacca
atatgaataaatctatgagtatgcatataaggcaataaatcatgctgagacaataaattaaaagtagcat
gatcactatcaatcgtgatagtttcttcagtagttttaactacctgctttaaacccaatttttcaaaggt
tcctagctgg

Appendix 3.10 Approved sequence of *Caulimovirus-PHy*, the *Caulimovirus* member in *P. hybrida* by the Repbase database.

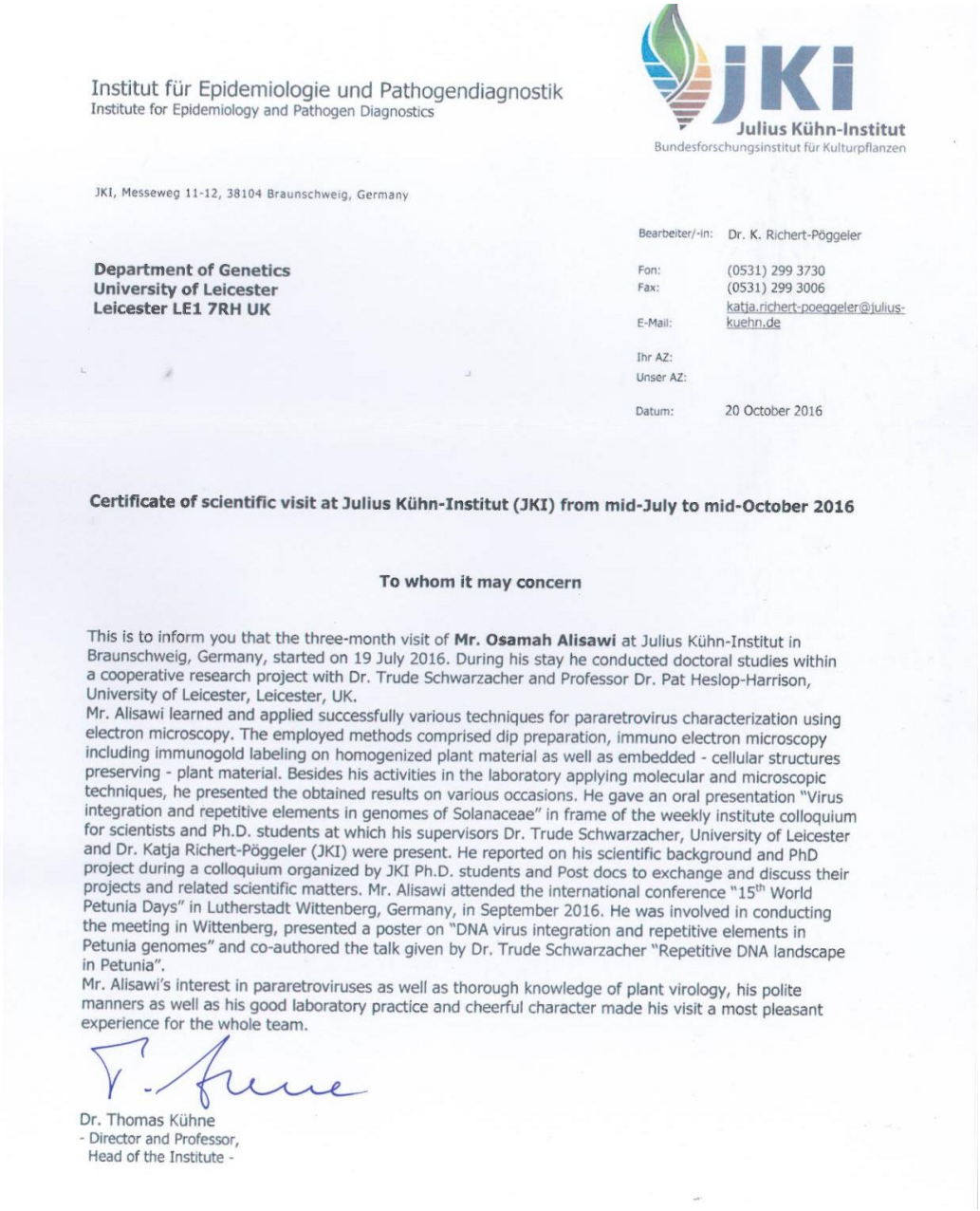
```
;ID Caulimovirus-PPa DNA ; PLN ; 7999 BP
;XX
;DE Integrated viral sequence of Caulimovirus in P. axillaris subsp
parodii.
;XX
;AC .
;XX
;DT 04-Apr-2018 (Rel. -1, Created)
;DT 04-Apr-2018 (Rel. -1, Last updated, Version 1)
;XX
;KW Caulimoviridae; Integrated Virus; Caulimovirus-PPa.
;XX
;OS Petunia axillaris subsP. axillaris subsp parodii
;XX
;OC Petunia axillaris subsP. axillaris subsp parodii
;OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
;OC Spermatophyta; Magnoliophyta; eudicotyledons; Gunneridae;
;OC Pentapetalae; asterids; lamiids; Solanales; Solanaceae;
;OC Petunioideae; Petunia.
;XX
;RN [1] (bases 1 to 7999)
;RA Alisawi,O., Richert-Poggeler,K.,
;RA Schwarzacher,T. and Heslop-Harrison,P.
;RT Integrated caulimoviral sequence within P. axillaris subsp parodii
S7 genome.
;RL Direct Submission to RR (04-Apr-2018)
;XX
;CC Consensus sequence.
;XX
;FT CDS 1437..5870
;FT /product="Caulimovirus-PPa_1p"
;FT /translation="KYLMDFTSTIDFSRGSPARKKISEDFYKPRWELFR
;FT KWFFKNFNKTEQNAFQEEFYKDLEKHNKLISFIPWFMGKYVANNIAVLE
;FT RQFEMSNGKIIDAVLPPQQPFKIKKEGKMIDFAAFSAMIEDNTLQITVR
;FT HINIMLKQQNYTNIYVHILGEHIVSLHDKVDKLCSIISSNTASTSGKEK
```

;FT LEKPPRVAIPTIQPPPDVENFKLKPDAYNIEKFLNEKFKELSLKPLINT
;FT EESDNAFSEDEVFEGDFPKLTYDQINKLKSSYSSKFADKPRQRMFYYPR
;FT PTPQDVLFEEQDDMYAPNSFSGKQIYEWNIDGLTNRQMYVVIHRMMMYS
;FT TICKTNGNSDKTVANMIIAGFTGQLKGWWDNYLTVDQRNEILNAVKSEP
;FT GTSTADNTNAVYFLVVNIIEHFSGRWSDNSETIRTLLQNMRCKTLTSFR
;FT LYKDTFLCRVMELPECNSTHWKSKFIDGLPNLFAERVKKALRQGGVHIN
;FT YDDHTYGSLVGICIQEGLALCNEIKLNQELKRHRLNERQQLGEFCGQFG
;FT MDDPRTRTKSKSHKLHRSKTGKKHDYQEWRQRYSERRTEKRTRRKKAFK
;FT ERKDDIKAKAKNPLNCWKCGRFGHRYKDCKVKERIKALNIDEDLKDSLC
;FT KILLNSEPEISDPGEENIPTKEELNKVFQDESDIVTSSEDECDPCKIGQ
;FT PCDNKENDEFYNLFSQFSLNTLDNDNLIELLKMVQDSEVRSKLIDQIKG
;FT NFEQKESEGLKKGFYAPMDNTPYTMKEVKRLLEQRKMSHSSSPTTSQDL
;FT EKEVYNLKQEIYNLKKENRILEQRISRIEGKGKEKDEILTPSNSENLDN
;FT IFSNMNEEFLTTLEVFTSQKFFIKITLLIDYNFKKEFIALVDSGADLNC
;FT VQEGLIPSRYFHKTTHSLRSANGEKMQIEYKLPKASICTNKVCIPQSFV
;FT LIKNMSNQVILGTPFLQKIFPITRIDKEGIIGTFQEKPIIFQFITEPIT
;FT RVLNQLKDMLAKKSQHINFLKQEIYTLNIDEILKNPKLREKIDFMINQF
;FT SLQICSEHPNAFWDRKKHVVTLPYEEEFSEANSIPTKARPCQMNSEYLE
;FT LCKKEIDSLLHKGLIRSSKSPWSCTAFYVNKHAEQERGVPRLVINYKPL
;FT NKVLKWIRYPIPNKKDLLDRLHNAIIFSKFDLKSGYWQIQIAEADRYKT
;FT AFNVPIGQFEWNVMPFGLKNAPSEFQKIMNDIFIPYSNFIIVYIDDILV
;FT FSNTIEMHFKHLELFKKIIIQNGLVISKPKMFLFQTKIRFLGHNIEKGK
;FT IIPINRSIEFASKFPDIITDKTQLQRFLGSLNYIAPFYENLAKDTAILY
;FT DRLKKIPKPWTDAHTKTVQKIKERVKNLPCLTLANPTWPKIIETDASDI
;FT GYGGILKQYSPEEKIEYLIQFHSGKWNNSQKNYATVAKEILAIVKCVLK
;FT FQGDLYNQKFIIKTDCQSAKFMFNKDCKHDVSKQMFARWQALLAPFDFE
;FT IHYKKGEDNSLPDFLTREYLAS"
;XX
;DR [1] (Consensus)
;XX
;SQ Sequence 7999 BP; 2984 A; 1120 C; 1395 G; 2500 T; 0 other;
Caulimovirus-PPa
tatttggcctaaatgataggctaaaatcattaagtgttgagctctatcacactaggctactggggtaaaa
tgattctaataacagattatgattagaatctcagcaggtggtggtgttaaaggtatggtttgatagacaa
cccaggcggaaggttgaatgccagaaccaaggttctgttgtttgcccgcttagccgaggatggcgcttag
ggctgttgttgaacctaatttgggttcagccactccttgccagggttgtgtcaccagtaaaatttcctg
ttacccccatgctgttaagatccgtcgaaatttggtatcagagccatggtgcacagtagcaataggtgaa
gagttcaacacttatctttctaatgaatatcattagaaacaaaactactgttaaaaacagtagaaaagag
gaatatgatattccacaacacttagatttacttaataaatggactattcctaaagttcaacctaagttaa
tatatcagctaggaacctttgaaaaattgggtttgaaacaagtagttaaaactacagaagatactatcac
tattgataatgatcatgctacttttaatttattgactcagcatgatttattgccttatatgcatactcat
agatttattcatattggtttaattcaagttgcatttaaacctttaactttaagaggattaaatgaaacgt
ttatagctgctctcagggatggtagaaaccataactggaaaaaatcccttatagggattgttcaatcaag
tttggcctatggcccagtatattttaatgcttatccgaatttacagatatcattgtcggatgaaaatgct
ttaaaagctttagttttaaatgttaaattacatggttatgattatatgcctaatacagaagttatatgta
tctgttataggatttatttaagccattatatactttaaatcctatgtgtaaaattatggatttaaaaa
tgagactattttaatagaaactaattttggtaaaactaaagttactaccagaagacctatcaagtgggat
gaaatagatttccctaaagagtgggtcatggaggatgctgctcctccaaaagatttgaaaagtactgatt
taactggaattgaacagactccagaaggtaatgtgaaaatcagatttgaatttgatgatagaataaatat
gggatcggatagatctagctctaaaaatttatctagatctaattctttgtattattctcatatctctcca
attgattatattgttcaaactccttctagagcttctacttctcaaataagagaagaagtgcctagatata
atagtttgaaaattgataaaaataatattgttactcctgttgctcctataagagaacaacaaatagacga
tgcggatataactgcatctgaaatggattttggtgaaaatatttaatggattttacatccacaattgatt
ttagtcggggatcacccgcaagaaagaaaatatctgaggatttttataaacccagatgggaacttttag
aaaatggtttttaaaaatttaataaaactgagcaaaatgcttttcaagaagaattttataaagatctt
gaaaagcataataaacttatatcttttatcccatggtttatgggaaaatatgtagcaaataatattgctg
ttttggaaagacaatttgaaatgtcaaatggtaaaatcattgacgctgtattacctccacagcaacctt
taaaataaagaaggagggaaaaatgattgattttgcagcttttctgcaatgattgaagataatactctt

caaattactgttagacacattaatataatgttgaaacagcagaactatactaatatttatgttcatatat
tgggtgaacatattgtatctttacatgataaggtagataaattatgttctataataagcagtaatactgc
tagcacatcaggaaaagaaaaattggaaaagccacctcgagtagccattccaactattcagcctcctcct
gatgttgaaaatttaaattaaaacctgatgcatataatattgaaaagtttttaaatgagaaatttaaag
agttatcactaaagcctctaattaatacggaggaaagtgataatgcattttctgaagatgaagtttttga
aggtgattttccaaaactaacttatgatcaaattaataaattaaaatcttcttatagttcgaaatttgct
gataaaccaaggcaaaggatgttttactatcccaggcctactcctcaggatgttctattcgaagagcagg
atgatatgtatgctccgaatagtttcagcggtaagcaaatatatgaatggaatattgatggccttaccaa
taggcaaatgtatgtagttatacatagaatgatgatgtatagcacgatttgcaagactaatggtaatagt
gataaaactgttgccaatatgattattgcaggttttactggtcaattaaaaggatggtgggataattatc
tcactgttgaccagcgaaatgaaattttgaatgcagttaaaagtgaaccaggaacttctacagctgataa
tactaatgctgtttatttttggttgttaatataattgaacactttcaggaagatggtctgataatagt
gagactattcgaaccctcttacagaatatgagatgtaagactttaacttcatttagactttataaagata
catttctttgtagagtcatggagttaccagagtgtaatagcactcattggaaatctaaatttatagatgg
acttcctaacctctttgcagaaagagttaagaaagctcttagacaaggaggagttcatattaattatgat
gaccatacttatggaagtttggttggaatatgtatccaagaaggattggcgttatgtaatgaaattaagc
taaatcaggagcttaaacgccatcgtcttaatgaaagacaacagttgggtgaattttgtggtcaattcgg
catggatgatccacgaacccgaaccaaatcaaagtctcataagcttcatagaagcaagactggtaagaaa
catgattaccaggaatggagacaaagatattcggagagacgcacagaaaagcgtactagaagaaaaaagg
cttttaaagagcgaaaggatgacattaaagccaaagccaagaatccactaaattgttggaaatgtggaag
gtttggtcatcgatataaagactgcaaggtcaaagagagaatcaaagctcttaatatagatgaagatctg
aaggattctttatgcaagatcttattaaattctgaaccagaaatatcggatccaggtgaggaaaatatcc
ctaccaaggaagaactaaataaagtcttccaggatgaaagtgatattgtcactagttctgaggatgaatg
tgatccttgcaagattggacaaccttgtgataataaagaaaatgatgaattttataatttattttctcaa
tttagtcttaacactttggataatgataatttaatagaattattgaagatggtacaggattccgaagtga
gatctaaacttatcgatcagatcaaaggaaatttcgaacagaaggaatccgaaggattgaaaaagggttt
ctatgcccctatggataatactccatacaccatgaaagaagttaaaagacttctagaacaaagaaaaatg
agtcatagtagttcacctactacgtctcaggatttggaaaaagaggtttataaccttaaacaggagattt
ataatctcaaaaaggaaaatagaatattagaacaaaggatttccagaattgaaggaaaaggtaaagaaaa
ggatgaaatccttaccccatctaattctgaaaatcttgataatatttttagcaatatgaatgaagagttt
ttaactaccctagaagtatttacttctcaaaagtttttataaaaattactcttcttattgattataatt
tcaaaaaagaatttatagctcttgttgatagtggagcagatttaaattgtgttcaagaaggattaattcc
ttcacgatattttcataaaactactcatagtttaaggtctgctaatggtgagaaaatgcagattgagtac
aaattgccaaaagcttctatatgcacaaataaggtatgtatacctcaaagctttgttttaataaaaaata
tgtctaaccaagttatacttggaacacccttttacaaaaaatatttcctattactaggattgataaaga
agggataataggaacttttcaagaaaaacctattatttttcagtttataactgaacctattactagagtt
ttaaaccagttaaaagatatgctagctaagaagagtcagcatatcaattttcttaaacaagaaatttata
ctttaaatatcgatgagatttttaaaaaatcctaaattgagggaaaaaattgattttatgataaatcagtt
ttcattacaaatttgtagtgaacatccaaatgctttttgggataggaaaaaacatgtagttacacttcca
tatgaggaagagttttctgaagctaatagtatccctactaaggcaagaccttgccagatgaattcagaat
atttggaattatgcaaaaaggaaatagattctttattgcataaaggtctcatacgatcatcgaaatcacc
atggtcttgtacagcgttctatgttaataaacatgctgaacaagaaagaggagttccaaggttagtaata
aattataaacctttgaataaagttttgaaatggatcaggtatccaattcctaataaaaaggatttattgg
ataggcttcataatgctattatatttttctaaatttgatttaaaatcaggatactggcagattcaaatagc
tgaagctgacaggtataaaactgcttttaatgtcccaataggacagttcgaatggaatgttatgccattt
ggattaaaaaatgctccttctgaatttcagaaaattatgaatgatatattcattccttatagtaatttta
ttatcgtttatattgacgatatattagtattttcaaatactattgaaatgcattttaagcatttagaatt
atttaaaaaaattattatacaaaatggtttggttatatctaaacctaaaatgtttttgtttcaaactaaa
ataagatttcttggccataatattgaaaaggggaaaattattcctattaatagaagtatagaatttgctt
ctaaatttccagatattataacagataaaactcaattacagagatttctaggaagtttaaattatattgc
tccattttatgaaaatcttgctaaagatacagctatcttatatgatagattgaaaaagattccaaagcca
tggactgatgctcatactaaaacagtccagaagatcaaggaaagagtaaaaaatctcccttgtcttacat
tggctaatcctacatggccaaaaataatagaaactgatgcttctgacattggttatggaggtattttaaa
gcaatactctccagaagaaaaaattgaatatttgatacaatttcattctggtaaatggaataatagccag
aaaaattatgcaactgtagcaaaagaaattcttgctatagtaaaatgtgttttaaaatttcaaggtgatt
tatacaatcagaagtttattataaagactgattgccaatcagctaaatttatgtttaataaagattgtaa
acatgatgtttctaaacaaatgtttgctagatggcaggcactattagccccatttgattttgaaatacat
tataaaaaaggagaagataatagccttcctgatttttaactagagaatatttggctagttaatgataat

aatcattattttaatgtgcaggttatttaggatgtgcagaagtaaaaatatgctctattattgtccctac
tgccagcaccagattatatcaagacaaatgagaatattatattgggatctgcttaatacaaaaaagaaat
taaaatactggaaaaatgaatccaaggatttgttatcacatgctgatcacttttataaaggagaaataga
tcatatcataagtatcctgaaaaggactattgatagaatgaaagaaataaatacggacataataggaatc
cgtaaagtattggaaacatatccaatgtatttagcctgacaaaattgtgatttatcgaatgcaggacatg
caccctggtcacggaaacccaacctctaatagaggaaggggaaacagaggaagaggcactggtagaggtg
ctgtaatagcccaaatgggtaagcaaaggcttatagccgaagacttaacaggatcatcttcaaagaataa
taaaaaggatgatactatgtcataccaacagtatcaagaattcctgaaatggaaaaattctcaagaaaag
caaaagggaacagaggataatatgcctccatcatattcttctattataatagaagatgataatgaaaata
tccagtcttatattaagacagattacaaggagttaatattattccttgaagatagagaccttcaatggaa
ggatgagccatggaaattaatggcaaaatatctgaataatacatcatacccaagctgtacttataagtac
agagctcattatgagatgatactcaaaggcactgaatcttgcgaaatatcgcatttttatccagcaaata
caaaaaaagtttataacttttccaaaatgataatcaaacggattatatccgcagaagaatggggtttaag
taccctgactgaaagaagttataatcatccagaaatgaaaacaccggtaaaatacaattactgggattat
gtggatagcttttataaagcattattatatgagaatgataacaagaaacattcttggtttataaaaatat
gctcaaatgtttataaacagagcaatataccaaattggttttcacaatggtggatcacttatggtcctac
cgtaaaaatattaccagaaccattgctcagtttatatactgattgggtggatgtatctccaaaaattcaa
gaattacaaagaaataatactcttattgaaggaagttcttatatgaatttctttattgaatattcaattc
catggataatgaaatggagtgttgaagtggattatactccacataatattcctgctttatacaggacctt
ttatacaaggttttggaaaaaaatgatccagaaaacagctgatggtaatctgcagtgccaggatacaatt
gatttaataaaacaatcactggaagaatataaacagattgatcatatgaagcaacttaatacaccaagtc
cttatcaacatataattcataaattaaagatgaagaagggactaatgacaaaagaagaaatgttagcttc
ttatttggaagaagtaaaaaaagatttaataaaaaatcttgaagttgaaattcagtcagacacatcaatg
gcttctgcaaataatacagaagatgaaggtgaatattgtctgccaggagaatcccagtcaatcacagaag
aagaagtaaatgattatatcacaaatatcataacacaggtgaataaagaagaatcttctatattacaaag
caccaaagggaaagataagatctgacgcggatgcattattgaagattccctcaactttcataaagtacct
tttataaagcagctttgataaagttgtaatgggccatctatcggccatttttctttctttaattttcaac
cggccacttgtaaagtttgggccatattagttgttttattctttgaaatttttctataaatagggtagga
gtttgagtagagaaggggagacaacccctcactctcattgtaaatctccccattgtaaaaatctttcagt
ttataaaaattctctattgcctgagcgcaaactaaggtattcataattgttccttattttgtttaactt
atataattttaaatattgc

Appendix 3.11 Approved sequence of *Caulimovirus-PPa*, the *Caulimovirus* member in *P. axillaris* subsp *parodii* S7 by the Repbase database.



Institut für Epidemiologie und Pathogendiagnostik
Institute for Epidemiology and Pathogen Diagnostics

Julius Kühn-Institut
Bundesforschungsinstitut für Kulturpflanzen

JKI, Messeweg 11-12, 38104 Braunschweig, Germany

Department of Genetics
University of Leicester
Leicester LE1 7RH UK

Bearbeiter/-in: Dr. K. Richert-Pöggeler
Fon: (0531) 299 3730
Fax: (0531) 299 3006
E-Mail: katja.richert-poeggeler@julius-kuehn.de
Ihr AZ:
Unser AZ:
Datum: 20 October 2016

Certificate of scientific visit at Julius Kühn-Institut (JKI) from mid-July to mid-October 2016

To whom it may concern

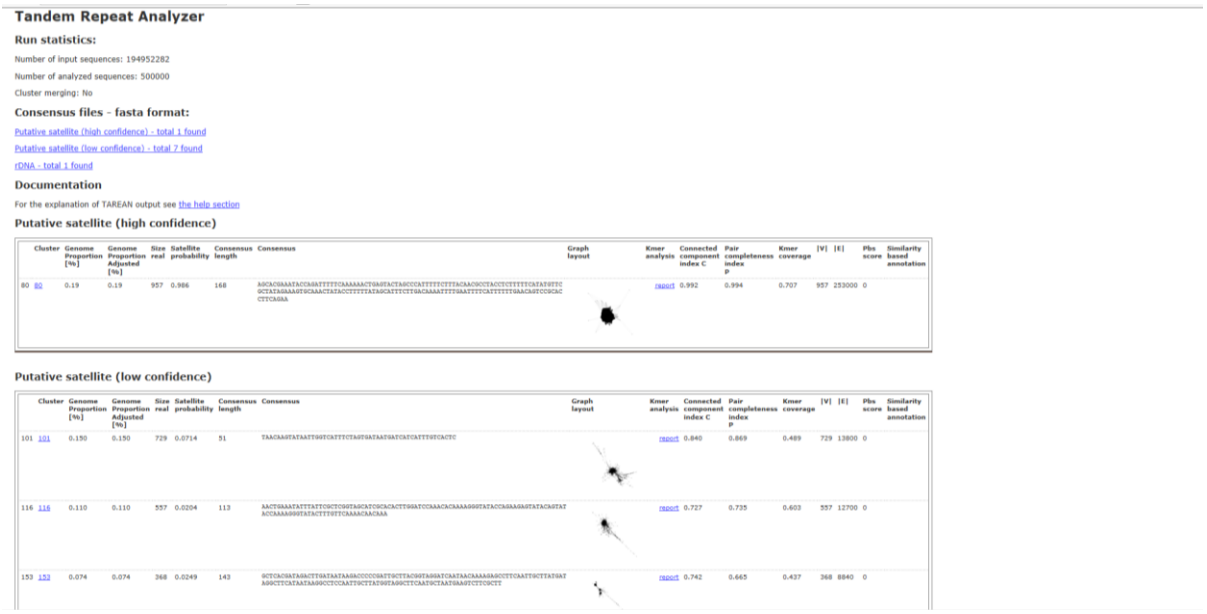
This is to inform you that the three-month visit of **Mr. Osamah Alisawi** at Julius Kühn-Institut in Braunschweig, Germany, started on 19 July 2016. During his stay he conducted doctoral studies within a cooperative research project with Dr. Trude Schwarzacher and Professor Dr. Pat Heslop-Harrison, University of Leicester, Leicester, UK.
Mr. Alisawi learned and applied successfully various techniques for pararetrovirus characterization using electron microscopy. The employed methods comprised dip preparation, immuno electron microscopy including immunogold labeling on homogenized plant material as well as embedded - cellular structures preserving - plant material. Besides his activities in the laboratory applying molecular and microscopic techniques, he presented the obtained results on various occasions. He gave an oral presentation "Virus integration and repetitive elements in genomes of Solanaceae" in frame of the weekly institute colloquium for scientists and Ph.D. students at which his supervisors Dr. Trude Schwarzacher, University of Leicester and Dr. Katja Richert-Pöggeler (JKI) were present. He reported on his scientific background and PhD project during a colloquium organized by JKI Ph.D. students and Post docs to exchange and discuss their projects and related scientific matters. Mr. Alisawi attended the international conference "15th World Petunia Days" in Lutherstadt Wittenberg, Germany, in September 2016. He was involved in conducting the meeting in Wittenberg, presented a poster on "DNA virus integration and repetitive elements in Petunia genomes" and co-authored the talk given by Dr. Trude Schwarzacher "Repetitive DNA landscape in Petunia".
Mr. Alisawi's interest in pararetroviruses as well as thorough knowledge of plant virology, his polite manners as well as his good laboratory practice and cheerful character made his visit a most pleasant experience for the whole team.

Dr. Thomas Kühne
- Director and Professor,
Head of the Institute -

Appendix 4.1 certificate of scientific visit at Julius Kühn-Institut (JKI), Braunschweig, Germany for three months.

Appendix 4.2 Embedding steps of selected samples that embedded in freshly prepared and added resin to half the volume of the labelled form and then polymerized overnight at 40°C in a drying oven (Memmert).

Appendix 4.3 The selected *Petunia* species (*P. axillaris* subsp *parodii* S7+PVCV), showing vein clearing symptoms maintained in Murashige skoog medium (MS) (tissue culture condition) for embedding later.



Tandem Repeat Analyzer

Run statistics:

Number of input sequences: 194952282

Number of analyzed sequences: 500000

Cluster merging: No

Consensus files - fasta format:

Putative satellite (high confidence) - total 1 found

Putative satellite (low confidence) - total 7 found

rDNA - total 1 found

Documentation

For the explanation of TAREAN output see the help section

Putative satellite (high confidence)

	Cluster	Genome Proportion [%]	Genome Proportion Adjusted [%]	Size real	Satellite probability	Consensus length	Consensus	Graph layout	Kmer analysis	Connected component index C	Pair completeness index P	Kmer coverage	|V|	|E|	Pbs score	Similarity based annotation
80	80	0.19	0.19	957	0.986	168	AGCACGAAATACCAGATTTTTCAAAAAACTGAGTACTAGCCCATTTTCTTTACAACGCTACCTCTTTTTCATATGTTC GCTATAGAAAGTGCAAACTATACCTTTTATAGCATTTCTTGACAAAATTTGAATTTCATTTTTGAACAGTCCGCAC CTTCAGAA		report	0.992	0.994	0.707	957	253000	0	

Putative satellite (low confidence)

	Cluster	Genome Proportion [%]	Genome Proportion Adjusted [%]	Size real	Satellite probability	Consensus length	Consensus	Graph layout	Kmer analysis	Connected component index C	Pair completeness index P	Kmer coverage	|V|	|E|	Pbs score	Similarity based annotation
101	101	0.150	0.150	729	0.0714	51	TAACAAGTATAAATTGGTCATTTCTAGTGATAATGATCATCATTTGTCACTC		report	0.840	0.869	0.489	729	13800	0	
116	116	0.110	0.110	557	0.0204	113	AACTGAAAATATTTATTCGCTCGGTAGCATCGCACACTTGGATCCAAACACAAAAGGGTATACGAAAGAGTATACAGTAT ACCAAAAGGGTATATACTTTGTTCAAAACAACAAA		report	0.727	0.735	0.603	557	12700	0	
153	153	0.074	0.074	368	0.0249	143	GCTCACGATAAACTTGATAATAAGACCCCCGATTGCTTACGGTAGGATCAATAACAAAAGACCTTCAATTGCTTATGAT AGGCTTCATAATAAGGCTCCAATTGCTTATGATAGCTTCAATGCTAATGAAGTCTTCGCTT		report	0.742	0.665	0.437	368	8840	0	

Appendix 5.1 The first page of tandem repeat analyzer (TAREAN) report showing a list of tandem repeat clusters with genome proportions and graphs.