Studies on the spatial and temporal evolution of *Turnip mosaic virus*

（カブモザイクウイルスの空間的・時間的進化に関する研究）

HUY DUC NGUYEN

2013

Studies on the spatial and temporal evolution of *Turnip mosaic virus*

（カブモザイクウイルスの空間的・時間的進化に関する研究）

By

Huy Duc Nguyen

A thesis submitted to the Faculty of Agriculture, The United Graduate School of Agricultural Sciences, Kagoshima University in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

In

Agricultural Science

2013

CONTENT

Page

I. SUMMARY

Studies on the spatial and temporal evolution have been reported for animal and plant viruses, however little information was known for those of potyviruses. *Turnip mosaic virus* (TuMV) is a species of the genus *Potyvirus* in the family *Potyviridae*. TuMV is probably the most widespread and damaging virus that infects cultivated brassicas worldwide. This study focused on the spatial and temporal evolution of TuMV populations.

Spatial evolution was analysed using 30 Vietnamese and 105 worldwide isolates of TuMV. The Vietnamese isolates were collected from *Brassica* and *Raphanus* plants during 2006-2008. Sequence-based phylogenetic and population genetic analyses were made of the complete polyprotein coding sequences, and of four non-recombinogenic regions of those sequences [the helper component-proteinase protein (HC-Pro), protein 3 (P3), nuclear inclusion b protein (NIb) and coat protein (CP)]. These were used to assess the subpopulation differentiation and divergence between Vietnamese TuMV populations and those of China and Japan. Nine inter- and intralineage recombination type patterns were identified in the genomes of the Vietnamese isolates, of which seven were novel. All the Vietnamese non-recombinant isolates fell into the world-B group and clustered with Chinese isolates. The estimates of genetic differentiation and gene flow revealed that the TuMV populations of Vietnam, China and Japan were genetically linked but have clear local founder effects. These results show for the first population genetic study of TuMV in Southeast Asia.

Temporal evolution was analysed using 155 isolates of TuMV collected mostly from Brassicaceae during 1968-2007 from worldwide. A sister lineage from European wild orchids (Orchis isolates) to the brassica-infecting TuMV (BI isolates) was identified. Extensive host-range tests showed that all of the Orchis isolates were biologically similar, but distinct from

BI isolates and did not readily infect brassicas. Bayesian coalescent analyses were applied for analyses of temporal evolution using a combination of novel and published sequences data from three TuMV protein coding regions; HC-Pro, P3 and NIb. The dating analyses of those coding regions indicated that the BI isolates diverged from Orchis group isolates around 1000 years ago. Only 150 years later, the four lineages (basal-B, basal-BR, Asian-BR and world-B groups) of the present global population of BI isolates diverged from one another. These dates are congruent with historical records of the spread of agriculture in West Europe.

II. INTRODUCTION

The Brassicaceae (Cruciferae) family, including many economically important vegetables such as broccoli, cabbage, cauliflower, Chinese cabbage, kale, kohlrabi, mustard, radish and turnip were grown widely around the world. Brassicaceae vegetables are highly regarded for their nutritional value. They provide large amounts of vitamin C which humans could not synthesize it and must obtain through their daily diet.

During growing season, the low quality and productivity of Brassicaceae plants could be caused by several abiotic and biotic stresses. Brassicaceae vegetables are susceptible to a large number of insects and pathogens, such as root rot disease, bacterial leaf blight and viruses. Among the viruses, *Turnip mosaic virus* (TuMV) was ranked second only to *Cucumber mosaic virus* (CMV) as the most important virus infecting filed-grown vegetables in a survey of virus disease in 28 countries and regions (Tomlinson, 1987).

Studies of the genetic structure of populations of plant RNA viruses are important for understanding the evolution of virus/host interactions (García-Arenal *et al*., 2001; Gibbs *et al*., 2008b; Gibbs and Ohshima, 2010), because plant RNA viruses are very variable enabling them to adapt rapidly to new or resistant hosts (Tsompana *et al*., 2005; Ohshima *et al*., 2010). There are several reports on the genetic structures of potyvirus populations, notably those on *Potato virus Y* (PVY) (Karasev *et al*., 2011; Ogawa *et al*., 2008, 2012), *Soybean mosaic virus* (SMV) (Seo *et al*., 2009), *Tobacco vein banding mosaic virus* (TVBMV) (Zhang *et al*., 2011), *Zucchini yellow mosaic virus* (ZYMV) (Lecoq *et al*., 2009) and TuMV (Ohshima *et al*., 2002, 2009; Tomimura *et al*., 2004; Tomitaka and Ohshima, 2006). These reports showed that virus populations had been shaped by selections, founder effects and genetic recombination.

The possibility of controlling a pathogen is improved if we know when, where, and how it first became established in the host population, namely its 'centre of emergence'. This is analogous to the 'centre of diversity' of crop species (Harlan, 1998; Vavilov, 1926). It is valuable to identify the centre because it may still contain the pathogen and host populations most closely related to those involved in the emergence. Therefore, these populations might have been interacting with the pathogen longer than others, leading to the greatest diversity in the genes controlling that interaction. As a consequence, such populations might be useful in the design of gene-based control strategies. The previous phylogeographic studies of TuMV showed that this virus was probably the world's most widespread and damaging virus of domesticated species of the family Brassicaceae, both crop and ornamental (Ohshima *et al*, 2002; 2007; Tomimura *et al*, 2003; 2004). These studies clearly indicated that present-day TuMV populations came from a founder population in western Eurasia, namely Europe, Asia Minor, and the Middle East. However, the source virus, source populations and the timing of that emergence remained unknown.

There are many previous studies on the biological chracteristics and molecular evolution of TuMV (Ohshima *et al*., 2002; Tomimura *et al*., 2003; 2004; Tan *et al*., 2004; 2005; Tomitaka and Ohshima, 2006; Tomitaka *et al*., 2007; Farzadfar *et al*., 2008). Recombination and mutation were the two major driving forces for the evolution of this virus (Tan *et al*., 2004; 2005). In this study, the analyses of spatial and temporal evolution of TuMV population were conducted. Spatial evolution was analysed using the populations of Vietnam and East Asian countries, while temporal evolution was analysed using the populations from worldwide collected for 39 years. Parts of studies written in this thesis were already published in Virus Research (Nguyen *et al*., 2013) and PLoS ONE (Nguyen *et al*., 2013).

III. LITERATURE REVIEW

1. Evolution of animal viruses

1.1. *Human immunodeficiency virus*

Acquired immune deficiency syndrome (AIDS) is one of the most important diseases of the human immune system (Leeper and Reddi, 2010; UNAIDS, 2010). The disease caused by human immunodeficiency virus (HIV) (Barre-Sinoussi *et al*., 1983; Gallo *et al*., 1983; Popovic *et al*., 1984; Sarngadharan *et al*., 1984; Sepkowitz, 2001). The virus is a global pandemic (Cohen *et al*., 2008). There was an average of 33.4 million people who were living with HIV in the world up to 2011, two million infected people died as reported by World Health Organization (WHO), and more than 6.6 million people infected in low and middle income countries as of 2010 (UNAIDS, 2010). HIV is a member of genus *Lentivirus* which belongs to subfamily *Lentivirinae* in the family *Retroviridae* (Chiu *et al*., 1985; Wain-Hobson *et al*., 1985; Vogt, 1987). The genomic structure of HIV is positive sense single-stranded RNA, and genome is about 9.7 kb in length. HIV is transmitted by three main routes: sexual, body-fluids and mother-to-child. HIV was found that originated in non-human primates in Sub-Saharan Africa. The virus was infected humans during the late 19th or early 20th century. HIV was divided into two types HIV-1 and HIV-2. HIV-1 was introduced into human populations in the early 20th century following multiple transmissions of a chimpanzee virus (SIVcpz) (Gao *et al*., 1999; Worobey *et al*., 2008), whereas HIV-2 resulted from transmissions of a virus found in sooty mangabeys (SIVsm) (Hirsch *et al*., 1989). The age of SIV was estimated using phylogenetics produced widely difference, but all relate to the current (Wertheim and Worobey, 2009; Sharp *et al*., 2000). The discovery of a full-length

endogenous SIV in the genomes of lemurs revealed that lentiviruses were present in prosimians approximately four million years ago (MYA) (Gifford *et al*., 2008; Gilbert *et al*., 2009). Moreover, a date to calibrate phylogenetics of SIV in Old World Monkey (OWM) was made with the identification of SIV strains endemic disease to the African island of Bioko. Each virus discovered on this island shares ancestor with a mainland virus. Their respective hosts belong to the same genus, revealed that lentiviruses infected OWM for tens of thousands of years ago (Worobey *et al*., 2010).

HIV-1 is more virulent than HIV-2. The virus is easily transmitted and is the cause of the huge majority of HIV infections globally (Reeves and Doms, 2002). The origin of HIV-1 found in the subspecies *Pan troglodytes troglodytes* of the chimpanzees (Gao *et al*., 1999), which lives in the forests of the Central African nations of Cameroon, Equatorial Guinea, Gabon, Republic of Congo (or Congo-Brazzaville), and Central African Republic. HIV-1 infection continues to spread rapidly world-wide. HIV-1 is characterized by high genetic variability, diversification and rapidly evolution (Hahn *et al*., 1984; Seillier-Moiseiwitsch *et al*., 1994). Actually, recombination combined with the high error rate of the reverse transcriptase, and the rapid turnover of HIV-1 in infected individuals, were at the origin of the high genetic variability of the virus (Peeters and Sharp, 2000). Futhermore, studies origin and timescale of HIV-1 subtype B showed that the viuses circulating in South America belong to the 'pandemic' clade that migrated out of Haiti around 1969 and spread through the word (Gilbert *et al*., 2007). HIV-2 is less virulent and is largely confined to West Africa (Reeves and Doms, 2002; Santiago *et al*, 2005). This virus is less transmittable and is largely confined to West Africa, along with its closest relative, a virus of the sooty mangabey (*Cercocebus atys atys*), an Old World monkey inhabiting southern Senegal, Guinea-Bissau, Guinea, Sierra Leone, Liberia, and western Ivory Coast (Reeves and Doms, 2002; Santiago *et al*, 2005). In

Guinea-Bissau, HIV-2 subtype A strains was diverged approximately 1940 ± 16, where as B strains was diverged around 1945 ± 14. The epidemic history of HIV-2 subtype A show a transition from constant size to rapid exponential growth around 1955 – 1970 (Lemey *et al*., 2003).

1.2. *Influenza virus*

Influenza is an infectious disease of birds and mammals. The disease is commonly known as 'the flu'. Influenza virus belongs to RNA viruses and makes up three of five genera in the family *Orthomyxoviridae* (Kawaoka, 2006); *Influenzavirus A*, *Influenzavirus B* and *Influenzavirus C*, *Isavirus* and *Thogotovirus*, whereas *Influenzaviruses B* and *Influenzavirus C* are less common than *Influenzavirus A*. The genus *Influenzavirus A* has one species, *Influenza A virus*. This species could be subdivided into different subtypes based on the antibody response to these viruses (Hay *et al*., 2001). The different serotypes are: H1N1 (Spanish flu in 1918, swine flu in 2009), H2N2 (Asian flu in 1957), H3N2 (Hong Kong flu in 1968), H5N1(bird flu in 2004), H7N7 (unusual zoonotic potential) (Fouchier *et al*, 2004), H1N2 (endemic in humans, pigs and birds), H9N2, H7N2, H7N3, H10N7 and H7N9. The A/H1N1pdm influenza virus has spread rapidly in humans, with over 5,700 human deaths. H1N1pdm was central to understanding the evolution and spatial spread of the current pandemic, and to predict its future impact on human populations. H1N1pdm had already diversified into distinct viral lineages with defined spatial patterns (Nelson *et al*., 2009).

Phylogenetic analysis can help to determine past viruses and their patterns as well as determining a common ancestor of the virus. Previous studies showed that an avian virus spread to pigs and then to humans approximately 100 years ago (Scholtissek, 1995). This resulted in human lineages further evolving and becoming more prominent and stable

(Scholtissek, 1995). Analysis can also feature relationships between species. The 1918 Spanish influenza virus demonstrated this. The hemagglutinin (HA) gene of the 1918 pandemic virus was closer in sequence to avian strains than other mammalian ones. Despite this genetic similarity, it was obviously a mammalian virus (Reid *et al*., 1999). The gene might have been adapting in humans even prior to 1918 (Reid *et al*., 1999). Breaking down the phylogenetic history of the influenza virus showed that there was a common ancestor that reaches back before the 1918 outbreak that links the current human virus to the swine virus (Gorman *et al*., 1990). The ancestor was derived from an avian host (Scholtissek, 1995).

The genus *Influenzavirus B* has one species, *Influenza B virus*. *Influenza B virus* almost exclusively infects humans. Seal and the ferret are known to be susceptible to *Influenza B virus*. The genus *Influenzavirus C* has one species, *Influenza C virus*. This species infected humans, dogs and pigs. Estimation of amino acid substitution rate for A virus HAs from duck was 3.19×10^{-4} subs/site/year, and slower than that for human and swine A virus HAs. However, substitution rate was similar to that for influenza B and C virus HAs (HEs) (Suzuki and Nei, 2002). The estimation of the time to the most recent common ancestor (TMRCA) was estimated to be about 2,000 years ago. The A virus HA gene diverged from the B virus HA gene about 4,000 years ago and from the C virus HE gene about 8,000 years ago. Recent studies on the evolution dynamics of local pandemic H1N1/2009 influenza virus revealed that the divergences time of this virus in United Kingdom clusters was around April-June 2009 (Baillie *et al*., 2012). Study on the antigenic evolution of influenza showed that several epidemiological features of influenza and its serological and molecular profiles are consistent with this model of 'antigenic thrift', and that identifying the protective epitopes of low variability predicted by a theoretical model (Wikramaratna *et al*., 2013).

1.3. *Hepatitis virus*

Liver inflammation caused by Hepatitis viruses included *Hepatitis A*, *Hepatitis B*, *Hepatitis C*, *Hepatitis D,* and *Hepatitis E*. Among these viruses, *Hepatitis B* and *C* virus are common which were studied for evolution deeply.

Hepatitis B virus (HBV) is a member of the genus *Orthohepadnavirus* in the family of *Hepadnaviridae* (Hunt, 2007). The genome was made of circular DNA. This virus caused the disease hepatitis B (Hassan *et al*., 2008). Like all other viruses, study on the early evolution of the *Hepatitis B* was difficult to establish. The divergence time of *Orthohepadnavirus* and *Avihepadnavirus* occurred about 125,000 years ago (van Hemert *et al*., 2011) or 25,000 years ago (van Hemert *et al*., 2011). Human strains had a most recent common ancestor dating back to 7,000 – 10,000 years ago. The evolution rate of nonsynonymous mutations in this virus was estimated to be about 2×10^{-5} amino acid replacements site/per/year, whereas the mean of evolution rate was about 7.9×10^{-5} nucleotide subs/site/year (Osiowy *et al*., 2006). Furthermore, the estimation of the origin of this virus suggested a TMRCA of the human strains evolved about 1500 years ago (Zhou and Holmes, 2007) while TMRCA of the avian strains was about 6000 years ago. The mutation rate was estimated to be 10^{-6} subs/site/year. The newest study on the evolution of HBV suggested that HBV jumped into humans around 33,600 years ago with mean of substitution rate was about 2.2×10^{-6} subs/site/year. HBV had been co-expanding and co-migrating with human populations for the last 40,000 years (Paraskevis *et al*., 2013).

Hepatitis C virus (HCV) belongs to the family *Flaviviridae* and had enveloped, positive-sense single-stranded RNA. HCV is about 55 – 65 nm in size. The genome of HCV, like all RNA viruses, evolves rapidly during replication. This high mutation rate was thought in part to

explain its persistence in the human host. HCV was classified into seven genotypes, and several subtypes (Ohno *et al*., 2007; Nakano and Tatsunori, 2011), whereas subtypes 1a and 1b were found worldwide. The earlier studies of origin of HCV suggested that it might have evolved in South East Asia and was spread to West Africa by traders from Western Europe (Salemi and Vandamme, 2002). Genotypes 1 and 4 appeared to share a common origin (Sarwar *et al*., 2011). Using a Bayesian analysis indicated that the major genotypes diverged about 300 – 400 years ago from the ancestor virus (Pybus *et al*., 2009). The minor genotypes diverged about 200 years ago from their major genotypes. All of the extant genotypes appeared to have evolved from genotype 1 subtype 1b. A study of genotype 6 strains suggested an earlier date of evolution 1,100 – 1,350 years before the present (Simmonds and Smith, 1997). The estimated rate of mutation was 1.8×10^{-4}. This genotype might be the ancestor of the other genotypes (Simmonds and Smith, 1997).

A study of European, USA and Japanese isolates suggested that the date of origin of genotype 1b was about 1925 (Magiorkinis *et al*., 2009). The estimated dates of origin of types 2a and 3a were 1917 and 1943 respectively. The time of divergence of types 1a and 1b was estimated to be 200 – 300 years.

1.4. *Rabies virus*

Rabies virus (RABV) is one of the most virulent diseases of humans and animals. RABV was reported in the Old World before 2300 BC (Steele and Fernandez, 1991; Theodorides, 1986). The virus is a negative single-stranded with a genome size of about 12 kb in length. The genus *Lyssavirus* included a number of important zoonotic bat viruses; phylogenetic analyses had defined both the evolutionary relationships among these lyssaviruses as well as the existence of seven distinct genotypes, although this number was likely to increase with

more intensive sampling (Bourhy *et al.*, 1993; Gould *et al.*, 1998; Kuzmin *et al.*, 2005). As a group, the lyssaviruses were distinguished by their ecological association with specific mammalian species, which act as vectors for their transmission, such that a number of phylogenetic lineages co-circulate among a range of mammalian hosts (Bourhy *et al*., 1999, Davis *et al.*, 2005; Holmes *et al.*, 2002; Kissi *et al.*, 1995).

Bayesian coalescent approaches and phylogenetic were employed to reveal both the timescale of RABV evolution in dogs and, for the first time, to explore the global phylogeography of this important human and wildlife pathogen (Bourhy *et al*., 2008).

In earlier study, rates of population growth and nucleotide substitution were approximately 2.5 – 4 x 10^{-4} substitutions/site/year in both bats and terrestrial mammals (Davis *et al*., 2006).

There were seven genotypes of rabies virus. In Eurasia cases are due to three of these genotype 1 (classical rabies) and to a lesser extent genotypes 5 and 6 (European bat lyssaviruses type-1 and -2) (McElhinney *et al*., 2006). Genotype 1 evolved in Europe in the 17th century and spread to Asia, Africa and the Americas as a result of European exploration and colonization (Kuzmina *et al*., 2013). Studies of spatial dynamics using 62 sequences (Bourhy *el al*., 2008) and 192 isolates (GenBank database) of dog RABV from 55 countries during 37 collection years revealed that a strong population subdivision by geographical region. Bayesian coalescent approach for temporal dynamics reported that the mean rate of nucleotide substitution estimated for N and G genes were 2.3 x 10^{-4} and 3.9 x 10^{-4} subs/site/year (Hervé *el al*., 2008). Morever, the TMRCAs of dog RABV originated within the past 1500 years (Hervé *el al*., 2008). Phylogenetic analysis of a collection of rabies viruses that currently circulate in Canadian big brown bats (*Eptesicus fuscus*) identified five distinct lineages which emerged from a common ancestor that existed over 400 years ago (Nadin-

Davis and Real, 2011). To investigate the evolutionary dynamics of RABV in western and central Africa, 92 isolates sampled from 27 African countries over 29 years were collected and sequenced. This revealed that RABV currently circulating in dogs in this region fell into a single lineage designated 'Africa 2'. A detailed analysis of the phylogeographical structure of this Africa 2 lineage revealed strong population subdivision at the country level, with only limited movement of virus among localities, including a possible east-to-west spread across Africa. In addition, Bayesian coalescent analysis suggested that the Africa 2 lineage was introduced into this region of Africa only recently (approximately 200 years ago), in accordance with the timescale of expanding European colonial influence and urbanization, and then spread relatively slowly, perhaps occupying the entire region in a 100 year period (Talbi *et al*., 2009).

Further evolutionary history of RABV was estimated using a Bayesian Markov chain Monte Carlo method to understand the temporal and spatial dynamics of this virus. Results showed that rabies viruses in China and Southeast Asia share a common ancestor and form 2 clades with each being further divided into 3 lineages. The time of the most recent common ancestor of current RABV strains was estimated to be year 1654 (1514–1812) and the viruses circulating in Southeast Asia likely derived from China (Gong *et al*., 2010). Furthermore the rabies virus in China was primarily defined by two clades that exhibit distinct population subdivision and translocation patterns and that contributed to the epidemic in different ways. The younger clade originated around 1992 and had properties that closely match the observed spread of the recent epidemic. The older clade originated around 1960 and had a dispersion pattern that suggested it represents a strain associated with a previous outbreak that remained at low levels throughout the country and reemerged in the current epidemic (Yu *et al*., 2012).

2. Evolution of plant viruses

2.1. *Rice yellow mottle virus*

Rice yellow mottle virus (RYMV) was first reported in Kenya (Bakker, 1974), subsequently in many western and eastern African countries (Hull, 1988). It has been detected in many countries in the sub-Saharan Africa (Abo *et al*., 1997). RYMV is a member of genus *Sobemovirus*, of which *Southern bean mosaic virus* (SBMV) is the type member (Matthews, 1982). The genome is a single-stranded positive-sense RNA. The 5' terminus of the RNA has a genome-linked protein (VPg) and the 3' end is not polyadenylated (Sehgal, 1981; Francki *et al*., 1985; Hull, 1988). The complete nucleotide sequence of RYMV showed that the virus had 4450 nucleotides in length. Similarity of genome organization and sequence comparisons of the proteins predicted to be encoded by RYMV with those of SBMV and other plant RNA viruses confirm that RYMV is a member of the sobemovirus group and extend the knowledge of sobemovirus sequences (Ngon A Yassi *et al*., 1994). Phylogeographic studies suggested that RYMV diverged in East Africa and radiated in West/Central Africa (Abubakar *et al*., 2003; Fargette *et al*., 2004). RYMV originated and evolved in wild poaceous species and only recently infected cultivated rice (Fargette *et al*., 2004). Furthermore, studies on the rate of evolution of RYMV showed that the overall and synonymous evolution rates of RYMV were within the ranges of those of RNA animal viruses and evolves as rapidly as most RNA animal viruses (Fagette *et al*., 2008). TMRCA of RYMV using 253 dated CP sequences showed that the average substitution rates ranged from 5.1 x 10^{-4} to 12.3 x 10^{-4} nucleotide/site/year, and the divergence time of RYMV in East Africa was approximately 200 years ago (Fargette *et al*., 2008), whereas the symptom was found for the first time in 1966 in East Africa (Bakker, 1974) and in 1975 in West Africa (Fauquet and

Thouvenel, 1977). A recent study on the evolutionary time scale of RYMV in Madagascar and in the Zanzibar Archipelago showed that the virus predated by ca. 5-10 years the first filed detection in 1989 (Rakotomalala *et al*., 2013).

2.2. *Tomato yellow leaf curl virus*

Tomato yellow leaf curl virus (TYLCV) causes severe damage in tomato crops in both tropical and subtropical regions. TYLCV infected tomato crops worldwide (Czosnek and Laterrot, 1997; Moriones and Navas-Castillo, 2000). TYLCV belongs to the genus *Begomovirus* in the family *Geminiviridae.* The viral genomics has paired particles and circular, single-stranded monopartite DNA genomes of about 2.7–2.8 (Kheyr-Pour *et al.*, 1991; Navot *et al.*, 1991). The viruses are considered as emerging viruses because they spread rapidly to new areas, partially as a result of the expansion of their natural vector, *B. tabaci* (Brown et al., 1995).

Some of plant viruses, belonging to the genus *Begomovirus*, are well known for their economic importance (Pico *et al*., 1996; Moriones and Navas-Castillo, 2000; Seal *et al*., 2006) and their high propensity to recombine (Lefeuvre *et al*., 2007; Padidam *et al*., 1999). Particularly notorious amongst these viruses is the highly invasive TYLCV, which is transmitted-like all begomoviruses–in a circulative persistent manner by the whitefly *Bemisia tabaci* Genn (Moriones and Navas-Castillo, 2000).

Previous studies reported that begomoviruses were divided into two groups; Old-World (OW) and New-World. The viruses perhaps presented in the Americas at least 1.9 million years ago (MYA), and substitution rate of the begomovirus CP gene was lower than 6.0×10^{-7} subs/site/year (Gibbs *et al*., 2010). Morever, the high evolutionary rate of the

geminiviruses was not primarily due to frequent recombination and might explain their ability to emerge in novel hosts (Duffy and Holmes, 2008).

During the past 30 years, TYLCV spread from its native geographical range in the Eastern Mediterranean and the Middle East to numerous regions of the world following accidental introduction (Lefeuvre *et al*., 2010). Substitution rates of TYLCV were about 2.88 x 10^{-4} (full genome), 4.63 x 10^{-4} (CP) subs/site/year. TYLCVs probably arose somewhere in the Middle East between the 1930s and 1950s and then the global spread of TYLCV only began in the 1980s (Lefeuvre *et al*., 2010).

2.3. *Citrus tristeza virus*

Citrus tristeza virus (CTV) is serious disease of citrus trees. The virus can be spread rapidly and killed over 100 million trees of citrus-growing areas throughout the world (Moreno *et al*., 2008). CTV is a member of the genus *Closterovirus* in the family *Closteroviridae*. The genome of this virus is single-stranded, positive-sense RNA with 19.3 kb in length and encoding 12 open reading frames (Karasev *et al*., 1995). CTV is not transmitted through seed or pollen (Moreno *et al*., 2008), but transmitted by aphids. Studies on the evolution of CTV have been reported intensively at the molecular level during the 2000s. For instance, isolates with a common origin that were separated geographically for several decades depicted were very similar sequences (Albiach-Martí *et al*., 2000; Lbida *et al*., 2004). Furthermore, studies on the evolution of the two isolates p18 and p20 revealed no evidence of recombination playing a significant role during evolution in these isolates. Based on obtained results, the authors suggested that hosts could be an important evolutionary factor for CTV isolates (Sentandreu *et al*., 2006). In contrast, analyses of genetic variations, phylogenetic relationships (Martín *et al*., 2009) and comparative analysis of CTV codon usage

pattern (Cheng *et al*., 2012) suggested that negative selection, gene flow, sequence recombination, virulence or host plant codon usage selection might be important factors driving CTV evolution. In China, 11 CTV isolates from wild citrus belonged to different clusters in the phylogenetic tree. These isolates shared higher homology and closer relationships with other CTV isolates from cultivated citrus in different countries. However, the isolates showed similar biological characteristics usually located into the same clusters. Based on the results, pathogenicity was critical to evolution and origin of CTV (Yi *et al*., 2010). Recently, studies on the evolutionary rate showed that nucleotide changes were determined directly in the progeny of a CTV T36 isolate infectious clone, resulting in an evolutionary rate in the range $4 \times 10^{5} – 8 \times 10^{-5}$ nt/ site/year^{-1} (Weng *et al*., 2010). By using a Bayesian coalescent approach, the evolutionary rate of CP gene sequences of CTV was approximately 1.58×10^{-4} nt/site/year^{-1} based on the best-fitting evolutionary and population models. This indicated that this virus ranked among the rates of the slowest animal RNA viruses (Silva *et al*., 2012). It has also been reported that a very slow evolutionary rate of 6.67×10^{-5} nt/site/year of evolution in a single CTV isolates over the time under stable condition (Harper, 2013).

2.4. Potyviruses

2.4.1. Taxonomy

The genus *Potyvirus* belongs to family *Potyviridae* that consists of plant viruses with a single stranded, positive sense RNA genome and flexuous, filamentous particles. Genomes of the family *Potyviridae* have a VPg covalently linked to the 5’ end and the 3’ terminus has a poly(A) tail of about 200-250 adenosine residues. Genomes encode a large polyprotein that

is self-cleaved into a set of functional proteins. Gene order and protein sequences are conserved throughout the family.

Virions of *Potyvirus* are flexuous filaments without envelope and are 11–15 nm in diameter include a helical pitch of about 3.4 nm. Particle lengths of members of some of the six genera differ. Members of the genera *Potyvirus*, *Ipomovirus*, *Macluravirus*, *Rymovirus*, *Tritimovirus*, *Brambyvirus* and the unassigned viruses are monopartite with particle modal lengths of 650–900 nm; members of the genus *Bymovirus* are bipartite with particles of two modal lengths of 250–300 and 500–600 nm. The *Potyvirus* is the largest genus in the family contains viruses transmitted by aphids in a non-persistent manner (ICTV, 2011).

2.4.2. Genome organization and fuction

The genus *Potyvirus* genomes have a 5'untranslated region (UTR) which is linked to a genome-linked protein (VPg), a single open reading frame (ORF) flanked by terminal untranslated regions and a 3'UTR region terminating in polyA tail. The ORF is translated to a single large polyprotein, which is hydrolysed into at least ten proteins (Figure 1) by virus-encoded proteinases (Urcuqui-Inchima *et al.,* 2001); the first protein (P1), helper component protein (HC-Pro), the third protein (P3), 6K1 protein, cylindrical inclusion protein (CI), 6K2 protein, nuclear inclusion protein (NIa) including two domains of the VPg and protease (NIa-Pro), nuclear inclusion protein (NIb) and coat protein (CP)(Adams *et al*., 2005).

P1 protein is located at the beginning of the viral genome. The protein has been illustrated to be the most variable protein among potyviruses encoding proteins (Adams *et al*., 2005; Valli *et al*., 2007), and it is thus very representative of the evolutionary diversification that facilitated viral adaptation to a wide range of host species (Valli *et al*., 2007). P1 protein of

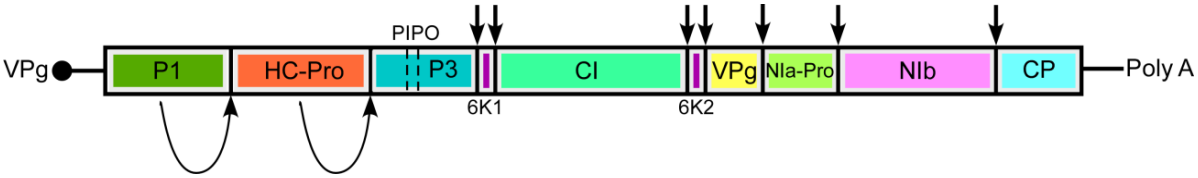


Figure 1. *Potyvirus* genome map. Arrows indicate cleavage sites of TuMV polyprotein. P1/HC-Pro cleavage site was cut by the P1 serine proteinase). HC-Pro/P3 cleavage site was cut by the HC-Pro cysteine proteinase. All other cleavage sites were cut by the NIa-Pro cysteine proteinase (modified from http://www.dpvweb.net/potycleavage/).

Tobacco vein mottling virus (TVMV) has been shown to interact with nucleic acids (Brantley and Hunt, 1993). The fusion of P1 and HC-Pro enhanced viral pathogenicity through suppression of posttranscriptional gene silencing (PTGS) in the host (Kasschau and Carrington, 1998). The 5'UTR-P1 region played a role for understanding the evolutionary history of *Leek yellow strip virus* (LYSV) isolates in garlic (Yoshida *et al*., 2012).

HC-Pro is a multifunctional protein required for viral acquisition by the vector, systemic, cell-to-cell movement and suppression of PTGS. HC-Pro could be divided into three regions, the N-terminal, central and C-terminal regions. HC-Pro was required for transmission through interaction with aphids and virions. Four aphid species *Myzus persicae*, *Lipaphis erysimi*, *Aphis gossypii* and *Myzus ascalonicus* were tested for their ability to transmit *Tobacco etch virus* (TEV) and TuMV. These aphids were acquired homologous virus/helper component mixtures through membranes. The results showed that *M. persicae* and *A. gossypii* transmitted both viruses efficiently from infected plants, whereas *L. erysimi* transmitted only TuMV and *M. ascalonicus* was a poor or non-transmitter of either virus (Wang *et al*., 1998). They repoted that the food canal of aphids differed in its ability to interact with HC-Pro, which, therefore, affected the ability of aphids to retain virions in the stylets. The other functions of HC-Pro were involved in systemic movement, cell-to-cell movement, and viral replication, suppression of gene silencing and proteinase activity. A mutation in the highly conserved CC/CS motif in the core region of the TEV HC-Pro was not capable of systemic movement. Systemic movement was restored, however, in transgenic plants provided with the intact HC-Pro (Cronin *et al*., 1995). The region responsible for activity of cell-to-cell movement was located in the C-terminal part of HC-Pro (Rojas *et al*., 1997).

P3 was shown to have a role in plant pathogenicity through interaction with other viral proteins. For example, the C-terminal region of the P3-6K1 complex carried a pathogenicity determinant of *Plum pox virus* (PPV) (Saenz *et al*., 2000). Similarly, a previous study showed that TuMV contained an important determinant in P3 C-terminal region, which conffered the ability of virus to infect different host (Suehiro *et al*., 2004; Tan *et al*., 2005).

PIPO, a frame-shift protein encoded by overlapping open reading frame pipo in the P3-encoding region called Pretty Interesting *Potyviridae* ORF, was first identified in an alignment of TuMV and other potyviruses (Chung *et al*., 2008; Wen and Hajimorad, 2010; Wei *et al*., 2010). This region has been reported to play roles in the infection and pathogenesis of potyviruses (Chung *et al*., 2008).

6K1 carried a determinant for the pathogenicity (Schaad *et al*., 1997). 6K1 protein encoded by cistrons that form the RNA replication module (Riechmann *et al.*, 1992). A study of 6K1 protein of the Pinellia isolate of SMV suggested that the potyvirus 6K1 protein played some role in viral cell-to-cell movement, but the lack of transmembrane domains suggested that it did not conform to currently-recognized patterns of viral movement proteins (Hong *et al.*, 2007).

CI is a major component of the replication complex. CI protein belongs to “super family 2” of helicase proteins that were characterized by seven conserved segments I, Ia, II, III, IV, V and VI (Kadare and Haenni, 1997). These fragments occupied the N-terminal half of the protein and had NTP binding, NTPase, RNA binding and RNA helicase activities (Fernandez and Garcia, 1996; Fernandez *et al*., 1995, 1997). CI was considered to be a major component of a multicomponent, membrane-associated replication complex of CI, VPg/NIa and NIb (Shukla *et al*., 1998). CI is involved in cell-to-cell movement. Although the CI was not movement protein (MP) in the same sense like CP and HC-Pro (Rojas *et al*., 1997), the

presence of ATPase activity in plasmodesmata of *Maize dwarf mosaic virus* (MDMV)-infected cell (Chen *et al*., 1994) suggested that cell-to-cell movement required energy released from ATP hydrolysis. Therefore, CI is the only virus-encoded protein that has ATPase activity and it may participate in this process. An analysis using alanine scanning mutagenesis based on the CI/TEV system supported a model in which CI interacts directly with plasmodesmata and CP-containing ribonucleprotein complex to facilities cell-to-cell movement (Carrington *et al*., 1998).

6K2 is required for genome amplification and that was anchors to the replication apparatus to the endoplasmic reticulum (Schaad *et al*., 1997). 6K2 protein of *Potato virus* A (PVA) affects viral long-distance movement and symptom induction independently and in a host-specific manner (Spetz and Valkonen, 2004). Futhermore, evidence for diversifying selection was obtained for the 6K2 protein of PVY, *Bean yellow mosaic virus* (BYMV) and *Yam mosaic virus* (YMV) (Moury *et al*, 2002). 6K2 genes with other proteins might have some involvement on host specificity and pathogenicity in Korean *Pepper mottle virus* (PepMoV) isolates (Kim *et al*, 2009).

The VPg domain of NIa has essential functions in viral replication and host genotype specificity. A study based on chimeric of TEV (Schaad *et al*., 1997) showed that VPg interacted either directly or indirectly with host components to facilitate long distance movement. The interaction between cellular factors with the VPg N-terminal region of a diverse range of potyviruses affected systemic symptoms involving both cell-to-cell movement and systemic movement in infected plants. The VPg interacted with plant translational initiation factors like eIF4E and eIF(iso)4E (Leonard *et al*., 2000; 2004; Wittmann *et al*., 1997). An avirulent determinant of VPg was identified based on observations that the resistance genes, at homozygous state, containing point mutants which interrupted the

interaction of eIF4E and VPg, created resistance phenotypes at different levels (Kang *et al*., 2005). The VPg interaction enhanced cap-independent translation by increasing the affinity of eIF4F for TEV RNA showed the first evidence of direct participation of VPg in translation initiation (Khan *et al*., 2008). Grzela *et al*. (2008) suggested that the VPg of PVY should be classified among intrinsically disordered proteins. Intrinsic disorder lied at the basis of VPg multifunctionality, which was necessary for virus survival in the host.

NIa is composed of two domains, the N-terminal VPg domain, and the C-terminal proteinase domain; these two are VPg and NIa-Pro. NIa is the major proteinase of potyviruses (Adam *et al*., 2005; Shukla *et al*., 1998; Fellers *et al*., 1998).

The function of NIb protein is an RNA dependent RNA polymerase (RdRp), this function was demonstrated in TVMV in which the TVMV NIb had poly(U) polymerase activity and was able to utilize full-length TVMV RNA as a template for RNA synthesis (Hong and Hun, 1996).

The CP gene was characterized in detail, and was roughly divided into three domains, the variable N- and C-terminal domains that are exposed on the surface of the particle and are sensitive to mild trypsin treatment, and the more conserved central or core domain. The N-terminal domain contained the major virus-specific epitopes. The CP was involved in aphid transmission, cell-to-cell and systemic movement, encapsidation of the viral RNA and finally in the regulation of viral RNA amplification (Shukla *et al*., 1998; López-Moya *et al*., 1999).

2.4.3. Evolution

Potyviruses infect plants and belong to the family *Potyviridae*. The genus is named after the type virus PVY. The *Potyvirus*, like the *Begomovirus*, have about 30% of the currently known plant viruses and have at least 180 definitive and possible members. Members of this

genus may cause significant losses in agricultural, pastoral, horticultural and ornamental crops.

This genus *Potyvirus* was studied deeply for the evolution in recent years (Bateson *et al*., 2002; Gibbs *et al*., 2008c; Gibbs and Ohshima, 2010; Simmons *et al*., 2008; Moury and Simon, 2011; Cuevas *et al*., 2012). An earlier phylogenetic analysis of 'coherently evolving' region of the potyvirus coat protein genes (cCP) indicated that the partial coat protein genes of potyviruses had an evolutionary rate of about 1.15×10^{-4} nucleotide substitutions/site/year, and the initial radiation of the potyviruses occurred about 6.600 - 7250 years ago, and hence concided with the dawn of agriculture (Gibbs *et al*., 2008c; Gibbs and Ohshima, 2010).

A phylogeographic analysis of the entire *Potyvirus* genus (Gibbs and Ohshima, 2010) indicated that the genus originated in western Eurasia and/or North Africa, and probably evolved from a virus of monocotyledonous plants. All species of the two earliest-diverging lineages of potyviruses were first isolated from monocotyledonous plants, which were first domesticated in the same region (Simmonds, 1976), as were all species of *Rymovirus*, the close sister genus to *Potyvirus.*

Furthermore, the analysis of the evolutionary dynamics of ZYMV using a Bayesian coalescent approach revealed that the coat protein of ZYMV has evolved at a mean rate of 5.0×10^{-4} nucleotide substitutions/site/year and the divergence time no more than 800 years. The results suggested that the rate is equivalent to those observed in animal RNA viruses and human activities played a central role in the dispersal of ZYMV (Simmons *et al*., 2008).

In an effort to understand the molecular mechanisms underlying such anomalous serological and molecular typing characteristics, partial 3' and 5' sequences was determined

for an atypical Turkish *Plum pox virus* (PPV) isolate (Glasa and Candresse, 2005). The results obtained indicated that its unusual typing behaviour was caused by point mutations affecting key restriction sites and epitopes and confirm that this isolate represents a divergent member of the PPV-M subgroup. In addition, analysis of the partial Ab-Tk genomic sequences demonstrated that the 5' region of the genome of this isolate had a mosaic structure resulting from recombination event(s), shedding new light on the evolutionary history of PPV.

A study on genetic structure of the population of SMV based on the analysis of complete genome sequences showed that recombination analyses identified 'clear' recombination events in the SMV population (Seo *et al.*, 2009). Furthermore, as several resistance-breaking strains were identified as recombinants, it appeared that recombination might contribute to overcome host resistance in SMV-soybean pathosystem. These findings suggested that recombination as well as mutation was an important evolutionary process in the genetic diversification of SMV population (Seo *et al.*, 2009).

Seventy-seven complete PVY genomes collected worldwide showed that both geographic and host-driven adaptations explain PVY diversification. Furthermore, purifying selection was the main force driving PVY evolution. Notably, the analysis of P3N-PIPO seems to show a variable length among the isolates analyzed, and this variability was explained by host-driven adaptation (Cuevas *et al*., 2012). By using mathematical models to analyse the within-host population dynamics of four PVY variants differing at most by two substitutions involved in pathogenicity properties. The models could accurately describe both selection and genetic drift processes shaping the evolutionary dynamics of viruses within their hosts (Fabre *et al*., 2012).

Another recently study indicated that the *Potyvirus* was a fine system to study roles and evolution of simple sequence repeats (SSRs) in viruses. Analyses of densities, relative abundances, compositions and evolutionary inferences of SSRs in 45 different *Potyvirus* genomes revealed that the densities and relative abundances of SSRs were similar in all those *Potyvirus* genomes. The variety of SSRs might be related to the genome diversity of *Potyvirus*. Perhaps *Potyvirus* and HIV genomes have the similar evolution mode and parallel evolution level (Zhao *et al*., 2011).

In Australia, potyviruses were divided into two groups. The viruses entered Australia after agriculture was established by migrants from Europe. Potyviruses are transmitted in seed more frequently than experimental evidence indicates, and shows that understanding the sources of emerging pathogens and the frequency with which they 'emerge' was essential for proper national biosecurity planning (Gibbs *et al*., 2008a).

2.5. *Turnip mosaic virus*

2.5.1. Biological properties

TuMV is a member of the genus *Potyvirus* in the family *Potyvirideae* (ICTV, 2011). TuMV has a wide natural host range of crop and weed plant species and geographically widespread. The virus is known to infect both dicot and monocot plants. Especially, it infects not less than 318 species from over 43 dicot families including Brassicaceae, Asteraceae, Chnopodiaceae, Fabaceae and Caryophyllaceae (Walsh and Jenner, 2002). TuMV was found from cultivated brassicaceous cash crops such as Chinese cabbage, radish, and smooth-leaf mustard in Asian countries (Sako, 1981), whereas from wide plants such as wild orchids in

Europe (Lesemann and Vetten, 1985), Brassicaceae weed in Iran (Farzadfar *et al*., 2008) and wild radish in Turkey (Korkmaz *et al*., 2008).

The previous studies of host reaction (Ohshima *et al*., 2002; Tomimura *et al*., 2003; 2004; Tomitaka and Ohshima, 2006) showed that TuMV isolates were of four host-infecting types; [(B)]-host type isolates infected *Brassica* plants latently and occasionally and did not infect *Raphanus* plants, [B]-host type isolates infected many of *Brassica* plants systemically causing mosaic of the systemically infected leaves but did not infect *Raphanus* plants, [B(R)]-host type isolates infected many *Brassica* plants systemically causing mosaic of systemically infected leaves but infected *Raphanus* plants latently and occasionally, and [BR]-host type isolates infected both *Brassica* and *Raphanus* plants systemically causing mosaic of the uninoculated leaves. Moreover, phylogenetic analyses using gene sequences of different virus isolates collected from around the world revealed four main TuMV genogroups called basal-B (*Brassica*), basal-BR (*Brassica*/*Raphanus*), Asian-BR and world-B. The basal-B cluster of (B) or B-host type isolates was most variable, was paraphyletic, and was isolated from both non-and Brassicaceae plants.

TuMV is transmitted by at least 89 species of aphids in the non-persistent manner, particularly by *Brevicoryne brassicae* and *Myzus persicae* (Walsh and Jenner, 2002). The virus can be acquired and inoculated by aphids in less than 1 min (Sylvester, 1953). There is no latent period. TuMV was retained by feeding vectors for less than 4 hrs (Sylvester, 1954). Up to date, there is no any report that TuMV is transmitted through the seed.

2.5.2. Serological properties

Serological techniques have been applied widely to the diagnosis of plant viruses. The serological properties of potyviruses have been found extremely useful for identifying and

classifying the individual members of this large group of plant viruses (Van Regenmortel, 1982; Van Regenmortel and Dubs, 1993).

For TuMV, antisera with titres of 1/512 were prepared by Shepherd and Pound (1960) and Tomlinson and Walkey (1967) using purified virus isolates obtained respectively from cabbage and rhubarb. In precipitin tube tests, 0.85% NaCl was used as diluent. Reactions in such tests gave flagellar-type precipitates. In gel-diffusion tests the virus normally gave no reaction but after ultrasonic treatment (Tomlinson *et al.,* 1965; Tomlinson and Walkey, 1967) diffusion of fragmented antigen results in a precipitation line (http://www.dpvweb.net/dpv/showdpv.php?dpvno=8).

There are several serology techniques; Enzyme-linked immunosorbent assay (ELISA), agglutination, precipitation, complement-fixation, and fluorescent antibodies. The ELISA technique (Clark and Adam, 1977) is used with a wide range of studies in many biological areas, especially in diagnosis of human, animal and plant diseases. The number of publications with ELISA mentioned in all science areas are increasing rapidly every year. For instance, a literature search in ScienceDirect database, there is about 253,000 publications using ELISA technique upto 2013.

2.5.3. Genome organization

TuMV, like other potyviruses, has flexuous filamentous particles 700-750 nm long, each of which contains a single copy of the genome, which is a single-stranded positive sense RNA molecule of about 9,800 nucleotides. This is translated into one large polyprotein which hydrolyzes itself into 10 proteins (King *et al*., 2012). Furthermore, a PIPO exists in the +2 reading frame within the protein 3 (P3 protein) encoding region (Chung *et al*., 2008).

2.5.4. Evolution

A phylogeny inferred from the polyproteins encoded by these genomes (Gibbs and Ohshima, 2010) revealed that there are at least 11 distinct lineages of potyviruses. One of these is the 'TuMV group', which comprises TuMV and at least five other species, all from monocotyledonous plants. These include *Japanese yam mosaic virus* (JYMV) (Fuji and Nakamae, 1999; 2000), *Narcissus yellow stripe virus* (NYSV) (Chen *et al*, 2003), and *Scallion mosaic virus* (ScMV) (Chen *et al*, 2002), all known from full genomic sequences. Sequence analyses of the full sequence have shown that the TuMV group also contains Indian narcissus potyvirus and *Narcissus late season yellows virus* (NLSYV).

Previous studies have showed that the different TuMV subpopulations probably emerged from the more ancient Eurasian subpopulations, such as those found in the Mediterranean region, including Southeast Europe, Asia Minor and mid-Eurasia (Ohshima *et al*., 2002; Tomimura *et al*., 2004). In these regions, brassicaceous crops are an important component of local agriculture; in Europe, the crops are mostly *Brassica* species; and in Asia Minor, both *Brassica* and *Raphanus* species are important.

Recent studies of the genetic structures of TuMV populations in East Asia were made using partial genome sequences, and showed that recent Chinese and Japanese TuMV isolates were parts of the same population but were a discrete lineage (Tan *et al*., 2004; Tomitaka and Ohshima, 2006).

Comparisons of recombinational, mutational and phylogenetical analyses between the East and West Eurasian subpopulations showed that the subpopulations could be considered to be different populations and were likely to have evolved independently from an ancestral population (Tomimura *et al.*, 2004). Compairisons of many isolates of TuMV in East Asia

found more recombination sites in the P1 and NIa-VPg genes than in the NIa-Pro and CP genes (Tan *et al.*, 2004). This study also showed that individual recombinants were dirived from parents in the same major lineage (intralineage recombination). These recombinations provided useful additional information about the migratory spread of the virus, perhap the first such report for a plant virus population. The study of recombination in 92 complete genomes of TuMV further supported our earlier conclusion that recombination was a dominant feature of TuMV evolution. It was probable that both recombination and mutation were required to produce new host-infecting type during TuMV evolution (Ohshima *et al.*, 2007).

In South East Asia, TuMV was recently reported in Vietnam (Ha *et al*., 2008), little data on the incidence of TuMV and its biological and molecular characteristics are available in Southeast Asian countries. Studies on the evolution of TuMV isolates in the Middle East reported that Turkish isolate TUR1 was an intralineage recombinant, and belonged to B-host infecting type, whereas isolate TUR9 belonged to BR-host infecting type and was a non-recombinant of Asian-BR group (Korkmaz *et al*., 2008). In contrast, two Iranian isolates IRNTRa6 and IRNSS5 were non-recombinants (Farzadfar *et al*., 2008). Both isolates fell into basal-B group in the phylogenetic tree and belonged to B-host infecting type. A recent study on the evolution of TuMV showed that the nucleotide diversity and the non-synonymous/synonymous ratio of the populations from the new host were higher than those from the old host. Molecular variance had significant differences among the populations from the old and new hosts (Ohshima *et al*., 2010).

VI. MATERIALS AND METHODS

1. Virus isolates

1.1. Vietnamese isolates

The Brassicaceae crop-producing areas of Vietnam, including Northwest, Northeast, Red river delta, North central coast, South central coast and Central highlands regions, were surveyed during the growing seasons of 2006–2008 (Figure 2). Details of the Vietnamese TuMV isolates, their place of origin, original host plant, symptom on collected plant, year of isolation, and host type are shown in Table 1. The Northeast region in Vietnam is one of major areas of Brassicaceae crop production.

All collected samples were tested by direct double antibody sandwich enzyme-linked immunosorbent assay (DAS-ELISA) (Clark and Adams, 1977)(Figure 3).

1.2. European isolates

European isolates used in this study were mostly collected from various host plants in European countries including Belgium, Denmark, France, Germany, Greece, Hungary, Italy, Poland, Portugal, Spain, The Netherlands, and the United Kingdom. Isolates were also collected in non-European countries including the United States of America. The isolates were collected from private gardens and fields with permission from the owner when collect samples from their properties. Some samples came from colleagues and these are listed in the Acknowledgements. None of the samples were from 'endangered' species. Details of the isolates are shown in Table 2.

Figure 2. A survey and collection of *Turnip mosaic virus* in a *Brassica* field of Vietnam in 2006.

Table 1. *Turnip mosaic virus* isolates collected in Vietnam.

Isolate	Original host	Location (City, Province/Distict)	Year of collection	Symptom on collected plant[a]	Host type[b]	Accession No.
VIET15	*Raphanus sativus*	Van Giang, Hung Yen	2006	M	B(R)	AB747286
VIET56	*Brassica juncea*	Moc Chau, Son La	2007	M	B	AB747287
VIET58	*B. juncea*	Moc Chau, Son La	2007	M	BR	AB747288
VIET65	*R. sativus*	Gia Lam, Ha Noi	2007	M	BR	AB747289
VIET66	*R. sativus*	Gia Lam, Ha Noi	2007	M	B	AB747290
VIET73	*R. sativus*	Van Giang, Hung Yen	2007	M	BR	AB747291
VIET79	*R. sativus*	Cam Giang, Hai Duong	2007	M	BR	AB747292
VIET80	*R. sativus*	Cam Giang, Hai Duong	2007	M	B	AB747293
VIET82	*R. sativus*	Ban Me Thuot, Dak Lak	2007	M	B(R)	AB747294
VIET83	*R. sativus*	Ban Me Thuot, Dak Lak	2007	M	B	AB747295
VIET89	*R. sativus*	Ban Me Thuot, Dak Lak	2007	M	BR	AB747296
VIET138	*B. juncea*	Thanh Long, Thua Thien Hue	2007	M, S	B(R)	AB747297
VIET153	*B. juncea*	Hoi An, Quang Nam	2007	M	B(R)	AB747298
VIET158	*B. juncea*	Gia Lam, Ha Noi	2007	M	B(R)	AB747299
VIET159	*B. juncea*	-, Lang Son	2007	M	B	AB747300
VIET160	*B. juncea*	Huu Lung, Lang Son	2007	M	B	AB747301
VIET164	*B. juncea*	Thuong Tin, Ha Tay	2007	M, S	B	AB747302
VIET166	*B. juncea*	Thuong Tin, Ha Tay	2007	M	B	AB747303
VIET167	*B. juncea*	Gia Lam, Ha Noi	2007	M	B	AB747304
VIET169	*B. juncea*	Vo Cuong, Bac Ninh	2007	M	B	AB747305
VIET170	*B. juncea*	Vo Cuong, Bac Ninh	2007	M	B(R)	AB747306
VIET172	*B. juncea*	Gia Lam, Ha Noi	2007	M	B	AB747307
VIET173	*B. juncea*	Viet Yen, Bac Giang	2007	M	B	AB747308
VIET174	*B. juncea*	Viet Yen, Bac Giang	2007	M	B	AB747309
VIET175	*B. juncea*	Viet Yen, Bac Giang	2007	M	B(R)	AB747310
VIET176	*B. juncea*	Vu Thu, Thai Binh	2007	M	B(R)	AB747311
VIET177	*B. juncea*	Vu Thu, Thai Binh	2008	M, S	B	AB747312
VIET178	*B. juncea*	Nam Truc, Nam Dinh	2008	M, S	B(R)	AB747313
VIET179	*B. juncea*	Nam Truc, Nam Dinh	2008	M	B(R)	AB747314
VIET180	*B. juncea*	Viet Yen, Bac Giang	2008	M	B(R)	AB747315

[a] M; mosaic, S; stunting

[b] Host type B; *Brassica*, isolates infected *B. rapa* cv. Hakatasuwari systemically giving mosaic symptoms. Host type BR; these isolates infected both *B. rapa* and *R. sativus* cvs. Akimasari and Taibyo-sobutori (Japanese radish) systemically giving symptoms. B(R); isolates infected *B. rapa* systemically giving mosaic symptoms and infected *R. sativus* cvs. Akimasari and Taibyo-sobutori only occasionally and latently.

Microplate (96 wells)(NUNC)

↓Coat with IgG diluted in 0.05M sodium carbonate bicarbonate buffer (pH 9.6)[1)]

↓ Incubate at 37°C for 3 hours

↓ Wash with 0.02M PBS-Tween[2)] for 3 minutes, repeat 3 times

Add 200µl sap homogenized with 0.02M PBS-Tween to plate

↓ Incubate at 4 °C for overnight

↓ Wash with 0.02M PBS-Tween for 3 minutes, repeats 3 times

Add conjugate diluted with 0.02M PBS-Tween

↓ Incubate at 37 °C for 3 hours

↓ Wash with 0.02M PBS-Tween for 3 minutes, repeat 3 times

Add substrate (*p*-nitrophenyl phosphate) diluted with 10% diethanol amine stock solution (pH9.8)[3)]

↓ Incubate at room temperature for 15 minutes to 2 or 3 hours

Measure absorbance at 405 nm

[1)] 0.05M sodium carbonate bicarbonate buffer (pH 9.6)

Na_2CO_3	1.59 g
$NaHCO_3$	2.93 g
NaN_3	0.20 g
DW	800 ml

Adjust pH with HCl to 9.6 and fill up to 1000 ml with DW.

0.1M PBS stock solution (pH 7.4)

NaCl	43.5 g
$Na_2HPO_4.12H_2O$	32.8 g
KH_2PO_4	0.8 g
KCl	1.0 g
NaN_3	1.0 g

Fill up to 1000 ml with DW

↓ Diluted 5 times with DW

0.02 M PBS

[2)] 0.02M PBS-Tween

0.02 M PBS	1000 ml
Tween 20	0.5 ml

[3)]10% Diethanol amine stock solution (pH 9.8)

Diethanol amine (2,2´-Iminodiethanol)	50 ml
DW	300 ml

Adjust pH with HCl to 9.8 and fill up to 500 ml with DW

Figure 3. Procedure for double antibody sandwich enzyme-linked immunosorbent assay (DAS-ELISA).

Table 2. *Turnip mosaic virus* isolates collected in Europe and The United State of America.

Isolate	Original host	Location (City, District, Country)	Year of collection	Host type[a]	Accession code
Europe					
AllA	*Alliaria officinalis*	-, -, Denmark	1991	B	this study, AB701694
ASP	*Allium* sp.	Gatersleben, -, Germany	1995	B	this study, AB701697
BEL 1	*Rorippa nasturtium-aquaticum*	Wavre, -, Belgium	1986	B	this study, AB701698
DEU 1	Not known	-, -, Germany	<1976	B	this study, AB701699
DEU 2	*Raphanus sativus*	-, -, Germany	<1993	B	this study, AB701700
DEU 4	*Lactuca sativa*	Stuttgart, -, Germany	1986	BR	this study, AB701701
DEU 5	*L. sativa*	Monchengladbach, -, Germany	1991	B	this study, AB701702
DEU 7	*L. sativa*	Frankfurt,-, Germany	1994	B	this study, AB701695
DNK 3	*Brassica rapa*	-, -, Denmark	1978	B	this study, AB701703
DNK 4	*B. rapa*	-, -, Denmark	1986	B	this study, AB701704
Eru 1D	*Eruca sativa*	Piedmont, Italy	1991	B	this study, AB701705
ESP 1	*Eruca vesicaria* subsp. sativa	-, -, Spain	2001	B	this study, AB701706
ESP 2	*Sisymbrium orientale*	Las Matas, -, Spain	2001	B	this study, AB701707
FRA 2	*B. napus*	-, -, France	<1994	B	this study, AB701708
GBR 7	*Rheum rhabarbarum*	Gloucestershire, -, UK	15.9.1993	B	this study, AB701709
GBR 8	*Lunaria annua*	Essex, -, UK	20.4.1994	B	this study, AB701710
GBR 27	*B. oleracea*	Kimmeridge, Dorset, UK	24.3.1999	B	this study, AB701711
GBR 30	*B. oleracea*	Kimmeridge, Dorset, UK	26.4.1999	B	this study, AB701712
GBR 31	*B. oleracea*	Chapman's Pool, Dorset, UK	26.4.1999	B	this study, AB701713
GBR 32	*B. oleracea*	Chapman's Pool, Dorset, UK	26.4.1999	B	this study, AB701714
GBR 38	*B. oleracea*	Winspit, Dorset, UK	15.7.1999	B	this study, AB701715
GBR 51	wild *B. oleracea*	Staithes, Yorkshire, UK	28.9.1999	B	this study, AB701742
GBR 57	wild *B. oleracea*	Llandudno, Conwy, UK	12.9.2000	B	this study, AB701716
GBR 83	wild *B. oleracea*	Llandudno, Conwy, UK	20.8.2002	B	this study, AB701717
GBR 91	wild *B. oleracea*	Llandudno, Conwy, UK	20.8.2002	B	this study, AB701718
GK1	*Matthiola incana*	-, -, Greece	<1989	B	this study, AB701696
HUN 1	*Alliaria petiolata*	-, -, Hungary	<1996	B	this study, AB701719
ITA 1A	*B. ruvo*	-, Campania, Italy	1990	B	this study, AB701720
ITA 2	*Cheiranthus cheiri*	-, Campania, Italy	1992	BR	this study, AB701721
ITA 4	*B. rapa*	-, Campania, Italy	1990	B	this study, AB701722
ITA 5	*B. ruvo*	-, Campania, Italy	1990	B	this study, AB701723
ITA 6	*M. incana*	-, Campania, Italy	1992	B	this study, AB701724
ITA 8	*Abutilon* sp.	Piedmont, Italy	09.1993	BR	this study, AB701725
ITA 9A	*Cucurbita pepo*	-, -, Italy	<1995	B	this study, AB701726
NLD 2	*B. oleracea*	-, -, The Netherlands	<1995	B	this study, AB701727
OM-A	*Orchis militaris*	Celle, -, Germany	1981	DI[b]	this study, AB701691
OM-N	*O. militaris*	Celle, -, Germany	1981	DI	this study, AB701690
ORM	*O. morio*	Celle, -, Germany	1983	(B)	this study, AB701692
OS	*O. simian*	Celle, -, Germany	1981	DI	this study, AB701693
POL 1	*B. napus oleifera*	Poznan, -, Poland	<4.10.1993	B	this study, AB701728
POL 2	*Papaver somniferum*	Czempin, -, Poland	<4.10.1993	B	this study, AB701731
POL 4	*B. napus oleifera*	Grabianowo, -, Poland	<4.10.1993	B	this study, AB701732
PRT 1	*B. oleracea acephala*	Madeira, -, Portugal	1993/1994	B	this study, AB701729
PV0054	*B. oleacea*	-, -, Germany	Not known	B	this study, AB701730
PV177	*Brassica* spp.	Cambridge, UK	1934[c]	B	this study, AB701733
TIGA	*Tigridia* sp.	Braunschweig, -, Germany	1983	(B)	this study, AB701734
TIGD	*Tigridia* sp.	Braunschweig, -, Germany	1983	(B)	this study, AB701735
UT	*Utricularia* sp.	Wuerzburg, -, Germany	1997	B	this study, AB701736
USA					
PV134	*Sesynibium* sp.	-, California, USA	<1960	B	this study, AB701737
PV389	*Tulipa gesnerana*	Beltsville, Maryland, USA	1986	B	this study, AB701738
USA 4	*B. pekinensis*	-, -, USA	1993	B	this study, AB701739
USA 5	*R. sativus*	San Francisco, California, USA	2002	BR	this study, AB701740
USA 6	*R. sativus*	San Francisco, California, USA	2002	BR	this study, AB701741

[a] Host type B; *Brassica*, isolates infected *B. rapa* cv. Hakatasuwari systemically giving mosaic symptoms. Host type (B); isolates infected *B. rapa* only occasionally. Host type BR; these isolates infected both *B. rapa* and *R. sativus* systemically giving mosaic symptoms. Host type B(R); isolates infected *B. rapa* systemically giving mosaic symptoms and infected *R. sativus* only occasionally.
[b] Difficult to infect brassica plants.
[c] Unclear

The orchid-infecting TuMV-like viruses were isolated from *Orchis militaris, Orchis morio*, and *Orchis simia* plants growing in a collection at Celle, Germany. These isolates were collected by Vetten and Lesemann, one of authors of the present study, and details of these isolates have already been published (Lesemann and Vetten, 1985). The isolates were collected by the permission of the owners. Orchis isolates were also found in nine other species of Orchidaceae in the same collection, so it is uncertain which species were the original sources of the TuMV-Orchis viruses, although all the orchids have overlapping natural distributions in eastern, central, and southern Europe. The collected samples were tested for the presence of TuMV by DAS-ELISA (Figure 4).

2. Host tests

All of the isolates were sap-inoculated to *Chenopodium quinoa* plants (Figure 5A) and serially cloned through single lesions at least three times. TuMV isolates were generally cloned by single lesion isolations in the earlier (Ohshima *et al*., 2002, 2007; Tomimura *et al*., 2003, 2004) and present studies because of the high frequency of mixed infections in the field, not only with other viruses but also other isolates of the same virus. Thus, biological cloning is mandatory when attempting to analyse recombination events and the genetic structure of populations. They were propagated in *Brassica rapa* cv. Hakatasuwari (Figure 5B) or *Nicotiana benthamiana* plants (Figure 5C). Plants infected systemically with each of the TuMV isolates were homogenized in 0.01 M potassium phosphate buffer (pH 7.0) and mechanically inoculated on to young plants of *B. rapa* cv. Hakatasuwari, and *Raphanus sativus* cvs Taibyo-sobutori (Figure 5D) and Akimasari. Infected leaves *B. rapa* cv. Hakatasuwari or *N. benthamiana* were harvested after two weeks or one month. They were kept in Silica gel in 4^{o}C or freeze dried and kept in -25^{o}C for further studies.

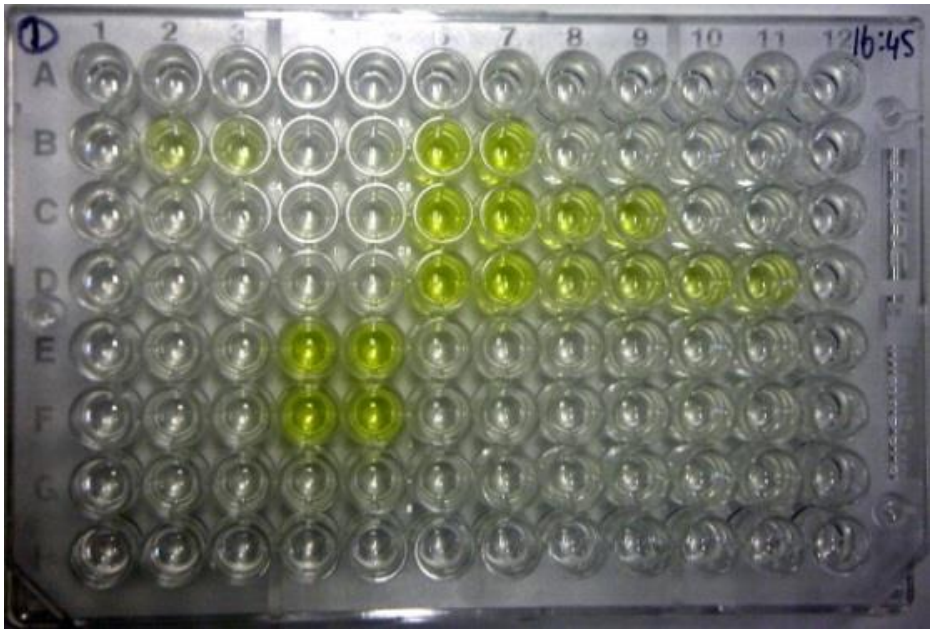


Figure 4. An example of DAS-ELISA test for detecting the presence of *Turnip mosaic virus* from collected samples. The samples with positive reactions appeared in yellow color.

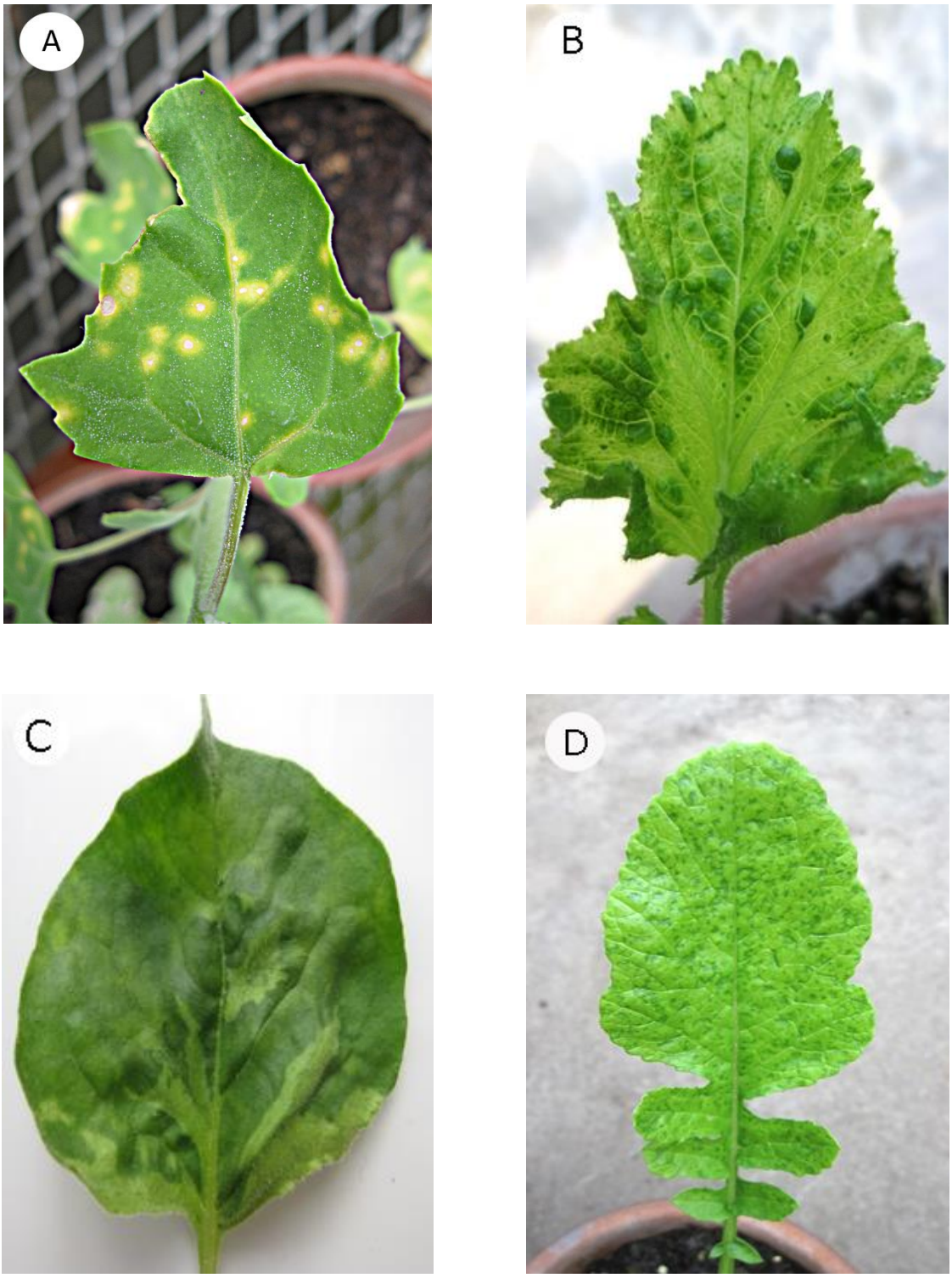


Figure 5. Symptoms produced by *Turnip mosaic virus* on *Chenopodium quinoa* (local lesion) (A), mosaic on *Brassica rapa* cv. Hakatasuwari (B), *Nicotiana benthamiana* (C) and *Raphanus sativus* cv. Taibyo-sobutori (D) after two or three weeks of inoculation.

The isolates collected in Vietnam were mechanically inoculated to young plants of *B. chinensis* cv. Choyo, *B. juncea* cv. Hakarashina, *B. napus* cv. Otsubu, *B. rapa* cv. Hakatasuwari, *B. oleracea* var. capitata cvs. Ryozan 2-go and Shinsei, *B. pekinensis* cvs. Nozaki 1-go and Kyoto 3-go, and *B. napus* cv. Norin-32go, *Eruca sativa* cv. Odyssey and *Lactuca sativa* cv. Emrap 231, as well as to *R. sativus* cvs. Akimasari, Taibyo-sobutori and Everest.

The isolates collected in European countries include isolates from *Orchis* and some other isolates, were also tested for host reactions using plants from a broader range of species (the name of hosts test will be presented in the results section). Inoculated plants were kept for at least four weeks in a glasshouse at 25^{o}C.

3. Viral RNA extraction, reverse transcription-polymerase chain reaction and sequencing

The genomic sequences from 30 isolates of TuMV representing different parts of Vietnam, 48 isolates from European countries and five isolates from USA were determined. The viral RNAs were extracted from TuMV-infected *B. rapa* cv. Hakatasuwari or *N. benthamiana* leaves using Isogen (Nippon Gene, Japan)(Figure 6). The RNAs were reverse transcribed by PrimeScript® *Moloney murine leukemia virus* (MMLV) reverse transcriptase (Takara, Japan) and amplified using high-fidelity Platinum™ *Pfx* DNA polymerase (Invitrogen, USA)(Figure 7). The reverse transcription and polymerase chain reaction (RT-PCR) products were separated by electrophoresis in agarose gels (Figure 8) and purified using the QIAquick Gel Extraction Kit (Qiagen K.K. Japan)(Figure 9). Sequences from each isolate were determined using four to five overlapping independent RT-PCR products (Figure 10) to cover the complete genome. The sequences of the RT-PCR products of adjacent regions of the genome overlapped by at least 200 bp to ensure that they were from the same genome and

0.05 ～ 0.1 g TuMV infected leaves

↓ Add 1 ml ISOGEN

↓ Homogenized with a small mortar and pestle

↓ Transfer into microcentrifuge tube 1.5 ml

↓ Store at room temperature for 5 minutes

↓ Centrifuge 14,000 rpm at 4°C for 10 minutes

Transfer supernatant into microcentrifuge tube 1.5 ml

↓ Add 200 µl chloroform

↓ Vortex for 20 seconds

↓ Store at room temperature for 3 minutes

↓ Centrifuge 14,000 rpm at 4 °C for 15 minutes

Transfer supernatant into microcentrifuge tube 1.5 ml

↓ Add 500 µl isopropanol

↓ Mix gently with hand at 10 times

↓ Store at room temperature for 10 minutes

↓ Centrifuge 14,000 rpm at 4 °C for 10 minutes

Pellet

↓ Add 1 µl of EtOH 70%

↓ Centrifuge 14,000 rpm at 4°C for 5 minutes

Pellet

↓ Dry up for 7 minutes

↓ Add 100 µ DEPC-Treated Water (Ambion)

↓ Add 10 µl 3M NaOAC

↓ Add 250 µl 100% EtOH

↓ Store at －80°C for 60 minutes

↓ Centrifuge 14,000 rpm at 4 °C for 15 minutes

Pellet

↓ Add 400 µl of EtOH 70%

↓ Centrifuge 14,000 rpm at 4 °C for 5 minutes

Pellet

↓ Dry up for 7 minutes

↓ Add 20-40 µl DEPC-Treated Water (Ambion)

Total RNAs use for reverse transcription and polymerase chain reaction (RT-PCR)

Figure 6. RNA extraction using ISOGEN (Nippon Gene).

Reverse transcription (RT) reaction mixture	
DEPC-Treated Water	X µl
10mM dNTPs	1.0 µl
5 × RT buffer	4.0 µl
Reverse primer (50 pmol/µl)	1.0 µl
RNA (1 µg/1 µl)	Y µl
PrimeScript™ reverse transciptase	1.0 µl
Total volume	20.0 µl

Incubate at 42℃ for 1 hour

Polymerase chain reaction (PCR) mixture	
DDW	29.7 µl
10mM dNTPs	1.0 µl
10 × *Pfx* Amplification buffer	5.0 µl
10 × PCR Enhancer Solution	5.0 µl
50mM $MgSO_4$	2.0 µl
1st cDNA (RT reaction mixture)	4.0 µl
Forward primer (10 pmol/µl)	1.5 µl
Reverse primer (10 pmol/µl)	1.5 µl
Platium™ *Pfx* DNA polymerase	0.33 µl
Total volume	50.0 µl

Vortex and flash centrifuge

PCR cycle:	
94℃ (2 min)	1 cycle
94℃ (15 sec) - 45℃ (30 sec) - 68℃ (3 min)	40 cycles
4℃ (10 min) - 25℃ (∞)	1 cycle

Figure 7**.** Reverse transcription and polymerase chain reaction (RT-PCR).

0.7% Agarose gel (20 ml)	
Agarose	0.14 g
Ethidium bromide (250 µg / µl)	20 µl
50xTAE	400 µl
DDW	20 ml

50xTris-Acetate-EDTA (TAE) buffer	
Tris	60.5 g
Acetic acid	14.275 ml
0.5M EDTA	25 ml
Fill up to 250 ml with DDW	

1xTAE buffer (400ml)	
DDW	392 ml
50xTAE	8 ml
Ethidium bromide (250 µg / µl)	400 µl

Figure 8. Compositions of Tris-Acetate-EDTA buffer and agarose gel.

Agarose gel

↓Excise the DNA fragment from the agarose gel

↓Add 3 volumes of Buffer QG

↓Incubate at 50℃ for 10 minutes

Apply the sample to the QIAquick column

↓Centrifuge 12,000 rpm at room temperatue for 1.5 minutes

Discard flow through

↓Add 500 µl of Buffer QG

↓Centrifuge 12,000 rpm at room temperatue for 1.5 minutes

Discard flow through

↓Add 750 µl of Buffer PE

↓Store at room temperatue for 5 minutes

↓Centrifuge 12,000 rpm at room temperatue for 1.5 minutes

Discard flow through

↓Centrifuge 13,000 rpm at room temperatue for 1.5 minutes

Place QIAquick column into 1.5 ml tube

↓Add 40-50 µl of 10mM Tris-HCl (pH 8.5)

↓Store at room temperatue for 1 minute

↓Centrifuge 13,000 rpm at room temperatue for 1.5 minutes

Electrophoresis in 0.7% agarose gel to assess amounts of the DNA for sequence reaction

Figure 9. DNA extraction by QIAquick Gel Extraction Kit (QIAGEN).

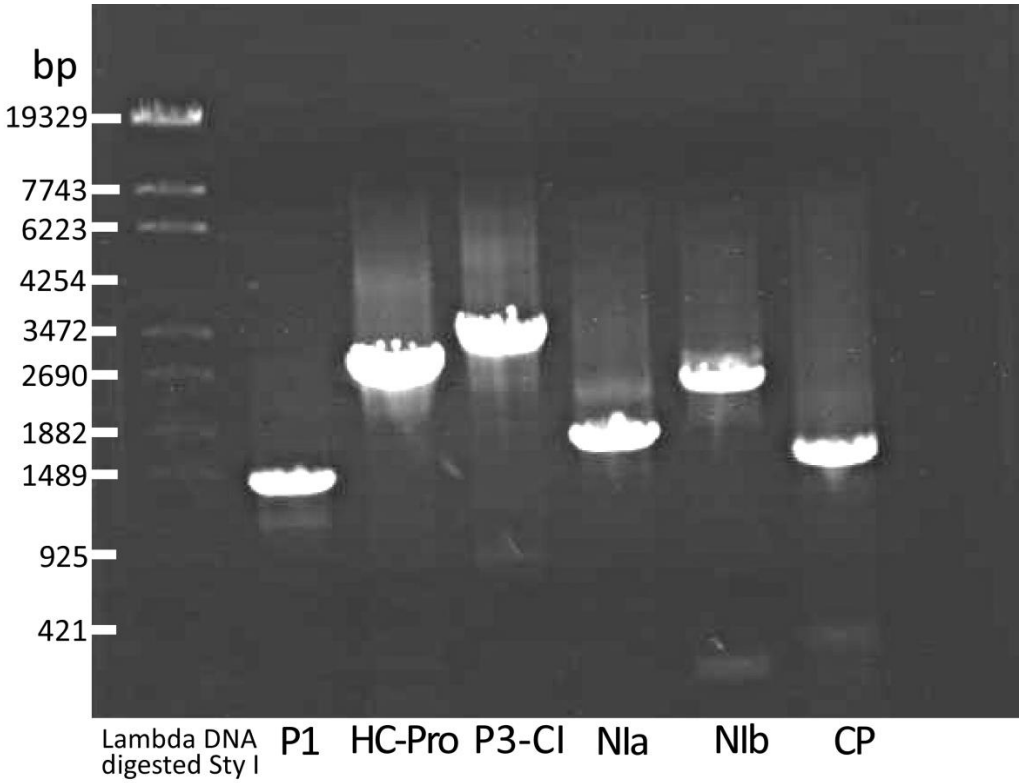


Figure 10. Reverse transcription-polymease chain reaction (RT-PCR) products of *Turnip mosaic virus*. Six overlapping regions were amplied; Protein 1 (P1), helper component-proteinase protein (HC-Pro), Protein 3–Cylindrical inclusion protein (P3-CI), nuclear inclusion a protein (NIa), nuclear inclusion b protein (NIb) and coat protein (CP). The overlapping regions were at least 200 nucleotides to ensure that all the regions come from the same virus.

were not from different components of a genome mixture. Each RT-PCR product was sequenced by primer walking in both directions (Table 3) using a BigDye Terminator v3.1 Cycle Sequencing Ready Reaction kit (Applied Biosystems, USA)(Figure 11) and an Applied Biosystems 310 Genetic Analyzer. Ambiguous nucleotides in any sequence were checked in sequences obtained from at least five other independent plasmids as described in the earlier studies (Ohshima *et al*, 2002, 2007; Tomimura *et al*, 2003, 2004).

Sequence data were assembled using BioEdit version 5.0.9 (Hall, 1999). The similarity of nucleotide sequences between TuMV isolates and group viruses was determined using SIMPLOT version 3.5.1 (Lole *et al*., 1999) with a window length of 200 and step size of 20.

4. Evolutionary analyses

4.1. Sequences alignment

4.1.1. Vietnamese isolates

The total 135 genomic sequences of 30 Vietnamese isolates (Table 1) together with details of the 105 isolates which complete genomic sequences have already been reported (Table 4) were used for spatial evolution analyses. Two sequences of *Japanese yam mosaic virus* (JYMV) (Fuji and Nakamae, 1999, 2000), one of *Scallion mosaic virus* (ScMV) (Chen *et al*., 2002) and one of *Narcissus yellow stripe virus* (NYSV) (Chen *et al*., 2003b) were used to align the TuMV genomic sequences as BLAST searches had shown them to be the sequences in the international sequence databases most closely and consistently related to those of TuMV; TuMV protein 1 (P1) genes were more closely related to those of JYMV than ScMV,

Table 3. Primers used in this study.

Primer name	Position
Tu5T4P	**1-26**
Tu5T6P	**1-25**
TuP150P	205-224
TuP12P	259-277
TuP11P	295-312
TuP113P	332-350
TuP137M	347-364
TuP135M	472-491
TuP112P	640-658
TuP16P	**649-667**
TuP128M	1042-1060
TuP146P	1153-1172
TuHC32P	1216-1237
TuHC30M	1381-1400
TuHC27M	**1381-1410**
TuHC13P	1483-1502
TuKA1HC11P	1834-1853
TuHC45P	2125-2144
TuHC40P	2161-2179
Tu59HC9P	2513-2533
TuP313M	2737-2756
TuP321M	2845-2865
TuP3OP1P	**3091-3112**
Tu59P311M	3207-3225
TuP3OP1M	**3208-3229**
TuDMP310M	3451-3469
Tu59CI11P	3812-3832
TuKA1CI13P	3883-3902
TuCI25M	4063-4082
TuKA1CI14M	4660-4680
Tu59CI6M	5050-5068
TuCI36M	5437-5456
Tu59CI9P	**5608-5627**
Tu596K24M	5791-5810
Tu596K23P	5824-5842
TuKA1VPG3P	5908-5927
TuVPG8M	**6076-6098**
Tu59NIA2P	6493-6511
TuNIA8P	6545-6565
TuNIA28P	6652-6671
TuNIA10M	6703-6722
TuNIA13M	6817-6837
TuNIA9P	6841-6860
TuNIA4P	**6862-6881**
TuNIA12M	7060-7079
Tu59NIA1M	7123-7141
Tu59NIA3P	7163-7183
TuNIB23P	7261-7279
Tu59NIB14M	**7279-7298**
TuNIB6P	7465-7484
TuNIB6M	7471-7489
TuNIB4P	7816-7835
Tu59NIB11P	7843-7861
TuNIB16P	8125-8144
TuNIB17P	**8383-8402**
Tu59CP1M	**9154-9173**
TuCP6M	9235-9254
TuCP10P	9319-9340
TuCP5P	9466-9485
TuCP8M	9541-9559
Tu3T9M	**9835-polyA**

The primes in bold letters are used for amplifying reverse transcription – polymerase chain reaction products.P: Plus sense, M: Minus sense

PCR for sequencing	
DDW	X µl
5 × Sequence buffer	2.6 µl
Primer (1 pmol / µl or 2 pmol / µl)	1.0 µl
cDNA (200-250ng)	Y µl
Ready Reaction Premix	1.0 µl
Total	20.0 µl
Vortex and flash centrifuge	

PCR cycle:	
96℃ (1 min)	1 cycle
96℃ (30 sec) - 43 ℃ (1 min) - 60℃ (4 min)	30 cycles
25℃ (∞)	1 cycle

Figure 11. Polymerase chain reaction for sequence reaction.

whereas for some other genomic regions between helper-component proteinase protein (HC-Pro) and nuclear inclusion b protein (NIb) sequence it was the converse, except that the TuMV coat protein (CP) gene is most closely related to that of NYSV. Therefore, all 135 P1 sequences with those of two JYMV isolates as the outgroup, the CP sequences with that of NYSV, and the remaining sequences with those of JYMV and ScMV were aligned using CLUSTAL_X2 (Larkin *et al*., 2007). However, this procedure resulted in some gaps that were not in multiples of 3 nts. Therefore the amino acid sequences corresponding to individual regions were aligned with the appropriate outgroups shown above using CLUSTAL_X2 with TRANSALIGN (kindly supplied by Georg Weiller) to maintain the degapped alignment of the encoded amino acids and then the aligned subsequences were reassembled to form complete sequences 9309 nt long. To check the junctions between 5' or 3' untranslated-region (UTR) and P1 or CP protein encoding regions, aligned 5' and 3' UTR sequences were joined with the polyprotein sequences of each isolate. This produced sequences 9,605 nucleotides long excluding the 35 nucleotides that were used to design the primer for RT-PCR amplification.

4.1.2. European isolates

The total 155 genomic sequences of the 48 isolates from European countries, five isolates from USA (Table 2) together with details of the 102 isolates (Table 4, but excluded three isolates BJ-C4, Lu2 and CAR51) were used for temporal evolution analyses. Two genomic sequences of JYMV (Fuji and Nakamae, 1999, 2000), one of ScMV (Chen *et al*., 2002), and one of NYSV (Chen *et al*., 2003b) were used as outgroups. The process of sequences alignment was done as the same discribled in aligment of Vietnamese isolates.

Table 4. *Turnip mosaic virus* isolates used in this study.

Isolate	Original host	Location (City, District, Country)	Year of collection	Host type[a]	Reference	Accession code
Asia						
1J	*Raphanus sativus*	Saga, Saga, Japan	1977	BR	Ohshima *et al*. (1996)	D83184
2J	*Brassica pekinensis*	-, Tochigi, Japan	1994	BR	Tomimura *et al*. (2003)	AB093622
59J	*R. sativus*	Saga, Saga, Japan	1996	BR	Tomimura *et al*. (2003)	AB093620
AD178J	*R. sativus*	Rokunohe, Aomori, Japan	1998	BR	Ohshima *et al*. (2007)	AB252094
AD181J	*R. sativus*	Tohoku, Aomori, Japan	1998	BR	Ohshima *et al*. (2007)	AB252095
AD853J	*R. sativus*	Ohhata, Aomori, Japan	2002	BR	Ohshima *et al*. (2007)	AB252096
AD855J	*R. sativus*	Ohminato, Aomori, Japan	2002	BR	Ohshima *et al*. (2007)	AB252097
AD860J	*R. sativus*	Sennai, Aomori, Japan	2002	BR	Ohshima *et al*. (2007)	AB252098
AKD161J	*R. sativus*	Ogachi, Akita, Japan	1998	BR	Ohshima *et al*. (2007)	AB252099
AKD934J	*R. sativus*	Hachiryu, Akita, Japan	2000	BR	Ohshima *et al*. (2007)	AB252100
AKH937J	*B. pekinensis*	Yuzawa, Akita, Japan	2000	BR	Ohshima *et al*. (2007)	AB252101
AT181J	*Eustoma russellianum*	Aomori, Aomori, Japan	<1998	BR	Ohshima *et al*. (2007)	AB252102
BJ-C4	*Brassicaceae* sp.	-, Beijing	<1987	Not known		HQ446217
C1	Not known	-, Taiwan , China	Not known	Not known		AF394601
C42J	*B. rapa*	Saga, Saga, Japan	1993	B	Tomimura *et al*. (2003)	AB093625
CH6	*R. sativus*	Zengjiang, Jiangsu, China	1999	BR	Ohshima *et al*. (2007)	AB252103
CHK16	*R. sativus*	Guilin, Guangxi, China	2000	BR	Ohshima *et al*. (2007)	AB252104
CHL13	*R. sativus*	Lushun, Liaoning, China	1999	BR	Ohshima *et al*. (2007)	AB252105
CHN 1	*Brassica* sp.	-, Taiwan, China	<1980	BR	Tomimura *et al*. (2003)	AB093626
CHN 12	Not known	-, -, China	<1990	B	Jenner *et al*. (2002)	AY090660
CHZJ26A	*B. campestris*	Jiande, Zhejiang, China	1999	B(R)	Ohshima *et al*. (2007)	AB252106
CP845J	*Calendula officinalis*	Kisarazu, Chiba, Japan	1997	BR	Tomimura *et al*. (2003)	AB093614
DMJ	*R. sativus*	-, Tochigi, Japan	1996	BR	Tomimura *et al*. (2003)	AB093623
FD27J	*R. sativus*	Fukuoka, Fukuoka, Japan	1998	BR	Tomimura *et al*. (2003)	AB093618
FKD001J	*R. sativus*	Sukagawa, Fukushima, Japan	2000	BR	Ohshima *et al*. (2007)	AB252109
FKD004J	*R. sativus*	Funehiki, Fukushima, Japan	2000	BR	Ohshima *et al*. (2007)	AB252110
FKH122J	*B. pekinensis*	Naraha, Fukushima, Japan	1998	BR	Ohshima *et al*. (2007)	AB252111
GFD462J	*R. sativus*	Yoro, Gifu, Japan	2001	BR	Ohshima *et al*. (2007)	AB252115
H1J	*R. sativus*	Hirosaki, Aomori, Japan	1996	BR	Ohshima *et al*. (2007)	AB252118
HOD517J	*R. sativus*	Kimobetsu, Hokkaido, Japan	1998	BR	Tomimura *et al*. (2003)	AB093617
HRD	*R. sativus*	Hongzhou, Zhejiang, China	1998	BR	Tomimura *et al*. (2003)	AB093627
HZ6	*Brassica* sp.	Xiaoshan, Zhejiang, China	1998	B	Ohshima *et al*. (2007)	AB252119
IWD032J	*R. sativus*	Iwaizumi, Iwate, Japan	2000	BR	Ohshima *et al*. (2007)	AB252120
IWD038J	*R. sativus*	Yahaba, Iwate, Japan	2000	BR	Ohshima *et al*. (2007)	AB252121
Ka1J	*B. pekinensis*	-, Tochigi, Japan	1994	BR	Tomimura *et al*. (2003)	AB093624
KD32J	*R. sativus*	Nankan, Kumamoto, Japan	1998	BR	Tomimura *et al*. (2003)	AB093621
KGD54J	*R. sativus*	Sendai, Kagoshima, Japan	1998	BR	Ohshima *et al*. (2007)	AB252123
KWB778J	*B. oleracea*	Takamatsu, Kagawa, Japan	2004	B	Ohshima *et al*. (2007)	AB252124
KWB779J	*B. rapa*	Takamatsu, Kagawa, Japan	2004	BR	Ohshima *et al*. (2007)	AB252125
KYD073J	*R. sativus*	Mineyama, Kyoto, Japan	2000	BR	Ohshima *et al*. (2007)	AB252126
KYD81J	*R. sativus*	Joyo, Kyoto, Japan	1998	BR	Tomimura *et al*. (2003)	AB093613
Lu2	*Brassica* sp.	-, Shandong	<1990	Not known		HQ446216
MED302J	*R. sativus*	Shiroyama, Mie, Japan	2001	BR	Ohshima *et al*. (2007)	AB252127
MYD013J	*R. sativus*	Yamamoto, Miyagi, Japan	2000	BR	Ohshima *et al*. (2007)	AB252128
MYD015J	*R. sativus*	Kesennuma, Miyagi, Japan	2000	BR	Ohshima *et al*. (2007)	AB252129
ND10J	*R. sativus*	Hirato, Nagasaki, Japan	1998	BR	Ohshima *et al*. (2007)	AB252130
NDJ	*R. sativus*	Takaki, Nagasaki, Japan	1997	BR	Tomimura *et al*. (2003)	AB093616
NID048J	*R. sativus*	Niitsu, Niigata, Japan	2000	BR	Ohshima *et al*. (2007)	AB252131
NID119J	*R. sativus*	Yuzawa, Niigata, Japan	1998	BR	Ohshima *et al*. (2007)	AB252132
NRD350J	*R. sativus*	Gojyo, Nara, Japan	2001	BR	Ohshima *et al*. (2007)	AB252134
RC4	*Zantedeschia* sp.	-, Taiwan, China	2000	BR	Chen *et al*. (2003)	AY134473
SGB088J	*B. rapa*	Hikone, Shiga, Japan	2000	BR	Ohshima *et al*. (2007)	AB252136
SGD311J	*R. sativus*	Nishiazai, Shiga, Japan	1998	BR	Tomimura *et al*. (2003)	AB093619

Table 4. Continued.

Isolate	Original host	Location (City, District, Country)	Year of collection	Host type[a]	Reference	Accession code
SMD060J	*R. sativus*	Gotsu, Shimane, Japan	2000	BR	Tomimura *et al*. (2003)	AB252137
TANX2	*R. sativus*	Tai'an, Shandong, China	2007	BR	Wang *et al*. (2009)	EU734433
TD88J	*R. sativus*	Tokyo, Tokyo, Japan	1998	BR	Tomimura *et al*. (2003)	AB093615
TRD052J	*R. sativus*	Akasaki, Tottori, Japan	2000	BR	Ohshima *et al*. (2007)	AB252138
TRD053J	*R. sativus*	Tomari, Tottori, Japan	2000	BR	Ohshima *et al*. (2007)	AB252139
Tu-2R1	*R. sativus*	-, Tochigi, Japan	Not known	BR	Suehiro *et al*. (2004)	AB105135
Tu-3	*B. oleracea*	-, Tochigi, Japan	Not known	B	Suehiro *et al*. (2004)	AB105134
TW	Not known	-, Taiwan, China	Not known	Not known		AF394602
YAD020J	*R. sativus*	Shirataka, Yamagata, Japan	2000	BR	Ohshima *et al*. (2007)	AB252140
YAL018J	*Lactuca sativa*	Sakae, Yamagata, Japan	2000	BR	Ohshima *et al*. (2007)	AB252141
YC5	*Zantedeschia* sp.	-, Taiwan, China	2000	BR	Chen *et al*. (2003)	AF530055
YMD069J	*R. sativus*	Misumi, Yamaguchi, Japan	2000	BR	Ohshima *et al*. (2007)	AB252142
YMD070J	*R. sativus*	Abu, Yamaguchi, Japan	2000	BR	Ohshima *et al*. (2007)	AB252143
WFLB06	*R. sativus*	Weifang, Shandong, China	2006	BR	Wang *et al*. (2009)	EU734434
Europe						
A102/11	*Anemone coronaria*	-, Liguria, Italy	1993	B	Tomimura *et al*. (2003)	AB093597
A64	*A. coronaria*	-, Liguria, Italy	1991	B	Tomimura *et al*. (2003)	AB093599
Al	*Alliaria officinalis*	-, Piedmont, Italy	1968	(B)	Tomimura *et al*. (2003)	AB093598
Cal1	*C. officinalis*	-, Liguria, Italy	1979	BR	Tomimura *et al*. (2003)	AB093601
CAR37	*Cochlearia armoracia*	-, -, Poland	2004	Not known	Kozubek *et al.* (2007)	DQ648592
CAR37A	*C. armoracia*	-, -, Poland	2004	Not known	Kozubek *et al.* (2007)	DQ648591
CAR39	*C. armoracia*	-, -, Poland	2004	Not known		EF374098
CAR51	*Armoracia rusticana*	-, -	2004	Not known		HQ637383
CZE 1	*B. oleracea*	Ruzyne, -, Czech Republic	1981	B	Tomimura *et al*. (2003)	AB093608
CZE 5	*B. rapa*	Ceske Budejovice, -, Czech Republic	1993	B	Ohshima *et al*. (2007)	AB252107
DNK 2	*B. napus*	-, -, Denmark	<1993	B	Ohshima *et al*. (2007)	AB252108
FRD 1	*B. oleracea*	-, -, Germany	1987	B	Ohshima *et al*. (2007)	AB252112
GBR 36	*B. oleracea*	Winspit, Dorset, UK	18.6.1999	B	Ohshima *et al*. (2007)	AB252113
GBR 50	wild *B. oleracea*	Staithes, Yorkshire, UK	28.9.1999	B	Ohshima *et al*. (2007)	AB252114
GBR 98	wild *B. oleracea*	Winspit, Dorset, UK	28.8.2002	B	Pallett *et al*. (2007)	EU861593
GRC 17	*B. oleracea*	Volos, -, Greece	1993	B	Ohshima *et al*. (2007)	AB252116
GRC 42	wild *Allium* sp.	-, -, Greece	1999	B	Ohshima *et al*. (2007)	AB252117
ITA 3	*B. ruvo*	-, Campania, Italy	1990	B	Ohshima *et al*. (2007)	AB252122
ITA 7	*R. raphanistrum*	-, Campania, Italy	1990	BR	Tomimura *et al*. (2003)	AB093600
NLD 1	*B. oleracea*	-, -, The Netherlands	<1995	B	Ohshima *et al*. (2007)	AB252133
PV0104	*L. sativa*	Stuttgart, -, Germany	1986	BR	Tomimura *et al*. (2003)	AB093603
PV376-Br	*B. napus*	Braunschweig, -, Germany	1970	B	Tomimura *et al*. (2003)	AB093604
Rn98	*Ranunculus asiaticus*	-, Liguria, Italy	1997	B	Ohshima *et al*. (2007)	AB252135
RUS 1	*Armoracia rusticana*	-, -, Russia	1993	B	Tomimura *et al*. (2003)	AB093606
RUS 2	*B. napus*	Moscow, -, Russia	Not known	B	Tomimura *et al*. (2003)	AB093607
St48	*Limonium sinuatum*	-, Toscana, Italy	1993	B	Tomimura *et al*. (2003)	AB093596
UK 1	*B. napus*	-, Warwickshire, UK	1975	B	Jenner *et al*. (2000)	AF169561
Other						
BZ1	*B. oleracea*	-, Federal, Brazil	1996	B	Tomimura *et al*. (2003)	AB093611
CDN 1	*B. napus napobrassica*	-, -, Canada	<1988	B	Jenner *et al*. (2003)	AY227024
IRNTRa6	*Rapistrum rugosum*	Varamin, Tehran, Iran	2004	B	Farzadfar *et al*. (2008)	AB440238
IRNSS5	*Sisymbrium loeselii*	Semnan, Semnan, Iran	2003	B	Farzadfar *et al*. (2008)	AB440239
IS1	*Allium ampeloprasum*	-, -, Israel	1993	B	Tomimura *et al*. (2003)	AB093602
KEN 1	*B. oleracea*	-, -, Kenya	1994	B	Tomimura *et al*. (2003)	AB093605
NZ290	*B. pekinensis*	-, Alan Stewart, New Zealand	1998	B	Tomimura *et al*. (2003)	AB093612
Q-Ca	*B. rapa*	-, -, Canada	Not known	Not known	Nicolas and Laliberté (1992)	D10927
TUR1	*B. oleracea*	Canakkale, Marmora, Turkey	2004	B	Korkmaz *et al*. (2008)	AB362512
TUR9	*R. sativus*	Balikesir, Marmora, Turkey	2005	BR	Korkmaz *et al*. (2008)	AB362513
USA 1	*B. oleracea*	-, -, USA	<1980	B	Tomimura *et al*. (2003)	AB093609

[a] Host type B; *Brassica*, isolates infected *B. rapa* cv. Hakatasuwari systemically giving mosaic symptoms. Host type (B); isolates infected *B. rapa* only occasionally. Host type BR; these isolates infected both *B. rapa* and *R. sativus* systemically giving mosaic symptoms. Host type B(R); isolates infected *B. rapa* systemically giving mosaic symptoms and infected *R. sativus* only occasionally.

Finally, sequences were then reassembled to form complete ORF sequences of 9321 nt. The aligned sequences of the 5' and 3' non-coding regions were then added.

4.2. Recombination analysis

Recombination in the genomic sequences was investigated using the split-decomposition method implemented in SPLITSTREE version 4.11.3 (Hunson and Bryant, 2006). Putative recombination breakpoints in all sequences were identified using RDP (Martin and Rybicki, 2000), GENECONV (Sawyer, 1999), BOOTSCAN (Salminen *et al*., 1995), MAXCHI (Maynard-Smith, 1992), CHIMAERA (Posada and Crandall, 2001) and SISCAN programs (Gibbs *et al*., 2000) implemented in the RDP3 package (Martin *et al*., 2010) and also the original PHYLPRO version 1 (Weiller, 1998) and SISCAN version 2 (Gibbs *et al*., 2000) programs. First, incongruent relationships were checked using the programs implemented in RDP3. These analyses were done using default settings for the different detection programs and a Bonferroni corrected *P*-value cut-off of 0.05 or 0.01, and then all isolates that had been identified as likely recombinants by the programs in RDP3, supported by three different methods with an associated *P*-value of $>1.0 \times 10^{-6}$, were re-checked using the original PHYLPRO version 1 and SISCAN version 2, not only with all nucleotide sites, but also with synonymous and non-synonymous sites separately. 100 and 50 nucleotides slices of all sequences for evidence of recombination were checked using these programs. These analyses also assessed which non-recombinant sequences had regions that were the closest to the regions of the recombinant sequences and hence indicated the likely lineages that provided those regions of the recombinant genomes. For convenience, they were called 'parental isolates' of recombinants. Having examined all sites with an associated *P*-value of $<1 \times 10^{-6}$ (i.e. the most likely recombination sites), the intralineage recombinants (parents

from the same major lineage) were retained and the interlineage recombinants (parents from different major lineages) removed by treating the identified recombination sites as missing data in subsequent analyses. The aligned 5' and 3' UTR sequences were added to the ends of the polyprotein sequences to reassemble nearly complete genomic sequences, which were then assessed again for evidence of recombination, especially for recombination sites in the UTRs. Finally, TuMV sequences were also aligned without an outgroup sequence, and directly checked for evidence of recombination using the programs.

4.3. Phylogenetic analysis

The phylogenetic relationships of the aligned sequences were inferred by the maximum likelihood method (ML) implemented in PhyML version 2.4.4 and 3.0 (Guindon and Gascuel, 2003) using the general time reversible (GTR) scheme of substitution with a gamma distribution (r_4) model of site rates variation (GTR +r_4). The best-fit model of nucleotide substitutions for each dataset was determined using jModeltest 0.1.1 (Posada, 2008). For ML analyses, branch support was evaluated by the bootstrap method based on 1000 pseudoreplicates. The calculated trees were displayed by TREEVIEW (Page, 1996).

4.4. Nucleotide difference and diversity

The nucleotide and amino acid difference were estimated using the Kimura two-parameter method (Kimura, 1980) and the Dayhoff PAM 001 matrix (Schwarz and Dayhoff, 1979), and the within-population diversities were assessed using MEGA version 5 (Tamura *et al*., 2011). The nucleotide identities of pairwise sequences were compared using Needle program availabled at http://www.ebi.ac.uk/Tools/psa/emboss_needle/.

4.5. Genetic differentiation and gene flow

Genetic differentiation between TuMV populations was examined by three permutation-based statistical tests, *K*s*, *Z* and Snn (Hudson, 2000). The extent of genetic differentiation or the level of gene flow between populations was measured by estimating F_{ST} (the interpopulational component of genetic variation or the standardized variance in allele frequencies across populations) using DnaSP version 5.0 (Librado and Rozas, 2009). The absolute value of F_{ST} ranges from 0 to 1 for undifferentiated to fully differentiated populations. Normally, an absolute value of $F_{ST} > 0.33$ suggests infrequent gene flow, while absolute value of $F_{ST} < 0.33$ suggests frequent gene flow (Librado and Rozas 2009; Wei *et al*., 2009).

The geographical and spatial patterns of genetic differentiation were also evaluated by performing analysis of molecular variance (AMOVA) using Arlequin version 3.0 (Excoffier *et al*, 2005). The statistical significance of the results was tested by 10 000 non-parametric data permutations.

4.6. Population demography analysis

DnaSP version 5.0 (Librado and Rozas, 2009) was used to estimate Tajima's *D* (Tajima, 1989), Fu and Li's *D** and *F** statistical tests (Fu and Li, 1993), and haplotype diversity. Tajima's *D*, Fu and Li's *D** and *F** tests hypothesize that all mutations are selectively neutral. Haplotype diversity refers to the frequency and number of haplotypes in the population. Nucleotide diversity estimates the average pairwise differences among sequences.

4.7. Substitution rates and divergence times

Substitution rates and divergence times were estimated from various subsets of the sequence data. Individual alignments of the HC-Pro, protein 3 (P3), NIb, and cCP genes, which included sequences from both Orchis and brassica-infecting TuMV (BI) isolates, were analysed. For some analyses, however, TuMV isolates from brassicas and those from non-brassicas were distinguished. The Bayesian phylogenetic method in BEAST version 1.4.7 (Drummond and Rambault, 2007) was used to estimate substitution rates and TMRCAs. Data sets were analysed using both strict and relaxed (uncorrelated exponential and uncorrelated lognormal) molecular clocks (Drummond *et al*., 2007). Path-O-Gen version 1.3 (http://tree.bio.ed.ac.uk/software/pathogen/) was used to test for clocklike evolution using regression of root-to-tip distances on viral sampling times. Small R-squared values were obtained for all alignments, indicating that it was necessary to use relaxed-clock models. Bayes factors were used to compare five demographic models (constant population size, expansion growth, exponential growth, logistic growth, and the Bayesian skyline plot), which were used as coalescent priors.

Posterior distributions of parameters, including the tree, were estimated by Markov chain Monte Carlo sampling. Samples were drawn every 10,000 steps over a total of 100 million steps, with the first 10% of samples discarded as burn-in. Sufficient sampling from the posterior and convergence to the stationary distribution were checked using Tracer (http://tree.bio.ed.ac.uk/software/tracer/).

To obtain reliable estimates of substitution rates and divergence times from time-stamped data, the range of sampling times needs to have been sufficient for genetic change to have occurred (Drummond *et al*., 2003). To test for adequate temporal structure in our

data sets, the rate estimates in this study was compared with those from ten date-randomized replicates. Following previous studies (Ho *et al*., 2011; Pagán and Holmes, 2010; Pagán *et al*., 2010), a data set was considered to have sufficient temporal structure when the mean rate estimate from the original data set was not contained in any of the 95% credibility intervals of the rates estimated from the date-randomized replicates.

V. RESULTS

1. Spatial evolution of *Turnip mosaic virus* population of Vietnam

1.1. Occurrence of *Turnip mosaic virus* in Brassicaceae crops in Vietnam

A total of 183 samples collected during the 2006–2008 growing seasons were tested by DAS-ELISA using the antiserum to isolate 59J (Ohshima *et al*., 1996; 2002). Of those tested, 139 plants (76%) including *B. juncea* (mustard) and *R. sativus* (radish) were found to be infected with TuMV (Table 5). No infection was detected in *B. oleracea* var. *capitata* (cabbage), *B. oleracea* var. *gongylodes* (kohlrabi) and *B. oleracea* vars *botrytis* (cauliflower). Viruses were found in commercial fields as well as in home gardens. When many symptomatic plants were found in a field or in a home garden, only some samples were randomly chosen for DAS-ELISA detection. This study only collected symptomatic plants in the present study, thus there might be plants which were latently infected by TuMV in the fields and home gardens.

1.2. Biological and molecular characteristics of the Vietnamese isolates

A total of 30 TuMV isolates collected in Vietnam were examined in this study; two from the Northwest, six from the Northeast, 17 from Red river delta, one from the North central coast, one from the South central coast and three from the Central highlands (Figure 12 and Table 1), thus isolates were collected from most parts of Vietnam where Brassicaceae crops are cultivated. Nine were isolated from radish and 21 were from mustard.

All the Vietnamese isolates from radish and mustard infected *B. juncea* cv. Hakarashina

Table 5. Occurrence of *Turnip mosaic virus* in Brassicaceae crops in different provinces of Vietnam[a].

Province	Host (common name)	No. of samples infected/collected	Collection year
Central highland			
Dak Lak	*Raphanus sativus* (radish)	10/10 (100)[b]	2006
North central coast			
Thua Thien Hue	*Brassica juncea* (mustard)	10/13 (76.9)	2006, 2007
	R. sativus	0/3 (0.0)	
Northeast			
Bac Giang	*B. juncea*	9/9 (100)	2007
Ha Giang	*R. sativus*	0/1 (0.0)	2006
Lang Son	*B. juncea*	5/5 (100)	2008
Northwest			
Son La	*B. juncea*	9/9 (100)	2006
Red river delta			
Bac Ninh	*B. juncea*	12/17 (70.6)	2006, 2007
	R. sativus	1/6 (16.7)	
Ha Noi	*B. juncea*	26/26 (100)	2006, 2007
	B. oleracea var. *capitata* (cabbage)	0/1 (0.0)	
	R. sativus	19/19 (100)	
Ha Tay	*B. juncea*	4/6 (66.7)	2007
Hai Duong	*B. juncea*	0/10 (0.0)	2006
	R. sativus	2/3 (66.6)	
Hung Yen	*B. oleracea* var. *botrytis* (cauliflower)	0/2 (0.0)	2006
	B. oleracea var. *gongylodes* (kohrabi)	0/3 (0.0)	
	B. juncea	4/6 (66.7)	
	R. sativus	5/5 (100)	
Nam Dinh	*B. juncea*	3/3 (100)	2008
Ninh Binh	*B. juncea*	0/3 (0.0)	2006
Thai Binh	*B. juncea*	10/10 (100)	2007
South central coast			
Quang Nam	*B. juncea*	5/5 (100)	2006, 2007
	R. sativus	5/8 (62.5)	
Total[c]		139/183 (76.0)	

[a] Identification is based on double antibody sandwich enzyme-linked immunosorbent assay (DAS-ELISA).

[b] Percent of virus infection rate.

[c] Average of virus infection.

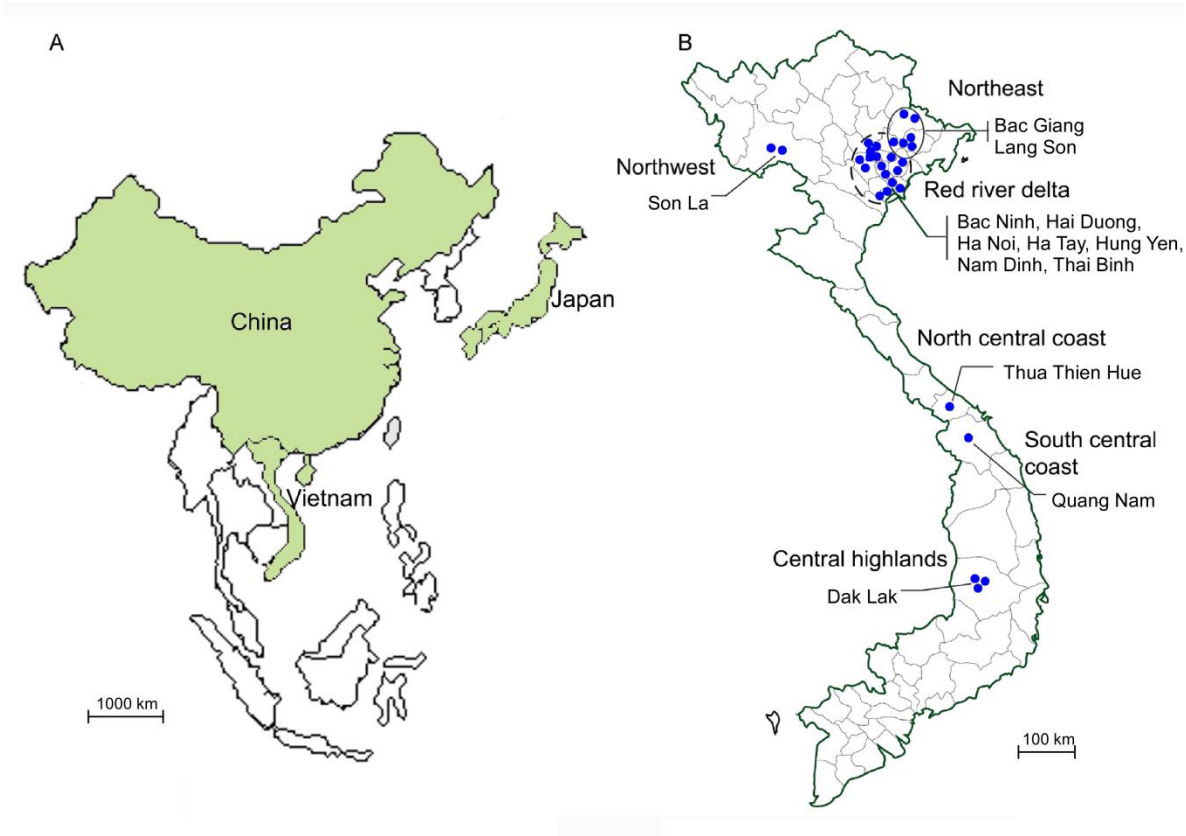


Figure 12. Location of the sampling sites of *Turnip mosaic virus* isolates used in this study; maps of (A) Asia including China and Japan (B) Vietnam, where blue dots indicate the sampling sites.

and *B. rapa* cv. Hakatasuwari plants but did not infect *B. oleracea* var. capitata cvs. Ryozan 2-go and Shinsei. The isolates from radish did not always infect Japanese radish (*R. sativus* cvs. Akimasari and Taibyo-soubutori), and those from mustard only occasionally infected them. Because there is no local radish cultivar in Vietnam and Chinese radish cultivars are most widely grown. Chinese radish (*R. sativus* cv. Everest) was inoculated with the Vietnamese isolates and found to be susceptible to most isolates producing mosaic symptoms (data not shown). This indicates the importance of using local cultivars for inoculation tests.

The 30 full genomes sequenced in this study were analyzed, as well as 105 full genomic TuMV sequences obtained from the public DNA sequence databases. All the genomes of Vietnamese isolates of TuMV were 9798 nucleotides excluding 5' end 35 nt primer sequences; the regions encoding the P1, HC-Pro, P3, PIPO, 6Kda 1 protein (6K1), cylindrical inclusion (CI), 6Kda 2 protein (6K2), genome linked viral protein (VPg), nuclear inclusion a-proteinase protein (NIa-Pro), NIb and CP proteins were 1086, 1374, 1065, 177, 156, 1932, 159, 576, 729, 1551 and 864 nucleotides long respectively, and, for most analyses, the 5' end primer sequences were omitted leaving only the region encoding the polyprotein.

All of the motifs reported for different potyvirus encoded proteins were found. The new genomic sequences determined in this study are available in DBJ/EMBL/GenBank databases with Accession Codes AB747286–AB747315.

1.3. Recombination

The polyprotein-encoding gene sequences of 30 Vietnamese isolates and 105 sequences obtained from the public DNA sequence databases were assessed for evidence of recombination.

Eight unequivocal recombination sites were found in the genome of 14 Vietnamese isolates (Table 6 and Figure 13). These were located in the P1, HC-Pro, P3, VPg and CP encoding regions. They were supported by clear *P*-values (<1 x 10^{-6}) in the RDP, GENECONV, BOOTSCAN, MAXCHI, CHIMAERA and SISCAN programs of the RDP3 software. One tentative recombination site was found in the genome of VIET176 isolate located in the CI encoding region; this recombination site gave *P*-values in the RDP (1.33 x 10^{-7}), MAXCHI (3.07 x 10^{-13}), CHIMAERA (5.60 x 10^{-10}) and SISCAN (5.29 x 10^{-6}) programs but did not give significant *P*-values (<1 x 10^{-3}) in the BOOTSCAN and GENECONV in RDP3, and was not detected by original SISCAN (*Z*-value, <3.0) and PHYLPRO programs. Thus nine recombination patterns were identified in genomes from the Vietnamese population. Two recombinants were interlineage and the remaining seven were intralineage recombinants with world-B and Asian-BR parents. Six of the recombination sites (nt 1511, 1631, 2499, 3461, 3534, 5536) had not been found before. The two interlineage recombination sites (nt 776, 8907), which were in the P1 and CP encoding regions were identical to two found in earlier studies of Asian (Chinese and Japanese) populations (Ohshima *et al*., 2002, 2007; Tomitaka and Ohshima, 2006), and the remaining one recombination site (an intralineage site at nt 5989) had also been found before in an east European (Czech Republic) population, (Ohshima *et al*., 2007).

Table 6. Recombination sites in Vietnamese *Turnip mosaic virus* genomes.

Protein encoding region and recombination site[a]	Isolate	Type of recombinant[b]	Parental isolate	Programs detected by[c]	*Z*-value[d]	*P*-value[e]
P1						
nt 776	VIET82	wB x ABR (interlineage)	VIET159 x HRD	RG**B**MCS$_R$S$_O$P	6.76	3.58 x 10^{-45}
	VIET83	wB x ABR (interlineage)	VIET159 x HRD	RG**B**MCS$_R$S$_O$P	6.57	3.36 x 10^{-45}
	VIET89	wB x ABR (interlineage)	VIET159 x HRD	RG**B**MCS$_R$S$_O$P	6.57	3.82 x 10^{-45}
	VIET15	wB x ABR (interlineage)	VIET159 x HRD	RG**B**MCS$_R$S$_O$P	7.70	8.96 x 10^{-79}
	VIET65	wB x ABR (interlineage)	VIET159 x HRD	RG**B**MCS$_R$S$_O$P	7.70	1.44 x 10^{-77}
	VIET66	wB x ABR (interlineage)	VIET159 x HRD	RG**B**MCS$_R$S$_O$P	7.70	8.12 x 10^{-76}
	VIET73	wB x ABR (interlineage)	VIET159 x HRD	RG**B**MCS$_R$S$_O$P	7.00	4.87 x 10^{-74}
HC-Pro						
nt 1511	VIET166	wB x wB (intralineage)	VIET153 x VIET172	**R**GBMCS$_R$S$_O$	5.35	1.51 x 10^{-27}
nt 1631	VIET159	wB x wB (intralineage)	YAL018J x HZ6	RBMC**S$_R$**S$_O$	3.76	1.99 x 10^{-24}
nt 2499	VIET79	wB x wB (intralineage)	VIET173 x VIET172	RGBMC**S$_R$**S$_O$	3.43	5.50 x 10^{-28}
P3						
nt 3461	VIET169	wB x wB (intralineage)	VIET164 x VIET158	RGBMC**S$_R$**S$_O$	4.72	2.27 x 10^{-17}
	VIET159	wB x wB (intralineage)	HZ6 x YAL018J	RGBMC**S$_R$**S$_O$	3.76	1.03 x 10^{-19}
nt 3534	VIET167	wB x wB (intralineage)	CHZJ26A x VIET158	**R**GBMCS$_R$S$_O$	5.74	5.66 x 10^{-31}
CI						
nt 5536 (tentative)	VIET176	wB x wB (intralineage)	VIET80 x BJ-C4	R**M**CS$_R$	<3.00	3.07 x 10^{-13}
VPg						
nt 5989	VIET160	wB x wB (intralineage)	VIET153 x VIET175	R**B**CS$_R$S$_O$	4.45	5.57x10^{-09}
	VIET164	wB x wB (intralineage)	VIET153 x VIET175	R**B**MCS$_R$S$_O$	4.47	1.21x10^{-07}
CP						
nt 8907	VIET82	ABR x wB (interlineage)	CHL13 x HZ6	RG**B**MS$_R$S$_O$P	6.63	5.48 x 10^{-39}
	VIET83	ABR x wB (interlineage)	CHL13 x HZ6	RG**B**MS$_R$S$_O$P	6.63	5.48 x 10^{-39}
	VIET89	ABR x wB (interlineage)	CHL13 x HZ6	RG**B**MS$_R$S$_O$P	6.63	5.84 x 10^{-39}

[a] Recombination sites detected in the TuMV genomes by the recombination detecting programs (listed in column 5), from the aligned sequences of the likely recombinant and its 'parental isolates'. The nucleotide position shows locations of individual genes numbered as in UK1 genome (Jenner *et al*., 2000).

[b] wB; world-B, ABR; Asian-BR

[c] Recombinant isolates identified by the recombination detecting programs: R (RDP), G (GENECONV), B (BOOTSCAN), M (MAXCHI), C (CHIMAERA) and S_R (SISCAN) programs in RDP3 software (programs greater than *P*-values 1.0 x 10^{-6} are listed), S_O (SISCAN total nucleotide site analysis) in original SISCAN version 2 and P (PHYLPRO) programs. The analyses were done using default settings and a Bonferroni-corrected *P*-values cut-off of 0.05 or 0.01 in RDP3 software.

[d] Both of the parental isolates of the recombination sites had *Z*-values greater than 3 in total nucleotide site analysis in S_o of the SISCAN version 2 program ; thus lower *Z*-value of one of the parents identified are shown.

[e] The reported *P*-value is for the program in red bold type and underlined in RDP3 and is the greatest *P*-value among the isolates calculated for the region in question.

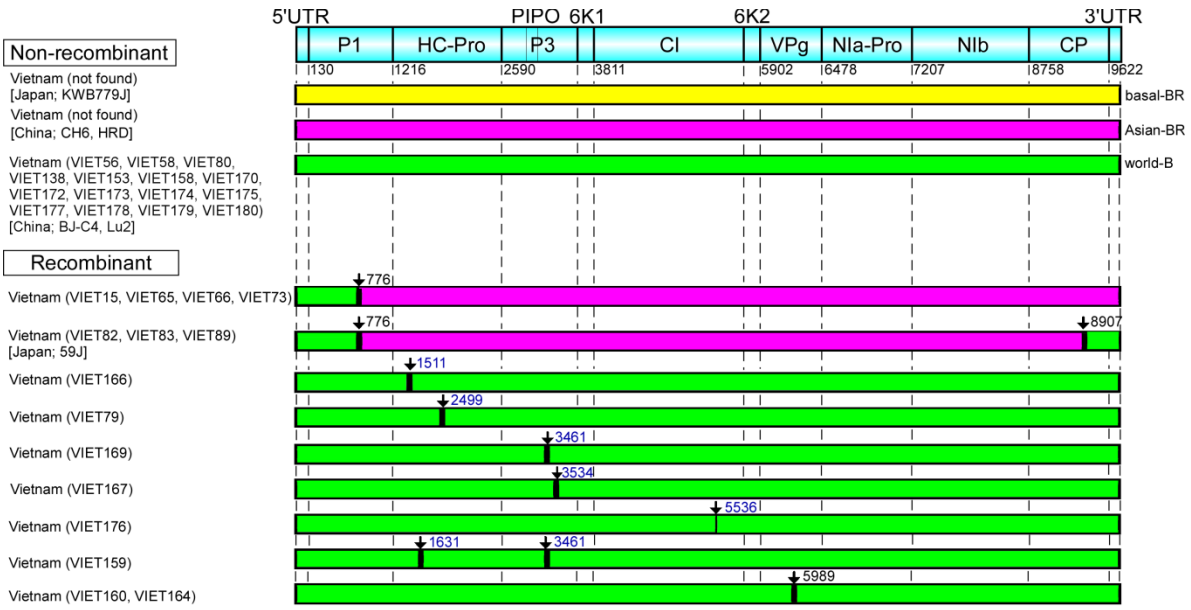


Figure 13. Recombination maps of *Turnip mosaic virus* genomes of Vietnamese isolates; the estimated nucleotide positions of the recombination sites are shown relative to the 5' end of the genome using numbering of the sequence of the UK1 isolate (Jenner *et al*., 2000). The vertical arrows show estimated recombination sites. The yellow, pink and green boxes indicate, respectively, basal-BR, Asian-BR and world-B parents. Bold and thin vertical lines show 'clear' and 'tentative' recombination sites; 'clear recombination sites' were those detected by four or more different programs and two or more types of method, whilst 'tentative recombination sites' were those detected by three or fewer programs and only one or two types of method. The recombination sites newly identified in the present study (acronyms in blue) or those identified in earlier studies (acronyms in black) are separately listed. Some isolates with each recombination pattern, and previously found in Chinese and Japanese population, are also shown.

1.4. Phylogenetic relationships

Phylogenetic trees were initially calculated from the genome sequences of the 135 isolates, including all of the recombinants identified in this study. However, the resulting trees had poor bootstrap support for some lineages, as found previously (Ohshima *et al*., 2002, 2007). Therefore, the trees were recalculated from the genomic sequences of only the 33 isolates that were not recombinants. The relationships of these isolates were investigated by ML and the resulting tree is shown in Figure 14. It partitioned most of the sequences into the same four major consistent genetic groups: basal-B, basal-BR, Asian-BR and world-B, as reported previously (Ohshima *et al*., 2002, 2007). Fifteen out of 30 Vietnamese isolates (50%) were non-recombinants and fell into world-B group. The world-B group further split into three subgroups; East Europe, worldwide and Asia. All the Vietnamese non-recombinant isolates fell into the Asia subgroup and clustered with Chinese isolates. Basal-BR and Asian-BR non-recombinants which had previously been found in Japanese and Chinese population respectively were not found in Vietnamese population.

To check this conclusion in more detail we, therefore, constructed phylogenetic trees of 118–129 non-recombinant isolates using the genomic regions that gave the least evidence of recombination (HC-Pro, P3, NIb and CP encoding regions) which were called the non-recombinogenic regions (Figure 15A-D). These trees showed that none of the Vietnamese isolates had regions from basal-B or basal-BR parental lineages, whereas some Chinese and Japanese isolates had genomic regions that came from basal-BR group parents, as indicated in earlier studies (Tomitaka and Ohshima, 2006; Tomitaka *et al*., 2007; Wang *et al*., 2009).

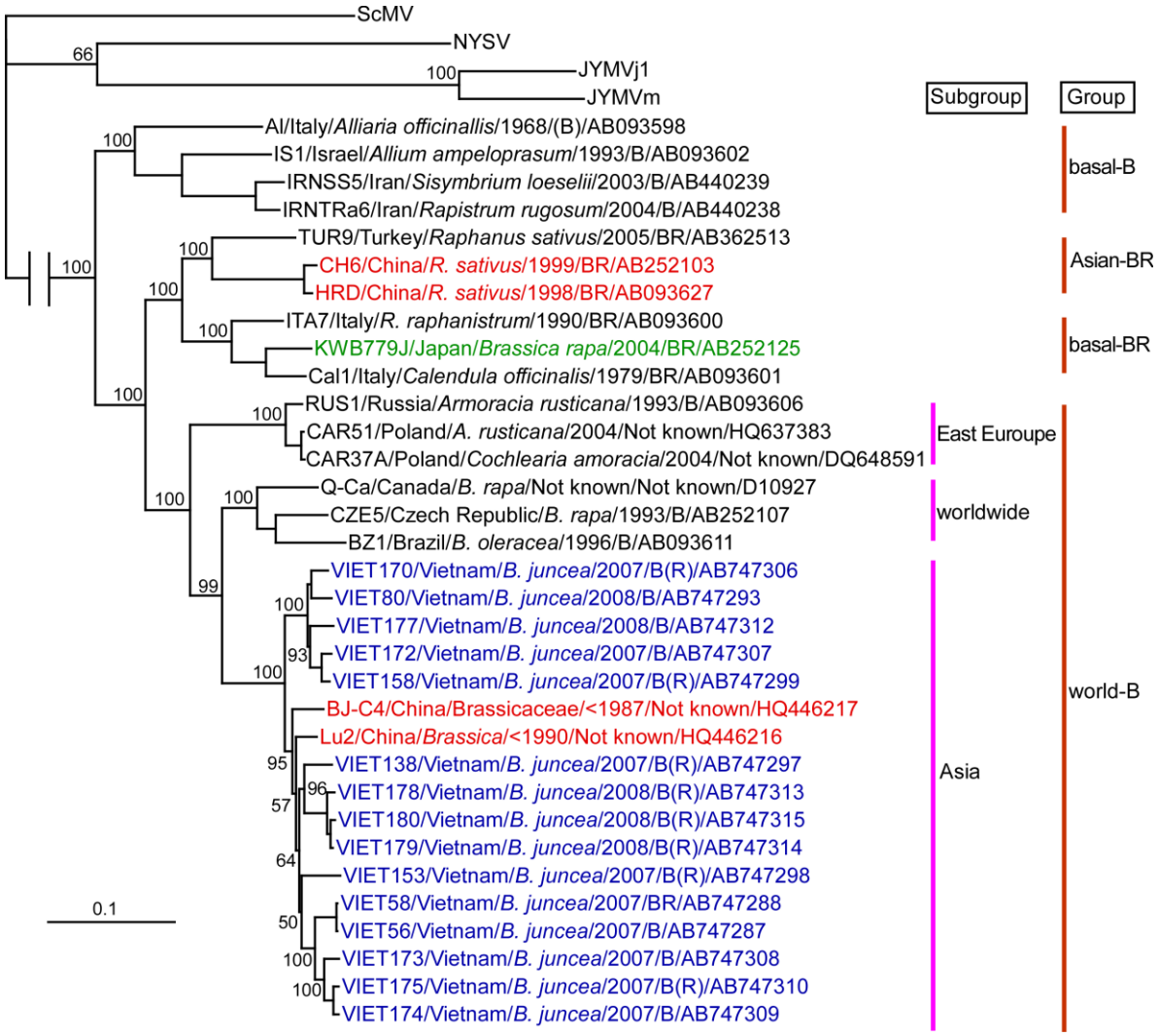
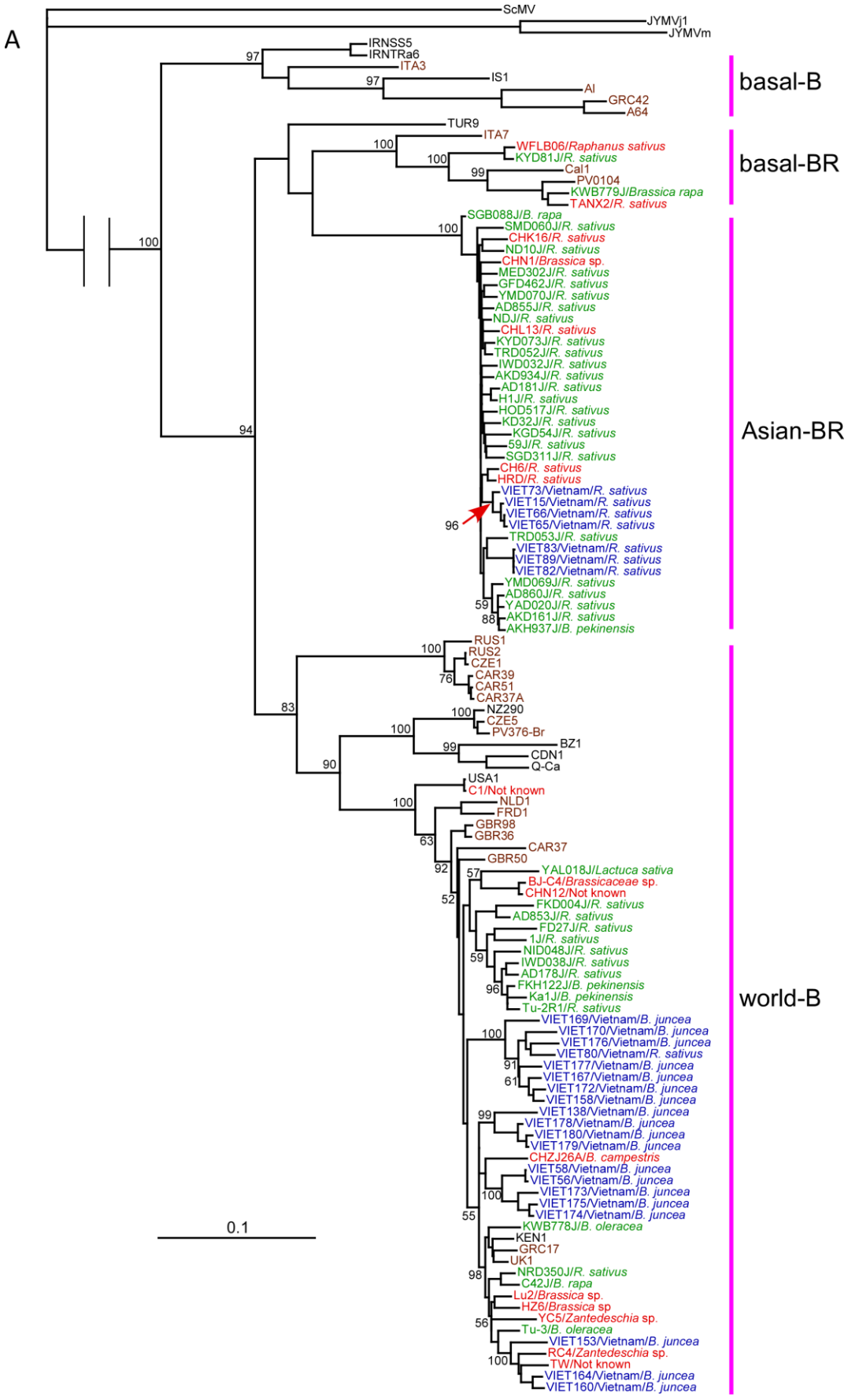
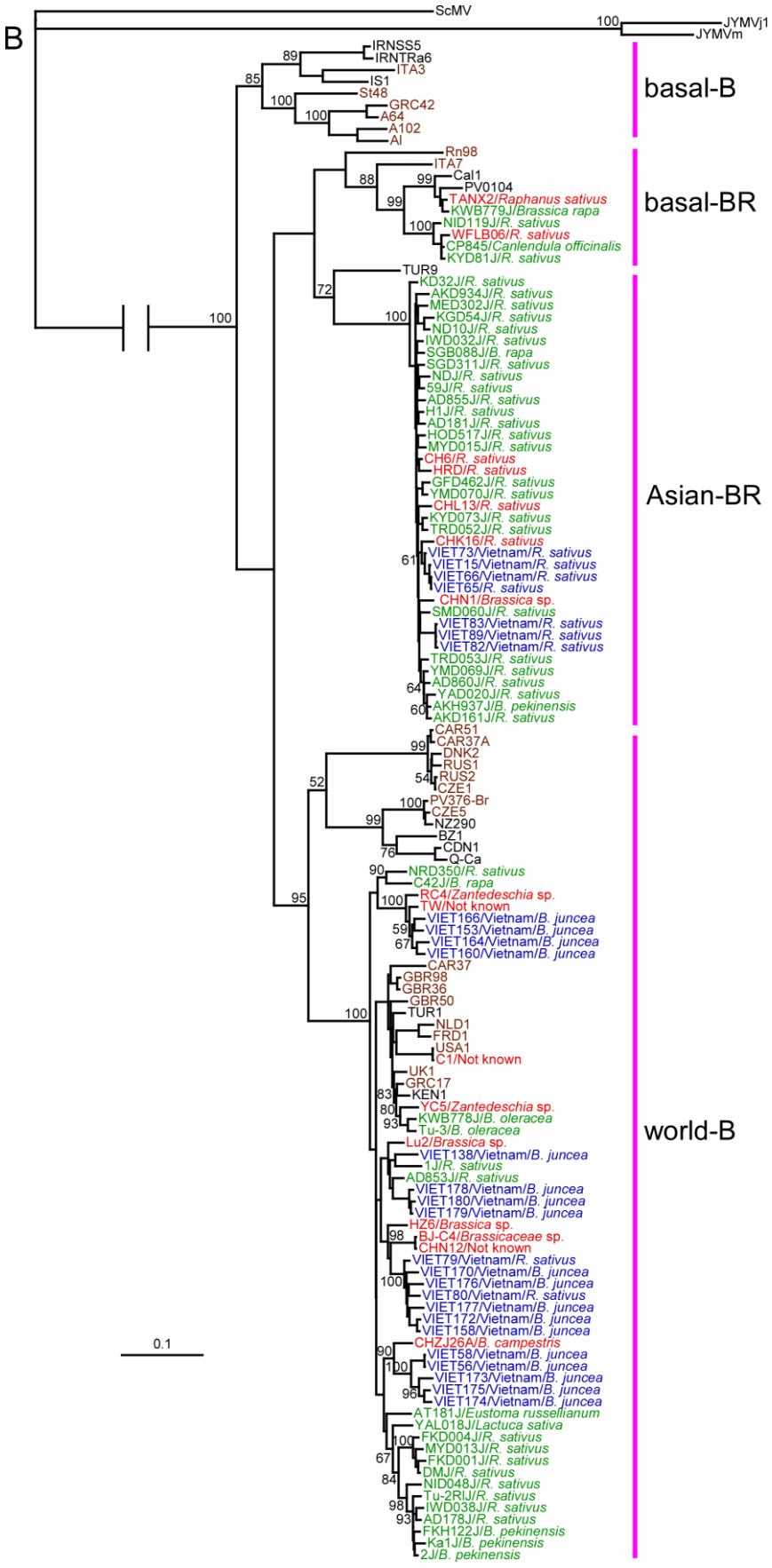


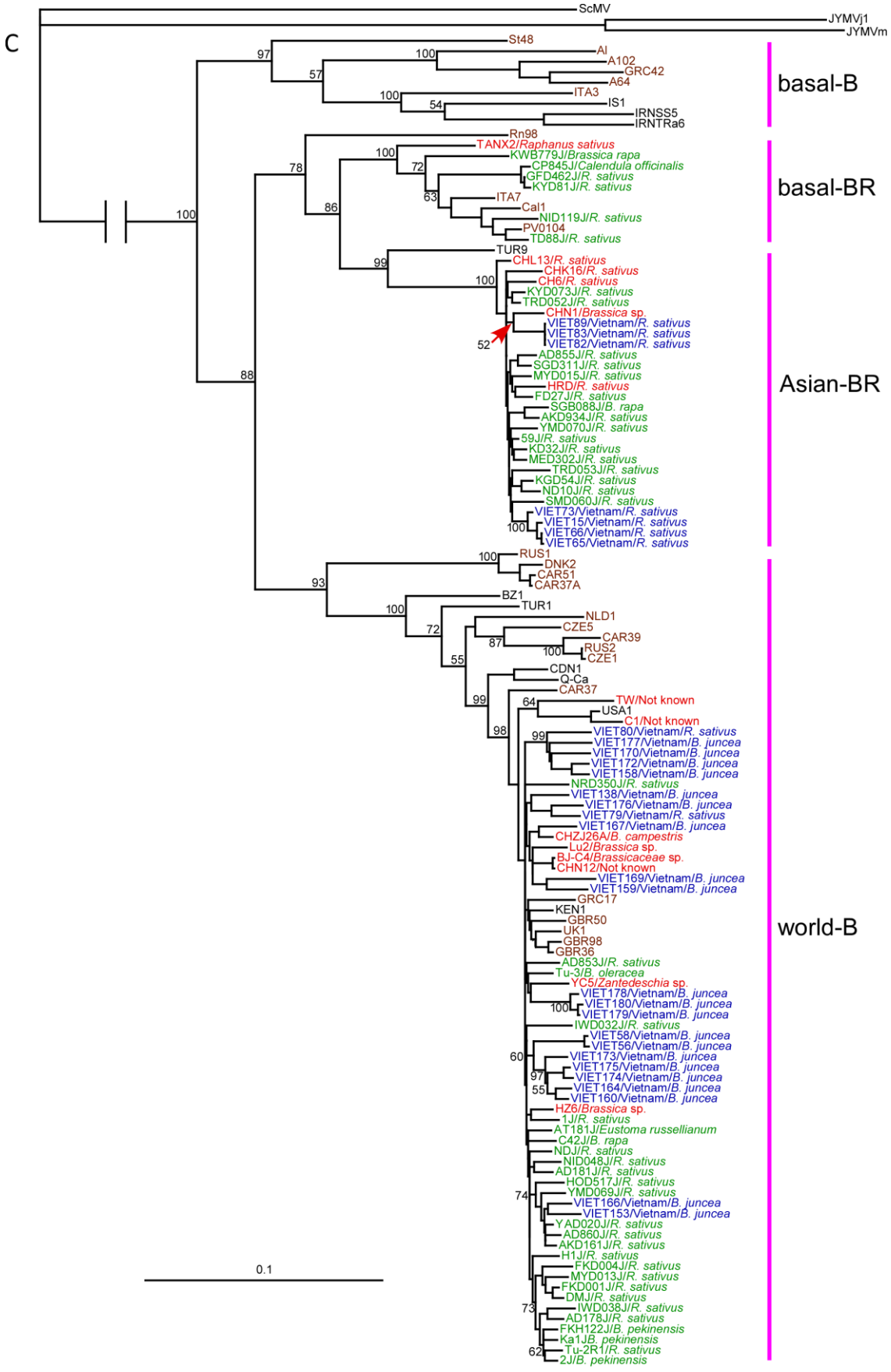
Figure 14. Maximum-likelihood tree calculated from the polyprotein sequences of 33 isolates of *Turnip mosaic virus*, after recombinants identified in this study and earlier study (Ohshima *et al*., 2007) had been discarded. Numbers at each node indicate bootstrap percentages based on 1000 psuedoreplicates (only percentages greater than and equal to 50 are shown) in maximum-likelihood tree. Horizontal branch length is drawn to scale with the bar indicating 0.1 nt replacements per site. The homologous two sequences of *Japanese yam mosaic virus* (JYMV) (Fuji and Nakamae, 1999, 2000), one of *Scallion mosaic virus* (ScMV) (Chen *et al*., 2002) and one of *Narcissus yellow stripe virus* (NYSV) (Chen *et al*., 2003b) were used as the outgroup. The isolates collected in China (acronyms in red), Japan (acronyms in green), Vietnam (acronyms in blue) and other county (acronyms in black) are separately listed. For the detailed of isolates, refer Tables 1 and 4.

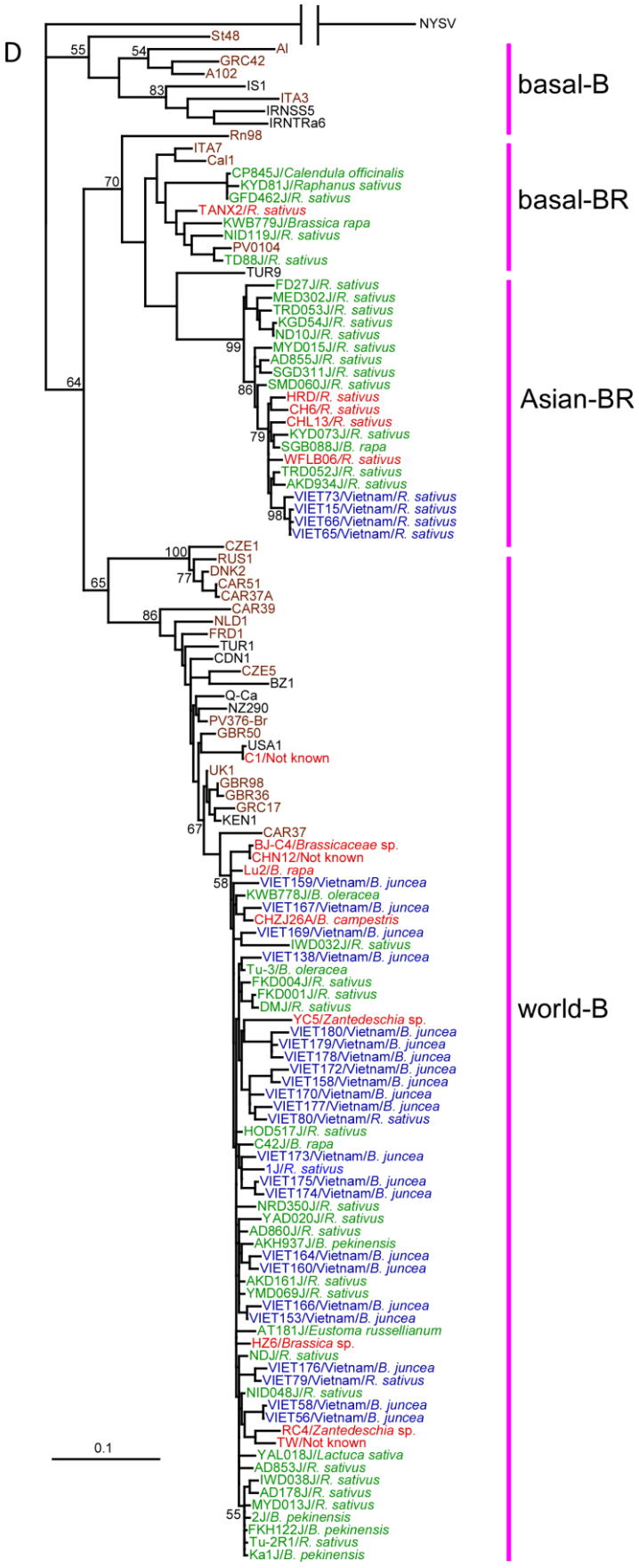
Figure 15. Maximum-likelihood trees calculated from the partial genomic sequences of non-recombinant sequences: (A) helper component-proteinase protein (HC-Pro), (B) protein 3 (P3), (C) nuclear inclusion b protein (NIb) and (D) coat protein (CP). Numbers at each node indicate bootstrap percentages based on 1000 psuedoreplicates (only percentages greater than and equal to 50 are shown) in maximum-likelihood tree. Horizontal branch length is drawn to scale with the bar indicating 0.1 nt replacements per site. The homologous two sequences of *Japanese yam mosaic virus* (JYMV) (Fuji and Nakamae, 1999, 2000), one of *Scallion mosaic virus* (ScMV) (Chen *et al*., 2002) and one of *Narcissus yellow stripe virus* (NYSV) (Chen *et al*., 2003b) were used as the outgroup for HC-Pro, P3 and CP trees, whereas that of NYSV for CP tree. The isolates collected in China (acronyms in red), Europe (acronyms in black), Japan (acronyms in green), Vietnam (acronyms in blue) and other county (acronyms in brown) are separately listed. The original host plants of the isolates collected in Asia are also listed. For the detailed of isolates, refer Tables 1 and 4.

→


A
ScMV
JYMVj1
JYMVm
IRNSS5
IRNTRa6
ITA3
IS1
AI
GRC42
A64
basal-B
TUR9
ITA7
WFLB06/Raphanus sativus
KYD81J/R. sativus
Cal1
PV0104
KWB779J/Brassica rapa
TANX2/R. sativus
basal-BR
SGB088J/B. rapa
SMD060J/R. sativus
CHK16/R. sativus
ND10J/R. sativus
CHN1/Brassica sp.
MED302J/R. sativus
GFD462J/R. sativus
YMD070J/R. sativus
AD855J/R. sativus
NDJ/R. sativus
CHL13/R. sativus
KYD073J/R. sativus
TRD052J/R. sativus
IWD032J/R. sativus
AKD934J/R. sativus
AD181J/R. sativus
H1J/R. sativus
HOD517J/R. sativus
KD32J/R. sativus
KGD54J/R. sativus
59J/R. sativus
SGD311J/R. sativus
CH6/R. sativus
HRD/R. sativus
VIET73/Vietnam/R. sativus
VIET15/Vietnam/R. sativus
VIET66/Vietnam/R. sativus
VIET65/Vietnam/R. sativus
TRD053J/R. sativus
VIET83/Vietnam/R. sativus
VIET89/Vietnam/R. sativus
VIET82/Vietnam/R. sativus
YMD069J/R. sativus
AD860J/R. sativus
YAD020J/R. sativus
AKD161J/R. sativus
AKH937J/B. pekinensis
Asian-BR
RUS1
RUS2
CZE1
CAR39
CAR51
CAR37A
NZ290
CZE5
PV376-Br
BZ1
CDN1
Q-Ca
USA1
C1/Not known
NLD1
FRD1
GBR98
GBR36
CAR37
GBR50
YAL018J/Lactuca sativa
BJ-C4/Brassicaceae sp.
CHN12/Not known
FKD004J/R. sativus
AD853J/R. sativus
FD27J/R. sativus
1J/R. sativus
NID048J/R. sativus
IWD038J/R. sativus
AD178J/R. sativus
FKH122J/B. pekinensis
Ka1J/B. pekinensis
Tu-2R1/R. sativus
VIET169/Vietnam/B. juncea
VIET170/Vietnam/B. juncea
VIET176/Vietnam/B. juncea
VIET80/Vietnam/R. sativus
VIET177/Vietnam/B. juncea
VIET167/Vietnam/B. juncea
VIET172/Vietnam/B. juncea
VIET158/Vietnam/B. juncea
VIET138/Vietnam/B. juncea
VIET178/Vietnam/B. juncea
VIET180/Vietnam/B. juncea
VIET179/Vietnam/B. juncea
CHZJ26A/B. campestris
VIET58/Vietnam/B. juncea
VIET56/Vietnam/B. juncea
VIET173/Vietnam/B. juncea
VIET175/Vietnam/B. juncea
VIET174/Vietnam/B. juncea
KWB778J/B. oleracea
KEN1
GRC17
UK1
NRD350J/R. sativus
C42J/B. rapa
Lu2/Brassica sp.
HZ6/Brassica sp
YC5/Zantedeschia sp.
Tu-3/B. oleracea
VIET153/Vietnam/B. juncea
RC4/Zantedeschia sp.
TW/Not known
VIET164/Vietnam/B. juncea
VIET160/Vietnam/B. juncea
world-B
0.1


B
ScMV
JYMVj1
JYMVm
IRNSS5
IRNTRa6
ITA3
IS1
St48
GRC42
A64
A102
AI
basal-B
Rn98
ITA7
Cal1
PV0104
TANX2/Raphanus sativus
KWB779J/Brassica rapa
NID119J/R. sativus
WFLB06/R. sativus
CP845/Canlendula officinalis
KYD81J/R. sativus
basal-BR
TUR9
KD32J/R. sativus
AKD934J/R. sativus
MED302J/R. sativus
KGD54J/R. sativus
ND10J/R. sativus
IWD032J/R. sativus
SGB088J/B. rapa
SGD311J/R. sativus
NDJ/R. sativus
59J/R. sativus
AD855J/R. sativus
H1J/R. sativus
AD181J/R. sativus
HOD517J/R. sativus
MYD015J/R. sativus
CH6/R. sativus
HRD/R. sativus
GFD462J/R. sativus
YMD070J/R. sativus
CHL13/R. sativus
KYD073J/R. sativus
TRD052J/R. sativus
CHK16/R. sativus
VIET73/Vietnam/R. sativus
VIET15/Vietnam/R. sativus
VIET66/Vietnam/R. sativus
VIET65/R. sativus
CHN1/Brassica sp.
SMD060J/R. sativus
VIET83/Vietnam/R. sativus
VIET89/Vietnam/R. sativus
VIET82/Vietnam/R. sativus
TRD053J/R. sativus
YMD069J/R. sativus
AD860J/R. sativus
YAD020J/R. sativus
AKH937J/B. pekinensis
AKD161J/R. sativus
Asian-BR
CAR51
CAR37A
DNK2
RUS1
RUS2
CZE1
PV376-Br
CZE5
NZ290
BZ1
CDN1
Q-Ca
NRD350/R. sativus
C42J/B. rapa
RC4/Zantedeschia sp.
TW/Not known
VIET166/Vietnam/B. juncea
VIET153/Vietnam/B. juncea
VIET164/Vietnam/B. juncea
VIET160/Vietnam/B. juncea
CAR37
GBR98
GBR36
GBR50
TUR1
NLD1
FRD1
USA1
C1/Not known
UK1
GRC17
KEN1
YC5/Zantedeschia sp.
KWB778J/B. oleracea
Tu-3/B. oleracea
Lu2/Brassica sp.
VIET138/Vietnam/B. juncea
1J/R. sativus
AD853J/R. sativus
VIET178/Vietnam/B. juncea
VIET180/Vietnam/B. juncea
VIET179/Vietnam/B. juncea
HZ6/Brassica sp.
BJ-C4/Brassicaceae sp.
CHN12/Not known
VIET79/Vietnam/R. sativus
VIET170/Vietnam/B. juncea
VIET176/Vietnam/B. juncea
VIET80/Vietnam/R. sativus
VIET177/Vietnam/B. juncea
VIET172/Vietnam/B. juncea
VIET158/Vietnam/B. juncea
CHZJ26A/B. campestris
VIET58/Vietnam/B. juncea
VIET56/Vietnam/B. juncea
VIET173/Vietnam/B. juncea
VIET175/Vietnam/B. juncea
VIET174/Vietnam/B. juncea
AT181J/Eustoma russellianum
YAL018J/Lactuca sativa
FKD004J/R. sativus
MYD013J/R. sativus
FKD001J/R. sativus
DMJ/R. sativus
NID048J/R. sativus
Tu-2RIJ/R. sativus
IWD038J/R. sativus
AD178J/R. sativus
FKH122J/B. pekinensis
Ka1J/B. pekinensis
2J/B. pekinensis
world-B
0.1


C
ScMV
JYMVj1
JYMVm
St48
AI
A102
GRC42
A64
ITA3
IS1
IRNSS5
IRNTRa6
basal-B
Rn98
TANX2/Raphanus sativus
KWB779J/Brassica rapa
CP845J/Calendula officinalis
GFD462J/R. sativus
KYD81J/R. sativus
ITA7
Cal1
NID119J/R. sativus
PV0104
TD88J/R. sativus
basal-BR
TUR9
CHL13/R. sativus
CHK16/R. sativus
CH6/R. sativus
KYD073J/R. sativus
TRD052J/R. sativus
CHN1/Brassica sp.
VIET89/Vietnam/R. sativus
VIET83/Vietnam/R. sativus
VIET82/Vietnam/R. sativus
AD855J/R. sativus
SGD311J/R. sativus
MYD015J/R. sativus
HRD/R. sativus
FD27J/R. sativus
SGB088J/B. rapa
AKD934J/R. sativus
YMD070J/R. sativus
59J/R. sativus
KD32J/R. sativus
MED302J/R. sativus
TRD053J/R. sativus
KGD54J/R. sativus
ND10J/R. sativus
SMD060J/R. sativus
VIET73/Vietnam/R. sativus
VIET15/Vietnam/R. sativus
VIET66/Vietnam/R. sativus
VIET65/Vietnam/R. sativus
Asian-BR
RUS1
DNK2
CAR51
CAR37A
BZ1
TUR1
NLD1
CZE5
CAR39
RUS2
CZE1
CDN1
Q-Ca
CAR37
TW/Not known
USA1
C1/Not known
VIET80/Vietnam/R. sativus
VIET177/Vietnam/B. juncea
VIET170/Vietnam/B. juncea
VIET172/Vietnam/B. juncea
VIET158/Vietnam/B. juncea
NRD350J/R. sativus
VIET138/Vietnam/B. juncea
VIET176/Vietnam/B. juncea
VIET79/Vietnam/R. sativus
VIET167/Vietnam/B. juncea
CHZJ26A/B. campestris
Lu2/Brassica sp.
BJ-C4/Brassicaceae sp.
CHN12/Not known
VIET169/Vietnam/B. juncea
VIET159/Vietnam/B. juncea
GRC17
KEN1
GBR50
UK1
GBR98
GBR36
AD853J/R. sativus
Tu-3/B. oleracea
YC5/Zantedeschia sp.
VIET178/Vietnam/B. juncea
VIET180/Vietnam/B. juncea
VIET179/Vietnam/B. juncea
IWD032J/R. sativus
VIET58/Vietnam/B. juncea
VIET56/Vietnam/B. juncea
VIET173/Vietnam/B. juncea
VIET175/Vietnam/B. juncea
VIET174/Vietnam/B. juncea
VIET164/Vietnam/B. juncea
VIET160/Vietnam/B. juncea
HZ6/Brassica sp.
1J/R. sativus
AT181J/Eustoma russellianum
C42J/B. rapa
NDJ/R. sativus
NID048J/R. sativus
AD181J/R. sativus
HOD517J/R. sativus
YMD069J/R. sativus
VIET166/Vietnam/B. juncea
VIET153/Vietnam/B. juncea
YAD020J/R. sativus
AD860J/R. sativus
AKD161J/R. sativus
H1J/R. sativus
FKD004J/R. sativus
MYD013J/R. sativus
FKD001J/R. sativus
DMJ/R. sativus
IWD038J/R. sativus
AD178J/R. sativus
FKH122J/B. pekinensis
Ka1JB. pekinensis
Tu-2R1/R. sativus
2J/B. pekinensis
world-B
0.1


D
NYSV
St48
AI
GRC42
A102
IS1
ITA3
IRNSS5
IRNTRa6
Rn98
ITA7
Cal1
CP845J/Calendula officinalis
KYD81J/Raphanus sativus
GFD462J/R. sativus
TANX2/R. sativus
KWB779J/Brassica rapa
NID119J/R. sativus
PV0104
TD88J/R. sativus
TUR9
FD27J/R. sativus
MED302J/R. sativus
TRD053J/R. sativus
KGD54J/R. sativus
ND10J/R. sativus
MYD015J/R. sativus
AD855J/R. sativus
SGD311J/R. sativus
SMD060J/R. sativus
HRD/R. sativus
CH6/R. sativus
CHL13/R. sativus
KYD073J/R. sativus
SGB088J/B. rapa
WFLB06/R. sativus
TRD052J/R. sativus
AKD934J/R. sativus
VIET73/Vietnam/R. sativus
VIET15/Vietnam/R. sativus
VIET66/Vietnam/R. sativus
VIET65/Vietnam/R. sativus
CZE1
RUS1
DNK2
CAR51
CAR37A
CAR39
NLD1
FRD1
TUR1
CDN1
CZE5
BZ1
Q-Ca
NZ290
PV376-Br
GBR50
USA1
C1/Not known
UK1
GBR98
GBR36
GRC17
KEN1
CAR37
BJ-C4/Brassicaceae sp.
CHN12/Not known
Lu2/B. rapa
VIET159/Vietnam/B. juncea
KWB778J/B. oleracea
VIET167/Vietnam/B. juncea
CHZJ26A/B. campestris
VIET169/Vietnam/B. juncea
IWD032J/R. sativus
VIET138/Vietnam/B. juncea
Tu-3/B. oleracea
FKD004J/R. sativus
FKD001J/R. sativus
DMJ/R. sativus
YC5/Zantedeschia sp.
VIET180/Vietnam/B. juncea
VIET179/Vietnam/B. juncea
VIET178/Vietnam/B. juncea
VIET172/Vietnam/B. juncea
VIET158/Vietnam/B. juncea
VIET170/Vietnam/B. juncea
VIET177/Vietnam/B. juncea
VIET80/Vietnam/R. sativus
HOD517J/R. sativus
C42J/B. rapa
VIET173/Vietnam/B. juncea
1J/R. sativus
VIET175/Vietnam/B. juncea
VIET174/Vietnam/B. juncea
NRD350J/R. sativus
YAD020J/R. sativus
AD860J/R. sativus
AKH937J/B. pekinensis
VIET164/Vietnam/B. juncea
VIET160/Vietnam/B. juncea
AKD161J/R. sativus
YMD069J/R. sativus
VIET166/Vietnam/B. juncea
VIET153/Vietnam/B. juncea
AT181J/Eustoma russellianum
HZ6/Brassica sp.
NDJ/R. sativus
VIET176/Vietnam/B. juncea
VIET79/Vietnam/R. sativus
NID048J/R. sativus
VIET58/Vietnam/B. juncea
VIET56/Vietnam/B. juncea
RC4/Zantedeschia sp.
TW/Not known
YAL018J/Lactuca sativa
AD853J/R. sativus
IWD038J/R. sativus
AD178J/R. sativus
MYD013J/R. sativus
2J/B. pekinensis
FKH122J/B. pekinensis
Tu-2R1/R. sativus
Ka1J/B. pekinensis
0.1
basal-B
basal-BR
Asian-BR
world-B

1.5. Nucleotide difference and diversity

The similarities of the sequences, both nucleotide and encoded polyprotein, of basal-B, basal-BR, Asian-BR and world-B genomes were calculated. The within group arithmetic means of all individual pairwise distances between isolates of the Asian-BR and world-B groups were 0.10552 and 0.10092 in nucleotide and 0.02591 and 0.03950 in amino acid sequences respectively. In contrast, the differences between the isolates in basal-B group were 0.15828 in nucleotide and 0.04946 in amino acid sequences. Thus there was little difference in the within-group variation of the Asian-BR and world-B groups, whereas the the basal-B group was 1.5 times more variable. Furthermore, the nucleotide and amino acid differences between subgroups were calculated (Table 7); the between-subgroup differences in world-B group were 0.17276–0.14612 in nucleotide and 0.06421–0.05323 in amino acid. Again the variation between basal-B and other groups/subgroups was much greater; 0.23030 and over in nucleotide sequences, and 0.08102 and over in amino acid sequences.

1.6. Genetic differentiation and gene flow

Three independent statistical tests of population differentiation were applied (Hudson, 2000) to estimate whether geographical isolates are genetically differentiated populations (Table 8). The analysis focused on the geographical groups of isolates, namely Vietnamese, Chinese and Japanese, because these are almost neighboring countries in Asia (Table 8A), and as the power of the tests increases as the subpopulation size also increases and these are the largest geographical groupings covered in the present study.

In the Asian-BR group, the null hypothesis was rejected for the majority of comparisons, with all the test statistics (K_S*, *Z*, and Snn) supported by *P*-values less than 0.05

Table 7. Comparison of nucleotide and amino acid differences in polyprotein encoding region within and between subpopulations of *Turnip mosaic virus*[a]

(A) Within group and subgroup

Group[b]/Subgroup	Nucleotide		Amino acid	
	differences	Standard error	differences	Standard error
Group				
basal-B (n=4)[c]	0.15828	0.00497	0.04946	0.00281
basal-BR (n=3)	0.08871	0.00283	0.01988	0.00194
Asian-BR (n=3)	0.10552	0.00391	0.02591	0.00212
world-B (23)	0.10092	0.00252	0.03950	0.00183
Subgroup				
East Europe (n=3)	0.01982	0.00114	0.01148	0.00147
worldwide (n=3)	0.10341	0.00402	0.03646	0.00276
Asia (n=18)	0.05420	0.00157	0.02519	0.00164

(B) Between group and subgroup

Group[b]		basal-B	basal-BR	Asian-BR	world-B			
	Subgroup					East Europe	worldwide	Asia
basal-B (n=4)[c]		-	0.23030	0.23507	0.23678	0.24428	0.23802	0.23524
basal-BR (n=3)		0.08161	-	0.16368	0.19528	0.20098	0.19723	0.19393
Asian-BR (n=3)		0.08102	0.04374	-	0.20022	0.20835	0.20475	0.19799
world-B (23)		0.09169	0.06574	0.06572	-	-	-	-
	East Europe (n=3)	0.09745	0.07320	0.07136	-	-	0.16976	0.17276
	worldwide (n=3)	0.09566	0.07171	0.07214	-	0.06421	-	0.14612
	Asia (n=18)	0.08997	0.06337	0.06359	-	0.06138	0.05323	-

[a] The nucleotide (above diagonal) and amino acid (below diagonal) similarities were assessed using Kimura two-parameter (Kimura, 1980) and Dayhoff PAM 001 matrix (Schwarz and Dayhoff, 1979).
[b] basal-B, basal-BR, Asian-BR and world-B groups isolates, and East Europe, worldwide and Asia subgroup isolates are listed in Figure 14.
[c] The number of sequences.

Table 8. Genetic differentiation and gene flow of *Turnip mosaic virus* populations.

(A) Between Chinese, Japanese and Vietnamese populations.

Protein encoding region[a]	Country (the number of sequences[b])	Parameter[c]				
		*Ks** (*P*-value)	*Z* (*P*-value)	*S*nn (*P*-value)	F_{ST}	*Nm*
Asian-BR						
HC-Pro	China (n=5) vs. Japan (n=25)	3.58275 (0.2020)	212.15699 (0.1440)	0.93333 (0.0010††)	0.04733	5.03
	China (n=5) vs. Vietnam (n=7)	3.05291 (0.0270†)	29.34524 (0.1020)	0.91667 (0.0100†)	0.20489	0.97
	Japan (n=25) vs. Vietnam (n=7)	3.32707 (0.0000†††)	208.76117 (0.0000†††)	1.00000 (0.0000†††)	0.24048	0.79
P3	China (n=5) vs. Japan (n=26)	3.29634 (0.0270†)	219.93829 (0.0190†)	0.69892 (0.6620)	0.03178	7.62
	China (n=5) vs. Vietnam (n=7)	2.94821 (0.0300†)	28.34405 (0.0700)	0.75000 (0.0710)	0.17460	1.18
	Japan (n=26) vs. Vietnam (n=7)	3.17308 (0.0000†††)	225.34025 (0.0000†††)	1.00000 (0.0000†††)	0.23172	0.83
NIb	China (n=5) vs. Japan (n=16)	3.45523 (0.0080††)	97.48725 (0.0190†)	0.76190 (0.1670)	0.03594	6.71
	China (n=5) vs. Vietnam (n=7)	3.07666 (0.0190†)	29.39107 (0.0870)	0.87500 (0.0030††)	0.17995	1.14
	Japan (n=16) vs. Vietnam (n=7)	3.23039 (0.0000†††)	106.17212 (0.0000†††)	1.00000 (0.0000†††)	0.22173	0.88
CP	China (n=4) vs. Japan (n=13)	2.71460 (0.0530)	62.22682 (0.0430†)	0.76471 (0.1260)	0.15366	1.38
	China (n=4) vs. Vietnam (n=4)	1.94935 (0.0330†)	5.87500 (0.0330†)	1.00000 (0.0330†)	0.44848	0.31
	Japan (n=13) vs. Vietnam (n=4)	2.53370 (0.0000†††)	52.58235 (0.0010††)	1.00000 (0.0010††)	0.48520	0.27
world-B						
HC-Pro	China (n=9) vs. Japan (n=15)	4.19975 (0.0020††)	119.27613 (0.0030††)	0.91667 (0.0000†††)	0.11087	2.00
	China (n=9) vs. Vietnam (n=20)	4.34650 (0.0080††)	190.56089 (0.0400†)	0.93103 (0.0000†††)	0.08933	2.55
	Japan (n=15) vs. Vietnam (n=20)	4.26170 (0.0000†††)	251.91014 (0.0000†††)	1.00000 (0.0000†††)	0.15706	1.34
P3	China (n=9) vs. Japan (n=19)	4.13867 (0.0000†††)	169.81695 (0.0050††)	0.82143 (0.0080††)	0.12030	1.83
	China (n=9) vs. Vietnam (n=19)	4.27571 (0.0300†)	186.46771 (0.2330)	0.82143 (0.0080††)	0.04147	5.78
	Japan (n=19) vs. Vietnam (n=19)	4.16531 (0.0000†††)	325.69737 (0.0020††)	0.94737 (0.0000†††)	0.13458	1.61
NIb	China (n=8) vs. Japan (n=26)	3.68684 (0.0000†††)	247.05609 (0.0000†††)	0.92647 (0.0000†††)	0.09791	2.30
	China (n=8) vs. Vietnam (n=23)	4.05893 (0.0160†)	242.42230 (1.0000)	0.87097 (0.0030††)	0.05434	4.35
	Japan (n=26) vs. Vietnam (n=23)	3.85237 (0.0000†††)	531.22282 (0.0000†††)	0.87075 (0.0000†††)	0.07804	2.95
CP	China (n=9) vs. Japan (n=27)	2.53378 (0.0000†††)	284.83272 (0.0000†††)	0.80556 (0.0080††)	0.06516	3.59
	China (n=9) vs. Vietnam (n=23)	2.95112 (0.0030††)	238.63409 (0.0210†)	0.82813 (0.0050††)	0.05219	4.54
	Japan (n=27) vs. Vietnam (n=23)	2.63673 (0.0000†††)	576.39216 (0.0000†††)	0.85933 (0.0000†††)	0.06028	3.90

(B) Between Asian and European populations.

Protein encoding region[a]	Country (the number of sequences)	Parameter[b]				
		*Ks** (*P*-value)	*Z* (*P*-value)	*S*nn (*P*-value)	F_{ST}	*Nm*
world-B						
HC-Pro	China (n=9) vs. Europe (n=14)	4.59695 (0.0220†)	117.45041 (0.0280†)	0.95652 (0.0000†††)	0.14962	1.42
	Japan (n=15) vs. Europe (n=14)	4.43169 (0.0000†††)	176.08261 (0.0000†††)	0.94253 (0.0000†††)	0.19546	1.03
	Vietnam (n=20) vs. Europe (n=14)	4.51533 (0.0000†††)	247.34455 (0.0000†††)	1.00000 (0.0000†††)	0.17807	1.15
P3	China (n=9) vs. Europe (n=13)	4.47785 (0.0150†)	103.93174 (0.0120†)	0.81818 (0.0080††)	0.15261	1.39
	Japan (n=19) vs. Europe (n=13)	4.27160 (0.0000†††)	192.91320 (0.0000†††)	1.00000 (0.0000†††)	0.24969	0.75
	Vietnam (n=19) vs. Europe (n=13)	4.40941 (0.0000†††)	215.88233 (0.0000†††)	1.00000 (0.0000†††)	0.18956	1.07
NIb	China (n=8) vs. Europe (n=12)	4.34342 (0.0090††)	87.26935 (0.0080††)	0.92500 (0.0000†††)	0.13330	1.63
	Japan (n=26) vs. Europe (n=12)	3.91941 (0.0000†††)	292.63737 (0.0000†††)	0.97368 (0.0000†††)	0.16540	1.26
	Vietnam (n=23) vs. Europe (n=12)	4.26600 (0.0000†††)	277.92666 (0.0010††)	0.91429 (0.0000†††)	0.12607	1.73
CP	China (n=9) vs. Europe (n=13)	3.27772 (0.0010††)	102.14939 (0.0030††)	0.95455 (0.0000†††)	0.16898	1.23
	Japan (n=27) vs. Europe (n=13)	2.71270 (0.0000†††)	298.26086 (0.0000†††)	1.00000 (0.0000†††)	0.30630	0.57
	Vietnam (n=23) vs. Europe (n=13)	3.10024 (0.0000†††)	234.88682 (0.0000†††)	1.00000 (0.0000†††)	0.25326	0.74

[a] HC-Pro; helper component-proteinase protein, P3; protein 3, NIb; nuclear inclusion b protein, CP; coat protein.

[b] For the numbers of sequences, refer Table 1, Table 4 and Figure 15.

[c] *Ks** and *Z* are the sequence-based statistics considered by Hudson *et al*. (2000). *S*nn is the nearest-neighbor statistic. F_{ST} is the interpopulation component of genetic variation of the standardized variance in allele frequencies across populations. *N* is the population size of each subpopulation. *m* is the migration fraction per generation recombination rate between the ends of the segment sequenced.

† 0.01<P<0.05, †† 0.001<P<0.01, ††† P<0.001, determined using 1000 permutations.

only in some cases the K_S*, *Z* and Snn statistics were not significant. The failure of the tests to differentiate the two geographical groups may be attributed to lack of statistical power. However, most of other test statistics were significant for the comparisons, supporting the conclusion that there has been partial genetic differentiation between the Vietnamese, Chinese and Japanese geographical groups in agreement with the results of the phylogenetic analyses (Figure 15). Namely, the absolute values of F_{ST} were smaller than 0.33, indicating that the gene flow of TuMV populations between Vietnam, China and Japan were frequent in most of protein encoding regions, the only exception being the CP encoding region, where comparisons between Vietnamese and Chinese/Japanese isolates gave valuses greater than 0.33. However, the gene flow between China and Japan seems to be more frequent than between those countries and Vietnam.

In the world-B group, all the values of F_{ST} were smaller than 0.33, indicating that the gene flow of TuMV populations between Vietnam, China and Japan were frequent. The smaller values of F_{ST} were seen in all four protein encoding regions. The values of F_{ST} were the smallest between China and Vietnam in all four protein encoding regions, indicating that the gene flow between China and Vietnam in the world-B group is the most frequent. As the world-B group contains many European countries, genetic differentiations between Asian and European populations were also analyzed. All the values of F_{ST} were smaller than 0.33 but greater than the values between Asian countries, indicating less gene flow of TuMV populations between Asian and European countries than those within Asian countries. The results are in agreement with the phylogenetic relationships of each population (Figures 14 and 15).

The statistical tests of population differentiation were also applied to estimate whether geographical isolates are genetically differentiated populations, considering the

original host plants, *Brassica* and *Raphanus*, of the isolates (Table 9). Most of other test statistics were significant for the comparisons, supporting the conclusion that there has been gene flow between the Vietnamese, Chinese and Japanese geographical groups, especially between the subpopulations from the same original hosts, in agreement with the results of the phylogenetic analyses (Figure 15).

1.7. Levels and patterns of intraspecific polymorphism

The patterns of molecular diversity were evaluated using Tajima's *D*, Fu and Li's *D** and *F** statistical test at segregating sites (Table 10), and haplotype diversity and nucleotide diversity at all sites (Table 11). The statistical tests are expected to have negative values for background selection, genetic hitchhiking and demographic expansion, and the negative values indicate that a population maintained low frequency polymorphism. Because selection events such as genetic hitchhiking and background selection affect relatively small fractions of the genome, a multilocus trend of negative statistical values would indicate that demographic forces are acting on the population (Tajima, 1989; Hey and Harris 1999; Tsompana *et al*., 2005). For the majority of geographical groups across the genome values of Tajima's *D*, Fu and Li's *D** and *F** statistical test are negative, indicating the occurrence of recent TuMV population expansions. Particularly, *F* statistical tests supported demographic expansion of the Japanese isolates in two of four protein encoding regions of Asian-BR and world-B groups.

Table 9. Genetic differentiation and gene flow of *Turnip mosaic virus* isolates, considering their original host plants.

Host and protein encoding region[a]	Country (the number of sequences[b])	Parameter[c]				
		*Ks** (*P*-value)	*Z* (*P*-value)	*Snn* (*P*-value)	F_{ST}	*Nm*
Between *Brassica* and *Raphanus* isolates						
HC-Pro	Asia[d]/Br[e] (n=31) vs. Asia/Rs[f] (n=47)	4.47249 (0.0000†††)	1185.34854 (0.0000†††)	0.80128 (0.0000†††)	0.38529	0.40
	China/Br (n=4) vs. China/Rs (n=6)	4.63606 (0.0730)	15.68333 (0.0380†)	0.90000 (0.0300†)	0.30631	0.57
	China/Br (n=4) vs. Japan/Rs (n=33)	4.42367 (0.0190†)	314.13697 (0.0230†)	0.94595 (0.0020††)	0.19243	1.05
	China/Rs (n=6) vs. Japan/Br (n=8)	4.78542 (0.0400†)	38.53524 (0.0330†)	0.71429 (0.0800)	0.21279	0.92
	China/Rs (n=6) vs. Vietnam/Br (n=19)	4.39375 (0.0000†††)	95.71353 (0.0000†††)	1.00000 (0.0000†††)	0.54036	0.21
	China/Br (n=4) vs. Vietnam/Rs (n=8)	3.80359 (0.0050††)	25.95833 (0.0110†)	0.83333 (0.0420†)	0.38333	0.40
	Japan/Br (n=8) vs. Japan/Rs (n=33)	4.48700 (0.0160†)	397.32760 (0.0480†)	0.71951 (0.3510)	0.13252	1.64
	Japan/Rs (n=33) vs. Vietnam/Br (n=19)	4.38403 (0.0000†††)	483.89686 (0.0000†††)	0.98077 (0.0000†††)	0.47535	0.28
	Japan/Br (n=8) vs. Vietnam/Rs (n=8)	4.20555 (0.0020††)	51.64286 (0.0120†)	0.81250 (0.0140†)	0.31966	0.53
	Vietnam/Br (n=19) vs. Vietnam/Rs (n=8)	4.12526 (0.0000†††)	105.13211 (0.0000†††)	0.88889 (0.0030††)	0.63519	0.14
P3	Asia/Br (n=31) vs. Asia/Rs (n=52)	4.43672 (0.0000†††)	1396.50515 (0.0000†††)	0.78313 (0.0000†††)	0.32095	0.53
	China/Br (n=4) vs. China/Rs (n=6)	4.58102 (0.0350†)	15.21667 (0.0120†)	0.90000 (0.0280†)	0.28682	0.62
	China/Br (n=4) vs. Japan/Rs (n=37)	4.38776 (0.0260†)	382.83914 (0.0110†)	0.87805 (0.1290)	0.14538	1.47
	China/Rs (n=6) vs. Japan/Br (n=9)	4.66516 (0.0490†)	42.80960 (0.0220†)	0.66667 (0.1870)	0.23383	0.82
	China/Rs (n=6) vs. Vietnam/Br (n=18)	4.31958 (0.0000†††)	85.55359 (0.0000†††)	1.00000 (0.0000†††)	0.52611	0.23
	China/Br (n=4) vs. Vietnam/Rs (n=9)	3.97924 (0.0150†)	35.27315 (0.0880)	0.80769 (0.0390)	0.21777	0.90
	Japan/Br (n=9) vs. Japan/Rs (n=37)	4.43281 (0.0140†)	492.55043 (0.0190†)	0.76449 (0.0860)	0.12629	1.73
	Japan/Rs (n=37) vs. Vietnam/Br (n=18)	4.33829 (0.0000†††)	565.96787 (0.0000†††)	0.96364 (0.0000†††)	0.42242	0.34
	Japan/Br (n=9) vs. Vietnam/Rs (n=9)	4.26027 (0.0040††)	72.19444 (0.1450)	0.77778 (0.0310†)	0.21474	0.91
	Vietnam/Br (n=18) vs. Vietnam/Rs (n=9)	4.11820 (0.0000†††)	125.99460 (0.0000†††)	0.85185 (0.0040††)	0.49565	0.25
NIb	Asia/Br (n=32) vs. Asia/Rs (n=53)	4.47808 (0.0000†††)	1541.64100 (0.0000†††)	0.80784 (0.0000†††)	0.25503	0.73
	China/Br (n=4) vs. China/Rs (n=5)	4.31887 (0.0250†)	11.75000 (0.0250†)	0.88889 (0.0250†)	0.40078	0.37
	China/Br (n=4) vs. Japan/Rs (n=39)	4.61099 (0.4270)	446.51773 (0.1730)	0.95349 (0.0040††)	-0.00948	-26.63
	China/Rs (n=5) vs. Japan/Br (n=7)	4.34525 (0.0020††)	23.33958 (0.0090††)	0.75000 (0.0720)	0.42479	0.34
	China/Rs (n=5) vs. Vietnam/Br (n=21)	4.10182 (0.0000†††)	108.36667 (0.0000†††)	1.00000 (0.0000†††)	0.66761	0.12
	China/Br (n=4) vs. Vietnam/Rs (n=9)	3.96064 (0.0040††)	34.43364 (0.0760)	0.76923 (0.0290†)	0.25982	0.71
	Japan/Br (n=7) vs. Japan/Rs (n=39)	4.59515 (0.1850)	512.83577 (0.1860)	0.82609 (0.0870)	0.02042	12.00
	Japan/Rs (n=39) vs. Vietnam/Br (n=21)	4.43705 (0.0000†††)	769.77580 (0.0000†††)	0.88611 (0.0000†††)	0.26828	0.68
	Japan/Br (n=7) vs. Vietnam/Rs (n=9)	4.06779 (0.0080††)	50.50099 (0.0370†)	0.75000 (0.0230†)	0.30854	0.56
	Vietnam/Br (n=21) vs. Vietnam/Rs (n=9)	4.01121 (0.0000†††)	145.38039 (0.0000†††)	0.90000 (0.0010††)	0.56017	0.20
CP	Asia/Br (n=33) vs. Asia/Rs (n=45)	3.32482 (0.0000†††)	1341.70343 (0.0000†††)	0.62265 (0.0330†)	0.25164	0.74
	China/Br (n=3) vs. China/Rs (n=5)	2.97347 (0.0080††)	5.96667 (0.0200†)	1.00000 (0.0200†)	0.72155	0.10
	China/Br (n=3) vs. Japan/Rs (n=34)	3.51051 (0.5130)	331.88430 (0.3120)	0.91892 (0.0860)	0.26416	0.70
	China/Rs (n=5) vs. Japan/Br (n=9)	3.03618 (0.0060††)	30.82056 (0.0040††)	0.50000 (0.5590)	0.51327	0.24
	China/Rs (n=5) vs. Vietnam/Br (n=21)	2.95490 (0.0000†††)	108.35238 (0.0000†††)	1.00000 (0.0000†††)	0.67887	0.12
	China/Br (n=3) vs. Vietnam/Rs (n=6)	2.91091 (0.0300†)	11.50000 (0.0040††)	0.77778 (0.0730)	0.52290	0.23
	Japan/Br (n=9) vs. Japan/Rs (n=34)	3.44397 (0.2140)	446.49652 (0.1840)	0.68837 (0.3850)	0.05270	4.49
	Japan/Rs (n=34) vs. Vietnam/Br (n=21)	3.31296 (0.0010††)	686.63561 (0.0010††)	0.83576 (0.0000†††)	0.22867	0.84
	Japan/Br (n=9) vs. Vietnam/Rs (n=6)	3.00204 (0.0090††)	39.20934 (0.0080††)	0.80000 (0.0080††)	0.28483	0.63
	Vietnam/Br (n=21) vs. Vietnam/Rs (n=6)	2.94211 (0.0000†††)	139.46087 (0.0000†††)	0.81481 (0.0280†)	0.46255	0.29
Between *Brassica* isolates						
HC-Pro	China (n=4) vs. Japan (n=8)	4.87002 (0.7770)	34.11310 (0.7930)	0.58333 (0.4410)	-0.10516	-2.63
	China (n=4) vs. Vietnam (n=19)	4.38814 (0.0650)	126.94537 (0.4930)	0.95652 (0.0060††)	0.01244	19.85
	Japan (n=8) vs. Vietnam (n=19)	4.49620 (0.0030††)	168.75779 (0.0880)	1.00000 (0.0000†††)	0.11110	2.00
P3	China (n=4) vs. Japan (n=9)	4.79136 (0.5970)	40.54784 (0.8330)	0.69231 (0.2330)	-0.07945	-3.40
	China (n=4) vs. Vietnam (n=18)	4.34428 (0.0900)	117.60367 (0.7210)	0.77273 (0.2000)	-0.02752	-9.33
	Japan (n=9) vs. Vietnam (n=18)	4.43599 (0.0000†††)	169.61505 (0.1030)	0.96296 (0.0000†††)	0.11433	1.94
NIb	China (n=4) vs. Japan (n=7)	4.44943 (0.5350)	26.45578 (0.3080)	0.75758 (0.0680)	-0.09919	-2.77
	China (n=4) vs. Vietnam (n=21)	4.12496 (0.1310)	157.99887 (0.9910)	0.76000 (0.4180)	0.01521	16.19
	Japan (n=7) vs. Vietnam (n=21)	4.15799 (0.0170)	198.27738 (1.0000)	0.85714 (0.0070††)	0.07237	3.20
CP	China (n=3) vs. Japan (n=9)	2.94923 (0.4960)	33.28472 (0.5620)	0.75000 (0.1460)	0.03075	7.88
	China (n=3) vs. Vietnam (n=21)	2.91200 (0.9090)	145.63798 (0.9850)	0.75000 (0.7710)	0.02606	9.34
	Japan (n=9) vs. Vietnam (n=21)	2.95030 (0.0220†)	223.35630 (0.9980)	0.79444 (0.0090)	0.03568	6.76
Between *Raphanus* isolates						
HC-Pro	China (n=6) vs. Japan (n=33)	4.42501 (0.1720)	362.95246 (0.1080)	0.92308 (0.0010††)	0.03714	6.48
	China (n=6) vs. Vietnam (n=8)	3.93228 (0.0340†)	43.02238 (0.1090)	0.92857 (0.0040††)	0.06404	3.65
	Japan (n=33) vs. Vietnam (n=8)	4.25642 (0.0230†)	388.46996 (0.0150†)	0.97154 (0.0000†††)	0.07735	2.98
P3	China (n=6) vs. Japan (n=37)	4.37287 (0.1830)	442.36585 (0.0930)	0.74806 (0.6380)	0.04352	5.49
	China (n=6) vs. Vietnam (n=9)	4.00069 (0.0720)	49.10429 (0.1000)	0.86667 (0.0160†)	0.05453	4.33
	Japan (n=37) vs. Vietnam (n=9)	4.25878 (0.0180†)	505.67111 (0.0580)	1.00000 (0.0000)	0.02162	11.31
NIb	China (n=5) vs. Japan (n=39)	4.58612 (0.0350†)	461.13349 (0.0710)	0.88636 (0.0660)	0.24356	0.78
	China (n=5) vs. Vietnam (n=9)	3.92618 (0.0850)	44.22583 (0.2590)	0.89286 (0.0000†††)	0.03536	6.82
	Japan (n=39) vs. Vietnam (n=9)	4.48855 (0.0040††)	549.51294 (0.0760)	0.95833 (0.0000†††)	0.13747	1.57
CP	China (n=5) vs. Japan (n=34)	3.50328 (0.0070††)	341.87504 (0.0010††)	0.87179 (0.0320†)	0.29907	0.59
	China (n=5) vs. Vietnam (n=6)	3.04607 (0.0480†)	23.05714 (0.0280†)	1.00000 (0.0030††)	0.10772	2.07
	Japan (n=34) vs. Vietnam (n=6)	3.47987 (0.0110†)	373.07158 (0.0290†)	0.95000 (0.0000†††)	0.10352	2.16

[a] HC-Pro; helper component protein, P3; protein 3, NIb; nuclear inclusion b protein, CP; coat protein.
[b] For the numbers of sequences, refer Table 1, Table 4 and Figure 15.
[c] *Ks** and *Z* are the sequence-based statistics considered by Hudson *et al*. (2000). *Snn* is the nearest-neighbor statistic. F_{ST} is the inter-population component of genetic variation of the standardized variance in allele frequencies across populations. *N* is the population size of each subpopulation. *m* is the migration fraction per generation recombination rate rate between the ends of the segment sequenced.
[d] Isolates from China, Japan and Vietnam.
[e] Isolates originated from *Brassica* sp.
[f] Isolates originated from *Raphanus* sp.
[†] 0.01<P<0.05, [††] 0.001<P<0.01, [†††] P<0.001, determined using 1000 permutations.

Table 10. Neutrality tests of each *Turnip mosaic virus* populations.

(A) Chinese, Japanese and Vietnamese populations

Protein encoding region[a] and country	n[b]	Tajima's *D*	Fu & Li's *F**	Fu & Li's *D**
Asian-BR				
HC-Pro				
China	5	-0.63139	-0.68532	-0.63139
Japan	25	-1.92660†	-2.82820†	-2.62744†
Vietnam	7	1.54363	1.21783	0.97184
P3				
China	5	-0.62215	-0.67523	-0.62215
Japan	26	-1.83537†	-2.55928	-2.33249
Vietnam	7	1.84380	1.59546	1.32611
NIb				
China	5	-0.72045	-0.75033	-0.72045
Japan	16	-0.96453	-1.32024	-1.22268
Vietnam	7	-0.81734	-0.79606	-0.81734
CP				
China	4	-0.87890	-1.01629	-0.94922
Japan	13	-1.76968	-2.26937	-2.03877
Vietnam	4	1.86133	1.58501	1.30789
world-B				
HC-Pro				
China	9	-0.48432	-0.58533	-0.53221
Japan	15	-0.83024	-1.18974	-1.11001
Vietnam	20	-0.16900	-0.23720	-0.21796
P3				
China	9	-0.19375	-0.21170	-0.18653
Japan	19	-0.82492	-1.25131	-1.18001
Vietnam	19	-0.04508	0.14576	0.19785
NIb				
China	8	-1.00796	-1.20539	-1.10071
Japan	26	-2.03745†	-3.02177†	-2.81582†
Vietnam	23	-1.34435	-1.46206	-1.19949
CP				
China	9	-1.18468	-1.25299	-1.09289
Japan	27	-2.31499††	-3.57982††	-3.38344††
Vietnam	23	-1.56482	-1.53754	-1.19468
Polyprotein				
Vietnam	15	-0.39358	-0.55695	-0.51760

(B) European populations

Protein encoding region and country	n	Tajima's *D*	Fu & Li's *F**	Fu & Li's *D**
world-B				
HC-Pro				
Poland	4	-0.83333	-0.89745	-0.83333
UK	4	0.24153	0.17328	0.14492
P3				
Poland	3	ND[c]	ND	ND
UK	4	-0.07240	-0.16964	-0.17860
NIb				
Poland	4	0.59727	0.74904	0.71694
UK	4	-0.29977	-0.31893	-0.29977
CP				
Poland	4	0.37113	0.55263	0.54862
UK	4	-0.18114	-0.18471	-0.18114

[a] HC-Pro; helper component-proteinase protein, P3; protein 3, NIb; nuclear inclusion b protein, CP; coat protein.

[b] The number of sequences.

[c] Not determined. The number of sequences was less than four isolates because one isolate in Poland, had recombination site in P3 encoding region (Ohshima *et al*., 2007).

†P<0.05, ††P<0.01; Tajima's *D* test compares the nucleotide diversity with the proportion of polymorphic sites, which are expected to be equal under selective neutrality. Fu & Li's *D** test is based on the differences between the numbers of singletons (mutations appearing only once among the sequences) and the total number of mutations. Fu & Li's *F** test is based on the differences between the number of singletons and the average number of nucleotide differences between pairs of sequences.

Table 11. Haplotype and nucleotide diversities of subpopulations in *Turnip mosaic virus* groups.

(A) Chinese, Japanese and Vietnamese populations

Group and country	Protein encoding region[a]														
	Polyprotein			HC-Pro			P3			NIb			CP		
	n[b]	H[c]	π[d]	n	H	π	n	H	π	n	H	π	n	H	Π
All groups															
Asia	20	1.000 ± 0.016	0.0819 ± 0.0017	85	1.000 ± 0.002	0.1171 ± 0.0034	91	1.000 ± 0.002	0.1355 ± 0.0042	92	0.999 ± 0.002	0.0870 ± 0.0031	87	1.000 ± 0.002	0.0511 ± 0.0036
China	4	1.000 ± 0.177	0.1186 ± 0.0025	16	1.000 ± 0.022	0.1280 ± 0.0047	16	1.000 ± 0.022	0.1493 ± 0.0057	14	1.000 ± 0.027	0.0937 ± 0.0040	14	1.000 ± 0.001	0.0754 ± 0.0042
Japan	1	ND[e]	ND	42	1.000 ± 0.005	0.1056 ± 0.0037	49	1.000 ± 0.004	0.1278 ± 0.0046	48	1.000 ± 0.004	0.0887 ± 0.0033	46	1.000 ± 0.005	0.0535 ± 0.0037
Vietnam	15	1.000 ± 0.024	0.0511 ± 0.0014	27	0.997 ± 0.011	0.1054 ± 0.0038	26	0.994 ± 0.013	0.1210 ± 0.0046	30	0.993 ± 0.012	0.0768 ± 0.0034	27	1.000 ± 0.010	0.0393 ± 0.0036
basal-BR															
Asia	1	ND	ND	4	1.000 ± 0.177	0.0695 ± 0.0050	6	1.000 ± 0.096	0.0533 ± 0.0044	7	1.000 ± 0.076	0.0555 ± 0.0039	7	1.000 ± 0.076	0.0365 ± 0.0042
China				2	1.000 ± 0.500	0.0992 ± 0.0085	2	1.000 ± 0.500	0.0840 ± 0.0089	1	ND	ND	1	ND	ND
Japan	1	ND	ND	2	1.000 ± 0.500	0.0992 ± 0.0085	4	1.000 ± 0.177	0.0515 ± 0.0053	6	1.000 ± 0.096	0.0493 ± 0.0038	6	1.000 ± 0.096	0.0368 ± 0.0043
Asian-BR															
Asia	2	1.000 ± 0.500	0.0164 ± 0.0013	37	0.998 ± 0.007	0.0241 ± 0.0029	38	0.999 ± 0.006	0.0356 ± 0.0041	28	0.992 ± 0.013	0.0218 ± 0.0026	21	1.000 ± 0.015	0.0195 ± 0.0030
China	2	1.000 ± 0.500	0.0164 ± 0.0013	5	1.000 ± 0.126	0.0225 ± 0.0029	5	1.000 ± 0.126	0.0287 ± 0.0038	5	1.000 ± 0.126	0.0246 ± 0.0029	4	1.000 ± 0.177	0.0145 ± 0.0032
Japan				25	1.000 ± 0.011	0.0226 ± 0.0029	26	1.000 ± 0.011	0.0247 ± 0.0034	16	1.000 ± 0.022	0.0195 ± 0.0025	13	1.000 ± 0.030	0.0191 ± 0.0030
Vietnam				7	0.952 ± 0.096	0.0189 ± 0.0021	7	0.952 ± 0.096	0.0209 ± 0.0025	7	0.857 ± 0.137	0.0167 ± 0.0018	4	1.000 ± 0.177	0.0043 ± 0.0018
world-B															
Asia	17	1.000 ± 0.020	0.0508 ± 0.0014	44	1.000 ± 0.005	0.0611 ± 0.0035	49	0.999 ± 0.004	0.0743 ± 0.0044	57	1.000 ± 0.003	0.0433 ± 0.0034	59	1.000 ± 0.003	0.0199 ± 0.0035
China	2	1.000 ± 0.500	0.0389 ± 0.0021	9	1.000 ± 0.052	0.0584 ± 0.0042	9	1.000 ± 0.052	0.0756 ± 0.0052	8	1.000 ± 0.063	0.0361 ± 0.0033	9	1.000 ± 0.052	0.0270 ± 0.0039
Japan				15	1.000 ± 0.024	0.0472 ± 0.0036	19	1.000 ± 0.017	0.0600 ± 0.0050	26	1.000 ± 0.011	0.0242 ± 0.0029	27	1.000 ± 0.010	0.0138 ± 0.0032
Vietnam	15	1.000 ± 0.024	0.0511 ± 0.0014	20	1.000 ± 0.016	0.0612 ± 0.0036	19	0.994 ± 0.019	0.0746 ± 0.0043	23	1.000 ± 0.013	0.0340 ± 0.0032	23	1.000 ± 0.013	0.0227 ± 0.0034

(B) European populations

Group and country	Protein encoding region[a]														
	Polyprotein			HC-Pro			P3			NIb			CP		
	n	H	π	n	H	π	n	H	π	n	H	π	n	H	π
All groups															
Europe	7	1.000 ± 0.076	0.1443 ± 0.0024	23	1.000 ± 0.013	0.1557 ± 0.0046	26	1.000 ± 0.011	0.1856 ± 0.0056	25	1.000 ± 0.011	0.1310 ± 0.0041	25	1.000 ± 0.011	0.0777 ± 0.0048
Czech Republic	1	ND	ND	2	1.000 ± 0.500	0.1678 ± 0.0111	2	1.000 ± 0.500	0.1992 ± 0.0137	2	1.000 ± 0.500	0.0542 ± 0.0059	2	1.000 ± 0.500	0.0756 ± 0.0097
Germany				2	1.000 ± 0.500	0.1809 ± 0.0115	2	1.000 ± 0.500	0.2210 ± 0.0144	1	ND	ND	2	1.000 ± 0.500	0.0843 ± 0.0102
Greece				2	1.000 ± 0.500	0.2181 ± 0.0126	2	1.000 ± 0.500	0.2238 ± 0.0145	2	1.000 ± 0.500	0.1960 ± 0.0112	2	1.000 ± 0.500	0.1016 ± 0.0112
Italy	3	1.000 ± 0.272	0.1555 ± 0.0034	5	1.000 ± 0.126	0.1852 ± 0.0075	8	1.000 ± 0.063	0.1954 ± 0.0077	8	1.000 ± 0.063	0.1501 ± 0.0056	7	1.000 ± 0.076	0.0932 ± 0.0066
Poland	2	1.000 ± 0.500	0.0045 ± 0.0007	4	1.000 ± 0.177	0.0857 ± 0.0061	3	1.000 ± 0.272	0.1149 ± 0.0085	4	1.000 ± 0.177	0.0996 ± 0.0057	4	1.000 ± 0.177	0.0630 ± 0.0062
Russia	1	ND	ND	2	1.000 ± 0.500	0.0292 ± 0.0046	2	1.000 ± 0.500	0.0179 ± 0.0179	2	1.000 ± 0.500	0.1251 ± 0.0090	1	ND	ND
UK				4	1.000 ± 0.177	0.0269 ± 0.0032	4	1.000 ± 0.177	0.0307 ± 0.0040	4	1.000 ± 0.177	0.0191 ± 0.0026	4	1.000 ± 0.177	0.0093 ± 0.0025
basal-B															
Europe	1	ND	ND	4	1.000 ± 0.177	0.1329 ± 0.0072	5	1.000 ± 0.126	0.1352 ± 0.0076	5	1.000 ± 0.126	0.1157 ± 0.0060	4	1.000 ± 0.177	0.0768 ± 0.0072
Italy	1	ND	ND	3	1.000 ± 0.272	0.1593 ± 0.0087	5	1.000 ± 0.126	0.1513 ± 0.0080	5	1.000 ± 0.126	0.1408 ± 0.0064	4	1.000 ± 0.177	0.0919 ± 0.0077
basal-BR															
Europe	2	1.000 ± 0.500	0.0887 ± 0.0031	3	1.000 ± 0.272	0.1155 ± 0.0074	3	1.000 ± 0.272	0.1045 ± 0.0080	3	1.000 ± 0.272	0.0389 ± 0.0041	3	1.000 ± 0.272	0.0314 ± 0.0051
Italy	2	1.000 ± 0.500	0.0887 ± 0.0031	2	1.000 ± 0.500	0.1262 ± 0.0096	3	1.000 ± 0.271	0.1580 ± 0.0098	3	1.000 ± 0.272	0.0933 ± 0.0063	3	1.000 ± 0.272	0.0558 ± 0.0068
world-B															
Europe	3	1.000 ± 0.001	0.1393 ± 0.0013	16	1.000 ± 0.022	0.1117 ± 0.0043	16	1.000 ± 0.022	0.1279 ± 0.0052	15	1.000 ± 0.024	0.0823 ± 0.0038	16	1.000 ± 0.022	0.0496 ± 0.0042
Czech Republic	1	ND	ND	2	1.000 ± 0.500	0.1678 ± 0.0111	2	1.000 ± 0.500	0.1992 ± 0.0137	2	1.000 ± 0.500	0.0545 ± 0.0059	2	1.000 ± 0.500	0.0756 ± 0.0097
Poland	2	1.000 ± 0.500	0.0045 ± 0.0007	4	1.000 ± 0.177	0.0857 ± 0.0061	3	1.000 ± 0.272	0.1149 ± 0.0085	4	1.000 ± 0.177	0.0996 ± 0.0057	4	1.000 ± 0.177	0.0630 ± 0.0062
Russia	1	ND	ND	2	1.000 ± 0.500	0.0292 ± 0.0046	2	1.000 ± 0.500	0.0179 ± 0.0179	2	1.000 ± 0.500	0.1251 ± 0.0090	1	ND	ND
UK				4	1.000 ± 0.177	0.0269 ± 0.0032	4	1.000 ± 0.177	0.0307 ± 0.0040	4	1.000 ± 0.177	0.0191 ± 0.0026	4	1.000 ± 0.177	0.0093 ± 0.0025

[a] HC-Pro; helper component-proteinase protein, P3; protein 3, NIb; nuclear inclusion b protein, CP; coat protein.
[b] The number of sequences.
[c] Haplotype diversity.
[d] Nucleotide diversity was estimated by the average pairwise difference between sequences in a sample, based on all sites.
[e] Not determined.

Haplotype and nucleotide diversities of Asian and European subpopulations in TuMV phylogenetic groups were compared (Table 11). The haplotype diversity in most groups analyzed had a value of 1.000. In most cases, haplotype diversity values are high and nucleotide diversity values are low. Nucleotide diversities of Europe in the all genogroups showed greater diversity values than those of Asia in all the protein encoding regions of HC-Pro, P3, NIb and CP, nonetheless the number of European isolates used for the calculations is small. The results suggested that these genetic differences indicate that the European populations of TuMV are the oldest. Furthermore, the analysis of within-subpopulation diversities between the Asian countries in Asian-BR group, HC-Pro encoding region (0.0225 in China, 0.0226 in Japan and 0.0189 in Vietnam), P3 encoding region (0.0287 in China, 0.0247 in Japan and 0.0209 in Vietnam), NIb encoding region (0.0246 in China, 0.0195 in Japan and 0.0167 in Vietnam) and CP encoding region (0.0145 in China, 0.0191 in Japan and 0.0043 in Vietnam), showed that Vietnamese subpopulation is smaller. In contrast, within subpopulation diversities in all the protein encoding regions of world-B group in Vietnamese subpopulation were greater than those of Chinese and Japanese subpopulations, suggesting that there was different evolution in world-B and Asian-BR groups in Asia. Overall, the deviations from the neutral equilibrium model for the four analyzed genes, together with the combination of high haplotype diversity and overall lack of nucleotide diversity within individual geographical groups are consistent with a model of recent population expansion events.

1.8. AMOVA for haplotype distribution

In order to estimate the contribution of various factors to the genetic differentiation of the TuMV populations in China, Japan and Vietnam using four different genes; HC-Pro, P3,

NIb and CP, the analysis of molecular variance (AMOVA) was performed. The results of AMOVA analyses for haplotype distribution were presented in Table 12.

The estimation of the TuMV population beetwen China and Japan to the total variation was not significant in two genes; NIb (-1.21510% genetic variation, F_{st} = -0.01215) and CP (-1.32812% genetic variation, F_{st} = -0.01328). The variation contributed by the TuMV populations was 2.37760% - 11.09128% (F_{st} = 0.02378 - 0.11091), whilst the largest contribution to the variance (85.06582% - 99.05276%) was between TuMV isolates within individual fields (F_{st} = 0.00947% - 0.14934) (Table 12).

The AMOVA analysis showed that the most important component in the observed variability in TuMV populations is due to the variation within populations in the individual fields. This accounted for 85.06582% - 99.05276% of the total variance observed in the spatial, respectively. This could be explained that why TuMV populations in different regions were highly differentiated, with limited gene flow between them, resulting in fewer haplotypes shared between the regions. In case of plant RNA viruses, predominantly negative selection, population bottlenecks during movement, transmission and population differentiation during plant growth and development have been shown to be responsible for maintaining population diversity (Li and Roossinck., 2004).

Table 12. AMOVA for haplotype distribution of protein encoding regions of *Turnip mosaic virus* populations.

Protein encoding region[a]	Group of population	d.f.[b]	Sum of squares	Variance components	Percentage of variation	F_{st}[c]
HC-Pro	China vs. Japan					
	Between populations	1	0.500	0.00000 Va	5.44274	0.05443**
	Within population	56	28.00	0.50000 Vb	94.55726	
	China vs. Vietnam					
	Between populations	1	0.514	0.00073 Va	2.39087	0.02391**
	Within population	41	20.463	0.49910 Vb	97.60913	
	Japan vs. Vietnam					
	Between populations	1	0.523	0.00070 Va	14.93418	0.14934**
	Within population	67	33.463	0.49945 Vb	85.06582	
P3	China vs. Japan					
	Between populations	1	0.500	0.00000 Va	3.76390	0.03764**
	Within population	63	31.500	0.50000 Vb	96.23610	
	China vs. Vietnam					
	Between populations	1	0.529	0.00158 Va	0.94724	0.00947*
	Within population	40	19.923	0.49808 Vb	99.05276	
	Japan vs. Vietnam					
	Between populations	1	0.550	0.00151 Va	11.09128	0.11091**
	Within population	73	36.423	0.49895 Vb	88.90872	
NIb	China vs. Japan					
	Between populations	1	0.500	0.00000 Va	-1.21510	-0.01215 ns
	Within population	60	30.000	0.50000 Vb	101.21510	
	China vs. Vietnam					
	Between populations	1	0.532	0.00179 Va	2.37760	0.02378**
	Within population	42	20.900	0.49762 Vb	97.62240	
	Japan vs. Vietnam					
	Between populations	1	0.561	0.00170 Va	4.28545	0.04285**
	Within population	75	37.400	0.49867 Vb	95.71455	
CP	China vs. Japan					
	Between populations	1	0.500	0.00000 Va	-1.32812	-0.01328 ns
	Within population	58	29.000	0.50000 Vb	101.32812	
	China vs. Vietnam					
	Between populations	1	0.500	0.00000 Va	5.21075	0.05211**
	Within population	39	19.500	0.50000 Vb	94.78925	
	Japan vs. Vietnam					
	Between populations	1	0.500	0.00000 Va	5.91760	0.05918**
	Within population	71	35.500	0.50000 Vb	94.08240	

[a] HC-Pro; helper component-proteinase protein, P3; protein 3, NIb; nuclear inclusion b, CP; coat protein

[b] Degree of freedom

ns; not significant ($P > 0.05$), * $P < 0.01$, ** $P < 0.001$

[c] Fixation indices (*F*-statistics)

1.9. Discussion

There are fewer studies of the genetic structure and variability of plant viruses than of human and animal viruses. This study utilized molecular population genetics techniques to characterize the evolution of a plant RNA virus. Studies of the population structures of several potyviruses have shown some details of genetic variations at regional or global scales (Ohshima *et al*., 2002; Tomimura *et al*., 2003, 2004; Moreno *et al*., 2004; Tomitaka and Ohshima, 2006; Ogawa *et al*., 2008, 2012; Lecoq *et al*., 2009; Seo *et al*., 2009; Joannon *et al*., 2010; Zhang *et al*., 2011). In this study, the full genomic sequences of 30 TuMV isolates from Vietnam located in Southeast Asia were determined and the genetic variability of populations was compared to those of nearby Asian countries. The results in this study showed that (1) TuMV is widely distributed in Vietnam, (2) seven novel recombination patterns were present in the Vietnamese population of TuMV, (3) non-recombinant Vietnamese isolates were from the world-B group and clustered with Chinese isolates, (4) the world-B group isolates dominate the Vietnamese population and (5) there was evidence of frequent gene flow between TuMV isolates from Vietnam and other Asian countries.

In studies of plant virus evolution it is very important to clone the viral isolates being studied before they are sequenced. TuMV isolates were generally cloned by single lesion isolations in the earlier and present studies because of the high frequency of mixed infections in the field, not only with other viruses, especially CMV, but also other isolates of the same virus (approximately 2% in Kyushu, Japan) (Tomitaka and Ohshima, 2006). It is considered that mutations occur in each infected plant so that each contains a ‘‘mutant cloud’’ or “quasispecies” (Roossinck, 2008) of slightly different genomes within a single population. However, each field isolate may also contain mixtures of populations, even

when the virus has been transmitted by aphids. Thus, biological cloning is mandatory when attempting to analyze recombinational events and the genetic structure of populations (Ohshima *et al*., 2002; Moreno *et al*., 2004). This study found very few mismatches in the sequences between the overlapping RT-PCR products after the virus populations had been cloned by single lesion isolations.

Recombination is an important source of genetic variation not only for TuMV (Ohshima *et al*., 2002, 2007) but also for the other potyvirus species (Gibbs and Ohshima, 2010). This study has found recombination patterns in Vietnamese TuMV that have not been found before in isolates from European and Asian countries (Figure 13). Non-recombinants from the world-B group were mostly found in the Vietnamese TuMV population (50%) and also the Chinese TuMV population. The most frequent type of recombinant in the Vietnamese population is intralineage recombinants of world-B parents (eight intralineage recombinants out of 30 isolates, 26.6%). Most of the recombination sites were unique to Vietnam, and not found in Chinese and Japanese populations in the earlier studies (Tomimura *et al*., 2003; Ohshima *et al*., 2007; Wang *et al*., 2009). One recombination site nt 5989 seems to be identical to the site previously found in Czech Republic population (Ohshima *et al*., 2007). Therefore, a total of 73% of Vietnamese population has world-B parents and this group is dominant in Vietnam around 2006–2008. On the other hand, non-recombinants of Asian-BR and basal-BR lineages were not found in the Vietnamese population, although they have already been found in Chinese and Japanese population. Two interlineage recombinant type patterns of Asian-BR and world-B parents were found in Vietnam, and one of two patterns was identical to the Japanese population, indicating that such isolates may be now geographically dispersed in Asian countries and there may have been founder effects during the establishment of the subpopulations in Asia.

The genetic differentiation analyses together with the phylogenetic analyses reveal that TuMV has a structure in Asian countries that is partially congruent with geography (Table 8 and Figure 15), even though gene flow of the TuMV populations between Asian countries was frequent, especially between China and Japan of Asian-BR group isolates. This pattern might reflect characteristics of TuMV transmission and the physical and quarantine barriers between Asian countries. TuMV is transmitted by aphids in a non-persistent manner and they are able to only carry the virus for a short time. The sea, country-dependent agriculture crops and growing conditions of crops may form an obstacle for the spread of aphids, and thus limit spread of the virus. Physical obstacles have also been reported to be a reason responsible for the strain localization of *Rice yellow mottle virus* (Traore *et al*., 2005) and TVBMV (Zhang *et al*., 2011).

Phylogenetic trees represent the evolutionary relationships among populations. Our phylogenetic trees from partial genomic and polyprotein regions showed that there are at least four major groups and three subgroups in the world population of TuMV (Figure 14). All the Vietnamese non-recombinant world-B group sequences clustered with BJ-C4 and Lu2 of Chinese isolates in the polyprotein phylogenetic tree (Figure 14). Of course, this study cannot distinguish whether the Vietnamese isolates originated in Vietnam or were introduced from other countries on one or more occasions and then spread. It is clear that Vietnamese and Chinese populations had strong evolutionary relationships in the past. However, it is possible that TuMV has been introduced to Vietnam via China, although this study cannot exclude the possibility that TuMV could have been introduced from Vietnam into China. The results are supported by the analysis of gene flow between Vietnam, China and Japan subpopulations (Table 11).

Some subpopulations of basal-BR and Asian-BR groups of TuMV are probably recently emerged (Tomimura *et al*., 2003; Tomitaka and Ohshima, 2006). However, the present analysis showed no evidence of a Vietnam-specific emergent TuMV population and most (73%) of non-recombinants and intralineage recombinants belonged to world-B group, and the subpopulation might have existed for a relatively long time. Therefore, the increasing incidence of TuMV in Vietnamese Brassicaceae crops result from changed agronomic factors, such as increased and widespread planting of more susceptible cultivars. Interestingly, TuMV was not detected in some Brassicaceae production areas in Vietnam (data not shown). Although the reason for this is not clear; local differences in cultivars and geographical environments including climate in Southeast Asia may account for this situation.

In conclusion, the results of evolutionary studies have shown that TuMV populations in Vietnam, China and Japan were partial discrete lineages confined to individual districts, although these may have been modified subsequently by clinal genetic drift. As shown in this work, the introduction of new isolates and the appearance of new genetic types not only by recombination but also by mutation, represent a high potential risk, and should be considered in the design of efficient control strategies. This study shows for the first time the genetic structure of the TuMV population in a Southeast Asia country, and illustrates that, when designing control strategies, it is important to understand the genetic structure of the virus population at different geographic scales; country, region and locality.

2. Temporal evolution of *Turnip mosaic virus*

2.1. Biological and molecular characteristics

The 74 European, 65 East Asian, and 16 other TuMV isolates examined in this study are listed in Tables 2 and 4. Most *Brassica* plants, but not *Raphanus sativus*, were systemically infected by most isolates (Ohshima *et al*., 2002; Tomimura *et al*., 2003; this study). Thus, they were of the B-infecting host-type (pathotype), although they had minor differences in pathogenicity. Only six European isolates were among this group: Cal1, DEU 4, ITA 2, ITA 7, ITA 8 and PV0104. These had been collected in Italy and Germany and had been isolated from *Abutilon* sp., *Calendula officinalis*, *Cheiranthus cheeri*, *R. raphanistrum*, *R. sativus*, and *Lactuca sativa*. In contrast, approximately 90% of isolates from Europe were B host-type (Ohshima *et al*., 2007; Tomimura *et al*., 2003; this study).

Nineteen of 74 isolates collected in Europe were from non-brassicas. Four isolates, which were called Orchis isolates, came from *Orchis* spp.: OM-N was isolated from *Orchis militaris* and OM-A was isolated from OM-N by single-lesion isolations, whereas ORM and OS were isolated from *Orchis morio* and *Orchis simia* plants, respectively. The fact that Orchis isolates have only been found in orchids in a single glasshouse collection supports the conclusion (Lesemann and Vetten, 1985) that Orchis isolates did not come from brassica plants growing near the glasshouse, but came from one or more of the orchids in the collection, all of which were from Central or Mediterranean Europe.

Isolates OM-N and OS did not infect brassica test plants despite repeated testing (Table 13). By contrast, isolate ORM occasionally infected *Brassica rapa* and *B. juncea*, but was only detected by reverse transcription-polymerase chain reaction (RT-PCR) and double antibody sandwich enzyme-linked immunosorbent assay (DAS-ELISA) in the upper two uninoculated-

Table 13. Host reactions of some isolates of *Turnip mosaic virus* from *Orchis* spp. and other hosts.

Host plant				Symptom[a]			
Family	**Species**	**Common name**	**Seed origin**	**OM-N[b]**	**OS**	**ORM**	**Standard brassica isolate[c]**
Amaranthaceae	*Gomphrena globosa* cv. Bicolor rose	Globe amaranth	France	-/-	(CS/-)	ND	(CS/-)
	G. globosa cv. Strawberry field	Globe amaranth	UK	-/-	-/-	ND	(CS/-)
Amaryllidaceae	*Narcissus tazetta* var. *chinensis* cv. Scarlet Gem	Narcissus	Japan	-/-	-/-	-/-	-/-
	N. tazetta var. *chinensis* cv. The Winston Churchill	Narcissus	Japan	-/-	-/-	-/-	-/-
Asteraceae	*Lactuca sativa* cv. Emrap 231	Lettuce	USA	-/-	-/-	-/-	-/-
	L. sativa cv. Salinas 88	Lettuce	Australia	-/-	-/-	-/-	(LI/M, St, D)
Brassicaceae	*Brassica chinensis* cv. Choyo	Qing geng cai	Denmark	-/-	-/-	-/-	NLL/BSS, PSF
	B. juncea cv. Hakarashina	Mustard	Italy	-/-	-/-	(LI/M, CS)[u]	NLL/M, LD
	B. napus cv. Norin-32 go	Oilseed rape	Japan	-/-	-/-	-/-	NLL/-
	B. narinosa cv. Tatsuai	Rosette pakchoi	Australia	-/-	-/-	ND	LI/M
	B. oleracea var. botrytis cv. Snow crown	Cauliflower	Australia	-/-	-/-	-/-	(LI/M)
	B. oleracea var. botrytis cv. Snow queen	Cauliflower	Chile	-/-	-/-	-/-	(LI/M)
	B. oleracea var. capitata cv. Ryozan-2 go	Green cabbage	Japan	-/-	-/-	-/-	(LI/M)
	B. oleracea var. capitata cv. Sinsei	Green cabbage	Japan	-/-	-/-	-/-	(LI/M)
	B. oleracea var. capitata cv. Yalova-1	White cabbage	Turkey	-/-	ND	ND	(LI/M)
	B. oleracea var. capitata cv. Zencibasi	Red cabbage	Turkey	-/-	ND	ND	(LI/M)
	B. oleracea var. capitata cv. Mohrenkopt	Red cabbage	Turkey	-/-	ND	ND	(LI/M)
	B. oleracea L. var. *acephala* DC.	Kale	USA	-/-	-/-	-/-	(LI/M)
	B. oleracea var. italica cv. Challenger	Broccoli	Japan	-/-	-/-	-/-	(LI/M)
	B. oleracea var. italica cv. Pixcel	Broccoli	Chile	-/-	-/-	-/-	(LI/M)
	B. pekinensis cv. Kyoto-3 go	Chinese cabbage	Japan	-/-	-/-	-/-	(LI/M, Y)
	B. pekinensis cv. Nozaki-1 go	Chinese cabbage	Japan	-/-	-/-	-/-	(LI/M, Y)
	B. pekinensis cv. unknown	Chinese cabbage	Turkey	-/-	ND	ND	(LI/M, Y)
	B. rapa cv. Hakatasuwari	Turnip	Japan	-/-	-/-	(LI/CS)[u]	(LI/M)
	B. rapa cv. Ada-202	Turnip	Turkey	-/-	ND	ND	(LI/M)
	Camelina sativa cv. Calena	Gold of Pleasure	Austria	LI/VC, M, St	LI/VC,M, St	LI/VC,M, St	LI/VC,M, St
	Camelina sativa cv. Suneson	Gold of Pleasure	USA	(LI/VC, M, St)	ND	ND	LI/VC,M, St
	Cheiranthus cheiri cv. Vega Yellow	Wallflower	Japan	-/-	ND	-/-	(LI/M)
	Eruca sativa cv. Odyssey	Rocket	Italy	(CS, NS/CS, NS, M)	-/-	(LI/M,CS)	(LI/M)
	E. sativa cv. Unknown	Rocket	Denmark	CS, NS/CS,M,St	-/-	(LI/mM)	(LI/M)
	E. sativa cv. Izmir	Rocket	Turkey	(CS, NS/CS, NS, M)	ND	ND	(LI/M)
	E. sativa cv. Balikesir	Rocket	Turkey	(CS, NS/CS, NS, M)	ND	ND	(LI/M)
	E. sativa cv. Unknown	Rocket	Turkey	(CS, NS/CS, NS, M)	-/-	ND	(LI/M)
	E. sativa cv. unknown (Greek local variety)	Rocket	Greece	-/-	ND	ND	-/-
	Matthiola incana cv. Hatsubeni	Stock	Japan	-/-	ND	ND	LI/Mo, M
	Nasturtium officinale cv. unknown	Watercress	Denmark	-/-	ND	-/-	(LI/M)
	Raphanus sativus cv. Akimasari	Japanese radish	Japan	-/-	-/-	-/-	(LI/M)
	R. sativus cv. Everest	Chinese radish	USA	-/-	-/-	-/-	(LI/M)
	R. sativus cv. Karagulle	Black radish	Turkey	-/-	ND	ND	(LI/M)
	R. sativus cv. Kartopu	White radish	Turkey	-/-	ND	ND	(LI/M)
	R. sativus cv. Kuromaru-kun	Black radish	Italy	-/-	ND	ND	(LI/M)
	R. sativus cv. Comet	Red small radish	USA	-/-	-/-	-/-	(LI/M)
	R. sativus cv. Taibosobutori	Japanese radish	New Zealand	-/-	-/-	-/-	(LI/M)
	R. sativus cv. Tenan-koshin	Chinese radish	Italy	-/-	-/-	-/-	(LI/M)
	Rapistrum rugosum (L.) All.	Annual bastard cabbage	Iran	-/-	ND	ND	(LI/M)
	Sisymbrium loeselii L.	Small tumbleweed-mustard	Iran	-/-	ND	ND	(LI/M)
Chenopodiaceae	*Chenopodium quinoa*	Quinoa	Japan	CS/-	CS/-	CS/-	CS/(M)
	C. amaranticolar	*Chenopodium album* var. album	Japan	CS/-	CS/-	CS/-	CS/(M)
Dioscoreaceae	*Dioscorea japonica*	Japanese yam	Japan	-/-	-/-	-/-	-/-
Liliaceae	*Allium fistulosum* cv. White tower	Green onion	Chile	-/-	-/-	-/-	-/-
	A. fistulosum cv. Kujyo-futo	Green onion	Chile	-/-	-/-	-/-	-/-
Orchidaceae	*Orchis graminifolia*	grass-like leaved orchis	Japan	-/-	-/-	-/-	-/-
Solanaceae	*Nicotiana benthamiana*		Japan	LI/M	NS/M	LI/M	LI/M
	N. clevelandii		Japan	NR/NR, Mo	LI/M	NR/NR, M	NR/NR, M
	N. tabacum cv. Sumsun		Japan	ND	CS/-	CS/-	(CS/-)
	N. tabacum cv. Xanthi nc		Japan	-/-	CS/-	CS/-	(CS/-)

[a] Reaction of inoculated leaves/uninoculated upper leaves. At least three plants were inoculated. BSS; Black stem streak, CS; Chlorotic spot, D; Dead, LD; Leaf deformation, LI; Latent infection, M; Mosaic, Mo; Mottle, ND: Not detected, NLL; Necrotic local lesion, NS; Necrotic spot, NR; Necrotic ringspot, PSF; Pod set failure, St; Stunting, VC; Vein clearing, -; No infection, (); Occasionally. All the leaves were examined for *Turnip mosaic virus* (TuMV) infection by DAS-ELISA. All the absorbance values at 405nm of DAS-ELISA were greater than 1.0 2.5 hrs after adding substrate.

[b] OS and ORM isolates are aphid transmissible but OM-N isolate is not.

[c] Typical symptom for most of TuMV brassica isolates.

[d] Viruses were detected only in upper two leaves, and not in the third.

leaves of the inoculated plants and not in leaves that emerged subsequently. Hence, as no systemic infection with clear symptoms was produced by Orchis isolates in standard brassica test plants, it was clear that they were biologically distinct from BI isolates. Orchis isolates were also inoculated to plants of *Camelina sativa* ('gold of pleasure') and *Eruca sativa* (rocket), both of which are West Eurasian brassicas that have been grown, unselected, as crop plants for several thousand years. OM and ORM isolates infected both *C. sativa* and *E. sativa*, but the OS isolate only infected *C. sativa*.

Isolates OS and ORM were transmitted by *Myzus persicae* (Sulzer) and *Aphis gossypii* (Glover) in a non-persistent manner, whereas OM-N and OM-A isolates were not, even though they were tested several times using more than 100 apterous aphids for each test; *E. sativa* and *C. sativa* were used as source and test plants. The three Orchis isolates were also mechanically inoculated to, but failed to infect, leaves of *Narcissus tazetta* var. *chinensis* (narcissus), *Dioscorea japonica* (Japanese yam), and *Allium fistulosum* (green or Welsh onion, scallion), which are the original hosts of NYSV, JYMV, and ScMV, respectively, and form the sister group to all TuMVs. Standard brassica isolates could be infected most of plants in Brassicaceae family and produced mosaic symptoms on the upper leaves.

2.2. Genome sequences

The genomes of 48 TuMV isolates from Europe amd five from USA were sequenced. These were analysed along with 102 other genomic TuMV sequences, mostly of East Asian isolates (Ohshima *et al*., 2007), obtained from the international sequence databases (Table 4). Most were 9798 nt in length; a few were one to three nucleotides shorter in the terminal UTRs, whereas those of Orchis isolates were 6 nt (i.e., two codons) shorter in the polyprotein region. All of the motifs reported in potyvirus genes, encoded proteins, and the 'Pretty

Interesting *Potyviridae* ORF' (P3N-PIPO) (Chung *et al*., 2008) were present. The P1 gene and its encoded protein were the most variable, and these had few totally conserved residues or compact motifs. The P3 gene and its encoded protein were only slightly less variable (Gibbs and Ohshima, 2010; Ohshima *et al*., 2007). The sequences are available in the GenBank, EMBL, and DDBJ databases with Accession Codes AB701690-AB701742.

2.3. Phylogenetic relationships

Separate phylogenetic trees were estimated for the polyproteins and for the individual HC-Pro, P3, NIb, and CP genes/regions of the 155 isolates. Inconsistent and poorly supported relationships among the resulting trees indicated that some isolates were recombinants, as found previously (Ohshima *et al*., 2002; Tan *et al*., 2005). Accordingly, the sequences for recombination using split decomposition (Hudson and Bryant, 2006) were checked. The Orchis group isolates formed a single non-recombinant lineage distinct from all the BI lineages (basal-B, basal-BR, Asian-BR and world-B), and closest to the basal-B lineage (Figure 16). The extent of the reticulations at the base of the world-B and Asian-BR lineages suggests that most of these sequences are recombinants.

Further analyses using recombination-detection methods confirmed that many of the sequences were recombinants, with only 37 of the sequences showing no significant evidence of recombination. Fifty sequences were interlineage recombinants (i.e., had 'parents' from different lineages; red names in Figure 16, whereas 68 sequences were intralineage recombinants (i.e., had 'parents' from the same lineage; blue names in Figure 16. When the interlineage recombinant sequences were removed, and the remaining sequences analysed again by split decomposition, the branching patterns of the major lineages were resolved (Figure 17).

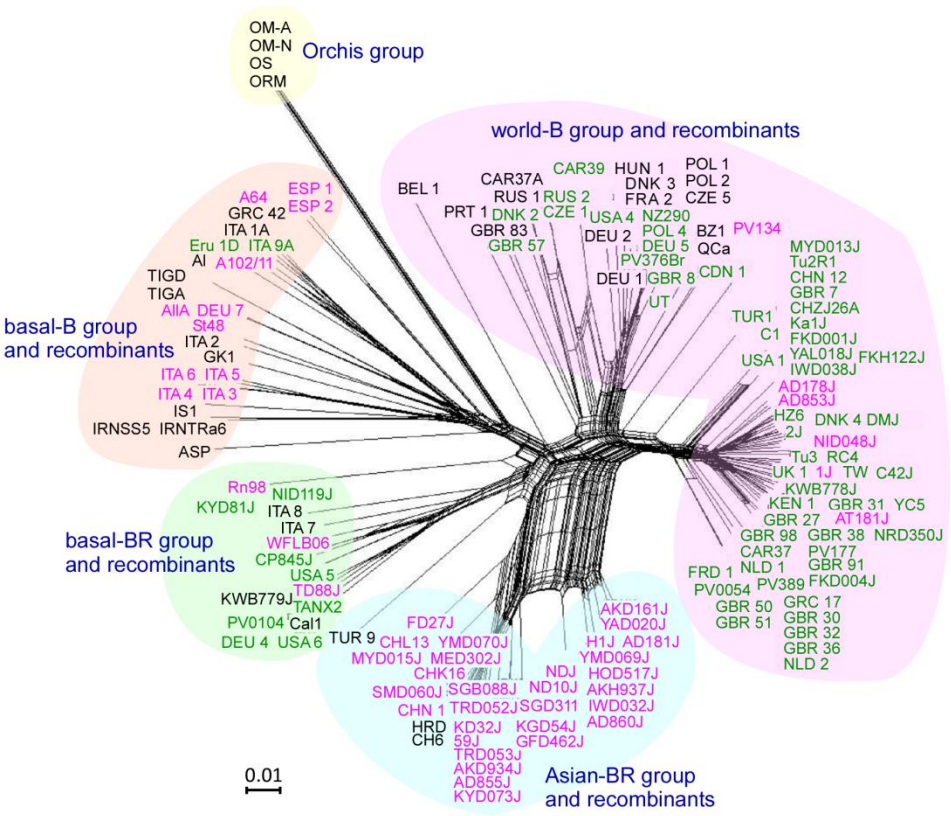


Figure 16. Split-decomposition phylogenetic networks. Networks inferred from polyprotein sequences of 155 *Turnip mosaic virus* (TuMV) isolates. The isolates of non-recombinants (acronyms in black), intrarecombinants (acronyms in green) and interrecombinants (acronyms in pink) are separately listed.

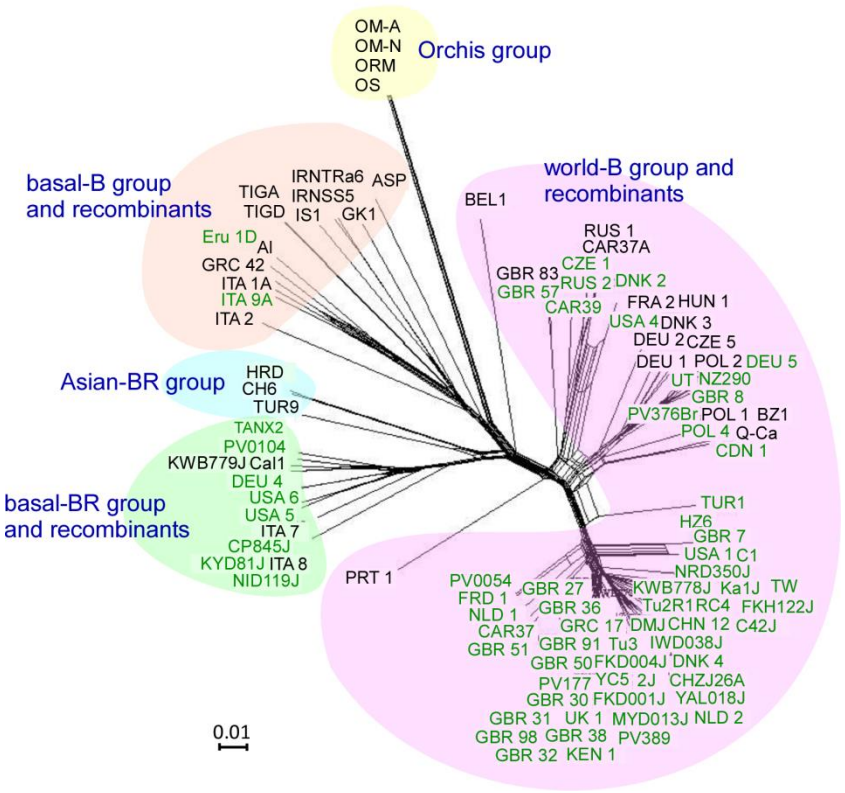


Figure 17. Split-decomposition phylogenetic networks. Networks inferred from polyprotein sequences of 105 sequences remaining after removing interlineage recombinants. The isolates of non-recombinants (acronyms in black) and intrarecombinants (acronyms in green) are separately listed.

Figure 18 shows a maximum-likelihood tree of the amino acid sequences encoded by the few genomic sequences that had no evidence of recombination. It confirms that the OM isolates form a monophyletic lineage that is sister to the BI lineages and closest to the basal-B group. No recombination sites were detected in the 5' and 3' non-coding regions. However, they were found throughout the coding regions of many of the genomes (Table 14), especially in the P1 gene and CI-VPg regions as reported previously (Ohshima *et al*., 2007). Therefore, the HC-Pro, P3, NIb, and CP genes/regions of the genomes that were not intralineage recombinants, and showed no intragenic recombination, were selected for phylogenetic analyses using maximum likelihood in PhyML version 3 (Guindon and Gascuel, 2003) and neighbor-joining in PHYLIP (Felsenstein, 1993). The resulting trees grouped the BI sequences into the four major groups previously reported (Ohshima *et al*., 2002, 2007), with the Orchis isolates grouped as a monophyletic sister lineage to all of the other TuMV lineages. An exception to this occurred in the HC-Pro trees, where the Orchis group lineage was sister to the basal-B lineage. All of these topologies were supported by high bootstrap values. These results raised the question as to whether the Orchis isolates were closer to the BI isolates or to the outgroup viruses. In a maximum-likelihood phylogeny of 37 non-recombinant TuMV-BI genomes, the four Orchis isolates genomes, and the four outgroup genomes, the mean patristic distances between the outgroup sequences and the BI and Orchis isolates genomes were 7.97 ± 1.06 and 7.63 ± 1.09 subs/site, respectively, but only 1.01 ± 0.03 subs/site between the BI and Orchis group genomes. Thus, the BI and Orchis isolates genomes represent distinct populations that are much closer to one another than to the outgroup viruses. Figure 19 shows in detail that the sequences of the Orchis group isolates are closer to those of TuMV isolates throughout the genome, and both are very different from the outgroup genomes. They are also closer in terms of other characteristics,

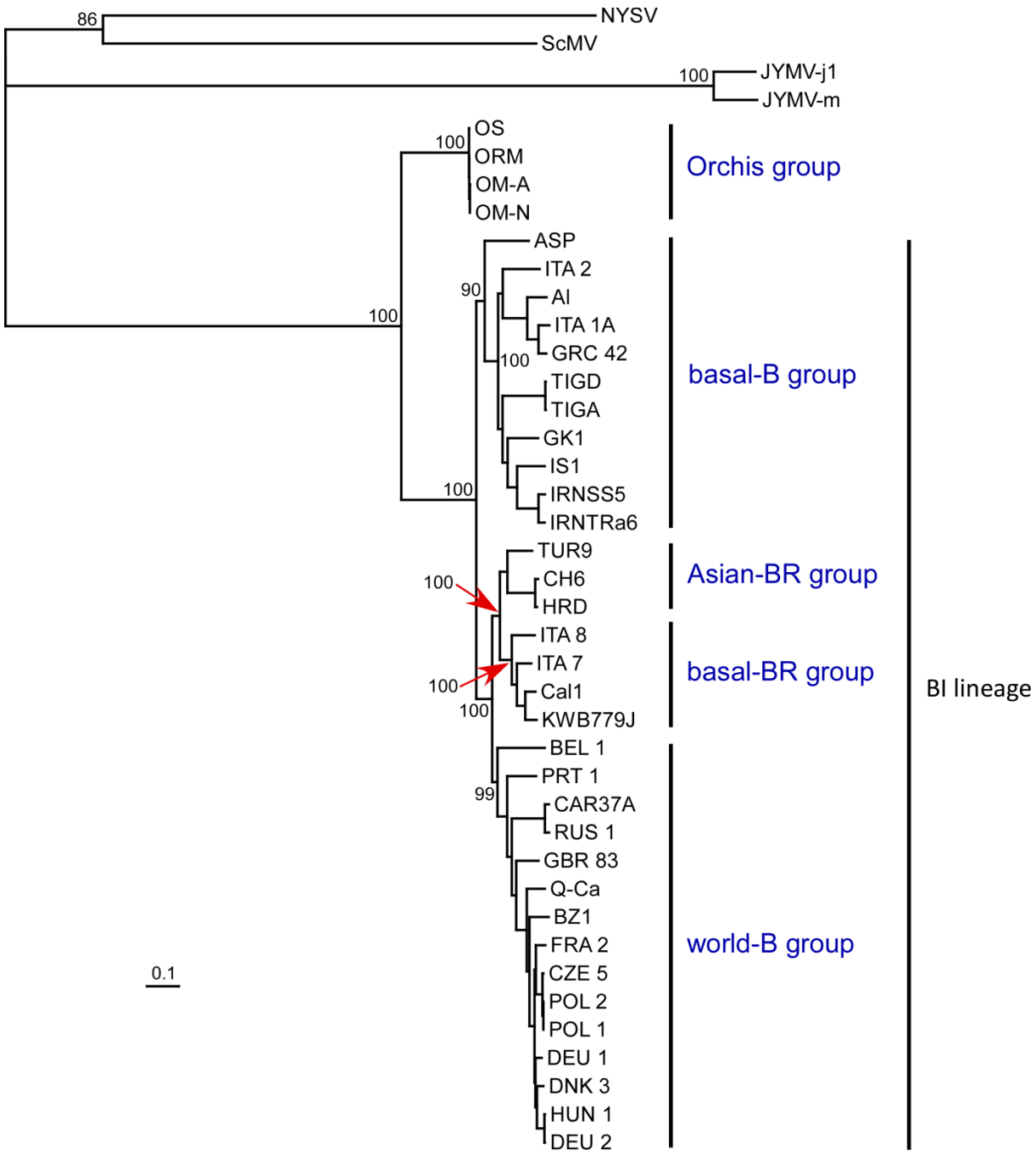


Figure 18. Maximum-likelihood tree of the complete polyprotein sequences of 37 non-recombinant *Turnip mosaic virus* isolates**.** Nodes are labelled with bootstrap support percentages.

Table 14. Recombination sites in full genomic sequences.

Isolate	Recombinant type	Recombinant type pattern	Recombination site	Parental sequence (group)	RDP	GENECONV	BootScan	MaxChi	Chimaera	SiScan	PhylPro
					P-value					P and Z-value	
1J	Inter- recombinant	ABR x wB3 x ABR x wB3 x wB2	183	CH6(ABR) x FKH122J(wB3)	1.15×10^{-6}	2.90×10^{-6}	7.23×10^{-7}	4.33×10^{-6}	1.11×10^{-4}	1.56×10^{-6}	
			750	DNK4(wB3) x CHN1(ABR)	2.63×10^{-22}	1.98×10^{-20}	1.63×10^{-22}	1.55×10^{-8}	2.18×10^{-8}	1.54×10^{-11}, Z>6.52	
			948	YAD020J(ABR) x GBR7(wB3)	2.38×10^{-27}	8.89×10^{-27}	2.32×10^{-37}	1.93×10^{-17}	5.30×10^{-18}	1.80×10^{-20}, Z>6.52	Detected
			6293	GBR36(wB3) x CDN1/Q-Ca(wB2)	1.28×10^{-20}	1.65×10^{-11}	9.49×10^{-20}	6.58×10^{-8}	1.41×10^{-9}	1.29×10^{-13}, Z>4.2	
2J	Intra-recombinant	wB2 x wB3 x wB3 x wB2	643	PV134(wB3) x Ka1J(wB3)	1.15×10^{-32}	3.69×10^{-19}	3.69×10^{-33}	1.47×10^{-6}	9.65×10^{-8}	7.80×10^{-14}, Z>3.98	
			2194	MYD013J(wB3) x Tu2R1(wB3)	6.03×10^{-14}	4.91×10^{-13}	3.58×10^{-14}	6.03×10^{-12}	2.96×10^{-7}	2.21×10^{-15}	
			6293	GBR36(wB3) x CDN1/Q-Ca(wB2)	1.63×10^{-13}	1.39×10^{-3}	9.49×10^{-20}	9.56×10^{-4}	8.78×10^{-6}	8.37×10^{-9}, Z>4.2	
59J	Inter- recombinant	ABR x wB3 x ABR x wB3	200	CH6(ABR) x USA4(wB3)	1.50×10^{-46}	6.76×10^{-22}	1.30×10^{-46}	5.72×10^{-14}	3.63×10^{-14}	2.67×10^{-14}, Z>5,78	
			756	USA4(wB3) x CH6(ABR)	1.50×10^{-46}	6.76×10^{-22}	1.30×10^{-46}	5.72×10^{-14}	3.63×10^{-14}	2.67×10^{-14}, Z>6.57	
			9152	CHL13(ABR) x HZ6(wB3)	9.98×10^{-17}	5.69×10^{-13}	6.54×10^{-18}	1.40×10^{-3}	3.71×10^{-1}	1.36×10^{-4}, Z>3.44	
C42J	Intra-recombinant	wB3 x wB3 x wB2	739	YAD020J(wB3) x GBR7(wB3)	6.64×10^{-29}	1.78×10^{-16}	2.33×10^{-29}	6.65×10^{-16}	1.47×10^{-16}	3.06×10^{-15}	
			6300	GBR36(wB3) x CDN1(wB2)	3.68×10^{-22}	5.71×10^{-14}	1.54×10^{-22}	7.84×10^{-8}	8.28×10^{-9}	5.03×10^{-13}, Z>6.5	
DMJ	Intra-recombinant	wB2 x wB3 x wB3 x wB3 x wB2	695	DNK2(wB2) x Ka1J(wB3)	1.21×10^{-22}	2.95×10^{-9}	6.28×10^{-23}	1.26×10^{-5}	4.55×10^{-6}	2.68×10^{-9}	
			1227	FKD004J(wB3) x HZ6(wB3)	1.03×10^{-12}	6.68×10^{-4}	2.10×10^{-12}	3.58×10^{-7}	6.57×10^{-9}	1.19×10^{-11}, Z>4.35	
			2056	HZ6(wB3) x FKD004J(wB3)	1.03×10^{-12}	6.68×10^{-4}	2.10×10^{-12}	3.58×10^{-7}	6.57×10^{-9}	1.19×10^{-11}, Z>4.35	
			6293	GBR36(wB3) x CDN1/Q-Ca(wB2)	3.80×10^{-21}	1.97×10^{-12}	1.60×10^{-21}	6.16×10^{-8}	2.34×10^{-9}	2.15×10^{-12}, Z>4.2	
Ka1J	Intra-recombinant	wB3 x wB3 x wB2	748	YAD020J(wB3) x GBR7(wB3)	3.37×10^{-32}	7.49×10^{-23}	6.57×10^{-32}	2.10×10^{-14}	1.93×10^{-14}	1.17×10^{-16}	
			6293	GBR36(wB3) x CDN1/Q-Ca(wB2)	1.29×10^{-16}	6.66×10^{-5}	1.17×10^{-16}	9.20×10^{-7}	7.76×10^{-8}	1.30×10^{-10}, Z>4.2	
NDJ	Inter-recombinant	ABR x wB3 x ABR x wB3	225	CH6(ABR) x USA4(wB3)	2.76×10^{-45}	6.70×10^{-32}	2.48×10^{-45}	1.55×10^{-10}	4.70×10^{-12}	9.55×10^{-14}	
			728	USA4(wB3) x CH6(ABR)	2.76×10^{-45}	6.70×10^{-32}	2.48×10^{-45}	1.55×10^{-10}	4.70×10^{-12}	9.55×10^{-14}	
			6450	MYD015J(ABR) x 2J(wB3)	2.18×10^{-144}	3.70×10^{-137}	5.20×10^{-144}	2.68×10^{-39}	2.63×10^{-28}	4.96×10^{-47}	
FD27J	Inter-recombinant	ABR x wB3 x ABR x wB3 x ABR	183	CH6(ABR) x FKH122J(wB3)	1.10×10^{-8}	3.37×10^{-8}	7.54×10^{-9}	8.05×10^{-3}	2.36×10^{-3}	1.40×10^{-7}, Z>6.07	
			750	DNK4(wB3) x CHN1(ABR)	7.46×10^{-15}	3.71×10^{-16}	5.12×10^{-15}	8.69×10^{-6}	8.78×10^{-6}	1.67×10^{-11}, Z>6.07	
			931	HRD(ABR) x FKD004J(wB3)	1.43×10^{-73}	1.08×10^{-64}	1.30×10^{-70}	2.68×10^{-35}	2.16×10^{-26}	1.85×10^{-33}, Z>6.07	Detected
			2989	FKD004J(wB3) x HRD(ABR)	1.43×10^{-73}	1.08×10^{-64}	1.30×10^{-70}	2.68×10^{-35}	2.16×10^{-26}	1.85×10^{-33}, Z>6.07	Detected
KD32J	Inter-recombinant	ABR x wB3 x ABR x wB3	224	CH6(ABR) x USA4(wB3)	2.21×10^{-41}	2.45×10^{-22}	3.08×10^{-42}	2.92×10^{-11}	6.01×10^{-12}	3.97×10^{-13}	
			728	USA4(wB3) x CH6(ABR)	2.21×10^{-41}	2.45×10^{-22}	3.08×10^{-42}	2.92×10^{-11}	6.01×10^{-12}	3.97×10^{-13}	
			9175	CHL13(ABR) x HZ6(wB3)	1.74×10^{-15}	5.48×10^{-12}	1.27×10^{-16}	5.01×10^{-3}	3.71×10^{-1}	3.82×10^{-4}, Z>3.25	
KYD81J	Intra-recombinant	bBR x bBR	867	TANX2(bBR) x USA6(bBR)	8.04×10^{-10}	2.81×10^{-3}	2.70×10^{-9}	5.01×10^{-8}	3.84×10^{-7}	1.33×10^{-10}	
TD88J	Inter-recombinant	bBR x ABR x bBR x bBR	2449	DEU4(bBR) x ND10J(ABR)	5.16×10^{-76}	3.43×10^{-66}	3.35×10^{-73}	1.41×10^{-17}	8.70×10^{-20}	3.64×10^{-22}	Detected
			3016	ND10J(ABR) x DEU4(bBR)	5.16×10^{-76}	3.43×10^{-66}	3.35×10^{-73}	1.41×10^{-17}	8.70×10^{-20}	3.64×10^{-22}	
			6541	TANX2(bBR) x NID119J(bBR)	1.13×10^{-29}	2.69×10^{-17}	7.31×10^{-29}	1.66×10^{-13}	6.36×10^{-11}	3.68×10^{-18}	
CP845J	Intra-recombinant	bBR x bBR x bBR	867	TANX2(bBR) x USA6(bBR)	1.60×10^{-12}	4.08×10^{-5}	1.12×10^{-12}	6.91×10^{-10}	2.82×10^{-9}	1.23×10^{-13}	
			1713	ITA7(bBR) x KWB779J/PV0104(bBR)	1.26×10^{-9}	1.11×10^{-1}	5.15×10^{-10}	1.13×10^{-8}	5.12×10^{-10}	4.39×10^{-20}, Z<3.0	
HOD517J	Inter-recombinant	wB3 x ABR x wB3 x ABR x wB3	756	USA4(wB3) x CH6(ABR)	1.14×10^{-47}	3.16×10^{-13}	2.52×10^{-47}	3.43×10^{-17}	6.81×10^{-17}	3.84×10^{-15}	
			5284	SMD060J(ABR) x FKD001J(wB3)	2.04×10^{-31}	2.63×10^{-22}	2.03×10^{-30}	1.89×10^{-7}	1.80×10^{-7}	5.35×10^{-9}	
			5732	FKD001J(wB3) x SMD060J(ABR)	2.04×10^{-31}	2.63×10^{-22}	2.03×10^{-30}	1.89×10^{-7}	1.80×10^{-7}	5.35×10^{-9}	Detected
			6097	MYD015J(ABR) x 2J(wB3)	1.00×10^{-103}	2.17×10^{-51}	6.87×10^{-100}	1.70×10^{-37}	7.89×10^{-4}	1.21×10^{-44}	Detected
SGD311J	Inter-recombinant	ABR x wB3 x ABR x wB3 x ABR	190	CH6(ABR) x USA4(wB3)	7.19×10^{-36}	1.11×10^{-15}	1.25×10^{-35}	3.93×10^{-11}	1.41×10^{-11}	7.89×10^{-13}	
			745	USA4(wB3) x CH6(ABR)	7.19×10^{-36}	1.11×10^{-15}	1.25×10^{-35}	3.93×10^{-11}	1.41×10^{-11}	7.89×10^{-13}	
			5013	SMD060J(ABR) x FKD001J(wB3)	6.83×10^{-43}	9.92×10^{-30}	1.53×10^{-41}	8.59×10^{-11}	2.88×10^{-12}	5.14×10^{-15}	Detected
			5726	FKD001J(wB3) x SMD060J(ABR)	6.83×10^{-43}	9.92×10^{-30}	1.53×10^{-41}	8.59×10^{-11}	2.88×10^{-12}	5.14×10^{-15}, Z>6.58	Detected
Al	Non-recombinant	basal-B1 (bB1)									
A64	Inter-recombinant	bB1 x wB3	9285	Al(bB1) x CZE1(wB3)	8.54×10^{-10}	8.58×10^{-4}	4.74×10^{-11}	4.71×10^{-1}	ND	2.56×10^{-8}, Z>3.88	
A102/11	Inter-recombinant	wB3/bBR x bB1 x bB1 x wB1	605	Rn98(wB3/bBR?) x GRC42(bB1)	1.23×10^{-7}	9.62×10^{-13}	1.98×10^{-8}	2.22×10^{-3}	1.42×10^{-1}	1.38×10^{-15}	
			2020	ITA1A(bB1) x Eru1D(bB1)	3.49×10^{-54}	1.10×10^{-83}	1.24×10^{-50}	6.34×10^{-38}	1.34×10^{-22}	2.88×10^{-45}	
			9183	Al(bB1) x CZE1(wB1)	2.97×10^{-10}	2.56×10^{-3}	1.28×10^{-10}	4.89×10^{-4}	ND	7.60×10^{-6}	
QCa	Non-recombinant	world-B2 (wB2)									

Table 14. Continued.

Isolate	Recombinant type	Recombinant type pattern	Recombination site	Parental sequence (group)	RDP	GENECONV	BootScan	MaxChi	Chimaera	SiScan	PhylPro
					P-value					P and Z-value	
CDN1	Intra-recombinant	wB2 x wB3 x wB2	5200	QCa(wB2) x YC5 (wB3)						Z>8.57	
			5300	YC5 (wB3) x QCa (wB2)						Z>5.93	Detected
CHN12	Intra-recombinant	wB3 x wB3 x wB2	748	YAD020J(wB3) x GBR7(wB3)	4.06×10^{-33}	4.91×10^{-14}	2.51×10^{-33}	4.44×10^{-13}	8.53×10^{-15}	7.74×10^{-16}	
			6293	GBR36(wB3) x CDN1/Q-Ca(wB2)	5.40×10^{-24}	6.79×10^{-18}	6.10×10^{-25}	3.68×10^{-8}	5.55×10^{-9}	1.41×10^{-13}, Z>4.2	
NZ290	Intra-recombinant	wB2 x wB2	7992	DNK3(wB2) x CDN1(wB2)	1.23×10^{-15}	3.79×10^{-7}	9.91×10^{-16}	5.93×10^{-7}	1.43×10^{-7}	3.53×10^{-12}	
UK1	Intra-recombinant	wB3 x wB3 x wB2	738	YAD020J(wB3) x GBR7(wB3)	2.28×10^{-35}	1.46×10^{-21}	5.57×10^{-36}	5.32×10^{-15}	2.19×10^{-15}	6.67×10^{-16}	
			6293	GBR36(wB3) x CDN1/Q-Ca(wB2)	1.31×10^{-21}	4.64×10^{-12}	7.06×10^{-22}	1.94×10^{-8}	6.92×10^{-9}	2.58×10^{-13}, Z>4.2	
CHN1	Inter-recombinant	ABR x wB3	8768	MYD015J(ABR) x 2J(wB3)	4.27×10^{-47}	1.14×10^{-76}	6.50×10^{-45}	1.04×10^{-34}	2.55×10^{-27}	1.97×10^{-84}	Detected
HRD	Non-recombinant	Asian-BR (ABR)									
KEN 1	Intra-recombinant	wB3 x wB2	6293	GBR36(wB3) x CDN1/Q-Ca(wB2)	2.98×10^{-16}	7.56×10^{-9}	1.74×10^{-16}	5.27×10^{-7}	3.09×10^{-7}	2.46×10^{-10}, Z>4.2	
PV376Br	Intra-recombinant	wB3 x wB2 x wB2	7968	DNK3(wB2) x GBR7(wB2)	1.47×10^{-18}	2.04×10^{-1}	4.08×10^{-18}	4.50×10^{-8}	ND	2.30×10^{-9}, Z>4.95	
			8004	DNK3(wB2) x CDN1(wB2)	4.55×10^{-13}	5.93×10^{-2}	1.11×10^{-13}	1.30×10^{-8}	3.93×10^{-10}	3.06×10^{-10}, Z>4.91	
BZ1	Non-recombinant	world-B2 (wB2)									
Cal1	Non-recombinant	basal-BR (bBR)									
IS1	Non-recombinant	basal-B2 (bB2)									
PV0104	Intra-recombinant	bBR x bBR x bBR	922	ITA7(bBR) x TANX2(bBR)	7.98×10^{-29}	2.29×10^{-25}	8.19×10^{-30}	4.63×10^{-26}	2.70×10^{-19}	1.40×10^{-44}	
			5965	KWB779J(bBR) x ITA7(bBR)	6.74×10^{-17}	2.95×10^{-15}	2.93×10^{-16}	1.56×10^{-22}	4.26×10^{-19}	1.02×10^{-5}	
St48	Inter-recombinant	wB3 x bB1	1543	FRD1(wB3) x A64(bB1)	7.13×10^{-6}	1.03×10^{-5}	3.54×10^{-6}	8.73×10^{-6}	3.12×10^{-7}	1.36×10^{-18}, Z>6.04	
Rn98	Inter-recombinant	wB3 x bB1	1543	TANX2(bBR) x TIGD(bB1)	2.31×10^{-16}	ND	1.88×10^{-13}	3.40×10^{-3}	6.20×10^{-6}	8.25×10^{-11}	
			1600 (tentative)	FRD1 x A64	4.77×10^{-5}	4.30×10^{-2}	4.56×10^{-5}	1.18×10^{-1}	1.31×10^{-1}	1.98×10^{-16}, Z>3.8	
ITA 7	Non-recombinant	basal-BR (bBR)									
CZE 1	Intra-recombinant	wB3 x wB2 x wB3	6009,	DNK2(wB3) x DNK3(wB2)	7.67×10^{-76}	1.77×10^{-71}	1.62×10^{-73}	2.08×10^{-34}	3.35×10^{-20}	7.78×10^{-42}	
			8767	DNK3(wB2) x DNK2(wB3)	3.72×10^{-76}	8.50×10^{-72}	7.88×10^{-74}	1.01×10^{-34}	1.62×10^{-20}	2.82×10^{-42}	
RUS 1	Non-recombinant	world-B1 (wB1)									
RUS 2	Intra-recombinant	wB3 x wB2 x wB3	6009	DNK2(wB3) x DNK3(wB2)	1.88×10^{-79}	3.19×10^{-74}	4.96×10^{-77}	1.47×10^{-34}	3.35×10^{-20}	3.02×10^{-42}	
			8791	DNK3(wB2) x DNK2(wB3)	9.11×10^{-80}	1.60×10^{-74}	2.40×10^{-77}	7.12×10^{-35}	1.62×10^{-20}	1.57×10^{-42}	
C1	Intra-recombinant	wB3 x wB2 x wB3 x wB2 x wB2	4016	USA1(wB3) x QCa(wB2)	3.51×10^{-41}	3.07×10^{-35}	3.18×10^{-41}	4.80×10^{-10}	7.63×10^{-10}	5.69×10^{-11}	Detected
			4445	QCa(wB2) x USA1(wB3)	1.70×10^{-41}	1.49×10^{-35}	1.54×10^{-41}	2.33×10^{-10}	9.58×10^{-10}	1.86×10^{-12}	Detected
			5551	QCa(wB2) x USA1(wB3)	5.44×10^{-20}	9.56×10^{-18}	5.36×10^{-20}	8.40×10^{-3}	9.43×10^{-2}	6.81×10^{-4}	
			6293	GBR36(wB3) x CDN1/Q-Ca(wB2)	4.71×10^{-11}	1.44×10^{-2}	1.63×10^{-11}	6.29×10^{-7}	2.64×10^{-7}	3.98×10^{-13}, Z>4.2	
TW	Intra-recombinant	wB3 x wB3 x wB2	686	YAD020J(wB3) x GBR7(wB3)	4.64×10^{-28}	9.31×10^{-14}	5.32×10^{-28}	1.22×10^{-13}	9.20×10^{-15}	3.04×10^{-15}	
			6293	GBR36(wB3) x CDN1/Q-Ca(wB2)	2.95×10^{-20}	5.26×10^{-9}	1.13×10^{-20}	6.54×10^{-7}	3.21×10^{-8}	4.85×10^{-13}, Z>4.2	
Tu3	Intra-recombinant	wB3 x wB3 x wB2	738	YAD020J(wB3) x GBR7(wB3)	2.78×10^{-27}	5.11×10^{-18}	1.81×10^{-27}	5.30×10^{-13}	2.59×10^{-14}	2.47×10^{-13}	
			6293	GBR36(wB3) x CDN1/Q-Ca(wB2)	3.07×10^{-23}	1.43×10^{-12}	1.38×10^{-23}	1.10×10^{-9}	1.22×10^{-9}	5.35×10^{-13}, Z>4.2	
Tu2-R1	Intra-recombinant	wB3 x wB3 x wB2	748	YAD020J(wB3) x GBR7(wB3)	3.40×10^{-29}	1.36×10^{-19}	4.44×10^{-29}	2.01×10^{-14}	1.32×10^{-14}	7.07×10^{-16}	
			6293	GBR36(wB3) x CDN1/Q-Ca(wB2)	4.50×10^{-14}	4.02×10^{-3}	4.74×10^{-14}	9.04×10^{-6}	1.53×10^{-7}	4.25×10^{-9}, Z>4.2	
RC4	Intra-recombinant	wB3 x wB3 x wB2 x wB2/wB3	747	YAD020J(wB3) x GBR7(wB3)	7.02×10^{-26}	1.21×10^{-10}	3.73×10^{-26}	4.41×10^{-13}	4.02×10^{-15}	2.56×10^{-14}	
			6293	GBR36(wB3) x CDN1/Q-Ca(wB2)	1.24×10^{-19}	1.30×10^{-9}	4.49×10^{-20}	4.30×10^{-8}	5.37×10^{-10}	6.80×10^{-13}, Z>4.2	
			7406-7414	DNK2(wB3) x NLD1(wB2/wB3)	7.12×10^{-8}	1.43×10^{-3}	7.88×10^{-8}	5.05×10^{-6}	4.05×10^{-6}	2.42×10^{-10}, Z<3.0	
YC5	Intra-recombinant	wB3 x wB3 x wB2	747	YAD020J(wB3) x GBR7(wB3)	5.63×10^{-29}	1.43×10^{-16}	1.02×10^{-29}	3.85×10^{-15}	2.22×10^{-16}	2.42×10^{-14}	
			6293	GBR36(wB3) x CDN1/Q-Ca(wB2)	1.92×10^{-22}	1.93×10^{-11}	7.15×10^{-23}	1.55×10^{-9}	8.36×10^{-10}	5.90×10^{-15}, Z>4.2	
GBR 36	Intra-recombinant	wB3 x wB3	746	YAD020J(wB3) x GBR7(wB3)	1.22×10^{-36}	4.11×10^{-19}	2.38×10^{-36}	3.34×10^{-15}	3.60×10^{-15}	2.47×10^{-15}	
GBR 50	Intra-recombinant	wB3 x wB3 x wB2	760	YAD020J(wB3) x GBR7(wB3)	1.17×10^{-39}	7.30×10^{-26}	4.45×10^{-40}	1.61×10^{-15}	1.77×10^{-16}	1.06×10^{-16}	
			6293	GBR36(wB3) x CDN1/Q-Ca(wB2)	4.86×10^{-16}	1.06×10^{-4}	2.75×10^{-16}	2.85×10^{-8}	4.04×10^{-8}	9.97×10^{-12}, Z>4.2	
GRC 17	Intra-recombinant	wB3 x wB3 x wB2	738	YAD020J(wB3) x GBR7(wB3)	3.59×10^{-30}	9.43×10^{-20}	1.20×10^{-30}	8.27×10^{-15}	9.47×10^{-15}	2.32×10^{-14}	
			6293	GBR36(wB3) x CDN1/Q-Ca(wB2)	1.23×10^{-22}	6.96×10^{-10}	6.81×10^{-23}	2.35×10^{-10}	6.02×10^{-11}	9.84×10^{-14}, Z>4.2	
GRC42	Non-recombinant	basal-B1 (bB1)									
ITA 3	Inter-recombinant	bBR x bB2	644	ITA8(bBR) x IS1(bB2)	3.45×10^{-24}	1.10×10^{-16}	1.66×10^{-23}	2.67×10^{-6}	3.08×10^{-6}	3.61×10^{-17}	
DNK 2	Intra-recombinant	wB2 x wB1	2453	DNK3(wB2) x CAR37A(wB1)	3.41×10^{-124}	7.55×10^{-117}	2.07×10^{-121}	1.28×10^{-39}	3.38×10^{-26}	3.71×10^{-42}	
CZE 5	Non-recombinant	world-B2									
FRD 1	Intra-recombinant	wB2 x wB2/wB3 x wB2	711	NLD1(wB2) x DNK3(wB2)	5.52×10^{-16}	1.14×10^{-4}	2.35×10^{-16}	4.76×10^{-5}	5.02×10^{-7}	8.27×10^{-7}, Z>4.02	
			6550	GBR36(wB3) x CDN1(wB2)	7.33×10^{-10}	5.68×10^{-7}	1.49×10^{-9}	1.04×10^{-4}	4.03×10^{-5}	6.87×10^{-12}, Z>5.17	
			7406-7414(tentative)	DNK2 x NLD1	ND	ND	ND	ND	ND	Z>4.2	

Table 14. Continued.

Isolate	Recombinant type	Recombinant type pattern	Recombination site	Parental sequence (group)	RDP	GENECONV	BootScan	MaxChi	Chimaera	SiScan	PhylPro
					P-value					P and Z-value	
NLD 1	Intra-recombinant	wB2 x wB2	6293	GBR36(wB3) x CDN1/Q-Ca(wB2)	4.69×10^{-17}	6.40×10^{-8}	1.66×10^{-17}	4.54×10^{-9}	7.42×10^{-10}	3.23×10^{-14}, Z>4.2	
H1J	Inter-recombinant	wB3 x ABR x wB3 x wB2 x ABR x wB3	758	USA4(wB3) x CH6(ABR)	2.17×10^{-47}	4.21×10^{-27}	1.83×10^{-46}	4.41×10^{-17}	6.26×10^{-17}	3.22×10^{-16}	
			5207	MYD015J(ABR) x 2J(wB3)	2.78×10^{-114}	1.80×10^{-108}	5.77×10^{-112}	6.56×10^{-45}	2.26×10^{-26}	1.80×10^{-49}	
			6293	GBR36(wB3) x CDN1/Q-Ca(wB2)	1.23×10^{-22}	6.96×10^{-10}	6.81×10^{-23}	2.35×10^{-10}	6.02×10^{-11}	9.84×10^{-14}, Z>4.2	
			8788	YAL018J(wB3) x TRD052J(ABR)	1.34×10^{-27}	1.31×10^{-24}	5.66×10^{-28}	5.93×10^{-11}	9.34×10^{-12}	3.96×10^{-14}	Detected
			9304	TRD052J(ABR) x YAL018J(wB3)	1.34×10^{-27}	1.31×10^{-24}	5.66×10^{-28}	5.93×10^{-11}	9.34×10^{-12}	3.96×10^{-14}	
AD178J	Inter-recombinant	wB3 x ABR x wB2	748	GBR7(wB3) x YAD020J(ABR)	6.43×10^{-30}	2.68×10^{-21}	1.35×10^{-29}	4.55×10^{-14}	1.07×10^{-14}	2.33×10^{-16}	
			6293	GBR36(wB3) x CDN1/Q-Ca(wB2)	1.17×10^{-16}	ND	5.00×10^{-17}	3.26×10^{-8}	1.14×10^{-8}	4.24×10^{-11}, Z>4.2	
AD181J	Inter-recombinant	wB3 x ABR x wB3 x wB2 x ABR x wB3	756	USA4(wB3) x CH6(ABR)	4.01×10^{-55}	1.25×10^{-32}	3.22×10^{-49}	2.70×10^{-17}	3.80×10^{-17}	3.22×10^{-17}	
			5207	MYD015J(ABR) x 2J(wB3)	1.17×10^{-106}	3.03×10^{-101}	8.54×10^{-105}	1.83×10^{-44}	1.40×10^{-25}	3.51×10^{-47}	
			6293	GBR36(wB3) x CDN1/Q-Ca(wB2)	1.23×10^{-22}	6.96×10^{-10}	6.81×10^{-23}	2.35×10^{-10}	6.02×10^{-11}	9.84×10^{-14}, Z>4.2	
			8788	YAL018J(wB3) x TRD052J(ABR)	6.37×10^{-35}	1.76×10^{-32}	3.05×10^{-35}	3.01×10^{-12}	1.82×10^{-12}	3.21×10^{-15}	Detected
			9304	TRD052J(ABR) x YAL018J(wB3)	6.37×10^{-35}	1.76×10^{-32}	3.05×10^{-35}	3.01×10^{-12}	1.82×10^{-12}	3.21×10^{-15}	
AD853J	Inter-recombinant	wB3 x ABR x wB2	739	GBR7(wB3) x YAD020J(ABR)	7.61×10^{-34}	1.24×10^{-22}	7.70×10^{-34}	1.55×10^{-13}	6.55×10^{-15}	9.76×10^{-16}, Z>6.41	
			6293	GBR36(wB3) x CDN1/Q-Ca(wB2)	4.85×10^{-25}	4.44×10^{-14}	1.99×10^{-25}	6.57×10^{-10}	9.73×10^{-11}	2.83×10^{-15}, Z>5.99	
AD855J	Inter-recombinant	ABR x wB3 x ABR	224	CH6(ABR) x USA4(wB3)	2.04×10^{-48}	3.63×10^{-29}	3.20×10^{-48}	6.82×10^{-13}	3.56×10^{-13}	5.99×10^{-14}	
			729	HZ6(wB3) x HRD(ABR)	1.57×10^{-64}	9.24×10^{-54}	1.45×10^{-63}	5.75×10^{-14}	6.60×10^{-15}	9.55×10^{-18}	
AD860J	Inter-recombinant	wB3 x ABR x wB3 x wB2	756	USA4(wB3) x CH6(ABR)	3.40×10^{-42}	4.38×10^{-9}	1.18×10^{-41}	1.59×10^{-16}	9.01×10^{-17}	1.82×10^{-14}	
			5122	MYD015J(ABR) x 2J(wB3)	4.50×10^{-100}	7.79×10^{-121}	6.95×10^{-99}	5.54×10^{-45}	1.04×10^{-24}	1.58×10^{-50}	Detected
			6293	GBR36(wB3) x CDN1/Q-Ca(wB2)	1.23×10^{-22}	6.96×10^{-10}	6.81×10^{-23}	2.35×10^{-10}	6.02×10^{-11}	9.84×10^{-14}, Z>4.2	
AT181J	Inter-recombinant	wB3 x ABR x wB3 x wB2	756	USA4(wB3) x CH6(ABR)	1.52×10^{-14}	2.59×10^{-11}	9.96×10^{-14}	1.56×10^{-16}	9.01×10^{-8}	6.25×10^{-16}	
			1559	YAD020J(ABR) x GBR7(wB3)	3.58×10^{-100}	4.73×10^{-96}	8.21×10^{-104}	2.51×10^{-34}	1.17×10^{-23}	1.01×10^{-46}	Detected
			6293	GBR36(wB3) x CDN1/Q-Ca(wB2)	8.70×10^{-24}	2.77×10^{-15}	1.76×10^{-24}	1.50×10^{-9}	8.01×10^{-10}	1.44×10^{-15}, Z>4.2	
AKD161J	Inter-recombinant	wB3 x ABR x wB3	756	USA4(wB3) x CH6(ABR)	3.36×10^{-41}	3.63×10^{-12}	2.58×10^{-40}	1.82×10^{-16}	4.92×10^{-17}	9.84×10^{-14}	
			5122	MYD015J(ABR) x 2J(wB3)	3.15×10^{-100}	6.03×10^{-117}	1.96×10^{-99}	1.16×10^{-44}	5.60×10^{-24}	3.09×10^{-51}	Detected
AKD934J	Inter-recombinant	wB3 x ABR	902	FKD004J(wB3) x HRD(ABR)	1.20×10^{-75}	3.38×10^{-45}	6.04×10^{-74}	7.81×10^{-22}	1.43×10^{-20}	2.05×10^{-20}	Detected
AKH937J	Inter-recombinant	wB3 x ABR x wB3 x wB2 xwB2/wB3	756	USA4(wB3) x CH6(ABR)	3.68×10^{-43}	3.27×10^{-12}	1.74×10^{-42}	2.49×10^{-16}	7.01×10^{-17}	6.88×10^{-15}	
			5122	MYD015J(ABR) x 2J(wB3)	7.88×10^{-99}	3.62×10^{-115}	1.20×10^{-97}	2.88×10^{-44}	3.87×10^{-24}	3.69×10^{-50}	Detected
			6293	GBR36(wB3) x CDN1/Q-Ca(wB2)	1.23×10^{-22}	6.96×10^{-10}	6.81×10^{-23}	2.35×10^{-10}	6.02×10^{-11}	9.84×10^{-14}, Z>4.2	
			7406-7414	DNK2(wB2) x NLD1(wB2/wB3)	6.04×10^{-18}	1.11×10^{-16}	3.94×10^{-6}	1.82×10^{-5}	2.12×10^{-6}	3.33×10^{-5}, Z<3.0	
IWD032J	Inter-recombinant	wB3 x ABR x wB3	716	USA4(wB3) x CH6(ABR)	1.86×10^{-49}	4.03×10^{-13}	3.95×10^{-48}	6.47×10^{-16}	5.30×10^{-15}	1.13×10^{-13}	
			5991	MYD015J(ABR) x 2J(wB3)	4.24×10^{-104}	1.08×10^{-109}	2.84×10^{-100}	2.93×10^{-42}	4.50×10^{-23}	8.14×10^{-43}	Detected
IWD038J	Intra-recombinant	wB3 x wB3 x wB2	748	YAD020J(wB3) x GBR7(wB3)	6.49×10^{-31}	2.96×10^{-23}	1.59×10^{-30}	4.49×10^{-14}	3.92×10^{-15}	8.91×10^{-16}	
			6293	GBR36(wB3) x CDN1/Q-Ca(wB2)	9.93×10^{-17}	3.56×10^{-6}	5.04×10^{-17}	1.90×10^{-6}	6.10×10^{-8}	6.49×10^{-11}, Z>4.2	
FKD001J	Intra-recombinant	wB2/wB3 x wB3 x wB3 x wB3 x wB2	695	DNK3(wB2) x Ka1J (wb3)	3.68×10^{-24}	8.58×10^{-16}	1.78×10^{-24}	2.55×10^{-6}	1.96×10^{-6}	9.88×10^{-10}	
			1194	FKD004J(wB3) x HZ6(wB3)	5.00×10^{-12}	1.51×10^{-2}	8.95×10^{-12}	3.14×10^{-6}	1.83×10^{-7}	1.00×10^{-11}	
			2056	HZ6(wB3) x FKD004J(wB3)	1.73×10^{-12}	2.12×10^{-2}	3.02×10^{-12}	8.82×10^{-6}	6.83×10^{-7}	3.54×10^{-7}	
			6293	GBR36(wB3) x CDN1/Q-Ca(wB2)	2.27×10^{-17}	1.53×10^{-9}	9.36×10^{-18}	4.37×10^{-8}	4.74×10^{-9}	1.82×10^{-12}, Z>4.2	
FKD004J	Intra-recombinant	wB2 x wB3 x wB2	643	DNK3(wB2) x Ka1J(wB3)	8.89×10^{-24}	3.07×10^{-11}	4.98×10^{-25}	1.63×10^{-6}	4.15×10^{-7}	9.34×10^{-11}	
			6293	GBR36(wB3) x CDN1/Q-Ca(wB2)	3.23×10^{-15}	3.02×10^{-7}	1.20×10^{-15}	9.15×10^{-6}	4.13×10^{-8}	6.81×10^{-12}, Z>4.2	
FKH122J	Intra-recombinant	wB2 x wB3 x wB2	695	DNK3(wB2) x Ka1J(wB3)	3.87×10^{-37}	4.37×10^{-26}	2.30×10^{-37}	1.38×10^{-7}	1.54×10^{-8}	3.66×10^{-10}	
			6293	GBR36(wB3) x CDN1/Q-Ca(wB2)	2.89×10^{-15}	1.08×10^{-2}	2.57×10^{-15}	7.34×10^{-8}	6.84×10^{-9}	1.96×10^{-11}, Z>4.2	
MYD013J	Intra-recombinant	wB2 x wB3 x wB3 x wB2	643	DNK3(wB2) x Ka1J (wb3)	3.68×10^{-24}	8.58×10^{-16}	1.78×10^{-24}	2.55×10^{-6}	1.96×10^{-6}	9.88×10^{-10}	
			1194	FKD004J(wB3) x HZ6(wB3)	1.51×10^{-15}	2.79×10^{-5}	9.12×10^{-16}	2.11×10^{-8}	3.04×10^{-9}	4.40×10^{-14}, Z>4.35	
			2056	HZ6(wB3) x FKD004J(wB3)	2.89×10^{-16}	6.66×10^{-5}	1.73×10^{-15}	5.81×10^{-8}	6.67×10^{-7}	1.02×10^{-10}, Z>4.35	
MYD015J	Inter-recombinant	wB2 x wB3 x ABR	701	DNK2(wB2) x YAL018J(wB3)	2.12×10^{-11}	2.83×10^{-8}	3.98×10^{-11}	1.69×10^{-8}	3.89×10^{-1}	6.60×10^{-12}, Z>5.89	
			1621	FKD004J(wB3) x HRD(ABR)	1.18×10^{-132}	1.16×10^{-122}	2.80×10^{-131}	2.87×10^{-34}	2.13×10^{-31}	2.40×10^{-41}	Detected

Table 14. Continued.

Isolate	Recombinant type	Recombinant type pattern	Recombination site	Parental sequence (group)	RDP	GENECONV	BootScan	MaxChi	Chimaera	SiScan	PhylPro
					P-value					P and Z-value	
YAL018J	Intra-recombinant	wB3 x wB3 x wB2 x wB2/wB3	748	YAD020J(wB3) x GBR7(wB3)	2.15×10^{-34}	1.69×10^{-23}	1.24×10^{-34}	2.29×10^{-15}	1.34×10^{-16}	2.19×10^{-17}	
			6293	GBR36(wB3) x CDN1/Q-Ca(wB2)	1.19×10^{-20}	1.49×10^{-8}	4.41×10^{-21}	3.86×10^{-8}	6.31×10^{-6}	7.34×10^{-14}, Z>4.2	
			7406-7414	DNK2(wB2) x NLD1(wB2/wB3)	5.28×10^{-9}	7.62×10^{-5}	7.46×10^{-9}	5.53×10^{-5}	4.28×10^{-5}	2.04×10^{-11}, Z<3.0	
YAD020J	Inter-recombinant	wB3 x ABR x wB3	756	USA4(wB3) x CH6(ABR)	2.03×10^{-42}	5.64×10^{-14}	4.38×10^{-42}	1.25×10^{-16}	3.29×10^{-17}	3.01×10^{-15}	
			5122	MYD015J(ABR) x 2J(wB3)	4.02×10^{-99}	5.97×10^{-117}	6.59×10^{-98}	1.63×10^{-44}	3.79×10^{-24}	2.93×10^{-47}	Detected
NID119J	Intra-recombinant	bBR x bBR x bBR x bBR x bBR	157	USA6(bBR) x TANX2(bBR)	1.03×10^{-10}	2.29×10^{-7}	1.59×10^{-10}	1.53×10^{-5}	5.06×10^{-7}	2.18×10^{-10}	
			878	TANX2(bBR) x USA5(bBR)	2.72×10^{-10}	1.97×10^{-7}	5.01×10^{-10}	1.63×10^{-5}	6.49×10^{-7}	2.96×10^{-10}	
			1713	ITA7(bBR) x KWB779J/PV0104(bBR)	1.94×10^{-6}	1.99×10^{-1}	2.45×10^{-7}	4.80×10^{-9}	2.17×10^{-9}	5.92×10^{-18}, Z<3.0	
			6621	CP845J(bBR) x DEU4(bBR)	6.69×10^{-37}	4.66×10^{-17}	4.62×10^{-36}	1.20×10^{-21}	1.19×10^{-17}	2.82×10^{-18}	
NID048J	Inter-recombinant	ABR x wB3 x wB2	242	CH6(ABR) x FKH122J(wB3)	1.77×10^{-26}	9.92×10^{-25}	1.09×10^{-26}	8.11×10^{-6}	4.25×10^{-6}	2.66×10^{-7}	
			643	PV134(wB3) x Ka1J(wB3)	1.23×10^{-14}	1.49×10^{-10}	1.36×10^{-15}	7.36×10^{-3}	3.90×10^{-9}	5.40×10^{-9}	
			6293	GBR36(wB3) x CDN1/Q-Ca(wB2)	2.84×10^{-22}	2.45×10^{-14}	1.12×10^{-22}	1.18×10^{-8}	7.70×10^{-11}	$8,73 \times 10^{-15}$, Z>4.2	
MED302J	Inter-recombinant	ABR x wB3 x ABR x wB3	224	CH6(ABR) x USA4(wB3)	2.22×10^{-45}	2.49×10^{-28}	7.20×10^{-46}	1.02×10^{-11}	1.38×10^{-12}	8.89×10^{-14}	
			719	USA4(wB3) x CH6(ABR)	2.22×10^{-45}	2.49×10^{-28}	7.20×10^{-46}	1.02×10^{-11}	1.38×10^{-12}	8.89×10^{-14}	
			9175	CHL13(ABR) x HZ6(wB3)	8.69×10^{-13}	2.79×10^{-8}	7.32×10^{-14}	ND	9.70×10^{-1}	5.77×10^{-4}	
TRD052J	Inter-recombinant	ABR x wB3 x ABR	224	CH6(ABR) x USA4(wB3)	3.47×10^{-44}	8.58×10^{-25}	1.15×10^{-43}	3.63×10^{-11}	1.72×10^{-11}	1.41×10^{-11}	
			728	USA4(wB3) x CH6(ABR)	3.47×10^{-44}	8.58×10^{-25}	1.15×10^{-43}	3.63×10^{-11}	1.72×10^{-11}	1.41×10^{-11}	
TRD053J	Inter-recombinant	ABR x wB3 x ABR x wB3	236	HRD(ABR) x FKD004J(wB3)	2.63×10^{-36}	3.34×10^{-10}	4.40×10^{-36}	2.07×10^{-11}	1.18×10^{-10}	1.28×10^{-9}	
			775	FKD004J(wB3) x HRD(ABR)	2.63×10^{-36}	3.34×10^{-10}	4.40×10^{-36}	2.07×10^{-11}	1.18×10^{-10}	1.28×10^{-9}	
			9175	CHL13(ABR) x HZ6(wB3)	2.30×10^{-17}	8.10×10^{-16}	1.42×10^{-17}	4.95×10^{-1}	ND	1.44×10^{-4}	
SMD060J	Inter-recombinant	ABR x wB3 x ABR	190	CH6(ABR) x USA4(wB3)	7.06×10^{-47}	2.78×10^{-20}	1.93×10^{-46}	9.25×10^{-13}	4.05×10^{-12}	5.33×10^{-14}	
			758	USA4(wB3) x CH6(ABR)	7.06×10^{-47}	2.78×10^{-20}	1.93×10^{-46}	9.25×10^{-13}	4.05×10^{-12}	5.33×10^{-14}	
YMD069J	Inter-recombinant	ABR x wB3 x ABR x wB3 x ABR x wB3	224	CH6(ABR) x USA4(wB3)	4.22×10^{-38}	3.36×10^{-38}	6.43×10^{-11}	3.17×10^{-12}	3.13×10^{-12}	5.99×10^{-14}	Detected
			728	USA4(wB3) x CH6(ABR)	4.22×10^{-38}	3.36×10^{-38}	6.43×10^{-11}	3.17×10^{-12}	3.13×10^{-12}	5.99×10^{-14}	
			4996	SMD060J(ABR) x FKD001J(wB3)	6.84×10^{-47}	1.18×10^{-35}	4.06×10^{-46}	9.72×10^{-13}	4.39×10^{-13}	2.28×10^{-15}	Detected
			5723	FKD001J(wB3) x SMD060J	6.84×10^{-47}	1.18×10^{-35}	4.06×10^{-46}	9.72×10^{-13}	4.39×10^{-13}	2.28×10^{-15}	Detected
			6097	MYD015J(ABR) x 2J(wB3)	5.99×10^{-89}	2.40×10^{-76}	1.36×10^{-87}	1.58×10^{-36}	4.79×10^{-16}	9.85×10^{-42}	Detected
YMD070J	Inter-recombinant	ABR x wB3 x ABR x wB3	239	CH6(ABR) x USA4(wB3)	2.26×10^{-47}	3.59×10^{-26}	1.01×10^{-43}	2.18×10^{-10}	1.41×10^{-11}	3.50×10^{-13}	
			728	USA4(wB3) x CH6(ABR)	2.26×10^{-47}	3.59×10^{-26}	1.01×10^{-43}	2.18×10^{-10}	1.41×10^{-11}	3.50×10^{-13}	
			9175	CHL13(ABR) x HZ6(wB3)	2.31×10^{-14}	2.20×10^{-11}	1.14×10^{-14}	1.42×10^{-3}	1.96×10^{-1}	1.52×10^{-3}, Z>3.08	
SGB088J	Inter-recombinant	ABR x wB3 x ABR	214	CH6(ABR) x USA4(wB3)	9.77×10^{-44}	1.29×10^{-18}	9.79×10^{-44}	8.64×10^{-13}	1.07×10^{-12}	5.26×10^{-14}	
			756	USA4(wB3) x CH6(ABR)	9.77×10^{-44}	1.29×10^{-18}	9.79×10^{-44}	8.64×10^{-13}	1.07×10^{-12}	5.26×10^{-14}	
KYD073J	Inter-recombinant	ABR x wB3 x ABR	224	CH6(ABR) x USA4(wB3)	6.53×10^{-50}	1.35×10^{-26}	3.00×10^{-46}	1.04×10^{-10}	1.52×10^{-11}	6.80×10^{-13}, Z>6.26	
			748	USA4(wB3) x CH6(ABR)	6.53×10^{-50}	1.35×10^{-26}	3.00×10^{-46}	1.04×10^{-10}	1.52×10^{-11}	6.80×10^{-13}, Z>6.95	
GFD462J	Inter-recombinant	ABR x wB3 x ABR x wB3 x ABR x bBR	224	CH6(ABR) x USA4(wB3)	1.40×10^{-42}	1.47×10^{-26}	1.55×10^{-42}	5.36×10^{-11}	2.89×10^{-12}	1.28×10^{-13}	Detected
			748	USA4(wB3) x CH6(ABR)	1.40×10^{-42}	1.47×10^{-26}	1.55×10^{-42}	5.36×10^{-11}	2.89×10^{-12}	1.28×10^{-13}	
			5509	SMD060J(ABR) x FKD001J(wB3)	2.30×10^{-16}	5.07×10^{-12}	4.65×10^{-15}	5.80×10^{-1}	2.15×10^{-1}	1.84×10^{-4}	Detected
			5722	FKD001J(wB3) x SMD060J(ABR)	2.30×10^{-16}	5.07×10^{-12}	4.65×10^{-15}	5.80×10^{-1}	2.15×10^{-1}	1.84×10^{-4}	Detected
			5974	IWD032J(ABR) x CP845J(bBR)	6.59×10^{-186}	1.68×10^{-172}	4.19×10^{-191}	4.30×10^{-49}	6.30×10^{-31}	9.26×10^{-64}	Detected
NRD350J	Intra-recombinant	wB3 x wB3 x wB2	738	YAD020J(wB3) x GBR7(wB3)	1.66×10^{-26}	5.44×10^{-18}	5.54×10^{-29}	5.65×10^{-13}	7.44×10^{-15}	3.73×10^{-15}	
			6300	GBR36(wB3) x CDN1(wB2)	4.53×10^{-22}	1.91×10^{-10}	2.22×10^{-22}	2.96×10^{-8}	3.37×10^{-9}	2.44×10^{-15}, Z>5.34	
KWB778J	Intra-recombinant	wB3 x wB3 x wB2 x wB2/wB3	738	YAD020J(wB3) x GBR7(wB3)	1.15×10^{-31}	1.92×10^{-18}	2.74×10^{-32}	2.30×10^{-13}	2.74×10^{-13}	6.64×10^{-15}	
			6293	GBR36(wB3) x CDN1/Q-Ca(wB2)	1.21×10^{-19}	2.30×10^{-9}	6.37×10^{-20}	3.09×10^{-8}	4.77×10^{-9}	1.82×10^{-12}, Z>4.2	
			7406-7414	DNK2(wB2) x NLD1(wB2/wB3)	4.80×10^{-8}	6.08×10^{-5}	7.81×10^{-8}	5.02×10^{-4}	2.46×10^{-4}	1.54×10^{-5}, Z<3.0	
KWB779J	Non-recombinant	basal-BR (bBR)									
ND10J	Inter-recombinant	ABR x wB3 x ABR x wB3 x ABR x wB3	190	CH6(ABR) x USA4(wB3)	2.05×10^{-35}	5.89×10^{-16}	1.69×10^{-37}	6.56×10^{-12}	2.43×10^{-13}	7.11×10^{-15}	
			756	USA4(wB3) x CH6(ABR)	2.05×10^{-35}	5.89×10^{-16}	1.69×10^{-37}	6.56×10^{-12}	2.43×10^{-13}	7.11×10^{-15}	
			4632	SMD060J(ABR) x DMJ(wB3)	1.21×10^{-82}	3.76×10^{-61}	5.04×10^{-82}	8.98×10^{-22}	8.13×10^{-24}	1.01×10^{-25}	Detected
			5990	DMJ(wB3) x SMD060J(ABR)	1.21×10^{-82}	3.76×10^{-61}	5.04×10^{-82}	8.98×10^{-22}	8.13×10^{-24}	1.01×10^{-25}	Detected
KGD54J	Inter-recombinant	ABR x wB3 x ABR x wB3 x ABR x wB3	184	CH6(ABR) x USA4(wB3)	1.85×10^{-49}	1.63×10^{-24}	4.18×10^{-43}	8.56×10^{-14}	3.21×10^{-14}	1.92×10^{-15}	
			758	USA4(wB3) x CH6(ABR)	1.85×10^{-49}	1.63×10^{-24}	4.18×10^{-43}	8.56×10^{-14}	3.21×10^{-14}	1.92×10^{-15}	
			4632	SMD060J(ABR) x DMJ(wB3)	3.15×10^{-86}	6.45×10^{-64}	1.66×10^{-86}	8.66×10^{-24}	6.30×10^{-25}	4.40×10^{-27}	Detected
			6001	DMJ(wB3) x SMD060J(ABR)	3.15×10^{-86}	6.45×10^{-64}	1.66×10^{-86}	8.66×10^{-24}	6.30×10^{-25}	4.40×10^{-27}	Detected

Table 14. Continued.

Isolate	Recombinant type	Recombinant type pattern	Recombination site	Parental sequence (group)	RDP	GENECONV	BootScan	MaxChi	Chimaera	SiScan	PhylPro
					P-value					P and Z-value	
CHZJ26A	Intra-recombinant	wB3 x wB3 x wB2	748	YAD020J(wB3) x GBR7(wB3)	2.08×10^{-30}	4.90×10^{-16}	6.20×10^{-30}	1.19×10^{-14}	4.04×10^{-15}	4.55×10^{-14}	
			6293	GBR36(wB3) x CDN1/Q-Ca(wB2)	9.23×10^{-27}	4.36×10^{-18}	3.06×10^{-27}	3.16×10^{-12}	3.94×10^{-12}	6.41×10^{-16}, Z>4.2	
CHL13	Inter-recombinant	ABR x wB3 x ABR x bBR x ABR	250	HRD(ABR) x PV376Br (wB3)	1.20×10^{-31}	3.17×10^{-12}	4.44×10^{-31}	2.49×10^{-15}	2.38×10^{-15}	2.48×10^{-19}, Z>4.51	Detected
			728	USA4(wB3) x CH6(ABR)	5.94×10^{-38}	5.51×10^{-16}	1.53×10^{-38}	8.90×10^{-11}	4.54×10^{-11}	2.52×10^{-12}	
			5672	TRD052J(ABR) x TANX2(bBR)	4.92×10^{-35}	1.45×10^{-33}	3.48×10^{-35}	3.80×10^{-7}	3.13×10^{-7}	9.41×10^{-11}, Z>6.9	
			5950	TANX2 (bBR) x TRD052J(ABR)						Z>6.9	Detected
CH6	Non-recombinant	Asian-BR (ABR)									
CHK16	Inter- recombinant	wB3 x ABR x wB3	745	FKD004J(wB2-wB3) x HRD(ABR)	5.81×10^{-38}	3.8×10^{-8}	2.90×10^{-37}	2.94×10^{-14}	2.25×10^{-13}	7.02×10^{-11}, Z>6.76	
			9123	CHL13(ABR) x HZ6(wB3)	2.01×10^{-20}	7.43×10^{-13}	1.60×10^{-18}	1.90×10^{-2}	8.95×10^{-1}	9.22×10^{-5}, Z>3.77	
HZ6	Intra-recombinant	wB3 x wB3 x wB2	672	YAD020J(wB3) x GBR7(wB3)	9.38×10^{-36}	1.73×10^{-23}	3.37×10^{-36}	1.66×10^{-13}	7.22×10^{-16}	3.68×10^{-16}	
			6293	GBR36(wB3) x CDN1/Q-Ca(wB2)	1.31×10^{-18}	1.30×10^{-7}	1.45×10^{-18}	1.22×10^{-8}	7.56×10^{-9}	1.03×10^{-12}, Z>4.2	
USA 1	Intra-recombinant	wB3 x wB2	6300	GBR36(wB3) x CDN1(wB2)	1.25×10^{-16}	1.14×10^{-5}	4.29×10^{-17}	2.79×10^{-10}	1.33×10^{-7}	1.69×10^{-14}, Z>4.55	
CAR37	Intra-recombinant	wB3 x wB2 x wB2/wB3 x wB2	243	Ka1J(wB3) x DNK3(wB2)	1.96×10^{-23}	3.37×10^{-13}	1.43×10^{-23}	4.49×10^{-7}	7.36×10^{-8}	$1.03x\ 10^{-12}$	
			1037	Ka1J(wB3) x PV134(wB3)	1.10×10^{-19}	5.52×10^{-7}	5.81×10^{-20}	1.68×10^{-9}	3.32×10^{-11}	2.64×10^{-15}	
			6350	GBR36(wB3) x CDN1(wB2)	3.00×10^{-19}	4.34×10^{-9}	4.39×10^{-19}	3.56×10^{-8}	2.69×10^{-6}	1.08×10^{-13}, Z>6.08	
CAR37A	Non-recombinant	world-B1 (wB1)									
CAR39	Intra-recombinant	wB2 x wB2	2641	CAR37A(wB2) x DNK3(wB2)	8.72×10^{-155}	8.88×10^{-149}	1.10×10^{-155}	7.61×10^{-45}	1.53×10^{-31}	8.71×10^{-54}, Z>8.92	
TUR1	Intra-recombinant	wB2 x wB3 x wB2	2505	FRA2(wB2) x KEN1(wB3)	1.15×10^{-57}	1.33×10^{-56}	5.23×10^{-56}	1.68×10^{-31}	3.73×10^{-18}	1.79×10^{-47}, Z>6.33	Detected
			5899	GRC17(wB3) x PV376Br(wB2)	9.62×10^{-62}	3.25×10^{-55}	1.91×10^{-58}	6.95×10^{-33}	2.47×10^{-21}	2.07×10^{-48}, Z>5.42	
TUR9	Non-recombinant	Asian-BR (ABR)									
IRNTRa6	Non-recombinant	basal-B2 (bB2)									
IRNSS5	Non-recombinant	basal-B2 (bB2)									
WFLB06	Inter- recombinant	ABR x bBR x ABR x bBR x ABR	475	NID048J(ABR) x CP845J(bBR)	1.19×10^{-36}	6.76×10^{-13}	2.57×10^{-36}	3.94×10^{-10}	6.49×10^{-9}	3.47×10^{-6}	
			4132	CP845J(bBR) x AKD934J(ABR)	2.20×10^{-52}	2.37×10^{-46}	1.70×10^{-52}	4.04×10^{-12}	9.23×10^{-12}	3.36×10^{-15}	Detected
			4502	HRD(ABR) x CP845J(bBR)	3.33×10^{-50}	9.45×10^{-45}	2.34×10^{-50}	8.55×10^{-13}	5.97×10^{-12}	2.98×10^{-14}	Detected
			7809	CP845J(bBR) x HRD(ABR)	2.09×10^{-73}	8.70×10^{-63}	3.68×10^{-73}	2.75×10^{-23}	3.77×10^{-24}	1.49×10^{-26}	
TANX2	Intra-recombinant	bBR x bBR x bBR	1107	ITA7(bBR) x KWB779J(bBR)	7.05×10^{-38}	9.78×10^{-41}	1.01×10^{-42}	9.21×10^{-31}	9.95×10^{-13}	1.87×10^{-56}	
			5968	KWB779J(bBR) x ITA7(bBR)	3.41×10^{-38}	4.79×10^{-41}	4.90×10^{-43}	7.56×10^{-31}	4.81×10^{-13}	9.06×10^{-57}	
GBR 98	Intra-recombinant	wB3 x wB3 x wB2	731	YAD020J(wB3) x GBR7(wB3)	4.13×10^{-37}	6.63×10^{-22}	1.11×10^{-36}	7.78×10^{-15}	3.51×10^{-15}	1.28×10^{-15}	
			6767	GBR51(wB3) x GBR27(wB3)	1.66×10^{-13}	ND	1.00×10^{-13}	3.66×10^{-5}	2.38×10^{-5}	1.91×10^{-4}, Z>3.72	Detected
AllA	Inter- recombinant	wB3 x bB1	485	TUR1(wB3) x TIGA(bB1)	3.12×10^{-16}	1.47×10^{-6}	1.01×10^{-15}	4.22×10^{-1}	8.38×10^{-4}	1.30×10^{-13}, Z>5.17	Detected
ASP	Non-recombinant	basal-B0 (bB0)									
BEL 1	Non-recombinant	world-B0 (wB0)									
DEU 1	Non-recombinant	world-B2 (wB2)									
DEU 2	Non-recombinant	world-B2 (wB2)									
DEU 4	Intra-recombinant	bBR x bBR		KWB779J(bBR) x ITA7(bBR)	2.39×10^{-18}	1.13×10^{-16}	2.13×10^{-17}	1.20×10^{-22}	4.26×10^{-19}	4.22×10^{-6}, Z>5.4	
DEU 5	Intra-recombinant	wB2 x wB2	7675	DNK3(wB2) x CDN1(wB2)	3.75×10^{-14}	2.81×10^{-2}	3.52×10^{-13}	2.21×10^{-7}	6.47×10^{-8}	9.62×10^{-11}	
DEU 7	Inter- recombinant	wB3 x bB1	612	TUR1(wB3) x TIGA(bB1)	3.84×10^{-21}	7.06×10^{-12}	3.53×10^{-19}	6.55×10^{-6}	4.63×10^{-8}	2.09×10^{-17}, Z>7.04	Detected
DNK 3	Non-recombinant	world-B2 (wB2)									
DNK 4	Intra-recombinant	wB3 x wB3 x wB2	744	YAD020J(wB3) x GBR7(wB3)	1.17×10^{-46}	3.86×10^{-41}	5.56×10^{-47}	3.39×10^{-15}	1.38×10^{-16}	5.62×10^{-21}	
			6300	GBR36(wB3) x CDN1(wB2)	1.12×10^{-24}	2.60×10^{-12}	1.42×10^{-24}	4.63×10^{-10}	4.61×10^{-11}	2.60×10^{-15}, Z>5.47	
Eru1D	Intra-recombinant	bB1 x bB1	5533	A64(bB1) x Al(bB1)	2.74×10^{-9}	1.11×10^{-2}	1.06×10^{-13}	4.23×10^{-10}	4.26×10^{-12}	2.29×10^{-12}, Z>4.84	
ESP 1	Inter- recombinant	wB1 x bB1	698	PRT1(wB1) x A64(bB1)	3.89×10^{-37}	1.24×10^{-41}	8.28×10^{-40}	2.18×10^{-11}	3.13×10^{-10}	8.33×10^{-23}	
ESP 2	Inter- recombinant	wB1 x bB1	686	PRT1(wB1) x A64(bB1)	3.77×10^{-30}	1.03×10^{-36}	1.44×10^{-32}	1.30×10^{-9}	7.70×10^{-10}	4.40×10^{-24}, Z>6.1	
FRA 2	Non-recombinant	world-B2 (wB2)									
GBR 7	Intra-recombinant	wB3 x wB2	6300	GBR36(wB3) x CDN1(wB2)	1.45×10^{-21}	6.92×10^{-10}	4.72×10^{-22}	9.66×10^{-10}	1.46×10^{-10}	2.44×10^{-15}, Z>5.3	
GBR 8	Intra-recombinant	wB2 x wB3	6628	POL2(wB2) x KEN1(wB3)	2.95×10^{-74}	2.98×10^{-49}	4.67×10^{-74}	1.91×10^{-22}	6.88×10^{-23}	1.44×10^{-23}	
GBR 27	Intra-recombinant	wB3 x wB3 x wB	738	YAD020J(wB3) x GBR7(wB3)	1.50×10^{-37}	1.32×10^{-24}	6.69×10^{-38}	4.67×10^{-16}	1.06×10^{-16}	1.10×10^{-15}	
			6200	GBR36(wB3) x CDN1(wB2)	1.02×10^{-23}	7.32×10^{-14}	4.89×10^{-24}	1.30×10^{-11}	2.05×10^{-11}	1.58×10^{-15}, Z>5.99	

Table 14. Continued.

Isolate	Recombinant type	Recombinant type pattern	Recombination site	Parental sequence (group)	RDP	GENECONV	BootScan	MaxChi	Chimaera	SiScan	PhylPro
					P-value					P and Z-value	
GBR 30	Intra-recombinant	wB3 x wB3 x wB2	738	YAD020J(wB3) x GBR7(wB3)	6.85×10^{-41}	3.87×10^{-30}	6.91×10^{-41}	9.63×10^{-15}	2.56×10^{-16}	6.21×10^{-16}	
			6300	GBR36(wB3) x CDN1(wB2)	4.14×10^{-22}	ND	2.42×10^{-22}	1.09×10^{-9}	9.64×10^{-11}	4.47×10^{-13}, Z>6.01	
GBR 31	Intra-recombinant	wB3 x wB3 x wB2	748	YAD020J(wB3) x GBR7(wB3)	4.04×10^{-40}	2.85×10^{-28}	1.32×10^{-40}	2.18×10^{-15}	2.92×10^{-16}	1.16×10^{-16}	
			6350	GBR36(wB3) x CDN1(wB2)	5.30×10^{-23}	1.67×10^{-14}	3.01×10^{-23}	3.68×10^{-8}	1.02×10^{-8}	8.38×10^{-13}, Z>6.08	
GBR 32	Intra-recombinant	wB3 x wB3 x wB2	738	YAD020J(wB3) x GBR7(wB3)	6.85×10^{-41}	3.87×10^{-30}	6.91×10^{-41}	9.63×10^{-15}	2.56×10^{-16}	6.21×10^{-19}	
			6300	GBR36(wB3) x CDN1(wB2)	1.40×10^{-22}	8.94×10^{-14}	8.00×10^{-23}	1.08×10^{-9}	9.64×10^{-11}	4.47×10^{-13}, Z>6.01	
GBR 38	Intra-recombinant	wB3 x wB3 x wB2	702	YAD020J(wB3) x GBR7(wB3)	5.35×10^{-37}	2.62×10^{-19}	1.59×10^{-36}	1.21×10^{-13}	9.08×10^{-15}	5.00×10^{-15}	
			6800	GBR36(wB3) x CDN1(wB2)						Z>4.66	Detected
GBR 51	Intra-recombinant	wB3 x wB3	682	YAD020J(wB3) x GBR7(wB3)	2.47×10^{-38}	1.55×10^{-18}	2.99×10^{-38}	4.95×10^{-14}	1.75×10^{-14}	2.87×10^{-14}	
GBR 57	Intra-recombinant	wB0 x wB3	8817	GBR83(wB0) x NLD2(wB3)	7.79×10^{-33}	7.01×10^{-29}	5.23×10^{-34}	2.41×10^{-8}	8.41×10^{-5}	1.16×10^{-13}	
GBR 83	Non-recombinant	world-B0 (wB0)									
GBR 91	Intra-recombinant	wB3 x wB3 x wB2	748	YAD020J(wB3) x GBR7(wB3)	3.01×10^{-40}	1.70×10^{-28}	4.16×10^{-41}	1.59×10^{-14}	5.40×10^{-16}	2.55×10^{-16}	
			6300	GBR36(wB3) x CDN1(wB2)	1.16×10^{-21}	2.26×10^{-14}	6.20×10^{-22}	3.03×10^{-10}	4.11×10^{-11}	1.69×10^{-14}, Z>5.99	
GK1	Non-recombinant	basal-B2 (bB2)									
HUN 1	Non-recombinant	world-B2 (wB2)									
ITA1 A	Non-recombinant	basal-B1 (bB1)									
ITA 2	Non-recombinant	basal-B1 (bB1)									
ITA 4	Inter- recombinant	bBR x bB2	657	ITA8(bBR) x IS1(bB2)	4.91×10^{-25}	4.15×10^{-21}	4.06×10^{-24}	2.95×10^{-6}	5.36×10^{-6}	2.57×10^{-17}	
ITA 5	Inter-recombinant	bBR x bB2	644	ITA8(bBR) x IS1(bB2)	2.46×10^{-23}	1.09×10^{-14}	1.29×10^{-22}	5.25×10^{-6}	3.78×10^{-7}	5.88×10^{-16}	
ITA 6	Inter- recombinant	bBR x bB2	654	ITA8(bBR) x IS1(bB2)	2.76×10^{-26}	6.94×10^{-22}	5.32×10^{-26}	1.57×10^{-6}	5.37×10^{-7}	3.18×10^{-18}	
ITA 8	Non-recombinant										
ITA 9A	Intra-recombinant	bB1 x bB1	1106	ITA1A(bB1) x A64(bB1)	1.53×10^{-63}	3.55×10^{-62}	6.85×10^{-62}	1.54×10^{-24}	4.40×10^{-16}	1.49×10^{-36}, Z>6.22	
			2468	GRC42(bB1) x A64(bB1)	2.51×10^{-8}	1.48×10^{-7}	5.42×10^{-10}	1.41×10^{-10}	1.64×10^{-5}	5.47×10^{-32}	
NLD 2	Intra-recombinant	wB3 x wB3 x wB2	738	YAD020J(wB3) x GBR7(wB3)	5.12×10^{-33}	7.49×10^{-20}	3.40×10^{-33}	2.86×10^{-14}	1.75×10^{-14}	1.66×10^{-14}	
			6300	GBR36(wB3) x CDN1(wB2)	4.42×10^{-22}	1.16×10^{-9}	2.29×10^{-22}	4.59×10^{-9}	8.16×10^{-10}	3.23×10^{-14}, Z>5.4	
OM-N	Non-recombinant	Orchis									
OM-A	Non-recombinant	Orchis									
ORM	Non-recombinant	Orchis									
OS	Non-recombinant	Orchis									
POL 1	Non-recombinant	world-B2 (wB2)									
PRT 1	Non-recombinant	world-B0 (wB2)									
PV0054	Intra-recombinant	wB2 x wB2/wB3(?) x wB2	711	DNK3(wB2) x NLD1(wB2/wB3)	4.51×10^{-16}	4.78×10^{-6}	1.02×10^{-16}	4.40×10^{-5}	5.40×10^{-7}	1.18×10^{-6}, Z>5.5	
			6550	GBR36(wB3) x CDN1(wB2)	2.58×10^{-10}	4.85×10^{-7}	1.23×10^{-10}	5.59×10^{-5}	1.35×10^{-4}	6.48×10^{-11}, Z>5.17	
POL 2	Non-recombinant	world-B2 (wB2)									
POL 4	Intra-recombinant	wB2 x wB2	6484	NZ290(wB2) x DEU1(wB2)	2.02×10^{-29}	4.78×10^{-26}	1.28×10^{-28}	6.02×10^{-18}	5.23×10^{-13}	2.75×10^{-20}	
PV177	Intra-recombinant	wB3 x wB3 x wB2	730	YAD020J(wB3) x GBR7(wB3)	2.88×10^{-32}	1.62×10^{-16}	2.28×10^{-32}	5.99×10^{-14}	5.04×10^{-14}	9.50×10^{-14}	
			6300	GBR36(wB3) x CDN1(wB2)	7.87×10^{-24}	2.90×10^{-9}	4.22×10^{-24}	6.60×10^{-10}	3.61×10^{-11}	9.44×10^{-16}, Z>5.61	
TIGA	Non-recombinant	basal-B1 (bB1)									
TIGD	Non-recombinant	basal-B1 (bB1)									
UT	Intra-recombinant	wB2 x wB2/wB3	7694	DNK3(wB2) x GBR7(wB2)	2.58×10^{-17}	2.12×10^{-3}	4.15×10^{-17}	7.66×10^{-8}	2.43×10^{-9}	4.36×10^{-10}	
PV134	Inter- recombinant	wB3 x bB1/bB2 x wB3 x wB2	1019	KEN1(wB3) x AllA(bB1)	9.93×10^{-47}	2.51×10^{-2}	5.19×10^{-42}	5.37×10^{-24}	6.92×10^{-18}	1.32×10^{-24}, Z>6.11	
			3340	AllA(bB1) x GBR27(wB3)	5.53×10^{-48}	5.37×10^{-1}	2.16×10^{-43}	3.79×10^{-23}	4.41×10^{-18}	1.13×10^{-19}, Z>6.11	Detected
			6144	GBR51(wB3) x DNK3(wB2)	6.97×10^{-10}	ND	1.09×10^{-9}	3.44×10^{-6}	4.18×10^{-1}	8.04×10^{-7}, Z>6.17	
PV389	Intra-recombinant	wB3 x wB3 x wB2	730	YAD020J(wB3) x GBR7(wB3)	1.60×10^{-32}	6.09×10^{-18}	1.28×10^{-32}	1.01×10^{-13}	9.28×10^{-14}	8.64×10^{-14}	
			6300	GBR36(wB3) x CDN1(wB2)	3.83×10^{-23}	3.50×10^{-9}	2.05×10^{-23}	2.43×10^{-9}	2.94×10^{-10}	1.88×10^{-15}, Z>5.61	
USA 4	Intra-recombinant	wB3 x wB2	1745	GBR91(wB3) x HUN1(wB2)	2.63×10^{-46}	1.31×10^{-10}	1.70×10^{-47}	1.16×10^{-19}	5.83×10^{-19}	4.58×10^{-20}	
USA 5	Intra-recombinant	bBR x bBR x bBR	3973	CP845J(bBR) x Cal1(bBR)	7.75×10^{-14}	1.20×10^{-3}	9.12×10^{-13}	2.42×10^{-12}	7.30×10^{-14}	2.18×10^{-17}, Z>4.46	
			5767	Cal1(bBR) x CP845J(bBR)	7.75×10^{-14}	1.20×10^{-3}	9.12×10^{-13}	2.42×10^{-12}	7.30×10^{-14}	2.18×10^{-17}, Z>4.46	
USA 6	Intra-recombinant	bBR x bBR x bBR	3973	CP845J(bBR) x Cal1(bBR)	8.81×10^{-15}	7.77×10^{-6}	3.67×10^{-14}	2.97×10^{-13}	8.97×10^{-10}	7.15×10^{-19}, Z>4.46	
			5774	Cal1(bBR) x CP845J(bBR)	8.81×10^{-15}	7.77×10^{-6}	3.67×10^{-14}	2.97×10^{-13}	8.97×10^{-10}	7.15×10^{-19}, Z>4.46	

Major groups: bB; basal-B, bBR; basal-BR, wB; world-B, ABR; Asian-BR. Number followed by major groups indicates the subgroup (sublineage) in the major groups [2].

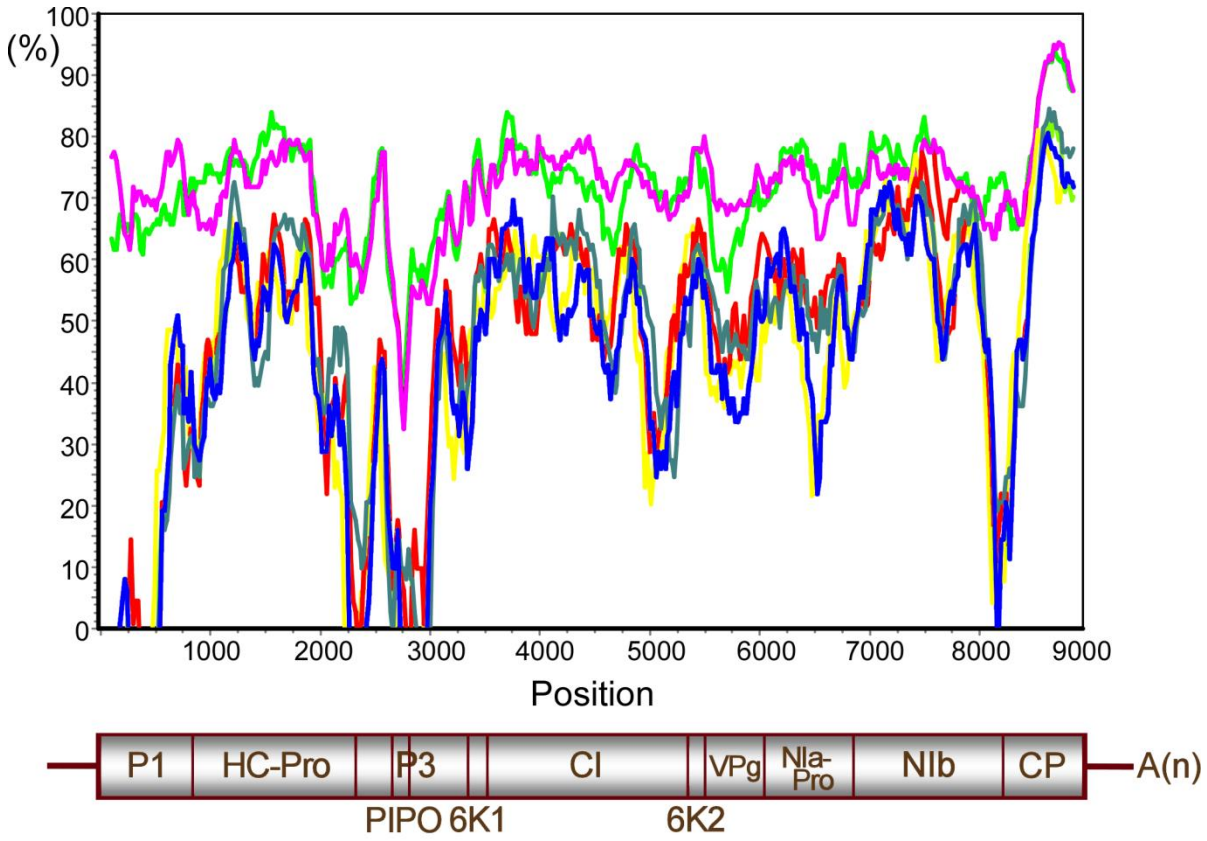


Figure 19. Similarity plot with OM-N genome sequence as the query isolate. Graph of the similarities between the genome sequence of OM and those of Al (red) and UK1 (blue) isolates, and *Japanese yam mosaic virus* (JYMV) (light green), *Scallion mosaic virus* (ScMV) (pink), and *Narcissus yellow stripe virus* (NYSV) (dark green). The similarities were estimated using SIMPLOT 3.5.1 with a window size of 200 nt.

such as the lengths of the genomes and each gene, especially those of the P1 and CP genes (Table 15). In addition, the protein cleavage sites of the Orchis and BI isolates are more similar to each other than to those of outgroups (Table 16). Furthermore, comparions of nucleotide identities for most of protein encoding regions and polyprotein between Orchis and BI isolates ranged from 76.3% to 82.8 %, excluded P3 (67.8% - 70.0%) and 6K2 (66.7% - 72.3%) (Table 17). Thus, this study concludes that Orchis isolates are close to, although biologically distinct from, the BI isolates.

Several of the BI isolates were isolated from non-brassica hosts other than orchids, but like other BI isolates, they infected most brassicas. Most had been collected in Europe and most were from the basal-B lineage. However, a Monte Carlo 'provenance randomization' test (Gibbs *et al*., 2008c; Simmons *et al*., 2008) showed that they did not significantly cluster in maximum-likelihood trees.

2.4. Evolutionary rates and timescales

This study used a Bayesian phylogenetic approach to estimate TMRCAs and nucleotide substitution rates of the individual genes of the BI and Orchis isolates. For all four data sets, a relaxed clock model was found to fit the data better than the strict clock model (Table 18). Notably, for all four data sets, similar posterior means were obtained with all demographic models and for both uncorrelated lognormal and uncorrelated exponential clock models. The best-supported demographic model for the HC-Pro, P3, and NIb TuMV sequences was that of constant population size, whereas for the cCP region (Gibbs *et al*., 2008c, 2010; Ward *et al*., 1995) a model of exponential growth received the strongest support (Table 18 and 19).

Table 15. Comparisons of the lengths of the genes of *Japanese yam mosaic virus* (JYMV), *Narcissus yellow stripe virus* (NYSV), *Scallion mosaic virus* (ScMV) and *Turnip mosaic virus* (TuMV).

	JYMV				TuMV		
Region/protein[a]	j1	mild	NYSV	ScMV	OM-N	Al	UK1
5' NCR	156[b]	153	127	109	130[c]	130	130
P1	975	981	951	633	1086	1086	1086
HC-Pro	1374	1374	1374	1371	1374	1374	1374
P3	1065	1065	1062	1062	1065	1065	1065
6K1	156	156	156	156	156	156	156
CI	1932	1932	1932	1932	1932	1932	1932
6K2	159	159	159	159	159	159	159
VPg	576	576	573	576	576	576	576
NIa-Pro	729	729	729	729	729	729	729
NIb	1554	1554	1551	1551	1551	1551	1551
CP	870	870	822	834	858	864	864
3' NCR	211	211	214	212	211	211	212

[a] NCR; Non-coding region, P1; Protein 1, HC-Pro; Helper component-proteinase protein, P3; Protein 3, 6K1; 6Kda 1 protein, CI; Cylindrical inclusion protein, 6K2; 6Kda 2 protein, VPg; Genome-linked viral protein; NIa-Pro; Nuclear inclusion a-proteinase protein, NIb; Nuclear inclusion b protein, CP; Coat protein.

[b] Nucleotide numbers in blue shows different from those from TuMV.

[c] Thirty-five nucleotide sequence used for primer is included.

Table 16. Comparisons of the amino acids at the polyprotein cleavage sites of *Japanese yam mosaic virus* (JYMV), *Narcissus yellow stripe virus* (NYSV), *Scallion mosaic virus* (ScMV) and *Turnip mosaic virus* (TuMV).

	JYMV				TuMV		
Proteins and stop codon	j1	mild	NYSV	ScMV	OM-N	Al	UK1
P1 / HC-Pro[a]	ITHY/S	IVRF/A	MKHY/S	IRQF/S	IIHF/S[b]	IVHF/S	IVHF/S
HC-Pro / P3	YLIG/G	YLVG/G	YLVG/G	YAVG/G	YRVG/G	YRVG/G	YRVG/G
P3 / 6K1	Q/A	Q/A	Q/T	Q/S	Q/A	Q/A	Q/A
6K1 / CI	Q/A	Q/G	Q/S	Q/A	Q/T	Q/A	Q/T
CI / 6K2	Q/S	Q/S	Q/S	Q/T	Q/S	Q/S	Q/N
6K2 / VPg	E/A	E/A	E/A	E/A	E/A	E/A	E/A
VPg / NIa-Pro	E/S	E/S	E/S	E/S	E/S	E/S	E/S
NIa-Pro / NIb	Q/M	Q/M	Q/M	Q/M	Q/T	Q/T	Q/T
NIb / CP	Q/S	Q/S	Q/S	Q/A	Q/A	Q/A	Q/A
Stop codon	TAA	TAA	TAA	TGA	TGA	TGA	TGA

[a] P1; Protein 1, HC-Pro; Helper component-proteinase protein, P3; Protein 3, 6K1; 6Kda 1 protein, CI; Cylindrical inclusion protein, 6K2; 6Kda 2 protein, VPg; Genome-linked viral protein; NIa-Pro; Nuclear inclusion a-proteinase protein, NIb; Nuclear inclusion b protein, CP; Coat protein.
[b] The amino acids between the cleavage sites were same to other TuMV isolates.

Table 17. Comparisons of the nucleotide identities (%) between Orchis group isolates and those of *Japanese yam mosaic virus* (JYMV), *Narcissus yellow stripe virus* (NYSV), *Scallion mosaic virus* (ScMV), and *Turnip mosaic virus* (TuMV).

Protein encoding region[a]	Orchis group isolates vs							
	JYMV-j1	JYMV-m	NYSV	ScMV	BI isolates			
					basal-B	basal-BR	Asian-BR	world-B
Polyprotein	62.8	62.7	63.2	62.7	76.5-77.1	76.4-76.5	76.8-77.0	76.8-77.7
P1	49.7	48.1	46.8	36.7	72.0-74.8	78.2-79.0	78.1-76.9	77.4-80.9
HC-Pro	64.2	65.0	65.0	65.2	79.3-79.6	76.5-77.3	76.5-76.6	76.8-77.6
P3	54.8	54.3	55.9	56.1	67.8-68.4	69.6-70.0	69.1-70.3	68.3-69.4
6K1	65.4	61.5	64.7	64.7	74.4-77.6	72.4-72.8	73.1-75.0	71.8-76.3
CI	65.8	66.0	66.7	67.1	77.5-78.8	78.2-78.7	77.5-78.2	78.6-78.7
6K2	57.4	60.1	53.2	58.6	67.9-72.3	69.2-69.8	71.7-72.3	66.7-71.7
VPg	63.6	64.8	64.9	66.6	73.5-74.1	75.0-77.8	75.2-77.1	76.9-77.8
NIa-Pro	64.2	62.6	65.9	67.3	77.9-78.1	75.5-76.3	76.5-76.6	76.3-77.1
NIb	68.8	69.8	69.5	69.8	78.2-78.7	77.1-77.7	76.5-76.8	76.9-77.5
CP	69.6	68.3	69.2	70.0	81.6-82.8	81.7-82.1	80.4-81.8	82.3-82.5

[a] P1; Protein 1, HC-Pro; Helper component-proteinase protein, P3; Protein 3, 6K1; 6Kda 1 protein, CI; Cylindrical inclusion protein, 6K2; 6Kda 2 protein, VPg; Genome-linked viral protein; NIa-Pro; Nuclear inclusion a-proteinase protein, NIb; Nuclear inclusion b protein, CP; Coat protein.
Viruses/isolates were listed in Figure 18.

Table 18. Detailed results of the Bayesian coalescent analysis.

Model		Marginal likelihood	Bayes factor	TMRCA (95% HPD lower – upper)	Substitution rate (subs/site/year)	95% HPD rate – lower (subs/site/year)	95% HPD rate – upper (subs/site/year)	Population size	95% HPD, population size (lower, upper)	Population growth rate	95% HDP, growth rate (lower, upper)
Helper component-proteinase protein (HC-Pro)											
Strict clock	Constant Size	-19873.560	1.542	3428 (987 – 6788)	1.76×10^{-4}	3.65×10^{-5}	3.25×10^{-4}	5.60×10^{3}	1.32×10^{3}, 1.11×10^{4}	N/A[a]	N/A
	Expansion Growth	-19872.830	3.186	3512 (1137 – 7027)	1.55×10^{-4}	4.07×10^{-5}	2.80×10^{-4}	4.61×10^{4}	2.03×10^{3}, 1.24×10^{5}	5.81×10^{-2}	7.13×10^{-3}, 1.25×10^{-1}
	Exponential Growth	-19873.990		2555 (1067 – 4734)	1.86×10^{-4}	6.53×10^{-5}	3.21×10^{-4}	5.02×10^{3}	1.88×10^{3}, 9.77×10^{3}	9.13×10^{-4}	1.41×10^{-4}, 1.76×10^{-3}
	Bayesian Skyline	-19873.090	2.477	5203 (1276 – 13860)	1.18×10^{-4}	1.14×10^{-5}	2.22×10^{-4}	2.21×10^{4}	3.77×10^{3}, 9.58×10^{4}	N/A	N/A
Relaxed Exponential	Constant Size	-19775.180	8E+42	819 (258 - 1643)	1.07×10^{-3}	5.82×10^{-4}	1.56×10^{-3}	6.76×10^{2}	3.33×10^{2}, 1.12×10^{3}	N/A	N/A
	Expansion Growth	-19775.340	7E+42	703 (254 - 1401)	1.07×10^{-3}	5.93×10^{-4}	1.58×10^{-3}	1.26×10^{4}	3.26×10^{2}, 2.85×10^{4}	5.13×10^{-1}	3.63×10^{-5}, 0.13×10^{-1}
	Exponential Growth[b]	-19775.120	9E+42	495 (174 - 924)	1.13×10^{-3}	6.71×10^{-4}	1.65×10^{-3}	7.76×10^{2}	3.70×10^{2}, 1.25×10^{3}	4.44×10^{-3}	9.00×10^{-4}, 1.12×10^{-2}
	Bayesian Skyline	-19776.410	2E+42	1203 (224 - 3166)	7.68×10^{-4}	1.97×10^{-4}	1.33×10^{-3}	5.51×10^{3}	5.87×10^{2}, 2.75×10^{4}	N/A	N/A
Relaxed Lognormal	Constant Size	-19778.470	3E+41	1232 (513 - 2248)	4.24×10^{-4}	1.74×10^{-4}	6.90×10^{-4}	1.77×10^{3}	7.09×10^{2}, 3.21×10^{3}	N/A	N/A
	Expansion Growth	-19776.520	2E+42	1173 (528 - 3438)	3.12×10^{-4}	8.36×10^{-5}	5.45×10^{-4}	3.98×10^{4}	9.14×10^{2}, 1.34×10^{5}	1.09×10^{-1}	1.73×10^{-3}, 2.31×10^{-1}
	Exponential Growth	-19779.190	1E+41	1063 (479 – 1891)	4.38×10^{-4}	2.02×10^{-4}	6.87×10^{-4}	2.01×10^{3}	8.61×10^{2}, 3.66×10^{3}	1.99×10^{-3}	1.30×10^{-4}, 4.03×10^{-3}
	Bayesian Skyline	-19779.210	1E+41	3708 (752 - 10744)	1.96×10^{-4}	1.66×10^{-5}	3.79×10^{-4}	1.72×10^{4}	2.36×10^{3}, 8.07×10^{4}	N/A	N/A
Protein 3 (P3)											
Strict clock	Constant Size	-18067.697	2.155	5132 (1614 – 10489)	2.20×10^{-4}	5.48×10^{-5}	3.73×10^{-4}	4.83×10^{3}	1.41×10^{3}, 1.01×10^{4}	N/A	N/A
	Expansion Growth	-18066.530	6.920	6299 (1878 – 13906)	1.66×10^{-4}	3.57×10^{-5}	3.06×10^{-4}	7.65×10^{4}	2.74×10^{3}, 2.37×10^{5}	6.55×10^{-2}	7.11×10^{-3}, 1.39×10^{-1}
	Exponential Growth	-18068.465		3796 (1615 – 6674)	2.32×10^{-4}	9.25×10^{-5}	3.82×10^{-4}	4.11×10^{3}	1.58×10^{3}, 7.48×10^{3}	5.96×10^{-4}	4.36×10^{-6}, 1.23×10^{-3}
	Bayesian Skyline	-18066.570	6.653	10248 (1933 – 31252)	1.34×10^{-4}	5.51×10^{-6}	2.71×10^{-4}	3.53×10^{4}	4.41×10^{3}, 1.74×10^{5}	N/A	N/A
Relaxed Exponential	Constant Size	-18020.298	8E+20	1071 (279 – 2511)	1.08×10^{-3}	5.59×10^{-4}	1.60×10^{-3}	8.38×10^{2}	3.83×10^{2}, 1.35×10^{3}	N/A	N/A
	Expansion Growth	-18021.032	4E+20	1013 (230 – 2248)	1.07×10^{-3}	3.01×10^{-4}	1.70×10^{-3}	1.88×10^{4}	3.39×10^{2}, 4.15×10^{4}	4.31×10^{-1}	1.25×10^{-5}, 0.12×10^{1}
	Exponential Growth	-18020.515	7E+20	612 (205– 1239)	1.08×10^{-3}	4.94×10^{-4}	1.68×10^{-3}	1.05×10^{3}	4.37×10^{2}, 1.91×10^{3}	4.01×10^{-3}	8.00×10^{-4}, 1.01×10^{-2}
	Bayesian Skyline	-18021.024	4E+20	1985 (243 – 5847)	6.82×10^{-4}	3.88×10^{-5}	1.31×10^{-3}	9.83×10^{3}	6.27×10^{2}, 6.55×10^{4}	N/A	N/A
Relaxed Lognormal	Constant Size	-18026.949	1E+18	3194 (1158 –5950)	3.06×10^{-4}	1.24×10^{-4}	5.00×10^{-4}	2.81×10^{3}	1.15×10^{3}, 5.22×10^{3}	N/A	N/A
	Expansion Growth	-18026.731	1E+18	4639 (1297 – 10290)	2.25×10^{-4}	5.32×10^{-5}	3.99×10^{-4}	5.96×10^{4}	1.35×10^{3}, 1.87×10^{5}	8.05×10^{-2}	8.83×10^{-3}, 1.65×10^{-1}
	Exponential Growth	-18026.664	1E+18	2434 (962 – 4499)	3.26×10^{-4}	1.40×10^{-4}	5.14×10^{-4}	2.87×10^{3}	1.24×10^{3}, 5.35×10^{3}	9.04×10^{-4}	1.0×10^{-1}, 2.09×10^{-3}
	Bayesian Skyline	-18026.312	2E+18	8552 (1179 – 27913)	1.74×10^{-4}	7.78×10^{-6}	3.62×10^{-4}	3.00×10^{4}	3.26×10^{3}, 1.60×10^{5}	N/A	N/A
Nuclear inclusion b protein (NIb)											
Strict clock	Constant Size	-21182.111		2257 (1387 – 3306)	2.51×10^{-4}	1.55×10^{-4}	3.47×10^{-4}	2.51×10^{3}	1.49×10^{3}, 3.77×10^{3}	N/A	N/A
	Expansion Growth	-21179.868	9.424	2846 (1502 – 4623)	2.05×10^{-4}	1.05×10^{-4}	3.05×10^{-4}	5.72×10^{4}	5.33×10^{3}, 1.50×10^{5}	7.92×10^{-2}	2.85×10^{-2}, 1.37×10^{-1}
	Exponential Growth	-21181.779	1.394	2074 (1275 – 3073)	2.56×10^{-4}	1.60×10^{-4}	3.62×10^{-4}	2.83×10^{3}	1.59×10^{3}, 4.31×10^{3}	1.28×10^{-3}	3.97×10^{-4}, 2.24×10^{-3}
	Bayesian Skyline	-21179.648	11.734	4068 (1581 – 7632)	1.63×10^{-4}	5.27×10^{-5}	2.66×10^{-4}	1.52×10^{4}	3.61×10^{3}, 4.94×10^{4}	N/A	N/A
Relaxed Exponential	Constant Size	-21157.000	8E+10	1330 (342 – 2920)	7.04×10^{-4}	3.73×10^{-4}	1.00×10^{-3}	1.04×10^{3}	5.39×10^{2}, 1.71×10^{3}	N/A	N/A
	Expansion Growth	-21157.376	6E+10	1157 (273 – 2579)	6.46×10^{-4}	2.76×10^{-4}	1.07×10^{-3}	6.58×10^{4}	4.37×10^{2}, 2.13×10^{5}	1.46×10^{-1}	3.63×10^{-3}, 3.27×10^{-1}
	Exponential Growth	-21157.492	5E+10	475 (184 – 875)	7.60×10^{-4}	4.06×10^{-4}	1.13×10^{-3}	1.36×10^{3}	6.56×10^{2}, 2.23×10^{3}	7.01×10^{-3}	6.27×10^{-4}, 1.53×10^{-2}
	Bayesian Skyline	-21158.010	3E+10	1058 (240 – 2431)	6.20×10^{-4}	2.28×10^{-4}	9.82×10^{-4}	8.53×10^{3}	9.77×10^{2}, 4.35×10^{4}	N/A	N/A
Relaxed Lognormal	Constant Size	-21162.639	3E+08	2195 (1161 – 3494)	2.71×10^{-4}	1.52×10^{-4}	3.87×10^{-4}	2.36×10^{3}	1.31×10^{3}, 3.76×10^{3}	N/A	N/A
	Expansion Growth	-21164.348	5E+07	2794 (1209 – 4855)	2.18×10^{-4}	1.01×10^{-4}	3.42×10^{-4}	5.68×10^{4}	3.65×10^{3}, 1.56×10^{5}	8.30×10^{-2}	2.49×10^{-2}, 1.48×10^{-1}
	Exponential Growth	-21162.878	2E+08	1833 (906 – 2936)	2.76×10^{-4}	1.51×10^{-4}	4.05×10^{-4}	2.72×10^{3}	1.41×10^{3}, 4.49×10^{3}	1.54×10^{-3}	4.11×10^{-4}, 2.96×10^{-3}
	Bayesian Skyline	-21165.873	1E+07	4084 (1440 – 8034)	1.67×10^{-4}	5.14×10^{-5}	2.80×10^{-4}	9.65×10^{3}	1.93×10^{3}, 3.42×10^{4}	N/A	N/A

Table 18. Detailed results of the Bayesian coalescent analysis.

Model		Marginal likelihood	Bayes factor	TMRCA (95% HPD lower – upper)	Substitution rate (subs/site/year)	95% HPD rate – lower (subs/site/year)	95% HPD Rate – upper (subs/site/year)	Population Size	95% HPD, Population Size (lower, upper)	Population Growth Rate	95% HDP, Growth Rate (lower, upper)
Coherently-evolving coat protein region (cCP)											
Strict clock	Constant Size	-6988.608	1.792	1069 (569–1704)	2.57×10^{-4}	1.36×10^{-4}	3.73×10^{-4}	1.18×10^{3}	6.17×10^{2}, 1.95×10^{3}	N/A	N/A
	Expansion Growth	-7021.757	3.671	1723 (539–3188)	2.02×10^{-4}	5.77×10^{-5}	3.45×10^{-4}	3.25×10^{4}	6.04×10^{2}, 8.93×10^{4}	8.23×10^{-2}	2.69×10^{-4}, 0.22×10^{-1}
	Exponential Growth	-6987.824		932 (507–1482)	2.55×10^{-4}	1.35×10^{-4}	3.79×10^{-4}	1.48×10^{3}	7.36×10^{2}, 2.49×10^{3}	3.18×10^{-3}	8.81×10^{-4}, 5.75×10^{-3}
	Bayesian Skyline	-7189.656	4.401	2765 (561–7333)	1.66×10^{-4}	2.66×10^{-6}	2.83×10^{-4}	7.37×10^{3}	8.53×10^{2}, 4.39×10^{4}	N/A	N/A
Relaxed Exponential	Constant Size	-6857.122	3E+16	672 (179–1450)	5.88×10^{-4}	2.25×10^{-4}	8.46×10^{-4}	5.35×10^{2}	2.79×10^{2}, 8.39×10^{2}	N/A	N/A
	Expansion Growth	-6866.461	5E+16	624 (155– 1365)	5.42×10^{-4}	2.21×10^{-4}	8.66×10^{-4}	2.57×10^{4}	2.09×10^{2}, 8.35×10^{4}	3.41×10^{-1}	5.32×10^{-5}, 0.13×10^{1}
	Exponential Growth	-6850.188	6E+16	271 (127–470)	6.12×10^{-4}	3.41×10^{-4}	8.93×10^{-4}	7.94×10^{2}	3.63×10^{2}, 1.31×10^{3}	1.13×10^{-2}	1.18×10^{-3}, 2.26×10^{-2}
	Bayesian Skyline	-6877.483	5E+15	603 (110–1730)	6.84×10^{-4}	6.49×10^{-5}	1.06×10^{-3}	4.25×10^{3}	3.16×10^{2}, 2.31×10^{4}	N/A	N/A
Relaxed Lognormal	Constant Size	-6918.626	4E+09	866 (295–1534)	3.74×10^{-4}	2.04×10^{-4}	5.41×10^{-4}	8.03×10^{2}	4.30×10^{2}, 1.26×10^{3}	N/A	N/A
	Expansion Growth	-6947.521	2E+09	1116 (318–2250)	2.96×10^{-4}	1.04×10^{-4}	4.91×10^{-4}	2.96×10^{4}	4.14×10^{2}, 1.07×10^{5}	1.24×10^{-1}	4.00×10^{-3}, 3.01×10^{-1}
	Exponential Growth	-6914.409	4E+09	492 (193–884)	3.80×10^{-4}	1.96×10^{-4}	5.73×10^{-4}	1.11×10^{3}	5.18×10^{2}, 1.85×10^{3}	6.74×10^{-3}	1.12×10^{-3}, 1.33×10^{-2}
	Bayesian Skyline	-7060.160	2E+09	2431 (229–7687)	2.14×10^{-4}	4.21×10^{-6}	3.98×10^{-4}	9.95×10^{3}	7.03×10^{2}, 7.20×10^{4}	N/A	N/A
Coherently-evolving coat protein region + 16 codons (cCP+16)											
Strict clock	Constant Size	-7040.533	4.489	5188 (786 – 13623)	1.00×10^{-4}	1.85×10^{-6}	2.00×10^{-4}	7.00×10^{3}	1.00×10^{3}, 1.87×10^{4}	N/A	N/A
	Expansion Growth	-7041.444	1.806	6917 (1146 – 18685)	6.57×10^{-5}	3.58×10^{-6}	1.38×10^{-4}	1.08×10^{5}	2.56×10^{3}, 3.46×10^{5}	3.26×10^{-2}	5.25×10^{-4}, 8.18×10^{-2}
	Exponential Growth	-7042.035		2767 (630 – 6914)	1.14×10^{-4}	1.63×10^{-5}	2.19×10^{-4}	5.40×10^{3}	1.09×10^{3}, 1.38×10^{4}	1.57×10^{-3}	1.81×10^{-4}, 3.19×10^{-3}
	Bayesian Skyline	-7040.704	3.785	12891 (1208 – 41307)	4.13×10^{-5}	2.37×10^{-6}	1.03×10^{-4}	6.96×10^{4}	7.65×10^{3}, 2.90×10^{5}	N/A	N/A
Relaxed Exponential	Constant Size	-6999.220	4E+18	1122 (227 – 2597)	4.39×10^{-4}	1.80×10^{-4}	7.04×10^{-4}	9.44×10^{2}	3.69×10^{2}, 1.75×10^{3}	N/A	N/A
	Expansion Growth	-6994.719	4E+20	4070 (335 – 11643)	1.14×10^{-4}	4.98×10^{-6}	2.50×10^{-4}	2.83×10^{6}	1.57×10^{3}, 1.04×10^{7}	7.49×10^{-2}	4.13×10^{-3}, 1.63×10^{-1}
	Exponential Growth	-6999.261	4E+18	371 (133 – 717)	4.29×10^{-4}	1.44×10^{-4}	7.24×10^{-4}	1.83×10^{3}	5.37×10^{2}, 3.90×10^{3}	1.16×10^{-2}	2.10×10^{-3}, 2.25×10^{-2}
	Bayesian Skyline[d]	-6994.360	5E+20	4497 (347 – 14184)	1.04×10^{-4}	7.15×10^{-6}	2.30×10^{-4}	7.60×10^{4}	7.10×10^{3}, 3.02×10^{5}	N/A	N/A
Relaxed Lognormal	Constant Size	-7018.724	1E+10	2442 (379 – 5458)	1.98×10^{-4}	2.59×10^{-5}	3.74×10^{-4}	2.76×10^{3}	5.49×10^{2}, 5.98×10^{3}	N/A	N/A
	Expansion Growth	-7014.287	1E+12	6441 (592 – 17548)	8.17×10^{-5}	3.73×10^{-6}	1.79×10^{-4}	1.84×10^{5}	1.40×10^{3}, 6.13×10^{5}	4.32×10^{-2}	1.56×10^{-3}, 9.87×10^{-2}
	Exponential Growth	-7017.211	6E+10	1161 (233 – 2629)	1.93×10^{-4}	3.84×10^{-5}	3.62×10^{-4}	3.55×10^{3}	9.14×10^{2}, 7.91×10^{3}	4.19×10^{-3}	1.69×10^{-4}, 8.69×10^{-3}
	Bayesian Skyline	-7010.831	4E+13	9767 (571 – 32075)	5.72×10^{-5}	2.49×10^{-6}	1.42×10^{-4}	6.61×10^{4}	6.83×10^{3}, 2.91×10^{5}	N/A	N/A

[a] Not applicable.

[b] Exponential-growth model (estimates for HC-Pro) is not much better than the constant-size model. So its better to favour the simpler model (constant population size).

[c] Bayesian Skyline (estimates for cCP+16); clearly the uncorrelated exponential relaxed clock is the best of the three different clock models. However, the expansion model is possibly better than the Bayesian skyline model. This is because the Bayes factor supporting the Bayesian skyline over the expansion model is not very large, and also because the expansion model has far fewer parameters. Given that there is not strong evidence to favour the Bayesian skyline, it is better to choose the simpler model (expansion).

Where models are compared the best-fit model is shaded grey.

The Logistic Growth model was not applicable to the data analysed in this study because the parameter estimates did not converge to reliable values during the computation (data not shown).

Table 19. Details of the data sets used for estimation of nucleotide substitution rate and time to the most recent common ancestor.

Parameter	Region[a]				
	HC-Pro	P3	NIb	cCP	cCP+16
Sequence length (nt)	1374	1065	1551	711	759
No. of sequences	108	109	115	113	113
Sampling date range	1968 – 2007	1968 – 2007	1968 – 2007	1968 – 2007	1968 – 2007
Best-fit substitution model	GTR + I + r	GTR + I + r	GTR + I + r	GTR + I + r	GTR + I + r
Best-fit molecular clock model	Relaxed Uncorrelated Exponential	Relaxed Uncorrelated Exponential	Relaxed Uncorrelated Exponential	Relaxed Uncorrelated Exponential	Relaxed Uncorrelated Exponential
Best-fit population growth model	Constant Size	Constant Size	Constant Size	Exponential Growth	Expansion Growth

[a] HC-Pro; Helper component-proteinase protein. P3; Protein 3. NIb; Nuclear inclusion b protein. cCP; Coherently-evolving coat protein. cCP+16; Sequences that include the 16 codons (48 nucleotides) at 5'-terminus immediately adjacent to cCP.

Preliminary analyses of all 108-115 sequences in each dataset revealed large differences (up to five-fold) in the mean TMRCAs estimated for different genes (Table 20; mixed hosts), despite care having been taken to remove recombinant sequences. Thus, these results were unable to provide a reliable estimate of the emergence time of TuMV, which was the principal objective of this project. The results were no more consistent when the analyses were confined to genes from the 28 complete genomic sequences that had collection date and no significant recombination signals (data not shown). This study found that different regions of the CP gene sequences gave quite different TMRCA estimates: the cCP region (Gibbs *et al*., 2008c; Ward *et al*., 1995) yielded much smaller TMRCAs than sequences that included the 16 codons (48 nts) immediately adjacent to the cCP region (Table 20).

To investigate whether there is a temporal signal in the data sets, this study used two available methods. First, the correlation between root-to-tip distances and sampling date were investigated for temporal signal using Path-O-Gen were calculated (Figure 20). Second, replicate data sets in which the ages of the sequences were randomized, as done in a number of recent studies (Ho *et al*., 2011; Pagán and Holmes, 2010; Pagán *et al*., 2010). Parameter estimates from each dataset were only considered reliable if the mean posterior rate obtained using the original sample dates was outside the 95% credibility intervals of rates estimated from the date-randomized data. The HC-Pro, P3, NIb, and cCP, but not the cCP+16, data sets passed this date-randomization test. In addition, the rate estimates from the original data sets had much smaller 95% credibility intervals than those from the date-randomized data sets (Figure 21). The mean estimated substitution rates for the three largest genes of all TuMVs, excluding the results for the non-brassica isolates, are HC-Pro 1.11 x 10^{-3}, P3 1.11 x 10^{-3}, and NIb 0.78 x 10^{-3} subs/site/year (Table 20).

Table 20. Estimates of the substitution rates and times to the most recent common ancestor (TMRCA) for TuMVs isolated from various hosts, from only Brassicaceae or from non-Brassicaceae, or from orchids and Brassicaceae.

TuMV isolates from:	Region[b]	No. of sequences	Nucleotide substitution rate[a]		TMRCA (years)	
			Mean	95% CI[c]	mean	95% CI
a mixture of hosts	HC-Pro	108	1.07×10^{-3}	$5.82 \times 10^{-4} – 1.56 \times 10^{-3}$	819	258 – 1643
	P3	109	1.08×10^{-3}	$5.59 \times 10^{-4} – 1.60 \times 10^{-3}$	1071	279 – 2511
	NIb	115	7.04×10^{-4}	$3.73 \times 10^{-4} – 1.00 \times 10^{-3}$	1330	342 – 2920
	cCP	113	6.12×10^{-4}	$3.41 \times 10^{-4} – 8.93 \times 10^{-4}$	271	127 – 470
	cCP+16	113	1.14×10^{-4}	$4.98 \times 10^{-6} – 2.50 \times 10^{-4}$	4070	335 – 11643
Brassicaceae only	HC-Pro	88	1.14×10^{-3}	$4.49 \times 10^{-4} – 1.75 \times 10^{-3}$	800	192 – 1775
	P3	88	1.17×10^{-3}	$4.97 \times 10^{-4} – 1.89 \times 10^{-3}$	829	196 – 1815
	NIb	93	7.61×10^{-4}	$3.26 \times 10^{-4} – 1.18 \times 10^{-3}$	928	191 – 1936
	cCP	91	5.63×10^{-4}	$1.57 \times 10^{-4} – 9.57 \times 10^{-4}$	446	110 – 973
	cCP+16	92	9.91×10^{-5}	$1.84 \times 10^{-6} – 2.59 \times 10^{-4}$	4027	157 – 12650
non-Brassicaceae	HC-Pro	20	1.97×10^{-3}	$9.91 \times 10^{-6} – 4.31 \times 10^{-3}$	765	63 – 1855
	P3	21	2.71×10^{-3}	$3.81 \times 10^{-6} – 6.19 \times 10^{-3}$	959	56 – 1967
	NIb	22	1.19×10^{-3}	$2.65 \times 10^{-7} – 3.16 \times 10^{-3}$	3485	66 – 6997
	cCP	22	1.61×10^{-3}	$4.44 \times 10^{-4} – 2.90 \times 10^{-3}$	186	49 – 414
	cCP+16	21	1.65×10^{-3}	$3.76 \times 10^{-4} – 3.02 \times 10^{-3}$	178	46 – 405
Orchids and Brassicaceae	HC-Pro	92	1.13×10^{-3}	$6.31 \times 10^{-4} – 1.68 \times 10^{-3}$	754	223 – 1555
	P3	92	1.09×10^{-3}	$4.93 \times 10^{-4} – 1.76 \times 10^{-3}$	1030	241 – 2219
	NIb	97	8.63×10^{-4}	$4.44 \times 10^{-4} – 1.28 \times 10^{-3}$	1025	240 – 2365
	cCP	95	6.85×10^{-4}	$3.85 \times 10^{-4} – 1.02 \times 10^{-3}$	284	114 – 533
	cCP+16	95	2.47×10^{-4}	$4.66 \times 10^{-6} – 6.47 \times 10^{-4}$	3041	170 – 10631

[a] substitutions/site/year.
[b] HC-Pro; Helper component-proteinase protein. P3; Protein 3. NIb; Nuclear inclusion b protein. cCP; Coherently-evolving coat protein. cCP+16; Sequences that include the 16 codons (48 nucleotides) at 5'-terminus immediately adjacent to cCP.
[c] 95% credibility interval.

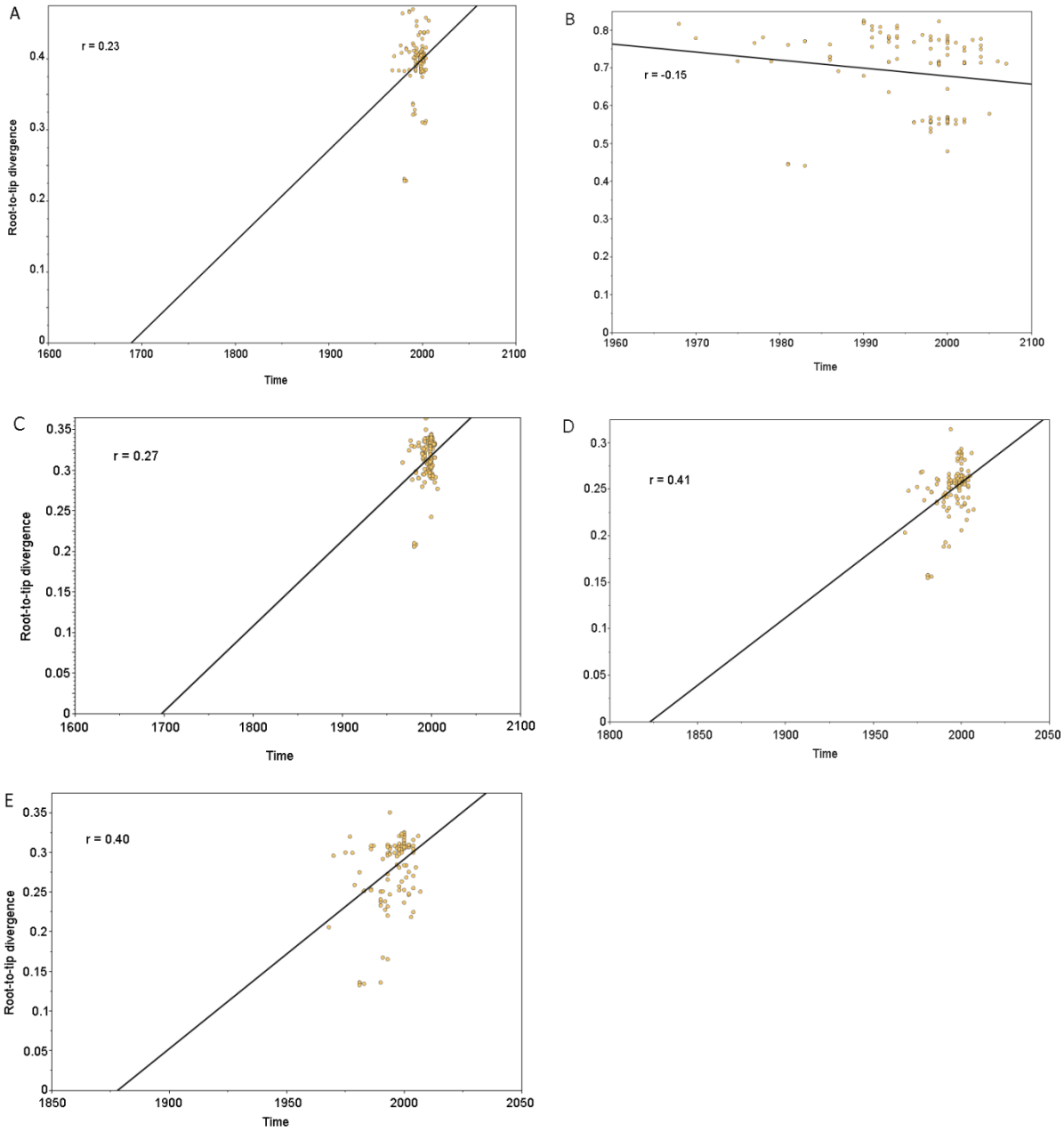


Figure 20. Genetic distance verus sampling year for four coding regions of *Turnip mosaic virus*; helper component proteinase (HC-Pro) genes (A), protein 3 (P3) genes (B), nuclear inclusion b protein (NIb) genes (C), coherently-evolving CP (cCP) genes (D), and cCP+16 genes (E). Week temporal (low r values) or no evidence (negative r value) of temporal structure within the sampled period was found in those coding regions using this method.

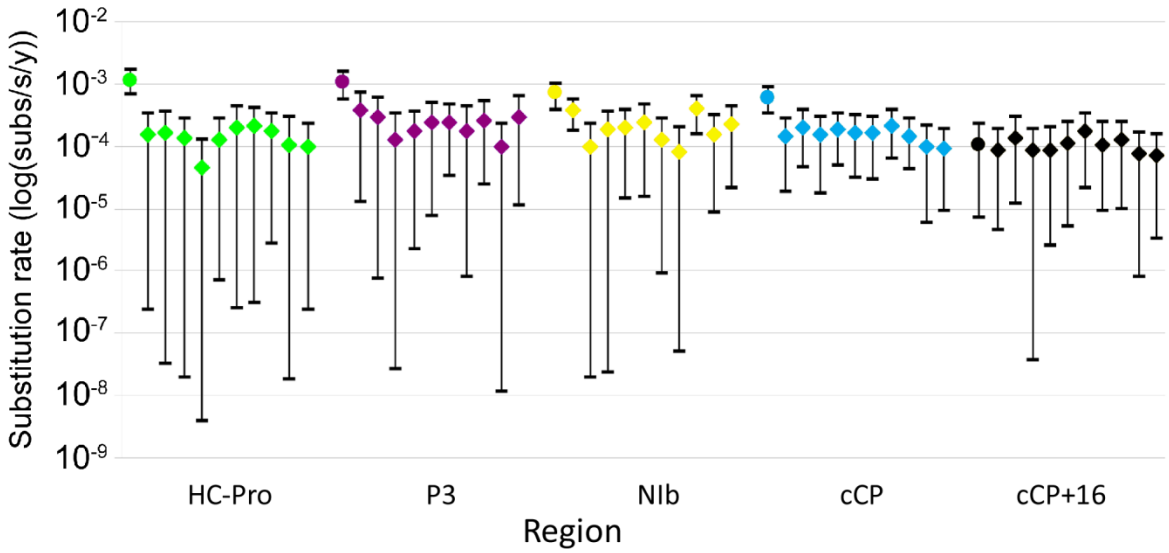


Figure 21. Estimates of nucleotide substitution rates. Mean estimates and 95% credibility intervals are shown. These were estimated from 108 helper component proteinase (HC-Pro) genes, 109 protein 3 (P3) genes, 115 nuclear inclusion b protein (NIb) genes, 113 coherently-evolving CP (cCP) genes, and 113 cCP+16 genes (see text) from non-recombinant and dated gene sequences of isolates obtained from species of non-brassicas and brassicas. In each set of estimates, the first is based on the original data, whereas the remaining ten values are from date-randomized replicates. The 95% credibility intervals of the estimates from the date-randomized replicates do not overlap with the mean posterior estimate from the original data set. In addition, the lower tails of the credibility intervals are long and tend towards zero. These features suggest that there is sufficient temporal structure in the original data sets for rate estimation.

Analysis of sequences of BI isolates from non-brassicas gave more variable results (Table 20; mean TMRCA estimates cCP 186 years to NIb 3485 years) than those from the isolates from brassicas (Table 20; mean TMRCA estimates cCP 446 years to NIb 928 years). This might be due to sampling variability, however, because there were only around 21 isolates in each non-brassica dataset compared with around 90 isolates in brassica datasets. We checked this by estimating the TMRCAs of 10 subsets of 21 sequences randomly selected from the 88 P3 genes from brassica isolates. These subsets produced widely variable mean estimates of TMRCAs (Figure 22). Most of these estimates are lower than that from the complete 88 sequence set, but two were considerably higher. For example, mean TMRCA estimates ranged from 99 years to 10,531 years (mean, 1385 years; logarithmic mean 354 years), compared with a mean estimate of 829 years for the complete set of 88 P3 sequences. Our analyses showed that the estimated TMRCAs and evolutionary rates for the three largest genes of each data set (i.e., HC-Pro, P3, and NIb) were always more similar to each other than to those from the cCP gene. We suggest that the temporal signal in the cCP sequences is less reliable because the cCP gene is around half the length of the others and has the lowest evolutionary rate (Table 20). Accordingly, we excluded the estimates obtained from this gene when inferring the time of divergence of the BI lineages and Orchis group isolates.

The three largest genes provided two estimates of the divergence time of the BI and Orchis isolates. The sequences from the “mixture of hosts” isolates (Table 20) gave a mean date estimate of 1073 YBP. The sequences from the “brassicas and orchids, but not non-brassicas” (Table 20), namely all isolates except those from non-Brassicaceae, gave a mean date estimate of 936 YBP. Even though the latter data set comprised a smaller number of sequences, the estimates differed less among genes. Therefore, in the absence of a clear-

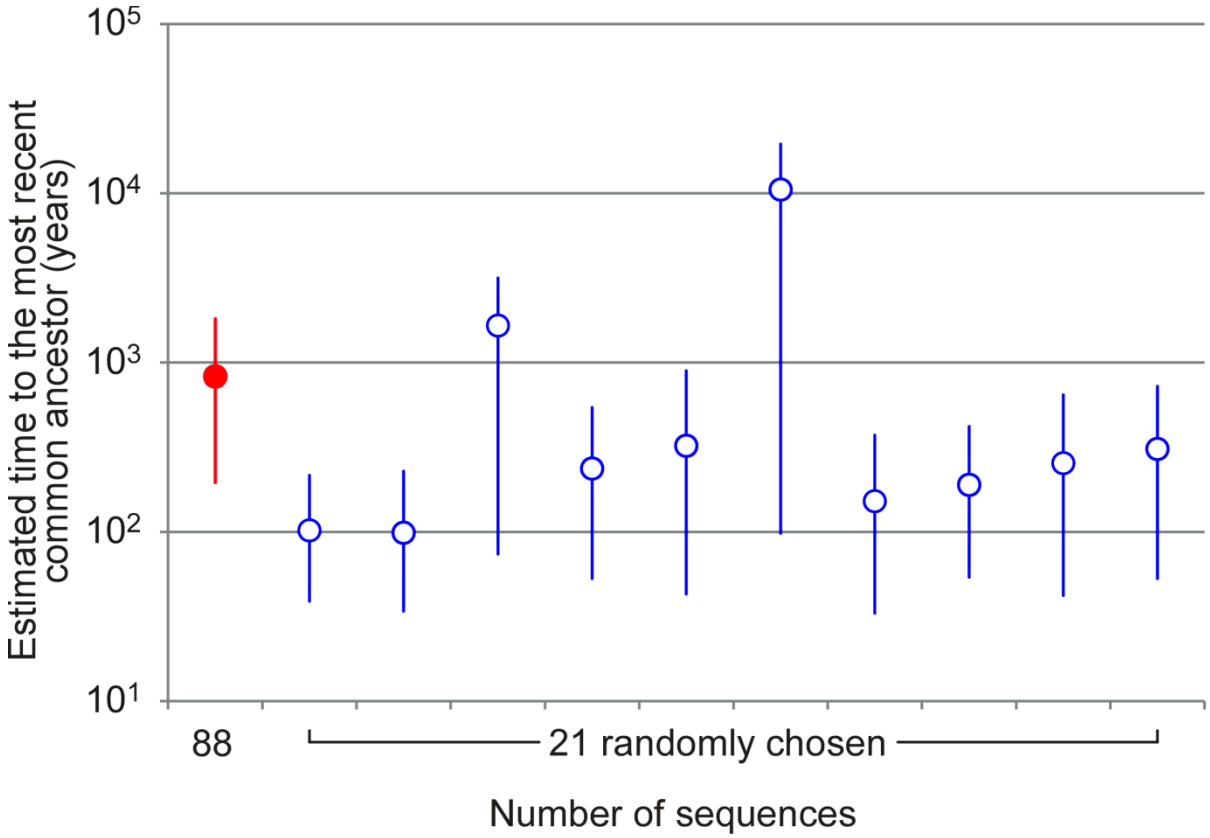


Figure 22. Estimated times to the most recent common ancestors of 88 P3 gene sequences and randomly selected subsets of 21 sequences. The leftmost data point shows the estimate from the original 88 P3 sequences. The remainining 10 data points show the estimates for each of 10 randomly selected sets, each comprising 21 sequences. Error bars indicate 95% credibility intervals.

reason for distinguishing between these two sets of results, we take the mean of these two values, 1005 YBP (mean 95% credibility interval from 264 YBP to 2203 YBP), as the most likely date of emergence of BI isolates. In addition, the “brassicas only” data set gives, for the three largest genes, a consistent estimate of the divergence time of the main lineages of BI of about 852 YBP (mean 95% CI from 193 YBP to 1842 YBP; mean TMRCAs of 800 YBP for HC-Pro, 829 YBP for P3, and 928 YBP for NIb). In this analysis, an example of maximum clade credibility (MCC) tree of NIb mixed sequences was showed (Figure 23). The MCC tree divided NIb mixed sequences into five genogroups; orchis group, basal-B, basal-BR, Asian-BR and world-B. TMRCA was about 1330 year ago.

2.5. Discussion

The phylogenetic analyses have shown that a group of isolates from European orchids form a small monophyletic sister group to all the TuMV isolates that readily infect brassicas. The same phylogeny was inferred by all analytical methods (maximum likelihood, neighbor-joining, and Bayesian inference), except for some analyses of the HC-Pro data, and was also inferred with the full genomes, the encoded polyproteins, or individual genes. The Orchis and BI isolates form phylogenetically distinct sister lineages within the TuMV group of potyviruses, which also includes the more distantly related JYMV, NYSV, and ScMV.

The Orchis group isolates came from orchids grown for two years in a glasshouse collection of geophyte orchids at Celle, Germany. It is uncertain whether one or more of these species are hosts of Orchis group viruses in nature because the same virus was also found in other orchid species in the same collection. Nonetheless, all of the infected plants have overlapping natural distributions in central and southern Europe within the region identified as the likely ‘centre of emergence’ of BI isolates.

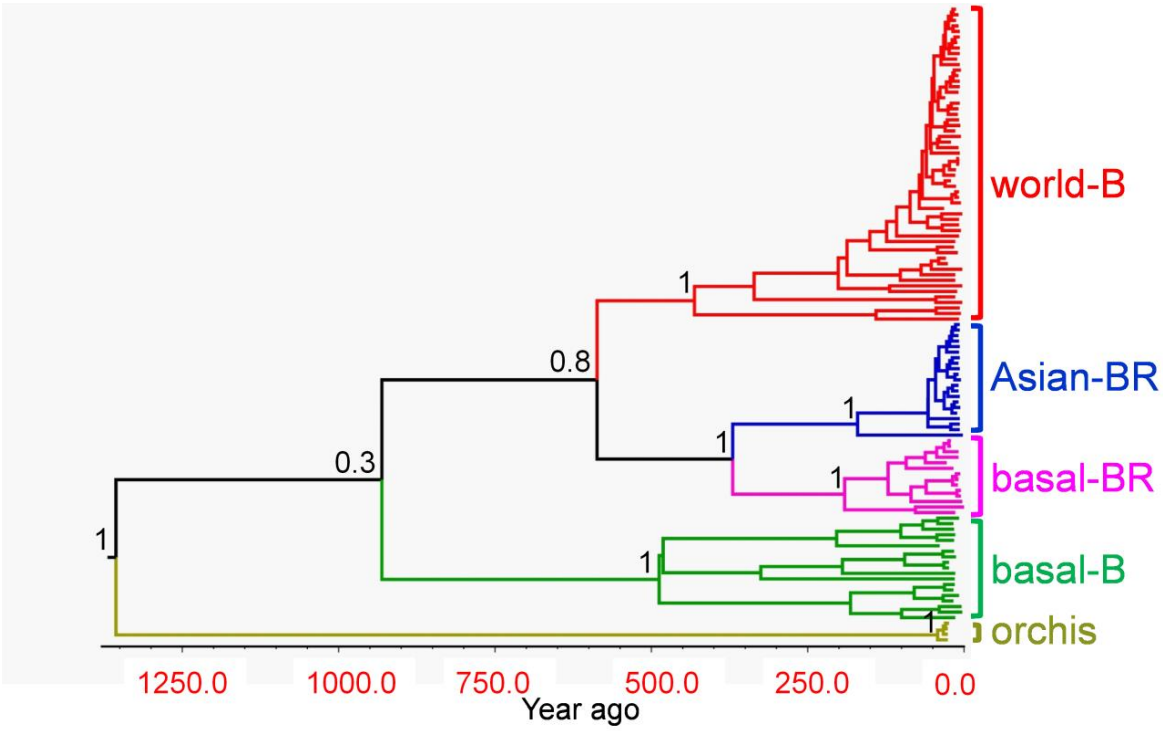


Figure 23. Maximum clade credibility (MCC) phylogeny of TuMV based on NIb data set using 115 sequences. The tree was calculated from posterior distribution of trees generated by Bayesian MCMC coalescent analysis with BEAST. Posterior probabilities are indicated above branches.

Earlier detailed serological tests had shown that TuMV-OM was most closely related to, but distinct from, TuMV (Lesemann and Vetten, 1985). This was supported by sequence analysis of the coat protein gene of the Orchis isolate (Gibbs *et al*., 2000). In the present study, we have confirmed and extended these earlier results and found that the Orchis isolates are biologically distinct from all BI isolates. Only one orchid isolate infected brassicas systemically, but did not produce symptoms, and none infected the monocotyledonous hosts of the outgroup viruses, JYMV, NYSV, and ScMV. Hence, the Orchis isolates are biologically and phylogenetically distinct from BI isolates, and should perhaps be considered a separate viral species (http://ictvonline.org/codeOfVirusClassification_2002.asp).

The immediate ancestor of the sister lineages were probably either brassica-infecting and spread to orchids, or orchid-infecting and spread to brassicas. In the absence of direct evidence, we conclude that the latter scenario is more likely because it involves fewer host changes. Also the Orchis virus population has the characteristics one would expect of the source of the BI lineage: it is phylogenetically distinct but closely related; it is found in a host more closely related to those of known outgroup viruses than to BI, although some isolates can infect species of Brassicaceae; and it is from western Eurasia, matching the likely location of the emergent virus population.

When attempting to estimate the timing of the divergence between BI and Orchis lineages, our initial Bayesian results were inconsistent but instructive. We found sufficient temporal structure among samples from the genome region encoding the coherently-evolving C-terminal part of the CP, but this data set still gave TMRCA estimates that differed substantially from the other genes. This was possibly due to the small size of the data set, in terms of both sequence length and sample size. Some of the variation in rates among genes might be due to differences in sites under purifying selection, which appears to have had a

strong effect on the four genes. However, estimates of evolutionary rates were very similar among the three largest genes. The effects of purifying selection, whereby younger branches of the tree tend to carry an elevated number of transient polymorphisms, might have led to an overestimation of the mutation rate in our analysis (Ho *et al*., 2011; Holmes, 2003). However, our rate estimates are very close to the mean rate found in a survey of virus studies (Sanjuán, 2012), but higher than those reported earlier for ZYMV (5.0×10^{-4} subs/site/year) and for a group of potyviruses (1.15×10^{-4} subs/site/year) (Gibbs *et al*., 2008c; Simmons *et al*., 2008).

The OM and BI lineages were estimated to have diverged around 1005 YBP; long before TuMV was first isolated in North America in 1921(Gardner and Kendrick, 1921). The phylogenetic trees (Figures 1 and 2) indicate that the BI lineage subsequently radiated around 850 YBP to give the four present-day lineages. The basal-B lineage, isolates of which have been found most often in Europe, radiated soon after the BI lineages were established. The other three lineages diversified more recently to be found in other parts of the world.

Although the TuMV population of the world has been more thoroughly sampled than that of any other plant virus, it is important to realize that only four Orchis isolates have been examined so far. Accordingly, our conclusions about its role in the evolution of TuMV must be treated with caution. It is difficult to obtain independent evidence to corroborate our estimate of the time of divergence of TuMV and Orchis group isolates. Nevertheless, the conclusion that BI lineage emerged in Western Europe around 1000 YBP is congruent with historical records of conditions existing at that time. Before then agriculture was small-scale and the landscape was dominated by natural ecosystems. Eurasian agriculture had first developed around 11,000 – 10,500 YBP in the region bordering the eastern Mediterranean from Turkey to Egypt (i.e., the Levant), and was based on the domestication of cereals,

legumes, flax and grazing animals. It dispersed north and west into Eurasia around 8,500 YBP and finally to the north and west fringes of Europe around 6,000 YBP (Bartlett, 1994; Bellwood, 2005; Pinhasi at el., 2005).

The genetically complex group of brassica hybrids now grown as crops, including turnips, was domesticated from about 4000 YBP, probably around the Mediterranean and cooler parts of northwest Europe (Crisp, 1995; Hemingway, 1995; Hodgkin, 1995; MacNaughton, 1995a, 1995b), and became a staple of the diet of humans and domesticated animals. In the Medieval Warm Period, 1050 – 750 YBP, there was a warming of the climate in West Eurasia and a great increase in both the human population and the extent of agriculture; forests and marshes were cleared and cultivated (Bartlett, 1994). This corresponds to the period during which, we suggest, the first BI lineage emerged. The landscape changed from isolated farming settlements set in woodlands to a landscape of contiguous farmlands with minor dispersed woodlands, and this would have fostered the spread of crop diseases. The conditions would have been ideal for a potyvirus like Orchis isolates, able to infect brassicas, to emerge from wild hosts and adapt to the increased population of brassicas provided by crops and their weeds.

Furthermore, potyviruses are spread by aphids as they non-specifically probe plants seeking suitable hosts on which to breed. Most potyviruses are spread by aphids from Aphidineae, a group that is unusual among phytophagous insects in that most species alternate between woody winter hosts and herbaceous summer hosts, and thus aphidines were also fostered by the conditions of early broadscale agriculture (Gibbs *et al*., 2008c). The spread of BI lineage to produce the present global distribution of the virus probably had to wait until international maritime trade was established after the discovery of the Americas and routes to the Far East by European adventurers around 500 YBP. Similar analyses of the

relationships and evolutionary timescales of the *Bean common mosaic* and *Potato virus Y* groups of potyviruses also concluded that most species had arisen in the last 1000 to 3000 years, and had involved similarly unknown host:virus:landscape specificities (Gibbs and Ohshima, 2010; Gibbs *et al*, 2008d).

Further testing of wild populations of European orchids and brassicas will enable us to determine which species are the primary hosts or hosts of Orchis isolates, and also whether any intermediates in the adaptation of this virus to brassicas have survived. Furthermore, because the primary hosts of all potyviruses are monocotyledonous bulb and grass species from Western Eurasia, a broad survey of such plants might reveal other relictual potyvirus populations and provide insight into the intermediate stages of potyvirus evolution.

VI. GENERAL DISCUSSION AND CONCLUSIONS

The spatial and temporal dynamics of RNA viruses are often reflected by their phylogenetic structures (Biek *et al.*, 2006; Grenfell *et al.*, 2004; Holmes, 2004). Recently, phylogenetic analyses or a Bayesian coalescent approach for evolutionary rate and timescale were reported for both animal and plant viruses. In this study, the major findings are (1) the TuMV populations of Vietnam, China and Japan had clear local founder effects and frequent gene flow, (2) a sister lineage from European orchids was the original host of TuMV and (3) the evolutionary rate of this virus was about 1.0×10^{-3} nucleotide subs/site/year and it spread to Eurasian brassica crops from wild ochids about 1000 years ago.

Spatial evolution analyses have been reported as many for both animal and plant viruses. Plant RNA viruses are very variable enabling them to adapt rapidly to new or resistant hosts. Privious studies on the population genetic structure reported that virus populations have been shaped by selection, founder effects and genetic recombination (Tsompana *et al*., 2005; Ohshima *et al*., 2010; Karasev *et al*., 2011; Ogawa *et al*., 2008, 2012; Seo *et al*., 2009; Zhang *et al*., 2011; Lecoq *et al*., 2009; Ohshima *et al*., 2002; Tomimura *et al*., 2004; Tomitaka and Ohshima, 2006). In this study, spatial evolution was analyzed using TuMV populations in Vietnam, China, and Japan. The results indicate that it is strongly influenced by founder effects and clear evidence of gene flow among Chinese, Japanese and Vietname TuMV populations. The results described that frequent gene flow in TuMV populations was major driving forces for the spatial evolution of TuMV.

During 2008-2012, temporal evolution of many animal and plant viruses was analyzed. The rate of evolution of an RNA plant virus has been estimated using temporally spaced sequence data as the first report (Fargette *et al*., 2008). The later estimation was reported

using CP genes for ZYMV (Simmons *et al*., 2008) and for Potyviruses (Gibbs *et al*., 2008c; 2010). Evolution rate of fifty partial coat protein genes of Potyviruses was estimated to be 1.15 x 10^{-4} subs/site/year (Gibbs *et al*., 2008c). Furthermore, estimation rate of single species using CP gene, ZYMV, was 5.0 x 10^{-4} subs/site/year (Simmons *et al*., 2008). Especially, TMRCA of potyviruses was 7250 years ago (Gibbs *et al*., 2008c) and ZYMV was 500 years ago for (Simmons *et al*., 2008). In this study, evolutionary rate of TuMV was estimated to be 6.12 x 10^{-4} subs/site/year for cCP genes. However, this is the first report that estimation rate is not only for cCP but also HC-Pro, P3 and NIB genes. The results showed that substitution rates were 1.11 x 10^{-3} (HC-Pro), 1.11 x 10^{-3} (P3), and 0.78 x 10^{-3} (NIb) subs/site/year. The TMRCA of TuMV was estimated about 1000 years ago. The study here indicated that HC-Pro, P3 and NIb were suitable encoding regions for temporal evolution analyses because they have clear sufficient temporal structure, whereas, cCP region was more variable at the beginning of the sequences and it had unclear sufficient temporal structure. Moreover, the results here suggested that a large number of sequences, broad collection and wide range of collection year were needed for the analysis of temporal evolution.

These studies on the both spatial and temporal evolution of TuMV are the first report for using different protein encoding regions. Understanding of the mechanisms and processes of viral evolution will help the virologists to predict what new infections and in what regions will be emerged and spread in the future. Consequently, studies of viral epidemiology and evolution should be integrated to acquire a better explanation of viral emergence and establishment for guide preventation, control and treatment of viral diseases. For future studies of evolution analyses, the different experimental procedures, including full sequences of TuMV would be required to use for analysing both spatial and temporal evolution.

VII. ACKNOWLEDGEMENTS

This PhD thesis is the result of my challenge during a long journey. Looking back through the journey, my PhD is the end of my journey and success with my main supervisors, subsupervisors, collabotors and supporters.

First and foremost, I would like to express my heartfelt and cincerest gratitude to Professor Dr. Kazusato Ohshima, my main supervisor at Laboratory of Plant Viology, Saga University, who is very kind to help me during my PhD journey. I am grateful for your thoughtful and detailed comments. His wide knowledge and his logical way of thinking have been of great value for me and his guidance helped me in all the time of research and writing manuscript of the PhD thesis. I also would like to thank you deeply for giving to my wife, Mrs. Hoa Tran Thi Nhu - a researcher at Hanoi University of Agriculture, a chance to do research in your laboratory. I expect that he continues to support my post-graduate researches on the plant viruses.

Besides my main supervisor, my sincerest grateful to other two subsupervisors, Associate Professor Dr. Motoaki Kusaba, Laboratory of Plant Mycology, Department of Applied Biological Sciences, Faculty of Agriculture, Saga University and Professor Dr. Hisashi Iwai, Department of Plant Pathology, Kagoshima University for their helps during my study. I would like to thank to them for their carefully reading the manuscript of thesis and as well as your suggestions to make my PhD thesis going to end.

To Professor Dr. Zenichi Moromizato, Faculty of Agriculture, Ryukyu Uniersity and Associate Professor Dr. Makoto Tokuda, Faculty of Agriculture, Saga University for your careful readings of the manuscript of my thesis. I would like to express my deepest thanks to them for their help and giving me suggestions for improving the PhD thesis.

To my dear supporters, Professor Dr. Adrian J. Gibbs (Emeritus Faculty, Australian National University, Canberra, ACT 0200, Australia), Associate Professor Dr. Simon S.W. Ho and Dr. Sebastián Duchêne (School of Biological Sciences, University of Sydney, Australia), very many thanks for their kind collaborations and support techniques of molecular analyses.

My sincerely thanks also to Professors Drs. Heinrich-Josef Vetten and Dietrich Lesemann (Julius Kuehn Institute, Federal Research Centre for Cultivated Plants (JKI), Institute of Epidemiology and Pathogen Diagnostics, Messeweg 11-12, 38104, Braunschweig, Germany), and Dr. John A. Walsh (Life Sciences, University of Warwick, Wellesbourne, Warwick, CV35 9EF, UK) for their kind collaboration and supplied useful TuMV samples.

I also thank Drs. A. Ragozzino, B. McKoewn, C. Kerlan, C. Lu, F. Ponz, I. Cooper, J. Verhoeven, J. Vollman, K. Ostrowka, M. Verhoven, P. Hunter, P. Roggero, R. Kramer, R. Provvidenti, S. Korkmaz, T. van der Horst, V. Lisa, X. Kyriakopoulou, and X. Weidemann, for providing TuMV isolates and plant materials.

I would like to thank Dr. J. Armstrong, Dr. Y. Tomitaka, H. Takeshita, A. Sadayuki, R. Koga, K. Tomimura, K. Fukumoto, and M. Umeo (Saga University) for their careful technical assistance. I also express sincere thanks to Giang Ha (Hanoi University of Agriculture) and Chung Van Phan (Dak Lak) for supplying some TuMV isolates in Vietnam.

To the official staffs of Saga University and Kagoshima University, I am grateful for your help to complete all necessary documents for this course as well as their supporting all the importance information of the lectures, seminars.

In my daily experiments, I have been blessed with a friendly and cheerful group of Japanese students, my vekindly lab mates, for everything we have shared during my three-year long studying and doing research works here.

Next, I would like to express my sincerest gratitude to my family members whose encouragement inspired me while living far away from fatherland.

I would like to express my extremely thanks to my colleagues in Department of Plant Pathology, Faculty of Agronomy, Hanoi University of Agriculture, for their help as well as great discussions with me.

Finally, I am greateful to Japanese Government for supporting Monbukagakusho Scholarship for school fee, living cost during my study in Japan.

VIII. REFERENCES

Adams, M.J., Antoniw, J., Fauquet, C.M., 2005. Molecular criteria for genus and species discrimination within the family *Potyviridae*. Archives of Virology 150, 459–479.

Abo, M.E., Sy, A.A., Alegbejo, M.D., 1997. Rice Yellow Mottle Virus (RYMV) in Africa: Evolution, Distribution, Economic Significance on Sustainable Rice Production and Management Strategies. Journal of Sustainable Agriculture. 11, 85–111.

Aboa, M.E., Syb, A.A., Alegbejoc, M.D., 1997. Rice yellow mottle virus (RYMV) in Africa: Evolution, distribution, economic significance on sustainable rice production and management strategies. Journal of Sustainable Agriculture 11, 85–111.

Abubakar, Z., Ali, F., Pinel, A., Traoré, O., N'Guessan, P., Notteghem, J.L., Kimmins, F., Konaté, G., Fargette, D., 2003. Phylogeography of *Rice yellow mottle virus* in Africa. Journal of General Virology 84, 733–743.

Albiach-Martí, M.R., Mawassi, M., Gowda, S., Satyanarayana, T., Hilf, M.E., Shanker, S., Almira, E.C., Vives, M.C., López, C., Guerri, J., Flores, R., Moreno, P., Garnsey, S.M., Dawson, W.O., 2000. Sequences of *Citrus tristeza virus* separated in time and space are essentially identical. Journal of Virology 74, 6856–6865.

Baillie, G.J., Galiano, M., Agapow, P.M., Myers, R., Chiam, R., Gall, A., Palser, A.L., Watson, S.J., Hedge, J., Underwood, A., Platt, S., McLean, E., Pebody, R.G., Rambaut, A., Green, J., Daniels, R., Pybus, O.G., Kellam, P., Zambon, M., 2012. Evolutionary dynamics of local pandemic H1N1/2009 influenza virus lineages revealed by whole-genome analysis. Journal of Viology 86, 11–18.

Bakker, W., 1974. Characterization of and ecological aspects of rice yellow mottle virus in Kenya. Agricultural Research Report 829. Wageningen. The Netherlands Publishing and Documentation. 152 pp.

Barre-Sinoussi, F., Chermann, J., Rey, F., Nugeyre, M., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vézinet-Brun, F., Rouzioux, C., Rozenbaum, W., Montagnier, L., 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science 220, 868–871.

Bartlett, R., 1994. The Making of Europe: Conquest, Colonization, and Cultural Change 950-1350. Princeton University Press. 447 pp.

Bateson, M.F., Lines, R.E., Revill, P., Chaleeprom, W., Ha, C.V., Gibbs, A.J., Dale, J.L., 2002. On the evolution and molecular epidemiology of the potyvirus *Papaya ringspot virus*. Journal of General Virology 83, 2575–2585.

Bedhomme, S., Lafforgue, G., Elena, S.F., 2013. Genotypic but not phenotypic historical contingency revealed by viral experimental evolution. BMC Evolutionary Biology 13:46. doi: 10.1186/1471-2148-13-46.

Bellwood, P., 2005. First farmers: the origins of agricultural societies. Oxford: Blackwell Publishing. 360 pp.

Biek, R., Drummond, A.J., Poss, M., 2006. A virus reveals population structure and recent demographic history of its carnivore host. Science 311, 538–541.

Bourhy, H., Kissi, B., Tordo, N., 1993. Molecular diversity of the *Lyssavirus* genus. Virology 194, 70–81.

Bourhy, H., Kissi, B., Audry, L., Smreczak, M., Sadkowska-Todys, M., Kulonen, K., Tordo, N., Zmudzinski, J.F., Holmes, E.C., 1999. Ecology and evolution of rabies virus in Europe. Journal of General Virology 80, 2545–2557.

Bourhy, H., Reynes, J.M., Dunham, E.J., Dacheux, L., Larrous, F., Huong, V.T., Xu, G., Yan, J., Miranda, M.E., Holmes, E.C., 2008. The origin and phylogeography of dog rabies virus. Journal of General Virology 89, 2673–2681.

Brantley, J.D., Hunt, A.G., 1993. The N-terminal protein of the polyprotein encoded by the potyvirus tobacco vein mottling virus is an RNA-binding protein. Journal of General Virology 74, 1157–1162.

Brown, J.K., Frohlich, D.R., Rosell, R.C., 1995. The sweetpotato or silverleaf white flies: biotypes of Bemisia tabaci or a species complex? Annual Review of Entomology 40,511–534.

Carrington, J.C., Jensen, P.E., Schaad, M.C., 1998. Genetic evidence for an essential role for potyvirus CI protein in cell-to-cell movement. Plant Journal 14, 393-400.

Chen, C.C., Chao, C.H., Chen, C.C., Yeh, S.D., Tsai, H.T., Chang, C.A., 2003. Identification of *Turnip mosaic virus* isolates causing yellow stripe and spot on calla lily. Plant Disease 87, 901–905.

Chen, J., Zheng, H.Y., Chen, J.P., Adams, M.J., 2002. Characterisation of a potyvirus and a potexvirus from Chinese scallion. Archives of Virology 147, 683–693.

Chen, J., Chen, J.P., Langeveld, S.A., Derks, A.F.L.M., Adams, M.J., 2003. Molecular characterization of carla- and potyviruses from *Narcissus* in China. Journal of Phytopathology 151, 26–29.

Chen, S., Das, P., Hari, V., 1994. In situ localization of ATPase activity in cells of plants infected by maize dwarf mosaic potyvirus. Archives Virology 134, 433–439.

Cheng, X.F., Wu, X.Y., Wang, H.Z., Sun, Y.Q., Qian, Y.S., Luo, L., 2012. High codon adaptation in citrus tristeza virus to its citrus host. Virology Journal 9:113. doi: 10.1186/1743-422X-9-113.

Chiu, I.M., Yaniv, A., Dahlberg, J.E., Gazit, A., Skuntz, S.F., Tronic, S.R., Asronson, S.A., 1985. Nucleotide sequence evidence for relationship of AIDS retrovirus to lentiviruses. Nature 317, 366–368.

Chung, B.Y.-W., Miller, W.A., Atkins, J.F., Firth, A.E., 2008. An overlapping essential gene in the *Potyviridae*. Proceedings of the National Academy of Sciences of the United States of America 105, 5897–5902.

Clark, M.F., Adams, A.N., 1977. Characteristics of the microplate method of enzyme-linked immunosorbent assay for the detection of plant viruses. Journal of General Virology 34, 475–483.

Cohen, M.S., Hellmann, N., Levy, J.A., DeCock, K., Lange, J., 2008. The spread, treatment, and prevention of HIV-1: evolution of a global pandemic. The Journal of Clinical Investigation 118, 1244–1254.

Crisp, P., 1995. Radish, *Raphanus sativus* (Cruciferae). In: Smartt J, Simmonds NW, editors. Evolution of Crop Plants. 2nd ed. UK, Harlow: Longman scientific & technical. pp. 86-89.

Cronin, S., Verchot, J., Haldeman-Cahill, R., Schaad, M.C., Carrington, J.C., 1995. Long-distance movement factor: a transport function of the Potyvirus helper component proteinase. Plant Cell 7, 549–559.

Cuevas, J.M., Delaunay, A., Rupar, M., Jacquot, E., Elena, S.F., 2012. Molecular evolution and phylogeography of *Potato virus Y* based on the CP gene. Journal of General Virology 93, 2496–2501.

Cuevas, J.M., Delaunay, A., Visser, J.C., Bellstedt, D.U., Jacquot, E., Elena, S.F, 2012. Phylogeography and molecular evolution of *Potato virus Y* . PLoS ONE 7:e37853.

Czosnek, H., Laterrot, H., 1997. A worldwide survey of tomato yellow leaf curl viruses. Archives of Virology, 142, 1391–1406.

Davis, P.L., Bourhy, H., Holmes, E.C., 2006. The evolutionary history and dynamics of bat rabies virus. Infection, Genetics and Evolution 6, 464–473.

Drummond, A.J., Pybus, O.G., Rambaut, A., Forsberg, R., Rodrigo, A.G., 2003. Measurably evolving populations. Trends in Ecology & Evolution 18, 481–488.

Drummond, A.J., Ho, S.Y.W., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. PLoS Biology 4: e88.

Drummond, A.J., Rambault, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology 7: 214. doi:10.1186/1471-2148-7-214.

Duffy, S., Holmes., E.C., 2008. Phylogenetic evidence for rapid rates of molecular evolution in the single-stranded DNA Begomovirus Tomato yellow leaf curl virus. Journal of Virology, 82, 957–965.

Excoffier, L., Laval, G., Schneider, S., 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evolution Bioinformatics Online 1, 47–50.

Fabre, F., Montarry, J., Coville, J., Senoussi, R., Simon, V., Moury, B., 2012. Modelling the evolutionary dynamics of viruses within their hosts: a case study using high-throughput sequencing. PLoS Pathogens 8: e1002654.

Fargette, D., Pinel, A., Abubakar, Z., Traoré, O., Brugidou, C., *et al*., 2004. Inferring the evolutionary history of *Rice yellow mottle virus* from genomic, phylogenetic and phylogeographic studies. Journal of Virology 78, 3252–3261.

Fargette, D., Pinel-Galzi, A., Sérémé, D., Lacombe, S., Hébrard, E., Traoré, O., Konaté, G., 2008. Diversification of *Rice yellow mottle virus* and related viruses spans the history of agriculture from the neolithic to the present. PLoS Pathogens 4: e1000125.

Farzadfar, S., Tomitaka, Y., Ikematsu, M., Golnaraghi, A., Pourrahim, R., Ohshima, K., 2008. Molecular characterisation of *Turnip mosaic virus* isolates from Brassicaceae weeds. European Journal of Plant Pathology 124, 45–55.

Fauquet, C., Thouvenel, J.-C., 1977. Isolation of the rice yellow mottle virus in Ivory Coast. Plant Diseases Rep 61, 443–446.

Fellers, J.P., Collins, G.B., Hunt, A.G., 1998. The NIa-Proteinase of different plant potyviruses provides specific resistance to viral infection. Crop Science 38, 1309–1319.

Felsenstein, J., 1993. PHYLIP (Phylogeny interference package), version 3.5c. Department of Genetics, University of Washington, Seattle.

Fernandez, A., Laín, S., García, J.A., 1995. RNA helicase activity of the plum pox potyvirus CI protein expressed in *Escherichia coli*. Mapping of an RNA binding domain. Nucleic Acids Research 23, 1327–1332.

Fernandez, A., García, J.A., 1996. The RNA helicase CI from plum pox potyvirus has two regions involved in binding to RNA. FEBS Lett 388, 206–210.

Fernandez, A., Guo, H.S., Sáenz, P., Simón-Buela. L., Gómez de Cedrón, M., García, J.A., 1997. The motif V of plum pox potyvirus CI RNA helicase is involved in NTP hydrolysis and is essential for virus RNA replication. Nucleic Acids Research 25, 4474–4480.

Fouchier, R.A.M., Schneeberger, P.M., Rozendaal, F.W., Broekman, J.M., Kemink, S.A., Munster, V., Kuiken, T., Rimmelzwaan, G.F., Schutten, M., Van Doornum, G.J., Koch, G., Bosman, A., Koopmans, M., Osterhaus, A.D., 2004. Avian influenza A virus (H7N7) associated with human conjunctivitis and a fatal case of acute respiratory distress syndrome. Proceedings of the National Academy of Sciences 101, 1356–1361.

Francki, R.I.B., Milne, R.G., Hatta, T., 1985. Atlas of Plant Viruses. Boca Raton: CRC Press. pp. 153-169.

Fu, Y.X., Li, W.H., 1993. Statistical tests of neutrality of mutations. Genetics 133, 693–709.

Fuji, S., Nakamae, H., 1999. Complete nucleotide sequence of the genomic RNA of a Japanese yam mosaic virus, a new potyvirus in Japan. Archives of Virology 144, 231–240.

Fuji, S., Nakamae, H., 2000. Complete nucleotide sequence of the genomic RNA of a mild strain of Japanese yam mosaic potyvirus. Archives of Virology 145, 635–640.

Gallo, R.C., Sarin, P.S., Gelmann, E.P., Robert-Guroff, M., Richardson, E., Kalyanaraman, V.S., Mann, D., Sidhu, G.D., Stahl, R.E., Zolla-Pazner, S., Leibowitch, J., Popovic, M., 1983. Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). Science 220, 865–867.

Gao, F., Bailes, E., Robertson, D.L., Chen, Y., Rodenburg, C.M., Michael, S.F., Cummins, L.B., Arthur, L.O., Peeters, M., Shaw, G.M., Sharp, P.M., Hahn, B.H., 1999. Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. Nature 397, 436–441.

García-Arenal, F., Fraile, A., Malpica, J.M., 2001. Variability and genetic structure of plant virus populations. Annual Review of Phytopathology 39, 157–186.

Gardner, M.W., Kendrick, J.B., 1921. Turnip mosaic. Journal of Agricultural Research 22, 123–124.

Gibbs, A., Mackenzie, A., Blanchfield, A., Cross, P., Wilson, C.R., Kitajima, E., Nightingale, M., Clements, M., 2000. Viruses of orchids in Australia; their identification, biology and control. Australian Orchid Review 65, 10–21.

Gibbs, A.J., Mackenzie, A.M., Wei, K.J., Gibbs, M.J., 2008a. The potyviruses of Australia. Archives of Virology 153, 1411–1420.

Gibbs, A.J., Ohshima, K., Gibbs, M., García-Arenal, F., 2008b. More about plant virus evolution; past, present and future. In Origin and Evolution of Viruses, 2nd ed, pp. 229–249. Edited by E. Domingo, C. R. Parrish & J. J. Holland. Elsevier Academic Press.

Gibbs, A.J., Ohshima, K., Phillips, M.J., Gibbs, M.J., 2008c. The prehistory of potyviruses: their initial radiation was during the dawn of agriculture. PLoS ONE 3: e2523.

Gibbs, A.J., Trueman, J.W.H., Gibbs, M.J., 2008d. The bean common mosaic virus lineage of potyviruses; where did it arise and when? Archives of Virology 153, 2177–2187.

Gibbs, A.J., Fargette, D., García-Arenal, F., Gibbs, M.J., 2010. Time – the emerging dimension of plant virus studies. Journal of General Virology 91, 13–22.

Gibbs, A.J., Ohshima, K., 2010. Potyviruses and the digital revolution. Annual Review of Phytopathology 48, 205–223.

Gibbs, M.J., Armstrong, J.S., Gibbs, A.J., 2000. Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. Bioinformatics 16, 573–582.

Gifford, R.J., Katzourakis, A., Tristem, M., Pybus, O.G., Winters, M., Shafer, R.W., 2008. A transitional endogenous lentivirus from the genome of a basal primate and implications for lentivirus evolution. Proceedings of the National Academy of Sciences of the United States of America 105, 20362–20367.

Gilbert., M.T., Rambaut, A., Wlasiuk, G., Spira, T.J., Pitchenik, A.E., Worobey, M., 2007. The emergence of HIV/AIDS in the Americas and beyond. Proceedings of the National Academy of Sciences of the United States of America 104, 18566–18570.

Gilbert, C., Maxfield, D., Goodman, S., Feschotte, C., Malik, H.S., 2009. Parallel Germline Infiltration of a Lentivirus in Two Malagasy Lemurs. PLoS Genetics 5: e1000425.

Glasa, M., Candresse, T., 2005. Partial sequence analysis of an atypical Turkish isolate provides further information on the evolutionary history of *Plum pox virus* (PPV). Virus Research 108, 199–206.

Gong, W., Jiang, Y., Za, Y., Zeng, Z., Shao, M., Fan, J., Sun, Y., Xiong, Z., Yu, X., Tu, C., 2010. Temporal and spatial dynamics of rabies viruses in China and Southeast Asia. Virus Research 150, 111–118.

Gorman, O., Donis, R., Kawaoka, Y., Webster, R., 1990. Evolution of influenza A virus PB2 genes: implications for evolution of the ribonucleoprotein complex and origin of human influenza A virus. Journal of Virology 64, 4893–4902.

Gould, A.R., Hyatt, A.D., Lunt, R., Kattenbelt, J.A., Hengstberger, S., Blacksell, S.D., 1998. Characterisation of a novel lyssavirus isolated from Pteropid bats in Australia. Virus Research 54, 165–187.

Grenfell, B.T., Pybus, O.G., Gog, J.R., Wood, J.L., Daly, J.M., Mumford, J.A., Holmes, E.C., 2004. Unifying the epidemiological and evolutionary dynamics of pathogens. Science 303, 327–332.

Grzela, R., Szolajska, E., Ebel, C., Madern, D., Favier, A., Wojtal, I., Zagorski, W., Chroboczek, J., 2008. Virulence factor of *Potato virus Y*, genome-attached terminal protein VPg, is a highly disordered protein. Journal of Biological Chemistry 283, 213–221.

Guindon, S., Gascuel, O., 2003. A simple fast and accurate algorithm to estimate large phylogenies by miximum likelihood. Systematic Biology 52, 696–704.

Ha, C., Revill, P., Harding, R.M., Vu, M., Dale, J.L., 2008. Identification and sequence analysis of potyviruses infecting crops in Vietnam. Archives Virology 153, 45–60.

Hahn, B.H., Shaw, G.M., Arya, S.K., Popovic, M., Gallo, R.C., WongStaal, F., 1984 Molecular cloning and characterization of the HTLV-III virus associated with AIDS. Nature 312, 166–169.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41, 95–98.

Harlan, J.R., 1998. Distributions of agricultural origins: A global perspective. In: Damania AB, Valkoun J, Willcox G, Qualset CO, editors. Origins of Agriculture and Crop Domestication. ICARDA, Aleppo, Syria. pp. 1–2.

Harper, S.J., 2013. *Citrus tristeza virus*: Evolution of complex and varied genotypic groups. Frontiers in Microbiology 4:93. doi: 10.3389/fmicb.2013.00093.

Hassan, M.M., Li, D., El-Deeb, A.S., Wolff, R.A., Bondy, M.L., Davila, M., Abbruzzese, J.L., 2008. Association between hepatitis B virus and pancreatic cancer. Journal of Clinical Oncology 26, 4557–4562.

Hay, A.J, Gregory, V., Douglas, A.R., Lin, Y.P., 2001. The evolution of human influenza viruses. Philosophical Transactions of the Royal Society B 356, 1861–1870.

Hemingway, J.S., 1995. Mustards, *Brassica* spp. and *Sinapis alba* (Cruciferae). In: Smartt J, Simmonds NW, editors. Evolution of Crop Plants. 2nd ed. UK, Harlow: Longman scientific & technical. pp. 82–86.

Hey, J., Harris, E., 1999. Population bottlenecks and patterns of human polymorphism. Molecular Biology and Evolution 16, 1423–1426.

Hirsch, V.M., Olmsted, R.A., Murphey-Corb, M., Purcell, R.H., Johnson, P.R., 1989. An African primate lentivirus (SIVsm) closely related to HIV-2. Nature 339, 389–392.

Ho, S.Y.W., Lanfear, R., Bromham, L., Phillips, M.J., Soubrier, J., Rodrigo, A.G., Cooper, A., 2011. Time-dependent rates of molecular evolution. Molecular Ecology 20, 3087–3101.

Ho, S.Y.W., Lanfear, R., Phillips, M.J., Barnes, I., Thomas, J.A., Kolokotronis, S.O., Shapiro, B., 2011. Bayesian estimation of substitution rates from ancient DNA sequences with low information content. Systematic Biology 60, 366–375.

Hodgkin, T., 1995. Cabbages, kales, etc. *Brassica oleracea* (Cruciferae). In: Smartt J, Simmonds NW, editors. Evolution of Crop Plants. 2nd ed. UK, Harlow: Longman scientific & technical. pp. 76–82.

Holmes, E.C., Woelk, C.H., Kassis, R., Bourhy, H., 2002. Genetic constraints and the adaptive evolution of rabies virus in nature. Virology 292, 247–257.

Holmes, E.C., 2003. Molecular clocks and the puzzle of RNA virus origins. Journal of Virology 77, 3893–3897.

Holmes, E.C., 2004. The phylogeography of human viruses. Molecular Ecology 13, 745–756.

Hong, X.Y., Chen, J., Shi, Y.H., Chen, J.P., 2007. The '6K1' protein of a strain of *Soybean mosaic virus* localizes to the cell periphery. Archives of Virology 152, 1547–1551.

Hong, Y., Hun, A.G., 1996. RNA polymerase activity catalyzed by a potyvirus-encoded RNA-dependent RNA polymerase. Virology 226, 146–151.

Hudson, R.R., 2000. A new statistic for detecting genetic differentiation. Genetics 155, 2011–2014.

Hull, R., 1988. The Sobemovirus Group. In: Plant Viruses, Polyhedral Virions with Monopartite Genomes, Koenig, R., ed. Plenum Press, New York, pp. 113–146.

Hunt, R., 2007. Hepatitis viruses. University of Southern California, Department of Pathology and Microbiology. Retrieved 2008-03-13.

Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution 23, 254–267.

Jenner, C.E., Sanchez, F., Nettleship, S.B., Foster, G.D., Ponz, F., Walsh, J.A., 2000. The cylindrical inclusion gene of *Turnip mosaic virus* encodes a pathogenic determinant to the brassica resistance gene *TuRB01*. Molecular Plant-Microbe Interactions 13, 1102–1108.

Jenner, C.E., Tomimura, K., Ohshima, K., Hughes, S.L., Walsh, J.A., 2002a. Mutations in *Turnip mosaic virus* P3 and cylindrical inclusion proteins are separately required to overcome two *Brassica napus* resistance genes. Virology 300, 50–59.

Jenner, C.E., Wang, X., Tomimura, K., Ohshima, K., Ponz, F., Walsh, J.A., 2002b. The dual role of the potyvirus P3 protein of *Turnip mosaic virus* as a symptom and avirulence determinant in brassicas. Molecular Plant-Microbe Interactions 16, 777–784.

Joannon, B., Lavigne, C., Lecoq, H., Desbiez, C., 2010. Barriers to gene flow between emerging populations of *Watermelon mosaic virus* in southeastern France. Phytopathology 100, 1373–1379.

Kadare, G., Haenni, A.L., 1997. Virus-encoded RNA helicases. Journal of Virology 71, 2583–2590.

Kang, B.-C., Yeam, I., Jahn, M.M., 2005. Genetics of plant virus resistance. Annual Review of Phytopathology 43, 581–621.

Karasev, A.V., Boyko, V.P., Gowda, S., Nikolaeva, O.V., Hilf, M.E., Koonin, E.V., Niblett, C.L., Cline, K., Gumpf, D.J., Lee, R.F., Garnsey, S.M., Lewandowski, D.J., Dawson, W.O., 1995. Complete sequence of Citrus tristeza virus RNA genome. Virology 208, 511–520.

Karasev, A.V., Hu, X., Brown, C.J., Kerlan, C., Nikolaeva, O.V., Crosslin, J.M., Gray, S.M., 2011. Genetic diversity of the ordinary strain of *Potato virus Y* (PVY) and origin of recombinant PVY strains. Phytopathology 101, 778–785.

Kasschau, K.D., Carrington, J.C., 1998. A counter defensive strategy of plant viruses: suppression of posttranscriptional gene silencing. Cell 95, 461–70.

Kawaoka, Y. (ed.), 2006. Influenza Virology: Current Topics. Caister Academic Press. 368pp.

Kennedy, J.S., Day, M.F., Eastop, V.F., 1962. A Conspectus of aphids as vectors of plant viruses. Commonwealth Institute of Entomology. The Eastern Press Ltd., London and Reading. 114 pp.

Khan, M.A., Miyoshi, H., Gallie, D.R., Goss, D.J., 2008. Potyvirus genome-linked protein, VPg, directly affects wheat germ in vitro translation. Journal of Biological Chemistry 283, 1340–1349.

Kheyr-Pour, A., Bendahmane, M., Matzeit, N., Accotto, G.P., Crespi, S., Gronemborn, B., 1991. Tomato yellow leaf curl virus from Sardinia is a whitefly-transmitted geminivirus. Nucleic Acids Research 19, 6763–6769.

Kim, Y.J., Jonson, M.G., Choi, H.S., Ko, S.J., Kim, K.H., 2009. Molecular characterization of Korean *Pepper mottle virus* isolates and its relationship to symptom variations. Virus Research 144, 83–88.

Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution 16, 111–120.

King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J., 2012. Virus taxonomy: Classification and nomenclature of viruses. Ninth report of the International Committee on Taxonomy of Viruses. San Diego: Elsevier/Academic Press. 1327 pp.

Kissi, B., Tordo, N., Bourhy, H., 1995. Genetic polymorphism in the rabies virus nucleoprotein gene. Virology 209, 526–537.

Korkmaz, S., Tomitaka, Y., Onder, S., Ohshima, K., 2008. Occurrence and molecular characterization of Turkish isolates of *Turnip mosaic virus*. Plant Pathology 57, 1155–1162.

Kozubek, E., Irzykowski, W., Lehmann, P., 2007. Genetic and molecular variability of a *Turnip mosaic virus* population from horseradish (*Cochlearia armoracia* L.). Journal of Applied Genetics 48, 295–306.

Kuzmina, N.A., Kuzmin, I.V., Ellison, J.A., Taylor, S.T., Bergman, D.L., Dew, B., Rupprecht, C.E., 2013. A reassessment of the evolutionary timescale of bat rabies viruses based upon glycoprotein gene sequences. Virus Genes. doi: 10.1007/s11262-013-0952-9.

Kuzmin, I.V., Hughes, G.J., Botvinkin, A.D., Orciari, L.A., Rupprecht, C.E., 2005 Phylogenetic relationships of Irkut and West Caucasian bat viruses within the *Lyssavirus* genus and

suggested quantitative criteria based on the N gene sequence for lyssavirus genotype definition. Virus Research 111, 28–43.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. CLUSTAL W and CLUSTAL X version 2.0. Bioinformatics 23, 2947–2948.

Lbida B., Fonseca F., Santos C., Zemzami M., Bennani A., Nolasco G. (2004). Genomic variability of *Citrus tristeza virus* (CTV) isolates introduced into Morocco. Phytopathologia Mediterranea 43, 205–210.

Lecoq, H., Wipf-Scheibel, C., Chandeysson, C., Lê Van, A., Fabre, F., Desbiez, C., 2009. Molecular epidemiology of *Zucchini yellow mosaic virus* in France: an historical overview. Virus Research 141, 190–200.

Leeper, S., Reddi, A., 2010. United States global health policy: HIV/AIDS, maternal and child health, and The President's Emergency Plan for AIDS Relief (PEPFAR). AIDS 24, 2145–2149.

Lefeuvre, P., Delatte, H., Naze, F., Dogley, W., Reynaud, B., Lett, J.M., 2007. A new tomato leaf curl virus from the Seychelles archipelago. Plant Pathology 56, 342.

Lefeuvre, P., Martin, D.P., Harkins, G., Lemey, P., Gray, A.J.A., Meredith, S., Lakay, F., Monjane, A., Lett, J.M., Varsani, A., Heydarnejad, J., 2010. The spread of tomato yellow leaf curl virus from the Middle East to the World. PLoS Pathogens 6: e1001164.

Lemey, P., Pybus, O.G., Wang, B., Saksena, N.K., Salemi, M., Vandamme, A.M., 2003. Tracing the origin and history of the HIV-2 epidemic. Proceedings of the National Academy of Sciences 100, 6588–6592.

Leonard, S., Plante, D., Wittmann, S., Daigneault, N., Fortin, M.G., Laliberte, J.-F., 2000. Complex formation between potyvirus VPg and translation eukaryotic initiation factor 4E correlates with virus infectivity. Journal of Virology 74, 7730–7737.

Leonard, S., Viel, C., Beauchemin, C., Daigneault, N., Fortin, M.G., Laliberte, J.F., 2004. Interaction of VPg-Pro of *Turnip mosaic virus* with the translation initiation factor 4E and the poly(A)-binding protein in planta. Journal of General Virology 85, 1055–1063.

Lesemann, D.E., Vetten, H.J., 1985. The occurrence of tobacco rattle and turnip mosaic viruses in *Orchis* ssp., and of an unidentified potyvirus in *Cypripedium calceolus*. Acta Horticulturae 164, 45–54.

Li, H.Y., Roossinck, M.J., 2004. Genetic bottlenecks reduce population variation in an experimental RNA virus population. Journal of Virology 78, 10582–10587.

Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25, 1451–1452.

Lole, K.S., Bollinger, R.C., Paranjape, R.S., Gadkari, D., Kulkarni, S.S., Novak, N.G., Ingersoll, R., Sheppard, H.W., Ray, S.C., 1999. Full-length human immuno deficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. Journal of Virology 73, 152–160.

López-Moya, J.J., García, J.A., 1999. Potyviruses (*Potyviridae*). In Encyclopedia of Virology, 2nd edn. Edited by Granoff, A., Webster, R.G. San Diego: Academic Press. pp. 1369–1375.

Luo, A., Qiao, H., Zhang, Y., Shi, W., Ho, S.Y., Xu W, Zhang, A., Zhu, C., 2010. Performance of criteria for selecting evolutionary models in phylogenetics: a comprehensive study based on simulated datasets. BMC Evolution Biology 10: 242. doi:10.1186/1471-2148-10-242.

MacNaughton, I.H., 1995. Turnip and relatives, *Brassica campestris* (Cruciferae). In: Smartt, J., Simmonds, N.W. (eds). Evolution of Crop Plants. 2nd ed. UK, Harlow: Longman scientific & technical. pp. 62–68.

MacNaughton, I.H., 1995. Swedes and rapes, *Brassica napus* (Cruciferae). In: Smartt, J., Simmonds, N.W. (eds). Evolution of Crop Plants. 2nd ed. UK, Harlow: Longman scientific & technical. pp. 68–75.

Magiorkinis, G., Magiorkinis, E., Paraskevis, D., Ho, S.Y., Shapiro, B., Pybus, O.G., Allain, J.P., Hatzakis, A., 2009. The global spread of hepatitis C virus 1a and 1b: a phylodynamic and phylogeographic analysis. PLoS Medicine 6: e1000198.

Martin, D., Rybicki, E., 2000. RDP: detection of recombination amongst aligned sequences. Bioinformatics 16, 562–563.

Martin, D.P., Lemey, P., Lott, M., Moulton, V., Posada, D., Lefeuvre, P., 2010. RDP3: a flexible and fast computer program for analyzing recombination. Bioinformatics 26, 2462–2463.

Martín, S., Sambade, A., Rubio, L., Vives, M.C., Moya, P., Guerri, J., Elena, S.F., Moreno, P., 2009. Contribution of recombination and selection to molecular evolution of *Citrus tristeza virus*. Journal of General Virology 90, 1527–1538.

Matthews, R.E.F., 1982. Classification and nomenclature of viruses. Fourth report of the International Committee on Taxonomy of Viruses. Intervirology 17, 1–199.

Maynard-Smith, J., 1992. Analyzing the mosaic structure of genes. Journal of Molecular Evolution 34, 126–129.

McElhinney, L.M., Marston, D., Johnson, N., Black, C., Matouch, O., Lalosevic, D., Stankov, S., Must, K., Smreczak, M., et al., 2006. Molecular epidemiology of rabies viruses in Europe. Developmental Biology 125, 17–28.

Moreno, I.M., Malpica, J.M., Diaz-Pendon, J.A., Moriones, E., Fraile, A., García-Arenal, F., 2004. Variability and genetic structure of the population of watermelon mosaic virus infecting melon in Spain. Virology 318, 451–460.

Moreno, P., Ambrós, S., Albiach-Martí, M.R., Guerri, J., Peña, L., 2008. *Citrus tristeza virus*: a pathogen that changed the course of the citrus industry. Molecular Plant Pathology 9, 251–68.

Moriones, E., Navas-Castillo, J., 2000. Tomato yellow leaf curl virus, an emerging virus complex causing epidemics worldwide. Virus Research 71, 123–134.

Moury, B., Morel, C., Johansen, E., Jacquemond, M., 2002. Evidence for diversifying selection in *Potato virus Y* and in the coat protein of other potyviruses. Journal of General Virology 83, 2563–2573.

Moury, B., 2010. A new lineage sheds light on the evolutionary history of *Potato virus Y*. Molecular Plant Pathology 11, 161–168.

Moury, B., Simon, V., 2011. dN/dS-based methods detect positive selection linked to trade-offs between different fitness traits in the coat protein of *Potato virus Y*. Molecular Biology and Evolution 28, 2707–2717.

Nadin-Davis, S.A., Real, L.A., 2011. Molecular phylogenetics of the lyssaviruses - insights from a coalescent approach. Advances in Virus Research 79, 203–238.

Nakano, T., Lau, G.M.G., Lau, G.M.L., Sugiyama, M., Mizokami, M., 2011. An updated analysis of hepatitis C virus genotypes and subtypes based on the complete coding region. Liver International 32, 339–345.

Navot, N., Pichersky, E., Zeidan, M., Zamir, D., Czosnek, H., 1991. Tomato yellow leaf curl virus: a whitefly-transmitted geminivirus with a single genomic component. Virology 185, 151–161.

Nelson, M., Spiro, D., Wentworth, D., Beck, E., Fan, J., Ghedin, E., Halpin, R., Bera, J., Hine, E., Proudfoot, K., Stockwell, T., Lin, X., Griesemer, S., Kumar, S., Bose, M., Viboud, C., Holmes, E., Henrickson, K., 2009. The early diversification of influenza A/H1N1pdm. PLoS Currents 3;1:RRN1126.

Ngon A Yassi, M., Ritzenthaler, C., Brugidou, C., Fauquet, C., Beachy, R.N., 1994. Nucleotide sequence and genome characterization of rice yellow mottle virus RNA. Journal of General Virology 75, 249–257.

Nguyen, H. D., Tomitaka, Y., Ho, S. Y., Duchêne, S., Vetten, H. J., Lesemann, D., Walsh, J. A., Gibbs, A. J., Ohshima, K., 2013. Turnip mosaic potyvirus probably first spread to Eurasian brassica crops from wild orchids about 1000 years ago. PLoS One 8: e55336. doi: 10.1371/journal.pone.0055336.

Nguyen, H. D., Tran, H.T., Ohshima, K., 2013. Genetic variation of the *Turnip mosaic virus* population of Vietnam: a case study of founder, regional and local influences. Virus Research 171, 138–149.

Nicolas, O., Laliberté, J.F., 1992. The complete nucleotide sequence of turnip mosaic potyvirus RNA. Journal of General Virology 73, 2785–2793.

Ogawa, T., Tomitaka, Y., Nakagawa, A., Ohshima, K., 2008. Genetic structure of a population of *Potato virus Y* inducing potato tuber necrotic ring spot disease in Japan; comparison with North American and European populations. Virus Research 131, 199–212.

Ogawa, T., Nakagawa, A., Hataya, T., Ohshima, K., 2012. The genetic structure of populations of *Potato virus Y* in Japan; Based on the analysis of 20 full genomic sequences. Journal of Phytopathology 160, 661–673.

Ohno, O., Mizokami, M., Wu, R.R., Saleh, M.G., Ohba, K., Orito, E., Mukaide, M., Williams, R., Lau, J.Y., 2007. New hepatitis C virus (HCV) genotyping system that allows for identification of HCV genotypes 1a, 1b, 2a, 2b, 3a, 3b, 4, 5a, and 6a. Journal of Clinical Microbiology 35, 201–207.

Ohshima, K., Tanaka, M., Sako, N., 1996. The complete nucleotide sequence of turnip mosaic virus RNA Japanese strain. Archives of Virology 141, 1991–1997.

Ohshima, K., Yamaguchi, Y., Hirota, R., Hamamoto, T., Tomimura, K., Tan, Z., Sano, T., Azuhata, F., Walsh, J.A., Fletcher, J., Chen, J., Gera, A., Gibbs, A.J., 2002. Molecular

evolution of *Turnip mosaic virus*: evidence of host adaptation, genetic recombination and geographical spread. Journal of General Virology 83, 1511–1521.

Ohshima, K., Tomitaka, Y., Wood, J.T., Minematsu, Y., Kajiyama, H., Tomimura, K., Gibbs, A.J. 2007. Patterns of recombination in turnip mosaic virus genomic sequences indicate hotspots of recombination. Journal of General Virology, 88, 298–315.

Ohshima, K., Akaishi, S., Kajiyama, H., Koga, R., Gibbs, A.J., 2010. Evolutionary trajectory of turnip mosaic virus populations adapting to a new host. Journal of General Virology 91, 788–801.

Osiowy, C., Giles, E., Tanaka, Y., Mizokami, M., Minuk, G.Y., 2006. Molecular evolution of hepatitis B virus over 25 years. Joural of Virology 80, 10307–10314.

Pagán, I., Firth, C., Holmes, E.C., 2010. Phylogenetic analysis reveal rapid evolutionary dynamics in the plant RNA virus genus *Tobamovirus*. Journal of Molecular Evolution 71, 298–307.

Pagán, I., Holmes, E.C., 2010. Long-term evolution of the *Luteoviridae*: time scale and mode of virus speciation. Journal of Virology 84, 6177–6187.

Page, R.D.M., 1996. Treeview: an application to display phylogenetic trees on personal computer. Computer Applications in the Biosciences 12, 357–358.

Pallett, D.W., Cooper, J.I., Wang, H., Reeves, J., Luo, Z., Machado, R., Obermeier, C., Walsh, J. A., Kearsey, M. J., 2007. Variation in the pathogenicity of two *Turnip mosaic virus* isolates in wild *Brassica rapa* provenances. Plant Pathology 57, 401–407.

Paraskevis, D., Magiorkinis, G., Magiorkinis, E., Ho, S.Y., Belshaw, R., Allain, J.P., Hatzakis, A., 2013. Dating the origin and dispersal of hepatitis B virus infection in humans and primates. Hepatology 57, 908-916.

Parrella, G., Lanave, C., 2009. Identification of a new pathotype of *Bean yellow mosaic virus* (BYMV) infecting blue passion flower and some evolutionary characteristics of BYMV. Archives of Virology 154, 1689–1694.

Peeters, M., Sharp, P.M., 2000. Genetic diversity of HTV-1: the moving target. AIDS 14, 129–140.

Pico, B., Diez, M.-J. & Nuez, F., 1996. Viral diseases causing the greatest economic losses to the tomato crop. II. The Tomato yellow leaf curl virus - a review. Scientia Horticulturae 67, 151–196.

Pinhasi, R., Fort, J., Ammerman, A.J., 2005. Tracing the origin and spread of agriculture in Europe. PLoS Biology 3: e410.

Popovic, M., Sarngadharan, M.G., Read, E., Gallo, R.C., 1984. Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. Science 224, 497–500.

Posada, D., 2008. jModelTest: phylogenetic model averaging. Molecular Biology and Evolution 25, 1253–1256.

Posada, D., Crandall, K.A., 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proceedings of the National Academy of Sciences of the United States of America 98, 13757–13762.

Pybus, O.G., Rambaut, A., 2009. Evolutionary analysis of the dynamics of viral infectious disease. Nature Reviews Genetics 10, 540–550.

Rakotomalala, M., Pinel-Galzi, A., Mpunami, A., Randrianasolo, A., Ramavovololona, P., Rabenantoandro, Y., Fargette, D., 2013. *Rice yellow mottle virus* in Madagascar and in the Zanzibar Archipelago; island systems and evolutionary time scale to study virus emergence. Virus Research 171, 71–79.

Reeves, J.D., Doms, R.W., 2002. Human Immunodeficiency Virus Type 2. Journal of General Virology 83, 1253–65.

Reid, A., Fanning, T., Hultin, J., Taubenberger, J., 1999. Origin and evolution of the 1918 "Spanish" influenza virus hemagglutinin gene". Proceedings of the National Academy of Sciences USA 96, 1651–1656.

Riechmann, J.L., Lain, S., Garcia, J.A., 1992. Highlights and prospects of potyvirus molecular biology. Journal of General Virology 73, 1–16.

Rojas, M.R., Zerbini, F.M., Allison, R.F., Gilbertson, R.L., Lucas, W.J., 1997. Capsid protein and helper component-proteinase function as potyvirus cell-to-cell movement proteins. Virology 237, 283–295.

Roossinck, M.J., Schneider, W.L., 2006. Mutant clouds and occupation of sequence space in plant RNA viruses. In: Dominogo, E. (ed.), Quasispecies: Concepts and Implications for Virology. Springer, 337–348.

Saenz, P., Cervera, M.T., Dallot, S., Quiot, L., Quiot, J.-B., Riechmann, J.L., Garcia, J.A., 2000. Identification of a pathogenicity determinant of *Plum pox virus* in the sequence

encoding the C-terminal region of protein P3+6K1. Journal of General Virology 81, 557–566.

Salemi, M., Vandamme, A.M., 2002. Hepatitis C virus evolutionary patterns studied through analysis of full-genome sequences. Journal of Molecular Evolution 54, 62–70.

Salminen, M.O., Carr, J.K., Burke, D.S., McCutchan, F.E., 1995. Identification of breakpoints in intergenotypic recombinants of HIV type 1 by Bootscanning. AIDS Research and Human Retroviruses 11, 1423–1425.

Sanjuán, R., 2012. From molecular genetics to phylodynamics: evolutionary relevance of mutation rates across viruses. PLoS Pathogens 8: e1002685.

Santiago, M.L., Range, F., Keele, B.F., Li, Y., Bailes, E., Bibollet-Ruche, F., Fruteau, C., Noë, R., Peeters, M., Brookfield, J.F., Shaw, G.M., Sharp, P.M., Hahn, B.H., 2005. Simian immunodeficiency virus infection in free-ranging sooty mangabeys (*Cercocebus atys atys*) from the Tai Forest, Cote d'Ivoire: implications for the origin of epidemic human immunodeficiency virus type 2. Journal of Virology 79, 12515–12527.

Sarngadharan, M.G., Popovic, M., Bruch, L., Schüpbach, J., Gallo, R.C., 1984. Antibodies reactive with human T-lymphotropic retroviruses (HTLV-III) in the serum of patients with AIDS. Science 224, 506–508.

Sarwar, M.T., Kausar, H., Ijaz, B., Ahmad, W., Ansar, M., Sumrin, A., Ashfaq, U.A., Asad, S., Gull, S., Shahid, I., Hassan, S., 2011. NS4A protein as a marker of HCV history suggests that different HCV genotypes originally evolved from genotype 1b. Journal of Virology 8: 317.

Sawyer, S.A., 1999. GENECONV: a computer package for the statistical detection of gene conversion. Distributed by the Author. Department of Mathematics. Washington University, St Louis, available at http://www.math.wustl.edu/~
sawyer.

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. EMBO Journal 16, 4049–4059.

Schliep, K.P., 2011. Phangorn: phylogenetic analysis in R. Bioinformatics 27, 592–593.

Scholtissek, C., 1995. Molecular evolution of influenza viruses. Virus Genes 11, 209–215.

Schwarz, R., Dayhoff, M., 1979. Matrices for detecting distant relationships. In Dayhoff M, editor, *Atlas of protein sequences*. National Biomedical Research Foundation. pp. 353–358.

Seal, S.E., vandenBosch, F., Jeger, M.J., 2006. Factors influencing begomovirus evolution and their increasing global significance: implications for sustainable control. Critical Reviews in Plant Sciences 25, 23–46.

Sehgal, O.P., 1981. Southern bean mosaic virus group. Handbook of plant virus infection and diagnosis. Elsevier, Cold Spring Harbor, New York. pp. 91–121.

Seillier-Moiseiwitsch, F., Margolin, B.H., Swanstrom, R., 1994. Genetic variability of the human immunodeficiency virus: statistical and biological issues. Annual Review of Genetics 28, 559–596.

Sentandreu, V., Castro, J.A., Ayllón, M.A., Rubio, L., Guerri, J., González-Candelas, F., Moreno, P., Moya, A., 2006. Evolutionary analysis of genetic variation observed in citrus tristeza virus (CTV) after host passage. Archives of Virology 151, 875–894.

Seo, J.K., Ohshima, K., Lee, H.G., So, M., Choi, H.S., Lee, S.H., Sohn, S.H., Kim, K.H., 2009. Molecular variability and genetic structure of the population of *Soybean mosaic virus* based on the analysis of complete genome sequences. Virology 393, 91–103.

Sepkowitz, K.A., 2001. AIDS-the first 20 years. The New England Journal of Medicine 344, 1764–1772.

Sharp, P.M., Bailes, E., Gao, F., Beer, B.E., Hirsch, V.M., Hahn, B.H., 2000. Origins and evolution of AIDS viruses: estimating the time-scale. Biochemical Society transactions 28, 275–282.

Shepherd, R.J., Pound, G.S., 1960. Purification of turnip mosaic virus. Phytopathology 50, 797–803.

Shukla, D.D., Ward, C.W., Brunt, A.A., Berger, P.H., 1998. Potyviridae family. In “Description of Plant Viruses, No. 366”. Association of Applied Biologists (AAB).

Silva, G., Marques, N., Nolasco, G., 2012. The evolutionary rate of citrus tristeza virus ranks among the rates of the slowest RNA viruses. Journal of General Virology 93, 419–429.

Simmonds, N.M., ed., 1976. Evolution of Crop Plants. London: Longman. 339 pp.

Simmonds. P., Smith, D.B., 1977. Investigation of the pattern of diversity of hepatitis C virus in relation to times of transmission. Journal of Viral Hepatistic 4, 69–74.

Simmonds, P., Smith, D.B., 1997. Investigation of the pattern of diversity of hepatitis C virus in relation to times of transmission. Journal of Viral Hepatitis 4, 69–74.

Simmons, H.E., Holmes, E.C., Stephenson, A.G., 2008. Rapid evolutionary dynamics of zucchini yellow mosaic virus. Journal of General Virology 89, 1081–1085.

Smarrt, J., Simmonds, N.M., eds., 1995. Evolution of Crop Plants. 2nd Edition. London: Longman. 340 pp.

Spetz, C., Valkonen, J.P.T., 2004. Potyviral 6K2 protein long distance movement and symptom-induction functions are independent and host-specific. Molecular Plant-Microbe Interactions 17, 502–510.

Steele, J.H., Fernandez, P.J., 1991. History of rabies and global aspects. In The Natural History of Rabies, 2nd edn. Edited by G. M. Baer. Boca Raton, USA: CRC Press. pp. 1–26.

Suehiro, N., Natsuaki, T., Watanabe, T., Okuda, S., 2004. An important determinant of the ability of *Turnip mosaic virus* to infect *Brassica* spp. and/or *Raphanus sativus* is in its P3 protein. Journal of General Virology 85, 2087–2098.

Sun, H., ShenTu, S., Xue, F., Duns, G., Chen, J., 2008. Molecular characterization and evolutionary analysis of soybean mosaic virus infecting Pinellia ternata in China. Virus Genes 36, 177–190.

Suzuki, Y., Nei, M., 2002. Origin and evolution of Influenza virus hemagglutinin genes. Molecular Biololy and Evolution 19, 501–509.

Sylvester, E.S., 1953. Host range and properties of *Brassica nigra* virus. Phytopathology 43, 541–546.

Sylvester, E.S, 1954. Aphid transmission of nonpersistent plant viruses with special reference to *Brassica nigra* virus. Hilgardia 23, 53–98.

Tajima, F., 1989. Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism. Genetics 123, 585–595.

Talbi, C., Holmes, E. C., de Benedictis, P., Faye, O., Nakoune, E., Gamatie, D., Diarra, A., Elmamy, B. O., Sow, A., Adjogoua, E. V., Sangare, O., Dundon, W. G., Capua, I., Sall, A. A. & Bourhy, H., 2009. Evolutionary history and dynamics of dog rabies virus in western and central Africa. Journal of General Virology 90, 783–791.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: Molecular Evolutionary genetics analysis using maximum likelihood, evolutionary

distance, and maximum parsimony methods. Molecular Biology and Evolution 28, 2731–2739.

Tan, Z., Wada, Y., Chen, J., Ohshima, K., 2004. Inter- and intralineage recombinants are common in natural populations of *Turnip mosaic virus*. Journal of General Virology 85, 2683–2696.

Tan, Z., Gibbs, A.J., Tomitaka, Y., Sánchez, F., Ponz, F., Ohshima, K., 2005. Mutations in *Turnip mosaic virus* genomes that have adapted to *Raphanus sativus*. Journal of General Virology 86, 501–510.

Theodorides, J., 1986. Histoire de la rage. Cave canem. Fondation Single Polignac Masson. 289 pp.

Tomimura, K., Gibbs, A.J., Jenner, C.E., Walsh, J.A., Ohshima, K., 2003. The phylogeny of *Turnip mosaic virus*; comparisons of 38 genomic sequences reveal a Eurasian origin and a recent 'emergence' in East Asia. Molecular Ecology 12, 2099–2111.

Tomimura, K., Spăk, J., Katis, N., Jenner, C.E., Walsh, J.A., Gibbs, A.J., Ohshima, K., 2004. Comparisons of the genetic structure of populations of *Turnip mosaic virus* in West and East Eurasia. Virology 330, 408–423.

Tomitaka, Y., Ohshima, K., 2006. A phylogeographic study of the *Turnip mosaic virus* population in East Asia reveals an ‘emergent’ lineage in Japan. Molecular Ecology 15, 4437–4457.

Tomitaka, Y., Yamashita, T., Ohshima, K., 2007. The genetic structure of populations of *Turnip mosaic virus* in Kyushu and central Honshu, Japan. Journal of General Plant Pathology 73, 197–208.

Tomlinson., J.A., Walkey, D.G.A., Hughes, D.E., Watson, D.H., 1965. Multiple transverse breakage of the filamentous particles of Turnip mosaic virus by ultrasonic vibration. Nature 207, 495–497.

Tomlinson, J.A., Walkey, D.G.A., 1967. Effects of ultrasonic treatment on turnip mosaic virus and potato virus X. Virology 32, 267–278.

Tomlinson, J.A., 1987. Epidemiology and control of virus diseases of vegetables. Annals of Applied Biology 110, 661–681.

Torres-Barceló, C., Daròs, J.A., Elena, S.F., 2009. Compensatory molecular evolution of HC-Pro, an RNA-silencing suppressor from a plant RNA virus. Molecular Biology and Evolution 27, 543–551.

Traore, O., Sorho, F., Pinel, A., Abubakar, Z., Banwo, O., Maley, J., Hebrard, E., Winter, S., Sere, Y., Konate, G., Fargette, D., 2005. Processes of diversification and dispersion of *Rice yellow mottle virus* inferred from large-scale and high-resolution phylogeographical studies. Molecular Ecology, 14, 2097–2110.

Tsompana, M., Abad, J., Purugganan, M., Moyer, J.W., 2005. The molecular population genetics of the *Tomato spotted wilt virus* (TSWV) genome. Molecular Ecology 14, 53–66.

Tugume, A.K., Cuéllar, W.J., Mukasa, S.B., Valkonen, J.P., 2010. Molecular genetic analysis of virus isolates from wild and cultivated plants demonstrates that East Africa is a hotspot for the evolution and diversification of sweet potato feathery mottle virus. Molecular Ecology 19, 3139–3156.

UNAIDS, 2010. "Global Report: Fact Sheet", "Sub-Saharan Africa: Fact Sheet" (23 November). Geneva: United Nations Progaram on HIV/AIDS. http://www.unaids.org/en/resources /presscentre/factsheets/.

Urcuqui-Inchima, S., Haenni, A.-L. & Bernardi, F., 2001. Potyvirus proteins: a wealth of functions. Virus Research 74, 157–175.

Valli, A., López-Moya, J.J., García, J.A., 2007. Recombination and gene duplication in the evolutionary diversification of P1 proteins in the family *Potyviridae*. Journal General Virology 88, 1016–1028.

Van Hemert, F.J., van de Klundert, M.A., Lukashov, V.V., Kootstra, N.A., Berkhout, B., Zaaijer, H.L., 2011. Protein X of hepatitis B virus: origin and structure similarity with the central domain of DNA glycosylase". PLoS ONE 6: e23392.

Vavilov, N., 1926. Origin of cultivated plants. Leningrad: Publication of the Bureau of Applied Botany. 248 pp.

Vogt, M.W., Hartshorn, K.L., Furman, P.A., Chou, T.C., Fyfe, J.A., Coleman, L.A., Crumpacker, C., Schooley, R.T., Hirsch, M.S., 1987. Ribavirin antagonizes the effect of azidothymidine on HIV replication. Science 235, 1376–1379.

Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S., Alizon, M., 1985. Nucleotide sequence of the AIDS virus, LAV. Cell 40, 9–17.

Walsh, J.A., Jenner, C.E., 2002. *Turnip mosaic virus* and the quest for durable resistance. Molecular Plant Pathology 3, 289–300.

Wang, R.Y., Powell, G., Hardie, J., Pirone, T.P., 1998. Role of the helper component in vector-specific transmission of potyviruses. Journal of General Virology 79, 1519–1524.

Wang, H.Y., Liu, J.L., Gao, R., Chen, J., Shao, Y.H., Li, X.D., 2009. Complete genomic sequence analyses of *Turnip mosaic virus* basal-BR isolates from China. Virus Genes 38, 421–428.

Ward, C.W., Weiller, G., Shukla, D.D., Gibbs, A.J., 1995. Molecular systematics of the Potyviridae, the largest plant virus family. In: Gibbs, A.J., Calisher, C.H., García-Arenal, H., editors. Molecular Basis of Virus Evolution. Cambridge: Cambridge University Press. pp. 477–500.

Wei, T., Huang, T.-S., McNeil, J., Laliberte, J.-F., Hong, J., Nelson, R.S., Wang, A., 2009. Sequential recruitment of the endoplasmic reticulum and chloroplasts for plant potyvirus replication. Journal of Virology 84, 799–809.

Weiller, G.F., 1998. Phylogenetic profiles: a graphical method for detecting genetic recombinations in homologous sequences. Molecular Biololy and Evolution 15, 326–355.

Weng, Z., Liu, X., Gowda, S., Barthelson, R.A., Galbraith, D.W., Dawson, W.O., Xiong, Z., 2010. Extreme genome stability of *Citrus tristeza virus*. In Abstracts of the XVIII Conference of the IOCV, 7–12 November 2010, Campinas, SP, Brazil, pp. 1–129.

Wertheim, J.O., Worobey, M., 2009. Dating the age of the SIV lineages that gave rise to HIV-1 and HIV-2. PLoS Computational Biology 5: e1000377.

Wikramaratna, P.S., Sandeman, M., Recker, M., Gupta, S., 2013. The antigenic evolution of influenza: drift or thrift? Philosophical Transactions of the Royal Society B 368: 20120200. http://dx.doi.org/10.1098/rstb.2012.0200.

Wittmann, S., Chatet, H., Fortin, M.G., Labiberte, J.I., 1997. Interaction of the viral protein genome linked of turnip mosaic potyvirus with the translation eukaryotic initiation factor (iso) 4E of Arabidopsis thaliana using the yeast two-hybrid system. Virology 234, 84–92.

Worobey, M., Gemmel, M., Teuwen, D.E., Haselkorn, T., Kunstman, K., Bunce, M., Muyembe, J.J., Kabongo, J.M., Kalengayi, R.M., Van Marck, E., Gilbert, M.T., Wolinsky, S.M., 2008. Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. Nature 455, 661–664.

Worobey, M., Telfer, P., Souquière, S., Hunter, M., Coleman, C.A., Metzger, M.J., Reed, P., Makuwa, M., Hearn, G., Honarvar, S., Roques, P., Apetrei, C., Kazanji, M., Marx, P.A., 2010. Island biogeography reveals the deep history of SIV. Science 329: 1487.

Yi, L., Zhou, C.-Y., Zhou, Y., Li, Z.-A., 2010. Genetic evolution analysis on wild Isolates of *Citrus tristeza virus* originated in China based on coat protein genes sequences. Agricultural Sciences in China 9, 1623–1629.

Yoshida, N., Shimura, H., Yamashita, K., Suzuki, M., Masuta, C., 2012. Variability in the P1 gene helps to refine phylogenetic relationships among leek yellow stripe virus isolates from garlic. Archives of Virology 157, 147–153.

Yu, F., Zhang, G., Xiao, S., Fang, L., Xu, G., Yan, J., Chen, H., Fu, Z.F., 2012. Complete genome sequence of a street rabies virus isolated from a rabid dog in China. Journal of Virology 86, 10890–10891.

Zhang, C.L., Gao, R., Wang, J., Zhang, G.M., Li, X.D., Liu, H.T., 2011. Molecular variability of *Tobacco vein banding mosaic virus* populations. Virus Research 158, 188–198.

Zhao, X., Tan, Z., Feng, H., Yang, R., Li, M., Jiang, J., Shen, G, Yu, R., 2011. Microsatellites in different Potyvirus genomes: survey and analysis. Gene 488, 52–56.

Zhou, Y., Holmes, E.C., 2007. Bayesian estimates of the evolutionary rate and age of hepatitis B virus. Joural of Molecular and Evolution 65, 197–205.