 doi:10.20944/preprints201705.0109.v1

The Genetic Structure of *Turnip mosaic virus* Population Reveals the Rapid Expansion of a New Emergent Lineage in China

Xiangdong Li[1*], Tiansheng Zhu[2*], Xiao Yin[1*], Chengling Zhang[4], Jia Chen[1] , Yanping Tian[1#], Jinliang Liu[3#]

[1] Laboratory of Plant Virology, Department of Plant Pathology, College of Plant Protection, Shandong Agricultural University, Tai'an, Shandong 271018, China;

[2] College of Plant Science and Technology, Tarimu University, Alar 843300, China;

[3] College of Plant Sciences, Jilin University, Changchun 130062, China;

[4] Xuzhou Sweet Potato Research Center of Jiangsu, Suzhou 221121, China.

[*] These authors contributed equally to this work.

[#] Correspondence: yanping.tian@sdau.edu.cn, jlliu@jlu.edu.cn; Tel: 86 538 8242523; Fax: 86 538 8226399

 doi:10.20944/preprints201705.0109.v1

Abstract: *Turnip mosaic virus* (TuMV) is one of the most widespread and economically important virus infecting both crop and ornamental species of the family *Brassicaceae*. TuMV isolates can be classified to five phylogenetic lineages, basal-B, basal-BR, Asian-BR, world-B and Orchis. To understand the genetic structure of TuMV from radish in China, the 3′-terminal genome of 90 TuMV isolates were determined and analyzed with other Chinese isolates available. The results showed that the Chinese TuMV isolates from radish formed three groups: Asian-BR, basal-BR and world-B. More than half of these isolates (52.54%) were clustered to basal-BR group, and could be further divided into three sub-groups. The TuMV basal-BR isolates in the sub-groups I and II were genetically homologous with Japanese ones, while those in sub-group III formed a distinct lineage. Sub-populations of TuMV basal-BR II and III were new emergent and in a state of expansion. The Chinese TuMV radish populations were under negative selection. Gene flow between TuMV populations from Tai'an, Weifang and Changchun was frequent.

Keywords: *Turnip mosaic virus*; *Potyvirus*; Genetic structure; Population; China

1. Introduction

Due to the error-prone nature of their RNA-dependent RNA polymerases, populations of plant RNA viruses are genetically heterogeneous and the genetic structure may change with time and environment[1, 2]. Studies of the genetic structure of viruses will provide information about the mechanisms and factors driving their evolution and help us to understand the molecular evolutionary history of viruses in relation to their dispersion and emergence of new epidemics[3].

Turnip mosaic virus (TuMV) is a species of the largest plant virus genus *Potyvirus* (family *Potyviridae*). TuMV has flexuous filamental particles of 700-750 nm long and can be transmitted by 40-50 species of aphids in a non-persistent manner [4, 5]. The TuMV genome consists of one single-stranded positive sense RNA molecule of approximately 9830 nucleotides (nt) and contains a large open reading frame (ORF)[6]. The ORF encodes a polyprotein that will be hydrolyzed to produce ten mature products[7, 8]. A frame-shift protein, P3N-PIPO, was reported to be involved in the pathogenesis and movement of TuMV[9, 10].

 doi:10.20944/preprints201705.0109.v1

TuMV can infect plants of 300 species in 43 families, and is probably the most widespread and economically important virus infecting both crop and ornamental species of family *Brassicaceae*[11, 12]. In an extensive survey conducted in 28 countries, TuMV ranked second for crop yield losses[4]. TuMV is a highly variable and has many biological and serological strains[13-16]. According to its host range, TuMV isolates can be classified to two pathotypes, B (mainly infects plants of the genus *Brassica*) and BR (infects plants of both *Brassica* and *Raphanus*). The brassica-infecting TuMV isolates were categorized into four phylogenetic lineages, basal-B, basal-BR, Asian-BR and world-B, which correlated well with their differences in pathogenicity and geographical origin[17]. Most recently, a monophyletic sister lineage called 'Orchis group' was detected from wild orchids-infecting TuMV isolates, which are more likely the ancestor of TuMV[18]. As in other potyviruses[19], recombination is a frequent event in the evolution of TuMV. Intra- and inter-lineage recombinants are common in natural populations of TuMV and can be detected throughout the genome[6, 20-22]. The Chinese and Japanese TuMV isolates are part of the same population but are a discrete lineage[22, 23]. The gene flow between sub-populations of TuMV from Vietnam, Japan and China are frequent[20]. The basal-BR isolates have occurred over the whole Japanese islands and have evolved into four sub-lineages[23-25].

Previous studies showed that the TuMV isolates of China can be clustered to world-B and Asian-BR groups[10, 17, 24, 26, 27]. However, we have detected the existence of basal-BR isolates in China and reported the complete genomic sequences of two basal-BR isolates that represented two novel recombination patterns[6, 28]. Here, we studied the genetic structure of TuMV population in China and found that the basal-BR group of TuMV was expanding in China.

2. Materials and Methods

2.1. Virus samples, RNA extraction and sequencing

Samples were collected from radish from Heilongjiang, Jilin and Shandong provinces from 2005 to 2010. All the samples were biologically cloned by three cycles of single lesion isolation in *Chenopodium amaranticolor a*nd propagated in *B. rapa*. Inoculated plants were maintained in a glasshouse at 25 ºC.

 doi:10.20944/preprints201705.0109.v1

Total RNAs were extracted from 100 mg TuMV-infected *B. rapa* leaves with the Invitrogen Trizol Kit following instructions of the manufacturer. The 3′-terminal 1.1 kb of TuMV genome were amplified with RT-PCR using primers CP-F (5′-ATC TTC GAA GAT TAC GAA GA-3′) and CP-R (5′-CCT TGC TTC CTA TCA AAT G-3′) [29]. The fragments were cloned into pMD18-T vector (TaKaRa Biotechnology Dalian Co, Ltd) and sequenced by a ABI PRISM^{TM} 377 DNA Sequencer. For each isolate, at least four clones from two separate PCR were sequenced. In case of any inconsistence, at least two more clones will be sequenced to obtain the consensus sequence.

2.2. Recombination analysis

The sequences of 90 TuMV isolates and other 28 obtained from the databank were subjected to recombination analyses using the software package RDP3, which assembled programs RDP[30], GENECONV[31], BOOTSCAN[32], MAXCHI[33], CHIMEARA[34] and SISCAN. The sequences were analyzed using the default settings for different detection programs and a Bonferroni-corrected *P*-value cut off of 0.05. The potential recombinants identified by the programs in RDP3 were re-checked using PHYLPRO[35]. The RDP, BOOTSCAN and SISCAN programs were based on phylogenetic methods, whereas GENECONV, MAXCHI and CHIMAERA programs were substitution methods, and the PHYLPRO program was a distance comparison method. Only those sequences with recombination supported by at least three programs or two kinds of methods and with *P*-value $<1.0\times10^{-6}$ were regarded as ‘clear’ recombinants; otherwise, they were called as ‘tentative’ recombinants[23, 25].

2.3. Phylogenetic analysis of the TuMV population

Sequence alignments were performed using the CLUSTAL W program (Thompson *et al*., 1994). Phylogenetic tree of TuMV isolates excluding the recombinant ones was constructed using methods including Maximum Likelihood (ML) method that are packaged in the MEGA6.0[36]. The CP gene of one *Narcissus yellow stripe virus* (NYSV) isolate was used as outgroup[37]. Bootstrap analysis was repeated 1000 times to evaluate the significance of the internal branches.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2017 doi:10.20944/preprints201705.0109.v1

2.4. Sequences diversity and population demography analysis

DnaSP version 5.10 was used to calculate the values of nucleotide diversity, Tajima's D, Fu and Li's D and F tests, haplotype diversity and nucleotide diversity[38-40]. Tajima's D, Fu and Li's D and F tests hypothesize that all mutations are selectively neutral. Tajima's D test depends on the differences between the numbers of segregating sites and the average number of nucleotide differences. Fu and Li's D test is related the differences between the number of singletons (mutations appearing only once among the sequences) and the total numbers of mutations. Fu and Li's F test is based on the differences between the numbers of singletons and the average number of nucleotide differences among all pairs of sequences. Haplotype diversity refers to the frequency and number of haplotypes in the population. Nucleotide diversity estimates the average pairwise differences among sequences. The nucleotide diversities were calculated within and between groups. DnaSP version 5.10[40] was also used to estimate the frequency distribution of the number of pairwise differences among all sequences. Mismatch distribution of all populations were estimated on all pairs of haplotypes present in a population[40]. Mismatch distribution analysis was based on 1 000 simulated samples and used to evaluate whether a population had undergone sudden expansion or maintained constant size. In a recently expanded and still intact population, the majority of lineage coalescence events were expected to produce a smooth unimodal Poisson distribution around the time of expansion; otherwise, multimodal and ragged distribution was expected.

2.5. Selection pressure, genetic differentiation and gene flow

The selection pressure was estimated by d_N/d_S ratio, where d_N represented the average number of non-synonymous substitutions per non-synonymous site and d_S represented the average number of synonymous substitutions per synonymous site. The values of d_N and d_S were estimated separately by using the PBL method[41, 42] implemented in MEGA 6.0. When d_N/d_S ratio =1, it means that neutral selection had occurred; when $d_N/d_S<1$ or >1, it means that negative (purifying) or positive (diversifying) selection, respectively, had occurred. Genetic distances were calculated by Pamilo-Bianchi-Li (PBL) methods[41, 42].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2017 doi:10.20944/preprints201705.0109.v1

Genetic differentiation between populations was examined by three permutation-based statistical tests, *Ks**, Z and Snn[43, 44]. $P < 0.05$ was considered as the criterion for rejecting the null hypothesis that there is no genetic differentiation between two subpopulations. The level of gene flow between populations was measured by estimating F_{st} (the inter-populational component of genetic variation or the standardized variance in allele frequencies across populations) and *Nm* using DnaSP 5.10[40]. F_{st} ranges from 0 to 1 for undifferentiated to fully differentiated populations, respectively. Normally, an absolute value of F_{st} >0.33 or *Nm* <1 suggests infrequent gene flow, while absolute value of F_{st} <0.33 or *Nm* >1 suggests frequent gene flow.

3. Results

3.1. Identities between TuMV isolates from radish in China

We collected and biologically cloned 101 TuMV isolates from radish from 2004 to 2010 from Beijing, Hebei, Heilongjiang, Henan, Jilin and Shandong provinces, 94 of which are first reported here. A fragment of 1082 bp covering partial NIb gene (28 bp), complete CP gene (867 bp) and 3'-UTR (187 bp) was amplified from these isolates. The geographical origin of each isolate is listed in Table 1.

Table 1. Recombination sites and possible parent-like isolates

Isolate	Major parent	Minor parent	B-E	Software*	*P*-value	Z-value#
CHBJ1	WF1-04	WF7-06	15-1051	GBS3	2.07×10^{-18}	9.7
CHBJ2	WF1-04	WF7-06	15-1051	GBS3	2.07×10^{-18}	9.65
CHK16	WF-05	R4	17-387	BMCS3	1.028×10^{-14}	9.43
CHK51	WF-05	R4	17-387	BMCS3	1.028×10^{-14}	9.38
R	WF-05	R4	17-697	MCS3	1.028×10^{-14}	8.92
R5	WF-05	R4	17-451	MCS3	1.028×10^{-14}	10.6
WF2-06	WF1-04	WF7-06	15-1051	GBS3	2.07×10^{-18}	10.2
WF3-06	WF1-04	WF7-06	10-1051	GBS3	2.07×10^{-18}	10.2
WF3-07	WF1-04	WF7-06	10-1047	GBS3	2.07×10^{-18}	10.2
WF8-08	WF1-04	WF7-06	10-1051	GBS3	2.07×10^{-18}	10.1
WF10-07	TA15-08	WF3-08	91-786	MCS3	1.18×10^{-8}	7.1
WFLB3	WF7-06	WF1-04	377-657	GBS3	2.07×10^{-18}	10.6

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2017 doi:10.20944/preprints201705.0109.v1

The recombination crossover sites within CP-UTR of turnip mosaic virus were detected by the recombination detecting programs.

*The programs supporting recombination event. R(RDP), G (Geneconv), B (Bootscan), M (Maxchi), C (Chimaera), S (Siscan) and 3 (3seq). The analysis was carried out with default settings for the different detection methods and a Bonferroni-corrected cutoff of 0.05. The program that has the greatest *P*-value was marked in bold. B-E represents the beginning and ending point of recombination.

The cloned sequences excluding primers shared identities of 89.6% - 100% at nt level with other 28 Chinese TuMV radish sequences available in Genbank. These 129 CP gene sequences showed identities of 88.2% - 100% at nt level and 91.3% - 100% at aa level. The identities of 44 TuMV isolates from Weifang were 89.6% -100% at nt level and 95.1% - 100% at aa level. Those of 37 isolates from Tai'an were 88.9% - 100% at nt level and 94.1 - 100% at aa level. The 12 Changchun isolates shared identities of 90.3% - 100% at nt level and 95.8% - 100% at aa level.

3.2. Recombination analyses

Possible recombination events in the CP-UTR region of 129 radish isolates from China were detected with the program package RDP. Twelve of the sequences (11.4%) analyzed had 'clear' recombination. Among the recombinant isolates, WF0710 was the within-group recombinant of basal-BR isolates , with TA0815 as its major parent and WF0803 as the minor parent; others were between-group recombinants of Asian-BR and world-B isolates, most with WF-05 or WF1-04 of world-B as the major parent and WF7-06 or R4 of Asian-BR as minor parent; WFLB had WF7-06 as its major parent and WF1-04 as minor parent (Table 1).

The recombination pattern can be classified into six types (Fig. 1). More than 50% recombinants belong to recombination pattern 1, with the recombination site located within UTR. WF0710 belonged to pattern II, WFLB3 to pattern III, CHK16 and CHK51 to pattern IV, R5 to pattern V and R to pattern VI (Fig. 1).

3.3. Phylogenetic analyses

Using ML method, a phylogenetic tree was constructed with the 118 CP-UTR sequences of TuMV (excluding the 11 between-group recombinants) from radish in China. These TuMV isolates were clustered to three lineages corresponding to world-B, Asian-BR

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2017 doi:10.20944/preprints201705.0109.v1

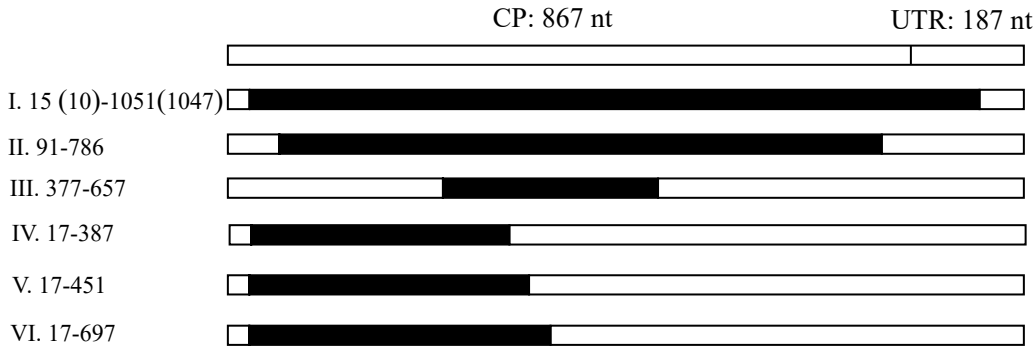


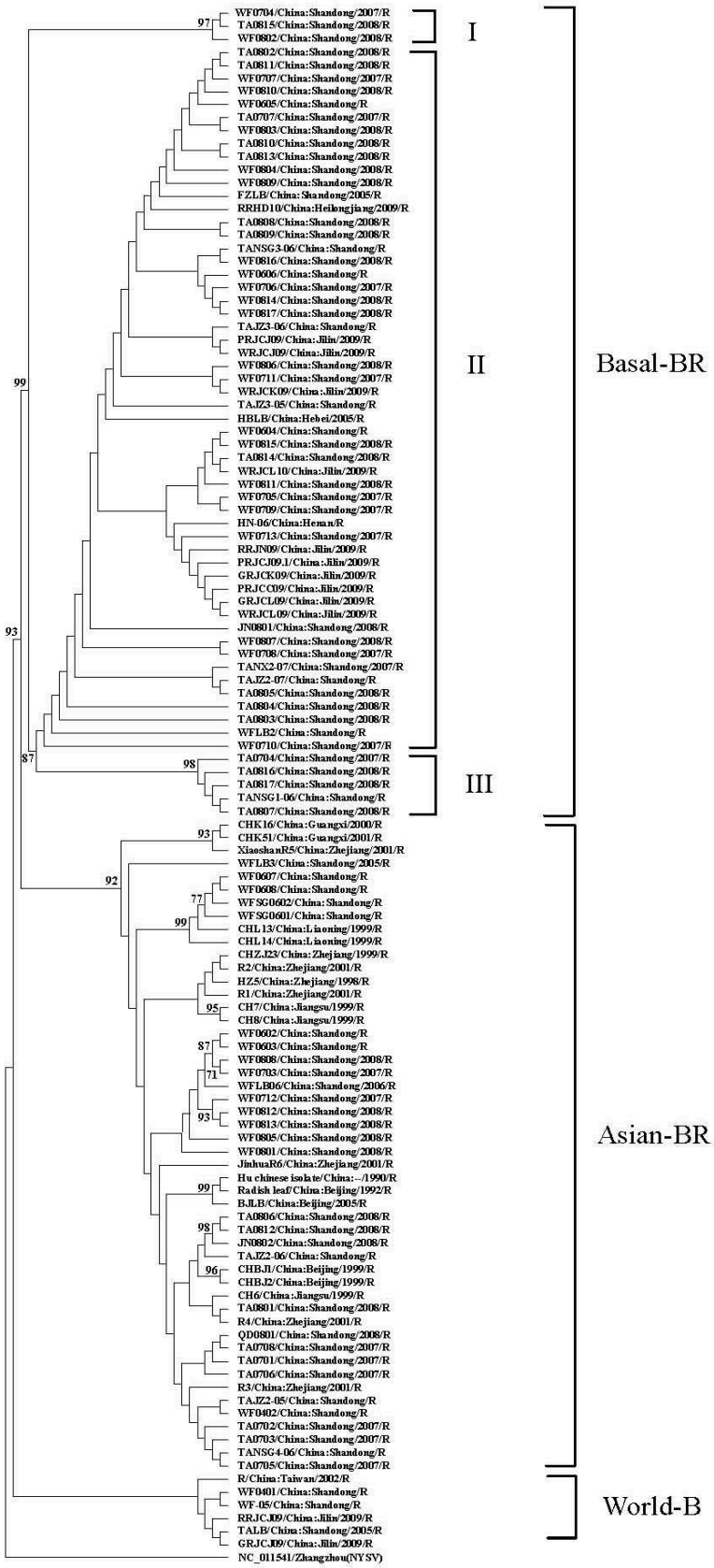
Figure 1. Recombination patterns in the CP-UTR region of TuMV from radish in China. Twelve recombinants were divided into 6 recombination patterns. I: CHBJ1, CHBJ2, WF2-06, WF3-06, WF8-08, WF3-07; II: WF10-07; III: WFLB3; IV: CHK16, CHK51; V: R5; VI: R.

and basal-BR (Fig. 2). The world-B lineage contained only six isolates (R, WF0401, WF-05, TALB, GRJCJ09 and RRJCJ09). The Asian-BR lineage consisted of 50 isolates. The Basal-BR lineage included 62 isolates which can be further divided into 3 sub-lineages. Sub-lineage Basal-BR I had three isolates (WF0704, WF0802 and TA0815), all of which were from Shandong province. Basal-BR II consisted of 54 isolates. Among which 41 were from Shandong, ten from Jilin, each from Henan, Heilongjiang and Hebei. Basal-BR III contained five isolates, all of which were found in Tai'an, Shandong province. The genetic distance values within groups ranged from 0.014 to 0.026, which were 4 to 5 times lower than those between groups (0.067 to 0.094) (Table 2). The genetic distance values between sub-groups of basal-BR were 0.033 to 0.041, which were higher than those within sub-groups but lower than those between groups.

The phylogenetic tree constructed with the CP gene could also be divided into three groups corresponding to world-B, Asian-BR and basal-BR. The genetic distance values between groups ranged from 0.076 to 0.091, which were higher than those within groups (0.015 to 0.049). The genetic distance values between sub-groups of basal-BR were 0.032 to 0.049, which were remarkably higher than those within sub-groups(0.004 to 0.015) but lower than those between groups. Therefore, the classification of these TuMV isolates into three groups and basal-BR into three sub-groups was reasonable.

To further study the genetic structure of TuMV basal-BR sub-populations from China and Japan, we constructed phylogenetic trees with basal-BR isolates available from both countries using ML method (Fig. 3). These TuMV isolates were clustered into four lineages,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2017 doi:10.20944/preprints201705.0109.v1



211

212 **Figure 2.** Maximum Likelihood tree of TuMV isolates from radish in China calculated from the CP

213 -UTR sequences.

Table 2. Estimates of genetic differentiation among sites (F_{ST}) within each region

Gene	Parameter	WF between WF	TA between TA	WF between TA	WF between CC	TA between CC	CC between CC
CP	Ks(*P*-value)	3.16314 (1.0000)	3.12325 (1.0000)	3.14501 (0.1200)	3.01985 (0.0080**)	2.90505 (0.0120*)	2.41801 (1.0000)
	Z (*P*-value)	1935.22093 (1.0000)	1367.97222 (1.0000)	1618.33343 (0.2990)	740.62579 (0.0080**)	534.81963 (0.0060**)	143.22727 (1.0000)
	Snn (*P*-value)	0.08081 (1.0000)	0.12416 (1.0000)	0.66143 ** (0.0030)	0.88492 (0.0000***)	0.85305 (0.0000***)	0.10000 (1.0000)
	F_{ST}	-0.02326	-0.02778	-0.00436	0.15467	0.15485	-0.09091
	*N*m	-11.00	-9.25	-57.56	1.37	1.36	-3.00
UTR	Ks*(*P*-value)	1.19299 (1.0000)	0.96193 (1.0000)	1.08658 (0.5460)	0.94143 (0.0570) ns	1.13222 (0.0420*)	0.88305 (0.9280) ns
	Z (*P*-value)	1844.07143 (1.0000)	1362.22222 (1.0000)	1581.52519 (0.4120)	552.71858 (0.0680) ns	723.74704 (0.0470*)	140.95455 (1.0000) ns
	Snn (*P*-value)	0.34906 (1.0000)	0.39139 (1.0000)	0.52678 (0.1430)	0.68796 (0.0230*)	0.69754 (0.0780) ns	0.31111 (1.0000) ns
	F_{ST}	-0.02381	-0.02778	-0.01268	0.08893	0.11459	-0.09091
	*N*m	-10.75	-9.25	-19.96	2.56	1.83	-3.00

ns, not significant; *, 0.01<P<0.05; **, 0.001<P<0.01; ***, P<0.001. Determined using 1000 permutations.

WF, isolates from Weifang of Shandong province; TA, isolates from Tai'an of Shandong province; CC, isolates from Changchun of Jilin province.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2017 doi:10.20944/preprints201705.0109.v1

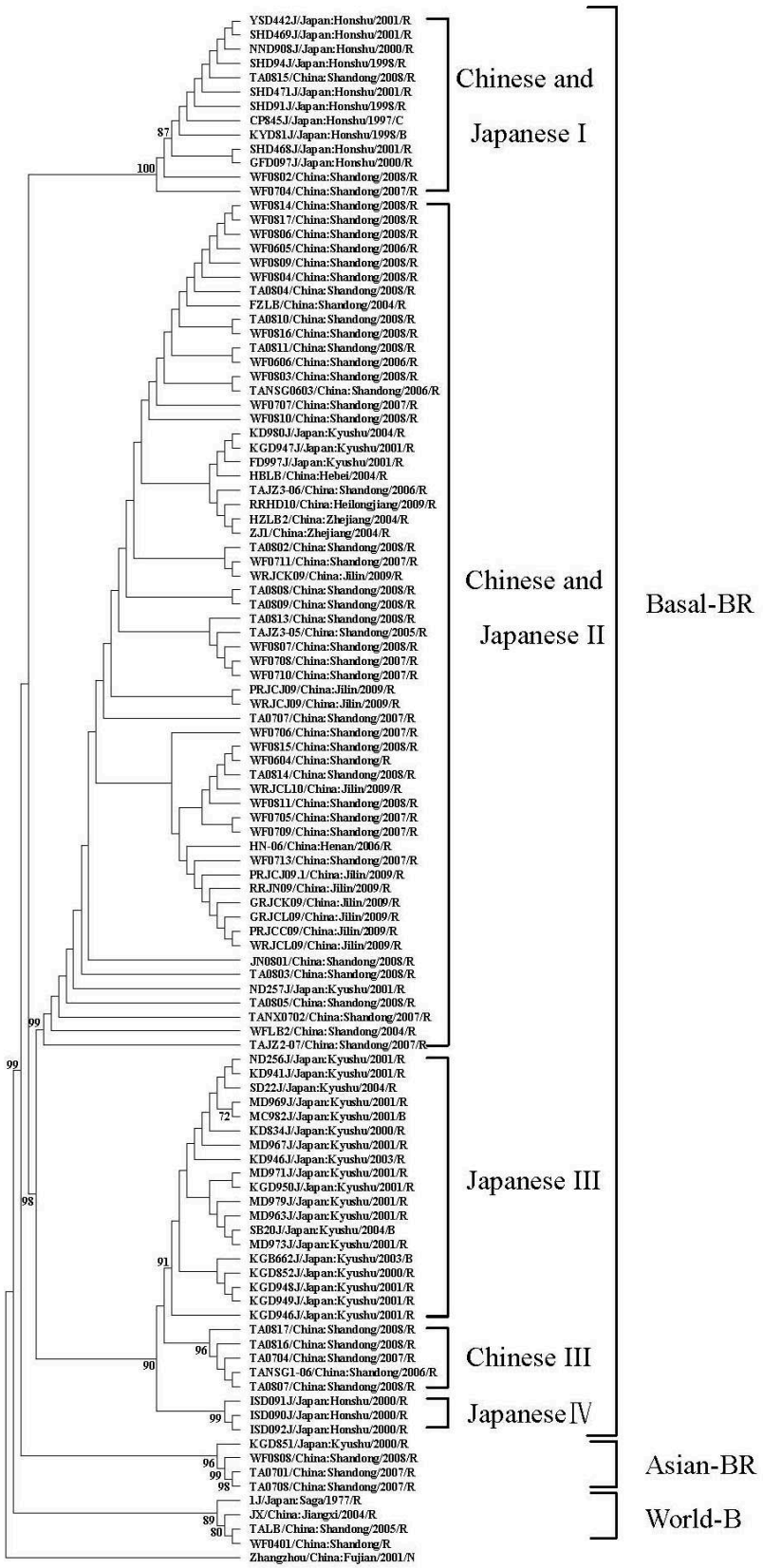


Figure 3. Phylogenetic trees were constructed by using the Maximum Likelihood method for CP gene nucleotide sequences of TuMV isolates of basal-BR collected from China and Japan. One *Narcissus yellow stripe virus* (NYSV) isolate (accession number: AJ311372) was used as outgroup. Isolates in the

 doi:10.20944/preprints201705.0109.v1

figure were listed by isolate name/ location of origin/ year of collection /original host (R is for *Raphanus*).

corresponding to the ones reported by Tomitaka *et al*.[25]. Interestingly, the TuMV basal-BR isolates of sub-groups I and II from both China and Japan formed common clusters, which indicated sub-populations of basal-BR I and II from these two countries were genetically identical. However, those of sub-group III from China and Japan formed separate clusters, indicating that China and Japan had different sub-populations of basal-BR III. Sub-group IV consisted of isolates from Japan only. No Chinese TuMV isolate fell into this sub-group.

3.4. Selective pressures acting on TuMV CP genes

To estimate the selection pressure acting on TuMV CP genes, we calculated the d_N/d_S ratios for TuMV sub-populations of different collection regions using Pamilo-Bianchi-Li (PBL) method assembled in MEGA version 6.0[36]. The d_N values for TuMV isolates from Weifang, Tai'an and Changchun were 0.005±0.001, 0.006±0.001 and 0.002±0.001, respectively, which were less than the d_S values (0.070±0.007, 0.098±0.012 and 0.014±0.006). Therefore, the values of the d_N / d_S ratio for TuMV *cp* genes were <1, indicating that purifying (negative) selection was acting on TuMV *cp* genes. The nucleotide distances were 0.023±0.002, 0.029±0.003 and 0.005±0.001, respectively, and showed no significant difference.

3.5. Genetic differentiation and gene flow

Genetic differentiation and gene flow between and within populations was examined by five permutation-based statistical tests, Ks*, Z and Snn or *F*st and Nm. The results showed no genetic differentiation between or within TuMV sub-populations from Tai'an and Weifang in the CP genes or UTR (Table 2).

The absolute values of F_{ST} between or within TuMV populations of Tai'an, Weifang and Changchun were all below 0.33, indicating that the gene flow between or within TuMV populations of Tai'an and Weifang, and that with TuMV population of Changchun is most frequent; however, the gene flow between Changchun and Tai'an, and Changchun and Weifang is less frequent. The absolute values of $Nm>1$ also support the conclusion on gene flow.

3.6. Population dynamics

The Tajima's *D*, Fu & Li's *D**, Fu & Li's *F** values for TuMV basal-BR II sub-population from Weifang and Tai'an of Shandong province were negative and the data is significant, which indicated that these sub-populations were in state of increasing (Table 3).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2017 doi:10.20944/preprints201705.0109.v1

Sub-populations of Asian-BR and basal-BRIII from Tai'an, Asian-BR from Weifang, and basal-BR from Changchun were also in a state of increasing, but the data was not significant (Table 3). Haplotype diversity, ranging from 0.890 to 1.000, had little difference betweens groups or sub-groups. The basal-BR II isolates from Tai'an had the lowest nucleotide diversity of 0.00322, while the Asian-BR isolates from Weifang had the highest one of 0.01159.

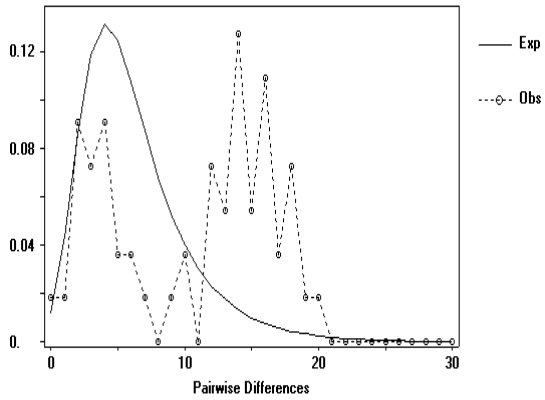
Table 3. Neutrality tests, haplotype and nucleotide diversity of *Turnip mosaic virus* sub-populations

Group		Haplotype diversity	Nucleotide diversity	Tajima's D	Fu and Li's D	Fu and Li's F
Tai'an	basal-BR II	0.890 (0.073)	0.00322 (0.00062)	-2.19554**	-2.72246*	-2.72246**
	basal-BR III	1.000 (0.126)	0.00417 (0.00073)	-1.18441	-1.18441	-1.24511
	Asian-BR	0.939 (0.058)	0.00661 (0.00130)	-0.95449	-0.79711	-0.95432
Weifang	basal-BR II	0.960 (0.031)	0.00490 (0.00073)	-2.39975**	-3.45534 **	-3.66541**
	Asian-BR	0.987 (0.023)	0.01159 (0.00161)	-1.22279	-1.79120	-1.88684
Changchun	basal-BR	0.982 (0.046)	0.00478 (0.00095)	-1.63909	-2.06292	-2.21575

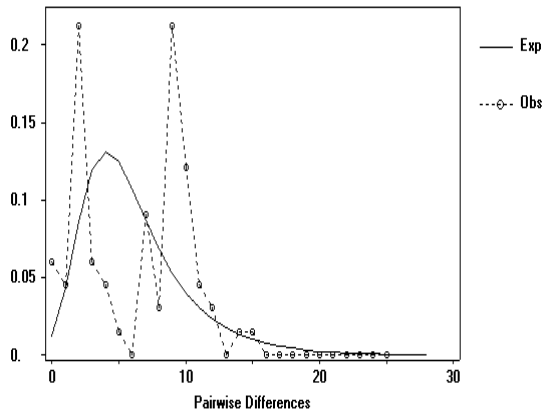
The sub-populations less than four isolates were not included. *: $P < 0.05$, **: $P < 0.02$.

The mismatch distribution of TuMV CP gene and 3'-UTR for the basal-BR II isolates collected from Weifang, Tai'an, Changchun and basal-BR III were unimodual and smooth, and fit well with the expected model of sudden expansion, indicating that these sub-populations were new emergent (Fig. 4). The Asian-BR isolates from Weifang and Tai'an of Shandong province and Zhejiang were multiple-peaked, ragged, indicating that these sub-populations were long-existing ones (Fig. 4).

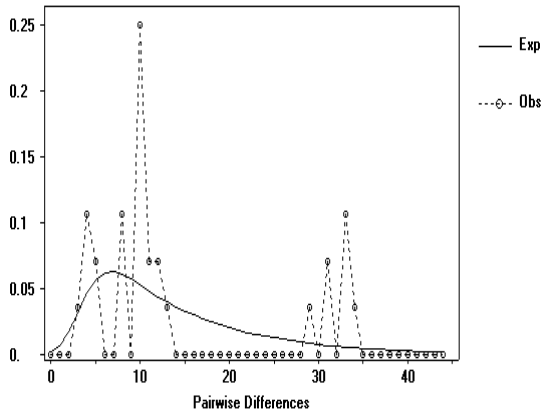
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2017 doi:10.20944/preprints201705.0109.v1



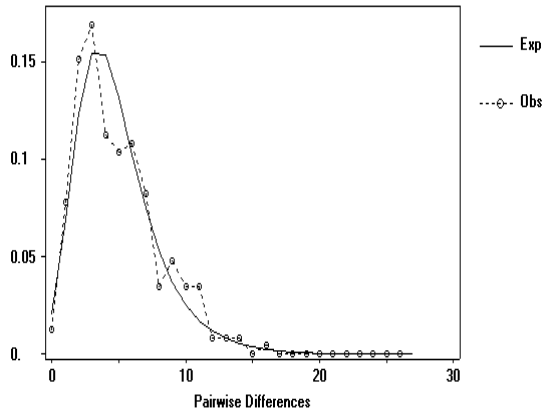
A. Weifang Asian-BR group



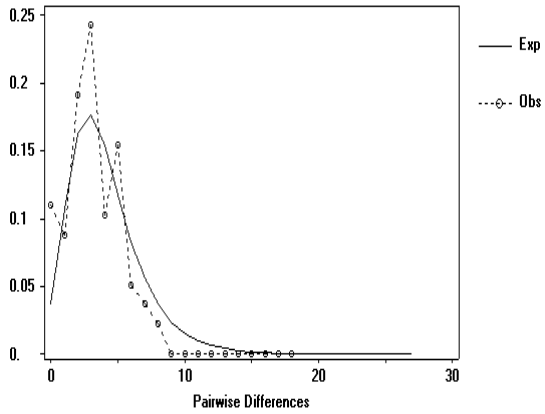
B. Taian Asian-BR group



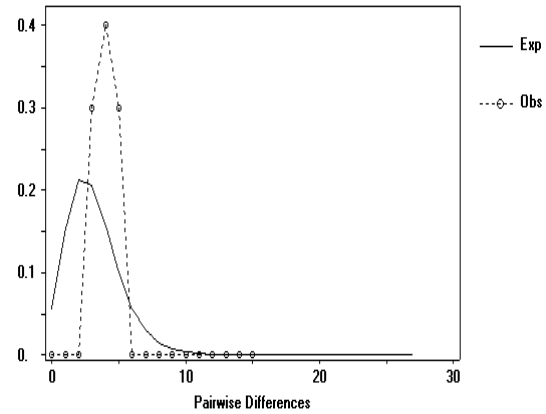
C. Zhejiang Asian-BR group



D. Weifang basal-BR II group

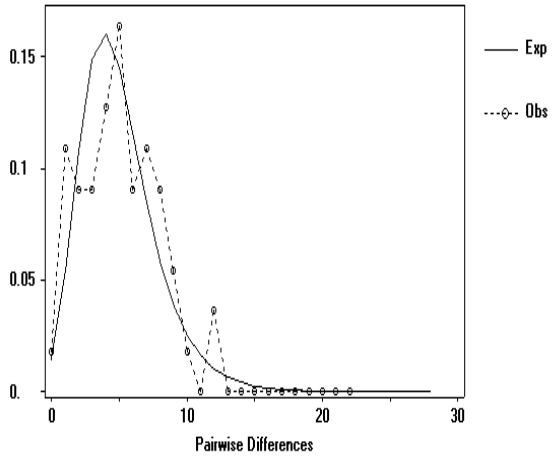


E. Tai'an basal-BR II group



F. basal-BR III

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2017 doi:10.20944/preprints201705.0109.v1



G. Changchun basal-BR group

Fig. 4. The frequency distribution of the number of pairwise nucleotide differences obtained from CP gene nucleotide sequences. (A) basal-BR II group of Weifang; (B) Asian-BR group of Weifang; (C) basal-BR II group of Tai′an, (D) basal-BR III group of Tai′an; (E) basal-BR II group of Tai′an, (F) basal-BR III group; (G) Changchun isolates of basal-BR group. Broken line represents the observed data and unbroken line represents the expected data. The sub-populations less than four isolates were not included.

4. Discussion

In this paper, we studied the molecular structure of TuMV population from China by analyzing the CP gene sequence of 129 TuMV isolates from radish and comparing them with 41 isolates of basal-BR group from Japan. Our results show that (1) about one-tenth of the TuMV isolates characterized are recombinants; (2) sub-population of basal-BR expands rapidly and accounts to more than one half of the isolates detected; (3) isolates of basal-BR in China evolve to three sub-groups, with sub-groups I and II genetically homologous with Japanese ones, while sub-group III a distinct lineage; (4) Sub-populations of TuMV basal-BR II and III are new emergent and in a state of expansion; (5) the TuMV population of China is under negative selection; (6) frequent gene flow is detected between TuMV sub-populations from Weifang, Taian and Changchun.

Recombination is important in virus evolution and has been detected in many potyvirus species[19, 20, 24, 45-48]. The percentages of recombinant isolates may account to ten to sixty-five of isolates studied[24, 47]. Intra- and inter-lineage recombination is very common in TuMV[6, 18, 22]. The hotspots of recombination sites of TuMV genome are located in the P1 and CI/6K2/VPg region[21]. Ohshima and colleagues have detected 37 recombination patterns[6, 21]. Novel recombination patterns of TuMV are increasing[6, 18]. About 10% of

 doi:10.20944/preprints201705.0109.v1

297 the TuMV isolates characterized in this study experienced 'clear' recombination event. The
298 percentage is a little lower than previous studies[18, 22]. The underlying reason is that we
299 just analyzed the CP-UTR region, where the crossover sites of TuMV are scarce[18, 22, 23].
300 If longer sequences or the whole genome is included, there will be more recombination
301 events detected.

302 The d_N/d_S ratios are often used to estimate the selection pressure under which viral
303 gene(s) suffered[10, 49]. Positive selection ($d_N/d_S > 1$) may endow the virus more fitness to
304 adapt a new host or environment. However, rapid divergence driven by positive selection has
305 been rarely demonstrated[50]. Like the case of most virus genes, our results show that
306 negative (purifying) selection dominates the evolution of TuMV CP genes. If selection
307 pressure on single residue is estimated, amino acids under positive selection may be sorted
308 out[49].

309 Basal-BR is a new emergent in east Asia and has been detected in Japan and China [6,
310 17, 23, 28]. So far, there has been no basal-BR isolates reported in Vietnam[20]. After its
311 first detection in 2005[28], the population of basal-BR isolates increased rapidly in China
312 and showed characteristics of founder effect. As reported in this research, Basal-BR isolates
313 were detected from samples from Hebei, Henan, Jilin and Shandong provinces, and
314 accounted to more than half of the isolates from Shandong and Jilin provinces. The Chinese
315 basal-BR isolates have evolved to three sub-groups. Among the 48 basal-BR isolates from
316 Weifang and Tai'an of Shandong province, 40 belonged to sub-group II, which represents
317 the prevalent cluster in those areas. Basal-BR III was detected after 2006 and only found in
318 Tai'an of Shandong province. What's more interesting, sub-groups of basal-BR I and II are
319 genetically homologous to those of Japanese isolates, while sub-groups of basal-BR III
320 from China and Japan are genetically distinct and form separate clusters, indicating that
321 China and Japan had different sub-populations. Another difference is that the prevalent
322 subgroup of Basal-BR is II in China but III in Japan.

323 The gene flow between TuMV isolates of basal-BR II and III from Weifang, Tai'an and
324 Changchun is frequent. But TuMV is transmitted by aphid in a non-persistent manner and
325 there is no evidence of seed transmission reported[4, 5]. It remains unknown how TuMV
326 isolates, especially the new emergent, spread to other places[18, 23, 25]. But TuMV isolates
327 of basal-BR II are prevalent and expanding rapidly in Weifang and Tai'an of Shandong and
328 Changchun of Jilin. Therefore, a program should be launched to evaluate the resistance of
329 commercial available cultivars of cruciferous crops to TuMV isolates, particularly basal-BR
330 II.

Acknowledgments: This research was supported by National Natural Scientific Foundation of China (NSFC, 31571984 and 31201485), Shandong Provincial Natural Science Foundation, China (SDNSF, Z2007D04 and ZR2011CM019), The National Key Research and Development Program of China (2016YFC0501202) and Jiangsu Provincial Natural Science Foundation, China (BK20140230).

Author Contributions: Xiangdong Li, Tiansheng Zhu, Xiao Yin and Jinliang Liu had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis; Study conception and design: Xiangdong Li and Yanping Tian; Acquisition of data: Tiansheng Zhu, Xiao Yin, Chengling Zhang, Jia Chen and Jinliang Liu; Analysis and interpretation of data: Xiangdong Li, Tiansheng Zhu, Xiao Yin, Yanping Tian and Jinliang Liu.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Roossinck, M. J., Mechanisms of plant virus evolution. *Annual Review of Phytopathology* **1997,** 35, (35), 191.
2. Roossinck, M. J., *Plant Virus Evolution*. Springer Berlin Heidelberg: 2008; p 157-164.
3. Garcíaarenal, F.; Fraile, A.; Malpica, J. M., Variability and genetic structure of plant virus populations. *Annual Review of Phytopathology* **2001,** 39, (1), 157.
4. Tomlinson, J. A., Epidemiology and control of virus diseases of vegetables. *Annals of Applied Biology* **1987,** 110, (3), 661-681.
5. Shukla, D. D.; Ward, C. W.; Brunt, A. A., The Potyviridae. **1994**.
6. Wang, H. Y.; Liu, J. L.; Gao, R.; Chen, J.; Shao, Y. H.; Li, X. D., Complete genomic sequence analyses of Turnip mosaic virus basal-BR isolates from China. *Virus Genes* **2009,** 38, (3), 421-8.
7. Riechmann, J. L.; Laín, S.; García, J. A., Highlights and prospects of potyvirus molecular biology. *Journal of General Virology* **1992,** 73 (Pt 1), (4), 1-16.
8. Urcuqui-Inchima, S.; Haenni, A. L.; Bernardi, F., Potyvirus proteins: a wealth of functions. *Virus Research* **2001,** 74, (1-2), 157.
9. Chung, B. Y.; Miller, W. A.; Atkins, J. F.; Firth, A. E., An overlapping essential gene in the Potyviridae. *Proceedings of the National Academy of Sciences of the United States of America* **2008,** 105, (15), 5897-5902.
10. Wei, T.; Zhang, C.; Hong, J.; Xiong, R.; Kasschau, K. D.; Zhou, X.; Carrington, J. C.; Wang, A., Formation of complexes at plasmodesmata for potyvirus intercellular movement is mediated by the viral protein P3N-PIPO. *Plos Pathogens* **2010,** 6, (6), e1000962.
11. Edwardson, J. R.; Christie, R. G., The Potyvirus Group. *University of Florida Monograph* **1991,** 3, 699-712.
12. Shattuck, V. I., The Biology, Epidemiology, and Control of Turnip Mosaic Virus. *Horticultural Reviews* **2010,** 14, 199-238.
13. Green, S. K.; Deng, T. C., Turnip mosaic virus strains in cruciferous hosts in Taiwan. *Plant Disease* **1985,** 69, (69), 28-31.
14. Liu, X. P.; Lu, W. C.; Li, J. L.; Liu, Y. K., A study on TuMV strain differentiation of cruciferous vegetables from ten provinces in China- Selection of new identification host and strains portion. *Science Bulletin* **1990,** 35, (20), 1734-1739.
15. Jenner, C. E.; Keane, G. J.; Jones, J. E.; Walsh, J. A., Serotypic variation in turnip mosaic virus. *Plant Pathology* **1999,** 48, (1), 101–108.
16. Jenner, C. E.; Walsh, J. A., Pathotypic variation in turnip mosaic virus with special reference to European isolates. *Plant Pathology* **1996,** 45, (5), 848-856.
17. Ohshima, K. Y. Y.; Hirota, R.; Hamamoto, T.; Tomimura, K.; Tan, Z. S. T.; Azuhata, F.; Walsh, J. A.; Fletcher, J.; Chen, J. S.; Gera, A., Molecular evolution of Turnip mosaic virus: evidence of host

adaptation, genetic recombination and geographical spread. *Journal of General Virology* **2002,** 83, (6), 1511-1521.

18. Nguyen, H. D.; Tomitaka, Y.; Ho, S. Y. W.; Duchêne, S.; Vetten, H. J.; Lesemann, D.; Walsh, J. A.; Gibbs, A. J.; Ohshima, K., Turnip Mosaic Potyvirus Probably First Spread to Eurasian Brassica Crops from Wild Orchids about 1000 Years Ago. *Plos One* **2013,** 8, (2), e55336.
19. Chare, E. R.; Holmes, E. C., A phylogenetic survey of recombination frequency in plant RNA viruses. *Archives of Virology* **2006,** 151, (5), 933-946.
20. Nguyen, H. D.; Tran, H. T.; Ohshima, K., Genetic variation of the Turnip mosaic virus population of Vietnam: a case study of founder, regional and local influences. *Virus Research* **2013,** 171, (1), 138-149.
21. Ohshima, K.; Tomitaka, Y.; Wood, J. T.; Minematsu, Y.; Kajiyama, H.; Tomimura, K.; Gibbs, A. J., Patterns of recombination in turnip mosaic virus genomic sequences indicate hotspots of recombination. *Journal of General Virology* **2007,** 88, (Pt 1), 298-315.
22. Tan, Z.; Wada, Y.; Chen, J.; Ohshima, K., Inter- and intralineage recombinants are common in natural populations of Turnip mosaic virus. *Journal of General Virology* **2004,** 85, (9), 2683-96.
23. Tomitaka, Y.; Ohshima, K., A phylogeographical study of the Turnip mosaic virus population in East Asia reveals an 'emergent' lineage in Japan. *Molecular Ecology* **2007,** 15, (14), 4437-4457.
24. Tomimura, K.; Gibbs, A. J.; Jenner, C. E.; Walsh, J. A.; Ohshima, K., The phylogeny of Turnip mosaic virus; comparisons of 38 genomic sequences reveal a Eurasian origin and a recent 'emergence' in east Asia. *Molecular Ecology* **2003,** 12, (8), 2099-2111.
25. Tomitaka, Y.; Yamashita, T.; Ohshima, K., The genetic structure of populations of Turnip mosaic virus in Kyushu and central Honshu, Japan. *Journal of General Plant Pathology* **2007,** 73, (3), 197-208.
26. Shi, M. L., Cloning and sequence analysis of HC-Pro genes of Turnip mosaic virus Eurasian isolates. *Acta Phytopathologica Sinica* **2007**, 383-389.
27. Song, Y., Cloning and sequence analysis of coat protein genes of turnip mosaic virus isolates obtained from shandong. *Scientia Agricultura Sinica* **2005,** 58, (3), 504-510.
28. Tian, Y. P.; Zhu, X. P.; Liu, J. L.; Yu, X. Q.; Du, J.; Kreuze, J.; Li, X. D., Molecular Characterization of the 3'-Terminal Region of Turnip mosaic virus Isolates from Eastern China. *Journal of Phytopathology* **2007,** 155, (155), 333-341.
29. Liu, J. L.; Yu, X. Q.; Tian, Y. P.; Du, J.; Zhu, X. P.; Li, X. D.; Yan, D. Y., Molecular characterization and coat protein gene expression of a turnip mosaic virus isolate from radish in Weifang. *Acta Horticulturae Sinica* **2006,** 33, 84-88.
30. Martin, D. P.; Lemey, P.; Lott, M.; Moulton, V.; Posada, D.; Lefeuvre, P., RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* **2010,** 26, (19), 2462-2463.
31. Sawyer, S. A., GENECONV: a computer package for the statistical detection of gene conversion. *Washington University in St. Louis* **1999,** http://www.math.wustl.edu/~sawyer.
32. Salminen, M. O.; Carr, J. K.; Burke, D. S.; Mccutchan, F. E., Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. *Aids Research & Human Retroviruses* **1995,** 11, (11), 1423-5.
33. Smith, J. M., Analyzing the mosaic structure of genes. *Journal of Molecular Evolution* **1992,** 34, (2), 126-9.
34. Posada, D.; Crandall, K. A., Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proceedings of the National Academy of Sciences of the United States of America* **2001,** 98, (24), 13757-62.
35. Weiller, G. F., Phylogenetic profiles: a graphical method for detecting genetic recombinations in homologous sequences. *Molecular Biology & Evolution* **1998,** 15, (3), 326-35.
36. Tamura, K.; Stecher, G.; Peterson, D.; Filipski, A.; Kumar, S., MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology & Evolution* **2013,** 44, (4), 2725.
37. Chen, J.; Chen, J. P.; Langeveld, S. A.; Derks, A. F. L. M.; Adams, M. J., Molecular characterization of carla- and potyviruses from Narcissus in China. *Journal of Phytopathology* **2003,** 151, (1), 26-29.
38. Fu, Y. X.; Li, W. H., Statistical tests of neutrality of mutations. *Genetics* **1993,** 133, (3), 693.
39. Tajima, F., Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **1989,** 123, (3), 585-95.
40. Librado, P.; Rozas, J., DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **2009,** 25, (11), 1451-1452.
41. Pamilo, P.; Bianchi, N. O., Evolution of the Zfx and Zfy genes: rates and interdependence between the genes. *Molecular Biology & Evolution* **1993,** 10, (2), 271.
42. Li, W. H., Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *Journal of Molecular Evolution* **1993,** 36, (1), 96-99.

 doi:10.20944/preprints201705.0109.v1

43. Hudson, R. R., A new statistic for detecting genetic differentiation. *Genetics* **2000,** 155, (4), 2011.
44. Hudson, R. R.; Boos, D. D.; Kaplan, N. L., A statistical test for detecting geographic subdivision. *Molecular Biology & Evolution* **1992,** 9, (1), 138.
45. Moreno, I. M.; Malpica, J. M.; Díaz-Pendón, J. A.; Moriones, E.; Fraile, A.; García-Arenal, F., Variability and genetic structure of the population of watermelon mosaic virus infecting melon in Spain. *Virology* **2004,** 318, (1), 451-60.
46. Alla G Gagarinova, M. B., Martina V Strömvik, Aiming Wang, Recombination analysis of Soybean mosaic virus sequences reveals evidence of RNA recombination between distinct pathotypes. *Virology Journal* **2008,** 5, (1), 1-8.
47. Tugume, A. K.; Cuéllar, W. J.; Mukasa, S. B.; Valkonen, J. P. T., 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* **2010,** 19, (15), 3139-3156.
48. Zhang, C. L.; Gao, R.; Wang, J.; Zhang, G. M.; Li, X. D.; Liu, H. T., Molecular variability of Tobacco vein banding mosaic virus populations. *Virus Research* **2011,** 158, (1-2), 188.
49. Tian, Y. P.; Liu, J. L.; Zhang, C. L.; Liu, Y. Y.; Wang, B.; Li, X. D.; Guo, Z. K.; Valkonen, J. P., Genetic diversity of Potato virus Y infecting tobacco crops in China. *Phytopathology* **2011,** 101, (3), 377.
50. Nielsen, R., Changes in ds/dn in the HIV-1 env gene. *Molecular Biology & Evolution* **1999,** 16, (5), 711-714.