Acta virologica **53**: 89 – 97, 2009 doi:10:4149/av_2009_02_89

Molecular variability of five Cucumber mosaic virus isolates from China

Y.Y. LIU[1,2$], S.L. YU[2$], Y.F. LAN[1], C.L. ZHANG[1], S.S. HOU[1], X.D. LI[1*], X.Z. CHEN[1,3], X.P. ZHU[1]

[1]Department of Plant Pathology, College of Plant Protection, Shandong Agricultural University, Tai'an, Shandong 271018, P.R. China; [2]Shandong Institute of Peanut, Qingdao, Shandong 266100, P.R. China; [3]Mengyin Tobacco Company, Mengyin, Shandong 276000, P.R. China

Received November 20, 2008; accepted March 27, 2009

Summary. – Cucumber mosaic virus (CMV) isolates are currently divided into two main groups, I and II according to their genomic sequences. The group I is further divided into two subgroups IA and IB. We performed a phylogenetic analysis of the genome regions containing 1a, 2a, 2b, coat protein (CP), and movement protein (MP) genes of 5 CMV isolates from China and other 28 CMV isolates available in the GenBank. The results indicated that CMV isolates could be genetically divided into three groups I, II, and III according to the genes encoding MP, CP, 1a, and 2a proteins and to the 2 groups according to the gene 2b. Group I could be further divided into two subgroups (IA and IB) according to the genes encoding CP, MP, 2a, and 2b proteins and to the three subgroups (IA, IB, and IC) according to the gene encoding 1a protein. Four of 5 examined Chinese CMV isolates belonged to the subgroup IB, while the remaining isolate was a natural inter-subgroup reassortant. We found that the 2b gene of CMV was under positive selection, while the other genes were under negative selection. No evidence of the selection associated with a host adaptation or geographic distribution was found.

Keywords: Cucumber mosaic virus; phylogenetic analysis; nucleotide diversity; reassortant

Introduction

CMV, the type species of the genus *Cucumovirus,* the family *Bromoviridae*, is probably the most widespread and successful plant virus. It has extremely broad host range and could infect plants of more than 1,000 species including many economically important crops (Palukaitis *et al*., 1992, 2003). CMV can be transmitted by at least 75 species of aphids in a non-persistent manner and through seeds. Its genome consists of three positive-sense RNAs that are termed as RNA1, 2, and 3 in a decreasing order of size. The RNAs have 5 ORFs packaged in the separate particles (Palukaitis *et al*., 2003). RNA1 encodes only one protein signed 1a, which contains methyl transferase and helicase motifs and functions as RNA-dependent RNA polymerase (Roossinck, 2001). RNA2 encodes two proteins signed 2a and 2b. The 2a protein together with 1a protein forms viral component of the replicase complex (Hayes and Buck, 1990; Roossinck, 2002). The 2b protein takes part in a host-specific long-distance movement, symptom severity, and suppression of RNA silencing (Ding *et al*., 1994, 1995, 1996; Brigneti *et al*., 1998). A recent study has shown that 2b gene determined the selection of inter-viral recombination by affecting the crossover site, acceptor RNA, and selection rate (Shi *et al*., 2008). RNA3 also encodes two proteins, 3a and 3b. The 3a protein is a cell-to-cell MP. The 3b protein is the CP that is involved in several processes including virus assembly, cell-to-cell and vascular movement, and aphid-mediated transmission (Suzuki *et al*., 1991; Palukaitis *et al*., 1992; Canto *et al*., 1997; Roossinck, 2001).

*Corresponding author. E-mail: xdongli@sdau.edu.cn; fax: +86-5388226399.

$These authors contributed equally to this paper.

Abbreviations: CMV = Cucumber mosaic virus; CP = coat protein; d_N = average number of non-synonymous substitutions per non-synonymous site; d_S = average number of synonymous substitutions per synonymous site; ME = minimum evolution; MP = movement protein; NJ = neighbor-joining; NTR = non-translated region; PBL = Pamilo-Bianchi-Li; RDP = recombination detection program; UPGMA = unweighted pair-group method using arithmetic averages

The CMV isolates can be divided into two main groups I and II according to the results of serologic reaction, peptide mapping, nucleic acid hybridization, and nucleotide sequence similarity (Palukaitis *et al*., 1992). Analysis of the CP gene of several isolates from group I suggested that they can be further divided into two subgroups (Chaumpluk *et al*., 1996). Analysis of the 5'-nontranslated region (NTR) of RNA3 and the CP gene of many CMV isolates support the further division of group I into subgroups IA and IB (Roossinck *et al*., 1999). However, the results of phylogenetic analyses of all ORFs did not completely support the classification inferred from the CP ORF and indicated that different RNAs had independent evolutionary histories (Roossinck, 2002). This is consistent with the hypothesis that the reassortment mechanism played an important role in the evolution of CMV (Roossinck, 2002).

The data of genetic sequences of plant viruses are quickly increasing, what is important for a new classification of individual virus species. In this paper, we determined the genomic sequences encoding 1a, 2a, 2b, MP, and CP proteins of 5 CMV isolates from China. Then, we performed the identity, recombination, phylogenetic, and genetic diversity analyses of 5 Chinese isolates with other 28 CMV isolates available in GenBank.

Materials and Methods

Virus isolates. Five CMV isolates were obtained from tobacco (*Nicotiana tabacum*) or petunia (*Petunia hybrida*) plants showing yellowing and/or mosaic symptom in regions Laiwu (isolate LW), Qingzhou (isolate QZ), Rizhao (isolate RZ) and Tai'an (isolates ND1 and ND2) of Shandong Province, China. The isolates were biologically cloned through single lesions on *Chenopodium amaranticolor* leaves for at least three times before propagation in *N. tabacum* cv. Samsun plants.

RNAs extraction and RT-PCR. Total RNAs were extracted from the leaves of inoculated *N. tabacum* cv. Samsun plants with $TRI_{ZOL}^{®}$ Reagent (Invitrogen) and kept at -70˚C until used.

Primers (Table 1) used to amplify different genomic regions of CMV were designed according to the sequences deposited in GenBank except for CMVR1-R and CMVMP-R were adopted from Lin *et al*. (2004). First-strand cDNAs were synthesized by Moloney murine leukemia virus reverse transcriptase (Promega). PCR were conducted using *Taq* DNA polymerase (TaKaRa). The program used to amplify 1a and MP genes was 94˚C/3 mins, 4 cycles of 94˚C/45 secs, 37˚C/50 secs and 72˚C/2 mins, then 27 cycles of 94˚C/45 secs, 57˚C/50 secs and 72˚C/2 mins. The PCR program for 2a and 2b genes was 94˚C/3 mins, 4 cycles of 94˚C/45 secs, 37˚C/50 secs and 72˚C/4 mins, 27 cycles of 94˚C/45 secs, 60˚C/50 secs, and 72˚C/4 mins. The reaction conditions for amplifying CP gene was 94˚C/3 mins, 30 cycles of 94˚C/1 min, 63˚C/50 secs, and 72˚C/1 min. All the programs were terminated with final extension at 72˚C for 10 mins.

Cloning and sequencing. PCR products with the expected size were purified from agarose gel, ligated into the pMD18-T vector (TaKaRa) using T4-DNA ligase (TaKaRa), and then transformed into competent cells of *Escherichia coli* DH5α. Recombinant plasmids were identified by electrophoresis and PCR analysis. Clones containing insertion of foreign fragment were sequenced by Biosune Biotechnology Company. For each fragment, two clones from independent PCR reactions were sequenced twice from both ends. If there was any discrepancy at any position of the sequences, at least two more clones were sequenced to determine correctly the base at the position concerned (Yu *et al*., 2007).

Recombination analysis. Possible recombination events were detected with software RDP3 and Phylipro. RDP3 included programs Bootscanning, Geneconv, Maxchi (Maximum Chi Square method), Chimaera, Siscan (Sister Scanning Method), 3SEQ, and Lard (Likelihood Assisted Recombination Detection) (Maynard Smith, 1992; Salminen *et al*., 1995; Weiller, 1998; Holmes *et al*., 1999; Padidam *et al*., 1999; Posada *et al*., 2001; Gibbs *et al*., 2000; Martin *et al*., 2005; Boni *et al*., 2007). RDP3 was used with default settings and Bonferroni corrected *P*-value cut-off of 0.05 or 0.01. Only those recombination events supported by more than 4 programs and with *P*-value $<1.0 \times 10^{-6}$ were regarded as "clear" recombination event. Otherwise, they were regarded as "tentative" recombination events (Tomitaka *et al*., 2006).

Table 1. Primers used to amplify different genomic regions of CMV

Primer name	Genomic region	Positions[a]	Expected size (bp)	Nucleotide sequence[b]
CMVR1-F	1a	13–33	1077	5'-GAGCGTACGGTTCAAYCCCTG-3'
CMVR1-R		1070–1089		5'-TGTCGAATGAGTTCGGGTGG-3'
CMVR2-F		13–33		5'-GAGCGTACGGTTCAAYCCCTG-3'
	2a and 2b		2860	
CMVR2-R		2848–2872		5'-CGTWAGCTGGATGGACAACCCGTTC-3'
CMVMP-F	MP	97–118	1171	5'-CAGTGTGTTAGAWTWCCCGAGG-3'
CMVMP-R		1247–1267		5'-GATTTGTCCATGACTCGACTC-3'
CMVCP-F	CP	1257–1279	657	5'-ATGGACAAATCYGRATCAACCAG-3'
CMVCP-R		1888–1913		5'-TCAAACTGGGAGCACNCCNGATGTGG-3'

[a]Nucleotide positions of the primers were designated according to the genomic RNAs of CMV-Fny.
[b]R = A or G ; W = A, T; Y = C, T; N = A, C, G, T.

Phylogenetic and diversity analyses. Multiple nucleotide sequence alignments were performed by using Clustal W program (Thompson *et al*., 1994). Sequence analyses and comparisons were performed using the programs DNASTAR 6.0 (DNAStar) and DNAMAN (Lynnon Biosoft).

Phylogenetic analyses were performed employing the neighbor-joining (NJ) method, minimum evolution (ME) method and unweighted pair-group method using arithmetic averages (UPGMA) that are packaged in the MEGA3.1 software (Kumar *et al*., 2004). All branches with bootstrap value support <50% were collapsed. Bootstrap value >80% were listed. Some proven standard isolates were taken as reference isolates, e.g. isolates Fny, MF, and Y for subgroup IA, isolates CTL, NT9, SD, and TFN for subgroup IB, and isolates LY, Q, S, and Trk7 for subgroup II (Roossinck *et al*., 1999; Lin *et al*., 2004; Chen *et al*., 2007).

Genetic diversity within and between groups of host and geographical origin were calculated with Pamilo-Bianchi-Li (PBL) method (Pamilo and Bianchi, 1993; Li, 1993). Confidence estimates for non-synonymous (amino acid-altering) and synonymous (silent) substitution rates (d_N and d_S, respectively) were calculated separately by using the bootstrap method with 500 replicates (Nei and Kumar, 2000; Kumar *et al*., 2004). A value of $d_N/d_S = 1$ meant that neutral evolution had occurred. When $d_N/d_S <1$ or >1, it meant that negative (purifying) or positive (diversifying) selection had occurred.

Results

The amplified genomic sequences of five CMV isolates

The genomic sequences of 5 CMV isolates determined in this research have been submitted to GenBank and allocated under Acc. Nos of DQ916111 and EF159146, EU414784 to EU414801 (Table 2).

The amplified regions of RNA1 for isolates LW and RZ were both 1037 nt in length, while those of isolates ND1 and ND2 were 1036 nt. The RNA1 fragment of isolate QZ was 3167 nt long and was amplified by the primers for RNA2 accidentally. However, in the following analyses, only the 5'-part of corresponding position and similar size with other four isolates was considered. The amplified regions of RNA2 varied from 2810 to 2859 nt with variable 5'-NTR of 45–73 nt. The 2a and 2b genes for isolates LW, ND1, ND2, and QZ were 2577 nt and 336 nt long, respectively. However, both genes were 3 nt shorter for isolate RZ, what meant that the corresponding proteins 2a and 2b of this isolate fell short of one amino acid. For all five isolates the sizes of MP and CP genes were 840 nt and 657 nt, respectively.

Sequence identity analyses

The genomic sequences of other 28 CMV isolates from the GenBank (Table 2) were downloaded and analyzed with the sequences of 5 CMV isolates examined in our research.

For gene 1a, the identities of isolates LW, ND1, ND2, QZ, and RZ with other 28 isolates varied from 81.8%–99.6% at the nt level. LW showed the highest nt identity of 98.4% with isolate D8 from Japan. ND1 and ND2 had the highest nt similarities of 99.6% with each other. QZ shared the highest nt identity of 95.1% with the isolate Phy. Isolate RZ had the highest nt identity of 95.1% with isolate IA.

The genes 2a and 2b examined in our 5 CMV isolates shared nt identities of 69.6%–98.7% and 63.7%–98.8%, respectively, in comparison with the 28 CMV isolates from the GenBank. Specifically, LW shared the highest nt identities of 98.7% (2a gene) and 98.8% (2b gene) with ND2. ND1 shared the highest nt identities of 97.2% and 96.7% with ND2. QZ shared the highest nt identities of 98.6% and 98.8%, respectively with SD. RZ shared the highest nt identities of 98.4% and 98.2%, respectively with Tsh. Interestingly, the nt identities for 2a genes of 5 CMV isolates examined in this research were 91.7%–91.8%, while those for 2b genes were only 87.8%–88.1%.

When MP gene was considered, these 5 isolates shared nt identities of 78.9%–99.4% with other 28 CMV isolates. LW had the highest nt identity of 99% with SD. ND1 and ND2 had the highest nt identity of 99.4%. The identity of QZ and CTL is 97.3%. RZ shared the highest nt identities of 98.3% with Ri-8, Fny, and Y.

For CP gene, these 5 isolates shared identities of 75.2%–98.6% with other 28 CMV isolates. LW shared the highest nt identity of 98.6% with SD. ND1 and ND2 had the highest nt identity of 98.5%. Isolates QZ and TFN had the highest nt identity of 95.4%. Isolate RZ shared the highest nt identity of 98.3% with isolate Y.

Recombination analyses

Altogether 33 isolates were included for possible recombination analyses using methods of Bootscanning, Geneconv, Maxchi, Chimaera, Siscan, 3SEQ, and Lard assembled in software RDP3. Several isolates were predicted to have experienced some recombination events, but these events were supported only by less than four programs and the *P*-value was bigger than 1.0×10^{-6}. Therefore, these isolates were designated as "tentative" recombinants. Nevertheless, these isolates were re-checked using Phylipro program and no recombination event was confirmed.

Phylogenetic analyses

Phylogenetic trees constructed using NJ, ME, and UPGMA methods were identical in topology and only the trees constructed with NJ method were presented (Fig. 1). In the phylogenetic tree of CP gene, the examined 33 isolates were clustered to 3 groups (Fig. 1a). Group I was the largest group and consisted of 25 isolates. This group could be further

Table 2. Acc. Nos, host, and geographical location of CMV isolates

Isolate	Acc. No.			Host	Geographical location
	RNA1	RNA2	RNA3 (MP/ CP)		
LW	EU414792	EU414797	EU414787/EU414784	*N. tabacum*	China
ND1	EU414793	EU414798	EU414788/EU414785	*N. tabacum*	China
ND2	EU414794	EU414799	EU414789/EU414786	*P. hybrida*	China
QZ	EU414795	EU414800	EU414790/DQ916111	*N. tabacum*	China
RZ	EU414796	EU414801	EU414791/EF159146	*N. tabacum*	China
BX	DQ399548	DQ399549	DQ399550	*Pinellia ternata*	China
Ca	AY429434	AY429433	AY429432	*Arachis hypogaea*	China
Cb7	EF216866	DQ785470	EF216867	*Lycopersicon esculentum*	China
CS	AY429435	AY429436	AY429437	*Arachis hypogaea*	China
CTL	EF213023	EF213024	EF213025	*Brassica chinensis*	China
D8	AB179764	AB179765	AB004781	*Raphanus sativus*	Japan
Fny	D00356	D00355	D10538	*Cucumis melo*	USA
IA	AB042292	AB042293	AB042294	n.a.	Indonesia
Ixora	U20220	U20218	U20219	*L. esculentum*	Philippines
LS	AF416899	AF416900	AF127976	*Lactuca sativa*	USA
LY	AF198101	AF198102	AF198103	*Lupinus angustifolius*	Australia
Ly2	AJ535913	AJ535914	AJ296154	*Lilium longiflorum*	Korea
MF	AJ276479	AJ276480	AJ276481	*Melandryum firmum*	Korea
Ns	AJ580953	AJ511989	AJ511990	*N. glutinosa*	Hungary
NT9	D28778	D28779	D28780	*L. esculentum*	China
P1-1	AM183114	AM183115	AM183116	*L. esculentum*	Spain
Pepo	AB124834	AB124835	AF103991	*Cucurbita pepo*	Japan
Phy	DQ402477	DQ412731	DQ412732	n.a.	China
Q	X02733	X00985	M21464	*Capsicum annuum*	Australia
Ri-8	AM183117	AM183118	AM183119	*L. esculentum*	Spain
Rs	AJ511988	AJ517801	AJ517802	*Raphanus sativus*	Hungary
S	Y10884	Y10885	U37227/AF063610	*Cucurbita pepo*	South Africa
SD	AF071551	D86330	AB008777	*N. tabacum*	China
TFN	Y16924	Y16925	Y16926	*L. esculentum*	Italy
TN	AB176849	AB176848	AB176847	*L. esculentum*	Japan
Trk7	AJ007933	AJ007934	L15336	*Trifolium repens*	Hungary
Tsh	EF202595	EF202596	EF202597	*L. esculentum*	China
Y	D12537	D12538	D12499	*N. tabacum*	Japan

n.a. = not applicable.

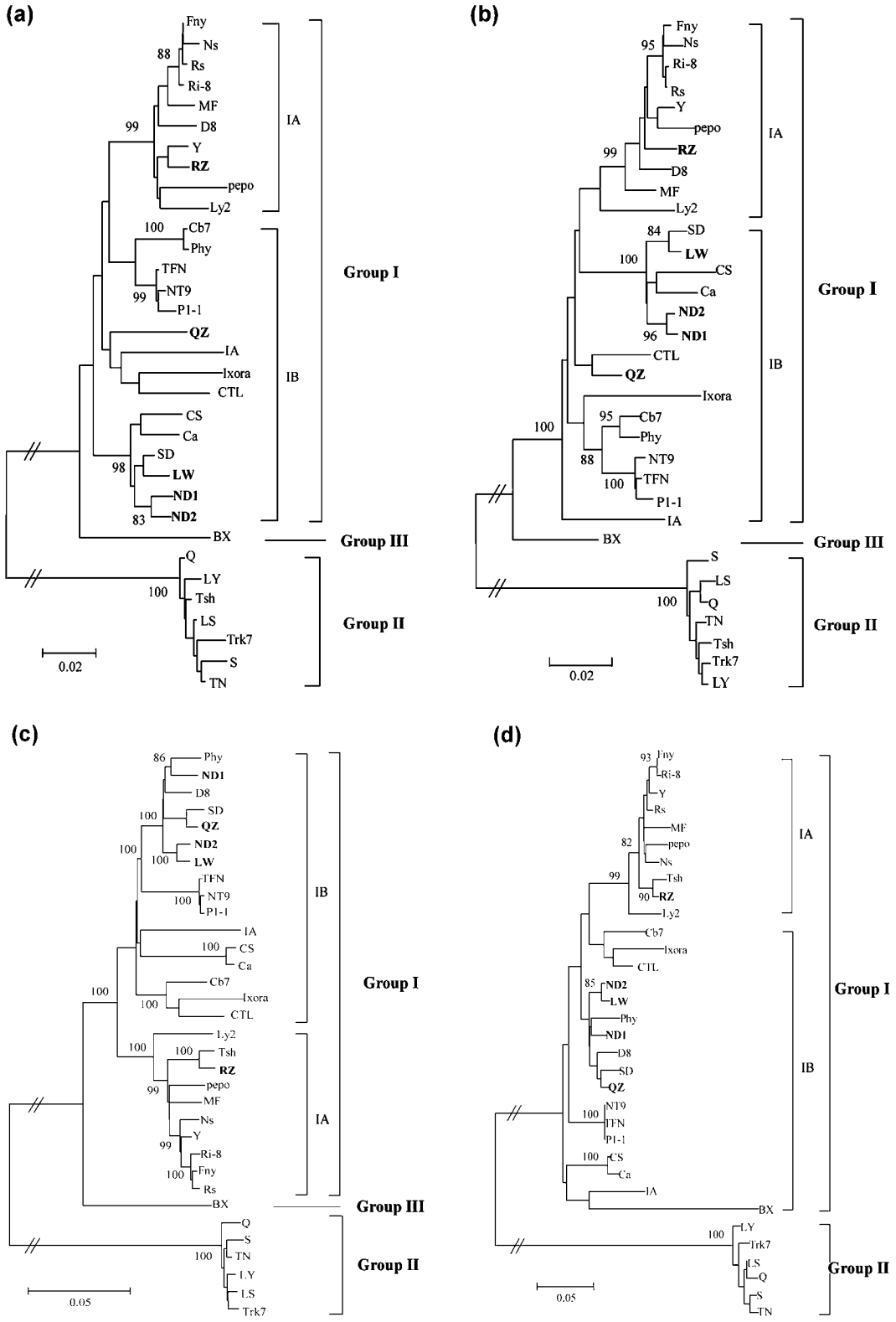

Fig. 1

Phylogenetic analysis of CP (a), MP (b), 2a (c), 2b (d), and 1a (e) genomic regions of CMV

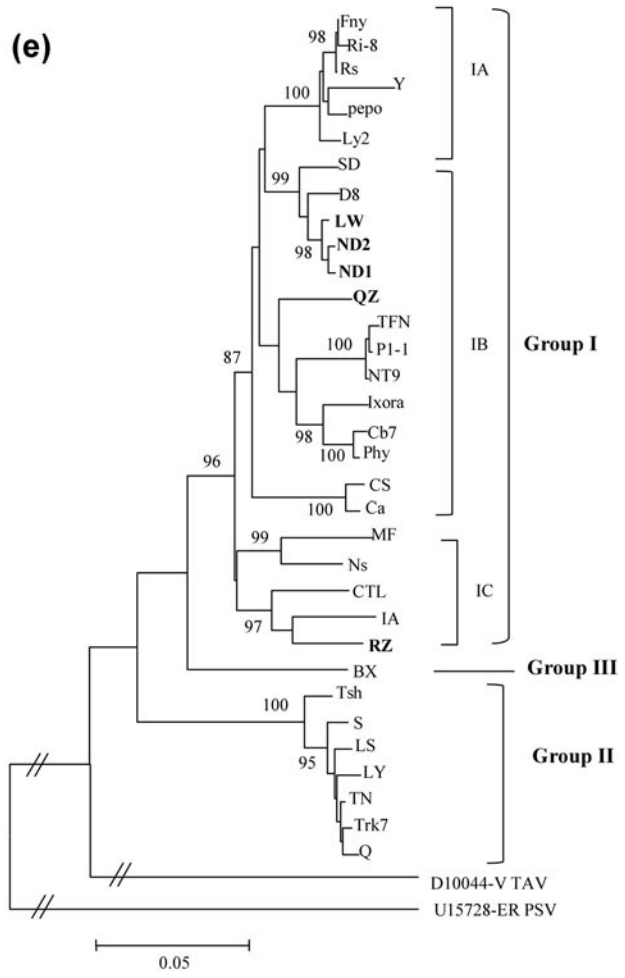

Fig. 1 – continued

divided into two subgroups corresponding to the previous subgroups IA and IB, respectively. Subgroup IA contained 10 isolates including the reference isolates Fny, MF, Y, and the isolate RZ examined in this research. Subgroup IB contained 15 isolates including the other 4 examined isolates LW, QZ, ND1, and ND2 and the reference isolates CTL, NT9, SD, and TFN. Group II contained 7 isolates including reference isolates LY, Q, S, and Trk7 and corresponded to the former established subgroup II. Group III contained only isolate BX. The isolates of subgroup IB were highly variable and should be divided into several subgroups. The results of phylogenetic analysis of CP gene were not related to the geographical or host origin.

According to the phylogenetic results of MP gene, the CMV isolates were divided into three groups that were identical with the grouping obtained with CP gene (Fig. 1b). RZ was clustered to the subgroup IA and other four isolates clustered to the subgroup IB. LW, ND1, and ND2 still formed a sub-branch with Chinese isolates CS, Ca, and SD. QZ formed a sub-branch with isolate CTL, but isolates IA and Ixora were clustered to other sub-branches. Similarly, the isolates of subgroup IB could be further divided into several subgroups. However, the clustering results were different from those obtained with CP gene.

Isolate BX formed a III group according to the results of phylogenetic analysis of 2a gene, while it clustered to the subgroup IB according to results of 2b gene (Fig. 1c and d). In both cases, RZ belonged to subgroup IA and formed a sub-branch with Tsh, which was clustered to the subgroup II according to CP and MP genes (Fig. 1a and b). The isolates LW, ND1, ND2, and QZ belonged to the subgroup IB. The phylogenetic results also implied the further division of subgroup IB isolates into more subgroups.

In the phylogenetic tree of 1a gene (Fig. 1e), the 33 CMV isolates were clustered to three groups that were identical to CP, MP, and 2a clustering. The isolates of group I could be divided into three subgroups. Besides the previously reported subgroups IA and IB, a third subgroup, signed IC was proposed. Subgroup IC included the isolates MF, Ns, CTL, IA, and RZ. The other four isolates characterized in this research besides RZ were clustered to the subgroup IB. The 1a gene of RZ was clustered to the newly proposed subgroup IC, while the other genes of isolate RZ belonged to the subgroup IA. Therefore, RZ was an inter-subgroup reassortant.

We also constructed phylogenetic trees with amino acid sequence of each isolate (data not shown). The results were consistent with those obtained with nucleotide sequences in topology and supported the conclusion that CMV isolates clustered into three groups and the group I can be divided into two or three subgroups.

Genetic diversity

To understand further the direction and degree of the selective constraints imposed on different coding regions, the ratio of the average numbers of non-synonymous and synonymous substitutions per site (d_N/d_S ratio) was estimated using PBL method.

The d_N/d_S ratios for 1a, 2a, MP, and CP genes of the isolates were less than 1.0 (Table 3) indicating that they were all subjected to negative (or purifying) selection. The d_N/d_S ratio for the 2b gene was >1 (1.1029), what meant that 2b gene was under positive or diversifying selection and the protein 2b was considerably more tolerant for amino acid changes. The results of d_N/d_S ratios also implied that the selection pressure imposed on the 2a gene of CMV was three times of that for MP and CP genes.

In order to analyze whether the host adaptation and geographic distribution could constitute selection factors,

Table 3. Nucleotide diversities for different genes of CMV

Gene	d_N	S.E.[a]	d_S	S.E.[a]	d_N/d_S
ORF1 (1a)	0.038	0.005	0.396	0.035	0.09596
ORF2 (2a)	0.112	0.005	0.269	0.016	0.4164
ORF3 (2b)	0.225	0.030	0.204	0.041	1.1029
ORF4 (MP)	0.05	0.007	0.361	0.035	0.1385
ORF5 (CP)	0.055	0.008	0.410	0.042	0.1341

[a]S.E. = standard error; d_N = non-synonymous substitution; d_S = synonymous substitution.

nucleotide diversity values were estimated between and within CMV subpopulations considering a subpopulation as the group of isolates that were originally collected from a given host species or a given geographical area. Diversity values between subpopulations were similar in their order of magnitude to those corresponding to within-subpopulation diversity values (data not shown). This indicated that there was no evidence of selection associated with the host adaptation and geographic distribution.

Discussion

In this research, we characterized the genome regions encoding five proteins of 5 CMV isolates from China among which the isolate RZ was a natural inter-subgroup reassortant. Analyses of the sequences of Chinese isolates with those of other 28 isolates from the GenBank revealed that CMV could be genetically divided into two groups for 2b gene and into three groups I, II, and III for other genes. The group I could be further divided into two (or three for 1a gene) subgroups. There was no evidence of selection associated with host adaptation and geographic distribution. The 2b gene of CMV was under positive (or diversifying) selection, while the other genes were under negative (or purifying) selection.

The phylogenetic analyses of CMV have been conducted with either nucleotide or amino acid sequences (Chaumpluk *et al*., 1996; Roossinck *et al*., 1999; Roossinck, 2002; Lin *et al*., 2004; Chen *et al*., 2007; Balaji *et al*., 2008). We analyzed studied isolates with both nucleotide and amino acid sequences and we obtained consistent results. CMV isolates could be classified into three groups according to the phylogenetic results of 1a, 2a, CP, and MP genes, but into two groups for the 2b gene. Isolates of group I could be divided into two subgroups according to CP, MP, 2a, and 2b genes, and into three subgroups according to the 1a gene. This is consistent with the hypothesis that different RNAs of CMV might have independent evolutionary histories (White *et al*., 1995; Roossinck, 2002). The isolates belonging to the subgroup IB could have been further divided. However, the clustering of different isolates was not consistent and therefore we did not admit the further division of the group.

Recombination and reassortment are major evolutionary forces for viruses including CMV (White *et al*., 1995; Bonnet *et al*., 2005; de Wispelaere *et al*., 2005). However, we did not find any evidence for recombination in the isolates analyzed in this research. The same conclusion was drawn by Lin *et al*. (2004). The 1a gene of isolate RZ formed a sub-branch with isolate IA, which belongs to subgroup IB according to the previous classification (Roossinck *et al*., 2002; Lin *et al*., 2003; Chen *et al*., 2007). In the present study, it was included in the newly proposed subgroup IC. But other genes of isolate RZ came from ancestor of subgroup IA. Therefore, whether according to the previous classification or results of the present study, the isolate RZ was an inter-subgroup reassortant. The fraction of reassortants between CMV subgroups IA and IB was larger than that of the reassortants between groups I and II. To date, there had been only two isolates of similar kind of reassortants reported (Bonneta *et al*., 2005; Chen *et al*., 2007). This phenomenon might be associated with the incidence of the genetic divergence between different subgroups (Chen *et al*., 2007). However, the isolates of group II were scarce in China (Xi *et al*., 2006) and it was understandable that reassortants between groups I and II were rarely detected. Reassortment might generate new isolates, expand host range, and recover isolates bearing deleterious mutations. Reassortants between highly divergent isolates might be competent unfavorably and would be eliminated soon due to the strong natural selection (Bonneta *et al*., 2005).

Some viruses could be divided into definite groups according to their host or geographical origin. (Chen *et al*., 2002; Ohshima *et al*., 2002; Tian *et al*., 2007). However, the phylogenetic results of CMV isolates were not correlated with the host or geographical origin, what was consistent with the results of Lin *et al*. (2003).

For most virus genes, the d_N/d_S ratio was smaller than 1.0, what was consistent with the negative selection against protein change. Although CP protein of CMV had little specific interaction with the host or aphid vectors, the MP protein interacted specifically with plasmodesmata and host factors to mediate the movement and host determination (Hong *et al*., 2007). However, d_N/d_S ratios for MP and CP genes were below 1.0 and almost identical. On the other hand, d_N/d_S ratios for 2a and 2b genes were quite different. The 2a gene of CMV was under negative selection, while the 2b gene was under positive selection. Surprisingly, 2b genes of CMV isolates from California were under negative selection (Lin *et al*., 2004). We analyzed the isolates with different host and geographical origins, while Lin *et al*. (2004) analyzed a subpopulation from California that shared low diversity and close evolutionary relationship. It is noteworthy that the d_N/d_S ratios were slightly higher than 1.0, when the whole

gene was considered. Conversely, when the gene part encoding 81–93 aa was selected for analysis, the d_N/d_S ratios was 5.2 (Lin *et al*., 2004). The positive selection in 2b gene would confer a higher protein tolerance against mutation. The protein could be more flexible in the interaction with other proteins and able to fulfill its functions in virus movement, symptom expression, host determination, and suppression of RNA silencing. This result also supported the hypothesis that 2b was a relatively recently emerging gene that is still undergoing intensive evolution (Lin *et al*., 2004).

The analyses of the genomic sequences of 33 CMV isolates revealed some distinctive characteristics that were important for their proper grouping. Since the recombination and reassortment play a significant role in the evolution of viruses with multiple genomes, one cannot obtain precise phylogenetic results based on the analysis of a single gene (Roossinck, 2002; Lin *et al*., 2004; Balaji *et al*., 2008). The phylogenetic and genetic diversity data obtained in this study provided insight into the evolution, gene function, and plant-virus interaction of CMV and related viruses.

Acknowledgement. Xiangdong Li received grants from the Program for New Century Excellent Talents in Universities (NCET-07-0520), Shandong Natural Scientific Foundation (Z2007D04), Special Research Fund for the Doctoral Program of Higher Education (20050434003, 20080434006), and Shandong Tobacco Company.

References

Balaji S, Bhat AI, Eapen SJ (2008): A phylogenetic reexamination of Cucumber mosaic virus isolates based on 1a, 2a, 3a, and 3b proteins. Indian J. Virol. 19, 27–33.

Boni MF, Posada D, Feldman MW (2007): An exact nonparametric method for inferring mosaic structure in sequence triplets. Genetics 176, 1035–1047. doi:10.1534/genetics.106.068874

Bonnet J, Fraile A, Sacristán S, Malpica JM, García-Arenal F (2005): Role of recombination in the evolution of natural populations of Cucumber mosaic virus, a tripartite RNA plant virus. Virology 332, 359–368. doi:10.1016/j.virol.2004.11.017

Brigneti G, Voinnet O, Li WX, Ji LH, Ding SW, Baulcombe DC (1998): Viral pathogenicity determinants are suppressors of transgene silencing in Nicotiana benthamiana. EMBO J. 17, 6739–6746. doi:10.1093/emboj/17.22.6739

Canto T, Prior DAM, Hellwald KH, Oparka KJ, Palukaitis P (1997): Characterization of cucumber mosaic virus IV. Movement protein and coat protein are both essential for cell-to-cell movement of cucumber mosaic virus. Virology 237, 237–248. doi:10.1006/viro.1997.8804

Chaumpluk P, Sasaki N, Nakajima H, Nagano I, Nakamura K, Suzuki K, Mise K, Inouye M, Okuno T, Furusawa I (1996): Six new subgroup I members of Japanese cucumber mosaic virus as determined by nucleotide sequence analysis of RNA3's cDNA. Ann. Phytopathol. Soc. Jpn. 62, 40–44.

Chen J, Chen JP, Adams MJ (2002): Variation between Turnip mosaic virus isolates in Zhejiang Provinces, China and evidence for recombination. J. Phytopathol. 150, 142–145. doi:10.1046/j.1439-0434.2002.00731.x

Chen Y, Chen J, Zhang H, Tang X, Du Z (2007): Molecular evidence and sequence analysis of a natural reassortant between Cucumber mosaic virus subgroup IA and II strains. Virus Genes 35, 405–413. doi:10.1007/s11262-007-0094-z

de Wispelaere M, Gaubert S, Trouilloud S, Belin C, Tepfer M (2005): A map of the diversity of RNA3 recombinants appearing in plants infected with Cucumber mosaic virus and Tomato aspermy virus. Virology 331, 117–127. doi:10.1016/j.virol.2004.10.017

Ding SW, Anderson BJ, Haase HR, Symons RH (1994): New overlapping gene encoded by the cucumber mosaic virus genome. Virology 198, 593–601. doi:10.1006/viro.1994.1071

Ding SW, Li W X, Symons R H (1995): A novel naturally occurring hybrid gene encoded by a plant RNA virus facilitates long distance virus movement. EMBO J. 14, 5762–5772.

Ding SW, Shi BJ, Li WX, Symons RH (1996): An interspecies hybrid RNA virus is significantly more virulent than either parental virus. Proc. Natl. Acad. Sci. USA 93, 7470–7474. doi:10.1073/pnas.93.15.7470

Gibbs MJ, Armstrong JS and Gibbs AJ (2000): Sister-Scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. Bioinformatics 16, 573–582. doi:10.1093/bioinformatics/16.7.573

Hayes RJ, Buck KW (1990): Complete replication of a eukaryotic virus RNA in vitro by a purified RNA-dependent RNA polymerase. Cell 63, 363–368. doi:10.1016/0092-8674(90)90169-F

Holmes EC, Worobey M, Rambaut A (1999): Phylogenetic evidence for recombination in Dengue virus. Mol. Biol. Evol. 16, 405–409.

Hong JS, Ohnishi S, Masuta C, Choi JK, Ryu KH (2007): Infection of soybean by cucumber mosaic virus as determined by viral movement protein. Arch. Virol. 152, 321–328. doi:10.1007/s00705-006-0847-3

Kumar S, Tamura K, Nei M (2004): MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Briefings in Bioinformatics 5, 150–163. doi:10.1093/bib/5.2.150

Li WH (1993): Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J. Mol. Evol. 36, 96–99. doi:10.1007/BF02407308

Lin HX, Rubio L, Smythe A, Jiminez M, Falk BW (2003): Genetic diversity and biological variation among California isolates of Cucumber mosaic virus. J. Gen. Virol. 84, 249–258. doi:10.1099/vir.0.18673-0

Lin HX, Rubio L, Smythe AB, Falk BW (2004): Molecular population genetics of Cucumber mosaic virus in California: evidence for founder effects and reassortment. J. Virol. 78, 6666–6675. doi:10.1128/JVI.78.12.6666-6675.2004

Martin DP, Williamson C, Posada D (2005): RDP2: recombination detection and analysis from sequence alignments. Bioinformatics 21, 260–262. doi:10.1093/bioinformatics/bth490

Maynard Smith J (1992): Analyzing the mosaic structure of genes. J. Mol. Evol. 34, 126–129.

Nei M, Kumar S (2000): Molecular Evolution and Phylogenetics. Oxford University Press, New York.

Ohshima K, Yamaguchi Y, Hirota R, Hamamoto T, Tomimura K, Tan Z, Sano T, Azuhata F, Walsh JA, Fletcher J, Chen J, Gera A, Gibbs A (2002): Molecular evolution of Turnip mosaic virus: evidence of host adaptation, genetic recombination and geographical spread. J. Gen. Virol. 83, 1511–1521.

Padidam M, Sawyer S, Fauquet CM (1999): Possible emergence of new geminiviruses by frequent recombination. Virology 265, 218–225. doi:10.1006/viro.1999.0056

Palukaitis P, García-Arenal F (2003): Cucumoviruses. Adv. Virus Res. 62, 241–323. doi:10.1016/S0065-3527(03)62005-1

Palukaitis P, Roossinck MJ, Dietzgen RG (1992): Cucumber mosaic virus. Adv. Virus Res. 41, 281–348. doi:10.1016/S0065-3527(08)60039-1

Pamilo P, Bianchi NO (1993): Evolution of the Zfx and Zfy genes: rates and interdependence between the genes. Mol. Biol. Evol. 10, 271–281.

Posada D, Crandall KA (2001): Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. Proc. Natl. Acad. Sci. USA 98, 13757–13762. doi:10.1073/pnas.241370698

Roossinck M J, Zhang L, Hellwald K (1999): Rearrangement in the nontranslated region and phylogenetic analyses of Cucumber mosaic virus RNA3 indicate radial evolution of three subgroups. J. Virol. 73, 6752–6758.

Roossinck MJ (2001): Cucumber mosaic virus, a model for RNA virus evolution. Mol. Plant Pathol. 2, 59–63. doi:10.1046/j.1364-3703.2001.00058.x

Roossinck MJ (2002): Evolutionary history of Cucumber mosaic virus as deduced by phylogenetic analyses. J. Virol. 76, 3382–3387. doi:10.1128/JVI.76.7.3382-3387.2002

Salminen MO, Carr JK, Burke DS, McCutchan FE (1995): Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. AIDS Res. Hum. Retrov. 11, 1423–1425. doi:10.1089/aid.1995.11.1423

Shi B-J, Symons RH, Palukaitis P (2008) The cucumovirus 2b gene drives selection of inter-viral recombinants affecting the crossover site, the acceptor RNA and the rate of selection. Nucleic Acids Res. 36, 1057–1071. doi:10.1093/nar/gkm1036

Suzuki M, Kuwata S, Kataoka J, Masuta C, Nitta N, Takanami Y (1991): Functional analysis of deletion mutants of cucumber mosaic virus RNA3 using an in vitro transcription system. Virology 183, 106–113. doi:10.1016/0042-6822(91)90123-S

Thompson JD, Higgins DG, Gibson TJ (1994): CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680. doi:10.1093/nar/22.22.4673

Tian Y, Zhu X, Liu J, Yu X, Du J, Kreuze J, Li X (2007): Molecular characterization of the 3'-terminal region of Turnip mosaic virus isolates from Eastern China. J. Phytopathol. 155, 333–341. doi:10.1111/j.1439-0434.2007.01236.x

Tomitaka Y, Ohshima K (2006): A phylogeographical study of the Turnip mosaic virus population in east Asia reveals an 'emergent' lineage in Japan. Mol. Ecol. 15, 4437–4457. doi:10.1111/j.1365-294X.2006.03094.x

Weiller GF (1998): Phylogenetic profiles: a graphical method for detecting genetic recombinations in homologous sequences. Mol. Biol. Evol. 15, 326–335.

White PS, Morales FJ, Roossinck MJ (1995): Interspecific reassortment in the evolution of a cucumovirus. Virology 207, 334–337. doi:10.1006/viro.1995.1088

Xi D, Lin H, Xiang B (2006): Subgrouping identification and strain differentiation of two Cucumber mosaic virus isolates. Acta Phytopathol. Sin. 36, 232–237.

Yu X, Lan Y, Wang H, Liu J, Zhu X, Valkonen J, Li X (2007) The complete genomic sequence of Tobacco vein banding mosaic virus and its similarities with other potyviruses. Virus Genes 35, 801–806. doi:10.1007/s11262-007-0135-7