UDC 578.233

O.I. Gordeyeichik, eng., I.C. Shcherbatenko, D. Sci.

SEARCHING FOR SIMILAR NUCLEOTIDE SITES IN GENOMIC SEQUENCES OF PLANT VIRUSES

Zabolotny Institute of Microbiology and Virology of the NAS of Ukraine, Kyiv.

Computer database search for similar nucleotide sites in genomes of plant pathogenic viruses has been conducted by subsequent comparison of pairs of genome sequences with increasing shift value of their starting postions.

Introduction.The sequence similarity of viral genomes or some nucleotide sites is quantified as persentage of nucleotide sequence identity and widely used for the elucidation of the evolutionary histories of viruses, detecting conserved functionally importent sequences and secondary structure elements as well as comparing virus relationship and taxonomic groupings [1,9 13,14,15].

Similarity searches are commonly performed with numerous available programs by sequence alignment. Although a variety of algorithms and improvements of the widely used programs and new software applications for sequence alignment have been developed [2, 3, 6, 7, 10, 16], sequence similarity analysis still presents certain difficulties [11, 12, 17].To overcome this problem new algorithms, alternative to alignment, are of particular interest [4, 18].The objective of this study was the computational search for similar nucleotide sites in genomes of plant viruses by successive disposition of two genomic sequences with increasing displacement of their initial positions.

Materials and methods. *Sequences***.** Tests were run on 3 random nucleotide sequences and 189 genomic sequences of (+)ssRNA-containing plant viruses from Tombusviridae (Avena-, Aureus-, Carmo-, Necro-, Tombusvirus), Flexiviridae (Allexi-, Capillo-, Carla-, Fovea-, Potex-, Tricho-, Vitivirus), Luteoviridae (Luteo-, Polerovirus), Tymoviridae (Tymovirus) and Potyviridae (Potyvirus) families as well as Tobamovirus and Sobemovirus genera. All the genomic sequences used in this study were derived from the DDBJ, EMBL, and GenBank nucleotide sequence databases. The 6400 nucleotide long random sequences were generated using the random number generator: basic function int(RND^{x}4+1), producing a random number between 1 and 4 . The standard nucleotides (A, T, G and C) were assigned to 1, 2, 3 and 4, accordingly.

Similarity searches. Searches of the similar nucleotide sites in viral genomes was performed by successive disposition of two sequences with increasing displacement of their initial position: 1/1 (displacement = 0); 1/2 (displacement = +1); 2/1 (displacement = -1); 1/3 (displacement = +2); 3/1 (displacement = -2). 1/maxd; maxd/1 (maxd (maximal displacement) = genome length - site length). In all dispositions of sequences were revealing the similar nucleotide sites, which length and persentage of sequence identity were not less then preliminary user-specified. Searching results:number, length and genomic positions of the similar nucleotide sites were graphical visualizing and tabular displaing.

Software application. A simple program we have termed Similar% was written at qBASIC and tested on two random sequence, in which some similar nucleotide sites was inserted. Extensive simulation studies show that the program used: 1) is able to find all similar sites inserted; 2) the length, positions and identity persentage of finding sites are equal to that of inserting ones. The values of the mean of sites quantity and length were obtained from at least 3 to 88 pairs of genomic sequences.

Results and discussion. *Revealing of non-random nucleotide coincidence in sequences*. To determine the parameters of non-random sequence similarity the dependence between nucleotide sites length and nucleotide coincidence in random sequences was investigated. It was shown that 5-nucleotide sites with 100% identity are appeared in random sequences from 8577 to 8711 times, 10-nucleotide sites - from 8 to 15 times, 12-nucleotide sites - from 0 to 2 times, and 15-nucleotide sites are not appeared (Tabl. 1). So, parameter limits of non-random similarity of short sites are the site length not less then 15 nucleotides and nucleotide coincidence (sequence identity) not less then 100%. The limit of non-random similarity for 20-nucleotide sites is not less then 90% sequence identity, 30-nucleotide - 70%, 40-nucleotide - 60%, 50-nucleotide - 58%, 100-nucleotide - 46%, 500-nucleotide - 34%, 1000-nucleotide - 32%, 2000-nucleotide - 32. So far as non-random similarity of the short sites sharply depends on their length (15 nucleotides - 100%, 20 nucleotides - 90%), just as the long sites - on their identity persentage (1000 nucleotides - 32%, 2000 nucleotides - 30%), the 40 nucleotide site length and 60% sequence identity was chosen as starting user-specified parameter for simile sites revealing.

Table 1. **The number of similar sites in random nucleotide sequences depending on the site length and nucleotide coincidence**

Site length, nucleotides	Nucleotide coincidence, %	Pairs of random sequences *		
		rand1/rand2	rand1/rand3	rand2/rand3
5	100	8711	8577	8654
10	100	8	15	10
12	100	0	2	0
15	100	0	0	0
	70	18	14	21
20	80	0	1	0
	90	0	0	0
	60	5	11	7
30	65	1	2	0
	70	0	0	0
	54	17	28	19
40	58	1	1	1
	60	0	0	0

	52	9	17	13
50	56	0	2	0
	58	0	0	0
	42	18	17	18
100	44	5	1	3
	46	0	0	0
	30	216	178	191
500	32	11	14	8
	34	0	0	0
	28	227	237	229
1000	30	2	2	0
	32	0	0	0
	26	811	765	802
2000	28	10	7	4
	30	0	0	0

* The length of random sequences rand1, rand2 and rand3 is 6400 nucleotides

Searching for the similar nucleotide sites in viral genomes. The similar nucleotide sites was firstly searching in the genomes of viruses from 16 genera, wich had not less then 6 sequences available. Tests were carried out on 5 pairs of samples - one sequence from six available was compared with the rest five ones. (Tabl. 2). It was determined that the same genus viruses have a large clusters of similar sites varying in the mean quantity from 5.2 (Sobemovirus) to 20.8 (Allexivirus), and in the mean length - from 938.4 (Potexvirus) to 4912.5 (Foveavirus). The maximal length of similar sites reaches 1206 - 8743 nucleotides, and their maximal quantity differ from 9 to 23.

The similar nucleotide sites of viruses from different genera but the same family differ in mean size from 7.0 to 12.1 nucleotides, and in mean length - from 677.8 to 836.3 (Tabl. 3). The maximal quantity of the similar sites reaches 13 - 14, and the maximal length varying from 1509 to 5451 nucleotides.

Table 2. **The similar nucleotide sites in genomes of plant viruses from the same genus**

Virus genus	No. of sites		Site length	
	mean*	maximal	mean*	maximal
Allexivirus	20,8	23	2315,0	3391
Carlavirus	18,8	22	1399,4	1584
Tobamovirus	17,4	20	1055,8	1404
Potyvirus	17,0	19	1198,8	1424
Potexvirus	14,6	23	938,4	1206
Vitivirus	13,7	20	2972,3	6812
Luteovirus	12,0	14	2043,8	4710
Tymovirus	11,6	16	2187,0	3264
Polerovirus	10,6	13	1219,2	2197
Foveavirus	10,3	19	4912,5	8743
Trichovirus	9,4	13	3675,6	5283
Aureusvirus	8,3	11	1559,0	2692
Capillovirus	5,8	14	3948,4	6478
Necrovirus	5,6	9	1799,0	3683
Carmovirus	5,4	9	1820,4	4186
Sobemovirus	5,2	10	1027,5	1606

*- The mean values from 5 virus pairs of earch genus

Table 3. **The similar nucleotide sites in genomes of plant viruses from the different genera**

Families and genera	Virus pair		No. of site		Site length	
	No. of tested	% with similar sites	mean	max	mean	max
Tombusviridae	29	100	7,0	13	767,0	5451
Flexiviridae	69	100	12,1	24	836,3	2567
Luteoviridae	43	100	7,5	14	677,8	1509
Luteovirus/Tombusviridae	48	100	3,8	7	255,4	438
Luteoviridae/ Sobemovirus	30	100	4,4	8	281,1	407
Flexiviridae/Luteoviridae	88	58,7	0,9	5	41,4	227
Tymovirus/Flexiviridae	31	93,5	1,7	4	89,9	224
Tymovirus/Potexvirus	16	100	4,2	8	227,8	468

Comparisons of the similar sites in genomic sequences showed that the lisianthus necrosis virus from Necrovirus genus is more similar with Tombusvirus members (fig.1, samples 1,2) then with Necrovirus ones (fig.1, samples 3-8). In addition to sequence similarity the genome length and ORFs of this virus are also typical for tombusviruses (fig. 2). As a result, there are worthy arguments for taking the lisianthus necrosis virus to Tombusvirus genus.

Fig. 1. Similarity of Lisianthus necrosis virus with tombusviruses (1, 2) and necroviruses (3, 8). Tombusviruses: 1 - carnation Italian ringspot virus, 2 - tomato bushy stunt virus. Necroviruses: 3 - beet black scorch virus, 4 - leek white stripe virus, 5 - olive latent virus 1, 6 - olive mild mosaic virus, 7 - tobacco necrosis virus A, 8 - tobacco necrosis virus D

Viruses of different genera from different families usually have no similar sites, but in some cases they can have some similarity: clearly expressed clasters of a relatively long nucleotide sites located in a certain genomic region and numerous short sites scattered along the viral genome (fig. 3); diffuse clasters of a short sites located in a several genomic positions (fig. 4) and a little clasters or a single nucleotide sites located in a certain genome region (fig. 5).

Fig. 2. The ORFs of Lisianthus necrosis virus (3) and tombusviruses (1, 2) and necrovirus (4, 5. Tombusviruses: 1- carnation Italian ringspot virus, 2 - tomato bushy stunt virus. Necroviruses: 4 - tobacco necrosis virus A, 5 - olive latent virus 1. Gray lines - genome length

These experiments reveale the phylogenetic relationships among some viruses from Luteoviridae, Tombusviridae, Flexiviridae, Sobemovirus, Tymovirus and Tobamovirus. Our results are in accordance with the comparisons of the viral genomes [8], ORFs [15], coat and replications proteins [5, 6], tRNA-like structures [9] and a cap-independent translation element [14].

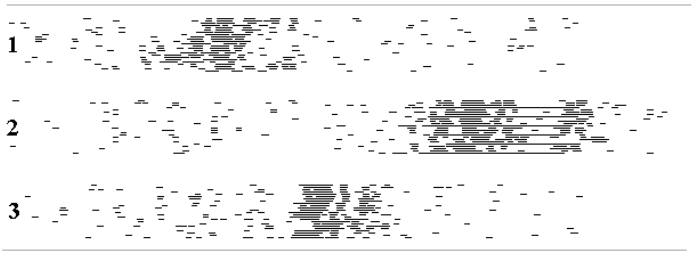

Fig. 3. Similar nucleotide sites of tombusviruses and luteoviruses (1); poleroviruses and luteoviruses (2); poleroviruses and sobemoviruses. Tombusviruses: carnation Italian ringspot virus, tomato bushy stunt virus, lisianthus necrosis virus, cucumber leaf spot virus, pothos latent virus, beet black scorch virus. Poleroviruses: beet chlorosis virus, beet mild yellowing virus, beet western yellows virus, carrot red leaf virus, cereal yellow dwarf virus, turnip yellows virus. Luteoviruses: barley yellow dwarf virus - GAV, barley yellow dwarf virus - PAS, bean leafroll virus, soybean dwarf virus. Sobemoviruses: ryegrass mottle virus, rice yellow mottle virus, cocksfoot mottle virus, southern bean mosaic virus

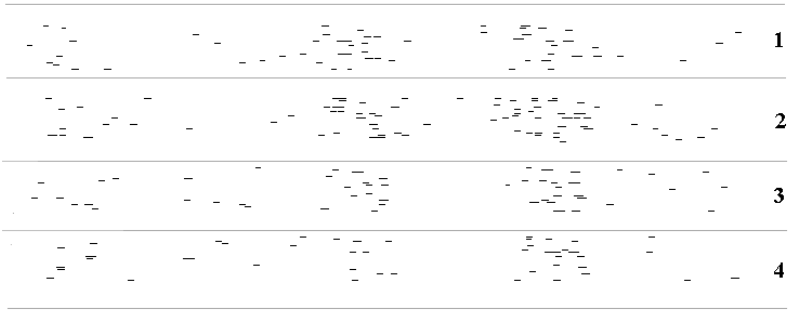

Fig. 4. Similar nucleotide sites of 22 potexviruses and anagyris vein yellowing virus (1); nemesia ring necrosis virus (2); plantago mottle virus (3); scrophularia mottle virus (4). Potexviruses: Potato Virus X, Alstroemeria virus X,Clover yellow mosaic virus, Pepino mosaic virus, Strawberry mild yellow edge virus, Scallion virus X, Foxtail mosaic virus, Potato aucuba mosaic virus, Cactus virus X, Bamboo mosaic virus, Zygocactus virus X, White clover mosaic virus(WCIMV1), White clover mosaic virus (WCIMV2), Cymbidium mosaic virus, Papaya mosaic virus, Cassava common mosaic virus, Foxtail mosaic virus, Narcissus mosaic virus, Plantago asiatica mosaic virus, Hydrangea ringspot virus,Lily virus X, Opuntia virus X

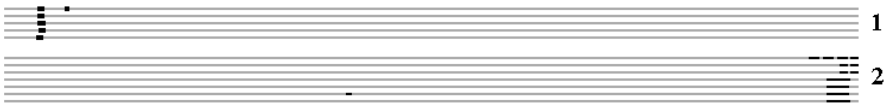

Fig. 5. Similarity of nemesia ring necrosis virus with allexiviruses (1) and tobamoviruses (2). Allexiviruses: garlic virus A, garlic virus C, garlic virus E, garlic virus X, shallot virus X. Tobamoviruses: odontoglossum ringspot virus, tobacco mosaic virus (vulgare), tomato mosaic virus, crucifer tobamovirus, tobacco mosaic virus (crucifer), ribgrass mosaic virus, turnip vein-clearing virus

Of particular interest is the location of the similar nucleotide sites in viral genomes. The distances between genomic positions of the all or the majority similar sites are equal to x.0 or x.33 or x.66 triplets (tabl. 4), which are equal to 3x, 3x+1 or 3x+2 nucleotides (tabl. 5), where x is an integer number from 0 to 300 - 500. The results suggest that the similar nucleotide sites represent an in-frame clusters of conserved sequence elements encoding some functionally important components. As such, the distances between the similar sites in the clusters may only be equal to an integer number of triplete codones, suggesting that any frameshift between similar sites is letal for virus. A large differences in the distances between the similar sites equal to 3x nucleotides might be the result of insertions/deletions/duplications/recombinations without frameshifting, as 3x+1 and 3x+2 nucleotide distances are due to additional non-integer triplete distances between starting codones of the genes containing the clusters of similar sites.

Table 4. **The distances between positions of similar nucleotide sites in genomes of two potexviruses**

Site length	Pos1*	Pos2**	Distances, nucleotides	Distances, triplets
62	815	834	19	6,3
72	1896	2704	808	269,3
65	2084	2898	814	271,3
80	2277	3091	814	271,3
60	2442	3262	820	273,3
259	2789	3612	823	274,3
344	3394	4217	823	274,3
69	4953	5820	867	289,0
65	4207	5080	873	291,0

*Pos1 - positions of similar nucleotide sites in genome of foxtail mosaic virus

**Pos2 - positions of similar nucleotide sites in genome of Scallion virus X

Table 5. **Displacement of the distances between positions of similar nucleotide sites in genomes of tobamo- and potyviruses**

Virus pair	No. of similar sites			
	All	with displacement		
		0	1	2
Tobamovirus genus				
Tobacco mild green mosaic virus / Cucumber green mottle mosaic virus	19	0	17	2
Pepper mild mottle virus / Ribgrass mosaic virus	14	2	8	4
Tomato mosaic virus / Crucifer tobamovirus	16	1	3	12
Cucumber fruit mottle mosaic virus / Odontoglossum ringspot virus	20	2	4	14
Potyvirus genus				
Potato virus A / Potato virus Y	19	0	17	2
Scallion mosaic virus / Sugarcane mosaic virus	18	1	16	1
Bean common mosaic virus / Cocksfoot streak virus	17	17	0	0
Soybean mosaic virus / Yam mosaic virus	16	1	14	1

* Displacement is equal to: or integer number of triplets (3x nucleotides, displacement 0), or 3x+1 nucleotides (displacement 1), or 3x+2 nucleotides (displacement 2)

Overall, we tested a new method for the computational search of the similar nucleotide sites in genomes by successive disposition of two genomic sequences with increasing displacement of their initial positions and revealing in all dispositions the similar nucleotide sites, which length and persentage of sequence identity were not less then preliminary user-specified.

We have determined a parameter limits of non-random nucleotide coincidence in sequences as well as the presence and localisation peculiarity of similar nucleotide sites in viral genomes. Our results also demonstrate that the lisianthus necrosis virus is the more closely related to tombusviruses then to necroviruses.

The computational search used provides an alternative sequence similarity analysis without a time-consuming aligning procedure. The method can be used as a simple, fast and obvious way for comparing sequence relatedness with graphical visualization of the similar nucleotide or aminoacid sites.

1. Aus dem Siepen M., Pohl J.O., Koo B.J., Wege C., Jeske H. Poinsettia latent virus is not a cryptic virus, but a natural polerovirus-sobemovirus hybrid // Virology.- 2005.- 336, N 2.- P. 240-250. 2. Catanho M., Mascarenhas D., Degrave W., de Miranda A.B. BioParser: a tool for processing of sequence similarity analysis reports // Appl. Bioinformatics.- 2006.- 5, N 1.- P. 49-53. 3. Collyda C., Diplaris S., Mitkas P., Maglaveras N., Pappas C. Enhancing the quality of phylogenetic analysis using fuzzy hidden Markov model alignments // Medinfo.- 2007.- 12, N 2.- P. 1245-1249. 4. Didier G., Debomy L., Pupin M., Zhang M., Grossmann A., Devauchelle C., Laprevotte I. Comparing sequences without using alignments: application to HIV/SIV subtyping // BMC Bioinformatics.- 2007.- 8.- P.1. 5. Ding S., Keese P., Gibbs A. The nucleotide sequence of the genomic RNA of kennedya yellow mosaic tymovirus-Jervis Bay isolate: relationships with potex- and carlaviruses // J. Gen. Virol.- 1990.- 71, N 4.- P. 925-931. 6. Dolja V. V., Koonin E. V. Phylogeny of capsid proteins of small icosahedral RNA plant viruses // J. Gen. Virol.- 1991.- 72, N 7.- P. 1481–1486. 7. Esteban D.J., Syed A., Upton C. Organizing and Updating Whole Genome BLAST Searches with ReHAB // Methods Mol. Biol.- 2007.- 395.- P. 187-194. 8. Hataya T., Uchino K., Arimoto R., Suda N., Sano T., Shikata E., Uyeda I. Molecular characterization of Hop latent virus and phylogenetic relationships among viruses closely related to carlaviruses // Arch.Virol.- 2000.- 145, N 12.- P. 2503-2524. 9. Koenig R., Barends S., Gultyaev A.P., Lesemann D.E., Vetten H.J., Loss S., Pleij C.W. Nemesia ring necrosis virus: a new tymovirus with a genomic RNA having a histidylatable tobamovirus-like 3' end // J. Gen. Virol.- 2005.- 86, N 6.- P. 1827-1833. 10. Lee D., Choi J.H., Dalkilic M.M., Kim S. COMPAM :visualization of combining pairwise alignments for multiple genomes // Bioinformatics.- 2006.- 22, N 2.- P. 242-244. 11. Miziara M.N., Riggs P.K., Amaral M.E. Comparative analysis of noncoding sequences of orthologous bovine and human gene pairs // Genet. Mol. Res.- 2004.- 3, N 4.- P. 465-473. 12. Nix D.A., Eisen M.B. GATA: a graphic alignment tool for comparative sequence analysis // BMC Bioinformatics.- 2005.- 6.- P.9. 13. 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 // J. Gen. Virol.- 2007.- 88, N 1.- P. 298-315. 14. Shen R. and Miller W.A. The 3' Untranslated Region of Tobacco Necrosis Virus RNA Contains a Barley Yellow Dwarf Virus-Like Cap-Independent Translation Element // J. Virol.- 2004.- 78, N 9.- P. 4655-4664. 15. Smith G.R., Borg Z., Lockhart B.E., Braithwaite K.S., Gibbs M.J. Sugarcane yellow leaf virus: a novel member of the Luteoviridae that probably arose by inter-species recombination // J. Gen. Virol.- 2000.- 81, N 7.- P.1865-1869. 16. Tamura K., Dudley J., Nei M., Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0 // Mol. Biol. Evol.- 2007.- 24, N 8.- P. 1596-1599. 17. Uchiyama I., Higuchi T., Kobayashi I. CGAT: a comparative genome analysis tool for visualizing alignments in the analysis of complex evolutionary changes between closely related genomes // BMC Bioinformatics.- 2006.- 7.- P.472. 18. Zhou H., Wang H., Huang L.F., Naylor M., Clifford P. Heterogeneity in codon usages of sobemovirus genes // Arch.Virol.- 2005.- 150, N 8.- P. 1591-1605.

Надійшла до редколегії: 23.01.08

UDC 578.85/.86

J. Staniulis[1], Dr. habil., M. Žižytė[1], PhD student, T. Norkus[1], bacaalaureate, L. Yusko[2], PhD student, H. Snihur[2], PhD, I. Budzanivska[2], PhD

INCIDENCE OF PLUM POX VIRUS IN LITHUANIA AND UKRAINE

[1] Institute of Botany, Plant Virus Laboratory, Vilnius, Lithuania;[2] Taras Shevchenko` Kyiv National University, Ukraine.

According to the internationally agreed joint project "Investigation of distribution of plum pox virus in Lithuania and Ukraine and development of biotechnological methods of obtaining of virus-free planting material" 95 samples of plum, apricot and peach from several locations of Ukraine were collected in June 2007 and checked for the presence of PPV by DAS-ELISA. Presence of PPV was detected in all explored regions. Four isolates PPV from plum of Kyivsky and Transkarpathian regions were transferred to herbaceous plants and the presence of the virus in these plants was confirmed by DAS-ELISA and IEM investigation. The presence of PPV in samples from Lithuania was confirmed in 2 cases from 20. PPV positive samples were collected from Jonava and Sakiai regions.

Introduction. Stone fruit trees are natural hosts of several widespread viruses. Thirty virus diseases of stone and pome fruits have been described in former Yugoslavia [Šutič et al., 1999]. Plum pox virus or 'Sharka' disease agent *Plum pox potyvirus* (PPV) originally described in Bulgaria on *Prunus domestica* gradually has spread throughout European borders [Šutič et al., 1999]. PPV infects virtually all cultivated fruit tree species of the genus *Prunus*. The disease also affects some wild *Prunus* species, especially blackthorn (*Prunus spinosa*) which has been a natural source of infection in many countries. PPV is now reported in most European countries, in parts of Asia and northern Africa, and in South America. Plum