ISSN 0026-8933, Molecular Biology, 2008, Vol. 42, No. 4, pp. 543–553. © Pleiades Publishing, Inc., 2008.
Original Russian Text © I.V. Babkin, T.S. Nepomnyashchikh, R.A. Maksyutov, V.V. Gutorov, I.N. Babkina, S.N. Shchelkunov, 2008, published in Molekulyarnaya Biologiya, 2008, Vol. 42, No. 4, pp. 612–624.

GENOMICS. TRANSCRIPTOMICS. PROTEOMICS

UDC 575.86:575.852'113

Comparative Analysis of Variable Regions in the Variola Virus Genome

I. V. Babkin, T. S. Nepomnyashchikh, R. A. Maksyutov, V. V. Gutorov, I. N. Babkina, and S. N. Shchelkunov

VECTOR State Research Center of Virology and Biotechnology, Kol'tsovo, Novosibirsk region, 630559 Russia; e-mail: babkin@vector.nsc.ru

Received January 23, 2008
Accepted for publication February 22, 2008

Abstract—Two segments of the variable terminal regions of the variola virus (VARV) genome were sequenced in 22 strains from the Russian collection, including about 13.5 kb of the left segment and about 10.5 kb of the right segment. The total length of the sequences was over 540 kb. Phylogenetic analysis of the new and published data determined the relationships among 70 VARV strains, the character of their clustering, and the intergroup and intragroup variation of the strain clusters. Loci with the highest polymorphism rate were identified and proved to map to noncoding regions or to regions of damaged open reading frames, characteristic of the ancestral virus. These loci offer attractive possibilities for developing a strategy of VARV strain genotyping. Recombination analysis by different methods did not detect, except for a single case, significant recombination events in the VARV strains examined.

DOI: 10.1134/S0026893308040092

Key words: variola virus, sequencing, phylogenetic analysis, genetic divergence, recombination

INTRODUCTION

Viruses of the genus *Orthopoxvirus* have a linear double-stranded DNA genome with covalently linked ends and a molecular mass varying from 120 MDa for the variola virus (VARV) to 145 MDa for the cowpox virus (CPXV) [1]. The VARV genome is the shortest among all of the orthopoxvirus genomes studied. Presumably, CPXV originated from an ancestor virus as a result of genomic deletions and mutations, which affected the terminal variable regions, leaving the 102-kb central part of the genome virtually unchanged [2, 3]. It is known that the case fatality levels of smallpox epidemics varied depending on the time and geographic region [3–5]. Therefore, it seems relevant to study the VARV genome variation among different virus isolates.

We have earlier performed a genome-wide RFLP analysis of the VARV strains from the Russian collection and found that, although the VARV genome is, in general, highly conserved, its left and right terminal regions still contain relatively variable sequences [6, 7]. Based on these data, we sequenced regions 11546–24790 (the left segment) and 149427–160265 (the right segment; genome positions as in VARV strain IND67) in 22 VARV strains from the Russian collection. In addition, the data analysis involved the sequences recently published for the same segments of 47 VARV strains from the US collection [5] and strain IND67, whose genome has been sequenced earlier [8, 9].

EXPERIMENTAL

DNA samples were obtained from VARV strains of the Russian collection (Vector) (table). Experiments with noninfectious DNA were conducted in biosafety level 2 labs. The study was approved by the WHO as a part of the international program for VARV research.

Oligonucleotide primers were synthesized using an automated ABI-394 synthesizer (Applied Biosystems, United States) in the Institute of Chemical Biology and Fundamental Medicine (Novosibirsk). Two primer pairs were designed to obtain two overlapping DNA fragments that together covered the left segment of the VARV genome, and 67 oligonucleotides were designed to determine the sequence of this segment. For the right genome segment, other two primer pairs served to obtain two overlapping amplicons, and 68 primers were designed for their sequencing.

PCR was performed using the hot start technique with an XL PCR kit (Applied Biosystems) in a GeneAmp PCR System 9700 (Applied Biosystems) under the following conditions: initial denaturation, 94°C, 1 min; 15 cycles (93°C, 15 s; 58°C, 15 s; 68°C, 15 min); 15 cycles (93°C, 15 s; 68°C, 15 min plus additional

VARV strains whose sequences were used in this work

No.	Name	Geographic region of strain isolation	Country of isolation	Year of isolation	GenBank number of the left genome segment	GenBank number of the right genome segment
1.	Mary	Africa	Tanzania	1962	EF611200	EF611212
2.	Ngami	″	″	1962	EF611202	EF611217
3.	13/62	″	″	1962	EF611189	EF611221
4.	Helder	″	″	1962	EF611194	EF611226
5.	Congo 9	″	Congo	1970	EF611192	EF611224
6.	Congo 2	″	″	1970	EF611193	EF611225
7.	Rw 18	″	Rwanda	1970	EF611203	EF611218
8.	Brazil 131	America	Brazil	unknown	EF611191	EF611223
9.	Aziz	Asia	Pakistan	1970	EF611190	EF611222
10.	Khateen	″	″	1970	EF611198	EF611230
11.	Wsim Ahmed	″	″	1970	EF611204	EF611219
12.	6-58	″	″	1958	EF611188	EF611220
13.	Kuw 5	″	Kuwait	1967	EF611205	EF611209
14.	Ind 3a	″	India	1967	EF611195	EF611227
15.	Ind 4a	″	″	1967	EF611196	EF611228
16.	India 71	″	″	1975	EF611197	EF611229
17.	M A 60*	Asia/Europe	Russia	1960	EF611199	EF611210
18.	M Abr 60*	″	″	1960	EF611207	EF611211
19.	M Bl 60*	″	″	1960	EF611208	EF611213
20.	M N 60*	″	″	1960	EF611206	EF611215
21.	M Gavr 60*	″	″	1960	EF611187	EF611214
22.	M Sur 60*	″	″	1960	EF611201	EF611216
23.	IND67	Asia	India	1967	NC_001611	NC_001611
24.	BEN68	Africa	Benin	1968	DQ441416	DQ441416
25.	BOT72	″	Botswana	1972	DQ441417	DQ441417
26.	BOT73	″	″	1973	DQ441418	DQ441418
27.	CNG70 46	″	Congo	1970	DQ437583	DQ437583
28.	CNG70 227	″	″	1970	DQ441423	DQ441423
29.	ETH72 16	″	Ethiopia	1972	DQ441424	DQ441424
30.	ETH72 17	″	″	1972	DQ441425	DQ441425
31.	GUI69	″	Guinea	1969	DQ441426	DQ441426
32.	NIG69	″	Niger	1969	DQ441434	DQ441434
33.	SAF65 102	″	South Africa	1965	DQ441435	DQ441435
34.	SAF65 103	″	″	1965	DQ441436	DQ441436
35.	SLN68	″	Sierra-Leone	1969	DQ441437	DQ441437
36.	SOM77ALI	″	Somali	1977	DQ437590	DQ437590
37.	SOM77 1252	″	″	1977	DQ441438	DQ441438
38.	SOM77 1605	″	″	1977	DQ441439	DQ441439
39.	SUD47JUB	″	Sudan	1947	DQ441440	DQ441440
40.	SUD47RUM	″	″	1947	DQ441441	DQ441441
41.	TAN65	″	Tanzania	1965	DQ441443	DQ441443
42.	AFG70	Asia	Afghanistan	1970	DQ437580	DQ437580
43.	BSH74NUR	″	Bangladesh	1974	DQ441420	DQ441420
44.	BSH74SHZ	″	″	1974	DQ441421	DQ441421
45.	BSH74SOL	″	″	1974	DQ441422	DQ441422

Table. (Contd.)

No.	Name	Geographic region of strain isolation	Country of isolation	Year of isolation	GenBank number of the left genome segment	GenBank number of the right genome segment
46.	BSH75BANU	″	″	1975	DQ437581	DQ437581
47.	BSH75	″	″	1975	L22579	L22579
48.	CHN48	″	China	1948	DQ437582	DQ437582
49.	IND53MAD	″	India	1953	DQ441427	DQ441427
50.	IND53NDEL	″	″	1953	DQ441428	DQ441428
51.	IND64VEL4	″	″	1964	DQ437585	DQ437585
52.	IND64VEL5	″	″	1964	DQ437586	DQ437586
53.	IRN72	″	Iran	1972	DQ437587	DQ437587
54.	JAP46YAM	″	Japan	1946	DQ441429	DQ441429
55.	JAP51HRPR	″	″	1951	DQ441430	DQ441430
56.	JAP51STWL	″	″	1951	DQ441431	DQ441431
57.	KOR47	″	Korea	1947	DQ441432	DQ441432
58.	KUW67	″	Kuwait	1967	DQ441433	DQ441433
59.	NEP73	″	Nepal	1973	DQ437588	DQ437588
60.	SUM70 222	″	Sumatra	1970	DQ437591	DQ437591
61.	SUM70 228	″	″	1970	DQ441442	DQ441442
62.	SYR72	″	Syria	1972	DQ437592	DQ437592
63.	GER58	Europe	Germany	1958	DQ437584	DQ437584
64.	UNK44HARV	″	United Kingdom	1944	DQ441444	DQ441444
65.	UNK46HIND	″	″	1946	DQ441445	DQ441445
66.	UNK47HIG	″	″	1947	DQ441446	DQ441446
67.	UNK52BUT	″	″	1952	DQ441447	DQ441447
68.	YUG72	″	Yugoslavia	1972	DQ441448	DQ441448
69.	BRZ66	America	Brazil	1966	DQ441419	DQ441419
70.	GAR66	″	″	1966	Y16780	Y16780

Note: Strains 1–23 are from the Russian collection; strains 24–70 are from the collection of the Center for Disease Control and Prevention (Atlanta, United States).

* VARV strains of the Moscow outburst [6].

15 s in each subsequent cycle); final elongation, 72°C, 20 min.

Sequencing was performed with a BigDye Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems); the reaction products were purified with a DyeEx Spin kit (Qiagen, United States) and separated using an automated ABI 310 genetic analyzer (Applied Biosystems). The nucleotide sequences were analyzed using the Sequencher v.4.0.5 software (Gene Codes, United States).

Sequences of VARV strains from the collection of the Center for Disease Control and Prevention (Atlanta, United States) were obtained from GenBank (http://www.ncbi.nih.gov) (table). The sequences of other orthopoxviruses were obtained from various databases. The DNA sequences of camelpox virus strains CMS (AY009089) and M-96 (AF438165), CPXV strain GER91 (DQ437593), monkeypox virus strain SL (AY741551), horsepox virus strain HSPV (DQ792504), taterapox virus strain TATV (NC_008291), and ectromelia virus strain MOS (AF012825) were obtained from GenBank. The sequence of ectromelia strain NAV is available at www.sanger.ac.uk/Projects/Ectromelia_virus. The genomes of CPXV strain GRI90 (X94355) and monkeypox virus strain Z96 (AF380138) have been sequenced earlier in our center.

Phylogenetic analysis. Nucleotide sequences were aligned using the BioEdit v. 7.0 [10] and ClustalX v.1.8 software [11]. Phylogenetic analysis of nucleotide sequences of 70 VARV strains was done using the maximum likelihood method with the Paup4.0b10

software (Sinauer Associates, United States). The models of nucleotide substitutions, the substitution matrix parameters, the nucleotide frequencies, the proportion of invariant sites, and the parameters of the γ-distribution for Paup4.0b10 were initially determined by the likelihood ratio test [12], using the Modeltest v. 3.7 software [13]. An heuristic search was performed by dividing and reassembling the tree and by branch exchange with addition of random sequences (100 replicates of additional sequences); the analysis was based on the initial tree obtained by neighbor-joining. The significance of the tree was tested by permutation analysis with 100 replicates, using the same strategy and parameters. The trees were constructed using the TreeView [14] and Mega v.3.1 software packages [15].

Phylogenetic analysis of 80 different orthopoxvirus sequences, including those of VARV strains, was performed by neighbor-joining using the Mega v.3.1 software. The significance of the tree was tested by permutation analysis with 1000 replicates.

Genetic distances between groups of VARV strains were calculated with the Mega v.3.1 software, using Kimura's biparametric model of nucleotide substitutions [16].

Genetic divergence π and the number of polymorphic sites in the left and right genome segments of the 70 VARV strains were calculated using the DNAsp v.4.10.9 software [17] according to Nei and Miller [18] with a window size of 100 nt and a step of 25 nt.

Recombination analysis was performed according to Hudson and Kaplan [19] with the DNAsp v.4.10.9 software [17]. For the recombination Phi-test [20], the SplitsTrees4 v.4.8 software was used [21].

RESULTS

Smallpox eradication, declared by the World Health Organization in 1980, was the first impressive human victory over a dangerous viral pathogen [22]. However, the molecular biological methods of that time were inadequate for a comprehensive comparative investigation of the VARV genome, open reading frame (ORF) structure, the functions of viral proteins, and the relationship between the genetic structure and the biological properties of the virus. Research in VARV structure and function should provide insights both into the functioning of this dangerous virus and into the possible medical use of the properties of VARV proteins. Moreover, although VARV has been completely eradicated, there still exists a range of other widespread orthopoxviruses, including CPXV, whose genomes code for a complete set of orthologs of VARV proteins [3, 23]. Therefore, there remains a potential danger of a VARV-like virus arising as a result of mutations in some existing *Orthopoxvirus* species.

It is known that VARV isolates from different geographic locations differ biologically and epidemiologically. American strains with a low case fatality rate were classified as *variola minor alastrim*, while *variola major* strains have a high case fatality [4, 24, 25]. We have earlier characterized the genome variation in 21 VARV strains from the Russian collection [6], which allowed us to choose the genome segments with the highest heterogeneity for further investigation.

PCR primers were based on the published genomic sequences of VARV strains IND67 (GenBank X69198), GAR66 (GenBank Y16780), and BSH75 (GenBank L22579) and were designed with Oligo 3.3 software. PCR was performed with r*Tth* DNA Polymerase XL (Applied Biosystems), which has proofreading activity and ensures DNA synthesis with a mutation frequency of about 1×10^{-6} mistakes/nucleotide, about two orders of magnitude lower than that of *Taq* DNA polymerase [26].

The nucleotide sequences of 22 VARV strains were determined with 135 primers. The sequencing profiles were visually verified twice with the Sequencher v.4.0.5 software package; each position was read in both strands. Altogether, over 540 kb were analyzed. For each nucleotide sequence, we determined ORFs in both strands; the data were deposited in GenBank (table).

Phylogenetic Analysis of VARV Genomic Sequences

For 47 VARV strains from the American collection and strain IND67, the sequences homologous to the left and right genome segments sequenced in VARV strains from the Russian collection were obtained from GenBank (table). In addition, we used the sequences of ten other orthopoxvirus strains.

Multiple sequence alignment of the left and right genome segments of the 70 VARV strains was performed with the ClustalX v. 1.8 software, and the results were verified manually with the BioEdit v. 7.0 software. The alignment length was 13 591 bp for the left genome segment and 10 979 bp for the right genome segment. The alignments were further concatenated to form a single 24 570-bp alignment. In a similar manner, we constructed nucleotide sequence alignments that included, in addition to the 70 VARV strains, ten orthopoxviruses of six species.

At the first step of phylogenetic analysis, we examined 80 aligned orthopoxvirus sequences of the combined left and right genome segments by the neighbor-joining method, using the Mega v. 3.1 software (Fig. 1). The phylogenetic tree demonstrates the peculiarities of the orthopoxvirus clade organization. For

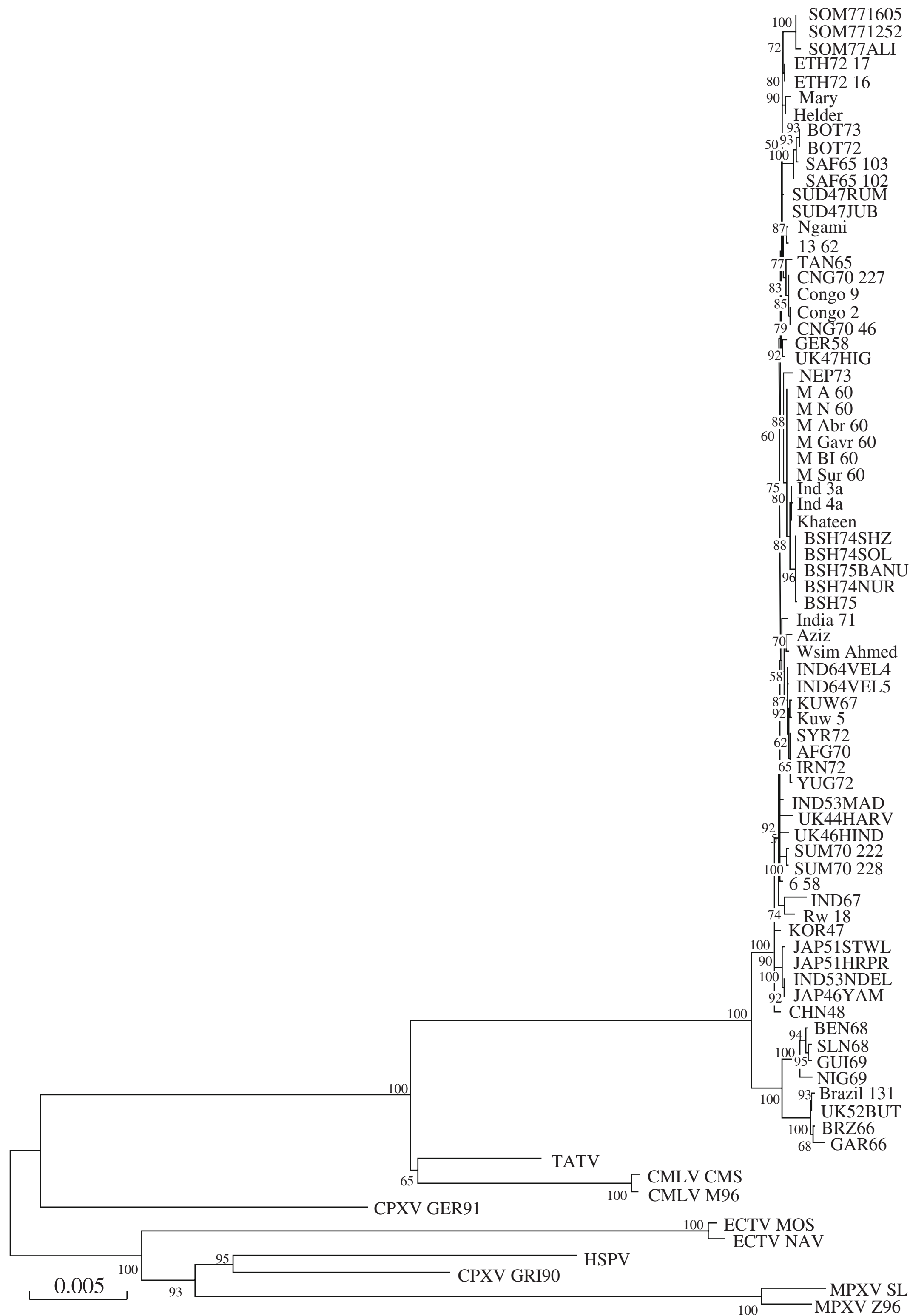

Fig. 1. Rootless tree constructed by neighbor-joining with the sequences of the left and right segments of orthopoxvirus genomes. The node confidence is shown when exceeding 50%. Here and in Figs. 2 and 3: The divergence scale (substitutions/site) is given at the bottom. Here and in Figs. 2–4, the strains are designated as in the table.

example, the camelpox virus and tateropoxvirus are in one group with VARV; CPXV strain GRI90 is combined with the horsepox virus, and the group of monkeypox strains is adjacent to them. A tree with a similar topology has been obtained by analyzing the central conserved region of orthopoxvirus genomes [27]. However, in our phylogram (Fig. 1), in contrast to the tree constructed for the central conserved region of orthopoxvirus genomes, CPXV strain GER91 was in the same clade as VARV, camelpox virus, and taterapox virus. The tree constructed using the maximum likelihood method with the Paup4.0b10 software confirmed this result. Using other orthopoxviruses as an outgroup for VARV strains, we determined the topology of the root of the VARV strain tree (Fig. 1).

At the next step, the phylogenetic analysis of the left and right genome segments of the 70 VARV strains was performed with the Paup4.0b10 software, using a more powerful and precise maximum likelihood method. The test parameters were initially chosen with the Modeltest v. 3.7 software package as described above. The resulting trees were rooted (Figs. 2, 3). The root position was determined from the previous phylogram, which included other orthopoxviruses in addition to VARV (Fig. 1). The root topology agreed with the results obtained in earlier phylogenetic studies with the central conserved region of orthopoxvirus genomes and the sequences of poxvirus DNA polymerase genes [27].

As it follows from our trees (Figs. 2, 3), all VARV strains are grouped according to their geographic origin. Noteworthy, South American and West African strains come into a common clade and differ to the greatest extent from other VARV strains. The strains closest to this group are from the Far East: Japan, China, and Korea, and strain IND53NDEL. The group of other African strains includes subgroups of strains coming from East, Central, and South Africa. Congolese strains form a common cluster with Tanzanian isolates TAN65, 13/62, and Ngami at a high confidence level. Two other Tanzanian strains, Helder and Mary, group with a strain from Rwanda, Rw 18.

Among Asian strains, separate clusters are formed by isolates from Bangladesh, Central Asia and the Middle East (Afghanistan, Iran, Iraq, Syria, and Kuwait), Sumatra, and Moscow (table). The closest relatives of the six strains from the Moscow outburst are isolates from India (Ind 3a and Ind 4a) and Pakistan (Khateen). Strain 6-58 (Pakistan) is, with a high significance, in the same group as IND67 [9]. Strains from India and Pakistan are present in different clusters of the Asian group, apparently, because this region used to be the main source of VARV biodiversity in the past century.

Although the positions of VARV strains Aziz and Rw 18 are somewhat different, the topology of the tree nodes with a more than 70% confidence is identical on the phylograms for the left and right segments of the VARV genome (Figs. 2, 3). The only exception is strain NEP73, whose unique sequence groups with a high confidence with the strains from India, Pakistan, and Bangladesh on the tree of the right segment and with the strains from Central Asia and the Middle East on the tree of the left segment. Apparently, this strain originated via molecular recombination of different VARV variants.

The combined nucleotide sequences of the left and right genome segments of the 70 VARV strains were divided into four major geographic groups: South American, West African, African (excluding West African strains), and Asian (Figs. 2, 3). Using the Mega v.3.1 software to determine the average inter- and intragroup genetic distances, we found that the evolutionary distance between the strains from South America and West Africa and the other VARV strains studied was the highest. The genetic distances between these strain groups and other VARV isolates were 0.0044 to 0.0048 nucleotide substitutions per site. The genetic distance between the isolates from South America and West Africa was also significant, 0.0026 substitutions per site. Asian and African (West African excluded) strains are the genetically closest groups (0.0010 substitutions per site). By intragroup divergence analysis, the highest variation was observed among VARV strains from Asia (0.00088 substitutions per site), while isolates from South America, whose evolutionary history is the shortest [27], had the lowest variation (0.00037 substitutions per site).

Intergenomic Differences among VARV Strains

Based on the model of VARV strain clustering (Figs. 2, 3), multiple sequence alignments of the left and right genome segments of 70 VARV isolates were modified by ordering the strains according to their grouping in the phylogram. Next, to study the interstrain heterogeneity, all positions containing conserved nucleotides or deletions were removed from both sequence alignments to obtain condensed alignments containing only variable positions. An example of such an alignment for the left genome segment is given in Fig. 4. Analyzing the condensed sequence alignments of both genome segments, we found that VARV strains from South America and West Africa had the highest number of substitutions. Furthermore, these two groups have the same nucleotides in a number of positions as in the subgroup containing VARV strains from China, Japan, and Korea and Indian strain IND53NDEL. It is noteworthy that all nucleotide substitutions characteristic of the strains of the Far East subgroup were also present in the strains from South America and West Africa, suggesting a monophyletic origin of these viruses.

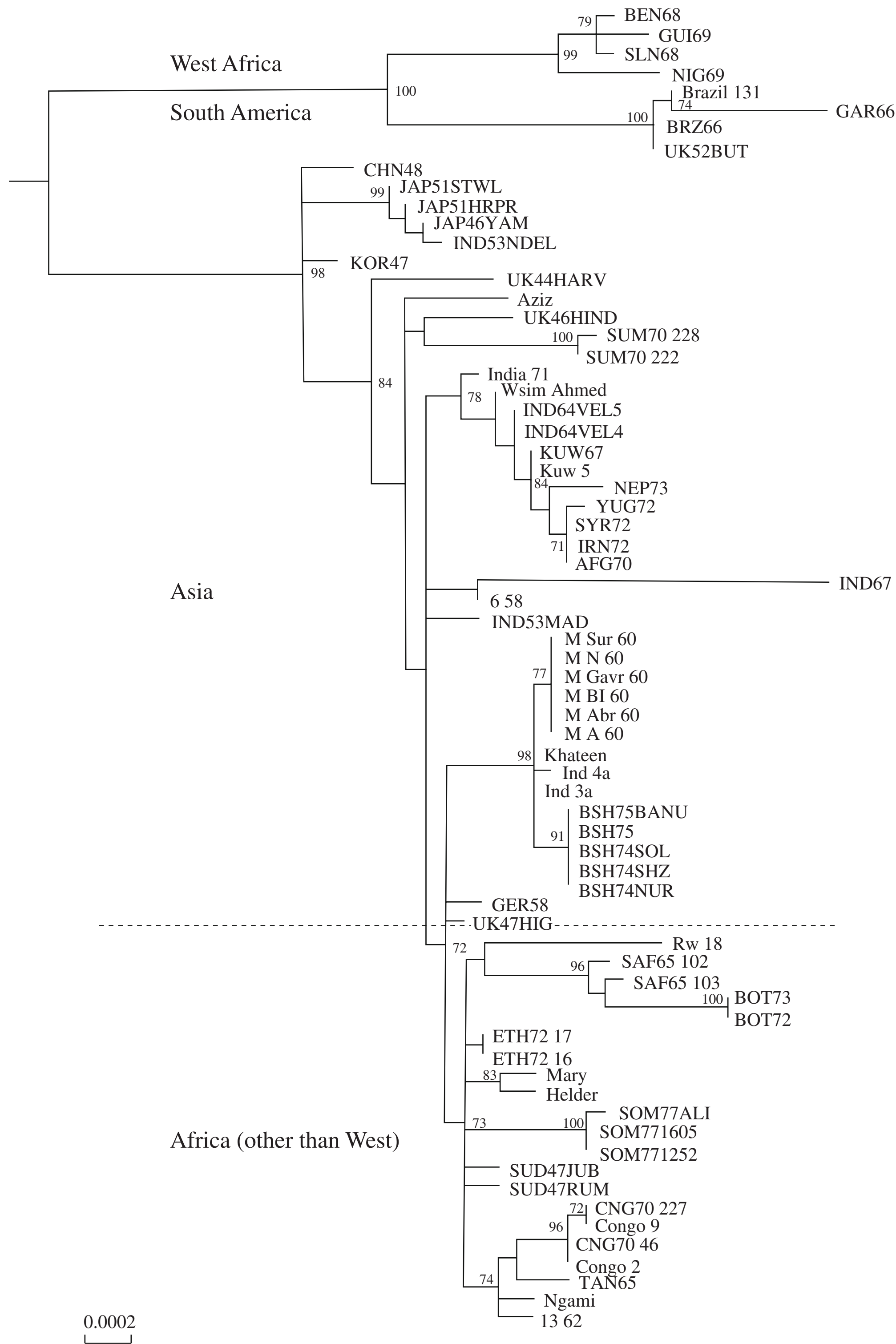

Fig. 2. Rooted tree constructed by the maximum likelihood method with the left segment of the VARV genome. Here and in Fig. 3, the node confidence is shown when exceeding 70%.

Regions with deletions, which are frequent in orthopoxvirus genomes [28], were excluded from the analysis, because their genomic location is not associated with the geographic clustering of VARV strains. As earlier analysis of sequence alignments has shown, genetically close isolates may contain deletions or insertions of different lengths.

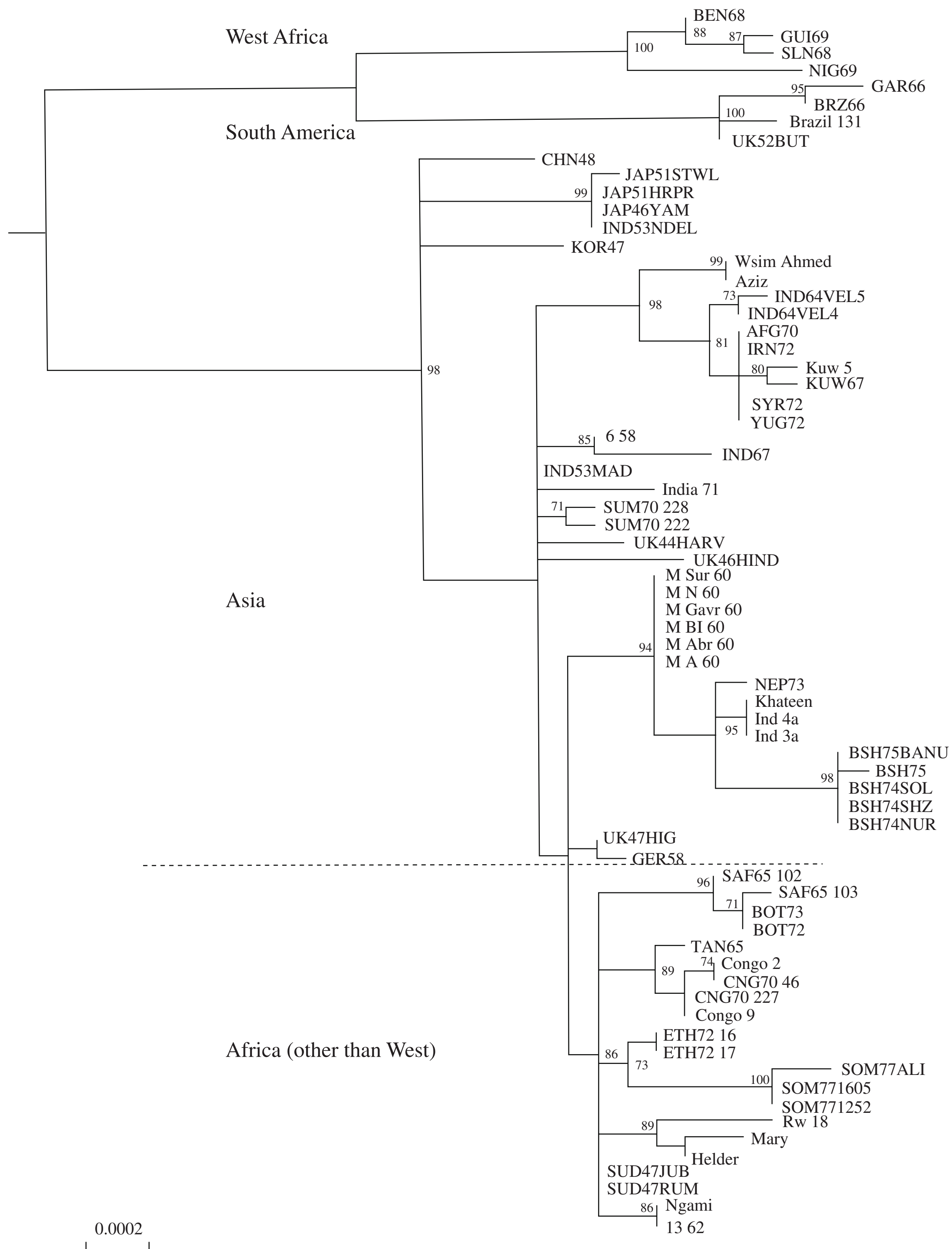

Fig. 3. Rooted tree constructed by the maximum likelihood method with the right segment of the VARV genome.

With the above method of condensed alignments, strain grouping according to the topology preliminarily determined by phylogenetic analysis enables intergroup genotyping of VARV strains.

Condensed sequence alignments do not reflect the actual genomic distribution of nucleotide substitutions. To detect the most variable genomic regions, interstrain differences were studied with the DNAsp

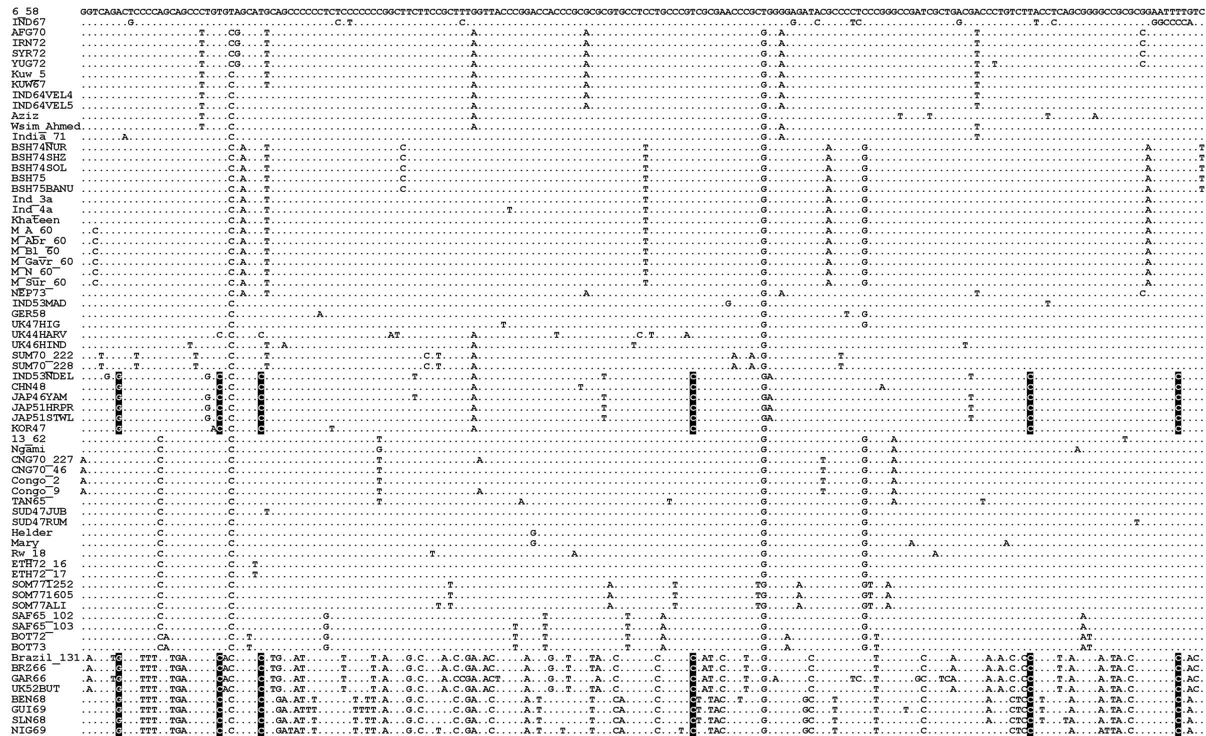

Fig. 4. Condensed sequence alignment of the left genomic segment (13 591 bp) of 70 VARV strains. Only positions with single nucleotide interstrain variations are considered. Positions characteristic of only VARV strains from the Far East and West Africa/South America are shaded.

software by computing the divergence π, which is the average number of nucleotide substitutions per site, and the number of polymorphic sites. Deletion-containing positions were excluded from the analysis (Fig. 5). Then, the computations were repeated with the deletion-containing positions (data not shown). The divergence graphs obtained with or without deletions were similar, indicating that the regions with the highest densities of deletions and substitutions are the same. Thus, we detected the regions with the highest polymorphism rates. In the VARV genome segments under study, the highest heterogeneity was observed in the left segment regions located at the beginning and at the end of ORF D13L (IND67 nomenclature), which is a fragment of a damaged ORF encoding a kelch-like protein in CPXV [1, 3, 29] (Fig. 5). Damage to the viral gene resulted in a loss of function and, as a consequence, to an accumulation of mutations. These regions seem promising for VARV strain genotyping. In the left segment, a divergence of 0.008 was observed for the noncoding region between ORFs C3L and C4R. In CPXV strain GRI90, this intergenic region corresponds to a fragment of ORF M5L, most of which is deleted in VARV [3]. In the right segment, the only region with the same divergence (0.008) is between ORFs B3L and B4L (IND67 nomenclature) and is a damaged analog of ORF B2R of CPXV strain GRI90 [3]. Thus, mutational hotspots in the VARV genome are in functionally inactive DNA regions.

Recombination between VARV Strains

To detect potential recombination events, we studied the phylograms constructed for each putative ORF of the left and right genome segments (data not shown). There are 17 ORFs in the left genome segment and 12 to 15 predicted ORFs in the right segment (depending on the strain). For all ORFs studied, all VARV strains, except for NEP73, tend to group according to their geographic origin. Significant recombination events were not detected.

At the next step, using the method by Hudson and Kaplan [19] with the DNAsp v.4.10.9 software, we studied the aligned sequences of the left and right segments of 69 VARV strains, excluding NEP73 as a supposed recombinant. We detected 13 potential recombination sites in the left segment and 5 sites in the right one. Then, potential recombination was assessed with the combined set of aligned sequences by using the Phi-test with the SplitsTrees4 software. This test did not detect significant recombination events between VARV strains. When orthopoxviruses were considered, the phylograms built for each ORF of the left and right genome segments revealed a complex picture of recombination events (data not shown).

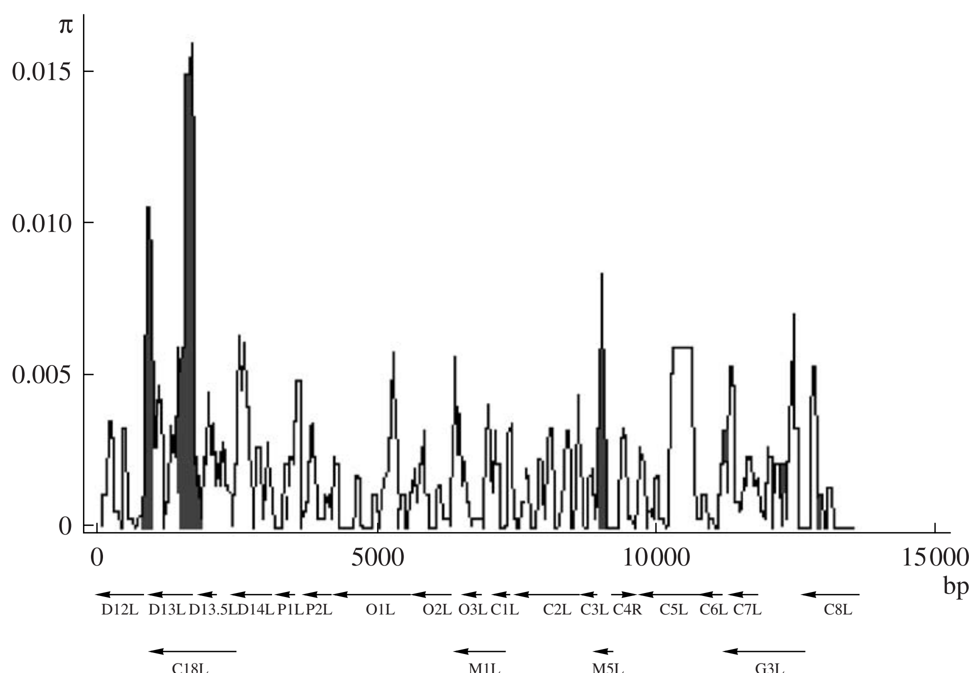

Fig. 5. Genetic divergence π of the left genomic segment of 70 VARV strains. X axis, nucleotide positions in the segment. Positions containing deletions were excluded from the analysis. In the region of deletions, the graph is parallel to the X axis. Graph peaks with $\pi > 0.008$ are marked black. VARV and differing CPXV ORFs are shown below. ORF boundaries and directions are indicated with arrows.

Apparently, the evolution of different orthopoxvirus species involved multiple recombination events in both genome segments studied.

DISCUSSION

This work presents a detailed analysis of the relationships among the VARV strains. Considering the inter- and intragroup heterogeneity of the VARV strain clusters established, we mapped the regions with the highest polymorphism in the two segments of the VARV genome and showed that they are located in noncoding regions and in the regions of damaged ORFs of the ancestral virus. These regions can used to develop a strategy for VARV strain genotyping. In spite of the potential possibility of recombination, different methods of recombination analysis did not detect, except for a single case, significant recombination events in the VARV genomes studied.

In general, it should be noted that the genome of our VARV strains, isolated from 1944 to 1977 in different geographic locations, is highly conserved. The most pronounced differences were found between West African strains and their South American descendants [27] on the one side and the rest of the strains on the other (Figs. 2, 3); this confirms our earlier results suggesting that the former isolates belong to an individual genetic VARV subtype [7].

Unexpectedly, the genome regions of VARV isolates from the Far East and West Africa/South America have common differences from the other VARV strains (Fig. 4), suggesting an evolutionary relationship between these two geographically distant VARV variants. In addition, almost every Asian subgroup of VARV isolates contains strains of the Indian origin (Figs. 2, 3).

Phylogenetic analysis (Fig. 1) suggests that VARV originated from an ancestor orthopoxvirus with a wide host range (for example, a rodent virus, termed CPXV). Geographically, VARV must have appeared in the Middle East [27] and further spread eastwards to Japan and westwards to the West African coast [24]. A distinct VARV subtype developed in West Africa owing to its relative geographic isolation.

The largest smallpox epidemics of the past century occurred in India [29]. Apparently, this region at the time was an evolutionary hot spot for VARV, where new genetic variants of VARV arose to spread to other geographic zones of Asia; this is why different Asian phylogenetic groups of VARV include strains that were isolated in India at different times.

Interstrain differences are primarily concentrated in noncoding regions of the two VARV genome segments, that is, in intergenic regions or in damaged nonfunctional genes, whose functional counterparts are present in the CPXV genome (Fig. 5).

By detailed sequence analysis of the two variable genome segments of 70 VARV strains, we characterized their variation and detected the regions useful for molecular biologic typing of VARV isolates.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project no. 08-04-00443).

REFERENCES

1. Marennikova S.S., Shchelkunov S.N. 1998. *Patogennye dlya cheloveka ortopoksvirusy* (Orthopoxviruses Pathogenic for Humans). Moscow: KMK.
2. Lefkowitz E.J., Wang C., Upton C. 2006. Poxviruses: Past, present and future. *Virus Res*. **117**, 105–118.
3. Shchelkunov S.N., Marennikova S.S., Moyer R.W. 2005. *Orthopoxviruses Pathogenic for Humans*. Berlin: Springer.
4. Fenner F., Wittek R., Dumbell K.R. 1989. *Orhopoxviruses*. San Diego: Academic Press.
5. Esposito J.J., Sammons S.A., Frace A.M., Osborne J.D., Gisen-Rasmussen M., Zhang M., Govil D., Damon I.K., Kline R., Laker M., Li Y., Smith G.L., Meyer H., LeDuc J.W., Wohlhueter R.M. 2006. Genome sequence diversity and clues to the evolution of variola (smallpox) virus. *Science*. **313**, 807–812.
6. Babkina I.N., Babkin I.V., Marennikova S.S., Sandakhchiev L.S., Shchelkunov S.N. 2004. Comparative restriction enzyme analysis of the genome in variola virus strains from the Russian collection. *Mol. Biol*. **38**, 429–436.
7. Babkina I.N., Babkin I.V., Li Yu., Ropp S., Cline R., Damon I., Esposito G., Sandakhchiev L.S., Shchelkunov S.N. 2004. Phylogenetic comparison of the genomes of different variola virus strains. *Dokl. Akad. Nauk*. **398**, 316–319.
8. Shchelkunov S.N., Blinov V.M., Sandakhchiev L.S. 1993. Genes of variola and vaccinia viruses needed for overcoming of the host protective mechanisms. *FEBS Lett*. **319**, 80–83.
9. Shchelkunov S.N., Massung R.F., Esposito J.J. 1995. Comparison of the genome DNA sequences of Bangladesh-1975 and India-1967 variola viruses. *Virus Res*. **36**, 107–118.
10. Hall T.A. 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids. Symp. Se*r. **41**, 95–98.
11. Thompson J.D., Gibson T.J., Plewniak F., Jeanmougin F., Higgins D.G. 1997. The ClustalX windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res*. **24**, 4876–4882.
12. Huelsenbeck J.P., Rannala B. 1997. Maximum likelihood estimation of phylogeny using stratigraphic data. *Paleobiology*. **23**, 174–180.
13. Posada D., Crandall K.A. 1998. MODELTEST: Testing the model of DNA substitution. *Bioinformatics*. **14**, 817–818.
14. Page R.D. 1996. TreeView: An application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci*. **12**, 357–358.
15. Kumar S., Tamura K., Nei M. 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief. Bioinform*. **5**, 150–163.
16. Kimura M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol*. **16**, 111–120.
17. Rozas J., Sanchez-DelBarrio J. C., Messeguer X., Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*. **19**, 2496–2497.
18. Nei M., Miller J.C. 1990. A simple method for estimating average number of nucleotide substitutions within and between populations from restriction data. *Genetics*. **125**, 873–879.
19. Hudson R.R., Kaplan N.L. 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics*. **111**, 147–164.
20. Bruen T., Philippe H., Bryant D. 2005. A simple and robust statistical test for detecting the presence of recombination. *Genetics*. **172**, 2665–2681.
21. Huson D.H., Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol*. **23**, 254–267.
22. WHO. 1980. *The global eradication of smallpox. Final report of the Global Commission for the certification of smallpox eradication*. Geneva: World Health Organization.
23. Shchelkunov S.N., Safronov P.F., Totmenin A.V., Petrov N.A., Ryazankina O.I., Gutorov V.V., Kotwal G.J. 1998. The genomic sequence analysis of the left and right species-specific terminal regions of a cowpox virus strain reveals unique sequences and a cluster of intact ORFs for immunomodulatory and host range proteins. *Virology*. **243**, 432–460.
24. Fenner F., Henderson D.A., Arita I., Jezek Z., Ladnyi I.D. 1988. *Smallpox and its Eradication*. Geneva: World Health Organization.
25. Shchelkunov S.N., Totmenin A.V., Loparev V.N., Safronov P.F., Gutorov V.V., Chizhikov V.E., Knight J.C., Parsons J.M., Massung R.F., Esposito J.J. 2000. Alastrim smallpox variola minor virus genome DNA sequences. *Virology*. **266**, 361–386.
26. Cheng S., Fockler C., Barnes W.M., Higuchi R. 1994. Effective amplification of long targets from cloned inserts and human genomic DNA. *Proc. Natl. Acad. Sci*. USA. **91**, 5695–5699.
27. Babkin I.V., Shchelkunov S.N. 2006. Time scale of poxvirus evolution. *Mol. Biol*. **40**, 20–24.
28. Shchelkunov S.N., Totmenin A.V. 1995. Two types of deletions in orthopoxvirus genomes. *Virus Genes*. **9**, 231–245.
29. Shchelkunov S.N., Totmenin A.V., Kolosova I.V. 2002. Species-specific differences in organization of orthopoxvirus Kelch-like proteins. *Virus Genes*. **24**, 157–162.
30. Basu R.N., Jezek Z., Ward N.A. 11979. *The eradication of smallpox from India*. New Delhi: World Health Organization Southeast Asia Regional Office.