Paediatrica Indonesiana

VOLUME 48 March • 2008 NUMBER 2

Original Article

Differences of nucleotide and amino acid sequences of nucleoprotein (N) gene between wild-type measles virus and vaccine virus in Indonesia

Made Setiawan,[1] Agus Sjahrurachman,[2] Fera Ibrahim,[2] Agus Suwandono.[3]

Abstract

Background Measles virus is a member of genus morbilivirus which belongs to family paramyxovirus with negative, single-strand RNA genome. RNA is packed by nucleocapsid (N) protein. The N protein is very important for RNA replication and translation. Abnormality in N protein will induce abnormality in virus replication.

Objective This study aimed to explore the differences of nucleotide sequence of N gene and amino acid sequences of N protein between wild-type measles virus (G2, G3 and D9) and vaccine virus (CAM-70, Schwarz and Edmonston-wt)

Methods The exctraction and amplification of the gene were conducted in the laboratory using biomolecular technology. The gene and protein analysis were conducted using the bioinformatic technology.

Results The results showed that more differences were found between nucleotide sequences of N gene of wild-type measles virus against CAM-70 vaccine virus (77 – 79 nucleotides) compared against Schwarz and Edmonston-wt vaccine virus (71-74 nucleotides). Likewise, more differences were also observed between amino acid sequences of N protein of wild-type measles virus against CAM-70 vaccine virus (18-24 residues) compared against Schwarz and Edmonston-wt vaccine virus (17-23 residues). **[Paediatr Indones 2008;48:81-87]**.

Keywords: *measles, wild-type, vaccine-type, n gene.*

Measles virus, a member of genus morbilivirus from family paramyxovirus, is an enveloped virus containing negative, single-strand RNA genome. The RNA genome encodes for N, F, P/V/C, H, M and L proteins.[1,2] Viral RNA is packed with nucleocapsid protein to become ribonucleoprotein particles that have helical form or nucleocapsid. The combination of RNA and N protein is very stable as nucleocapsid can withstand against high salt concentration and nuclease enzyme digestion. Nucleoprotein is one of the important components for viral replication. During viral replication, nucleoprotein works together with P (phosphoprotein) and L (polymerase) proteins in replication and translating viral RNA. The first nucleoprotein mRNA is transcripted from genome, and N protein is the most abundant amongst other proteins in measles virus.[3-5]

From the Department of Child Health, Infectious Diseases Hospital Prof. Sulianti Saroso, Jakarta, Indonesia (MS).[1] From the Department Microbiology, Medical School, University of Indonesia, Jakarta, Indonesia (AS, FI).[2] From the National Institute Health Research and Development, Indonesia (AS).[3]

Reprint requests to: Made Setiawan, MD, Department of Child Health, Infectious Diseases Hospital Prof. Sulianti Saroso, Jakarta, Indonesia. Tel. +62-21-6506568, Fax. +62-21-6506568.

Based on its function, N protein can be divided into three domains, which are binding site for P protein to form N-P encapsidation complex, binding site for other N protein to assemble nucleocapsid, and binding site for RNA to start the elongation during RNA assembly. When abnormality occurs in the structure of N protein, the viral replication process will be disturbed.[6]

There is one much conserved region and there are two very variable regions within the N gene.[7] The most protected amino acid sequence of the N protein is the N-terminal where the P binding site resides, while the variable regions are residue number 130-150 and the 125-residues of C-terminal.[8,9] Data clearly showed that there are many branches of wild-type measles virus. All current wild-type measles virus have undergone genetic drift relatively against vaccine strains and wild-type strains isolated in 1950 and 1960. Overall, the 456 nucleotide COOH terminal of N gene have almost 9% differences on the nucleotide level, and this sequence contains region with the most genetic differences.[10] The variability of the nucleotide sequence of N gene leads to the classification of 21 genotypes.[11] This molecular-based classification of measles virus is very valuable for exploring the molecular epidemiology of measles virus in the world.[12]

Moreover, when measles virus infects a host, the immune system of the host will respond to N protein with either cellular or hummoral immune responses. Antibodies against N protein can be used to diagnose measles disease, as the occurring antibody titer is relatively higher compared to antibody against other proteins.[13]

Currently there are three known CTL epitopes at N protein, which are amino acids number 52-59, 81-88 and 281-289.[14] Two epitopes for helper T-cell are amino acids number 57-98 and 457-525.[9,15] It also has been known that there were genetic and antigenic property differences between wild-type measles virus and vaccine virus in United States.[16]

There are three known measles virus genotypes in Indonesia which are G2, G3 and D9. Meanwhile, the most common vaccines currently being used are CAM-70 and Schwarz.[17] The purpose of this research is to observe the nucleotide sequence differences of N gene and amino acid sequence of N protein between wild-type virus (G2, G3 and D9 genotype) and vaccine measles virus (CAM-70 and Schwarz) in Indonesia.

Methods

Wild-type measles virus isolates being used were provided by Litbangkes RI. The genotypes of each isolates, which are G2, G3 and D9, have been determined by ICDC (Atlanta-USA). The genotypes of each isolates were also redetermined in Indonesia with same results.[18] The sequences of N gene for Schwarz and Edmonston-wt vaccine virus were downloaded from Genbank (**Table 1**).

Viral RNA extraction was performed using QIAmp Viral RNA Mini Kit from QIAGEN (cat. no. 5204) following the instruction from the manufacturer. Specimens being used were culture products. The method is as follow. Cells infected by virus were taken from the bottle and were centrifuged by 1500X G at 4 degree C for 15 minutes. The resultant supernatant was discarded, while the pellet was taken for further RNA extraction. RNA was extracted by means of pipetting 560 µl AVL buffer containing RNA Carrier into 1.5 ml Eppendorf tube. About 140 µl virus sample was added and was vortexed for 10-15 seconds, incubated for minimum of 10 minutes, and then centrifuged at 6000X g for 30 seconds. Absolute alcohol was added for 560 µl, further vortex was performed and centrifugation was performed for a

Table 1. Wild-type isolates being analyzed

Code	Address	IgM	Age (yr)	Sex	Genotype
MVi/INA/06.02/161Yo	Subang-Jabar	+	4.5	L	D9
MVi/INA/05.02/Ba	Gresik-Jatim	+	4	P	G3
Mvi/INA/03.04/362 Sep	Pekalongan	+	3.8	P	G2
Virus vaksin CAM-70	Bio Farma				
Virus Vaksin Schwarz*	NCBI Genbank				
Virus Edmonston-wt*	NCBI Genbank				

* Sequence data obtained from Genbank

moment. As much as 630 µl of the solution was transferred to the second spin column. Centrifugation was performed at 6000X for 1 minute. The collection tube containing the filtrate was disposed away and replaced with a clean collection tube. AW1 wash buffer was added for 500 µl, and further centrifugation was performed at 6000X g for 1 minute. Collection tube was then replaced with the new one with additional 500 µl AW2 wash buffer added to the tube, and then another centrifugation at 20,000X g was performed for 3 minutes. Collection tube was replaced with the clean one, and then further centrifugation was performed at 20,000X g for 1 minute (to clean up the spin column for the buffer and alcohol). Collection tube was replaced with 1.5 ml Eppendorf tube. AVE buffer was added as much as 60 ul (right in the middle of the filter). Incubation was performed at room temperature for 1 minute. Centrifugation was next performed at 6000X g for 1 minute. The RNA from elution was kept at temperature -20° C or -70° C.

RNA was amplified by PCR machine using primers designed from the analysis of reference sequence from Edmonston-wt substrain AIK-C with accession number AB046218.[19] The program used to design the primer is Primer-3.

Reverse transcription reaction was performed by SuperScript III Reverse Transcriptase (Invitrogen) using the protocol supplied by the manufacturer. The mix of reverse transcription reaction was made by adding 1.5 µl random hexamer, 10 mM 1 µl dNTPs, 0.5 µl dH20 and 10 µl RNA. The mix was heated at 65° C for 5 minutes, and then incubated in ice for 1 minute. The mix was centrifuged for a moment, and then was added with 4 µl 5X first-Strand buffer, 1 ul 0.1 M DTT, 1 ul RNAsin, 1 ul Superscript III RT. The mix was stirred carefully with a pipette, and then was incubated at 25°C for 5 minutes, at 50°C for 30-60 minutes, and the reaction was stopped by heating at 70°C for 15 minutes. The obtained cDNA was capable to be used as template for PCR amplification.

The cDNA from reverse transcription reaction was amplified with PCR method using Platinum Taq DNA Polymerase (Invitrogen). The primers used were showed in **Table 2**. The reaction mix was made by adding 10 µl cDNA, 1 µl 20 uM primer1, 1 µl 20 uM primer2, 1.5 µl MgCl2, 5 µl 10 mM dNTPs, 5 µl 10X PCR buffer and water/aquadest was added to fill up to 50 µl, 1 ul Platinum Tag, and was stirred carefully and was added with 1 drop of mineral oil. The reaction mix was heated at 94° C fro 2 minutes for denaturation process. The PCR reaction was performed for 40 cycles, consisted of denaturation for 30 second at 94° C, annealing for 60 seconds at 56° C, and extension for 120 seconds at 72° C, and ended with incubation at 72° C for 7 minute.

PCR product was analyzed with gel electrophoresis whose detection was based on molecular weight or length of nucleotide chain. The N gene has DNA nucleotide with length of 1685 bp. To detect the PCR product, the conditions were 2% agarose gel mixed with 10 µl (1µg/ml) ethidium bromide, with 100 Volt voltages for 40 minutes.[20] DNA product was purified using QIAquick Gel Extraction Purification (QIAGEN catalog no. 28704), where the PCR product was run in the low melting agarose first before purification. The protocol used in the experiment followed the instruction from the manufacturer.

The purified PCR product was cloned using TOPO TA Cloning kit from Invitrogen (catalog no. KA4500-01) following the protocol supplied by the manufacturer. Plasmid extraction was performed using QIAprep Spin Miniprep (QIAGEN catalog no. 27104) by following the protocol from the manufacturer.

The PCR product was sequenced by direct sequencing using Perkin Elmer ABI 337 Automatic DNA Sequencer which utilized Sanger method (dideoxy termination method). The primers used for sequencing were similar with those for PCR amplification with ABI PRISM Big Dye Terminator

Table 2. Primers being used for PCR and sequencing

No.	Primer name	Nucleotide sequences	Position
1	Primer N 1	CAG GGA CAA gAg CAg gAT TA	75-95
	2	gTT CCT CAC CAC ATC CAA CC	751-770
2	Primer N 3	gAA ggT ggA TAA AgT ACA CCC AAC	688-711
	4	CTT TgA TCA CCg TgT AgA AAT gAT	1358-1382
3	Primer N 5	CAC ATT ggC ATC TgA ACT Cg	1241-1260
	6	CAA TgA Tgg Agg gTA ggC Ag	1714-1734

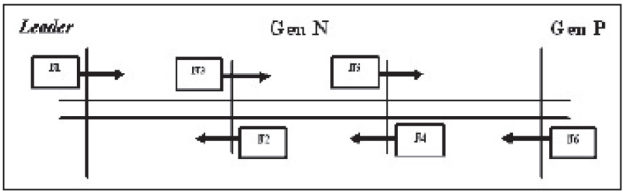

Figure 1. Strategy for sequencing N protein in overlapping regions

Cycle Ready Reaction Kit for the PCR sequencing reaction (**Figure 1**).

The raw data generated by the computer reading were reanalyzed manually by examining the electrophoresis trace and comparing with the available sequence data (Edmonston-wt, CAM-70, and Schwarz). The examined and edited fragments were assembled to create the whole nucleotide sequence with the help of Genetyx and Bioedit software.

The completed gene sequences were compared with one another and with the sequence data obtained from NCBI Genbank. All comparison was performed using Genetyx, ClustalW and Bioedit, which can be downloaded from http://www.mbio.nesu.edu/Bioedit/bioedit.html.

The overall mean distance was performed with Mega 2.1 using Tamura-Nei method, while the amino acid overall distance was calculated using Poisson correction method.

Results

Nucleotide sequence differences

To observe the nucleotide sequence differences of N gene, alignments of N gene sequences from G2, G3, D9 genotypes with Edmonston-wt vaccine virus were made. Alignment between G2 genotype and Edmonston-wt showed 71 nucleotide differences (4.5%) which leaded to 18 residues of amino acid substitutions. Alignment between G3 genotype and Edmonston-wt showed 71 nucleotide differences (4.5%) which leaded to 19 residues of amino acid substitutions. Alignment between D9 genotype and Edmonston-wt showed 72 nucleotide differences (4.6%) which leaded to 17 residues of amino acid substitutions.

Next, alignments of nucleotide sequences of N gene from G2, G3 and D9 genotypes with Schwarz were made. Alignment between G2 genotype and Schwarz showed 73 nucleotide differences (4.6%) which leaded to 17 residues of amino acid substitutions. Alignment between G3 genotype and Schwarz showed 74 nucleotide differences (4.6%) which leaded to 23 residues of amino acid substitutions. Alignment between D9 genotype and Schwarz showed 74 nucleotide differences (4.6%) which leaded to 19 residues of amino acid substitutions.

Alignment between G2 genotype and CAM-70 vaccine virus showed 77 nucleotide differences (4.9%) which leaded to 19 residues of amino acid substitutions. Alignment between G3 genotype and CAM-70 vaccine virus showed 79 nucleotide differences (5.0%) which leaded to 24 residues of amino acid substitutions. Alignment between D9 genotype and CAM-70 vaccine virus showed 79 nucleotide differences (5.0%) which leaded to 18 residues of amino acid substitutions.

Those data showed that the greatest nucleotide differences was between G2, G3 and D9 genotypes and CAM-70 vaccine virus, followed by Schwarz and Edmonston-wt vaccine virus (**Table 3**).

Amino acid sequence differences

Types of amino acid substitutions deeply influence the secondary structure of protein, and changes of secondary structure affect the antigenic property of antigen proteins. After performing analysis of amino acid differences based on the amino acid property

Tabel 3. Nucleotide and amino acid differences between wild-type and vaccine virus

Wild-type genotyps	Nucleotide differencs			Amino acid differences		
	Edmons	Schwaz	CAM70	Edmons	Schwaz	CAM70
G2	71 (4.5 %)	73 (4.6 %)	77 (4.9 %)	18 (3.4 %)	17 (3,2 %)	19 (3,6 %)
G3	71 (4.5 %)	74 4.6 %)	79 5.0 %)	19 (3.6 %)	23 (4,4 %)	24 (4,4 %)
D9	72 (4.6 %)	74 (4.6 %)	79 (5.0 %)	17 (3.2 %)	19 (3,6 %)	18 (3,4 %)

groups, there were 35 amino acid positional differences on N protein between wild-type measles virus (G2, G3 and D9) and CAM-70 vaccine virus, and from those differences there were 21 residues which had different amino acid property groups. This could lead to differences in the secondary structure of N protein between wild-type measles virus and vaccine virus.

Amino acid differences of N protein between wild-type measles virus and Schwarz vaccine virus were found in 34 positional residues, and about 22 residues of those belonged to different amino acid property groups. This could cause secondary structure differences between wild-type and vaccine virus. From this analysis, it was predicted that wild-type measles virus in Indonesia had closer relationship to Schwarz vaccine virus compared to CAM-70 vaccine virus (**Table 4** and **Table 5**).

Discussion

Analysis results of nucleotide and amino acid sequence alignments of N protein showed that there were differences between G2, G3 and D9 genotypes with CAM-70, Schwarz and Edmonston-wt vaccine virus. Nucleotide and amino acid sequence differences were observed highest against CAM-70 vaccine virus (77-79 nucleotides), followed by Schwarz vaccine virus (73-74 nucleotides) and Edmonston-wt (71-72 nucleotides). All these led to predicted amino acid substitutions around 17-24 residues. From the differed amino acids, around 59% were found to have different amino acid properties. Predicted many difference amino acids could affect the protein structures which in turn could affect the biological properties of N protein from each virus. Taylor, *et al*, (1991)[10] and Rota *et al*, (1994)[9,15] found many lineages of wild-type measles virus which was caused as the differences in the region of 456 bp at the COOH-terminal and differences at whole nucleotide sequence of H gene, so that virus could be classified into several viral genotypes. These viral genotypes which based on the molecular level had been used to study the measles virus epidemiologically around the world.[12]

Table 4. Amino acid differences between wild-type virus and CAM-70

No	G2	G3	D9	CAM-70		No	G2	G3	D9	CAM-70	
10	L	S	L	L	+	405	R	R	R	K	-
						406	I	T	I	I	+
52	L	G	L	L	-	431	G	G	G	R	+
55	L	W	L	L	-	434	G	G	G	A	-
58	L	L	L	S	+	438	K	K	R	R	-
132	L	L	S	S	+	439	R	R	K	R	-
136	S	S	S	P	-	456	H	P	S	P	+
137	S	S	S	P	-	460	N	N	S	S	-
139	G	S	G	S	+	462	A	A	E	A	+
142	S	S	F	S	+	466	D	H	D	H	+
149	S	S	S	F	+	470	G	G	S	G	+
215	C	W	W	W	+	481	F	F	S	S	+
242	R	G	R	R	+	482	S	S	G	S	+
247	N	K	K	K	+	505	S	S	L	S	+
312	Y	Y	Y	C	-	510	S	L	L	S	+
320	I	L	L	I	-	512	M	M	T	T	+
324	F	F	F	L	-	516	R	R	K	I	-
357	F	L	F	F	-	522	D	D	D	N	+

Note:+ amino acid with different groups/property
- amino acid with same groups/property

Tabel 5. Amino acid differences between wild-type virus and Schwarz

No	G2	G3	D9	Schw.	
10	L	S	L	L	+
52	L	G	L	L	-
55	L	W	L	L	-
132	L	L	S	S	+
136	S	S	S	P	-
137	S	S	S	I	+
139	G	S	G	S	+
142	S	S	F	S	+
144	S	S	S	F	+
148	E	E	E	G	+
215	C	W	W	W	+
224	P	A	A	A	-
242	R	G	R	R	+
247	N	K	K	K	+
320	I		L	L	I
357	F	L	F	F	-
405	R	R	R	K	-
406	I	T	I	I	+
431	G	G	G	R	+
438	K	K	R	R	-
439	R	K	R	R	-
456	H	P	S	P	+
460	N	N	S	S	-
462	A	A	E	A	+
466	D	H	D	H	+
470	G	G	S	G	+
479	S	S	S	T	-
481	F	F	S	S	+
482	G	S	S	S	+
505	S	S	L	S	+
510	S	L	L	S	+
-	512	M	M	T	T +
516	R	R	K	I	-
522	D	D	D	N	+

Note:+ amino acid with different groups/property

- amino acid with same groups/property

Cytotoxic T lymphocyte (CTL) response is very important for the recovery process of measles disease. During the process, the most important CTL response is against N protein compared to the glycoproteins. It was said that CTL response against glycoproteins increased the mortality number of the guinea pigs, while response against N protein could decrease the mortality number. Based on the results of the researches, N protein could induce CTL twice as high compared to glycoproteins.[14]

Cytotoxic T lymphocyte CD8+ anti measles virus is very important during and after the acute infection happens. The N protein is the only antigen of the measles virus that could raise strong CTL response.[14] CTL epitope in the N protein is the most efficient in sensitifying cell. When several amino acids of the epitope changes, the ability to sensitively cell will differ, hence the ability to destroy virus will differ too.[14]

T-cell helper cell is very important in inducing B-cell to produce antibodies and increase CTL function.[21] T-cell helper specific epitope of N protein resides in the amino acid 57-98.[9,15] If different amino acid sequences were found on the T-cell, this would cause differences in T-cell helper sensitivity.

From the results of analysis of N-gene nucleotide sequences and N-protein amino acid sequences, it could be concluded that:

1. differences of the nucleotide sequences of N gene were greatest between wild-type and CAM-70 vaccine virus, compared to Schwarz and Edmonston-wt,
2. differences of amino acid sequences of N protein were greatest between wild-type and CAM-70 vaccine virus compared to Schwarz and Edmonston-wt with 60% of those differ amino acids had different properties which could lead to changes in the conformation and antigenicity of the protein.

Hence, based on the nucleotide and amino acid sequences, wild-type measles virus (G2, G3 and D9) were genetically closer to Schwarz vaccine virus compared to CAM-70 vaccine virus.

Based on the results, it was suggested to:

1. perform studies on the configuration of N protein

to observe the relationship between the variability of nucleotide and amino acid sequence with the protein structures,

2. perform both serologic and genetic analysis of vaccine virus before deciding on the vaccine that would be used nationally, in order to have vaccine that suitable with measles virus currently distributing in Indonesia and have more optimal serological results.

Acknowledgments

The authors would like to thank Dr. I Nyoman Kandum MPH, the director of P2M & PLP who had funded this study. The authors would also like to thank Mr. Harun, Joko, Bambang at Litbangkes, Diah Iskandriati, Joko Pamungkas, Uus Saefullah, Silmi at PSSP IPB Bogor who had helped in finishing this study.

References

1. Tyrell DIJ, Norrby F. Structural polypeptides of measles virus. J Gen Virol 1978;39:219-229.
2. Rima BK. The proteins of morbilli viruses. J Gen Virol 1983;64:1205-1219.
3. Bellini WJ, Rota JS, Rota PA. Virology of measles virus. J Infct Dis 1994;170:S15-23.
4. Lamb RA, Kolakofsky D. Paramyxoviridae: the viruses and their Replication; In Field Virology; Third ed. Vol.1; Philadelphia-New York: Lippincott-Raven, 1996; p.1177-1204.
5. Griffin, DE, Bellini, WJ. Measles Virus. In: Fields Virology. 3rd ed. Philadelpia-New York: Lippincott-Raven, 1996; p. 1267-1312.
6. Bankamp B, Horikami SM, Thompson PD, Huber M, Billeter M, Moyer SA. Domains of the measles virus N protein required for binding to P protein and self-asembly. Virology 1996;216:272-7.
7. Giraudon P, Jacquier MF, Wild TF. Antigenic analysis of african measles virus field isolates: identification and localization of one conserved and two variable epitope site on the NP protein. Virus Res 1988;18:137-152.
8. Baczko K, Pardowitz I, Rima BK, Meulen VT. Constan and variable regions of measles virus proteins encoded by the nucleocapsid and phosphoprotein gen derived from lytic and persistent viruses. Virology 1992;190:469-474.
9. Rota PP, Bloom AE, Vanehiere JA, Bellini WJ. Evolution of the nucleoprotein and matrix gene of wild-type strains of measles virus isolated from recent epidemics. Virology 1994; 198:724-730.
10. Taylor MJ, Godfrey E, Baczko K, ter Meulen V, Wild TF, Rima BK. Identification of several different lineages of measles virus. J Gen Virol 1991;72:83-8.
11. WHO: Expanded program on immunization. Standardization of the nomenclature for describing the genetic characterization of wild-type measles viruses. Weekly Epidemiological report 2001;76:241-248.
12. Rota JS, Heath JL, Rota AP, King GE, Celma ML, Carabana J, *et al*. Molecular epidemiology of measles virus: Identification of pathways of transmission and implications for measles elimination. J Infec Dis 1996;173:32-7.
13. Grave M, Griffin DE, Johnson RT, Hirsch RL, Soriano L, Roedenbeck S, *et al*. Development on antibody to measles virus polypeptides during complicated and uncomplicated measles virus infections. J Virol 1984;49:409-412.
14. Beauverger P, Chadwick J, Buckland, R, Wild TF. Serotype-spesific and canine distemper virus cross-reactive H-2KK-restricted cytotoxic T lymphocyte epitopes in the measles virus nucleoprotein. Virology 1994;203:172-7.
15. Rota JS, Wang ZD, Rota PA, Bellini WJ. Comparison of sequnces of the H, F, and N coding genes of measles virus vaccin strains. Virus Res 1994;31:317-30.
16. Tamin A, Rota PA, Wang Z, Heath JL, Anderson LJ, Bellini WJ. Antigenic analysis of current wild type and vaccine strains of measles virus. J Infect Dis 1994;170:795-801.
17. Litbangkes DepKes RI.Laporan hasil genotipe virus campak yang dikirim oleh WHO. (2002)
18. Setiawan M. Analisis genetik dan antigenik virus campak liar dan virus vaksin di Indonesia [Dissertation]. Jakarta: Fakultas Kedokteran Universitas Indonesia; 2005.
19. Komase K, Suzuki N, Nakayama T, Miki K, Kawanishi R, and Fukuda K. Genom sequence of measles virus. NCBI no. accession: AB046218 (2001).
20. Coligan JE, Kruisbeek AM, Margulies DH, Shevach, EM, Strober W. Current protocols in immunology. Vol. I. National Institut of Health; Published by Current Protocols Wiley, (1996).
21. Flint SJ, Enquist LW, Krug RM, Racaniello VR, Skalka AM. Principle of virology, moleculer biology, pathogenesis and control. Washington DC: ASM press; 2000. p. 479-515.