Genetic characterization of wild-type measles viruses circulating in suburban Khartoum, 1997–2000

H. S. El Mubarak,[1,2] M. W. G. van de Bildt,[1] O. A. Mustafa,[2] H. W. Vos,[1] M. M. Mukhtar,[2] S. A. Ibrahim,[2] A. C. Andeweg,[1] A. M. El Hassan,[2] A. D. M. E. Osterhaus[1] and R. L. de Swart[1]

[1] Institute of Virology, Erasmus MC, PO Box 1738, 3000 DR Rotterdam, The Netherlands

[2] Institute of Endemic Diseases, University of Khartoum, PO Box 102, Khartoum, Sudan

Measles remains endemic in many East African countries, where it is often associated with high morbidity and mortality. We collected clinical specimens from Sudanese measles patients between July 1997 and July 2000. Sequencing of the 3′ 456 nucleotides of the nucleoprotein gene from 33 measles virus (MV) isolates and 8 RNA samples extracted from clinical specimens demonstrated the presence of a single endemic MV strain with little sequence variation over time (overall nucleotide divergence of 0 to 1·3%). This was confirmed by sequencing of the complete H gene of two isolates from 1997 and two from 2000, in which the overall divergence ranged between 0 and 0·5%. Comparison with MV reference strains demonstrated that the viruses belonged to clade B, genotype B3, and were most closely related to a set of viruses recently isolated in Nigeria. Our study demonstrates a remarkable genetic stability of an endemically circulating MV strain.

Measles virus (MV) is considered a monotypic and antigenically stable virus, which has long masked the presence of nucleotide variation (Schneider-Schaulies & ter Meulen, 1999). Sequence analysis of vaccine and wild-type MV isolates reveals the presence of nucleotide variation over the entire genome, the most variable genes being the nucleoprotein (N) and the haemagglutinin (H) genes with about 7% nucleotide variability between the most distantly related sequences (Taylor *et al.*, 1991; WHO, 1998, 2001). The 456 nucleotides encoding the carboxyl terminus of the nucleoprotein are the most variable region of the MV genome, as the nucleotide variability in this region can exceed 12% between different genotypes (Taylor *et al.*, 1991; WHO, 2001). This variability is unaffected by propagating the virus in tissue culture, making it

a suitable tool for the molecular characterization of wild-type MV isolates (Rima *et al.*, 1995). This region has now been designated by the WHO as the minimum sequence required to genotype a wild-type MV isolate (WHO, 1998). Complete H gene sequences should be obtained from viruses that may represent new genotypes, or from representative viruses of a larger set of viruses (WHO, 2001). The generation of a database of sequences of different wild-type MV isolates allowed phylogenetic studies, resulting in the identification of eight clades and a still increasing number of genotypes (WHO, 1998, 2001). Some of these are considered inactive, but the majority are still actively circulating, and to some extent geographically restricted (Kreis *et al.*, 1997; Bellini & Rota, 1998). Genetic characterization of wild-type isolates, when combined with standard epidemiological methods, may enable the identification of source and transmission pathways of wild-type viruses and permit differentiation between indigenous and imported viruses (Rima *et al.*, 1995; Rota *et al.*, 1996; WHO, 1998; Hanses *et al.*, 2000). Characterization of globally circulating MV strains, particularly from countries where measles remains endemic, will therefore contribute to these studies.

One of the regions of the world where measles remains endemic is East Africa. However, no MV isolates from this region have been sequenced, except for the NY 94 and NY 96 isolates which were isolated in New York and epidemiologically linked to Kenya (Rota *et al.*, 1996; Truong *et al.*, 1999). The sequence database of the available African MV isolates shows the complexity of the distribution of virus genotypes. In Southern Africa viruses of clades A and D were found to predominate (Kreis *et al.*, 1997; WHO, 1998; Truong *et al.*, 1999). Clade B viruses were shown to predominate in Western and Central Africa. This clade comprises three genotypes, B1, B2 and B3 (Rota *et al.*, 1994, 1996; Hanses *et al.*, 1999; Truong *et al.*, 1999; WHO, 2001).

We studied the phylogenetic characteristics of 41 wild-type MV sequences obtained in suburban Khartoum between July 1997 and July 2000. The majority of the patient samples (*n* = 18) had been collected in 1997, and the rest in 1998 (*n* = 10), 1999 (*n* = 8) and 2000 (*n* = 6) (see Table 1). Heparinized

Author for correspondence: Rik de Swart.

Fax +31 10 408 9485. e-mail deswart@viro.fgg.eur.nl

Table 1. Summary of sequences determined

| Patient ID | Sequence available (WHO nomenclature) | GenBank accession no. | | Material* | Year |
		N gene	H gene		
SM11	MVi/Khartoum.SUD/23.97	AF149065		vi	1997
SM23	MVi/Khartoum.SUD/26.97	AF311785		vi	1997
SM24	MVi/Khartoum.SUD/28.97	AF311786		vi	1997
SM25	MVi/Khartoum.SUD/30.97/1	AF311787		vi	1997
SM26	MVi/Khartoum.SUD/30.97/2	AF149064		vi	1997
SM27	MVi/Khartoum.SUD/31.97/1	AF149066		vi	1997
SM29	MVi/Khartoum.SUD/33.97	AF311788	AF453430	vi	1997
SM31	MVi/Khartoum.SUD/34.97/1	AF149061		vi	1997
SM32	MVi/Khartoum.SUD/34.97/2	AF149063		vi	1997
SM34	MVi/Khartoum.SUD/34.97/3	AF311789		vi	1997
SM37	MVi/Khartoum.SUD/35.97	AF149062		vi	1997
SM40	MVi/Khartoum.SUD/36.97/1	AF311790		vi	1997
SM42	MVi/Khartoum.SUD/36.97/2	AF311791	AF453431	vi	1997
SM44	MVi/Khartoum.SUD/37.97	AF311792		vi	1997
SM45	MVi/Khartoum.SUD/38.97/1	AF311793		vi	1997
SM46	MVi/Khartoum.SUD/38.97/2	AF311794		vi	1997
SM54†	MVs/Khartoum.SUD/40.97	AF311795		fp	1997
SM64†	MVs/Khartoum.SUD/44.97	AF311796		ts	1997
SM111†	MVs/Khartoum.SUD/6.98	AF311797		fp	1998
SM120	MVs/Khartoum.SUD/21.98	AF311798		ts	1998
SM125	MVi/Khartoum.SUD/22.98	AF311799		vi	1998
SM133	MVi/Khartoum.SUD/26.98	AF311800		vi	1998
SM135	MVi/Khartoum.SUD/27.98/1	AF311801		vi	1998
SM136	MVi/Khartoum.SUD/27.98/2	AF311802		vi	1998
SM149†	MVs/Khartoum.SUD/33.98	AF311803		fp	1998
SM154†	MVs/Khartoum.SUD/34.98	AF311804		fp and ts	1998
SM161†	MVs/Khartoum.SUD/38.98	AF311805		fp	1998
SM170	MVi/Khartoum.SUD/1.99/1	AF311806		vi	1999
SM172	MVi/Khartoum.SUD/5.99	AF311807		vi	1999
SM176	MVi/Khartoum.SUD/8.99	AF311808		vi	1999
SM177	MVi/Khartoum.SUD/11.99	AF311809		vi	1999
SM180	MVi/Khartoum.SUD/13.99	AF311810		vi	1999
SM184	MVi/Khartoum.SUD/22.99	AF311811		vi	1999
SM192†	MVs/Khartoum.SUD/23.99	AF311812		ts	1999
SM194	MVi/Khartoum.SUD/24.99	AF311813		vi	1999
SM00-1	MVi/Khartoum.SUD/28.00/1	AF311814		vi	2000
SM00-2	MVi/Khartoum.SUD/28.00/2	AF311815		vi	2000
SM00-3	MVi/Khartoum.SUD/28.00/3	AF311816		vi	2000
SM00-4	MVi/Khartoum.SUD/28.00/4	AF311817		vi	2000
SM00-5	MVi/Khartoum.SUD/28.00/5	AF311818	AF453432	vi	2000
SM00-6	MVi/Khartoum.SUD/28.00/6	AF311819	AF453433	vi	2000

* vi, virus isolate; fp, blood sample spotted on filter paper; ts, throat swab.
† Samples cloned into PCR 2.1 vector system prior to sequencing.

blood samples, throat swabs and blood samples spotted on filter paper were collected from clinically diagnosed measles patients in suburban Khartoum in the framework of a prospective measles study in suburban Khartoum, as previously described (El Mubarak *et al.*, 2000; De Swart *et al.*, 2001a). Samples were collected within 7 days after onset of rash, upon having obtained informed consent of the parents or guardians. MV isolates ($n = 33$) were obtained by co-cultivation of phytohaemagglutinin-activated peripheral blood mononuclear cells of the Sudanese patients with a human Epstein–Barr virus-transformed B-lymphoblastic cell line of a healthy donor as previously described (El Mubarak *et al.*, 2000). MV isolates were frozen after two or maximally three passages in the same cell line.

Total viral RNA was isolated from MV isolates or throat swab samples using the High Pure viral RNA isolation kit

(Roche Diagnostics). RNA from filter-paper blood samples was extracted by a modification of the High Pure viral nucleic acid isolation kit (Roche), as previously described (De Swart *et al.*, 2001b). First strand cDNA synthesis was primed by random hexanucleotides (Promega) as previously described (El Mubarak *et al.*, 2000). Primers used for amplification of the 456 nucleotides encoding the carboxyl terminus of the N protein were 5′ TTAGGGCAAGAGATGGTAAGG 3′ (MV-N1, positions 1090–1110) and 5′ TTATAACAATGATGGAGGG 3′ (MV-N2, positions 1633–1615). The PCR products from the virus isolates were then directly sequenced using the Dye Terminator cycle sequencing kit (Perkin Elmer), with the same primer set and following the manufacturer's protocol. Sequencing was performed on a 373A automated sequencer (Applied Biosystems). The RT–PCR products obtained from throat swabs or blood spotted on filter paper were in some cases not sufficient for direct sequencing. These samples, indicated in Table 1 by a dagger (†), were cloned into the PCR 2.1 vector system (Invitrogen) prior to sequencing. Mini-preparations of plasmid DNA were isolated by the alkali method (Sambrook *et al.*, 1989) and sequenced with M13 forward and reverse primers. The complete coding region of the H gene was amplified in three overlapping PCR fragments using a slight modification of the primers used by Truong *et al.* (1999) and Chibo *et al.* (2000). The primer combinations were as follows: 5′ TTAGGGTGCAAGATCATCCACA 3′ (MV-H1, positions 7249–7270) and 5′ GACCATTACTGACTGGTTGC 3′ (MV-H6, positions 8120–8100); 5′ CACCTCAGAGATTCACTGACC 3′ (MV-H3, positions 7590–7610) and 5′ GAGCCGTGTGTGATCAATGG 3′ (MV-H8, positions 8586–8567); 5′ GTACCGAGTGTTTGAAGTAGG 3′ (MV-H5, positions 8023–8043) and 5′ GGGTGACATCATGTGATTGG 3′ (MV-H2, positions 9157–9138). The PCR products amplified from the virus isolates were then directly sequenced as described for the N gene, using the above primers and the following additional primers: 5′ GATCTCTGAAGTCGTACTCC 3′ (MV-H4, positions 7676–7657) and 5′ GATCTGAGTCTGACAGTTGAGC 3′ (MV-H7, positions 8516–8537). The sequences determined are available from the GenBank database (accession numbers are shown in Table 1).

The sequences obtained were aligned with the reference sequences representing the different genotypes described by WHO (2001) by using the Clustal-W function of the BioEdit program (T. Hall, Department of Microbiology, North Carolina State University, USA). Distance matrices were calculated using maximum likelihood function of the PHYLIP package (Felsenstein, 1993) and the phylogenetic relation was then inferred using the neighbour joining method of the PHYLIP package (Felsenstein, 1993) in combination with bootstrap analysis (100 replications).

The 41 partial N gene sequences studied were closely related, as over the 3 year period a divergence of only 0 to 1·3 % was found in this hypervariable region of the MV genome. Half of the point mutations were silent. One nucleotide mutation, ^{1370}A → G (numbering according to Mori *et al.*, 1993), was found in five out of nine 1998 sequences, seven out of eight 1999 sequences and five out of six 2000 sequences, but in none of the sequences obtained in 1997. This mutation was not found in any MV strain described in GenBank (BLAST search). Since the data set was found to be homogeneous, we randomly selected two virus isolates from 1997 and two from 2000 for which the nucleotide sequence of the H gene was determined. Analysis of these sequences confirmed the observed high homology, as the 1997 isolates differed from the 2000 isolates by only four nucleotide mutations (0·2 %), all of which were silent. As shown in Fig. 1, the Sudanese sequences clustered with the reference strain of genotype B3.

For a more comprehensive analysis of the relationship between the Sudanese viruses and other clade B virus strains, a dendrogram was made of a larger group of the Sudanese N-sequences and a large number of previously described clade B N-sequences (Fig. 2). From this analysis, it became apparent that the Sudanese sequences were most closely related to a group of sequences which was recently described in Nigeria, and assigned to genotype B3, cluster 1 (Hanses *et al.*, 1999).

Although our study area was rather restricted, the highly conserved sequences among the isolates over a 3 year period strongly suggest that our isolates represent an indigenous and probably dominant strain circulating in greater Khartoum and probably elsewhere in Sudan. The continuous arrival of new refugees from the south and west of Sudan in the main study area suggests that at least in these areas the same strain predominates.

The observed high conservation over the 3 year period suggests a relatively high genetic stability of wild-type MV isolates. These results are in agreement with previously observed low mutation rates in H gene sequences of MV isolates collected in Madrid from 1993 to 1996 (Rima *et al.*, 1997). Similarly, during the last 3 years of endemic MV circulation in the USA, H and N gene sequences differed by less than 0·5 % (Rota *et al.*, 1996). However, in a previous study in Nigeria and Ghana a much higher variability was found between different virus isolates analysed (Hanses *et al.*, 1999; Truong *et al.*, 1999). This difference may be explained by two factors. Firstly, travel in and out of Sudan is more restricted than travel in and out of Nigeria, which reduces the chances of importation of new MV strains into Sudan. Secondly, the level of sequence variation in a situation of endemic virus circulation will be directly related to the absolute number of simultaneously infected patients (i.e. chains of transmission), since each infection may potentially result in the occurrence of new mutations. The numbers of MV-infected patients in Nigeria may indeed be higher than in Sudan, since reported MV vaccination coverage in Nigeria is much lower than that in Sudan, 26 % and 63 % respectively (UNICEF, 2000; De Swart *et al.*, 2001a). Other phylogenetic studies in areas with high levels of measles transmission (e.g. China, Vietnam)
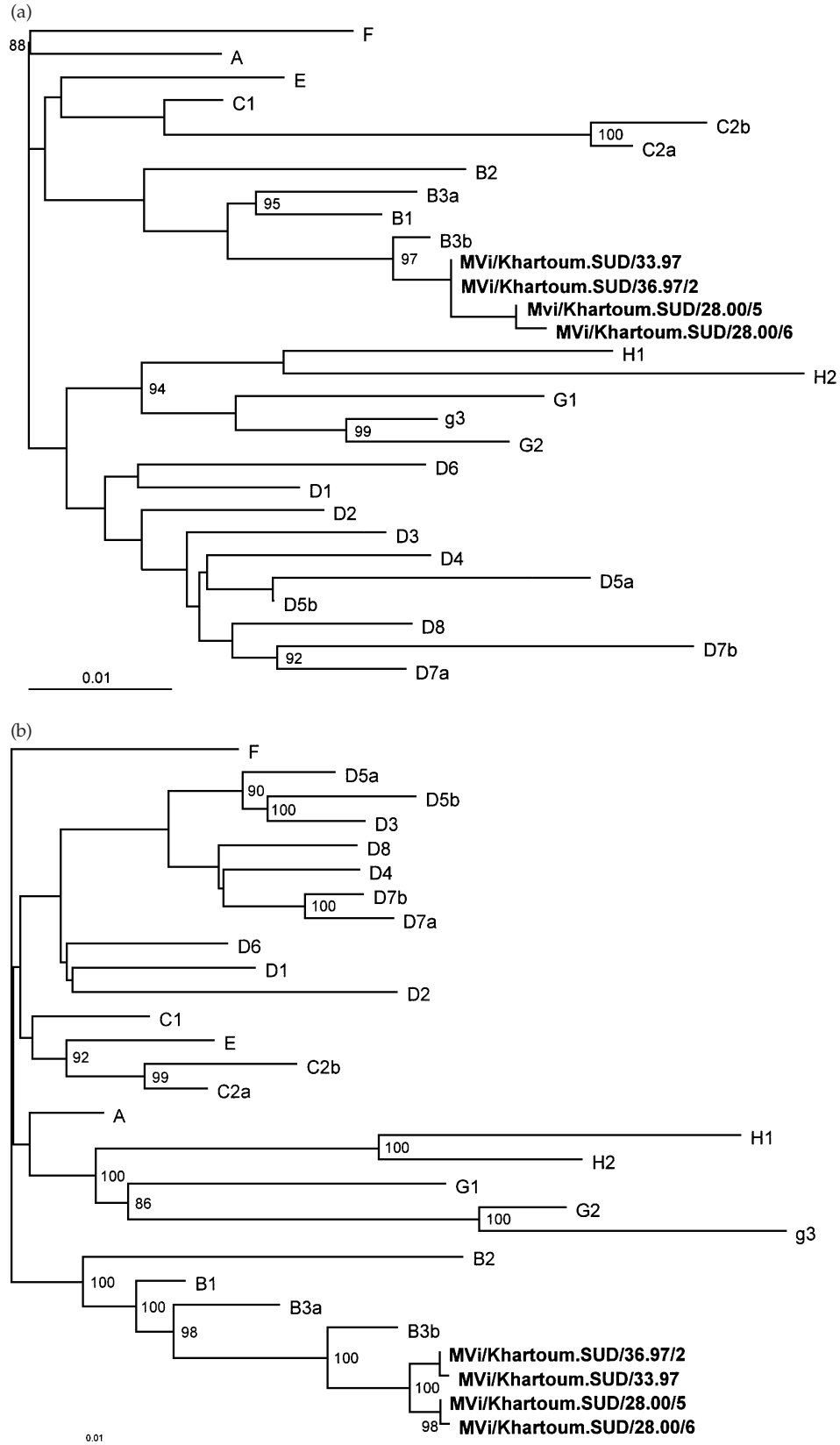
Fig. 1. Phylogenetic comparison of sequences of the 3′ 456 nucleotides of the N gene (a) or the complete H gene (b) of four selected Sudanese MV isolates (in bold) with the MV reference strains. Designated reference strains were as published (WHO, 2001). The phylogram was generated by analysing 100 bootstrap replicates; bootstrap percentages are shown when above 80%.
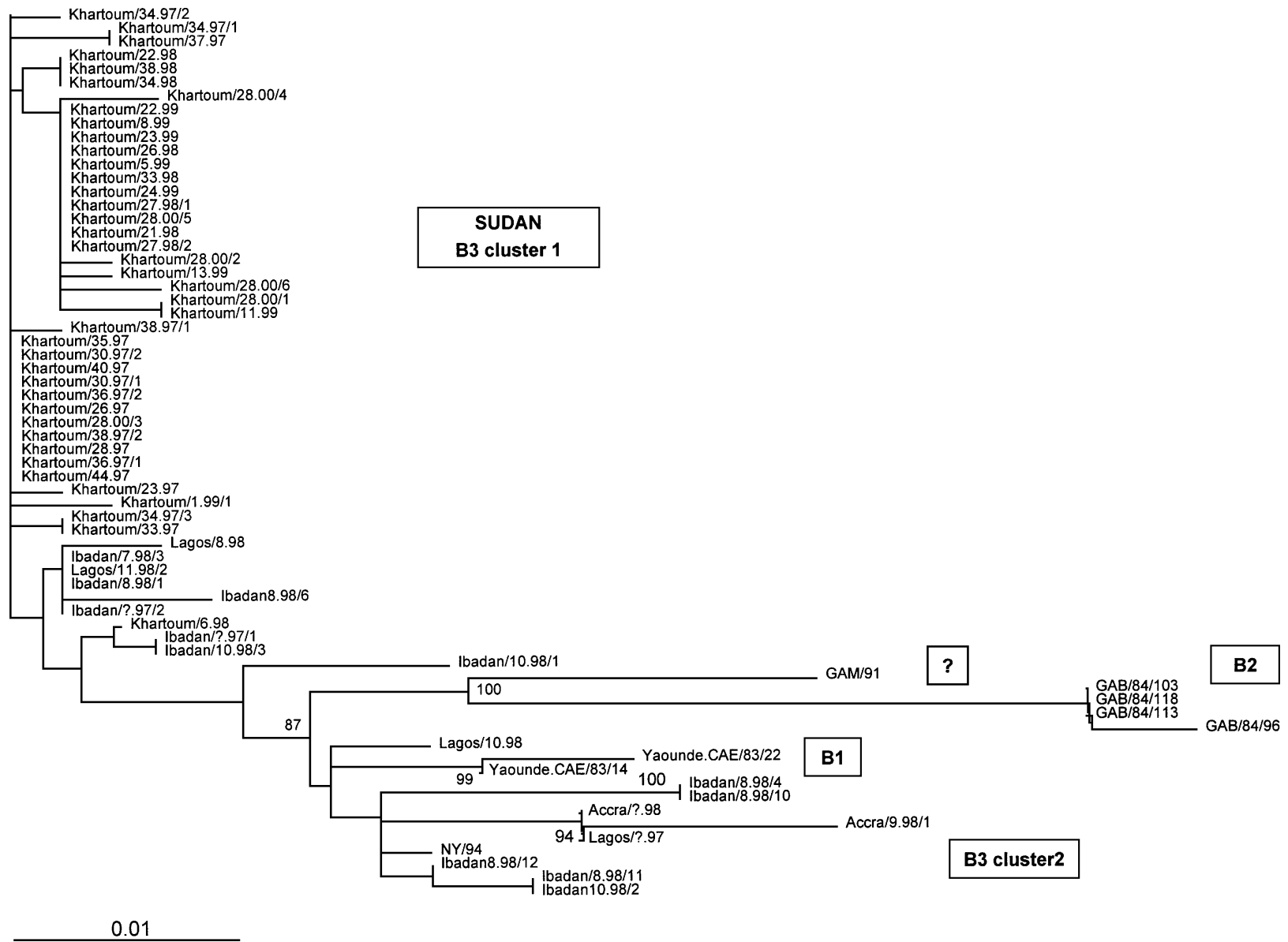
Fig. 2. Phylogenetic comparison of the 3′ N gene sequences of an extended set of clade B MV isolates.

also found a relatively high variability within the locally circulating genotype (Xu *et al.*, 1998; Liffick *et al.*, 2001).

Comparison of the Sudanese isolates with the reference sequences of the known wild-type MV clades and genotypes allowed the assignment of the isolates to clade B, in which isolates from Central and Western Africa had been placed (WHO, 1998). The Sudanese isolates were clearly distinct from the reference sequences of the known genotypes of clade B. Analysis of all the isolates of clade B for which N sequences were available allowed the Sudanese isolates to be placed with those from Nigeria within the genotype B3, cluster 1.

Until recently no standard criteria for the definition of clades and genotypes were available. However, two approaches were proposed by Kreis *et al.* (1997) and Hanses *et al.* (1999) based on the percentage divergence and common characteristic mutations between the isolates, respectively. The second approach provides more information about the common evolutionary background of viruses, but it requires the presence of larger numbers of isolates (Hanses *et al.*, 1999). The WHO now proposes a minimum nucleotide divergence of 2·5% for the partial N sequence and 2·0% for the complete H sequence from the most closely related strain as standard criteria for defining new genotypes (WHO, 2001). According to the above criteria, our isolates, together with the Nigerian B3, cluster 1 isolates, would form a new genotype, with a divergence of 3·4 % and a total of five set specific mutations in the COOH-N and a divergence of 2·0% in the H gene. Sequences of the proposed genotype B3 cluster 2 and genotype B1 are closely associated. They form, however, a heterogeneous group with a nucleotide divergence of up to 2·7% in the N gene, which may comprise more than one group or subgroup. More sequences from different African countries will thus be required for a better definition of clade B genotypes.

Globally, different MV clades show a certain degree of geographical restriction. The movement of infected people within and between geographical areas will determine virus epidemiology. The close similarity between our MV isolates and those isolated in Nigeria may therefore be explained by the historical and continuing links between Sudan and Northern Nigeria.

References

Bellini, W. J. & Rota, P. A. (1998). Genetic diversity of wild-type measles viruses: implications for global measles elimination programs. *Emerging Infectious Diseases* **4**, 29–35.

Chibo, D., Birch, C. J., Rota, P. A. & Catton, M. G. (2000). Molecular characterization of measles viruses isolated in Victoria, Australia, between 1973 and 1998. *Journal of General Virology* **81**, 2511–2518.

De Swart, R. L., El Mubarak, H. S., Vos, H. W., Mustafa, O. A., Abdallah, A., Groen, J., Mukhtar, M. M., Zijlstra, E. E., El Hassan, A. M., Wild, T. F., Ibrahim, S. A. & Osterhaus, A. D. M. E. (2001a). Prevention of measles in Sudan: a prospective study on vaccination, diagnosis and epidemiology. *Vaccine* **19**, 2254–2257.

De Swart, R. L., Nur, Y., Abdallah, A., Kruining, H., El Mubarak, H. S., Ibrahim, S. A., van den Hoogen, B., Groen, J. & Osterhaus, A. D. M. E. (2001b). Combination of reverse transcriptase PCR analysis and immunoglobulin M detection on filter paper blood samples allows diagnostic and epidemiological studies of measles. *Journal of Clinical Microbiology* **39**, 270–273.

El Mubarak, H. S., Van de Bildt, M. W. G., Mustafa, O. A., Vos, H. W., Mukhtar, M. M., Groen, J., El Hassan, A. M., Niesters, H. G. M., Ibrahim, S. A., Zijlstra, E. E., Wild, T. F., Osterhaus, A. D. M. E. & De Swart, R. L. (2000). Serological and virological characterization of clinically diagnosed cases of measles in suburban Khartoum. *Journal of Clinical Microbiology* **38**, 987–991.

Felsenstein, J. (1993). Phylip (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle, WA, USA.

Hanses, F., Truong, A. T., Ammerlaan, W., Ikusika, O., Adu, F., Oyefolu, A. O., Omilabu, S. A. & Muller, C. P. (1999). Molecular epidemiology of Nigerian and Ghanaian measles virus isolates reveals a genotype circulating widely in western and central Africa. *Journal of General Virology* **80**, 871–877.

Hanses, F., Van Binnendijk, R. S., Ammerlaan, W., Truong, A. T., De Rond, L., Schneider, F. & Muller, C. P. (2000). Genetic variability of measles viruses circulating in the Benelux. *Archives of Virology* **145**, 541–551.

Kreis, S., Vardas, E. & Whistler, T. (1997). Sequence analysis of the nucleocapsid gene of measles virus isolates from South Africa identifies a new genotype. *Journal of General Virology* **78**, 1581–1587.

Liffick, S. L., Thoung, N. T., Xu, W., Li, Y., Lien, H. P., Bellini, W. J. & Rota, P. A. (2001). Genetic characterization of contemporary wild-type measles viruses from Vietnam and the People's Republic of China: identification of two genotypes within clade H. *Virus Research* **77**, 81–87.

Mori, T., Sasaki, K., Hashimoto, H. & Makino, S. (1993). Molecular cloning and complete nucleotide sequence of genomic RNA of the AIK-C strain of attenuated measles virus. *Virus Genes* **7**, 67–81.

Rima, B. K., Earle, J. A. P., Yeo, R. P., Herlihy, L., Baczko, K., ter Meulen, V., Carabana, J., Caballero, M., Celma, M. L. & Fernandez-Munoz, R. (1995). Temporal and geographical distribution of measles virus genotypes. *Journal of General Virology* **76**, 1173–1180.

Rima, B. K., Earle, J. A. P., Baczko, K., ter Meulen, V., Liebert, U. G., Carstens, C., Carabana, J., Caballero, M., Celma, M. L. & Fernandez-Munoz, R. (1997). Sequence divergence of measles virus haemagglutinin during natural evolution and adaptation to cell culture. *Journal of General Virology* **78**, 97–106.

Rota, P. A., Bloom, A. E., Vanchiere, J. A. & Bellini, W. J. (1994). Evolution of the nucleoprotein and matrix genes of wild-type strains of measles virus isolated from recent epidemics. *Virology* **198**, 724–730.

Rota, J. S., Heath, J. L., Rota, P. A., King, G. E., Celma, M. L., Carabana, J., Fernandez-Munoz, R., Brown, D., Jin, L. & Bellini, W. J. (1996). Molecular epidemiology of measles virus: identification of pathways of transmission and implications for measles elimination. *Journal of Infectious Diseases* **173**, 32–37.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). Small scale preparations of plasmid DNA. In *Molecular Cloning: a Laboratory Manual*, 2nd edn, pp. 1.25–1.31. Edited by C. Nolan, N. Ford & M. Ferguson. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Schneider-Schaulies, S. & ter Meulen, V. (1999). Measles virus (Paramyxoviridae). In *Encyclopedia of Virology*, 2nd edn, pp. 952–960.

Edited by A. Granoff & R. G. Webster. New York & London: Academic Press.

Taylor, M. J., Godfrey, E., Baczko, K., ter Meulen, V., Wild, T. F. & Rima, B. K. (1991). Identification of several different lineages of measles virus. *Journal of General Virology* **72**, 83–88.

Truong, A. T., Kreis, S., Ammerlaan, W., Hartter, H. K., Adu, F., Omilabu, S. A., Oyefolu, A. O., Berbers, G. A. M. & Muller, C. P. (1999). Genotypic and antigenic characterization of African measles virus isolates. *Virus Research* **62**, 89–95.

UNICEF (2000). *The State of the World's Children* 2000. New York: UNICEF.

WHO (1998). Expanded programme on immunization (EPI). Standardization of the nomenclature for describing the genetic characteristics of wild-type measles viruses. *Weekly Epidemiological Record* **73**, 265–272.

WHO (2001). Nomenclature for describing the genetic characteristics of wild-type measles viruses (update). *Weekly Epidemiological Record* **76**, 242–251.

Xu, W., Tamin, A., Rota, J. S., Zhang, L., Bellini, W. J. & Rota, P. A. (1998). New genetic group of measles virus isolated in the People's Republic of China. *Virus Research* **54**, 147–156.