Journal of General Virology (1997), **78**, 1581–1587. Printed in Great Britain SHORT COMMUNICATION

Sequence analysis of the nucleocapsid gene of measles virus isolates from South Africa identifies a new genotype

Stephanie Kreis, Eftyhia Vardas and Toni Whistler

National Institute for Virology, Private Bag X4, Sandringham 2131, Johannesburg, South Africa

Sequence analysis was performed on 20 measles virus (MV) isolates from South Africa, five of which were obtained between 1986 and 1989 and 15 isolates collected during the 1994/95 measles season. A 590 bp fragment of the carboxyl terminus of the nucleocapsid (N) was amplified by PCR and subjected to sequence and phylogenetic analysis. Comparison of the South African MV strains with those previously described revealed that at least two distinct groups of wild-type (wt) MV exist, one of which has been circulating since 1986. The major genotype (I) was represented by the more recent isolates which showed three characteristic amino acid substitutions. Furthermore, three vaccine-like viruses with sequences very similar to the Edmonston wt strain were identified. Phylogenetic analysis of 100 MV strains allowed the assignment of new definitions for MV genotypes and subgroups. Employing these definitions, the majority of South African isolates analysed here formed a new genotype.

Measles virus (MV) is a negative-stranded RNA virus of the genus *Morbillivirus* in the family *Paramyxoviridae* and only one serotype has been described. Recent MV isolates have undergone significant genetic drift relative to the Edmonston (Edm) wild-type (wt) strain, which was isolated in 1954 (Rozenblatt *et al.*, 1985). Given the increased mutation rates over time in wt MVs (Rima *et al.*, 1995*b*; P. A. Rota *et al.*, 1994; Taylor *et al.*, 1991), the possibility of emerging escape mutants which might render the current vaccines ineffective cannot be excluded.

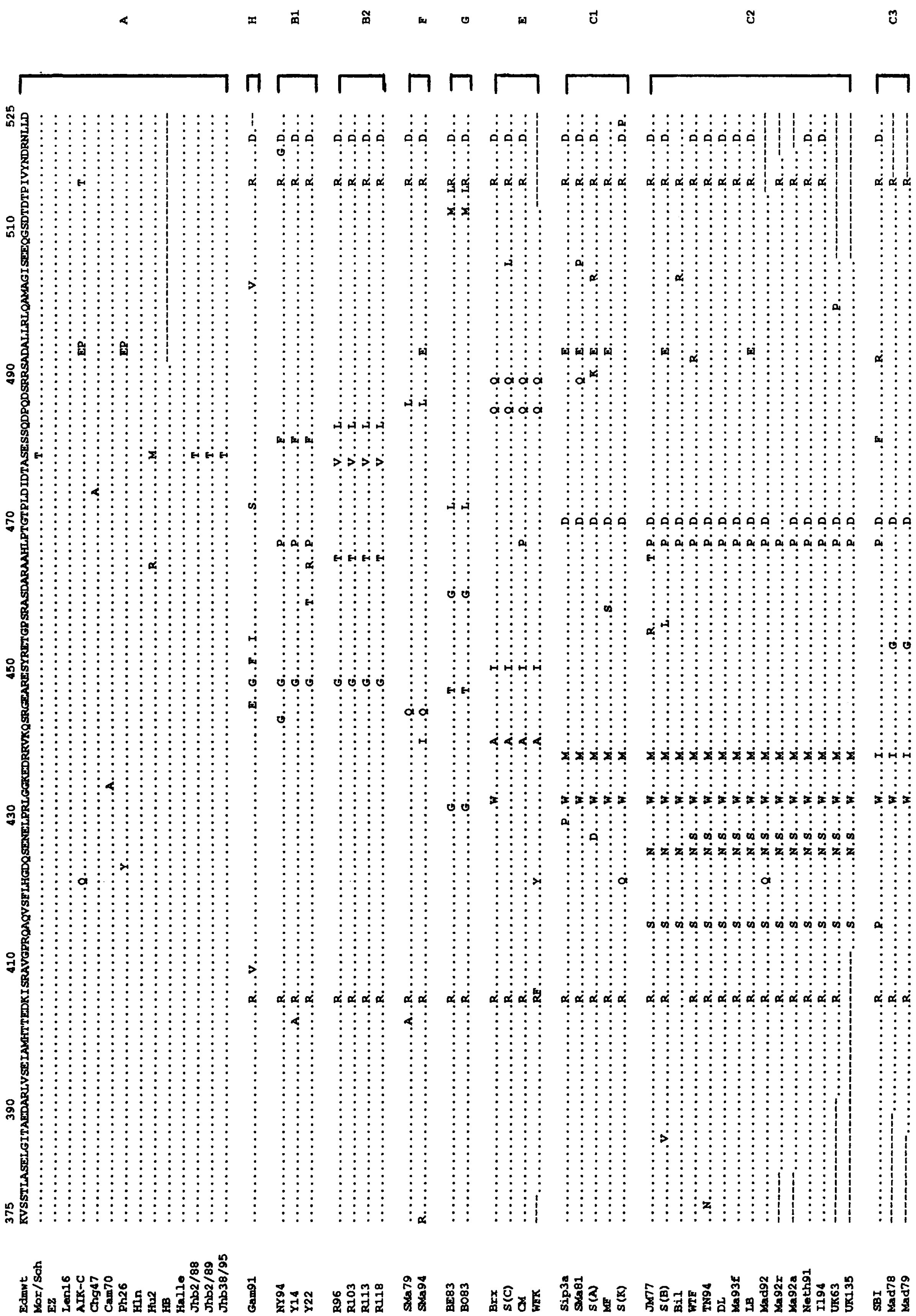
The majority of published sequences have been obtained from MV strains originating in Europe and the USA. Studies have focussed on the carboxyl-terminal 456 nt of the N gene, which are the most variable of the MV genome, and have proved suitable for strain characterization. Depending on the criteria used for allocation of MVs into separate genotypes, seven or eight different genotypes have been described to date (Rima *et al.*, 1995*b*). Several of these appear to be geographically restricted and still circulating, whereas others seem to be extinct (Rima *et al.*, 1995*a*; P. A. Rota *et al.*, 1994). However, the real geographical distribution and the question of how many lineages exist in the world can only be determined by including more sequences from Asia, Australia and Africa. Currently, the only sequences available from the African continent are derived from isolates made in Gabon (1983), Cameroon (1984) (Giraudon *et al.*, 1988; P. A. Rota *et al.*, 1994) and Gambia (1994) (Rota *et al.*, 1996); these show little relation to each other or to any previously described strains.

The endemic occurrence of MV in South Africa offers an opportunity to study a greater number of isolates and to monitor the evolution of the virus under circumstances of relatively low vaccine coverage, a condition which can still be found in some parts of the country (Department of Health, South Africa, 1995). To examine the genetic variability of MV isolates from South Africa, sequence analysis was performed on the carboxyl-terminal 456 nt of the N gene of 20 isolates. Of these, five were isolated in 1986, 1988 and 1989, originally passaged in Vero cells, and for this study adapted to B95a cells (Kobune *et al.*, 1990). Those isolates made in 1994/95 (Table 1) were cultivated in B95a cells. Total RNA was extracted from infected B95a cells using the acid guanidinium thiocyanate–phenol extraction technique (Chomczynski & Sacchi, 1987). Table 1 includes the 20 South African strains sequenced for this study and 12 isolates from various countries which have not previously been included in phylogenetic studies (sequences were retrieved from GenBank). Other sequences used in this phylogenetic analysis have been described previously (J. S. Rota *et al.*, 1994, 1996; P. A. Rota *et al.*, 1994; Rima *et al.*, 1995*b*). Predicted amino acid sequences were aligned (Fig. 1) and nucleotide sequence data were analysed with the ClustalW program (Thompson *et al.*, 1994).

To date, no general consensus has been reached as to how to define MV genotypes or subgroups. By analysing 100 MV

Author for correspondence: Stephanie Kreis.
Fax +27 11 882 0596. e-mail stephk@niv.ac.za

The GenBank accession numbers of the sequences reported in this paper are U64577–U64596.


375
390
410
430
450
470
490
510
525
KVSSTLASELGITAEDARLVSEIAMHTTEDKISRAVGPRQAQVSFLHGDQSENELPRLGGKEDRRVKQSRGEARESYRETGPSRASDARAAHLPTGTPLDIDTASESSQDPQDSRRSADALLRLQAMAGISEEQGSDTDTPIVYNDRNLLD
Edmwt
Mor/Sch
EZ
Len16
AIK-C
Chg47
Cam70
Ph26
Hln
Hu2
HB
Halle
Jhb2/88
Jhb2/89
Jhb38/95
Gam91
NY94
Y14
Y22
R96
R103
R113
R118
SMa79
SMa94
BE83
BO83
Brx
S(C)
CM
WTK
Sip3a
SMa81
S(A)
MF
S(K)
JM77
S(B)
Bil
WTF
TN94
DL
Ma93f
LB
Mad92
Ma92r
Ma92a
Neth91
I194
UK63
UK135
SBI
Mad78
Mad79
A
H
B1
B2
F
G
E
C1
C2
C3

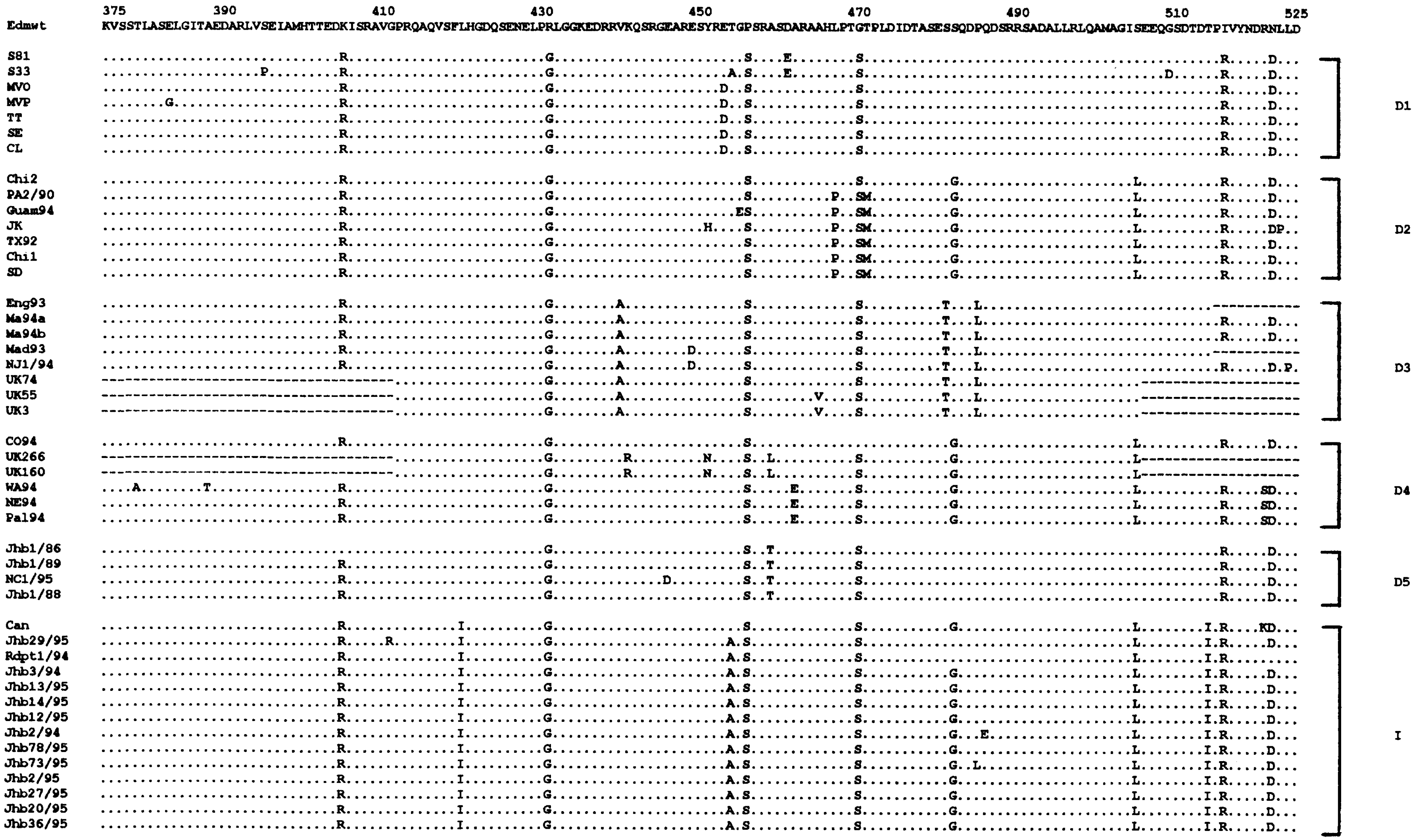


Fig. 1. Alignment of the predicted amino acid changes within the carboxyl-terminal 151 amino acids of 100 MV N proteins. Strains are as described in Table 1 or as published previously (J. S. Rota *et al.*, 1994, 1996; P. A. Rota *et al.*, 1994; Rima *et al.*, 1995). A dot indicates the same residue as in the Edm wt sequence; a dash was used where sequence data were not available. Genotype and subgroup designations are given on the right-hand side. Numbering of the coding sequence of the entire N gene is according to P. A. Rota *et al.* (1994).

Table 1. Origins of MV strains included in this study, not described previously

Strain[*]	Year	Location[‡]	Description	Accession no.	Group
Jhb2/88	1988	Johannesburg, RSA	Vaccine-like	U64583	A
Jhb2/89	1989	Johannesburg, RSA	Vaccine-like	U64584	A
Jhb38/95	1995	Johannesburg, RSA	Vaccine-like	U64585	A
Neth91	1991	Bilthoven, Netherlands	wt	L46748	C2
Mad92	1992	Madrid, Spain	wt	L46742	C2
UK63	1994	UK	wt	U29317	C2
UK135	1994	UK	wt	U29283	C2
Eng93	1993	England	wt	L46730	D3
Mad93	1993	Madrid, Spain	wt	L46744	D3
Ma94a	1994	Madrid, Spain	wt		D3
UK74	–[†]	UK	wt	U29327	D3
UK55	–	UK	wt	U29311	D3
UK3	1994	UK	wt	U29308	D3
UK266	1994	UK	wt	U29300	D4
UK160	1994	UK	wt	U29289	D4
Jhb1/86	1986	Johannesburg, RSA	wt	U64580	D5
Jhb1/88	1988	Johannesburg, RSA	wt	U64582	D5
Jhb1/89	1989	Johannesburg, RSA	wt	U64586	D5
NC1/95	1995	Newcastle, Natal, RSA	wt	U64581	D5
Jhb2/94	1994	Johannesburg, RSA	wt	U64587	I
Jhb3/94	1994	Johannesburg, RSA	wt	U64577	I
Rdpt1/94	1995	Roodepoort, RSA	wt	U64589	I
Jhb2/95	1995	Johannesburg, RSA	wt	U64590	I
Jhb12/95	1995	Johannesburg, RSA	wt	U64588	I
Jhb13/95	1995	Johannesburg, RSA	wt	U64578	I
Jhb14/95	1995	Johannesburg, RSA	wt	U64579	I
Jhb20/95	1995	Johannesburg, RSA	wt	U64591	I
Jhb27/95	1995	Johannesburg, RSA	wt	U64592	I
Jhb29/95	1995	Johannesburg, RSA	wt	U64593	I
Jhb36/95	1995	Johannesburg, RSA	wt	U64594	I
Jhb73/95	1995	Johannesburg, RSA	wt	U64595	I
Jhb78/95	1995	Johannesburg, RSA	wt	U64596	I

[*] Bold type denotes South African strains sequenced for this study.

[†] –, Date not recorded.

[‡] Both Johannesburg and Roodepoort are in Gauteng province.

strains, we have defined sequences as belonging to one genotype by having no more than 15 nucleotide differences within the carboxyl-terminal N region (⩾ 3·3 % nt divergence). Viruses within a genotype were denoted as a subgroup when their sequences showed fewer than 11 nucleotide changes (⩽ 2·2 %). These definitions proved to accommodate previously reported groupings (Rima *et al*., 1995 *b*) and allowed the allocation of new isolates to genotypes.

Comparing the position and number of amino acid substitutions (Fig. 1), three different groups of South African viruses were evident. Three isolates (Jhb2/88, Jhb2/89 and Jhb38/95) proved to be closely related to the Edm wt strain (genotype A), showing one amino acid substitution at position 478 of the N gene (S → T), which is also present in the Schwarz vaccine strain currently used in South Africa (Rimevax, SmithKline Beecham). Genotype A, represented by MV

Fig. 2. Phylogenetic analysis of the carboxyl-terminal 456 nt of 100 MV N genes. A 589 bp fragment was amplified by RT–PCR as described previously (J. S. Rota *et al*., 1994), using primers 5′ GCTATGCCATGGGAGTAGGAGTGG 3′ (nt position 1001 to 1024) and 5′ GGCCTCTCGCACCTAGTCTAG 3′ (nt position 1590 to 1570); positions correspond to the N gene of the Edm wt strain. Direct sequencing of PCR products in both directions was performed using the Sequenase PCR Product Sequencing Kit (United States Biochemical) according to the manufacturer's instructions. Sequences were obtained using the PCR primers and two additional primers: 5′ CTTGTTTCAGAGATTGCAATGCAT 3′ (nt position 1123 to 1146) and 5′ ACCTTCGACTGTCCTGCGGATCT 3′ (nt position 1449 to 1471). The phenogram is unrooted and was generated by the CLUSTALW program. South African isolates are given in italic letters. The scale indicates 1% nucleotide differences.

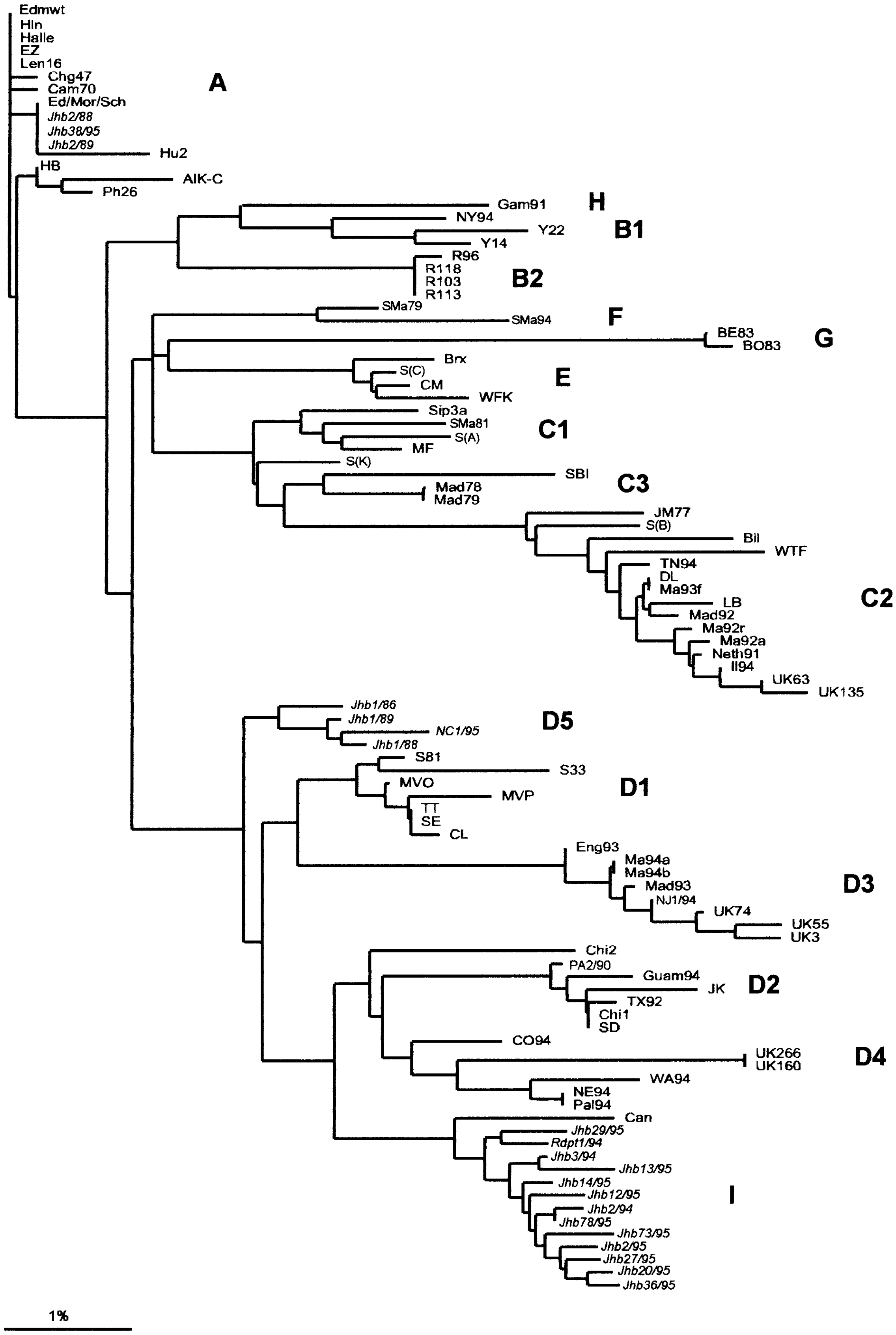


Fig. 2. For legend see facing page.

vaccine strains and isolates made before 1971, has been analysed extensively (J. S. Rota *et al*., 1994). The present study adds three strains of South African origin to this genotype. Although detailed patients' histories and information about possible contact to vaccinees was not available, these strains reflect cases of a wt measles infection caused by a vaccine-like virus. So far, clinical signs of infection following vaccination have been reported to be subclinical, milder and shorter than those of a natural infection (Osterhaus *et al*., 1994). Further characterization of these isolates including sequencing of the fusion and haemagglutinin genes is in progress. Emphasis will be placed on two amino acids, at positions 451 and 481 of the haemagglutinin gene, which have recently been shown to be important phenotypic markers for differentiation of vaccine and recent wt MV strains (Lecouturier *et al*., 1996).

A second group of South African wt MVs included four isolates, three of which were made between 1986 and 1989 in Johannesburg, and one recent isolate from Natal (Table 1). The amino acid alignment illustrates the resemblance of these viruses to genotype D viruses (Fig. 1). The majority of genotype D viruses and the 'earlier' South African wt viruses shared nine nucleotide changes of which six were coding. Evaluation of nucleotide divergence revealed that these four viruses form a new subgroup within genotype D, designated D5, differing on average by no more than 14 nucleotides (3·1%) from other genotype D viruses. The amino acid substitution at position 459 (A → T), characteristic for D5 viruses, was shared only by an isolate made in Cameroon (Y22) in 1983.

Recent MVs isolates from the greater Johannesburg area formed a new genotype, designated I. These isolates displayed more variation relative to the Schwarz vaccine strain than D5 virus strains, carrying on average 21 group-specific nucleotides (4·6% divergence), 20 of which were non-synonymous and 11 resulted in the replacement of an amino acid. These recent South African isolates carried three group-specific amino acids at positions 420 (L → I), 454 (T → A) and 514 (T → I), two of which were also common to a Canadian (Can) wt, isolated in 1989 (Fig. 1). Can proved to be closely related to the recent South African MVs, sharing 10 out of 11 amino acid substitutions, and was therefore moved from the D3 subgroup (Rima *et al*., 1995*b*) into genotype I. As indicated in Fig. 1, amino acid replacements at positions 431 (R → G), 456 (P → S) and 470 (G → S) were common to South African strains (I and D5) and other genotype D viruses from America and Europe. Two substitutions at positions 482 (S → G) and 505 (S → L) were seen in genotype I and subgroups D2 and D4, which represent mainly American strains isolated between 1989 and 1994. Regardless of the apparent resemblance between genotypes I and D, sequences differed by 15 to 25 nt, making the recent South African MVs a distinct genotype. A further 27 isolates made in 1995 in Johannesburg and Middleburg (Mpumalanga province) were shown to group within genotype I, as determined by the heteroduplex mobility assay (Kreis & Whistler, 1997). This suggests that genotype I is responsible for the majority of measles cases seen in these regions during the 1995 season.

Most of the recently isolated MVs from the USA, Europe and the strains from South Africa described here are grouping within genotypes D and I. These viruses share three characteristic amino acid substitutions (at positions 431, 456 and 470), which may indicate a common evolutionary relationship for these two genotypes. It appears that D and I represent the dominant genotypes currently circulating in many parts of the world. Furthermore, the subgroups of genotype D seem to show some geographical restriction (Fig. 2). D1 represents UK strains isolated between 1974 and 1991, D2 includes American isolates from a MV outbreak in 1989/90 and three recently described viruses from the USA, D3 consists of recent European strains, D4 of 1994 isolates from the USA and two recent UK isolates and finally, D5 represents four South African MVs, isolated between 1986 and 1995. The only recent isolate (NC1/95) within subgroup D5 was isolated from a patient hospitalized in Johannesburg who had recently travelled from Natal where he had probably become infected. It is possible that this lineage was more prevalent in the country in the late 1980s and that it may have been partly replaced by the group I viruses, which appear to be more abundant and show a higher mutation rate. Relative to the Schwarz vaccine strain, amino acid variation reached 7·3% in genotype I viruses, whereas subgroup D5 wt MVs had a lower rate of amino acid divergence, with only seven changes (4·6%). Thus, genotype I displays a greater number of mutations and may have evolved from the 'earlier' South African viruses in subgroup D5, signifying a genetic drift. However, the amino acid substitution at position 459 of the N gene was characteristic for D5 viruses only. Given that this mutation was not seen in any of the recent genotype I strains and assuming that a reversion of this particular mutation did not occur, it is more plausible to consider an independent evolution for both lineages.

Phylogenetic analysis was performed on 100 MV sequences and the resulting phenogram is shown in Fig. 2. Overall, the phenogram presented here confirms the previously described groupings (Rima *et al*., 1995*b*; Rota *et al*., 1996). However, the following changes occurred as a result of applying our criteria for genotype and subgroup definitions. The two Spanish isolates Mad78 and Mad79, formerly denoted as C1 viruses, now group together with the German isolate SBI, which previously formed a subgroup on its own (C3) (Figs 1 and 2). Although SBI represents a rather unique sequence, it meets the criteria by which a subgroup has been defined for this study, carrying fewer than 11 nucleotide changes relative to the two Spanish strains. The Gambian isolate (Gam91) forms a new genotype, H. This sequence showed 18 and 19 nucleotide differences from the Cameroon strains Y22 and Y14 (B1) respectively, and a minimum of 19 nucleotide changes from the Gabon isolates (B2). A recent study (Rota *et al*., 1996), which focussed on the full coding sequences of the N and haemagglutinin genes, grouped Gam91 together with the west African strains (B1 and B2) and a new isolate (NY94) into one genotype. However, considering the number of unique substitutions in Gam91, a distinct genotype for this isolate is suggested here. The NY94 strain, which was isolated in New York, grouped together with Cameroon isolates in subgroup B1. This is not surprising, as the virus was isolated from a child who most likely became infected in Kenya before travelling to New York (Rota *et al*., 1996). Some changes also affected genotype D, with the three UK strains CL, SE and TT now grouping in the D1 subgroup (formerly of D2). The D3 subgroup includes numerous recent isolates from Europe and one USA MV strain (NJ1/94). The subgroup D2 consists only of MVs isolated in the USA between 1989 and 1994. The majority of USA MV isolates made in 1994, and two recent strains from the UK, now form a new subgroup, D4, and as discussed here, D5 represents the four South African isolates.

In summary, sequence analyses of MV strains from various parts of the world have previously revealed seven or eight distinct genotypes. By expanding the number of sequences analysed here and by modifying the criteria for MV classification, we have described a new genotype (I), consisting of recent South African isolates, and a new subgroup (D5), representing three 'earlier' isolates and one recent isolate from South Africa. Furthermore, three vaccine-like isolates were shown to group within genotype A. Whether the two distinct wt lineages described here are the only ones circulating in South Africa at present will have to await further studies focussing on isolates from other regions within the country. Including Gam91, which forms the new genotype H, the number of genotypes is now nine, with a total of ten subgroups in genotypes B, C and D.

This work was supported by the Poliomyelitis Research Foundation. We thank Dr S. Patz and staff from the Alexandra Health Clinic, Dr M. Andre from Rietfontein Hospital, and all viral watch doctors who took part in the collection of MV specimens. We are also grateful to Dr Lynn Morris and Claudia Chezzi for reading of this manuscript.

References

Chomczynski, P. & Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction. *Analytical Biochemistry* **162**, 156–159.

Department of Health South Africa (1995). *Epidemiological Comments* **22**(9), 190–202.

Giraudon, P., Jacquier, M. F. & Wild, T. F. (1988). Antigenic analysis of African measles field isolates: identification and localization of one conserved and two variable epitope sites on the NP protein. *Virus Research* **10**, 137–153.

Kobune, F., Sakata, H. & Sugiura, A. (1990). Marmoset lymphoblastoid cells as a sensitive host for isolation of measles virus. *Journal of Virology* **64**, 700–705.

Kreis, S. & Whistler, T. (1997). Rapid identification of measles virus strains by the heteroduplex mobility assay. *Virus Research* (in press).

Lecouturier, V., Fayolle, J., Caballero, M., Carabaña, J., Celma, M. L., Fernandez-Muñoz, R., Wild, T. F. & Buckland, R. (1996). Identification of two amino acids in the hemagglutinin glycoprotein of measles virus (MV) that govern hemadsorption, HeLa cell fusion, and CD46 down-regulation: phenotypic markers that differentiate vaccine and wild-type MV strains. *Journal of Virology* **70**, 4200–4204.

Osterhaus, A. D. M. E., De Vries, P. & Van Binnendijk, R. S. (1994). Measles vaccines: novel generations and new strategies. *Journal of Infectious Diseases* **170**(Suppl. 1), S42–55.

Rima, B. K., Earle, J. A. P., Baczko, K., Rota, P. A. & Bellini, W. J. (1995*a*). Measles virus strain variation, pp. 65–83. In *Measles Virus*. Edited by V. ter Meulen & M. A. Billeter. Berlin: Springer Verlag.

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

Rota, J. S., Wang, Z.-D., Rota, P. A. & Bellini, W. J. (1994). Comparison of sequences of the H, F, and N coding genes of measles virus vaccine strains. *Virus Research* **31**, 317–330.

Rota, J. S., Heath, J. L., Rota, P. A., King, G. E., Celma, M. L., Carabaña, J., Fernandez-Muñoz, 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.

Rota, P. A., Bloom, A. E., Vanchieri, 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.

Rozenblatt, S., Eizenberg, O., Ben-Levy, R., Lavie, V. & Bellini, W. J. (1985). Sequence homology within the morbilliviruses. *Journal of Virology* **53**, 684–690.

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.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673–4680.

Received 2 October 1996; Accepted 26 February 1997