Journal of General Virology (2002), **83**, 1419–1430. Printed in Great Britain

Increased positive selection pressure in persistent (SSPE) versus acute measles virus infections

Christopher H. Woelk,[1] Oliver G. Pybus,[1] Li Jin,[2] David W. G. Brown[2] and Edward C. Holmes[1]

[1] Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

[2] Enteric, Respiratory and Neurological Virus Laboratory, Central Public Health Laboratory, London NW9 5HT, UK

We compared the extent of positive selection acting on acute and persistent strains of measles virus (MV). Far stronger positive selection was found in the fusion (F) and haemagglutinin (H) genes from subacute sclerosing panencephalitis (SSPE) compared to acute MV cases. Most of the positively selected sites identified in these surface glycoprotein genes from SSPE cases correspond to structural, functional or antigenic areas, and could not be explained by the effects of cell passaging. The correlations between selected sites and functional studies of MV are discussed in detail with reference to the maintenance of persistent infection. No positive selection was found in the matrix (M) gene from acute cases of MV and the effects of including hypermutated SSPE M gene sequences in phylogenetic inference were also explored. Finally, using H gene data, we estimated the rate of molecular evolution for SSPE strains as $3{\cdot}4 \times 10^{-4}$ substitutions/site/year, which is similar to previous estimates obtained for acute strains.

Introduction

Measles virus (MV) belongs to the morbillivirus subgroup of the paramyxovirus family of single-stranded negative-sense RNA viruses. Acute MV infection is a regular occurrence throughout the world and rare complications can lead to subacute sclerosing panencephalitis (SSPE) and measles inclusion body encephalitis (MIBE). Both are lethal syndromes affecting the central nervous system (CNS) through cerebral degradation and consequent deterioration of sensory and motor function. SSPE occurs at low frequency (4 per 100000 cases in England and Wales) in adolescents and young adults with normal immune functions 7–10 years after initial MV infection (Miller *et al.*, 1992; ter Meulen *et al.*, 1983). MIBE occurs more often, but only in immunologically compromised children (e.g. leukaemia patients treated with cytostatic drugs) after a shorter incubation period measured in months (Ohuchi *et al.*, 1987; Roos *et al.*, 1981).

The fusion (F) and haemagglutinin (H) glycoproteins of MV are type I and type II integral membrane proteins, respectively, and reside on the viral envelope. The F protein is primarily synthesized as an inactive precursor (F_0), which is cleaved by host cell proteolytic enzymes to form the biologically active disulphide-linked F_1 and F_2 subunits (Scheid & Choppin, 1974).

Author for correspondence: Christopher Woelk.
Fax +44 1865 310447. e-mail Christopher.Woelk@zoo.ox.ac.uk

N-Glycosylation of the F_2 subunit at Asn residues 29, 61 and 67 produces the fusion active conformation of the F protein (Alkhatib *et al.*, 1994; Hu *et al.*, 1995). The mature form of the H protein also results from disulphide bridge formation and *N*-glycosylation of Asn residues 168, 187, 200 and 215 (Hu *et al.*, 1994; Langedijk *et al.*, 1997). The fusion process is thought to result from aggregation of the envelope proteins such that H contacts the receptor of a neighbouring cell and F controls the actual membrane fusion (Billeter *et al.*, 1994; Lamb, 1993). The CD46 and SLAM host cell receptors have been identified as the targets for H protein interaction (Manchester *et al.*, 2000; Tatsuo *et al.*, 2000). The matrix (M) protein is not an integral membrane protein but it does associate with membranes and is found in dense concentrations underlying the lipid bilayer after host cell infection (Lamb & Kolakofsky, 1996). The M protein is thought to play a pivotal role in the formation of budding viral particles because it interacts with the cytoplasmic tails of H and F, the lipid bilayer and the nucleocapsids (N) (Lamb & Kolakofsky, 1996).

The development of SSPE is thought to be determined by a combination of host cell factors, immune system activity and mutations in the MV genome. For example, mutations in the M gene of SSPE cases result in sites of cryptic initiation and premature termination, as well as adenosine to inosine (A/I) hypermutations, which lead to instability and limited interaction with the N protein. Truncations in the cytoplasmic domain of the F protein impede efficient virus assembly and

budding (Cattaneo & Rose, 1993), and loss of glycosylation in the H protein leads to inefficient membrane transport (Schneider-Schaulies *et al.*, 1995). These features often lead to reduced amounts of envelope protein on the cell surface as well as dysfunctional M protein, which maintain persistence by hampering the assembly and budding of mature infectious MV particles (Liebert, 1997).

The body overcomes acute MV infection through activation of T-cell immunity and protection against reinfection results from induction of humoral responses. High titres of antibodies against the F, H, N and P proteins are found in the serum and cerebrospinal fluid (CSF) of SSPE patients (Liebert, 1997; Schneider-Schaulies *et al.*, 1995). Therefore, it appears that virus spread through the CNS in SSPE cases occurs in the presence of a considerable immune response but without detectable virion formation or virus fusion (Billeter *et al.*, 1994). It is thought that, despite downregulation of the envelope glycoproteins, the fusion function is still maintained in the CNS, possibly by axonal transport through synaptic clefts. Establishment of virus persistence probably results from interactions between MV and cells of the CNS, whereas sequence alterations are thought to lead to the maintenance of persistent infections over long incubation periods (Schneider-Schaulies & Liebert, 1991).

This study aims to investigate the differences between acute and persistent MV infection in terms of sequence evolution and specifically to determine if any mutations (beyond those currently identified) that occur in viral genes facilitate persistence either through aberrant protein function or avoidance of the immune system. Sites under positive selection (adaptive evolution) in genes encoding envelope proteins of SSPE patients are good candidates for such mutations because they may represent substitutions in epitopes that abolish immune recognition, or changes in structural domains that result in abnormal proteins. Sites under putative positive selection can be identified by calculating the number of synonymous (d_S) and nonsynonymous (d_N) substitutions per site, whose ratio (d_N/d_S) is termed ω. An $\omega < 1$ suggests purifying (negative) selection, whereas an $\omega > 1$ is indicative of positive selection because the rate of fixation is higher than the background rate of mutation, which defies neutrality ($\omega = 1$) (Yang & Bielawski, 2000). A maximum likelihood (ML) method using models of sequence evolution can be employed to calculate ω ratios and to identify amino acid sites as conserved, neutral or positively selected (Yang *et al.*, 2000). This method accounts for phylogenetic structure, biases in codon usage and the transition/transversion (T_S/T_V) ratio, and has already been applied to the H and L genes of MV (Woelk *et al.*, 2001). This previous research investigated the difference in sites under positive selection between acute and persistent infections but lack of sequence data led to inconclusive results. In this work, new SSPE sequence data for the H gene have been obtained and F gene sequences from both acute and persistent patients have been analysed. In addition, phylogenetic methods are used to estimate the rate of evolution and the extent of hypermutation in genes from SSPE cases.

Methods

■ **Virus strains.** H gene sequences from SSPE isolates [UK86/UK/74(60s), UK87/UK/74(69) and UK111/UK/98(80s)] were determined by extracting nucleic acid from frozen brain specimens using the silica–guanidinium thiocyanate method (Boom *et al.*, 1990) followed by RT–PCR as described previously (Jin *et al.*, 1996, 1998). The GenBank accession numbers of these newly sequenced isolates are: AY077712–AY077714. Michiko Watanabe kindly provided sequence data for the F and H genes of the Biken SSPE isolate.

■ **Multiple alignments and phylogenetic analysis.** Complete sequences of H, F and M genes were downloaded from GenBank and aligned after the removal of identical, vaccine and highly passaged strains. The M gene alignment of 35 sequences includes five SSPE variants of isolate S(B) (Baczko *et al.*, 1993). Within this alignment, sequences were divided into Acute-M ($n = 20$) data sets and SSPE-M [$n = 10$; S(B) variants and the S(C) MIBE isolate removed]. The F gene alignment (Total-F) contains 22 sequences and was further divided into Acute-F ($n = 12$) and SSPE-F [$n = 9$; S(C) MIBE isolate removed]. Amino acids after site 523 were stripped from the Total-F and SSPE-F alignment due to the presence of frameshifts and premature stop codons. Three H gene alignments of SSPE isolates were created; SSPE-H ($n = 20$), SSPE-H(NP) ($n = 12$) and SSPE-H(CP) ($n = 7$). The latter two data sets contain SSPE isolates that were non-passaged (NP) and cell passaged (CP), respectively. For comparison to SSPE, we incorporated an alignment of 37 H genes from acute strains of MV (Acute-H) that was not analysed separately in our previous paper (Woelk *et al.*, 2001). A full list of the sequences used in this study is presented in Table 1.

The PAUP* package was used to construct maximum likelihood (ML) trees under the HKY85 + Γ model of nucleotide substitution (Swofford, 2000). An ML tree for the Total-M data set is presented in Fig. 1 and estimates of the T_S/T_V ratio are included in Table 2. An ML tree representing the Acute-H data set was presented in Woelk *et al.* (2001).

■ **Evolutionary rate analysis.** An alignment of 13 H gene sequences from SSPE cases of known isolation which had not undergone extensive cell passaging was compiled so that a phylogeny could be obtained using the PAUP* method described earlier. This phylogeny was analysed using the TipDate program (Rambaut, 2000), which estimates the rate of molecular evolution (i.e. nucleotide substitution) for sequences that have been isolated at different dates. An ML estimate of the rate can be obtained using the SRDT (single rate, dated tip) model, which assumes a constant rate of substitution (i.e. molecular clock) but accommodates differences in the dates of isolation. The significance of the rate obtained can be evaluated by comparing the likelihood estimate from the SRDT model with that obtained when rates are allowed to vary along each branch (different rate, DR model) with a likelihood ratio test (LRT).

■ **Selection analyses.** Yang *et al.* (2000) have developed an ML method for determining selection pressures in protein coding DNA sequences. This method originally employed 14 models that use statistical distributions to account for variable ω (d_N/d_S) ratios among codon sites. Subsequent research has shown that models M0, M1, M2, M3, M7 and M8 are sufficient for accurate selection analysis (Woelk & Holmes, 2001; Woelk *et al.*, 2001). Models M0, M1 and M7 do not allow for the existence of positively selected sites. M0 calculates a single ω ratio (between the bounds 0 and 1) averaged over all sites, whereas M7 uses a discrete beta distribution (between the same bounds) to model different ω ratios among sites. The shape of the beta distribution is governed by

the parameters p and q. M1 accounts for neutral evolution by estimating the proportion of conserved ($\omega = 0$) and neutral ($\omega = 1$) sites. Alternatively, models M2, M3 and M8 account for positive selection using parameters that estimate $\omega > 1$. Models M2 and M8 extend M1 and M7, respectively, through the addition of two parameters (p_2 and ω_2 for M2 and p_1 and ω_1 for M8) that have the potential to estimate $\omega > 1$ for an extra class of sites. M3 provides the most sensitive test for positive selection by estimating an ω ratio for a predetermined number of classes. Three classes were used in this analysis (p_0, p_1 and p_2) such that three corresponding ω ratios (ω_0, ω_1 and ω_2) were estimated. The first step in the identification of amino acid sites under positive selection is to test whether sites exist with $\omega > 1$ by comparing nested models using likelihood ratio tests (LRTs). M0 and M1 are both special cases of M2 and M3, while M7 is a special case of M8, and such nested models can be compared with LRTs. Once positively selected sites have been shown to exist, the second step uses Bayesian methods to locate their position. Sites having high posterior probabilities ($> 90\%$) of belonging to a site class with $\omega > 1$ are good candidates for positively selected sites. Posterior probabilities are conditional on the observed data such that they refer to the probability that a site, given the data at that site, is from a particular site class. The methods and models described here were implemented using the CODEML program of the PAML package, version 3.0c (Yang, 1997).

Results

Analysis of evolutionary rate

The 13 H genes from SSPE cases analysed using TipDate were temporally widely dispersed with isolation dates from 1965 to 1998. A mean rate of molecular evolution of $3{\cdot}4 \times 10^{-4}$ substitutions/site/year (95% confidence intervals: $2{\cdot}1 \times 10^{-4}$–$4{\cdot}8 \times 10^{-4}$) was estimated for this data set, which fits a molecular clock since the model where a single substitution rate was applied (SRDT) was not rejected by the one where each branch had a different substitution rate (DR model).

Hypermutation investigation and selection analysis of the M gene

To determine the extent of hypermutation in genes of SSPE isolates, the T_S/T_V ratio (Table 2) and substitution rate parameters for all possible transitions and transversions (not shown) were estimated for the acute and SSPE alignments of the F, H and M genes. The T_S/T_V ratio was comparable between acute and SSPE alignments for the F and H genes but was approximately double in SSPE-M when compared to Acute-M. Further analysis reveals that the increased ratio in SSPE-M isolates predominantly results from U to C changes. Selection analysis of isolates of the M protein from acute patients (results not shown) indicated that this protein is strongly conserved, which is not surprising considering its critical functional role during lytic infection. Selection analysis of the M gene from SSPE cases was not performed because sequence data contained too many frameshifts and premature stop codons in the N terminus.

Selection analysis of the F gene

Selection analysis of the Acute-F data set did not identify any positively selected sites. M0 estimated a similar likelihood compared to models able to account for positive selection (M2, M3 and M8) and suggested that sites are strongly conserved ($\omega = 0{\cdot}098$) (results not shown). In contrast, positive selection was identified in the SSPE-F data set by M2, M3 and M8, which rejected models M0, M1 and M7 in likelihood ratio tests (LRTs) (Table 3). M2 and M3 produced similar results and estimated a large proportion of conserved sites and a small proportion of positively selected sites ($\omega > 5$). Using Bayesian methods, M2 only assigned site 62 to the positively selected class with a posterior probability of 0·9356. M3 and M8 both assigned sites 62, 167, 446 and 461 to the positively selected class with posterior probabilities $> 99\%$ for M3 and $> 95\%$ for M8. Although M3 could not reject M2 in an LRT, sites 167, 446 and 461 should still be considered under possible positive selection because they were also identified by M8. M3 further identified sites 462 and 465 as under potential positive selection but this was with posterior probabilities of $\approx 90\%$ and, since they were not confirmed by M8, these two sites are not discussed further.

Selection analysis of the H gene

Both M3 and M8 estimated a positively selected class of sites in the H gene from acute strains of MV (Acute-H). M2 was unable to estimate a positively selected class because its extra parameters were used to account for a large class (68·1%) of fairly conserved sites ($\omega_2 = 0{\cdot}159$). M3 was not restricted in this way and predicted that 1·7% of sites are under weak positive selection ($\omega_2 = 2{\cdot}222$). M3 was able to reject both M0 and M1 in LRTs but was unable to reject M2. Fortunately M8, which also predicted a small class of sites with a similar positive selection pressure ($\omega_1 = 2{\cdot}114$), was able to reject M7 and confirm the significance of positive selection in Acute-H. Using Bayesian methods, M3 and M8 both predicted that sites 481, 546, 562 and 575 belong to the positively selected class with posterior probabilities $> 90\%$.

Stronger evidence for positive selection was found in the SSPE-H alignment. M2 and M8 estimated similar parameters for positive selection such that a small proportion of sites is under a very strong force of positive selection ($\omega = 10{\cdot}704$ and 9·350, respectively) (Table 3). The M3 model suggested that two classes of positive selection with different ω ratios exist but the significance of this could not be confirmed because M3 was unable to reject M2. Bayesian methods assigned sites 12, 62 and 618 with posterior probabilities $> 99\%$ to the positively selected class estimated by M2 and M8.

When SSPE strains detected directly from brain material [SSPE-H(NP)] were analysed, M2 and M3 were able to reject M0 but not M1, suggesting that a neutral class of sites fits the

Table 1. The 22 F gene, 35 M gene and 20 SSPE H gene sequences used for analysis

Accession numbers of H gene sequences shown in bold were used to estimate the evolutionary rate of SSPE. Sequence names are in the format: traditional name (when applicable)/country of isolation/year of isolation/isolate number (when applicable). Genotype designations are from the World Health Organization (2001). IU, information unavailable; HK, human kidney cell; HEK, human embryonic kidney cell; CV-1 and Vero, African green monkey kidney cells. For SSPE strains, the first date refers to the date of isolation and the date in brackets refers to the original MV infection.

Accession no.	Isolate name	Genotype	Passaging history	Reference
F gene				
Acute strains				
AB003178	Kar.PAK/89/1	D4	IU	Nishimura (direct submission)
AB003181	Kar.PAK/89/2	D4	IU	Nishimura (direct submission)
AB012948	9301B/Japan/⩽89	D3	B95a	Takeda *et al.* (1998)
AB016162	Ichinose-B95a/Japan/84	D3	B95a	Takeuchi *et al.* (2000)
AF179430	Masusako/Japan/83	C1	IU	Ning *et al.* (direct submission)
AF179432	Toyoshima/Japan/59	A	IU	Ning *et al.* (direct submission)
AJ133108	WTF/Germany/90	C2	BJAB	Johnston *et al.* (1999)
D63926	Nagahata/Japan/71	C1	HEK & Vero	Watanabe (direct submission)
K01711	Edm/US/54	A	IU	Billeter *et al.* (1984)
M81901	SD/US/89	D3	Vero	Rota *et al.* (1992)
M81903	Chi1/US/89	D3	Vero	Rota *et al.* (1992)
Y17840	Lys-1/France/96	IU	B95a	Fayolle *et al.* (1999)
SSPE strains				
AF179433	Osaka-1/Japan/93(68)	C1	HEL	Ning *et al.* (direct submission)
AF179436	Osaka-2/Japan/94(84)	C1	B95a	Ning *et al.* (direct submission)
AF179439	Osaka-3/Japan/95(71)	C1	B95a	Ning *et al.* (direct submission)
D00090*	Hallé/US/⩽71(⩽71)	A	Vero	Buckland *et al.* (1987)
D10548	Yamagata-1/Japan/82(71)‡	C1	Vero	Komase *et al.* (1990b)
L46871	Mantooth/IU/70s(IU)	A	IU	Palmer *et al.* (direct submission)
X16567	S(A)/Germany/80s(IU)	C1	None	Cattaneo *et al.* (1988)
X16568	S(B1)/Austria/80s(IU)	C2	None	Cattaneo *et al.* (1988)
Unpublished	Biken/Japan/75(68)	C1	HEL	Watanabe (personal communication)
MIBE strain				
X16569	S(C)/US/78	E	None	Cattaneo *et al.* (1988)
M gene				
Acute strains				
AB002682	Toyoshima/Japan/59	A	Vero	Ayata *et al.* (1998)
AB002683	Masusako/Japan/83	C1	Vero	Ayata *et al.* (1998)
AB002684	Iijima/Japan/84	C1	Vero	Ayata *et al.* (1998)
AB002685	Sobahani/Japan/85	C1	Vero	Ayata *et al.* (1998)
AB002686	Yokota/Japan/87	D3	Vero	Ayata *et al.* (1998)
AB002687	Nagahama/Japan/87	D3	Vero	Ayata *et al.* (1998)
AB002688	Taguchi/Japan/91	D5	Vero	Ayata *et al.* (1998)
AB012948	9301B/Japan/⩽89	D5	B95a	Takeda *et al.* (1998)
AB016162	Ichinose-B95a/Japan/84	D3	B95a	Takeuchi *et al.* (2000)
D00338	Hu2/N.Ireland/71	A	Vero	Curran & Rima (1988)
K01711	Edm/US/54	A	IU	Billeter *et al.* (1984)
M72147	Nagahata/Japan/71	C1	HEK, HEL, Vero or CV-1	Wong *et al.* (1991)
U01975	Berkley/US/83	G1	Vero	Rota *et al.* (1994)
U01979	Chi2/US/88	D3	Vero	Rota *et al.* (1994)
U01980	Chi1/US/89	D3	Vero	Rota *et al.* (1994)
U01982	HAL/Finland/62	A	IU, Vero	Rota *et al.* (1994)
U01983	JM/US/77	C2	Vero	Rota *et al.* (1994)
U01985	Bo83/US/83	D7	IU, Vero	Rota *et al.* (1994)
U01986	PH26/US/70	A	IU, Vero	Rota *et al.* (1994)
U01997	SD/US/89	D3	Vero	Rota *et al.* (1994)

Table 1 (*cont.*)

Accession no.	Isolate name	Genotype	Passaging history	Reference
SSPE strains				
AB002689	Osaka-1/Japan/93(68)	C1	Vero	Ayata *et al.* (1998)
AB002690	Osaka-2/Japan/94(84)	C1	Vero	Ayata *et al.* (1998)
AB002691	Osaka-3/Japan/95(71)	C1	Vero	Ayata *et al.* (1998)
D00493	Niigata-1/Japan/IU(IU)	C1	Vero or HeLa,	Enami *et al.* (1989)
D00494	ZH/Japan/IU(IU)	C1	Vero	Enami *et al.* (1989)
D00495	Biken/Japan/75(68)	C1	HEL & Vero	Enami *et al.* (1989)
M30810	Yamagata-1/Japan/82(71)‡	C1	Vero	Wong *et al.* (1989)
X16567	S(A)/Germany/80s(IU)	C1	None	Cattaneo *et al.* (1988)
X16568	S(B1)/Austria/80s(IU)	C2	None	Cattaneo *et al.* (1988)
X54068*	Hallé/US/⩽71(⩽71)	A	HeLa & Vero	Buckland *et al.* (1990)
Z25393†	S(B2)/Austria/80s(IU)	C2	None	Baczko *et al.* (1993)
Z25394†	S(B3)/Austria/80s(IU)	C2	None	Baczko *et al.* (1993)
Z25395†	S(B4)/Austria/80s(IU)	C2	None	Baczko *et al.* (1993)
Z25396†	S(B5)/Austria/80s(IU)	C2	None	Baczko *et al.* (1993)
MIBE strain				
X16569	S(C)/US/78	E	None	Cattaneo *et al.* (1988)
SSPE H gene				
AB045302	Osaka-1/Japan/93(68)	C1	Vero	Furukawa *et al.* (2001)
AB045303	Osaka-2/Japan/94(84)	C1	Vero	Furukawa *et al.* (2001)
AB045306	Osaka-3/Japan/95(71)	C1	Vero	Furukawa *et al.* (2001)
AF399848	UK44/UK/97(80s)	D7	None	Woelk *et al.* (2001)
AF399849	UK83/UK/65(56)	A	None	Woelk *et al.* (2001)
AF399850	UK85/UK/65(56)	C1	None	Woelk *et al.* (2001)
AF399851	UK88/UK/65(55)	C1	None	Woelk *et al.* (2001)
AY077712	UK86/UK/74(60s)	D1	None	Present study
AY077713	UK87/UK/74(69)	D1	None	Present study
AY077714	UK111/UK/98(80s)	B	None	Present study
D10549	Yamagata-1/Japan/82(71)‡	C1	Vero	Komase *et al.* (1990a)
L46870	Mantooth/IU/70s(IU)	A	IU	Palmer *et al.* (direct submission)
X04720*	Hallé/US/⩽71(⩽71) SSPE	A	Vero	Gerald *et al.* (1986)
X16567	S(A)/Germany/80s(IU)	C1	None	Cattaneo *et al.* (1989)
X16568	S(B1)/Austria/80s(IU)	C2	None	Cattaneo *et al.* (1989)
X68043	LEC-WI/US/70(IU)	E	Vero	Hu *et al.* (1993)
Z80810	Sma81/Spain/81(IU)	C1	None	Rima *et al.* (1997)
Z80828	Sma79/Spain/79(IU)	F	None	Rima *et al.* (1997)
Z80830	Sma94/Spain/94(IU)	F	None	Rima *et al.* (1997)
Unpublished	Biken/Japan/75(68)	C1	CV-1 and HEL	Watanabe (personal communication)

* The Hallé strain of SSPE was previously suggested to be an Edmonston contaminant (Cattaneo *et al.*, 1989) but was included for analysis in this paper.
† Hypermutated variants of S(B1).
‡ Original infection of MV for the Yamagata-1/Japan/82(71) SSPE strain was listed as 1980 in Woelk *et al.* (2001). This was a typographical error and should read 1971.

data just as well as a class of positively selected of sites (Table 3). Interestingly, the M7 vs M8 comparison contradicted this, since M8 was able to reject M7. This lends significance to positive selection ($\omega_1 = 3{\cdot}331$) at sites 7, 8, 12, 62, 302, 423, 560 and 575 estimated by M8 with posterior probabilities $> 90\%$. Although the alignment of cell-passaged SSPE strains [SSPE-H(CP)] contained fewer sequences than SSPE-H(NP), models accounting for positive selection were consistently strongly supported (Table 3). Similar to the results obtained for SSPE-H, M3 estimated two classes of positively selected sites, but since it could not reject M2 the existence of both classes is unconfirmed. M2 and M8 both estimated a very strong positive selection pressure for site 62 ($\omega_2 = 15{\cdot}314$ and $\omega_1 = 14{\cdot}211$, respectively), which was the only site that could be assigned to the positively selected class with a posterior probability $> 90\%$ (Table 3).

Table 2. Maximum likelihood parameters for the phylogenetic trees of the F, H and M gene data sets

T_s/T_v, transition/transversion ratio.

Data set	Gene	Taxa	Codons	T_s/T_v
Total-F	F	22	523	5·104
Acute-F	F	12	550	4·422
SSPE-F	F	9	523	5·683
Acute-H	H	37	617	5·125
SSPE-H	H	20	617	5·210
SSPE-H(NP)	H	12	617	5·730
SSPE-H(CP)	H	7	617	4·753
Total-M	M	35	335	16·011
Acute-M	M	20	335	6·261
SSPE-M	M	10	335	11·559

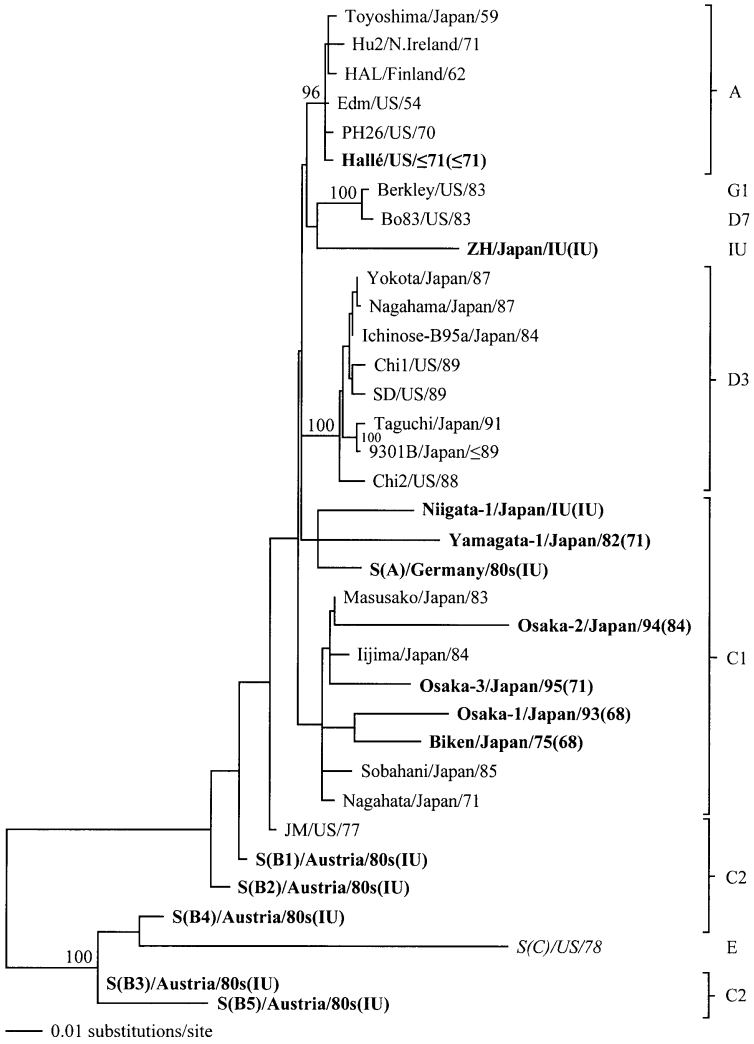

Fig. 1. Maximum likelihood (ML) tree of the M gene alignment including the five variants of SSPE strain S(B). The phylogeny was midpoint rooted for clarity and bootstrap values > 75% are indicated at the nodes (1000 replicates). Sequences in bold were sampled from SSPE cases and sequence S(C)/US/78 was from a patient with MIBE. Sequence names are in the format: traditional name/country of isolation/year of isolation. For the SSPE cases, the date in parentheses refers to the instance of initial MV infection (IU, information unknown).

Table 3. Parameter estimates and likelihood ratio tests (LRTs) for selection analysis of the F and H gene data sets

Data set*/model	Parameter estimates	LRT†	*p*-value
Acute-H			
M0	$\omega = 0{\cdot}208$	M0 vs M2	**< 0·000**
M1	$p_0 = 0{\cdot}667$, $p_1 = 0{\cdot}333$	M1 vs M2	**< 0·000**
M2	$p_0 = 0{\cdot}224$, $p_1 = 0{\cdot}095$, $p_2 = 0{\cdot}681$	M0 vs M3	**< 0·000**
	$\omega_2 = 0{\cdot}159$		
M3	$p_0 = 0{\cdot}751$, $p_1 = 0{\cdot}232$, $p_2 = 0{\cdot}017$	M1 vs M3	**< 0·000**
	$\omega_0 = 0{\cdot}076$, $\omega_1 = 0{\cdot}505$, $\omega_2 = 2{\cdot}222$	M2 vs M3	0·154
M7	$p = 0{\cdot}316$, $q = 1{\cdot}194$		
M8	$p = 0{\cdot}604$, $q = 2{\cdot}803$	M7 vs M8	**0·003**
	$p_1 = 0{\cdot}019$, $\omega_1 = 2{\cdot}114$		
SSPE-H			
M0	$\omega = 0{\cdot}482$	M0 vs M2	**< 0·000**
M1	$p_0 = 0{\cdot}586$, $p_1 = 0{\cdot}414$	M1 vs M2	**< 0·000**
M2	$p_0 = 0{\cdot}567$, $p_1 = 0{\cdot}421$, $p_2 = 0{\cdot}012$	M0 vs M3	**< 0·000**
	$\omega_2 = 10{\cdot}705$		
M3	$p_0 = 0{\cdot}895$, $p_1 = 0{\cdot}100$, $p_2 = 0{\cdot}005$	M1 vs M3	**< 0·000**
	$\omega_0 = 0{\cdot}235$, $\omega_1 = 2{\cdot}414$, $\omega_2 = 20{\cdot}658$		
M7	$p = 0{\cdot}008$, $q = 0{\cdot}010$	M2 vs M3	0·161
M8	$p = 0{\cdot}134$, $q = 0{\cdot}205$	M7 vs M8	**< 0·000**
	$p_1 = 0{\cdot}015$, $\omega_1 = 9{\cdot}350$		
SSPE-H(NP)			
M0	$\omega = 0{\cdot}424$	M0 vs M2	**< 0·000**
M1	$p_0 = 0{\cdot}618$, $p_1 = 0{\cdot}382$	M1 vs M2	0·102
M2	$p_0 = 0{\cdot}630$, $p_1 = 0{\cdot}349$, $p_2 = 0{\cdot}022$	M0 vs M3	**< 0·000**
	$\omega_2 = 5{\cdot}186$		
M3	$p_0 = 0{\cdot}585$, $p_1 = 0{\cdot}339$, $p_2 = 0{\cdot}077$	M1 vs M3	0·206
	$\omega_0 = 0{\cdot}214$, $\omega_1 = 0{\cdot}216$, $\omega_2 = 3{\cdot}324$		
M7	$p = 0{\cdot}005$, $q = 0{\cdot}009$	M2 vs M3	0·512
M8	$p = 132{\cdot}576$, $q = 482{\cdot}362$	M7 vs M8	**0·048**
	$p_1 = 0{\cdot}076$, $\omega_1 = 3{\cdot}331$		
SSPE-H(CP)			
M0	$\omega = 0{\cdot}508$	M0 vs M2	< 0·000
M1	$p_0 = 0{\cdot}571$, $p_1 = 0{\cdot}429$	M1 vs M2	**0·004**
M2	$p_0 = 0{\cdot}561$, $p_1 = 0{\cdot}430$, $p_2 = 0{\cdot}008$	M0 vs M3	**< 0·000**
	$\omega_2 = 15{\cdot}314$		
M3	$p_0 = 0{\cdot}935$, $p_1 = 0{\cdot}063$, $p_2 = 0{\cdot}002$	M1 vs M3	**0·012**
	$\omega_0 = 0{\cdot}280$, $\omega_1 = 3{\cdot}832$, $\omega_2 = 41{\cdot}686$		
M7	$p = 0{\cdot}007$, $q = 0{\cdot}012$	M2 vs M3	0·359
M8	$p = 0{\cdot}022$, $q = 0{\cdot}029$	M7 vs M8	**0·004**
	$p_1 = 0{\cdot}009$, $\omega_1 = 14{\cdot}211$		
SSPE-F			
M0	$\omega = 0{\cdot}313$	M0 vs M2	**< 0·000**
M1	$p_0 = 0{\cdot}745$, $p_1 = 0{\cdot}255$	M1 vs M2	**0·022**
M2	$p_0 = 0{\cdot}767$, $p_1 = 0{\cdot}210$, $p_2 = 0{\cdot}022$	M0 vs M3	**< 0·000**
	$\omega_2 = 5{\cdot}923$		
M3	$p_0 = 0{\cdot}473$, $p_1 = 0{\cdot}497$, $p_2 = 0{\cdot}030$	M1 vs M3	**0·042**
	$\omega_0 = 0{\cdot}163$, $\omega_1 = 0{\cdot}163$, $\omega_2 = 5{\cdot}589$		
M7	$p = 0{\cdot}005$, $q = 0.\ 012$	M2 vs M3	0·328
M8	$p = 0{\cdot}008$, $q = 0{\cdot}131$	M7 vs M8	**0·033**
	$p_1 = 0.\ 093$, $\omega_1 = 2{\cdot}751$		

* For simplicity the ω ratios estimated between 0 and 1 for M7 and M8 have been omitted.

† Likelihood ratio tests are performed by taking twice the difference in log likelihood between two models and comparing the value obtained with a χ^2 distribution (degrees of freedom equal to the difference in the number of parameters between the models). Degrees of freedom are 2 for the M0 vs M2, M1 vs M2, M2 vs M3 and M7 vs M8 comparisons, and 4 for the M0 vs M3 and M1 vs M3 comparisons. *p*-values in bold indicate comparisons where the null hypothesis can be rejected in favour of the alternative hypothesis (i.e. the model on the left is rejected in favour of the one on the right).

Discussion

In this study, we investigated the evolutionary differences between the F, H and M protein coding genes from acute and persistent (SSPE) cases. The approximate doubling of the T_S/T_V ratio between the Acute-M and SSPE-M data sets clearly depicts the influence of A/I hypermutation in the M gene of SSPE patients (Table 2). It is important to note that A/I hypermutation was found in the H gene of the SSPE isolate IP-3-Ca (Cattaneo *et al.*, 1989) but the validity of this isolate to clinical SSPE infection has been questioned because its isolation involved multiple cell passaging and the formation of an SSPE-infected cell line (Billeter *et al.*, 1994). Therefore, it appears that A/I hypermutation is predominantly a phenomenon of the M gene from SSPE patients. Five variants of the M gene from SSPE patient S(B), designated S(B1) to S(B5), were isolated from several different brain regions by Baczko *et al.* (1993). Fig. 1 depicts a good example of convergent evolution in that variants S(B3) to S(B5), which are extensively hypermutated, cluster with high bootstrap support with the hypermutated MIBE strain S(C), instead of with their parental isolates S(B1) and S(B2). Conversely, all the S(B) variants grouped together with high bootstrap support (not shown) when sites of hypermutation were removed from the alignment. It appears that hypermutation can lead to similarity between unrelated strains such that M gene sequences from SSPE isolates should not be used for phylogenetic inference, whereas F and H gene sequences can be used for such a purpose.

The estimated rate of molecular evolution for SSPE strains ($3{\cdot}4 \times 10^{-4}$ substitutions/site/year) is very close to the acute rate of $4{\cdot}0 \times 10^{-4}$ substitutions/site/year estimated for epidemics of genotypes C2 and D3 in Spain (Rima *et al.*, 1997) and by using TipDate on a diverse set of sequences (Jenkins *et al.*, 2002). This similarity in rates is probably because a large proportion of the evolutionary history linking SSPE strains to one another has occurred within acute infections, such that the transition from acute to persistent infection occurs quite recently in the overall phylogenetic tree.

Selection in the F gene

Positive selection appears to have affected the evolution of the F gene in persistent (amino acid sites 62, 167, 446 and 461) but not acute MV cases (Table 4). The F_1 portion of the F protein contains two heptad repeat regions termed A (140–188) and B (463–484) (Lamb & Kolakofsky, 1996). Heptad repeat A is located directly after the fusion peptide (Chambers *et al.*, 1990) while heptad repeat B resides just prior to the transmembrane domain and contains a leucine zipper

Table 4. Sites of positive selection in the F and H genes

Positively selected site*	Data set	Comment†
F gene		
62	SSPE-F	Within potential BCE (56–70), within TCE (51–65)
167	SSPE-F	Within heptad repeat A (140–188), adjacent to potential BCE (151–165)
446	SSPE-F	Subsequent to TCE (435–443)
461	SSPE-F	Prior to heptad repeat B (463–484), adjacent to TCE (451–460)
H gene		
7	SSPE-H(NP)	Within cytoplasmic domain
8	SSPE-H(NP)	Within cytoplasmic domain
12	SSPE-H, SSPE-H(NP)	Within cytoplasmic domain
62	SSPE-H, SSPE-H(NP), SSPE-H(CP)	Within stem 1 domain
302	SSPE-H(NP)	Unknown relevance
423	SSPE-H(NP)	Within potential BCE (411–425)
481	Acute-H	Natural selection for binding of the CD46 receptor
546	Acute-H	Vero cell selection for binding of the CD46 receptor
560	SSPE-H(NP)	Unknown relevance
562	Acute-H	Within potential BCE (561–575)
575	Acute-H, SSPE-H(NP)	Within potential BCE (561–575)
618	SSPE-H	Selection for abolished stop codon

* Sites were identified to be under possible positive selection in the corresponding data sets with posterior probabilities above the 90% level.
† Comments for sites 302, 423, 481, 546, 560, 562 and 575 of the H gene were taken from Woelk *et al.* (2001).

motif (Buckland *et al.*, 1992). In parallel to observations concerning other viral envelope glycoproteins (Gething *et al.*, 1986; Kreis & Lodish, 1986; Spruce *et al.*, 1991), heptad repeat regions are thought to form long helices integral in tetramerization and higher oligomerization of the F protein which in turn govern the properties of fusion pore formation and cellular transport. Mutations in the leucine zipper motif of heptad repeat B have been shown to abolish fusion activity without disrupting the oligomeric form or surface expression of the protein (Chambers *et al.*, 1990). Putative positively selected site 167 lies within heptad repeat A and site 461 resides just prior to heptad repeat B. It is possible that substitutions at these sites have been selected to maintain persistence of MV within cells because they can lead to disruption of fusion pore formation and limit transport of the F protein to the cell surface. For instance, substitutions at site 167 appear to be from Ala to Thr or Val. Ala and Val are classified as hydrophobic but Thr contains a hydroxyl group, which increases hydrophilicity, and such a change could disrupt helix structure.

A number of studies using synthetic peptide approaches have attempted to identify B-cell epitopes (BCEs) and T-cell epitopes (TCEs) in the F protein of MV (Atabani *et al.*, 1997; Muller *et al.*, 1993, 1995, 1996; Partidos & Steward, 1990; van Binnendijk *et al.*, 1993; Wiesmüller *et al.*, 1992). The possible positive selection at site 461 may in fact be more relevant to immune selection because it is associated with a CD4 TCE identified by van Binnendijk *et al.* (1993). They noted that the minimal peptide needed for TCE recognition encompassed amino acids 451–460, to which site 461 is adjacent [the numbering of amino acids in van Binnendijk *et al.* (1993) is 3 more than it should be and this has been corrected here], and displays substitutions from the small hydrophilic Thr to the larger hydrophobic Ile. Van Binnendijk *et al.* (1993) also identified a CTL epitope (435–443) just prior to putative positively selected site 446. Substitutions at site 446 are from hydrophobic to more hydrophilic residues (Ile to Asn and Thr). Whether the changes discussed for sites 446 and 461 in close proximity to TCEs can abolish recognition by the cellular immune response requires more investigation. A further CD4 TCE (51–65) has been identified by Muller *et al.* (1996) and contains putative positively selected site 62. Using site-directed Ala substitutions, this study determined that sites 51, 54, 59, 63 and 65 interact with MV-induced T-cell lines and along with residues 52, 56 and 57 were important for stimulation of peptide specific T-cell lines. Since site 62 does not correspond to any of these positions it is unclear if substitutions at this site affect T-cell recognition. Substitutions at site 62 appear to be from the large hydrophobic Ile to Thr, which is more hydrophilic and unlikely to be representative of the Ala substitution used by Muller *et al.* (1996).

Putative positive selection at sites 62 and 167 also correlates with synthetic peptides shown to react with antibodies in human sera from donors previously exposed to MV (Muller *et al.*, 1993; Wiesmüller *et al.*, 1992). Whether the peptides identified by these studies correspond to bona fide BCEs is unclear because no immune function was shown for the antigenic regions identified and some reactions may represent non-specific binding of sera to peptide since experiments were performed in the absence of anti-MV-negative antisera (Atabani *et al.*, 1997). However, site 167 has already been discussed in association with structural features of the F protein and site 62 was found to correlate with a TCE.

Selection in the H gene

We have previously presented an analysis of a combined acute and SSPE data set of H gene sequences (Woelk *et al.*, 2001). That study identified weak positive selection at sites 12, 62, 211, 303, 348, 423, 476, 481, 546, 562 and 575, and demonstrated that removal of SSPE isolates from the analysis led to the loss of positive selection at seven of them – 12, 62, 211, 303, 348, 423, and 476. Here we show that positive selection in the remaining acute isolates could only be confirmed at sites 481, 546, 562 and 575 (Table 4). Sites 481 and 546 are thought to result from selection for the CD46 receptor and sites 562 and 575 were shown to correlate with potential BCEs. Previously we were unable to confirm which sites were positively selected in SSPE isolates because of a lack of data. The larger data set of H gene sequences from SSPE cases (SSPE-H) presented here overcomes this limitation and it appears that sites 12, 62 and 618 are under extremely strong positive selection (Table 4). Therefore, we suggest that different sites are under positive selection in SSPE and acute cases with selection at sites 12, 62 and 618 exclusive to SSPE isolates and selection at sites 481, 546, 562 and 575 exclusive to acute isolates. Positive selection at sites 211, 303, 348, 423 and 476 appears to have affected the evolution of both types of infection since a combined data set of SSPE and acute isolates is required for their identification.

The exact effect of the positive selection in the H gene at sites exclusive to SSPE isolates is unclear, but plausible explanations exist. Site 12 lies in the cytoplasmic domain and it is possible that substitutions at this site could affect M protein interaction during virus budding and hence lead to the maintenance of MV persistence. Substitutions at site 12 are from an ancestral Tyr (highly hydrophobic) to His, which can exist as uncharged or positively charged (hydrophilic) depending on its environment. This change could potentially hinder hydrophobic contacts between proteins. In our previous analysis (Woelk *et al.*, 2001) we were unable to find an antigenic feature that correlated with site 62 and the reason for positive selection at this site remains unclear. Site 62 lies within the stem 1 domain of the H protein, which projects the neuraminidase head above the envelope (Langedijk *et al.*, 1997). Substitutions at site 62 are from an ancestral hydrophilic Arg to either the highly hydrophobic Trp, or the extremely small Gln. Whether such changes have structural implications

for the H protein is not known but major distortions of the stem 1 region could affect efficient H protein function and help to maintain persistence. SSPE strains Osaka (1, 2 and 3), Yamagata-1, Sma79 and Sma94 have H proteins of abnormal length resulting from substitutions at site 618 which abolish the stop codon at this position. Whether this assists the maintenance of persistence in some way through irregular function or limited cellular transport is unclear.

Several SSPE strains were sequenced after various degrees of cell passaging (CP) and we were concerned that such strains were not truly representative of natural SSPE infection. For instance, the essence of persistent viruses is that they cannot replicate lytically and SSPE strains that have been cell passaged must have overcome this barrier. We therefore compared SSPE isolates that had been passaged with those that had not. Since most CP strains were passaged through Vero cells, it was surprising that sites known to interact with the CD46 receptor (see Woelk *et al.*, 2001) were not identified as being under positive selection. Some evidence for positive selection, although not significant, was observed at sites involved in CD46 interaction (481 and 546). Only positive selection at site 62 could be confirmed when cell passaged SSPE strains were analysed separately. This site has not previously been implicated in CD46 interaction (Langedijk *et al.*, 1997) and since it was also identified in the data set that was not cell passaged [SSPE-H(NP)] it is unlikely to be the result of Vero passaging. When this non-passaged data set was analysed, it was initially unclear as to whether a small class of sites was positively selected or neutral, but further testing (M7 vs M8) provided some evidence for positive selection at sites 7, 8, 12, 62, 302, 423, 560 and 575. Hence, this result suggests that selection at site 575 may not be exclusive to acute MV strains. Selection at sites 12 and 62 in the H gene have been discussed above, while the selection at sites 7 and 8 may be due to the same pressures hypothesized for site 12 (i.e. hindrance to M protein interaction). Sites 423 and 575 correspond to antigenic areas recognized by antibodies (Woelk *et al.*, 2001), but the relevance of selection at sites 302 and 560 is unknown (Table 4).

In summary, different selection pressures appear to be affecting the evolution of the envelope glycoproteins from acute and persistent cases of MV. For the F protein, positive selection was not evident in acute isolates but strong selection was seen in persistent cases. The ω value estimated under the M3 model of codon substitution – 5·589 – is equivalent to that seen in the haemagglutinin gene of influenza virus (ω = 4–7) (Yang, 2000; Yang *et al.*, 2000). Passaging was not a major problem for the F gene data set used here because nonsynonymous substitutions were seen at all of the positively selected sites in non-passaged strains S(A) and S(B). A more complicated scenario of positive selection exists in the H gene. Acute strains are undergoing moderate selection ($\omega > 2$), whereas analysis of all the available SSPE isolates suggests the influence of an extremely high force of positive selection (with ω ranging from 9·350 to 10·705), comparable to that seen in the V3 region of the HIV-1 envelope protein (ω = 6–30) (Nielsen & Yang, 1998). However, SSPE-H contains passaged strains such that this high positive selection pressure may not be indicative of natural SSPE evolution. Larger alignments of non-passaged isolates should help to obtain a more accurate picture of positive selection in the H gene of SSPE isolates. In any case, it is clear that far stronger positive selection pressure exists in SSPE cases compared to acute strains. Cattaneo *et al.* (1993) have proposed that single or multiple mutations in the envelope glycoproteins may be pivotal in the establishment as well as the maintenance of SSPE cases by impeding virus assembly and possibly reducing cell surface expression. This proposition correlates with amino acid sites under possible selection in the H protein from SSPE cases (sites 12, 62 and 618), which have the potential to reduce M protein interaction and cellular transport, as well as distort structure. Sites under potential selection in the F protein (62, 167, 446 and 461) show limited association with structural and antigenic features. Further investigation is needed to confirm the exact roles of the sites under potential positive selection in these envelope glycoproteins in establishing SSPE infection.

We thank Dr Michiko Watanabe for providing sequence data for the Biken isolate of SSPE, Dr Conall McCaughey for SSPE samples, two referees for useful suggestions, and Drs Andrew Rambaut and Robert Freckleton for analytical advice. The BBSRC, FAWCO, AWBS, The Wellcome Trust and The Royal Society funded this research.

References

Alkhatib, G., Shen, S. H., Briedis, D., Richardson, C., Massie, B., Weinberg, R., Smith, D., Taylor, J., Paoletti, E. & Roder, J. (1994). Functional analysis of N-linked glycosylation mutants of the measles virus fusion protein synthesized by recombinant vaccinia virus vectors. *Journal of Virology* **68**, 1522–1531.

Atabani, S. F., Obeid, O. E., Chargelegue, D., Aaby, P., Whittle, H. & Steward, M. W. (1997). Identification of an immunodominant neutralizing and protective epitope from measles virus fusion protein by using human sera from acute infection. *Journal of Virology* **71**, 7240–7245.

Ayata, M., Kimoto, T., Hayashi, K., Seto, T., Murata, R. & Ogura, H. (1998). Nucleotide sequences of the matrix protein gene of subacute sclerosing panencephalitis viruses compared with local contemporary isolates from patients with acute measles. *Virus Research* **54**, 107–115.

Baczko, K., Lampe, J., Liebert, U. G., Brinckmann, U., ter Meulen, V., Pardowitz, I., Budka, H., Cosby, S. L., Isserte, S. & Rima, B. K. (1993). Clonal expansion of hypermutated measles virus in a SSPE brain. *Virology* **197**, 188–195.

Billeter, M. A., Baczko, K., Schmid, A. & ter Meulen, V. (1984). Cloning of DNA corresponding to four different measles virus genomic regions. *Virology* **132**, 147–159.

Billeter, M. A., Cattaneo, R., Spielhofer, P., Kaelin, K., Huber, M., Schmid, A., Baczko, K. & ter Meulen, V. (1994). Generation and properties of measles virus mutations typically associated with subacute sclerosing panencephalitis. *Annals of the New York Academy of Sciences* **724**, 367–377.

Boom, R., Sol, C. J. A., Salimans, M. M. M., Jansen, C. L., Wertheim van Dillen, P. M. E. & van der Noordaa, J. (1990). Rapid and simple method

for purification of nucleic acids. *Journal of Clinical Microbiology* **28**, 495–503.

Buckland, R., Gerald, C., Barker, R. & Wild, T. F. (1987). Fusion glycoprotein of measles virus – nucleotide sequence of the gene and comparison with other paramyxoviruses. *Journal of General Virology* **68**, 1695–1703.

Buckland, R., Cheynet, V., Beauverger, P. & Wild, F. (1990). Cloning of the matrix gene of measles virus (Hallé strain). *Nucleic Acids Research* **18**, 5283.

Buckland, R., Malvoisin, E., Beauverger, P. & Wild, F. (1992). A leucine zipper structure present in the measles virus fusion protein is not required for its tetramerization but is essential for fusion. *Journal of General Virology* **73**, 1703–1707.

Cattaneo, R. & Rose, J. K. (1993). Cell fusion by the envelope glycoproteins of persistent measles viruses which caused lethal human brain disease. *Journal of Virology* **67**, 1493–1502.

Cattaneo, R., Schmid, A., Eschle, D., Baczko, K., ter Meulen, V. & Billeter, M. A. (1988). Biased hypermutation and other genetic changes in defective measles viruses in human brain infections. *Cell* **55**, 255–265.

Cattaneo, R., Schmid, A., Spielhofer, P., Kaelin, K., Baczko, K., ter Meulen, V., Pardowitz, J., Flanagan, S., Rima, B. K., Udem, S. A. & Billeter, M. A. (1989). Mutated and hypermutated genes of persistent measles viruses which caused lethal human brain diseases. *Virology* **173**, 415–425.

Chambers, P., Pringle, C. R. & Easton, A. J. (1990). Heptad repeat sequences are located adjacent to hydrophobic regions in several types of virus fusion glycoproteins. *Journal of General Virology* **71**, 3075–3080.

Curran, M. D. & Rima, B. K. (1988). Nucleotide sequence of the gene encoding the matrix protein of a recent measles virus isolate. *Journal of General Virology* **69**, 2407–2411.

Enami, M., Sato, T. A. & Sugiura, A. (1989). Matrix protein of cell-associated subacute sclerosing panencephalitis viruses. *Journal of General Virology* **70**, 2191–2196.

Fayolle, J. L., Verrier, B., Buckland, R. & Wild, T. F. (1999). Characterization of a natural mutation in an antigenic site on the fusion protein of measles virus that is involved in neutralization. *Journal of Virology* **73**, 787–790.

Furukawa, K., Ayata, M., Kimura, M., Seto, T., Matsunaga, I., Murata, R., Yamano, T. & Ogura, H. (2001). Hemadsorption expressed by cloned H genes from subacute sclerosing panencephalitis (SSPE) viruses and their possible progenitor measles viruses isolated in Osaka, Japan. *Microbiology and Immunology* **45**, 59–68.

Gerald, C., Buckland, R., Barker, R., Freeman, G. & Wild, T. F. (1986). Measles virus haemagglutinin gene: cloning, complete nucleotide sequence analysis and expression in COS cells. *Journal of General Virology* **67**, 2695–2703.

Gething, M. J., McCammon, K. & Sambrook, J. (1986). Expression of wild-type and mutant forms of influenza haemagglutinin – the role of folding in intracellular transport. *Cell* **46**, 939–950.

Hu, A. Z., Sheshberadaran, H., Norrby, E. & Kovamees, J. (1993). Molecular characterization of epitopes on the measles virus haemagglutinin protein. *Virology* **192**, 351–354.

Hu, A. H., Cattaneo, R., Schwartz, S. & Norrby, E. (1994). Role of *N*-linked oligosaccharide chains in the processing and antigenicity of measles virus haemagglutinin protein. *Journal of General Virology* **75**, 1043–1052.

Hu, A. Z., Cathomen, T., Cattaneo, R. & Norrby, E. (1995). Influence of *N*-linked oligosaccharide chains on the processing, cell surface expression and function of the measles virus fusion protein. *Journal of General Virology* **76**, 705–710.

Jenkins, G. M., Rambaut, A., Pybus, O. G. & Holmes, E. C. (2002). Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. *Journal of Molecular Evolution* **54**, 152–161.

Jin, L., Richards, A. & Brown, D. W. G. (1996). Development of a dual target-PCR for detection and characterization of measles virus in clinical specimens. *Molecular and Cellular Probes* **10**, 191–200.

Jin, L., Knowles, W. A., Rota, P. A., Bellini, W. J. & Brown, D. W. G. (1998). Genetic and antigenic characterisation of the haemagglutinin protein of measles virus strains recently circulating in the UK. *Virus Research* **55**, 107–113.

Johnston, I. C. D., ter Meulen, V., Schneider-Schaulies, J. & Schneider-Schaulies, S. (1999). A recombinant measles vaccine virus expressing wild-type glycoproteins: consequences for viral spread and cell tropism. *Journal of Virology* **73**, 6903–6915.

Komase, K., Haga, T., Yoshikawa, Y., Sato, T. A. & Yamanouchi, K. (1990a). Molecular analysis of structural protein genes of the Yamagata-1 strain of defective subacute sclerosing panencephalitis virus. III. Nucleotide sequence of the haemagglutinin gene. *Virus Genes* **4**, 163–172.

Komase, K., Haga, T., Yoshikawa, Y., Sato, T. A. & Yamanouchi, K. (1990b). Molecular analysis of structural protein genes of the Yamagata-1 strain of defective subacute sclerosing panencephalitis virus. IV. Nucleotide sequence of the fusion gene. *Virus Genes* **4**, 173–181.

Kreis, T. E. & Lodish, H. F. (1986). Oligomerization is essential for transport of vesicular stomatitis viral glycoprotein to the cell surface. *Cell* **46**, 929–937.

Lamb, R. A. (1993). Paramyxovirus fusion: a hypothesis for changes. *Virology* **197**, 1–11.

Lamb, A. L. & Kolakofsky, D. (1996). *Paramyxoviridae*: the viruses and their replication. In *Fields Virology*, 3rd edn, pp. 577–604. Edited by B. N. Fields, D. M. Knipe & P. M. Howley. Philadelphia: Lippincott–Raven.

Langedijk, J. P. M., Daus, F. J. & van Oirschot, J. T. (1997). Sequence and structure alignment of *Paramyxoviridae* attachment proteins and discovery of enzymatic activity for a morbillivirus haemagglutinin. *Journal of Virology* **71**, 6155–6167.

Liebert, U. G. (1997). Measles virus infections of the central nervous system. *Intervirology* **40**, 176–184.

Manchester, M., Eto, D. S., Valsamakis, A., Liton, P. B., Fernandez-Munoz, R., Rota, P. A., Bellini, W. J., Forthal, D. N. & Oldstone, M. B. A. (2000). Clinical isolates of measles virus use CD46 as a cellular receptor. *Journal of Virology* **74**, 3967–3974.

Miller, C., Farrington, C. P. & Harbert, K. (1992). The epidemiology of subacute sclerosing panencephalitis in England and Wales 1970–1989. *International Journal of Epidemiology* **21**, 998–1006.

Muller, C. P., Handtmann, D., Brons, N. H. C., Weinmann, M., Wiesmuller, K. H., Spahn, G., Wiesneth, M., Schneider, F. & Jung, G. (1993). Analysis of antibody-response to the measles virus using synthetic peptides of the fusion protein – evidence of nonrandom pairing of T-cell and B-cell epitopes. *Virus Research* **30**, 271–280.

Muller, C. P., Bunder, R., Mayser, H., Ammon, S., Weinmann, M., Brons, N. H. C., Schneider, F., Jung, G. & Wiesmuller, K. H. (1995). Intramolecular immunodominance and intermolecular selection of H2(D)-restricted peptides define the same immunodominant region of the measles virus fusion protein. *Molecular Immunology* **32**, 37–47.

Muller, C. P., Ammerlaan, W., Fleckenstein, B., Krauss, S., Kalbacher, H., Schneider, F., Jung, G. & Wiesmuller, K. H. (1996). Activation of T cells by the ragged tail of MHC class II-presented peptides of the measles virus fusion protein. *International Immunology* **8**, 445–456.

Nielsen, R. & Yang, Z. (1998). Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* **148**, 929–936.

Ohuchi, M., Ohuchi, R., Mifune, K., Ishihara, T. & Ogawa, T. (1987). Characterization of the measles virus isolated from the brain of a patient with immunosuppressive measles encephalitis. *Journal of Infectious Diseases* **156**, 436–441.

Partidos, C. D. & Steward, M. W. (1990). Prediction and identification of a T-cell epitope in the fusion protein of measles virus immunodominant in mice and humans. *Journal of General Virology* **71**, 2099–2105.

Rambaut, A. (2000). Estimating the rate of molecular evolution: incorporating non-contemporaneous sequences into maximum likelihood phylogenies. *Bioinformatics* **16**, 395–399.

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.

Roos, R. P., Graves, M. C., Wollmann, R. L., Chilcote, R. R. & Nixon, J. (1981). Immunological and virologic studies of measles inclusion body encephalitis in an immunosuppressed host: the relationship to subacute sclerosing panencephalitis. *Neurology* **31**, 1263–1270.

Rota, J. S., Hummel, K. B., Rota, P. A. & Bellini, W. J. (1992). Genetic variability of the glycoprotein genes of current wild-type measles isolates. *Virology* **188**, 135–142.

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

Scheid, A. & Choppin, P. (1974). Identification of biological activities of paramyxovirus glycoproteins. Activation of cell fusion, hemolysis, and infectivity of proteolytic cleavage of an inactive precursor protein of Sendai virus. *Virology* **57**, 475–490.

Schneider-Schaulies, S. & Liebert, U. G. (1991). Pathogenic aspects of persistent measles virus infections in brain tissue. *Seminars in the Neurosciences* **3**, 149–155.

Schneider-Schaulies, S., Schneider-Schaulies, J., Dunster, L. M. & ter Meulen, V. (1995). Measles virus gene expression in neural cells. In *Measles Virus*, pp. 101–116. Edited by V. ter Meulen & M. A. Billeter. Berlin: Springer.

Spruce, A. E., Iwata, A. & Almers, W. (1991). The first milliseconds of the pore formed by a fusogenic viral envelope protein during membrane fusion. *Proceedings of the National Academy of Sciences, USA* **88**, 3623–3627.

Swofford, D. L. (2000). Phylogenetic analysis using parsimony (*and other methods). Version 4.0b6. Sunderland, MA: Sinauer Associates.

Takeda, M., Kato, A., Kobune, F., Sakata, H., Li, Y., Shioda, T., Sakai, Y., Asakawa, M. & Nagai, Y. (1998). Measles virus attenuation associated with transcriptional impediment and a few amino acid changes in the polymerase and accessory proteins. *Journal of Virology* **72**, 8690–8696.

Takeuchi, K., Miyajima, N., Kobune, F. & Tashiro, M. (2000). Comparative nucleotide sequence analyses of the entire genomes of B95a cell-isolated and Vero cell-isolated measles viruses from the same patient. *Virus Genes* **20**, 253–257.

Tatsuo, H., Ono, N., Tanaka, K. & Yanagi, Y. (2000). SLAM (CDw150) is a cellular receptor for measles virus. *Nature* **406**, 893–897.

ter Meulen, V., Stephenson, J. R. & Kreth, H. W. (1983). Subacute sclerosing panencephalitis. *Comprehensive Virology* **18**, 105–159.

van Binnendijk, R. S., Versteeg-van Oosten, J. P. M., Poelen, M. C. M., Brugghe, H. F., Hoogerhout, P., Osterhaus, A. D. M. E. & Uytdehaag, F. G. C. M. (1993). Human HLA class I-restricted and HLA class II-restricted cloned cytotoxic T-lymphocytes identify a cluster of epitopes on the measles virus fusion protein. *Journal of Virology* **67**, 2276–2284.

WHO (2001). Nomenclature for describing the genetic characteristics of wild-type measles virus (update). *WHO Weekly Epidemiological Report* **76**, 241–247.

Wiesmüller, K. H., Spahn, G., Handtmann, D., Schneider, F., Jung, G. & Muller, C. P. (1992). Heterogeneity of linear B-cell epitopes of the measles virus fusion protein reacting with late convalescent sera. *Journal of General Virology* **73**, 2211–2216.

Woelk, C. H. & Holmes, E. C. (2001). Variable immune-driven natural selection in the attachment (G) glycoprotein of respiratory syncytial virus (RSV). *Journal of Molecular Evolution* **52**, 182–192.

Woelk, C. H., Jin, L., Holmes, E. C. & Brown, D. W. G. (2001). Immune and artificial selection in the haemagglutinin (H) glycoprotein of measles virus. *Journal of General Virology* **82**, 2463–2474.

Wong, T. C., Ayata, M., Hirano, A., Yoshikawa, Y., Tsuruoka, H. & Yamanouchi, K. (1989). Generalized and localized biased hypermutation affecting the matrix gene of a measles virus strain that causes subacute sclerosing panencephalitis. *Journal of Virology* **63**, 5464–5468.

Wong, T. C., Ayata, M., Ueda, S. & Hirano, A. (1991). Role of biased hypermutation in evolution of subacute sclerosing panencephalitis virus from progenitor acute measles virus. *Journal of Virology* **65**, 2191–2199.

Yang, Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in the Biosciences* **13**, 555–556.

Yang, Z. (2000). Maximum likelihood estimation on large phylogenies and analysis of adaptive evolution in human influenza virus A. *Journal of Molecular Evolution* **51**, 423–432.

Yang, Z. & Bielawski, J. P. (2000). Statistical methods for detecting molecular adaptation. *Trends in Ecology & Evolution* **15**, 496–503.

Yang, Z., Nielsen, R., Goldman, N. & Pedersen, A. M. K. (2000). Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**, 431–449.

Received 8 November 2001; Accepted 25 January 2002