Paper III

Clinical herpes simplex virus type 2 isolates from Tanzania and Norway cluster in two different genogroups

Kasubi MJ[1,2,3], Norberg P[4], Nilsen A[5], Liljeqvist JA[4], Langeland N[6], Bergström T[4], Haarr L[3]

[1] Center for International Health, University of Bergen, Bergen, Norway.

[2] Department of Microbiology and Immunology, Muhimbili University College of Health Sciences, Dar Es Salaam, Tanzania.

[3] Department of Microbiology and Immunology, The Gade Institute, University of Bergen, Norway.

[4] Department of Infectious Diseases, Göteborg University, Göteborg, Sweden.

[5] Department of Dermatology Haukeland University Hospital and University of Bergen, Bergen, Norway.

[6] Department of Medicine, Haukeland University Hospital and University of Bergen, Norway

ABSTRACT

Genetic diversity among clinical strains of herpes simplex virus type 2 (HSV-2), of potential importance for diagnostics and vaccine development, is little known. Here, we have sequenced the genes US 4, 7 and 8 encoding glycoproteins (g) G, I and E from eleven Tanzanian strains and ten Norwegian isolates. Overall, the sequence variation was low, with an overall similarity between the two most distant isolates of 99.6%. By phylogenetic analysis the strains clustered in two genogroups, arbitrarily designated as African (A) and European (E), with high bootstrap values. For gE/gI as well as for gG, all of the strains belonging to genogroup A (eight and five, respectively) were isolated in Tanzania, while the E genogroup contained both Tanzanian and Norwegian strains. We interpret the finding of both genogroups in Africa only as compatible with a later introduction of the E genogroup into that continent. In addition, three Tanzanian strains that were placed in genogroup A for gE/gI belonged to genogroup E as judged from gG phylogeny, consistent with intergenic recombination. In all isolates a region of 14 codons of the gE gene was different from that of the laboratory strain HSV-2 HG52. The epitope in gG-2 mimicked in the peptide 55 ELISA for specific detection of HSV-2 antibodies (Marsden et al., 1998) was genetically stable among Tanzanian and Norwegian isolates, suggesting that the ELISA would be reliable in different parts of the world.

INTRODUCTION

Herpes simplex virus type 2 (HSV-2), a member of the alphaherpesvirinae subfamily of human herpesviridae, is regarded as the most common source of genital ulcers globally. Development of methods to control the spread of sexually transmitted pathogens is of utmost importance and is urgently needed, especially in some developing countries. Recent studies have underscored that a high prevalence of genital ulcers caused by HSV-2 constitute a major risk factor for infection by HIV. Often, both viruses are transmitted simultaneously (Ramjee et al., 2005). Control measures of HSV-2 spread in Africa in form of antiviral prophylaxis has been suggested as a mean to limit spread also of HIV (Cohen, 2005). Such measures require serological testing for optimal usage of anti-herpetic drugs.

Somewhat surprisingly, seroprevalence reports on HSV-2 differ markedly between Tanzania and Scandinavia in the younger age groups, with substantial antibody positivity in the former country (Tunbäck et al. 2003; Kasubi et al. 2006). The reasons for that are unknown, but could include other (i.e. extragenital) means of spread of HSV-2 in Africa, or that genetic differences exist in HSV-1 and HSV-2 strains from this continent as compared to the Western world. Such genetic differences could bias type-specific diagnosis since current antigens for detection of serum antibodies are based on amino acid sequences of gG from Western isolates. Hence, a focus on HSV-2 seroprevalence investigations in African populations requires background knowledge on local genetic variability of this virus, and it's nearest relative HSV-1, and the consequences thereof for antibody reactivity. We have recently shown that a colinear sequence of gG-1 and gG-2 constitute important epitope regions that can be used for type specific serodiagnosis, but that this specificity heavily depends on a few

key residues (Nilsen et al., 2003; Kasubi et al., 2005; Tunbäck et al., 2000; Tunbäck et al., 2005).

In the present study, we have compared DNA sequences from a limited number of African and European isolates of HSV-2, derived from patients with genital herpes. We have focussed on the US4 gene encoding gG-2 due to the above mentioned reasons, and in addition on US7 (gI) and US8 (gE), since the latter were successfully utilized to discriminate three genogroups in HSV-1 through phylogenetic analysis (Norberg et al., 2004). In the present study we found, despite a very low overall interstrain diversity, evidence of two distinct genogroups, denoted African (A) and European (E), and recombinants thereof. The findings form a base for further study of the genetic relationship of Tanzanian and Scandinavian HSV-2 strains and seroreactivity to these strains.

Table 1a. Primers used for amplification and sequencing of the gG-2 gene

Gene	Nucleotide position[a]	Type[b]	Sequence
gG-2	4082-4100	S	5′GCACAAAAAGACGCGGCCC3′
	4361-4379	S	5′CGTCGTCCGTCACGAGCCC3′
	4445-4427[c]	AS	5′TGCGCCAAATCCGCGTACC3′
	4654-4673[c]	S	5′ACCGCGCCGCGGAGACATTC3′
	4739-4721	AS	5′CTCCCCGCCCACCTCTACC3′
	4986-5005[c]	S	5′TTTATTCGGATGGCACGACC3′
	5020-5038[c]	AS	5′CAATGGGGCGGCAGGACCC3′
	5191-5210[c]	S	5′CCTCCGATTCGCCTACGTCC3′
	5335-5316[c]	AS	5′GTTGCGGCTTGTGTGGCCAT3′
	5444-5461 [c]	S	5′CACGAACACCAGCAGCGC3′
	5489-5471[c]	AS	5′TGTGGTGGGGGGCGTTTTC3′
	5569-5551[c]	AS	5′GGAGGGGTTGTTTGGGGCC3′
	5759-5738[c]	AS	5′GTGTGGGTGCGTCTTTGGGTCC3′
	5900-5919[c]	S	5′GAACCCCAACAAACCACCCC3′
	6278-6258	AS	5′TCCCGTCCTTCATCGTTTCTC3′

[a]For the US region

[b]S, sense; AS, antisense.

[c]Used only for sequencing.

Table 1b. Primers used for amplification and sequencing of the gI-2 and gE-2 genes

Gene	Nucleotide position[a]	Type[b]	Sequence
gI-2	8604-8620	S	5′GCTGTCCCGACGATTAG3′
	8812-8832[c]	S	5′TGCGTGTTTTCGGGGAGCTTC3′
	9111-9128[c]	S	5′GTCGGCAGCGCGACGAAC3′
	9166-9148[c]	AS	5′TTGGCAGAGAGCGCCACCC3′
	9375-9394[c]	S	5′ATAGCCCCGCCCAATTCCAC3′
	9661-9680[c]	S	5′CGTCCACGACCATGCCTTCC3′
	9715-9734[c]	S	5′CAGTCGTGCTGCTGTCCGTC3′
	9719-9698[c]	AS	5′GACTGGACCTGGCTCCGATTCC3′
gE-2	10058-10078 [c]	S	5′CAGCTAGTCTCCGATCTGCCC3′
	10165-10148[c]	AS	5′CCGCCAGGCACGATACGA3′
	10379-10400[c]	S	5′CGCTCGCCATAGCATACAGTCC3′
	10454-10472[c]	S	5′GCGTAGCCGTGGTCAACGA3′
	10578-10560[c]	AS	5′GACCAGAACCACCGACGCC3′
	10740-10757[c]	S	5′CCACGTGCGCGGGGTAAC3′
	10879-10897[c]	S	5′CGGTTTGACGTGCCGTCCT3′
	10920-10900[c]	AS	5′GTAGATCCGCATATCGGCGCA3′
	11006-10988[c]	AS	5′AGGCGGTACGCCCAGGAAC3′
	11236-11254[c]	S	5′AACGCGGTGGTGGAACAGC3′
	11418-11401[c]	AS	5′GCAGGTCATGCACGCCCA3′
	11786-11766	AS	5′TGGCAATCAGTTCATCGCCGA3′

[a]For the US region

[b]S, sense; AS, antisense.

[c]Used only for sequencing

MATERIALS AND METHODS

Clinical isolates of HSV-2.

HSV-2 isolates were collected from patients with genital lesions attending sexually transmitted disease clinics in Bergen, Norway, or in Dar es Salaam, Tanzania. Ten isolates were from Norway and 11 from Tanzania. Sterile dacron swabs were used to collect materials from the bases of either ulcers or vesicles which were deliberately ruptured prior to specimen collection. The swabs were immediately put in a liquid virus transport medium. Tanzanian samples were stored at –80^{o}C until further analysis, while virus was cultivated immediately from the Norwegian samples.

Virus stocks were prepared by infecting baby hamster kidney (BHK) cells which were grown in Eagles minimal essential medium supplemented with 2% calf serum and antibiotics. All virus isolates were confirmed to be HSV-2 by using nested PCR targeting either the type-specific promoter region of the gD-1 gene for HSV-1 or coding sequences within the gG-2 gene of HSV-2, as described previously (Aurelius et al., 1993; Cinque et al., 1996). Laboratory strain HSV-2 HG52 (McGeoch et al., 1987) was used as reference. The virus isolates are listed in Table 2

PCR amplification and sequencing.

Two regions of the HSV-2 genome were amplified prior to sequencing. One set of primers was designed to amplify a 2197 bp fragment spanning from 58 bp upstream of the start codon of the US4 gene (encoding gG-2) to 39 bp downstream of the termination codon. Amplification with the other set of primers resulted in a 3182 bp fragment starting from 57 bp upstream of the start codon of the US7 gene (encoding gI-2) and extending to 47 bp

downstream of the stop codon for the US8 gene (encoding gE-2), thus including non-coding sequences between the two genes. The primers are shown in Table 1.

Each PCR reaction was performed in a total volume of 50µl using an automated Gene AMP system 2400 thermal cycler (Perkin-Elmer Corporation, Norwalk CT/USA). The reaction mixture contained 17.5µl H_2O, 1µl of each of the reverse and forward primer, at a concentration of 1 pmole/µl, 0.5 µl Tfl DNA polymerase (Boule Nordic, Stockholm), 25µl buffer solution GN (Boule Nordic, Stockholm) and 5µl of diluted purified DNA extract. The incubation steps were initial denaturation at 96^0C for 5 minutes followed by 30 cycles of denaturation at 95^0C for 1 minute, annealing of primers for 1 min at 60^0C for US4 and 57^0C for US7/US8, elongation for 3 minutes at 68^0C, and finally an extension cycle of 68^0C for15 minutes. The sizes of amplified PCR products were analysed by electrophoresis of 10 µl samples in 1% agarose gels stained with ethidium bromide. PCR products were purified using the QIA-quick Purification Kit (Qiagen. Germany) according to the manufacturer's instruction.

Sets of overlapping primers, as shown in Table 1a and 1b , were then used for sequencing.
The ABI Prism Big Dye terminator Cycle Sequencing Read Reaction kit (Applied Biosystems) was used. The reaction mixture contained 1µl 5X sequencing buffer, 2µl Big-dye, 4.4 µl H_20, 1µl PCR product and 1.6 µl primer at a concentration of 1 pmole/µl and 10µl deionized H_20 in a total volume of 20µl. Incubation was according to the following program: 1 min at 96^0C followed by 10 s at 96^0C, 5 s at 50^0C and 4 min at 60^0C for 25 cycles. The reaction mixtures were then treated with sequencing Reaction Cleanup Kit (Biomek 2000) according to manufacturer's protocol. Sequencing of both strands was performed in a AB1 Prism 3700 DNA analyzer (Applied Biosystem). For all parts of either the gG gene or the

combined gI and gE genes a minimum of two sequences were obtained in parallel experiments. Sequences were assembled using DNA Sequence Assembly Software version 3.7 (Applied Biosystems).

The sequences were further analyzed by using the Staden sequence analysis package (Staden 1994). Multiple sequences were aligned by the Clustal W (1.86) program and compared with those of the reference strain HG52.

Phylogenetic analysis. Due to the conserved nature of HSV-2, all sequences were easily aligned manually. To avoid inference by possible hypervariable regions, all gaps in the alignment were excluded prior to all analyses. Phylogenetic trees were constructed using the maximum likelihood method included in the Phylip package (Felsenstein, 1993). To estimate the significance of the trees, the calculations were based on 100 bootstrap replicates of each alignment and the consensus trees were selected as the most likely.

RESULTS

PCR amplification and DNA sequencing. Due to high content of nucleotides G and C, several overlapping primer systems had to be used. All strains could then be successfully amplified and sequenced. Resequencing and overlapping sequencing showed identical results throughout the study. The similarity between the two most distant isolates, Tz-2557 and Ng-No6, (Fig. 1) were approximately 99,6 %, which indicates that HSV-2 is more conserved than HSV-1 but slightly less conserved than VZV. Overall genetic distances between the investigated strains, based upon the sequences of the gG-2 gene, are shown in Figure 1.

Detailed analysis of the gene encoding gG-2

The coding sequences of the gG-2 gene consist of 2097 nucleotides. Comparison of the sequences from the clinical isolates with those of the gG-2 gene of HSV-2 HG52 (McGeoch et al., 1987) is shown in Table 3. All 21 isolates harboured three different types of mutations. Substitution of one nucleotide resulted either in a silent mutation or replacement of one amino acid. Deletion of a codon led to deletion of an amino acid without affecting the reading frame. Rather than being scattered randomly, the mutations were located to certain positions along the gene. Some mutations affected several isolates, whereas others were only observed in one or two strains. All isolates had a T→A substitution at codon 1048 and deletion of one codon (for A) after nucleotide 1282, Alterations, in comparison with HG52, that were detected in all strains were regarded as the "consensus sequence". Several identical mutations were observed in some of the Norwegians as well as in some of the Tanzanian isolates. These included amino acid substitutions at nucleotides 104 (S→N) 611 (P→L), 872 (E→G), 1013 (R→L)) and 1268 (L→P). Furthermore, silent mutations at nucleotide positions 274, 930 and 1116 affected several Tanzanian and Norwegian isolates. Some mutations were only detected in a small number of Norwegian strains. These mutations included 11 different amino acid substitutions (at nucleotide positions 172, 329, 635, 729, 1045, 1324, 1499, 1568, 1646, 1722 and 1853), 6 silent mutations (at nucleotide positions 432, 891, 1746, 1758, 1800 and 2008) and deletion of amino acid V (after nucleotide 877). Likewise, other mutations were specific for some of the Tanzanian isolates. These included 6 different amino acid substitutions (at nucleotides 127, 406, 499, 1115, 1215 and 1627) and 2 silent mutations (at nucleotide positions 711 and 1581). Variations at nucleotides 1115 and 1116 were observed in a number of isolates from both countries, leading to either silent mutations or amino acid substitutions. Three overlapping and linear type-specific epitopes in the gG-2 protein have been identified by using monoclonal antibodies (Liljeqvist et al. 1998). One of

these epitopes was immunodominant and is marked in Fig. 2. Furthermore, Marsden et al. (1998) used overlapping, synthetic oligopeptides to identify a type-specific epitope within the same protein. This epitope (detected with peptide 55) is also shown in Fig. 2. Mutations within the epitope regions 1 and 11 identified by monoclonal antibodies were detected in 2 Norwegian isolates only. Norwegian isolates 5 and 7 harboured single mutations leading to amino acid substitutions S→L and E→D at positions 549 (nucleotide position 1646) and 574 (nucleotide position 1722), respectively. Position 549 is outside the peptide 55-region, and position 574 close to its C-terminal. These results indicate a genetic stability within the type-specific epitopes identified in one way or another.

Detailed analysis of the genes encoding gI-2 and gE-2 and the non-coding sequences between the genes

The coding sequences for the genes gI-2 and gE-2 contain 1116 and 822 nucleotides, respectively. The non-coding region between the genes is 325 nucleotides long. These sequences of all isolates were again compared with the corresponding sequences of the laboratory strain HSV-2 HG52, and the results are shown in Tables 4, 5 and 6. All clinical isolates harboured mutations. As for the gG-2 gene, these mutations were limited to certain sites. Analysis of the gI-2 gene is shown in Table 4. All isolates had a silent mutation at nucleotide position 198. Two identical mutations were observed in a few Norwegian and Tanzanian isolates. These were at positions 476-477 and 643, leading to amino acid substitutions R→L and P→S, respectively. All Tanzanian and 5 Norwegian isolates had the same mutation at position 716 so that amino acid T was sbustituted by R. Some mutations were only detected in a few Norwegian strains. These included 6 different amino substitutions (at nucleotide positions 338, 448, 503, 618, 716 and 1043, and 2 silent mutations (at positions 618 and 798). Other mutations occurred in Tanzanian isolates only, These were 4 different

amino acid substitutions (positions 29, 51, 476-477 and 644) and 1 silent mutation (position 1062), The most frequent Tanzanian mutation was the amino acid substitution R→L (positions 476 and 477). Nucleotides 716 and 717 were highly variable, leading to three different amino acid substitutions observed in a number of isolates. Different types of mutations were also detected at positions 476-477 and 643-644.

More extensive differences between the laboratory strain and the clinical isolates were observed in the gE-2 gene, as shown in Table 5. All isolates had amino acid substitutions at nucleotide positions 671 (L→P), 804 (D→E) and 1621 (P→S). However, at positions 671 and 804 all isolates are identical to strain G of HSV-2 (Choi et al, 1996). Furthermore, all strains differed from HG52 but were similar to strain G by an insertion of amino acids GPE between positions 129 and 130. gE of HG52 is consequently 3 amino acids shorter than that of the other strains. A frameshift was detected in all isolates by an inseretion between nucleotides 541 and 542. Except for isolate T9, a second shift after 14 codons brought the translation back in frame. The deduced amino acids in this region are shown in table 5. Strain T9 was back in frame after 23 codons. Identical mutations in Norwegian and Tanzanian isolates were observed at nucleotide positions 570 and 1211 (amino acid substitution H→P). Two differrent amino acid substitutions (nucleotide positions 59 and 396), and 7 different silent mutations (positions 384, 531, 1140, 1245, 1273 and 1551) were present in Norwegian isolates only. Mutations specific for Tanzanian isolates were 4 different amino acid substitutions (nucleotide positions 127, 342, 392 and 605) and 1 silent mutation (position 138). The L→R transition at nucleotide position 392 was detected in the majority of the Tanzanian isolates.

Particularly two regions in the non-coding sequences between the gI-2 and gE-2 genes were subjected to variations, as shown in Table 6. In the first region, which is extremely rich in C

and is delimited by nucleotides 60 – 69, there were deletions of 9 or 10 nucleotides in several isolates from either country. In contrast, insertions of a viariable number of G nucleotides were detected in the second region, delimited by nucleotides 178 and 179 in the reference strain. Insertions had occurred in both Tanzanian and Norwegian isolates.

Phylogenetic analysis

All detected sequences, including the three genes as well as non-coding regions, were analysed by the maximum likelyhood method and phylogenetic trees were obtained. The tree for the genes encoding gI-2/gE-2 (Fig 3) divided the isolates into two main groups designated E (European) and A (African). Group A contained sequences of Tanzanian origin only, while group E had both Norwegian and Tanzanian isolates. Three isolates of Tanzanian origin , specifically Tz 3034, Tz 2302 and Tz 3111, were apparently closely related to Norwegian isolates.

Results from analysis of the gG-2 sequences (Fig. 3) confirm the separation into the genetic groups E and A. Group A consisted of five Tanzanian isolates. Group E contained six Tanzanian and ten Norwegian isolates. Three isolates from Tanzania, specifically Tz 3034, Tz 2302 and Tz 3111, clustered in genetic group E in both trees, whereas three other isolates of the same origin (Tz 2032, Tz 1855 and Tz 2737) were in the A group in the gE-2/gI-2 tree, but in the E group in the gG-2 tree. These results indicate that at least three isolates (Tz 2032, Tz 1855 and Tz 2737) were recombinants, derived from recombination events between other isolates belonging to genotypes A and E, respectively.

DISCUSSION

Two sexually transmitted viruses, HIV and HSV-2, constitute a major health problem in developing countries, and may enhance spread of each other. Different strategies such as use of microbicides have been advocated to reduce spread of both viruses (Cohen, 2005). Hitherto, attempts to create vaccines have been futile. One important limitation is the genetic variability of HIV, which makes choices of selection of optimal sequences difficult. The genetic diversity among HSV-2 strains is currently unknown, and the present study aims to increase such knowledge.

Genetic variation among strains of human herpes viruses has been used to distinguish viral genotypes and provides useful information for defining molecular epidemiology of infection (Franti et al., 1998; Meng et al., 1999; Shepp et al., 1998; Umene et al., 1997, Zweygberg et al., 1998). Human viruses of the alphaherpesvirinae subfamily have been reported to display a comparatively low degree of genetic inter-strain variability. Despite their conserved genetics, distinctive genogroups have been described for herpes simplex virus type 1 (HSV-1) as well as for varicella-zoster virus (VZV) (Muir et al., 2002; Norberg et al., 2004; Loparev et al., 2004). Interestingly, despite that a limited part of the US region of the HSV-1 genome was sequenced; several recombinants were detected, suggesting that most or all HSV-1 strains are mosaic of recombinants (Norberg et al., 2004). Similar results were presented by Bowden et al. (2004) who suggested that co-infection was the evolutionary source of the high recombination rate found among HSV-1 strains. Further phylogenetic work on whole genome sequences of VZV strains has revealed four instead of three genogroups, and recombination is a prominent feature also for this virus (Norberg et al., 2006). Based on these previous studies, recombination seems to be a major mechanism responsible for viral genetic diversity of human alphaherpesvirinae.

In the present study, our aim was to discern potential genetic differences between African and European genital isolates of HSV-2. Although a limited number of viral strains were obtained from only two distinct geographical regions, and rather small parts of the genome were sequenced and shown to have a low degree of genetic diversity, two genogroups were detected. The groups were arbitrarily designated as A (African) and E (European), for phylogenetic trees based on the gG-2 gene as well as on the gE-2 and gI-2 genes and neighbouring non-coding sequences. The A genogroup consisted of isolates from Tanzania only, why the E genogroup contained strains from Norway and from Tanzania, a finding compatible with a reintroduction of E genogroup strains into Africa, maybe by seafarers and early colonizers.

Detection of three isolates that were classified into different genotypes in the two trees strongly indicates that these strains were derived from recombination events between two other isolates belonging to genotype A and E, respectively. A frequency of three recombinants out of 21 HSV-2 strains investigated, with only a very limited part of the genome investigated, suggest that recombination is a common event also during the evolution of HSV-2.

The possibility of introducing changes in DNA sequences during multiple passages of viral isolates should be considered. In the present study all 21 clinical isolates were passaged four times in cell culture during the process of plaque purification and preparation of viral stocks. However, other investigators (Terhune et al., 1998; Liljeqvist et al., 2000) have presented evidence that multiple passages of HSV through cell culture do not alter the DNA sequences. Although genetic variation introduced by PCR during amplification and sequencing cannot be ruled out, such variations would be expected to be randomly distributed over the genes, which

was not the case. Furthermore, every specified sequence in the present work was derived from identical results in at least two independent reactions. Interestingly, the laboratory strain HG52 harboured several unique sequence alterations as compared to a suggested consensus sequence of all clinical isolates. Of special interest were frameshifts in the gE-2 genes of all clinical isolates so that a sequence of 14 consecutive codons were different from those in the gE-2 gene of HG52. These findings suggest that laboratory strains, which often have been repeatedly passaged, may diverge differently from clinical strains.

Diversity within the gG-2 gene might affect the performances of type-specific tests for HSV-2 antibodies, since such tests are based upon the antigenicities of either the complete protein or epitopes therein. We have previously shown that the human B-cell epitopes on the mature portion of gG-2 are fully conserved (Liljeqvist et al., 1998; 2000). Another epitope is mimicked in the peptide 55 ELISA for detection of HSV-2 antibodies (Oladepo et al., 2000, Nilsen et al, 2003). The only variation observed in the present study within the peptide 55 epitope was a transition from glutamic acid (E) to aspartic acid (D) at position 574 in one Norwegian isolate. The same substitution has been reported in two Swedish isolates, as well as in the laboratory strain 333 (Liljeqvist et al, 2000). Since both E and D are acidic residues differing by one CH_2- group only, one would not expect this variation to have a great impact. However, the significance of this change near the C-terminal end of the epitope has not been investigated further. Nevertheless, it seems reasonable to conclude that the epitope is quite stable among Scandinavian as well as Tanzanian HSV2- isolates, so that the peptide 55 ELISA would be expected to perform well in studies performed both in Europe and Africa. This has been confirmed by analysis of sera from Norwegian adults (Nilsen et al., 2003), and from Tanzanian children and young persons (Kasubi et al. 2006)

In conclusion, the here reported dicothomy in two genogroups of clinical strains of HSV-2, of which one was solely demonstrated in Africa, demonstrate the importance of genetic studies of African isolates of sexually transmitted viruses. Such studies may prove decisive for future attempts of creating effective vaccines against HSV-2 and HIV.

ACKNOWLEDGEMENT

We thank Kjerstin Jakobsen and Anette Roth for skilful technical assistance. Financial support was received from the Western Norway Regional Health Authority and the Swedish International Development Agency and the Swedish Research Council.

REFERENCES

Aurelius, E., Johansson, B., Sköldenberg, B, and Forsgren, M. 1993. Encephalitis in immunocompetent patients due to herpes simplex virus type 1 and 2 as determined by type-specific polymerase chain reaction and antibody assay of cerebrospinal fluid. J. Med. Virol. 39:179-186.

Bowden R, Sakaoka H, Donelly P, Ward R. 2004. High recombination rate in herpes simplex virus type 1 natural populations suggests significant co-infection. Infect Genet Evol. 4,115-123.

Choi S, Seong YR, Lee EK, Chon SK, Yoo W, Lee C-K, Im D-S. 1996. The nucleotide sequence of the glycoprotein E gene of herpes simplex type 2 and its structural chracteristics in comparison wuth the gE of herpes simplex virus type 1. Mol. Cells 6, 145-152.

Cinque, P., Vago, L., Dahl, H., Brytting, M., Terreni, M.R., Fornara, C., Racca, S., Castanga, A., D'Arminio Monforte, A., Wahren, B., Lazzarin, A, and Linde, A. 1996. Polymerase chain reaction on cerebrospinal fluid for diagnosis of virus-associated opportunistic diseases of the central nervous system in HIV-infected patients. AIDS 10, 951-958.

Cohen J. 2005. Prevention cocktails: combining tools to stop HIV spread. Science 309,1002-1005.

Dolan A, Jamieson FE, Cunningham C, Barnett BC, McGeoch D. 1998. The genome sequence of herpes simplex virus type 2. J. Virol. 72, 2010-2021.

Felsenstein, J. 1993. PHYLIP inference package, version 3.5 Department of Genetics, University of Washington, Seattle.

Franti, M., Aubin, J.-T., Poirel, L., Gautheret-Dejean, A., Candotti, D., Huraux, J.-M, and Agut, H. 1998. Definition and distribution analysis of glycoprotein B gene alleles of human herpesvirus 7. J. Virol. 72, 8725-8730.

Kasubi MJ, Nilsen A, Marsden HS, Bergström T, Langeland N, Haarr L. 2005. A branched, synthetic oligopeptide corresponding to a region of glycoprotein G of HSV-1 reacts sensitively and specifically with HSV-1 antibodies in an ELISA. J Virol. Methods125,137-143.

Kasubi MJ, Nilsen A, Marsden HS, Bergström T, Langeland N, Haarr L. 2006. Prevalence of antibodies against herpes simplex viruses type 1 and type 2 in children and young people in an urban region in Tanzania. J Clin Microbiol. 44, 2801-2807.

Liljequist J-Å, Trybala E, Svennerholm B, Jeansson S, Sjögren-Jansson E, Bergström T. 1998. Localization of type-specific epitopes of herpes simplex virus type 2 glycoprotein G by human and mouse antibodies. J Gen. Virol.7 9, 1215-1224.

Liljeqvist JÅ, Svennerholm B, Bergström T. Conservation of type-specific B-cell epitopes of glycoprotein G in clinical herpes simplex virus type 2 isolates. 2000 J Clin Microbiol 38, 4517-4522.

Loparev VN, Gonzalez A, Deleon-Carnes M, Tipples G, Fickenscher H, Torfason EG, Schmid DS. 2004. Global identification of three major genotypes of varicella-zoster virus: longitudinal clustering and strategies for genotyping. J. Virol. 78:8349-8358.

Marsden HS, MacAulay K, Murray J, Smith IW. 1998. Identification of an immunodominant sequential epitope in glycoprotein G of herpes simplex type 2 that is useful for serotype-specific diagnosis. J. Med. Virol. 56, 79-84.

McGeoch DJ, Moss HWM, McNab D, Frame MC. 1987. DNA sequences and genetic conten of the *Hind* III *l* region in the short unique componenet of the herpes simplex virus type 2

genome: identification of the gene encoding glycoprotein G, and evolutionary comparisons. J. Gen. Virol. 68, 19-38.

Meng YX, Spira TJ, Bhat GJ, Birch CJ, Druce JD, Edlin BR, Edwards R, Gunthel C, Newton R, Stamey FR, Wood C, Pellett PE. 1999. Individuals from North America, Australasia and Africa are infected with different genotypes of human herpesvirus 8. Virology 261, 106-119.

Muir WB, Nichols R, Breuer I. 2002. Phylogenetic analysis of varicella zoster virus: evidence of intercontinental spread of genotypes and recombination. J. Virol. 76, 1971-1979.

Nilsen A, Ulvestad E, Marsden H, Langeland N, Myrmel H, Matre R, Haarr L. 2003. Performance characteristics of a glycoprotein G based oligopeptide (peptide 55) and two different methods using the complete glycoproteimn as assays for detection of anit-HSV-2 antibodies in human sera. J. Virol. Methods 107, 21-27.

Norberg P, Bergstrom T, Rekabdar E, Lindh M, Liljeqvist J-A. Phylogenetic analysis of clinical herpes simplex virus type 1 isolates identified three genetic groups and recombinant viruses. J Virol 2004;78:10755-19764.

Norberg P, Liljeqvist J-A, Bergström T, Sammona S, Schmid DS, Loparev VN. Complete-genome phylogenetic approach to varicella-zoster virus evolution: genetic divergence and evidence for recombination. J Virol 2006; 80:9569-9576.

Oladepo DK, Klapper PE, Marsden HS. 2000. Peptide based enzymelinked immunoassays for detection of anti-HSV-2 IgG in human sera. J. Virol. Methods 87, 63-70.

Ramjee G et al. The impact on incident and prevalent herpes simplex virus–2 infection among commercial sexworkers in South Africa. J Aquir Immune Defic Syndr 2005;39:333-9.

Shepp DH, Match M, Lipson SM, Pergolizzi RG. 1998. A fifth human cytomegalovirus glycoprotein B genotype. Res. Virol. 148, 109-114.

Staden, R. 1994. The STADEN package. Methods Mol. Biol. 25:9-170.

Terhune SS, Coleman KT, Sekulovich R, Burke RL, Spear PG. 1998. limitedvariability of glycoprotein gene sequences and neutralization targets in herpes simplex type 2 isolates and stability on passage in cell culture. J. Infect. Dis. 178, 8-15.

Tunbäck P, Liljeqvist JÅ, Löwhagen GB, Bergström T. 2000. Glycoprotein G of herpes simplex virus type 1 - identification of type-specific human epitopes. J. Gen. Virol., 81, 1033-1040.

Tunbäck P, Bergström T, Andersson AS, Nordin P, Krantz I, Löwhagen GB. 2003. Prevalence of herpes simplex virus antibodies in childhood and adolescence – a cross-sectional study. Scand. J. Infect. Dis. 35, 498-502.

Tunbäck P, Bergström T, Löwhagen G-B, Hoebeke J, Liljeqvist J-Å. 2005. Type-specific reactivity of anti-glycoprotein G antibodies from herpes simplex virus infected patients is maintained by single or dual type-specific residues. J. Gen. Virol. 86, 247-251.

Umene K, Sakaoka H. 1997. Populations of two eastern countries of Japan and Korea and with a related history share a predominant genotype of herpes simplex virus type 1. Arch. Virol. 142, 1953-1961.

Zweygberg L, Wirgart B, Brytting M, Linde A, Wahren B, Grillner L, 1998. Sequence variation within three important cytomegalovirus gene regions in isolates from four different patient populations. J. Clin. Micorbiol. 36, 3663-3669.

FIGURE LEGENDS

Fig 1. Genetic distances, shown as nucleotide substitutions per site, between all investigated HSV-2 strains.

Fig. 2. The amino acid sequence of the relevant region of the gG-2 gene of HSV-2 HG52 is shown on the top. Letters in bold indicate the epitope regions I and II identified by Liljeqvist et al. (1998)*, and the immunodominant one is underlined. Asterisks mark the sequence which is mimicked in peptide 55 (Marsden et al., 1998)**. Clinical isolates are shown in which the amino cids, as numbered from the first methionine, are substituted.

Fig. 3.

Phylogenetic trees showing the separation of the clinical HSV-2 isolates into two genetic groups designated A (African) and E (European). One tree was based upon the nucleotide sequences of the genes encoding the gI-2 and gE-2 genes, as well as the sequences between these genes. The other tree was derived from the sequences of the gG-2 gene. Numbers indicate bootstrap values.

* There was previously a printing error, stating that letters in bold indicate the immunodominant region identified by Liljeqvist et al. (1998).

** There was previously a printing error, showing wrong positions of asterics. This has now been corrected.

Table 2. Strains included in the present study

Isolates from Norway	Isolates from Tanzania
N1 or Ng no 1 : 21531551	T1 : Tz 20-2032
N2 or Ng no 2 : 21698861	T2 : Tz 25-2557
N3 or Ng no 3 : 21609612	T3 : Tz 34-2737
N4 or Ng no 4 : 21708171	T4 : Tz 41-3041
N5 or Ng no 5 : 21759551	T5 : Tz 42-1855
N6 or Ng no 6 : 21663631	T6 : Tz 53-4587
N7 or Ng no 7 : 21822792	T7 : Tz 54-3111
N8 or Ng no 8 : 21544951	T8 : Tz 66-2357
N9 or Ng no 9 : 21544911	T9 : Tz 67-3034
N10 or Ng no 10 : 2168985	T10 : Tz 9-2302
	T11 : Tz 2211-12

Table 3. Comparison of the gG-2 genes of the isolates with that of HSV-2 HG52

Nucleotide position(s)[a]	Codon	Mutation	Amino acid substitution	Mutated HSV-2 strain(s)
104	AGC	AAC	S→N	N1-10, T1-T6, T8, T9, T11
127	TCC	GCC	S→A	T7, T10
172	CCC	ACC	P→T	N6, N10
274	TTG	CTG	None	N3, N6, N7, N10, T7, T10
329	CGT	CAT	R→H	N3, N6, N7, N10
406, 407	GAG	TCG	E→S	T9
432	GGG	GGC	None	N1, N8, N9
499	CTG	ATG	L→M	T11
611	CCC	CTC	P→L	N1, N2, N8, N9, T1, T3, T5, T9
635	CGT	CAT	R→H	N8, N9
711	TCC	TCA	None	T1, T3, T5 (should it be T7 instead of T3?)
729	TTT	TTG	F→L	N6
872	GAA	GGA	E→G	N1, T2, T4, T6, T8, T11
877 - 879	GTC	* * *	V-deletion	N3, N6, N7, N10
891	CCG	CCA	None	N5
930	TAC	TAT	None	N1-3, N6-10, T1-11
1013	CGA	CTA	R→L	N5, T2, T4, T6, T8, T11
1045	GAC	AAC	D→N	N2
1048	ACG	GCG	T→A	All strains
1115, 1116	GCA	GTG	A→V	T2, T4, T6, T8, T11
1116	GCA	GCG	None	N1-10, T1, T3, T5, T7, T9, T10
1215	GAG	GAC	E→D	T7, T10
1268	CTC	CCC	L→P	N3, N6, N7, N10, T9
1282 - 1284	GCG	* * *	A-deletion	All strains
1324	TCC	CCC	S→P	N5
1499	TCG	TTG	S→L	N6, N10
1568	GGA	GAA	G→E	N8, N9
1581	ACC	ACT	None	T11
1627	GCG	ACG	A→T	T2, T4, T6, T8, T11
1646	TCG	TTG	S→L	N7
1722	GAG	GAT	E→D	N5
1746	CTC	CTT	None	N2
1758	ACT	ACC	None	N1
1800	ATC	ATT	None	N1
1853	CGC	CAC	R→H	N4
2008	TTG	CTG	None	N4

a; Numbered from the first nucleotide in the initiation codon

Table 4. Comparison of the gI-2 genes of the isolates with that of HSV-2 HG52

Nucleotide position(s)[a]	Codon	Mutation	Amino acid substitution	Mutated HSV-2 strain(s)
29	GCG	GTG	A→V	T2, T6, T8, T11
51	GCC	ACC	A→T	T11
198	CAT	CAC	None	All strains
338	CAC	CGC	H→R	N3, N4
448	TGG	AGG	W→R	N9
476	CGG	CTG	R→L	N5, T9
476, 477	CGG	CTA	R→L	T1-6, T8, T11
503	GCC	GTC	A→V	N3
618	CGA	CGG	None	N5
643	CCC	TCC	P→S	N5, T7, T9, T10
644	CCC	CTC	P→L	T2, T6, T8, T11
716	ATA	AGA	I→R	N3-5, N8, N9, T1-11
716	ATA	AAA	I→K	N2
716, 717	ATA	AGC	I→S	N1, N6, N7, N10
798	CCG	CCT	None	N2
1043	TCG	TTG	S→L	N10
1062	GTG	GTC	None	T5

a: Numbered from the first nucleotide in the initiation codon

Table 5. Comparison of the gE-2 genes of the isolates with that of HSV-2 HG52

Nucleotide position(s)[a]	Codon	Mutation	Amino acid substitution	Mutated HSV-2 strain(s)
59	GCG	GGG	A→G	N3
127	GCG	ACG	A→T	T2, T6, T8
Between 129 and 130		Insert GGGCCGGAG	Insert GPE	All strains
138	ACC	ACT	None	T2, T6, T8, T11
342	TGG	TGT	W→C	T2, T8
384	TAC	TAT	None	N1
392	CTG	CGG	L→R	T1-6, T8, T11
396	GAG	GAT	E→D	N9
531	GAC	GAT	None	N8, N9
Between 541 and 542	ACG	AGCG	Insert, frameshift	All strains
570		GCA	A	N2
570		ACA	T	T1, T3-5
570		CCA	P	N1, N3-10, T2, T6-11
583	CGT	* GT	Delete, framshift	N1-10, T1-8, T10-11
605	CCG	CAG	P→Q	T2, T6, T8, T11
Between 610 and 611		Insert CCCCCCGACGC	Insert PPRR, frameshift	T9
671	CTG	CCG	L→P	All strains
804	GAT	GAG	D→E	All strains
1140	GTG	GTT	None	N3
1146	GAA	GAG	None	N3, N6, N7, N10
1211	CAT	CCT	H→P	N1, N5-7, N10, T9
1245	CTC	CTT	None	N8, N9
1273	CTG	TTG	None	N3
1551	CGT	CGC	None	N1
1621	CCG	TCG	P→S	All strains

a: Numbered from the first nucleotide in the initiation codon

Altered amino acids due to frameshift

Amino acid number	180														195
HSV-2 HG52	V	T	N	A	R	R	A	A	F	P	P	Q	P	P	P	R
All strain except	V	S	E	R	T	P	V	S	V	P	P	P	T	P	P	R

Table 6. Comparison of the of the sequences between the gI-2 and gE-2 genes with shose of HSV-2 HG52

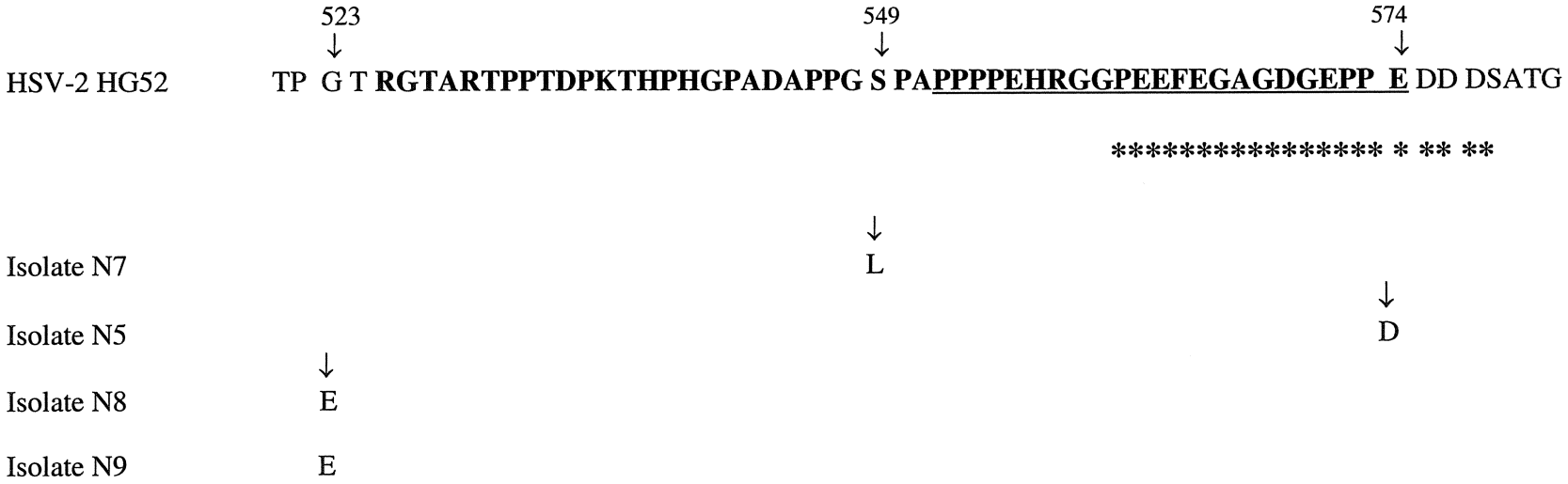
Nucleotide position(s)[a]	Mutation	Mutated HSV-2 satrain(s)
60 - 69	10 nucleotides deleted	N1, N8, N9, T1-6, T8
61 - 69	9 nucleotides deleted	N2-7, N10, T7, T9, T10
Between 178 - 179	Insert, G	T9, T11
Between 178 - 179	Insert, GG	N5, T1-6, T8
Between 178 - 179	Insert, GGG	N7
Between 178 - 179	Insert, GGGG	N4, N6, N10
Between 178 - 179	Insert, GGGGGGGG	N3
308	G→A	T1, T3-5

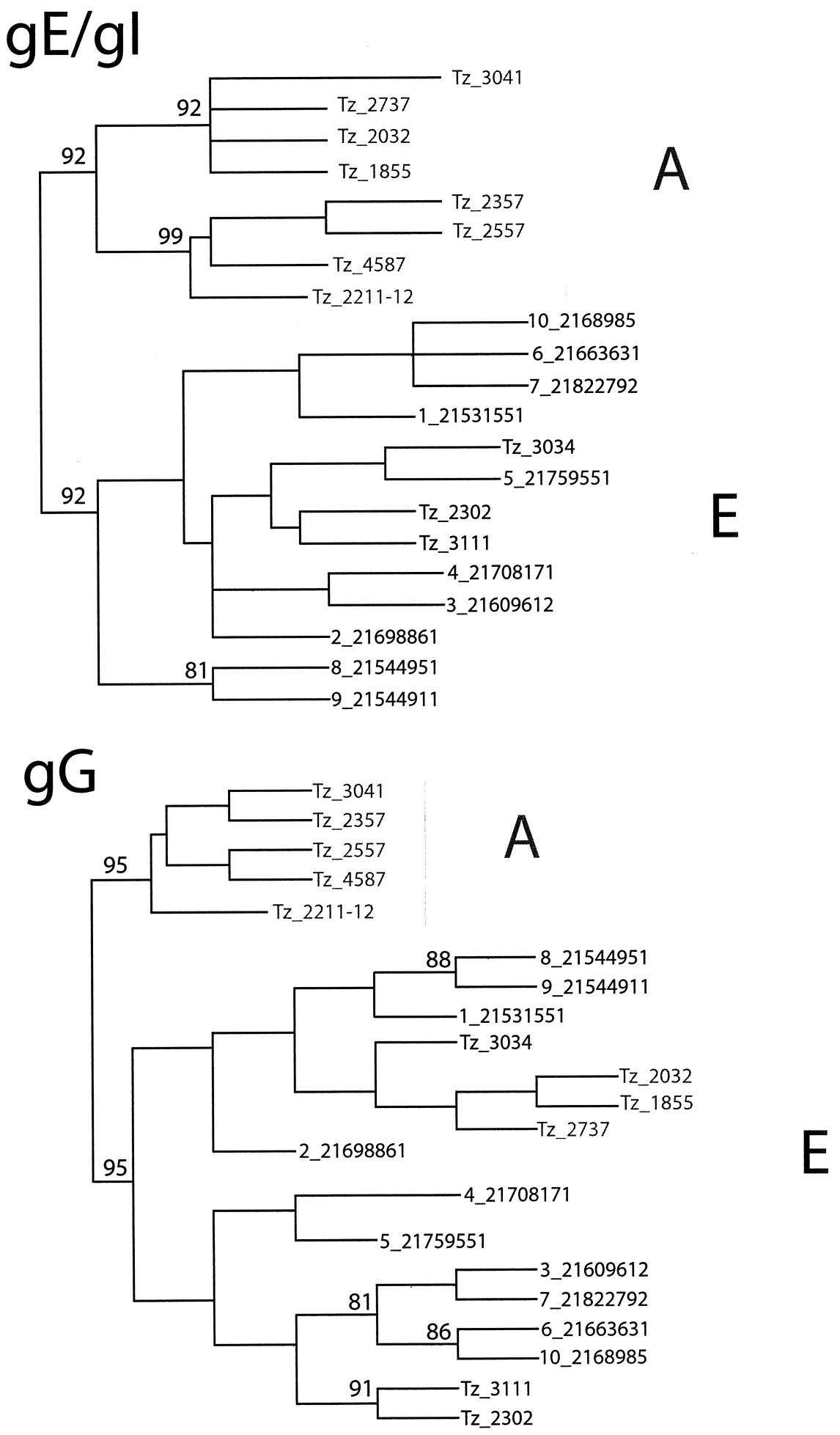
a: Numbered from the first nucleotide after the stop codon of the gI-2 gene.

	NgG2_no.1	NgG2_no.2	NgG2_no.3	NgG2_no.4	NgG2_no.5	NgG2_no6	NgG2_no.7	NgG2_no.8	NgG2_no.9	NgG2_no.10	TZ-2032	TZ-2557	TZ-2737	TZ-3041	TZ-1855	TZ-4587	TZ-3111	TZ-2357	TZ-3034
NgG2_no.1																			
NgG2_no.2	0.002519																		
NG2_no.3	0.003497	0.002913																	
NG2_no.4	0.002519	0.001935	0.002125																
NgG2_no.5	0.003103	0.002913	0.003497	0.002125															
NgG2_no.6	0.002913	0.003103	0.002125	0.002709	0.003293														
NgG_no.7	0.002329	0.002519	0.001555	0.002125	0.002709	0.000971													
NgG2_no.8	0.002125	0.001935	0.002913	0.001935	0.002913	0.003103	0.002519												
NgG2_no.9	0.002519	0.002329	0.003293	0.002329	0.003293	0.003497	0.002913	0.000387											
NgG_no.10	0.002709	0.002913	0.001935	0.002519	0.003103	0.000577	0.000781	0.002913	0.003293										
Tz-2032	0.002709	0.001935	0.003103	0.002125	0.002709	0.003293	0.002709	0.002125	0.002519	0.003103									
Tz-2557	0.003877	0.003687	0.004271	0.003293	0.003483	0.004461	0.003877	0.003687	0.004067	0.004271	0.002709								
Tz-2737	0.002709	0.001935	0.003103	0.002125	0.002709	0.003293	0.002709	0.002125	0.002519	0.003103	0.000007	0.002709							
Tz-3041	0.003103	0.002709	0.003497	0.002519	0.002709	0.003687	0.003103	0.002913	0.003293	0.003497	0.001161	0.001555	0.001161						
Tz-1855	0.002913	0.002125	0.003293	0.002329	0.002913	0.003497	0.002913	0.002329	0.002709	0.003293	0.000197	0.002913	0.000197	0.001351					
Tz-4587	0.003687	0.003483	0.004067	0.003103	0.003293	0.004271	0.003687	0.003483	0.003877	0.004067	0.002519	0.000197	0.002519	0.001351	0.002709				
Tz-3111	0.002709	0.002125	0.002329	0.001745	0.002329	0.002519	0.001935	0.002125	0.002519	0.002329	0.002329	0.003497	0.002329	0.002709	0.002519	0.003293			
Tz-2357	0.003877	0.003687	0.004271	0.003293	0.003497	0.004461	0.003877	0.003687	0.004081	0.004271	0.002709	0.000007	0.002709	0.001555	0.002913	0.000197	0.003497		
Tz-3034	0.002329	0.002125	0.002709	0.002125	0.001935	0.002519	0.001935	0.002125	0.002519	0.002329	0.001935	0.003497	0.001935	0.002709	0.002125	0.003293	0.001935	0.003497	
Tz-2302	0.002709	0.002125	0.002329	0.001745	0.002329	0.002519	0.001935	0.002125	0.002519	0.002329	0.002329	0.003497	0.002329	0.002709	0.002519	0.003293	0.000007	0.003497	0.001935

Fig 1

Fig. 2. Immunodominant regions in the gG-2 protein




gE/gI
Tz_3041
92
Tz_2737
Tz_2032
92
Tz_1855
A
Tz_2357
Tz_2557
99
Tz_4587
Tz_2211-12
10_2168985
6_21663631
7_21822792
1_21531551
Tz_3034
5_21759551
92
Tz_2302
E
Tz_3111
4_21708171
3_21609612
2_21698861
81
8_21544951
9_21544911
gG
Tz_3041
Tz_2357
A
Tz_2557
95
Tz_4587
Tz_2211-12
88
8_21544951
9_21544911
1_21531551
Tz_3034
Tz_2032
Tz_1855
Tz_2737
E
2_21698861
95
4_21708171
5_21759551
3_21609612
81
7_21822792
86
6_21663631
10_2168985
91
Tz_3111
Tz_2302