MOLECULAR VARIATION, EVOLUTION AND GEOGRAPHICAL DISTRIBUTION OF LOUPING ILL VIRUS

G.F. GAO[1*], P.M. DE A. ZANOTTO[1], E.C. HOLMES[2], H.W. REID[3], E.A. GOULD[1]

[1]Institute of Virology and Environmental Microbiology, Mansfield Road, Oxford OX1 3SR; [2]Wellcome Centre for Epidemiology, Institute of Infectious Disease, Department of Zoology, University of Oxford, Oxford; [3]Moredun Research Institute, Edinburgh, UK

Received August 11, 1997; revised August 18, 1997

Summary. – Following the demonstration that the tick-borne encephalitis (TBE) subgroup viruses are distributed as a cline across the Northern Hemisphere (Zanotto *et al.*, 1995), we have analysed the dispersal pattern of louping ill (LI) virus, the most westerly located member in the cline. A total number of 21 LI or LI-related virus E gene sequences have been used for a detailed molecular analysis of the evolution, phylogeny and geographical distribution of LI virus in the British Isles and Ireland. The results show that LI virus is genetically stable in general but minor differences enable its separation into four genetically distinct subtypes (genotypes) with clear geographical correlation, designated Type 1 in Scotland and England, Type 2 in Scotland, Type 3 in Wales and Type 4 in Ireland. These data demonstrate that geographically independent evolution of LI viruses has occurred. The molecular systematics and substitutional parameters analyses combined with the clinal distribution of the TBE virus complex allow the assignment of the origin for both Negishi (NEG) virus and a Norwegian isolate to the British Isles. Moreover, proposals for the classification of LI and LI-like viruses which cause encephalomyelitis in sheep, goat or cattle are presented.

Key words: louping ill virus; molecular epidemiology; evolution

Introduction

LI virus, a tick-transmitted member of the TBE virus group, genus *Flavivirus,* family *Flaviviridae* (Francki *et al.*, 1991), is the only arthropod-transmitted virus in the British Isles and Ireland with a recognised disease in vertebrates. It causes encephalitis or encephalomyelitis in sheep and red grouse (*Lagopus scoticus*), and occasionally affects other domestic animals and humans. Other wildlife species can also become infected with LI virus if they serve as hosts for the feeding of tick vector *Ixodes ricinus*, but these vertebrates may be dead-end hosts for the virus because they do not develop high-titer viraemia (for review, see Reid, 1988). LI virus is prevalent in the upland sheep grazing areas of Scotland, northern England, Wales, southwest England and Ireland, and on the basis of plaque morphology, virus neutralization (NT) and haemagglutination-inhibition (HAI) tests, these viruses have been described as indistinguishable (Reid, 1984).

LI is conventionally considered to be restricted to the British Isles (Reid, 1988) although a sheep or goat disease caused by LI-like viruses was reported also in Bulgaria (Pavlov, 1968), Turkey (Hartley *et al.*, 1969), Greece (Papadopoulos *et al.*, 1971), Spain (Gonzalez *et al.*, 1987) and Norway (Reid, 1988). Our previous studies (Venugopal *et al.*, 1992; Gao *et al.*, 1993a, 1993b; Marin *et al.*, 1995a) and those of others (Whitby *et al.*, 1993) using nucleotide (nt) sequenc-

*Present address: Molecular Immunology Group, Institute of Molecular Medicine, University of Oxford, Headington, Oxford OX3 9DS, UK

Abbreviations: aa = amino acid; HA = haemagglutination; HAI = HA-inhibition; HIV = human immunodeficiency virus; LI = louping ill; NEG = Negishi; nt = nucleotide; NT = neutralization; RT-PCR = reverse transcription/polymerase chain reaction; SMB = suckling mouse brain; SSE = Spanish sheep encephalomyelitis; TBE = tick-borne encephalitis; ts:tv = transitions:transversions; TSE = Turkish sheep encephalomyelitis

ing and immunological comparisons have shown that viruses which are antigenically very similar to LI virus and were isolated from sheep or goats in Norway, Turkey, Greece and Spain are genetically different, though closely related, from LI virus, and two new virus names have been proposed to recognise these differences, i.e., Turkish sheep encephalomyelitis (TSE) virus in Turkey and Greece (Gao *et al.*, 1993a; Marin *et al.*, 1995a) and Spanish sheep encephalomyelitis (SSE) virus in Spain (Marin *et al.*, 1995a). It has also been proposed that the Norwegian isolate (Gao *et al.*, 1993b) and NEG virus (Venugopal *et al.*, 1992), a tick-transmitted virus isolated from Japan, are LI viruses. These unexpected results have raised questions concerning the identity, geographical distribution and evolution of LI virus which have been addressed in this communication by carrying out an extensive molecular analysis of LI and closely related viruses.

In general, a clear understanding of the genetic relatedness among different strains (isolates, varieties) of a virus and of their evolution can provide information on the origins, dispersal patterns and geographical distribution of particular genotypes (sometimes phenotypes as well), on their routes of transmission and for the development of vaccines. Therefore such an analysis of LI virus would enormously help solve the LI and LI-like virus puzzle as successfully used in influenza virus, human immunodeficiency virus (HIV) and more recently in hepatitis C virus (for review, see Leigh Brown, 1994).

In common with all other flaviviruses, LI virus contains a single-stranded RNA genome of positive polarity which encodes three structural proteins viz. capsid (C), membrane (M) and envelope (E) (Shiu *et al.*, 1991) and seven non-structural (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5) proteins (K. Venugopal, personal communication). Among the structural proteins, the E glycoprotein is responsible for the following biological functions: haemagglutination (HA), NT, receptor binding, neuropathogenicity, membrane fusion and induction of protective antibody responses (for reviews, see Heinz, 1986; Gould *et al.*, 1990). Since the flaviviruses have been grouped mainly on the basis of HA and NT tests, their classification largely reflects properties of the E proteins (Calisher *et al.*, 1989). Moreover, the E gene of flaviviruses was also shown to be a reliable phylogenetic marker for all flaviviruses since its phylogeny is congruent to that of the RNA-dependent RNA polymerase (NS5) which was successfully used for phylogenetic analysis (Koonin, 1991; Marin *et al.*, 1995b). We have cloned and sequenced the E genes of six additional LI viruses including an isolate from Wales and one from Ireland that cause sheep or cattle encephalomyelitis, respectively, and although the sequence data have been used in our previous report (Zanotto *et al.*, 1995), the detailed methods of sequencing and phylogenetic analysis of LI viruses were not addressed previously. Therefore, in this report, we describe the methods of sequencing for the six known LI viruses and detailed phylogenetic analysis by addressing LI viruses (totally 12 LI virus E genes are included) in Britain and Ireland which are the known LI virus epidemic regions. Our results show that genetic variation occurs in LI viruses and the strain of LI virus isolated from Ireland is genetically distinct from British mainland LI viruses. In evolutionary terms, the Irish virus is positioned between western (European) TBE (WTBE) and British LI viruses. We propose that there are four genotypes of LI virus in the British Isles and Ireland, each of which occupies a distinct ecological niche.

Materials and Methods

Viruses. The original isolates were identified as LI virus at the Moredun Research Institute, Edinburgh, using conventional virological and immunological methods. The details of the six LI viruses sequenced in this and Zanotto's (Zanotto *et al.*, 1995) studies and another six sequenced LI viruses are listed in Table 1. All the virus isolates have received less than 3 serial passages in suckling mouse brain (SMB) since their isolation from field samples. All other viruses and their sequence accession numbers used for comparison were described by Marin *et al.* (1995b). Other viruses compared in this study are as follows: SSE virus;, TSE virus; NEG virus; WTBE virus, strain Neudorfl; far-eastern subtype of TBE (FETBE) virus, strain Sofjin; Omsk haemorrhagic fever (OHF) virus; Langat (LGT) virus; Kyasanur forest disease (KFD) virus; Powassan (POW) virus.

RT-PCR, cloning and nucleotide sequencing. Ten percent SMB suspensions of plaque-purified viruses were used directly for the extraction of viral RNA followed by reverse transcription and polymerase chain reaction (RT-PCR) as previously described (Gao *et al.*, 1993b; Gao, 1995). Briefly, a SMB suspension was treated with proteinase K and extracted with phenol and ether. First-strand cDNA was synthesised using the conserved downstream primer 5'-CGGGTAGTATGCATAGTT-3', complementary to LI virus genome nt 2447-2449 (Shiu *et al.*, 1991) and the E gene was subsequently amplified by standard PCR (*Taq* polymerase and annealing temperature of 50°C for 30 cycles) using the upstream primer 5'-CTGGAGAGTGTGGTGAC-3', (nt 805-821 of the LI virus genome) and the same downstream primer as was used in the synthesis of the first-strand cDNA. The PCR product was directly cloned into the pGEM-T vector by following the product specifications (Promega) and sequenced in both directions by double-stranded dideoxynucleotide sequencing (Sanger *et al.*, 1977) using USB Sequenase® Version 2 Kit. Two batches of PCR product for each virus were cloned and at least three transformed subclones from each virus were sequenced to overcome any problems arising from PCR errors. All other molecular manipulations followed the standard methods (Sambrook *et al.*, 1989).

Sequence data analysis. Unless otherwise stated below, the nt and amino acid (aa) sequence analyses were performed using

Table 1. LI viruses sequenced and analysed in this study

Virus designation	Geographical origin	Year of isolation	Host of isolation	Sequence Accession No.[c] source
LI/31	Scotland	1931	Sheep	Gao *et al.*,1993 D12937
LI/369	Ayrshire, Scotland	1963	Tick[a]	Shiu *et al.*,1991 M37687
LI/SB526	Oban, Scotland	1968	Sheep	Venugopal *et al.*, 1992 M94957
LI/MA54	Ireland	1968	Cattle	This paper X86784
LI/G	Mull, Scotland	1979	Pig	This paper X86788
LI/M	Scotland	ND[b]	ND[b]	This paper X71872
LI/K	Grampian, Scotland	1980	Grouse	Gao *et al.*,1993 D12935
LI/I	Wales	1980	Sheep	This paper X86785
LI/A	Devon, England	1980	Sheep	Gao *et al.*,1994 X69975
LI/NOR	Norway	1984	Sheep	Gao *et al.*,1993 D12936
LI/917	Penrith, England	1985	Sheep	This paper X86786
LI/261	Newcastle, England	1987	Sheep	This paper X86787

[a]*Ixodes ricinus*. [b]No data available. [c]GenBank database accession No.

CLUSTAL programme, Staden programme and the University of Wisconsin Genetics Computer Group (GCG) package.

Molecular systematics of the virus E genes. To determine optimal parameters for the molecular systematic analysis and the genetic structure of the viruses, several pairwise comparisons for substitutional parameters, e.g. synonymous, non-synonymous, transitions and transversions (Li and Graur, 1990; Nei, 1987) were plotted using DIPLOMO programme (Weiller and Gibbs, 1993). To determine an adequate transitions:transversions (ts:tv) ratio for the E gene dataset, estimations of the highest value for the natural logarithm of the maximum likelihood were determined using different values of ts:tv (1, 2, 4, 5, 7 and 10) (Felsenstein, 1993). With the optimal ts:tv, the likelihood values obtained for the best trees under equal and unequal assumption rates were used to determine evolutionary rate variation via a χ^2 test (Felsenstein, 1993).

In addition to maximum likelihood, the phylogeny of the E genes was determined using distance- and character-based methods with nucleic acids and protein primary sequence data using PAUP (Swofford, 1990), MacClade (Maddison and Maddison, 1992), PHYLIP (Felsenstein, 1993) and MEGA programmes (Kumar *et al.*, 1993). The level of support from the data to the phylogenetic estimations was determined using the bootstrap methods.

Results

Genetic variation of LI viruses isolated from the British Isles and Ireland

The E genes of six serologically recognised LI viruses (Table 1) were sequenced and the sequences were deposited in the GenBank. The viruses were selected to represent different geographical regions in the British Isles and Ireland. The sequence analysis showed that all LI virus E glycoproteins contained 496 aa (1488 nt). No insertion or deletion was detected, even among the TBE subgroup viruses except for POW virus which had an additional codon initiating at position 1006 generating a 1491 nt long data set. The positions of the twelve conserved flavivirus cysteine residues were identical and the three potential N-linked glycosylation sites (one N-X-T, two N-P-Ts) were also conserved among LI viruses. These results demonstrate

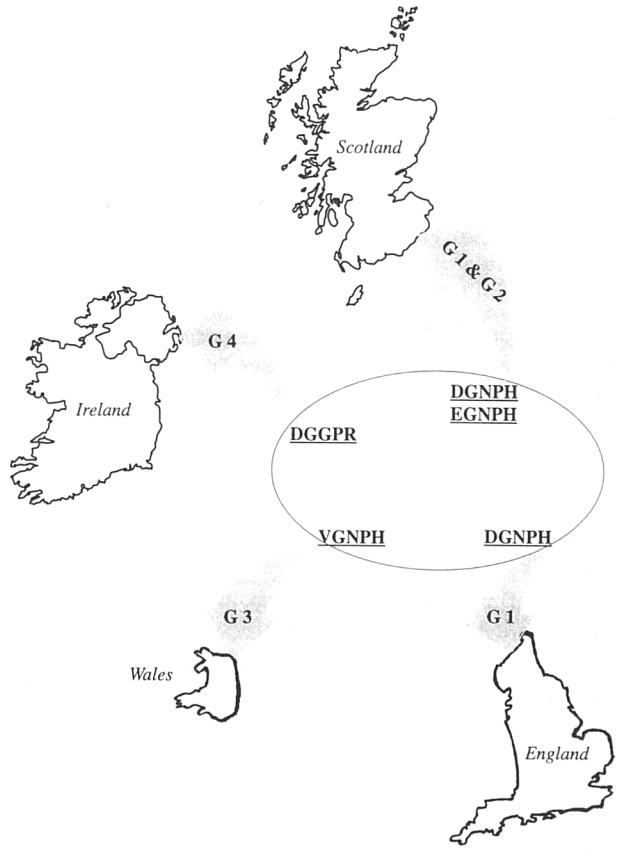

Fig. 1
Schematic representation of LI virus genotypes, G1-G4, together with the appropriate genetic marker motifs

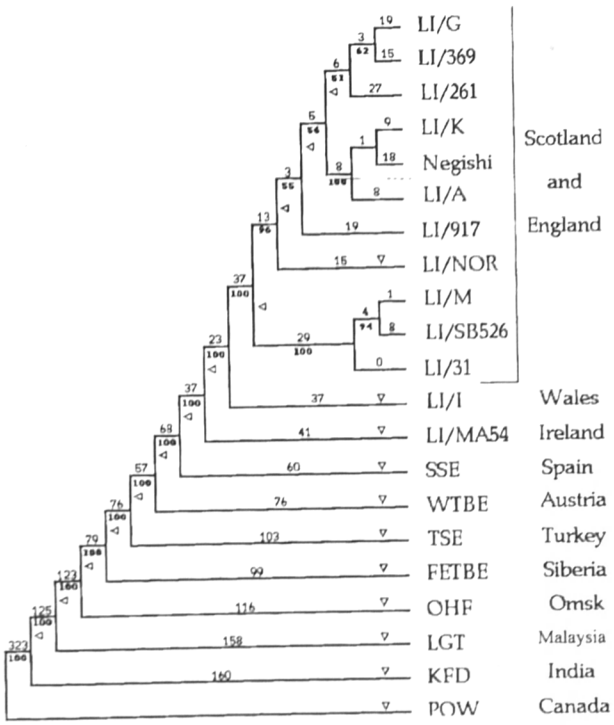

Fig. 2
Phylogenetic tree illustrating the evolutionary relationships of LI viruses and other viruses in the TBE virus complex
The tree was obtained by maximum likelihood, parsimony analysis (PAUP, MEGA, PHYLIP) and distance methods (neighbouring-joining and Fitch-Margaliash). Branch lengths are the number of most parsimonious reconstructed changes along the branches and are shown above the branches. Bootstrap percent values are shown below the branches. The position where NEG virus was located in the tree for topology tests is shown by triangle.

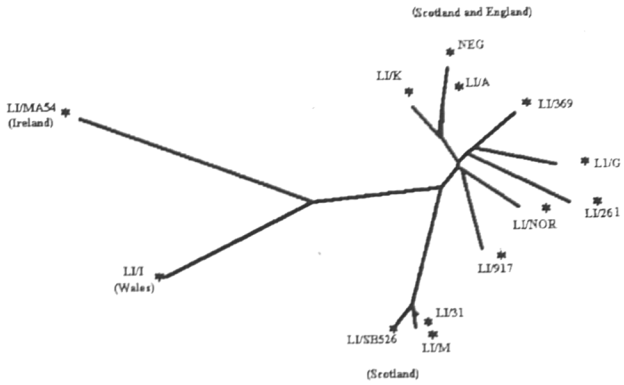

Fig. 3
Unrooted tree (using J-C distance) for LI viruses shows the genotype clusters of the isolates

the structural stability of LI viruses, at least in terms of the E gene.

A specific analysis of the aa sequences showed that both the recognised hexapeptide and pentapeptide genetic marker motifs, EHLPTA (aa 207-212 in the LI virus E gene) and DSGHD (aa 320-324 in the LI virus E gene) which identify TBE subgroup viruses (Shiu *et al.*, 1991; Gao *et al.*, 1993a), and the LI-specific tripeptide genetic marker motif, NPH (aa 232-234 in LI virus E gene) (Shiu *et al.*, 1992), were totally conserved in all the sequenced LI viruses except the Irish isolate (LI/MA54) which had a unique tripeptide genetic marker (GPR) at this position, implying the possibility of a distinct identity for this virus as compared with other LI viruses. This is also supported by the evidence that the percentage identity in aa of LI virus E genes with the tripeptide marker motif NPH was at least 96.7% while for LI/MA54 virus the percentage identity with the other LI viruses ranged from 95.4% to 96.4%, implying a direct correlation between the genetic marker motif and the genetic difference of this virus from other LI viruses. Interestingly, the highest variation among the British mainland LI viruses was observed between LI/I, the Welsh isolate, and the English and Scottish isolates, ranging from 96.7 to 98.2% aa identity. All English and Scottish isolates showed 98.0 – 99.8% identity. Moreover, when the tripeptide genetic marker motif was extended to a pentapeptide (aa 230-234) as originally proposed by Shiu *et al.* (1992), the mainland LI viruses could then be divided into three types, each with a unique aa in the first position of the pentapeptide motif (Fig. 1), producing the genetic marker motifs *D*GNPH for Scottish and English isolates, *E*GNPH for other Scottish isolates or *V*GNPH for the Welsh isolate. When the viruses were subdivided on the basis of these pentapeptide motifs, the percentage identity within each type was over 98.0%. Therefore, using the 98.0% aa identity and the unique genetic marker motif described above as the cut-off point, four LI virus genotypes were identified in the British Isles and Ireland. Genotype 1 included isolates from Scotland and England (LI/369, LI/K, LI/A, LI/G, LI/261, LI/917) with the genetic marker motif DGNPH. Genotype 2 with the genetic marker motif EGNPH included only Scottish isolates, LI/31, LI/M and LI/SB526. The Welsh isolate LI/I was designated as genotype 3 with the genetic marker motif VGNPH. The genetically most distinct isolate LI/MA54 from Ireland was designated as genotype 4. The validity of the pentapeptide motifs for the classification of genotypes was confirmed by phylogenetic tree reconstructed with parsimony and maximum likelihood methods (Fig. 2) and by distance-based unrooted tree (Fig. 3).

Molecular systematics of LI virus and its relationships to other members of the TBE subgroup viruses

Pairwise comparisons of the substitutional parameters of the E genes of LI viruses were done. In addition, other members in the TBE subgroup for which the E gene sequences

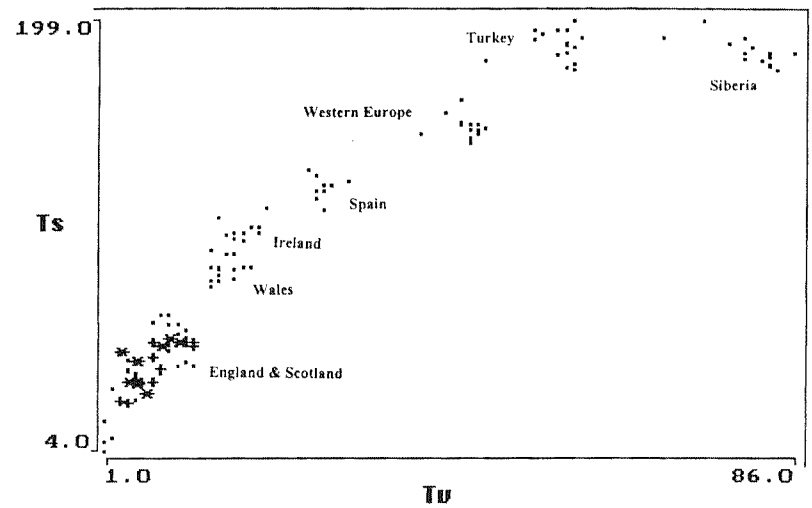

Fig. 4

Pairwise comparisons of value differences of the transitions versus transversions

They indicate that both types of substitutions accumulate gradually and show that NEG virus (*) and the Norwegian LI isolate (+) cluster within the English and Scottish LI virus isolates.

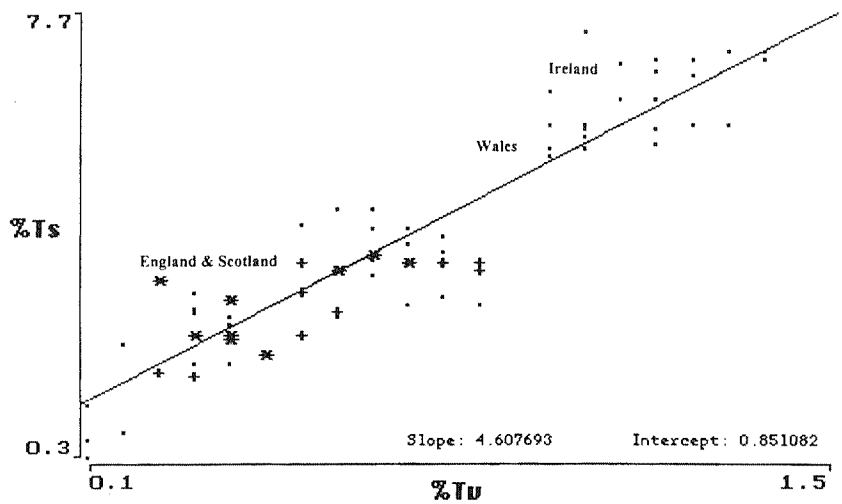

Fig. 5

Determination of the transition to transversion ratios for LI viruses

Maximum likelihood curve determined by analysis of sequences of LI viruses and other TBE subgroup viruses with the DNAML programme (Felsenstein, 1993), indicating that the maximum values of ts:tv are between 4 and 5.

were available, SSE, TSE, WTBE, FETBE, OHF, KFD, LGT and POW viruses, were included. For these viruses the sequence data were obtained from GenBank. The POW virus E gene was chosen as the outgroup for the phylogenetic analysis based on previous determinations of the TBE virus complex root in relation to its sister group including the TYU and SRE tick-borne viruses (Marin *et al.*, 1995b). The ts:tv values (G or A to C or T, and C or T to G or A) (Fig. 4) indicated the lack of significant nt saturation when only LI viruses or when LI, SSE, TSE and WTBE viruses were compared as was also shown by the linear relationship between the percentage of transition over transversion (Fig. 5). The values for synonymous (without aa changes) to non-synonymous (with aa changes) differences between the viruses showed a linearly increasing number of synonymous differences (Fig. 6). This indicates the absence of the saturation in the TBE virus population in Europe. Nevertheless, the saturation was detectable when comparing LI virus with

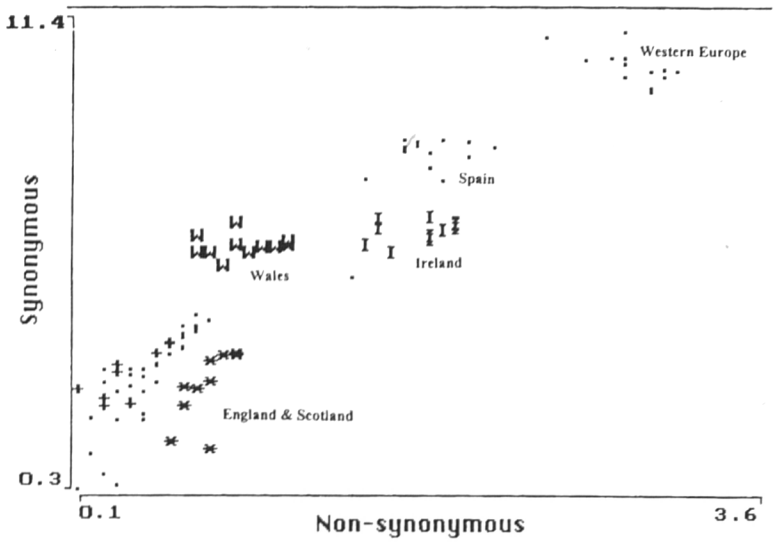

Fig. 6

Pairwise comparisons of synonymous versus non-synonymous differences

They show a similar pattern to the ts:tv difference plot but reveal that the Welsh isolate, though with similar values for synonymous differences, has accumulated fewer aa changes in relation to the English and Scottish isolates than to the Irish isolate; moreover, NEG virus shows more aa changes than all LI virus isolates (Scottish, English, Welsh, Irish and Norwegian) with fewer nt changes. The higher values of transitions over transversions and synonymous over non-synonymous substitutions and gradual accumulation of changes are indicative of a genetic continuum in the TBE virus complex.

FETBE virus (Fig. 4). In Fig. 6, it can also be seen that the ratio of synonymous to non-synonymous substitutions is higher with the Welsh isolate than with the other viruses. This may reflect the separate origin and evolutionary pattern of LI/I virus. Therefore, the Welsh isolate appears to be confirmed as a separate genotype which evolved in a distinct ecological niche. The maximum likelihood curve (Fig. 7) pointed to a maximum between 4 and 5 and the regression of ts on tv had a slope of 4.61 which was also shown by maximum likelihood analysis to be close to the maximum value. Based on the determination of substitutional parameters, the optimal ratio was therefore assumed to be 4.61 for LI virus. For the parsimony analysis, this value was rounded to 5 in a nucleotide inter-conversion stepmatrix (Swofford, 1993). Distance-based maximum parsimony and maximum likelihood analyses all produced data confirming that the viruses converge on a single highly asymmetrical phylogenetic tree (Fig. 2).

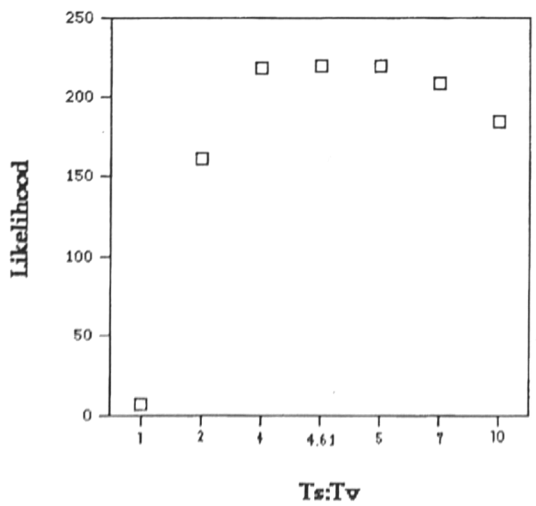

Fig. 7

Determination of the ts:tv rate

Linear regression ($r^2 = 0.98$) of percentage transition and percentage transversion differences indicates a slope of 4.61. This value approaches the maximum determined by the above maximum likelihood analysis.

Norwegian LI-like virus and NEG virus are LI virus

The evidence presented above has also enabled conclusions to be drawn concerning the phylogenetic status of LI/NOR and NEG viruses. The percentage identity/genetic marker motif and the systematic molecular analysis demonstrated that the Norwegian isolate and the Japanese NEG virus are strains of LI virus as previously reported (Venugopal *et al.*, 1992; Gao *et al.*, 1993b). The phylogenetic trees (Figs. 2 and 3) show that NEG, LI/K and LI/A viruses cluster as a trichotomy in the consensus tree (bootstrap value 100%), implying that NEG virus originated from Great Britain. The evolutionary invariant results show no significant statistical support for pairs formed by either NEG

virus or the Norwegian isolate with continental TBE other than SSE or LI virus and the results confirm the pairwise substitution comparisons and global phylogenetic analyses. The origin of the Norwegian isolate from LI virus in the British Isles is supported by the clinal distribution of the TBE virus complex (Zanotto *et al.*, 1995) and could be explained by conventional dispersal mechanisms. However, the origin of NEG virus is more controversial since in the presence of a geographically structured genetic continuum, NEG virus should be more closely related to FETBE isolates. Therefore the NEG virus origin was further investigated by global phylogenetic statistical analysis on rival trees implying different possible origins of the isolate. The statistical results from the maximum likelihood analysis for tree topology comparisons with NEG virus branching from the points indicated by triangles in Fig. 2 show that placing NEG virus outside the (LI/K, LI/A) group diminishes the likelihood drastically causing trees to differ by at least 18.4 S.D. from the best tree (Fig. 2), by 30 S.D. when it is placed in WTBE group, and the figure increases to 45.7 S.D. when placed in FETBE group. These results unequivocally reject the hypothesis that NEG virus originated outside the British Isles. Both NEG and LI/NOR viruses were included in genotype 1 and therefore have the genetic marker motif DGNPH and 98% identity compared with other members of the genotype.

Discussion

All LI viruses have been considered an identical pathogenetic entity for almost a century, at least in the British Isles and Ireland, and the disease louping ill has generally been considered to have been caused by one virus species. The work described here together with our previous observations (Venugopal *et al.*, 1992; Gao *et al.*, 1993a; 1993b; Marin *et al.*, 1995a) has shown that a complex of closely related but distinct viruses may cause louping ill and the disease is not confined to the British Isles. In this extended analysis, nt and aa sequence alignments of the LI virus E protein have revealed characteristic sequence similarities for LI viruses isolated from distinct geographical localities. By comparing genetic marker motifs that have been shown to be a reliable phylogenetic indicator for other flaviviruses (Shiu *et al.*, 1992), the viruses could be subdivided into 4 distinct genotypes, viz. genotype 1, isolates from England and Scotland with the genetic marker motif DGNPH; genotype 2, isolates from Scotland with the genetic marker motif EGNPH; genotype 3, isolates from Wales with the genetic marker motif VGNPH; genotype 4, isolates from Ireland with the genetic marker motif DGGPR. The subdivision of these viruses in this way is in agreement with the results obtained by systematic phylogenetic analysis. The viruses were grouped into four genotypes with bootstrap values near (96%) or equal to 100%. This does not conflict with our previous observations in which the TBE subgroup viruses were shown to form a cline (Zanotto *et al.*, 1995). Indeed, the results demonstrate that individual species have diverged along corridors and then evolved in distinct niches.

It is interesting to note the significant difference in the pentapeptide genetic marker motifs between LI virus from Ireland and the British LI virus isolates. The precise function of the domain that encodes the pentapeptide has not yet been identified. As the crystal structure of TBE virus E protein showed that these amino acids are likely to lie on the virus surface (Rey *et al.*, 1995), it is tempting to speculate that this domain may have an important role in the pathogenicity especially with regard to the findings that the Irish LI virus was more virulent for mice and produced larger plaques (4-5 mm versus 2-4 mm) than the British LI viruses (Gao *et al.*, 1994, and unpublished data). It may be important to note that Irish LI virus is the only one of those sequenced to have been isolated from cattle. Whether or not this is significant in the selection of variant virus strains remains to be determined. These data support the conclusion that the Irish isolate is distinct from the other mainland British LI viruses. Therefore, genotype 4 should be considered an Irish topotype. Partial E gene sequences covering the genetic maker motif regions of some more Irish isolates have confirmed the distinct characteristic of Irish LI virus isolates (McGuire, personal communication). These data lead one to ask: what is a suitable cut-off point for the differentiation of LI virus and other members of the TBE subgroup? We previously proposed the inclusion of two new members (TSE virus and SSE virus) in the TBE subgroup (Gao *et al.*, 1993b; Marin *et al.*, 1995a) but the genetic distance of British LI virus from SSE virus is comparable with that between British LI and the Irish topotype. In addition to the suggestion of a topotype or genotype for the Irish LI virus, there are at least two other optional ways in which LI and so-called LI-like viruses could be differentiated. Firstly, the Irish topotype could be defined as a new member of the TBE virus subgroup in the same way as SSE and TSE viruses were defined. This would lead to the classification of three genotypes of LI virus in Great Britain and three new members of the TBE subgroup, viz. SSE virus in Spain, TSE virus in Turkey and Greece, and Irish virus. Secondly, TSE and SSE viruses could also be considered as topotypes of LI/TBE virus. This would mean that there are four topotypes of LI virus, viz. the British LI virus topotype (in which three genotypes were recognised), Irish LI virus topotype, Spanish LI virus topotype and Turkish/Greek LI virus topotype. This second alternative is probably less satisfactory since the aa similarity/identity data would appear "blurred" because in this second system there would be an

overlap between TSE virus and TBE viruses, i.e., TSE virus is not justified as a topotype of LI virus because it is closer to WTBE and FETBE viruses than to LI virus (Gao *et al.*, 1993a; Whitby *et al.*, 1993). Moreover, the Greek strain of TSE virus (GGE virus or Vergina strain, see Marin *et al.*, 1995a) was once defined as type III TBE virus (WTBE and FETBE as type I and II) by using the traditional serological/pathogenetic methods (Rubin *et al.*, 1975; Papodoupoulos, 1980) and confirmed later by DNA hybridisation by Pogodina *et al.* (1993), which indicates the distinct identity of TSE virus but has not yet been formally accepted by all flavivirologists. The data at this stage seem to support the first of the two options, i.e. LI virus primarily causes LI in Great Britain, and other viruses (Irish, SSE and TSE viruses) cause animal encephalomyelitis in Ireland, Spain, Turkey and Greece. More sequence data will be required before a final decision can be taken, especially the Bulgarian sheep LI-like isolate (Pavlov, 1968) which is the only unsequenced virus of known LI-like viruses in this time.

The principles of comparison by aa alignment and also by molecular systematic analysis have also been applied to demonstrate the origins of two LI-related viruses, which were reported to have been isolated outside the British Isles and Ireland. Firstly, the Norwegian strain of LI virus almost certainly represents a British strain of LI virus (Gao *et al.*, 1993b) that was transported presumably via sheep, birds or ticks from Scotland to Norway since it is clearly very closely related to viruses of genotype 1 and Norway is geographically close to Scotland for such an exchange to occur. On the other hand, it would be extraordinary if NEG virus really did represent a strain of LI virus that occurs in Japan as transport of infection from Britain to Japan by natural means would appear to be improbable. It is also important to note that there has never been another report of NEG virus in any region of the world. A plausible explanation for this apparent anomaly is that NEG virus represents a laboratory contaminant which arose during the identification procedure that took place when NEG virus was first isolated and characterised (Ando *et al.*, 1952).

From the data presented in this study LI virus appears to have evolved in distinct localities within the British Isles. There is very little evidence of significant interaction/interchange between strains of LI virus from different areas of the British Isles and Ireland despite very significant internal trade of livestock in the country. This is perhaps not surprising since ticks are relatively immobile arthropods, i.e., they do not move horizontally and when they attach to an animal such as a sheep, deer or hare, their movement is restricted to the area in which the animal moved. Therefore, the tick ecology and physiology explain the correlation between the genotypes of LI viruses and their geographical distribution. As a general rule, the rates of evolution (mutation fixation) for all the arboviruses are relatively lower by an order of magnitude or more than are the rates for non-arthropod-borne viruses (Strauss and Strauss, 1994; Weaver *et al.*, 1994). The evolutionary rates of mosquito-borne flaviviruses are higher than those of their tick-borne counterparts (Zanotto *et al.*, 1995, 1996). The asymmetrical phylogenetic tree for the whole TBE subgroup suggests a geographically consistent dispersal pattern and no radiation dispersal. The tick factor alone cannot be completely responsible for that; other factors, e.g., the pathogenesis of these viruses in humans/animals might be an important factor too.

We have had a cut-off point of aa identity in the E protein for LI virus genotypes of 98.0% in this study. Heinz (1990) observed a similar percentage for WTBE virus isolates from central Europe. We propose that such a cut-off point could be appropriate for other tick-borne flavivirus genotypes.

The molecular genetic structure of LI virus has been investigated using viruses from most regions in which LI virus is considered to be present in the British Isles and Ireland. Our results indicate the extent to which these viruses and their closest relatives have diversified in the recent past. These types of analysis should contribute to our understanding of the dispersal of such viruses and may enable prediction to be made of their future behaviour in the environment. Further study in the future for the dispersal pattern of other viruses in the TBE subgroup, e.g., WTBE or FETBE viruses, are needed to unravel further determinants of the stasis of the TBE cline. It is clear that it would not be appropriate to use other methods (e.g. monoclonal antibodies as did Stephenson *et al.* (1984) and Hubálek *et al.* (1995)) than nt and/or aa sequencing to draw any picture on the relationships among the viruses in such a genetic continuum or cline as TBE serocomplex.

Acknowledgements. The authors thank Drs. W.R. Jiang, W.H. Zhang, B. Lewis, M.W. Gaunt (IVEM, NERC, Oxford) for useful advice during the preparation of this manuscript and Dr. A. Gibbs (The Australian National University) for providing the DIPLOMO and DISCALC programmes for the analysis of the substitution patterns. The first author (G.F. Gao) was supported by an SBFSS fellowship of the British Council and Chinese State Education Commission.

References

Ando K, Kuratsuka K, Arima S, Hironaka N, Honda Y, Ishii K (1952): Studies on the viruses isolated during epidemic of Japanese B encephalitis in 1948 in Tokyo area. *Kitasata Arch. Exp. Med.* **24**, 49–61 (in Japanese).

Calisher CH, Karabatsos N, Dalrymple JM, Shope RE, Porterfield JS, Westaway EG, Brandt WE (1989): Antigenic relationships between flaviviruses as determined by cross-neutralization tests with polyclonal antisera. *J. Gen. Virol.* **70**, 37–43.

Francki RIB, Fauquet CM, Knudson DL, Brown F (Eds) (1991): *Classification and Nomenclature of Viruses. Fifth Report of the International Committee on Taxonomy of Viruses.* Wien-New York, Springer Verlag.

Felsenstein J (1993): *PHYLIP (Phylogeny Inference Package) Version 3.5p.* University of Washington, Seattle.

Gao GF (1995): Molecular biological and immunological studies of tick-borne flaviviruses. University of Oxford. *DPhil Thesis.*

Gao GF, Hussain MH, Reid HW, Gould EA (1993a): Classification of a new member of the TBE flavivirus subgroup by its immunological, pathogenetic and molecular characteristics: identification of subgroup-specific pentapeptides. *Virus Res.* **30,** 129–144.

Gao GF, Hussian MH, Reid HW, Gould EA (1994): Identification of naturally occurring monoclonal antibody escape variants of louping ill virus. *J. Gen. Virol.* **75**, 609–614.

Gao GF, Jiang WR, Hussain MH, Venugopal K, Gritsun TS, Reid HW, Gould EA (1993b): Sequencing and antigenic studies of a Norwegian virus isolated from encephalomyelitic sheep confirm the existence of louping ill virus outside Great Britain and Ireland. *J. Gen. Virol.* **74**, 109–114.

Gonzalez L, Reid HW, Pow I, Gilmour JS (1987): A disease resembling louping-ill in sheep in the Basque region of Spain. *Vet. Rec.* **121**, 12–13.

Gould EA, Buckley A, Higgs S, Gaidamovich S (1990): Antigenicity of flaviviruses. *Arch. Virol.*, Suppl. **1**, 137–152.

Hartley WJ, Martin WB, Hakioglu F, Chifney STE (1969): A viral encephalitis of sheep in Turkey. *Pendik Institute J.* **2**, 89–95.

Heinz FX (1986): Epitope mapping of flavivirus glycoproteins. *Adv. Virus Res.* **31**, 103–168.

Heinz FX, Mandl CW, Guirakhoo F, Holzmann H, Tuma W, Kunz C (1990): The envelope protein E of tick-borne encephalitis virus and other flaviviruses: structure, function and evolutionary relationships. *Arch. Virol.*, Suppl. **2**, 125–135.

Hubálek Z, Pow I, Reid HW, Hussain MH (1995): Antigenic similarity of central European encephalitis and louping-ill viruses. *Acta Virol.* **39**, 251–256.

Koonin EV (1991): The phylogeny of RNA-dependent RNA polymerase of positive-strand RNA viruses. *J. Gen. Virol.* **72**, 2197–2206.

Kumar S, Tamura K, Nei M (1993): *MEGA: Molecular Evolutionary Genetic Analysis, Version 1.0.* Pennsylvania State University, University Park.

Leigh Brown AJ (1994): Methods of evolutionary analysis of viral sequences. In Morse SS (Ed.): *The Evolutionary Biology of Viruses.* Raven Press, New York, pp. 74–85.

Li W-H, Graur D (1991): *Fundamentals of Molecular Evolution.* Sinauer Association, Sunderland, p. 48.

Maddison WP, Maddison DR (1992): *MacClade: Analysis of Phylogeny and Character Evolution, Version 3.0.* Sinauer Association, Sunderland.

Marin MS, McKenzie J, Gao GF, Reid HW, Antoniadis A, Gould EA (1995a): The virus causing encephalomyelitis in sheep in Spain: a new member of the tick-borne encephalitis group. *Res. Vet. Sci.* **58**, 11–13.

Marin MS, Zanotto PM de A, Gritsun TS, Gould EA (1995b): Phylogeny of TYU, SRE, and CFA virus: different evolutionary rates in the genus *Flavivirus*. *Virology* **206**, 1133–1139.

Nei M (1987): *Molecular Evolutionary Genetics.* Columbia University Press, New York, pp. 79–86.

Papadopoulos O, Paschaleri-Papadopoulou E, Deligaris N, Doukas G (1971): Isolation of tick-borne encephalitis virus from a flock of goats with abortions and fatal diseases. *Vet. News-Greece* **3**, 112–114.

Papadopoulos O (1980): Arbovirus problems in Greece. In Vesenjak-Hirjan J, Porterfield JS, Arslanagic E (Eds): *Arboviruses in the Mediterranean Countries.* Gustav Fischer Verlag, Stuttgart-New York, pp. 117–121.

Pavlov P (1968): Studies of tick-borne encephalitis of sheep and their natural foci in Bulgaria. *Zentralb. Bakt. Parasit. Infekt. Hyg.* **206**, 360–367.

Pogodina VV, Bochkova NG, Zlobin VI, Frolova MP, Drokin DA, Levina LS, Dzhioev IuP (1993): The genetic characteristics of the Vergina serotype of the tick-borne encephalitis virus and its pathogenetic traits. *Vopr. Virusol.* **38**, 158–162 (in Russian).

Pool WA, Brownlee A, Wilson RD (1930): The aetiology of louping ill. *J. Pathol. Therap.* **43**, 253–290.

Reid HW (1984): Epidemiology of louping ill. In Mayo MA, Harrap KA (Eds): *Vectors in Virus Biology.* Academic Press, London, pp. 161–178

Reid HW (1988): Louping ill. In Monath TP (Ed.): *The Arboviruses: Epidemiology and Ecology.* Vol. 3, CRC Press, Boca Raton, pp. 117–135

Rey FA, Heinz FX, Mandl C, Kunz C, Harrison SC (1995): The envelope glycoprotein from tick-borne encephalitis virus at 2 resolution. *Nature* **375**, 291–298.

Rubin SG, Chumakov MP, Papadopoulos O (1975): The antigenic type of Vergina strain of tick-borne encephalitis virus isolated in Greece from sick goat in 1969. *Proc. Sci. Ser. Institute of Poliomyelitis and Viral Encephalitides.* Academy of Medical Sciences of USSR, Moscow, p. 353.

Sagner F, Nicklen S, Coulson AR (1977): DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.

Sambrook J, Maniatis T, Fritsch EF (1989): *Molecular Cloning: A Laboratory Manual.* 2nd Ed. Cold Spring Harbor Laboratory, Cold Spring Harbor.

Shiu SYW, Ayres M, Gould EA (1991): Genomic sequence of the structural protein of louping ill virus: comparative analysis with tick-borne encephalitis virus. *Virology* **180**, 411–415.

Shiu SYW, Jiang WR, Porterfield JS, Gould EA (1992): Envelope protein sequences of dengue virus isolates TH-36 and TH-Sma, and identification of a type-specific genetic marker for dengue and tick-borne flaviviruses. *J. Gen. Virol.* **73**, 207–212.

Strauss JH, Strauss EG (1994): The alphaviruses: gene expression, replication, and evolution. *Microbiol. Rev.* **58**, 491–562.

Stephenson JR, Lee JM, Wilton-Smith PD (1984): Antigenic variation among members of the tick-borne encephalitis complex. *J. Gen. Virol.* **65**, 81–89.

Swofford DL (1993): *PAUP: Phylogenetic Analysis Using Parsimony, Version 3.1.* Illinois Natural History Survey, Champaign.

Venugopal K, Buckley A, Reid HA, Gould EA (1992): Nucleotide sequence of the envelope glycoprotein of Negishi virus shows very close homology to louping ill virus. *Virology* **190**, 515–521.

Weaver SC, Hagenbaugh A, Bellew LA, Gousset L, Mallampalli V, Holland JJ, Scott TC (1994): Evolution of alphaviruses in the eastern equine encephalomyelitis complex. *J. Virol.* **68**, 158–169.

Weiller GF, Gibbs AJ (1993): *DIPLOMO: Distance Plot Monitor, Version 1.0.* Australian National University, Canberra.

Whitby JE, Whitby SN, Jennings AD, Stephenson JR, Barrett ADT (1993): Nucleotide sequence of the envelope protein of a Turkish isolate of tick-borne encephalitis (TBE) virus is distinct from other viruses of the TBE virus complex. *J. Gen. Virol.* **74**, 921–924.

Zanotto PM de A, Gao GF, Gritsun TS, Marin MS, Jiang WR, Venugopal K, Reid HW, Gould EA (1995): An arbovirus cline across the Northern Hemisphere. *Virology* **210**, 152–159.

Zanotto PM de A, Gould EA, Gao GF, Harvey PH, Holmes EC (1996): Population dynamics of flaviviruses revealed by phylogenetic analysis. *Proc. Natl. Acad. Sci. USA* **93**, 548–553.