Evolution of Human α-Herpesviruses

Peter Norberg

GÖTEBORG UNIVERSITY

Dept. of Clinical Virology,

2007

Peter Norberg
Evolution of human alpha-herpesviruses

Cover by: Björn Norberg

© 2007 Peter Norberg

ISBN 978-91-628-7094-2

To Eva, Hugo and Elliot

ABSTRACT

Herpesviridae is a large virus family with more than 100 members, which are highly disseminated among animals. Three sub-families have been classified; alpha-herpesviruses, beta-herpesviruses and gamma-herpesviruses. Eight herpesviruses have hitherto been identified in humans of which three belong to the alpha-herpesviruses; (i) herpes simplex virus type 1 (HSV-1), which is a ubiquitous pathogen causing mainly oral or genital lesions, (ii) herpes simplex virus type 2 (HSV-2), which is closely related to HSV-1, and is the most common sexually transmitted virus globally, causing mainly genital lesions, and (iii) Varicella zoster virus (VZV), which is the cause of chicken pox and shingles. All alpha-herpesviruses give lifelong infections and establish latency in the sensory ganglia. In the present work, the genetic variability of clinical HSV-1, HSV-2 and VZV isolates was investigated.

Twenty-eight clinical HSV-1 isolates were collected from patients suffering from oral or genital lesions or encephalitis and compared with the laboratory strains F, KOS321 and 17. Phylogenetic analyses based on the genes US4, US7 and US8 divided the isolates into three genogroups, arbitrarily designated as A, B and C, differing in DNA sequences by approximately 2%. In addition, seven clinical isolates as well as strain 17 were classified as recombinants. To facilitate further genotyping of clinical isolates an assay was developed based on restriction enzyme cleavage of PCR-products. Furthermore, a polymorphic tandem repeat (TR) region was detected in US7. The region encodes the amino acids serine, threonine and proline, which are targets for O-linked glycosylation. Using a synthetic peptide, containing two of the repeated blocks, it was shown that the described TR-region is a substrate for massive O-linked glycosylation, and hence codes for a mucin region. Mucin regions have not been described previously within herpesvirus-encoded proteins.

The corresponding genes were sequenced and investigated for 45 clinical HSV-2 isolates collected in Sweden, Norway and Tanzania. Phylogenetic analysis revealed a divergence of the isolates in one Tanzanian and one European genogroup, arbitrarily designated as A and E, differing by approximately 0.4%. In addition, analyses using recombination networks, the BootsScan method and the phi-test, suggested that most HSV-2 isolates are mosaic recombinants.

The complete genome was sequenced for two VZV isolates and compared with the laboratory strains MSP, Dumas, BR, p-Oka and the vaccine strain v-Oka. The results show a division of VZV into four genogroups, designated as

E, J, M1 and M2, of which M1 and M2 were suggested to be recombinants derived from ancient recombination events between viruses from the E and J genogroups.

In conclusion, the results presented here demonstrate that clinical isolates, for all three investigated human alpha-herpesviruses, can be divided into different genogroups. Estimations of evolutionary timescales suggest that the divergence of the three HSV-1 genogroups may have occurred approximately 500,000 Myears BP, i.e. prior to the emergence of *Homo sapiens*. Furthermore, it is evident that intrastrain recombination is a prominent feature of the evolutionary history of these viruses. Thus, homologous recombination is suggested to be a powerful evolutionary mechanism for human alpha-herpesviruses to exchange genetic segments between different viral strains, as well as to create variability of TR-regions.

List of papers

This thesis is based on the following papers:

I. **Norberg, P., Bergström, T., Rekabdar, E., Lindh, M. & Liljeqvist, J-Å.** Phylogenetic analysis of clinical herpes simplex virus type 1 isolates identified three genetic groups and recombinant viruses. J Virol 2004; 78: 10755-10764.

II. **Norberg, P., Bergström, T. & Liljeqvist, J-Å.** Genotyping of clinical herpes simplex virus type 1 isolates by use of restriction enzymes. J Clin Microbiol 2006; 44: 4511-4514

III. **Norberg, P., Olofsson, S., Agervig Tarp, M., Clausen, H., Bergström, T. & Liljeqvist J-Å.** Glycoprotein I of herpes simplex virus type 1 contains a unique polymorphic tandem repeated mucin region. Submitted for publication.

IV. **Norberg, P., Kasubi, M.J., Haarr, L., Bergström, T. & Liljeqvist, J-Å.** Evolution of herpes simplex virus type 2 –Identification of two genogroups and multiple recombinants. In manuscript.

V. **Norberg P., Liljeqvist, J-Å., Bergström, T., Sammons, S., Schmid, D.S. & Loparev V.N.** Complete-genome phylogenetic approach to varicella-zoster virus evolution: genetic divergence and evidence for recombination. J Virol 2006; 80: 9569-9576.

CONTENTS

ABBREVIATIONS

aa	Amino acid(s)
bp	Base pairs
BP	Before present
DNA	Deoxyribonucleic acid
EBV	Epstein-Barr virus
EHV-1	Equine herpesvirus 1
EHV-4	Equine herpesvirus 4
GalNAc	N-acetyl galactosamine
gG, gI, gE	Glycoprotein G, I, E
HBV	Hepatitis B virus
HCMV	Human cytomegalovirus
HHV-6	Human herpesvirus 6
HHV-7	Human herpesvirus 7
HHV-8	Human herpesvirus 8
HIV	Human immunodeficiency virus
HSV	Herpes simplex virus
HSV-1	Herpes simplex virus type 1
HSV-2	Herpes simplex virus type 2
HTU	Hypothetical taxonomic unit
kb	Kilobases
nt	Nucleotide(s)
OTU	Operational taxonomic unit
PCR	Polymerase chain reaction
PrV	Pseudorabies virus
P	Proline
RNA	Ribonucleic acid
S	Serine
snp	Single nucleotide polymorphism
T	Threonine
TR	Tandem repeats
UL	Unique long
US	Unique short
VNTR	Variable number of tandem repeats
VZV	Varicella-zoster virus

GENERAL BACKGROUND

The herpesvirus family

The name Herpes is derived from the Greek *Herpein* -'to creep', which refers to its ability to give recurrent eruptions.

Herpesviridae is a large virus family with more than 100 members, which are highly disseminated among animals (Roizman, 1996a). Eight herpesviruses have hitherto been identified in humans: herpes simplex virus type 1 (HSV-1) and type 2 (HSV-2), human cytomegalovirus (HCMV), Epstein-Barr virus (EBV), varicella-zoster virus (VZV), human herpesvirus type 6 (HHV-6), type 7 (HHV-7) and type 8 (HHV-8), where HHV-8 is the most recently reported member. Molecular phylogeny of the human herpesviruses clearly establishes three subfamilies (McGeoch *et al.*, 1995) (Fig. 1). These three groups correspond to the current taxonomic classification based on biological properties and include *alphaherpesvirinae* (α), *betaherpesvirinae* (β), and *gammaherpesvirinae* (γ). HSV-1, HSV-2 and VZV belong to the *alphaherpesvirinae* and have a wide host cell range, efficient and rapid reproductive cell cycle, and the capacity to establish latency in the sensory ganglia (Roizman, 1996b).

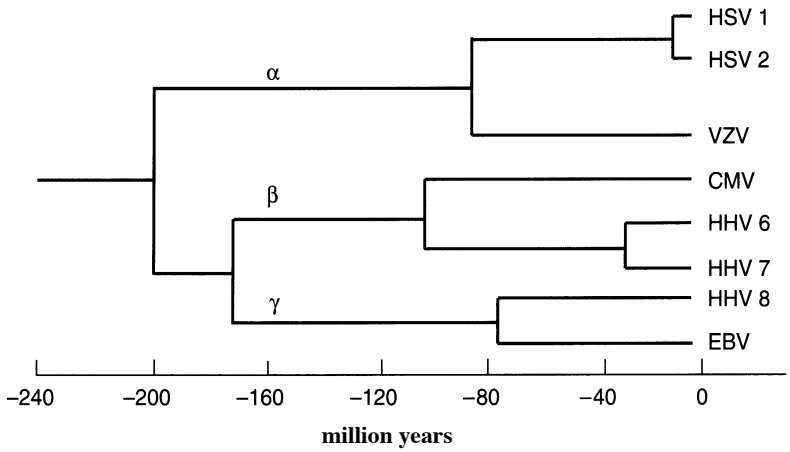

Fig. 1. *Phylogenetic tree over the eight herpesviruses that have been identified in humans,* (McGeoch *et al.*, 1995).

Herpesviruses are large and complex DNA-viruses, which have evolved over a period of at least 400 million years (McGeoch et al., 2000; Weir, 1998). Among them, the α-herpesvirus subfamily diverged 180-210 million years ago (McGeoch et al., 1995). The genomes of *Herpesviridae* differ widely with a size ranging from 124 kb for simian varicella virus from the *α-herpesvirinae* (Gray et al., 2001) to 241 kb for chimpanzee cytomegalovirus from the β-*herpesvirinae* (Davison et al., 2003). The number of genes encoded by the genomes ranges from 70 to 200, where HSV-1 and HSV2 encode at least 74 genes (Dolan et al., 1998) and VZV encodes at least 70 genes (Davison, 2000; Kemble et al., 2000). In addition, the G+C content ranges widely, from 32% to 75% (Honess, 1984). Despite those differences, approximately 40 genes are common to all mammalian herpesviruses as regards conservation of encoded amino acid sequences and local gene layout (Chee et al., 1990; Davison & Taylor, 1987; McGeoch, 1989). The herpesvirus viron consists of a core containing the DNA, an icosahedral capsid (100-125 nm in diameter), the tegument and the surrounding lipid envelope containing the viral glycoproteins (Fig. 2). Although there is a large variation in the genomic sequence and the encoded proteins, the viral structure is similar for all herpesviruses and it is difficult to distinguish them in electron micrographs.

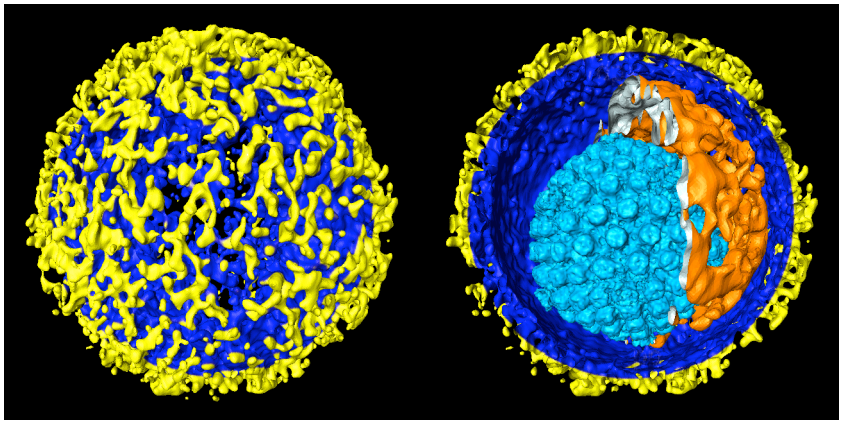

Fig. 2. *Three-dimensional structure of HSV from cryo-electron tomography (Grunewald et al., 2003).*

The genomes consist of a unique long (UL) and a unique short (US) segment, which are flanked by inverted repeat regions (Sheldrick & Berthelot, 1975; Wadsworth et al., 1975). Because of these repeats, the unique regions are rearranged during replication into mixtures of four different isomers with different orientations of UL and US segments. The repeated regions are

variable and vary in size up to 10 kbp between particular viruses. The replication of herpesviruses in the nucleus of the host cell in combination with a sophisticated viral DNA replication machinery lead to an efficient proofreading activity (Crute & Lehman, 1989; Drosopoulos et al., 1998; Kato *et al*., 1994). The rate of synonymous nucleotide substitutions has been estimated to 3 x 10^{-8} substitutions per site per year (Sakaoka et al., 1994), which is about 20 times higher than the rate in mammalian genomes (Dolan et al., 1998; Hughes, 2002; Markine-Goriaynoff et al., 2003), although significantly lower than described for most RNA-viruses.

HSV-1 is the most well-studied virus within the α-herpes subgroup and is usually associated with oral lesions. However, genital lesions have recently become more commonly detected and HSV-1 is now considered as a major cause of genital lesions in several western world countries including Sweden. Although oral or genital lesions are usually harmless, more severe symptoms may occur such as encephalitis, myelitis, meningitis, facial palsy and keratitis. HSV-2 is the most common sexually transmitted pathogen, usually associated with genital lesions. In similarity with HSV-1, HSV-2 can also induce severe symptoms like meningitis and a devastating neonatal infection.

HSV spreads via direct contact. Although virus particles are present in enormous amounts in lesions, a recent study (Liljeqvist et al., unpublished) showed that asymptomatic shedding occurs frequently in HSV-1-positive individuals. No vaccines against HSV are present on the market today, although efforts are made to develop a vaccine against HSV-2 since genital lesions is a major risk factor for the transmission of HIV, especially in developing countries.

VZV is the cause of chicken pox and can also reactivate later in life causing herpes zoster (shingles). Like HSV, VZV can also cause more severe symptoms like pneumonia, meningitis, encephalitis and keratitis. Owing to better and more accurate diagnostic methods, new findings reveal that VZV seems to be a much more common reason for encephalitis than previously described (Bergström et al., unpublished). Contrary to HSV, VZV can be transmitted via air, especially during the first days of chicken pox infection. An attenuated vaccine strain (v-Oka) is at the moment being introduced in the world.

Herpesvirus glycoproteins

All α-herpesviruses possess several glycoprotein-encoding genes. HSV-1 and HSV-2 encode at least eleven glycoproteins (g) B, C, D, E, G, H, I, J, K, L and M, whereas VZV lacks the gD-gene and has no apparent gD functional homologue. All encoded glycoproteins except gK (Hutchinson et al., 1995) are attached on the virus envelope as well as on virus-infected cell membranes. The herpes viral glycoproteins are involved in several functions, such as in the virus entrance into the host cell through fusion with the lipid envelope or in cell-to-cell fusion, cell-to-cell spread and in the escape of the host's immune system (Haarr & Skulstad, 1994). There are various degrees of sequence homology between HSV-1 and HSV-2, where the most conserved glycoproteins, gD and gB, differ by only 15%. In contrast, the sequence homology of VZV gE and HSV gE is only 27% (Litwin et al., 1992). Another example is the gG-gene, of which a large portion is deleted in the HSV-1 (714 nt) in comparison with the HSV-2 gG gene (2097 nt). Although the fundamental functions of HSV-1, HSV-2 and VZV glycoproteins are similar, several functional differences have been demonstrated; VZV-mediated cell-to-cell fusion requires only a combination of two glycoproteins, either gH and gL (Duus & Grose, 1996; Duus et al., 1995) or gB and gE (Duus & Grose, 1996; Duus et al., 1995), whereas HSV-1 and HSV-2 require the combination of the four glycoproteins gH, gL, gB and gD for cell-to-cell fusion (Muggeridge, 2000; Turner et al., 1998). In addition, while VZV gE is essential for virus replication in cell-cultures (Mallory et al., 1997), HSV-1 gE can be dispensable for replication in cell-cultures (Longnecker et al., 1987; Longnecker & Roizman, 1987).

The glycoprotein genes of HSV-1 and HSV-2 studied in this work code for gG, gI and gE. All three genes US4, US7 and US8 encoding the gG, gI and gE, respectively, are located in the US-segment of the HSV genome (Fig. 3). The VZV genome contains genes corresponding to US7 and US8, but lacks a counterpart of US4 (Davison, 1983; Davison & Scott, 1986). Comparisons of the gene DNA sequences in the US-segment have demonstrated that US4 and US7 are similar, and have probably evolved by duplication and divergence of the gD gene (McGeoch, 1990). Although US8 is distinct from US4 and US7, a more distant relationship has been suggested based on conservation of two clusters of cysteine residues.

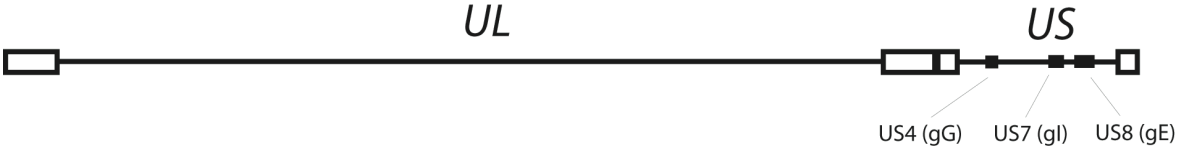

Fig. 3. *Schematic illustration of the HSV genome.*

It has been demonstrated that gI binds to and forms a complex with gE, and that the gE/gI complex is involved in cell-to-cell spread in epithelial (Dingwell & Johnson, 1998) and neuronal tissue (Dingwell et al., 1995) as well as in the virus escape from the host immune system by Fc-receptor binding of IgG-antibodies (Chapman et al., 1999; Dubin et al., 1990; Hanke et al., 1990; Johnson & Feenstra, 1987; Johnson et al., 1988). The known function of gG (in HSV-1) is that it facilitates entry through apical polarized cell surfaces. The functions of gG, gE and gI in HSV-2 have not been described.

Mucins

Mucins are a family of highly glycosylated proteins, usually secreted or present on apical cell membranes of human epithelial cells. They are produced from the mammary and salivary glands, digestive and respiratory tracts, bladder, kidney, prostate, uterus and testis. The non-globular protein backbone contains both highly glycosylated and unglycosylated regions. The glycosylated regions contain high levels of serine, threonine, alanine, glycine and proline residues but, in contrast, low amounts of aromatic- and sulphur-containing amino acids. The serine and threonine residues on mucins are modified by the addition of GalNAc residues, catalyzed by polypeptide GalNAc transferases (GalNAcT), which results in an O-linked oligosaccharide or O-glycan. Typically, the glycosylated regions in mucins vary drastically in size due to variable numbers of tandem repeats (VNTR) rich in O-glycosylation sites. Although the exact functions of mucins are unknown, several possible functions have been proposed. Mucins present on the cell surface may act as a barrier between the cell surface and the surrounding environment, which may protect the cell against microorganisms, toxins or proteolytic attack. Changes in recognition pattern for microorganisms by the extension of the VNTR backbone may also occur. Other possible functions are prevention of proteolytic degradation, facilitation of fatty-acid uptake, lubrication of epithelial surfaces and regulation of cell growth by mimicking high cell density. In addition, the size of VNTR regions and differences in glycosylation may affect tumor cell recognition and promote metastasis in mammals. Although no specific consensus sequence for O-glycosylation has been demonstrated, some predictive abilities have been achieved. Several algorithms based on databases with known mucins have been developed (Gupta et al., 1999). These algorithms use recognition-patterns achieved from known mucin genes to predict possible O-glycosylation sites or mucin regions. However, the accuracy of those

predictions has been questioned. No mucin region has previously been described for human α-herpesviruses.

INTRODUCTION TO VIRAL EVOLUTION

The null-hypothesis of evolution of all diploid organisms, is that no evolution occurs, and hence, that the allele or genotype frequency remains constant between generations within a population. Under such conditions the genotype frequencies for diploid organisms can be calculated from $p^2 + 2pq + q^2$, where p and q are the respective frequency in a gene pool of two possible alleles on a particular locus (the Hardy-Weinberg equilibrium). As most viruses are haploid organisms, i.e. only have one set of each gene, the genotype frequency is equal to the allele frequency. If the frequency of genotypes changes from one generation to the next, the population is evolving. When an allele frequency has reached 100% in a population, that allele is said to be *fixed* in that population, and that alternative alleles are lost. Evolution, or changes in the allele frequency (=genotype frequency for most viruses), can be caused by mutations, natural selection and/or random sampling error (genetic drift, founder effect or bottlenecks).

Mutations

Mutations in DNA or RNA sequences are the fundamental basis for the evolutionary process in all living organisms. The most common mutation is a single nucleotide shift, but there also exist insertions and deletions in the genome where an entire region is deleted or inserted. Mutations are introduced in the genome either spontaneously - randomly caused by error in the replication machinery - or as a result of exposure to toxic materials, nuclear or ultraviolet radiation or specific chemicals.

Natural selection

Natural selection, or survival of the fittest as Charles Darwin stated (1859), is the process where organisms with favorable traits have a higher probability to survive and reproduce than organisms with unfavorable traits. Genetic events such as point mutations, deletions, insertions or recombinants can be beneficial, harmful or neutral. When a specific, usually random, mutation arises, at least three possibilities exist. If the mutation is harmful, the mutant will most likely disappear quickly from the population due to the selection pressure in favor of more “biologically fit” individuals. If the

mutation is neutral, i.e. does not lead to any amino acid shift, or if the change does not interfere with essential functions of the organism, the outcome of that mutant will depend on how the genetic drift affects its frequency in the population (see below). If, on the other hand, the mutation is beneficial, natural selection will most likely favor the mutant, which will have a higher probability to survive and/or reproduce than the parental organism. However, it is important to note that a mutant with less biological fitness under normal conditions may be favored and selected for under special circumstances, for example when the environment changes. Examples of viruses under selection pressure from antiviral drugs have been frequently reported and antiviral drug-resistance of herpesvirus mutants has been described, especially acyclovir resistance due to a mutation in the thymidine kinase-coding gene in HSV (for reviews see Aymard, 2002; Collins, 1993; Crumpacker, 1988; True & Carter, 1984).

Genetic drift and the founder effect

A powerful mechanism behind the evolution of all organisms is genetic drift. Random genetic drift is independent of natural selection (in contrast to the Darwinian "survival of the fittest" theorem) and is a stochastic process and the evolutionary equivalent to sampling error. Consequently, genetic drift results in a random increase or decrease of the frequency of specific alleles or genotypes transferred from one generation to the next.

The size of the population in which sampling errors take place is of great importance. In large populations, genetic drift will have little effect since the random nature of the sampling errors will often average out. Small populations, on the other hand, are much more sensitive to sampling errors, so the effect of genetic drift can be very rapid and highly significant.

Suzuki et al. (1989) explain genetic drift in the following way:

"If a population is finite in size (as all populations are) and if a given pair of parents have only a small number of offspring, then even in the absence of all selective forces, the frequency of a gene will not be exactly reproduced in the next generation because of sampling error. If in a population of 1000 individuals the frequency of "a" is 0.5 in one generation, then it may by chance be 0.493 or 0.0505 in the next generation because of the chance production of a few more or less progeny of each genotype. In the second generation, there is another sampling error based on the new gene frequency, so the frequency of "a" may go from 0.0505 to 0.501 or back to

0.498. This process of random fluctuation continues generation after generation, with no force pushing the frequency back to its initial state because the population has no "genetic memory" of its state many generations ago. Each generation is an independent event. The final result of this random change in allele frequency is that the population eventually drifts to p=1 or p=0. After this point, no further change is possible; the population has become homozygous. A different population, isolated from the first, also undergoes this random genetic drift, but it may become homozygous for allele "A", whereas the first population has become homozygous for allele "a". As time goes on, isolated populations diverge from each other, each losing heterozygosity. The variation originally present within populations now appears as variation between populations."

Although the reproduction of viruses differs from that of diploid organisms, genetic drift has the same powerful impact on virus populations. All living organisms are subjected to different stochastic processes. Two important examples, where sampling error plays a critical role and which drastically enhances the effect of genetic drift, are the bottleneck effect and the founder effect.

The bottleneck effect refers to random, usually accidental, events that reduce the population size and hence also randomly influence allele and genotype frequencies of the population. Examples of such events are natural disasters like earthquakes, floods, storms and fires, which lead to the survival of only a small fraction of the population. Although survival from natural disasters might sometimes be influenced by selection of the fittest, the mortality is usually unselective. Typically, the size of a population is usually restored within a relatively short time after a bottleneck period. However, the longer the population remains at a reduced size, the higher the impact of genetic drift on the allele or genotype frequency.

The founder effect is an alternative cause of decreased population size and occurs when a small cohort of a population breaks off and forms a smaller population in another geographic region. Because of sampling errors, new populations (founder populations) tend to have different allele frequencies than their parental populations and, in addition, the limited size of the new population drastically increases the power of genetic drift (Fig. 4). Typically, the small size of a founder population tends to remain for a longer time than the small size resulting from a bottleneck event.

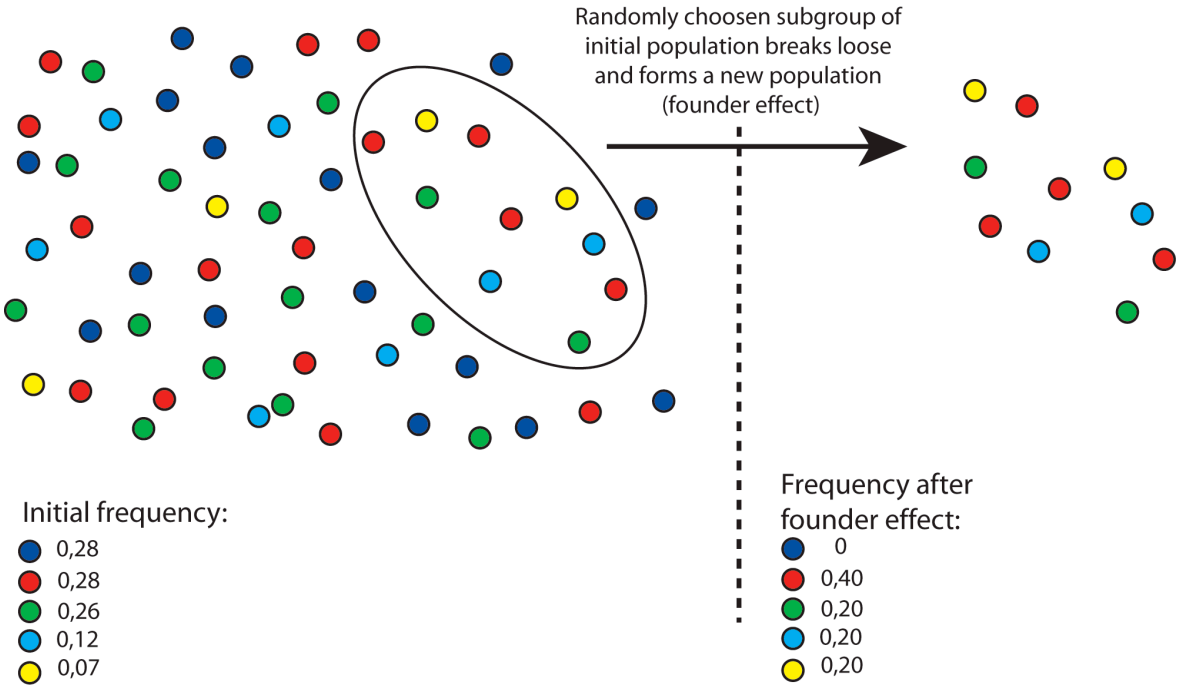

Fig. 4. *The founder effect.*

In conclusion, genetic drift, bottlenecks and the founder effect, in addition to mutations and natural selection, are forces behind the evolution of organisms. According to the neutral theory of evolution (Kimura, 1979; 1987), the majority of all mutations present in nature is caused by random fixation rather than Darwinian natural selection. It has also been shown that the neutral theory has been in operation at least for the viruses HIV, hepatitis B virus (HBV) and influenza A viruses (Gojobori et al., 1990). However, the most important consequences of genetic drift, caused by random sampling errors, are the loss of genetic variability within populations, but an enhanced genetic divergence between populations.

Recombination

Genetic recombination is the molecular process, which generates new combinations of genetic material (Leach, 1996). Similarly, viral recombination is a phenomenon that occurs when two viruses of different parent strains co-infect the same cell and interact during replication to generate progeny, the genomes of which consist of genetic segments obtained from both parental strains. Two main mechanisms can mix viral genetic material: independent assortment and recombination (incomplete linkage).

Independent assortment is exclusive for viruses with segmented genomes, for example the influenza viruses. In such viruses, loci on different segments are unlinked. During a co-infection of the host cell by viruses with segmented genomes, different genetic segments can be mixed. When progeny virus particles are created, the segmented genome can consist of segments obtained from different parental strains (FIG 5).

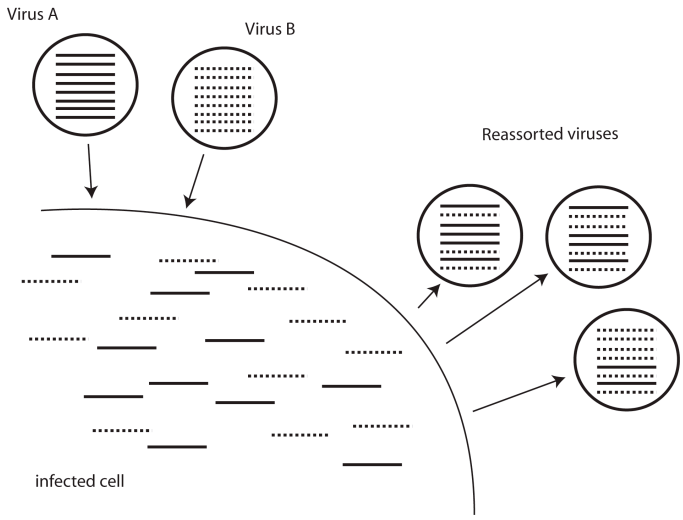

Fig. 5. *Independent assortments of viruses with segmented genomes.*

Recombination (incomplete linkage) is a more complicated process where at least four general types of mechanisms have been described: (i) *homologous recombination*, which involves a reciprocal exchange where a pair of homologous DNA sequences breaks and rejoins in a crossover; (ii) *site specific recombination*, which occurs between DNA molecules with low or no homology by the binding of certain proteins to specific DNA sequences, for example, the non-homologous insertion of DNA into a chromosome, which often occurs during viral genome integration of the host; (iii) *transposition*, which occurs for specific DNA sequences, recognized by so-called transposon-encoded proteins; and (iv) *illegitimate recombination*, in which recombination occurs despite the absence of sequence homology or specific identified sequences. Illegitimate recombination is also sometimes referred to as *non-homologous recombination* (Kowalczykowski et al., 1994; Leach, 1996). Of these four types of recombination, two have been described for herpesviruses; homologous recombination and illegitimate recombination (Umene, 1999).

Homologous recombination is carried out by break-rejoin mechanisms, which require a break in the double-stranded DNA, followed by the invasion of a homologous DNA molecule with a single-stranded DNA end. These homologous DNA sequences are paired and migrate forming a so-called

Holliday junction. The final step is an isomerization of the flanking sequences (FIG 6). Also models requiring single-stranded breaks have been proposed (FIG 7).

Illegitimate recombination occurs among sequences with no or low homology (Leach, 1996) and is normally less common than homologous recombination. The joining of two DNA molecules with no homology is an important mechanism involved in the repair of breaks in the DNA and is divided into two classes (Shimizu et al., 1997); (i) a short-homology independent class related to the action of enzymes affecting DNA, and (ii) a short-homology dependent class, where DNA breaks are ligated after processing and annealing of DNA ends. Illegitimate recombination is believed to be essential for DNA rearrangement, which can lead to duplications of specific genetic regions in the genome (Umene, 1998).

Independently of which mechanism that is responsible for the recombination process, recombinants are interesting from a biological and evolutionary viewpoint. For example, if the genome of a viral recombinant C is a mixture of the genomes from virus A and virus B, the evolutionary step can be enormous as compared with point mutations. Thus, new behaviors and features can appear in a single step. Single or multiple recombination events including several parental virus strains may result in progeny viruses with mosaic genomes consisting of a randomized pattern of genetic blocks originating from different parental strains. If the recombination process is free and randomized, the numbers of different combinations are almost unlimited. By the mechanisms of recombination, one single virus can obtain beneficial mutations from several parental genomes and thereby receive several beneficial functions that would be very unlikely to occur in a single genome without recombination. In addition, harmful mutations can be deleted from a genome by the act of recombination.

Recombination is also an interesting phenomenon from a bioinformatician's point of view since the measurement of the recombination frequency between different loci can be used as a tool to map genomes. If several loci in a genome are sequenced, typed or marked, linkage analysis can be utilized to order the loci in the genome using the recombination frequency between all loci, typically from an *in vitro* recombination assay. Since recombination between two loci is more likely to occur when the distance increases between them, loci closely together in the genome (linked) represent fewer recombination events than loci located far from each other (incomplete linkage). This is valid up to a certain distance, where the genes are regarded as unlinked. In such case, the probability to detect a recombination between

two loci is 0.5, i.e. the probability of detecting a recombinant by investigating two loci will never exceed 0.5 because an even number of multiple recombination points between the loci will leave the recombination event undetected.

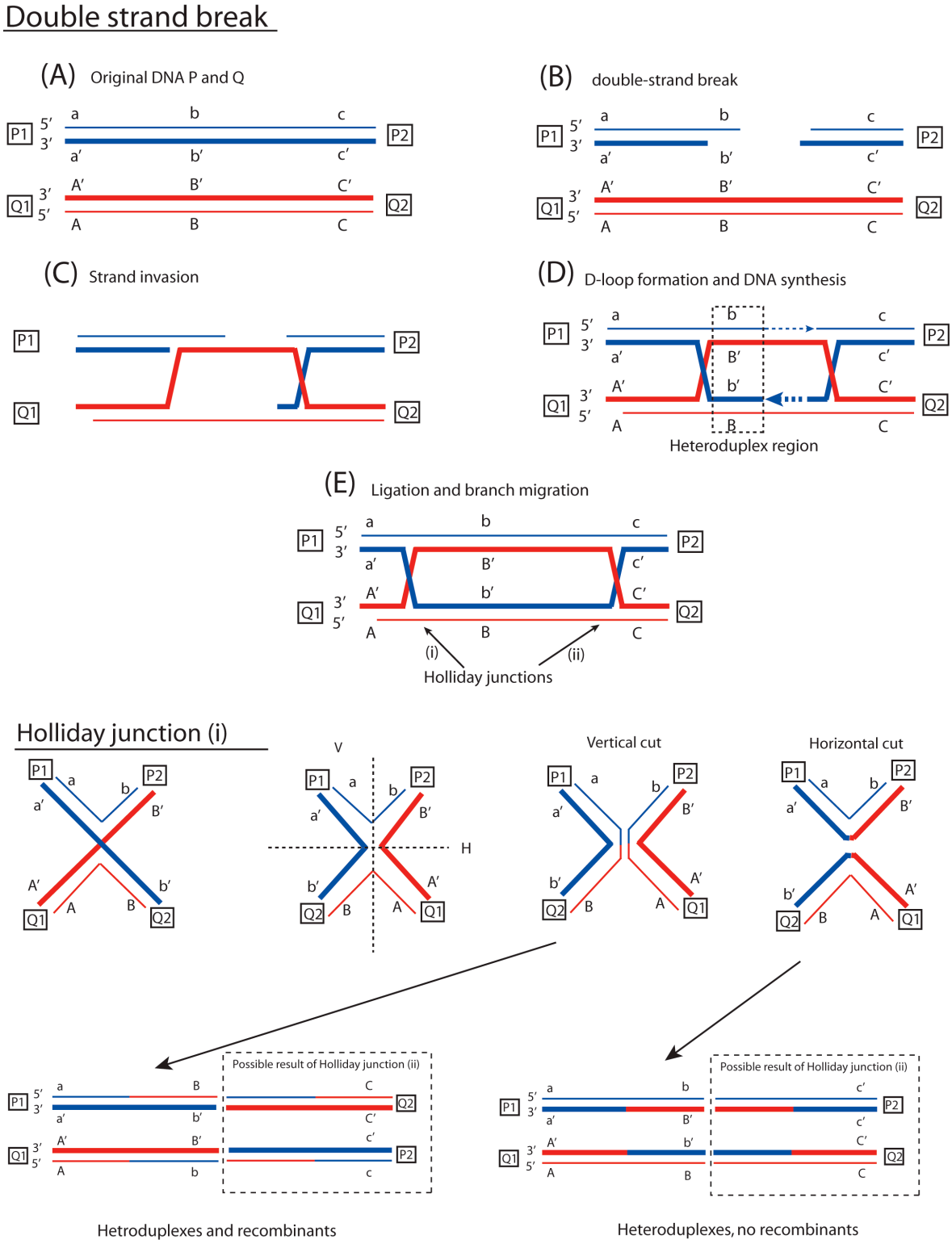

Fig. 6. *Homologous recombination by double strand break.*

Single strand break

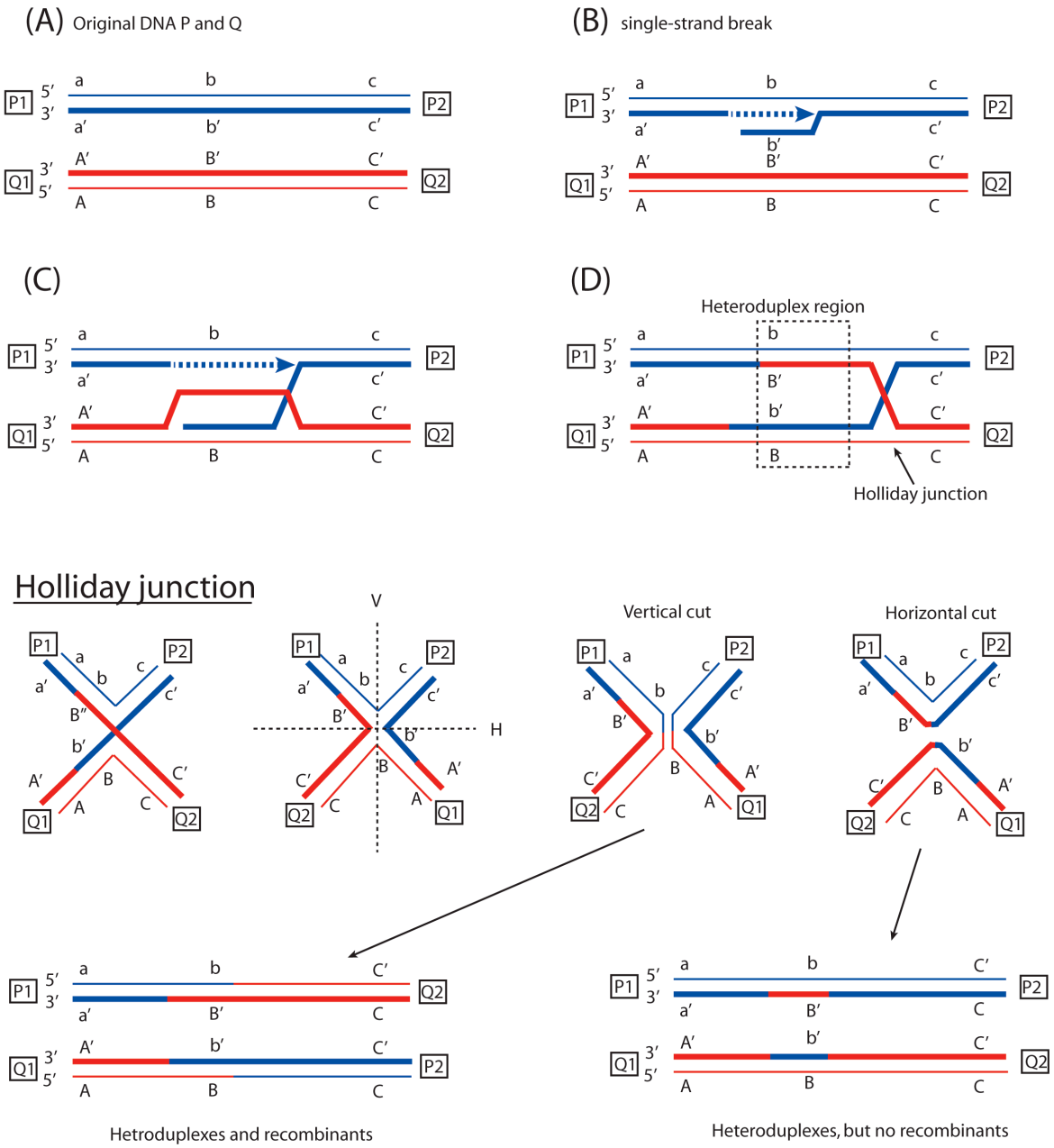

Fig. 7. *Homologous recombination by single strand break.*

Another important function of recombination is the maintenance of the hyper-variability of VNTR regions. VNTR are genomic regions of various size, consisting of repeated genetic blocks. VNTR:s has been shown to be common features of the HSV-genome and is localized within the direct repeat regions at the genomic termini as well as within the internal repeat region separating the L and S segments (Davison & Wilkie, 1981; Mocarski & Roizman, 1981; Perry & McGeoch, 1988). VNTR:s have also been detected within the coding sequences UL36 gene (McGeoch et al., 1988), the US10 gene (Davison & McGeoch, 1986) and for the ICP34.5 gene (Bower et al., 1999; Mao & Rosenthal, 2003). The variability of the length of VNTR:s is usually caused by unequal crossover during homologous recombination (FIG 8), although illegitimate recombination has been proposed to play a certain role for inverted repeat regions and VNTR in the HSV-genome (Umene, 1998).

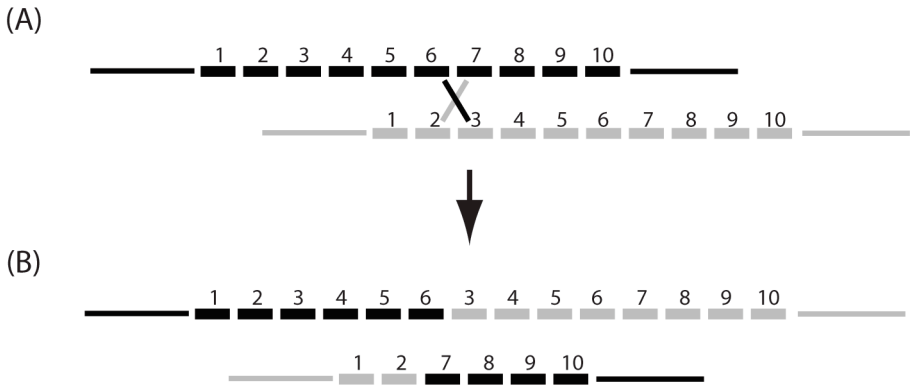

Fig. 8. *Two TR regions, each containing 10 repeats, recombine by homologous recombination (A). Because of an unequal crossover the progeny recombinants contain 6 and 14 repeats, respectively (B).*

INTRODUCTION TO PHYLOGENETIC ANALYSIS

Molecular phylogenetic analysis reconstructs the evolutionary history of different organisms. By analyzing DNA-sequences of different isolates or species, conclusions about evolutionary relationships can be drawn. These relationships can be presented in phylogenetic trees, which are bifurcating graphs consisting of nodes and branches, where only one branch connects any two adjacent nodes. The nodes represent the taxonomic units, which can be populations, individuals or single genes. The nodes can be either terminal or internal. The terminal nodes (also called leafs) represent the taxonomic units under comparison, called Operational Taxonomic Units (OTU) and the internal nodes represent the inferred ancestral units. Because we usually do not have data on those units they are referred to as Hypothetical Taxonomic Units (HTU). The input values of phylogenetic algorithms are usually DNA, RNA or protein sequences, which are prepared, sorted and aligned prior to analysis. It is, however, not always trivial how to correctly align multiple sequences. Furthermore, the quality of the alignments are highly critical for the quality and reliability of the resulting trees.

Several different techniques, theories and algorithms, typically based on mutations in the DNA, RNA or protein sequences, can be applied to construct phylogenetic trees. They all have in common that they construct trees or graphs that represent the evolutionary relationship or history of the different organisms, species, or isolates that are under investigation. Phylogenetic trees can be either rooted or unrooted.

All rooted trees have a particular node, from which a unique directed path leads to any other node in the tree. The root node is the HTU that is supposed to be the ancestor of all other HTU:s and OTU:s represented in the tree. By following the path from the root to any of the OTU:s, all evolutionary steps to that particular OTU will be passed. Rooted trees have n terminal nodes, n-1 internal nodes, n-2 internal branches and n external branches.

An unrooted tree may be considered more as a bifurcating graph than a tree. Similar to rooted trees, unrooted trees represent the evolutionary relationships among different OTU:s and HTU:s. However, since there is no root in the tree it does not illustrate in which order the different evolutionary steps took place. Hence, although unrooted trees represent evolutionary relationships, no conclusion about common ancestors can be drawn. An unrooted tree has n terminal nodes and n-2 internal nodes, n-3 internal branches and n external branches.

A limitation of most phylogenetic algorithms is that they do not take account of non tree-like evolutionary events. Recombination is such an event, which produces a child sequence by crossing two parent sequences. Recombinants are difficult to insert correctly in a phylogenetic tree if the tree is based on a genomic region consisting of segments from both parental genomes. Instead, there are two different correct trees that represent the evolutionary history, one for some segments of the genome and the other for the remaining part (Fig. 9).

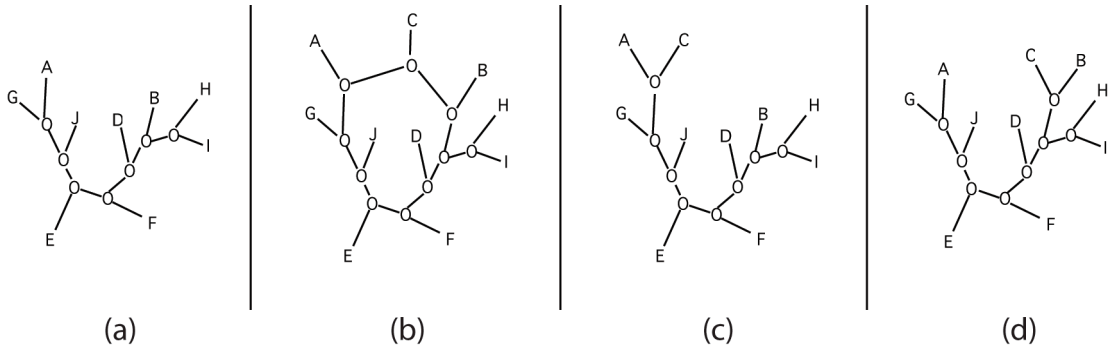

Fig. 9. *OTU A and B in tree (a) recombine to form C. A correct way to illustrate the evolutionary history would be like tree (b). Since this is not a legal bifurcating phylogenetic tree and the common algorithms are unable to construct such trees, two different trees may be constructed, (c) and (d). Each tree is correct and represents the evolutionary history of different segments.*

Algorithms and theories

Several different phylogenetic theories have been proposed to deal with how the evolutionary history should be reconstructed based on sequence data. The most common theories are based on neighbor joining and distance matrices, maximum parsimony, maximum likelihood and Bayesian inference. A limitation though, is that most algorithms based on those methods evaluate all possible different trees in an attempt to select the one that best represents the evolutionary history (given the theory the algorithm is based on). The number of different possible phylogenetic trees grows extremely rapidly when the number of OTU:s increases. The number of bifurcating unrooted trees (N_U) for n OTU:s is given by

$$N_U = (2n - 5)! / 2^{n-3}(n - 3)!$$

whereas the number of bifurcating rooted trees (N_R) for n OTU:s is given by

$$N_R = (2n - 3)! / 2^{n-2}(n - 2)!$$

which gives the equality

$$N_U(n) = N_R(n-1)$$

A consequence of the equation is that, for a set of only 20 OTU:s, nearly 10^{22} rooted trees exist that have to be evaluated! Since it is not uncommon with analyses of more than 100 OTU:s it is in fact impossible to test all possible trees even with a modern computer. To overcome this computational problem, additional efficient algorithms are necessary.

Step matrices

The rows and columns in a step matrix can consist of either the DNA letters A, T, C, G, the RNA letters A, U, C, G or the amino acids. The elements represent the minimal number of nucleotide substitutions required for a state in the column to the state in the row. The values in the amino acid matrix can vary between 1 and 3 depending on how many DNA substitutions that are required for that particular step. All matrices can also be weighted to reflect the different probabilities for each substitution to occur. Step matrices can be used to calculate the minimum number of substitutions from one OTU to another.

Distance matrices

The distance matrix method is based on the computed sequential distances between all pairs of taxonomic units. A sequential distance is usually based on the number of nucleotide substitutions or amino acid replacements between the two taxonomic units calculated using a step matrix. A tree-constructing algorithm can then be applied to those data.

One of the most commonly used algorithms based on distance matrices is the *unweighted pair-group method with arithmetic means (UPGMA)* (Sokal & Michener, 1958). UPGMA is a sequential clustering algorithm, which builds the tree in a stepwise manner. First, the algorithm locates the two OTU:s that are most similar to each other with regard to the distance between them. These OUT:s are then considered as one single OTU and the distance matrix is recalculated and the next two most similar OTU:s are chosen. In this way the tree is constructed stepwise until two OTU:s remain. These are connected to a root. The branching point between two OTU:s is calculated as follows:

$l_{ij} = d_{ij} / 2$ and $l_{(i)(jm)} = ((d_{ij} + d_{im}) / 2) / 2$

Algorithm: UPGMA

Initialization:

Assign each sequence *i* to its own cluster C

Define one leaf of *T* for each sequence, and place at height zero.

Iteration:

Determine the two clusters *i, j* for which d_{ij} is minimal.

Define a new cluster *k* by $C_k = C_i \cup C_j$, and define d_{kl} for all *l* by $d_{kl} = (d_{il}|C_i|+d_{jl}|C_j|) / (|C_i|+|C_j|)$.

Define a node *k* with daughter nodes *i* and *j*, and place it at height $d_{ij}/2$.

Add *k* to the current clusters and remove *i* and *j*.

Termination:

When only two clusters *i, j* remain, place the root at height $d_{ij}/2$.

UPGMA is one of the few methods that give a rooted tree as a result. Unfortunately, the algorithm does not always produce the correct tree regarding distances in evolution as the distance used in two clustered OTU:s is calculated as a mean value of the two. That is, the algorithm does not take into account the possibility of unequal substitution rates along the different branches.

To overcome problems with unequal substitution rates, a correction method called *transformed distance method* has been proposed (Farris, 1977; Klotz et al., 1979). The idea is to use an *outgroup* as a reference in order to make corrections for the unequal rates of evolution along the lineages. UPGMA is then applied to the new distance matrix to calculate the topology of the tree. The *outgroup* is an OTU or a group of OTU:s that is known to have diverged from the common ancestor prior to the rest of the OTU:s which is called the *ingroup* taxa. The distance is calculated as

$$d'_{ij} = (d_{ij} - d_{iD} - d_{jD}) / 2\, d''_{D} ,$$

where d'_{ij} is the transformed distance between OTU *i* and *j* and d''_{D} is the correction term regarding the outgroup. The latter can be calculated as

$$d''_D = \sum_{k=1}^{n} dkD / n ,$$

where n is the number of OTU:s in the ingroup. The reason for the existence of d''_D is to avoid negative distances.

However, one problem remains; which OTU:s should be placed in the outgroup? To solve this problem a two-step method has been proposed (Li, 1981). First, the topology of the tree is calculated with UPGMA. Then the taxa on one side of the root are used as an outgroup to calculate the correct topology of the other side. The same operation is then applied with the other side used as the outgroup.

Another commonly used algorithm is the neighbour joining algorithm (Saitou & Nei, 1987), which also uses distance matrices to calculate the distance between two OTU:s. The neighbour joining algorithm produces unrooted trees.

Algorithm: Neighbour joining

Initialization:

Define T to be the set of nodes, one for each given sequence,
and set $L = T$.

Iteration:

Pick a pair i, j in L for which D_{ij} is minimal.
Define a new node k and set $d_{km}=1/2\ (d_{im}+d_{jm}-d_{ij})$, for all m in L.
Add k to T with edges of lengths $d_{ik}=1/2\ (d_{ij}+r_i-r_j)$, $d_{jk}=d_{ij}-d_{ik}$, joining k to i and j.
Remove i and j from L and add k.

Termination:

When L consists of two leaves i and j add the remaining edge between i and j,
with length d_{ij}.

Maximum parsimony

Several parsimony methods have been developed for handling different types of data (Eck & 1966, 1966; Felsenstein, 1982; Fitch, 1977). The idea of maximum parsimony is to identify the tree that requires the lowest number of substitutions along the paths from the root to the OTU:s. The maximum parsimony method is based on so-called informative sites, which are found by performing a multiple alignment over the different sequences followed by the localization of variable sites, i.e. the sites where the characters differ between the sequences. Variable sites can be informative or uninformative. A site is phylogenetically informative if and only if it favors a subset of trees over the other possible trees. The maximum parsimony tree, of all possible trees, is calculated as follows:

1. Locate the informative sites.
2. For each possible tree, calculate the minimum number of substitutions required at each informative site.
3. Sum the number of changes over all informative sites for each possible tree.
4. Choose the tree that requires the smallest number of substitutions.

Sometimes, two or more trees with the same (lowest) number of changes will be identified. These trees are called equally parsimonious. The total number of substitutions (at informative and uninformative sites) in a tree is called the tree length.

Although this method may give interesting results regarding biological aspects, the algorithm handles all substitutions equally. Typically, substitutions occur with different probabilities, e.g. transitions are more common than transversions. A way to include different probabilities of occurrence is to give some substitutions a higher weight than others and construct the tree with respect to these weights. This method is referred to as *weighted parsimony* and tends to result in different and more evolutionary accurate trees than *unweighted parsimony*.

Algorithm: Weighted parsimony

Initialization:

Set $k = 2n-1$, the number of the root node.

Recursion:

Compute $S_k(a)$ for all a as follows:

If k is a leaf node:

Set $S_k(a) = 0$ for $a = x^k_u$ otherwise $S_k(a) = \infty$.

If k is not a leaf node:

Compute $S_i(a)$, $S_j(a)$ for all a at the daughter nodes i, j, and define $S_k(a) = \min_b(S_i(b)+S(a,b))+\min_b(S_j(b)+S(a,b))$.

Termination:

Minimal cost of tree = $\min_a S_{2n-1}(a)$.'

Although the *exhaustive search* described above always gives the correct maximum parsimony tree, a major problem is that the method is time-consuming in practice, since we usually are interested in more than just a few taxa (the maximum number of taxa that are virtually possible to include in an exhaustive search is approximately 12 with today's computers). If the need for the absolute optimal tree is essential, a simple branch and bound algorithm (Hendy & Penny, 1982) may be applied to decrease the number of possible trees to evaluate. The upper bound is typically calculated as the minimum number of substitutions (L) for a tree obtained from a faster algorithm (for example a distance algorithm). A tree can then be excluded if the number of substitutions is higher than L (often before the tree is fully constructed), and all the sub-trees that will grow from that tree do not need further evaluation. If any of the calculated complete trees has a number of substitutions that are lower than L during the evaluation, assign L to this new number. The maximum parsimony algorithm with branch and bound optimization can be used to find the maximum parsimony tree for up to 20 OTU:s in a reasonable time with modern computers.

When more then 20 OTU:s are included in the maximum parsimony analysis, more sophisticated algorithms are needed to speed up the process. Although faster algorithms may not always find the absolute optimal tree, several methods exist that produce “reasonably good” trees, i.e. trees that are likely to be the optimal parsimony trees but may also be good but not optimal. *Heuristic search algorithms* usually start with a tree obtained from some basic algorithm (for example the distance algorithm) as an initial tree. This step is typically followed by the examination of a subset of all trees that have a similar topology as the initial tree. Such algorithms will often result in a better

tree than the initial one, but most likely the algorithm will end up in a so-called local minimum. This is due to the fact that the most optimal tree for diverged sequences typically has a very different topology from the initial tree. An effective method that has been proposed to overcome the problem with local minimum is branch swapping. Randomized algorithms can be used to choose how and when to swap the branches (like the MCMC algorithm). However, some of the unsolved and difficult problems are how to give the different weights and running time, and when to perform large or small swaps. Modern maximum parsimony algorithms usually include a mix of different advanced methods to speed up the running time while keeping the accuracy of the results at a sufficiently high level.

Maximum likelihood

The maximum likelihood method was first developed for gene frequency data (Cavalli-Sforza & Edwards, 1967) and later for nucleotide and amino acid sequences (Felsenstein, 1973; Felsenstein, 1981). (Li, 1997; Swofford et al., 1996). Maximum likelihood is a probabilistic approach to phylogeny and is a time consuming method where the probability of observing the nucleotide sequences under a given tree is calculated. The likelihood (L) is calculated as

$$L = P(data \mid tree),$$

where the data is typically aligned DNA sequences. Modern maximum likelihood algorithms also include a model and additional parameters of evolution, which attribute each substitution a certain probability. This model and parameters can include a wide range of different properties such as unequal substitution rates, unequal expected frequency of the nucleotides (for DNA sequences), unequal rates of transitions and transversions and constant or gamma-distributed rates among sites. The likelihood (L) with an evolutionary model included is calculated as

$$L = P(data \mid tree, model).$$

At each site the probability of all possible reconstructions of ancestral states is calculated. Since L is calculated as the product of all individual likelihoods, the calculating computer will end up with numbers that are difficult to handle. To overcome this the log(L) is calculated, which gives a summation over all individual likelihoods, which is easier to handle. The exhaustive likelihood method can only handle a relatively small number of taxa. However, similar methods to those described above for the maximum

parsimony method to speed up the algorithm may also be applied to maximum likelihood algorithms. A simple maximum likelihood algorithm is Felsenstein's algorithm for likelihood.

Algorithm: Felsenstein´s algorithm for likelihood

Initialization:

Set $k = 2n-1$.

Recursion:

Compute $P(L_k|a)$ for all a as follows:

If k is a leaf node:

Set $P(L_k|a) = 1$ if $a = x^k_u$, $P(L_k|a) = 0$ if $a \neq x^k_u$.

If k is not a leaf node:

Compute $P(L_i|a)$, $P(L_j|a)$ for all a at the daughter nodes i, j, and set $P(L_k|a) = \sum_{b,c} P(b|a,t_i)P(L_i|b)P(c|a,t_j)P(L_j|c)$.

Termination:

Likelihood at site $u = P(x^*_u|T,t_*) = \sum_a P(L_{2n-1}|a)q_a$.

Bayesian inference

A phylogenetic method of growing popularity is the Bayesian inference method. In contrast to likelihood, which is the probability of the observed data given a tree and a model, Bayesian inference of phylogeny is based upon a quantity called the posterior probability distribution of trees, which is the probability of a certain tree and a model given the data. Using Bayes' formula, this is stated as

$$P(tree, model \mid data) = (P(tree, model) * \underline{P(data \mid tree, model)}) / P(data),$$

where the underlined part is the likelihood described above. As for likelihood, the posterior probability distribution of trees is impossible to calculate in reality for all possible trees. Most (or all) phylogenetic programs based on Bayesian inference instead perform simulations by using sophisticated Markov Chain Monte Carlo (MCMC) algorithms for an approximation of the posterior probabilities of trees. If this algorithm runs for a "sufficiently long time", the algorithm will end up with the optimal result. Although well-designed MCMC algorithms usually present good results, there is always a risk of ending up in "local optima". One of the solutions proposed to avoid local optima and to speed up the running time is the use of Metropolis

Coupled MCMC (MCMCMC or MC^3). MC^3 algorithms run several chains simultaneously, of which one is "normal" and the other makes bigger steps in an attempt to locate other optima. One of the advantages of using algorithms based on Bayesian inference is that the result also gives an estimation of the robustness of the resulting tree and, hence, no bootstrapping is necessary. However, the reliability of these estimations have been questioned, especially when concatenated genomic segments are used (Suzuki et al., 2002). One of the first research groups implementing Bayesian inference into phylogeny was Huelsenbeck et. al. (Huelsenbeck & Ronquist, 2001; Huelsenbeck et al., 2001), and they stated that *"Bayesian inference is roughly equivalent to maximum likelihood analysis with bootstrapping, but much faster"*. For reviews of Bayesian inference on phylogeny see Holder & Lewis (2003) and Lewis (2001)

Algorithm: Simplified MCMC algorithm for Bayesian inference of phylogeny

Initialization:
    Start with a random tree and parameters.
Recursion:
    (Randomly, according to predefined rules about MCMC graph) choose
        either new tree or new parameters.
        If P(new tree) > P(current tree)
            accept move.
        If P(new tree) < P(current tree)
            accept move with a probability of P(new tree) / P(current tree)
    Every *k* generation, save tree and all parameters.
Termination:
    After *n* generations, summarize samples using histograms, means,
        credibility intervals, etc.

Phylogenetic networks

As described above, evolutionary relationships between organisms are typically based on sequence data and represented as bifurcating phylogenetic trees. However, a limitation with most traditional phylogenetic algorithms is the assumption that the evolutionary history only consists of mutations and speciation. If the evolutionary history is more complex and involves reticulate events, such as recombination, horizontal gene transfer, hybridization, gene

duplication or loss of genes, traditional phylogenetic trees tend to be inadequate to illustrate the evolution.

Several methods have been proposed to detect recombinants using traditional phylogenetic algorithms applied to sub-genomic regions using sliding window protocols (see for example Lole et al., 1999). However, such algorithms are usually restricted to evaluate single or a few recombinant candidates among a set of non-recombinants, or present evidence for recombination but not the evolutionary history. The more complex the pattern of recombination events and crossovers is in the dataset, the more sophisticated algorithms are needed due to the incompatible signals in the data set. Despite this complexness, algorithms constructing phylogenetic or reticulate networks, may be applied to construct a visual representation of the evolutionary relationships among taxa with a history of recombination. There are two types of reticulate networks, hybridization networks and recombination networks, where recombination networks are used to describe evolution in the presence of recombination events. Recombination networks are based on binary sequences, which can be used to illustrate DNA sequences (Huson & Kloepper, 2005). A newly developed software for construction of phylogenetic networks is the SplitsTree4 program (Huson, 1998; Huson & Bryant, 2006).

Rooting unrooted trees

Most algorithms produce unrooted trees, which do not give the evolution a direction in time, but rather an evolutionary relationship between the taxa. To be able to decide where to place the root, an outgroup may be included in the analysis. The outgroup should be evolutionarily conserved and separated from, but still not too distant from, the ingroup. The root is then placed between the outgroup and the node connecting to the ingroup. If it is possible to find more than one outgroup this will usually lead to a more reliable result.

Sometimes it can be difficult or even impossible to select an appropriate outgroup. A (poor) solution to this problem is to assume that the rate of evolution has been approximately uniform over all the branches. The root is then placed in the midpoint of the longest pathway between two OTU:s.

Reliability of a tree

The definition “reliability” of a certain tree has two meanings. First, the reliability may refer to the correctness of the tree regarding the applied method (i.e. if it really is the maximum parsimonious tree or not). Second, although the tree is optimal, or close to optimal, in the first respect, does the tree reflect the correct evolutionary history? There are several methods to validate the robustness of a certain tree, of which the bootstrap method is the most commonly used.

Bootstrapping

The bootstrap method (Efron, 1982) is a statistic method to calculate a value on reliability. The method was introduced to phylogenetics by Felsenstein (1985) and is widely used as a method of assessing the significance of some phylogenetic feature, such as the segregation of a particular set of species on their own branch. Given an original dataset consisting of a multiple alignment, the bootstrap method generates a number of new sets (normally 100 to 1000) of the same size as the original set by resampling. The new sets are constructed in the following way: Randomly pick columns from the original set and put these in the new artificial sets until these are of the same size as the original set. A consequence of random sampling (due to normal sampling error) is that one specific column can appear multiple times in one artificial dataset and not at all in another. A phylogenetic algorithm may then be applied to the new sets, one at a time, to construct new trees. The frequency by which a phylogenetic feature appears among the artificial trees is then shown as bootstrap values in the original tree, or alternatively, in the consensus tree of all new trees. The bootstrap values shown in each HTU are the number of trees, often expressed in percent, that have the same OTU:s in its sub-tree as the corresponding node in the original tree. Typically, a sub-tree supported by a bootstrap value above 70 is considered as relatively robust.

Summary of phylogenetic methods

Each method described above has its own pros and cons and no single method is clearly superior for all data sets. Depending on the complexness of the input data, more or less sophisticated algorithms may be used and it is advantageous to apply several methods for comparison. Another aspect is the running time of the calculation. The more diverged the input data is, the more

the results tend to differ between the different algorithms, and hence the more sophisticated methods are needed. The most sophisticated methods available are probably the maximum likelihood and the Bayesian inference methods with appropriate and well-estimated evolutionary models included in the analysis. These are, on the other hand, the most time consuming methods and using the distance matrix algorithms might be an effective and rapid way to get an indication of the phylogenetic topology or an initial state for a more advanced algorithm. However, for diverged and complex datasets, the topology obtained from distance matrix algorithms may be very different from a topology obtained from a maximum likelihood algorithm including advanced evolutionary models. In addition, most traditional algorithms are not adequate for analysing taxa with a history of recombination, which demands more sophisticated algorithms designed for analysing recombinants. Although bootstrapping increases the computational burden, it is a very good method to estimate the robustness of a certain tree. A tree in which all HTU:s are supported by poor bootstrap values says very little about the evolutionary relationship among, or history of, the OUT:s. However, it should be kept in mind that a tree based on an inappropriate algorithm or evolutionary model may still be a poor representative of the true evolutionary history even though the bootstrap values are high. However, by using Bayesian inference the need of bootstrapping is obliterated, which might ease the computational burden. For an evaluation of bootstrapping and Bayesian posterior probabilities see Erixon et al. (2003).

Aims

The aims of the studies presented here were to increase the understanding of the evolution of human alpha-herpesviruses, more specific, to

- Describe genetic variability of clinical isolates collected from different geographic regions and from patients with different clinical entities.
- Perform phylogenetic analyses based on selected genomic regions to reconstruct the evolutionary history and investigate responsible mechanisms thereof.

RESULTS AND DISCUSSION

General considerations

Herpesviruses are among the most extensively studied DNA viruses, and the evolutionary relationships among the different herpesviruses infecting humans, reptiles and other vertebrates as well as invertebrates have been investigated and reported in several studies (for a detailed review, see McGeoch et al., 2006). However, most studies have focused on the evolutionary relationships among different herpesviruses, which can be traced back ten to hundreds of million years ago. Although genetic variation and classification into different genogroups have been described for limited genomic regions of clinical isolates from certain herpesviruses such as VZV (Muir et al., 2002), EBV (Sample et al., 1990), HCMV (Chou, 1992; Chou & Dennison, 1991) HHV 6 (Clark, 2000), HHV 7 (Franti et al., 1998) and HHV 8 (Meng et al., 1999), data on genetic variability based on DNA sequencing of clinical HSV-1 and HSV-2 isolates are limited. In addition, complete genome analysis of VZV has hitherto not been described.

In the present study, selected genomic regions of clinical HSV-1 and HSV-2 isolates, as well as the complete genome for clinical VZV isolates, were sequenced in an attempt to increase the knowledge about genetic variability, evolution and evolutionary mechanisms.

Herpes simplex virus type 1 (paper I, II and III)

In paper I, 28 clinical HSV-1 isolates were collected from male and female patients in Sweden, suffering from oral lesions, genital lesions or encephalitis. Sequence comparison and phylogenetic analysis of the genes US4, US7 and US8 revealed three genotypes, arbitrarily designated as A, B and C, supported by high bootstrap values. The genetic distance between the most distant isolates was approximately 2% (on average of the three investigated genes). Several isolates were also identified as recombinants derived from isolates from the different genotypes. The recombinants were classified by observing phylogenetic topologies from different genes as well as form the same gene by using the BoosScan method. Approximately 20% of the investigated isolates were recombinants, and crossovers were detected within as well as between the genes US4, US7 and US8. Furthermore, evidences of recent as well as ancient recombination events were found. In

addition, tandem repeat (TR) regions were detected in all investigated genes and were shown to be polymorphic among clinical isolates. The TR region of US7 (the gI gene) is likely to be a result of homologous recombination events with unequal crossovers, and codes for two to eight repeats of three blocks of the amino acids serine, threonine and proline (aa´s STPSTTT, STPSTTI or PAPSTTI). These residues are typical constituents for *O*-linked glycosylation, suggesting that the region functions as a mucin tract. The gI sequences for the alpha-herpesviruses HSV-2, VZV, simian varicella virus, pseudorabies virus (PrV), equine herpesvirus 1 (EHV-1) and 4 (EHV-4), bovine herpesvirus 1 and monkey B virus were analysed using the "Tandem Repeats Finder" program (Benson, 1999). A tandem repeat region was only detected for monkey B virus comprising 2.8 copy numbers with a period size of 33 nucleotides at position nt 686-778. However, this region does not code for residues involved in a potential mucin tract (aa´s AAPPTPGAEGT). Thus, the polymorphic TR region of gI therefore seems to be an exclusive feature for HSV-1 among the alpha-herpesviruses.

Simple methods for genotyping clinical isolates may be of interest as a screening tool in the search of associations between genotype and clinical manifestations and in epidemiological studies from different geographical regions. For such purposes, a rapid genotyping method distinguishing between the three HSV-1 genotypes A, B and C was developed (paper II). Two targets for genotyping were selected (US4 and US7). With this approach it was also possible to detect recombinants with different genotypic origins. The method utilizes single nucleotide substitutions specific for each genotype and is based on restriction enzyme cleavage of PCR amplicons from US4 and US7. Two restriction enzymes, using the same buffer and temperature, were selected for each PCR product. The cleaved products were finally separated on a metaphor agarose gel. The cleavage patterns are similar for both genes; a single band for genotype A isolates, two bands for genotype B isolates and three bands for genotype C isolates.

Mucin genes contain regions coding for a high number of the amino acids serine (S), threonine (T) and proline (P), of which S and T are suitable targets for O-linked glycosylation. In this study (paper III) the potential mucin region in HSV-1 gI was examined to investigate the variability of the TR region and possible O-glycosylation.

To investigate the evolutionary rate of duplications and deletions in the TR region, the number of repeats was compared with the phylogenetic topology of a tree based on the US7 gene excluding the TR region. However, no correlation was detected between genotype identity of the clinical isolates

and the number of repeated blocks in the TR region, indicating that the evolution in the TR region is more rapid compared with nucleotide substitutions in the US7 gene (Fig. 10).

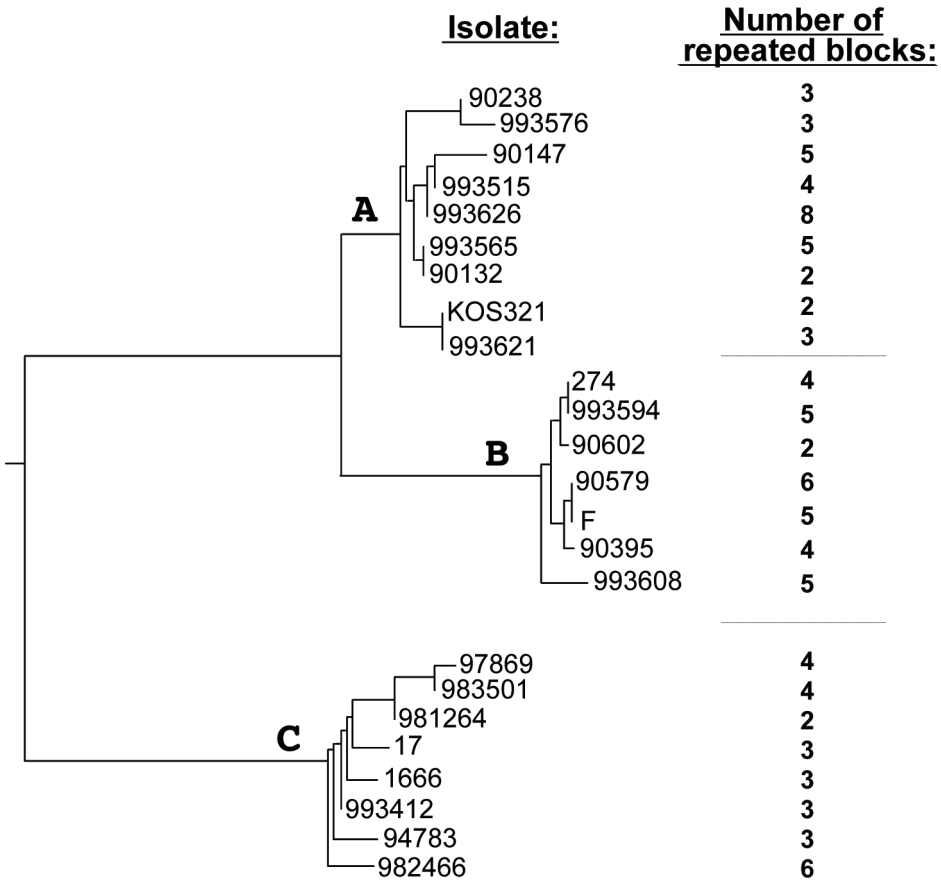

Fig. 10. *Phylogenetic tree of clinical HSV-1 isolates, as well as the laboratory strains F, KOS321 and 17, and the number of repeated blocks in the TR-region in US7.*

A western blot of virus infected cells, using a monospecific rabbit anti-gI serum, of the gI protein from isolates presenting 2, 4, 6 and 8 DNA repeats was performed. The isolate with two repeats presented the lowest molecular weight, followed by the isolates with 4, 6, and 8 repeats. In addition, an *in vitro* assay was set up to investigate whether a synthetic peptide (KPNPA STPSTTT STPSTTT PAPK) with two repeats (underlined) of the TR region could be utilized as substrate for O-linked glycosylation for individual recombinantly produced human GalNAc-T1,-T2,-T3,-T4 and T11 transferases. The results showed that the peptide is a universal substrate for

massive O-linked glycosylation, not only for the two most commonly expressed GalNAc-T1 and -T2, but also for GalNAc-T3, -T4 and -T11. Furthermore, an additional western blot was performed to investigate which kinds of saccharides that might be attached to the TR region. The results hereof indicated that monosaccharides, but not disaccharides, are attached to the gI. It was concluded that the TR codes for a mucin region. Such a polymorphic mucin region has not previously been described for alpha-herpesviruses.

Herpes simplex virus type 2 (paper IV)

Genetic variability and phylogenetic analyses of clinical HSV-2 isolates have not previously been described. In the present study (paper IV) a total of 45 clinical HSV-2 isolates were collected and compared with the laboratory strains HG52 and B4327. Twentysix isolates were collected in Dar es Salaam in Tanzania, ten isolates were collected in Bergen, Norway and nine isolates were collected in Gothenburg, Sweden. The genes US4, US7 and US8 as well as the non-coding region between US7 and US8 were sequenced for all isolates.

The sequences were aligned an analyzed using the maximum likelihood and maximum parsimony methods applied on 100 bootstrap replicates. In addition, recombination analyses were performed by using the SplitsTree program for constructing phylogenetic networks, the BootScan method included in the simplot program and the phi-test for statistical test for recombination (Bruen et al., 2006).

Sequence comparison and phylogenetic analysis based on the complete concatenated sequence alignments of US4 and US7 to US8 revealed a star phylogeny and a different genetic divergence than described for HSV-1. However, although the genetic distances were shorter than for HSV-1, two main clusters, arbitrarily designated A and E, were present in the tree. The divergence into the two clusters was not as obvious as the different genotypes described for HSV-1, but was supported by high bootstrap values. The A cluster consisted of isolates collected in Africa and the E cluster contained isolates collected in Europe. In addition, the E type was also frequently present in Tanzania, probably due to an ancient co-existence, or as a result of a recent introduction of the E type into this region. In a further analysis, two other trees were based solely on US4 or the US7 to US8 genomic region. The phylogenetic topologies obtained were similar to the tree based on the

complete concatenated sequences, except that three isolates belonging to the E cluster in the tree based on US4, belonged to the A cluster in the tree based on US7 to US8. After analyzing those three isolates separately by using the BootScan method, they were classified as recombinants derived from viruses from the A and E clusters. Furthermore, after extensive recombination analyses including all isolates, it was concluded that all or most HSV-2 isolates are recombinants derived from isolates within each cluster. There are several lines of evidence for this conclusion; (i) the phylogenetic analysis results in a star phylogeny that is typical for organisms that have been under the influence of free and random recombination, i.e. the wide and star-shaped genetic variability within each cluster may be explained by previous existence of genetic subtypes, which are now difficult to detect due to the high rate of recombinants, (ii) phylogenetic network analysis present reticulate topologies, which is consistent with recombination, (iii) three isolates were classified as probable recombinants by using the BootScan method, and (iv) the phi-test presented statistically significant evidence for recombination.

Varicella Zoster virus (paper V)

Previously published studies on genetic variability of clinical VZV isolates have proposed a divergence into three distinct genotypes of which one was suggested to be derived from a recombinant strain. In addition, a recent study designated those genotypes as European (E), Japanese (J) and mixed (M) (Loparev et al., 2004). However, these studies were only based on short genomic regions with few nucleotide substitutions. In the present study (paper V), the complete genome of two selected isolates belonging to different clusters of the M-genotype were sequenced and compared with the European laboratory strains BC, DR and DUMAS as well as with the Japanese laboratory strain p-Oka and the vaccine strain v-Oka. The strain v-Oka is a live vaccine strain, which has been attenuated during passages in cell culture (Oxman et al., 2005), and is now globally introduced as a vaccine against VZV.

A complete genomic phylogenetic analysis of the isolates revealed a divergence into four distinct genotypes, E, J, M1 and M2. However, comparison of phylogenetically informative sites in the complete genome indicated that the topology obtained from the phylogenetic analysis based on the complete genome was not uniformly supported, indicating recombination. To gain further support for recombination events in the evolutionary history of VZV, and to investigate which isolates that might be recombinants, several

sub-genomic analyses were performed. Phylogenetic analyses on different genomic regions revealed that the phylogenetic topology differed and was dependent on which genomic region that was analyzed. The conclusions drawn from these results were that of the four genotypes E, J, M1 and M2, the genotypes M1 and M2 are recombinants derived from E and J at different occasions. Furthermore, the lengths of the branches separating M1 and M2 from E and J in the phylogenetic tree indicate that these are derived from ancient recombination events followed by independent evolution of the four genotypes.

Recombination

The mechanisms responsible for recombination in herpesviruses are poorly understood but are associated with DNA replication (Umene, 1999) and different cell factors (Dutch *et al.*, 1995; Taylor & Knipe, 2004). Several studies have shown that homologous recombination occurs frequently under experimental conditions (Meurens et al., 2004a; Schynts et al., 2003) and for example, HSV-1 recombinants have been detected *in vitro* (Brown & Ritchie, 1975; Brown *et al.*, 1992; Honess *et al.*, 1980; Umene, 1985) as well as in animal models (Lingen et al., 1997; Wildy, 1955). Recombination has also been shown experimentally between varicelloviruses, i.e. VZV, PrV, bovine herpesvirus-1 and feline herpesvirus (Dohner et al., 1988; Fujita et al., 1998; Henderson et al., 1990; Schynts et al., 2003). Furthermore, recombinants were recently demonstrated between bovine herpesvirus 1 mutants after co-inoculation of calves by the natural route of infection (Schynts et al., 2003).

Although recombination may be relatively easy to study under laboratory conditions, it is more difficult under natural conditions. However, wild-type recombinants have recently been described for HSV-1 (Bowden et al., 2004), VZV (Muir et al., 2002) and PrV (Christensen & Lomniczi, 1993). Based on clinical data, the results presented in this thesis (Paper I, IV and V) clearly demonstrate a divergence into different genotypes. By defining genetic markers specific for each genotype, these could be utilized as markers in the search for recombinants. Furthermore, by applying different algorithms for recombination analysis on sequences from clinical isolates, it was here possible to perform a thorough analysis of recombination under natural conditions. The results support the conclusion that recombination has been, and is, a prominent feature of the evolution of the human alpha-herpesviruses HSV-1, HSV-2 and VZV. These conclusions are based on data supporting ancient as well as recent recombination events within all three studied viruses

HSV-1, HSV-2 and VZV.

The frequency of recombination in herpesviruses has been shown to be high *in vitro* with a rate of 25% – 32% for bovine herpesvirus 1 (Meurens et al., 2004b) and 10% – 21% for feline herpesvirus 1 (Fujita et al., 1998). The frequency of detected recombinants among clinical isolates in this study was 20% for HSV-1 (paper I), 50% for VZV (old recombination events which created the genotypes M1 and M2, paper V) and probably almost 100% for the HSV-2 isolates (paper IV). However, the high number of recombinants for VZV may be explained by only a few numbers of ancient recombination events, inherited to all members of the M1 and M2 genotypes. Furthermore, in an ongoing study (Norberg et al., unpublished), including 200 clinical isolates collected from patients suffering from HSV-1 induced keratitis, were investigated. All isolates were genotyped in the US4 and US7 genes using the genotyping method described in paper II. Despite the short distance between US4 and US7, as much as 30% of the isolates could be classified as recombinants due to different genotype classification in the two genes. In an additional study (Norberg et al., unpublished), an *in vitro* assay has been set up in an attempt to investigate the recombination frequency of HSV-1 in two different cell types; human fibroblast cells and a DNA-ligase 1 defect cell line. Each cell line was infected with a mix of three viruses, one from each HSV-1 genotype A, B and C. The viruses were passaged ten times and plaque purified. After genotyping of the US4 and US7 genes using the method described in paper II, five of 25 plaques from the fibroblast cell line and eight of 25 plaques from the DNA-ligase 1 defect cell line contained recombinant viruses presenting different genotype identity in US4 and US7, i.e. the recombinant frequency of HSV-1 after ten passages in cell-culture ranged from 20% to 32%.

However, it is likely that the number of recombinant viruses, at least for HSV-1, is underestimated. The methods described in this study for detecting recombinants within HSV-1 and HSV-2 are mainly based on short genomic regions and specific markers. A limitation of such analysis is that recombination can either occur outside the investigated genomic region, or the recombination event can produce an even number of crossovers between two markers implying that both markers in the recombinant will be inherited from the same parental strains. In addition, recombination can also occur between strains that differ in specific nucleotides, but present identical markers (for example two HSV-1 isolates from the same genotype). In either case the recombination events will be undetected. It is only by the comparison of complete genome sequences that all recombination events can be detected (except, of course, for recombination events involving completely identical

sequences, which will still remain undetected).

A prerequisite for homologous recombination is a certain amount of sequence homology of the recombining strains. However, the more diverse the strains are, the higher evolutionary impact may the recombinant have. Besides sequence homology, another variable that needs consideration is the triggering factors for replication, i.e. if the different viruses replicate simultaneously in the cell, which is necessary for homologous recombination to occur. Despite relatively high sequence divergence between different herpesviruses, recombination has been demonstrated experimentally *in vitro* and *in vivo* between HSV-1 and HSV-2 (Esparza *et al.*, 1976; Halliburton, 1980; Halliburton *et al.*, 1977; Preston *et al.*, 1978; Timbury & Subak-Sharpe, 1973; Yirrell *et al.*, 1992). In contrast, no recombination was reported between HSV-1 and BoHV-2 or PrV (Halliburton, 1980; Halliburton *et al.*, 1977; Timbury & Subak-Sharpe, 1973; Yirrell *et al.*, 1992). However, interspecific recombinants rarely survive under natural conditions and none of the investigated HSV-1, HSV-2 or VZV isolates in this study was classified as an interspecific recombinant. In fact, the only natural interspecific recombinant hitherto detected is EHV-1, which was recently described as a recombinant between progenitors of EHV-1 and EHV-4 (Pagamjav *et al.*, 2005).

Genetic distance and evolutionary timescale

Although genogroups were detected for HSV-1, HSV-2 and VZV, the intergenotypical variability between these viruses differed almost by a factor of ten. The most diverged virus was HSV-1 with a genetic distance of approximately 2 %, and the least diverged virus was VZV with a variability of only 0.3 %. HSV-2 was less variable than HSV-1, but slightly more variable than VZV with a variability of approximately 0.4%.

An intriguing question is when the described events, i.e. the divergence into different genogroups, occured. An evolutionary timescale for herpesviruses has previously been proposed (McGeoch & Cook, 1994; McGeoch *et al.*, 1995). The estimated timescale was based on paleontological data using divergence times of mammalian taxa and included analyses of several conserved genomic regions among the different herpesviruses. The divergence of HSV-1 and HSV-2 was proposed to be dated 8.5 to 8.4 Myears before present (BP).

In an ongoing study (Norberg, et al., unpublished), efforts were made to

estimate a timescale of the different evolutionary events in the history of HSV-1, HSV-2 and VZV. The genetic distance between the HSV-1 genotypes, based on the US7 to US8 genomic region, was estimated by using the F84 model and the dnadist program included in the Phylip package. The Kimura and Jukes-Cantor models were used in parallel for comparison. Bowden et al. (2004) recently estimated the average mutation rate for HSV-1 to 5 x 10^{-8} nucleotide substitutions per site per year. By applying this mutation rate, the divergence of the genotypes AB and C was estimated to ~145,000 years BP and the divergence of genotypes A and B was estimated to ~84,000 years BP. However, it is unlikely that the substitution rate in the US7 to US8 region is equal to the average substitution rate in the HSV-1 genome. To define a more accurate estimation of the substitution rate, the genetic distance between the US7 to US8 region between HSV-1 and HSV-2 was calculated using the same method as described above. The choice of model, or inclusion of gamma-distributed rates across sites, had low impact on the result. Based on the assumption that HSV-1 and HSV-2 diverged 8.5 to 8.4 Myears BP (McGeoch et al., 1995), the refined substitution rate was estimated to 1.44 x 10^{-8} nucleotide substitutions per site per year. By applying this new substitution rate to the HSV-1 genotypes, the divergence of genotypes AB and C was estimated to ~500,000 years BP and the divergence of genotype A and B was estimated to ~445,000 years BP. In conclusion, the estimation of the divergence of the HSV-1 genotype AB and C varied from 145,000 to 500,000 years BP depending on which evolutionary rate that was applied. This estimation highly depends on the accuracy of the estimated substitution rate and since older estimations have been based on several simplifications and assumptions, the presented dates are to be considered as rough estimations. In addition, it seems unlikely that the substitution rate has been constant since the environmental conditions for HSV-1 and HSV-2 probably have varied during the time since divergence. Furthermore, several important factors differ among HSV-1, HSV-2, and VZV that may explain different evolutionary rates, such as differences in transmission of virus from host to host, frequency of reactivation, seroprevalence and replication rate. However, based on preliminary data, it seems likely that the divergence into the different HSV-1 genotypes not only occurred before the human emigration from Africa approximately 100.000 years BP, but also prior to the emergence of *Homo sapiens*, i.e. in *Homo heidelbergensis*. If so, the genetic variability described here may be effects of founder populations, evolutionary bottlenecks, natural selection and genetic drift in small populations within Africa and/or in other geographic regions occupied by *Homo heidelbergensis*.

Differences in recombination rates

A striking difference between HSV-1 and HSV-2 is that most or all HSV-2 isolates appear to be mosaic recombinants, while only 20%-30% of the clinical HSV-1 isolates presented evidence for recombination. Furthermore, the HSV-2 recombinants presented a significantly higher level of crossovers than the HSV-1 recombinants, indicating that each HSV-2 recombinant has been involved in higher numbers of recombination events. The underlying reasons for the high rate of recombination events in HSV-2 are speculative, but could be explained by differences in replication, transmission, or a higher probability of re-infections.

A prerequisite for recombination is that two viruses simultaneously replicate in the same cell, which requires a multi-strain infected host. Such infections can be obtained either by the transmission of a heterogeneous viral population, or by re-infection. In addition, it seems reasonable that a transmission of a high number of virus particles during infection facilitates both these scenarios. However, possible differences in transmission doses between HSV-1 and HSV-2 are hitherto unknown and need further study.

It is probable that HSV-2 has presented a similar genotypic divergence as HSV-1, which may be concealed due to a history of numerous recombination events. Assuming that HSV-1 and HSV-2 have the same tendency to recombine, a possible explanation for the higher recombination rate in HSV-2 may be that this virus has been allowed to recombine for a longer period of time than HSV-1. Furthermore, it is likely that the different genotypes of HSV-1 as well as of HSV-2 have been geographically separated during history, followed by a recent transmission between continents. Since HSV-1 spreads orally via direct contact it is usually transmitted within the family in an early stage of life. In contrast, HSV-2 is transmitted sexually, usually later in life. A somewhat speculative explanation may be that HSV-2 strains have been transmitted earlier between continents during ancient plundering expeditions and that the high rate of recombinants is a tragic remnant of rapes during such raids.

Biological implications of the presented results

Results in this thesis present a divergence into different genotypes for all investigated viruses HSV-1, HSV-2 and VZV. Although clinical HSV-1 isolates in this study were collected from patients with different clinical

symptoms, i.e. oral or genital lesions or encephalitis, no association was found between genotype identity in US4, US7 or US8 and clinical manifestation. Nevertheless, due to the high frequency of recombination it is not excluded that association between genotype and phenotype may be present in other loci of the HSV-1 genome. In addition, several other biological aspects, such as the capacity to infect different cell-types, reactivate, replicate and escape from the immune system, may depend on genotype identity.

There are several evolutionary and biological aspects to consider regarding recombination. Since nucleotide substitutions are rare events for herpesviruses, recombination may act as a powerful and essential driving force of evolution. Recombination can break down associations between deleterious and beneficial mutations at different loci (negative disequilibrium). A genome can collect beneficial mutations from several other genomes, which is advantageous when different individuals in a population carry different beneficial mutations (Felsenstein & Yokoyama, 1976). A consequence hereof is that recombination can increase the additive genetic variance and, by Fisher's fundamental theorem for natural selection, can increase the rate of adaptation (Edwards, 1994; Fisher, 1930). In addition, all organisms randomly introduce harmful mutations in their genomes and a recent study has demonstrated that recombination is a powerful mechanism for the deletion of such mutations (Keightley & Otto, 2006).

It is reasonable to believe that herpesviruses can benefit from recombination. It has, for example, experimentally been shown that two avirulent herpes simplex viruses may generate lethal recombinants *in vivo* (Javier et al., 1986). In addition, it is likely that most virus strains occasionally introduce harmful but not lethal mutations. Theoretically, new strains without any harmful mutations can be created by recombination. If two viruses, each containing harmful, neutral and beneficial mutations, infect the same cell and are allowed to recombine, several recombinants with different mosaic patterns may be created. By the randomness of crossovers, some recombinants will contain fewer harmful and more beneficial mutations than the others, and by natural selection, the best-fitted recombinants will have the highest chance to survive and reproduce. A direct consequence of such selection may be that recombinants are overall favored and selected for in a population, which may explain the high rate of detected wild-type recombinants in this study, especially for HSV-2. In fact, an additional explanation of the higher rate of recombinants among clinical HSV-2 isolates could be that HSV-2 may be more sensitive to mutations than HSV-1, and hence is more dependent on recombination to delete deleterious mutations.

Another aspect of recombination is selection of neutral mutations. If a virus by recombination acquires a genomic region with a beneficial mutation, this virus will typically be naturally selected and favored over less fit viruses lacking the beneficial mutation. However, in addition to the beneficial mutation, the genomic region might also contain several neutral mutations. Due to this "hitch-hiking", also neutral mutations can be favored in the evolutionary process. Such selection of neutral mutations may have consequences in studies searching for positive selection and for biologically important regions involving specific genes, since selection of neutral mutations might lead to the conclusion that also such mutations are beneficial and biologically important.

A live attenuated VZV vaccine strain (v-Oka), derived from the Japanese strain p-Oka during numerous passages in cell-culture, is currently introduced worldwide. In the present study (paper V) it was shown that VZV is divided into at least four different genotypes and that v-Oka belongs to genotype J. In addition, a recent study demonstrates that the three genotypes E, J and M are geographically separated with the E –type represented in Europe and North America, the J-type in northern Asia and the M-type represented in warmer climates, such as Africa and Central America (Loparev et al., 2004). By introducing a live, but attenuated, J-type strain in geographic regions where other genotypes are dominant, it is possible that novel and virulent J and E, or J and M1 or M2 recombinants are emerge. Although the genetic distances between the VZV genotypes are relatively small, possible consequences of such recombinants cannot be neglected.

ACKNOWLEDGEMENTS

I wish to express my sincere appreciation and gratitude to:

Jan-Åke Liljeqvist, my supervisor and friend, for endless support, scientific guidance, interesting discussions and for making my time as a PhD student instructive, interesting and highly enjoyable.

Tomas Bergström, my co-supervisor, for friendship, for always being encouraging and for essential scientific guidance.

Anette Roth, for invaluable and skillful technical assistance and for generating considerable parts of the DNA-data presented in this thesis.

Magnus Lindh, co-author (paper I), for technical support and interesting discussions about phylogeny.

Sigvard Olofsson, co-author (paper III), for all help and for instructive discussions about O-glycans.

Mads Agervig Tarp and Henrik Clausen, co-authors (paper III) at the University of Copenhagen, Denmark.

Mabula Kasubi and Lars Haarr, co-authors (paper IV) at the University of Bergen, Norway.

Vladimir Loparev, Scott Schmid and Scott Sammons, co-authors (paper V) at the Centers for Disease Control and Prevention, USA.

Elham Rekabdar, co-author (Paper I).

Maria Johansson, Carolina Gustafsson, Ann-Sofie Tylö and Mona Brantefjord, for skillful technical assistance, for always being friendly and for guiding me in the right directions in the lab.

Staffan Görander, for friendship, interesting discussions about everything, musical evenings, medical council and for a nice motorcycle trip to Smögen.

Edward Trybala and Kicki Bergefall, for technical support about virus purification.

Kristina Nyström, for technical assistance with RNA-purification and for

being a nice travel companion.

Per Elias, for collaboration, interesting discussions and revision of manuscripts.

Ka Wei Tang, for ongoing and future collaboration.

Henrik Nilsson, for technical assistance and interesting discussions about phylogeny.

Petra Tunbäck, for an interesting collaboration.

Gaby Helbok and Sabina Wagner, for safely guiding me through the bureaucratic jungle.

Dan Groth, for multi-disciplinary technical assistance.

Catherine Brinkley, for friendship and for linguistic revision.

The staff at the departments of tissue-culture and PCR, for technical assistance.

George Verjans, for an interesting collaboration.

Ana, Beata, Carla, Charles, Elin, Eric, Eva-Corina, Fredrik, Maria and Sebastian, for sharing the time as a PhD student and for interesting discussions about science and everyday life.

Ulla Lindhe and Åke Norberg, my parents, for encouragement, for always believing in me and for being excellent scientific role models.

Björn Norberg, my brother, for friendship, encouragement and for skillful technical assistance with illustrations.

Eva, Hugo and Elliot, my family for whom this thesis is dedicated, for endless love, support, and patience during the last months. Words cannot express how much I owe you for everything.

REFERENCES

Aymard, M. (2002). [Current epidemiology of herpes]. *Pathologie-biologie* **50**, 425-435.

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* **27**, 573-580.

Bowden, R., Sakaoka, H., Donnelly, 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.

Bower, J. R., Mao, H., Durishin, C., Rozenbom, E., Detwiler, M., Rempinski, D., Karban, T. L. & Rosenthal, K. S. (1999). Intrastrain variants of herpes simplex virus type 1 isolated from a neonate with fatal disseminated infection differ in the ICP34.5 gene, glycoprotein processing, and neuroinvasiveness. *J Virol* **73**, 3843-3853.

Brown, S. M. & Ritchie, D. A. (1975). Genetic studies with herpes simplex virus type 1. Analysis of mixed plaque-forming virus and its bearing on genetic recombination. *Virology* **64**, 32-42.

Brown, S. M., Subak-Sharpe, J. H., Harland, J. & MacLean, A. R. (1992). Analysis of intrastrain recombination in herpes simplex virus type 1 strain 17 and herpes simplex virus type 2 strain HG52 using restriction endonuclease sites as unselected markers and temperature-sensitive lesions as selected markers. *J Gen Virol* **73 (Pt 2)**, 293-301.

Bruen, T. C., Philippe, H. & Bryant, D. (2006). A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172**, 2665-2681.

Cavalli-Sforza, L. L. & Edwards, A. W. (1967). Phylogenetic analysis. Models and estimation procedures. *American journal of human genetics* **19**, Suppl 19:233+.

Chapman, T. L., You, I., Joseph, I. M., Bjorkman, P. J., Morrison, S. L. & Raghavan, M. (1999). Characterization of the interaction between the herpes simplex virus type I Fc receptor and immunoglobulin G. *The Journal of biological chemistry* **274**, 6911-6919.

Chee, M. S., Bankier, A. T., Beck, S., Bohni, R., Brown, C. M., Cerny, R., Horsnell, T., Hutchison, C. A., 3rd, Kouzarides, T., Martignetti, J. A. & et al. (1990). Analysis of the protein-coding content of the sequence of human cytomegalovirus strain AD169. *Current topics in microbiology and immunology* **154**, 125-169.

Chou, S. (1992). Comparative analysis of sequence variation in gp116 and gp55 components of glycoprotein B of human cytomegalovirus. *Virology* **188**, 388-390.

Chou, S. W. & Dennison, K. M. (1991). Analysis of interstrain variation in cytomegalovirus glycoprotein B sequences encoding neutralization-related epitopes. *J Infect Dis* **163**, 1229-1234.

Christensen, L. S. & Lomniczi, B. (1993). High frequency intergenomic recombination of suid herpesvirus 1 (SHV-1, Aujeszky's disease virus). *Archives of virology* **132**, 37-50.

Clark, D. A. (2000). Human herpesvirus 6. *Rev Med Virol* **10**, 155-173.

Collins, P. (1993). Mechanisms of herpesvirus resistance. *Annals of medicine* **25**, 441-445.

Crumpacker, C. S. (1988). Significance of resistance of herpes simplex virus to acyclovir. *Journal of the American Academy of Dermatology* **18**, 190-195.

Crute, J. J. & Lehman, I. R. (1989). Herpes simplex-1 DNA polymerase. Identification of an intrinsic 5'----3' exonuclease with ribonuclease H activity. *The Journal of biological chemistry* **264**, 19266-19270.

Darwin, C. (1859). The origin of species. *J Murray, London*.

Davison, A. J. (1983). DNA sequence of the US component of the varicella-zoster virus genome. *The EMBO journal* **2**, 2203-2209.

Davison, A. J. (2000). Molecular evolution of alphaherpesviruses. In: Arvin, A.M., Gershon, A.A. (Eds.), Varicella-Zoster Virus. *Cambridge University Press, Cambridge*, 25-50.

Davison, A. J., Dolan, A., Akter, P., Addison, C., Dargan, D. J., Alcendor, D. J., McGeoch, D. J. & Hayward, G. S. (2003). The human cytomegalovirus genome revisited: comparison with the chimpanzee cytomegalovirus genome. *J Gen Virol* **84**, 17-28.

Davison, A. J. & McGeoch, D. J. (1986). Evolutionary comparisons of the S segments in the genomes of herpes simplex virus type 1 and varicella-zoster virus. *J Gen Virol* **67 (Pt 4)**, 597-611.

Davison, A. J. & Scott, J. E. (1986). The complete DNA sequence of varicella-zoster virus. *J Gen Virol* **67 (Pt 9)**, 1759-1816.

Davison, A. J. & Taylor, P. (1987). Genetic relations between varicella-zoster virus and Epstein-Barr virus. *J Gen Virol* **68 (Pt 4)**, 1067-1079.

Davison, A. J. & Wilkie, N. M. (1981). Nucleotide sequences of the joint between the L and S segments of herpes simplex virus types 1 and 2. *J Gen Virol* **55**, 315-331.

Dingwell, K. S., Doering, L. C. & Johnson, D. C. (1995). Glycoproteins E and I facilitate neuron-to-neuron spread of herpes simplex virus. *J Virol* **69**, 7087-7098.

Dingwell, K. S. & Johnson, D. C. (1998). The herpes simplex virus gE-gI complex facilitates cell-to-cell spread and binds to components of cell junctions. *J Virol* **72**, 8933-8942.

Dohner, D. E., Adams, S. G. & Gelb, L. D. (1988). Recombination in tissue culture between varicella-zoster virus strains. *Journal of medical virology* **24**, 329-341.

Dolan, A., Jamieson, F. E., Cunningham, C., Barnett, B. C. & McGeoch, D. J. (1998). The genome sequence of herpes simplex virus type 2. *J Virol* **72**, 2010-2021.

Drosopoulos, W. C., Rezende, L. F., Wainberg, M. A. & Prasad, V. R. (1998). Virtues of being faithful: can we limit the genetic variation in human immunodeficiency virus? *Journal of molecular medicine (Berlin, Germany)* **76**, 604-612.

Dubin, G., Frank, I. & Friedman, H. M. (1990). Herpes simplex virus type 1 encodes two Fc receptors which have different binding characteristics for monomeric immunoglobulin G (IgG) and IgG complexes. *J Virol* **64**, 2725-2731.

Dutch, R. E., Bianchi, V. & Lehman, I. R. (1995). Herpes simplex virus type 1 DNA replication is specifically required for high-frequency homologous recombination between repeated sequences. *J Virol* **69**, 3084-3089.

Duus, K. M. & Grose, C. (1996). Multiple regulatory effects of varicella-zoster virus (VZV) gL on trafficking patterns and fusogenic properties of VZV gH. *Journal of virology* **70**, 8961-8971.

Duus, K. M., Hatfield, C. & Grose, C. (1995). Cell surface expression and fusion by the varicella-zoster virus gH:gL glycoprotein complex: analysis by laser scanning confocal microscopy. *Virology* **210**, 429-440.

Eck, R. V. & 1966, M. O. D. (1966). Atlas of Protein Sequence and Structure. *National Biomedical Research Foundation, Silver Spring, MD*.

Edwards, A. W. (1994). The fundamental theorem of natural selection. *Biological reviews of the Cambridge Philosophical Society* **69**, 443-474.

Efron, B. (1982). The Jackknife, the Bootstrap, and Other Resampling Plans. *Society for Industrial and Applied Mathematics, Philadelphia*.

Erixon, P., Svennblad, B., Britton, T. & Oxelman, B. (2003). Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. *Systematic biology* **52**, 665-673.

Esparza, J., Benyesh-Melnick, B. & Schaffer, P. A. (1976). Intertypic complementation and recombination between temperature-sensitive mutants of herpes simplex virus types 1 and 2. *Virology* **70**, 372-384.

Farris, J. S. (1977). On the phenetic approach to vertebrate classification. *M K Hecht, P C Goody, and B M Hecht (eds), Major patterns in Vertebrate Evolution Plenum, New York*, 823-850.

Felsenstein, J. (1973). Maximum-likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters. *Syst Zool* **22**, 240-249.

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of molecular evolution* **17**, 368-376.

Felsenstein, J. (1982). Numerical methods for inferring evolutionary trees. *Q Rev Biol* 379-791.

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**, 783-791.

Felsenstein, J. & Yokoyama, S. (1976). The evolutionary advantage of recombination. II. Individual selection for recombination. *Genetics* **83**, 845-859.

Fisher, R. A. (1930). The Genetical Theory of Natural Selection. *Oxford Univ Press, Oxford*.

Fitch, W. M. (1977). On the problem of discovering the most parsimonious tree. *Am Nat* 223-257.

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

Fujita, K., Maeda, K., Yokoyama, N., Miyazawa, T., Kai, C. & Mikami, T. (1998). In vitro recombination of feline herpesvirus type 1. *Archives of virology* **143**, 25-34.

Gojobori, T., Moriyama, E. N. & Kimura, M. (1990). Molecular clock of viral evolution, and the neutral theory. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 10015-10018.

Gray, W. L., Starnes, B., White, M. W. & Mahalingam, R. (2001). The DNA sequence of the simian varicella virus genome. *Virology* **284**, 123-130.

Grunewald, K., Desai, P., Winkler, D. C., Heymann, J. B., Belnap, D. M., Baumeister, W. & Steven, A. C. (2003). Three-dimensional structure of herpes simplex virus from cryo-electron tomography. *Science* **302**, 1396-1398.

Gupta, R., Birch, H., Rapacki, K., Brunak, S. & Hansen, J. E. (1999). O-GLYCBASE version 4.0: a revised database of O-glycosylated proteins. *Nucleic acids research* **27**, 370-372.

Haarr, L. & Skulstad, S. (1994). The herpes simplex virus type 1 particle: structure and molecular functions. Review article. *Apmis* **102**, 321-346.

Halliburton, I. W. (1980). Intertypic recombinants of herpes simplex viruses. *J Gen Virol* **48**, 1-23.

Halliburton, I. W., Randall, R. E., Killington, R. A. & Watson, D. H. (1977). Some properties of recombinants between type 1 and type 2 herpes simplex viruses. *J Gen Virol* **36**, 471-484.

Hanke, T., Graham, F. L., Lulitanond, V. & Johnson, D. C. (1990). Herpes simplex virus IgG Fc receptors induced using recombinant adenovirus vectors expressing glycoproteins E and I. *Virology* **177**, 437-444.

Henderson, L. M., Katz, J. B., Erickson, G. A. & Mayfield, J. E. (1990). In vivo and in vitro genetic recombination between conventional and gene-deleted vaccine strains of pseudorabies virus. *American journal of veterinary research* **51**, 1656-1662.

Hendy, M. D. & Penny, D. (1982). Branch and bound algorithms to determine minimal evolutionary trees. *Math Biosci* **59**, 277-290.

Holder, M. & Lewis, P. O. (2003). Phylogeny estimation: traditional and Bayesian approaches. *Nature reviews* **4**, 275-284.

Honess, R. W. (1984). Herpes simplex and 'the herpes complex': diverse observations and a unifying hypothesis. The eighth Fleming lecture. *J Gen Virol* **65 (Pt 12)**, 2077-2107.

Honess, R. W., Buchan, A., Halliburton, I. W. & Watson, D. H. (1980). Recombination and linkage between structural and regulatory genes of herpes simplex virus type 1: study of the functional organization of the genome. *J Virol* **34**, 716-742.

Huelsenbeck, J. P. & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics (Oxford, England)* **17**, 754-755.

Huelsenbeck, J. P., Ronquist, F., Nielsen, R. & Bollback, J. P. (2001). Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* **294**, 2310-2314.

Hughes, A. L. (2002). Origin and evolution of viral interleukin-10 and other DNA virus genes with vertebrate homologues. *Journal of molecular evolution* **54**, 90-101.

Huson, D. H. (1998). SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics (Oxford, England)* **14**, 68-73.

Huson, D. H. & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular biology and evolution* **23**, 254-267.

Huson, D. H. & Kloepper, T. H. (2005). Computing recombination networks from binary sequences. *Bioinformatics (Oxford, England)* **21 Suppl 2**, ii159-ii165.

Hutchinson, L., Roop-Beauchamp, C. & Johnson, D. C. (1995). Herpes simplex virus glycoprotein K is known to influence fusion of infected cells, yet is not on the cell surface. *Journal of virology* **69**, 4556-4563.

Javier, R. T., Sedarati, F. & Stevens, J. G. (1986). Two avirulent herpes simplex viruses generate lethal recombinants in vivo. *Science* **234**, 746-748.

Johnson, D. C. & Feenstra, V. (1987). Identification of a novel herpes simplex virus type 1-induced glycoprotein which complexes with gE and binds immunoglobulin. *J Virol* **61**, 2208-2216.

Johnson, D. C., Frame, M. C., Ligas, M. W., Cross, A. M. & Stow, N. D. (1988). Herpes simplex virus immunoglobulin G Fc receptor activity depends on a complex of two viral glycoproteins, gE and gI. *J Virol* **62**, 1347-1354.

Kato, N., Ootsuyama, Y., Sekiya, H., Ohkoshi, S., Nakazawa, T., Hijikata, M. & Shimotohno, K. (1994). Genetic drift in hypervariable region 1 of the viral genome in persistent hepatitis C virus infection. *J Virol* **68**, 4776-4784.

Keightley, P. D. & Otto, S. P. (2006). Interference among deleterious mutations favours sex and recombination in finite populations. *Nature* **443**, 89-92.

Kemble, G. W., Annunziato, P., Lungu, O., Winter, R. E., Cha, T. A., Silverstein, S. J. & Spaete, R. R. (2000). Open reading frame S/L of varicella-zoster virus encodes a cytoplasmic protein expressed in infected cells. *J Virol* **74**, 11311-11321.

Kimura, M. (1979). The neutral theory of molecular evolution. *Scientific American* **241**, 98-100, 102, 108 passim.

Kimura, M. (1987). Molecular evolutionary clock and the neutral theory. *Journal of molecular evolution* **26**, 24-33.

Klotz, L. C., Komar, N., Blanken, R. L. & Mitchell, R. M. (1979). Calculation of evolutionary trees from sequence data. *Proc Natl Acad Sci USA* 4516-4520.

Kowalczykowski, S. C., Dixon, D. A., Eggleston, A. K., Lauder, S. D. & Rehrauer, W. M. (1994). Biochemistry of homologous recombination in Escherichia coli. *Microbiological reviews* **58**, 401-465.

Leach, D. R. F. (1996). Genetic Recombination. *Blackwell Science, Oxford.*

Lewis, P. O. (2001). Phylogenetic systematics turns over a new leaf. *Trends in Ecology and Evolution* **16**, 30-37.

Li, W. H. (1981). Simple method for constructing phylogenetic trees from distance matrices. *Proceedings of the National Academy of Sciences of the United States of America* **78**, 1085-1089.

Li, W. H. (1997). Molecular Evolution. *Sinauer Associates, Sunderland, MA*.

Lingen, M., Hengerer, F. & Falke, D. (1997). Mixed vaginal infections of Balb/c mice with low virulent herpes simplex type 1 strains result in restoration of virulence properties: vaginitis/vulvitis and neuroinvasiveness. *Med Microbiol Immunol* **185**, 217-222.

Litwin, V., Jackson, W. & Grose, C. (1992). Receptor properties of two varicella-zoster virus glycoproteins, gpI and gpIV, homologous to herpes simplex virus gE and gI. *J Virol* **66**, 3643-3651.

Lole, K. S., Bollinger, R. C., Paranjape, R. S., Gadkari, D., Kulkarni, S. S., Novak, N. G., Ingersoll, R., Sheppard, H. W. & Ray, S. C. (1999). Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* **73**, 152-160.

Longnecker, R., Chatterjee, S., Whitley, R. J. & Roizman, B. (1987). Identification of a herpes simplex virus 1 glycoprotein gene within a gene cluster dispensable for growth in cell culture. *Proceedings of the National Academy of Sciences of the United States of America* **84**, 4303-4307.

Longnecker, R. & Roizman, B. (1987). Clustering of genes dispensable for growth in culture in the S component of the HSV-1 genome. *Science* **236**, 573-576.

Loparev, V. N., Gonzalez, A., Deleon-Carnes, M., Tipples, G., Fickenscher, H., Torfason, E. G. & Schmid, D. S. (2004). Global identification of three major genotypes of varicella-zoster virus: longitudinal clustering and strategies for genotyping. *J Virol* **78**, 8349-8358.

Mallory, S., Sommer, M. & Arvin, A. M. (1997). Mutational analysis of the role of glycoprotein I in varicella-zoster virus replication and its effects on glycoprotein E conformation and trafficking. *J Virol* **71**, 8279-8288.

Mao, H. & Rosenthal, K. S. (2003). Strain-dependent structural variants of herpes simplex virus type 1 ICP34.5 determine viral plaque size, efficiency of glycoprotein processing, and viral release and neuroinvasive disease potential. *J Virol* **77**, 3409-3417.

Markine-Goriaynoff, N., Georgin, J. P., Goltz, M., Zimmermann, W., Broll, H., Wamwayi, H. M., Pastoret, P. P., Sharp, P. M. & Vanderplasschen, A. (2003).

The core 2 beta-1,6-N-acetylglucosaminyltransferase-mucin encoded by bovine herpesvirus 4 was acquired from an ancestor of the African buffalo. *J Virol* **77**, 1784-1792.

McGeoch, D. J. (1989). The genomes of the human herpesviruses: contents, relationships, and evolution. *Annual review of microbiology* **43**, 235-265.

McGeoch, D. J. (1990). Evolutionary relationships of virion glycoprotein genes in the S regions of alphaherpesvirus genomes. *J Gen Virol* **71 (Pt 10)**, 2361-2367.

McGeoch, D. J. & Cook, S. (1994). Molecular phylogeny of the alphaherpesvirinae subfamily and a proposed evolutionary timescale. *Journal of molecular biology* **238**, 9-22.

McGeoch, D. J., Cook, S., Dolan, A., Jamieson, F. E. & Telford, E. A. (1995). Molecular phylogeny and evolutionary timescale for the family of mammalian herpesviruses. *J Mol Biol* **247**, 443-458.

McGeoch, D. J., Dalrymple, M. A., Davison, A. J., Dolan, A., Frame, M. C., McNab, D., Perry, L. J., Scott, J. E. & Taylor, P. (1988). The complete DNA sequence of the long unique region in the genome of herpes simplex virus type 1. *J Gen Virol* **69 (Pt 7)**, 1531-1574.

McGeoch, D. J., Dolan, A. & Ralph, A. C. (2000). Toward a comprehensive phylogeny for mammalian and avian herpesviruses. *J Virol* **74**, 10401-10406.

McGeoch, D. J., Rixon, F. J. & Davison, A. J. (2006). Topics in herpesvirus genomics and evolution. *Virus research* **117**, 90-104.

Meng, Y. X., Spira, T. J., Bhat, G. J., Birch, C. J., Druce, J. D., Edlin, B. R., Edwards, R., Gunthel, C., Newton, R., Stamey, F. R., Wood, C. & Pellett, P. E. (1999). Individuals from North America, Australasia, and Africa are infected with four different genotypes of human herpesvirus 8. *Virology* **261**, 106-119.

Meurens, F., Keil, G. M., Muylkens, B., Gogev, S., Schynts, F., Negro, S., Wiggers, L. & Thiry, E. (2004a). Interspecific recombination between two ruminant alphaherpesviruses, bovine herpesviruses 1 and 5. *J Virol* **78**, 9828-9836.

Meurens, F., Schynts, F., Keil, G. M., Muylkens, B., Vanderplasschen, A., Gallego, P. & Thiry, E. (2004b). Superinfection prevents recombination of the alphaherpesvirus bovine herpesvirus 1. *J Virol* **78**, 3872-3879.

Mocarski, E. S. & Roizman, B. (1981). Site-specific inversion sequence of the herpes simplex virus genome: domain and structural features. *Proc Natl Acad Sci U S A* **78**, 7047-7051.

Muggeridge, M. I. (2000). Characterization of cell-cell fusion mediated by herpes simplex virus 2 glycoproteins gB, gD, gH and gL in transfected cells. *J Gen Virol* **81**, 2017-2027.

Muir, W. B., Nichols, R. & Breuer, J. (2002). Phylogenetic analysis of varicella-zoster virus: evidence of intercontinental spread of genotypes and recombination. *J Virol* **76**, 1971-1979.

Oxman, M. N., Levin, M. J., Johnson, G. R., Schmader, K. E., Straus, S. E., Gelb, L. D., Arbeit, R. D., Simberkoff, M. S., Gershon, A. A., Davis, L. E., Weinberg, A., Boardman, K. D., Williams, H. M., Zhang, J. H., Peduzzi, P. N., Beisel, C. E., Morrison, V. A., Guatelli, J. C., Brooks, P. A., Kauffman, C. A., Pachucki, C. T., Neuzil, K. M., Betts, R. F., Wright, P. F., Griffin, M. R., Brunell, P., Soto, N. E., Marques, A. R., Keay, S. K., Goodman, R. P., Cotton, D. J., Gnann, J. W., Jr., Loutit, J., Holodniy, M., Keitel, W. A., Crawford, G. E., Yeh, S. S., Lobo, Z., Toney, J. F., Greenberg, R. N., Keller, P. M., Harbecke, R., Hayward, A. R., Irwin, M. R., Kyriakides, T. C., Chan, C. Y., Chan, I. S.,

Wang, W. W., Annunziato, P. W. & Silber, J. L. (2005). A vaccine to prevent herpes zoster and postherpetic neuralgia in older adults. *The New England journal of medicine* **352**, 2271-2284.

Pagamjav, O., Sakata, T., Matsumura, T., Yamaguchi, T. & Fukushi, H. (2005). Natural recombinant between equine herpesviruses 1 and 4 in the ICP4 gene. *Microbiology and immunology* **49**, 167-179.

Perry, L. J. & McGeoch, D. J. (1988). The DNA sequences of the long repeat region and adjoining parts of the long unique region in the genome of herpes simplex virus type 1. *J Gen Virol* **69 (Pt 11)**, 2831-2846.

Preston, V. G., Davison, A. J., Marsden, H. S., Timbury, M. C., Subak-Sharpe, J. H. & Wilkie, N. M. (1978). Recombinants between herpes simplex virus types 1 and 2: analyses of genome structures and expression of immediate early polypeptides. *J Virol* **28**, 499-517.

Roizman, B. (1996a). Herpesviridae. In *Virology*, pp. 2221-2230. Edited by B. N. Fields, D. M. Knipe, P. M. Howley, et al.,. Philadelphia: Lippincott-Raven.

Roizman, B., A. E. Sears. (1996b). Herpes simplex viruses and their replication. In *Virology*, pp. 2231-2295. Edited by B. N. Fields, D. M. Knipe, P. M. Howley, et al.,. Philadelphia: Lippincott-Raven.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution* **4**, 406-425.

Sakaoka, H., Kurita, K., Iida, Y., Takada, S., Umene, K., Kim, Y. T., Ren, C. S. & Nahmias, A. J. (1994). Quantitative analysis of genomic polymorphism of herpes simplex virus type 1 strains from six countries: studies of molecular evolution and molecular epidemiology of the virus. *J Gen Virol* **75 (Pt 3)**, 513-527.

Sample, J., Young, L., Martin, B., Chatman, T., Kieff, E. & Rickinson, A. (1990). Epstein-Barr virus types 1 and 2 differ in their EBNA-3A, EBNA-3B, and EBNA-3C genes. *J Virol* **64**, 4084-4092.

Schynts, F., Meurens, F., Detry, B., Vanderplasschen, A. & Thiry, E. (2003). Rise and survival of bovine herpesvirus 1 recombinants after primary infection and reactivation from latency. *J Virol* **77**, 12535-12542.

Sheldrick, P. & Berthelot, N. (1975). Inverted repetitions in the chromosome of herpes simplex virus. *Cold Spring Harbor symposia on quantitative biology* **39 Pt 2**, 667-678.

Shimizu, H., Yamaguchi, H., Ashizawa, Y., Kohno, Y., Asami, M., Kato, J. & Ikeda, H. (1997). Short-homology-independent illegitimate recombination in Escherichia coli: distinct mechanism from short-homology-dependent illegitimate recombination. *Journal of molecular biology* **266**, 297-305.

Sokal, R. R. & Michener., C. D. (1958). A statistical method for evaluating systematic relationships. *Univ Kansas Sci Bull* **28**, 1409-1438.

Suzuki, D. T., Griffiths, A. J. F., Miller, J. H. & Lewontin, R. C. (1989). *An Introduction to Genetic Analysis 4th ed WH Freeman*, 704.

Suzuki, Y., Glazko, G. V. & Nei, M. (2002). Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 16138-16143.

Swofford, D. L., Olsen, G. J., Waddell, P. J. & Hillis, D. M. (1996). Phylogenetic inference. *D M Hillis, C Moritz, and B K Mable (eds), Molecular Systematics, 2nd Ed Sinauer Associates, Sunderland, MA*, 407-459.

Taylor, T. J. & Knipe, D. M. (2004). Proteomics of herpes simplex virus replication compartments: association of cellular DNA replication, repair, recombination, and chromatin remodeling proteins with ICP8. *J Virol* **78**, 5856-5866.

Timbury, M. C. & Subak-Sharpe, J. H. (1973). Genetic interactions between temperature-sensitive mutants of types 1 and 2 herpes simplex viruses. *J Gen Virol* **18**, 347-357.

True, B. L. & Carter, B. L. (1984). Update on acyclovir: oral therapy for herpesvirus infections. *Clinical pharmacy* **3**, 607-613.

Turner, A., Bruun, B., Minson, T. & Browne, H. (1998). Glycoproteins gB, gD, and gHgL of herpes simplex virus type 1 are necessary and sufficient to mediate membrane fusion in a Cos cell transfection system. *J Virol* **72**, 873-875.

Umene, K. (1985). Intermolecular recombination of the herpes simplex virus type 1 genome analysed using two strains differing in restriction enzyme cleavage sites. *J Gen Virol* **66 (Pt 12)**, 2659-2670.

Umene, K. (1998). Herpesvirus: Genetic Variability and Recombination. *Touka Shobo, Fukuoka*.

Umene, K. (1999). Mechanism and application of genetic recombination in herpesviruses. *Reviews in medical virology* **9**, 171-182.

Wadsworth, S., Jacob, R. J. & Roizman, B. (1975). Anatomy of herpes simplex virus DNA. II. Size, composition, and arrangement of inverted terminal repetitions. *J Virol* **15**, 1487-1497.

Weir, J. P. (1998). Genomic organization and evolution of the human herpesviruses. *Virus genes* **16**, 85-93.

Wildy, P. (1955). Recombination with herpes simplex virus. *J Gen Microbiol* **13**, 346-360.

Yirrell, D. L., Rogers, C. E., Blyth, W. A. & Hill, T. J. (1992). Experimental in vivo generation of intertypic recombinant strains of HSV in the mouse. *Archives of virology* **125**, 227-238.