

Microbes coevolving with human host and ancient human migrations

Angelo Pavesi

Department of Genetics, Anthropology & Evolution, University of Parma, Parco Area delle Scienze 11/A, I-43100 Parma, Italy; e-mail: angelo.pavesi@unipr.it

Summary – Several studies have tried to trace the history of human populations using parasites as genetic markers. The main benefit of using pathogens and other microorganisms as proxies for human variability is that they show more genetic variation with respect to human genes. However, the hypothesis that infectious agents can reconstruct the history of our ancestors does require a careful examination of their biological features, such as the mechanism of transmission, the pathogen power, the rate of variation, and the intensity and form of selection. The purpose of this paper is to review critically the studies on human history that are focused on genetic analyses of a number of infectious agents. Protozoan parasites such as *Plasmodium falciparum*, the aetiological agent of malaria, and metazoan parasites such as lice (*Pediculus humanus*) and tapeworms (*Taenia saginata* and *Taenia asiatica*) were first considered. The ability of *Helicobacter pylori*, a Gram-negative bacterium that colonizes the stomach of most humans, to trace past migrations was then evaluated. Two double-strand DNA viruses (human papillomavirus and human polyomavirus JC) and two positive-strand RNA viruses (human hepatitis G virus and human T-cell lymphotropic virus) were also examined. A broad perspective on pathogens and human evolution, concerning for example the role of parasites in determining and maintaining the high genetic diversity found in human populations, is finally presented.

Keywords – Bacteria, Parasites, Viruses, Coevolution, Human evolution, Human migration.

Parasites can be a powerful tool for reconstructing host evolutionary history because they provide data that are independent of host data. The close link between the taxonomy of parasites and their host (Hafner & Nadler 1988) has supported the hypothesis that the corresponding phylogenetic trees should be topologically identical. In accordance with these observations, several studies have tried to trace the history of human populations using parasites as genetic markers. The main benefit of using microbes as proxies for human variability is that they show more genetic variation with respect to human genes. This means that the microorganisms, with their rapid generation time, can accurately reconstruct past events, like those involved with ancient or more recent human migrations.

However, the hypothesis that infectious agents can reflect human history does require a careful examination of their biological features, such as the mechanism of transmission, the pathogen

power, the rate of variation, and the intensity and form of selection. The geographical origin and the time of emergence of a human parasite is a further significant feature. In fact, reliable inferences regarding ancient human migrations should be provided by parasites that infected humans in Africa since prehistoric times.

The purpose of the present paper is to review critically the studies on human history that are focused on genetic analyses of a number of parasites (protozoan and metazoan parasites, bacteria and viruses).

An overview of human pathogens

Protozoan parasites

Plasmodium falciparum is a major human pathogen associated with 200-300 million clinical cases of malaria and 1-3 million deaths each year worldwide. It is a hermaphroditic protozoan, with haploid asexual replication in the human host and

a brief diploid sexual phase in the mosquito vector. Haploid parasites divide mitotically in the human host, and some cells differentiate into male and female stages. Male and female gametes fuse in the mosquito host (*Anopheles gambiae*) to form a short-lived diploid zygote. Meiotic division then gives rise to haploid cells that develop into infective sporozoites, which migrate to the mosquito salivary glands and infect humans during mosquito blood-feeding.

The closest relative of *P. falciparum* is the chimpanzee parasite *P. reichenowi*. Sequence analysis at two gene loci (the small subunit rRNA and the circumsporozoite protein genes) has demonstrated that these two species diverged from each other around 8-12 million years ago, which is roughly consistent with the time of divergence between the two host species, human and chimpanzee (Escalante & Ayala 1994; Escalante *et al.*, 1995). This finding suggests that *P. falciparum* is an ancient parasite, associated with our ancestors since the divergence of the hominids from the great apes.

A study of the microsatellite haplotypes has documented a strong geographical population structure of *P. falciparum*, with the highest level of variation occurring in Africa (Anderson *et al.*, 2000). The hypothesis of an African origin of the parasite is also supported by a genetic analysis of 100 worldwide mitochondrial DNA sequences (Joy *et al.*, 2003). Analysis of non-coding and synonymous sites provides an estimate of the time to the most recent common ancestor (TMRCA) of all extant parasite population comprised between 70,000 and 90,000 years ago. A sudden increase in the African population of *P. falciparum*, dating 10,000 years ago, has also been inferred. This latter date is concurrent with the emergence of slash-and-burn agriculture in the African rainforest, which could have provided suitable expansion conditions for the mosquito vectors of *P. falciparum* and adequate human population size to maintain transmission (Joy *et al.*, 2003).

A different estimate of the TMRCA of *P. falciparum* is provided by other studies (Rich *et al.*, 1998; Volkman *et al.*, 2001). The virtual absence of synonymous substitutions in some protein-coding genes, combined with the lack of Small Nuclear Polymorphisms (SNPs) in housekeeping-gene introns, is indicative of a recent expansion from a

single progenitor comprised between 6,000 and 8,000 years ago. The most probable reason for the observed monomorphism is a recent bottleneck in the *P. falciparum* population, followed by a rapid expansion from its African tropical origins to the tropical and subtropical regions of the world (Rich *et al.*, 1998).

It is obvious that further research, mainly focused on a genome-wide examination of SNPs, could explain these conflicting results. However, both views - the recent common ancestor hypothesis and the ancient common ancestor hypothesis (Fig. 1) - support the notion that endemic malaria arose some 5,000 - 10,000 years ago in Africa, favored by the environmental changes brought by agriculture. Consistent with this view, is the time of origin (3,000 - 10,000 years) of two malaria-resistant alleles of the gene for glucose-6-phosphate dehydrogenase in human red blood cells (Tishkoff *et al.*, 2001). Likewise, the malaria-protective human alleles that cause sickle cell disease and alpha-thalassaemia also appear to have origins within this timeframe (Curat *et al.*, 2002).

The link between human populations and *P. falciparum* can be summarized as follows. The parasite was imported into the Mediterranean region during the first migration of modern humans out of Africa. However, it is difficult to think how *P. falciparum*, with its high pathogenicity and short-term survival in the host, could have survived in small, dispersed groups of hunters and gatherers. The parasite probably attained its full means of expansion with the Neolithic revolution, 8,000 to 10,000 BC. Sedentary life, the formation of the first villages and small towns, clearing and irrigating crops, and the increase in human population certainly favored the spread of malaria (Hume *et al.*, 2003a). Initially, malaria transmission in Neolithic Africa was probably carried out by more zoophilic *Anopheles* species. The evolution of more anthropophilic taxa occurred later and, in the case of *Anopheles gambiae*, was a direct consequence of the impact of humans on the surrounding environment (Coluzzi 1999; Hume *et al.*, 2003b).

These observations support the hypothesis that *P. falciparum* could be an adequate marker for tracing human migrations dating to Neolithic times.

Fortunately, a large number of SNPs has been

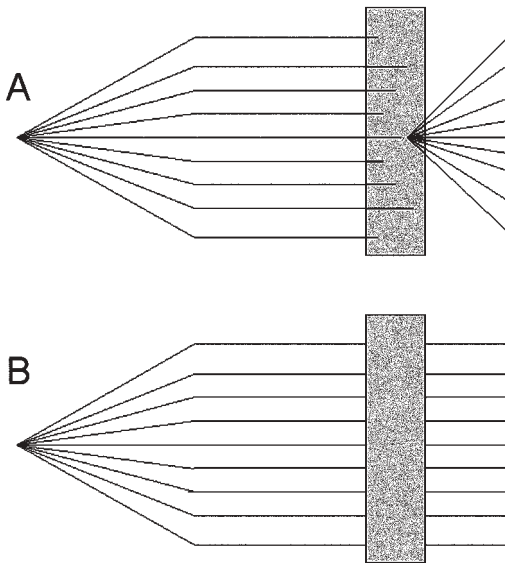


Fig. 1 - Different estimates of the time to the most common ancestor (TMRCA) of all extant populations of *P. falciparum*. (a) The recent common ancestor model suggests a recent bottleneck in the parasite population, characterized by the persistence of only a few (and perhaps only one) of the ancient lineages. Under this hypothesis, a TMRCA comprised between 6,000 and 8,000 years was inferred (b) The ancient common ancestor model suggests the persistence of many ancient lineages. Under this hypothesis, a TMRCA comprised between 70,000 and 90,000 years was inferred (redrawn from Hartl *et al.*, 2002, with permission from Elsevier).

identified by Mu and co-workers (2002), thus demonstrating that the extant parasite population shows an adequate level of genetic diversity. The sequencing of the entire genome of *P. falciparum* (Berry *et al.*, 2004; <http://PlasmoDB.org>) should allow for the detection of the genomic fragments most suitable for inferring the evolutionary history of the parasite. Sequence analysis of these regions could clarify the pattern of past human migrations that occurred in geographical areas where endemic malaria still persists.

Metazoan parasites

Lice are obligate ectoparasites of mammals that

complete their entire life cycle on the body of the host. They cannot survive more than a few hours or days off the host. Humans are parasitized by two species of lice: head/body lice (*Pediculus humanus*) and pubic lice (*Phthirus pubis*). The two forms of *P. humanus* are morphologically similar, but ecologically distinct. Body lice live primarily in clothing and move onto the skin to feed once or twice a day. Head lice are confined to the scalp and feed more frequently.

Kittler *et al.* (2003) used a molecular-clock approach to date the origin of body lice, under the assumption that this should roughly correspond with the use of clothing by humans. They obtained both nuclear and mitochondrial DNA sequence data from a worldwide sample of 40 head and body lice, and from a chimpanzee louse to use as outgroup. Data analysis yielded a phylogenetic tree whose basal clade includes nine African head lice, supporting the hypothesis of an African origin of the parasite.

The finding that the other clade of the tree contains both body and head lice suggests that body lice originated from head lice. The root of this clade was estimated to be 72,000 +/- 42,000 years old, suggesting a link with the global expansion of modern humans out of Africa in the last 100,000 years. This result implies that clothing was a recent innovation, probably associated with the spread of early modern humans into cooler regions.

Remarkable findings on human history have been provided by a further genetic study of head/body lice (Reed *et al.*, 2004). On the basis of a phylogenetic analysis of mitochondrial DNA sequences from five louse species (a 1,525 bp combined fragment of the cytochrome oxidase subunit I and the cytochrome b genes), the authors estimated a divergence time of 5.6 millions of years for the split between *P. schaeffi* from chimpanzee and *P. humanus* from human. The similarity between this divergence time and that estimated for the divergence between the two host species (human and chimpanzee) suggests that there is a history of co-speciation in the primate-lice interaction.

Phylogeny within *P. humanus* itself yielded a tree with two distinct lineages, whose time of divergence was predicted to be of 1.18 millions of years (My). This estimate predates by a considerable margin the origin of modern *H. sapiens* (0.15-0.20

My), as inferred from mitochondrial DNA (Ingman *et al.*, 2000) or fossil evidence (White *et al.*, 2003). One lineage is distributed worldwide and contains both head and body louse forms (WW clade). The other lineage is restricted to the New World and contains only the head louse form (NW clade).

With the aim to explain why the TMRCA for *P. humanus* is an order of magnitude older than that of its human host, Reed *et al.* raise the hypothesis that the two lineages of *P. humanus* must have diverged at about the same time that two lineages of humans - perhaps Asian *H. erectus* and the African ancestors of *H. sapiens* - became established (Fig. 2). The strong genetic difference between the two parasite clades suggests that they had little or no contact with each other. Thus, the lice probably living on *H. erectus* (the NW clade) must have jumped to *H. sapiens* at some point before *H. erectus* became extinct, perhaps as late as 30,000 years ago. The shift probably occurred through skin-to-skin contact, as might occur for example during fighting.

The WW clade of *P. humanus* seems to have a common evolutionary history with modern *H. sapiens*, because it shows features indicative of a population expansion dating 0.11 My. This date is similar to the time of expansion (0.10 My) of modern humans out of Africa (Harpending *et al.*, 1993; Cavalli-Sforza & Feldmann, 2003).

Finally, Hoberg *et al.* (2001) report that two tapeworm species parasitizing humans (*Taenia saginata* and *Taenia asiatica*) show a divergence time and a geographical pattern amazingly concordant with those proposed for the two subgroups of *P. humanus*. As stated by Pennisi (2004), genetic data from complex parasites such as lice and tapeworms should get insights into the possibility of close contacts between distinct species of early humans.

Bacteria

Helicobacter pylori is a Gram-negative bacterium that was discovered about 20 years ago by Marshall & Warren (1984). Its complete genome sequence is available at the World-Wide Web site <http://genolist.pasteur.fr/PyloriGene>. *H. pylori* colonizes the stomach of about half of the human population and is associated with chronic gastritis and peptic/duodenal ulcer. However, clinical

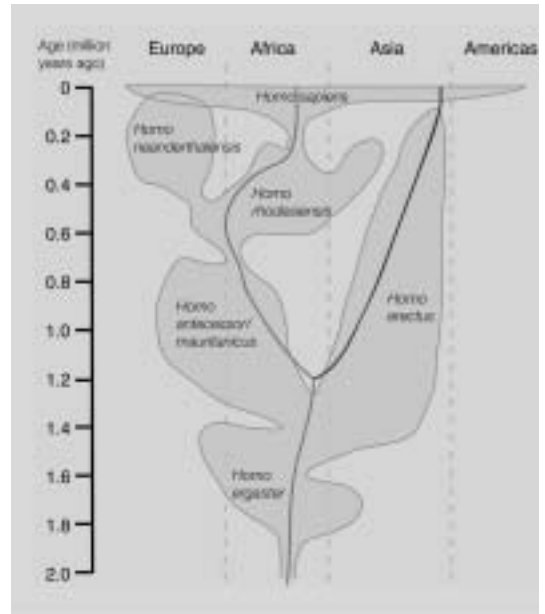


Fig. 2 - Superposition of the temporal distribution of the two divergent lineages of *P. humanus* on the hominid tree. Such a tree depicts one view of human evolutionary history based on fossil data (Stringer, 2003). The worldwide (WW) clade is shown in red and the New World (NW) clade in blue (from Reed *et al.*, 2004)

disease only occurs in a minority of infected individuals, and in most humans *H. pylori* has many of the features of a commensal that colonizes the gastric mucosa. Children acquire the bacterium as infants,

and epidemiological studies have shown that transmission occurs predominantly within families, and probably mostly from mother to child (Kivi *et al.*, 2003).

The quasi-vertical mode of transmission should lead, over long periods of time, to different human populations being colonized by distinct sets of bacterial genotypes. However, the main problem is that *H. pylori* is an unusually sexual bacterium: the genotype of a resident *H. pylori* isolate accumulates many very small recombinational replacements from genetically distinct coinfecting isolates, which rapidly diversify its genome (Falush *et al.*, 2001). Horizontal genetic exchange occurs via DNA transformation between

multiple strains that colonize a single individual, resulting in a heterogeneous pool of isolates that are closely related but differ in alleles that have been acquired by recombination (Suerbaum *et al.*, 1998).

Despite the evidence for frequent recombination, phylogeographical differences exist within *H. pylori* that seem to reflect co-evolution with the host (Covacci *et al.*, 1999). Moreover, Ghose *et al.* (2002) have demonstrated that the *H. pylori* genotypes found in aboriginal South-American populations are largely of East Asian origin, whereas isolates from South American cities are of European origin. These findings are consistent with an ancient colonization of modern humans by *H. pylori*, so ancient that bacteria were first introduced into the Americas 13,000-20,000 years ago from North-East Asia via the Bering land bridge. The high frequency of European isolates in South-America is probably due to the post-Columbus colonization of the New World by peoples of Caucasian origin. A similar study, based on 41 strains from native New World populations inhabiting Alaska and Colombia, has confirmed that *H. pylori* was present in the Americas before Columbus (Yamaoka *et al.*, 2002).

Falush and colleagues (2003a) have provided further evidence that *H. pylori* can be a valuable tool for tracing ancient human migrations. The authors obtained sequences of eight genes (seven house-keeping genes and one virulence-associated gene) from 370 *H. pylori* strains recovered from 27 different human populations, grouped according to geographical location or ethnicity. Of the 3850 nucleotides determined for each isolate, 1418 were polymorphic and were used to define bacterial populations. Thanks to the use of the computer-program STRUCTURE (Falush *et al.*, 2003b; <http://pritch.bsd.uchicago.edu>), the authors obtained five different *H. pylori* populations, which correspond to the human populations from which the isolates were recovered.

On the basis of the genetic diversity of *H. pylori*, several migratory routes of humans were inferred. The pattern of human migrations concerns ancient events such as the first peopling of Americas and the Bantu expansion in South Africa, as well as more recent events such as the colonial expansion from Europe to the Americas, South-Africa, and Australia (Fig. 3).

In some *H. pylori* isolates, all the polymorphisms found in the eight gene segments were assigned with confidence to a single human population. This is the case, for example of an isolate from the Maori population, all of whose eight gene segments are assigned to East-Asia with high probability. However, there are examples of population mosaics, where a small number of polymorphisms in a gene segment, or a whole gene segment, appears to have been introduced by a recombinational replacement from a coinfecting *H. pylori* derived from a different population (Falush *et al.*, 2003; Spratt 2003).

Recently, Wirth *et al.* (2004) have demonstrated that sequence information from *H. pylori* shows a high resolution and discriminatory power for clarifying the history of a human population living in the Ladakh province of northern India. The Ladakh is inhabited by two ethnic groups (Buddhists and Muslims), who have coexisted for about 1,000 years but remain isolated due to religious and cultural differences.

The *H. pylori* isolates from Buddhists were mosaics of nucleotides from ancestral East-Asia and ancestral Europe, consistent with the introduction of Buddhism by Tibetans in an ancient Ladakhi population. By contrast, almost all isolates from Muslims were of European origin, indicating that the Islamic religion was introduced by few individuals rather than extensive migrations. Interestingly, these conclusions are consistent with the known history of the Ladakh province.

The main limitations in the use of *H. pylori* as a marker of human history lie in the fact that the sampling requires microbiological cultivation from biopsies taken during gastroendoscopy, an invasive procedure. Much of the current sampling is biased towards isolates from humans with clinical symptoms. *H. pylori* is not universally present and its prevalence is decreasing in industrialized countries. Inferences on human history are mainly restricted to human migrations in the last few thousand years (Fig. 3). Finally, the African origin of *H. pylori* is a mere assumption, since no phylogenetic tree with a root including only African isolates has been obtained so far.

Viruses

Although the origin, progress and impact of viral diseases has long been of interest to

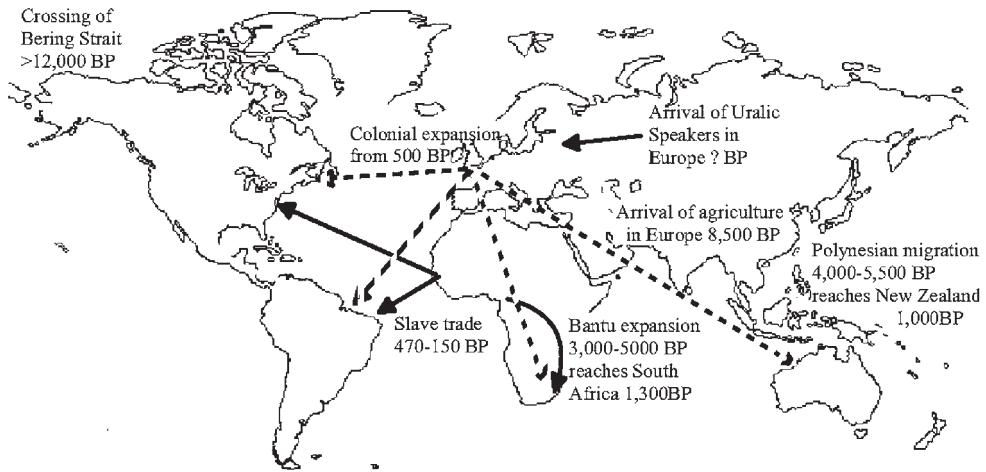


Fig. 3 - Putative modern and ancient migrations of *H. pylori*. Arrows indicate specific migrations of humans and *H. pylori* populations (from Falush *et al.*, 2003, with permission from the American Association for the Advancement of Science).

epidemiologists, the advent of rapid and large-scale gene sequencing has allowed remarkable new insights into the patterns of their spread through human populations (Holmes 2004). The geographical distribution of viruses can reveal patterns of human migration, especially in the cases in which they move in narrow, ethnically restricted streams of infection. The examples reported here include DNA viruses, such as the JC polyomavirus and the human papillomavirus, and RNA viruses, such as the human hepatitis G virus and the human T-cell lymphotropic virus.

Polyomavirus JC

The human JC polyomavirus (JCV) is a small, slowly evolving DNA virus belonging to the *Polyomaviridae* family (Frisque *et al.*, 1984). The virus is ubiquitous in populations worldwide and is excreted harmlessly in urine by a large percentage of individuals. After primary infection, usually occurring during childhood, the virus persists in the kidneys without obvious symptoms. Reactivation of latent JCV is frequent in adults, as proved by the detection of viral progeny in the urine of a high percentage (40-60%) of healthy individuals (Chang *et al.*, 2002).

Only in a small percentage (5-10%) of immunocompromised patients, JCV can cause a fatal demyelinating disease, called progressive

multifocal leukoencephalopathy. Multiple infections rarely occur, as proved by the detection of the same viral strain from urine samples collected from the same patients at different times (Kitamura *et al.*, 1997). It follows that genetic recombination between different JCV strains seems to be very unlikely. Most of the virus transmission among humans occurs by paternal or maternal route (Kunitake *et al.*, 1995). Since the virus seems to be spread largely within the family or in the immediate community early in the life, the infection is transmitted primarily within the particular ethnic group.

Two groups of investigators (Agostini *et al.*, 1997; Sugimoto *et al.*, 1997) first found JCV to be useful for tracing human history. Human migrations, with additional local bottlenecks and genetic drift, have led to the occurrence of at least seven distinct viral genotypes that correlate with human populations in different geographic areas. The major genotypes so far identified include Types 1 and 4, which are of European origin (Agostini *et al.*, 2001), and Types 3 and 6, which are of African origin (Sugimoto *et al.*, 1997). Types 2 and 7 are found in Asia (Sugimoto *et al.*, 1997; Agostini *et al.*, 1998), and Type 8 has been isolated in Papua New Guinea and the Pacific Islands (Yanagihara *et al.*, 2002).

The presence of distinctive types of JCV in the

main ethnic groups suggests an ancient interaction with the human host. This view was first proposed by Agostini and co-workers (1997), who detected Asian genotypes of JCV in samples from Native Americans, a finding consistent with the Asian origin of these populations. On the basis of the small diversity between Amerindian and North-East Asian strains, the authors propose a close co-evolution of JCV with the major lineages of early humans, who migrated out of Africa around 100,000 years ago to people Asia and Europe.

This latter view, however, has been doubted by Wooding (2001). The main criticism lies in the fact that the phylogenetic reconstruction based on the complete genome sequence of JCVs from Old World populations yields a tree which places on its lowest branch the European lineage (Sugimoto *et al.*, 1997; Agostini *et al.*, 2001; Yanagihara *et al.*, 2002). This result is inconsistent with trees constructed from human genes, whose first split separates African and non-African populations (Cavalli-Sforza *et al.*, 1988).

The reason for this incongruity has been investigated by a whole-genome phylogenetic analysis focused on the distinction between slow- and fast-evolving sites (Pavesi 2003). By this approach, it has been proposed that the association of JCV with humans originated in Africa, since the African Type 6 was predicted to be the ancestral genotype. It was also demonstrated how Type 6 gave rise to two independent evolutionary lineages: one including Types 1 and 4, the other Types 2, 3, 7, and 8.

The diffusion in the world of both lineages has been investigated through the analysis of over one thousand sequences of the genomic region of JCV with the highest variation rate (Pavesi 2004). By using synthetic maps, the expansion of *Homo sapiens* from Africa was hypothesized to be mediated by two migration waves, each carrying a different virus lineage. This finding is a valuable one, because it sheds new light on the pattern of human evolution yielded so far by human genes, which supports the hypothesis of one single expansion from Africa into Asia and from there to the other continents (Cavalli-Sforza & Feldman, 2003).

The view that the dual exit of JCV from Africa mirrors two migrations on the part of our ancestors is appealing. However, the objection can be raised

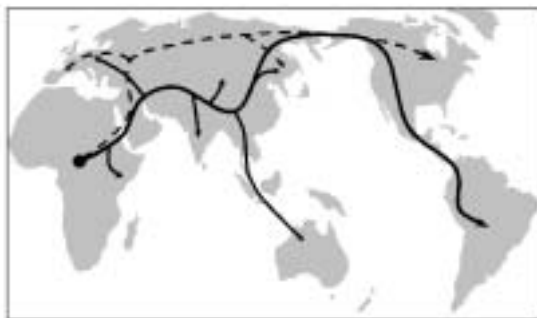


Fig. 4 - The two-migration model of the expansion from Africa of *Homo sapiens* suggested by JCV. The migration traced with a solid line is compatible with that yielded by human genes. The migration traced with a dashed line indicates the additional route of expansion suggested by JCV but undetectable with human genes (from Pavesi, 2005, with permission from the Society for General Microbiology).

that the present genetic diversity between the two virus lineages - one (Types 1 and 4) mainly diffused in the northern areas of the world and the other (Types 2, 3, 7, and 8) in the central and southern areas - is the result of selective pressures favoring adaptation to different climates. In this case, large-scale inferences regarding human evolution should be treated with caution, since a reliable reconstruction of human history is based on phenomena such as genetic drift or migration, and not natural selection (Cavalli-Sforza *et al.*, 1994).

A possible response to this objection has been provided by a multivariate statistical analysis of about three hundred complete genome sequences of JCV, with the aim of characterizing the type of nucleotide substitutions causing the deep divergence between the two virus lineages (Pavesi 2005). Correspondence analysis yielded evidence that the variation pattern of JCV meets with the expectations of the neutral theory of evolution, which is entirely described by random genetic drift, mutation and migration (Kimura 1983). Thus, it can be hypothesized that the diffusion in the world of the two virus lineages really mirrors two early expansions of humans from Africa (Fig. 4).

Since JCV exhibits the unusual feature of a dual exit from Africa (Pavesi 2003), it could shed new light on the number of migrations leading to the

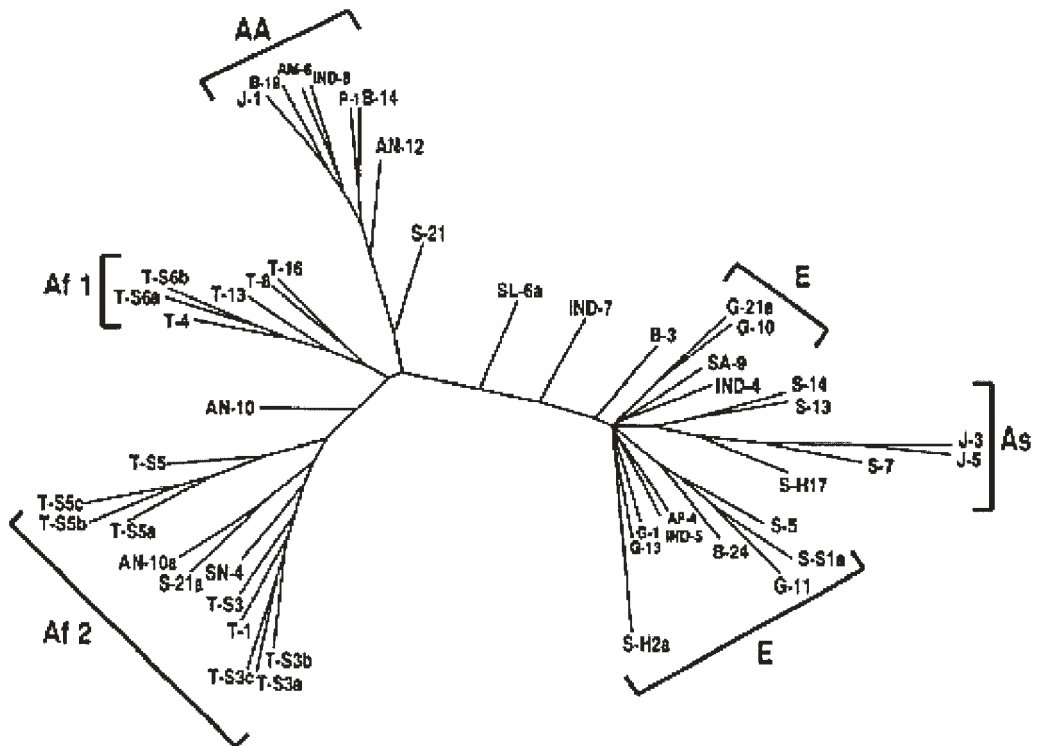


Fig. 5 - Consensus phylogenetic tree highlighting the relationship between 48 genomic variants of HPV-16. Af1 and Af2 indicate the two African branches, AA the Asian/American branch, E the European branch, and As the East Asian branch (from Ho *et al.*, 1993, with permission from the American Society for Microbiology).

peopling of the various continents. The virtual lack of pathogenous power, the absence of genetic recombination, the strong ethnicity due to a transmission mechanism within the family or in the same community, and the easy detection in healthy individuals due to the high frequency of urinary excretion support the utility of JCV as a marker of human history.

Papillomavirus

The human papillomavirus (HPV) is a member of the Papillomavirus genus and consists of a circular, double-stranded DNA with a genomic size of about 8 kb. HPV is spread both horizontally through sexual contact and vertically from mother-to-child, and infections are rarely persistent (Hildesheim *et al.*, 1994).

HPV has been isolated in every human

population sampled to date, with a global prevalence of about 6-7% (Munoz *et al.*, 2000). It has been classified into more than 90 types, some of which are more often associated with cervical cancers than others (Chan *et al.*, 1995). The mutation rate in HPV has been found to be roughly equivalent to that of host cellular DNA, and recombination is rare (Yamada *et al.*, 1997). The strong stability of the HPV genome has led a number of authors to investigate the possibility of co-divergence between virus and humans. Some studies have focused on two common types, type 16 (HPV-16) and type 18 (HPV-18), which seem to reflect better past human migrations.

Investigation of genetic variation in HPV-16, carried out by sequence analysis of the control region, has first discerned three geographically

associated lineages, that is lineages specific to Africa, Brazil and Eurasia (Chan *et al.*, 1992). A further analysis of the same region of 301 HPV-16 isolates from 25 different human populations (Ho *et al.*, 1993) has provided a tree consisting of five major lineages: two predominantly present in Africa (Af1 and Af2), two in Asia (AA and As) and one mainly found in Indo-Europe (E). Unfortunately, the lack of an appropriate outgroup precluded the determination of the most ancestral lineage, since the tree remains unrooted.

However, examination of the HPV-16 tree (Fig. 5) reveals some discrepancies with respect to what is known from human genetic data. For example, the As lineage (Japan and Singapore) is closely related to the European E cluster, rather than the other Asian cluster. An unexpected link between the European lineage E and many isolates from Amerindian populations is detectable. Since these latter ethnic groups originated from North-Asiatic emigrants via the Bering land bridge, their HPV-16 strains should all be associated with the AA lineage. Even more surprising is the detection of the Af1 lineage from one Inuit (Native American population living in North-Canada) and the Af2 lineage from Navajos.

A study of the genetic diversity of HPV-18, based on a sequence analysis of the long control region, has demonstrated that the corresponding phylogenetic tree has a topology that suggests co-evolution of the virus within the major ethnic groups: Africans, Caucasians, and East Asians (Ong *et al.*, 1993). Using HPV-45 as outgroup, the HPV-18 isolates from Africa were placed at the root of the phylogenetic tree, thus suggesting an African origin. By using the close genetic relationship observed between isolates from Amazonian Indians and Japanese populations, it was also possible to evaluate the speed of the molecular clock during HPV-18 evolution (one point mutation fixed every 12,000 years). This estimate suggests an age of approximately 200,000 years for the intra-type diversity of HPV-18, an order of magnitude similar to that of the evolution of *H. sapiens* (Ong *et al.*, 1993).

Although HPV-18 has been found to be associated with cervical neoplasia, this is a rare outcome of HPV infection that generally occurs after child-bearing age. Through most of human evolution, few women lived to the age where they

would be susceptible to HPV-induced cancers. So it is unlikely that the pathogenic potential of HPV has influenced its co-evolution with human host. The main drawbacks in the use of HPV-18 as a marker of human history concern the mechanism of transmission (frequently horizontal) and the low prevalence (1-2%) in human populations (Taira *et al.*, 2004).

Hepatitis G virus

In 1996 two groups of investigators reported the identification of two novel infectious agents of humans called hepatitis G virus (HGV) and hepatitis GB virus C (GBV-C) in American and West African patients, respectively (Linnen *et al.*, 1996; Leary *et al.*, 1996). The molecular characterization of these two agents showed them to be different isolates of the same virus which was designated GBV-C/HGV. GBV-C/HGV is a RNA virus belonging to the Flaviviridae family, with a genomic size of about 9.3 kb.

The virus shows an extremely low infection rate, ranging from 1-4% in industrialized countries to 10% in Brazil (Cheung *et al.*, 1997) and 13% in Bolivia (Konomi *et al.*, 1999). GBV-C/HGV is transmitted parenterally, and highly prevalent in patients exposed to transfused blood (Desai *et al.*, 2004). The virus can also be transmitted horizontally by sexual contact, or vertically from mother to baby (Cheng *et al.*, 2003). It has been detected in the saliva and serum of infected patients, but not in the urine (Seemayer *et al.*, 1998).

The low mutation rate found in the virus (Suzuki *et al.*, 1999) reflects a strong selection against nonsynonymous changes. The finding that selective pressures occur also at third codon position has been explained as a result of functional constraints due to extensive RNA secondary structures (Simmonds & Smith, 1999) and the presence of a new potential overlapping gene (Pavesi, 2000).

Since its first description, many groups have investigated the genetic diversity of GBV-C/HGV isolates from globally distributed human populations (Smith *et al.*, 2000 and references therein). Initially, three GBV-C/HGV genotypes were described originating from Africa (genotype 1), North-America and Europe (genotype 2), and Japan (genotype 3). Two novel isolates, discovered

in South-East Asia and South-Africa, were later classified as genotypes 4 and 5, respectively (Naito *et al.*, 1999; Sathar *et al.*, 1999).

The detection of a large insertion (12 amino acids) in the NS5A gene of GBV-C/HGV strains isolated from Pygmy and Bantu populations of Central Africa proved to be particularly interesting. This feature, combined with a phylogenetic analysis based on the NS5A gene, suggested an African origin of hepatitis G virus (Tanaka *et al.*, 1998). This finding, however, was not confirmed when the analysis was extended to the entire coding region of 27 genomic sequences (Suzuki *et al.*, 1999). The same study also raised the possibility that the above insertion is the result of a later duplication event, rather than an ancestral feature. The difficulty in detecting the ancestral lineage was apparent in the study by Smith *et al.* (2000), based on a phylogenetic analysis of 33 complete genome sequences.

Some light on the origin of GBV-C/HGV has been provided by a phylogenetic analysis of the

genomic regions showing the lowest rate of synonymous substitutions (Pavesi, 2001). The corresponding phylogenetic tree (Fig. 6) supported the hypothesis of an African origin for the virus. By using a multivariate statistical method, and extending the analysis to the complete coding region, further details of the evolutionary history of GBV-C/HGV were clarified. By this approach, isolates from Southeast Asia appeared to be the most closely related to those of African origin, suggesting a major route of human migrations from Africa to south-eastern areas of the Asian continent during prehistoric times (Pavesi, 2001).

As a marker of human history, GBV-C/HGV presents several advantages such as the probable African origin, the lack of pathogenic power and a mechanism of transmission - from mother to infant by breast feeding (Takayama *et al.*, 1999) - that mimics the Mendelian inheritance of mitochondrial DNA. Disadvantages mainly concern a low rate of infection in human populations, making virus isolation rather

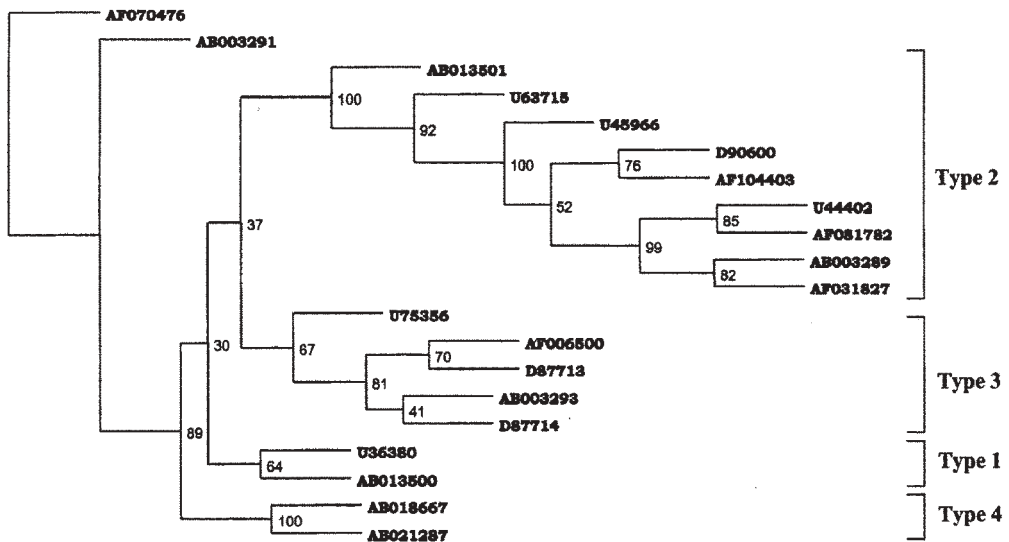


Fig. 6 - Consensus phylogenetic tree of GBV-C/HGV obtained from an analysis of the four most slowly evolving genome regions. The sequence AF70476, which corresponds to GBV-C/HGV from chimpanzee, was used as outgroup. The position of the sequence AB003291, which correspond to a GBV-C/HGV isolate taken from an African patient, supports the hypothesis of an African origin of hepatitis G virus. The internal node numbers represent the bootstrap values (expressed as a percentage of all trees) obtained from 1000 replicates (from Pavesi, 2001, with permission from Springer-Verlag).

difficult, and a transmission mechanism also by sexual contact (Scallan *et al.*, 1998) that can obscure the link between virus distribution and human migrations. The finding that GBV-C/HGV can recombine (Worobey & Holmes, 2001) raises further doubts on its efficacy as a marker of human history.

T-cell lymphotropic virus

Human T-cell lymphotropic virus type 1 (HTLV-I) is the aetiological agent of the adult T-cell leukaemia (Yoshida *et al.*, 1982) and tropical spastic paraparesis, a neurological syndrome (Osame *et al.*, 1987). It infects an estimated 11 to 20 million persons worldwide, with an approximate 5% lifetime risk of serious disease. HTLV-I is transmitted by vertical infection from mother to child, but also by horizontal infection via sexual contact, blood transfusion or intravenous drug abuse.

The HTLV-I is a single-stranded RNA virus, with a genomic size of about 9 kb. The genome is transcribed into DNA when the virus enters a cell and it integrates into the host genome. The strong genetic stability of HTLV-I can be partially explained by the finding that provirus replication via clonal expansion of the infected cells is preferred to virus replication by reverse transcription (Wattell *et al.*, 1995).

A first phylogenetic study (Ureta Vidal *et al.*, 1994) led to the detection of five major geographical genotypes: Cosmopolitan (C) genotype widespread all over the world, Japanese (J) genotype, West African (WA) genotype, Central African (CA) genotype and Melanesian (M) genotype. The same study yielded a phylogenetic tree whose most basal clade is not found in Africa but instead among isolates from Melanesia.

On the basis of the molecular characterization of the long terminal repeat region, Miura *et al.* (1994) discerned three major lineages (subtypes A, B and C) within the Cosmopolitan genotype. Interestingly, subtype A suggests a close connection of the Caribbean and South American natives with the Japanese and thereby a possible migration of the virus lineage to the American continent via Beringia in the Paleolithic era. Subtype C consists of the West African and other Caribbean isolates, indicating that part of the Caribbean strains directly originated from West Africa probably

during the period of slave trade. However, the construction of the corresponding phylogenetic tree yielded again a basal clade consisting of isolates from New Guinea and Melanesia.

Even more surprising is the finding that, in the same tree (see Fig. 1 in Miura *et al.*, 1994), simian T-lymphotropic type I viruses (STLV-I) are more closely related to other human HTLV-I genotypes than they are to the Melanesian HTLV-I genotype. This finding suggests the existence of multiple viral reservoirs in higher primates from which transmission to humans occurred on independent occasions. This hypothesis was further substantiated by the identification of an additional subtype (D) in Africa, which shows a close affiliation with isolates from mandrills and baboons (Mahieux *et al.*, 1998). Thus, the phylogenetic analysis of HTLV-I supports the hypothesis of multiple, independent transmissions from different ape species to local populations of modern humans. This type of transmission raises strong doubts for the use of sequence diversity within HTLV-I to trace the history of human populations.

The human T-cell lymphotropic virus type II (HTLV-II) shows a mechanism of transmission and a genome organization similar to HTLV-I. Sequence analysis at the genome level reveals a difference between HTLV-I and HTLV-II of 37%. The virus has been found in some American Indian (Biggar *et al.*, 1996) and Pygmy tribes (Gessain *et al.*, 1995) where it is considered endemic, and in intravenous drug users world-wide (Lee *et al.*, 1989).

A phylogenetic analysis of the entire long terminal repeat (LTR) sequence supports an African origin of HTLV-II, based on the very deep branching of an HTLV-II strain isolated from an Efe Bambuti Pygmy (Vandamme *et al.*, 1998). A similar phylogenetic study (Neel *et al.*, 1994) demonstrates that the HTLV-II strain present in Amerindian tribes is not present in ethnic groups of eastern Siberia, but rather in the indigeneous population of Mongolia. This finding suggests the ancestors of the first migrants to the New World were not derived from North and Central Siberia, but from populations inhabiting the regions of Mongolia and Manchuria.

The main problem lies in the fact that culturally and geographically isolated ethnic groups inhabiting different continents share a similar form

of the virus. The best example is the identical LTR sequence between a Pygmy population from Bakola and a Wayuu aboriginal population from Colombia (Switzer *et al.*, 1995), a finding hard to explain under the traditional model of human expansion.

These conflicting results can be due to the high rate of substitution of LTR, which was estimated to be only two orders of magnitude lower than that of HIV, the most rapidly evolving virus discovered so far (Salemi *et al.*, 1998). It is likely that the LTR sequence may be saturated by polymorphisms and contain multiple homoplasies, which would reduce the information content (Wirth *et al.*, 2005). On the other hand, HTLV phylogeny from sequence analysis of the envelope gene could be questioned, because envelope proteins are subjected to strong selective pressure due to the host immune system response. These observations, combined with a mechanism of transmission often horizontal, raise doubts on the overall utility of HTLVs for reconstructing human movement and history.

A critical evaluation of evidence

The studies reported here suggest that the genetic information carried by microbes can really get insight into the history of our ancestors. A

parallel example of the utility of these indirect markers is provided by mitochondrial DNA (mtDNA). Although mitochondria descended from bacteria that were incorporated into the eukaryotic cell, their DNA is extensively used in human population studies (Cann *et al.*, 1987; Cavalli-Sforza & Feldmann, 2003), because it exhibits a mutation rate that is an order of magnitude faster than nuclear genes. Therefore, human parasites showing a nucleotide diversity even higher than that of mtDNA can be expected to provide novel information on the history of human populations. For example, the polyomavirus JC shows a mean nucleotide diversity double than that of mtDNA. The greater amount of silent changes in JCV, with respect to human mitochondria, can be appreciated by comparing the trends of the mean synonymous diversity along the entire protein-coding region (Fig. 7).

Moreover, novel statistical methods have been developed to test the hypothesis of host-parasite codivergence. For example, reconciliation analysis (Treemap version 2.0, software distributed by Michael Charleston & Roderic Page; <http://taxonomy.zoology.gla.ac.uk/rod/treemap.html>) determines whether the two phylogenies are more congruent than expected by chance, based on randomization of both the host and parasite

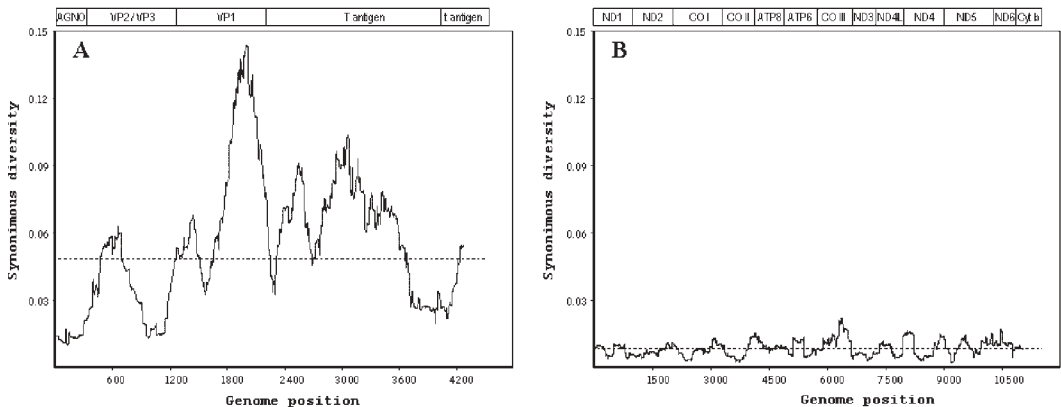


Fig. 7 - Trend of the rate of synonymous substitution, averaged over a sliding-window region of 100 codons, along the entire coding sequence of 273 JCV strains (a) and 156 human mitochondria (b). The dashed line indicates the mean value of the rate of synonymous change over the entire coding sequence (1524 codons for JCV and 3778 codons for mtDNA). At the top of the figure the organization of the coding region of JCV and mtDNA is presented (from Pavesi, 2005, with permission from the Society for General Microbiology).

phylogeny. Significant congruence between host and parasite phylogenies is interpreted as being the result of a shared evolutionary history.

Since the efficacy of parasites as surrogates for human genetic diversity depends on multiple factors, aim of the present paper was a critical evaluation of virtues and limitations of a number of human microbes, such as *P. falciparum*, *P. humanus*, *H. pylori*, two double-strand DNA and two positive-strand RNA viruses. *P. falciparum* could be useful for deciphering the pattern of human migrations dating back to Neolithic times, since it is probable that repeats 'imports' of plasmodia from moving populations have accompanied human history in the last thousand years. Obviously, this task could be carried out only in geographical areas where the problem of malaria still persists.

By contrast, head/body lice (*H. sapiens*) or tapeworms (*Taenia saginata* and *Taenia asiatica*) seem to provide a promising approach in probing even the origin of modern humans. In particular, inferences drawn from louse mtDNA (Reed *et al.*, 2004) suggest that there must have been some contact between distinct species of early humans. As proposed in a commentary paper (Pennisi 2004), further research focused on the pubic louse (*Phthirus pubis*) could suggest what kind of contact likely occurred among early human species.

Utility of *H. pylori* seems to be restricted to ancient human migrations dating to 10,000-15,000 years, or to migrations occurred in the last few thousand years (Fig. 3). The failure to obtain a phylogenetic tree with a basal clade including only African *H. pylori* strains does not allow to drawn inferences concerning the earliest human dispersal out of Africa.

By contrast, some viruses such as GBV-C/HGV or JCV have been found to be useful for tracing human migrations likely going back to the origin of modern humans. Moreover, primate evolutionary history could be revised by using GBV-C/HGV, which has been present in New World and Old World apes for more than 10 My. Several features of JCV (lack of pathogenous power, absence of genetic recombination, strong ethnicity, and easy detection) suggests it as the most suitable candidate to provide novel information on human history. Indeed, the geographical distribution of the two main lineages of JCV (Fig. 4) supports the

hypothesis that the expansion of *Homo sapiens* from Africa was mediated by two distinct migration waves. If validated, this hypothesis sheds really new light on the pattern of human migrations yielded so far by human genes, which supports the view of one single expansion from Africa into Asia and from there to the other continents (Cavalli-Sforza & Feldman, 2003).

A broader perspective on pathogens and human evolution

The interaction between pathogenous microbes and human host probably played an important role in determining and maintaining the surprising genetic diversity found in human populations. However, the study of human evolution has concentrated on humans and their hominid ancestors, without as much attention to other organisms also evolving in the same environments.

As significant causes of morbidity and mortality, parasites were particularly important to influence the evolution of human host. For example, the high degree of polymorphism in the major histocompatibility complex (MHC) is due to the selective pressure driven by disease (Reche *et al.*, 2003). In fact, MHC codes for membrane glycoproteins that play an important role in the immune system by binding fragments of infectious origin and presenting them to T-cells.

Malaria has exerted a powerful effect on human evolution by means of its aetiological agent (*P. falciparum*) and the associate mosquito vector (*Anopheles gambiae*). Selection for resistance to malaria has led to the appearance and persistence of a number of inherited diseases (thalassaemias, sickle cell disease, ovalocytosis and glucose-6-phosphate dehydrogenase) because of protective effects against illness and death (reviewed by Cooke *et al.*, 2004). Interaction between humans and infectious agents has also been the subject of theoretical studies (Fisher *et al.*, 1998).

Recent advances in molecular microbiology have provided a large amount of information about the evolutionary history of the most important human microbes, thus allowing us to decipher when they might have entered our line (Zimmer, 2001). For example, phylogenetic and divergence date analyses indicate that the occurrence of *Taenia* tapeworms in humans pre-dates the development

of agriculture, animal husbandry and domestication of cattle or swine (Hoberg *et al.*, 2001). Since the analysis of the amount of genetic variation among different species of tapeworms suggests a common ancestor living as long as 1 million years old, it is probable that tapeworms first infected our hominid ancestors as they scavenged grazing mammals on the African plains.

The ease with which our species acquires emerging infections from other animals points to the importance of these zoonoses as a source of new human diseases. Hirsch *et al.* (1989) showed that HIV-2, one major form of the virus that causes AIDS, is closely related to a virus that infects sooty mangabeys in West Africa. The monkey virus probably jumped into humans from sooty mangabeys kept as pets or hunted for food. More recently, Hahn *et al.* (2000) constructed a phylogeny of HIV-1, the far more common form of the virus, indicating that it crossed the species barrier from West African chimpanzees to humans around 1930. An emerging field of research concerns the detection of ancient microbial DNA in human and animal remains (Zink *et al.*, 2002). This approach, combined with phylogeny of extant infectious agents, has the potential to contribute significantly to a better understanding of the spread of infectious diseases in historic and modern times.

Obviously, the types of viruses that have afflicted humans throughout their evolutionary history changed in time. As human populations progressed technologically, viruses grew in numbers and density. As a consequence, different viruses found suitable conditions to thrive and establish long-lasting associations with man (Leal & Zanotto, 2000).

Most of the slowly evolving DNA viruses are ancient and have coevolved in close association with their hosts. Families such as the *Herpesviridae*, *Papovaviridae*, and *Adenoviridae* have cospeciated with vertebrates, and even the earliest hominids undoubtedly had their share. However, the life-long, mostly asymptomatic infections caused by these viruses are unlikely to have had a significant selective effect, unless introduced into a new host. Therefore, these viruses diversified and migrated along with human populations. Production of latent and subclinical infections allowed these pathogens to persist in the small populations that were probably typical of earlier hominids (Van

Blerkom, 2003).

Many RNA viruses, with their rapid rate of nucleotide substitution, were more recently acquired during the Neolithic, when humans and domesticated animals entered into intimate contact, and agricultural surpluses and permanent housing attracted disease-carrying rodents. Candidate viruses that thrived since the agricultural revolution up to recent historical times were mainly those causing acute transient infections, such as respiratory viruses, measles, smallpox, and so on. Post-Neolithic viruses include the diarrheal calicivirus and rotaviruses, the coronaviruses of common colds, and ortho- and paramyxoviruses that cause a variety of other respiratory disease (Van Blerkom, 2003). Finally, during the last few hundred years, viral disease previously confined to specific geographical areas (e.g., dengue, yellow fever, etc.) were able to spread widely, due to the increasing mobility of humans (Leal & Zanotto, 2000).

In conclusion, many aspects of human history can be clarified by accurate analyses of the molecular data obtained from human parasites. Molecular phylogenies may also provide means of mapping amino acid substitutions during the succession of endemic influenza strains (Fitch *et al.*, 1991). The use of viruses as experimental models can get insight into the long-lasting debate on selectionism and neutralism (Zanotto *et al.*, 1999). Finally, understanding the processes that give rise to the phylogeographic patterns observed in viruses (Holmes, 2004) can also be of great epidemiological and clinical importance.

Acknowledgements

The author is grateful to Franco Conterio for his support and encouragement. A special thank you goes to Adriano Cosenza for preparing the figures and to Maria Cristina Cignatta for language revision. David Reed (Florida Museum of Natural History, Gainesville), Mark Achtman (Max-Planck Institut für Infektionsbiologie, Berlin), and Thierry Wirth (University Konstanz) are also gratefully acknowledged for providing the electronic files of Figures 2, 3, and 5. David Reed (Florida Museum of Natural History), Mark Achtman (Max-Planck Institut, Berlin) and Thierry Wirth (University of Konstanz) are gratefully acknowledged for providing

the electronic files of figures 2,3 and 5. This work was financed by the MURST (Ministero dell'Università e della Ricerca Scientifica e Tecnologica).

Info on the Web

<http://PlasmoDB.org>

PlasmoDB is the official database of the *P. falciparum* genome sequencing consortium. It contains the complete nucleotide sequence of the *P. falciparum* genome. Sequence information is integrated with annotation and other genomic-scale data, including gene expression analysis from EST, SAGE and microarray projects and proteomics studies.

<http://genolist.fr/PyloriGene>

PyloriGene is a database dedicated to the analysis of the entire genome sequence of the Gram-negative bacterium *H. pylori*. It provides a complete a complete dataset of DNA and protein sequences derived from two different strains: 26695 and J99, linked to the relevant annotations and functional assignments.

<http://pritch.bsd.uchicago.edu>

The program STRUCTURE is a free software package for using multi-locus genotype data to investigate population structure. Its uses include inferring the presence of distinct populations and assigning individuals (or bacterial and viral strains) to populations. It can be applied to most of the commonly-used genetic markers, including microsatellites, RFLPs and SNPs.

<http://bioinfo.ernet.in/virgen/virgen.html>

VirGen is a comprehensive viral genome resource that organizes the information from viral genomes in a structured fashion (Kulkarni-Kale *et al.*, 2004). It has been developed with the objective of serving as an annotated and curated database comprising complete genome sequences of viruses. It also provides phylogenetic trees of viral species computed using whole-genome sequence data.

<http://taxonomy.zoology.gla.ac.uk/rod/tree-map.html>

Treemap is designated as a simple tool for visually comparing host and parasite phylogenies. It provides the host and parasite trees, reconstructs

the history of the host-parasite association, performs randomization tests of tree similarity, and finally compares branch lengths in the two trees.

References

- Agostini H.T., Yanagihara R., Davis V., Ryschkewitsch C.F. & Stoner G.L. 1997. Asian genotypes of JC virus in Native Americans and in a Pacific Island population: Markers of viral evolution and human migration. *Proc. Natl. Acad. Sci. USA*, 94: 14542-14546.
- Agostini H.T., Shishido-Hara Y., Baumhefner R.W., Singer E.J., Ryschkewitsch C.F. & Stoner G.L. 1998. JC virus Type 2: Definition of subtypes based on DNA sequence analysis of ten complete genomes. *J. Gen. Virol.*, 79: 1143-1151.
- Agostini H.T., Deckhut A., Jobes D.V., Girones R., Schlunck G., Prost M.G., Frias C., Pérez-Trallero E., Ryschkewitsch C.F. & Stoner G.L. 2001. Genotypes of JC virus in East, Central and Southwest Europe. *J. Gen. Virol.*, 82: 1221-1231.
- Anderson T.J.C., Haubold B., Williams J.T., Estrada-Franco J.G., Richardson L., Mollinedo R., Bockarie M., Mokili J., Mharakurwa S., French N., Whitworth J., Velez I.D., Brockman A.H., Nosten F., Ferreira M.U. & Day K.P. 2000. Microsatellite markers reveal a spectrum of population structure in the malaria parasite *Plasmodium falciparum*. *Mol. Biol. Evol.*, 17: 1467-1482.
- Berry A.E., Gardner M.J., Caspwr G.J., Roos D.S. & Berriman M. 2004. Curation of the *Plasmodium falciparum* genome. *Trends Parasitol.*, 20: 548-552.
- Biggar R.J., Taylor M.E., Neel J.V., Hjelle B., Levine P.H., Black F.L., Sgaw G.M., Sharp P.M. & Hahn B.H. 1996. Genetic variants of human T-lymphotropic virus type II in American Indian groups. *Virology*, 216: 165-173.
- Cann R.L., Stoneking M. & Wilson A.C. 1987. Mitochondrial DNA and human evolution. *Nature*, 325: 31-36.
- Cavalli-Sforza L.L., Piazza A., Menozzi P. & Mountain J. 1988. Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data. *Proc. Natl. Acad. Sci. USA*, 85: 6002-6006.

- Cavalli-Sforza L.L., Menozzi P. & Piazza A. 1994. *Genetic history of world populations*. In *The history and geography of human genes*, pp. 60-157. Princeton, Princeton University Press.
- Cavalli-Sforza L.L. & Feldman M.W. 2003. The application of molecular genetic approaches to the study of human evolution. *Nat. Genet.*, 33: 266-275.
- Chan S.Y., Ho L., Ong C.K., Chow V., Drescher B., Durst M., ter-Meulen J., Villa L., Luande J., Mgaya H.N. & Bernard H.U. 1992. Molecular variants of human papillomavirus type 16 from four continents suggest ancient pandemic spread of the virus and its coevolution with humankind. *J. Virol.*, 66: 2057-2066.
- Chang H., Wang M., Tsai R.T., Lin H.S., Huan J.S., Wang W.C. & Chang D. 2002. High incidence of JC viruria in JC-seropositive older individuals. *J. Neurovirol.*, 8: 447-451.
- Cheng P.N., Chang T.T., Jen C.M., Ko A.W., Young D.C. & Wu H.L. 2003. Molecular evidence for transmission of GB virus-C/hepatitis G virus infection within family: close relationship between mother and child. *Hepatogastroenterol.*, 50: 151-156.
- Cheung R.C., Keeffe E.B. & Greenberg H.B. 1997. Hepatitis G virus: Is it a hepatitis virus? *West. J. Med.*, 167: 23-33.
- Coluzzi M. 1999. The clay feet of the malaria giant and its African roots: hypotheses and inferences about origin, spread and control of *Plasmodium falciparum*. *Parassitologia*, 41: 277-283.
- Cooke B.M., Mohandas N. & Coppel R.L. 2004. Malaria and the red blood cell membrane. *Semin. Hematol.*, 41: 173-188.
- Covacci A., Telford J.L., Del Giudice G., Parsonnet J. & Rappuoli R. 1999. *Helicobacter pylori* virulence and genetic geography. *Science*, 284: 1328-1333.
- Curat M., Trabuchet G., Rees D., Perrin P., Harding R.M., Clegg J.B., Langaney A. & Excoffier L. 2002. Molecular analysis of the beta-globin gene cluster in the Niokholo Mandenka population reveals a recent origin of the beta(S) Senegal mutation. *Am. J. Hum. Genet.*, 70: 207-223.
- Desai M.M., Pal R.B. & Banker D.D. 2004. GB virus C/hepatitis G virus infection in Indian blood donors and high risk groups. *Transfus. Apheresis Sci.*, 30: 111-117.
- Escalante A.A. & Ayala F.J. 1994. Phylogeny of the malarial genus *Plasmodium*, derived from rRNA gene sequences. *Proc. Natl. Acad. Sci. USA*, 91: 11373-11377
- Escalante A.A., Barrio E. & Ayala F.J. 1995. Evolutionary origin of human and primate malaria: evidence from the circumsporozoite protein gene. *Mol. Biol. Evol.*, 12: 616-626.
- Falush D., Kraft C., Taylor N.S., Correa P., Fox J.G., Achtman M. & Suerbaum S. 2001. Recombination and mutation during long-term gastric colonization by *Helicobacter pylori*: estimates of clock rates, recombination size, and minimal age. *Proc. Natl. Acad. Sci. USA*, 98: 15056-15061.
- Falush D., Wirth T., Linz B., Pritchard J.K., Stephens M., Kidd M., Blaser M.J., Graham D.Y., Vacher S., Perez-Perez G.I., Yamaoka Y., Mégraud F., Otto K., Reichard U., Katzowitsch E., Wang X., Achtman M. & Suerbaum S. 2003a. Traces of human migrations in *Helicobacter pylori* populations. *Science*, 299: 1582-1585.
- Falush D., Stephens M. & Pritchard J.K. 2003b. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, 164: 1567-1588.
- Fisher C., Jock B. & Vogel F. 1998. Interplay between humans and infective agents: a population genetic study. *Hum. Genet.*, 102: 415-422.
- Fitch W.M., Leiter J.M.E., Li X. & Palese P. 1991. Positive Darwinian evolution in human influenza A viruses. *Proc. Natl. Acad. Sci. USA*, 88: 4270-4274.
- Frisque G.L. & Cannella M.T. 1984. Human polyomavirus JC virus genome. *J. Virol.*, 51: 458-469.
- Gessain A., Maucélé P., Froment A., Biglione M., Le Hesran J.Y., Tekaiia F., Millan J. & de Thé G. 1995. Isolation and molecular characterization of a human T-cell lymphotropic virus type II (HTLV-II), subtype B, from a healthy pygmy living in a remote area of Cameroon: an ancient origin for HTLV-II in Africa. *Proc. Natl. Acad. Sci. USA*, 92: 4041-4045.
- Ghose C., Perez-Perez G.I., Dominguez-Bello M.G., Pride D.T., Bravi C.M. & Blase M.J.

2002. East Asian genotypes of *Helicobacter pylori* strains in Amerindians provide evidence for its ancient human carriage. *Proc. Natl. Acad. Sci. USA*, 99: 15107-15111.
- Hafner M.S. & Nadler S.A. 1988. Phylogenetic trees support the coevolution of parasites and their hosts. *Nature*, 332: 258-259.
- Hahn B.H., Shaw G.M., De-Cock K.M. & Sharp P.M. 2000. AIDS as a zoonosis: scientific and public health implications. *Science*, 287: 607-614.
- Hall W.W., Ishak R., Zhu S.V., Novoa P., Eiraku N., Takahashi H., Ferreira C., Azevedo V., Ishak M.O., Monken C. & Kurata T. 1996. Human T lymphotropic virus type II (HTLV-II): epidemiology, molecular properties, and clinical features of infection. *J. Acquir. Immuno Defic. Syndr. Hum. Retrovirol.*, 13: S204-S214.
- Harpending H.C., Sherry S.T., Rogers A.R. & Stoneking M. 1993. The genetic structure of ancient human populations. *Curr. Anthropol.*, 34: 483-496.
- Hartl D.L., Volkman S.K., Nielsen K.M., Barry A.E., Day K.P., Wirth D.F. & Winzeler E.A. 2002. The paradoxical population genetics of *Plasmodium falciparum*. *Trends Parasitol.*, 18: 266-272.
- Hildesheim A., Schiffman M.H., Gravitt P.E., Glass A.G., Greer C.E., Zhang T., Scott D.R., Rush B.B., Lawler P., Sherman M.E., Kurman R.J. & Manes M.M. 1994. Persistence of type-specific human papillomavirus infection among cytologically normal women. *J. Infect. Dis.*, 169: 235-240.
- Hirsch V.M., Olmsted R.A., Murphey-Corb M., Purcell R.H. & Johnson P.R. 1989. An African primate lentivirus (SIVsm) closely related to HIV-2. *Nature*, 339: 389-392.
- Ho L., Chan S.Y., Burk R.D., Das B.C., Fujinaga K., Icenogle J.P., Kahn T., Kiviat N., Lancaster W. W., Mavromara-Nazos P., Labropoulou V., Mitrani-Rosenbau T., Radhakrishna Pillai M., Stoerker J., Syrjaenen K., Syrjaenen S., Tay S.K., Villa L.L., Wheeler C.M., Williamson A.L. & Bernard H.U. 1993. The genetic drift of human papillomavirus type 16 is a means of reconstructing prehistoric viral spread and the movement of ancient human populations. *J. Virol.*, 67: 6413-6423.
- Hoberg E.P., Alkire N.L., de Queiroz A. & Jones A. 2001. Out of Africa: origins of the *Taenia* tapeworms in humans. *Proc. R. Soc. Lond. B.*, 268: 781-787.
- Holmes E.C. 2004. The phylogeography of human viruses. *Mol. Ecol.*, 13: 745-756.
- Hume J., Lyons E.J. & Day K.P. 2003a. Malaria in antiquity: a genetics perspective. *World Archaeol.*, 35: 180-192.
- Hume J., Lyons E.J. & Day K.P. 2003b. Human migration, mosquitoes and the evolution of *Plasmodium falciparum*. *Trends Parasitol.*, 19: 144-149.
- Ingman M., Kaessmann H., Paabo S. & Gyllensten U. 2000. Mitochondrial genome variation and the origin of modern humans. *Nature*, 408: 708-713.
- Joy D.A., Feng X., Mu J., Furuya T., Chotivanich K., Krettli A.U., Ho M., Wang A., White N.J., Suh E., Beerli P., Su X.Z. 2003. Early origin and recent expansion of *Plasmodium falciparum*. *Science*, 300: 318-321.
- Kimura M. 1983. *The Neutral Theory of Molecular Evolution*. Cambridge, Cambridge University Press.
- Kitamura T., Sugimoto C., Kato A., Ebihara H., Suzuki M., Taguchi F., Kawabe K. & Yogo Y. 1997. Persistent JC virus (JCV) infection is demonstrated by continuous shedding of the same JCV strains. *J. Clin. Microbiol.*, 35: 1255-1257.
- Kittler R., Kayser M. & Stoneking M. 2003. Molecular evolution of *Pediculus humanus* and the origin of the clothing. *Curr. Biol.*, 13: 1414-1417.
- Konomi N., Miyoshi C., La Fuente Zerain C., Li T.C., Arakawa Y. & Abe K. 1999. Epidemiology of hepatitis B, C, E, and G virus infections and molecular analysis of hepatitis G virus isolates in Bolivia. *J. Clin. Microbiol.*, 37: 3291-3295.
- Kivi M., Tindberg Y., Sorberg M., Casswall T.H., Befrits R., Hellstrom P.M., Bengtsson C., Engstrand L. & Granstrom M. 2003. Concordance of *Helicobacter pylori* strains within families. *J. Clin. Microbiol.*, 41: 5604-5608.
- Kulkarni-Kale U., Bhosle S., Sunitha Manjari G. & Kolaskar A.S. 2004. VirGen: a comprehensive viral genome resource. *Nucleic Acids Res.*, 32:

- D289-D292.
- Kunitake T., Kitamura T., Guo J., Taguchi F., Kawabe K. & Yogo Y. 1995. Parent-to-child transmission is relatively common in the spread of the human polyomavirus JC. *J. Clin. Microbiol.*, 33: 1448-1451.
- Leal E.S. & Zanotto P.M. 2000. Viral diseases and human evolution. *Mem. Inst. Oswaldo Cruz*, 1: 193-200.
- Leary T.P., Muerhoff A.S., Simons J.N., Pilot-Matia T.J., Erker J.C., Chalmers M.L., Schlauder G.G., Dawson G.J., Desai S.M. & Mushahwar I.K. 1996. Sequence and genomic organization of GBV-C: a novel member of the *Flaviviridae* associated with human non-A-E hepatitis. *J. Med. Virol.*, 48: 60-67.
- Lee H., Swanson P., Shorty V.S., Zack J.A., Rosenblatt J.D. & Chen I.S.Y. 1989. High rate of HTLV-II infection in seropositive IV drug abusers in New Orleans. *Science*, 244: 471-475.
- Linnen J., Wages J., Zhang-Keck Z.Y., Fry K.E., Krawczynski K.Z., Alter H., Koonin E., Gallagher M., Alter M., Hadziyannis S., Karayiannis P., Fung K., Nakatsuji Y., Shih J.W.K., Young L., Piatak M., Hoover C., Fernandez J., Chen S., Zou J.C., Morris T., Hyams K.C., Ismay S., Lifson J.D., Hess G., Fong S.K.H., Thomas H., Bradley D., Margolis H. & Kim J.P. 1996. Molecular cloning and disease association of hepatitis G virus: a transfusion-transmissible agent. *Science*, 271: 505-508.
- Mahieux R., Chappey C., Georges-Courbot M.C., Dubreuil G., Mauclore P., Georges A. & Gessain A. 1998. Simian lymphotropic virus type 1 from *Mandrillus sphinx* as a simian counterpart of human T-cell lymphotropic virus type 1 subtype D. *J. Virol.*, 72: 10316-10322.
- Marshall B.J. & Warren J.R. 1984. Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. *Lancet*, 1: 1311-1315.
- Miura T., Fukunaga T., Igarashi T., Yamashita M., Ido E., Funahashi S.I., Ishida T., Washio K., Ueda S., Hashimoto K.I., Yoshida M., Osame M., Singhal B.S., Zaninovic V., Cartier L., Sonoda S., Tajima K., Ina Y., Gojobori T. & Hayami M. 1994. Phylogenetic subtypes of human T-lymphotropic virus type I and their relations to the anthropological background. *Proc. Natl. Acad. Sci. USA*, 91: 1124-1127.
- Mu J., Duan J., Makova K.D., Joy D.A., Huynh C.Q., Branch O.H., Li W.H. & Su X.Z. 2002. Chromosome-wide SNP reveal an ancient origin for *Plasmodium falciparum*. *Nature*, 418: 323-326.
- Munoz N. 2000. Human papillomavirus and cancer: the epidemiological evidence. *J. Clin. Virol.*, 19: 1-5.
- Naito H., Win K.M. & Abe K. 1999. Identification of a novel genotype of hepatitis G virus in Southeast Asia. *J. Clin. Microbiol.*, 37: 1217-1220.
- Ong, C. K., Chan, S. Y., Campo, M. S., Fujinaga K., Maovromara-Nazos P., Labropoulou V., Pfister H., Tay S.K., ter Meulen J., Villa L.L. & H.U. Bernard 1993. Evolution of human papillomavirus type 18: an ancient phylogenetic root in Africa and intratype diversity reflect coevolution with human ethnic groups. *J. Virol.*, 67: 6424-6431.
- Osame M., Matsumoto M., Usuku K., Izumo S., Ijichi N., Amitani H., Tara M. & Igata A. 1987. Chronic progressive myelopathy associated with elevated antibodies to human T-lymphotropic virus type 1 and adult T-cell leukaemia-like cells. *Annals Neurol.*, 21: 117-122.
- Pavesi A. 2000. Detection of signature sequences in overlapping genes and prediction of a novel overlapping gene in hepatitis G virus. *J. Mol. Evol.*, 50: 284-295.
- Pavesi A. 2001. Origin and evolution of GBV-C/hepatitis G virus and relationships with ancient human migrations. *J. Mol. Evol.*, 53: 104-113.
- Pavesi A. 2003. African origin of polyomavirus JC and implications for prehistoric human migrations. *J. Mol. Evol.*, 56: 564-572.
- Pavesi A. 2004. Detecting traces of prehistoric human migrations by geographic synthetic maps of Polyomavirus JC. *J. Mol. Evol.*, 58: 304-313.
- Pavesi A. 2005. Utility of JC polyomavirus in tracing the pattern of human migrations dating to prehistoric times. *J. Gen. Virol.*, 86: 1315-1326.
- Pennisi E. 2004. Louse DNA suggests close contact between early humans. *Science*, 306: 210.
- Reche P.A. & Reinherz E.L. 2003. Sequence variability analysis of human class I and class II MHC molecules: functional and structural correlates of amino acid polymorphisms. *J. Mol.*

- Biol.*, 331: 623-641.
- Reed D.L., Smith V.S., Hammond S.L., Rogers A.R. & Clayton D.H. 2004. Genetic analysis of lice supports direct contact between modern and archaic humans. *PLOS Biol.*, 2: 1972-1983.
- Rich S.M., Licht M.C., Hudson R.R. & Ayala F. 1998. Malaria's Eve: evidence of a recent population bottleneck throughout the world populations of *Plasmodium falciparum*. *Proc. Natl. Acad. Sci. USA*, 95: 4425-4430.
- Salemi M., Vandamme A.M., Gradozzi C., Van Laethem K., Cattaneo E., Taylor G., Casoli C., Goubau P., Desmyter J. & Bertazzoni U. 1998. Evolutionary rate and genetic heterogeneity of human T-cell lymphotropic virus type II using isolates from European injecting drug users. *J. Mol. Evol.*, 46: 602-611.
- Sathar M.A., Soni P., Naicker S., Conradie J., Lockhat F. & Gouws E. 1999. GB virus C/hepatitis G virus infection in KwaZulu Natal, South Africa. *J. Med. Virol.*, 59: 38-44.
- Scallan M.F., Clutterbuck D., Jarvis L.M., Scott G.R. & Simmonds P. 1998. Sexual transmission of GB virus C/hepatitis G virus. *J. Med. Virol.*, 55: 203-208.
- Seemayer C.A., Viazov S., Philipp T. & Roggendorf M. 1998. Detection of GBV-C/HGV RNA in saliva and serum, but not in urine of infected patients. *Infection*, 26: 39-41.
- Simmonds P. & Smith D.B. 1999. Structural constraints on RNA virus evolution. *J. Virol.*, 73: 5787-5794.
- Slattery J.P., Franchini G. & Gessain A. 1999. Genomic evolution, patterns of global dissemination, and interspecies transmission of human and simian T-cell leukemia/lymphotropic viruses. *Genome Res.*, 9: 525-540.
- Smith D.B., Basaras M., Frost S., Haydon D., Cucenau N., Prescott L., Kamenka C., Millband D., Sathar M.A. & Simmonds P. 2000. Phylogenetic analysis of GBV-C/HGV virus. *J. Gen. Virol.*, 81: 769-780.
- Spratt B.G. 2003. Stomachs out of Africa. *Science*, 299: 1528-1529.
- Stringer C.B. 2003. Out of Ethiopia. *Nature*, 423: 692-695.
- Suerbaum S., Bapumia K., Morelli G., Smith N.H., Kunstmann E., Dyrek I. & Achtman M. 1998. Free recombination within *Helicobacter pylori*. *Proc. Natl. Acad. Sci. USA*, 95: 12619-12624.
- Sugimoto C., Kitamura T., Guo J., Al-Ahdal M.N., Shchelkunov S.N., Otova B., Ondrejka P., Chollet J.Y., El-Safi S., Ettayebi M., Grésenguet G., Kocagöz T., Chaiyarasamee S., Thant K.Z., Thein S., Moe K., Kobayashi N., Taguchi F. & Yogo Y. 1997. Typing of urinary JC virus DNA offers a novel means of tracing human migrations. *Proc. Natl. Acad. Sci. USA*, 94: 9191-9196.
- Suzuki Y., Katayama K., Fukushi S., Kageyama T., Oya A., Okamura H., Tanaka Y., Mizokami M. & Gojobori T. 1999. Slow evolutionary rate of GB virus C/hepatitis G virus. *J. Mol. Evol.*, 48: 383-389.
- Switzer W., Owen S., Pleniasek D., Nerurkar V., Duenas-Barajas E., Heneine W. & Lal R. 1995. Molecular analysis of human T-cell lymphotropic virus type II from Wayuu Indians of Colombia demonstrates two subtypes of HTLV-IIb. *Virus Genes*, 10: 153-162.
- Taira A.V., Nukermans C.P. & Sanders G.D. 2004. Evaluating human papillomavirus vaccination programs. *Emerg. Infect. Dis.*, 10: 1915-1923.
- Takayama S., Miura T., Tominaga T., Taki M., Matuso S., Sugi S. & Shimotohno K. 1999. Partial nucleotide sequencing of the NS3/helicase region of hepatitis G virus to prove vertical transmission. *FEMS Microbiol. Lett.*, 175: 273-279.
- Tanaka Y., Mizokami M., Orito E., Ohba K.I., Kato T., Kondo Y., Mboudjeka I., Zekeng L., Kaptue L., Bikandou B., M'Pele P., Takehisa J., Hayami M., Suzuki Y. & Gojobori T. 1998. African origin of GB virus C/hepatitis G virus. *FEBS Lett.*, 423: 143-148.
- Tishkoff S.A., Varkonyi R., Cahinhinan N., Abbes S., Argyropoulos G., Desro-Bisol G., Drousiotou A., Dangerfield B., Lefranc G., Loiselet J., Piro A., Stoneking M., Tagarelli A., Tagarelli G., Touma E.H., Williams S.M. & Clark A.G. 2001. Haplotype diversity and linkage disequilibrium at human G6PD: recent of alleles that confer malarial resistance. *Science*, 293: 455-462.
- Ureta Vidal A., Gessain A., Yoshida M., Tekaja F., Garin B., Guilleman B., Schulz T., Farid R. & De Thé G. 1994. Phylogenetic classification of human T cell leukaemia/lymphoma virus type I genotypes in five major molecular and

- geographical subtypes. *J. Gen. Virol.*, 75: 3655-3666.
- Van Blerkom L.M. 2003. Role of viruses in human evolution. *Am. J. Phys. Anthropol.*, 46: 14-46.
- Vandamme A.M., Salemi M., Van Brussel M., Liu H.F., Van Laethem K., Van Rast M., Michels L., Desmyter J. & Goubau P. 1998. African origin of human T-lymphotropic virus type 2 (HTLV-2) supported by a potential new HTLV-2d subtype in Congolese Bambuti Efe Pygmies. *J. Virol.*, 72: 4327-4340.
- Volkman S.K., Barry A.E., Lyons E.J., Nielsen K.M., Thomas S.M., Choi M., Thakore S.S., Day K.P., Wirth D.F. & Hartl D.L. 2001. Recent origin of *Plasmodium falciparum* from a single progenitor. *Science*, 293: 482-484.
- Wattel E., Vartanian J.P., Pannetier C. & Wain-Hobson S. 1995. Clonal expansion of human T-cell leukemia virus type I infected cells in asymptomatic and symptomatic carriers without malignancy. *J. Virol.*, 69: 2863-2868.
- White T.D., Asfaw B., DeGusta D., Gilbert H., Richards G.D., Suwa G. & Howell F.C. 2003. Pleistocene *Homo sapiens* from Middle Awash, Ethiopia. *Nature*, 423: 742-747.
- Wirth T., Wang X., Linz B., Novick R.P., Lum J.K., Blaser M., Morelli G., Falush D. & Achtman M. 2004. Distinguishing human ethnic groups by means of sequences from *Helicobacter pylori*: lessons from Ladakh. *Proc. Natl. Acad. Sci. USA*, 101: 4746-4751.
- Wirth T., Meyer A. & Achtman M. 2005. Deciphering host migrations and origins by means of their microbes. *Mol. Ecol.*, 14: 3289-3306.
- Wooding S. 2001. Do human and JC virus show evidence of host-parasite codemography? *Infect. Genet. Evol.*, 1: 3-12.
- Worobey M. & Holmes E.C. 2001. Homologous recombination in GB virus C/hepatitis G virus. *Mol. Biol. Evol.*, 18: 254-261.
- Yamada T., Manos M.M., Peto J., Greer C.E., Munoz M., Bosch F.X. & Wheeler C.M. 1997. Human papillomavirus type 16 sequence variation in cervical cancers: a worldwide perspective. *J. Virol.*, 71: 2463-2472.
- Yamaoka Y., Orito E., Mizokami M., Gutierrez O., Saitou N., Kodama T., Osato M.S., Kim J.G., Ramirez F.C., Mahachai V. & Graham D.Y. 2002. *Helicobacter pylori* in North and South America before Columbus. *FEBS Lett.*, 517: 180-184.
- Yanagihara R., Nerurkar V.R., Scheirich I., Agostini H.T., Mgone C.S., Cui X., Jobes D.V., Cubitt C.L., Ryschewitsch C.F., Hrady D.B., Friedlaender J.S., Stoner G.L. 2002. JC virus genotypes in the Western Pacific suggest Asian mainland relationships and virus association with early population movements. *Hum. Biol.*, 74: 473-488.
- Yoshida M., Miyoshi I. & Hinuma Y. 1982. Isolation and characterization of retrovirus from cell lines of human adult T-cell leukaemia and implications in the disease. *Proc. Natl. Acad. Sci. USA*, 79: 2031-2035.
- Zanotto P.M.A., Kallas E.G., Souza R.F. & Holmes E.C. 1999. Evidence for positive selection in the *nef* gene of HIV-1. *Genetics*, 153: 1077-1089.
- Zimmer C. 2001. Genetic trees reveal disease origins. *Science*, 292: 1090-1093.
- Zink A.R., Reischl U., Wolf H. & Nerlich A.G. 2002. Molecular analysis of ancient microbial infections. *FEMS Microbiol. Lett.*, 213: 141-147.