Network analysis of the hominin origin of Herpes Simplex virus 2 from fossil data

Simon Underdown[1,2], **Krishna Kumar**[3], **and Charlotte Houldcroft**[4,5,*]

[1]Human Origins and Palaeoenvironmental Research Group (HOPE), Department of Anthropology & Geography, Oxford Brookes University, Oxford, OX3 0BP, UK.
[2]Leverhulme Centre for Human Evolutionary Studies, University of Cambridge, Henry Wellcome Building, Fitzwilliam Street, Cambridge, CB2 1QH, UK.
[3]Computational Geomechanics, Cambridge University Engineering Department, Trumpington Street, Cambridge, CB2 1PZ, UK.
[4]Division of Biological Anthropology, Department of Archaeology & Anthropology, University of Cambridge, Cambridge, CB2 3QG, UK.
[5]McDonald Institute for Archaeological Research, University of Cambridge, Downing Street, Cambridge, CB2 3ER, UK.
[*]ch504@cam.ac.uk

ABSTRACT

Herpes simplex virus 2 is a human herpesvirus found worldwide that causes genital lesions and more rarely causes encephalitis. This pathogen is most common in Africa, and particularly in central and east Africa, an area of particular significance for the evolution of modern humans. Unlike HSV1, HSV2 has not simply co-speciated with humans from their last common ancestor with primates. HSV2 jumped the species barrier between 1.4 and 3 MYA, most likely through intermediate but unknown hominin species.

In this paper, we use probability-based network analysis to determine the most probable transmission route between intermediate hosts of HSV2, from the ancestors of chimpanzees to the ancestors of modern humans, using paleo-environmental data on the distribution of African tropical rainforest over the last 3 million years and data on the age and distribution of fossil species of hominin present in Africa between 1.4 and 3 MYA. Our model identifies *Paranthropus boisei* as the most likely intermediate host of HSV2, while *Homo habilis* may also have played a role in the initial transmission of HSV2 from the ancestors of chimpanzees to *P. boisei*.

Keywords network analysis; human evolution; infectious disease; epidemiology; archaeology

Introduction

Herpes simplex virus 2 (HSV2) is a sexually transmitted human pathogen that causes genital lesions and, rarely, encephalitis (eg Tang et al. (2003)), and is associated with increased risk of HIV acquisition (Freeman et al., 2006). After primary infection, the virus adopts a life cycle of latency punctuated by periods of lytic replication when new hosts can be infected through genital contact. The virus is related to the human oral pathogen herpes simplex virus 1 (HSV1). Both HSV1 and HSV2 are alphaherpesviruses, which are found in many primates (Wertheim et al., 2014). HSV1 primarily causes infection, and sporadic lesions, within the oral cavity and establishes latency in the trigeminal ganglia, while HSV-2 is associated with infection of the genitalia and surrounding skin, and establishes latency in the sacral ganglia (**Whitley2007**). Both simplex viruses can infect either body cavity, although there is viral shedding data from co-infected individuals to suggest that HSV1 is a more successful oral and HSV2 a more successful genital pathogen (**Kim2006**). The genetic basis of this difference in tropism isn't fully understood: some studies of recombinant HSV1 x HSV2 strains have highlighted the importance of the latency-associated transcript (LAT) in how successfully the two simplex viruses reactivate from latency within different nerve types (**Bertke2007**). HSV1 and 2 also differ in the length of the glycoprotein G (US4) open reading frame, which may play a role in tropism (**Baines2007**).

HSV2 was originally thought to have co-speciated with humans when our lineage diverged from that of the ancestors of chimpanzees and bonobos (anc-chimps). Comparisons of the HSV1, HSV2 and chimpanzee herpes virus 1 (ChHV1) genomes (Tang et al., 2003) suggest that HSV2 is more closely related to ChHV1 than HSV1 (Wertheim et al., 2014). This analysis also found that ChHV1 and HSV2 diverged from one another between 1.4 and 3 MYA, and the authors inferred that an unknown hominin (or hominins) was infected with HSV2 before it switched host to the ancestors of modern humans.

We hypothesise that by combining fossil data on when and where different hominin species were likely to be present in Africa, the geographical range of modern chimpanzees and bonobos, and the reconstructed distribution of tropical rainforest habitat as a proxy for the past range of the chimpanzee/bonobo ancestor (anc-chimps) it will be possible to develop a model to statistically infer the species that facilitated the host-switch of HSV2 in the modern human lineage defined here as beginning with *Homo erectus* (Anton et al., 2016).

Before *Homo erectus* it is unclear which hominins are the direct ancestors of anatomically modern humans. However, once HSV2 has reached *Homo erectus*, no further host-switches are required for HSV2 to be considered present in the ancestors of modern humans.

HSV2 is found in all living populations (Looker et al., 2015), is accepted as having an African origin (**Koelle2017**; Burrel, Boutolleau, et al., 2017), and has patterns of genetic divergence consistent with having diverged with human populations as they spread out of Africa (**Koelle2017**). Current genetic, archaeological and fossil evidence suggests 100 KYA as a plausible (although not universally accepted) earliest date for anatomically modern humans (AMH) to have left Africa Mirazón Lahr et al., 2016. We therefore infer that HSV2 must have been in the population of AMH before they left Africa, in order to take it with them when migrating to the rest of the world.

Methods

HSV2 prevalence data, hominin fossil data and chimpanzee and tropical rainforest geographic range data

HSV2 is most closely related to ChHV1 which infected the ancestors of modern chimpanzees. The GIS data on the range of modern *Pan troglodytes* and *Pan paniscus*, provided by the IUCN Red List (Oates et al., 2008), is shown in fig. 3a. Only one ancestral chimpanzee fossil, dated to c. 500 KYA, is currently known (**Mcbrearty2005**). This means that the habitat range of anc-chimps is not directly measurable from the fossil record but is inferred to be a larger geographical range. Therefore, we have used the paleo-tropical rainforest range during the period of 1.4 - 3 MYA Köppen-Geiger climate classification dataset Peel, Finlayson, and McMahon, 2007 as a proxy for the ancient range of chimpanzees and combined this with data based on modern great ape distribution patterns and range size estimates (**Thompson2003**), and is shown in fig. 3b.

To identify which species could have been involved in the host-switch of HSV2 from the ancestors of chimpanzees to the ancestors of modern humans, we collated spatio-temporal data on African hominin fossil species extant between 100 KYA and 3 MYA. Latitude and longitude of site location was used to provide a spatial data point for each species. Temporally, published fossil dating evidence was used to provide a first appearance datum (FAD) and last appearance datum (LAD) for each species (see supplementary Table 1).

Data on the prevalence of herpes simplex virus 2 between 2000 and 2015 CE was taken from the supplementary materials of (Looker et al., 2015), and plotted to demonstrate the distribution of HSV2 across Africa (Supplementary Figure A. 1).

Network analysis

To establish the most probable transmission route of HSV2 from anc-chimps to modern humans, it is important to identify potential species that may have been intermediary hosts. Initially, all hominins in Africa with temporal ranges within the 1.4 - 3 MYA confidence window (Wertheim et al., 2014) of the chimp-hominin transmission were identified (see Supplementary Table 1). Their distance to ancient tropical rainforest was calculated, and only those species whose fossil remains were found within 400 km of tropical rainforest were considered as putative species for initial ancestral-chimp-hominin HSV2 transmission. This reflects the distance that would have been covered by hominins employing three possible strategies: (a) a broadly omnivorous subsistence strategy based on scavenging, (b) hunting using carnivore and herbivore movement patterns and (c) modern hunter-gatherer range sizes (**Foley1978**; **Grant1992**). A matrix of spatio-temporal distances was then calculated to map the distances between the nearest neighbours of each species and also to calculate the temporal overlap between species using the fossil record. Fossils from four genera (*Ardipithecus*, *Kenyanthropus*, *Orrorin* and *Sahelanthropus*) were excluded from the analysis on the basis that there is no fossil evidence that they persisted after 3 MYA.

A network of possible transmission routes of HSV2 from anc-chimps to humans through different hominins was developed as a directed acyclic graph (DAG) $G = (V, E)$ comprising of a set of nodes (V), representing potential intermediary hosts, and edges (E) connecting the nodes, which represents the direction of transmission between species (See Figures 1 and 2). The DAG comprised of the anc-chimps as the start node, *H. erectus* as the target node, and other potential species forming secondary nodes in the graph. The edges are typically weighted; in this analysis, the weights are based on the inverse probability of transmission.

If HSV2 was transmitted to *Homo erectus*, no further cross-species transmission event is needed to explain the infection of modern humans. Simple vertical mother-to-child or horizontal (sexual) transmission of the virus through the genus *Homo* from this point would be sufficient as the ancestor-descendent path from *Homo erectus* to *Homo sapiens* is relatively secure (Maslin, Shultz, and Trauth, 2015).

Bayesian inference

A Bayesian network or a belief network is a graphical structure that allows us to represent and reason about an uncertain domain. In a Bayesian network, the species (nodes) are variables, and the transmission routes (edges) represent direct links between species. The process of conditioning (also called probability propagation or inference) is performed via a "flow of decisions" through the network, which involves computing the posterior probability distribution for a set of query nodes, given values for some evidence (or observation) nodes. An important consideration in all Bayesian-based methods is the choice of a prior. An empirical Bayesian method that estimates the likelihood of HSV2 infection using a prior beta distribution is adopted, which is an approximation to a hierarchical Bayesian approach (Farine and Strandburg-Peshkin, 2015; Murphy, 2012).

Bayesian networks provide full representations of probability distributions over their variables, which allows us to infer upon any subset of variables. A Bayesian network is created using the DAG described above. Each node on the graph represents a potential intermediary host that has a probability of transmitting HSV2. The probability of infection transmission is represented as a beta distribution, with shape parameters *alpha* representing the time period of the species in 100,000 years and *beta* representing the distance to the neighbour in kilometres. A *Conditional Probability Table* (CPT) is generated for all possible combinations of a dichotomous outcome for each variable (HSV2 infection is true or false). The "*Combined Inference*" approach is adopted to evaluate the probability of intermediary hosts transmitting HSV2, by conditioning the ancestral-chimpanzee and *Homo erectus* nodes for the presence of HSV2 (Korb and Ann E. Nicholson, 2003). The Bayesian inference analysis was performed using the AI space decision network tool (Poole and Mackworth, 2010).

Optimal path traversal

A* is an informed search algorithm that searches through all possible paths to the target that yields the smallest cost (Hart, Nilsson, and Raphael, 1968). This is done by combining information by favouring vertices that are close to the starting point and to the target. At each time-step the A* algorithm selects the path at a given vertex n that has the lowest $f(n) = g(n) + h(n)$, where $g(n)$ represents the exact cost of the path from the starting point to any vertex n, and $h(n)$ represents the heuristic estimated cost from vertex n to the goal. A* balances the two as it moves from the starting node to the target node. The most probable transmission route is evaluated by minimising the traversal costs based on the edge cost and nodal heuristics of the transmission network graph. A Python script with the *NetworkX* package was used to implement the A* path-finding algorithm on the DAGs to identify the most probable HSV2 transmission (optimal path) route.

anc-chimps and Homo erectus formed the start node and the target node of the DAG. The edges represent the direction of HSV2 transmission and are weighted based on the inverse probability of transmission between the species (nodes). The weights (edge costs) are determined based on two probability models: Infection Prevalence (HSV2-IP) and Infection Transmission (HSV2-IT) based on the temporal and geographic range of the species. More details about the probability models are discussed in the next section. A Conditional Probability Table (CPT) is used to determine the nodal heuristics.

HSV2-Infection Prevalence (HSV2-IP) model

The Infection Prevalence model is a local model of probability, which assumes each transmission route and species to be independent. The model uses a beta distribution to determine the probability of a species transmitting/infected-by HSV2 based on the proximity of the species to the rainforest habitat and the duration (difference between the first and the last appearance datum) of the species. Both the probability of transmission and infection are described using beta distributions. The probability density function for a beta distribution is defined as:

$$PDF : \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} \qquad \text{where} \qquad B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \tag{1}$$

Where, Γ defines a gamma distribution function, α is the time period of existence of the species in 100,000 years and β is the spatial distance of the fossil from the rainforest in kilometres. The probability of transmission through a particular route is determined as the combined probability of the species (nodes) forming the edge. For example, the probability of transmission route from ancestral-chimps ($P(A)$) to *P. boisei* ($P(B)$) is the combined probability of both these species $P(A*B)$. The edge costs were estimated by Monte-Carlo simulations of the combined probabilities of two species (nodes), forming the edge, by sampling from their respective beta distributions.

Stochastic modelling of infection transmission

The Susceptible - Infection - Recovered (SIR) (**Chen2008**) is one of the most widely used models in epidemiology to predict infection transmission. In the stochastic version of the SIR model, the continuous variables are replaced by discrete numbers, and the process rates are replaced by process probabilities. At time 't' the probability that a new susceptible host is infected is modelled as an exponential distribution, which is epidemiologically incorrect for most diseases (Wearing, Rohani, and Keeling, 2005; Bailey, 1975; Sartwell et al., 1950), i.e., the rate of leaving the exposed is independent of the time spent on the host.

Wearing, Rohani, and Keeling (2005) suggested a more realistic distribution of latent and infectious periods, with a stronger central tendency:

# of infectives at time (t)	=	Initial number of infectives who are still infectious at time (t)	+	Those who have acquired the infection in the time interval $[0,t]$ and are still infectious at time (t)

More realistic distributions can be obtained by choosing probability density function of the infectious period, p(t) to be a gamma probability density function (Blythe and Anderson, 1988; Lloyd, 2001).

Infection Transmission (HSV2-IT) model

The proposed Infection Transmission model considers the history of transmission, and the probability of routes (edges) and species (nodes) are dependant on the parent nodes and routes. The model utilises the temporal overlap between hominin species and their geographic proximity to one another to determine the probability of a transmission route, using a gamma distribution. The probability density function of a gamma distribution is given as:

$$PDF : \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} \exp^{-\beta x} \tag{2}$$

in terms of shape α and rate β. The shape parameter α is defined as the ratio of the time period in 1000 years / distance in kilometres and the rate parameter β is defined as the normalised time period Y/x, where x is the time period of the species in 1000 years and Y is the time period of anc-chimps. Monte-Carlo simulations were performed to evaluate the conditional probability of transmission between species, as mutually exclusive events, considering the parent nodes by sampling from the respective probability distributions. For example, the probability of *P. boisei* transmitting HSV2 to *H. erectus* depends on the probabilities of *P. Boisei* being infected by the anc-chimps and/or *H. habilis* and/or *H. rudolfensis*.

Sensitivity analysis

Sensitivity analysis is performed to identify the impact of the probability distribution models in estimating the edge costs and in turn the optimal route of transmission. A variance-based global sensitivity analysis was performed using the Saltelli method, a variation of the Sobol technique, to evaluate the sensitivity indices of the transmission model (**Sobol2001**; **Saltelli2010**). The sensitivity of each input is represented by a numeric value, called the sensitivity index. These indices are used to estimate the influence of individual variables or groups of variables on the model output, by decomposing the variance of the model output into fractions. A 10% variance in the model inputs was assumed. A parametric space of edge costs was generated for the Sobol analysis by sampling 50,000 times from the probabilistic distribution for each edge. This generated a total of 2.3 million analyses for the sensitivity analysis for each transmission model. The first-order, second-order and the total-order indices were measured using a Python script with the SALib package.

Results

The A* optimal graph traversal algorithm was used to evaluate the most probable intermediary host(s) of HSV2 between the ancestors of modern chimpanzees and bonobos (ancestral-chimpanzees or anc-chimps) and the ancestors of modern humans. Directed Acyclic Graphs (DAGs) linking potential HSV2 transmission hominins were developed, with edges representing the possible direction of transmission between different hominins. Two mathematical models: the Infection Prevalence (HSV2-IP) model and the Infection Transmission (HSV2-IT) model, were developed for estimating the probability of HSV2 infection/transmission between species based on their geo-spatial and temporal distributions and their proximity to rainforest. These probability models were used to populate the DAG edge costs and node heuristics. Bayesian network inference of the DAG was also used to determine the most probable intermediary hosts. Both the A* algorithm figs. 1 and 2 and the Bayesian network inference identified *Homo rudolfensis* as the most probable intermediary host that transmitted HSV2 from ancestral-chimpanzees to the ancestors of modern humans. Table 1 shows the probable HSV2 intermediary routes and their rankings based on the infection transmission models and Bayesian network inference.

Fossils from four genera (Ardipithecus, Kenyanthropus, Orrorin and Sahelanthropus) were excluded from the analysis on the basis that there is no fossil evidence that they persisted after 3 MYA. Our range of candidate species can be restricted to those in supplementary Table A.1.

We then used further data on geo-temporal proximity of one species to another to develop mathematical models to quantify the probability of HSV2 infection transmission so as to assess the likelihood that each of these species further transmitted HSV2 to another hominin (see methodology section for further detail of the model). If HSV2 was transmitted to *Homo erectus*,

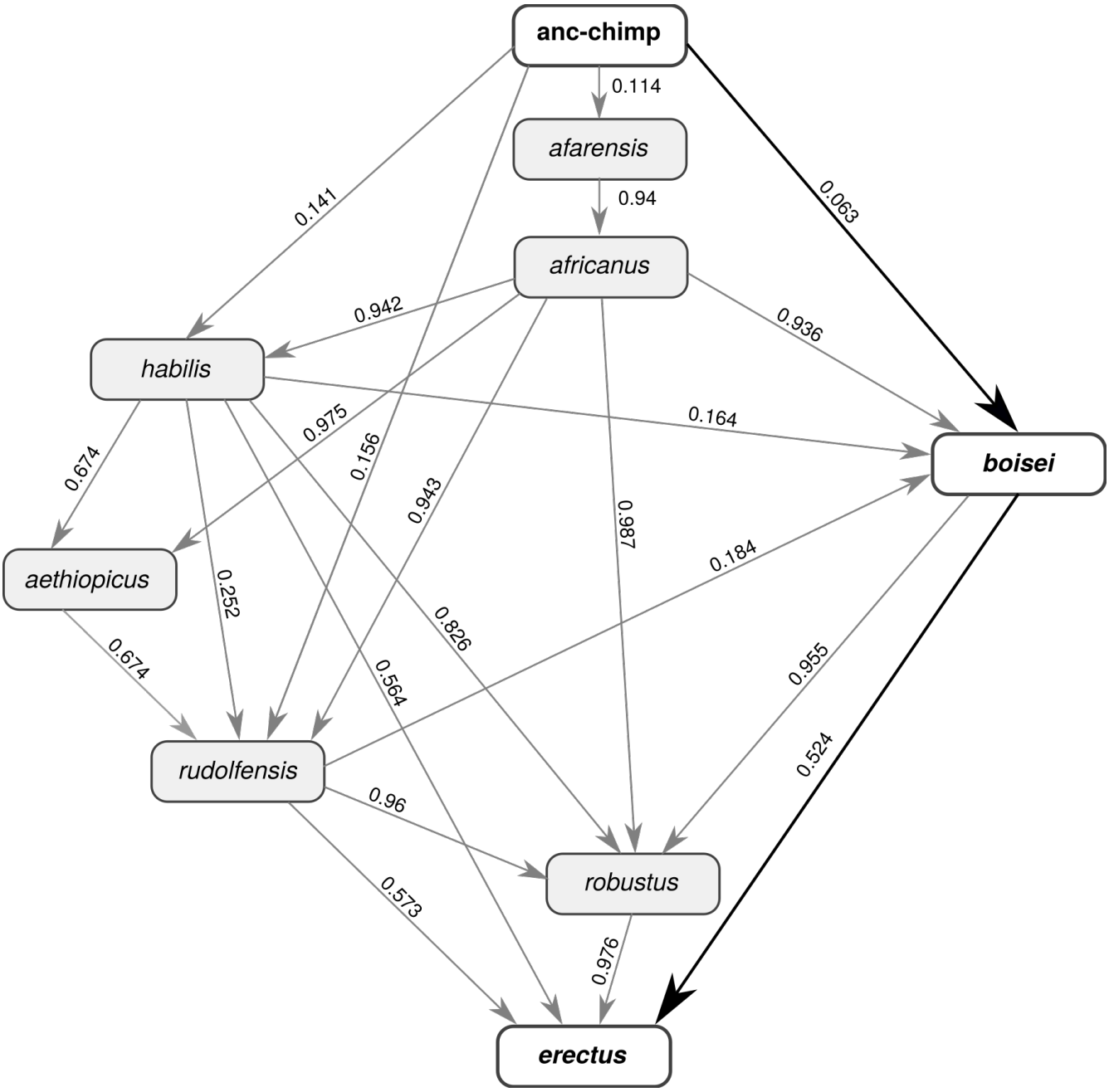


Figure 1. the A* shortest path route(s) for the Infection Prevalence model. The lines with arrows are the possible transmission paths. The values on the lines are the edge costs (inverse probability of transmission). This model predicts that the host switch of HSV2 occurred through the route ancestral-chimps, *P. boisei* and *H. erectus*. The route remained unchanged in the sensitivity analysis.

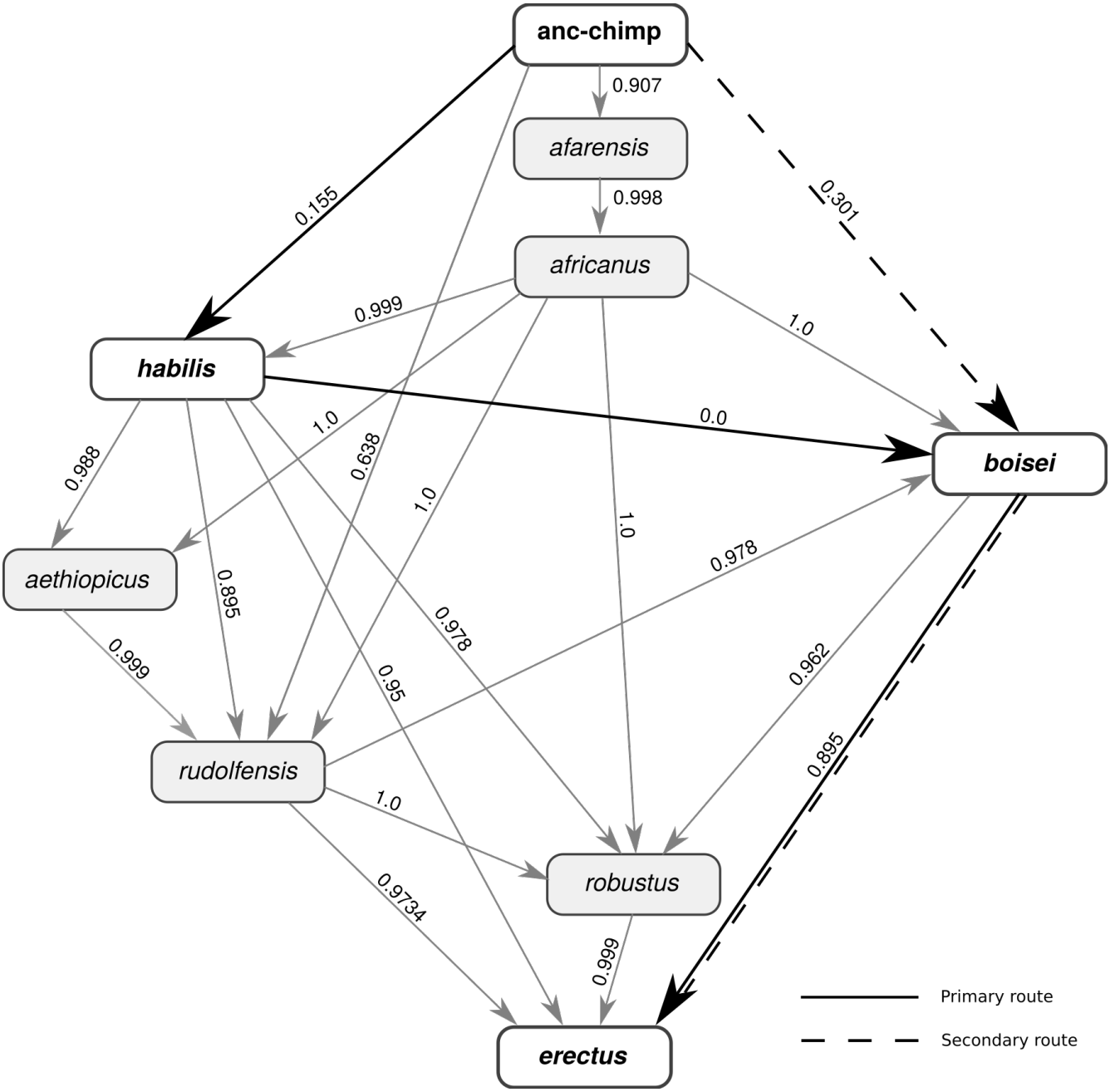
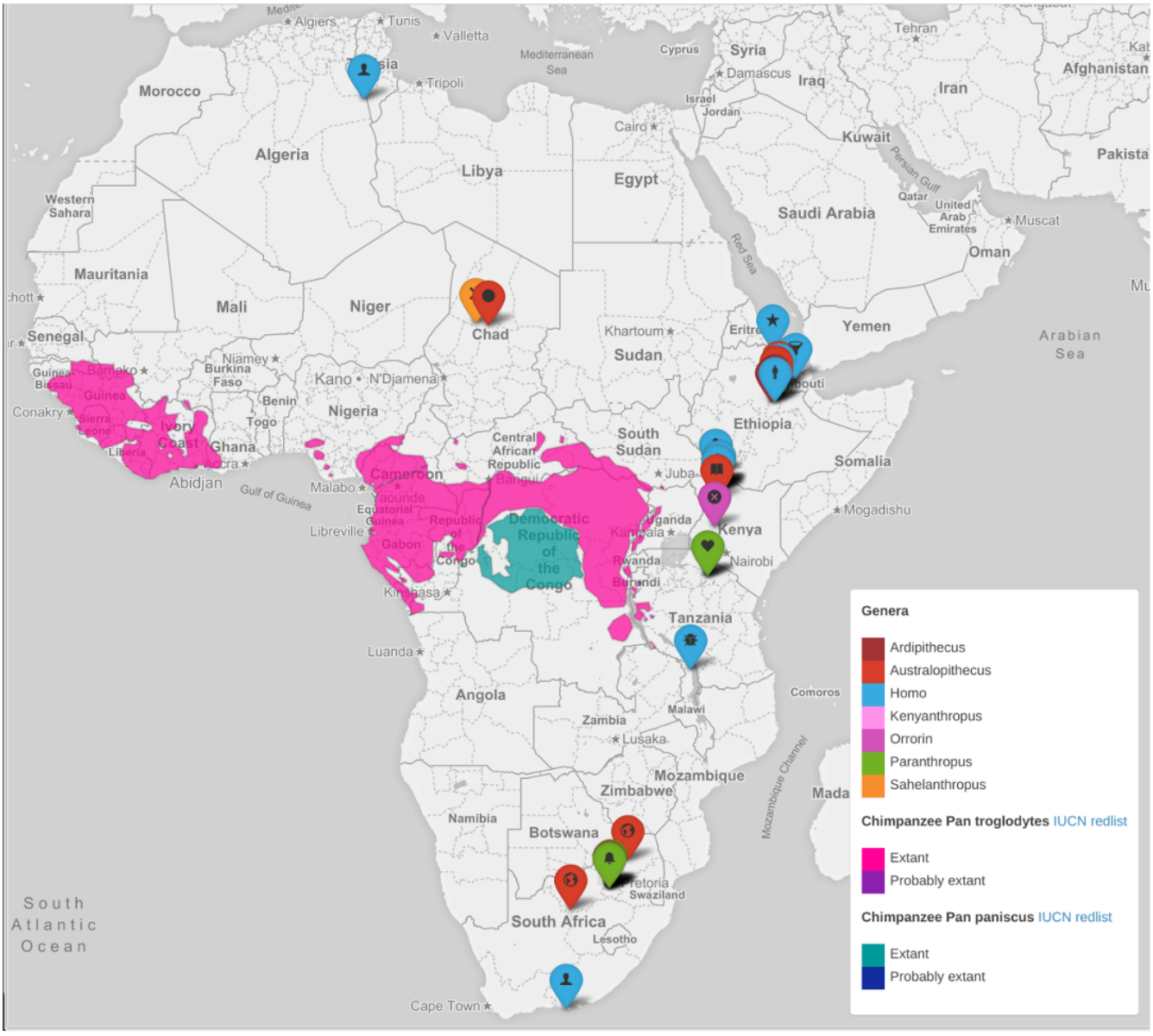
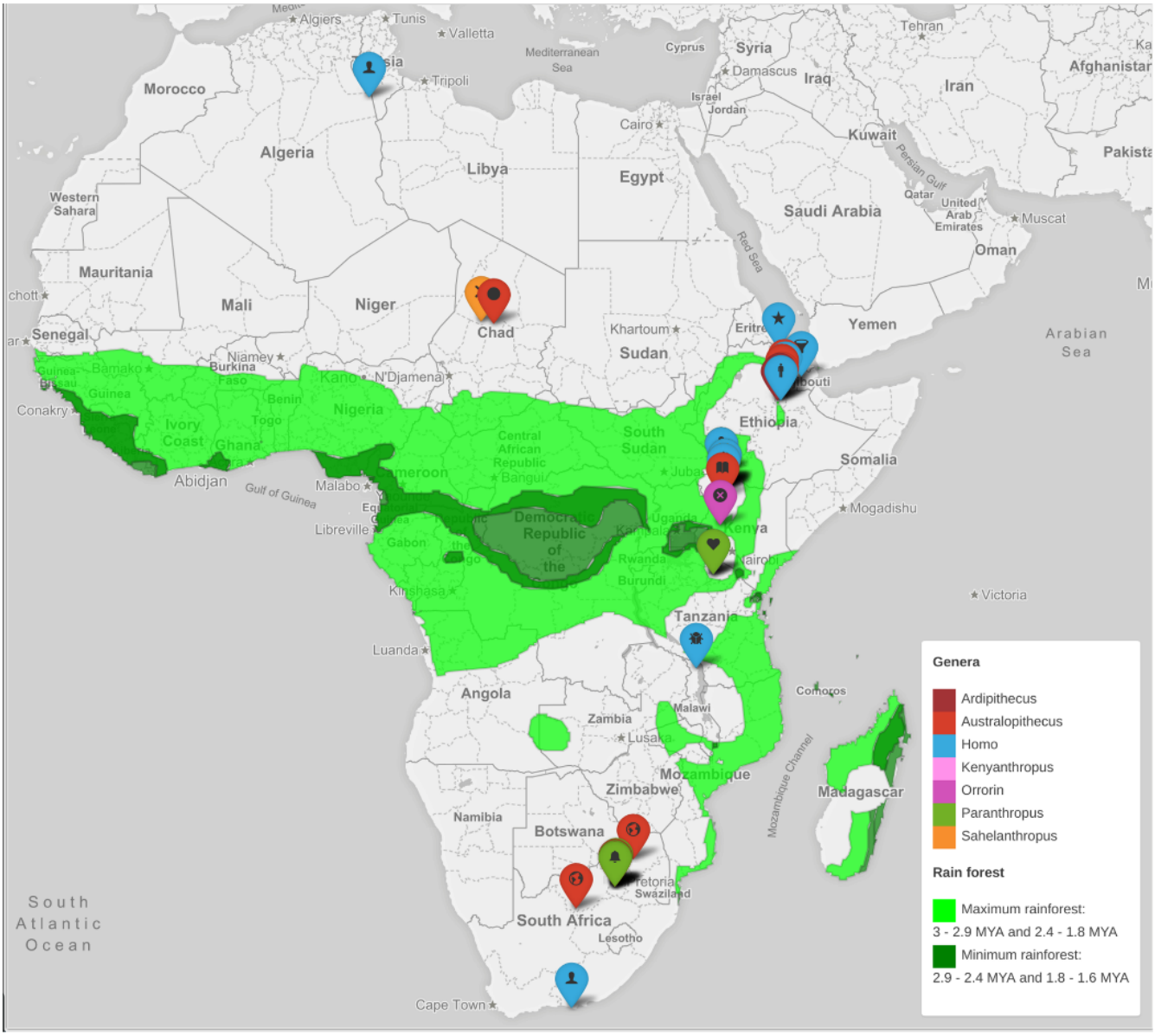


Figure 2. the A* shortest route for the Infection Transmission model. The lines with arrows are the possible transmission paths. The values on the lines are the edge costs (inverse probability of transmission). This model predicts that the host switch of HSV2 occurred through the primary route of ancestral-chimps, *H. Habilis*, *P. boisei* and *H. erectus*. Sensitivity analysis revealed the primary route occurred 60% as opposed to the secondary route of ancestral-chimps, *P. boisei* and *H. erectus* (which occurred the remaining 40%).



(a) Distribution of extant chimpanzee (Pan troglodytes) and bonobo (Pan paniscus) populations [IUCN redlist http://maps.iucnredlist.org/map.html?id=15933, http://maps.iucnredlist.org/map.html?id=15932]. The locations of hominin fossils [supplementary table 1] are shown with markers. The colour of the marker indicates the hominin genus; the symbol represents the species.



(b) Location of hominin fossils relative to ancient minimum and maximum rainforest distributions (Peel, Finlayson, and McMahon, 2007). These figure are available interactively: https://wadhamite.github.io/hsv-mapping/.

Figure 3. Location of hominin fossils, extant chimpanzees, and ancient minimum and maximum rainforest distributions.

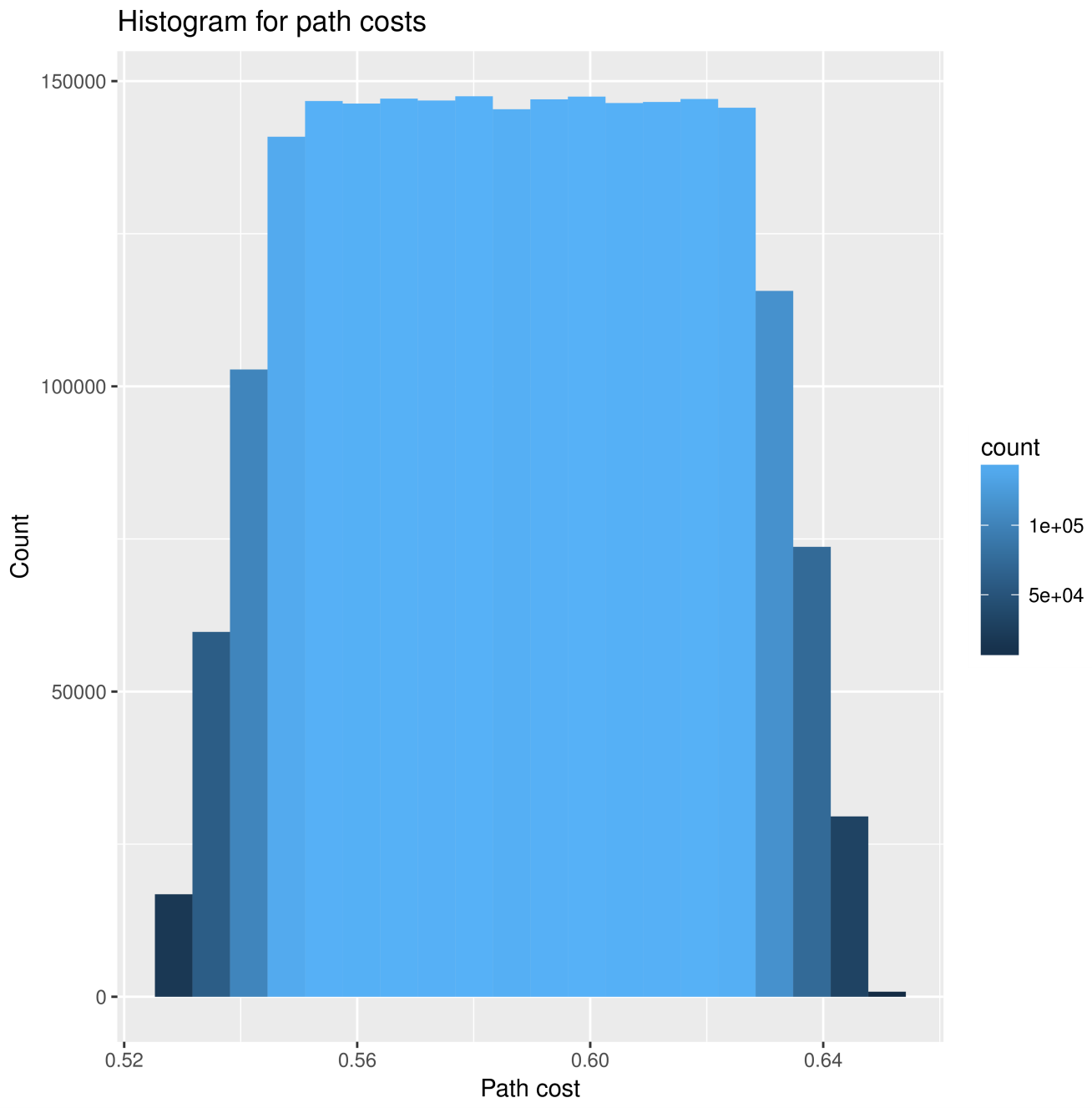


Figure 4. Path costs distribution for Infection Prevalence model. The shortest path is 'anc-chimp ⇒ *P. boisei* ⇒ *H. erectus*' with an average path cost 0.587.

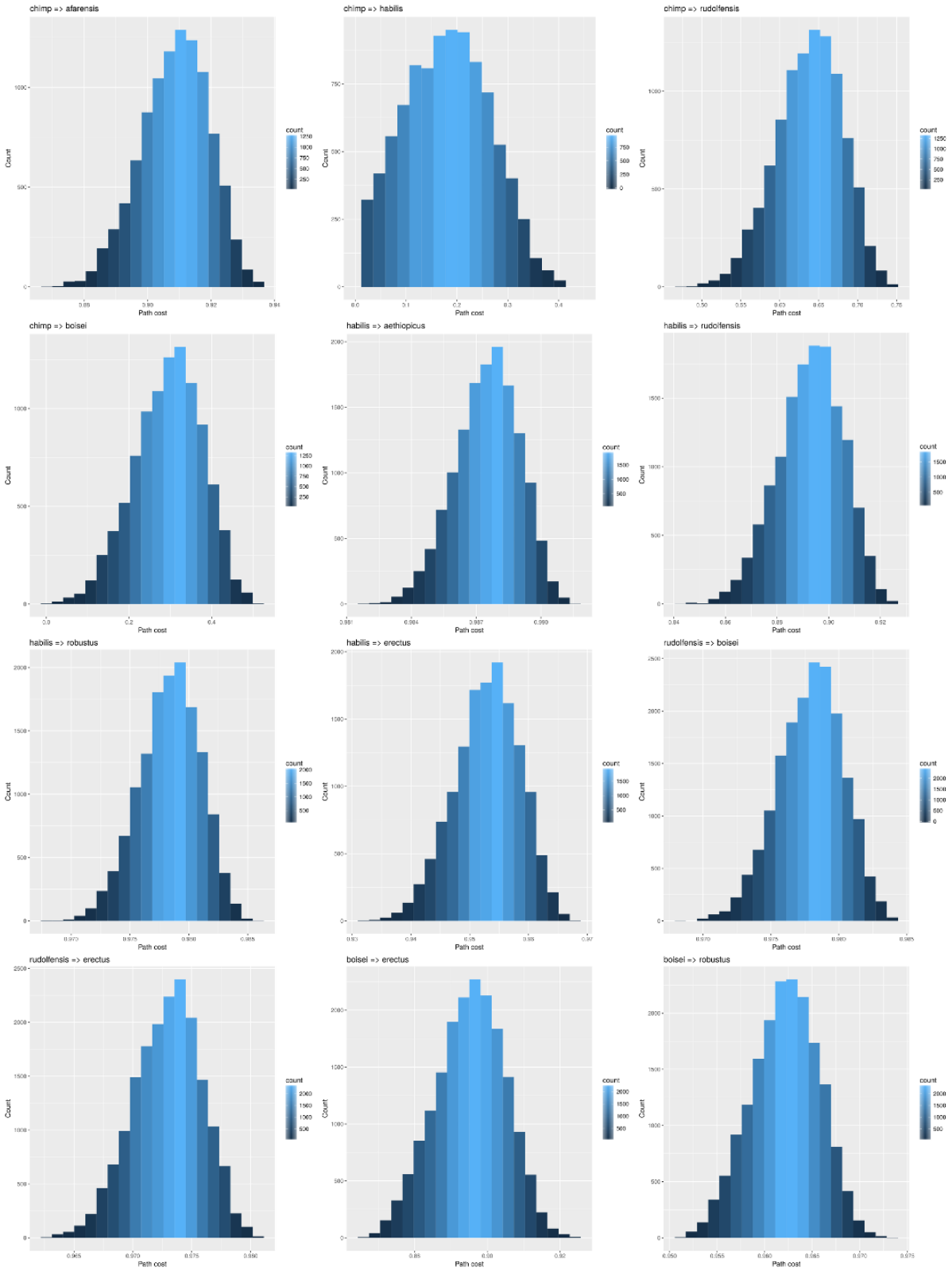


Figure 5. Distribution of selected edge costs for Infection Transmission model. Each histogram shows the distribution of edge costs for the IT model of HSV2 transmission. Colour contours indicate the number of occurrences of a path cost in the sensitivity analysis. The lowest edge costs are seen for 'anc-chimp ⇒ *H. habilis*', 'anc-chimp ⇒ *P. boisei*' and '*P. boisei* ⇒ *H. erectus*'.

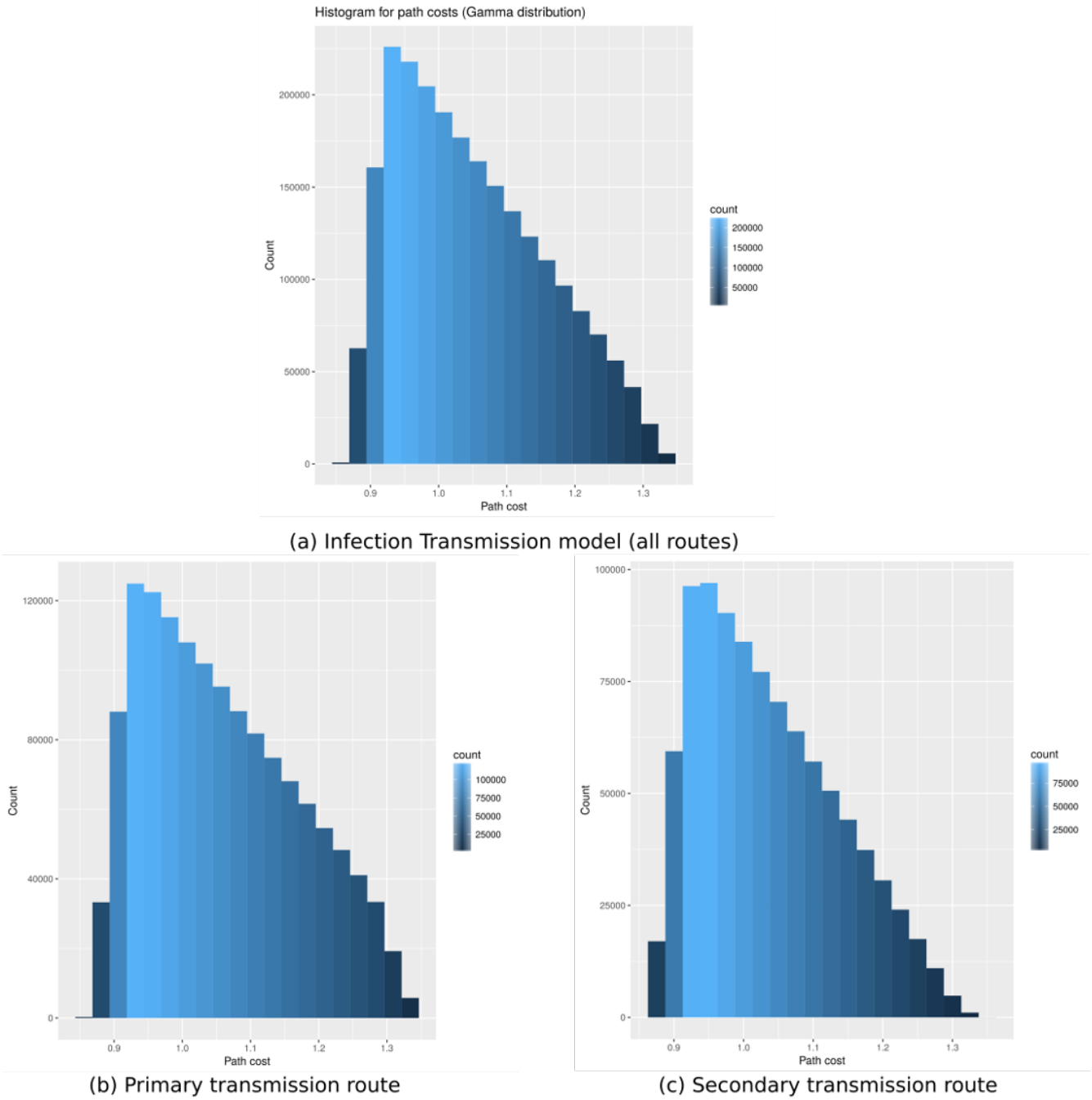


Figure 6. Path costs distribution for Infection Transmission model. Each histogram shows the distribution of path costs for the IT model of HSV2 transmission. Colour contours indicate the number of occurrences of a path cost in the sensitivity analysis. The optimal transmission route is 'anc-chimp ⇒ *H. habilis* ⇒ *P. boisei* ⇒ *H. erectus*'. (a) Total distribution of path costs for HSV2-IT model, (b) distribution path costs for the primary transmission route ('anc-chimp ⇒ *H. habilis* ⇒ *P. boisei* ⇒ *H. erectus*'), and (c) distribution path costs for the secondary transmission route ('anc-chimp ⇒ *P. boisei* ⇒ *H. erectus*').

Table 1. Probable HSV2 transmission routes

Disease Transmission Route	HSV2-IP route ranking (normalised path cost)	HSV2-IT route ranking (normalised path cost)
anc-chimp ⇒ *P. boisei* ⇒ *H. erectus*	**1 (1.0)**	3 (1.14)
anc-chimp ⇒ *H. habilis* ⇒ *P. boisei* ⇒ *H. erectus*	4 (1.13)	**1 (1.0)**
anc-chimp ⇒ *H. habilis* ⇒ *H. erectus*	2 (1.20)	2 (1.05)
anc-chimp ⇒ *H. rudolfensis* ⇒ *H. erectus*	3 (1.24)	4 (1.54)

* The most probable route has a cost of 1. The path costs are normalised to the shortest path.

Table 2. Sobol indices of transmission paths for HSV2-IT model using gamma distribution

Disease Transmission Route	Sobol total-order indices	Sobol first-order indices
anc-chimp ⇒ *P. boisei*	0.640	0.454
anc-chimp ⇒ *H. habilis*	0.519	0.334
P. boisei ⇒ *H. erectus*	0.0262	0.0262

parsimony militates against the need for a further cross-species transmission event to be invoked to explain the infection of modern humans by HSV2. Simple vertical mother-to-child or horizontal (sexual) transmission of the virus through the genus Homo from this point would be sufficient as the ancestor-descendent path from *Homo erectus* to *Homo sapiens* is relatively secure Maslin, Shultz, and Trauth, 2015.

Discussion

Transmission of HSV2

Our analysis suggests that *Homo rudolfensis* was the most likely intermediate host of HSV2 between chimpanzees and the ancestors of *Homo sapiens*. The combined inference analysis of a Bayesian Network, considering the presence of HSV2 in both ancestral-chimpanzees and *Homo erectus*, revealed *A afarensis* (90.33%), *H habilis* (60.85%), *P boisei* (57.3%), and *H rudolfensis* (42.6%) to be the likely intermediary hosts. However, *A afarensis* could only transmit to *A africanus*, which had a mere 6% chance of being infected by HSV2. This narrowed the possibility of transmission through *H habilis*, *P boisei* and *H rudolfensis*, either directly or as intermediary hosts. Infection Prevalence (HSV2-IP) describes the probability of HSV2 infection based on the proximity to the rain forest habitat and the duration (difference between the first and the last appearance datum) using a beta distribution. While, the Infection Transmission (HSV2-IT) model assigns probability values (edge costs) to the DAG utilising the temporal overlap between hominin species and their geographic proximity to one another. The A* algorithm estimated the minimum path costs of 0.252 and 0.87 for the transmission route through *Homo rudolfensis* for both HSV2-IP and HSV2-IT based DAGs, respectively. Bayesian network inference analysis estimated the highest probability of HSV2 transmission (11.6%) to the ancestors of *Homo sapiens* through *rudolfensis*. Transmission through *boisei* and *habilis* was estimated to be 9.3% and 7.8% probable. A* optimal path traversals also predicted *boisei* and *habilis* to be the second (5% less likely than *rudolfensis*) and the third (7% less likely than *rudolfensis*) most probable routes. These algorithms predict a single intermediate hominin to be more likely than multiple-levels of transmissions. The A* algorithm and the Bayesian network inference have clearly demonstrated that *Homo rudolfensis* is the most likely intermediate hominin host.

Tropical refugia during hot dry periods may have driven chimpanzees into higher concentrations in certain areas, driving them into contact and competition with *Homo rudolfensis*. Violent confrontation or predation in combination with hunting and butchery practices would have provided a viable route of transmission for HSV2. While no stone tools have to date been found in the same archaeological strata as *Homo rudolfensis* fossils it is a logical assumption that they would have been extensively used by this species Ungar, Grine, and Teaford, 2006; Torre et al., 2003 as Oldowan or Mode 1 stones tools are widely distributed in the same spatial-temporal localities as *Homo rudolfensis* fossils Stern et al., 1993; Leakey et al., 2012. *Paranthropus aethiopicus*, *P. boisei*, and *P. robustus* are associated with the Olduwan stone tool complex De Heinzelin et al.,

1999, and *P boisei* explicitly with butchery Domínguez-Rodrigo, Pickering, Baquedano, et al., 2013 lending support to the hypothesis that bushmeat hunting and butchery may have led to the initial transmission of HSV2 to the hominins. *A afarensis*, although a tool user, can be excluded both on statistical grounds and by the lack of evidence for hunting and butchery within this species which instead used tools to scavenge the carcasses left behind by savannah predators Domínguez-Rodrigo, Pickering, and Bunn, 2010; Domalain, Bertin, and Daver, 2016.

Although ChHV1 causes outbreaks of oral and pharyngeal lesions in chimpanzees in a manner similar to HSV1, in hominins contracting the virus which was to become HSV2, the oral niche was already occupied by HSV1. This may have protected the hominin first infected with HSV2: pre-existing infection with HSV1 reduces the likelihood that subsequent infection with HSV2 will be symptomatic Langenberg et al., 1999, and also reduces the risk of HSV2 meningitis Aurelius et al., 2012. HSV2 may have been forced to adapt to a different mucosal niche in order to reduce competition from the co-evolved, native HSV1.

We suggest that the mode of transmission of HSV2 into hominins was most likely through hunting injuries (eg chimpanzee bites or cuts sustained during meat processing), although onwards transmission into the ancestors of *Homo sapiens* could have been sexual (horizontal) or a result of hunting injuries (vertical). There are many reports of transmission of B virus (*Cercopithecine herpesvirus 1*), the cercopith homolog of HSV1 and ChHV1, to humans, where disease ranges from mild to fatal. Transmission has occurred from bites and scratches, needle sticks and even scratches from cage bars that are contaminated with B virus-positive bodily fluids Huff and Barry, 2003. Onwards transmission between humans has been reported to occur Centers for Disease Control (CDC), 1987. Human herpesvirus 1 can infect other primates, from gorillas Gilardi et al., 2014 to owl monkeys Melendez et al., 1969, typically causing fatal disease in species more distantly related to Homo sapiens, while causing oral lesions and milder disease in great apes such as Gorilla *beringei graueri* Gilardi et al., 2014. We therefore infer that herpes simplex-like viruses spread relatively easily between individuals even across species barriers, increasing the chances of transmission between hominins and other primates from close contact such as hunting, butchery, inter-personal violence or sexual contact. Evidence for close hominin-hominid contact is also found in other 'heirloom' human pathogens Houldcroft et al., 2017.

The high prevalence of HSV2 in central and eastern Africa (see Figure A.1) and interactive maps at https://wadhamite.github.io/hsv-mapping) is consistent with the limited genetic data available from African HSV2 isolates. A study from Burrel and colleagues Burrel, Desire, et al., 2015 showed that HSV2 can be divided into African and worldwide lineages on the basis of diversity in gene UL30. Furthermore, they found evidence of gene flow from HSV1 into HSV2, and speculate that the flow of HSV1 loci into the worldwide HSV2 lineage may have helped this lineage of HSV2 to further adapt to human hosts, and so spread more successfully around the world from around 41kya Burrel, Boutolleau, et al., 2017. Two recent studies have significantly increased the number of whole HSV2 genomes available for analysis, contributing to our knowledge of HSV2 diversity Szpara et al., 2014; Kolb et al., 2015, however precise geographic origins of each sample are not available. More whole HSV2 genomes from central and east Africa, with clear country of origin data, are needed to enrich this picture, as our model predicts that individuals from east Africa are also likely to carry ancient HSV2 lineages.

The time-depth of ancient DNA analysis is dually limited by technology and preservation of DNA. Similarly the archaeological and fossil records suffer from differential rates of preservation and gaps that can never be filled because the material has simply not survived. Our analysis has allowed the reconstruction of hominin/human-disease interaction well beyond the horizon of ancient DNA and at a level that is invisible to the fossil and archaeological records. Demonstrating the potential for using modern disease genetics to better understand the evolutionary interaction between humans and disease in deep time.

Data presentation

The distribution maps were created with custom JavaScripts codes using Leaflet.js library [http://leafletjs.com/] and interactive maps can be found at https://wadhamite.github.io/hsv-mapping

Author contributions

SJU and CH conceived the study and contributed data. SJU, KK and CH performed the analyses. SJU, KK and CH wrote the paper. All authors approved the publication of the manuscript.

Acknowledgements

SJU was funded by Oxford Brookes University. KK and CH were funded by the University of Cambridge. KK is a college research associate at King's College, Cambridge. The authors would like to thank C Ruis (University College London) for helpful discussion.

References

Anton, Susan C et al. (2016). "Morphological variation in Homo erectus and the origins of developmental plasticity". In: *Philosophical Transactions of the Royal Society, B* 371, p. 20150236. ISSN: 1471-2970. DOI: 10.1098/rstb.2015.0236.

Aurelius, E. et al. (2012). "Long-term valacyclovir suppressive treatment after herpes simplex virus type 2 meningitis: A double-blind, randomized controlled trial". In: *Clinical Infectious Diseases* 54.9, pp. 1304–1313. ISSN: 10584838. DOI: 10.1093/cid/cis031.

Bailey, Norman T. J. (1975). *The mathematical theory of infectious diseases and its applications. 2nd edition*. Vol. 413, p. 413. ISBN: 0852642318.

Blythe, S P and R M Anderson (1988). "Distributed incubation and infectious periods in models of the transmission dynamics of the human immunodeficiency virus (HIV)." In: *IMA journal of mathematics applied in medicine and biology* 5.1, pp. 1–19. ISSN: 0265-0746.

Burrel, Sonia, David Boutolleau, et al. (2017). "Ancient Recombination Events between Human Herpes Simplex Viruses". In: *Molecular Biology and Evolution* 25, pp. 1910–1920. ISSN: 0737-4038. DOI: 10.1093/molbev/msx113.

Burrel, Sonia, Nathalie Desire, et al. (2015). "Genetic Diversity within Alphaherpesviruses: Characterization of a Novel Variant of Herpes Simplex Virus 2." In: *Journal of virology* 89.24, pp. 12273–12283. ISSN: 1098-5514 (Electronic). DOI: 10.1128/JVI.01959-15.

Centers for Disease Control (CDC) (1987). "B-virus infection in humans–Pensacola, Florida." In: *MMWR. Morbidity and mortality weekly report* 36.19, pp. 289–90, 295–6. ISSN: 0149-2195.

De Heinzelin, Jean et al. (1999). "Environment and Behavior of 2.5-Million-Year-Old Bouri Hominids". In: *Science* 284111185.23, pp. 625–629.

Domalain, Mathieu, Anne Bertin, and Guillaume Daver (2016). "Was Australopithecus afarensis able to make the Lomekwian stone tools? Towards a realistic biomechanical simulation of hand force capability in fossil hominins and new insights on the role of the fifth digit". In: *Comptes Rendus Palevol*. ISSN: 16310683. DOI: 10.1016/j.crpv.2016.09.003.

Domínguez-Rodrigo, Manuel, Travis Rayne Pickering, Enrique Baquedano, et al. (2013). "First Partial Skeleton of a 1.34-Million-Year-Old Paranthropus boisei from Bed II, Olduvai Gorge, Tanzania". In: *PLoS ONE* 8.12. Ed. by Darren Curnoe, e80347. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0080347.

Domínguez-Rodrigo, Manuel, Travis Rayne Pickering, and Henry T Bunn (2010). "Configurational approach to identifying the earliest hominin butchers." In: *Proceedings of the National Academy of Sciences of the United States of America* 107.49, pp. 20929–34. ISSN: 1091-6490. DOI: 10.1073/pnas.1013711107.

Farine, Damien R and Ariana Strandburg-Peshkin (2015). "Estimating uncertainty and reliability of social network data using Bayesian inference." In: *Royal Society open science* 2.9, p. 150367. DOI: 10.1098/rsos.150367.

Freeman, Esther E et al. (2006). "Herpes simplex virus 2 infection increases HIV acquisition in men and women: systematic review and meta-analysis of longitudinal studies." In: *AIDS (London, England)* 20.1, pp. 73–83. ISSN: 0269-9370. DOI: 10.1097/01.AIDS.0000198081.09337.A7.

Gilardi, Kirsten et al. (2014). "Human herpes simplex virus type 1 in confiscated gorilla". In: *Emerging Infectious Diseases* 20.11, pp. 1883–1886. ISSN: 1080-6059. DOI: 10.3201/eid2011.140075.

Hart, Peter, Nils Nilsson, and Bertram Raphael (1968). "A Formal Basis for the Heuristic Determination of Minimum Cost Paths". In: *IEEE Transactions on Systems Science and Cybernetics* 4.2, pp. 100–107. ISSN: 0536-1567. DOI: 10.1109/TSSC.1968.300136.

Houldcroft, Charlotte J. et al. (2017). "Germs on a journey: what pathogens can tell us about population movements and human evolution". In: In press.

Huff, Jennifer L. and Peter A. Barry (2003). *B-virus (Cercopithecine herpesvirus 1) infection in humans and macaques: Potential for zoonotic disease*. DOI: 10.3201/eid0902.020272.

Kolb, Aaron W et al. (2015). "Genomic, Phylogenetic, and Recombinational Characterization of Herpes Simplex Virus 2 Strains." In: *Journal of virology* 89.12, pp. 6427–6434. ISSN: 1098-5514. DOI: 10.1128/JVI.00416-15.

Korb, Kevin B. and Ann E. Nicholson (2003). *Bayesian Artificial Intelligence*, pp. 29–54. ISBN: 1-58488-387-1. DOI: doi:10.1201/b10391-4. arXiv: arXiv:1011.1669v3.

Langenberg, Andria G.M. et al. (1999). "A Prospective Study of New Infections with Herpes Simplex Virus Type 1 and Type 2". In: *New England Journal of Medicine* 341.19, pp. 1432–1438. ISSN: 0028-4793. DOI: 10.1056/NEJM199911043411904.

Leakey, Meave G. et al. (2012). "New fossils from Koobi Fora in northern Kenya confirm taxonomic diversity in early Homo". In: *Nature* 488.7410, pp. 201–204. ISSN: 0028-0836. DOI: 10.1038/nature11322. arXiv: ;homohabilis;KoobiFora.

Lloyd, Alun L (2001). “Destabilization of epidemic models with the inclusion of realistic distributions of infectious periods”. In: *Proceedings of the Royal Society of London B.* 268, pp. 985–993. ISSN: 0962-8452. DOI: 10.1098/rspb.2001.1599.

Looker, Katharine J. et al. (2015). “Global estimates of prevalent and incident herpes simplex virus type 2 infections in 2012”. In: *PLoS ONE* 10.1. ISSN: 19326203. DOI: 10.1371/journal.pone.0114989.

Maslin, Mark a, Susanne Shultz, and Martin H Trauth (2015). “A synthesis of the theories and concepts of early human evolution.” In: *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 370, pp. 1–12. ISSN: 1471-2970. DOI: 10.1098/rstb.2014.0064.

Melendez, L V et al. (1969). “Natural herpes simplex infection in the owl monkey (Aotus trivirgatus).” In: *Laboratory animal care* 19.1, pp. 38–45. ISSN: 0094-5331.

Mirazón Lahr, Marta et al. (2016). “The shaping of human diversity: filters, boundaries and transitions.” In: *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 371.1698, pp. 62–108. ISSN: 1471-2970. DOI: 10.1098/rstb.2015.0241.

Murphy, Kevin P. (2012). “Machine Learning: A Probabilistic Perspective”. In: *MIT Press*, p. 25. ISSN: 0036-8733. DOI: 10.1007/978-3-642-21004-4_10. arXiv: 0-387-31073-8.

Oates, J.F. et al. (2008). *Pan troglodytes (Chimpanzee, Common Chimpanzee, Robust Chimpanzee)*.

Peel, B.L., B. L. Finlayson, and T. a. McMahon (2007). “Updated world map of the Köppen-Geiger climate classification”. In: *Hydrology and Earth System Sciences* 11.2001, pp. 1633–1644. ISSN: 09412948. DOI: 10.5194/hess-11-1633-2007.

Poole, David L and Alan K Mackworth (2010). *Artificial Intelligence: foundations of computational agents*. Cambridge University Press.

Sartwell, Philip E et al. (1950). “The distribution of incubation periods of infectious disease.” In: *American Journal of Hygiene* 51, pp. 310–318.

Stern, Nicola et al. (1993). “The structure of the lower pleistocene archaeological record: a case study from the Koobi fora formation [and comments and reply]”. In: *Current Anthropology* 34.3, pp. 201–225.

Szpara, Moriah L. et al. (2014). “Evolution and Diversity in Human Herpes Simplex Virus Genomes”. In: *Journal of Virology* 88.2, pp. 1209–1227. ISSN: 0022-538X, 1098-5514. DOI: 10.1128/JVI.01987-13.

Tang, J W et al. (2003). “Brain stem encephalitis caused by primary herpes simplex 2 infection in a young woman.” In: *Journal of neurology, neurosurgery, and psychiatry* 74.9, pp. 1323–5. ISSN: 0022-3050. DOI: 10.1136/JNNP.74.9.1323.

Torre, Ignacio de la et al. (2003). “The Oldowan industry of Peninj and its bearing on the reconstruction of the technological skills of LowerPleistocene hominids”. In: *Journal of Human Evolution* 44.2, pp. 203–224. ISSN: 00472484. DOI: 10.1016/S0047-2484(02)00206-3.

Ungar, Peter S, Frederick E Grine, and Mark F Teaford (2006). “Diet in Early Homo: A Review of the Evidence and a New Model of Adaptive Versatility”. In: *Annual Review of Anthropology* 35, pp. 209–228. ISSN: 00846570. DOI: 10.2307/25064922.

Wearing, Helen J, Pejman Rohani, and Matt J Keeling (2005). “Appropriate models for the management of infectious diseases.” In: *PLoS medicine* 2.7, e174. ISSN: 1549-1676. DOI: 10.1371/journal.pmed.0020174.

Wertheim, Joel O. et al. (2014). “Evolutionary origins of human herpes simplex viruses 1 and 2”. In: *Molecular Biology and Evolution* 31.9, pp. 2356–2364. ISSN: 15371719. DOI: 10.1093/molbev/msu185.

Supplementary material

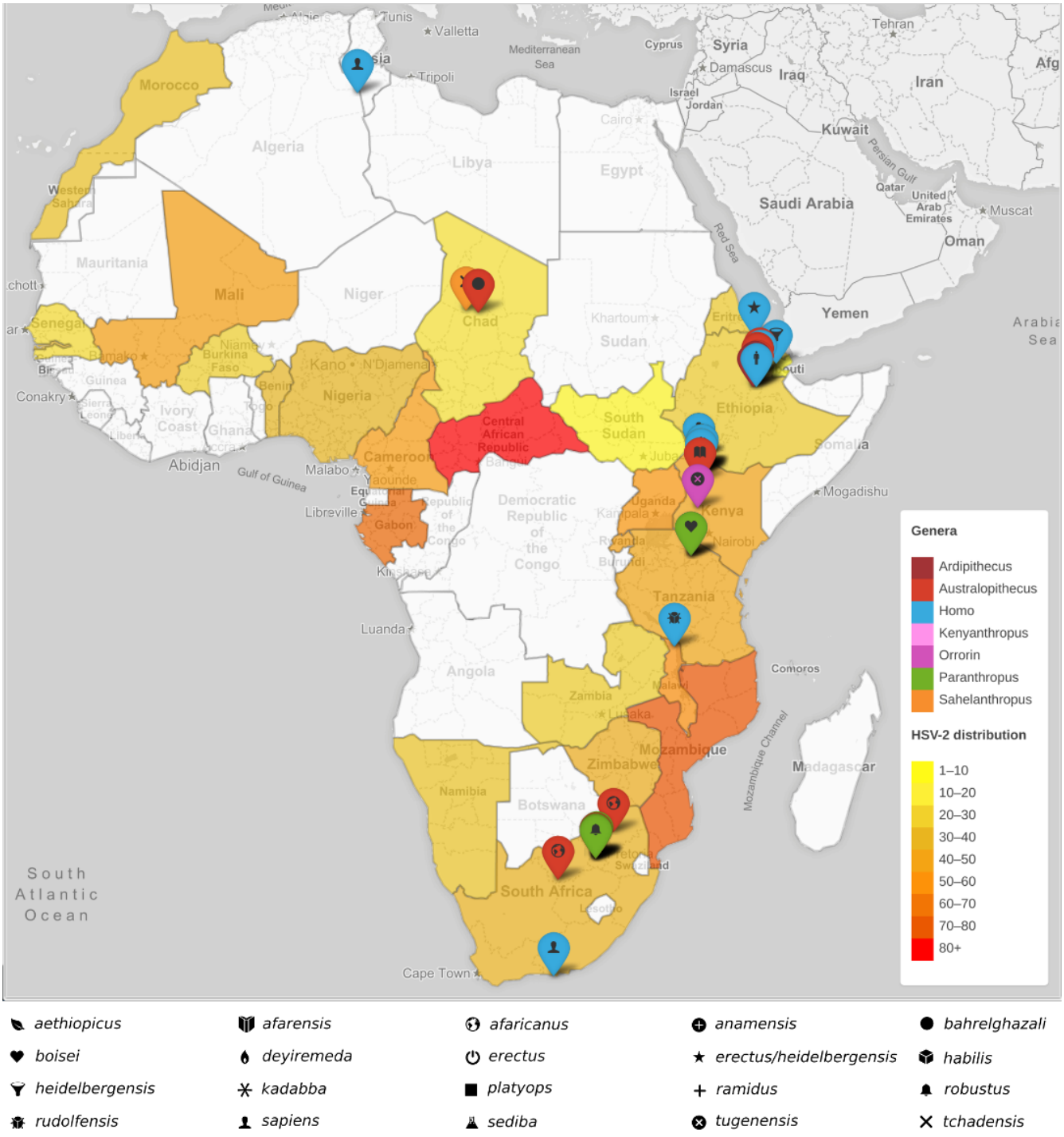


Figure A.1. Prevalence of HSV2 in Africa

Table A.1. Fossil hominins with the potential to acquire - HSV2 from ancestral chimpanzee

Species	Type specimen	Locations (type specimen)	Age (000s BP)	Publication (suggesting a new species)
Australopithecus afarensis	LH4 (mandible)	East Africa (Laetoli, Tanzania)	3900 - 2800	Johanson DC, White TD and Coppens Y 1978 A new species of the genus Australopithecus (Primates: Hominidae) from the Pliocene of eastern Africa. *Kirtlandia 28: 2-14.*
Australopithecus africanus	Taung 1 (cranium)	Southern Africa (Taung, SA)	2800 - 2300	Dart RA 1925 Australopithecus africanus: The man-ape of South Africa. *Nature 115: 195-199.*
Australopithecus gahri	BOU-VP-12/130 (cranium)	East Africa (Bouri, Ethiopia)	2500	Asfaw B, White T, Lovejoy O, Latimer B,. Simpson S and Suwa G 1999 Australopithecus garhi: A new species of early hominid from Ethiopia. *Science 284: 629-635.*
Paranthropus aethiopicus	Omo 18-1967-18 (mandible)	East Africa (Omo, Kenya)	2700 - 2300	Arambourg C and Coppens Y 1967 Sur la découverte, dans le Pléistocène inférieur de la vallée de l'Omo (Éthiopie), d'une mandibule d'australopithécien. *Comptes Rendus de l'Académie des Sciences, Paris, series D, 265: 589-590.*
Paranthropus robustus (including *Paranthropus crassidens*, SK 48)	TM 1517 (various)	Southern Africa (Kromdraai, SA)	1800 - 1000	Broom R 1938 The Pleistocene anthropoid apes of South Africa. *Nature 142: 377-379.*
Paranthropus boisei	OH 5 (cranium)	East Africa (Olduvai Gorge, Tanzania)	1750 - 1200	Leakey LSB 1959 A new fossil skull from Olduvai. *Nature 184: 491-493.*

Continued on next page

Species	Type specimen	Locations (type specimen)	Age (000s BP)	Publication (suggesting a new species)
Homo habilis (Australopithecus?)	OH 7 (mandible)	East Africa (Olduvai Gorge, Tanzania)	2500 - 1600	Leakey LSB, Tobias PV and Napier JR 1964 A new species of the genus Homo from Olduvai Gorge, Tanzania. *Nature 202: 308-312.*
Homo rudolfensis (Kenyanthropus?)	KNM-ER 1470 (cranium)	East Africa (Koobi Fora, Kenya)	2400 - 1800	Leakey RE 1973 Evidence for an advanced Plio-Pleistocene hominid from east Rudolf, Kenya. *Nature 242: 447-450.*
Homo erectus	KNM-ER 992 (mandible)	Africa (Koobi Fora, Kenya)	2000 - 600	Groves CP and Mazák V 1975 An approach to the taxonomy of the Hominidae: Gracile Villafranchian hominids of Africa. *Casopis pro Mineralogii Geologii 20: 225-247.*