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]Department of Archaeology, 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 path 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 (Whitley, Kimberlin, and Prober, 2007). 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 (Kim et al., 2006). 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 (Bertke, Patel, and Krause, 2007). HSV1 and 2 also differ in the length of the glycoprotein G (US4) open reading frame, which may play a role in tropism (Baines and Pellett, 2007).

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 (Koelle et al., 2017; Burrel, Boutolleau, et al., 2017), and has patterns of genetic divergence consistent with having diverged with human populations as they spread out of Africa (Koelle et al., 2017). 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

The process of identifying the most probable intermediary species that transmitted proto-HSV2 (henceforth HSV2) from anc-chimps to modern humans involves four steps: (a) data acquisition, which involves collecting information on the prevalence of HSV2 in Africa, fossil records of potential intermediary hominins, their geographical range and time-period, and the extent of tropical rainforest data, (b) network generation, where a network of possible transmission paths of HSV2 from anc-chimps to humans through different hominins is created, (c) probabilistic network analysis, which involves assigning weights to different transmission paths on the graph network based on a certain probability distribution, and (d) optimal path traversal, which evaluates the shortest path from anc-chimps to modern humans based on the probabilistic weights assigned to different paths.

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. 3. Only one ancestral chimpanzee fossil, dated to c. 500 KYA, is currently known (McBrearty and Jablonski, 2005). 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 than that of modern Pan. We can use the GIS data on the range of modern Pan troglodytes and Pan paniscus, provided by the IUCN Red List (Oates et al. 2008), shown in Figure 1, to make some inferences about the range of anc-chimps, but a measure of paleo-tropical rainforest range is likely to be a better proxy for the range of anc-chimps Elton, 2008. Therefore, we have used the paleo-tropical rainforest range during the period of 1.4 - 3 MYA from the 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 (Myers Thompson, 2003). This is shown in fig. 3.

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 (see supplementary fig. A.1).

Network analysis

To establish the most probable transmission path 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 initial threshold on the distance 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 (Robert A. Foley, 1978; Grant, Chapman, and Richardson, 1992). 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 (see supplementary tables A.2 and A.3). 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 paths 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 figs. 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 which allows us to represent and reason about an uncertain domain. In a Bayesian network, the species (nodes) are variables, and the transmission paths (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 (hominins which could have transmitted HSV2), given values for some evidence (or observation) nodes (anc-chimps and *H. erectus*). The presence of HSV2 in the common ancestors of modern humans, and anc-chimps being the source species of the virus, provides the evidence of the presence of HSV2 in anc-chimps and *H. erectus*, and these species form the observation nodes in the belief network. The process of conditioning involves computing the probability of different nodes (hominins) for the acquisition and transmission of HSV2 through all possible transmission paths in the network, such that the observations of HSV2 in anc-chimps (origin node) and *H. erectus* (target node) are true. 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).

For example, to evaluate the Conditional Probability Table of node *Paranthropus boisei* we identified its parent nodes on the network graph as the following species: anc-chimps, *A. africanus*, *H. habilis*, and *H. rudolfensis*. To calculate the probability of transmission from anc-chimps to *P. boisei* we assume a beta distribution. The parameter α, in the beta distribution, represents the time-period overlap between the two species, in this case 700 Kyr and β represents the distance between the two species (35 kilometers between the fossils of *P. boisei* and the extent of anc-chimps range). Based on this beta distribution, Monte Carlo simulations are run to sample from the distribution. This yields a probability of *P. boisei* acquiring HSV2 from anc-chimps of 0.954. Similarly, the probability of other hominins infecting *P. boisei* are evaluated for *A. africanus*, *H. habilis* and *H. rudolfensis* as 0.096, 0.996 and 0.34, respectively. A Conditional Probability Table is constructed for the node *P. Boisei* (see supplementary table A.4).

Conditional Probability Tables are evaluated for each node on the graph network. The Bayesian inference analysis was performed on the network using the AI space (`http://aispace.org/bayes/`) decision network tool (Poole and Mackworth, 2010). The "*Combined Inference*" approach is adopted to evaluate the probability of intermediary hosts transmitting HSV2, by conditioning the anc-chimp and *Homo erectus* nodes for the presence of HSV2 (Korb and Ann E. Nicholson, 2003), i.e. they are treated as being infected by HSV2.

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), which allows us to identify the most probable HSV2 transmission path. 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 cost $f(n) = g(n) + h(n)$, 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 path 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* pathfinding algorithm on the DAGs to identify the most probable (optimal) HSV2 transmission path (see `https://gist.github.com/wadhamite/e8991aff7d8d3a48d5bc846c747de6d8`).

Two species, 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 path and species to be independent: the probability of a species being infected by HSV2 is affected by how long the species persisted and where the species was located, not the probability that other species were infected.

The model uses a beta distribution to determine the probability of a species transmitting / being 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 path is determined as the combined probability of the species (nodes) forming the edge. For example, the probability of transmission path from anc-chimps ($P(A)$) to *P. boisei* ($P(B)$) is the combined probability of both these species ($P(A*B)$). The probability of transmission from anc-chimps is evaluated using a beta distribution with as 6000 (corresponding to 6 Million years of existence), and as 100 (distance to the rainforest in km). The beta density functions of HSV2 transmission / infection for anc-chimps and *P. boisei* are shown in supplementary fig. A.2. The probability of *P. boisei* being infected also follows the same beta-distribution with the parameters as 700 (corresponding to 700,000 years of existence), and as 35 (distance to the rainforest in km). 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. A typical run of a Monte-Carlo simulation is shown as a blue line in supplementary fig. A.2, where the probability of transmission by anc-chimps ($P(A) = 0.97$) and the probability of *P. boisei* being infected ($P(B) = 0.9375$), evaluates the combined probability of the transmission path, for that run, as 0.909. Monte-Carlo simulation results are averaged over 1000 simulation runs. Monte-Carlo simulations evaluating combined probabilities $P(A*B)$ give an edge cost of 0.937 for the edge (path) connecting anc-chimps to *P. boisei*, i.e., the probability of HSV2 being transmitted from anc-chimps to *P. boisei* is 0.937.

Stochastic modelling of infection transmission

Infection transmission in epidemiology can be predicted using mathematical models: a popular approach is the Susceptible - Infection - Recovered (SIR) model (Chen et al., 2008). 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 paths (edges) and species (nodes) are dependent on the parent nodes and paths. The probability of a parent node being infected with HSV2 affects the probability of a child node being infected with HSV2 in addition to the proximity and time period overlap between the parent and child nodes.

The model utilises the temporal overlap between hominin species and their geographic proximity to one another to determine the probability of a transmission path, 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. The probability density distribution of transmission of HSV2 from anc-chimps is shown in supplementary fig. A.3. 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*. The probability of transmission follows the gamma distribution. The gamma distribution for anc-chimp is defined with a shape parameter α of 6, and a rate parameter β of 10. Similarly, the gamma probability distribution is evaluated for other parent nodes of *P. boisei* (i.e., *H. habilis* and/or *H. rudolfensis*) based on their time-period overlap, and distance. The path from *A. afarensis* is ignored in this analysis due to its low probability of transmitting HSV2. A conditional probability of transmitting HSV2 from *P. boisei* to *H. erectus*, given the probability distributions of all possible paths capable of infecting *P. boisei* that is $P(C|A \cup B \cup C)$, is evaluated as 0.152.

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 path 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 (Sobol′, 2001; Saltelli et al., 2010). 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. While spatial and temporal range is often presented as a hard number in biology the complexities found on the ground are much more nuanced both in terms of daily interactions and evolutionary trajectories (Benton and Hitchin, 1996). These complexities are, naturally, further compounded when dealing with a fragmentary fossil record. In order to allow for natural uncertainty to be introduced into the models the level of variation for sensitivity analysis was set at 10% to mirror diffusion between temporal and spatial borders. 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 results from the Bayesian inference and the optimal path traversal using the Infection Prevalence and Infection Transmission models are presented below, beginning with the Bayesian inference analysis, the optimal path traversal results and finally the sensitivity analysis of the different models.

Bayesian inference

By conditioning anc-chimps and H. erectus for the presence of HSV2, the combined Bayesian inference of the DAG reveals *A. afarensis* (90%), *H. habilis* (67.8%), *P. boisei* (50%), and *H. rudolfensis* (55.8%) to be the likely intermediary hosts for the transmission of HSV2. These are the most probable hosts to have acquired HSV2, and a transmission path through these hosts is more likely.

Optimal path traversal

The optimal traversal path of the DAG for HSV2 transmission was modelled using the Infection Prevalence and the Infection Transmission models. The A* algorithm using the weighted DAGs from both models identified *Paranthropus boisei* as the most probable intermediary hosts that transmitted HSV2 from anc-chimps to the ancestors of modern humans (see fig. 1). In addition to the direct transmission of HSV2 from anc-chimps to *P. boisei*, the Infection Transmission model also identified Homo habilis as an intermediary host that transmitted HSV2 to *P. boisei*, which subsequently transmitted to the ancestors of modern humans (see fig. 2). Table 1 shows the probable HSV2 intermediary paths and their rankings based on the edge weights from the probabilistic models.

Table 1. Probable HSV2 transmission paths

Disease Transmission Route	HSV2-IP path ranking (normalised path cost)	HSV2-IT path 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 path has a cost of 1. The path costs are normalised to the shortest path.

Sensitivity analysis

We performed sensitivity analysis allowing the input parameters of the probabilistic distributions for both the models to vary by 10%. A total of 2.3 million analyses were performed for each model by sampling 50,000 times from the respective probability distributions for each edge. The cost of transmission of HSV2 along all possible paths (path costs) from anc-chimps to *H. erectus* were calculated. The path cost represents the ease of transmitting HSV2 through the path. A path cost is calculated as an arithmetic sum of the inverse probability of transmission through each edge in a given transmission path. The most probable transmission path is always the one that has the lowest cost (i.e., the sum of the probabilities of all of the edges in the path is the highest).

HSV2-IP model

The input parameters to this model were: distance to the rainforest and the overlap in the time-period was allowed to vary by 10%. Sensitivity analysis of HSV2-IP model always predicted the transmission path as 'anc-chimps → *P. boisei* → *H. erectus*'. The average path cost was 0.587, and it varied between 0.526 and 0.649. The distribution of path costs obtained from the sensitivity analysis is shown in fig. 4. The total-order and first-order Sobol indices of the 'anc-chimp → *boisei*' path is 4.53E-2 and 4.53E-2, respectively. The '*P. boisei* → *H. erectus*' path has a total index of 9.55E-1 and a first-order index of 9.55E-1. All other paths had negligible or zero index values. The frequency of occurrence of different transmission paths from the sensitivity analyses are summarised in supplementary table A.5. When the critical path, defined as the path that costs the least to traverse, through *P. bosei* is not available, the transmission path from anc-chimps to *H. erectus* was through H. habilis (normalised path costs, defined as the ratio of the path cost to the critical path cost in the graph, 1.08 - 1.32 of the critical path cost) for 67% of the cases and through *H. rudolfensis* (normalised path costs of 1.11 - 1.32 of the critical path cost) for the remaining 33%.

HSV2-IT model

In this analysis, the input parameters of the gamma distribution were: the proximity of species and the time-period overlap between species, varied by 10%. The distribution of individual path costs used to populate the edge costs of the DAG used in the sensitivity analysis are shown in fig. 5. Sensitivity analysis predicted a primary transmission path (60% of the cases) of lysis predicted a primary transmission path (60% of the cases) of 'anc-chimps → H. habilis → P. boisei → H. erectus' and a secondary transmission path (remaining 40%) through 'anc-chimps → P. boisei → H. erectus'. The distribution of the shortest path costs for both the primary and secondary transmission paths are shown in fig. 6. The Sobol indices for the HSV2-IT model are presented in table 2.

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

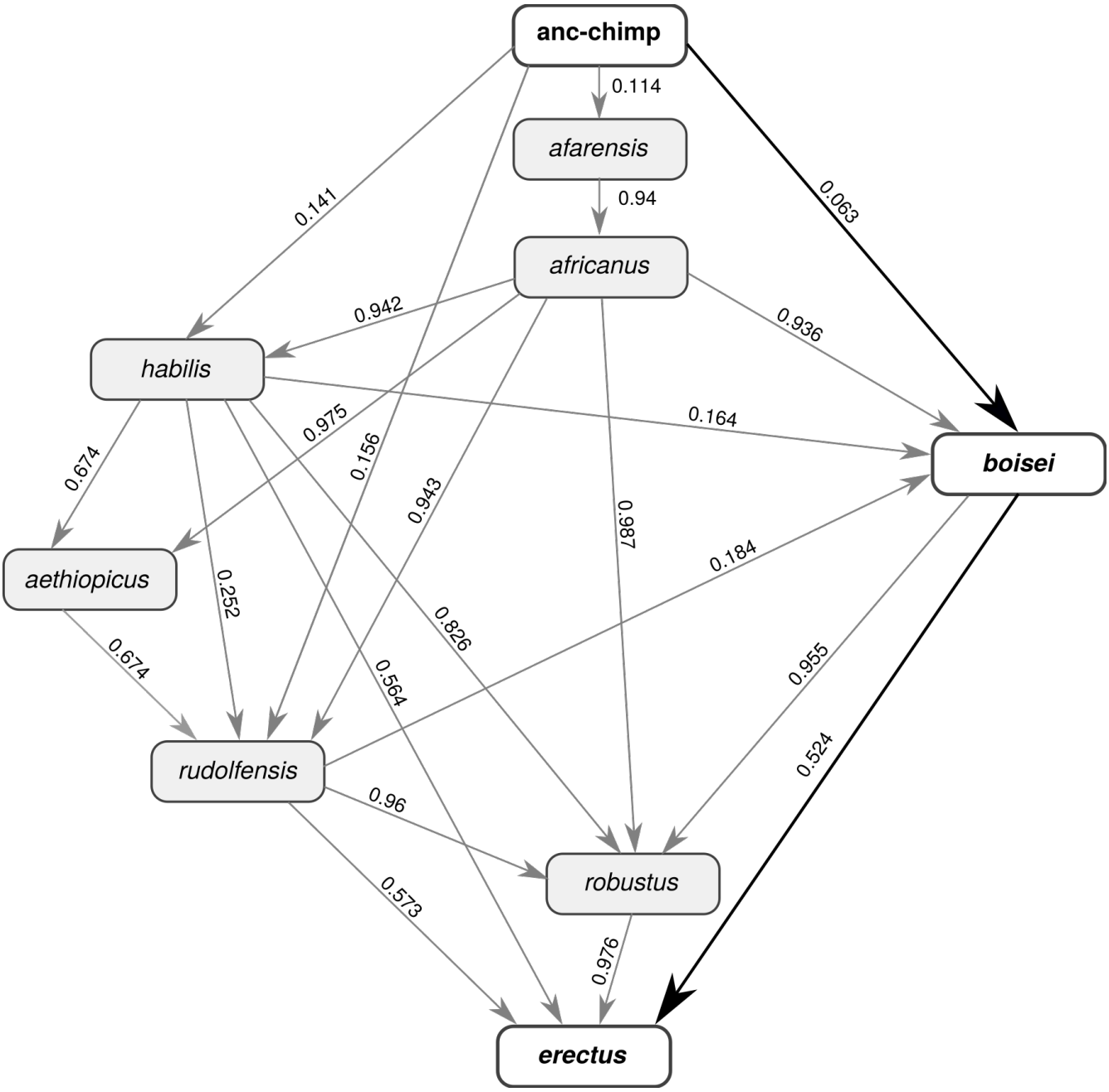


Figure 1. the A* shortest path path(s) for the Infection Prevalence model. The lines with arrows are the possible transmission paths. The values on the lines are the edge costs (inverse probability of transmission). This model predicts that the host switch of HSV2 occurred through the path ancestral-chimps,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

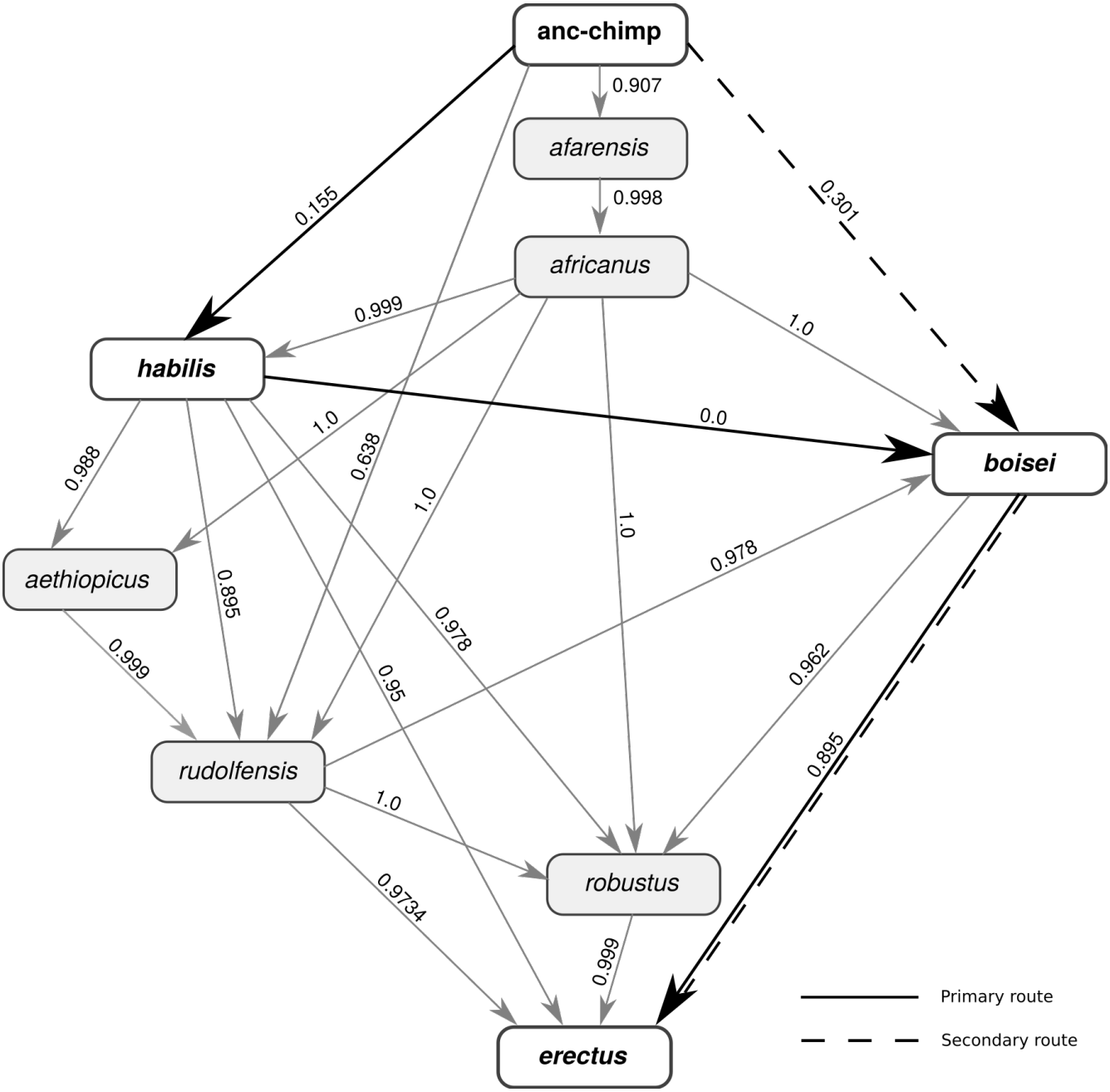


Figure 2. the A* shortest path for the Infection Transmission model. The lines with arrows are the possible transmission paths. The values on the lines are the edge costs (inverse probability of transmission). This model predicts that the host switch of HSV2 occurred through the primary path of ancestral-chimps,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

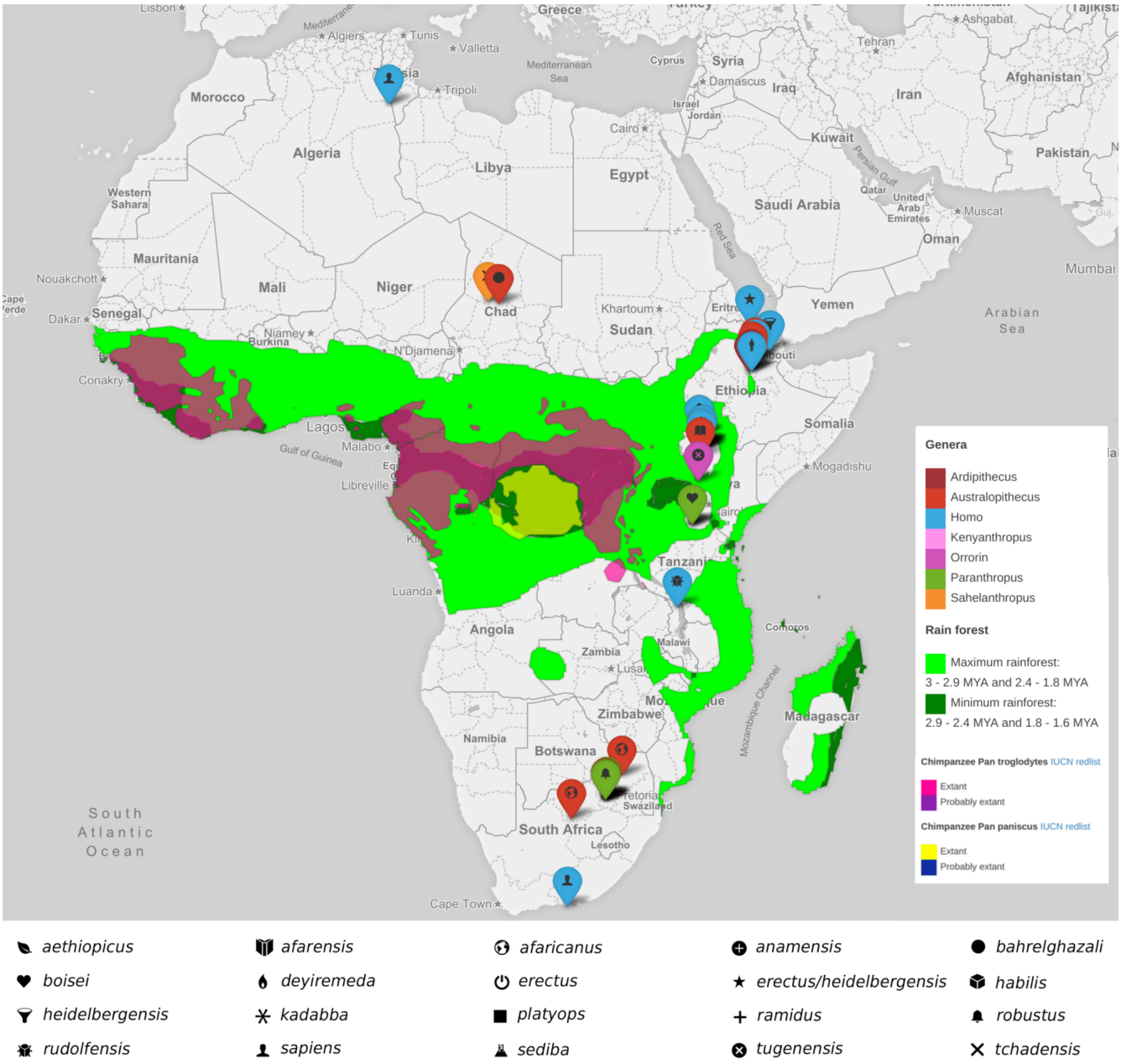


Figure 3. Map showing the distribution of extant chimpanzee (Pan troglodytes) and bonobo (Pan paniscus) populations [IUCN redlist http://maps.iucnredlist.org/map.html?id=15933, http://maps.iucnredlist.org/map.html?id=15932]; the locations of hominin fossils [supplementary table 1] are shown with markers. The colour of the marker indicates the hominin genus; the symbol represents the species. The map also shows the location of hominin fossils relative to ancient minimum and maximum rainforest distributions (Peel, Finlayson, and McMahon, 2007). This figure is available interactively: https://wadhamite.github.io/hsv2-map/.

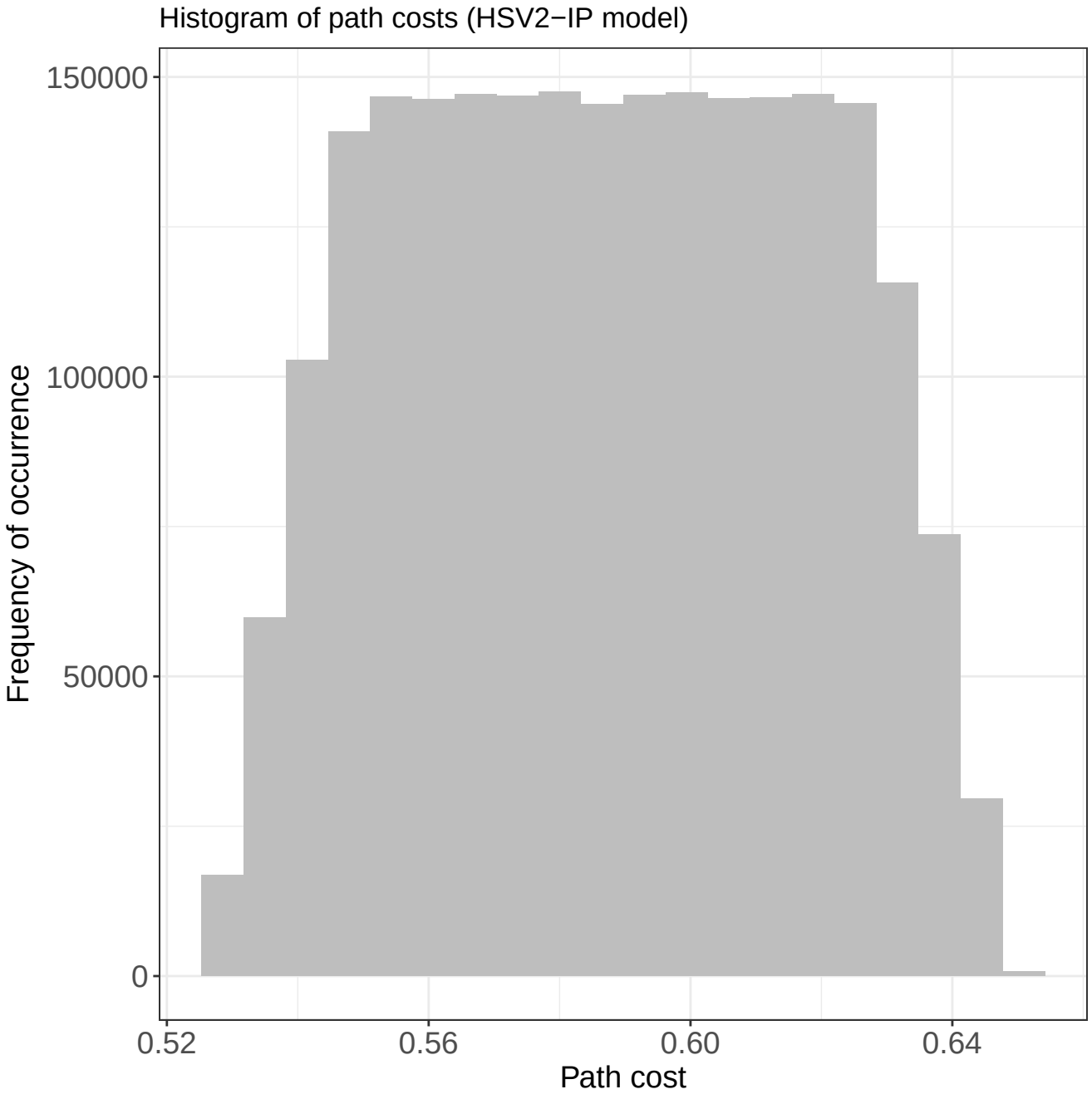


Figure 4. Path costs distribution for Infection Prevalence model. The shortest path is 'anc-chimp ⇒ *P. boisei* ⇒ *H. erectus*' with an average path cost 0.587.

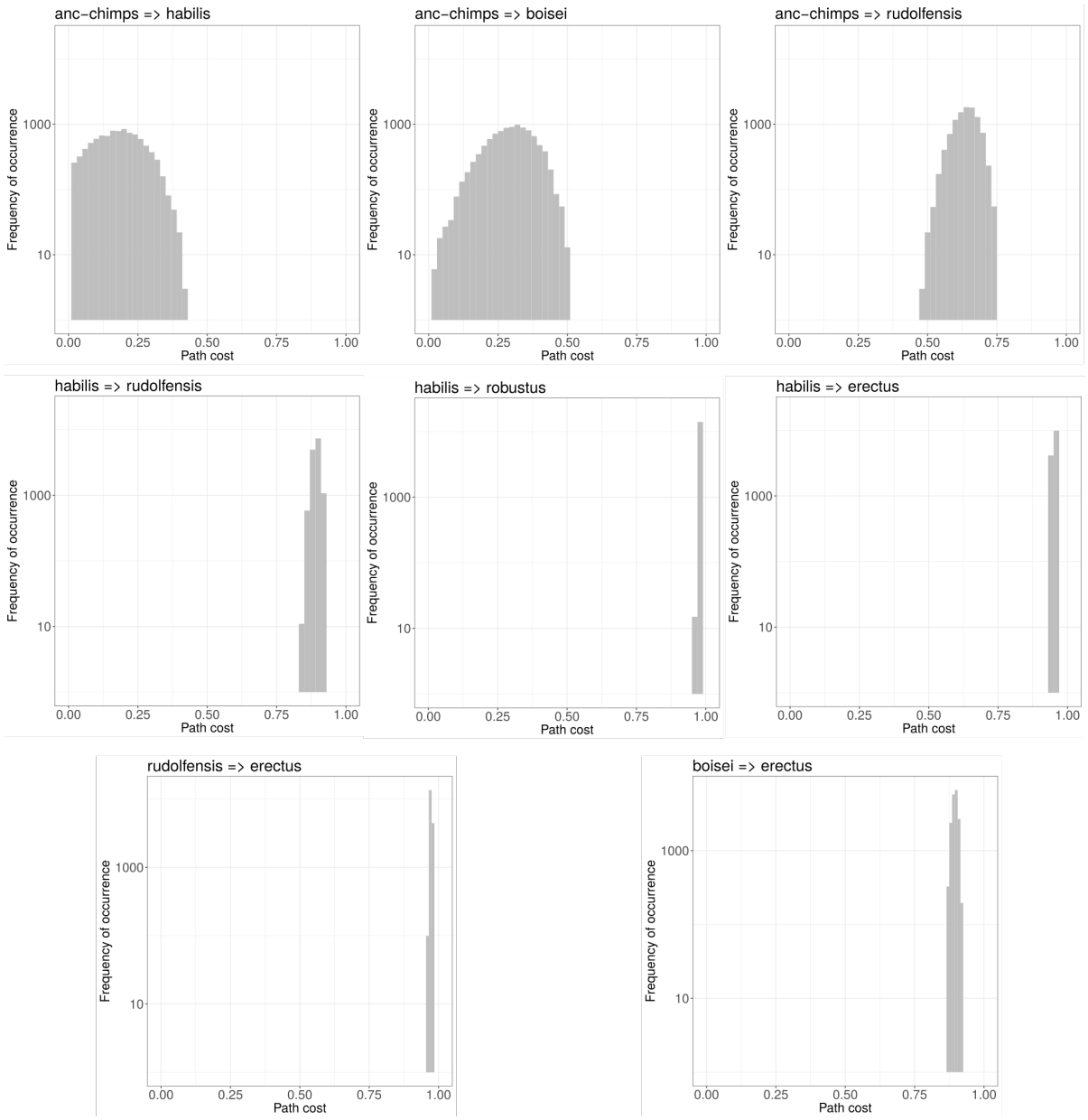
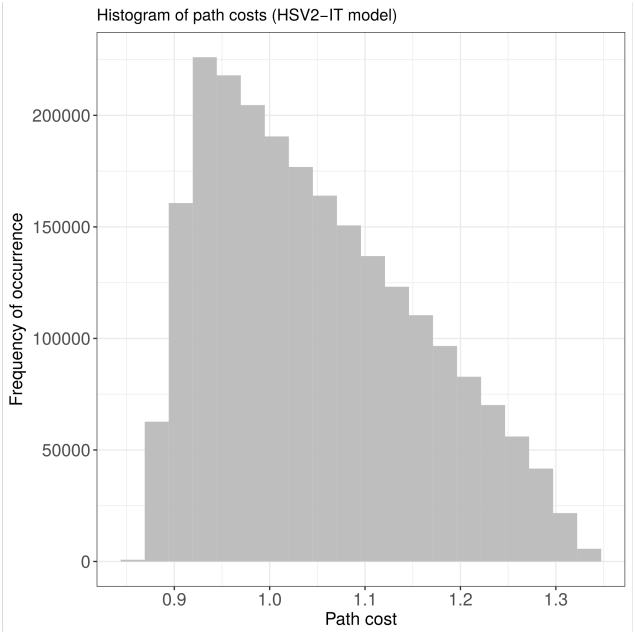
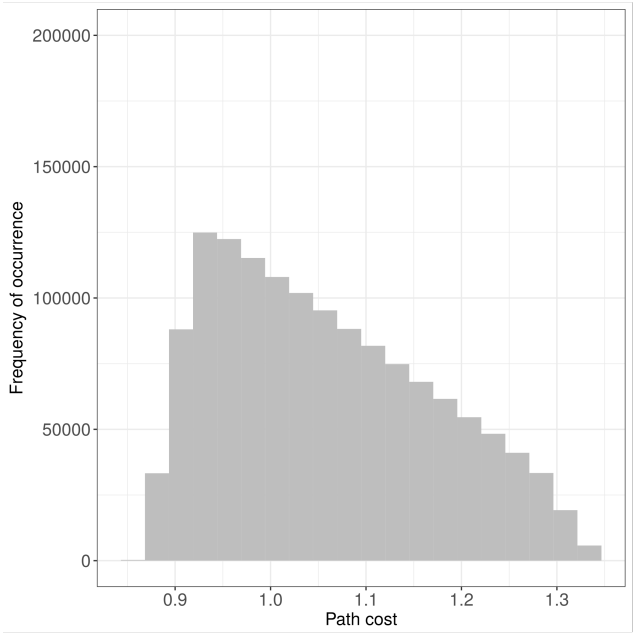


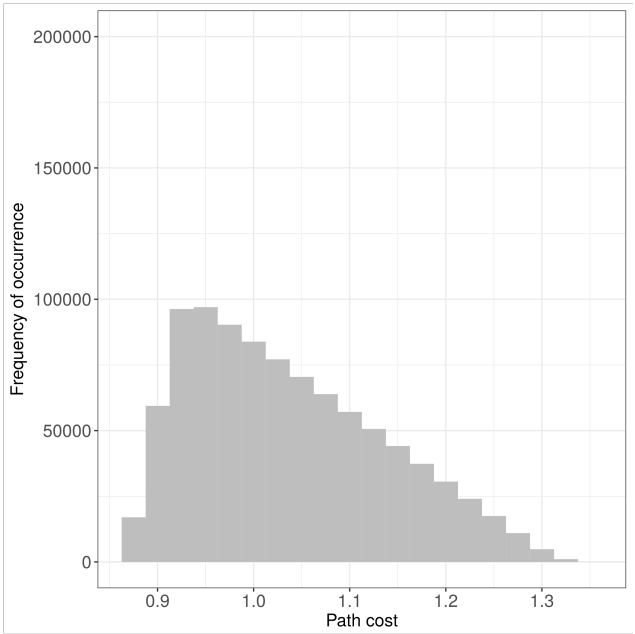
Figure 5. Distribution of selected edge costs for Infection Transmission model. Each histogram shows the distribution of edge costs for the IT model of HSV2 transmission. Colour contours indicate the number of occurrences of a path cost in the sensitivity analysis. The lowest edge costs are seen for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



(a) Infection Transmission model (all routes)



(b) Primary transmission route



(c) Secondary transmission route

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 path 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 path ('anc-chimp ⇒ *H. habilis* ⇒ *P. boisei* ⇒ *H. erectus*'), and (c) distribution path costs for the secondary transmission path ('anc-chimp ⇒ *P. boisei* ⇒ *H. erectus*').

Discussion

Transmission of HSV2

The combined inference analysis of a Bayesian Network, considering the presence of HSV2 in both anc-chimps and Homo erectus, revealed *A. afarensis* (90%), *H. habilis* (67.8%), *P. boisei* (50%), and *H. rudolfensis* (55.8%) to be the likely intermediary hosts. Although, *A. afarensis* had a close proximity to the geographical range of anc-chimps, it could only transmit HSV2 to *A. africanus*. However, *A. africanus* has a mere 3% chance of being infected by HSV2 due its geographic location. This limits the possibility of HSV2 transmission to *H. habilis*, *P. boisei* and *H. rudolfensis*, either directly or as intermediary hosts.

The Infection Prevalence (HSV2-IP) is a local model that describes the probability of HSV2 infection based on the proximity to the rainforest habitat and the duration a species persisted in the fossil record using a beta distribution. In a beta distribution, the parameters alpha (time period) and beta (proximity to the rainforest) are weighted equally, i.e., an increase in the distance to the rainforest by 1 kilometer has the same effect on probability as a decrease of 1000 years in the time period. Thus the beta distribution assumes a constant rate of infection transmission. The HSV2-IP model considers each transmission as an independent event, i.e., the probability of a species transmitting HSV2 is independent of previous events leading to the HSV2 infection of that species.

The A* algorithm using the weighted DAG with the HSV2-IP model predicted the transmission of HSV2 from anc-chimps to *H. erectus* through *P. boisei* with an average path cost of 0.587. Sensitivity analysis did not show any variation in the path (i.e., varying the location and the time-period of the probable species by 10% did not affect the optimal path). Non-zero values of sensitivity indices were observed only along the critical transmission path of 'anc-chimp → P. boisei → *H. erectus*'. Although, the '*P. boisei* → *H. erectus*' path accounted for 95.5% variation in the path costs, the path cost of 'anc-chimp → *P. boisei*' was the critical value that controlled the transmission path. Hence, the optimal path through *P. boisei* remained unaffected. Further analysis of the path revealed a significant influence (88 - 90% of the variation in the results) of *P. boisei* (its proximity and duration) on the path cost in contrast to the effect of proximity and duration of anc-chimps and *H. erectus* on the path costs. The infection prevalence model uses a beta distribution, which assumes equal weights for the shape parameters: *alpha* and *beta*. As expected, Sobol analysis showed equal influence for the distance and time-period parameters (a first-order Sobol indices of 0.493 and 0.492 respectively) on the path costs. Due to the nature of beta distribution, care should be taken in defining the rate of transmission (ratio of the shape parameters, alpha and beta), as the probability value estimated from the distribution is dependent on the ratio of the shape parameters. The intermediary host identified as *P. boisei* is only 35 kilometers from the rainforest, which results in high probability of transmission, and hence results in the lowest cost of transmission through this species. HSV2-IP is a local model, which is independent of other events and does not consider the history of transmission and the associated probabilities.

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 HSV2-IT model is history dependent, i.e., the probability of a child node transmitting a disease depends on the probability of the parent nodes infecting the child node in a DAG. The Infection Transmission model predicts an initial transmission of HSV2 from anc-chimps to *H. habilis*. Further, *H. habilis* transmitted the virus to *P. boisei*, which then infected *H. erectus*. The average shortest path cost was estimated as 1.05. Sensitivity analysis of the HSV2-IT model revealed two transmission paths: (a) Primary transmission path (60% of the cases) of 'anc-chimps → *H. habilis* → *P. boisei* → *H. erectus*' and (b) a secondary transmission path (remaining 40% of the cases) through 'anc-chimps → *P. boisei* → *H. erectus*'. However, the distribution of path costs through both the paths were similar, showing that both transmission paths could have been possible. Sensitivity analysis shows that the 'anc-chimp → *H. habilis*' and 'anc-chimp → *P. boisei*' have quite a significant effect on the transmission path (total path cost). About 70% the variation in the path of 'anc-chimp → *H. habilis*' is due to the direct effect of edge weight determined by the gamma distribution in comparison to 64% of variation for 'anc-chimp → P. boisei' path. Secondary interactions (interactions between different paths) accounts for 30 - 35% of the overall routing. The probability of a species transmitting, and another being infected had similar effects (similar Sobol indices), which explains that the model did not have a bias either toward the proximity or the temporal overlap between species. Because the HSV2-IT model considers the history of transmission, it is a better predictor than the HSV2-IP model.

Our analysis suggests that *Paranthropus boisei* was the most critical intermediate host for transmitting HSV2 between anc-chimps and the ancestors of *Homo sapiens*. The transmission of HSV2 from anc-chimps to *P. boisei* could have happened directly or through *Homo habilis*. However, our analysis did not predict *Homo habilis* transmitting HSV2 directly to *Homo erectus*. Our results suggest the initial transmission to *P. boisei* could have happened from *H. habilis* 60% of the time due (most likely through direct sexual contact rather than consumption of infected tissue) and the remaining 40% was directly transmitted from anc-chimps via hunting or scavenging.

P. boisei would have been well placed to act as an intermediate host for HSV2. It most likely contracted the infection through hunting or more likely scavenging infected ancestral-chimpanzee meat. Processing (with or without tools) and consumption of raw meat would act as a simple path for ChHV1 to have crossed into *P. boisei* via open cuts or sores. Tropical refugia during hot

dry periods may have driven chimpanzees into higher concentrations in certain areas, driving them into contact and competition with *P. boisei* and *Homo habilis* as the margins of tropical forest blended into more open savannah like habitats (Julier et al., 2017). Violent confrontation or hunting/scavenging and butchery practices would have provided a viable path of transmission for HSV2. *Homo habilis* remains have been recovered from the same layers as stone tools and bones carrying evidence of butchery, supporting a possible transmission-through-hunting/scavenging hypothesis for the initial anc-chimp to *H. habilis* transmission (Clarke, 2012). *Paranthropus aethiopicus*, *P. boisei*, and *P. robustus* are associated with the Oldowan stone tool complex (De Heinzelin et al., 1999), and *P. boisei* explicitly with butchery (Domínguez-Rodrigo et al., 2013) lending support to the hypothesis that bushmeat hunting/scavenging and butchery may have led to the initial transmission of HSV2 to the hominins.

Both *Homo erectus* and *Paranthropus boisei* are known from sites around Lake Turkana in Kenya that are contiguous in age (Anton et al., 2016; Wood and Constantino, 2007) and it is likely that close contact between the species would have been relatively common especially around water sources. The appearance of *Homo erectus* 2.0 MYA is accompanied by evidence of active hunting and butchery, and from 1.76 MYA increasingly sophisticated stone tools (Cachel and Harris, 1998). This behavioural shift towards active hunting displayed by *Homo erectus*, combined with direct archaeological evidence of hunting, provides a credible path for direct transmission of HSV2 from *P. boisei* to *H. erectus* through contact and/or consumption of infected material processed from *P. boisei* carcasses.

In this study, *Homo erectus* is chosen as the target node for the transmission of HSV2 for a number of reasons. Morphological adaptation for bipedal locomotion is used as the primary trait for assigning fossils to the hominin sub-family, especially between 7.0 - 4.5 MYA, but is not an effective tool for determining patterns of ancestor-descendant relationships in the fossil record. The hominin sub-family currently contains seven genera but the exact taxonomic relationship between each genus is not clearly definable because of the fragmentary nature of the fossil record. Similarly, patterns of intra- and inter-species variation are difficult to define (Robert A Foley, 2016). The adaptive radiation of the genus Australopithecus between 4.5 - 2.0 MYA represents the first morphologically coherent group of fossil hominin species but its relationship to the genus Homo is not yet clear. The appearance of Homo erectus circa 2.0 MYA in East Africa marks the first appearance of recognisably 'human' morphology, life history and brain development and represents a secure most recent common ancestor (MRCA) for all subsequent Homo species (with the possible exception of *Homo floresiensis* (Argue et al., 2017; Anton et al., 2016).). Therefore, the ancestor-descendant relationship between *Homo erectus* at c. 2.0 MYA and *Homo sapiens* c. 200 KYA is an evolutionarily secure path of transmission for HSV2 to leave Africa as a modern human-borne virus.

Although ChHV1 causes outbreaks of oral and pharyngeal lesions in chimpanzees in a manner similar to HSV1, in hominins contracting 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. However, both simplexviruses remain capable of infecting both the oral and genital niche in modern humans (Kim et al., 2006; Whitley, Kimberlin, and Prober, 2007), and of causing co-infection in both niches (eg 10% to 15% of herpes labialis (oral lesions) is caused by HSV-2 (Glick and Siegel, 1999)).

We suggest that the mode of transmission of HSV2 to 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 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 (Disease Control, Prevention, et al., 1987). HSV1 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 (Looker et al., 2015) (see supplementary fig. A.1 and interactive maps at `https://wadhamite.github.io/hsv2-map`) is consistent with the limited genetic data available from African HSV2 isolates. A study from (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, two groups found evidence of gene flow from HSV1 into HSV2, and speculated 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 41 KYA (Burrel, Desire, et al., 2015; Koelle et al., 2017). Recent studies have significantly increased the number of whole HSV2 genomes available

for analysis, contributing to our knowledge of HSV2 diversity (Kolb et al., 2015; Szpara et al., 2014); but there is conflicting evidence on whether the most basal HSV-2 genotypes are from west and central (Burrel, Boutolleau, et al., 2017) or east (Koelle et al., 2017) Africa. Our analyses predict that individuals from east Africa are likely to carry the most 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 (Allentoft et al., 2012) 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. There are other ancient human pathogens that have switched between different primate and hominin hosts over the last 6 million years and these transmission paths could be further explored with this methodology. For example, human pubic lice (*Pthirus pubis*) were introduced by an unknown hominin through contact with the ancestor of gorillas around 3.3 MYA (D. L. Reed et al., 2007). There is also evidence of Neanderthal to human transmission of human papillomavirus genotypes (Pimenoff, Oliveira, and Bravo, 2017), and of hominin to human transmission of body louse genotypes (D. D. L. Reed et al., 2004). The studies and the findings presented here demonstrate the potential for using modern disease genetics to understand better the evolutionary interaction between humans and disease in deep time.

Data presentation

The distribution maps were created with custom JavaScript codes using Leaflet.js library [http://leafletjs.com/] and interactive maps can be accessed at https://wadhamite.github.io/hsv2-map

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

The authors would like to thank C Ruis (University College London), and JB Ramond and RF Rifkin (University of Pretoria) for helpful discussion. 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. CH is a post-doctoral affiliate at Christ's College, Cambridge.

References

Allentoft, Morten E. et al. (2012). "The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils". In: *Proceedings of the Royal Society of London B: Biological Sciences* 279.1748.

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.

Argue, Debbie et al. (2017). "The affinities of Homo floresiensis based on phylogenetic analyses of cranial, dental, and postcranial characters". In: *Journal of Human Evolution* 107, pp. 107–133. ISSN: 00472484. DOI: 10.1016/j.jhevol.2017.02.006.

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.

Baines, Joel D. and Philip E. Pellett (2007). *Genetic comparison of human alphaherpesvirus genomes*. Cambridge University Press. ISBN: 9780521827140.

Benton, Michael J and Rebecca Hitchin (1996). "Testing the quality of the fossil record by groups and by major habitats". In: *Historical Biology* 12.2, pp. 111–157.

Bertke, Andrea S, Amita Patel, and Philip R Krause (2007). "Herpes simplex virus latency-associated transcript sequence downstream of the promoter influences type-specific reactivation and viral neurotropism." In: *Journal of virology* 81.12, pp. 6605–13. ISSN: 0022-538X. DOI: 10.1128/JVI.02701-06.

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.

Cachel, S and JWK Harris (1998). “The lifeways of Homo erectus inferred from archaeology and evolutionary ecology: a perspective from East Africa”. In: *Early human behaviour in global context:*

Chen, Yiping et al. (2008). “Finding a Better Immunization Strategy”. In: *Physical Review Letters* 101.5, p. 058701. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.101.058701.

Clarke, R.J. (2012). “A Homo habilis maxilla and other newly-discovered hominid fossils from Olduvai Gorge, Tanzania”. In: *Journal of Human Evolution* 63.2, pp. 418–428. ISSN: 00472484. DOI: 10.1016/j.jhevol.2011.11.007.

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

Disease Control, Centers for, Prevention, et al. (1987). “B-virus infection in humans–Pensacola, Florida”. In: *MMWR: Morbidity and mortality weekly report* 36.19, pp. 289–90.

Domínguez-Rodrigo, Manuel 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.

Elton, Sarah (2008). “The environmental context of human evolutionary history in Eurasia and Africa”. In: *Journal of Anatomy* 212.4, pp. 377–393.

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.

Foley, Robert A. (1978). “Incorporating sampling into initial research design: some aspects of spatial archaeology”. In: *Sampling in contemporary British archaeology*. Ed. by John F. Cherry, Clive. Gamble, and Stephen. Shennan. BAR British Series, pp. 49–66.

Foley, Robert A (2016). “Mosaic evolution and the pattern of transitions in the hominin lineage”. In: *Phil. Trans. R. Soc. B* 371.1698, p. 20150244.

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.

Glick, Michael and Michael A. Siegel (1999). “VIRAL AND FUNGAL INFECTIONS OF THE ORAL CAVITY IN IMMUNOCOMPETENT PATIENTS”. In: *Infectious Disease Clinics of North America* 13.4, pp. 817–831. ISSN: 08915520. DOI: 10.1016/S0891-5520(05)70110-0.

Grant, J.W.A., C.A. Chapman, and K.S. Richardson (1992). “Defended versus undefended home range size of carnivores, ungulates and primates”. In: *Behavioral Ecology and Sociobiology* 31.3, pp. 149–161. ISSN: 0340-5443. DOI: 10.1007/BF00168642.

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). “Migrating microbes: what pathogens can tell us about population movements and human evolution”. In: *Annals of Human Biology* In press, pp. 1–11. ISSN: 0301-4460. DOI: 10.1080/03014460.2017.1325515.

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.

Julier, A. et al. (2017). “Characterisation and differentiation of the modern pollen-vegetation relationships of sites within a forest-savannah mosaic landscape in tropical West Africa (Ghana)”. In: *Palynology* In press.

Kim, H. Nina et al. (2006). “Oral Herpes Simplex Virus Type 2 Reactivation in HIV-Positive and -Negative Men”. In: *The Journal of Infectious Diseases* 194.4, pp. 420–427. ISSN: 0022-1899. DOI: 10.1086/505879.

Koelle, David M. et al. (2017). “Worldwide circulation of HSV-2 $\times$ HSV-1 recombinant strains”. In: *Scientific Reports* 7, p. 44084. ISSN: 2045-2322. DOI: 10.1038/srep44084.

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.

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.

McBrearty, Sally and Nina G. Jablonski (2005). "First fossil chimpanzee". In: *Nature* 437.7055, pp. 105–108. ISSN: 0028-0836. DOI: 10.1038/nature04008.

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.

Myers Thompson, Jo A (2003). "A model of the biogeographical journey from Proto-pan to Pan paniscus." In: *Primates; journal of primatology* 44.2, pp. 191–7. ISSN: 0032-8332. DOI: 10.1007/s10329-002-0029-1.

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.

Pimenoff, Ville N, Cristina Mendes de Oliveira, and Ignacio G Bravo (2017). "Transmission between Archaic and Modern Human Ancestors during the Evolution of the Oncogenic Human Papillomavirus 16." In: *Molecular biology and evolution* 34.1, pp. 4–19. ISSN: 1537-1719. DOI: 10.1093/molbev/msw214.

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

Reed, David L et al. (2007). "Pair of lice lost or parasites regained: the evolutionary history of anthropoid primate lice". In: *BMC Biology* 5.1, p. 7. ISSN: 17417007. DOI: 10.1186/1741-7007-5-7.

Reed, DL David L et al. (2004). "Genetic analysis of lice supports direct contact between modern and archaic humans." In: *PLoS biology* 2.11. Ed. by Nick Barton, e340. ISSN: 1545-7885. DOI: 10.1371/journal.pbio.0020340.

Saltelli, Andrea et al. (2010). "Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index". In: *Computer Physics Communications* 181.2, pp. 259–270. ISSN: 00104655. DOI: 10.1016/j.cpc.2009.09.018.

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

Sobol′, I.M (2001). "Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates". In: *Mathematics and Computers in Simulation* 55.1-3, pp. 271–280. ISSN: 03784754. DOI: 10.1016/S0378-4754(00)00270-6.

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.

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.

Whitley, Richard, David W. Kimberlin, and Charles G. Prober (2007). *Pathogenesis and disease*. Cambridge University Press. ISBN: 9780521827140.

Wood, Bernard and Paul Constantino (2007). “Paranthropus boisei: Fifty years of evidence and analysis”. In: *American Journal of Physical Anthropology* 134.S45, pp. 106–132. ISSN: 00029483. DOI: 10.1002/ajpa.20732.

Supplementary material

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.*

Table A.2. Matrix of separation distance (in kms) between hominins. Red indicates a longer distance of separation, while dark- green indicates closer proximity between species. When the the time-periods of two hominins do not overlap, they are represented as '-', i.e., there is no possibility of transmission, hence the separation distance between those hominins are not computed.

Distance (km)	*afarensis*	*africanus*	*habilis*	*aethiopicus*	*rudolfensis*	*boisei*	*robustus*	*sediba*	*erectus*
anc-chimps	11	-	174	-	100	35	-	-	-
afarensis		3301.571	-	-	-	-	-	-	-
africanus		-	2442	3173	1624	2442	243	-	-
habilis				736.2	777.4	3.74	2675	2672	776.6
aethiopicus					41.19	-	-	-	-
rudolfensis						777.3	3323	3439	1 or 40
boisei							2675	2671	737
robustus								4	3403
sediba									-

Table A.3. Matrix of temporal overlap between hominins (time in Myr). Red indicates least amount of over-lap, while dark green shows the most time-period overlap between hominins. No overlaps in the time-periods of hominins are shown as '-'.

Time overlap (Myr)	*afarensis*	*africanus*	*habilis*	*aethiopicus*	*rudolfensis*	*boisei*	*robustus*	*sediba*	*erectus*
anc-chimps	0.1	-	1.2	-	0.6	0.7	-	-	-
afarensis		0.13	-	-	-	-	-	-	-
africanus		-	0.76	0.2	0.46	0.26	0.04	-	-
habilis				0.2	0.6	0.8	0.5	0.01	0.4
aethiopicus					0.1	-	-	-	-
rudolfensis						0.4	0.1	0.01	0.1
boisei							0.8	0.01	0.7
robustus								0.01	0.7
sediba									-

Table A.4. The probability of transmission of HSV2 from anc-chimps, *H. africanus*, *H. habilis* and *H. rudolfensis* to *P. boisei* are computed as 0.954, 0.096, 0.996 and 0.34 respectively, using a beta distribution. The Conditional Probability Table for *P. boisei* is shown below.

H. africanus	*H. habilis*	*H. rudolfensis*	anc-chimps	True	False
True	True	True	True	0.9963	0.0037
True	True	True	False	0.0326	0.9674
True	True	False	True	0.0914	0.9086
True	True	False	False	0.0959	0.9041
True	False	True	True	0.0312	0.9688
True	False	True	False	0.0327	0.9673
True	False	False	True	0.0918	0.9082
True	False	False	False	0.09622	0.9038
False	True	True	True	0.3229	0.6771
False	True	True	False	0.3385	0.6615
False	True	False	True	0.9501	0.0499
False	True	False	False	0.9963	0.0037
False	False	True	True	0.3241	0.6759
False	False	True	False	0.33981	0.66019
False	False	False	True	0.95368	0.04632
False	False	False	False	0	1

Table A.5. Summary of sensitivity analyses showing the frequency of occurrence of different transmission routes using both HSV2-IP and HSV2-IT infection models. In total 4.6 million analyses were performed for each model.

Model	Primary transmission path			Second transmission path		
	Intermediary species	Path cost **(normalised)**	% occurrence	Intermediary species	Path cost **(normalised)**	% occurrence
HSV2-IP	*P. boisei*	0.587 **(1.0)**	100%	-	-	-
HSV2-IP (no critical path)	*H. habilis*	0.634 - 0.775 **(1.08 - 1.32)**	67%	*H. rudolfensis*	0.56 - 0.775 **(1.11 - 1.32)**	33%
HSV2-IT	*H. habilis* → *P. boisei*	0.866 - 1.34 **(1.0 - 1.547)**	60%	*P. boisei*	0.866 - 1.34 **(1.0 - 1.547)**	40%

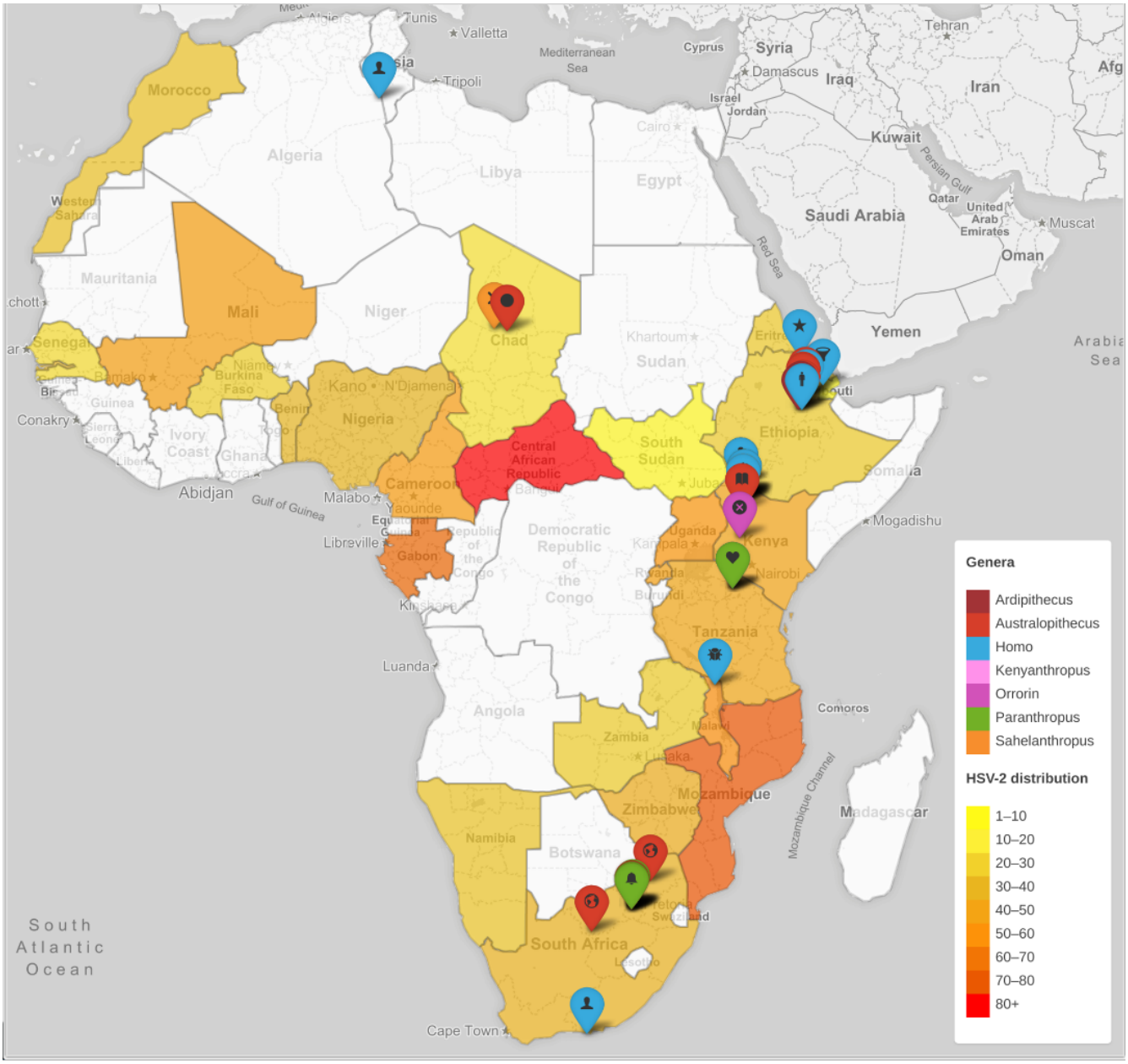
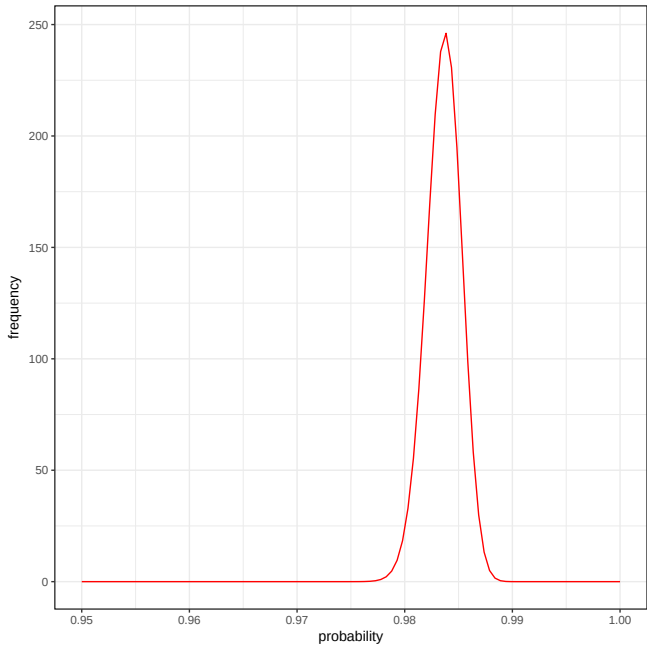
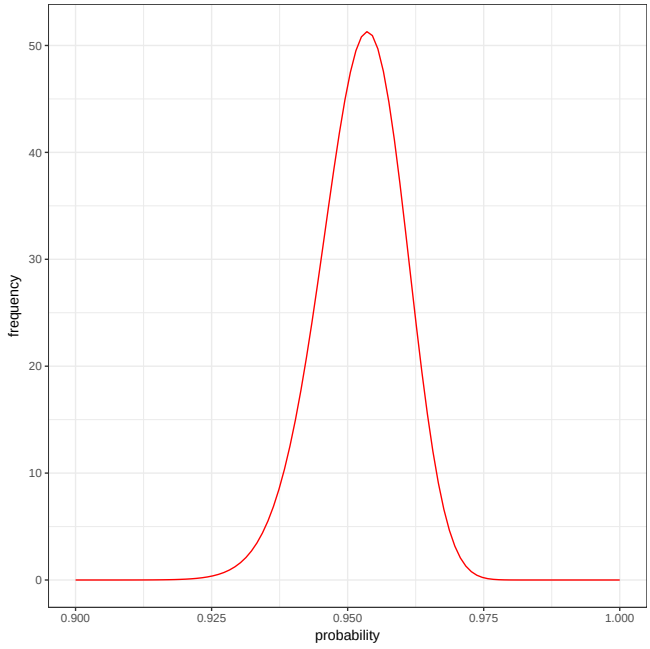


Figure A.1. Map showing the prevalence of HSV2 in Africa, using data from Looker (2015). The locations of hominin fossils [supplementary table A.1] are shown with markers. The colour of the marker indicates the hominin genus; the symbol represents the species. This figure is available interactively: https://wadhamite.github.io/hsv2-map/



(a) anc-chimp



(b) *P. boisei*

Figure A.2. Probability density of anc-chimps and *P. boisei* being infected with HSV2 and transmitting the virus. This figure shows a single run of a Monte Carlo simulation (shown as a blue line). In a single run of a Monte-Carlo simulation, a random probability of transmission of HSV2 from the beta-distribution for anc-chimp is obtained as 0.97 ($P(A)$), while a probability of *P. boisei* being infected is obtained as 0.9375 ($P(B)$). The combined probability is the probability of transmission from anc-chimps to *P. boisei*, $P(A \text{ and } B)$ which is 0.909. In this study, 1000 runs of MC simulations are carried out and the average probability of the anc-chimp to *P. boisei* path is evaluated as 0.937.

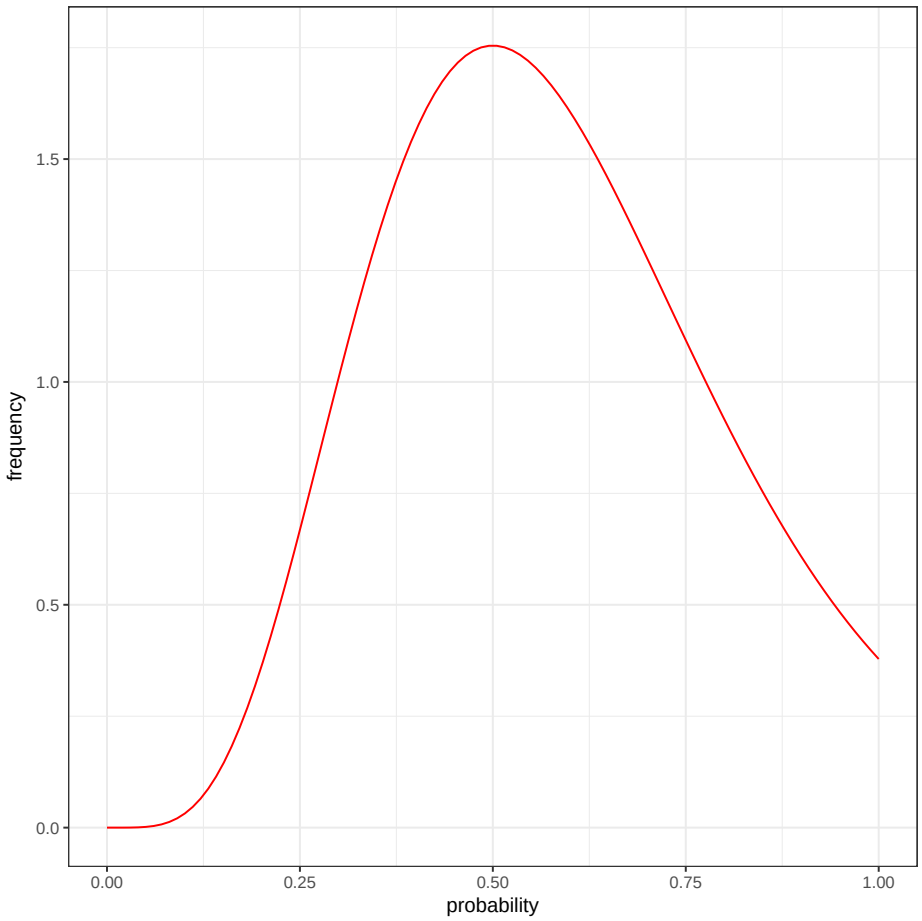


Figure A.3. A typical gamma distribution of probability density of transmission from anc-chimps.The shape parameter α is defined as the ratio of the time period in 1000 years / distance in kilometres ($\alpha = 6$) and the rate parameter β (= 10) is defined as the normalised time period Y/x where x is the time period of the species in 1000 years and Y is the time period of anc-chimps.