Title (14/15 words): Network analysis of the hominin origin of Herpes Simplex virus 2 from fossil data

Short title: Hominin origin of HSV2

Authors: Simon J. Underdown1, 2, Krishna Kumar3, and Charlotte Houldcroft4, 5*

*Corresponding author: Dr Charlotte Houldcroft Ch504[at]camDOTacDOTuk

Affiliations

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.

Abstract 147/150 words

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 an intermediate but unknown hominin species.

In this paper, we use probability-based network analysis to determine the probable path between intermediate hosts of HSV2, from chimpanzees to modern humans, using paleoenvironmental 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 *Homo rudolfensis* as the most likely intermediate host of HSV2

Main text (3325/5000 words)

Herpes simplex virus 2 (HSV2) is a sexually transmitted human pathogen that causes genital lesions and, rarely, encephalitis (eg [1]), and is associated with increased risk of HIV acquisition [2]. 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[3].

HSV2 was originally thought to have co-speciated with humans when our lineage diverged from that of chimpanzees, but recent comparisons of the HSV1, HSV2 and chimpanzee herpesvirus 1 (ChHV1) genomes suggests that HSV2 is in fact more closely related to ChHV1 than HSV1[3]. This analysis also

found that HSV2 diverged from ChHV1 between 1.4 and 3 MYA, and the authors inferred that an intermediate hominin served as a host for HSV2 before it introgressed into the ancestors of modern humans.

We argue 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 chimp/bonobo ancestor, it will be possible to develop a model to statistically infer the species that facilitated the introgression of HSV2 in the modern human lineage defined here as beginning with *Homo erectus*[4]. The ancestor-descendant relationship between *Homo erectus* at c. 2.0 MYA and *Homo sapiens* c. 200 KYA is an evolutionarily secure route of transmission for HSV2 to leave Africa as a modern human-borne virus. This produces a window of infection time that closes with the dispersal of *Homo sapiens* out of Africa no later than 100 KYA[5].

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 (figures 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.

Disease Transmission Route	**HSV2-IP route ranking (normalised path cost)**	**HSV2-IT route ranking (normalised path cost)**	**Bayesian inference (probability)**
anc-chimp --> *rudolfensis* -> *erectus*	**1 (0.0 %)**	**1 (0.0 %)**	**11.59%**
anc-chimp --> *boisei* --> *erectus*	2 (6.24%)	4 (4.03%)	9.28%
anc-chimp --> *habilis* --> *boisei* --> *erectus*	3 (7.97%)	2 (1.92%)	5%
anc-chimp --> *habilis* --> *erectus*	4 (10.91%)	3 (3.39%)	7.8%

* % path cost represents the normalised value with respect to the most and the least probable route. The most probable route has a 0% cost, and least probable route has an 100% cost.

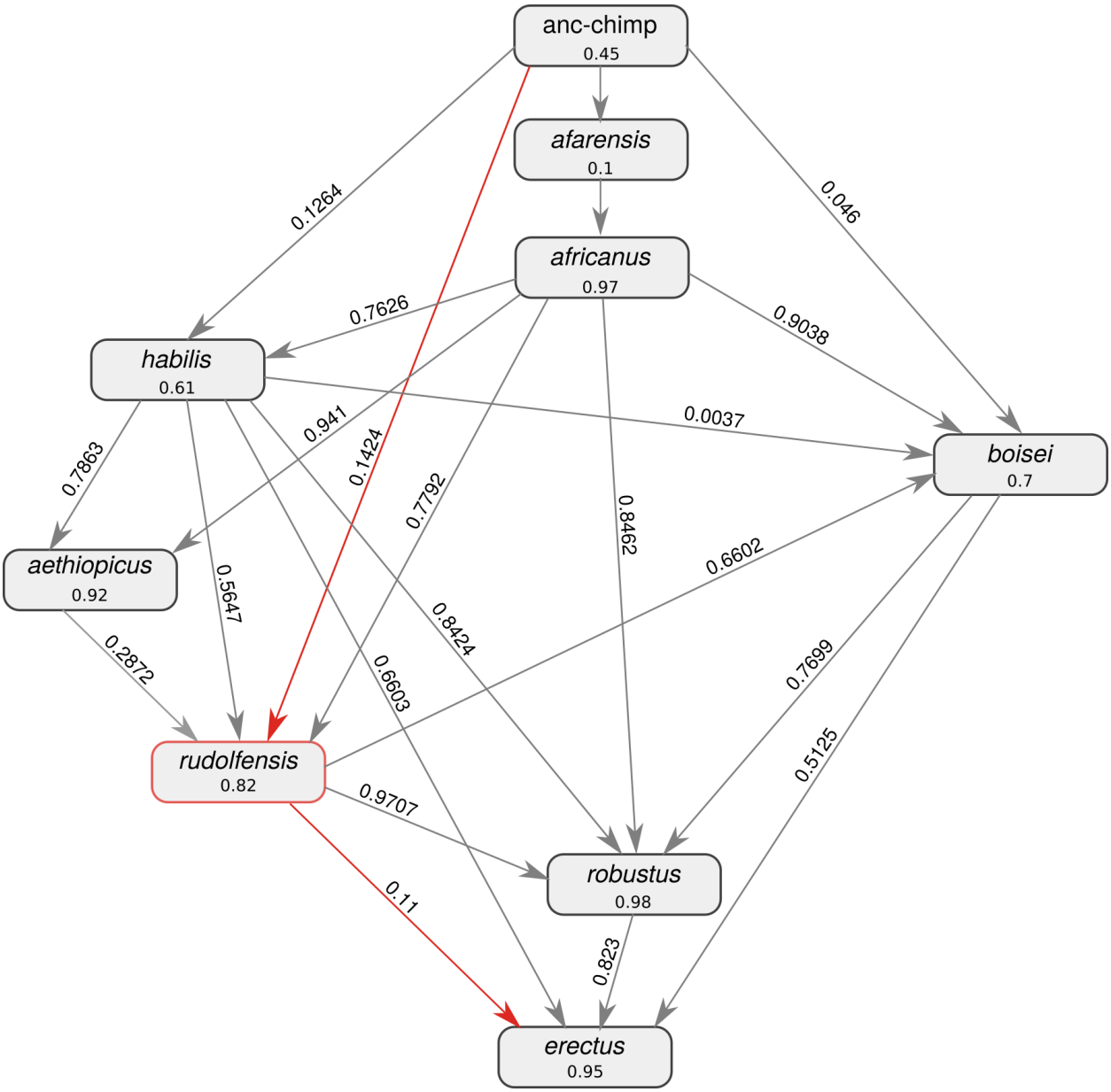

Figure 1: A* shortest path for Infection Prevalence model

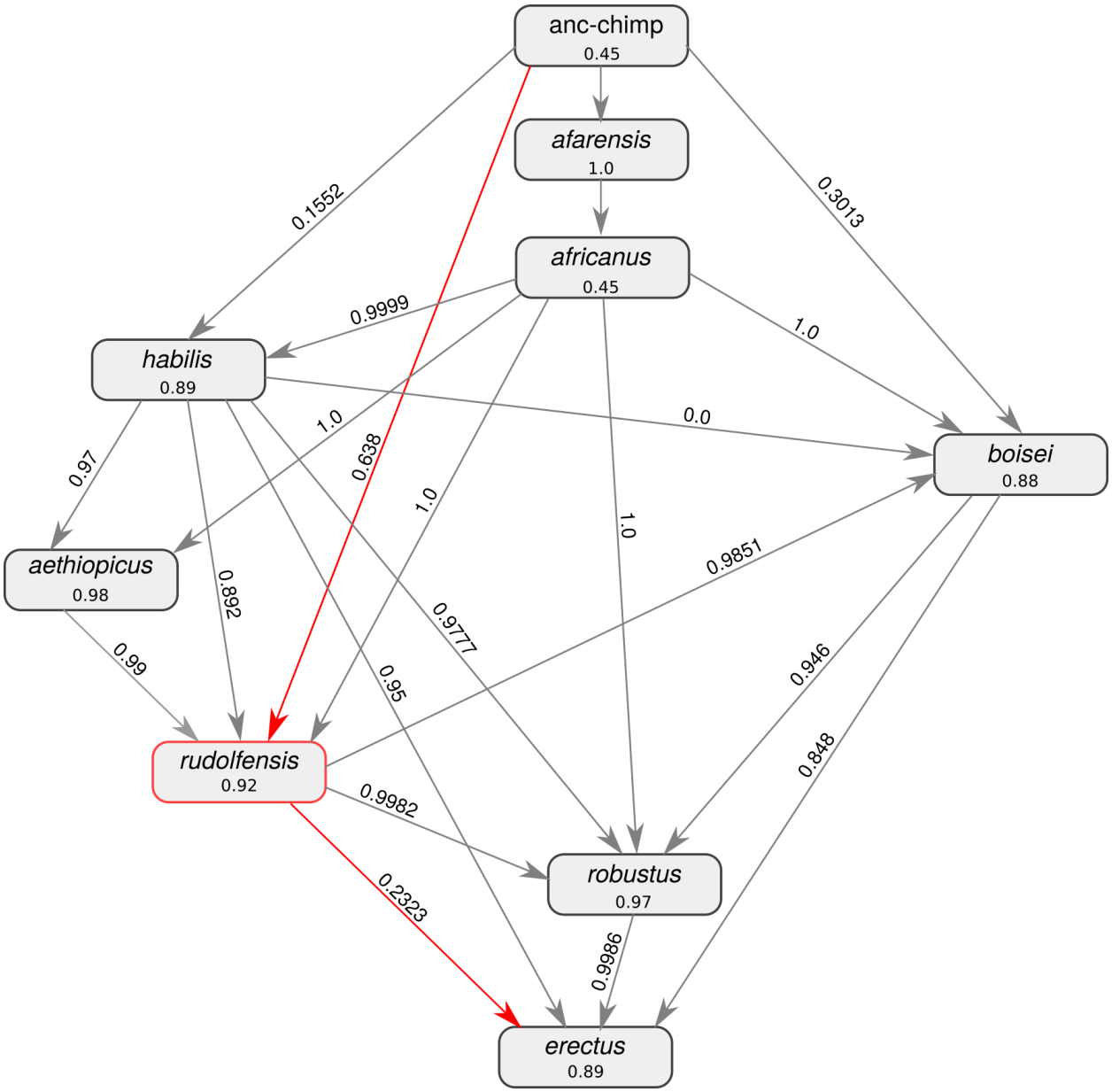

Figure 2: A* shortest path for Infection Transmission model

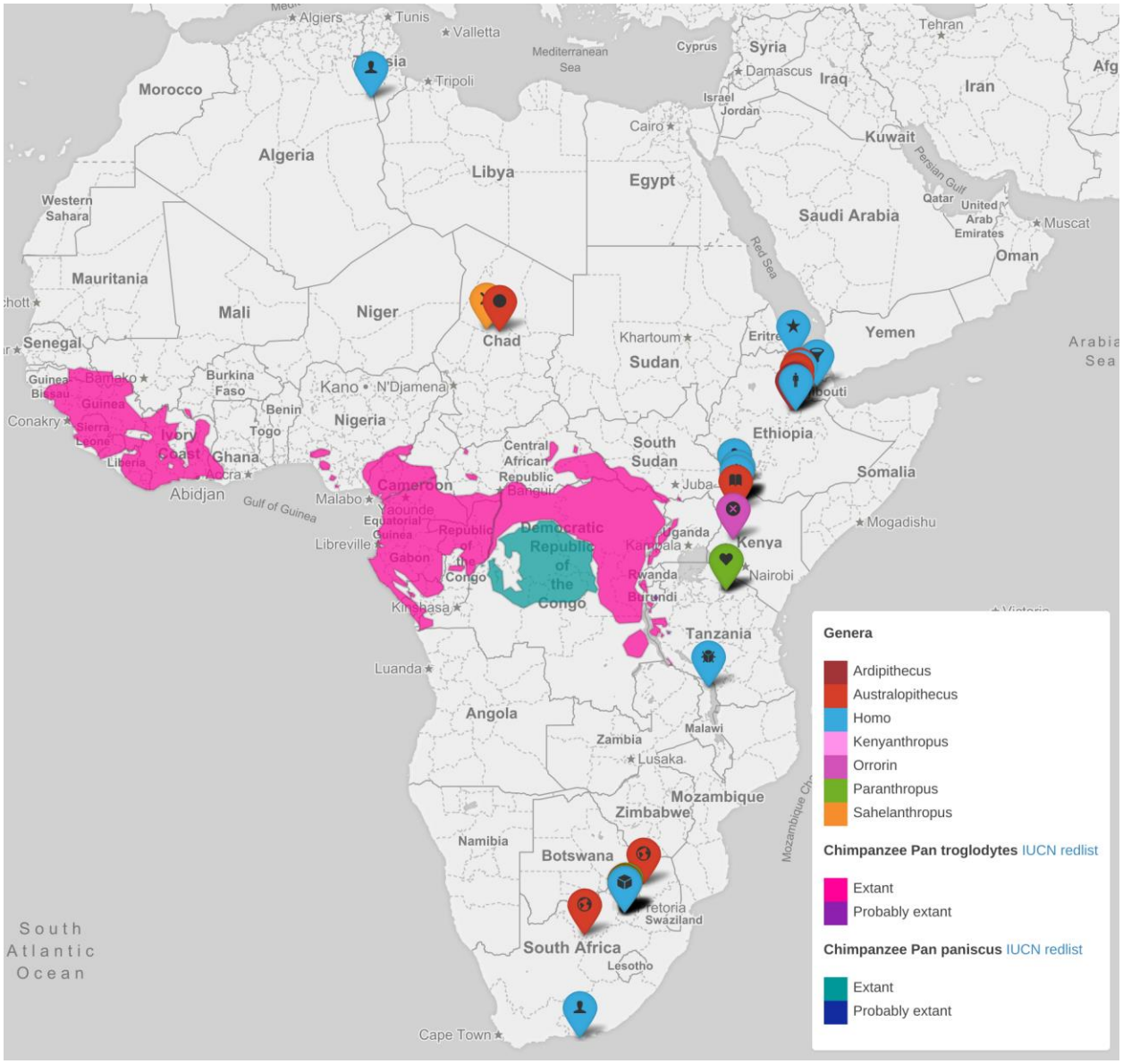

Figure 3: Distribution of Chimpanzee [IUCN redlist http://maps.iucnredlist.org/map.html?id=15933, http://maps.iucnredlist.org/map.html?id=15932] and fossils

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 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*, 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[6].

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[7,8] as Oldowan or Mode 1 stones tools are widely distributed in the same spatial-temporal localities as *Homo rudolfensis* fossils[9,10]. *Paranthropus aethiopicus*, *P. boisei*, and *P. robustus* are associated with the Olduwan stone tool complex[11], and *P boisei* explicitly with butchery[12] 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[13,14].

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[15], and also reduces the

risk of HSV2 meningitis[16]. 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[17]. Onwards transmission between humans has been reported to occur[18]. Human herpesvirus 1 can infect other primates, from gorillas[19] to owl monkeys[20], 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*[19]. 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, Ramond, Rifkin & Underdown, 2017, in press].

The high prevalence of HSV2 in central and eastern Africa (supplementary figure 1 and interactive maps at https://github.com/wadhamite/hsv-mapping) is consistent with the limited genetic data available from African HSV2 isolates. A study from Burrel and colleagues[21] showed that HSV2 can be divided into African and worldwide lineages on the basis of diversity in gene UL30. Furthemore, 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[22]. Two recent studies have significantly increased the number of whole HSV2 genomes available for analysis, contributing to our knowledge of HSV2 diversity[23,24], 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.

Methods

<u>HSV2 prevalence data, hominin fossil data and chimpanzee and tropical rainforest geographic range data</u>

We collated African hominin spatial-temporal data on hominin fossil species extant between 100 KYA and 3 MYA. Spatially, latitude and longitude of site location was used to provide a data point for each species. Temporally, dating was used to provide a first appearance datum (FAD) and last appearance datum (LAD) for each species [supplementary data]. GIS data on the range of *Pan troglodytes* and *Pan paniscus* was provided by the IUCN Red List[25]. Tropical rainforest distribution between 1.4 and 3 MYA was acquired from the Köppen-Geiger climate classification dataset[26]. This was treated as a proxy for the ancient range of chimpanzees. Data on the prevalence of herpes simplex virus 2 between 2000 and 2015

was taken from the supplementary materials of Looker[27], with additional country data (see supplementary material).

Network analysis

To establish the most probable transmission route of HSV2 from chimpanzees to modern humans, it is important to identify potential species that may have been intermediary hosts. Firstly, hominins present within the 1.4 - 3 MYA confidence window of chimp-hominin transmission were identified. Their distance to tropical rainforest was calculated: only species with fossil remains found within 400 km of tropical rainforest were considered as putative species for initial ancestral-chimp-hominin HSV2 transmission. A matrix of spatio-temporal distances was then calculated to map distances between nearest neighbours of each species and also to calculate their temporal overlap within the fossil record.

Optimal path traversal:

A network of possible transmission routes of HSV2 from ancestral-chimpanzees to humans through different potential hominims was developed as a Directed Cyclic 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. Infection Prevalence (HSV2-IP) and Infection Transmission (HSV2-IT) models assign probability value (edge costs) to the DAG utilising the temporal overlap between hominin species and their geographic proximity to rainforest habitat and one another. More details about the probability models are discussed in the next section. A Conditional Probability Table (CPT) is calculated to determine the nodal heuristics. Ancestral-chimpanzees and *Homo erectus* formed the start node and the target node of the DAG. The A* pathfinding algorithm[28] was used on the DAG to identify the most probable transmission (optimal path) route for HSV2. The A* evaluates the most probable transmission route, by minimising the traversal costs based on the edge cost and nodal

heuristics of the transmission network graph. See the supplementary section for a brief overview of the A* algorithm.

Bayesian inference:

A Bayesian network or a belief network is a graphical structure that allows us to represent and reason about an uncertain domain. The nodes are variables, and edges represent direct links between nodes. 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, which is an approximation to a hierarchical Bayesian approach is adopted[29,30].

Bayesian networks provide full representations of probability distributions over their variables, which allows to infer upon any subset of variables. 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 probable intermediary hosts, by conditioning the ancestral-chimpanzee and *Homo erectus* nodes for the presence of HSV2. A "Diagnostic reasoning"[31] was carried out to evaluate the most probable intermediate hosts for the transmission of HSV2 from ancestral-chimpanzees to *Homo sapiens.*

HSV2-Infection Prevelance (HSV2-IP) model:

This model 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. The probability density function is defined as:

PDF: $\frac{\square^{(\square-1)}(1-\square)^{(\square-1)}}{\square(\square,\square)}$ where $\square(\square,\square) = \frac{\square(\square)\square(\square)}{\square(\square+\square)}$

Where, Γ defines a gamma function, α is the time period of existence of the species in 100,000 years and β is the distance of the fossil from the rain forest in kilometers. Monte-carlo simulations were performed to sample from the distributions to populate the edge costs. A minimum overlap threshold of 10% is considered when evaluating the possibility of transmission HSV2 from one species to another.

Stochastic modelling of infection transmission

The Susceptible - Infection - Recovered (SIR)[32] is one of the most widely used model in epidemiological system. 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 of a new susceptible host is infected is modelled as an exponential distribution, which is epidemiologically incorrect for most diseases[33–35], i.e., the rate of leaving the exposed is independent of the time spent on the host. Wearing et al[33] suggested a more realistic distribution of latent and infectious periods, with a stronger central tendency:

Number 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[36,37].

Infection Transmission (HSV2-IT) model:

This model utilising the temporal overlap between hominin species and their geographic proximity to one another to determine the probability of transmission using a gamma distribution. The probability density function of a gamma distribution is given as:

$\frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$ in terms of shape alpha and rate beta

The parameters shape α is defined as the ratio of the time period in 1000 years / distance in kilometers and rate 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-chimpanzees. Monte-carlo simulations were performed to evaluate the combined probability of transmission between species is evaluated as mutually exclusive events based on the probability distributions.

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://github.com/wadhamite/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

1. Tang, J. W. *et al.* Brain stem encephalitis caused by primary herpes simplex 2 infection in a young woman. *J. Neurol. Neurosurg. Psychiatry* **74,** 1323–5 (2003).

2. Freeman, E. E. *et al.* Herpes simplex virus 2 infection increases HIV acquisition in men and women: systematic review and meta-analysis of longitudinal studies. *AIDS* **20,** 73–83 (2006).

3. Wertheim, J. O., Smith, M. D., Smith, D. M., Scheffler, K. & Kosakovsky Pond, S. L. Evolutionary origins of human herpes simplex viruses 1 and 2. *Mol. Biol. Evol.* **31,** 2356–2364 (2014).

4. Anton, S. C. *et al.* Morphological variation in Homo erectus and the origins of developmental plasticity. *Philos. Trans. R. Soc. B* **371,** 20150236 (2016).

5. Mirazón Lahr, M. *et al.* The shaping of human diversity: filters, boundaries and transitions. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **371,** 62–108 (2016).

6. Maslin, M. a, Shultz, S. & Trauth, M. H. A synthesis of the theories and concepts of early human evolution. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **370,** 1–12 (2015).

7. Ungar, P. S., Grine, F. E. & Teaford, M. F. Diet in Early Homo: A Review of the Evidence and a New Model of Adaptive Versatility. *Annu. Rev. Anthropol.* **35,** 209–228 (2006).

8. de la Torre, I., Mora, R., Domínguez-Rodrigo, M., de Luque, L. & Alcalá, L. The Oldowan industry of Peninj and its bearing on the reconstruction of the technological skills of LowerPleistocene hominids. *J. Hum. Evol.* **44,** 203–224 (2003).

9. Stern, N. *et al.* The structure of the lower pleistocene archaeological record: a case study from the Koobi fora formation [and comments and reply]. *Curr. Anthropol.* **34,** 201–225 (1993).

10. Leakey, M. G. *et al.* New fossils from Koobi Fora in northern Kenya confirm taxonomic diversity in early Homo. *Nature* **488,** 201–204 (2012).

11. De Heinzelin, J. *et al.* Environment and Behavior of 2.5-Million-Year-Old Bouri Hominids. *Science (80-.).* **284111185,** 625–629 (1999).

12. Domínguez-Rodrigo, M. *et al.* First Partial Skeleton of a 1.34-Million-Year-Old Paranthropus boisei from Bed II, Olduvai Gorge, Tanzania. *PLoS One* **8,** e80347 (2013).

13. Domínguez-Rodrigo, M., Pickering, T. R. & Bunn, H. T. Configurational approach to identifying the earliest hominin butchers. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 20929–34 (2010).

14. Domalain, M., Bertin, A. & Daver, G. 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. *Comptes Rendus Palevol* (2016). doi:10.1016/j.crpv.2016.09.003

15. Langenberg, A. G. M., Corey, L., Ashley, R. L., Leong, W. P. & Straus, S. E. A Prospective Study of New Infections with Herpes Simplex Virus Type 1 and Type 2. *N. Engl. J. Med.* **341,** 1432–1438 (1999).

16. Aurelius, E. *et al.* Long-term valacyclovir suppressive treatment after herpes simplex virus type 2 meningitis: A double-blind, randomized controlled trial. *Clin. Infect. Dis.* **54,** 1304–1313 (2012).

17. Huff, J. L. & Barry, P. A. B-virus (Cercopithecine herpesvirus 1) infection in humans and macaques: Potential for zoonotic disease. *Emerging Infectious Diseases* **9,** 246–250 (2003).

18. Centers for Disease Control (CDC). B-virus infection in humans--Pensacola, Florida. *MMWR. Morb. Mortal. Wkly. Rep.* **36,** 289–90, 295–6 (1987).

19. Gilardi, K. *et al.* Human herpes simplex virus type 1 in confiscated gorilla. *Emerg. Infect. Dis.* **20,** 1883–1886 (2014).

20. Melendez, L. V, España, C., Hunt, R. D., Daniel, M. D. & Garcia, F. G. Natural herpes simplex infection in the owl monkey (Aotus trivirgatus). *Lab. Anim. Care* **19,** 38–45 (1969).

21. Burrel, S. *et al.* Genetic Diversity within Alphaherpesviruses: Characterization of a Novel Variant of Herpes Simplex Virus 2. *J. Virol.* **89,** 12273–12283 (2015).

22. Burrel, S. *et al.* Ancient recombination events between human herpes simplex viruses. *bioRxiv* doi:10.1101/093641

23. Szpara, M. L. *et al.* Evolution and Diversity in Human Herpes Simplex Virus Genomes. *J. Virol.* **88,** 1209–1227 (2014).

24. Kolb, A. W., Larsen, I. V, Cuellar, J. a & Brandt, C. R. Genomic, Phylogenetic, and Recombinational Characterization of Herpes Simplex Virus 2 Strains. *J. Virol.* **89,** 6427–6434 (2015).

25. Oates, J. F. *et al.* Pan troglodytes (Chimpanzee, Common Chimpanzee, Robust Chimpanzee). *The IUCN Red List of Threatened Species* (2008). Available at: http://www.iucnredlist.org/details/15933/0. (Accessed: 2nd February 2017)

26. Peel, B. L., Finlayson, B. L. & McMahon, T. a. Updated world map of the Koppen-Geiger climate classification.pdf. *Hydrol. Earth Syst. Sci.* **11,** 1633–1644 (2007).

27. Looker, K. J. *et al.* Global estimates of prevalent and incident herpes simplex virus type 2 infections in 2012. *PLoS One* **10,** (2015).

28. Hart, P., Nilsson, N. & Raphael, B. A Formal Basis for the Heuristic Determination of Minimum Cost Paths. *IEEE Trans. Syst. Sci. Cybern.* **4,** 100–107 (1968).

29. Murphy, K. P. Machine Learning: A Probabilistic Perspective. *MIT Press* 25 (2012). doi:10.1007/978-3-642-21004-4_10

30. Farine, D. R. & Strandburg-Peshkin, A. Estimating uncertainty and reliability of social network data using Bayesian inference. *R. Soc. open Sci.* **2,** 150367 (2015).

31. Korb, K. B. & Ann E. Nicholson. *Bayesian Artificial Intelligence*. *CRC* (2003). doi:doi:10.1201/b10391-4

32. Callahan, J. The spread of a contagious illness. *Smith Coll.*

33. Wearing, H. J., Rohani, P. & Keeling, M. J. Appropriate models for the management of infectious diseases. *PLoS Med.* **2,** e174 (2005).

34. Bailey, N. T. J. *The mathematical theory of infectious diseases and its applications. 2nd edition*. *The mathematical theory of infectious diseases and its applications 2nd edition* **413,** (1975).

35. Sartwell, P. E. & others. The distribution of incubation periods of infectious disease. *Am. J. Hyg.* **51,** 310–318 (1950).

36. Blythe, S. P. & Anderson, R. M. Distributed incubation and infectious periods in models of the transmission dynamics of the human immunodeficiency virus (HIV). *IMA J. Math. Appl. Med. Biol.* **5,** 1–19 (1988).

37. Lloyd, A. L. Destabilization of epidemic models with the inclusion of realistic distributions of infectious periods. *Proc. R. Soc. London B.* **268,** 985–993 (2001).