Controlled fire use in early humans might have triggered the evolutionary emergence of tuberculosis

Rebecca H. Chisholm[1,2], James M. Trauer[3], Darren Curnoe[4], Mark M. Tanaka[1,2]

[1] School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney 2052, Australia
[2] Evolution & Ecology Research Centre, University of New South Wales, Sydney 2052, Australia
[3] School of Public Health and Preventive Medicine, Monash University, Melbourne 3004, Australia
[4] Palaeontology, Geobiology and Earth Archives Research Centre, University of New South Wales, Sydney 2052, Australia

SI Appendix

SI Appendix A

We consider the change in dynamics of our model when we allow for genetic predisposition to mycobacterial disease. Here we add the assumption that exposure to the pathogen over time decreases the susceptibility of the human population to infection due to an increase in frequency of protective alleles against intracellular infection [1].

Assume that a protective allele that confers partial resistance can become fixed in the host population between pathogen introduction events. If it takes on average $1/r$ introductions of the pathogen into the human population from the environmental reservoir for an allele to become fixed in the host population, then the number X of introductions required until fixation is a geometrically distributed random variable with parameter $1/r$ and probability mass function

$$\Pr(X = x) = (1 - r)^{x-1} r$$

where $x = 1, 2, \ldots$. When fixation occurs, we assume the probability of emergence after a single introduction reduces to p_{r} where $p_{\mathrm{r}} < p_{\mathrm{emerge}}$.

We are interested in how many introductions are required for emergence to occur. Therefore we consider the process where, if the protective allele is not yet fixed in the host population, a single introduction will result in emergence with probability p_{emerge}. If emergence does not occur, then with probability r the protective allele will become fixed in the population and the probability of emergence will reduce to p_{r}. In this scenario, the number Y of introductions required for emergence to occur is a random variable with probability mass function:

$$\begin{aligned}\Pr(Y = y) &= \sum_{x=1}^{\infty} \Pr(Y = y | X = x) \Pr(X = x) \\ &= \underbrace{\sum_{x=1}^{y-1} (1 - p_{\mathrm{emerge}})^x (1 - p_{\mathrm{r}})^{y-x-1} p_{\mathrm{r}} \Pr(X = x)}_{\text{emergence occurs after fixation}} + \underbrace{\sum_{x=y}^{\infty} (1 - p_{\mathrm{emerge}})^{y-1} p_{\mathrm{emerge}} \Pr(X = x)}_{\text{emergence occurs before fixation}}.\end{aligned}$$

After some algebra, it can be shown that the mean value of Y is given by the expression

$$\mathbb{E}(Y) = \sum_{y=1}^{\infty} y \Pr(Y = y) = \frac{1}{p_{\text{emerge}}} \left[1 + \left(\frac{p_{\text{emerge}} - p_{\text{r}}}{p_r} \right) \left(1 - \frac{p_{\text{emerge}}}{p_{\text{emerge}} + r(1 - p_{\text{emerge}})} \right) \right]. \quad (1)$$

Therefore, if host susceptibility changes at a much slower rate than pathogen emergence ($r \ll p_{\text{emerge}}$), the expected value approaches that from the original model where host susceptibility remains constant, *i.e.*,

$$\mathbb{E}(Y) \approx 1/p_{\text{emerge}}, \quad r \ll p_{\text{emerge}}.$$

If, on the other hand, host susceptibility changes at a much faster rate than pathogen emergence ($r \gg p_{\text{emerge}}$), then

$$\mathbb{E}(Y) \approx 1/p_{\text{r}}, \quad r \gg p_{\text{emerge}}.$$

Simulations of this process shown in Fig. S5 illustrate these results.

SI Appendix Figures

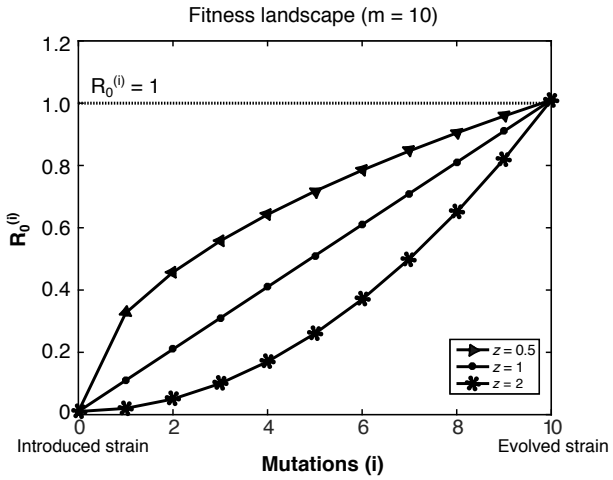

Fig. S1: Fitness landscapes that govern the evolution of introduced strains in our model of infectious disease emergence. Here, ten mutations are required for the introduced strain to evolve an $R_0^{(i)} > 1$ ($m = 10$).

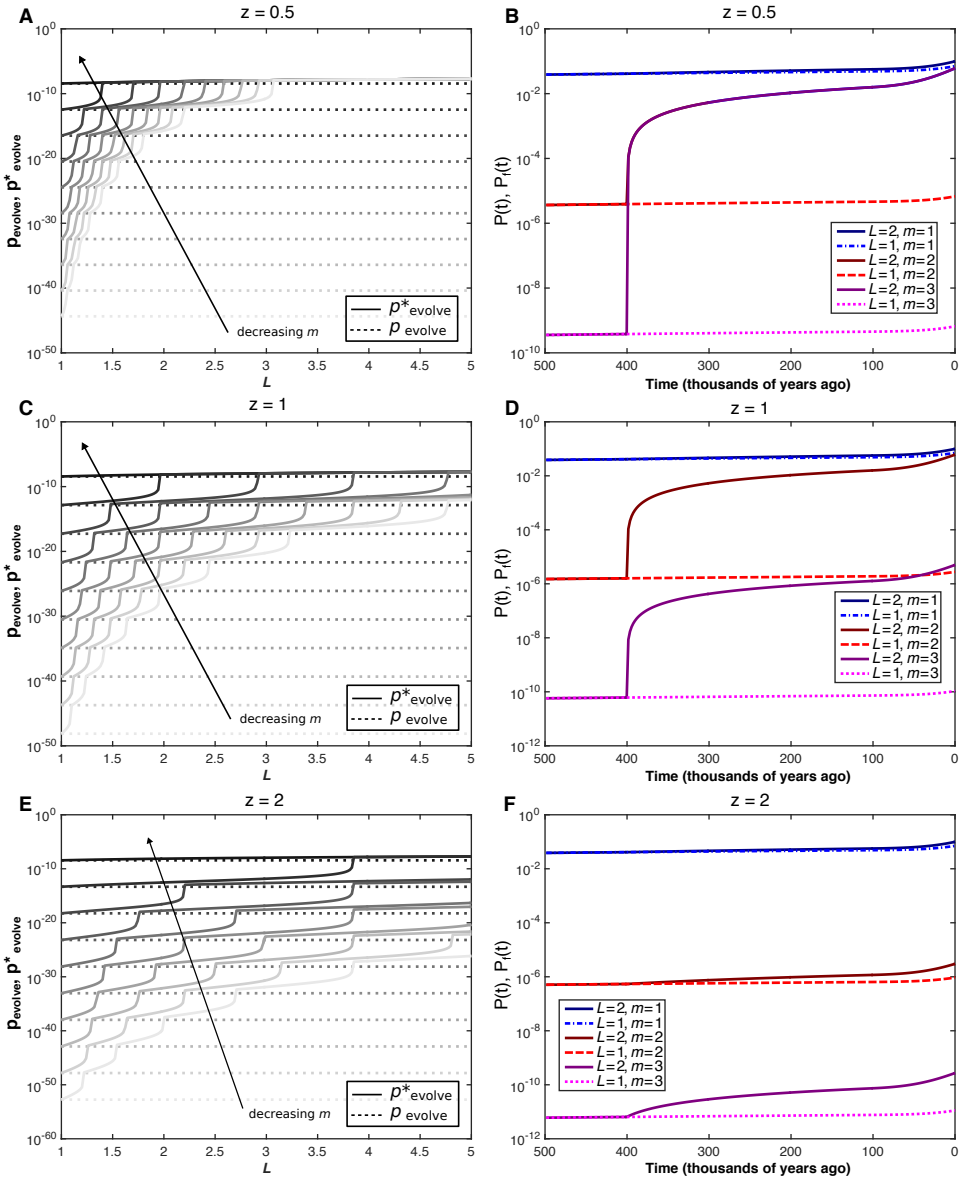

Fig. S2: Model quantities that determine the cumulative probability of TB emergence through time. The number of mutations required for an introduced strain to evolve an $R_0 > 1$ (m), the shape of the fitness landscape (z) and the factor of increase in the basic reproductive number due to fire use (L) influence the probabilities p_{evolve} and p^*_{evolve} an introduced strain evolves an $R_0 > 1$ and the cumulative probabilities $P(t)$ and $P_{\text{f}}(t)$ of TB emergence through time. In Panels A, C and E, dotted lines correspond to p_{evolve} and solid lines represent p^*_{evolve}, while increasingly light coloured lines represent p_{evolve} and p^*_{evolve} calculated with increasing values of $m = 1, 2, \ldots, 10$. Notice that p_{evolve} is constant with respect to L. The steep increases in the probabilities p^*_{evolve} which lead to the jumps seen in the cumulative probabilities $P_{\text{f}}(t)$ (Panels B, D, F) occur for a given value of m when L becomes sufficiently high to cause a reduction in the number of mutations required for R_0 to become greater than unity. This reduces m^* and the number of terms in Eq. (6) by one each time. In Panels B, D and F, solid lines correspond to $P_{\text{f}}(t)$ with $L = 2$, broken lines to $P(t)$ (*i.e.*, $L = 1$), while colours indicates the value of m: blue for $m = 1$, red for $m = 2$, and purple for $m = 3$. The values of unspecified parameters are provided in Table S1.

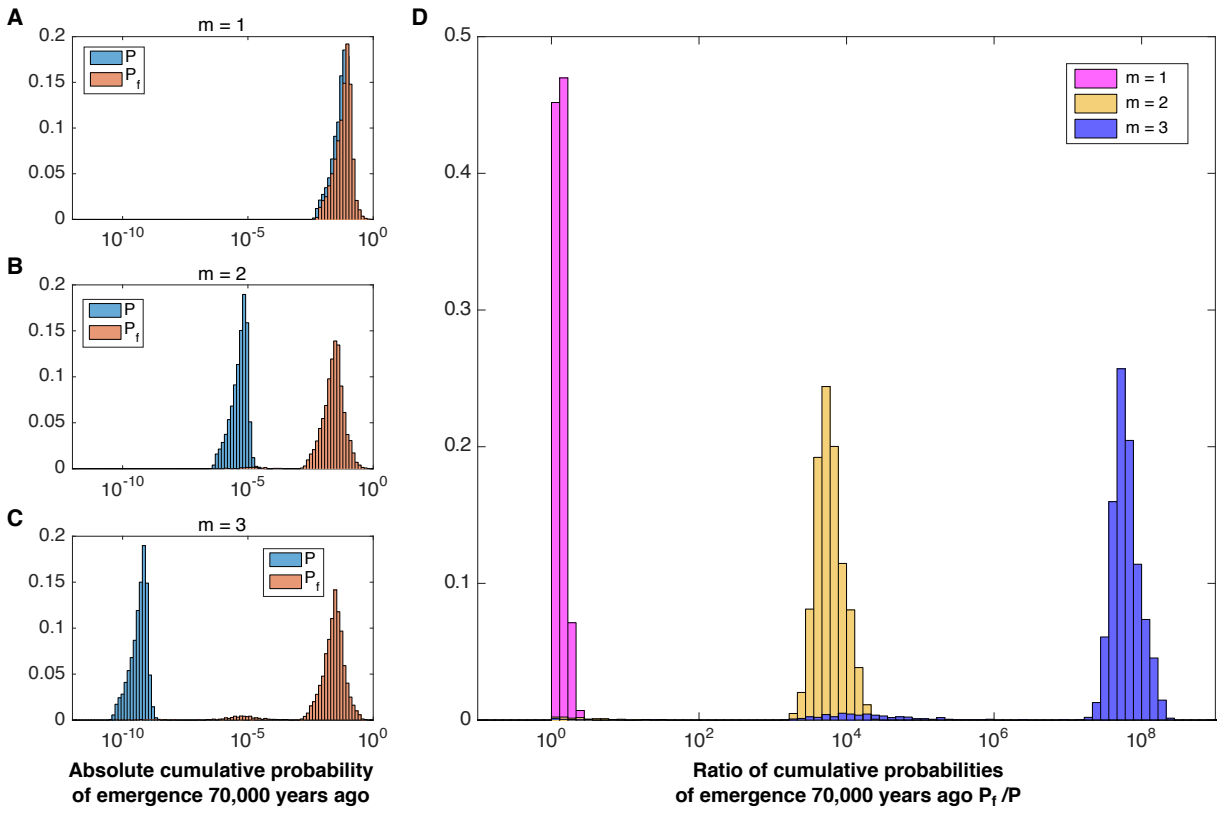

Fig. S3: Sensitivity analysis of the cumulative probability of TB emergence 70,000 years ago. (A)–(C) The frequency distributions of the absolute cumulative probabilities P_f (blue) and P (red) of the emergence of MTBC 70,000 years ago for $z = 0.5$ and $m = 1, 2, 3$. This data determines the frequency distributions of the ratio P_f/P of cumulative probabilities of the emergence of MTBC 70,000 years ago shown in Panel D.

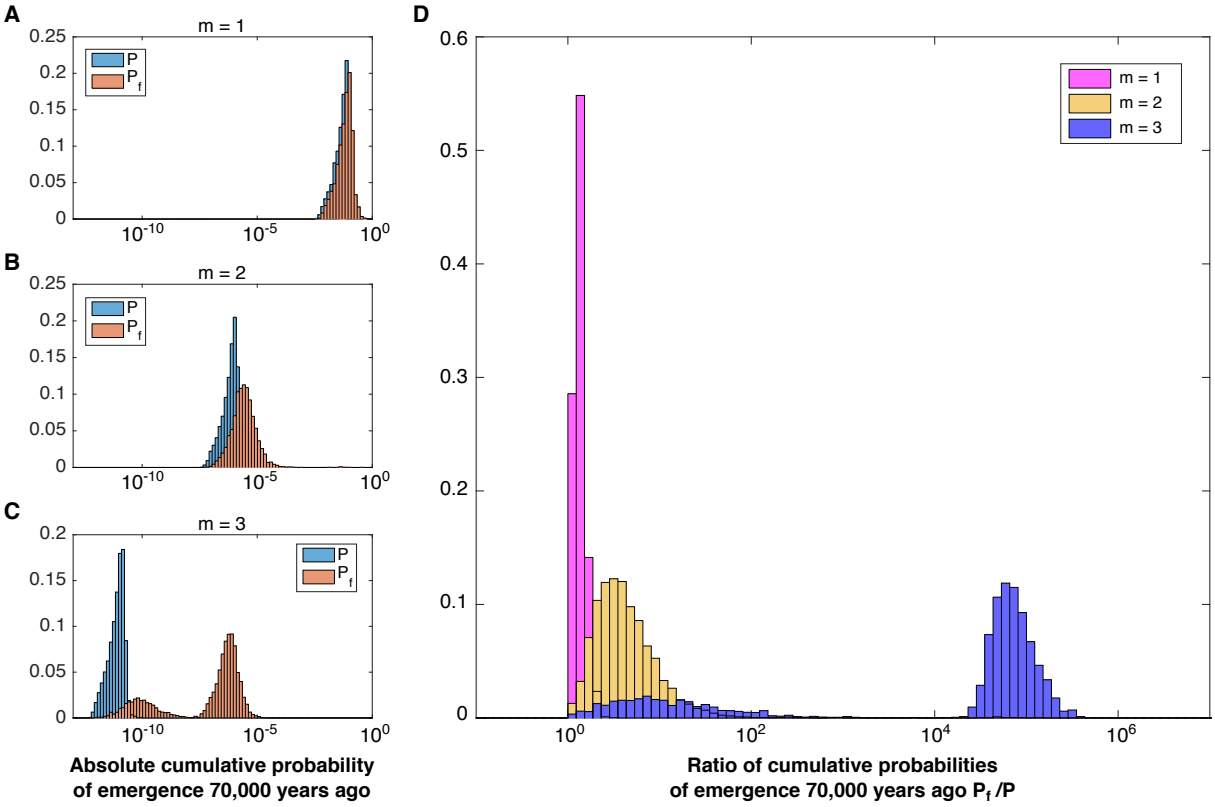

Fig. S4: Sensitivity analysis of the cumulative probability of TB emergence 70,000 years ago. (A)–(C) The frequency distributions of the absolute cumulative probabilities P_f (blue) and P (red) of the emergence of MTBC 70,000 years ago for $z = 2$ and $m = 1, 2, 3$. This data determines the frequency distributions of the ratio P_f/P of cumulative probabilities of the emergence of MTBC 70,000 years ago shown in Panel D.

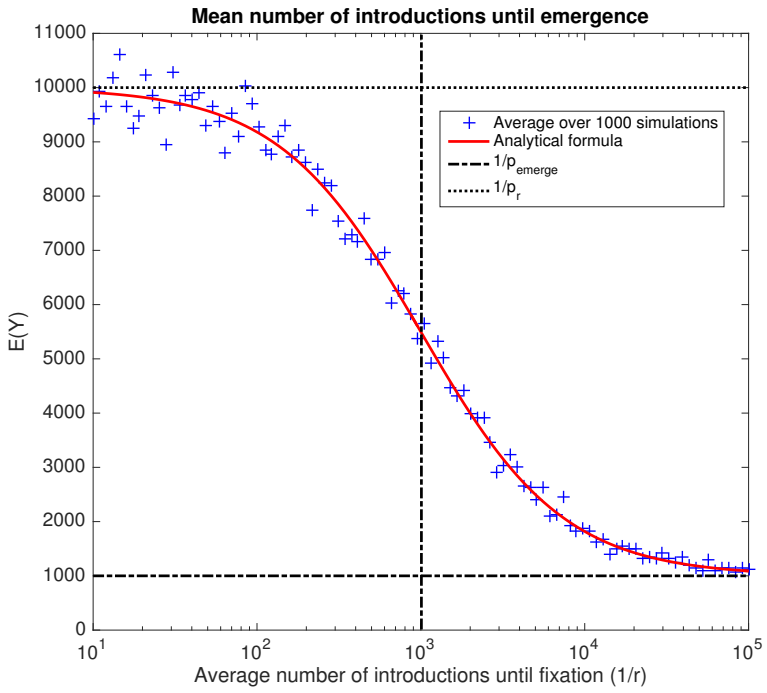

Fig. S5: The expected number of introductions required for emergence $\mathbb{E}(Y)$ as a function of the mean number $(1/r)$ of introductions until fixation of a protective allele conferring partial immunity to the host population as determined by Eq. (1) (solid red line) and averaged over 1000 simulations of the discrete stochastic process outlined above in SI Appendix A (blue crosses) when $p_{\text{emerge}} = 10^{-3}$, $p_{\text{r}} = 10^{-4}$ and $1/r \in [10^1, 10^5]$. Here, the dash-dot lines indicate $1/r = 1/p_{\text{emerge}}$ and $\mathbb{E}(Y) = 1/p_{\text{emerge}}$ and the dotted black line indicates $\mathbb{E}(Y) = 1/p_{\text{r}}$. For $r \gg p_{\text{emerge}}$ (or equivalently $1/r \ll 1/p_{\text{emerge}}$), $\mathbb{E}(Y) \to 1/p_{\text{r}}$ while for $r \ll p_{\text{emerge}}$ (or equivalently $1/r \gg 1/p_{\text{emerge}}$), $\mathbb{E}(Y) \to 1/p_{\text{emerge}}$.

SI Appendix Tables

Table S1: **Model parameters**

Symbol	Parameter name	Value	Reference
$R_0^{(0)}$	Basic reproductive number of introduced pathogen	0.01	—
$R_0^{(m)}$	Basic reproductive number of evolved pathogen	1.01	—
μ	Mutation rate per year (mutation rate $\times$ generation time $\times 10^3$)	3.65×10^{-5}	[2]
C	Disease prevalence (cases per capita per year)	7.2×10^{-5}	[3]
N_0	Census population size for hunter gatherers (effective size $\times$ 10)	1×10^5	[4]
α	Host growth rate per generation (25 years) during population expansion	8.5×10^{-4}	[5]
t_e	Onset time of population expansion (years before present)	10^5	[4]
t_f	Onset time of widespread controlled fire use (years before present)	4×10^5	[6]
L	Factor of increase in basic reproductive number due to fire use	2	[7]

Table S2: **Sample distributions used in the sensitivity analysis**

Parameter	Distribution	Reference
$R_0^{(0)}$	U(0.001,0.02)*	—
N_0	$\mathcal{N}(10^5, 1.5 \times 10^4)^\dagger$	[4]
α	$\mathcal{N}(7 \times 10^{-4}, 1.5 \times 10^{-4})$	[5]
t_e	U(200,000 years ago, 20,000 years)	[4]
t_f	$\mathcal{N}$(400,000 years ago, 66,000 years)	[6]
L	$\mathcal{N}(2.5, 0.5)$	[7]

* $U(a, b)$ is the uniform distribution with lower bound a and upper bound b.

† $\mathcal{N}(c, d)$ is the normal distribution with mean c and standard deviation d.

References

[1] Barnes I, Duda A, Pybus OG, Thomas MG (2011) Ancient urbanization predicts genetic resistance to tuberculosis. *Evolution* 65(3):842–848.

[2] Ford CB et al. (2013) *Mycobacterium tuberculosis* mutation rate estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis. *Nat. Genet.* 45(7):784–790.

[3] Cassidy PM, Hedberg K, Saulson A, McNelly E, Winthrop KL (2009) Nontuberculous mycobacterial disease prevalence and risk factors: a changing epidemiology. *Clin. Infect. Dis.* 49(12):e124–e129.

[4] Kim HL et al. (2014) Khoisan hunter-gatherers have been the largest population throughout most of modern-human demographic history. *Nat. Commun.* 5:5692.

[5] Cox MP et al. (2009) Autosomal resequence data reveal Late Stone Age signals of population expansion in sub-Saharan African foraging and farming populations. *PLoS ONE* 4(7):e6366–e6366.

[6] Roebroeks W, Villa P (2011) On the earliest evidence for habitual use of fire in Europe. *Proc. Natl. Acad. Sci. U.S.A.* 108(13):5209–5214.

[7] Lin H, Ezzati M, Murray M (2007) Tobacco smoke, indoor air pollution and tuberculosis: a systematic review and meta-analysis. *PLoS Med.* 4(1):e20.